







JWST/NIRISS and HST: Exploring the improved ability to characterize exoplanet atmospheres in the JWST era

Chloe Fisher ¹★, Jake Taylor ¹★, Vivien Parmentier,² Daniel Kitzmann,^{3,4} Jayne L. Birkby ¹,
Michael Radica ⁵, Joanna Barstow ⁶, Jingxuan Yang ⁷ and Giuseppe Morello⁸

¹*Astrophysics, Department of Physics, University of Oxford, Denys Wilkinson Building, Keble Road, Oxford, OX1 3RH, UK*

²*Université Côte d'Azur, Observatoire de la Côte d'Azur, CNRS, Laboratoire Lagrange, 06903, France*

³*Space Research and Planetary Sciences, Physics Institute, University of Bern, Gesellschaftsstrasse 6, CH-3012 Bern, Switzerland*

⁴*Center for Space and Habitability, University of Bern, Gesellschaftsstrasse 6, CH-3012 Bern, Switzerland*

⁵*Trottier Institut de Recherche sur les Exoplanètes (iREx), Université de Montréal, 1375 Avenue Thérèse-Lavoie-Roux, Montréal, QC, H2V 0B3, Canada*

⁶*School of Physical Sciences, The Open University, Walton Hall, Milton Keynes, MK7 6AA, UK*

⁷*Atmospheric, Oceanic and Planetary Physics, Department of Physics, University of Oxford, Oxford OX1 3PU, UK*

⁸*Instituto de Astrofísica de Andalucía (IAA-CSIC), Glorieta de la Astronomía s/n, E-18008 Granada, Spain*

Accepted 2024 September 24. Received 2024 September 4; in original form 2024 July 19

ABSTRACT

The Hubble Space Telescope has been a pioneering instrument for studying the atmospheres of exoplanets, specifically its WFC3 and STIS instruments. With the launch of JWST, we are able to observe larger spectral ranges at higher precision. NIRISS/SOSS covers the range 0.6–2.8 microns, and thus, it can serve as a direct comparison to WFC3 (0.8–1.7 microns). We perform atmospheric retrievals of WFC3 and NIRISS transmission spectra of WASP-39 b in order to compare their constraining power. We find that NIRISS is able to retrieve precise H₂O abundances that do not suffer a degeneracy with the continuum level due to the coverage of multiple spectral features. We also combine these data sets with spectra from STIS and find that challenges associated with fitting the steep optical slope can bias the retrieval results. In an effort to diagnose the differences between the WFC3 and NIRISS retrievals, we perform the analysis again on the NIRISS data cut to the same wavelength range as WFC3. We find that the water abundance is in strong disagreement with both the WFC3 and full NIRISS retrievals, highlighting the importance of wide wavelength coverage. Finally, we carry out mock retrievals on the different instruments, which shows further evidence of the challenges in constraining water abundance from the WFC3 data alone. Our study demonstrates the vast information gain of JWST's NIRISS instrument over WFC3, highlighting the insights to be obtained from our new era of space-based instruments.

Key words: planets and satellites: atmospheres.

1 INTRODUCTION

The arrival of data from JWST presents a huge increase in the resolution and wavelength coverage of exoplanet atmosphere spectra we can obtain from space-based observatories. Spanning from 0.6 μm all the way to 28 μm across four different instruments with spectroscopic capabilities, the gain in potential information for exoplanet atmospheres is no doubt extensive. This new wavelength domain has facilitated the discovery of new molecules and confirmed previous detections (e.g. Bell et al. 2023), improved our analysis of cloudy objects (e.g. Grant et al. 2023; Kempton et al. 2023), and pushed the limits of the smallest planets we can observe (e.g. Lustig-Yaeger et al. 2023; Moran et al. 2023; Kirk et al. 2024). In the previous few decades, our space-based spectroscopic analyses of exoplanet atmospheres have been limited to the Hubble Space Telescope (HST) and Spitzer Space Telescopes. One of the most powerful tools of

characterization with HST has been via transmission spectra. HST is able to provide transmission spectra in the 0.2–1.7 μm range with the Wide Field Camera 3 (WFC3) instrument via its UVIS and NIR channels (0.2–1.0 μm and 0.85–1.7 μm, respectively). Its Space Telescope Imaging Spectrograph (STIS) can also measure transmission spectra in the 0.115–1.0 μm range.

Transmission spectra from WFC3 have been used considerably over the years to characterize the atmospheres of exoplanets, with the first observation being the super-Earth GJ 1214b (Berta et al. 2012), yielding a featureless spectrum. Over the last 10 yr, the majority of the WFC3 spectral analysis has focused on measuring the water (or methane) feature at 1.4 μm (e.g. Deming et al. 2013; Wakeford et al. 2013; Kreidberg et al. 2014b). However, various studies have also looked at the signatures from clouds and hazes in WFC3 transmission spectra (e.g. Gibson et al. 2012; Kreidberg et al. 2014a; Sing et al. 2015). As the number of published transmission spectra from WFC3 grew, statistical studies emerged, comparing results across a range of exoplanets, to explore potential trends between parameters such as metallicity and mass, water abundance, clouds, and temperature (e.g. Sing et al. 2016; Barstow et al. 2017; MacDonald & Madhusudhan

* E-mail: chloe.fisher@physics.ox.ac.uk (CF);
jake.taylor@physics.ox.ac.uk (JT)

2017; Fisher & Heng 2018; Tsiaras et al. 2018; Pinhas et al. 2019; Welbanks et al. 2019).

For a single JWST instrument, the closest comparison to WFC3 is the Near Infrared Imager and Slitless Spectrograph (NIRISS; Doyon et al. 2023), covering 0.6–2.8 μm with its Single Object Slitless Spectroscopy (SOSS) mode (Albert et al. 2023), capable of providing transmission spectra. Note that although the G140 mode of the Near Infrared Spectrograph (NIRSpec) covers 0.7–1.84 μm and could also be compared to WFC3, there are no exoplanet transmission observations using it, either current or planned, since NIRISS is a more optimal instrument for the majority of cases. Due to its two filters, G140 can only observe either 0.7–1.27 μm or 0.97–1.84 μm , so multiple transits would need to be taken to cover the full wavelength range. Nevertheless, NIRISS can face issues with contaminated fields or saturation with bright stars, so may not be suitable for all targets. However, newly available in the upcoming JWST cycle 4 is the Near Infrared Camera’s (NIRCam) short wavelength grism time-series mode (Schlawin et al. 2016), allowing spectroscopic observations of exoplanets around bright stars at 0.6–2.3 μm . This could provide an ideal option for targets on which NIRISS would saturate.

Approximately 27 proposals for exoplanet transmission in JWST cycles 1, 2, and 3 will use NIRISS/SOSS, leading to many opportunities to take advantage of what this instrument has to offer. In light of previous WFC3 results and current and upcoming NIRISS observations, a comparative question about the instruments presents itself – can we quantify the information gained from NIRISS compared to WFC3? And furthermore, can we determine the source of this information? These are the key questions motivating the current study.

The hot gas giant WASP-39 b was chosen as a case study for the JWST Transiting Exoplanet Early-Release Science (ERS) programme, which provided the community with transmission spectra of this planet from several instruments, including NIRISS. These data, along with previous observations of WASP-39 b with HST (Wakeford et al. 2018), provide the opportunity to perform a direct comparison of NIRISS and WFC3 data. In this study, we compare constraints from atmospheric retrievals of the NIRISS and WFC3 transmission spectra of WASP-39 b. Furthermore, we attempt to determine the driving factor behind the information gain for each model parameter. Since the NIRISS wavelength range entirely encompasses that of WFC3, by cutting the NIRISS spectra down to this range, we can test whether the increase in wavelength coverage or spectral resolution is more important. We stress that the intention of this study is to compare the information content of WFC3 and NIRISS rather than to make confident conclusions about the nature of the atmosphere of WASP-39 b.

Although JWST has a number of instruments spanning a wide wavelength range, it lacks coverage of the optical wavelengths. Several studies have shown the importance of optical data in characterizing exoplanets through transmission spectroscopy (e.g. Lecavelier Des Etangs et al. 2008; Benneke & Seager 2012; Griffith 2014), and recently Fairman, Wakeford & MacDonald (2024) demonstrated the complimentary nature of HST/STIS (covering the 0.2–1.0 μm range) when combined with JWST data, in particular for constraining cloud parameters and alkali species. However, combining spectra from different instruments can be challenging (e.g. Yip et al. 2021), particularly if the data are reduced in distinct ways, assuming different orbital parameters for the planetary system. In this work, we test the effect of combining the NIRISS and WFC3 transmission spectra of WASP-39 b with data from STIS to establish if this can add useful information in either case and assess the complications

that can arise from combining data sets with inconsistent reduction methods.

1.1 Exoplanets in transmission with NIRISS/SOSS

Since the start of JWST observations, NIRISS has been used in a number of exoplanet transmission studies, the first one was obtained through the Early Release Observations (ERO) of HAT-P-18b (Fu et al. 2022) and WASP-96b (Radica et al. 2023). As part of the Early Release Science programme, Feinstein et al. (2023) presented the first NIRISS exoplanet transmission spectrum of WASP-39b. A single-transit observation of the Saturn-mass exoplanet WASP-39 b showed clear water features in multiple bands, the potassium doublet, and evidence of clouds. NIRISS was also used to observe the rocky exoplanet Trappist-1 b (Lim et al. 2023), which displayed signatures of stellar contamination in its transmission spectra. A follow-up study of NIRISS/SOSS observations of WASP-96b presented in Radica et al. (2023) was able to constrain the abundances of H₂O, CO₂, and K in its atmosphere (Taylor et al. 2023). Madhusudhan et al. (2023) combined NIRISS/SOSS and NIRSpec/G395H transmission spectra of the habitable zone sub-Neptune K2-18 b and found strong detections of CH₄ and CO₂, but a lack of H₂O, contradicting the original interpretations from this planet’s WFC3 spectrum. Radica et al. (2024) found muted spectral features in the atmosphere of the hot Neptune LTT 9779 b using NIRISS/SOSS, suggesting either an extremely high-metallicity atmosphere or high-altitude clouds. Fournier-Tondreau et al. (2024) reanalysed the ERO NIRISS/SOSS transmission spectrum for the Saturn-mass planet HAT-P-18 b. They demonstrated that the wavelength coverage provided by these data enables one to break the degeneracy between a cloudy atmosphere and one with a high metallicity and distinguish between water features due to the planet’s atmosphere and those from unocculted star-spots. Benneke et al. (2024) used NIRISS/SOSS and NIRSpec/G395H observations of the two-earth-radius exoplanet TOI-270 d to detect signals of CH₄, CO₂, and H₂O, as well as potential signs of SO₂ and CS₂. The abundances of multiple molecules allowed them to constrain the mean-molecular weight of the atmosphere and infer the metal mass fraction. Cadieux et al. (2024) presented two transits of LHS 1140 b, a planet in the radius valley, and found evidence of unocculted faculae, with possible signs of Rayleigh scattering from an N₂-dominated atmosphere on the planet.

Over the past few years, NIRISS/SOSS has been highly successful in characterizing the atmospheres of transiting exoplanets, enabling precise measurements of water features, detecting other molecules and clouds, and showing clear signatures of stellar activity. Of all the instruments on JWST, NIRISS/SOSS is the only one designed specifically for obtaining spectra of exoplanets. Whilst an increase in constraining power of this state-of-the-art instrument over WFC3 is to be expected, the extent of this improvement is unclear. This motivates our comparative study of the capabilities of NIRISS/SOSS and WFC3 for exoplanet characterization.

1.2 A case study of WASP-39b

As part of the JWST Transiting Exoplanet ERS programme (Stevenson et al. 2016; Bean et al. 2018), the hot gas giant exoplanet WASP-39 b was chosen to be observed with four of JWST’s instruments: NIRCam (F322W2 filter, 2.4–4.0 μm , $R \sim 2000$; Ahrer et al. 2023), NIRISS/SOSS (0.6–2.8 μm , $R \sim 1000$; Feinstein et al. 2023), NIRSpec G395H (2.87–5.14 μm , $R \sim 2700$; Alderson et al. 2023), and NIRSpec PRISM (0.6–5.3 μm , $R \sim 100$; Rustamkulov et al. 2023). WASP-39 b proved an ideal target for the ERS programme due to its

low density, providing large features in the transmission spectrum, and previous H₂O detection (Wakeford et al. 2018), as well as optimal timing with the observations. The ERS data provided a first look for the community at the capabilities of each instrument and immediately revealed new insights into the atmosphere of WASP-39 b, such as absorption from CO₂ (JWST Transiting Exoplanet Community Early Release Science Team et al. 2023) and the first detection of SO₂ (Tsai et al. 2023).

Previous analyses of the pre-JWST spectra of WASP-39 b had already revealed its large spectral features, such as the strong sodium and potassium absorption lines with pressure broadened wings (Fischer et al. 2016; Nikolov et al. 2016; Sing et al. 2016), which provided evidence of a predominantly clear atmosphere. Analyses of the infrared spectra of WASP-39 b offered a range of possible H₂O abundances and metallicities. Barstow et al. (2017) used STIS and *Spitzer* observations and obtained an upper limit on the water abundance of about $1 \times$ solar. Wakeford et al. (2018) used a combination of data from WFC3, STIS, the Very Large Telescope’s (VLT) Focal Reducer and low dispersion Spectrograph (FORIS2) instrument, and *Spitzer* and constrained the water abundance to $100\text{--}200 \times$ solar. Using only the WFC3 data, Fisher & Heng (2018) found a water abundance of $4\text{--}24 \times$ solar. Welbanks et al. (2019) used the STIS, WFC3, and *Spitzer* data to constrain the water abundance to $4\text{--}13 \times$ solar. Tsiaras et al. (2018) rereduced the WFC3 data and found a substantially reduced water abundance of $0.0001\text{--}0.01 \times$ solar. Pinhas et al. (2019) again used the combination of STIS, WFC3, and *Spitzer* to constrain the water abundance to $0.01\text{--}0.05 \times$ solar. Min et al. (2020) also used these instruments and found a water abundance of $2\text{--}50 \times$ solar. Kirk et al. (2019) further combined these data sets with a spectrum from the William Herschel Telescope and recovered a metallicity of $224\text{--}347 \times$ solar. This wide array of values highlights the possible uncertainty in constraining abundances from space-based data pre-JWST and stresses the potential information to be gained from the increase in precision, resolution, and wavelength coverage of the JWST instruments. The current study provides an in-depth analysis of the information content of JWST’s NIRISS instrument in comparison with WFC3 and determines the driving factor behind the improved constraints on a planet’s atmospheric properties in the JWST era.

1.3 Layout of the study

In Section 2, we discuss the methods we use, including the origin of the data sets and the atmospheric retrieval codes, along with their respective setups. In Section 3, we analyse our results across the various retrievals, comparing the constraints from WFC3 and NIRISS, and the effect of adding STIS. Additionally, we compare the results from the WFC3 retrievals with retrievals on NIRISS spectra cut to the same wavelength range. We also perform mock retrievals on simulated data in order to test our conclusions. In Section 4, we discuss the implications of our results and make recommendations for future best practices. Finally, in Section 5, we summarize our conclusions.

2 METHODS

2.1 Data

In this study, we consider the transmission spectra of the hot-Jupiter WASP-39 b obtained from HST/WFC3, HST/STIS, and JWST NIRISS/SOSS. The WFC3 and STIS data are sourced from Wakeford et al. (2018) and Fischer et al. (2016) and Sing et al. (2016), respectively, whilst the NIRISS data come from the Exoplanet Early

Release Science Team (Feinstein et al. 2023). For consistency with Carter et al. (2024), we use the `supreme-SPOON`¹ reduction from Feinstein et al. (2023). As described in Feinstein et al. (2023), the raw, uncalibrated data frames were reduced using a combination of steps from the default JWST calibration pipeline and custom routines, including a group-level correction of time-variable $1/f$ noise as is commonly performed for JWST’s other instruments (e.g. Alderson et al. 2023; Rustamkulov et al. 2023). The stellar spectra were extracted from the 2D data frames using the `ATOCA` and `APPLESOSS` algorithms (Darveau-Bernier et al. 2022; Radica et al. 2022) to explicitly model the overlap of the first and second diffraction orders on the detector. The light curves were fit at the pixel-level (that is, one light curve per detector column) using the `juliet` library (Espinoza, Kossakowski & Brahm 2019). Along with the `BATMAN` transit model, we included in the fits a systematics model consisting of a linear trend with the x -position of the spectral trace on the detector. The fitted transit depths were then binned to a constant $R = 100$.

These data sets were reduced using different techniques, pipelines, and prior assumptions, which could lead to inhomogeneities such as offsets across the instruments (Mugnai et al. 2024). However, we remain agnostic to these differences in an effort to highlight some of the issues involved with combining spectra from different papers, when a homogeneous reanalysis of the raw data may not be possible. Furthermore, even in the case of homogeneously analysed data, differences can still arise, such as offsets between instruments (Madhusudhan et al. 2023; Benneke et al. 2024; Welbanks et al. 2024), changes in stellar activity between epochs (Rosich et al. 2020), or variability in the planetary atmosphere (Morello et al. 2023; Parviainen 2023; Changeat et al. 2024).

Fig. 1 shows the three transmission spectra for WASP-39 b. Whilst spectral features are clearly visible in all three spectra, there are some obvious discrepancies between the data sets. First, the WFC3 data have a slight offset from the NIRISS data at the $1.4 \mu\text{m}$ water feature. Secondly, many of the STIS data points appear inconsistent with the blue end of the NIRISS data, although the error bars on these are large.

2.2 Retrieval setup

In order to perform a comparison across all three data sets, we apply atmospheric retrievals to these spectra. This involves using a Bayesian sampling algorithm, such as nested sampling or a Markov chain Monte Carlo (MCMC), coupled with an atmospheric model, to search parameter space for the values that could explain the spectroscopic observations (e.g. Madhusudhan & Seager 2009; Benneke & Seager 2013). There are many choices involved in building a retrieval code, such as the model assumptions, molecular line-lists, and sampling method. We test two different retrieval codes in this work: `BeAR` (formerly known as `Helios-r2`; Kitzmann et al. 2020) and `CHIMERA` (Line et al. 2013). For a comprehensive table of current atmospheric retrieval codes for exoplanets, see MacDonald & Batalha (2023).

2.2.1 BeAR

`BeAR`² is a 1D, open-source, graphics processing unit (GPU) - accelerated atmospheric retrieval code (Kitzmann et al. 2020), using

¹This pipeline is now known as `exoTEDRF` (Radica 2024)

²The open-source code `Helios-r2` has recently been superseded by a new version that has been renamed to `BeAR`, the Bern Atmospheric Retrieval

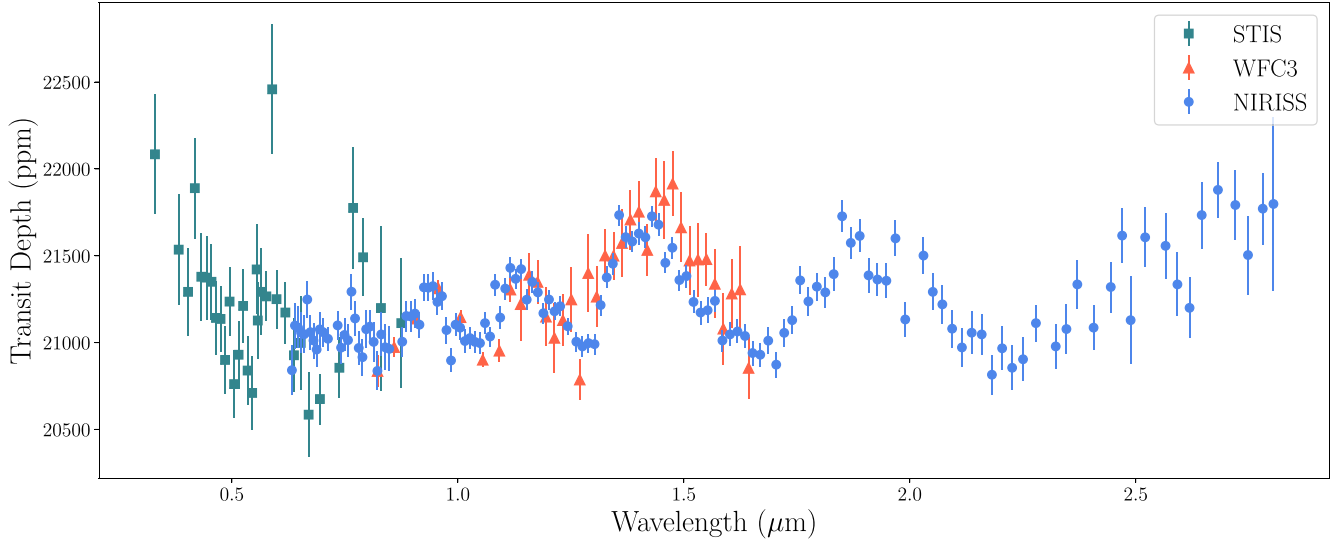


Figure 1. STIS, WFC3, and NIRISS transmission spectra for WASP-39 b. The data are plotted as they are published in Fischer et al. (2016), Sing et al. (2016; STIS), Wakeford et al. (2018; WFC3), and Feinstein et al. (2023; NIRISS), with no offsets added.

nested sampling via the MULTINEST package (Feroz & Hobson 2008; Feroz, Hobson & Bridges 2009). It was originally developed for analysing the emission spectra of brown dwarfs, to which it has been applied extensively (Lueber et al. 2022). It has since been updated for exoplanets, for both transmission and secondary eclipse spectra. BeAR uses line-by-line opacity sampling to describe the wavelength-dependent absorption and scattering. In this work, we include the following molecules and their associated EXOMOL and HITEMP line-lists: H₂O (Polyansky et al. 2018), CO (Li et al. 2015; Somogyi & Yurchenko 2021), CO₂ (Rothman et al. 2010), and CH₄ (Yurchenko et al. 2013; Yurchenko & Tennyson 2014). Additionally, we include the alkalis, Na and K, for which we use the Kurucz line-lists (Kurucz & Bell 1995). The pressure broadening of their strong resonance lines are calculated with data from Allard et al. (2019) and Allard, Spiegelman & Kielkopf (2016), see Kitzmann et al. (2020) for details. The opacities are calculated using HeLIOS-k (Grimm & Heng 2015; Grimm et al. 2021) at a resolution of 0.01 cm⁻¹ and can be found on the DACE opacity data base.³ In this study, the models are sampled at a resolution of 1.0 cm⁻¹. We assume the atmosphere is hydrogen dominated, with a solar He/H₂ ratio of 0.17 (Asplund et al. 2009). We include collision-induced absorption (CIA) from H₂-H₂ (Abel et al. 2011) and H₂-He (Abel et al. 2012), using the HITRAN data base, as well as Rayleigh scattering due to H₂ (Vardya 1962). A plot of all the opacity sources used by BeAR in this study is shown in Fig. A1.

BeAR can be run using the chemical equilibrium model FastChem⁴ (Stock et al. 2018; Stock, Kitzmann & Patzer 2022; Kitzmann, Stock & Patzer 2024), or using free chemistry, with independently retrieved molecular volume mixing ratios X . We opt for the latter, in order to avoid assumptions about the chemistry of the atmosphere. BeAR also has the option to include variable abundance profiles for the chemical species, though we do not use this. The

code, which is used in this study. This code can be found here: <https://github.com/newstrangeworlds/bear>.

³<https://dace.unige.ch>

⁴The open-source code FastChem can be found here: <https://github.com/newstrangeworlds/FastChem>.

atmosphere in BeAR is divided into equal layers in log-pressure space. We use a top pressure of 10⁻⁶ bar and a bottom pressure of 10 bar, with 200 layers in between. This bottom, or ‘reference’, pressure corresponds to the level of the planet radius R_p .

When considering clouds, BeAR has the option to include either a grey or non-grey cloud layer. For the grey-cloud, one retrieves a cloud-top pressure P_{cloudtop} , a cloud optical depth τ_{cloud} , and a cloud-bottom pressure $P_{\text{cloudbottom}}$, set by a parameter b_c such that $P_{\text{cloudbottom}} = b_c P_{\text{cloudtop}}$. b_c effectively sets the size of the cloud layer. In the non-grey cloud model, one still retrieves the cloud top and bottom, but instead of a constant optical depth, we now have an equation depending on wavelength,

$$\tau(\lambda) = \tau_{\text{ref}} \frac{Q_0 x_{\lambda_{\text{ref}}}^{-a_0} + x_{\lambda_{\text{ref}}}^{0.2}}{Q_0 x_{\lambda}^{-a_0} + x_{\lambda}^{0.2}}, \quad (1)$$

where $x_{\lambda} = 2\pi r_{\text{cloud}}/\lambda$. The additional free parameters are then Q_0 (a proxy for the cloud composition), a_0 (the power law index in the small particle limit), and r_{cloud} (the cloud particle radius). For a detailed explanation of this cloud model, see Kitzmann & Heng (2018) or Lueber et al. (2022).

In addition to the atmospheric parameters mentioned above, BeAR also retrieves the planet and star parameters, such as the surface gravity $\log g$, the planet radius R_p at the bottom of the assumed atmospheres, and the radius of the host star R_* . For the purposes of this study, we fix the surface gravity and stellar radius [$\log g = 2.63$, $R_* = 0.9R_{\odot}$ (Faedi et al. 2011)], and only retrieve the planet radius at a pressure of 10 bar. The priors used for all the possible model parameters are found in Table 1.

2.2.2 CHIMERA

We use CHIMERA⁵ to perform both free and chemically consistent spectral retrievals. CHIMERA is the only framework in this study that uses the correlated- k approach (Lacis & Oinas 1991) when

⁵The open-source CHIMERA code can be found here: <https://github.com/mrlinc/CHIMERA>

Table 1. Prior distributions for all possible model parameters for BeAR.

Parameter	Unit	Prior distribution	Limits
R_p	R_J	Uniform	[1.1, 1.4]
$X_{\text{H}_2\text{O}}$	–	Log-uniform	$[10^{-12}, 1.0]$
X_{CO}	–	Log-uniform	$[10^{-12}, 1.0]$
X_{CO_2}	–	Log-uniform	$[10^{-12}, 1.0]$
X_{CH_4}	–	Log-uniform	$[10^{-12}, 1.0]$
X_{Na}	–	Log-uniform	$[10^{-12}, 1.0]$
X_{K}	–	Log-uniform	$[10^{-12}, 1.0]$
T (isothermal)	K	Uniform	[500, 3000]
T_{irr} (non-isothermal)	K	Uniform	[500, 3000]
κ_{IR}	cm^2g^{-1}	Log-uniform	$[10^{-4}, 10]$
γ	–	Log-uniform	$[10^{-2}, 100]$
τ_{cloud} (grey)	–	Log-uniform	$[10^{-5}, 20]$
τ_{ref} (non-grey)	–	Log-uniform	$[10^{-5}, 20]$
Q_0	–	Uniform	[1, 100]
a_0	–	Uniform	[3, 6]
r_{cloud}	cm	Log-uniform	$[10^{-7}, 10^{-1}]$
P_{cloudtop}	bar	Log-uniform	$[10^{-5}, 10]$
b_c	–	Log-uniform	$[1, 10^3]$

computing transmission through the atmosphere. The k -tables are computed at a resolution of $R = 1000$; the line-by-line data used to calculate the k -tables are from the following sources: H_2O (Freedman et al. 2014; Polyansky et al. 2018), CO_2 (Freedman et al. 2014), CO (Rothman et al. 2010), CH_4 (Rothman et al. 2010), Na (Kramida et al. 2018; Allard et al. 2019), and K (Kramida et al. 2018; Allard et al. 2016); and they were computed following the methods described in Gharib-Nezhad et al. (2021) and Grimm et al. (2021). Again, we assume the atmosphere is dominated by H_2 , with a solar He/H_2 ratio of 0.1764; therefore, we also model the H_2 - H_2 and H_2 - He CIA (Richard et al. 2012). The CHIMERA framework has been extensively used to analyse the transmission spectra of exoplanets obtained with HST/WFC3 and JWST (e.g. Kreidberg et al. 2014b; Taylor et al. 2023) and is part of the cutting edge frameworks being used in the Early Release Science collaboration (Feinstein et al. 2023).

In CHIMERA, we treat the inclusion of clouds in three ways: no cloud, grey cloud, or non-grey cloud. For our grey cloud, we fit for a cloud top pressure. At pressures larger than this, we set the atmosphere to be completely opaque. For our non-grey model, we employ the Ackerman & Marley (2001) cloud model to produce realistic vertical droplet profiles over a broad range of droplet sizes (for a specific condensate), given a sedimentation efficiency (f_{sed}), eddy diffusivity (K_{zz} cm^2/s), cloud base pressure, and the condensate mixing ratio at the cloud base. Here we assume enstatite (MgSiO_3) physical/optical properties for the condensate as a representative non-grey cloud at short wavelengths, following the conclusions of Gao et al. (2020). However, the cloud top pressure is a separate free parameter, not linked with the condensation curve of enstatite.

2.2.3 Temperature-pressure profiles

Previous studies have shown that, even in the case of transmission spectra, non-isothermal profiles can be required to accurately retrieve the molecular abundances from JWST data (e.g. Rocchetto et al. 2016). This motivates us to test both isothermal and non-isothermal retrievals in our study. Both BeAR and CHIMERA have the option to include different temperature-pressure parametrizations, and for this study, we adopt the ‘Guillot’ profile (Guillot 2010; Parmentier & Guillot 2014). In this work, the parametrization follows Guillot (2010; the semi-grey case), where the parameters are the infrared

opacity κ_{IR} , the irradiation temperature T_{irr} , the internal temperature T_{int} , the ratio between optical and IR opacity γ , and the redistribution factor f . In our retrievals, we fix the internal temperature to 200 K, and f to 0.25 to correspond to an averaging of the incoming flux over the planetary surface. This leaves three free parameters, the priors for which are shown in Table 1.

2.2.4 Model complexity

For each of our retrieval codes, we test a range of different retrieval models, with increasing complexity. We start with an isothermal, cloud-free model, then add a grey-cloud, and a non-grey cloud or haze, then do the same for the non-isothermal model. This leads to six different models to test for each individual data set and various combinations for each code. It is worth noting that the two codes have different parametrizations for the non-grey cloud model. Although the ERS study of the NIRISS spectra of WASP-39b found evidence for spatially inhomogeneous cloud cover in the planet’s atmosphere (Feinstein et al. 2023), we opt not to include this level of complexity in our model, in order to limit the number of test cases, and because the wavelength region driving this model selection (2.0–2.3 microns) is not encompassed in the WFC3 range.

The Bayesian Evidence, obtained from a nested sampling retrieval, can be compared to determine when one model is more favoured by the data than another. The Bayesian Evidence penalizes the addition of parameters that do not sufficiently improve the fit to the data, thus providing a method of formally implementing Occam’s Razor. Trotta (2008) gives the difference in log Bayesian Evidence (or the logarithm of the Bayes factor, $\ln B_{ij}$) corresponding to inconclusive, weak, moderate, and strong evidence for model i over model j as $\ln B_{ij} < 1.0$, $1.0 < \ln B_{ij} < 2.5$, $2.5 < \ln B_{ij} < 5.0$, and $\ln B_{ij} > 5.0$, respectively.

3 RESULTS

3.1 BeAR versus CHIMERA

In this section, we compare results from our two retrieval codes. Fig. 2 shows a summary of the retrieved water abundances and temperatures for each of the codes applied to the WFC3 and NIRISS data for the different models. The codes generally agree to within 1- or 2σ , with the largest disagreement occurring for the most complex model (non-isothermal + non-grey cloud), applied to the NIRISS data. Given the similarity of the T-P profiles in the cloud-free case for NIRISS, this disagreement is likely due to the different parametrizations of clouds/hazes (see Sections 2.2.1 and 2.2.2). Though the two codes use different line lists for some of the species, this is unlikely to cause major differences across the WFC3 and NIRISS wavelengths, since the water line list is consistent between the two, and this is the main absorber in this spectral region. These discrepancies in the codes’ results highlight the sensitivity of retrievals to minor differences in the models, particularly for data with the precision of JWST.

Our study focuses on a comparison of the different types of data rather than aiming to make confident claims about the atmosphere of WASP-39 b. Therefore, in the rest of the paper, we will only use BeAR to perform the retrievals to save additional computations. However, this code comparison demonstrates the need to compare results from multiple retrieval codes with different assumptions to ensure any claims are not model dependent. This is particularly important when analysing JWST data due to its increased precision and sensitivity. For a comprehensive comparison of different retrieval codes applied to JWST data, see Welbanks et al. (in preparation).

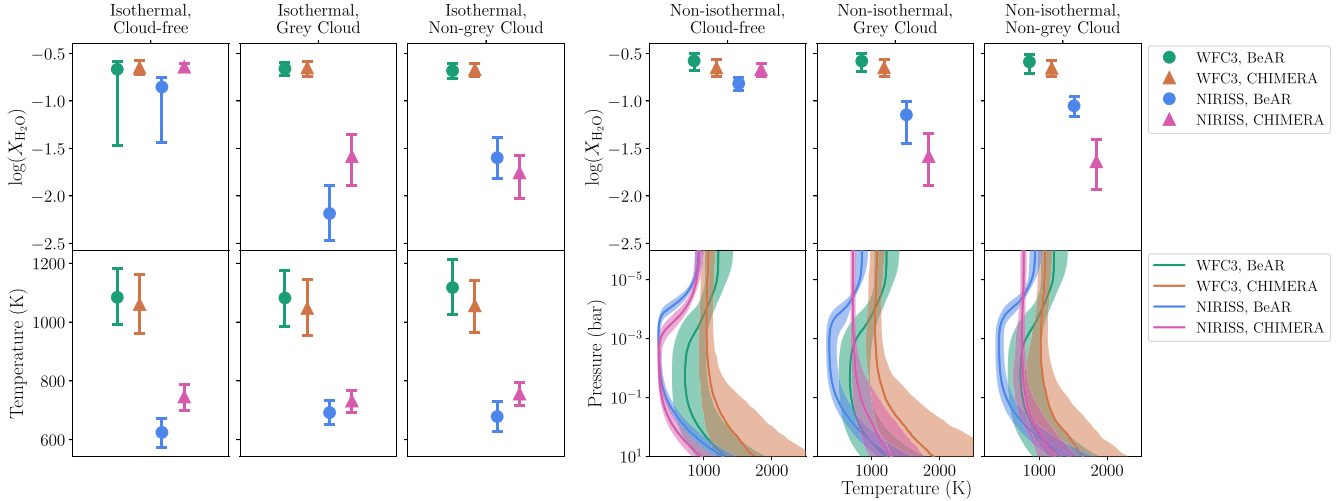


Figure 2. A summary of the retrieval results for the different retrievals applied to the WFC3 and NIRISS data using BeAR and CHIMERA. The symbols and errorbars denote the median and 1σ values from the posterior distributions. The top row shows the retrieved water abundances for each case. The bottom row shows the retrieved temperatures, where the three left-hand panels display single isothermal temperatures and the three right-hand panels show the full T-P profiles for the non-isothermal cases. It can be seen that the two codes agree to within 1σ for the majority of water abundances and isothermal temperatures. The non-isothermal temperatures slightly deviate owing to the differences in the codes, such as cloud parametrization.

3.2 WFC3 versus NIRISS

In this section, we compare the results from the BeAR retrievals on the WFC3 and NIRISS data. The left-hand panels of Fig. 3 show the Bayesian evidence hierarchy for the retrievals on these individual data sets. In the WFC3 case, the Bayesian model comparison is inconclusive across all models, implying there is no real preference to fit this data set. For NIRISS, on the other hand, the non-isothermal + non-grey cloud model is the best-fitting model, whilst the other models are either weakly, moderately, or, in the case of the cloud-free models, strongly ruled out by the Bayesian evidence. With this in mind, Fig. 4 compares the retrievals for the non-isothermal + non-grey cloud model for the WFC3 and NIRISS spectra.

There is a clear discrepancy between the water abundance posteriors from the two retrievals. WFC3 gives an extremely high value of $\log X_{\text{H}_2\text{O}} = -0.59^{+0.08}_{-0.12}$ corresponding to an atmosphere of $\sim 25\%$ water, whilst NIRISS gives a lower value of $\log X_{\text{H}_2\text{O}} = -1.05^{+0.10}_{-0.11}$. This is likely due a difference in shape of the water feature, seen in Fig. 1. Without the wider wavelength coverage, the WFC3 data hold little information about the continuum, providing a challenge for the retrieval to fit spectral features accurately. This high water abundance causes the mean molecular weight of the atmosphere to increase (to ~ 6.4), causing a decrease in the atmospheric scale height, squashing spectral features, and thus allowing an extremely high abundance to fit a moderately-sized water feature. Previous studies have shown that, in the first-order analytical case and when only one absorber is present, it is not possible to measure its abundance from the transmission spectra due to degeneracies (Lecavelier Des Etangs et al. 2008; Benneke & Seager 2012; Line & Parmentier 2016), except in the case of very high abundance, due to the mean-molecular weight effect, or very low abundance, where the continuum due to CIA sets a minimum (Welbanks et al. 2019). This degeneracy is broken when the effects of temperature and pressure on the species' opacity is taken into account. However, these effects are quite weak, and the large error bars and low-resolution of WFC3 data make them hard to distinguish. This is consistent with our extremely high value for the

water abundance retrieved from WFC3 data alone. The difference in shape of the water feature between WFC3 and NIRISS could be due to different systematics in the instruments, but the lack of additional features in WFC3 likely make the retrieval more sensitive to these changes.

Since the narrow wavelength range of WFC3 restricts any detections of molecules to H_2O only, the other molecules unsurprisingly have unconstrained posteriors. In contrast, the NIRISS retrieval shows some evidence for CO_2 and K and upper limits for CO, CH_4 , and Na. These additional absorbers allow for abundance constraints not only at extreme values, as suggested by Line & Parmentier (2016). To compute the detection significances of the constrained species, we ran additional retrievals for the non-isothermal + non-grey cloud case, omitting each molecule in turn. For the WFC3 retrievals, we find a detection significance of 9.0σ for H_2O . In comparison, the NIRISS retrievals obtain significances of 21.3σ for H_2O , 2.8σ for CO_2 , and 2.1σ for K. This demonstrates a quantifiable gain in information in the NIRISS data when compared to WFC3, providing a more robust detection of water, as well as the ability to probe additional species.

For the clouds, the WFC3 retrieval shows unconstrained posteriors for every cloud parameters. This is unsurprising as the cloud-free models are equally as preferred by the Bayesian evidence, suggesting clouds are not detected. The posteriors spanning the entire prior range for each cloud parameter also indicates that clouds would not be detected in this data, even if they were present in the atmosphere. This is likely due to the extremely high retrieved water abundance, preventing any visible effect of a cloud. In contrast, the NIRISS retrieval shows a clear cloud detection, with a significant reference optical depth, and a cloud top pressure of ~ 10 mbar, placing it in the region predicted to be probed by transmission spectroscopy.

For the temperature-pressure profile, the WFC3 retrieval shows a fairly unconstrained profile, with some signs of an inversion in the upper atmosphere. In the NIRISS case, the profile appears to have a similar shape, but more tightly constrained and reaching cold temperatures in the middle of the atmosphere. These cold temperatures may not be realistic for a planet with an equilibrium

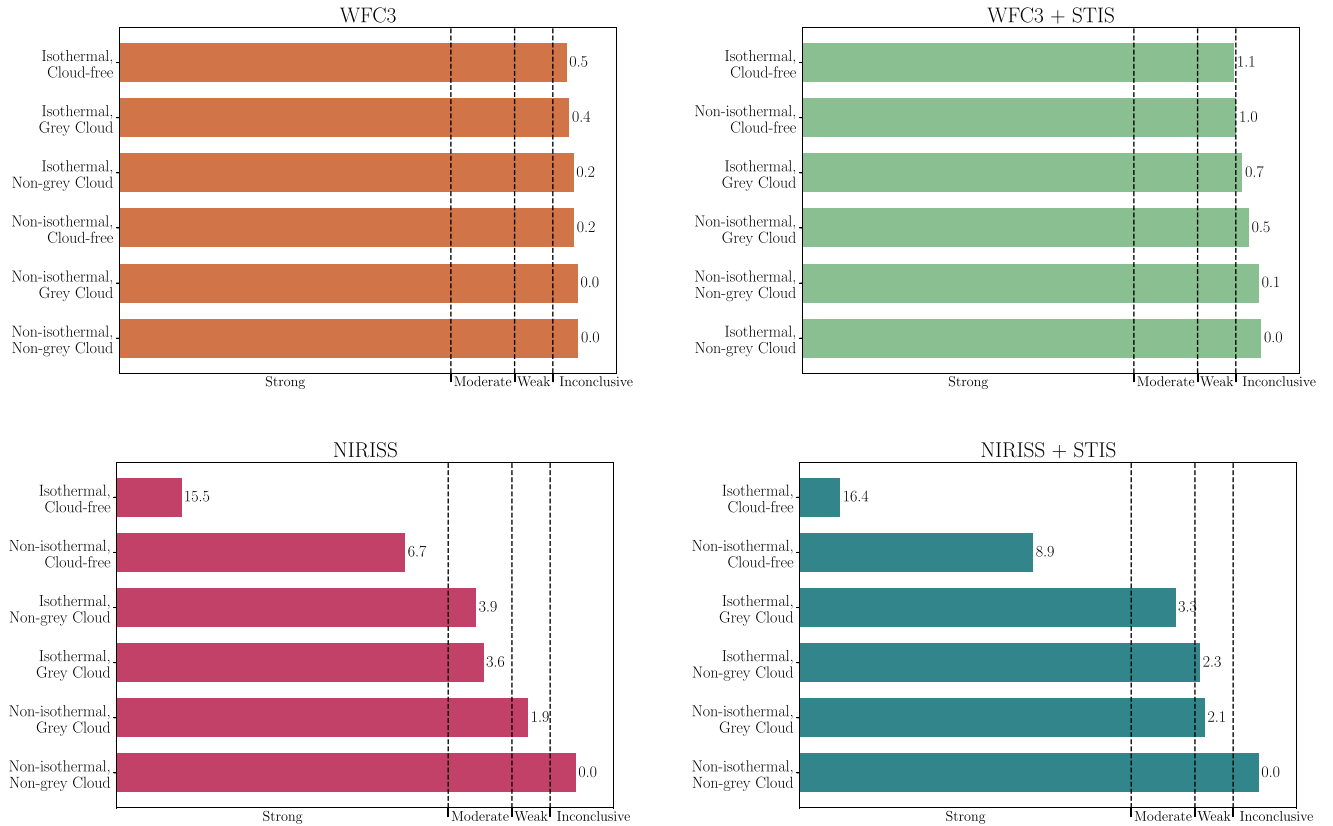


Figure 3. Bayesian evidence hierarchy for the different models used in each retrieval for each data set. The models are sorted by Bayesian evidence, with the most favoured model at the bottom. Bayes factor values are displayed next to the bars, comparing each model to the most favoured model. The x -axis shows the division between inconclusive, weak, moderate, and strong evidence for ruling out models with respect to the most favoured model. These boundaries correspond to significances of 2.1σ (weak), 2.7σ (moderate), and 3.6σ (strong).

temperature of 1100 K. The inversions seen at high altitudes are likely caused by the model requiring a hotter temperature to fit the alkalis (potassium in this case) high in the atmosphere. However, this model is not entirely consistent as it fits a pressure-dependent temperature alongside molecular abundances that are constant in altitude. It also assumes a single temperature profile across the entire terminator, neglecting variations in longitude and latitude. Therefore, these retrieved T-P profiles are unlikely to be an accurate representation of the vertical structure of the atmosphere. Furthermore, there is only moderate evidence favouring this model over several isothermal models, so this level of complexity is not necessarily required to fit the data. In general, WFC3 retrievals obtain higher temperatures than those of NIRISS, as shown in Fig. 5. The issues with the reliability of the temperature profiles represent a limitation of simplified 1D retrieval models that are commonly used in exoplanet analysis, which has been discussed in detail in, e.g. MacDonald, Goyal & Lewis (2020).

Our NIRISS retrievals give a water abundance of $\log X_{\text{H}_2\text{O}} = -2$ to -1 , depending on the model used. In comparison to literature values, this is generally in good agreement with other studies of the JWST data. Feinstein et al. (2023) constrain the C/O ratio and metallicity of the planet by grid-fitting self-consistent equilibrium models and find values that correspond to a log water abundance of -1.86 . Rustamkulov et al. (2023) use similar methods to analyse the planet’s atmosphere from the NIRSpec PRISM data and find values corresponding to slightly lower log water abundances (-3 to -2.5). This deviation between results could be due to the

assumption of chemical equilibrium and strong absorption features in the PRISM data from other molecules, such as CO_2 . Ahler et al. (2023) analyse the NIRCcam data and find C/O ratios and metallicities that correspond to $\log X_{\text{H}_2\text{O}} = -3.3$ to -1.4 . Lueber et al. (2024) retrieve a log water abundance of -1.45 to -1.13 from the NIRISS data, depending on the model complexity they use. Their analysis of the PRISM and NIRCcam spectra, however, also give lower abundances ($\log X_{\text{H}_2\text{O}} = -3.53$ to -3.10 and $\log X_{\text{H}_2\text{O}} = -4.15$ to -4.10 , respectively). Even more depleted H_2O abundance values are found from their analysis of the NIRSpec G395H data, but they discuss how the lack of full coverage of the water feature combined with a flexible non-grey cloud could be biasing those results. Another study of the PRISM spectrum of WASP-39b, Constantinou, Madhusudhan & Gandhi (2023), also found a lower water abundance ($\log X_{\text{H}_2\text{O}} = -4.85$ to -3.14). However, they only use the 3–5 micron range of the PRISM data and combine it with WFC3. Although not JWST data, Wakeford et al. (2018) use a combination of WFC3, STIS, VLT, and *Spitzer* data and find a log water abundance of -1.37 from their free chemistry retrievals. On the whole, this demonstrates the wide range of H_2O abundances that are still consistent with the data, even in the era of JWST.

3.3 Adding STIS

In this section, we look at the effect of adding the STIS data to each of the NIRISS and WFC3 spectra in turn. Again, we stress that the different reduction techniques could cause discrepancies between the

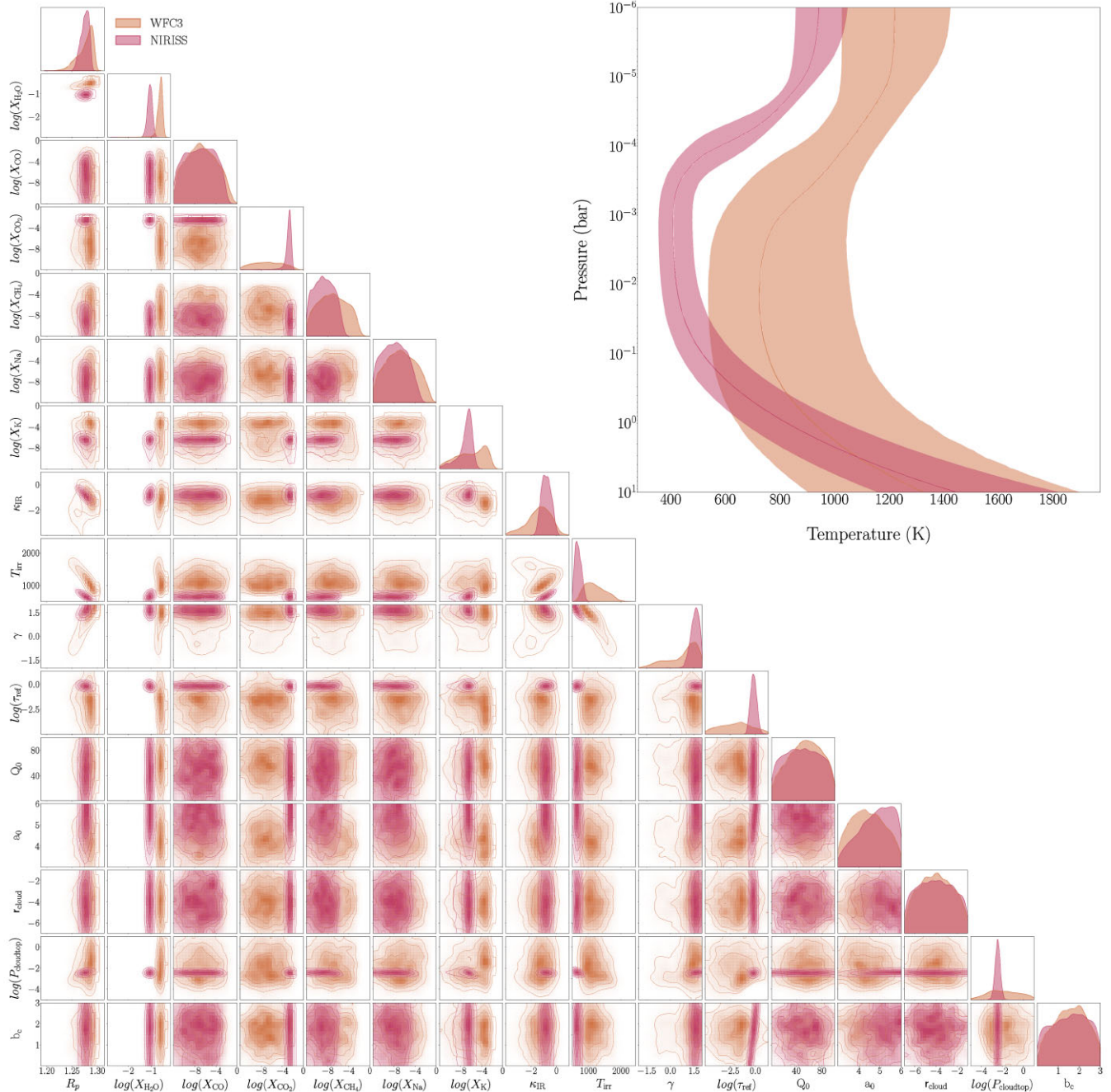


Figure 4. Retrievals using the non-isothermal + non-grey cloud model for the WFC3 (orange) and the NIRISS (red) data. The median and 1σ limits of the retrieved temperature-pressure profiles are shown in the top right corner.

data, but since we wish to test potential effects of this, we do not re-reduce the data before combining them.

The Bayesian evidence for these retrievals is shown in the two right-hand panels of Fig. 3. For WFC3 + STIS, we are now able to weakly rule out the cloud-free models, as the STIS data have widened the wavelength coverage and given us some information about the bluer points, where clouds have their strongest effect. However, the rest of the models are all comparable in terms of the evidence. For NIRISS + STIS, the evidence is similar to the NIRISS-only case, favouring the non-isothermal + non-grey cloud case and strongly ruling out cloud-free models.

3.3.1 Water and temperature

To compare the effect of adding STIS across all retrievals, Fig. 5 shows a summary of the retrieved water abundance and temperature for every model. For H_2O , adding STIS to WFC3 almost always brings the abundance down substantially, because it adds more information about the continuum and because the large features from Na and K, along with the visible optical slope (see Fig. 1), prevent a high mean-molecular weight atmosphere from fitting the spectrum. For the non-isothermal + non-grey cloud model, the cloud top pressure is driven to unphysically high values (see Fig. A2), hitting the edge of the prior at 10^{-5} bar (one order of magnitude below

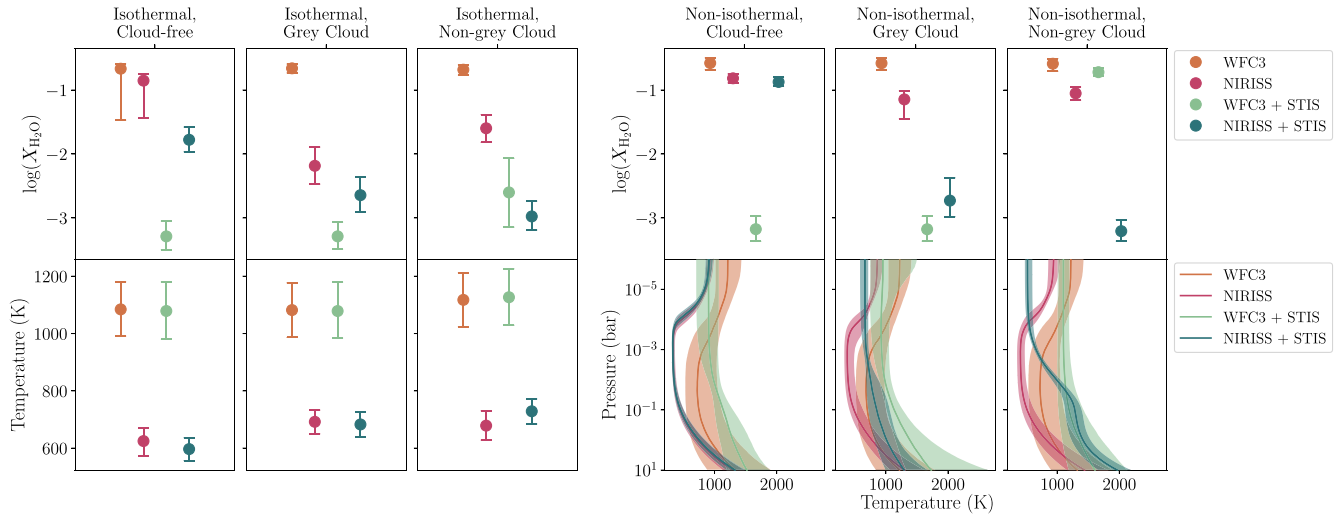


Figure 5. Summary of the retrieval results for the water abundance and temperature for all the models and all the data set combinations.

the top of the atmosphere in the model). In contrast, the retrieval obtains a low value for the cloud optical depth (τ_{ref}), balancing the high cloud top pressure. This is likely due to the extremely steep slope on the edge of the STIS data.

For NIRISS, adding STIS also brings the water abundance down, though with varying impact. For the isothermal models, adding STIS to either WFC3 or NIRISS has little-to-no effect on the retrieved temperature. For the non-isothermal models, adding STIS to WFC3 seems to increase the temperature at the bottom of the atmosphere and decrease it at the top, removing the temperature inversion and favouring a more isothermal profile. For NIRISS, the same effect appears for the cloudy models, but no change occurs when adding STIS for the cloud-free case. The smaller effect of adding STIS in the cloud-free cases can be explained by the model’s inability to adjust the fit to the STIS data without the cloud. When considering the non-isothermal cases, it is worth noting that we do not expect temperature profiles to be robustly retrieved from transmission spectra due to their narrow probed-pressure range, and the fact that the temperature only enters into the equation in two places – the scale height and the temperature-dependent opacities. The latter is a weak effect at low resolution, so the temperature profile constraints rely on probing different scale heights across the spectrum. Despite this, some of the T-P profiles appear very tightly constrained, even for altitudes at which there is no information from the transmission spectrum. It is possible these models are overfitting the data. It is also likely that the choice of T-P profile parametrization will have a large effect on the results (Blecic, Dobbs-Dixon & Greene 2017). Furthermore, since our retrievals assume constant mixing ratios with altitude, the temperature could be compensating for varying mean molecular weight in the atmosphere, though we do not explore this.

As with the WFC3 and NIRISS retrievals, we also run additional retrievals on the data sets combined with STIS, omitting particular species in order to compute their detection significances. For WFC3 + STIS, we find significances of 8.8σ for H_2O , 3.8σ for Na, and 2.6σ for K. In comparison, for NIRISS + STIS, we obtain significances of 21.3σ for H_2O and 2.3σ for Na. Interestingly, when K is removed, the Bayesian evidence of the retrieval improves, despite it having a well-constrained posterior (see Fig. A2). Possibly, the overlapping K feature in the NIRISS and STIS data proves

challenging for the retrieval to fit, thus removing the species is not detrimental.

Fig. 6 shows the reduced χ^2 values for the different retrievals for the different sections of data. In general, the addition of STIS to the retrieval does not seem to change the reduced χ^2 value for the WFC3 or NIRISS spectra. However, for the NIRISS + STIS retrievals, the reduced χ^2 values for the STIS section are further from 1 in all cases than the NIRISS section. This indicates that, although the addition of STIS data does seem to affect certain model parameters, such as the water abundance, the fit to the data is poor, and thus the effects may not be physical. This is further demonstrated in Fig. 7, which shows the posterior fits to the data from the NIRISS + STIS retrieval for the non-isothermal + non-grey cloud model. It is evident that the STIS slope, in particular, is not fitted well by this model, despite it having the highest Bayesian evidence. It is unknown whether the poor fit to the STIS data is due to instrument systematics or missing model complexity, such as stellar contamination or a high-altitude haze. Although our non-grey cloud model could fit a thick haze, this would worsen the fit in the infrared region, so is not favoured by the retrieval. In order to fit this very steep optical slope, Fairman et al. (2024) require an extreme scattering slope of $\gamma < -10$, far beyond the small particle limit of Mie theory (-4; Bohren & Huffman 1998). In turn, this drives the water abundance up to the top of their prior ($\log X_{\text{H}_2\text{O}} = -1.40^{+0.29}_{-0.76}$), though this value is consistent with our results. However, it is possible to produce a super-Rayleigh slope that could fit very steep optical data with a detached haze layer and a clear atmosphere beneath (Barstow et al. 2017), or with a photochemical haze formed in a vigorously mixing atmosphere (Kawashima & Ikoma 2019; Ohno & Kawashima 2020), both of which could provide an explanation for the STIS data. Recent work also considered the possibility of fitting these data with unocculted active regions in the star and found the retrieved molecular abundances for the planet were unchanged – a reassuring result for retrievals on JWST data (Welbanks et al., in preparation). If the slope is instead due to instrument systematics, this could affect the retrieved molecular abundances. We suggest that one should exercise caution when directly combining STIS data with new JWST transmission spectra, especially when the data were not reduced in a consistent way.

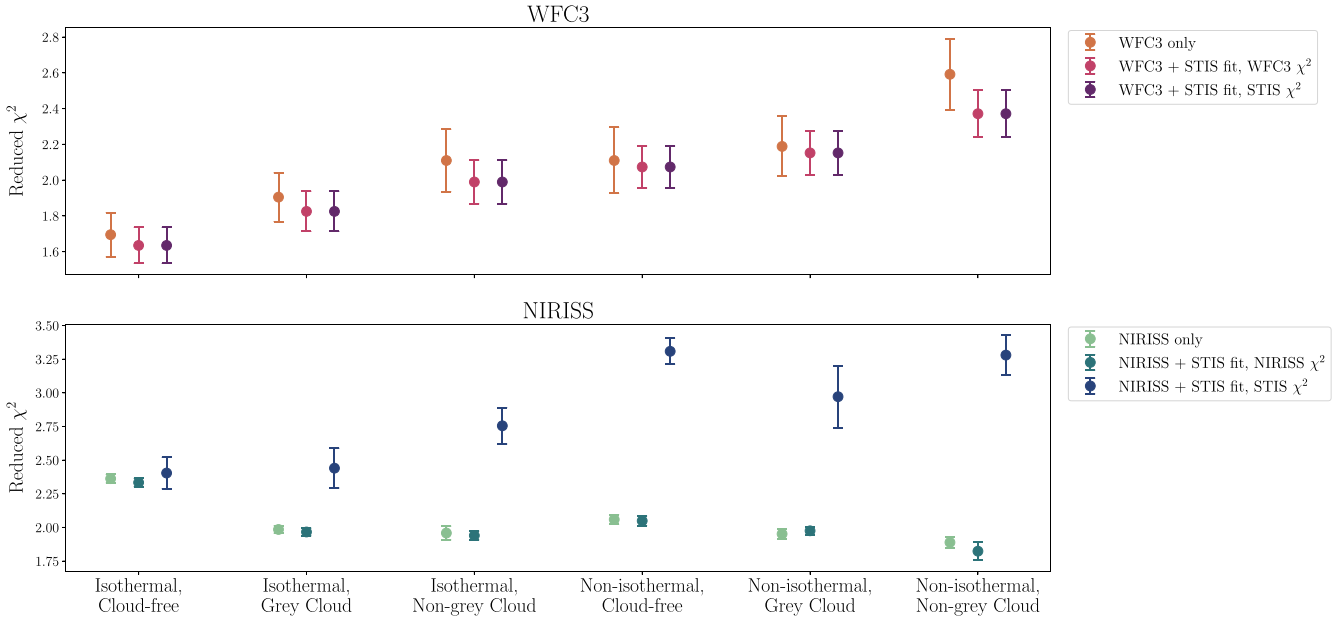


Figure 6. Reduced χ^2 values for retrievals with each model. The top panel shows the reduced χ^2 values for the retrievals on WFC3 data alone, as well as for the joint WFC3 + STIS data. In the joint case, the reduced χ^2 is shown for the WFC3 and STIS sections separately. The bottom panel shows the equivalent for the NIRISS and NIRISS + STIS retrievals.

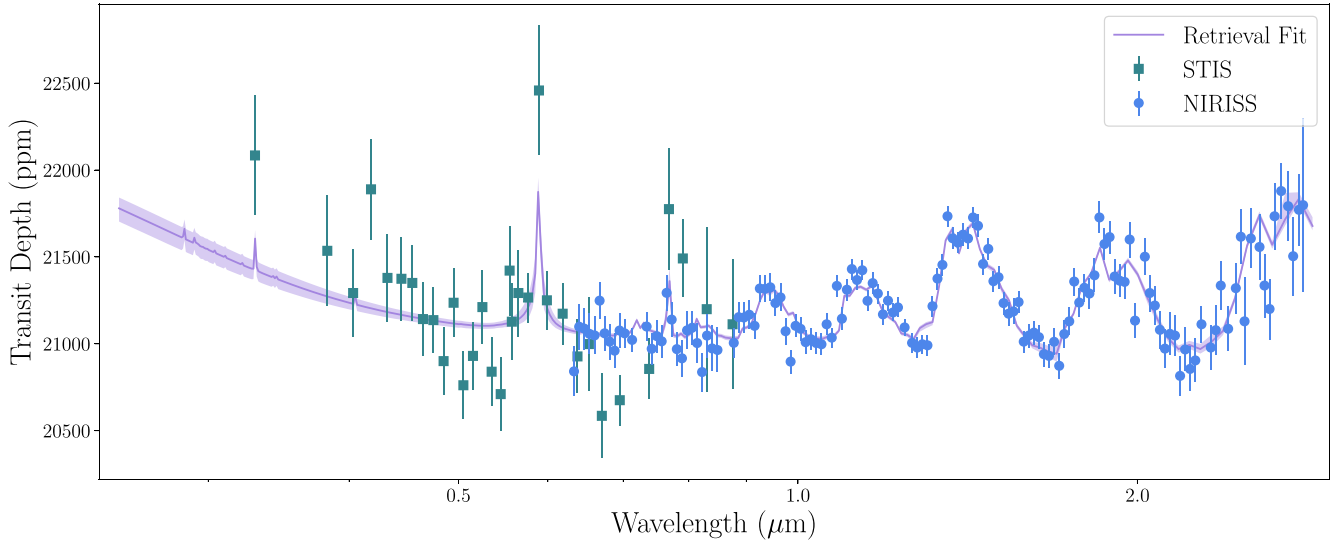


Figure 7. Retrieval fit for NIRISS + STIS using the non-isothermal, non-grey cloud model. The purple line and shaded region show the median and 1σ transit depth values.

3.3.2 Shifting STIS

It is a known issue that spectra from different observing instruments can be offset from one another (Yip et al. 2021). To test this issue, we add a parameter to all our retrievals that shifts the STIS data up or down with respect to the WFC3 or NIRISS data. It turns out that allowing for a shift has little-to-no effect on the retrievals, and in all cases, the STIS shift parameter is consistent with zero (see Fig. A3). This suggests that the issue with fitting the STIS data is not caused by only an offset between the data sets.

3.3.3 Removing sodium

Our combined retrievals demonstrate that the addition of STIS data has a substantial effect on the constraints for water abundance. However, the source of the constraints in retrievals is not always clear and requires further analysis (e.g. Welbanks et al. 2023). The STIS data encompasses a strong and very broad sodium line, as well as an optical slope from clouds/hazes or Rayleigh scattering. In order to test which of these properties is affecting the constraint on water abundance, we perform a set of retrievals on the combined data sets but remove the STIS data points around the sodium feature.

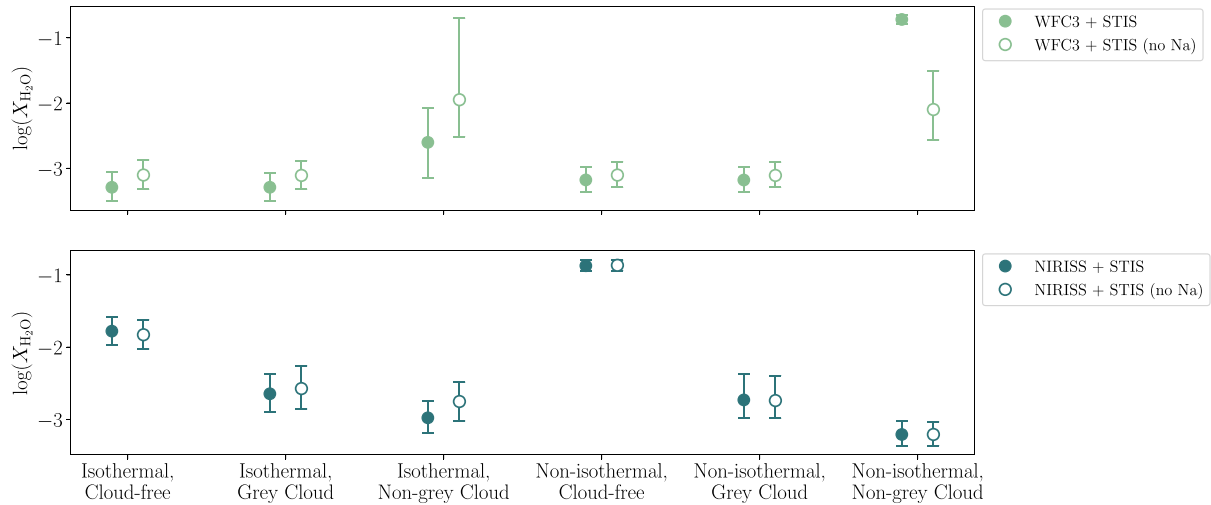


Figure 8. Retrieved water abundances for the full data (filled circles) and for the data with the sodium line removed (empty circles). In the latter case, the sodium abundance and opacity is not included in the retrieval.

We also remove the sodium opacity from the model, and the sodium abundance is no longer a free parameter in the retrieval. The results for the retrieved water abundances are shown in Fig. 8. In the WFC3 cases, the water abundance remains consistent when sodium is removed, with the exception of the non-grey cloud retrievals. Since these models have more freedom to fit slopes in the optical, it seems that the sodium feature is important for constraining the slope and thus the spectral continuum. However, in the NIRISS cases, no such effect appears. This is likely due to the larger number of points covering the continuum in the NIRISS data set, meaning the effect of removing sodium is minimized. Overall, this demonstrates that the optical slope is driving the water abundance constraint when STIS data are included, but that flexible cloud models benefit from the additional information from the sodium feature.

3.3.4 Alkali detections

The ERS study of NIRISS data found evidence for K, but no evidence for Na due to the sodium doublet lying outside the NIRISS wavelength range (Feinstein et al. 2023). Theoretically, the slope of the Na resonance line wings at the blue end of NIRISS could provide some constraints, but it may also be degenerate with other atmospheric properties, such as clouds (e.g. Radica et al. 2023; Taylor et al. 2023). The ability to compute refractory-to-volatile ratios could be key for placing constraints on the formation history of planet (Lothringer et al. 2021; Schneider & Bitsch 2021; Chachan et al. 2023), making detections of these alkalis particularly interesting. Fig. 9 shows the retrieval results for sodium and potassium for the different data set combinations. In agreement with the ERS study, we find that, in order to detect sodium, the retrieval requires the STIS wavelengths to cover the sodium doublet. The simplest model, isothermal + cloud-free, does show a detection for the NIRISS-only case; however, the Bayesian evidence strongly ruled out this model (Fig. 3). The retrieved Na abundance appears fairly consistent across the NIRISS + STIS retrievals, obtaining values of $\sim 10^{-6}$. For WFC3 + STIS, the abundance of Na is fairly consistent across models, with the exception of the non-isothermal + non-grey cloud case, issues of which are explained in Section 3.3.1. The abundance is generally slightly lower than the NIRISS + STIS retrievals, probably due to the higher continuum fitting the WFC3 data. When comparing the

detection significances for Na, as mentioned above, we obtain 3.8σ for WFC3 + STIS and 2.3σ for NIRISS + STIS. However, note that this is only for the non-isothermal + non-grey cloud retrieval, which obtains substantially higher alkali abundances than other retrievals on WFC3 + STIS.

For potassium, we find the expected result that the NIRISS data alone is able to detect it, though the abundance depends on the model, with values ranging from $\sim 10^{-6}$ to $\sim 10^{-8}$. The results are mostly consistent when the STIS data are added, showing that the necessary information is already encompassed in the NIRISS data. For WFC3, however, potassium is unsurprisingly not detected, and only when adding STIS and using certain models does one retrieve a constrained abundance for K. When it is constrained, the abundance appears to be consistent with the NIRISS retrievals, again with the exception of the non-isothermal + non-grey cloud case. Thus, NIRISS alone can be a powerful tool for diagnosing the content in exoplanetary atmospheres in order to constrain formation histories. Alternatively, one could combine space-based near-infrared data with ground-based optical data in order to constrain the alkali abundances. For the detection significances for K, we obtain 2.6σ for WFC3 + STIS, and 2.1σ for NIRISS. As discussed previously, for NIRISS + STIS, the Bayesian evidence favours the retrieval without K, likely due to the overlapping but inconsistent K feature in the two data sets. This again provides evidence that blindly combining the two data sets could be detrimental to the retrieval results.

3.4 Cutting NIRISS

In this section, we want to address the question of whether the biggest gain in information from NIRISS over WFC3 comes from the increased resolution and precision, or the extended wavelength coverage. In order to test this, we perform the retrievals on the NIRISS data cut to the wavelength range of the WFC3 data. Fig. 10 shows the summary of the results.

For the water abundance, the NIRISS-cut retrievals obtain substantially lower values than either the WFC3 or full NIRISS retrievals. This is likely due to the lower relative NIRISS points in the $1.4\ \mu\text{m}$ water feature, possibly due to different instrument systematics, and the lack of wavelength coverage required to provide the continuum information. As shown in previous studies, and discussed in Sec-

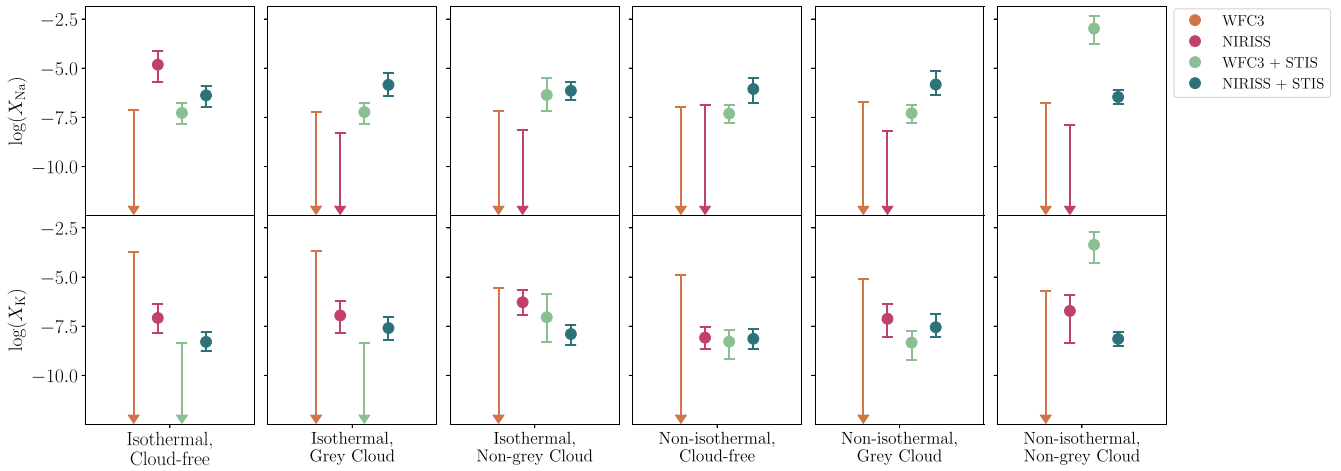


Figure 9. Summary of the retrieval results for the sodium and potassium abundances for all the models and all the data set combinations. The downward arrows indicate upper limits only.

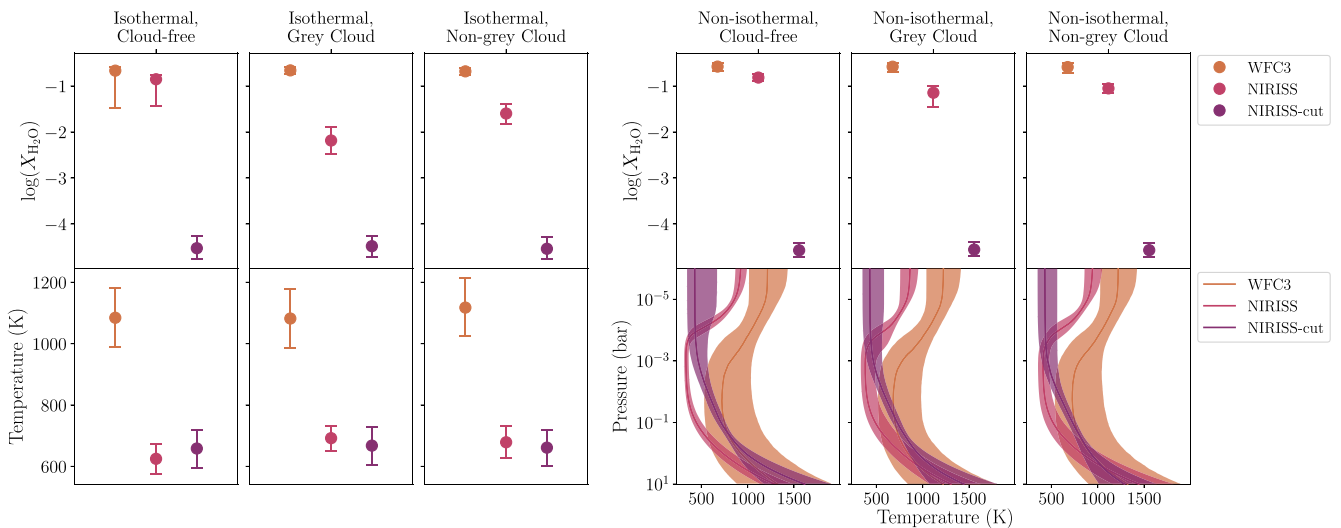


Figure 10. Summary of the retrieval results for all the models for WFC3, NIRISS, and NIRISS cut to the wavelength range of WFC3.

tion 3.2, the water abundance cannot be constrained in the limited wavelength range of WFC3, except in the cases of very high or very low abundance (Line & Parmentier 2016; Welbanks et al. 2019). The WFC3 retrieval represents the very high abundance case, where the mean molecular weight is affected, whereas the NIRISS-cut retrieval represents the other extreme, where low water abundance reveals the continuum from CIA at the troughs of the water feature. In reality, the water abundance is likely in between the two extremes, but this can only be constrained when additional features are encompassed by a wider wavelength range (i.e. the full NIRISS case). This result has serious implications for previous constraints on abundances from WFC3 data, as they are unlikely to be robust and are potentially very sensitive to the instrument systematics. It also demonstrates how, when the data are scarce, retrievals can struggle to provide good estimates for the errors on their constraints.

For the isothermal cases, the temperature for the NIRISS-cut retrieval is consistent with the full NIRISS retrieval. This suggests that the temperature is governed by the resolution and precision of the

data. For the non-isothermal cases, the picture is more complicated. At the base of the atmosphere, all three profiles agree fairly well, though with quite loose constraints. In the middle of the atmosphere, where the transmission spectrum is probing, the NIRISS-cut profile transitions from agreeing more closely with the WFC3 profile, at pressures around 0.1 bar, to the NIRISS profile, at pressures around $10^{-3} - 10^{-4}$ bar. At the top of the atmosphere, the NIRISS-cut profile is substantially colder than either of the others. Interestingly, the NIRISS-cut profiles show no evidence for an inversion, despite this appearing in all the other profiles. These results suggest that in the non-isothermal case, a combination of wavelength coverage and resolution contribute to the temperature profile's constraints. However, as noted before, we do not expect there to be sufficient information in transmission spectra to accurately constrain all levels of the temperature profile for the planet.

An additional interesting comparison comes from the Bayesian evidences. In the NIRISS-cut retrievals, the models all obtain a Bayes factor of 0.6 or less, giving inconclusive evidence for favouring any of them. This is comparable to the Bayesian evidence from the WFC3

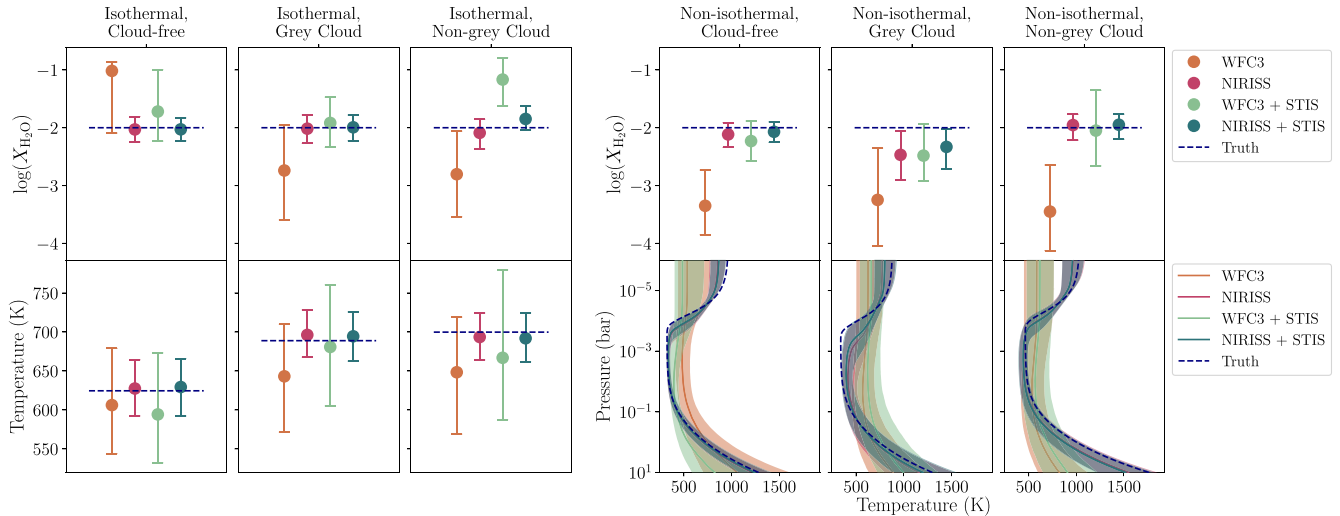


Figure 11. Mock retrieval results for the water abundance and temperature. In this case, the same model was used to generate the mock spectrum and perform the retrieval. The input parameters were set to the best-fitting values from the real retrievals on the NIRISS data, and the water abundance was fixed to 10^{-2} . The true values are shown by the navy dashed lines. The error bars on the mock spectra are assumed to be the same as the real data for WASP-39 b, but no noise was added to the data points. This idealistic case allows us to assess the retrieval’s performance in the absolute best-case scenario.

case (upper-left panel of Fig. 3), suggesting that the wavelength coverage is key for distinguishing between different models.

3.5 Mock retrievals

In order to test the conclusions from Sections 3.2, 3.3, and 3.4, we perform mock retrievals on simulated data. This removes the differences seen in the real data potentially due to systematics. For each of the six models (isothermal/non-isothermal + cloudfree/grey cloud/non-grey cloud), we simulate the WFC3, NIRISS, and STIS data, assuming the system parameters of WASP-39 b and setting the model parameters to the best-fitting values from the retrievals on the real NIRISS data. This is with the exception of the water abundance, which is fixed to the same value across all models. This allows us to test each model’s ability to retrieve the water abundance accurately in a consistent way. The error bars are assumed to be the same as for the real WASP-39 b data, though no noise was added to the simulated points. This idealistic case allows us to focus on differences due to only the wavelength coverage and resolution/precision, without being affected by a random noise draw.

3.5.1 Full data

Fig. 11 shows the results for the WFC3, NIRISS, WFC3 + STIS, and NIRISS + STIS mock retrievals for a fixed water abundance of 10^{-2} . For both water abundance and temperature, the NIRISS and NIRISS + STIS retrievals are able to accurately retrieve the input values across all models. However, for WFC3 and WFC3 + STIS, the water abundance is not always correctly retrieved. In fact, for a number of models, the 1σ constraints from the WFC3 retrievals significantly underestimate the water abundance. Adding the STIS data helps correct this in most cases. Fig. 12 shows the retrieval fits to the mock WFC3 data for the retrievals on WFC3 and WFC3 + STIS for the non-isothermal + non-grey cloud case. The water features from both fits clearly lie within the WFC3 error bars at $1.4\ \mu\text{m}$, despite corresponding to an abundance difference of 2–3 orders of magnitude. Fig. A4 shows the full cornerplot for these

retrievals. The true abundances are consistent with the posteriors for WFC3 (as expected due to the lack of noise), but lie in the tail of the distribution for the water abundance. It appears the degeneracy with the cloud top pressure is what leads to the lower water abundances, and only extreme abundances at the high and low end are ruled out [as predicted by Line & Parmentier (2016)]. This experiment also demonstrates how the addition of the STIS wavelengths can improve your retrieval constraints. However, it is the noise from the real STIS data, due to systematics or stellar activity, that can cause problematic biases, as shown in Section 3.3.

These mock retrieval tests show how the WFC3 retrieval adjusts differently to fit the data and further emphasizes the need for extended wavelength coverage in order to accurately constrain molecular abundances. In the isothermal cases, the temperature is accurately retrieved, though with a looser constraint than for the NIRISS retrievals. In the non-isothermal cases, the profiles appear to overlap with the true profile in the areas probed by transmission spectra (i.e. $10^{-1} - 10^{-4}$ bar). However, there are discrepancies in the 1σ constraints from the WFC3 and WFC3 + STIS retrievals, likely due to the temperature compensating for degenerate water abundances by adjusting the scale height.

3.5.2 Cut NIRISS

In addition to the mock retrievals on the full data ranges, we also perform a mock retrieval on the NIRISS data cut to the same wavelength range as WFC3, as in Section 3.4. Again, the water abundance for the simulated spectrum is fixed to 10^{-2} , and the other parameters are set to the best-fitting values from the retrievals on the real full NIRISS data. Fig. 13 shows the comparison of the retrieved water abundance and temperature for the WFC3 and NIRISS mock retrievals (same as Fig. 11) and the NIRISS-cut mock retrieval. In the isothermal cases, the water abundance from the NIRISS-cut retrievals agrees with the true value and has a similar constraint to the full NIRISS retrieval. Though the WFC3 retrieval also agrees with the true value here, the constraints are, as expected, much wider due to the larger error bars on the data. However, in the non-isothermal

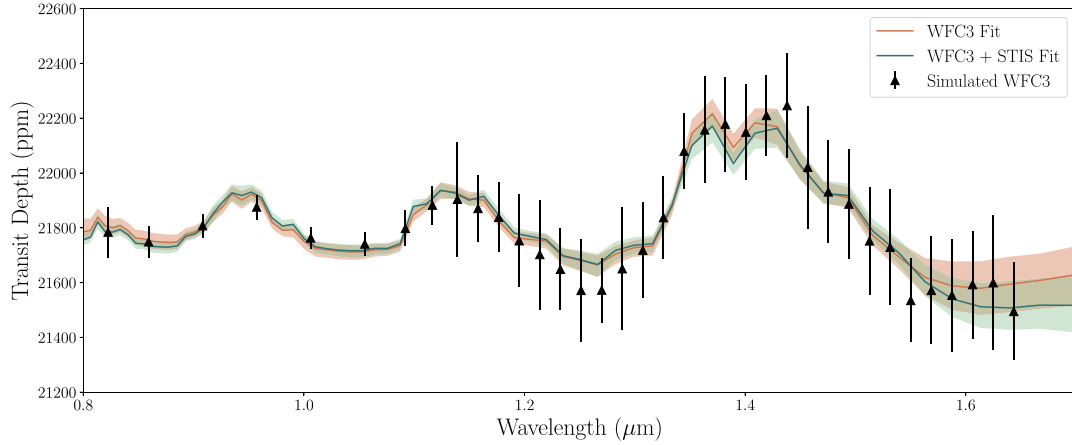


Figure 12. Spectral fits to the WFC3 data for the mock retrievals of the simulated data, assuming a water abundance of 10^{-2} , using the non-isothermal + non-grey cloud model. The pink line and shaded region show the fit and 1σ error from the retrieval on the simulated WFC3 data alone, respectively. The red line and shaded region show the fit and 1σ error from the retrieval on the simulated WFC3 and STIS data combined, respectively.

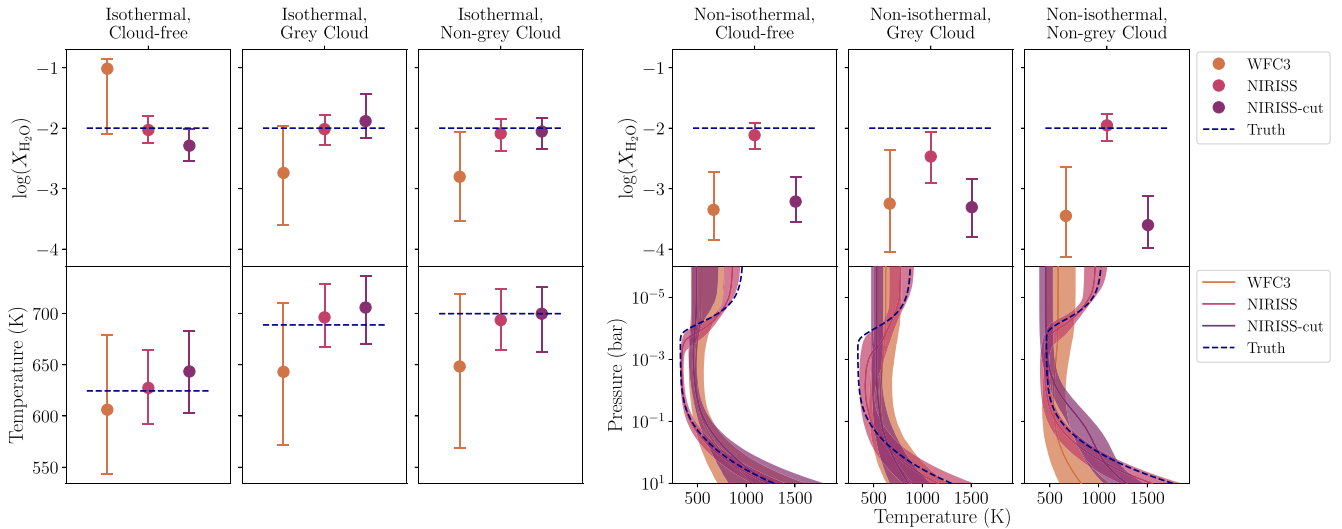


Figure 13. Mock retrieval results for the water abundance and temperature, for the WFC3 and NIRISS-simulated data (same as Fig. 11), but now including the cut-NIRISS simulated data case as well.

cases, the NIRISS-cut results suffer from the same problem as the WFC3 retrievals, deviating from the true value. Again, the correct water abundance lies in the very edge of the tail of the posterior distribution (not shown). These results suggest that a degeneracy with the temperature profile can also drive the water abundance to different values, even with the higher resolution and smaller error bars associated with NIRISS spectra. As before, the presence of clouds exacerbates the issue. These results provide further evidence in favour of extended wavelength coverage for accurately determining molecular abundances.

4 DISCUSSION

The start of the JWST era brings major gains in both resolution and wavelength coverage for exoplanet spectra. This yields a substantial increase in the confidence of our analysis of these atmospheres. It also enables us to probe new molecular features, inaccessible from previous instruments. This helps to build a much more complete picture of these distant planets. However, this increase in detail and

precision brings its own issues. The various assumptions made by standard retrieval models may no longer be sufficient to fit these data (e.g. Feng et al. 2016; MacDonald et al. 2020; Taylor et al. 2020; Yang, Irwin & Barstow 2023), and care must be taken when drawing conclusions from retrieval results. To help mitigate these issues, standard practice has developed to include multiple data reductions and retrievals, using different pipelines and forwards models with distinct origins (e.g. Grant et al. 2023; Moran et al. 2023; Welbanks et al. 2024; Yang et al. 2024). A thorough exploration of the different approaches to data reduction and forward modelling assumptions in retrieval models can help us fully explore the information content of the JWST era spectra.

With an abundance of archival exoplanet data and the current operation of HST/STIS and HST/UVIS that access wavelengths not seen by JWST, it is logical to want to combine these data together in order to expand our wavelength coverage and break degeneracies in retrievals. Our mock retrieval results show that, even in the absence of noise, WFC3 data alone is not always able to retrieve the correct water abundance, and the inclusion of STIS wavelengths can help.

However, when the real data are used, effects from the instrument's noise can lead to biases in the retrievals when adding STIS data. This highlights some of the issues that can arise when combining data sets from different instruments, especially given they are reduced using different methods with different assumptions. Furthermore, potential variability in the planet can also cause discrepancies in the measurements if they are taken at different times (e.g. Meier Valdés et al. 2023), leading to issues when combining different epochs of data. It is common to perform retrieval studies, especially population studies, by extracting spectra from the literature instead of homogeneously reducing the data. We have shown that the retrieval results can be biased by one instrument, even when the model lacks the complexity to fit that instrument's data accurately.

A recent study by Lueber et al. (2024) carried out a detailed analysis of the ERS data of WASP-39b from the different instruments. They found that flexible non-grey cloud models were able to fit spectral features with insufficient wavelength coverage, thus impacting the retrieval results significantly. In the current study, when STIS data are included, the non-grey cloud adjusts to fit the steep slope, affecting the retrieved abundances. However, it's possible that the slope has an alternative source, such as stellar contamination or instrument systematics, and the fit from the non-grey cloud model could be inaccurate.

As part of the ERS papers, Welbanks et al. (in preparation) performed a model synthesis study comparing retrieval results from different codes on the full 0.6–12 micron JWST spectra of WASP-39 b. They found a metallicity of 10–50× solar, depending on modelling assumptions. This is slightly lower than our results, which, for the NIRISS non-isothermal + non-grey cloud retrieval, for example, corresponds to a metallicity of $\sim 200\text{--}350\times$ solar. The wider spectral data likely provide additional constraints on the water vapour abundance beyond that available from NIRISS. Welbanks et al. (in preparation) note that determining the water abundance of this planet is uniquely challenging, since it occupies a region of parameter space where relatively large changes in the water abundance have a very subtle effect on the spectrum. This is due to the change in mean molecular weight, and therefore scale height, compensating for a change in the absorption due to water.

5 CONCLUSION

In this work, we perform atmospheric retrievals on the JWST/NIRISS and HST/WFC3 spectra of WASP-39 b to quantify the difference in information content across the two instruments (Section 3.2). We find that even the water abundance, the only molecule confidently probed by the WFC3 wavelength range, can be biased to spurious values due to the lack of continuum information in the short WFC3 window. The NIRISS data provide this essential information, as well as additional absorbers, such as K, and clouds. Furthermore, our analysis reveals a substantial gain in the confidence of the water detection for NIRISS over WFC3 (21.3σ versus 9.0σ).

We also combine these spectra with STIS data to expand our comparison (Section 3.3). These additional data affect the retrieved water abundance and hold further information on clouds and alkali absorption. However, the STIS data also have the power to bias the retrieval results, for NIRISS, for example, despite the poor fit to its spectrum. In order to determine the source of the effect on the constraints from STIS, we tried removing the sodium feature from the data. This generally did not affect the retrieved water abundance, suggesting that the optical slope from STIS is driving the constraint, as opposed to the sodium line wings. Adding the STIS data to NIRISS also weakens the detection significance of

CO₂, as well as removing the detection of K altogether, which shows further evidence for biasing. Additionally, we test if including a shift parameter applied to the STIS data can mitigate some of these issues in the retrievals (Section 3.3.2). However, we find this has little effect, and the shift parameter is consistent with zero, indicating that issues with combining these data sets does not stem from offsets alone. Based on our results, we recommend exercising caution when combining future NIRISS and archival STIS exoplanet spectra.

The ability to measure refractory-to-volatile ratios could be key for constraining planet formation models. We consider the detectability of alkali species (Section 3.3.4) from both the NIRISS and STIS spectra and find that, although Na relies on the bluer wavelengths covered by STIS, NIRISS has the ability to constrain K. The opportunity to measure the abundances of K and H₂O could motivate a survey of exoplanets using NIRISS to provide a statistical study of K/O ratios and thus planet formation histories.

We test our retrievals on the NIRISS spectra cut to the wavelength range covered by WFC3 to determine whether each model parameter is affected by the wavelength coverage or the resolution (Section 3.4). We find that the water abundance is strongly affected by the wavelength range, as the inclusion of wider wavelengths is essential for setting the continuum. In this limited wavelength range, the water abundance is extremely sensitive to instrument systematics, which has serious implications for previous constraints from WFC3 data. In the isothermal case, the temperature is driven solely by the resolution of the data, with little information gained from wider wavelengths. In the non-isothermal case, the picture is less clear, though these models may be overly complex.

Finally, we perform mock retrievals on simulated WFC3, NIRISS, and STIS data. Our results show that, even with no noise added to the data points, the retrieval struggles to accurately constrain the water abundance from WFC3 data alone, further emphasizing the need for the wider wavelength coverage provided by the other data sets. These results also provide more robust evidence of our earlier conclusions that the WFC3 retrievals can obtain incorrect results, as only in the mock retrieval case do we know the ground truth.

Our results demonstrate the vast information gained by the advent of JWST, improving our knowledge of the conditions of exoplanet atmospheres. It remains an exciting period for exoplanet characterization, and the development of more accurate and reliable analysis techniques must be a priority in order to maximize what we can learn from these data.

ACKNOWLEDGEMENTS

We would like to thank Adina Feinstein for useful comments on the manuscript.

CF acknowledges financial support from the Swiss National Science Foundation (SNSF) Mobility Fellowship under grant no. P500PT_203110. CF and JLB acknowledge financial support from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme under grant agreement no. 805445. JT acknowledges the support provided by Schmidt Sciences, LLC. MR acknowledges financial support from the Natural Sciences and Engineering Research Council of Canada and the Fonds de recherche du Québec—Nature et technologies. JKB is supported by a Science and Technology Facilities Council Ernest Rutherford Fellowship (ST/T004479/1). GM acknowledges funding from the Severo Ochoa grant CEX2021-001131-S and the Ramón y Cajal grant RYC2022-037854-I funded by MCIN/AEI/10.13039/501100011033 and FSE +.

Software: ASTROPY (Astropy Collaboration et al. 2013, 2018, 2022), NUMPY (Harris et al. 2020), SCIPY (Virtanen et al. 2020), MATPLOTLIB (Hunter 2007), PYMULTINEST (Buchner et al. 2014), CORNER (Foreman-Mackey 2016), and SEABORN (Waskom 2021).

DATA AVAILABILITY

The data used in this study are publicly available and published in Sing et al. (2016), Wakeford et al. (2018), and Feinstein et al. (2023). The open-source retrieval codes BeAR and CHIMERA can be found at <https://github.com/newstrangeworlds/bear> and <https://github.com.mrline/CHIMERA>, respectively.

REFERENCES

- Abel M., Frommhold L., Li X., Hunt K. L. C., 2011, *J. Phys. Chem. A*, 115, 6805
- Abel M., Frommhold L., Li X., Hunt K. L. C., 2012, *J. Chem. Phys.*, 136, 044319
- Ackerman A. S., Marley M. S., 2001, *ApJ*, 556, 872
- Ahrer E.-M. et al., 2023, *Nature*, 614, 653
- Albert L. et al., 2023, *PASP*, 135, 075001
- Alderson L. et al., 2023, *Nature*, 614, 664
- Allard N. F., Spiegelman F., Kielkopf J. F., 2016, *A&A*, 589, A21
- Allard N. F., Spiegelman F., Leininger T., Molliere P., 2019, *A&A*, 628, A120
- Asplund M., Grevesse N., Sauval A. J., Scott P., 2009, *ARA&A*, 47, 481
- Astropy Collaboration et al., 2013, *A&A*, 558, A33
- Astropy Collaboration et al., 2018, *AJ*, 156, 123
- Astropy Collaboration et al., 2022, *ApJ*, 935, 167
- Barstow J. K., Aigrain S., Irwin P. G. J., Sing D. K., 2017, *ApJ*, 834, 50
- Bean J. L. et al., 2018, *PASP*, 130, 114402
- Bell T. J. et al., 2023, *Nature*, 623, 709
- Benneke B., Seager S., 2012, *ApJ*, 753, 100
- Benneke B., Seager S., 2013, *ApJ*, 778, 153
- Benneke B. et al., 2024, preprint
- Berta Z. K. et al., 2012, *ApJ*, 747, 35
- Blecic J., Dobbs-Dixon I., Greene T., 2017, *ApJ*, 848, 127
- Bohnen C. F., Huffman D. R., 1998, *Absorption and Scattering of Light by Small Particles*. WILEY-VCH, Weinheim
- Buchner J. et al., 2014, *A&A*, 564, A125
- Cadieux C. et al., 2024, *ApJ*, 970, L2
- Carter A. L. et al., 2024, *Nature Astron.*, 8, 1008
- Chachan Y., Knutson H. A., Lothringer J., Blake G. A., 2023, *ApJ*, 943, 112
- Changeat Q. et al., 2024, *ApJS*, 270, 34
- Constantinou S., Madhusudhan N., Gandhi S., 2023, *The ApJ*, 943, L10
- Darveau-Bernier A. et al., 2022, *PASP*, 134, 094502
- Deming D. et al., 2013, *ApJ*, 774, 95
- Doyon R. et al., 2023, *PASP*, 135, 098001
- Espinoza N., Kossakowski D., Brahm R., 2019, *MNRAS*, 490, 2262
- Faedi F. et al., 2011, *A&A*, 531, A40
- Fairman C., Wakeford H. R., MacDonald R. J., 2024, *AJ*, 167, 240
- Feinstein A. D. et al., 2023, *Nature*, 614, 670
- Feng Y. K., Line M. R., Fortney J. J., Stevenson K. B., Bean J., Kreidberg L., Parmentier V., 2016, *ApJ*, 829, 52
- Feroz F., Hobson M. P., 2008, *MNRAS*, 384, 449
- Feroz F., Hobson M. P., Bridges M., 2009, *MNRAS*, 398, 1601
- Fischer P. D. et al., 2016, *ApJ*, 827, 19
- Fisher C., Heng K., 2018, *MNRAS*, 481, 4698
- Foreman-Mackey D., 2016, *J. Open Source Softw.*, 1, 24
- Fournier-Tondreau M. et al., 2024, *MNRAS*, 528, 3354
- Freedman R. S., Lustig-Yaeger J., Fortney J. J., Lupu R. E., Marley M. S., Lodders K., 2014, *ApJS*, 214, 25
- Fu G. et al., 2022, *ApJ*, 940, L35
- Gao P. et al., 2020, *Nature Astron.*, 4, 951
- Gharib-Nezhad E., Iyer A. R., Line M. R., Freedman R. S., Marley M. S., Batalha N. E., 2021, *ApJS*, 254, 34
- Gibson N. P. et al., 2012, *MNRAS*, 422, 753
- Grant D. et al., 2023, *ApJ*, 956, L32
- Griffith C. A., 2014, *Phil. Trans. R. Soc.*, 372, 20130086
- Grimm S. L., Heng K., 2015, *ApJ*, 808, 182
- Grimm S. L. et al., 2021, *ApJS*, 253, 30
- Guillot T., 2010, *A&A*, 520, A27
- Harris C. R. et al., 2020, *Nature*, 585, 357
- Hunter J. D., 2007, *Comput. Sci. Engineer.*, 9, 90
- JWST Transiting Exoplanet Community Early Release Science Team et al., 2023, *Nature*, 614, 649
- Kawashima Y., Ikoma M., 2019, *ApJ*, 877, 109
- Kempton E. M. R. et al., 2023, *Nature*, 620, 67
- Kirk J., López-Morales M., Wheatley P. J., Weaver I. C., Skillen I., Louden T., McCormac J., Espinoza N., 2019, *AJ*, 158, 144
- Kirk J. et al., 2024, *AJ*, 167, 90
- Kitzmann D., Heng K., 2018, *MNRAS*, 475, 94
- Kitzmann D., Heng K., Oreshenko M., Grimm S. L., Apai D., Bowler B. P., Burgasser A. J., Marley M. S., 2020, *ApJ*, 890, 174
- Kitzmann D., Stock J. W., Patzer A. B. C., 2024, *MNRAS*, 527, 7263
- Kramida A., Ralchenko Yu., Reader J., NIST ASD Team, 2028, NIST Atomic Spectra Database, Version 5.6.1 (accessed 2019 February 6), available at: <https://physics.nist.gov/asd>
- Kreidberg L. et al., 2014a, *Nature*, 505, 69
- Kreidberg L. et al., 2014b, *ApJ*, 793, L27
- Kurucz R., Bell B., 1995, Robert Kurucz CD-ROM No.23
- Lacis A. A., Oinas V., 1991, *J. Geophys. Res.*, 96, 9027
- Lecavelier Des Etangs A., Pont F., Vidal-Madjar A., Sing D., 2008, *A&A*, 481, L83
- Li G., Gordon I. E., Rothman L. S., Tan Y., Hu S.-M., Kassi S., Campargue A., Medvedev E. S., 2015, *ApJS*, 216, 15
- Lim O. et al., 2023, *ApJ*, 955, L22
- Line M. R., Parmentier V., 2016, *ApJ*, 820, 78
- Line M. R. et al., 2013, *ApJ*, 775, 137
- Lothringer J. D., Rustamkulov Z., Sing D. K., Gibson N. P., Wilson J., Schlafman K. C., 2021, *ApJ*, 914, 12
- Lueber A., Kitzmann D., Bowler B. P., Burgasser A. J., Heng K., 2022, *ApJ*, 930, 136
- Lueber A., Novais A., Fisher C., Heng K., 2024, *A&A*, 687, A110
- Lustig-Yaeger J. et al., 2023, *Nature Astron.*, 7, 1317
- MacDonald R. J., Batalha N. E., 2023, *Res. Notes Am. Astron. Soc.*, 7, 54
- MacDonald R. J., Madhusudhan N., 2017, *The ApJ*, 850, L15
- MacDonald R. J., Goyal J. M., Lewis N. K., 2020, *ApJ*, 893, L43
- Madhusudhan N., Seager S., 2009, *ApJ*, 707, 24
- Madhusudhan N., Sarkar S., Constantinou S., Holmberg M., Piette A. A. A., Moses J. I., 2023, *ApJ*, 956, L13
- Meier Valdés E. A. et al., 2023, *A&A*, 677, A112
- Min M., Ormel C. W., Chubb K., Helling C., Kawashima Y., 2020, *A&A*, 642, A28
- Moran S. E. et al., 2023, *ApJ*, 948, L11
- Morello G., Changeat Q., Dyrek A., Lagage P. O., Tan J. C., 2023, *A&A*, 676, A54
- Mugnai L. V., Swain M. R., Estrela R., Roudier G. M., 2024, *MNRAS*, 531, 35
- Nikolov N., Sing D. K., Gibson N. P., Fortney J. J., Evans T. M., Barstow J. K., Kataria T., Wilson P. A., 2016, *ApJ*, 832, 191
- Ohno K., Kawashima Y., 2020, *ApJ*, 895, L47
- Parmentier V., Guillot T., 2014, *A&A*, 562, A133
- Parviainen H., 2023, *A&A*, 671, L3
- Pinhas A., Madhusudhan N., Gandhi S., MacDonald R., 2019, *MNRAS*, 482, 1485
- Polyansky O. L., Kyuberis A. A., Zobov N. F., Tennyson J., Yurchenko S. N., Lodi L., 2018, *MNRAS*, 480, 2597
- Radica M., 2024, *JOSS*, 9, 6898
- Radica M. et al., 2022, *PASP*, 134, 104502
- Radica M. et al., 2023, *MNRAS*, 524, 835
- Radica M. et al., 2024, *ApJ*, 962, L20
- Richard C. et al., 2012, *J. Quant. Spectrosc. Radiat. Transfer*, 113, 1276

- Rocchetto M., Waldmann I. P., Venot O., Lagage P. O., Tinetti G., 2016, *ApJ*, 833, 120
- Rosich A., Herrero E., Mallonn M., Ribas I., Morales J. C., Perger M., Anglada-Escudé G., Granzer T., 2020, *A&A*, 641, A82
- Rothman L. S. et al., 2010, *J. Quant. Spectrosc. Radiat. Transfer*, 111, 2139
- Rustamkulov Z. et al., 2023, *Nature*, 614, 659
- Schlawin E. et al., 2016, *PASP*, 129, 015001
- Schneider A. D., Bitsch B., 2021, *A&A*, 654, A72
- Sing D. K. et al., 2015, *MNRAS*, 446, 2428
- Sing D. K. et al., 2016, *Nature*, 529, 59
- Somogyi W., Yurchenko S. N., 2021, International Symposium on Molecular Spectroscopy, Talk TF03
- Stevenson K. B. et al., 2016, *PASP*, 128, 094401
- Stock J. W., Kitzmann D., Patzer A. B. C., Sedlmayr E., 2018, *MNRAS*, 479, 865
- Stock J. W., Kitzmann D., Patzer A. B. C., 2022, *MNRAS*, 517, 4070
- Taylor J., Parmentier V., Irwin P. G. J., Aigrain S., Lee E. K. H., Krissansen-Totton J., 2020, *MNRAS*, 493, 4342
- Taylor J. et al., 2023, *MNRAS*, 524, 817
- Trotta R., 2008, *Contemp. Phys.*, 49, 71
- Tsai S.-M. et al., 2023, *Nature*, 617, 483
- Tsiaras A. et al., 2018, *AJ*, 155, 156
- Vardya M. S., 1962, *ApJ*, 135, 303
- Virtanen P. et al., 2020, *Nature Methods*, 17, 261
- Wakeford H. R. et al., 2013, *MNRAS*, 435, 3481
- Wakeford H. R. et al., 2018, *AJ*, 155, 29
- Waskom M. L., 2021, *J. Open Source Softw.*, 6, 3021
- Welbanks L., Madhusudhan N., Allard N. F., Hubeny I., Spiegelman F., Leininger T., 2019, *ApJ*, 887, L20
- Welbanks L., McGill P., Line M., Madhusudhan N., 2023, *AJ*, 165, 112
- Welbanks L. et al., 2024, *Nature*, 630, 836
- Yang J., Irwin P. G. J., Barstow J. K., 2023, *MNRAS*, 525, 5146
- Yang J. et al., 2024, *MNRAS*, 532, 460
- Yip K. H., Changeat Q., Edwards B., Morvan M., Chubb K. L., Tsiaras A., Waldmann I. P., Tinetti G., 2021, *AJ*, 161, 4
- Yurchenko S. N., Tennyson J., 2014, *MNRAS*, 440, 1649
- Yurchenko S. N., Tennyson J., Barber R. J., Thiel W., 2013, *J. Mol. Spectrosc.*, 291, 69

APPENDIX A: ADDITIONAL PLOTS

Fig. A1 shows the opacities of all the species considered in this study over the wavelength range covered by NIRISS, WFC3, and STIS. Some clear features can be matched to the data in Fig. 1 by-eye, such as the sodium and potassium lines, as well as the water absorption features.

Fig. A2 shows the retrieval results for the non-isothermal + non-grey cloud model applied to the combination of WFC3 + STIS data and the NIRISS + STIS data. The addition of STIS to WFC3 is causing some of the cloud parameters to be forced to extreme values, hitting the edges of their priors.

Fig. A3 shows the retrieved median values and 1σ error bars of the shift parameter applied to the STIS data for all the different models when combined with the WFC3 and NIRISS data. Adding this parameter does not affect the retrieval results, and the consistency with zero demonstrates that there is no evidence for an absolute shift between the data sets.

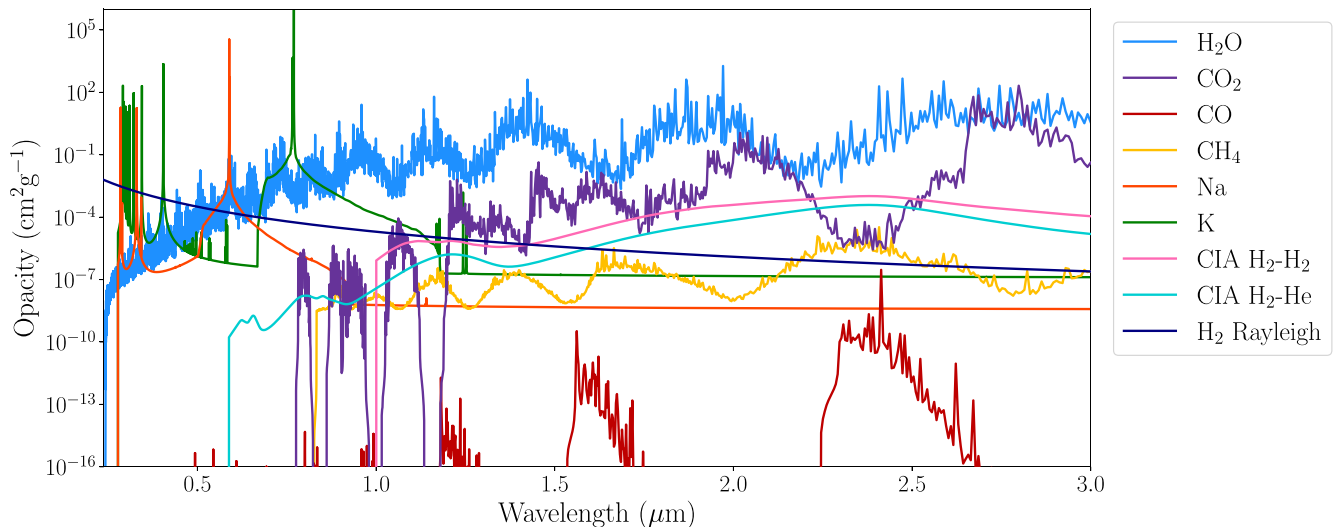


Figure A1. Opacities as a function of wavelength for the opacity sources in the current study, computed at a temperature of 1000 K and a pressure of 0.01 bar. The opacities are weighted by their mass mixing ratios corresponding to the best-fitting model from the non-isothermal + non-grey cloud retrieval on the NIRISS data. They are shown over the wavelength range covered by the three instruments in this study – NIRISS, WFC3, and STIS.

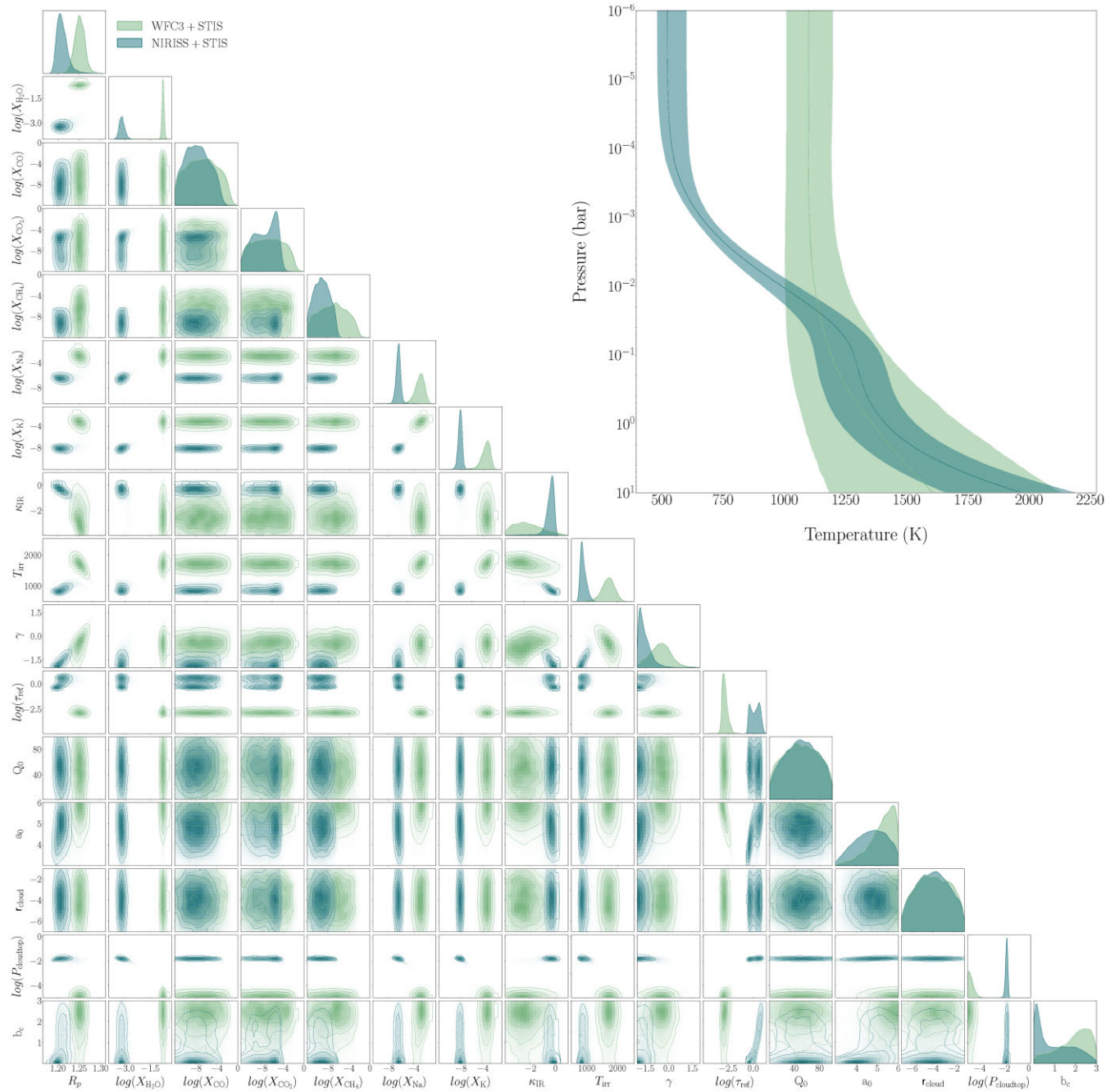


Figure A2. Retrievals using the non-isothermal + non-grey cloud model for the WFC3 + STIS (light green) and the NIRISS + STIS (dark green) data. The median and 1σ limits of the retrieved temperature-pressure profiles are shown in the top right corner.

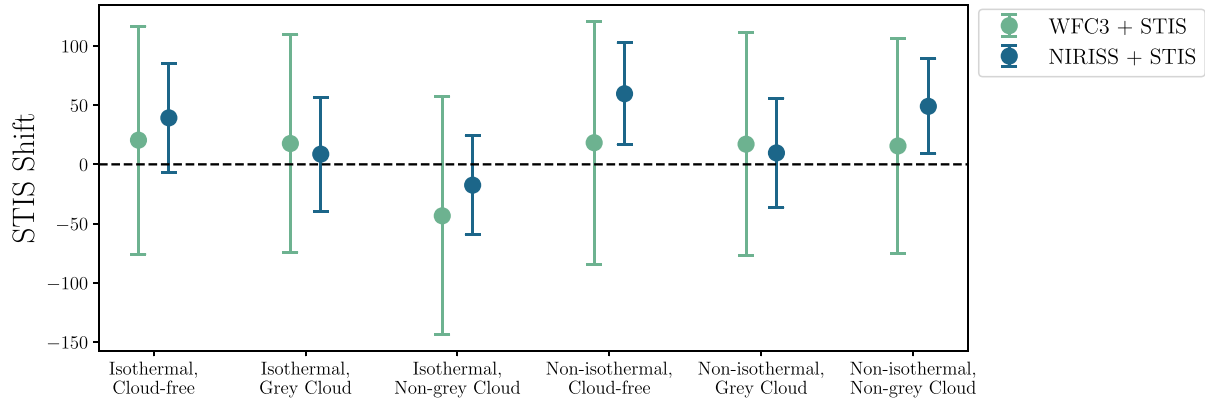


Figure A3. Median and 1σ error bars on a shift parameter applied to the STIS data for the retrievals combined with WFC3 and NIRISS for all six models.

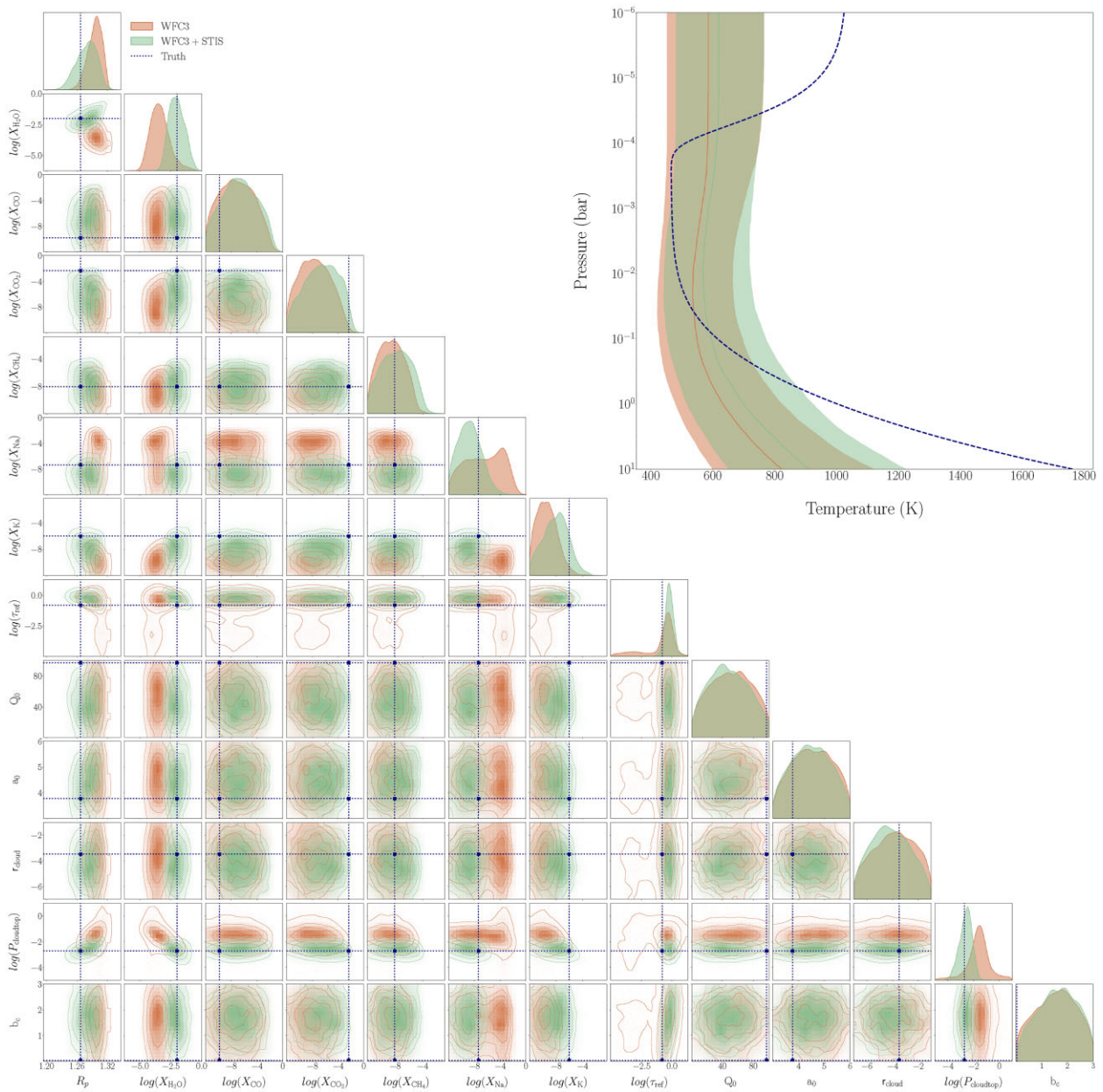


Figure A4. Mock retrieval on the WFC3 (purple) and WFC3 + STIS (green) simulated data, assuming the non-isothermal + non-greycloud model. The navy dashed lines show the true values used for the simulation.

APPENDIX B: RETRIEVED VALUES

Tables B1 and B2 show the retrieved values for each of the data sets for their best-fitting models.

Table B1. Retrieved radius, temperature, and molecular abundances for the best model (favoured by the Bayesian Evidence) for each data set. For temperature, if the best model is non-isothermal, the quoted temperature is the value at the 1 mbar pressure level.

Data set	Best model	R_P (R_J)	T (K)	$\log(X_{\text{H}_2\text{O}})$	$\log(X_{\text{CO}})$	$\log(X_{\text{CO}_2})$	$\log(X_{\text{CH}_4})$	$\log(X_{\text{Na}})$	$\log(X_{\text{K}})$
WFC3	Non-isothermal non-grey cloud	$1.28^{+0.01}_{-0.02}$	808^{+241}_{-219}	$-0.59^{+0.08}_{-0.12}$	$-6.86^{+3.16}_{-3.01}$	$-6.83^{+3.1}_{-3.05}$	$-7.06^{+2.99}_{-2.93}$	$-6.75^{+3.08}_{-3.02}$	$-5.71^{+2.55}_{-3.35}$
WFC3 + STIS	Isothermal non-grey cloud	1.28 ± 0.01	1126^{+99}_{-97}	$-2.6^{+0.53}_{-0.54}$	$-0.84^{+0.07}_{-0.08}$	$-7.54^{+2.92}_{-2.95}$	$-8.22^{+2.5}_{-2.45}$	$-6.36^{+0.89}_{-0.84}$	$-7.05^{+1.19}_{-1.25}$
NIRISS	Non-isothermal non-grey cloud	1.28 ± 0.01	409^{+69}_{-53}	$-1.05^{+0.1}_{-0.11}$	$-7.01^{+2.99}_{-3.12}$	$-2.72^{+0.31}_{-0.51}$	$-8.74^{+2.1}_{-2.05}$	$-7.88^{+2.56}_{-2.61}$	$-6.72^{+0.79}_{-1.64}$
NIRISS + STIS	Non-isothermal non-grey cloud	1.21 ± 0.01	658^{+37}_{-32}	$-3.21^{+0.18}_{-0.16}$	$-7.77^{+2.64}_{-2.42}$	$-7.33^{+2.11}_{-2.72}$	$-9.21^{+1.63}_{-1.6}$	$-6.46^{+0.33}_{-0.36}$	$-8.14^{+0.33}_{-0.36}$

Table B2. Retrieved cloud parameters for the best model (favoured by the Bayesian Evidence) for each data set.

Data set	Best model	$\log(\tau_{\text{cloud}})$	Q_0	a_0	r_{cloud}	$\log(P_{\text{cloudtop}})$	b_c
WFC3	Non-isothermal non-grey cloud	$-1.88^{+1.61}_{-1.79}$	$54.25^{+26.97}_{-29.21}$	$4.4^{+0.91}_{-0.81}$	-4.01 ± 1.75	$-1.9^{+1.68}_{-1.41}$	$1.6^{+0.8}_{-0.88}$
WFC3 + STIS	Isothermal non-grey cloud	$-3.31^{+0.5}_{-0.43}$	$51.02^{+32.01}_{-32.92}$	$5.02^{+0.68}_{-1.07}$	$-4.03^{+1.94}_{-1.93}$	$-4.26^{+1.36}_{-0.53}$	$1.4^{+1.03}_{-0.93}$
NIRISS	Non-isothermal non-grey cloud	$-0.18^{+0.32}_{-0.31}$	$49.82^{+31.34}_{-30.98}$	$4.89^{+0.75}_{-0.98}$	$-4.03^{+1.86}_{-1.88}$	-2.45 ± 0.24	$1.58^{+0.9}_{-1.03}$
NIRISS + STIS	Non-isothermal non-grey cloud	$0.24^{+0.45}_{-0.56}$	$51.46^{+28.4}_{-28.37}$	$4.71^{+0.75}_{-0.85}$	$-4.01^{+1.69}_{-1.68}$	$-1.81^{+0.08}_{-0.1}$	$0.73^{+1.3}_{-0.63}$

This paper has been typeset from a \LaTeX file prepared by the author.