

Hot Rocks Survey

III. A deep eclipse for LHS 1140c and a new Gaussian process method to account for correlated noise in individual pixels

Mark Fortune^{1,*}, Neale P. Gibson¹, Hannah Diamond-Lowe^{2,3}, João M. Mendonça^{3,4,5}, Amélie Gressier², Daniel Kitzmann⁶, Natalie H. Allen⁷, Prune C. August³, Jegug Ih², Erik Meier Valdés⁸, Merlin Zraggen⁹, Lars A. Buchhave³, Brice-Olivier Demory^{6,9,10}, Néstor Espinoza^{2,7}, Kevin Heng^{11,10,12,13}, Kathryn Jones⁹, and Alexander D. Rathcke³

(Affiliations can be found after the references)

Received 20 February 2025 / Accepted 27 May 2025

ABSTRACT

Time-series photometry at mid-infrared wavelengths is becoming a common technique to search for atmospheres around rocky exoplanets. This method constrains the brightness temperature of the planet to determine whether heat redistribution is taking place, which would be indicative of the presence of an atmosphere, or whether the heat is reradiated from a low-albedo bare rock. By observing at $15\mu\text{m}$, we are also highly sensitive to CO_2 absorption. We observed three eclipses of the rocky super-Earth LHS 1140c, using MIRI/Imaging with the F1500W filter. We found a significant variation in the initial settling ramp for these observations and identified a potential trend between the detector settling and the previous filter used by MIRI. We analysed our data using aperture photometry, however, we also developed a novel approach, which performs a joint fit of the pixel light curves using a shared eclipse model and a flexible multi-dimensional Gaussian process which can model changes in the PSF over time. Using simulated data, we demonstrate that our method has the ability to weight away from particular pixels that exhibit increased systematics, allowing for the recovery of eclipse depths in a more robust and precise way. Both methods, as well as an independent analysis, have detected the eclipse at $>5\sigma$, while recovering an eclipse depth consistent with a low-albedo bare rock. We measured a dayside brightness temperature of $T_{\text{day}} = 561 \pm 44$ K, close to the theoretical maximum of $T_{\text{day; max}} = 537 \pm 9$ K. We rule out a wide range of atmospheric forward models to $>3\sigma$, including pure CO_2 atmospheres with surface pressure ≥ 10 mbar and pure H_2O atmospheres with surface pressure ≥ 1 bar. Our strict constraints on potential atmospheric composition, in combination with future observations of the exciting outer planet LHS 1140b, could provide a powerful benchmark for understanding atmospheric escape around M dwarfs.

Key words. methods: data analysis – methods: statistical – techniques: photometric – planets and satellites: atmospheres – stars: individual: LHS 1140

1. Introduction

The search for atmospheres on rocky exoplanets is one of the major science goals defining the early JWST era. However, we still lack any definitive evidence of an atmosphere around any rocky planet, with only a few hints of signals reported to date (e.g. August et al. 2025; Hu et al. 2024; Patel et al. 2024; Gressier et al. 2024; Banerjee et al. 2024). At present, M-dwarf systems offer the best hope of characterising rocky exoplanet atmospheres using current facilities and they have therefore become the targets of numerous JWST programs (e.g. Greene et al. 2023; Zieba et al. 2023b; Weiner Mansfield et al. 2024; Alam et al. 2025). It remains an open question as to how many M-dwarf targets can retain atmospheres due to the significant flaring and intense EUV emission these stars are believed to exhibit in their first ~ 1 Gyr (Shields et al. 2016). However, there is a large degree of uncertainty around the extent of this activity, additionally, a planet might retain a secondary atmosphere after this active phase finishes (Krissansen-Totton et al. 2024) or could even form a new secondary atmosphere from volcanic outgassing (Kite & Barnett 2020; Tian & Heng 2024). Ultimately, the only way we

can answer this question is by observing a sample of M-dwarf targets and measuring whether or not they host atmospheres.

There are multiple approaches to inferring the presence of an atmosphere, with different methods varying in observational cost. We used eclipse photometry over the $15\mu\text{m}$ wavelength band offered by the MIRI F1500W filter. This approach assumes that the planet being studied is tidally locked to its host star (which is expected for our target from dynamical modelling of the LHS 1140 system; Gomes & Ferraz-Mello 2020) and asserts that the surfaces most likely to exist on rocky worlds within a given temperature range ($300\text{ K} < T_{\text{eq}} < 880\text{ K}$) would possess low surface albedos of $A < 0.2$ (Koll et al. 2019; Mansfield et al. 2019). As there should be insignificant heat redistribution on a planet without a substantial atmosphere (Joshi et al. 1997; Selsis et al. 2011; Koll 2022), we can expect this to result in a hot dayside, which would then produce strong thermal emission at mid-infrared (MIR) wavelengths. Alternatively, a planet with a substantial atmosphere should display more efficient heat redistribution to its nightside, cooling the dayside and reducing its thermal emission. In addition, carbon dioxide (CO_2) hosts a strong absorption band at $15\mu\text{m}$, which can further reduce this eclipse depth if present.

* Corresponding author: fortunma@tcd.ie

The advantages of infrared eclipse photometry are that it can require fewer observing hours to identify an atmosphere compared to observing full phase curves (Koll et al. 2019). Unlike transmission spectroscopy, our observations are not affected by the transit light source effect (TLS; Rackham et al. 2018) as we observe the planet as it passes behind its star, avoiding degeneracies in distinguishing a planetary atmosphere from stellar surface inhomogeneities. Compared with eclipse spectroscopy, photometry offers a higher signal-to-noise ratio (S/N), which helps reduce the observational cost – albeit at the expense of increased degeneracy in our interpretations (Hammond et al. 2025).

These benefits are the motivation behind the Hot Rocks Survey (GO 3730 PI: Diamond-Lowe, Co-PI: Mendonça), a large JWST Cycle 2 programme aimed at observing nine rocky planets with equilibrium temperatures ranging from $T_{\text{eq}} \sim 420\text{--}910$ K, using $15\mu\text{m}$ eclipse photometry. Of these targets, LHS 1140c has the lowest equilibrium temperature ($T = 422 \pm 7$ K; Cadieux et al. 2024a) and lowest instellation ($S \approx 5 S_{\oplus}$; similar to Mercury at aphelion) of the sample. With a mass of $M = 1.91 \pm 0.06 M_{\oplus}$ and radius of $1.272 \pm 0.026 R_{\oplus}$ it can be considered a super-Earth. This gives it a reasonably high escape velocity of $v = 13.7 \pm 0.3$ km/s (using constraints from Cadieux et al. 2024a), which is larger than any Solar System rocky body. Combining these properties together makes it a high priority target as the cosmic shoreline model suggests atmospheres are more likely to be retained by planets with low instellation (in particular low integrated EUV flux) and high escape velocities (Zahnle & Catling 2017). Results from models of atmospheric escape suggests LHS 1140c might well have retained a secondary atmosphere (Kite & Barnett 2020; Chatterjee & Pierrehumbert 2024) and observations from Swift (Gehrels et al. 2004) have shown the LHS 1140 system shows relatively low levels of present day NUV flux and estimated FUV flux for a star of this type (Spinelli et al. 2019). However, we note recent work suggests mid-type M dwarfs such as LHS 1140 may have had longer early active periods than previous models predicted, making atmospheric retention more challenging (Pass et al. 2025).

LHS 1140c was originally discovered by Ment et al. (2019) and is the innermost known planet of the system. The only other confirmed planet in the system, LHS 1140b, lies within its star’s habitable zone and has a low-density with a mass of $M = 5.60 \pm 0.19 M_{\oplus}$ and radius of $1.730 \pm 0.025 R_{\oplus}$, potentially consistent with a water world or mini-Neptune (Cadieux et al. 2024a). Transit observations using NIRISS/SOSS (Cadieux et al. 2024b) and NIRSpec (Damiano et al. 2024) suggest that LHS 1140b is inconsistent with the mini-Neptune case due to the flat transmission spectrum recovered. This implies that it could be a water world with a high mean molecular weight atmosphere. The NIRISS/SOSS observations happened to capture a transit of our target LHS 1140c entirely within a transit of LHS 1140b, although there was a low S/N coming from just a single transit and the transmission spectrum was consistent with a flat line with no molecular features or scattering-slope from hazes detected. Additionally, it has previously been noted that there is a 4σ discrepancy between the transit depths of LHS 1140c from TESS observations and a single Spitzer observation at $4.5\mu\text{m}$. This result suggests possible absorption from CO_2 (Cadieux et al. 2024a). However, these recent NIRISS/SOSS observations appear to show that the TESS observations underestimated the transit depth, with the new observations lying between the TESS and Spitzer transit depths. Eclipse observations using the F1500W filter can further resolve this discrepancy as our observations are highly sensitive to the presence of CO_2 .

One of the lessons we have learned is that systematics affect at least some of the observations in our survey (e.g. August et al. 2025; Meier Valdés et al. 2025). In addition, there can be significant variation in the strength and sign of the initial settling ramp which covers at least the first $\sim 30\text{--}60$ minutes of our observations. We present a possible explanation for these variations related to the previous filter used by MIRI, as we note it remains in place throughout target acquisition. In addition, we present a new Gaussian process (GP) based approach which fits the eclipse depth at the level of individual pixels without performing an aperture extraction. We also retrieve information about how the point spread function (PSF) changes in time and characterise the strength of interpixel capacitance. Our method has the benefit of being able to identify and weight away from systematics found in individual pixels, which we demonstrate using simulated data. In one of our observations, this method identifies a possible persistence effect caused by a cosmic ray.

Our work is laid out as follows: first we introduce our new pixel-based approach to account for systematics in Section 2 and we demonstrate the advantages of this method using simulated data in Section 3. Those primarily interested in the observations may move straight to Section 4 which presents the analysis of our observations using the standard approach of aperture photometry as well as using our pixel-based approach. It also includes our findings on MIRI detector settling. Section 5 describes our modelling of LHS 1140c and the interpretation of our recovered eclipse depth. Section 6 presents a discussion of the implications of our findings and elaborate on the strengths and weaknesses of our new method and where it may be extended. Our conclusions are outlined in Section 7.

2. Methods of fitting time-series photometry

2.1. Motivation for a new pixel-fitting method

There have been a number of methods developed to analyse time-series photometric observations. After the data cleaning, a standard approach is to define an aperture region where flux from the target star is greatest and add up the flux within this region with a uniform weighting on each pixel. This classic approach is known as aperture extraction and can perform well for bright targets where any remaining background noise from imperfections in the background subtraction is minimal. It has a couple of limitations, the choice of the aperture region and shape is quite arbitrary, but choosing too wide an aperture reduces the S/N due to increased background noise while too small an aperture reduces S/N from missed flux. Typically an intermediate aperture radius is chosen which maximises S/N. However, applying a uniform weighting for each pixel is generally not optimal as brighter pixels will be less impacted by background noise than outer pixels and might ideally be more strongly weighted.

Optimal extraction attempts to remedy these issues by fitting for the PSF shape and estimating the Poisson noise in each pixel (Horne 1986). This makes it possible to weight each pixel in a way which maximises S/N. For observations which contain a lot of background noise, optimal extraction may result in a significant improvement in the S/N. This is because optimal extraction weights each pixel based on its particular S/N value and can be mathematically shown to maximise the S/N in the presence of uncorrelated noise.

However, the situation is more complicated in the presence of time-correlated noise. The Spitzer Space Telescope was known to suffer from systematics due to ‘intrapixel sensitivity’, where the efficiency of each pixel varies across the surface

of the pixel. In combination with instability in the telescope pointing and undersampling of the PSF, this resulted in systematics that could be different in each pixel (Ingalls et al. 2012). The first work to measure an exoplanet eclipse depth used detrending of the PSF centroid position to correct for this effect (Charbonneau et al. 2005). The technique of BiLinearly-Interpolated Subpixel Sensitivity (BLISS) mapping was later developed to account for this, directly mapping the sensitivity across each pixel’s surface (Stevenson et al. 2012). In addition, pixel-level decorrelation (PLD) also corrected for this issue by detrending the systematics in the extracted light curve against normalised pixel light curves (Deming et al. 2015). These techniques were used in early work characterising rocky exoplanet heat redistribution using phase curves from Spitzer (Demory et al. 2016; Kreidberg et al. 2019). Other relevant techniques applied to Spitzer photometric time series include GPs which use centroid positions as input variables (Evans et al. 2015). Independent component analysis (ICA) – introduced to exoplanet time series in Waldmann (2012); Waldmann et al. (2013) – has been applied to disentangle astrophysical and systematic effects and works by calculating statistically independent trends in the pixel time series (Morello et al. 2014; Morello 2015; Morello et al. 2016). A separate approach used to account for systematics in Kepler time series is the ‘causal pixel model’, which can distinguish systematics in each pixel by comparing them to multiple other stars on the detector (Wang et al. 2016). However, this approach is not applicable to our observations of a single star.

Time-series photometry with the MIRI/Imaging mode is expected to suffer significantly less from intrapixel sensitivity variations compared to Spitzer due to the PSF being spread out over more pixels and the telescope pointing being more stable. There is also a difference in detector technology, the Si:As detectors on Spitzer were not significantly affected by intrapixel sensitivity while the InSb detectors were (Knutson et al. 2008), MIRI uses Si:As detectors. We performed some simple tests simulating a PSF with a similar width to F1500W observations and modeling the pointing stability of JWST, which confirmed that effects such as intrapixel sensitivity and errors in flat fielding should be completely negligible compared to the photometric precision of these observations – even when significant intrapixel or interpixel sensitivity variations were modelled (see Appendix B for more details).

However, at our early stage in understanding systematics from JWST instruments, we should be careful in assuming our observations are completely free from systematics. In addition, the more approaches we can develop to analyse these data, the less sensitive we are to the limitations of any individual method. With that in mind, we introduce our new pixel-fitting approach to analyse these data, which is radically different from existing approaches such as PLD or BLISS mapping. This method combines the benefits of optimal extraction by fitting for the shape of the PSF and accounting for the amplitude of Poisson noise in each pixel. However, it also fits for the amplitude of correlated noise in each pixel. This allows it to weight away from pixels that suffer from increased systematics, in addition to weighting based on reduced Poisson noise. This method is the first we are aware of which applies GPs to joint-fit pixel time series in astronomical data. Our approach can be more precise and robust when systematics are present in single pixels and can be used as an alternative to aperture photometry or optimal extraction for eclipse depth measurements. It also quantifies multiple detector effects in the process, such as constraining the amplitude of background noise and the amplitude of correlated Poisson noise resulting from interpixel capacitance.

2.2. Introduction to our new pixel-fitting method

Gaussian processes are a method often used to fit correlated noise (also known as systematics) in exoplanet light curves (e.g. Gibson et al. 2012, 2013; Evans et al. 2015; Foreman-Mackey et al. 2017; Aigrain & Foreman-Mackey 2023). To apply them to an eclipse light curve, we must first model the eclipse signal using a deterministic transit model (e.g. using equations from Mandel & Agol 2002). This deterministic model is referred to as the ‘mean function’, μ , and the underlying assumption of the method is that we can model the noise around this signal as a random draw from a covariance matrix built using a kernel function. That is to say, we model the noise as having a given amplitude or height scale, h , and the covariance in the noise between any two points is described by a kernel function which typically decreases as a function of the separation of the two points using a specific metric, such as time. For example, the exponential kernel combined with white noise is a common choice for a kernel function:

$$k(t, t') = h^2 \exp\left(-\frac{|t - t'|}{l}\right) + \sigma^2 \delta_{tt'}, \quad (1)$$

where t, t' could be the timestamps of two datapoints, h is the height scale and l the length scale of correlated noise, σ is the amplitude of (uncorrelated) white noise, and $\delta_{tt'}$ is the Kronecker delta which equals unity when $t = t'$ and zero otherwise.

If we want to fit multiple time series simultaneously (e.g. joint-fitting individual pixel light curves), then we can use a 2D GP model, where the dataset lies along a 2D grid of pixels-by-time. In this case, our kernel function can describe correlation as a function of separation in time and as a function of distance away on the detector. As the computational cost of GPs can scale as the cube of the total number of datapoints, namely, $O(N_p^3 N_t^3)$ for N_p pixels and N_t time points, we need to exploit a particular GP optimisation to make this method computationally tractable for our datasets. We chose the optimisation originally identified in Rakitsch et al. 2013, and introduced to exoplanet time series by Fortune et al. (2024), as it has sufficient flexibility in the choice of kernel functions we can use and it scales efficiently as $O(2N_p^3 + 2N_t^3 + N_p N_t (N_p + N_t))$, making the analysis computationally feasible. The method calculates the exact log likelihood value without approximations and works by breaking up the expensive decomposition of the covariance matrix into separate covariance matrices for each dimension, see Fortune et al. (2024) for more details. It is implemented in an open-source Python package called LUAS which is available on GitHub¹ along with documentation and tutorials (Fortune et al. 2024).

2.3. Pixel-fitting mean function

First, we need to model the PSF and account for how it will move in each integration due to telescope pointing instability. The PSF was not found to be well-characterised by simple parameterisations such as a 2D Gaussian profile or using the models generated by the package WEBBPSF (Perrin et al. 2014) and so instead we directly fit for the average flux in time for each pixel included in the fit. We also fit for a linear trend in flux for each pixel (with the slope denoted as $T_{\text{grad};(i,j)}$) as different pixels can show quite distinct slopes in time throughout the observations. To account for movements in telescope pointing, we fit for shifts in the PSF position along the x and y directions for each integration. Finally, the amplitude of the PSF as a function of time is fit with an eclipse model with a single shared set of eclipse

¹ <https://github.com/markfortune/luas>

parameters featuring the eclipse depth, $d = f_p/f_*$; period, P ; system scale, a/R_* ; radius ratio, R_p/R_* ; impact parameter, b ; central transit time, T_0 ; eccentricity, e , and argument of periastron, ω . We also included a shared exponential ramp model between all pixels with height scale, h_{ramp} , and length scale, l_{ramp} , to account for the sharp changes in flux typically seen in the first ~60 minutes of the observations; however, this ramp does appear quite different between pixels in the first ~30–45 minutes of the observations (as discussed in Section 4.5). Thus, we avoided fitting the first 45 minutes of data with this method. Alternatively it may be possible to fit this with a separate ramp for each pixel but we did not explore this in this work.

Combining the eclipse and settling effects, we model the light curve for each pixel (i, j) using:

$$f(i, j, t) = f_{\text{eclipse}}(T_0, P, a/R_*, R_p/R_*, b, d, \sqrt{e} \cos \omega, \sqrt{e} \sin \omega, t) \\ * f_{(i,j)}(\mathbf{f}, \Delta\mathbf{x}(t), \Delta\mathbf{y}(t)) \\ * \left[1 + T_{\text{grad}(i,j)}(t - t_{\text{mid}}) + h_{\text{ramp}} \exp\left(-\frac{t - t_{\text{init}}}{l_{\text{ramp}}}\right) \right], \quad (2)$$

where $f_{(i,j)}(\mathbf{f}, \Delta\mathbf{x}(t), \Delta\mathbf{y}(t))$ is the flux of pixel (i, j) , as calculated by a 2D interpolating function which takes a grid of average flux values located at the central position of each pixel and linearly interpolates to a new grid offset by $\Delta\mathbf{x}(t), \Delta\mathbf{y}(t)$; namely, the x and y shift in PSF position for the integration at a time, t ; $\mathbf{f}$ may be thought of as the baseline flux for each pixel light curve, while $\Delta\mathbf{x}(t)$ and $\Delta\mathbf{y}(t)$ fit for the time series of telescope pointing movements. Then, $\mathbf{f}$, $\Delta\mathbf{x}(t)$ and $\Delta\mathbf{y}(t)$ were all fit for as free parameters. We note we chose to fit for $\sqrt{e} \cos \omega$ and $\sqrt{e} \sin \omega$ instead of e and ω as it puts a simple uniform prior on e (Eastman et al. 2013).

2.4. Pixel-fitting kernel function

There are multiple noise processes which are correlated between detector pixels that are not normally required to be accounted for when fitting an aperture extracted light curve. However, if we are trying to fit an eclipse depth directly using the pixel light curves, then we must account for the correlation in the noise between different pixels in order to avoid underestimating or overestimating our eclipse depth uncertainty. We present each noise process to be accounted for in this section.

We must account for how each noise process is correlated both between different pixels and also in time. To meet the restrictions of the 2D GP optimisation from Fortune et al. (2024), we must assume that these correlations are separable and that the full kernel function of the dataset can be represented as a sum of two kernel functions separable as a product of pixel and time kernel functions:

$$k(i, j, t, i', j', t') = k_p(i, j, i', j')k_t(t, t') + s_p(i, j, i', j')s_t(t, t'). \quad (3)$$

Here, we define two pixel kernel functions, k_p and s_p , and two time kernel functions, k_t and s_t . The terms in k_p and k_t describe the systematics which are correlated over time, while the terms in s_p and s_t describe white noise or correlations between pixels which are instantaneous in time. We introduce each noise process in turn over the rest of this section, but the overall form of our kernel function is

$$k(i, j, t, i', j', t') = [k_{p,\text{FCS}} + k_{p,\text{IPS}} + k_{p,\text{CS}}] \exp\left(-\frac{|t - t'|}{l_t}\right) \\ + [s_{p,\text{wn+IPC}} + s_{p,\text{backgr}}] \delta_{tt'}, \quad (4)$$

where we have three time-correlated systematics processes. Due to the restrictions of this GP optimisation, they must all share the same time kernel function with a shared length scale (chosen to be the exponential kernel), but with different correlations as a function of separation of pixels on the detector. These correlations between pixels are denoted by the pixel kernel functions to be described in this section: $k_{p,\text{FCS}}$, $k_{p,\text{IPS}}$, and $k_{p,\text{CS}}$. We also have multiple noise processes that we modeled as independent in time, including correlations associated with the background subtraction, as well as white noise which must be adjusted to account for interpixel capacitance (IPC). These are described by the pixel kernel functions $s_{p,\text{backgr}}$ and $s_{p,\text{wn+IPC}}$. Note that our kernel function is in the form of Equation (3) and that none of the pixel kernel functions can have any dependence on time.

Depending on the noise process, it can either be simpler to describe the kernel function in terms of the covariance matrices that it builds or by writing out the kernel function explicitly; thus, we sometimes switch between these notations. Our kernel function restriction may equivalently be expressed with Equation (5), where the full covariance matrix of the dataset must be expressible as the sum of two Kronecker products

$$\mathbf{K} = \mathbf{K}_p \otimes \mathbf{K}_t + \mathbf{\Sigma}_p \otimes \mathbf{\Sigma}_t. \quad (5)$$

Here, we have defined two pixel covariance matrices $\mathbf{K}_p$ and $\mathbf{\Sigma}_p$ (generated using the kernel functions k_p and s_p) and two time covariance matrices $\mathbf{K}_t$ and $\mathbf{\Sigma}_t$ (generated using k_t and s_t). The Kronecker product is an operation where each element in the matrix left of the $\otimes$ sign is multiplied by the full matrix on the right. For two matrices of shape (N_p, N_p) and (N_t, N_t) , this will result in a matrix of shape $(N_p N_t, N_p N_t)$. Since our choice of GP optimisation only requires calculating the separate pixel and time covariance matrices, we avoid memory issues from building the full $(N_p N_t, N_p N_t)$ covariance matrix. In addition, instead of needing to decompose the full covariance matrix $\mathbf{K}$ (e.g. using Cholesky factorisation or eigendecomposition), we only need to separately decompose the pixel and time covariance matrices (specifically using eigendecomposition), significantly reducing the computational costs.

2.4.1. Flux-conserved systematics (FCS)

Examining the pixel light curves reveals significant systematics that do not appear obvious in the aperture extracted light curves (see Figure 1). They are particularly strong in the highest flux pixels at the centre of the PSF. This could be interpreted as the PSF continuously changing in time on timescales of ~30 minutes, perhaps due to thermal distortion of mirrors on JWST (Rigby et al. 2023) or alternatively may be due to charge migration between pixels. In principle, we could examine whether these systematics arise from a global process (e.g. from distortion of the mirrors) or from a local process (e.g. from charge migration) if we have another bright star on the detector to compare the pixel light curves to (similar to the ‘causal pixel model’; Wang et al. 2016). Unfortunately, we could not find an example of this in a MIRI/Imaging time-series dataset. The systematics not being apparent in the aperture extracted light curve suggests that they may be flux-conserving; namely, a bump up in one pixel may be paired with a bump down in a neighbouring pixel. The autocorrelation of neighbouring pixel light curves from our observations hints that this may be true (see Appendix H). Regardless of the underlying cause, we modelled this using anti-correlated GPs between neighbouring pixels and we fit for the amplitude of this effect. This allows the model to

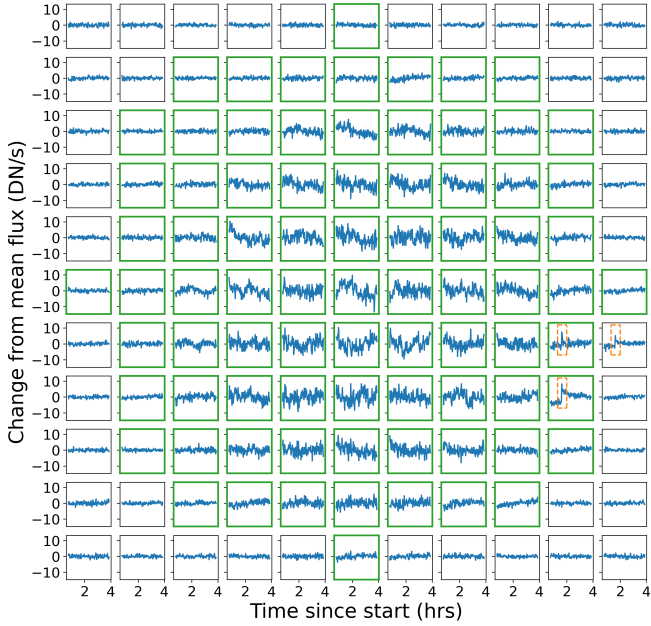


Fig. 1. Pixel light curves centred on the PSF for the second eclipse, excluding the first 45 minutes dominated by settling and binned for clarity. A sharp persistence effect from a cosmic ray is highlighted with orange dashed boxes (see Appendix A for details). Pixels highlighted with green boundaries were fit using the pixel-fitting method except where specified we excluded the pixels containing the highlighted flux jump. Significant systematics are visible in the central light curves which may be flux-conserved between pixels or be independent for each pixel.

fit these bumps in individual pixel light curves while keeping the overall aperture sum flux-conserved. We can then fit for additional non-flux-conserving systematics, which may arise from other causes that we might be able to weight away from.

We used the exponential kernel to describe this correlation as a function of time and our pixel kernel function $k_{p,FCS}$ to describe how these systematics are (anti-)correlated between pixels. To model the flux as conserved we assumed that the amount of flux ‘moving between’ two neighbouring pixels, i and j , is proportional to the geometric mean of the baseline flux falling on each pixel, $\sqrt{f_i f_j}$. We can then model the correlation in flux between both pixels with a covariance matrix as:

$$\mathbf{K}_p = \begin{bmatrix} h_{FCS}^2 f_i f_j & -h_{FCS}^2 f_i f_j \\ -h_{FCS}^2 f_i f_j & h_{FCS}^2 f_i f_j \end{bmatrix} = h_{FCS}^2 f_i f_j \begin{bmatrix} +1 & -1 \\ -1 & +1 \end{bmatrix}. \quad (6)$$

This covariance matrix can describe correlated noise which is perfectly anti-correlated between two pixels, meaning that the sum of this correlated noise will be zero if the pixels are summed within an aperture extraction².

Each pixel actually ends up surrounded by eight neighbours, four pixels which share an edge and four pixels which share a corner. For simplicity, we might assume that flux movement occurs with equal amplitude in all directions including between adjacent and diagonal pixels. We can describe this by summing the covariance terms corresponding to flux movement between each neighbouring pair of pixels together. Given these assumptions, we can describe the pixel kernel function of flux-conserved

systematics for all pixels using:

$$k_{p,FCS}(i, j, i', j') = h_{FCS}^2 f_{(i,j)} \left(f_{\text{neigh};(i,j)} \delta_{ii'} \delta_{jj'} - f_{(i',j')} n_{(i,j),(i',j')} \right), \quad (7)$$

where we define h_{FCS} as the amplitude of flux-conserving systematics, $f_{(i,j)}$ is the average flux on pixel (i, j) , $f_{\text{neigh};(i,j)}$ is defined as the sum of the fluxes of all neighbouring pixels to pixel (i, j) , and we define $n_{(i,j),(i',j')}$ analogously to the Kronecker delta but for neighbouring pixels:

$$n_{(i,j),(i',j')} = \begin{cases} 1, & \text{if } (i, j), (i', j') \text{ are neighbours,} \\ 0, & \text{otherwise.} \end{cases} \quad (8)$$

We note that we chose not to fit all the pixels on the detector, which means we ended up fitting pixels that have neighbours not included in our fit. As the flux may still ‘move between’ these pixels, we still account for the additional variance experienced by these border pixels by including these pixel fluxes in the sum of neighbouring fluxes, $f_{\text{neigh};(i,j)}$. However, there are no corresponding covariance terms included in the covariance matrix. The consequence of this is that the flux moving between pixels being fit and pixels not being fit will not be conserved; although this only affects pixels at the edges of our boundary, which will typically experience low flux. This effect can be reduced by including more pixels in the fit, although this comes at an increased computational cost.

2.4.2. Independent pixel systematics

Each pixel may also have systematics which are independent to neighbouring pixels, such as due to persistence from cosmic ray hits or unmasked bad pixels. For example, the pixel light curves from the second LHS 1140c eclipse appear to show a jump in flux in only a couple of pixels (as highlighted in Figure 1). For these systematics, we assume they can be modelled with the same kernel function in time k_t as for flux-conserving systematics and have a pixel kernel function given by:

$$k_{p,IPS}(i, j, i', j') = h_{IPS,(i,j)}^2 f_{(i,j)}^2 \delta_{ii'} \delta_{jj'}, \quad (9)$$

where we refer to these as independent pixel systematics (IPS) with height scale, h_{IPS} . Since the amplitude of systematics may vary significantly between pixels (e.g. due to persistence effects from a cosmic ray hitting specific pixels), we may choose to fit for the amplitude, $h_{IPS,(i,j)}$, separately for each pixel. Checking which pixels recover large values of $h_{IPS,(i,j)}$ could help identify pixels exhibiting increased systematics.

2.4.3. Common systematics between all pixels

We may also encounter systematics which have the same shape in all pixels and are proportional to the flux on each pixel. For example, if the target star flares, that could produce the same systematic shape in all pixels. These systematics may therefore represent stellar activity or alternatively could result from variations in the throughput of the detector which are correlated in time. We note that these systematics should be similar to fitting the aperture extracted light curve with a GP as they are not conserved and each pixel is treated similarly. Our pixel kernel function for these systematics is described by:

$$k_{p,CS}(i, j, i', j') = h_{CS}^2 f_{(i,j)} f_{(i',j')}. \quad (10)$$

² Note: while this matrix is only positive semi-definite and not invertible, once white noise is included in the overall kernel then the full covariance matrix $\mathbf{K}$ will be invertible.

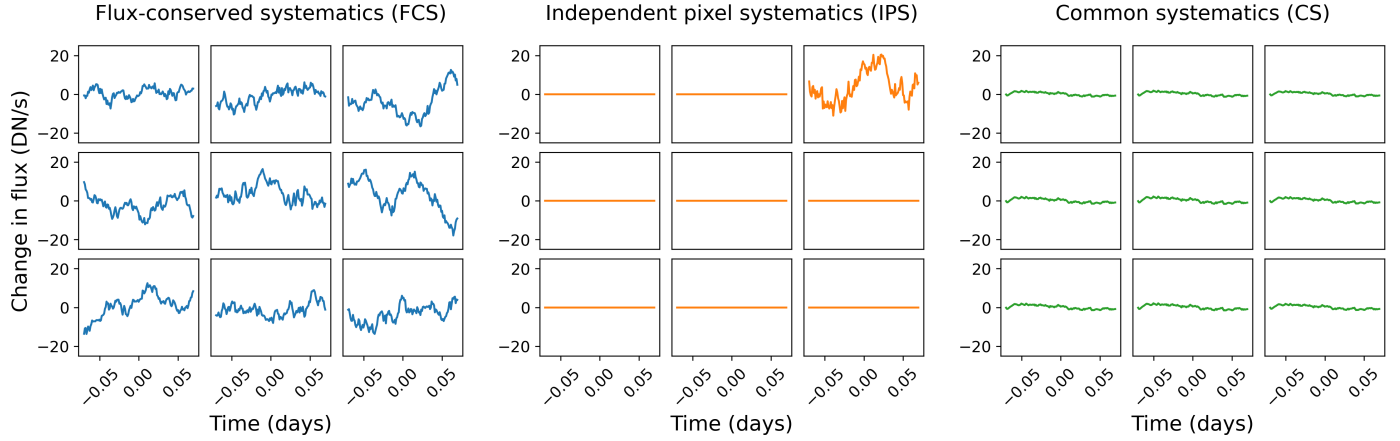


Fig. 2. Three different random draws of time-correlated systematics for a grid of 3×3 pixels. Left grid shows flux-conserved systematics whose sum is zero after aperture extraction. The second grid shows systematics independent to a particular pixel. The third grid shows common systematics with the same shape and with an amplitude proportional to the flux on each pixel. While the amplitude of common systematics may appear small, their effect combines together over multiple pixels to reach a similar amplitude to the independent pixel systematics after aperture extraction.

We refer to these as common systematics (CS) with height scale h_{CS} . The limitations of our GP optimisation require these systematics to also share the same time kernel function as the flux-conserved systematics and independent pixel systematics. A comparison between the three previous systematics described is shown in Figure 2.

2.4.4. Interpixel capacitance (IPC)

Interpixel capacitance is an effect which leads to correlated Poisson noise between neighbouring pixels. It is not to be confused with charge migration which is when charge carriers generated in one detector pixel are read by a neighbouring pixel. Charge diffusion does not result in a correlation in Poisson noise because the charge is exclusively read by the pixel the charge migrated to. Interpixel capacitance is where the potential difference on a given detector pixel is coupled to neighbouring pixels due to a small amount of capacitance between these pixels. This results in an observed correlation in the noise between neighbouring pixels (Moore et al. 2006).

This effect can be modelled by taking the image we expect to receive (including Poisson noise) and using a 3×3 convolution kernel, H (not to be confused with a GP kernel function), which blurs the signal received in each pixel into its neighbouring eight pixels. The kernel will always sum to unity which results in no change to the overall flux received. We can parameterise this convolution kernel using Equation (11) with α_x the coupling between neighbouring pixels to the left or right, α_y for pixels above and below and α_{xy} for diagonal pixels. More complex coupling is also possible but this parameterisation is common in the literature (Moore et al. 2006; Morrison et al. 2023) as follows:

$$H = \begin{bmatrix} \alpha_{xy} & \alpha_y & \alpha_{xy} \\ \alpha_x & (1 - 2\alpha_x - 2\alpha_y - 4\alpha_{xy}) & \alpha_x \\ \alpha_{xy} & \alpha_y & \alpha_{xy} \end{bmatrix}. \quad (11)$$

Morrison et al. (2023) suggested that for the MIRI detector $\alpha_x \approx \alpha_y \approx 3\%$ while α_{xy} is significantly smaller yet still non-zero. The exact extent of IPC is difficult to quantify and is claimed to vary between observations so we rely on fitting for α_x , α_y and α_{xy} for each observation.

We need to account for both the amplitude of white noise for each pixel and the effect of IPC in our combined pixel kernel function $s_{p;wn+IPC}$. However, as IPC can be modelled using linear transformations, it is more natural to describe this by the covariance matrix generated by this kernel function, $S_{p;wn+IPC}$. Suppose each pixel (i, j) observes Poisson noise with variance $\sigma_{(i,j)}^2$ - which in the absence of IPC is independent between pixels. This can be described by a diagonal covariance matrix, $S_{p;wn}$, which contains the variance of each pixel $\sigma_{(i,j)}^2$ along its diagonal (the pixels must be sorted into some arbitrary order). We can simulate IPC by taking a random draw z from $S_{p;wn}$ and then convolving with our IPC kernel H from Equation (11). Note we can construct a matrix T_{IPC} which performs this convolution for each pixel via the matrix multiplication: $z' = T_{IPC}z$, where z' is our random draw of white noise with the effect of IPC simulated. The covariance matrix of z' is then given by:

$$S_{p;wn+IPC} = \mathbb{E}((z' - \mathbb{E}(z'))(z' - \mathbb{E}(z'))^T) = \mathbb{E}(z'z'^T), \quad (12)$$

$$= \mathbb{E}(T_{IPC}zz^TT_{IPC}^T), \quad (13)$$

$$= T_{IPC}S_pT_{IPC}^T, \quad (14)$$

where we have used the fact that $\mathbb{E}(z') = 0$, which follows from $\mathbb{E}(z) = 0$. For notating our overall kernel function, we define the value of our pixel kernel function $s_{p;wn+IPC}$ calculated between points located at (i, j) and (i', j') as $S_{p;wn+IPC;(i,j),(i',j')}$, the element of the covariance matrix corresponding to these two pixels.

2.4.5. Background flux and systematics along rows and columns

There is significant background flux in these observations, due to a variety of sources including thermal emission from JWST itself (Glasse et al. 2015). The exact level of background flux will vary in time, resulting in a correlation in the scatter between all pixels on the detector which can be partially corrected for using a background subtraction. In addition, due to effects such as the reset anomaly and the intricacies of the way pixels are read out by the MIRI instrument, there could be a correlation in the noise between pixels which share the same row or column.

We account for this correlation using our pixel kernel function $s_{p;backgr}$. We assume any residuals from the background

subtraction will be independent in time and so these processes can share the same time kernel function $s_t = \delta_{t'}$ as for white noise.

We can describe the pixel covariance between pixels located at (i, j) and (i', j') using:

$$s_{p;backgr}(i, j, i', j') = h_{back}^2 + h_{row}^2 \delta_{ii'} + h_{col}^2 \delta_{jj'}, \quad (15)$$

where h_{back} is the amplitude of overall background scatter, h_{row} is the amplitude of correlated scatter in each row and similarly h_{col} for each column. By fitting for the amplitude of h_{back} , h_{row} and h_{col} , we can constrain how much remaining background scatter is left by the background subtraction. We could potentially use this to compare different background subtraction procedures. We note that this does not give us the ability to detect whether the background has been over-subtracted or under-subtracted. Instead, it simply allows us to see whether there is a lot of correlated scatter between rows, columns, or all pixels.

2.4.6. Combined kernel

Our full kernel function for the pixel-fitting method is therefore given as:

$$k(i, j, t, i', j', t') = \left[h_{FCS}^2 f_{(i,j)} \left(f_{neigh;(i,j)} \delta_{ii'} \delta_{jj'} - f_{(i',j')} n_{(i,j),(i',j')} \right) + h_{IPS;(i,j)}^2 f_{(i,j)}^2 \delta_{ii'} \delta_{jj'} + h_{CS}^2 f_{(i,j)} f_{(i',j')} \right] \exp\left(-\frac{|t-t'|}{t_t}\right) + \left[s_{p;wn+IPC;(i,j),(i',j')} + h_{back}^2 + h_{row}^2 \delta_{ii'} + h_{col}^2 \delta_{jj'} \right] \delta_{t'}. \quad (16)$$

3. Testing on simulated data

To better examine how the pixel-fitting method compares to aperture photometry, we generated synthetic observations for which the true eclipse depth was known. This then allows us to examine how precisely and robustly each method recovers the eclipse depth in different circumstances. We examined how well each method performed when there were only flux-conserved systematics present, when systematics were added in only a single pixel light curve, and when there were systematics present across all pixels with the same shape (e.g. from a flare).

3.1. Simulating the target signal

The simulations were based on the first eclipse of LHS 1140c, which consisted of 1262 integrations and covered ~ 233 minutes in total (for more details see Section 4), although we assume the first 30 minutes have been clipped. To minimise the computational cost, we simulated eight times fewer integrations. This permitted each pixel-fitting analysis to run in ~ 40 minutes on a CPU cluster node, and we could run eight of these analyses in parallel. As we were simulating observations with eight times the integration time, we reduced the amplitude of Poisson noise by a factor of $\sqrt{8}$ to maintain similar constraints on the eclipse depth.

The PSF model was taken from the median frame of the first LHS 1140c eclipse after clipping the first 30 minutes. This PSF model was then interpolated in x and y position for each integration with a change in position distributed as $\mathcal{N}(0, 0.005)$ pixels around the central location. No detector settling or linear slope in flux was introduced meaning that the baseline flux in each pixel was constant.

The eclipse model was generated using the package JAXO-PLANET (Hattori et al. 2024), which implements the equations

of Mandel & Agol (2002) in JAX (Bradbury et al. 2018). The eclipse model was fixed for all simulations with the mean values of P , a/R_* , R_p/R_* , b from literature values of LHS 1140c (as listed in Table D.1). The central eclipse time was fixed to the midpoint of the time series and the eccentricity was fixed to zero. For all fits of the simulated data, the eclipse depth was the only eclipse model parameter being fit for with every other eclipse parameter held fixed to the true values.

3.2. Simulating white noise and systematics

White noise was generated with the amplitude for each pixel determined by taking the standard deviation of each pixel time series in the first LHS 1140c eclipse. This noise was then convolved with the IPC convolution kernel from Equation (11) using values of $\alpha_x = 3\%$, $\alpha_y = 2\%$ and $\alpha_{xy} = 1\%$, before then being added to the simulated signal. Background noise was also introduced to simulate residual background scatter not removed by a background subtraction. Each integration had a constant value added to all pixels, drawn from $\mathcal{N}(0, 0.1)$ DN/s, which (given the flux in the target signal) works out to adding scatter with amplitude of ~ 135 ppm for a 5px radius circular aperture. We also added scatter which was constant along each row and column but independent between each row and column. This was added by drawing from $\mathcal{N}(0, 0.4)$ DN/s for the rows and $\mathcal{N}(0, 0.3)$ DN/s for the columns. Our analysis of the real observations did not constrain the presence of these row or column systematics but it was a possibility we wanted to include which both aperture photometry and the pixel-fitting method should be able to account for.

We simulated the flux-conserved systematics by taking a random draw from a GP with an exponential kernel and a length scale of 35 minutes and adding it to one pixel and subtracting it from a neighbouring pixel. This was performed for each pair of neighbouring pixels (i.e. 8 times per pixel) and the amplitude of the draw was scaled to be the geometric mean of the flux on each of the pair of pixels multiplied by the constant $h_{FCS} = 990$ ppm. By adding and subtracting the same signal, it means that these systematics are flux-conserving with the only non-flux-conserving effects due to the finite area of the aperture (i.e. there is ‘flux movement’ across the fixed aperture boundary), although the amplitude of this effect was small. For a pixel surrounded by eight pixels with the same flux values, the amplitude of systematics in this pixel would work out to $\sqrt{8}h_{FCS} = 2800$ ppm. We note that h_{FCS} was meant to be set to 700 ppm – consistent with the values recovered in the observations – but the addition and subtraction was accidentally performed twice for each pair of pixels, effectively increasing the amplitude of this effect by $\sqrt{2}$. We generated 200 of these simulations and we refer to them as the flux-conserved systematics (FCS) scenario. The other two sets of simulations included additional systematics which were not flux-conserved. The second set of simulations – called the independent pixel systematics (IPS) scenario – took each simulation from the FCS scenario and injected a $h_{IPS} = 5000$ ppm height scale systematic into a single pixel time series. The pixel was randomly chosen for each simulation to be one of the nine brightest pixels at the centre of the PSF. As these pixels tend to make up about 3–4.5% of total flux, this results in a ~ 150 – 225 ppm systematic in the aperture extracted light curve. While aperture extraction with this pixel masked out may be a good approach, the flux-conserved systematics make it difficult to distinguish these additional systematics from flux-conserved systematics. This can be seen in Figure 3, which features a plot

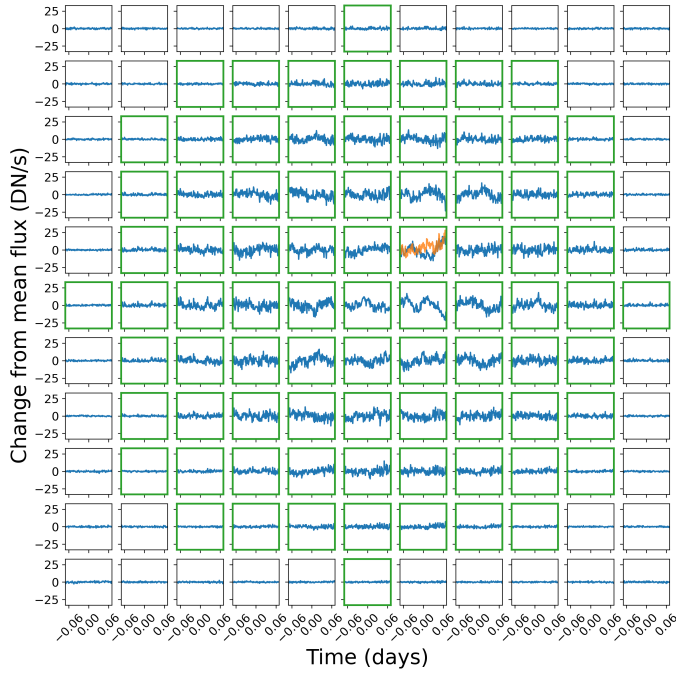


Fig. 3. Pixel light curves for the first simulation in the flux-conserved systematics (FCS) and independent pixel systematics (IPS) scenarios. The only difference is the pixel light curve in orange which had independent systematics added to it. The pixels included within the pixel-fits are highlighted with green boundaries.

of the pixel light curves centred on the PSF for the first simulated dataset in the FCS scenario and the IPS scenario.

The third set of simulations added a common systematic shape to all pixels at an amplitude equal to $h_{CS} = 200$ ppm multiplied by the flux on each pixel. Here, this is referred to as the common systematics (CS) scenario. This could represent real astrophysical systematics such as a flare or instrumental systematics which affect the throughput of the observation. The height scale was chosen as it resulted in a similar amplitude of systematics in an aperture extracted light curve as the IPS scenario, but with the systematics spread out over all pixels instead of originating entirely from a single pixel.

For each scenario, the synthetic data were aperture extracted with a radius of 5 pixels centred on the best-fit centroid for each integration – as performed for the real observations. This radius minimised scatter in the aperture extracted light curves compared to radii of 4.5 or 5.5 pixels, matching the real observations. In Figure 4, we give an example of a simulated dataset from the FCS and IPS scenarios after aperture extraction. The pixel-fitting method only included pixels with centres within 5 pixels from the PSF centre, resulting in 81 pixels being included, which is similar to the $78.5\pi r^2$ area for a circular aperture with a radius of 5 pixels.

3.3. Different fitting approaches tested

We used MCMC to recover the eclipse depth with each method, modelling the posterior as Gaussian and using the mean and standard deviation of the chains. For each method, we ran four chains of No U-Turn Sampling (NUTS; Hoffman & Gelman 2014) as implemented in PYMC (Salvatier et al. 2016) – with each chain having 1000 tuning steps and 1000 draws. We also used slice sampling (Neal 2003) in combination with NUTS, by performing blocked Gibbs sampling (Jensen et al. 1995) on

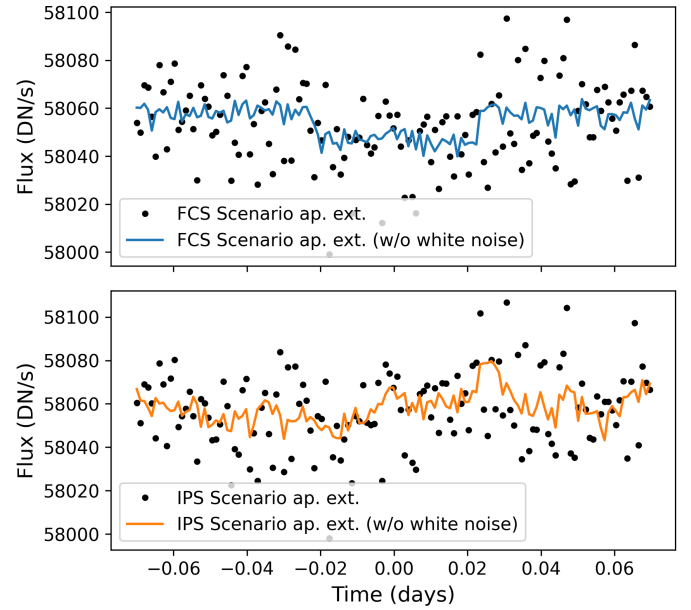


Fig. 4. Aperture extracted light curves from simulated datasets shown in Figure 3. Aperture extractions were also performed without including white noise to visualise systematics more clearly (some jitter remains from imperfections in centroiding). The systematics injected into one pixel has dramatically affected the aperture extracted light curve.

some parameters close to their prior bounds (e.g. h_{back} , h_{row} , h_{col} , l_t , h_{CS} , h_{IPS}). This was performed because it was found that running NUTS alone could sometimes result in chains getting stuck on these parameters. The Gelman-Rubin statistic (Gelman & Rubin 1992) was used to measure convergence of the chains. A slightly higher than normal threshold of $\hat{r} \leq 1.02$ was used to determine convergence for each simulation in order to reduce computational costs arising from the large number of MCMCs performed.

For each set of simulations, we tested aperture extraction with and without a GP. These MCMCs fit for the eclipse depth, d , the flux out-of-(secondary) transit, F_{oot} , the white noise amplitude, σ , and in the case of the GP fits also the height scale, h , and length scale, l_t , of the GP. We also tested using a GP with l_t fixed to the true simulated value of 35 minutes, as this may be a fairer comparison to the pixel-fitting method – which obtains extra information about the time length scale as it fits for flux-conserved systematics with the same time length scale.

For all pixel-fitting approaches, we fit for the eclipse depth, d , baseline flux for each pixel, $f_{(i,j)}$ (for 81 pixels), white noise amplitude for each pixel $\sigma_{(i,j)}$, change in centroid position at each integration Δx and Δy (for 137 integrations), IPC parameters α_x , α_y , and α_{xy} , background scatter parameters, h_{back} , h_{row} and h_{col} , height scale and time length scale of flux-conserving systematics h_{FCS} and l_t . This describes the ‘only flux-conserving systematics’ approach (abbreviated as ‘only FCS’) and includes 445 free parameters which were all fit for. We also included a height scale for systematics with a common shape in all pixels h_{CS} for the ‘common systematics’ approach and height scales for independent systematics in each pixel $h_{IPS,i}$ for the ‘independent systematics’ approach. Finally, as this independent systematics approach turned out to give quite conservative predictions, we also reran each of these approaches where only the pixels recovered to have large height scales (specifically where the 3^d percentile value was greater than a threshold of $10^{-4.8} \approx 16$ ppm) were fit with independent height scales and the remaining pixels

all shared a joint height scale $h_{\text{IPS};\text{shared}}$. We note these pixel systematics were still treated as independent but with a shared amplitude (i.e. they could still have completely different shapes). This is listed in tables as the ‘shared independent systematics’ approach. We also included methods which could account for both common and independent pixel systematics as these were robust to a range of systematics.

3.4. Comparing the performance of each method

We measured the precision of each method using the root mean square error (RMSE) of the mean eclipse depth for each simulation. The smaller the RMSE, the more precise the eclipse depth measurement is. If the method is robust, then the average standard deviation of the eclipse depth measurement, $\bar{\sigma}_d$, should approximately match the RMSE. If the standard deviation is too big then our results are overly conservative and if it is too small then the results are unreliable. This can be summarised with the reduced chi-squared statistic χ_r^2 , which should equal unity when the uncertainties are correctly estimated, less than one for uncertainties which are too large/conservative and greater than one for uncertainties which are too small/unreliable³. For each scenario (consisting of 200 simulations), it is calculated as follows:

$$\chi_r^2 = \frac{1}{200} \sum_{i=1}^{200} \left(\frac{d_i - d_{\text{true}}}{\sigma_i} \right)^2. \quad (17)$$

Here, $d_{\text{true}} = 200$ ppm is the true simulated eclipse depth and d_i , σ_i are the mean and standard deviation of the recovered eclipse depth for the i^{th} simulation.

Finally, we examined whether any method was biased towards measuring eclipse depths which were too large or too small by calculating whether the mean eclipse depth across all 200 simulations was consistent with $d_{\text{true}} = 200$ ppm. We found that all methods tested were within 2.5σ of the expected mean values across each of the scenarios. Given that this metric has little impact on the comparison between methods, and is typically very similar for methods within the same scenario, we chose to exclude it from each table of results.

3.5. Flux-conserved systematics (FCS) scenario results

This scenario only contained flux-conserved systematics, therefore, aperture extraction should result in very clean light curves and a GP should not be required to robustly analyse these observations. However, for real observations we can never be sure that the light curves are completely free from systematics and so we are interested in how well each method performs in this scenario. The methods that marginalise over systematics will naturally tend to give more conservative constraints on the eclipse depths (even when systematics are not present); this is true both for aperture extraction with a GP and for various pixel-fitting approaches. However, this will be balanced out by their more reliable performance when systematics are present.

The results in Table 1 demonstrate this very clearly. The most precise methods (i.e. smallest RMSE) are aperture extraction without a GP and the pixel-fitting approach which only accounts for flux-conserved systematics. They both exhibit χ_r^2 values close to unity across these simulations, showing their uncertainty estimates are reliable. Aperture extraction with a GP has a slightly worse precision and larger uncertainties. The significantly smaller value of $\chi_r^2 = 0.74$ shows that this method

Table 1. Results of the 200 simulations in the Flux-Conserved Systematics (FCS) scenario.

Method	RMSE	$\bar{\sigma}_d$	χ_r^2
Ap. Ext. w/o GP	52	53	0.98
Px-fit (only FCS)	51	52	0.94
Ap. Ext. w/ GP	54	64	0.74
Ap. Ext. w/ GP (fixed l_i)	55	66	0.73
Px-fit (common systematics)	53	64	0.7
Px-fit (indep. systematics)	63	81	0.59
Px-fit (shared indep. sys.)	57	63	0.82
Px-fit (common+indep. sys.)	68	88	0.57
Px-fit (common+shared indep. sys.)	64	73	0.71

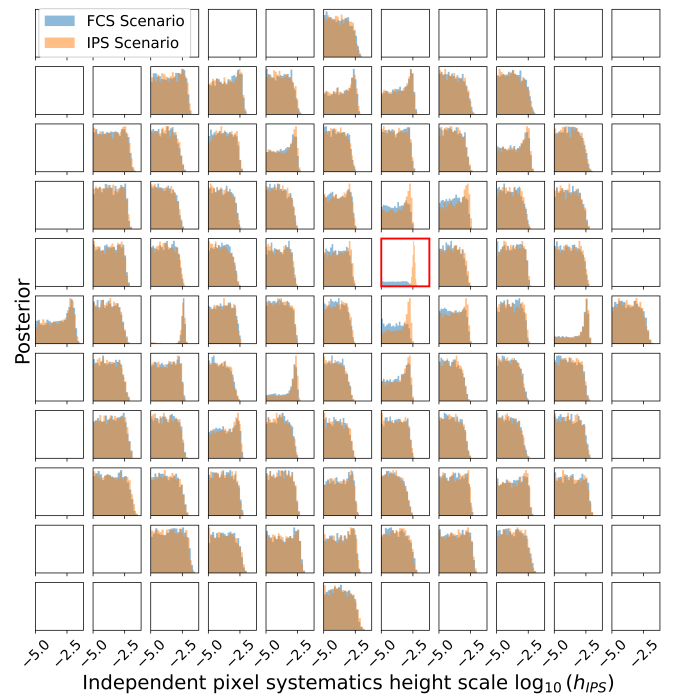


Fig. 5. Constraints on the Independent Pixel Systematics (IPS) height scale for each pixel for the first simulation in the FCS and IPS scenarios (shown in Figure 3). Highlighted in red is the one pixel which had independent pixel systematics injected into it for the IPS scenario. This pixel has been correctly identified by the pixel-fitting method and would be weakly weighted in the eclipse fit.

is quite conservative. Pixel-fitting only accounting for common systematics performs similarly.

However, we see that the other pixel-fitting approaches which can account for independent systematics in each pixel have worse accuracy and even larger uncertainty estimates. This is similar to aperture extraction with a GP and in general we do expect a more flexible systematics model to result in larger errorbars as it marginalises over a wider space of models (Gibson 2014). The pixel-fitting method is particularly flexible as it can fit for potential systematics in every pixel. Despite the vast majority of these pixel systematics height scales being consistent with the minimum prior bound, they still marginalise over a range of amplitudes which are consistent with the data. See Figure 5 for the posterior of each independent pixel systematics height scale, h_{IPS} , for the first simulation in the FCS and IPS scenarios.

³ This assumes the posteriors for the eclipse depths are Gaussian.

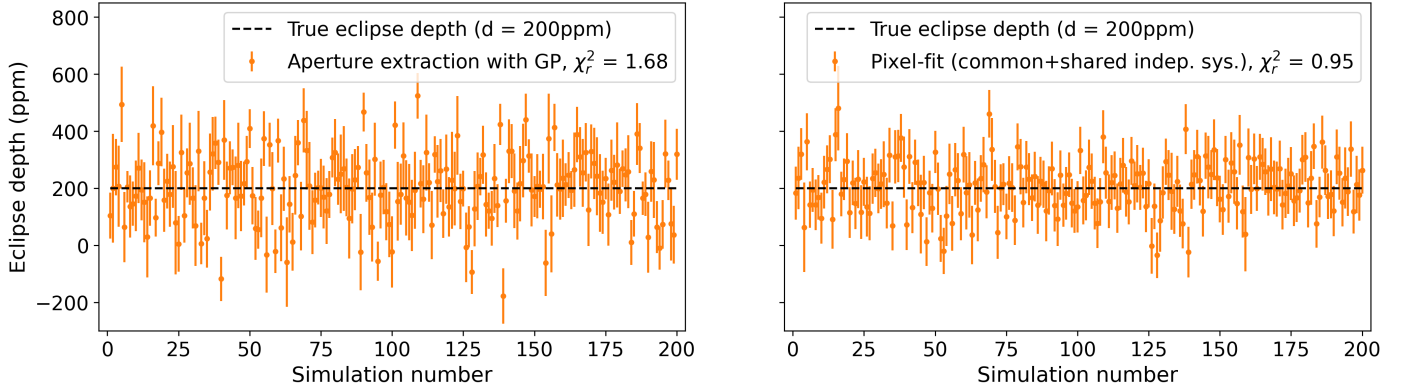


Fig. 6. Eclipse depth constraint for all simulations in the IPS scenario, which each had a single pixel contaminated with time-correlated systematics. The left plot shows the results from using aperture extraction with a GP. The right plot shows pixel-fitting accounting for both independent pixel systematics (with a shared height scale for cleaner pixels) and common systematics. Pixel-fitting outperforms aperture extraction here in both precision (smaller RMSE) and reliability (χ_r^2 closer to unity) as it can weight away from the single pixel contaminated by systematics.

Table 2. Results of the 200 simulations in the independent pixel systematics (IPS) scenario.

Method	RMSE	$\bar{\sigma}_d$	χ_r^2
Ap. Ext. w/o GP	143	59	6.07
Ap. Ext. w/ GP	122	107	1.68
Ap. Ext. w/ GP (fixed l_t)	119	112	1.46
Px-fit (indep. systematics)	87	93	0.89
Px-fit (shared indep. sys.)	84	81	1.12
Px-fit (common+indep. sys.)	92	101	0.82
Px-fit (common+shared indep. sys.)	89	92	0.95

We see that sharing the amplitude of the smallest independent pixel systematics height scales does reduce how conservative the method is, with smaller average uncertainties and a χ_r^2 value closer to one. Accounting for common systematics as well as independent pixel systematics offers the most conservative result; however this method is the most flexible and can be robustly applied to all three sets of simulations generated.

3.6. Independent pixel systematics (IPS) scenario results

The IPS scenario simulations were the exact same as the FCS scenario but with a single GP draw added to one of the brightest central nine pixels for each simulation. In this case, the pixel-fitting method has the potential to outperform aperture extraction as it can weight away from the pixel with systematics injected into it. However, it must be able to constrain the presence of independent systematics on top of the flux-conserved systematics. Since the method can fit for independent pixel systematics in all of the pixels, the results can still be a bit conservative given that the systematics are only present in a single pixel.

We see from Table 2 that all of the pixel-fitting approaches outperform aperture extraction with a GP on precision (RMSE), have tighter constraints (smaller $\bar{\sigma}_d$) and have more reliable constraints (χ_r^2 closer to unity). This difference is also visualised in Figure 6, which shows the eclipse depth constraint for all 200 simulations in the IPS scenario for both aperture extraction with a GP and pixel-fitting (accounting for common systematics and shared independent systematics). Aperture extraction without a GP performs particularly poorly, with the lowest accuracy and

with uncertainties which are far too small ($\chi_r^2 = 6.07$) because it cannot account for the systematics injected into the data. Aperture extraction with a GP does a reasonable job of accounting for the systematics but with quite low precision compared to pixel-fitting, as it has no way to weight away from the pixel with injected systematics. The chi-squared value of $\chi_r^2 = 1.68$ is also quite high, with a slightly better value of $\chi_r^2 = 1.46$ when the length scale in time is fixed to the true simulated value.

For aperture extraction with a GP, the height scale of systematics h is not constrained to be greater than the minimum prior bound in many of these simulations. This means that varying the location of this prior bound will increase or decrease the weighting of the area of the posterior where systematics are negligible. For example, we get $\chi_r^2 = 1.14$ if we increase the minimum bound to $h > 10^{-4}$ and if we instead lower the minimum bound to $h > 10^{-6}$, we get $\chi_r^2 = 1.88$. Increasing the minimum prior bound on the height scale can be interpreted as a more conservative analysis where the presence of systematics is more likely, and this would help lower the χ_r^2 here. However, setting $h > 10^{-4}$ also lowers the χ_r^2 value for the first set of simulations to just $\chi_r^2 = 0.46$, which was already overly conservative at $h > 10^{-5}$. Overall, for our normal prior bound of $h > 10^{-5}$, the χ_r^2 from combining the FCS and IPS scenarios is $\chi_r^2 = 1.21$ when fitting for l_t and $\chi_r^2 = 1.10$ when fixing it; thus, the method performs well over an even mix of clean and systematics-contaminated datasets.

The pixel-fitting methods which account for common systematics on top of accounting for independent pixel systematics have slightly lower precision and larger uncertainties. However, they are accounting for additional systematics that could potentially exist in real data – such as those from variations in throughput and detector settling – so this is a more conservative approach that might make sense when we do not know the origin of systematics in the data.

3.7. Common systematics (CS) scenario results

Finally, the simulations in the common systematics scenario differ from the IPS scenario in that the systematics were injected into all pixels with a common shape instead of into a single pixel. The amplitude of these systematics was scaled by the flux in each pixel, so this could represent systematics inherent from the signal, such as stellar granulation or flaring. It could also arise from factors affecting the throughput which vary over time. The

Table 3. Results of the 200 simulations in the common systematics (CS) scenario.

Method	RMSE	$\bar{\sigma}_d$	χ_r^2
Ap. Ext. w/o GP	155	60	6.86
Ap. Ext. w/ GP	133	116	1.73
Ap. Ext. w/ GP (fixed l_r)	133	120	1.54
Px-fit (common systematics)	127	118	1.54
Px-fit (common+indep. sys.)	128	121	1.24
Px-fit (common+shared indep. sys.)	128	120	1.37

amplitude of this was such that aperture extraction with a GP performs similarly to the IPS scenario but with a different systematics origin. In the IPS scenario, the pixel-fitting approaches could outperform aperture extraction by weighting towards the cleaner pixels, however, this will not work when all the pixels suffer from a similar proportion of systematics. Pixel-fitting should therefore perform similarly to aperture extraction with a GP.

We can see from Table 3 that all methods have low precision and most have large uncertainties. Aperture extraction without a GP has a similarly poor performance as it had for the IPS scenario, with uncertainties which are too small as it cannot robustly account for the systematics present. Any of the pixel-fitting methods which can account for common systematics approximately match the performance of aperture extraction with a GP. Similar to the IPS scenario, aperture extraction with a GP and also the pixel-fitting methods have quite high χ_r^2 values. This could also be explained by the trade-off between performance on clean data and systematics-contaminated data when the systematics are not fully constrained to be present. The slightly more robust χ_r^2 values when independent pixel systematics are also accounted for may arise from the method just being generally more conservative across all the simulations rather than it more closely modelling the systematics.

3.8. Summary of the simulations

The results demonstrate that pixel-fitting can outperform aperture extraction when systematics are large in a particular pixel, but the method is overly conservative when only flux-conserving systematics are present. It also matches the performance of aperture extraction with a GP when systematics have a common shape across all pixels. See Figure 7 for a summary of the results in Tables 1, 2, and 3 for aperture extraction with and without a GP as well as for the general pixel-fitting approach of accounting for common systematics across all pixels and systematics in individual pixels. We can see that compared with aperture extraction with a GP, pixel-fitting performs very similarly when systematics are common between all pixels but much better when there are systematics present in specific pixels; for instance, after a cosmic ray hit (as shown in Figure 1).

4. Observations

The three eclipses of LHS 1140c were observed on 27th November 2023, 7th July 2024, and 19th July 2024. Each eclipse used the SUB256 subarray, read 36 groups for each integration and had 1262 integrations. The integrations had an 11.1s cadence, with each observation lasting ~233 minutes. This followed a strategy of leaving 60 minutes of flexible start time plus 30 minutes

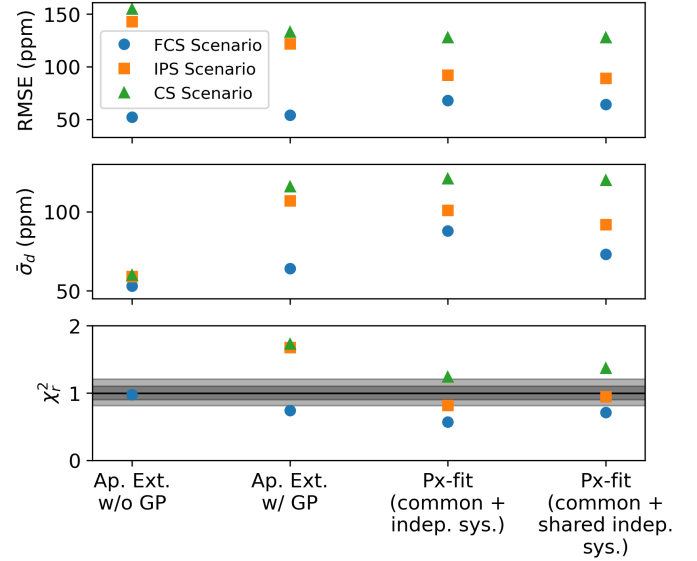


Fig. 7. Results from Tables 1, 2, and 3 comparing the RMSE, mean uncertainty in eclipse depth, and χ_r^2 of recovered eclipse depths for various methods. Aperture extraction with a GP is shown to perform similarly for the IPS and CS scenarios as the aperture extracted light curves contain similar amplitude systematics. Pixel-fitting achieves a better performance for the IPS scenario as it can weight away from the contaminated pixel.

of settling time followed by observing for two eclipse durations, with one eclipse duration (~67 minutes) before and after expected mid-eclipse for a circular orbit. There was also an additional 9 minutes of padding to help catch slightly more eccentric orbits.

4.1. Primary data reduction

Our reduction began with a download of the uncalibrated ‘uncal’ FITS files for our three eclipses from the Mikulski Archive for Space Telescopes (MAST), which had already been pre-processed by version 2024_1a of the Science Data Processing (SDP) system. We ran the JWST calibration pipeline version 1.15.1 with the default settings for steps 1 and 2, except we changed the jump detection threshold to 7σ , excluded both the first and last groups from each integration when reading up-the-ramp (the default is to only exclude the last group) and skipped the PHOTOM step in stage 2 to keep the flux units in data numbers per second (DN/s). We excluded the first and last groups to avoid a known issue where these groups can include difficult to correct reset effects (Morrison et al. 2023). The jump detection threshold was decided upon by calculating the average amplitude of white noise recovered across all three eclipse fits when only white noise and an exponential ramp + linear model was fit. It was found that 7σ minimised this out of integer thresholds from 4–10 σ .

One issue we encountered was that running the pipeline separately on each uncal data segment resulted in offsets in the flux time series between segments, despite these segments not representing any actual change in the observation but instead are simply a way of reducing the total filesize of any given file. To remedy this, for all our reductions we first combined the uncal segments together into a single FITS file and then ran the JWST pipeline on this combined uncal file, which avoided these offsets.

Once we had our stage 1 and 2 calibrated ‘calints’ files, we applied a median clipping filter to remove any additional outliers.

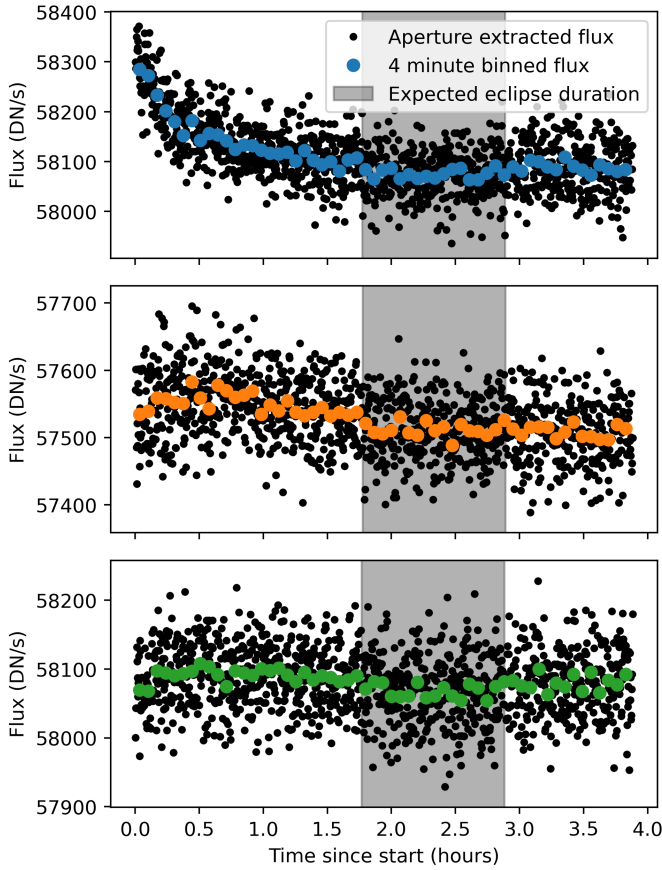


Fig. 8. Three aperture extracted eclipse light curves from the primary reduction. Observations are binned into four minute bins for clarity. Expected eclipse location for a circular orbit is shaded in gray. There are differences in initial detector settling in the first half hour, where the first eclipse shows a strong decrease in flux, while the other two eclipses increase in flux.

A background subtraction was then performed on each integration by taking the median flux values from an annulus region centred around the target PSF with lower and upper radii of 50 to 80 pixels from the annulus centre. This annulus region was chosen as there still appeared to be a gradient in flux as far out as about 50 pixels from the centre while the flux appears to remain constant from 50 to 80 pixels.

4.1.1. Primary reduction: Aperture extraction analysis

Aperture extraction was performed with a circular aperture using the package PHOTUTILS (Bradley et al. 2024). For each integration, the centroid position was fit for using a best-fit to a 2D Gaussian. A range of radii from 3.0px to 7.0px were used in 0.5px steps to extract the light curves and it was found a 5px radius minimised the amplitude of scatter in the light curves. The 5px aperture extracted flux for each light curve is shown in Figure 8 with only the first integration of each eclipse removed as these show significantly lower flux (consistent with the ‘first exposure effect’; Morrison et al. 2023).

We performed a joint fit of the three eclipses using priors on the central (primary) transit time, T_0 ; system scale, a/R_* ; impact parameter, b ; and radius ratio, R_p/R_* , from the NIRISS/SOSS transit observation of LHS 1140c, as analysed in Cadieux et al. (2024b). As the NIRISS/SOSS analysis assumed a circular orbit

with fixed period, we used priors for the period, P , and eccentricity parameters, $\sqrt{e} \cos \omega$ and $\sqrt{e} \sin \omega$, from the joint fit of radial velocity and previous transit observations (performed before the JWST transit observation) in Cadieux et al. (2024a). Including the JWST transit observation from Cadieux et al. (2024b) within a joint fit of the RV data could help to tighten these constraints further but is beyond the scope of this paper. Table D.1 shows a summary of these priors. As the time of primary transit is tightly constrained with this prior, the central eclipse time is largely determined by the eccentricity parameters (see Appendix C).

To deal with detector settling, we clipped between 30 and 60 minutes (162–324 integrations) from the start of each observation, with all joint-fits clipping 45 minutes (243 integrations). We experimented with including an exponential ramp in addition to a linear slope, modelling the baseline flux as:

$$f_{\text{base}}(t) = F_{\text{oot}} + T_{\text{grad}}(t - t_{\text{mid}}) + h_{\text{ramp}} \exp\left(-\frac{t - t_{\text{init}}}{l_{\text{ramp}}}\right), \quad (18)$$

where t_{init} is the time stamp of the first integration being fit and t_{mid} the midpoint of the observation, F_{oot} is the baseline flux out-of-(secondary) transit, T_{grad} is a linear slope in baseline flux, h_{ramp} is the amplitude of the settling ramp and l_{ramp} is the lengthscale of the ramp. When fitting for a combined exponential ramp and linear slope, there were issues with degeneracies between long exponential ramps with large l_{ramp} values and the linear slope T_{grad} ; thus, in practice, we also subtracted the tangent line to the exponential ramp at the last integration, which helped reduce this degeneracy when l_{ramp} was large.

When fitting the aperture extracted light curve with a GP, we chose an exponential kernel plus white noise:

$$\mathbf{K}_{ij} = h^2 \exp\left(-\frac{|t_i - t_j|}{l_t}\right) + \sigma^2 \delta_{ij}, \quad (19)$$

where h is the height scale, l_t the length scale of correlated noise, and σ is the amplitude of white noise.

4.1.2. Primary reduction: Pixel-fitting analysis

For the pixel-fitting method, the aperture extraction was not needed and the reduced data were used directly. The same eclipse model and priors were used as for the aperture extraction method.

Figure 9 shows the choice of pixels within 5 px of the PSF centre for the pixel-fitting method overlaid over the median frame of the PSF. It is also shown how it compares to the 5px circular aperture used for aperture extraction. We note that unlike aperture extraction, pixel-fitting can weight away from lower S/N pixels similar to optimal extraction, so we can choose to include more pixels to try improve our constraints. However, including more pixels becomes more computationally expensive particularly as there are many more parameters to recover, and the change in constraints was found to be quite negligible when pixels 8.5px away from the centre were included (covered in Appendix G).

The computational cost of joint-fitting the three eclipse light curves using pixel-fitting was quite high and took approximately 20 hours for each analysis when running on one NVIDIA A100 GPU. This was largely due to the relatively large number (1019) of integrations fit for each eclipse which required the decomposition of a large time covariance matrix $\mathbf{K}_t$ in each MCMC step. In addition, there was a huge number of free parameters to be fit, especially because each integration in all three

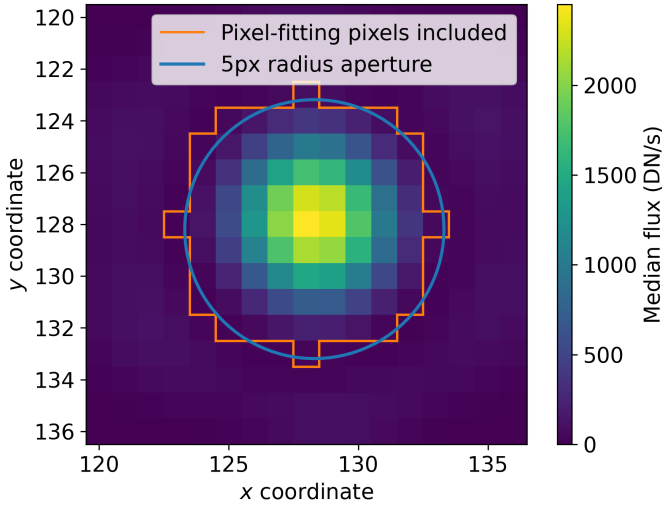


Fig. 9. Median frame of the first observation centred on the target PSF. Aperture extraction was performed with a 5px circular aperture in blue, centred for each integration. Pixels used for the pixel-fitting are enclosed in orange.

datasets had two parameters which marginalised over the uncertainty in the change in centroid position (resulting in over 6 000 free parameters). As these parameters were all well constrained and had Gaussian-like posteriors, they were efficiently sampled using NUTS and all had an $\hat{r} \leq 1.01$. The computational cost could have been reduced substantially by binning in time, parallelising each MCMC chain over more GPUs or running on each dataset separately. Running the method on a single dataset and parallelising over 4 GPUs on a GPU cluster took approximately one hour, even without binning in time. This means the method is better suited to examining pixel systematics in individual eclipses but poorly suited to joint-fitting large numbers of eclipses at high time-resolution.

4.2. Secondary data reduction

We performed a secondary reduction using a completely independent pipeline, which has been previously used in August et al. (2025) and will be described in more detail in the upcoming paper (Gressier et al., in prep.). This pipeline employs the `transitspectroscopy` (Espinoza 2022) wrapper for the initial stages of the `jwst` pipeline, skipping the reference pixel correction and incorporating a custom jump correction. The identified 10 Hz heater noise is corrected. Once the groups are converted to slopes, we proceed to Stage 2. We compute a median frame, which is then used for the following step. To estimate the background shape, we mask pixels within a box centered around the source. The masked pixels are interpolated using a linear interpolation with neighboring pixels. This frame is then smoothed using a Gaussian filter with a standard deviation of 7 pixels. The resulting median background image is subtracted from all frames. Pixel light curves are extracted, and outliers are identified using a median filter. These outliers are replaced with the median value of each pixel. Once we obtain a background corrected, cleaned image, we proceed to flux extraction.

We fit a 2D Gaussian to each frame to determine the centroid position and then apply a circular mask with a 10 pixel aperture to isolate the flux. Flux extraction is performed using two techniques: optimal extraction and classic aperture photometry with a 5 pixel radius aperture. For optimal extraction, the PSF is

Table 4. Eclipse depths from the joint-fits of all three eclipses from the primary reduction analysis.

Method	Depth (ppm)
Ap. Ext. w/o GP	272 ± 40
Px-fit (only FCS)	262 ± 40
Ap. Ext. w/ GP	273 ± 43
Px-fit (common systematics)	257 ± 45
Px-fit (common+indep. sys.)	253 ± 55
Px-fit (common+shared indep. sys.)	253 ± 49

modelled by fitting a 2D Gaussian to the background-corrected cleaned median frame. The normalized PSF (scaled to unity) is used as weights to extract the flux within the 10 pixel circular mask. The uncertainties in the flux are estimated through error propagation. We apply the same reduction process to all three observations.

4.2.1. Secondary reduction analysis

We use `juliet` (Espinoza et al. 2019) to fit the extracted light curves from this secondary reduction. The three observations are jointly fitted using five detrending methods: a linear slope in time (L), a single exponential (E), a linear slope with a single exponential (L+E), a linear slope with a GP (L+GP), and a linear slope with a single exponential and a GP (L+E+GP). We use the Matérn 3/2 kernel, which is described by:

$$k_{\text{sec.}}(t, t') = \xi^2 \left(1 + \sqrt{3} \frac{|t - t'|}{\rho} \right) \exp \left(-\sqrt{3} \frac{|t - t'|}{\rho} \right) + \sigma^2 \delta_{tt'}, \quad (20)$$

with ξ the height scale of systematics, ρ the length scale in time and σ the amplitude of white noise.

We trim the first 200 integrations of each observation, corresponding to 37 minutes. Each detrending model is applied independently to each observation, while the eclipse depth and orbital parameters are jointly fitted. Table D.2 presents the priors and distributions of the planetary parameters included in the joint fit. We provide details on the parameters of the detrending methods and their corresponding priors.

4.3. Eclipse depth results

4.3.1. Results from primary reduction analysis

To joint-fit the three eclipse datasets only the eclipse model parameters (T_0 , P , a/R_* , R_p/R_* , b , f_p/f_* , $\sqrt{e} \cos \omega$, $\sqrt{e} \sin \omega$) were shared between datasets for all methods, with all systematics and PSF parameters being separate for each dataset. For all joint-fits we cut the first 45 minutes to remove the strongest settling effects and fit the remaining settling with the exponential ramp model.

The results from joint-fitting the light curves are shown in Table 4. While the results are all highly consistent, the pixel-fitting method tends to be a bit more conservative with larger uncertainties - which is similar to the simulations where there were only flux-conserving systematics present (i.e. FCS scenario; see Section 3.5). We note that pixel-fitting while only accounting for flux-conserving systematics performs almost identically to aperture extraction without a GP, so any differences in the size of uncertainties are likely arising purely from the constraints on systematics which are not flux-conserved.

Table 5. Comparison of individual eclipse depths using aperture extraction from the primary reduction analysis.

Model	t_{cut}	E1	E2	E3	Comb.
L	30	499±53	287±51	313±52	364±30
	45	446±53	273±52	316±52	344±30
	60	412±56	248±53	309±55	320±32
L+E	30	305±71	235±65	291±61	276±38
	45	327±76	211±66	278±71	267±41
	60	324±76	228±76	244±80	266±45
L+GP	30	389±126	284±64	312±60	309±41
	45	421±85	272±62	313±59	319±38
	60	404±70	249±60	307±62	312±37
L+E+GP	30	306±79	244±75	288±69	279±43
	45	327±82	215±75	272±79	268±45
	60	321±85	226±80	243±88	262±49

Notes. Each model included some combination of a linear slope (L), exponential ramp (E) and/or a GP. Initial time cut is denoted in minutes.

We also tested fitting for the eclipse depth individually for each dataset. In this case, we assumed a circular orbit and used constraints on the central eclipse time T_{sec} from the other two eclipses as a prior on each individual fit as each eclipse alone gave poor constraints on the eclipse time. These priors are listed in Table C.1. We experimented with varying the initial time being cut (t_{cut}) and the effect of fitting for an exponential ramp or using a GP. Due to the computational cost of the pixel-fitting method, we only tested these variations with aperture extraction. The results are presented in Table 5 showing the eclipse depths individually for each eclipse as well as the combined result from the three eclipses calculated using error propagation (i.e. not a joint-fit).

The individual eclipse depths show that fitting for an exponential ramp has a large effect on the result. Given the strong settling ramp seen at the beginning of the first eclipse (see Figure 8), the eclipse depth may be biased if this ramp is not accounted for – even after cutting a full hour at the start. This would explain why the linear model without a GP (denoted ‘L’ in the Model column) recovers very deep eclipses (>400ppm) inconsistent with the other two eclipses. The linear model with a GP (L+GP) performs slightly differently and recovers large uncertainties likely because the GP height scale is increasing to fit this initial settling ramp. As initial settling effects are not in line with a stationary Gaussian process, we would prefer not to use a GP to fit them and so our preferred choice of model is the linear with exponential ramp and a GP (L+E+GP) model. This performs very similarly to the linear with exponential ramp without a GP (L+E) model but including a GP is slightly more conservative as it can account for additional systematics which could potentially lie in these data. We see that both these models give eclipse depths that are all within 2σ of each other and varying the time cut between 30 and 60 minutes has minimal impact on the result.

4.3.2. Results from secondary reduction analysis

The secondary analysis, conducted using an independent pipeline and the *juliet* package to fit the light curve, yields consistent results. Table 6 presents the eclipse depth results

Table 6. Comparison of jointly fitted eclipse depths for various methods from the secondary analysis.

Model	Extraction	Comb. depth (ppm)	$\ln \mathcal{Z}$
L	Optimal	423±32	17995
	Classic	283±33	17849
E	Optimal	257±33	18048
	Classic	258±35	17869
L+E	Optimal	320±40	18012
	Classic	211±37	17845
L+GP	Optimal	254±53	18026
	Classic	257±45	17841
L+E+GP	Optimal	235±70	18016
	Classic	268±51	17828

Notes. 200 integrations are cut at the beginning of each observations – equivalent to 37 minutes.

Table 7. Eclipse depths for the second eclipse either including or excluding the pixels containing a strong cosmic ray jump.

Cosmic ray pixels:	Excluded	Included
Ap. Ext. w/o GP	211 ± 66	181 ± 61
Px-fit (only FCS)	195 ± 73	185 ± 64
Ap. Ext. w/ GP	215 ± 75	189 ± 70
Px-fit (common systematics)	194 ± 80	188 ± 74
Px-fit (common+indep. sys.)	208 ± 99	214 ± 99
Px-fit (common+shared indep. sys.)	204 ± 85	203 ± 87

for various detrending methods in the combined fit. The combined eclipse depths remain consistent within 1σ of those from the primary analysis, which used classic aperture photometry or pixel-fitting, except for the L and L+E models in the secondary analysis. However, these models are strongly disfavored, compared to the other models. The use of GPs inflates the uncertainty.

4.4. Including the masked cosmic ray pixels

One interesting test case for the pixel-fitting method is to include the pixels which appear to suffer a large persistence effect from a cosmic ray jump in the second eclipse (highlighted in Figure 1). These pixels were masked out of all pixel-fits and aperture extraction fits for the primary reduction analysis. The eclipse depths recovered including or excluding these pixels are given in Table 7 for aperture extraction and pixel-fitting methods. We can see that including these pixels improves the constraints on the eclipse depth for most of the methods even though these pixels do not appear to contain clean data, and the aperture extraction methods also get their mean values shifted down by 25–30 ppm. However, the pixel-fitting methods which account for systematics in individual pixels (labelled ‘common+indep. sys’ and ‘common+shared indep. sys.’) are not affected by the inclusion of these pixels. This could be analogous to the simulations in the IPS scenario (Section 3.6), where the pixel-fitting method would weight away from pixels containing independent pixel systematics and by doing so the eclipse depth was recovered more precisely.

We also experimented with using the pixel-fitting GP predictive mean to visualise the systematics in individual pixels.

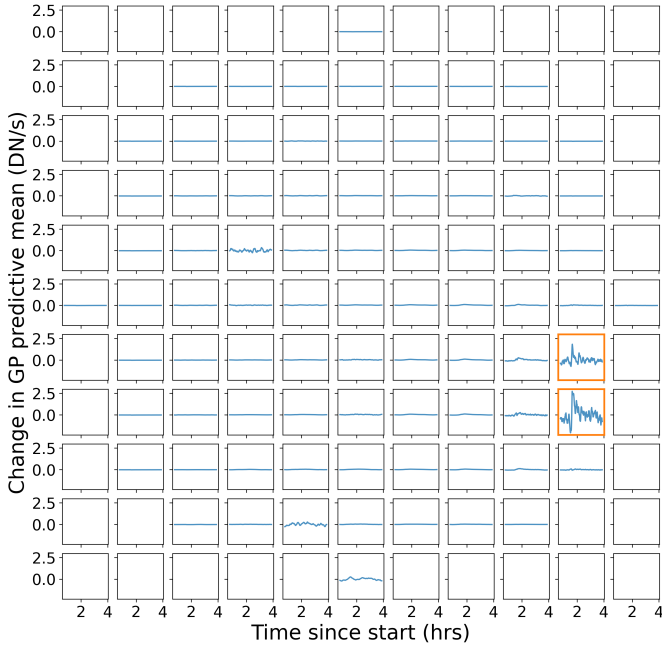


Fig. 10. Difference in GP predictive mean from pixel-fitting the second eclipse due to fitting for independent pixel systematics. Pixels affected by the strong cosmic ray persistence effect from Figure 1 are highlighted in orange. By fitting for independent pixel systematics the shape of this persistence effect is modelled by the GP and these pixels are also weighted away from.

We calculated the GP predictive mean fitting the pixel light curves both accounting for independent pixel systematics or not accounting for them (i.e. setting all $h_{\text{IPS};(i,j)} = 0$). Getting the difference between these two GP means gives an idea of how the inclusion of accounting for independent pixel systematics changes the fit to the pixel light curves, potentially revealing some of the shape of these systematics being fit. This is shown in Figure 10 and the pixels featuring the cosmic ray jump are highlighted.

4.5. Variations in detector settling

We investigated potential explanations for the variations in detector settling seen at the start of our three LHS 1140c observations (shown in Figure 8). It was noticed that the previous MIRI observation for the first eclipse was a calibration observation which was taking darks for MIRI/Imaging. This observation also shows a strong negative slope very different from the start of the other two eclipses. We note that this calibration observation ended >5 hours before the start of the first LHS 1140 observation (the last observation of any kind was carried out using NIR-Spec/IFU). However, MIRI typically leaves the previous filter it used in position until almost immediately before the next MIRI observation begins. In this case, the OPAQUE filter was kept in place until switching to the F1500W filter, leaving a long time for each pixel to settle without receiving flux. For the other two eclipses, the F1000W and F770W filters were left in place until right before the observation.

We note that there is no shutter on MIRI, as the detector is continuously observing the sky and constantly resetting after each read; this means that the telescope will spend time observing LHS 1140 with the previous filter in the time after slewing to the target but before switching filter. It has already been noted

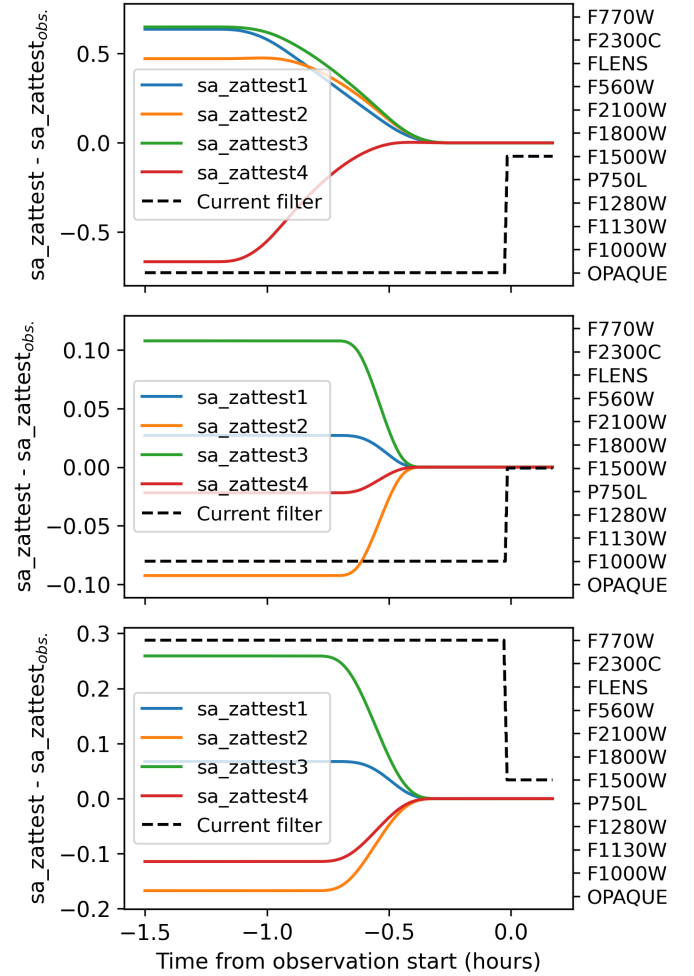


Fig. 11. JWST mnemonics which track telescope orientation (sa_zattest) overlaid with the mnemonic tracking current MIRI filter position. Left y-axis shows difference in telescope pointing parameters from our observations. Filter wheel position is plotted as black dashed line and corresponds to right y-axis, including any intermediate filters rotated through. In all three eclipses, coarse slewing towards the target finishes ~15 minutes before the observation starts, leaving time for persistence effects from the last filter used.

that this change in filters can lead to persistence effects from different amounts of background flux in each filter (Dicken et al. 2024). However, the PSF of LHS 1140 observed through the different wavelength filters may also have sufficient time to produce persistence effects and affect the beginning of our observations. Examining this using the JWST engineering mnemonics for telescope orientation and MIRI filter position appears to support this possibility, with the telescope pointing in a similar direction to the observation period for about 10–15 minutes before the filter is switched to F1500W (see Figure 11).

To test this idea, each of the observations in the Hot Rocks survey which used the SUB256 subarray were reduced in the same manner as described in Section 4.1. The four observations in the survey which used either the SUB64 or SUB128 subarrays would require a different background subtraction and occur on completely separate pixels on the detector so we restricted our comparison to just the SUB256 observations. After we had performed the same background subtraction and 5px radius aperture extraction as for the LHS 1140c eclipses, we took the first 30 minutes of each observation (clipping the first integration to

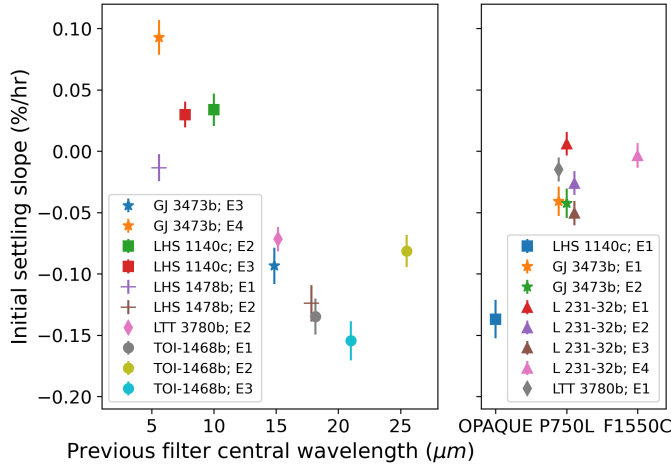


Fig. 12. Possible trend in detector settling based on the previous filter used by MIRI. The mean and uncertainty in the slope of the first 30 minutes of the aperture extracted light curve is plotted for various Hot Rocks observations. Left plot shows previous imaging filters based on their central wavelength, the right plot includes the OPAQUE position for darks, the P750L prism for LRS and a coronagraphic filter. Observations which shared the same previous filters often display similar detector settling.

avoid the ‘first exposure’ effect) and fit a best-fit line to it. We then compare the slope of the first 30 minutes of each observation to the previous filter used. The previous filter used was found with the `IMIR_HK_FW_CUR_POS` engineering mnemonic, which provides the MIRI filter currently in position. In all cases, the observations are being switched to the F1500W filter. See Figure 12 for a plot of these results.

There appears to be a pattern where the observations which previously used short wavelength filters (i.e. F560W, F770W, F1000W) show flat or positive slopes at the start, while longer wavelength filters up to the F2000W filter show increasingly negative slopes. The F2550W filter doesn’t quite match this trend but still shows a negative slope. The OPAQUE filter performs similarly to the long wavelength filters with a strong negative slope, while observations which previously had the P750L prism used for MIRI/LRS have slightly negative or flat slopes. The fourth eclipse of L 231-32b was the only observation which previously used a coronagraphic filter (F1550C), with this observation being consistent with a flat slope. We note that observations which both used the same previous filter tend to have quite consistent slopes, even when the targets being observed are different. This suggests that it may be possible to improve the consistency of the initial detector settling by switching filter earlier (e.g. before slewing or before guide star acquisition), which might enable better characterisation of detector settling and allow us to clip less data.

However, the two observations which previously used the F560W filter show inconsistent settling, suggesting that there may be other important factors. For example, examining engineering mnemonics for these observations (not shown but similar to those for LHS 1140c in Figure 11) suggests that they differ significantly in terms of the time spent on target with the previous filter. The fourth eclipse of GJ 3473b spends ~45 minutes on target with the F560W filter while the first eclipse of LHS 1478b only spends ~10 minutes. This could explain the strong positive slope seen for the fourth eclipse of GJ 3473b, as it has more time to build up a significant persistence effect with the previous filter compared to the first eclipse of LHS-1478b.

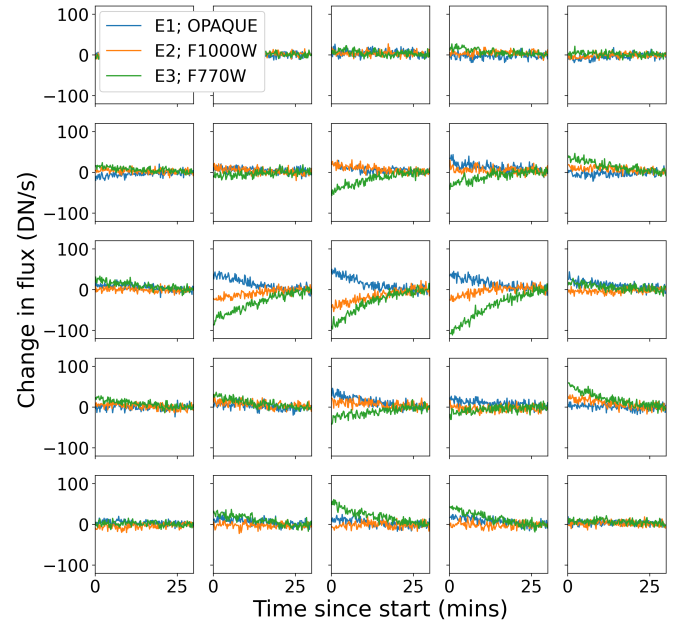


Fig. 13. Central pixel light curves for the first 30 minutes of each LHS 1140c eclipse. Each plot shows the change in flux relative to the last ten integrations of the first 30 minutes. The legend specifies which previous filter was in place for each eclipse. Note the strong differences in the central pixel light curves, particularly between the first eclipse (which switched from the OPAQUE filter) and the third eclipse (which switched from the F770W filter).

It is worth noting that if we avoid performing any background subtraction we recover similar slopes in the aperture extracted light curves. Most settling slopes are shifted by less than 0.01%/hr if a background subtraction is not performed, with the most affected observation being the first LHS 1140c eclipse which is shifted 0.035%/hr flatter (i.e. $-0.137\%/hr$ to $-0.102\%/hr$). This implies that any trend is primarily driven by the strong settling effects seen in the central pixels of the PSF and not related to persistence effects in the background pixels. The strong settling of the central pixels is visualised in Figure 13, which shows the pixel light curves centred on the PSF for the initial 30 minutes of each LHS 1140 observation. Consider that when the F770W filter is in position prior to the third observation, this filter should produce a 1.8 times narrower PSF which has about seven times greater total flux compared to the F1500W filter (based on the filter bandpasses and BT-Settl models). This combination of a narrower PSF and a bandpass which collects more flux could likely explain the significant differences in the persistence effects seen on the central pixels compared to an observation which previously had the OPAQUE filter in and hence was exposed to very little light in the time immediately before the observation began. One caveat is that there is some time required to switch filter ($\approx 8s$ per filter position; Wright et al. 2015), so the filters which are in between the previous filter and the F1500W filter could also introduce some persistence effects, although there is significantly less time for this to occur.

4.6. Pixel-fitting parameters recovered

In addition to recovering the eclipse depth, the pixel-fitting method also recovers hyperparameters which quantify the level of different detector systematics. These hyperparameters have been described in Section 2.4 as parameters of the kernel

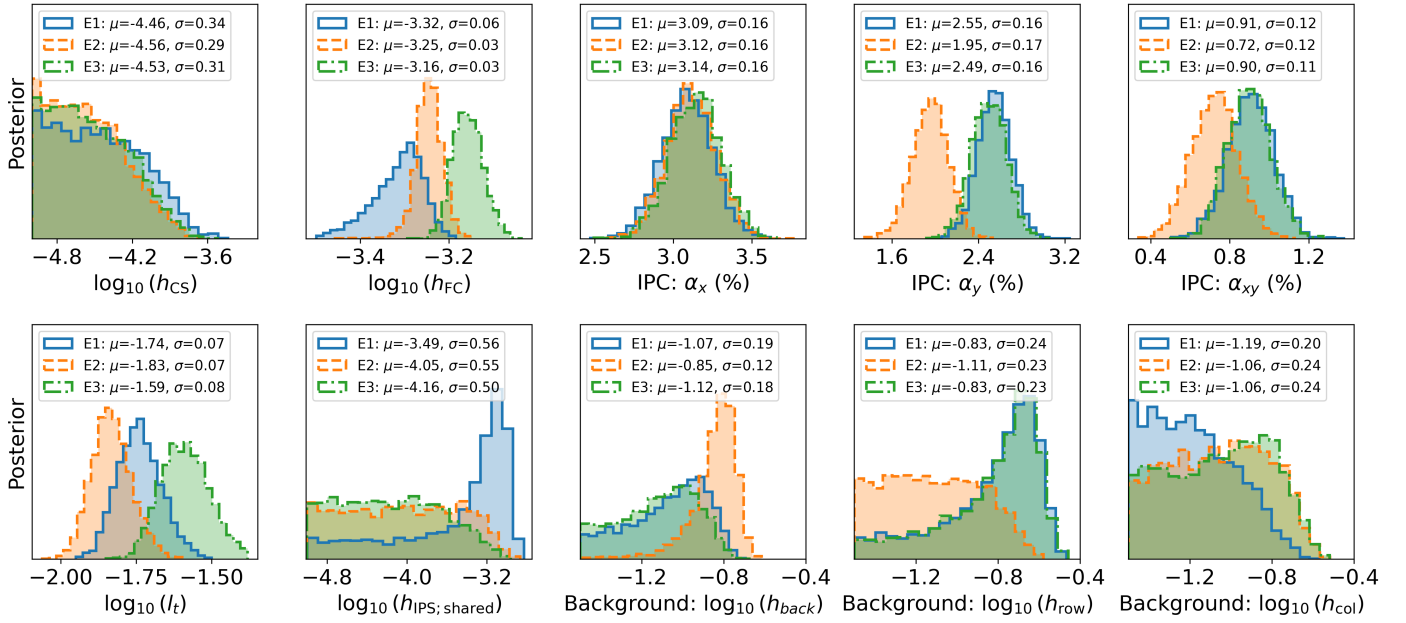


Fig. 14. Marginal posterior distributions of various pixel-fitting systematics parameters for each eclipse. Values from joint-fit of three eclipses with independent systematics parameters for each eclipse. Common systematics and correlated background scatter were not strongly detected in any eclipse and were consistent with lower bounds. Flux-conserved systematics, interpixel capacitance and the time length scale of systematics show consistent values across each observation.

function, used to describe the correlation between different noise processes in the data. We have plotted the posteriors of some of these parameters in Figure 14, all of which come from the joint-fit of the three eclipses using the pixel-fitting method which was accounting for both common systematics and independent systematics with a partially shared height scale. As stated earlier, while these datasets were being joint-fit, only the eclipse model parameters were being shared between datasets and the systematics parameters were independent for each dataset.

It is notable that the systematics parameters are all consistent between each dataset. None of the datasets constrain the presence of common systematics (with height scale h_{CS}) or found strong evidence of independent pixel systematics from the shared heightscale ($h_{IPS;shared}$). The amplitude of the flux-conserving systematics (h_{FCS}) – which could be interpreted as the amplitude of the PSF changing shape over time – appears to be well constrained. The correlations in noise arising from interpixel capacitance in the horizontal (α_x), vertical (α_y), and diagonal (α_{xy}) directions are also tightly constrained and greater than zero. Combining these constraints through error propagation we get $\alpha_x = 3.12 \pm 0.09\%$, $\alpha_y = 2.35 \pm 0.09\%$, and $\alpha_{xy} = 0.84 \pm 0.07\%$ – consistent with expected values (Morrison et al. 2023).

The length scale in time of the systematics, l_t , is also well constrained. However, there was a prior placed on each l_t parameter taken from the posterior of the joint-fit accounting for common systematics but not independent pixel systematics. This was because the addition of independent pixel systematics could create a degeneracy between very gradual systematics in each pixel and the slope in each pixel being fit for in the mean function. As a result, the constraints on l_t are more closely related to the timescale of the flux-conserving systematics and so could represent the rate at which the PSF is changing shape (in units of days).

Finally, the parameters describing correlations in background scatter do not definitively constrain correlations in background noise either across all pixels (h_{back}) or along the rows

(h_{row}) or columns (h_{col}). The inclusion of fitting for correlations in noise along each row and column was largely a conservative choice as there are a number of row/column effects seen with MIRI and the lack of a detection does not rule these systematics out but instead puts an upper bound on their presence in these data. However, the background subtraction is not perfect so h_{back} almost certainly should be non-zero and all three eclipses do appear to favour the value being at least somewhat larger than the minimum prior bound of $h_{back} = 10^{-1.5} \approx 0.032$ DN/s. By including more pixels or if our datasets had more integrations, this might provide the method more information to constrain the remaining level of background scatter in the data. This could be interesting as a way to compare different background subtraction strategies and identify which method best minimises the background scatter.

4.7. Absolute flux calibration

As our eclipse depth measurements correspond to the planet-to-star flux ratio, f_p/f_* , our measurement of the dayside planetary flux f_p is sensitive to our models of flux from the host star across the F1500W filter bandpass. In addition, stellar models are required to determine the amount of radiation the dayside is exposed to, which will determine the expected flux emitted for each model. Inaccuracies in our stellar models may therefore affect our eclipse spectra models in two different ways. While we do not have spectra to examine the accuracy of these models across near to mid-infrared, we can use our observations to perform an absolute flux calibration and examine how closely the BT-Settl models match the observed flux across the F1500W bandpass.

We performed the absolute flux calibration following the procedure used for the JWST Absolute Flux Calibration program (Gordon et al. 2025). This involved running the standard JWST calibration pipeline but skipping the EMI correction step EMICORR and turning off the after-jump flagging and shower

detection steps of the JUMP step. We skipped the PHOTOM step in stage 2 and instead used the values given in [Gordon et al. \(2025\)](#) to convert from DN/s to millijanskys (mJy). PHOTUTILS was used for the centroiding and aperture extraction, with an aperture radius of $r = 5.69\text{px}$ and background annulus radii of 8.63px and 11.45px – matching the radii used for the program. A conversion factor is used to convert the aperture extracted flux using a finite aperture to the total flux expected from an ‘infinite’ aperture and this factor also accounts for the oversubtraction of background flux due to some of the PSF being included within the background annulus.

We performed the flux calibration for each integration of our observations and calculated the mean flux. This resulted in flux values of 9.17 ± 0.04 mJy, 9.19 ± 0.04 mJy, and 9.17 ± 0.04 mJy for each eclipse respectively, with the uncertainties determined using the reported uncertainty in the calibration factor of 0.48% (the uncertainties due to Poisson noise are negligible in comparison). In comparison, when we calculate the expected flux weighted across the F1500W bandpass using a grid of BT-Settl (CIFIST) models ([Allard et al. 2011](#)), we find an expected flux of 8.33 ± 0.32 mJy. The uncertainty in this value propagates uncertainties in the stellar distance $d = 14.9861 \pm 0.0153$ pc from Gaia DR3 ([Gaia Collaboration 2021](#)) and in the stellar radius, R_* , stellar effective temperature, T_{eff} , and surface gravity, $\log(g)$, from [Cadieux et al. \(2024a\)](#).

This makes the measured flux $\sim 2.6\sigma$ larger than the flux expected from the BT-Settl models after marginalising over T_{eff} and $\log(g)$ – or $(10.3 \pm 3.2)\%$ greater than the interpolated BT Settl model with T_{eff} and $\log(g)$ fixed to mean values⁴. If this increased stellar flux is real and only deviates from the BT-Settl models in a region around the F1500W bandpass, then we may expect this to roughly shift our bare rock and atmospheric eclipse models to be $\sim 10\%$ shallower by affecting the f_p/f_* ratio. In principle, this should not dramatically affect our conclusions given our $\sim 16\%$ uncertainty in the measured eclipse depth (and in fact may provide stronger evidence on the lack of an atmosphere). However, this does not account for how stellar model inaccuracies may affect the radiative transfer and energy balance equations or how the BT-Settl models may deviate from the stellar spectrum at other wavelengths. We must therefore advise some caution in the interpretation of our eclipse models. We expect that future observations of M-dwarf spectra at near to mid-infrared will help us improve the accuracy of these models as there are currently significant challenges facing the modeling their spectra ([Shields et al. 2016](#)).

Our result of $\sim 10\%$ increased flux is in contrast with the flux calibration results for TRAPPIST-1, as [Greene et al. \(2023\)](#) reported a 13% reduced flux compared to PHOENIX models, while [Ducrot et al. \(2024\)](#) and [Ih et al. \(2023\)](#) report only $\sim 7\%$ reduced flux – both using SPHINX models. On the other hand, [Zieba et al. \(2023b\)](#) report an expected flux using SPHINX models which is consistent within $\sim 1\%$ of the absolute flux calibrations. These works differ significantly in their methods of estimating flux from stellar models and in their choice of stellar parameters and stellar models used. However, all absolute flux calibrations performed were consistent within $\sim 1\%$, although their methods differ from [Gordon et al. 2025](#) and this work. We tested our own expected flux for TRAPPIST-1 using BT-Settl models and using the stellar parameters from

[Agol et al. \(2021\)](#); [Gaia Collaboration \(2021\)](#) and we obtained a flux of 2.507 ± 0.077 mJy, which similar to the [Zieba et al. \(2023b\)](#) result is within 1–3% of all reported absolute flux calibrations. Therefore, our flux estimates from BT-Settl models are consistent with the TRAPPIST-1 flux calibrations yet differ from our LHS 1140 flux calibrations.

We may note that the calibration factor from the program was determined using much shorter duration observations often with only one or two integrations. Therefore these observations may have been affected by the reduced flux often seen in the first integration of each observation ([Morrison et al. 2023](#)). If we calculate the flux using just the first integration of each eclipse, we do get reduced flux values of 9.10 ± 0.04 mJy, 9.09 ± 0.04 mJy and 9.09 ± 0.04 mJy for each eclipse, however these are still $\sim 2.3\sigma$ deeper than expected so this makes little difference.

The absolute flux value can be used in combination with the eclipse depth to calculate the brightness temperature of LHS 1140c. This is performed by multiplying the eclipse depth, absolute flux value and the square of the ratio of the distance the system is away to the radius of the planet. This is compared to integrating Planck’s law over the F1500W bandpass and finding a brightness temperature T_B which corresponds to this level of emission. We performed this by calculating the emission from Planck’s law over a range of brightness temperatures between 200 and 800 K in 0.1K intervals and interpolating to invert emission to brightness temperature. Marginalising over the uncertainty in the eclipse depth from aperture extraction with a GP ($d = 273 \pm 43$ ppm) as well as the system distance and planetary radius, we recover a brightness temperature of $T_B = 561 \pm 44$ K. This is close to the maximum brightness temperature we might expect from a (smooth) zero albedo bare rock with no heat redistribution of $T_{B,\text{max}} = 537 \pm 9$ K, while full heat redistribution even with zero albedo would be inconsistent with this value at only $T_{B,\text{full}} = 421 \pm 7$ K (and higher albedos would be even colder). However, this is only a simple calculation which does not take into account effects such as the greenhouse effect, scattering, or thermal beaming (discussed in Section 5). This calculation does have the benefit of not requiring the use of (potentially unreliable) stellar models while the atmospheric modelling described in Section 5 requires the use of stellar models to account for the wavelength-dependent effects of absorption and emission as well as scattering.

5. Modelling

For the no atmosphere case, we can calculate the dayside temperature by balancing thermal emission and absorbed stellar radiation:

$$\epsilon T_{\text{day}}^4 = T_*^4 \left(\frac{R_*}{a} \right)^2 (1 - A_{\text{surf}})(f/4) \quad (21)$$

where T_* is the stellar effective temperature, R_* is the stellar radius, a is the semi-major axis, A_{surf} is the surface albedo (i.e. the Bond albedo of the surface), ϵ is the surface emissivity, and f is the heat redistribution factor, where $f = 1/4$ represents full heat redistribution, while $f = 2/3$ represents zero heat redistribution ([Hansen 2008](#)). This equation is adapted from [Malik et al. \(2019a\)](#) and assumes there is no extra heat emitted due to a high interior planetary temperature. For the bare rock case we set $f = 2/3$ as we expect insignificant heat redistribution without an atmosphere ([Joshi et al. 1997](#); [Selsis et al. 2011](#); [Koll 2022](#)). For all simulations, we assumed that the surface albedo was constant in wavelength, which implies $\epsilon = 1 - A_{\text{surf}}$ (as emission and absorption can occur at different wavelengths this is not

⁴ One caveat is that as we are scaling the BT-Settl models based on the stellar radius and system distance, it is possible that the BT-Settl models are accurate but for example the stellar radius value from [Cadieux et al. \(2024a\)](#) is inaccurate.

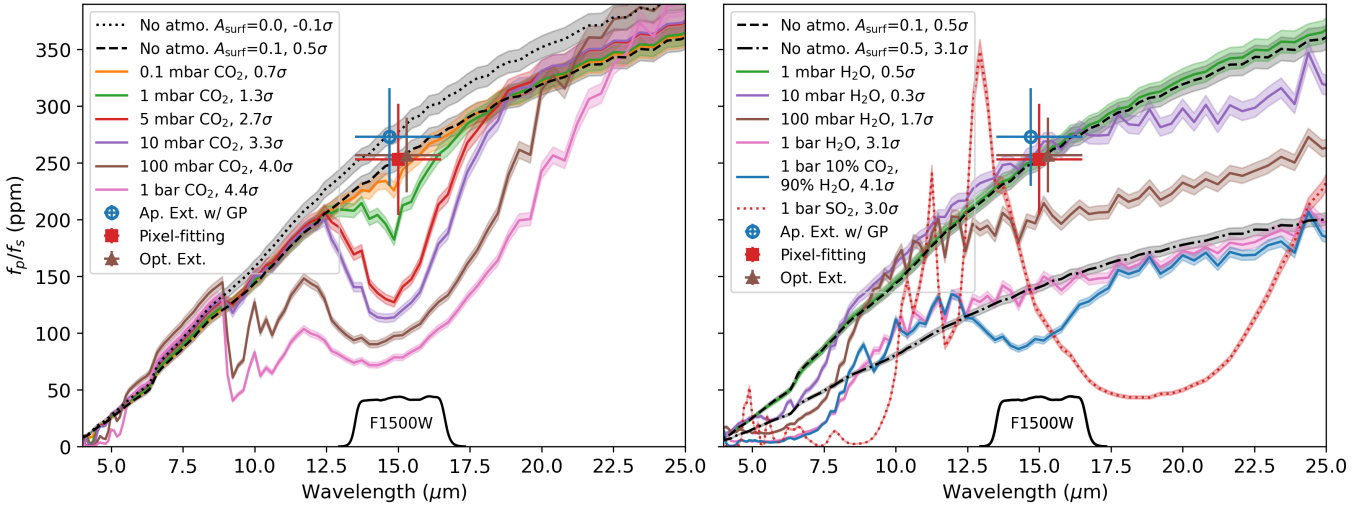


Fig. 15. Atmospheric forward models compared to eclipse depth measurements using various analyses. Spectra are compared to eclipse depths recovered using aperture extraction with a GP, pixel-fitting or a secondary analysis fit using optimal extraction without a GP. Left: emission spectra of bare rock and pure CO₂ models at varying surface pressures. Statistical significance given relative to joint-fitting aperture extracted light curves using GPs. Right: similar plot for pure H₂O, pure SO₂ or H₂O/CO₂ mixture atmospheres. All atmospheric models account for heat redistribution using the f-factor parameterisation in Koll (2022) and use a surface albedo of $A_{\text{surf}} = 0.1$.

necessarily true otherwise). Emission was then calculated using Planck’s law with the dayside temperature T_{day} and scaling by the emissivity ϵ .

We also added the reflected light component,

$$\frac{f_{\text{refl}}}{f_*} = A_g \left(\frac{R_p}{a} \right)^2, \quad (22)$$

where A_g is the geometric albedo. For a Lambertian reflector we get $A_g = (2/3)A_{\text{surf}}$ (Heng et al. 2021). Note for our target $(R_p/a)^2 \approx 4$ ppm so reflection is only a minor contribution.

For the atmospheric cases, we used the 1D atmospheric model HELIOS (Malik et al. 2017, 2019b) with a solid surface bottom layer (Malik et al. 2019a; Whittaker et al. 2022) to calculate the planet’s temperature structure and its synthetic emission and reflection spectra. The temperature profile of a planet’s atmosphere is primarily determined by the balance between radiative, convective processes and heat redistribution. The model considers well-mixed atmospheres with various gas compositions consisting of CO₂, H₂O, SO₂, N₂ and O₂. It employs k-distribution tables for opacities, integrating radiative fluxes over 386 spectral bands and 20 Gaussian points. Opacities are calculated using the HELIOS-K (Grimm & Heng 2015) code and using cross-sections for CO₂ from Rothman et al. (2010), H₂O from Barber et al. (2006), SO₂ from Underwood et al. (2016) and O₂ from Gordon et al. (2022). Additionally, we incorporate Rayleigh scattering cross-sections for H₂O, CO₂, N₂, and O₂ (Cox 2000; Snee & Ubachs 2005; Thalman et al. 2014), but did not include scattering for SO₂. For all atmospheric models, the surface albedo was fixed to $A_{\text{surf}} = 0.1$.

Heat redistribution by winds also plays an important role in determining the temperature structure of the atmosphere. The heat redistribution in HELIOS is controlled by the f-factor (the same f parameter used in Equation (21) for the no atmosphere case). The f-factor was approximated by the scaling equation from Koll (2022). This analytical equation depends on the long-wave optical depth at the surface, the surface pressure and the equilibrium temperature.

Stellar spectra are taken from BT-Settl (CIFIST) models (Allard et al. 2011) and linearly interpolated for the stellar parameters. We did not choose to adjust these stellar models based on the absolute flux calibration performed in Section 4.7 so note that these stellar models may be underestimating flux by $\sim 10\%$ in the F1500W bandpass.

Figure 15 shows a range of models including pure CO₂, pure H₂O, pure SO₂, and a CO₂/H₂O mixture. Each model was run 100 times with the input planetary parameters R_p , a , g_p (planetary surface gravity) and stellar parameters R_* , T_{eff} , $\log(g)$ randomly drawn from their constraints listed in Cadieux et al. (2024a) to propagate the uncertainty in these parameters. The figure displays the emission spectra resulting from the modelling, with the shaded regions representing the 1σ range of the computed values.

We find that the eclipse depths from aperture photometry, pixel-fitting and the independent analysis using optimal extraction are all in excellent agreement with a low-albedo bare rock ($A_{\text{surf}} \sim 0.1$). Higher albedo surfaces of $A_{\text{surf}} = 0.5$ diverge at $>3\sigma$ (uncertainties given relative to the joint-fit using aperture extraction with GPs; $d = 273 \pm 43$ ppm). Extremely tenuous (<1 mbar) CO₂ or H₂O atmospheres naturally behave similarly to bare rocks and cannot be distinguished using these observations.

However, we do find that a pure CO₂ atmosphere with a surface pressure of only $P_s = 5$ mbar is discrepant with observations at 2.7σ , making a Mars-like atmosphere (~ 6 mbar, 95% CO₂) unlikely. These constraints are only weakly affected by our heat redistribution model as fixing $f = 2/3$ (i.e. no heat redistribution) has a negligible effect on the 5 mbar result and only slightly weakens the 100 mbar and 1 bar pure CO₂ results to be between 3 and 4σ . This implies that our constraints on the lack of a thick pure CO₂ atmosphere largely derive from the lack of CO₂ absorption rather than from the lack of heat redistribution. We compare the 5 mbar and 1 bar pure CO₂ eclipse depths to our phase-folded light curve fit in Figure 16 to show how far they diverge from the light curve data.

Pure H₂O atmospheres are unlikely from a theoretical perspective as models suggest they are highly vulnerable to atmospheric escape for planets around M dwarfs

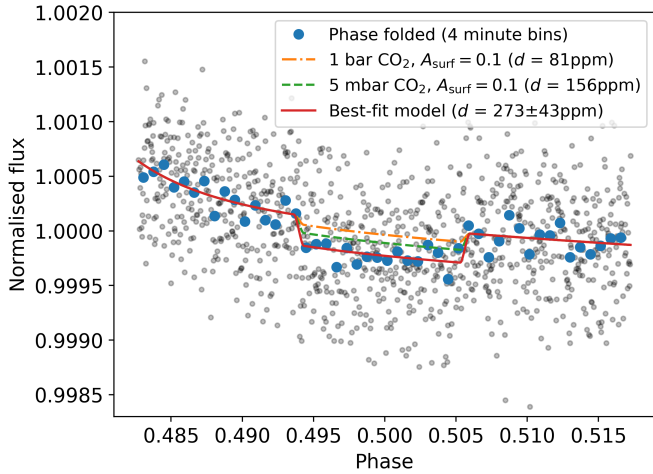


Fig. 16. Phase-folded observations compared to model light curves of pure CO₂ atmospheres. Observations plotted without binning (black) and with binning (blue circles). In red is the phase folded best-fit eclipse model from joint-fitting the aperture extracted light curves with a GP. The orange and green dashed lines show expected eclipse depths for different pressure pure CO₂ atmospheres.

(Luger & Barnes 2015; Kreidberg et al. 2019). From our data, we cannot rule out tenuous 10 mbar models but we can rule out 1 bar H₂O atmospheres at 3.1σ . In contrast to the pure CO₂ atmospheres, this result is completely dependent on our heat redistribution model and is not the result of a particular absorption feature. All atmospheric models plotted in Figure 15 use the analytic f-factor parameterisation from (Koll 2022), but if we instead fix $f = 2/3$ then all pure H₂O atmospheres up to 1 bar are within 1σ of our results (not plotted). We have also included a mixed 10% CO₂, 90% H₂O 1 bar atmosphere which appears to behave similarly to the pure H₂O 1 bar atmosphere but with the addition of a CO₂ absorption feature over the 15 μ m region observed, making it less consistent with observations.

As we may expect volcanically outgassed secondary atmospheres to include sulphur-based molecules such as SO₂, we also tested a range of pure SO₂ atmospheres. In reality, sulphur chemistry is complex and SO₂ is very photochemically sensitive; thus, more sophisticated photochemical modelling would need to be performed to understand what the dominant form of sulphur would be for this planet (see e.g. Bello-Arufe et al. 2025). It is likely that a dense SO₂ atmosphere would require constant replenishment from volcanic outgassing, which given the low eccentricity of the planet ($e < 0.0172$ to 95% confidence, see Appendix C) would not be expected from tidal forces, unlike suggestions for the L 98-59 system (Seligman et al. 2024). Noting these caveats, we recover that pure 1 bar SO₂ atmospheres are ruled out at 3σ when modelling heat redistribution (plotted in Figure 15); however, without any heat redistribution, this result would decrease to 2.3σ . Atmospheres of 100 mbar or weaker are all within 2σ regardless of the effect of heat redistribution. We note that longer wavelengths such as those covered by the F1800W filter might provide significantly tighter constraints for thinner SO₂ atmospheres. We have included models of SO₂ in a CO₂ background in Appendix E which suggest that combinations of these molecules may produce particularly strong signatures which we can rule out down to 1 mbar. However, we again note that these models exclude any photochemical or geochemical modelling and we hope future work can examine to what significance physically plausible volcanically outgassed atmospheres are ruled out to.

Table 8. Statistical significance in std. dev. of various atmospheres containing CO₂ with N₂ as a background gas.

Pressure	1 ppm CO ₂	100 ppm CO ₂	1000 ppm CO ₂
0.1 mbar	0.6	0.6	0.6
1 mbar	0.6	0.6	0.7
5 mbar	0.6	0.7	0.8
10 mbar	0.6	0.7	0.9
50 mbar	0.6	1.0	1.9
100 mbar	0.7	1.5	2.7
500 mbar	0.9	3.1	4.0
1 bar	1.2	3.7	4.3
5 bar	2.7	4.5	4.8
10 bar	3.4	4.8	5.0

Notes. Results given relative to the aperture extraction joint-fit with a GP ($d = 273 \pm 43$ ppm). Assumes a surface albedo of $A_{\text{surf}} = 0.1$.

We also tested a pure 1 bar O₂ atmosphere but found it indistinguishable from a bare rock model of the same albedo and have therefore not included it in Figure 15. This makes sense as O₂ is only weakly absorbing at these wavelengths and our heat redistribution model predicts minimal heat redistribution as the atmosphere is very optically thin. We highlight that the atmospheres this method is sensitive to are not simply any atmospheres with high surface pressures, they must also be reasonably optically thick for significant heat redistribution to occur (Koll 2022; Ih et al. 2023). We may expect a similar result for pure N₂ atmospheres.

In addition, we tested various concentrations of CO₂ in a background of N₂ at various surface pressures. The statistical significance of each model is listed in Table 8. We found that even low concentrations of CO₂ are ruled out with high confidence at sufficiently high surface pressure. For example, 100 ppm CO₂ in a 1 bar N₂ atmosphere is rejected at 3.7σ . We also ran each model with O₂ instead of N₂ and every result was affected by less than 0.2σ . This makes sense as these molecules are both weakly absorbing and are simply acting as inert background gases with similar mean molecular weights. This implies that an Earth-like atmosphere of 78% N₂, 21% O₂, ~400 ppm CO₂ at 1 bar is also strongly rejected by our observations.

One caveat is that our bare rock models do not account for surface roughness, which can make a bare rock surface appear hotter at eclipse relative to a smooth spherical surface. This happens because the surfaces tilted towards the observer are also tilted towards its host star and are therefore hotter (Hapke 1993; Coy et al. 2024). This effect – referred to as “thermal beaming” – has been observed for both the Moon (Wohlfarth et al. 2023) and Mercury (Emery et al. 1998). It is also the focus of a Cycle 2 program to help characterise the surface of the hot rocky exoplanet LHS 3844b (GO 4008; Zieba et al. 2023a). The effect may not be very significant at 15 μ m based on simulations of Mercury (a similar temperature planet) in Wohlfarth et al. (2023), but more modelling work would need to be performed to confirm this.

6. Discussion

Our observations rule out a wide range of possible atmospheres on LHS 1140c. While we cannot rule out very optically thin atmospheres such as pure O₂ atmospheres (which were a possible outcome of atmospheric escape modelling for TRAPPIST-1b Krissansen-Totton et al. 2024), these atmospheres would be

difficult to detect using any technique. The consistency with a bare rock despite our high sensitivity to CO₂ is in stark contrast to rocky planet atmospheres within our Solar System and also appears inconsistent with volcanically outgassed atmospheres. This work further backs up the finding in Cadieux et al. (2024b) that the 4 σ discrepancy between transit depths measured with Spitzer and TESS (previously noted in Cadieux et al. 2024a) may be resolved by the transit depth being underestimated by the TESS observations and is unlikely to be due to CO₂ absorption.

These constraints are very informative because the planet's low equilibrium temperature of 422 ± 7 K (Cadieux et al. 2024a), its relatively mild present-day radiation environment (Spinelli et al. 2019), and the reported age of the system (>5 Gyr; Dittmann et al. 2017), make it one of the better known candidates to have an outgassed atmosphere or to have retained some of its primary atmosphere (Kite & Barnett 2020; Chatterjee & Pierrehumbert 2024). In contrast, our result is in line with the suggestion that historic levels of XUV fluence may be underestimated for mid-type M dwarfs such as LHS 1140, making atmospheric retention more difficult (Pass et al. 2025). As this star also hosts LHS 1140b – a larger Super-Earth which likely still retains an atmosphere⁵ – our new atmospheric constraints are expected to help make the LHS 1140 system an important benchmark for understanding atmospheric escape and the cosmic shoreline (Zahnle & Catling 2017).

We have introduced a new technique, which can act as an alternative to aperture photometry or optimal extraction. Section 3 demonstrates that our method presents a clear advantage when we have individual pixels contaminated by systematics, as it can weight away from them towards cleaner pixels. For the JWST observations, we see in the case of our second eclipse that the pixel-fitting method could weight away from two discrepant pixels apparently affected by a persistence effect from a cosmic ray. While the method has disadvantages, such as its increased computational cost and complexity to implement, it provides a new way to analyse and explore data at the pixel-level which can identify problematic pixels in an empirical way. It is also important to note that while our observations do not appear heavily impacted by systematics, an observation of another target in our survey (LHS 1478b) appears to be significantly impacted by the presence of systematics (August et al. 2025), which our method may be able to alleviate. The fact that our method is capable of distinguishing between systematics in individual pixels or across all pixels could also help to diagnose the cause of these systematics.

Pixel-fitting therefore provides a powerful alternative to aperture photometry, but we do not claim it is a replacement for it. Highly flexible systematics models can come at the cost of providing more conservative results on clean data. This is a common trade-off when we analyse data for which we do not understand the extent of systematics present and the best approach is to use a range of methods varying in how conservative they are. By utilising both aperture photometry and pixel-fitting, we can test the robustness of our results and learn about potential systematics in our datasets. Having more independent techniques to analyse these data is particularly important given the decision for 500 hours of Director's Discretionary Time to be spent extending this technique to even more rocky exoplanets (Redfield et al. 2024).

The pixel-fitting method has huge potential as it could be applied to a wide range of existing time-series observations

including spectroscopic time series. There is currently a move within exoplanet atmospheres research to analyse JWST spectroscopic time series at the native or pixel resolution, yet there is concern over how much correlation this introduces in the noise between neighbouring light curves, with standard analysis approaches typically ignoring correlations between light curves (see Ih & Kempton 2021 and Fortune et al. 2024 for how this can affect atmospheric retrievals). We note that MIRI/LRS uses the same detector as MIRI/Imaging (but on a subarray that is different from that of our observations); thus, many of the detector systematics seen in our observations (e.g. the flux-conserving systematics) might affect MIRI/LRS observations too. Future work extending pixel-fitting to spectroscopy could examine whether this anti-correlation between neighbouring pixel light curves could explain the large increase in scatter in transmission spectra noted when extracting spectroscopic light curves with one pixel-wide bins (Bell et al. 2024). Implementing pixel-fitting for spectroscopy would differ as the eclipse depth would not be shared across all pixels – but only between pixels at the same wavelength. The computational cost could be a barrier to implementation as the spectral trace is spread across more pixels. Strategies such as binning in time and parallelising over multiple GPUs could help to mitigate this.

Finally, we believe the relationship between the initial detector settling slope and the previous filter used by MIRI warrants further study as a significant fraction of MIRI TSOs are affected by this. Therefore, mitigating these effects could provide a serious improvement to the efficiency of the instrument. A strategy such as changing when the filter is moved into place and/or changing how MIRI is clocked in idle mode could have a large effect on the initial settling slope.

7. Conclusions

Our three eclipses of LHS 1140c using MIRI/Imaging at 15 μ m have allowed us to rule out a wide range of possible atmospheres. We have found that LHS 1140c is consistent with a low-albedo bare rock and is unlikely to host an atmosphere capable of substantial heat redistribution or an atmosphere containing significant levels (>100 ppm) of CO₂. Using an absolute flux calibration, we measured its dayside brightness temperature to be $T_B = 561 \pm 44$ K, close to the maximum expected dayside temperature of $T_{\text{day; max}} = 537 \pm 9$ K. These results are fully consistent with the scenario that LHS 1140c lacks a substantial atmosphere. There are alternative possibilities, including dense atmospheres that are very optically thin, but these models may require some fine-tuning to explain the lack of strong absorbers such as CO₂. If LHS 1140c does indeed lack an atmosphere, then it could have strong implications for the cosmic shoreline model of atmospheric escape, as LHS 1140c is a super-Earth with one of the lowest instellations of any target this technique has been applied to so far (after TRAPPIST-1b and c; Greene et al. 2023; Zieba et al. 2023b). Given that the outer planet LHS 1140b is also a super-Earth, but one that is likely to host an atmosphere (Damiano et al. 2024; Cadieux et al. 2024b), this could place each planet on either side of the cosmic shoreline.

We have identified a possible relationship in the strength and sign of the initial settling slope with the previous filter used by MIRI. This may help to inform ways to mitigate this effect in future. We have also demonstrated a new method of analysing time series photometry, which produces similar results to aperture photometry on these datasets. The benefits of this method include:

⁵ There is still some discussion on the composition of LHS 1140b, see Cadieux et al. (2024b) and Damiano et al. (2024) for evidence distinguishing rocky, water world, or cloudy mini-Neptune scenarios.

- identifying and weighting away from systematics arising in individual pixels, potentially leading to more precise and robust measurements.
- identifying whether systematics arise from individual pixels or are common to all pixels, helping to determine the source of the systematics.
- weighting towards higher S/N pixels, similar to optimal extraction.
- retrieving additional information about the detector, such as how the PSF behaves over time and the strength of interpixel capacitance.
- constraining information about the amplitude of background scatter remaining after the background subtraction.

Overall, we hope this paper provides greater insight into the systematics present in MIRI data and provides useful approaches for dealing with them. This may help to robustly characterise more exoplanets in the future.

Acknowledgements. We are grateful to the anonymous referee for the careful reading and helpful suggestions which improved this manuscript. We would like to thank Charles Cadieux for providing orbital constraints on the LHS 1140 system and Karl Gordon for his advice in performing the absolute flux calibration. We also thank Bibiana Prinoth for a careful reading of the manuscript and helping to improve the clarity of this work. We acknowledge the input from the following individuals to the GO 3730 proposal: Can Akin, Adam Burgasser, Chloe Fisher, Matthew Hooton, Andrea Guzmán Mesa, Anna Lueber, Nicholas Borsato, Bibiana Prinoth, Jens Hoeijmakers, Meng Tian, Mette Baungaard. Based on observations with the NASA/ESA/CSA James Webb Space Telescope. The data were obtained from the Mikulski Archive for Space Telescopes at the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Incorporated, under NASA contract NAS 5-03127. Support for program number 3730 was provided through a grant from the STScI under NASA contract NAS5-03127. The raw data described here may be obtained from <https://dx.doi.org/10.17909/11g0-pd60>. The authors wish to acknowledge the Irish Centre for High-End Computing (ICHEC) for the provision of computational facilities and support. Many of the calculations in this work were performed on the astro01 system maintained by the Trinity Centre for High Performance Computing (Research IT). This system is co-funded by the School of Physics and by the Irish Research Council grant award IRCLA/2022/3788. N.P.G. gratefully acknowledges support from Science Foundation Ireland and the Royal Society through a University Research Fellowship (URFR\201032). H.D.L. and P.C.A. acknowledge support from the Carlsberg Foundation, grant CF22-1254. J.M.M. acknowledges support from the Horizon Europe Guarantee Fund, grant EP/Z00330X/1. N.H.A. acknowledges support by the National Science Foundation Graduate Research Fellowship under Grant No. DGE1746891. E.M.V. acknowledges financial support from the Swiss National Science Foundation (SNSF) Mobility Fellowship under grant no. P500PT_225456/1. B.-O.D. acknowledges support from the Swiss State Secretariat for Education, Research and Innovation (SERI) under contract number MB22.00046. A.D.R. acknowledges support from the Carlsberg Foundation, grant CF22-1548. We are grateful to the developers of NUMPY, SCIPY, MATPLOTLIB, ASTROPY, PHOTUTILS, PANDAS, IPYTHON, ARVIZ, CORNER, JAX, PYMC, JAXOPLANET and EXOPLANET as these packages were used extensively in this work (Harris et al. 2020; Virtanen et al. 2020; Hunter 2007; Astropy Collaboration 2022; Bradley et al. 2024; The pandas development team 2020; Pérez & Granger 2007; Kumar et al. 2019; Foreman-Mackey 2016; Bradbury et al. 2018; Salvatier et al. 2016; Hattori et al. 2024; Foreman-Mackey et al. 2021).

References

- Agol, E., Dorn, C., Grimm, S. L., et al. 2021, *Planet. Sci. J.*, **2**, 1
- Aigrain, S., & Foreman-Mackey, D. 2023, *ARA&A*, **61**, 329
- Alam, M. K., Gao, P., Adams Redai, J., et al. 2025, *AJ*, **169**, 15
- Allard, F., Homeier, D., & Freytag, B. 2011, *ASP Conf. Ser.*, **448**, 91
- Alonso, R. 2018, in *Handbook of Exoplanets*, eds. H. J. Deeg, & J. A. Belmonte (Berlin: Springer), 40
- Astropy Collaboration (Price-Whelan, A. M., et al.) 2022, *ApJ*, **935**, 167
- August, P. C., Buchhave, L. A., Diamond-Lowe, H., et al. 2025, *A&A*, **695**, A171
- Banerjee, A., Barstow, J. K., Gressier, A., et al. 2024, *ApJ*, **975**, L11
- Barber, R. J., Tennyson, J., Harris, G. J., & Tolchenov, R. N. 2006, *MNRAS*, **368**, 1087
- Bell, T. J., Cruzet, N., Cubillos, P. E., et al. 2024, *Nat. Astron.*, **8**, 879
- Bello-Arufe, A., Damiano, M., Bennett, K. A., et al. 2025, *ApJ*, **980**, L26
- Bradbury, J., Frostig, R., Hawkins, P., et al. 2018, JAX: composable transformations of Python+NumPy programs, <http://github.com/google/jax>
- Bradley, L., Sipőcz, B., Robitaille, T., et al. 2024, <https://doi.org/10.5281/zenodo.12585239>
- Cadieux, C., Plotnikov, M., Doyon, R., et al. 2024a, *ApJ*, **960**, L3
- Cadieux, C., Doyon, R., MacDonald, R. J., et al. 2024b, *ApJ*, **970**, L2
- Charbonneau, D., Allen, L. E., Megeath, S. T., et al. 2005, *ApJ*, **626**, 523
- Chatterjee, R. D., & Pierrehumbert, R. T. 2024, *ApJ*, submitted [arXiv:2412.05188]
- Cox, A. N. 2000, *Allen's Astrophysical Quantities* (London: Athlone Press)
- Coy, B. P., Ih, J., Kite, E. S., et al. 2024, arXiv e-prints [arXiv:2412.06573]
- Damiano, M., Bello-Arufe, A., Yang, J., & Hu, R. 2024, *ApJ*, **968**, L22
- Deming, D., Knutson, H., Kammer, J., et al. 2015, *ApJ*, **805**, 132
- Demory, B.-O., Gillon, M., de Wit, J., et al. 2016, *Nature*, **532**, 207
- Dicken, D., Marín, M. G., Shvaid, I., et al. 2024, *A&A*, **689**, A5
- Dittmann, J. A., Irwin, J. M., Charbonneau, D., et al. 2017, *Nature*, **544**, 333
- Ducrot, E., Lagage, P.-O., & Min, M. 2024, in *EGU General Assembly Conference Abstracts, EGU General Assembly Conference Abstracts*, 20183
- Eastman, J., Gaudi, B. S., & Agol, E. 2013, *PASA*, **125**, 83
- Emery, J. P., Sprague, A. L., Witteborn, F. C., et al. 1998, *Icarus*, **136**, 104
- Espinoza, N. 2022, <https://doi.org/10.5281/zenodo.6960924>
- Espinoza, N., Kossakowski, D., & Brahm, R. 2019, *MNRAS*, **490**, 2262
- Evans, T. M., Aigrain, S., Gibson, N., et al. 2015, *MNRAS*, **451**, 680
- Foreman-Mackey, D. 2016, *J. Open Source Softw.*, **1**, 24
- Foreman-Mackey, D., Agol, E., Ambikasaran, S., & Angus, R. 2017, *AJ*, **154**, 220
- Foreman-Mackey, D., Luger, R., Agol, E., et al. 2021, *J. Open Source Softw.*, **6**, 3285
- Fortune, M., Gibson, N. P., Foreman-Mackey, D., et al. 2024, *A&A*, **686**, A89
- Gaia Collaboration (Brown, A. G. A., et al.) 2021, *A&A*, **649**, A1
- Gehrels, N., Chincarini, G., Giommi, P., et al. 2004, *ApJ*, **611**, 1005
- Gelman, A., & Rubin, D. B. 1992, *Stat. Sci.*, **7**, 457
- Gibson, N. P. 2014, *MNRAS*, **445**, 3401
- Gibson, N. P., Aigrain, S., Roberts, S., et al. 2012, *MNRAS*, **419**, 2683
- Gibson, N. P., Aigrain, S., Barstow, J. K., et al. 2013, *MNRAS*, **428**, 3680
- Glasse, A., Rieke, G. H., Bauwens, E., et al. 2015, *PASP*, **127**, 686
- Gomes, G. O., & Ferraz-Mello, S. 2020, *MNRAS*, **494**, 5082
- Gordon, I. E., Rothman, L. S., Hargreaves, R. J., et al. 2022, *J. Quant. Spec. Radiat. Transf.*, **277**, 107949
- Gordon, K. D., Sloan, G. C., Garcia Marin, M., et al. 2025, *AJ*, **169**, 6
- Greene, T. P., Bell, T. J., Ducrot, E., et al. 2023, *Nature*, **618**, 39
- Gressier, A., Espinoza, N., Allen, N. H., et al. 2024, *ApJ*, **975**, L10
- Grimm, S. L., & Heng, K. 2015, *ApJ*, **808**, 182
- Hammond, M., Guimond, C. M., Lichtenberg, T., et al. 2025, *ApJ*, **978**, L40
- Hansen, B. M. S. 2008, *ApJS*, **179**, 484
- Hapke, B. 1993, *Theory of Reflectance and Emittance Spectroscopy*, Topics in Remote Sensing (Cambridge: Cambridge University Press)
- Harris, C. R., Millman, K. J., van der Walt, S. J., et al. 2020, *Nature*, **585**, 357
- Hattori, S., Garcia, L., Murray, C., et al. 2024, <https://doi.org/10.5281/zenodo.10736936>
- Heng, K., Morris, B. M., & Kitzmann, D. 2021, *Nat. Astron.*, **5**, 1001
- Hoffman, M. D., & Gelman, A. 2014, *J. Mach. Learn. Res.*, **15**, 1593
- Horne, K. 1986, *PASP*, **98**, 609
- Hu, R., Bello-Arufe, A., Zhang, M., et al. 2024, *Nature*, **630**, 609
- Hunter, J. D. 2007, *Comp. Sci. Eng.*, **9**, 90
- Ih, J., & Kempton, E. M. R. 2021, *AJ*, **162**, 237
- Ih, J., Kempton, E. M. R., Whittaker, E. A., & Lessard, M. 2023, *ApJ*, **952**, L4
- Ingalls, J. G., Krick, J. E., Carey, S. J., et al. 2012, *SPIE Conf. Ser.*, **8442**, 84421Y
- Jensen, C. S., Kjærulff, U., & Kong, A. 1995, *Int. J. Human-Comp. Studies*, **42**, 647
- Joshi, M., Haberle, R., & Reynolds, R. 1997, *Icarus*, **129**, 450
- Kite, E. S., & Barnett, M. N. 2020, *Proc. Natl. Acad. Sci.*, **117**, 18264
- Knutson, H. A., Charbonneau, D., Allen, L. E., Burrows, A., & Megeath, S. T. 2008, *ApJ*, **673**, 526
- Koll, D. D. B. 2022, *ApJ*, **924**, 134
- Koll, D. D. B., Malik, M., Mansfield, M., et al. 2019, *ApJ*, **886**, 140
- Kreidberg, L., Koll, D. D. B., Morley, C., et al. 2019, *Nature*, **573**, 87
- Krissansen-Totton, J., Wogan, N., Thompson, M., & Fortney, J. J. 2024, *Nat. Commun.*, **15**, 8374
- Kumar, R., Carroll, C., Hartikainen, A., & Martin, O. 2019, *J. Open Source Softw.*, **4**, 1143
- Luger, R., & Barnes, R. 2015, *Astrobiology*, **15**, 119
- Mahajan, A. S., Eastman, J. D., & Kirk, J. 2024, *ApJ*, **963**, L37
- Malik, M., Groscheintz, L., Mendonça, J. M., et al. 2017, *AJ*, **153**, 56
- Malik, M., Kempton, E. M. R., Koll, D. D. B., et al. 2019a, *ApJ*, **886**, 142
- Malik, M., Kitzmann, D., Mendonça, J. M., et al. 2019b, *AJ*, **157**, 170
- Mandel, K., & Agol, E. 2002, *ApJ*, **580**, L171

- Mansfield, M., Kite, E. S., Hu, R., et al. 2019, *ApJ*, **886**, 141
- Meier Valdés, E. A., Demory, B.-O., Diamond-Lowe, H., et al. 2025, *A&A*, **698**, A68
- Ment, K., Dittmann, J. A., Astudillo-Defru, N., et al. 2019, *AJ*, **157**, 32
- Moore, A. C., Ninkov, Z., & Forrest, W. J. 2006, *Opt. Eng.*, **45**, 076402
- Morello, G. 2015, *ApJ*, **808**, 56
- Morello, G., Waldmann, I. P., & Tinetti, G. 2016, *ApJ*, **820**, 86
- Morello, G., Waldmann, I. P., Tinetti, G., et al. 2014, *ApJ*, **786**, 22
- Morrison, J. E., Dicken, D., Argyriou, I., et al. 2023, *PASP*, **135**, 075004
- Neal, R. M. 2003, *Ann. Stat.*, **31**, 705
- The pandas development team 2020, <https://doi.org/10.5281/zenodo.3509134>
- Pass, E. K., Charbonneau, D., & Vanderburg, A. 2025, *ApJ*, accepted [arXiv:2504.01182]
- Patel, J. A., Brandeker, A., Kitzmann, D., et al. 2024, *A&A*, **690**, A159
- Pérez, F., & Granger, B. E. 2007, *Comp. Sci. Eng.*, **9**, 21
- Perrin, M. D., Sivaramakrishnan, A., Lajoie, C.-P., et al. 2014, *SPIE Conf. Ser.*, **9143**, 91433X
- Rackham, B. V., Apai, D., & Giampapa, M. S. 2018, *ApJ*, **853**, 122
- Rakitsch, B., Lippert, C., Borgwardt, K., & Stegle, O. 2013, in *Advances in Neural Information Processing Systems* (USA: NIPS), Vol. 26
- Redfield, S., Batalha, N., Benneke, B., et al. 2024, arXiv e-prints [arXiv:2404.02932]
- Rigby, J., Perrin, M., McElwain, M., et al. 2023, *PASP*, **135**, 048001
- Rothman, L. S., Gordon, I. E., Barber, R. J., et al. 2010, *J. Quant. Spec. Radiat. Transf.*, **111**, 2139
- Salvatier, J., Wiecki, T. V., & Fonnesbeck, C. 2016, *PeerJ Comp. Sci.*, **2**, e55
- Seligman, D. Z., Feinstein, A. D., Lai, D., et al. 2024, *ApJ*, **961**, 22
- Selsis, F., Wordsworth, R. D., & Forget, F. 2011, *A&A*, **532**, A1
- Shields, A. L., Ballard, S., & Johnson, J. A. 2016, *Phys. Rep.*, **663**, 1
- Sneep, M., & Ubachs, W. 2005, *J. Quant. Spec. Radiat. Transf.*, **92**, 293
- Spinelli, R., Borsa, F., Ghirlanda, G., et al. 2019, *A&A*, **627**, A144
- Stevenson, K. B., Harrington, J., Fortney, J. J., et al. 2012, *ApJ*, **754**, 136
- Thalman, R., Zarzana, K., Tolbert, M., & Volkamer, R. 2014, *J. Quant. Spec. Radiat. Transf.*, **147**, 171
- Tian, M., & Heng, K. 2024, *ApJ*, **963**, 157
- Tulasi Ram, S., Veenadhari, B., Dimri, A. P., et al. 2024, *Space Weather*, **22**, 2024SW004126
- Underwood, D. S., Tennyson, J., Yurchenko, S. N., et al. 2016, *MNRAS*, **459**, 3890
- Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, *Nat. Methods*, **17**, 261
- Waldmann, I. P. 2012, *ApJ*, **747**, 12
- Waldmann, I. P., Tinetti, G., Deroo, P., et al. 2013, *ApJ*, **766**, 7
- Wang, D., Hogg, D. W., Foreman-Mackey, D., & Schölkopf, B. 2016, *PASP*, **128**, 094503
- Weiner Mansfield, M., Xue, Q., Zhang, M., et al. 2024, *ApJ*, **975**, L22
- Whittaker, E. A., Malik, M., Ih, J., et al. 2022, *AJ*, **164**, 258
- Wohlfarth, K., Wöhler, C., Hiesinger, H., & Helbert, J. 2023, *A&A*, **674**, A69
- Wright, G. S., Wright, D., Goodson, G. B., et al. 2015, *PASA*, **127**, 595
- Zahnle, K. J., & Catling, D. C. 2017, *ApJ*, **843**, 122
- Zieba, S., Hu, R., Kreidberg, L., et al. 2023a, *The Search for Regolith on the Airless Exoplanet LHS 3844 b* (USA: JWST Proposal. Cycle 2), ID. #4008
- Zieba, S., Kreidberg, L., Ducrot, E., et al. 2023b, *Nature*, **620**, 746

-
- ¹ School of Physics, Trinity College Dublin, University of Dublin, Dublin 2, Ireland
 - ² Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, 21218, MD, USA
 - ³ Department of Space Research and Space Technology, Technical University of Denmark, Elektrovej 328, 2800 Kgs. Lyngby, Denmark
 - ⁴ Department of Physics and Astronomy, University of Southampton, Highfield, Southampton SO17 1BJ, UK
 - ⁵ School of Ocean and Earth Science, University of Southampton, Southampton, SO14 3ZH, UK
 - ⁶ Space Research and Planetary Sciences, Physics Institute, University of Bern, Gesellschaftsstrasse 6, 3012 Bern, Switzerland
 - ⁷ Department of Physics and Astronomy, Johns Hopkins University, 3400 N. Charles Street, Baltimore, MD 21218, USA
 - ⁸ Department of Physics, University of Oxford, Keble Road, Oxford, OX1 3RH, UK
 - ⁹ Centre for Space and Habitability, University of Bern, Gesellschaftsstrasse 6, 3012 Bern, Switzerland
 - ¹⁰ ARTORG Center for Biomedical Engineering Research, University of Bern, Murtenstrasse 50, 3008, Bern, Switzerland
 - ¹¹ Ludwig Maximilian University, Faculty of Physics, Scheinerstr. 1, Munich 81679, Germany
 - ¹² University College London, Department of Physics & Astronomy, Gower St, London, WC1E 6BT, UK
 - ¹³ University of Warwick, Department of Physics, Astronomy & Astrophysics Group, Coventry CV4 7AL, UK

Appendix A: Cosmic ray persistence effect

To verify that the persistence effect seen in Figure 1 was caused by a cosmic ray, we took the uncalibrated ‘uncal’ files for the second eclipse observation and calculated the change in data number between the groups before and after the cosmic ray hit. This occurred between groups 7 and 8 of the 524th integration. To correct for the expected increase in data number due to flux from the target star, thermal background, etc. we subtracted the median frame of all integrations divided by the number of groups. This procedure had minimal effect as this expected increase in data number was far smaller (of the order of hundreds of data numbers) than the size of the jump from the cosmic ray ($\approx 16,000$ data numbers). In Figure A.1 we overlay this increase in data number between these two groups with the pixel light curves centred around this cosmic ray event. These pixel light curves are from the reduction described in Section 4.1 which used the jump step correction and also performed an additional custom median outlier clipping step on the stage 1 and 2 processed data. These steps remove the significant jump in flux observed in many of these light curves in the integration containing the cosmic ray hit, but they do not perform any correction for this longer duration persistence artifact.

The large jump in data number of $\approx 16,000$ DN seen in Figure A.1 is larger than any other cosmic ray which hit within the 5px aperture during our three eclipse observations, suggesting these events are relatively rare. Examining pixels outside this aperture region which experience large (e.g. $>5,000$ DN) cosmic ray jumps also often correspond to similar but typically smaller persistence artifacts. To get a measure of the frequency of these events, we took the uncal files for all 22 observations in the Hot Rocks Survey and performed a similar test comparing the difference between successive groups across the full subarray (excluding the outermost rows and columns to avoid possible edge effects). Note this approach ignores cosmic rays which hit the first or last group which could also cause persistence effects. We found that jumps $>5,000$ DN occur at an average rate of ~ 0.0027

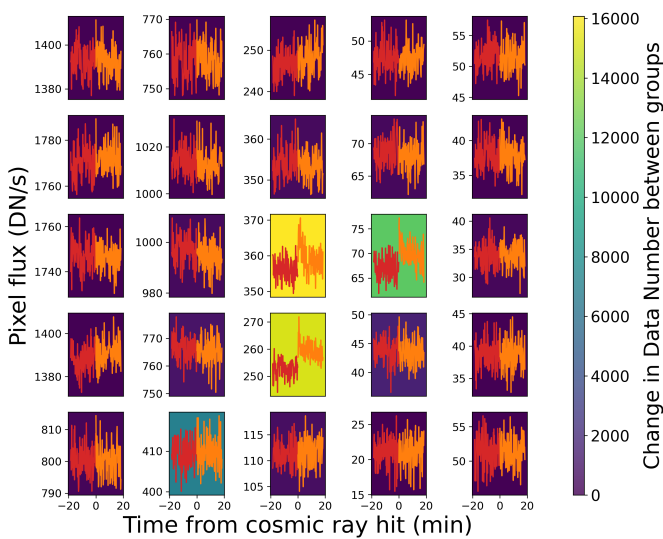


Fig. A.1. Pixel light curves centred on the pixels hit by a strong cosmic ray in the second eclipse. Light curves coloured in red before cosmic ray hit and in orange after. The background colour gives the change in data number between the groups before and after the cosmic ray hit for that pixel (after correcting for expected flux). Large data number jumps correspond to pixel light curves with strong persistence effects.

hits/(pixel hr) across each observation (excluding the eclipse of GJ-357b which was an outlier). Within a 5px radius aperture, we should expect jumps $>5,000$ DN every 4-7 hours with the rate varying between the observations, while hits $>16,000$ DN only occur every ~ 70 hours. Therefore, smaller cosmic ray hits that could cause persistence effects may affect every observation or two, while effects as strong as Figure A.1 should be quite rare. However, one exception is the eclipse of GJ 357b (Zraggen et al., in prep), which was a huge outlier and for which hits $>5,000$ DN occurred every ~ 13 minutes within a 5px radius aperture. This can likely be attributed to the observation being taken on 11th May 2024, during one of the most intense Solar storms recorded in recent times (Tulasi Ram et al. 2024).

In summary, persistence effects due to cosmic rays hitting individual pixels may be a widespread phenomenon for which the pixel-fitting method could be beneficial for weighting away from them. If these persistence effects could be modelled then it may be possible to directly fit them in the mean function of the pixel-fitting method (e.g. fitting an exponential ramp to the affected pixels) rather than weight away from them, potentially improving S/N. This could be more important if a strong cosmic ray were to affect one of the central pixels of the PSF or in the case of a high rate of cosmic ray hits as in the eclipse of GJ-357b.

Appendix B: Effect of intrapixel sensitivity variations and flat fielding errors

As described in Section 2.1, there are multiple reasons why intrapixel sensitivity should be significantly reduced in MIRI time series compared to Spitzer. To confirm that it can be assumed to be negligible for MIRI, we performed tests to examine to what level intrapixel sensitivity could affect our observations. A related issue is flat fielding, while we did run the FLAT-FIELD step in the JWST pipeline for the pixel-fitting analyses, any inaccuracies in the flat fielding would result in uncorrected for variations in sensitivity between different pixels (in contrast to intrapixel sensitivity which describes sensitivity variations within a pixel). We can examine both of these effects using simulated observations, similar to what has been performed for Spitzer (Morello 2015).

All of our tests follow the same procedure. First, we generate a possible pixel sensitivity map. As we were unable to find any measurements of intrapixel sensitivity for MIRI in the literature, we tested various possibilities, shown in Figure B.1. We considered a "gradient" map in which the pixel surface smoothly varies in sensitivity with a constant gradient from a relative efficiency of 95% in the top left corner to 105% in the bottom right corner. We also considered a "quadratic" map where the efficiency varied as a function of the square of the distance from the centre of the pixel. The normalised map has an efficiency ranging from 103% at the pixel centre to 93% at the edges. Our 'random' sensitivity map was generated using a GP with a 2D squared-exponential kernel function with correlation a function of the x and y position within a pixel. The height scale for this GP produced sensitivity variations between 95% to 105% relative efficiency. Each map was generated as a grid of (250, 250) points within each pixel. We then copied this map for a (15, 15) grid of pixels, creating an overall (3750, 3750) sub-pixel resolution sensitivity map simulating 225 total pixels.

In addition to testing the scenario where all pixels share the same sensitivity map, we also generated maps which used randomised gradient maps, quadratic maps and random 2D GP maps which were independent for each pixel. These maps

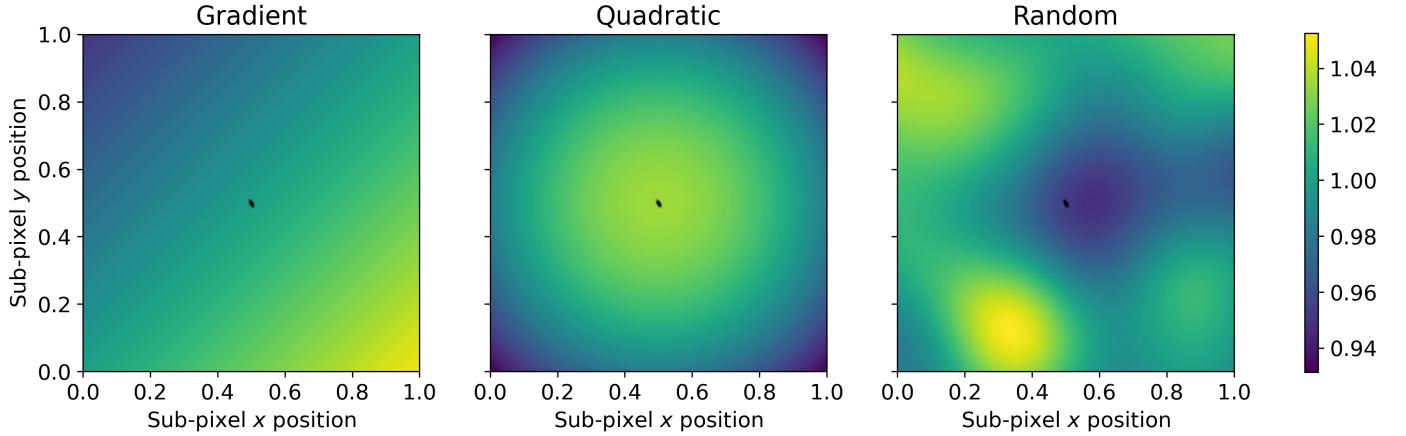


Fig. B.1. Various normalised sub-pixel sensitivity maps tested to examine intrapixel sensitivity. The colour scale gives the relative efficiency at each sub-pixel location. We visualise the level of pointing variation across an observation by plotting black dots for each centroid position in the first eclipse around the centre of each pixel map.

also exhibit $\sim 10\%$ variations in efficiency across each pixel. To test flat fielding errors, we generated (250, 250) sub-pixel maps which had constant efficiency across each surface, but we varied this constant for each pixel with a standard deviation of 1% and centred around 100%. The MIRI reference file for flat fielding exhibits $\sim 1\%$ level variations between pixels, so if we run the FLATFIELD step then we are modelling these maps being incorrect with a standard deviation of 1%.

After generating all of these possible sensitivity maps, we generated a symmetric 2D Gaussian profile with the same FWHM as the MIRI F1500W PSF (FWHM = 4.436 pixels; Dicken et al. 2024). We used a 2D Gaussian profile as it can easily be calculated to arbitrary sub-pixel resolution and roughly matches the inner core of the MIRI PSF. To get a realistic level of pointing stability, we used the x and y centroid values from the first observation, which were calculated by fitting a 2D Gaussian profile to the PSF for each integration (see Section 4.1.1 for more details). This likely overestimates the level of telescope pointing variation as this includes random scatter in fitting for the centroid position on top of the actual pointing instability. We discarded the centroid values from the first 45 minutes as these showed strong settling effects, likely arising from different pixels settling at different rates (see Section 4.5). For each of the centroid values, we shifted the location of our 2D Gaussian profile by these x and y values and then multiplied our shifted Gaussian profile by each sub-pixel sensitivity map. We also did this for a completely uniform sensitivity map for reference. We then binned up these sub-pixel resolution values to pixel resolution and performed aperture extraction with a 5px radius aperture and using the centroid values which were used to shift the Gaussian profile. The resulting aperture extracted light curves were then subtracted from the light curves generated using a uniform map, and the standard deviation of this difference was calculated. This gives us a measure of how much additional noise is produced from the non-uniform pixel surfaces. Since we did not simulate any noise, the aperture extracted light curve using the uniform map was almost completely noise-free (<0.1 ppm RMS). We also examined the difference made to each pixel light curve to understand how this may affect the pixel-fitting method. In this case, we subtracted the pixel-resolution light curves generated with each map from the values using the uniform map. Calculating the standard deviation for each pixel light curve then gave

Table B.1. Simulated noise introduced by various (sub)-pixel sensitivity maps for both aperture extraction and for the pixel light curve most affected.

Sensitivity map	Ap. Ext. (ppm)	Max pixel (ppm)
Flat field errors	2.43	245.1
Gradient (shared)	0.07	207.1
Quadratic (shared)	0.00	48.2
Random (shared)	0.02	100.6
Gradient (indep.)	0.26	115.5
Quadratic (indep.)	0.01	17.1
Random (indep.)	0.56	222.9

us a measure of how much additional noise would be present in each pixel light curve for each map.

Our results were that all of the intrapixel sensitivity maps and flat fielding errors were completely negligible, never producing greater than a couple ppm of additional noise in our aperture extracted light curves (see Table B.1). The pointing variation shows minimal correlation in time so most of this variation behaves like white noise and is tiny compared to the level of photometric noise in our observations (~ 800 ppm). We note that the noise would be added in quadrature to photon noise, so even 40 ppm of noise from intrapixel sensitivity would only increase the total aperture extracted noise by ~ 1 ppm in combination with 800 ppm of photon noise. These effects are therefore more than an order of magnitude below the level that would impact our observations, meaning that significantly more extreme sub-pixel sensitivity variations would be needed to have any noticeable effect. While the maximum effect in a particular pixel could be a couple hundred ppm (see the ‘Max pixel (ppm)’ column in Table B.1), this was largely only in the outer regions of the PSF with the lowest S/N and the photon noise in any individual pixel for the real observations is never better than 3000 ppm even in the central pixels, so this effect should also be negligible for the pixel-fitting method. The exact values in Table B.1 would vary if different random draws of sensitivity maps were generated, but as the effect is so far below the level of photon noise it appears unlikely this would ever have a significant impact.

It is worth repeating that the Si:As detector technology used for MIRI is expected to show very minor intrapixel sensitivity, so extreme variations across each pixel surface is expected to be highly unlikely. This is similar to what was found with the Si:As

detectors on Spitzer, which did not display the same level of intrapixel sensitivity as the shorter wavelength 3.6 μ m and 4.5 μ m channels which used InSb detectors (Knutson et al. 2008). The narrower PSF widths from the shorter wavelength filters (e.g. F560W) may show slightly more variation, although rerunning the same tests with a reduced FWHM = 1.882px matching that of the F560W filter (and setting the aperture radius to match this FWHM), we still find that all maps produce < 20 ppm of noise. This demonstrates that the exquisite pointing stability of JWST is enough to result in minimal intrapixel sensitivity, even if the pixels do demonstrate 10% level variations in sensitivity across their surfaces. Therefore, it appears unlikely that intrapixel sensitivity or flat fielding errors have any significant effect on time series observations with MIRI/Imaging.

Appendix C: Eclipse timing and eccentricity constraints

Given that the time of primary transit is already tightly constrained by transit observations, and our observations are the first to detect the eclipse of LHS 1140c, this should permit a tight constraint on the eccentricity parameter $\sqrt{e} \cos \omega$. This is because the time gap from primary to secondary transit Δt (ignoring light delay) is strongly dependent on $\cos \omega$. This time gap is given in Alonso (2018) as:

$$\Delta t = P \left[\frac{1}{2} + \frac{\left(1 + \frac{1}{\sin^2 i}\right) e \cos \omega}{\pi} \right], \quad (\text{C.1})$$

where i is the orbital inclination ($89.80^{+0.14}_{-0.19}^\circ$ for our target; Cadieux et al. 2024a). A negative value of $\cos \omega$ therefore leads to the eclipse occurring earlier than expected (relative to a zero eccentricity orbit) while a positive value would result in a later than expected time.

We marginalise over $\sqrt{e} \sin \omega$ too, although our observations provide little improvement to the constraint on this value as it mostly affects the duration of the transit and eclipse and the existing transit data already constrains this very well. We also account for light delay in our eclipse model as this adds an approximate 22 second delay to the expected eclipse time.

The eccentricity is constrained to be closer to zero than the previous analysis from Cadieux et al. (2024a). This is highlighted by Figure C.1, which shows that $\sqrt{e} \cos \omega$ is constrained much tighter compared to the posterior from Cadieux et al. (2024a) (C. Cadieux, priv. comm.), while $\sqrt{e} \sin \omega$ is relatively unaffected. Note the posterior from Cadieux et al. (2024a) was used as the priors on $\sqrt{e} \cos \omega$ and $\sqrt{e} \sin \omega$ in our analysis as our observations alone would not put good constraints on $\sqrt{e} \sin \omega$, while removing the prior on $\sqrt{e} \cos \omega$ has negligible impact. While we have constrained the eccentricity to be very small ($e < 0.0172$ to 95% confidence), the observations are quite inconsistent with a zero eccentricity orbit - as can be seen in the scatter plot of $\sqrt{e} \sin \omega$ against $\sqrt{e} \cos \omega$. This is because all three eclipses favour the eclipse occurring slightly early - as shown in Appendix C.1 - resulting in a negative value of $\sqrt{e} \cos \omega$. By using Equation C.1 with our posterior values, we recover that the eclipse occurs 2.8 ± 0.9 minutes early relative to a circular orbit. Our very small eccentricity value is consistent with dynamical modelling which predicts a short circularisation timescale for both LHS 1140b and LHS 1140c (Gomes & Ferraz-Mello 2020).

⁶ And in principle on other stellar parameters such as the stellar density (e.g. see Mahajan et al. 2024), although we do not explore this within this work

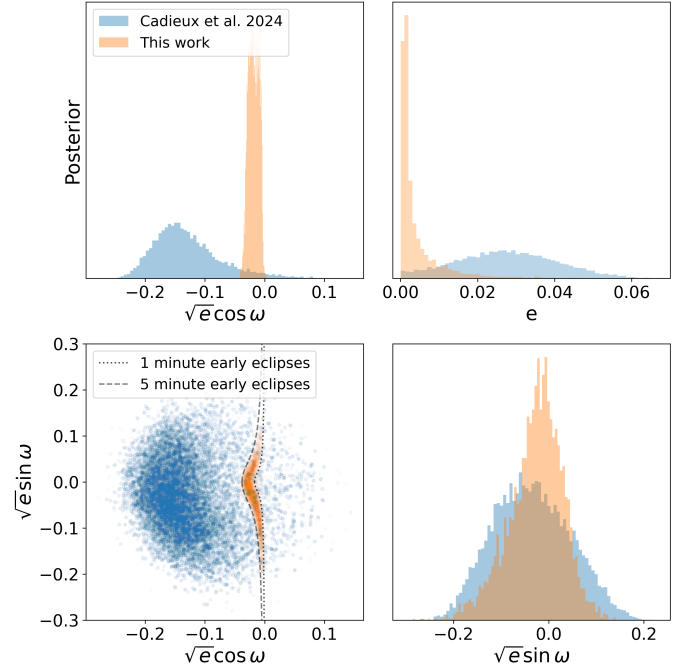


Fig. C.1. Comparison of eccentricity constraints between Cadieux et al. (2024a) (in blue; C. Cadieux, priv. comm.) and this work (in orange). The posterior from Cadieux et al. (2024a) was used as the prior for our eccentricity parameters. The additional constraint from the eclipse time significantly narrows the range of $\sqrt{e} \cos \omega$ (top left), which helps tighten the eccentricity constraint much closer to zero (top right). In the bottom left plot, we see that the new joint-posterior of $\sqrt{e} \cos \omega$ and $\sqrt{e} \sin \omega$ is largely constrained to lie between values which correspond to 1 to 5 minute early eclipses (relative to a circular orbit). The constraints on $\sqrt{e} \sin \omega$ are mostly unchanged from the prior values however (bottom right).

C.1. Individual eclipse time constraints

We also tested fitting each eclipse completely independently using aperture extraction with a GP, as opposed to Section 4.3.1 where we used eclipse times from the other two eclipses to constrain each eclipse time. Figure C.2 shows the constraints on the central eclipse times (given relative to a circular orbit) and eclipse depths for each of the three eclipses fit independently. We see that while each individual dataset does not strongly detect an eclipse, the highest probability central eclipse times all peak close to the expected time for a circular orbit and the eclipse depths are all closely matching. This consistency offers strong evidence that our observed eclipses are real and not the result of systematics.

In addition to testing a fully independent fit to each of the three datasets, we also tested a joint-fit with a shared eclipse depth but separate eclipse times (using aperture extraction with GPs). This could be thought of as a joint-fit which assumes a circular orbit but accounts for transit timing variations (TTVs), although these are probably unlikely given the large separation between the two known planets in the system. Unfortunately, due to the low S/N of each individual eclipse, the central eclipse time is not strongly constrained by each individual dataset and so a uniform prior was chosen bounded within 30 minutes either side of the expected eclipse time for a circular orbit. We found that all three eclipses were favoured to occur a couple of minutes before the expected time with zero eccentricity, which explains

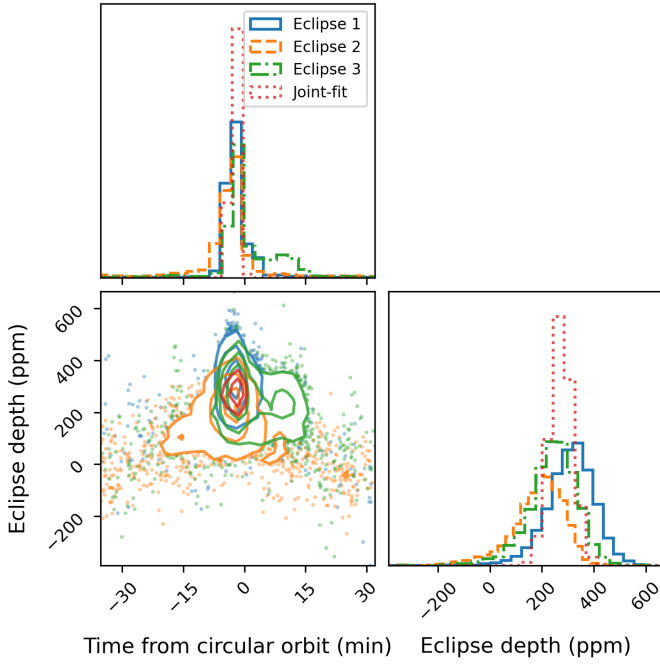


Fig. C.2. Posterior of central eclipse time (relative to expected time for a circular orbit) versus eclipse depth from individual eclipse fits and a joint-fit. Each dataset was fit using aperture extraction with a GP and the values from the joint-fit also using aperture extraction with a GP are overplotted in red. All individual fits are highly consistent with the joint-fit.

Table C.1. Central eclipse time constraints from each eclipse.

	T_{sec} (BJD TDB - 2460276)
Eclipse 1	$0.3716^{+0.0020}_{-0.0013}$
Eclipse 2	$223.2694^{+0.0013}_{-0.0017}$
Eclipse 3	$234.6048^{+0.0062}_{-0.0013}$

the negative $\sqrt{e} \cos \omega$ constraint when joint-fitting eccentricity parameters to all three datasets.

Table C.1 gives the central eclipse time constraints for each eclipse from this joint-fit without adjusting for light delay. Figure C.3 plots a histogram of the posterior for each of the central eclipse times. We note that Table C.1 approximates these posteriors using the median value and upper and lower Gaussian uncertainties. The accuracy of these approximations is visualised by overlaying the probability distributions they produce in orange over the true posteriors in blue. The posteriors deviate substantially from symmetric Gaussian distributions but the different upper and lower uncertainties appear to do a reasonable job at tracing the true posterior.

We note that the constraints in Table C.1 were used as priors when fitting individual eclipse depths as described in Section 4.3.1 and Appendix G. The way this was performed was by using the central eclipse time constraints from the two eclipses not being fit as priors on the expected eclipse time for the eclipse which was being fit (after adding/subtracting the correct number of periods between them). This helped constrain the individual eclipse fits which otherwise would have struggled to constrain the central eclipse time from a single dataset.

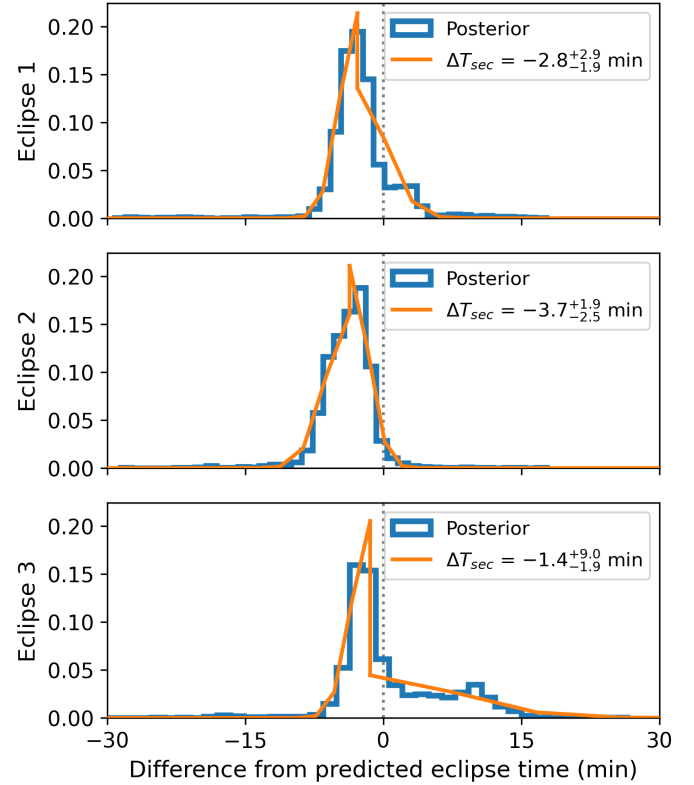


Fig. C.3. Central eclipse time posteriors from a joint-fit of all three eclipses using a shared eclipse depth but separate eclipse times. These posteriors were approximated using the median and 1σ upper and lower uncertainties (highlighted in orange) and used to constrain the eclipse times when fitting individual eclipses. The predicted eclipse time was calculated using values in Table D.1 but assumes a circular orbit and has been adjusted for light delay. All observations favour the eclipse to occur slightly early.

Table D.1. Priors used for eclipse parameters in the primary reduction analysis.

Parameter	Distribution	Prior
f_p/f_*	Uniform	$[-\infty, +\infty]$
T_0	Normal	$2460304.70775 \pm 0.00004$
P	Normal	3.777940 ± 0.000002
a/R_*	Normal	$27.1^{+0.2}_{-0.3}$
R_p/R_*	Normal	0.05312 ± 0.00028
b	Trunc. Normal	0.09 ± 0.06 & $b > 0$
$\sqrt{e} \cos \omega$	Normal	$-0.14^{+0.06}_{-0.04}$
$\sqrt{e} \sin \omega$	Normal	$-0.04^{+0.09}_{-0.08}$

Notes. Values taken from Cadieux et al. (2024b) for T_0 , a/R_* , R_p/R_* , b and Cadieux et al. (2024a) for P , $\sqrt{e} \cos \omega$ and $\sqrt{e} \sin \omega$.

Appendix D: Tables of priors for both reduction analyses

The choice of priors chosen for various parameters in the primary and secondary reductions analyses are included in Tables D.1 and D.2.

Table D.2. Priors used for the secondary analysis fit.

Parameter	Distribution	Prior
f_p/f_*	Uniform	$[-0.01, 0.01]$
T_0	Normal	$(2460304.70775, 0.00004)$
P	Fixed	3.777940
a/R_*	Normal	$(27.1, 0.1)$
R_p/R_*	Fixed	0.05312
b	Trunc. Normal	$(0.09, 0.1)$ & $b > 0$
$\sqrt{e} \cos \omega$	Normal	$(-0.14, 0.1)$
$\sqrt{e} \sin \omega$	Normal	$(-0.04, 0.1)$
m_{flux}^a	Normal	$(0.0, 0.1)$
σ^b	LogUniform	$(10, 10000)$
Linear (L)		
θ	Uniform	$[-100, 100]$
Exponential (E)		
A	Uniform	$[-1, 1]$
τ	Uniform	$[-1, 1]$
Gaussian Process (GP)		
GP_ξ	LogUniform	$(10^{-8}, 10^{-2})$
GP_ρ	LogUniform	$(10^{-8}, 10^2)$

^a Out-of-transit flux.^b White noise (ppm).

Notes. Values taken from Cadieux et al. (2024b) for T_0 , a/R_* , R_p/R_* , b and Cadieux et al. (2024a) for P , $\sqrt{e} \cos \omega$ and $\sqrt{e} \sin \omega$. We indicate the priors and distributions for the parameters of the detrending methods. Each parameter is fitted independently for each observation.

Appendix E: Atmospheres containing both SO₂ and CO₂

In addition to the models described in Section 5, we tested various concentrations on SO₂ in a CO₂ background. These are

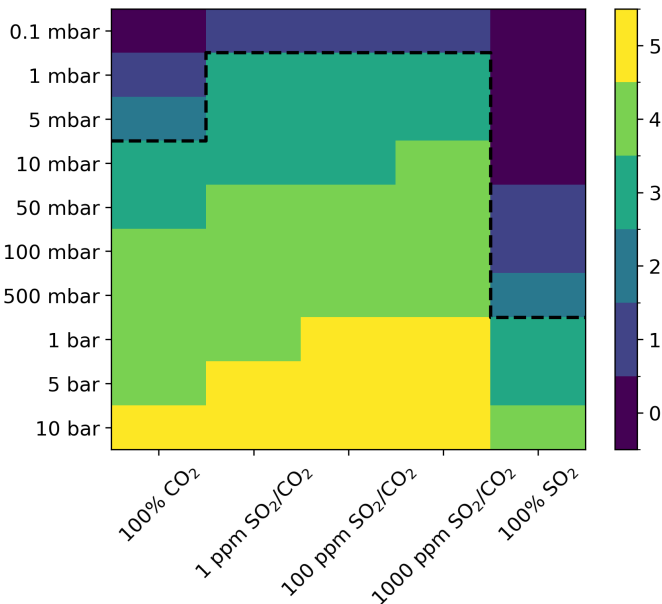


Fig. E.1. Minimum significance different atmospheres are ruled out to based on surface pressure and composition. Pressures below the black dotted line are all ruled out to at least 3σ . These models suggest various combinations of SO₂ and CO₂ are ruled out down to pressures of only 1 mbar, stronger than the results for either pure CO₂ or pure SO₂.

meant to represent possible secondary atmospheres which may arise from volcanic outgassing. We refer back to the limitations of these models discussed in Section 5 and we recommend that future modelling work be conducted to understand how photochemistry, atmospheric escape and realistic levels of outgassing may affect physically plausible secondary atmospheres.

Our models suggest that the combination of SO₂ at concentrations between 1 ppm - 1000 ppm in a CO₂ background are ruled out to very low pressures of just 1 mbar to at least 3σ . Figure E.1 summarises these constraints and highlights which compositions and pressures are ruled out at $\geq 3\sigma$. We note that these models all include heat redistribution as modelled by the analytic parameterisation in Koll (2022). These results are intriguing as they suggest that eclipse photometry at 15 μm may be particularly sensitive to volcanic atmospheres. The addition of small concentrations of SO₂ appears to make the CO₂ absorption feature more prominent in the model spectra, likely due to the effect of SO₂ on the temperature-pressure profile of the atmosphere.

Appendix F: Allan deviation plots

We took the last 1,024 integrations of each observation (cutting the first ~ 44 minutes) and subtracted our best-fit exponential ramp and linear model from our aperture extracted light curve fits (which were joint-fit using a GP). We then calculated the standard deviation of these residuals (referred to as the RMS) at different binnings increasing in powers of two - starting from not binning at all and up to 256 points per bin. We also performed the same operation for 10,000 independent draws of white noise with standard deviation scaled to match the real data and each with 1,024 points. This was in order to examine the expected level of deviation of uncorrelated white noise from the expected $1/\sqrt{N}$ for N binned points. We may expect sufficiently strong correlated noise to cause the RMS value to increase faster than $1/\sqrt{N}$ as N increases.

We see that for all three observations our Allan deviation plots do not increase any faster than $1/\sqrt{N}$. We do see some deviation less than $1/\sqrt{N}$, although this is not indicative of correlated noise and it does not deviate beyond the 3σ boundaries of the simulated white noise draws. It is however possible that this effect is related to an imperfect best-fit model subtracted from the data i.e. the exponential ramp and linear slope fit is off, resulting in some anti-correlation between datapoints at either end of the dataset.

We note that while our residuals are consistent with white noise according to this Allan deviation test, this should not be used to assume that these observations lack any correlated noise. For example, we can perform a similar test for each white noise draw but adding time-correlated systematics from an exponential kernel with a timescale of 30 minutes and amplitude of 50ppm. In this case, we find that only $\sim 3\%$ of these simulations will increase above $1/\sqrt{N}$ by $>3\sigma$ for at least one choice of binning. This is similar to the $\sim 1\%$ of cases that the white noise draws will deviate by $>3\sigma$ for at least one choice of binning (as multiple binnings are being tested for each draw this is larger than the expected $\sim 0.14\%$ of deviations if testing at any specific binning). While we could use a lower threshold of 2σ for detecting correlated noise, note that $\sim 12\%$ of white noise draws will increase more than expected by $>2\sigma$ for at least one binning, so this would result in many false positive detections of correlated noise. Also we could increase the number of binnings tested to increase the detection rate of correlated noise, but this would also likely increase the rate of false positives for uncorrelated noise.

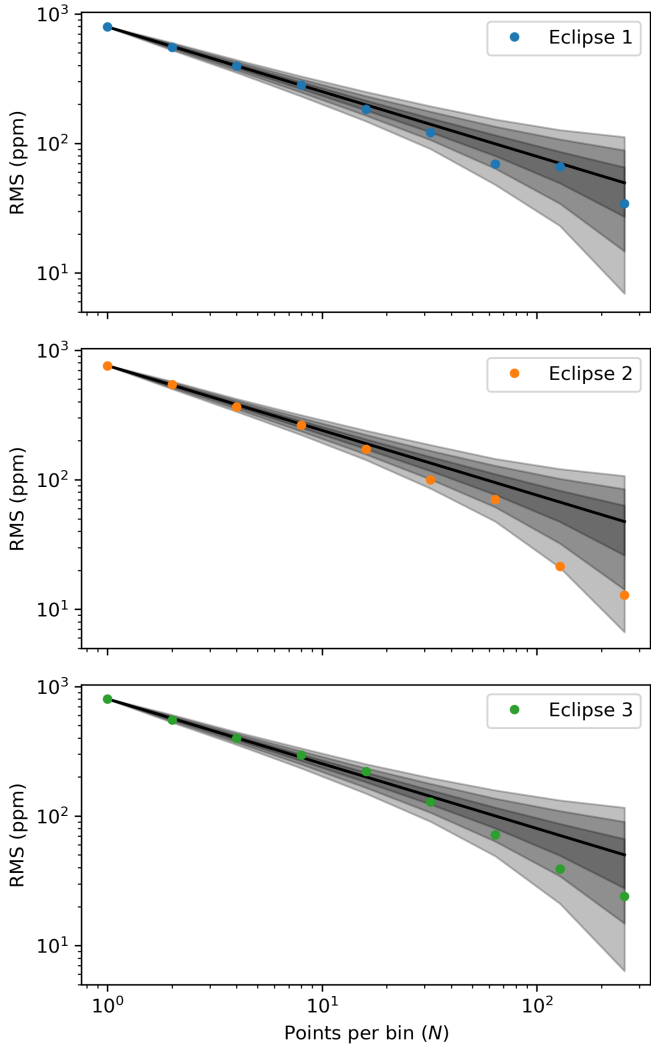


Fig. F.1. Allan deviation plot for each eclipse. The black line denotes the $1/\sqrt{N}$ trend expected for uncorrelated noise, while the shaded regions denote the 1σ , 2σ and 3σ confidence intervals for each particular binning calculated using 10,000 white noise draws. We note that only 75% of the white noise draws stay within the 2σ intervals for all bin sizes tested. We see that there can be a particularly large range of deviation from $1/\sqrt{N}$ for large bin sizes even for actual white noise.

As the amplitude of correlated noise increases, the detection rate at $>3\sigma$ will increase. For 100 ppm of correlated noise, we detect these systematics $\sim 22\%$ of the time. 150 ppm of correlated noise is detected $\sim 54\%$ of the time and 200 ppm detected 78% of the time. Therefore, quite significant levels of correlated noise can be added and will still frequently follow the $1/\sqrt{N}$ trend for uncorrelated noise. This means that it is always important to experiment with methods which can account for the effect of correlated noise when the signal being measured is so small (~ 250 ppm).

Appendix G: Individual eclipse depths for different pixel-fits

We tested a number of individual eclipse pixel-fits including varying some of the choices of the method. These are listed in Table G.1. For example, the PSF is modelled using the mean

function and is always allowed to interpolate to shift its x and y position for each integration, but the PSF was not allowed to change width in previous fits. We tested allowing the PSF to change width by linearly scaling it along the x and y -axis for each integration, introducing two additional parameters to be fit per integration. We found this had negligible impact on all three eclipses, although we just tested it with only flux-conserved systematics accounted for. The constraints appeared to be slightly tighter from this run, possibly because the PSF model is fitting the data slightly more accurately.

While running pixel-fitting including the various components of pixel systematics as chosen in Section 3 and 4 such as including common systematics and independent pixel systematics, we also tested including more than just the 81 pixels which are 5px away from the centre of the PSF. We chose a larger radius of 8.5px to examine how this affects the results. Similar to optimal extraction, the method should be able to weight away from lower flux pixels so adding more pixels should always either tighten the constraints or leave them unchanged. We found that increasing from 5px away to 8.5px away (225 total pixels) had a negligible impact on the result other than slightly tightening the eclipse depth constraints. This was at a substantial computational cost as for each pixel we fit for the baseline flux F_{out} and slope in flux T_{grad} , white noise amplitude σ , and height scale of independent pixel systematics h_{IPS} (unless there is a shared height scale). Therefore this required fitting for more than 500 extra parameters and resulted in little change in the result. However, it is reassuring that our choice of pixels to include only had a small effect on the result. We note that for these pixel-fitting analyses a slightly larger convergence criterion of $\hat{r} < 1.02$ was used (instead of $\hat{r} < 1.01$ for the results presented in Section 4) in order to reduce computational costs.

G.1. Testing different reduction steps

We also tested varying individual reduction choices one-by-one to examine if any choice has a large effect. We applied aperture extraction with a GP to each individual eclipse in all cases. The ‘reference’ reduction is the primary reduction as described in Section 4.1. To reiterate it here, it uses the standard JWST calibration pipeline except it modifies the jump rejection threshold to 7σ instead of the default of 4σ , removes both the first and last group instead of just the last group and aperture extracts the light curve using a 5px radius circular aperture. For stage 2 we only skipped the photom step, although this should have no effect on the eclipse depth as this just linearly scales the light curve and we normalise the light curves which will cancel out this correction.

We tested varying the radius aperture between 4px and 6px instead of 5px but it had minimal effect on the eclipse depth as can be seen in Table G.2. We also examined just running stage 1 and 2 of the pipeline with its default settings, using the JUMP step with the 4σ default threshold or skipping the JUMP step entirely. In addition, we tested skipping the EMICORR step which corrects for electromagnetic interference in the raw data, skipping the flat-fielding step in stage 2 as well as including all groups (i.e. skipping the steps which exclude the first and last groups). We did not find any dramatic difference for any of the various reductions we tested which suggests our results are not highly sensitive to these reduction choices.

Appendix H: Autocorrelation of pixel light curves

We can autocorrelate the pixel light curves for each eclipse to empirically inform our choice of kernel function. In practice this

Table G.1. Eclipse depths for individual eclipse depth fits using different pixel-fitting assumptions with aperture extraction results included for reference.

Method	E1	E2	E3	Combined
Ap. Ext. w/o GP	327±76	211±66	278±71	267±41
Ap. Ext. w/ GP	327±82	215±75	272±79	268±45
Px-fit (only FCS systematics)	310±72	195±73	261±70	256±41
Px-fit (varying PSF width, only FCS systematics)	316±73	188±67	282±67	259±40
Px-fit (common systematics)	304±88	194±80	258±77	249±47
Px-fit (common+indep. sys.)	283±101	208±99	246±93	245±56
Px-fit (8.5px; common+indep. sys.)	289±97	203±93	243±89	243±54
Px-fit (common+shared indep. sys.)	294±95	204±85	257±78	249±49
Px-fit (8.5px; common+shared indep. sys.)	291±94	203±81	256±75	247±47

Table G.2. Eclipse depths for individual eclipse depth fits using aperture extraction with different reductions.

Method	E1	E2	E3	Combined
Reference	323±82	213±73	276±78	266±45
Reference, 4px rad.	329±83	213±83	270±80	271±47
Reference, 4.5px rad.	319±83	211±75	277±76	266±45
Reference, 5.5px rad.	332±82	219±73	282±75	273±44
Reference, 6.0px rad.	317±83	219±76	301±78	276±46
Default pipeline	315±83	216±74	262±76	261±45
Default jump step	323±84	217±73	276±76	267±45
Skip jump step	325±85	216±72	271±77	265±45
Skip emicorr step	323±83	211±76	272±75	266±45
Skip flat-fielding	318±83	214±73	275±77	265±45
Include all groups	328±79	207±72	264±74	263±43

is somewhat complicated to implement as our pixel time series can be thought of as a 3D data cube with two spatial dimensions and a time dimension (X, Y, t). Figure H.1 attempts to examine correlations in this space. In the centre of the (5, 5) subplot is the autocorrelation of each pixel light curve with itself as a function of separation in time, averaged together across all pixels included within the pixel fit (5px from the PSF centre) and averaged across all three eclipse datasets. This is plotted in blue when performed on the real JWST data. To examine our kernel function choice we also took our best-fit model from pixel-fitting the real dataset (accounting for both common systematics and independent pixel systematics) and generated random draws of noise for each of the three real datasets using the covariance matrix generated by our kernel function. We could then autocorrelate this noise in the same manner as for the real data which was then plotted in orange. Finally, instead of performing this once for each eclipse and averaging together, the autocorrelation plotted in green was recovered by generating noise 100 times for each eclipse dataset and averaging together the autocorrelation, giving a better idea for how the kernel function converges for a lot of synthetic observations.

For the plot immediately left or right of the centre plot, the autocorrelation was taken between horizontally adjacent pixels as a function of separation in time. Similarly the pixels above and below the central plot are for vertically adjacent pixels and so on for diagonally adjacent pixels. The outer plots are all for pixels that are separated with one pixel gap between them i.e. two pixels apart. We note that for our kernel choice, we expect neighbouring pixels to show an anti-correlation in time as we expect them to be flux-conserved. This is clear in the green light curves as there has been one hundred covariance matrix draws to average over. For both the blue and orange light curves they only have three eclipse datasets to average over so they may only show moderate evidence of an anti-correlation. However there

does appear to be some sign of this anticorrelation, particularly in the adjacent neighbouring pixels where the blue light curves show signs of the V-shape seen in the green light curves. The pixels separated by more than one pixel show such a weak correlation in all cases that there is no evidence of anything to adjust in our kernel function for pixels at this separation.

The central subplot shows that our simulated noise does not quite match the correct kernel function as the real autocorrelation in blue is slightly sharper than either of the simulated autocorrelation. As this is the autocorrelation in time of pixels with themselves, this plot is for zero pixel separation and therefore the autocorrelation should just trace the kernel function in time (note we have excluded the point at zero separation in pixels and time which trivially gives a correlation of one). Our choice of the exponential kernel is able to match a lot of the sharpness of the autocorrelation of the real data however and we can clearly see that a smoother kernel such as the squared exponential kernel would be a less accurate match.

Appendix I: Corner plots

We have included the corner plot featuring all parameters from the joint-fit of all three eclipses using aperture extraction with GPs. This is in Figure I.1. In addition, we have a corner plot from each of the individual pixel-fits applied to each eclipse in Figures I.2, I.3 and I.4. These are all using the method which included common systematics and a shared independent pixel systematics height scale for some pixels. As there are approximately 2,300 parameters in each of these fits, only a subset of the parameters were included, with none of the changes in PSF position included (two parameters per integration) and only the parameters for the central pixel of the PSF included for parameters which are fit individually for each pixel.

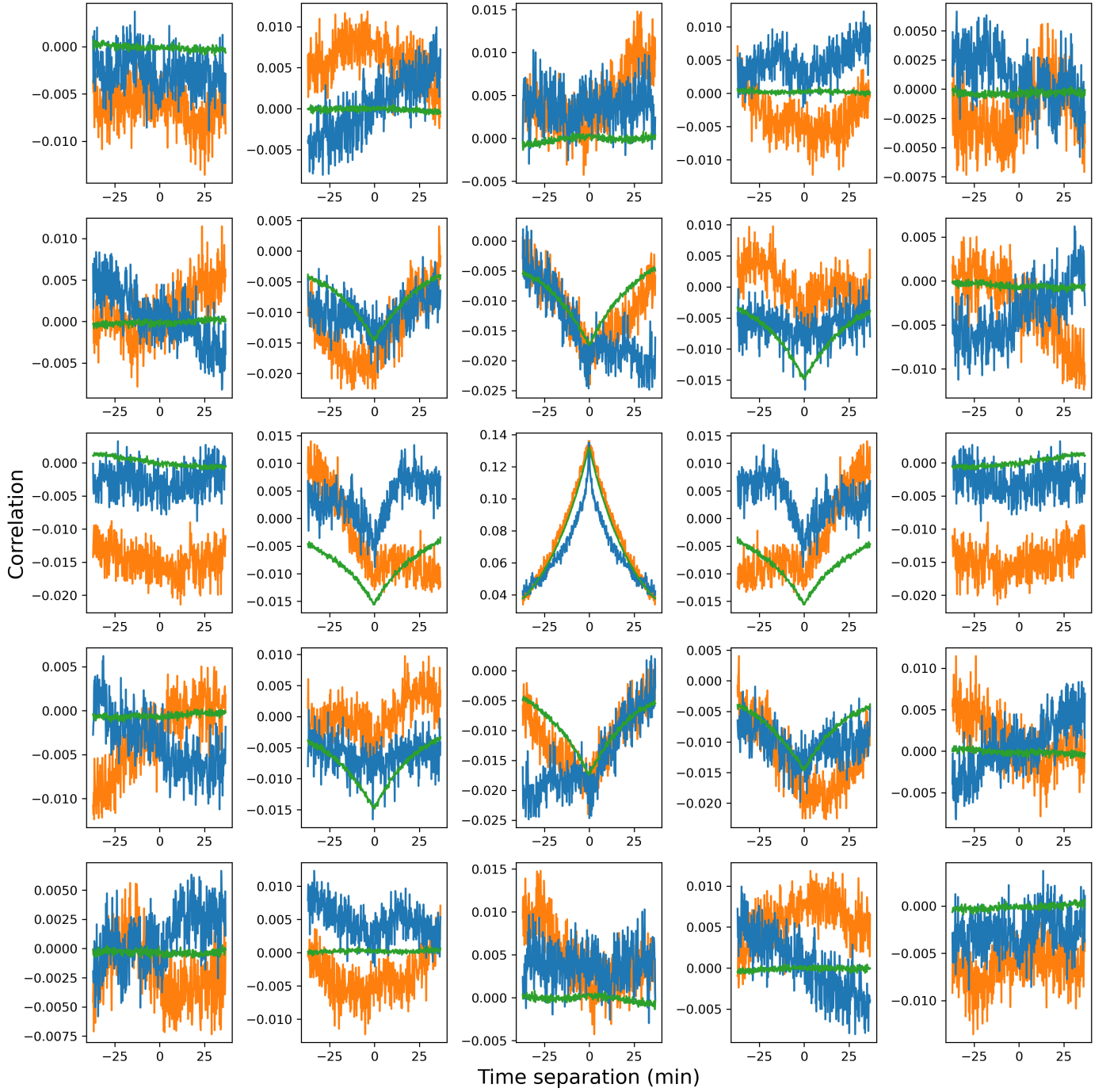


Fig. H.1. Autocorrelation of pixel light curves averaged across all the three eclipse datasets in blue, as well as autocorrelation of simulated pixel light curves for each eclipse in orange. The green is the averaged autocorrelation of 100 simulated eclipse datasets using our chosen kernel function. The autocorrelation of the real data is largely consistent with the general trend of the simulated data for three eclipses.

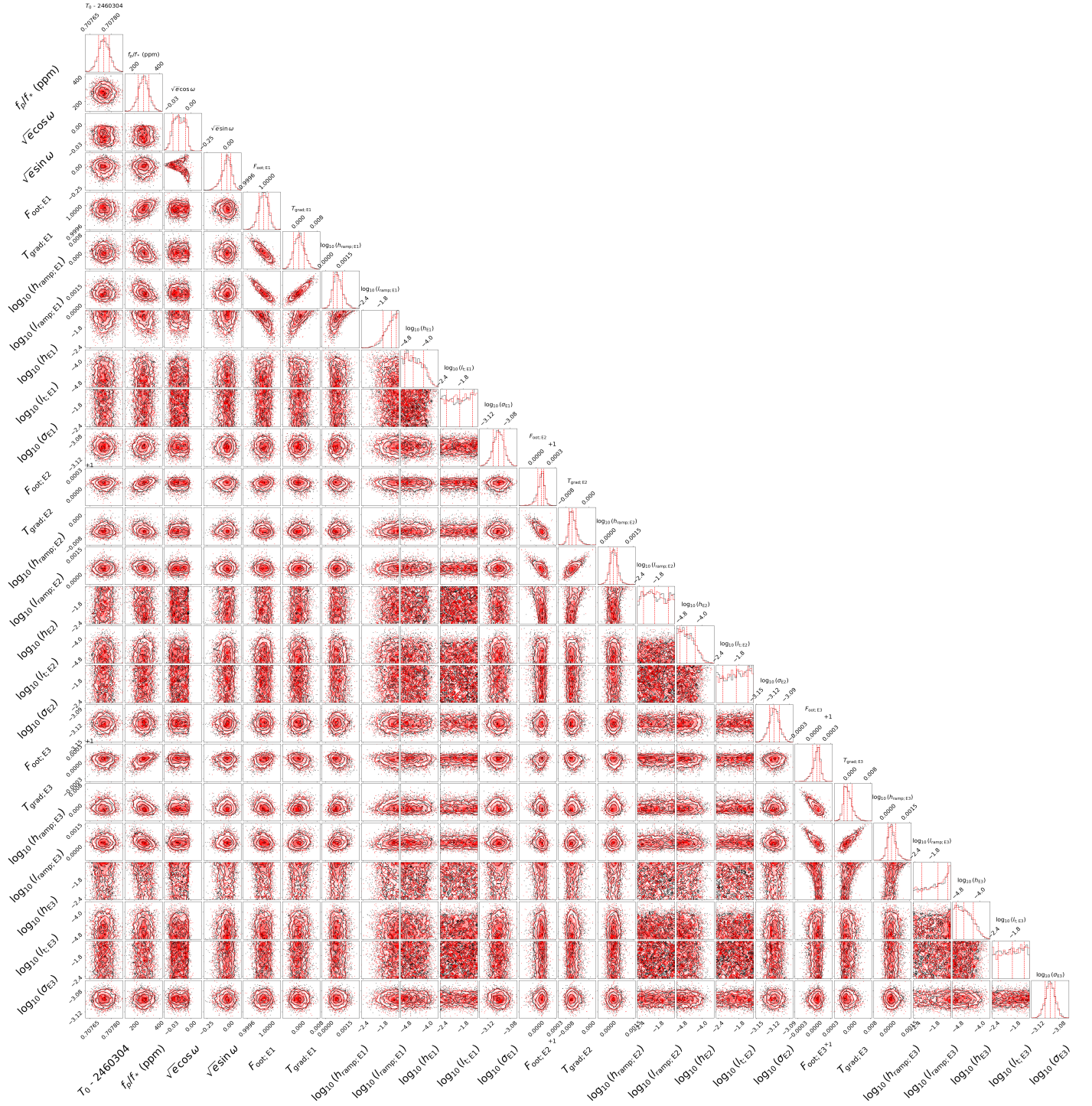


Fig. I.1. Corner plot of the joint-fit of aperture extracted light curves with GPs. The eclipse depth f_p/f_* can be seen to be slightly correlated with many of the parameters which describe the baseline flux or detector settling in each eclipse. The slope in baseline flux is also somewhat degenerate with the initial detector settling slope for each eclipse. For many of the hyperparameters a log-uniform prior was used and so the logarithm of these parameters is plotted. Note the unusual relationship between the two eccentricity parameters which is explained in Appendix C.

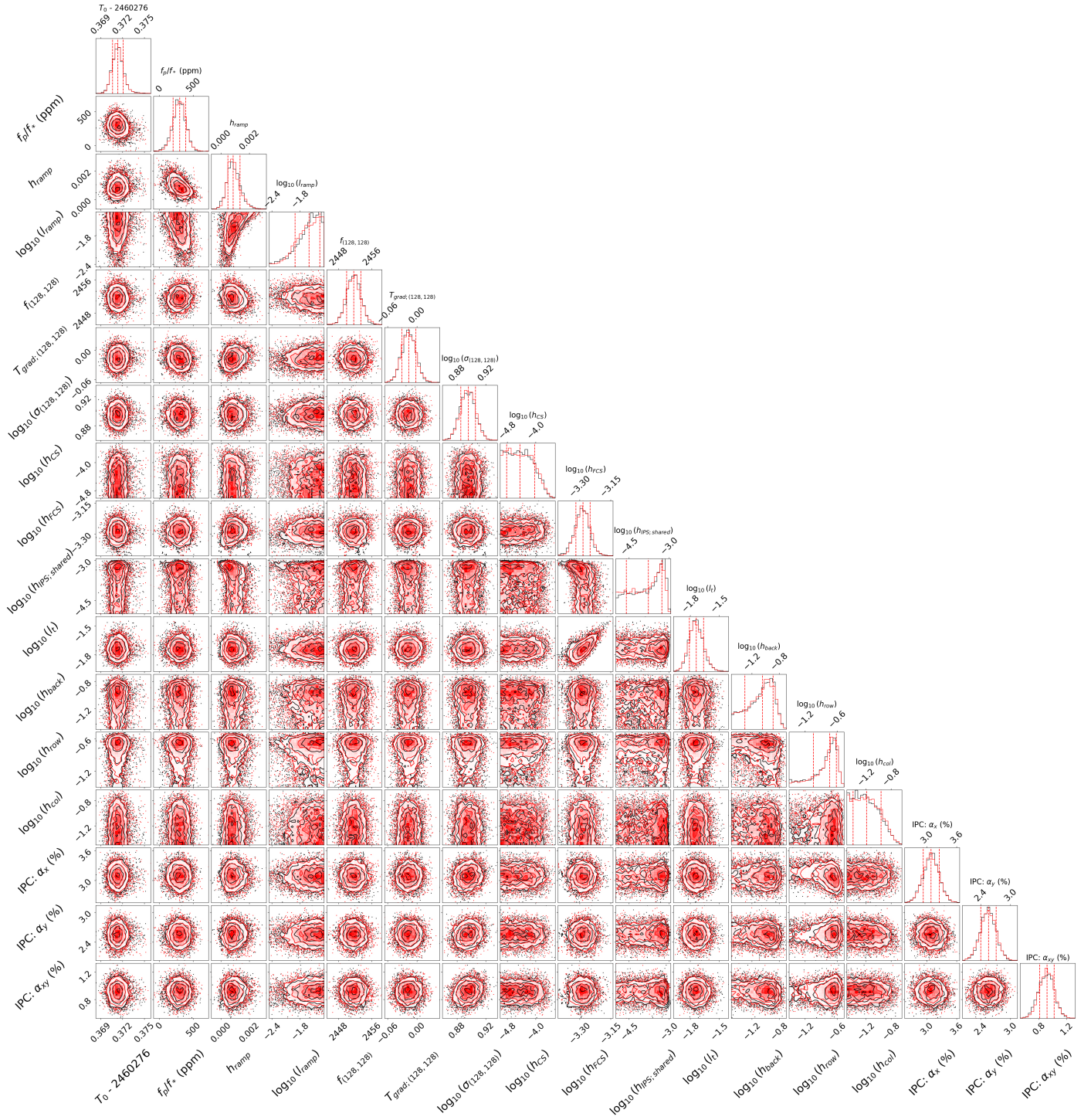


Fig. I.2. Corner plot of the first eclipse with the pixel-fitting method. Many parameters appear independent but we still find the eclipse depth is correlated with the detector settling parameters h_{ramp} and l_{ramp} , similar to Figure I.1. In addition, when the shared height scale of independent pixel systematics $h_{\text{IPS,shared}}$ is high, the height scale of flux-conserving systematics h_{FCS} is reduced. This makes sense as the "bumps" seen in each pixel light curve in Figure 1 could in principle be fit as either flux-conserved systematics or independent systematics, but the model does clearly prefer to fit a non-zero flux-conserving systematics height scale h_{FCS} while $h_{\text{IPS,shared}}$ is consistent with zero. This means the model favours that these "bumps" are flux-conserving and not independent.

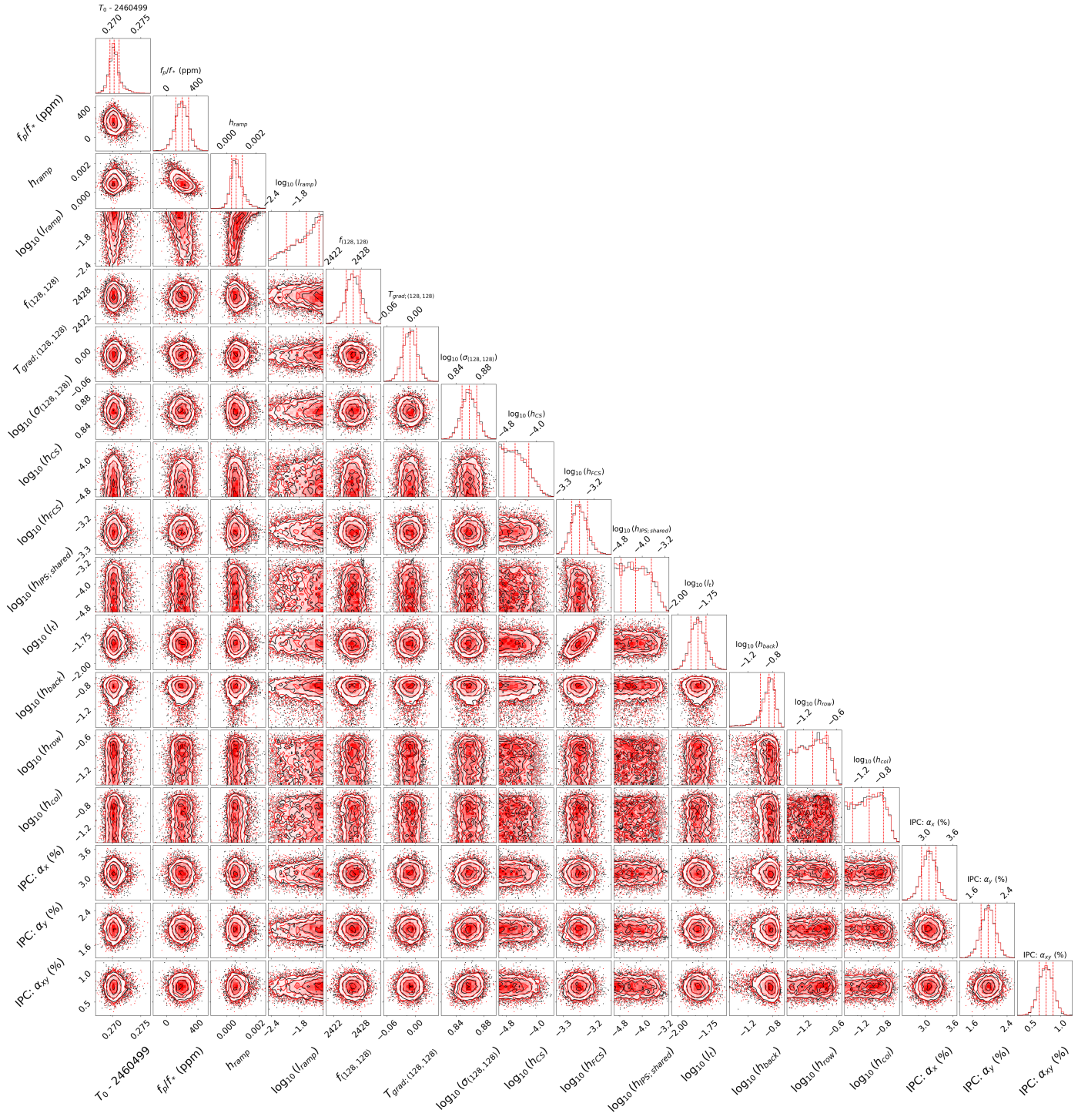


Fig. I.3. Same as Figure I.2 but for the second eclipse.

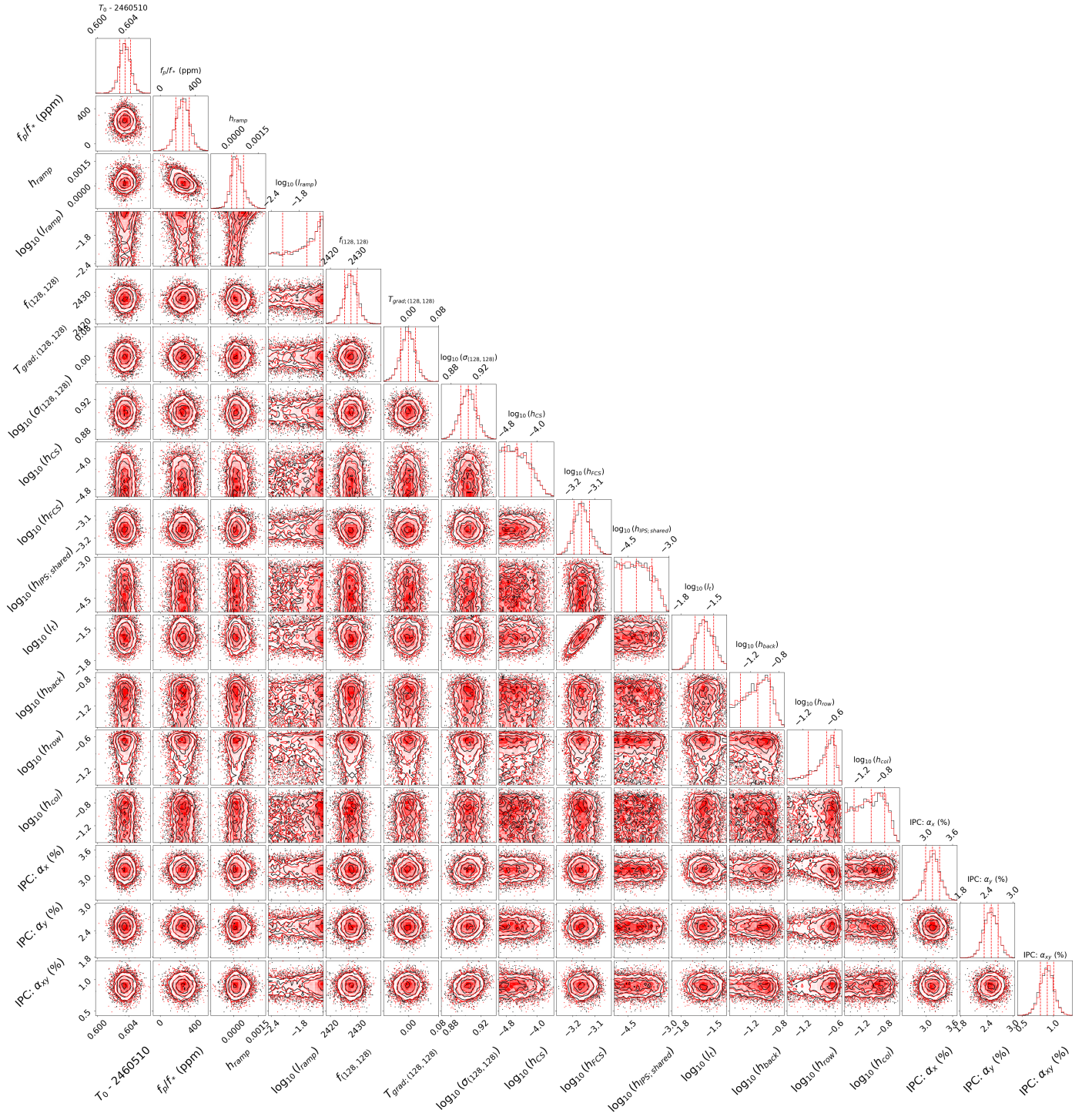


Fig. I.4. Same as Figure I.2 but for the third eclipse.