Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **531,** 2731–2756 (2024)
Advance Access publication 2024 May 11

https://doi.org/10.1093/mnras/stae1230

Exoplanet transit spectroscopy with *JWST* NIRSpec: diagnostics and homogeneous case study of WASP-39 b

Subhajit Sarkar [1★], Nikku Madhusudhan [2], Savvas Constantinou [2] and Måns Holmberg [2]

[1] *School of Physics and Astronomy, Cardiff University, Queen's Buildings, The Parade, Cardiff CF24 3AA, UK*
[2] *Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK*

Accepted 2024 May 4. Received 2024 May 4; in original form 2023 July 27

ABSTRACT

The *JWST* has ushered in a new era of exoplanet transit spectroscopy. Among the *JWST* instruments, the Near-Infrared Spectrograph (NIRSpec) has the most extensive set of configurations for exoplanet time-series observations. The NIRSpec Prism and G395H grating represent two extremes in NIRSpec instrument modes, with the Prism spanning a wider spectral range (0.6–5.3 μm) at lower resolution ($R \sim 100$) compared to G395H (2.87–5.14 μm; $R \sim 2700$). In this work, we develop a new data reduction framework, JexoPipe, to conduct a homogeneous assessment of the two NIRSpec modes for exoplanet spectroscopy. We use observations of the hot Saturn WASP-39 b obtained as part of the *JWST* Transiting Exoplanet Early Release Science programme to assess the spectral quality and stability between the two instrument modes at different epochs. We explore the noise sources, effect of saturation, and offsets in transmission spectra between the different instrument modes and also between the two G395H NRS detectors. We find an inter-detector offset in G395H of ∼40–50 ppm, consistent with recent studies. We find evidence for correlated noise in the Prism white light curve. We find the G395H spectrum to be of higher precision compared to the Prism spectrum at the same resolution. We also compare the JexoPipe spectra with those reported from other pipelines. Our work underscores the need for robust assessment of instrument performance and identification of optimal practices for *JWST* data reduction and analyses.

Key words: instrumentation: spectrographs – techniques: spectroscopic – planets and satellites: atmospheres.

1 INTRODUCTION

Exoplanetary atmospheres are key to understanding the physical conditions, chemical composition, and origins of exoplanets. Transmission spectroscopy (Seager & Sasselov 2000; Brown 2001) has emerged over the past two decades as the most widely applied technique used to obtain exoplanet spectra. Since the first detection of sodium in the upper atmosphere of the hot Jupiter HD 209458 b (Charbonneau et al. 2002), transmission spectroscopy has progressed with dozens of exoplanet atmospheres probed, and remains an extremely promising method of planetary remote sensing (Madhusudhan 2019). In particular, over the last decade, the Hubble Wide Field Camera 3 (WFC3) near-infrared (NIR) instrument G141 grism (1.1–1.7 μm) has delivered high-precision detections of H_2O in numerous hot Jupiters (e.g. Deming et al. 2013; McCullough et al. 2014; Sing et al. 2016) as well as in the sub-Neptune K2-18 b (Benneke et al. 2019; Tsiaras et al. 2019).

Transmission spectroscopy depends on wavelength-dependent absorption and scattering of stellar light by atmospheric atomic and molecular species as the planet transits in front of its host star. During the transit, a proportion of the star's light is effectively blocked in the line of sight by the atmosphere, which can be represented as an opaque annulus that adds an apparent extra height to the bulk radius of the planet. Due to the wavelength dependence of the various opacities, the transit depth varies with wavelength and thus the apparent $(R_p/R_s)^2$ (where R_p is the planet radius and R_s is the star radius). This technique probes the high-altitude atmosphere at the planet day–night terminator. In addition, atmospheric modelling and spectral retrieval methods have progressed to allow increasingly more sophisticated interpretation of these spectra (Madhusudhan & Seager 2009; Madhusudhan 2018). Transmission spectroscopy is particularly powerful in the visible and NIR wavelength ranges, where the host star flux is maximal (minimizing fractional photon noise), and where there are numerous atomic and molecular spectral signatures of molecules expected in planetary atmospheres.

The *JWST* promises to revolutionize our understanding of exoplanet atmospheres providing the highest quality transmission spectra ever obtained in terms of both precision (through its 6.5-m primary mirror in comparison to the 2.4-m Hubble primary mirror) and unparalleled wavelength coverage through its suite of four instruments: Near-Infrared Imager and Slitless Spectrograph (NIRISS; Doyon et al. 2012), Near-Infrared Camera (NIRCam; Beichman et al. 2012), Near-Infrared Spectrograph (NIRSpec; Ferruit et al. 2014), and Mid-Infrared Instrument (MIRI; Rieke et al. 2015). Combined these instruments provide a potential wavelength coverage ranging from 0.6 to 24 μm. In the past year, *JWST* has delivered several transmission spectra of exoplanets (e.g. Bell et al. 2023; Grant et al.

★ E-mail: subhajit.sarkar@astro.cf.ac.uk

© 2024 The Author(s).
Published by Oxford University Press on behalf of Royal Astronomical Society. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Downloaded from https://academic.oup.com/mnras/article/531/2/2731/7670622 by guest on 03 July 2024

2023; Kempton et al. 2023; Lim et al. 2023; Lustig-Yaeger et al. 2023; Madhusudhan et al. 2023; May et al. 2023; Moran et al. 2023; Dyrek et al. 2024; Kirk et al. 2024) in addition to those completed under the Early Release Science (ERS) (Bean et al. 2018) and Early Release Observations (ERO) (Pontoppidan et al. 2022) programmes (Ahrer et al. 2023; Alderson et al. 2023; Feinstein et al. 2023; JWST Transiting Exoplanet Community Early Release Science Team 2023; Radica et al. 2023; Rustamkulov et al. 2023; Taylor et al. 2023; Fournier-Tondreau et al. 2024). These observations have led to confident detections of several prominent molecules, such as CO_2, H_2O, and SO_2 in the hot Saturn WASP-39 b (e.g. Alderson et al. 2023; Constantinou, Madhusudhan & Gandhi 2023; Feinstein et al. 2023; Rustamkulov et al. 2023; Tsai et al. 2023), CH_4 in the warm Jupiter WASP-80 b (Bell et al. 2023), and CO_2 and CH_4 in the candidate Hycean world K2-18 b (Madhusudhan et al. 2023).

While only MIRI provides substantial wavelength coverage in the mid-infrared beyond 5 μm, there is a choice of three instruments in the NIR range of ∼1–5 μm. In addition, each of these NIR instruments has different configurations, i.e. combinations of dispersion element, filter, detector subarray, and readout pattern, which further expands the choices presented to an observer. While the choice of configuration may at least partly depend on wavelength coverage, there are overlapping wavelength ranges between the different NIR instrument configurations (see fig. 1 in Sarkar et al. 2021). One of the key questions in this early stage of *JWST* operations is the relative performance of the different instruments and their configurations particularly in regions of overlapping wavelength and also the optimal data reduction methods to process each configuration.

Of the available NIR instruments, NIRSpec presents the largest number of possible configurations with seven different dispersive elements available. Two of these elements, the Prism and the G395H grating, present two extremes. The Prism mode gives the broadest wavelength coverage in a single pass (0.6–5.3 μm) with high pixel count rates, but has the lowest spectral resolving power (R) of ∼100, with the spectrum concentrated within a 512-pixel-wide detector subarray on the NRS1 detector (the SUB512 subarray, which is 32 × 512 pixels in size), which causes it to saturate more easily than other modes. In contrast, G395H has a high R of ∼2700 but covers about half the wavelength range of Prism (2.87–5.14 μm), with the spectrum dispersed widely over the two NRS (2048 × 2048 pixel) detectors (using the SUB2048 subarray, which is 32 × 2048 pixels in size over each of the two detectors).

In this paper, we aim to compare these two key NIRSpec modes and uncover how similar the final results are between the two. We do this by performing a homogenized analysis of two observations taken of the hot Saturn-mass planet, WASP-39 b, obtained as part of the *JWST* Transiting Exoplanet Community ERS programme 1366 (PI: N. Batalha), one with NIRSpec Prism and the other with G395H. To facilitate this work, we developed an end-to-end data processing framework, JexoPipe, which takes the raw uncalibrated files and produces final transmission spectrum. We minimize differences in processing between the two configurations, allowing us to better control and assess the causes of any differences between the final spectra in their overlapped regions. We further examine JexoPipe against previously developed pipelines by comparing the final spectra obtained. The ERS NIRSpec Prism observation of WASP-39 b was originally presented in JWST Transiting Exoplanet Community Early Release Science Team (2023) and Rustamkulov et al. (2023) and the G395H observation in Alderson et al. (2023).

In the following, we first describe WASP-39 b and summarize previous results in Section 2 and then review NIRSpec and its performance in Section 3. We describe JexoPipe in Section 4. In Section 5, we present the final spectra obtained and compare these to those from other pipelines. In Section 6, we discuss any differences in the spectra obtained from the different NIRSpec modes. Section 7 describes application of an atmospheric forward model to the data. We summarize our findings and conclusions in Section 8.

2 WASP-39 B

WASP-39 b is a highly inflated Saturn-mass planet discovered in 2011 (Faedi et al. 2011) transiting a G8 star {[Fe/H] = −0.12} located at a distance of 213.982 pc.[1] It has an equilibrium temperature of 1166 K with an orbital period of 4.055 2941 d (Mancini et al. 2018). With a radius of 1.279 R_J and mass of 0.281 M_J, its bulk density is roughly four times less than that of Saturn at 0.167 g cm^{-3} (Mancini et al. 2018).

Prior to *JWST*, NIR transmission spectra had been obtained with Hubble WFC3 (Wakeford et al. 2018), Space Telescope Imaging Spectrograph (STIS), and *Spitzer* Infrared Array Camera (IRAC) (Fischer et al. 2016; Sing et al. 2016).Ground-based facilities have also contributed with an ultraviolet (UV)–optical transmission spectrum from the Very Large Telescope FOcal Reducer/low dispersion Spectrograph 2 (VLT-FORS2) (Nikolov et al. 2016) and UV–optical multiband photometry (Ricci et al. 2015). These spectra indicated a cloud-free atmosphere with the detection of sodium and potassium (Fischer et al. 2016; Nikolov et al. 2016; Sing et al. 2016). Water vapour was also detected with an estimated metallicity of 151^{+48}_{-46} × solar (Wakeford et al. 2018).

A suite of *JWST* ERS NIR observations has been performed on WASP-39 b. Using the NIRISS Single Object Slitless Spectroscopy (SOSS) mode (0.6–2.8 μm), Feinstein et al. (2023) reported a supersolar metallicity with results varying 10–30 × solar dependent on the atmospheric models applied, a subsolar C/O ratio, and supersolar K/O ratio. Using the G395H grating, Alderson et al. (2023) reported a somewhat lower metallicity of 3–10 × solar, with a subsolar to solar C/O ratio. They also report detection of CO_2, CO, H_2O, and SO_2. Rustamkulov et al. (2023) and JWST Transiting Exoplanet Community Early Release Science Team (2023) reported results using the Prism configuration (0.6–5 microns), with detection of CO_2, CO, H_2O, SO_2, and Na. The best-fitting model corresponded supersolar metallicity and supersolar C/O ratio with moderate cloud opacity. The SO_2 feature has also been discussed in Tsai et al. (2023) who reported that the feature could be explained by the photochemical breakdown of H_2S. Using NIRCam *F*322*W*2 (2–4 μm), Ahrer et al. (2023) reported detection of water vapour and that best-fitting chemical equilibrium models favoured a metallicity of 1–100 × solar with a substellar C/O ratio.

3 NIRSPEC

NIRSpec is a complex instrument with multiple configurations that cover (in total) the wavelength range from 0.6 to 5.3 μm. Due to the wide wavelength coverage in the NIR (covering the spectral features of key atmospheric molecules at high signal-to-noise ratio (SNR) as well as the Rayleigh scattering slope) and the many choices of configuration (giving flexibility for different targets and brightnesses), NIRSpec will be one of the principal instruments used in *JWST* transmission spectroscopy. It incorporates an integral field unit and a microshutter array for multi-object spectroscopy and five slits (apertures) for individual spectroscopy. The observing template for exoplanet time series is the bright object time-series mode. This utilizes the

[1] https://exoplanetarchive.ipac.caltech.edu/overview/WASP-39

Downloaded from https://academic.oup.com/mnras/article/531/2/2731/7670622 by guest on 03 July 2024

large aperture S1600A (1600 × 1600 mas^2) minimizing slit losses, in combination with up to nine different disperser/filter combinations. There are three high-resolution ($R \sim 2700$), each with an associated filter, and three medium-resolution ($R \sim 1000$) gratings, where G140H and G140M have two possible filters, and the others have one filter. This gives eight possible grating/filter configurations. In addition, the Prism mode is combined with the CLEAR filter to give the ninth configuration. Two Teledyne H2RG 2048 × 2048 HgCdTe detector arrays are used, called NRS1 and NRS2. Spectra from the medium-resolution gratings and Prism project only on to NRS1.

A choice of detector subarrays exists with different frame times. The detectors are read up-the-ramp producing non-destructive reads (NDRs) separated by the frame time. Up-the-ramp sampling allows for cosmic ray (CR) detection and possible recovery of affected slopes (Giardino et al. 2019). Most time-series observations are expected to use the NRSRAPID read pattern, where there is no onboard frame averaging, giving one NDR or frame per 'group', and where one group time per integration ramp is lost to reset. The observation proceeds as a sequence of integrations constituting an exposure. Thermal settling at the beginning of the exposure can occur (Birkmann et al. 2022a). If the SUB2048 subarray is used (with the high-resolution gratings), both detectors are implemented and a small gap will appear in any spectrum due to the physical separation between the two detectors. While the detectors are identical in manufacture, the commissioning study by Espinoza et al. (2023) using G395H and HAT-P-14 b as the target found differences in the slope of the systematic trend (being stronger in NRS1) and how closely the measured scatter in light curves matched the calculated noise (being a closer match in NRS1). All NIRSpec spectra have a slight curvature and tilt on the detector.

One of the main sources of systematics in NIRSpec is '1/*f* noise', which correlates counts during readout mainly in the fast readout direction (and to a lesser extent in the slow direction) and manifests as vertical banding seen in raw images. Espinoza et al. (2023) recommended producing light curves at the sampling resolution of the instrument (rather than binning across pixel columns) to minimize the degradation in SNR from correlated 1/*f* noise across columns, and then binning the results at a post-processing stage. However, Holmberg & Madhusudhan (2023) tested this strategy for NIRISS and found that the result is the same, regardless of the order of the binning, if one takes into account the covariance.

While pointing jitter and drift combined with intra-pixel variations were considered another possible correlated noise source, commissioning studies found that the line-of-sight pointing stability is very good ∼1 mas radial (Lallo & Hartig 2022). This minimizes the need to apply de-jittering algorithms to the data.

While persistence is expected in H2RG detectors producing a similar form of temporal effect as seen in the Hubble WFC3 detectors, the effect is expected to be smaller in the newer *JWST* detectors (Birkmann et al. 2022a). Commissioning studies indicate that the effect of persistence may not be a major concern for science programmes (Böker et al. 2023).

Bad pixels can exist of various types, e.g. 'hot' pixels with high dark signal, pixels with poor response to light, or dead pixels. From commissioning, Böker et al. (2023) found an operability rate of pixels of 99.59 per cent on NRS1 and 99.80 per cent on NRS2. These give 16 948 and 8275 non-operable pixels on NRS1 and NRS2, respectively. CRs are another source of abnormal pixel counts during an observation. Böker et al. (2023) report an average CR hit rate of about 5.5 cm^{-2} s^{-1} with a typical hit area of about 10.5 pixels. In addition, the so-called snowballs (Birkmann et al. 2022b) have been identified: CR events with a heavily saturated core of 2–5 pixels in radius surrounded by a halo. The large number of bad pixels combined with CR hits require pipelines to have a robust management method for dealing with bad pixel counts.

Time-dependent systematics and time-correlated noise are important to characterize for time-series exoplanet observations. Espinoza et al. (2023) performed an Allan deviation analysis of the band-integrated light-curve fit residuals using NIRSpec G395H. These were consistent with uncorrelated noise. The Alderson et al. (2023) analysis of the G395H data for WASP-39 b found minimal systematics, but there was a mirror segment tilt event that caused a change in point spread function and a jump in flux. For the Prism, Rustamkulov et al. (2023) found a high-gain antenna (HGA) movement event that affected a few integrations during the WASP-39 b Prism observation.

4 JEXOPIPE

JexoPipe is a recently developed data reduction framework that incorporates selected *JWST* Science Calibration Pipeline steps, which it combines with its own customized steps, procedures, and pathways. JexoPipe remains under development and will evolve to be applied for different instrument configurations. In this paper, we describe its use for the Prism and G395H configurations.

We applied JexoPipe to the Prism and G395H observations of WASP-39 b from the ERS programme 1366 (PI: N. Batalha). Details of these observations are given in JWST Transiting Exoplanet Community Early Release Science Team (2023), Alderson et al. (2023), and Rustamkulov et al. (2023). To summarize, the Prism observation of WASP-39 b began at 15:30:59 on 2022 July 10 and ended at 23:37:13 UTC. The NRSRAPID readout mode was used with group time of 0.226 16 s and five groups per integration ramp, with a total of 21 500 integrations. The G395H observation began at 22:04:06 on 2022 July 30 and ended at 06:20:26 on 2022 July 31 UTC. NRSRAPID was used with a group time of 0.902 s and 70 groups per integration, with a total of 465 integrations. The duty cycle efficiencies were 82 and 98.6 per cent, respectively, for Prism and G395H.

The pipeline is summarized in Fig. 1, with four stages of processing. To allow a homogenized approach to facilitate comparison of spectra from the two configurations, we keep pathways for the two configurations as similar as possible. The main differences arise in Stage 1 due to management of saturated pixels in the Prism data. The G395H pathway also applies a *Reference Pixel Correction* step and *Jump Detection* step in Stage 1, which are not applied in the Prism pathway. Reference files will also differ between the two configurations (and between the two G395H NRS detectors), such as the superbias file. The final transmission spectra were produced using steps from version 1.11.3 of the *JWST* Science Calibration Pipeline.

4.1 Stage 1

Stage 1 begins with .uncal Flexible Image Transport System (FITS) files obtained from the Mikulski Archive for Space Telescopes (MAST) archive.[2] These contain the uncalibrated NDRs (i.e. group-level images) per integration, and are provided as contiguous segments: four for the Prism and three for the G395H grating data. The G395H data are additionally divided between the two detectors: NRS1 and NRS2. The end products of Stage 1 are .rateints FITS files containing the count rate (in DN s^{-1}) per pixel per integration.

[2] https://mast.stsci.edu/portal/Mashup/Clients/Mast/Portal.html

Downloaded from https://academic.oup.com/mnras/article/531/2/2731/7670622 by guest on 03 July 2024

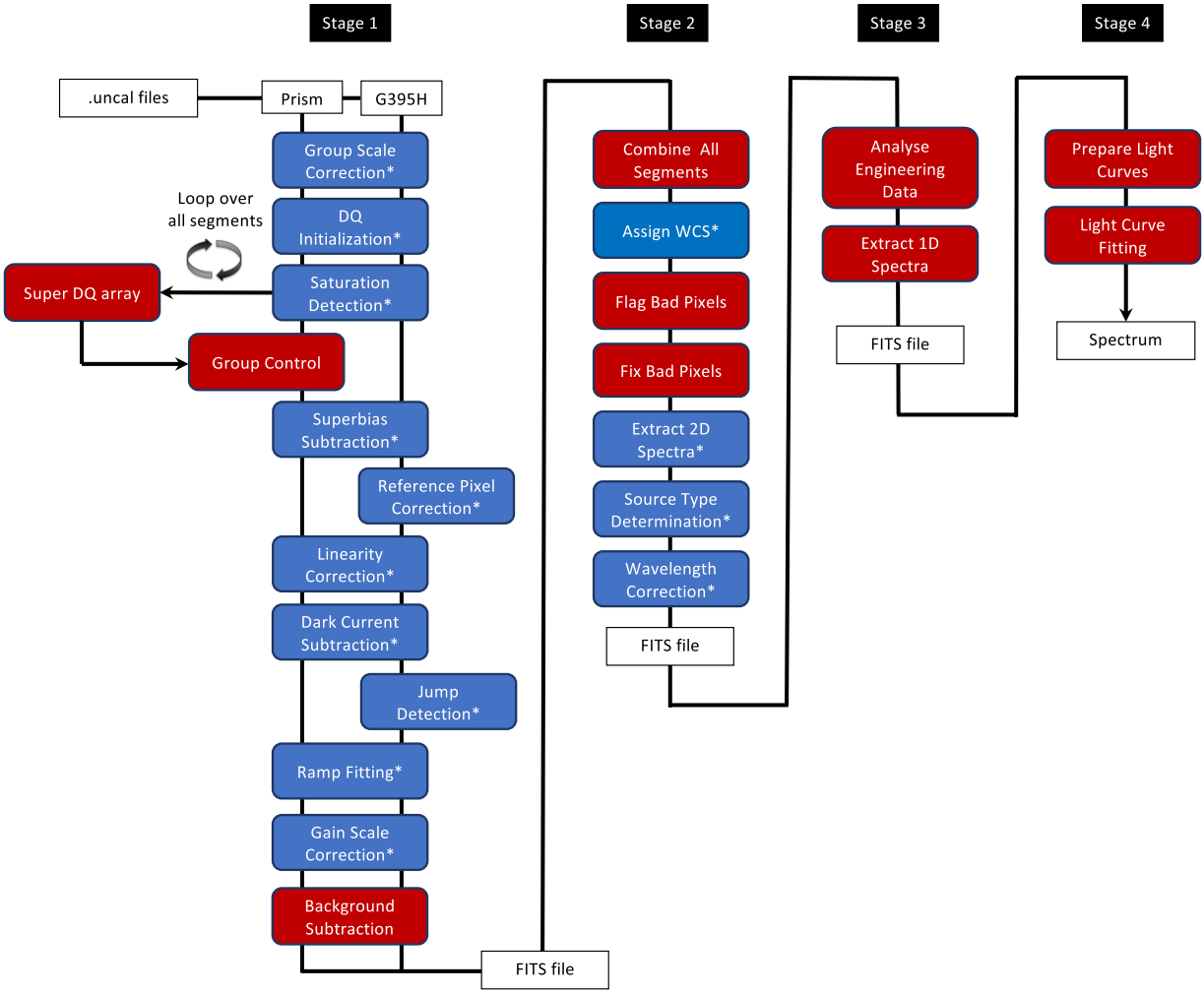

Figure 1. JexoPipe: the data reduction framework used in this work. Boxes with asterisks are steps utilized from the *JWST* Science Calibration Pipeline. The remaining boxes are customized steps. In Stage 1, the Prism and G395H pathways differ in the handling of saturated pixels, the application of reference pixels, and jump detection. Note that for G395H NRS2, the dark step was applied but automatically 'skipped' because the dark reference frame does not have enough groups. We expect this framework to evolve as our understanding of *JWST* performance improves and to be tailored for different instrument configurations.

For both Prism and G395H, we process each segment through the steps shown in Fig. 1. The functions of the official *JWST* pipeline steps (shown with asterisks) are described in the *JWST* pipeline package documentation.[3] We use the default settings for these steps except where described below.

For the Prism, we first run the pipeline for all segments up to completion of the *Saturation Detection* step. At that point, the group data quality (DQ) arrays from all the segments are combined into a 'super DQ array'. This super DQ array is subsequently used in the *Group Control* step.

The pipeline then proceeds to the *Superbias Subtraction* step. We used the default superbias files in all cases. Given previous concern with the G395H NRS1 superbias file as cited in Alderson et al. (2023) for this mode, we tried using a custom superbias created from the median of all first group images. However, we found no significant difference in the final spectrum offset or noise for G395H compared to using the default file; for Prism, there was a very small increase in the average spectrum transit depth when using the custom file (Section 6).

The G395H pipeline applies the *Reference Pixel Correction* step as side reference pixels are available; however, the top and bottom reference pixels are not available due to the subarray cutting these off. The SUB512 subarray used in the Prism observation has no reference pixels available (either at the sides or at the top and bottom) to apply the *Reference Pixel Correction* step. The *Linearity Correction* and *Dark Current Subtraction* steps are then applied. We found no significant difference in the spectral baseline if the dark current step was omitted compared to that if it was included (Section 6).

In common with the pipelines in Rustamkulov et al. (2023), the *Jump Detection* step is omitted for Prism (but not for G395H), as it results in a large number of false positives. Instead, CRs in Prism are managed by detecting outliers in Stage 2 (in the *Flag Bad Pixels* step, which is also applied to G395H). For G395H, we trialled the default rejection threshold of 4σ and also a threshold of 15σ, as used in Alderson et al. (2023). There were statistically insignificant differences in the final spectrum (Section 6). For the final baseline case, we use the results from the 15σ threshold.

[3] https://jwst-pipeline.readthedocs.io/en/latest/index.html

Downloaded from https://academic.oup.com/mnras/article/531/2/2731/7670622 by guest on 03 July 2024

The *Background Subtraction* step is performed to mitigate 1/*f* noise and remove diffuse and sky backgrounds. This is applied at the group level in both the Prism and G395H modes. Column-by-column background subtraction at the group level has been proposed to mitigate the vertical banding that arises from 1/*f* noise particularly for the Prism where no reference pixels are available (Birkmann et al. 2022a). Background subtraction is performed as follows. For the Prism, a median image is first obtained from the median of all the final groups in each integration (per data segment). A spatial profile of the median image is obtained by taking the sum in the *y*-axis and is used to identify the pixel row with the maximum count. In each group image, we mask out the spectrum, leaving the peripheral five rows on the top and bottom of the image. We also mask out bad pixels identified in the group DQ array and the pixel DQ array. Then, we mask out outliers to minimize the impact of CRs: half of the 16th–84th percentile range of all unmasked pixel values is taken to be $\sigma_{\rm bg}$ and pixels $\pm 10\sigma_{\rm bg}$ beyond the median value of all unmasked pixels are then masked. The column mean of unmasked pixels is then subtracted from all pixels in that column. We use this method since the spectral trace for the Prism is nearly linear and parallel to the *x*-axis of the detector.

For G395H, a median image is obtained in the same way as for the Prism. While the Prism spectral trace is roughly linear and parallel to the *x*-axis of the detector, the trace for G395H is slightly curved. As a result, the application of the mask for G395H images required first that the trace be mapped on the detector. On the median image, we divide the subarray into 10-pixel-wide columnwise slices and obtain a median profile in the *y*-axis for each slice. The pixel row with the maximum count in this profile is identified for each slice, and this is used to assign an approximate initial maximum point for each individual pixel column. We then fit a fourth-order polynomial to these points. This polynomial refines the pixel row in each column closest to the maximum of the spectrum and also allows the trace to be extended to the edges of the image where the signal is low. In each group-level image, a mask is applied ± 10 pixels around this central pixel per column. Bad pixels and outliers are masked out in the same way as for the Prism, and the mean of the unmasked pixels in each column is then subtracted from all pixels in that column.

As noted in Radica et al. (2023), while photons from diffuse backgrounds are affected by detector non-linearity, 1/*f* noise is not, so ideally 1/*f* noise correction should occur before the *Linearity Correction* step and diffuse background subtraction after. For NIRISS SOSS, Radica et al. (2023) present a way to do this with multiple steps; however, since the main purpose of this study was to compare two data sets with the same pipeline we decided implementing a single stage of background subtraction that happens after the *Linearity Correction* step, to be sufficient, thus conflating 1/*f* noise and diffuse background subtraction. Radica et al. (2023) noted that if 1/*f* correction was performed after the non-linearity correction, it did not result in any biases in the spectrum but increased noise. Thus, this single-step approach may result in increased noise.

4.1.1 Saturated pixels

Prism data present a challenge in that a large number of pixels near the peak of the stellar spectrum are saturated. We identify a 'persistently saturated region' of pixels, which are pixels where at least 10 per cent of the integrations in the timeline have at least one saturated group. This defines a region around the peak of the spectrum between 0.69 and 1.91 μm (Fig. 2). The management of this region is challenging and final results from this saturated region need to be interpreted with added caution.

We take a different approach to managing saturation compared to the pipelines in Rustamkulov et al. (2023). All four pipelines used in that paper expand the saturation flags along entire columns if a pixel in that column was saturated. Instead, here we make use of the *Saturation Detection* step argument `n_pix_grow_sat`, which we describe further below.

The *Saturation Detection* step compares the count on each pixel (before linearity correction) to a pre-determined saturation level in a reference file. If during an integration a pixel exceeds its reference level in a particular group, then it is flagged as saturated in the DQ array (DQ flag = 2) in that group and all remaining groups after that, *but not in the preceding groups*. Then at the *Ramp Fitting* step, the saturated groups are ignored when fitting the ramp. If all but the first group is saturated, there is the option of using this first group alone to obtain a 'ramp' value. To do this, the *Ramp Fitting* step argument `suppress_one_group` must be changed to False from the default of True. Since the 'slope' value using just one group may be quite different compared to that from two or more groups, this can potentially lead to noise in pixel timelines if a pixel flips between a single group and two groups in different integrations in its timeline. To reduce the noise impact of such an effect, we implement a *Group Control* step for the Prism pipeline, which is described further below.

If a pixel becomes saturated, exceeding its full well capacity, charge can leak from the pixel diffusing into surrounding pixels. This horizontal charge overflow from saturated pixels is known as 'blooming' (Cohen, Olaes & Grogin 2020). These neighbouring pixels may not be truly saturated (as defined by exceeding their saturation thresholds); however, due to this leaked charge, their electron counts now become unreliable in relation to their incident photon counts. By default, the *Saturation Detection* step flags all neighbouring pixels within 1 pixel of a truly saturated pixel as 'saturated' in the same groups as the truly saturated pixel; i.e. 3×3 box of pixels centred on the affected pixel is flagged as saturated. The size of this box is controlled by the step argument `n_pix_grow_sat`, which has the default value of 1.

We ran our pipeline for the Prism data varying the value of `n_pix_grow_sat` between 1 and 4, and the final spectra obtained are shown in Fig. 3 up to 3 μm for three different conditions: (a) no group control and `suppress_one_group` set to True (excludes single group integrations), (b) no group control and `suppress_one_group` set to False (includes single group integrations), and (c) with group control and `suppress_one_group` set to False (includes single group integrations). We explain group control below. In both cases (a) and (b), there are wide point-to-point variations in the spectrum within the persistently saturated region for all values of `n_pix_grow_sat`, while outside this region, all values of `n_pix_grow_sat` result in similar final spectra. In case (c), implementing group control greatly reduces the point-to-point variations; however, we also notice that increasing `n_pix_grow_sat` to values >1 reduces the spectrum baseline, which leads to the emergence of distinct peaks at 1.4 and 1.9 μm consistent with water bands (as well as further reducing point-to-point variation). The fact that increasing `n_pix_grow_sat` to values >1 appears to bring out some spectral features would suggest that the effect of the saturated pixels influences the counts on neighbouring pixels greater than 1 pixel away. We realize, however, that this is a somewhat subjective assessment of the improvement in the spectrum, and thus we exercise caution in the interpretation of the saturated region.

Downloaded from https://academic.oup.com/mnras/article/531/2/2731/7670622 by guest on 03 July 2024

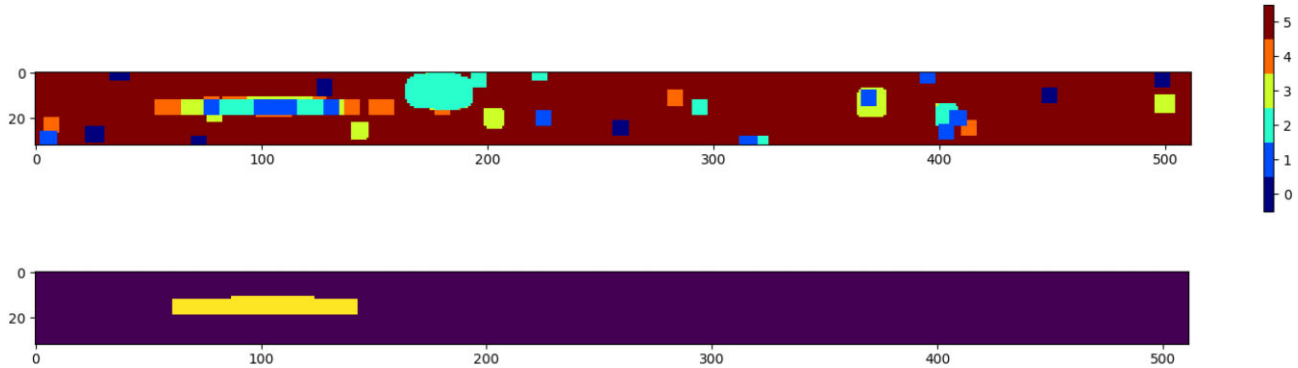

Figure 2. Determining a 'persistently saturated region' on the Prism subarray. Top: per pixel, the minimum number of unsaturated groups in any integration after completing the *Saturation Detection* step as implemented in this framework. Bottom: region shown where pixels where ≥10 per cent of the timeline have at least one saturated group. We term this the 'persistently saturated region'.

We decided to proceed in the final analysis using a value of `n_pix_grow_sat` = 3 for all saturated pixels in the Prism data. However, we use the default value of 1 for the G395H data. This is because count rates in the G395H data are much lower than in the Prism, and rare persistently saturated pixels are those isolated pixels with unusually low saturation thresholds rather than those with excessively high photon counts in large groups. Thus, we assume that the leakage into surrounding pixels is less of an issue. However, to be sure, we also ran the pipeline with `n_pix_grow_sat` = 3 for G395H and found no significant difference in the spectrum transit depths or noise (Section 6).

4.1.2 Group control step and 'flipping' noise

We find that some pixels in the Prism data exhibit a type of noise in their timelines that seems to result from flipping of the number of unsaturated groups per integration, and which can be controlled if the number of groups fitted for in the *Ramp Fitting* step is kept constant for the entire timeline (Fig. 4). This also appears to give the wide point-to-point variations seen in Fig. 3 in cases (a) and (b). For flips between 1 and 2 groups, this 'flipping noise' is also controlled by the default ramp fit step argument `suppress_one_group`, which is by default set to True, where 1 group integrations are given Not a Number (NaN) values, and which explain the slightly less noisy spectra seen for case (a) in Fig. 4 compared to case (b).

For this reason, and also to manage pixels where there are a significant number of single unsaturated groups, we apply a *Group Control* step for the Prism data. For pixels in the persistently saturated region, the super DQ array is used to find the minimum number of unsaturated groups in any integration (Fig. 2, top). If the minimum number of unsaturated groups is 2 or more, all integrations of that pixel are forced to use this number of groups. For pixels where the minimum number of unsaturated groups is 1, if the pixel exhibits ≥10 per cent of its timeline as single unsaturated groups, then the full timeline is fixed to a single group per integration. If a pixel has single groups for <10 per cent of its timeline, these groups are given NaN values (resulting in NaN values for the integration after the ramp fit step) (and flagged as 'bad' in the group DQ array) and the number of groups fixed at 2 for the remainder of the timeline.

The NaNs act as effective flags for correction in the Stage 2 *Fix Bad Pixels* step. Outside of the persistently saturated region, we do not apply group control as saturation is less frequent. If a pixel outside the persistently saturated region has a rare integration where the number of groups is different from the majority of its integrations, this may manifest as an outlier in the timeline, and thus managed in the same way as for other outliers in the Stage 2 *Flag Bad Pixels* step. There are a few pixels outside the persistently saturated region that have single unsaturated groups but these are for only 1–2 integrations (possibly from CR hits) and are flagged by applying NaN values at this stage to these groups.

When we apply group control, the point-to-point variations seen in the persistently saturated region are largely suppressed: Fig. 3 (case c). The *Group Control* step was not applied to the G395H data as saturation was much less frequent, and the longer ramp probably favours more stability against changes in the ramp gradient (and thus flipping noise) if the number of groups changes from integration to integration.

4.2 Stage 2

Stage 2 begins with the .rateints FITS files for each segment and first combines them into one file for the entire observation (Fig. 1). This facilitates the production of a rolling median image used in the *Fix Bad Pixels* step (see below). On completion of Stage 2, a .calints FITS file is produced containing calibrated, wavelength-assigned 2D slope images in flux units of DN s^{-1}.

4.2.1 Flag bad pixels

After combining all segments, the *Assign WCS* (World Coordinate System) step is applied. This step does not change the science data.We then apply a custom *Flag Bad Pixels* step. We flag all pixels that have abnormal DQ flags. We, however, do not include pixels flagged just for saturation or jumps as these effects will have been managed in the *Ramp Fitting* step, except in cases where additional flags, e.g. 'DO NOT USE', have been generated. NaN values are applied to these flagged pixels. Next, we check for outliers not picked up in DQ flagging in each integration image on a row-by-row basis. For each row in each integration image, we obtain a rolling median and rolling standard deviation (σ_1) ±5 pixels around a given pixel in the x-direction. We also find a line 'sigma' (σ_2) based on the half of the 16th–84th percentile range in the entire row. We flag as outliers (and give NaN values to) any pixel $\pm 3\sigma_1$ or $\pm 3\sigma_2$ from its corresponding rolling median value. This is iterated three times. For our baseline case, before this step, the proportion of NaN values (for

Downloaded from https://academic.oup.com/mnras/article/531/2/2731/7670622 by guest on 03 July 2024

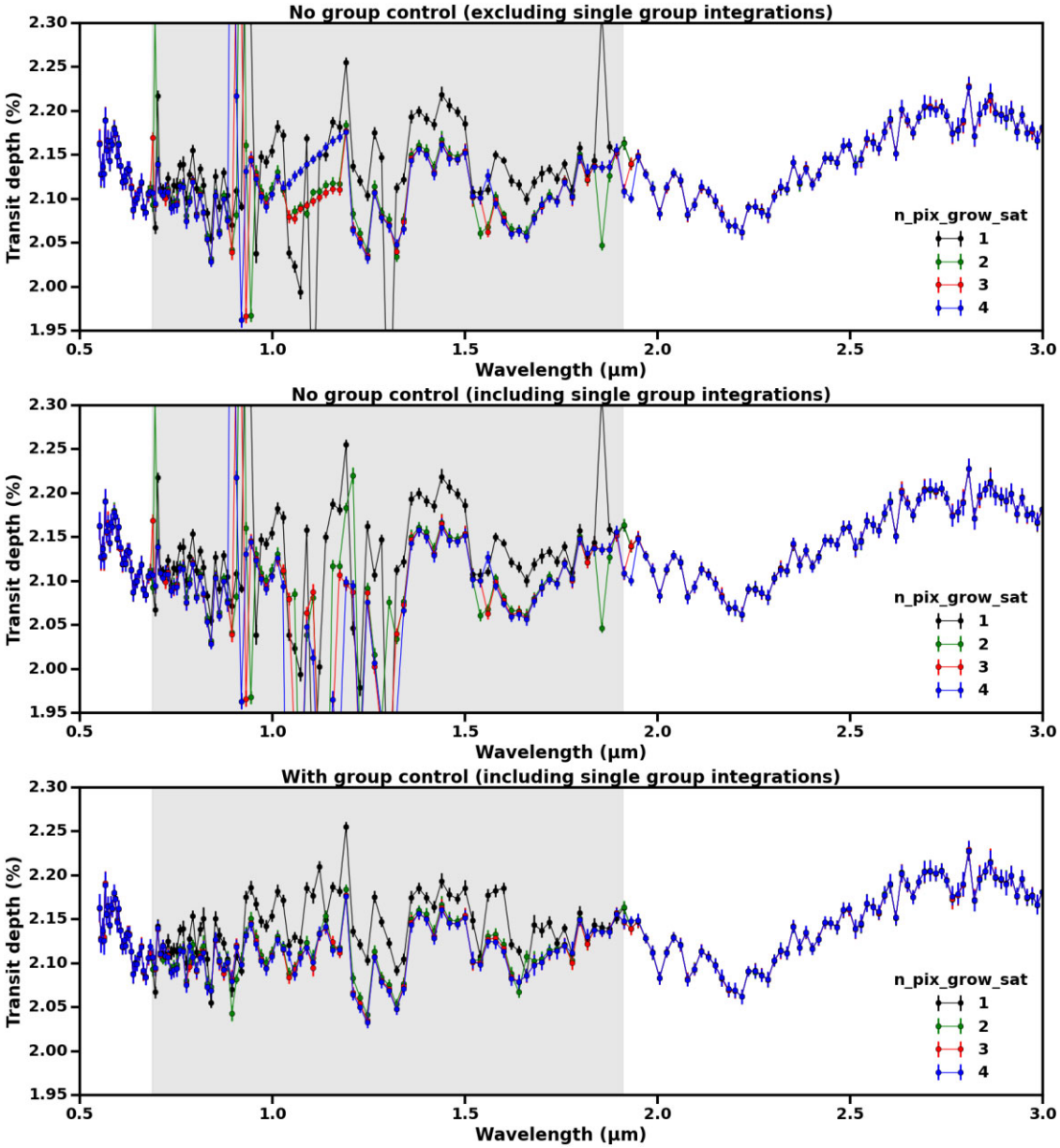

Figure 3. The effect of changing the *Saturation Detection* step argument, `n_pix_grow_sat`, on the transmission spectrum for Prism. The shaded area shows the persistently saturated region, which extends from 0.69 to 1.91 μm. Top: case (a) no group control and `suppress_one_group` set to True. Middle: case (b) no group control and `suppress_one_group` set to False. Bottom: case (c) with group control and `suppress_one_group` set to False. We settle on using an `n_pix_grow_sat` value of 3 for our final reduction.

all pixels over all integrations) is 0.6 per cent in Prism, 0.8 per cent in G395H NRS1, and 0.9 per cent in G395H NRS2. After the *Flag Bad Pixels*, this increases to 3.5 per cent in Prism, 3.6 per cent in G395H NRS1, and 1.6 per cent in G395H NRS2.

4.2.2 Fix bad pixels

We adopt the approach of filling in NaNs that have flagged bad pixels and outliers, rather than leaving these values open. This contrasts with the approach where such values are not filled and bad pixels are de-weighted. We fill in any NaN values in two ways. First, if a pixel timeline has <10 per cent NaN values, these are filled in by linear interpolation of good values. If a pixel timeline has ≥ 10 per cent NaN values, then the full timeline for that pixel is made NaN. Secondly, any remaining NaNs are filled in spatially by linear interpolation of good values on a row-by-row basis in each integration image. Finally, we deal with any remaining pixel-level light-curve outliers thus far not identified. For each pixel on each integration image, a rolling median value is produced from neighbouring images spanning ± 100 integrations for the Prism data set and ± 10 integrations for the G395H data. In addition for each pixel, we obtain a rolling standard deviation ($\sigma_{\rm roll}$) over the same range, and also the median of the rolling standard deviation values ($\sigma_{\rm median}$). We identify pixel values that are either $\pm 5\sigma_{\rm roll}$ or $\pm 5\sigma_{\rm median}$ beyond the corresponding rolling median value. These outlier pixel values are replaced by their values in the corresponding rolling median image.

Downloaded from https://academic.oup.com/mnras/article/531/2/2731/7670622 by guest on 03 July 2024

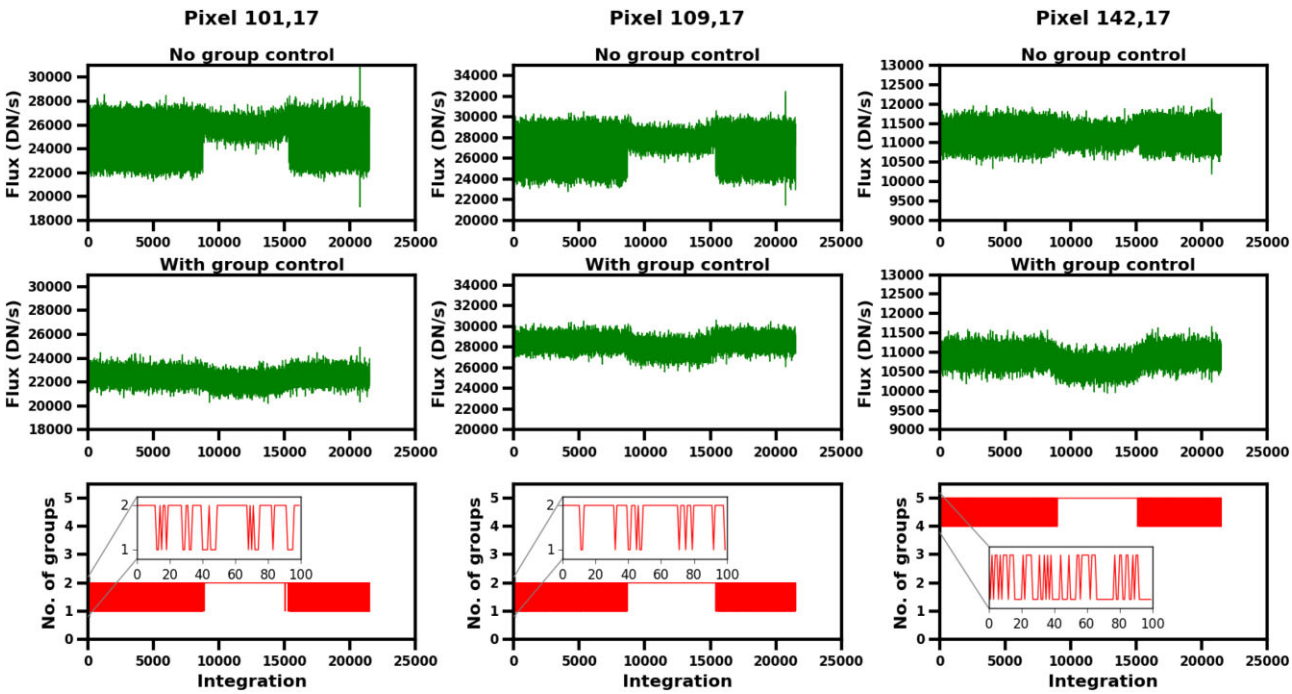

Figure 4. 'Flipping' noise as the rationale for the *Group Control* step in Prism processing. Three pixels are shown, all of which fall in the persistently saturated region. `suppress_one_group` was set to False. The top plots show the timelines of pixel counts without any group control. Middle plots show the timelines with the *Group Control* step implemented. The timelines are shown after the *Fix Bad Pixels* step in Stage 2. The bottom panel shows the number of unsaturated groups per integration (and thus the default ramp length). For pixel (101, 17), the number of groups flips between 1 and 2 during the OOT phase. As explained in the text, and as can be seen in the top row of figures, this leads to noise in the timeline. The number of single group integrations is $\geq$10 per cent (20 per cent) of the timeline for this pixel, so as per the rules adopted we force all the integrations to use only 1 group. This gives the result shown in the middle panel, where we see that the noise has been mitigated. Pixel (109, 17) also flips between 1 and 2 groups with the same kind of associated noise in the OOT phase. However, it has $<$10 per cent (8.4 per cent) of its timeline with single group integrations. For such cases, we flag the single group integrations as 'bad' by giving them NaN values in Stage 1, and these are later filled in during the Stage 2 *Fix Bad Pixels* correction step as described in the text. Pixel (142, 17) occurs at the edge of the persistently saturated region, and flips between 4 and 5 unsaturated groups in its OOT phase. This again results in noise that is controlled by fixing all integrations to 4 groups (middle panel). The noise resulting from flipping between different groups per integration may be due in part to the small number of total groups per integration operating near the saturation threshold, giving noticeably different slopes per integration.

4.2.3 Remaining steps

We then proceed to utilize steps from the *JWST* Science Calibration Pipeline that apply the 2D wavelength solution (Fig. 1) ending with the *Wavelength Correction* step. The flux units after this step remain as DN s^{-1}. The *Photometric Calibration* step in the *JWST* Science Calibration Pipeline is not strictly required since we are interested in relative flux changes.

In common with other pipelines, we have not applied a flat-field step as the reference files were not complete (Alderson et al. 2023; Lustig-Yaeger et al. 2023). The effect of not applying a flat-field may result in increased noise due to uncorrected pixel quantum efficiency and gain variations that manifests during movement (jitter) of the spectral image on the detector (Sarkar et al. 2021). The *JWST* pointing system has line-of-sight pointing stability of $\sim$1 mas (Lallo & Hartig 2022), which is 1/100th the pixel scale of NIRSpec (0.1 arcsec). The jitter noise impact of not applying the flat-field should thus be negligible.

4.3 Stage 3

Stage 3 begins with a manual examination of engineering data to identify relevant events in the timeline that may require addressing at the light-curve level. The .calints file from Stage 2 is then opened and extraction of the 1D stellar spectra performed. Stage 3 generates a FITS file containing 1D stellar spectra per integration.

4.3.1 Analysis of engineering data

We analyse the engineering data associated with each observation,[4] in particular the guide star *x* and *y* centroids, guide star flux, and HGA movement flags. As previously noted in Rustamkulov et al. (2023), in the Prism data there is HGA movement, and we find that this occurs between 59770.9722 and 59770.9723 MJD UTC. Therefore, in Stage 4, we exclude any light-curve points that fall into this time period. As explained below, we bin the Prism timeline to every 25 integrations, so this amounts to exclusion of two of the final binned light-curve points. In the G395H data, there are no HGA movements; however, we can see the impact of the previously noted (Alderson et al. 2023) mirror segment tilt event as a step change in the guide star flux (Fig. 5). This is reflected in the light curves of individual pixels, where the change in count can be positive or negative and varies

[4]https://jwst-docs.stsci.edu/methods-and-roadmaps/jwst-time-series-observations/jwst-time-series-observations-noise-sources

Downloaded from https://academic.oup.com/mnras/article/531/2/2731/7670622 by guest on 03 July 2024

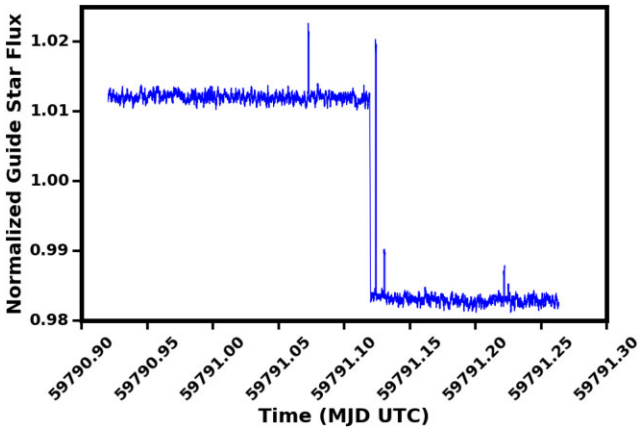

Figure 5. Guide star flux during the G395H observation. Flux has been normalized to the mean value, and smoothed by convolving with a 1000-step box-shaped kernel for clarity.

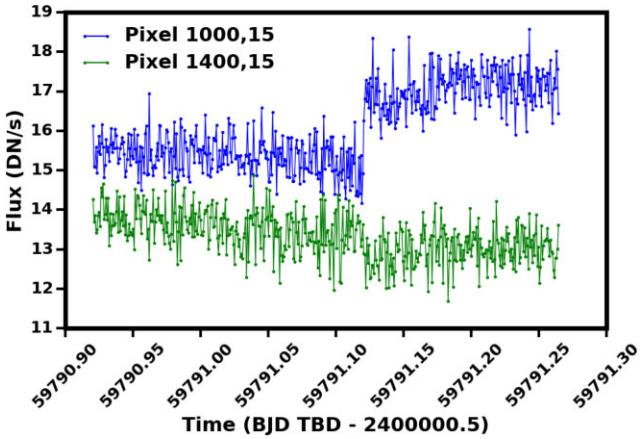

Figure 6. G395H mirror segment tilt event at pixel level. Two example pixel timelines from NRS2 are shown, with a step clearly visible and corresponding to that in the guide star flux timeline. In pixel (1000, 15) there is an increase in flux, and in pixel (1400, 15) there is a decrease in flux after the event. The amount and direction of flux change is thus pixel-dependent.

between pixels (Fig. 6). We did not detect an associated change in the guide star centroid positions. The tilt event is corrected in the Stage 4 *Prepare Light Curves* step.

4.3.2 1D spectral extraction

Next, a custom *Extract 1D Spectra* step is applied. Extraction is obtained through a 'box' extraction method and also an optimal extraction method.

For box extraction, using the assigned 2D wavelength solution per pixel, we obtain the mean wavelength per pixel column giving us a nominal 1D wavelength grid. We then assume wavelength bins centred on these values with boundaries being the mean of adjacent values. We then sum up the counts on all pixels that fall in a given wavelength bin. This method allows flexibility to obtain the 1D spectra from potentially tilted or curved spectral traces. We found the wavelength variations over a pixel column to be as follows: in Prism the standard deviation of the wavelength ranged from 0.002 to 0.013 μm per column, and for G395H this was ∼0.0003 μm per column in both NRS1 and NRS2. The resulting 1D spectra look virtually

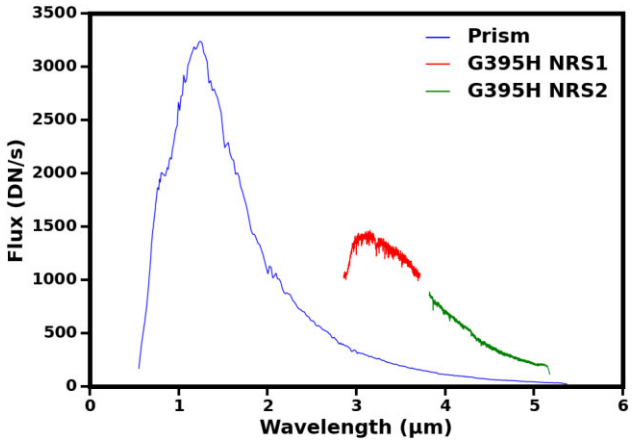

Figure 7. Example 1D stellar spectra from the first integration of each time series. These are the end result of Stage 3 of the pipeline. The flux counts on the Prism have been reduced by a factor of 100 to allow all spectra to be shown on same plot.

identical to those obtained through simple summation of the counts per pixel column. These 1D spectra, the wavelength grid, and timing information are saved together in a FITS file, which forms the output of Stage 3. Example 1D spectra from the first integration in each time series are shown in Fig. 7. We use the box extraction results for our baseline case.

For optimal extraction, we apply the principles in Horne (1986) to obtain the spectra and variance for each integration image. We use a rolling median image of 12 integrations for G395 and 100 integrations for Prism to provide the columnwise profile for each integration image. No background subtraction is included, and 5σ outliers were rejected iteratively. The final optimal extraction transmission spectra did not show any significant differences in offset, and had comparable or slightly less noise compared to the baseline cases (Section 6).

4.3.3 Error propagation

Regarding error propagation up to the end of Stage 3, the ERR (error) array in Stages 1 and 2 propagates the calculated photon noise and reads noise errors through the different steps. We do not adjust the error in the custom *Background Subtraction* step since the proportion of signal removed is small and will not impact the photon noise substantially. In the *Fix Bad Pixels* step, the variances of bad pixels are corrected in the same way as the pixel values themselves. In Stage 3, box extraction of the 1D spectra proceeds with quadrature addition of the corresponding ERR array values, whereas in optimal extraction, the errors are obtained directly from the optimal extraction algorithm.

4.4 Stage 4

In Stage 4, the 1D spectra from Stage 3 are extracted from the Stage 3 FITS file to produce a data cube of 1D spectra versus time for each data set. In the time axis, this constitutes the set of native (i.e. pixel-level) resolution spectral light curves. These spectral light curves are then fitted with light-curve models to extract the transit depth per wavelength, resulting in the final output consisting of the planet transmission spectrum.

In Stage 4, when calculating the errors on white light curve data points, these are obtained by taking the quadrature sum of the errors

Downloaded from https://academic.oup.com/mnras/article/531/2/2731/7670622 by guest on 03 July 2024

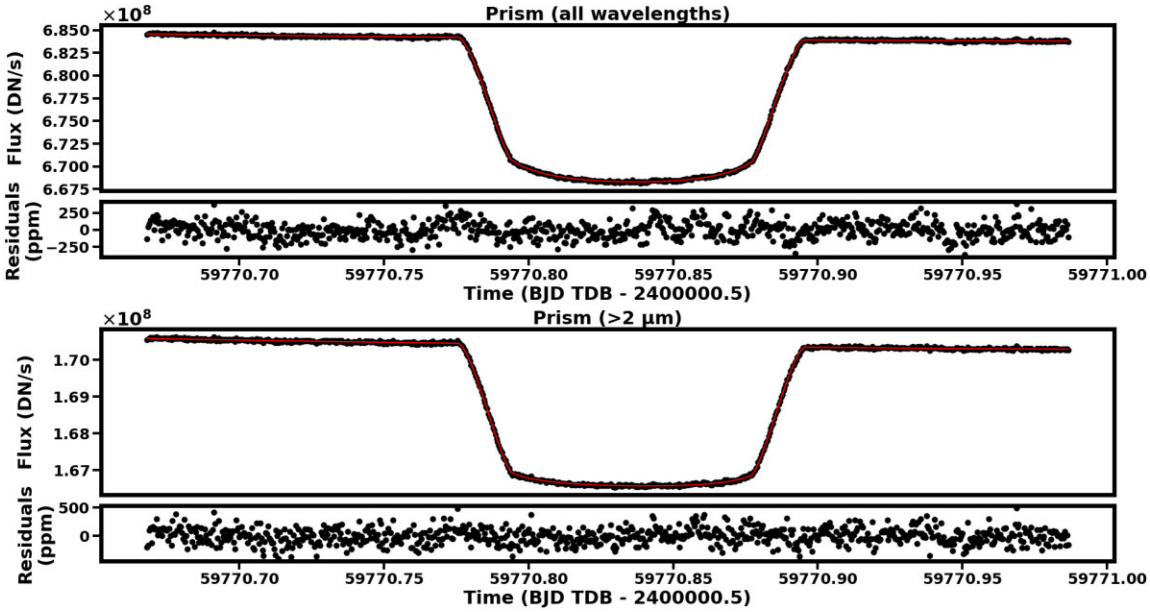

Figure 8. Prism white light curves and best-fitting solutions using median values from the MCMC posterior distribution, with residuals.

on the spectral data points that are summed to produce the white light curve point. Further error handling is discussed below.

4.4.1 *Prepare light curves*

Prior to light-curve fitting, we prepare the light curves in various ways, e.g. removal of any unwanted sections or data points, and/or binning in time or wavelength space as needed.

We perform one more stage of outlier removal by identifying white light curve outliers. A rolling median and rolling standard deviation (σ) of 200 data points in Prism and 20 data points in G395H are obtained, and outliers identified as being those $\pm 4\sigma$ from the median. No outliers were identified for G395H this way and five data points identified in Prism. The 1D spectrum corresponding to the outlying integrations was rejected and replaced by the average of the two integrations on either side of it.

For the Prism, the white light curve shows a clear non-linear trend in the out-of-transit (OOT) data. By excluding the first 1500 points, we found that we could reasonably fit this trend with a second-order polynomial, so this clipping was applied to all spectral light curves as well as the white light curve. Given the high cadence rate of the Prism data, we bin down each spectral light curve in time every 25 points, obtaining the mean mid-BJD timestamp of the 25 binned points. The time-step between the binned points is 34.438 s. Two points are then excluded around the HGA movement event as previously mentioned.

For G395H, we do not bin down the timelines, the time-step between points being 64.056 s. Although systematic trend appears more linear than in the Prism light curves, we fitted the trend with a second-order polynomial to keep the pipeline as consistent with the Prism pipeline as possible.

A short 'hook' appears at the start of the G395H white light curves, which we remove by excluding the first 10 integrations. To correct for the mirror segment tilt event in G395H, we first identify visually the event on the white light curves, and exclude three integrations around the tilt event. The effect of the event is then corrected during light-curve fitting as explained below.

4.4.2 *White light curve fitting*

The white light curve for the Prism was obtained by co-adding the Stage 4 spectral light curves. We trialled this twice: (1) including only wavelengths above 2 μm (and thus excluding the persistently saturated region), and (2) including all wavelengths. For light-curve fitting, we use a Mandel–Agol transit model applied using PYLIGHTCURVE[5] within a Monte Carlo Markov chain (MCMC) algorithm applied using EMCEE (Foreman-Mackey et al. 2013). In addition to the planet–star radius ratio (R_p/R_s), we fit for mid-transit time (t_0), the ratio of semimajor axis to star radius (a'/R_s), the inclination angle (i), and two quadratic limb-darkening coefficients [LDCs (c_1 and c_2)]. We also fit for the systematic trend and OOT baseline using a second-order polynomial [and thus three coefficients, a, b, and c, where the trend is represented by $a(1 + bt + ct^2)$, where t is time since the first integration included]. The error on the data is obtained by fitting the OOT data only with the systematic model (using a least-squares method), dividing this out and obtaining the standard deviation of the resulting points. The error bars on the individual light-curve data points are then scaled by a factor such that the average error bar matches the standard deviation of points.

This initial fit to the systematic trend also gives the signs of the coefficients c and b. These signs are held in variables and applied later in the final model fit, while the positive magnitudes of c and b are converted to natural logarithmic values for the MCMC algorithm. A log-likelihood function is used with uniform priors for all free parameters. We fit for the natural logarithms of all free parameters except t_0, c_1, and c_2. The period (P) is fixed to 4.055 2941 d (Mancini et al. 2018). Eccentricity is set to 0 and argument of periastron to 90°. We use 64 walkers with a burn-in of 1000 steps, followed by 4000 steps for the production run. Mean acceptance fractions of 44 and 43 per cent, respectively, for the all wavelength and the >2 μm cases are obtained consistent with good convergence. Fig. A1 shows the posterior distributions for these cases, and Fig. 8 shows

[5]https://github.com/ucl-exoplanets/pylightcurve

Downloaded from https://academic.oup.com/mnras/article/531/2/2731/7670622 by guest on 03 July 2024

Table 1. Summary of retrieved Prism white light curve parameters.

Parameter	Prism all wavelengths	Prism >2 μm only
R_p/R_s	$0.145\,22 \pm 0.000\,15$	$0.146\,04 \pm 0.000\,14$
a'/R_s	11.46 ± 0.02	11.43 ± 0.02
i (°)	87.83 ± 0.02	87.792 ± 0.025
t_0 (BJD TDB – 2400000.5)	59770.83563 ± 0.00001	59770.83566 ± 0.00001
c_1	0.17 ± 0.01	0.06 ± 0.02
c_2	0.28 ± 0.03	0.22 ± 0.03
a (DN s^{-1})	$6.845\,75 \pm 0.000\,09 \times 10^8$	$1.705\,84 \pm 0.000\,03 \times 10^8$
b (s^{-1})	-0.0057 ± 0.0003	-0.0081 ± 0.0003
c (s^{-2})	0.0067 ± 0.0008	0.009 ± 0.001

Table 2. Summary of retrieved G395H white light curve parameters.

Parameter	G395H NRS1	G395H NRS2	G395H NRS1 independent	G395H NRS2 independent
R_p/R_s	$0.145\,85 \pm 0.000\,07$	0.1465 ± 0.0001	0.1463 ± 0.0002	0.1466 ± 0.0002
a'/R_s	Fixed to Prism value	Fixed to Prism value	11.41 ± 0.03	11.39 ± 0.04
i (°)	Fixed to Prism value	Fixed to Prism value	87.74 ± 0.03	87.75 ± 0.04
t_0 (BJD TDB – 2400000.5)	59791.11203 ± 0.00002	59791.11214 ± 0.00003	59791.11203 ± 0.00002	59791.11214 ± 0.00003
c_1	0.04 ± 0.02	0.06 ± 0.02	0.09 ± 0.03	$0.07^{+0.04}_{-0.03}$
c_2	0.21 ± 0.02	0.13 ± 0.03	0.10 ± 0.05	0.12 ± 0.06
a (DN s^{-1})	$1.609\,34 \pm 0.000\,02 \times 10^6$	$8.7624 \pm 0.0002 \times 10^5$	$1.609\,34 \pm 0.000\,02 \times 10^6$	$8.7624 \pm 0.0002 \times 10^5$
b (s^{-1})	$0.000\,01^{+0.000\,23}_{-0.000\,01}$	$0.0002^{+0.0003}_{-0.0002}$	$0.000\,01^{+0.000\,20}_{-0.000\,01}$	0.0003 ± 0.0003
c (s^{-2})	$-0.0011^{+0.0007}_{-0.0008}$	$-0.000\,02^{+0.000\,01}_{-0.000\,20}$	$-0.0009^{+0.0007}_{-0.0008}$	$-0.000\,02^{+0.000\,02}_{-0.000\,24}$
Shift	$0.000\,97^{+0.000\,04}_{-0.000\,03}$	$0.000\,58^{+0.000\,06}_{-0.000\,04}$	$0.000\,99 \pm 0.000\,04$	$0.000\,59 \pm 0.000\,05$

the white light curves with best-fitting model and residuals. Table 1 summarizes the parameter estimation results from the Prism white light fits. We find that there are no significant differences between the system parameters obtained (a'/R_s and i) using all wavelengths versus just >2 μm; however, given the uncertainties in managing the saturated region, we choose to use the >2 μm values for the rest of the study. The residuals for the all wavelength case have a standard deviation of 120 ppm, which is ∼1.8 × the estimated noise based on propagating the ERR array values, and for the >2 μm case, the residuals are 140 ppm, 1.6 × the estimated noise.

For G395H, the white light curves for NRS1 and NRS2 were obtained by co-adding all spectral light curves. We performed two sets of fits: (1) where the system parameters a'/R_s and i are fixed to those from the Prism >2 μm result, and (2) 'independent' white light curve fits where a'/R_s and i were not fixed to the Prism values but obtained directly from the fits. In both cases, we fitted for t_0, c_1, c_2, and the natural logarithms of R_p/R_s and three polynomial coefficients, a, b, and c. To correct for the tilt event, a 'shift' parameter is added to the light-curve model that adds an offset to the post-event section of the light-curve model prior to multiplication by the systematic model.

Other aspects of the fit were as for the Prism above. We obtain mean acceptance rates of 31 and 32 per cent for NRS1 and NRS2 fixed-to-Prism cases, respectively, and 29 per cent for the independent cases. The results are summarized in Table 2. Fig. A2 shows the posterior distributions and Fig. 9 shows the light-curve fits. For the fixed-to-Prism cases, the residuals for NRS1 and NRS2 have standard deviations of 142 and 195 ppm, respectively (1.4 and 1.5 × the estimated noise). For the independent cases, these are almost the same, 141 and 195 ppm, respectively (1.4 and 1.5 × the estimated noise).

We find that the system parameters a'/R_s and i have somewhat lower median values in the 'independent' G395H fit than in the Prism fit; however, these are both highly correlated in the corner plots. To allow an unbiased comparison of Prism and G395H final spectra, we thus need to control for a'/R_s and i, which we do by adopting the Prism >2 μm values for a'/R_s and i for all final spectral light-curve fits. The median R_p/R_s results for the G395H results are somewhat greater than for the Prism all wavelength result, but closer to the Prism >2 μm result. This could be explainable by the high-amplitude spectral features in the G395H range that increase the average apparent radius over the NRS1 and NRS2 wavelength ranges.

4.4.3 Allan deviation analysis

We also performed an Allan deviation analysis of white light curves to examine for correlated noise. We look at the following cases: (1) Prism all wavelengths, (2) Prism >2 μm, (3) Prism 0.65–2 μm, (4) Prism <0.65 μm, (5) G395H independent NRS1, and (6) G395H independent NRS2. The Allan deviation plots are shown in Fig. 10. To produce these plots, the residuals are binned into progressively larger bin sizes. At each bin size, the fractional noise is calculated from the standard deviation of the binned counts divided by the binned signal. The latter is obtained from the coefficient a multiplied by the number of binned points per bin. For uncorrelated noise, the gradient of the log of the fractional noise versus the log of the bin size should be −0.5. Shallower gradients may indicate correlated noise in the timeline. In Fig. 10, the plots have been normalized to the first point, to allow easier comparison of the relative deviation from the uncorrelated expectation. The dotted line with a gradient of

Downloaded from https://academic.oup.com/mnras/article/531/2/2731/7670622 by guest on 03 July 2024

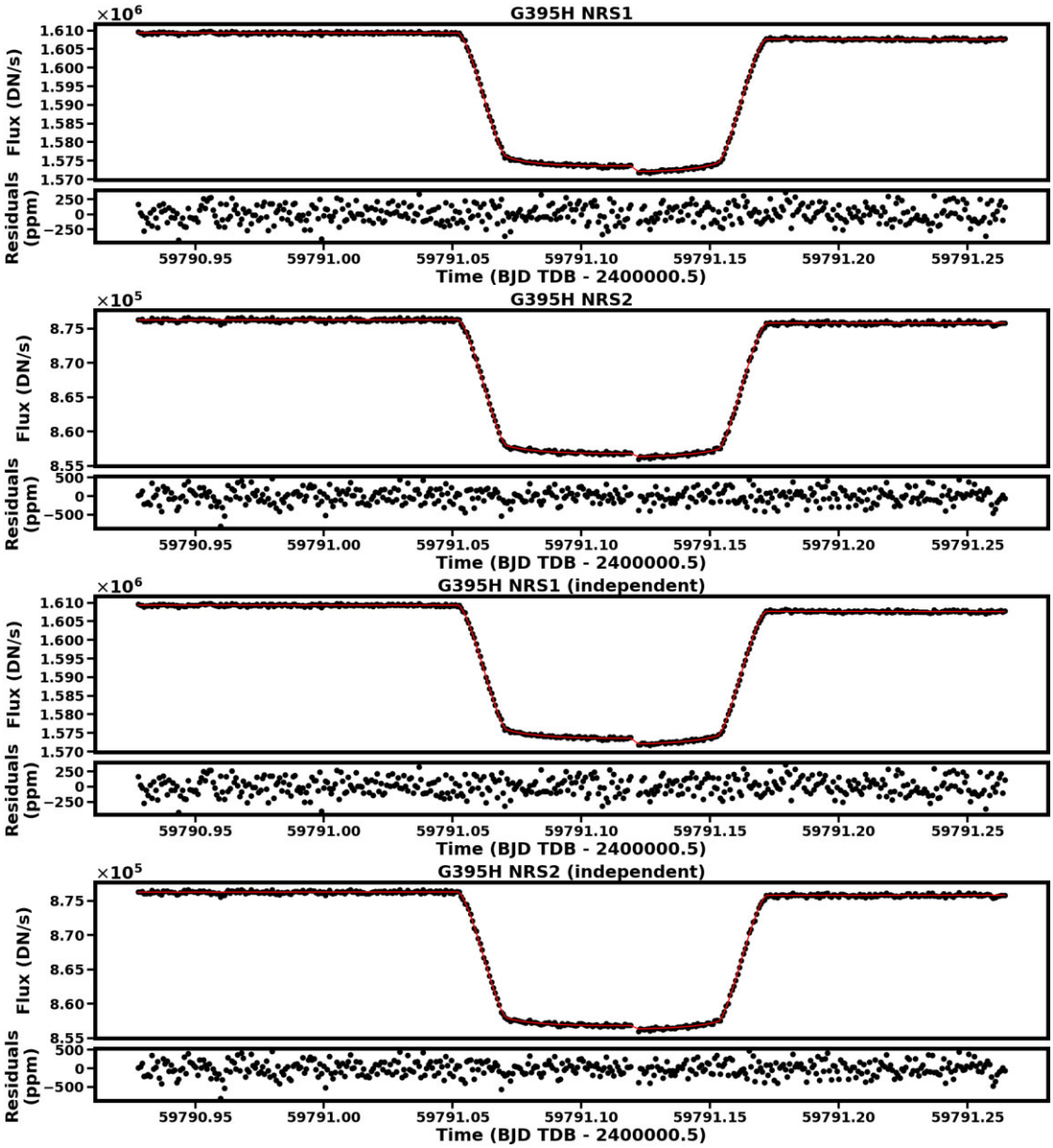

Figure 9. G395H white light curves and best-fitting solutions using median values from the MCMC posterior distribution, with residuals for G395H.

−0.5 is shown as an indicator of how uncorrelated noise would bin down.

We find that the G395H noise bins down fairly closely to what we would expect for uncorrelated noise, consistent with Espinoza et al. (2023). For the Prism, when we include all wavelengths in the white light curve, there is a markedly shallower gradient than expected for uncorrelated noise. For the fractionated white light curves, we see that the deviation is worst for 0.65–2 μm, which encompasses the persistently saturated region. The deviation is less for >2 μm and lesser still for <0.65 μm. These results suggest that the saturated region suffers the greatest correlated noise, which may be related to the saturation or our data reduction methodology for that region. However, some correlated noise extends beyond this region and this is therefore not attributable to saturation. It may be related to $1/f$ noise, our chosen systematic model or an additional unidentified cause but further investigation is needed to elucidate this. In Fig. 8, we see a slight upward 'bump' in the residuals around the start of ingress (more noticeable in the all wavelength case). We repeated the Prism white light curve fits with this region (containing 43 light-curve points) removed, and found the Allan deviations for the different cases to be unaffected, indicating that the correlated noise is not due to a poorer model fit in this region.

We use the Monte Carlo prayer bead method (Gillon et al. 2007; Manjavacas et al. 2018) to estimate the error inflation if we account for the correlated noise. For each case, we perform 50 trials. In each trial, we produce 1000 model realizations of the white light curve and then fit for all system (except period, eccentricity, and argument of periastron, which are set to the previously stated values), light-curve, and systematic parameters using LMFIT (Newville et al. 2016). Each realization is produced as follows. The residuals from the MCMC best-fitting model and the light curve are obtained. The residuals are then shifted in sequence by a random value picked from a uniform distribution ranging from zero up to the total number of light-curve points. The shifted residuals are added to the best-fitting

Downloaded from https://academic.oup.com/mnras/article/531/2/2731/7670622 by guest on 03 July 2024

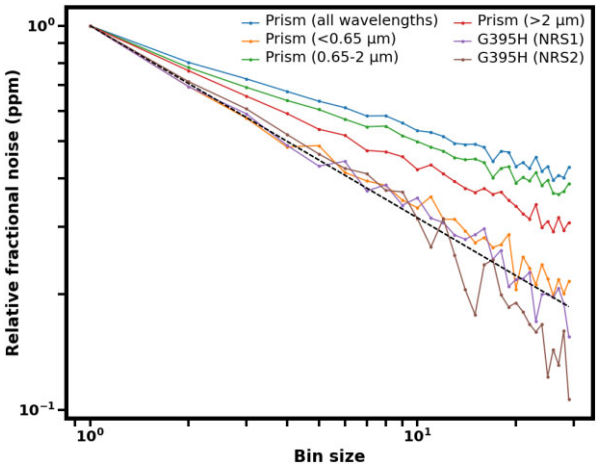

Figure 10. Allan deviation analysis of white light curve residuals. The fractional noise is normalized to the first data point for comparison purposes. The dotted black line has a gradient of −0.5 in log–log space and indicates how uncorrelated noise would appear to integrate down.

model to construct a 'new' light curve. This light curve is then fitted using LMFIT to extract the planet–star radius ratio. 1000 values of the planet–star radius ratio are thus obtained and we take the standard deviation of the distribution as an estimate of the 1σ uncertainty that incorporates correlated noise. We then take the ratio of this to the error given by LMFIT on the original light curve, to obtain an error inflation factor. We then obtain the mean and standard deviation of the inflation factor from the 50 trials. The inflation in errors on R_p/R_s as a result of this is as follows: Prism (all wavelengths) $\times$ 1.57 $\pm$ 0.03, Prism (>2 μm) $\times$ 1.36 $\pm$ 0.03, Prism (0.65–2 μm) $\times$ 1.45 $\pm$ 0.03, and Prism (<0.65 μm) $\times$ 1.29 $\pm$ 0.03. We confirm the relative lack of correlated noise in the G395H data by finding inflation factors of $\times$ 1.09 $\pm$ 0.02 in NRS1 and $\times$ 1.119 $\pm$ 0.025 in NRS2.

4.4.4 *Use of model four-factor limb-darkening coefficients*

To see whether any improvement in accuracy or precision occurs using model LDCs, we repeated the following cases using model four-factor LDCs obtained from the Exoplanet Characterization ToolKit (ExoCTK)[6]: Prism (all wavelengths), Prism (>2 μm), and G395H NRS1 and NRS2 (independent of Prism values). The resulting corner plots and light-curve fits are shown in Figs A3 and 11, with the parameters summarized in Table 3. For the Prism, there is close agreement with the baseline case in the >2 μm case. In the all wavelength case, the transit depth, a'/R_s, and i are distinct at 1σ but not at 2σ. In G395H NRS1, the transit depths again agree at 2σ, while in NRS2 they agree at 1σ. The residuals on the Prism show more systematic variation than in the baseline case indicating a poorer fit to the data: the standard deviations of the residuals are 133 and 141 ppm for the all wavelength and >2 μm cases, respectively. For G395H, the fits appear similar to the baseline cases with residual standard deviations of 145 and 196 ppm for NRS1 and NRS2, respectively. As a goodness-of-fit test to compare the baseline case with the four-factor LDC case, we employ the reduced chi-squared, keeping the noise constant between the two comparisons. We use the absolute noise from the baseline cases for this purpose (i.e. the standard deviation of the residuals). Table 4 summarizes the results. The reduced chi-square is larger in the four-factor model cases in all four cases, with the greatest increase being in the Prism all wavelength case. This would indicate that the empirically derived quadratic LDCs give a better fit to the data than the model four-factor LDCs.

4.4.5 *Spectral light-curve fitting*

We obtained model quadratic LDCs from the ExoCTK website, which were used for preliminary studies and to give initial values for fits; however, for the final analysis, we chose to obtain empirical LDCs for the full wavelength range covered by both Prism and G395H from light-curve fitting. This is partially motivated by the fact that for white light curves the empirical quadratic fit gave a better fit to data than the four-factor model LDCs. Empirical four-factor fits would be challenging due to the additional two free parameters required, and quadratic fits have been used in previous fits to WASP-39 data (Ahrer et al. 2023; Alderson et al. 2023; Rustamkulov et al. 2023). In addition, Rustamkulov et al. (2023) found that WASP-39 is 6 per cent brighter at the limb than predicted by models, indicating that an empirical approach is preferred.

The spectra from Prism and G395H were binned to $R = 66$ (slightly lower than the lowest native spectral power across the Prism subarray) to optimize the SNR. We then fitted each binned spectral light curve as described below and obtained the two LDCs, c_1 and c_2, for the binned wavelengths. Fig. 12 shows the extracted LDCs with error bars. Despite differences in instrument transmission, we find that the empirical LDCs for G395H match those from Prism within the 1σ errors. As a result, we proceed using just the Prism-derived LDCs for both modes.

A smoothing function was then applied to the Prism LDC vs wavelength plots to obtain the final LDCs[7] Fig. 13 shows the final empirically derived LDCs and those from the ExoCTK model. On running the final spectral light-curve fits for Prism and G395H, the LDCs were obtained for each spectral channel by interpolating the empirical LDC versus wavelength plots to the central wavelength of each light curve.

We fit the spectral light curves for both Prism and G395H at their native (pixel column level) resolutions. To allow comparison with the Prism, G395H spectral light curves were also binned to the Prism resolution. We exclude NRS2 wavelengths >5.1 μm, due to the rapid fall-off in the transmission above that wavelength resulting in low SNR (Fig. 7). Example posterior plots and light-curve fits for full resolution cases are shown in Fig. B1. In fitting the spectral light curves for each configuration (and NRS detector), we fix t_0 to the values in Tables 1 and 2 (i.e. to the Prism >2 μm case, and the G395H fixed-to-Prism cases). We fix the values for a'/R_s and i to those obtained from the Prism white light curve fit (>2 μm case) and fix the period to 4.055 2941 d (Mancini et al. 2018). Eccentricity is set to 0 and argument of periastron to 90°. For Prism and both G395H detectors, we fit for R_p/R_s and three polynomial coefficients for the systematic fit, a, b, and c. Additionally, for G395H, we fit for the light-curve 'shift' correcting for the mirror tilt event.

Fig. 14 shows how the systematic coefficients a, b, and c and the shift parameter in G395H vary with wavelength. The results have been binned to $R = 60$ for clarity. a follows the shape of the stellar spectrum, giving an OOT baseline flux. The variation of b and c is complex in Prism, and they appear negatively correlated to each

[6]https://exoctk.stsci.edu/limb_darkening

[7]This involved a convolution with a 10-point wide box function, with 5 points at each end filled in with values from a polynomial fit.

Downloaded from https://academic.oup.com/mnras/article/531/2/2731/7670622 by guest on 03 July 2024

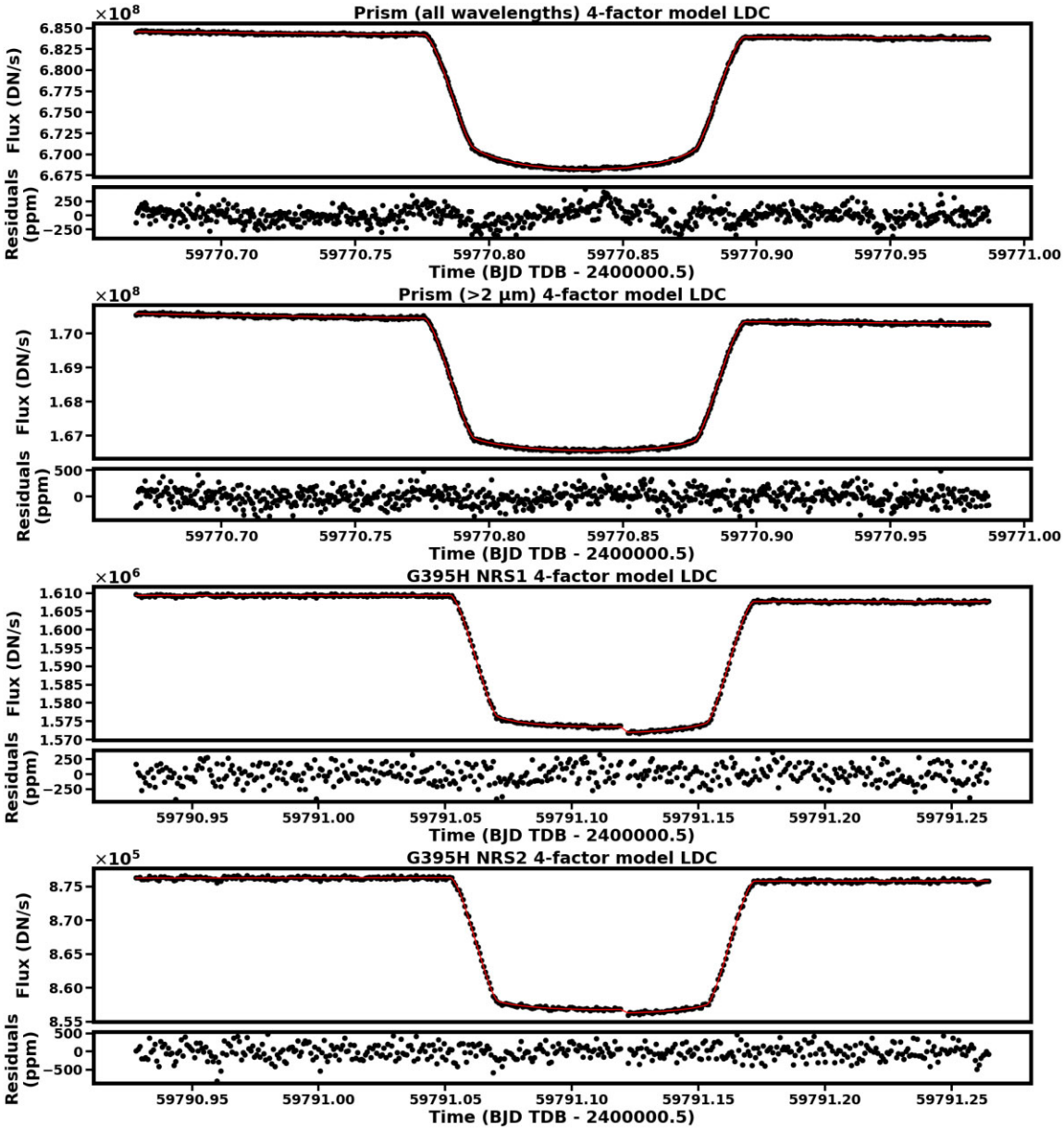

Figure 11. White light curves and best-fitting solutions when using model four-factor LDCs.

Table 3. Summary of retrieved white light curve parameters using model four-factor LDCs.

Parameter	Prism all wavelengths	Prism >2 μm only	G395H NRS1	G395H NRS2
R_p/R_s	$0.144\,87 \pm 0.000\,06$	$0.146\,09 \pm 0.000\,06$	$0.145\,84\,^{+0.000\,07}_{-0.000\,06}$	$0.146\,53 \pm 0.000\,09$
a'/R_s	11.51 ± 0.02	11.43 ± 0.02	11.448 ± 0.025	11.41 ± 0.03
i (°)	$87.90 + 0.02$	87.79 ± 0.02	87.810 ± 0.025	87.77 ± 0.03
t_0 (BJD TDB – 2400000.5)	59770.83563 ± 0.00001	59770.83566 ± 0.00001	59791.11202 ± 0.00002	59791.11213 ± 0.00003
a (DN s^{-1})	$6.845\,73 \pm 0.000\,09 \times 10^8$	$1.705\,84 \pm 0.000\,03 \times 10^8$	$1.609\,33 \pm 0.000\,02 \times 10^6$	$8.7623 \pm 0.0002 \times 10^5$
b (s^{-1})	-0.0057 ± 0.0003	-0.0081 ± 0.0003	$0.000\,008^{+0.000\,127}_{-0.000\,007}$	0.0005 ± 0.0003
c (s^{-2})	0.0066 ± 0.0008	0.009 ± 0.001	$-0.0002^{+0.0002}_{-0.0008}$	$-0.000\,02^{+0.000\,02}_{-0.000\,26}$
Shift	N/A	N/A	$0.001\,03^{+0.000\,02}_{-0.000\,03}$	$0.000\,62^{+0.000\,05}_{-0.000\,06}$

Downloaded from https://academic.oup.com/mnras/article/531/2/2731/7670622 by guest on 03 July 2024

Table 4. White light curve goodness-of-fit for four-factor model LDCs versus a quadratic LDC fit.

	Noise (DN s^{-1})	Baseline (quad. LDC fit) χ^2_ν	Model four-factor LDC χ^2_ν
Prism (all wavelengths)	82 194.90	1.011	1.234
Prism (>2 μm)	23 817.34	1.011	1.033
G395H NRS1	226.93	1.028	1.075
G395H NRS2	170.87	1.023	1.028

other in wavelength space, but not obviously correlated to spectral features. In G395H, b and c are within 1σ of zero, consistent with no time-dependent systematic trend. The shift parameter appears fairly constant with wavelength in NRS1, but displays an increasing trend with wavelength in NRS2.

5 TRANSMISSION SPECTRA

The final transmission spectra obtained are shown in Fig. 15. The uppermost plot shows the unbinned 'native resolution' spectra. In the second plot, the G395H spectra are rebinned to the resolution of the Prism spectrum, with the residuals between the two configurations and the average errors shown in the lower two plots.

5.1 Prism

Using the strategy of group control and increasing `n_pix_grow_sat` to 3, we were able to recover data from the persistently saturated region including the 1.4-μm H_2O feature. However, given there is no consensus on the optimal strategy to manage the saturated region, the results in this region should be considered tentative. We recover the previously observed water features at 1.8 and 2.8 μm, the large CO_2 feature at 4.3 μm, and the likely SO_2 feature at 4 μm. Fig. 16 compares our Prism spectrum with those publicly available spectra from the four pipelines used in the study by Rustamkulov et al. (2023). The errors on the residuals are the quadrature sum of the 1σ error bars from the comparison pipeline spectrum and those from the rebinned JexoPipe spectrum. There is some disagreement within the persistently saturated region, where JexoPipe gives a shallower transit depth compared to the other pipelines on the blueward side of the 1.4-μm water feature, but a deeper transit depth over the rest of the region. This may reflect the difference in approach taken in processing the saturated region between JexoPipe and the four other pipelines. Another consideration for the difference in the saturated region could be possible correlations between our empirically derived LDCs and transit depth, which could have resulted in a transit depth bias in the saturated region. Further investigation is needed to fully understand the differences.

Excluding the saturated region by comparing wavelengths above 2 μm only, if we consider the residuals (i.e. JexoPipe minus the alternate pipeline), the closest match is to Eureka where the average residual above 2 μm is 28 ± 20 ppm. When comparing the average errors on the transit depths (at the binned resolution of the comparison pipeline spectrum), the smallest difference is with Eureka (JexoPipe average error above 2 μm is 5 ppm higher), and the largest difference is with Tiberius (JexoPipe average error is 24 ppm lower).

Limb darkening: Given the structure of the residuals presented in Figs 8 and 11 (no evident symmetry around the mid-time), we believe that the differences between the baseline analysis and the quadratic or four-factor model LDC cases are due to the correlated noise. In practice, by fitting for empirical LDCs, the model is probably absorbing part of the correlated noise caused by the saturated pixels, resulting in better statistics. At a first glance, this can be overlooked but we are concerned that it can cause biases in the final spectrum. Such a bias can be seen in the saturated region of the Prism data. In Fig. A1, we can see that c_1 is correlated with depth, while c_2 and depth are anti-correlated. Around 1.4 μm, the empirical c_1 is significantly lower than the theoretical c_1 and the empirical c_2 is significantly higher than the theoretical c_2. By combining the two observations we can expect that if the LDC calculations are biased, then the spectrum will be negatively biased. There are two data points in the current analysis that are consistently lower compared to any other result close to 1.4 μm (Fig. 16). Therefore, we believe that at least these two points are biased due to the choice of empirical LDCs instead of model LDCs. We do not expect the authors to change their analyses based on this observation, but they could note it in the text as something to further investigate in the future.

5.2 G395H

The much higher native resolution of G395H compared to Prism is evident in Fig. 15 (upper plot). To make it easier to visualize the spectral features and to allow comparison to the Prism, the spectral light curves were rebinned to the Prism native resolution (Fig. 15, second plot). The SO_2 and CO_2 features are evident. The gap in the middle of the spectrum is due to the physical division between the NRS1 and NRS2 detectors.

In Figs 17 and 18, we compare the JexoPipe spectrum to 12 publicly available spectra from the 11 pipelines and weighted mean used in the study by Alderson et al. (2023). The errors on the residuals are the quadrature sum of the 1σ error bars from the comparison pipeline spectrum and those from the rebinned JexoPipe spectrum. We find good agreement with the amplitudes of spectral features obtained; however, there are differences in the spectrum baseline that vary depending on the pipeline being compared. There is a broad range of spectrum baselines across the 11 comparison pipelines. The closest matches in terms of average offset are with Espinoza-transitspectroscopy (−8 ± 16 ppm) and Evans (−4 ± 17 ppm). The JexoPipe result has a lower spectrum baseline than the weighted mean result with an average offset of −194 ± 17 ppm. Comparing average errors on the transit depths, JexoPipe has similar errors to the weighted mean result, the average error being 12 ppm lower in JexoPipe.

6 COMPARISON OF PRISM AND G395H SPECTRA

We now examine the differences in the baseline Prism and G395H spectra (Fig. 15). An offset is visible between the Prism and G395H results. The average offset (G395H–Prism) is −188 ± 20 ppm for NRS1 and −139 ± 31 for NRS2. We can see that when binned to the same resolution, the transit depth precision is higher for G395H than for Prism (Fig. 15, lowest plot). The average difference in noise compared to Prism is −29 ppm in G395H NRS1 and −50 ppm in G395H NRS2. The noise in Prism is on average ∼1.3× that in G395H at the Prism resolution.

The cause of the offset between the two modes is not immediately apparent. Since we homogenized this study by using the same LDCs and system parameters, a'/R_s and i, for both data sets, we can attribute any differences to instrumental or astrophysical variations between the two observations or differences in data processing.

Downloaded from https://academic.oup.com/mnras/article/531/2/2731/7670622 by guest on 03 July 2024

Table 5. Effects of changing data reduction and light-curve fitting steps on relative spectral transit depths. Baseline case is described in the text and corresponds to spectra shown in Fig. 15.∗ First three columns give the mean difference in ppm between spectral data points (binned to the Prism wavelength grid) between that case and the baseline case. The next two columns give the average offset between Prism and G395H spectra binned to the Prism resolution, for the given case, to enable a statistical comparison between the two spectra (Prism versus G395H). The average noise (when binned to Prism resolution) on the spectrum transit depth is given in the final three columns. Changes in Stage 1 reduction steps are shown above the line, and changes in Stages 3–4 extraction and light-curve fitting are shown below the line. The Prism results are for wavelengths >2 μm only. 'Background pre-linearity' applies the background subtraction stage just prior to the linearity correction and omits dark current subtraction.

	Change from baseline (ppm)			Comparison with Prism (ppm)		Average noise (ppm)		
	Prism	G395H NRS1	G395H NRS2	G395H NRS1	G395H NRS2	Prism	G395H NRS1	G395H NRS2
Baseline case	N/A	N/A	N/A	-185 ± 20	-139 ± 31	234	104	232
Background pre-linearity	-102 ± 21	1 ± 17	1 ± 28	-96 ± 20	-1 ± 31	237	100	217
No dark subtraction	-7 ± 22	-4 ± 18	2 ± 29	-191 ± 20	-103 ± 32	239	102	232
No reference pixel stage	N/A	-74 ± 18	-2 ± 29	-259 ± 21	-142 ± 32	234	107	233
n_pix_grow_sat = 3 (G395H)	N/A	-1 ± 18	-11 ± 29	-186 ± 20	-151 ± 32	234	104	233
CR rej. threshold = 4σ (G395H)	N/A	-1 ± 18	1 ± 29	-186 ± 20	-138 ± 31	234	104	230
Custom superbias	28 ± 21	6 ± 18	2 ± 29	-223 ± 20	-138 ± 32	236	105	231
Optimal extraction	6 ± 21	2 ± 18	5 ± 28	-189 ± 21	-140 ± 30	227	104	228
Quadratic model LDC	-12 ± 21	-7 ± 18	-4 ± 29	-183 ± 20	-140 ± 32	235	104	232
Four-factor model LDC	11 ± 21	19 ± 18	12 ± 29	-183 ± 20	-138 ± 32	234	105	232
Null systematic (G395H)	N/A	-6 ± 18	-5 ± 28	-191 ± 20	-144 ± 31	234	102	226

Note. ∗To permit comparison with the other cases, the baseline case was repeated here with the shorter MCMC runs used for the comparisons; hence, the G395H NRS1 offset is slightly different from that quoted in the text for the longer MCMC run used for Fig. 15.

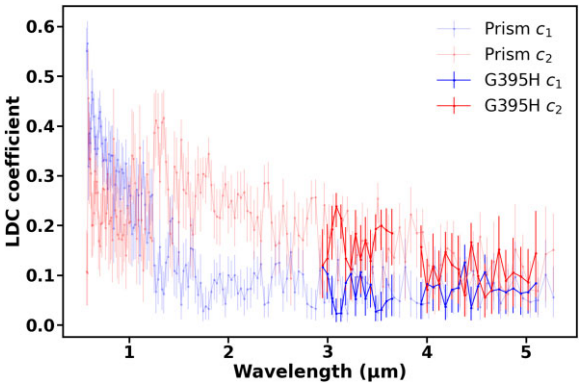

Figure 12. Empirically derived quadratic LDCs from $R = 66$ Prism data (faded points) and those from $R = 66$ G395H data (bolder points). The values between Prism and G395H match within the 1σ error bars.

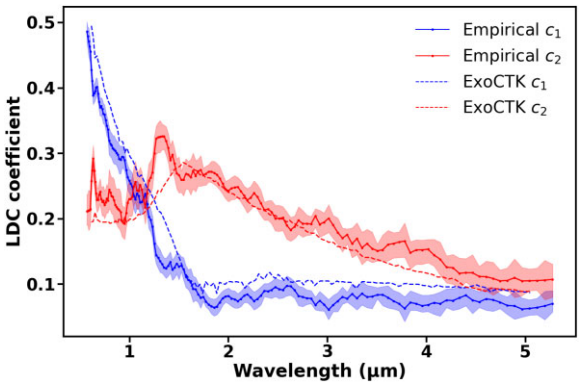

Figure 13. Empirically derived quadratic LDCs from Prism data and those from the ExoCTK website model. The latter uses the Kurucz ATLAS9 model grid ($T_{\rm eff} = 5000$ K, $\log g = 4.45$, and Fe/H = 0.01). Shaded regions give the 1σ error on the empirical values.

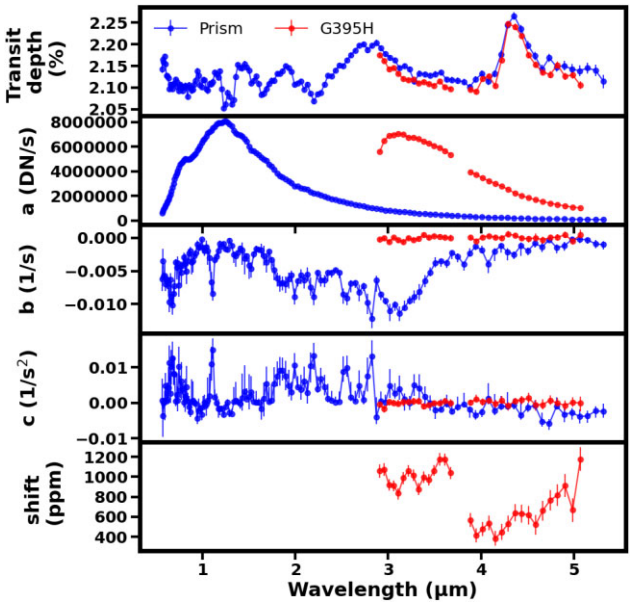

Figure 14. Systematic coefficients. Data have been binned to $R = 60$. Transmission spectra are shown for comparison in top panel. For G395H, values for a have been increased by a factor of 5000 for clarity. The light-curve shift applied during light-curve fitting for G395H to correct for the tilt event is shown in the bottom panel.

We kept the data processing differences to a minimum; however, there were differences in Stage 1, where the *Reference Pixel Correction* step was applied in the G395H pathway but not in the Prism pathway, and where the Prism pathway applied the *Group Control* step and a different value of `n_pix_grow_sat`. Group control, however, is only applied to the persistently saturated region of Prism, and so would not be a cause of the offsets seen with G395H, which are at wavelengths beyond this region.

We investigated the effects of various changes to pipeline processing and light-curve fitting from the baseline case. The results are summarized in Table 5. The light-curve fits for these comparisons

Downloaded from https://academic.oup.com/mnras/article/531/2/2731/7670622 by guest on 03 July 2024

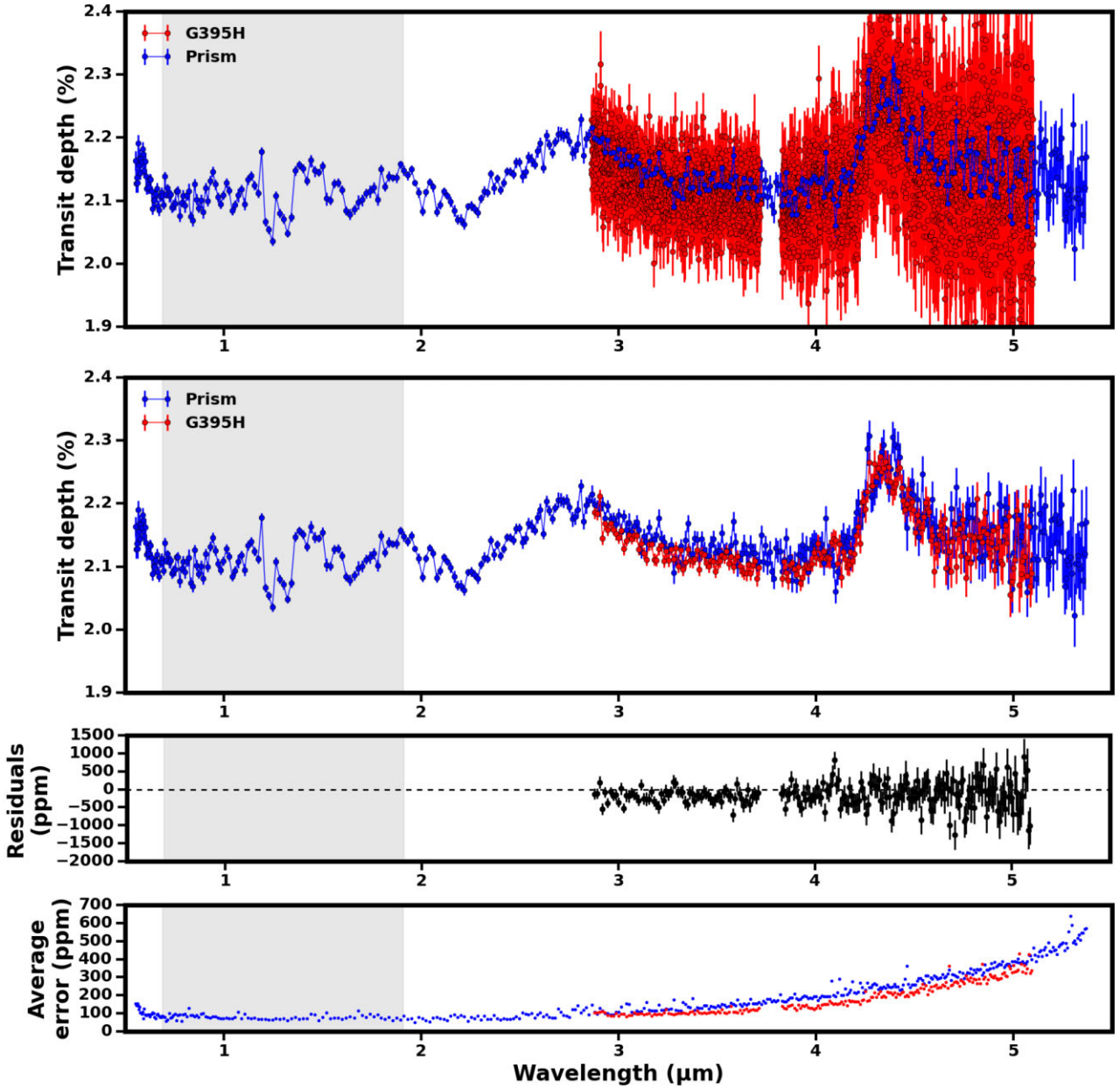

Figure 15. Transmission spectra of WASP-39 b obtained using JexoPipe. Uppermost plot shows the spectra obtained at the native resolution without any wavelength binning. The shaded area indicates the region of persistent saturation for the Prism. Central points are the median of the posterior distributions and the error bars span the 16th–84th percentile range. In the second plot down, the Prism spectrum is shown at native resolution, while the G395H spectra are binned to the Prism wavelength grid to allow comparison of the transit depths with wavelength. To permit a difference comparison of central points and quadrature sum of errors, the error bars show the average error (half the 16th–84th percentile range) and the central points are the average of the 16th and 84th percentiles in the posterior distribution. The third plot gives the difference between the binned G395H spectrum and the Prism spectrum central points (i.e. residuals = G395H minus Prism). The error bars on the residuals are the quadrature sum of the error bars as shown in the second plot. The fourth plot gives the average error on the transit depth at the native resolution of the Prism.

were performed as previously described but with modified MCMC parameters (400 burn-in / 2000 production steps, 32 walkers), and Prism spectra were obtained only at >2 μm, with G395H binned to Prism resolution to allow comparison. We re-ran the baseline case with the changed MCMC run parameters and obtained an offset of -185 ± 20 ppm between G395H NRS1 and Prism as compared to -188 ± 20 ppm mentioned above, while the offset between G395H NRS2 and Prism is unchanged.

The noise in all the cases is comparable to the baseline cases. In all cases, the G395H NRS2–Prism offset is greater than the G395H NRS1–Prism offset, indicating an offset between the two G395H NRS detectors, which is ∼40–50 ppm in most cases, including the baseline case. We note that this is consistent with inter-detector offsets reported in previous studies (Madhusudhan et al. 2023; Moran et al. 2023).

Downloaded from https://academic.oup.com/mnras/article/531/2/2731/7670622 by guest on 03 July 2024

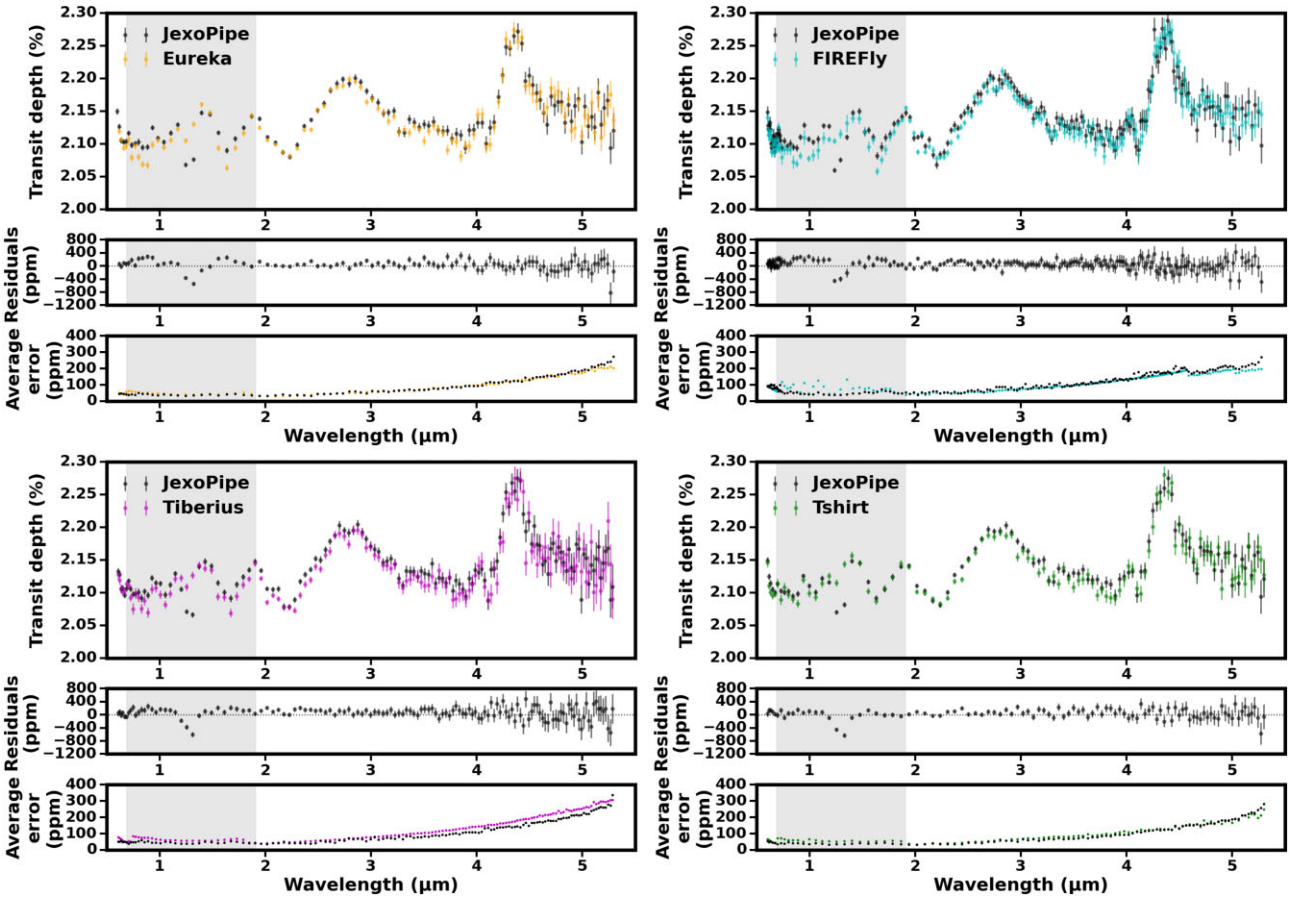

Figure 16. Comparison of our Prism result with other Prism pipelines presented in Rustamkulov et al. (2023). The JexoPipe result is rebinned to the wavelength grid of the comparison pipeline's publicly available spectrum. The shaded area indicates the region of persistent saturation. Residuals are JexoPipe minus the comparison pipeline spectrum. The average errors on the transit depths for the comparison spectrum and the JexoPipe spectrum when binned to the resolution of the comparison spectrum are shown in the lowest plots.

An interesting finding is that if background subtraction is performed just before the linearity correction step,[8] then the offset between Prism and G395H is greatly reduced or eliminated. Compared to the baseline, the G395H NRS1–Prism offset falls from -185 ± 20 to -96 ± 20 ppm and the G395H NRS2–Prism offset falls from -139 ± 31 to -1 ± 31 ppm (Table 5). This reduction in offset is due to a change in the Prism spectrum from the baseline case (changing on average by -102 ± 21 ppm). There is no significant change in the G395H spectra from baseline. The reason for this change in Prism (and not G395H) is not immediately clear.

Omitting the reference pixel stage impacts G395H NRS1 by causing a significant negative offset compared to the baseline case of -74 ± 18 ppm, and increasing the G395H NRS1–Prism offset to -259 ± 21 ppm. However, omitting the reference pixel stage does not seem to impact NRS2, where the change from the baseline case is not significant (-2 ± 29 ppm). The reason for this discrepant response between the two NRS detectors is not clear. Applying `n_pix_grow_sat` = 3 to G395H also does not significantly change the offset (see Table 5). The other variations in Table 5 do not show significant changes from the baseline case in terms of average offsets and noise. These include using a 'null' systematic for G395H where no time-dependent trends are included (i.e. we fit only for systematic coefficient a and not b or c) and the use of model LDCs.

We thus find offsets between Prism and G395H and also between the two G395H detectors. The causes of the inter-modal and inter-detector offsets are not immediately clear. The G395H–Prism offsets could be due to astrophysical or instrumental variations between the two observations. However, they may also be due to differences in data processing, e.g. use of different reference files. Alternate systematic models to the second-order polynomial used here may be investigated to see whether these might affect the offsets seen, though as noted above use of a null systematic model (with no time-dependent trend) in G395H gives similar results to the baseline case. While the offset between Prism and G395H has several possible causes, we can rule out astrophysical variation for the G395H inter-detector offset between NRS1 and NRS2, which we detect here through the comparison with the Prism spectrum. While the cause of offsets like these needs continued investigation, they represent a systematic uncertainty that needs to be accounted for when analysing such spectra, and when comparing spectra taken at different times and/or with different instrument modes.

7 ATMOSPHERIC MODELLING

We use the AURA framework (Pinhas et al. 2018) to generate a number of simulated transmission spectra to compare to the obtained

[8]Dark current was not subtracted to prevent the dark current being subtracted twice.

Downloaded from https://academic.oup.com/mnras/article/531/2/2731/7670622 by guest on 03 July 2024

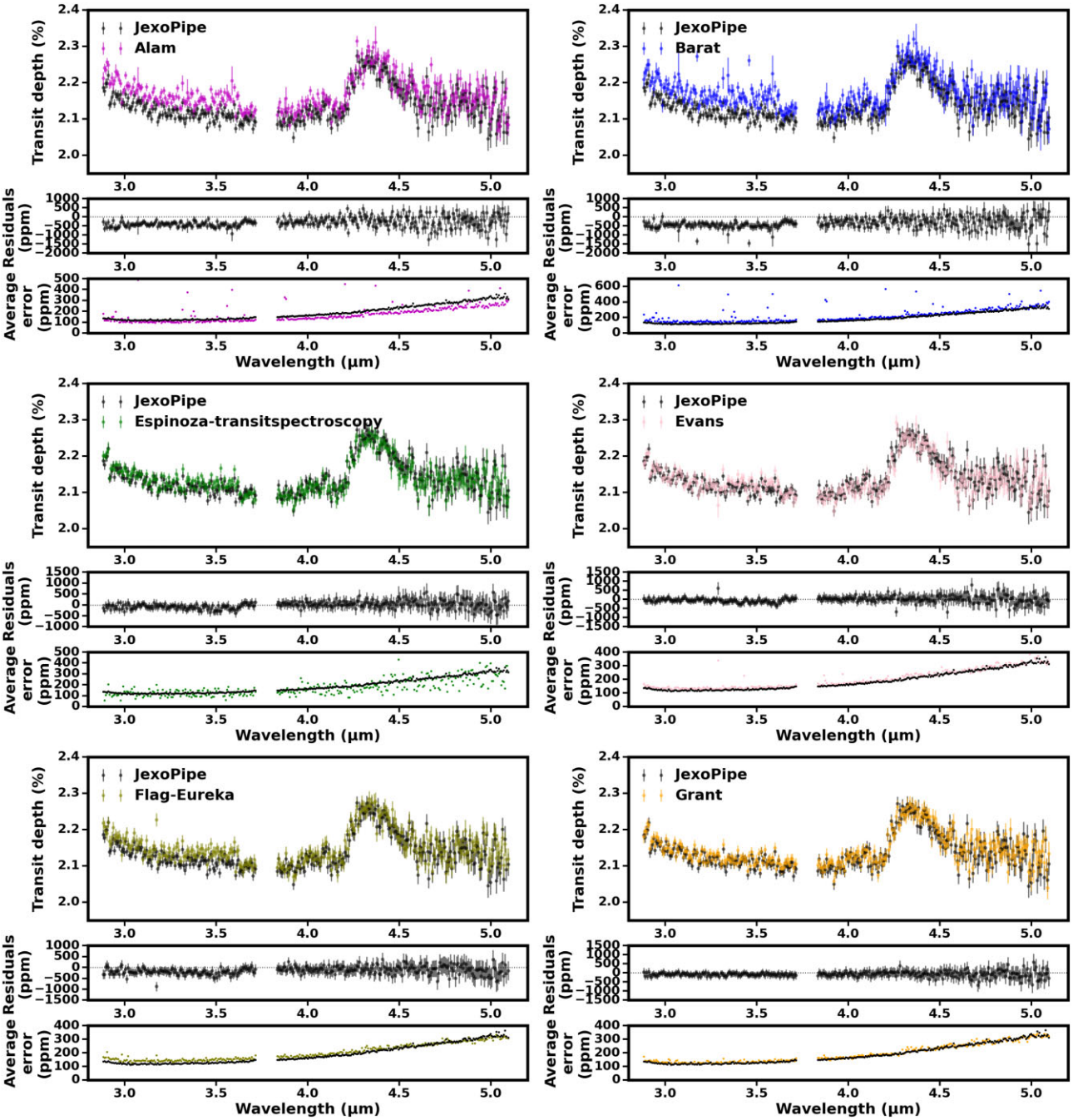

Figure 17. Comparison of our G395H result with six other G395H pipelines (listed by author and/or pipeline name) included in the study by Alderson et al. (2023). The JexoPipe result is rebinned to the wavelength grid of the comparison pipeline's publicly available spectrum. Residuals are JexoPipe minus the comparison pipeline spectrum. The average errors on the transit depths for the comparison spectrum and the JexoPipe spectrum when binned to the resolution of the comparison spectrum are shown in the lowest plots.

JWST NIRSpec observations. AURA treats the terminator as a 1D plane-parallel atmosphere in hydrostatic equilibrium with a uniform composition. For the models considered in this work, we treat the atmospheric regions giving rise to the transmission spectrum as isothermal.

We consider atmospheric opacity contributions from a number of gaseous species previously reported in the atmosphere of WASP-39 b (e.g. Fischer et al. 2016; Nikolov et al. 2016; Wakeford et al. 2018; Kirk et al. 2019; Ahrer et al. 2023; Alderson et al. 2023; Constantinou et al. 2023; Feinstein et al. 2023; JWST Transiting Exoplanet Community Early Release Science Team 2023; Rustamkulov et al. 2023). We specifically consider H_2O (Barber et al. 2006; Rothman et al. 2010), CO (Li et al. 2015), CO_2 (Tashkun et al. 2015), SO_2 (Underwood et al. 2016), H_2S (Azzam et al. 2016; Chubb et al. 2018), Na (Allard et al. 2019), and K (Allard, Spiegelman & Kielkopf 2016). We also include atmospheric extinction arising from Mie scattering through ZnS (Querry 1987) aerosols, again motivated by prior findings (Constantinou et al. 2023; Feinstein et al. 2023). We lastly include opacity contributions from H_2–H_2 and H_2–He collision-induced absorption (Borysow, Frommhold & Birnbaum

Downloaded from https://academic.oup.com/mnras/article/531/2/2731/7670622 by guest on 03 July 2024

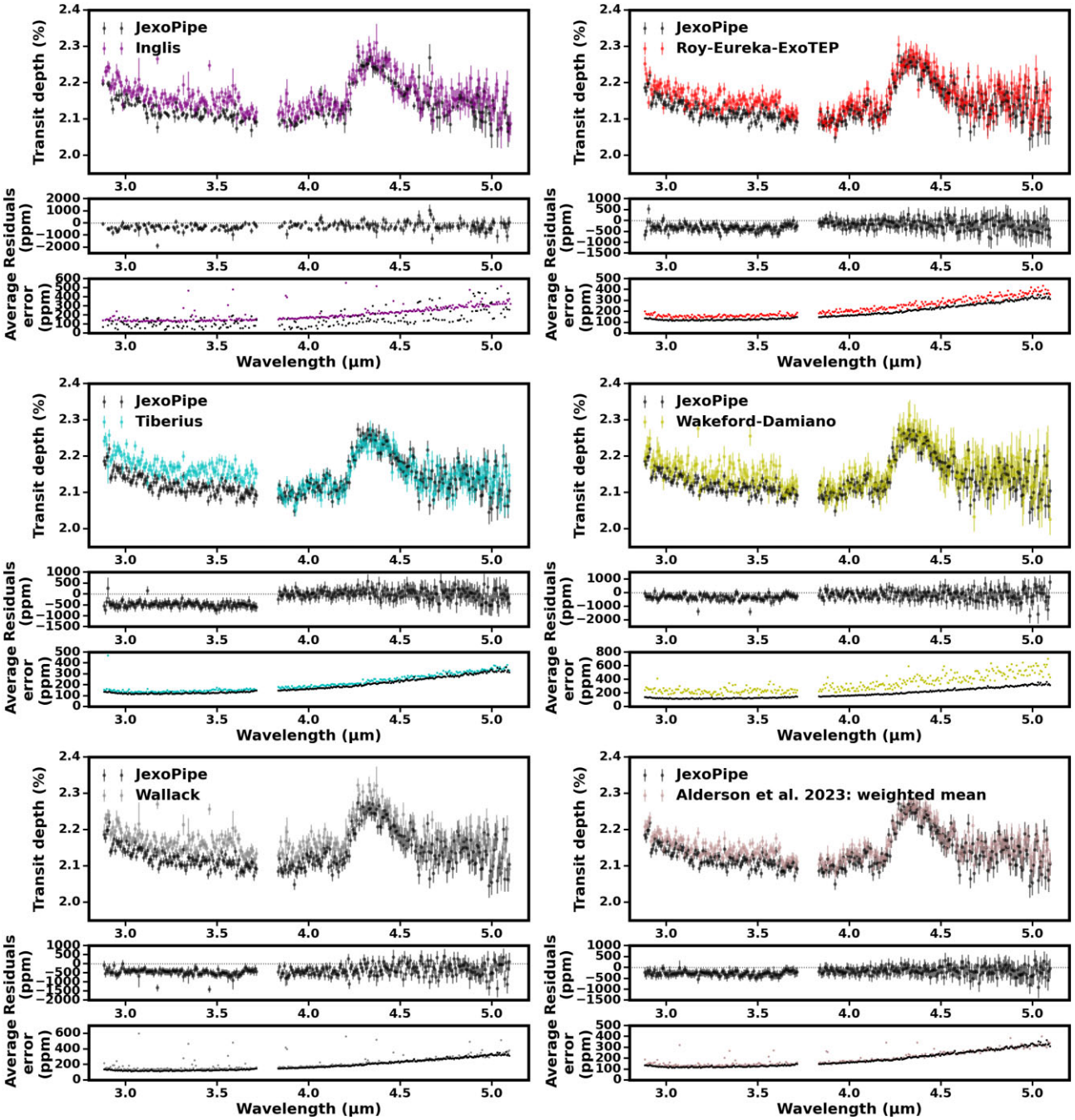

Figure 18. Comparison of our G395H result with five further G395H pipelines (listed by author and/or pipeline name) included in the study by Alderson et al. (2023) and the weighted mean result in that study. The JexoPipe result is rebinned to the wavelength grid of the comparison pipeline's publicly available spectrum. Residuals are JexoPipe minus the comparison pipeline spectrum. The average errors on the transit depths for the comparison spectrum and the JexoPipe spectrum when binned to the resolution of the comparison spectrum are shown in the lowest plots.

1988; Orton et al. 2007; Abel et al. 2011; Richard et al. 2012), which set the spectral baseline in the absence of aerosols.

For the nominal model shown in Fig. 19, the abundances of all gaseous species except SO_2 correspond to a 10× enhancement over solar values under thermochemical equilibrium (Burrows & Sharp 1999; Lodders & Fegley 2002; Madhusudhan & Seager 2011; Moses et al. 2013). For SO_2, which is the product of disequilibrium processes (Zahnle et al. 2009; Wang, Miguel & Lunine 2017; Polman et al. 2023; Tsai et al. 2023), we use a volume mixing ratio of 10^{-5}, which is largely consistent with literature findings (Constantinou et al. 2023; Rustamkulov et al. 2023; Tsai et al. 2023). We set the atmosphere's isothermal temperature to 900 K. The ZnS aerosols have a 40 per cent coverage fraction of the terminator and full vertical extent, with a mixing ratio of 5×10^{-7} and a modal particle radius of 0.01 μm.

The forward model shown in Fig. 19 provides a good fit to the observations and displays prominent spectral features consistent with previous studies (Alderson et al. 2023; Constantinou et al. 2023; Rustamkulov et al. 2023; Tsai et al. 2023). The data and model show absorption features from H_2O near 0.9, 1.1, 1.4, 1.9, and 2.9

Downloaded from https://academic.oup.com/mnras/article/531/2/2731/7670622 by guest on 03 July 2024

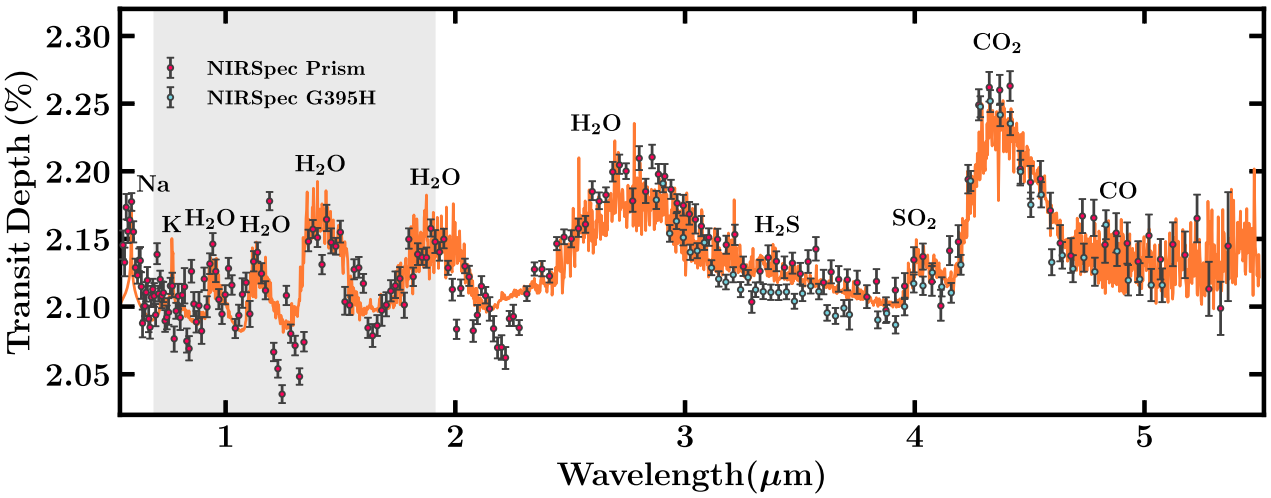

Figure 19. The obtained NIRSpec Prism and G395H transmission spectra of WASP-39 b are shown with error bars. Also shown is a nominal model assuming 10× solar elemental abundances and Mie scattering aerosols, as discussed in Section 7. The observations are binned to $R \sim 100$ for visual clarity, while the shaded area denotes the persistently saturated region for NIRSpec Prism as shown in Fig. 2.

μm. The latter feature also contains secondary spectral contributions from CO_2 and H_2S. Additionally, the blue end of the observations prominently shows a significant absorption feature from Na. Both NIRSpec Prism and G395H data sets show a highly prominent CO_2 absorption feature at ∼4.3 μm, which is reproduced by the forward model. There are also spectral contributions from SO_2, CO, and H_2O on either side of the large CO_2 feature. Specifically, SO_2 is responsible for a small peak at 4 μm, while CO and H_2O provide atmospheric opacity towards the red end of the spectrum. Lastly, ZnS aerosols are responsible for significant truncation of spectral features, particularly those of H_2O at wavelengths smaller than 2 μm.

While the above forward modelling does not provide a definitive retrieval of the present data set, the findings are broadly consistent with prior analyses of *JWST* observations of WASP-39 b with NIRSpec (Alderson et al. 2023; Constantinou et al. 2023; JWST Transiting Exoplanet Community Early Release Science Team 2023; Niraula et al. 2023; Rustamkulov et al. 2023) and other instruments (Ahrer et al. 2023; Alderson et al. 2023; Feinstein et al. 2023). Specifically, prior works also find highly prominent spectral features from H_2O and CO_2, with additional contributions from SO_2, CO, and H_2S. Moreover, many of the above works find that the observations are best explained by a supersolar atmospheric metallicity, with additional spectral contributions from non-grey clouds. A more comprehensive retrieval analysis of the present data can confirm this agreement with prior works.

As noted earlier, we find the greatest time-correlated noise in the Prism white light curve composed of wavelengths between 0.65 and 2 μm, which would suggest that the transit depth error bars in this range may be underestimated, precisely where we find largest discrepancy between the forward model and the data.

8 CONCLUSIONS

In this work, we present JexoPipe, a newly developed *JWST* pipeline for exoplanet transit spectroscopy. We applied it to observations of the warm Saturn WASP-39 b obtained in the ERS programme 1366 using the NIRSpec instrument in two contrasting configurations, Prism and G395H. We use JexoPipe to apply consistent pipeline procedures, LDCs, and system parameters to both data sets, enabling a comparative analysis of the spectra and instrument configurations.

We find a significant offset between Prism and G395H spectra, which is more pronounced for G395H NRS1. The Prism data baseline also reveals an offset between G395H NRS1 and NRS of the order of 40–50 ppm. This is consistent with an intra-detector offset reported by Moran et al. (2023) and supports the incorporation of detector offsets when interpreting such spectra (Madhusudhan et al. 2023). We cannot rule out astrophysical causes for the G395H–Prism offset; however, instrumental changes or differences in processing may be potential causes.

We note that the effect of omitting the reference pixel stage on the spectral baseline was more pronounced for NRS1 than NRS2. We also found that performing the group-level background subtraction before the linearity correction step resulted in a significant fall in the Prism spectrum baseline, such that the offsets with G395H were reduced (NRS1) or eliminated (NRS2), but no such change is seen for G395H itself. Further investigation of these differences is warranted.

We address the Prism saturation through a combination of increasing the `n_pix_grow_sat` argument to counteract detector 'blooming' and a custom group control stage. In the Prism, we find 'flipping' noise, which appears to result from the variation in number of groups with integration. Group control is used to mitigate this.

On choosing between Prism and G395H, each offers advantages and disadvantages. The Prism avoids any inter-detector offsets, but the final spectra have somewhat more noise, and correlated noise was detected in this study. The average errors on the spectrum transit depths are about 1.3× higher in Prism compared to those on G395H when binned to the Prism resolution. It also saturates easily. G395H is superior in terms of noise and allows for higher resolution spectra; however, there is the potential for inter-detector spectrum baseline offsets of the order of tens of ppm.

Using a nominal forward atmospheric model with 10× solar elemental abundances, we show that we recover water peaks at 1.1, 1.4, and 1.9 μm within the saturated region, although the scatter and deviation from the model are somewhat higher in this region than outside. We note that this is also the region where time-correlated noise is highest likely leading to an underestimation of the error bars on the spectrum, which might explain at least some of this added

Downloaded from https://academic.oup.com/mnras/article/531/2/2731/7670622 by guest on 03 July 2024

deviation. It remains to be seen whether the method used in this paper provides a more or less accurate recovery of the saturated region than previous methods.

The examination of JexoPipe involved a comparison with spectra obtained from previously developed pipelines. The greatest disagreement with pipelines used in Rustamkulov et al. (2023) occurs in the persistently saturated region. JexoPipe has reasonably good agreement with two pipelines used in Alderson et al. (2023) but has appreciable baseline differences with others.

In this early stage of *JWST* observations, development of independent pipelines such as JexoPipe allows us to compare results, ultimately leading to more robust scientific conclusions and the elucidation of optimal and best practice approaches for the processing of data from *JWST* exoplanet transit observations.

ACKNOWLEDGEMENTS

The authors acknowledge the *JWST* Transiting Exoplanet Community ERS team (PI: Batalha) and the ERO team (PI: Pontoppidan) for developing their observing programs with a zero-exclusive-access period. We thank NASA, ESA, CSA, STScI, and everyone whose efforts have contributed to the *JWST*, and the exoplanet science community for the thriving current state of the field. NM and MH acknowledge support from the MERAC Foundation, Switzerland, and the UK Science and Technology Facilities Council (STFC) Center for Doctoral Training (CDT) in Data Intensive Science at the University of Cambridge (STFC grant number ST/P006787/1), towards the doctoral studies of MH. We thank the anonymous reviewer for their helpful comments and suggestions.

DATA AVAILABILITY

This work is based on observations made with the NASA/ESA/CSA *JWST*. The publicly available data were obtained from the MAST 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*. These observations are associated with programmes 1091, 1366, 1541, and 2734. The data underlying this article will be shared on reasonable request to the corresponding author.

REFERENCES

Abel M., Frommhold L., Li X., Hunt K. L. C., 2011, J. Phys. Chem. A, 115, 6805
Ahrer E.-M. et al., 2023, Nature, 614, 653
Alderson L. et al., 2023, Nature, 614, 664
Allard N. F., Spiegelman F., Kielkopf J. F., 2016, A&A, 589, A21
Allard N. F., Spiegelman F., Leininger T., Molliere P., 2019, A&A, 628, A120
Azzam A. A. A., Tennyson J., Yurchenko S. N., Naumenko O. V., 2016, MNRAS, 460, 4063
Barber R. J., Tennyson J., Harris G. J., Tolchenov R. N., 2006, MNRAS, 368, 1087
Bean J. L. et al., 2018, PASP, 130, 114402
Beichman C. A., Rieke M., Eisenstein D., Greene T. P., Krist J., McCarthy D., Meyer M., Stansberry J., 2012, in Clampin M. C., Fazio G. G., MacEwen H. A., Oschmann J. M. J., eds, Proc. SPIE Conf. Ser. Vol. 8442, Space Telescopes and Instrumentation 2012: Optical, Infrared, and Millimeter Wave. SPIE, Bellingham, p. 84422N
Bell T. J. et al., 2023, Nature, 623, 709
Benneke B. et al., 2019, ApJ, 887, L14
Birkmann S. M. et al., 2022a, A&A, 661, A83
Birkmann S. M. et al., 2022b, in Coyle L. E., Matsuura S., Perrin M. D., eds, Proc. SPIE Conf. Ser. Vol. 12180, Space Telescopes and Instrumentation 2022: Optical, Infrared, and Millimeter Wave. SPIE, Bellingham, p. 121802P
Böker T. et al., 2023, PASP, 135, 038001
Borysow J., Frommhold L., Birnbaum G., 1988, ApJ, 326, 509
Brown T. M., 2001, ApJ, 553, 1006
Burrows A., Sharp C. M., 1999, ApJ, 512, 843
Charbonneau D., Brown T. M., Noyes R. W., Gilliland R. L., 2002, ApJ, 568, 377
Chubb K. L. et al., 2018, J. Quant. Spectrosc. Radiat. Transfer, 218, 178
Cohen Y., Olaes M. K., Grogin N. A., 2020, Instrument Science Report ACS 2020-7, Unusual Horizontal Charge Overflow from Saturated CCD Pixels in the ACS WFC: Discovery and Remediation. Space Telescope Science Institute Baltimore
Constantinou S., Madhusudhan N., Gandhi S., 2023, ApJ, 943, L10
Deming D. et al., 2013, ApJ, 774, 95
Doyon R. et al., 2012, in Clampin M. C., Fazio G. G., MacEwen H. A., Oschmann J. M. J., eds, Proc. SPIE Conf. Ser. Vol. 8442, Space Telescopes and Instrumentation 2012: Optical, Infrared, and Millimeter Wave. SPIE, Bellingham, p. 84422R
Dyrek A. et al., 2024, Nature, 625, 51
Espinoza N. et al., 2023, PASP, 135, 018002
Faedi F. et al., 2011, A&A, 531, A40
Feinstein A. D. et al., 2023, Nature, 614, 670
Ferruit P., Birkmann S., Böker T., Sirianni M., Giardino G., de Marchi G., Alves de Oliveira C., Dorner B., 2014, in Oschmann J. M. J., Clampin M., Fazio G. G., MacEwen H. A., eds, Proc. SPIE Conf. Ser. Vol. 9143, Space Telescopes and Instrumentation 2014: Optical, Infrared, and Millimeter Wave. SPIE, Bellingham, p. 91430A
Fischer P. D. et al., 2016, ApJ, 827, 19
Foreman-Mackey D., Hogg D. W., Lang D., Goodman J., 2013, PASP, 125, 306
Fournier-Tondreau M. et al., 2024, MNRAS, 528, 3354
Giardino G. et al., 2019, PASP, 131, 094503
Gillon M. et al., 2007, A&A, 471, L51
Grant D. et al., 2023, ApJ, 956, L29
Holmberg M., Madhusudhan N., 2023, MNRAS, 524, 377
Horne K., 1986, PASP, 98, 609
JWST Transiting Exoplanet Community Early Release Science Team, 2023, Nature, 614, 649
Kempton E. M. R. et al., 2023, Nature, 620, 67
Kirk J., López-Morales M., Wheatley P. J., Weaver I. C., Skillen I., Louden T., McCormac J., Espinoza N., 2019, AJ, 158, 144
Kirk J. et al., 2024, AJ, 167, 90
Lallo M., Hartig G., 2022, Technical Report JWST-STScI-008271, JWST Line-of-Sight Jitter Measurement during Commissioning.Space Telescope Science Institute, Baltimore
Li G., Gordon I. E., Rothman L. S., Tan Y., Hu S.-M., Kassi S., Campargue A., Medvedev E. S., 2015, ApJS, 216, 15
Lim O. et al., 2023, ApJ, 955, L22
Lodders K., Fegley B., 2002, Icarus, 155, 393
Lustig-Yaeger J. et al., 2023, Nat. Astron., 7, 1317
McCullough P. R., Crouzet N., Deming D., Madhusudhan N., 2014, ApJ, 791, 55
Madhusudhan N., 2018, Atmospheric Retrieval of Exoplanets. Springer, Cham, p. 1
Madhusudhan N., 2019, ARA&A, 57, 617
Madhusudhan N., Seager S., 2009, ApJ, 707, 24
Madhusudhan N., Seager S., 2011, ApJ, 729, 41
Madhusudhan N., Sarkar S., Constantinou S., Holmberg M., Piette A. A. A., Moses J. I., 2023, ApJ, 956, L13
Mancini L. et al., 2018, A&A, 613, A41
Manjavacas E. et al., 2018, AJ, 155, 11
May E. M. et al., 2023, ApJ, 959, L9
Moran S. E. et al., 2023, ApJ, 948, L11
Moses J. I. et al., 2013, ApJ, 777, 34
Newville M., Stensitzki T., Allen D. B., Rawlik M., Ingargiola A., Nelson A., 2016, Astrophysics Source Code Library, record ascl:1606.014
Nikolov N., Sing D. K., Gibson N. P., Fortney J. J., Evans T. M., Barstow J. K.,Kataria T., Wilson P. A. , 2016, ApJ, 832, 191

Downloaded from https://academic.oup.com/mnras/article/531/2/2731/7670622 by guest on 03 July 2024

Niraula P., de Wit J., Gordon I. E., Hargreaves R. J., Sousa-Silva C., 2023, ApJ, 950, L17
Orton G. S., Gustafsson M., Burgdorf M., Meadows V., 2007, Icarus, 189, 544
Pinhas A., Rackham B. V., Madhusudhan N., Apai D., 2018, MNRAS, 480, 5314
Polman J., Waters L. B. F. M., Min M., Miguel Y., Khorshid N., 2023, A&A, 670, A161
Pontoppidan K. M. et al., 2022, ApJ, 936, L14
Querry M., 1987, Chemical Research, Development and Engineering Center, U.S. Army Armament Munitions Chemical Command, Aberdeen Proving Ground, Maryland
Radica M. et al., 2023, MNRAS, 524, 835
Ricci D. et al., 2015, PASP, 127, 143
Richard C. et al., 2012, J. Quant. Spectrosc. Radiat. Transfer, 113, 1276
Rieke G. H. et al., 2015, PASP, 127, 584
Rothman L. et al., 2010, J. Quant. Spectrosc. Radiat. Transfer, 111, 2139
Rustamkulov Z. et al., 2023, Nature, 614, 659
Sarkar S., Pascale E., Papageorgiou A., Johnson L. J., Waldmann I., 2021, Exp. Astron., 51, 287
Seager S., Sasselov D. D., 2000, ApJ, 537, 916
Sing D. K. et al., 2016, Nature, 529, 59
Tashkun S. A., Perevalov V. I., Gamache R. R., Lamouroux J., 2015, J. Quant. Spectrosc. Radiat. Transfer, 152, 45
Taylor J. et al., 2023, MNRAS, 524, 817
Tsai S.-M. et al., 2023, Nature, 617, 483
Tsiaras A., Waldmann I. P., Tinetti G., Tennyson J., Yurchenko S. N., 2019, Nat. Astron., 3, 1086
Underwood D. S., Tennyson J., Yurchenko S. N., Huang X., Schwenke D. W., Lee T. J., Clausen S., Fateev A., 2016, MNRAS, 459, 3890
Wakeford H. R. et al., 2018, AJ, 155, 29
Wang D., Miguel Y., Lunine J., 2017, ApJ, 850, 199
Zahnle K., Marley M. S., Freedman R. S., Lodders K., Fortney J. J., 2009, ApJ, 701, L20

APPENDIX A: WHITE LIGHT CURVE POSTERIORS

Here, we show the joint posterior distributions from the MCMC parameter estimations using white light curves from the different instrument modes. Fig. A1 shows the results for the Prism baseline case with all wavelengths included and also when only wavelengths >2 μm are included. Fig. A2 shows the results for the G395H baseline case (with a'/R_s and i fixed to Prism values), and for the 'independent' cases (where a'/R_s and i are fitted for). Fig. A3 shows the results for cases where four-factor model LDCs were used.

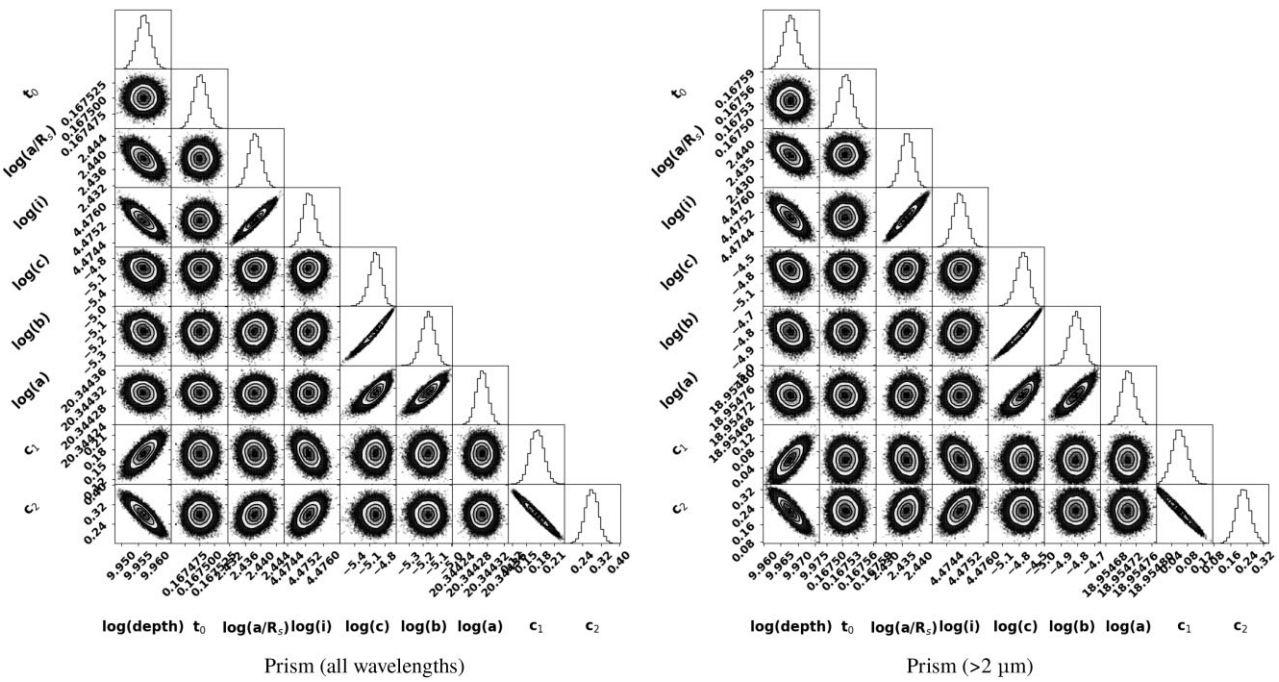

Figure A1. Joint posterior probability distributions from white light curve MCMC processing for Prism.

Downloaded from https://academic.oup.com/mnras/article/531/2/2731/7670622 by guest on 03 July 2024

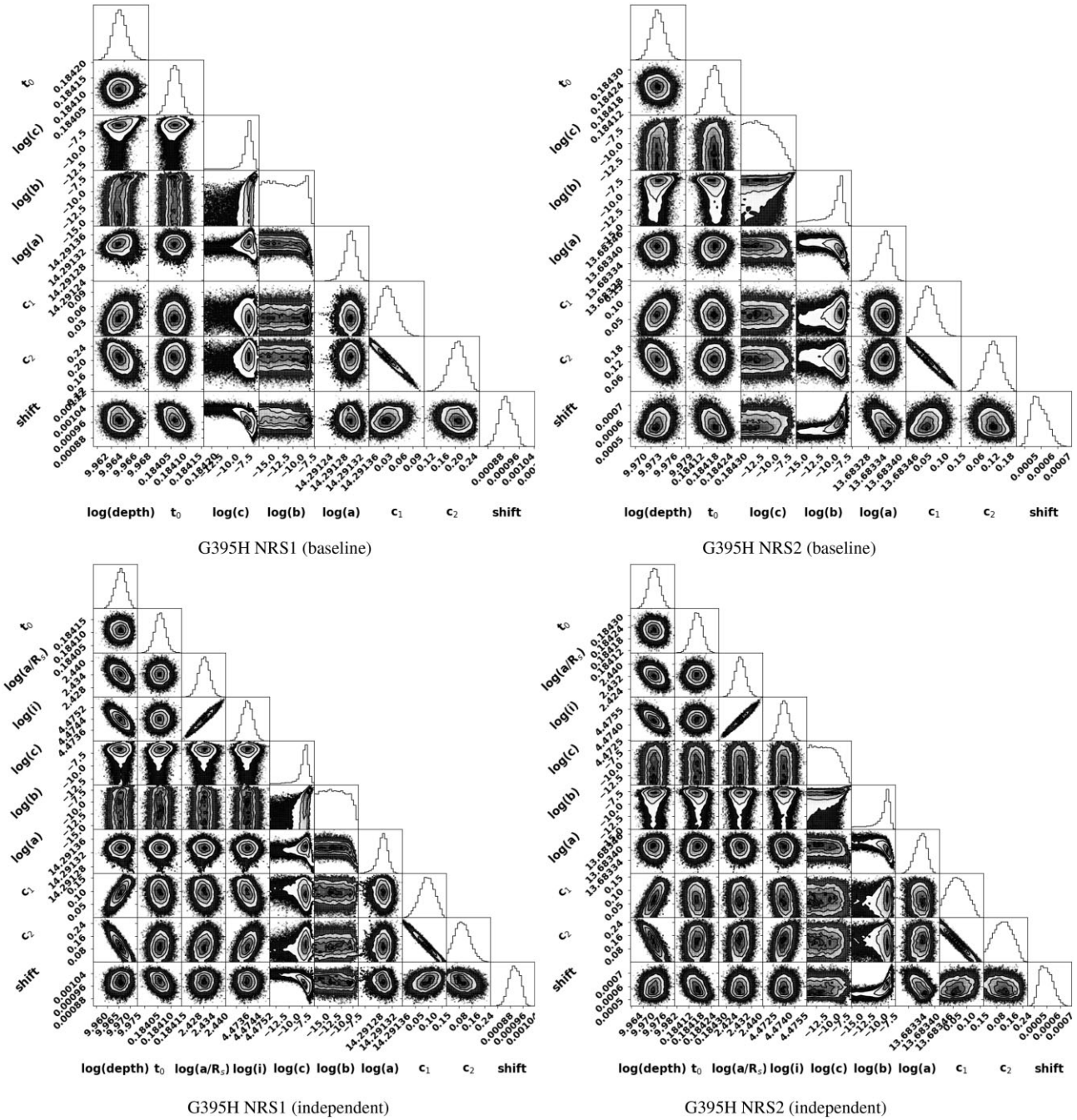

Figure A2. Joint posterior probability distributions from white light curve MCMC processing for G395H.

Downloaded from https://academic.oup.com/mnras/article/531/2/2731/7670622 by guest on 03 July 2024

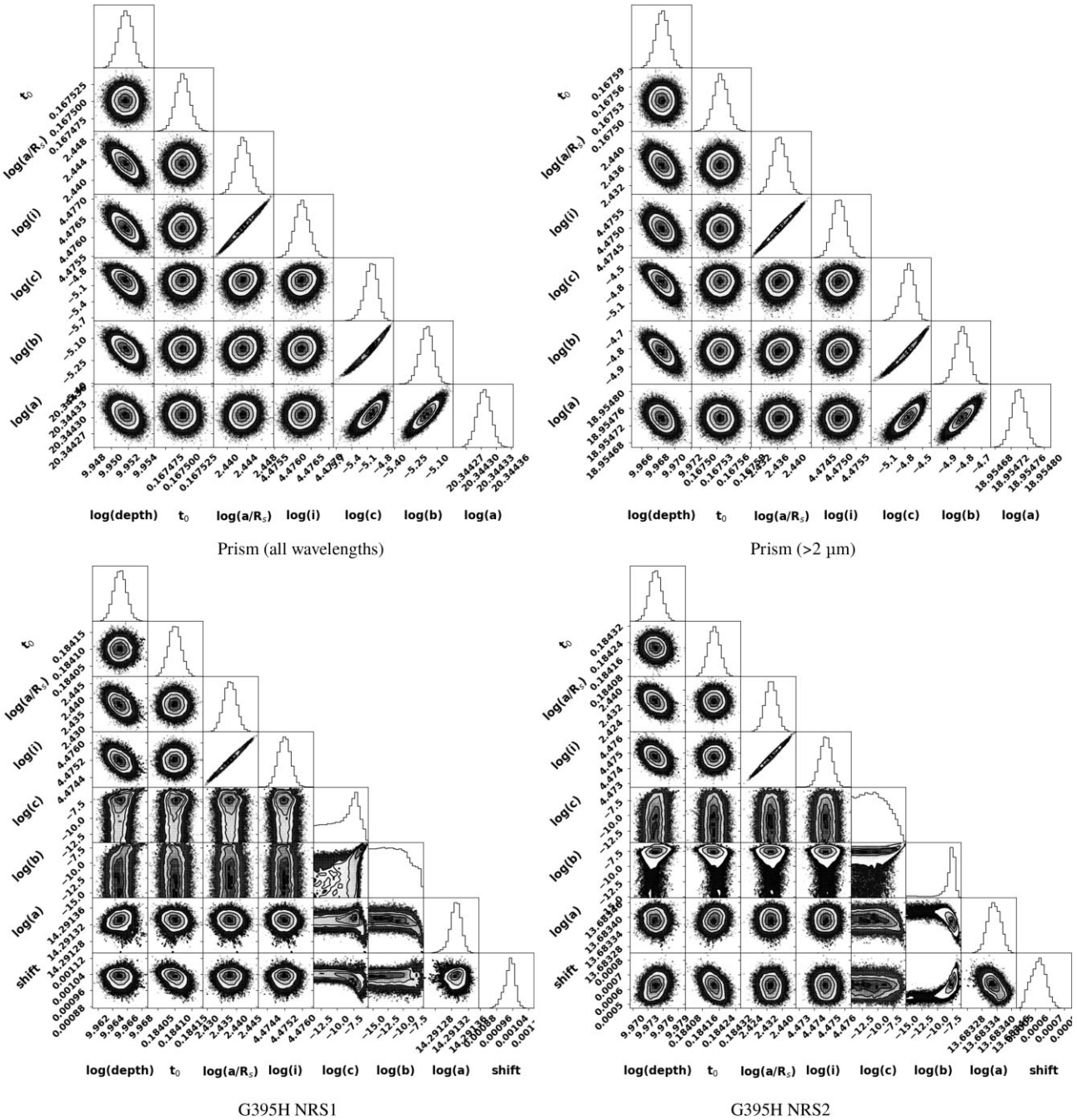

Figure A3. Joint posterior probability distributions from white light curve MCMC processing when using model four-factor LDCs.

APPENDIX B: SPECTRAL LIGHT-CURVE FITS AND POSTERIORS

Here, we show the light-curve fits and corresponding joint posterior distributions from the MCMC parameter estimations for three example spectral light curves at full resolution. Fig. B1 shows examples from the baseline cases for Prism (top), G395H NRS1 (middle), and G395H NRS2 (bottom) with central wavelengths shown above each light-curve plot.

Downloaded from https://academic.oup.com/mnras/article/531/2/2731/7670622 by guest on 03 July 2024

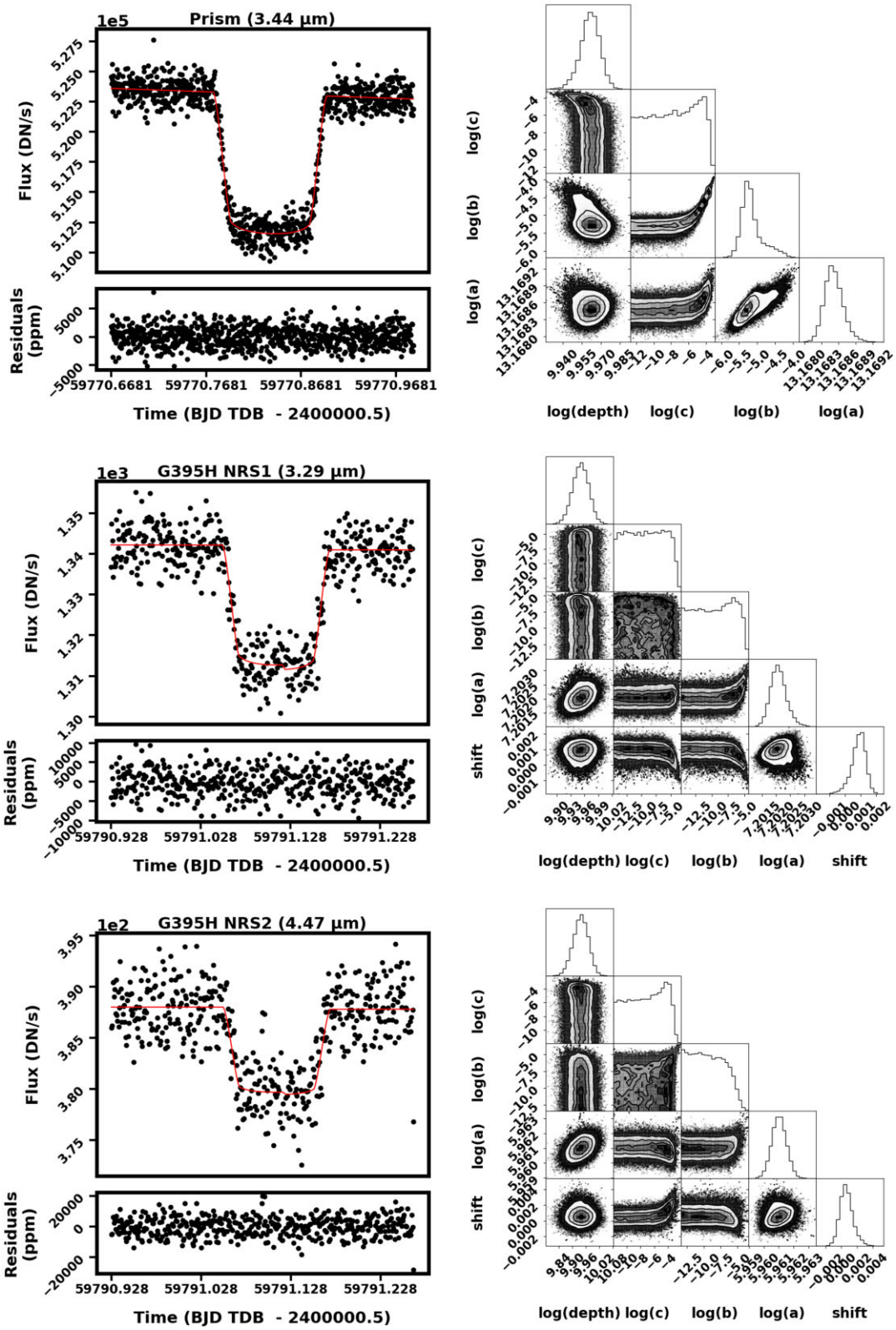

Figure B1. Joint posterior probability distributions and light-curve fits for three example spectral light-curve fits at full resolution.

This paper has been typeset from a TEX/LATEX file prepared by the author.

Downloaded from https://academic.oup.com/mnras/article/531/2/2731/7670622 by guest on 03 July 2024

© 2024 The Author(s).
Published by Oxford University Press on behalf of Royal Astronomical Society. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.