



“Understanding Predictions made by Machine Learning for Spectroscopic Atmospheric Characterisation”

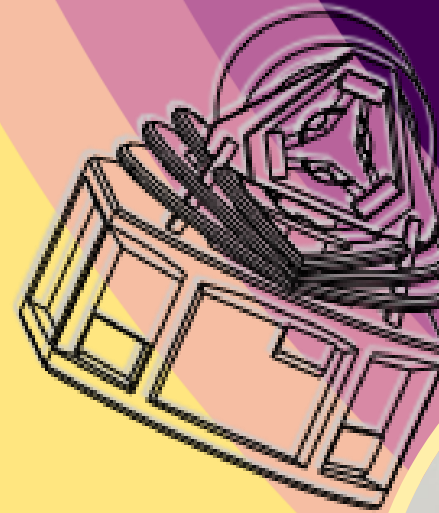
**Jools Clarke¹, Gordon Yip², and Nikos
Nikolaou³**

¹ jools.clarke.23@ucl.ac.uk

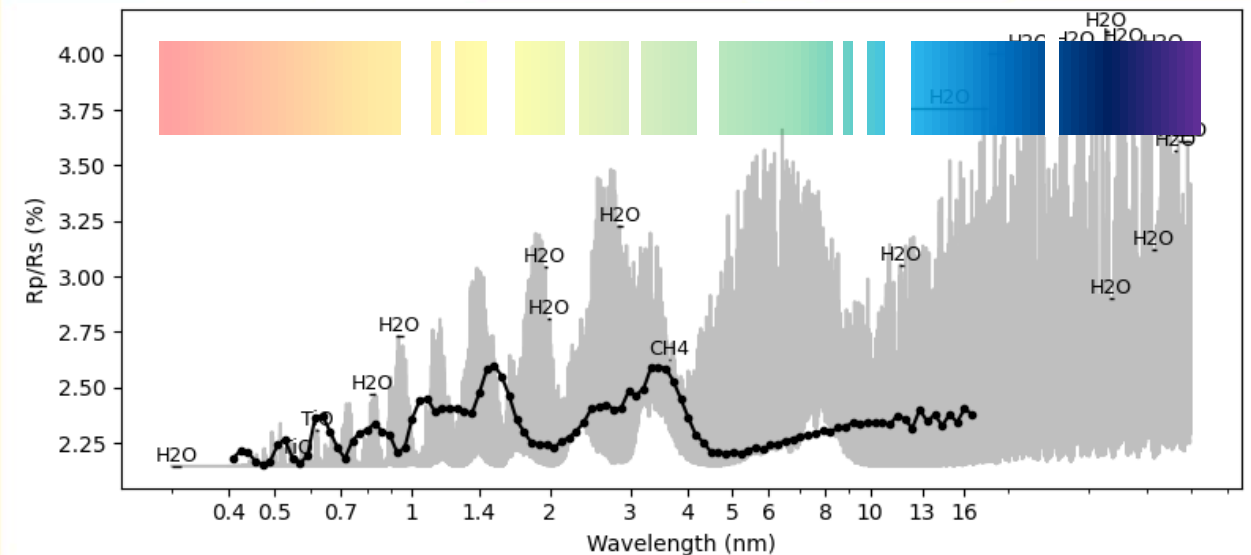
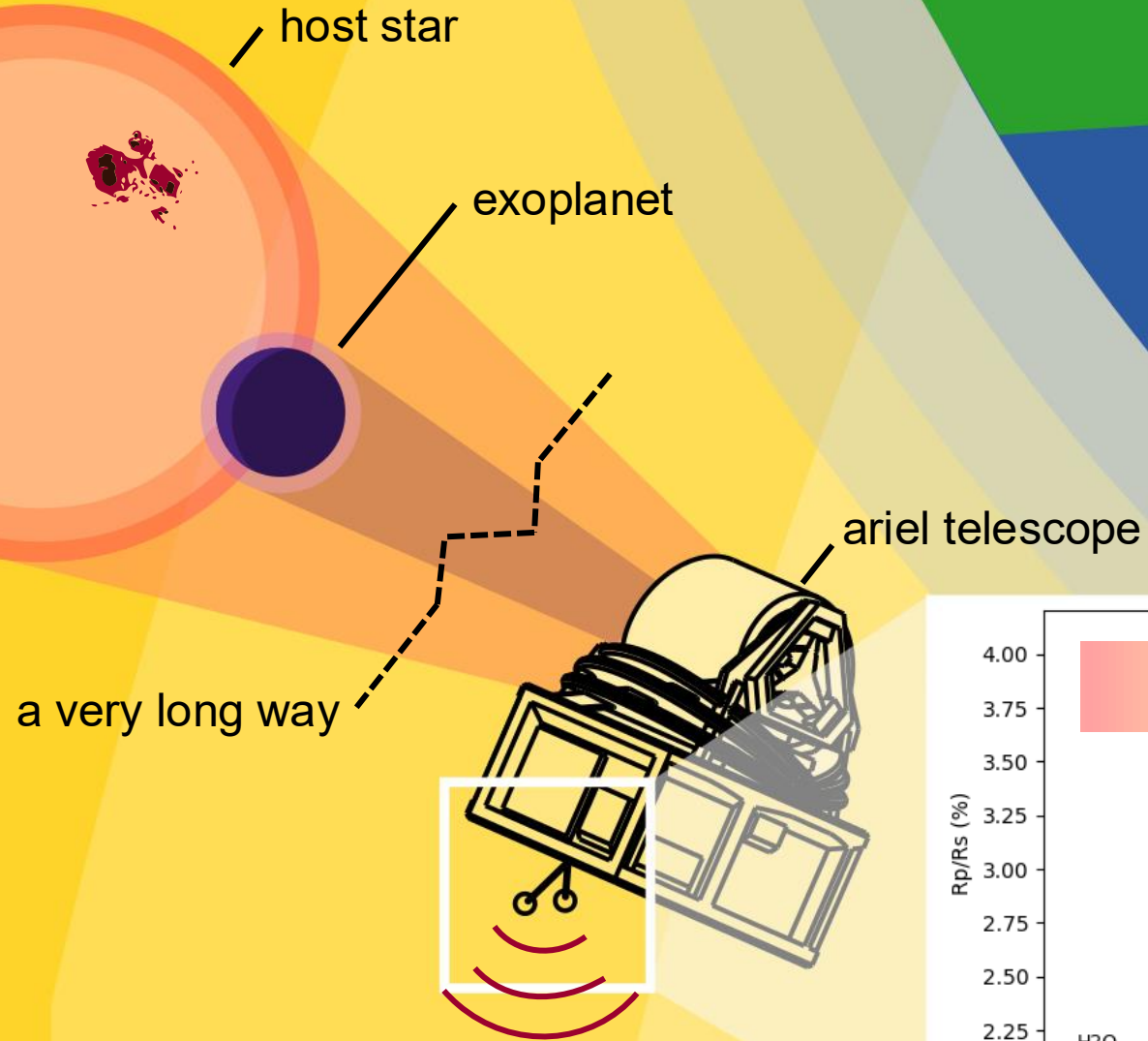
² kai.hou.yip@ucl.ac.uk

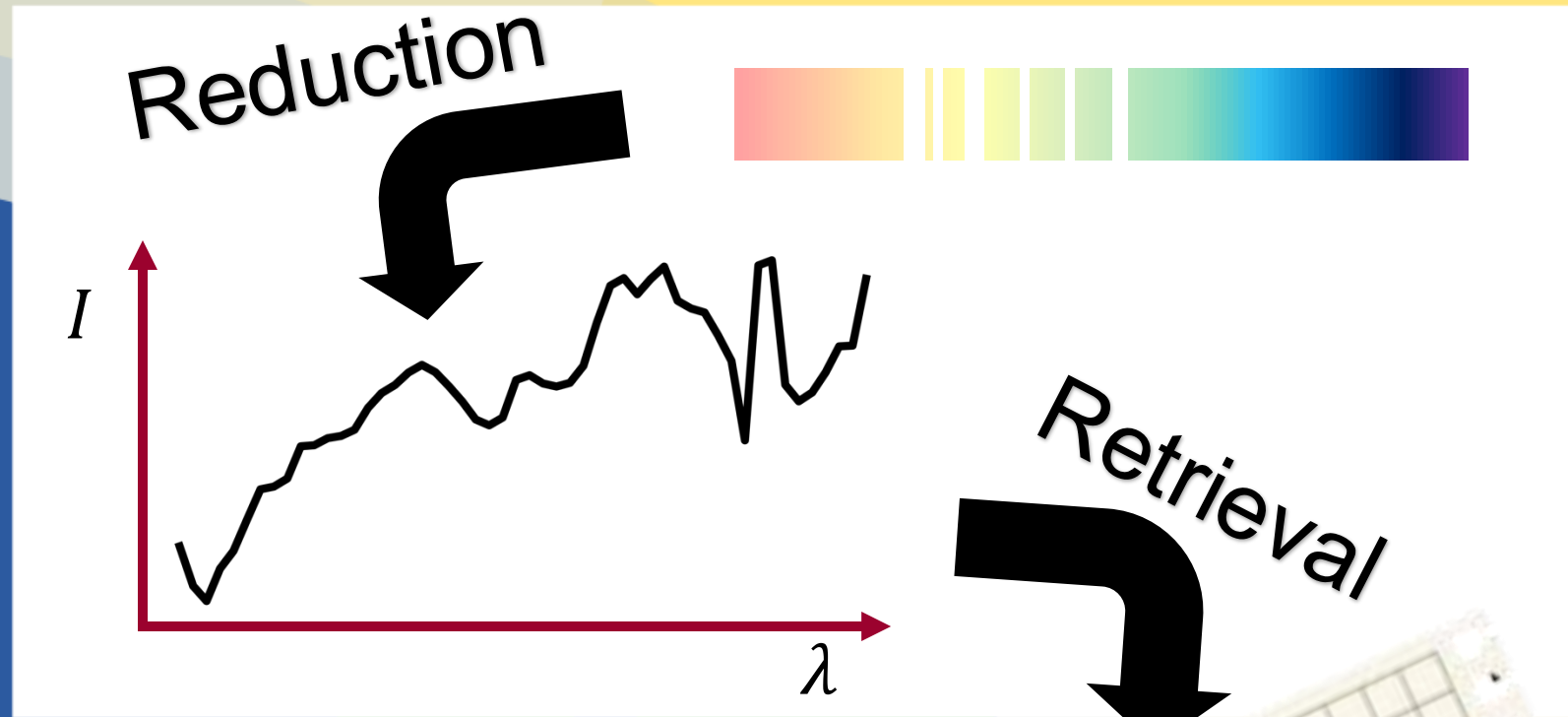
³ n.nikolaou@ucl.ac.uk

*With thanks to **Sushuang Ma** for the data*

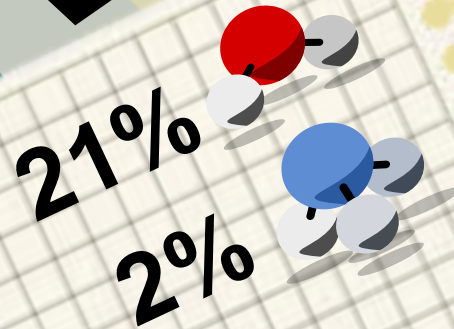







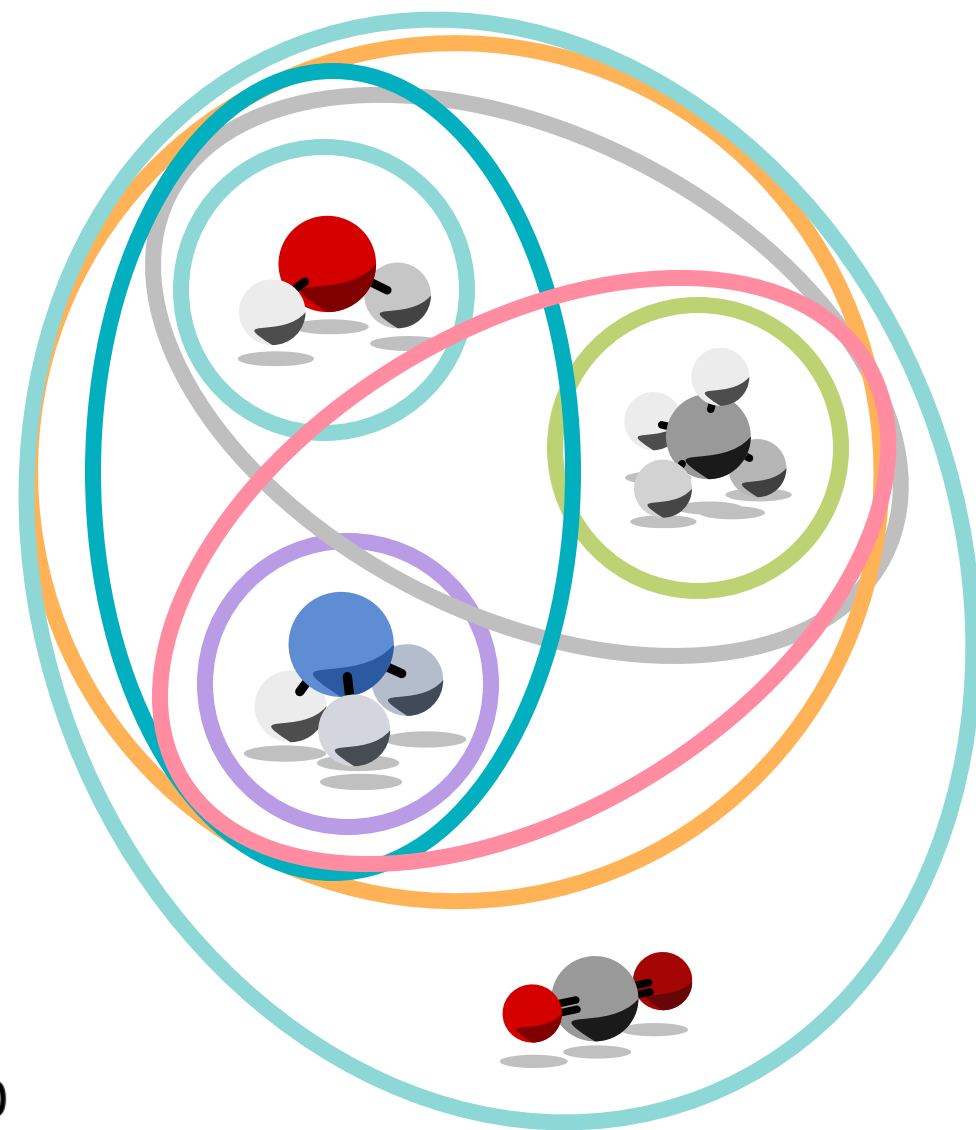
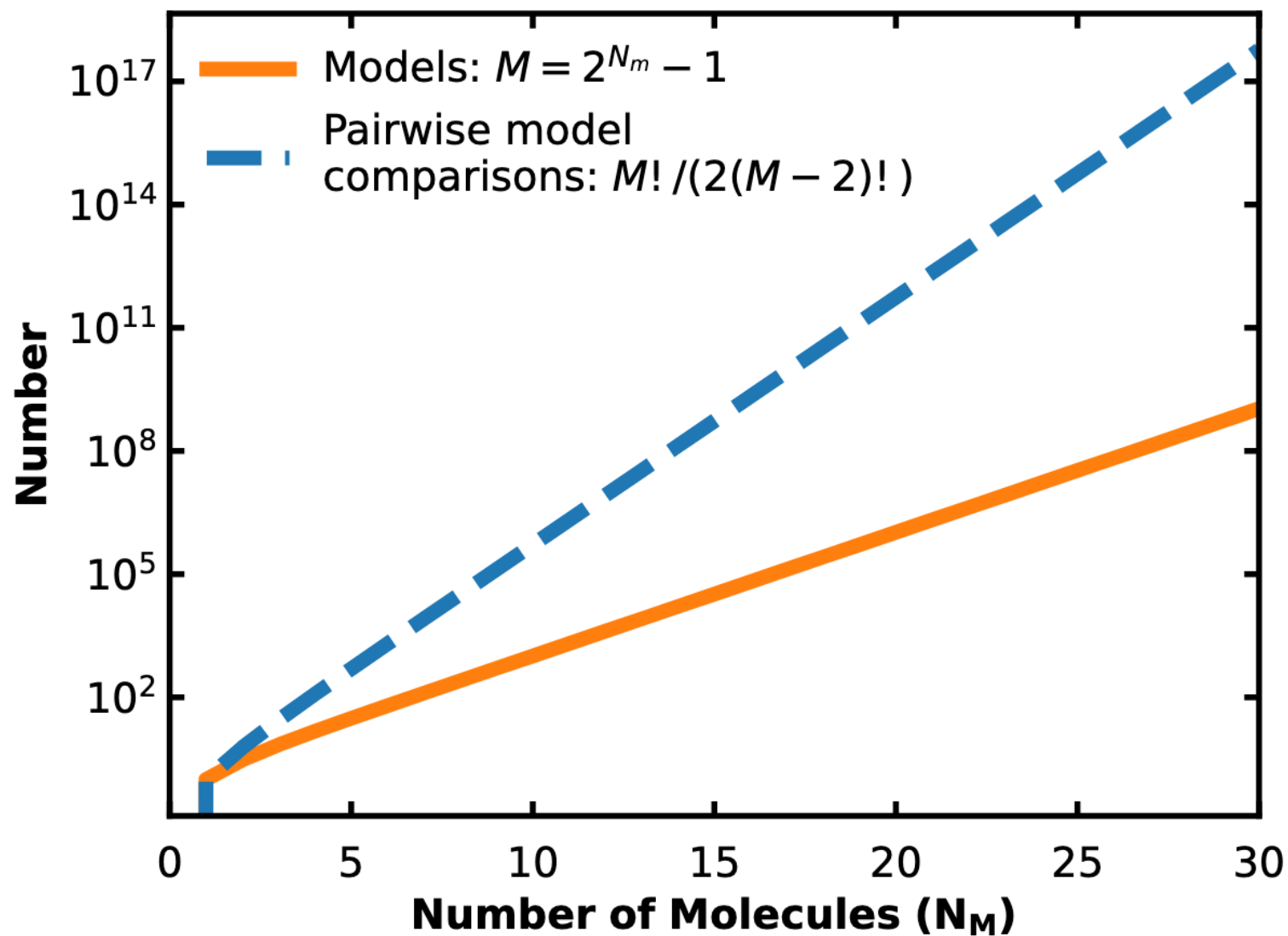
The problem with transit spectroscopy?





Forward Model

21%   
2%   



(Welbanks *et al.*, 2025)

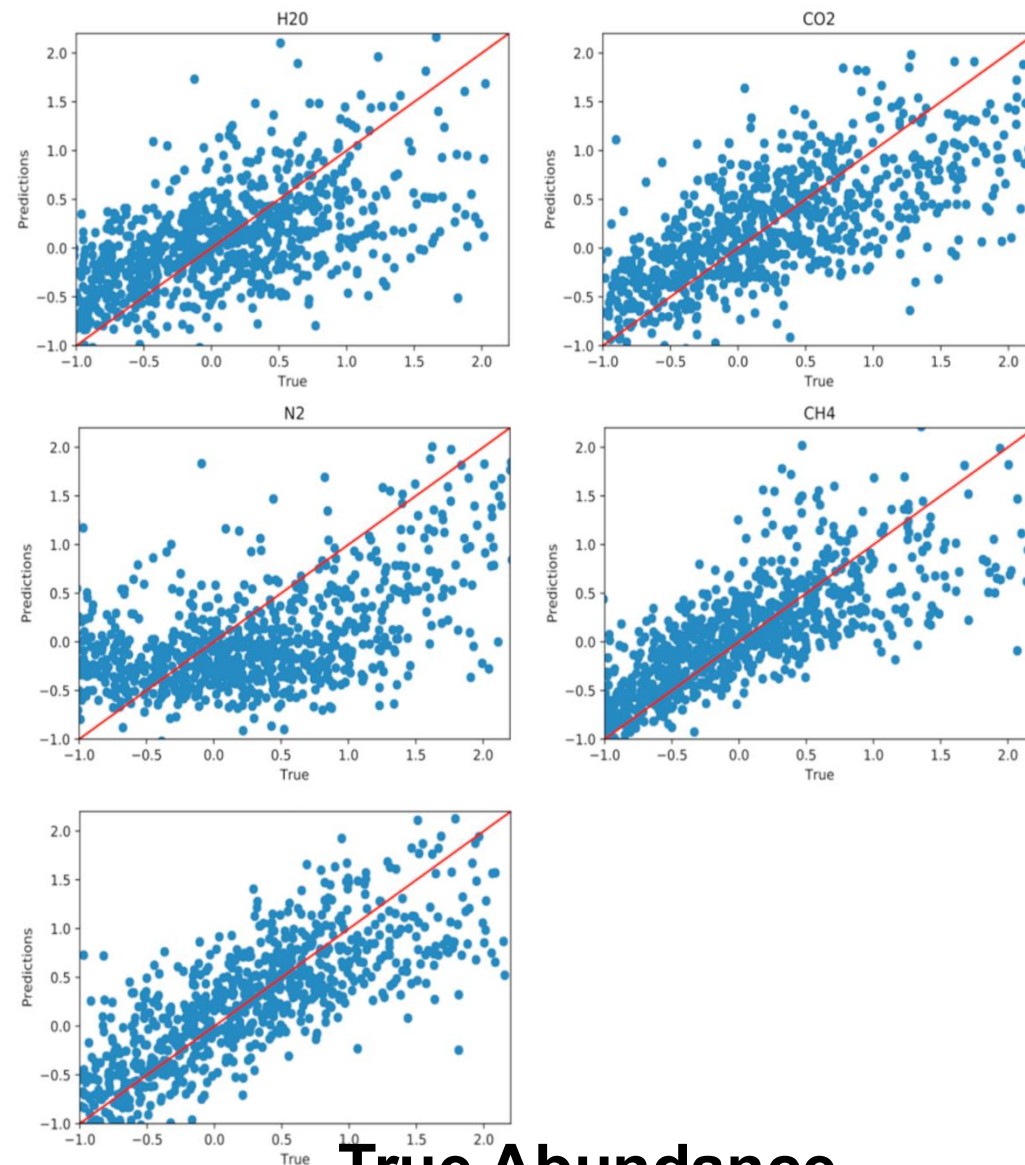
INARA

Comprehensive grid search
over model architectures and
hyperparameters

Established **1D CNNs** as best
model architecture

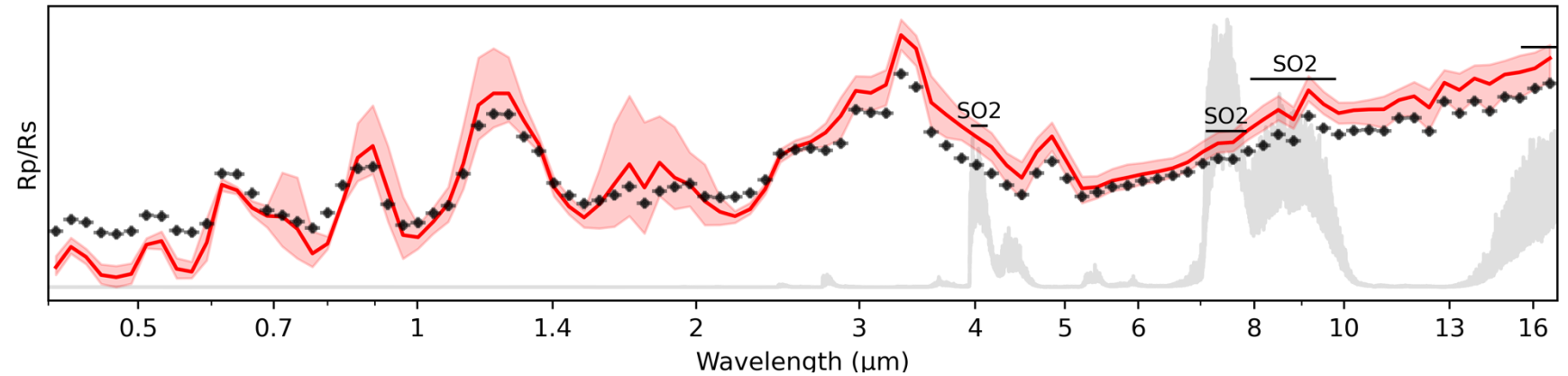
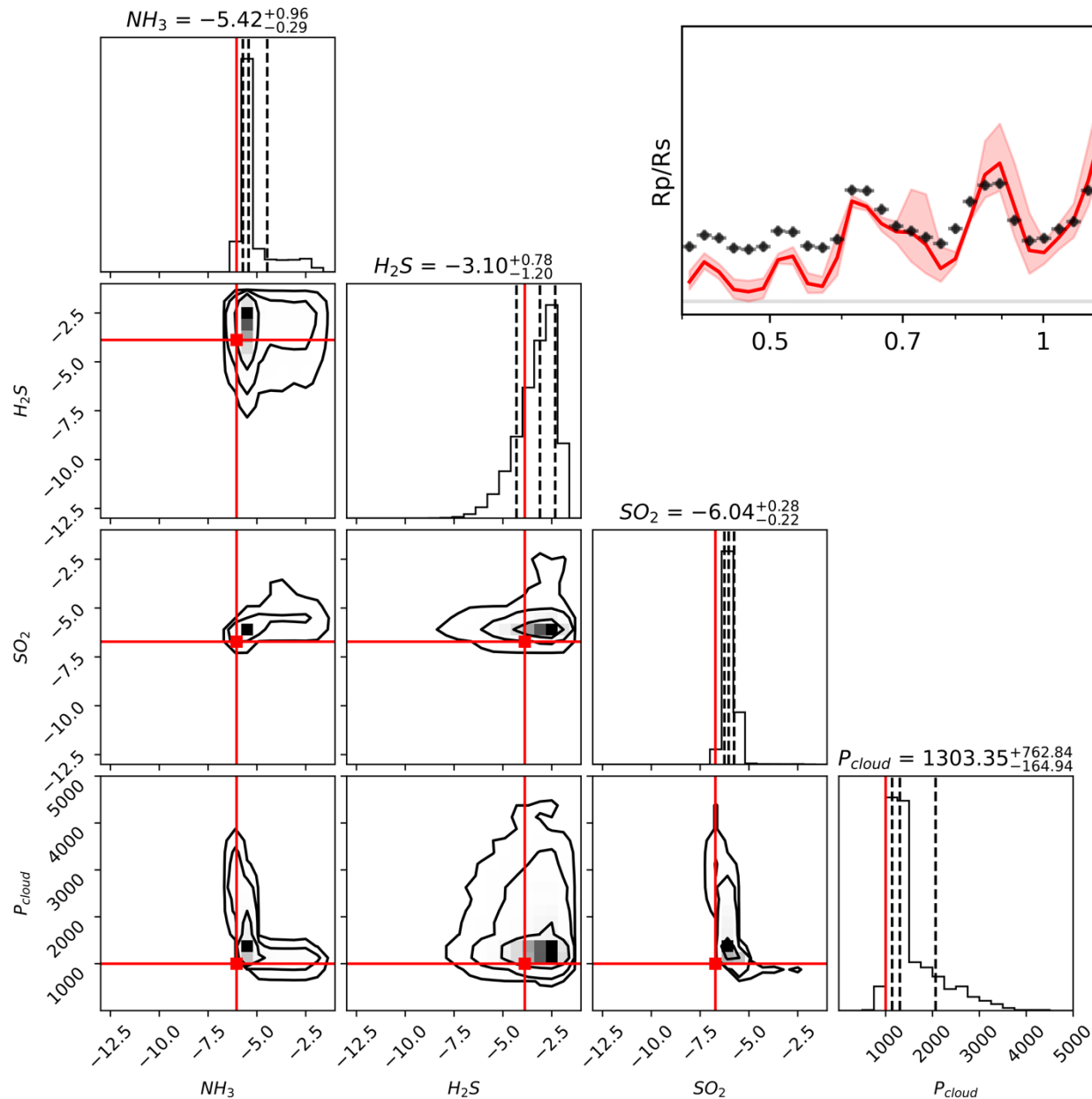
Significantly **accelerated**
compared to traditional
Bayesian retrievals

Prediction



True Abundance

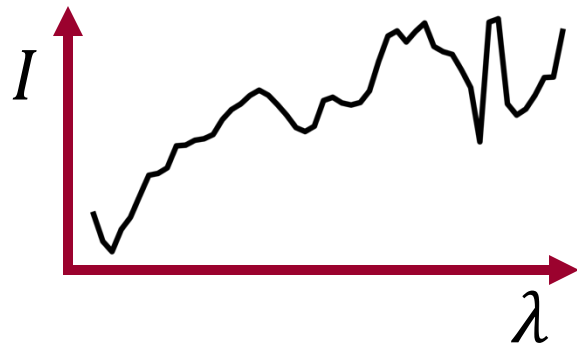
(Soboczenski *et al.*, 2018)



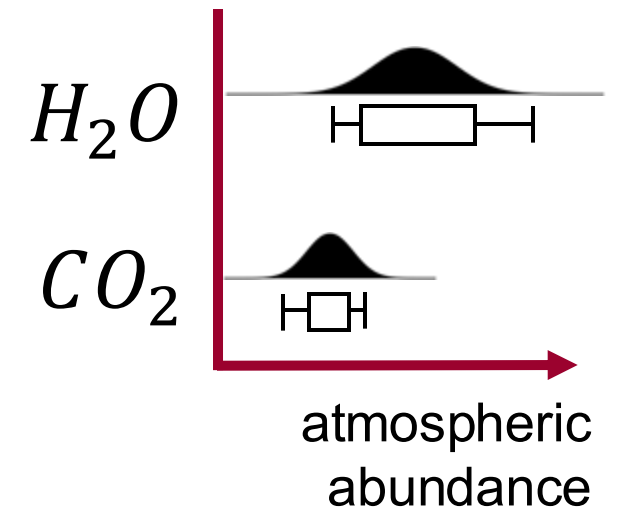
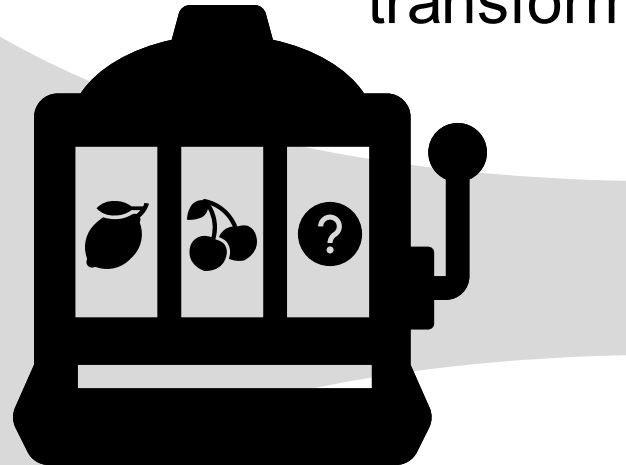
PERTURB

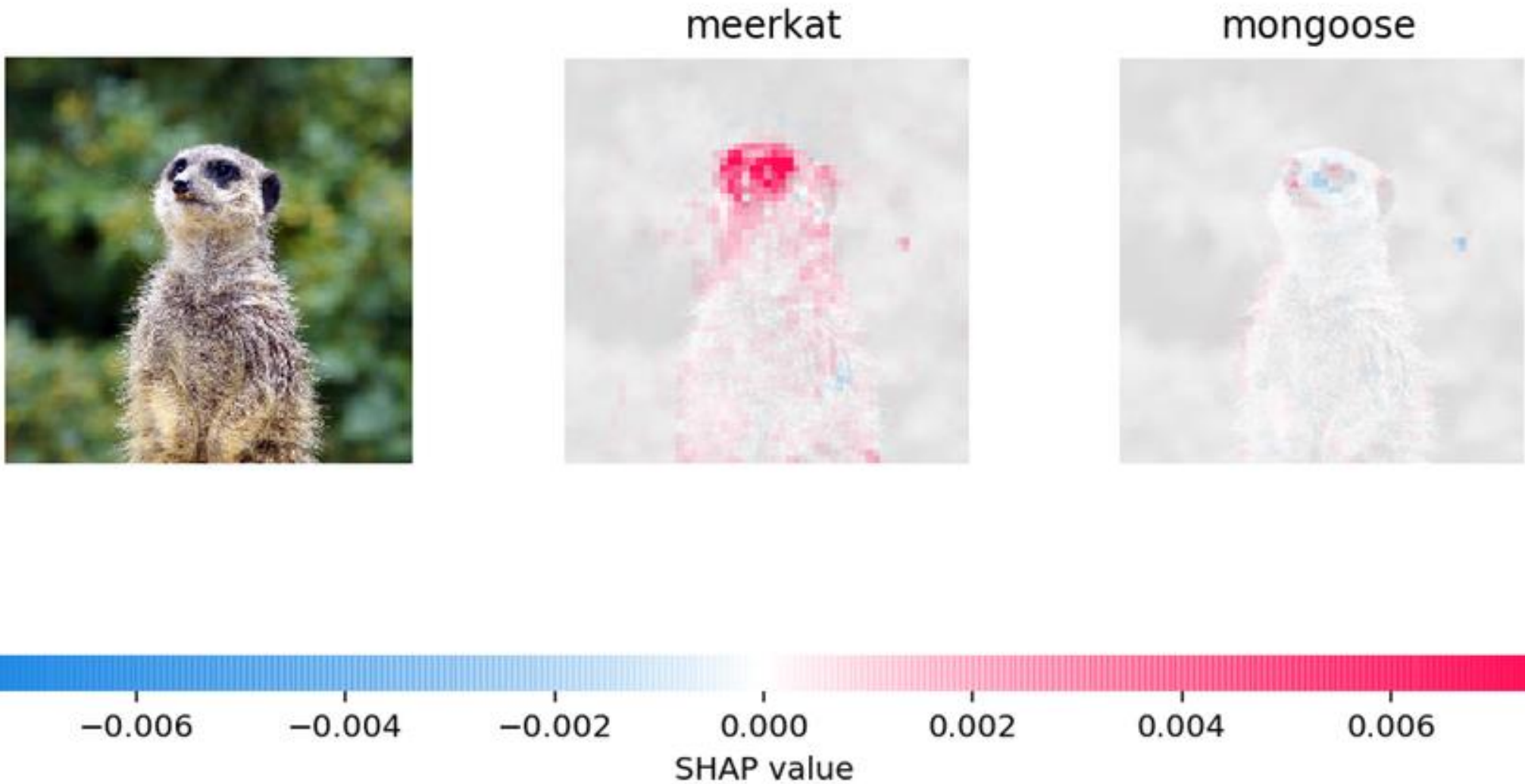
Accuracy of these retrievals
 on simulated data is
sufficient for making useful
 statements about a system

(Clarke *et al.* 2025b, in prep.)

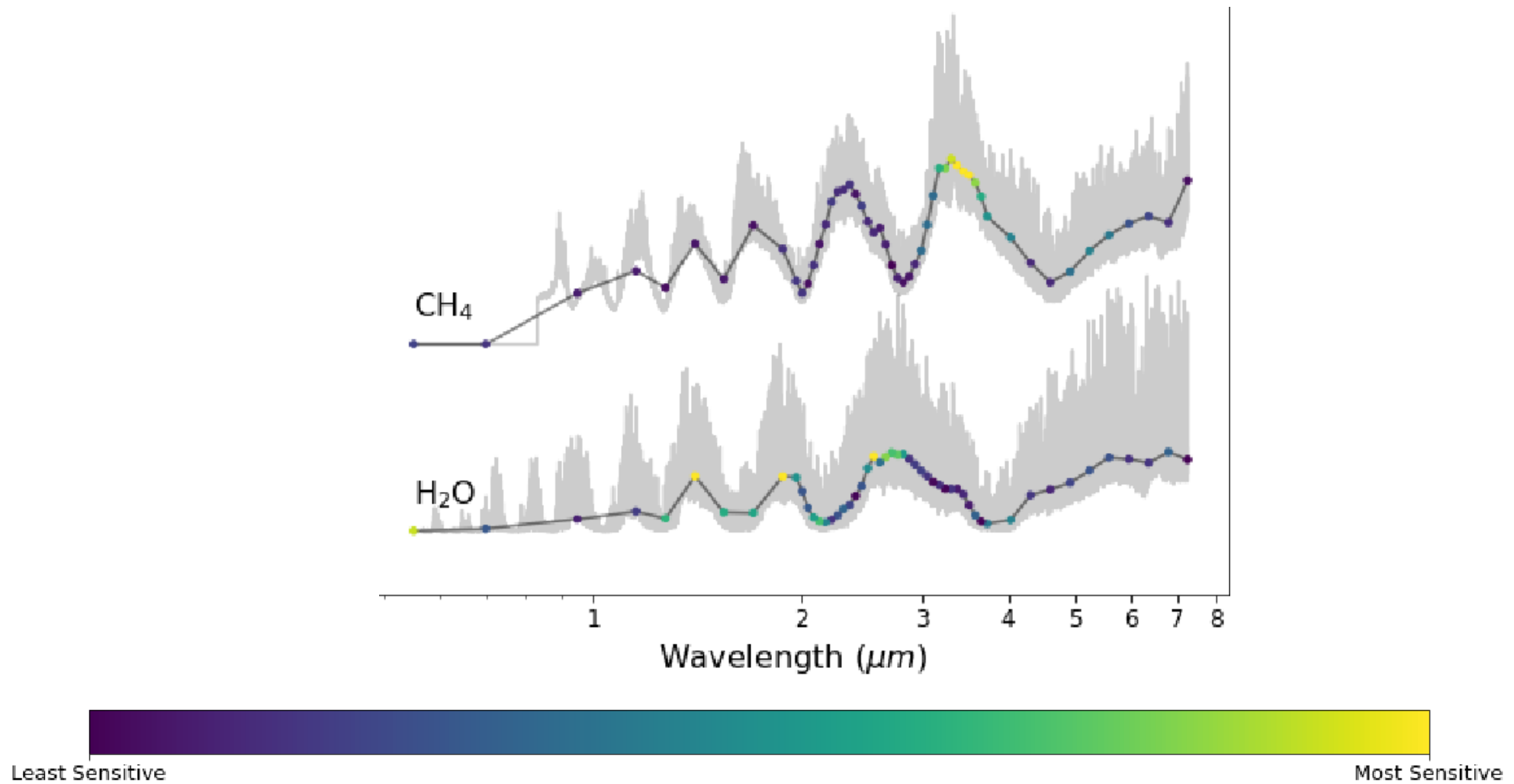


Set of **non-linear** matrix
transformations

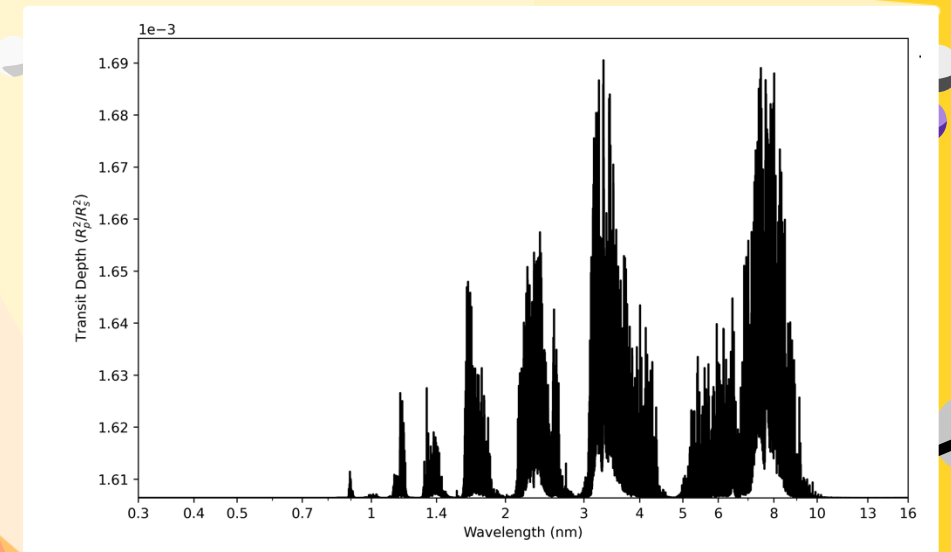
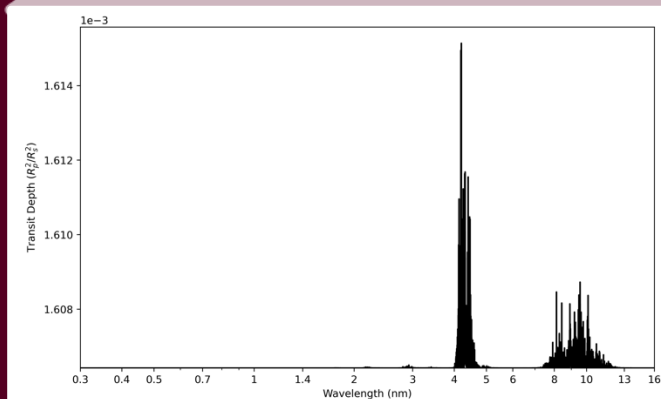
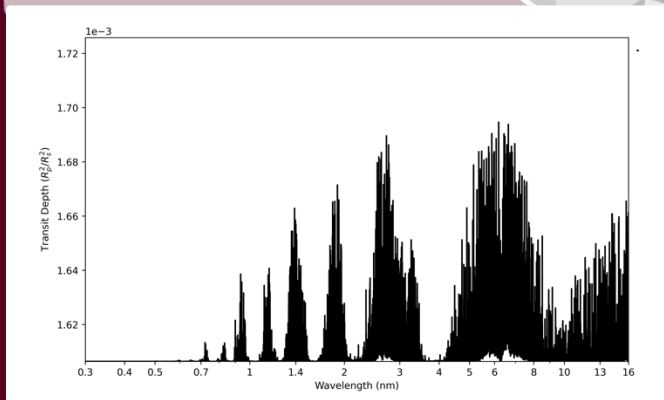
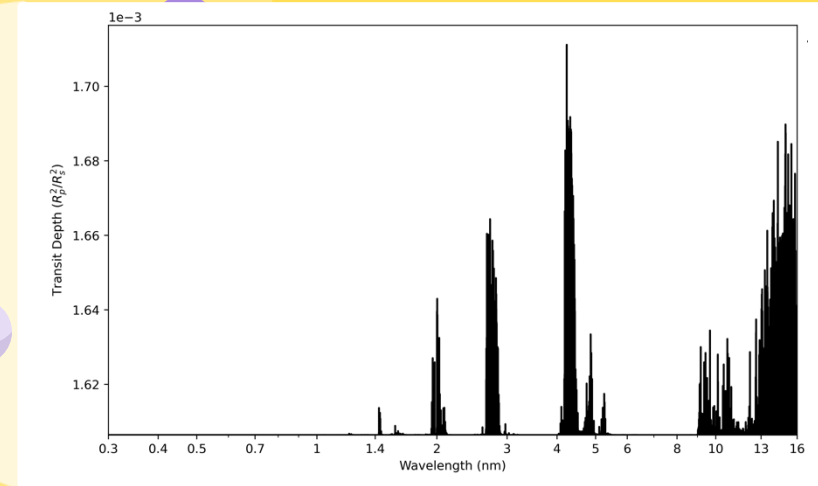
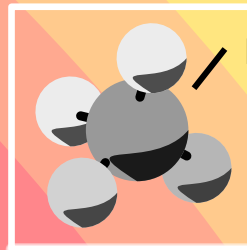
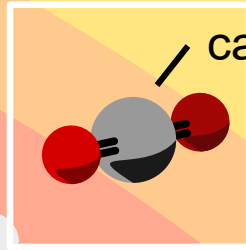
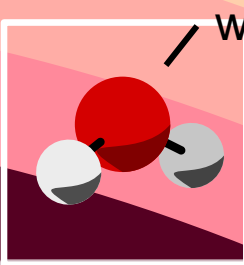




SHAP paper: (Lundberg and Lee, 2017), image: (Shojaeinasab *et al.*, 2024)



(Yip *et al.*, 2021)



Introducing



P hysical
E xplainability
R anking
T echniques for
U nderstanding
R etrieval
B lack-boxes

A new **lightweight** method for ML retrieval model interpretability based on **noise injection** and **response regression**.

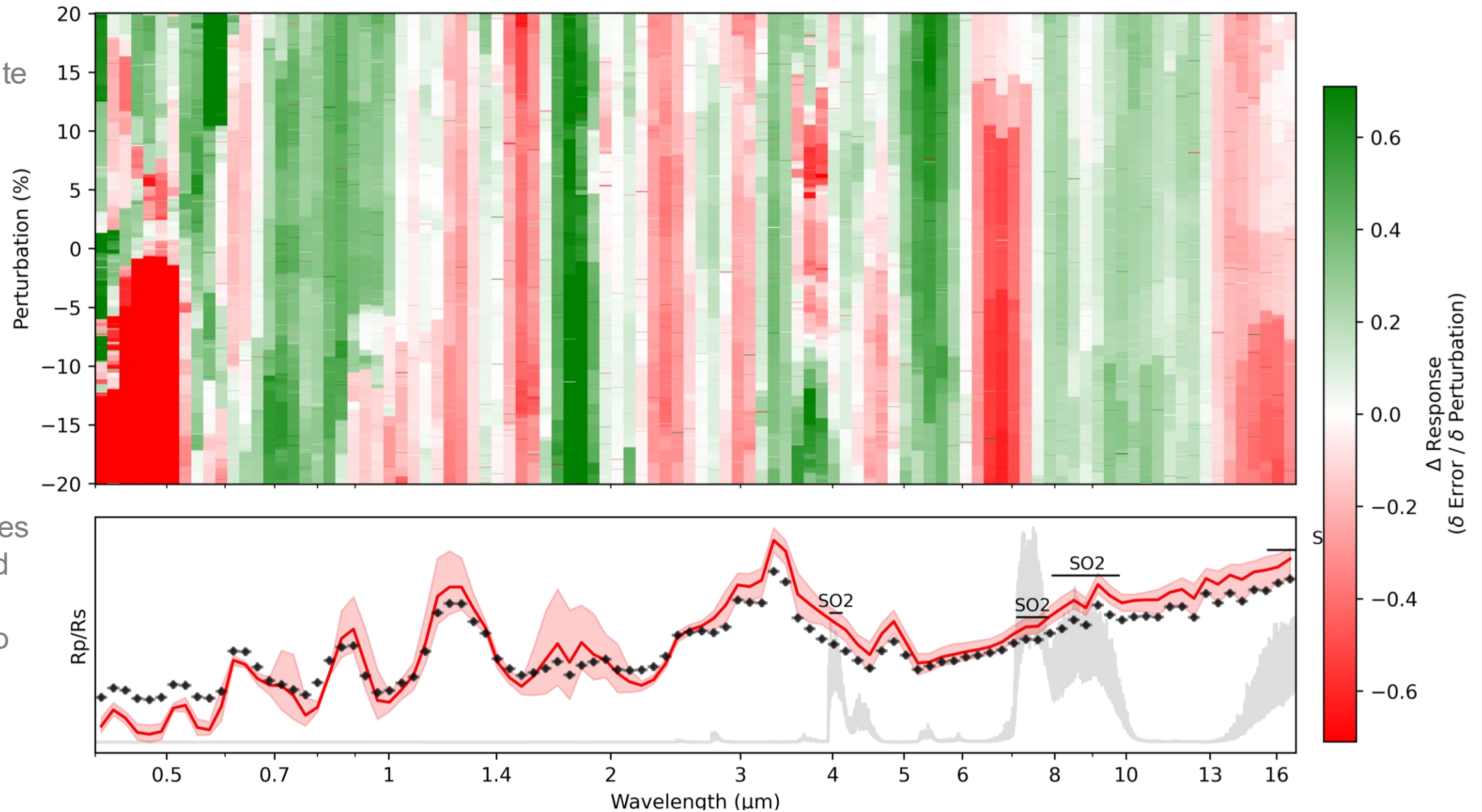
Built to be **modular** and **tunable**

Model agnostic, can be applied to any retrieval methodology

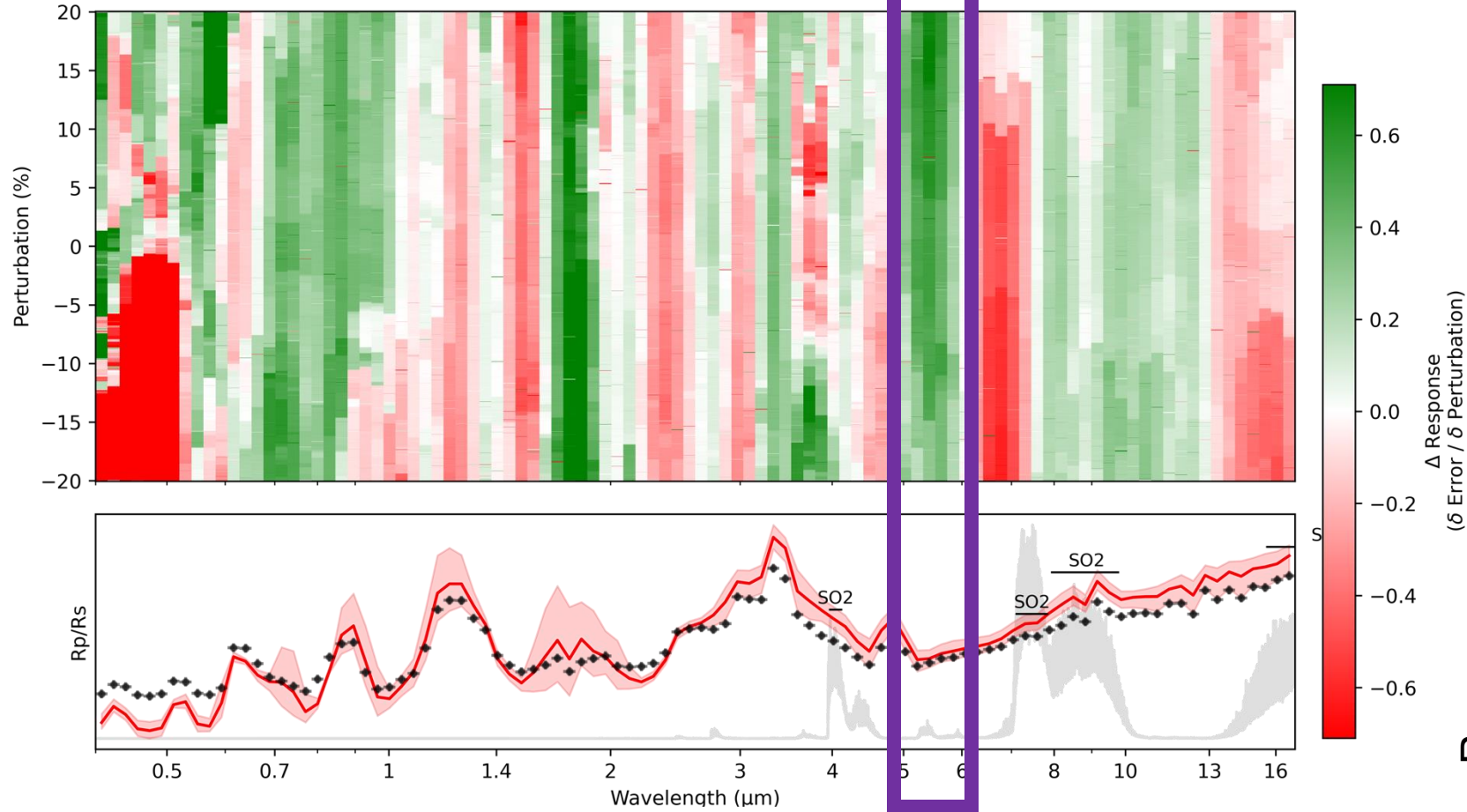
(Clarke *et al.* 2025a, in prep.)



Fast to
generate

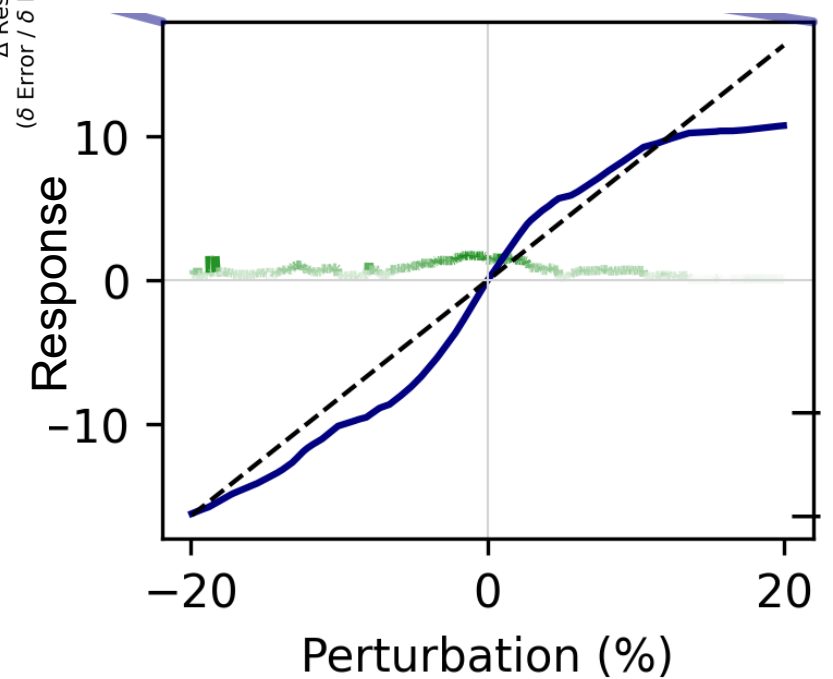


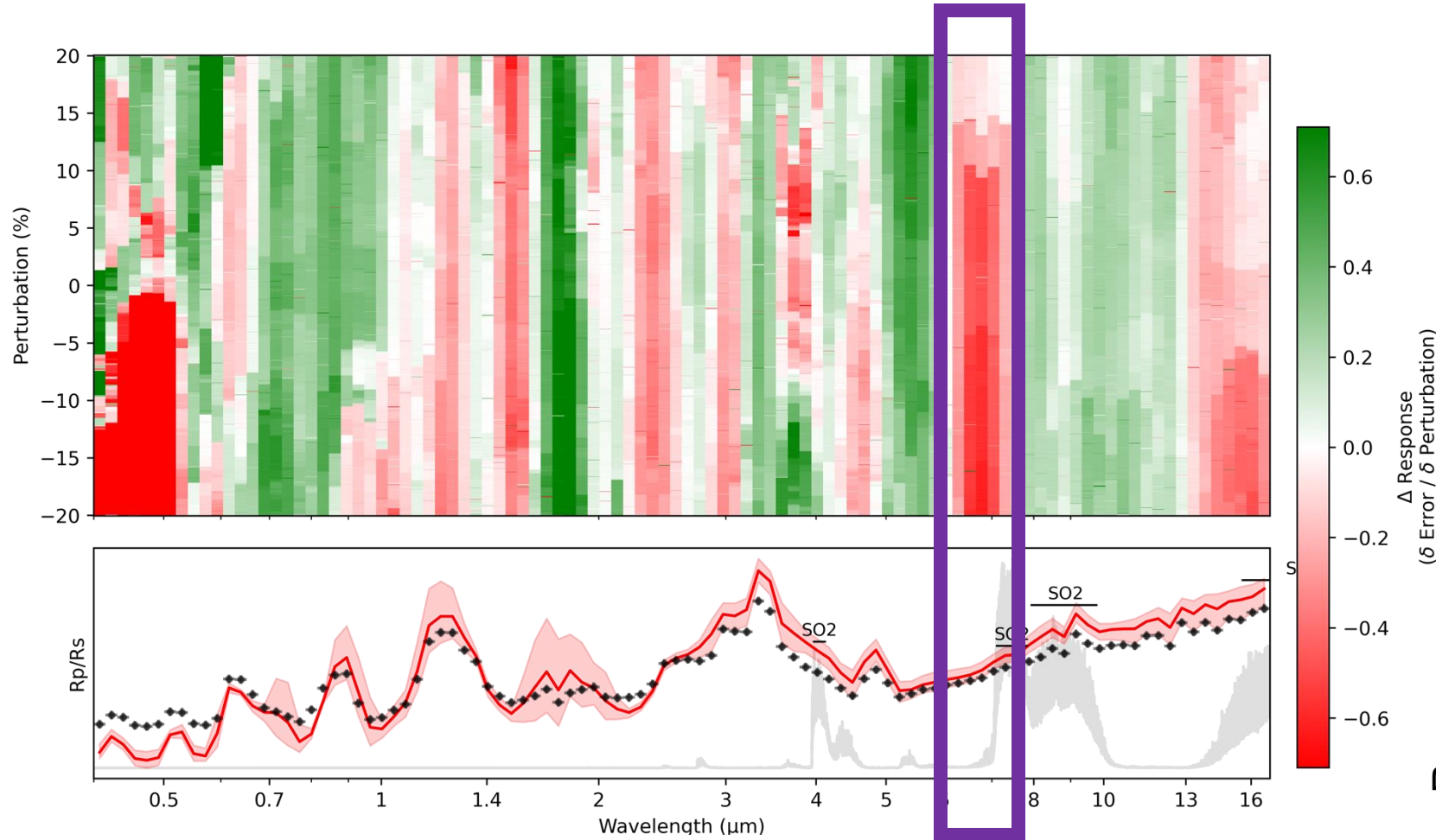
Requires
forward
model
calls so
slow



Positive response:

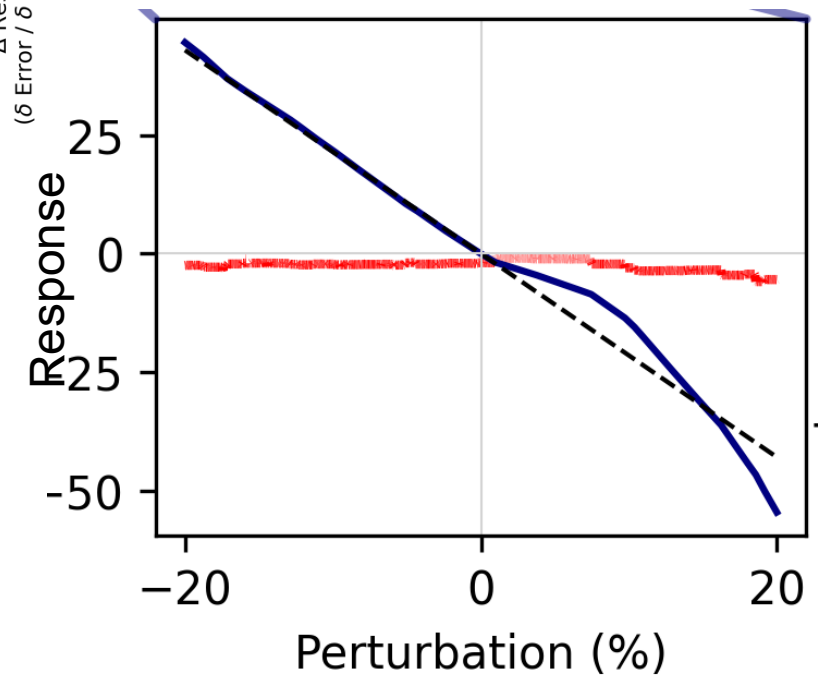
Underestimating these datapoints results in **underestimated abundance** AKA where the information is

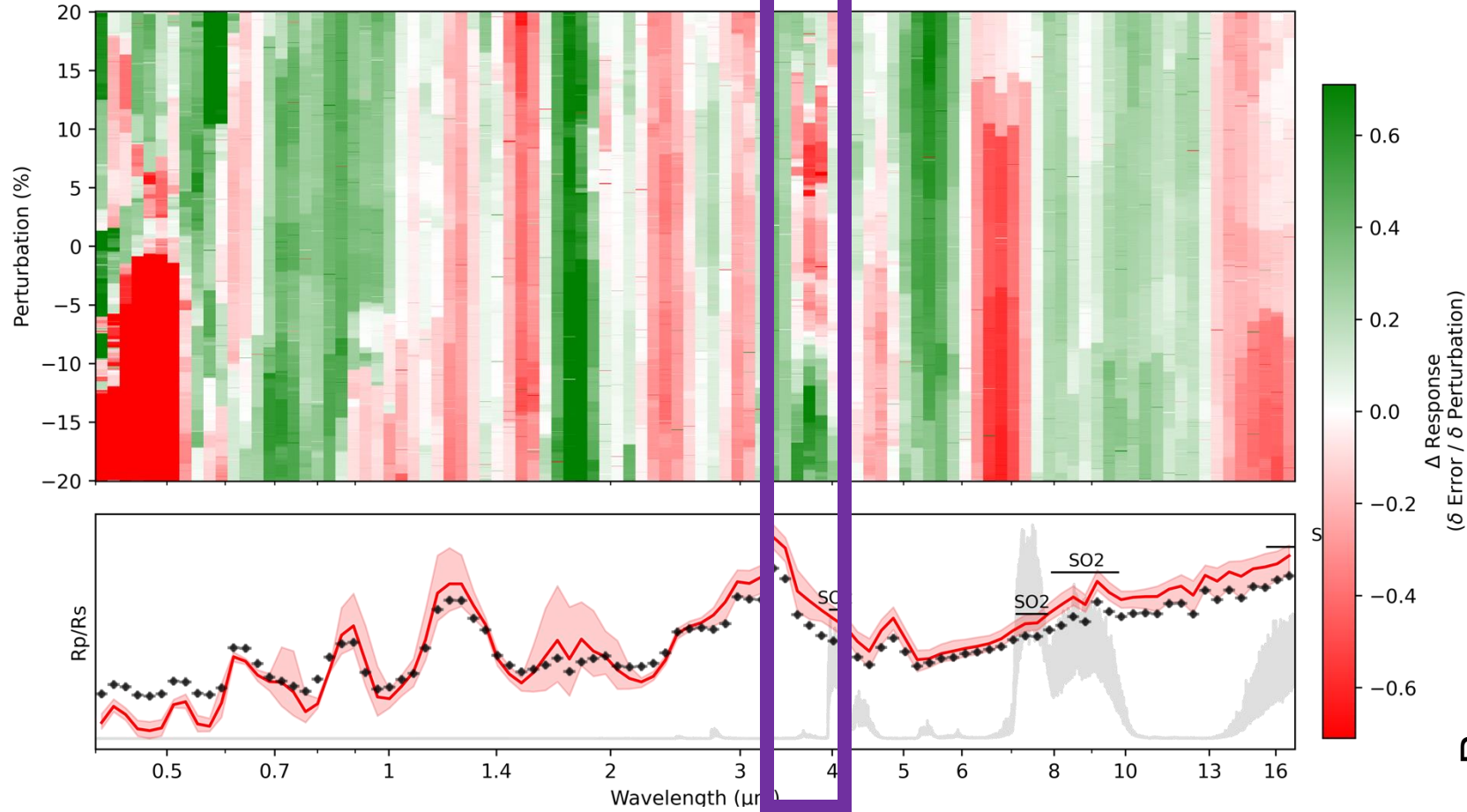




Negative response:

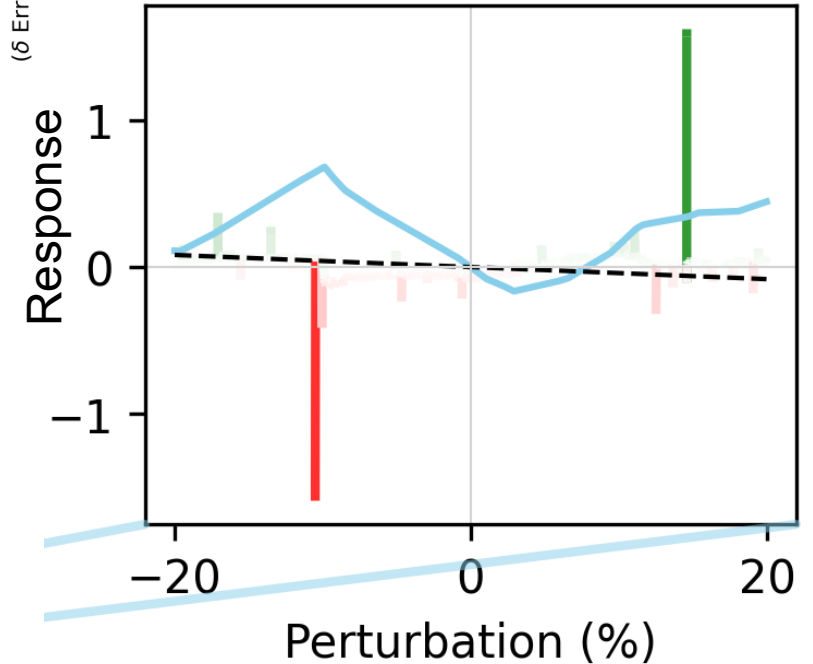
Underestimating these
datapoints results in
overestimated abundance
AKA where the information isn't





Non-Linear response:

Retrieved abundance is very **sensitive** to any data uncertainty in this region, likely



Often where the fit is poor, areas of degeneracy, or where the model is against limits

**Help shape PERTURB for your
research! Fill out the survey**

or read more about the project



**Jools Clarke¹, Gordon Yip², and Nikos
Nikolaou³**

¹ jools.clarke.23@ucl.ac.uk

² kai.hou.yip@ucl.ac.uk

³ n.nikolaou@ucl.ac.uk

*With thanks to **Sushuang Ma** for the data*

