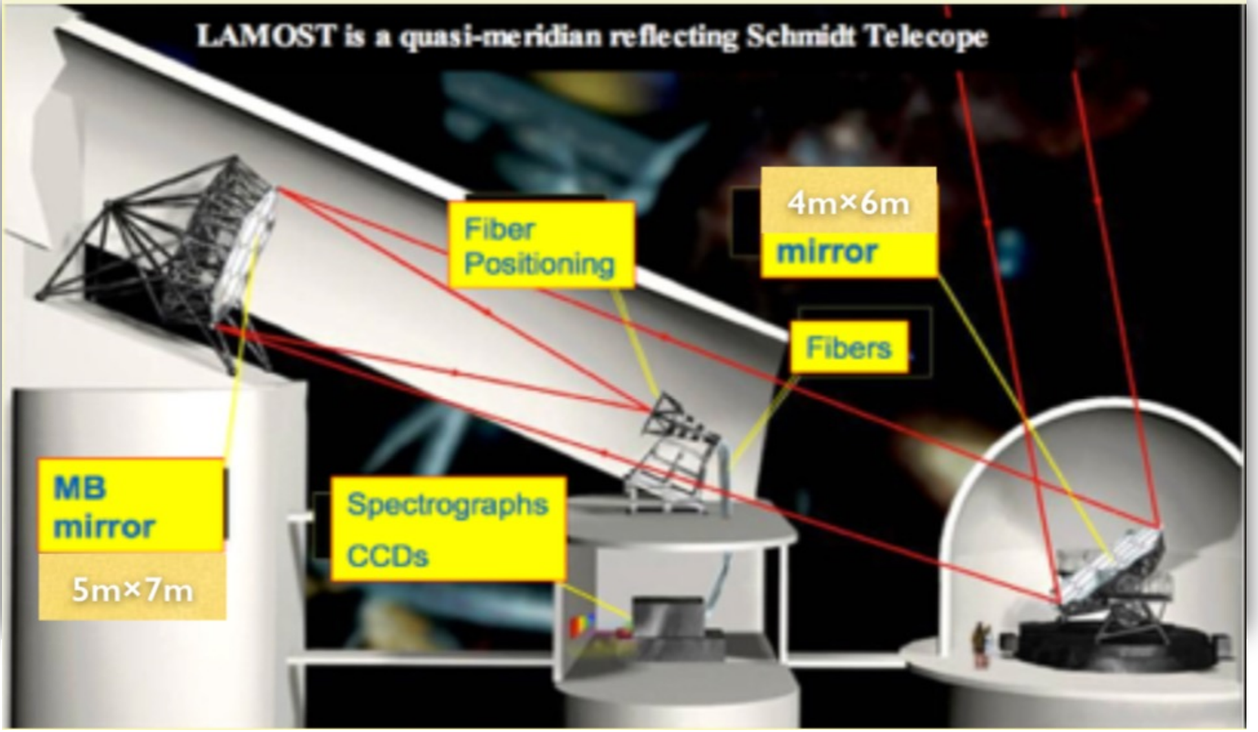
Synergy between SPHEREx and LAMOST

Yang Huang (UCAS/NAOC)

Eff. Aper.	3.6 – 4.88 m
Focal length	20 m
FoV	20 sq. deg.
Sky cover.	Dec > -10 deg.
Num. of fibers	4000
Image quality	80% energy in 2”
Resolution	1800(LRS) / 7500(MRS)
Wavelength coverage	3700 – 9000 Å (LRS) 4950 – 5350 & 6300 – 6800 Å (MRS)

LAMOST timeline:

- 2009.09 – 2011.06 First light and commissioning phase
- 2011.09 – 2012.06 Pilot survey
- 2012.09 – 2017.06 Phase-I Regular survey (LRS)
- 2017.09 – 2018.06 Phase-II Pilot survey (MRS testing)
- 2018.09 – 2023.06 Phase-II Regular survey (LRS & MRS)
- 2023.09 – present Phase-III Regular survey (LRS & MRS)

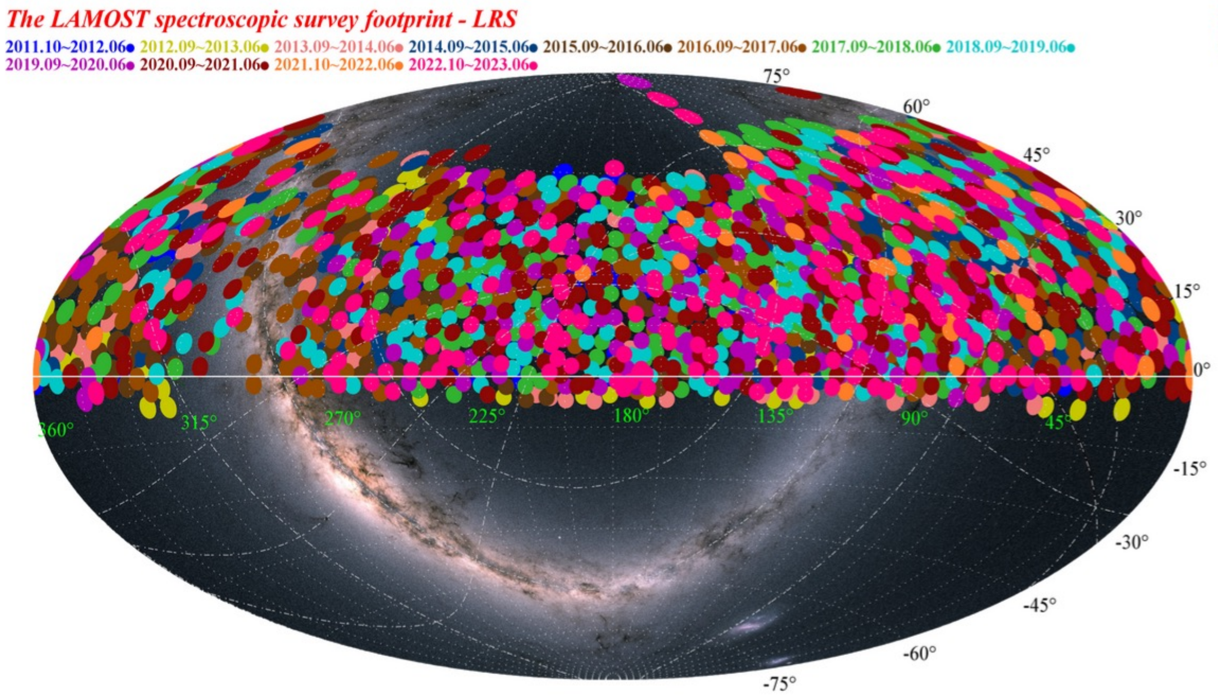


Focal length	20 m
FoV	20 sq. deg.
Sky cover.	Dec > -10 deg.
Num. of fibers	4000
Image quality	80% energy in 2”
Resolution	1800(LRS) / 7500(MRS)
Wavelength coverage	3700 – 9000 Å (LRS) 4950 – 5350 & 6300 – 6800 Å (MRS)

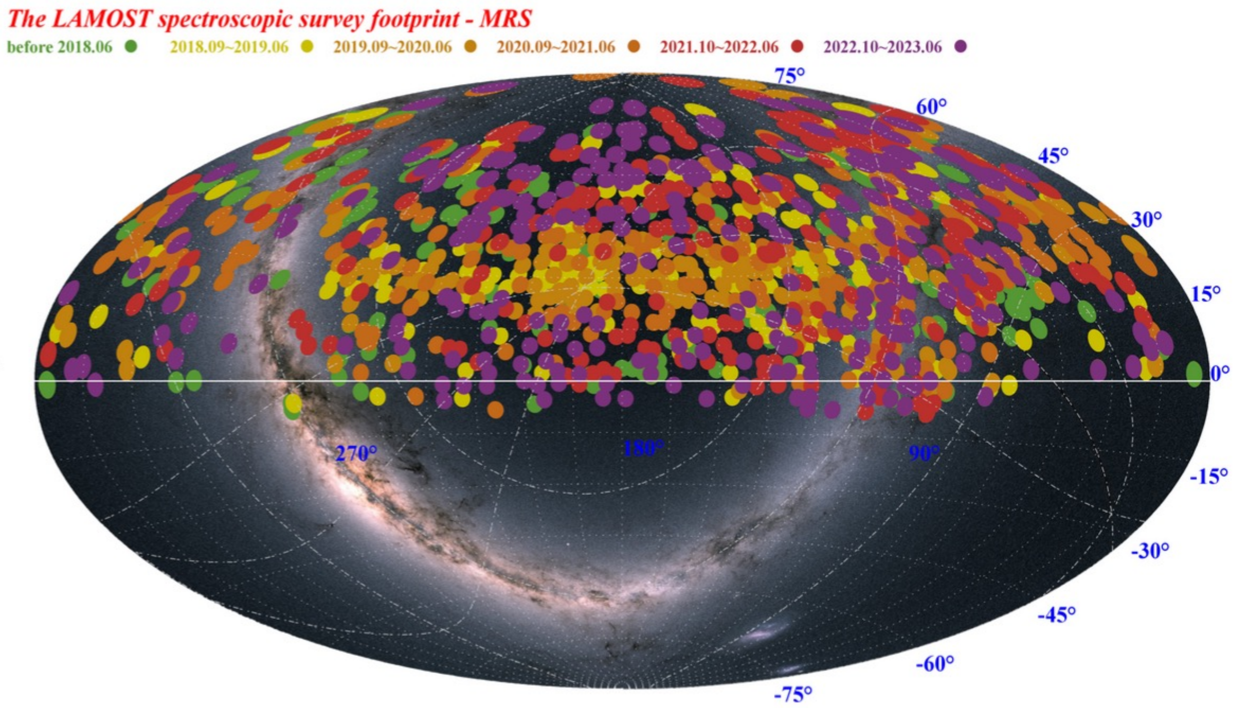
LAMOST timeline:

- 2009.09 – 2011.06 First light and commissioning phase
- 2011.09 – 2012.06 Pilot survey
- 2012.09 – 2017.06 Phase-I Regular survey (LRS)
- 2017.09 – 2018.06 Phase-II Pilot survey (MRS testing)
- 2018.09 – 2023.06 Phase-II Regular survey (LRS & MRS)
- 2023.09 – present Phase-III Regular survey (LRS & MRS)

Internationally released DR11

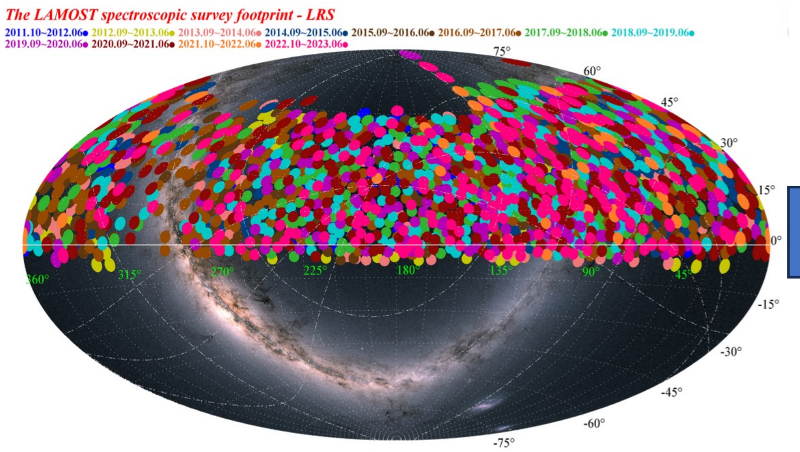


11.93 Million LRS spectra



13.15 Million MRS spectra

Stellar parameter estimates



Wu et al. 2014, IAUS, 206, 340
Luo et al. 2015, RAA, 15, 1095

LAMOST stellar parameter pipeline (LASP)

Template matching with ELODIE to derive RVs and atmospheric parameters

The LAMOST stellar parameter pipeline at Peking University (LSP3)

Template matching with ELODIE (for RVs), MILES (for atmospheric parameters) and Kurucz spectra (for [alpha/Fe])

Xiang et al. 2015, MNRAS, 448, 822 (LSP3)
Li et al. 2016, RAA, 16, 10 ([α/Fe])

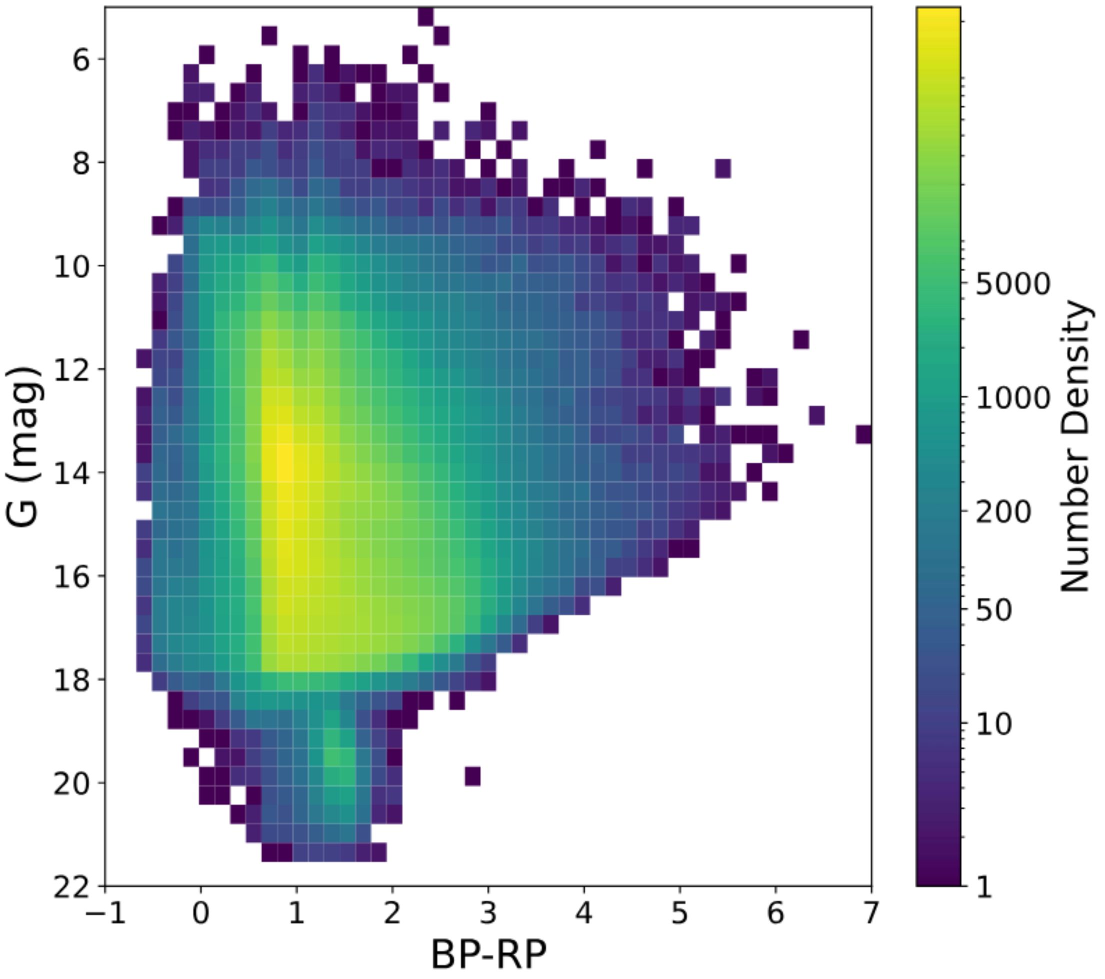
Data driven based on KPCA (improved log *g* for giants, absolute magnitudes, 4 elemental abundances)

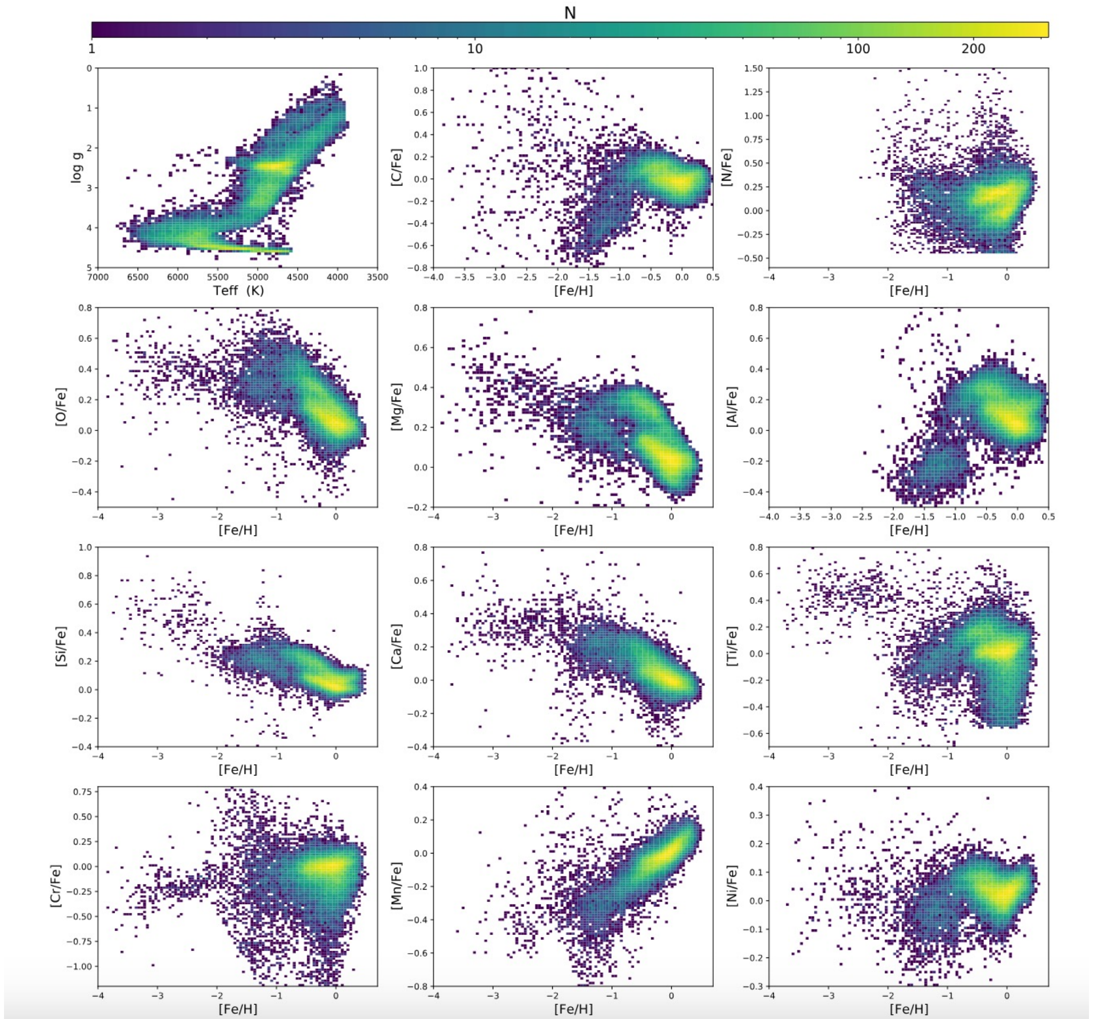
Xiang et al. 2017 MNRAS, 463, 3657

DD-Payne (atmospheric parameters + ~20 element abundances)

Xiang et al. 2019, ApJS, 245, 34
Zhang et al. 2025, ApJS, 279, 5

Stellar parameter estimates

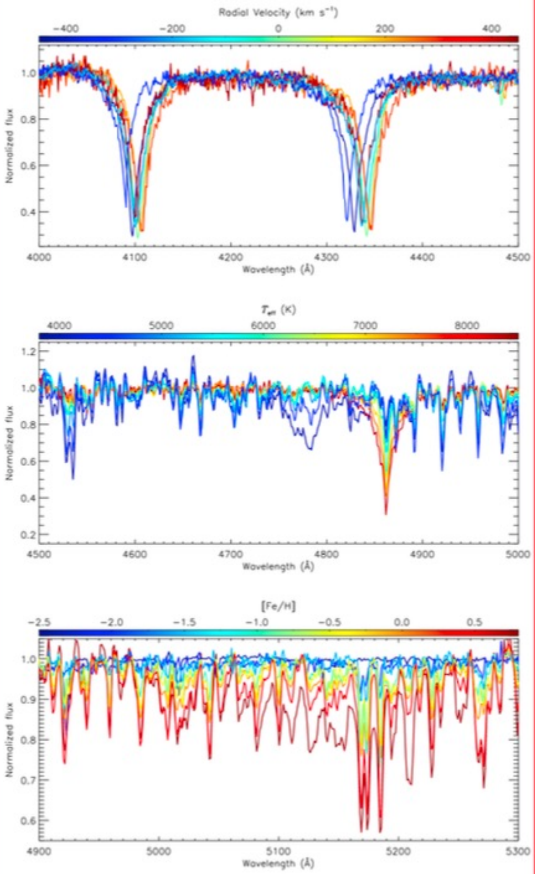




Zhang et al. 2025, ApJS, 279, 5

SpecCLIP: the first version

LAMOST Foundation Model



Data-driven "law" from line features (unsupervised!)

R~2000
Full optical range

"more is different"

SpecCLIP

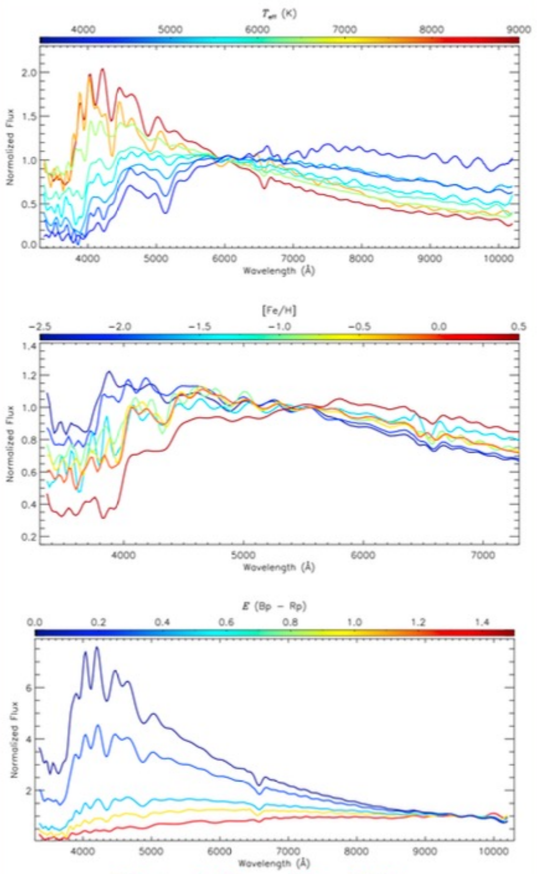
Spectra to parameter

Parameter to spectra

Spectra to spectra

More …

GaiaXP Foundation Model



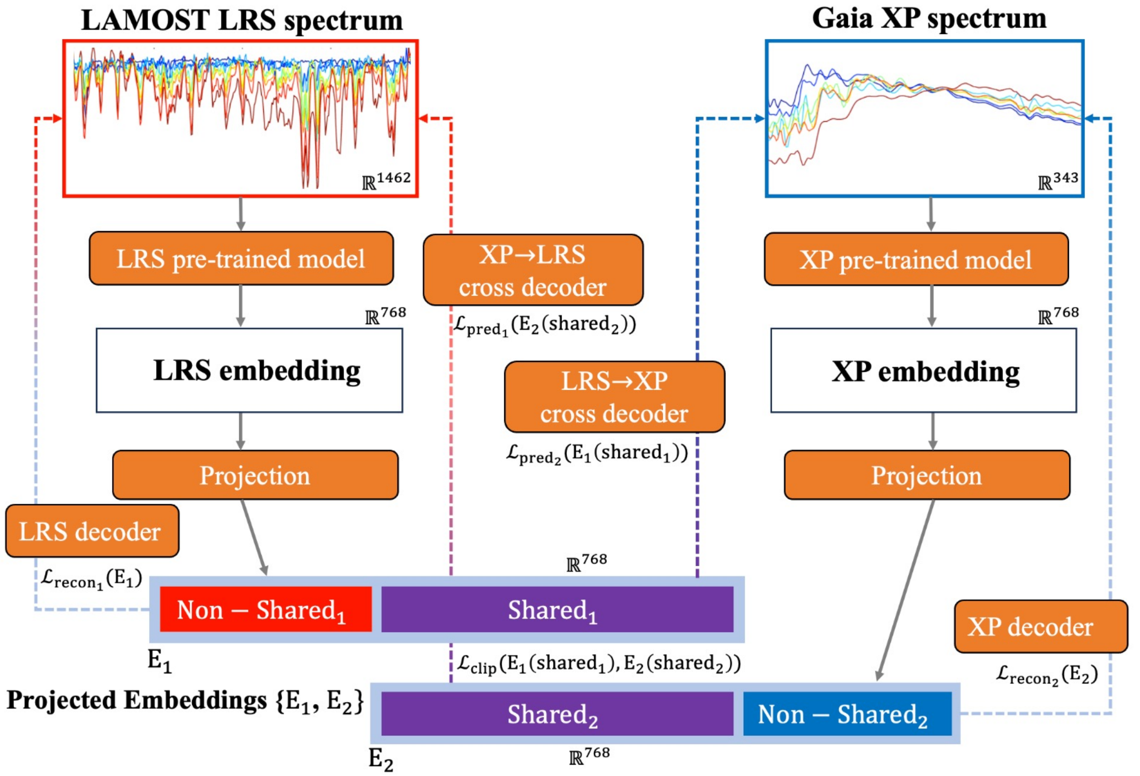
Data-driven "law" from multi-colors & local features (unsupervised!)

R~30-50
Full optical range

SpecCLIP: the first version

Cross-modality prediction, increasing the mutual information between embeddings and spectra

LRS pre-trained model:
Transformer
with mask modeling



XP pre-trained model:
1. Auto-encoders with MLP
2. Transformer with mask modeling

CLIP with contrastive loss:
"maximize the similarity between positive pairs and minimize it between negative pairs"
CLIP (Radford et al. 2021)
AstroCLIP (Parker et al. 2023)

Different CLIP:
Model1: without decoders, original CLIP model — "CLIP",
Model2: with only cross decoders for prediction — "CLIP-p",
Model3: with all decoders, implicitly contains the shared and non-shared space in the embedding — "CLIP-pr",
Model4: Explicitly split into shared and non-shared space in the embedding — "CLIP-split".

SpecCLIP: One-sentence summary!

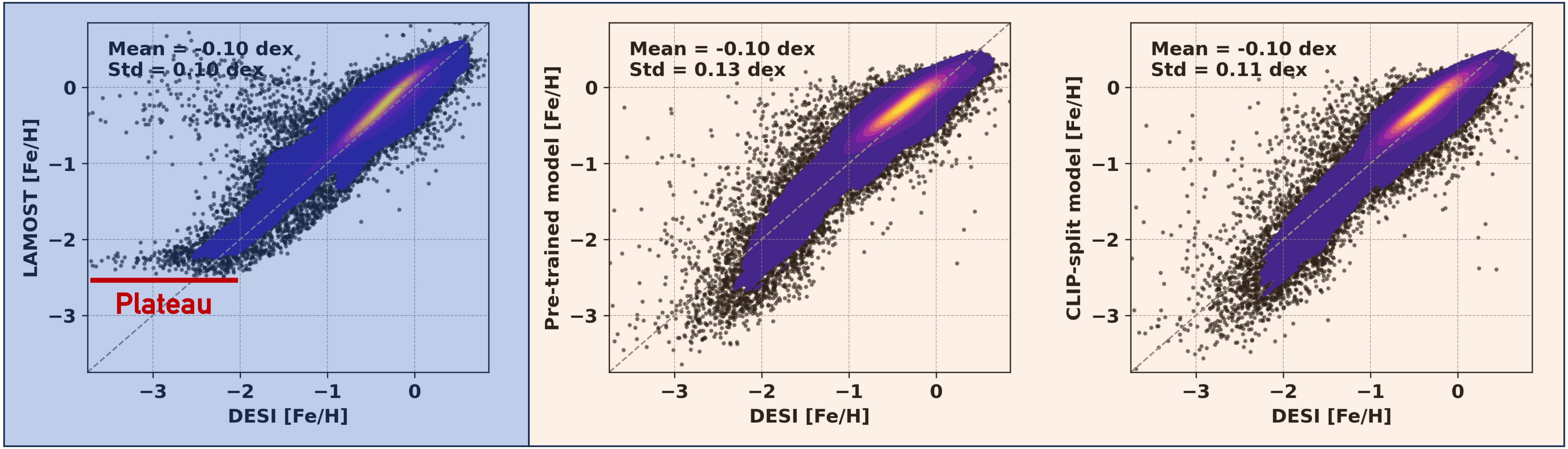
✓ Self-supervised learning to summarize empirical relationships/correlations from large spectral datasets

SpecCLIP is a foundation model designed to align stellar spectra across different modalities using the CLIP-like algorithm.

✓ Robust and informative embeddings for down-stream tasks
✓ Transform between different modalities

[Fe/H] Estimation: Plateau bias is eliminated

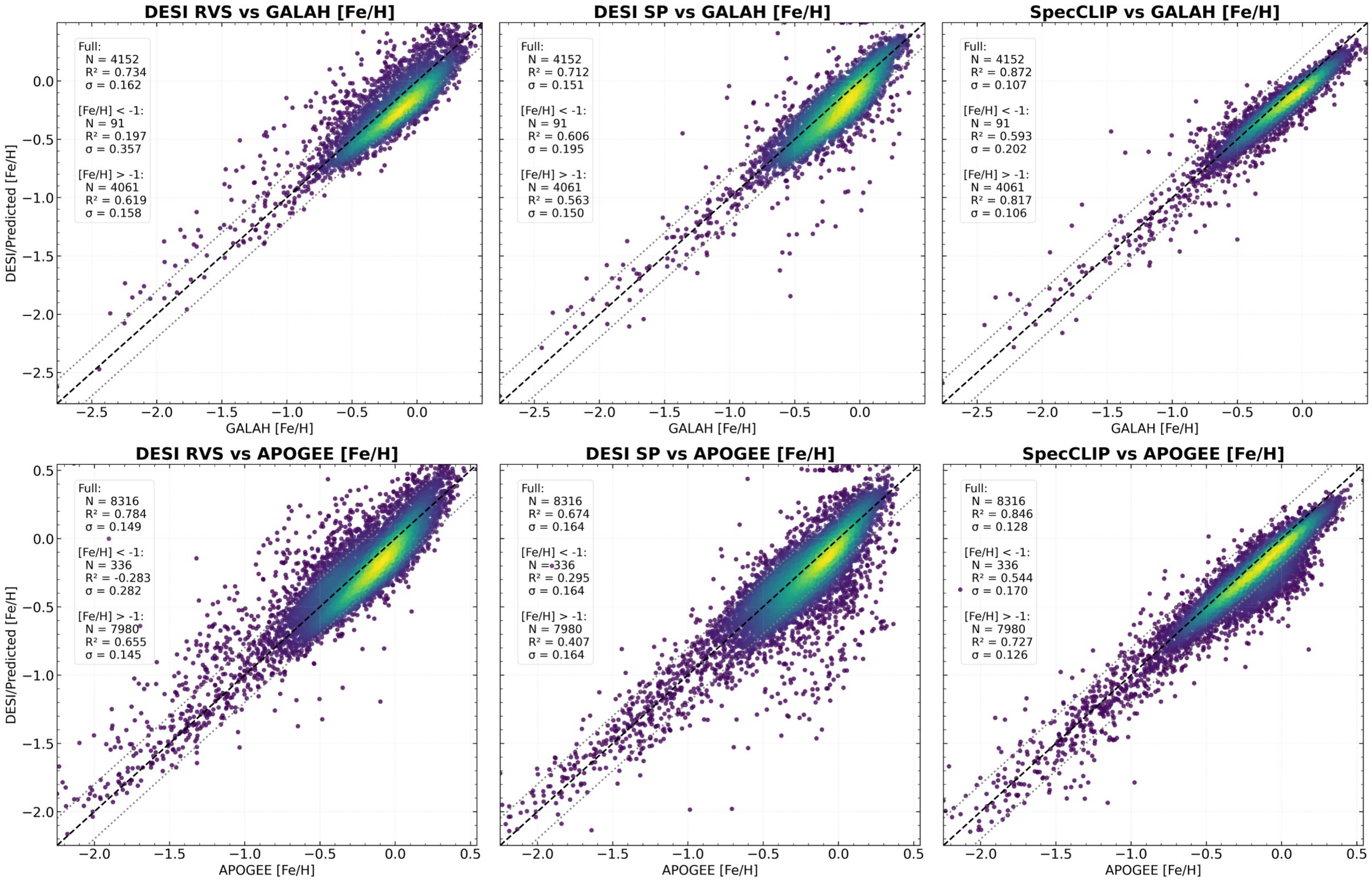
LAMOST versus DESI



LAMOST official pipeline
(Template matching)

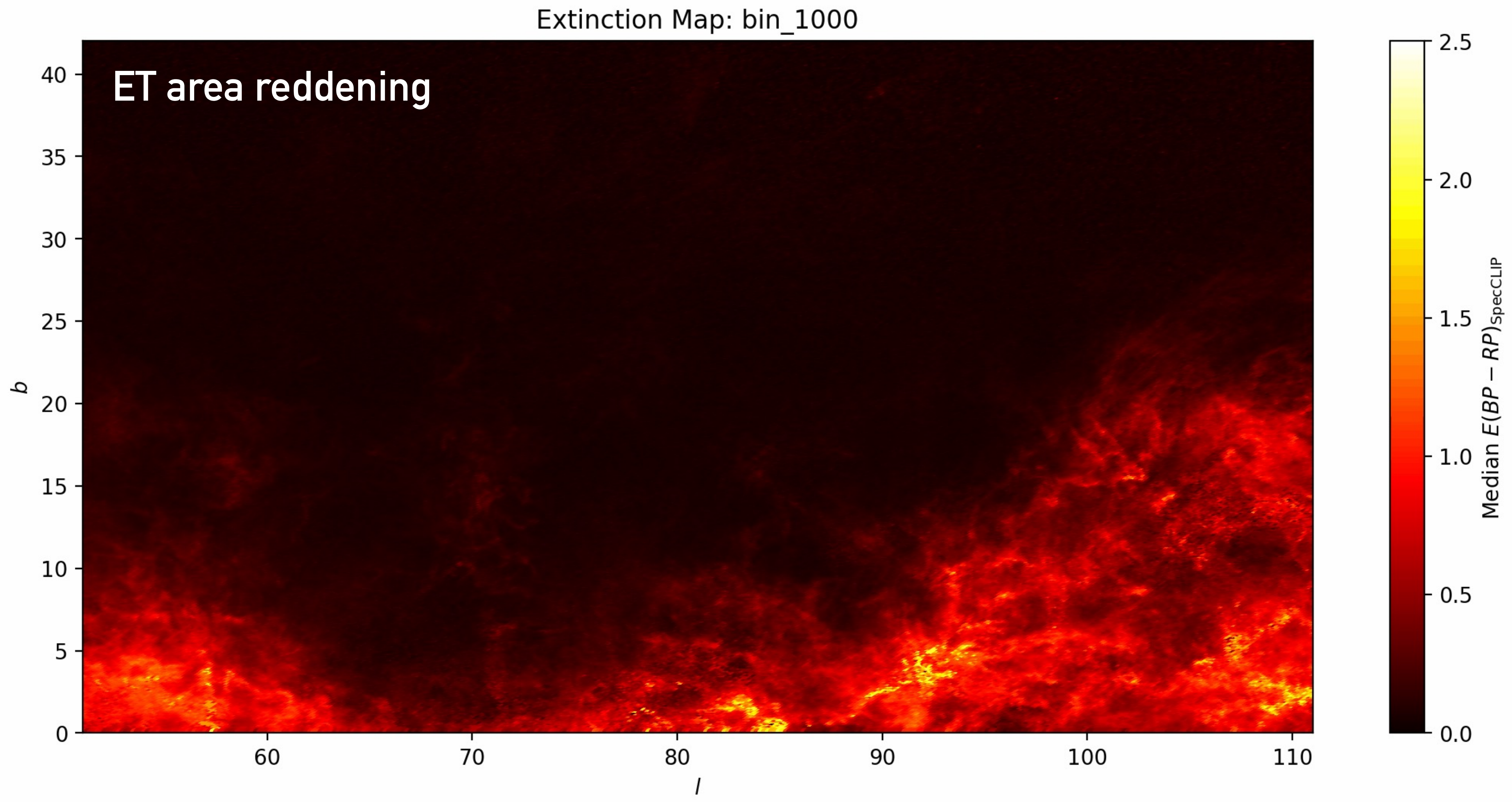
This work

Extremely excellent scalability: zero-shot on DESI DR1



RVSpecfit FERRE SpecCLIP

SpecCLIP: A foundation model for stellar spectral analysis



SpecCLIP: A foundation model for stellar spectral analysis

XS Zhao, Y Huang* et al. arXiv: 2507.01939

	LAMOST LRS Spectra	Gaia XP Spectra
Model size	43 Million	43 Million
Downstream tasks	● Atmospheric parameters (including RV) ● Elements abundance ratios (10) ● Asteroseismology parameters (for giant stars only) ● Reddening ● …	● Atmospheric parameters ● Elements abundance ratios (3) ● Reddening ● …
Value-added catalog	7 Million (for spectra with SNR>20)	>200 million
Performances	● Robust and efficient for different kinds of downstream tasks ● The speed of RV estimation even quicker than CCF	● Robust and efficient for different kinds of downstream tasks ● Excellent performance for EMP/UMP ([Fe/H] < -4.0)

Apply SpecCLIP to LAMOST+SPHEREx

The infrared bands indeed carry some information about metallicity!

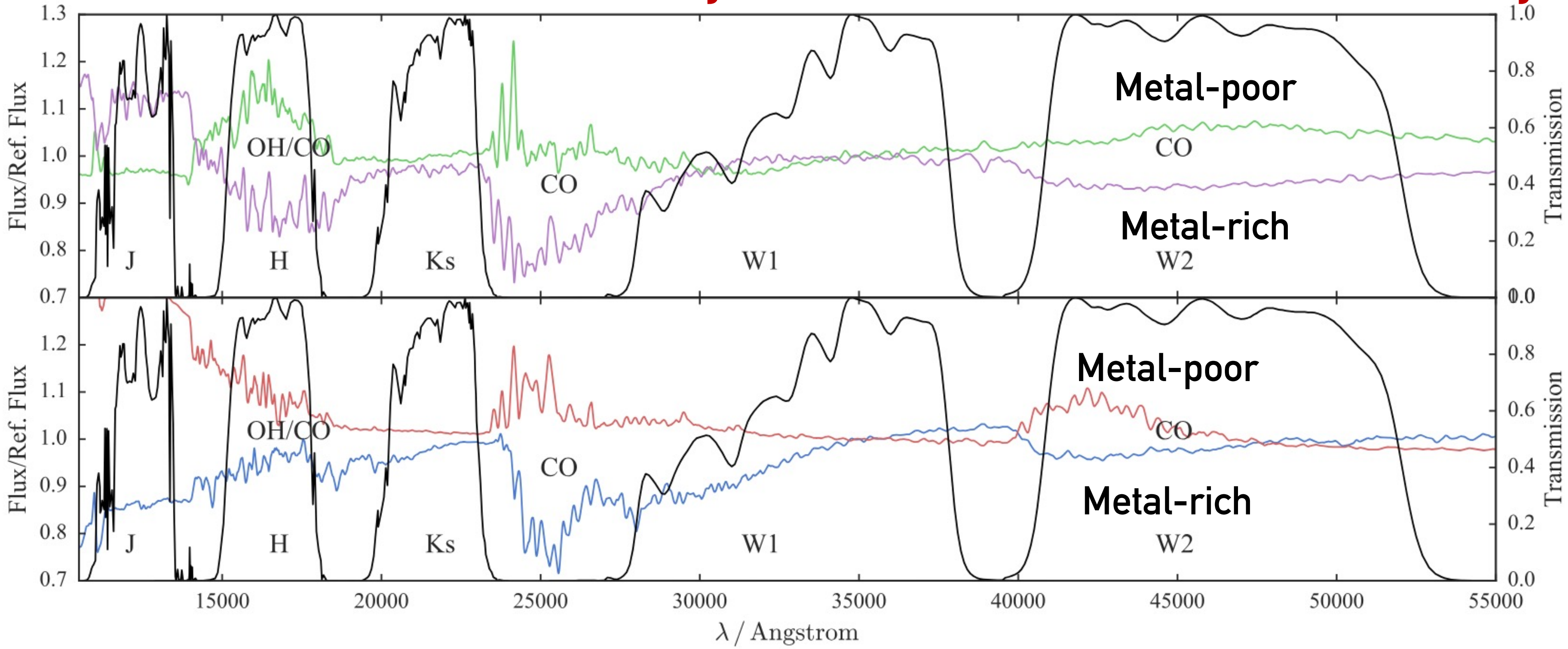
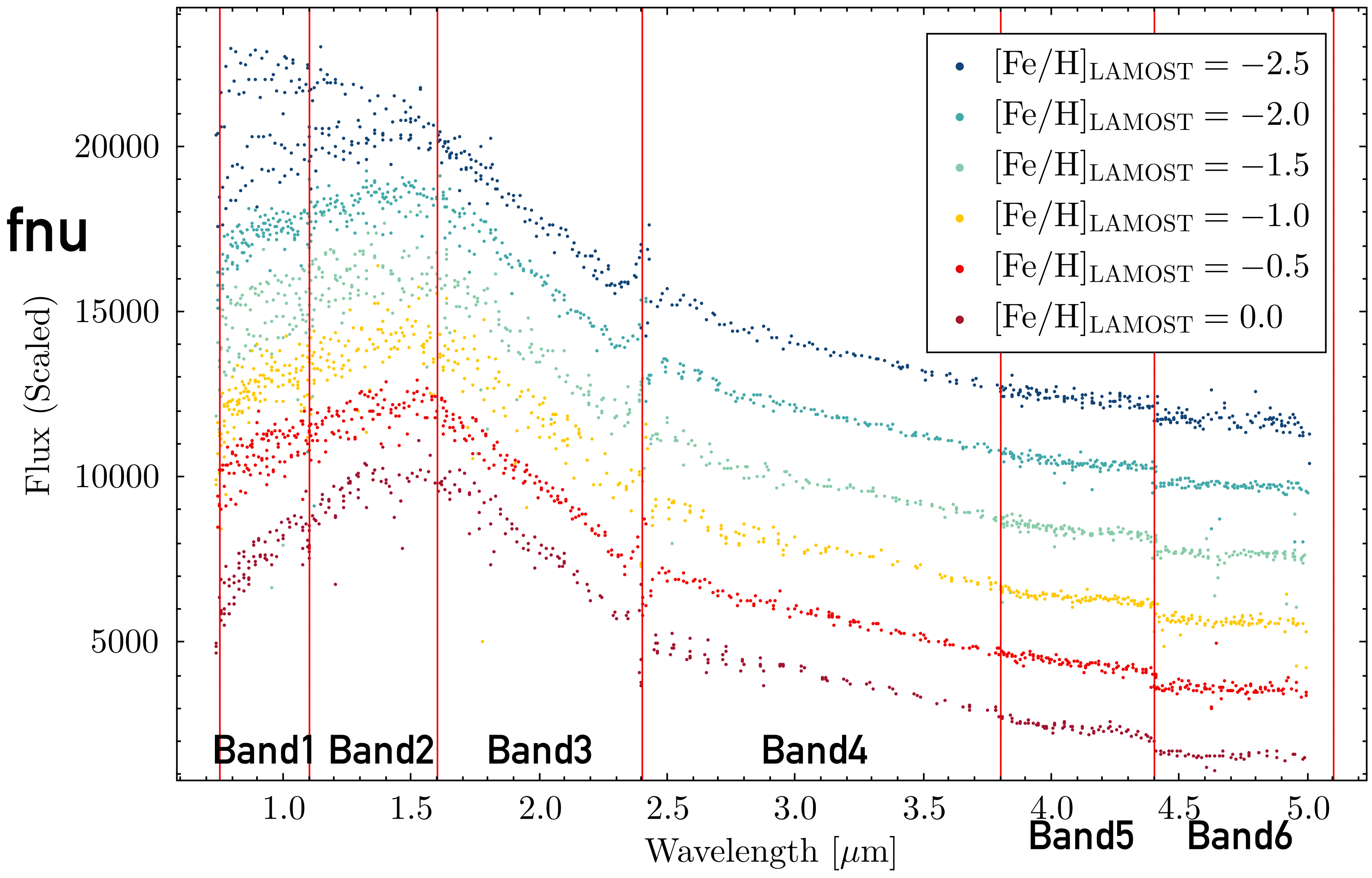


Figure 1. The sensitivity of WISE to stellar metallicity: Ratios of MARCS models with infrared bandpasses overplotted normalized at $\lambda = 35\,000$ Å. The reference model has $T_{\rm eff} = 3500$ K, $\log g = 0$ dex, and [M/H] $= 0$ dex. The green and purple lines show models with [M/H] $= -1$ dex and [M/H] $= 1$ dex, and blue and red lines $T_{\rm eff} = 3200$ K and $T_{\rm eff} = 4000$ K. Note the strong gradients in *W2* due to the CO feature. The effects of temperature and metallicity variations in *W2* can be distinguished using the bluer 2MASS bands.

Fallows et al. 2022

Apply SpecCLIP to LAMOST+SPHEREx

A more accurate flux calibration is required!

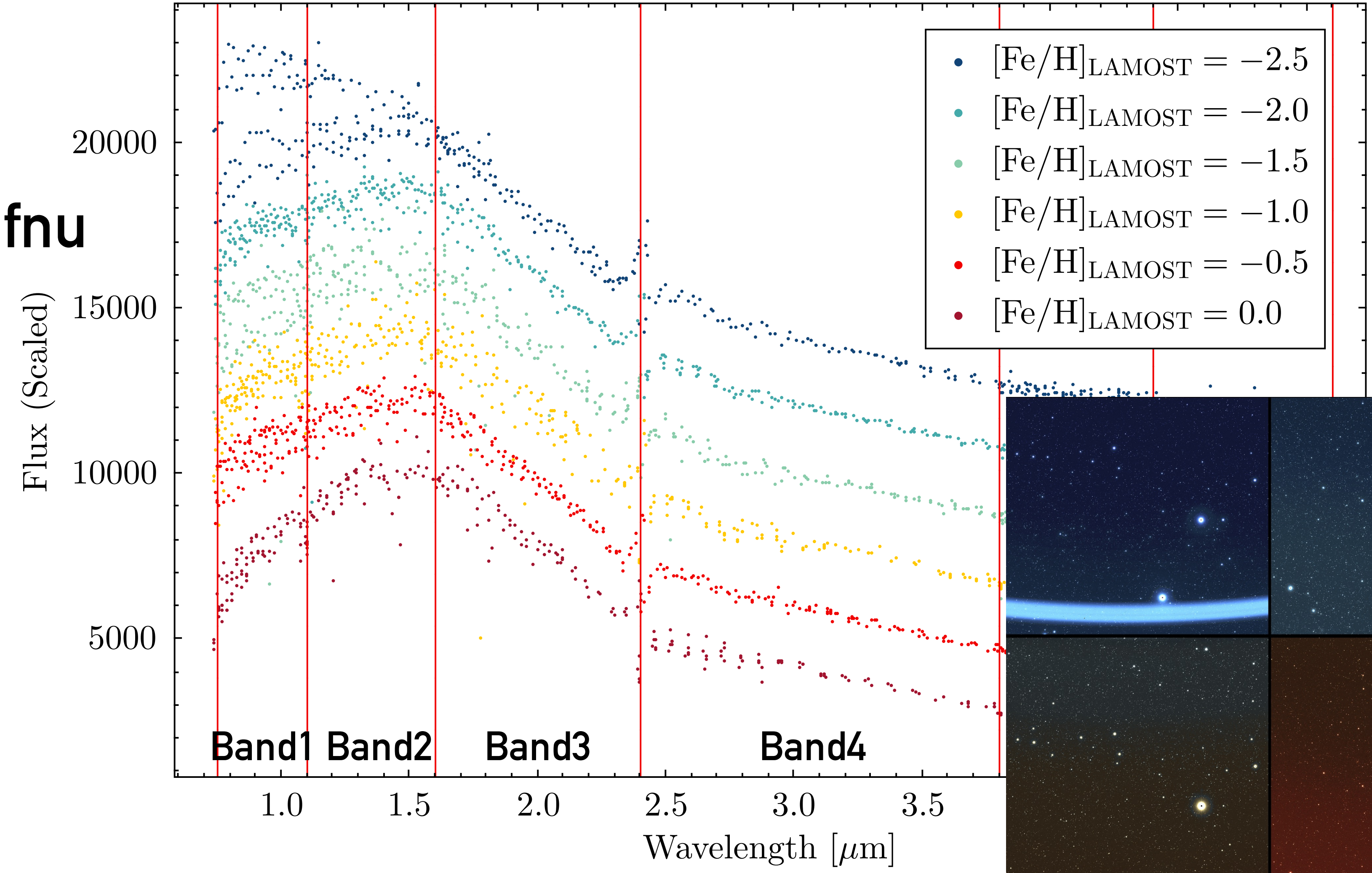


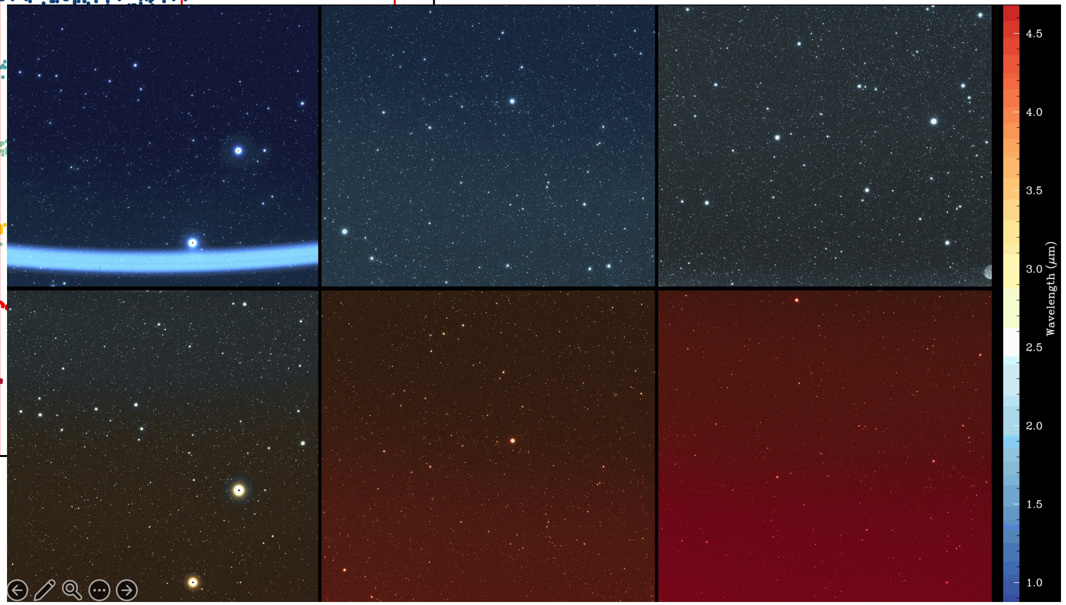
Extremely important for:

- **High reddening region (e.g., bulge)**
- **MW outskirt (red giant – far away)**

Apply SpecCLIP to LAMOST+SPHEREx

A more accurate flux calibration is required!



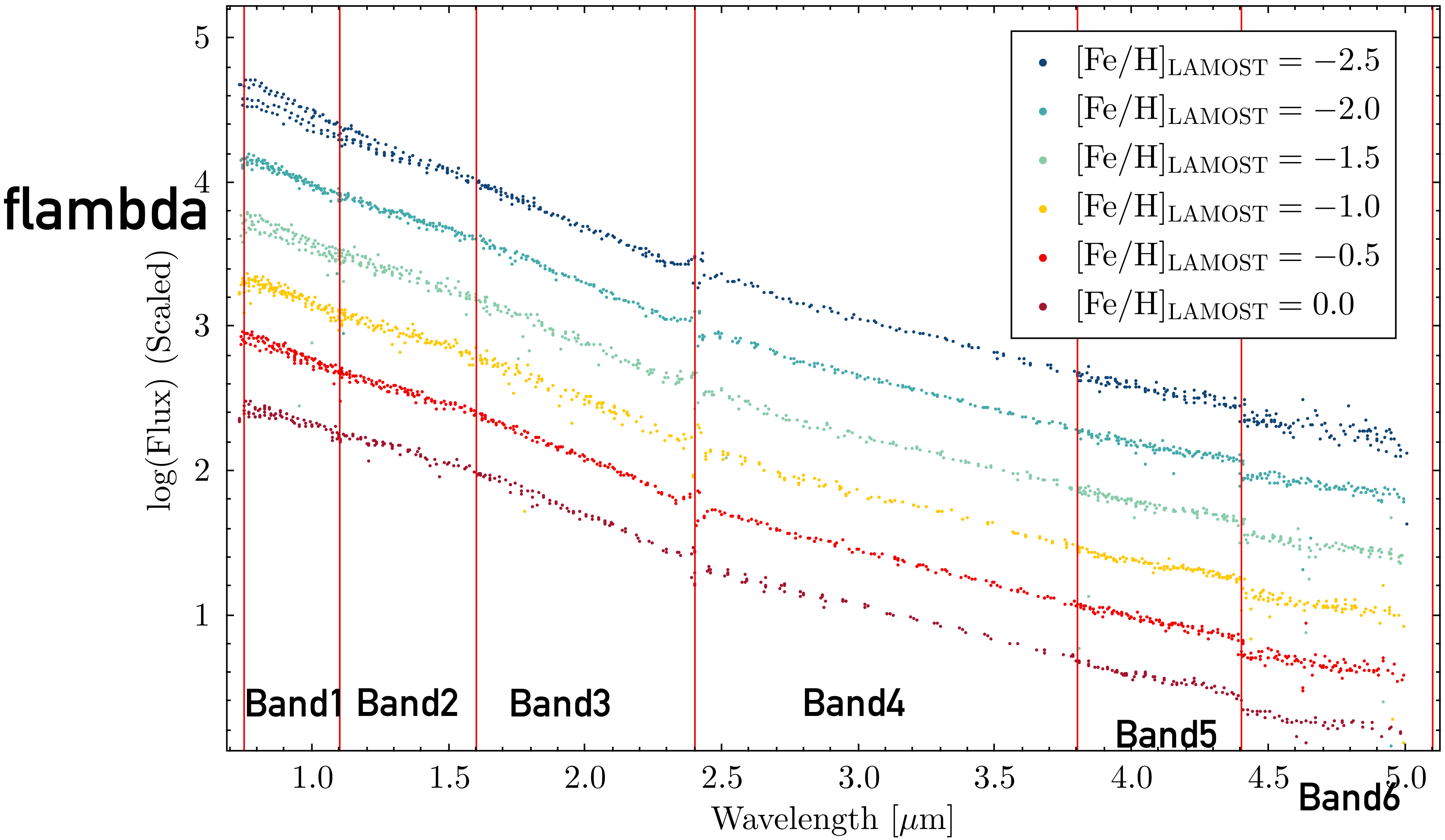


Extremely important for:

- **High reddening region**
- **MW outskirt (red giant – far away)**

Apply SpecCLIP to LAMOST+SPHEREx

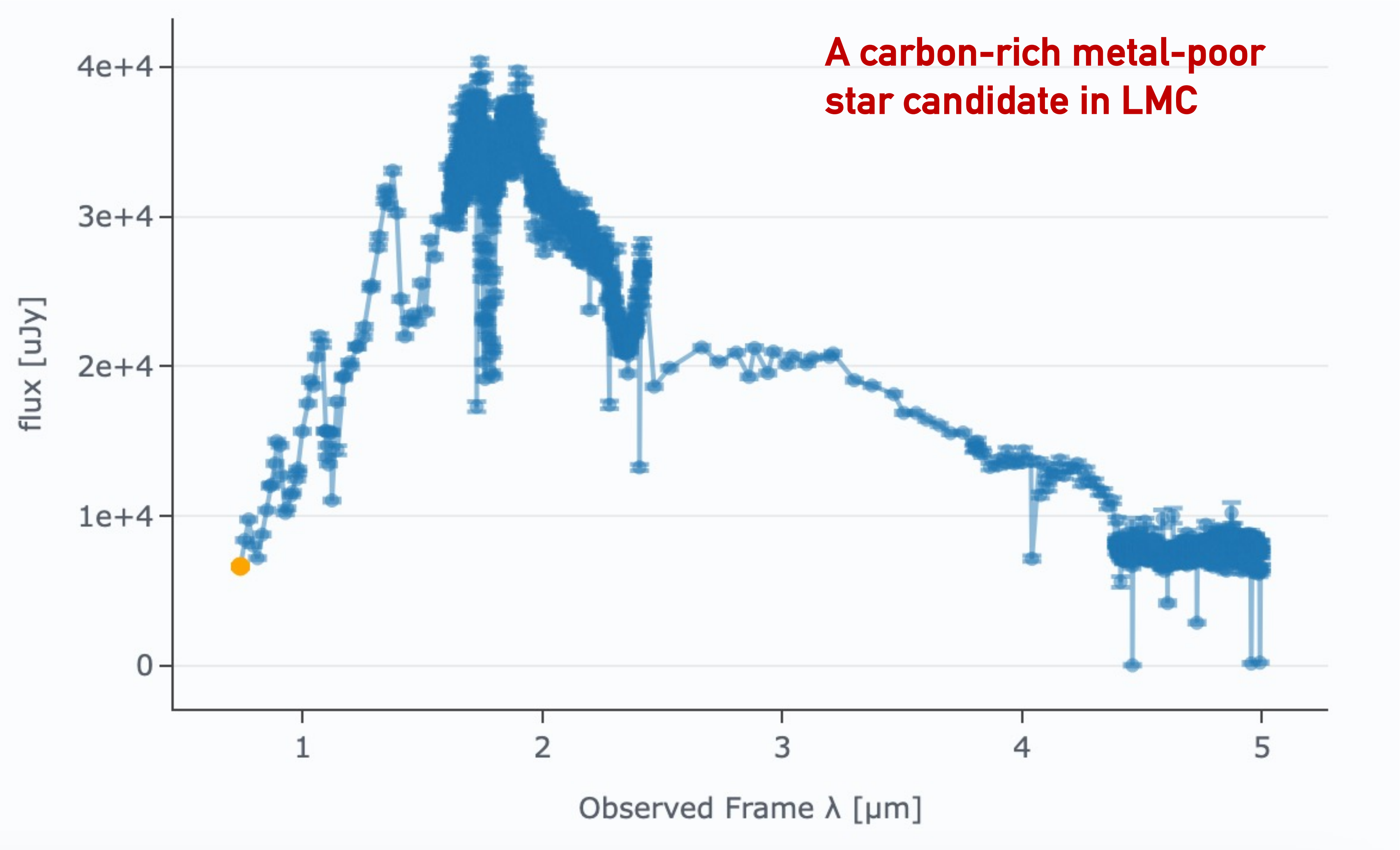
A more accurate flux calibration is required!



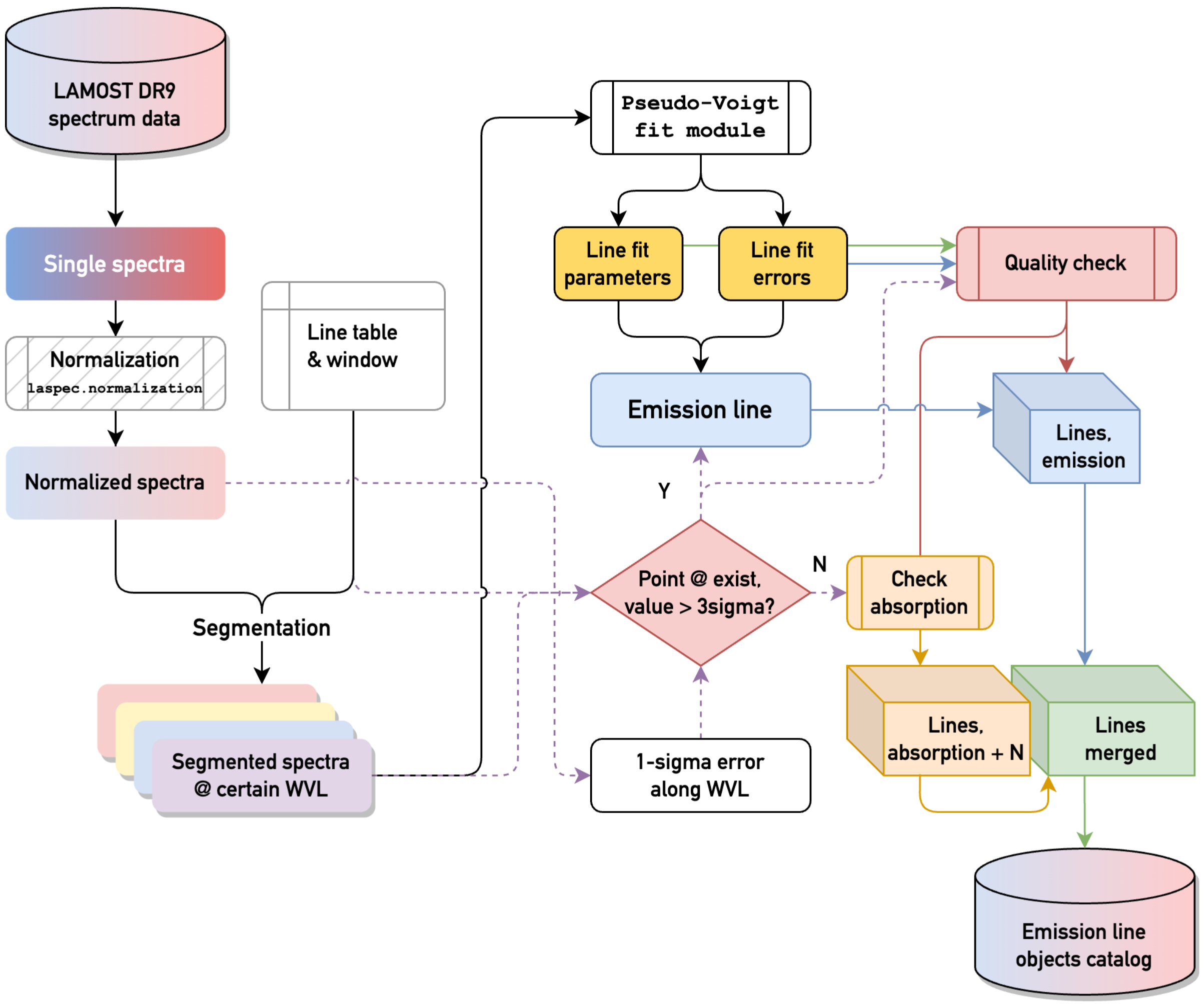
Extremely important for:

- **High reddening region (e.g., bulge)**
- **MW outskirt (red giant – far away)**

Apply SpecCLIP to LAMOST+SPHEREx



A pipeline to identify emission line objects from LAMOST



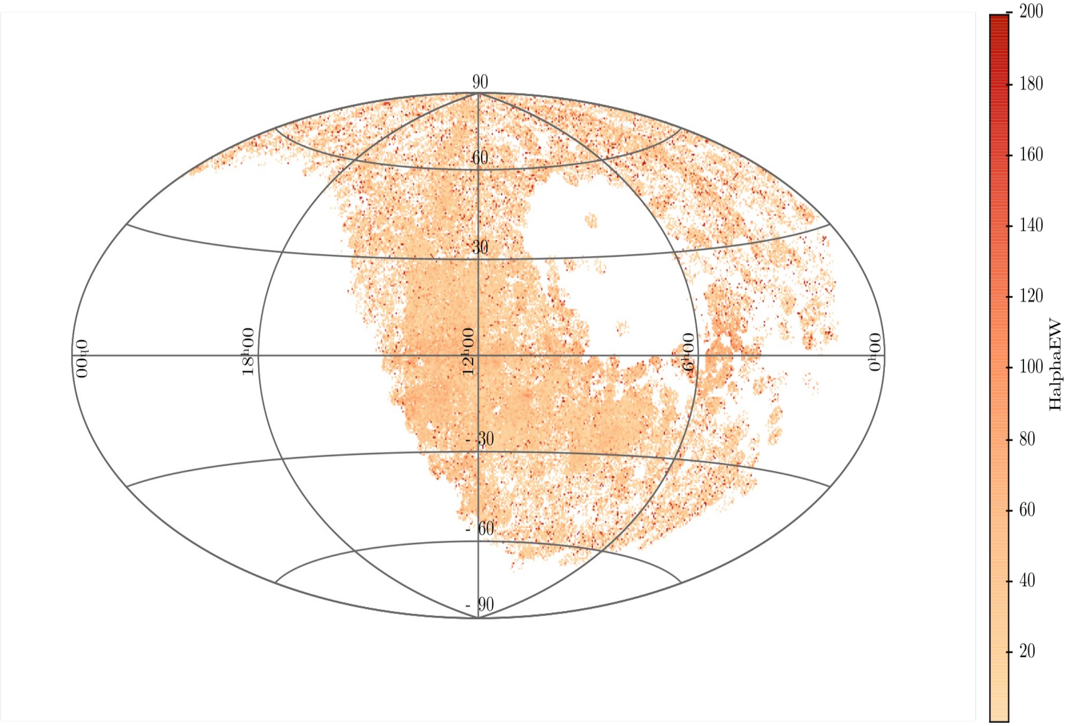
Li, Huang et al. In prep.

A pipeline to identify emission line objects from LAMOST

The catalog was extracted from v1.0 catalog with the following criteria:

Ha_6564_emission == true;
Ha_6564_EW > 0; bad_data == false;
high_redshift == false

A total of 241,618 sources were extracted.

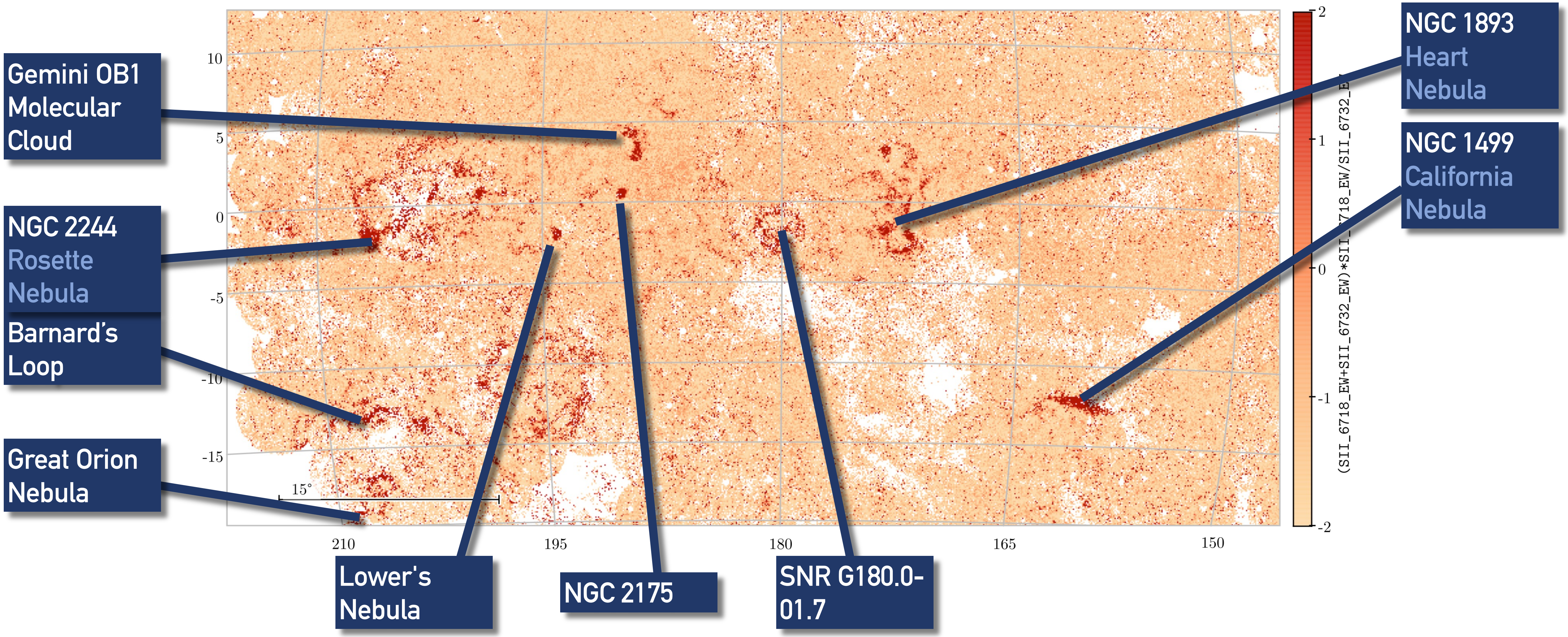


Sky coverage of H-alpha emission catalog

Li, Huang et al. In prep.

A pipeline to identify emission line objects from LAMOST

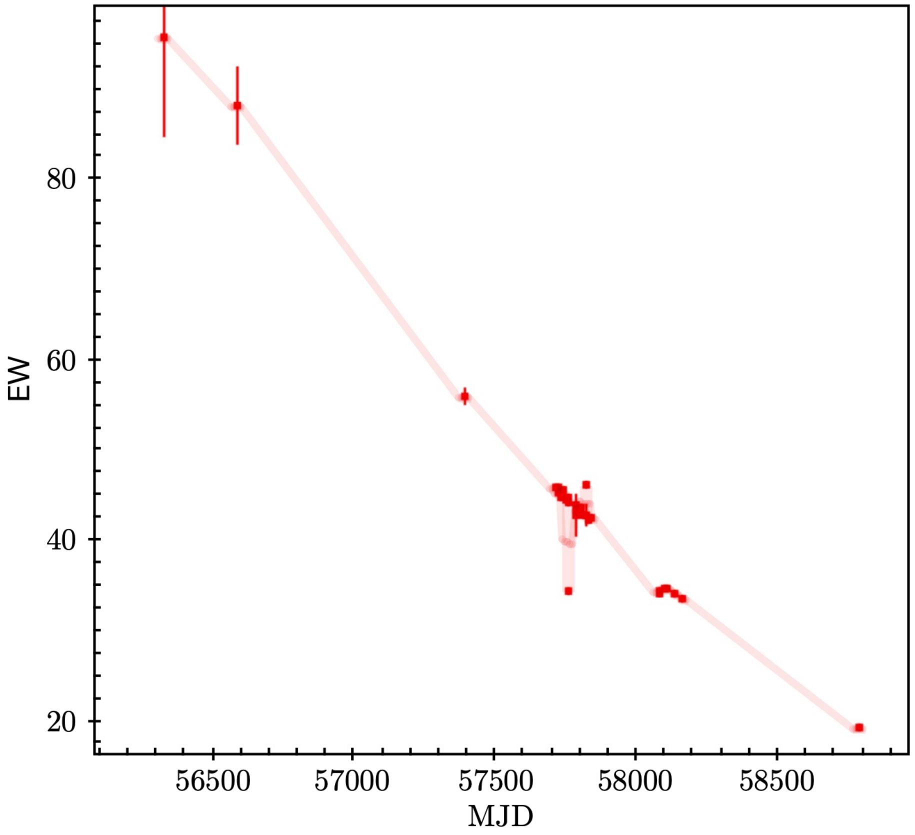
SII forbidden lines (6716/6731) near the galactic anti-center



A pipeline to identify emission line objects from LAMOST

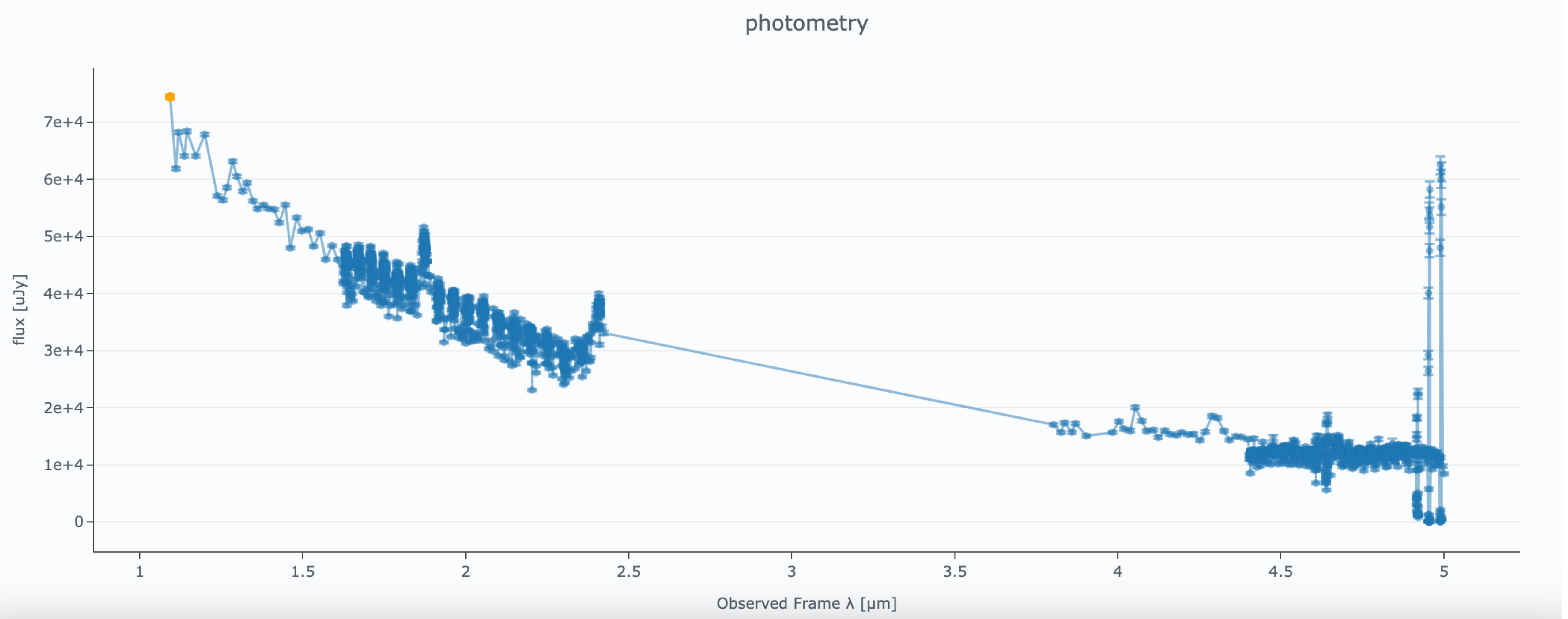
Time-domain monitoring

- A dynamic universe has various manifestations. Variations in the spectra of diverse objects are embedded in it.
- As a showcase, we demonstrate that the catalog is capable of tracking the changing of H-alpha EW.



A showcase of a Be star with a time-changing H-alpha EW

Systematic studying emission line objects with LAMOST+SPHEREx



SPHEREx of a Be star

Systematic studying emission line objects with LAMOST+SPHEREx

Lots of sciences: e.g., YSO, CV, Be, eruption stars …