SoS: the Survey of Surveys

homogenization of stellar parameters across large spectroscopic datasets

Authors: **Aleksandra Avdeeva, Elena Pancino, Alessio Turchi, StarDance Team**
INAF - Osservatorio Astrofisico di Arcetri

Prologue: Many voices in the sky

Large spectroscopic surveys such as APOGEE, GALAH, LAMOST, RAVE and Gaia RVS have measured all together millions of stars, providing estimates of their temperatures, surface gravities and chemical abundances. However, each survey uses different instruments and analysis pipelines and, as a consequence, their results are not always directly comparable. To study stars from different surveys together, we need to place their measurements onto a common scale.

In this work, we present a method that achieves this in several steps: first, we define a reference scale using large and well-characterized surveys; then we map additional surveys onto this scale; and finally, we validate the results using reliable benchmark stars.

The result is **a unified catalog** that enables consistent analysis of stellar populations and provides a clearer view of the structure of our Galaxy.

Act I: Building the bridge

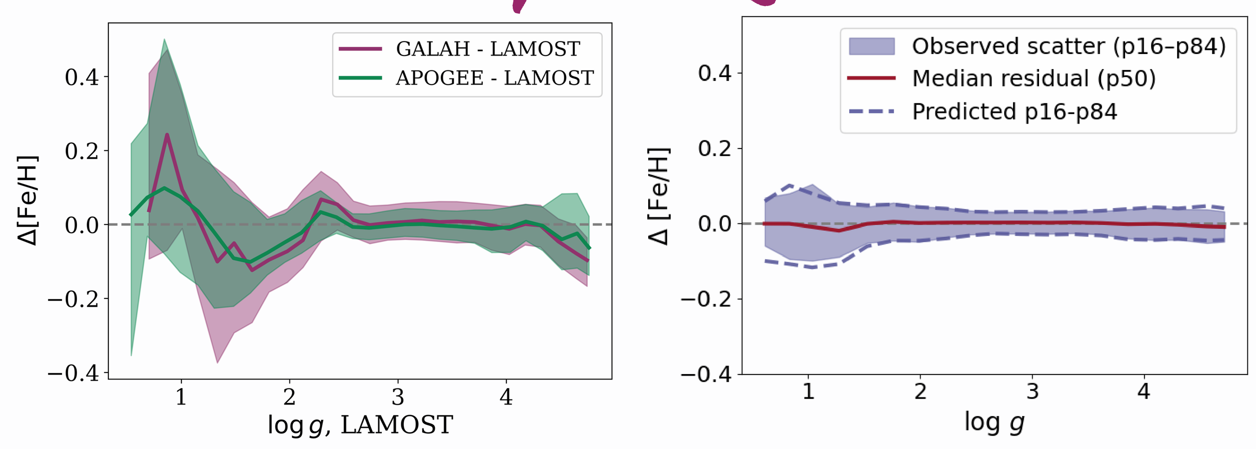

Figure 1: Systematic offsets in [Fe/H] are removed after calibration, bringing LAMOST onto the APOGEE+GALAH scale.

We begin with three major spectroscopic surveys – **APOGEE**, **GALAH**, and **LAMOST** – which have substantial overlap in targets. Aligning these datasets onto a common scale provides the foundation for further expansion. Stars observed in multiple surveys allow direct parameter comparisons (**Fig. 1**, **left panel**).

For the reference system, we adopt the APOGEE+GALAH scale for logg and [Fe/H], and the APOGEE scale for Teff, with GALAH and LAMOST adjusted accordingly. Systematic differences in Teff, logg, and [Fe/H] are modeled using **XGBoost** [1] **with pinball loss**, and the resulting predictions are calibrated with **MAPIE** [2] to obtain reliable uncertainties (**Fig. 1**, **right panel**). We monitor both global agreement and residual trends, balancing calibration accuracy with model stability to avoid overfitting.

After correction, survey-to-survey comparisons show consistent parameters with minimal residual trends, forming the basis of the reference scale. The Kiel diagram of the unified dataset shows a clear improvement compared to the individual survey data (**Fig. 2**).

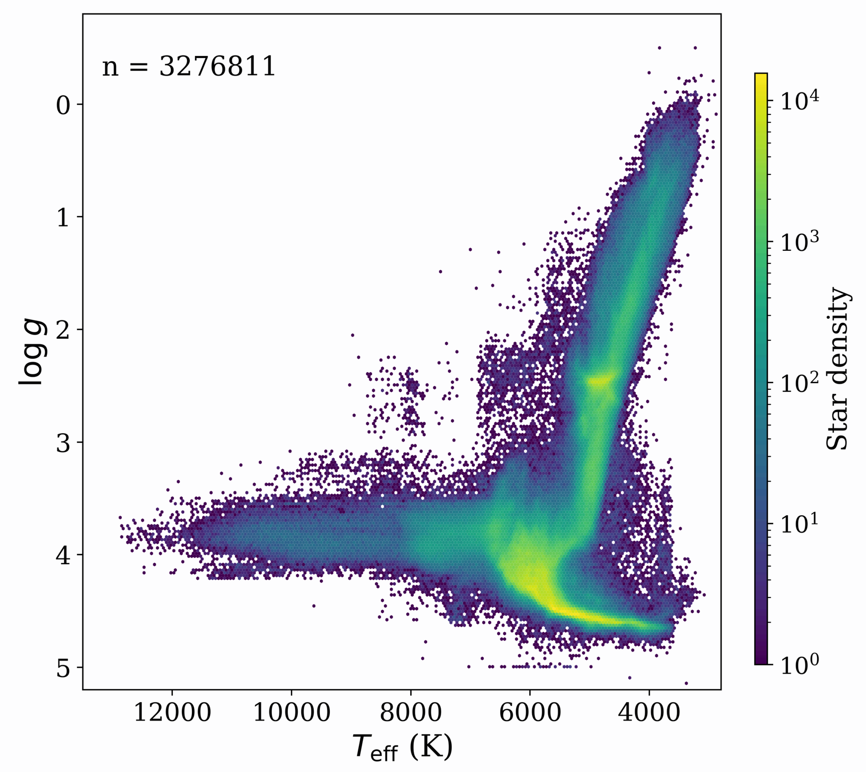

Figure 2: The calibrated reference scale shows clean evolutionary sequences, compared to pre-calibration scatter.

Act II: Expanding the conversation

With the reference scale in place, we can now extend it to additional surveys. **Gaia RVS**, **RAVE**, and **LAMOST MRS** are aligned using stars observed in common with the reference set during this step.

The same XGBoost regressor is applied to model the systematic differences and propagate the corresponding corrections to these datasets.

Each new connection expands the reference scale, ensuring that parameters derived from different instruments and analysis pipelines can be compared consistently.

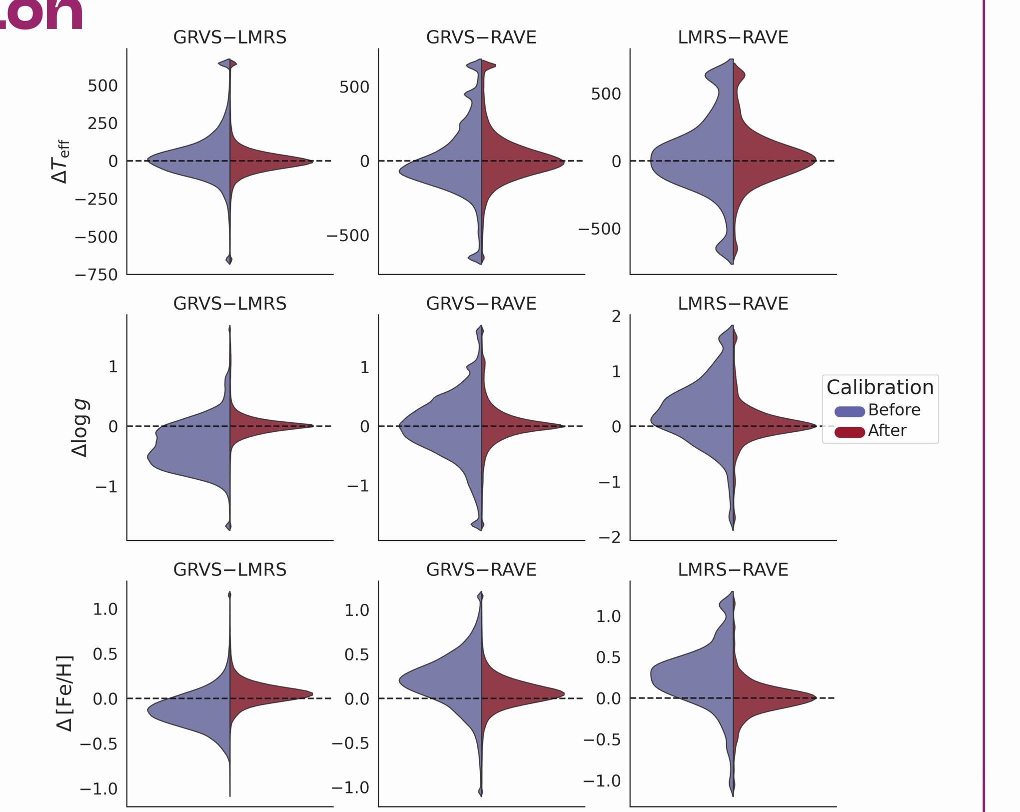

Figure 3: Residual differences between Gaia RVS, LAMOST MRS, and RAVE are reduced after calibration.

Interlude: Error propagation

Different surveys report uncertainties in different forms, so we propagate them consistently using Monte Carlo sampling: Gaussian distributions for symmetric errors, multivariate Gaussians for correlated parameters, and split-normal distributions for asymmetric percentile-based values. In regions where surveys overlap, multiple measurements are combined by minimizing the split-normal negative log-likelihood, and asymmetric 68% credible intervals are derived from the points where ΔNLL = 0.5.

Act III: Testing the harmony

After calibration, we test internal consistency by comparing each overlapping survey pair (LAMOST MRS, Gaia RVS, RAVE) in Teff, log g, and [Fe/H]. We assess residual scatter and trends as a function of stellar parameters, S/N, and magnitude, and verify continuity in the Kiel diagram, ensuring that evolutionary sequences remain consistent across surveys.

These tests show that the surveys now lie on a unified scale with improved internal agreement.

Fig.3 illustrates the increased coherence in the calibrated parameters; stars used for training are excluded from this comparison. **Fig.4** presents the resulting Kiel diagrams of the calibrated surveys.

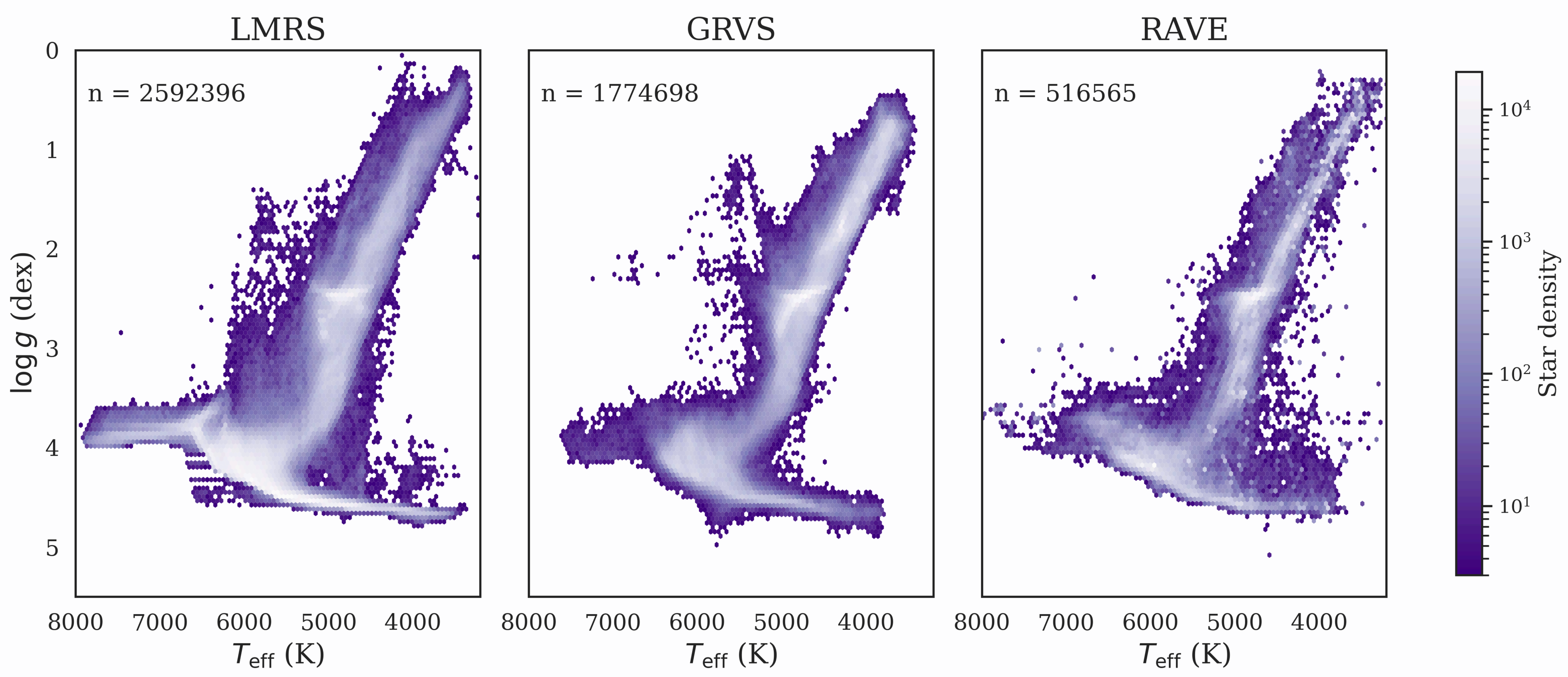

Figure 4: Calibrated surveys show consistent Kiel diagrams.

Epilogue: Looking forward

With internal consistency established, the next step is to **extend** the unified scale **to smaller** and **more specialized spectroscopic datasets**. Many of these surveys have limited direct overlap, so they will be incorporated through intermediate links to the existing reference. This gradual process expands the calibrated parameter space and accommodates a wider range of spectral resolutions and stellar types. Each new dataset added in this way strengthens the framework and moves us closer to a complete and coherent view of the Milky Way.

The remaining challenge is **external consistency**: validating the scale against benchmark stars and well-studied stellar clusters to ensure astrophysical reliability beyond internal comparisons.

Sources

[1] Chen, T. & Guestrin, C. (2016). "XGBoost: A Scalable Tree Boosting System". arXiv:1603.02754.
[2] Cordier, T. (2023). "Flexible and systematic uncertainty estimation with conformal prediction via the mapie library". Conformal and Probabilistic Prediction with Applications. PMLR, 2023.

European Research Council
Established by the European Commission

Our latest catalogue of 23 million stars (SoS-Spectro and SoS-ML) is already available online and can be found here →