Alfnoor: A Retrieval Simulation of the Ariel Target List.

Q. Changeat,1 A. Al-Refaie,1 L. V. Mugnai,2 B. Edwards,1 I. P. Waldmann,1 E. Pascale,2 and G. Tinetti1

1Department of Physics and Astronomy
University College London
Gower Street, WC1E 6BT London, United Kingdom
2Dipartimento di Fisica
La Sapienza Universita di Roma
Piazzale Aldo Moro 2, 00185 Roma, Italy

(Accepted to AJ)

ABSTRACT

In this work, we present Alfnoor, a dedicated tool optimised for population studies of exoplanet atmospheres. Alfnoor combines the latest version of the retrieval algorithm TauREx3, with the instrument noise simulator ArielRad and enables the simultaneous retrieval analysis of a large sample of exo-atmospheres. We applied this tool to the Ariel list of planetary candidates and focus on hydrogen dominated, cloudy atmospheres observed in transit with the Tier-2 mode (medium Ariel resolution).

As a first experiment, we randomised the abundances – ranging from $10^{-7}$ to $10^{-2}$ – of the trace gases, which include H$_2$O, CH$_4$, CO, CO$_2$ and NH$_3$. This exercise allowed to estimate the detection limits for Ariel Tier-2 and Tier-3 modes when clouds are present.

In a second experiment, we imposed an arbitrary trend between a chemical species and the effective temperature of the planet. A last experiment was run requiring molecular abundances being dictated by equilibrium chemistry at a certain temperature.

Our results demonstrate the ability of Ariel Tier-2 and Tier-3 surveys to reveal trends between the chemistry and associated planetary parameters. Future work will focus on eclipse data, on atmospheres heavier than hydrogen and will be applied also to other observatories.

1. INTRODUCTION

In the last decade, the field of extra-solar planets has very rapidly grown and matured. The NASA Kepler mission and other dedicated surveys from the ground have revolutionised our understanding of these extraterrestrial worlds. We are now aware of the ubiquity and vast diversity of planets outside our solar system, ranging from ultra-hot giant planets (Gaudi et al. 2017; Delrez et al. 2016; Cameron et al. 2010) to more temperate Earths and Super-Earths (Gillon et al. 2016; Ment et al. 2019). With TESS (Ricker et al. 2014), GAIA (Gaia Collaboration et al. 2016) Cheops (Broeg et al. 2013), SPHERE (Beuzit et al. 2019), GPI (Macintosh et al. 2014), Espresso (Pepe et al. 2010) currently operating and space missions like PLATO (Rauer et al. 2016) and WFIRST (Bennett et al. 2018) soon to come online, the statistics of planets in our galaxy will evolve even further in the next decade.

Current studies of exoplanetary atmospheres have been largely conducted using general observatories from space – Hubble Space Telescope and Spitzer Space Telescope – or from the ground – e.g. VLT-Crires, NASA-IRTF, TNT-
Giano, VLT-SPHERE, Gemini-GPI, Subaru—
and thus results are often sparse and only avail-
able for a limited number of the discovered plan-
et. As a result, most atmospheric retrieval 
studies have focused so far on the analysis of 
individual planets (Line et al. 2016; Tsiaras 
et al. 2016; Kreidberg et al. 2014; Tsiaras et al. 
2019) with only a few papers having attempted 
a consistent spectral analysis of multiple targets 
(Tsiaras et al. 2018; Pinhas et al. 2019; Barstow 
et al. 2016; Sing et al. 2015). In the next decade, 
a new generation of observatories from space 
and the ground and dedicated missions (Gard-
ner et al. 2006; Edwards et al. 2019b; Tinetti 
et al. 2018; Gilmozzi & Spyromilio 2007; Skid-
more 2015) will come online, offering a broader 
spectral coverage, higher signal-to-noise ratio 
(SNR) and the ability to study a significantly 
larger number of targets. The ESA-Ariel mis-
sion alone has been designed to deliver transit, 
eclipse and phase-curve spectra for hundreds of 
planets, providing, for the first time, the chance 
to conduct a statistically significant survey of 
exoplanet atmospheres (Edwards et al. 2019a). 

In most fields of astronomy (supernovae, 
brown dwarfs, black holes), revolutions in our 
understanding of the main processes often came 
from the study of the statistical behaviour using 
large samples as opposed to individual studies. 
As the next generation of space telescopes come 
online, we will reach this important step for 
exo-atmospheres and it is therefore critical to 
be aware of the challenges associated with large 
scale studies.

In this paper, we describe our integrated algo-

rithm, Alfnoor, which combines the open source atmospheric retrieval code TauREx 3 (Al-Refaie et al. 2019) and the Ariel noise simulator ArielRad (Mugnai et al. 2020b) with the aim to facil-
itate the spectral analysis and interpretation of 
populations of exoplanetary atmospheres (§2). 
Current Ariel’s strategy is to observe planets in 
accordance to a four tier structure, where the 
aim of the second tier (Tier-2) of observations 
is to extract the key atmospheric constituents 
(Edwards et al. 2019a). In this paper we simu-
lated Ariel Tier-2 and Tier-3 performances for 
a large sample of planets provided in Edwards et al. (2019a). For the selected targets, differ-
ent, randomised atmospheric compositions were 
assumed and an automated retrieval analysis for 
each planet was performed. We then compared 
and discussed the results of the posterior dis-
tributions, as provided by the retrievals, to the 
ground-truth to assess Ariel’s ability to recover 
accurately and precisely the abundances of the 
key trace-gases and identify arbitrary injected 
chemical trends (§3). Finally we discuss these 
results in light of new facilities coming on line 
soon and next steps needed to progress further 
in our understanding of population studies (§4).

2. METHODOLOGY AND SOFTWARE 
DESCRIPTION

2.1. Description of the software

To study large samples of exoplanetary spec-
tra, we built a new tool: Alfnoor. Alfnoor 
combines the highly flexible next generation re-
trieval code TauREx3 with the ArielRad noise 
simulator to provide a unique framework dedi-
cated to the study of exoplanetary populations 
with Ariel.

TauREx 3 (Al-Refaie et al. 2019) is the new 
version of TauREx (Waldmann et al. 2015a,b). 
This complete rewrite takes the form of a li-
ibrary and is designed to make customisation 
and external code integration easy. It uses 
the highly accurate line-lists from the Exo-
Mol (Tennyson et al. 2016), HITRAN (Roth-
man & Gordon 2014) and HITEMP (Gordon 
et al. 2016) databases to build forward and re-
trieval models. A large number of options are 
available in terms of forward models (trans-
mission, emission), chemical profiles (constant 
as a function of pressure, two-layer, equilib-
rium chemistry), temperature profiles (isother-
mal, NPoints, Guillot (2010)) and cloud param-
eterisations (Grey, Lee et al. (2013),Bohren & 
Huffman (2008)).

ArielRad (Mugnai et al. 2020b) estimates 
Ariel performances to observe a certain target 
when stellar, planetary and orbital parameters 
are specified. It also calculates the required 
number of observations to match the require-
Table 1. Wavelength coverage ($\lambda$) and resolutions (R) of each spectrometers (NIRSpec, AIRS-CH0 and AIRS-CH1) for the Ariel tiers. We also show the photometers (VISPhot, FGS1 and FGS2).

| Instrument | $\lambda$ ($\mu$m) | R - Tier 1 | R - Tier 2 | R - Tier 3 |
|------------|-------------------|----------------|----------------|----------------|
| VISPhot    | 0.5 - 0.6         | 0             | 0             | 0             |
| FGS1       | 0.6 - 0.8         | 0             | 0             | 0             |
| FGS2       | 0.8 - 1.1         | 0             | 0             | 0             |
| NIRSpec    | 1.1 - 1.95        | 1             | 10            | 20            |
| AIRS-CH0   | 1.95 - 3.9        | 3             | 50            | 100           |
| AIRS-CH1   | 3.9 - 7.8         | 1             | 10            | 20            |

In our study, we focused on Tier-2 observations, which is the core of the mission, and aims at characterising the key chemical species, thermal structure and the cloud properties of the selected atmospheres. Ariel observations are expected to cover the wavelengths from 0.5 $\mu$m to 7.8 $\mu$m. The telescope has 3 photometers: a Visible Photometer (VISPhot) and two Fine Guidance Sensors (FGS1 and FGS2) that are also used for the observations. The telescope also has two spectrometers: the Near Infrared Spectrometer (NIRSpec) and the Ariel Infrared Spectrometer (AIRS). The resolution of the spectrometers is adapted to the Tier levels. A description of the resolution achieved for each Tier can be found in Tinetti et al. (2018), Edwards et al. (2019a) and Mugnai et al. (2020b). It is summarised in Table 1.

The function Alfnoor-forward simulated high-resolution transit spectra with TauREx 3 for all the targets. Next it called ArielRad to calculate the Ariel error bars, wavelength bins and the number of required observations to reach Tier-2 performances for all the targets. The function alfnoor-inverse took the Tier-2 spectra generated by alfnoor-forward and performed atmospheric retrievals using TauREx 3 in fitting mode.

Tier 1 observations are studied in detail in Mugnai et al. (2020a). Our sample of planets consists of the 146 planets observed in transit at Tier 2 from the Mission Reference Sample presented in Edwards et al. (2019a). Of these planets, 14 of them qualify for observations in Tier 3. The simulated planets are built to represent the entire parameter space. In our sample 20 planets have radius smaller than $2 R_E$, 29 are between $2 - 5 R_E$ and 97 have radius $> 5 R_E$. For a more detailed description of the methodology used to build this target list, we refer the reader to Edwards et al. (2019a). Future studies will concentrate on eclipse observations and / or secondary atmospheres.

2.2. Approach and initial setups

In all the models, the atmosphere is composed of H$_2$ and He with a ratio He/H$_2$ = 0.17. For the trace-gases, the list and sources of the opacities used in this paper are presented in Table 2. Collision Induced Absorption for H$_2$-H$_2$ and H$_2$-He and Rayleigh scattering are included. For the retrievals, unless specified otherwise, we used the same assumptions: mixing ratios constant with pressure, temperature constant with pressure, grey opaque clouds. While temperature variations with altitude are crucial for eclipse observations, in the case of transmission spectra, most studies assume isothermal temperature profiles. This is justified by the narrow wavelength coverage and signal-to-noise in available observations (with the Hubble Space Telescope) which is not allowing to probe large pressure regions in the planet atmosphere. The temperature variations in transmission act as a second order parameter and the spectrum is most sensitive to the mean temperature value, which directly appear in the scale height. However, Barstow et al. (2012), Rocchetto et al. (2016) and Changeat et al. (2019a) highlighted the impact of temperature variations for high signal-to-noise and broad wavelength coverage cases, indicating that JWST and Ariel would be able to retrieve more complex temperature structures from transit spectra. As this study focuses on the capabilities of Ariel to recover chemical species, we do not investigate further the impact of non-isothermal temperature structures. We however note that this assumption could introduce biases to our results. Parameters that are traditionally determined using external methods are fixed to
Table 2. List of opacities used in this work

| Opacity | References |
|---------|------------|
| \( \text{H}_2\text{H}_2 \) | Abel et al. (2011), Fletcher et al. (2018) |
| \( \text{H}_2\text{He} \) | Abel et al. (2012) |
| \( \text{H}_2\text{O} \) | Barton et al. (2017), Polyansky et al. (2018) |
| \( \text{CH}_4 \) | Hill et al. (2013), Yurchenko & Tennyson (2014) |
| \( \text{CO} \) | Li et al. (2015) |
| \( \text{CO}_2 \) | Rothman et al. (2010) |
| \( \text{NH}_3 \) | Yurchenko et al. (2011), Tennyson & Yurchenko (2012) |

Table 3. List of the fit parameters and their priors for the retrievals (Alffnoor-inverse). We take a conservative approach and select larger bounds than the ones used to randomly generate the planets in forward mode. The chemical abundances are expressed in Vertical Mixing Ratios (VMR).

| Parameters | Priors | Scale |
|------------|--------|-------|
| radius \( (R_J) \) | \( \pm 50\% \) | linear |
| cloud pressure (bar) | 10 - 10^{-7} | log |
| \( T \) (K) | \( \pm 50\% \) | linear |
| \( \text{H}_2\text{O} \) (VMR) | 10^{-12} - 10^{-1} | log |
| \( \text{CH}_4 \) (VMR) | 10^{-12} - 10^{-1} | log |
| \( \text{CO} \) (VMR) | 10^{-12} - 10^{-1} | log |
| \( \text{CO}_2 \) (VMR) | 10^{-12} - 10^{-1} | log |
| \( \text{NH}_3 \) (VMR) | 10^{-12} - 10^{-1} | log |

In this study, we aim to explore two particular aspects of the Ariel mission:

1. the ability of Ariel to detect molecular species and the detection limits for these molecules in the context of cloudy primary atmospheres observed in transit. This task can be easily achieved by performing retrievals on an unbiased dataset of planets where the atmospheric composition is randomised and by assessing the cases that have been successfully recovered.

2. the ability of Ariel to reveal chemical trends in exoplanet populations. To assess this possibility, a biased sample can be used as input where an artificial trend is introduced.

We describe below the actual implementation of this plan.

1. **Unbiased sample.** We built the forward model by using the stellar and planetary basic parameters from Edwards et al. (2019a) for the Ariel Target list. We randomised the chemistry, temperature and cloud parameters so that a unique set of these parameters is adopted for each planet of each sample. For the chemistry, we considered constant profiles with pressure for the mixing ratios of \( \text{H}_2\text{O}, \text{CH}_4, \text{CO}, \text{CO}_2 \) and \( \text{NH}_3 \) and chose a random abundance in logarithmic scale from \( 10^{-7} \) to \( 10^{-2} \). For clouds, we generated grey opaque clouds with random top pressures varying in log-scale from 10 bar (equivalent to no clouds) to \( 10^{-3} \) bar. Finally, the atmospheric temperatures were also randomly generated and allowed to assume values between \( 0.7 \times T_{\text{eff}} \) and \( 1.05 \times T_{\text{eff}} \), where \( T_{\text{eff}} \) is the effective temperature in the Ariel target list of Edwards et al. (2019a). The temperature was consciously selected biased towards lower values to account for differences between effective temperature and the terminator temperature (Caldas et al. 2019). We repeated the generation of the observed spectra twice to build un-scattered and scattered datasets. In the un-scattered set, we conserve the theoretical simulated spectra as-is. As scatter generally arises from the random realisation of observations, we apply a Gaussian scatter to a second dataset using the true value as mean and the simulated noise as variance. This scattered dataset better describes what would be obtained in an actual observation by the telescope but cannot be used to characterise retrieval biases as unfortunate runs could lead to large discrepancies between true and retrieved values (Feng et al. 2018; Changeat et al. 2019a). Unscattered spectra are more suitable for the study of retrieval biases and intrinsic correlations between the at-
mospheric parameters (Feng et al. 2018). On the opposite, scattered spectra can inform us on the stability and the redundancy in the information content of Ariel spectra. Previous studies have used both types to simulate observations by future telescopes (Barstow et al. 2012; Tinetti et al. 2015; Feng et al. 2016; Rocchetto et al. 2016; Mollire et al. 2017; Batalha et al. 2017; Tinetti et al. 2018; Blumenthal et al. 2018; Edwards et al. 2019b; Changeat et al. 2019a; Lustig-Yaeger et al. 2019; Changeat et al. 2019b). Feng et al. (2018) results predicted that the retrieved uncertainties should be similar in both scattered and un-scattered runs but that the retrieved mean could be different. Here, we use our two datasets to investigate these predictions keeping in mind that if Ariel spectra contain enough information content redundancy, we should not see large differences in the retrieved mean values.

2. Biased samples. We imposed first a linear relationship between the logarithmic abundance of water and the temperature. We enacted this correlation water-temperature by requiring a mixing ratio of $10^{-4}$ for an effective temperature of 1000K and $10^{-3}$ for an effective temperature of 2000K.

We then tried a more realistic example where the atmospheres were assumed to be in chemical equilibrium and simulated accordingly the chemical abundances and profiles (Agndez et al. 2012). We used the same solar C/O ratio and metallicity for all the planets in the sample. To recover the input profiles, we used in the retrievals both free, constant with altitude chemical profiles and profiles which are forced to follow chemical equilibrium prescriptions. We did not test the entire sample with the two-layer chemistry retrieval scheme as presented in Changeat et al. (2019a), but we have run an example to show the expected improvements of this scheme over the pressure constant chemical profiles.

3. RESULTS

3.1. Unbiased sample

We show in Figure 1 both the observed and retrieved spectra for a subset of the simulated Ariel Tier-2 observations, along with the correlation map between water abundance and temperature with their $1\sigma$ uncertainties. The distance between the true and the retrieved value is visualised by the colour of the point. The retrieved parameters are represented by the median chemical or temperature profiles weighted by the contribution function. The contribution function is defined as the wavelengths averaged variations of the optical depth with pressure. This choice ensures that the values reported well reflect the conditions in the atmospheric regions probed by observations. In order to better visualise the Ariel detection limits in Tier 2, we also provide complementary plots of the retrieved abundances versus their true values for each molecule. The H$_2$O map is presented in Figure 2.

The water-temperature map in Figure 1 clearly shows that our unbiased population is randomly spread in the parameter space, as expected. The retrieved temperature is very precise across the whole parameter space, showcasing the ability of Ariel Tier-2 to study a wide range of planets. It also illustrates that the retrieved values are mostly accurate for water abundances higher than $10^{-6}$: with the exception of a few cases, the retrieved values for water and temperature fall well within the $1\sigma$ error bars (blue to green in the colour scale). We notice for water a rapid change in the posteriors for abundances smaller than $10^{-5}$, marked by large error bars on the left side of the plot. Indeed, when the abundance is too low, the retrievals are not able to distinguish well the features and provide only upper limits. This is an expected behaviour and an indication of the Ariel detection limit for our sample of planets. This exercise was repeated for other molecules to assess Ariel ability to detect different sets of
molecules in Tier-2 mode. Other temperature-molecule maps, as well as the radius-clouds map, are reported in Appendix (Figures 8, 9, 10, 11, 12).

The detection limits are best visualised in the retrieved versus true abundances (see Figure 2). In the same Figure, we also show the retrieved uncertainties versus input abundances as this allows us to distinguish 3 regimes. The first regime corresponds to low abundances where molecular detections are not possible: for example, between $10^{-7}$ and $10^{-6}$ for water no detections seem possible with Ariel. Other molecules are presented in Figures 13: CH$_4$; Figure 14:
Figure 2. Map of the H$_2$O retrieved abundance versus the true value for the unbiased sample. The colour-scale of the 1-$\sigma$ retrieved error bars represents the distance to the true value in units of 1-$\sigma$.

CO; Figure 15: CO$_2$ and Figure 16: NH$_3$. It is interesting to note that when the molecules are not detected, the retrieved errors ($\sigma$) are dominated by the size of the priors and the location of the detection limit: for water no-detection errors are between 2 and 4 orders of magnitude. The second regime for intermediate abundances displays a mix between successful detections and lack of evidence for the molecules. This corresponds to the region with large ranges in the retrieved errors (between $10^{-6}$ and $10^{-5}$ for water). In general, this variability is due to the other constituents in the planet that are susceptible to mask the signal of interest (e.g: clouds,
Table 4. Detection limits for each molecule in Ariel Tier-2 and Tier-3 samples considered here. The detection limits correspond to the lowest value we were able to extract abundance with less than 1 order of magnitude errors. Tier-3 sample includes only 14 planets.

| Molecule      | Tier 2 | Tier 3 |
|---------------|--------|--------|
| log(H₂O)      | −6.5   | < −7   |
| log(CH₄)      | −7     | < −7   |
| log(CO)       | −5.5   | −6     |
| log(CO₂)      | −7     | < −7   |
| log(NH₃)      | −6.5   | < −7   |

other molecules). Finally, for the highest abundances, the retrieved uncertainties are low (less than 1 order of magnitude in the mixing ratios), which indicate that these abundances are always retrieved, regardless of the other constituents in the atmosphere.

Additionally, the map exploring the correlation between planetary radius and cloud top pressure shows that Ariel can separate well these parameters, most likely thanks to the FGS optical channels.

We repeated the same experiment with the second run composed of ‘scattered’ spectra. Each planet is simulated with a new set of randomised parameters. As previously stated, the observed values of the transit depth are assumed to follow a normal distribution (the mean is the simulated transit depth and the standard deviation is the instrumental noise), which better reproduces a real observation. Figure 3 shows the water-temperature map. The other chemical parameters are reported in Appendix (Figures 8, 9, 10, 11, 12). From the analysis of the scattered spectra, we appreciate that the scattering of the data points around their true value does not necessary introduce biases in Ariel Tier-2 retrieval studies. Indeed, this result, which has already been explored in Feng et al. (2018); Changeat et al. (2019a), naturally arises from the redundancy of the information relative to each molecule in the Ariel spectra and the fact that in most cases N repeated observations are needed to obtain Tier-2 requirements, therefore reducing by 1/√N the scattering amplitude around their true value. Feng et al. (2018) highlighted that, to avoid potential biases arising from individual noise instances, one would essentially have to produce multiple retrievals with different noise instances and average the obtained results. As this was not computationally feasible, they chose not to scatter the spectra and use the true value as an approximation, stating that the shape of the posteriors would be accurate but that the position may be optimistically centred. For Ariel Tier-2 observations, the information content of the spectra is redundant enough to ensure that the retrieved values are not affected by this phenomenon and these are mostly centred around the true value in both scattered and non-scattered scenarios. For all molecules, we find that the correlation maps are very similar in both cases and the detection limits remain unchanged from the non-scattered runs. For the clouds, however, we note an overall increase in the distance to the true value (see correlation map in Appendix Figure 12). We note that in the simulations presented here, we considered fully opaque grey cloud cover, which is essentially the worst case scenario as no cloud features are detectable and it is well known to be degenerate with radius (Changeat et al. 2019b). More realistic cloud simulations will be considered in a future paper to test more thoroughly this case.

We summarise in Table 4 the approximate detection limits for each molecule considered. These represent the regions where our retrieval analysis have been able to extract constraints on the given atmospheric constituents. The stated detection limit corresponds to the lowest value that was successfully recovered with less than 1 order of magnitude uncertainties. As seen before, 3 regions of the parameter space can be identified: region 1 with no possible detections of the molecule, region 2 with detections depending on the other atmospheric properties and region 3 where the molecule is always detected. This means the stated values do not represent a guaranty of detection, but rather the lowest limit we can hope to detect the molecule. In addition, we plot the contribu-
tion of each molecule individually in Figure 17 (each spectrum only contains $10^{-5}$ of the considered molecule) to show the features span by each molecule. In general, Ariel Tier-2 spectra should enable molecular detections down to mixing ratios of $10^{-6}$. In our simulations, only CO appears to be difficult to detect at abundances smaller than $10^{-4}$. CO presents two features that are overlapping with CO$_2$ at 4.5$\mu$m and with CH$_4$ at 2.5$\mu$m and are relatively weak. In a real scenario (equilibrium chemistry), we believe CO could be more easily distinguishable as our unbiased assumption underestimates the CO abundance and overestimates the CO$_2$ abundance by design (Agndez et al. 2012; Venot et al. 2012; Venot & Agüindez 2015). We also note that H$_2$O and CH$_4$ have a large number of anti-correlated features, which may give rise to
more featureless spectra when the two molecules are present. For all parameters, Ariel Tier-2 spectra provide accurate and precise estimates, as most of the retrieved error bars are less than 1-sigma away from the true value. This statement applies to both non-scattered and scattered spectra.

For completeness, we performed additional retrievals for 14 benchmark planets in Tier-3 mode (Edwards et al. 2019a). The benchmark planets achieve a high signal-to-noise ratio in a very limited number of transits and are re-observed at different times to allow for temporal and spatial variability studies. In the examples presented here, we combined five transit observations to reach the required signal-to-noise for Tier-3 (Edwards et al. 2019a; Tinetti et al. 2018). The retrievals were performed on the scattered spectra and are illustrated in Figure 4. The retrieval maps for the 14 Ariel Tier-3 cases are reported in Appendix (Figures 18, 19, 20) and the molecular detection limits in Table 4. The detection limit for Ariel Tier-3 spectra is very low, typically mixing ratios equal or smaller than \(10^{-7}\) can be retrieved. Even CO at mixing ratios of \(\sim 10^{-6}\) appears to be detectable. Due to the limited number of studied cases, the Tier-3 detection limits reported here should be taken with caution and will be refined in a separate paper dedicated to the study of Tier-3 planets.

3.2. Biased sample: linear water-temperature trend

When we imposed an arbitrary linear trend between the water abundance and the effective temperature, we obtained the water-temperature map shown in Figure 5. Here the imposed trend is easily recovered by our retrieval analysis. Both scattered and unscattered spectra allow to recover the imposed trend down to water abundances \(\sim 10^{-6}\). In the scattered example, a few cases have larger departures from the true value compared to the non-scattered one but this does not affect the conclusions on the entire population. Additionally, we note that this analysis has been done without retrieval fine tuning.

3.3. Biased sample: equilibrium chemistry atmospheres

When an equilibrium chemistry model was used for both the forward model and the retrievals, we obtained the water-temperature map shown in Figure 6, top, where the trend is very accurately and precisely recovered. Since the molecular abundances are varying with altitude, the values stated correspond to the average weighted by the atmospheric contribution function (the optical depth variations collapsed over wavelengths). Being the model generating and retrieving the data the same, this is an optimistic result, as we should not expect all atmospheres to satisfy the equilibrium chemistry assumption.

Also the free, constant with pressure chemistry retrievals (Figure 6, bottom) allow to recover the equilibrium chemistry trend. The retrieved parameters, however, have large distances from the true value, in some cases the offsets are greater than 2\(\sigma\), meaning that the model confidently recovers a biased value. This behaviour, also present in other chemical species (see Appendix Figures 21, 22, 23, 24), is particularly noticeable for temperatures between 600K and 1100K: this region is known to exhibit large chemical gradients with altitude as the balance in the \(\text{CH}_4/\text{CO}\) reaction changes. These variations in the chemical profiles cannot be captured by our simplistic constant chemistry retrieval model.

It has been shown in Changeat et al. (2019a) that Ariel and JWST will be sensitive to chemical vertical gradients and that retrieval techniques such as the two-layer parametrisation would be essential for the analysis of these next generation spectra.

We show in Figure (7) a comparison between the various retrieval techniques: the two-layer parametrisation (Changeat et al. 2019a) well captures the departure of the methane profile from the constant with altitude case without strong prior assumptions, as opposed to the case of the equilibrium chemistry retrieval.

4. DISCUSSION
In all simulated cases, retrieval analyses were performed without any fine tuning. Also our simulations are simplified compared to real atmospheres, which are expected to have disequilibrium effects, 3D effects and other complexities.

Recently, self-consistent methods, such as the equilibrium chemistry retrieval adopted in a few examples here, have been implemented in retrieval tools. Embedding these chemical schemes in atmospheric retrievals is very tempting as they allow to describe complex chemistry while maintaining a low dimensionality. However, we should be careful in using these techniques to interpret unknown atmospheres, as they do not reflect the information content of the observed spectra. In other terms, if the assumptions made by the retrieval model are not correct, the results will likely be biased (Miller-Ricci Kempton et al. 2012; Rocchetto et al. 2016; Agndez et al. 2012; Changeat et al. 2019a). This issue has been discussed in the literature and should always be remembered when using such techniques.

Other approaches which let the chemical species assume arbitrary values, may allow to discover unexpected trends in the data. However, the model complexity should be adapted to the data, which is not known a priori. A too simplistic model will tend to be biased, while a too complex model will tend to overfit. In this paper, Section 3.3 highlighted a case where the free constant with pressure chemistry retrieval did not adequately describe the input chemical profiles (which were using equilibrium chemistry), thus biasing our results. A more sophisticated description of the chemical profiles in retrievals is presented in Changeat et al. (2019a).

We illustrate this point by comparing different chemical schemes on an observed spectrum taken from our previously made equilibrium chemistry dataset. Figure 7 demonstrates that all three chemical schemes (equilibrium, constant, two-layer) are able to match the observed spectrum. The contribution function (solid blue line on the right figure) shows how the models try to reproduce the input abundances for CH\textsubscript{4} in the region where the contribution function is maximum. The equilibrium and two-layer scenarios are better describing the input profiles in general, while the retrieved uncertainties are more representative. The retrieved constant chemical profile with altitude only averages the input CH\textsubscript{4} abundance, providing limited details on the atmospheric chemical processes. As expected, we find that the input retrieved weighted abundance is best approxi-
Figure 5. Biased sample: linear water-temperature trend. Top: retrieved water-temperature map from the non-scattered spectra. Bottom: retrieved water-temperature map from the scattered spectra. We show the retrieved 1-σ error bars on the retrieved parameters. The colour-scale represents the distance to the true value in units of 1-σ. The dashed grey line indicates the input trend.

The behaviour seen in this example explains the large distances to the true value and the general overconfidence in the retrieved chemistry of our free constant with altitude scenario in Figure 6.

5. CONCLUSION
Figure 6. Biased sample: equilibrium chemistry atmospheres. Correlation map of the retrieved abundance of water and the temperature. Results obtained with equilibrium chemistry retrievals (top) and with free, constant with altitude chemistry retrievals (bottom). We show the retrieved 1-σ error bars on the retrieved parameters. The colour-scale represents the distance to the true value in units of 1-σ.

This work assessed the capabilities of Ariel to identify chemical trends – if present – in exoplanet populations through the study of their atmospheres. We developed a dedicated software, Alfnoor, to perform atmospheric retrievals on the entire Ariel list of planetary candidates. Among the key results obtained, we found the detection limits for H$_2$O, CH$_4$, CO$_2$ and NH$_3$ to be $\sim 10^{-6}$ in the case of Tier-2 and $< 10^{-7}$ in the case of Tier-3 transit observations. CO, though, has higher detection thresholds, i.e. $\sim 10^{-4}$ for Tier-2 observations and $\sim 10^{-6}$ for Tier-3.
Figure 7. Results for our retrievals with three different chemical profiles (equilibrium, constant with altitude and two-layer). The input forward model is taken from the Alfnoor run with equilibrium chemistry. Left: Simulated observations and retrieval best fit models; Right: Comparison of the retrieved CH$_4$ profiles. The contribution function in the atmosphere, corresponding to d$T$/dP, is also provided. The global log evidence, which qualifies the preference shown by the data for a given model, is 400 for the equilibrium model, 397.5 for the two-layer retrieval and only 395 for the constant chemistry retrieval. In comparing models, a difference of 2 indicate a strong preference towards the model of higher value (Kass & Raftery 1995). We calculate the abundances weighted by the contribution function (log) to be: $-5.29$ for the forward model; $-5.27 \pm 0.21$ for the equilibrium model; $-4.64 \pm 0.15$ for the constant chemistry model; $-4.96 \pm 0.39$ for the two-layer model.

We also confirmed the potentials of Ariel to recover chemical trends in exoplanetary atmospheres. We tested correlations between chemical species and temperature and a planet population whose chemical composition is entirely determined by equilibrium chemistry.

Limitations in our assumptions for the chemistry, temperature and cloud models imply that additional work still needs to be done to fully understand the degeneracies associated with these techniques and how to fully automate retrieval strategies. In the future, we aim to simulate more realistic scenarios using self consistent forward models (e.g. including disequilibrium chemistry) and more complex thermal and cloud assumptions. While this work was inspired by the Ariel mission, similar large scale simulations could also help prioritising the use of other observatories from space and the ground and provide a great tool for the preparation of observational campaigns.

6. ACKNOWLEDGEMENT

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 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. ASI grant n. 2018.22.HH.O. UCL London-Rome Cities Partnerships Program.
REFERENCES

Abel, M., Frommhold, L., Li, X., & Hunt, K. L. 2011, The Journal of Physical Chemistry A, 115, 6805
—. 2012, The Journal of chemical physics, 136, 044319
Agndez, M., Venot, O., Iro, N., et al. 2012, Astronomy & Astrophysics, 548, A73.
http://dx.doi.org/10.1051/0004-6361/201220365
Al-Refaie, A. F., Changeat, Q., Waldmann, I. P., & Tinetti, G. 2019, TauREx III: A fast, dynamic and extendable framework for retrievals, , , arXiv:1912.07759
Barstow, J. K., Aigrain, S., Irwin, P. G. J., et al. 2012, On the potential of the EChO mission to characterise gas giant atmospheres, , , arXiv:1212.5020
Barstow, J. K., Aigrain, S., Irwin, P. G. J., & Sing, D. K. 2016, The Astrophysical Journal, 834, 50.
http://dx.doi.org/10.3847/1538-4357/834/1/50
Barton, E. J., Hill, C., Yurchenko, S. N., et al. 2017, Journal of Quantitative Spectroscopy and Radiative Transfer, 187, 453
Batalha, N. E., Mandell, A., Pontoppidan, K., et al. 2017, Publications of the Astronomical Society of the Pacific, 129, 064501.
http://dx.doi.org/10.1088/1538-3873/aa65b0
Bennett, D. P., Akeson, R., Anderson, J., et al. 2018, The WFIRST Exoplanet Microlensing Survey, , , arXiv:1803.08564
Beuzit, J. L., Vigan, A., Mouillet, D., et al. 2019, A&A, 631, A155
Blumenthal, S. D., Mandell, A. M., Hbrard, E., et al. 2018, The Astrophysical Journal, 853, 138.
http://dx.doi.org/10.3847/1538-4357/aa9e51
Bohren, C. F., & Huffman, D. R. 2008, Absorption and scattering of light by small particles (John Wiley & Sons)
Broeg, C., Fortier, A., Ehrenreich, D., et al. 2013, in European Physical Journal Web of Conferences, Vol. 47, European Physical Journal Web of Conferences, 03005
Caldas, A., Leconte, J., Selsis, F., et al. 2019, Astronomy & Astrophysics, 623, A161
Cameron, A. C., Guenther, E., Smalley, B., et al. 2010, Monthly Notices of the Royal Astronomical Society, 407, 507514.
http://dx.doi.org/10.1111/j.1365-2966.2010.16922.x
Changeat, Q., Edwards, B., Waldmann, I. P., & Tinetti, G. 2019a, The Astrophysical Journal, 886, 39.
http://dx.doi.org/10.3847/1538-4357/ab4a14
Changeat, Q., Keyte, L., Waldmann, I. P., & Tinetti, G. 2019b, Impact of planetary mass uncertainties on exoplanet atmospheric retrievals, , , arXiv:1908.06305
Delrez, L., Santerne, A., Almenara, J.-M., et al. 2016, Monthly Notices of the Royal Astronomical Society, 458, 40254043.
http://dx.doi.org/10.1093/mnras/stw522
Edwards, B., Mugnai, L., Tinetti, G., Pascale, E., & Sarkar, S. 2019a, AJ, 157, 242
Edwards, B., Rice, M., Zingales, T., et al. 2019b, Experimental Astronomy, 47, 29
Feng, Y. K., Line, M. R., Fortney, J. J., et al. 2016, The Astrophysical Journal, 829, 52.
http://dx.doi.org/10.3847/0004-637X/829/1/52
Feng, Y. K., Robinson, T. D., Fortney, J. J., et al. 2018, AJ, 155, 200
Fletcher, L. N., Gustafsson, M., & Orton, G. S. 2018, The Astrophysical Journal Supplement Series, 235, 24
Gaia Collaboration, Prusti, T., de Bruijne, J. H. J., et al. 2016, A&A, 595, A1
Gardner, J. P., Mather, J. C., Clampin, M., et al. 2006, Space Science Reviews, 123, 485606.
http://dx.doi.org/10.1007/s11214-006-8315-7
Gaudi, B. S., Stassun, K. G., Collins, K. A., et al. 2017, Nature, 546, 514518.
http://dx.doi.org/10.1038/nature22392
Gillon, M., Jehin, E., Lederer, S. M., et al. 2016, Nature, 533, 221224. http://dx.doi.org/10.1038/nature17448
Gilmozzi, R., & Spyromilio, J. 2007, The Messenger, 127, 11
Gordon, I., Rothman, L. S., Wilzewski, J. S., et al. 2016, in AAS/Division for Planetary Sciences Meeting Abstracts, Vol. 48, AAS/Division for Planetary Sciences Meeting Abstracts #48, 421.13
Guillot, T. 2010, A&A, 520, A27
Hill, C., Yurchenko, S. N., & Tennyson, J. 2013, Icarus, 226, 1673
Kass, R. E., & Raftery, A. E. 1995, Journal of the american statistical association, 90, 773
Kreidberg, L., Bean, J. L., Dsert, J.-M., et al. 2014, The Astrophysical Journal, 793, L27.
http://dx.doi.org/10.1088/2041-8205/793/2/L27
Lee, J.-M., Heng, K., & Irwin, P. G. J. 2013, The Astrophysical Journal, 778, 97.
https://doi.org/10.1088/2041-8205/793/2/L27
Li, G., Gordon, I. E., Rothman, L. S., et al. 2015, The Astrophysical Journal Supplement Series, 216, 15
Line, M. R., Stevenson, K. B., Bean, J., et al. 2016, The Astronomical Journal, 152, 203.
http://dx.doi.org/10.3847/0004-6256/152/6/203
Lustig-Yaeger, J., Meadows, V. S., & Lincowski, A. P. 2019, The Astronomical Journal, 158, 27.
http://dx.doi.org/10.3847/1538-3881/ab21e0
7. APPENDIX
Figure 8. Unbiased sample: correlation map between the temperature and the retrieved abundances of CH$_4$, with the 1-σ retrieved error bars. The colour-scale represents the distance to the true value in units of 1-σ. Top: Non-scattered spectra. Bottom: Scattered spectra.
Figure 9. Unbiased sample: correlation map between the temperature and the retrieved abundances of CO, with the 1-σ retrieved error bars. The colour-scale represents the distance to the true value in units of 1-σ. Top: Non-scattered spectra. Bottom: Scattered spectra.
Figure 10. Unbiased sample: correlation map between the temperature and the retrieved abundances of CO$_2$, with the 1-$\sigma$ retrieved error bars. The colour-scale represents the distance to the true value in units of 1-$\sigma$. Top: Non-scattered spectra. Bottom: Scattered spectra.
Figure 11. Unbiased sample: correlation map between the temperature and the retrieved abundances of NH$_3$, with the 1-$\sigma$ retrieved error bars. The colour-scale represents the distance to the true value in units of 1-$\sigma$. Top: Non-scattered spectra. Bottom: Scattered spectra.
Figure 12. Unbiased sample: correlation map between the clouds and the radius, with the 1-σ retrieved error bars. The colour-scale represents the distance to the true value in units of 1-σ. Top: Non-scattered spectra. Bottom: Scattered spectra.
Figure 13. Top: Map of the CH$_4$ retrieved abundances versus their the values for the unbiased sample. Bottom: Error retrieved as a function of the input abundances. The colour-scale of the 1-σ retrieved error bars represents the distance to the true value in units of 1-σ.
Figure 14. Top: Map of the CO retrieved abundances versus their the values for the unbiased sample. Bottom: Error retrieved as a function of the input abundances. The colour-scale of the 1-σ retrieved error bars represents the distance to the true value in units of 1-σ.
Figure 15. Top: Map of the CO$_2$ retrieved abundances versus their the values for the unbiased sample. Bottom: Error retrieved as a function of the input abundances. The colour-scale of the 1-$\sigma$ retrieved error bars represents the distance to the true value in units of 1-$\sigma$. 
Figure 16. Top: Map of the NH$_3$ retrieved abundances versus their the values for the unbiased sample. Bottom: Error retrieved as a function of the input abundances. The colour-scale of the 1-$\sigma$ retrieved error bars represents the distance to the true value in units of 1-$\sigma$. 
Figure 17. Absorption in Ariel of the molecular species considered in this paper. Each simulation is for a 1 R_J, 1 M_J planet and 1 R_⊙ star with 10^{-5} of the considered molecule as only absorber. The models are offset for better visibility. The shaded region is the full resolution contribution, while the solid lines and black points correspond to Ariel resolutions.
Figure 18. Correlations maps obtained for Ariel Tier-3 scattered spectra: H$_2$O-T (top) and CH$_4$-T (bottom).
Figure 19. Correlations maps obtained for Ariel Tier-3 scattered spectra: CO-T (top) and CO$_2$-T (bottom).
Figure 20. Correlations maps obtained for Ariel Tier-3 scattered spectra: NH$_3$-$T$. 
Figure 21. Biased sample: equilibrium chemistry atmospheres. Correlation map of the retrieved abundance of CH$_4$ and the temperature, with the 1-σ retrieved error bars. Results obtained with equilibrium chemistry retrievals (top) and with free, constant chemistry with pressure retrievals (bottom). The colour-scale represents the distance to the true value in units of 1-σ.
Figure 22. Biased sample: equilibrium chemistry atmospheres. Correlation map of the retrieved abundance of CO and the temperature, with the 1-σ retrieved error bars. Results obtained with equilibrium chemistry retrievals (top) and with free, constant with pressure chemistry retrievals (bottom). The colour-scale represents the distance to the true value in units of 1-σ.
Figure 23. Biased sample: equilibrium chemistry atmospheres. Correlation map of the retrieved abundance of CO$_2$ and the temperature, with the 1-$\sigma$ retrieved error bars. Results obtained with equilibrium chemistry retrievals (top) and with free, constant with pressure chemistry retrievals (bottom). The colour-scale represents the distance to the true value in units of 1-$\sigma$. 
Figure 24. Biased sample: equilibrium chemistry atmospheres. Correlation map of the retrieved abundance of NH$_3$ and the temperature, with the 1-$\sigma$ retrieved error bars. Results obtained with equilibrium chemistry retrievals (top) and with free, constant with pressure chemistry retrievals (bottom). The colour-scale represents the distance to the true value in units of 1-$\sigma$. 