

Gemini/GMOS optical transmission spectroscopy of WASP-121b: signs of variability in an ultra-hot Jupiter?

Jamie Wilson,¹★ Neale P. Gibson^{1,2}, Joshua D. Lothringer,³ David K. Sing,^{3,4} Thomas Mikal-Evans^{1,5}, Ernst J. W. de Mooij,¹ Nikolay Nikolov^{1,6} and Chris A. Watson¹

¹*Astrophysics Research Centre, School of Mathematics and Physics, Queens University Belfast, Belfast BT7 1NN, UK*

²*School of Physics, Trinity College Dublin, The University of Dublin, Dublin 2, Ireland*

³*Department of Physics and Astronomy, Johns Hopkins University, Baltimore, MD 21210, USA*

⁴*Department of Earth and Planetary Sciences, Johns Hopkins University, Baltimore, MD 21210, USA*

⁵*Kavli Institute for Astrophysics and Space Research, Massachusetts Institute of Technology, Cambridge, MA 02139, USA*

⁶*Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, MD 21218, USA*

Accepted 2021 March 7. Received 2021 February 19; in original form 2020 December 18

ABSTRACT

We present ground-based, spectroscopic observations of two transits of the ultra-hot Jupiter WASP-121b covering the wavelength range $\approx 500\text{--}950$ nm using Gemini/GMOS. We use a Gaussian process framework to model instrumental systematics in the light curves, and also demonstrate the use of the more generalized Student's-T process to verify our results. We find that our measured transmission spectrum, whilst showing overall agreement, is slightly discrepant with results obtained using *HST*/STIS, particularly for wavelengths shortwards of ≈ 650 nm. In contrast to the STIS results, we find evidence for an increasing bluewards slope and little evidence for absorption from either TiO or VO in our retrieval, in agreement with a number of recent studies performed at high-resolution. We suggest that this might point to some other absorbers, particularly some combination of recently detected atomic metals, in addition to scattering by hazes, being responsible for the excess optical absorption and observed vertical thermal inversion. Our results are also broadly consistent with previous ground-based photometry and 3D GCM predictions, however, these assumed different chemistry to our retrievals. In addition, we show that the GMOS observations are repeatable over short periods (days), similar to the *HST*/STIS observations. Their difference over longer periods (months) could well be the result of temporal variability in the atmospheric properties (i.e. weather) as predicted by theoretical models of ultra-hot Jupiters; however, more mundane explanations such as instrumental systematics and stellar activity cannot be fully ruled out, and we encourage future observations to explore this possibility.

Key words: methods: data analysis – techniques: spectroscopic – stars: individual (WASP-121) – planetary systems.

1 INTRODUCTION

Close-in giant exoplanets are amongst the most promising targets for detailed atmospheric studies of planets outside our own Solar System, owing to their favourable combinations of short periods, extended atmospheres, and large planet-to-star radius ratios – all of which helps us to maximize the achievable signal to noise of what are very challenging measurements. The extreme temperatures experienced by the most highly irradiated hot Jupiters may also lead to a simplification of the atmospheric chemistry on these planets, due to the thermal dissociation of a large fraction of molecular species into their constituent elements – aiding in the interpretation of their atmospheric observations (Kitzmann et al. 2018; Lothringer, Barman & Koskinen 2018). The technique of transmission spectroscopy, which measures the variations in the effective size of the planet as a function of wavelength, is an enormously successful method for investigating the chemical composition and physical structure

of a planet's atmosphere (Seager & Sasselov 2000; Brown 2001). Transmission spectroscopy observations have provided many high-significance detections of atomic and molecular species for a range of exoplanets, including detections of Na, K, and H₂O both from the ground and using space-based instrumentation (e.g. Charbonneau et al. 2002; Deming et al. 2013; Kreidberg et al. 2014; Stevenson et al. 2014; Sing et al. 2015; Evans et al. 2016, 2018; Nikolov et al. 2016, 2018; Spake et al. 2018; Carter et al. 2020). However, other observations have resulted in partially muted or even completely featureless spectra due to the influence of clouds and/or hazes (e.g. Bean et al. 2011; Sing et al. 2011, 2016; Gibson et al. 2013b; Pont et al. 2013).

A number of recent studies have also focused on investigating the temperature structures of highly irradiated exoplanets and the presence of vertical thermal inversions (e.g. Knutson, Howard & Isaacson 2010; Madhusudhan et al. 2011; Mollière et al. 2015; Beatty et al. 2017; Lothringer et al. 2018; Parmentier et al. 2018; Gandhi & Madhusudhan 2019). These inversion layers occur high in the planet's atmosphere where the temperature is increasing with altitude and are present in the atmospheres of almost all of the Solar System planets (Gillett, Low & Stein 1969; Ridgway 1974; Wallace,

* E-mail: jwilson34@qub.ac.uk

Prather & Belton 1974). The Earth’s stratosphere for example is due to absorption of UV photons high in the atmosphere by ozone and hazes. For very highly irradiated exoplanets it has been predicted that the molecules titanium oxide (TiO) and/or vanadium oxide (VO) could provide both the required opacity to incoming optical radiation and the necessary abundances to be strong candidates for the drivers of these thermal inversions (Hubeny, Burrows & Sudarsky 2003; Fortney et al. 2008). Though thus far, and despite much effort, detections exist only for a limited number of planets (e.g. Nugroho et al. 2017; Sedaghati et al. 2017), and in most cases have been disputed (e.g. Espinoza et al. 2019; Herman et al. 2020). The presence of a thermal inversion can be inferred from planetary emission spectra by observing bands in emission rather than absorption, and this was first achieved for the hot Jupiter WASP-121b from observations of H₂O on the dayside (Evans et al. 2017), and has since been verified using TESS optical phase curves and other observations (Daylan et al. 2021; Bourrier et al. 2020b).

WASP-121b is a highly irradiated ultra-hot Jupiter discovered by Delrez et al. (2016). It orbits a bright F6-type star ($V = 10.4$) and has a highly inflated atmosphere, both of which make it particularly amenable to transmission spectroscopy observations. It has a mass similar to that of Jupiter and a radius significantly larger – $1.18 M_J$ and $1.7 R_J$, respectively, and an equilibrium temperature above 2400 K. WASP-121b is amongst the most extensively studied of the transiting planets with numerous observations having been acquired from the ground and from space and at both low-resolution and using high-resolution Doppler-resolved spectroscopy. At low-resolution, observations obtained with the *Hubble Space Telescope* (HST) have resulted in detections of H₂O along with excess optical absorption which was tentatively attributed to TiO/VO (Evans et al. 2016). Follow-up observations revealed further evidence for VO (Evans et al. 2018) along with an unknown blue absorber suggested to be attributed to the SH molecule. However, additional secondary eclipse measurements have not been able to confirm the previous VO detection (Mikal-Evans et al. 2019, 2020) and instead suggest H⁻ emission.

It has not yet been possible to definitively show that either VO or TiO are responsible for the thermal inversion in the atmosphere of WASP-121b. Additionally, Beatty et al. (2017) and Parmentier, Showman & Lian (2013) both show how cold-trap processes can interfere with the circulation of TiO/VO and suppress the formation of inversions, and it has also been demonstrated (Lothringer et al. 2018; Parmentier et al. 2018) that for the very hottest planets much of the TiO and VO will be thermally dissociated on the dayside and in these cases thermal inversions could instead be driven by NUV and optical absorption by gas phase metals such as Fe and Mg. Indeed, at high-resolution, the atmosphere of WASP-121b has been observed using both the HARPS (Bourrier et al. 2020a; Cabot et al. 2020) and UVES (Gibson et al. 2020; Merritt et al. 2020) spectrographs, which has resulted in significant detections of the atomic metals Fe I and Na I and absorption from H-alpha but non-detections of both TiO and VO. Fe I in particular is a strong optical absorber and its presence could therefore provide the source of the heating required to produce the observed temperature inversion (Gibson et al. 2020; Pino et al. 2020).

Even more recent observations at high-resolution have confirmed the previous metal detections and also uncovered evidence for multiple other atomic species including Mg, Na, Ca, Cr, Fe, Ni, V (Ben-Yami et al. 2020; Hoeijmakers et al. 2020) and a very recent detection of Li using ESPRESSO on the VLT (Borsa et al. 2020). In the UV, SWIFT/UVOT observations revealed tentative evidence of metal ions high in the atmosphere (Salz et al. 2019), whilst HST/STIS observations detected strong absorption lines from iron and magnesium atoms (Sing et al. 2019) extending far higher than

the optical and near-infrared features and demonstrating the existence of an extended and possibly escaping atmosphere. In this study, we report optical transmission spectroscopy observations of WASP-121b using the Gemini Multi-Object Spectrograph (GMOS). Our observations aim to search for evidence of absorption by TiO/VO or other metals in order to infer the main driver of the thermal inversion by resolving the shape of these molecular features.

This paper is structured as follows: we describe our observations and data reduction steps in Section 2 and detail our light-curve analysis in Section 3; in Section 4 we describe our atmospheric retrieval approach and discuss our results in Section 5. Finally, we offer our conclusions in Section 6.

2 GMOS OBSERVATIONS AND DATA REDUCTION

Two transits of the ultra-hot Jupiter WASP-121b were observed on the nights of 2017 January 4 and 9 (hereafter Transit 1 and 2, respectively) using the 8-m Gemini-South telescope with the Gemini Multi-Object Spectrograph (GMOS; Hook et al. 2004) as part of programme GS-2016B-Q-42 (PI: Gibson). GMOS consists of three $2k \times 4k$ CCDs arranged side by side and separated by small detector gaps and has an imaging field of view of 5.5×5.5 arcmin squared. Both transits were observed using the R400 grism + OG515 filter with a central wavelength of 725 nm in 2×2 binning mode covering the spectral range of 515–940 nm. To reduce readout time, we read out only three regions of interest (ROI)¹ on the chip including the target and two comparison stars. For Transit 1 we obtained 580 science exposures over a total period of 340 min, whilst for Transit 2 we obtained a total of 505 science exposures covering a total period of 290 min. For both transits we used exposure times of 10 s with a readout time of 20 s. For Transit 1 the on-sky separation of the Moon from the target was ≈ 97 deg with lunar illumination of ≈ 44 per cent. For Transit 2 the Moon was separated by ≈ 63 deg with illumination of 93 per cent.

To enable differential spectroscopy, we observed two comparison stars simultaneously with WASP-121 for each transit and used a custom mask with wide slit widths (40×15 arcsec) in order to reduce the impact of differential slit losses. Only the brighter of our two comparison stars was used in the subsequent analysis since the other was found to be significantly fainter. The average FWHM for Transit 1 was found to be ≈ 4 pixels but varied between a maximum of ≈ 7 pixels and a minimum of ≈ 3 pixels, whilst for Transit 2 the average was ≈ 5 pixels, also varying between a maximum of ≈ 8 pixels and a minimum of ≈ 4 pixels. The seeing-limited spectral resolution was found to be $R \approx 690$ –1500 and $R \approx 600$ –1200 for Transit 1 and 2, respectively, and we observed the target from an airmass of 1.51 to 1.01 for Transit 1 and 1.25 to 1.01 for Transit 2.

The standard GMOS pipeline contained in the Gemini IRAF²/PYRAF³ package was used to carry out the initial data reduction. We first converted the ROI images to standard GMOS format, and then proceeded with basic reduction procedures to de-bias and flat-field the raw images using the calibration frames obtained at the beginning and end of each night. To extract the spectra we used a custom pipeline in IRAF/PYRAF and experimented with various

¹To speed up the readout time only pixels containing flux from the slits were recorded into separate regions of interest (ROIs).

²IRAF is distributed by the National Optical Astronomy Observatory, which is operated by the Association of Universities for Research in Astronomy (AURA) under cooperative agreement with the National Science Foundation.

³PYRAF is a product of the Space Telescope Science Institute, which is operated by AURA for NASA.

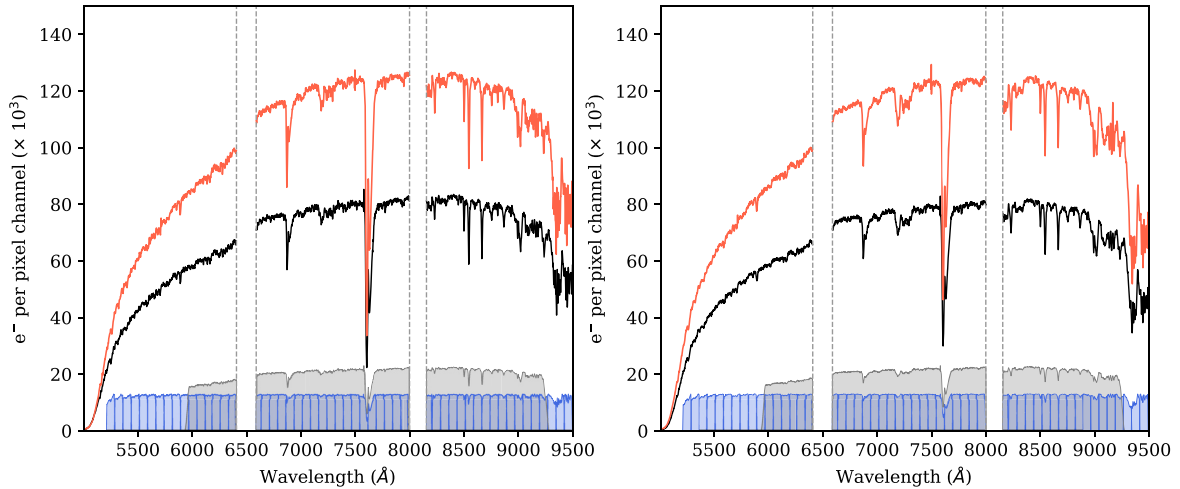


Figure 1. Example spectra of WASP-121 (black) and the comparison star (red) for Transit 1 (left-hand panel) and Transit 2 (right-hand panel). Coloured regions indicate the spectral bins used for extraction of the white light curves (grey) and spectroscopic light curves (blue). The dashed lines indicate the locations of the GMOS detector gaps.

aperture widths and background regions in order to minimize the residual scatter. For the final analysis, we used an optimum aperture radius of 18 pixels and two background regions located 80–100 pixels either side of the spectral trace. The background contribution was estimated by taking the median value within these regions. Fig. 1 shows example spectra for WASP-121 and the comparison star for both transits obtained after spectral extraction.

To align each of our spectra we use the x-shifts obtained from a process of cross-correlation using suitable absorption features in the stellar spectra after normalizing the continua. We repeated this cross-correlation procedure for a number of different absorption features but found the results to be insensitive to the specific feature chosen, in addition, we did not notice any significant stretching of the spectra when compared to the width of our bins over the time-series, and we use the alignment obtained using the O_2 feature in our final analysis.

To calibrate the wavelength scales, we first tried the wavelength solution derived from arc frames obtained with a calibration mask with much narrower 1 arcsec slit widths. However, inspection of our spectra revealed residual offsets between the target and comparison spectra and we instead constructed our solution by first identifying a set of well-resolved absorption lines in the mean spectrum, and then fitting a Gaussian to each of these lines to accurately retrieve the position of the line centres. Our solution is obtained by fitting these line centres using a Gaussian process (GP) with a standard squared exponential kernel (see Section 3.1 for a detailed description of our implementation of GPs) and we use the solution derived from this fit in our final analysis.

To produce the differential white light curves for each transit we bin the flux along the dispersion axis for the target and comparison star over the wide wavelength ranges shown in grey in Fig. 1, and then divide the flux of the target star by the flux of the comparison star for the entire time-series to correct for variations in observing conditions and telluric effects. Multiple spectroscopic light curves are constructed in a similar way by integrating over each of the narrower bins shown in blue (avoiding areas which overlapped the detector gaps). We extracted 53 spectroscopic channels for both transits and the resulting differential white light curves and spectral light curves are shown in Figs 2, 3, and 4.

Our limb darkening coefficients and their associated uncertainties were obtained using the PYLDTK toolkit (Parviainen & Aigrain 2015),

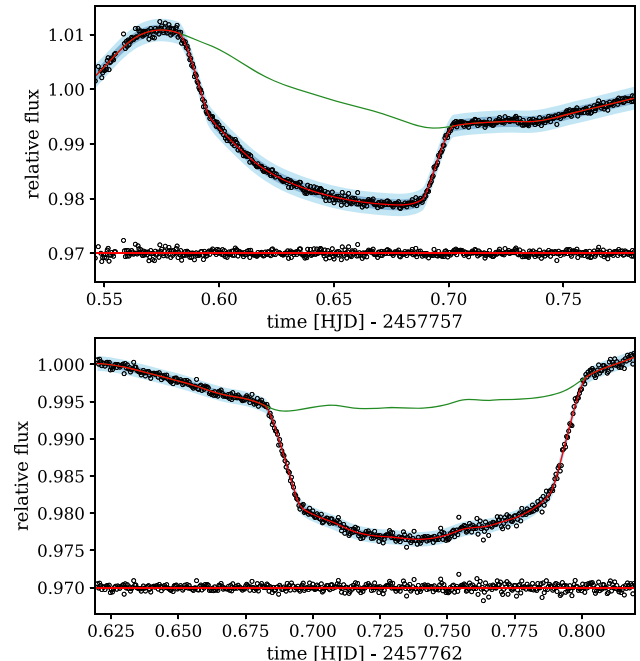


Figure 2. White light curves of WASP-121b obtained with the R400 grism + OG515 filter for Transit 1 (top) and Transit 2 (bottom). The red line shows the best-fitting model with blue shading indicating plus/minus two standard deviations. The green line shows the systematics model derived from the GP fit. Residuals are indicated below the light curves.

which uses the PHOENIX models of Husser et al. (2013). For this we used the stellar parameters for WASP-121 and the system parameters and uncertainties as given in Delrez et al. (2016). Finally, we also examined various diagnostic measurements, including the FWHM of the spectral trace and the shifts in the dispersion and spatial axes to check for correlations with the instrumental systematics (e.g. Brown 2001; Gilliland & Arribas 2003; Pont et al. 2007; Swain et al. 2009; Stevenson et al. 2010; Gillon et al. 2012; Huitson et al. 2013; Nikolov et al. 2016), though in this case no obvious correlations were found.

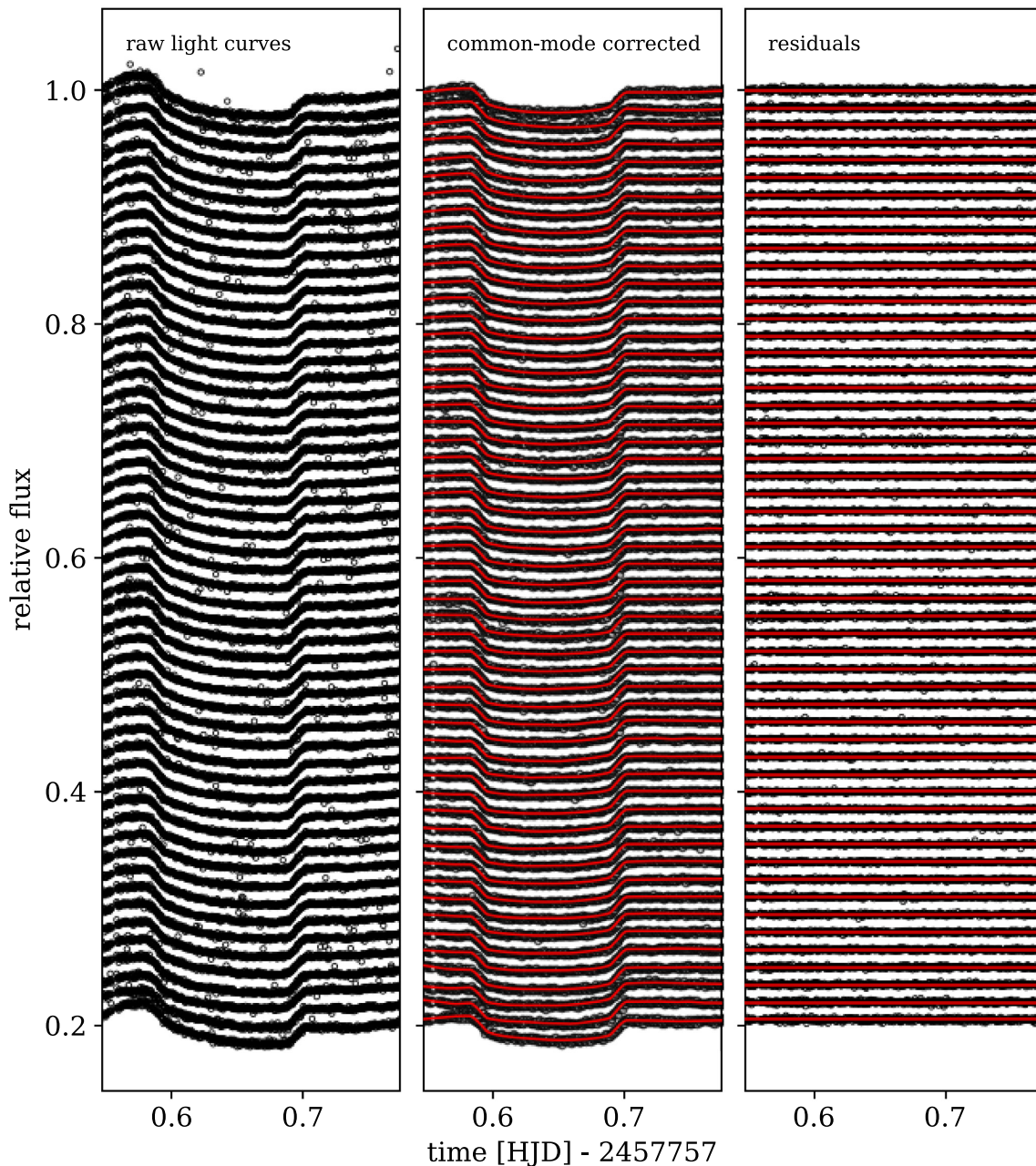


Figure 3. Spectral light curves for the R400 grism + OG515 filter for Transit 1 corresponding to the spectral channels shown in blue in Fig. 1 (left-hand panel). The left-hand panel shows the raw light curves before correction. The middle panel shows the light curves with best-fitting GP model after the common-mode correction. The right-hand panel shows the residuals from the best-fitting model.

3 ANALYSIS

3.1 White light-curve analysis

To model our white light curves we use a very similar procedure to that previously outlined in Wilson et al. (2020), which implements the methodology introduced in Gibson et al. (2012). In brief, we fit using a time-dependent GP⁴ simultaneously with the analytic transit model of Mandel & Agol (2002) with quadratic limb darkening. GPs have

⁴For the implementation of our Bayesian inference we made extensive use of the Python modules GEEPEA and INFER which are freely available from <https://github.com/neaalgibson>.

been successfully implemented in a large number of similar studies (e.g. Gibson et al. 2013a; Evans et al. 2017, 2018; Nikolov et al. 2018) and have been shown to be extremely useful for modelling correlated noise in exoplanet time-series, whilst offering robust uncertainty estimates. Our joint posterior distribution is given by the multivariate Gaussian:

$$p(\mathbf{f}|\mathbf{t}, \boldsymbol{\phi}, \boldsymbol{\theta}) = \mathcal{N}(T(\mathbf{t}, \boldsymbol{\phi}), \boldsymbol{\Sigma}(\mathbf{t}, \boldsymbol{\theta})), \quad (1)$$

where \mathbf{t} and \mathbf{f} are the vectors of time and flux measurements, respectively, T is the analytic transit mean function with parameters $\boldsymbol{\phi}$, and $\boldsymbol{\Sigma}$ is the covariance matrix with hyperparameters $\boldsymbol{\theta}$. Correlations between data points are described using the covariance matrix which is calculated using a kernel function with parameters $\boldsymbol{\theta}$ (for a more

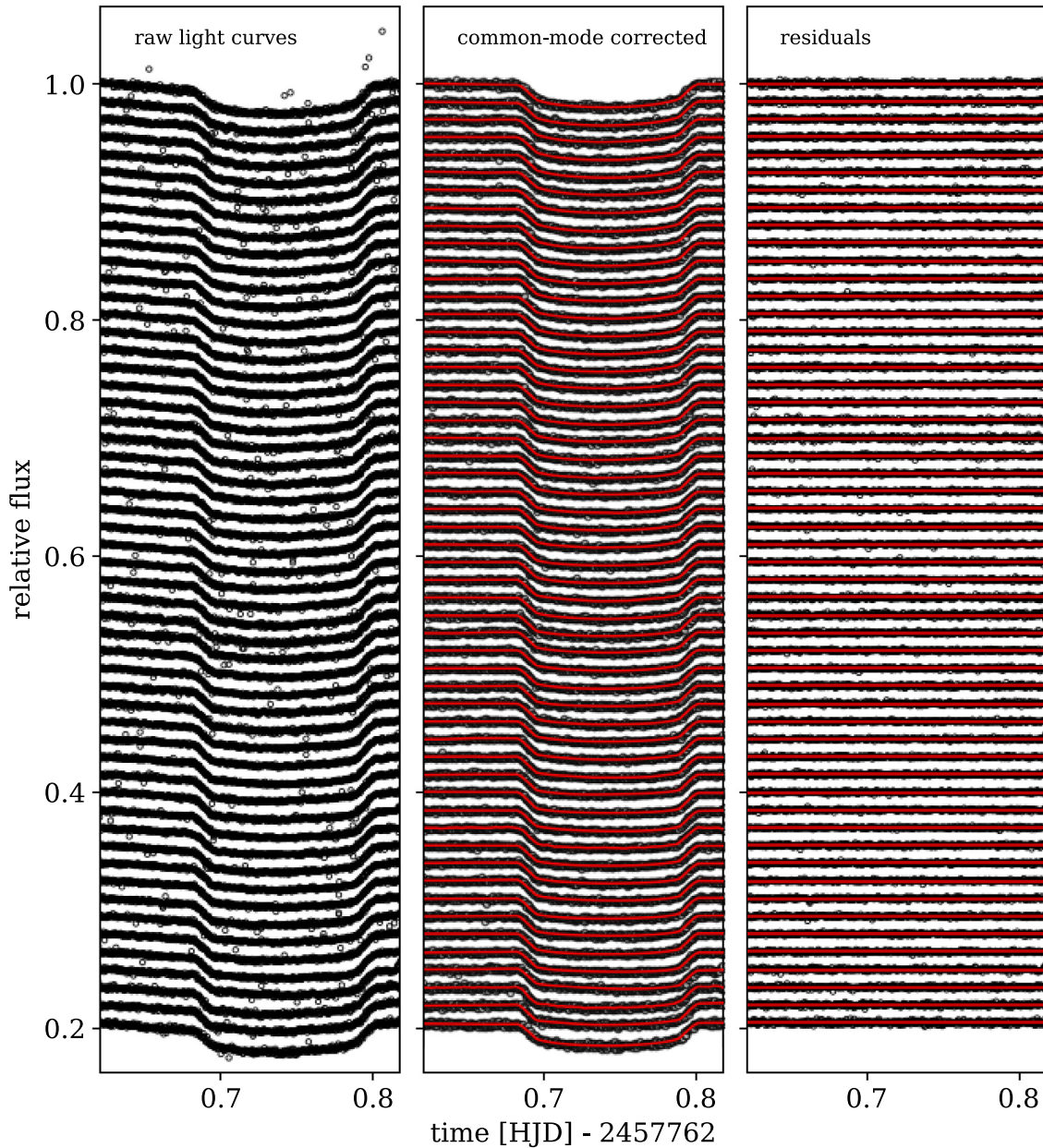


Figure 4. Same as Fig. 3 but for Transit 2. The left-hand panel shows the raw light curves before correction. The middle panel shows the light curves with best-fitting GP model after the common-mode correction. The right-hand panel shows the residuals from the best-fitting model.

detailed review of GPs and kernels see Rasmussen & Williams 2005). In this work, we use the Matérn 3/2 kernel which can be viewed as a less smooth version of the more commonly used squared exponential kernel and which is defined as

$$k(t_n, t_m | \theta) = \xi^2 (1 + \sqrt{3} \eta \Delta t) \exp(-\sqrt{3} \eta \Delta t) + \delta_{nm} \sigma^2, \quad (2)$$

where ξ is the height scale, Δt is the time difference of observations, η is the inverse length scale, δ_{nm} is the Kronecker delta, and σ is the white noise term which is identical for each data point. As a check we repeated our analysis using a squared exponential kernel but found that our results were unaffected. Since no obvious correlations were observed between the form of the systematics in the light curves and the diagnostic measurements, we proceeded by using time as our only input variable, which has the added advantage of requiring fewer assumptions. We obtain the posterior probability distributions

of the parameters of interest by first specifying prior distributions for the hyperparameters of the kernel function and then multiplying by the marginal log-likelihood.

Our GP mean function assumes a circular orbit with period fixed to that reported by Delrez et al. (2016). For each white light curve we allow the mid-transit time (T_c) and linear baseline parameters (f_{foot} , T_{grad}) to vary as free parameters and fixed the system scale (a/R_*) and impact parameter b to the values previously constrained in Evans et al. (2018). For the planet-to-star radius ratio ($\rho = R_p/R_*$) we use a Gaussian prior centred on the reported value in Evans et al. (2018). Constraining the white light-curve parameters to previously reported values allows us to more easily compare the respective results and should help us to improve the accuracy of our systematics models.

To account for the effects of stellar limb darkening, we used a quadratic limb darkening law (Claret 2000) and placed Gaussian

Table 1. Assumed transit parameter values used in the fitting of the white light curves. The orbital period, system scale, and impact parameter were held fixed and Gaussian priors were placed on the following parameters with the mean and standard deviations given below.

Parameter	Value
P	1.2749255 d (fixed)
a/R_*	3.86 (fixed)
b	0.06 (fixed)
ρ	0.1219 ± 0.0005
c_1	0.395 ± 0.003
c_2	0.141 ± 0.004

priors on the quadratic coefficients c_1 and c_2 with a mean and standard deviation determined using PYLDTK. We also tried repeating our analysis having both fixed the limb darkening parameters to their best-fitting values and leaving the parameters completely free in the fits, but found that this variation in treatment did not affect our results (see Section 5 where we discuss the impact of our limb darkening treatment further). We summarize the assumed parameter values for the white light curves in Table 1. Similar to previous studies (e.g. Evans et al. 2017; Gibson et al. 2017), we fit for $\log \eta$ and $\log \xi$ with uniform priors and constrain the length scales to lie between the cadence and twice the total duration of our observations.

We performed a joint fit to both individual transits, allowing for only a single planet-to-star radius ratio and limb darkening coefficient pair, whilst allowing the hyperparameters to vary separately for each transit. To check the validity of our assumed parameter values, we also carried out independent fits to both of our individual white light curves finding the values to be consistent within 1 sigma to those of Evans et al. (2018) for both transits, though the measured planet-to-star radius ratio was found to be higher for one transit and slightly lower for the other (discussed further in Section 5). This ultimately results in an offset in the mean levels of the individual transmission spectra, though it does not affect their relative values and we find that both transits produce spectra with a consistent shape. We further tested this by again fitting each transit independently, but this time using the same priors as for our joint fit (shown in Fig. 5). We confirm that the shape of the individual spectra are not affected by imposing these priors and find that both transits produce spectra which are in excellent agreement, though with quite different average uncertainties. We computed the reduced χ^2 between the two spectra to be 0.7 by taking the difference between each pair of points and adding the uncertainties in quadrature.⁵ The overall mean level of the transmission spectrum can also be influenced by inaccuracies in the common-mode correction, though this similarly should not affect the relative values, but can potentially lead to offsets, especially when including multiple data sets from different instruments, and needs to be carefully considered when interpreting results. The individual transmission spectra for Transits 1 and 2 are shown in Fig. 5 and the results from the joint fits are shown in Fig. 6. We discuss these results further in Section 5.

To find the best-fitting models for each of our white light curves we began by using a differential evolution algorithm to optimize the log-likelihood with respect to the transit and kernel parameters using the values from Delrez et al. (2016) and Evans et al. (2018) as the initial points. Next, we used a Nelder–Mead simplex algorithm (Nelder & Mead 1965) to fine-tune these estimated values and, finally, we ran a

⁵where the difference between pairs of data points will follow a normal distribution $\Delta x \sim \mathcal{N}(0, \sigma_{x_1}^2 + \sigma_{x_2}^2)$, hence distributed according to χ^2 .

Markov Chain Monte Carlo (MCMC) method to marginalize our posterior distributions. Four independent chains with length 80 000 were initiated for each of our light curves with the first 40 per cent of samples in the chain being considered as burn-in and discarded. We used the Gelman–Rubin statistic to confirm mutual convergence for each free parameter. We show the best fit white light-curve models and derived systematics models and residuals for both transits in Fig. 2.

3.2 Spectroscopic light-curve analysis

To construct our spectroscopic light curves, we began by binning over the narrow wavelength regions shown in blue in Fig. 1. For each transit, we extracted 53 individual light curves using a uniform bin width of 75 Å whilst avoiding the GMOS detector gaps. We show the resulting raw light curves in the left-hand panels of Figs 3 and 4. The raw light curves for both transits show significant systematics which are largely independent of wavelength. This is a similar situation to that encountered for the FORS2 data set in Wilson et al. (2020) and we proceeded in much the same way by dividing each of the spectroscopic light curves by the common-mode corrections derived from the white light curves. Additionally, to remove any remaining high-frequency systematics, we also subtract the residuals from the white light curves fits. The net result of this procedure is to provide significant improvements to the precision of our transmission spectrum whilst preserving the relative values of the planet-to-star radius ratios.

After applying the common-mode correction described above we find the best-fitting model for each spectroscopic light curve using the same process and the same systematics model as described in Section 3.1. The main difference is that we fix the mid-transit time to the best-fitting value from the white light-curve analysis, whilst allowing the planet-to-star radius ratio, limb darkening parameters, linear baseline parameters and kernel hyperparameters to vary for each fit. As for the white light curves, we perform a joint analysis allowing for only a single planet-to-star radius ratio and limb darkening coefficient pair for each individual wavelength channel, but allow the hyperparameters to vary separately for each transit. To add another layer of flexibility to our model, we use wide Gaussian priors with a standard deviation of 0.1 for the limb darkening coefficients, with a mean given by the best-fitting values determined using PYLDTK. We again used the same differential evolution algorithm to optimize the log-likelihood as for the white light curves, and fine-tuned using a Nelder–Mead simplex algorithm. Outliers deviating more than 4σ from the predictive distribution were removed for each fit (this procedure typically clipped 1–2 points per light curve) and we used the same MCMC method described above to marginalize our posterior distributions. The best-fitting common-mode corrected GP models for the spectroscopic light curves are shown in the middle panels of Figs 3 and 4 and we list the measured planet-to-star radius ratios and uncertainties in Table 2. Our GMOS transmission spectrum for WASP-121b is shown in Fig. 6 and the results are discussed further in Section 5.

3.3 Modelling systematics with Student’s- T processes

To investigate the robustness of the analysis outlined above, we performed some additional modelling using a Student’s- T Process (STP), implemented with a modified version of our GEEPEA code. STPs have already been successfully applied to Bayesian optimization problems and aerospace design (e.g. Shah 2013; Tracey & Wolpert 2018) and have been shown to come at no additional computational cost over GPs (Shah, Wilson & Ghahramani 2014)

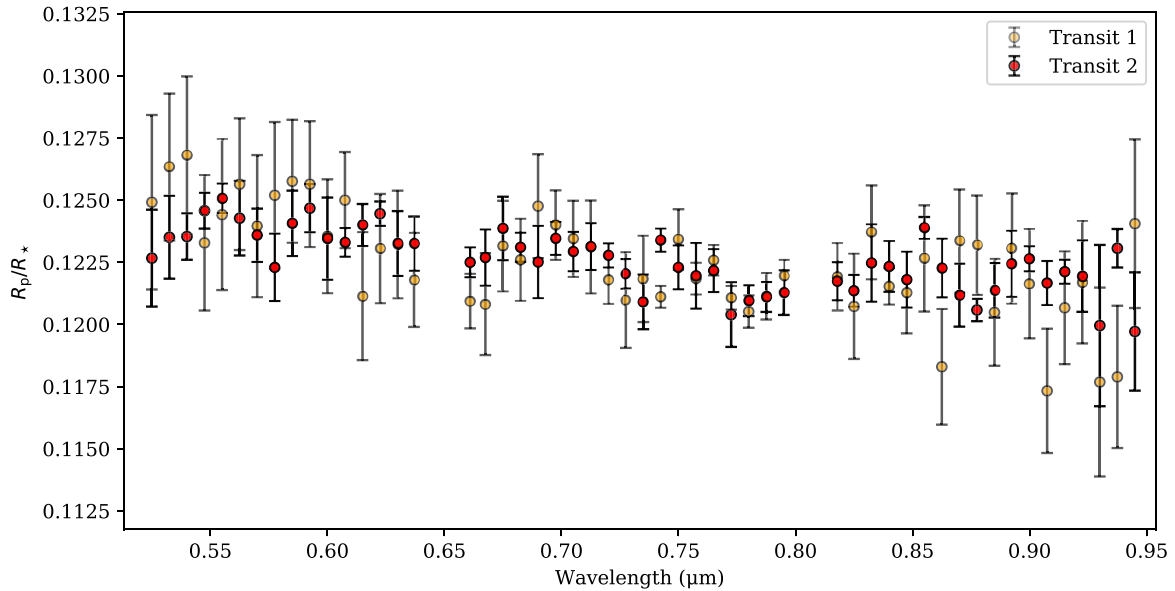


Figure 5. Individual Gemini/GMOS transmission spectra of WASP-121b for both transits. The orange points are the results of an independent analysis of the Transit 1 light curves using the same prior as for our joint analysis. The red points show the same but for Transit 2. The results from the joint fits are shown in Fig. 6 below.

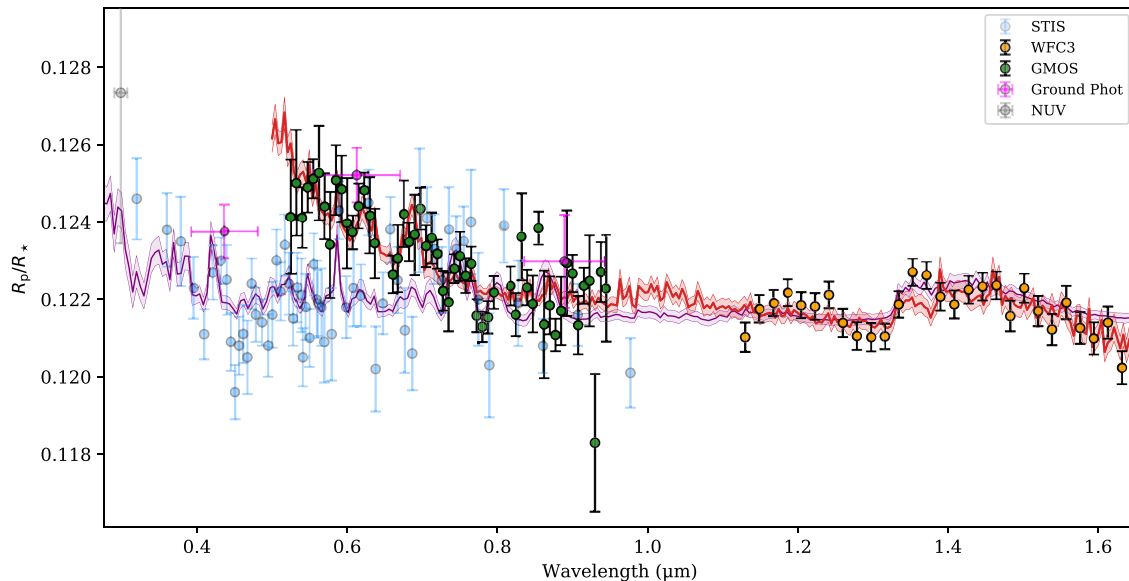


Figure 6. Combined optical-infrared transmission spectrum of WASP-121b obtained from Gemini/GMOS and WFC3 observations. The green points are the results from the joint fits to the GMOS light curves. The orange points are the WFC3 results from Evans et al. (2016) which have been slightly shifted to match the best-fitting offset found in the retrieval. The light blue points show the STIS results from Evans et al. (2018), magenta points show the results of a re-analysis of the ground-based photometry measurements from Delrez et al. (2016) and the grey point on the far left indicates the longest wavelength NUV measurement from Sing et al. (2019). The red line shows the best-fit from our GMOS/WFC3 retrieval analysis using PETRA along with 1σ significance contours (light red). We also show the best-fitting model from the STIS/WFC3 retrieval (purple) described in Lothringer et al. (2020a) for comparison.

but, to our knowledge, have not yet been applied to the analysis of systematics in transit light curves.

Whilst the basis of a GP is the multivariate Gaussian distribution, the basis of an STP is instead the multivariate Student’s- T distribution (Genz & Bretz 2009) and therefore, similarly to a GP, an STP defines a prior over functions. An STP differs however from a GP in two ways: first, a Student’s- T distribution can have a higher kurtosis than the corresponding Gaussian distribution, and therefore an STP prior

can assign higher probability to extreme outliers. Secondly, with a GP, the conditioned posterior variance (excluding white noise) depends exclusively on the location of the input variables, whilst with an STP the posterior variance will also depend explicitly on the observed values. To account for this extra complexity, an STP has an additional parameter, ν , which describes the degrees of freedom of the distribution. As the degrees of freedom approaches infinity, the multivariate Student’s- T distribution converges to the corresponding

multivariate Gaussian distribution. Thus, a STP can be considered as a generalization of a GP with an additional parameter. Following the derivation outlined in Tracey & Wolpert (2018), the pdf of a STP is given by

$$\mathcal{T}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu) = \frac{\Gamma((\nu + d)/2)}{\Gamma(\nu/2) \nu^{d/2} \pi^{d/2} |\boldsymbol{\Sigma}|^{1/2}} \times \left(1 + \frac{1}{\nu} (\mathbf{y} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{y} - \boldsymbol{\mu}) \right)^{-(\nu+d)/2}, \quad (3)$$

where $\boldsymbol{\mu}$ and ν are the mean and degrees of freedom of the distribution, respectively, and d is the dimension. $\boldsymbol{\Sigma}$ is a symmetric positive definite shape parameter which is related to the covariance matrix by

$$E[(\mathbf{y} - \boldsymbol{\mu})^T (\mathbf{y} - \boldsymbol{\mu})] = \frac{\nu}{\nu - 2} \boldsymbol{\Sigma}. \quad (4)$$

Using the notation common in the GP literature and given a set of observations $\mathbf{D} = \{(x_1, y_1), (x_2, y_2), (x_3, y_3), \dots\}$, the posterior mean and covariance will be given by

$$\hat{\boldsymbol{\mu}} = \mathbf{K}_* \mathbf{K}^{-1} \mathbf{y}, \quad (5)$$

and

$$\hat{\mathbf{K}} = \frac{\nu + \mathbf{y}^T \mathbf{K}^{-1} \mathbf{y} - 2}{\nu + |\mathbf{D}| - 2} (\mathbf{K}_{**} - \mathbf{K}_* \mathbf{K}^{-1} \mathbf{K}_*^T). \quad (6)$$

These two equations are the same as that which would be obtained for the equivalent GP, except for an additional term modifying the covariance, which can be seen as a corrective factor which depends explicitly on the observed values. In short, if the observed values are consistent with a GP then the posterior covariance of the STP will be roughly equal to that of a GP. Conversely, if the variation in the observed values is greater or smaller than expected then the STP posterior covariance will be larger or smaller, respectively. In other words, for the same kernel, the posterior means will be identical, but the variance will differ, depending on the observed values.

For our implementation of STPs we used the same Matérn 3/2 kernel and hyperparameter priors as for our GP analysis described in Sections 3.1 and 3.2. For the degrees of freedom parameter, we set a prior condition of $\nu > 4$ to ensure that the distribution has both finite variance and finite kurtosis. We then carry out the same procedure as before to perform a joint fit to both transits. We compare the results from the STP analysis with those from our GP analysis in Fig. A1 in the appendix. As can be seen from the figure, our results are almost identical across the full spectrum, with a mean uncertainty that is inflated by only a few percent and a posterior mean for ν that was typically a very large number. Our results therefore verify that the observed data can indeed be well described by a GP in this case, and gives us greater confidence in our estimated uncertainties.

For the analysis as implemented here we should acknowledge that the value for ν is common for both the stochastic component, and for the noise. Ideally, these should be treated independently with separate instances of ν for each. However, the fact that we recover almost identical results to our GP analysis is reassuring. We will return to examine this potential shortcoming in greater detail in future work. A more detailed discussion on the comparisons between STPs and GPs and their respective strengths and weaknesses is beyond the scope of this study.

4 ATMOSPHERIC MODELLING WITH PETRA

To obtain constraints on the vertical temperature structure and terminator composition of WASP-121b, we performed an atmospheric

Table 2. Transmission spectrum for WASP-121b recovered from the GMOS low-resolution spectroscopic light curves.

Wavelength Centre [Range] (Å)	Radius ratio R_p/R_*	Limb darkening	
		c_1	c_2
5250 [5213–5288]	0.12413 ± 0.00148	0.584	0.113
5325 [5288–5363]	0.12501 ± 0.00136	0.576	0.115
5400 [5363–5438]	0.12411 ± 0.00079	0.561	0.117
5475 [5438–5513]	0.12490 ± 0.00065	0.557	0.120
5550 [5513–5588]	0.12512 ± 0.00050	0.549	0.124
5625 [5588–5663]	0.12527 ± 0.00122	0.537	0.126
5700 [5663–5738]	0.12440 ± 0.00085	0.527	0.130
5775 [5738–5813]	0.12341 ± 0.00137	0.523	0.133
5850 [5813–5888]	0.12508 ± 0.00090	0.509	0.137
5925 [5888–5963]	0.12485 ± 0.00086	0.508	0.134
6000 [5963–6038]	0.12397 ± 0.00117	0.498	0.136
6075 [6038–6113]	0.12373 ± 0.00045	0.493	0.135
6150 [6113–6188]	0.12441 ± 0.00058	0.485	0.131
6225 [6188–6263]	0.12482 ± 0.00045	0.474	0.135
6300 [6263–6338]	0.12417 ± 0.00099	0.473	0.136
6370 [6333–6408]	0.12345 ± 0.00090	0.464	0.138
6610 [6580–6640]	0.12264 ± 0.00048	0.424	0.155
6675 [6650–6700]	0.12305 ± 0.00086	0.438	0.142
6750 [6713–6788]	0.12421 ± 0.00086	0.433	0.141
6825 [6788–6863]	0.12348 ± 0.00049	0.429	0.138
6900 [6863–6938]	0.12367 ± 0.00104	0.424	0.144
6975 [6938–7013]	0.12434 ± 0.00054	0.418	0.144
7050 [7013–7088]	0.12337 ± 0.00054	0.414	0.139
7125 [7088–7163]	0.12358 ± 0.00066	0.405	0.141
7200 [7163–7238]	0.12317 ± 0.00037	0.405	0.140
7275 [7238–7313]	0.12223 ± 0.00051	0.398	0.140
7350 [7313–7388]	0.12193 ± 0.00076	0.395	0.140
7425 [7388–7463]	0.12279 ± 0.00037	0.388	0.141
7500 [7463–7538]	0.12311 ± 0.00063	0.384	0.142
7575 [7538–7613]	0.12261 ± 0.00044	0.379	0.141
7650 [7613–7688]	0.12292 ± 0.00044	0.373	0.141
7725 [7688–7763]	0.12158 ± 0.00043	0.369	0.142
7800 [7763–7838]	0.12129 ± 0.00040	0.366	0.141
7875 [7838–7913]	0.12154 ± 0.00042	0.362	0.142
7950 [7913–7988]	0.12218 ± 0.00042	0.358	0.140
8180 [8143–8218]	0.12234 ± 0.00050	0.338	0.144
8250 [8225–8275]	0.12160 ± 0.00055	0.330	0.147
8325 [8288–8363]	0.12362 ± 0.00112	0.332	0.142
8400 [8363–8438]	0.12230 ± 0.00046	0.317	0.150
8475 [8438–8513]	0.12188 ± 0.00073	0.312	0.147
8550 [8513–8588]	0.12384 ± 0.00043	0.307	0.147
8625 [8588–8663]	0.12135 ± 0.00139	0.307	0.151
8700 [8663–8738]	0.12184 ± 0.00075	0.306	0.148
8775 [8738–8813]	0.12108 ± 0.00042	0.304	0.150
8850 [8813–8888]	0.12170 ± 0.00088	0.288	0.157
8925 [8888–8963]	0.12294 ± 0.00136	0.329	0.144
9000 [8963–9038]	0.12266 ± 0.00048	0.279	0.158
9075 [9038–9113]	0.12133 ± 0.00075	0.320	0.140
9150 [9113–9188]	0.12235 ± 0.00046	0.331	0.139
9225 [9188–9263]	0.12248 ± 0.00117	0.258	0.164
9300 [9263–9338]	0.11829 ± 0.00178	0.325	0.138
9375 [9338–9413]	0.12271 ± 0.00077	0.328	0.132
9450 [9413–9488]	0.12228 ± 0.00138	0.320	0.137

retrieval using an adaptation of the retrieval code PETRA (PHOENIX Exoplanet Retrieval Algorithm; Lothringer & Barman 2020). In addition to the optical GMOS data presented in Table 2, our retrieval also incorporates the near-IR *HST*/WFC3 data of Evans et al. (2016). Atmospheric retrieval codes typically comprise a forward model which is used to generate the predicted atmospheric spectrum and include a statistically robust inference method for parameter

estimation. PETRA incorporates a modified version of the well-tested and self-consistent atmosphere model PHOENIX as its forward model, which has been widely used to study the atmospheres of both stellar and sub-stellar atmospheres (Hauschildt, Baron & Allard 1997; Hauschildt, Allard & Baron 1999; Barman, Hauschildt & Allard 2001; Allard, Homeier & Freytag 2011; Barman et al. 2011; Lothringer et al. 2018; Lothringer & Barman 2019).

The model assumes hydrostatic equilibrium and a plane-parallel atmosphere and solves for radiative transfer in a line-by-line manner for transmission geometry. PHOENIX incorporates an expansive line list data base including CIA, bound-free, and free-free opacities. We include the prominent absorbers expected for a hydrogen-rich atmosphere in the observed spectral range: H₂O, Na, K, TiO, VO, HCN, and other metal oxides, hydrides, and atomic metal opacity. Chemical equilibrium is calculated with PHOENIX’s Astrophysical Chemical Equilibrium Solver (ACES) to calculate the equation of state for each species with elemental abundances scaled uniformly by a free parameter for the bulk metallicity. We treat H₂O, TiO, and VO abundances as free and independent parameters. Line profiles are calculated using Voigt profiles. Details of the sources of the various opacity cross-sections can be found in Lothringer et al. (2018) and Lothringer & Barman (2019).

For the p–T profile, we make use of the three-channel Eddington approximation outlined in Parmentier & Guillot (2014) as implemented in Line et al. (2013). This approach has five free parameters which describe the Planck mean thermal IR opacity, two independent downwelling visible channels of radiation, the partition of the flux between the two visible streams, and a catch-all term which describes the albedo, emissivity, and day–night redistribution. Our model also considers the contributions from cloud/haze coverage as described in MacDonald & Madhusudhan (2017), with parameters P_{cloud} (the cloud-deck altitude in bars), a parameter which describes the scattering enhancement factor and a parameter which describes the scattering slope. For a full description of the retrieval framework see Lothringer & Barman (2020).

We use a differential evolution Markov Chain (DEMC) algorithm (Ter Braak 2006) with ‘snooker’ updates (Ter Braak & Vrugt 2008) as a statistical sampling algorithm to obtain the posterior distributions of the forward model parameters, checking for convergence using the Gelman–Rubin statistic. Lastly, we also include an offset between the GMOS/HST data sets as another free parameter to account for potential discrepancies. This is particularly important given that the Evans et al. (2016) analysis adopted slightly different values for the transit shape parameters to those adopted in Evans et al. (2018). We used wide uniform priors for the temperature parameters, the H₂O, TiO, and VO abundances (between 10^{-12} and 10^{-1}) and [Fe/H] (between 10^{-1} and 10^4). For our final parameter and uncertainty estimates, we also inflated the error bars by ~ 13 per cent so that our fits were approaching a chi-square of 1.0.

Our retrieval results are summarized in Table 3. We find little evidence for either TiO or VO with 1σ upper limits on the best-fitting absolute terminator abundances of $\log(X_{\text{TiO}}) < -9.60$ and $\log(X_{\text{VO}}) < -10.25$, respectively. We also constrain the H₂O abundance to $\log(X_{\text{H}_2\text{O}}) = -5.05^{+0.19}_{-0.18}$ and find a best-fitting mean temperature of 2664^{+120}_{-100} K. In contrast to the results in Evans et al. (2018), our retrieval favours a scattering/haze slope rather than absorption by TiO/VO, with best-fitting scattering enhancement and slope of $2.43^{+0.31}_{-0.29}$ and $-7.51^{+1.72}_{-1.86}$, respectively. This is perhaps not all that surprising given the overall smoother, sloping shape of our transmission spectrum. For our best-fitting model, we reach a reduced χ^2 of 1.485 (for 14 free parameters) which is comparable to the value in Evans et al. (2018). Interestingly, our retrieval also tries to fit the

Table 3. Retrieved atmospheric parameters using PETRA, where $\mathcal{U}(a, b)$ defines a uniform distribution between a and b .

Parameter	Prior	Value
$\log(X_{\text{H}_2\text{O}})$	$\mathcal{U}(-12, -1)$	$-5.05^{+0.19}_{-0.18}$
$\log(X_{\text{TiO}})$	$\mathcal{U}(-12, -1)$	$-9.98^{+0.38}_{-0.65}$
$\log(X_{\text{VO}})$	$\mathcal{U}(-12, -1)$	$-10.74^{+0.49}_{-0.56}$
T_0	$\mathcal{U}(600, 4200)$	2664^{+120}_{-100}
Scattering enhancement factor	$\mathcal{U}(-4, 8)$	$2.43^{+0.31}_{-0.29}$
Scattering slope	$\mathcal{U}(-15, 2)$	$-7.51^{+1.72}_{-1.86}$
$\log(P_{\text{cloud}})$ [bar]	$\mathcal{U}(-4, 8)$	$6.09^{+0.87}_{-0.79}$

metal hydride CaH to the apparent feature at ~ 0.7 microns, although we find only weak evidence for it ($\Delta\text{BIC} = 2.14$). CaH absorption is observed in some cool stars and brown dwarfs (e.g. Kirkpatrick et al. 1999; Reiners et al. 2007; West et al. 2011).

The best fit from our GMOS/WFC3 retrieval along with 1 sigma confidence contours are shown in Fig. 6 and the full marginalized posterior probability densities are presented in Fig. A3 in the appendix. A similar retrieval using PETRA was also carried out on the original STIS and WFC3 data and is presented in detail in Lothringer et al. (2020a). In that study the Fe/H, H₂O, and VO abundances were found to be slightly reduced from that reported in Evans et al. (2018) though the overall interpretation remained largely the same. We also show the best-fitting model from this retrieval in Fig. 6 for comparison.

5 DISCUSSION

WASP-121b is a close-in, ultra-hot Jupiter with a mass and radius of $1.18 M_J$ and $1.7 R_J$, respectively, an equilibrium temperature over 2400 K and an orbital period of 1.27 d (Delrez et al. 2016). In Evans et al. (2017) strong evidence was found for the presence of a vertical thermal inversion using secondary eclipse measurements and this was subsequently confirmed in Bourrier et al. (2020b). The cause of such thermal inversions in the atmospheres of hot Jupiters has traditionally been suggested to be due to absorption by TiO and/or VO (Fortney et al. 2008), though a convincing detection of either molecule has yet to be made for WASP-121b. Evans et al. (2018) presented a STIS optical transmission spectrum of WASP-121b which showed features consistent with absorption by VO. However, follow-up secondary eclipse observations and recent observations at high-resolution have been unable to confirm this detection (Merritt et al. 2020; Mikal-Evans et al. 2020). This has led to some speculation that other species, such as the plethora of atomic metals including Fe I that have recently been detected, might be the drivers for WASP-121b’s temperature inversion (Lothringer et al. 2018; Gandhi & Madhusudhan 2019; Gibson et al. 2020).

The transmission spectrum obtained from our joint analysis of two transits of WASP-121b is shown in Fig. 6, with the STIS data of Evans et al. (2018) and the results of a re-analysis of the ground-based photometry from Delrez et al. (2016) (as presented in Evans et al. 2016) overplotted. For a small number of our points, the values and uncertainties appear to deviate slightly from what would be expected from a simple weighted average of the two individual transits. This is likely due to more accurate limb darkening constraints from the joint fits, which should also improve the constraints on the corresponding systematics models. Overall, our transmission spectrum shows some agreement with that of Evans et al. (2018) for wavelengths longwards of ~ 650 nm, with some similar features

being apparent in the middle portion of the spectrum. However, the results begin to diverge noticeably at wavelengths shorter than about ~ 650 nm, with the GMOS values being consistently larger than the STIS measurements across this short wavelength portion of the spectrum. We considered a number of factors that might be responsible for this apparent disagreement. The first of these is the treatment of limb darkening, which can potentially introduce small discrepancies in the transmission spectrum if not treated correctly, and will be particularly true at shorter wavelengths where the effects are strongest. For the analysis we have presented here we used the best-fitting values calculated using PYLDTK and imposed wide Gaussian priors with a width of 0.1 to add some flexibility to the model fits. However, we also tried both fixing the limb darkening coefficients to the best-fitting values and leaving them completely free in the fits, and found that this made almost no difference to our final transmission spectrum. In the Evans et al. (2018) study they used 3D limb darkening coefficients from the STAGGER grid (Magic et al. 2013), so, as a check to ensure that our results were not overly dependent on our particular choice of limb darkening treatment, we re-extracted all light curves using identical bins as used for the G750L data set in Evans et al. (2018), and repeated the analyses having fixed the limb darkening coefficients to the same best-fitting values obtained from the STAGGER 3D stellar model. A comparison of our original spectrum and that produced from this subsequent analysis is shown in Fig. A2 in the appendix. We find that adopting the STAGGER 3D limb darkening coefficients has a minimal influence on the overall shape of our transmission spectrum and is unable to explain the discrepancy between the GMOS and STIS data sets. This, together with the fact that we see little variation from fixing the limb darkening parameters or leaving them free, leads us to conclude that our results appear to be relatively insensitive to the treatment of limb darkening.

Stellar activity, including the possibility of unocculted spots or plages, has the potential to introduce both systematic offsets and wavelength-dependent biases in the measured transmission spectrum, with the effect being strongest for shorter wavelengths. In addition, changes in the wavelength dependence, as a result of time-variable spot coverages, could plausibly explain the observed discrepancy between the GMOS and STIS data sets given the separation in time of the individual observations. In Rackham, Apai & Giampapa (2019) the estimated contamination for an F6 dwarf was found to be around a factor of ~ 1.001 , which corresponds to an offset in transit depth of ~ 0.0025 per cent. The authors conclude that, whilst the effects of stellar contamination are more pronounced at shorter wavelengths, unocculted spots on typically active FGK dwarfs should only result in minor transit depth changes. In the WASP-121b discovery paper, Delrez et al. (2016) used the 60 cm TRAPPIST telescope to investigate the photometric variability of WASP-121 over a six-week period, finding standard deviations of 1.6 mmag in the *B* band, 1.3 mmag in the *V* band and 1.1 mmag in the *z'* band for the nightly photometry, and found no evidence for periodic variability above the ~ 1 mmag level. Despite finding the photometry to be quiet, they none the less found that WASP-121 shows high scatter in its RV residuals, CCF bisector spans and FWHM from an analysis of CORALIE spectra, and suggested that this could point to the photosphere being plage dominated, since the lower flux ratios with respect to spots would result in smaller brightness variations (e.g. Dumusque, Boisse & Santos 2014).

Similarly, Evans et al. (2018) also carried out photometric monitoring of WASP-121 using the Automated Imaging Telescope (AIT) during two separate campaigns in 2017 and 2018 and found a standard deviation about the yearly mean of 4.6 mmag for the

2017 campaign and 3.0 mmag for the 2018 campaign, with no significant periodicity detected for either campaign. They concluded that host star activity was unlikely to have significantly affected their transmission spectrum. In addition, they demonstrated that unocculted spots were unable to explain the shape of their measured spectrum under reasonable assumptions. Given this relatively quiet photometry and considering that both of our visits (separated by five days) analysed individually result in transmission spectra with very similar shapes, we likewise do not expect stellar activity to have significantly affected our results. However, taking into account the signatures of activity presented in the discovery paper, together with the temporal separation of the observations, we cannot entirely rule out the possibility that variable activity levels could at least partially be responsible for the discrepancies between our results and those obtained using STIS.

In order to quantify the potential effect of varying activity, we follow a similar approach to that outlined in Evans et al. (2018), which is an equivalent method to calculating spot correction factors as used in previous analyses (e.g. Sing et al. 2011; Huitson et al. 2013). Using this approach we fit for the chromatic bias $\kappa(\lambda)$, which we here define as the difference between the measured GMOS transmission spectrum and the weighted average of the G750L STIS points and is given by:

$$\kappa(\lambda) = D \left(\frac{\alpha[1 - \beta(\lambda)]}{1 - \alpha[1 - \beta(\lambda)]} \right), \quad (7)$$

where α is the fractional spot coverage (also known as the filling factor), β is the wavelength-dependent spot-to-photosphere flux ratio, and D is the true transit depth in the absence of spots. In our fits we set D equal to the weighted average of the STIS measurements and allow α to vary as a free parameter. This simple model assumes a fractional spot coverage and fixed spot temperature, and we use a model atmosphere obtained from the Phoenix grid (Husser et al. 2013) with parameters $T_* = 6500$ K, $\log g = 4.0$ cgs, $[\text{Fe}/\text{H}] = 0$ dex to estimate the flux for the stellar photosphere. We use the same model to estimate the spot fluxes for a range of spot temperatures from 3500 K up to 6000 K and repeat our fits for each spot temperature. We found that none of our bias models were able to adequately explain the difference between the GMOS and STIS data sets, though the fits improved as the spot temperature was increased, however, this also required the spot fractions to increase from at least 2 per cent coverage to about 8 per cent coverage. We repeated this process subtracting the data sets in the opposite order, finding similar results. Given the suggestion in Delrez et al. (2016) that the activity might be dominated by plage, we also tried fitting for higher temperature spots but found that a similar level of coverage was required in this case. This level of spot coverage variation would almost certainly be expected to result in larger features in the long-term monitoring provided by TRAPPIST, AIT, and TESS which, although limited by both photometric precision and temporal coverage, suggest modulations no larger than about 5 mmag. Given the complexity involved in modelling the effects of stellar activity, we cannot completely rule out the possibility of changing activity levels with the simple model used here, although we consider it unlikely that it could account for the full variability observed between the data sets.

Another likely explanation for the disagreement is the possibility of remaining, unaccounted-for systematics in one or both of the GMOS transits, or within the STIS data sets, and perhaps even within both. Although independent analyses of our two individual transits already reveals an overall offset between them, we none the less recover transmission spectra with a consistent shape for both (see

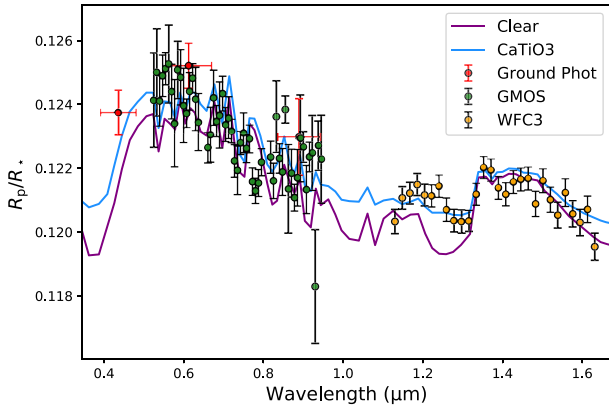


Figure 7. The GMOS/WFC3 data compared to the GCM predictions from Parmentier et al. (2018) with (blue) and without (purple) CaTiO_3 clouds. The ground-based photometry from Delrez et al. (2016) is also shown.

Fig. 5), and we have shown that this does not change with our choice of priors for the transit shape parameters. It therefore seems strange that unaccounted-for systematics could conspire to alter the shape of our final transmission spectrum, unless they affected both individual transits in a similar way. However, this is equally true of the STIS data sets (which were analysed using a similar Gaussian process technique), where good agreement was found for separate fits to the individual G430L visits. We therefore cannot rule out the possibility that some unaccounted-for systematics in the GMOS and/or STIS transits are the cause of the discrepancy at short wavelengths. We further discuss the potential causes of the disagreement between data sets in Section 5.1.

Whilst we fail to detect any significant evidence for TiO or VO absorption in our PETRA retrieval, our transmission spectrum clearly shows excess absorption in the optical relative to that in the near-IR as reported in previous studies. Our results would therefore tend to support the hypothesis that other species are responsible for this optical excess. In particular, a number of atomic metals, including Fe I, have already been detected in the atmospheres of WASP-121b and several other ultra-hot Jupiters at high-resolution (e.g. Cauley et al. 2019; Hoeijmakers et al. 2019; Sing et al. 2019; Gibson et al. 2020; Hoeijmakers et al. 2020; Nugroho et al. 2020) and would be expected to be strong sources of optical absorption (Fortney et al. 2008; Lothringer et al. 2020b). The best fit from our retrieval analysis favours a scattering/haze slope, and absorption by Fe I and other atomic metals could potentially help us to strengthen this slope. Additionally, UV observations of WASP-121b have shown evidence of atmospheric escape (Sing et al. 2019) and this could also enhance the effect of metals on the transmission spectrum if there is substantial material escaping from the atmosphere.

It has not been possible to uniquely identify these absorbing species in our low-resolution observations, though it seems likely that one or more of the reported species could be contributing to the observed optical excess and driving the thermal inversion (Evans et al. 2017; Gibson et al. 2020; Lothringer et al. 2020b). It is also possible that our non-detection of either TiO or VO could be explained simply by these molecules being condensed out on the cooler nightside and terminator regions. In Fig. 7 we show a comparison of the GMOS transmission spectrum and WFC3 data from Evans et al. (2016) with the predictions of a 3D general circulation model (GCM) including TiO/VO/FeH opacity from Parmentier et al. (2018). The two models correspond to a temperature map produced by a solar composition SPARC/MITgcm simulation with asymmetric temperatures between

the west and east limbs and assume either a clear atmosphere or the presence of CaTiO_3 clouds. It is clear from this comparison that the GMOS data is in broad agreement with both the ground-based photometry and GCM predictions, even though the models have not themselves been fine-tuned to the data. Given this broad agreement we cannot entirely rule out TiO/VO despite not resolving such features in our transmission spectrum. Additionally, the reasonably strong haze slope favoured by PETRA is inconsistent with the short wavelength turn down in the GCM. There thus remains a number of plausible scenarios that could explain the current data and further observations at shorter wavelengths may help to determine if the absorption keeps rising (as favoured by PETRA) or turns back down (as favoured by the GCM).

A clearer picture of the overall atmospheric composition of WASP-121b may also emerge with broader wavelength coverage including in the IR with e.g. *JWST*. Accurate modelling of the full optical-infrared continuum is crucial for reliable retrievals, helping to break some of the degeneracies between model parameters (Pinhas et al. 2018; Wakeford et al. 2018). A further way to break some of these degeneracies and help constrain the atmospheric properties could be achieved by combining the low-resolution observations with those obtained at high-resolution, either by joint fitting or by combining the posterior distributions of parameters, for example using direct likelihood evaluation techniques (e.g. Brogi & Line 2019; Gibson et al. 2020).

5.1 Variability in the transmission spectrum?

A further, and somewhat intriguing, interpretation of the differences in the STIS and GMOS transmission spectra is that they are direct evidence of temporal variability in the atmospheric properties of WASP-121b, on time-scales similar to that of the separation between the individual data sets. Such time-variability has previously been evoked to explain the changing phase curve offsets observed for HAT-P-7b, Kepler-76b, and WASP-12b (Armstrong et al. 2016; Bell et al. 2019; Jackson et al. 2019), and also the changing secondary eclipse depths for WASP-12b (Hooton et al. 2019; von Essen et al. 2019), although no such variations have yet been confirmed in transit observations.

In Parmentier et al. (2013), the authors use a 3D GCM to simulate dynamical mixing and vertical settling for the well-studied hot Jupiter HD 209458b, and show that strong variability of condensable chemical constituents is expected both spatially and temporally, driven mainly by the large day-to-night temperature contrast, and could result in observable variations during transit measurements. Furthermore, these results may also apply to silicate cloud/haze coverage, which could also display variability from epoch to epoch. Depending on the size of the condensates involved, the simulations predict an expected period of fifty to one hundred days for the largest amplitude variations, which is similar to the separation between the STIS/GMOS observations.

Similarly, in Komacek & Showman (2020), the authors used GCM simulations to investigate the time-variability of hot Jupiter atmospheres, finding ~ 0.1 per cent–1 per cent variations in global-average, dayside-average, and nightside-average temperatures, and ~ 1 per cent–10 per cent variations in globally averaged wind speeds, even when ignoring variability from magnetic effects and clouds. They also show how these variations could result in time-variable secondary eclipse depths, phase-curve amplitudes, and offsets and terminator-averaged wind speeds. From phase-curve retrievals, WASP-121b is expected to have wind speeds of $\sim 7 \text{ km s}^{-1}$ and a p–T profile which lies near the condensation curves of a

number of species (e.g. Parmentier et al. 2018; Daylan et al. 2021). It is therefore perhaps not all that surprising that small temperature fluctuations could result in significant spatial and temporal variations in atmospheric constituents and could lead to measurable variations in transit measurements.

Taken together the STIS/GMOS observations represent a somewhat unique data set in that they both consist of sets of repeat observations separated by several days, with a longer period of ~ 60 d between these sets. The STIS data set comprises two repeat observations using the G430L grating and one with the G750L grating with individual analyses of the G430L data giving consistent results and showing a good level of agreement both with the NUV data from Sing et al. (2019) and with the G750L data set across the overlapping wavelength range. In their analysis Evans et al. (2018) carried out a comprehensive range of tests to evaluate the robustness of their transmission spectrum, including a sensitivity analysis of their limb darkening treatment, the inclusion of additional GP inputs and a consideration of the effects of stellar activity. Evans et al. (2018) also noticed a small discrepancy between their results and the ground-based photometry results obtained by Delrez et al. (2016), and also briefly speculated that intrinsic variability of the atmosphere from epoch to epoch could be responsible (these observations were separated by over 100 d). Both sets of individual observations for each instrument show excellent repeatability on short time-scales, whilst the results over a longer period do not. It therefore seems plausible that the discrepancies between the STIS, GMOS, and ground-based photometry could be the signature of time-variable weather conditions in the atmosphere of WASP-121b. However, we cannot entirely rule out the possibility that coherent systematics could be affecting the individual observations for one or both of the data sets in a similar way. Whilst a number of studies have shown some disagreement between ground/space-based instrumentation likely due to systematics (e.g. Gibson et al. 2017, 2019), others have shown good agreement (e.g. Alam et al. 2020) and we therefore consider temporal variability to represent a credible and intriguing possibility.

The fact that both the GMOS and STIS data sets can be fit with physically plausible PETRA models, are repeatable (over \sim day time-scales), appear insensitive to limb darkening treatment and are a reasonable match to theoretical GCM predictions, means that we have no strong reason to discount the validity of either data set and is therefore suggestive of the presence of time-variability. However, neither can we fully rule out the possibility of more mundane causes, namely the influence of stellar activity and instrumental systematics, given the difficulty in robustly accounting for them even with our best instrumentation (e.g. Gibson et al. 2017, 2019; Espinoza et al. 2019). We would therefore highly recommend further transit observations in the optical regime using appropriate time-scales to specifically search for signs of such variable weather conditions.

6 CONCLUSION

We have presented ground-based Gemini/GMOS observations of the ultra-hot Jupiter WASP-121b covering two full transits and extracting a transmission spectrum over the wavelength range ≈ 500 – 950 nm using the technique of differential spectrophotometry. We used a Gaussian process to simultaneously model the deterministic transit component and the instrumental systematics thereby avoiding the need to specify a specific functional form. We also introduced a new analysis technique for dealing with instrumental systematics in exoplanet transit light curves using Student's- T processes, which can be thought of as a generalization of the more commonly employed Gaussian processes. We used this new method to verify the analysis

we have described here, finding that the results are in excellent agreement. We used the derived systematics models to correct our spectroscopic light curves by the application of common-mode corrections to improve the precision of our transmission spectrum. In contrast to the STIS results, we find evidence for an increasing blueward slope and little evidence for absorption from either TiO or VO from an atmospheric retrieval using PETRA, in agreement with a number of recent studies performed at high-resolution. We suggest that this might point to other absorbers, such as some combination of atomic metals, being responsible for the optical excess and the vertical thermal inversion observed for WASP-121b. The scattering/haze slope favoured by our retrieval could be enhanced by both absorption from Fe I and other atomic metals and from the atmospheric escape of material.

We considered a number of factors that might be responsible for the apparent discrepancy between the GMOS and STIS observations and conclude that, while it is still possible that the disagreement could be caused by either contamination from host star activity or some unaccounted-for systematic effects, the excellent repeatability of both individual data sets over day-long periods, and difference over longer month-long periods suggests that it is plausible that this discrepancy represents time-variable weather conditions in the atmosphere of WASP-121b, as predicted by theoretical models of ultra-hot Jupiters. We recommend further observations to both confirm this time-variability and help to further constrain the atmospheric properties, ultimately revealing the driver of the vertical thermal inversion.

ACKNOWLEDGEMENTS

This work is based on observations obtained at the international Gemini Observatory, a program of NSF's NOIRLab, which is managed by the Association of Universities for Research in Astronomy (AURA) under a cooperative agreement with the National Science Foundation. On behalf of the Gemini Observatory partnership: the National Science Foundation (United States), National Research Council (Canada), Agencia Nacional de Investigación y Desarrollo (Chile), Ministerio de Ciencia, Tecnología e Innovación (Argentina), Ministério da Ciência, Tecnologia, Inovações e Comunicações (Brazil), and Korea Astronomy and Space Science Institute (Republic of Korea). These data were obtained under programme GS-2016B-Q-42. We thank Vivien Parmentier for providing the models shown in Fig. 7 from Parmentier et al. (2018). JW would like to acknowledge funding from the Northern Ireland Department for the Economy. NPG gratefully acknowledges support from Science Foundation Ireland and the Royal Society in the form of a University Research Fellowship. We are grateful to the developers of the NUMPY, SCIPY, MATPLOTLIB, IPYTHON, and ASTROPY packages, which were used extensively in this work (Hunter 2007; Perez & Granger 2007; Astropy Collaboration 2013; Harris et al. 2020; Virtanen et al. 2020).

DATA AVAILABILITY

The data in this article are available from the Gemini Observatory Archive (<https://archive.gemini.edu>) with program ID GS-2016B-Q-42. The data products generated from the raw data are available upon request from the author.

REFERENCES

- Alam M. K. et al., 2020, *AJ*, 160, 51
 Allard F., Homeier D., Freytag B., 2011, in Johns-Krull C., Browning M. K., West A. A., eds, ASP Conf. Ser. Vol. 448, 16th Cambridge Workshop on

- Cool Stars, Stellar Systems, and the Sun. Astron. Soc. Pac., San Francisco, p. 91
- Armstrong D. J., de Mooij E., Barstow J., Osborn H. P., Blake J., Sanjeev N. F., 2016, *Nat. Astron.*, 1, 0004
- Astropy Collaboration et al., 2013, *A&A*, 558, A33
- Barman T. S., Hauschildt P. H., Allard F., 2001, *ApJ*, 556, 885
- Barman T. S., Macintosh B., Konopacky Q. M., Marois C., 2011, *ApJ*, 733, 65
- Bean J. L. et al., 2011, *ApJ*, 743, 92
- Beatty T. G., Madhusudhan N., Tsiaras A., Zhao M., Gilliland R. L., Knutson H. A., Shporer A., Wright J. T., 2017, *AJ*, 154, 158
- Bell T. J. et al., 2019, *MNRAS*, 489, 1995
- Ben-Yami M., Madhusudhan N., Cabot S. H. C., Constantinou S., Piette A., Gandhi S., Welbanks L., 2020, *ApJ*, 897, L5
- Borsa F. et al., 2020, *A&A*, 645, A24
- Bourrier V. et al., 2020a, *A&A*, 635, A205
- Bourrier V. et al., 2020b, *A&A*, 637, A36
- Brogi M., Line M. R., 2019, *AJ*, 157, 114
- Brown T. M., 2001, *ApJ*, 553, 1006
- Cabot S. H. C., Madhusudhan N., Welbanks L., Piette A., Gandhi S., 2020, *MNRAS*, 494, 363
- Carter A. L. et al., 2020, *MNRAS*, 494, 5449
- Cauley P. W., Shkolnik E. L., Ilyin I., Strassmeier K. G., Redfield S., Jensen A., 2019, *AJ*, 157, 69
- Charbonneau D., Brown T. M., Noyes R. W., Gilliland R. L., 2002, *ApJ*, 568, 377
- Claret A., 2000, *A&A*, 363, 1081
- Daylan T. et al., 2021, *AJ*, 161, 131
- Delrez L. et al., 2016, *MNRAS*, 458, 4025
- Deming D. et al., 2013, *ApJ*, 774, 95
- Dumusque X., Boisse I., Santos N. C., 2014, *ApJ*, 796, 132
- Espinoza N. et al., 2019, *MNRAS*, 482, 2065
- Evans T. M. et al., 2016, *ApJ*, 822, L4
- Evans T. M. et al., 2017, *Nature*, 548, 58
- Evans T. M. et al., 2018, *AJ*, 156, 283
- Fortney J. J., Lodders K., Marley M. S., Freedman R. S., 2008, *ApJ*, 678, 1419
- Gandhi S., Madhusudhan N., 2019, *MNRAS*, 485, 5817
- Genz A., Bretz F., 2009, *Computation of Multivariate Normal and Probabilities, Vol. 195*, Springer-Verlag, Berlin, Germany
- Gibson N. P., Aigrain S., Roberts S., Evans T. M., Osborne M., Pont F., 2012, *MNRAS*, 419, 2683
- Gibson N. P., Aigrain S., Barstow J. K., Evans T. M., Fletcher L. N., Irwin P. G. J., 2013a, *MNRAS*, 428, 3680
- Gibson N. P., Aigrain S., Barstow J. K., Evans T. M., Fletcher L. N., Irwin P. G. J., 2013b, *MNRAS*, 436, 2974
- Gibson N. P., Nikolov N., Sing D. K., Barstow J. K., Evans T. M., Kataria T., Wilson P. A., 2017, *MNRAS*, 467, 4591
- Gibson N. P., de Mooij E. J. W., Evans T. M., Merritt S., Nikolov N., Sing D. K., Watson C., 2019, *MNRAS*, 482, 606
- Gibson N. P. et al., 2020, *MNRAS*, 493, 2215
- Gillett F. C., Low F. J., Stein W. A., 1969, *ApJ*, 157, 925
- Gilliland R. L., Arribas S., 2003, High Signal-to-Noise Differential NICMOS Spectrophotometry, Space Telescope NICMOS Instrument Science Report, Space Telescope Science Institute, Baltimore, MD
- Gillon M. et al., 2012, *A&A*, 542, A4
- Harris C. R. et al., 2020, *Nature*, 585, 357
- Hauschildt P. H., Baron E., Allard F., 1997, *ApJ*, 483, 390
- Hauschildt P. H., Allard F., Baron E., 1999, *ApJ*, 512, 377
- Herman M. K., de Mooij E. J. W., Jayawardhana R., Brogi M., 2020, *AJ*, 160, 93
- Hoeijmakers H. J. et al., 2019, *A&A*, 627, A165
- Hoeijmakers H. J. et al., 2020, *A&A*, 641, A123
- Hook I. M., Jørgensen I., Allington-Smith J. R., Davies R. L., Metcalfe N., Murowinski R. G., Crampton D., 2004, *PASP*, 116, 425
- Hooton M. J., de Mooij E. J. W., Watson C. A., Gibson N. P., Galindo-Guil F. J., Clavero R., Merritt S. R., 2019, *MNRAS*, 486, 2397
- Hubeny I., Burrows A., Sudarsky D., 2003, *ApJ*, 594, 1011
- Huitson C. M. et al., 2013, *MNRAS*, 434, 3252
- Hunter J. D., 2007, *Comput. Sci. Eng.*, 9, 90
- Husser T. O., Wende-von Berg S., Dreizler S., Homeier D., Reiners A., Barman T., Hauschildt P. H., 2013, *A&A*, 553, A6
- Jackson B., Adams E., Sandidge W., Kreyche S., Briggs J., 2019, *AJ*, 157, 239
- Kirkpatrick J. D. et al., 1999, *ApJ*, 519, 802
- Kitzmann D. et al., 2018, *ApJ*, 863, 183
- Knutson H. A., Howard A. W., Isaacson H., 2010, *ApJ*, 720, 1569
- Komacek T. D., Showman A. P., 2020, *ApJ*, 888, 2
- Kreidberg L. et al., 2014, *ApJ*, 793, L27
- Line M. R. et al., 2013, *ApJ*, 775, 137
- Lothringer J. D., Barman T., 2019, *ApJ*, 876, 69
- Lothringer J. D., Barman T. S., 2020, *AJ*, 159, 289
- Lothringer J. D., Barman T., Koskinen T., 2018, *ApJ*, 866, 27
- Lothringer J. D., Rustamkulov Z., Sing D. K., Gibson N. P., Wilson J., Schlaufman K. C., 2020a, preprint ([arXiv:2011.10626](https://arxiv.org/abs/2011.10626))
- Lothringer J. D., Fu G., Sing D. K., Barman T. S., 2020b, *ApJ*, 898, L14
- MacDonald R. J., Madhusudhan N., 2017, *MNRAS*, 469, 1979
- Madhusudhan N., Mousis O., Johnson T. V., Lunine J. I., 2011, *ApJ*, 743, 191
- Magic Z. et al., 2013, *A&A*, 557, A26
- Mandel K., Agol E., 2002, *ApJ*, 580, L171
- Merritt S. R. et al., 2020, *A&A*, 636, A117
- Mikal-Evans T. et al., 2019, *MNRAS*, 488, 2222
- Mikal-Evans T., Sing D. K., Kataria T., Wakeford H. R., Mayne N. J., Lewis N. K., Barstow J. K., Spake J. J., 2020, *MNRAS*, 496, 1638
- Mollière P., van Boekel R., Dullemond C., Henning T., Mordasini C., 2015, *ApJ*, 813, 47
- Nelder J. A., Mead R., 1965, *Comput. J.*, 7, 308
- Nikolov N., Sing D. K., Gibson N. P., Fortney J. J., Evans T. M., Barstow J. K., Kataria T., Wilson P. A., 2016, *ApJ*, 832, 191
- Nikolov N. et al., 2018, *Nature*, 557, 526
- Nugroho S. K., Kawahara H., Masuda K., Hirano T., Kotani T., Tajitsu A., 2017, *AJ*, 154, 221
- Nugroho S. K., Gibson N. P., de Mooij E. J. W., Watson C. A., Kawahara H., Merritt S., 2020, *MNRAS*, 496, 504
- Parmentier V., Guillot T., 2014, *A&A*, 562, A133
- Parmentier V., Showman A. P., Lian Y., 2013, *A&A*, 558, A91
- Parmentier V. et al., 2018, *A&A*, 617, A110
- Parviainen H., Aigrain S., 2015, *MNRAS*, 453, 3821
- Perez F., Granger B. E., 2007, *Comput. Sci. Eng.*, 9, 21
- Pinhas A., Rackham B. V., Madhusudhan N., Apai D., 2018, *MNRAS*, 480, 5314
- Pino L. et al., 2020, *ApJ*, 894, L27
- Pont F. et al., 2007, *A&A*, 476, 1347
- Pont F., Sing D. K., Gibson N. P., Aigrain S., Henry G., Husnoo N., 2013, *MNRAS*, 432, 2917
- Rackham B. V., Apai D., Giampapa M. S., 2019, *AJ*, 157, 96
- Rasmussen C. E., Williams C. K. I., 2005, Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning). The MIT Press, Cambridge, MA
- Reiners A., Homeier D., Hauschildt P. H., Allard F., 2007, *A&A*, 473, 245
- Ridgway S. T., 1974, *ApJ*, 187, L41
- Salz M., Schneider P. C., Fossati L., Czesla S., France K., Schmitt J. H. M. M., 2019, *A&A*, 623, A57
- Seager S., Sasselov D. D., 2000, *ApJ*, 537, 916
- Sedaghati E. et al., 2017, *Nature*, 549, 238
- Shah A., Wilson A., Ghahramani Z., 2013, Bayesian Optimization using Student-t Processes, NIPS Workshop on Bayesian Optimization, [Available at: http://www.cs.cmu.edu/~andrewgw/tptoptimisation.pdf](http://www.cs.cmu.edu/~andrewgw/tptoptimisation.pdf)
- Shah A., Wilson A., Ghahramani Z., 2014, PMLR. Reykjavik, Iceland, p. 877. [Available at: http://proceedings.mlr.press/v33/shah14.html](http://proceedings.mlr.press/v33/shah14.html)
- Sing D. K. et al., 2011, *MNRAS*, 416, 1443
- Sing D. K. et al., 2015, *MNRAS*, 446, 2428
- Sing D. K. et al., 2016, *Nature*, 529, 59
- Sing D. K. et al., 2019, *AJ*, 158, 91
- Spake J. J. et al., 2018, *Nature*, 557, 68
- Stevenson K. B. et al., 2010, *Nature*, 464, 1161

Stevenson K. B., Bean J. L., Seifahrt A., Désert J.-M., Madhusudhan N., Bergmann M., Kreidberg L., Homeier D., 2014, *AJ*, 147, 161
 Swain M. R., Vasisth G., Tinetti G., Bouwman J., Chen P., Yung Y., Deming D., Deroo P., 2009, *ApJ*, 690, L114
 Ter Braak C. J. F. T., 2006, *Stat. Comput.*, 16, 239
 Ter Braak C. J. F., Vrugt J. A., 2008, *Stat. Comput.*, 18, 435
 Tracey B. D., Wolpert D., 2018, *Upgrading from Gaussian Processes to Student's-T Processes*. Available at: <https://arc.aiaa.org/doi/pdf/10.2514/6.2018-1659>
 Virtanen P. et al., 2020, *Nat. Methods*, 17, 261
 von Essen C. et al., 2019, *A&A*, 628, A115

Wakeford H. R. et al., 2018, *AJ*, 155, 29
 Wallace L., Prather M., Belton M. J. S., 1974, *ApJ*, 193, 481
 West A. A., Bochanski J. J., Bowler B. P., Dotter A., Johnson J. A., Lépine S., Rojas-Ayala B., Schweitzer A., 2011, in Johns-Krull C., Browning M. K., West A. A., eds, ASP Conf. Ser. Vol. 448, 16th Cambridge Workshop on Cool Stars, Stellar Systems, and the Sun. Astron. Soc. Pac., San Francisco, p. 531
 Wilson J. et al., 2020, *MNRAS*, 497, 5155

APPENDIX A:

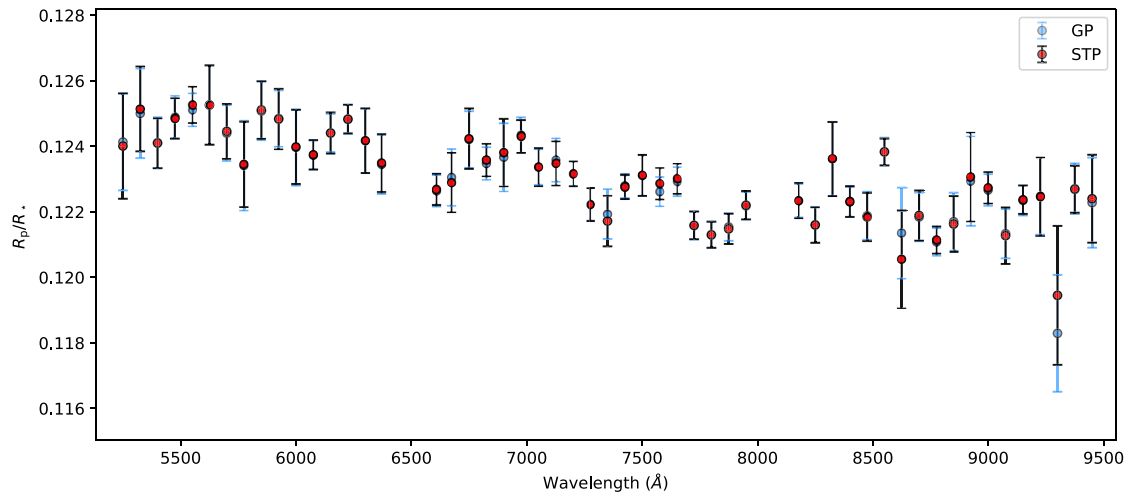


Figure A1. GMOS transmission spectrum of WASP-121b recovered from our GP and STP analyses. The blue points are the results of the GP analysis, whilst the red points are the result of the STP analysis. Both analyses used the same priors for the white light curves.

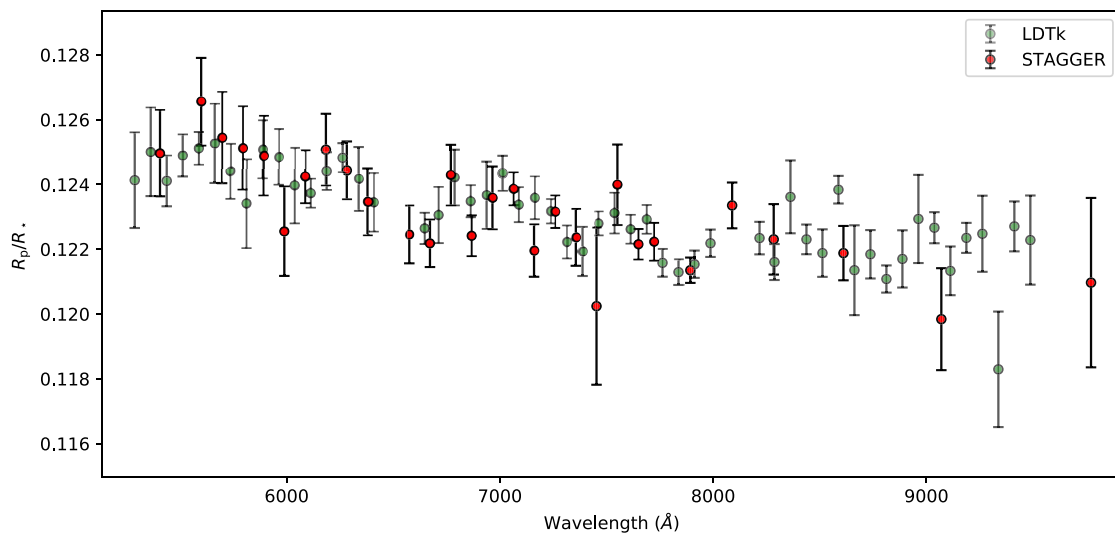


Figure A2. Our original GMOS transmission spectrum of WASP-121b compared with the result of a re-analysis using identical bins and the same STAGGER 3D limb darkening coefficients as in Evans et al. (2018). The green points show our original transmission spectrum whilst the red points are the result of the re-analysis.

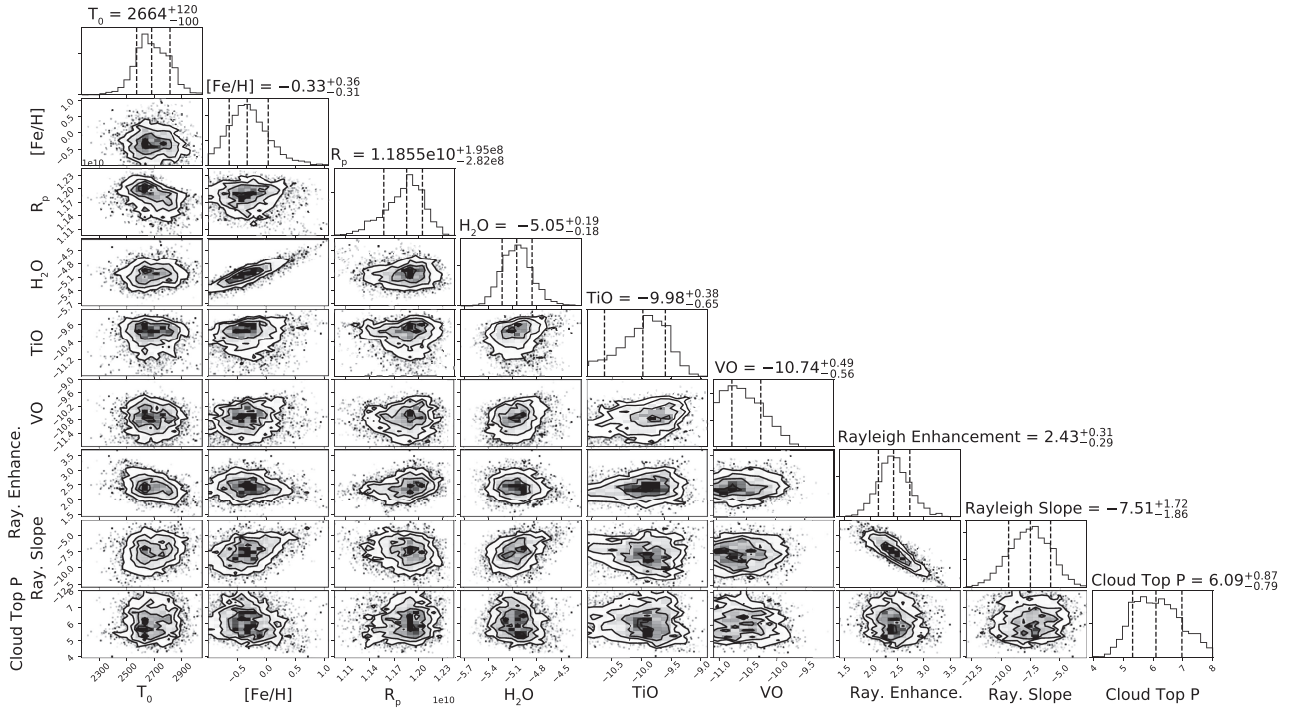


Figure A3. Marginalized posterior probability densities for the atmospheric retrieval performed using the PETRA code on the combined optical-infrared transmission spectrum of WASP-121b.

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