

Optimising spectroscopic observations of transiting exoplanets

Linn Boldt-Christmas¹, Fabio Lesjak², Ansgar Wehrhahn¹, Nikolai Piskunov¹, Adam D. Rains¹, Lisa Nortmann², and Oleg Kochukhov¹

¹ Observational Astrophysics, Department of Physics and Astronomy, Uppsala University, Sweden
e-mail: linn.boldt-christmas@physics.uu.se

² Institut für Astrophysik und Geophysik, Georg-August-Universität, Friedrich-Hund-Platz 1, 37077 Göttingen, Germany

Received 7 July 2023 / Accepted 13 December 2023

ABSTRACT

Context. When observing the atmospheres of transiting exoplanets using high-resolution spectroscopy, the aim is to detect well-resolved spectral features with high signal-to-noise ratios (S/Ns), as is possible today with modern spectrographs. However, obtaining such high-quality observations comes with a trade-off: a lower cadence of fewer, longer exposures across the transit collects more photons thanks to reduced overheads, enhancing the S/N of each observation, while a higher cadence of several shorter exposures minimises spectral feature smearing due to the continuously changing radial velocity of the planet.

Aims. Considering that maximising S/N and minimising smearing are both beneficial to analysis, there is a need to identify the optimal compromise between the two for a given target. In this work, we aim to establish where this compromise lies for a typical exoplanet transit observation in order to benefit future data collection and subsequent interpretation.

Methods. We modelled real transit events based on targets as they would be observed with VLT/CRIRES+ at Paranal Observatory, Chile. Creating four hypothetical scenarios, we simulated each set of transmission spectra across 100 realisations of the same transit event in order to vary the time resolution only. We removed telluric and stellar lines from these data sets using the SYSREM algorithm and analysed them through cross-correlation with model templates, measuring how successfully each time resolution and case detected the planetary signal and exploring how the results vary.

Results. We demonstrate that there is a continuous change in the significance of the cross-correlation detection based on the trade-off between high and low time resolutions, and that, averaged over a large number of realisations, the function of this significance has clear maxima. The strength and location of these maxima vary depending on, for example, planet system parameters, instrumentation, and the number of removal iterations. We discuss why observers should therefore take several factors into account using a strategy akin to the ‘exposure triangle’ employed in traditional photography where a balance must be struck by considering the full context of the observation. Our method is robust and may be employed by observers to estimate the best observational strategies for other targets.

Key words. methods: observational – methods: statistical – techniques: spectroscopic – planets and satellites: atmospheres – planetary systems – infrared: planetary systems

1. Introduction

As exoplanets orbit their host stars, the chemical and physical structures of their atmospheres can be studied using spectroscopic observations. New exoplanet candidates are continually being discovered by transit-searching surveys, including both ground-based surveys, such as WASP (Pollacco et al. 2006), HATNet (Bakos et al. 2004), HATSouth (Bakos et al. 2013), KELT (Pepper et al. 2007), TRAPPIST (Jehin et al. 2011), and MASCARA (Talens et al. 2017), and space-based surveys, such as Kepler (Borucki et al. 2010), K2 (Howell et al. 2014), and TESS (Ricker et al. 2015). Together, such surveys have resulted in over 5500 confirmed exoplanets on record today, with over 4500 of those being discovered in the last ten years¹. Thanks to this progress, alongside our increasing ability to study these planets in detail with new instruments and improvements in our methods, the relatively new field of exoplanet atmosphere characterisation is rapidly maturing.

Spectroscopic characterisation can be carried out through transmission observations of the nightsides of exoplanets or through emission and/or reflection observations of their daysides. Most confirmed exoplanetary systems are observed

edge-on, meaning that planets can be seen regularly transiting across the face of their host stars (a primary eclipse), during which we observe the dusk–dawn terminator on their nightsides and collect light that has been transmitted through the upper layers of the planet atmosphere. Beyond this eclipse, the planets can also be studied as they continue along their orbit and are illuminated by their host star, with this dayside reflecting light from the host star together with whatever emission features come from the irradiated planet. This type of observation is ideally performed immediately prior to or following the moment the planet disappears behind the host (a secondary eclipse), but is possible during a much larger window of opportunity in time and has the advantage that it can be done for non-transiting systems too.

As such, transmission studies face unique challenges in their physical time constraints. The two branches of transit spectroscopy, namely low-resolution from space and high-resolution from the ground, both provide impressive results thanks to their specific and complementary strengths. Low spectral resolution has a drawback in that one cannot distinguish individual spectral lines, but space-based observations benefit from the lack of telluric contamination and flux-calibrated spectra with high photometric precision across a wide spectral range. This means that the contrast between wavelengths with strong absorption (where the exoplanet atmosphere is opaque) and those with

¹ NASA Exoplanet Archive: <https://exoplanetarchive.ipac.caltech.edu/>

little absorption is diminished with low spectral resolution, but the remaining effect can be detected with high photometric precision. This is possible from space, and characterisation has been regularly achieved with space-based telescopes such as the *Hubble* Space Telescope (as Seager & Sasselov (2000) predicted would be possible) for over 20 years now (e.g. Charbonneau et al. 2002; Vidal-Madjar et al. 2004; Swain et al. 2008; Pont et al. 2008; Sing et al. 2008, 2011; Kreidberg et al. 2014, 2018; Stevenson 2016; MacDonald & Madhusudhan 2017; Benneke et al. 2019; Zhou et al. 2022) and more recently with JWST (e.g. Rustamkulov et al. 2023; JWST Transiting Exoplanet Community Early Release Science Team 2023).

Comparatively, the mirror size and associated instrumentation required for high spectral resolution observations across a large wavelength range poses an engineering challenge that is currently unachievable (in part financially) from space, and therefore these observations are restricted to being ground-based. High spectral resolution allows lines to be more easily separated, and for ground-based observations, where spectral resolution can be high across a large wavelength range, the high contrast between the effective blocking area of the planet in different wavelengths provides the ability to look between the cores of strong telluric absorption lines that dominate most of the near-infrared. High-resolution ground-based results have been consistently obtained for over a decade (e.g. Redfield et al. 2008; Snellen et al. 2008, 2010; Bean et al. 2010; Astudillo-Defru & Rojo 2013; Di Gloria et al. 2015; Allart et al. 2017; Brogi et al. 2018; Diamond-Lowe et al. 2018; Merritt et al. 2021; Nikolov et al. 2021; Maimone et al. 2022; Maguire et al. 2023), and more recently, even in collaboration with space-based results (e.g. Boucher et al. 2023; Spyros et al. 2023).

Together, these efforts are contributing to an expanding catalogue of confirmed exoplanets and an evolving understanding of their composition, revealing a large diversity in the architecture of exoplanetary systems. With the advent of highly stable spectrographs on the largest ground-based telescopes, combining wide wavelength range with high spectral resolution, such as HARPS (Mayor et al. 2003), IGRINS (Park et al. 2014), SPIRou (Artigau et al. 2014), MAROON-X (Seifahrt et al. 2020), ESPRESSO (Pepe et al. 2021), and CRIRES+ (Dorn et al. 2023), the number of exoplanets well characterised with high-resolution spectroscopy is steadily increasing. Cross-correlation analysis with realistic templates initially made it possible to perform analyses of the atmospheric composition and basic dynamics of hot Jupiters (Brogi et al. 2016; Hoeijmakers et al. 2018), and today, the quality of observations and the advances of the analysis techniques (for many instruments, including benefits from adaptive optics systems) have reached a point where we can even speculate about more detailed meteorological effects. This is possible because we are starting to identify horizontal and vertical stratification on exoplanets as well as dynamic weather patterns as inferred from the shape of individual lines, which are resolvable with high-quality, high-resolution spectroscopy (e.g. Ehrenreich et al. 2020; Pino et al. 2022; Prinot et al. 2022; Gandhi et al. 2023; Yan et al. 2023).

This ability to identify spectral features in detail is not only enabling more detailed characterisation, but is also relevant as studies indicate that the occurrence rates of exoplanets are high around cooler stars such as FGK stars (Kunimoto & Matthews 2020) and M-dwarfs (Hardegree-Ullman et al. 2019; Kanodia et al. 2019). As the photospheres of cooler stars allow complex molecular species to exist, these colder planet-hosting stars have more chemically complicated atmospheres whose absorption features flood the continuum, creating greater

challenges for disentangling the planetary and stellar spectra (Wakeford et al. 2019) that may benefit from well-resolved planetary features. M-dwarfs also see an overabundance of low-mass planets with shorter periods in particular (Sabotta et al. 2021), meaning they are especially suitable for transit studies.

Great efforts are going into extending the success so far achieved in characterising gas giants to the domain of low-mass planets in potentially habitable zones. This task remains a challenge for (at least) observations and so in the present paper we explore one important question of high-resolution transit spectroscopy: how to obtain the best possible data by considering the trade-off between signal-to-noise ratio (S/N) and the time resolution of observations.

During the observation of a transit, there are two possible strategies that may be employed: exposures can either be taken at a lower cadence with fewer, longer exposures or at a higher cadence of several shorter exposures. Across the transit, longer exposures collect more photons thanks to reduced overheads, which enhances the S/N of each exposure. Meanwhile, shorter exposures minimise the effect whereby spectral features get smeared due to the continuously changing radial velocity of the planet across the single exposure. Considering that both maximising the S/N per exposure and minimising the effect of smearing will benefit analysis, there is a need to identify the optimal compromise between the two for a given target.

In the present paper, we explore this balance by investigating what range of time resolutions results in the strongest possible planetary signal with traditional cross-correlation analysis (as a measure of what can be considered an ‘optimal’ time resolution). We demonstrate that there is a continuous change in the significance of the cross-correlation detection based on the trade-off between a small number of higher-S/N exposures and a large number of lower-S/N exposures of a transiting target, and that averaged over a large number of realisations, the function of this significance has a clear maximum between the two. By testing different case studies, we find that the locations of these maxima depend on a number of factors, including target and system parameters. With this work, we are also presenting a robust method to determine the ideal time resolution for observing other given targets using simulated spectra. Using the methods presented here may help observers plan what exposure cadence to use prior to obtaining their data, therefore optimising their own transit spectroscopy observations.

In Sect. 2, we describe how we explored this question using model spectra (i.e. simulated observations). We describe how these spectra are generated to create a set of data that is representative of the observations obtainable with CRIRES+ while still maintaining full control of external variables for the purpose of comparability and repeatability. In Sect. 3, we describe how our simulated observations are then analysed (following the same methodology that would be used on real data), removing the stellar signal and the telluric contamination with the commonly used SYSREM method before performing cross-correlation analysis on the remaining planetary signal. We present these results in Sect. 4, followed by some further discussion regarding other considerations in Sect. 5. Final conclusions and recommendations are presented in Sect. 6.

2. Method

2.1. Time resolution optimisation problem

A transmission spectrum observed from Earth consists of three components: (i) stellar spectral lines from the exoplanet host

star, (ii) telluric spectral lines from the Earth's atmosphere, and (iii) the minor contribution of spectral lines from the exoplanet atmosphere. Importantly, the spectral and telluric lines are relatively stationary across the transit in comparison to the lines of the orbiting exoplanet; as the exoplanet moves towards and then away from the observer, its lines will vary from being blueshifted at the beginning of its transit and redshifted towards the end. In analysis, this difference in Doppler shifts between lines of different origins is exploited to help distinguish the planetary component from the stellar and telluric component (see Sect. 3.1).

The planetary component is only found in the spectra of exposures taken during the actual eclipse of the host star by the moving planet. This is limited in time, and so it sets a firm limitation on the possible maximum S/Ns of the data collected with a given instrument. In practice, observations are also taken before and after the transit to assess the stability of the instrument and to extend the baseline for following data trends such as the strong telluric features, but this data does not contain any planetary signal.

Consider the extreme case where a transit observation consists of a single exposure over an entire transit. In this case, light collection and thus S/N is maximised (much like taking a long-exposure photograph in the dark on Earth) but all information about radial velocity shifts of the planet is lost as one cannot see the Doppler shift of the planetary component. This would render the data useless because subsequent analysis will not be able to tell apart planetary, telluric, and stellar features. The total transit exposure must therefore be split up into some number of subdivisions in order to track this shift, even if this results in some loss of S/N.

On the other end of extremes, in the case of too many subdivisions, there must also be a lower limit of S/N below which the cross-correlation signal will be dominated by the noise. This is because if one ignores readout noise and exposure overheads, the total S/N is (approximately) preserved while subdividing exposures; but in reality, what sets the limit is the background and the readout noise, and both grow linearly in combined data with the number of subexposures.

Establishing an appropriate number of exposures across the transit is not straightforward. Observing a transit at a lower time resolution, meaning we take a smaller number of longer exposures, increases total light collection per exposure as less time is lost to overheads, which increases S/N. However, this also results in an effect where resolved spectral features captured over a longer period are ‘smeared’ during the exposure due to the changing radial velocity of the moving target. A schematic illustrating how this smearing effect arises is shown in Fig. 1.

This smearing effect makes analysis and the identification of spectral lines more challenging or even impossible, especially when trying to retain the true line profile shape for more detailed atmospheric characterisation. Conversely, at a higher time resolution, meaning a larger number of shorter exposures, there is less smearing of spectral features but more time is lost in overheads between exposures (decreasing total exposure time) which gives a lower total light collection and lower S/N on the single exposure. As retaining line profiles with a high time resolution and obtaining high S/N with a low time resolution are both beneficial for analysis, there is a need to strike a balance between the two.

2.2. Generating model CRIRES+ spectra

Our objective is to analyse observational data of different time resolutions to determine where the balance between high and

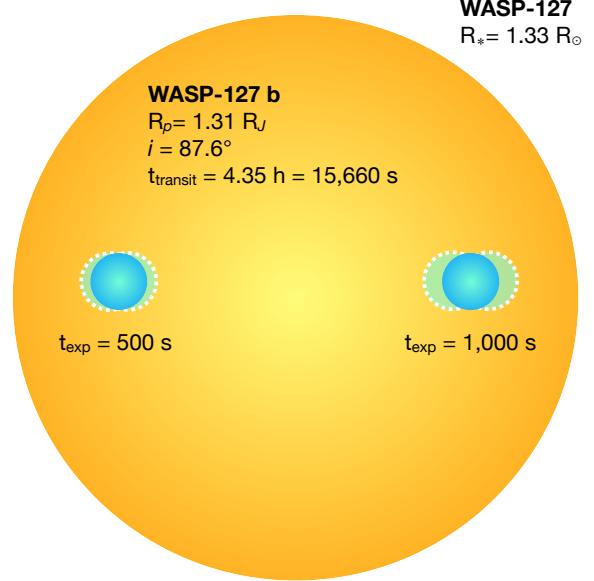


Fig. 1. Schematic demonstrating how the smearing effect arises due to increased exposure length. Using an illustration of exoplanet WASP-127 b and its host star (to scale) as an example, consider the target's transit time $t_{\text{transit}} = 4.35 \text{ h} = 15\,660 \text{ s}$. A shorter exposure time of $t_{\text{exp}} = 500 \text{ s}$ (left) covers approx. 3% of the total transit time, versus a longer exposure time of $t_{\text{exp}} = 1000 \text{ s}$ (right) which covers approx. 6%. In this schematic, the location of the planet at the beginning and end of an exposure of length t_{exp} is shown by the shaded, dotted regions. The relative distance between these regions illustrates how much the target moves across a single exposure of the given time. This movement is greater for the longer exposure and thus will give rise to a larger smearing effect.

low exposure cadence lies. In order to analyse an ideal scenario with all uncertainties controlled, we produced simulated model spectra to conduct our analysis on. This is because if one was to attempt an investigation of our time resolution optimisation problem using real observational data only, the immediate obstacle would be that one cannot test different time resolutions simultaneously. Consequently, one would have to settle for observations of the same target using different time resolutions taken across different nights – and all would have different weather conditions, visibility, lunar contamination, etc. As such, using simulated spectra ensures that all other parameters unrelated to our study are fully controlled, and allows us to play with the level of the planetary signal for different targets to follow detection deterioration.

As the results are instrument- and target-specific, the model spectra were designed to be representative of the type and quality of data that can be delivered by CRIRES+ at the VLT at Paranal Observatory in Chile. CRIRES+ is the upgrade project of the previous CRIRES instrument (CRyogenic InfraRed Echelle Spectrograph, in operation until 2014) that was recently completed with first science observations conducted in October 2021. CRIRES+ is a cross-dispersed spectrograph with a nominal spectral resolution of up to approximately $R = 100\,000$ across wavelengths from $0.95 \mu\text{m}$ to $5.3 \mu\text{m}$ ($YJHKLM$ bands), covering the near-infrared region and partial mid-infrared region. This range is vital for resolving the spectral lines of several molecular species (such as CO, CO₂, H₂O, NH₃, and CH₄), and notably, CRIRES+ is rare in its ability to cover the domain of $3.0 \mu\text{m}$ to $5.2 \mu\text{m}$ at high-resolution (Dorn et al. 2023). The range of CRIRES+ is also highly complementary with the spectroscopic

capabilities of the recently launched JWST, namely the NIRSpec instrument which covers $0.6\text{ }\mu\text{m}$ to $5.3\text{ }\mu\text{m}$ (Birkmann et al. 2022; Jakobsen et al. 2022).

Each simulated data set corresponds to a real transit event of a hypothetical target planet as visible from Paranal, and comes with realistic time-dependent local conditions such as air mass variation, barycentric velocities for each exposure, and the telluric absorption computed for the zenith distance of the target at the moment of observations. Target and system parameters were taken from a selected case study target, WASP-127 b. This target is a puffy gas giant of radius $1.31 R_J$ and mass $0.165 M_J$, found at a semi-major axis distance of 0.0484 au from its G5 host star with an orbital period of 4.18 days and a transit duration of 4.35 h (Seidel et al. 2020b) meaning it can reasonably be studied using a single transit.

The transit observations were simulated using a custom script² that generates spectral data by following the path of the light from the stellar surface to the spectrograph detector. All simulation parameters can be found in Table 1.

Our spectral simulator first sets up the Keplerian orbits of both the planet and host star around the barycentre according to the catalogued ephemerides, and the orbits are used to determine positions and velocities of the two bodies for each exposure. An appropriate stellar template spectrum I_\star is then selected from the PHOENIX spectral library (Husser et al. 2013). For the planetary spectrum, we use a reference exoplanet atmosphere transmission spectrum that has been computed using the radiative transfer package petitRADTRANS (Mollière et al. 2019) and the HITEMP line list for CO and H₂O (Rothman et al. 2010), modelling a theoretical atmosphere based on parameters for a generic hot Jupiter atmosphere (Bouchy et al. 2005; Addison et al. 2019). This planetary spectrum is then imprinted onto the stellar spectrum to create the simulated total transmission spectrum, I_{tr} :

$$I_{\text{tr}}(\lambda, t) = I_\star(\lambda, t) - I_b(\lambda, t) \quad (1)$$

where I_\star is the template that represents the out-of-transit or baseline stellar spectrum, and I_b is the stellar flux that is blocked by the planetary disk and its surrounding atmosphere i.e. the spectrum of exoplanetary absorption features. Spectra are initially in units of energy/wavelength bin ($\text{erg cm}^{-2} \text{ s}^{-1} \text{ cm}^{-1}$) and are converted into units of photons per pixel at a later step. At the end, planetary and stellar components of I_{tr} are Doppler-shifted according to the radial velocities of the planet, star, and barycentric velocity.

To calculate I_b , the planet can be thought of as a single opaque disk with a wavelength-dependent variable for the blocked radius $R_b(\lambda)$. The area of this radius covers a region of the stellar surface, with a representative spectrum of the blocked stellar region given by $I_{\star b}(\lambda, t)$ for each exposure during the transit. This blocked spectrum varies in intensity and rest frame from the overall stellar spectrum I_\star due to the limb darkening and rotation of the stellar surface, with the varying amount of occultation during the ingress and egress also accounted for here. Thus, the radiation that is blocked by the planet and its atmosphere I_b can be expressed as:

$$I_b(\lambda, t) = I_{\star b}(\lambda, t) \cdot \left(\frac{R_b(\lambda)}{R_\star} \right)^2. \quad (2)$$

Our spectral simulator makes the following assumptions: (i) the star has no surface inhomogeneities such as stellar spots

Table 1. All parameters used for generating simulated spectra of our reference case study, a fiducial planet (based on exoplanet WASP-127 b) as would be observed with VLT/CRIRES+.

Star (WASP-127)		
Parameter	Units	Value
Mass ^(a)	$M_*(M_\odot)$	0.950 ± 0.020
Radius ^(a)	$R_*(R_\odot)$	1.333 ± 0.027
Spectral class ^(b)		G5
Magnitude (K -band) ^(c)	m_{K_s}	8.641 ± 0.019
Effective temperature ^(d)	T_{eff} (K)	$5828.010^{+107.743}_{-104.525}$
Metallicity ^(d)	(M/H)	-0.193 ± 0.014
Log. gravity ($\log g$) ^(d)	\log_{10} (cm s $^{-2}$)	$4.200^{+0.0771}_{-0.064}$
Systemic radial velocity ^(e)	(km s $^{-1}$)	-8.248 ± 0.892
Planet (WASP-127 b)		
Parameter	Symbol	Value
Mass ^(a)	$M_p(M_J)$	$0.1647^{+0.0214}_{-0.0172}$
Radius ^(a)	$R_p(R_J)$	$1.311^{+0.025}_{-0.029}$
Semi-major axis ^(a)	a (au)	$0.0484^{+0.0014}_{-0.0010}$
Orbital period ^(a)	P (days)	$4.1780620^{+0.0000009}_{-0.0000005}$
Orbital eccentricity ^(a)	e	0
Orbital inclination ^(a)	i (deg)	$87.84^{+0.36}_{-0.33}$
Radial velocity amplitude ^(a)	K (m s $^{-1}$)	22^{+3}_{-2}
Transit duration ^(a)	t_t (h)	$4.3529^{+0.0084}_{-0.0139}$
Transit depth ^(a)	δ (%)	$1.021^{+0.005}_{-0.029}$
Assumed eq. temperature ^(a)	T_{eq} (K)	1400 ± 24
Observing (VLT/CRIRES+)		
Parameter	Symbol	Value
Wavelength setting ^(f)		K2148
Wavelength coverage ^(g)	μm	$2.007\text{--}2.491$
No. of echelle orders ^(f)		6
Air mass variation ^(h)		$1.58\text{--}1.06\text{--}1.99$
Seeing ⁽ⁱ⁾		0.7
Readout time ^(g)	t_{RO} (s)	14

References. ^(a)Seidel et al. (2020b), ^(b)Lam et al. (2017), ^(c)2MASS All-Sky Catalog (Cutri et al. 2003), ^(d)TESS Input Catalog (Stassun et al. 2019), ^(e)Gaia DR2 (Gaia Collaboration 2018), ^(f)ESO/CRIRES+ User Manual, ^(g)Dorn et al. (2023), ^(h)Calculated using `pyasl.airmassPP` from PyAstronomy, ⁽ⁱ⁾Average Paranal seeing in 2019 according to data from the ESO Differential Image Motion Monitor (Kornilov et al. 2007).

or flares and emits a constant limb-darkened spectrum; (ii) the planet surface and atmosphere are homogeneous and produce a time-independent planetary spectrum; (iii) the observation is photon noise limited, and other sources of noise can be ignored. Extinction is accounted for according to the air mass at each exposure. Telluric lines are imprinted in the combined spectrum of star and planet using a synthetic model spectrum of the standard atmosphere at Paranal, which was modeled using Molecfit (Smette et al. 2015; Kausch et al. 2015) and then scaled to the airmass value (without accounting for a scaling of H₂O lines

² <https://github.com/fabioles/Eclspecl>

with the amount of precipitable water). The model also accounts for the Rossiter-McLaughlin effect as the simulation removes the received stellar flux and Doppler shifts this spectrum according to the projected rotational velocity of the stellar surface behind the planet, before subtracting it from the overall spectrum.

All spectra were simulated for observations in the CRIRES+ *K*-band ($2.0\text{ }\mu\text{m}$ to $2.5\text{ }\mu\text{m}$) using wavelength setting K2148. Instrumental effects acting on the spectra include the instrumental profile, which is assumed to be of Gaussian shape with a FWHM corresponding to a spectral resolution of 100 000 (i.e. approximately 3 pixels), and the spectra are scaled according to the blaze functions measured for CRIRES+. For more technical details regarding the specifications of the CRIRES+ instrument, please consult the publicly available CRIRES+ User Manual from ESO³.

The expected S/N per pixel for a given observation was estimated using values computed using the adaptive optics (AO) setup with the CRIRES Exposure Time Calculator⁴, which were fitted to yield the following relation for the *K*-band:

$$S/N = 247.31 \times 10^{-m_K/5} \times 10^{\kappa(1-\text{AM})/5} \cdot \sqrt{t_{\text{exp}}}, \quad (3)$$

with the *K*-band magnitude m_K , the extinction coefficient assumed to be $\kappa = 0.05$, the airmass AM, and the exposure time t_{exp} in seconds. After scaling the spectra according to the S/N for the desired exposure time, artificial photon noise drawn from a normal distribution (with a width of the S/N) was added.

Throughout a transit, the planet's atmospheric lines will be smeared over multiple detector pixels due to the planetary motion. We simulate this effect for longer exposures by computing spectra for several shorter subexposures (assuming no additional readout overhead) that are then co-added with the proper Doppler shifts.

The generated model data differs from real data in that it does not go through the standard CRIRES+ data reduction pipeline. In reality, raw data from CRIRES+ is 2D (in the dispersion and spatial directions) and the reduction pipeline converts these observations into 1D spectra. Our simulations generate 1D spectra and so the 2D extraction step is not needed. Several other effects that are handled by the reduction pipeline – such as correction for dark current, bias, cosmic rays, and other instrumental effects – are not included in the model data.

At each time resolution, 100 realisations of the same transit event were generated in order to create a larger statistical sample to analyse. Between realisations, only random data noise and start time of the observation relative to the transit mid-point (varying by 1 exposure length) changes. Model observations were also generated for imagined variations of this reference case in order to investigate if and how results would vary for other theoretical planets: one case where the planetary signal was halved (fainter target); one case where the transit duration was halved (same semi-major axis, but shorter period); and one case where the target was observed at a lower spectral resolution ($R = 50\,000$ instead of $R = 100\,000$). Observational data was simulated for each of the four cases at nine different time resolutions, ranging from approximately 2–50 min-long exposures, for 100 simulated realisations of the same transit event; in total, this results in 3600 simulations (see Sect. 4).

³ <https://www.eso.org/sci/facilities/paranal/instruments/cires/doc.html>

⁴ <https://etc.eso.org/observing/etc/cires>

3. Analysis

It is important that the simulated observational data is treated as equally as possible to genuine observational data in order to retain realism. As such, the subsequent analysis for the model data follows standard analysis procedures as closely as possible, starting with the removal of stellar and telluric contamination followed by cross-correlation.

3.1. SYSREM

The first step is to isolate the exoplanetary atmosphere spectrum. As described in Sect. 2.1, the planetary component of the transmission spectrum can be distinguished from the stellar and telluric components by its comparatively large radial velocity shift. This shift is sufficiently significant that the movement of planetary lines should be notable from exposure to exposure across the transit (to an extent that naturally depends on the time resolution) while stellar and telluric lines are systematically present with nearly negligible offset between exposures. For example, in the generated spectra for our reference case, the stellar lines shift by 0.82 km s^{-1} during the observation (due to barycentric velocity); the telluric lines are completely stationary; and the planet by $\pm 16.45\text{ km s}^{-1}$ throughout the transit. The difference between these can be used to identify and remove non-planetary lines.

The SYSREM algorithm first described in Tamuz et al. (2005) was originally designed to correct the systematic variations of light curve observations, and it has since been used in a wide range of exoplanet cross-correlation studies for removing stellar and telluric features (e.g. Birkby et al. 2013, 2017; Hawker et al. 2018; Sánchez-López et al. 2019). A major advantage of this algorithm is that it does not require any a priori knowledge of the observational features that might influence the measurements. In the visible wavelength regions that are less contaminated by tellurics, alternative approaches to using iterative algorithms like SYSREM generally include methods that model tellurics with radiative transfer codes (such as Molecfit, mentioned in Sect. 2.2) and then mask the affected spectral regions before dividing out the stellar signal using baseline measurements from before or after the transit (e.g. Allart et al. 2017; McCloat et al. 2021; Mounzer et al. 2022); however, in the infrared wavelengths, the spectrum is so flooded with tellurics that this is often not a practical approach.

SYSREM achieves its goal by representing the data with a model f that consists of two components: a spectrum S that is constant in time (but dependent on wavelength) and the time variation A (constant in wavelength but dependent on time) as expressed by:

$$f(\lambda, t) = S(\lambda) \cdot A(t). \quad (4)$$

Both components S and A are fitted consecutively to the data by minimising the sum of the residuals squared meaning the spectrum S will contain both the stellar and the telluric spectrum, and the time variation A includes several factors like the seeing variation, changes in air mass, etc. The planet atmosphere signal is, however, not included in the model as it changes in time due to the Doppler shift. As such, after removing the SYSREM model f from the observations, one should be left with the planet signal, the residuals of time variability not associated with systematic Doppler shift (e.g. variations of seeing), and the noise. This method is closely related to the method of principal component analysis (PCA) and can be interpreted as removing the N largest components from the observations.

Usually this algorithm is iterated several times to gradually remove features present at the same wavelengths. The number of iterations must be carefully selected; an insufficient number of iterations can result in unwanted contamination being retained, yet an excess of iterations will eventually remove the planet signal, particularly in the phases where the Doppler-shifted planetary lines fall on the wavelengths of strong telluric features. As the algorithm is not formulated as an optimisation problem, there remains an element of subjectivity in selecting the number of iterations. This choice is not always obvious, so one example of a more robust approach involves injecting an artificial (known) signal and testing how many PCA iterations were required to remove it as tested by Cheverall et al. (2023). In this work, we tested a wide range of SYSREM iterations in order to track the impact of selecting this iteration number.

3.2. Cross-correlation

Using cross-correlation is by now a standard practice in the field of characterising exoplanetary atmospheres with high-resolution spectroscopy (e.g. Birkby et al. 2013, 2017; Kok et al. 2014; Brogi et al. 2017; Hawker et al. 2018; Giacobbe et al. 2021; Prineth et al. 2022). After applying SYSREM to remove stellar and telluric signals, one is left with a minute planetary signal buried within the noise in the form of residuals for each exposure. It is these residuals that are then cross-correlated with a template spectrum of the target atmosphere at a range of radial velocity offsets, which effectively combines the signal from all planet atmosphere absorption lines and boosts the S/N of the atmosphere detection.

The template is a simulated transmission spectrum computed using a highly simplified planet atmosphere model and atomic/molecular data of the species expected to be present at the given physical conditions. Provided that the modelled features in the template do in fact appear in the planetary spectrum, a cross-correlation of the observation and template should confirm their presence even though the predicted relative strength of different lines may be incorrect. For real data, the cross-correlation is carried out between this template and the reduced observational data. In this work, the cross-correlation is carried out between the template and the reduced simulated data, for which the planetary signal is modelled using the very same template (created with petitRADTRANS as detailed in Sect. 2.2). This means that in our subsequent analysis, we know that the features definitely exist in our data, giving us the largest possible cross-correlation peak. What we evaluate here is therefore not merely the existence of a cross-correlation detection, but rather its significance relative to the noise of our simulations across the parameter space we intend to explore.

Generally, cross-correlation studies are performed using templates of individual atomic or molecular species at a time, and the strength of each species/template's detection is calculated respectively. Once this is completed, it can be possible to use this information to infer an overall atmospheric composition using a suite of retrievals based on the detection strength of each chemical species, and the fit of this retrieved model can be measured by then cross-correlating the observed spectrum with the retrieved spectrum as shown by e.g. Brogi & Line (2019); Gibson et al. (2020); Lesjak et al. (2023); Prineth et al. (2023). Effectively, in this work, as we are certain that we know the true global transmission spectrum *a priori* (i.e. our template), it is this final step of likelihood-fitting that is being simulated and whose performance is measured.

The weighted cross-correlation function (CCF) dependent on velocity v and time t can be expressed as:

$$\text{CCF}(v, t) = \sum_{i=0}^N \frac{R_i(t) \cdot M_i(v)}{\sigma_{R,i}(t)^2} \quad (5)$$

where i is the pixel index, R is the spectrum of residuals (after SYSREM), M is the template spectrum from petitRADTRANS, and σ_R is the uncertainty of R . σ_R is usually obtained from the pipeline, but for our model, this is estimated by assuming a constant error of every pixel and subsequently applying the same correction during the steps of normalisation and SYSREM to these as to the data.

For each exposure, the template spectrum is shifted in 1 km s^{-1} steps across a range of $\pm 200 \text{ km s}^{-1}$ around the systemic velocity v_{sys} of the host star. The top panel of Fig. 2 shows a plot of all exposures across the planet's orbital phases along the vertical axis (i.e. time, where each row is one exposure). At each v_{sys} step along the horizontal axis, the cross-correlation value can be determined with the SYSREM residuals of each exposure, where a high value indicates matching spectral features between those in the observation and those in the template spectrum. Thus, the transit of the planet reveals itself as a diagonal line of high cross-correlation values: the first, bottom rows of exposures show no significant cross-correlation peak (pre-transit). This is followed by a line of high cross-correlation values, which is slanted due to the orbital motion of the planet being first blueshifted (negative velocities) and then redshifted (positive velocities). Finally, the top rows of exposures again show no significant cross-correlation peak (post-transit).

By shifting every cross-correlation function in this plot to the planetary rest frame, the slanted line becomes vertical. The shift corresponds to the planetary radial velocity semi-amplitude ($K_p = 126 \text{ km s}^{-1}$, which can be calculated from orbital parameters of WASP-127 b) and collapsing the plot vertically will then amplify the cross-correlation peak. Without a priori setting the correct K_p , one can also try different values of K_p from a large range (e.g. 0 – 400 km s^{-1}); this gives results that will each be similar to the top panel of Fig. 2 but with different slant angles, and realisations producing more vertical lines will result in narrower and higher peak after collapsing the image vertically.

Stacking the results of different shifts as sorted by K_p produces an image similar to the middle panel of Fig. 2. This is a ‘detection map’ as the planetary atmosphere detection will show up as a bright, central region that is centred at $v_{\text{sys}} = 0 \text{ km s}^{-1}$ and $K_p = 126 \text{ km s}^{-1}$. Overlaying each K_p value in a single plot in this way gives an intuitive, visual representation of the detection strength as the brightness indicates where cross-correlation values are high, i.e. where there is a peak in the cross-correlated S/N, as seen at the bottom of Fig. 2.

3.3. Measure of optimisation

Considering this work is in the pursuit of ‘optimal’ observations of an exoplanet atmosphere, we must explicitly describe what we mean by this potentially subjective term. As the field of exoplanetary atmosphere characterisation is presently largely focused on attempts to identify and characterise chemical species present in exoplanetary atmospheres using comparisons to templates, the definition employed in this paper for what constitutes as an ‘optimal observation’ is: whatever time resolution results in a cross-correlation detection map, i.e. the K_p – v_{sys} plot, with the most significant detection. Using this value as a proxy for ‘optimal’ is valid as it is not only the value arguably most relevant

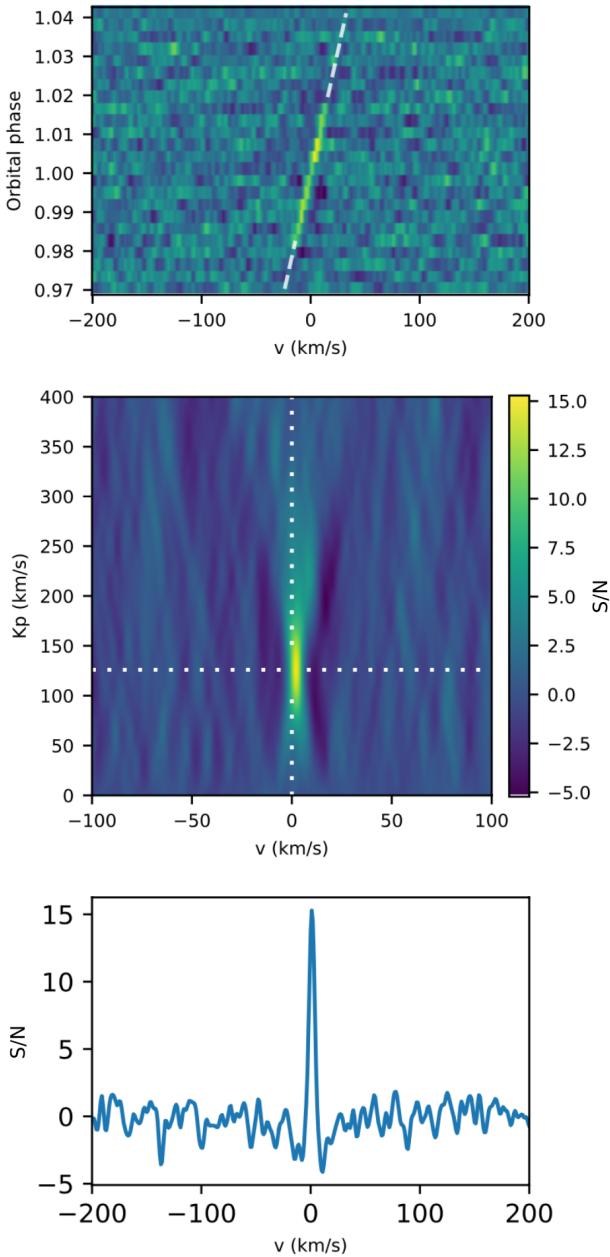


Fig. 2. Examples of data plots at three different points in the analysis. (a) *Top panel*: exposures are stacked along the vertical axis in time (where orbital phase $\phi = 1.0$ is the middle of the primary eclipse). The dotted white lines indicate before and after the transit. (b) *Middle panel*: $K_p - v_{\text{sys}}$ detection map, where all values have been shifted into the planetary rest frame. The white dotted lines indicate rest frame system velocity ($v_{\text{sys}} = 0 \text{ km s}^{-1}$) and the semi-amplitude of the radial velocity of the planet ($K_p = 126 \text{ km s}^{-1}$). (c) *Bottom panel*: example of what the cross-correlation function looks like at $K_p = 126 \text{ km s}^{-1}$, i.e. at the horizontal white line from (b).

to real observational measurements (as it is the value generally used to determine whether a detection has or has not been made), but also because it will vary with both parameters, as it should improve with both increased S/N per exposure (low time resolution) and with reduced smearing (high time resolution).

This relationship can be understood more explicitly by considering how to calculate planetary S/N . For a close-in exoplanet, assuming negligible telluric interference, the S/N for the planet

can be estimated to the first order to depend on:

$$S/N_p = \left(\frac{s_p}{s_\star} \right) S/N_\star \sqrt{N_{\text{lines}}} \quad (6)$$

where subscripts \star and p refer to the star and planet respectively; s is signal strength; and N_{lines} is the number of detected (resolvable) lines in the given wavelength range, which accounts for both how many lines are identified and their depths. By detecting multiple spectral lines, the S/N is boosted by the factor of $\sqrt{N_{\text{lines}}}$ since retaining both depth and plurality of spectral lines gives greater confidence in our identification of each unique combination of line patterns as being those of a particular molecule or species (Snellen et al. 2015; Birkby 2018). This is ultimately why S/N is both negatively affected by smearing and positively affected by boosted signal strength: smearing spectral lines reduces the S/N of the detection, as this results in fewer resolvable lines, while an increase in planetary signal strength improves the s_p/s_\star ratio.

The measure of detection strength in each detection map, denoted by \mathcal{S} , is given by the maximum cross-correlation value resulting from dividing the cross-correlation peak by the standard deviation of the detection map (calculated by excluding the central peak, i.e. everything that is further than 50 km s^{-1} to the left or right in the detection map). Physically, the \mathcal{S} -value is therefore an estimate of significance of the peak in each plot in comparison to random fluctuations. In order to confirm that this sampling method is robust against the fact that individual samples in the $K_p - v_{\text{sys}}$ should not be statistically independent, the performance of this method was compared to that of two other sampling methods (first sampling at only one row, namely the row of the signal's K_p , and then sampling at every tenth K_p row to avoid small number statistics). These three sampling methods were found to give very similar results with an average standard deviation of $\sigma = 0.078$ from each other (ranging from $0.007 \leq \sigma \leq 0.192$), indicating that the choice of exact sampling method is of negligible impact.

In order to provide more intuitive results to the reader, the measure of time resolution is given in S/N per exposure (S/N_{exp}) in this paper. As described, a high time resolution (shorter exposures) gives lower S/N per exposure, and a low time resolution (longer exposures) gives higher S/N per exposure. Considering that different systems require different total exposure times (depending on e.g. transit duration for a given exoplanet), citing a time resolution as an exposure length of n seconds does not automatically evoke an idea of whether this is high or low; instead, we cite total S/N per exposure to facilitate the comparison of relative exposure cadences across different systems. Here, a low total S/N_{exp} represents a high time resolution (short exposures) and a high total S/N_{exp} represents a low time resolution (long exposures).

4. Results

For the ‘target’ of each hypothetical case (reference study case and its variations), simulations were generated for nine different time resolutions. These ranged from the highest resolution at 186 exposures of 132 s each (plus 14 s overhead for each exposure) to the lowest resolution at 9 exposures of 2998 s each for the entire observation including baseline out-of-transit exposures. The nine time resolutions resulted in S/N_{exp} ranging between 50 and 250, increasing in increments of $S/N = 25$ for each set of longer exposures. For each combination, we computed 100 realisations of the same transit event, and the results of these

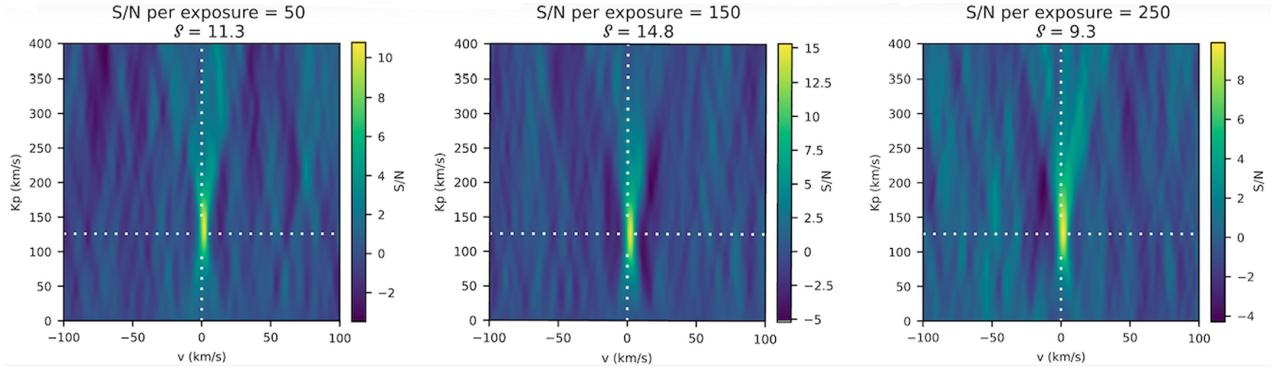


Fig. 3. Example of three $K_p - v_{\text{sys}}$ detection maps for different time resolutions ('high' resolution for $S/N_{\text{exp}} = 50$, 'medium' for $S/N_{\text{exp}} = 150$, and 'low' for $S/N_{\text{exp}} = 250$) for the WASP-127 b case at SYSREM = 5. For each detection map, its respective measure of detection strength \mathcal{S} is cited (see Sect. 3.3 for definition) with a higher \mathcal{S} -value implying a stronger cross-correlation detection.

Table 2. Differences between the four hypothetical cases, noting what their respective deviations are from the reference case (which uses parameters from Table 1).

For all cases		
Parameter	Symbol	Value
Maximum time resolution (t.r.)	$n_{\text{exp}}; t_{\text{exp}} \text{ (s)}$	186; 132
Maximum t.r. total observation	$t_{\text{obs}} \text{ (h)}$	6.82
Minimum t.r.	$n_{\text{exp}}; t_{\text{exp}} \text{ (s)}$	9; 2998
Minimum t.r. total observation	$t_{\text{obs}} \text{ (h)}$	7.50
Reference case		
Signal strength	$(R_p/R_*)^2$	1.70×10^{-4}
Transit duration	$t_t \text{ (h)}$	4.3529
Spectral resolution	R	100 000
Half signal case		
Signal strength	$(R_p/R_*)^2$	8.50×10^{-5}
Transit duration	$t_t \text{ (h)}$	4.3529
Spectral resolution	R	100 000
Half transit duration case		
Signal strength	$(R_p/R_*)^2$	1.70×10^{-4}
Transit duration	$t_t \text{ (h)}$	2.1765
Spectral resolution	R	100 000
Half spectral resolution case		
Signal strength	$(R_p/R_*)^2$	1.70×10^{-4}
Transit duration	$t_t \text{ (h)}$	4.3529
Spectral resolution	R	50 000

realisations were averaged in order to obtain a mean solution as a reference and to assess differences between individual realisations. Each set was then filtered by SYSREM and $K_p - v_{\text{sys}}$ detection maps were generated after each SYSREM iteration. An example of what different single-realisation detection maps looked like for different time resolutions can be seen in Fig. 3.

An overview of the different parameters for each case is detailed in Table 2. The four cases were constructed as follows:

(i) Reference case: This refers to the fiducial planet that is representative of the type of target that is nominally observable with our selected instrument. For the target, we use the system and transit parameters of exoplanet WASP-127 b and its host star.

(ii) Half signal: This refers to the case of halved planetary signal strength, which is achieved by reducing the line depths of the planetary model by 50%. This case effectively simulates a range of circumstances that could result in the reduction of signal strength, such as planet size, host star magnitude, distance, data lost to weather, etc.

(iii) Half transit duration: This refers to the case of the halved transit duration, which is modelled by reducing the orbital period only. The transit duration was nominally 4.35 h, and so for this case, it was reduced to 2.18 h. For this case, the same number of exposures and the same exposure lengths were used; effectively, the change here is that a smaller number of exposures contains planetary signal.

(iv) Half spectral resolution: This refers to the case of the halved spectral resolution, simulating observations taken at $R = 50\,000$ instead of $R = 100\,000$. All other instrumental effects remain the same.

4.1. Idealised case: no stellar/telluric signal, no SYSREM

For certain types of observations, removal of stellar and telluric lines is significantly less complicated and does not require algorithms like SYSREM. For example, in the optical regime, telluric lines are much more sparse as the dense bands of molecular absorption do not appear until regions of longer wavelengths. As mentioned, at these spectral ranges, telluric removal can often be done by simply masking out the comparatively small number of telluric lines as needed while the stellar signal can be divided out based on data from out-of-transit exposures. However, in the infrared regime, this is not possible due to the vast breadth of wavelength regions that would have to be masked – and so, having a method for removing stellar and telluric signal in the more targeted way that SYSREM offers is very necessary.

This means that for any observations that are simulated specifically for a spectrograph like CRIRES+, algorithms like SYSREM make up an important step in analysis that is needed to be retained for realism, but this also means that any and all interpretations of the results have to be made through the lens of this method. In order to disentangle our results from the effects of SYSREM, we first explore an idealised case (based on the same parameters of the so-called reference case) where the stellar spectrum is replaced with a simple limb-darkened black body and where no tellurics have been added, also simulated over 100 realisations. This is essentially a case where one can assume that all stellar and telluric signal have been successfully removed

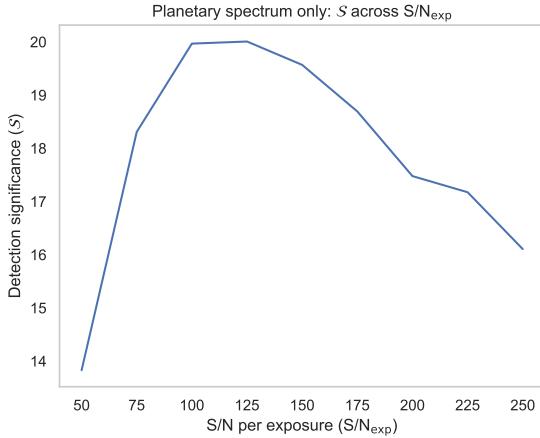


Fig. 4. Idealised reference case of no stellar or telluric signal, averaged over 100 realisations. Here, all stellar and telluric contamination are assumed to have been removed successfully (or be otherwise absent, e.g. for telluric contamination in space-based observations) through methods other than SYSREM, which is how stellar and telluric signal will be removed for the subsequent non-ideal cases.

through alternative methods (meaning this case also represents those of space-based observations). Figure 4 shows a plot for this idealised scenario, demonstrating that under these conditions, the maximum for a trade-off does indeed exist and it lies between $S/N_{\text{exp}} = 100\text{--}125$.

4.2. CRIRES+ simulations using SYSREM: averaged realisations

The value of S for each detection map was recorded, and the average S -values over 100 transit realisations for the four cases were plotted on a grid as heatmaps as seen in Fig. 5. The brighter areas indicate higher values of S for that particular combination of time resolution (S/N_{exp} value) along the vertical axis, and the number of SYSREM iterations along the horizontal axis. As the first two SYSREM iterations are not sufficient to remove telluric features, these were not included and only data from iterations 3–10 were kept for the results seen in this section.

For the reference case (top left), i.e. simulations of planet WASP-127 b over 100 realisations, there is a well-defined parametric space of maximum S -values that indicates robust detection located in the range of approximately $S/N_{\text{exp}} = 50\text{--}150$ across 4–7 SYSREM iterations. Specifically for $S/N_{\text{exp}} = 75\text{--}100$, S -values peak at $S \gtrsim 15$. Horizontally, S -values appear to peak around SYSREM $\approx 4\text{--}5$ iterations, with subsequent iterations seeing a steady decrease across all time resolutions (as per the ‘triangular’ shape of the heat map). This is an indication that at lower time resolutions (high S/N_{exp}), even a low number of SYSREM iterations will start to destroy the planetary signal.

For the case of the halved signal (top right), S -values are approximately halved ($S \gtrsim 7$) compared to the reference case as one might intuitively expect – but notably, while the numerical values of S have scaled accordingly by approximately half, the location of the parameter space maximum has not moved significantly. In this case, the approximate best range of S/N is also $S/N_{\text{exp}} = 50\text{--}150$ with a similar decay following SYSREM iterations beyond ~ 4 . In this case, the maxima has not shifted in location or breadth, but only in strength.

For the case of the halved transit duration (bottom left), the number of exposures that are taken during the transit is less than

for the previous two cases; the total observing time and the exposure length vs. exposure time configuration have not changed, but the number of baseline exposures that are taken in out-of-transit has increased since the transit is shorter in time. Here, the region of maximised S -values is shifted down towards higher time resolutions, i.e. larger number of exposures of $S/N \lesssim 100$, with the highest S -values here being approximately $S \sim 13$. This can be explained by the interpretation that as the radial velocity of the planet increases, the smearing effect per exposure also increases, and the need for a larger number of exposures becomes more pertinent. As the planet transit time is reduced, so is the total S/N collected across the transit, which is why the maximum S -values of this case is less than that of the reference case. For this case, SYSREM performance still tends to decay at higher iterations, but the ‘slope’ of this triangular shape is flatter than for the previous two cases.

For the case of the halved spectral resolution (bottom right), the overall S -values are globally lower compared to those of the reference case and the halved transit case, with highest values of $S \sim 11$. These S -values are still inevitably higher than for the case of the halved signal, but the contrast between this case and the reference case greatly emphasises the benefit of high spectral resolution for ground-based observations. While this case also shows SYSREM performance decay with more iterations at longer exposures, it also shows a much more uniform distribution of S -values across the vertical axis. This reflects the reduced impact of the smearing due to the change of planet line-of-sight velocity in comparison to the spectral resolution of the instrument. In this case, the detection significance is already low, and shorter exposures are not able to recover lost spectral resolution. In a sense, as the contrast of planetary spectral lines is lost due to lower resolution, this loss dominates the detection strength of an observation more significantly than any strength lost or gained due to changes in time resolution. As such, the loss of sensitivity to atmospheric features due to reduced spectral resolution cannot be fully recovered with improved time resolution.

The effect of smearing being more pronounced for faster targets with shorter transits yet less pronounced for observations with lower spectral resolution can also be seen by plotting kernel density estimates (KDE) for our data sets as in Fig. 6. In this plot, the range of S -values is shown for the halved transit duration case and for the halved spectral resolution case compared to the reference case.

Here, the reference case’s time resolutions of $S/N_{\text{exp}} = 75, 100$ is shown to indeed return the highest S -values, with $S/N_{\text{exp}} = 50$ faring next best, followed by $S/N_{\text{exp}} = 125, 150$. However, for the halved transit duration case, the range of S -values are much more spread; here, $S/N_{\text{exp}} = 50\text{--}100$ all fare similarly well (as confirmed by the heatmaps) but subsequent time resolutions quickly deteriorate with each successive time resolution resulting in a decreased peak of S -values due to increased smearing. Comparatively, for the halved spectral resolution, this is not true; here, $S/N_{\text{exp}} = 50\text{--}100$ still results in the highest S -values but $S/N_{\text{exp}} = 125\text{--}175$ are not far behind. For all three cases, the time resolutions of $S/N_{\text{exp}} = 200\text{--}250$ fares significantly worse as is also seen in the heatmaps of Fig. 5, in part due to the diminished success of SYSREM at higher iterations.

This spread can also be identified by calculating the variance (σ^2) of these data sets. A pronounced smearing effect should register as a larger variance across the data set (due to larger differences between performance of shorter versus longer exposures) and a diminished smearing effect should register as a

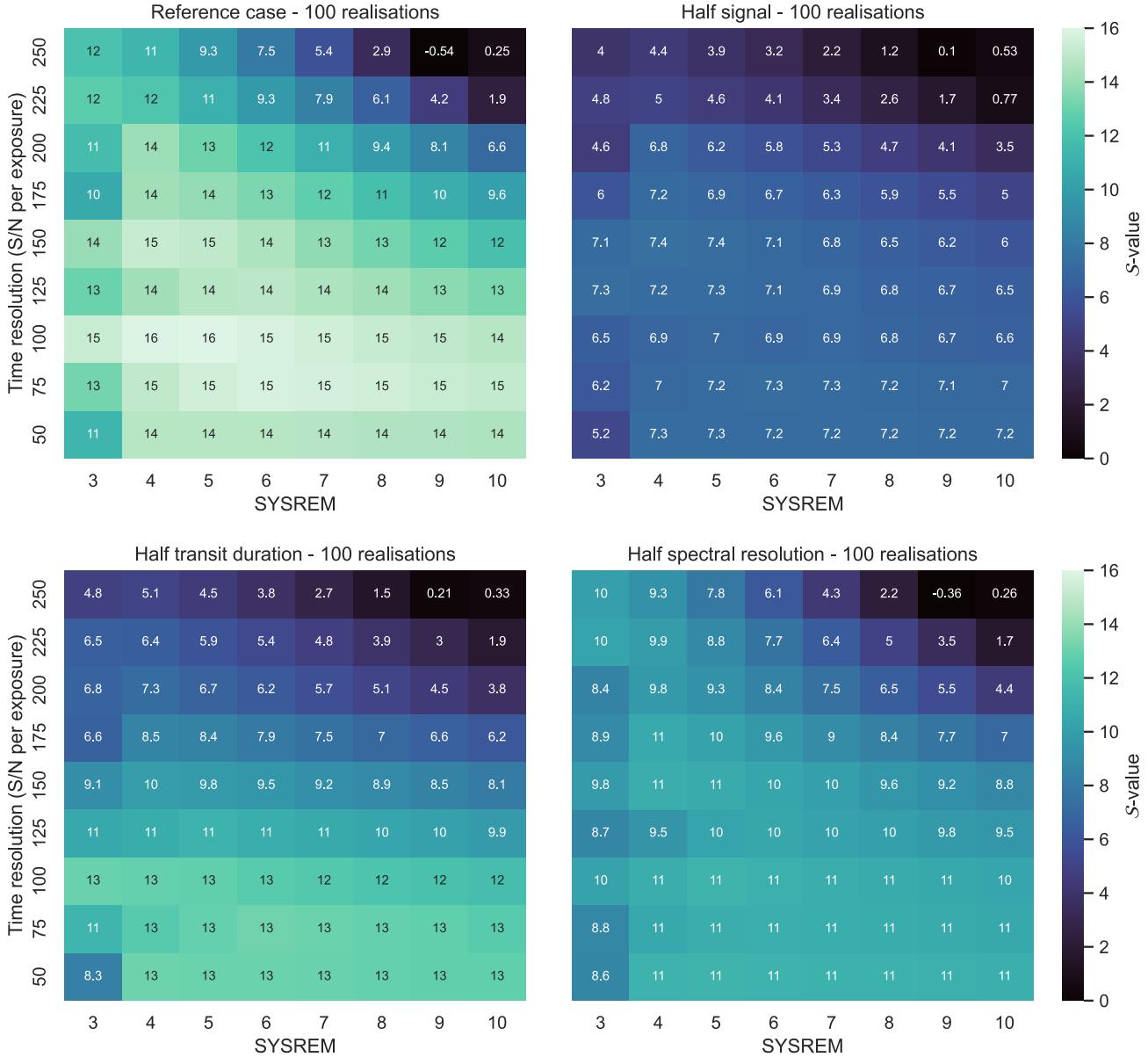


Fig. 5. How the cross-correlation detection strength, S , varies across time resolution and SYSREM iterations for four hypothetical cases averaged across 100 realisations of the same observing night. (i) *Top left*: reference case of a hypothetical observation ($R = 100\,000$) based on fiducial planet WASP-127 b. (ii) *Top right*: scenario where the reference case's planetary signal strength is halved. (iii) *Bottom left*: scenario where the reference case's transit duration is halved. (iv) *Bottom right*: scenario where the reference case is observed with half the spectral resolution ($R = 50\,000$). In these plots, a large number of short exposures corresponds to a lower S/N_{exp} (bottom of vertical axis) and a small number of long exposures corresponds to a higher S/N_{exp} (top of vertical axis).

smaller variance. Two values for variance are given in the titles of the plots of Fig. 6. The first value $\sigma_{\text{s.i.}}^2$ comes from calculating the variance across all time resolutions in a single SYSREM iteration (effectively the variance of S -values in a single column of the heatmaps of Fig. 5). Here, $\sigma_{\text{s.i.}}^2$ is calculated for $\text{SYSREM} = 4$ as this iteration consistently performs well for the three respective cases. The second value $\sigma_{\text{t.r.}}^2$ is a calculation of variance across time resolutions for all SYSREM iterations (effectively a measure of how much the KDE plots for different time resolutions shift). This value represents the amount of variation between each row of the heatmaps – and considering all $S/N_{\text{exp}} \geq 200$ fare consistently badly due to poor SYSREM performance at higher iterations, $\sigma_{\text{t.r.}}^2$ excludes $S/N_{\text{exp}} = 200–250$.

For both $\sigma_{\text{s.i.}}^2$ and $\sigma_{\text{t.r.}}^2$, the trend seen in the heatmaps holds: the smearing effect is larger (higher σ) for the halved transit case, and smaller (lower σ) for the halved resolution case.

4.3. CRIRES+ simulations using SYSREM: comparison between single realisations

In contrast to these S -values based on detection maps averaged over 100 realisations, the S -values of a single realisation can be seen in Fig. 7. Here, grids of three randomly selected realisations out of 100 (realisation number 10, 50, and 80) for each case are shown in order to illustrate how much individual realisations can deviate from the mean.

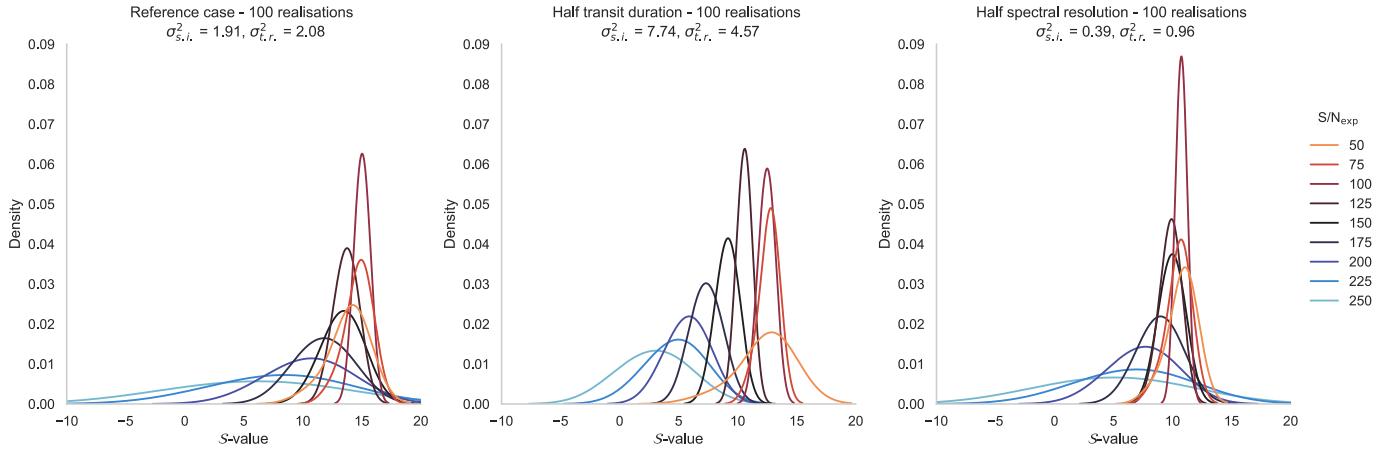


Fig. 6. Kernel density estimates of S -values for all different time resolutions $S/N_{\text{exp}} = 50\text{--}250$ across all SYSREM iterations, seen for the reference case (left) compared to the halved transit duration (middle) case and the halved spectral resolution range (right) using the data presented in Fig. 5. The variances are calculated for each case across all time resolutions at a selected iteration of $\text{SYSREM} = 4$ ($\sigma^2_{\text{s.i.}}$) and across all SYSREM iterations for time resolutions $S/N_{\text{exp}} = 50\text{--}175$ ($\sigma^2_{\text{t.r.}}$).

Importantly, we find that while the individual realisations' maxima are still confined within the range of maxima seen for the averaged realisations, the exact range of optimal time resolutions do not show up as a clear maxima in a single realisation in the same way. This contrast between the individual realisations and the averaged realisations clearly underlines how small the planetary signal is relative to noise and fluctuations associated with observations, meaning that observations from a given night can – and often will – vary significantly from the mean. This can also be seen in the presence of horizontal stripes of higher S -values in the single realisation plots: as noise was generated anew for each time resolution (vertical axis), each row has its own maxima of peak S -values in a SYSREM range. There are notable variations between each row as each have their own randomly generated noise, impacting each data set independently.

This is an important finding as we can see that two instances of the same exact circumstances – an observation of the same planetary system, with the same instrument, at the same time resolution – can seemingly in the worst case result in either a detection or a non-detection, with favourable or unfavourable random variation governing the outcome. For example, the individual realisations of the reference case (first row of plots in Fig. 7) gives a high S -value detection at time resolution $S/N_{\text{exp}} = 75$ at $\text{SYSREM} = 4\text{--}5$ (which the averaged realisation confirms is a good choice of parameters) for realisation no. 10, but a detection that is comparatively much more marginal for realisation no. 50 and no. 80. Similar variations can be seen between all cases' different individual realisations, indicating that this is a persistent, systematic issue. In order to eliminate the possibility that this result could be merely a manifestation of S being calculated from the peak value (which is itself noisy), this was also tested by using a Gaussian fitting technique, but this method still replicated these results for both the averaged realisations in Fig. 5 and the variation seen in the single realisations of Fig. 7.

same method to simulate spectra for their own specific target, transit event, or instrument to experiment with resulting cross-correlation S -values as presented here. While probing every category of such variations is certainly outside the scope of this paper, there are certain trends that we expect to hold for any high-spectral resolution transit spectroscopy observation, with certain parameters impacting the range of optimal time resolutions. These are discussed below in the context of our results.

5.1. Summary of the results

1. The SYSREM + cross-correlation approach tends to work well with relatively low S/N observations. When planning observations, one should favour higher time resolution versus higher S/N_{exp} as long as the data reaches S/N_{exp} of 50 or so. While higher S/N per exposure helps mitigate the large variation of detection significance that we see across the individual realisations, we usually do not have the luxury of combining many transits – as such, striving for higher S/N should not happen at the cost of time resolution beyond this point. Another aspect that could motivate longer (higher S/N) exposures is the overhead time that can be comparable to exposure times for large CCD detectors (see Sect. 5.6).
2. Chances of detecting planetary signal are increased by using the highest spectral resolution available. The reason for this is two-fold: firstly, higher resolution improves contrast and effectively boosts N_{lines} (from Eq. (6)). Secondly, high spectral resolution minimises the impact of telluric contamination (for ground-based observations) and opens the possibility of combining multiple transits in a single analysis if one can observe between telluric lines across multiple nights. The downside of high resolution is the difficulty of achieving high S/N and so the compromise must be found for each target and telescope/instrument combination. In our case, the width of the instrumental profile (1 km s^{-1}) is a factor of 10 smaller than the variation of the planet line-of-sight velocity during the transit, which increases the potential of reliably detecting atmospheric signal without losing too much light on the slit.
3. The use of SYSREM has a clear impact on results, with variation in optimal time resolution seen both between cases that do and do not use it (the ideal case of Sect. 4.1 versus the

5. Discussion

In this paper, we have used observations simulated for a specific instrument and targets. In order to explore how our findings would vary for other cases and conditions, observers may use the

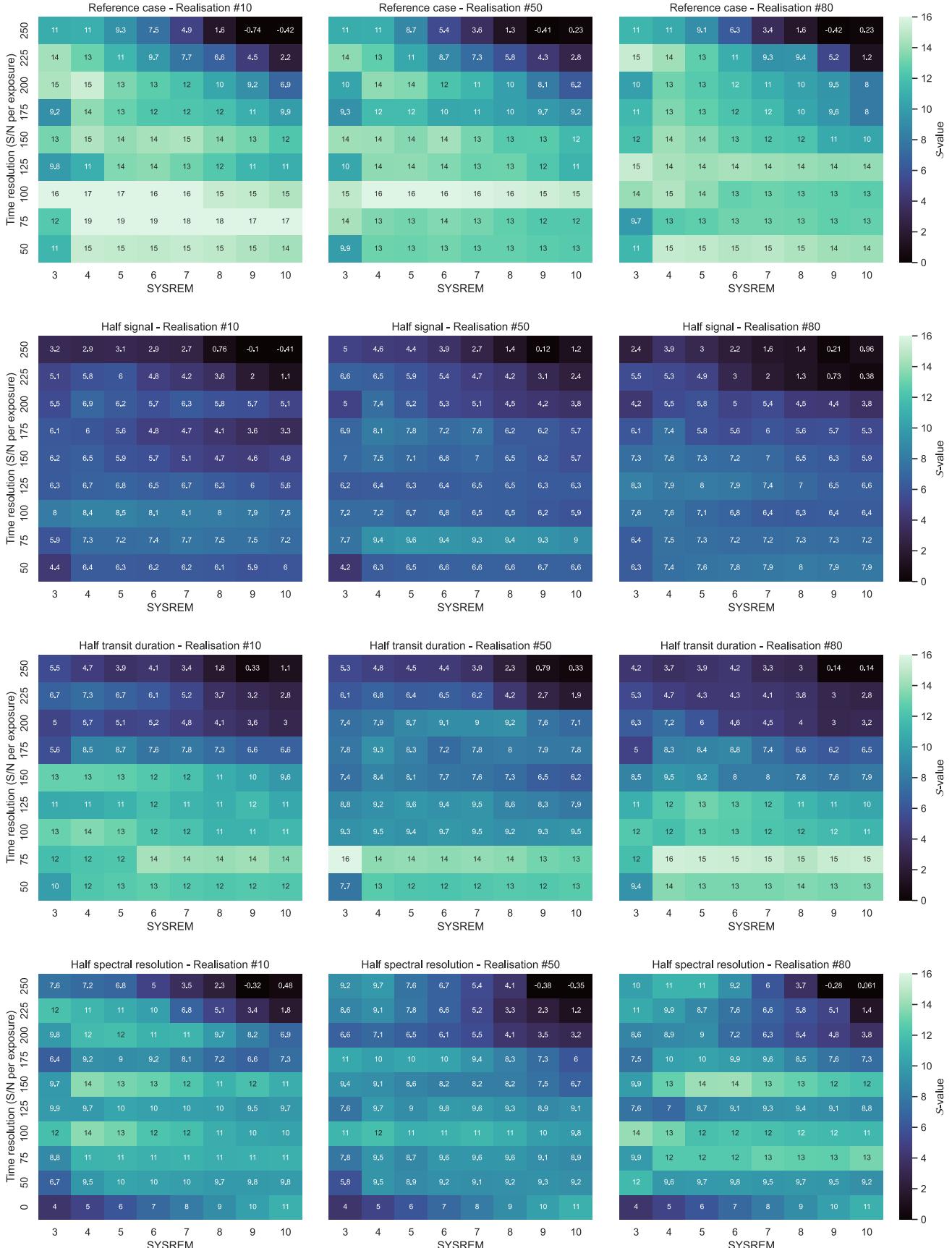


Fig. 7. How the cross-correlation detection strength, S , varies across time resolution and SYSREM iterations for individual realisations of the four hypothetical cases. Each row represents (i) the reference case, (ii) the case of the halved planetary signal, (iii) the case of the halved transit duration, and (iv) the case of the halved spectral resolution, following the same format and parameters as in Fig. 5. Note that some values are beyond the colour bar's scale, which has been locked to facilitate comparison to Fig. 5.

reference case of Sect. 4.2), as well as between different choices of iterations within the same time resolution and case (along the horizontal axes of Figs. 5 and 7). In the results of Sect. 4.2, SYSREM cleans the spectra of telluric and stellar features sufficiently already after 4–6 iterations. Generally, it can therefore be said that a large number of iterations is seemingly detrimental to the recovery of the atmospheric signal (as SYSREM may then be struggling to identify what is linear and not with such a small number of exposures) and should be avoided. In our cases, SYSREM performance is more affected by the time resolution of the data than by its S/N, leading to point 1 above.

5.2. Stacking multiple nights

If one can observe and combine data from multiple transits, this would improve the total S/N for a given target by stacking observations from different nights – in particular if done with the same instrumental setup. Such observations have been done for both transmission and emission studies in the past using a range of ground-based instruments (e.g. Giacobbe et al. 2021; Kesseli et al. 2022; Scandariato et al. 2023). It has also been shown that stacking multiple observations will be required for future characterisation of Earth-like and super-Earth exoplanets from space (e.g. Kaltenegger & Traub 2009; Rauer et al. 2011; Wunderlich et al. 2019), and this has indeed been achieved with recent JWST data (Lustig-Yaeger et al. 2023). However, stacking of ground-based observations in the mid-infrared will require more advanced telluric removal algorithms than SYSREM considering that the barycentric velocity change between nights and weather variations will result in large changes of telluric features between transits.

As shown in Sect. 4, there is a notable difference in the distribution of \mathcal{S} -values for a single realisation compared to the distribution of the multi-realisation \mathcal{S} -values; however, these effectively show an average of multiple observations, which is not the same as stacking observations to amplify the S/N. Differences in observing conditions (seeing, telluric line strength) may set a limit to the number of transits that can be properly combined with a gain to the planetary signal detection.

If the quality of observations of individual transits is similar, the stacked data set will be effectively similar to ‘oversampling’ planet transit with more spectra of the same time and spectral resolution. In any case, stacking multiple transits should result in higher values of \mathcal{S} provided that SYSREM is modified to be used with such data set.

5.3. Variable exposure lengths

In this work, we have explored different time resolutions with simulations ranging from several shorter exposures to fewer longer exposures. However, it may be possible that a variation exists in what should be prioritised during different sections of the transit event; is the trade-off balance the same during ingress/egress as it is during the middle of the transit? If not, further optimisation could be possible if exposure lengths were varied over time, with either shorter exposures during ingress/egress and longer exposures during mid-transit, or vice versa.

Considering that the background spectrum of the star changes fastest in time close to the limb, this might suggest that shorter exposures during ingress and egress would be beneficial. On the other hand, the specific intensity is much lower across the limbs than in disk centre, which may suggest the opposite. This

exact balance – and how it may be affected by further parameters e.g. planetary orbit eccentricity (Van Eylen et al. 2019) – could be explored in future work, but because the two effects are working against each other, one could hope that equal exposure times are not far from the optimal strategy. Furthermore, it is reasonable to assume that the effect will nonetheless be significantly smaller compared to other, more dominant factors such as the number of SYSREM iterations. This, alongside the balance of the two opposite influences, discouraged further exploration of variable exposure lengths as we do not expect it to have a significant effect on the detectability of planetary signal for our particular cases.

5.4. Exoplanet atmospheric features

Several exoplanet atmospheric features have been found to affect the transmission spectra of exoplanets. Thanks to the progress in the field, we have been increasingly obliged to acknowledge that exoplanetary atmospheres are complicated and three-dimensional systems that should not be overly simplified; more chemically complicated species like aerosols are now known to be commonly found in many types of exoplanetary atmospheres, creating clouds and hazes that make characterisation difficult at the lower atmosphere layers (e.g. Helling 2019; Gao et al. 2021), and the formation of other known meteorological phenomena such as precipitation and winds are also being studied in great detail (e.g. Heng & Showman 2015; Seidel et al. 2020a; Fortney et al. 2021; Loftus & Wordsworth 2021). Recent years have also seen an rise in adopting 3D general circulation models, or global climate models (GCM), for use with exoplanets in order to demonstrate the importance of atmospheric dynamics and stratification for interpretation of exoplanets’ transmission spectra (e.g. Ding & Wordsworth 2019; Wolf et al. 2019; Way & Del Genio 2020; Fauchez et al. 2022).

The cross-correlation detection map is indeed sensitive to many such atmospheric effects, with evidence of high-altitude winds manifesting as broadening or off-sets. The simulations in this work did not take these effects into account, including e.g. rotational broadening – an effect that would be present most notably on tidally-locked planets with shorter orbital periods – which would presumably lead to an increase in a type of smearing that arises separately from choice of time resolution. As such, in the case of strong atmospheric features, there may be a supplementary sensitivity in what time resolutions result in the optimal cross-correlation detection but the exact nature of this relationship would need to be explored before further conclusions can be reached. This relationship may be investigated by testing the dependence of inferred input atmospheric parameters on simulated observing choices as per e.g. Savel et al. (in prep.).

5.5. Host star activity

As mentioned in Sect. 1, cooler stars such as M-dwarfs are likely hosts of exoplanets and thus commonly observed. M-dwarfs are known to be active stars with magnetic fields that may affect close-in orbiting exoplanets (e.g. Morin et al. 2010; Airapetian et al. 2017; Kochukhov 2021), and this high level of activity means that they are likely to have starspots appear on their photospheres. While such activity of M-dwarf host stars can in cases be exploited to aid exoplanet atmosphere studies (e.g. Diamond-Lowe et al. 2023), there has been concern regarding what the influence may be of starspots present on the surface of host stars during transits – and unfortunately, the fear of such spots creating false spectral features appears to be well-founded especially

at lower spectral resolutions (Rackham et al. 2018; Apai et al. 2018; Barclay et al. 2021; Moran et al. 2023; Libby-Roberts et al. 2023). Considering the simulated transmission spectra in this work assumes a uniform stellar disk with no features, the risk of contamination due to potential activity of host stars should be taken into account when considering the validity of such an assumption. The situation could be different for stars of later spectral type, where the impact of the combination of larger level of activity and smaller convective cells must be further explored.

5.6. Differences between instruments

By now, there are tens of astronomy instruments available globally of world-class quality that are regularly used to take spectroscopic observations of exoplanets. As discussed in Sect. 1, the different traits of the instruments across this wide range all provide advantages and disadvantages regarding high or low spectral resolution, ground- or space-based, mirror size, wavelength coverage, and many more. All of these will combine uniquely to create an exclusive case for each given instrument and target, and so it is possible that the range of optimal time resolutions varies distinctly when using different instruments/telescopes. Furthermore, differences between instruments could be taken advantage of in the case that observers explore the possibility of stacking multiple observations from multiple instruments, as done by e.g. Ridden-Harper et al. (2023).

One important aspect to consider when selecting between different instruments is how large the observing overheads are. CRIRES+ has the benefit of being able to observe with relatively small overheads, requiring only a few seconds between sequential exposures. Other instruments have notably longer overheads of e.g. 100 s for each pair of frames for the Gemini-N/MAROON-X instrument, and either 45 s (fast readout mode) or 68 s (slow readout mode) for the VLT/ESPRESSO instrument.

Accounting for overhead time per exposure (t_{OH}) is important because large overheads will eat into valuable light collection during the transit, with a larger number of exposures (n_{exp}) consuming more observing time. The amount of observing time lost to overheads for a given instrument is given by $t_{\text{OH}}(n_{\text{exp}} - 1)$, which means that there will be a larger difference between how much total light collection is lost for a high n_{exp} and low n_{exp} if the instrument has a high t_{OH} . As such, for instruments with notably large overheads, faint targets requiring a low number of n_{exp} need to be reasonably slow, and fast targets requiring a high number of n_{exp} need to be reasonably bright. For targets that have both short transit times and low signal strengths, observers will benefit from using instruments with relatively small overheads instead.

5.7. Alternatives to PCA-based approaches

The major drawback of ground-based observations in the infrared is telluric contamination. Its strength is, ironically, the reason why we observe at these wavelengths as the species present in our atmosphere are often those we search for in our targets. Considering that various techniques of removing telluric features such as SYSREM and other functionally-similar tools have several limitations, their iterative nature makes it difficult to control their impact on the planetary atmosphere features. As such, it is very difficult to assess this impact even in the analysis of the cross-correlation results.

An alternative way to approach this problem of isolating the planetary contribution is to instead model the observations as

a whole, treating the components (stellar, telluric, and planetary spectra) as unknown functions. This will lead to an inverse problem that can be solved for a global minimum assuming that the orbital elements and the corresponding Doppler shifts of the components are well known for all phases of the transit. While both algorithmically and computationally more complex, an inverse problem approach has the benefit of producing a single global planet transmission spectrum, rather than a set of residuals for each phase as PCA-based approaches do. Such a technique is currently under development by Piskunov et al. (in prep.), building on preliminary theoretical work by Aronson & Waldén (2015).

5.8. Implications for previous studies

With the finding that the detection strength of individual realisations vary significantly from the mean (and from each other), there is a need to assess what this means for current methodologies as the choice of time resolution or number of SYSREM iteration could be the difference between a detection and non-detection.

Many previous studies have acknowledged the lack of consistent treatment and the challenges associated with using iterative detrending methods such as SYSREM, including the potential for false positives (e.g. Hawker et al. 2018; Cabot et al. 2019; Cheverall et al. 2023). Our findings further underline this need for caution, and for all results (new or revisited), care should be taken to avoid the risk of both false positives and false negatives due to noise variation per observing night. One potential mitigation would be to carry out PCA-analysis injection tests (with their proven success record) over multiple realisations to obtain an averaged result similar to our work as this might improve iteration selection – especially for PCA-based approaches that do not use the same noise model for the injection and for the data – but further work is needed to develop and test the performance of such a method.

5.9. An ‘exposure triangle’ rule for planning observations

Our results indicate that the concept of a trade-off in the context of transit spectroscopy observations is a multi-dimensional problem. By varying different parameters, the maxima of the ‘optimal’ time resolution’s range can in response vary in location, strength, and size – and only clearly so over a large number of realisations. With so many variables that can influence the result, it is not surprising that changes do not manifest in a linear fashion across the breadth of parameters tested.

It is therefore our finding that while observers will fare best by creating their own simulations and investigating specific scenarios as needed, as a rule of thumb, a strategy akin to the so-called ‘exposure triangle’ may be recommended. In traditional photography, the rule of the exposure triangle is approximately that three elements must be considered prior to taking a photograph: aperture (how widely the camera lens’ diaphragm opens), shutter speed (how long the camera shutter opens for), and film speed (how much light can be registered on the camera’s film or digital sensor). These three controls balance to determine how much light enters (through aperture and shutter speed) based on how sensitive the instrument is (film speed). Depending on several conditions – e.g. the subject’s motion, or the lighting of the subject’s location – photographers may be restrained in one or more of these dimensions, and considering the exposure triangle allows them to determine how to compensate for those restraints through adjusting one or both of the others. For example, if one

is taking a photograph of a fast-moving car, one needs to reduce the shutter speed to capture it without blurring, but can compensate for that loss of light by increasing the aperture or increasing film speed.

By considering an analogous scenario for capturing transiting exoplanets, a similar rule may be employed: if one is restrained by a particular observational parameter – a short transit, a faint target, or an instrument of low spectral resolution – one may consider the other variables to determine approximately which parameters to prioritise or adjust. As such, there is no single range of time resolutions that will consistently deliver optimal results; like photography, the way to centre your exposure within this set of parameters will depend on several factors as discussed in this section and above. As per our results, certain adjustments are fairly intuitive – for example, that smearing effects are more pronounced for shorter periods but less pronounced at lower spectral resolutions – but others require more consideration or sacrifice.

While exact priorities must therefore be considered for each given case, one can still produce a general recommendation for a single target: smearing is more of a concern when the transit duration is short (because the effect is more pronounced) and less of a concern when resolving power is low (because spectral resolution cannot be regained by shorter exposures). Therefore, observers should first consider the parameters of their target system and chosen spectrograph/instrument. Smearing will be of minimal concern when their target has a relatively long transit duration, and when their instrument has a relatively low spectral resolution; in this case, observers should prioritise a longer exposure for maximum S/N. Inversely, smearing will be of maximal concern when their target has a relatively short transit duration, and their instrument has a relatively high spectral resolution; in this case, observers should prioritise a shorter exposure for minimal smearing.

In the case where neither of these factors is of notable effect, i.e. either because both factors are effectively balanced (long duration with high R , or short duration with low R) or because both are of roughly equal detriment (neither factor stands out as being of particularly great impact), one should instead consider the signal strength. Per our simulations, as long as the S/N_{exp} can reach $\sim 50-75$, longer exposures do not appear to improve detections sufficiently to be worthwhile, especially considering that SYSREM clearly performs more poorly with a lower number of spectra (longer exposures). This means that for systems that can guarantee higher signal strength, where observers can trust that they can receive a relatively high S/N even in a shorter exposure due to e.g. high contrast, they may prioritise a shorter exposure in order to retain as much spectral resolution as possible. However, if the signal strength is expected to be weak, and observers know they can only obtain sufficiently high signal strengths by longer exposures, they are recommended to do so in spite of increased smearing.

This generalised recommendation is outlined schematically in Fig. 8, whose first plot shows a simple flow chart that may be consulted for broader decision-making, with the second plot illustrating the logic of the flow chart structure. What exact values will qualify as being a short/long transit duration, or of high/low spectral resolution, or a high/low signal strength cannot be established explicitly here for every combination of possible targets and instruments; as such, decisions based on this recommendation require a judgement call that must be made by astronomers based on their science goals (e.g. detection or characterisation) and on their global understanding of their upcoming observation. It is this evaluation, where one judges what may or

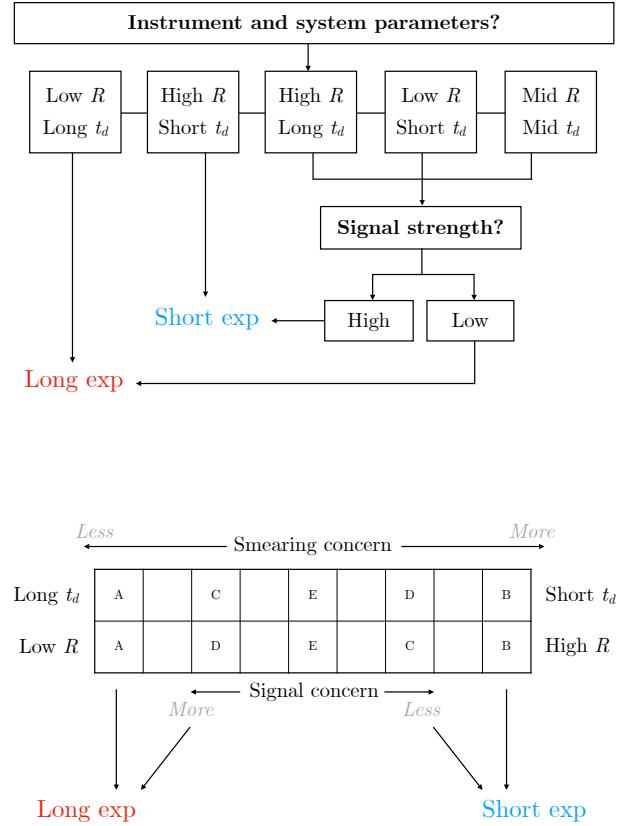


Fig. 8. Generalised recommendation for a single target. *Top panel:* a flow chart that observers may consult while planning their observations. First, observers should consider their instrument and system parameters. For lower spectral resolutions (R) and longer transit durations (t_d), longer exposures may be prioritised; for the inverse case, shorter exposures are more suitable. For other cases, the choice is determined by what signal strength is possible. *Bottom panel:* the logic of the top schematic is illustrated here. Smearing is more of a concern for low R and long t_d (case A) and less of a concern for the inverse (case B). For all other cases (C, D, E), a larger concern for signal (i.e. lower signal strength) encourages longer exposures and a lesser concern encourages shorter exposures.

may not be of greatest priority for a particular case, that can be considered as the use of an ‘exposure triangle’ approach in this context. Again, if resources permit, more specific boundaries on time resolution may be calculated if observers choose to create their own simulations following the methodology outlined in this work, but the recommendation above can still be employed as a short-hand guide.

6. Conclusions

In this work, we simulated multiple CRIRES+ observations of four hypothetical targets based on fiducial planet WASP-127 b and used the PCA-like SYSREM algorithm – a standard analysis tool in exoplanetary atmosphere studies – to investigate the trade-off between S/N and time cadence when conducting transmission spectroscopy observations.

(1) We demonstrate that time resolution significantly impacts the strength of cross-correlation detection for transiting exoplanets. Across our range of data sets, there is a clear trade-off between observing a transit with a small number of higher-S/N exposures and a large number of lower-S/N exposures, with the

range of optimal time resolution and overall performance varying based on the characteristics of the system being observed. The exact location, strength, and size of this range varies based on multiple variables, which means that observers need to consider the entire context of the observation when evaluating the best strategy (akin to the ‘exposure triangle’ rule of traditional photography), and supplementary techniques such as stacking may return different maxima ranges. We caution those planning observations to robustly simulate target systems ahead of time in order to take full advantage of the time limitations imposed both by transit duration and telescope access.

(2) Based on our simulations, variation in target signal strength does not appear to shift the location of these maxima in the parameter space, but the benefits of prioritising a larger number of exposures should be taken into account for targets with shorter transit times. At lower spectral resolution, the smearing effect associated with faster transiting targets appears to be less pronounced, although a higher spectral resolution clearly results in more significant detection at the optimal time resolution. For the four hypothetical cases explored in this paper, as a general rule, time resolutions of $S/N_{\text{exp}} \approx 50\text{--}100$ appear to provide relatively high detection strength for all cases, with no case achieving an optimal resolution of $S/N_{\text{exp}} > 200$.

(3) There also appears to be a minimum threshold of time resolution below which SYSREM may destroy the planetary signal after only a small number of iterations; however, this threshold needs to be determined more precisely for specific targets and instrument configurations. For our cases, time resolutions of $S/N_{\text{exp}} \gtrsim 200$ appear to be especially sensitive, and SYSREM $\approx 4\text{--}6$ generally seems to yield the best results.

(4) We also find that night-specific variations in seeing and data noise, for example, have a significant impact on the results of that night. As this work has explored ideal cases, this impact is reasonably assumed to be even greater for real-life circumstances such as variations in weather conditions. Considering that many studies are based on single-night observations, this needs to be accounted for during analysis (including the steps pertaining to stellar and/or telluric removal) and in evaluating the validity of single-night detections.

As the field of exoplanetary atmosphere characterisation makes solid progress in every direction – improved instrumentation quality, analytical methods, data reduction, and modelling methods for generating template spectra for cross-correlation – we can reasonably expect good progress in the relatively near future concerning characterisation of lower-mass exoplanets to match our achievements with higher-mass exoplanets. With such developments making rapid strides, it is important to monitor our tools over time to ensure that they do not become obsolete; and so, by monitoring the variation of optimal time resolution for different observations, we can verify that we are continuously obtaining the best possible data for our scientific progress.

Acknowledgements. L.B.-Ch., A.D.R., and N.P. acknowledge support by the Knut and Alice Wallenberg Foundation (grant 2018.0192). O.K. acknowledges support by the Swedish Research Council (grant agreement no. 2019-03548), the Swedish National Space Agency, and the Royal Swedish Academy of Sciences. F.L. and L.N. acknowledge the support by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project number 314665159. This research has made use of the NASA Exoplanet Archive, which is operated by the California Institute of Technology, under contract with the National Aeronautics and Space Administration under the Exoplanet Exploration Program. This work has also made use of the following Python packages: Astropy (Astropy Collaboration 2013), iPython (Perez & Granger 2007), Matplotlib (Hunter 2007), NumPy (Harris et al. 2020), Pandas (McKinney 2010; The pandas development team

2023), PyAstronomy (Czesla et al. 2019), SciPy (Virtanen et al. 2020), and Seaborn (Waskom et al. 2017). We also thank the anonymous referee for their comments as their suggestions have contributed significantly to this work.

References

- Addison, B., Wright, D. J., Wittenmyer, R. A., et al. 2019, *PASP*, **131**, 115003
 Airapetian, V. S., Glocer, A., Khazanov, G. V., et al. 2017, *ApJ*, **836**, L3
 Allart, R., Lovis, C., Pino, L., et al. 2017, *A&A*, **606**, A144
 Apai, D., Rackham, B. V., Giampapa, M. S., et al. 2018, arXiv e-prints [arXiv:1803.08708]
 Aronson, E., & Waldén, P. 2015, *A&A*, **578**, A133
 Artigau, E., Kouach, D., Donati, J.-F., et al. 2014, *Proc. SPIE*, **9147**, 914715
 Astropy Collaboration (Robitaille, T. P., et al.) 2013, *A&A*, **558**, A33
 Astudillo-Defru, N., & Rojo, P. 2013, *A&A*, **557**, A56
 Bakos, G., Noyes, R. W., Kovács, G., et al. 2004, *PASP*, **116**, 266
 Bakos, G. A., Csabry, Z., Penev, K., et al. 2013, *PASP*, **125**, 154
 Barclay, T., Kostov, V. B., Colón, K. D., et al. 2021, *AJ*, **162**, 300
 Bean, J. L., Miller-Ricci Kempton, E., & Homeier, D. 2010, *Nature*, **468**, 669
 Benneke, B., Wong, I., Piaulet, C., et al. 2019, *ApJ*, **887**, L14
 Birkby, J. L. 2018, arXiv e-prints [arXiv:1806.04617]
 Birkby, J. L., de Kok, R. J., Brogi, M., et al. 2013, *MNRAS*, **436**, L35
 Birkby, J. L., Kok, R. J. d., Brogi, M., Schwarz, H., & Snellen, I. A. G. 2017, *AJ*, **153**, 138
 Birkmann, S. M., Ferruit, P., Giardino, G., et al. 2022, *A&A*, **661**, A83
 Borucki, W. J., Koch, D., Basri, G., et al. 2010, *Science*, **327**, 977
 Boucher, A., Lafrenière, D., Pelletier, S., et al. 2023, *MNRAS*, **522**, 5062
 Bouchy, F., Udry, S., Mayor, M., et al. 2005, *A&A*, **444**, L15
 Brogi, M., & Line, M. R. 2019, *AJ*, **157**, 114
 Brogi, M., de Kok, R. J., Albrecht, S., et al. 2016, *ApJ*, **817**, 106
 Brogi, M., Line, M., Bean, J., Désert, J.-M., & Schwarz, H. 2017, *ApJ*, **839**, L2
 Brogi, M., Giacobbe, P., Guilluy, G., et al. 2018, *A&A*, **615**, A16
 Cabot, S. H. C., Madhusudhan, N., Hawker, G. A., & Gandhi, S. 2019, *MNRAS*, **482**, 4422
 Charbonneau, D., Brown, T. M., Noyes, R. W., & Gilliland, R. L. 2002, *ApJ*, **568**, 377
 Cheverall, C. J., Madhusudhan, N., & Holmberg, M. 2023, *MNRAS*, **522**, 661
 Cutri, R. M., Skrutskie, M. F., van Dyk, S., et al. 2003, VizieR Online Data Catalog, II/246
 Czesla, S., Schröter, S., Schneider, C. P., et al. 2019, Astrophysics Source Code Library [record ascl:1906.010]
 Diamond-Lowe, H., Berta-Thompson, Z., Charbonneau, D., & Kempton, E. M. R. 2018, *AJ*, **156**, 42
 Diamond-Lowe, H., Mendonça, J. M., Charbonneau, D., & Buchhave, L. A. 2023, *AJ*, **165**, 169
 Di Gloria, E., Snellen, I. A. G., & Albrecht, S. 2015, *A&A*, **580**, A84
 Ding, F., & Wordsworth, R. D. 2019, *ApJ*, **878**, 117
 Dorn, R. J., Bristow, P., Smoker, J. V., et al. 2023, *A&A*, **671**, A24
 Ehrenreich, D., Lovis, C., Allart, R., et al. 2020, *Nature*, **580**, 597
 Fauchez, T. J., Villanueva, G. L., Sergeev, D. E., et al. 2022, *Planet. Sci. J.*, **3**, 213
 Fortney, J. J., Dawson, R. I., & Komacek, T. D. 2021, *J. Geophys. Res. Planets*, **126**, e2020JE006629
 Gaia Collaboration (Brown, A. G. A., et al.) 2018, *A&A*, **616**, A1
 Gandhi, S., Kesseli, A., Zhang, Y., et al. 2023, *AJ*, **165**, 242
 Gao, P., Wakeford, H. R., Moran, S. E., & Parmentier, V. 2021, *J. Geophys. Res. (Planets)*, **126**, e06655
 Giacobbe, P., Brogi, M., Gandhi, S., et al. 2021, *Nature*, **592**, 205
 Gibson, N. P., Merritt, S., Nugroho, S. K., et al. 2020, *MNRAS*, **493**, 2215
 Hardegree-Ullman, K. K., Cushing, M. C., Muirhead, P. S., & Christiansen, J. L. 2019, *AJ*, **158**, 75
 Harris, C. R., Millman, K. J., van der Walt, S. J., et al. 2020, *Nature*, **585**, 357
 Hawker, G. A., Madhusudhan, N., Cabot, S. H. C., & Gandhi, S. 2018, *ApJ*, **863**, L11
 Helling, C. 2019, *Annu. Rev. Earth Planet. Sci.*, **47**, 583
 Heng, K., & Showman, A. P. 2015, *Annu. Rev. Earth Planet. Sci.*, **43**, 509
 Hoeijmakers, H. J., Ehrenreich, D., Heng, K., et al. 2018, *Nature*, **560**, 453
 Howell, S. B., Sobeck, C., Haas, M., et al. 2014, *PASP*, **126**, 398
 Hunter, J. D. 2007, *Comput. Sci. Eng.*, **9**, 90
 Husser, T.-O., Berg, S. W.-v., Dreizler, S., et al. 2013, *A&A*, **553**, A6
 Jakobsen, P., Ferruit, P., Oliveira, C. A. d., et al. 2022, *A&A*, **661**, A80
 Jehin, E., Gillon, M., Queloz, D., et al. 2011, *The Messenger*, **145**, 2
 JWST Transiting Exoplanet Community Early Release Science Team (Ahrer, E.-M., et al.) 2023, *Nature*, **614**, 649
 Kaltenegger, L., & Traub, W. A. 2009, *ApJ*, **698**, 519
 Kanodia, S., Wolfgang, A., Stefansson, G. K., Ning, B., & Mahadevan, S. 2019, *ApJ*, **882**, 38

- Kausch, W., Noll, S., Smette, A., et al. 2015, [A&A](#), **576**, A78
- Kesseli, A. Y., Snellen, I. A. G., Casasayas-Barris, N., Mollière, P., & Sánchez-López, A. 2022, [AJ](#), **163**, 107
- Kochukhov, O. 2021, [A&ARv](#), **29**, 1
- Kok, R. J. d., Birkby, J., Brogi, M., et al. 2014, [A&A](#), **561**, A150
- Kornilov, V., Tokovinin, A., Shatsky, N., et al. 2007, [MNRAS](#), **382**, 1268
- Kreidberg, L., Bean, J. L., Désert, J.-M., et al. 2014, [Nature](#), **505**, 69
- Kreidberg, L., Line, M. R., Thorngren, D., Morley, C. V., & Stevenson, K. B. 2018, [ApJ](#), **858**, L6
- Kunimoto, M., & Matthews, J. M. 2020, [AJ](#), **159**, 248
- Lam, K. W. F., Faedi, F., Brown, D. J. A., et al. 2017, [A&A](#), **599**, A3
- Lesjak, F., Nortmann, L., Yan, F., et al. 2023, [A&A](#), **678**, A23
- Libby-Roberts, J. E., Schutte, M., Hebb, L., et al. 2023, [AJ](#), **165**, 249
- Loftus, K., & Wordsworth, R. 2021, [J. Geophys. Res. Planets](#), **126**
- Lustig-Yaeger, J., Fu, G., May, E. M., et al. 2023, [Nat. Astron.](#), **7**, 1317
- MacDonald, R. J., & Madhusudhan, N. 2017, [MNRAS](#), **469**, 1979
- Maguire, C., Gibson, N. P., Nugroho, S. K., et al. 2023, [MNRAS](#), **519**, 1030
- Maimone, M. C., Brogi, M., Chiavassa, A., et al. 2022, [A&A](#), **667**, A106
- Mayor, M., Pepe, F., Queloz, D., et al. 2003, [The Messenger](#), **114**, 20
- McCloat, S., von Essen, C., & Fieber-Beyer, S. 2021, [AJ](#), **162**, 132
- McKinney, W. 2010, in [Proceedings of the 9th Python in Science Conference](#), 56
- Merritt, S. R., Gibson, N. P., Nugroho, S. K., et al. 2021, [MNRAS](#), **506**, 3853
- Mollière, P., Wardenier, J. P., Boekel, R. v., et al. 2019, [A&A](#), **627**, A67
- Moran, S. E., Stevenson, K. B., Sing, D. K., et al. 2023, [ApJ](#), **948**, L11
- Morin, J., Donati, J. F., Petit, P., et al. 2010, [MNRAS](#), **407**, 2269
- Mounzer, D., Lovis, C., Seidel, J. V., et al. 2022, [A&A](#), **668**, A1
- Nikolov, N., Maciejewski, G., Constantino, S., et al. 2021, [AJ](#), **162**, 88
- Park, C., Jaffe, D. T., Yuk, I.-S., et al. 2014, [Proc. SPIE](#), **9147**, 91471D
- Pepe, F., Cristiani, S., Rebolo, R., et al. 2021, [A&A](#), **645**, A96
- Pepper, J., Pogge, R. W., DePoy, D. L., et al. 2007, [PASP](#), **119**, 923
- Perez, F., & Granger, B. E. 2007, [Comput. Sci. Eng.](#), **9**, 21
- Pino, L., Brogi, M., Désert, J. M., et al. 2022, [A&A](#), **668**, A176
- Pollacco, D. L., Skillen, I., Collier Cameron, A., et al. 2006, [PASP](#), **118**, 1407
- Pont, F., Knutson, H., Gilliland, R. L., Moutou, C., & Charbonneau, D. 2008, [MNRAS](#), **385**, 109
- Prineth, B., Hoeijmakers, H. J., Kitzmann, D., et al. 2022, [Nat. Astron.](#), **6**, 449
- Prineth, B., Hoeijmakers, H. J., Pelletier, S., et al. 2023, [A&A](#), **678**, A182
- Rackham, B. V., Apai, D., & Giampapa, M. S. 2018, [ApJ](#), **853**, 122
- Rauer, H., Gebauer, S., Paris, P. V., et al. 2011, [A&A](#), **529**, A8
- Redfield, S., Endl, M., Cochran, W. D., & Koesterke, L. 2008, [ApJ](#), **673**, L87
- Ricker, G. R., Winn, J. N., Vanderspek, R., et al. 2015, [J. Astron. Telescopes Instrum. Syst.](#), **1**, 014003
- Ridden-Harper, A., de Mooij, E., Jayawardhana, R., et al. 2023, [AJ](#), **165**, 211
- Rothman, L. S., Gordon, I. E., Barber, R. J., et al. 2010, [JQSRT](#), **111**, 2139
- Rustamkulov, Z., Sing, D., Mukherjee, S., et al. 2023, [Am. Astron. Soc. Meeting Abstracts](#), **55**, 124.01
- Sabotta, S., Schlecker, M., Chaturvedi, P., et al. 2021, [A&A](#), **653**, A114
- Sánchez-López, A., Alonso-Floriano, F. J., López-Puertas, M., et al. 2019, [A&A](#), **630**, A53
- Scandariato, G., Borsa, F., Bonomo, A. S., et al. 2023, [A&A](#), **674**, A58
- Seager, S., & Sasselov, D. D. 2000, [ApJ](#), **537**, 916
- Seidel, J. V., Ehrenreich, D., Pino, L., et al. 2020a, [A&A](#), **633**, A86
- Seidel, J. V., Lendl, M., Bourrier, V., et al. 2020b, [A&A](#), **643**, A45
- Seifahrt, A., Bean, J. L., Stürmer, J., et al. 2020, [Proc SPIE](#), **11447**, 114471F
- Sing, D. K., Vidal-Madjar, A., Désert, J. M., Lecavelier des Etangs, A., & Ballester, G. 2008, [ApJ](#), **686**, 658
- Sing, D. K., Pont, F., Aigrain, S., et al. 2011, [MNRAS](#), **416**, 1443
- Smette, A., Sana, H., Noll, S., et al. 2015, [A&A](#), **576**, A77
- Snellen, I. A. G., Albrecht, S., de Mooij, E. J. W., & Le Poole, R. S. 2008, [A&A](#), **487**, 357
- Snellen, I. A. G., de Kok, R. J., de Mooij, E. J. W., & Albrecht, S. 2010, [Nature](#), **465**, 1049
- Snellen, I., Kok, R. d., Birkby, J. L., et al. 2015, [A&A](#), **576**, A59
- Spyratos, P., Nikolov, N. K., Constantinou, S., et al. 2023, [MNRAS](#), **521**, 2163
- Stassun, K. G., Oelkers, R. J., Paegert, M., et al. 2019, [AJ](#), **158**, 138
- Stevenson, K. B. 2016, [ApJ](#), **817**, L16
- Swain, M. R., Vasish, G., & Tinetti, G. 2008, [Nature](#), **452**, 329
- Talens, G. J. J., Albrecht, S., Spronck, J. F. P., et al. 2017, [A&A](#), **606**, A73
- Tamuz, O., Mazeh, T., & Zucker, S. 2005, [MNRAS](#), **356**, 1466
- The pandas development team 2023, <https://doi.org/10.5281/zenodo.8092754>
- Van Eylen, V., Albrecht, S., Huang, X., et al. 2019, [AJ](#), **157**, 61
- Vidal-Madjar, A., Désert, J. M., Lecavelier des Etangs, A., et al. 2004, [ApJ](#), **604**, L69
- Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, [Nat. Meth.](#), **17**, 261
- Wakeford, H. R., Lewis, N. K., Fowler, J., et al. 2019, [AJ](#), **157**, 11
- Waskom, M., Botvinnik, O., O'Kane, D., et al. 2017, <https://doi.org/10.5281/zenodo.883859>
- Way, M. J., & Del Genio, A. D. 2020, [J. Geophys. Res. Planets](#), **125**, e2019JE006276
- Wolf, E., Kopparapu, R., Airapetian, V., et al. 2019, Astro2020: Decadal Survey on Astronomy and Astrophysics, science white papers, 177
- Wunderlich, F., Godolt, M., Grenfell, J. L., et al. 2019, [A&A](#), **624**, A49
- Yan, F., Nortmann, L., Reiners, A., et al. 2023, [A&A](#), **672**, A107
- Zhou, L., Ma, B., Wang, Y., & Zhu, Y. 2022, [AJ](#), **164**, 203