Photometric metallicities using S-PLUS data

Eduardo Machado Pereira

Métodos para análise de grande volume de dados e Astroinformática

Summary

Metallicity

Southern Photometric Local Universe Survey (S-PLUS)

Data

LightGBM

Results

Metallicity

$$[\mathrm{Fe/H}] \equiv \log_{10}\left[\frac{(N_{\mathrm{Fe}}/N_{\mathrm{H}})_{\mathrm{star}}}{(N_{\mathrm{Fe}}/N_{\mathrm{H}})_{\odot}}\right]$$

N_{Fe}: iron atoms

N_{H}: hydrogen atoms

Metallicity

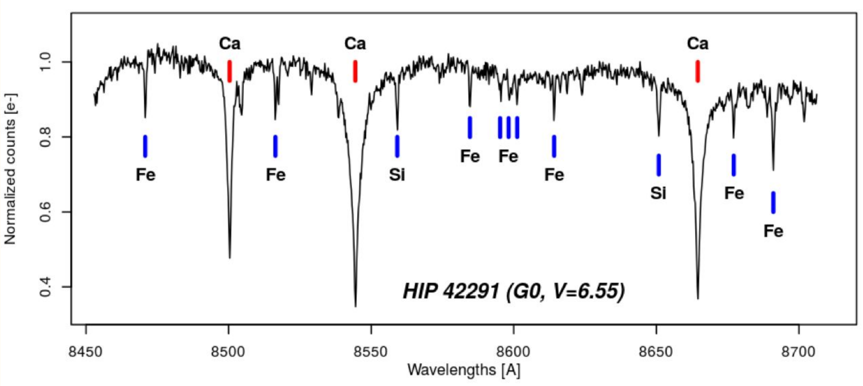
$$[\mathrm{Fe/H}] \equiv \log_{10}\left[\frac{(N_{\mathrm{Fe}}/N_{\mathrm{H}})_{\mathrm{star}}}{(N_{\mathrm{Fe}}/N_{\mathrm{H}})_{\odot}}\right]$$

N_{Fe}: iron atoms

N_{H}: hydrogen atoms

Sun

$$[\mathrm{Fe/H}] = \log_{10}\left(\frac{N_{\mathrm{Fe}}}{N_{\mathrm{H}}}\right)_{\mathrm{star}} - \log_{10}\left(\frac{N_{\mathrm{Fe}}}{N_{\mathrm{H}}}\right)_{\mathrm{sun}}$$



Sartoretti et al. (2018)

Metallicity

$$[\mathrm{Fe/H}] \equiv \log_{10}\left[\frac{(N_{\mathrm{Fe}}/N_{\mathrm{H}})_{\mathrm{star}}}{(N_{\mathrm{Fe}}/N_{\mathrm{H}})_{\odot}}\right]$$

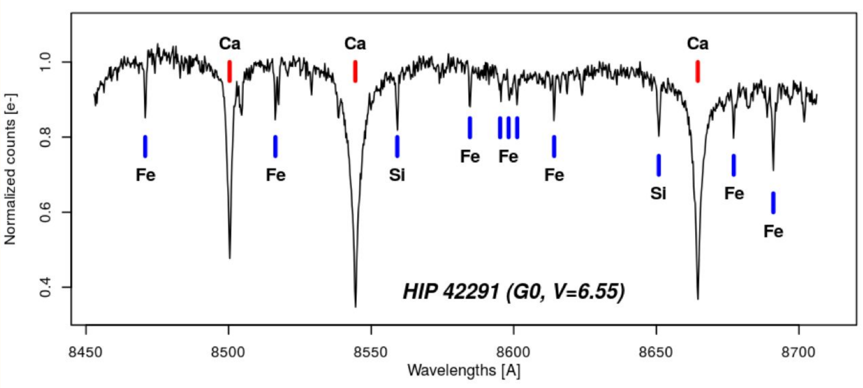
N_{Fe}: iron atoms

N_{H}: hydrogen atoms

Sun

$$[\mathrm{Fe/H}] = \log_{10}\left(\frac{N_{\mathrm{Fe}}}{N_{\mathrm{H}}}\right)_{\mathrm{star}} - \log_{10}\left(\frac{N_{\mathrm{Fe}}}{N_{\mathrm{H}}}\right)_{\mathrm{sun}}$$

- $[\mathrm{Fe/H}] = 0 \Rightarrow$ solar metallicity
- $-6 \lesssim [\mathrm{Fe/H}] \lesssim 1$
- metal poor stars → insights on Milky Way's past



Sartoretti et al. (2018)

Helix Nebula (NGC 7293)

HST (2004)

Metallicity

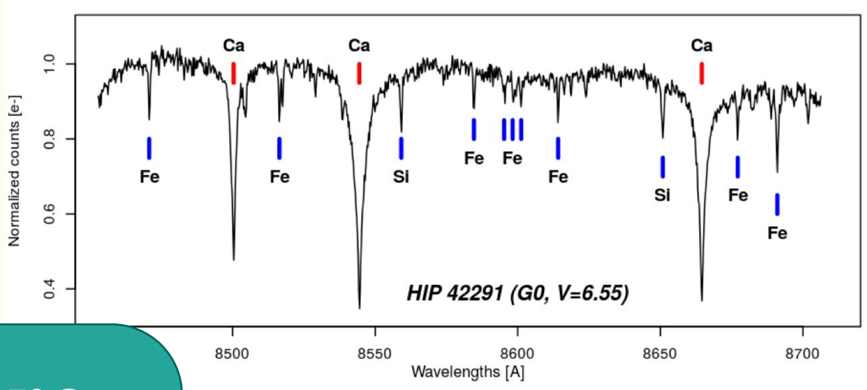
$$[\mathrm{Fe}/\mathrm{H}] \equiv \log_{10}\left[\frac{(N_{\mathrm{Fe}}/N_{\mathrm{H}})_{\mathrm{star}}}{(N_{\mathrm{Fe}}/N_{\mathrm{H}})_{\odot}}\right]$$

N_{Fe}: iron atoms

Sun

$$[\mathrm{Fe}/\mathrm{H}] = \log_{10}\left(\frac{N_{\mathrm{Fe}}}{N_{\mathrm{H}}}\right)_{\mathrm{star}} - \log_{10}\left(\frac{N_{\mathrm{Fe}}}{N_{\mathrm{H}}}\right.$$

- $[\mathrm{Fe}/\mathrm{H}] = 0 \Rightarrow$ solar metallicity
- -6 $\lesssim [\mathrm{Fe}/\mathrm{H}] \lesssim 1$
- metal poor stars → insights on Milky Way's past



Sartoretti et al. (2018)

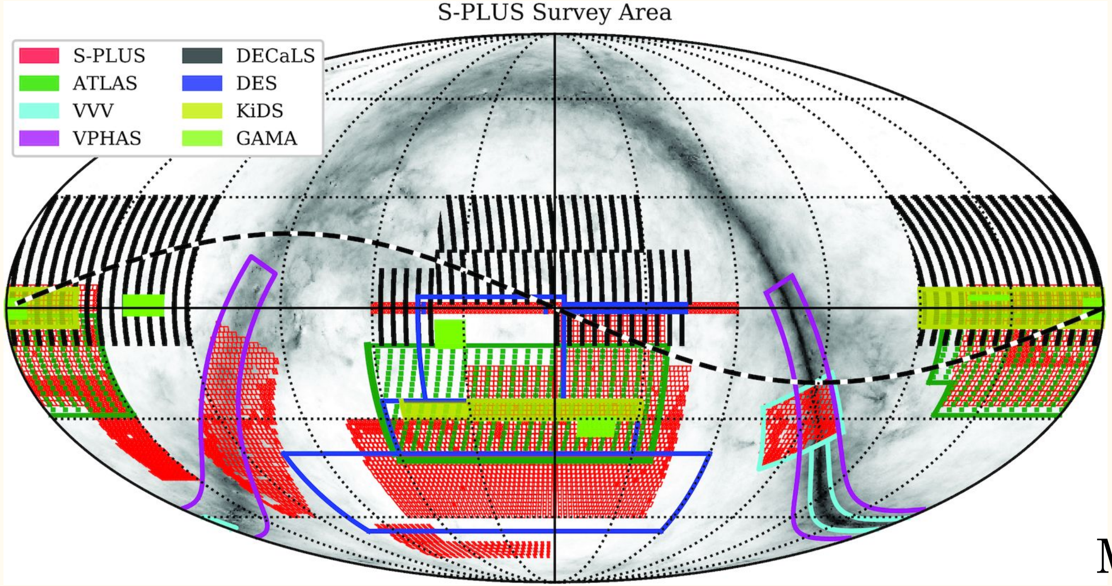
$pectra are expensive!

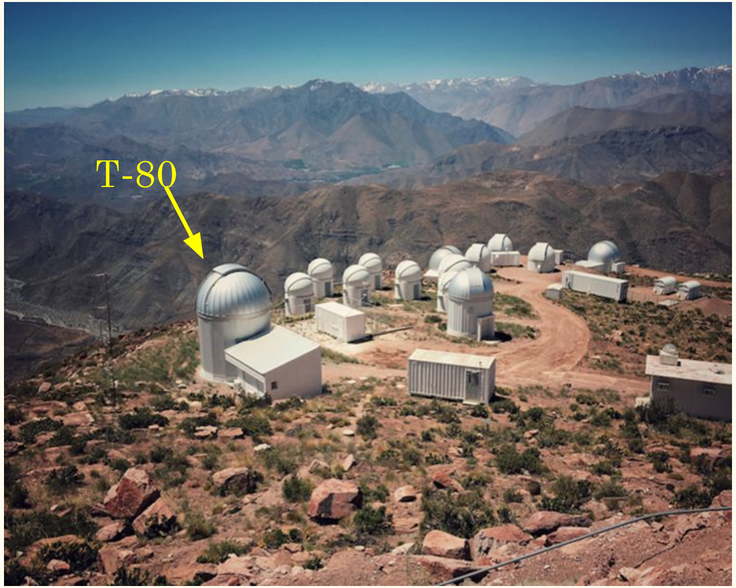
Helix Nebula (NGC 7293)

HST (2004)

Southern Photometric Local Universe Survey

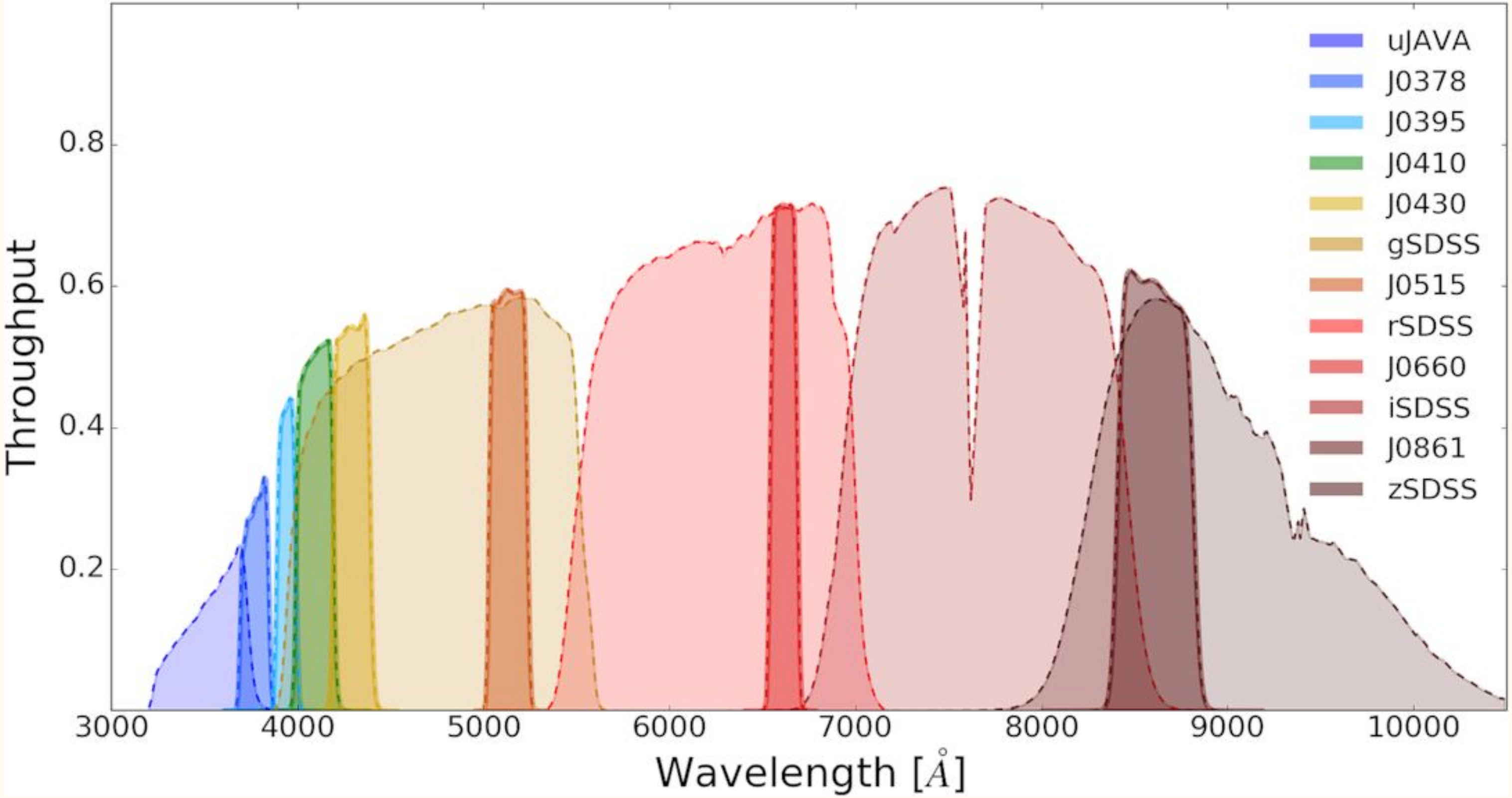
0.8 m telescope (T-80, Chile)
~9300 sq. deg
12 filters





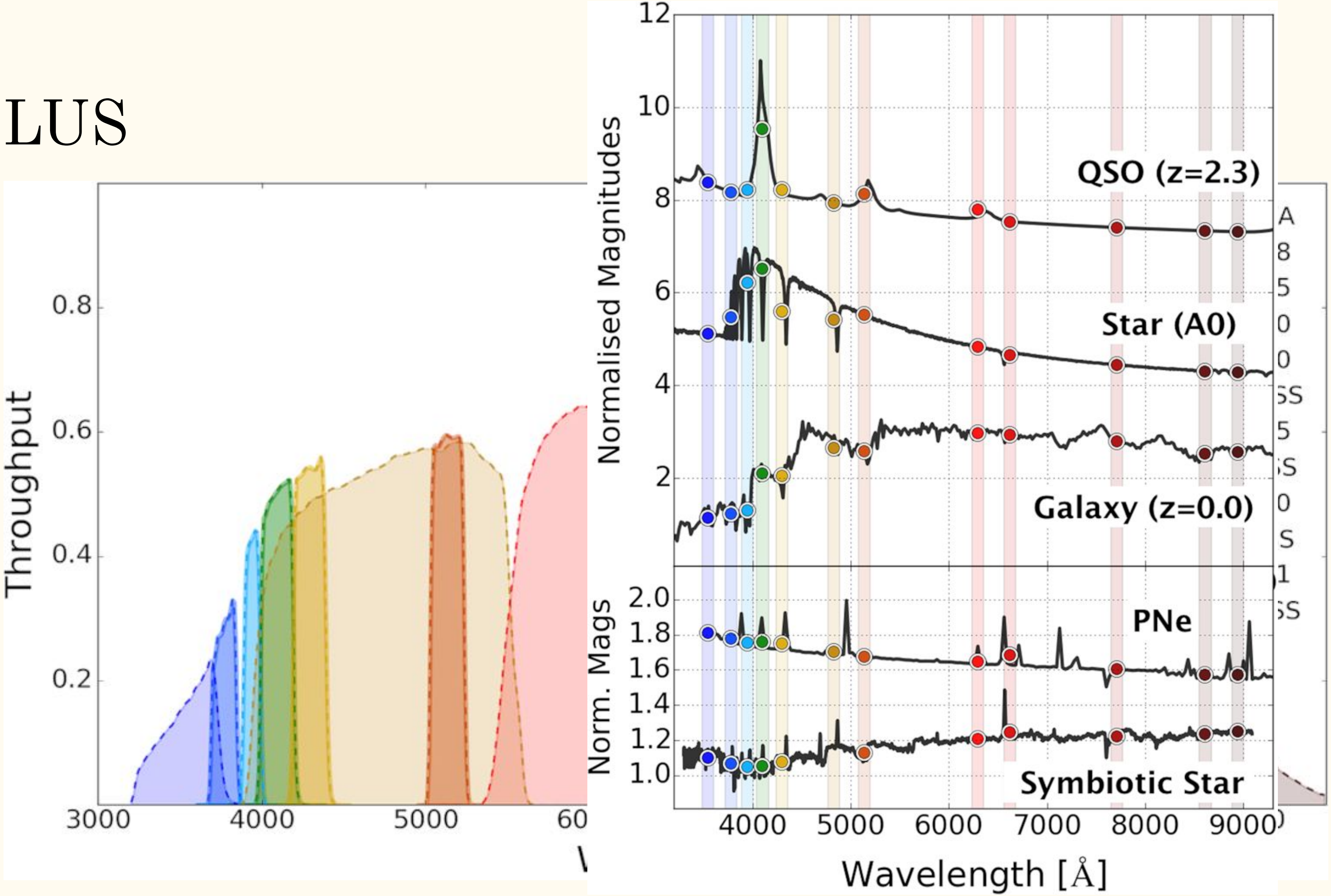
Mendes de Oliveira et al. (2019)

S-PLUS



uJAVA
J0378
J0395
J0410
J0430
gSDSS
J0515
rSDSS
J0660
iSDSS
J0861
zSDSS
Throughput
0.8
0.6
0.4
0.2
3000
4000
5000
6000
7000
8000
9000
10000
Wavelength [Å]

S-PLUS



Throughput
0.2
0.4
0.6
0.8
3000
4000
5000
Normalised Magnitudes
QSO (z=2.3)
Star (A0)
Galaxy (z=0.0)
2
4
6
8
10
12
Norm. Mags
PNe
Symbiotic Star
1.0
1.2
1.4
1.6
1.8
2.0
4000
5000
6000
7000
8000
9000
Wavelength [Å]

Data

S-PLUS main data access: https://splus.cloud/

- Python ADQL query (https://splus.cloud/documentation/python#adql)
  - DR2
    - Basic information with conditions on selecting stars: ~4.5 mi objects
    - (magnitudes $<$ 30 & errors $<$ 1): ~2.2 mi objects
  - Currently DR3 publicly available

Data

Spectroscopy $\Rightarrow$ chemical composition (among others) $\Rightarrow$ [Fe/H]

Large surveys era:

APOGEE

© 2MASS, IPAC/Caltech & U Mass
© 2000, Axel Mellinger

GaiaESO

大天区面积多目标光纤光谱天文望远镜
The Large Sky Area Multi-Object Fiber Spectroscopic Telescope (LAMOST)

GALAH
GALactic Archaeology with HERMES

SEGUE: Mapping the Outer Milky Way
Sloan Extension for Galactic Understanding and Exploration (SEGUE)

Data

Spectroscopy ⇒ chemical composition (among others) ⇒ [Fe/H]

Large surveys era:

APOGEE

© 2MASS, IPAC/Caltech & U Mass
© 2000, Axel Mellinger

GaiaESO

大天区面积多目标光纤光谱天文望远镜

The Large Sky Area Multi-Object Fiber Spectroscopic Telescope (LAMOST)

GALAH

GALactic Archaeology with HERMES

SEGUE: Mapping the Outer Milky Way

Sloan Extension for Galactic Understanding and Exploration (SEGUE)

Data

Spectroscopic constraints:

- **4500 K < $\mathbf{T_{eff}}$ < 7000K** ⇒ F, G, K stars
- **0 < log g < 5** ⇒ most of a star's life cycle
- **-7 < [Fe/H] < 1** ⇒ any metallicity

SEGUE: CasJobs (SQL query)

- ~260k objects

GALAH: GALAH DR3 catalogues or VizieR catalogues (conditions for search)

- ~500k objects

LAMOST: Low Resolution Search - LAMOST DR7 v2.0 (conditions for search)

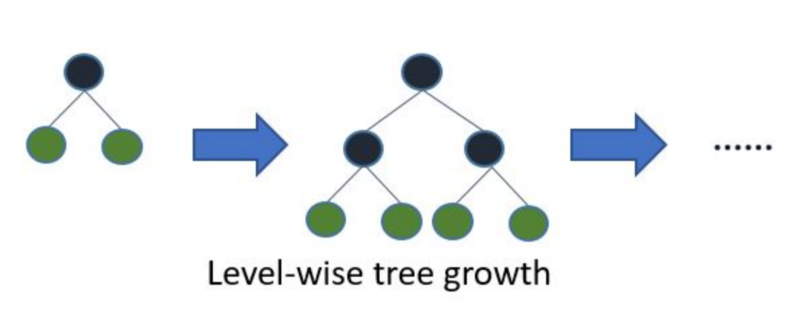
- ~5.5 mi objects

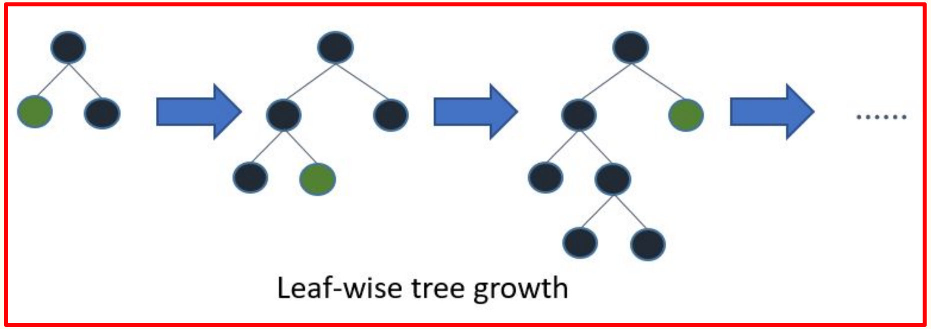
https://lightgbm.readthedocs.io/en/latest/

LightGBM

Tree based learning algorithms with the following advantages:

- Faster training speed and higher efficiency;
- lower memory usage;
- better accuracy;
- support of parallel, distributed, and GPU learning;
- capable of handling large-scale data.





https://lightgbm.readthedocs.io/en/latest/

LightGBM

LightGBM

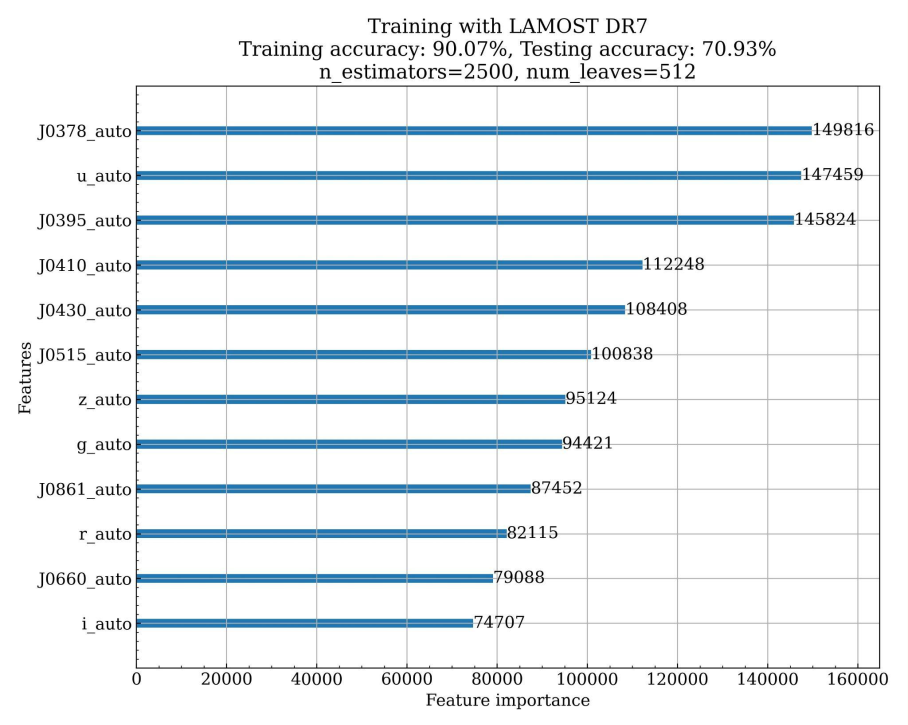
Model training: lightgbm.LGBMRegressor(boosting_type='gbdt', num_leaves=31, max_depth= -1, learning_rate=0.1, n_estimators=100, subsample_for_bin=200000, objective=None, class_weight=None, min_split_gain=0.0, min_child_weight=0.001, min_child_samples=20, subsample=1.0, subsample_freq=0, colsample_bytree=1.0, reg_alpha=0.0, reg_lambda=0.0, random_state=None, n_jobs= -1, importance_type='split', **kwargs)

2 training models:
LAMOST DR7 (more objects) and SEGUE (largest range in [Fe/H])

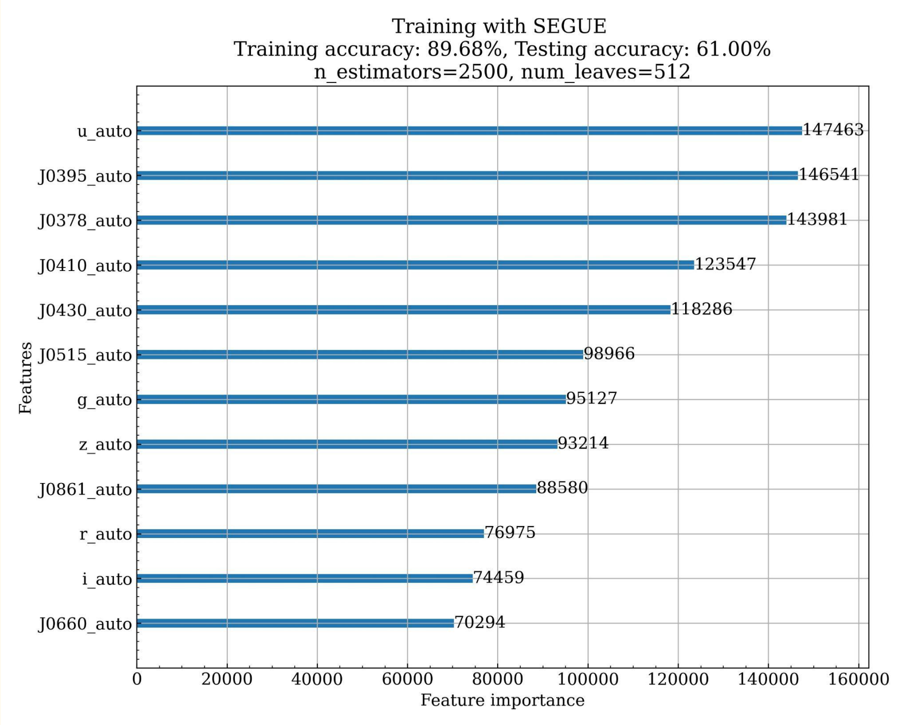
90% training + 5% test + 5% validation

	Objects	num_leaves	n_estimators	random_state	Training accuracy	Testing accuracy
LAMOST DR7	51391	512	2500	42	90.07%	70.93%
SEGUE	16546				89.68%	61.00%

Results

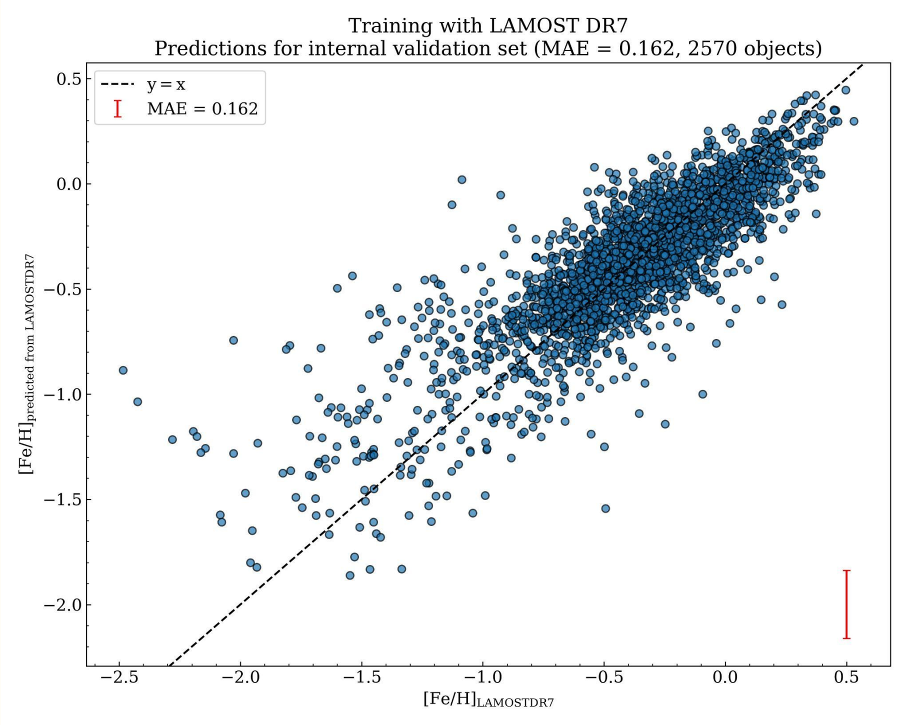


Training with LAMOST DR7
Training accuracy: 90.07%, Testing accuracy: 70.93%
n_estimators=2500, num_leaves=512
J0378_auto 149816
u_auto 147459
J0395_auto 145824
J0410_auto 112248
J0430_auto 108408
J0515_auto 100838
z_auto 95124
g_auto 94421
J0861_auto 87452
r_auto 82115
J0660_auto 79088
i_auto 74707
Features
0 20000 40000 60000 80000 100000 120000 140000 160000
Feature importance

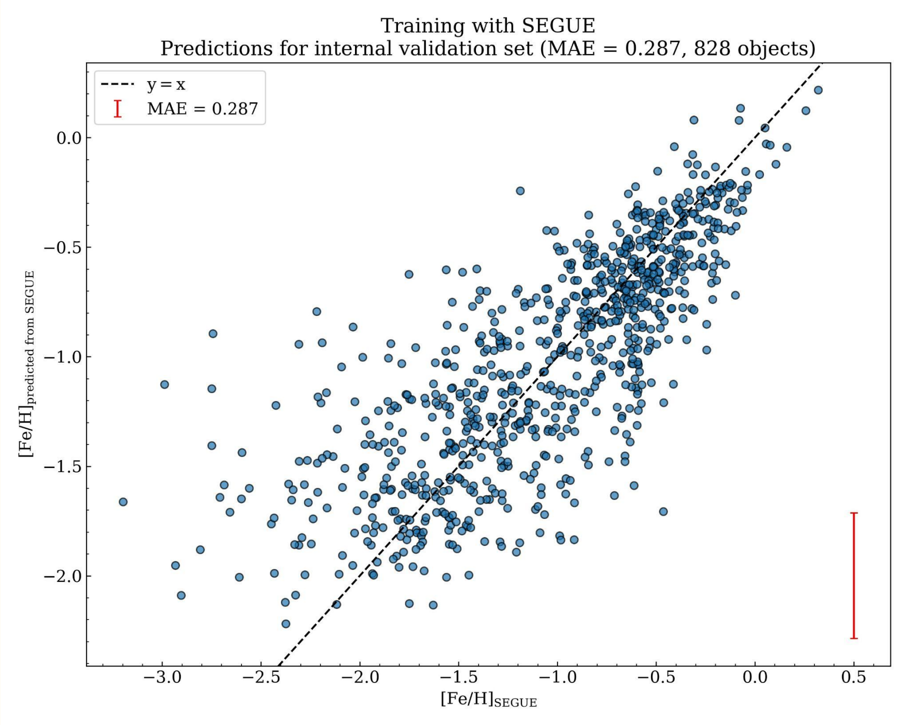


Training with SEGUE
Training accuracy: 89.68%, Testing accuracy: 61.00%
n_estimators=2500, num_leaves=512
u_auto 147463
J0395_auto 146541
J0378_auto 143981
J0410_auto 123547
J0430_auto 118286
J0515_auto 98966
g_auto 95127
z_auto 93214
J0861_auto 88580
r_auto 76975
i_auto 74459
J0660_auto 70294
Features
0 20000 40000 60000 80000 100000 120000 140000 160000
Feature importance

Results

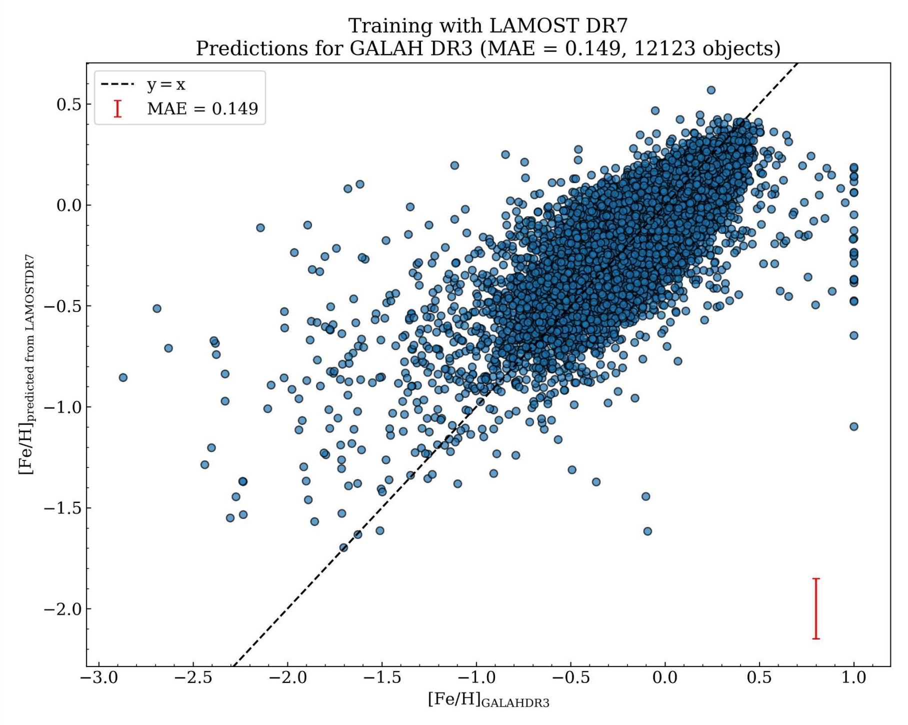


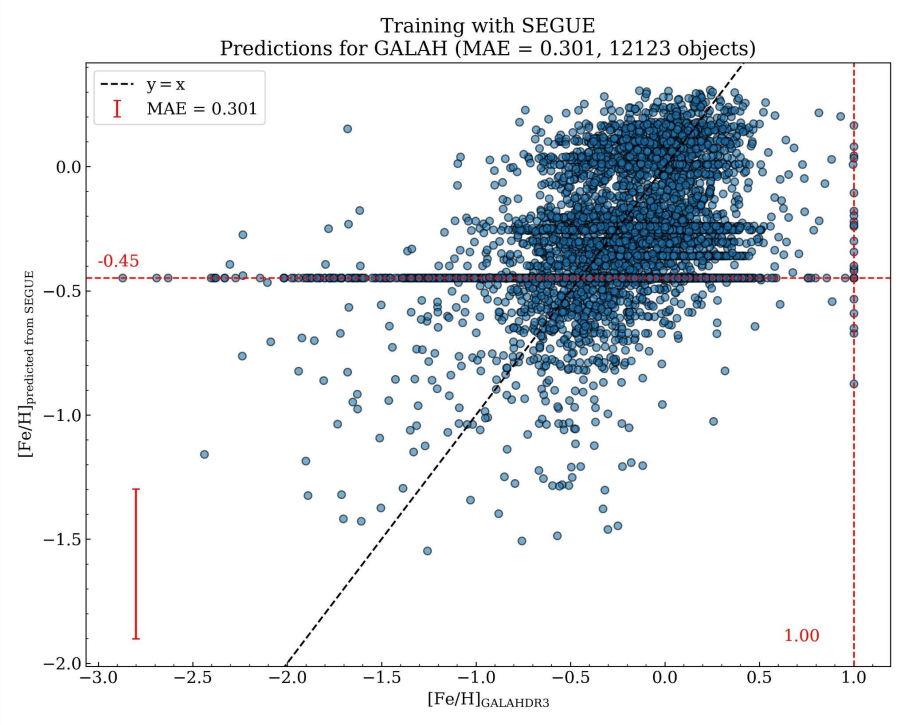
Training with LAMOST DR7
Predictions for internal validation set (MAE = 0.162, 2570 objects)
y = x
MAE = 0.162
$[Fe/H]_{\text{predicted from LAMOSTDR7}}$
$[Fe/H]_{\text{LAMOSTDR7}}$
0.5
0.0
−0.5
−1.0
−1.5
−2.0
−2.5



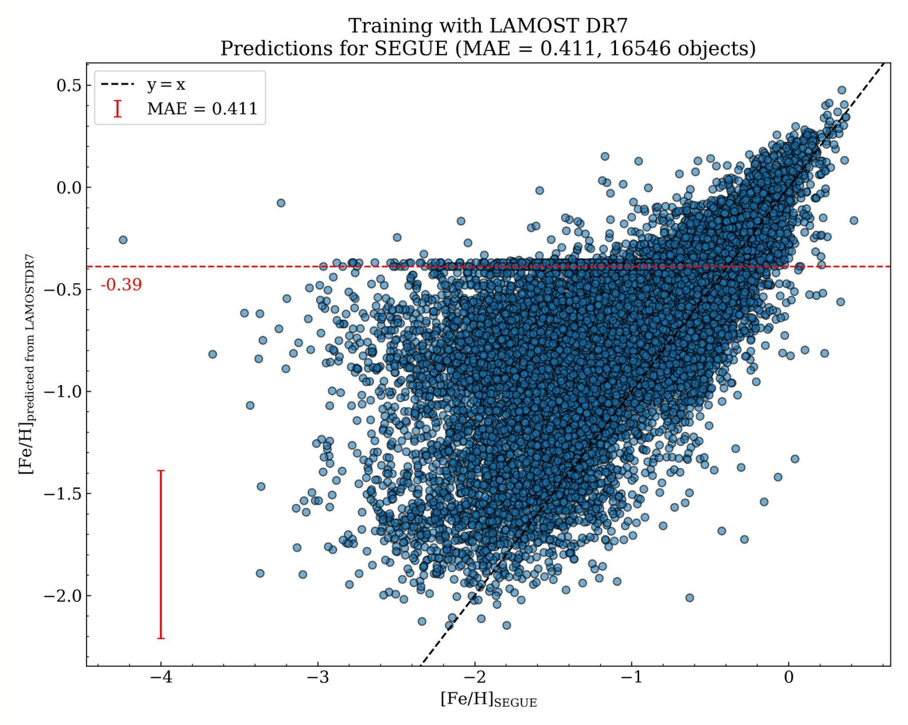
Training with SEGUE
Predictions for internal validation set (MAE = 0.287, 828 objects)
y = x
MAE = 0.287
$[Fe/H]_{\text{predicted from SEGUE}}$
$[Fe/H]_{\text{SEGUE}}$
0.0
−0.5
−1.0
−1.5
−2.0
−2.5
−3.0
0.5

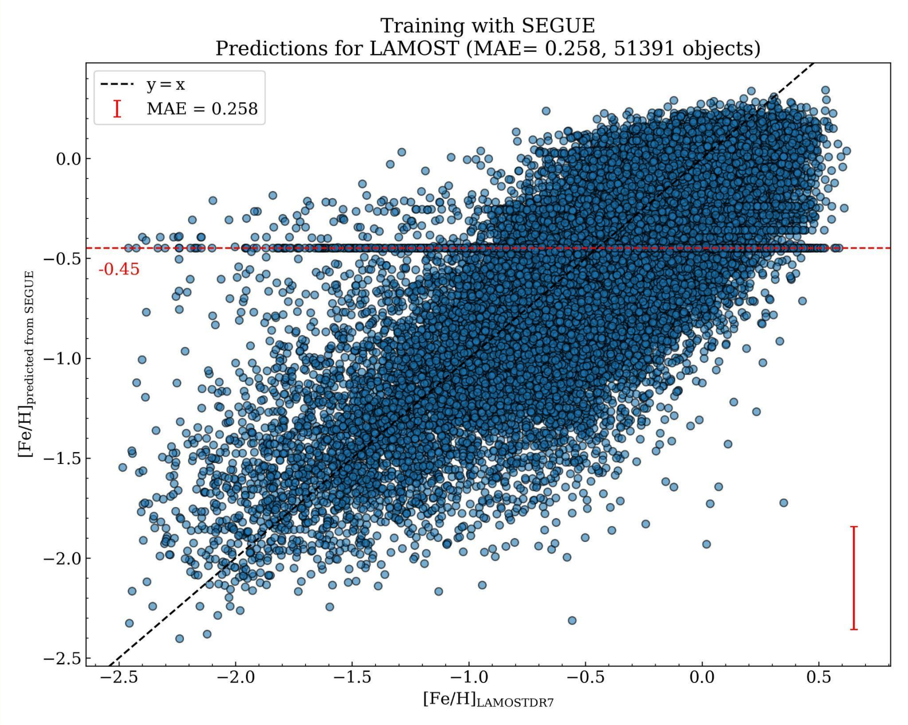
Results


Training with LAMOST DR7
Predictions for GALAH DR3 (MAE = 0.149, 12123 objects)
y = x
MAE = 0.149
$[Fe/H]_{\text{predicted from LAMOSTDR7}}$
$[Fe/H]_{\text{GALAHDR3}}$

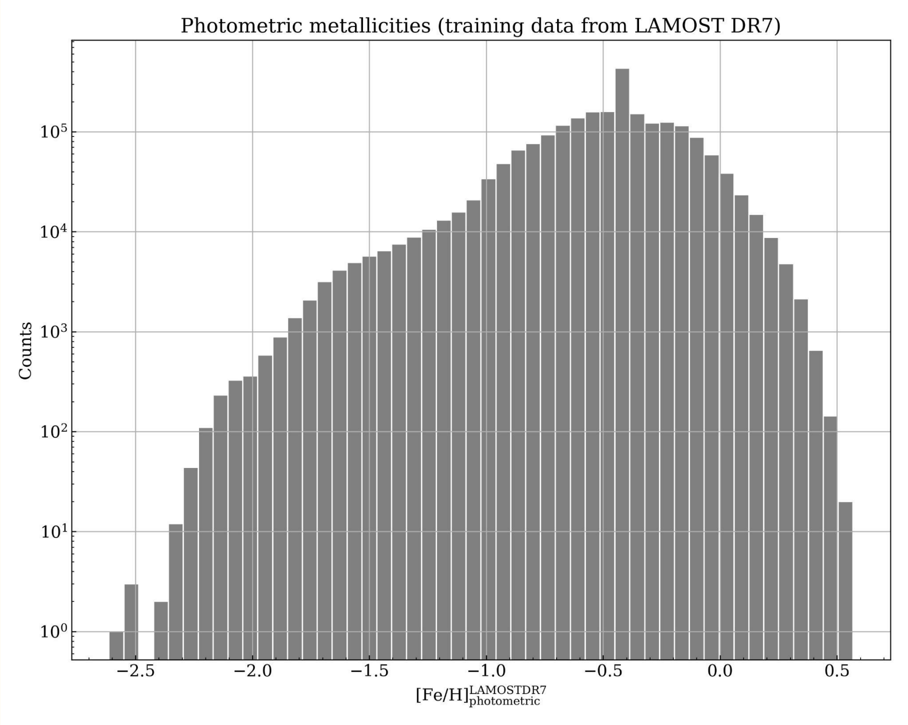

Training with SEGUE
Predictions for GALAH (MAE = 0.301, 12123 objects)
y = x
MAE = 0.301
-0.45
1.00
$[Fe/H]_{\text{predicted from SEGUE}}$
$[Fe/H]_{\text{GALAHDR3}}$

Results

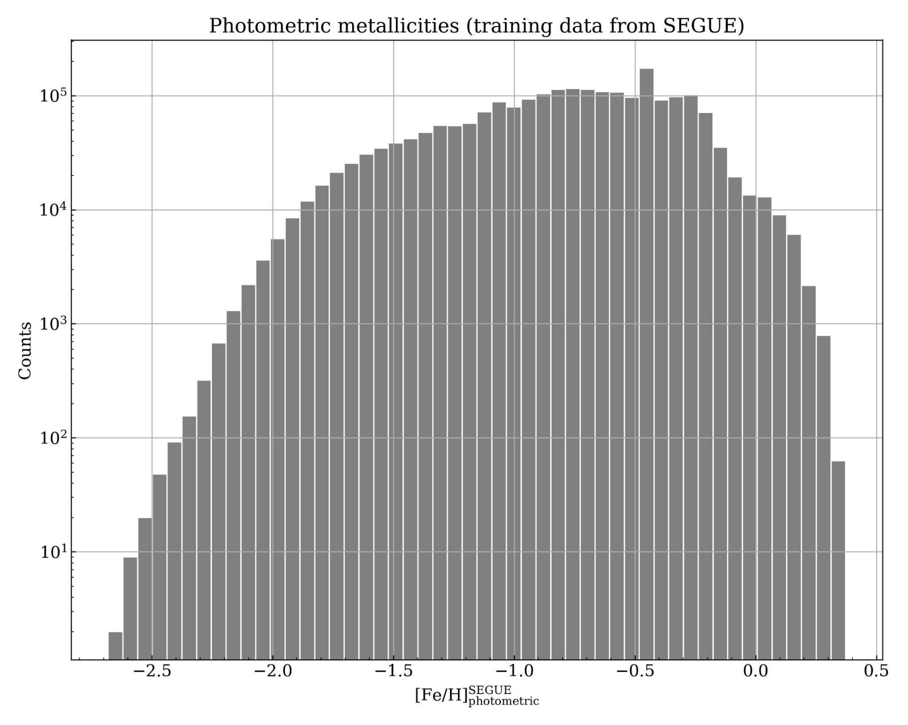

Training with LAMOST DR7
Predictions for SEGUE (MAE = 0.411, 16546 objects)
y = x
MAE = 0.411
-0.39
$[Fe/H]_{predicted\ from\ LAMOSTDR7}$
$[Fe/H]_{SEGUE}$


Training with SEGUE
Predictions for LAMOST (MAE= 0.258, 51391 objects)
y = x
MAE = 0.258
-0.45
$[Fe/H]_{predicted\ from\ SEGUE}$
$[Fe/H]_{LAMOSTDR7}$

Results

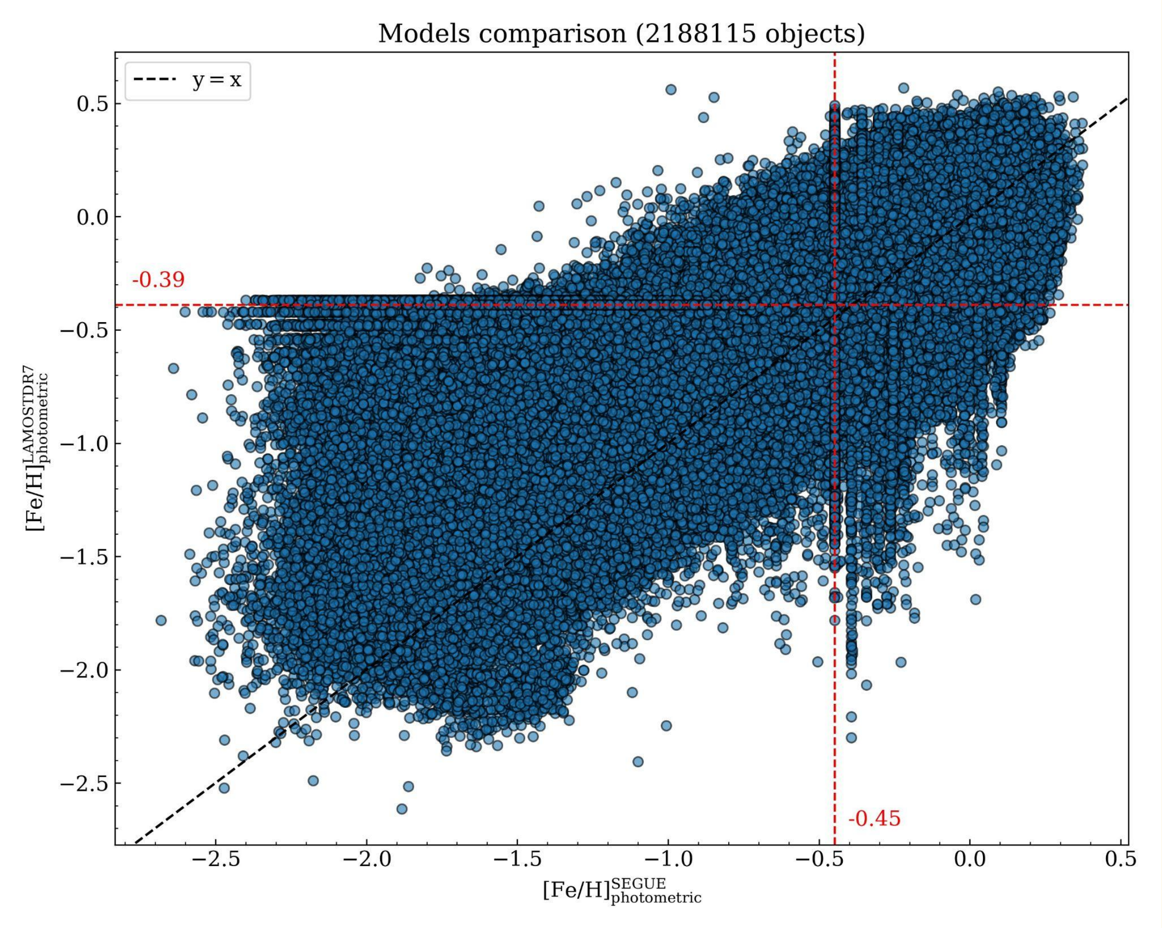


Photometric metallicities (training data from LAMOST DR7)
Counts
10^0
10^1
10^2
10^3
10^4
10^5
−2.5
−2.0
−1.5
−1.0
−0.5
0.0
0.5
$[Fe/H]^{LAMOSTDR7}_{photometric}$



Photometric metallicities (training data from SEGUE)
Counts
10^1
10^2
10^3
10^4
10^5
−2.5
−2.0
−1.5
−1.0
−0.5
0.0
0.5
$[Fe/H]^{SEGUE}_{photometric}$

Results



Models comparison (2188115 objects)
y = x
-0.39
-0.45
$[Fe/H]^{LAMOSTDR7}_{photometric}$
$[Fe/H]^{SEGUE}_{photometric}$
0.5
0.0
−0.5
−1.0
−1.5
−2.0
−2.5
−2.5
−2.0
−1.5
−1.0
−0.5
0.0
0.5