EXOCLIMES SIMULATION PLATFORM

EXO PLANETS EXOCLIMES GROUP

FNSNF

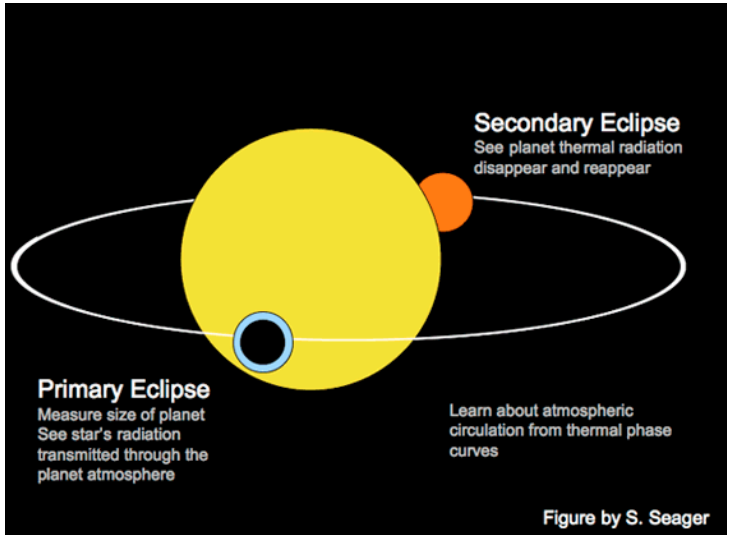
FONDS NATIONAL SUISSE
SCHWEIZERISCHER NATIONALFONDS
FONDO NAZIONALE SVIZZERO
SWISS NATIONAL SCIENCE FOUNDATION

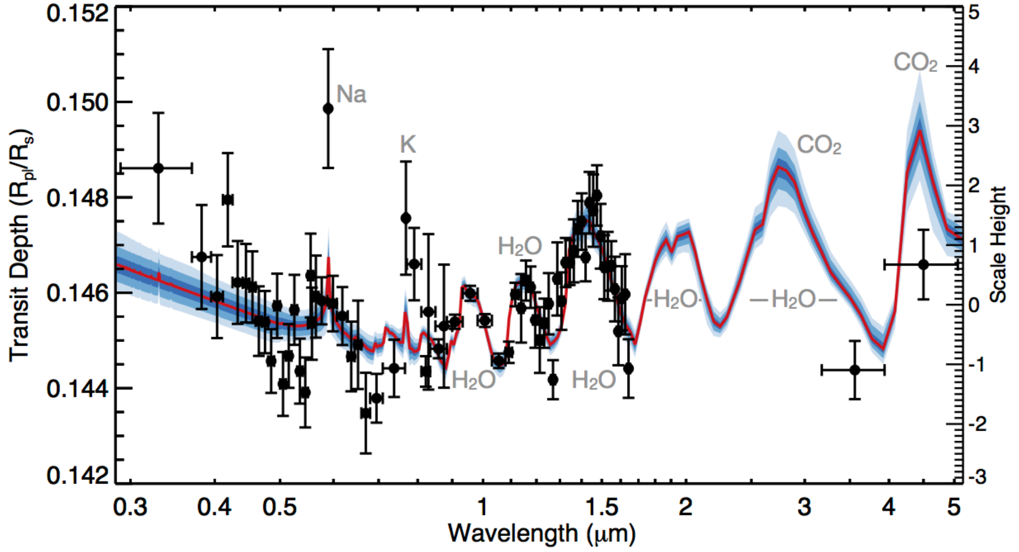
b
UNIVERSITÄT
BERN

Atmospheric Retrieval using Machine Learning and CCFs

Chloe Fisher

Transmission Spectra



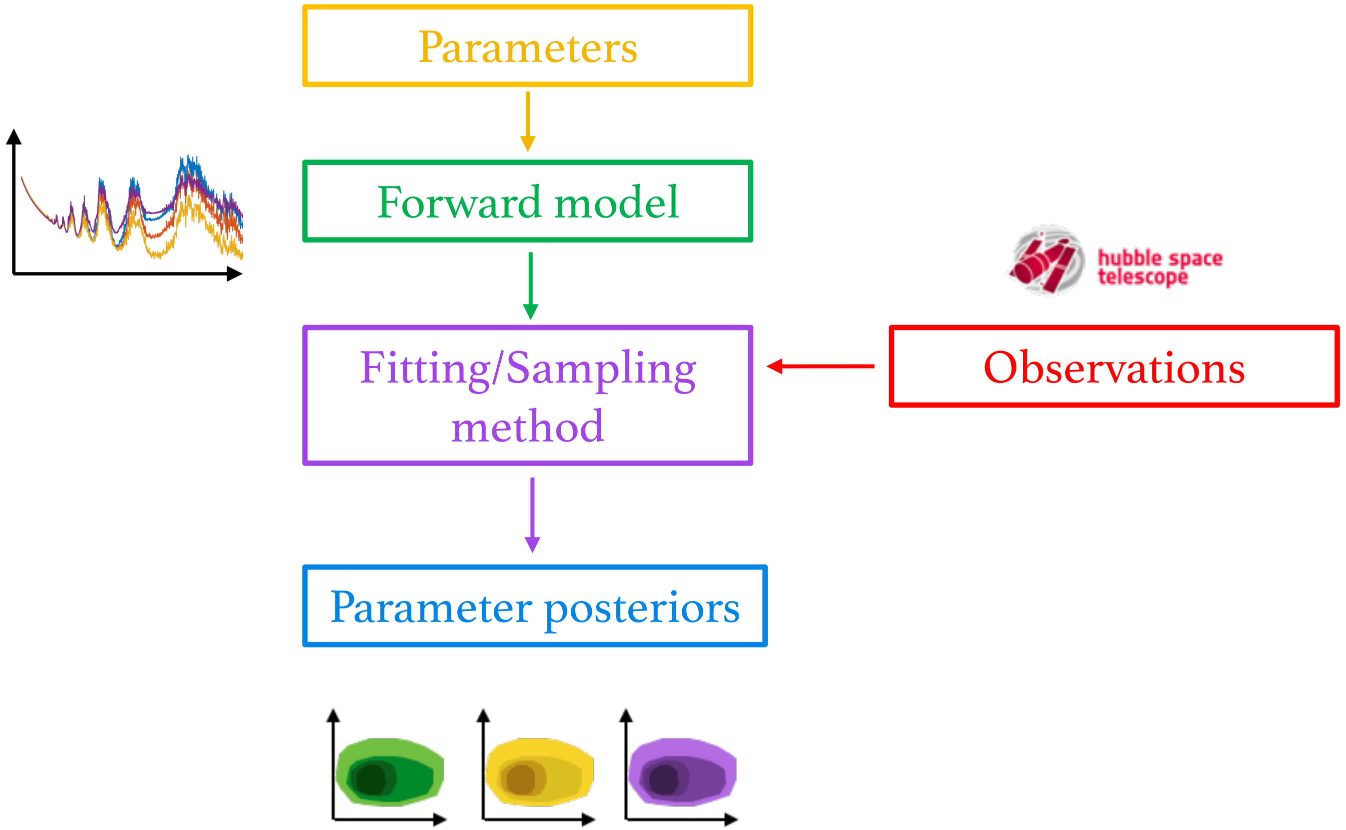


Wakeford et al. 2018

Atmospheric Retrieval

- Uses a sampling technique to search parameter space for the best-fitting model.

- Bayesian methods such as Nested-Sampling or MCMC.

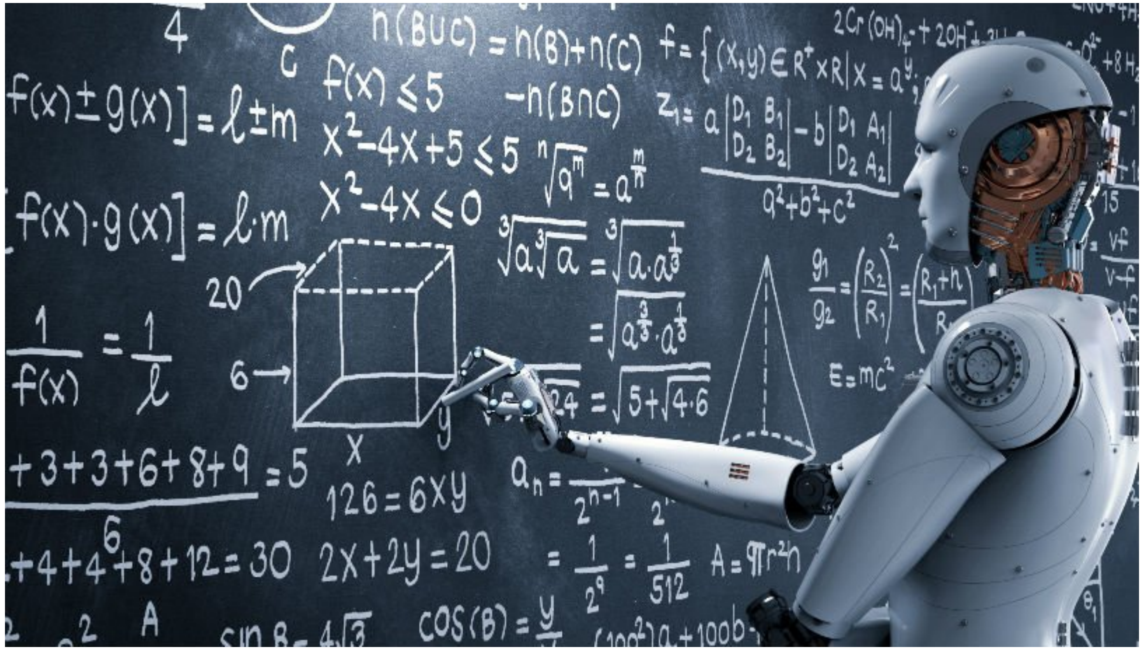


Machine Learning in Retrievals

- Uses a pre-computed training set of models.

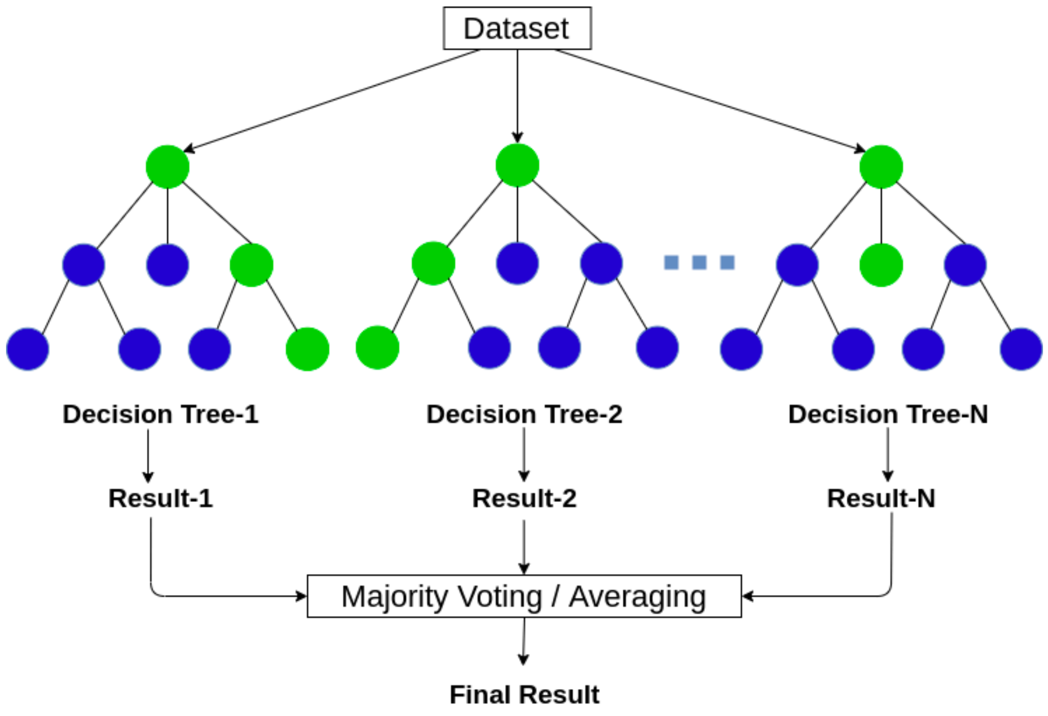
- Shifts computational burden offline, and models can be re-used.

- Doesn't require the code, only the models themselves.



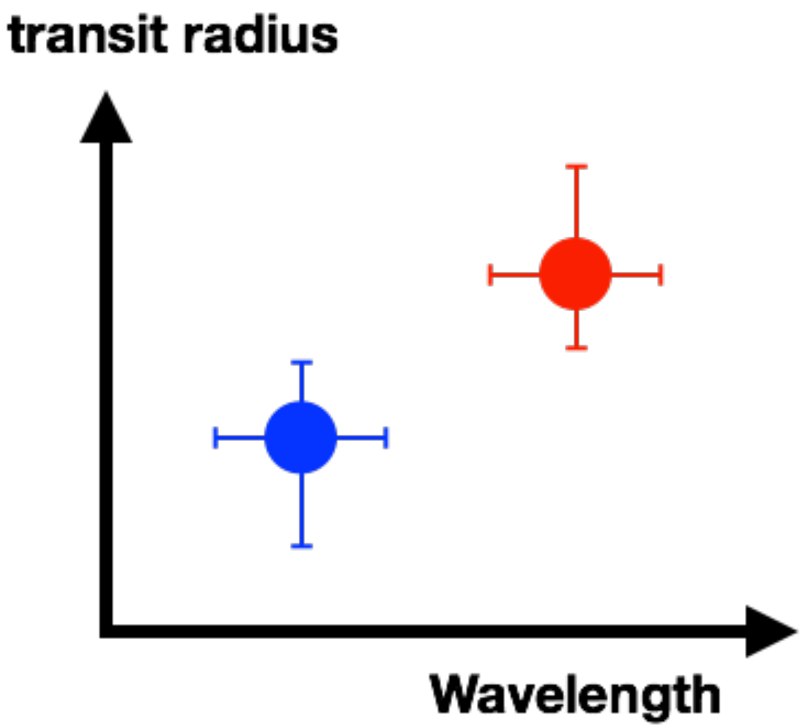
Machine Learning Retrievals

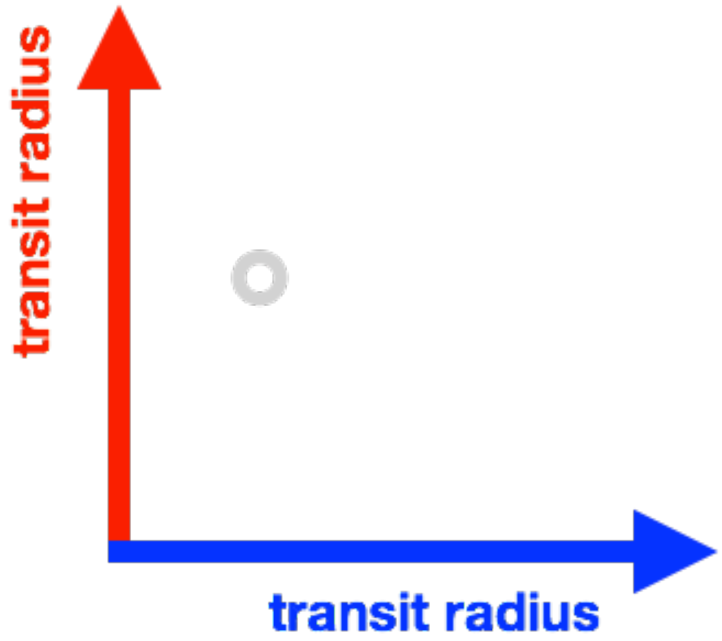
- Neural networks (Waldmann et al. 2015b, Waldmann 2016)
- Random Forests (Márquez-Neila et al. 2018)
- Generative Adversarial Networks (Zingales & Waldmann 2018)
- Bayesian Neural Networks (Cobb et al. 2019)



A Single Regression Tree

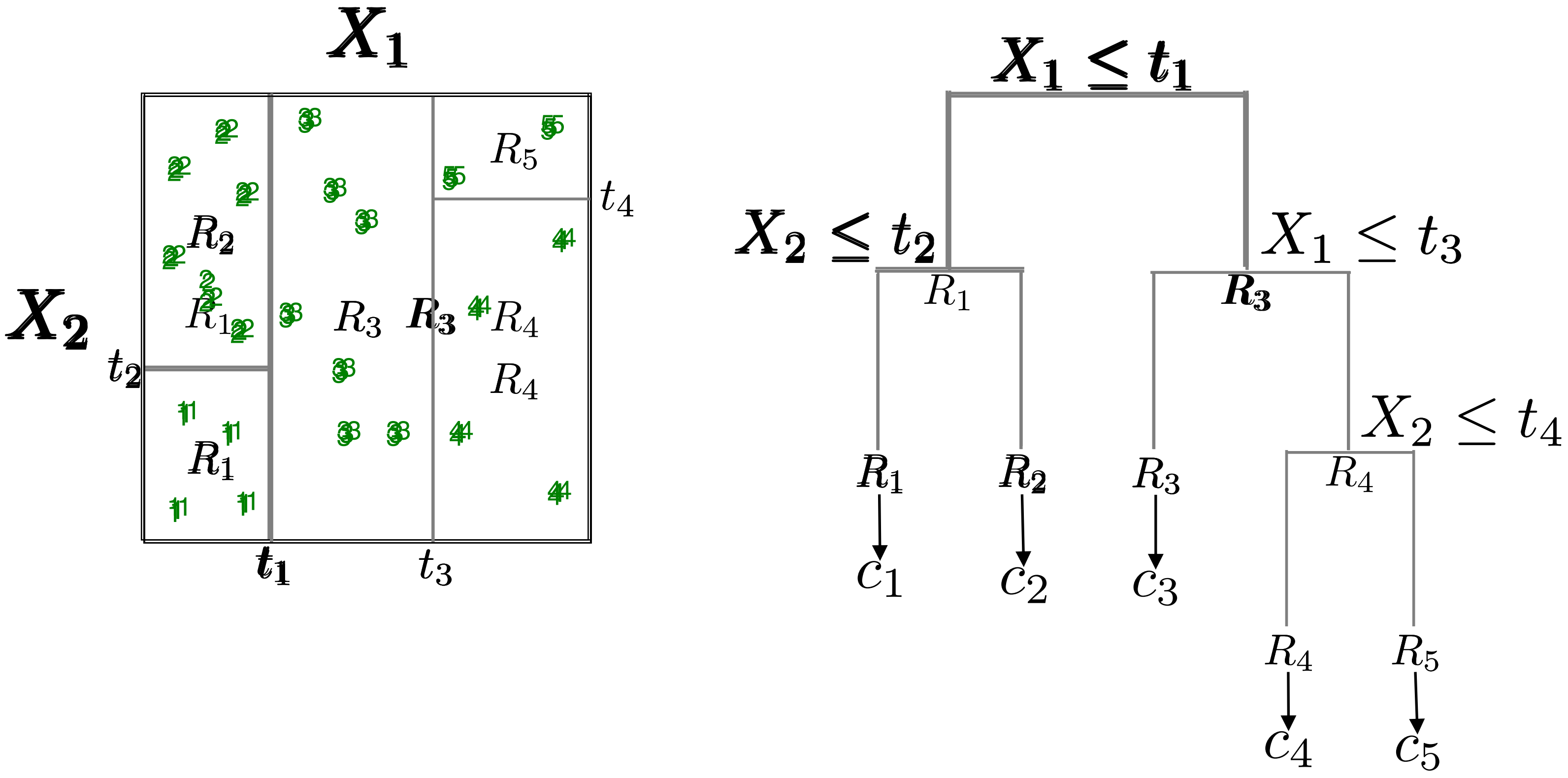
- Transmission Spectra mapping:





A Single Regression Tree - Training

- Parameter: temperature (1 is cold, 5 is hot)



A Random Forest

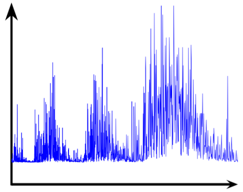
Every tree makes a 'vote' for the answer

- Collection of regression trees.
- Combining many trees solves bias-variance trade-off.
- Predictions from the trees can be combined into posteriors.
- ABC method = posteriors approximate Bayesian ones.

Training Set

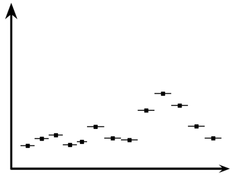
Choose parameters randomly from prior range

↓



Generate model

↓



Bin to data resolution

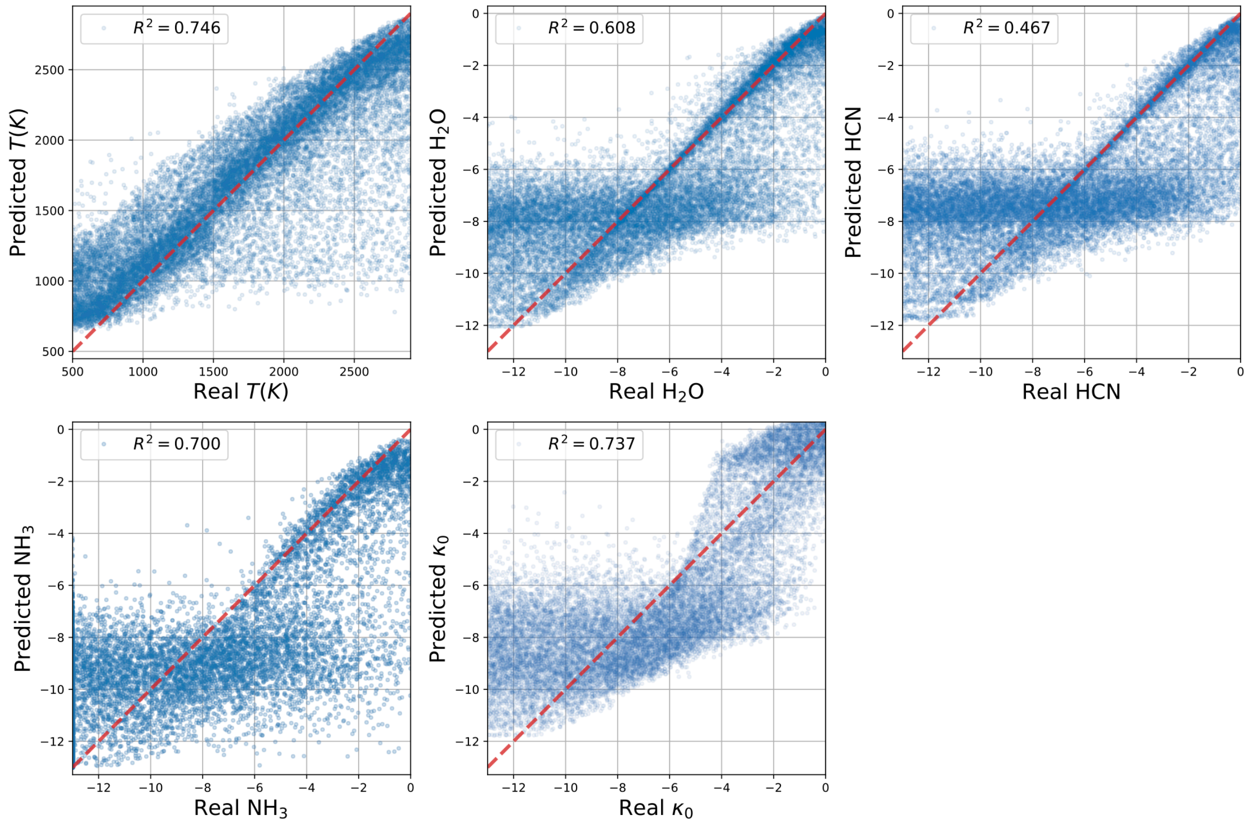
↓

Add random Gaussian noise

- Repeat for many models, split into training and testing.
- Computational burden shifted offline.
- Compute only once.

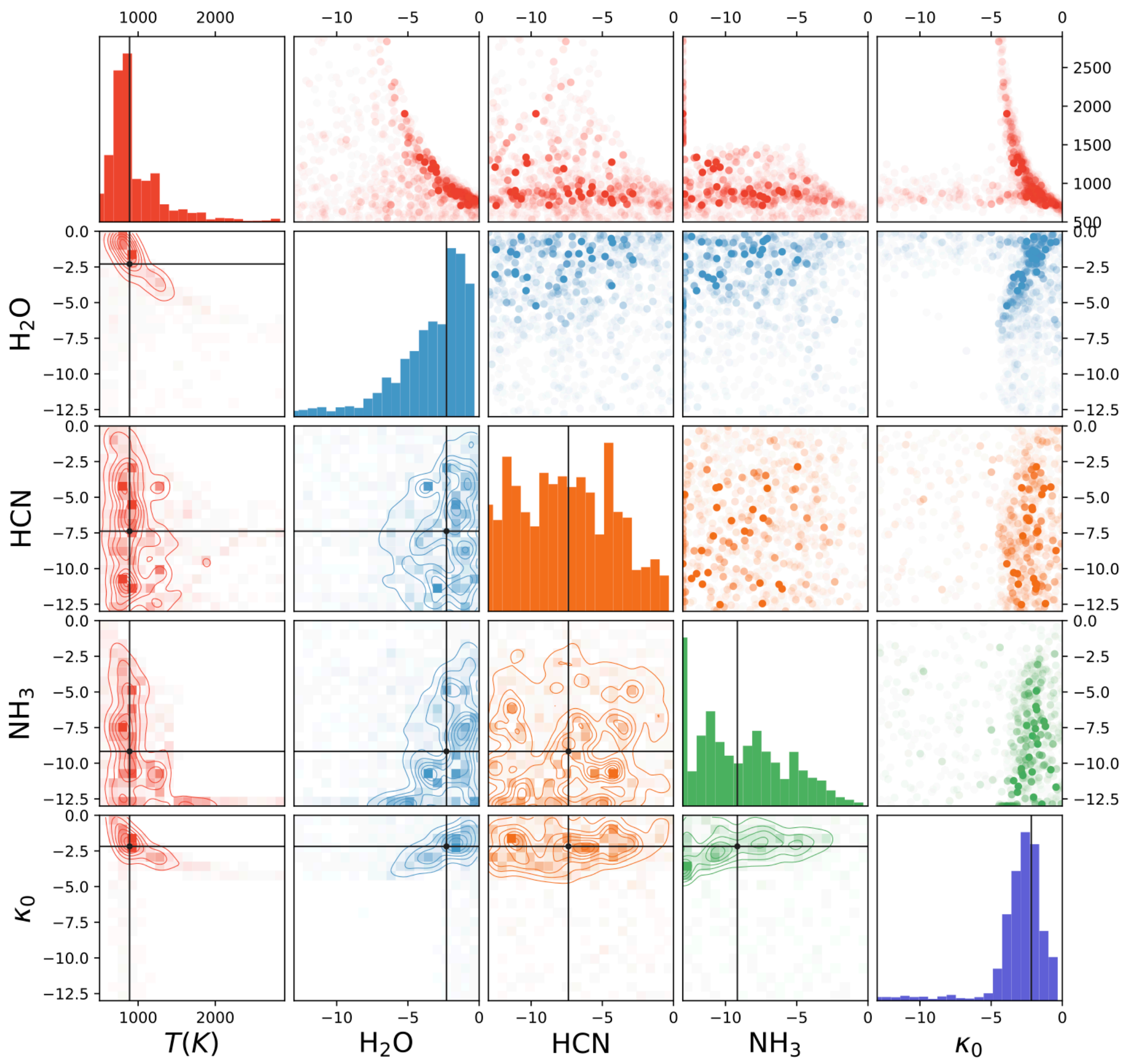
Testing Stage

- Single prediction very fast – allows for thousands of mock retrievals.



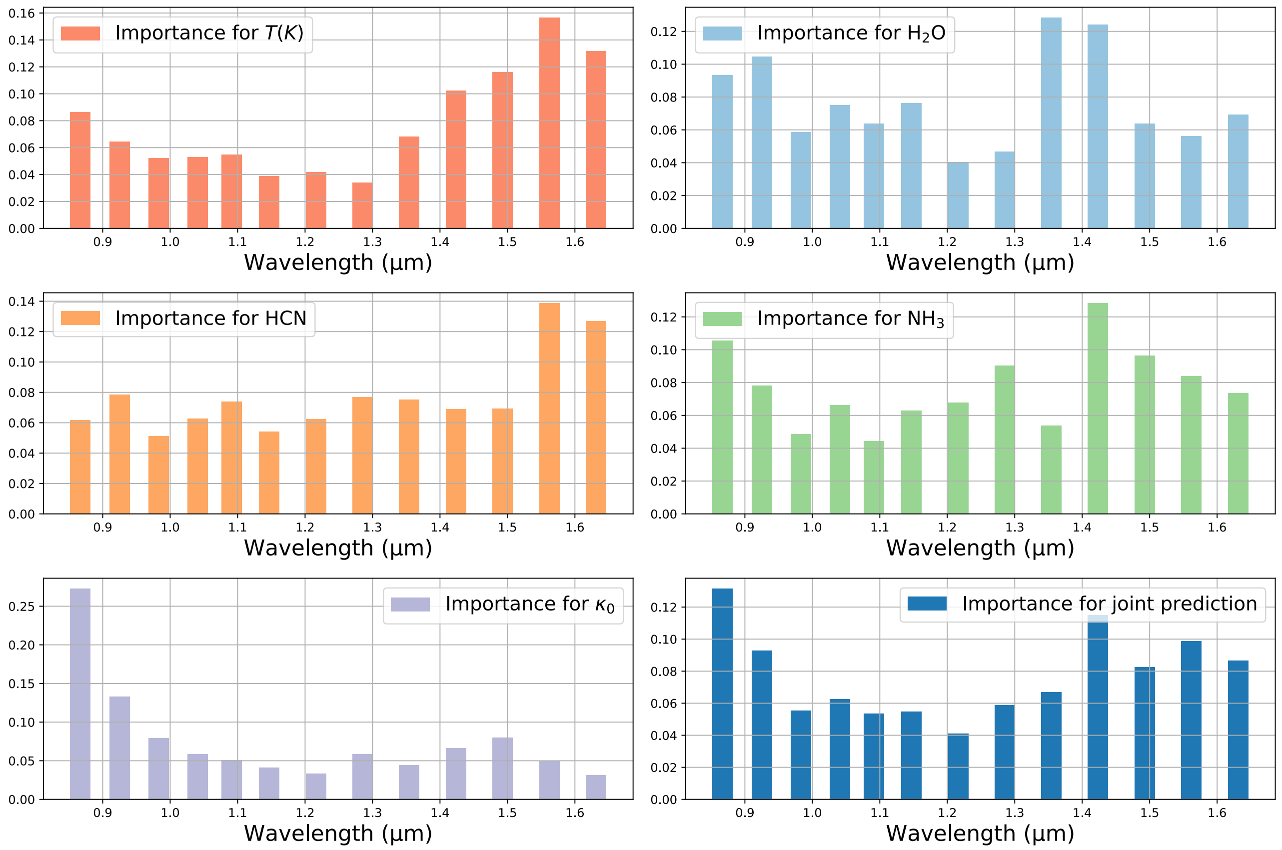
Predicting

- Each point comes from a single tree.
- Trees trained on different subsets of the training set.
- Points collected into distribution – Approximate Bayesian Computing.



Márquez-Neila et al. 2018

Feature Importance



Márquez-Neila et al. 2018

- Quantifies information content for each spectral point.
- Could be used for telescope proposals (see later).

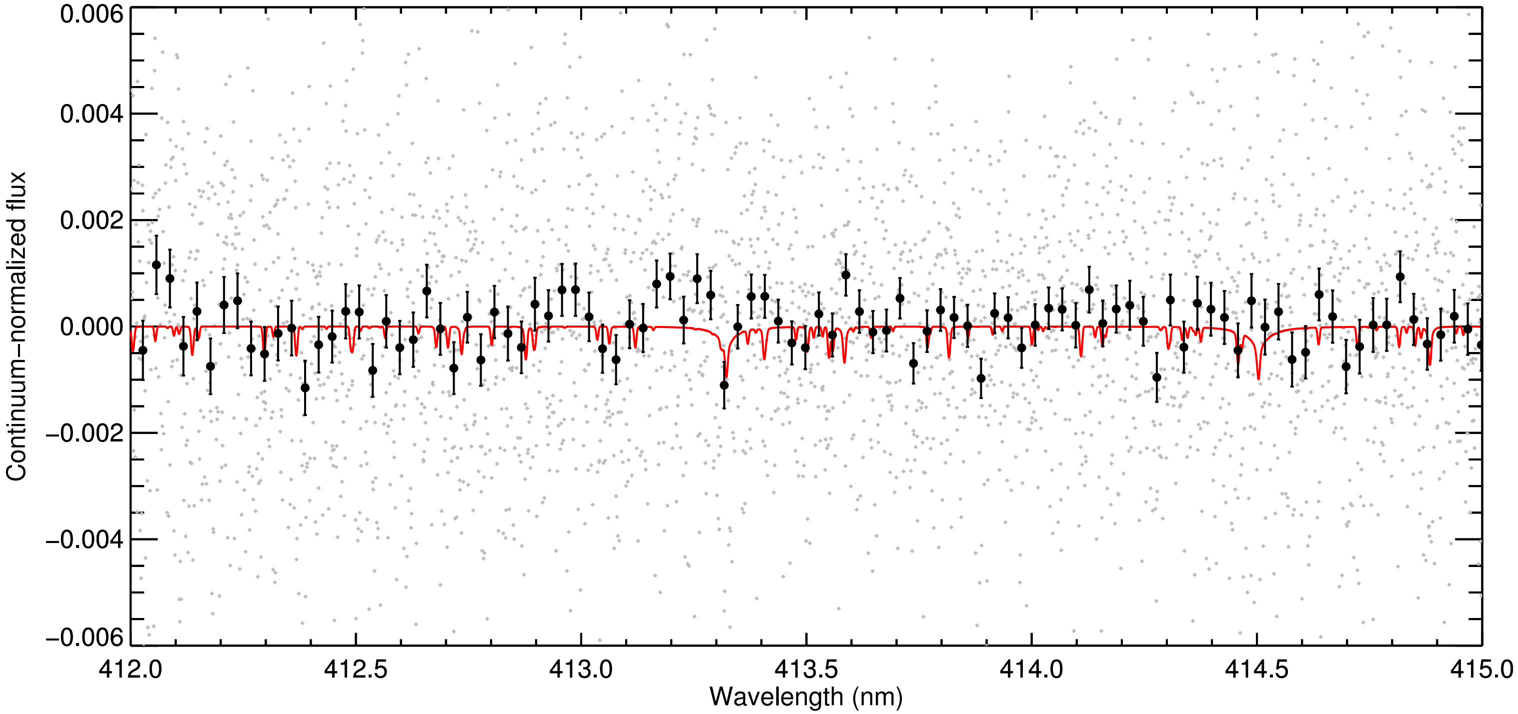
Colab

- Link in Slack channel

- Works through low-resolution retrieval with the Random Forest using HELA

- https://github.com/exoclime/HELA

High-Resolution Data



Zoomed-in HARPS-N spectrum with model (red)

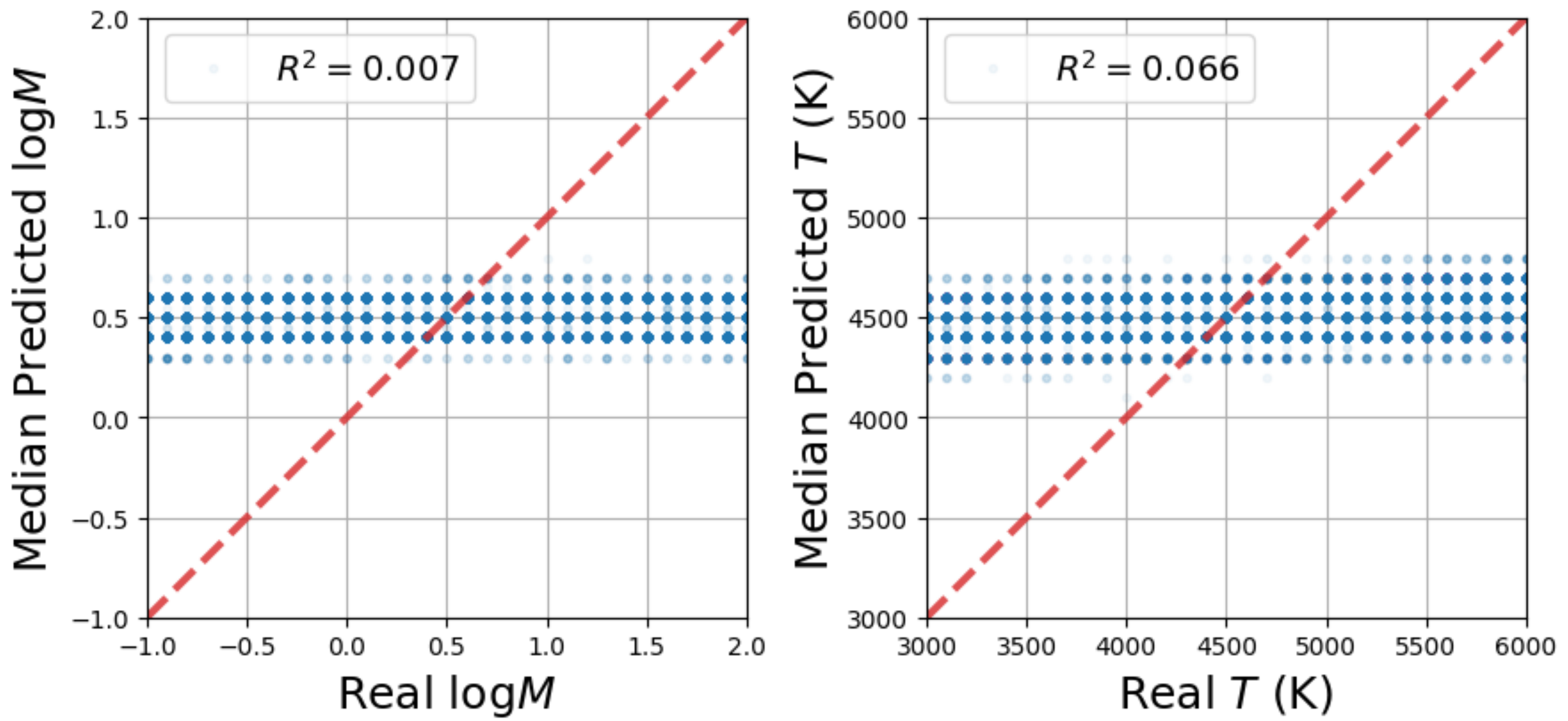
- Huge number of points.
- High level of noise.

KELT-9 b

- Ultra-hot Jupiter with detected iron and titanium.

- Q: Can we retrieve its metallicity and temperature from the high-resolution HARPS-N spectrum using the forest?

Direct Random Forest Retrieval

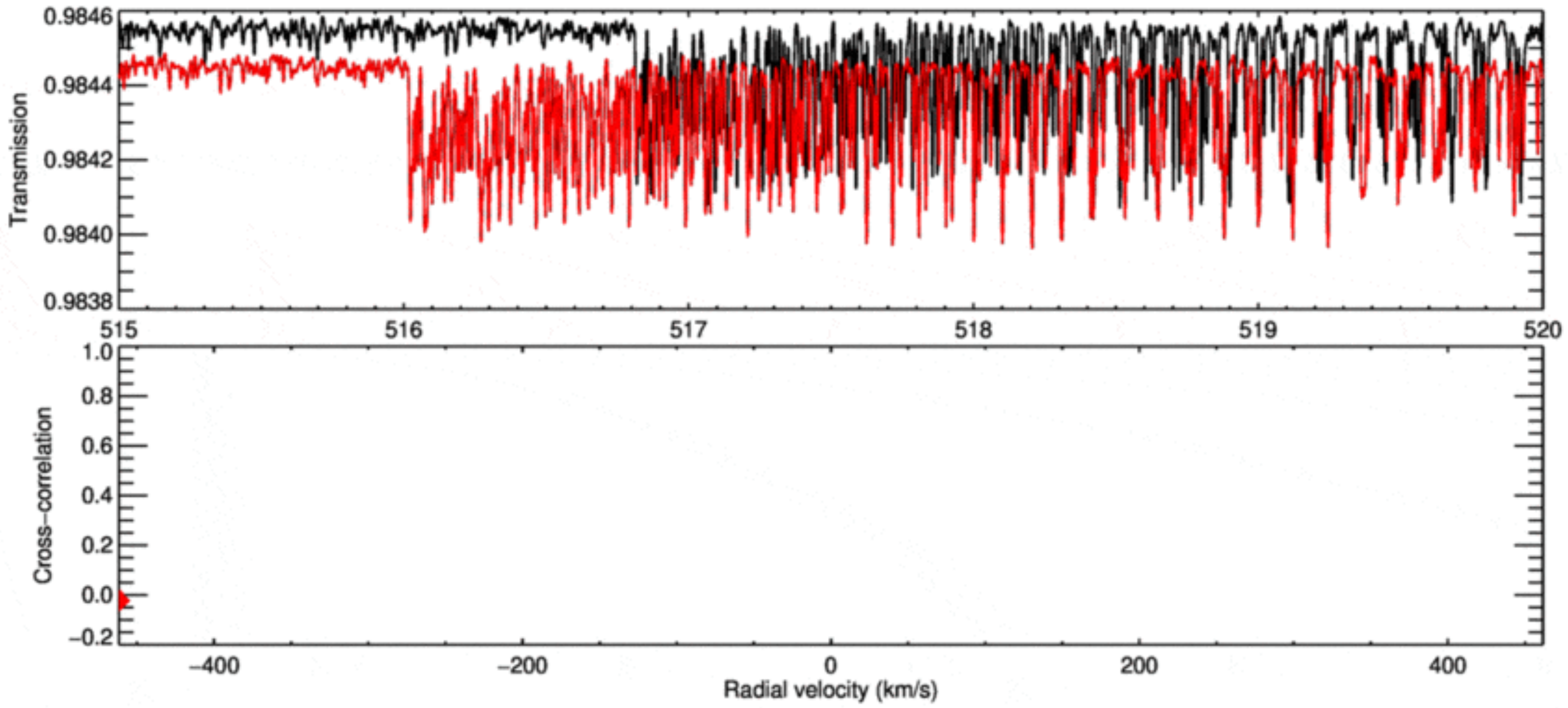


Fisher et al. 2020

Too much noise! (per spectral point)

Cross-Correlations

- Checks molecular templates against spectra.

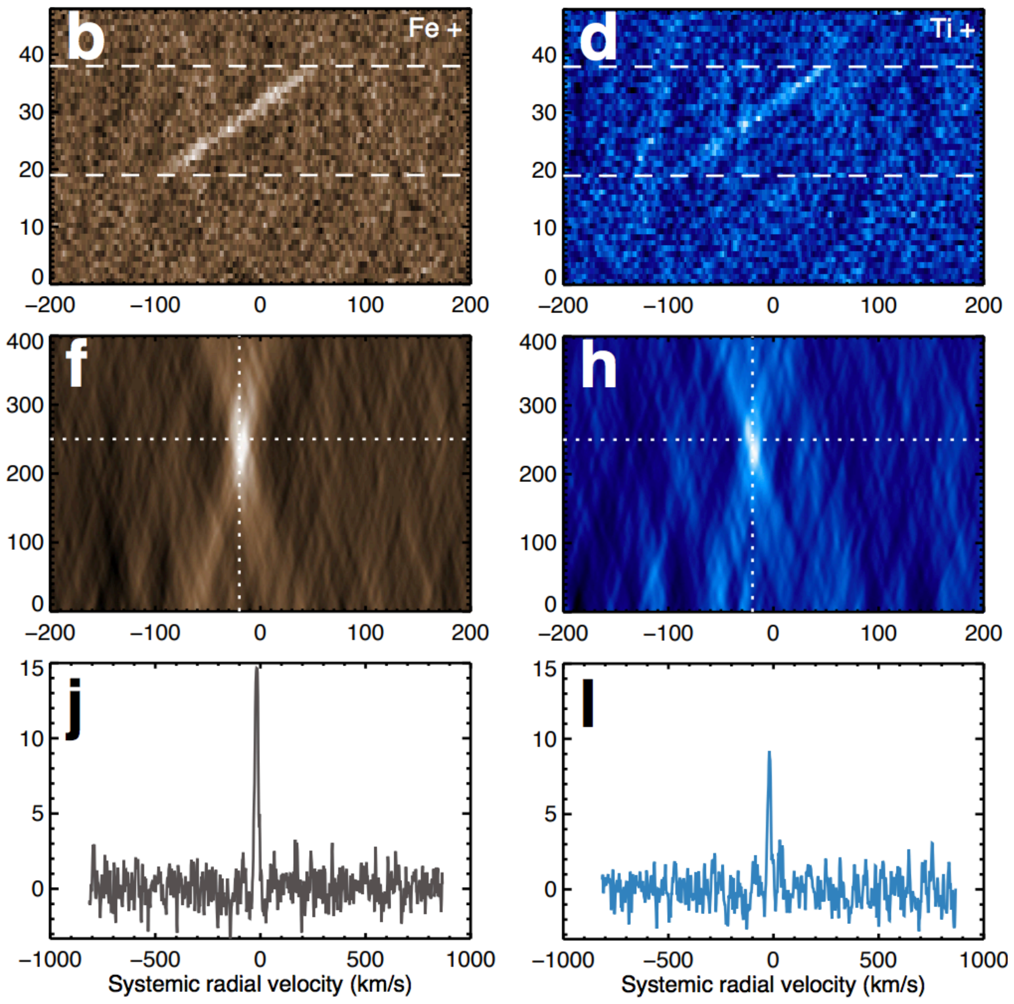


Or watch Jens Hoeijmaker's video lesson

Jens_Hoeijmakers_Astro on youtube

Cross-Correlations

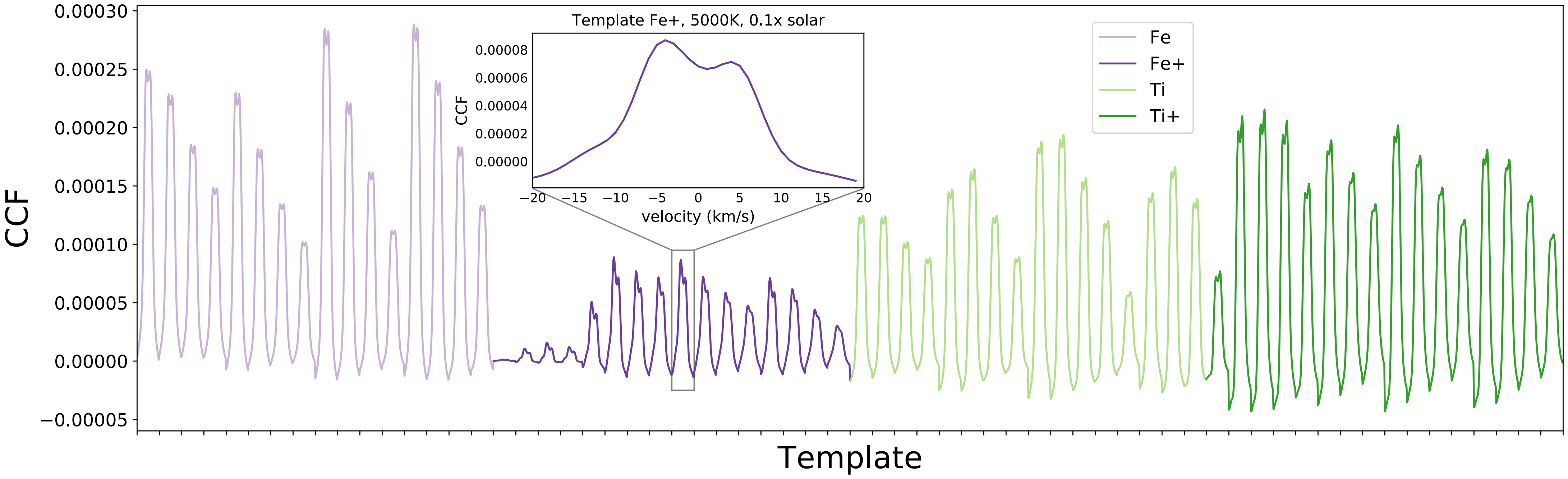
- Used primarily for detections.
- Could the shape of the CCF tell us something about the temperature/metallicity?



Hoeijmakers et al. 2018

CCF-Sequences

- Cross-correlate model/spectrum with 64 different templates (4 species, 4 temperatures, 4 metallicities).

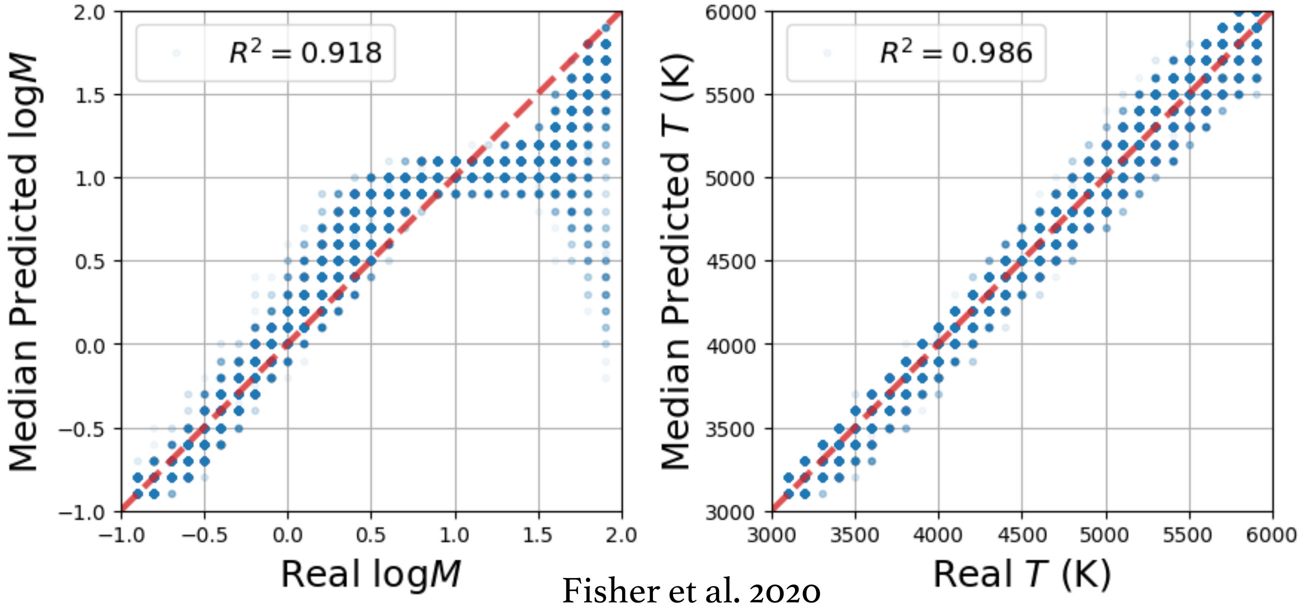


Fisher et al. 2020

CCF-Sequence Retrieval

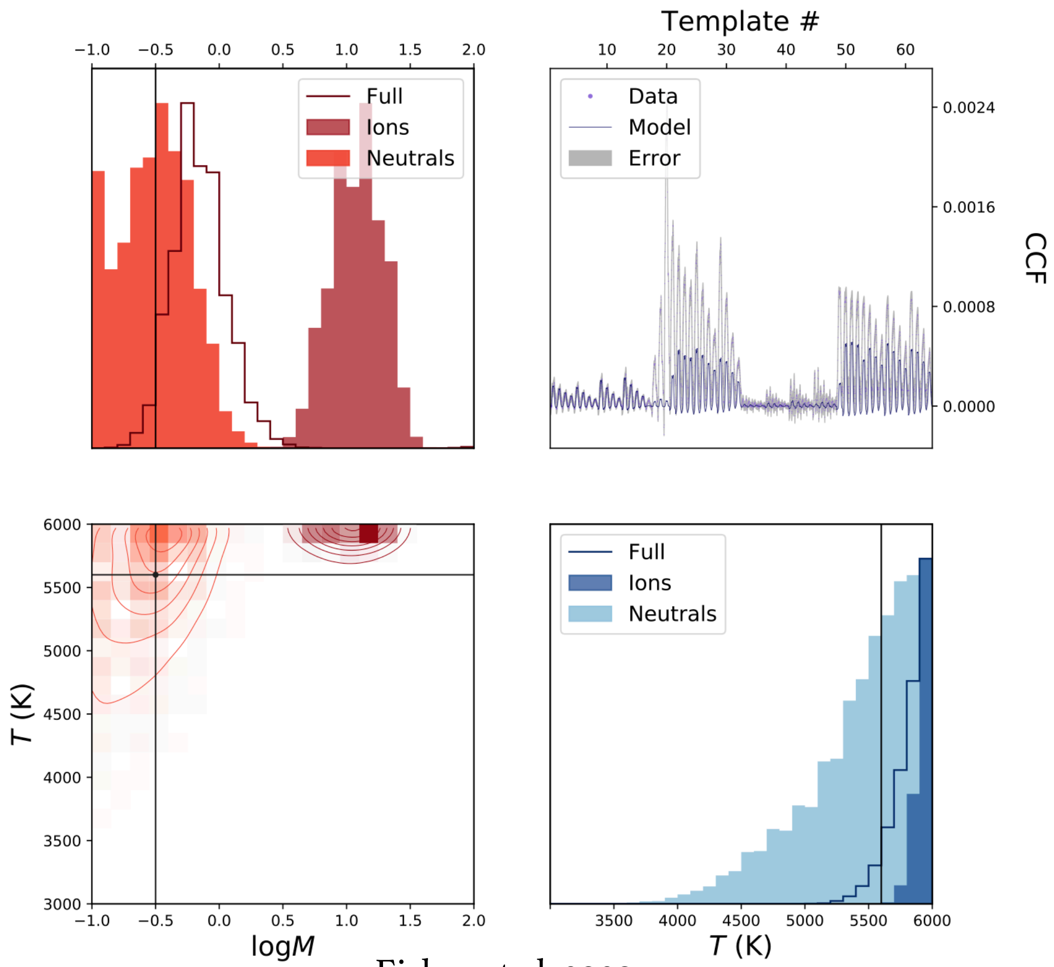
- Generated 1000 models, converted to CCF-Sequences.
- Added noise (easy with CCF).
- Trained and tested Random Forest.

$$C(v) = \frac{\sum_i F_i T_i(v)}{\sum_i T_i(v)}$$



Fisher et al. 2020

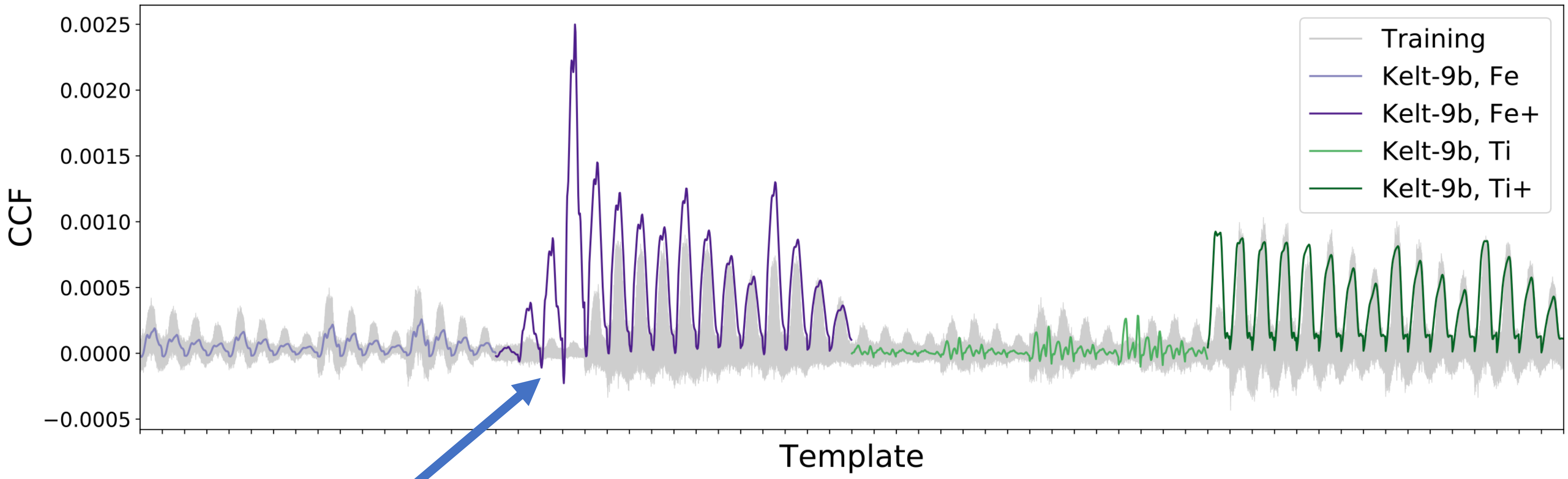
KELT-9b Retrieval



Fisher et al. 2020

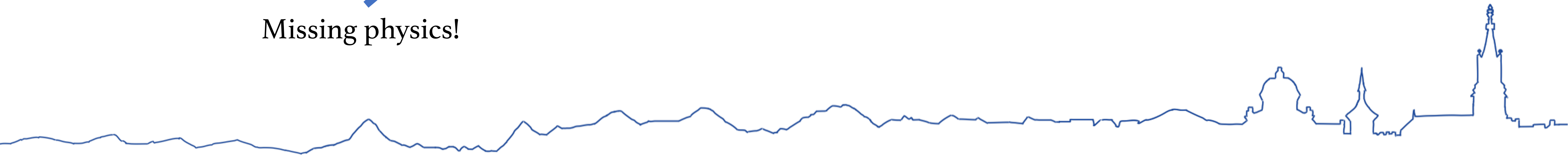
- Retrievals using full CCF-Sequence, just ions and just neutrals.
- High Fe+ CCFs drive retrieval to extreme temperatures.
- Requires more advanced physical models.

KELT-9b vs Training Set



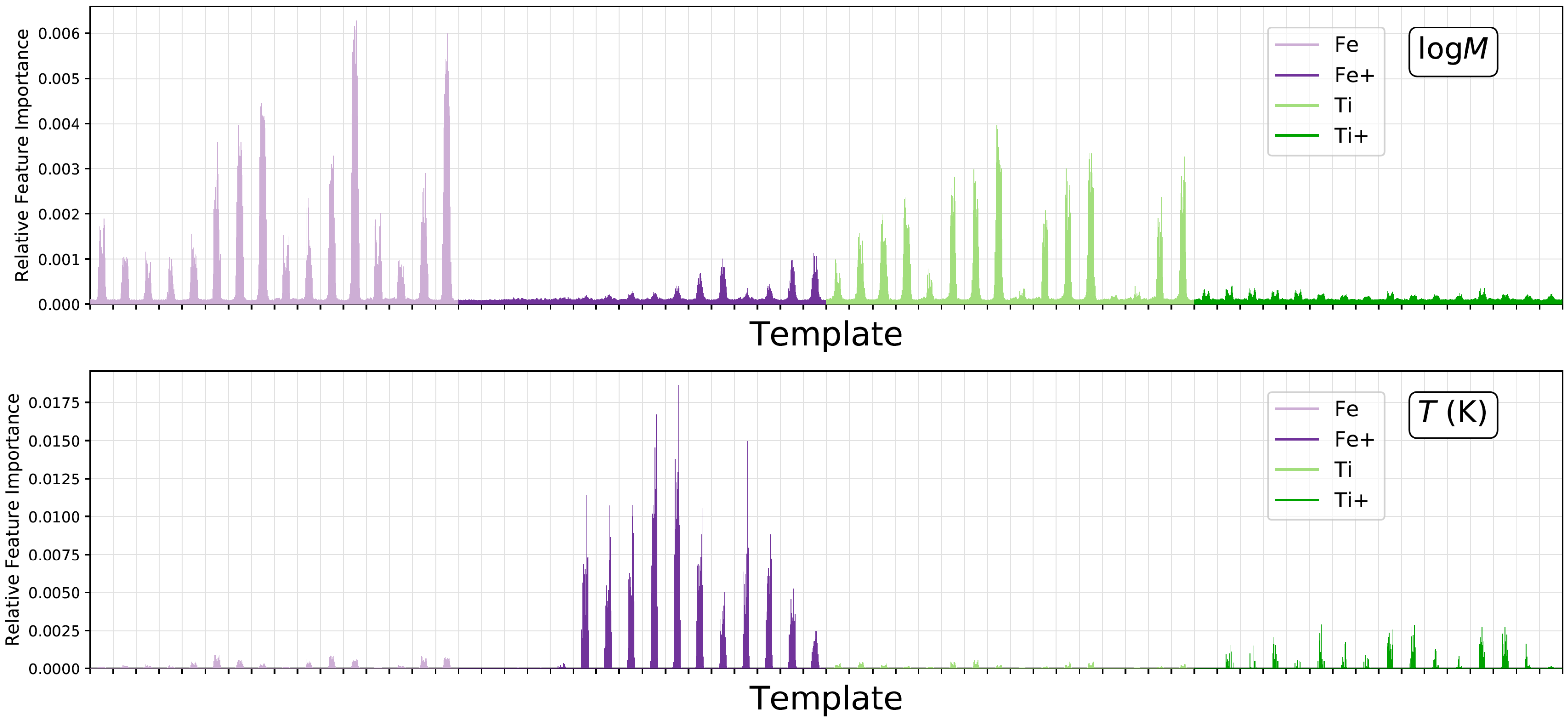
Fisher et al. 2020

Missing physics!



Feature Importance

- Metallicity driven by neutral species.
- Temperature driven by ions.



Fisher et al. 2020

Recap

- Ingredients:

  - grid of models or code

  - species templates

  - cross-correlation code

  - random forest

  - help? (chloe.fisher@unibe.ch)

e.g. Eleonora & Evert's lecture, Ryan & Natasha's lecture

Kitzmann, Hoeijmakers et al., in prep

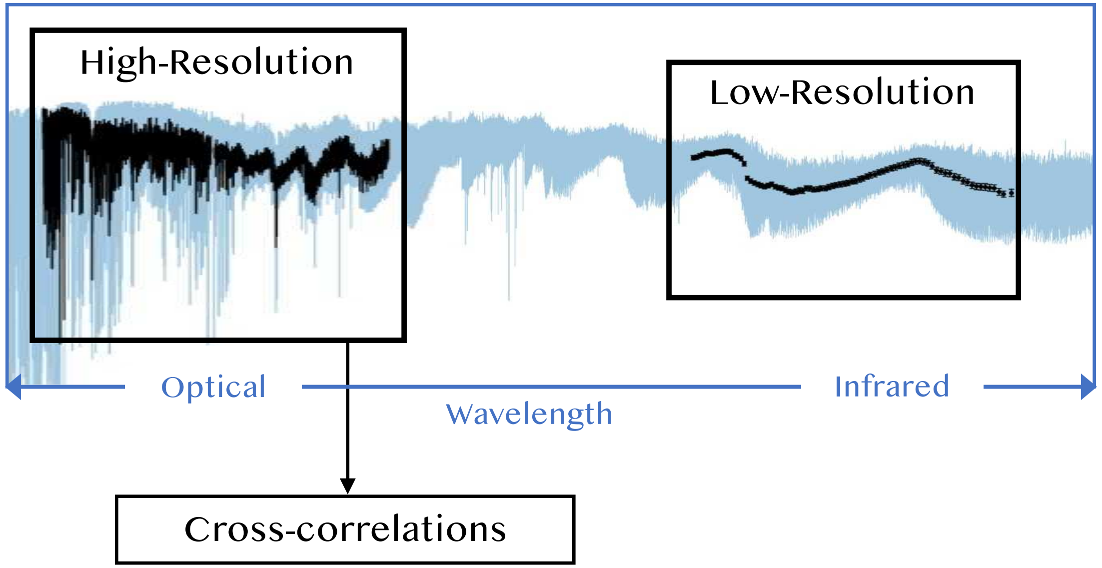
e.g. Jens & Matteo's lecture

HELA

Combining High- and Low-resolution Data

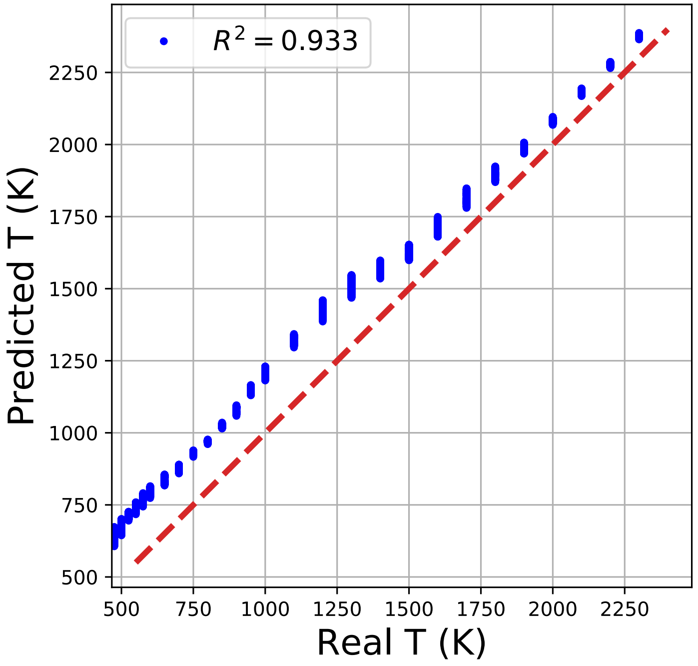
- As Sid showed, tighter constraints can be obtained.

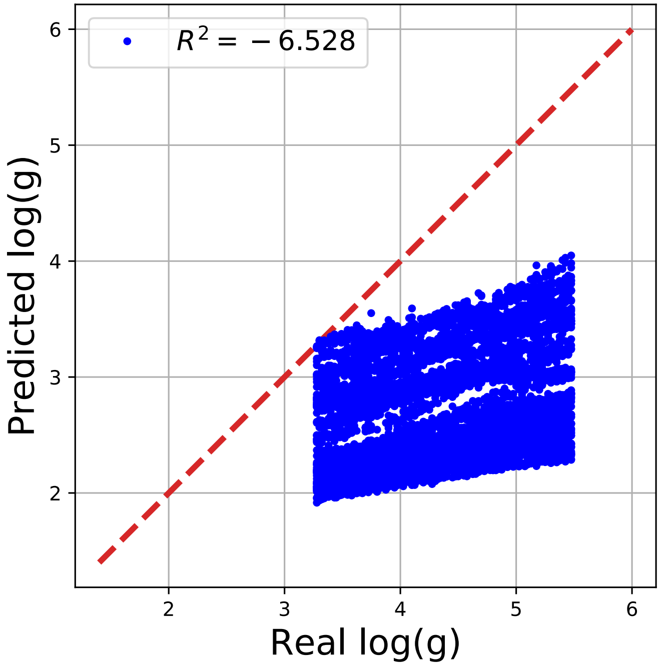
- Feature importance can tell us which dataset drives each parameter.



Brown Dwarf Grid Comparison

- Compare grids of Brown Dwarf models by training on one and testing on another (Oreshenko et al. 2020).

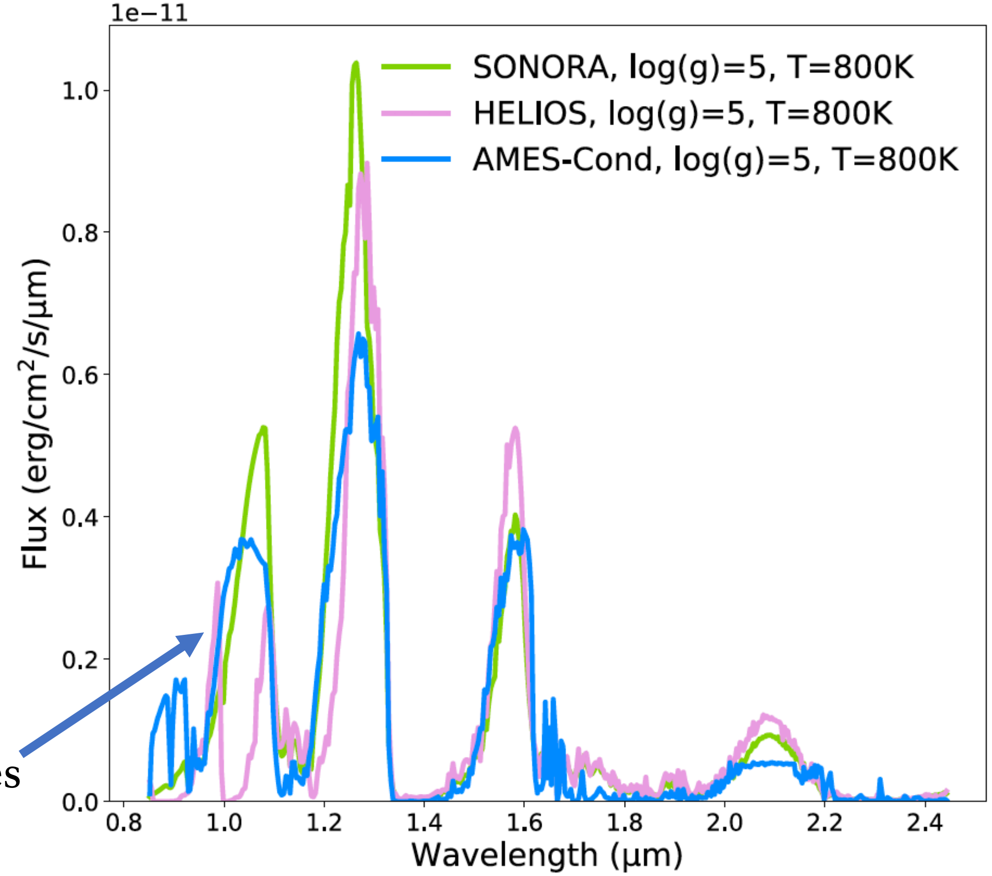




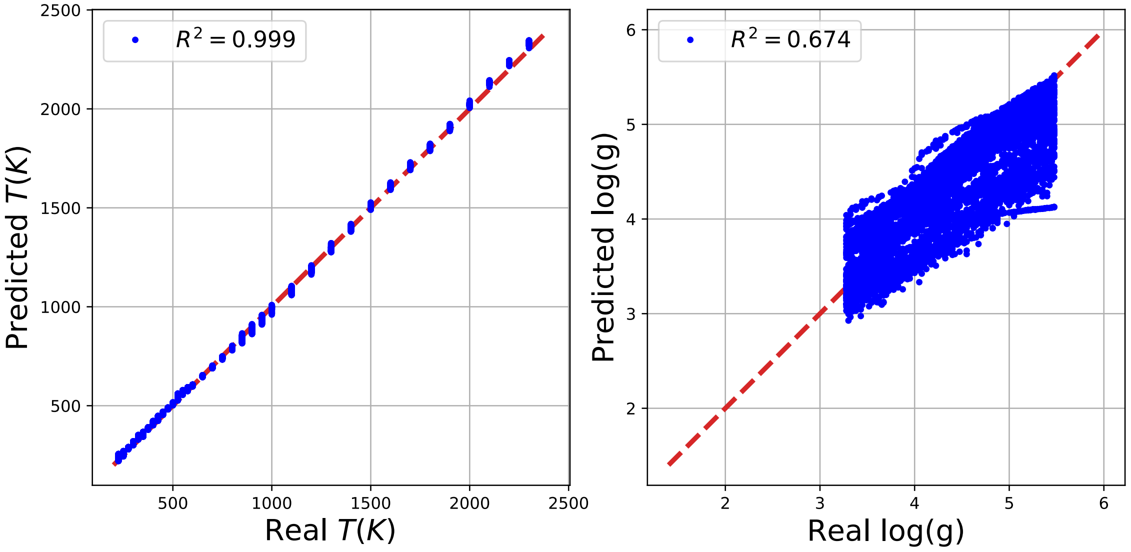
Trained on HELIOS, tested on SENORA

Brown Dwarf Grid Comparison

Maria Oreshenko



Alkali lines treated differently

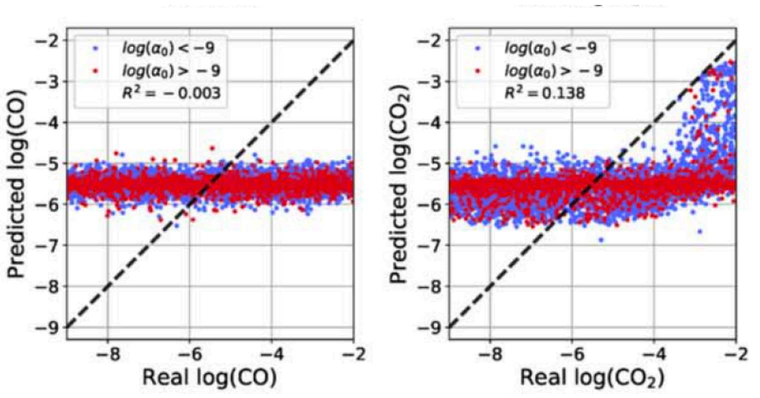


Trained on HELIOS, tested on SENORA, without < 1.2 μm

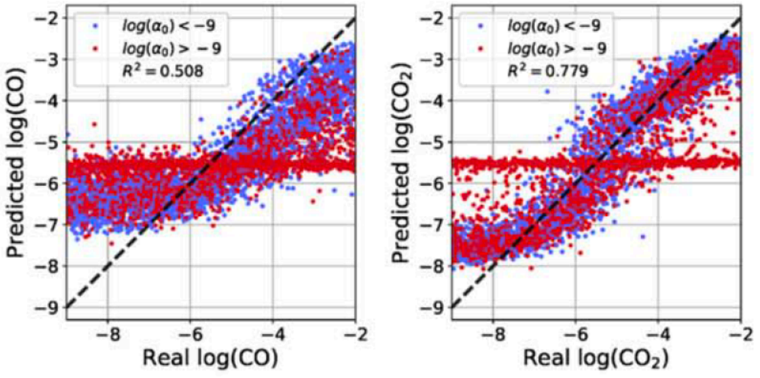
Oreshenko et al. 2020

JWST for Warm Neptunes

Andrea Guzmán Mesa



M1 Mode



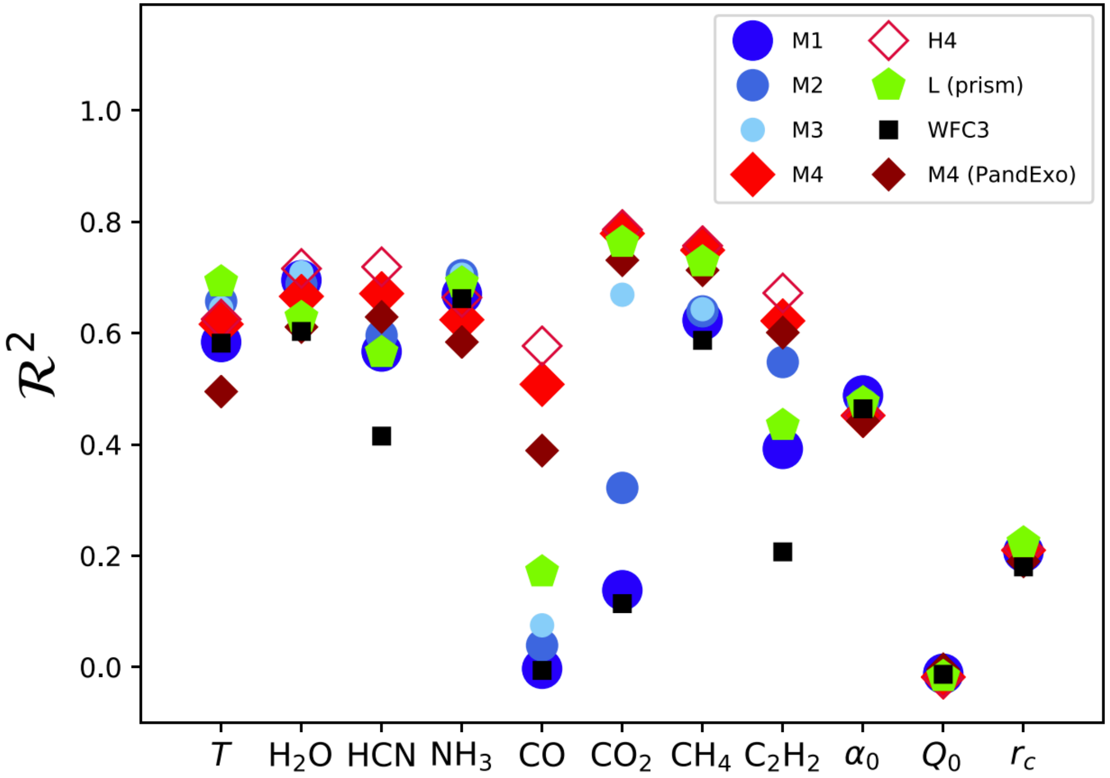
M4 Mode

- Use predictability and feature importance to determine which JWST NIRSpec modes are optimal for studying warm Neptunes (Guzmán Mesa et al. 2020).

JWST for Warm Neptunes

Andrea Guzmán Mesa

- Conclusion: medium resolution M4 mode is recommended.



Guzmán Mesa et al. 2020

Summary

- Machine learning can be used to speed up atmospheric retrieval, with added benefits.

- Random Forest and CCF combination can retrieve on high-resolution data.

- Forests can be used to compare model grids and optimise telescope mode selection.