



# Mass and Age from FLAME in Gaia DR3: comparison with an external code and quantification of a solar-metallicity assumption

---

prepared by: Creevey, Orlagh L., Lebreton, Yveline  
approved by: Rosanna Sordo  
reference: GAIA-C8-TN-OCA-OLC-036  
issue: 1  
revision: 0  
date: 2022-11-30  
status: I

## Abstract

The objective of this technical note is to quantify the performance of the CU8/Apsis/FLAME software which published masses and ages in Gaia Data Release 3. We compare the results from the FLAME software with those from the publically available code *SPInS*. We also test the robustness of retrieving ages and masses by running on simulated data. This latter includes testing and quantifying the impact of the solar metallicity prior that was imposed during DR3 operations. We additionally propose some empirical corrections to the published parameters to account for non-solar metallicity.

## Document History

| Issue | Revision | Date       | Author     | Comment                       |
|-------|----------|------------|------------|-------------------------------|
| D     | 0        | 2020-08-20 | OLC        | Start TN and work             |
| D     | 1        | 2021-04-18 | OLC/<br>YL | Sections 1 - 4                |
| D     | 2        | 2022-05-22 | OLC        | Section 5                     |
| D     | 3        | 2022-11-22 | OLC        | Feedback from coauthor and RS |

# Contents

|          |  |           |
|----------|--|-----------|
| <b>1</b> | <b>FLAME</b>   | <b>5</b>  |
| <b>2</b> | <b>Objectives</b>  | <b>5</b>  |
| <b>3</b> | <b>Assumptions &amp; Methods</b>                             | <b>6</b>  |
| 3.1      | Assumptions and Models . . . . .                             | 6         |
| 3.2      | SPInS . . . . .  | 7         |
| <b>4</b> | <b>Performance on test data</b>                              | <b>7</b>  |
| 4.1      | Generation of test data . . . . .                            | 7         |
| 4.2      | Results . . . . .  | 9         |
| 4.2.1    | SampleA_0 . . . . .  | 9         |
| 4.2.2    | SampleA_1 . . . . .  | 10        |
| 4.2.3    | SampleA_2 . . . . .  | 11        |
| 4.2.4    | SampleA_5 . . . . .  | 12        |
| 4.3      | Conclusions . . . . .  | 12        |
| <b>5</b> | <b>Comparison with an external code using real Gaia data</b> | <b>13</b> |
| 5.1      | Input Data . . . . .   | 13        |
| 5.2      | Comparison of results . . . . .                              | 13        |
| 5.2.1    | Mass . . . . .   | 14        |
| 5.2.2    | Age . . . . .  | 15        |
| 5.3      | Conclusions of test . . . . .                                | 15        |
| <b>6</b> | <b>Impact of solar metallicity prior on mass and age</b>     | <b>19</b> |
| 6.1      | Datasets . . . . .   | 19        |
| 6.2      | Results per metallicity . . . . .                            | 19        |
| 6.3      | Results per inferred parameter . . . . .                     | 28        |
| 6.3.1    | Mass . . . . .   | 28        |
| 6.3.2    | Evolution Index . . . . .                                    | 32        |
| 6.3.3    | Age . . . . .  | 35        |
| 6.4      | Conclusions . . . . .  | 35        |
| <b>7</b> | <b>Conclusions</b>   | <b>38</b> |
| <b>8</b> | <b>Acknowledgements</b>                                      | <b>38</b> |

## 1 FLAME

The Final Luminosity Mass Age Estimator FLAME is a CU8/Apsis software running at DPCC (CNES) which is used for the production of data for Gaia catalogues. It was first used in Gaia DR2 where it produced a catalogue of radii and luminosities for around 80 million stars, (Andrae et al. 2018, LL:GAIA-CU8-TN-OCA-OLC-001,LL:GAIA-CU8-TN-OCA-OLC-003-1). For Gaia DR3 it produced radii, luminosities, gravitational redshift, masses, ages and evolution index, along with two auxiliary data products: a bolometric correction and flags (Creevey et al., 2022; Fouesneau et al., 2022; Delchambre et al., 2022). Table 1 summarises the products in Gaia DR3. The results are found in two tables; 280 M results based on the GSP-Phot input parameters are found in the `gaiadr3.astrophysical_parameters` tables, while 5 M results based on the input parameters from GSP-Spec are found in the `gaiadr3.astrophysical_parameters_supp` table.

FLAME is divided into two modules: `Algo1` and `Algo2`. `Algo1` takes upstream data to produce luminosity  $\mathcal{L}$ , radius  $\mathcal{R}$ , and gravitational redshift  $rv_{GC}$ . `Algo2` uses the upstream  $\mathcal{T}_{\text{eff}}$  and  $[M/H]$  along with  $\mathcal{L}$  from `Algo1` to derive the mass  $\mathcal{M}$ , age  $\tau$ , and evolutionary stage  $\epsilon$ .

`Algo2` involves the comparison of the observed data  $\mathcal{L}$ ,  $\mathcal{T}_{\text{eff}}$ , and  $[M/H]$  with stellar models. It is based on a bootstrap Monte Carlo method where the input observations are perturbed based on the input uncertainties  $N$  times. The best-match for each  $i$  iteration is obtained through a  $\chi^2$  minimization process. The result is a distribution of  $N$  parameters. Using a prior  $P$  as a weight which depends on the position in the HR diagram, we calculate the 16, 50, and 84 percentile values of the stellar parameters from the weighted distribution. These are used to denote the lower confidence interval, the value of the parameter, and the upper confidence interval, respectively.

FLAME has been led and developed by Orlagh Creevey and Christophe Ordenovic since 2014, in collaboration with the FLAME team: Yveline Lebreton, Frédéric Thévenin, Bernard Pichon. FLAME was originally conceived in the early 2000s and was led by Yveline Lebreton until 2010, see Bailer-Jones (ICAP-CBJ-021) and CBJ-033, with the first version of the algorithms provided and implemented by Christophe Ordenovic in 2010.

## 2 Objectives

The objective of this technical note is to provide further scientific validation of the model parameters inferred using the FLAME software which appeared in Gaia Data Release 3. The results from FLAME validation on Gaia DR3 have been published in several internal documents (OLC-032, OLC-032; MFX-026, MFX-026), and in the [official online documentation](#), and DR3 Apsis papers (see above), and these are complementary to this technical note.

In this technical note we focus on the mass, age and evolutionary parameters that were published in Gaia DR3, see Table 1. In Section 3 we describe the assumptions and models, and give a brief overview of the method used in SPInS. We then address each of the following objectives:

Section 4 We test FLAME results on simulated data using the Gaia DR3 configuration

Section 5 We compare the output of FLAME with the publically available SPInS code using Gaia data (Lebreton & Reese, 2020)

Section 6 We test and quantify the impact of the solar metallicity prior imposed during operations.

### 3 Assumptions & Methods

In this section we describe the assumptions in the models used, along with a summary of the methods used in SPInS.

#### 3.1 Assumptions and Models

The FLAME software relies on the use of published stellar evolution models and isochrones. The BASTI grid of stellar evolution models (Hidalgo et al., 2018) were implemented in the FLAME software and details of their implementation is described in OLC-016.

Upon validation of FLAME masses and ages during the qualification stage before Operations, it was decided to impose a solar metallicity prior, see validation FLAME reports on ApsisVal3.5g

| gaiadr3.   | field                     |
|--|---------------------------|
| astrophysical_parameters<br>(based on GSP-Phot APs)      | radius_flame              |
|  | lum_flame                 |
|  | gravredshift_flame        |
|  | mass_flame                |
|  | age_flame                 |
|  | evolstage_flame           |
| astrophysical_parameters_supp<br>(based on GSP-Spec APs) | radius_flame_spec         |
|  | lum_flame_spec            |
|  | gravredshift_flame_spec   |
|  | mass_flame_spec           |
|  | age_flame_spec            |
|  | evolstage_flame_spec_spec |

TABLE 1: Evolutionary parameters from FLAME in Gaia DR3

OLC-030 and ApsisOPS3.2 OLC-032. In this TN we therefore restrict the analysis in Sections 4 and 5 to solar metallicity model assumptions. Given that the vast majority of the stars peak at solar metallicity, this assumption is acceptable, and see Sect. 6.3.3 for validity tests. This is properly documented in the archive datamodel, and has been reported in the published articles on Gaia DR3, see Creevey et al. (2022); Fouesneau et al. (2022).

The SPInS code runs more slowly than FLAME. All of our tests are therefore restricted to subgroups of approximately 600 stars, which provided a good balance between run time and statistics. Even though the comparison with SPInS is only in Section 5, we use the same sample size throughout this note to construct our simulated data.

### 3.2 SPInS

SPInS standing for “Stellar Parameters INferred Systematically” is a public tool that can be used to estimate stellar parameters and credible intervals/error bars. From a set of observational data (effective temperature, luminosity, metallicity, etc.) and associated probability distributions, SPInS searches in a precalculated grid of stellar models and applies an MCMC algorithm to infer stellar mass and age as well as other stellar parameters. SPInS can provide the mean, mode and median value of the parameters. In the present work we used the mean of the distribution to define the parameter value, and used the distribution of the results to define the error bars.

SPInS has been presented in detail in Lebreton & Reese (2020). It can be downloaded at <https://gitlab.obspm.fr/dreese/spins>.

## 4 Performance on test data

The general performance of FLAME is tested using simulated data based on the stellar models that are used for the inference (BaSTI).

### 4.1 Generation of test data

We used the BaSTI stellar models to create a set of simulated data. As we focus just on the performance in Gaia DR3, we therefore draw the samples from the solar metallicity models. This corresponds to the solar-scaled models with  $[\text{Fe}/\text{H}] = +0.06$ . We also restrict the stellar masses here to between 0.75 and  $10.0 M_{\odot}$ . The upper limit corresponds to the upper limit of the input models used in FLAME. The lower limit corresponds to stars with  $T_{\text{eff}}$  roughly above 4000 K. Below this value the stars are not expected to pass the subgiant phase in the assumed maximum 14 billion years for the age of the Universe. A focussed study could be performed on the lower mass stars ignoring their age determinations.

The data were generated by selecting randomly from the input models:

- 300 main sequence stars, denoted 'ms' in subsequent tables
- 100 subgiant stars, denoted as 'sg'
- 200 red giant stars, denoted as 'gt'

Then we created mock datasets by considering different simulation error levels – 0%, 1%, 2% and 5% uncertainties in  $\mathcal{T}_{\text{eff}}$  and  $\mathcal{L}$ . The mock datasets contain also  $\text{gFlux}$  as this is required on input for FLAME. For this we assumed a distance of 1000 pc, along with no extinction and no bolometric correction. These assumptions were done to isolate the results on the performance of the inference method. The mock datasets are named SampleA\_0.txt, SampleA\_1.txt, SampleA\_2.txt and SampleA\_5.txt. The distribution of the targets in the HR diagram is shown in Fig. 1. Colour-code indicates number of stars with the brighter colour indicating a higher number of stars.

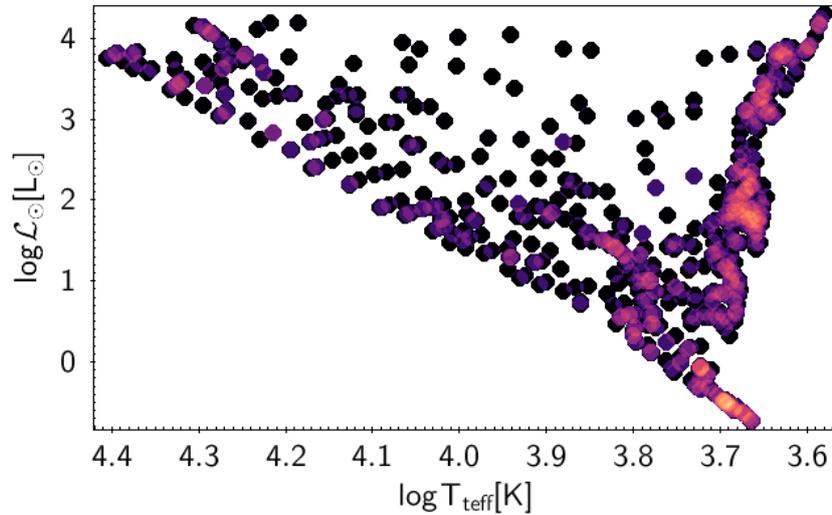


FIGURE 1: The HR diagram showing the sample of test stars, drawn randomly from the BASTI models considering their evolutionary stage.

In order to run these mock data sets through FLAME, a few minor configuration parameters needed to be set. The parallax was set to 1.0 mas and its error to  $1. \cdot 10^{-5}$  mas;  $A_G$  was set to  $1. \cdot 10^{-5}$  with the lowerValue = 0.0 and upperValue =  $2. \cdot 10^{-5}$ ; The input  $\mathcal{T}_{\text{eff}}$  uncertainty had to be converted to a lower and upper value. These minor modifications would have negligible effects on the results. Additionally, the cancelBc and cancelAg properties were set to True in the FLAME configuration file to remove the call to correct the luminosities for the bolometric correction and extinction (which were set to zero in the input data).

## 4.2 Results

In the following subsections and tables we separate the results into 'ms', 'sg' and 'gt'. We also report results for the three parameters  $P = \{M, \tau, \epsilon\}$ , where these are mass, age, and evolutionary stage, respectively. For the evolutionary stage which is an index running from 100 – 1200, we quantify the difference in absolute value.

We report the median of  $\Delta P = P_{\text{FLAME}} - P_{\text{input}}$ , the median absolute deviation of  $\Delta P$ ,  $\Delta P.MAD$ , the relative difference in terms of the true parameter  $\Delta P/P$ , and the relative difference in terms of the derived uncertainty  $\Delta P/\sigma(P)$ .

### 4.2.1 SampleA\_0

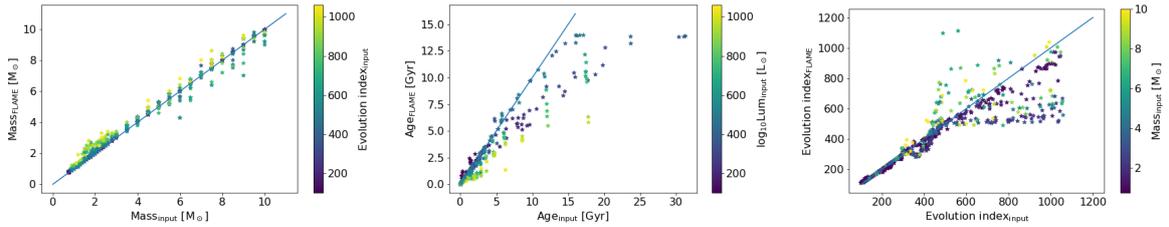


FIGURE 2: Comparison of input and output  $\mathcal{M}$  (left),  $\tau$  (middle),  $\epsilon$  (right) for SampleA\_0.

| $P$        | Stage | $\Delta P$ | $\Delta P.MAD$ | $\Delta P/P$ | $\Delta P/\sigma(P)$ |
|------------|-------|------------|----------------|--------------|----------------------|
| $M$        | ms    | 0.001      | 0.002          | 0.000        | 0.057                |
|            | sg    | 0.073      | 0.091          | 0.033        | 2.985                |
|            | gt    | 0.201      | 0.206          | 0.092        | 2.284                |
| $\tau$     | ms    | -0.000     | 0.003          | -0.004       | -0.083               |
|            | sg    | -0.030     | 0.044          | -0.091       | -3.644               |
|            | gt    | -0.158     | 0.169          | -0.234       | -3.280               |
| $\epsilon$ | ms    | 0.8        |                |              |                      |
|            | sg    | 1.1        | 29.7           |              |                      |
|            | gt    | -84.0      | 87.6           |              |                      |

TABLE 2: Median differences in stellar parameters  $\Delta P = P_{\text{FLAME}} - P_{\text{input}}$ , along with relative differences  $\Delta P/P$ , and in terms of  $\sigma_{\text{FLAME}}$  for SampleA\_0.

## 4.2.2 SampleA\_1

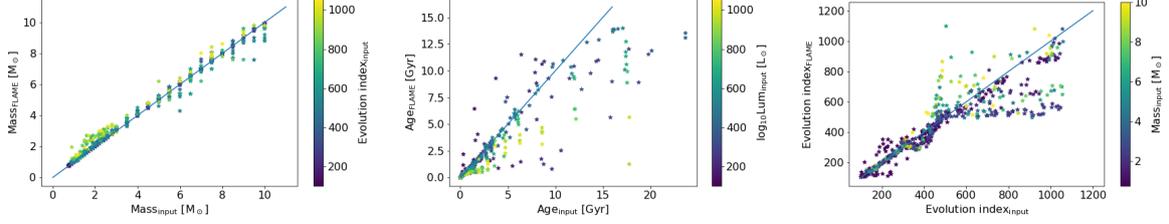


FIGURE 3: Comparison of input and output  $\mathcal{M}$  (left),  $\tau$  (middle),  $\epsilon$  (right) for SampleA\_1, colour-coded by different stellar parameters.

| $P$        | Stage | $\Delta P$ | $\Delta P.MAD$ | $\Delta P/P$ | $\Delta P/\sigma(P)$ |
|------------|-------|------------|----------------|--------------|----------------------|
| $M$        | ms    | 0.006      | 0.017          | 0.003        | 0.442                |
|            | sg    | 0.043      | 0.076          | 0.024        | 1.872                |
|            | gt    | 0.187      | 0.187          | 0.067        | 2.278                |
| $\tau$     | ms    | -0.004     | 0.019          | -0.039       | -0.526               |
|            | sg    | -0.016     | 0.040          | -0.067       | -2.086               |
|            | gt    | -0.123     | 0.136          | -0.150       | -3.091               |
| $\epsilon$ | ms    | -1.4       |                |              |                      |
|            | sg    | 2.1        | 36.8           |              |                      |
|            | gt    | -75.5      | 76.2           |              |                      |

TABLE 3: Median differences in stellar parameters  $\Delta P = P_{\text{FLAME}} - P_{\text{input}}$ , along with relative differences  $\Delta P/P$ , and in terms of  $\sigma_{\text{FLAME}}$  for SampleA\_1.

## 4.2.3 SampleA\_2

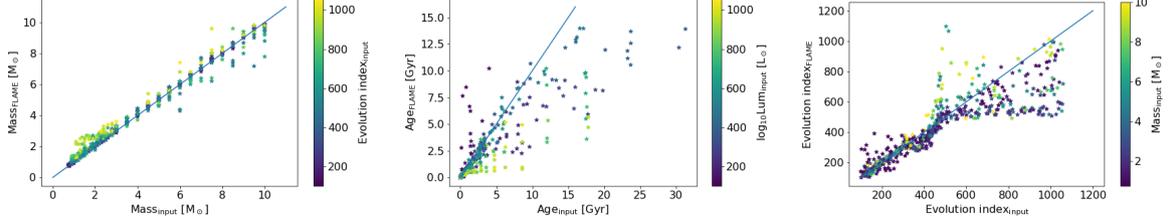


FIGURE 4: Comparison of input and output  $\mathcal{M}$  (left),  $\tau$  (middle),  $\epsilon$  (right) for SampleA\_2, colour-coded by different stellar parameters.

| $P$        | Stage | $\Delta P$ | $\Delta P.MAD$ | $\Delta P/P$ | $\Delta P/\sigma(P)$ |
|------------|-------|------------|----------------|--------------|----------------------|
| $M$        | ms    | 0.013      | 0.027          | 0.006        | 0.529                |
|            | sg    | 0.084      | 0.094          | 0.035        | 1.852                |
|            | gt    | 0.205      | 0.225          | 0.069        | 2.024                |
| $\tau$     | ms    | -0.006     | 0.038          | -0.051       | -0.611               |
|            | sg    | -0.029     | 0.049          | -0.111       | -2.251               |
|            | gt    | -0.085     | 0.141          | -0.176       | -2.636               |
| $\epsilon$ | ms    | -2.8       |                |              |                      |
|            | sg    | -5.9       | 32.5           |              |                      |
|            | gt    | -115.9     | 105.7          |              |                      |

TABLE 4: Median differences in stellar parameters  $\Delta P = P_{\text{FLAME}} - P_{\text{input}}$ , along with relative differences  $\Delta P/P$ , and in terms of  $\sigma_{\text{FLAME}}$  for SampleA\_2.

#### 4.2.4 SampleA\_5

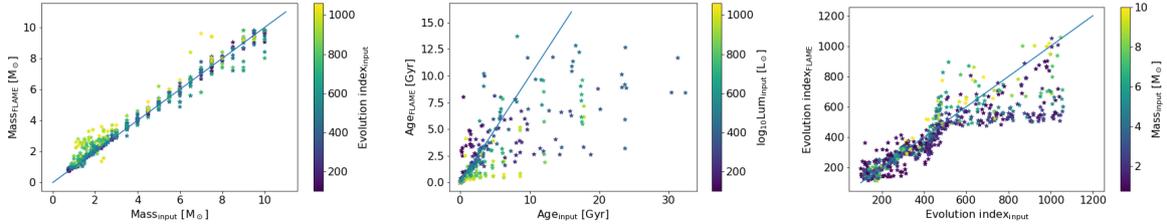


FIGURE 5: Comparison of input and output  $\mathcal{M}$  (left),  $\tau$  (middle),  $\epsilon$  (right) for SampleA\_5, colour-coded by different stellar parameters.

| $P$        | Stage | $\Delta P$ | $\Delta P.MAD$ | $\Delta P/P$ | $\Delta P/\sigma(P)$ |
|------------|-------|------------|----------------|--------------|----------------------|
| $M$        | ms    | 0.015      | 0.047          | 0.007        | 0.387                |
|            | sg    | 0.098      | 0.103          | 0.038        | 1.826                |
|            | gt    | 0.220      | 0.297          | 0.092        | 2.094                |
| $\tau$     | ms    | -0.009     | 0.055          | -0.064       | -0.490               |
|            | sg    | -0.050     | 0.057          | -0.115       | -2.017               |
|            | gt    | -0.107     | 0.136          | -0.212       | -3.041               |
| $\epsilon$ | ms    | -2.8       |                |              |                      |
|            | sg    | -12.2      | 39.7           |              |                      |
|            | gt    | -122.5     | 103.2          |              |                      |

TABLE 5: Median differences in stellar parameters  $\Delta P = P_{\text{FLAME}} - P_{\text{input}}$ , along with relative differences  $\Delta P/P$ , and in terms of  $\sigma_{\text{FLAME}}$  for SampleA\_5.

### 4.3 Conclusions

In all cases, FLAME retrieves the input stellar masses and ages to within  $3\sigma$  where  $\sigma$  is the derived uncertainty, with performance to better than  $0.5\sigma$  for main sequence stars,  $\sim 2\sigma$  for subgiants and  $\sim 3\sigma$  for giants. Masses are in general slightly overestimated by these quantities, resulting in ages that are underestimated. Confusion happens mostly for the giants where uncertainties in input data lead to degeneracies in the models that fit them (all of the models fall along one similar diagonal in the HR diagram), and while the results are statistically OK for masses below,  $3 - 4M_{\odot}$ , above this value we see a quick degradation of performance.

For 1%, 2%, and 5% input errors on  $\mathcal{T}_{\text{eff}}$  and  $\mathcal{L}$  (or flux) we retrieve the median mass to +0.3, +0.6, and +0.7% for main sequence, 2.4%, 3.5% and 3.8% for subgiants, and 7%, 7%, and 9% for giants. These numbers for the age are -3.9, -5.1, and -6.4% for main sequence, -6.7, -11.1 and -11.5% for subgiants, and -15, -17.6, and -21.2% for giants. These results show that FLAME performs statistically as it should do in Gaia DR3.

## 5 Comparison with an external code using real Gaia data

In this section we compare the FLAME mass and age using Gaia DR3 input data (these appear in the Gaia DR3 archive) with the performance of the SPInS software<sup>1</sup>

### 5.1 Input Data

We extracted random samples of sources from the FLAME validation data. FLAME validation data consists of approximately 100 000 chosen sources, used for validation with external catalogues, see [OLC-002](#), along with another 11 million sources chosen at random, see [JIRA:DCOPC-563](#), used for internal validation.

For the following tests we imposed that the relative parallax error is better than 50% and then we selected targets in the following  $G$ -band magnitude ranges, chosen to cover a broad range where FLAME provides results:

- Sample3:  $8.50 < G < 9.20$
- Sample4:  $10.95 < G < 11.00$
- Sample5:  $12.725 < G < 12.700$
- Sample6:  $15.10 < G < 15.12$

This resulted in approximately 30 000 sources for each sample, from which 600 sources were extracted at random.

In Fig. 6 we show the distribution of the input data of the sample by plotting the relative luminosity uncertainties as a function of the relative  $\mathcal{T}_{\text{eff}}$  uncertainties, colour-coded by  $G$ . It is clear that the higher magnitudes have generally lower quality data. For the rest of this section we will show the results by magnitude sample.

For each of these sources, the input  $\mathcal{T}_{\text{eff}}$ ,  $\mathcal{L}$ ,  $[M/H]$ , and associated uncertainties (we assume symmetric uncertainties) were provided in an input file and run through the SPInS software, in order to compare directly with the results from FLAME.

### 5.2 Comparison of results

In what follows we compare the mass and age results from the two stellar codes. On the x-axis we show always the results from SPInS, while on the y-axis we show the FLAME or FLAME -

<sup>1</sup>We have no information about postprocessing or removal of sources at the archive level for this test.

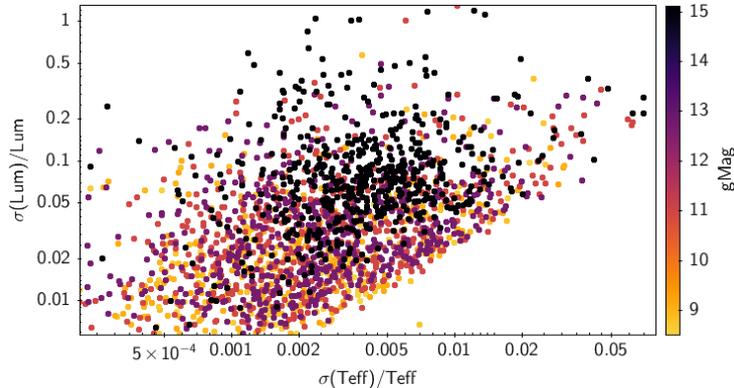


FIGURE 6: Distribution of the Gaia sources as a function of the quality of the input data; relative luminosity versus relative  $\mathcal{T}_{\text{eff}}$  uncertainty, colour-coded by G magnitude.

SPInS results. We do not show results on the evolution stage parameter because this parameter is not retrieved with the current setup of the SPInS code.

### 5.2.1 Mass

We present the comparison of the masses in Fig. 7. Each panel represents a comparison for a specific  $G$  magnitude bin. On the left we show mass from FLAME versus mass from SPInS, with the 1–1 line shown to guide the eye. The colour-code represents the surface gravity of the source as derived by FLAME (using models, we note that this is used for validation purposes only and is not published in data releases). The main sequence stars are the green to yellow coloured stars, while the giants are the dark blue. On the right panels we show histograms of the difference between the masses from FLAME and SPInS, divided by their joint uncertainties. We show two histograms: the blue represents the differences for all stars in the sample, while the orange represents the differences for the main sequence only.

We find that for the main sequence stars ( $\log g > 3.9$ ) we obtain almost perfect agreement between both methods for all magnitudes. For the more evolved stars, there is more scatter among the results, although the scatter is reduced as the  $G$  magnitude increases. This latter could be the effect of having larger uncertainties, which allows the codes to *be more flexible* in finding an adequate result. In all cases, we find that the differences in the giant results can be significant at the 4–5 $\sigma$  level, but most differences are within an expected 3 $\sigma$ . For the main sequence stars, the differences are very small as can be seen on the left panels, but they are also statistically equivalent results as can be seen by the right panels (orange histograms) showing differences of less than 3 $\sigma$ .

### 5.2.2 Age

We present the comparison of the results for ages in Fig. 9, similar to Fig. 7. The left and bottom right panels represent a comparison for a specific  $G$  magnitude bin, with the 1–1 line shown to guide the eye. The colour-code represents the evolution stage of the source as derived by FLAME. Here 100 – 420 represents main sequence, 420-490, the subgiant phase, and 490-1200 the red giant phase. On the right top two panels we show histograms of the difference between the ages from FLAME and SPINs, divided by their joint uncertainties, again separated into the full sample (blue) and the main sequence only (orange). We only show the results for  $G = 9, 11$  because the comparison for the fainter sources resulted in normalised uncertainties corresponding to the span of the parameter space.

Similar to the results for the mass, we find excellent agreement for the main sequence stars (dark blue on the scatter plots) and orange histograms on the histogram plots. Again for the giants we find results that are in worse agreement. These are mostly either FLAME or SPINs that derive very young results, lower than 0.2-0.3 Gyr, and these are in fact normally stars with large derived masses, see Fig. 10 which shows the same age–age plot but colour-coded by mass for  $G \sim 9$  and  $G \sim 13$ . So the confusion happens because of the difference in masses that are derived.

## 5.3 Conclusions of test

We conclude from these tests that for main sequence stars, performance on masses and ages is identical statistically for both FLAME and SPINs. We also found that the uncertainties are underestimated for FLAME, because when repeating the histograms shown in the above figures but considering only FLAME uncertainties, the number of "sigma" were much bigger. Keeping this latter point in mind, and the results from Section 4, we can conclude that the performances with FLAME using simulated data are better than the maximum "3 $\sigma$ " differences that are shown, because the  $\sigma$  are underestimated.

For the more evolved stars, the best agreements between both codes are always found for stars with masses  $< 3 - 4 \mathcal{M}_{\odot}$ , with only some outliers for  $1-3 \mathcal{M}_{\odot}$  stars. The most discrepant values are always when one of the codes estimates a very young age corresponding to a massive star, which the other code does not retrieve.

For the ages published in DR3, we stated that for the giant stars the results are only validated between  $1-2 \mathcal{M}_{\odot}$ , which is consistent with, in fact more strict than, our findings here.

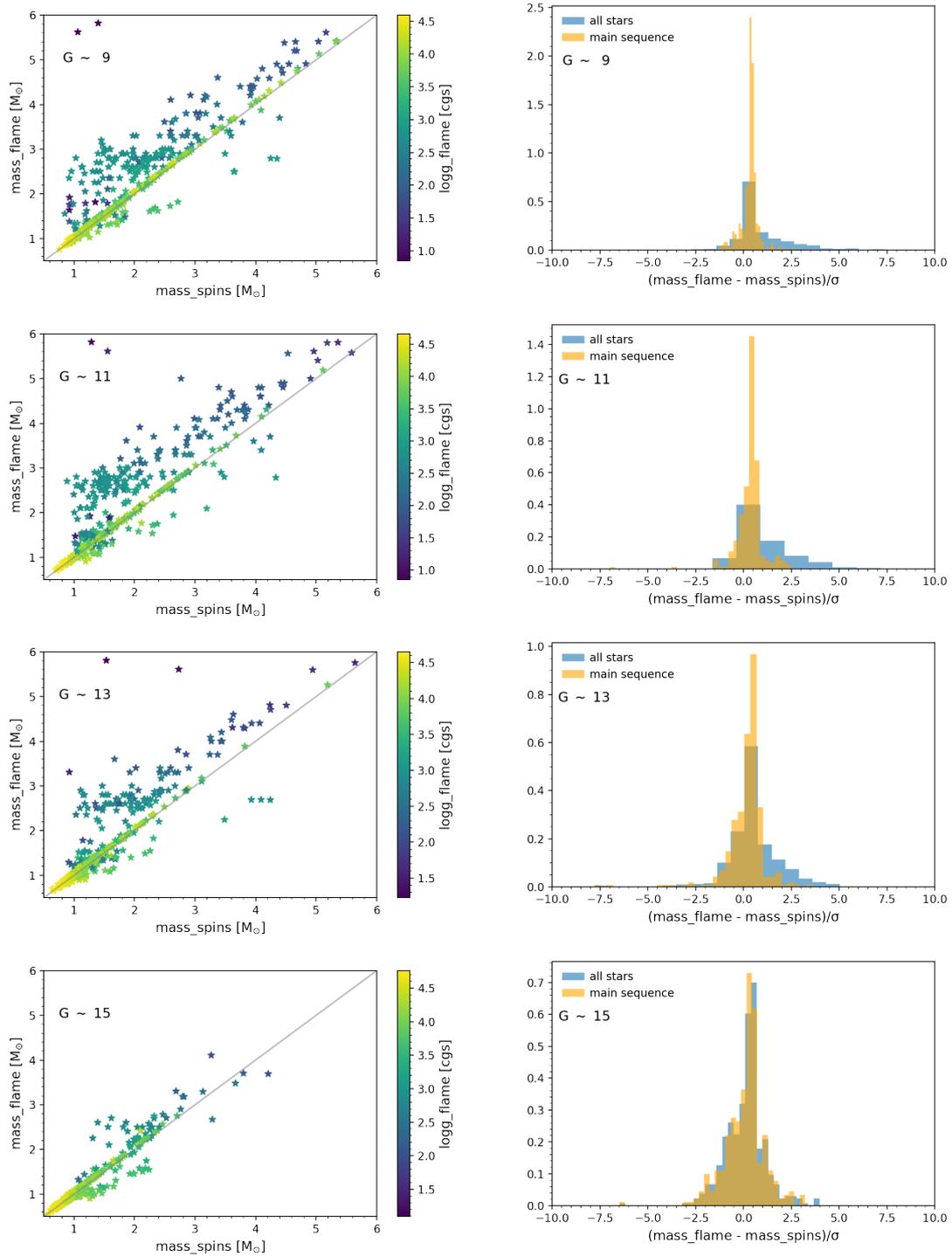


FIGURE 7: Comparison of FLAME and SPINs results for mass for different G (different rows). The left panels show a direct comparison of the values colour-coded by the  $\log g$  from FLAME (model value used for validation, unpublished). The right panels show the differences in masses between both results, but divided by the quadratically-summed uncertainties. In blue we show all stars, and in orange we show the main sequence stars only.

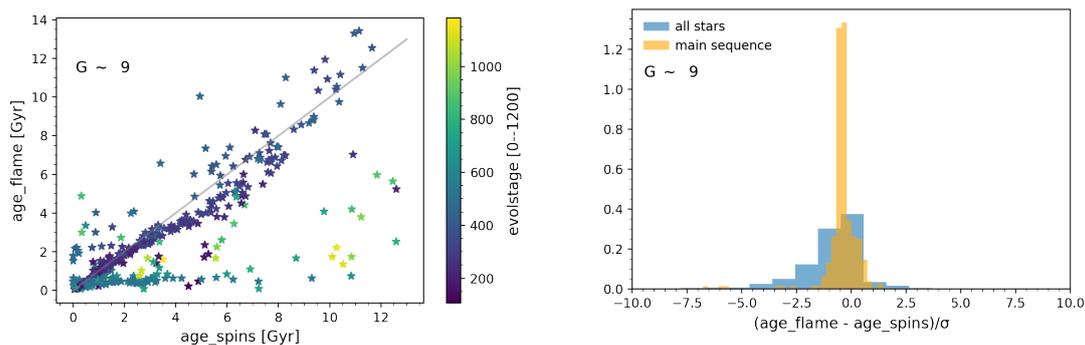
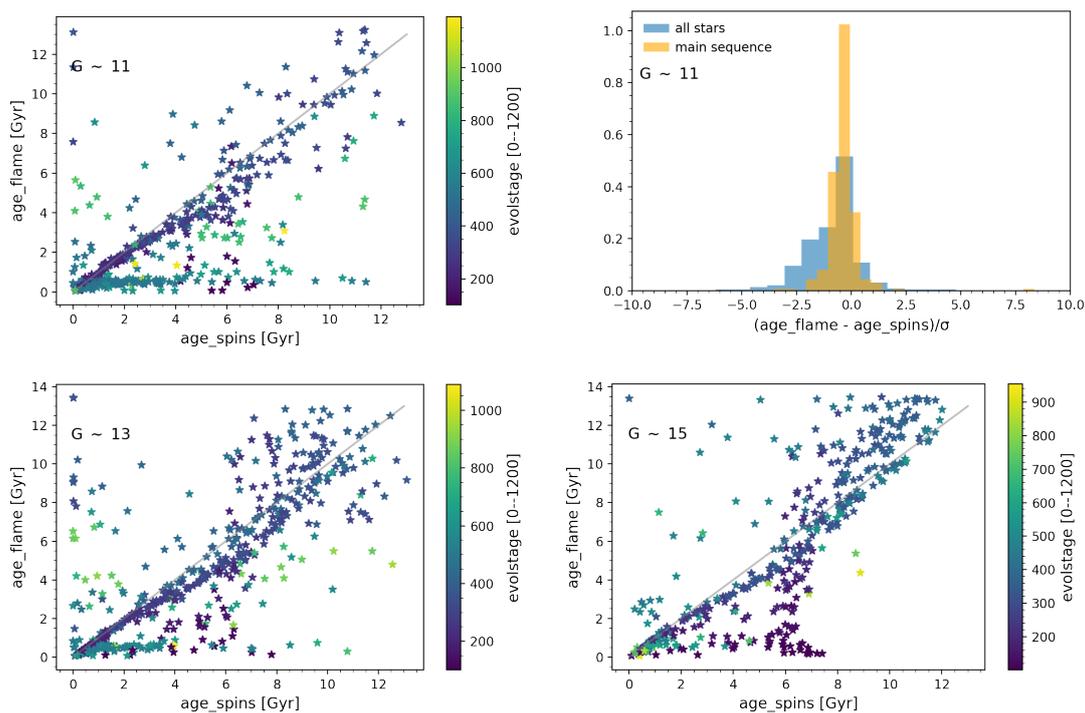
FIGURE 8: Comparison of FLAME and SPINs results for age for  $G \sim 9$ 

FIGURE 9: Comparison of FLAME and SPINs results for age for different  $G$  (different rows). The left and lower panels show a direct comparison of the values colour-coded by the evolution stage from FLAME. The top right panels show the differences in ages between both results, but divided by the quadratically-summed uncertainties. In blue we show all stars, and in orange we show the main sequence stars only.

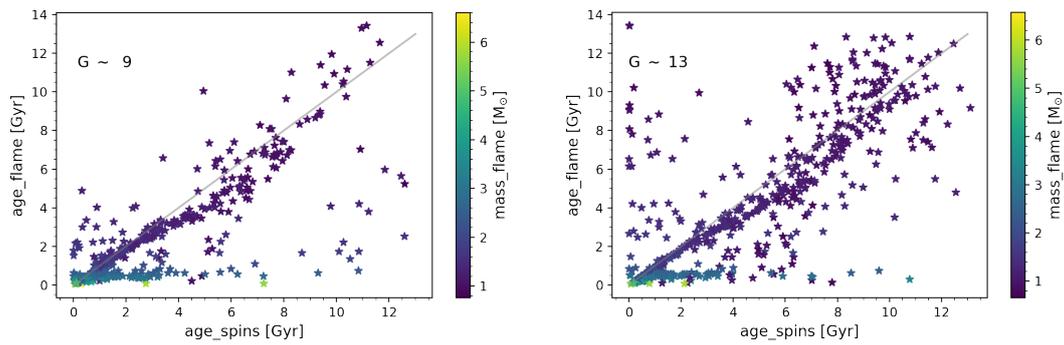


FIGURE 10: See Fig. 9 for caption, but this time the colour-code is mass\_flame.

## 6 Impact of solar metallicity prior on mass and age

Upon validation of initial DR3 operational data, it was decided to assume a solar-metallicity prior for the analysis of the mass, age and evolutionary stage. This decision was made due to some known but at the time non-quantified issues with the stellar metallicity from GSP-Phot. Today we know that GSP-Phot metallicities have a peak offset compared to the literature of  $-0.2$  dex (see Fig. 11 from [Andrae et al. 2022](#) where metallicities are compared to those in open clusters), and have uncertainties on the order of  $0.2 - 0.3$  dex, and are additionally not very reliable at metallicities below around  $-1.0$  dex. In fact, there is a proposed tool from GSP-Phot that exploits Gaia DR3 data and produces a calibrated metallicity<sup>2</sup>, see [the Gaia cosmos software tools](#). These issues, however, led to systematically offset masses and ages with respect to the literature. As this was found during the final pre-operations phase, it was therefore decided to impose a solar metallicity prior for operations, assuming that this supposition would hold for a large majority of the stars.

In this section, we explore the impact of this assumption on non-solar metallicity stars.

### 6.1 Datasets

To test the impact of a solar-metallicity assumption on the mass, age, and evolution stage parameters, we created mock datasets, as explained in Sect. 4.1. The data sets consisted of 600 stars comprising main sequence, sub-giant, and giant evolution states, derived from the `BaSTI` evolution tracks for 8 values of metallicity:  $+0.45$ ,  $+0.30$ ,  $-0.20$ ,  $-0.40$ ,  $-0.70$ ,  $-1.05$ ,  $-1.70$ ,  $-2.20$ . For each mock dataset, we created datasets with errors of 0%, 1%, 2%, and 5% errors in  $\mathcal{L}$  and  $\mathcal{T}_{\text{eff}}$  (see Sect. 4.1 for details), i.e.  $8 \times 4 = 32$  datasets.

For each of the 32 datasets, we used the FLAME software in the same configuration as that for Gaia Data Release 3, but adapting the parameters, just as in Sect. 4.1, in order to remove any errors introduced from the bolometric correction or extinction, i.e. the luminosity and  $\mathcal{T}_{\text{eff}}$  from the mock datasets were identically recovered in FLAME. This allowed us to investigate the impact of the metallicity assumption *only*, on the recovered stellar parameters.

### 6.2 Results per metallicity

The results for 28 of the 32 datasets are shown in Figs. 11–17 (we ignore the set for  $[M/H] = -2.20$  in this section because the results are worse than those for  $-1.70$ ). Each of the figures shows the result for one metallicity, beginning with  $[M/H] = -0.20$  to furthest from solar ( $-1.70$ ). Each panel in each figure shows a comparison of the FLAME-derived parameter (y-axis) versus the true input one (x-axis). The left, center and right panels correspond always to the mass, evolution index, and age parameter, respectively. From top to bottom, the errors on the input

<sup>2</sup><https://github.com/mpi-astronomy/gdr3apcal>

$\mathcal{T}_{\text{eff}}$  and  $\mathcal{L}$  are 0%, 1%, 2%, and 5%. A third variable is included to help interpret the results: the evolution index for the mass and age, and  $\log \mathcal{T}_{\text{eff}}$  for the evolution index parameter, with the lowest values always indicated by the brighter colours. For the third variable evolution index, the brighter yellow and orange colours indicate main sequence and sub-giant stars, while the darker purple are giants.

In all figures, it is clear to see that the only impact of large uncertainties is to have a larger scatter in the resulting comparison, but the actual mean trends are identical for each set of uncertainties. One can also clearly see that the results are worse for giants always, with a typical offset of 20-50% in mass and 50-100% offsets in ages, in the best case scenarios. Typically FLAME overestimates the masses for giants irrespective of the metallicity, which results in underestimated ages. We also see that results are generally in good agreement for all positive metallicities, while the results for negative metallicities degrade much quicker, e.g. compare Fig. 11 with 13, which show that the results for the latter ( $[M/H] = +0.45$ ) are in better agreement than the ones for  $[M/H] = -0.20$  for mass, evolution index, and age.

If we consider only the main sequence and sub-giant stars (i.e. evolution index  $< 490$ ) we find that we have good agreement, or quantifiable offsets, for all parameters for metallicities between  $-0.40$  and  $+0.45$  dex. Below  $[M/H] = -0.40$ , we can see that the age offset degrades significantly, while for the mass, even going as low as  $-1.05$  in metallicity, it is possible to quantify the offset and thus correct the published masses in Gaia DR3, see Sect. 6.3 for further details. For low metallicity stars for the evolution index, we can also still distinguish between main sequence / sub-giant stars and giants, although the actual values are incorrect. The real main sequence stars all appear close to ZAMS values ( $\sim 100$ ), while the real sub-giants have an index spanning 200 - 490. This erroneous information is still useful if one knows the true metallicity. For the giants, we consider that the results for the mass and evolution index are satisfactory for metallicities between  $-0.20$  and  $+0.45$  dex, but below  $[M/H] = -0.20$  there are quantifiable offsets in mass, see Sect. 6.3.

We additionally remark that a star with higher metallicity than the Sun will appear as a slightly underestimated mass by FLAME, and thus overestimated age, and viceversa for sub-solar metallicities. This means that if a user knows that a star has a metallicity of  $+0.30$  for example, then they should add a couple of percent of mass to the published `mass_flame` (in the `gaiadr3.astrophysical_parameters` table) or `mass_flame_spec` (in the `gaiadr3.astrophysical_parameters_supp` table). Likewise for a star that has a subsolar metallicity, a user should remove a few percent from the mass. The amount by which we should modify the parameters (if possible) is discussed in the next subsections.

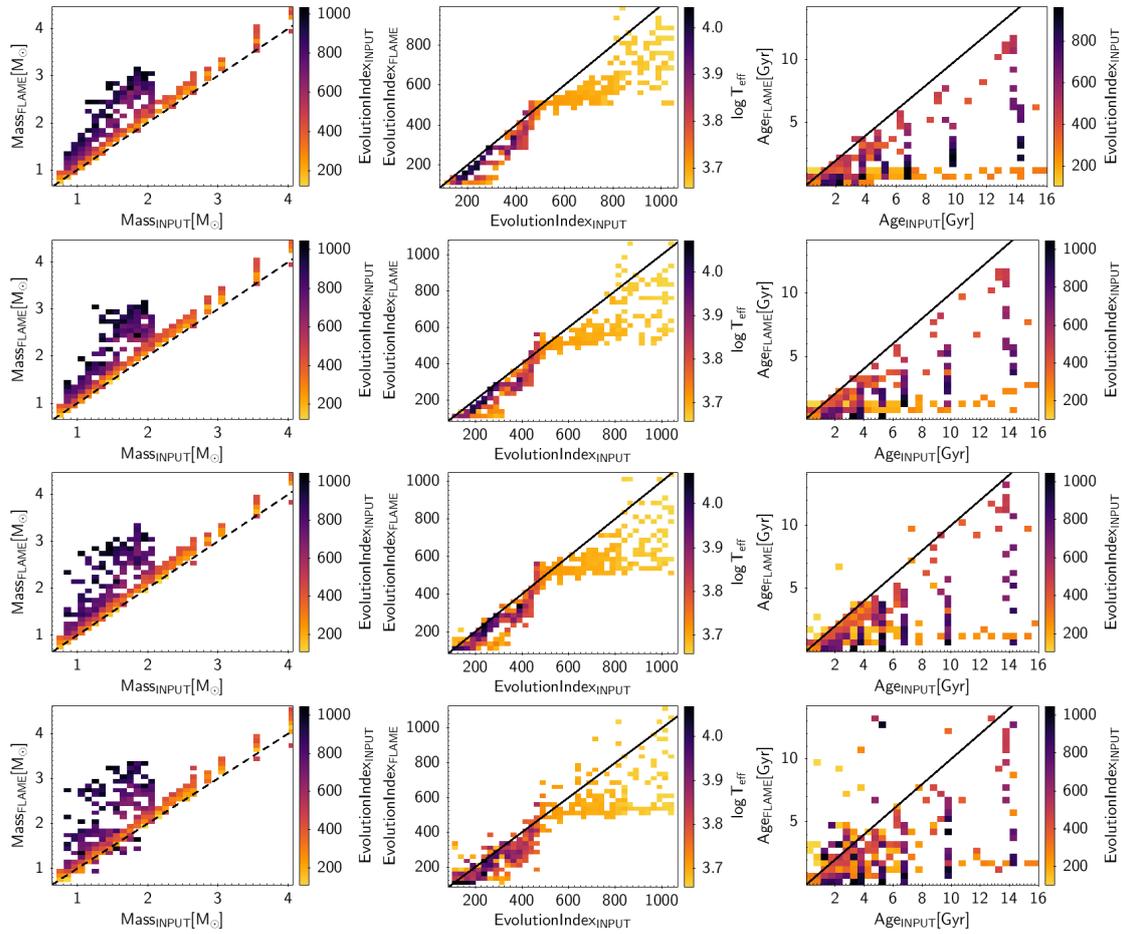


FIGURE 11: Comparison of FLAME results and the input values for mass (left columns), evolution index (middle columns), and age (right columns) for stars with metallicities of  $[M/H] = -0.20$ . Each row considers different simulated error level on the input  $T_{\text{eff}}$  and  $\mathcal{L}$ ; from top to bottom errors are 0%, 1%, 2%, and 5%. The plots are colour-coded according to evolution index (mass and age) or  $\log T_{\text{eff}}$  (evolution index plot).

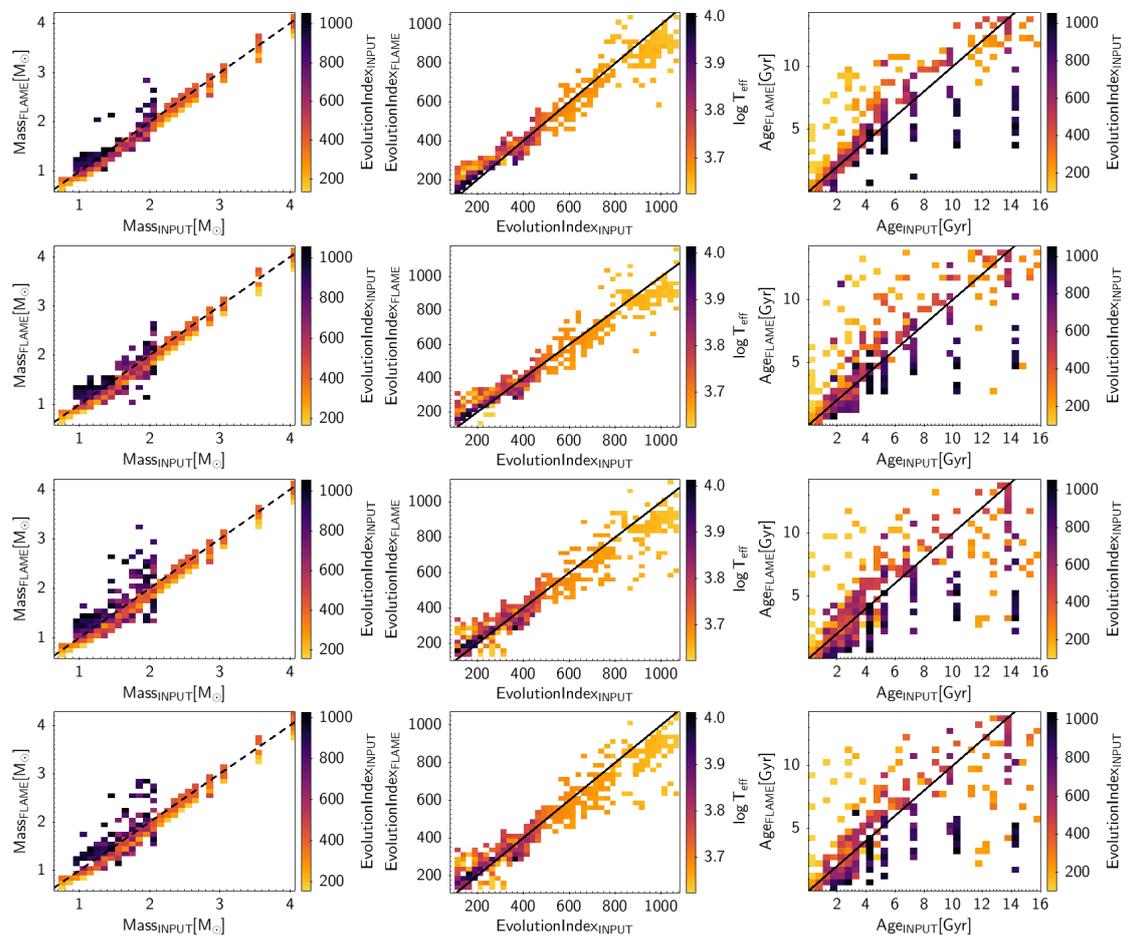
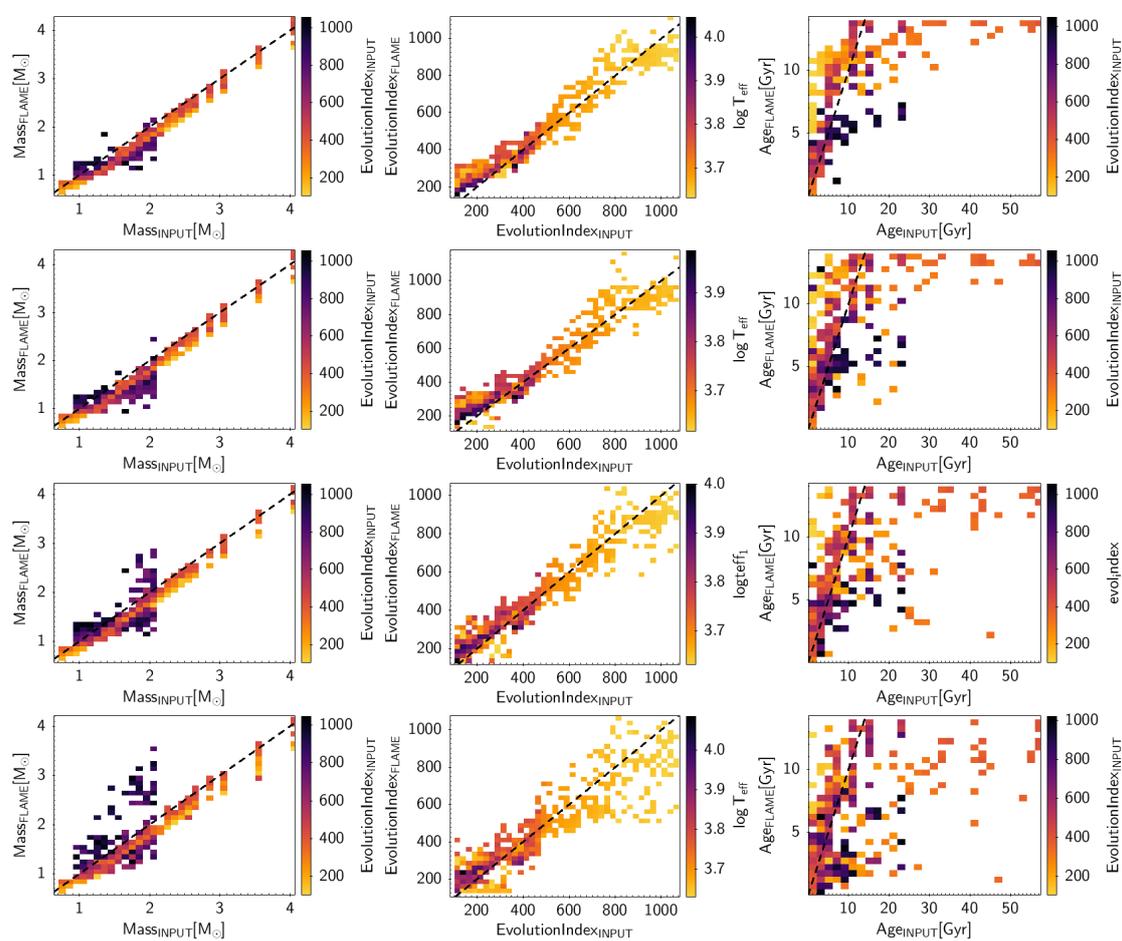
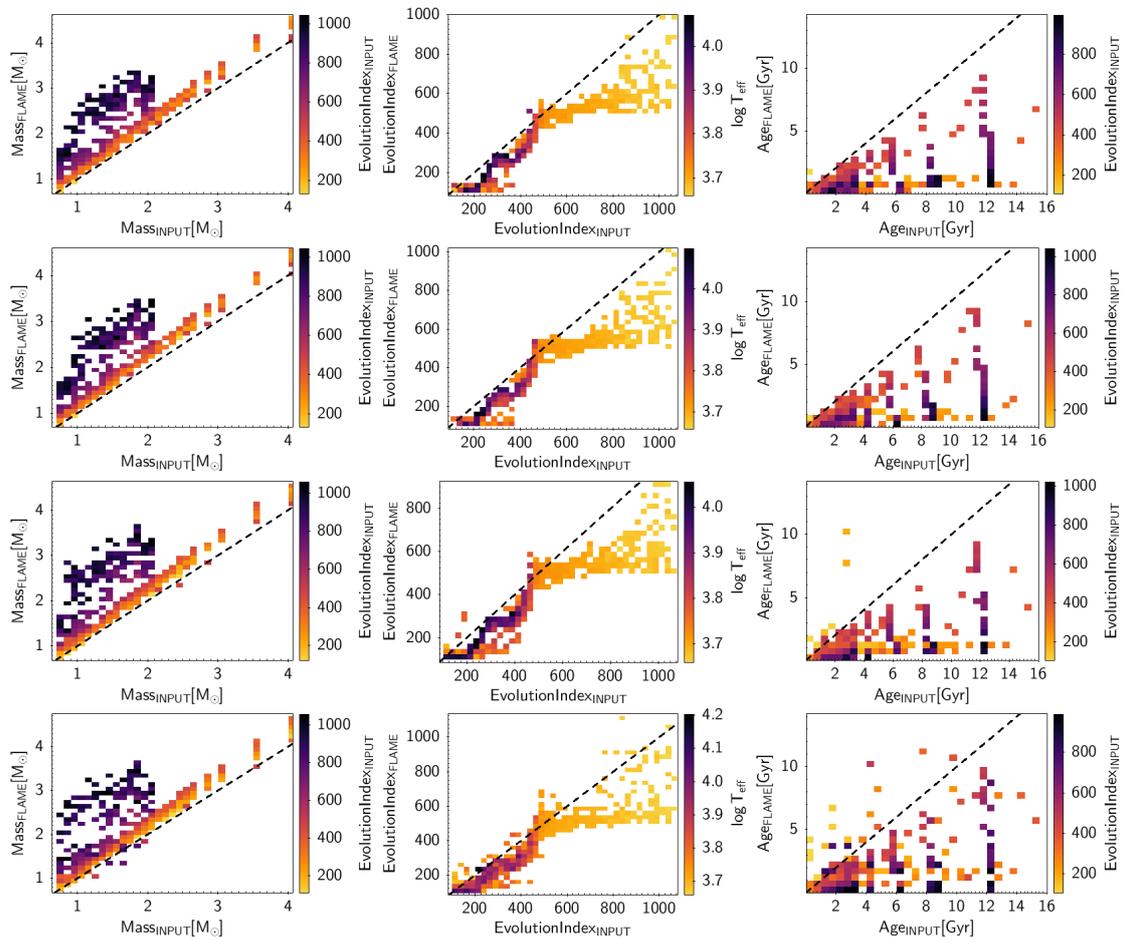
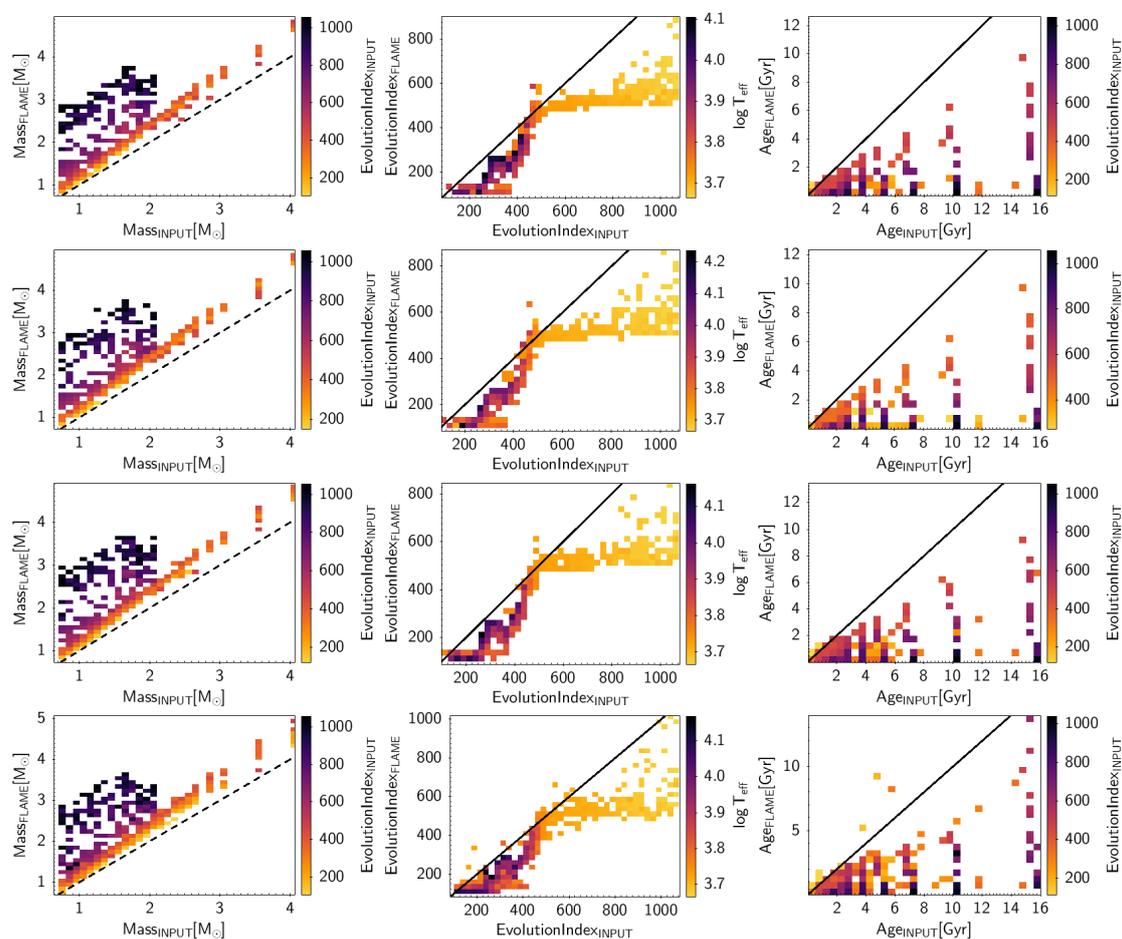
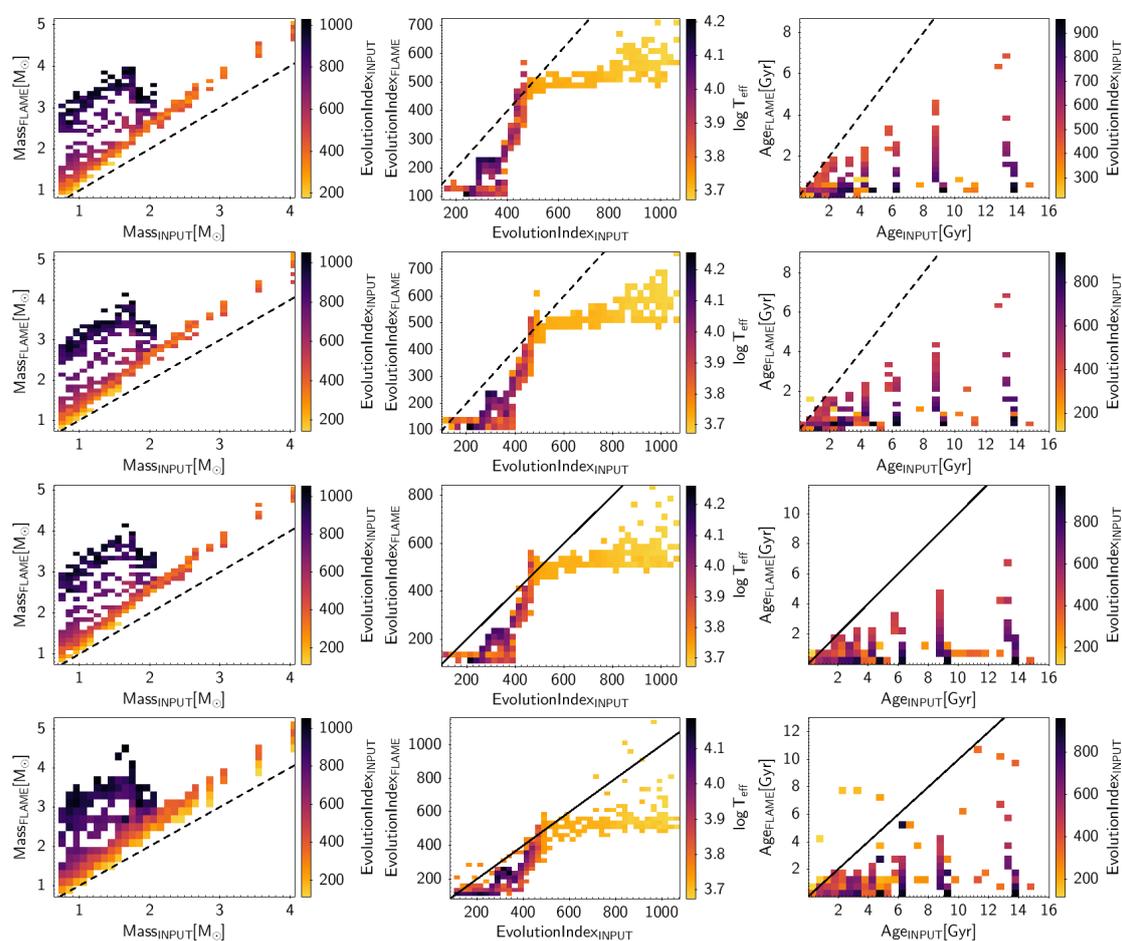


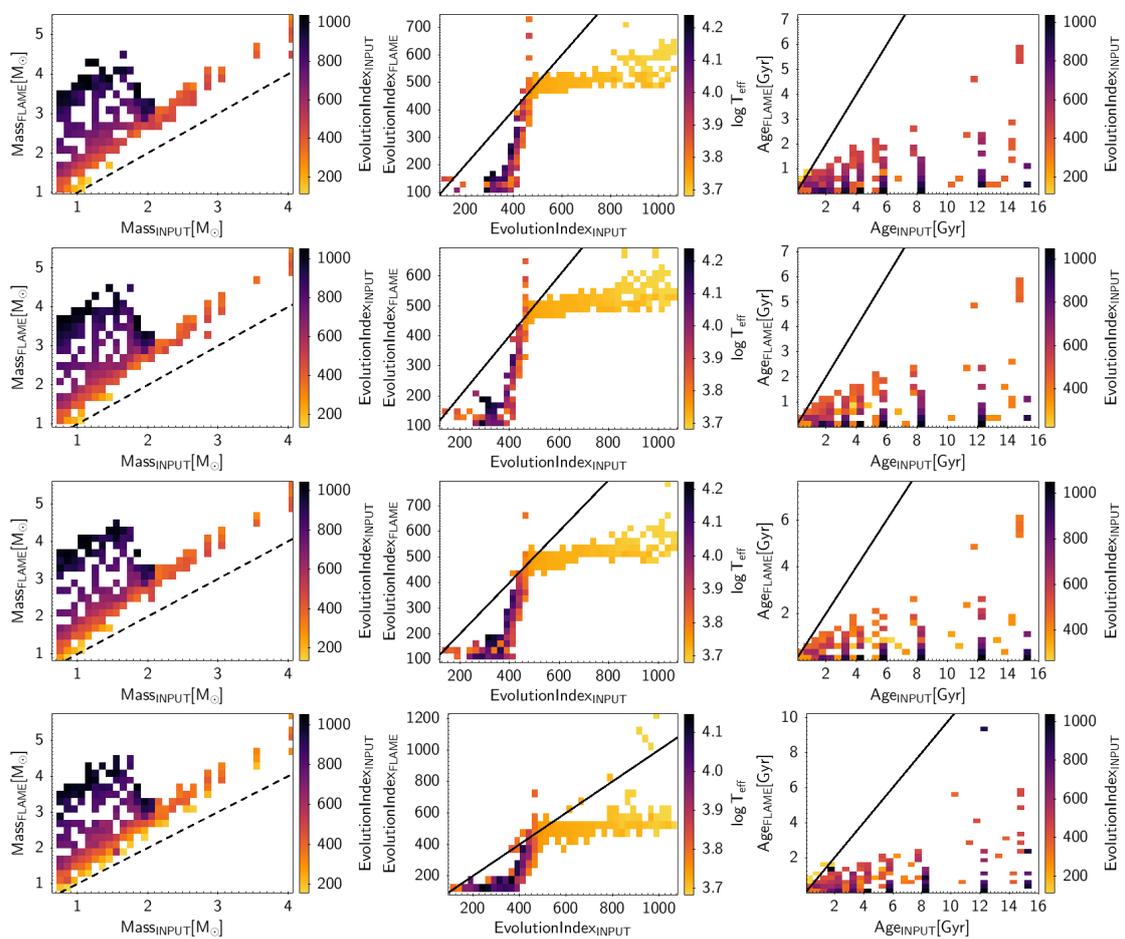
FIGURE 12: Same as caption Fig. 11 but for stars with metallicities of  $[M/H] = +0.30$ .

FIGURE 13: Same as caption Fig. 11 but for stars with  $[M/H] = +0.45$ .

FIGURE 14: Same as caption Fig. 11 but for stars with  $[M/H] = -0.40$ .

FIGURE 15: Same as caption Fig. 11 but for stars with  $[M/H] = -0.70$ .

FIGURE 16: Same as caption Fig. 11 but for stars with  $[M/H] = -1.05$ .

FIGURE 17: Same as caption Fig. 11 but for stars with  $[M/H] = -1.70$ .

### 6.3 Results per inferred parameter

In this section we attempt to quantify the offsets per parameter and provide recipes, where possible, to correct the published FLAME parameters (mass, evolution index, age) which relies on knowing the true metallicity. Each parameter is discussed separately. Additionally, as we have seen that the offsets are very similar for different simulated error levels, we only show the results for 0% and 5% errors, and we propose the same recipe for each metallicity to apply to all errors.

In all of the following figures, we show the parameter from FLAME (y-axis) versus the true input value (x-axis), colour-coded by a third variable. The nine panels correspond to from top left to bottom right:  $[M/H] = +0.06, +0.30, +0.45$  (top),  $-0.20, -0.40, -0.70$  (middle),  $-1.05, -1.70, -2.20$  (lower), the latter three being far outside of the FLAME validity range.

#### 6.3.1 Mass

Figures 18 and 19 compare the input and output masses, colour-coded by evolutionary index. The main sequence and sub-giant stars are the yellow to orange colours, and the giants are the deep red to purple. The two figures differ by input errors: 0% and 5% respectively.

Concerning the giants, we can say that for the FLAME masses between 1 and 2  $\mathcal{M}_{\odot}$ , the results are relatively good for metallicities between  $-0.40$  and  $+0.45$ , i.e. they can be used as an approximation. Below  $[M/H] = -0.40$  it is difficult to quantify the offset.

For the main sequence and sub-giant stars there is a clear agreement with the input data down to metallicities of even  $-1.05$ , although from  $-0.40$  there is a quantifiable offset between the input and output masses. We can therefore attempt to correct the FLAME published masses, if we know the metallicities by applying the following linear corrections to the mass:

$$\mathcal{M}_{\text{cor}} = \mathcal{M}_{\text{FLAME}} - f(\mathcal{M}_{\text{FLAME}}, [M/H]) \quad (1)$$

where  $f(\mathcal{M}_{\text{FLAME}}, [M/H]) = \sum_{i=0}^1 a_i x^i$  is a linear function to apply to the published mass,  $\mathcal{M}_{\text{FLAME}}$ , to derive a corrected mass  $\mathcal{M}_{\text{cor}}$ . The coefficients of the linear fit for each  $[M/H]$  are given in Table 6. The function has been evaluated separately for the main sequence (evolution index  $< 400$ ) and the sub-giants (evolution index  $420 - 490$ ), and the residual of the standard deviation of the data minus the fit is given under the column  $s$ . We note that the median absolute deviation is always a factor of around 30% smaller than the value of  $s$  but we do not report it here.

We show examples of the residuals of (FLAME – input) versus input masses (colour-coded in the figures), along with the residuals when including the correction  $f_{\text{cor}}$  (blue dots) in Fig. 20. As can be seen, the application of  $f_{\text{cor}}$  allows one to recover the true input values. We note that we can not overplot the linear function, because it is evaluated as a function of the FLAME

value (which the user has access to), while here it makes more sense to show the residuals as a function of the input value.

| [M/H] | main sequence |            |      | sub-giants |           |      |
|-------|---------------|------------|------|------------|-----------|------|
|       | $a_0$         | $a_1$      | $s$  | $a_0$      | $a_1$     | $s$  |
| +0.00 | –             | –          | –    | –          | –         | –    |
| +0.30 | -0.0672961    | -0.0282493 | 0.02 | –          | –         | –    |
| +0.45 | -0.1184745    | -0.0338001 | 0.04 | -0.1822361 | 0.0526909 | 0.08 |
| -0.20 | 0.0383209     | 0.0538070  | 0.02 | 0.0417425  | 0.0742857 | 0.10 |
| -0.40 | 0.0376970     | 0.1019328  | 0.02 | 0.0396883  | 0.1206176 | 0.08 |
| -0.70 | 0.0191003     | 0.1688453  | 0.04 | 0.1637709  | 0.1209222 | 0.08 |
| -1.05 | 0.0307472     | 0.2146826  | 0.04 | 0.1970984  | 0.1565529 | 0.07 |
| -1.70 | -0.1419530    | 0.3413357  | 0.06 | 0.4057037  | 0.1355701 | 0.08 |
| -2.20 | -0.2657367    | 0.3888628  | 0.04 | 0.4107980  | 0.1574575 | 0.08 |

TABLE 6: Coefficients of  $f(\mathcal{M}_{\text{FLAME}}, [\text{M}/\text{H}])$  to correct the FLAME masses for different metallicities. For the main sequence the stars with input masses between  $1.0 - 4.0 M_{\odot}$  are used, although we propose to apply this fit to lower main sequence stars, and for the sub-giants the input mass range is between  $0.75$  and  $4.0 M_{\odot}$ .

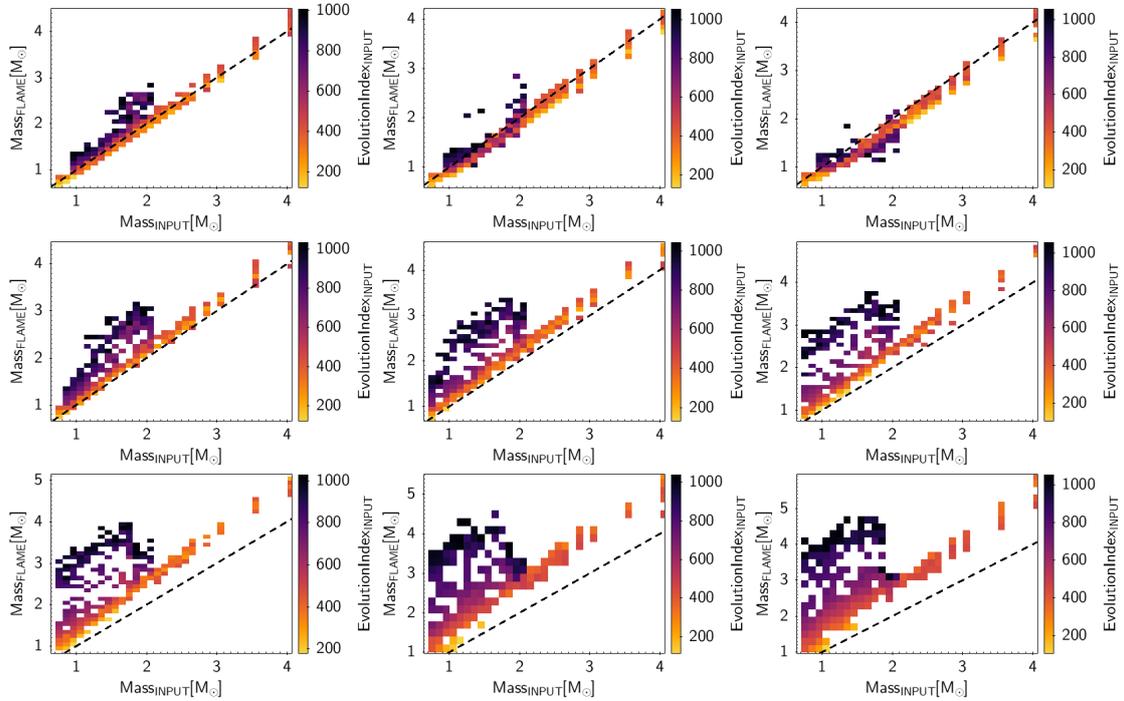


FIGURE 18: Results for mass with no errors. Top panel is increasing metallicity from solar to +0.30 to +0.45 dex. Middle panel: decreasing metallicity:  $-0.20$ ,  $-0.40$ ,  $-0.70$ . Lower panel: Outside of validity range:  $-1.05$ ,  $-1.70$ ,  $-2.20$

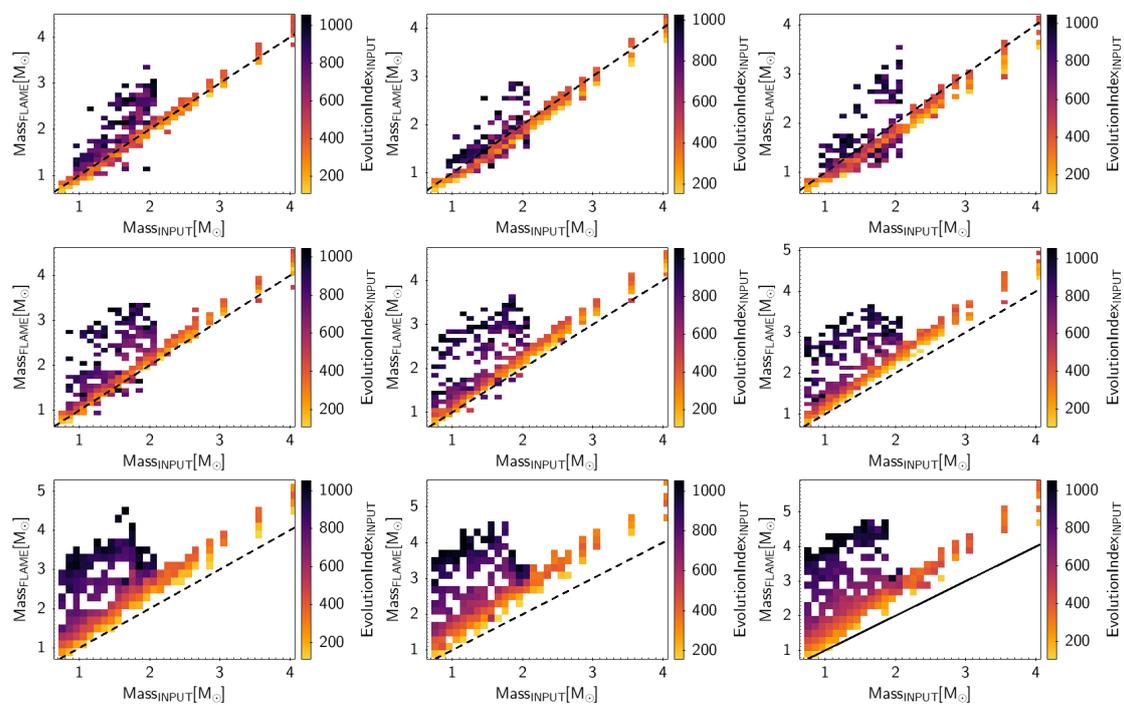


FIGURE 19: Same as Fig. 18 but for 5% errors.

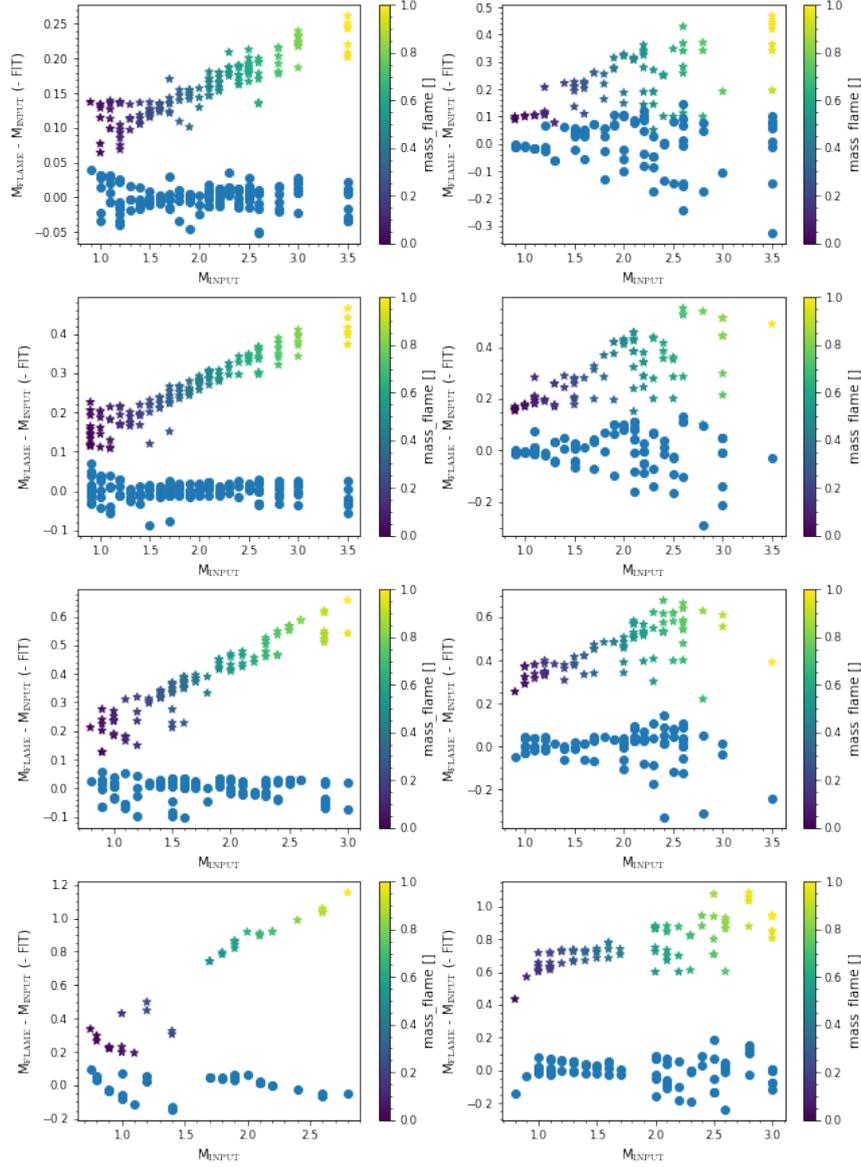


FIGURE 20: Residuals of  $\mathcal{M}_{\text{FLAME}}$  (colour-coded by mass) and  $\mathcal{M}_{\text{COR}}$  (blue), where for the latter we have applied the proposed functions to correct the masses. Left and right panels are main sequence and sub-giant stars, respectively. From top to bottom, we show the results for  $[\text{M}/\text{H}] = -0.20, -0.40, -0.70$  and  $-1.70$ .

### 6.3.2 Evolution Index

We now focus on the evolution index parameter in order to propose a recipe to the user for exploiting the `evolstage_flame(_spec)` parameter in the case of a non-solar metallicity star, assuming that we have external knowledge of the metallicity.

The comparison between the input and FLAME results are shown in Figs. 21 and 22, see caption of Fig. 18 for details. Similar to the mass, the results with different input uncertainties show the same offset trends, but just with more dispersion. In Table 7 for each metallicity, and for ms, sg, and gt evolutionary stages, we describe the proposed action to take in order to interpret the FLAME evolutionary parameter given the star's metallicity. Under "index offset" we give the mean offset of the (FLAME – input ) evolutionary index, and under "dispersion" we give the mean residual of the results for 0% – 5% input errors. For example, if the star is known to have solar metallicity, and we are interested in stars half way through their evolution, then we should look for sources with evolution index between 230 and 330, and going from 0% to 5% errors in the input data, we should expect a typical dispersion on the values between 1 and 16 points. As another example, if a star has a known metallicity of  $-0.70$ , and its evolution index is between  $\sim 350$  and 490, then it is most likely a sub-giant.

For metallicities outside of the valid range, the behaviour is slightly different, and in Table 8 we show how to interpret these results. As an example, if we know the star has  $[M/H] = -1.00$  and we would like to find sub-giants, then we should choose those stars with an index between 300 and 480. As another example, if the star has a  $[M/H] = -1.00$  and FLAME gives an evolution stage around 100, then it is most likely a main sequence star, but we can not tell in which stage it is in.

| [M/H] | true<br>evol stage | index<br>offset | dispersion<br>0% – 5% | observation                                 |
|-------|--------------------|-----------------|-----------------------|---|
| +0.06 | ms                 |                 | 1 – 16                |   |
|       | sg                 |                 | 15 – 20               |   |
|       | gt                 |                 | 45 – 78               |   |
| +0.30 | ms                 | +25             | 16 – 16               |   |
|       | sg                 | +5              | 12 – 14               |   |
|       | gt                 | –26             | 40 – 50               |   |
| +0.45 | ms                 | +58             | 23 – 28               |   |
|       | sg                 | +18             | 13 – 32               |   |
|       | gt                 | +16             | 57 – 80               |   |
| –0.20 | ms                 | –40             | 1 – 2                 |   |
|       | sg                 | –27             | 29 – 49               |   |
|       | gt                 | –160            | 70 – 116              | true giants tend to have indices around 500 |
| –0.40 | ms                 | –65             | 22 – 24               |   |
|       | sg                 | –32             | 23 – 42               |   |
|       | gt                 | –200            | 103 – 111             | true giants tend to have indices around 500 |
| –0.70 | ms                 | –90             | 22 – 24               |   |
|       | sg                 | –46             | 44 – 52               |   |
|       | gt                 | –240            | 113 – 127             | true giants tend to have indices around 500 |

TABLE 7: Interpretation of the published FLAME evolstage parameter ("index") for metallicities between –0.70 and +0.45.

| true<br>evol stage | [M/H] | observation                             |
|--------------------|-------|---|
| ms                 | –1.05 | flat index of $174 \pm 47$              |
|                    | –1.70 | linear increase in index from 300 – 480 |
|                    | –2.20 | flat index of $534 \pm 43$              |
| sg                 | –1.05 | flat index of $135 \pm 33$              |
|                    | –1.70 | linear increase in index from 300 – 490 |
|                    | –2.20 | flat index of $516 \pm 34$              |
| gt                 | –1.05 | flat index of $118 \pm 27$              |
|                    | –1.70 | linear increase in index from 290 – 490 |
|                    | –2.20 | flat index of $513 \pm 70$              |

TABLE 8: Interpretation of the published FLAME evolstage parameter ("index") for metallicities  $< -1.00$

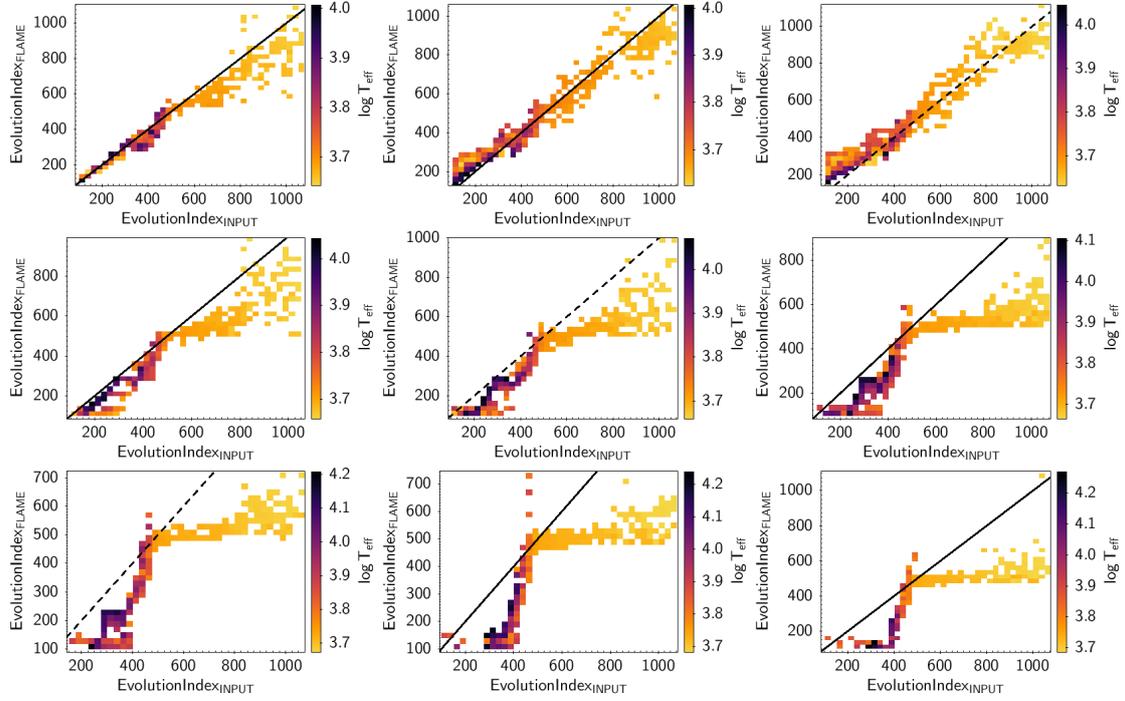


FIGURE 21: Same as Fig. 18 but for the evolution index parameter with 0% errors.

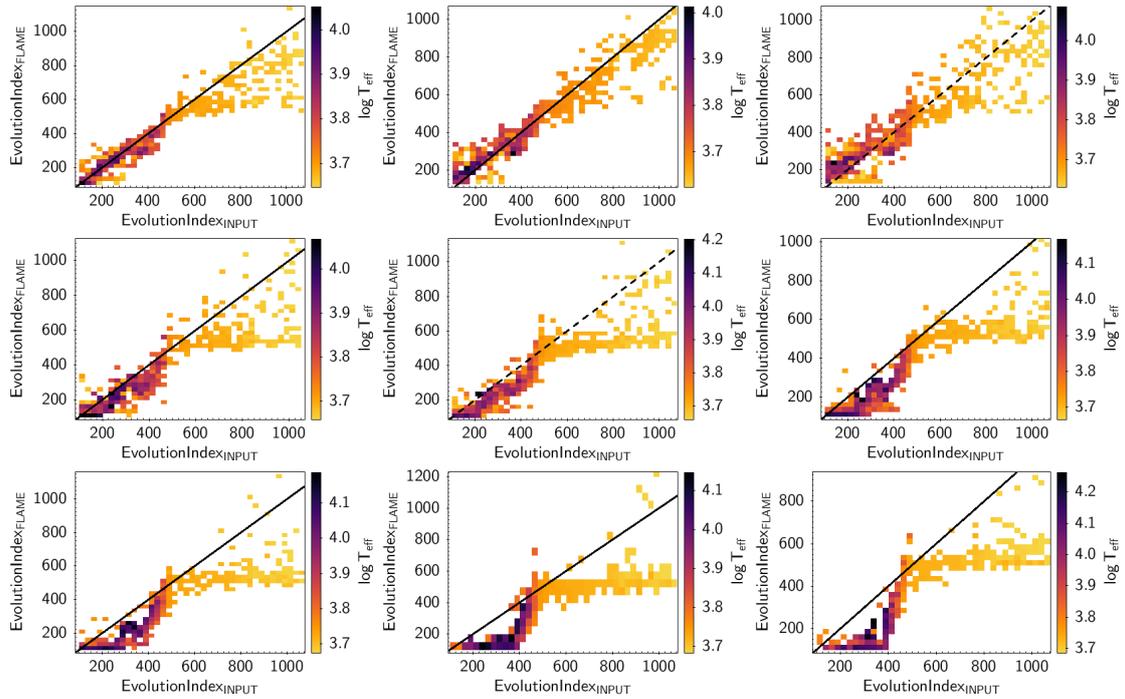


FIGURE 22: Same as Fig. 18 but for the evolution index parameter with 5% errors.

### 6.3.3 Age

Now we consider the results for the stellar age of a star. Just as in the previous sections, we show the results for different metallicities in Figs. 23 and 24, where the colour-code refers to the true evolution index (giants are dark red to dark purple, main sequence stars are yellow–orange).

The main trends that can be seen are:

- The ages of young main sequence stars can only be recovered or attempted to be corrected for stars with metallicities of +0.06, +0.30, and +0.45.
- For more evolved main sequence stars and sub-giants, these have the best chances of being recovered and we discuss this further
- For giants the ages are typically always underestimated, and attempting to correct them may be possible only for stars with  $[M/H]$  between +0.45 and  $-0.20$ .

We follow the approach taken to correct the masses and propose a linear fit to improve the ages

$$\tau_{\text{cor}} = \tau_{\text{FLAME}} - f(\tau_{\text{FLAME}}, [M/H]) \quad (2)$$

where  $f(\tau_{\text{FLAME}}, [M/H]) = \sum_{i=0}^1 a_i x^i$  is a linear function to apply to the published age,  $\tau_{\text{FLAME}}$ , to derive a corrected  $\tau_{\text{cor}}$  age. The coefficients of the linear fit for each  $[M/H]$  are given in Table 9. The function has been evaluated for the sub-giants (evolution index between 420 and 490), and the residual of the standard deviation of the data minus the fit is given under the column  $s$ . We can not give a proposed correction to the ages of the main sequence stars nor the giants because a simple correction is not sufficient.

As can be seen in the figures, by correcting  $\tau_{\text{FLAME}}$ , it is possible to recover the input ages, although we still end up with uncertainties on the order of  $\pm 0.50$  Ga. No corrections are proposed for the super-solar metallicity values.

## 6.4 Conclusions

In this section we evaluated the validity of the solar-metallicity assumption that we imposed during operations. We first showed the results of simulations showing the impact on mass, age, and evolutionary stage for non-solar metallicity models, when inferring the parameters with solar-metallicity assumptions. We found that around solar metallicity, e.g. between  $-0.40$  and  $+0.45$ , our results are valid, considering the uncertainties, but systematic differences start to be important beyond  $-0.50$  dex, in particular for the age and evolution index. For masses of main sequence and sub-giant stars, the offsets can be quantified up to  $-2.20$  dex and thus corrected for. For the age, only corrections for sub-giants have been suggested, because it was not possible to

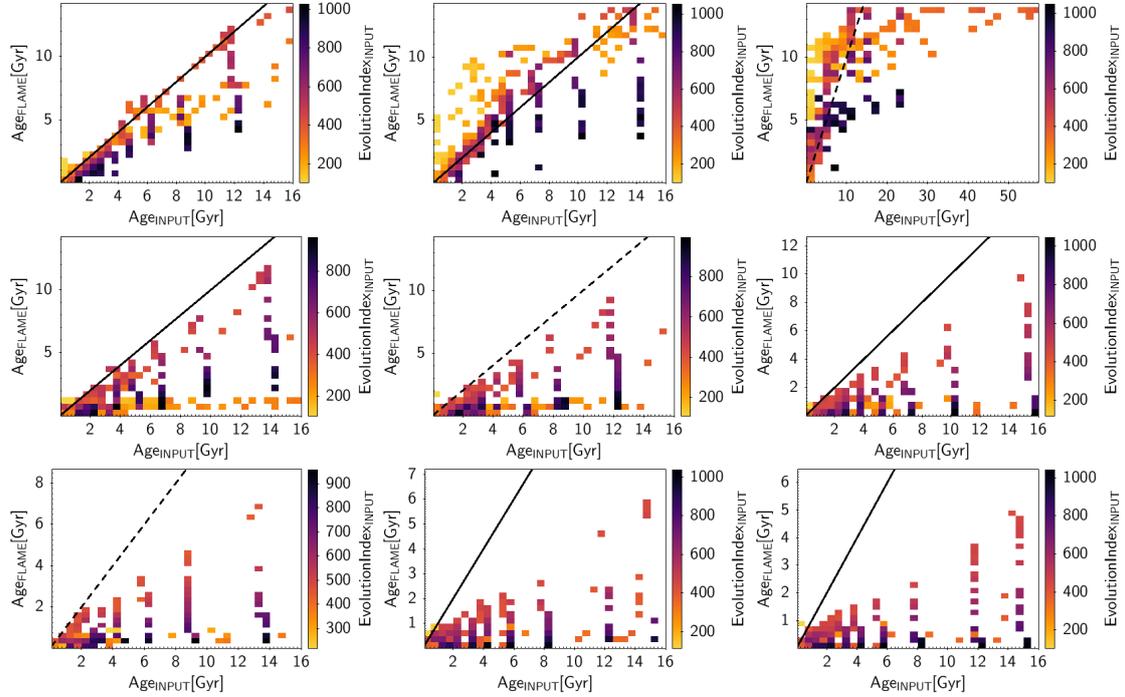


FIGURE 23: Same as Fig. 18 but for the age parameter with 5% errors.

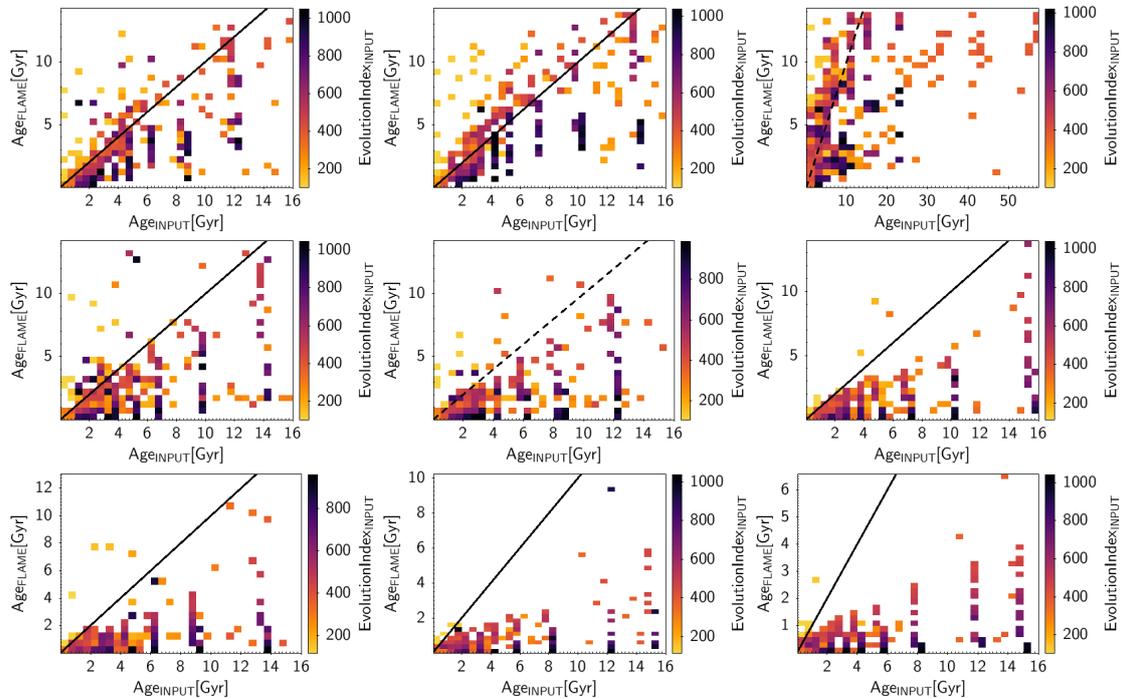


FIGURE 24: Same as Fig. 18 but for the age parameter with 5% errors.

| [M/H] | sub-giants |            |      |
|-------|------------|------------|------|
|       | $a_0$      | $a_1$      | $s$  |
| +0.00 | —          | —          | —    |
| +0.30 | —          | —          | —    |
| +0.45 | —          | —          | —    |
| -0.20 | 0.0042181  | -0.1914876 | 0.20 |
| -0.40 | 0.0797576  | -0.3347886 | 0.28 |
| -0.70 | 0.2083332  | -0.6618318 | 0.32 |
| -1.05 | 0.3251517  | -1.0439210 | 0.45 |
| -1.70 | 0.4839032  | -1.7493386 | 0.33 |
| -2.20 | 0.8684005  | -2.6220880 | 0.52 |

TABLE 9: Coefficients of  $f(\tau_{\text{FLAME}}, [\text{M}/\text{H}])$  to correct the FLAME ages for different metallicities for sub-giants.

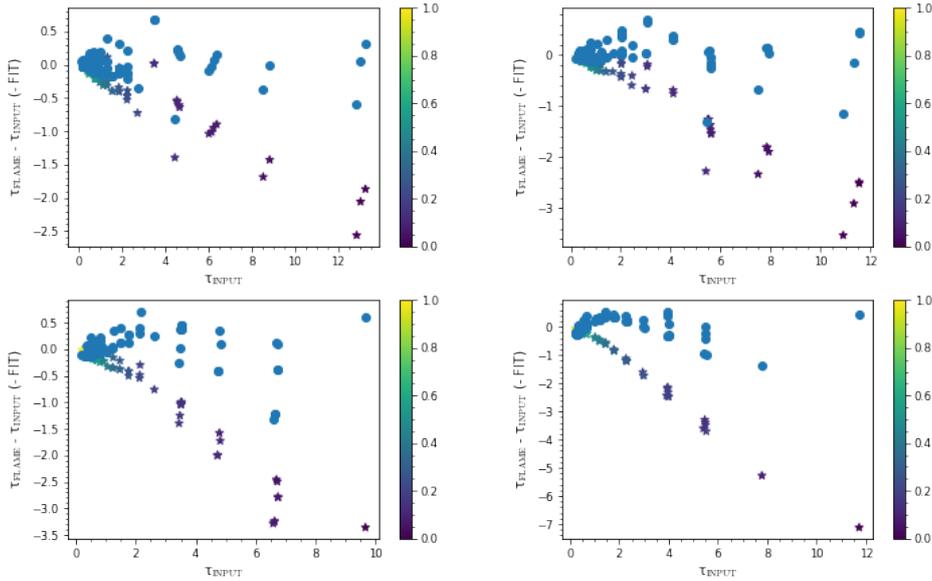


FIGURE 25: Residuals of  $\tau_{\text{FLAME}}$  (colour-coded by mass) and  $\tau_{\text{COR}}$  (blue), where for the latter we have applied the proposed functions to correct the ages. All panels show the results for sub-giants. From top left to bottom right, we show the results for  $[\text{M}/\text{H}] = -0.20, -0.40, -0.70$  and  $-1.70$ .

derive a simple correction for main sequence and giants. For the evolution index, we propose a way of interpreting the FLAME results based on the knowledge of an external metallicity. The proposed corrections to and interpretations of the parameters derived by FLAME are then given in Tables 6 – 9, if the metallicity of the star is known and is non-solar. We recommend using these corrections, in particular outside of the  $-0.50 - +0.50$  regime.

## 7 Conclusions

In this technical note we validated the results of the mass, age, and evolutionary index parameters presented in Gaia DR3. These products are found in the `gaiadr3.astrophysical_parameters` and `gaiadr3.astrophysical_parameters_supp` tables, see Table 1.

In Section 4 we showed a comparison of input and output test data for masses and ages and showed that FLAME performs as it should in the ideal conditions, while imposing the Gaia DR3 assumptions i.e. solar-metallicity. Reducing the precision to 5% on the input data had the effect of increasing the resulting offsets between the input and output to 0.7%, 3.8% and 9% for main sequence, subgiant, and giant stars, respectively.

In Section 5 we compared the performance of FLAME with a published external code, `SPINS`, on real Gaia data, and found that the results are identical statistically for main sequence stars. For the more evolved stars, the best agreements are always found for stars with masses less than  $3 - 4 M_{\odot}$ , we also validated the statement confirmed with observations in the Gaia DR3 documentation that for giants, that one should only use the results for masses between  $1 - 2 M_{\odot}$ . This statement is more conservative than the results we showed here.

Finally in Section 6 we explored the impact on the results for stars of non-solar metallicity of using a solar-metallicity prior. We showed that our results are valid for  $-0.50 < [M/H] + 0.50$ , but even beyond that the masses can still be used. We proposed empirical corrections to make to the mass and the age of the star based on the FLAME properties and an external knowledge of the metallicity for main sequence and sub-giants. These corrections are given in Tables 6 – 9.

## 8 Acknowledgements

This work has made extensive use of `Topcat` (Taylor, 2005). `Topcat` is actively developed by Mark Taylor and can be downloaded here: <http://www.starlink.ac.uk/topcat/>. The authors thank Rosanna Sordo for constructive feedback.

## References

Andrae, R., Fouesneau, M., Creevey, O., et al., 2018, *A&A*, 616, A8, [ADS Link](#)

Andrae, R., Fouesneau, M., Sordo, R., et al., 2022, arXiv e-prints, arXiv:2206.06138, [ADS Link](#)

[ICAP-CBJ-021], Bailer-Jones, C., 2005, *Astrophysical Parameters: Workpackage descrip-*

tions,

ICAP-CBJ-021,

URL <http://www.rssd.esa.int/cs/livelihood/open/526280>

Creevey, O.L., Sordo, R., Pailler, F., et al., 2022, arXiv e-prints, arXiv:2206.05864, [ADS Link](#)

Delchambre, L., Bailer-Jones, C.A.L., Bellas-Velidis, I., et al., 2022, arXiv e-prints, arXiv:2206.06710, [ADS Link](#)

Fouesneau, M., Frémat, Y., Andrae, R., et al., 2022, arXiv e-prints, arXiv:2206.05992, [ADS Link](#)

Hidalgo, S.L., Pietrinferni, A., Cassisi, S., et al., 2018, ApJ, 856, 125, [ADS Link](#)

Lebreton, Y., Reese, D.R., 2020, A&A, 642, A88, [ADS Link](#)

[**MFX-026**], MFX-026, 2021, *CU8 Validation of Outputs from APSISOPS3.2*,

GAIA-CU8-TN-OCA-MFX-026,

URL <http://www.rssd.esa.int/cs/livelihood/open/1555368>

[**OLC-032**], OLC-032, 2021, *FLAME Validation Report based on APSISOPS3.2*,

GAIA-CU8-TN-OCA-OLC-032,

URL <http://www.rssd.esa.int/cs/livelihood/open/1445478>

Taylor, M.B., 2005, In: Shopbell, P., Britton, M., Ebert, R. (eds.) *Astronomical Data Analysis Software and Systems XIV*, vol. 347 of Astronomical Society of the Pacific Conference Series, 29, [ADS Link](#)