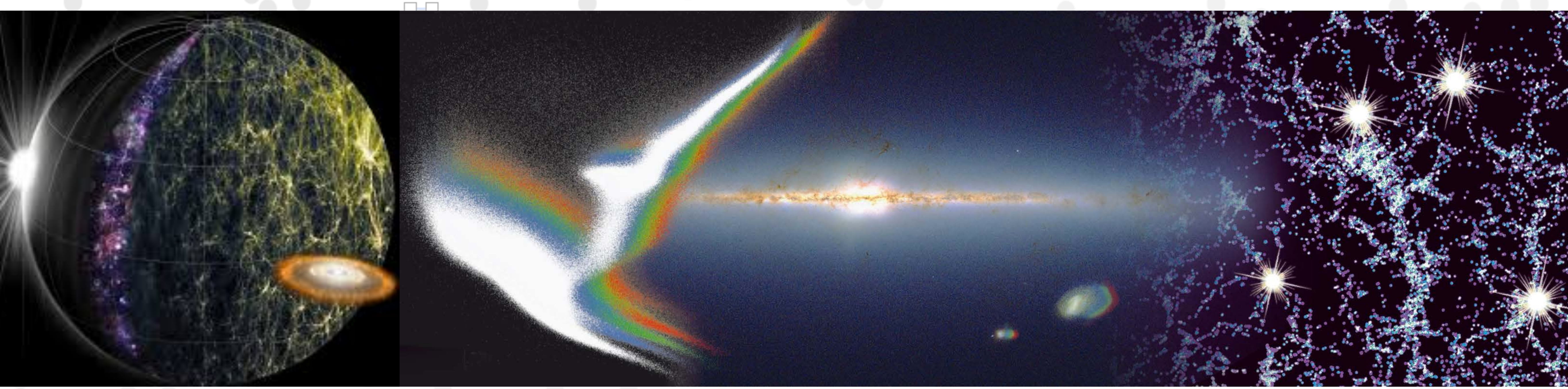


Galactic Archaeology with SPHEREx & SDSS-V (all the stars)

Melissa Ness, Columbia University, Feb 24, 2020
SPHEREx workshop @ Flatiron

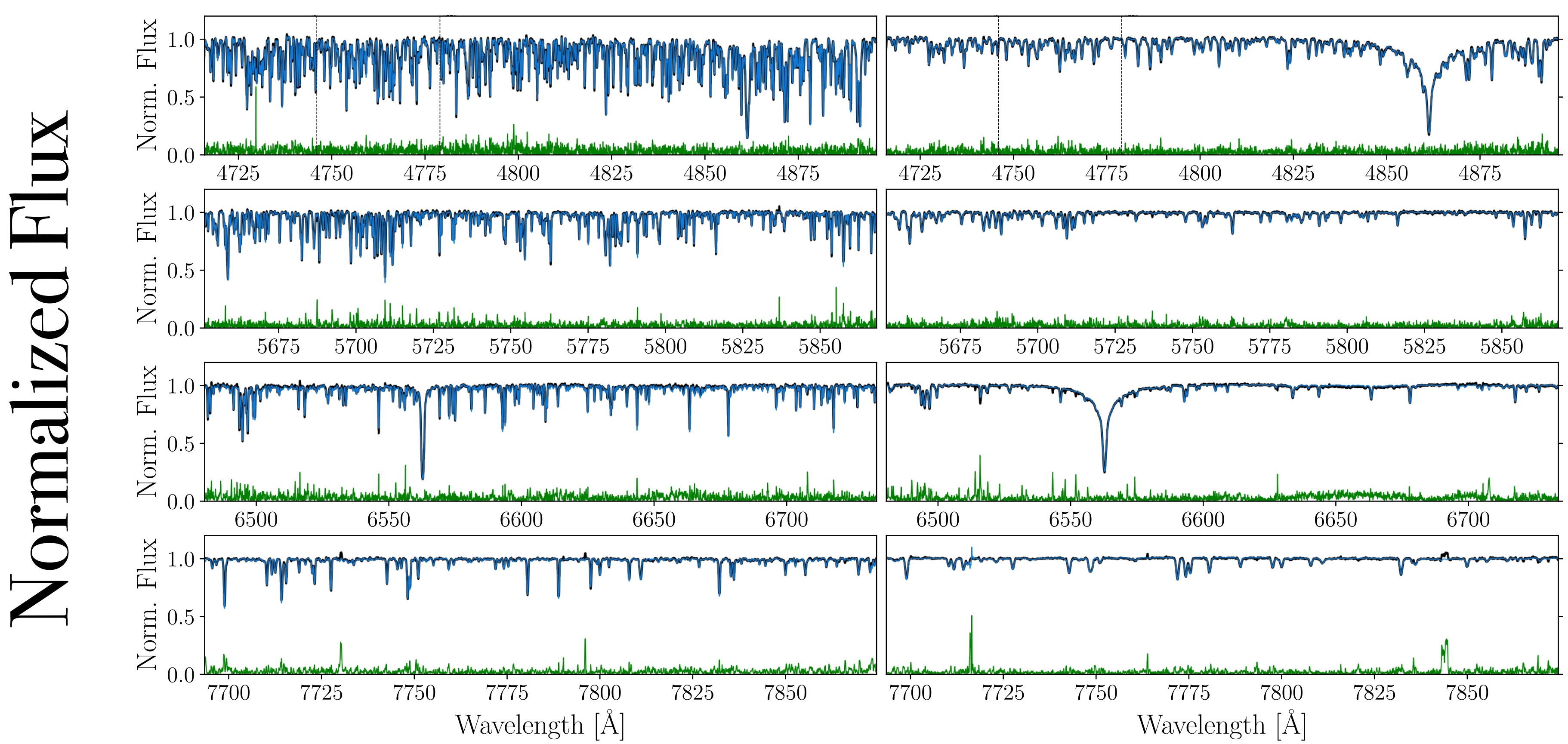
1

Who am I?

1) Survey Scientist – *SDSS V's Milky Way Mapper*
2) I derive measurements from stellar spectra (Teff, logg, age, chemical abundances): use these to say things (about Galaxy's formation)
3) Build (simple) data-driven (generative) models: R=30,000 to 1800

Who am I?

1) Survey Scientist – *SDSS V's Milky Way Mapper*
2) I derive measurements from stellar spectra (Teff, logg, age, chemical abundances): use these to say things (about Galaxy's formation)
3) Build (simple) data-driven (generative) models: R=30,000 to 1800

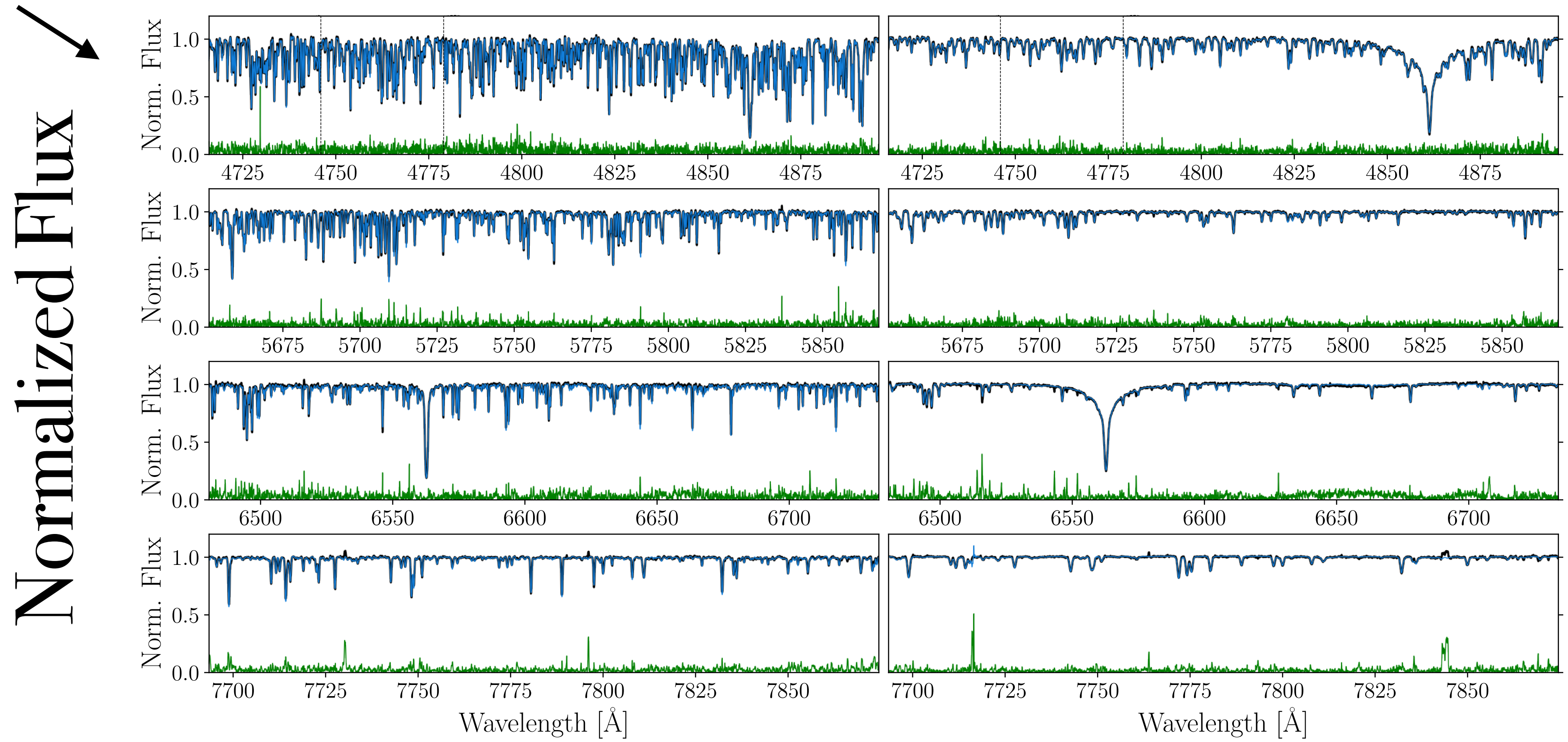


GALAH, Buder+ 2018

Who am I?

1) Survey Scientist – *SDSS V's Milky Way Mapper*
2) I derive measurements from stellar spectra (Teff, logg, age, chemical abundances): use these to say things (about Galaxy's formation)
3) Build (simple) data-driven (generative) models: R=30,000 to 1800

red giant



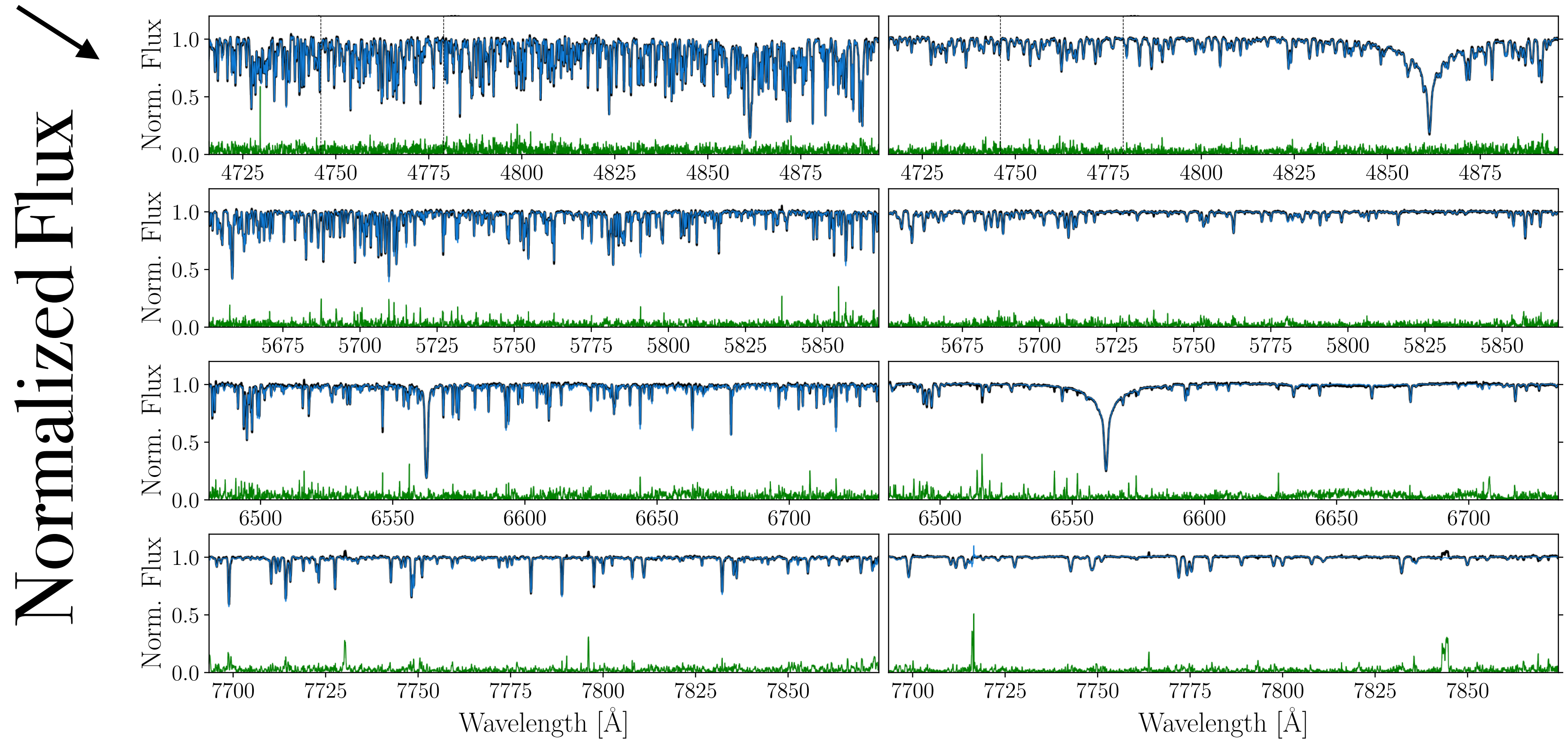
GALAH, Buder+ 2018

Who am I?

1) Survey Scientist – *SDSS V's Milky Way Mapper*
2) I derive measurements from stellar spectra (Teff, logg, age, chemical abundances): use these to say things (about Galaxy's formation)
3) Build (simple) data-driven (generative) models: R=30,000 to 1800

red giant

main sequence



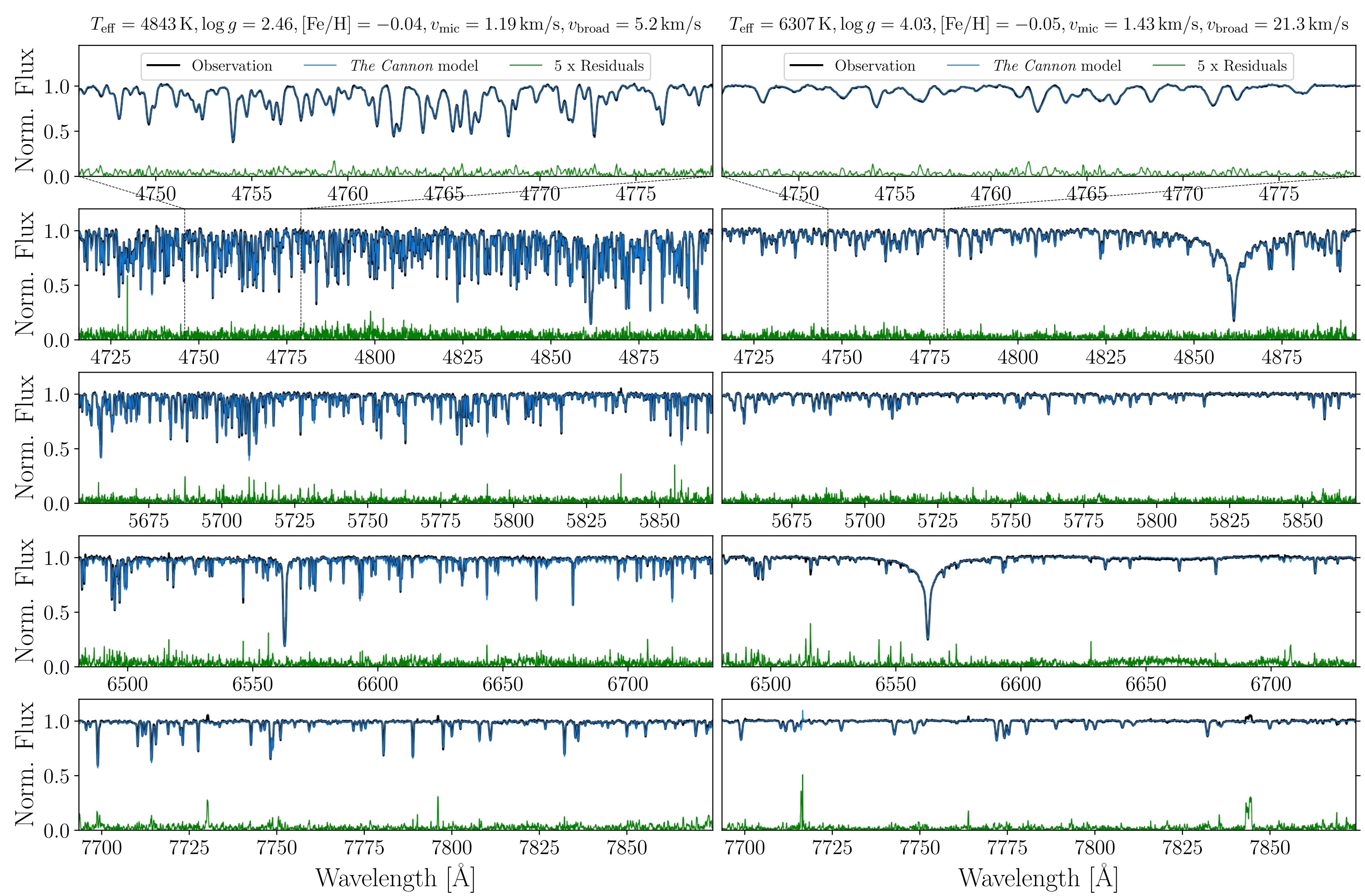
GALAH, Buder+ 2018

Who am I?

1) Survey Scientist – *SDSS V's Milky Way Mapper*
2) I derive measurements from stellar spectra (Teff, logg, age, chemical abundances): use these to say things (about Galaxy's formation)
3) Build (simple) data-driven (generative) models: R=30,000 to 1800

red giant

Normalized Flux



main sequence

GALAH, Buder+ 2018

What am I going to propose?

Now

Later

What am I going to propose?

Now

Characterise **what** information we can recover from *SPHEREx* (stellar) data

Later

What am I going to propose?

Now

Characterise **what** information we can recover from *SPHEREx* (stellar) data

→ Apply (data-driven) methods developed within SDSS

→ precision of inferred stellar properties (e.g. Teff, logg, [M/H]) across (evolutionary state, SNR)

Later

What am I going to propose?

Now

Characterise **what** information we can recover from *SPHEREx* (stellar) data

→ Apply (data-driven) methods developed within SDSS

→ precision of inferred stellar properties (e.g. Teff, logg, [M/H]) across (evolutionary state, SNR)

→ Build a theoretical / empirical / semi-empirical SPHEREx stellar library

Later

What am I going to propose?

Now

Characterise **what** information we can recover from *SPHEREx* (stellar) data

→ Apply (data-driven) methods developed within SDSS

→ precision of inferred stellar properties (e.g. Teff, logg, [M/H]) across (evolutionary state, SNR)

→ Build a theoretical / empirical / semi-empirical SPHEREx stellar library

Later

Where do data-driven models *not* match the data?

What am I going to propose?

Now

Characterise **what** information we can recover from *SPHEREx* (stellar) data

→ Apply (data-driven) methods developed within SDSS

→ precision of inferred stellar properties (e.g. Teff, logg, [M/H]) across (evolutionary state, SNR)

→ Build a theoretical / empirical / semi-empirical SPHEREx stellar library

Later

Where do data-driven models *not* match the data?

→ Non-stellar features in the spectrum (dust, molecules in ISM)

→ Peculiar chemical composition (dynamical interactions, planet ingestion)

SDSS-V

5 year program starting 2020 - both hemispheres

3 Science Programs

1) Milky Way Mapper (stars)
2) Black Hole Mapper
3) Local Volume Mapper (galaxies)

2 Infrastructure Investments

1) Fiber Positioning System
2) Local Volume Mapper

Kollmeier et al., 2017 – http://adsabs.harvard.edu/abs/2017arXiv171103234K

SDSS-V

5 year program starting 2020 - both hemispheres

3 Science Programs

1) Milky Way Mapper (stars)
2) Black Hole Mapper
3) Local Volume Mapper (galaxies)

2 Infrastructure Investments

1) Fiber Positioning System
2) Local Volume Mapper

Kollmeier et al., 2017 – http://adsabs.harvard.edu/abs/2017arXiv171103234K

SDSS-V

5 year program starting 2020 - both hemispheres

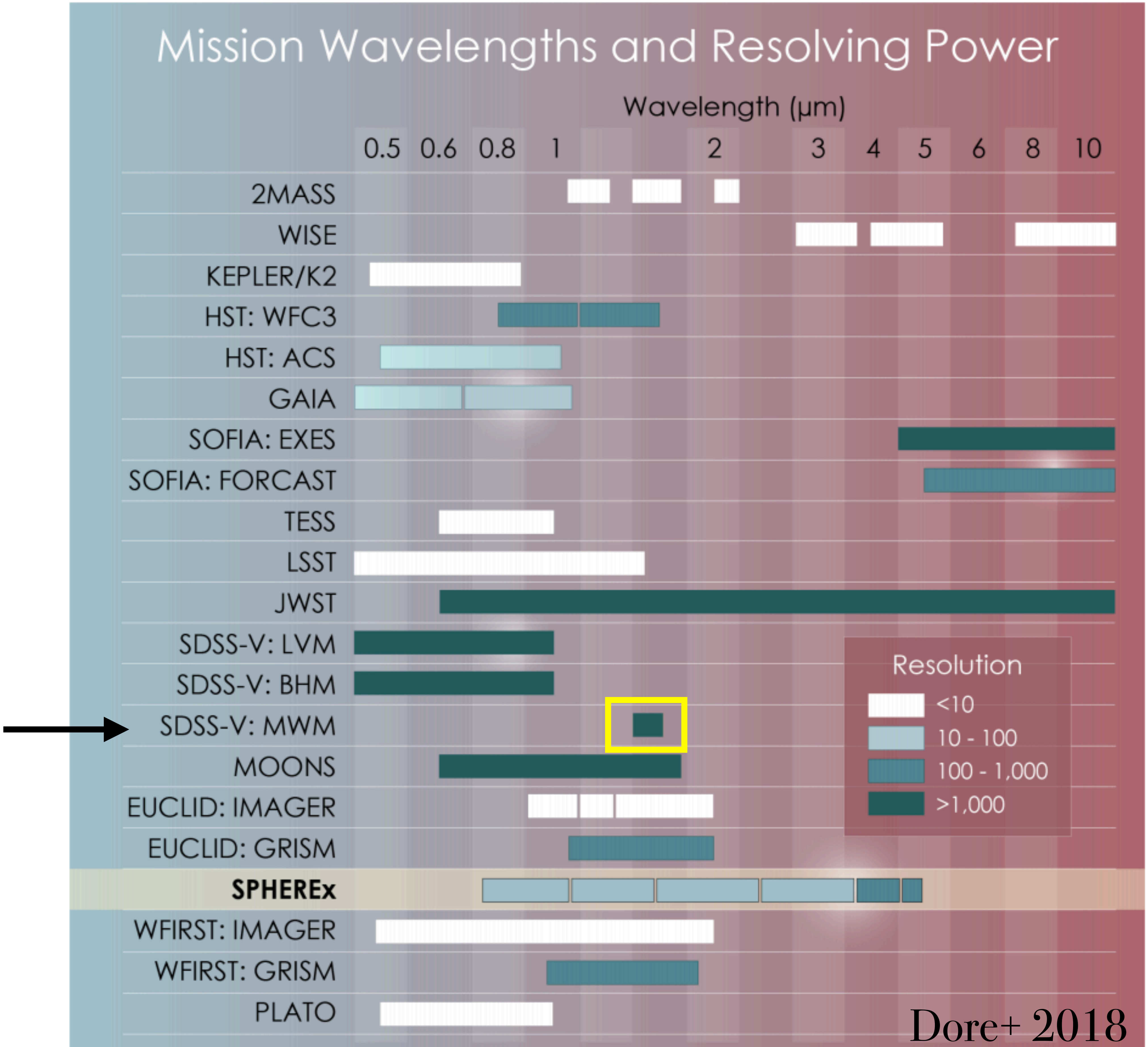
3 Science Programs

1) Milky Way Mapper (stars)
2) Black Hole Mapper
3) Local Volume Mapper (galaxies)

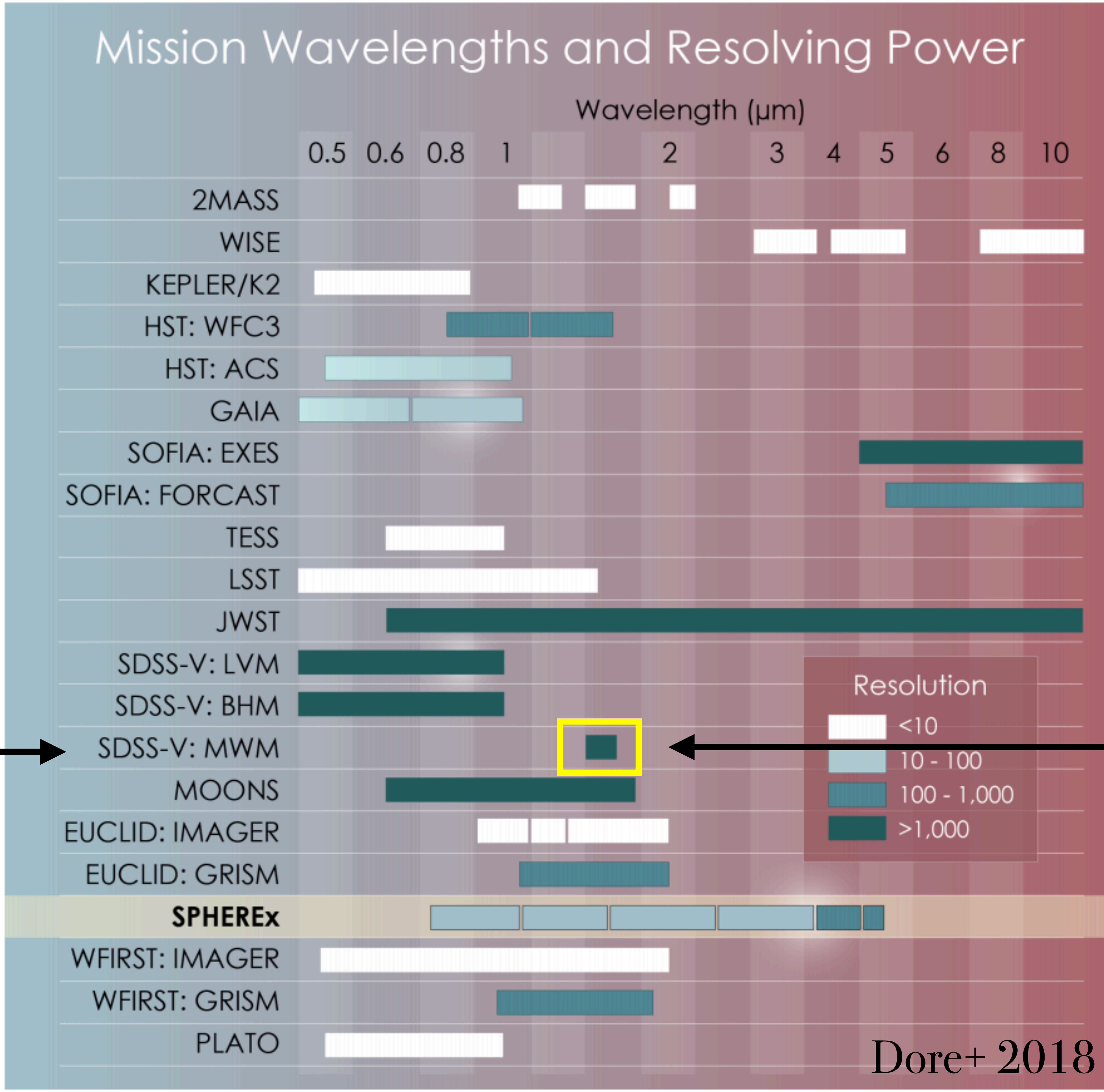
2 Infrastructure Investments

1) Fiber Positioning System
2) Local Volume Mapper

Kollmeier et al., 2017 – http://adsabs.harvard.edu/abs/2017arXiv171103234K



Mission Wavelengths and Resolving Power
Wavelength (μm)
0.5
0.6
0.8
1
2
3
4
5
6
8
10
2MASS
WISE
KEPLER/K2
HST: WFC3
HST: ACS
GAIA
SOFIA: EXES
SOFIA: FORCAST
TESS
LSST
JWST
SDSS-V: LVM
SDSS-V: BHM
SDSS-V: MWM
MOONS
EUCLID: IMAGER
EUCLID: GRISM
SPHEREx
WFIRST: IMAGER
WFIRST: GRISM
PLATO
Resolution
<10
10 - 100
100 - 1,000
>1,000
Dore+ 2018



R=22,500
1.5-1.7μm

R=1800
0.4-0.9μm

SDSS-V's Milky Way Mapper

- Target 5 million stars in the Milky Way with H < 11, G-H > 3.5
- Spans a spatial area of ~ 3000 deg^2

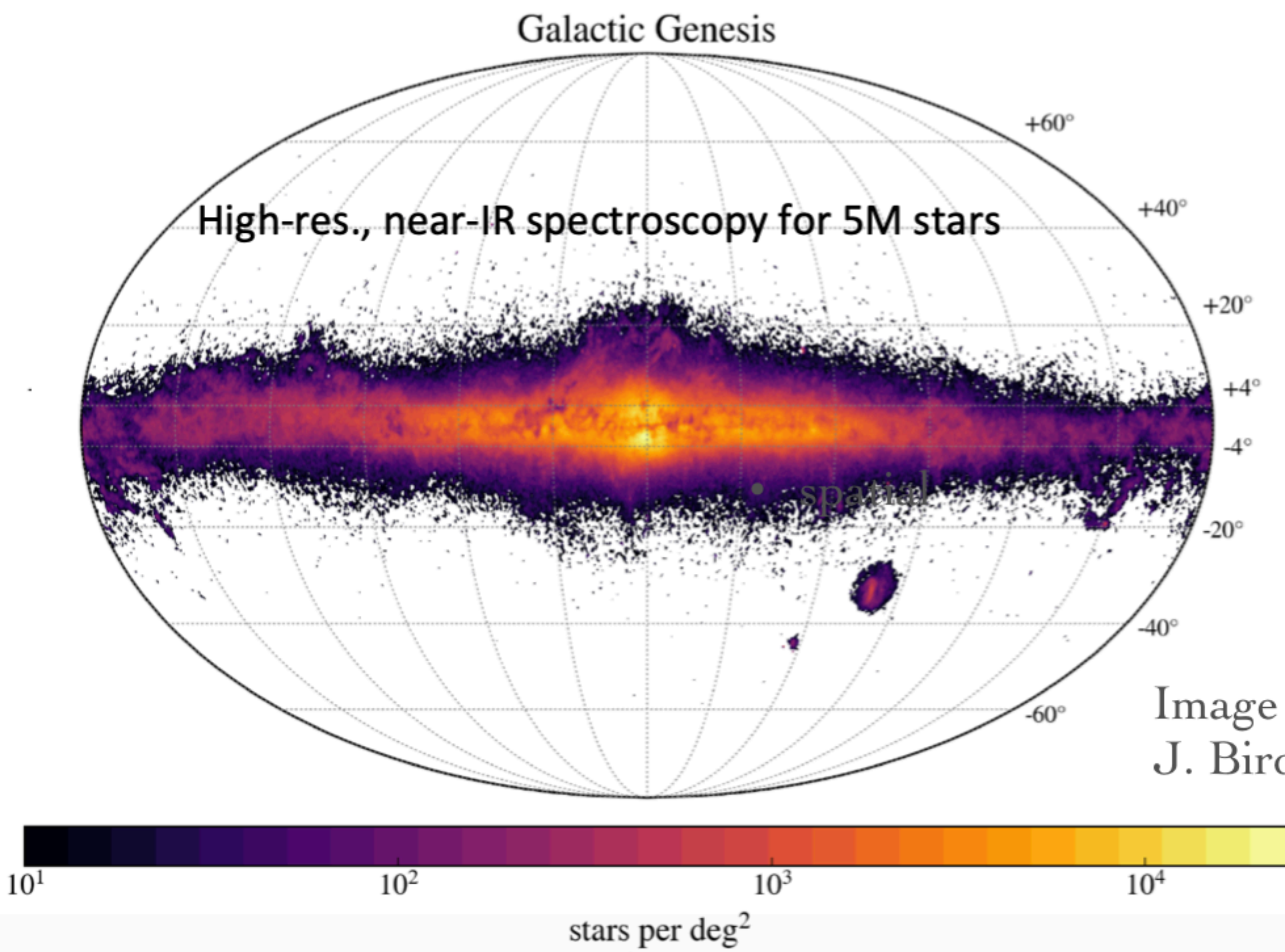


Image credit:
J. Bird (Vanderbilt)

SDSS-V's Milky Way Mapper

- Target 5 million stars in the Milky Way with H < 11, G-H > 3.5
- Spans a spatial area of ~ 3000 deg^2

<u>Spectra</u>
R=22,500 -APOGEE
(R=1800 - BOSS)

<u>Measure</u>
Teff, logg,[Fe/H],
25 x [X/Fe] (0.05dex)
mass (age) (< 40%)
distances (< 20%)

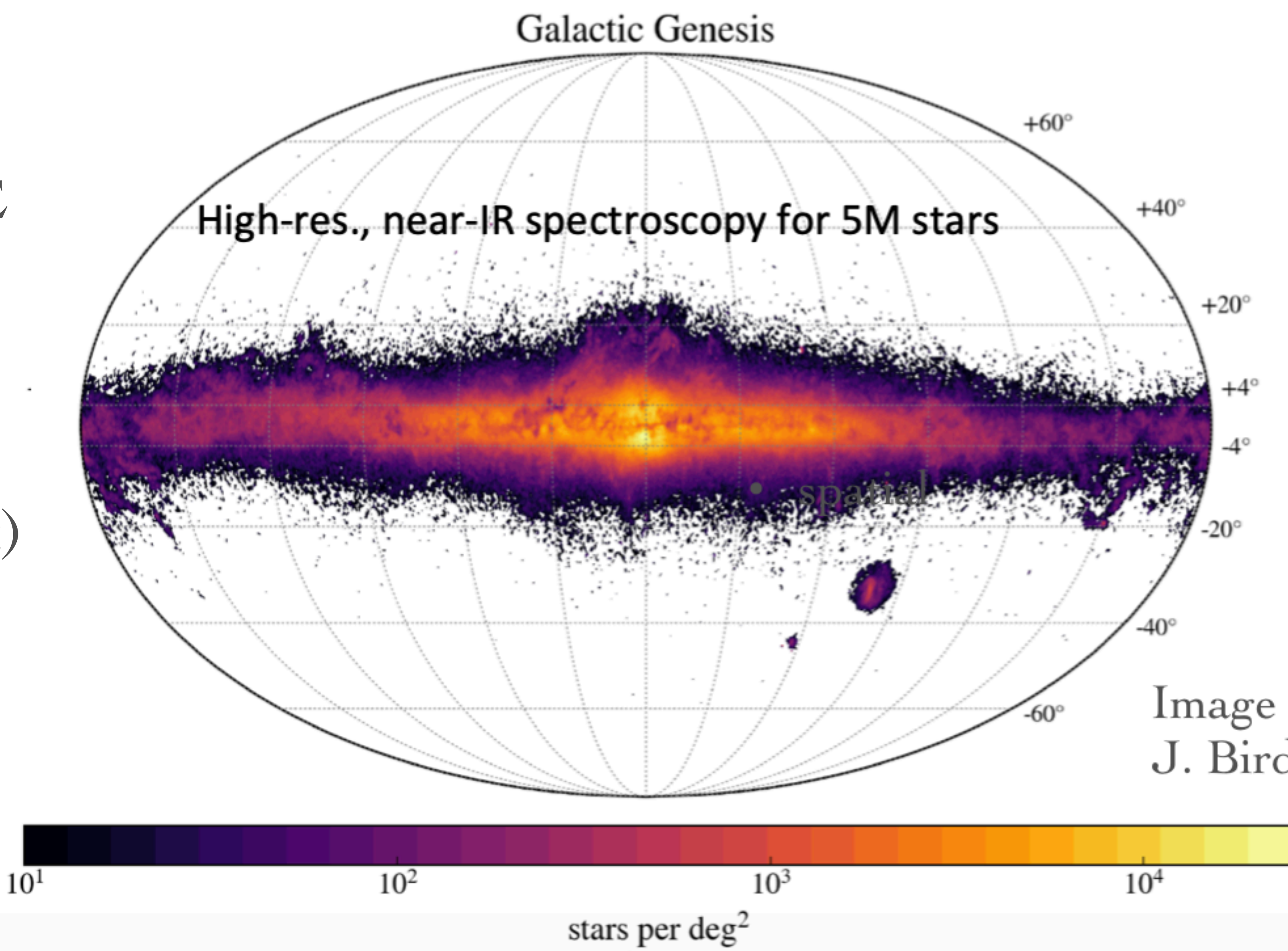


Image credit:
J. Bird (Vanderbilt)

Data-Driven Modeling of Stars

Approach to infer stellar parameters and individual abundances for stars in large surveys

[The Cannon (Ness+2015), The Payne (Ting+ 2017), AstroNN (Leung+ 2019)]

Data-Driven Modeling of Stars

Approach to infer stellar parameters and individual abundances for stars in large surveys

[The Cannon (Ness+2015), The Payne (Ting+ 2017), AstroNN (Leung+ 2019)]

- Relies on a subset of **reference** stars in survey - known **labels** (Teff, logg, [Fe/H]…)
- **Labels** - high resolution analyses, any λ
- Use reference objects to **build a model** — label full survey data

Data-Driven Modeling of Stars

Approach to infer stellar parameters and individual abundances for stars in large surveys

[The Cannon (Ness+2015), The Payne (Ting+ 2017), AstroNN (Leung+ 2019)]

- Relies on a subset of **reference** stars in survey - known **labels** (Teff, logg, [Fe/H]…)
- **Labels** - high resolution analyses, any λ
- Use reference objects to **build a model** — label full survey data

- **Propagate labels** - one survey to another
  - different surveys directly on the same scale,
  - good data to label worse data (high resolution labels to low resolution survey),
  - discover information (derive mass (age) from C,N features in spectra)

Data-Driven Modeling of Stars

Approach to infer stellar parameters and individual abundances for stars in large surveys

[The Cannon (Ness+2015), The Payne (Ting+ 2017), AstroNN (Leung+ 2019)]

- Relies on a subset of **reference** stars in survey - known **labels** (Teff, logg, [Fe/H]…)
- **Labels** - high resolution analyses, any λ
- Use reference objects to **build a model** — label full survey data

- **Propagate labels** - one survey to another
  - different surveys directly on the same scale,
  - good data to label worse data (high resolution labels to low resolution survey),
  - discover information (derive mass (age) from C,N features in spectra)

- **Precise** (2-3 times higher precision than previous approaches)
- **Computationally fast**
- Understand **where/how** information distributed

Data-Driven Modeling

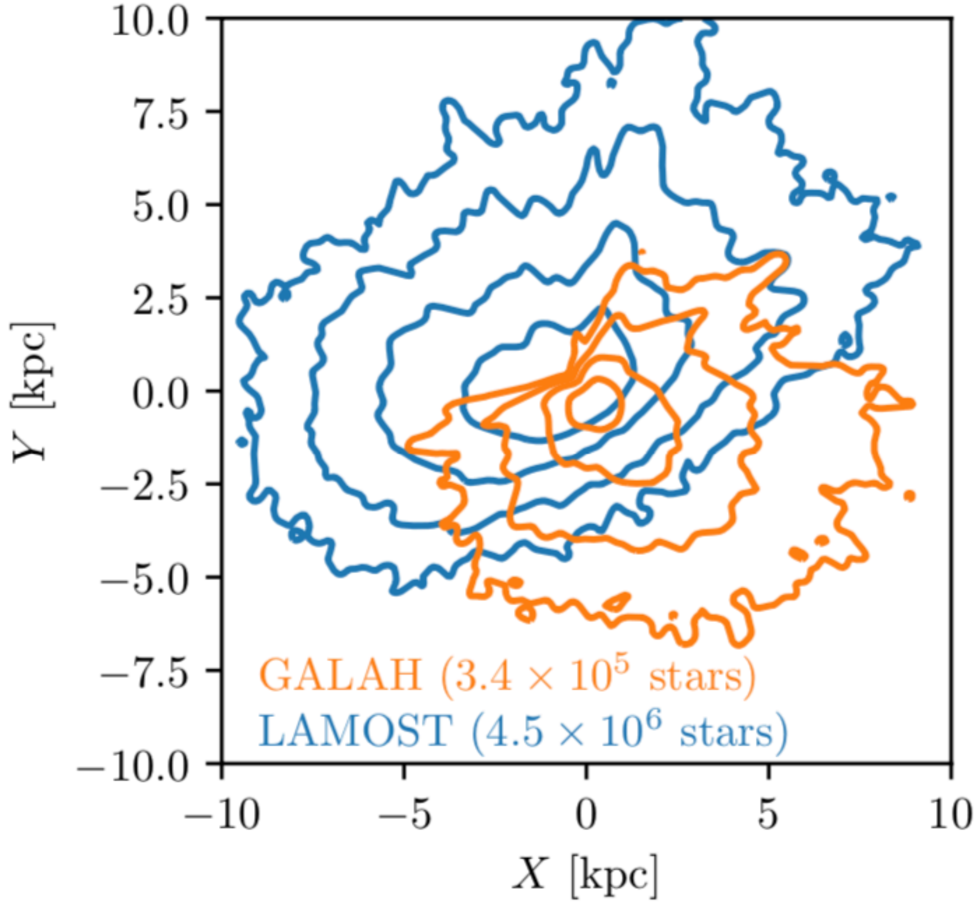
Approach to “betterize" spectra in large surveys — de-noise and in-paint missing regions

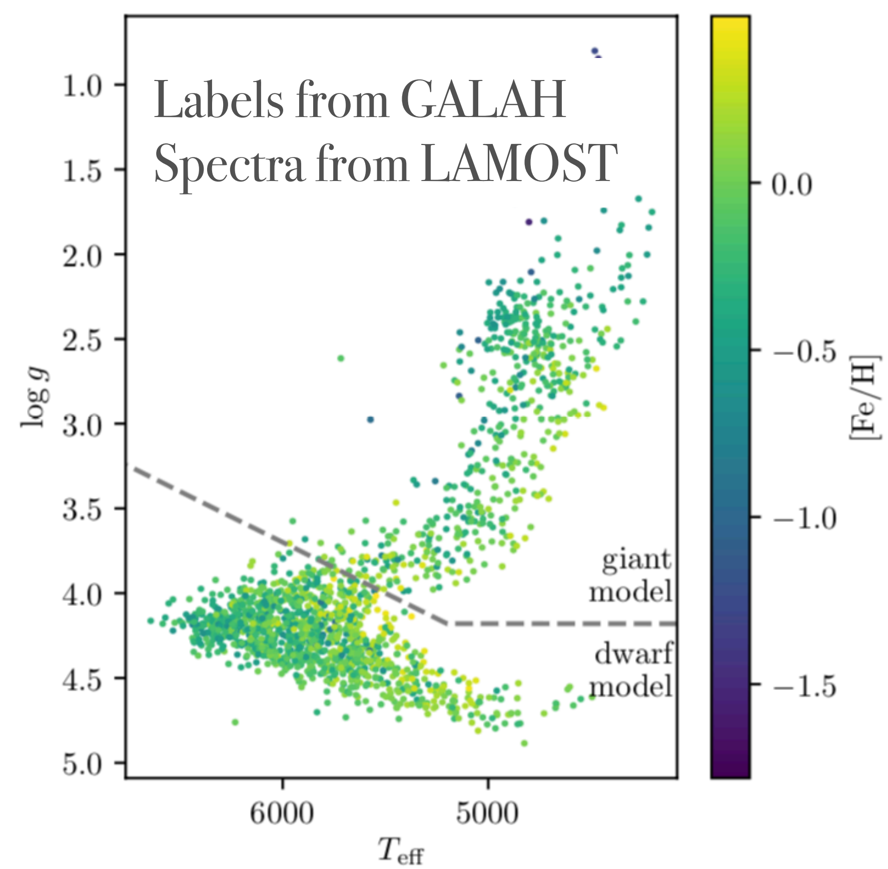
SSSpaNG! (Feeney, Wandelt, Ness 2020)

Spectra as data-driven non-Gaussian processes

The Cannon (for label transfer)

Determined [X/Fe] labels for LAMOST (R=1800) – using stars in common with GALAH (R=28,000)



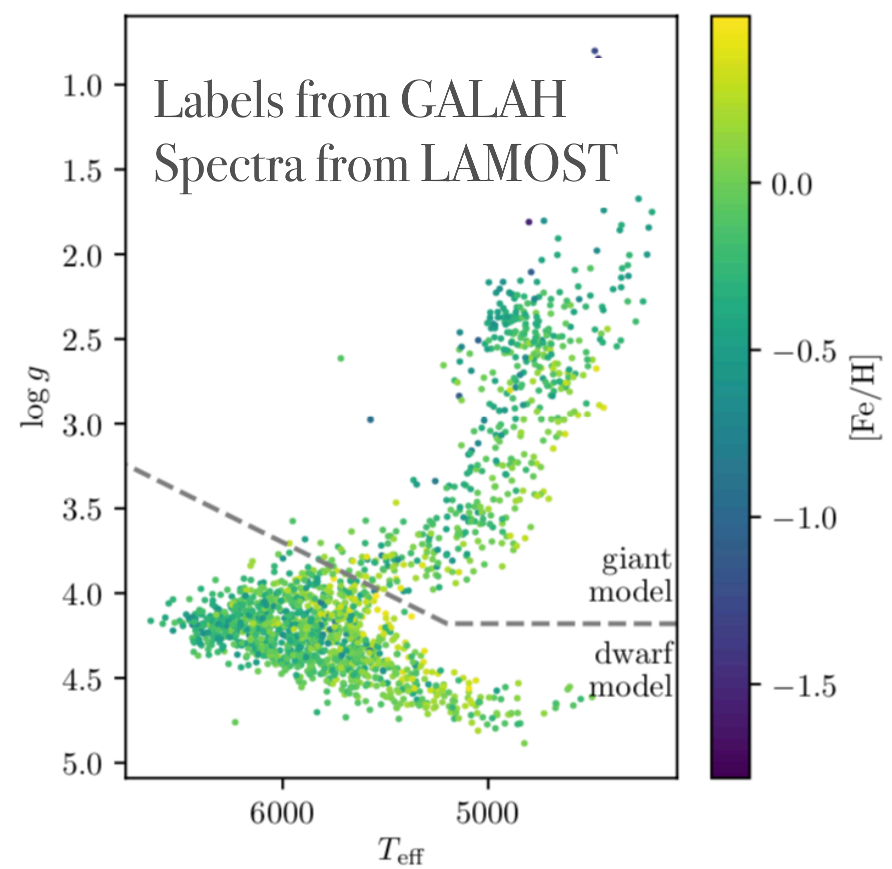


Adam Wheeler, Ness+ 2020

The Cannon (for label transfer)

Determined [X/Fe] labels for LAMOST (R=1800) – using stars in common with GALAH (R=28,000)

1. Uses n reference objects with known labels l to build a model *Training*



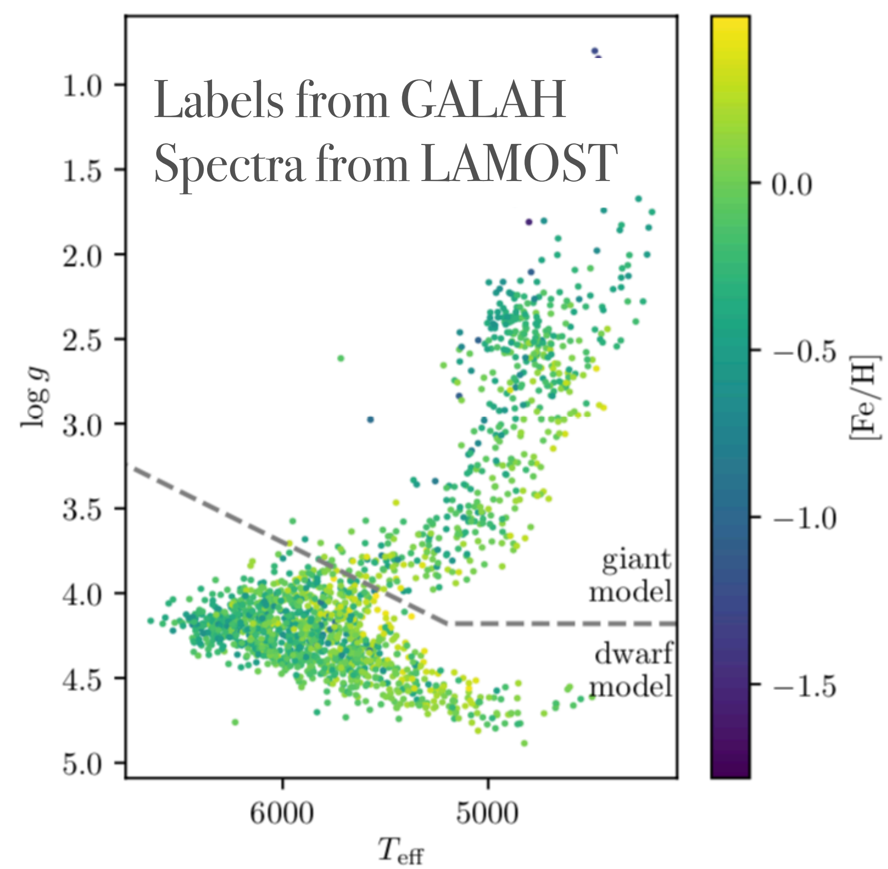
Adam Wheeler, Ness+ 2020

The Cannon (for label transfer)

Determined [X/Fe] labels for LAMOST (R=1800) – using stars in common with GALAH (R=28,000)

Adam Wheeler, Ness+ 2020

1. Uses n reference objects with known labels l to build a model *Training*



$$f_{n\lambda} = g(l_n \mid \theta_\lambda) + \text{noise}$$

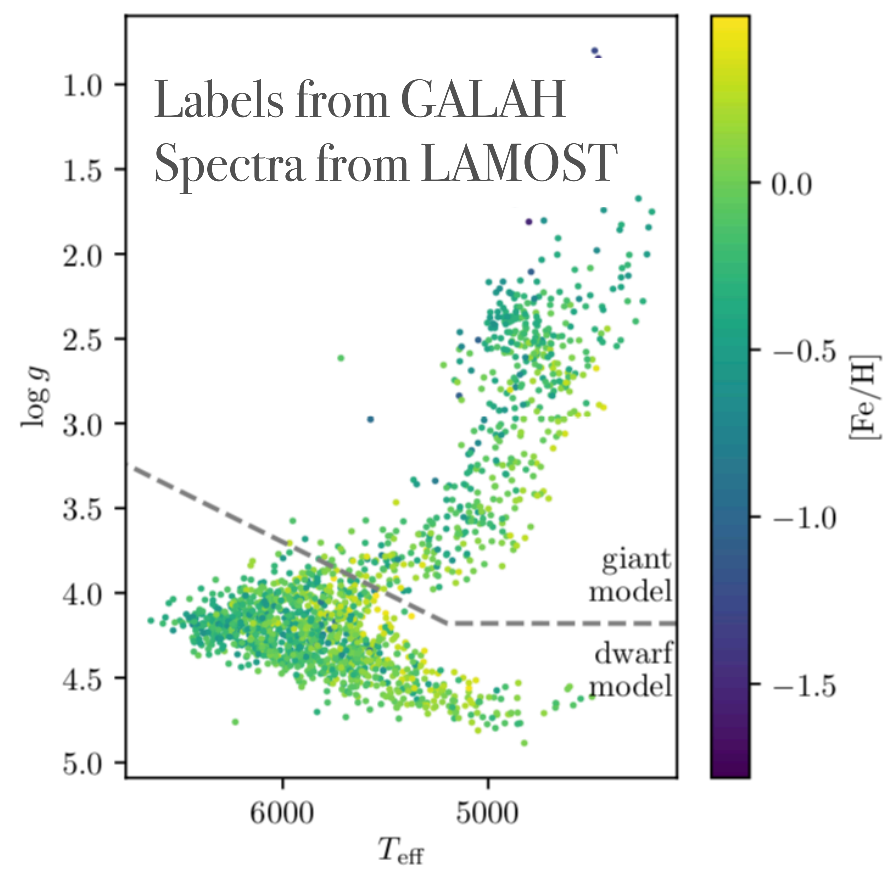
The Cannon (for label transfer)

Determined [X/Fe] labels for LAMOST (R=1800) – using stars in common with GALAH (R=28,000)

1. Uses n reference objects with known labels l to build a model *Training*

Teff, logg, [Fe/H]
[X/Fe]

$$f_{n\lambda} = g(l_n \mid \theta_\lambda) + \text{noise}$$



Adam Wheeler, Ness+ 2020

The Cannon (for label transfer)

Determined [X/Fe] labels for LAMOST (R=1800) – using stars in common with GALAH (R=28,000)

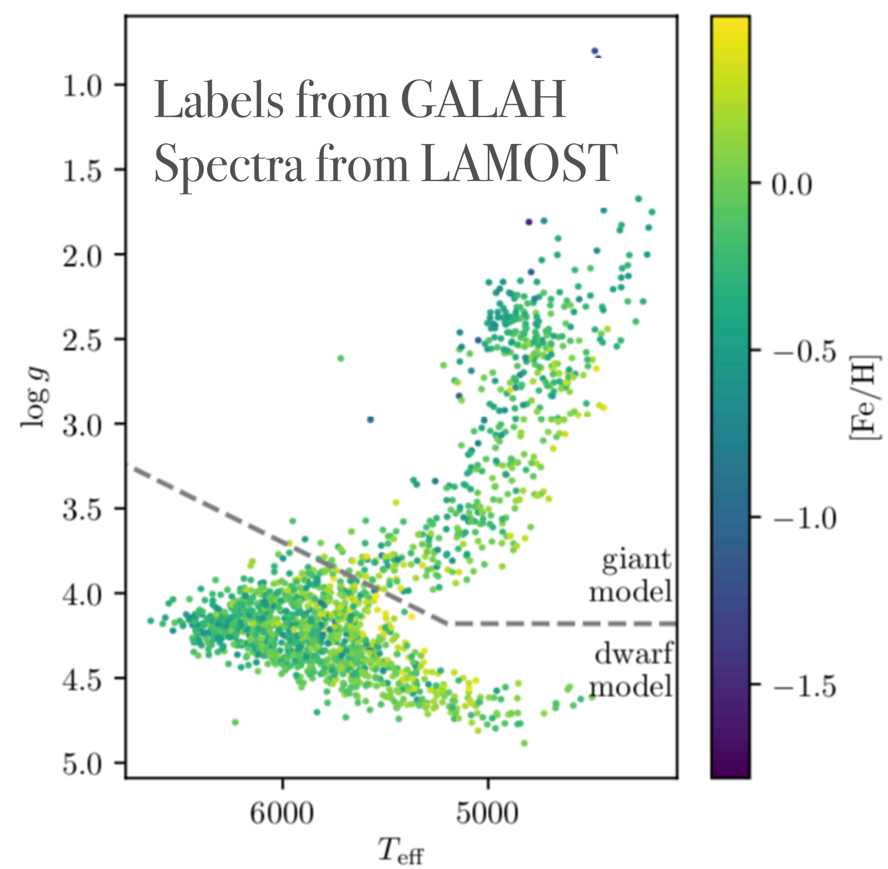
Adam Wheeler, Ness+ 2020

1. Uses n reference objects with known labels l to build a model *Training*

Teff, logg, [Fe/H]
[X/Fe]

$$f_{n\lambda} = g(l_n \mid \theta_\lambda) + \text{noise}$$

spectral model



The Cannon (for label transfer)

Determined [X/Fe] labels for LAMOST (R=1800) – using stars in common with GALAH (R=28,000)

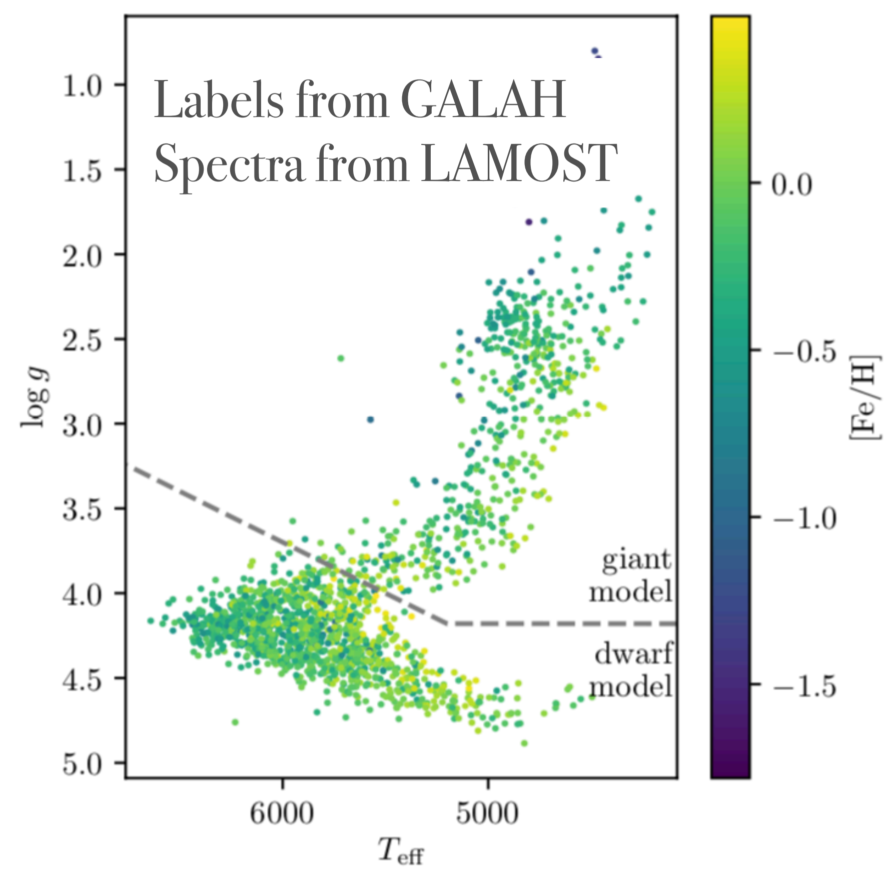
1. Uses n reference objects with known labels l to build a model *Training*

Teff, logg, [Fe/H] [X/Fe]

photon noise + fit of spectral model

$$f_{n\lambda} = g(l_n \mid \theta_\lambda) + \text{noise}$$

spectral model



Adam Wheeler, Ness+ 2020

The Cannon (for label transfer)

Determined [X/Fe] labels for LAMOST (R=1800) – using stars in common with GALAH (R=28,000)

Adam Wheeler, Ness+ 2020

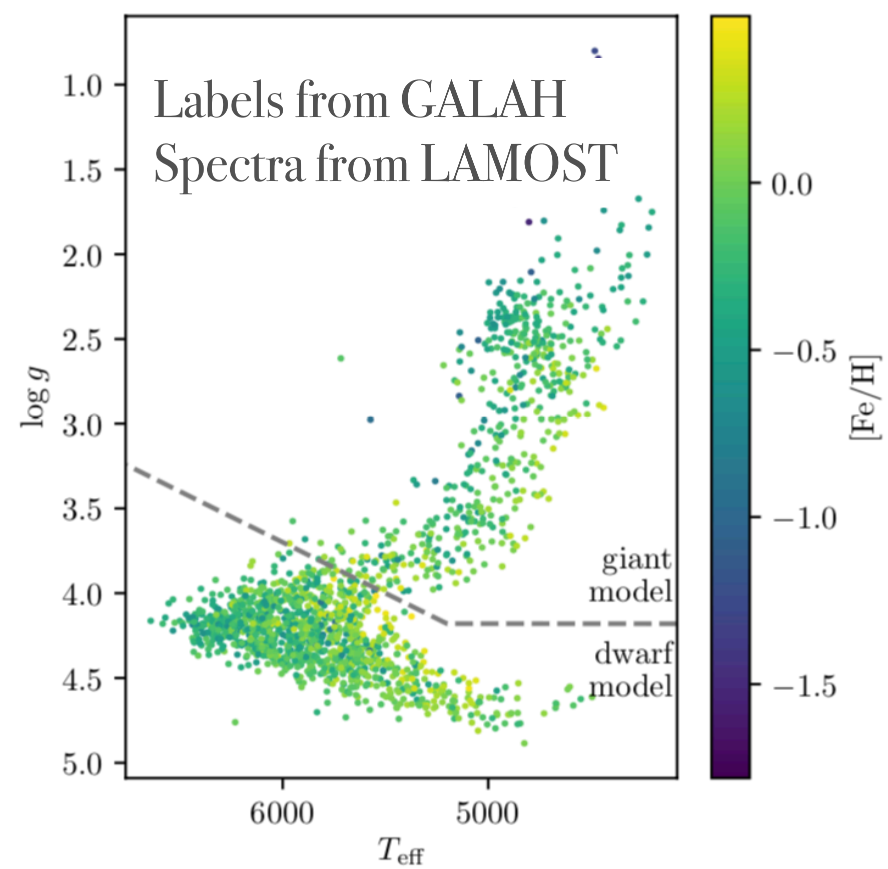
1. Uses n reference objects with known labels l to build a model *Training*

Teff, logg, [Fe/H] [X/Fe]

photon noise + fit of spectral model

$$f_{n\lambda} = g(l_n \mid \theta_\lambda) + \text{noise}$$

spectral model



2. Relates l to stellar flux f, at each wavelength λ.

The Cannon (for label transfer)

Determined [X/Fe] labels for LAMOST (R=1800) – using stars in common with GALAH (R=28,000)

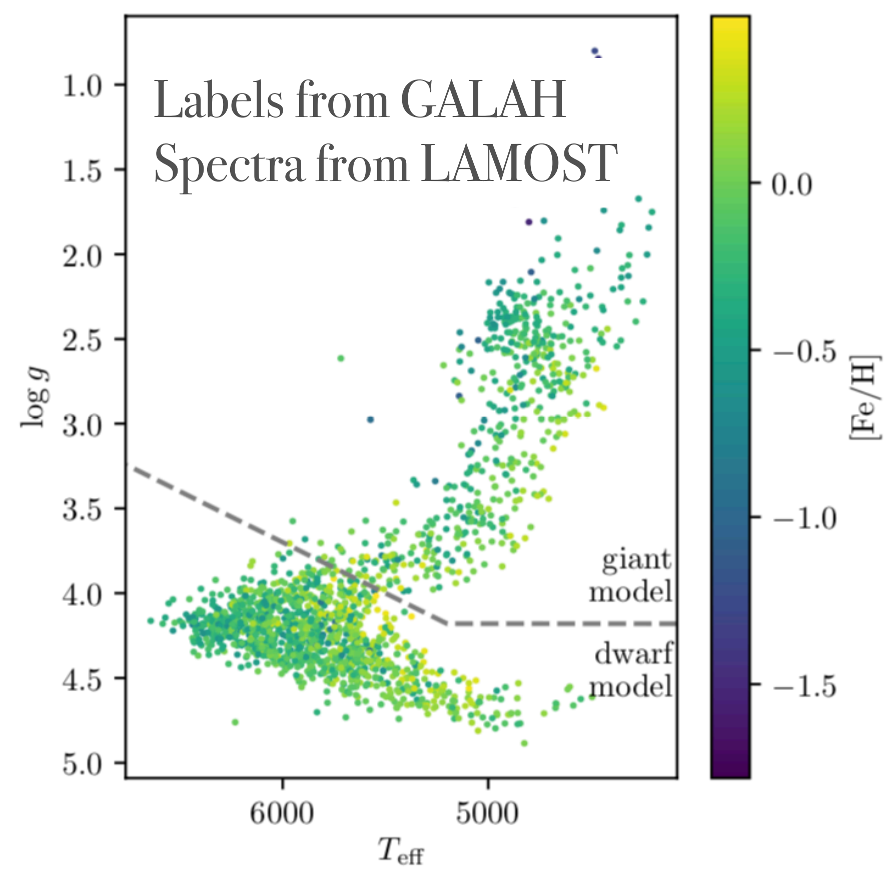
1. Uses **n** reference objects with known labels l to build a model *Training*

Teff, logg, [Fe/H] [X/Fe]

photon noise + fit of spectral model

$$f_{n\lambda} = g(l_n \mid \theta_\lambda) + \text{noise}$$

spectral model



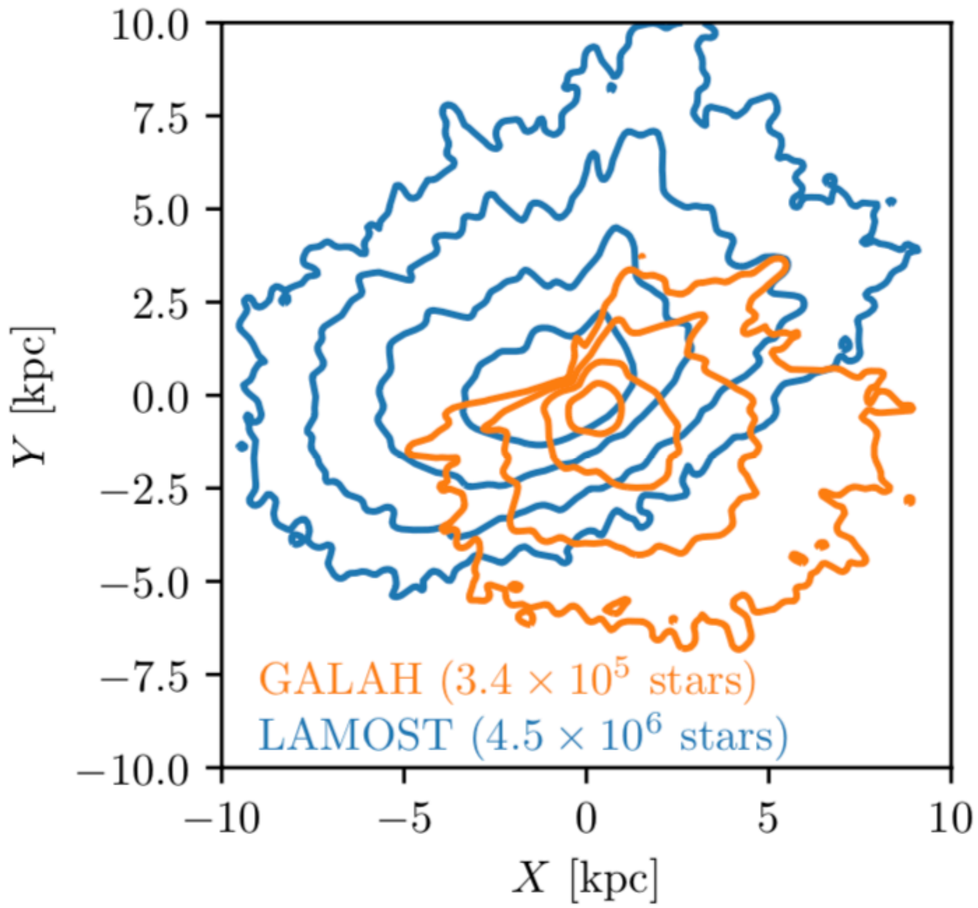
Adam Wheeler, Ness+ 2020

2. Relates l to stellar flux **f**, at each wavelength **λ**.

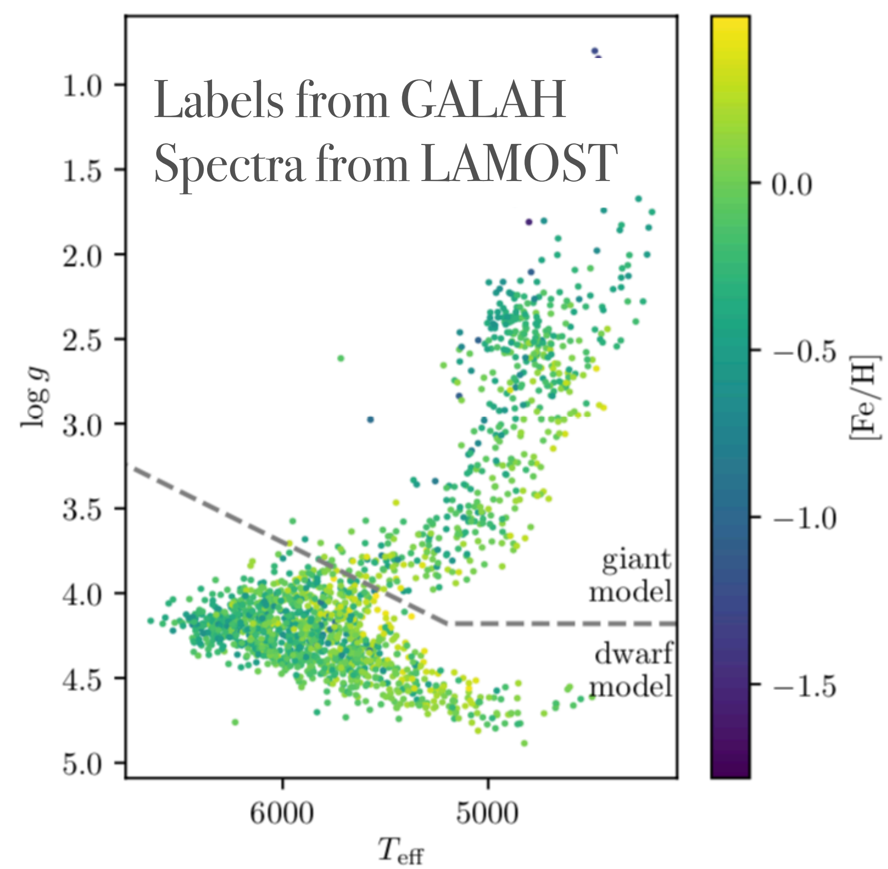
3. That model is then used to infer the stellar labels for the remaining stars in the survey *Test*

The Cannon (for label transfer)

Determined [X/Fe] labels for LAMOST (R=1800) – using stars in common with GALAH (R=28,000)



ɔdel



Adam Wheeler, Ness+ 2020

2. Relates l to stellar flux f, at each wavelength λ.

3. That model is then used to infer the stellar labels for the remaining stars in the survey *Test*

A simple polynomial model of the labels

$$f_{n\lambda} = g(l_n \,|\, \theta_\lambda) + \text{noise}$$

A simple polynomial model of the labels

$$f_{n\lambda} = g(l_n \,|\, \theta_\lambda) + \text{noise}$$

$$\boldsymbol{\ell}_n = \langle T_{\rm eff}, \log(g), v_{\rm mic}, [{\rm Fe/H}], [{\rm X}_1/{\rm Fe}], \ldots, [{\rm X}_N/{\rm Fe}] \rangle$$

A simple polynomial model of the labels

GALAH labels

$$f_{n\lambda} = g(\ell_n \,|\, \theta_\lambda) + \text{noise}$$

$$\boldsymbol{\ell}_n = \langle T_{\rm eff}, \log(g), v_{\rm mic}, [{\rm Fe/H}], [{\rm X}_1/{\rm Fe}], \ldots, [{\rm X}_N/{\rm Fe}] \rangle$$

A simple polynomial model of the labels

GALAH labels

$f_{n\lambda} = g(l_n \mid \theta_\lambda) + \text{noise}$

$$\ell_n = \langle T_{\text{eff}}, \log(g), v_{\text{mic}}, [\text{Fe/H}], [\text{X}_1/\text{Fe}], \ldots, [\text{X}_N/\text{Fe}] \rangle$$

$$\begin{aligned}
F_{n\lambda} = \theta_\lambda^0 & \quad \text{(constant term)} \\
+ \theta_\lambda^{T_{\text{eff}}} T_{\text{eff}} + \cdots + \theta_\lambda^{X_N} [X_N/\text{Fe}] & \quad \text{(linear terms)} \\
+ \theta_\lambda^{T_{\text{eff}}^2} T_{\text{eff}}^2 + \cdots + \theta_\lambda^{X_N^2} ([X_N/\text{Fe}])^2 & \quad \text{(squared terms)} \\
+ \theta_\lambda^{T_{\text{eff}} \log(g)} T_{\text{eff}} \log(g) + \ldots & \\
+ \theta_\lambda^{X_N X_{N-1}} [X_N/\text{Fe}][X_{N-1}/\text{Fe}] & \quad \text{(cross-terms)} \\
+ \text{error.} &
\end{aligned}$$

Wheeler,
Ness+ 2020

A simple polynomial model of the labels

GALAH labels

$$f_{n\lambda} = g(l_n \,|\, \theta_\lambda) + \text{noise}$$

$$\ell_n = \langle T_{\rm eff}, \log(g), v_{\rm mic}, [\mathrm{Fe/H}], [\mathrm{X_1/Fe}], \ldots, [\mathrm{X}_N/\mathrm{Fe}] \rangle$$

LAMOST Flux

$$
\begin{aligned}
F_{n\lambda} = {} & \theta_\lambda^0 && \text{(constant term)} \\
& + \theta_\lambda^{T_{\rm eff}} T_{\rm eff} + \cdots + \theta_\lambda^{X_N} [X_N/\mathrm{Fe}] && \text{(linear terms)} \\
& + \theta_\lambda^{T_{\rm eff}^2} T_{\rm eff}^2 + \cdots + \theta_\lambda^{X_N^2} ([X_N/\mathrm{Fe}])^2 && \text{(squared terms)} \\
& + \theta_\lambda^{T_{\rm eff} \log(g)} T_{\rm eff} \log(g) + \ldots \\
& \quad + \theta_\lambda^{X_N X_{N-1}} [X_N/\mathrm{Fe}][X_{N-1}/\mathrm{Fe}] && \text{(cross-terms)} \\
& + \text{error.}
\end{aligned}
$$

Wheeler,
Ness+ 2020

A simple polynomial model of the labels

GALAH labels

LAMOST Flux

$$f_{n\lambda} = g(l_n \,|\, \theta_\lambda) + \text{noise}$$

$$\ell_n = \langle T_{\rm eff}, \log(g), v_{\rm mic}, [\mathrm{Fe/H}], [\mathrm{X}_1/\mathrm{Fe}], \ldots, [\mathrm{X}_N/\mathrm{Fe}] \rangle$$

$$\begin{aligned}
F_{n\lambda} = {} & \theta_\lambda^0 && \text{(constant term)} \\
& + \theta_\lambda^{T_{\rm eff}} T_{\rm eff} + \cdots + \theta_\lambda^{X_N} [X_N/\mathrm{Fe}] && \text{(linear terms)} \\
& + \theta_\lambda^{T_{\rm eff}^2} T_{\rm eff}^2 + \cdots + \theta_\lambda^{X_N^2} ([X_N/\mathrm{Fe}])^2 && \text{(squared terms)} \\
& + \theta_\lambda^{T_{\rm eff} \log(g)} T_{\rm eff} \log(g) + \ldots \\
& \quad + \theta_\lambda^{X_N X_{N-1}} [X_N/\mathrm{Fe}][X_{N-1}/\mathrm{Fe}] && \text{(cross-terms)} \\
& + \text{error.}
\end{aligned}$$

Wheeler,
Ness+ 2020

A simple polynomial model of the labels

GALAH labels

LAMOST Flux

$$f_{n\lambda} = g(\ell_n \mid \theta_\lambda) + \text{noise}$$

$$\ell_n = \langle T_{\text{eff}}, \log(g), v_{\text{mic}}, [\text{Fe/H}], [\text{X}_1/\text{Fe}], \ldots, [\text{X}_N/\text{Fe}] \rangle$$

$$\begin{aligned}
F_{n\lambda} = \theta_\lambda^0 & \qquad \text{(constant term)} \\
+ \theta_\lambda^{T_{\text{eff}}} T_{\text{eff}} + \cdots + \theta_\lambda^{X_N} [X_N/\text{Fe}] & \qquad \text{(linear terms)} \\
+ \theta_\lambda^{T_{\text{eff}}^2} T_{\text{eff}}^2 + \cdots + \theta_\lambda^{X_N^2} ([X_N/\text{Fe}])^2 & \qquad \text{(squared terms)} \\
+ \theta_\lambda^{T_{\text{eff}} \log(g)} T_{\text{eff}} \log(g) + \ldots & \\
+ \theta_\lambda^{X_N X_{N-1}} [X_N/\text{Fe}][X_{N-1}/\text{Fe}] & \qquad \text{(cross-terms)} \\
+ \text{error}. &
\end{aligned}$$

Wheeler,
Ness+ 2020

A simple polynomial model of the labels

GALAH labels

$$f_{n\lambda} = g(l_n \,|\, \theta_\lambda) + \text{noise}$$

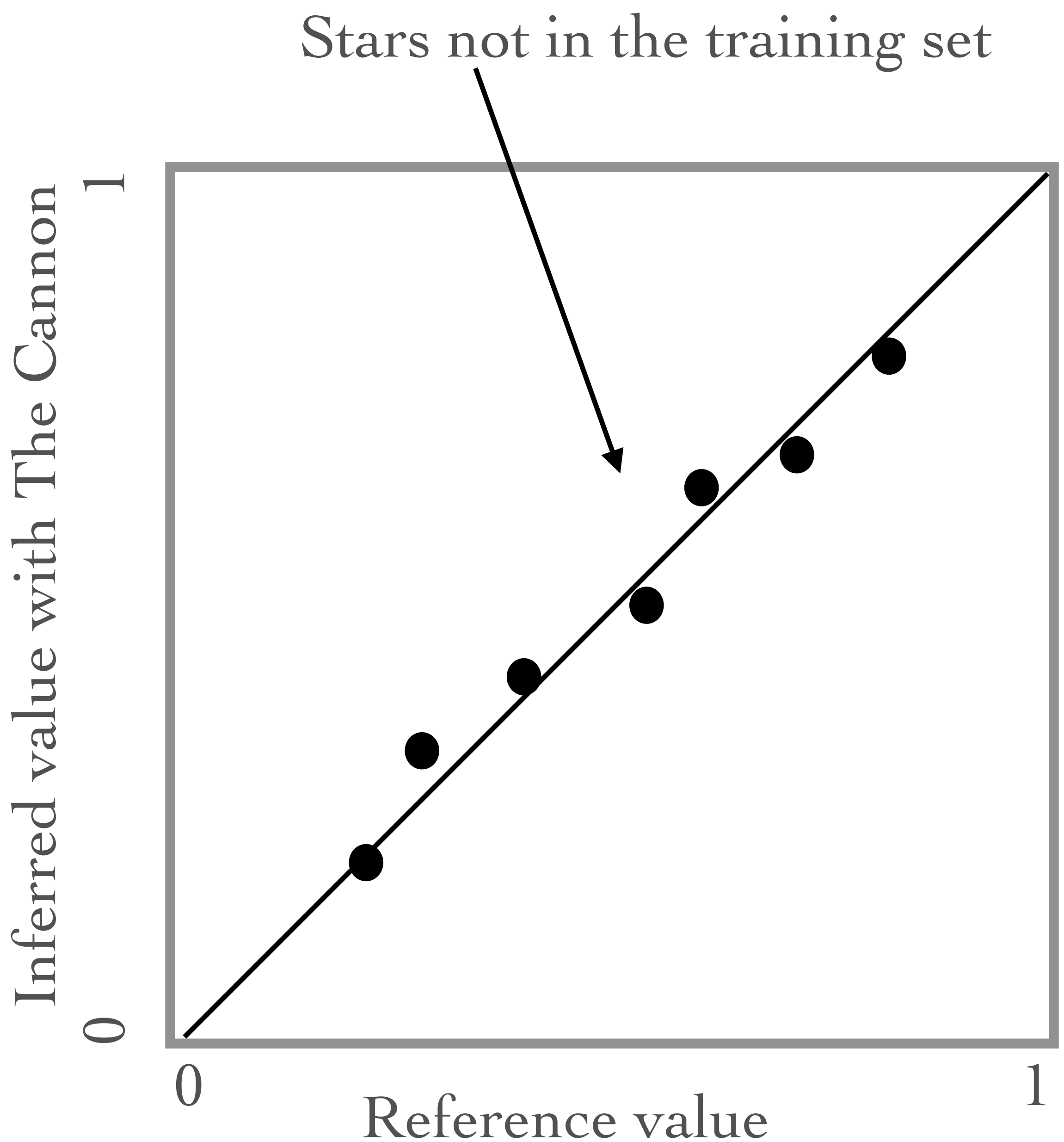
$$\ell_n = \langle T_{\rm eff}, \log(g), v_{\rm mic}, [{\rm Fe/H}], [{\rm X}_1/{\rm Fe}], \ldots, [{\rm X}_N/{\rm Fe}] \rangle$$

LAMOST Flux

$$\begin{aligned}
F_{n\lambda} = {} & \theta_\lambda^0 && \text{(constant term)} \\
& + \theta_\lambda^{T_{\rm eff}} T_{\rm eff} + \cdots + \theta_\lambda^{X_N} [X_N/{\rm Fe}] && \text{(linear terms)} \\
& + \theta_\lambda^{T_{\rm eff}^2} T_{\rm eff}^2 + \cdots + \theta_\lambda^{X_N^2} ([X_N/{\rm Fe}])^2 && \text{(squared terms)} \\
& + \theta_\lambda^{T_{\rm eff} \log(g)} T_{\rm eff} \log(g) + \ldots \\
& \quad + \theta_\lambda^{X_N X_{N-1}} [X_N/{\rm Fe}][X_{N-1}/{\rm Fe}] && \text{(cross-terms)} \\
& + \text{error.}
\end{aligned}$$

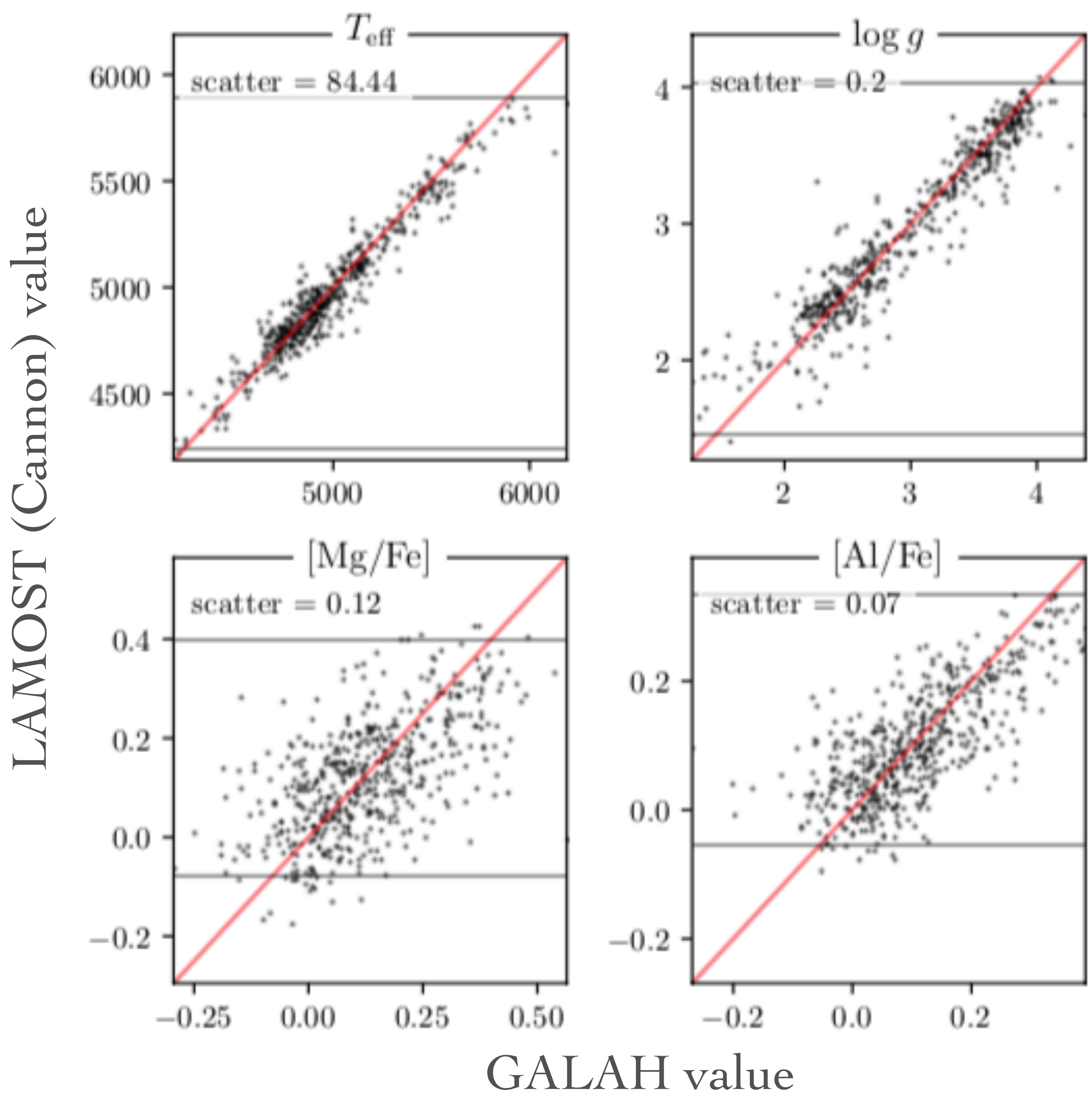
Wheeler, Ness+ 2020

Validation



Adam
Wheeler,
Ness+ 2020

Validation



Adam Wheeler, Ness+ 2020

Data-Driven Modeling with SPHEREx

Data-Driven Modeling with SPHEREx

- Propagate labels between SPHEREx and Sloan V data

Data-Driven Modeling with SPHEREx

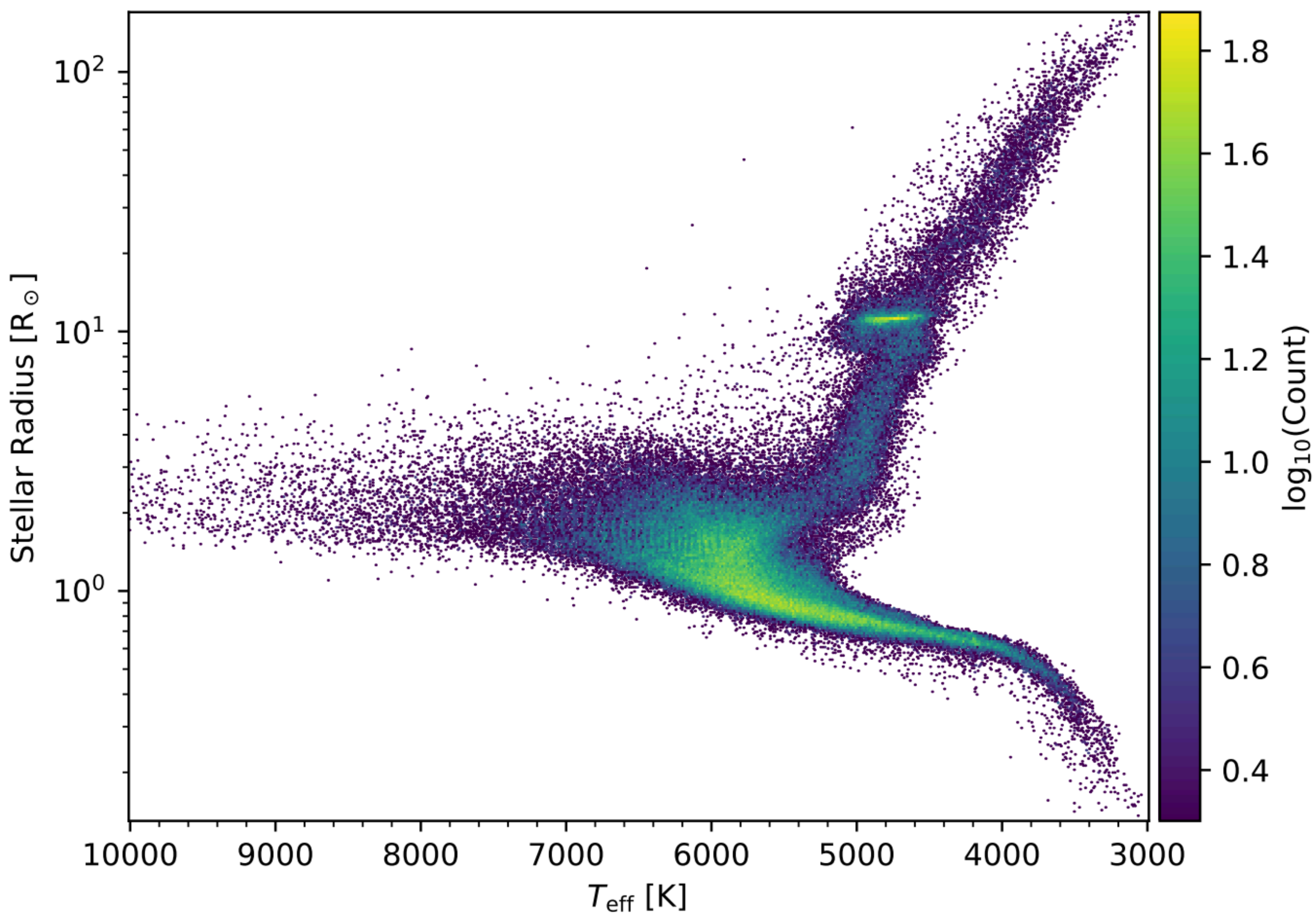
- Propagate labels between SPHEREx and Sloan V data
- What labels can we infer from SPHEREx spectra and to what precision

Data-Driven Modeling with SPHEREx

- Propagate labels between SPHEREx and Sloan V data
- What labels can we infer from SPHEREx spectra and to what precision
- What does stellar SPHEREx data look like?

Data-Driven Modeling with SPHEREx

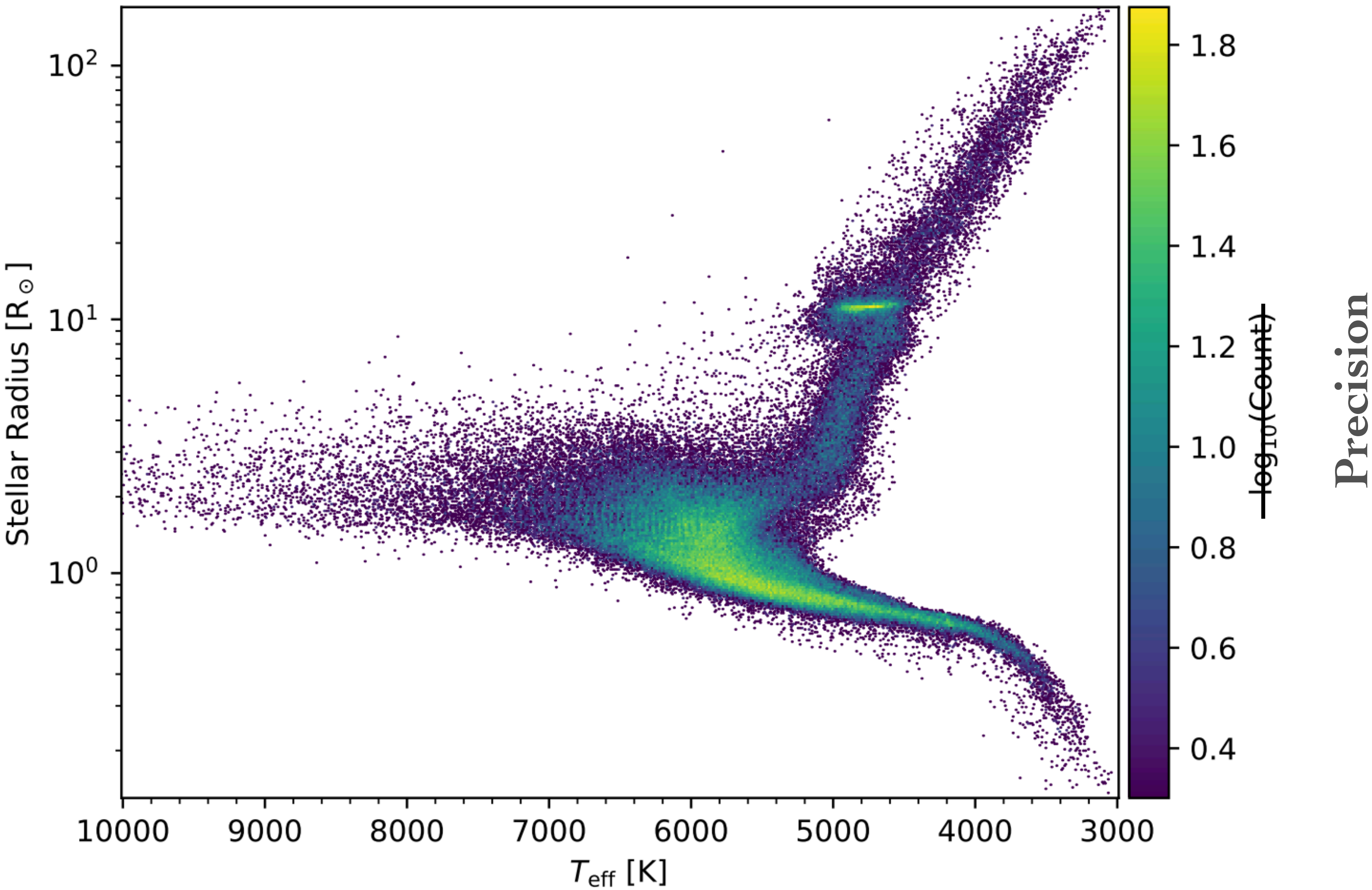
- Propagate labels between SPHEREx and Sloan V data
- What labels can we infer from SPHEREx spectra and to what precision
- What does stellar SPHEREx data look like?



Berger+ 2019

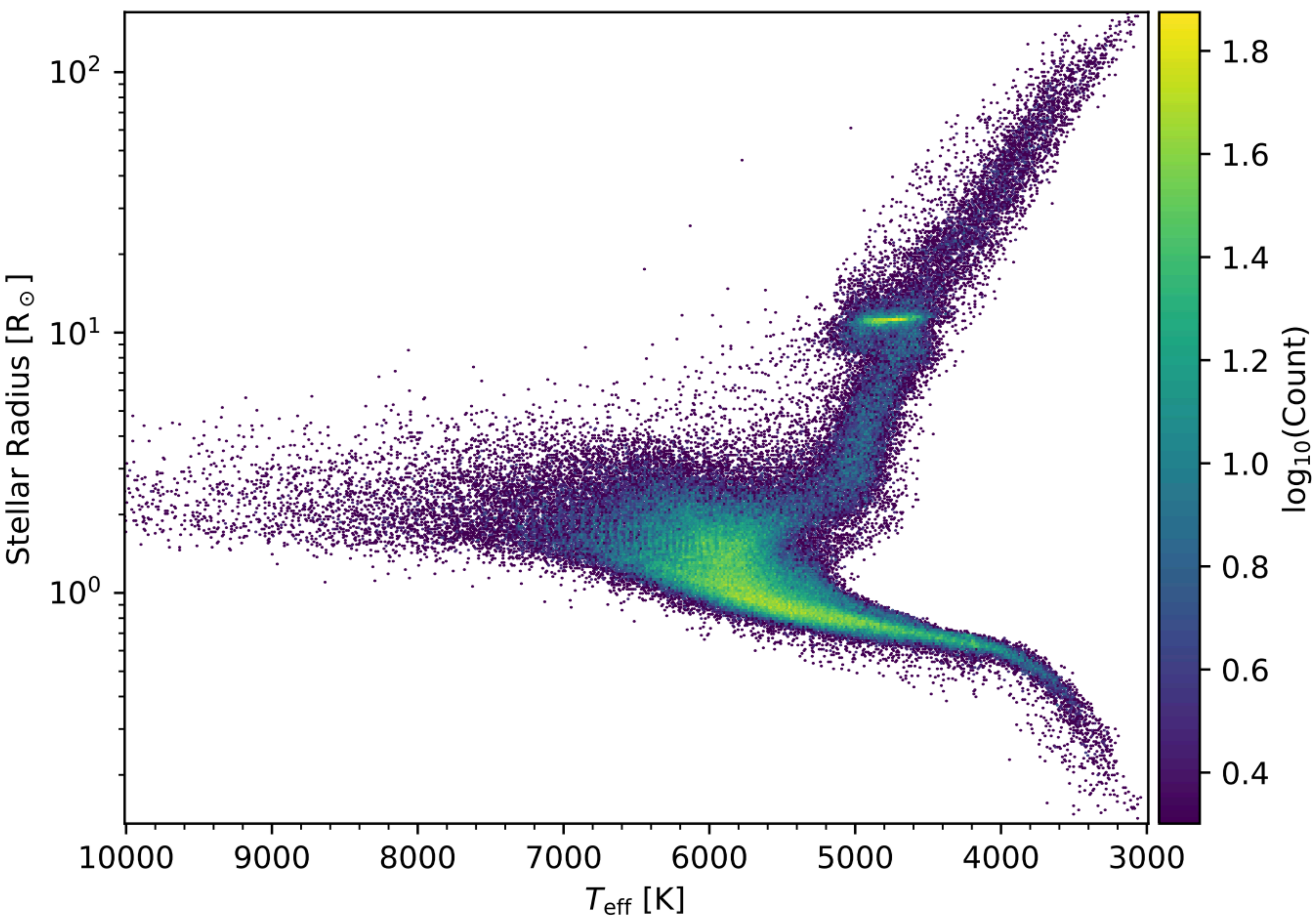
Data-Driven Modeling with SPHEREx

- Propagate labels between SPHEREx and Sloan V data
- What labels can we infer from SPHEREx spectra and to what precision
- What does stellar SPHEREx data look like?



Berger+ 2019

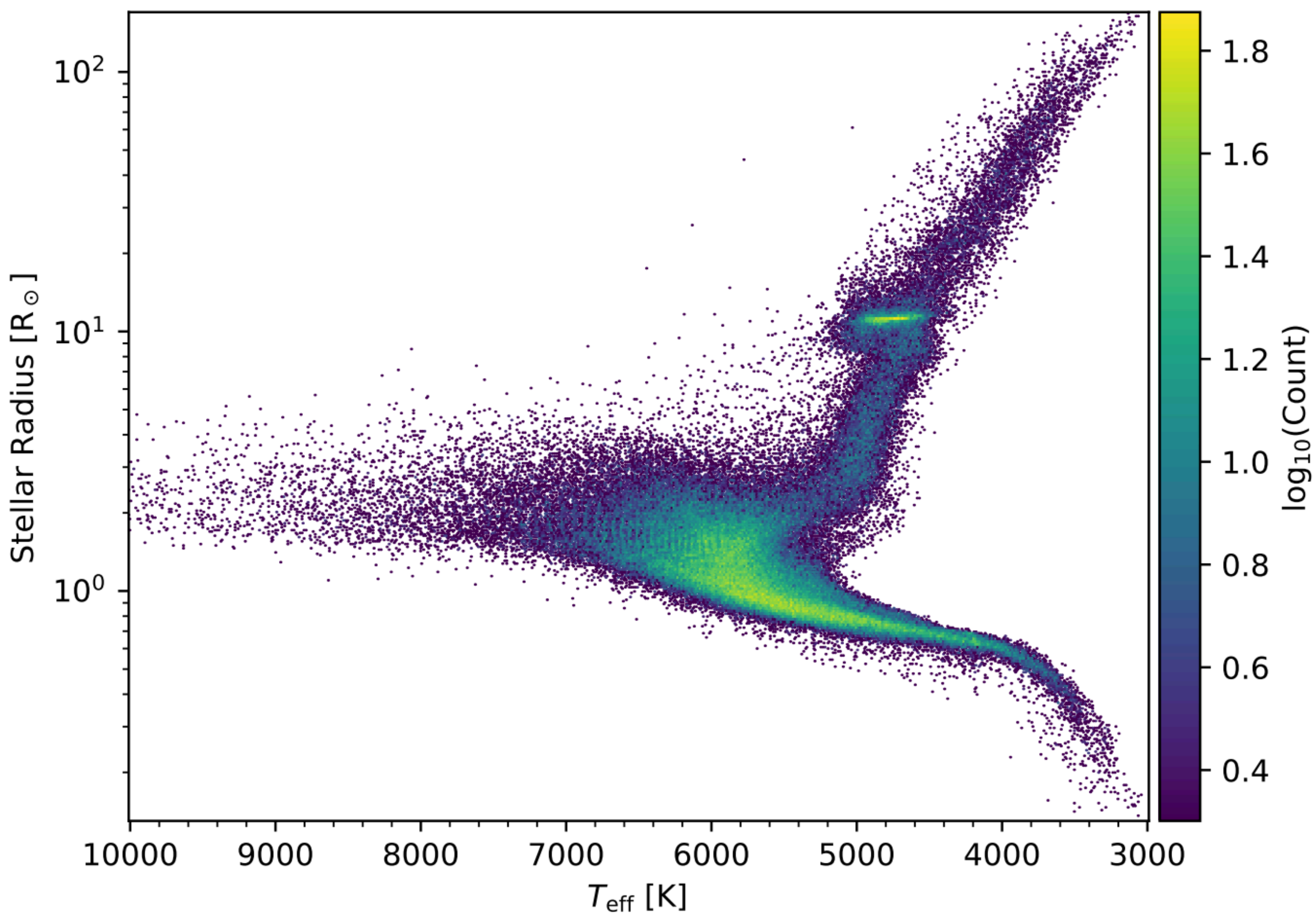
Stars as seen by SPHEREx?



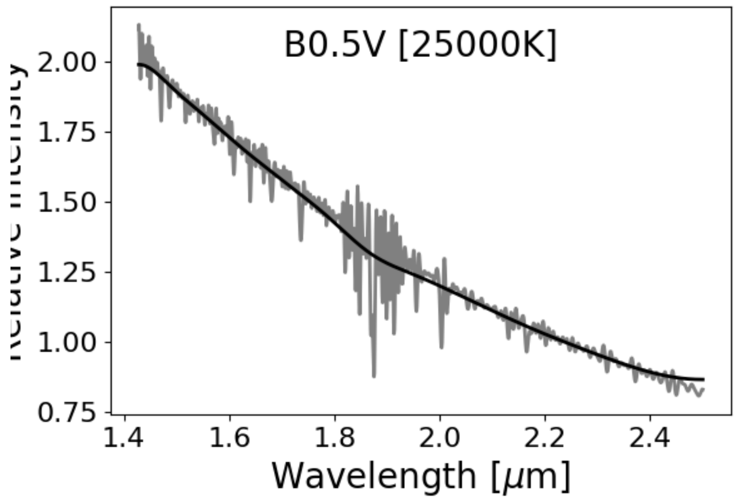
Berger+ 2019

Stars as seen by SPHEREx?

Lacon+ 1996 Near-IR spectra for 56 stars across the HR diagram R= 500 (grey)

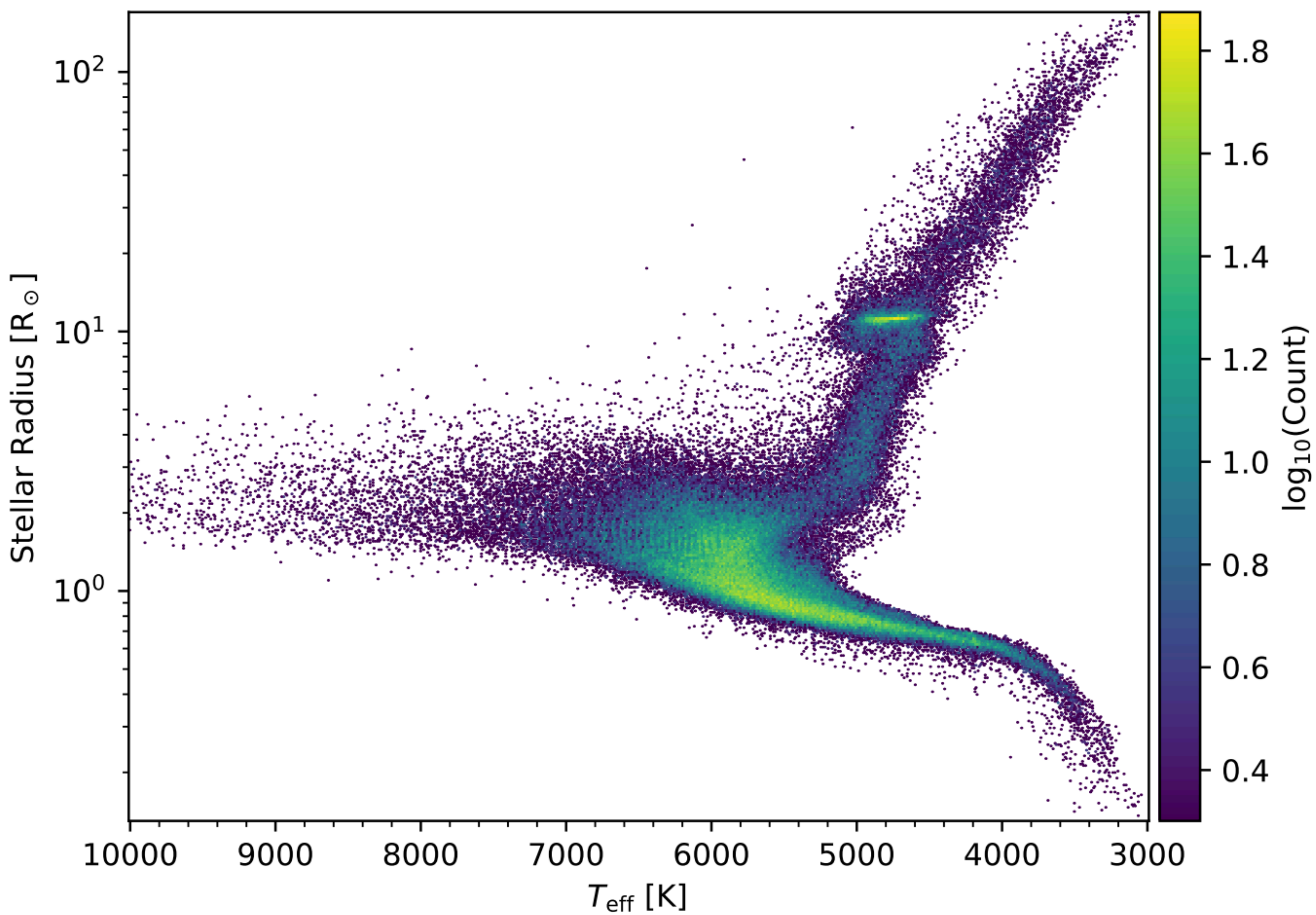


Berger+ 2019



ars as seen by SPHEREx?

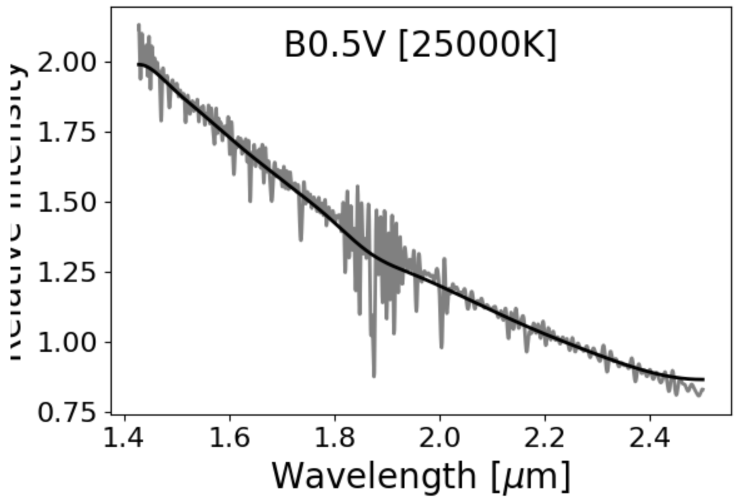
spectra for 56 stars across the HR diagram R= 500 (grey)

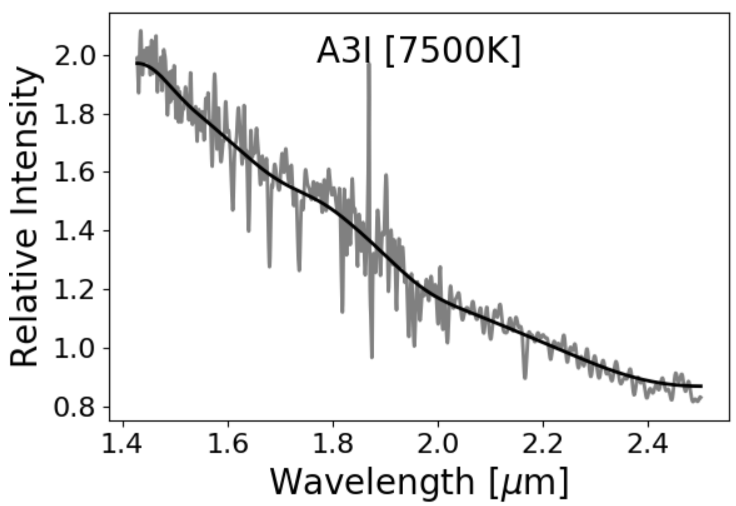


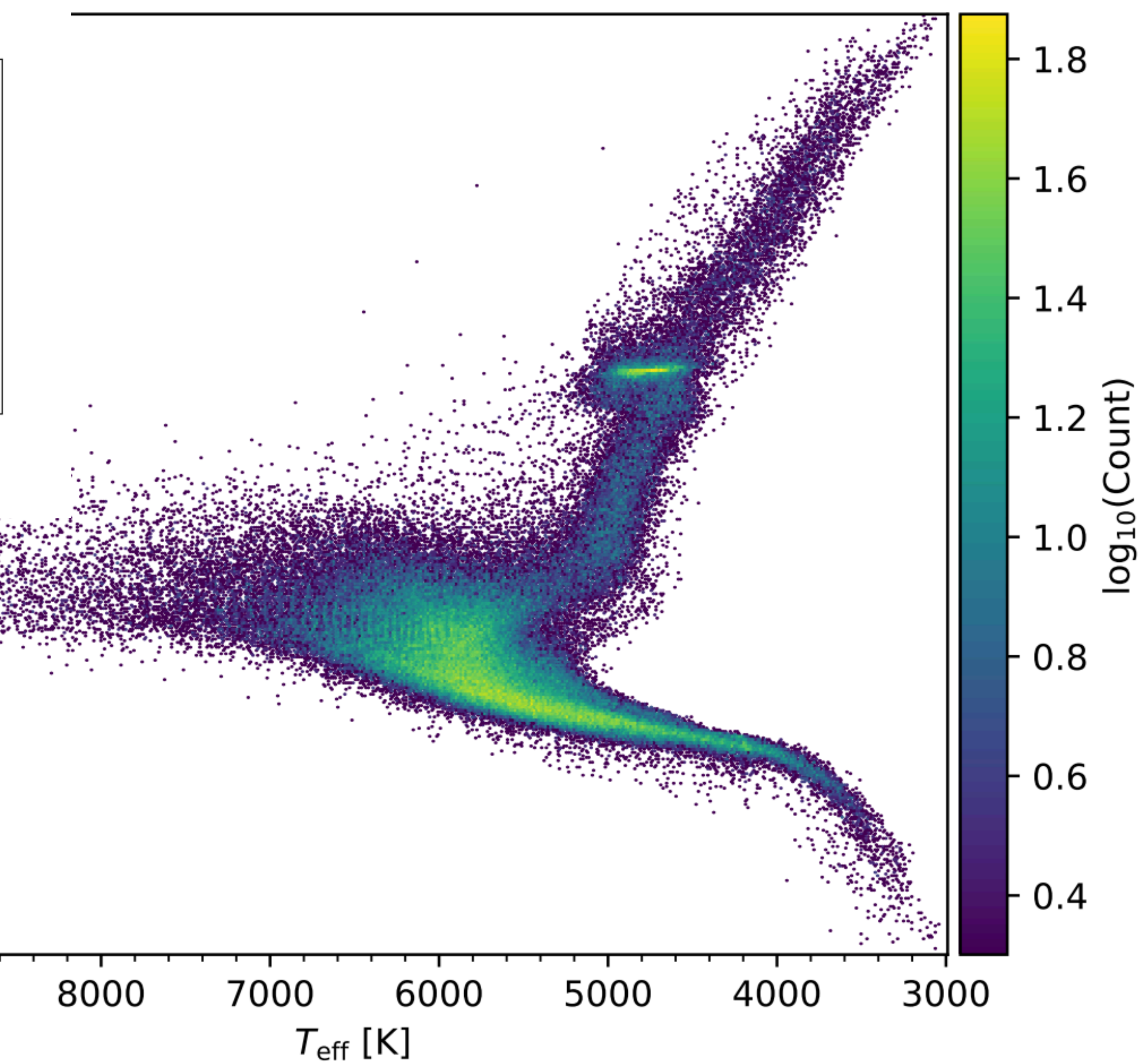
Berger+
2019

ars as seen by SPHEREx?

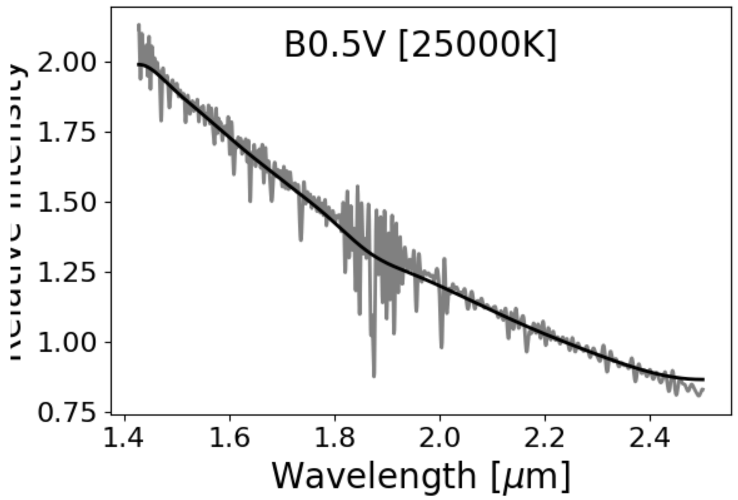
spectra for 56 stars across the HR diagram R= 500 (grey)





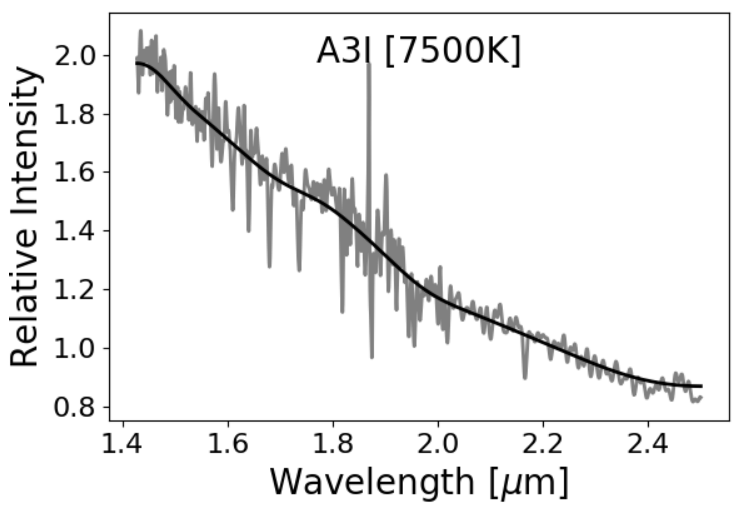


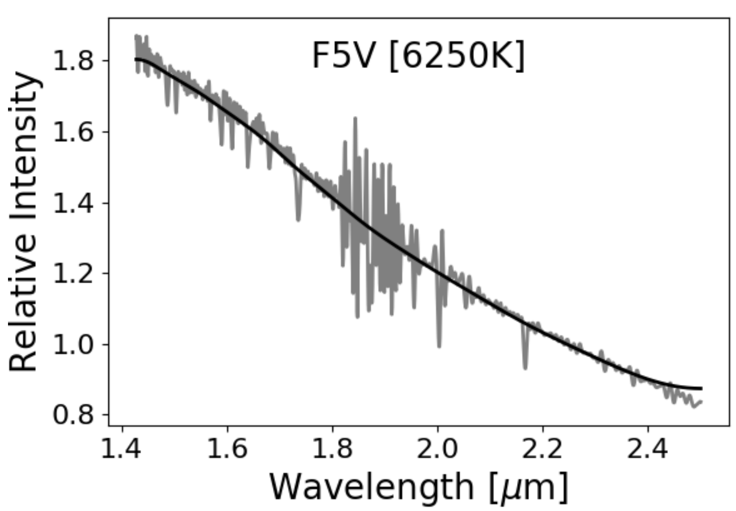
Berger+ 2019

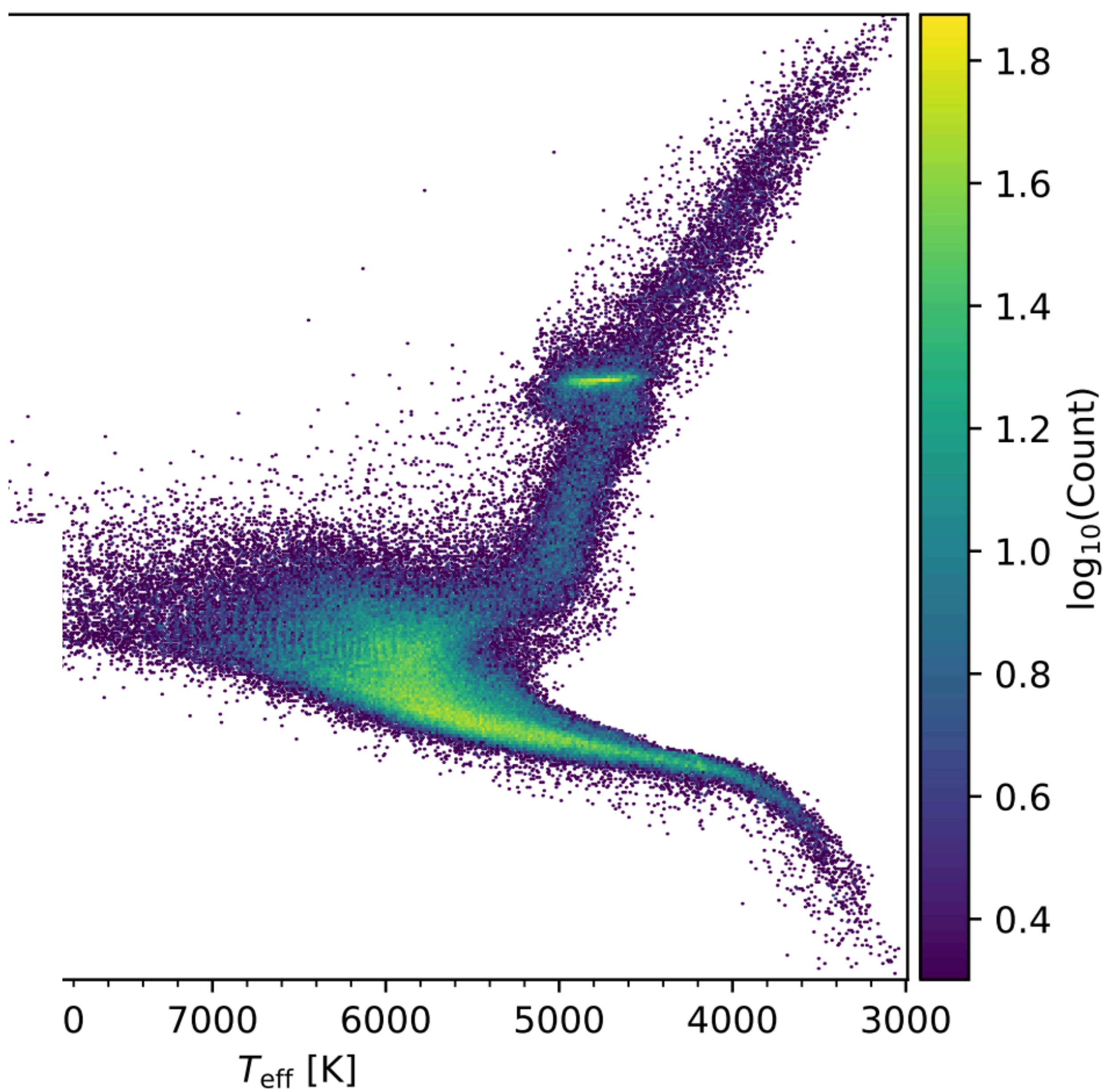


ars as seen by SPHEREx?

spectra for 56 stars across the HR diagram R= 500 (grey)



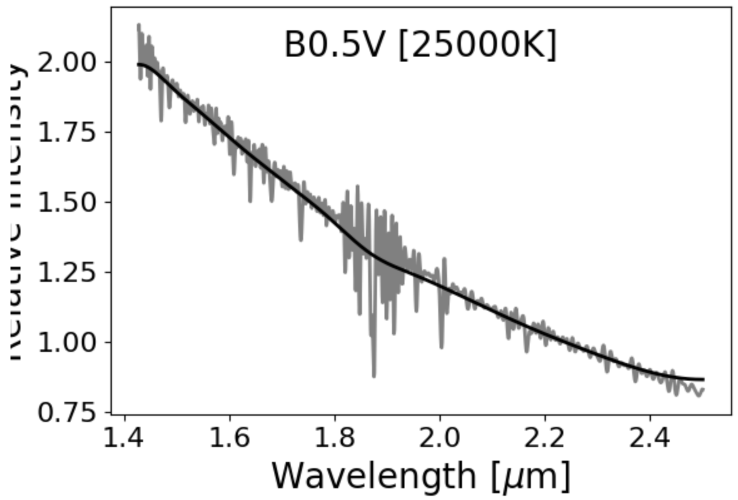


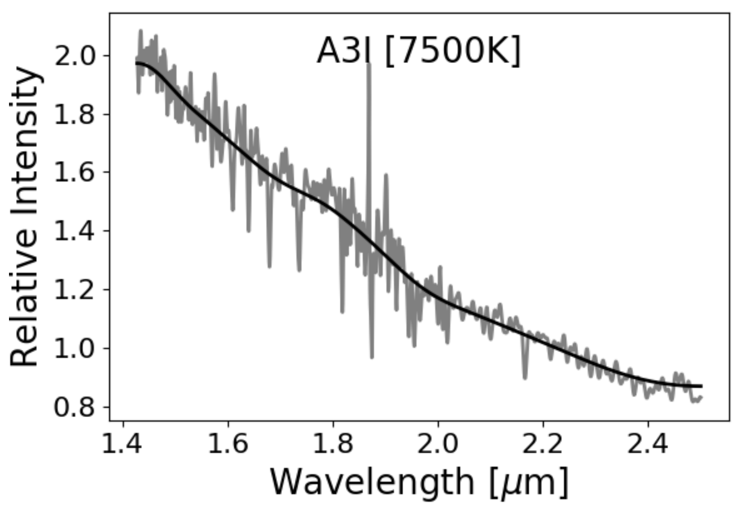


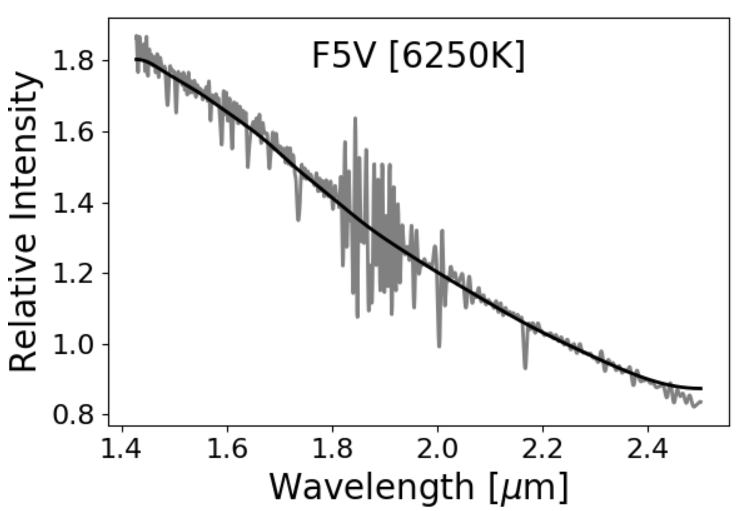
Berger+ 2019

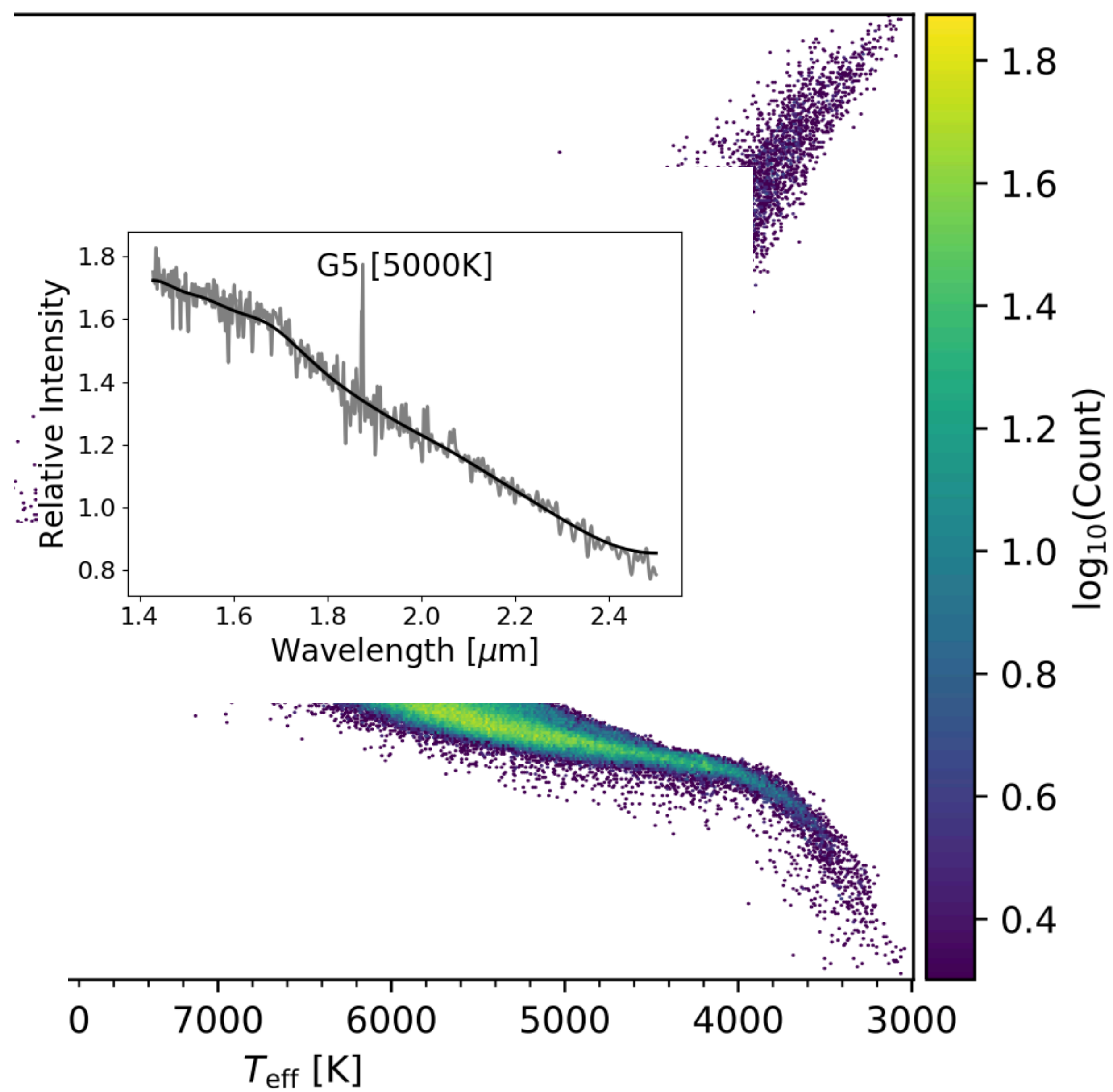
ars as seen by SPHEREx?

spectra for 56 stars across the HR diagram R= 500 (grey)







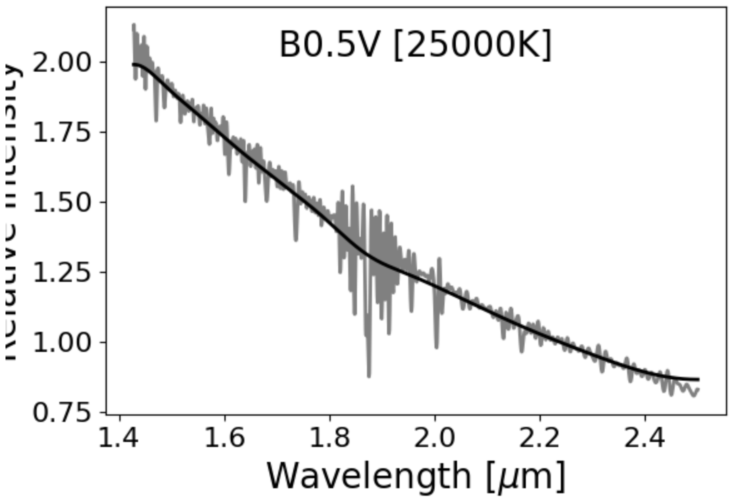


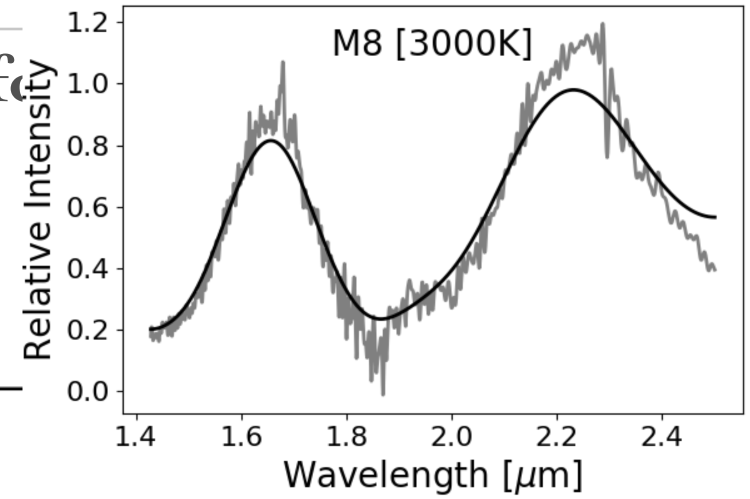
Berger+
2019

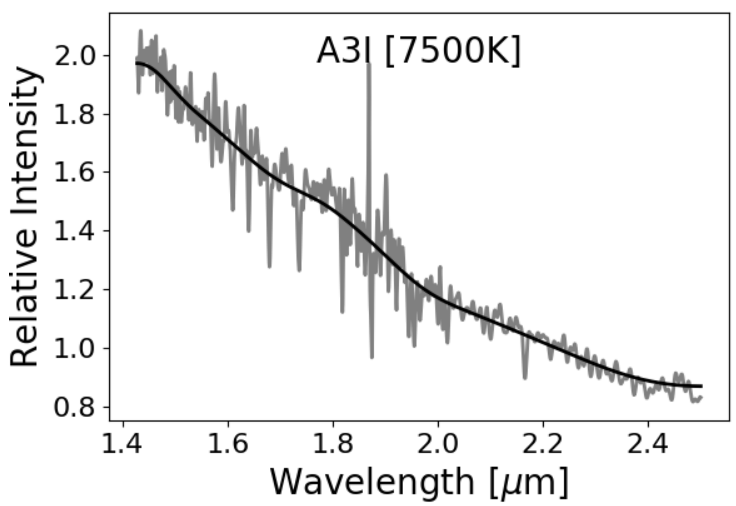
ars as seen by SPHEREx?

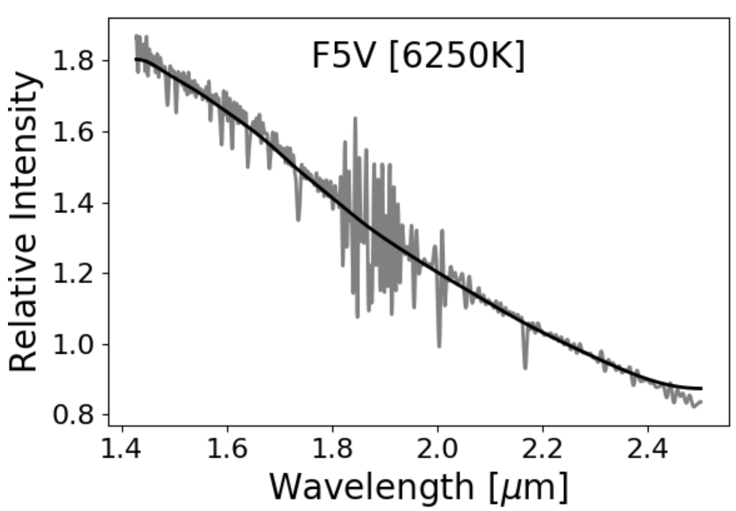
spectra fo

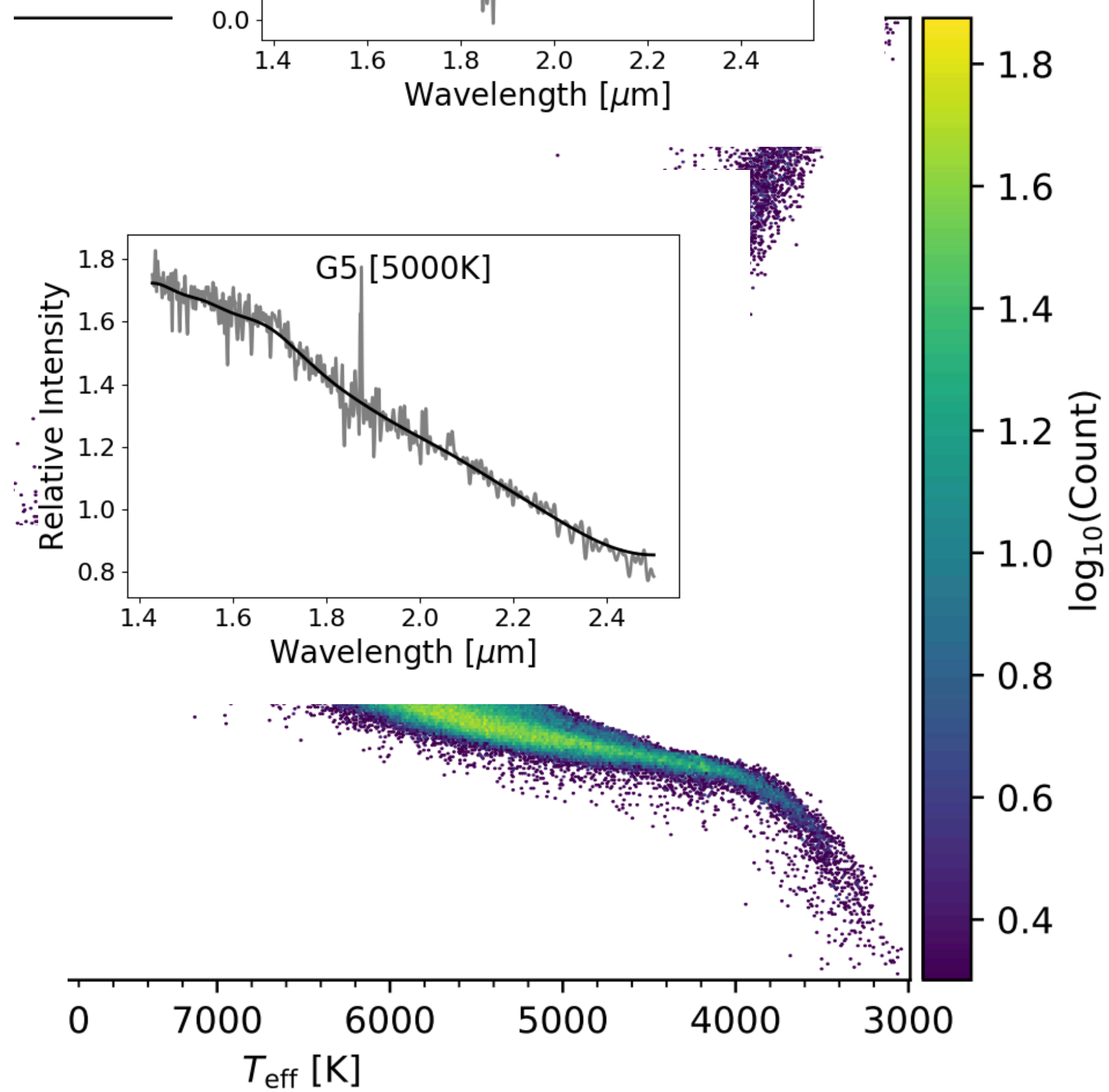
agram R= 500 (grey)







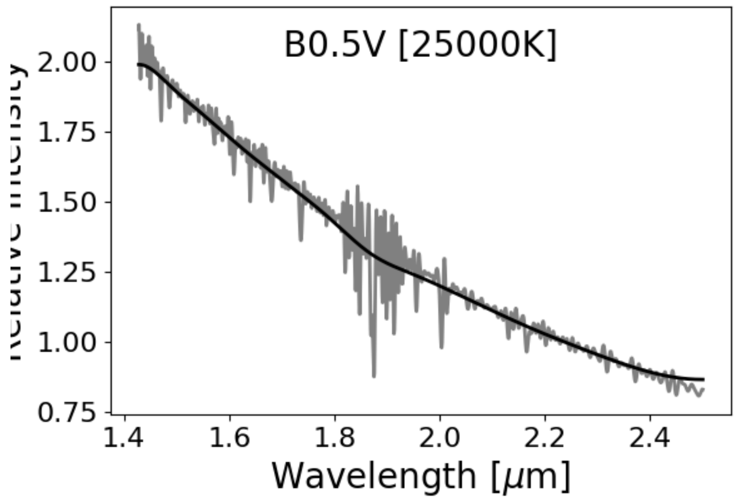


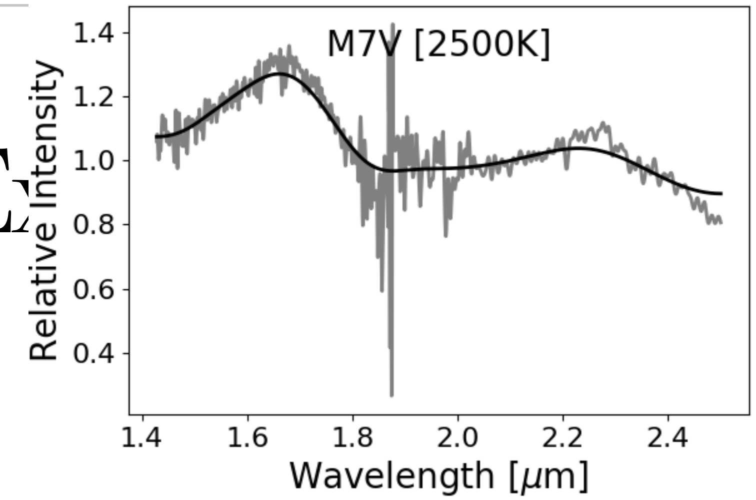


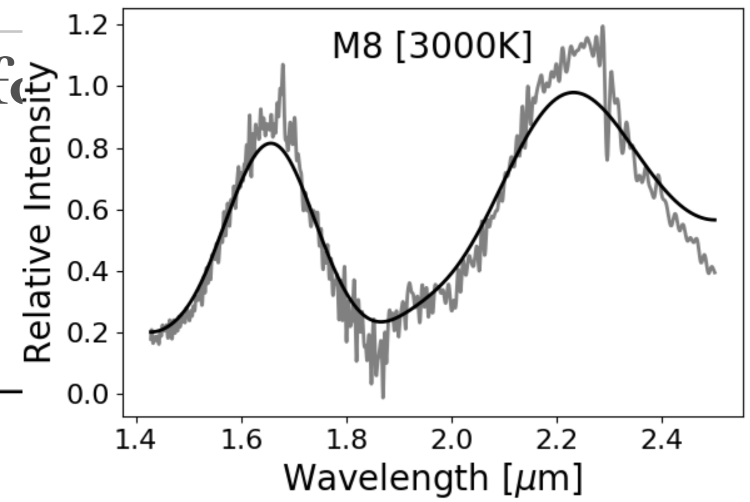
Berger+ 2019

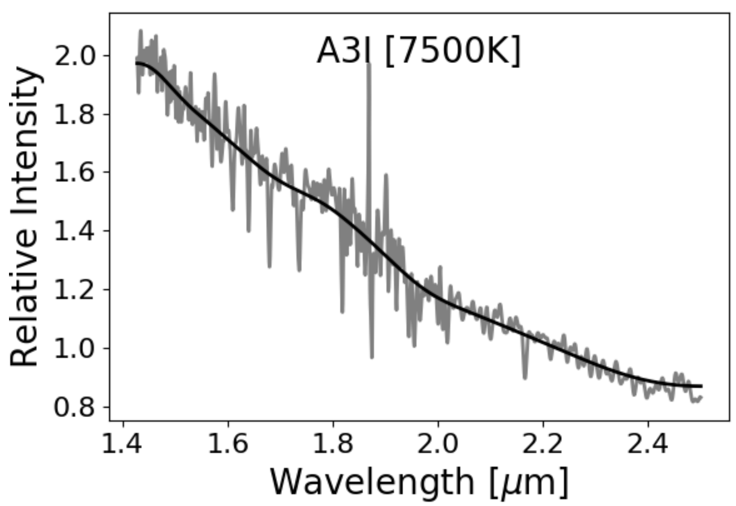
ars as seen by SPHERE

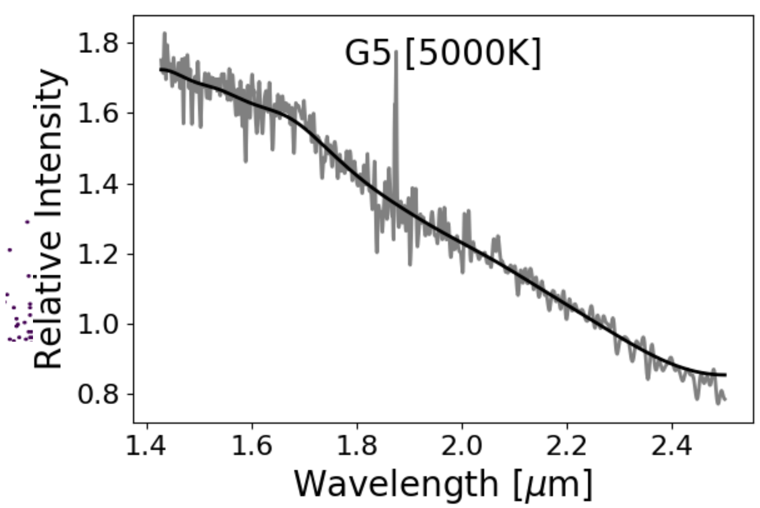
spectra f

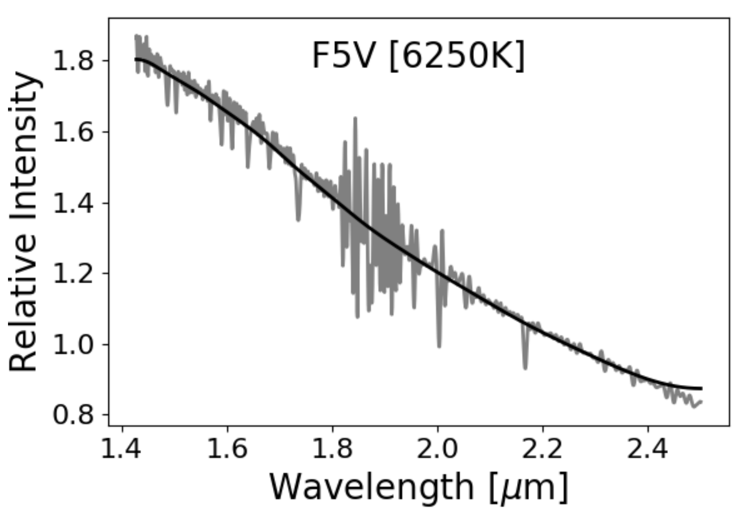


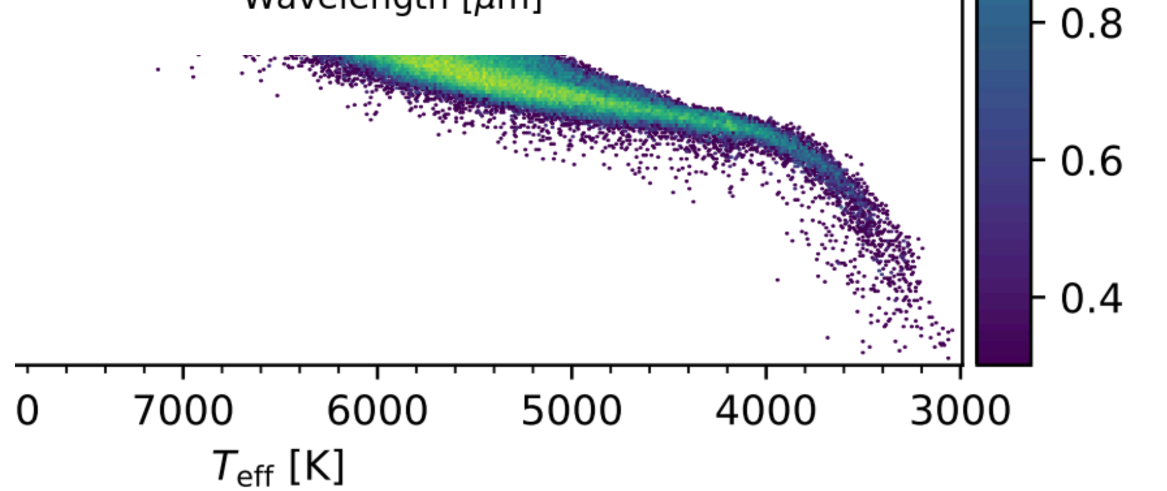








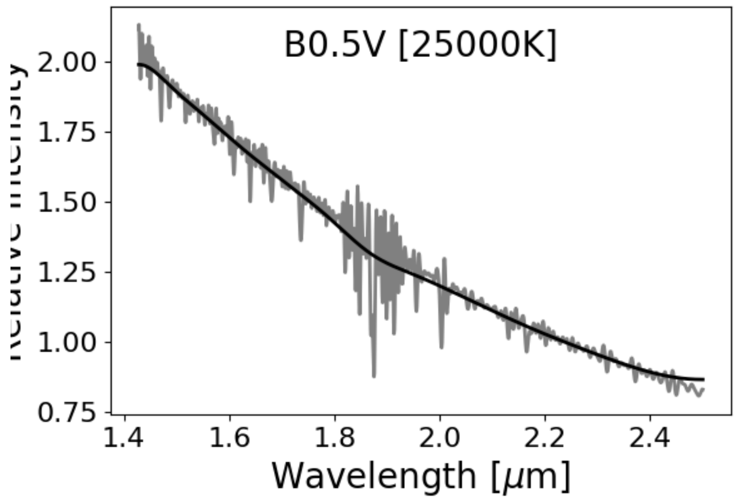


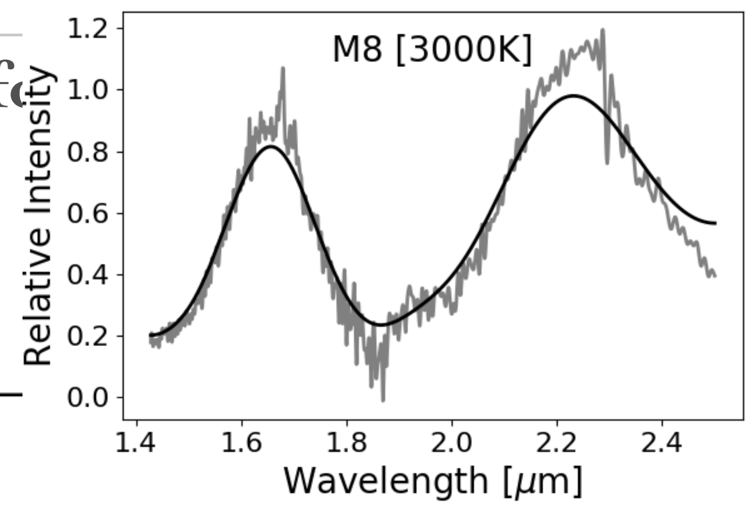


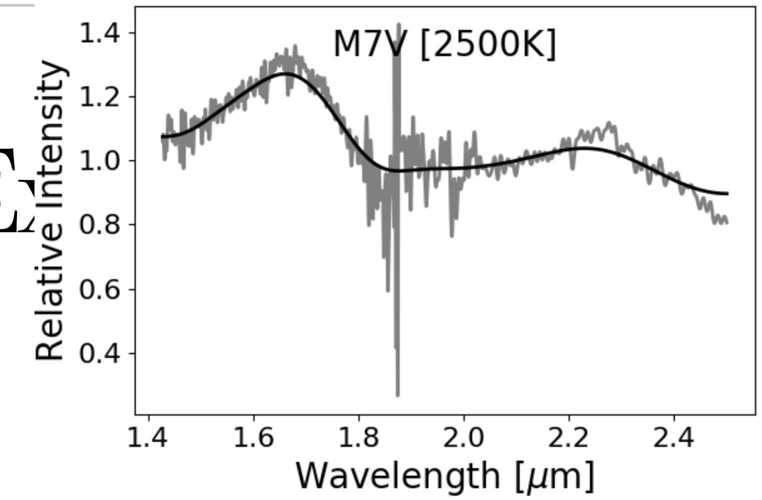
Berger+ 2019

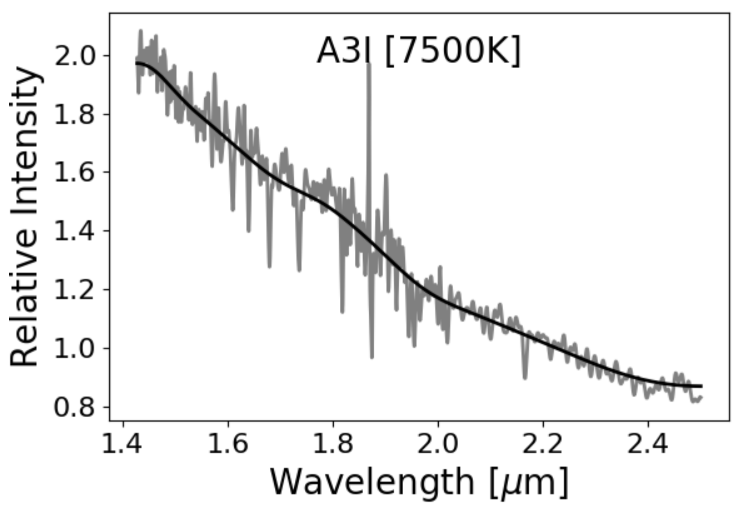
ars as seen by SPHERE

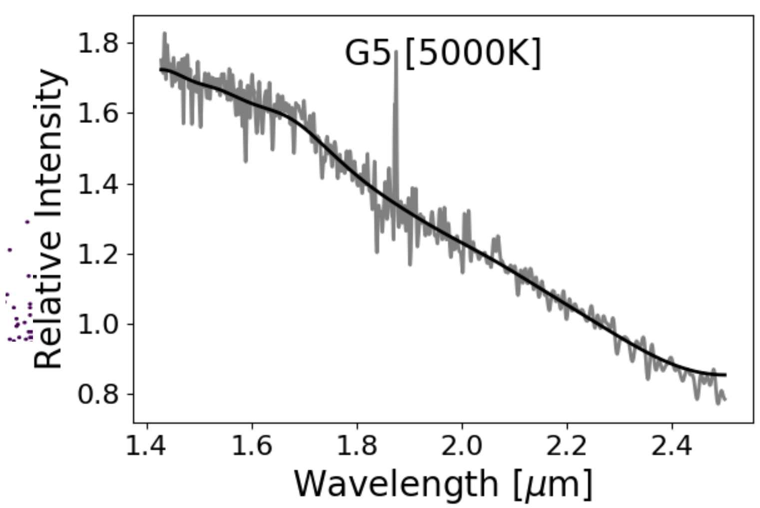
spectra fo

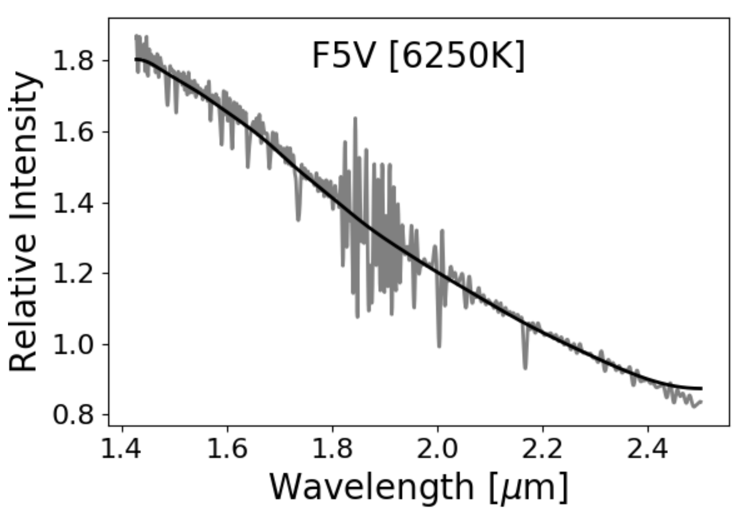


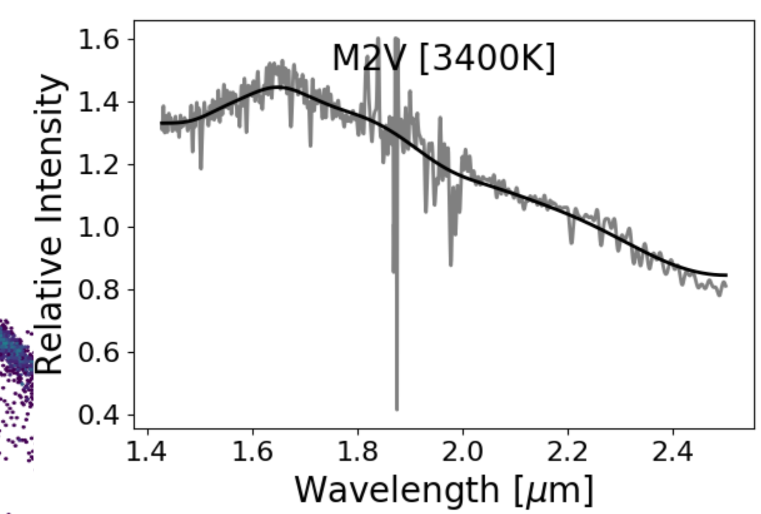


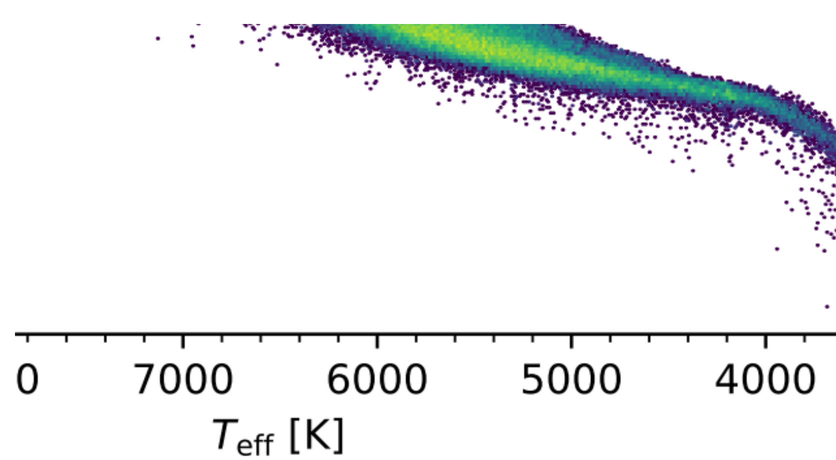












Berger+ 2019

Action item

→ Build a theoretical / empirical / semi-empirical stellar SPHEREx library

→ Characterise **what** information we can recover from *SPHEREx* (stellar) data

→ precision of inferred (Teff, logg, [M/H]) across (evolutionary state, SNR)

Action item

→ Build a theoretical / empirical / semi-empirical stellar SPHEREx library
→ Characterise **what** information we can recover from *SPHEREx* (stellar) data
→ precision of inferred (Teff, logg, [M/H]) across (evolutionary state, SNR)

Now

Apply (data-driven) methods developed within SDSS to SPHEREx

Future action item

Future action item

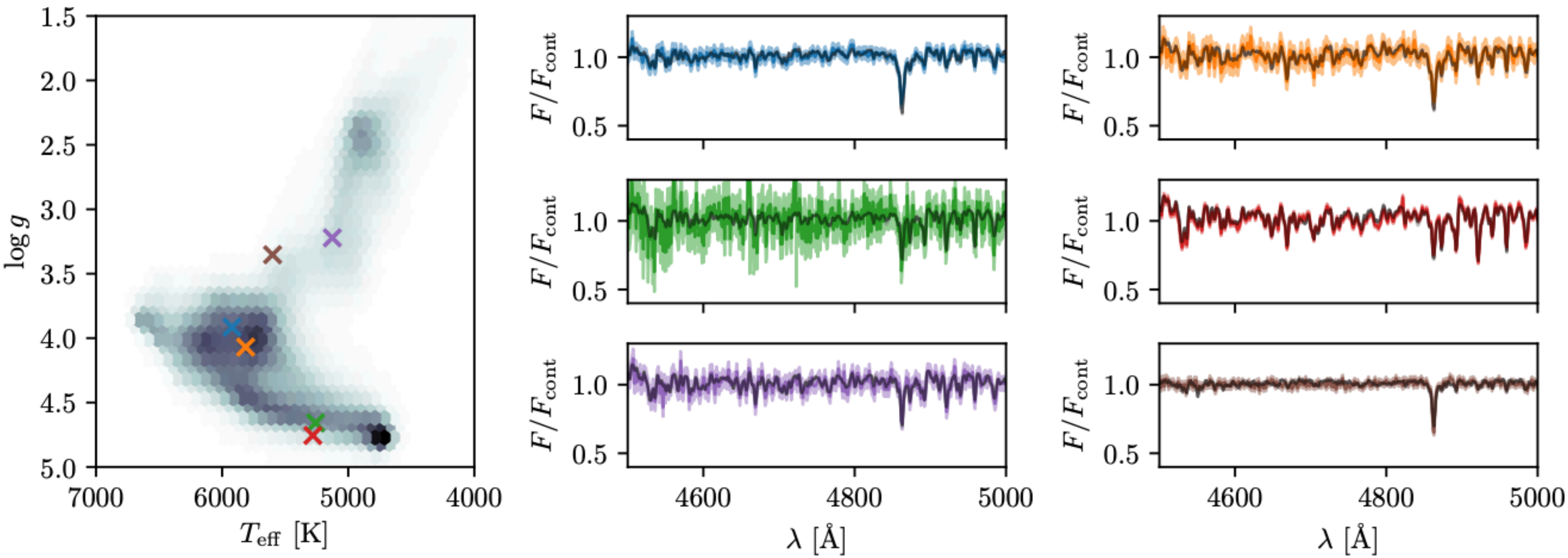
Where do data-driven models *not* match the data?

Future action item

Where do data-driven models *not* match the data?

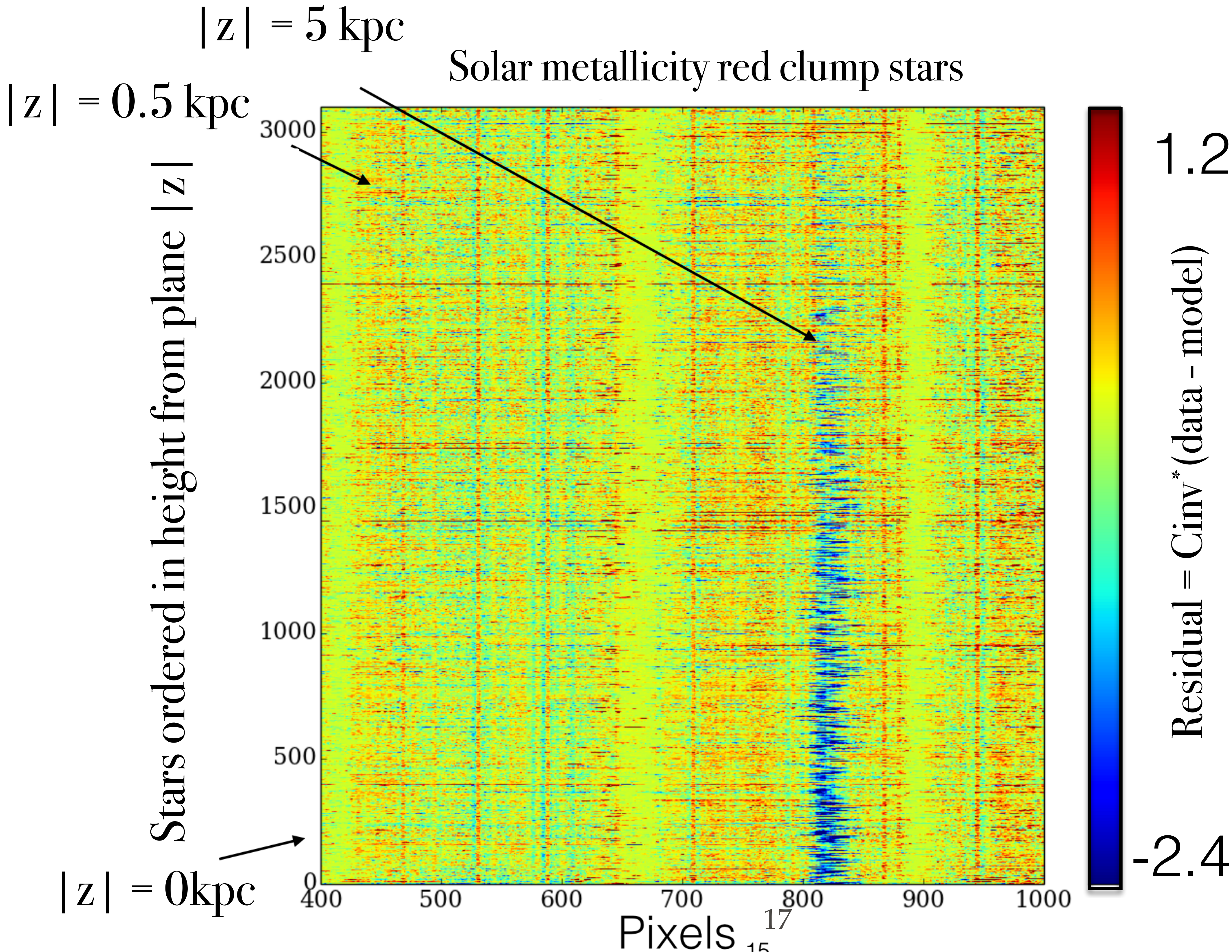
→ Non-stellar features in the spectrum (dust, molecules in ISM)
→ Peculiar chemical composition (dynamical interactions, planet ingestion)

The Cannon provides a remarkable match to the data

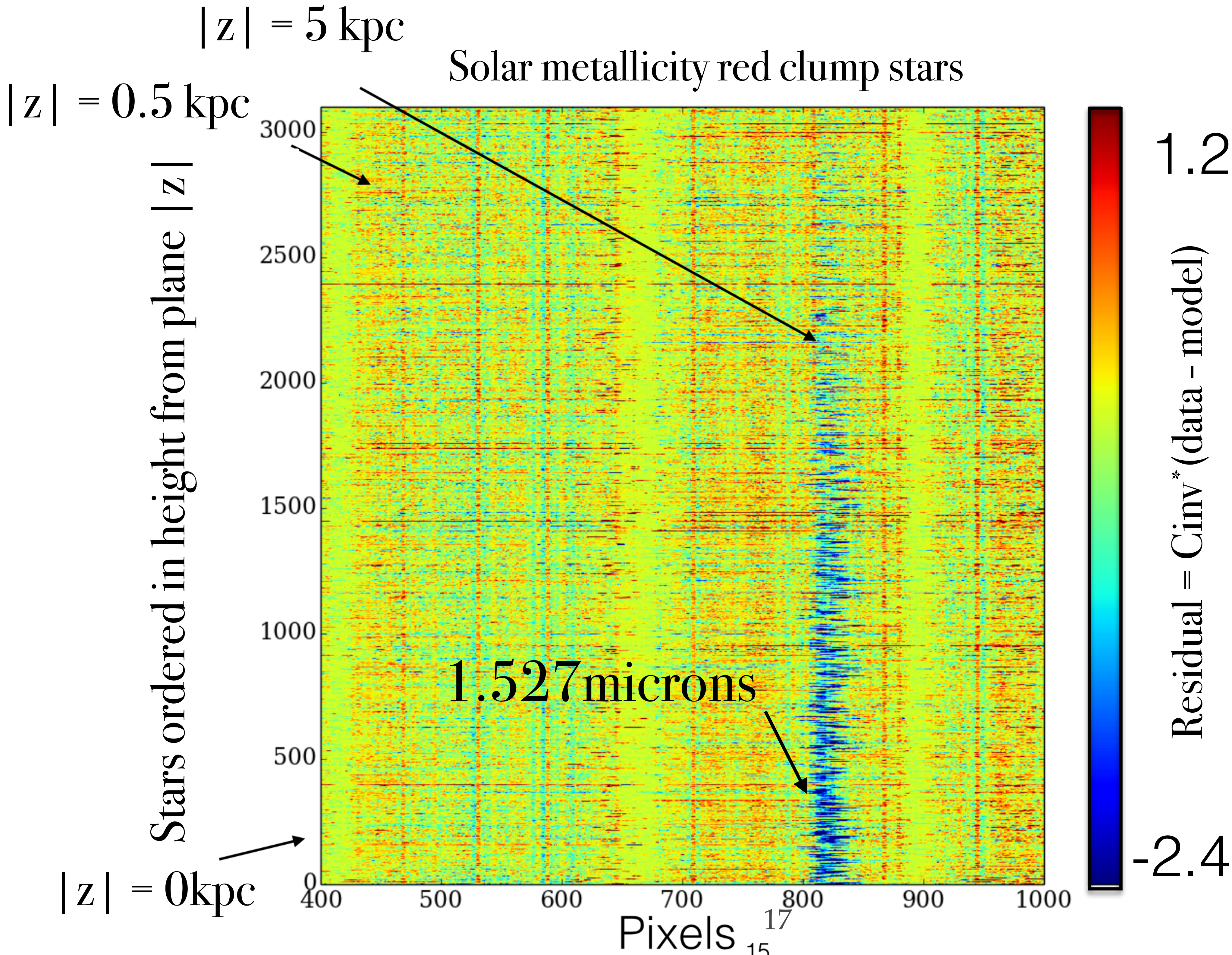


Wheeler,
Ness+ 2020

Failures: non-stellar



Failures: non-stellar



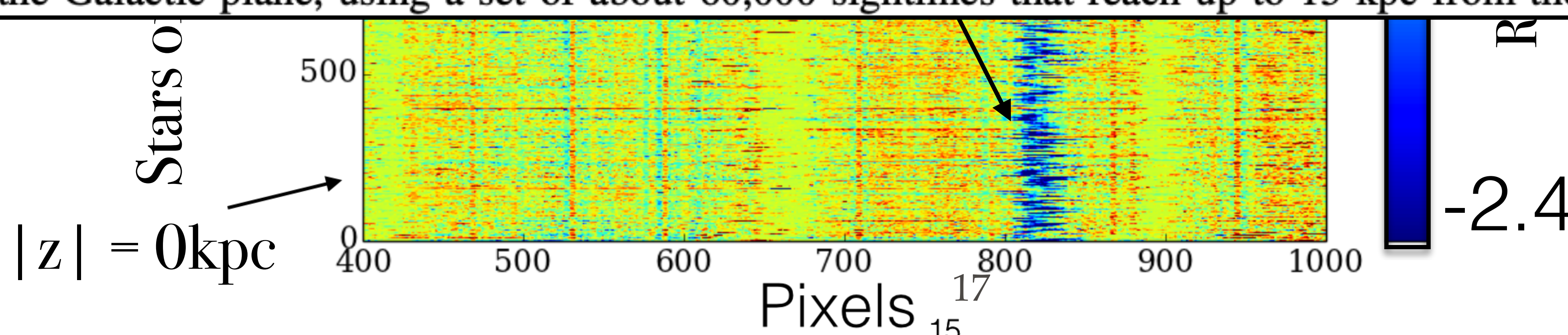
Failures: non-stellar

MAPPING THE INTERSTELLAR MEDIUM WITH NEAR-INFRARED DIFFUSE INTERSTELLAR BAND

G. Zasowski[1,11], B. Ménard[1,2,12], D. Bizyaev[3,4], D. A. García-Hernández[5,6], A. E. García Pérez[7], M. R. Hayden
J. Holtzman[4], J. A. Johnson[8], K. Kinemuchi[4], S. R. Majewski[7], D. L. Nidever[9], M. Shetrone[10], and J. C. Wilson
[1] Department of Physics & Astronomy, Johns Hopkins University, Baltimore, MD 21218, USA; gail.zasowski@gmail.com
[2] Kavli Institute for the Physics and Mathematics of the Universe, University of Tokyo, Kashiwa 277-8583, Japan
[3] Apache Point Observatory, Sunspot, NM 88349, USA
[4] Department of Astronomy, New Mexico State University, Las Cruces, NM 88003, USA
[5] Instituto de Astrofísica de Canarias, E-38205 La Laguna, Tenerife, Spain
[6] Departamento de Astrofísica, Universidad de La Laguna, E-38206 La Laguna, Tenerife, Spain
[7] Department of Astronomy, University of Virginia, Charlottesville, VA 22904, USA
[8] Department of Astronomy, The Ohio State University, Columbus, OH 43210, USA
[9] Department of Astronomy, University of Michigan, Ann Arbor, MI 48104, USA
[10] The University of Texas at Austin, McDonald Observatory, McDonald Observatory, TX 79734, USA
Received 2014 June 4; accepted 2014 October 20; published 2014 December 18

ABSTRACT

We map the distribution and properties of the Milky Way's interstellar medium as traced by diffuse interstellar bands (DIBs) detected in near-infrared stellar spectra from the SDSS-III/APOGEE survey. Focusing exclusively on the strongest DIB in the *H* band, at $\lambda \sim 1.527\ \mu$m, we present a projected map of the DIB absorption field in the Galactic plane, using a set of about 60,000 sightlines that reach up to 15 kpc from the Sun and probe up to



Failures: peculiar stars

Where data-driven model fails (around labels **not** modeled)

Failures: peculiar stars

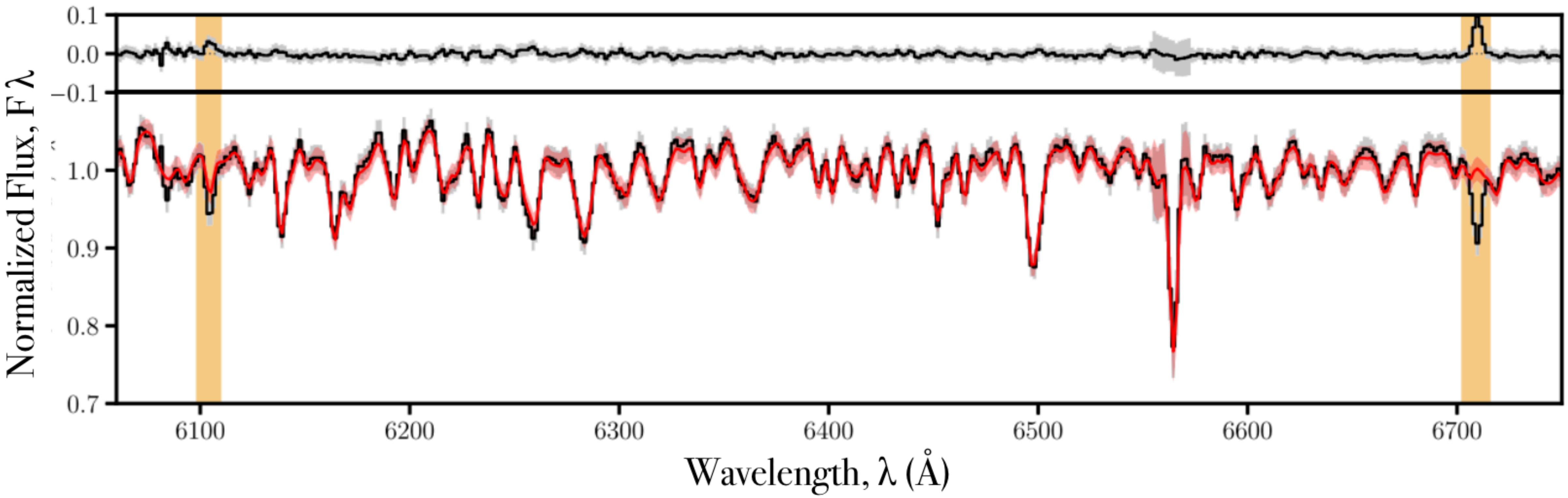
Where data-driven model fails (around labels **not** modeled)

LAMOST - two **Lithium** features

Failures: peculiar stars

Where data-driven model fails (around labels **not** modeled)

LAMOST - two **Lithium** features

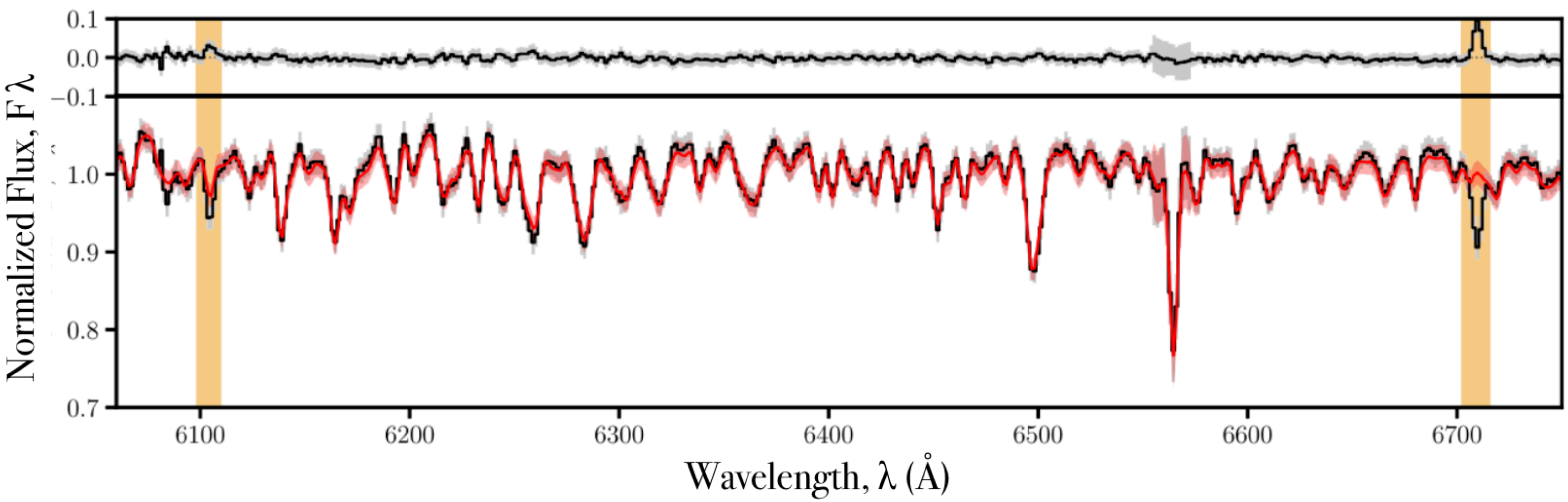


2300 Lithium rich giants discovered in LAMOST (Casey+ 2019, Ho+ 2017)
looking where Li lines >> data-driven model (no Li modeled),
signature of tidal spin up and in some cases planet consumption (Soares-Furtado+ 2020)

Failures: peculiar stars

Where data-driven model fails (around labels **not** modeled)

LAMOST - two **Lithium** features



2300 Lithium rich giants discovered in LAMOST (Casey+ 2019, Ho+ 2017)
looking where Li lines >> data-driven model (no Li modeled),
signature of tidal spin up and in some cases planet consumption (Soares-Furtado+ 2020)

Wheeler in prep (unsupervised)

What I proposed today

Now

Later

What I proposed today

Now

Characterise **what** information we can recover from *SPHEREx* (stellar) data

Later

What I proposed today

Now

Characterise **what** information we can recover from *SPHEREx* (stellar) data

→ *Build a stellar SPHEREx library*

→ *Apply data-driven approach to determine labels (Teff, logg, [M/H])*

→ *precision f(evolutionary state, SNR)*

Later

What I proposed today

Now

Characterise **what** information we can recover from *SPHEREx* (stellar) data

→ *Build a stellar SPHEREx library*

→ *Apply data-driven approach to determine labels (Teff, logg, [M/H])*

→ *precision f(evolutionary state, SNR)*

Later

Where do data-driven models *not* match the data?

What I proposed today

Now

Characterise **what** information we can recover from *SPHEREx* (stellar) data

→ *Build a stellar SPHEREx library*

→ *Apply data-driven approach to determine labels (Teff, logg, [M/H])*

→ *precision f(evolutionary state, SNR)*

Later

Where do data-driven models *not* match the data?

→ *Non-stellar features in the spectrum (dust, molecules in ISM)*

→ *Peculiar spectra due to perturbations (external work, planet ingestion, mass exchange, extra-galactic)*

Large coverage of the disk

APOGEE: 250K stars

Milky Way Mapper:
5 million stars

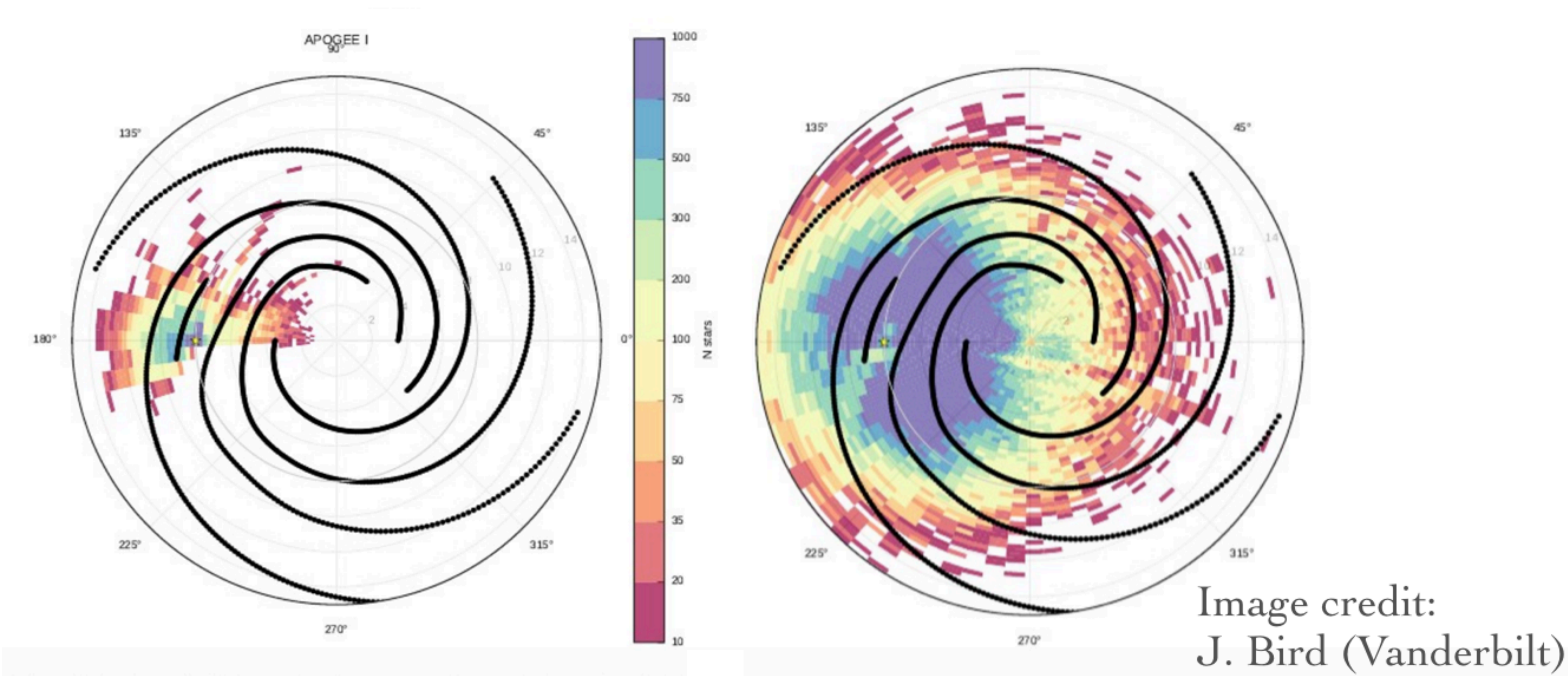


Image credit:
J. Bird (Vanderbilt)

Large coverage of the disk

APOGEE: 250K stars

Milky Way Mapper:
5 million stars

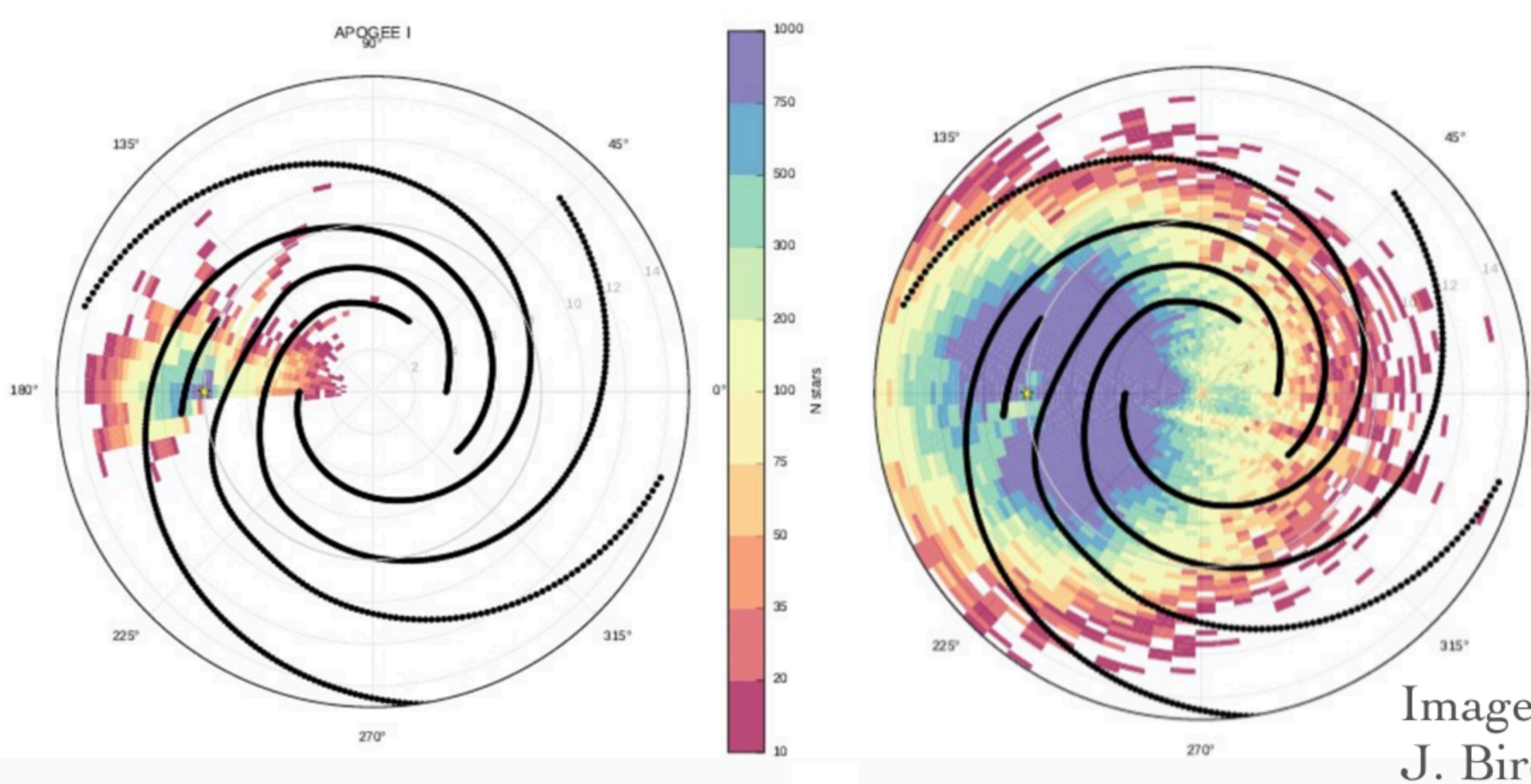


Image credit:
J. Bird (Vanderbilt)

Spectrophotmetric distances with data-driven models (in dust obscured crowded disk)
(Hogg+2019, Eilers+2019, Leung+2019)