The Gaia promise of Astrophysical Parameters in GDR3

Orlagh Creevey*
Rosanna Sordo, Morgan Fouesneau
And the Gaia-DPAC CU8 Team

*ocreevey@oca.eu, Gaia-DPAC-CU8 Lead
In the last episode (DR2),...

\[(G, B_p, R_p, \text{varpi}) \rightarrow T_{\text{eff}}, A_G, E(Bp-Rp), R, L\]

Gaia-DPAC-CU8- GSP-Phot / FLAME (Andrae+ 2018)
In the next episode (DR3), ...
Use of mean Bp and Rp Spectra

Bp and Rp spectra produced by Gaia-DPAC-CU5/DPCI
Astrophysical Parameters (APs) produced by Gaia-DPAC-CU8/DPCC
Use of mean RVS spectra

RVS spectra produced by Gaia-DPAC-CU6/DPCC; APs, Abundances, DIB produced by CU8/DPCC
Behind CU8 Astrophysical Parameters
Astrophysical Parameters from CU8

CU8 processing at DPCC (CNES)
Astrophysical Parameters from CU8

Stellar Parameters
$T_{\text{eff}}, A_G, E(Bp-Rp), \log g, [M/H], \ldots$

R, L, distance, mass, age, binarity, ..

Abundances, activity index, DIB, emission line stars, …

CU5 processing at DPCC (CNES)
Astrophysical Parameters from CU8

Extra-galactic, Interstellar medium & Outliers
Classifications, class probabilities
QSO and galaxy redshifts
2D total extinction map, Outlier classification

CU8 processing at DPCC (CNES)
The HR diagram for 500M stars

GSP-Phot derives $T_{\text{eff}}$, $\log g$, $[\text{M/H}]$, $A_G$ from XP spectra for stars:
- $2,500 < T_{\text{eff}} < 55,000$ K
- $0 < A_G < 10$
- $G < 19$
- Use of isochrones for distance
Insight into Hot-Star physics

Hot stars dominating the region between the ZAMS and TAMS. The more advanced stages (horizontal branch) can be clearly seen by the overdensity in the top right (ESP-HS).
Chemical abundances of FGK stars from RVS spectra by GSP-Spec (Recio-Blanco+ 2016)

N, Mg, Si, S, Ca, Ti, Cr, Fe, FeII, Ni, Zr, Ce, Nd
The DIBs are absorption features seen in high resolution stellar spectra.

In the RVS spectra the 862nm DIB is observed and its EW is measured by GSP-Spec for about 500,000 spectra, (He+ 2020).

**Figure:** Top-down view of the Milky Way in the solar neighbourhood showing the strength of the DIB features which increases with distance.
DSC (Bailer-Jones+ 2019) provides class probabilities of sources using XP spectra, photometry and astrometry:

- Star
- Galaxy
- Quasar
- White dwarf
- Binary

**Figure:** photometric indices of the sources classified as galaxies and quasars for a high completeness sample.
QSO emission lines BP/RP spectra \(17 \leq G \leq 20\) mag

QSO (Delchambre+2018) measures \(\sim 1.8M\) qso candidates with \(0 \leq z \leq 6\) from Bp and Rp spectra

Bp and Rp spectra ordered by QSO redshift \(z\). The colour-code indicates the normalised flux (more intense red lines are the redshifted emission lines)
CU8 harvest in Gaia DR3

- Classification probability → ~1.8G sources
- APs from XP spectra → ~500M stars
- Evolutionary parameters → ~330M stars
- APs from RVS spectra → ~4.8M stars
- Chemistry from RVS spectra → ~50K - 4.3M stars
- Redshifts → ~1.3M unresolved galaxies + ~1.8M QSOs
- Activity index, Ha EW, emission line stars classes, hot star classes, binary
- 2D map of total Galactic extinction
- Outlier analysis (self-organised map and statistics)

Quite rich and complex dataset, stored in 11 different archive tables
Where are all of these parameters?
Where are all of these parameters?
Where are all of these parameters?

Parameters

Multi-D

- AstrophysicalParameters
- AstrophysicalParametersSupp
- Dr2Neighbourhood
- EpochPhotometry
- GaiaSource
- Galaxy
- GalaxyCatalogueName
- McmcSamplesGspPhot
- McmcSamplesMsc
- McmcSource

- OaNeuronInformation
- OaNeuronXpSpectra
- Qso
- QsoCatalogueName
- RSMeanSpectrum
- ScienceAlerts
- QsoSource
- TotalGalacticExtinctionMap
- TotalGalacticExtinctionMapOpt

- Gspphot
- Gspspec
- Espucd
- Esphs
- Msc

- teff
- teff_gspphot, teff_gspphot_marcs
- teff_gspspec, teff_gspspec_ann
- teff_espucd, teff_espucd_uncertainty
- teff_esphs, teff_esphs_uncertainty
- teff_msc1, teff_msc2, teff_msc1_upper
Conclusions

- Gaia DR3 is the first data release for most of CU8
- Process Bp and Rp (mean) spectra, RVS spectra, astrometry, photometry
- Stellar parameterization
  - \( T_{\text{eff}}, A_G, \log g, [M/H], \text{mass}, \text{age}, \text{distances}, \text{classprob} \)
  - \( RVS (10^6s): T_{\text{eff}}, \log g, [M/H], [X/M], \text{DIB, activity index} \)
  - Multiple values of some APs: tailor to your specific need (hot, cool,...)
- Extra-galactic parameterization:
  - \( (10^6s) \) Unresolved galaxy and quasar redshifts, classlabels
- Multi-dimensional tables:
  - MCMCs, self-organised maps, 2D extinction map
- Supplementary material on the following slides
Supplementary material
GSP-Phot provides $T_{\text{eff}}$, $A_G$, $\log g$, $[\text{M/H}]$, $E(Bp-Rp)$, $A_0$ from XP spectra and, using isochrones, distance, absolute magnitude and radius.

500M stars ($G<19$) $2500 < T_{\text{eff}} < 55000$

**Figure:** $\text{teff}\_\text{gspphot}\_\text{upper} - \text{teff}\_\text{gspphot}\_\text{lower}$ (uncertainties) as a function of $\text{teff}\_\text{gspphot}$ and colour-coded by observed $G$ magnitude. The contours denote density of sources.
General Stellar Parametrizer (GSP-Phot)

Comparison of $\log g_{\text{gspphot}}$ with $\log g$ derived from asteroseismology for **dwarfs** and **giants**
Extended Stellar Parametrizer: Emission Line Stars (ESP-ELS)

Goals: measurement of Hα pseudo-equivalent width, spectral type tagging (carbon star, M, K, G, F, A, B, O), classification of emission-line stars in terms of 7 (candidate) classes: Wolf Rayet (WC, WN), Be star, Herbig Ae/Be, T Tauri, dwarf Me, planetary nebula

Method: random forest algorithms trained on synthetic spectra and observations

GDR3 parameters: Hα pEW on observed spectrum, Hα pEW theo. on closest synthetic spectrum (GSP-Phot APs), class label (WC,WN, ...) and corresponding probabilities, spectral type tag (carbon star, M, K ...)

**Figure:** Effective temperature variation of the Hα pseudo equivalent width (pEW) of targets with and without emission. Below Teff 5000 K, the value measured on the closest synthetic spectrum was subtracted from the observed value to take into account blends with other spectral lines and molecular bands (e.g., TiO).
Extended Stellar Parametrizer: Hot Stars (ESP-HS)

**Method:** Fitting of synthetic spectra to XP+RVS or XP-only data.

**GDR3 parameters:** $T_{\text{eff}}$, $\log(g)$, $A_0$, AG, E(BP-RP), $[v\cdot\sin(i)]$, assuming $[M/H] = 0$

**Validity range:** $> 7500$ K, $\log(g)$: 2.5/3 - 4.75, $A_0 < 10$

**Expected number of sources:** ~ 3 million A, B, and O targets. Completeness ~ 70%. $G < 17.6$

- **Uncertainty $T_{\text{eff}}$ [K]:**
  - A: ~300 to 500
  - B: ~1000 to 2000
  - O: ~4000 to 7000

- **Uncertainty $\log(g)$ [dex, c.g.s]:**
  - ~0.2 to 0.3
  - ~0.2 to 0.4
  - ~0.4 to 0.7

Kiel diagram of open clusters. Surface gravities and effective temperatures were derived by using **RVS and XP spectra**, or by using **XP spectra only**. Cluster membership and age are taken from Cantat-Gaudin et al, 2018. The closest isochrone (Bertelli et al. 1994) is shown in cyan.
The ESP-UCD module derives effective temperatures (with uncertainties and flags) for ultra-cool dwarfs.

**Expected Number of sources**: $\sim 10^5$ (many of these sources are on the hotter side of UCDs)

**Validity range**: 500 - 2700 K

Caption: CMD of category 0 sources (best). The colour code reflects the $\text{teff}_{\text{espucd}}$ inferred by the module. The spectral types shown correspond to a calibration based on the Gaia UltraCool Dwarf Sample (GUCDS; left) and Stephens+ (2009, right)
Final Luminosity Mass Age Estimator (FLAME)

FLAME derives the radius, luminosity, mass, age, evolution stage, and gravitational redshift for stars processed by GSP-Phot and GSP-Spec down to G<18.25

Validity ranges and Numbers:
300M sources based on GSP-Phot APs
   Validity range: 2500 < $T_{\text{eff}}$ < 20000
5M sources based on GSP-Spec APs
   Validity range: 4000 < $T_{\text{eff}}$ < 7500

**Figure:** Comparison of $\text{radius}_{\text{flame}}$ (upper) and $\text{mass}_{\text{flame}}$ (lower) with Pinsonneault et al 2018 and Casagrande et al. 2011, resp.
Extended Stellar Parametrizer: Cool Stars (ESP-CS)

ESP-CS provides the chromospheric activity index from the Ca IRT for 2.5 million late-type stars

**Validity range:**
3000 < teff < 8000, 3.0 < logg < 5.0, -1 < [M/H] < +0.5

Figure: Comparison of `activityindex_espcs` derived from RVS spectra and FEROS spectra. A small bias on the order of 0.014 nm is seen, with a scatter of 0.015 nm. The different colours indicate whether the spectroscopic parameters were constrained by **GSP-Phot** or **GSP-Spec**
Multiple Star Classifier (MSC)

MSC provides stellar parameters of sources (teff, logg, luminratio, [M/H], A0, AG, distance) assuming a non-resolved binary in Bp and Rp spectra. Two parameters are derived for $T_{\text{eff}}$ and logg, e.g. teff_msc1, teff_msc2.

In GDR3 there are results on ~300M sources, with only ~10% being binary. The MCMC quality indicators are useful for finding reliable parameter estimates.

Figure: Comparison of $T_{\text{eff}}$ of primary component from MSC $\text{teff\_msc1}$ (y-axis) with the $T_{\text{eff}}$ from Traven+2020 (~10k GALAH sources, x-axis). Colour-code indicates quality of fit.
DSC: probabilistic classification

Probabilistic classification using XP spectra, colours, astrometry, and G-band variability of 1.8 billion sources into 5 classes: galaxy, quasar, star, binary, white dwarf, in the archive as `classprob_dsc_xxxxx`.

From this we extract a high completeness, low purity, sample of nearly 10 million quasars and galaxies, together with a class label (`classlabel_dsc_joint`) indicating a purer subset of 550k quasars and 250k galaxies.

<table>
<thead>
<tr>
<th></th>
<th>completeness</th>
<th>purity</th>
<th>completeness</th>
<th>purity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quasar</td>
<td>0.916</td>
<td>0.240</td>
<td>0.385</td>
<td>0.621</td>
</tr>
<tr>
<td>Galaxy</td>
<td>0.935</td>
<td>0.219</td>
<td>0.826</td>
<td>0.638</td>
</tr>
</tbody>
</table>

The table above shows the completeness and purity for quasars and galaxies when `classprob_dsc_combmod > 0.5` and also when using `classlabelDscJoint`.
Outlier Analysis (OA)

The OA module performs unsupervised classification using Self-Organized Maps and a statistical description/quality assessment/class labeling for the neurons (group of similar objects). It analyses objects that have no class probability > 0.5 for DSC classes (galaxy, quasar, star, binary, white dwarf).

36% of high quality neurons represent 54% of the sources, and were labeled using reference templates for different astronomical types of objects.

47% of intermediate quality neurons represent 35% of the sources, present heterogeneous spectra, typically quasars.

17% of the neurons were assigned a low-quality index and they represent 11% of the sources, which have a few number of transits and very heterogeneous spectra, such as artifacts.
QSOC predicts the redshift ($z$) for $\sim 1.8$M qsos
Validity range: $0 \leq z \leq 6$

81% of the $G < 20$ mag quasars have $|\Delta z| < 0.1$
Resp. 96.5% if flagQsoc = 0

Figure: Comparison of the redshift as predicted by QSOC against values from the literature for $17 \leq G \leq 19$ mag quasars. The logarithmic redshift scale highlights the common degeneracies existing in the prediction of the redshift that are mostly due to mismatches between emission lines. The presence of a quality flag, flagQsoc, allows to mitigate the effect of these degeneracies without affecting too much the central peak.
The UGC module derives the galaxy redshift of non-resolved galaxies for ~1.1M

Validity range: 0.0 ≤ z ≤ 0.6

UGC implements Support Vector Machine (SVM) models using Gaia BP/RP spectra. Trained with ~6K Gaia observed “reference” galaxies with known redshifts (SDSS DR16).

UGC results validated against ~300K SDSS DR16 galaxies cross-matching Gaia observed sources. The standard error of the redshift prediction is ±0.037
"Thank you for your attention"