Research in Astronomy and Astrophysics, 25:035019 (10pp), 2025 March
https://doi.org/10.1088/1674-4527/adba29
CSTR: 32081.14.RAA.adba29

XGboost for Predicting Chemical Abundances of Open Clusters in LAMOST DR11 LRS

R. Zhang[1], Jianxing Chen[2], Chengyuan Wu[1], Luqian Wang[1], Shuai Zha[1], Sufen Guo[1,3], and Zhanwen Han[1]
[1] Yunnan Observatories, Chinese Academy of Sciences, Kunming 650216, China; ruyuan.zhang@studenti.unipd.it, wuchengyuan@ynao.ac.cn
[2] School of Physics and Astronomy, Beijing Normal University, Beijing 100875, China; jxchen@bnu.edu.cn
[3] School of Physical Science and Technology, Xinjiang University, Urumqi 830046, China
Received 2025 January 14; revised 2025 February 21; accepted 2025 February 24; published 2025 March 24

Abstract

The 11th data release of the LAMOST survey provides fundamental stellar parameters but lacks detailed abundance measurements for α-elements, which are crucial for understanding stellar populations and Galactic chemical evolution. In this study, we derive the abundances of oxygen (O), magnesium (Mg), silicon (Si), calcium (Ca), and titanium (Ti) for over 6.8 million stars using LAMOST DR11 low-resolution spectra. To ensure reliable measurements, we select 760 open clusters spanning a broad range of ages and apply Monte Carlo sampling for accurate abundance estimates. Additionally, we utilize over 30,000 stars from the GALAH DR4 catalog to train an XGBoost model for extracting α-element abundances from LAMOST DR11 spectra. Bayesian linear regression is employed to analyze the compositional distribution across the Galactic disk and infer chemical gradients as a function of Galactocentric distance. Our results indicate a general increase in chemical abundances with Galactocentric distance, with oxygen showing the steepest gradient. Our results confirm the overall increase of α-element abundances with Galactocentric distance, consistent with previous studies, while minor discrepancies in Mg, Ca, and Ti gradients likely arise from differences in sample selection, observational sensitivity, or Galactic enrichment processes.

Key words: stars: abundances – stars: atmospheres – Galaxy: abundances – Galaxy: disk – (Galaxy:) open clusters and associations: general

1. Introduction

The study of galaxy formation and evolution, particularly with respect to the Milky Way, is a cornerstone of modern astrophysics. Investigating fundamental stellar properties, such as effective temperature, surface gravity, metallicity, and elemental abundances, offers critical insights into the structure and evolutionary history of our Galaxy. The advent of large-scale spectroscopic surveys—such as the Large Sky Area Multi-Object Fiber Spectroscopic Telescope (LAMOST[4]; Zhao et al. 2012), the Sloan Extension for Galactic Understanding and Exploration (SEGUE; Yanny et al. 2009), the Apache Point Observatory Galactic Evolution Experiment (APOGEE; Majewski et al. 2017), and Galactic Archaeology with HERMES (GALAH[5] De Silva et al. 2015)—combined with Gaia's astrometric and photometric data has facilitated precise measurements for millions of stars. This vast data set provides an opportunity for detailed studies into the chemical evolution and dynamical processes in our Milky Way. Currently seeking answers to critical questions such as: how Galactic disks evolved, where stars are born, how stars migrate, how chemical elements originate in stars, and so on. In these studies, open clusters (OCs) are of prime importance as they act as a benchmark for stellar formation and evolution, helping to understand methods of assembly and development of the Milky Way.

OCs are collections of stars that are bound by gravitation and have formed from a single molecular cloud. While most OCs are less than 1 Gyr, some of the oldest clusters may survive for up to 8 Gyr before dispersion (Fujii & Portegies Zwart 2016). OCs are special because the stars therein tend to be of similar ages, chemistries, and distances, allowing for better estimations of their ages, metallicity, and chemical abundances. Distance determination is usually through main-sequence and isochrone fitting, with an additional constraint arising from the spatial concentration of the cluster members. Since all stars within an OC are formed from the same interstellar medium (ISM), detailed understanding of their metallicity and elemental composition can be gleaned from spectroscopic knowledge. OCs mainly comprise the Galactic disk and generally have solar-like metallicities. Gradually, these clusters tend to migrate away from the Galactic plane due to vertical heating processes

[4] https://www.lamost.org/
[5] https://www.galah-survey.org/; The GALAH data release includes high-resolution spectra with a resolving power of 28,000, spanning four distinct wavelength ranges: 4713–4903 Å, 5648–5873 Å, 6478–6737 Å, and 7585–7887 Å.

driven by interaction with Galactic structures (Piskunov et al. 2006; Sharma et al. 2021). The Gaia mission has immensely helped in identifying OCs and characterizing them; now over 7000 clusters are described in the literature. Recent studies, such as those by Chi et al. (2023) and Hunt & Reffert (2024), have further refined OC catalogs and examined the gravitationally bound nature of these stellar systems.

The distribution of chemical elements across the Galactic disk serves as a critical constraint for chemo-dynamical models of the Milky Way, with numerous studies emphasizing metallicity and elemental abundance ratios (Donor et al. 2020; Myers et al. 2022). While field stars have traditionally been the focus of such studies, their migration complicates the interpretation of chemical distributions. In contrast, OCs, being younger and less dynamically evolved, offer a more reliable tracer of the gradients in the disk where new stars form. The metallicity gradient derived from OCs is approximately $-0.060\,\mathrm{dex\,kpc^{-1}}$, though it flattens in the outer regions of the disk (Friel et al. 2010).

Beyond metallicity, the investigation of abundance ratios for elements such as α-elements, iron-peak elements, and neutron-capture elements provides deeper insights into nucleosynthesis processes occurring in diverse stellar environments. α-elements are key tracers for evaluating the production sites and timescales that depict the complex pathways of the Milky Way's chemical evolution. Recently, there have been works emphasizing OC migration to explain the observed trends predicted in chemo-dynamical models (Myers et al. 2022). Such observations highlight some of the complicated interactions that shape the Milky Way's evolutionary history.

The LAMOST project is a giant-scale astronomical survey launched to gather the largest spectral data of stars. The survey is performed in two modes, namely low-resolution spectroscopic survey (LRS), and medium-resolution spectroscopic survey (MRS). The LRS accounts for the larger share of the LAMOST observations and operates at a resolution of $R \approx 1800$, whereas the MRS, which was introduced in 2018, gets up to $R \approx 7500$. The surveys are conducted every year in the months from September to around June 15, with a summer break allocated for maintenance. The main aim of LAMOST spectroscopic survey is to study stars, the Milky Way, and galaxies, bringing significant progress to understand star populations and Galactic evolution. So far, LAMOST has released over 17 million spectra to the public, including new data from the latest observation seasons together with previously reprocessed data updated with the latest version of the LAMOST Stellar Parameter Pipeline (LASP). Moreover, there have been some very specific pipelines, like LSP3 (Xiang et al. 2015b) implemented at Peking University and SPAce (Boeche et al. 2018), including techniques customized for certain studies. More information about the data pipeline can be found in (Luo et al. 2015). The recent 11th data release (DR11) has added new data coverage with significant improvements. However, still with such a vast volume of data, LAMOST has too few precise measurements of chemical abundances, particularly for many elements untraceable. This points to the fact that there is still room for development of such a mode with additional ranges and abundances that could feed back into the work on studies of stellar populations together with Galactic chemical evolution.

LAMOST has successfully completed the first extensive spectroscopic survey of the Galactic anticenter disk, utilizing a well-defined target selection strategy based on systematic magnitude criteria. Strong collaborations with Gaia and other surveys have put many pieces back together about the Milky Way disk in terms of the basic parameters (Tian et al. 2018; Wang et al. 2019), stellar populations (Coronado et al. 2020; Yu et al. 2021), chemo-dynamical evolution (Xiang et al. 2015b), kinematics (Wu et al. 2021), and metallicity gradients (Huang et al. 2015; Vickers et al. 2021). Open clusters have all, meanwhile, been mapped by Fu et al. (2022) to reveal a threefold spread in [Fe/H] for nearby young clusters, calling into question rapid metal enrichment by efficient star formation, or uneven cloud sequence mixing giving rise to metallicity patches. Zhang et al. (2024) conducted another LAMOST DR8 LRS study covering 1131 OCs, under which there is practically no metallicity gradient with respect to many of the predictions modeled. Actually, their study hints that most OCs just close to the Sun were formed in the outer disk area, whereas a younger OC in the inner disk may have been disrupted via tidal interactions.

Spectroscopy from large-scale surveys faces big challenges in establishing the stellar labels with great precision and efficiency, usually affected by the caveats of classical model fit techniques (Nissen & Gustafsson 2018; Jofré et al. 2019). Noteworthy is that machine learning provides certain edge in efficiency, accuracy, and scalability while searching for complex patterns that go far beyond the limits of traditional methods. For example, Xiang et al. (2015b) utilized the DD-Payne method to monitor LAMOST DR5 low-resolution spectra, successfully producing stellar parameters and elemental abundances of over 8 million stars, with uncertainties between 0.05 and 0.3 dex. The obtained results were validated against the GALAH and APOGEE data sets, with confirmation of their reliability. Similarly, Li & Lin (2023) developed StarGRUNet, a neural network model that significantly enhances the precision and robustness of stellar parameter estimations from LAMOST-APOGEE data. These advancements illuminate the path for machine learning to transform spectral analyses of stellar populations and to understand our Milky Way structure better. Building on these advancements, this paper applies machine learning techniques to LAMOST DR11 data, using OCs as tracers to investigate the chemical distribution and evolutionary features of the Galactic thin disk.

The structure of this paper is as follows: Section 2 describes the sample selection, reference data, and membership

determination of OCs establishing the foundation for this study. Section 3 provides a comprehensive overview of the LAMOST data reduction process and the analysis of stellar parameters, which are essential for subsequent chemical abundance determinations. In Section 4, the XGBoost algorithm is introduced and applied to estimate the abundances of α elements in the LAMOST DR11 LRS catalog, demonstrating its efficacy for large-scale stellar abundance analyses. Section 5 examines the spatial distribution of chemical compositions within the Galactic disk, exploring the correlations between elemental abundances, Galactic structure, and evolutionary history. Finally, Section 6 summarizes the key findings and implications of this work.

2. Samples

The Gaia satellite has revolutionized OCs research, greatly enhancing the census of these clusters and our understanding of their physical and dynamical properties. Gaia has made significant contributions in several key areas since its first data release (Gaia Collaboration et al. 2018). Its high-precision astrometry enabled the identification and exclusion of asterisms erroneously designated as OCs, thus contributing positively toward the reliability of OC catalogs (Cantat-Gaudin et al. 2020b; Hunt & Reffert 2023, 2024). In parallel, Gaia enabled the discovery of thousands of new clusters augmenting the previously known OC population revealing comparatively fainter objects in obscured regions (Liu & Pang 2019).

Besides discovery, the high-precision data from Gaia allowed the direct estimation of important parameters about clusters such as their age, distance, and metallicity. Such advances have contributed profoundly toward our understanding of the structure, formation, and evolution of the Galactic disk (Bossini et al. 2019; Cantat-Gaudin & Anders 2020a). In combination, these results delineate Gaia's transformation in stellar cluster research, posing new markers in Galactic studies.

A recent study by Hunt & Reffert (2024) refined the classification of 6956 star clusters by calculating Jacobi radii-photometric masses permitting a clear distinction between gravitationally bound and unbound clusters. The work thus resulted in the largest catalog of Milky Way cluster masses yet prepared. It found that 79% of the clusters are gravitationally bound, and a fair number of nearby clusters were discovered to be unbound moving groups. Age functions, mass functions, and completeness of the OC census were improved in this study: it estimates that the Milky Way contains approximately 1.3×10^5 OCs, of which only about 4% have been identified. This result further indicates that a great many clusters still await their discovery in the Galaxy.

For this study, a total of 7435 stars from 1086 OCs in Hunt & Reffert (2024) were cross-matched with LAMOST DR11 LRS data for final sample construction. It contains important dynamical properties of OCs such as Galactocentric distance ($R_{\rm GC}$), radial velocity, age, and cluster extinction. This full data set provides a substantial basis for areas such as chemical abundances and dynamical characteristics of OCs due to the combination of these particular properties with spectroscopic data. As such, it gives new insight into the contribution of OCs as tracers of Milky Way structure and evolution.

3. Dataset

The LAMOST Stellar Parameter Pipeline is a robust system designed for determining the primary stellar atmospheric parameters such as effective temperature ($T_{\rm eff}$), surface gravity ($\log g$), and iron abundance ([Fe/H]) from low-resolution spectra. The data provide a basis for a fairly standard sequence of events: data reduction, where raw data are used to extract one-dimensional spectra, is corrected for instrumental and observational effects, and is co-added from the multiple sub-exposures in order to improve the signal-to-noise ratio (SNR).

The initial stellar parameters are delivered by the CFI method (Du et al. 2012), which uses correlations between certain observed spectra and pre-computed stellar model templates. Then the previous estimates are iteratively refined using a χ^2 minimization fitting procedure within the ULySS framework (Wu et al. 2011). Such a double-step implementation guarantees clarity and reliability in the derived parameters essential to research in stellar populations and Galactic chemical evolution. The spectral templates utilized in LASP are also documented in Wu et al. (2011).

LASP may excel in obtaining essential parameters of stars, but very low spectral signal-to-noise ratios often become limiting factors to that achievement and accuracy. Li & Lin (2023) and recent approaches manage to improve on estimates of atmospheric parameters, while most do address star's basic properties. Here we extend this method to include abundances of five α-elements, thus enhancing cosmic characterization of stellar parameters and contribute toward improved knowledge of Galactic chemical evolution.

In this paper, we combine three data sets: Gaia DR3, LAMOST DR11 LRS, and GALAH DR4, the first two of those give estimates of stellar atmospheric parameters, while the remnants can be employed efficiently for describing abundances of alpha-elements. To counterbalance these limitations, elemental abundance data from GALAH DR4 were added to further the analysis. Such an integrative approach has enabled us to refine the understanding of open cluster chemical abundance and gain insight into their better utility as tracers of Galactic structure and evolution.

4. Stellar Composition

This research discusses the abundance patterns of the α-elements O, Mg, Si, Ca, and Ti, exclusively measured by GALAH but not available in the LAMOST DR11 LRS catalog.

These elements are crucial in tracing stellar nucleosynthesis and provide a new insight into the formation and evolution of star populations and the Galaxy itself.

In alleviating the gap that exists between two surveys in the determination of α-element abundance values, this study's aim is to predict the α-element abundances existing in LAMOST DR11 values with respect to the corresponding measurements already available in GALAH. However, the correct matching in spectroscopic type and catalog parameters has been constrained by the difficulties that arise from countervailing evidence from differences between the instrumentation, calibrations, and quality between these two surveys, thus forcing their crossing of each other.

To resolve discrepancies, machine learning techniques were used in predicting α-element abundances in LAMOST DR11 catalog. In this work, this research employs the XGBoost algorithm, because of its adeptness in handling nonlinear correlations, features interactions, and high-dimensional data integration, already employed in previous astrophysical studies (Spina et al. 2021). At every model iteration in predicting overlapping data sets, high-resolution abundance measurements from the GALAH survey were iteratively used to refine predictions for LAMOST to improve the final accuracy and reliability of the abundance estimates.

This methodology presents synergy between the GALAH and LAMOST data sets to broaden the spatial information and possibly enhance the richness of chemical abundance data. By providing significant predictions on α-elements into the LAMOST DR11 LRS catalog, the present work lays a far-reaching basis for Galactic chemical evolution studies.

4.1. Model Structure

XGBoost-eXtreme Gradient Boosting-is a super-fast, scalable machine learning algorithm that builds an ensemble of decision trees by applying a gradient-boosting algorithm. This means that each new one is designed almost exclusively to correct the residual errors of the preceding ones, thereby increasing the model robustness and predictive accuracy. A stronger point for XGBoost is that it incorporates many advanced regularizer functionalities preventing overfitting and consequently improving the model level of generalization.

In astronomical data analysis, XGBoost has been demonstrated to be truly efficient in large, high-dimensional data sets. It shows great performance in stellar parameterization and classification because of its data structure management, handling missing values, and overfitting reduction. Such strengths make it a unique choice for the intrinsic nature of astronomical data-related problems because they allow for the most accurate and efficient analyses for myriad astrophysics applications.

XGBoost full mathematical form is built on three major components: the objective function, which contains the regularizer part, the scheme for calculating leaf-node weights, and the optimization algorithm, which utilizes second-order Taylor expansion and a greedy approach for selecting the best split points. The objective function, in this combination of loss function with a regularizer, states the complete optimization task and can be written as follows:

$$\mathcal{L}(\theta) = \sum_{i=1}^{n}(y_i - \hat{y_i})^2 + \lambda\sum_{t=1}^{T}\Omega(f_t)$$

where:

1. y_i is the measured abundances from GALAH and $\hat{y_i}$ is the predicted value.
2. The first term $\sum_{i=1}^{n}(y_i - \hat{y_i})^2$ is the Mean Squared Error (MSE) loss function, which quantifies the difference between the predicted and measured abundances.
3. $\Omega(f_t)$ is the regularization term, which penalizes the model's complexity to prevent overfitting.

An objective function or loss function measures how much the two numbers actual and predicted values are differing from each other. This guides the optimization of the model since this minimization of all errors leads to better predictive performance. The XGBoost regression model was used for the abundance homogenization of data from GALAH and APOGEE using MSE as the loss function. To mitigate overfitting and thus enhance generalization, the model is provided with the regularization term. It adorns the model's complexity with a penalty, and it is usually defined as follows:

$$\Omega(f_t) = \gamma T + \frac{1}{2}\lambda\sum_{j=1}^{T} w_j^2$$

where:

1. T is the number of leaves in the tree, and w_j is the weight of the jth leaf.
2. The parameters γ and λ serve as regularization terms, governing the model's complexity.

During training, the objective function is optimized by minimizing it, with the model parameters adjusted to enhance accuracy and prevent overfitting.

4.2. Model Training

During the process of training, the model makes use of observational data from GALAH, using $T_{\rm eff}$, $\log g$, and [Fe/H] as input features, while the target values are defined as chemical abundance corrections. The model adjusts its internal parameters by minimizing prediction errors so as to be able to learn the underlying relationships between the input features and target outputs. Models were developed in isolation for every target element, with a separate model trained for the prediction of the abundance of each element.

The training set was selected following the conditions suggested by Spina et al. (2021), which required temperatures to be between 4000 and 7000 K, surface gravity greater than 0 dex, and metallicity between −1 and 0.5 dex. The microturbulence had to be less than 2.5 km s^{-1}. To reduce discrepancies perceived between LAMOST and GALAH observations, strict checks for consistency were made to ensure data quality for model training. The range of acceptability for the difference in effective temperature was set at less than 150 K, the difference in surface gravity at less than 0.3 dex, and the metallicity difference was limited to within 0.1 dex. Moreover, GALAH's chemical composition was fixed to the range of [−0.5, 0.5] dex. For the Ca abundance, we adopted the range of abundances reported in (Yaz Gökçe et al. 2017), which spans approximately from −0.1 to 0.35 dex. To ensure high data quality and adequate sample size for comparison, we selected spectra from GALAH with S/N > 50 and from LAMOST with S/N > 30 as the cutoff criteria. After applying these thresholds, the final sample comprised 27,712 targets.

A grid search was performed to tune the parameters of the model using iron abundance normalization ([Fe/H]) as a reference. This approach aims at improving the consistency and reliability of abundance comparisons across data sets. This model employed the Mean Absolute Error (MAE) as a loss function and carried out some 5-fold cross-validation to add an element of stability to the outcomes, accuracy, and interpretability. This fairly reduces the effects of noise and outliers hence ensuring credible performance. The data were separated into training and testing sets of 75% versus 25% in order to carry out a proper analysis.

Figure 1 presents the relationship between the predicted and observed abundances, highlighting the model's performance. The MAE for the five α-elements ranges from 0.04 dex to 0.10 dex, demonstrating the model's accuracy. We observed that predicted values exhibit increased scatter in abundance ranges below 0 dex and above 0.3 dex, with the largest dispersion observed for targets with chemical abundances lower than −0.4 dex. This pronounced variability at both low and high abundance levels is likely attributed to the limited sample size in these regions. To further ensure robustness, the final model incorporates a regularization term to address potential overfitting. Subsequently, we identified sources from the LAMOST DR11 LRS data set that met the calibration criteria for atmospheric parameters. Using the trained machine learning model, we predicted the abundances of the five α-elements across the data set, providing a comprehensive chemical composition analysis for over 6.8 million stellar targets.

4.3. The Chemical Abundance for OCs

The Monte Carlo (MC) sampling method is employed within a Bayesian framework to enhance the precision of OCs metallicity estimates. By randomly sampling initial abundance estimates along with their uncertainties, considering both observational and model, the method iteratively compares chemical models with observed data. This process generates posterior distributions of chemical abundances, yielding robust estimates and uncertainties. By addressing data errors comprehensively, the MC approach significantly improves the reliability of chemical abundance analyses, facilitating a more accurate understanding of the chemical evolution of OCs.

Following the data quality control procedures outlined in the previous chapters, the initial OC sample is systematically refined and constrained to 760 clusters. This study adopts a rigorous approach to the compositional analysis of OCs, emphasizing membership selection, uncertainty quantification, and statistical sampling. Memberships are determined using the catalog of Hunt & Reffert (2024), providing a reliable foundation for the analysis. Elemental uncertainties are quantified by the dispersion of predicted values, with standard deviations of $\sigma_{\rm O} = 0.091$ dex, $\sigma_{\rm Mg} = 0.051$ dex, $\sigma_{\rm Si} = 0.037$ dex, $\sigma_{\rm Ca} = 0.044$ dex, and $\sigma_{\rm Ti} = 0.044$ dex. Assuming Gaussian distributions for these uncertainties, as outlined by Zhang et al. (2024) and Fu et al. (2022), simplifies the evaluation process.

To refine abundance estimates, 5000 random samples are generated for each member star, incorporating both observational and modeled uncertainties. The mean and median of the sampled distributions are used to define individual abundances, with stars exhibiting deviations beyond 2σ excluded. The retained stars' abundances are then employed to determine the overall composition of each cluster. The results, including detailed abundance data for the selected OCs, are presented in Table 1. The median standard deviation for each element across all OCs is summarized as follows: $\sigma_{\rm O,median} = 0.065$ dex, $\sigma_{\rm Mg,median} = 0.036$ dex, $\sigma_{\rm Si,median} = 0.026$ dex, $\sigma_{\rm Ca,median} = 0.031$ dex,and $\sigma_{\rm Ti,median} = 0.031$ dex.

5. Discussion

Elements in stellar atmospheres originate from various nucleosynthesis sites, including Type II and Ia supernovae, as well as asymptotic giant branch (AGB) stars, each contributing at different release rates to the ISM (Kobayashi et al. 2020). Analyzing the variation of elemental abundances across the Galactic disk provides valuable insights into the production mechanisms of these elements and the processes that drive the Milky Way's evolution (Spina et al. 2021; Zhang et al. 2022; Carbajo-Hijarrubia et al. 2024). The chemical composition of OCs, which spans a range of elements, ages, and Galactocentric radii, plays a pivotal role in the method of "chemical tagging." This approach posits that a star's chemical signature reflects the environment in which it formed, with stars sharing similar formation conditions clustering in chemical space (Freeman & Bland-Hawthorn 2002).

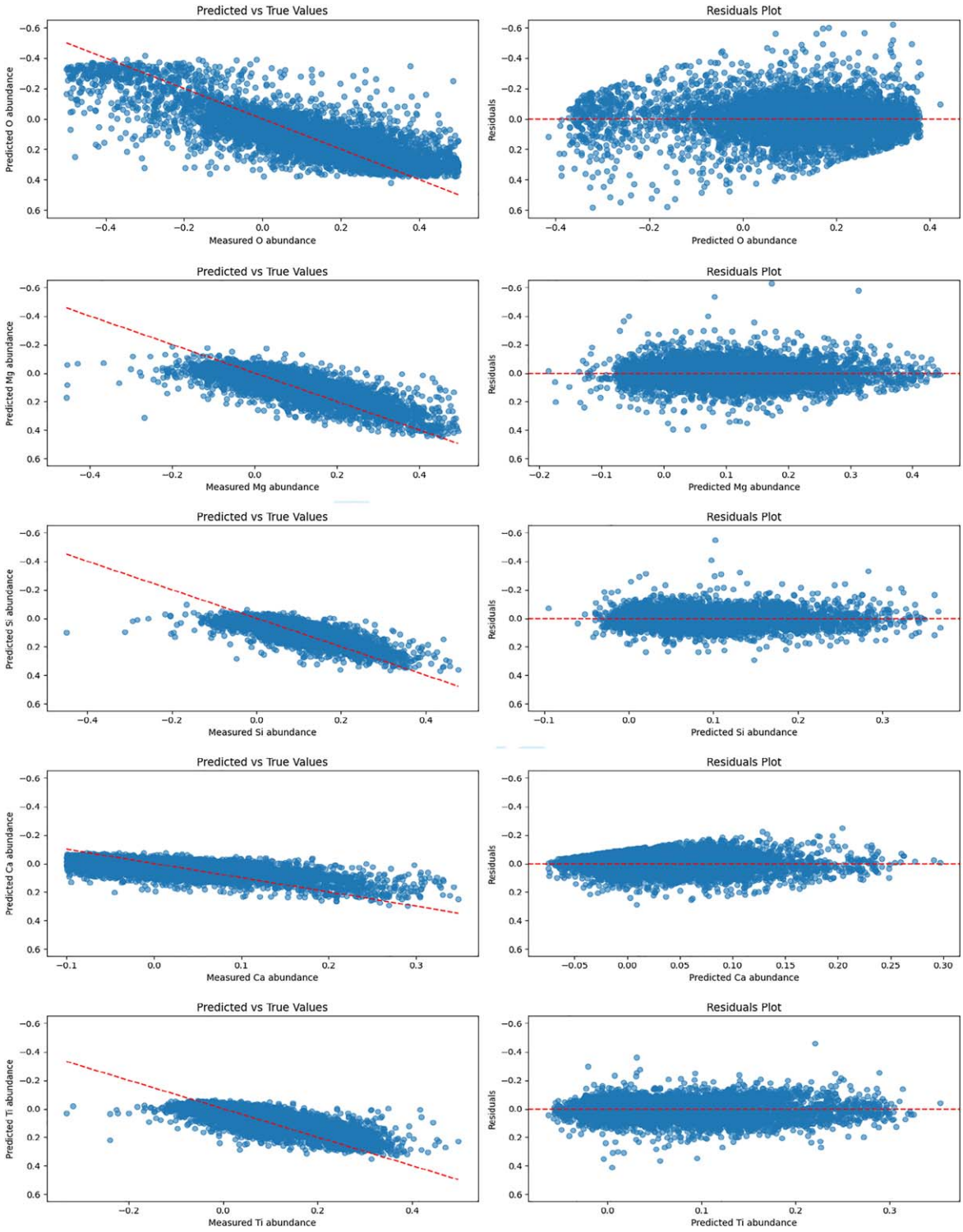

Figure 1. Left: The relationship between the measured abundances and predicted abundances of α-elements, illustrating the accuracy of the model's predictions. Right: The residuals, showing the differences between the true and predicted values.

Table 1
Chemical Abundances of OCs in the Current Sample (Extract)

Name	R.A. (H:M:S)	Decl. (D:M:S)	Rgc (pc)	[O/Fe] _mean	[O/Fe] _std	[Mg/Fe] _mean	[Mg/Fe] _std	[Si/Fe] _mean	[Si/Fe] _std	[Ca/Fe] _mean	[Ca/Fe] _std	[Ti/Fe] _mean	[Ti/Fe] _std
Alessi_12	20:43:28.99	+23:32:18.48	7933.253	0.090	0.066	0.004	0.039	0.013	0.027	−0.010	0.034	0.020	0.033
Alessi_144	07:30:22.02	+05:43:00.42	8820.936	0.009	0.111	0.004	0.048	0.012	0.028	−0.019	0.033	0.014	0.033
Alessi_145	08:03:48.86	+03:59:06.32	8603.044	0.043	0.087	0.004	0.040	0.007	0.027	−0.012	0.032	0.007	0.032
Alessi_170	06:41:49.55	−05:14:21.42	9243.102	0.066	0.066	0.002	0.036	0.010	0.026	−0.015	0.031	0.009	0.031
Alessi_19	18:18:08.20	+11:53:50.45	7701.673	0.054	0.064	−0.012	0.036	0.000	0.026	−0.023	0.031	0.001	0.031
Alessi_2	04:46:13.69	+55:11:02.65	8658.511	0.027	0.095	−0.005	0.044	0.005	0.028	−0.020	0.034	0.004	0.032
Alessi_20	00:10:34.44	+58:45:23.44	8326.947	0.033	0.078	0.043	0.046	0.046	0.042	−0.001	0.040	0.037	0.054
Alessi_22	23:48:41.35	+36:13:45.29	8233.496	−0.013	0.100	0.023	0.048	0.021	0.039	−0.010	0.036	0.014	0.044
Alessi_37	22:47:53.01	+46:17:24.62	8293.571	0.020	0.087	0.016	0.050	0.018	0.039	−0.011	0.037	0.013	0.043
.....													

Note. The full table is available at the ScienceDB (available at 10.57760/sciencedb.j00167.00012).

The effectiveness of chemical tagging relies on the homogeneity of chemical compositions within giant molecular clouds, as well as the spatial and temporal variability of elements in the ISM. Massive stars that end their lives in core-collapse supernova (SNe), as one very prominent source of α-elements. Being short-lived, core-collapse supernovae eject fresh material into the ISM, thereby influencing almost instantaneously, but significantly, the early chemical evolution of the Milky Way. Such immediate perturbation plays a stronger role in establishing earlier conditions under which galactic evolution goes into action. In this section, we shall examine the spatial distribution of these species across the Galactic disk.

Figure 2 illustrates the correlation between chemical abundances and Galactocentric distance for a sample of 760 OCs. Using the PYMC3 package (Salvatier et al. 2016), we performed Bayesian linear regression to investigate the relationship between α-elements and Galactocentric distance. This approach follows the methodology outlined by Spina et al. (2021), leveraging prior distributions informed by key findings from earlier studies (Spina et al. 2021; Myers et al. 2022; Zhang et al. 2022). The specified priors established a robust statistical foundation for parameter estimation, ensuring the reliability of the analysis. The resulting gradients and intercepts for each α-element, as presented in Table 2, provide critical insights into the spatial distribution and variation of α-element abundances across the Galactic disk. Importantly, our analysis was conducted without stratifying the data by age or Galactocentric distance, allowing for an unbiased assessment of the overarching trends and patterns in the chemical abundance gradients.

The chemical abundances exhibit a general increasing trend with R_{GC}, with oxygen displaying the steepest gradient as the distance from the Galactic center increases, However, the present calculation remains consistent with the findings of Donor et al. (2020), Zhang et al. (2022). The distribution patterns of Mg and Si are closely aligned with those reported in previous studies, such as Zhang et al. (2022), reinforcing consistency with well-established trends in the literature. Notably, Mg exhibits a slightly flatter gradient compared to the findings of Carbajo-Hijarrubia et al. (2024), which may suggest subtle variations in the underlying processes influencing its distribution.

Similarly, Ca shows a marginally steeper gradient relative to the results reported by Zhang et al. (2022), yet it aligns well with the theoretical calculations and observational results from Casamiquela et al. (2019), Donor et al. (2020), Myers et al. (2022). This consistency underscores the reliability of current models in capturing the chemical evolution of Ca across the Galactic disk.

In contrast, Ti demonstrates a relatively flatter gradient compared to high-resolution spectroscopic studies, potentially reflecting variations in observational sensitivity or sample selection effects. However, it remains in good agreement with the results presented by Myers et al. (2022), supporting the robustness of these calculations in describing Ti's distribution trends. These observations collectively contribute to a deeper understanding of the elemental abundance gradients and their implications for Galactic chemical evolution.

Our results confirm the overall increase in α-element abundances with Galactocentric distance, in agreement with previous studies. The gradients of Mg and Si are consistent with prior findings, though Mg appears slightly flatter than reported by Carbajo-Hijarrubia et al. (2024), while Ca exhibits a marginally steeper gradient than Zhang et al. (2022), both remaining within theoretical expectations. Ti presents a relatively flatter gradient compared to other α-elements; however, its slope remains consistent with Myers et al. (2022), suggesting that observational sensitivity may play a significant role in this discrepancy. These discrepancies may

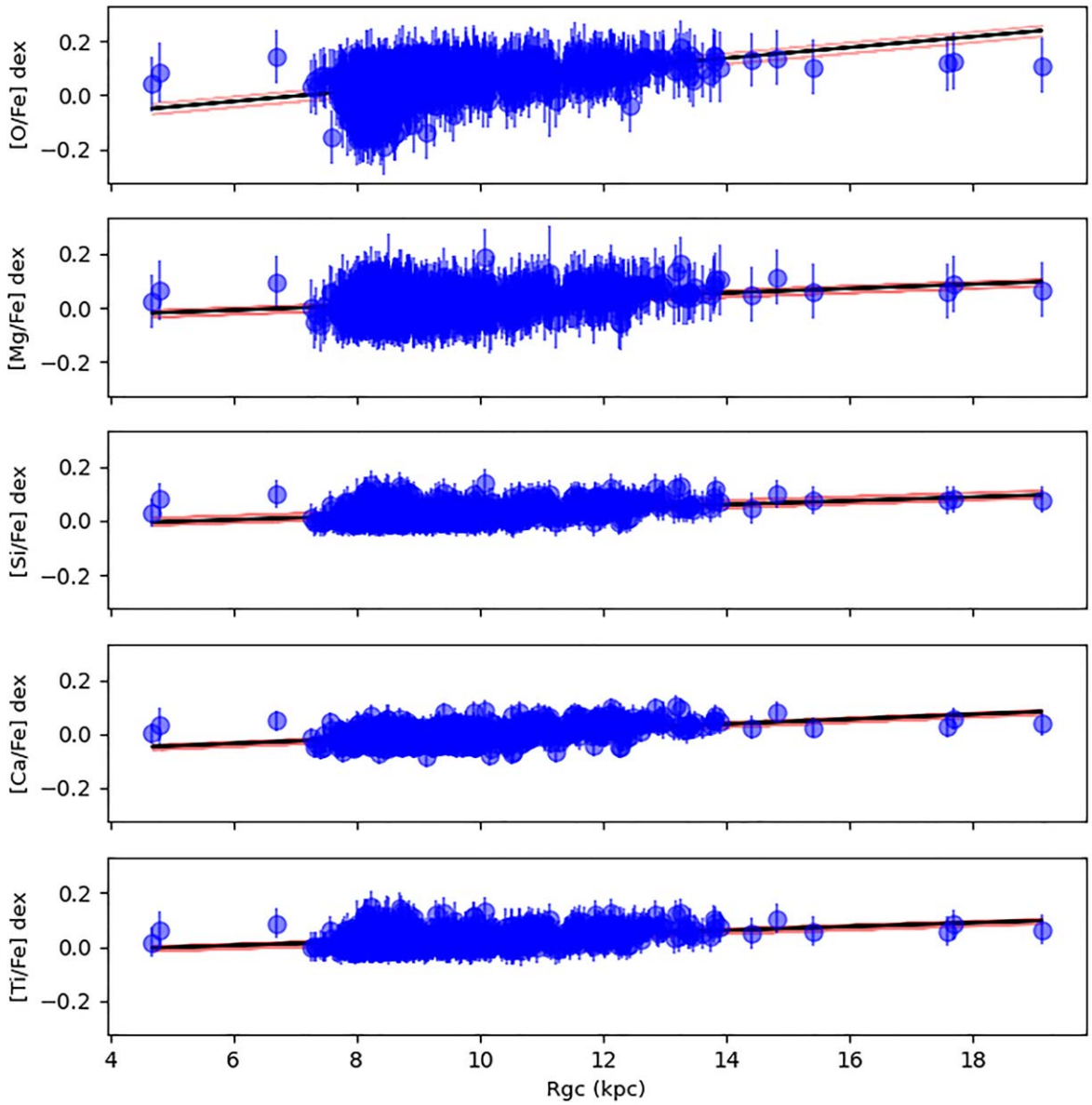

Figure 2. The distribution of abundance ratios for O, Mg, Si, Ca, and Ti as a function of Galactocentric distance is presented for all open clusters (OCs). Data points are accurately plotted, and a linear regression line with corresponding 95% confidence intervals is superimposed to illustrate the overall trend.

stem from differences in sample selection, statistical methodologies, and underlying astrophysical processes, as further discussed in the final conclusion.

6. Conclusions

LAMOST recently released its 11th data set, providing stellar parameters for low-resolution spectra analyzed using the LASP pipeline. While the catalog offers essential atmospheric parameters, such as effective temperature, surface gravity, and metallicity, it lacks detailed α-element abundances, which are crucial for exploring stellar populations and Galactic chemical evolution. To address this gap, this study employs low-resolution spectra from LAMOST DR11 to re-evaluate the α-element abundances-oxygen (O), magnesium (Mg), silicon (Si), calcium (Ca), and titanium (Ti)-for over 6.8 million stars. The chemical compositions derived from the LAMOST DR11 LRS are predicted using the XGBoost model, carefully calibrated against high-precision data from GALAH DR4, ensuring robust and accurate abundance determinations.

Total of 760 OCs were meticulously cross-matched and selected from the catalog provided by Hunt & Reffert (2024), ensuring a comprehensive representation of stellar ages. These

Table 2
Linear Regression Coefficients and the 95% Confidence Intervals for 5 α-elements

	Parameter	Mean	Uncertainty	95% C.I
[O/Fe]–R_{GC}	α	0.021	0.001	[0.018, 0.024]
	β	−0.151	0.013	[−0.177, −0.126]
[Mg/Fe]–R_{GC}	α	0.007	0.001	[0.005, 0.009]
	β	−0.052	0.008	[−0.068, −0.040]
[Si/Fe]–R_{GC}	α	0.006	0.000	[0.005, 0.008]
	β	−0.038	0.005	[−0.050, −0.026]
[Ca/Fe]–R_{GC}	α	0.008	0.000	[0.007, 0.010]
	β	−0.084	0.005	[−0.096, −0.072]
[Ti/Fe]–R_{GC}	α	0.007	0.000	[0.006, 0.008]
	β	−0.037	0.005	[−0.047, −0.028]

clusters span a wide age range, from 400 Myr, representing the younger end, to 6.73 Gyr, capturing substantial chronological diversity. To ensure the reliability and precision of chemical abundance measurements for each OC, MC sampling was employed as a rigorous quality control measure.

The primary objectives of this study are as follows: More than 30,000 stars were carefully selected from the GALAH DR4 catalog, with attention to key atmospheric parameters such as $T_{\rm eff}$, $\log g$, and [Fe/H]. Using this data set, an XGBoost-based model was developed, enabling the extraction of abundances for five α-elements: O, Mg, Si, Ca, and Ti, from the LAMOST DR11 LRS via a decision tree approach. To further enhance the accuracy of the abundance values, MC sampling was applied for the final validation of the chemical abundance measurements for each OC.

Subsequently, the compositional distribution across the Galactic disk was analyzed. Bayesian linear regression was employed to infer chemical abundance gradients as a function of R_{GC}. The results indicate that chemical abundances generally increase with Galactocentric radius, with oxygen exhibiting the steepest gradient. In contrast, the distributions of elements such as Mg, Si, Ca, and Ti reveal both similarities and differences when compared to previous studies, contributing to a deeper understanding of elemental abundance gradients and Galactic chemical evolution.

These discrepancies primarily arise from differences in data set characteristics, analytical methodologies, and underlying astrophysical processes. High-resolution spectroscopic studies, such as those by Casamiquela et al. (2019), Carbajo-Hijarrubia et al. (2024), Zhang et al. (2022), are constrained by smaller and predominantly younger samples, which may bias the inferred abundance gradients. In contrast, our study incorporates large-scale survey data, supplemented by machine learning techniques to predict missing values, thereby improving sample coverage. However, this approach may introduce systematic biases due to model assumptions and training data limitations. Additionally, methodological differences-including statistical frameworks, regression models, and the treatment of observational uncertainties-contribute to the observed variations. Despite these challenges, our findings provide valuable constraints on Galactic chemical evolution and underscore the importance of methodological consistency in abundance gradient studies. Future research incorporating improved modeling techniques and additional high-resolution spectroscopic data will be essential for further refining our understanding of chemical enrichment processes in the Milky Way.

Acknowledgments

This work is supported by the National Natural Science Foundation of China under program Nos. 12090040, 12090043, 12473031, and 12003025, as well as the Basic Research Program of Yunnan Province (No. 202401AT070142). We also acknowledge the generous support from the International Center of Supernovae, Yunnan Key Laboratory (No. 202302AN360001), and the Natural Science Foundation of Yunnan Province (No. 202201BC070003). J.X.C. acknowledges the support of the Postdoctoral Fellowship Program of CPSF under grant Nos. GZC20240124 and 2024M760242.

We sincerely thank the anonymous referees for their valuable and insightful comments, which have greatly improved this work. This research has utilized the Simbad, Vizier, and NASA-ADS databases, as well as the software TOPCAT (Taylor 2005).

The Guo Shou Jing Telescope, commonly referred to as LAMOST (Large Sky Area Multi-Object Fiber Spectroscopic Telescope), represents a significant national scientific endeavor developed by the Chinese Academy of Sciences. Its construction was financed by the National Development and Reform Commission. The telescope's operation and administration are overseen by the National Astronomical Observatories under the Chinese Academy of Sciences.

We also acknowledge the use of data from the European Space Agency (ESA) mission Gaia (https://www.cosmos.esa.int/Gaia). These data were processed by the Gaia Data Processing and Analysis Consortium (DPAC, https://www.cosmos.esa.int/web/Gaia/dpac/consortium, 02302AN360001).

ORCID iDs

Shuai Zha https://orcid.org/0000-0001-6773-7830

References

Boeche, C., Smith, M. C., Grebel, E. K., et al. 2018, AJ, 155, 181
Bossini, D., Vallenari, A., Bragaglia, A., et al. 2019, A&A, 623, A108
Cantat-Gaudin, T., & Anders, F. 2020a, A&A, 633, A99
Cantat-Gaudin, T., Anders, F., Castro-Ginard, A., et al. 2020b, A&A, 640, A1
Carbajo-Hijarrubia, J., Casamiquela, L., Carrera, R., et al. 2024, A&A, 687, A239
Casamiquela, L., Blanco-Cuaresma, S., Carrera, R., et al. 2019, MNRAS, 490, 1821

Chi, H., Wang, F., Wang, W., et al. 2023, ApJS, 266, 36
Coronado, J., Rix, H.-W., Trick, W. H., et al. 2020, MNRAS, 495, 4098
De Silva, G. M., Freeman, K. C., Bland-Hawthorn, J., et al. 2015, MNRAS, 449, 2604
Donor, J., Frinchaboy, P. M., Cunha, K., et al. 2020, AJ, 159, 199
Du, B., Luo, A., Zhang, J., et al. 2012, Proc. SPIE, 8451, 845137
Freeman, K., & Bland-Hawthorn, J. 2002, ARA&A, 40, 487
Friel, E. D., Jacobson, H. R., & Pilachowski, C. A. 2010, AJ, 139, 1942
Fu, X., Bragaglia, A., Liu, C., et al. 2022, A&A, 668, A4
Fujii, M. S., & Portegies Zwart, S. 2016, ApJ, 817, 4
Gaia Collaboration, Brown, A. G. A., Vallenari, A., et al. 2018, A&A, 616, A1
Huang, Y., Liu, X.-W., Zhang, H.-W., et al. 2015, RAA, 15, 1240
Hunt, E. L., & Reffert, S. 2023, A&A, 673, A114
Hunt, E. L., & Reffert, S. 2024, A&A, 686, A42
Jofré, P., Heiter, U., & Soubiran, C. 2019, ARA&A, 57, 571
Kobayashi, C., Karakas, A. I., & Lugaro, M. 2020, ApJ, 900, 179
Li, X., & Lin, B. 2023, MNRAS, 521, 6354
Liu, L., & Pang, X. 2019, ApJS, 245, 32
Luo, A.-L., Zhao, Y.-H., Zhao, G., et al. 2015, RAA, 15, 1095
Majewski, S. R., Schiavon, R. P., Frinchaboy, P. M., et al. 2017, AJ, 154, 94
Myers, N., Donor, J., Spoo, T., et al. 2022, AJ, 164, 85
Nissen, P. E., & Gustafsson, B. 2018, A&ARv, 26, 6
Piskunov, A. E., Kharchenko, N. V., Röser, S., et al. 2006, A&A, 445, 545
Salvatier, J., Wieckiâ, T. V., & Fonnesbeck, C., 2016 Astrophysics Source Code Library, ascl:1610.016
Sharma, S., Hayden, M. R., Bland-Hawthorn, J., et al. 2021, MNRAS, 506, 1761
Spina, L., Ting, Y.-S., De Silva, G. M., et al. 2021, MNRAS, 503, 3279
Taylor, M. B. 2005, Astronomical Data Analysis Software and Systems XIV, ASP Conf. Series, 347 (San Francisco, CA: ASP), 29
Tian, H.-J., Liu, C., Wu, Y., et al. 2018, ApJL, 865, L19
Vickers, J. J., Shen, J., & Li, Z.-Y. 2021, ApJ, 922, 189
Wang, C., Huang, Y., Yuan, H.-B., et al. 2019, ApJL, 877, L7
Wu, Y., Luo, A.-L., Li, H.-N., et al. 2011, RAA, 11, 924
Wu, Y., Xiang, M., Chen, Y., et al. 2021, MNRAS, 501, 4917
Xiang, M.-S., Liu, X.-W., Yuan, H.-B., et al. 2015, RAA, 15, 1209
Yanny, B., Rockosi, C., Newberg, H. J., et al. 2009, AJ, 137, 4377
Yaz Gökçe, E., Bilir, C., Karaali, S., et al. 2017, Ap&SS, 362, 185
Yu, Y., Wang, H.-F., Cui, W.-Y., et al. 2021, ApJ, 922, 80
Zhao, G., Zhao, Y.-H., Chu, Y.-Q., et al. 2012, RAA, 12, 723
Zhang, R., Lucatello, S., Bragaglia, A., et al. 2022, A&A, 667, A103
Zhang, R., Wang, G.-J., Lu, L., et al. 2024, A&A, 692, A212