Efforts to characterize exoplanet atmospheres in the presence of angry stars

Luis Welbanks

51 Pegasi b Fellow → Assistant Professor

luis.welbanks@asu.edu

Arizona State University

HEISING-SIMONS FOUNDATION

L. Calçada/ESO

The atmosphere is the most observable portion of an extra-solar body, therefore the information we learn about the processes and history of that object will come from its atmosphere.

The atmosphere is the most observable portion of an extra-solar body, therefore the information we learn about **the processes and history** of that object will come from its atmosphere.

Leveraging the Population to Understand Planet Formation

The atmosphere is the most observable portion of an extra-solar body, therefore the information we learn about **the processes and history** of that object will come from its atmosphere.

Leveraging the Population to Understand Planet Formation

Leveraging the Population to Understand Planetary Scale Processes

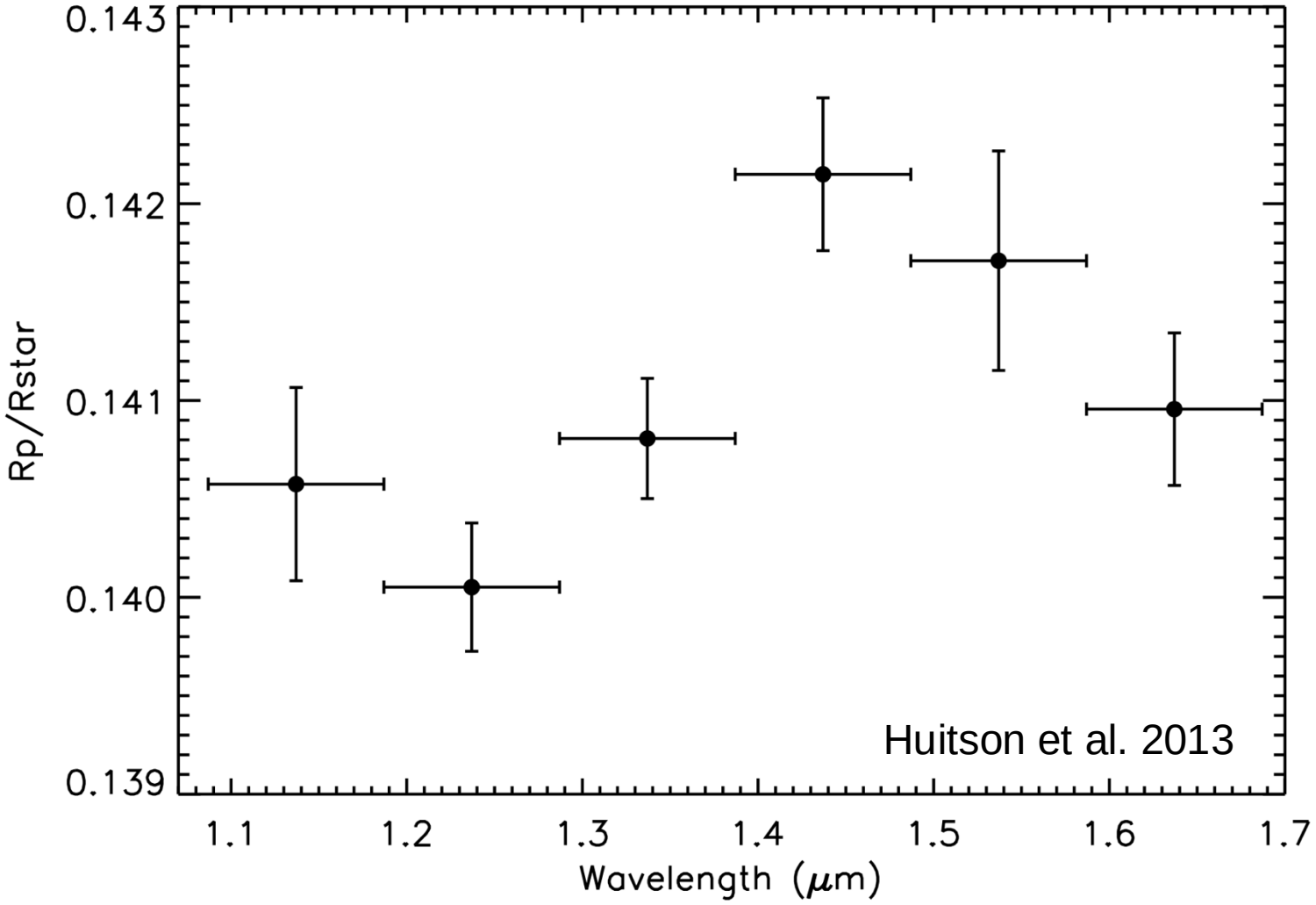
The atmosphere is the most observable portion of an extra-solar body, therefore the information we learn about **the processes and history** of that object will come from its atmosphere.

Leveraging the Population to Understand Planet Formation

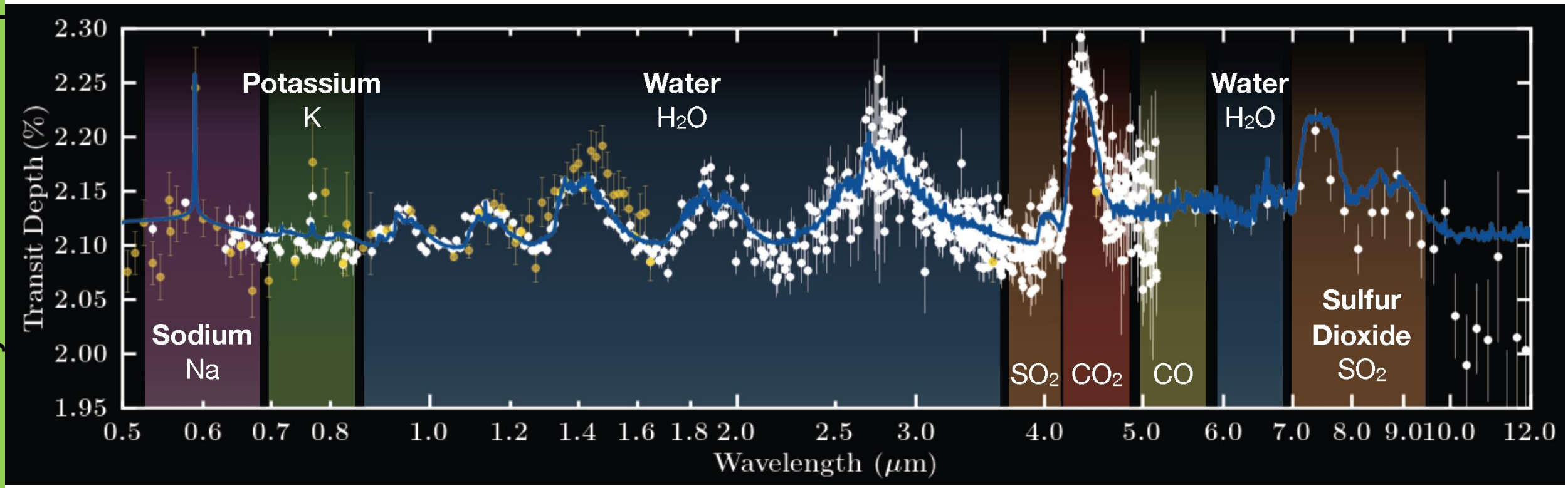
Leveraging the Population to Understand Planetary Scale Processes

… and yes, **life**

That was the way then…



That's still the way now…



Welbanks et al. in prep

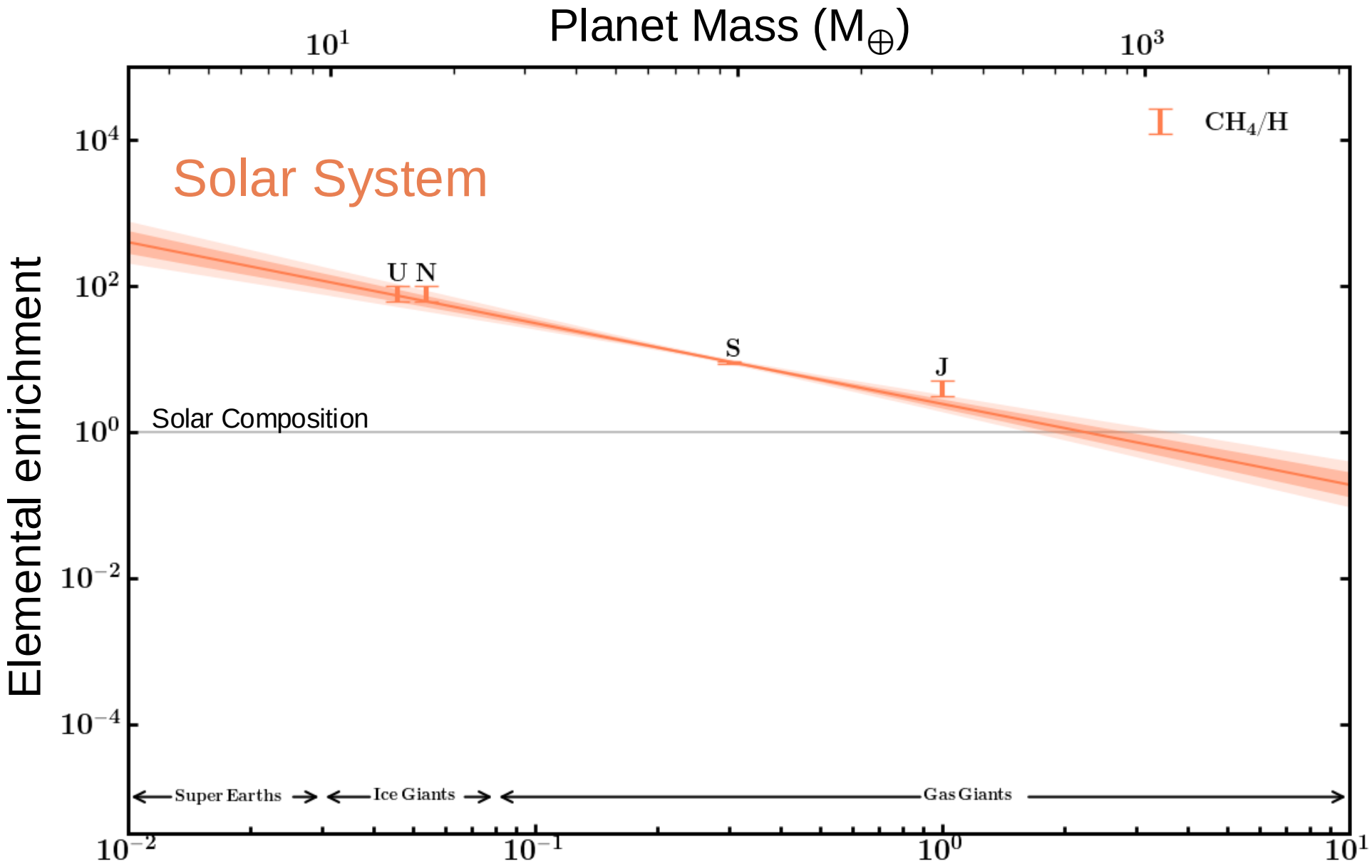
Testing Hypothesis on the Primordial Formation Pathways of Exoplanets.

How should atmospheric metallicity change with planetary mass?

Hypothesis: **Metallicity** of a planetary atmosphere should **increase** with **decreasing mass** (core-accretion model of planet formation)

Evidence: A trend that appears within the solar system Jovian planet population.

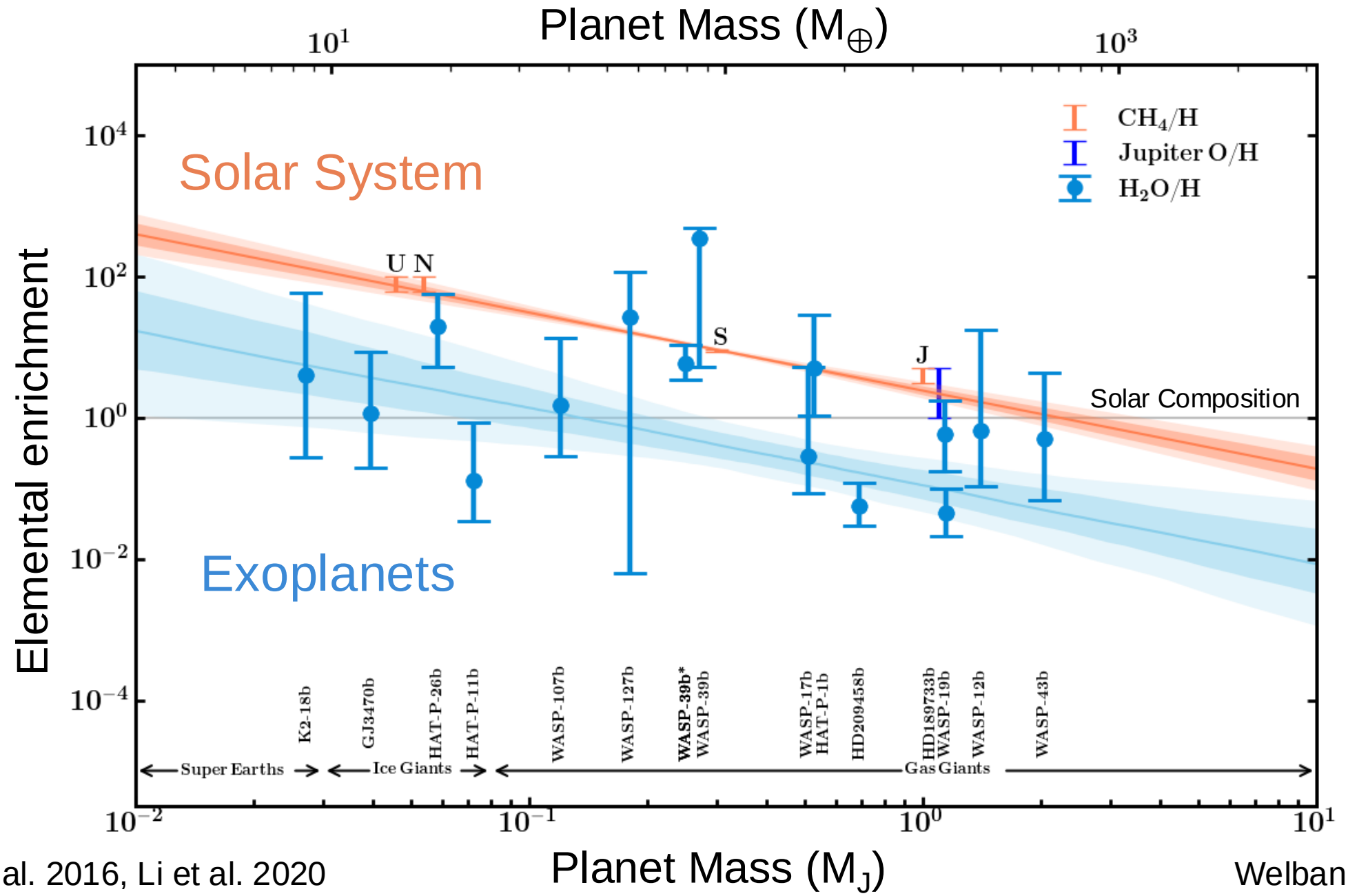
Comparing Exoplanets compositions to Solar System



Atreya et al. 2016

Welbanks et al. 2019b

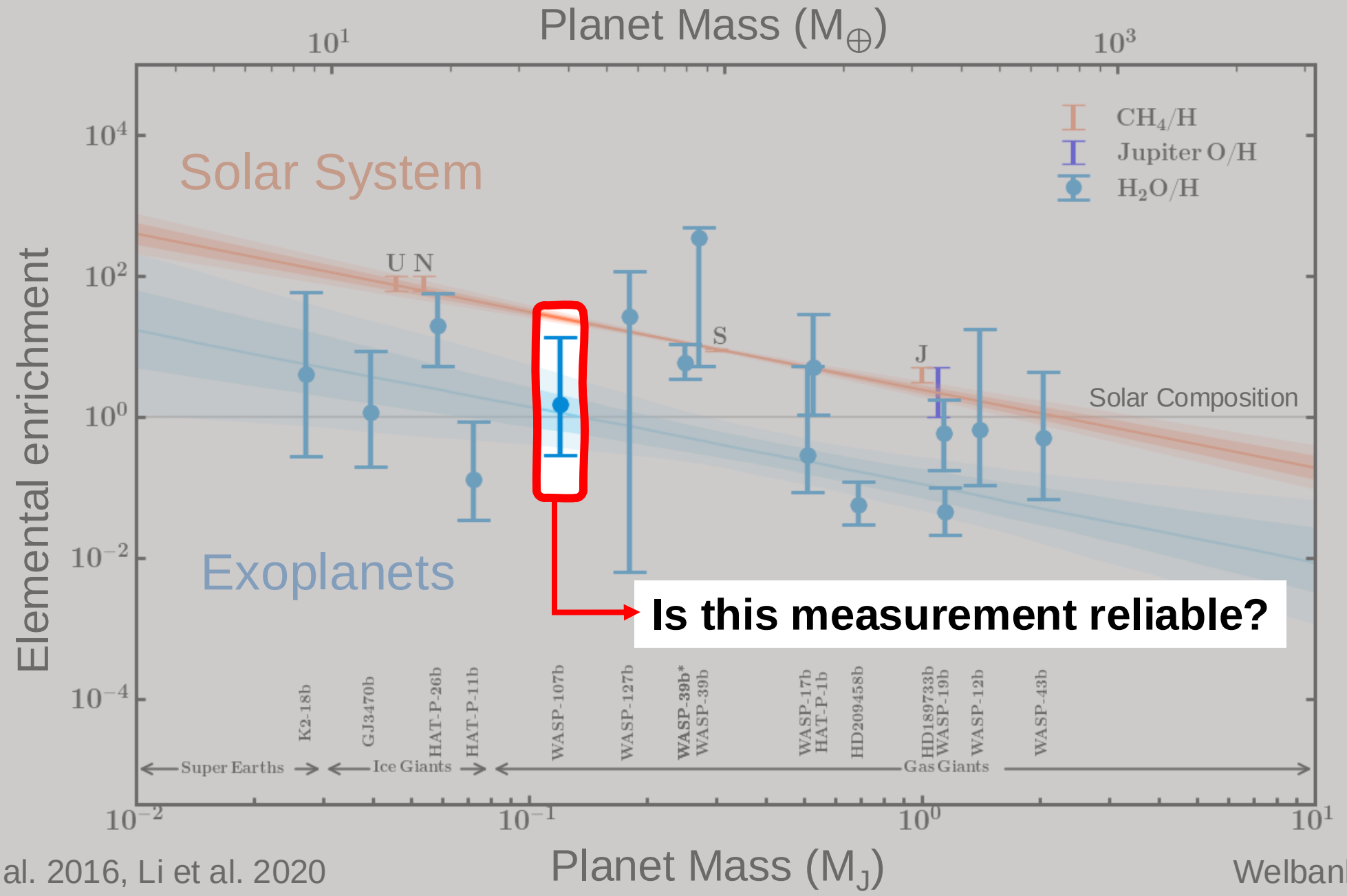
Comparing Exoplanets compositions to Solar System



Atreya et al. 2016, Li et al. 2020

Welbanks et al. 2019b

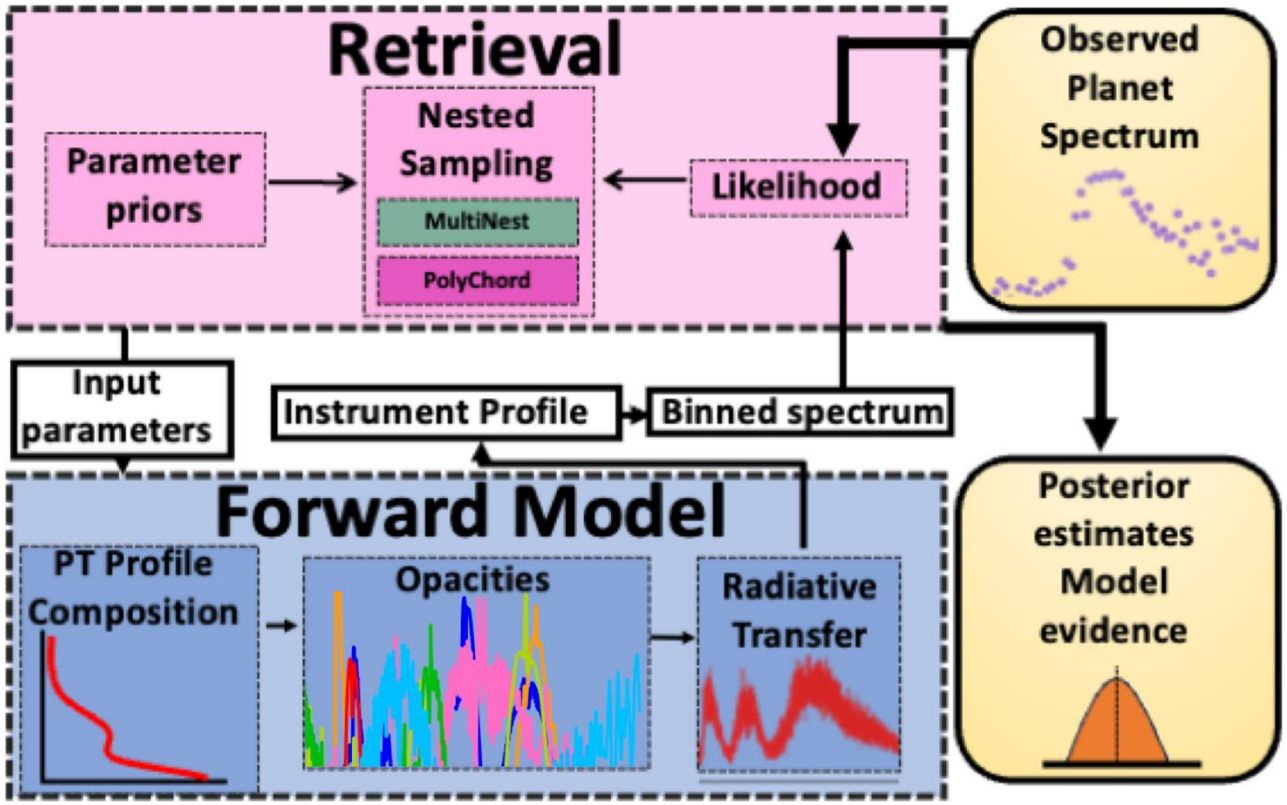
Comparing Exoplanets compositions to Solar System



Atreya et al. 2016, Li et al. 2020

Welbanks et al. 2019b

Remote Sensing of Exoplanets 101

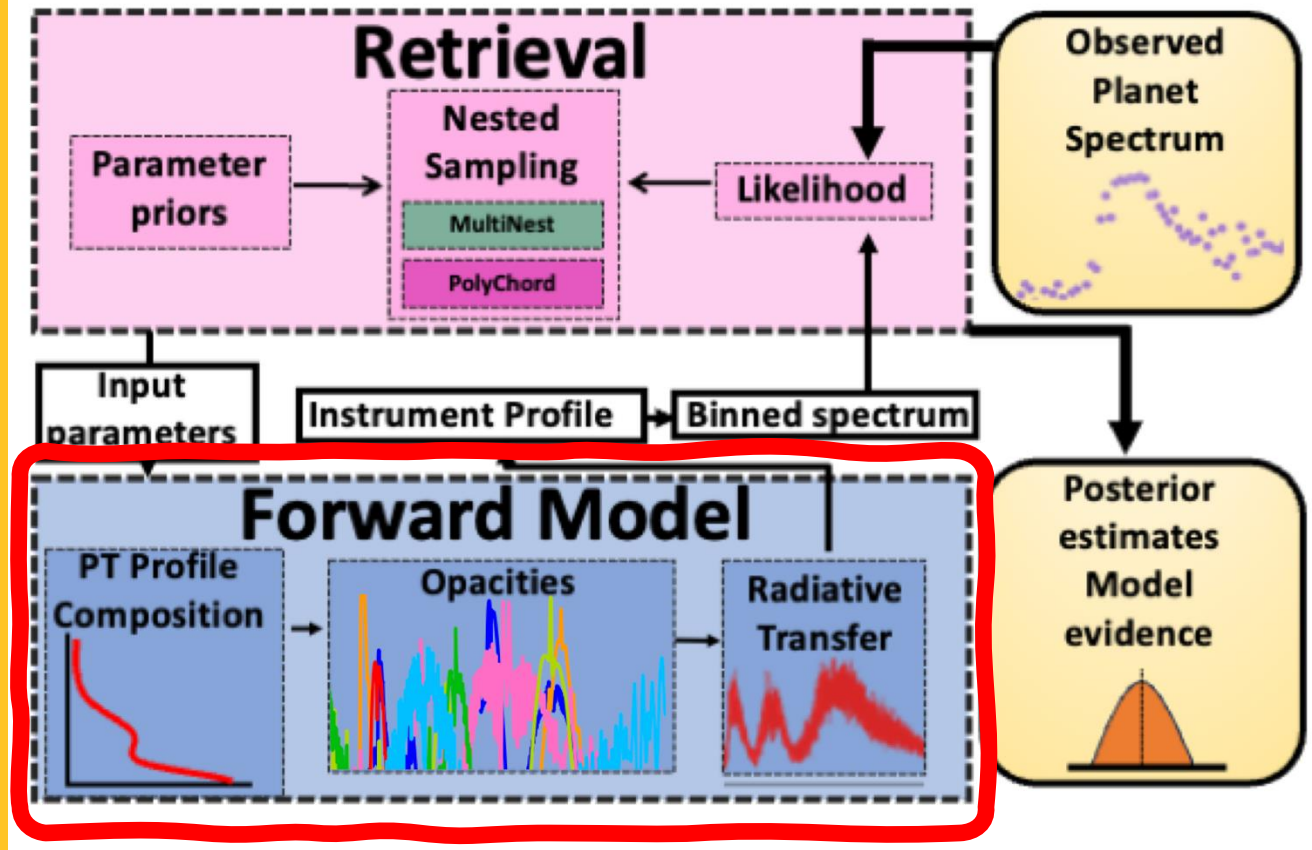


Welbanks 2021

- To test our theories we need constraints on the **chemical composition, vertical/horizontal temperature structure, aerosol properties**

- We use **Bayesian parameter estimation** involving a **large ensemble** of models (i.e., running radiative transfer thousands or millions of times) to explore the parameter space that can explain a set of observations.

Remote Sensing of Exoplanets 101

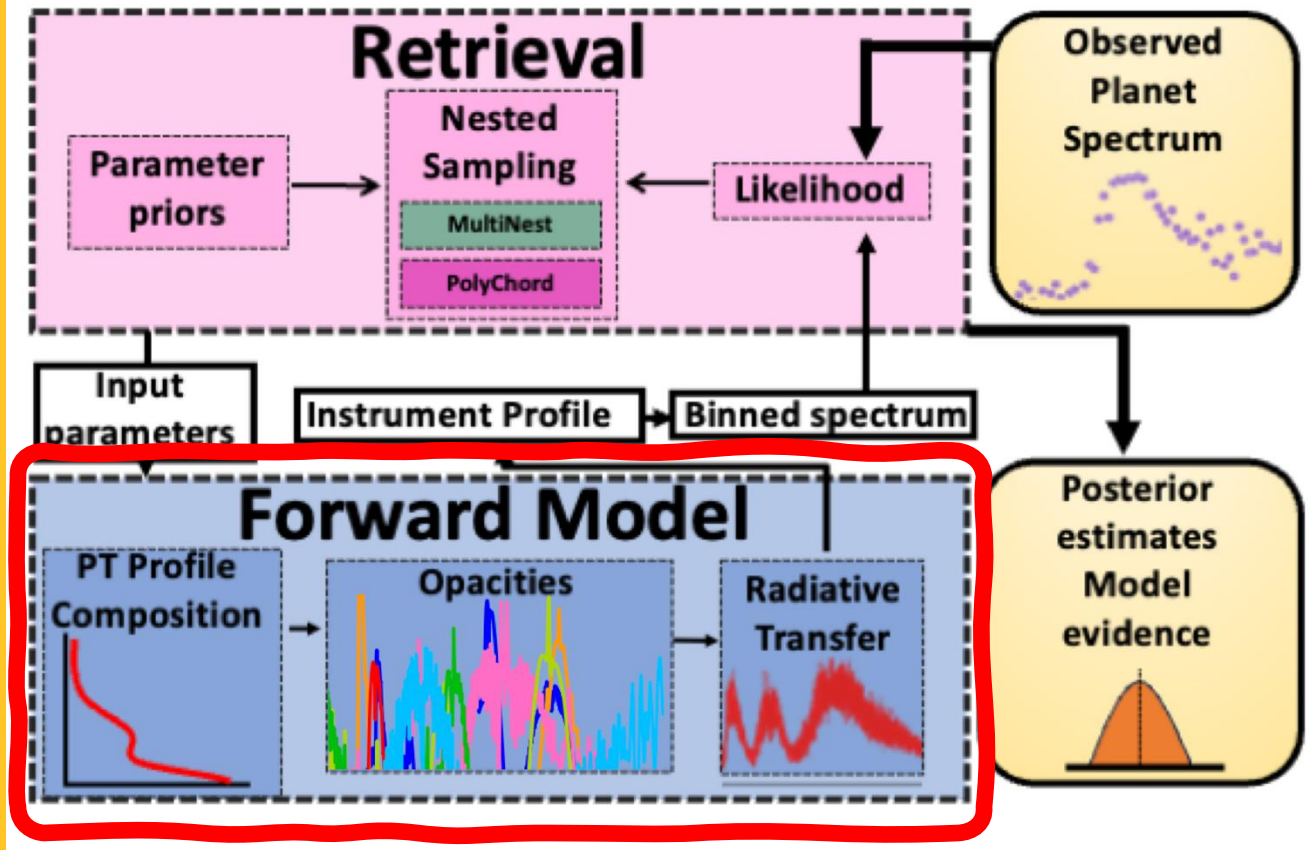


Welbanks 2021

- To test our theories we need constraints on the **chemical composition, vertical/horizontal temperature structure, aerosol properties**

- We use **Bayesian parameter estimation** involving a **large ensemble** of models (i.e., running radiative transfer thousands or millions of times) to explore the parameter space that can explain a set of observations.

Remote Sensing of Exoplanets 101

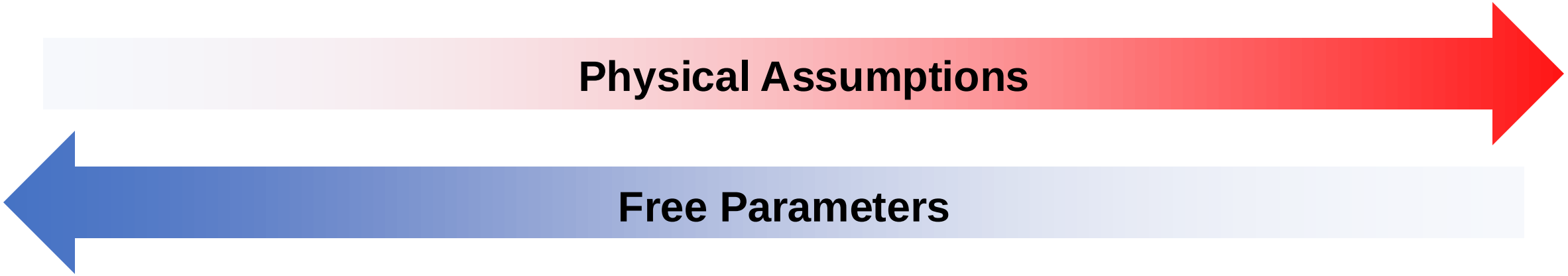


Welbanks 2021

Key points:

- Any data-model **interpretation** is only as good as the **model assumptions** themselves.
- Most interpretations from retrievals rely on the ‘art’ of **model comparison**

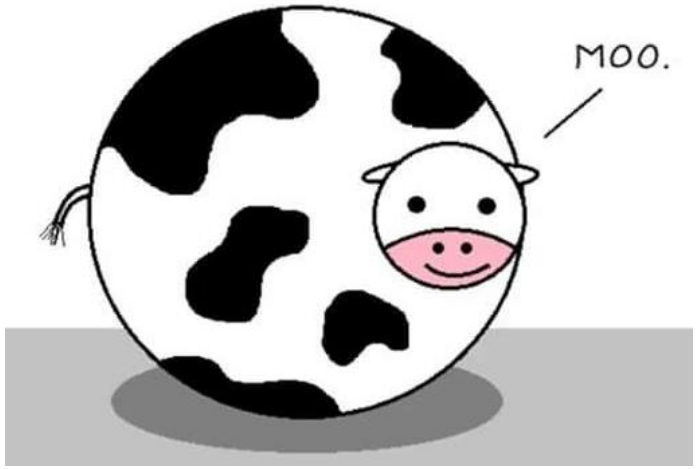
Different Modelling Strategies



Different Modelling Strategies

Physical Assumptions

Free Parameters



Modelling choices are largely ‘justified’ by the quality of our data

The ‘retrieval modeler’ evolution

Isothermal atmospheres

Cloud-free atmospheres

Single absorber atmospheres

Planets are 1D

Isobaric cross-sections

…

Stellar atmosphere is homogeneous

Thermal structure matters!

Clouds affect our interpretations!

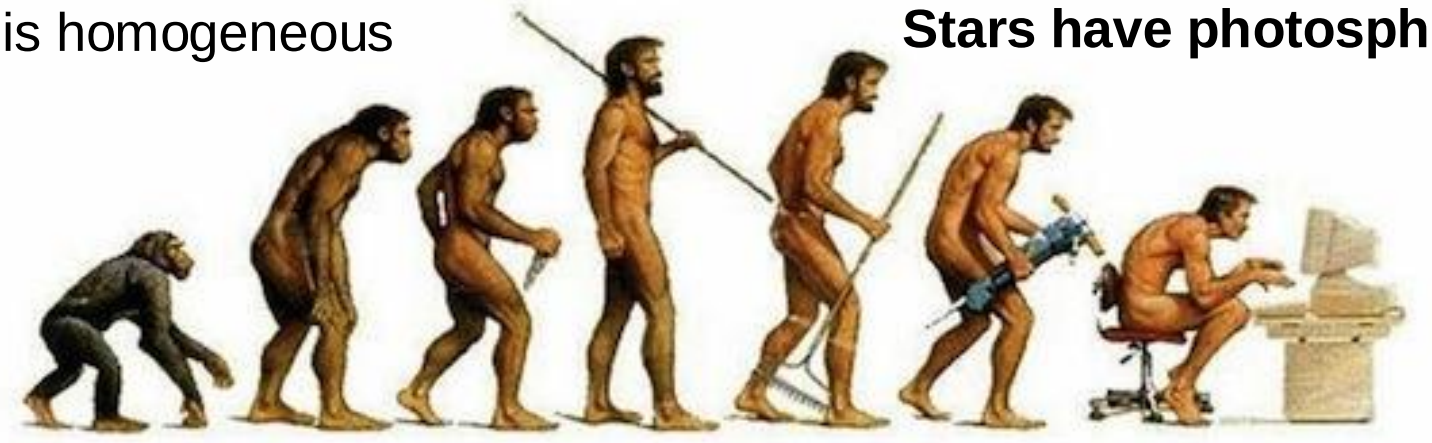
Need multiple absorbers!

Planets are *not* 1D

Cross-sections choices matter

…

Stars have photospheric heterogeneities



Model assumptions matter see e.g., Madhusudhan & Seager 2009, Fortney et al 2010, Benneke & Seager 2012, 2013, De Wit & Seager 2013, Line & Parmentier 2016, Betremieux 2016, Betremieux & Swain 2017, 2018, Heng & Kitzmann 2017, MacDonald & Madhusudhan 2017, Welbanks & Madhusudhan 2019, Caldas et al. 2019, Pluriel et al. 2020, Lacy & Burrows 2020, MacDonald et al. 2020, Espinoza & Jones 2021, Pluriel et al. 2022, Welbanks et al. 2022, Barstow 2020, and many more…

A retrieval paradigm for inferring both planetary and stellar properties from a transmission spectrum

Spots and Faculae in the star contaminate the transmission spectrum – Rackham + 18

A point acknowledged for over a decade see e.g., Pont+08, Sing+11, …

Pinhas, Rackham, Madhusudhan & Apai 2018 offer **a way to "retrieve" these effects**

A retrieval paradigm for inferring both planetary and stellar properties from a transmission spectrum

Spots and Faculae in the star contaminate the transmission spectrum – Rackham + 18

A point acknowledged for over a decade see e.g., Pont+08, Sing+11, …

Pinhas, Rackham, Madhusudhan & Apai 2018 offer **a way to "retrieve" these effects**

$$\underbrace{\left(\frac{\mathcal{R}_p(\lambda)}{\mathcal{R}_\star(\lambda)}\right)^2}_{\text{Observed spectrum}} = \underbrace{\left(\frac{R_p(\lambda)}{R_\star}\right)^2}_{\text{Planetary spectrum}} \underbrace{\left(1-\delta\left(1-\frac{\mathcal{S}_u(\lambda)}{\mathcal{S}_o(\lambda)}\right)\right)^{-1}}_{\text{Stellar contamination}}$$

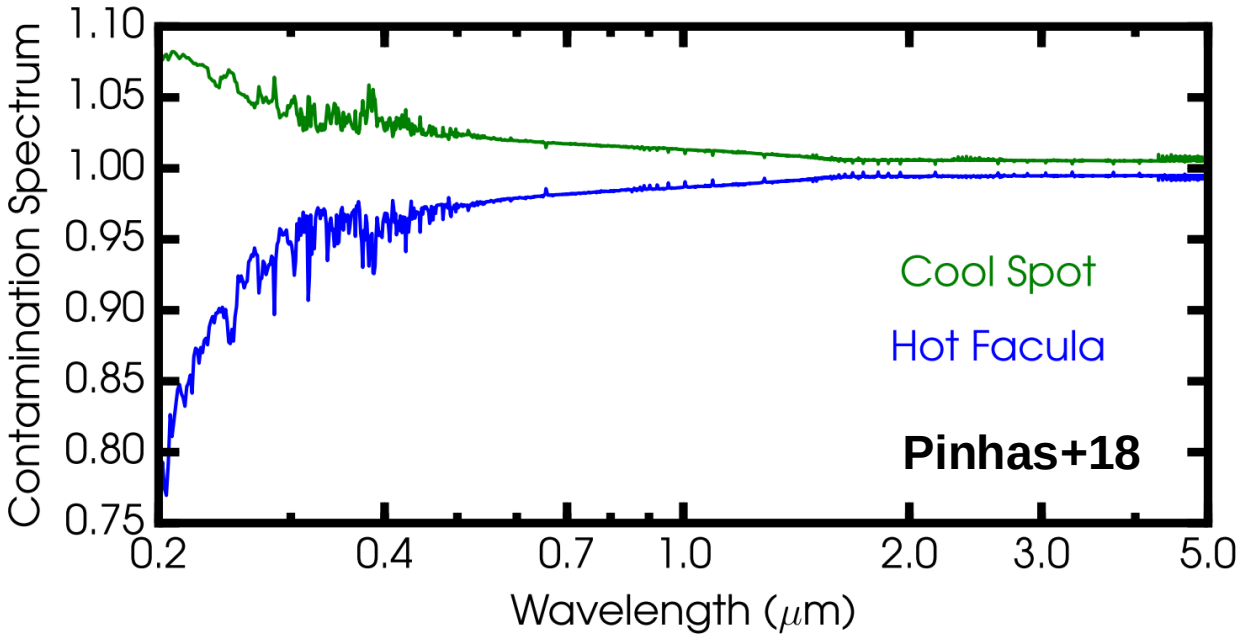
A retrieval paradigm for inferring both planetary and stellar properties from a transmission spectrum

Spots and Faculae in the star contaminate the transmission spectrum – Rackham + 18

A point acknowledged for over a decade see e.g., Pont+08, Sing+11, …

Pinhas, Rackham, Madhusudhan & Apai 2018 offer **a way to "retrieve" these effects**

$$\underbrace{\left(\frac{\mathcal{R}_p(\lambda)}{\mathcal{R}_\star(\lambda)}\right)^2}_{\text{Observed spectrum}} = \underbrace{\left(\frac{R_p(\lambda)}{R_\star}\right)^2}_{\text{Planetary spectrum}} \underbrace{\left(1-\delta\left(1-\frac{\mathcal{S}_u(\lambda)}{\mathcal{S}_o(\lambda)}\right)\right)^{-1}}_{\text{Stellar contamination}}$$



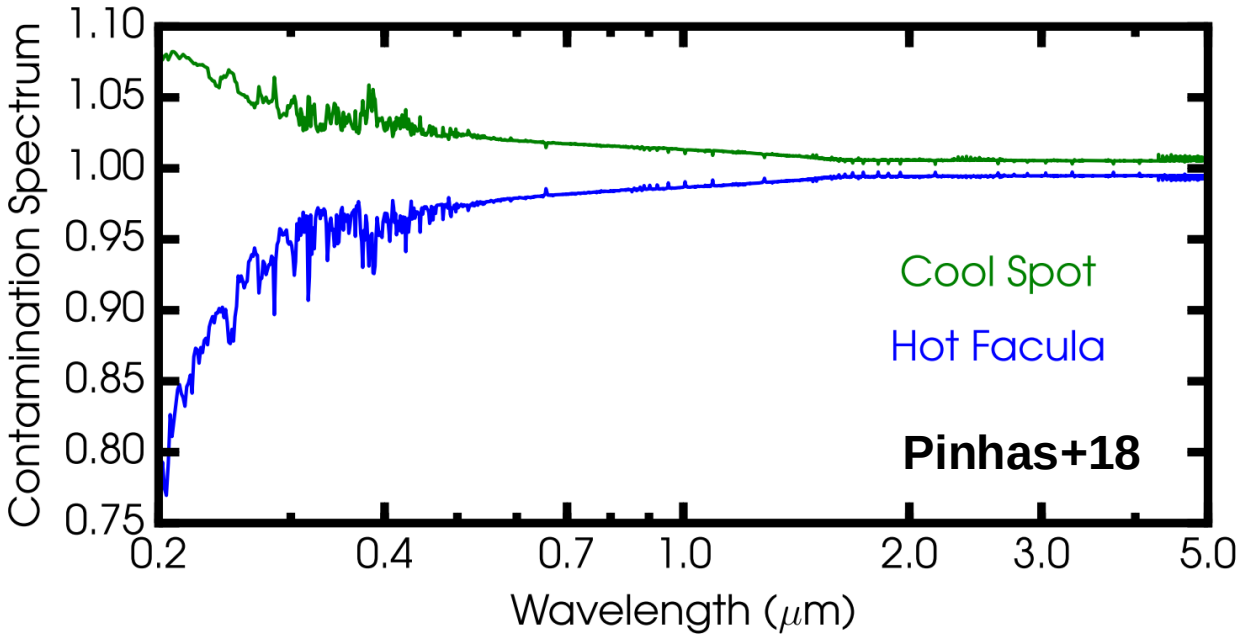
A retrieval paradigm for inferring both planetary and stellar properties from a transmission spectrum

Spots and Faculae in the star contaminate the transmission spectrum – Rackham + 18

A point acknowledged for over a decade see e.g., Pont+08, Sing+11, …

Pinhas, Rackham, Madhusudhan & Apai 2018 offer **a way to "retrieve" these effects**

$$\underbrace{\left(\frac{\mathcal{R}_p(\lambda)}{\mathcal{R}_\star(\lambda)}\right)^2}_{\text{Observed spectrum}} = \underbrace{\left(\frac{R_p(\lambda)}{R_\star}\right)^2}_{\text{Planetary spectrum}} \underbrace{\left(1-\delta\left(1-\frac{\mathcal{S}_u(\lambda)}{\mathcal{S}_o(\lambda)}\right)\right)^{-1}}_{\text{Stellar contamination}}$$



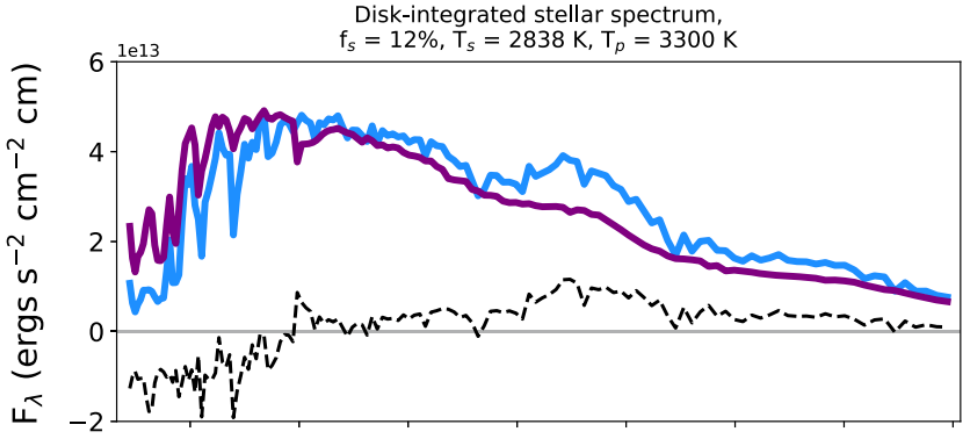
Use stellar atmospheric models (e.g., PHOENIX Husser+13), model the spectral components, retrieve stellar and planet properties.

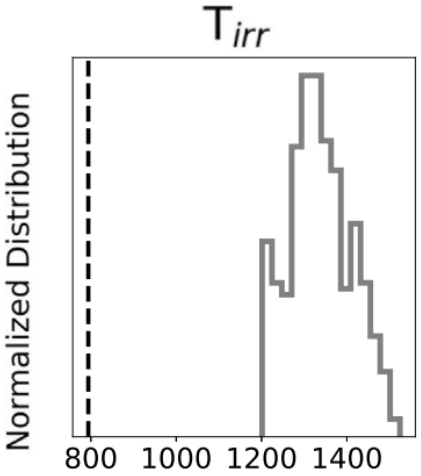
Excellent! We have an infallible method to correct our...

We note that joint retrievals of the stellar photosphere and exoplanetary atmosphere rely on the assumptions that the model components are reasonably correct and reasonably complete. Presently, however, even state-of-the-art models for stellar heterogeneity are based on very limited knowledge; furthermore, the hypothesized haze particles can provide similar spectral signatures in the optical. Thus, future efforts must collect better data to break the degeneracy between stellar contamination and possible atmospheric aerosols.

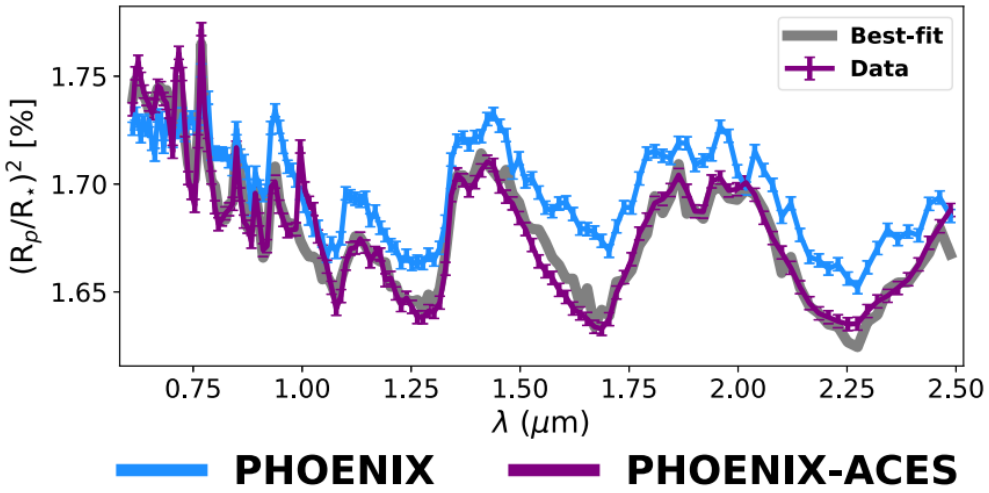
Pinhas+18

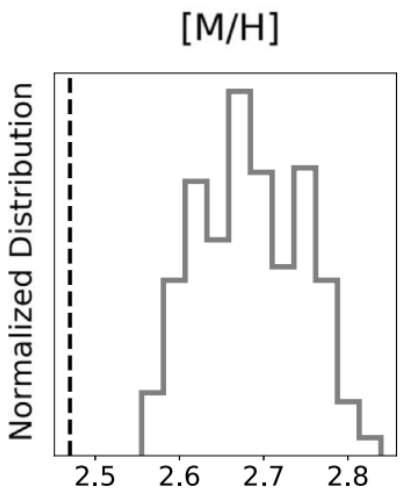
Biases in planet parameters persist despite all recommended combinations of corrections and retrievals—all mainly limited by stellar model fidelity

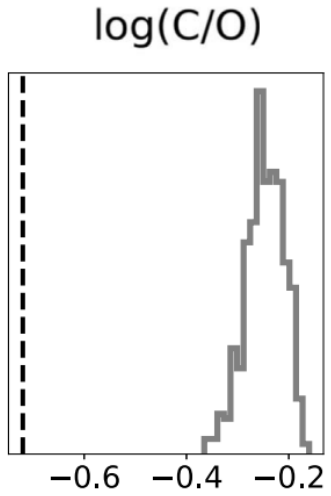




Iyer & Line 2020







Strong biases on the atmospheric properties that can occur when an inadequate stellar photosphere model is used for the spot correction

Our ability to infer atmospheric compositions in the presence of stellar activity is model-limited

TLSE treatments in retrievals rely on the fidelity of stellar atmosphere model grids.

Improper understanding of clouds/convection, incomplete molecular line-lists, correlated noise sources are a key bottleneck in modeling M-dwarf spectra (e.g., Iyer+23).

Our TLSE treatment is degenerate with spectral signatures of aerosols in the planet atmosphere.

Difficult to determine the complexity in the stellar contamination model (e.g., spots or faculae or both) and planetary model a priori. Model comparisons needed.

Our ability to infer atmospheric compositions in the presence of stellar activity is model-limited

TLSE treatments in retrievals rely on the fidelity of stellar atmosphere model grids.

Improper understanding of clouds/convection, incomplete molecular line-lists, correlated noise sources are a key bottleneck in modeling M-dwarf spectra (e.g., Iyer+23).

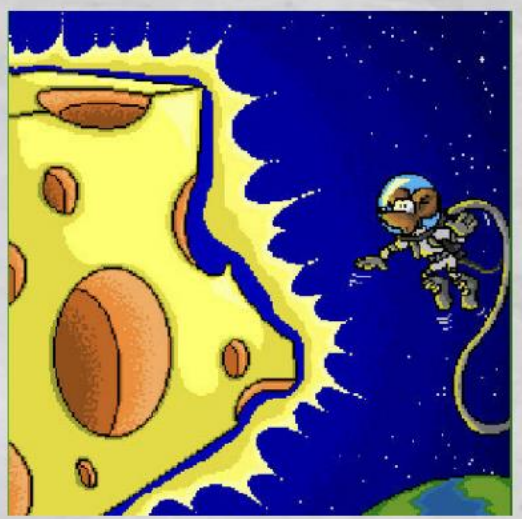
Our TLSE treatment is degenerate with spectral signatures of aerosols in the planet atmosphere.

Difficult to determine the complexity in the stellar contamination model (e.g., spots or faculae or both) and planetary model a priori. Model comparisons needed.

Astronomers find 6-sigma detection of cheese on the Moon!

By JOHN DOE

Why are you reading this fake news article? This is not part of the presentation. What I am saying should be more interesting... or so I hope



Reuters

The Daily Retrieval

15 JUL 2022

This 6-sigma detection of cheese could happen if you try to explain the composition of the moon with a cheese model and a sponge model

We must be careful when performing model comparisons.

Bayesian evidence is not the be-all and end-all

How do we assess what in the data drive our inferences?

Determine the **limits of our data and models** see e.g., LOO cross-validation Welbanks+23

How do we ensure our models are *'good'*?

Explore the **continuum of model complexity** and implement physical guardrails see e.g., ongoing efforts REVEAL, SPOTLESS, etc.

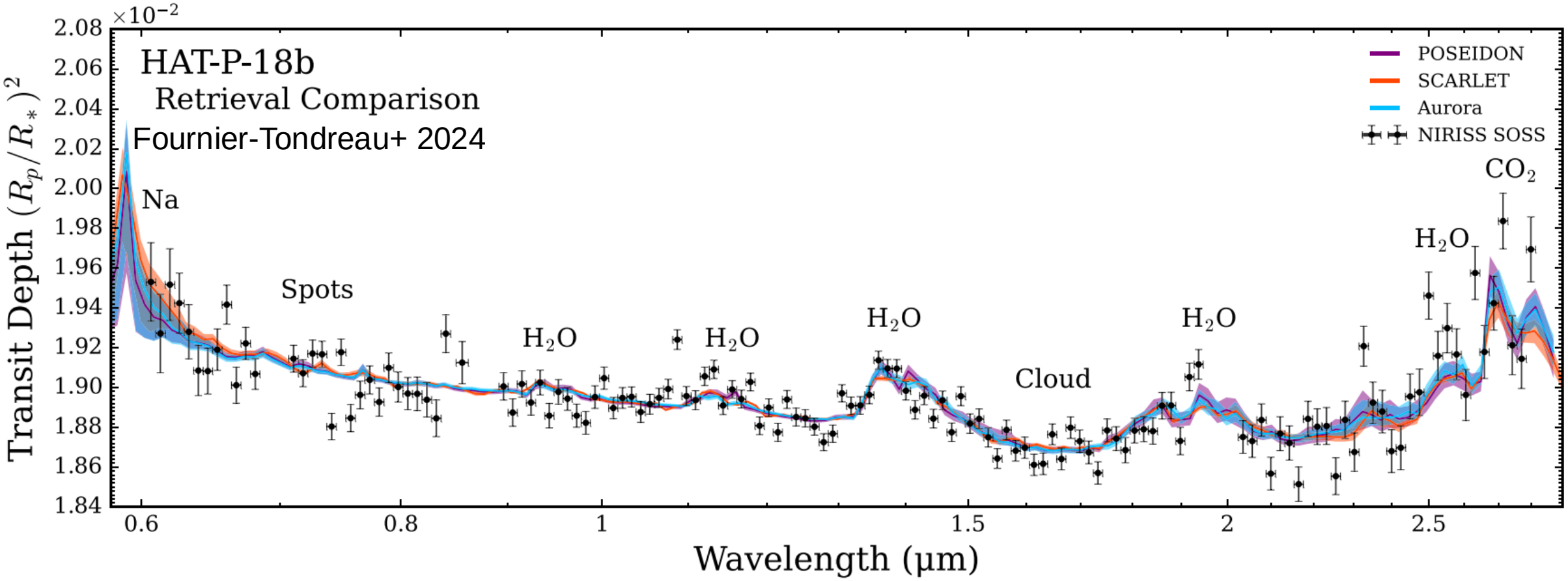
How do we prevent falling into a trap of comparing two '*bad*' models?

Provide **context to our claims** and acknowledge the limits of our data and models.

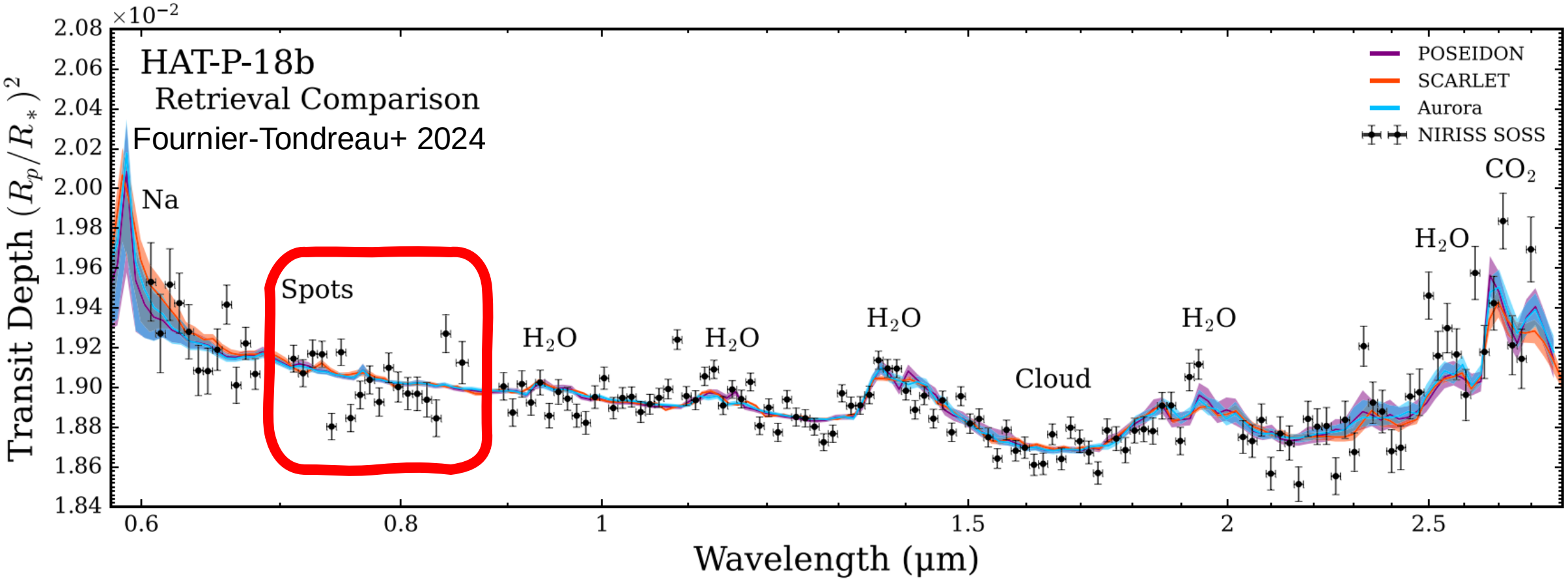
Early Lessons with JWST

1. **Stellar surface heterogeneities are ubiquitous in JWST data. Our interpretation is model-dependent.**

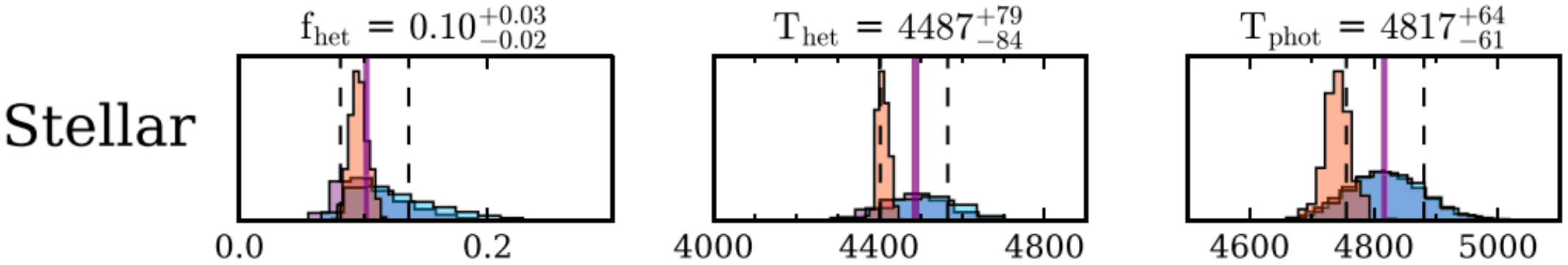
Early Release Observations



Early Release Observations



Early Release Observations

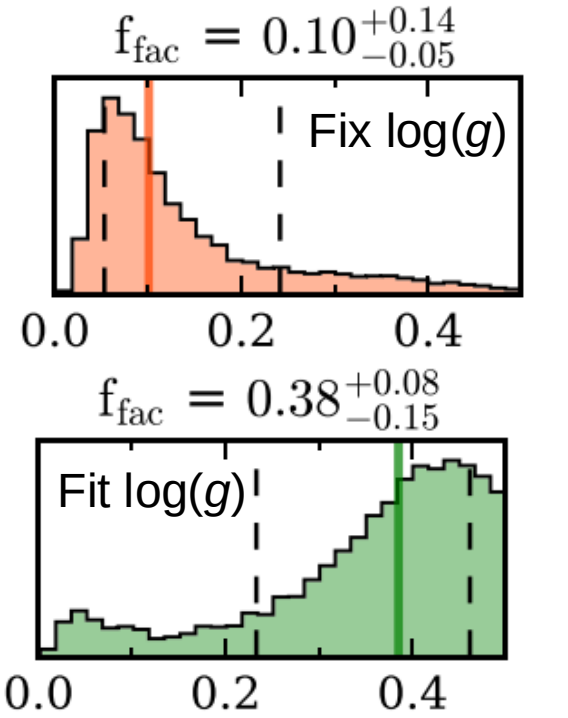


Fournier-Tondreau+ 2024

Same data, same set of atmospheric model assumptions, different choices in TLSE treatment

Inferred properties can still be different between retrieval codes

Early Release Observations



Fournier-Tondreau+ 2024

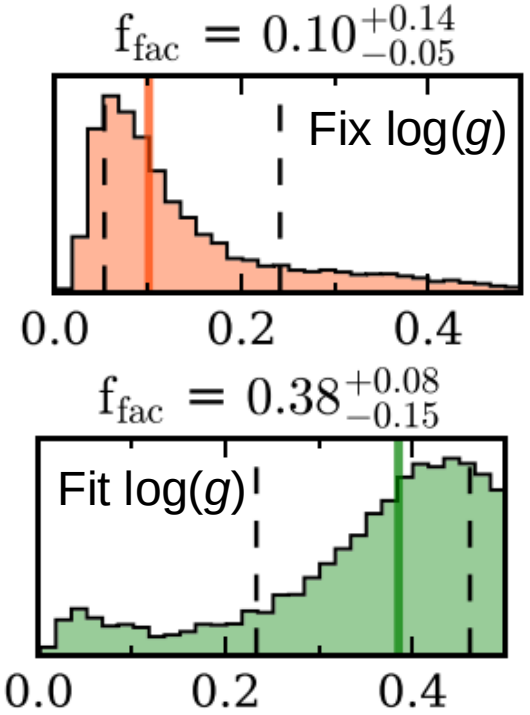
TLSE models of increasing complexity:

- Stellar spots only
- Stellar spots and stellar faculae

fit for the surface gravity of the active regions (proxy for magnetic pressure?).

Star and planet **inferences change with model assumptions**.

Early Release Observations



Fournier-Tondreau+ 2024

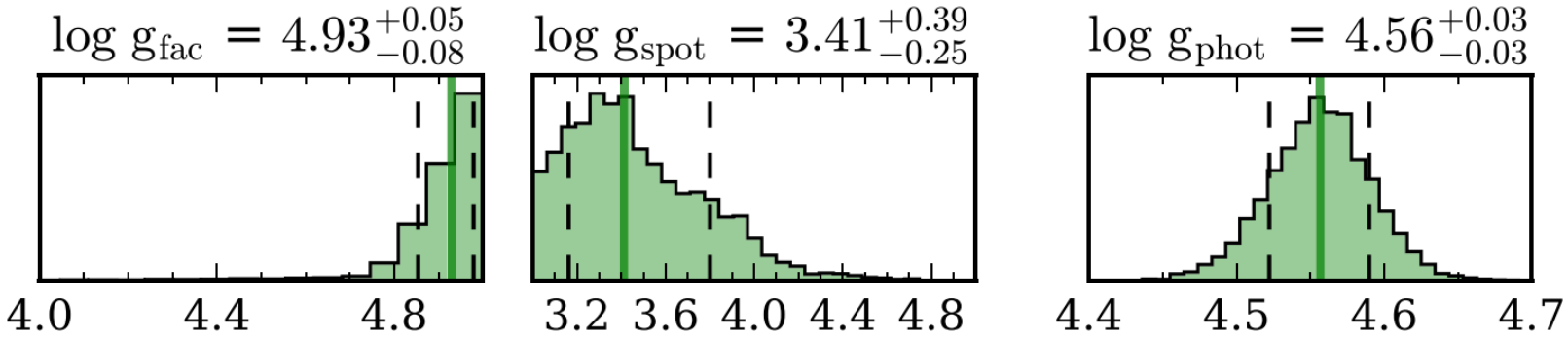
TLSE models of increasing complexity:

- Stellar spots only
- Stellar spots and stellar faculae

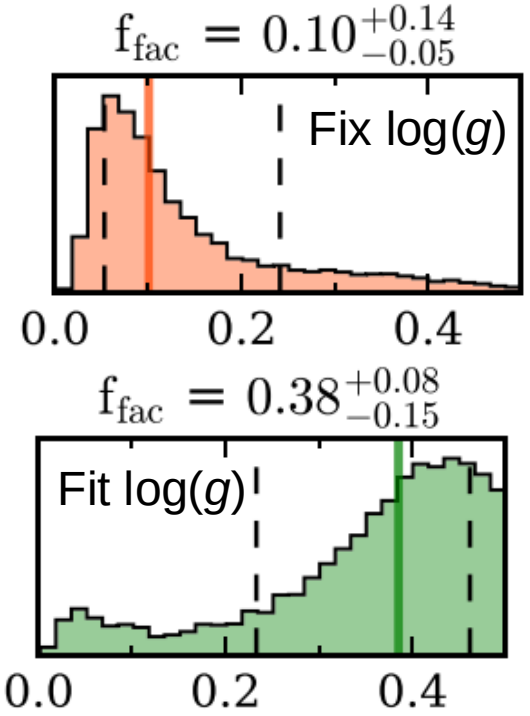
fit for the surface gravity of the active regions (proxy for magnetic pressure?).

Star and planet **inferences change with model assumptions**.

For HAT-P-18b, the model with two heterogeneities and different gravities between the spots, faculae, and photosphere is preferred.



Early Release Observations



Fournier-Tondreau+ 2024

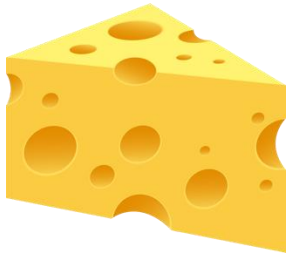
TLSE models of increasing complexity:

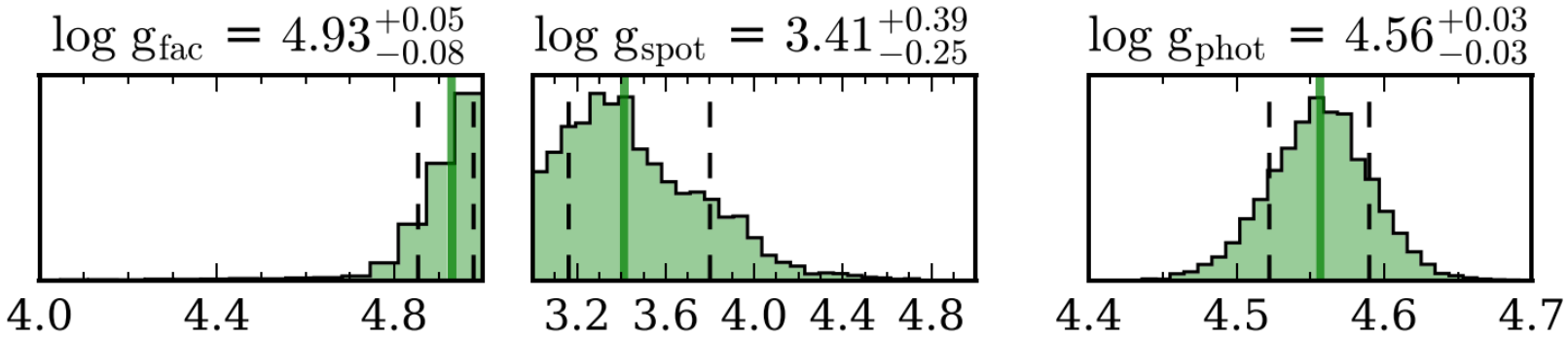
- Stellar spots only
- Stellar spots and stellar faculae

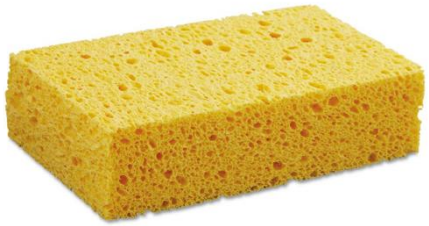
fit for the surface gravity of the active regions (proxy for magnetic pressure?).

Star and planet **inferences change with model assumptions**.

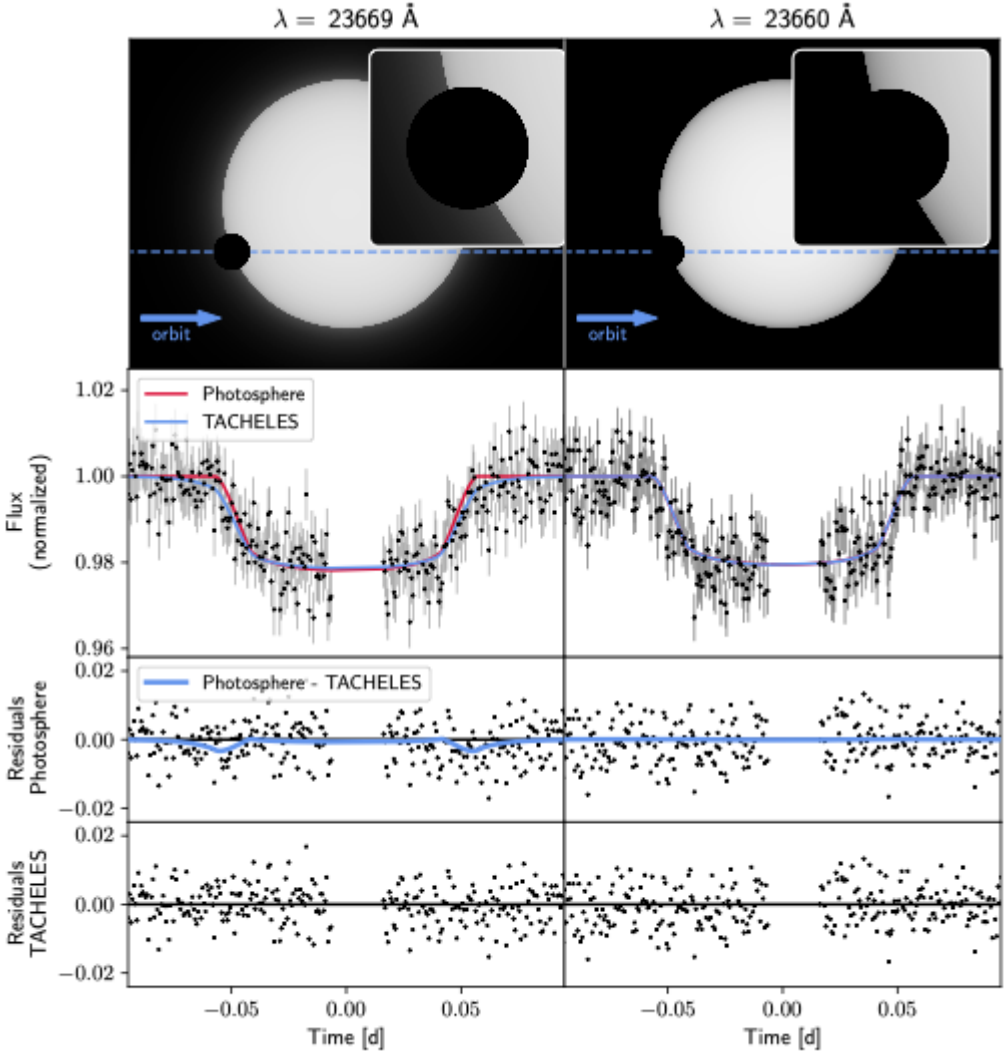
For HAT-P-18b, the model with two heterogeneities and different gravities between the spots, faculae, and photosphere is preferred.







Same pixels - Different answers



New result: Perdelwitz et al. (2025)

Consider the effects of chromosphere and a corona when performing a data reduction.

This previously overlooked consideration in the data leads to different results in our inferences.

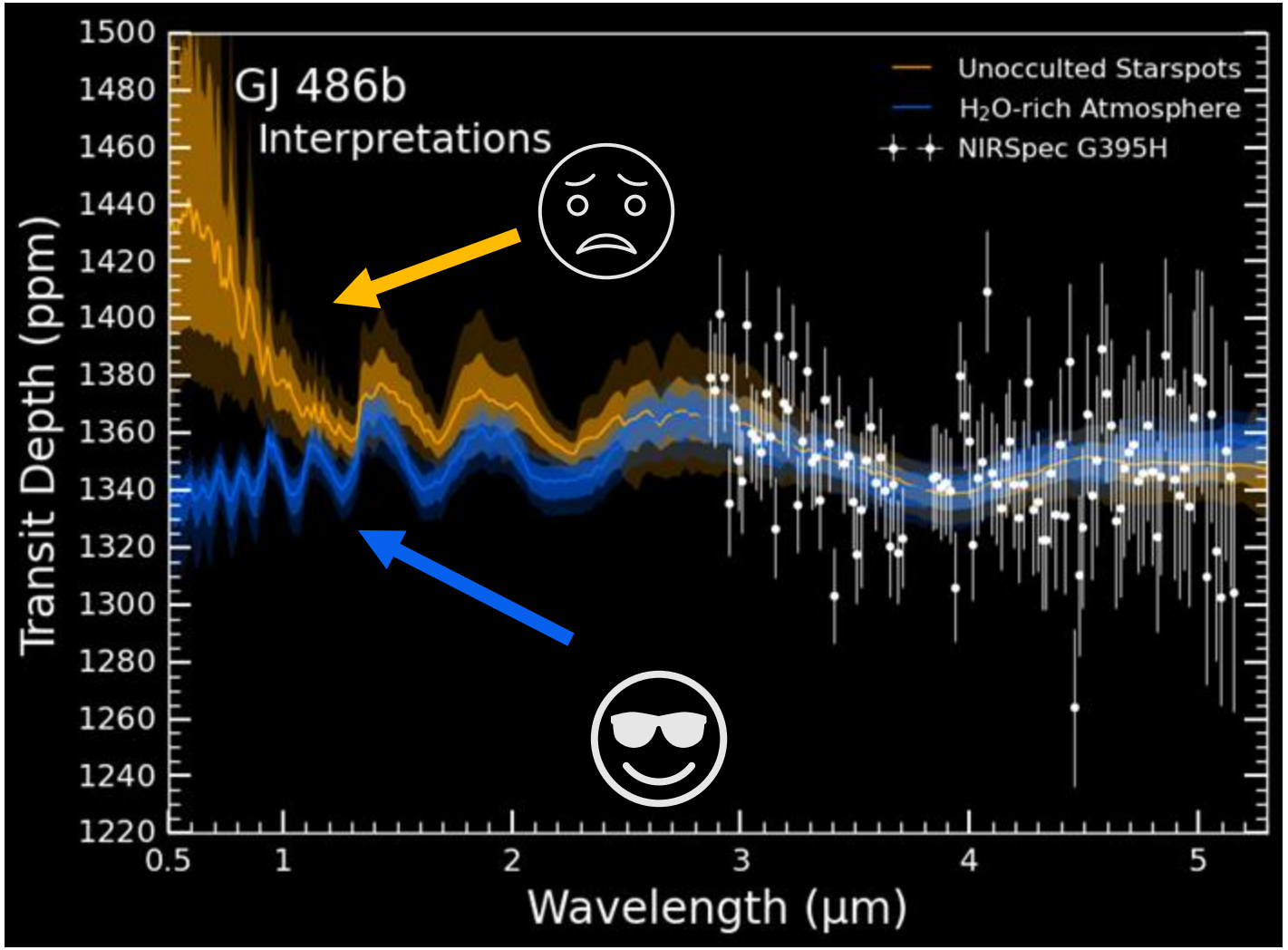
Same pixels → Different answer

Should we account for the effects 'chromospheric/coronal' contamination in our models?

Early Lessons with JWST

1. Stellar surface heterogeneities are ubiquitous in JWST data. Our interpretation is model-dependent.
2. **Stellar surface heterogeneities limit our ability to infer the nature of exoplanetary atmospheres. Our interpretation is model-limited.**

A water-rich atmosphere?



The NIRSpec/G395 data prefers a Gaussian-like feature at 2-3σ

The spectrum can be equally explained by a water-rich atmosphere or unocculted starspots

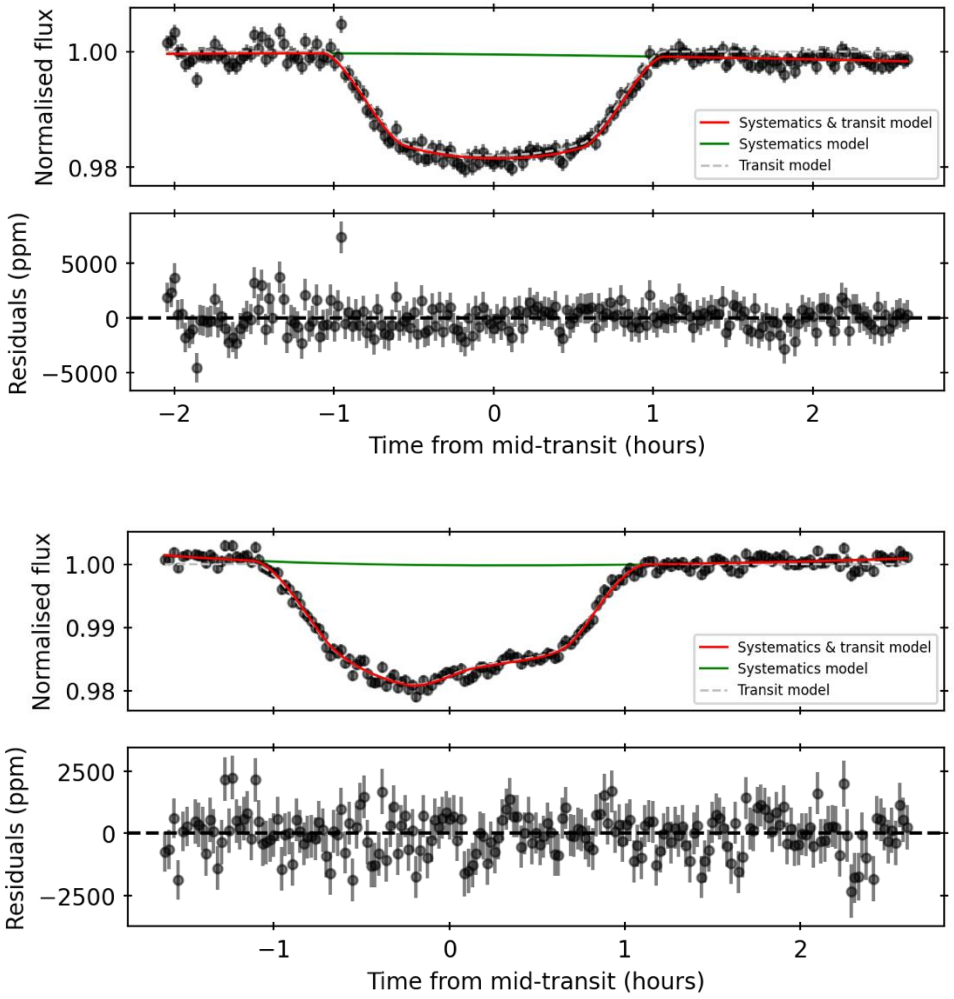
Forthcoming NIRISS observations should differentiate between these two scenarios (GO5866: PI MacDonald)

Moran & Stevenson et al. (2023) –Material courtesy of Kevin Stevenson & Ryan MacDonald

Early Lessons with JWST

1. Stellar surface heterogeneities are ubiquitous in JWST data. Our interpretation is model-dependent.
2. Stellar surface heterogeneities limit our ability to infer the nature of exoplanetary atmospheres. Our interpretation is model-limited.
3. **The TLSE is stochastic in nature. Planning against it is difficult. Need multiple visits?**

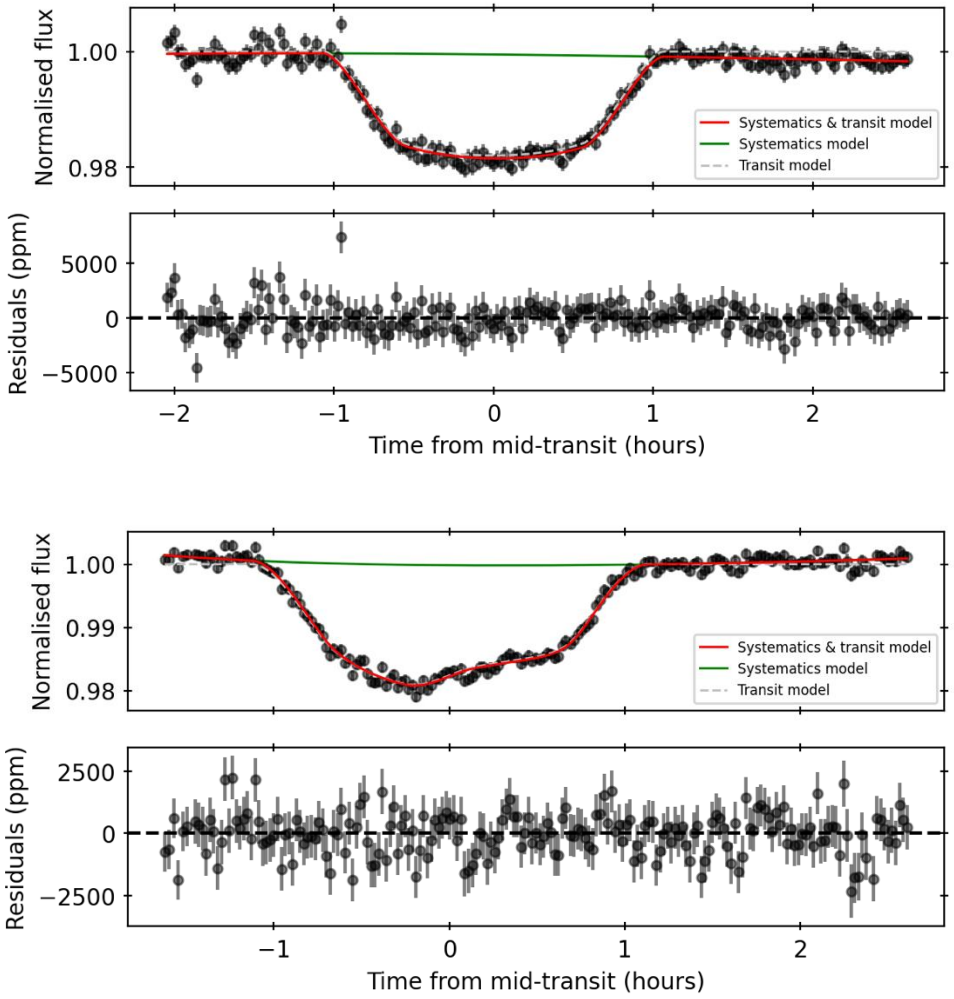
LRG-BEASTS: Combining visits is an untamed beast



Two transits
1 month apart

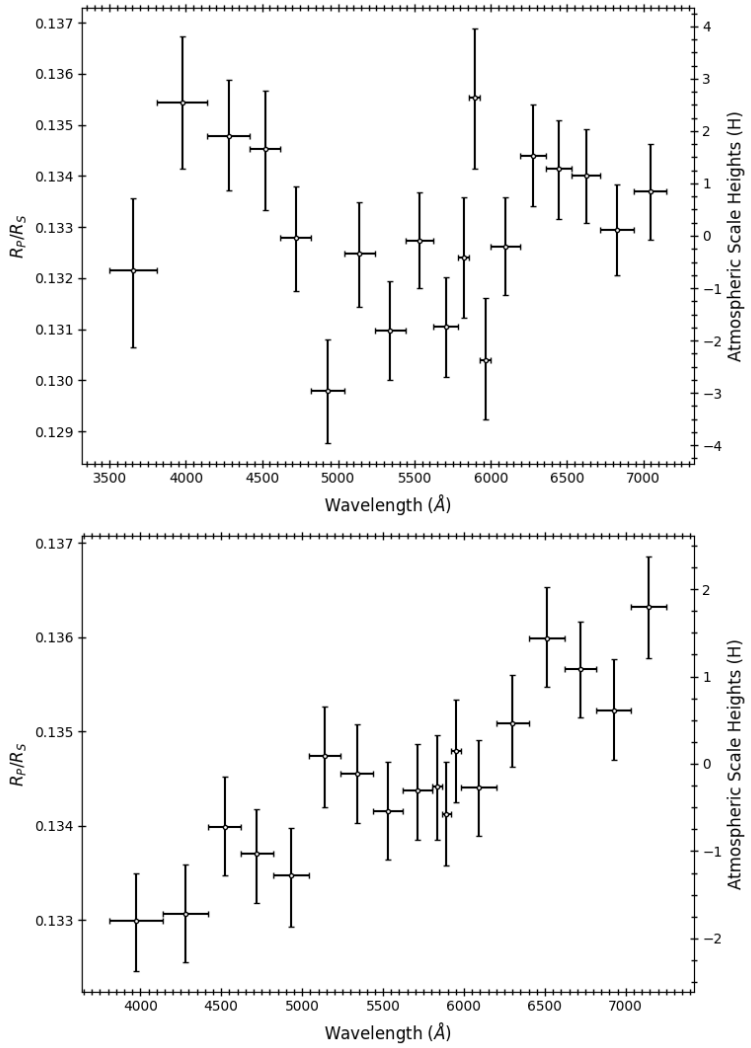
From Staudt et al. (in prep) courtesy of Eva-Maria Ahrer

LRG-BEASTS: Combining visits is an untamed beast



Two transits 1 month apart

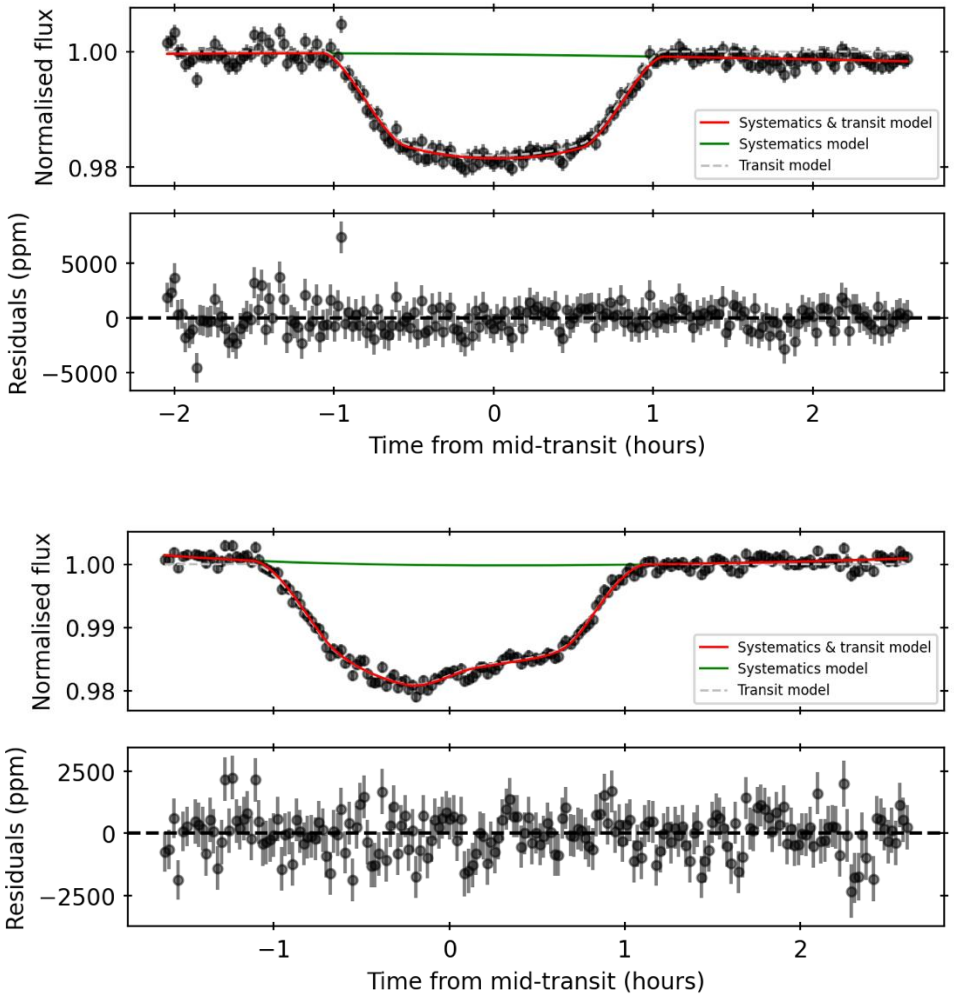
Two very different spectra



From Staudt et al. (in prep) courtesy of Eva-Maria Ahrer

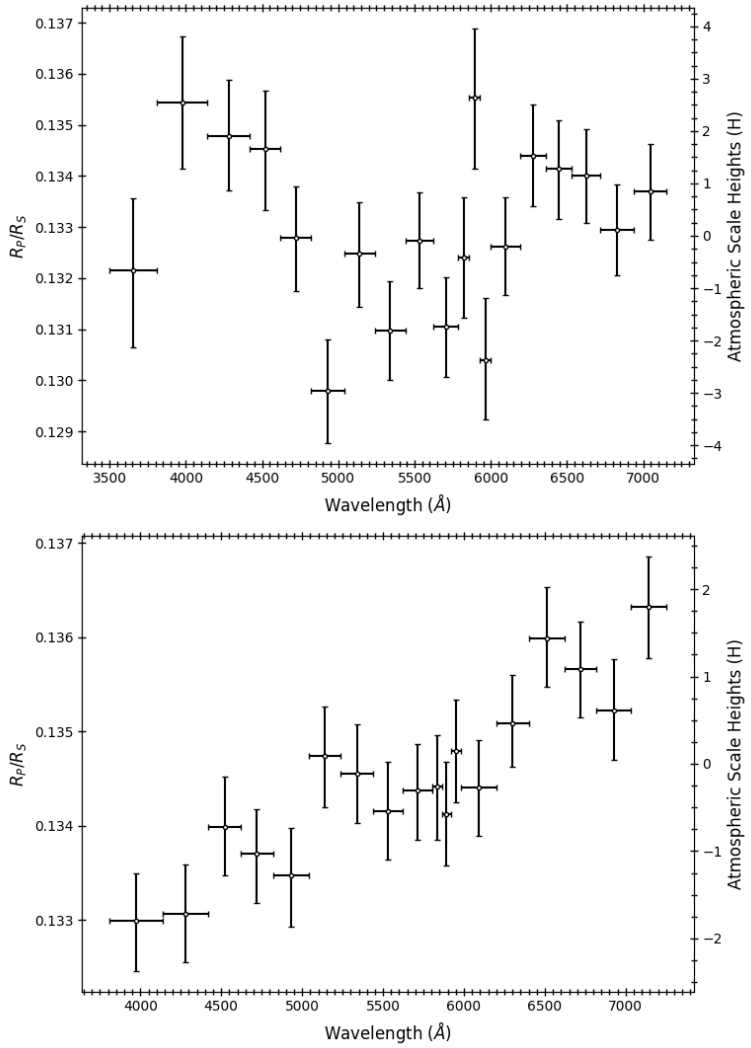
See also Radica+ 2024 Trappist-1c, May+2023 GJ1132b

LRG-BEASTS: Combining visits is an untamed beast



Two transits 1 month apart

Two very different spectra



From Staudt et al. (in prep) courtesy of Eva-Maria Ahrer

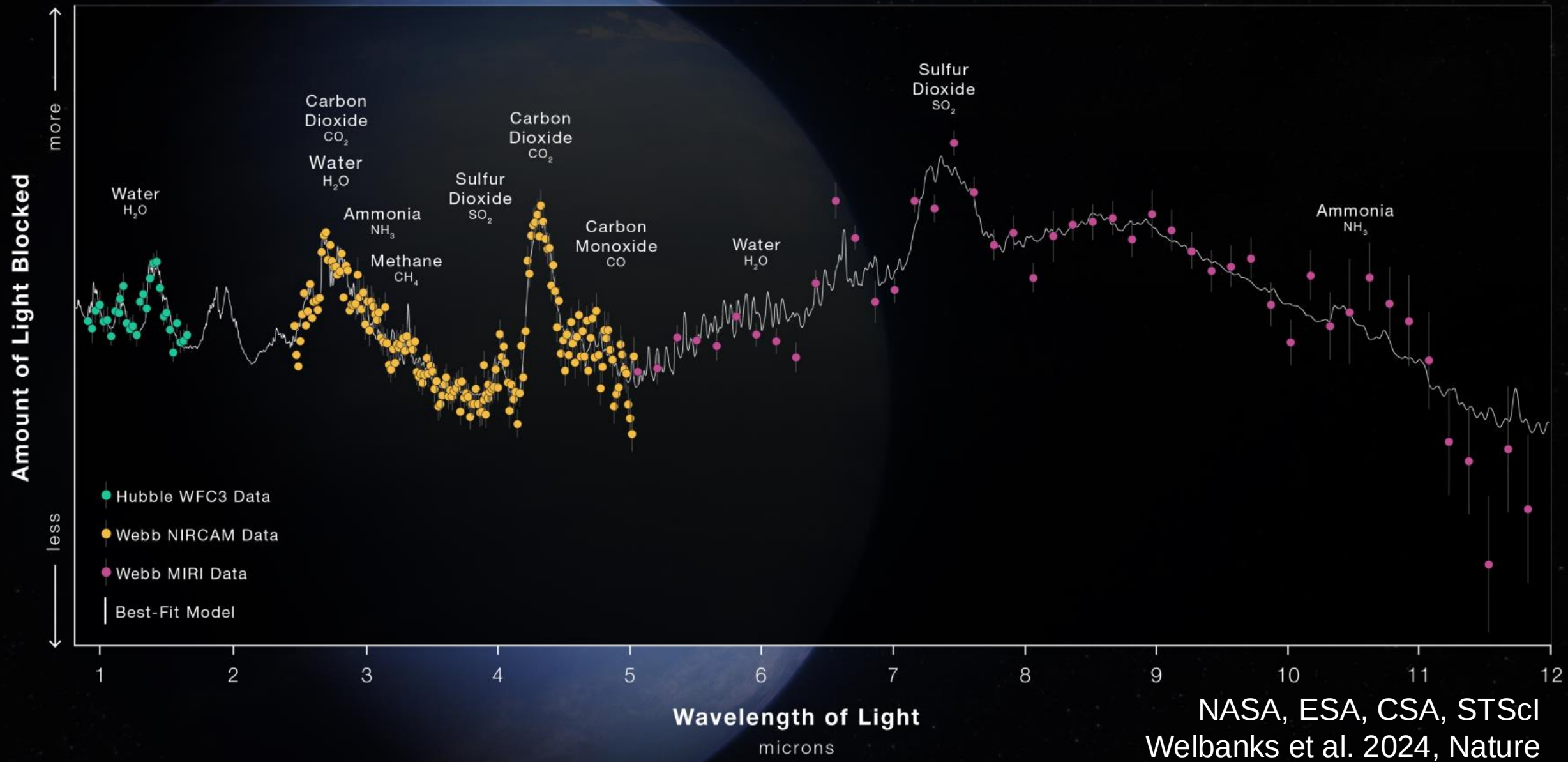
See also Radica+ 2024 Trappist-1c, May+2023 GJ1132b

How do we go about combining observations? How do we tease the effects of aerosols from the effects of stellar heterogeneities?

Early Lessons with JWST

1. Stellar surface heterogeneities are ubiquitous in JWST data. Our interpretation is model-dependent.
2. Stellar surface heterogeneities limit our ability to infer the nature of exoplanetary atmospheres. Our interpretation is model-limited.
3. The TLSE is stochastic in nature. Planning against it is difficult. Need multiple visits?
4. **Star spots are degenerate with claims of limb-asymmetries.**

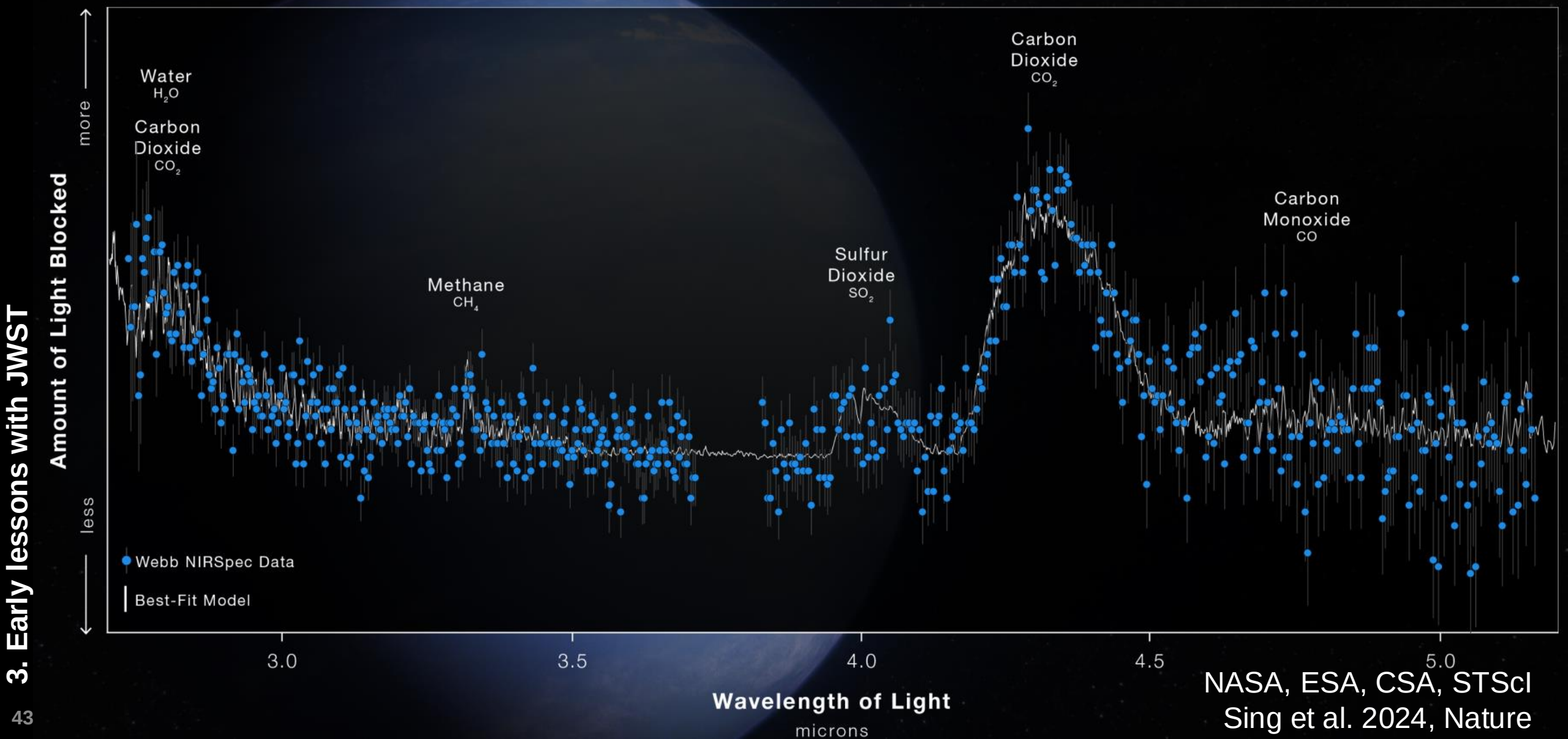
WASP-107b: The story of two programs



NASA, ESA, CSA, STScI
Welbanks et al. 2024, Nature

3. Early lessons with JWST

WASP-107b: The story of two programs



NASA, ESA, CSA, STScI
Sing et al. 2024, Nature

WASP-107b: The story of two programs

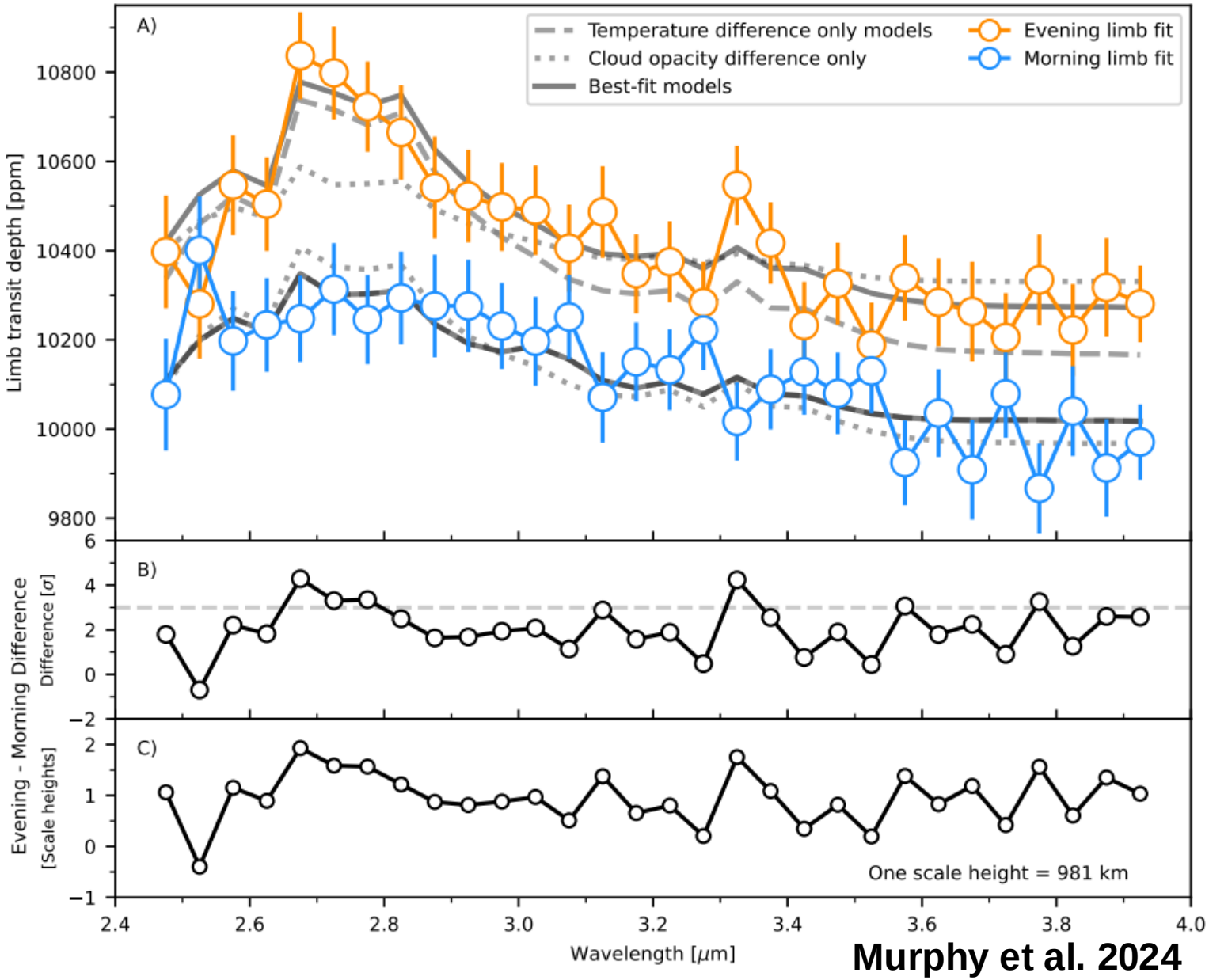
F444W taken close-in-time to the NIRSpec.
F322W2 further away in time.

F322W2/ F444W show no spot crossings. NIRSpec shows a stellar spot crossing.

Despite this, overall inferences are the same:
chemical disequilibrium, depleted methane, hot interior.

Luck?

WASP-107b: The story of two programs



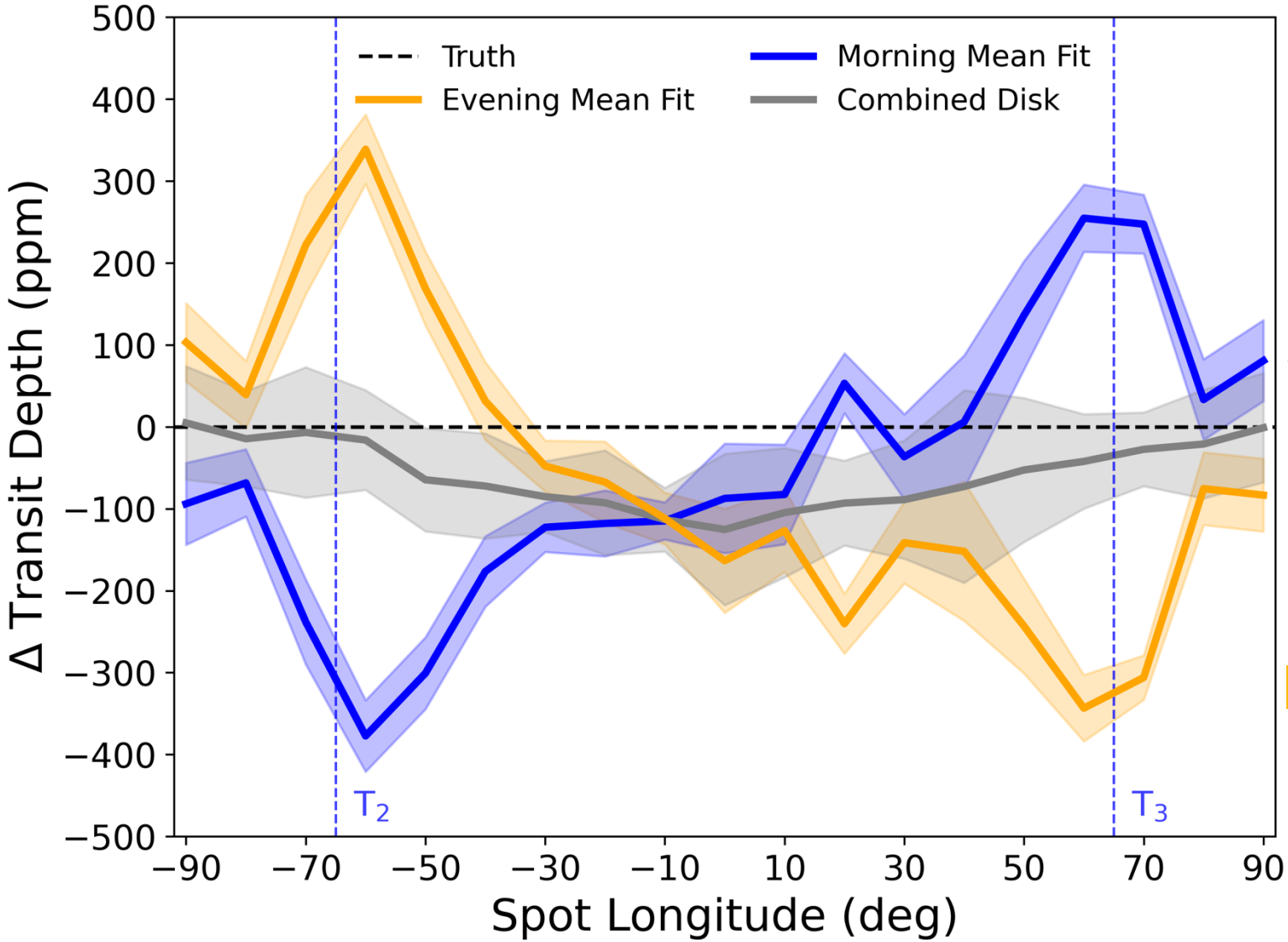
Murphy et al. 2024

Evidence for morning-to-evening limb asymmetry. NIRCam F322W2

Possible difference of ~180 K between the evening and morning limbs

Can this be confirmed using NIRSpec?

WASP-107b: The story of two programs



Occulted spots can introduce "*fake*" signatures of limb asymmetries.

Simulated example for WASP-107b.

A spot is injected at different locations along the stellar equator - the planet transits across it.

The resulting light curve is fit using Catwoman (Jones & Espinoza 2021)

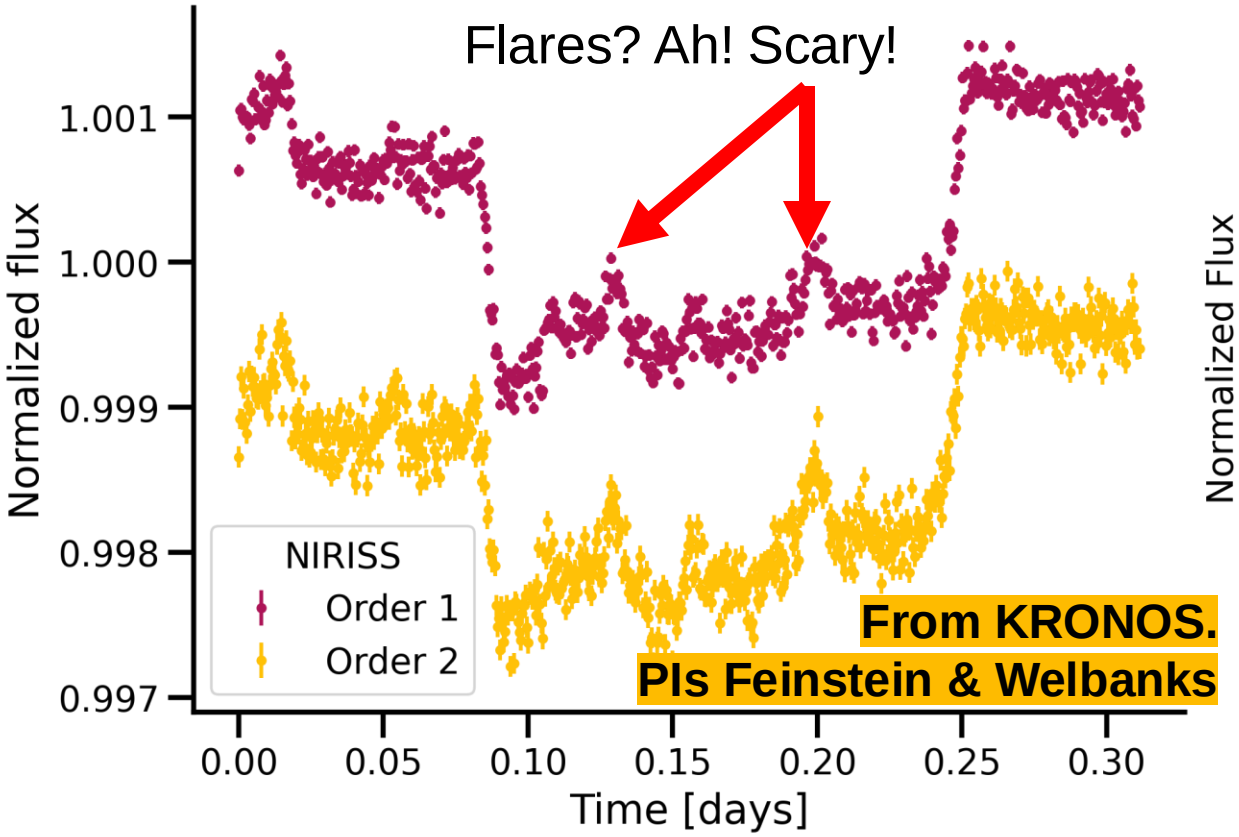
Starspot modeling is relevant for any method used to measure limb asymmetry.

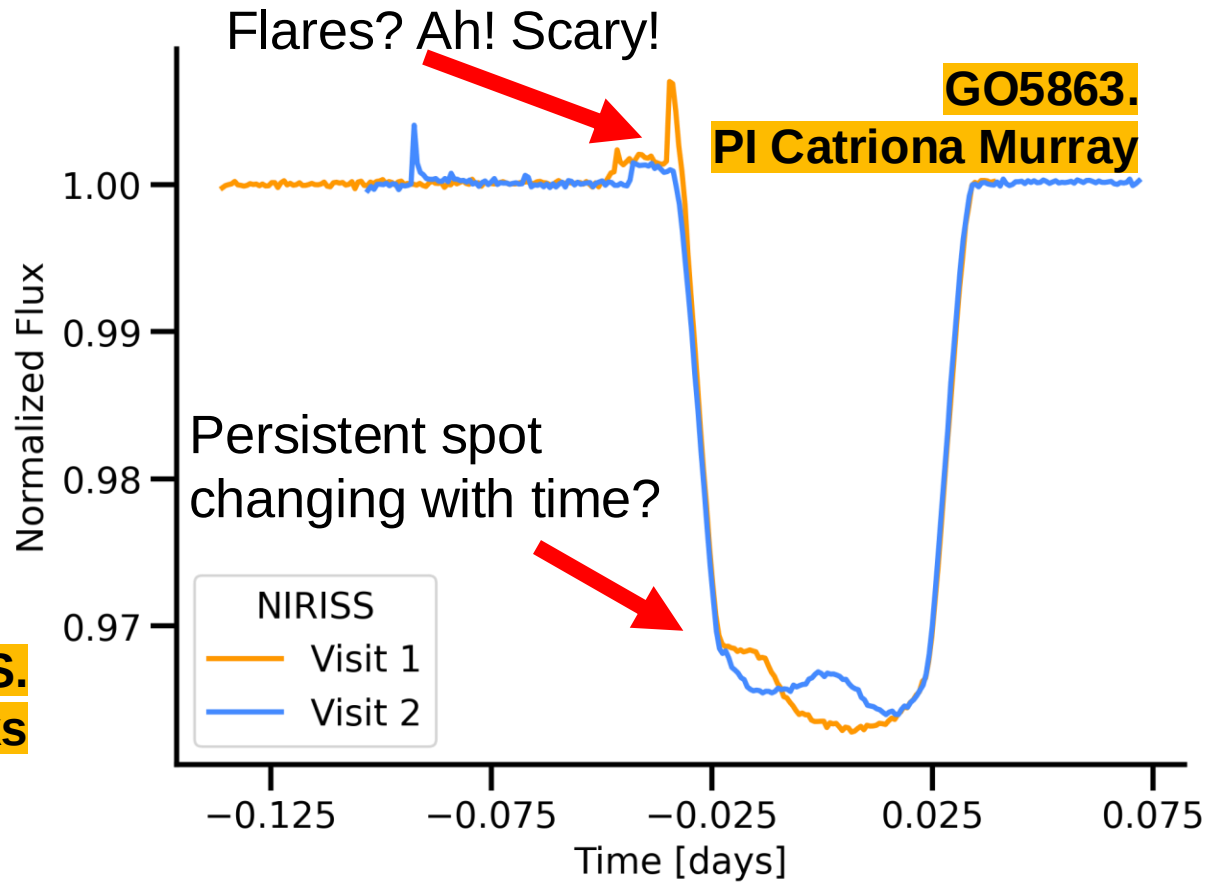
Courtesy of Matthew Murphy. Murphy et al. in prep.

Early Lessons with JWST

1. Stellar surface heterogeneities are ubiquitous in JWST data. Our interpretation is model-dependent.
2. Stellar surface heterogeneities limit our ability to infer the nature of exoplanetary atmospheres. Our interpretation is model-limited.
3. The TLSE is stochastic in nature. Planning against it is difficult. Need multiple visits?
4. Star spots are degenerate with claims of limb-asymmetries.
5. **There is more to stars than spots and faculae. Who would have thought?**

Stars are scary. How do we model this!?





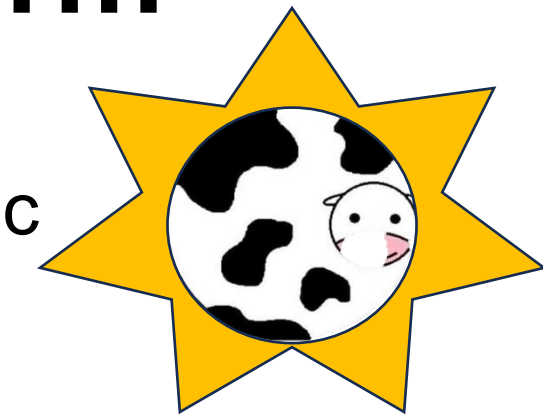
Flares are also present in our light-curves!
How do we model flares when producing a transmission spectrum?
How do we account for their imprint in our ‘retrievals’?

In summary, I am glad you are here

In summary, I am glad you are here

and not grading C4 proposals…

But while we are here together…



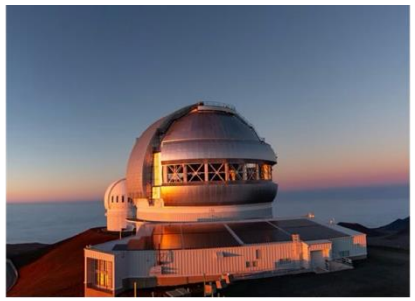
Let's think about making our "spherical cows" more realistic

Beyond removing star gunk:

- Stellar abundances for planet/star comparisons (see e.g., Sun+24)
- Which stellar parameters (Teff, log(g), Z) can we trust when modeling TLSE.
- What about the stellar UV spectrum and its role in photochemistry?
- How do we overcome the limits of our stellar grids and account for our lack of knowledge?
- What about *everything else we talked about*?

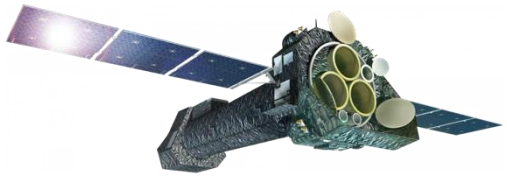
Multi-Wavelength observations to inform our models and inferences

24 orbits with HST



6.3 hours with Gemini-N/IGRINS-2

72 ks with XMM-Newton



80 hours with LCO

50 hours with JWST

We are taking a full multi-wavelength approach to understanding the contributions of spots AND flares to transmission spectra

PIs Feinstein & Welbanks

Take home message

1. **Stars are angry and scary**. They leave imprints in the measured transmission spectrum.
2. Atmospheric inferences (i.e., ‘retrievals’) are **model dependent**.
3. Retrievals have been using the TLSE prescription from Rackham+18 and treatment from Pinhas+18 as a ‘band-aid’ and we know this approach is limited. **Innovation is urgently needed** (see SAG21).
4. Lack of **fidelity in stellar models** bias retrieved atmospheric properties.
5. Stellar contamination are **degenerate** with aerosols in the planet and claims of limb-asymmetries.
6. Remember, is the moon made of cheese or sponge?