189. More metallicities and gravities

SINCE Data Release 3 in 2022 (which included a catalogue of metallicities, temperatures and gravities for 470 million sources), four community-generated catalogues of [M/H], T_{eff}, and $\log g$ have been made available, which I will describe here. To understand their context, let me start with some background.

ALONG WITH THE main astrometric field, and the radial velocity spectrometer (RVS) field, the Gaia focal plane includes two low-resolution objective prism fields, BP (blue photometer) and RP (red photometer), together generally denoted XP (essay 68). The prisms disperse the spectra over ~45 pixels in the along-scan direction, with the 60-pixel window allowing for sky subtraction. The spectral dispersion varies from 3–27 nm per pixel over 330–680 nm for BP, and from 7–15 nm per pixel over 640–1050 nm for RP.

For the majority of objects, $G \gtrsim 11.5$ mag, the BP and RP spectra are binned on-chip in the across-scan direction, over 12 pixels, to form one-dimensional, along-scan spectra. The dispersed spectra overlap in crowded regions, and BP/RP acquisition is accordingly limited to about 750 000 objects per square degree. The spectra are wavelength and flux calibrated, while their integrated fluxes provide calibrated multi-epoch photometry, designated G_{BP} and G_{RP}.

The 34-month Data Release 3 (essay 76) then comprises 1.812 billion sources with astrometric solutions to 21 mag, 1.806 billion with mean G magnitudes, 1.5 billion with mean G_{BP} and G_{RP} photometry, along with 219 million mean BP/RP spectra.

The task of source classification (as star, white dwarf, physical binary, quasar, or galaxy), and the 'astrophysical parameters inference system' (Apsis) is carried out by Coordination Unit 8 of DPAC, and I have given some details (and references) in essay 89. The 13 Apsis modules take relevant combinations of the RVS spectra (GSP–Spec) and BP/RP spectra (GSP–Phot), fit a number of astrophysical parameters to the data, and pass the results to the FLAME module which derives the evolutionary parameters radius, luminosity, mass, and age.

GSP–Phot uses the XP spectra to estimate T_{eff}, $\log g$, [M/H], M_{G}, radius R, distance, and extinctions by modelling the BP/RP spectra, G magnitude, and parallax (Andrae et al., 2023a; Creevey et al., 2023, §3.5). To match the XP spectra, Andrae et al. (2023a) used four stellar atmospheric models to cover different ranges of T_{eff}, the Fitzpatrick (1999) mean extinction law, and a grid of (PARSEC) isochrones to fix the absolute magnitudes.

As a result, astrophysical parameters in DR3 include T_{eff}, $\log g$, and [M/H] for 470 million sources using BP/RP (and 6 million using RVS); radius (470 million), mass (140 million), age (120 million), and spectral types (220 million), along with smaller catalogues of chemical abundances, diffuse interstellar bands, activity indices, Hα equivalent widths, and emission-line stars.

IN AN IMPORTANT application of the XP spectral database, Montegriffo et al. (2023) provided the tools to generate synthetic photometry, for all 220 million sources with mean BP and RP spectra, in any user-specified photometric system (see essay 187).

They also made available an associated Gaia Synthetic Photometry Catalogue (GSPC) which contains, for the majority of the 220 million stars with XP spectra from DR3, colours in 13 passbands. These include UBVRI in the Johnson–Kron–Cousins system, *ugriz* in the SDSS system, and two in the HST–ACS/WFC system.

LET ME EMPHASISE the size of the Gaia XP spectral database: there are more than 20 times as many as in the largest ground-based spectral survey, LAMOST, although at only 1/20th of the spectral resolution.

And the metallicities published with DR3 can still be improved, in part through improved calibration, more so through additional observations as planned for DR4, and also through better modelling of some key spectral lines given the low resolution of the XP spectra. Furthermore, the use of synthetic model spectra (to match the observed XP spectra) is most likely sub-optimal, leaving parameter estimates sensitive to inaccuracies and omissions in the underlying models.

THE ASTROPHYSICAL parameters generated by DPAC CU8 were never intended be the final word in classification or parameter estimation, and improvements in calibration, and in training sets and algorithms were considered inevitable.

The motivation to construct better metallicities is, of course, based on the pursuit of accurate chemical abundances across all stellar populations throughout the Galaxy, which are crucial inputs for studies of star formation, detailed nucleosynthesis modelling, Galactic chemical and dynamical evolution, and so on.

And already, since 2023, four other large catalogues of astrophysical parameters, focusing on [M/H], have been derived from the XP spectra, as follows:

ZHANG ET AL. (2023) used forward modelling to estimate $T_{\rm eff}$, $\log g$, and [Fe/H] for all 220 million DR3 stars with XP spectra. Their training set used atmospheric parameters from LAMOST, augmented by 2MASS and WISE photometry to reduce degeneracies and yield more precise estimates of $T_{\rm eff}$ and reddening. Their catalogue includes $T_{\rm eff}$, $\log g$, and [Fe/H], along with revised parallaxes and extinctions. It ignores binary stars, and does not cover all parts of the Hertzsprung–Russell diagram, notably white dwarfs.

ANDRAE ET AL. (2023b) building on previous work by Rix et al. (2022), employed a specific machine-learning algorithm, XGBoost. It was trained on 500 000 stars with stellar parameters from APOGEE, including those with CatWISE 3.4 μm and 4.6 μm infrared photometry to reduce the degeneracy between $T_{\rm eff}$ and reddening. The training set was augmented by some 300 very metal-poor stars from LAMOST (Li et al., 2022), and they included the Gaia parallaxes to assist constraints on $\log g$ and [M/H].

Although therefore tied to the parameter scale of the APOGEE survey, the resulting catalogue of 175 million stars has a mean precision of 0.1 dex in [M/H] and, obtained as by-products, 50 K in $T_{\rm eff}$, and 0.08 dex in $\log g$. They also provide a catalogue of 17 million bright ($G < 16$) red giants using more conservative cuts to ensure a higher data quality.

Chandra et al. (2023) describe an application of this catalogue in identifying the evolution of angular momentum in the Galaxy with metallicity.

HATTORI (2024) used tree-based machine-learning to estimate [M/H] and [α/Fe] for 48 million giants and dwarfs in low-extinction regions from the DR3 XP spectra. Again, the training set used APOGEE DR17 and the metal-poor stars of Li et al. (2022). It resulted in a mean precision of 0.09 dex for [M/H] and 0.04 dex for [α/Fe], with the most reliable values being for giants and metal-rich stars which dominate the training set.

YAO ET AL. (2024) focused on the much rarer very metal-poor stars, [Fe/H] < -2, also using XGBoost. For $G_{\rm BP} < 16$, they developed classifiers optimised for turn-off stars and for giant stars, finding 11 000 metal-poor turn-off stars, and 111 000 or 44 000 bright metal-poor giants depending on the target purity. For $G_{\rm BP} > 16$, they identified 38 000 additional turn-off candidates, and 41 000 additional metal-poor giant candidates.

Investigations by Witten et al. (2022), meanwhile suggest that metal-poor stars can be identified for $G < 16$ using the XP spectra, but that true detections will be overwhelmed by false positives at fainter magnitudes.

THE XP DATA will provide a magnificent resource for chemical abundance investigations. But it is also a substantial database for research into classification algorithms, and the trade-offs between 'physics-driven' (relying on synthetic stellar spectra) and 'data-driven' (based on machine-learning) classification.

Some of this path has already been trodden, for example by APOGEE (e.g. Ness et al., 2015; Leung & Bovy, 2019; Ting et al., 2019).

As a cautionary example, at least some algorithms which estimate [α/Fe] from the XP spectra do so by exploiting known correlations between [α/Fe] and other elements, rather than the direct effect of [α/Fe] on the spectrum (e.g. Gavel et al., 2021; Hattori, 2024).

THOSE INVOLVED in these kinds of studies will appreciate the contribution of Laroche & Speagle (2024). They argue that physics-driven models inevitably suffer from a 'synthetic gap', viz. a combination of theoretical and instrumental effects which together produce unresolvable differences between synthetic and observed spectra. At the same time, data-driven models which depend on 'labels' (a generic term here covering $T_{\rm eff}$, $\log g$, [M/H], and [α/Fe]) themselves suffer from 'label systematics' which decrease any model's performance.

They demonstrate this by applying a variational auto-encoder (unsupervised learning) to the XP spectra which (they argue) learns stellar properties directly from the data. They also show that the spectra *do* contain meaningful [α/Fe] information, by identifying α-bimodality in the absence of stellar label correlations.

They conclude: *'Label-dependent models are incapable of exploiting the entire astrophysical information in the XP data, because they are limited by the availability of stellar labels to train on. Novel data-driven techniques must be developed to tackle this big data problem'.*

Using the same techniques used by Large Language Models for AI, and applied to the XP spectra, Leung & Bovy (2024) argue that *'building and training a single foundation model without fine-tuning using data and parameters from multiple surveys to predict unmeasured observations and parameters is well within reach'.*