Atmospheric interpretation from cross correlation maps

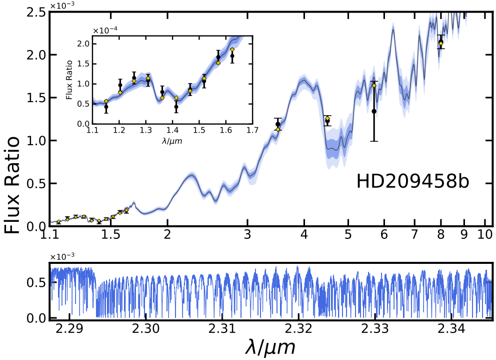
Sid Gandhi[1,2]

Siddharth.Gandhi@warwick.ac.uk

24th August 2021

[1]Department of Physics, University of Warwick, Coventry, CV4 7AL, UK
[2]Centre for Exoplanets and Habitability, University of Warwick, Gibbet Hill Road, Coventry CV4 7AL, UK

Low- vs High-Resolution Spectroscopy



For Low Resolution Spectra:

- Retrievals have been ubiquitous for over a decade
- Robustly detected and constrained species such as H_2O

We want to set up something similar with ground based high resolution spectra.

Low- vs High-Resolution Spectroscopy

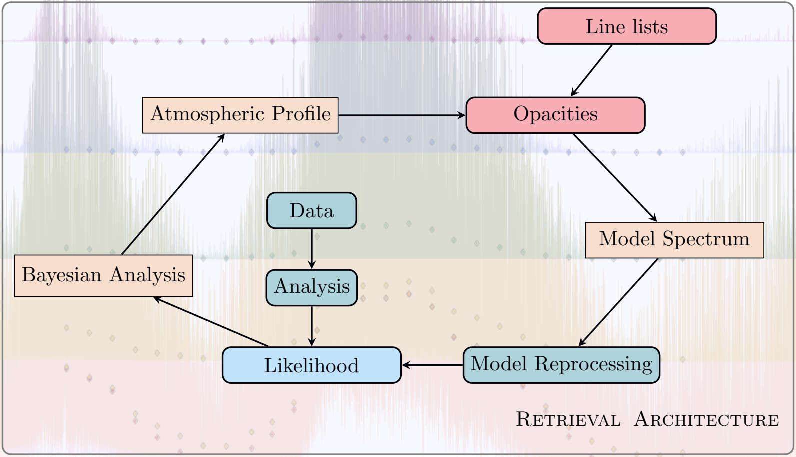
Low Resolution

- Retains continuum information
- Clear features can be seen
- Degeneracies occur due to overlapping bands
- Generally sensitive to a narrow range of pressures

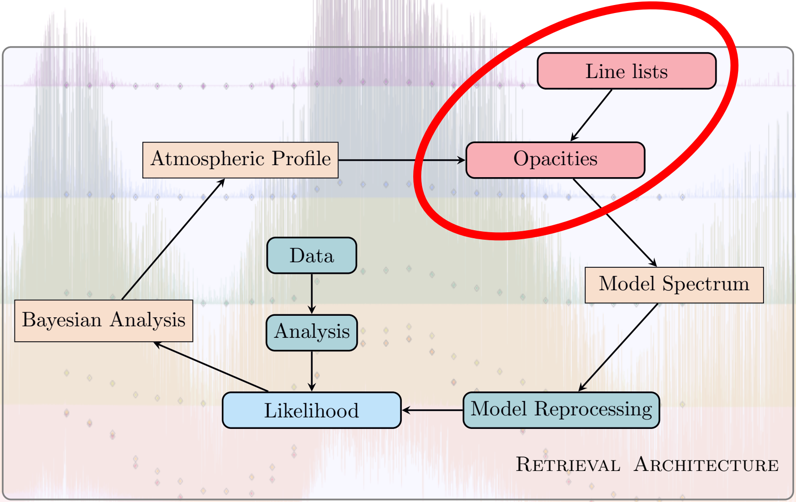
High Resolution

- Avoids degeneracies between species
- Probes wider range of altitudes
- Can lose some continuum information
- Can be difficult to extract planetary signal due to noise

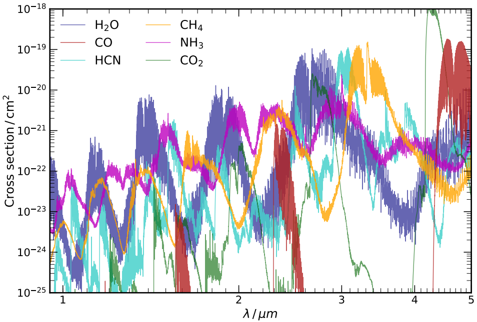
High Resolution Retrieval Architecture



High Resolution Retrieval Architecture



Opacity Considerations



As we are correlating many individual spectral lines, it is important that they are accurate and their profiles are well understood.

- The choice of line lists can make a big difference to whether we detect a species/get biased abundances.
- We also need accurate pressure broadening coefficients to correctly model the line profile.

How do we calculate the cross section?

For *every* line in the line list (can be $\gtrsim 10^{10}$ for some!) we must calculate the broadened profile at *every* pressure and temperature.

The pressure broadens the profile into a Lorentzian - the broadening width depends on the quantum numbers of each line and the broadening medium (e.g. H_2- or air-broadening).

$$f_L(\nu - \nu_0) = \frac{1}{\pi} \frac{\gamma_L}{(\nu - \nu_0)^2 + \gamma_L^2},$$

$$\gamma_L = P \sum_i \left(\frac{T_{\mathrm{ref}}}{T} \right)^{n_i} \gamma_i X_i$$

How do we calculate the cross section?

The temperature also broadens the profile into a Gaussian - caused by the velocity distribution of particles of each gas. This width depends on the molecular weight of the species we are trying to calculate the cross section for.

$$f_G(\nu - \nu_0) = \frac{1}{\gamma_G \sqrt{\pi}} \exp\left(-\frac{(\nu - \nu_0)^2}{\gamma_G^2}\right),$$

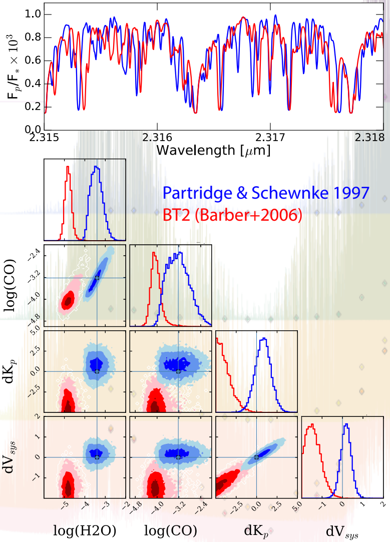
$$\gamma_G = \sqrt{\frac{2k_b T}{m}} \frac{\nu_0}{c},$$

The overall signal is a convolution of the Gaussian and Lorentzian, a *Voigt* profile:

$$f_V(\nu - \nu_0) = \int_{-\infty}^{\infty} f_G(\nu' - \nu_0) f_L(\nu - \nu') d\nu'.$$

The sum of the contributions from each line is the resultant cross section.

Lines Lists for HRS



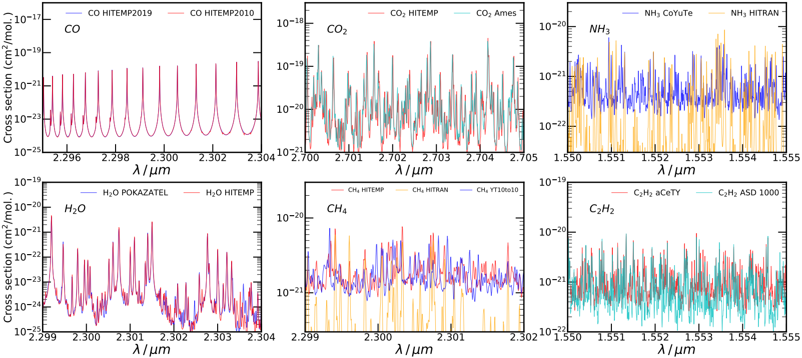
The choice of line list can significantly alter the spectrum at R∼100,000.

This can influence not only the detection of the species but also have a knock on effect on other parameters.

New high resolution lines lists from databases such as HITEMP and ExoMol often use ab initio calculations combined with empirical observations for accuracy.

Brogi and Line, *AJ, 157, 114 (2019)*

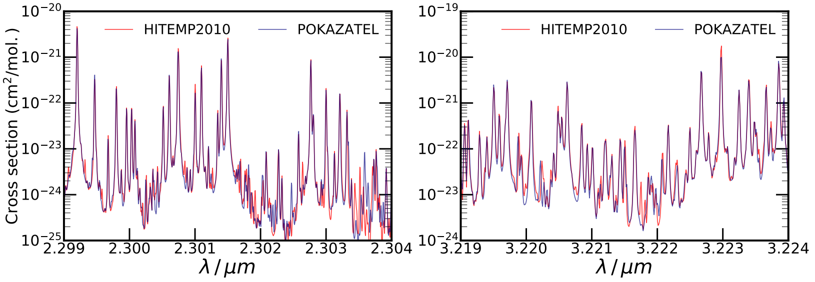
Line Lists for HRS



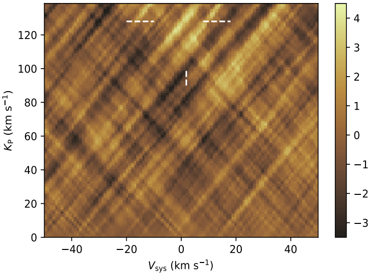
H_2O, CO and CO_2 line lists agree pretty well.

Newer high resolution line lists for CH_4, NH_3 and also C_2H_2 have very recently become available.

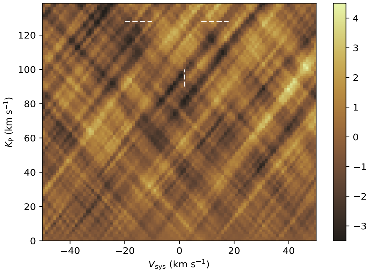
H_2O: HITEMP2010 vs POKAZATEL ExoMol



CH_4 in HD102195b



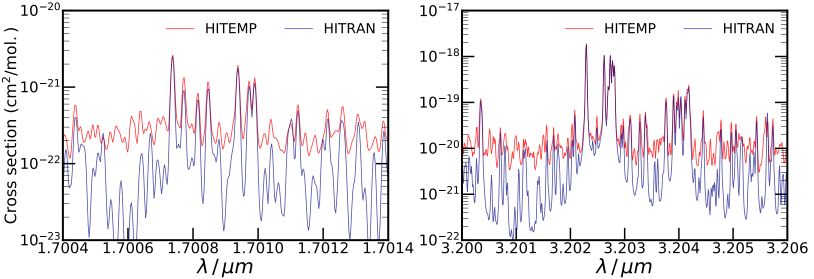
HITRAN2016



HITEMP

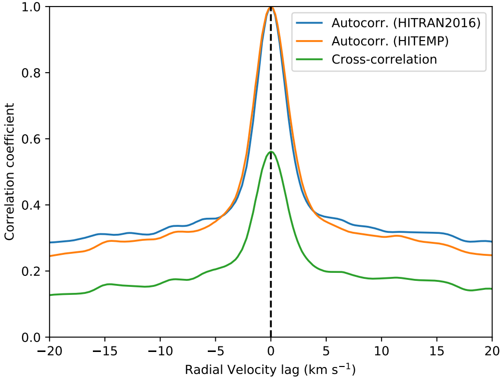
The HITEMP line list shows no significant signal for CH_4.

CH_4: HITEMP vs HITRAN



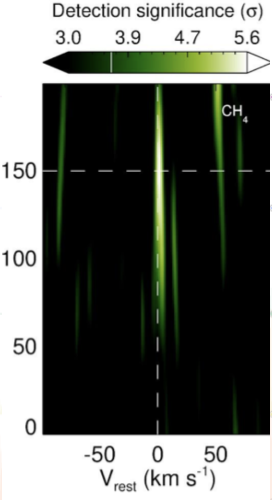
The strongest lines agree, but at high temperature the weaker lines which form the continuum are missed in HITRAN.

Correlation between spectra



- The cross correlation between spectra is only ∼0.55.
- The peak is aligned with the autocorrelation, indicating that there is no net offset between the line lists.

CH_4 in HD209458b



CH_4 has recently been detected in the atmosphere of HD209458b.

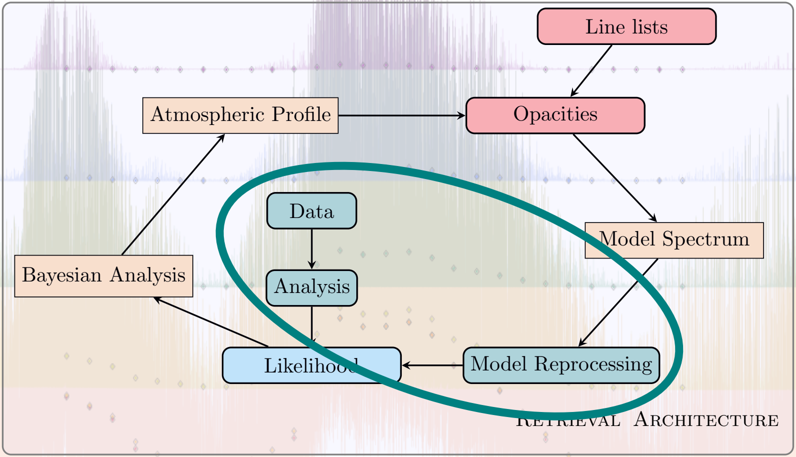
This was detected with 4 nights of primary transit observations using GIANO on the TNG.

The observations covered the H and K bands, where CH_4 has prominent opacity bands.

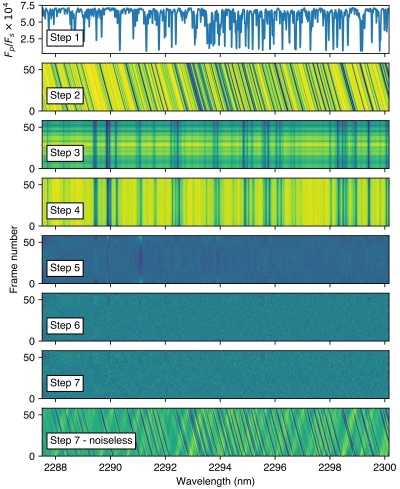
This detection was based on the HITEMP line list, using HITRAN the signal was much weaker (3.7σ).

Giacobbe et al., *Nature, 592, 204 (2021)*

High Resolution Retrieval Architecture



Data Analysis and Model Reprocessing



Brogi and Line, *AJ, 157, 114 (2019)*

As we saw from Jens and Matteo's talk yesterday, the observations contain many strong noisy signals which must be removed/accounted for. These include:

- The stellar signal (we measure $F_p + F_{star}$ for emission spectra).
- Telluric absorption features from species such as H_2O, CH_4 and CO_2.
- Airmass variation over the transit.

Data Analysis and Model Reprocessing

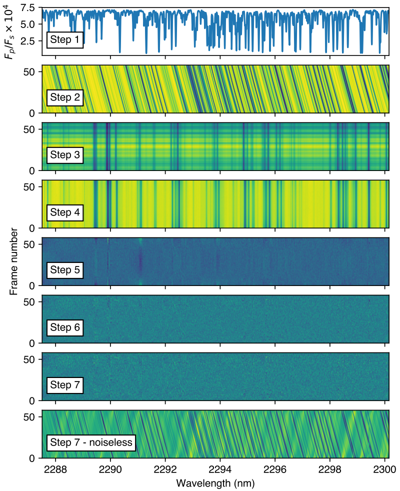
We remove these from the observations through masking/SVD as well as other data analysis techniques.

Removing this from the observations can also affect the the signal we want from the planet, F_p (or transit depth).

These can skew the planetary signal, reduce/rescale it and result in other artefacts that may result in a model spectrum not correlating as strongly.

Therefore, we must incorporate the same processing steps into the model in order to effectively match the model to the observations.

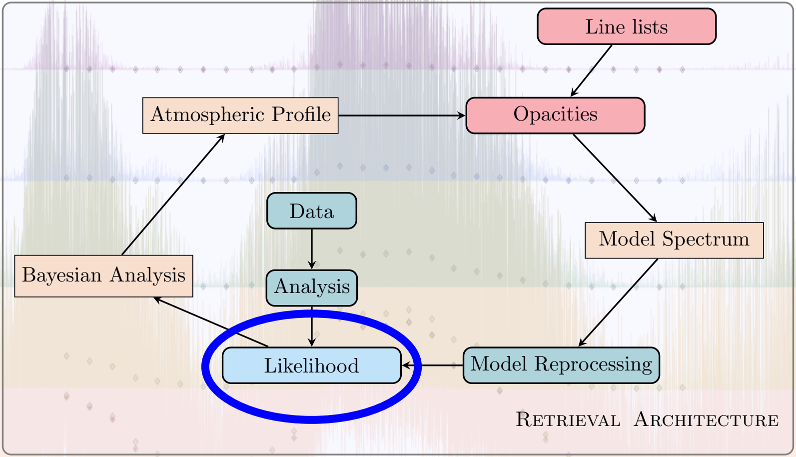
Data Analysis and Model Reprocessing



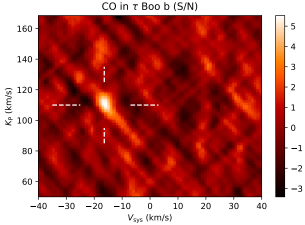
Brogi and Line, *AJ, 157, 114 (2019)*

1. Injected Model.
2. Shifted Model according to K_p and V_{sys}.
3. Telluric, airmass and instrumental throughput.
4. Wavelength Calibration.
5. Spectrum is averaged in time and mean fitted.
6. Temporal variability is fitted and divided out.
7. Noisy columns masked.

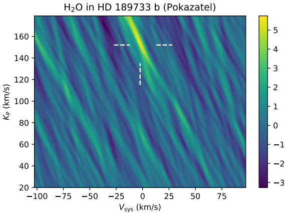
High Resolution Retrieval Architecture



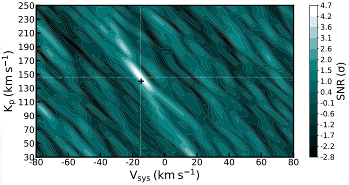
Detections with High Resolution Spectroscopy



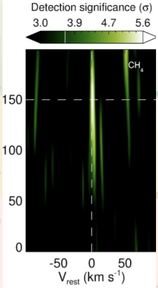
CO in τ Boo b



H_2O in HD 189733 b



HCN in HD 209458 b



CH_4 in HD 209458 b

These detections use the cross-correlation method.

Cross-Correlation

$$C(s) = \frac{\frac{1}{N}\sum_n f(n)g(n-s)}{\sqrt{\frac{1}{N}\sum_n f^2(n) \times \frac{1}{N}\sum_n g^2(n-s)}}$$

Behaviour

- A well matched model will result in C$\rightarrow$1 at some s.
- A poorly matched model will result in C very close to 0.

Pros

- Very easily allows us to determine which models and thus which species match the observations.

Cons

- Removes the scale of a model as the value is normalised by denominator.

Retrievals - Likelihood values

For low resolution spectra, we use the χ^2 value in the likelihood calculations.

$$\ln L = -\frac{1}{2}\chi^2 = -\frac{1}{2}\sum_i \frac{(y_i - y_{model,i})^2}{\sigma_i^2}$$

We need to convert our $C(s)$ into a likelihood for HRS.

I will briefly discuss 3 ways to do this, using the Zucker (2003) approach, the Brogi and Line (2019) approach and the Gibson et al (2020) approach.

Zucker Approach

We define

$$\ln L = -\frac{N}{2} \ln(1 - C^2)$$

Zucker, *MNRAS, 342, 1291 (2003)*

Behaviour

- A well matched model will result in $\ln L \to$ large and positive.
- A poorly matched model will result in $\ln L$ very close to 0.

Pros

- Easy to implement.
- Used already for stellar RVs.

Cons

- Anti-correlation cannot be distinguished from correlation (e.g. thermal inversions).
- We are still invariant of the scaling of the model.

Brogi and Line Approach

We define

$$\ln L = -\frac{N}{2} \ln \left(\frac{1}{N} \sum_n g^2(n-s) + \frac{\alpha^2}{N} \sum_n f^2(n) - \frac{2\alpha}{N} \sum_n f(n) g(n-s) \right)$$

Brogi and Line, *AJ, 157, 114 (2019)*

Behaviour

- A well matched model and variance $\ln L \to$ large and $+$ve.
- A poorly matched model/variance will be penalised.

Pros

- Model variance/scaling now accounted for.
- We are sensitive to the sign of the correlation.

Cons

- Does not account for wavelength-dependent noise.

Gibson et al Approach

$$\ln L = -\frac{N}{2}\ln\Big(\frac{1}{N}\sum_n \frac{g^2(n-s)}{\sigma_n^2} + \frac{\alpha^2}{N}\sum_n \frac{f^2(n)}{\sigma_n^2} - \frac{2\alpha}{N}\sum_n \frac{f(n)g(n-s)}{\sigma_n^2}\Big)$$

Gibson et al., *MNRAS, 493, 2215 (2020)*

Behaviour

- Similar to Brogi and Line 2019 method
- High uncertainty terms/noisy parts of the spectrum are penalised.

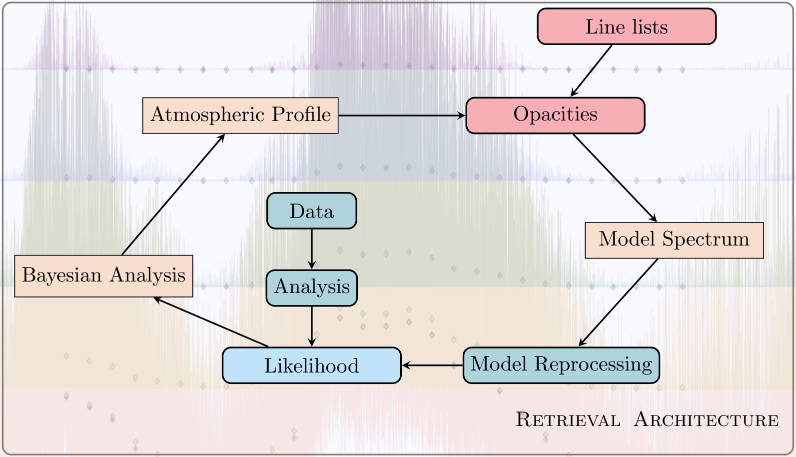
Pros

- Now includes wavelength dependence on noise.
- Allows for a scaling in the noise structure.

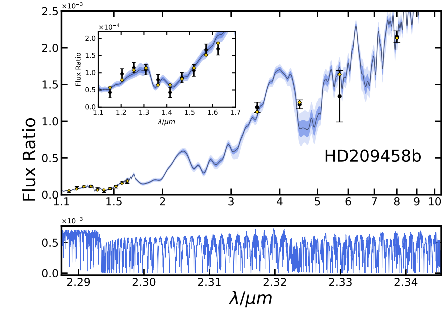
Cons

- Potentially need additional parameter β.

High Resolution Retrieval Architecture



Other Modelling Considerations



For HRS observations we must generate line-by-line models at $R \gtrsim 100{,}000$-$300{,}000$ in order to compare against the data.

- $\sim$100-1000$\times$ increase in R over HST spectra.
- Spectrographs cover a wide range in wavelength.
- Probe a much wider range of altitudes, so number of modelled atmospheric layers (P-T values) is also important.

Hybrid Retrievals

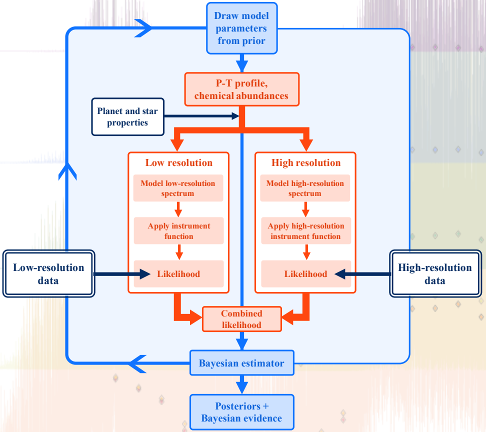
Once we are able to get a $\ln L$ value for HRS, it is straightforward to retrieve high resolution and low resolution spectra simultaneously.

$$\ln L_{tot} = \ln L_{HRS} + \ln L_{LRS}$$

Significantly,

- This combines the advantages of both approaches.
- Allows for robust detections and constraints.
- Also allows us to compare the differences between each approaches (if e.g. they probe different altitudes in the atmosphere).

Hybrid Retrievals

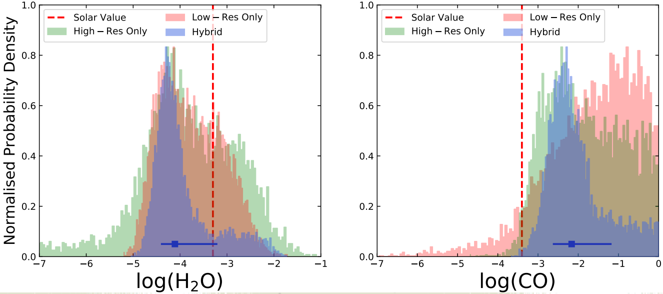


We tested this combined retrieval on low-resolution HST+Spitzer and high-resolution CRIRES/VLT observations of the dayside of HD209458b.

This uses the method from Brogi and Line, *AJ 157, 114 (2019)* to map the cross correlation onto a likelihood.

Gandhi et al., *AJ, 158, 228 (2019)*

Abundance Constraints

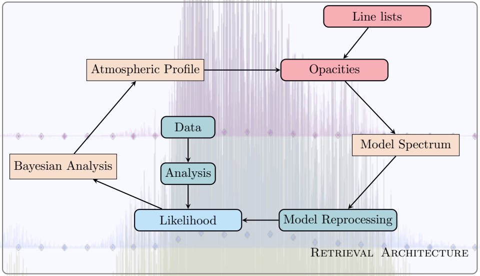


The combined constraints are more stringent than each individual dataset alone.

The detection significances for H_2O and CO have also increased in the hybrid retrieval.

These give us much more precise constraints on atmospheric properties such as the C/O ratio and metallicity, key parameters in formation models.

Conclusions



- We must carefully consider the opacity sources for HRS.
- The data analysis and model reprocessing are crucial in extracting planetary signatures.
- There are various ways of converting between CC to ln L. These allow for simultaneous retrieval of both LRS and HRS data.

Thanks for listening! Any questions?