Citation


URL

https://oro.open.ac.uk/82145/

License

(CC-BY 4.0) Creative Commons: Attribution 4.0

https://creativecommons.org/licenses/by/4.0/

Policy

This document has been downloaded from Open Research Online, The Open University’s repository of research publications. This version is being made available in accordance with Open Research Online policies available from Open Research Online (ORO) Policies

Versions

If this document is identified as the Author Accepted Manuscript it is the version after peer review but before type setting, copy editing or publisher branding
A retrieval challenge exercise for the Ariel mission

Joanna K. Barstow1 · Quentin Changeat2 · Katy L. Chubb3 · Patricio E. Cubillos4 · Billy Edwards2 · Ryan J. MacDonald5 · Michiel Min3 · Ingo P. Waldmann2

Received: 30 July 2020 / Accepted: 29 November 2021 / Published online: 9 February 2022 © The Author(s) 2022

Abstract
The Ariel mission, due to launch in 2029, will obtain spectroscopic information for 1000 exoplanets, providing an unprecedented opportunity for comparative exoplane
tology. Retrieval codes - parameteric atmospheric models coupled with an inversion algorithm - represent the tool of choice for interpreting Ariel data. Ensuring that reliable and consistent results can be produced by these tools is a critical preparatory step for the mission. Here, we present the results of a retrieval challenge. We use five different exoplanet retrieval codes to analyse the same synthetic datasets, and test a) the ability of each to recover the correct input solution and b) the consistency of the results. We find that generally there is very good agreement between the five codes, and in the majority of cases the correct solutions are recovered. This demonstrates the reproducibility of retrievals for transit spectra of exoplanets, even when codes are not previously benchmarked against each other.

Keywords First keyword · Second keyword · More

Joanna K. Barstow
jo.barstow@open.ac.uk

1 School of Physical Sciences, The Open University, Walton Hall, Milton Keynes MK7 6AA, UK
2 Department of Physics and Astronomy, University College London, Gower Street, WC1E 6BT London, UK
3 SRON, Sorbonnelaan 2, 3584 CA Utrecht, Netherlands
4 Space Research Institute, Austrian Academy of Science, Wolfgang BaumjohannSchmiedlstrasse 6, A-8042 Graz, Austria
5 Department of Astronomy, Cornell University, 404 Space Sciences Building, 14853 Ithaca, NY, USA
1 Introduction

In recent years, our knowledge of exoplanet atmospheres has been increasing rapidly. Recent highlights have included the detection of hazes/clouds in most exoplanets [1–4], the presence of ionised metals in the atmosphere of an ultra-hot Jupiter [5], and the discovery of water vapour in the atmosphere of a small, temperate planet [6, 7]. Studies of their 3-dimensional structures have also highlighted the complexity of these worlds [8–10]. In addition, atmospheric characterisation of directly imaged planets is providing interesting results [see e.g. 11].

The majority of planets for which we have obtained detailed atmospheric information transit their parent stars. Their atmospheres can be observed either during transit, when starlight passes through the limb of the atmosphere, or at eclipse, when a difference measurement between fluxes just outside of and during the eclipse reveals reflected light (in the optical) and thermal emission from the planet itself (in the infrared). For the most favourable targets, reflection or emission can be measured as a function of phase, providing a map of planetary conditions. These observations can be made using both space- and ground-based facilities.

Typically, transit, eclipse and phase curve spectra are analysed using so-called ‘retrieval’ modelling frameworks [12–28]. Retrieval models incorporate a simplified, parameterised radiative transfer model, usually one-dimensional, and an algorithm to explore the parameter space and recover the model solution that provides the best fit to the data. Generally, these models involve minimal physical assumptions, instead allowing the atmospheric parameters to vary freely; as a consequence, this technique is a data-driven approach to spectral analysis. This is particularly advantageous for exoplanets, which often boast extremes of temperature and irradiation that stretch our understanding of atmospheric physics.

Ariel, [29–31], a European Space Agency mission currently expected to launch in 2029, will perform the first census of transiting exoplanet atmospheres. Retrieval algorithms therefore have a critical role to play in support of the mission.

In this paper, we present a retrieval challenge conducted by the Spectral Retrievals Working Group for the Ariel Science Team. We use simulated Ariel observations to test our ability to recover a range of properties of transiting exoplanet atmospheres, using five independent retrieval frameworks that are currently used in the literature. We describe the basic properties of each model in Section 2; how the challenge was set up in Section 3.1; and the results in Section 3.2.

2 Retrieval codes

The main details of the five retrieval schemes used in this analysis are briefly summarised here. For detailed information, we recommend referring to the journal articles for each.
Each retrieval code conforms to the usual basic structure of a simple, parameterised radiative transfer model, coupled to an algorithm that samples the model parameters from a pre-defined prior distribution and converges towards the most likely solution. The versions of each model used in this work are 1D, and all contain the same model parameters, which are described in Section 3.1. All models except Pyrat Bay use a Nested Sampling approach for convergence, whilst Pyrat Bay uses an MCMC sampler.

2.1 ARCIS

The ARtful modelling Code for exoplanet Science (ARCiS) is a forward modelling and Bayesian retrieval code designed to include physical and chemical atmospheric processes [23]. The structure of the atmosphere can be defined through classical parameterisations or computed self consistently using various approximations. For the physical and chemical processes, computationally efficient methods are included that are parameterised where our physical knowledge is lacking. For a more detailed description of the modelling philosophy we refer to [23] and [32]. The radiative transfer is computed using correlated-k sampling of the molecular opacities. Many molecules are included, where available from the ExoMol database [33, 34]. Clouds can be included either parameterised or using the cloud formation concept from Ormel & Min (2019). The cloud opacities are computed from either Mie theory or using a model for irregularly shaped particles [35]. Efficient isotropic multiple scattering calculations can be included with a correction factor for anisotropic scattering. Full anisotropic scattering can be performed as well using a Monte Carlo scattering method. Processes that can be computed include chemical equilibrium and disequilibrium (Kawashima et al. in prep), and cloud and haze formation. Also the pressure temperature structure can be computed from radiative equilibrium with the stellar irradiation. The retrieval can be done using either optimal estimation or Multinest Bayesian sampling.

2.2 NEMESIS

NEMESIS is a retrieval scheme that works with both optimal estimation and nested sampling approaches. It incorporates a fast correlated-k radiative transfer model, where the correlated-k approximation is a way of pre-tabulating gas absorption coefficients within a wavelength interval, relying on the assumption that the strongest lines at one level in the atmosphere are correlated with the strongest lines at other levels. For further details see [17, 36, 37] and [38]. NEMESIS was originally developed for analysis of Solar System planets (e.g. [39, 40]) and has subsequently been extended to exoplanets (e.g. [41, 42]).

Line data are sourced primarily from the ExoMol project and provided in appropriate format for each model by [33]. H$_2$O is from [43], CO$_2$ from [44], CO from [45], CH$_4$ from [46] and TiO from [47].
2.3 Pyrat Bay

The Python Radiative Transfer in a Bayesian framework (Pyrat Bay, Cubillos & Blecic, in prep.), is a modular open-source code to model exoplanet spectra and retrieve the planet’s atmospheric properties. The atmospheric models consist of parameterized 1D profiles of the temperature, composition, and altitude (in hydrostatic equilibrium) as a function of pressure. For transmission geometry, Pyrat Bay solves the radiative transfer equation under the plane-parallel approximation, sampling the opacities at a constant resolving power over the wavelengths considered here.

Pyrat Bay considers opacities from the main sources expected for exoplanets at these wavelengths: molecular line transitions from HITRAN or ExoMol [48–50], collision-induced absorption from Borysow or HITRAN [51–55], resonant Na and K opacity models [56], Rayleigh scattering for H, He, and H$_2$ [57, 58], and several cloud models, from a simple gray cloud deck to complex Mie-scattering [59] models in thermal stability [60] or microphysical parameterization (Blecic et al., in prep.). Pyrat Bay handles the billion-sized line lists by compressing them with the Repack package [61], to extract only the dominating line transitions.

The code explores the parameter space via a differential-evolution MCMC sampler implemented in [62], checking on the Gelman–Rubin statistics for convergence [63].

2.4 TauREx

TauREx (Tau Retrieval for Exoplanets) is a fully Bayesian radiative transfer and retrieval framework [12, 26, 27]. TauREx can be used with the line-by-line cross sections from the Exomol project [34] and HITEMP [64] and HITRAN [65]. TauREx can be used to model both transmission and thermal emission. We also included absorptions from Rayleigh scattering and CIA for the couples H$_2$-H$_2$ and H$_2$-He [66–68]. The public version of TauREx is able to retrieve chemical composition of exoplanets by assuming constant abundances with altitude, parametric 2-layer variations [69], or equilibrium chemistry [70]. In the new version, TauREx 3 is particularly flexible, allowing users to redefine any part of the code with their own custom modules. To perform the retrieval, TauREx can use multiple sampling techniques. Here we use the nested sampling retrieval algorithm Multinest [71] in its python implementation PyMultinest [72].

2.5 POSEIDON

POSEIDON is a nested sampling retrieval code for exoplanet transmission spectra [21]. Radiative transfer is computed via the sampling of high spectral resolution ($R \sim 10^6$) cross sections onto intermediate resolution wavelength grids (typically 100× higher than the resolution of the observations being retrieved), producing a close representation of line-by-line radiative transfer. The atmospheric temperature
structure can be parameterised either via a 6-parameter function [22] or an isotherm. Inhomogenous ‘patchy’ clouds and hazes are included, allowing cloud fractions to be retrieved. Over 50 chemical species are currently supported as retrievable parameters, with molecular line data largely sourced from ExoMol [34], atomic data from VALD3 [73], and continuum data from HITRAN [55].

3 Retrieval challenge

3.1 Setup

Here, we present the results of a multi-code retrieval challenge, conducted using synthetic spectra generated by TauREx with retrievals by NEMESIS, ARCiS, Pyrat Bay, TauREx and POSEIDON. Four spectra, representing respectively a clear hot Jupiter (with system parameters similar to e.g. HAT-P-30b), a cloudy hot Jupiter (with system parameters similar to e.g. XO-2Nb), a clear warm Neptune (with system parameters similar to e.g. GJ436b) and a cloudy warm super-Earth (with system parameters similar to e.g. GJ1214b) were provided, with known atmospheric inputs. Appropriate noise for Ariel was generated using the radiometric model ArielRad [74], and added as a error envelope to each synthetic observation. This allowed each user to test, benchmark and modify their retrieval procedure. A further four spectra representing similar planets, but without known atmospheric properties, were also provided such that a blind retrieval test could be conducted with NEMESIS, ARCiS, Pyrat Bay and POSEIDON. All retrievals with TauREx were non-blind, as this model was used to generate the synthetic spectra. This also meant that for TauREx retrievals only included parameters known to be necessary for each case. For the other four codes, H$_2$O, CO$_2$, CO, CH$_4$, TiO and cloud are included in all retrievals, regardless of whether they were included in the input to the synthetic spectrum. Note that the only fundamental difference between the blind and non-blind retrievals is the prior knowledge of the person executing the retrievals and thus the inclusion of certain parameters.

<table>
<thead>
<tr>
<th>Planet</th>
<th>$R_p$ ($R_\odot$)</th>
<th>$M_p$ ($M_J$)</th>
<th>$R_c$ ($R_\odot$)</th>
<th>H$_2$O</th>
<th>CH$_4$</th>
<th>CO</th>
<th>Temp (K)</th>
<th>$P_{\text{cloud}}$ (bar)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.2</td>
<td>0.7</td>
<td>1.35</td>
<td>7.0×10$^{-6}$</td>
<td>0.0</td>
<td>4.5×10$^{-4}$</td>
<td>1900</td>
<td>None</td>
</tr>
<tr>
<td>1B</td>
<td>1.2</td>
<td>0.73</td>
<td>1.4</td>
<td>1.0×10$^{-5}$</td>
<td>0.0</td>
<td>1.0×10$^{-3}$</td>
<td>1600</td>
<td>None</td>
</tr>
<tr>
<td>2</td>
<td>1.2</td>
<td>0.7</td>
<td>1.35</td>
<td>7.0×10$^{-6}$</td>
<td>0.0</td>
<td>4.5×10$^{-4}$</td>
<td>1900</td>
<td>1.94×10$^{-2}$</td>
</tr>
<tr>
<td>2B</td>
<td>1.1</td>
<td>0.7</td>
<td>1</td>
<td>4.0×10$^{-5}$</td>
<td>0.0</td>
<td>7.0×10$^{-4}$</td>
<td>1500</td>
<td>8.74×10$^{-5}$</td>
</tr>
<tr>
<td>3</td>
<td>0.5</td>
<td>0.06</td>
<td>0.3</td>
<td>2.0×10$^{-6}$</td>
<td>5.0×10$^{-5}$</td>
<td>0.0</td>
<td>700</td>
<td>None</td>
</tr>
<tr>
<td>3B</td>
<td>0.45</td>
<td>0.07</td>
<td>0.35</td>
<td>1.0×10$^{-5}$</td>
<td>1.0×10$^{-5}$</td>
<td>0.0</td>
<td>750</td>
<td>None</td>
</tr>
<tr>
<td>4</td>
<td>0.3</td>
<td>0.01</td>
<td>0.11</td>
<td>2.0×10$^{-4}$</td>
<td>1.0×10$^{-4}$</td>
<td>0.0</td>
<td>1000</td>
<td>2.91×10$^{-2}$</td>
</tr>
<tr>
<td>4B</td>
<td>0.45</td>
<td>0.01</td>
<td>0.15</td>
<td>1.0×10$^{-6}$</td>
<td>1.0×10$^{-5}$</td>
<td>0.0</td>
<td>700</td>
<td>4.52×10$^{-2}$</td>
</tr>
</tbody>
</table>
The setup of all atmospheric forward and retrieval models consists of an isothermal atmosphere with constant abundances of the given molecular species. The atmospheres are assumed to be in hydrostatic equilibrium with the mean molecular weight computed using the molecular abundances plus 85% H$_2$ and 15% He. The cloud is modeled using a grey, infinite opacity cloud deck at all pressure above a certain pressure level. The spectra were simulated to represent Ariel Tier 2 data [for details we refer to Table 1 in 75].

The bulk planet properties for each case are listed in Table 1. For the blind retrievals, only the planet mass and stellar radius are assumed to be known.

### 3.2 Results

The combined retrieval results from all five codes are compared with the input

![Graphs showing retrieval results and spectral fits for Planet 1.](image)

**Fig. 1** Retrieval results and spectral fits for Planet 1. The colours represent the different retrievals used. Black lines on the parameter plots (left hand panels) indicate the input values for retrieved quantities. Where black lines and TaurREx retrieved values are absent, the gas/cloud was not included in the input model. Thick/thin error bars indicate the 1/2-σ limits respectively. The black points in the top right panel indicate the input spectrum with error bars. The difference spectra (bottom right) show input - model for each retrieval, with the black lines indicating the error envelope.

values, with 1σ error bars, in Figs. 1, 2, 3, 4, 5, 6, 7 and 8. We also present example corner plots for each code, for Planet 2, in Figs. 9, 10, 11, 12 and 13; these provide a better indication of correlations and degeneracies between parameters.
The results shown here display overall very good agreement between both spectra and retrieved parameters for all cases, both given and blind. In general, retrieved parameters are also correct to within 1σ from the input value.

Fig. 2 As Fig. 1 but for Planet 1B

Fig. 3 As Fig. 1 but for Planet 2
The quality of the spectral fits is generally extremely good. The $\chi^2$ values for each model and planet are presented in Table 2.

There are specific spectral regions where discrepancies emerge for some models. For example, for Pyrat Bay there is generally a discrepancy at around 5.5 $\mu$m;
PYRAT BAY underestimates the TauREx transit depth consistently here. Similarly, NEMESIS underestimates the transit depth at around 1.3 μm in Fig. 3, but overestimates it in Fig. 1. These discrepancies across small wavelength regions are most...
likely to be related to different treatments of absorption line data. ARCiS and NEMESIS use k tables, whilst the other codes use cross sections. These different methods for tabulating and binning absorption data both introduce some error over the (much less efficient) line-by-line approach. Additionally, different cross-section or k table grids could also be a source of error between codes that use the same method. This is explored in more detail in Section 4.3.

### Table 2

<table>
<thead>
<tr>
<th>Planet</th>
<th>NEMESIS</th>
<th>ARCiS</th>
<th>PyRat Bay</th>
<th>TauREx</th>
<th>POSEIDON</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>24.7</td>
<td>9.6</td>
<td>17.8</td>
<td>12.1</td>
<td>12.4</td>
</tr>
<tr>
<td>1B</td>
<td>22.3</td>
<td>7.4</td>
<td>15.2</td>
<td>10.0</td>
<td>10.6</td>
</tr>
<tr>
<td>2</td>
<td>11.7</td>
<td>5.4</td>
<td>8.8</td>
<td>5.8</td>
<td>8.9</td>
</tr>
<tr>
<td>2B</td>
<td>5.4</td>
<td>1.5</td>
<td>3.1</td>
<td>1.7</td>
<td>10.2</td>
</tr>
<tr>
<td>3</td>
<td>3.5</td>
<td>0.7</td>
<td>0.4</td>
<td>1.7</td>
<td>0.5</td>
</tr>
<tr>
<td>3B</td>
<td>5.6</td>
<td>2.1</td>
<td>1.4</td>
<td>3.2</td>
<td>2.2</td>
</tr>
<tr>
<td>4</td>
<td>8.2</td>
<td>5.9</td>
<td>7.1</td>
<td>17.6</td>
<td>6.6</td>
</tr>
<tr>
<td>4B</td>
<td>8.2</td>
<td>5.1</td>
<td>4.8</td>
<td>16.6</td>
<td>8.4</td>
</tr>
</tbody>
</table>
3.2.2 Retrieval results

Retrieved results are correct to within $2\sigma$ in all cases except for a few radius retrievals. Since the planet radius is the most precisely determined quantity, small offsets in the synthetic spectrum can produce large deviations in the measured radius with respect to the error bars. This does not however affect the accuracy of the other retrieved quantities.

The one remaining discrepant result is in the temperature retrieval for Planet 3B, for the ARCiS and Pyrat Bay models. This spectrum has relatively large error bars, allowing more flexibility in the model fit than some of the other examples. In addition, the NEMESIS and TauREx codes have been substantially benchmarked against each other ([76]) so NEMESIS may be expected to more faithfully reproduce a spectrum generated using TauREx than other models that have been less extensively calibrated. POSEIDON retrieves the correct input result despite a lack of extensive prior benchmarking, which is testament to the generally high reliability of current spectral retrieval schemes.

Where the retrieved gas is absent from the input model, each retrieval scheme with the exception of TauREx is effectively retrieving an upper limit on that gas. Generally, the error bars for these species are large, indicating little constraint. The exception to this is an apparently constrained TiO abundance for the clear hot Jupiter case from NEMESIS. This is likely to be a function of the fact that TiO features are not resolved by Ariel as they occur at short wavelengths, where the resolving power is low, so a small amount of TiO can be invoked to produce opacity at shorter wavelengths without there being confidence that TiO is actually the species concerned. Definitive detection of TiO with Ariel is likely to be a challenge as a result of this.

Where there is no cloud in the input model, an upper limit for the cloud top pressure is retrieved. This upper limit corresponds to the pressure at which the atmosphere becomes opaque in transmission (so any cloud deck sitting beneath this level would be invisible).

The generally very good agreement in both spectral fit and retrieved properties from five different retrieval models, and the accuracy of the retrieved solutions, demonstrate that a wealth of atmospheric information can be reliably recovered from Ariel spectra.

4 Discussion

Here, we discuss the retrieved results for each model in more detail, and investigate further some discrepancies that emerge. It is necessary to note here that differences in output spectra and in model parameters of the level that we see here are likely to be equivalent to discrepancies between models and real data/truth; in a real-observation scenario, these differences could be due to uncorrected instrument systematics or astrophysical noise (e.g. stellar activity that is not accounted for), as well as incompleteness of the model itself [76].
In Table 3 we present indices for the accuracy of each retrieval, in terms of its ability to correctly identify the value of the input parameters to within $1/\sigma$. This is defined in a similar way to the $\chi^2$ value. We only consider the parameters for each case that are constrained by the retrieval, so for example where a gas is not included in the input model the accuracy for the null detection is not calculated.

Our metric is given by:

$$a_{\text{index}} = \frac{\sum_{n}^{n}(x_{\text{ret},n} - x_{\text{input},n})^2 / \sigma^2_n}{n}$$

(1)

where $x_{\text{input}}$ and $x_{\text{ret}}$ are the input and retrieved values for each parameter, $\sigma$ is the (average) error on the parameter, and $n$ is the number of constrained parameters. We took logs for the volume mixing ratios and cloud pressure for this calculation.

As shown in Table 3, retrieval accuracy varies between codes and planets, but in general retrievals for different codes are consistent with each other for a given planet. TauREx generally has an extremely low accuracy index, which is expected, since TauREx was used to generate the input model.

The highest accuracy indices, indicating relatively poorer retrievals, are found for planets 1, 1B and 4B. Planets 1 and 1B also have relatively high $\chi^2$ values, indicating that these have some of the poorer spectral fits. By contrast, Planet 4B has a reasonable quality of fit, but since this is a cloudy planet this adds to the complexity of the retrieval.

Examining the retrieved values for planets 1 and 1B reveals that the temperatures are slightly overestimated for both planets outside of the $1\sigma$ error bars, except for the TauREx retrieval. This is likely to have resulted in the poorer accuracy values. For Planet 4B, the gas abundances are slightly overestimated for some codes, and the cloud top pressure is correspondingly underestimated, indicating that it is degeneracy between cloud pressure and gas abundances that is responsible for the relative lack of accuracy in this case.

For all other planets, the accuracy index is below 1 regardless of the retrieval model used (except Planet 4 for PyRat Bay), indicating that all retrieved values
are recovered correctly to within 1σ. The Pyrat Bay retrieval for Planet 4 suffers from similar cloud top pressure/gas abundance degeneracy to that seen more widely for Planet 4B.

Despite this, the retrievals are accurate to within 2σ across all models, for all planets, and in most cases the molecular abundances are recovered correctly to within 1σ. Considering that measurement of atmospheric composition is a key goal for Ariel, this finding provides confidence in the ability of the mission to deliver on its objectives.

Fig. 9 This figure shows the retrieved posterior probability distributions for NEMESIS on Planet 2. Input values are indicated by black lines. Dashed lines show the median and +/-σ values. In all cases, the true values fall within the +/-σ range.
For ease of comparison, we have so far shown simply the median, and 1- and 2-σ values for each retrieved property. This of course doesn’t show any correlations present between parameters, or fully capture the shape of the retrieved probability distribution.

To illustrate this, we show the full retrieved posteriors from each code for Planet 2 (Figs. 9–13). This planet was chosen as it is cloudy, allowing the effect of clouds on retrievals (and especially on parameter correlations) to be seen. Corner plots for NEMESIS, TauREx and ARCiS were generated using the corner.py routine [77].

All retrieved posteriors show that the abundances of the constrained gases H$_2$O and CO are inversely correlated with the cloud top pressure. Lower cloud top pressures correspond to cloud that sits higher in the atmosphere. The higher the cloud is,
the larger the fraction of the atmospheric features that are obscured, so more \( \text{H}_2\text{O} \) and CO are required to offset a higher cloud.

By contrast, 10-bar radius and cloud top pressure are correlated, because a lower cloud top pressure (higher cloud) means that a smaller radius is required to fit the observed spectrum.

\( \text{H}_2\text{O} \) and CO abundances are correlated. This is likely to be because the spectrum is dominated by absorption due to \( \text{H}_2\text{O} \). If the \( \text{H}_2\text{O} \) abundance increases, more CO is required for the feature to stand out against the \( \text{H}_2\text{O} \) absorption.

Finally, temperature and radius are inversely correlated. This is because both affect the atmospheric scale height in a similar way. Scale height is proportional to temperature, and inversely proportional to the gravitational acceleration. The gravitational acceleration \( g \propto r^{-2} \), so the scale height is proportional to the square of the

---

Fig. 11 As Fig. 9 but for Pyrat Bay. In this case, dark red lines indicate the input values and dashed lines the median retrieved values. Shading in the histograms shows the +/−1σ envelope, which is obtained from the 68% highest-posterior-density credible region [62, see Appendix in]
radius. The variation in transit depth is proportional to the radius multiplied by the scale height, so $\propto r^3$. An increase in radius can therefore be offset by a decrease in temperature, and vice versa.

### 4.3 Effects of cross section grids

The PyRAT Bay retrievals initially used different sampling for the gas absorption cross sections compared with the TauREx cross sections used to generate the input spectrum, which in the first iteration produced significantly discrepant results. The main difference was that the PyRAT Bay cross sections sampled the line transitions only up to

---

**Fig. 12** As Fig. 9 but for TauREx. Gases that were not included in the original model were not retrieved for in this case.
100 half-width at half maximum (HWHM) away from the line center, whereas the TauREx cross sections sampled the lines up to 500 HWHM or 25 cm$^{-1}$. The discrepancies are database-dependent (e.g., more significant for the CO molecule which has more sparse line transitions) and are more significant at longer wavelengths (due to the narrower Doppler line broadening). In Fig. 14 we show the effects of this on the retrieval. The PYRAT BAY run using the TauREx cross sections results in retrieved parameters that are much closer to the input values. Note that the difference in the cross sections arise only in the line sampling, as they both are computed from the same line lists. This is a good example of the way in which apparently minor variations in model set-up can affect retrieval outcomes.
Conclusions

We present a comparison of retrievals conducted by five different codes that provide overall very good agreement between them. We show that the parameters and uncertainties derived by these different codes are all comparable. One important aspect to note is that small differences in the forward model setup can lead to noticeable differences in the retrieval outcome. These systematic errors have to be considered when interpreting the absolute values of retrieval results, even though we show here that they are generally small.

Fig. 14 As Fig. 9 but for PyRaT Bay, with results plotted for both the original cross sections (yellow) and the TauREx cross sections (blue). Dark orange lines indicate the input values. Dashed lines indicate median and +/- σ values. It is clear that the run using the same cross sections as used to generate the input spectrum provides a much better match to the true values.

5 Conclusions

We present a comparison of retrievals conducted by five different codes that provide overall very good agreement between them. We show that the parameters and uncertainties derived by these different codes are all comparable.

One important aspect to note is that small differences in the forward model setup can lead to noticeable differences in the retrieval outcome. These systematic errors have to be considered when interpreting the absolute values of retrieval results, even though we show here that they are generally small.
6 Outlook

Ensuring model accuracy and completeness is key for developing tools that can be used to interpret data from missions such as Ariel. We have presented an example where line data tabulation substantially affected the accuracy of a retrieval; however, line data is just one aspect of modelling.

So far, we have adopted very simple treatments of atmospheric thermal structure and clouds, and further work is needed to fully investigate these effects; some progress has already been made on cloud parameterisation for transit spectra (see e.g. [78] and [79]), and more sophisticated temperature parameterisations are already being applied to existing data. In addition, we have made the assumption here that the terminator of each planet is homogeneous, which we know is unlikely to be the case. Work by [80] and more recently by [81–84] demonstrate the ways in which this assumption can introduce bias into retrieval solutions.

Comparative retrieval studies such as this one are required to understand the impacts of model differences on results, and new investigations with increased model complexity will certainly be required. Similar efforts for secondary eclipse spectra are also necessary and will form the subject of future studies by the Ariel Spectral Retrievals Working Group.

Acknowledgements We thank Patrick Irwin for the use of NEMESIS. JKB was supported by a Royal Astronomical Society Research Fellowship during this project. This project has received funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (grant agreement No 758892, ExoAI and grant agreement No 776403, ExoPANETS-A) and under the European Union’s Seventh Framework Programme (FP7/2007-2013)/ ERC grant agreement numbers 617119 (ExoLights). Furthermore, we acknowledge funding by the Science and Technology Funding Council (STFC) grants: ST/K502406/1, ST/P000282/1, ST/P002153/1 and ST/S002634/1. We acknowledge the availability and support from the High Performance Computing platforms (HPC) DIRAC and OzSTAR, which provided the computing resources necessary to perform this work.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References


32. Ormel CW, Min M (2019b) ARChIS framework for exoplanet atmospheres. The cloud transport model. Astronomy & Astrophysics 622, A121 https://doi.org/10.1051/0004-6361/201833678, Preprint at arXiv:1812.05053


