



A Comprehensive Reanalysis of K2-18 b's JWST NIRISS+NIRSpec Transmission Spectrum

Stephen P. Schmidt^{1,18}, Ryan J. MacDonald^{2,3,19}, Shang-Min Tsai^{4,5}, Michael Radica^{6,7,20}, Le-Chris Wang¹, Eva-Maria Ahrer⁸, Taylor J. Bell^{9,10,11}, Chloe Fisher¹², Daniel P. Thorngren¹, Nicholas Wogan¹³, Erin M. May¹⁴, Piero Ferrari¹⁵, Katherine A. Bennett¹⁶, Zafar Rustamkulov¹⁶, Mercedes López-Morales¹⁷, and David K. Sing^{1,16}

¹ William H. Miller III Department of Physics and Astronomy, Johns Hopkins University, Baltimore, MD 21218, USA; sschmi42@jh.edu

² Department of Astronomy, University of Michigan, 1085 S. University Ave., Ann Arbor, MI 48109, USA

³ School of Physics and Astronomy, University of St Andrews, North Haugh, St Andrews, KY16 9SS, UK

⁴ Institute of Astronomy & Astrophysics, Academia Sinica, Taipei 10617, Taiwan

⁵ Department of Earth and Planetary Sciences, University of California, Riverside, CA, USA

⁶ Department of Astronomy & Astrophysics, University of Chicago, 5640 South Ellis Ave., Chicago, IL 60637, USA

⁷ Institut Trotier de Recherche sur les Exoplanètes et Département de Physique, Université de Montréal, 1375 Ave. Thérèse-Lavoie-Roux, Montréal, QC, H2V 0B3, Canada

⁸ Max Planck Institute for Astronomy, Heidelberg, 69117, Germany

⁹ AURA for the European Space Agency (ESA), Space Telescope Science Institute, 3700 San Martin Dr., Baltimore, MD 21218, USA

¹⁰ Bay Area Environmental Research Institute, NASA Ames Research Center, M.S. 245-6, Moffett Field, 94035 CA, USA

¹¹ Space Science and Astrobiology Division, NASA Ames Research Center, M.S. 245-6, Moffett Field, 94035 CA, USA

¹² Astrophysics, University of Oxford, Denys Wilkinson Building, Keble Rd., Oxford, OX1 3RH, UK

¹³ NASA Ames Research Center, Moffett Field, CA 94035, USA

¹⁴ Johns Hopkins APL, Laurel, MD 20723, USA

¹⁵ HFML-FELIX, Radboud University, Nijmegen 6525 ED, The Netherlands

¹⁶ Morton K. Blaustein Department of Earth & Planetary Sciences, Johns Hopkins University, Baltimore, MD 21218, USA

¹⁷ Space Telescope Science Institute, 3700 San Martin Dr., Baltimore, MD 21218, USA

Received 2025 January 30; revised 2025 August 21; accepted 2025 August 28; published 2025 November 5

Abstract

Sub-Neptunes are the most common type of planet in our galaxy. Interior structure models suggest that the coldest sub-Neptunes could host liquid water oceans underneath their hydrogen envelopes—sometimes called “hycean” planets. JWST transmission spectra of the ~ 250 K sub-Neptune K2-18 b were recently used to report detections of CH₄ and CO₂, alongside weaker evidence of (CH₃)₂S (dimethyl sulfide, or DMS). Atmospheric CO₂ was interpreted as evidence for a liquid water ocean, while DMS was highlighted as a potential biomarker. However, these notable claims were derived using a single data reduction and retrieval modeling framework, which did not allow for standard robustness tests. Here, we present a comprehensive reanalysis of K2-18 b's JWST NIRISS SOSS and NIRSpec G395H transmission spectra, including the first analysis of the second-order NIRISS SOSS data. We incorporate multiple well-tested data reduction pipelines and retrieval codes, spanning 60 different data treatments and over 250 atmospheric retrievals. We confirm the detection of CH₄ ($\approx 4\sigma$), with a volume mixing ratio range $-2.14 \leq \log_{10} \text{CH}_4 \leq -0.53$, but we find no statistically significant or reliable evidence for CO₂ or DMS. Finally, we assess the retrieved atmospheric composition using photochemical-climate and interior models, demonstrating that our revised composition of K2-18 b can be explained by an oxygen-poor mini-Neptune without requiring a liquid water surface or life.

Unified Astronomy Thesaurus concepts: Exoplanets (498); Exoplanet atmospheres (487); Exoplanet structure (495); Habitable planets (695); Mini Neptunes (1063); Exoplanet atmospheric composition (2021); Ocean planets (1151); Exoplanet surfaces (2118); Extrasolar gaseous planets (2172)

1. Introduction

James Webb Space Telescope (JWST) transmission spectroscopy has revealed the atmospheric composition of dozens of transiting exoplanets in the 3 yr since its launch. The most amenable planets to atmospheric characterization are hot H₂-dominated giant exoplanets. Major JWST results from such hot giant planets include detections of SO₂ (e.g.,

L. Alderson et al. 2023; Z. Rustamkulov et al. 2023; A. Dyrek et al. 2024; D. Powell et al. 2024; D. K. Sing et al. 2024; L. Welbanks et al. 2024), which provides direct evidence for photochemistry in exoplanetary atmospheres (e.g., S.-M. Tsai et al. 2023), and the direct detection of aerosol species (D. Grant et al. 2023; A. Dyrek et al. 2024; J. Inglis et al. 2024). JWST is also enabling initial forays into population-level trends across properties of giant exoplanet atmospheres (e.g., G. Fu et al. 2025), which will provide deeper insights as more transmission spectra are analyzed in the coming years.

While JWST transmission spectra have succeeded in probing the giant exoplanet atmospheres, decisive results have been more elusive for the small and cold ends of the exoplanet population. For super-Earths and terrestrial worlds ($R_p < 1.7 R_\oplus$), JWST transmission spectra largely rule out thick H₂-dominated atmospheres (e.g., J. Lustig-Yaeger et al. 2023; O. Lim et al. 2023;

¹⁸ NSF Graduate Research Fellow.

¹⁹ NHFP Sagan Fellow.

²⁰ NSERC Postdoctoral Fellow.



L. Alderson et al. 2024; M. Radica et al. 2025) or are degenerate with unocculted stellar active regions (e.g., S. E. Moran et al. 2023; E. M. May et al. 2023). However, sub-Neptunes,²¹ planets with radii $1.7R_{\oplus} < R < 3.5R_{\oplus}$, likely possess significant atmospheres dominated by relatively light molecules (e.g., H_2 or H_2O), which aids their atmospheric detectability. While searches for atmospheres on several sub-Neptunes have yielded nondetections (e.g., N. L. Wallack et al. 2024) or weak evidence (e.g., C. Cadieux et al. 2024; M. Damiano et al. 2024), an emerging population now has definitive atmospheric detections. To date, JWST has detected atmospheres on five sub-Neptunes: GJ 3470 b (T. G. Beatty et al. 2024), GJ 9827 d (C. Piaulet-Ghorayeb et al. 2024), TOI-270 d (B. Benneke et al. 2024; M. Holmberg & N. Madhusudhan 2024), TOI-421 b (B. Davenport et al. 2025), and K2-18 b (N. Madhusudhan et al. 2023). Of these planets, the cold (~ 250 K) sub-Neptune K2-18 b has attracted significant observational and theoretical interest in its atmospheric composition and interior structure.

The K2-18 system was identified as a potential transiting exoplanet system in the first campaign of NASA’s K2 mission (B. T. Montet et al. 2015). Further transit observations by Spitzer (B. Benneke et al. 2017) and radial velocity follow-up (R. Cloutier et al. 2017; P. Sarkis et al. 2018) confirmed K2-18 b as a habitable zone planet, with a refined radial velocity extraction method further improving its mass precision (M. Radica et al. 2022a). The system is now known to host at least two planets (R. Cloutier et al. 2019; M. Radica et al. 2022a), although only K2-18 b transits its early M dwarf host star. The combination of its cool temperature (255 K, assuming an albedo of 0.3; B. Benneke et al. 2019), relatively large planetary radius ($2.61R_{\oplus}$), and small stellar radius ($0.4445R_{\odot}$) has led to K2-18 b being one of the best targets for the atmospheric characterization of a potentially habitable exoplanet.

Initial transmission spectroscopy observations of K2-18 b with the Hubble Space Telescope (HST) yielded a detection of a H_2 -dominated atmosphere, with evidence of either gas phase H_2O or CH_4 . The HST Wide Field Camera 3 (WFC3) transmission spectrum of K2-18 b was initially interpreted as showing a $>3\sigma$ detection of H_2O (B. Benneke et al. 2019; A. Tsiaras et al. 2019; N. Madhusudhan et al. 2020). However, later work by T. Barclay et al. (2021) argued that the low-resolution WFC3 data could alternatively be explained by CH_4 , due to the similar shapes of H_2O and CH_4 absorption over the $1.1\text{--}1.7\mu\text{m}$ WFC3 bandpass. These competing interpretations of K2-18 b’s atmospheric composition resulted in an inability to definitively constrain the nature of its atmosphere and interior through HST observations alone.

Atmospheric and interior models of K2-18 b have ignited a vigorous debate on the nature of this world. Early atmospheric models disfavored a water reservoir below the H_2 -dominated atmosphere as the source of the apparent H_2O vapor (M. Scheucher et al. 2020), while general circulation models pointed toward a misinterpretation of the HST data as caused by CH_4 rather than H_2O (D. Blain et al. 2021). Interior structure models of K2-18 b highlighted that the pressure–temperature (P–T) profile could allow a habitable liquid water ocean beneath a thin H_2 atmosphere (N. Madhusudhan et al. 2020; A. A. A. Piette & N. Madhusudhan 2020; M. C. Nixon & N. Madhusudhan 2021)—sometimes termed a “hycean”

planet. Sustaining such a liquid water ocean requires the planet to have a H_2 -rich atmosphere, an iron-rock core comprising $>10\%$ of the planet’s mass, and a H_2O layer with a mass fraction from 10% – 90% (N. Madhusudhan et al. 2021). Such an ocean can be considered a habitable surface, which renders K2-18 b an object of potential astrobiological interest in the search for extraterrestrial life.

Recently, the first JWST transmission spectrum of K2-18 b revealed multiple prominent atmospheric absorption features (N. Madhusudhan et al. 2023). The Near Infrared Imager and Slitless Spectrograph (NIRISS) Single Object Slitless Spectroscopy (SOSS) and Near Infrared Spectrograph (NIRSpec) G395H observations (spanning $0.8\text{--}5.3\mu\text{m}$) were explained by strong atmospheric CH_4 bands (detected at 5σ), a peak near $4.3\mu\text{m}$ attributed to CO_2 (at 3σ), and “potential signs” of $(CH_3)_2S$ (dimethyl sulfide, or DMS). The JWST detection of CH_4 , and nondetection of H_2O , revealed that the previous inference of H_2O from HST was indeed a case of mistaken identity. N. Madhusudhan et al. (2023) reported an atmospheric composition of $\sim 1\%$ CH_4 and $\sim 1\%$ CO_2 , which is challenging to explain under standard thermochemistry for a mini-Neptune with a deep atmosphere. They also reported a nondetection of NH_3 , a gas that would generally be expected in a mini-Neptune at K2-18 b’s temperature, with an upper limit of ≈ 30 ppm. N. Madhusudhan et al. (2023) argued that these two lines of evidence point toward a water ocean beneath the H_2 -dominated outer layer, since an ocean would increase the CO_2/CH_4 ratio and deplete NH_3 (R. Hu et al. 2021; S.-M. Tsai et al. 2021a).

Subsequent modeling efforts have argued that the JWST observations of K2-18 b do not require a habitable liquid water ocean. O. Shorttle et al. (2024) proposed that the NH_3 nondetection can be explained by a deep H_2 envelope above a magma ocean, which allows efficient dissolution of nitrogen. J. Yang & R. Hu (2024) found that the CO_2/CH_4 ratio of ~ 1 from N. Madhusudhan et al. (2023) can also be explained by a gas-rich mini-Neptune enriched in water vapor in the deep atmosphere. N. F. Wogan et al. (2024) used photochemical and climate models to investigate three scenarios for K2-18 b: (i) a lifeless hycean planet, (ii) an “inhabited” hycean planet with a substantial biological CH_4 flux into the atmosphere, and (iii) a $100\times$ solar metallicity mini-Neptune. They ruled out the first scenario (which would have < 1 ppm of CH_4), such that the $\sim 1\%$ CH_4 abundance inferred from JWST could be explained by either a methane-producing biosphere or, alternatively, a metal-enriched mini-Neptune. N. F. Wogan et al. (2024) argued that the mini-Neptune scenario is more plausible a priori, since this explanation works “out of the box” without requiring one to postulate the existence of life. Most recently, G. J. Cooke & N. Madhusudhan (2024) investigated the same three scenarios, arguing that the CO_2 abundance found by N. Madhusudhan et al. (2023) favors the “inhabited” hycean scenario over a mini-Neptune.

The astrobiological assessment of H_2 -dominated sub-Neptunes such as K2-18 b is at an early stage, and hence, any claims of life must be tempered with extreme caution. N. Madhusudhan et al. (2023) highlighted a low-significance inference of DMS ($\sim 1\sigma$ from their best-fitting model) as potential evidence for life, given DMS is significantly produced by marine phytoplankton on Earth (R. J. Charlson et al. 1987; I. Barnes et al. 2006). Such a DMS flux would have to be equivalent to ~ 20 times that produced by Earth’s marine life to reach detectable levels on a hycean planet with K2-18 b-like

²¹ We use “sub-Neptune” when categorizing planets by size (super-Earth), and “mini-Neptune” when categorizing planets by interior structure (hycean).

conditions (S.-M. Tsai et al. 2024). C. R. Glein (2024) showed that metabolic reactions consuming CO₂ and/or DMS and producing CH₄ can potentially produce sufficient free energy to sustain a biosphere (if one assumes a hycean scenario and the atmospheric abundances reported in N. Madhusudhan et al. 2023) on K2-18 b. However, DMS in this metabolic pathway cannot strictly be considered a biosignature due to it being an *energy source* for methane-producing life rather than a *by-product* of life. Furthermore, the recent detection of DMS on a comet by N. Hänni et al. (2024) demonstrates that DMS can also be produced in abiotic environments. These issues with DMS as potential evidence for a biosphere on K2-18 b in combination with the lack of detailed cross section data for it complicate any potential investigation of the planet’s habitability. Indeed, the state-of-the-art line data for DMS (HITRAN; S. W. Sharpe et al. 2004) lack pressure and temperature dependence, potentially biasing retrieval results that include it (K. Heng & D. Kitzmann 2017; L. Welbanks & N. Madhusudhan 2019; P. Niraula et al. 2022).

The plausibility of the hycean interpretation for K2-18 b relies critically on the retrieved chemical abundances reported by N. Madhusudhan et al. (2023). However, to date, there has been no independent analysis of the original JWST transmission spectrum of K2-18 b. The gold standard for interpreting JWST spectra of exoplanet atmospheres is to apply multiple data reduction techniques and retrieval codes to confirm the reproducibility and robustness of the results—this philosophy has been broadly adopted by the exoplanet community (e.g., L.-P. Coulombe et al. 2023; D. Grant et al. 2023; J. Taylor et al. 2023; D. Powell et al. 2024; A. Banerjee et al. 2024; A. Gressier et al. 2025; C. Piaulet-Ghorayeb et al. 2024; L. Welbanks et al. 2024). However, the K2-18 b analysis by N. Madhusudhan et al. (2023) investigated only a single reduction and retrieval model framework. Performing several reductions on the same data set quantifies the impact of choices such as outlier rejection, noise correction, and limb-darkening treatments that occur when using independent pipelines. Similarly, performing retrievals with different codes can illustrate how some model-level assumptions (e.g., opacity sources, aerosol parameterization) propagate into atmospheric inferences (e.g., C. Mai & M. R. Line 2019; L. Welbanks & N. Madhusudhan 2019; J. K. Barstow et al. 2020; P. Niraula et al. 2023; M. C. Nixon et al. 2024a). Atmospheric inferences that withstand these tests are considered robust. On the other hand, severe inconsistencies (e.g., a molecular detection seen only with one reduction) demonstrate that the statistical evidence for such an atmospheric inference must be carefully assessed. Here, we provide the first comprehensive analysis of K2-18 b’s JWST NIRISS SOSS and NIRSpec G395H transmission spectra, using multiple data reduction and retrieval codes, to offer an updated assessment of K2-18 b’s atmospheric composition and possible interior structures.

We note that, during the revision of this study, N. Madhusudhan et al. (2025) presented the MIRI LRS transmission spectrum of K2-18 b. While N. Madhusudhan et al. (2025) reported independent evidence for DMS and/or DMDS (dimethyl disulfide) at $\sim 3\sigma$, subsequent studies have demonstrated that the MIRI data are well fit by a flat line (J. Taylor 2025), that DMS and DMDS do not provide a uniquely good fit relative to other molecules (L. Welbanks et al. 2025; L. Pica-Ciamarra et al. 2025), and that DMS or DMDS is not detected when the MIRI data are analyzed in concert with the near-infrared data (R. Luque et al. 2025). These studies highlight the low

information content of the MIRI data relative to the shorter wavelength near-infrared data, which we focus on in this study.

Our study is structured as follows. We describe our new K2-18 b data reductions in Sections 2 and 3 for NIRISS SOSS and NIRSpec G395H, respectively. Our retrieval analysis and atmospheric composition constraints are presented in Section 4. We model plausible atmospheric and interior structures for K2-18 b in Section 5. Finally, we summarize our results and discuss the implications in Section 6.

2. NIRISS Data Reductions

The K2-18 system was observed by JWST’s NIRISS (R. Doyon et al. 2012, 2023) instrument using the SOSS (L. Albert et al. 2023) mode as part of JWST GO Program 2722 (PI: N. Madhusudhan; N. Madhusudhan 2023). The details of this observation are described thoroughly in N. Madhusudhan et al. (2023). We perform two reductions of this data set with two independent pipelines: FIREFLY and exoTEDRF.²² To allow a transparent comparison between our reduction approaches and N. Madhusudhan et al. (2023; who used the JEXORES pipeline M. Holmberg & N. Madhusudhan 2023), we provide an overview of the most important reduction-level configuration choices for NIRISS in Table 1. Both of our reductions fit the spectroscopic light curves at a 2 pixel binning level (the same binning strategy as N. Madhusudhan et al. 2023) as well as at resolving powers $R \approx 25$ and $R \approx 100$. We use 2 pixel binning for NIRISS SOSS rather than 1 pixel binning as its point-spread function (PSF) is 2 pixels wide. Additionally, we show the point-by-point discrepancies between reductions in Appendix C, as has been done previously for analyses involving multiple reductions where small point-by-point differences could yield different atmospheric interpretations (e.g., L. Alderson et al. 2024, 2025; M. K. Alam et al. 2025; J. Kirk et al. 2025; D. R. Louie et al. 2025).

2.1. FIREFLY

We use the FIREFLY pipeline (Z. Rustamkulov et al. 2022, 2023), which has been recently updated to better support NIRISS/SOSS observations (Liu et al. 2025; L. C. Wang et al. 2025, in preparation), to rereduce the K2-18 NIRISS/SOSS observations. A more thorough inventory of the steps we use is included in Table 1; here, we describe the most pertinent choices made. To correct the $1/f$ noise at the group level, we first subtract the zodiacal background at each group. This is done by scaling the flux jump caused by the reflection of zodiacal light off the pick-off mirror in the STScI SOSS background model to the flux jump present in the observed data and adding a constant offset to match the flux level of the model SOSS background before and after the jump to the observed data. We then mask out the bright spectral trace and notable zeroth-order contaminants present in the frame and performed $1/f$ destriping by subtracting the median of each PSF-masked and background-subtracted column. In addition, we subtract each image from the temporal median, which we create as a running median of a seven frame window, to reveal any remaining $1/f$ noise and subtract it. This removes high-frequency structure while preserving the slowly changing source flux. Because the background levels are real counts and

²² We do not perform NIRISS reductions with the Eureka! pipeline, as support for this instrument mode is currently pending.

Table 1
NIRISS Reduction Configuration Comparison

Step	FIREFLY (This Work)	ExoTEDRF (This Work)	N. Madhusudhan et al. (2023)	Reference
Stage 1				
Saturation	Default	Default	Default	
Superbias	Default	Default	Default	
Dark current	Skipped	Default	Skipped	
Background subtraction	Single-component STScI background	Two-component STScI background	Two-component STScI background	
(group level)	model	model	model	
1/f subtraction	Yes	scale-achromatic	Yes	
(group level)				
Linearity	Default	Default	Default	
Jump	Skipped	exoTEDRF-custom, 10 σ in time	Default, 5 σ	
Ramp fitting	Default	Default	Default	
Stage 2				
Wavelength calibration	pastasoss	Cross correlation with PHOENIX	?	pastasoss: (T. Baines et al. 2023)
	stellar model			
1/f subtraction (integration-level)	Yes	No	Yes	
Bad pixel cleaning	>5 σ outlier pixels	>5 σ outlier pixels in space and time	?	
Order tracing	pastososs	edgetrigger	similar to edgetrigger	
Stage 3				
Temporal outliers	>15 σ	>5 σ	?	
1D Spectrum Extraction	Box extraction, 30 pixel aperture for order 1, 23 pixel aperture for order 2 aperture	Box extraction, 30 pixel aperture	Multisorder	Optimal extraction: K. Horne (1986)
Exclusions	None	None	Columns with > 20% masked flux	
Stage 4				
Light-curve model	batman	batman	SPOTROD	
Spot-crossing event treatment	Integrations trimmed light-curve model	Gaussian in light-curve model	Four-parameter spot model in	
Limb darkening	Fixed to Stagger model with offsets based on white light-curve fit	Fit for each bin	Binned to $R \sim 20$, fit in each bin	
LD parameterization	$\{u_+, u_-\}$	$\{u_1, u_2\}$	$\{q_1, q_2\}$	
Systematic trends	linear	linear	linear	
Code used for white light-curve posterior	emcee	emcee	MultiNest	
Fitted WLC a/R^*	81.32 ± 1.74	...	$79.9^{+1.4}_{-1.4}$	
Fitted WLC b	0.6031 ± 0.0237	...	...	
Fitted WLC i	...	...	$89.550^{+0.021}_{-0.020}$	
Fitted WLC T_0	$60096.72942 \pm 6 \times 10^{-5}$	57264.3914	$60096.729368^{+0.00063}_{-0.00065}$	
Binning method	2 pixel, $R \approx 25$, and $R \approx 100$	2 pixel, $R \approx 25$, and $R \approx 100$	2 pixel level	
Code used for spectrophotometric fit	Levenberg–Marquardt; lmfit	MCMC; emcee	Levenberg–Marquardt	
Error bar inflation	Red noise via residual rms-bin size trend	Additive error inflation	?	
Code Availability	Not Public	exoTEDRF GitHub	Not Public	

Note. “Default” refers to the default handling of the step in the `jwst` pipeline. The system parameters (a/R_* , i , and T_0) are consistent across each reduction, although they were calculated in N. Madhusudhan et al. (2023) based on their reduction up to that point. The listed orbital parameters were used to either confirm consistency with N. Madhusudhan et al. (2023; FIREFLY) or show negligible differences between resulting spectra with an alternative orbital solution (exoTEDRF, from M. Radica et al. 2022a). This is not an exhaustive list of reduction steps. “?” indicates reduction configuration settings not mentioned in N. Madhusudhan et al. (2023).

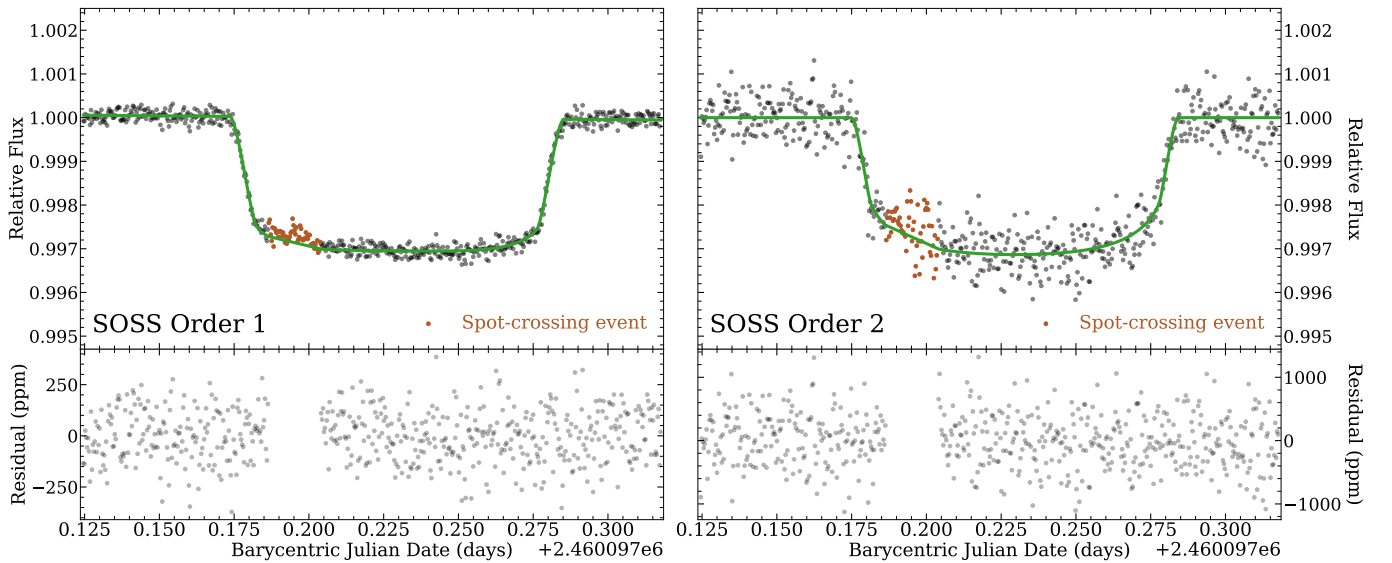


Figure 1. JWST NIRISS SOSS data and white light-curve fit for K2-18 b from the FIREFLY reduction. Left: first-order data. Right: second-order data. The photometric data points (points, top panels) are compared with the best-fit light curve (green lines), resulting in a 124 ppm residual scatter for order 1 and 414 ppm for order 2 (bottom panels). Integrations during a spot-crossing event are highlighted (brown points), which are trimmed in the white and spectroscopic light-curve fits.

not the detector’s bias level, we add the background back after subtracting the $1/f$ noise in order to ensure that the `jwst` pipeline’s bias correction and up-the-ramp fitting procedures are unaffected. We also apply an integration-level $1/f$ subtraction and clean $>5\sigma$ outlier bad/hot pixels.

In Stage 3, we apply a $>15\sigma$ temporal outlier cut, remove the background for the final time, and perform an additional $1/f$ noise removal at the integration level. Both the background removal and $1/f$ noise subtraction follow the same algorithm as the group level. To extract the white light curves, we sum all the flux in order 1 and the flux from wavelengths within $[0.6, 0.85] \mu\text{m}$ for order 2. We align frames and conduct box extractions on each order with aperture widths that minimize the scatter of the white light curve for each order.

Moving on to the light-curve fitting, we trim out integrations corresponding to a spot-crossing event in the NIRISS light curve. Informed by the Bayesian information criterion, we choose the systematics vector to be composed solely of a linear trend for both order 1 and order 2. We find that our fitted values are consistent with those reported by N. Madhusudhan et al. (2023); we therefore fix the orbital parameters to theirs for a consistent comparison in both the white light-curve and spectrophotometric fits. For limb darkening, we perform a two-step process of first fitting for spectroscopic limb-darkening coefficients and then fixing them to an offset stellar atmosphere model. We experiment with both the D. M. Kipping (2013) $\{q_1, q_2\}$ and the quadratic $u_+ = u_1 + u_2$ and $u_- = u_1 - u_2$ parameterizations, choosing $\{u_+, u_-\}$ as q_2 in particular tends to be uninformative about light-curve shape despite both spectra being virtually indistinguishable. Combining the fitted u_+ and u_- from order 1 and order 2, we find that these fitted limb-darkening coefficients trace well the limb-darkening coefficients from the Stagger grid of 3D stellar atmosphere models, with an offset of -0.14843 for u_+ and -0.07728 for u_- . We show the white light-curve fits for the NIRISS SOSS orders 1 and 2 separately in Figure 1, with the spot-crossing event highlighted. We account for red noise in our spectrophotometric fit by inflating error bars in quadrature for

wavelength bins where the residual rms-bin size trend is above log linear (F. Pont et al. 2006; J. N. Winn et al. 2008). This does not impact the error bars for most points, but for those that are, it is typically by a factor of no more than 2.

2.2. *exoTEDRF*

We also reduced the time-series observations using the publicly available *exoTEDRF* pipeline (A. D. Feinstein et al. 2023; L.-P. Coulombe et al. 2023; M. Radica et al. 2023; C. Cadieux et al. 2024; M. Radica 2024), starting from the raw, uncalibrated files available on MAST. We closely follow the procedure laid out in M. Radica et al. (2024), and we summarize the most pertinent points here. In `stage1`, we correct the column-correlated $1/f$ noise at the group level, using the `scale-achromatic` method. Before ramp fitting, we perform a time-domain outlier rejection (M. Radica et al. 2024), using a threshold of 10σ . We perform a piecewise correction of the SOSS background (O. Lim et al. 2023; M. Fournier-Tondreau et al. 2024), using the standard SOSS background model provided by STScI,²³ and a prestep and poststep scaling factor of 0.95784 and 0.92449, respectively, relative to the background model. Finally, we extract the stellar spectra using a simple box aperture extraction with a width of 30 pixels, since the order self-contamination is expected to be negligible (A. Darveau-Bernier et al. 2022; M. Radica et al. 2022b).

For the light-curve analysis, we first construct two separate white light curves by summing all the flux in order 1, and wavelengths $\lambda \in (0.6, 0.85]$ for order 2. We then jointly fit the white light curves using the flexible *exoUPRF* library (M. Radica 2024) following the procedure in M. Radica et al. (2024, 2025). The fitted light-curve model is composed of an astrophysical component (i.e., a `batman` transit model), and a systematics component. For the astrophysical component, we share the orbital parameters (i.e., time of

²³ <https://jwst-docs.stsci.edu/jwst-near-infrared-imager-and-slitless-spectrograph/niriss-observing-strategies/niriss-soss-recommended-strategies#gsc.tab=0>

midtransit, orbital inclination, and scaled semimajor axis, as well as the orbital eccentricity and argument of periastron) between the two orders, but separately fit the scaled planet radius and two parameters of the quadratic limb-darkening law to each. We put Gaussian priors on the orbital parameters based on the findings from N. Madhusudhan et al. (2023). The systematics component of our light-curve fit consists of a linear trend with time, fit independently to each order, as well as a scalar jitter term added in quadrature to the flux errors. Instead of commonly used prescriptions like *spotrod* (B. Béky et al. 2014; N. Madhusudhan et al. 2023; M. Fournier-Tondreau et al. 2024) to model the spot-crossing event, we simply model it with a Gaussian profile, which has been shown to produce equally accurate fits to spot-crossing events in transit light curves (Roy et al. 2025, in preparation). We fit the amplitude of the Gaussian independently to each order, but share the width and position of the Gaussian between the two. In all, our white light fit has 19 free parameters, and we use wide, uninformative priors for each parameter. For the fits, we use the affine-invariant sampler *emcee* with 40 chains and using 50,000 steps per chain. We discard the first 80% as burn-in.

For the spectrophotometric fits, we fix the orbital parameters to the best-fitting values from the white light-curve fits. We test fixing the orbital parameters to the updated solution presented in M. Radica et al. (2022a), but find negligible changes in the resulting transmission spectra. We also fix the location and width of the Gaussian spot model to our best-fitting white light-curve values, but fit the amplitude for each wavelength bin. Finally, we freely fit the quadratic limb-darkening parameters, systematic slope, and additive jitter term to each bin, resulting in a total of eight free parameters.

3. NIRSpec Data Reductions

The K2-18 system was also observed by JWST’s NIRSpec (P. Ferruit et al. 2012; S. M. Birkmann et al. 2014) instrument using the G395H grating as part of JWST GO Program 2722 (PI: N. Madhusudhan; N. Madhusudhan 2023). As with NIRISS, the details of these observations are discussed in N. Madhusudhan et al. (2023). We perform a total of four NIRSpec reductions: one with the *FIREFLY* pipeline, one with the *ExoTEDRF* pipeline, and two with the *Eureka!* pipeline. We bin the data in four ways: resolving powers $R \approx 100$, $R \approx 200$, and $R \approx 300$, and at native pixel resolution. We summarize the most relevant choices made in each reduction, as well as a comparison to the reduction choices made in N. Madhusudhan et al. (2023)’s *JEXORES* NIRSpec reduction, in Table 2. Similarly with NIRISS, we show the point-by-point discrepancies between reductions in Appendix C. Combining these reductions with our NIRISS reductions, we compare the low-resolution data variants from each reduction code in Figure 2 ($R \approx 25$ for NIRISS; $R \approx 100$ for NIRSpec) and directly compare our pixel-level data, binned to the same resolution as in N. Madhusudhan et al. (2023), in Figure 3.

3.1. FIREFLY

We also used the *FIREFLY* pipeline to conduct a full time-series analysis of the K2-18 NIRSpec/G395H data. We again list a more thorough inventory of the steps in Table 2, describing here the most pertinent choices made. We skip the jump step in Stage 1 to avoid cosmic-ray false positives due to a less well-constrained ramp linear slope, as typically occurring with data sets with fewer than 25 groups per

integration. In Stage 3, we clean bad pixels by flagging ones with sharp variance spikes of over 100σ using *lacosmic* (P. G. van Dokkum 2001) and other known bad pixels in NIRSpec G395H as the first part of the stellar extraction. Our box extraction’s aperture full-widths were selected as they minimize the scatter of the out-of-transit white light curve.

In the light-curve fitting step, we trim the first 575 pixel columns of NRS1, first 8 pixel columns of NRS2, and last 18 pixel columns of NRS2. We find that our fitted, weighted average orbital parameters are consistent with those reported by N. Madhusudhan et al. (2023); we therefore fix the orbital parameters to theirs for our white light-curve fit. For consistency with N. Madhusudhan et al. (2023), we experiment with including both a linear and quadratic term in time, but we find that only a linear term is favored; we therefore choose to only include a linear term for our systematics model. We show the results of our white light-curve fit in Figure 4.

For our spectrophotometric fit, we fix the limb-darkening parameters and linear systematic term to the fitted white light-curve values in addition to the already fixed orbital parameters. We elect to fix the spectrophotometric limb-darkening values to the white light-curve values as limb darkening is not expected to be strongly wavelength dependent at the redder wavelengths of NIRSpec (E. M. May et al. 2023; S. E. Moran et al. 2023). We account for red noise in our spectrophotometric fit by inflating error bars in quadrature for wavelength bins where the residual rms-bin size trend is above log linear. For most points, this does not result in error bar inflation, and for points that are inflated, it is typically by a factor of no more than 2. To demonstrate the fidelity of our fitting procedure given that *FIREFLY* is proprietary software, we show all of the spectroscopic light-curve fits and residuals for each bin of our $R \approx 100$ reduction in Appendix B.

3.2. exoTEDRF

To reduce the NIRSpec/G395H observations, we also use the *ExoTEDRF* pipeline (M. Radica et al. 2023; A. D. Feinstein et al. 2023; M. Radica 2024), which has recently been updated to support NIRSpec observations (M. Radica 2024; Ahrer et al. submitted). We start from the raw, uncalibrated data files and follow the procedure outlined in Ahrer et al. (submitted). For completeness, we summarize the key points here. We perform standard saturation and superbias corrections on the data frames (e.g., L. Alderson et al. 2023) before correcting the background and $1/f$ noise at the group level, using the median method and a trace mask width of 16 pixels. As with SOSS, we perform a time-domain outlier flagging (M. Radica et al. 2024) with a rejection threshold of 12σ . We then repeat the background and $1/f$ correction at the integration-level, after ramp fitting, in order to remove any remaining traces of the background flux. We then interpolate all pixels that have a nonzero data quality flag, or are flagged as 10σ temporal outliers using a median of the surrounding pixels in space and time, respectively. For the spectral extraction, we use the *edgetrigger* algorithm (M. Radica et al. 2022b) to determine the centroids of the NRS1 and NRS2 spectral traces, and then perform a box aperture extraction with a width of 8 pixels.

We follow the same light-curve fitting procedure as with the SOSS light curves—jointly fitting the NRS1 and NRS2 white light curves, sharing the planet’s orbital parameters between the two, and independently fitting for chromatic parameters like the scaled planet radius, limb darkening, and systematics.

Table 2
NIRSpec Reduction Configuration Comparison

Step	FIREFLY (This Work)	ExoTEDRF (This Work)	Eureka! A (This Work)	Eureka! B (This Work)	N. Madhusudhan et al. (2023)	Reference
Stage 1						
Saturation	Default	Default	Default	Default	Default	
Superbias	Default	Default	Default	Eureka! custom	Default	
Dark current	Skipped	Default	Default	Default	Default	
Background subtraction (group level)	FIREFLY custom	exoTEDRF custom	Eureka! custom	Eureka! custom	JExoRES custom	
1/f subtraction (group level)	FIREFLY custom	exoTEDRF custom	Eureka! custom	Eureka! custom	JExoRES custom	
Linearity	Default	Default	Default	Default	Default	
Jump	Skipped in time	exoTEDRF custom, 12σ	Skipped	Default, 15σ	Default, 5σ	
Ramp fitting	Default	Default	Default	Default	Default	
Stage 2						
Wavelength calibration	Default PHOENIX model	Cross correlation with	Default	Default	Default	
1/f correction (integration level)	Yes	Yes	No	No	No	
Stage 3						
Bad pixel mask and temporal outliers	lacosmic and known bad pixels	Outliers $>10\sigma$ in time and $>10\sigma$ in space	Outliers $>3\sigma$ in time and $>5 \times$ median in space	Outliers $>4\sigma$ in time and $>2.5\sigma$ in space	Data Quality flags only	
1D Spectrum extraction	Box extraction with 5.2 pixel (for NRS1) and 2.41 pixel (for NRS2) aperture	Box extraction with edgetrigger-defined centroids component PSF	spatial profile-based extraction, 3 pixel half-width	spatial profile-based extraction, 4 pixel half-width	spatial profile-based extraction with three principal	Optimal extraction: K. Horne (1986)
Exclusions	None masked flux	$\geq 5\sigma$ outliers	$\geq 4\sigma$ outliers	$\geq 4\sigma$ outliers	Columns with $> 20\%$	
Stage 4						
Trim	First 575 pixels in NRS1; first 8 and last 18 pixels in NRS2	None	None	First 25 integrations	First 5 minutes	
Light-curve model	batman	batman	batman	batman	SPOTROD	
Starspot treatment	None	None	None	None	Four-parameter spot model	
Limb darkening	Fixed to fitted white light-curve values for the respective detector	Fit for each bin calculated with exo- tic-ld pixels within the bin	MPS-ATLAS model set 2, calculated with exotic-ld	PHOENIX model, fit in each bin and all	Binned to $R \sim 20$, then	
LD parameterization	$\{u_+, u_-\}$	$\{u_1, u_2\}$	$\{u_1, u_2\}$	Quadratic	$\{q_1, q_2\}$	
Systematic trends	Linear correlation with position and spatial PSF width	Linear	Linear	Linear in time and linear	Quadratic and linear	

Table 2
(Continued)

Step	FIREFLY (This Work)	ExoTEDRF (This Work)	Eureka! A (This Work)	Eureka! B (This Work)	N. Madhusudhan et al. (2023)	Reference
Code used for white light-curve posterior	emcee	emcee	emcee	dynesty	MultiNest	
Fitted WLC a/R^*	80.27 ± 0.98	...	80.68 ± 0.10	81.56 ± 0.85	$80.92^{+0.78}_{-0.72}$	
Fitted WLC b	0.621 ± 0.013	...	...	...	...	
Fitted WLC i (deg.)	...	...	89.564 ± 0.002	89.577 ± 0.013	$89.567^{+0.012}_{-0.011}$	
Fitted WLC T_0	59964.96946 ± 0.00004	57264.3914	59964.96947 ± 0.00006	$59964.969451 \pm 0.000034$	$59964.969453^{+0.00035}_{-0.00034}$	
Binning method	Pixel-level, $R \approx 100$, $R \approx 200$, and $R \approx 300$	Pixel-level, $R \approx 100$, $R \approx 200$, and $R \approx 300$	Pixel-level, $R \approx 100$, $R \approx 200$, and $R \approx 300$	Pixel-level, $R \approx 100$, $R \approx 200$, and $R \approx 300$	Pixel-level	
Code used for spectrophotometric fit	Levenberg–Marquardt algorithm; lmfit	MCMC; emcee algorithm	MCMC; emcee	Nested Sampling; dynesty	Levenberg–Marquardt	
Error bar inflation	Red noise via residual rms-bin size trend	Jitter term	White noise term	White noise term	?	
Code Availability	Not Public	exoTEDRF GitHub	Eureka! GitHub	Eureka! GitHub	Not Public	

Note. “Default” refers to the default handling of the step in the `jwst` pipeline. The system parameters (a/R_* , i , and T_0) are the same for each reduction, and are calculated in N. Madhusudhan et al. (2023). The listed orbital parameters were used to either confirm consistency with N. Madhusudhan et al. (2023; FIREFLY, Eureka! B) or to show negligible differences between resulting spectra with an alternative orbital solution (exoTEDRF, from M. Radica et al. 2022a). Note that this is not an exhaustive list of reduction steps. “?” indicates reduction configuration settings not mentioned in N. Madhusudhan et al. (2023).

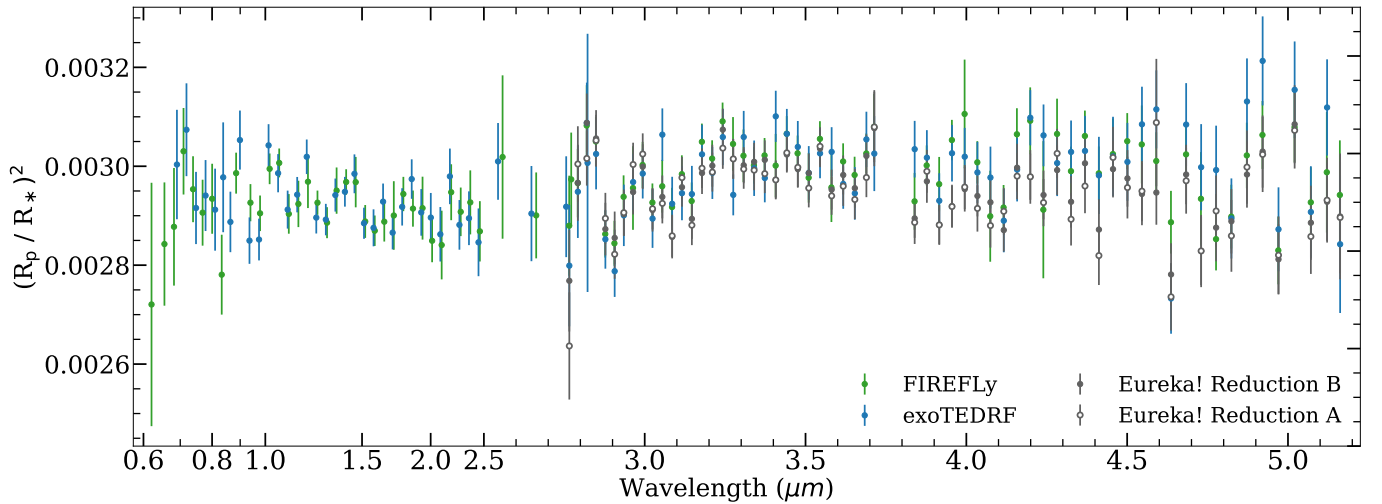


Figure 2. Four independently reduced JWST transmission spectra of K2-18 b. NIRISS SOSS and NIRSpec G395H data sets reduced using the FIREFLY pipeline (green error bars; Z. Rustamkulov et al. 2022, 2023; Liu & Wang et al. 2025, in preparation; Wang et al. 2025, in preparation) are compared with the same data reduced with the exoTDRF pipeline (blue error bars; A. D. Feinstein et al. 2023; M. Radica et al. 2023; M. Radica 2024) and two NIRSpec G395H reductions using the Eureka! pipeline (error bars with gray (B)/white (A) central points; T. Bell et al. 2022). NIRISS data are shown at $R \approx 25$, and NIRSpec data are shown at $R \approx 100$. The four reductions demonstrate broad agreement.

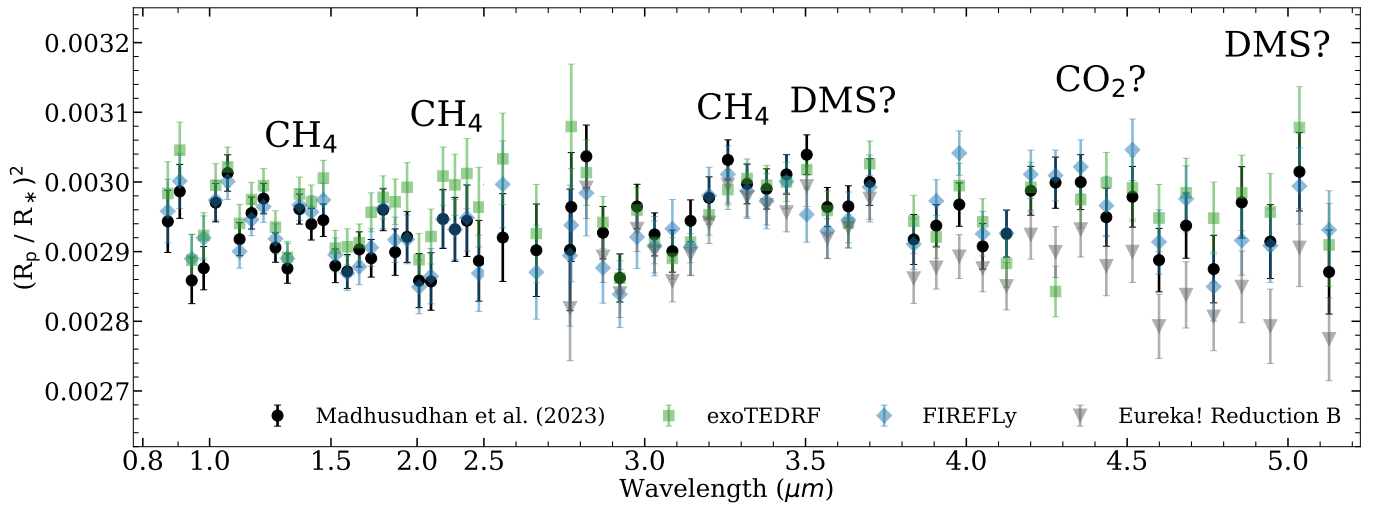


Figure 3. Comparison between our transmission spectra and the reduction from N. Madhusudhan et al. (2023). Three of our reductions, which were performed at the pixel-level and then binned to $R \approx 55$, are shown: exoTDRF (green squares), FIREFLY (blue diamonds), and the NIRSpec-only Eureka! reduction B (gray downward-pointing triangles). We show only Eureka! reduction B, since reduction A was performed at $R = 100$. We compare these binned pixel-level data to the low-resolution data from N. Madhusudhan et al. (2023), which were obtained using the JEXORES pipeline (black error bars; M. Holmberg & N. Madhusudhan 2023). The N. Madhusudhan et al. (2023) data were binned directly from their pixel-level data, so we bin our pixel-level data using the same binning approach. These $R \approx 55$ data sets are purely for visual comparison with N. Madhusudhan et al. (2023), and are not used for retrievals later in our study. Wavelength locations of CH_4 , CO_2 , and DMS absorption features are annotated.

As with SOSS, we fix the orbital parameters to the values used by N. Madhusudhan et al. (2023; there is, once again, no difference in the resulting spectra using the M. Radica et al. 2022a orbital solution). There is no spot crossing in the NIRSpec light curves, and so, we do not include a Gaussian spot model. However, we keep the linear slope and error inflation term added in quadrature to the flux errors that were included in the SOSS fits. In all, our white light-curve fits have 15 free parameters. We again use *emcee* for the fits, using the same amount of chains, steps, and burn-in as above.

3.3. Eureka! Reductions

We conduct two independent reductions of the NIRSpec data using the open-source Python package Eureka! (T. Bell et al. 2022) to test the extent to which the reduction choices

within the same pipeline can yield differing final results. The Eureka! control and parameter files we used for each reduction, allowing replication of our results, are available for download on Zenodo: doi:10.5281/zenodo.14735688.

For our first Eureka! reduction, which we henceforth refer to as reduction “A,” we start with the uncalibrated files and run Stage 1 and Stage 2, which are wrapped around the *jwst* pipeline. We use the default steps in both stages with the exception of the jump step and the photom step, which we skip. We also use Eureka!’s group-level background subtraction and a custom bias scale factor (using a smoothing filter with a window length of 30 integrations and group 1) in Stage 1.²⁴

²⁴ See also Moran & Stevenson et al. (2023).

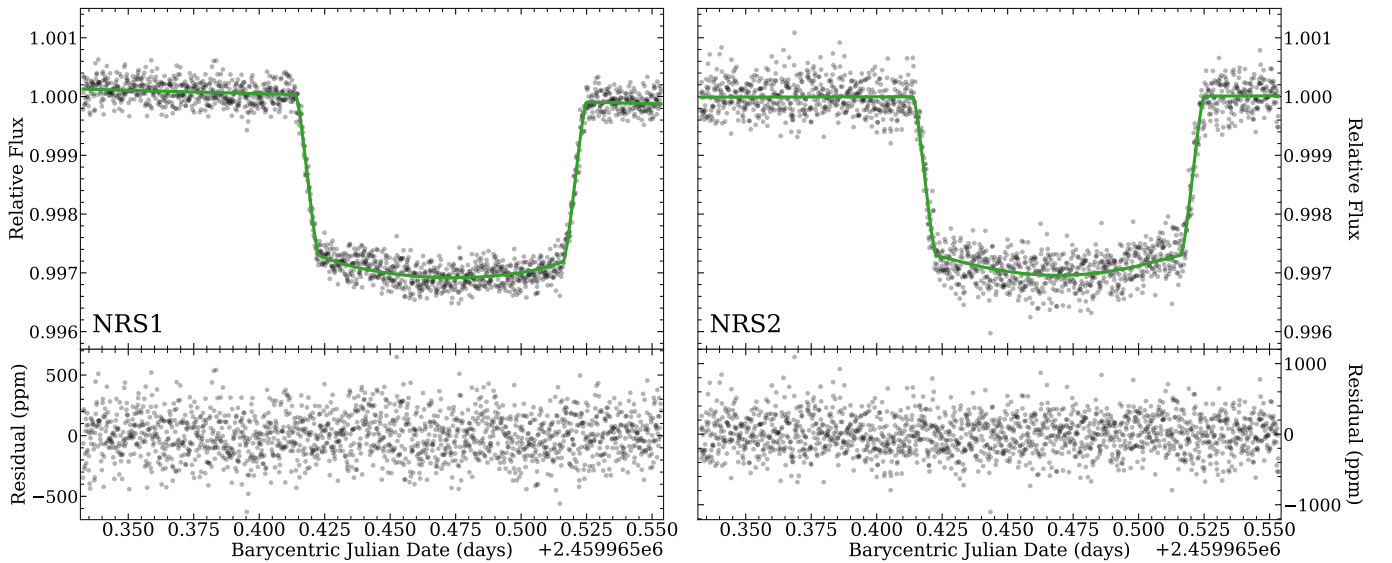


Figure 4. JWST NIRS1 G395H data and white light-curve fit for K2-18 b from the FIREFLY reduction. Left: NRS1 detector data. Right: NRS2 detector data. The photometric data points (points, top panels) are compared with the best-fit light curve (green lines), resulting in a 180 ppm residual scatter for NRS1 and 269 ppm for NRS2 (bottom panels).

In Stage 3, we extract the stellar time-series spectra. First, we correct the curved trace of the spectra and perform a background subtraction where we used the median of the entire frame, excluding the area within 5 pixels of the center of the trace. Prior to the calculation of the median background, we perform two iterations of outlier masking using thresholds of 3σ along the time and 5 times the median along the spatial axis. We use an aperture of 7 pixels, corresponding to a 3 pixel half-width, to extract the time-series spectra using the optimal extraction method (K. Horne 1986).

We generate pixel-level and binned light curves in Stage 4 of Eureka! following the same binning scheme as FIREFLY. We use a 3σ clipping compared to a 10 pixel rolling-median to reduce the number of outliers. We fit the light curves by fixing the system parameters to those reported in N. Madhusudhan et al. (2023) to allow for direct comparison, and we used the Markov Chain Monte Carlo (MCMC) Python package `emcee`. We opt to use a quadratic limb-darkening law, where we fix one parameter (u_1) to limb-darkening parameters generated using the `exotic-ld` Python package (D. Grant & H. Wakeford 2024) and the MPS-ATLAS (set 2) stellar atmosphere model grid (N. Kostogryz et al. 2023). We use a simple linear trend in time to account for a systematic trend in our light curves.

Our second Eureka! reduction, which we will henceforth refer to as reduction “B,” used Eureka! version 0.11.dev446 +gf5d684ee.d20240712; we started with the `_uncal` files produced with SDP_VER 2023_3b, and we used version 1.15.1 of the `jwst` pipeline with CRDS version 11.17.26 and CRDS context `jwst_1252`. Our reduction procedure largely followed the Eureka! reduction described in E. Schlawin et al. (2024); we summarize the important differences between this Eureka! reduction and that of E. Schlawin et al. (2024) below.

We increased the Stage 1 jump step rejection threshold to 15 (instead of 6). In Stage 3, we 4σ clipped background pixels along the time axis (instead of 10σ) and subtracted the mean background flux per column and per integration computed using only the pixels ≥ 6 pixels away from the center of the

spectral trace (instead of 7 pixels). Our optimal spectral extraction (K. Horne 1986) used a spatial profile with a half-width of 4 pixels (instead of 5 pixels). In Stage 4, we manually masked several pixel columns (18 in NRS1 and 9 in NRS2) that exhibited unusually high noise levels compared to their neighbors, likely due to unmasked bad pixels in an earlier stage. We then spectrally binned the data into multiple different spectral resolutions and masked any integration that was a $\geq 4\sigma$ outlier (instead of 3.5σ) with respect to a 20-integration wide boxcar filter, likely due to a missed cosmic ray.

When fitting the light curves, our astrophysical model consisted of a `batman` transit model with quadratic limb-darkening coefficients fixed to a `PHOENIX` limb-darkening model computed by `exotic-ld` (T. O. Husser et al. 2013; D. Grant & H. Wakeford 2024). We first verified that our broadband light curves were consistent with the orbital parameters reported by N. Madhusudhan et al. (2023), with our joint fit of the NRS1 and NRS2 broadband light curves finding $T_0 = 59964.969451 \pm 0.000034$ BMJD_TDB, $i = 89.577 \pm 0.013$, and $a/R_* = 81.56 \pm 0.85$ (we had fixed the orbital period to that of B. Benneke et al. 2019). We then fixed our orbital parameters to those of N. Madhusudhan et al. (2023). Our systematic model consisted of a linear trend in time and a linear correlation with the position and the width of the spatial PSF. We also fitted for a white noise multiplier to account for any excess white noise in the data beyond the estimated photon noise. We trimmed the first 25 integrations to remove any initial detector settling. To estimate the best-fit model values and their corresponding uncertainties, we used the `dynesty` nested sampling algorithm (J. S. Speagle 2020) with 121 live points, “multi” bounds, the “rwalk” sampling algorithm, and a convergence criterion of $\Delta \log \mathcal{Z} \leq 0.01$, where $\mathcal{Z}$ is the Bayesian evidence.

4. Atmospheric Retrieval Analysis of K2-18 b’s Transmission Spectrum

Here, we detail our atmospheric interpretation of K2-18 b’s JWST transmission spectrum. Our central goal is to quantify the sensitivity of atmospheric inferences to both data-level and

model-level choices, establishing which atmospheric properties are robust and which are not. To this end, we conduct an extensive atmospheric retrieval analysis across the 60 possible combinations of reanalyzed NIRISS and NIRSpec data produced in this study. We use two independent retrieval codes to verify the robustness of our results. In total, our analysis comprises over 250 retrievals. We describe our retrieval configurations in Section 4.1 and our results in Section 4.2.

4.1. Retrieval Configuration

We employ two open-source retrieval codes, POSEIDON and BeAR, to explore the range of atmospheric properties consistent with the revised K2-18 b’s transmission spectra derived above. We additionally show retrieval results using the data from N. Madhusudhan et al. (2023) in Appendix A. Our retrieval configurations are constructed to encompass the parameter space explored by N. Madhusudhan et al. (2023), while also considering additional factors (such as planetary mass uncertainties and including the second-order NIRISS SOSS data). Our model settings, retrieval configurations,²⁵ and a comparison with N. Madhusudhan et al. (2023) are summarized in Table 3.

4.1.1. POSEIDON

We first retrieved K2-18 b’s transmission spectrum using the retrieval code POSEIDON²⁶ (R. J. MacDonald & N. Madhusudhan 2017; R. J. MacDonald 2023). POSEIDON is widely applied to interpret JWST transmission spectra, including hot Jupiters (e.g., D. Grant et al. 2023; M. Fournier-Tondreau et al. 2025), sub-Neptunes (C. Piaulet-Ghorayeb et al. 2024), super-Earths (e.g., S. E. Moran et al. 2023; E. M. May et al. 2023), and terrestrial worlds (e.g., O. Lim et al. 2023; C. Cadieux et al. 2024). Here, we use POSEIDON to conduct a systematic retrieval survey of the sensitivity of K2-18 b’s atmospheric inferences to 60 different data variants (reduction codes and/or data spectral resolution). Our POSEIDON retrieval configuration is constructed to be as close as possible to that described in N. Madhusudhan et al. (2023), allowing a fair comparison, but we emphasize and justify below several areas where we adopted different model choices.

Our atmosphere model for K2-18 b assumes a H₂-He background gas (with a solar ratio of He/H₂ = 0.17; M. Asplund et al. 2009), commensurate with its planetary radius and mass ($\approx 2.6 R_{\oplus}$ and $\approx 8.6 M_{\oplus}$, respectively). We construct atmospheres ranging from 10^{-8} –10 bar with 100 layers spaced uniformly in log-pressure under hydrostatic equilibrium. We adopt 10^{-8} bar as the lowest atmosphere pressure (versus the 10^{-6} bar limit implied by N. Madhusudhan et al. 2023) to reduce the risk of strong bands saturating the atmosphere. We solve hydrostatic equilibrium using a boundary condition where the freely retrieved

reference pressure corresponds to the radial distance where $r = 2.61 R_{\oplus}$ (the radius from B. Benneke et al. 2019). The atmospheric P-T profile is freely retrieved following the prescription in N. Madhusudhan & S. Seager (2009). We consider the same potential molecular species as N. Madhusudhan et al. (2023), which includes the major O-, C-, and N-bearing species expected at K2-18 b’s equilibrium temperature (255 K; B. Benneke et al. 2019), H₂O, CH₄, NH₃, HCN, CO, and CO₂; as well as several disequilibrium species proposed as biosignatures in cool sub-Neptunes (N. Madhusudhan et al. 2021), C₂H₆S (DMS), CS₂, CH₃Cl, OCS, and N₂O. We fit the volume mixing ratios of these gases independently, with their number densities then determined via the ideal gas law. Given the large uncertainties on K2-18 b’s mass ($8.63 \pm 1.35 M_{\oplus}$; R. Cloutier et al. 2019), we prescribe the planet’s mass as a free parameter using a Gaussian prior (N. Madhusudhan et al. 2023 fixed K2-18 b’s mass), allowing mass uncertainties to propagate into the retrieved atmospheric composition (N. E. Batalha et al. 2019a). Finally, we consider aerosols following the inhomogeneous cloud and haze parameterization of R. J. MacDonald & N. Madhusudhan (2017).

We calculate model spectra by solving the equation of radiative transfer in a cylindrical coordinate system for 100 incident stellar rays attenuated by atmospheric opacity along the line of sight. The POSEIDON opacity database recently underwent a major update in POSEIDON v1.2 (E. Mullens et al. 2024), including a state-of-the-art new ExoMol CH₄ line list (S. N. Yurchenko et al. 2024). Our POSEIDON opacities are precomputed at a high resolution ($\Delta\nu = 0.01 \text{ cm}^{-1}$, equivalent to $R = \lambda/\Delta\lambda = 10^6$ at $1 \mu\text{m}$) for a grid of temperatures and pressures using the open-source Python package Cthulhu²⁷ (A. Agrawal & R. MacDonald 2024). Our K2-18 b retrievals use opacities derived from the following line lists/measured cross sections: H₂O (POKAZATEL; O. L. Polyansky et al. 2018), CH₄ (MM; S. N. Yurchenko et al. 2024), NH₃ (CoYuTe; P. A. Coles et al. 2019), HCN (Harris; R. J. Barber et al. 2014), CO (Li2015; G. Li et al. 2015), CO₂ (UCL-4000; S. N. Yurchenko et al. 2020), C₂H₆S (DMS; HITRAN measured xsec; S. W. Sharpe et al. 2004) CS₂ (HITRAN-2020; I. E. Gordon et al. 2022), CH₃Cl (HITRAN-2020; I. E. Gordon et al. 2022), OCS (OYT8; A. Owens et al. 2024), and N₂O (HITEMP; R. J. Hargreaves et al. 2019). We further include continuum collision-induced absorption from H₂-H₂ and H₂-He pairs (T. Karman et al. 2019) and Rayleigh scattering for all gases (R. J. MacDonald & N. K. Lewis 2022). We calculate model transmission spectra using opacity sampling on a wavelength grid from 0.58 to $5.3 \mu\text{m}$ at a spectral resolution of $R = 100,000$. We adopt this exceptionally high spectral resolution for all our POSEIDON retrievals for three reasons: (i) to ensure negligible opacity sampling errors when interpreting pixel-level data; (ii) to consistently sample cross sections for every data combination (even at lower resolutions); and (iii) to reduce the risk of opacity accuracy differences between our work and N. Madhusudhan et al. (2023; since the latter did not state their assumed sampling resolution). While our $R = 100,000$ models impose a greater computational burden than is strictly necessary for retrievals of JWST data,²⁸ this choice ensures high accuracy in our retrieved atmospheric properties and reduces the

²⁵ In this case, the “model” describes the structure of the planet’s atmosphere in the context of an observed transmission spectrum, with different potential “assumptions/settings” made for it. These include which potential aspects of the model to consider, the preferred parameterization of each component of the model, of which parameters are fixed as well as the values they are fixed to, and the priors on each free parameter in the model. On the other hand, the “retrieval framework” describes how the model is fit to the data. This includes the way that the planet’s atmosphere and transmission through it are numerically modeled, the data used in these numerical models, and the code used to explore the parameter space. The combination of these is what we describe as the overall “retrieval configuration.”

²⁶ <https://github.com/MartianColonist/POSEIDON>

²⁷ <https://github.com/MartianColonist/Cthulhu>

²⁸ Through the JWST Transiting Exoplanet Early Release Science Program, we found a useful heuristic that $R_{\text{model}} = 20,000$ suffices for $R_{\text{data}} = 100$, and $R_{\text{model}} = 60,000$ suffices for $R_{\text{data}} \sim 3000$ (pixel-level).

Table 3
Atmospheric Retrieval Configuration Comparison

Model Setting	POSEIDON (This Work)	BeAR (This Work)	AURA (N. Madhusudhan et al. 2023)	Reference
System Properties				
Stellar Radius	$0.4445 R_{\odot}$	$0.41 R_{\odot}$	$0.4445 R_{\odot}$	$0.41 R_{\odot}$: B. Benneke et al. (2017); $0.4445 R_{\odot}$: B. Benneke et al. (2019)
Planetary Radius	Fit (from $\log_{10} P_{\text{ref}}$)	Fit (free parameter)	Fit (from $\log_{10} P_{\text{ref}}$)	
Planetary Mass	Fit (free parameter)	Fit (from $\log_{10} g_{\text{p}}$)	$8.63 M_{\oplus}$	
Planetary Gravity	Fit (from M_{p})	Fit (free parameter)	12.4 ms^{-2}	
Atmospheric Model				
Pressure Grid	10^{-8} –10 bar	10^{-8} –10 bar	?	
Number of Layers	100	200	?	
Background Gas	$\text{H}_2 + \text{He}$	$\text{H}_2 + \text{He}$	$\text{H}_2 + \text{He}$	
Molecules Included	$\text{H}_2\text{O}, \text{CH}_4, \text{NH}_3, \text{HCN}, \text{CO}, \text{CO}_2, \text{DMS}, \text{CS}_2, \text{CH}_3\text{Cl}, \text{OCS}, \text{N}_2\text{O}$	$\text{H}_2\text{O}, \text{CH}_4, \text{NH}_3, \text{HCN}, \text{CO}, \text{CO}_2, \text{DMS}, \text{CS}_2, \text{CH}_3\text{Cl}, \text{OCS}$	$\text{H}_2\text{O}, \text{CH}_4, \text{NH}_3, \text{HCN}, \text{CO}, \text{CO}_2, \text{DMS}, \text{CS}_2, \text{CH}_3\text{Cl}, \text{OCS}, \text{N}_2\text{O}$	
He/H ₂ Ratio	0.17	0.17	?	
P-T Profile Treatment	Free Profile	Isotherm	Free Profile	Free Profile: N. Madhusudhan & S. Seager (2009)
		...		
Cloud Treatment	Patchy Clouds + Haze	Opaque Cloud	Patchy Clouds + Haze	Patchy Clouds + Haze: R. J. MacDonald & N. Madhusudhan (2017)
		...		
Hydrostatic Boundary Condition	$\log_{10} P_{\text{ref}}$	$R_{\text{p,ref}}$	$\log_{10} P_{\text{ref}}$	
Spectral Model				
Model Wavelength Grid	0.58 – $5.3 \mu\text{m}$	0.5996 – $5.176 \mu\text{m}$	?	
Native Opacity Resolution	0.01 cm^{-1}	0.01 cm^{-1}	?	
Opacity Sampling Resolution	$R = 100,000$	$R = 60,000$	?	
CH ₄ Line List	MM	YT10 to 10	Ames-2016	MM: S. N. Yurchenko et al. (2024); YT10 to 10: S. N. Yurchenko & J. Tennyson (2014); Ames-2016: X. Huang et al. (2013, 2017)
CO ₂ Line List	UCL-4000	UCL-4000	HITEMP	(S. N. Yurchenko et al. 2020), (R. J. Hargreaves et al. 2020)
DMS Cross Section (citation)	HITRAN	HITRAN	HITRAN	S. W. Sharpe et al. (2004)
Stellar Contamination Model				
Stellar Heterogeneity Treatment	One-het	Two-het	None ^a	
Stellar Heterogeneity Parameters	$T_{\text{phot}}, T_{\text{het}}, f_{\text{het}}$	$T_{\text{phot}}, \Delta T_{\text{spot}}, \Delta T_{\text{fac}}, f_{\text{spot}}, f_{\text{fac}}$	... ^a	
Stellar Model Grid	PHOENIX	blackbody	... ^a	
Stellar Metallicity ([Fe/H])	-0.08	...	... ^a	
Stellar Surface Gravity ($\log_{10} g_*$)	4.93 (cgs)	...	... ^a	
(citation)		...	... ^a	(H. M. Tabernero et al. 2024)
Retrieval Priors and Settings				
Planetary Mass (M_{p})	$\mathcal{N}(8.63, 1.35^2) M_{\oplus}$	...	$8.63 M_{\oplus}$ (fixed)	
Planetary Gravity ($\log_{10} g_{\text{p}}$)	...	$\mathcal{N}(3.19, 0.17^2)$ (cgs)	3.09 (fixed; cgs)	
Atmospheric Temperature (T_{ref})	$\mathcal{U}(0, 800) \text{ K}$	$\mathcal{U}(50, 1000) \text{ K}$	$\mathcal{U}(0, 800) \text{ K}$	
P-T Profile Curvature 1 (α_1)	$\mathcal{U}(0.02, 2.00) \text{ K}^{-\frac{1}{2}}$	...	$\mathcal{U}(0.02, 2.00) \text{ K}^{-\frac{1}{2}}$	
P-T Profile Curvature 2 (α_2)	$\mathcal{U}(0.02, 2.00) \text{ K}^{-\frac{1}{2}}$	...	$\mathcal{U}(0.02, 2.00) \text{ K}^{-\frac{1}{2}}$	
P-T Profile Region 1 ($\log_{10}(P_1 / \text{bar})$)	$\mathcal{U}(-8, 1)$	...	$\mathcal{U}(-6, 0)$	
P-T Profile Region 2 ($\log_{10}(P_2 / \text{bar})$)	$\mathcal{U}(-8, 1)$	...	$\mathcal{U}(-6, 0)$	

Table 3
(Continued)

Model Setting	POSEIDON (This Work)	BeAR (This Work)	AURA (N. Madhusudhan et al. 2023)	Reference
P-T Profile Region 3 ($\log_{10}(P_3 / \text{bar})$)	$\mathcal{U}(-2, 1)$	...	$\mathcal{U}(-2, 0)$	
Molecule Volume Mixing Ratio ($\log_{10} X_i$)	$\mathcal{U}(-12, -0.3)$	$\mathcal{U}(-12, -0.522)$	$\mathcal{U}(-12, -0.3)$	
Reference Radius ($R_{p,\text{ref}}$)	$2.61 R_{\oplus}$ (fixed)	$\mathcal{U}(2.0, 2.8) R_{\oplus}$	$2.61 R_{\oplus}$ (fixed)	
Reference Pressure ($\log_{10}(P_{\text{ref}} / \text{bar})$)	$\mathcal{U}(-6, 0)$	1.0 (fixed)	$\mathcal{U}(-6, 0)$	
Cloud Pressure ($\log_{10}(P_{\text{cloud}} / \text{bar})$)	$\mathcal{U}(-8, 1)$	$\mathcal{U}(-7, 0)$	$\mathcal{U}(-6, 1)$	
Cloud Fraction ($\bar{\phi}$)	$\mathcal{U}(0, 1)$	...	$\mathcal{U}(0, 1)$	
Haze Rayleigh Enhancement ($\log_{10} a$)	$\mathcal{U}(-4, 10)$	...	$\mathcal{U}(-4, 10)$	
Haze Slope (γ)	$\mathcal{U}(-20, 2)$	...	$\mathcal{U}(-20, 2)$	
Data Offset 1 ($\delta_{\text{rel},1}$)	$\mathcal{U}(-100, 100)$ ppm	$\mathcal{U}(-100, 100)$ ppm	$\mathcal{U}(-100, 100)$ ppm	
Data Offset 2 ($\delta_{\text{rel},2}$)	$\mathcal{U}(-100, 100)$ ppm	$\mathcal{U}(-100, 100)$ ppm	$\mathcal{U}(-100, 100)$ ppm	
T_{phot}	$\mathcal{N}(3563, 25^2)$ K	$\mathcal{N}(3457, 39^2)$ K	... ^a	
T_{het}	$\mathcal{U}(2300, 4275.6)$ K	...	... ^a	
f_{het}	$\mathcal{U}(0, 0.5)$	...	... ^a	
ΔT_{spot}	...	$\mathcal{U}(-1500, 0)$ K	...	
ΔT_{fac}	...	$\mathcal{U}(0, 1000)$ K	...	
f_{spot}	...	$\mathcal{U}(0, 0.5)$	...	
f_{fac}	...	$\mathcal{U}(0, 0.5)$	...	
MultiNest Live Points	1000	1000	?	
Retrieval Code Availability	POSEIDON GitHub	BeAR GitHub	Not Public	

Notes. “?” indicates model settings not specified in N. Madhusudhan et al. (2023). Line list references for other molecules included in each retrieval code are provided in Sections 4.1.1 and 4.1.2.

^a All the results presented in N. Madhusudhan et al. (2023) did not include stellar contamination, but the study stated they ran additional (not shown) retrievals considering a one-heterogeneity model. Gaussian priors are summarized as $\mathcal{N}(\mu, \sigma^2)$, where μ and σ are the mean and standard deviation, respectively. T_{ref} refers to the top-of-atmosphere temperature for POSEIDON and N. Madhusudhan et al. (2023), while for BeAR it represents the isothermal temperature.

risk of an insufficiently high opacity resolution inconsistent with the methodology of N. Madhusudhan et al. (2023).

Our POSEIDON retrievals additionally account for stellar contamination from unocculted active regions, commonly known as the transit light source effect (B. V. Rackham et al. 2018). We employ a one-heterogeneity model (e.g., A. D. Rathcke et al. 2021; O. Lim et al. 2023; M. Fournier-Tondreau et al. 2024), defined by the stellar photosphere temperature, the heterogeneity temperature (corresponding to spots/faculae if the heterogeneity is cooler/warmer than the photosphere), and the heterogeneity covering fraction. The stellar contamination contribution is calculated by interpolating PHOENIX models (T. O. Husser et al. 2013) using the PyMSG package (R. Townsend & A. Lopez 2023). For fixed stellar properties, we adopt the stellar radius from B. Benneke et al. (2019) ($R_* = 0.4445 R_{\odot}$) and the stellar surface gravity, metallicity, and effective temperature from H. M. Tabernero et al. (2024) ($[\text{Fe}/\text{H}] = -0.08$, $\log_{10} g_* = 4.93$, $T_{\text{eff}} = 3563$ K). These updated K2-18 stellar properties were published after N. Madhusudhan et al. (2023), so we adopt them for a more accurate characterization of the transit light source effect.

Our reference POSEIDON retrieval model has 28 free parameters, with priors as given in Table 3. We generally adopt the same priors as in N. Madhusudhan et al. (2023) for consistency, with a few exceptions: (i) we adjust the pressure parameters describing the P-T profile and cloud-top pressure to span our wider atmospheric pressure range; (ii) we apply a Gaussian prior on the planetary mass, using the standard deviation from R. Cloutier et al. (2019); and (iii) we use newly updated constraints on the host star’s effective temperature from H. M. Tabernero et al. (2024) to set the Gaussian prior on T_{phot} . We allow for two free offsets, one between the NIRISS SOSS and NIRSpec G395H NRS1 data and one between the NIRISS and NIRSpec G395H NRS2 data, as both N. Madhusudhan et al. (2023) and our Appendix A demonstrate the need for at least one free offset. Alongside the reference model, we additionally ran nested retrievals, with CH_4 , CO_2 , and DMS removed in turn, to calculate the Bayes factors (and hence the equivalent detection significances) for these molecules. Therefore, we ran 4 retrievals for each combination of NIRISS SOSS (6 variants, 2 reduction pipelines each with 3 resolutions) and NIRSpec G395H data (10 variants, 2

reduction pipelines each with 4 resolutions, and 2 reduction pipelines with 1 resolution), for a total of 240 retrievals spanning the 60 different data combinations. We use `PyMultiNest` (F. Feroz et al. 2009; J. Buchner et al. 2014) with 1000 live points to explore the parameter space. Each retrieval took 2–5 days on 24 CPUs—with longer runtimes for pixel-level data variants—for a total compute time of $\gtrsim 50$ CPU years.

4.1.2. BEAR

To validate our POSEIDON results, we also perform a second set of retrievals using the open-source Bern Atmospheric Retrieval code (BeAR²⁹). BeAR is an updated version of the retrieval code formerly known as `Helios-r2`, which was originally developed for retrievals on brown dwarf spectra (D. Kitzmann et al. 2020). BeAR has been adapted for use on exoplanet spectra (C. Fisher et al. 2024) and has the capabilities to analyze transmission, emission, and phase curve spectra. BeAR uses opacity sampling of line-by-line cross sections to calculate model spectra, coupled to `PyMultiNest` (F. Feroz et al. 2009; J. Buchner et al. 2014) for statistical exploration.

Our BeAR retrievals use a similar retrieval configuration to that described above, but here, we highlight differences in the model setup. First, we assume for simplicity an isothermal temperature throughout the atmosphere (with a uniform prior of 50–1000 K). The atmosphere is divided into 200 equal layers in log-pressure, ranging from 10 to 10^{-8} bar. Second, we fit for the 10 bar planet radius, rather, a reference pressure, as our hydrostatic boundary condition (uniform from 2.0 to $2.8R_{\oplus}$). We retrieve the planetary log-surface gravity as a proxy for the mass uncertainties (Gaussian with mean 3.19 and standard deviation 0.17, both in cgs units). Third, we restrict our aerosol model to an opaque cloud deck (with a log-uniform pressure prior of $1.0\text{--}10^{-7}$ bar). Fourth, we sample the opacities at a slightly lower resolution of $R = 60,000$, due to GPU memory limitations. We include the following molecules in our BeAR retrievals, with their associated line lists: H_2O (O. L. Polyansky et al. 2018), CH_4 (S. N. Yurchenko & J. Tennyson 2014), NH_3 (P. A. Coles et al. 2019), HCN (G. J. Harris et al. 2006; R. J. Barber et al. 2014), CO (G. Li et al. 2015), CO_2 (S. N. Yurchenko et al. 2020), DMS (S. W. Sharpe et al. 2004), CS_2 (S. W. Sharpe et al. 2004), CH_3Cl (A. Owens et al. 2018), and OCS (J. S. Wilzewski et al. 2016). We also include collision-induced absorption from $\text{H}_2\text{--H}_2$ (M. Abel et al. 2011) and $\text{H}_2\text{--He}$ (M. Abel et al. 2012), alongside Rayleigh scattering from H_2 (M. S. Vardya 1962). Our retrievals assume free chemistry, meaning the molecular abundances (aside from H_2 and He) are retrieved independently, making no assumptions about chemical processes. The abundances are drawn from log-uniform prior distributions, ranging from volume mixing ratios of 10^{-12} –0.3. The remaining atmosphere is filled with H_2 and He , assuming a solar He/H_2 ratio of 0.17 (M. Asplund et al. 2009). Finally, we include the effects of stellar contamination using a two-heterogeneity model with five free parameters: the stellar photosphere temperature (following a Gaussian prior with a mean of 3457 K and a standard deviation of 39 K—corresponding to the effective temperature values from R. Cloutier et al. 2017); the faculae fraction (uniform from 0–0.5) and temperature difference relative

to the photosphere (uniform from 0–1000 K); and equivalent parameters for spots (with a temperature difference uniform from $-1500\text{--}0$ K). The stellar contamination is calculated assuming blackbodies for the photosphere and active regions. Alongside the atmospheric and stellar parameters described above, our BeAR retrievals account for possible offsets between different filters and instruments (e.g., K. H. Yip et al. 2021; L. Alderson et al. 2023) by fitting two offsets—one between NIRISS SOSS and NIRSpec G395H, and one between the NRS1 and NRS2 filters of NIRSpec G395H (both uniform from $-100\text{--}100$ ppm). Table 3 provides a full comparison between our BeAR and POSEIDON retrieval setups, alongside that used in N. Madhusudhan et al. (2023).

We applied our BeAR retrievals to a subset of our reanalyzed NIRISS SOSS and NIRSpec G395H data. Specifically, we focus on low-resolution data combinations using a single reduction code ($R \approx 100$ FIREFLY and exoTDRF data for both NIRISS and NIRSpec) and the highest-resolution pixel-level data produced by a common reduction code. Therefore, we considered four data combinations with BeAR. For each data combination, we ran four retrievals, a reference model and three nested models, as above, to calculate detection significances of CH_4 , CO_2 , and DMS .

4.2. Retrieval Results

4.2.1. Molecular Detection Significances

Our atmospheric retrieval analysis confirms a strong detection of CH_4 in K2-18 b’s atmosphere. Figure 5 shows a grid of 180 Bayesian model comparisons from POSEIDON, colored by the detection significance corresponding to each Bayes factor, spanning every data combination produced in this study. We see that CH_4 is detected at $>2.7\sigma$ for all data combinations, with most combinations favoring CH_4 at $>4\sigma$. Considering our CH_4 significances as independent draws from the distribution of possible data-level choices, we find a detection significance of $4.2^{+0.7}_{-0.9}\sigma$. While slightly lower than the 5σ detection reported in N. Madhusudhan et al. (2023), we consider these results a strong confirmation that CH_4 is robustly present in K2-18 b’s atmosphere.

Figure 5 additionally contains a wealth of information pertaining to the sensitivity of CH_4 detections to both data resolution and reduction code. First, we see that increasing the NIRISS SOSS data resolution from $R \approx 25$ to $R \approx 100$ leads to a clear increase in the CH_4 detection significances from $\sim 3.3\sigma$ to $\sim 4.5\sigma$. This behavior is not surprising, since $R \approx 25$ NIRISS data make it harder for retrievals to uniquely resolve CH_4 bands from other molecules with similar absorption features (especially H_2O), resulting in lower Bayes factors and hence lower significances. We generally find higher CH_4 significances for the NIRISS data with exoTDRF (mean significances of 3.5σ , 5.0σ , and 4.8σ for the $R \approx 25$, $R \approx 100$, and pixel-level data, respectively) compared with FIREFLY (3.1σ , 3.8σ , and 4.4σ , for the same data resolutions). We note that the pixel-level exoTDRF results are in good agreement with the $R \approx 100$ exoTDRF results, while the $R \approx 100$ FIREFLY results lead to notably lower detection significances. A potential reason for the discrepancies at $R \approx 100$, but not at pixel-level, could be differences in how each reduction code handles systematics and limb darkening, which manifests more at lower resolutions than at pixel-level.

We do not find any reliable or significant evidence for CO_2 or DMS in the observed transmission spectrum of K2-18 b.

²⁹ <https://newstrangeworlds.github.io/BeAR>

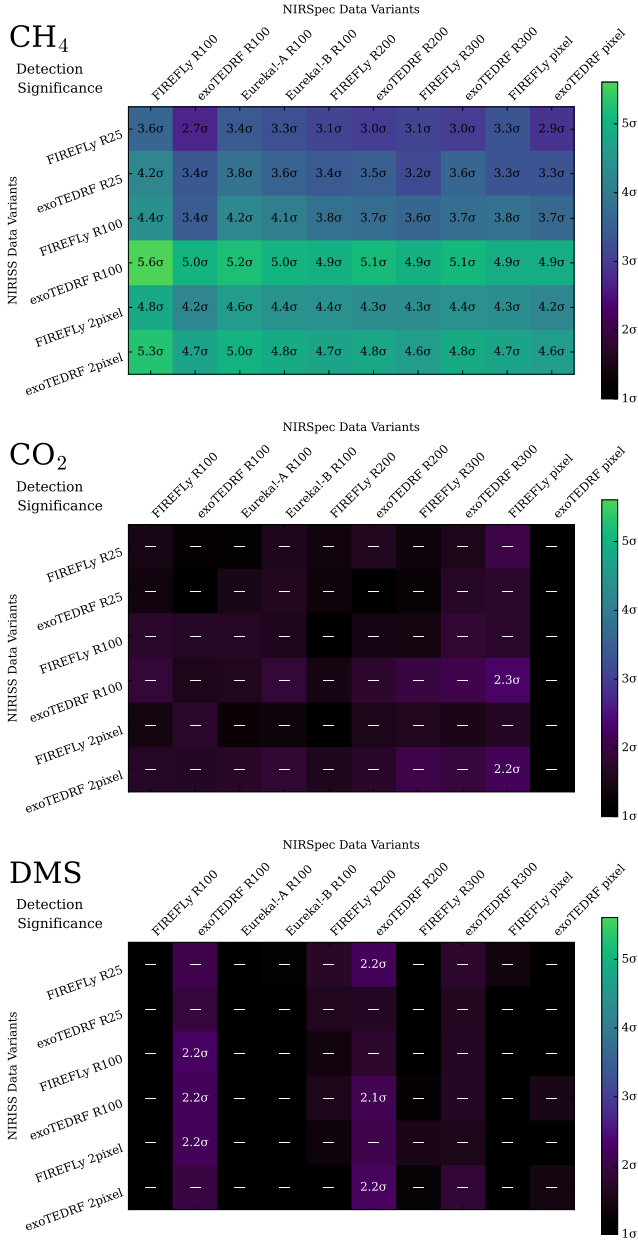


Figure 5. Detection significances of CH₄, CO₂, and DMS from POSEIDON retrievals of K2-18 b’s JWST NIRISS SOSS + NIRSpect G395H transmission spectrum. Each grid cell represents a Bayesian model comparison between two retrievals, one with and one without a molecule, for a specific NIRISS SOSS and NIRSpect G395H data variant pair. The data variants are ordered with increasing spectral resolution to the right (for NIRSpect) and down (for NIRISS). The cells are colored and annotated according to the detection significance equivalent to the calculated Bayes factor (e.g., R. Trotta 2008; B. Benneke & S. Seager 2013). Bayes factors < 3 (equivalent to < 2.1σ), corresponding to no evidence on the Jeffreys’ scale, are denoted by “—.” Our retrievals demonstrate strong evidence for CH₄, but find no reliable statistical evidence for CO₂ or DMS in K2-18 b’s atmosphere.

Figure 5 demonstrates that almost every combination of NIRISS and NIRSpect data results in no evidence for CO₂ or DMS (Bayes factors < 3, denoted by “—” in Figure 5), are “not worth mentioning” according to the Jeffreys’ scale of Bayesian model comparison—see R. Trotta 2008). Even choosing the data combination with the maximum Bayes factors results in 4.2 for CO₂ (2.3σ) and 3.9 for DMS (2.2σ), which are considered as “weak evidence at best” on the Jeffreys’ scale. However, one can see in Figure 5 that it is only the coupling

between the FIREFLY pixel-level NIRSpect data and the exoTEDRF $R \approx 100$ or 2 pixel NIRISS data that result in any marginal evidence for CO₂. Similarly, even the marginal evidence for DMS is only present for $R \approx 100$ or $R \approx 200$ NIRSpect data from exoTEDRF and disappears for higher data resolutions. Therefore, any marginal evidence for CO₂ or DMS depends on either the choice of a specific data reduction code or insufficient data resolution. In Appendix A, we demonstrate that if we run retrievals on the data from N. Madhusudhan et al. (2023), then we recover a maximum significance of 2.5σ for CO₂—this shows that the claimed CO₂ detection arises from the specific data treatments in that study. Our results demonstrate that, contrary to the findings of N. Madhusudhan et al. (2023), there is no robust detection of CO₂ from K2-18 b’s initial NIRISS SOSS and NIRSpect G395H transmission spectra, nor are there even potential signs of DMS. Figure 5 offers a clear visual representation of the difference between a robust detection and nondetections.

Our BeAR retrievals similarly find strong detections of CH₄ and nondetections of CO₂ and DMS. For the FIREFLY reductions, for CH₄, we obtain detection significances of 3.5σ and 4.5σ for the pixel-level and $R \approx 100$ retrievals, respectively. For CO₂ and DMS, the Bayes factor is < 3 for both resolutions, corresponding to no evidence. The results are extremely similar for the exoTEDRF reductions, for which we obtain 3.8σ and 4.4σ for CH₄ for the pixel-level and $R \approx 100$ retrievals, respectively, and Bayes factors < 3 for CO₂ and DMS. These results are in good agreement with our retrievals using POSEIDON (Figure 5), indicating that our conclusions are robust to the choice of retrieval code.

Having established that only CH₄ is evidenced by K2-18 b’s JWST NIRISS SOSS and NIRSpect G395H transmission spectrum, we now turn to present abundance constraints consistent with these data.

4.2.2. Molecular Abundance Constraints

We first consider the sensitivity of the atmospheric abundances of CH₄, CO₂, and DMS to the choice of data reduction code, data resolution, and retrieval code. Figure 6 provides a pairwise retrieval comparison of how these three factors can affect the inferred atmospheric composition of K2-18 b. First, we highlight a comparison between POSEIDON retrievals of the lowest-resolution FIREFLY and exoTEDRF data ($R \approx 25$ NIRISS; $R \approx 100$ NIRSpect). We selected these combinations to allow a clear demonstration of why the exoTEDRF NIRSpect $R \approx 100$ data lead to weak evidence of DMS (per Figure 5), but FIREFLY does not. We see in Figure 6 (top panel) that several exoTEDRF data points near 3.5 and 4.9 μm have greater transit depths than FIREFLY, which coincide with opacity maxima for DMS. These transit depth differences result in a peak in the DMS mixing ratio posterior near $\log_{10} X_{\text{DMS}} \sim -4$ for exoTEDRF, but an upper limit for FIREFLY. The nonrobustness of this DMS inference can be seen in the second pairwise comparison in Figure 6 (middle panel), where we show that our retrieval of the highest-resolution exoTEDRF data (2 pixel for NIRISS; pixel-level for NIRSpect) no longer favors DMS. We note that the higher model transit depths for the pixel-level retrieval arise from a higher mean transit depth in the exoTEDRF NIRISS data compared to the $R \approx 25$ data variant (resulting in smaller offsets between the NIRISS and NIRSpect data for the pixel-level retrieval). Finally, we demonstrate the excellent

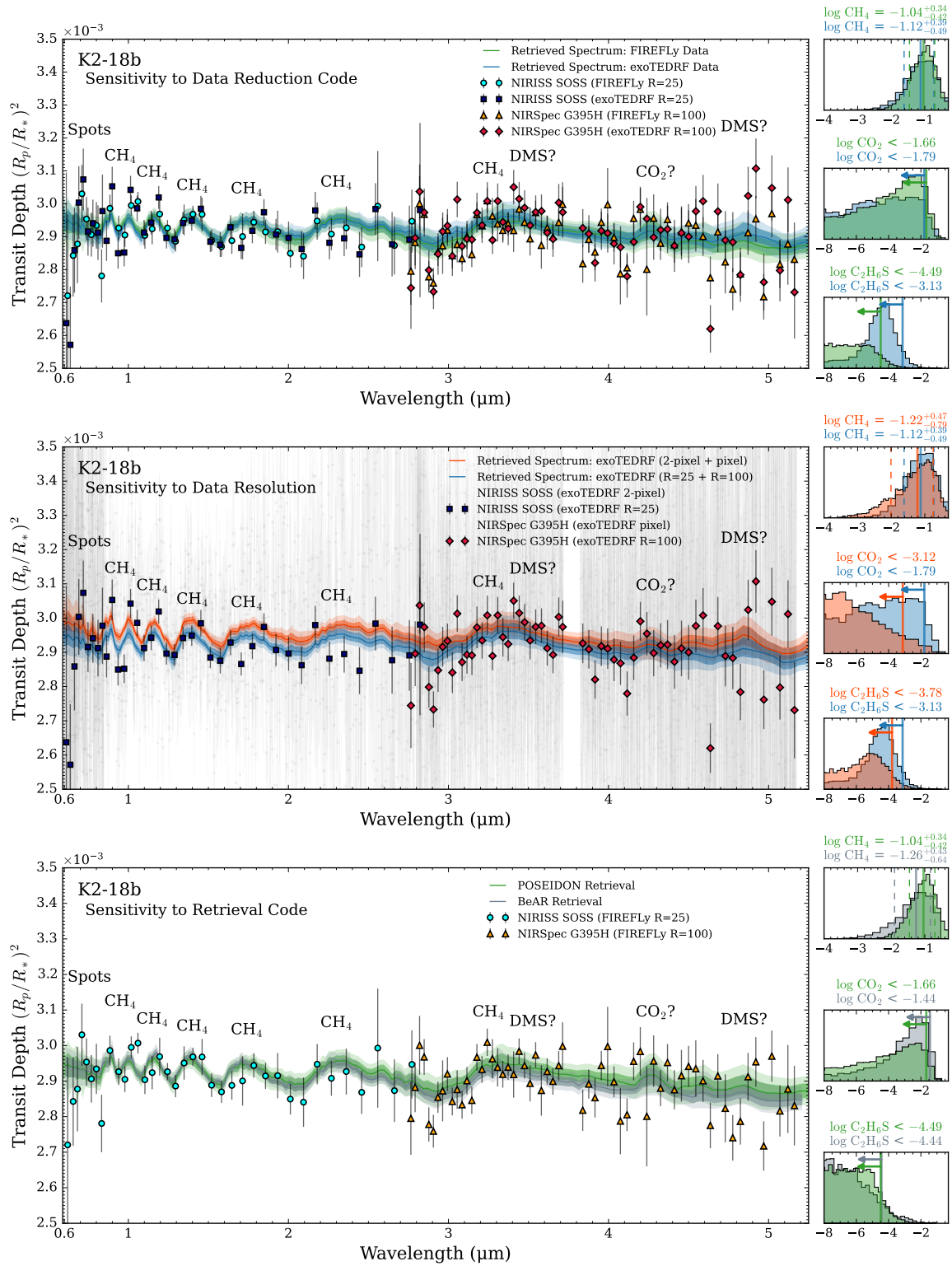


Figure 6. Sensitivity of K2-18 b's atmospheric composition to different data treatments and retrieval configurations. Top: comparison between POSEIDON retrievals of data reduced by the FIREFLY and exoTDRF pipelines. Middle: comparison between POSEIDON retrievals of low-resolution (NIRISS SOSS, $R \approx 25$; NIRSpec G395H, $R \approx 100$) and high-resolution (pixel-level) exoTDRF data. Bottom: comparison between the retrieval codes POSEIDON and BeAR on FIREFLY data. The retrieved model transmission spectra, plotted at $R \approx 100$ for clarity, show the median (solid lines) and $\pm 1\sigma$ and $\pm 2\sigma$ credible regions (contours) corresponding to the data set(s) shown in each panel. The median retrieved offsets have been applied to the NIRSpec G395H NRS1 and NRS2 data sets. Right: the retrieved CH_4 , CO_2 , and DMS abundances corresponding to each retrieval. Our retrievals find broadly consistent CH_4 abundances, indicating a robust detection (the median and $\pm 1\sigma$ abundance constraints are represented by solid and dashed lines, respectively). However, any suggestions of CO_2 and DMS depend on the specific data reduction and data resolution, indicating unreliable inferences (hence the 95% upper limits, represented by arrows).

agreement between POSEIDON and BeAR in Figure 6 (lower panel). This shows that our main results (CH_4 abundance constraints, upper limits on the CO_2 and DMS abundances) are robust to the different atmospheric modeling and retrieval configuration choices made by POSEIDON and BeAR (Table 3). We additionally provide the full corner plots from our POSEIDON and BeAR retrieval grids as supplementary material on Zenodo: doi:[10.5281/zenodo.14735688](https://doi.org/10.5281/zenodo.14735688).

CH_4 is the only molecule with an abundance lower limit across all data combinations (Figure 7), which indicates a robust detection. To account for the spread in abundances across different data treatments, we consider a conservative range spanning the lowest -1σ and highest $+1\sigma$ across all the CH_4 posteriors—similar to the approach by D. Powell et al. (2024) for SO_2 in WASP-39b. We find a 1σ range of $\log_{10}\text{CH}_4 = -2.14$ to -0.53 and a 2σ range of $\log_{10}\text{CH}_4 = -2.86$ to -0.35 , with the most probable CH_4 abundance near $\sim 10\%$. Our retrieved CH_4 is consistent with the value reported in N. Madhusudhan et al. (2023) using a single data reduction code ($\log_{10}\text{CH}_4 = -1.74^{+0.59}_{-0.69}$; see also Appendix A). The high abundance of CH_4 in K2-18 b’s atmosphere is consistent with a metal-enriched mini-Neptune (as discussed in Section 5.1) and is similar to the deep abundance of CH_4 on Neptune ($4\% \pm 1\%$; E. Karkoschka & M. G. Tomasko 2011). This high CH_4 abundance also raises the background mean molecular weight of K2-18 b’s atmosphere to $\mu_{\text{atm}} = 2.46\text{--}7.64$ amu.

Abundance inferences for CO_2 and DMS depend critically on the data combination adopted, with some combinations showing suggestive peaks but most showing broad posteriors with no evidence supporting their presence (Figure 7). We therefore report 2σ upper limits on the abundances of CO_2 and DMS in K2-18 b’s atmosphere, since neither of these molecules are robustly detected. Considering 2σ extrema upper limits, as for CH_4 above, we find $\log_{10}\text{CO}_2 < -1.02$ and $\log_{10}\text{DMS} < -2.51$. However, these limits are pulled to higher abundances by the few data combinations exhibiting posterior peaks (Figure 7). Therefore, we consider the median 2σ upper limits to provide a more representative statistic for the abundances of molecules that are not robustly detected: $\log_{10}\text{CO}_2 < -1.58$ and $\log_{10}\text{DMS} < -3.58$. We stress that these upper limits do not imply that either molecule is absent at *any* atmospheric abundance, as abundances lower than our 95% limits are still compatible with the present JWST transmission spectrum. Indeed, some mini-Neptune scenarios predict CO_2 abundances of $\log_{10}\text{CO}_2 \sim -3$ (Section 5.1), consistent with our upper limit.

We additionally report upper limits on several gases important for understanding the atmospheres and interiors of sub-Neptunes. Water and ammonia are depleted from the gas phase (95% upper limits of $\log_{10}\text{H}_2\text{O} < -2.44$ and $\log_{10}\text{NH}_3 < -4.70$), which can be explained by cloud condensation (R. Hu et al. 2021; N. Madhusudhan et al. 2023) and dissolution into a deep magma ocean (O. Shorttle et al. 2024), respectively. Similarly, we find an upper limit of $\log_{10}\text{CO} < -2.04$. We show in Section 5 that these nondetections do not require a surface liquid ocean on K2-18 b. Alongside our nondetection of DMS, we find upper limits on four other gases proposed as biosignatures (e.g., A. Segura et al. 2005; S. Seager et al. 2013; E. W. Schwieterman et al. 2018; N. Madhusudhan et al. 2021): $\log_{10}\text{CS}_2 < -2.63$, $\log_{10}\text{CH}_3\text{Cl} < -1.81$, $\log_{10}\text{OCS} < -2.77$, and $\log_{10}\text{N}_2\text{O} < -3.56$. Therefore, there is no evidence of biosignatures in the JWST

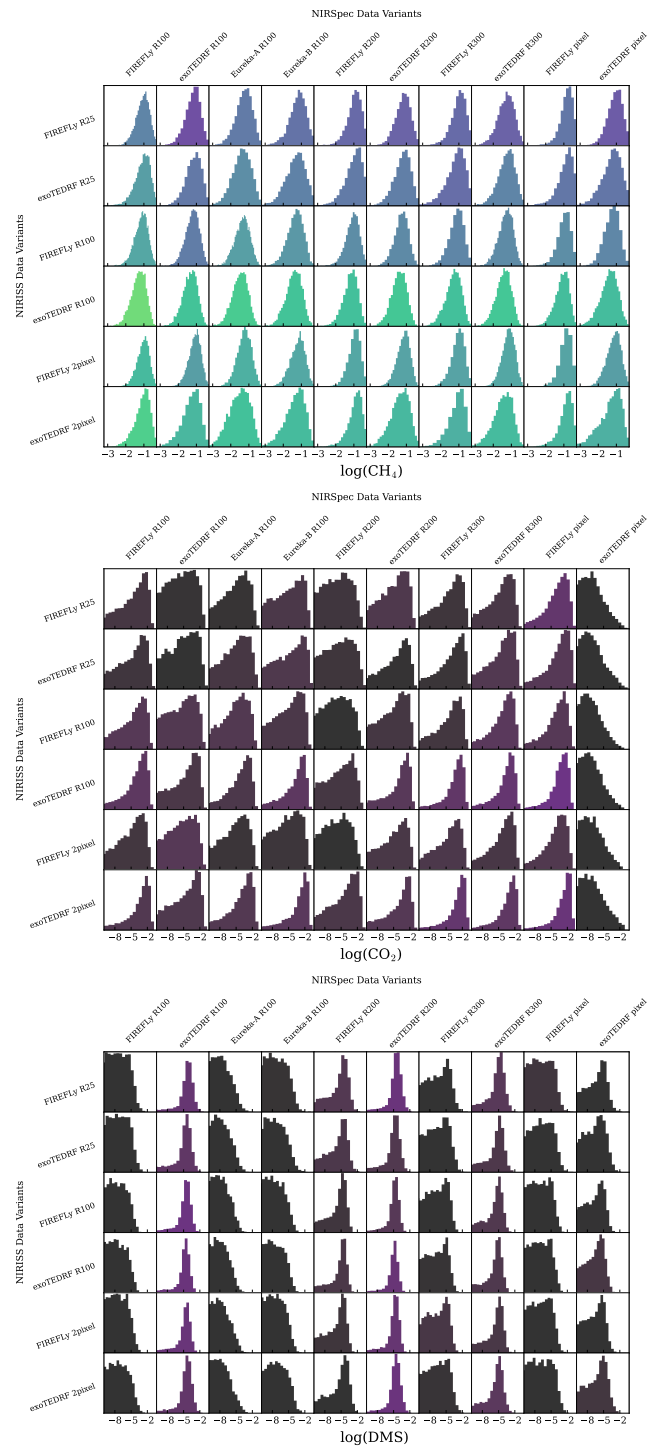


Figure 7. Grid of histograms showing the posteriors for CH_4 , CO_2 , and DMS for each of our retrievals. The arrangement and color map are the same as in Figure 5, with the color of the histogram representing the detection significance. While the CH_4 posteriors exhibit little variance across reduction choices, there are major differences for CO_2 and DMS that are indicative of a nondetection.

NIRISS SOSS + NIRSpec G395H transmission spectrum of K2-18 b.

4.2.3. Atmospheric Temperature

K2-18 b’s transmission spectrum is consistent with an isothermal upper-atmosphere temperature structure. While our

POSEIDON retrievals allow for temperature gradients, our top-of-atmosphere temperature ($T_{10 \text{ mbar}} = 98\text{--}339 \text{ K}$) is consistent with the temperature at significantly higher pressures (e.g., $T_{10 \text{ mbar}} = 120\text{--}358 \text{ K}$). Our temperature constraints are consistent with N. Madhusudhan et al. (2023), who reported $T_{10 \text{ mbar}} = 242^{+57}_{-79} \text{ K}$. Since low-resolution transmission spectra only probe atmospheric pressures from $\sim 10^{-5}\text{--}10^{-1} \text{ bar}$, we note that the lack of any detected temperature gradient does not preclude steep adiabatic temperature gradients at deeper pressures.

4.2.4. Aerosols

We do not detect any evidence of aerosols in K2-18 b's atmosphere. Our retrievals offer no constraints (either upper or lower limits) on the haze parameters or the terminator cloud fraction. N. Madhusudhan et al. (2023) similarly reported a nondetection of aerosols. Since scattering hazes can play an important role in moderating K2-18 b's temperature under a hycean interpretation (N. Madhusudhan et al. 2023; S. Jordan et al. 2025), our nondetection of a haze—based on data that include the bluer $0.6\text{--}0.8 \mu\text{m}$ second-order NIRISS SOSS data that are more sensitive to a haze slope—provides important context for the viability of a liquid water ocean. However, given the present signal-to-noise ratio at the shortest wavelengths (e.g., Figure 6), additional observations with NIRISS SOSS would be required before hazes can be definitively ruled out.

4.2.5. Unocculted Starspots

By including stellar contamination in our retrievals, we obtain constraints of unocculted active regions on K2-18. We find a small coverage fraction (2%–10%) of cool spots ($\Delta T \approx 500 \pm 350 \text{ K}$ cooler than the photosphere). While this may appear contrary to the finding of no stellar contamination reported by N. Madhusudhan et al. (2023), we note that our inclusion of the second-order NIRISS data provides coverage of shorter wavelengths where starspot contamination is more prominent (e.g., O. Lim et al. 2023; M. Fournier-Tondreau et al. 2024; C. Cadieux et al. 2024; M. Radica et al. 2025). We investigated the impact of stellar contamination by running several additional POSEIDON retrievals with stellar contamination excluded, with one example shown in Figure 8. We find that this potential low-level stellar contamination does not notably affect the inferred atmospheric composition of K2-18 b. However, given the stochastic nature of the transit light source effect, further repeat observations will be necessary to confirm this. Therefore, while it is likely that future transit observations of K2-18 b will be able to be stacked to increase the transmission spectrum precision, it would be prudent to continue observations with NIRISS SOSS (as opposed to only observing at longer wavelengths) to further constrain the possibility of stellar heterogeneities.

With our inferences from K2-18 b's JWST transmission spectrum in hand, we next turn to an assessment of viable atmospheric and interior structures.

5. Atmosphere and Interior Modeling: The Nature of K2-18 b

K2-18 b's mass and radius allow a wide range of planetary structures, with the most widely considered scenarios being a mini-Neptune or a hycean world. Here, we conduct a careful

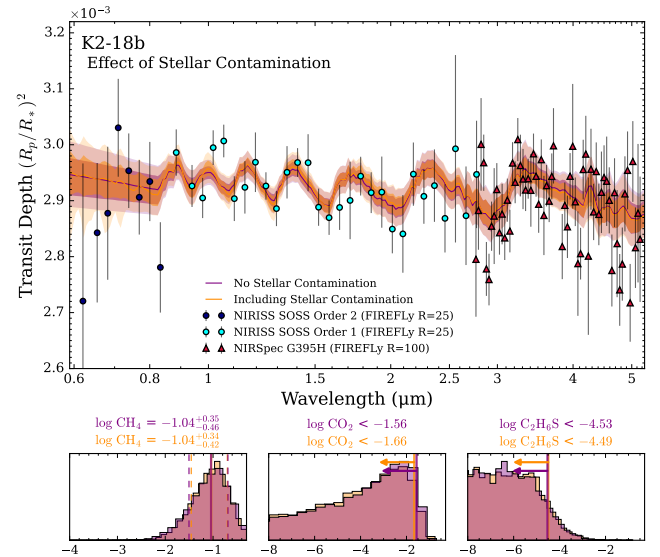


Figure 8. Influence of stellar contamination on K2-18 b retrievals. Top: two POSEIDON retrieval models for the low-resolution FIREFLY data variant, one including a planetary atmosphere and stellar contamination (orange) and an alternative atmosphere-only model (purple). Bottom: corresponding posterior distributions for the CH_4 , CO_2 , and DMS abundances. The inferred atmospheric composition is insensitive to whether or not stellar contamination is included.

assessment of the range of planetary structures and climates compatible with our retrieved composition (Section 5.1). We also calculate interior models to assess which planetary structures are compatible with the observed mass and radius (Section 5.2). We proceed to elucidate viable natures for K2-18 b.

5.1. Atmosphere Modeling

We first quantify the expected compositional and climate properties for K2-18 b across various proposed scenarios and compare them to our retrieved atmospheric abundance constraints. Our analysis builds upon a similar narrative to N. F. Wogan et al. (2024), but here, we use our revised retrieved abundances to evaluate if a hydrogen-rich mini-Neptune or a hycean planet with a liquid water surface is favored. We consider four scenarios: (i) a mini-Neptune with water clouds; (ii) an oxygen-poor mini-Neptune; (iii) a hycean planet; and (iv) a supercritical ocean.

5.1.1. Mini-Neptune with Water Clouds

First, we consider K2-18 b as a mini-Neptune with a thick hydrogen envelope. The key distinction from previous studies (S.-M. Tsai et al. 2021a; G. J. Cooke & N. Madhusudhan 2024; N. F. Wogan et al. 2024) is that here we compute temperature profiles consistent with the disequilibrium composition predicted by photochemistry and vertical mixing. We simulate mini-Neptune atmospheres with two independent sets of coupled photochemical-climate models: (1) Photochem (v0.6.2, N. Wogan 2024; N. F. Wogan et al. 2024) and PICASO (v3.3.0, N. Batalha et al. 2024; S. Mukherjee et al. 2025); and (2) VULCAN (S.-M. Tsai et al. 2017, 2021b) and HELIOS (M. Malik et al. 2017, 2019). For each simulation, we iterate between the photochemical and climate models, passing the chemical composition and temperature profiles between the codes until a photochemical-climate steady state is achieved.

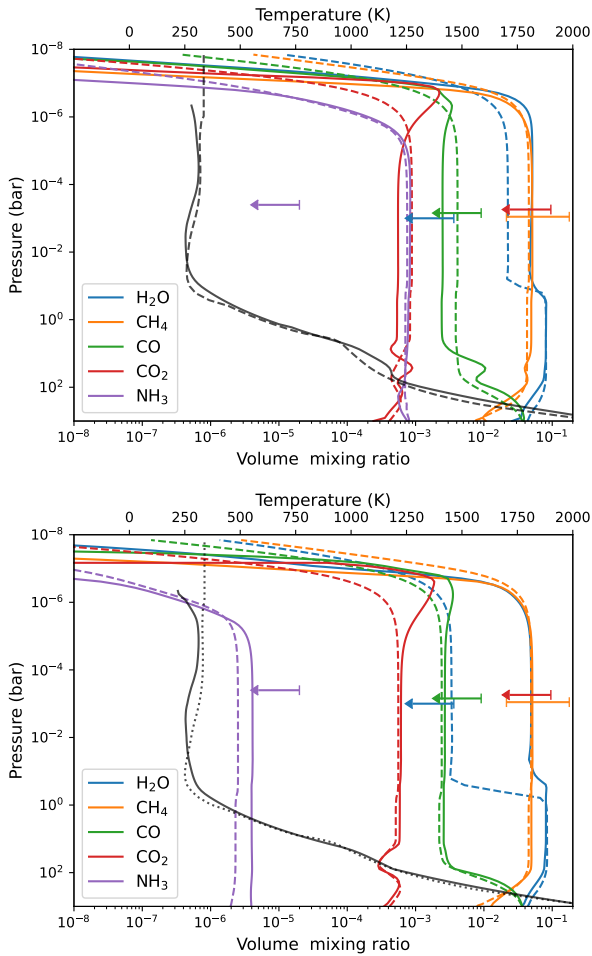


Figure 9. Self-consistent mini-Neptune models of K2-18 b’s atmosphere. Upper panel: $100\times$ solar metallicity case. Lower panel: the same metallicity, but with $5000\times$ nitrogen depletion. The temperature (black lines) and composition profiles (colored lines) from VULCAN-HELIOS (solid lines) and Photochem-PICASO (dashed lines) with the 1σ CH_4 range (orange bar) and 95% upper limits for other molecules (arrows).

Figure 9 (upper panel) illustrates the temperature and chemical composition computed by our self-consistent models, assuming $100\times$ solar metallicity, an intrinsic temperature of 60 K, and a uniform eddy diffusion coefficient (K_{zz}) of $10^7 \text{ cm}^2 \text{ s}^{-1}$ (similar to R. Hu et al. 2021; S.-M. Tsai et al. 2021a). The simulated CH_4 matches the retrieved abundance, while the modeled CO and CO_2 remain compatible with their respective 2σ upper limits of $\log_{10} \text{CO} = -2.45$ and $\log_{10} \text{CO}_2 = -1.70$. However, the modeled H_2O and NH_3 exceed our inferred upper limits. G. J. Cooke & N. Madhusudhan (2024) highlighted that K2-18 b’s temperature as a mini-Neptune with $100\times$ solar metallicity equilibrium composition is too high to condense enough water to meet the H_2O upper limit, unless the stellar irradiation is artificially reduced by 40%. Our self-consistent calculations yield similar results, with the temperature remaining too warm and the H_2O mixing ratio only reduced by a factor of a few due to condensation. The Photochem-PICASO model produces a slightly cooler temperature than VULCAN-HELIOS and, hence, a marginally lower water abundance. Without ad hoc irradiation reduction, the predicted H_2O abundance exceeds the 2σ upper limit by more than an order of magnitude for both simulations.

Second, the lack of detected NH_3 is one of the main arguments against a mini-Neptune scenario, as NH_3 is

hypothesized to be present at detectable abundances ($\log_{10} \text{NH}_3 \approx -3$, see Figure 9) in a thick H_2 -dominated atmosphere (R. Hu et al. 2021; S.-M. Tsai et al. 2021a; X. Yu et al. 2021). A potential explanation for lower NH_3 abundances, suggested by O. Shorttle et al. (2024), is that a reducing magma layer at the bottom of the gaseous envelope can efficiently dissolve nitrogen into the magma and hence deplete NH_3 from the gas phase.

We explore the extent of nitrogen depletion required to explain our nondetection of NH_3 in the lower panel of Figure 9. This second mini-Neptune model also has a $100\times$ solar metallicity atmosphere, but with the total nitrogen depleted by a factor of 5000 (i.e., $\text{N}/\text{H} = 0.02\times$ solar) to account for nitrogen dissolution into a magma ocean. Such a low N/H ratio is close to, but within the lower bound, of expected nitrogen abundances under low magma oxygen fugacity (F. E. Rigby et al. 2024; O. Shorttle et al. 2024). This nitrogen-depletion factor lowers the atmospheric NH_3 abundance at the pressure probed by transmission spectroscopy to < 10 ppm, compatible with our retrieved 95% upper limit. Furthermore, a depleted NH_3 abundance moves the stratosphere to a slightly higher pressure, which allows additional H_2O condensation (especially for the Photochem-PICASO model). The condensed water would form clouds around 0.1 bar, which is consistent with our retrieval constraints without damping the CH_4 features over the NIRISS spectral range.

Should K2-18 b’s atmosphere possess a scattering haze (not considered in these models), the tropopause could be cooler than shown in Figure 9 with a corresponding lower H_2O abundance in the upper atmosphere (A. A. A. Piette & N. Madhusudhan 2020). However, a fine balance is needed to avoid producing a sufficiently strong haze that would dampen the shortest wavelength CH_4 bands to a degree not seen in the JWST NIRISS transmission spectrum (J. Leconte et al. 2024). These considerations motivate additional modeling efforts under the mini-Neptune scenario to explore the sensitivity of the H_2O and CH_4 abundance to aerosols. Nevertheless, these results demonstrate that a mini-Neptune scenario for K2-18 b, including nitrogen depletion from a basal magma ocean, is a plausible interpretation of our retrieved atmospheric abundances—although an additional albedo source is required to cold-trap water.

5.1.2. Mini-Neptune with High C/O Ratio

An alternative way to explain the low water abundance on K2-18 b is an oxygen-poor atmosphere with a high carbon-to-oxygen (C/O) ratio. K2-18 b’s current orbit sits between the soot (J. Li et al. 2021) and water ice lines. If the planet migrated early to its present location within the ice line, it is plausible to form a carbon-rich and oxygen-poor envelope (E. A. Bergin et al. 2023; J. Yang & R. Hu 2024). We now consider the necessary C/O ratio to explain our nondetection of H_2O under the mini-Neptune scenario without an additional albedo source.

We first consider the simplest calculation of thermochemical equilibrium to illustrate the impact of the C/O ratio on the atmospheric CH_4 and H_2O abundances. We perform these equilibrium calculations using Fastchem (J. W. Stock et al. 2022; D. Kitzmann et al. 2024) over the 1–100 bar pressure range in the deep atmosphere for a range of temperatures. Since this pressure range encloses possible quench points,

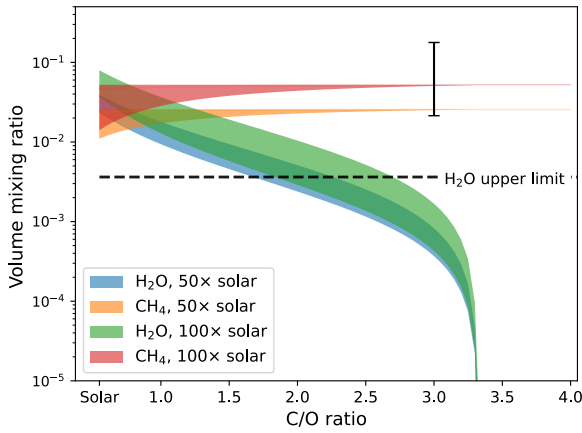


Figure 10. The equilibrium abundances of H_2O and CH_4 as a function of C/O ratio. The shaded regions represent the range across pressures from 100–1 bar and temperatures from 800–1000 K, which cover the expected quench levels. The dashed line is the 2σ upper limit of H_2O from our retrieval analysis, and the open circle with an error bar represents our retrieved 1σ CH_4 abundance.

these deep equilibrium abundances should broadly reflect the abundances from $\sim 10^{-4}$ – 10^{-1} bar probed by the JWST spectrum—as confirmed by the rather uniform vertical distributions of CH_4 and H_2O within this region in the absence of condensation (e.g., Figure 9; see also S.-M. Tsai et al. 2021a; X. Yu et al. 2021; G. J. Cooke & N. Madhusudhan 2024).

Figure 10 shows the equilibrium H_2O and CH_4 abundances as C/O varies. For simplicity, we fixed carbon and only changed the oxygen abundance. From the reasoning above, the equilibrium H_2O and CH_4 abundances shown by the shaded regions in Figure 10 can be considered as the observable abundances. We see that H_2O decreases with increasing C/O—well known from previous studies (e.g., N. Madhusudhan 2012; P. Mollière et al. 2015), as oxygen is preferentially sequestered within silicates—while CH_4 remains nearly constant. We require $\text{C/O} \gtrsim 3$ under equilibrium chemistry to lower the H_2O abundance below the 95% upper limit from our retrievals. This simple analysis suggests that an oxygen-depleted (high C/O ratio) mini-Neptune scenario, without effective water condensation, can be consistent with the observed H_2O and CH_4 constraints. Having demonstrated H_2O depletion under equilibrium conditions, we next turn to self-consistent photochemical-climate models for a high C/O ratio.

Figure 11 shows our self-consistent photochemical-climate models for K2-18 b under a high C/O ratio mini-Neptune scenario. The atmospheric properties are the same as Section 5.1.1 (100 $\times$ solar metallicity and 5000 $\times$ N depletion), but with $\text{C/O} = 3.25$. We see that this oxygen-poor scenario naturally explains the retrieved CH_4 abundance, alongside the nondetections of H_2O and CO_2 , without invoking water condensation. Additionally, CO_2 is negligible due to the scarcity of oxygen. Future observations with tighter constraints on CO_2 could help differentiate this oxygen-depleted scenario from the cold-trapped water scenario discussed in Section 5.1.1.

5.1.3. Hycean Scenario

Previous arguments for hycean atmospheres have emphasized the presence of CO_2 (R. Hu et al. 2021; N. Madhusudhan et al. 2023), since CO_2 is expected to be the dominant carbon-bearing molecule. Its abundance, possibly ranging from ~ 1 ppm to 1%, would be controlled by the solubility balance with the ocean (E. S. Kite & E. B. Ford 2018; R. Hu et al. 2021). Conversely,

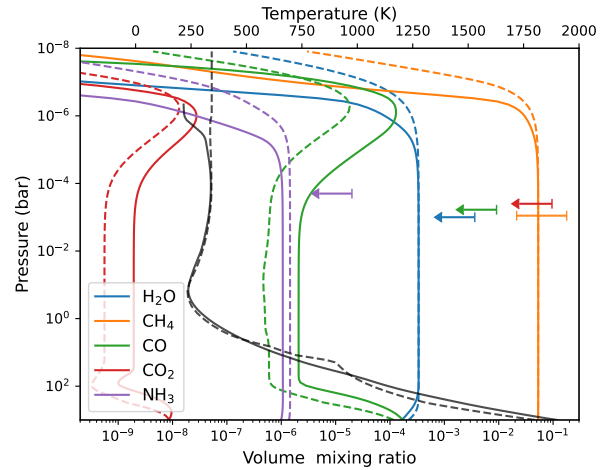


Figure 11. Self-consistent mini-Neptune models for K2-18 b with an oxygen-depleted atmosphere. The atmospheric conditions and models are the same as the lower panel in Figure 9, but with $\text{C/O} = 3.25$.

CH_4 is not photochemically stable in a small atmosphere. The high amount of CH_4 ($\sim 10\%$ from our retrievals) requires an additional supply, as demonstrated by N. F. Wogan et al. (2024).³⁰

Given our nondetection of CO_2 from K2-18 b’s JWST transmission spectrum, while CH_4 is robustly detected, we now consider whether a hycean scenario remains viable for K2-18 b. Since a hycean world possesses a thin H_2 atmosphere to sustain a liquid water ocean, CH_4 is expected to be photochemically converted to CO_2 in the upper atmosphere, as discussed in S.-M. Tsai et al. (2021a) and N. F. Wogan et al. (2024). To explore whether our CH_4 and CO_2 constraints fit the hycean scenario, we conducted a test to demonstrate how CH_4 undergoes such conversion. We adopted a representative temperature profile for a hycean atmosphere from S.-M. Tsai et al. (2024; assuming a surface albedo of 0.3) and fixed the surface CH_4 abundance at 1%, without any initial CO_2 , representing a surface source continuously replenishing CH_4 .

The resulting steady-state composition and the evolution are shown in the upper panel of Figure 12. Compared to the mini-Neptune scenario (Figures 9 and 11), CO_2 and CO exhibit significantly higher abundances.

The evolution in the lower panel of Figure 12 illustrates the photochemical conversion of CH_4 into CO_2 , occurring on a geologically short timescale of ~ 10 – 100 Myr. While the exact CO_2 abundance depends on the precise levels of H_2O , CH_4 , and the surface temperature, the main takeaway of our test is that CO_2 and CO are expected to have comparable abundances to CH_4 within a CH_4 -rich ($\sim 1\%$) and thin atmosphere. Therefore, regardless of the source of CH_4 , our well-constrained measured CH_4 abundance, which lies above the 95% upper limits for both CO_2 and CO , poses a challenge to reconcile the absence of CO_2 , as CH_4 should naturally convert to CO_2 in a hycean atmosphere.

Should K2-18 b indeed have a water-rich interior covered by a hydrogen-helium envelope, as allowed by the observed mass and radius, the water will most likely exist in a supercritical state. This water phase arises from the greenhouse effect of H_2 collision-induced absorption and the suppression of moist convection (R. T. Pierrehumbert 2023; H. Innes et al. 2023).

³⁰ See also G. J. Cooke & N. Madhusudhan (2024) for a thorough analysis of its dependence on various parameters.

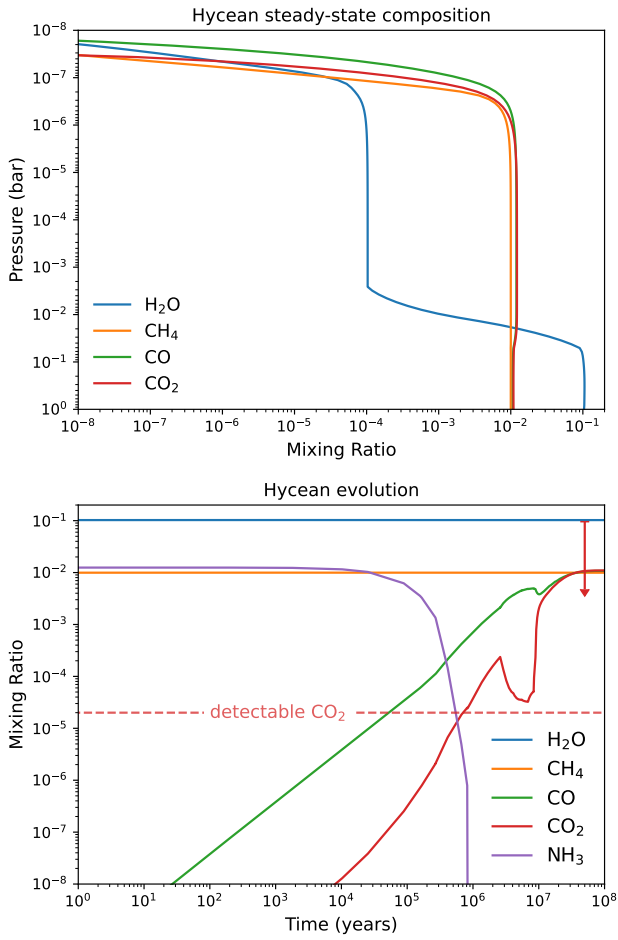


Figure 12. Atmospheric composition over time for a hycean scenario for K2-18 b. Upper panel: the steady-state composition profiles computed by VULCAN in our hycean scenario. Lower panel: the composition evolution at 0.1 bar. The red arrow corresponds to the 2σ upper limit of CO_2 from our retrievals, while the dashed line indicates the CO_2 abundance required to produce a potentially detectable (~ 100 ppm) transmission feature at $4.2 \mu\text{m}$.

Under such a superrunaway scenario, supercritical H_2O becomes miscible with H_2 (F. Soubiran & B. Militzer 2015; A. Gupta et al. 2025). This was recently proposed to be the case for TOI-270 d (B. Benneke et al. 2024) as well as a possible explanation for GJ 1214 b (M. C. Nixon et al. 2024b). Without an appropriate climate model to track the transition from the supercritical ocean to the subcritical upper atmosphere, it is unclear whether the temperature in the upper atmosphere allows a condensing layer to cold-trap water enough to match our retrieved abundances.

Under a supercritical ocean scenario, a wide range of CH_4/CO_2 ratios is allowed due to the unknown hydrogen molarity of K2-18 b (C. N. Luu et al. 2024). Consequently, it is not currently possible to conduct a detailed comparison between our retrieved CH_4 abundance and the CO_2 upper limit. Future observations that provide a tighter CO_2 abundance constraint would offer useful insights into the potential thermal properties of a supercritical ocean.

5.2. Interior Structure Modeling

To aid in evaluating the possibility of a liquid water ocean on K2-18 b, we also construct several interior structure models of the planet, which match its observed mass and radius

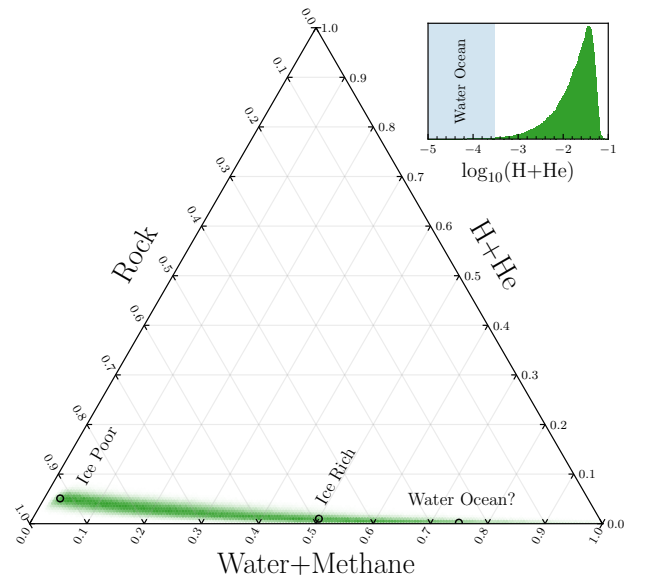


Figure 13. Potential interior compositions for K2-18 b given our retrieval analysis. The green shaded region on the ternary diagram shows an MCMC posterior of compositions consistent with K2-18 b's mass, radius, host star flux, atmosphere metallicity ($\sim 30\%$ CH_4), and system age. We group methane with water as volatile species for analysis in terms of ice-to-rock ratios, but is physically located in the H/He layer in our model. The three cases discussed in Section 5.2 are circled and labeled. The inset shows the posterior histogram of the $\log(\text{H}/\text{He})$ mass fraction, with the amount permitting a liquid water ocean (shaded region) containing only 0.58% of our posterior. Above that, the water layer is a supercritical-fluid.

($M_p = 8.63 \pm 1.35 M_\oplus$, $R_p = 2.61 \pm 0.087 R_\oplus$). We use the mass and radius presented by B. Benneke et al. (2019) as the basis of our models.³¹ We prefer this set of planet parameters due to the use of a Gaia parallax in their calculation. These interior models solve the equations of hydrostatic equilibrium, mass conservation, and the material equations of state (see D. P. Thorngren et al. 2016). This is similar to the analysis of N. Madhusudhan et al. (2020), but includes methane in the atmosphere to comport with the observed spectra, setting its mass fraction prior as normally distributed with the mean 0.3 and standard deviation 0.05. We used the G. Chabrier & F. Debras (2021) equation of state for H/He, S. Mazevet et al. (2019) for water, N. Nettelmann et al. (2016) and M. Bethkenhagen et al. (2017) for methane, ANEOS (S. L. Thompson 1990) for rock, and SESAME (S. P. Lyon & J. D. Johnson 1992) for iron. The methane and H/He are mixed in the atmosphere via the additive volume approximation, with the adiabats calculated through thermal integration (see D. P. Thorngren 2019, Equations (1.07)–(1.18)). We thermally evolve the planet using atmosphere models from J. J. Fortney et al. (2007) to regulate heat flow from the interior.

We used MCMC to find compositions of the planet, which match the planet's observed properties, especially the radius and the atmospheric methane abundance. Figure 13 shows the results as a ternary diagram of the mass fractions of rock (and iron, assuming an Earth-like ratio), nebular ices (here, water and methane), and H/He (at a solar ratio) for uniformly distributed composition priors. Our models find H/He makes up less than 10% of the planet's mass, with the exact limit

³¹ If the planet were smaller, as reported in earlier work, the required water fraction for a thin H/He atmosphere drops to a more plausible range.

Table 4
Mini-Neptune vs. Hycean Interpretations for K2-18 b

Scenario Constraint	Mini-Neptune	Hycean
Climate	Compatible	Requires high-albedo aerosols
Interior	Compatible	Requires high water-to-rock ratio
H ₂ O	Requires high-albedo aerosols or oxygen depletion	Requires high-albedo aerosols
CH ₄	Compatible	Requires external source
CO ₂	Compatible	Compatible, but not compatible with the coexistence with CH ₄
NH ₃	Requires high depletion from a basal magma ocean	Compatible through dissolution into the water

varying with the water abundance. Any of the models along the shaded region in Figure 13 fits the observed data, but we identify a few specific cases of interest (denoted with circles) for further discussion.

First, we consider the ice-poor case, which represents a typical Neptune-like planet that could form within the ice line, experiencing limited volatile accretion. In this case, approximately 6% of the mass must be H/He to match the observed radius. Such a planet features a deep envelope on top of a solid or molten rock surface. A more ice-rich planet could result from forming outside the ice line, with rock-to-water ratios around 50–50 as seen in the solar system icy moons (A. P. Showman & R. Malhotra 1999; O. L. Kuskov & V. A. Kronrod 2005). Stratified into separate layers, the planet would have a H/He envelope of about 1% of the planet’s mass. This results in the surface of the water layer occurring at about $2.21 R_{\oplus}$, a pressure of 100 kbar, and a temperature of 3200 K; a supercritical fluid rather than a liquid.

To support a liquid water ocean, the surface must not exceed the critical point of water. Our model atmospheres pass almost directly through the water critical point at 647 K and 220 bar. Thus, the H/He atmosphere must not pressurize the water beyond this, which corresponds to an H/He layer mass fraction of no more than 0.03%. Our posterior probability of this composition is just 0.59% (Figure 13, inset). Furthermore, accounting for the possibility of a runaway greenhouse effect could push this pressure limit down to tens of bars (H. Innes et al. 2023).

In order to match the observed radius, such a low H/He fraction must be compensated for by a high water fraction of at least 70%. This is improbable from a formation perspective—K. Lodders (2003) find the solar water-to-rock ratio to be 1.17:1 (54%), similar to the solar system icy moons (O. L. Kuskov & V. A. Kronrod 2005). R. A. Marcus et al. (2010) argue that this initial water-to-rock ratio provides an upper limit for the final planet composition, as giant impacts cannot enhance the water fraction, only deplete it. Including other ices (methane and ammonia) gives a max ice-to-rock ratio of 2.41:1 (70%), but this would require formation outside the *methane* ice line and subsequent migration, and methane’s low solubility in water suggests it would likely contribute to a thick atmosphere rather than permit a water ocean. Thus, while we cannot absolutely rule out a water ocean, such a structure is very improbable for K2-18 b due to its large radius.

6. Summary and Discussion

The nature of K2-18 b is of broad interest for sub-Neptune studies and for exoplanet habitability. We have conducted a comprehensive reanalysis of the JWST transmission spectrum of K2-18 b from NIRISS SOSS and NIRSpec G395H. Our

approach considers multiple well-tested data reduction and atmospheric retrieval codes to ensure robust and reliable results. Thus, our analysis assessed what can be confidently inferred about K2-18 b’s atmosphere and interior from the initial JWST observations. Our main results are as follows:

1. We confirm the detection of CH₄ in K2-18 b’s atmosphere ($\approx 4\sigma$), with a volume mixing ratio of $-2.14 \leq \log_{10} \text{CH}_4 \leq -0.53$ (1%–30%).
2. We do not detect CO₂ or DMS, contrary to the findings of N. Madhusudhan et al. (2023), placing 95% confidence upper limits of $\log_{10} \text{CO}_2 < -1.02$ and $\log_{10} \text{DMS} < -2.51$.
3. Any inferences of CO₂ and DMS from these JWST data are unreliable and of low significance, exhibiting high sensitivity to choices made during the reduction process and retrieval model assumptions (Figure 5).
4. K2-18 b’s retrieved atmospheric composition can be explained by a $100\times$ solar mini-Neptune with an oxygen-poor and nitrogen-depleted (e.g., from a magma ocean) composition.
5. A hycean scenario for K2-18 b is only viable with an anonymously high water-to-rock ratio, a low H/He fraction, atmospheric hazes, and positing a substantial CH₄ flux. We regard the necessary fine-tuning as an unlikely product of nature.

Based on our reanalysis, we summarize the observational constraints, challenges, and viability for each interpretation of K2-18 b in Table 4. Our findings raise many implications about K2-18 b, and the sub-Neptune population more broadly, as we now discuss.

6.1. Plausibility of Hycean Scenarios and Biosphere Viability for K2-18 b

Our revised atmospheric composition of K2-18 b, accounting for data-level uncertainties, can be naturally explained by a mini-Neptune without requiring a liquid water ocean or life. Our nondetection of CO₂ lowers the potential CO₂/CH₄ ratio from ~ 1 to < 0.3 (at 95% confidence). This lower ratio complicates a hycean interpretation for K2-18 b, as a high CO₂/CH₄ is one of the primary lines of evidence for a liquid water ocean underlying a thin H₂-dominated envelope (R. Hu et al. 2021; S.-M. Tsai et al. 2021a). However, absence of evidence is not evidence of absence, and we stress that even mini-Neptune scenarios will contain some level of CO₂ (e.g., Figure 9 predicts $\log_{10} \text{CO}_2 \sim -3$). Consequently, should future, more sensitive, JWST observations result in a CO₂ detection with an abundance lower than our upper limit ($\log_{10} \text{CO}_2 < -1.7$), this would be an additional line of evidence *against* a hycean interpretation (see also Section 6.4).

Table 5
Evidence for Molecular Signatures in K2-18 b’s JWST Transmission Spectrum

Molecule	This Work				N. Madhusudhan et al. (2023)				
	$\mathcal{B}_{\text{Ref}, i}$	$N\sigma_{\text{classic}}$	$N\sigma_{\text{intuitive}}$	Jeffreys’ Scale	$\mathcal{B}_{\text{Ref}, i}$	$N\sigma_{\text{classic}}$	$N\sigma_{\text{intuitive}}$	Their Classification	Jeffreys’ Scale
CH ₄	12–10 ⁶	2.7 σ –5.6 σ	1.8 σ –4.9 σ	Moderate– Strong evidence	11,000–44,000 ^a	4.7 σ –5.0 σ	3.9 σ –4.2 σ	Robust detection	Strong evidence
CO ₂	0.6–4.2	<2.3 σ	<1.3 σ	No evidence	18–40 ^a	2.9 σ –3.2 σ	1.9 σ –2.3 σ	Robust detection	Moderate evidence
DMS	0.6–3.9	<2.2 σ	<1.3 σ	No evidence	$\sim$ 1–5 ^a	<2.4 σ	<1.4 σ	Potential signs	No evidence– Weak evidence

Notes. $\mathcal{B}_{\text{Ref}, i}$ is the Bayes factor between models with and without molecule “i.” $N\sigma_{\text{classic}}$ uses the relation from B. Benneke & S. Seager (2013), while $N\sigma_{\text{intuitive}}$ is derived from equating $p = 1/(1 + \mathcal{B}_{\text{Ref}, i})$ (D. Kipping & B. Benneke 2025). The Jeffreys’ scale uses the following terms: “no evidence” ($\mathcal{B}_{\text{Ref}, i} < 3$), “weak evidence” ($3 < \mathcal{B}_{\text{Ref}, i} < 12$), “moderate evidence” ($12 < \mathcal{B}_{\text{Ref}, i} < 150$), and “strong evidence” ($\mathcal{B}_{\text{Ref}, i} > 150$).

^a Bayes factors were not provided by N. Madhusudhan et al. (2023), so we compute them from their quoted detection significances. The range of Bayes factors and significances quoted for N. Madhusudhan et al. (2023) correspond to their no offset, one offset, and two offsets cases. Our lowest Bayes factor for CH₄ corresponds to $R = 25$ NIRISS data, which hinders the CH₄ detectability (see Section 4.2.1 and Figure 5), but for $R \geq 100$ data, the minimum CH₄ Bayes factor is 86 ($N\sigma_{\text{intuitive}} = 2.5\sigma$). Since the formula mapping Bayes factors to $N\sigma_{\text{classic}}$ breaks down for low Bayes factors ($\lesssim 3$), we quote only the highest significance for CO₂ and DMS.

Our photochemical modeling offers an additional line of evidence against inhabited hycean scenarios. We show in Figure 12 that, even if one assumes a hycean scenario with a large unknown CH₄ flux (e.g., from methane-producing life), over time, the atmospheric CO₂ and CO abundances would build to levels comparable to or above our upper limits. Consequently, our nondetections of CO₂ and CO are not consistent with the observed CH₄ abundance under an inhabited hycean scenario. Our results therefore render the mini-Neptune scenario the most likely interpretation for K2-18 b given current observational constraints.

Our revised atmospheric composition also motivates a careful reevaluation of the viability of methanogenesis on K2-18 b. Our nondetections of CO₂ and DMS prompt a reassessment of whether a sufficient free energy source exists for methane-producing life under the metabolic pathway studied by C. R. Glein (2024), rendering it less plausible that K2-18 b sustains a biosphere. Further, our models in Figure 12 demonstrate that a hycean scenario with the CH₄ abundance we retrieve would necessarily have a high $\sim 1\%$ CO₂ abundance that is in slight tension with our nondetection of CO₂ (although it is consistent with our 2σ upper limit). Updated metabolic calculations that consider the minimum viable CO₂ abundance to sustain methanogenesis will be required to assess whether a biosphere is ruled out by our revised atmospheric composition.

6.2. On Establishing Robust and Reliable Molecular Detections from Atmospheric Retrievals

Our revised atmospheric composition for K2-18 b demonstrates the critical importance of using multiple data reductions and multiple retrieval codes to establish which molecular detections are robust. In particular, our nondetection of CO₂ stands in stark contrast to the finding of a robust detection reported by N. Madhusudhan et al. (2023). However, this is not the first time that reported molecular detections in K2-18 b’s atmosphere have proven unreliable. Pre-JWST, multiple transmissions reported detections of H₂O vapor from HST WFC3 transmission spectroscopy at compelling statistical significances: 3.6 σ (A. Tsaras et al. 2019), 3.9 σ (B. Benneke et al. 2019), and 3.3 σ (N. Madhusudhan et al. 2020). We now know from the JWST spectrum of K2-18 b that the spectral feature seen by WFC3 was, in fact, CH₄ masquerading as H₂O (likely

due to WFC3 edge effects near 1.7 μm). Here, we discuss here how future retrieval analyses can safeguard against overstating the statistical evidence of a molecular detection.

First, our grid of retrievals demonstrates that detection significances naturally possess their own uncertainty due to data-level assumptions. Figure 5 clearly illustrates a spread of CH₄ significances (clustered around 4 σ) as a function of the data reduction code and the degree of spectral binning preretrieval. We propose that visualizations such as Figure 5 are a valuable tool to intuitively establish if a molecular detection is robust (as with CH₄) or whether there is no reliable statistical evidence (as with CO₂ and DMS). An additional value of Figure 5 is the clear representation of lost information when retrieving NIRISS spectra at $R = 25$ rather than $R = 100$, resulting in a significantly lower CH₄ detection significance, which suggests $R = 100$ is a safe minimum resolution for inferring atmospheric properties from JWST transmission spectra. Our first takeaway is thus that the term “robust” must be reserved for molecular detections that hold across multiple data reduction and retrieval codes.

Second, we note that the retrieval literature tends to present overconfident detection significances and inconsistent terminology from Bayesian model comparisons. When establishing if a molecule is present in an atmosphere, the common practice is to conduct a Bayesian model comparison between a retrieval model containing the molecule in question and one with it removed, calculate the Bayes factor, and then convert the Bayes factor into an “equivalent detection significance” (using the inverse of Equation (27) in R. Trotta 2008). This approach was first introduced in the exoplanet retrieval context by B. Benneke & S. Seager (2013) and has been widely used in retrieval codes since (e.g., I. P. Waldmann et al. 2015; R. J. MacDonald & N. Madhusudhan 2017; L. Welbanks & N. Madhusudhan 2021). This procedure often provides useful insights, provided that (i) both the Bayes factor and equivalent significance are provided, and (ii) well-defined terminology is adopted (e.g., the Jeffreys’ scale; see R. Trotta 2008). Not providing Bayes factors can produce overconfidence in a result. For example, a Bayes factor of 3 corresponds to an equivalent significance, according to the approach from B. Benneke & S. Seager (2013), of 2.1 σ , when this model is actually only favored at a 3:1 odds ratio. Similarly, a claim of a 3 σ “detection” from a Bayesian model comparison is only a Bayes factor of 21. Recently, D. Kipping & B. Benneke (2025)

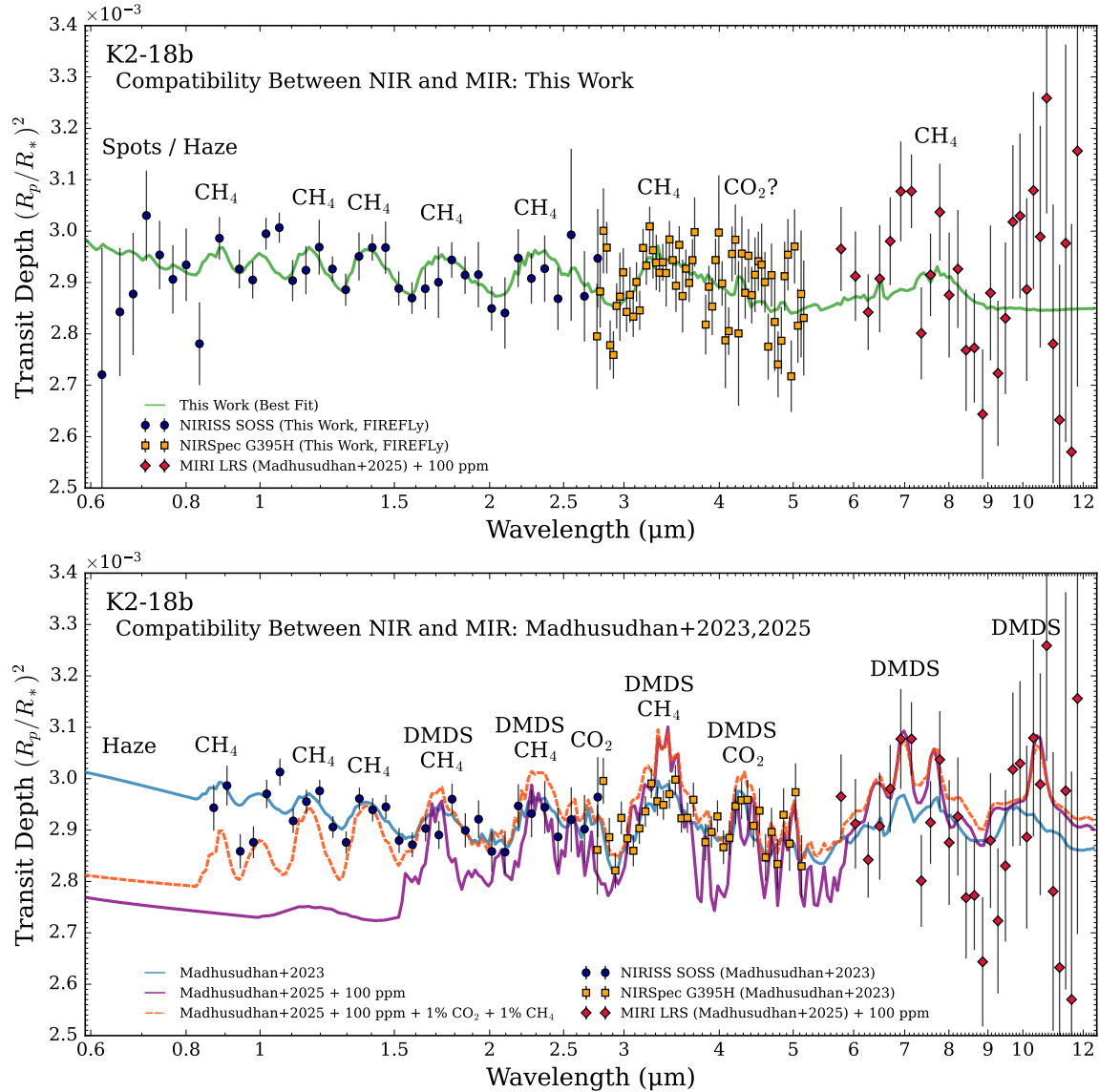


Figure 14. Extrapolated best-fit models compared to additional JWST transmission spectra of K2-18 b. Top: best-fitting POSEIDON model to the low-resolution FIREFLY near-infrared data (NIRISS SOSS, $R \approx 25$; NIRSpec G395H, $R \approx 100$) extrapolated into the mid-infrared for comparison with the MIRI LRS data from N. Madhusudhan et al. (2025) ($R \approx 40$). Bottom: reproduction of the median model spectra from N. Madhusudhan et al. (2023; near-infrared only) and N. Madhusudhan et al. (2025; mid-infrared only) with POSEIDON, each extrapolated over the wavelength range of the other data sets. All models are shown binned to $R = 100$ for clarity, and a 100 ppm offset is applied to the MIRI LRS data, and the models corresponding to N. Madhusudhan et al. (2025), to account for differences in the baseline transit depth. The model from this study, exhibiting no absorption features from DMS or DMDS, provides an adequate fit to the MIRI data. However, the median atmospheric properties from N. Madhusudhan et al. (2025) predict DMDS absorption features with amplitudes inconsistent with the near-infrared data. Near-IR (NIR); mid-IR (MIR).

highlighted this point and suggested that a more intuitive detection significance is given by equating the false-positive detection probability with the p -value (i.e., $p = 1/(1 + \mathcal{B})$ and deriving $p = \text{erfc}[N\sigma_{\text{intuitive}}/\sqrt{2}]$. This results in values of $N\sigma_{\text{intuitive}}$ that are approximately 1 lower than the relation from B. Benneke & S. Seager (2013; see also R. Trotta 2008). Using this logic, the “3 σ detection” of CO₂ in K2-18 b’s atmosphere reported by N. Madhusudhan et al. (2023; without quoted Bayes factors) should be better thought of as a 2 σ result, for which the term “detection” is too strong. It is for this reason that we report all Bayes factors < 3 (equivalent to $N\sigma_{\text{intuitive}} \lesssim 1$) as nondetections in Figure 5.

In Table 5, we summarize and compare the statistical evidence for CH₄, CO₂, and DMS in K2-18 b’s atmosphere from our study with N. Madhusudhan et al. (2023). We present the Bayes

factors, the “classical” detection significance (following the scheme from B. Benneke & S. Seager 2013), and the “intuitive” detection significance (D. Kipping & B. Benneke 2025). First, we consider the evidence for CH₄ in K2-18 b’s atmosphere. Figure 5 shows that 33/40 Bayesian model comparisons with $R_{\text{data}} \geq 100$ have $N\sigma_{\text{classic}} > 4.0\sigma$ —or $N\sigma_{\text{intuitive}} > 3.0\sigma$ —which corresponds to “strong evidence” on the Jeffreys’ scale (only the FIREFLY $R = 100$ NIRISS data result in slightly lower significances). At this level of model preference, the term “detection” is clearly warranted for CH₄ in both our study and N. Madhusudhan et al. (2023). Second, we see that the statistics reported for CO₂ by N. Madhusudhan et al. (2023) should be downgraded from “robust detection” to “moderate evidence” according to the Jeffreys’ scale, while our retrieval model comparisons are classified as “no evidence” for CO₂ (see Figure 5). These results

highlight the reliable and reproducible CH₄ inference, but show that the JWST data from N. Madhusudhan et al. (2023) do not provide sufficient evidence supporting a detection of CO₂ (see also Appendix A). Finally, we see that our retrievals yield “no evidence” for DMS, while the results from N. Madhusudhan et al. (2023) correspond to “no evidence” or “weak evidence” at most when using the Jeffreys’ scale. We therefore advocate for reporting more conservative statistical measures for detection significances from Bayesian model comparisons (e.g., $N\sigma_{\text{intuitive}}$, see D. Kipping & B. Benneke 2025), alongside exploring the sensitivity of detections to specific data regions (e.g., L. Welbanks et al. 2023), to avoid overstating the statistical evidence of molecules from JWST spectra of exoplanet atmospheres.

6.3. Compatibility between Near-infrared and Mid-infrared Transmission Spectra of K2-18 b

Here, we briefly examine the implications of our nondetections of DMS and CO₂ with the mid-infrared MIRI LRS spectrum of K2-18 b recently presented by N. Madhusudhan et al. (2025). In Figure 14 (upper panel), we show our best-fitting transmission spectrum to our FIREFLY low-resolution data (NIRISS SOSS at $R \approx 25$, NIRSpec G395H at $R \approx 100$) extrapolated to the mid-infrared. We see that this extrapolated model—which contains only a minor CH₄ feature and no DMS absorption in the mid-infrared—provides a reasonable fit to the MIRI LRS JEXORES data from N. Madhusudhan et al. (2025), consistent with there being no significant spectral features in the mid-infrared (J. Taylor 2025; L. Welbanks et al. 2025). However, the same is not true when extrapolating the transmission spectrum corresponding to the median retrieved atmospheric properties reported by N. Madhusudhan et al. (2025) backward to the near-infrared. We show in Figure 14 (lower panel) that the DMDS features reported by N. Madhusudhan et al. (2025) would produce significantly stronger absorption near 2.3 and 3.3 μm than is seen in the near-infrared data (or in the corresponding median retrieved models from N. Madhusudhan et al. 2023). This mismatch in scale heights is primarily driven by the MIRI LRS data favoring a temperature of ≈ 350 K, while the near-infrared data favor cooler temperatures of ≈ 240 K. We note that R. Luque et al. (2025) similarly found that simultaneous retrievals of the near-infrared and mid-infrared JWST data offer no significant evidence for DMS or DMDS in K2-18 b’s atmosphere.

6.4. Additional JWST Observations of K2-18 b

JWST will continue to observe spectra of K2-18 b, offering deeper insights into its atmospheric composition and internal structure. In particular, Program GO-2372 (PI: Renyu Hu) is obtaining new near-infrared observations of K2-18 b’s transmission spectrum. We briefly discuss how our results relate to potential outcomes from these upcoming observations.

Program GO-2372 is in the process of observing four additional transits with NIRSpec G395H (with two pending at the time of writing) and two transits with NIRSpec G235H. These observations will significantly improve the observational precision over the 1.7–5.2 μm wavelength range, offering additional sensitivity to the allowed abundances of molecules including CH₄, CO₂, DMS, CO, H₂O, NH₃, and HCN. The improved precision of these observations may allow

the detection of CO₂, which would be expected for a mini-Neptune with a slightly lower abundance than our present CO₂ upper limit (see Figure 9). Therefore, a detection of CO₂ in new multitransit JWST data would not be at odds with the nondetection of CO₂ from the single NIRISS SOSS and NIRSpec G395H transits we report here.

6.5. K2-18 b Is Unlikely to Harbor a Water Ocean

Our finding that K2-18 b is unlikely to be a hycean planet adds to the current dearth of observational evidence for the existence of hycean planets. A similar example is provided by JWST transmission spectra of the sub-Neptune TOI-270 d, which M. Holmberg & N. Madhusudhan (2024) identified as a hycean candidate using only NIRSpec data, but B. Benneke et al. (2024), using both NIRISS and NIRSpec data, found is better explained via a miscible supercritical envelope. Future JWST observations of sub-Neptunes will have a valuable role to play in assessing the existence of hycean worlds. Our results provide a cautionary tale for the value of exploring a spread of data-level and model-level approaches when retrieving sub-Neptune transmission spectra, with the consequence that there is currently no spectroscopic evidence for the existence of hycean planets in nature.

Acknowledgments

We thank the anonymous referee for the helpful comments. We also thank Adina Feinstein, Duncan Christie, Emma Esparza-Borges, Evert Nasedkin, and Jake Taylor for helpful comments. We thank Hamish Innes for an illuminating discussion on the supercritical ocean scenario. This work is based in part on observations made with the NASA/ESA/CSA JWST. The data were obtained from the Mikulski Archive for Space Telescopes at the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc., under NASA contract NAS 5-03127 for JWST. The specific observations analyzed can be accessed via MAST (N. Madhusudhan 2023). S.P.S. is supported by the National Science Foundation Graduate Research Fellowship Program under grant No. DGE2139757. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation. R.J.M. is supported by NASA through the NASA Hubble Fellowship grant HST-HF2-51513.001, awarded by the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc., for NASA, under contract NAS 5-26555. S.-M.T. acknowledges support from NASA Exobiology grant No. 80NSSC20K1437 and the University of California, Riverside. M.R. acknowledges support from the Natural Sciences and Engineering Research Council of Canada (NSERC). T.J.B. acknowledges funding support from the NASA Next Generation Space Telescope Flight Investigations program (now JWST) via WBS 411672.07.04.01.02. This research was supported in part through computational resources and services provided by Advanced Research Computing at the University of Michigan, Ann Arbor. This research has made use of the NASA Exoplanet Archive, which is operated by the California Institute of Technology, under contract with the National Aeronautics and Space Administration under the Exoplanet Exploration Program. This research has made use of NASA’s Astrophysics Data System.

Facilities: JWST, MAST, ADS.

Software: Eureka! (T. Bell et al. 2022), exoTDRF (A. D. Feinstein et al. 2023; M. Radica et al. 2023; M. Radica 2024), FIREFLY (Z. Rustamkulov et al. 2022, 2023), BeAR (D. Kitzmann et al. 2020), POSEIDON (R. J. MacDonald & N. Madhusudhan 2017; R. J. MacDonald 2023), Photochem (N. Wogan 2024; N. F. Wogan et al. 2024), PICASO (N. E. Batalha et al. 2019b; S. Mukherjee et al. 2023), VULCAN (S.-M. Tsai et al. 2017, 2021b), HELIOS (M. Malik et al. 2017, 2019), IPython (F. Perez & B. E. Granger 2007), astropy (Astropy Collaboration et al. 2013, 2018, 2022), batman (L. Kreidberg 2015), dynesty (J. S. Speagle 2020), emcee (D. Foreman-Mackey et al. 2013) ExoTiC-LD (D. Grant & H. Wakeford 2024), lacosmic (P. G. van Dokkum 2001), lmfit (M. Newville et al. 2024), matplotlib (J. D. Hunter 2007), numpy (C. R. Harris et al. 2020), PyMultiNest (F. Feroz et al. 2009; J. Buchner et al. 2014), scipy (P. Virtanen et al. 2020; E. Jones et al. 2024).

Appendix A

Reproduction of the Retrieval Results from N. Madhusudhan et al. (2023)

Here, we show that our atmospheric retrieval framework can reproduce consistent results to those presented in N. Madhusudhan et al. (2023), if we use the same full-resolution data presented in that study. This demonstrates that the nondetections of CO₂ and DMS presented above are primarily driven by differences in the data reduction and/or light-curve fitting approaches from our respective studies, rather than differences in the atmospheric modeling and/or retrieval approaches.

We used POSEIDON to replicate the results of N. Madhusudhan et al. (2023). For this purpose, we used the exact same atmospheric model configuration as the “canonical” model described in N. Madhusudhan et al. (2023). To summarize, this retrieval model fits for the log₁₀ volume mixing ratios of H₂O, CH₄, NH₃, HCN, CO, CO₂, DMS, CS₂, CH₃Cl, OCS, and N₂O, the six-parameter P-T profile from N. Madhusudhan & S. Seager (2009), the four-parameter inhomogeneous cloud and haze parameterization from R. J. MacDonald & N. Madhusudhan (2017), and the reference pressure corresponding to 2.61 $R_{\oplus}$. We consider three retrieval models: (i) no offset between the NIRISS and NIRSpec G395H data; (ii) one offset between NIRISS and NIRSpec G395H; and (iii) two offsets, one offset between NIRISS and the NIRSpec G395H NRS1 detector and one offset between NIRISS and the NIRSpec G395H NRS2 detector. Therefore, the three models have 22, 23, and 24 free parameters, respectively. We use identical priors for these parameters as in Table 4 from N. Madhusudhan et al. (2023). For fixed system properties (e.g., planetary mass, stellar radius), we used the values from B. Benneke et al. (2019) for consistency with N. Madhusudhan et al. (2023).

Several important retrieval model settings were not mentioned by N. Madhusudhan et al. (2023). Therefore, we adopted conservative choices to ensure the reliability of our retrieval results. Our opacities are sampled on a model wavelength grid with a resolution of $R = \lambda/\Delta\lambda = 100,000$, ensuring negligible errors from opacity sampling. We used 100 atmospheric layers from 10^{−6}–10 bar, spaced uniformly in log-pressure. We note that the minimum atmospheric pressure used here differs from the value of 10^{−8} bar we adopt for our main retrieval results, but we

Table 6

Bayesian Model Comparison Using the N. Madhusudhan et al. (2023) Data

Retrieval Model	Bayesian Evidence ($\ln \mathcal{Z}$)	Bayes Factor ($\mathcal{B}_{\text{Ref}, i}$)	Detection Significance	Classification
No Offset				
Ref.	28,826.8	...	...	...
No CH ₄	28,818.2	5541	4.5 σ	Strong Evidence
No CO ₂	28,825.7	3.1	2.1 σ	Weak Evidence
No DMS	28,825.6	3.2	2.1 σ	Weak Evidence
One Offset				
Ref.	28,830.5	...	...	...
No CH ₄	28,820.4	25,848	4.9 σ	Strong Evidence
No CO ₂	28,828.9	5.2	2.4 σ	Weak Evidence
No DMS	28,830.2	1.3	N/A	No Evidence
Two Offsets				
Ref.	28,829.9	...	...	...
No CH ₄	28,819.6	30,638	4.9 σ	Strong Evidence
No CO ₂	28,828.0	6.4	2.5 σ	Weak Evidence
No DMS	28,829.8	1.1	N/A	No Evidence

Note. $\mathcal{Z}$ is the Bayesian evidence of each retrieval, $\mathcal{B}_{\text{Ref}, i}$ is the Bayes factor between the reference model and the nested model without molecule “ i ,” and the “classification” follows the Jeffreys’ scale (e.g., R. Trotta 2008).

chose 10^{−6} bar here to match the value implied in Table 4 of N. Madhusudhan et al. (2023). The line lists in POSEIDON (described in the main text and Table 3) do differ from those in N. Madhusudhan in N. Madhusudhan et al. (2023) for several important molecules (in particular CH₄ and CO₂), but we stress that POSEIDON v1.2 is a state-of-the-art opacity database with the latest ExoMol line lists (as of 2024). Finally, we used 1000 MultiNest live points to ensure fine sampling of the parameter space.

Table 6 summarizes the retrieval model statistics from our analysis using the data from N. Madhusudhan et al. (2023). We find that only CH₄ is robustly detected, with a detection significance >4.5 σ across the different offset treatments. None of our retrievals provide notable evidence for CO₂ or DMS, with peak Bayes factors of ≈ 6 and ≈ 3 , respectively (compared to $\mathcal{B}_{\text{Ref}, i} > 12$ for “moderate evidence” and $\mathcal{B}_{\text{Ref}, i} > 150$ for “strong evidence” on the Jeffreys’ scale, see R. Trotta 2008). Our retrievals favor a single offset between the NIRISS SOSS and NIRSpec G395H data (44⁺¹³_{−12} ppm), consistent with N. Madhusudhan et al. (2023). The statistically preferred “one offset” model has a Bayes factor of ≈ 5 for CO₂ (2.4 σ), corresponding to “weak evidence” at best, while there is no evidence for DMS (Bayes factor ≈ 1).

Figure 15 provides a reproduction of N. Madhusudhan et al. (2023)’s Figure 4. We find broadly consistent posterior distributions to those in N. Madhusudhan et al. (2023) when using their data. Our CH₄ abundances are higher by ≈ 0.8 dex, but consistent within 1 σ . We note that our CO₂ posterior for the “no offset” case has a more prominent tail to low abundances than Figure 4 in N. Madhusudhan et al. (2023). This difference arises because N. Madhusudhan et al. (2023) renormalized the probability densities for each retrieval in their Figure 4 to have the same peak probability density, while our Figure 15 preserves the relative probability density differences between the retrievals. We highlight that our visualization method has the advantage of ensuring the same integrated

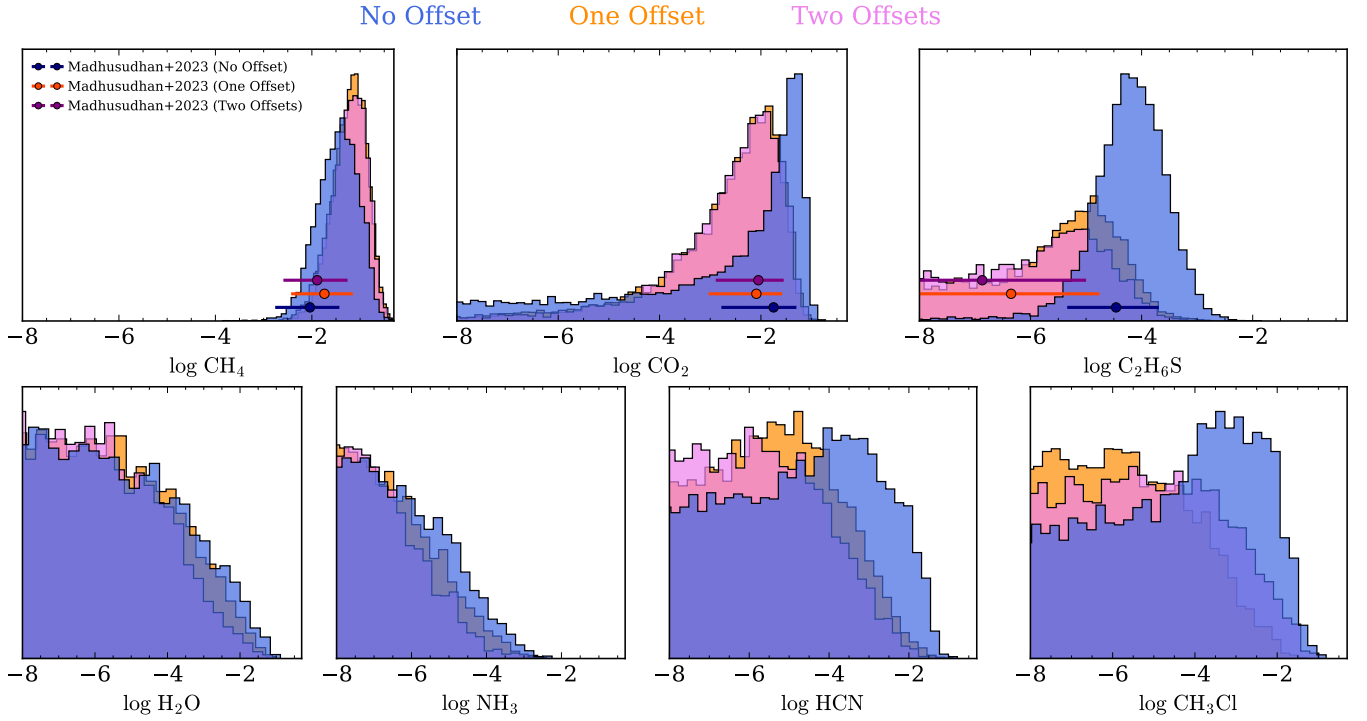


Figure 15. Reproduction of the results from N. Madhusudhan et al. (2023) using their full-resolution NIRISS + NIRSpec data. Retrieval results are overplotted for three retrieval models: (i) with no offset between NIRISS and NIRSpec (blue), (ii) an offset between NIRISS and NIRSpec (orange), and (iii) two offsets, one between NIRISS and NIRSpec G395H NRS1 and one between NIRISS and NIRSpec G395H NRS2 (violet). The statistically favored retrieval model has a single offset. Our POSEIDON retrieval framework produces consistent results with N. Madhusudhan et al. (2023) when using the same priors and data. Any differences in detections for our other retrievals are therefore due to differences in data reduction rather than differences inherent to our retrieval approach.

area for each distribution (i.e., total probability), which N. Madhusudhan et al. (2023)’s Figure 4 does not respect. We stress this point because it artificially suppresses the CO_2 tail for the “no offset” case in N. Madhusudhan et al. (2023)’s Figure 4, which consequently oversells the confidence of their claimed CO_2 detection.

Finally, Figure 16 shows correlations between notable atmospheric properties from these retrievals on the N. Madhusudhan

et al. (2023) data. We highlight a strong correlation between the CO_2 abundance and the temperature at the top of the atmosphere (T_{ref}), with most of the parameter space that favors a high CO_2 abundance corresponding to temperatures far in excess of the planetary equilibrium temperature (≈ 250 K). Thus, even if one assumes CO_2 is present in K2-18 b’s atmosphere, the planet would be far too warm to support habitable liquid water surface conditions.

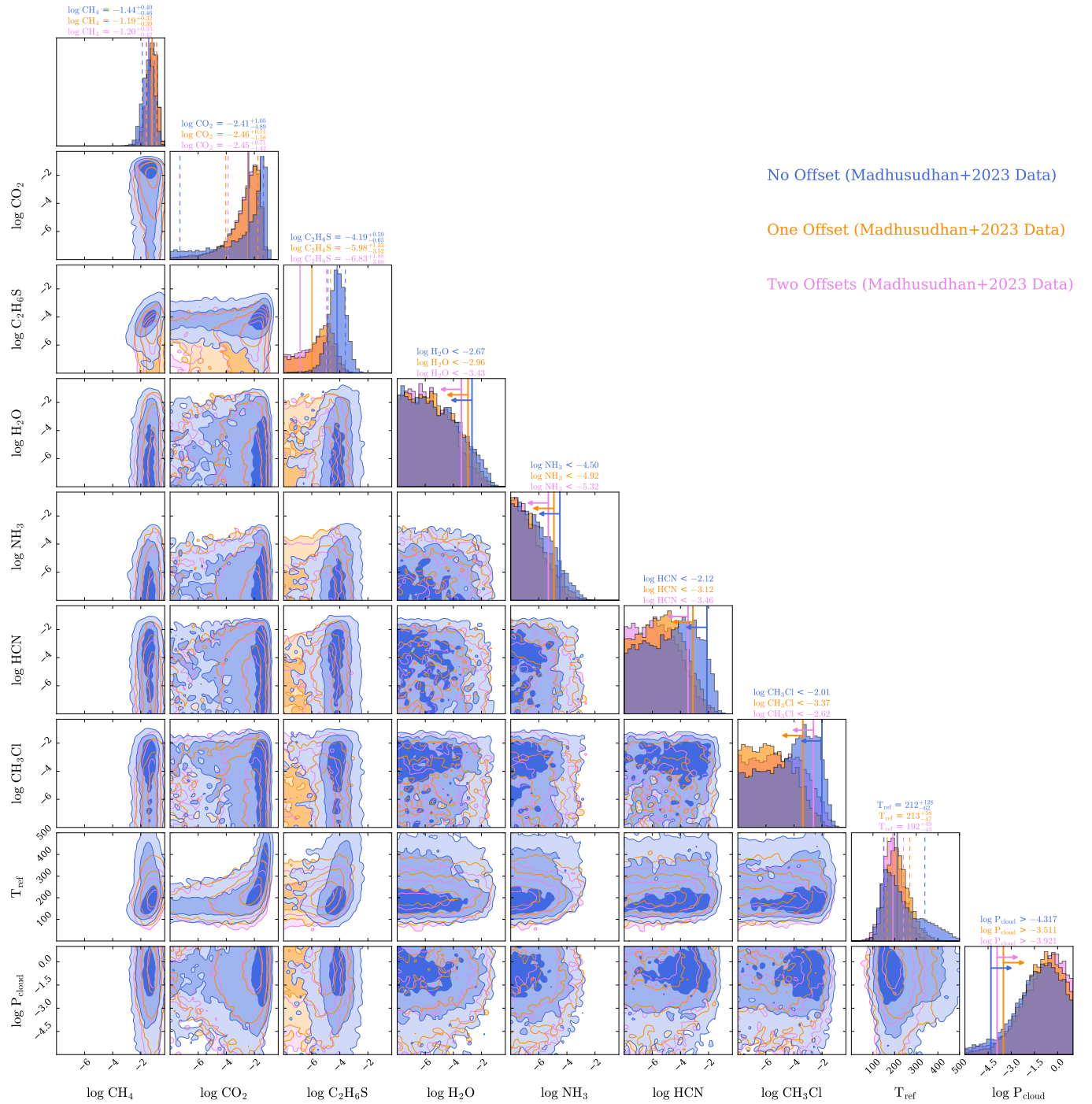


Figure 16. Corner plots corresponding to Figure 15, showing correlations between key atmospheric parameters. Parameters with only an upper or lower limit are marked by a vertical line and arrow at the 95% credible interval. We include 1σ credible regions for CO_2 and DMS , even though they are not robustly detected, for comparison with N. Madhusudhan et al. (2023).

Appendix B

Spectrophotometric Light-curve Fits with FIREFLY

Here, we present the spectroscopic light-curve fits and residuals for our $R \approx 100$ FIREFLY reduction. We show the NRS1 fits in Figure 17 and the NRS2 fits in Figure 18.

Additionally, we show the residuals of the NRS1 fits in Figure 19 and the residuals of the NRS2 fits in Figure 20. Each figure's panels all have the same x and y limits. The lack of trends in each bin demonstrates the fidelity of our spectrophotometric fits.

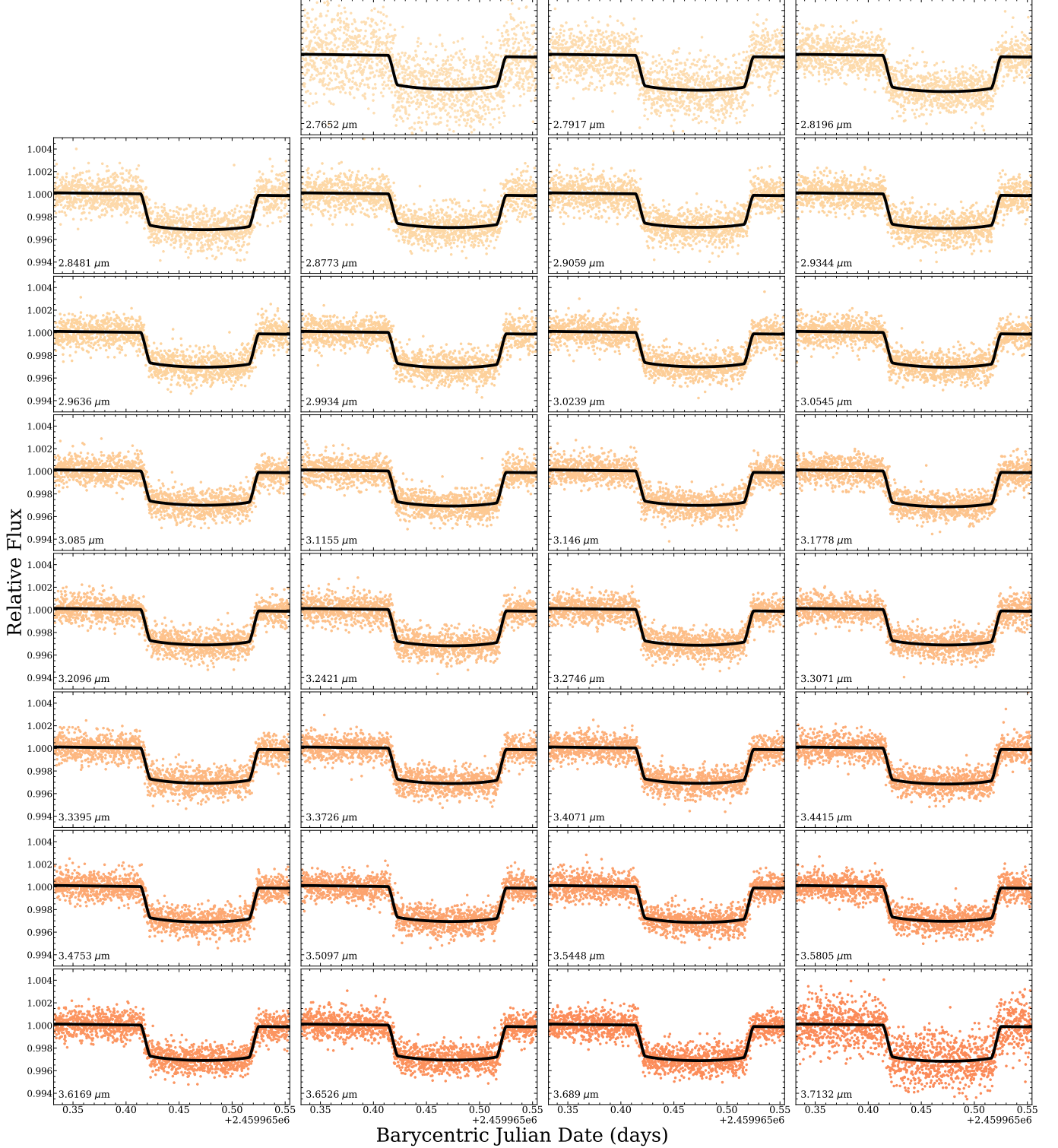


Figure 17. FIREFLY Spectroscopic light-curve fits for NRS1.

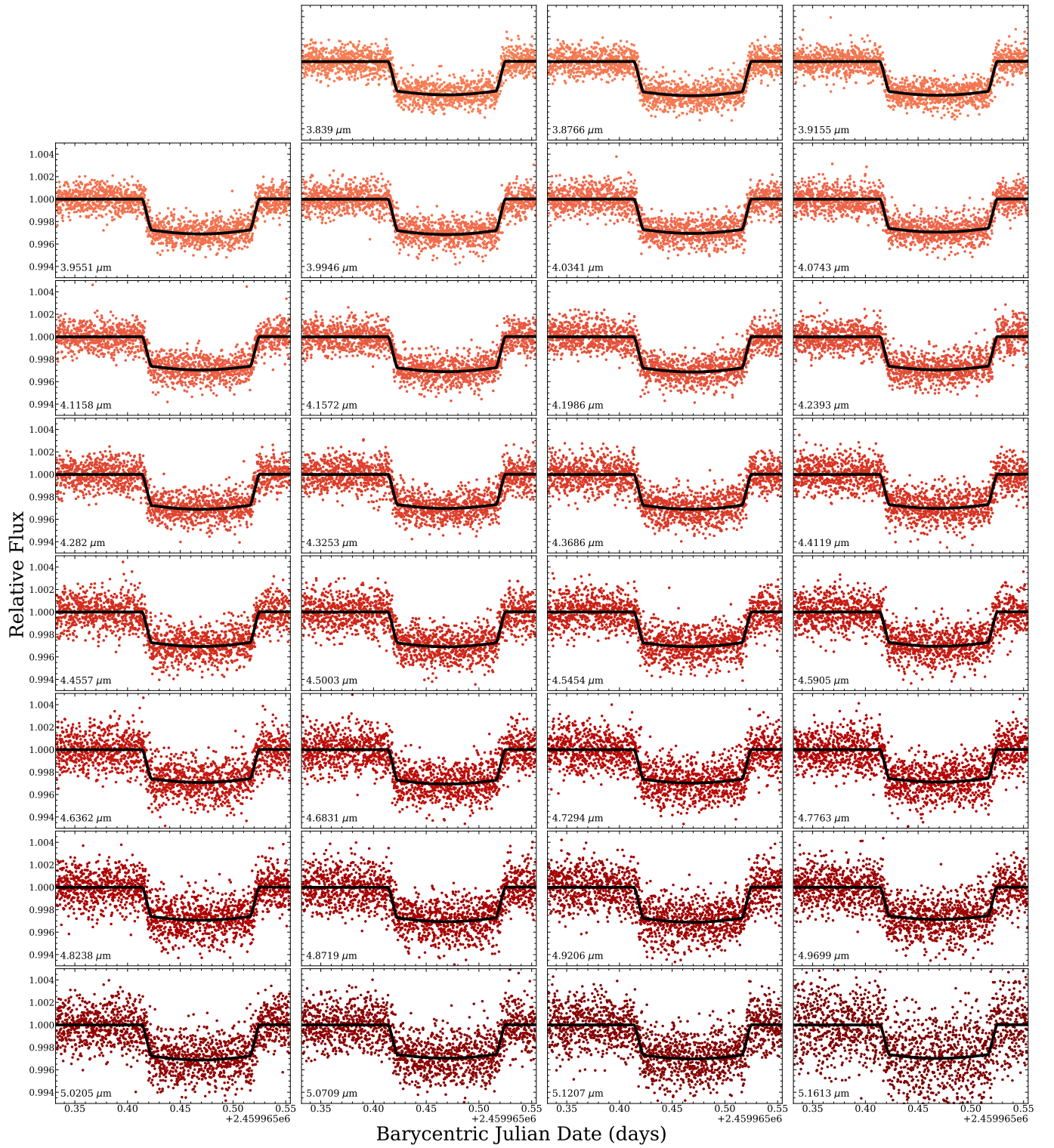


Figure 18. FIREFLY Spectroscopic light-curve fits for NRS2.

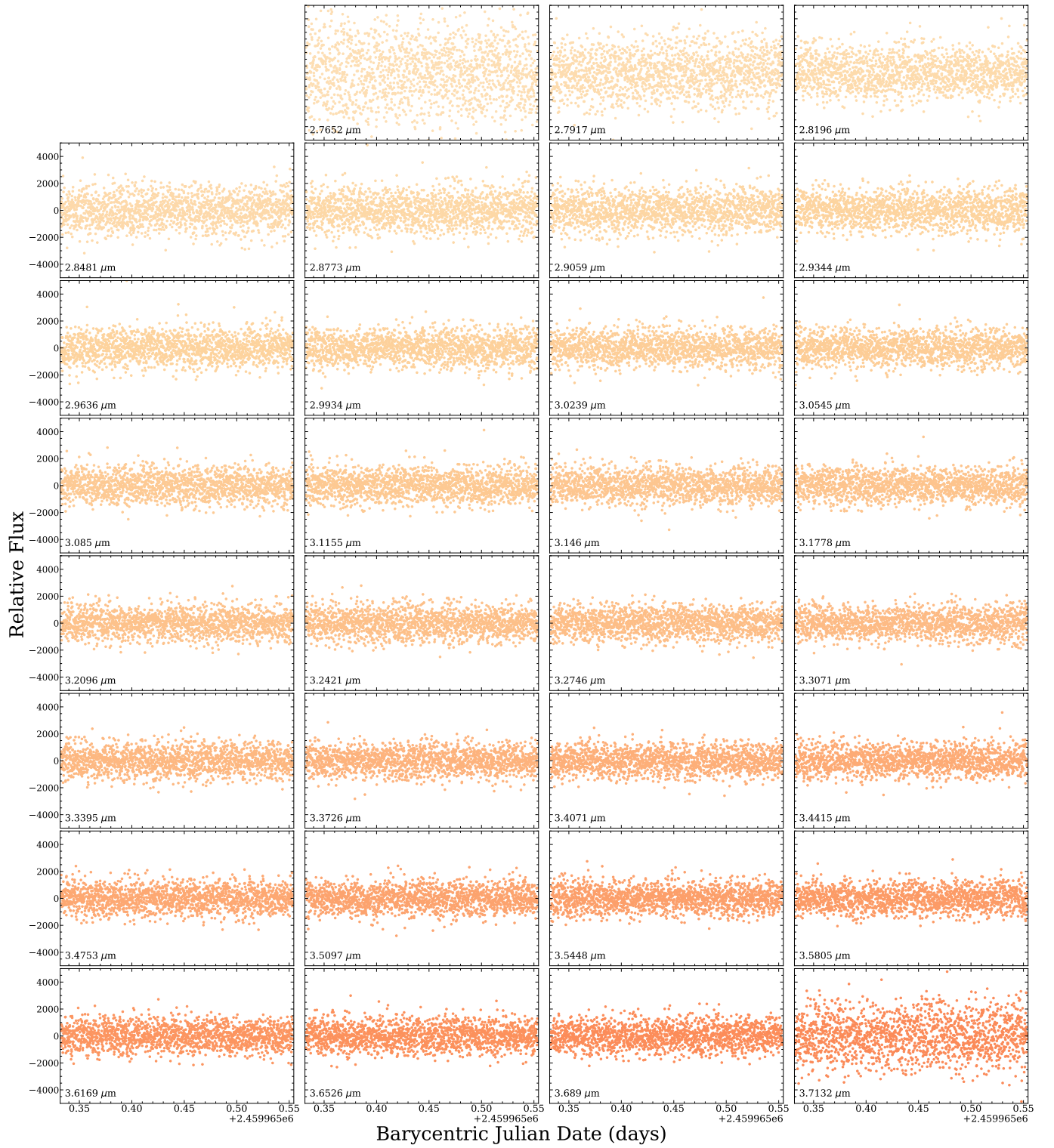


Figure 19. FIREFLY Spectroscopic light-curve residuals for NRS1.

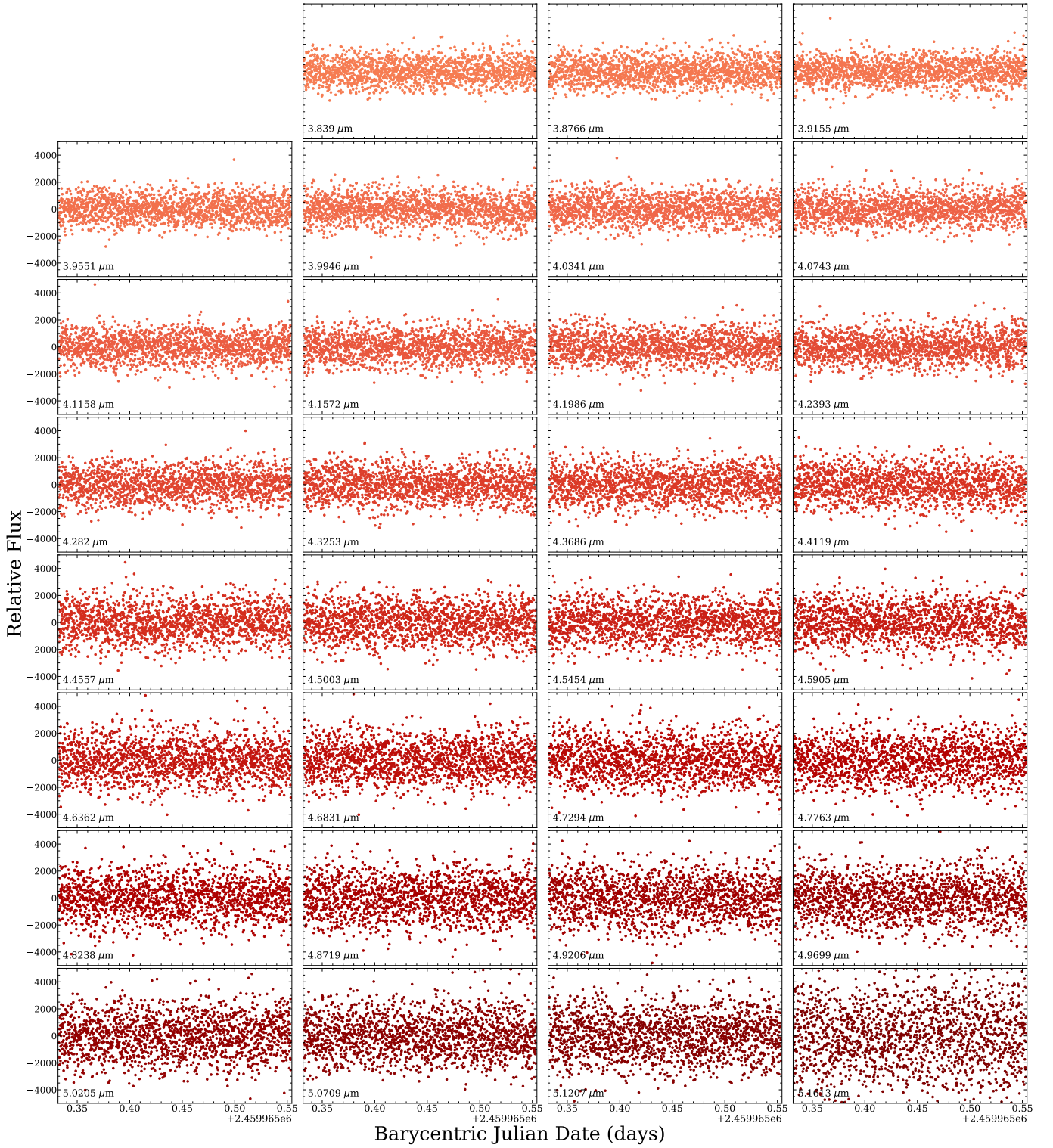


Figure 20. FIREFLY Spectroscopic light-curve residuals for NRS2.

Appendix C

Significance Comparison between NIRSpec and NIRISS Reductions

For illustrative purposes, we perform a comparison between reductions in which we calculate the significance of the discrepancy in the transmission spectrum between the FIREFLY NIRSpec reduction and the other three reductions at $R \approx 100$, $R \approx 200$, and $R \approx 300$. We do the same for NIRISS,

but at $R \approx 25$, $R \approx 100$, and 2 pixel binning. We show the results of this in Figure 21. We attribute the higher number of $>2\sigma$ outliers at shorter wavelengths between our $R \approx 25$ NIRISS reductions to slight differences in our NIRISS binning scheme that are more pronounced there. Our reductions are consistent with one another, typically within a 1σ to 1.5σ range of the FIREFLY reduction. The higher resolving power reductions have more outliers, but we expect this due to the

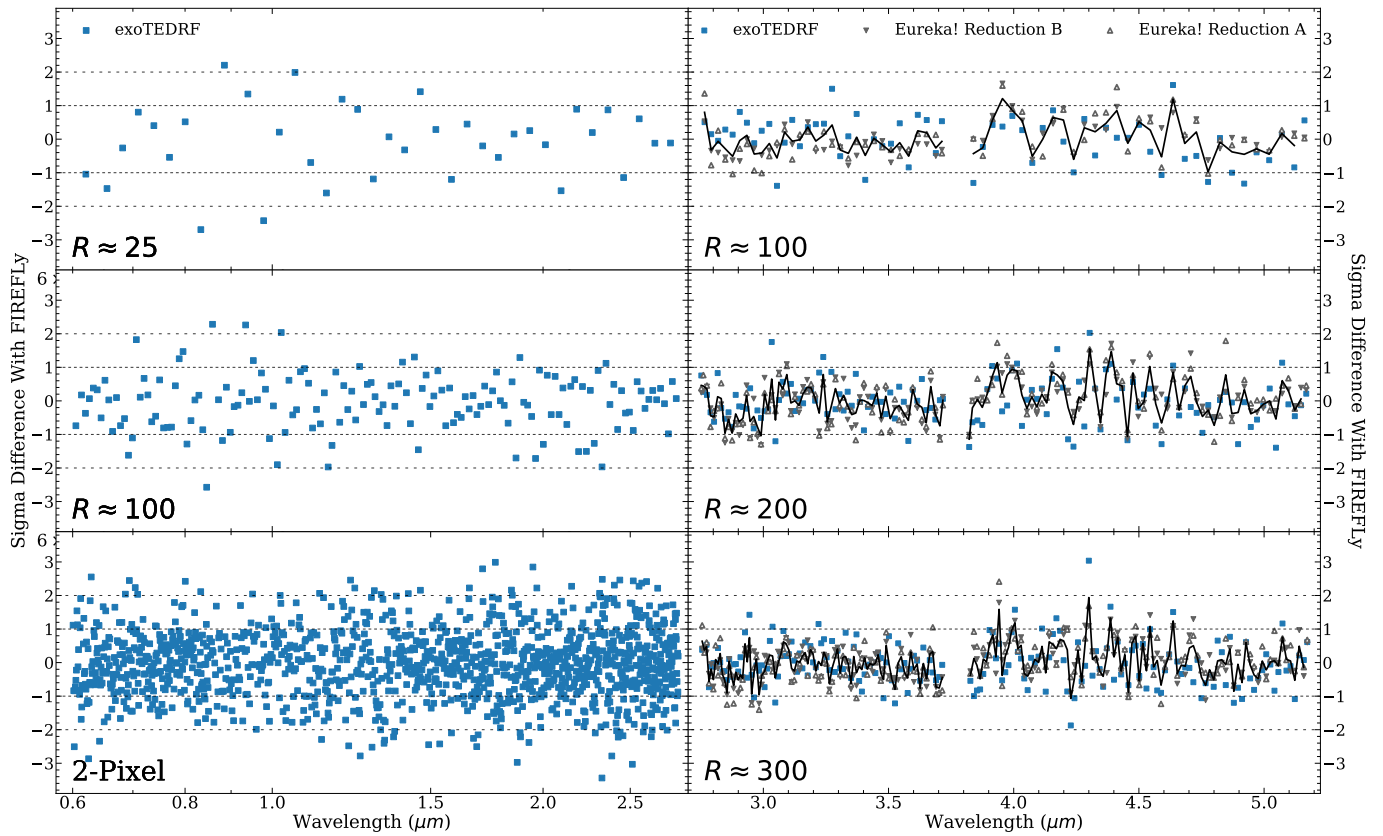


Figure 21. Significance of the discrepancy in transit depths between the FIREFLY reduction and the three other independent reductions. On the left three panels, we plot the significance between NIRISS reduction at $R \approx 25$ (top), $R \approx 100$ (middle), and at the 2 pixel level (bottom). On the right three panels, we plot the significance between NIRSpec reductions at $R \approx 100$ (top), $R \approx 200$ (middle), and $R \approx 300$ (bottom). Using the same color scheme as Figure 2, we plot as blue squares the exoTEDRF reduction and as gray triangles the Eureka! reductions, distinguishing between the two with unfilled downward-pointing (A) and filled upward-pointing (B) triangles. We plot as a black solid line on the right three panels the average deviation between the three, and as gray horizontal lines the 1σ (densely dashed) and 2σ (sparsely dashed) cutoffs. Our reductions are consistent with one another, although the rate of $>2\sigma$ outliers increases with higher resolution.

larger number of data points. This indicates that the differing choices we apply in our reductions each yield consistent results and are reasonable for the purposes of marginalizing over in our retrieval analysis.

ORCID iDs

Stephen P. Schmidt <https://orcid.org/0000-0001-8510-7365>
 Ryan J. MacDonald <https://orcid.org/0000-0003-4816-3469>
 Shang-Min Tsai <https://orcid.org/0000-0002-8163-4608>
 Michael Radica <https://orcid.org/0000-0002-3328-1203>
 Le-Chris Wang <https://orcid.org/0000-0002-6379-3816>
 Eva-Maria Ahrer <https://orcid.org/0000-0003-0973-8426>
 Taylor J. Bell <https://orcid.org/0000-0003-4177-2149>
 Chloe Fisher <https://orcid.org/0000-0003-0652-2902>
 Daniel P. Thorngren <https://orcid.org/0000-0002-5113-8558>
 Nicholas Wogan <https://orcid.org/0000-0002-0413-3308>
 Erin M. May <https://orcid.org/0000-0002-2739-1465>
 Piero Ferrari <https://orcid.org/0000-0001-6096-7772>
 Katherine A. Bennett <https://orcid.org/0000-0002-9030-0132>
 Zafar Rustamkulov <https://orcid.org/0000-0003-4408-0463>
 Mercedes López-Morales <https://orcid.org/0000-0003-3204-8183>
 David K. Sing <https://orcid.org/0000-0001-6050-7645>

References

- Abel, M., Frommhold, L., Li, X., & Hunt, K. L. C. 2011, *JPCA*, **115**, 6805
 Abel, M., Frommhold, L., Li, X., & Hunt, K. L. C. 2012, *JChPh*, **136**, 044319
 Agrawal, A., & MacDonald, R. 2024, *JOSS*, **9**, 6894
 Alam, M. K., Gao, P., Adams Redai, J., et al. 2025, *AJ*, **169**, 15
 Albert, L., Lafrenière, D., René, D., et al. 2023, *PASP*, **135**, 075001
 Alderson, L., Batalha, N. E., Wakeford, H. R., et al. 2024, *AJ*, **167**, 216
 Alderson, L., Moran, S. E., Wallack, N. L., et al. 2025, *AJ*, **169**, 142
 Alderson, L., Wakeford, H. R., Alam, M. K., et al. 2023, *Natur*, **614**, 664
 Asplund, M., Grevesse, N., Sauval, A. J., & Scott, P. 2009, *ARA&A*, **47**, 481
 Astropy Collaboration, Price-Whelan, A. M., Lim, P. L., et al. 2022, *ApJ*, **935**, 167
 Astropy Collaboration, Price-Whelan, A. M., Sipőcz, B. M., et al. 2018, *AJ*, **156**, 123
 Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., et al. 2013, *A&A*, **558**, A33
 Baines, T., Espinoza, N., Filippazzo, J., & Volk, K. 2023, Technical Report JWST-STScI-008448,
 Banerjee, A., Barstow, J. K., Gressier, A., et al. 2024, *ApJL*, **975**, L11
 Barber, R. J., Strange, J. K., Hill, C., et al. 2014, *MNRAS*, **437**, 1828
 Barclay, T., Kostov, V. B., Colón, K. D., et al. 2021, *AJ*, **162**, 300
 Barnes, I., Hjorth, J., Mihalopoulos, N., et al. 2006, *ChRv*, **106**, 940
 Barstow, J. K., Changeat, Q., Garland, R., et al. 2020, *MNRAS*, **493**, 4884
 Batalha, N., Rooney, C., Mukherjee, S., et al. 2024, Picaso: Release v3.3, Zenodo, doi:10.5281/zenodo.14160128
 Batalha, N. E., Lewis, T., Fortney, J. J., et al. 2019a, *ApJL*, **885**, L25
 Batalha, N. E., Marley, M. S., Lewis, N. K., & Fortney, J. J. 2019b, *ApJ*, **878**, 70
 Beatty, T. G., Welbanks, L., Schlawin, E., et al. 2024, *ApJL*, **970**, L10
 Béky, B., Kipping, D. M., Holman, M. J., et al. 2014, *MNRAS*, **442**, 3686
 Bell, T., Ahrer, E.-M., Brande, J., et al. 2022, *JOSS*, **7**, 4503

- Benneke, B., Roy, P.-A., Coulombe, L.-P., et al. 2024, arXiv:2403.03325
- Benneke, B., & Seager, S. 2013, *ApJ*, **778**, 153
- Benneke, B., Werner, M., Petigura, E., et al. 2017, *ApJ*, **834**, 187
- Benneke, B., Wong, I., Piaulet, C., et al. 2019, *ApJL*, **887**, L14
- Bergin, E. A., Kempton, E. M. R., Hirschmann, M., et al. 2023, *ApJL*, **949**, L17
- Bethkenhagen, M., Meyer, E. R., Hamel, S., et al. 2017, *ApJ*, **848**, 67
- Birkmann, S. M., Ferruit, P., Alves de Oliveira, C., et al. 2014, *Proc. SPIE*, **9143**, 914308
- Blain, D., Charnay, B., & Bézard, B. 2021, *A&A*, **646**, A15
- Buchner, J., Georgakakis, A., Nandra, K., et al. 2014, *A&A*, **564**, A125
- Cadieux, C., Doyon, R., MacDonald, R. J., et al. 2024, *ApJL*, **970**, L2
- Chabrier, G., & Debras, F. 2021, *ApJ*, **917**, 4
- Charlson, R. J., Warren, S. G., Lovelock, J. E., & Andreae, M. O. 1987, *Natur*, **326**, 655
- Cloutier, R., Astudillo-Defru, N., Doyon, R., et al. 2017, *A&A*, **608**, A35
- Cloutier, R., Astudillo-Defru, N., Doyon, R., et al. 2019, *A&A*, **621**, A49
- Coles, P. A., Yurchenko, S. N., & Tennyson, J. 2019, *MNRAS*, **490**, 4638
- Cooke, G. J., & Madhusudhan, N. 2024, *ApJ*, **977**, 209
- Coulombe, L.-P., Benneke, B., Challener, R., et al. 2023, *Natur*, **620**, 292
- Damiano, M., Bello-Arufe, A., Yang, J., & Hu, R. 2024, *ApJL*, **968**, L22
- Darveau-Bernier, A., Albert, L., Talens, G. J., et al. 2022, *PASP*, **134**, 094502
- Davenport, B., Kempton, E. M. R., Nixon, M. C., et al. 2025, *ApJ*, **984**, L44
- Doyon, R., Hutchings, J. B., Beaulieu, M., et al. 2012, *Proc. SPIE*, **8442**, 84422R
- Doyon, R., Willott, C. J., Hutchings, J. B., et al. 2023, *PASP*, **135**, 098001
- Dyrek, A., Min, M., Decin, L., et al. 2024, *Natur*, **625**, 51
- Feinstein, A. D., Radica, M., Welbanks, L., et al. 2023, *Natur*, **614**, 670
- Feroz, F., Hobson, M. P., & Bridges, M. 2009, *MNRAS*, **398**, 1601
- Ferruit, P., Bagnasco, G., Barho, R., et al. 2012, *Proc. SPIE*, **8442**, 84422O
- Fisher, C., Taylor, J., Parmentier, V., et al. 2024, *MNRAS*, **535**, 27
- Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, *PASP*, **125**, 306
- Fortney, J. J., Marley, M. S., & Barnes, J. W. 2007, *ApJ*, **659**, 1661
- Fournier-Tondreau, M., MacDonald, R. J., Radica, M., et al. 2024, *MNRAS*, **528**, 3354
- Fournier-Tondreau, M., Pan, Y., Morel, K., et al. 2025, *MNRAS*, **539**, 422
- Fu, G., Stevenson, K. B., Sing, D. K., et al. 2025, *ApJ*, **986**, 1
- Glein, C. R. 2024, *ApJL*, **964**, L19
- Gordon, I. E., Rothman, L. S., Hargreaves, R. J., et al. 2022, *JQSRT*, **277**, 107949
- Grant, D., Lewis, N. K., Wakeford, H. R., et al. 2023, *ApJL*, **956**, L32
- Grant, D., & Wakeford, H. 2024, *JOSS*, **9**, 6816
- Gressier, A., MacDonald, R. J., Espinoza, N., et al. 2025, *AJ*, **169**, 57
- Gupta, A., Stixrude, L., & Schlichting, H. E. 2025, *ApJL*, **982**, L35
- Hänni, N., Altwegg, K., Combi, M., et al. 2024, *ApJ*, **976**, 74
- Hargreaves, R. J., Gordon, I. E., Rey, M., et al. 2020, *ApJS*, **247**, 55
- Hargreaves, R. J., Gordon, I. E., Rothman, L. S., et al. 2019, *JQSRT*, **232**, 35
- Harris, C. R., Millman, K. J., van der Walt, S. J., et al. 2020, *Natur*, **585**, 357
- Harris, G. J., Tennyson, J., Kaminsky, B. M., Pavlenko, Y. V., & Jones, H. R. A. 2006, *MNRAS*, **367**, 400
- Heng, K., & Kitzmann, D. 2017, *MNRAS*, **470**, 2972
- Holmberg, M., & Madhusudhan, N. 2023, *MNRAS*, **524**, 377
- Holmberg, M., & Madhusudhan, N. 2024, *A&A*, **683**, L2
- Horne, K. 1986, *PASP*, **98**, 609
- Hu, R., Damiano, M., Scheucher, M., et al. 2021, *ApJL*, **921**, L8
- Huang, X., Freedman, R. S., Tashkun, S. A., Schwenke, D. W., & Lee, T. J. 2013, *JQSRT*, **130**, 134
- Huang, X., Schwenke, D. W., Freedman, R. S., & Lee, T. J. 2017, *JQSRT*, **203**, 224
- Hunter, J. D. 2007, *CSE*, **9**, 90
- Husser, T. O., Wende-von Berg, S., Dreizler, S., et al. 2013, *A&A*, **553**, A6
- Inglis, J., Batalha, N. E., Lewis, N. K., et al. 2024, *ApJL*, **973**, L41
- Innes, H., Tsai, S.-M., & Pierrehumbert, R. T. 2023, *ApJ*, **953**, 168
- Jones, E., Oliphant, T., Peterson, P., et al. 2024, SciPy: Open Source Scientific Tools for Python, <http://www.scipy.org/>
- Jordan, S., Shorttle, O., & Quanz, S. P. 2025, arXiv:2504.12030
- Karkoschka, E., & Tomasko, M. G. 2011, *Icar*, **211**, 780
- Karman, T., Gordon, I. E., van der Avoird, A., et al. 2019, *Icar*, **328**, 160
- Kipping, D., & Benneke, B. 2025, arXiv:2506.05392
- Kipping, D. M. 2013, *MNRAS*, **435**, 2152
- Kirk, J., Ahrer, E.-M., Claringbold, B., et al. 2025, *MNRAS*, **537**, 3027
- Kite, E. S., & Ford, E. B. 2018, *ApJ*, **864**, 75
- Kitzmann, D., Heng, K., Oreshenko, M., et al. 2020, *ApJ*, **890**, 174
- Kitzmann, D., Stock, J. W., & Patzer, A. B. C. 2024, *MNRAS*, **527**, 7263
- Kostogryz, N., Shapiro, A. I., Witzke, V., et al. 2023, *RNAAS*, **7**, 39
- Kreidberg, L. 2015, *PASP*, **127**, 1161
- Kuskov, O. L., & Kronrod, V. A. 2005, *Icar*, **177**, 550
- Leconte, J., Spiga, A., Clément, N., et al. 2024, *A&A*, **686**, A131
- Li, G., Gordon, I. E., Rothman, L. S., et al. 2015, *ApJS*, **216**, 15
- Li, J., Bergin, E. A., Blake, G. A., Ciesla, F. J., & Hirschmann, M. M. 2021, *SciA*, **7**, eabd3632
- Lim, O., Benneke, B., Doyon, R., et al. 2023, *ApJL*, **955**, L22
- Liu, R., Wang, L.-C., Rustamkulov, Z., & Sing, D. K. 2025, *AJ*, **169**, 335
- Lodders, K. 2003, *ApJ*, **591**, 1220
- Louie, D. R., Mullens, E., Alderson, L., et al. 2025, *AJ*, **169**, 86
- Luque, R., Piaulet-Ghorayeb, C., Radica, M., et al. 2025, *A&A*, **700**, A284
- Lustig-Yaeger, J., Fu, G., May, E. M., et al. 2023, *NatAs*, **7**, 1317
- Luu, C. N., Yu, X., Glein, C. R., et al. 2024, *ApJL*, **977**, L51
- Lyon, S. P., & Johnson, J. D. 1992, Technical Report LA-UR-92-3407
- MacDonald, R. J. 2023, *JOSS*, **8**, 4873
- MacDonald, R. J., & Lewis, N. K. 2022, *ApJ*, **929**, 20
- MacDonald, R. J., & Madhusudhan, N. 2017, *MNRAS*, **469**, 1979
- Madhusudhan, N. 2012, *ApJ*, **758**, 36
- Madhusudhan, N. 2023, JWST Transit Spectroscopy of K2-18 b, 1.0, MAST, doi:10.17909/3ds1-8z15
- Madhusudhan, N., Constantinou, S., Holmberg, M., et al. 2025, *ApJL*, **983**, L40
- Madhusudhan, N., Nixon, M. C., Welbanks, L., Piette, A. A. A., & Booth, R. A. 2020, *ApJL*, **891**, L7
- Madhusudhan, N., Piette, A. A. A., & Constantinou, S. 2021, *ApJ*, **918**, 1
- Madhusudhan, N., Sarkar, S., Constantinou, S., et al. 2023, *ApJL*, **956**, L13
- Madhusudhan, N., & Seager, S. 2009, *ApJ*, **707**, 24
- Mai, C., & Line, M. R. 2019, *ApJ*, **883**, 144
- Malik, M., Groscheintz, L., Mendonça, J. M., et al. 2017, *AJ*, **153**, 56
- Malik, M., Kitzmann, D., Mendonça, J. M., et al. 2019, *AJ*, **157**, 170
- Marcus, R. A., Sasselov, D., Stewart, S. T., & Hernquist, L. 2010, *ApJL*, **719**, L45
- May, E. M., MacDonald, R. J., Bennett, K. A., et al. 2023, *ApJL*, **959**, L9
- Mazevet, S., Licari, A., Chabrier, G., & Potekhin, A. Y. 2019, *A&A*, **621**, A128
- Mollière, P., van Boekel, R., Dullemond, C., Henning, T., & Mordasini, C. 2015, *ApJ*, **813**, 47
- Montet, B. T., Morton, T. D., Foreman-Mackey, D., et al. 2015, *ApJ*, **809**, 25
- Moran, S. E., Stevenson, K. B., Sing, D. K., et al. 2023, *ApJL*, **948**, L11
- Mukherjee, S., Batalha, N. E., Fortney, J. J., & Marley, M. S. 2023, *ApJ*, **942**, 71
- Mukherjee, S., Fortney, J. J., Wogan, N. F., Sing, D. K., & Ohno, K. 2025, *ApJ*, **985**, 209
- Mullens, E., Lewis, N. K., & MacDonald, R. J. 2024, *ApJ*, **977**, 105
- Nettelmann, N., Wang, K., Fortney, J. J., et al. 2016, *Icar*, **275**, 107
- Newville, M., Otten, R., Nelson, A., et al. 2024, Imfit/Imfit-py: 1.3.2, Zenodo, doi:10.5281/zenodo.598352
- Niraula, P., de Wit, J., Gordon, I. E., Hargreaves, R. J., & Sousa-Silva, C. 2023, *ApJL*, **950**, L17
- Niraula, P., de Wit, J., Gordon, I. E., et al. 2022, *NatAs*, **6**, 1287
- Nixon, M. C., & Madhusudhan, N. 2021, *MNRAS*, **505**, 3414
- Nixon, M. C., Piette, A. A. A., Kempton, E. M. R., et al. 2024b, *ApJL*, **970**, L28
- Nixon, M. C., Welbanks, L., McGill, P., & Kempton, E. M. R. 2024a, *ApJ*, **966**, 156
- Owens, A., Yachmenev, A., Thiel, W., et al. 2018, *MNRAS*, **479**, 3002
- Owens, A., Yurchenko, S. N., & Tennyson, J. 2024, *MNRAS*, **530**, 4004
- Perez, F., & Granger, B. E. 2007, *CSE*, **9**, 21
- Piaulet-Ghorayeb, C., Benneke, B., Radica, M., et al. 2024, *ApJL*, **974**, L10
- Pica-Ciamarra, L., Madhusudhan, N., Cooke, G. J., Constantinou, S., & Binet, M. 2025, arXiv:2505.10539
- Pierrehumbert, R. T. 2023, *ApJ*, **944**, 20
- Piette, A. A. A., & Madhusudhan, N. 2020, *ApJ*, **904**, 154
- Poliansky, O. L., Kyuberis, A. A., Zobov, N. F., et al. 2018, *MNRAS*, **480**, 2597
- Pont, F., Zucker, S., & Queloz, D. 2006, *MNRAS*, **373**, 231
- Powell, D., Feinstein, A. D., Lee, E. K. H., et al. 2024, *Natur*, **626**, 979
- Rackham, B. V., Apai, D., & Giampapa, M. S. 2018, *ApJ*, **853**, 122
- Radica, M. 2024, *JOSS*, **9**, 6898
- Radica, M. 2024, radcamc/exoUPRF: v1.0.1, Zenodo, doi:10.5281/zenodo.12628066
- Radica, M., Albert, L., Taylor, J., et al. 2022b, *PASP*, **134**, 104502
- Radica, M., Artigau, É., Lafrenière, D., et al. 2022a, *MNRAS*, **517**, 5050
- Radica, M., Coulombe, L.-P., Taylor, J., et al. 2024, *ApJL*, **962**, L20
- Radica, M., Piaulet-Ghorayeb, C., Taylor, J., et al. 2025, *ApJL*, **979**, L5
- Radica, M., Welbanks, L., Espinoza, N., et al. 2023, *MNRAS*, **524**, 835

- Rathcke, A. D., MacDonald, R. J., Barstow, J. K., et al. 2021, *AJ*, **162**, 138
- Rigby, F. E., Pica-Ciamarra, L., Holmberg, M., et al. 2024, *ApJ*, **975**, 101
- Rustamkulov, Z., Sing, D. K., Liu, R., & Wang, A. 2022, *ApJL*, **928**, L7
- Rustamkulov, Z., Sing, D. K., Mukherjee, S., et al. 2023, *Natur*, **614**, 659
- Sarkis, P., Henning, T., Kürster, M., et al. 2018, *AJ*, **155**, 257
- Scheucher, M., Wunderlich, F., Grenfell, J. L., et al. 2020, *ApJ*, **898**, 44
- Schlawin, E., Ohno, K., Bell, T. J., et al. 2024, *ApJL*, **974**, L33
- Schwieterman, E. W., Kiang, N. Y., Parenteau, M. N., et al. 2018, *AsBio*, **18**, 663
- Seager, S., Bains, W., & Hu, R. 2013, *ApJ*, **775**, 104
- Segura, A., Kasting, J. F., Meadows, V., et al. 2005, *AsBio*, **5**, 706
- Sharpe, S. W., Johnson, T. J., Sams, R. L., et al. 2004, *ApSpe*, **58**, 1452
- Shorttle, O., Jordan, S., Nicholls, H., Lichtenberg, T., & Bower, D. J. 2024, *ApJL*, **962**, L8
- Showman, A. P., & Malhotra, R. 1999, *Sci*, **296**, 77
- Sing, D. K., Rustamkulov, Z., Thorngren, D. P., et al. 2024, *Natur*, **630**, 831
- Soubiran, F., & Miltzer, B. 2015, *ApJ*, **806**, 228
- Speagle, J. S. 2020, *MNRAS*, **493**, 3132
- Stock, J. W., Kitzmann, D., & Patzer, A. B. C. 2022, *MNRAS*, **517**, 4070
- Tabernero, H. M., Shan, Y., Caballero, J. A., et al. 2024, *A&A*, **689**, A223
- Taylor, J. 2025, *RNAAS*, **9**, 118
- Taylor, J., Radica, M., Welbanks, L., et al. 2023, *MNRAS*, **524**, 817
- Thompson, S. L. 1990, *Technical Report* SAND-89-2951, 6939284, Sandia National Laboratory
- Thorngren, D. P. 2019, PhD thesis, Univ. California, Santa Cruz
- Thorngren, D. P., Fortney, J. J., Murray-Clay, R. A., & Lopez, E. D. 2016, *ApJ*, **831**, 64
- Townsend, R., & Lopez, A. 2023, *JOSS*, **8**, 4602
- Trotta, R. 2008, *ConPh*, **49**, 71
- Tsai, S.-M., Innes, H., Lichtenberg, T., et al. 2021a, *ApJL*, **922**, L27
- Tsai, S.-M., Innes, H., Wogan, N. F., & Schwieterman, E. W. 2024, *ApJL*, **966**, L24
- Tsai, S.-M., Lee, E. K. H., Powell, D., et al. 2023, *Natur*, **617**, 483
- Tsai, S.-M., Lyons, J. R., Grosheintz, L., et al. 2017, *ApJS*, **228**, 20
- Tsai, S.-M., Malik, M., Kitzmann, D., et al. 2021b, *ApJ*, **923**, 264
- Tsiaras, A., Waldmann, I. P., Tinetti, G., Tennyson, J., & Yurchenko, S. N. 2019, *NatAs*, **3**, 1086
- van Dokkum, P. G. 2001, *PASP*, **113**, 1420
- Vardya, M. S. 1962, *ApJ*, **135**, 303
- Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, *NatMe*, **17**, 261
- Waldmann, I. P., Tinetti, G., Rocchetto, M., et al. 2015, *ApJ*, **802**, 107
- Wallack, N. L., Batalha, N. E., Alderson, L., et al. 2024, *AJ*, **168**, 77
- Welbanks, L., Bell, T. J., Beatty, T. G., et al. 2024, *Natur*, **630**, 836
- Welbanks, L., & Madhusudhan, N. 2019, *AJ*, **157**, 206
- Welbanks, L., & Madhusudhan, N. 2021, *ApJ*, **913**, 114
- Welbanks, L., McGill, P., Line, M., & Madhusudhan, N. 2023, *AJ*, **165**, 112
- Welbanks, L., Nixon, M. C., McGill, P., et al. 2025, arXiv:2504.21788
- Wilzewski, J. S., Gordon, I. E., Kochanov, R. V., Hill, C., & Rothman, L. S. 2016, *JQSRT*, **168**, 193
- Winn, J. N., Holman, M. J., Torres, G., et al. 2008, *ApJ*, **683**, 1076
- Wogan, N. 2024, Photochem v0.6.2, Zenodo, doi:10.5281/zenodo.14032108
- Wogan, N. F., Batalha, N. E., Zahnle, K. J., et al. 2024, *ApJL*, **963**, L7
- Yang, J., & Hu, R. 2024, *ApJL*, **971**, L48
- Yip, K. H., Changeat, Q., Edwards, B., et al. 2021, *AJ*, **161**, 4
- Yu, X., Moses, J. I., Fortney, J. J., & Zhang, X. 2021, *ApJ*, **914**, 38
- Yurchenko, S. N., Mellor, T. M., Freedman, R. S., & Tennyson, J. 2020, *MNRAS*, **496**, 5282
- Yurchenko, S. N., Owens, A., Kefala, K., & Tennyson, J. 2024, *MNRAS*, **528**, 3719
- Yurchenko, S. N., & Tennyson, J. 2014, *MNRAS*, **440**, 1649