

SIMTERFERE: An optical interferometry simulator for quantifying the coherent flux stability of VLTI/GRAVITY+

Reaching per mill stability: Application to exoplanet spectroscopy

J. R. Sauter^{1,2,*}, A. von Stauffenberg^{1,2}, G. Bourdarot³, W. Brandner¹, F. Eisenhauer³, L. Kreidberg¹, L. Labadie⁴, S. Scheithauer¹, D. Trevascus^{1,2}, and R. van Boekel¹

¹ Max-Planck-Institut für Astronomie, Königstuhl 17, 69117 Heidelberg, Germany

² Fakultät für Physik und Astronomie, Universität Heidelberg, Im Neuenheimer Feld 226, 69120 Heidelberg, Germany

³ Max-Planck-Institut für extraterrestrische Physik, Gießenbachstraße 1, 85748 Garching bei München, Germany

⁴ I. Physikalisches Institut, Universität zu Köln, Zùlpicher Str. 77, 50937 Köln, Germany

Received 19 December 2025 / Accepted 19 February 2026

ABSTRACT

The implementation of the GRAVITY+ Adaptive Optics (GPAO) system at VLTI enables unprecedented sensitivity and stability in optical interferometry. This allows high-precision characterization of directly imaged exoplanets at medium spectral resolution, providing a new pathway for studying planetary atmospheres. We aim to quantify and characterize the short- and long-term stability of GRAVITY+ through a consecutive seven-hour observation of the bright and stable star β Pictoris, providing a benchmark for future exoplanet observations. We developed SIMTERFERE, a data-driven simulation tool that reproduces GRAVITY+ on-star observations using ancillary instrument and telemetry data. By comparing the simulations with the measured coherent fluxes, we traced the origins of systematic flux variations and assessed their impact on exoplanet contrast measurements. We find that the $\sim 10\%$ variations are dominated by throughput changes driven by variable fiber coupling, which depends on wavefront stability, atmospheric dispersion, and residual fiber offsets. These variations appear as smooth continuum changes across wavelength and can be effectively mitigated using second-order polynomial corrections. After removing these instrumental effects, the remaining $\sim 1\%$ variations are almost purely of telluric origin, which we can reliably correct down to the photon-noise limit (0.1% precision) using a contrast spectrum approach with linear airmass interpolation. The GRAVITY+ interferometric instrument is highly stable: low-order continuum and telluric variations can be corrected with high precision, making it uniquely capable of high-fidelity characterization of directly imaged exoplanets.

Key words. atmospheric effects – instrumentation: adaptive optics – instrumentation: high angular resolution – instrumentation: interferometers – techniques: interferometric – planets and satellites: general

1. Introduction

The GRAVITY instrument (GRAVITY Collaboration 2017) at the Very Large Telescope Interferometer (VLTI) enables adaptive optics (AO)-corrected, phase-referenced, near-infrared interferometric beam combination. During its first five years of operation, GRAVITY has revolutionized optical interferometry, enabling high-spatial-resolution studies of the Galactic Center, active galactic nuclei, exoplanets, and young stellar objects (GRAVITY Collaboration 2018a,b; GRAVITY Collaboration 2018; GRAVITY Collaboration 2019b).

This success motivated a series of upgrades unified under the GRAVITY+ project (GRAVITY+ Collaboration 2022), which aims to significantly improve sky coverage, sensitivity, and stability. A central component of this upgrade is the implementation of the new GRAVITY+ Adaptive Optics (GPAO) system (GRAVITY+ Collaboration 2026) on the four 8 m unit telescopes (UTs). The GPAO system provides state-of-the-art extreme-AO wavefront correction using visible or infrared natural and laser guide stars. Initial tests demonstrate a factor-of-two improvement in the Strehl ratio.

Measuring the emission spectrum of an exoplanet in the presence of its much brighter host star remains extremely challenging because of the high contrast (typically 10^{-4} – 10^{-6}) that direct imaging methods must overcome. Detectability is closely tied to the achievable angular resolution, which dictates the size of the unresolved stellar point spread function (PSF). For a diffraction-limited telescope of aperture size D , the angular resolution scales as λ/D , while for an interferometer it scales as $\lambda/2B$, where B is the longest baseline, providing an order-of-magnitude improvement in resolving power (Eisenhauer et al. 2023). However, GRAVITY has not yet fully exploited its 130-m baselines because sensitivity remained dominated by stellar photon and speckle noise coupled into the single-mode fiber. The improved performance of GPAO substantially reduces these noise sources and enhances the sensitivity to exoplanet emission by approximately an order of magnitude (GRAVITY+ Collaboration 2026).

These advancements enable precise and high-cadence spectral characterization of exoplanets using GRAVITY's high-resolution grating ($R \approx 4000$). The K-band (1.97–2.40 μm) offers a unique spectral window to probe the chemical and cloud structure of young gas giants, placing the first isotopologue detections and variability measurements of directly imaged

* Corresponding author: sauter@mpia.de

exoplanets within reach. Such detections offer transformative insights into exoplanet formation histories and atmospheric dynamics (Gandhi et al. 2023; Mulder et al. 2025; Biller et al. 2024; Chen et al. 2025).

Both isotopologue and atmospheric-variability measurements require multi-hour integrations to achieve high signal-to-noise ratios (S/Ns) and to track hour-scale changes in atmospheric structure (Ravet et al. 2025). While GRAVITY+ demonstrates the sensitivity required for these observations, the spectrometric stability over multi-hour timescales remains uncharacterized.

The recent upgrade of GRAVITY to GRAVITY+ opens new and promising opportunities for ground-based exoplanet characterization. However, as the upgraded instrument was only recently deployed, its actual performance and limitations remain largely unexplored.

Recent analyses of exoplanet spectra obtained with GRAVITY reveal flux variations at the level of approximately 10% between multi-epoch observations. These variations are attributed to instrumental effects that are not yet fully understood (Kammerer et al. 2025), which limits the reliability of spectroscopic measurements and their physical interpretation.

The primary goal of this work is to characterize the spectrometric flux stability of GRAVITY+ by systematically investigating the extent and properties of instrumental flux variations. In addition, we aim to identify the underlying causes of these variations and to develop and validate mitigation strategies within the exogravity data-reduction framework.

We developed a data-driven simulation tool that reproduces GRAVITY+ on-star observations using ancillary instrument and telemetry data. This approach isolates individual instrumental processes, enabling detailed analysis and interpretation of systematic effects present in the data. The resulting framework provides a flexible diagnostic tool that is publicly available to the community for targeted performance studies.

A deeper understanding of the spectrometric stability and performance of GRAVITY+ is essential to improve current observations and guide future instrument development and observing strategies.

The exogravity¹ pipeline (GRAVITY Collaboration 2019a; Nowak et al. 2020, 2024a) utilizes the high spatial resolution of GRAVITY to detect and characterize directly imaged exoplanets. Following the notations of Balmer et al. (2025), the measured interferometric observable is the coherent flux Γ of an object at baseline \mathbf{b} , time t , and wavelength λ is

$$\Gamma(\mathbf{b}, t, \lambda) = F(t, \lambda) \cdot \mathcal{V}(\mathbf{b}, \lambda) \cdot e^{i\frac{2\pi}{\lambda}[\alpha \cdot U(\mathbf{b}) + \delta \cdot V(\mathbf{b})]}, \quad (1)$$

where $F \in \mathbb{R}$ is the spectral flux, $\mathcal{V} \in \mathbb{C}$ is the complex visibility, $(\alpha, \delta) \in \mathbb{R}^2$ is the angular sky position, and $(U, V) \in \mathbb{R}^2$ are the projected baseline coordinates in the uv -plane.

The flux contrast $C \in (0, 1)$ between the planet p and the star s is given by

$$\begin{aligned} C(t, \lambda) &= \frac{F_p(t, \lambda)}{F_s(t, \lambda)} \\ &= \frac{\Gamma_p(\mathbf{b}, t, \lambda) \mathcal{V}_s(\mathbf{b}, \lambda)}{\Gamma_s(\mathbf{b}, t, \lambda) \mathcal{V}_p(\mathbf{b}, \lambda)} e^{i\frac{2\pi}{\lambda}[\Delta\alpha \cdot U(\mathbf{b}) + \Delta\delta \cdot V(\mathbf{b})]}. \end{aligned} \quad (2)$$

The observed stellar coherent flux $\Gamma_{os} \in \mathbb{C}$ is modified by atmospheric and instrumental transmissivity $G \in (0, 1)$:

$$\Gamma_{os}(\mathbf{b}, t, \lambda) = G(\mathbf{b}, t, \lambda) \cdot \Gamma_s(\mathbf{b}, t, \lambda). \quad (3)$$

¹ <https://gitlab.obspm.fr/mnowak/exogravity>

The observed planetary coherent flux Γ_{op} is additionally influenced by residual stellar leakage from the stellar PSF at the science channel (SC) fiber position. This leakage can be modeled as a low-order wavelength polynomial $\mathcal{Q}_\lambda \in \mathbb{R}$ applied to the stellar coherent flux:

$$\begin{aligned} \Gamma_{op}(\mathbf{b}, t, \lambda) &= G(\mathbf{b}, t, \lambda) \cdot \Gamma_p(\mathbf{b}, t, \lambda) + \mathcal{Q}(\mathbf{b}, t, \lambda) \cdot \Gamma_{os}(\mathbf{b}, t, \lambda) \\ &= \Gamma_{os}(\mathbf{b}, t, \lambda) \cdot \\ &\quad \left[C(t, \lambda) \frac{\mathcal{V}_p(\mathbf{b}, \lambda)}{\mathcal{V}_s(\mathbf{b}, \lambda)} e^{-i\frac{2\pi}{\lambda}[\Delta\alpha \cdot U(\mathbf{b}) + \Delta\delta \cdot V(\mathbf{b})]} + \mathcal{Q}(\mathbf{b}, t, \lambda) \right]. \end{aligned} \quad (4)$$

The exogravity pipeline determines the best-fitting position and contrast spectrum through chi²-minimization:

$$\chi^2(\Delta\alpha, \Delta\delta, \mathcal{Q}, C) = [\hat{\mathbf{r}}_{op} - \mathbf{r}_{op}]^T \times \mathbf{W}^{-1} \times [\hat{\mathbf{r}}_{op} - \mathbf{r}_{op}], \quad (5)$$

where \mathbf{r}_{op} is the coherent flux model in matrix form, $\hat{\mathbf{r}}_{op}$ are the measured data, and \mathbf{W} is the respective covariance matrix. In this work, we denote matrix multiplications as \times and transpose as T .

In practice, the star and the planet cannot be observed simultaneously in the SC. The SC fiber is therefore periodically switched between the stellar and planetary positions. During on-planet observations, the on-star coherent flux is estimated as the average of the on-star measurements taken immediately before and after:

$$\begin{aligned} \Gamma_{os}(\mathbf{b}, t, \lambda) &\approx \frac{\hat{\Gamma}_{os}(\mathbf{b}, t + \Delta t_2, \lambda) + \hat{\Gamma}_{os}(\mathbf{b}, t - \Delta t_1, \lambda)}{2} \\ &\approx \frac{G(\mathbf{b}, t + \Delta t_2, \lambda) + G(\mathbf{b}, t - \Delta t_1, \lambda)}{2} \Gamma_s(\mathbf{b}, \lambda). \end{aligned} \quad (6)$$

The modeled on-planet coherent flux in Eq. (4) is therefore only valid for the minimization if both the transmissivity and stellar flux vary slowly in time.

Hence, time-variable transmissivity directly biases the derived contrast spectrum, making it a strong candidate for the observed instrumental flux variation. This highlights the need to thoroughly characterize the transmissivity of GRAVITY and develop strategies to mitigate or correct its temporal variations.

β Pictoris is a young (23 ± 3 Myr), nearby (19.44 ± 0.05 pc), and therefore bright ($K = 3.48$ mag) intermediate-mass ($1.75 M_\odot$) A6V star that hosts a prominent edge-on debris disk and two confirmed gas-giant planets (Crifo et al. 1997; Ducati 2002; Van Leeuwen 2007; Mamajek & Bell 2014; Nielsen et al. 2020; Nowak et al. 2020). Although β Pictoris is a known δ -Scuti variable star (Koen et al. 2003), its variability is small $\lesssim 1\%$ in the visible (Koen et al. 2003) and $\ll 1\%$ in the K band.

A multi-hour observation of the β Pictoris system is therefore ideal for tracking the temporal behavior of atmospheric and instrumental transmissivity. β Pictoris' near-constant K-band brightness enables high-S/N measurements while ensuring that temporal variations in the coherent flux primarily reflect instrumental or atmospheric effects rather than intrinsic stellar variability.

The paper is organized as follows. Section 2 summarizes our GRAVITY observations of the β Pictoris system and the corresponding data reduction. Section 3 introduces SIMTERFERE, the simulation framework used for our analysis and characterization of the data. We present the results in Sect. 4, and we discuss the limitations of the simulation tool in Sect. 5. Section 6 provides an overall discussion and concludes the paper.

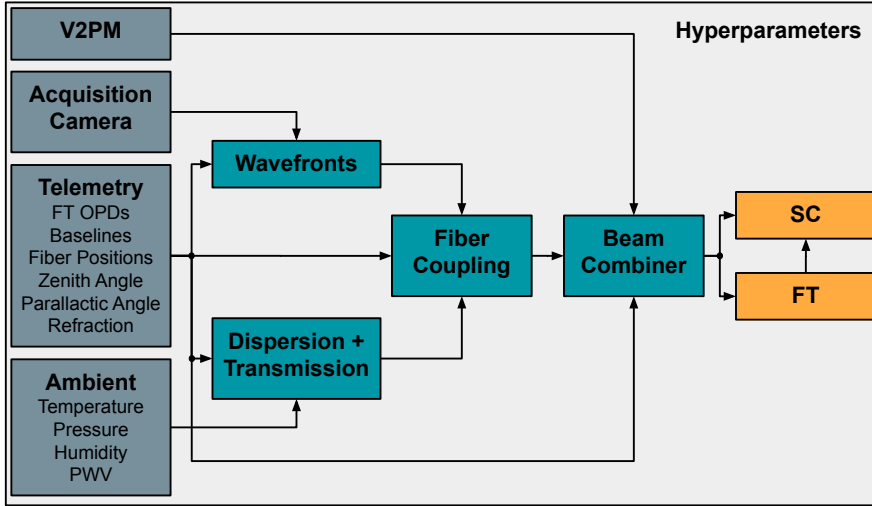


Fig. 1. Flow chart of the SIMTERFERE simulation framework. Simulated coherent fluxes (orange) are generated by injecting atmospheric and instrumental effects (blue), derived from the ancillary telemetry of the GRAVITY observations (dark gray). The overall configuration of the simulation and the subcomponents is controlled through user-defined hyperparameters (light gray).

Table 1. Exposure settings for the β Pictoris observations.

Target	#Exp.	#Int.	DIT (s)	Total (min)
β Pic	18	16	3	14.4
β Pic b	38	4	100	253.3
sky a	9	16	3	7.2
sky b	10	4	100	66.7

Notes. The different columns comprise the SC fiber pointings, total number of exposures, integrations per exposure, detector integration time (DIT), and combined total exposure times.

Table 2. Atmospheric conditions during observations of β Pictoris at the Paranal astronomical site.

Quantity	Value	Unit
Seeing ($\lambda = 500$ nm)	$0.69^{+0.09}_{-0.08}$	($''$)
Coherence time	$7.1^{+1.4}_{-1.3}$	(ms)
Wind speed	$1.0^{+0.7}_{-0.3}$	(m/s)
Precipitable water vapor	$1.56^{+0.04}_{-0.07}$	(mm)

Notes. Values are temporal medians with 16–84th percentile range uncertainties.

2. Data

We observed the β Pictoris system on 20 December 2024 for a total of seven consecutive hours using GRAVITY in the Guaranteed Time Observations (GTO) program 114.27JS (PI: L. Kreidberg), covering airmasses from 1.1 to 1.5. We carried out the observations with the four UTs in high-spectral-resolution mode ($R \approx 4000$) and dual-field on-axis mode configuration. In this setup, the fringe tracker (FT) fiber remains locked on the star for fringe tracking and phase referencing, while the SC fiber alternates between the predicted planet position, the host star, and dedicated sky locations. Table 1 summarizes the exposure settings for the SC fiber pointings.

As summarized in Table 2, the atmospheric conditions were generally good throughout the night. However, ground wind speeds were exceptionally low (<3 m/s), causing the low-wind effect (LWE; Milli et al. 2018; Pourré et al. 2022), which

degrades the quality and stability of the PSF. Sect. 4.2 discusses the impact of the LWE on our data.

We reduced the raw data using the ESO GRAVITY Public Release 1.9.4. pipeline and processed it up to the astroreduced intermediate data products. At this stage, individual integrations are kept separate rather than averaged, which is essential for our analysis of short-timescale stability. For on-star observations, the pipeline reports a median absolute coherent flux S/N of 180^{+60}_{-40} per spectral bin, integration, and baseline, with uncertainties corresponding to the 16–84th percentile range.

3. SIMTERFERE

The SIMTERFERE² tool is a publicly available Python-based data-driven interferometric simulator designed to reproduce VLTI/GRAVITY on-star observations using their associated ancillary data. The GRAVITY instrument is highly complex, with numerous astronomical, atmospheric, and instrumental processes acting simultaneously. Consequently, features observed in the coherent flux or visibility are often the product of multiple coupled effects, making it challenging to attribute observed systematics to individual subsystems. The SIMTERFERE package addresses this challenge by providing controlled access to the various contributors to the measurement process, allowing targeted diagnostic tests.

Using GRAVITY’s ancillary data, we can inject the behavior of key subsystems, such as AO (Arsenault et al. 2003; Scheithauer et al. 2016; GRAVITY+ Collaboration 2022, 2026), delay lines (Derie 2000; Pfuhl et al. 2012; Drescher et al. 2022), and fringe tracker (Lacour et al. 2019; Nowak et al. 2024b), directly into the simulation without requiring detailed modeling. This approach allows SIMTERFERE to include most instrumental effects while remaining computationally efficient, modular, and easily extendable.

Fig. 1 shows a schematic overview of the simulation workflow. The simulation mimics the propagation of stellar light through Earth’s atmosphere, telescopes, and the GRAVITY instrument, following the most prominent sequence of physical and instrumental effects that affect real observations.

Starting from a user-defined input spectrum, the simulation applies atmospheric transmission to account for wavelength-dependent absorption and throughput losses. The light is then

² <https://github.com/Manaffel/SIMTERFERE>

injected into the single-mode fibers of the four telescopes, where the simulation determines the coupling efficiency from measured wavefront stability and chromatic dispersion effects. The four beams are subsequently combined into the six interferometric baselines, and the recorded residual piston errors are applied to reproduce the temporal phase perturbations observed during real observations.

In a final step, the simulation interferes with the light using a linear model of the GRAVITY beam combiner, which includes the overall instrumental throughput. This procedure is performed independently for the FT and SC channels, taking into account their different integration times and spectral resolutions. The resulting simulated observables can then be directly compared to actual GRAVITY+ data, allowing instrumental effects to be isolated and studied in a controlled manner.

Each component of the simulation workflow is described in more detail in the following sections. For a comprehensive description of GRAVITY and its subsystems, we refer the reader to the instrument papers cited above.

3.1. Ancillary data

The GRAVITY ancillary data provides real-time measurements of atmospheric properties and instrumental performance, recorded as telemetry and ambient parameters. Using these data streams as direct inputs, SIMTERFERE reproduces the dominant effects seen in the observations without requiring full end-to-end simulations of each subsystem.

3.1.1. Visibility-to-pixel-matrix (V2PM)

The visibility-to-pixel-matrix $\mathbf{V2PM}(\lambda) \in \mathbb{C}^{6 \times 4 \times 10}$ (Tatulli et al. 2007) provides a linear description of the GRAVITY beam combiner at wavelength λ . The detected signal $\mathbf{P}(t, \lambda) \in \mathbb{R}_+^{6 \times 4}$ at time t , corresponding to the coherent fluxes $\mathbf{\Gamma}(t, \lambda) \in \mathbb{C}^6$ of the six baselines and the fluxes $\mathbf{f}(t, \lambda) \in \mathbb{R}_+^4$ of the four telescopes, is given by

$$\mathbf{P} = \mathbf{V2PM} \times [\mathbf{f}, \mathbf{\Gamma}]. \quad (7)$$

For simplicity, we omit time and wavelength dependence in the following equations. The GRAVITY instrument employs the ABCD-modulation scheme (Lacour et al. 2008), in which the light from each baseline is split into four beams with different phase shifts (i.e., $A : 0$, $B : \pi/2$, $C : \pi$, and $D : 3\pi/2$, for an ideal beam combiner). The measured $6 \times 4 = 24$ signals result from the combination of the six baselines with these four ABCD outputs.

During calibration, the instrument uses an internal light source. For each baseline combination $i = 1, \dots, 6$, only the respective two shutters are opened, and the 24 signals are measured while applying a phase modulation $\phi(t) \in (0, 2\pi)^6$ through the delay lines. The calibration signals $\hat{\mathbf{P}}(t, \lambda) \in \mathbb{R}_+^{6 \times 6 \times 4}$ are given by

$$\hat{P}_{i,j,k} = F_{i,\alpha_j,k} + F_{i,\beta_j,k} + 2\gamma_{i,j,k} \sqrt{F_{i,\alpha_j,k} F_{i,\beta_j,k}} \cos(\phi_i + \Delta\Phi_{i,j,k}), \quad (8)$$

where $\gamma(\lambda) \in (0, 1)^{6 \times 6 \times 4}$ is the coherence, $\alpha = [1, 1, 1, 2, 2, 3]$, $\beta = [2, 3, 4, 3, 4, 4]$ are the two interfering telescope indices for the six baselines $j = 1, \dots, 6$, and $k = 1, \dots, 4$ labels the four ABCD beams.

We obtain the V2PM by fitting the measured calibration signals to determine the fluxes $\mathbf{F}(\lambda) \in \mathbb{R}_+^{6 \times 4 \times 4}$, phase shifts $\Delta\Phi(\lambda) \in (0, 2\pi)^{6 \times 6 \times 4}$, and coherence $\gamma(\lambda) \in (0, 1)^{6 \times 6 \times 4}$. We then define the V2PM as

$$\mathbf{V2PM}_{i,k,l} = \begin{cases} F_{i,l,k} & l = 1, \dots, 4 \\ 2\gamma_{i,l-4,k} \sqrt{F_{i,\alpha_{l-4},k} F_{i,\beta_{l-4},k}} e^{i\Delta\Phi_{i,l-4,k}} & l = 5, \dots, 10, \end{cases} \quad (9)$$

and finally normalized to ensure unit throughput response:

$$\mathbf{V2PM}^*_{i,k,l} = \begin{cases} \frac{24 \cdot \mathbf{V2PM}_{i,k,l}}{\sum_{i,k=1}^{6,4} \mathbf{V2PM}_{i,k,l}} & l = 1, \dots, 4 \\ \frac{24 \cdot \mathbf{V2PM}_{i,k,l}}{\sqrt{\sum_{i,k=1}^{6,4} \mathbf{V2PM}_{i,k,\alpha_{l-4}} \sum_{i,k=1}^{6,4} \mathbf{V2PM}_{i,k,\beta_{l-4}}}} & l = 5, \dots, 10. \end{cases} \quad (10)$$

During science observations, the simulation computes the pixel-to-visibility-matrix $\mathbf{P2VM}(\lambda) \in \mathbb{C}^{10 \times 6 \times 4}$ as the pseudo-inverse of $\mathbf{V2PM}^*$ and applies to the measured signals to recover the coherent fluxes:

$$[\mathbf{f}, \mathbf{\Gamma}] = \mathbf{P2VM} \times \mathbf{P}. \quad (11)$$

In SIMTERFERE, individual V2PM matrices can be defined separately for the SC and FT. We then apply the P2VM following the procedure described above.

3.1.2. Acquisition camera

The GRAVITY acquisition camera (Anugu et al. 2014, 2016) performs beam alignment, active pupil and field stabilization, defocus correction, and engineering tasks. It comprises a pupil tracker, a field tracker, a pupil imager, and an aberration tracker, all recorded simultaneously on a single detector with a cadence of 0.7 s.

A dichroic in each telescope beam separates the H-band flux, directing it to the acquisition camera. The field images have a pixel scale of $p_{AC} = 17.8$ mas.

In SIMTERFERE, the four PSF images are extracted using fixed windows of size 75×75 pixel.

3.1.3. Telemetry and ambient parameters

The GRAVITY instrument software subsystem (ISS) computes relevant telemetric data in real time, including the zenith angle (for atmospheric dispersion and transmission), fiber offsets (for fiber coupling), baselines and fringe-tracker optical path differences (OPDs; for the beam combiner), and the parallactic angle (for wavefront sensing). It records these data in the FITS headers of the observation files.

The METEO and LHATPRO (Querel & Kerber 2014) systems at the VLT measure ambient parameters in real time. The simulations use ambient temperature, pressure, relative humidity, and perceptible water vapor (PWV) to compute atmospheric dispersion and transmission effects.

3.2. Wavefronts

The extracted acquisition camera field images of the star serve as underlying PSF references for the applied focal-plane wavefront sensing. For each image, we subtracted the temporally closest median sky observation to remove bias, dark current, and background. The fitting routine limits the maximum count to 40 000 ADU to mitigate saturation effects.

We modeled the VLT aperture using POPPY³ (Perrin et al. 2016) for wavefront reconstruction. Following Por et al. (2018), we modeled the primary mirror as a CircularAperture with radius 4.1 m, and the secondary mirror as an AsymmetricSecondaryObscuration with a radius of 0.558 m for UT1-3 and 0.6495 m for UT4. We placed four supporting spiders, each 0.04 m wide, around the mirror at angles of 39.5°, 140.5°, 219.5°, and 320.5°. We extracted the aperture using CompoundAnalyticOptic.get_transmission at $\lambda_H = 1.654 \mu\text{m}$ and placed the 33-pixel aperture model in the center of a 75×75 pixel zero-array to match the acquisition camera pixel scale at Nyquist sampling ($17.8 \text{ mas} \cdot 0.5D/\lambda \approx 75/33$).

We implemented the focal plane wavefront sensing following Feng et al. (2025) in PyTorch (Paszke et al. 2019). The simulation rotates the spiders in the aperture $\mathbf{A}_H(t) \in (0, 1)^{75 \times 75}$ according to the parallactic angle. It fits the phase $\Phi_H(t) \in (0, 2\pi)^{75 \times 75}$, amplitude $\text{amp}(t) \in \mathbb{R}_+$, and uniform background $\text{bg}(t) \in \mathbb{R}_+$ via χ^2 -minimization, assuming photon noise as the main error source in our loss function $\mathcal{L}(\text{amp}, \text{bg}, \Phi_H) \in \mathbb{R}_+$:

$$\mathcal{L} = \sum_{x=-37}^{37} \sum_{y=-37}^{37} \frac{(\text{amp} \cdot \text{PSF}_{x,y} + \text{bg} - \widehat{\text{PSF}}_{x,y})^2}{|\widehat{\text{PSF}}_{x,y}|}, \quad (12)$$

where $\text{PSF}(t) = \left| \mathcal{F} \left\{ \mathbf{A}_H \cdot e^{i\Phi_H} \right\} \right|^2 \in \mathbb{R}_+^{75 \times 75}$ is the PSF model, with \mathcal{F} denoting the discrete 2D Fourier transform. The simulation saturates the modeled PSF at 40 000 ADU to match the detector properties.

We initialized the minimization by first fitting a third-order 2D polynomial wavefront along with amplitude and background. We set the polynomial to zero, the amplitude to one, and the background to zero as an initial guess. The initial fit uses torch.optim.AdamW for 3000 iterations with a learning rate of 10^{-3} . The solution is then refined through a full phase minimization with 20,000 iterations, using the previous fit as the initial guess. For subsequent frames within an exposure, the simulation uses the solution from the previous frame as an initial guess to prevent phase jumps.

To extrapolate from H to K band, the simulation scales the K-band phase $\Phi_K(t) \in (0, 2\pi)^{99 \times 99}$ and aperture $\mathbf{A}_K(t) \in (0, 1)^{99 \times 99}$ by $\lambda_K/\lambda_H \approx 99/75$ by zero-padding at the edges.

Because the acquisition camera, fringe tracker, and science cadences are not synchronized, the simulation linearly interpolated the unwrapped phases for intermediate time steps.

Notably, the focal-plane wavefront reconstruction has sign ambiguity, meaning that the solution is not unique. Nevertheless, the obtained solutions reproduce statistical properties and PSF shapes, accurately modeling the effective throughput losses in the fiber coupler.

3.3. Dispersion

The simulation calculates atmospheric dispersion using an extension of the Cassini approximation (e.g. Young 2004), modeling a two-layer atmosphere with a moist layer at the observatory level and a dry layer above. We set the height of the moist layer to 2 km and adopted the refractive indices of moist and dry air from Mathar (2007). We set the ambient conditions of the observatory to the mean values during the observations: a temperature of 14.41°C and a pressure of 744.7 hPa.

³ <https://github.com/spacetelescope/poppy>

We precomputed a grid of diffraction curves for PWV values from 0.5 mm to 2.5 mm in 16 increments, and zenith angles from 39.7° to 54.0° in eight increments. We performed cubic interpolation of this regular grid using RegularGridInterpolator from SciPy (Virtanen et al. 2020) to retrieve the relative dispersion $(\text{disp}_x, \text{disp}_y)(t, \lambda) \in \mathbb{R}^2$ on the detector, where we adopted the detector xy-directions from the recorded telemetry refraction values.

3.4. Transmission

We computed the atmospheric transmission $G_{\text{telluric}}(t, \lambda) \in (0, 1)$ using LBLRTM⁴ via the Python wrapper TelFit⁵ (Clough et al. 1992, 2005; Gullikson et al. 2014). Following Eqs. (1) and (3), the transmitted flux $f_{\text{telluric}}(t, \lambda) \in \mathbb{R}_+$ is then given as

$$f_{\text{telluric}} = G_{\text{telluric}} f_s, \quad (13)$$

where $f_s(\lambda) \in \mathbb{R}_+$ denotes the input stellar spectrum.

We initialized the transmission using the Paranal observatory altitude (2635 m) and latitude (24°37'39" S). We used the recorded ambient pressure, temperature, relative humidity, and zenith angle as input parameters. The relative abundance of other molecular species (e.g. CO₂, CH₄, and CO) can either be set manually or retrieved using the fitting options of LBLRTM/TelFit.

3.5. Fiber coupling

We simulated the fiber as an ideal single-mode fiber using a Gaussian profile, with the fiber FWHM given by (Perrin & Woillez 2019)

$$\text{FWHM}_K = \frac{\lambda_K}{D}. \quad (14)$$

We computed the fiber profile $\mathbf{fp}(t, \lambda) \in \mathbb{C}^{99 \times 99}$ in pupil space as

$$\mathbf{fp} = e^{-2\pi^2 \sigma_{\text{fiber}}^2 (\mathbf{U}^2 + \mathbf{V}^2) + 2\pi i (x_{\text{fiber}} \mathbf{U} + y_{\text{fiber}} \mathbf{V})}, \quad (15)$$

where the fiber width σ_{fiber} scales continuously with wavelength (instead of the discretized aperture size):

$$\sigma_{\text{fiber}} = \frac{\text{FWHM}_K \lambda_K}{2 \sqrt{2 \ln 2} \lambda}. \quad (16)$$

Here, $\mathbf{U}, \mathbf{V} \in \mathbb{R}^{99 \times 99}$ are the detector coordinates in pupil space and $x_{\text{fiber}}/y_{\text{fiber}}(t) \in \mathbb{R}$ are the recorded fiber positions.

We calculated flux coupling $\eta(t, \lambda) \in (0, 1)$ as

$$\eta = \left| \sum_{x=-49}^{49} \sum_{y=-49}^{49} \mathbf{fp}_{x,y}^\dagger \mathbf{A}_{K,x,y} e^{i \left(\frac{\lambda_H}{\lambda} \Phi_{K,x,y} - \frac{2\pi}{99p_{AC}} [x \cdot \text{disp}_x + y \cdot \text{disp}_y] \right)} \right|. \quad (17)$$

3.6. Beam combiner

We computed the detected signals from the four telescope coupled fluxes $\boldsymbol{\eta}(t, \lambda) \in (0, 1)^4$ using the beam combination described by the V2PM components. Applying Eqs. (7) and (9), the output signals are given as

$$\mathbf{P}_{j,k} = \sum_{i=1}^6 (1 - \xi_{i,j,k}) \cdot (I_{i,\alpha_j,k} + I_{i,\beta_j,k}) + \xi_{i,j,k} I_{i,4+j,k}, \quad (18)$$

⁴ <https://github.com/AER-RC/LBLRTM>

⁵ <https://github.com/kgullikson88/Telluric-Fitter>

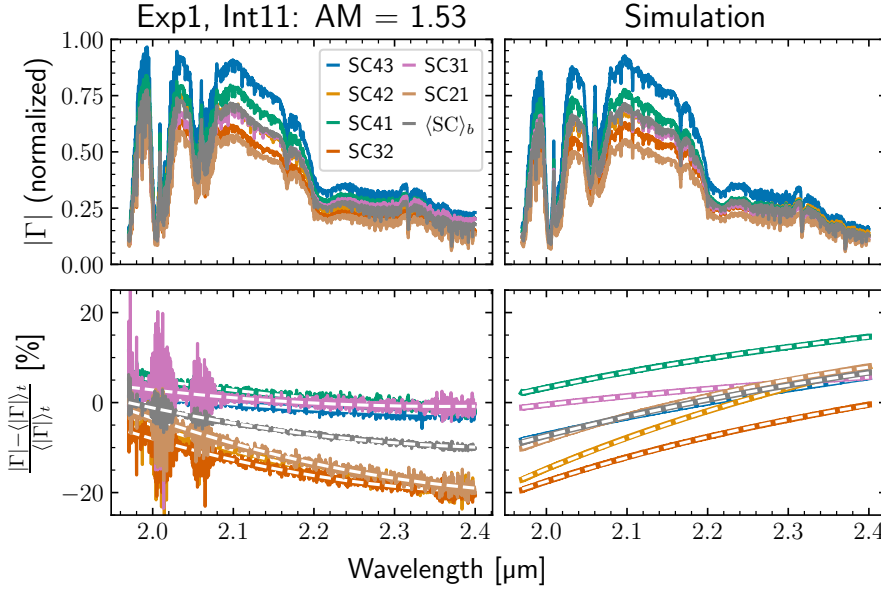


Fig. 2. *Top:* example frame (exposure 1/18, integration 11/16) showing measured (left) and simulated (right) absolute SC coherent flux per wavelength bin for the six baselines (43, ..., 12) and their baseline averages $\langle SC \rangle_b$ (different colors), taken at high airmass. Absolute coherent fluxes are normalized to the maximum flux in the exposure. *Bottom:* relative absolute coherent flux differences with respect to the time-averaged value for each baseline. Second order polynomial fits are indicated by dashed white lines. The systematic differences between the measured and simulated coherent fluxes are attributed to unknown fiber offsets, which are discussed in more detail in Sect. 4.1.2. The associated movie is available [online](#).

where

$$\xi_{i,j,k} = 6\gamma_{i,j,k} \sqrt{F_{i,\alpha_j,k} F_{i,\beta_j,k}}. \quad (19)$$

The uncombined signals $\mathbf{l}(t, \lambda) \in \mathbb{R}_+^{6 \times 10 \times 4}$ from the beam combiner output are

$$l_{i,l,k} = \begin{cases} |E_{i,l,k}|^2 & l = 1, \dots, 4 \\ |E_{i,\alpha_{l-4},k}|^2 + |E_{i,\beta_{l-4},k}|^2 + 2\Re \left\{ (E_{i,\alpha_{l-4},k})^\dagger E_{i,\beta_{l-4},k} e^{i(\phi_{\text{eff},i} \frac{\lambda_{\text{eff}}}{\lambda} + \Delta\Phi_{i,l-4,k})} \right\} & l = 5, \dots, 10, \end{cases} \quad (20)$$

with

$$E_{i,j,k} = \sqrt{F_{i,j,k} \eta_j} f_{\text{telluric}}. \quad (21)$$

Here $\phi_{\text{eff}}(t) \in (0, 2\pi)^6$ are the recorded FT residual piston errors.

Equations (18)–(21) describe how individual telescope fluxes, fiber coupling, atmospheric transmission, and residual piston errors combine to produce the final interferometric signals used by SIMTERFERE.

3.7. Hyperparameters

The hyperparameters are defined by the user to set the input stellar spectrum, specify telluric properties, adjust temporal and spectral resolution, modify sub-instrument performances, apply noise to the signals, and configure the V2PMs.

For this work, we adopted the NewEra PHOENIX model grid (Hauschildt et al. 2025) with an effective temperature of 8000 K, a logarithmic surface gravity of $\log g = 4$ cgs, and solar metallicity, approximately matching the properties of β Pictoris. We applied rotational broadening of 130 km/s (Royer et al. 2007) and a wavelength shift of -60 km/s or -0.45 nm to shift the Br γ absorption from vacuum to air (Mathar 2007, -80 km/s or -0.6 nm) and correct for the systematic velocity in the observed spectra (Gontcharov 2006, 20 km/s or 0.15 nm). We further binned the spectrum to match the GRAVITY SC and FT spectral resolution.

For telluric absorption, we empirically adjusted molecular abundances to match the measured spectra: H₂O at 130% of the measured ambient relative humidity, used as a proxy for the total abundance; CO₂ at 125%; and CH₄ at 140% of the default LBLRTM values. We left all other molecular abundances at their default settings.

The simulation subsamples each integration with 100 frames and OPDs. We empirically find that increasing the number of samples beyond 100 does not significantly change the simulated coherent fluxes. The simulation computes wavefronts and OPDs at this high cadence, whereas other quantities (e.g., dispersion and transmission) are evaluated only once per exposure.

No modifications to sub-instrument performance or additional noise are applied, allowing the focus to remain on systematic effects.

The simulation initializes the SC V2PM using the measured and normalized recorded V2PM from daytime calibration, consistent with the GRAVITY pipeline. It is then rescaled using the measured throughput for each baseline and wavelength stored in the calibration file’s C-Matrix. The FT V2PM is modeled as an ideal beam combiner.

4. Results

To characterize the transmissivity of GRAVITY and identify the sources of systematic effects, we simulated all on-star observations (18 exposures with each 16×3 s integration) using the configuration described in Sect. 3.7. By comparing these controlled simulations to the measured data, we can disentangle the contributions of instrumental, atmospheric, and fiber-coupling effects that are otherwise inseparable in the raw observations.

Figure 2 shows the measured and simulated absolute coherent fluxes for an example integration at the beginning of the night. A movie comprising all integrations of this exposure is available online. The simulations qualitatively reproduce the absolute flux levels, which are dominated by the combination of instrumental and atmospheric transmission.

Relative absolute coherent fluxes exhibit low-order variations on the order of 10%, reflecting systematic throughput fluctuations. These variations are primarily caused by time-dependent changes in fiber coupling, amplified by atmospheric

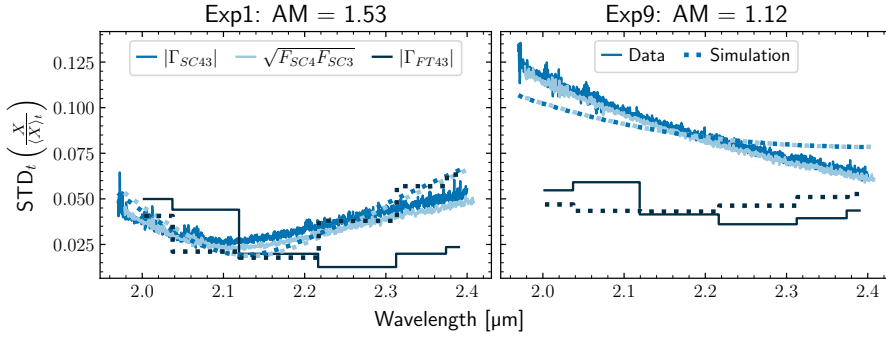


Fig. 3. *Left:* standard Deviation (STD) of the measured (solid lines) and simulated (dotted lines) relative SC (1628 spectral bins) absolute coherent flux $|\Gamma_{SC43}|$, the combined SC flux $\sqrt{F_{SC4}F_{SC3}}$, and the FT (6 spectral bins) absolute coherent flux $|\Gamma_{FT43}|$ (different shades) for an examples exposure at high airmass on baseline 43. *Right:* same as the left panel, but for an exposure at low airmass.

dispersion. They persist across all exposures and are well modeled by second-order wavelength polynomials $Q_{\lambda,2}$. Consequently, we can express the total effective transmissivity as

$$G(\mathbf{b}, t, \lambda) \approx Q_{\lambda,2}(\mathbf{b}, t, \lambda) \cdot G_{\text{instrument}}(\mathbf{b}, \lambda) \cdot G_{\text{telluric}}(t, \lambda). \quad (22)$$

The increased scatter in the measured normalized coherent fluxes arises from photon noise and chromatic dispersion, particularly in strong telluric absorption bands. We did not include these additional noise terms in the simulations.

4.1. Throughput variations

To quantify the impact of systematic low-order throughput variations on GRAVITY observations, we first examined variations within individual exposures and subsequently across multiple exposures.

4.1.1. Variations between integrations

Figure 3 shows the standard-deviations (STDs) of the measured and simulated normalized SC and FT absolute coherent fluxes, as well as combined SC flux-pairs, for two representative exposures at high and low airmass for baseline 43. We provide the complete plot with all baselines in Fig. A.1.

The simulations qualitatively reproduce the observed variations, capturing both the wavelength dependence and the airmass-related gradient introduced by throughput fluctuations. The FT generally shows better agreement with the simulations, whereas the SC exhibits higher STDs. We attribute this discrepancy primarily to fiber position offsets; we discuss this in more detail in Sect. 4.1.2.

The STDs of the absolute coherent fluxes closely match those of the combined fluxes. This similar behavior indicates that a common source causes the variations across all four telescopes. We observe residual excess noise in strong absorption bands, which likely originates from uncorrected chromatic dispersion. We therefore conclude that systematic throughput variations in each telescope flux are the primary source of scatter in both absolute and combined coherent fluxes.

The wavelength dependence of the STDs also exhibits a clear airmass effect: at high airmass, STDs show pronounced “v-shaped” gradients centered around the effective fiber coupling wavelength, reflecting increased atmospheric dispersion. At low airmass, the STDs decrease slightly with wavelength because reduced wavefront errors produce more stable PSFs.

4.1.2. Fiber offsets

Figure 4 illustrates that fiber positions play a critical role in the observed STD systematics. In the simulation, we computed fluxes using manually measured PSF centroids at $2.18 \mu\text{m}$

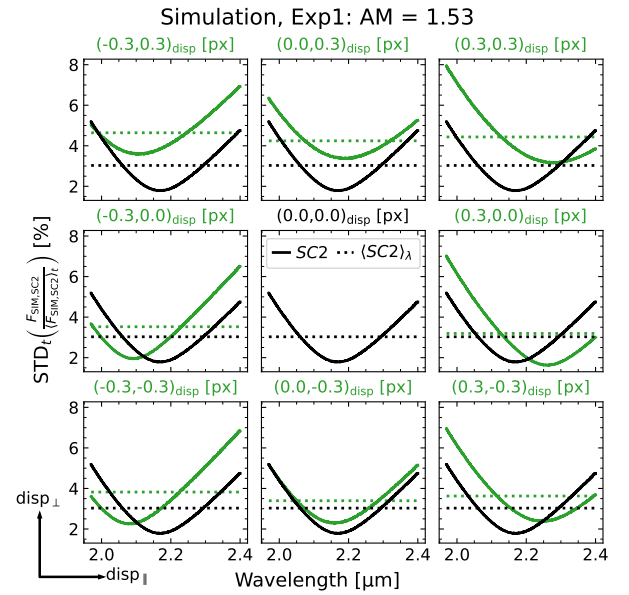


Fig. 4. Standard deviation (STD) of the simulated SC2 relative fluxes for different applied fiber offsets, either along or orthogonal to the dispersion direction. Black lines represent STDs with manual fiber tracking and no applied offsets, while green lines show STDs with applied offsets. The dotted lines indicate the respective wavelength average. An offset parallel to the dispersion direction shifts the characteristic “V-shape” toward positive wavelengths, and an antiparallel shift produces the opposite effect. Offsets orthogonal to the dispersion direction increase the overall STDs. The observed asymmetries between opposite offset directions arise from the asymmetry of the PSFs.

for each integration, rather than using the recorded fiber positions. We applied sub-pixel offsets (corresponding to ≈ 5 mas) to further probe the sensitivity of throughput to fiber alignment.

Offsets along the atmospheric dispersion direction shift the characteristic v-shaped STD pattern along the wavelength axis, while offsets orthogonal to the refraction direction increase the overall STD amplitude. These effects explain much of the discrepancy between the SC, FT, and simulated STDs, indicating that even small fiber misalignments can dominate systematic throughput variations.

Quantitatively, the STD shape and amplitude are extremely sensitive to fiber position: an offset of just 0.3 pixel on the acquisition camera can increase the variations by more than 50%. These findings suggest that the recorded fiber positions do not fully reflect the actual positions during the observations. The origin of these offsets remains unclear and warrants further investigation.

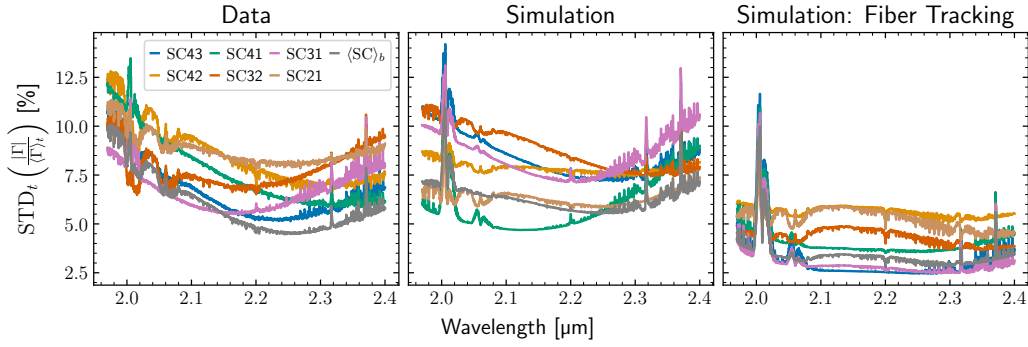


Fig. 5. *Left:* standard deviation (STD) of the measured relative absolute coherent SC fluxes for all exposures during the observation, shown for the six baselines (43, ..., 21) and the baseline average $\langle \cdot \rangle_b$ (different colors). *Middle:* same as the left panel, but showing the respective simulation results. *Right:* same as the middle panel, but with manual fiber tracking applied during the simulations.

Importantly, the mismatch between the SC and FT fiber positions prevents direct calibration of the SC fluxes using the FT.

4.1.3. Variations between exposures

Following the approach in Sect. 4.1.1, we computed the STDs of the measured and simulated normalized absolute coherent fluxes across all exposures over the full night. We also evaluated simulations with manual fiber tracking to assess the impact of fiber alignment.

Figure 5 shows that systematic throughput variations dominate on hour-long timescales, while telluric variations contribute only secondary, smaller fluctuations. The overall systematic variations of $\sim 10\%$ are consistent with previously reported multi-epoch differences in exogravity spectra (Kammerer et al. 2025).

The measured and simulated STDs agree qualitatively, but individual baselines show differing offsets and trends. Comparison of simulations with and without manual fiber tracking demonstrates that these discrepancies primarily result from misaligned fiber positions. Under ideal fiber tracking, observed throughput variations could be reduced by more than a factor of two.

4.2. Low wind effect

Throughout the night, low wind speeds (< 3 m/s) degraded the quality and stability of the PSF due to the LWE. Figures 6 and 7 show example PSF and wavefront fit frames under moderate and strong LWE conditions. Movies comprising all frames for the two exposures are available online. We derived the PSF fits using focal-plane wavefront sensing within SIMTERFERE. As the LWE strengthens, phase discontinuities appear between the four spider quadrants, causing increased light leakage into the side lobes of the PSF. This effect is particularly pronounced for UT1 and UT2, which lack specialized spider coatings (Pourré et al. 2022).

Figure 8 presents the measured and simulated relative UT fluxes for each exposure. The simulations reproduce the overall trends and agree reasonably well with the individual UTs and the average fluxes. The remaining differences are primarily due to misaligned fiber positions, as discussed in Sect. 4.1.2.

The flux trends show an overall complex behavior influenced by airmass, AO performance, and fiber positioning. The derived Strehl ratio, representative of AO performance, only loosely correlates with the seeing conditions because the LWE further

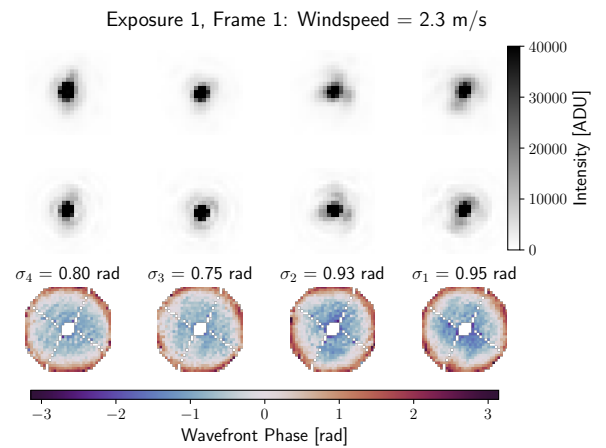


Fig. 6. *Top:* example PSF frames of the four UTs (from left to right: 4, ..., 1) from the acquisition camera at moderate wind speed, saturated at 40 000 ADU. *Center:* PSF fits to the data. *Bottom:* retrieved tip-tilt-subtracted wavefronts from the PSF fits. The wavefront errors are measured using the standard deviation. The associated movie is available [online](#).

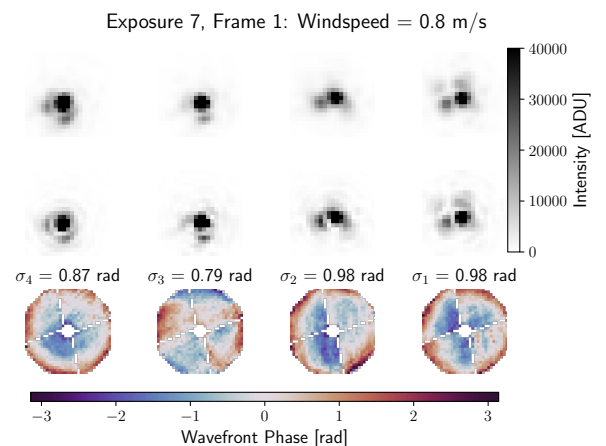


Fig. 7. Same as Fig. 6, but at very low wind speed. Strong phase discontinuities between the spiders are present in the UT1 and UT2 wavefronts, introducing pronounced side lobes in the PSFs. The associated movie is available [online](#).

degrades it. The UT3 and UT4 telescopes consistently exhibit higher Strehl ratios than UT1 and UT2, reflecting the improved PSF stability provided by the spider coatings.

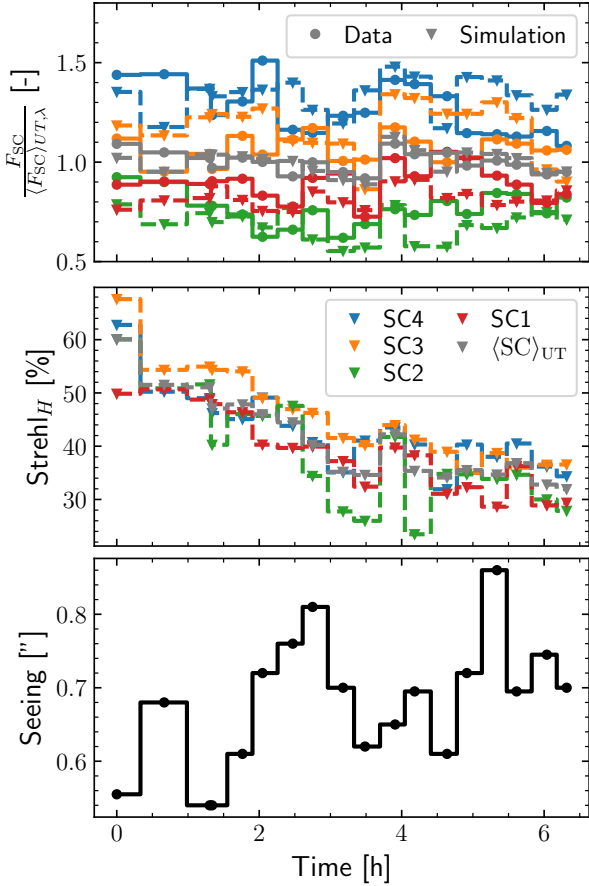


Fig. 8. *Top:* measured (dots, solid lines) and simulated (triangles, dashed lines) relative SC fluxes of the four UTs (4, ..., 1) and their average (different colors) for all exposures during the observation. *Center:* respective H-band Strehl ratios derived from focal-plane wavefront sensing. *Bottom:* recorded seeing during the observing night.

4.3. Polynomial correction

As described in Sect. 1, we extracted exoplanet spectra using the stellar spectrum measured immediately before and after each observation in exogravity. To quantify the precision of this approach, we computed the root-mean-square (RMS) relative differences of the absolute coherent fluxes between consecutive star exposures.

As shown in Sect. 4, the relative throughput variations can be modeled as low-order polynomials in wavelength. We therefore applied a throughput correction by fitting second-order polynomials $Q_{\lambda,2}(\mathbf{b}, t, \lambda) \in \mathbb{R}$:

$$Q_{\lambda,2} \approx \frac{|\Gamma|}{\langle |\Gamma| \rangle_t}, \quad (23)$$

which we then applied to the absolute fluxes to obtain the throughput corrected coherent fluxes $|\Gamma|_Q(\mathbf{b}, t, \lambda) \in \mathbb{C}$:

$$|\Gamma|_Q = \frac{|\Gamma|}{Q_{\lambda,2}}. \quad (24)$$

Figure 9 shows the measured and simulated relative SC RMS coherent flux differences before and after polynomial calibration. The mean relative RMS decreases from $6.5^{+0.7}_{-0.4}\%$ to $0.21^{+0.49}_{-0.13}\%$, with the residuals dominated by telluric variations.

Although the simulated fluxes with and without manual fiber tracking exhibit markedly different RMS trends prior to correction, the residuals converge after polynomial calibration. This result demonstrates the effectiveness of low-order polynomial correction in mitigating systematic throughput variations.

4.4. Telluric variations

To quantify residual variations after polynomial correction, we measured the baseline- and wavelength-averaged relative change in absolute coherent flux between consecutive polynomial-corrected exposures as a function of airmass change. We compared the measured and simulated data with a pure telluric model computed with the same atmospheric parameters used in the simulations.

Figure 10 shows that the coherent flux increases approximately linearly for large airmass changes but converges to the noise floor for small airmass changes. As expected, this noise floor is absent in both the simulations and the pure telluric model. We fit the behavior with a combination of a constant noise term and a linear term proportional to airmass change:

$$\left| \frac{\Delta|\Gamma|}{|\Gamma|} \right|(\lambda, \Delta AM) = \sqrt{\text{SNR}_{|\Gamma|}^{-2}(\lambda) + \left[\frac{\partial|\Gamma|}{\partial AM \cdot |\Gamma|}(\lambda) \cdot \Delta AM \right]^2}. \quad (25)$$

The three curves show good agreement with the fitted curves, with scattering dominated by non-airmass-dependent variations such as changes in pressure, temperature, and water vapor. Deviations at the largest airmass changes are attributed to small biases introduced by the polynomial correction.

Following Beer–Lambert’s law, the slope with respect to airmass directly measures the telluric transmissivity at unit airmass:

$$\frac{\partial|\Gamma|}{\partial AM \cdot |\Gamma|}(\lambda) = \frac{\partial \ln G_{\text{telluric}, AM=1}}{\partial AM}(\lambda) = \ln G_{\text{telluric}, AM=1}(\lambda). \quad (26)$$

Figure 11 shows the wavelength-dependent measured relative noise S/N^{-1} and airmass slopes $\partial|\Gamma|/\partial AM \cdot |\Gamma|$. The measured noise clearly lies above the pipeline prediction assuming no correlation (i.e., $1/\sqrt{6}$ scaling) and is only slightly below the fully correlated noise estimate at intermediate wavelengths. The increase at the low and high ends results primarily from polynomial correction biases.

These results highlight the strong correlation among the six baseline coherent fluxes, which must be taken into account in noise estimation and propagation for GRAVITY data.

Simulations, measurements, and pure telluric models show excellent agreement, indicating that polynomial-corrected fluxes are dominated by telluric variations. The measured RMS residuals of 0.019 and 0.013 correspond to relative flux errors of 0.05% and 0.035% at median airmass changes, well below the intrinsic data noise.

In the K band, the main telluric absorbers are CO_2 , CH_4 , and H_2O : CO_2 absorbs mainly below $2.1 \mu\text{m}$, CH_4 above $2.2 \mu\text{m}$, and H_2O below $2 \mu\text{m}$ and above $2.3 \mu\text{m}$. We included CO , an important exoplanet atmosphere tracer, as a reference. The CO telluric absorption from $2.3 \mu\text{m}$ to $2.4 \mu\text{m}$ (peak slope of 0.0035) is negligible when dealing with GRAVITY spectra.

The residual discrepancies between the simulated and telluric slopes are caused by the imperfect polynomial correction. Variable telluric line depths, uncorrelated with the systematic throughput, bias the polynomial fits, particularly at the low- and

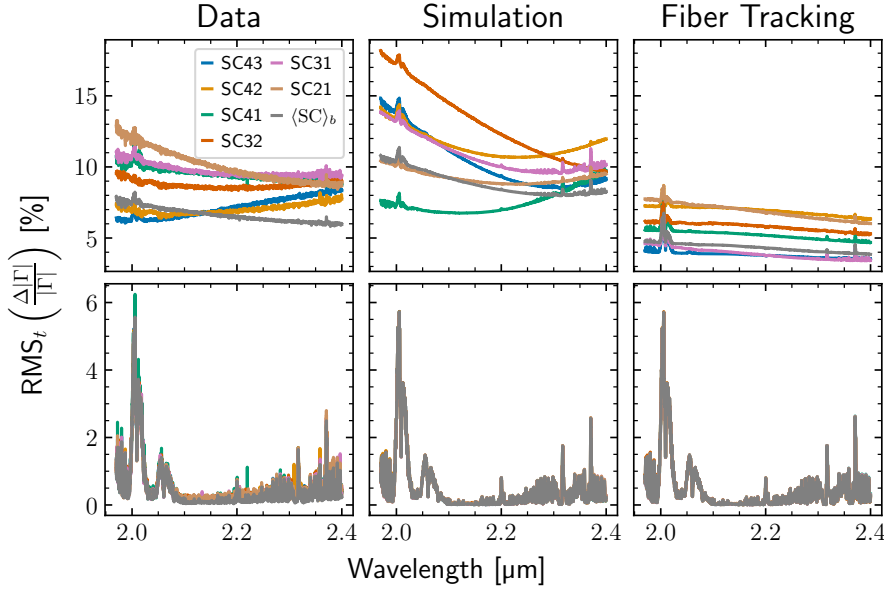


Fig. 9. *Top:* root-mean-square (RMS) of the relative absolute SC coherent flux between two subsequent exposures, shown per wavelength bin for the data (left), the simulations (center), and the simulations with applied fiber tracking, for the six baselines (43, ..., 21) and the baseline average $\langle \cdot \rangle_b$ (different colors). *Bottom:* same as top panel, but after applying the polynomial corrections.

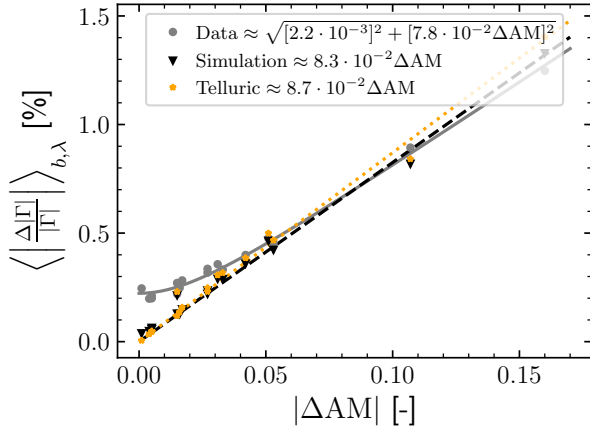


Fig. 10. Baseline- and wavelength-averaged relative absolute coherent flux change between two subsequent stellar pointings, shown as a function of absolute airmass change. The simulations (triangles, dashed lines) and the pure telluric models (stars, dotted lines) do not include noise, which results in a linear behavior. For the data (dots, solid lines), the flux change instead converges toward the noise floor at small airmass changes.

high-wavelength ends. To minimize this effect, we fit the polynomials only at wavelengths where the modeled transmission exceeds 99%, leaving the fits still imperfect.

4.5. Implications for exoplanet spectroscopy

As outlined in Sect. 1, the precision of the predicted on-star coherent flux during on-planet observations scales linearly with the retrieved contrast spectrum precision in exogravity. Quantifying both the accuracy and long-term stability of this prediction is therefore essential for high-fidelity exoplanet spectroscopy.

To evaluate prediction precision, we compared the measured polynomial-corrected absolute coherent fluxes $|\Gamma_i|$ with the predicted fluxes derived from the preceding and subsequent exposures $|\Gamma_{i-1/i+1}|$.

Following Eq. (6), in the current implementation of exogravity, we obtain the predicted coherent flux by

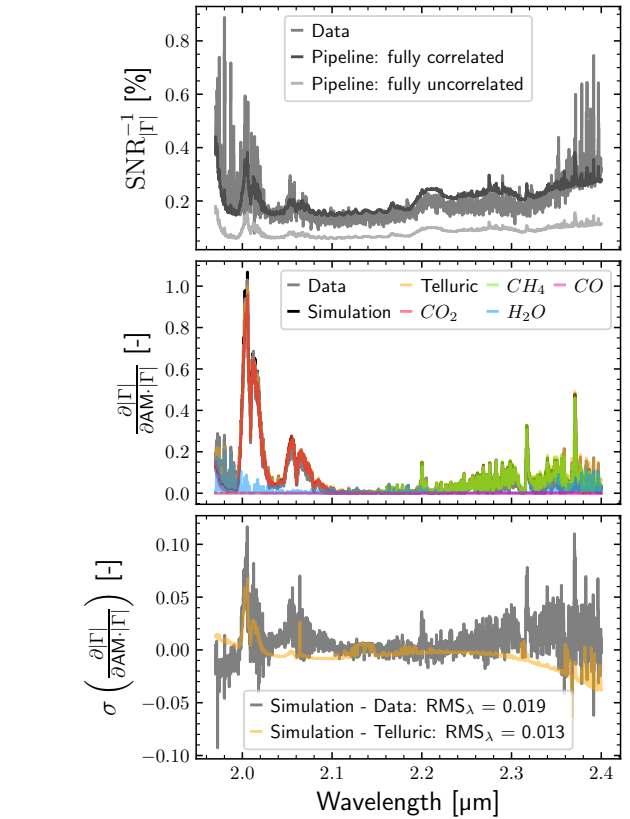


Fig. 11. *Top:* relative absolute coherent flux noise as a function of wavelength, measured in the data and estimated by the GRAVITY pipeline assuming fully correlated and uncorrelated baselines. *Center:* relative absolute coherent flux slopes as a function of wavelength for the data, the simulations, and the pure telluric models, including their respective molecular contributions. *Bottom:* residuals of the relative absolute coherent flux slopes between the simulations, the data, and the telluric models.

averaging the absolute coherent fluxes before and after the science exposure:

$$|\Gamma_i| \approx |\Gamma_{\text{Averaged}}| = \frac{|\Gamma_{i-1}| + |\Gamma_{i+1}|}{2}. \quad (27)$$

As derived in Sect. 4.4, we achieved a more accurate prediction by applying a weighted linear interpolation as a function of airmass:

$$\begin{aligned}
 |\Gamma_i| &\approx |\Gamma_{\text{Interpolated}}| \\
 &= w \cdot \left(\frac{(|\Gamma_{i+1}| - |\Gamma_{i-1}|)(\text{AM}_i - \text{AM}_{i+1})}{\text{AM}_{i+1} - \text{AM}_{i-1}} + |\Gamma_{i+1}| \right) \\
 &\quad + (1 - w) \cdot \Gamma_{\text{Averaged}},
 \end{aligned} \tag{28}$$

where the weight factor $w \in (0, 1)$ is defined as

$$w = \frac{(\text{AM}_{i+1} - \text{AM}_{i-1})^2}{\delta^2 + (\text{AM}_{i+1} - \text{AM}_{i-1})^2}. \tag{29}$$

This weighting scheme suppresses numerical instabilities for very small airmass changes ($|\text{AM}_{i+1} - \text{AM}_{i-1}| \rightarrow 0$), while smoothly transitioning to interpolation for large changes. We empirically find $\delta \approx 0.03$ to provide good improvements for large airmass changes while still suppressing noise amplification for low changes. For airmass changes below 0.03, photon noise dominates over telluric variability, as illustrated in Fig. 10.

We define the precision as the wavelength RMS of the relative difference between the measured True and Predicted absolute coherent flux. Figure 12 shows that the averaging approach exhibits a strong decrease in precision at high airmass or during periods of rapid airmass change, leading to increased temporal scatter. In contrast, the interpolation approach largely removes this dependence, yielding photon-noise-dominated residuals that remain stable throughout the full seven-hour observation for both individual baselines and the baseline-averaged fluxes. We observe no systematic drift over the full time series, demonstrating the long-term spectrometric stability of GRAVITY+ on multi-hour timescales.

Table 3 summarizes the median prediction precision achieved across the full time series, together with the corresponding 16–84th percentile ranges. For individual baselines, the interpolation approach yields a consistent improvement of approximately 10% compared to simple averaging. For the baseline-averaged fluxes, where the photon noise is further reduced, the improvement reaches approximately 30%. In this case, the averaging approach achieves a median precision of $0.16^{+0.11}_{-0.04}\%$, while the interpolation approach reaches a precision of $0.117^{+0.044}_{-0.010}\%$.

This combination of high precision, low intrinsic scatter, and temporal stability over the continuous seven-hour observation demonstrates that the polynomial-corrected and airmass-interpolated coherent-flux prediction provides a robust foundation for high-fidelity, long-duration spectroscopic characterization of exoplanets with GRAVITY+.

5. Limitations

As demonstrated in Sect. 4, SIMTERFERE qualitatively reproduces the absolute coherent fluxes and their short-term variations for our high-S/N (≈ 180), short-exposure (48 s) β Pictoris observations. On these timescales, the approximations made for time-discretized wavefronts, atmospheric dispersion, and transmission are thus adequate.

However, several limitations remain:

1. For longer exposures, atmospheric conditions (e.g., temperature, pressure, and water vapor) may vary during the integration, requiring a more detailed temporal treatment of wavefronts and telluric effects;

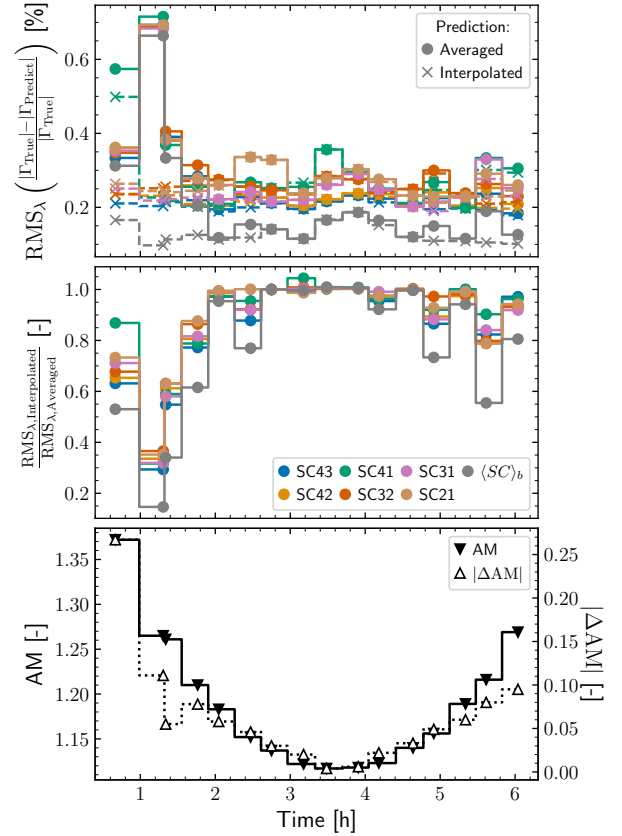


Fig. 12. *Top:* wavelength RMS of the relative difference between the measured True and Predicted absolute coherent flux as a function of time. The RMS values are shown for each interferometric baseline and for the baseline average (different colors), and are computed for each on-star exposure with at least one preceding and subsequent calibration exposure, using averaging (dots, solid lines) and interpolation (crosses, dashed lines) approaches. *Center:* ratio between the RMS obtained with the averaging and interpolation methods for each exposure. *Bottom:* airmass (filled triangles, solid lines) at the time of observation and absolute airmass difference (open triangles, dotted lines) between the preceding and subsequent calibration exposures. Because the observations are obtained at approximately constant time intervals, the airmass differences are strongly correlated with the airmass.

Table 3. Achieved absolute coherent flux precisions and corresponding ratios for the average and interpolation prediction method.

Baseline	Averaged (%)	Interpolated (%)	Ratio (-)
SC43	$0.224^{+0.089}_{-0.026}$	$0.204^{+0.011}_{-0.009}$	0.91
SC42	$0.235^{+0.086}_{-0.029}$	$0.215^{+0.017}_{-0.014}$	0.92
SC41	$0.27^{+0.10}_{-0.04}$	$0.25^{+0.05}_{-0.04}$	0.93
SC32	$0.269^{+0.065}_{-0.028}$	$0.248^{+0.027}_{-0.014}$	0.92
SC31	$0.25^{+0.09}_{-0.03}$	$0.224^{+0.033}_{-0.008}$	0.90
SC21	$0.28^{+0.07}_{-0.04}$	$0.245^{+0.059}_{-0.015}$	0.88
$\langle \text{SC} \rangle_b$	$0.16^{+0.11}_{-0.04}$	$0.117^{+0.044}_{-0.010}$	0.73

Notes. The precision is defined as RMS temporal median with 16–84th percentile range uncertainties per baseline and baseline average. The ratio is given as the quotient between the median averaged and interpolated value.

2. SIMTERFERE does not compute detector images containing the 24 spectral traces, but instead computes the signals directly. As a result, optical effects (e.g., line spread functions or curved spectral traces) and detector effects (e.g., bad pixels or flat fields) that are corrected by the GRAVITY pipeline are not included. Nevertheless, the agreement between the data and the simulations verifies the excellent performance of the pipeline;
3. SIMTERFERE currently does not implement measurement noise (i.e. photon noise, background noise, and detector readout noise) or its propagation through the P2VM method. These effects become particularly significant for low-S/N targets, such as faint exoplanets. In Sect. 4.1.1, we show that the dominant noise in the β Pictoris observations arises from chromatic dispersion, particularly in the CO₂ atmospheric absorption bands, which are also not included in SIMTERFERE;
4. The current implementation supports only on-star observations in the dual-field on-axis mode. Supporting on-planet simulations would require the implementation of a two-source mode, in which the stellar and planetary contributions are simulated separately and then combined. Additional optical effects such as anisoplanatism or anisoplanatism would also need to be taken into account.

6. Discussion and conclusion

The on-star coherent fluxes play a central role in measuring exoplanet contrast spectra. First, they provide the linear reference against which the on-planet coherent fluxes are fitted. Second, any systematic effects that influence the stellar fluxes are expected to influence the planetary fluxes. A robust characterization of the stellar coherent fluxes is therefore essential for quantifying the fidelity of extracted exoplanet spectra.

Our continuous seven-hour GRAVITY observation of the β Pictoris system offers a unique dataset for this purpose: a high-S/N, non-variable target that enables a systematic decomposition of the temporal and chromatic effects in the coherent fluxes. To interpret these data, we developed the data-driven simulation framework SIMTERFERE, which reproduces the observed flux behavior using the ancillary data recorded during the observations. This approach allows us to identify the physical origin of each systematic effect in isolation, which is otherwise inaccessible in the real data.

Our analysis demonstrates that the dominant source of coherent-flux variations arises from non-constant single-mode fiber coupling. The coupling efficiency is set by the interplay between PSF stability, atmospheric dispersion, and fiber-position offsets. These effects produce wavelength-dependent throughput variations on the order of 10%, a level consistent with the multi-epoch differences previously reported in exogravity exoplanet spectra (Kammerer et al. 2025). Importantly, SIMTERFERE shows that these variations are systematic, not stochastic, and largely predictable from the measured AO performance.

We find that these coupling-induced variations manifest primarily as low-order chromatic changes, which can be effectively removed through second-order polynomial throughput corrections. Once this correction is applied, the remaining inter-exposure variability is dominated by atmospheric transmission changes. These telluric variations follow the Beer-Lambert law with high precision. Consequently, instead of the current exogravity pipeline approach of averaging the on-star exposures taken before and after the planet observation, a linear

interpolation with airmass provides a more accurate and physically consistent reference for the stellar coherent flux.

After polynomial correction, the residual variations in the coherent fluxes fall to the level expected from telluric changes and noise, with no evidence for significant additional systematics. This demonstrates the excellent performance of the GRAVITY instrument and pipeline and shows that coherent-flux stability is fundamentally limited by atmospheric transmission rather than instrumental drifts once fiber-coupling systematics are removed.

We demonstrate that the GRAVITY+ instrument achieves exceptional long-term on-star stability, allowing the stellar coherent flux to be predicted with a precision of approximately 0.1% over the entire seven-hour observing sequence. This precision is limited only by photon noise, highlighting the capability of the instrument to support high-fidelity exoplanet spectroscopy over multi-hour observations.

In forthcoming work, we will extend our analysis to the on-planet observations within our β Pictoris dataset. By extending and applying SIMTERFERE, we will test the exogravity pipeline end-to-end, quantify possible biases in the planet extraction, and evaluate the impact of residual stellar and telluric systematics on the recovered planet spectrum.

Data availability

Supplementary data underlying the results presented in this article are available at <https://zenodo.org/records/17817680>. The SIMTERFERE framework is publicly available at <https://github.com/Manaffel/SIMTERFERE>. Movies associated to Figs. 2, 6–7 are available at <https://www.aanda.org>

Acknowledgements. The authors thank the anonymous referee for the constructive comments and suggestions, which helped to improve the quality of this work. This work made use of the Python packages: Astropy (Astropy Collaboration 2013, 2018, 2022), Numpy (Harris et al. 2020), LBLRTM/TelFit (Clough et al. 1992, 2005; Gullikson et al. 2014), Matplotlib (Hunter 2007), Poppy (Perrin et al. 2016), PyTorch (Paszke et al. 2019), Scipy (Virtanen et al. 2020), and Seaborn (Waskom 2021). Based on observations collected at the European Southern Observatory under ESO programmes 114.27JS.

References

- Anugu, N., Garcia, P. J. V., Wiegrecht, E., et al. 2014, in *24th Portuguese National Astronomy and Astrophysics Meeting*, eds. N. Gomes & J. Ascenso, 13
- Anugu, N., Garcia, P., Amorim, A., et al. 2016, in *Optical and Infrared Interferometry and Imaging V*, 9907, eds. F. Malbet, M. J. Creech-Eakman, & P. G. Tuthill, International Society for Optics and Photonics (SPIE), 990727
- Arsenault, R., Alonso, J., Bonnet, H., et al. 2003, in *Adaptive Optical System Technologies II*, 4839, eds. P. L. Wizinowich & D. Bonaccini, International Society for Optics and Photonics (SPIE), 174
- Astropy Collaboration (Robitaille, T. P., et al.) 2013, *A&A*, **558**, A33
- Astropy Collaboration (Price-Whelan, A. M., et al.) 2018, *AJ*, **156**, 123
- Astropy Collaboration (Price-Whelan, A. M., et al.) 2022, *AJ*, **935**, 167
- Balmer, W. O., Franson, K., Chomez, A., et al. 2025, *AJ*, **169**, 30
- Billler, B. A., Vos, J. M., Zhou, Y., et al. 2024, *MNRAS*, **532**, 2207
- Chen, X., Biller, B. A., Tan, X., et al. 2025, *MNRAS*, **539**, 3758
- Clough, S. A., Iacono, M. J., & Moncet, J.-L. 1992, *J. Geophys. Res.*, **97**, 15,761
- Clough, S. A., Shephard, M. W., Mlawer, E. J., et al. 2005, *JQSRT*, **91**, 233
- Crifo, F., Vidal-Madjar, A., Lallemand, R., Ferlet, R., & Gerbaldi, M. 1997, *A&A*, **320**, L29
- Derie, F. 2000, in *Interferometry in Optical Astronomy*, 4006, eds. P. J. Lena, & A. Quirrenbach, International Society for Optics and Photonics (SPIE), 25
- Drescher, A., Fabricius, M., Shimizu, T., et al. 2022, in *Optical and Infrared Interferometry and Imaging VIII*, 12183, eds. A. Mérand, S. Sallum, & J. Sanchez-Bermudez, International Society for Optics and Photonics (SPIE), 121830T
- Ducaiti, J. R. 2002, *yCat*, **2237**
- Eisenhauer, F., Monnier, J. D., & Pfuhl, O. 2023, *ARA&A*, **61**, 237

- Feng, B. Y., Ferrer-Chávez, R., Levis, A., et al. 2025, arXiv e-prints [arXiv:2501.01912]
- Gandhi, S., de Regt, S., Snellen, I., et al. 2023, arXiv e-prints [arXiv:2311.05349]
- Gontcharov, G. A. 2006, *Astron. Lett.*, **32**, 759
- GRAVITY Collaboration 2018, *Nature*, **563**, 657
- GRAVITY Collaboration (Abuter, R., et al.) 2017, *A&A*, **602**, A94
- GRAVITY Collaboration (Abuter, R., et al.) 2018a, *A&A*, **615**, L15
- GRAVITY Collaboration (Abuter, R., et al.) 2018b, *A&A*, **618**, L10
- GRAVITY Collaboration (Lacour, S., et al.) 2019a, *A&A*, **623**, L11
- GRAVITY Collaboration (Perraut, K., et al.) 2019b, *A&A*, **632**, A53
- GRAVITY+ Collaboration (Abuter, R., et al.) 2022, *Msngr.*, **189**, 17
- GRAVITY+ Collaboration (Allouche, F., et al.) 2026, *A&A*, **707**, A115
- Gullikson, K., Dodson-Robinson, S., & Kraus, A. 2014, *AJ*, **148**, 53
- Harris, C. R., Millman, K. J., Van Der Walt, S. J., et al. 2020, *Nature*, **585**, 357
- Hauschildt, P. H., Barman, T., Baron, E., Aufdenberg, J. P., & Schweitzer, A. 2025, *A&A*, **698**, A47
- Hunter, J. D. 2007, *CiSE*, **9**, 90
- Kammerer, J., Winterhalder, T. O., Lacour, S., et al. 2025, *A&A*, **704**, A318
- Koen, C., Balona, L. A., Khadaroo, K., et al. 2003, *MNRAS*, **344**, 1250
- Lacour, S., Jocou, L., Moulin, T., et al. 2008, in *Optical and Infrared Interferometry*, ed. M. Schöller, W. C. Danchi, & F. Delplancke, Vol. 7013, *Int. Soc. Opt. Photon. (SPIE)*, 701316
- Lacour, S., Dembet, R., Abuter, R., et al. 2019, *A&A*, **624**, A99
- Mamajek, E. E., & Bell, C. P. M. 2014, *MNRAS*, **445**, 2169
- Mathar, R. J. 2007, *J. Opt. A: Pure Appl. Opt.*, **9**, 470
- Milli, J., Kasper, M., Bourget, P., et al. 2018, in *Adaptive Optics Systems VI*, 10703, eds. L. M. Close, L. Schreiber, & D. Schmidt, International Society for Optics and Photonics (SPIE), 107032A
- Mulder, W., de Regt, S., Landman, R., et al. 2025, *A&A*, **694**, A164
- Nielsen, E. L., De Rosa, R. J., Wang, J. J., et al. 2020, *AJ*, **159**, 71
- Nowak, M., Lacour, S., Lagrange, A.-M., et al. 2020, *A&A*, **642**, L2
- Nowak, M., Lacour, S., Abuter, R., et al. 2024a, *A&A*, **687**, A248
- Nowak, M., Lacour, S., Abuter, R., et al. 2024b, *A&A*, **684**, A184
- Paszke, A., Gross, S., Massa, F., et al. 2019, arXiv e-prints [arXiv:1912.01703]
- Perrin, G., & Woillez, J. 2019, *A&A*, **625**, A48
- Perrin, M., Long, J., Douglas, E., et al. 2016, *Astrophysics Source Code Library* [record ascl:1602.018]
- Pfuhl, O., Haug, M., Eisenhauer, F., et al. 2012, in *Optical and Infrared Interferometry III*, 8445, eds. F. Delplancke, J. K. Rajagopal, & F. Malbet, International Society for Optics and Photonics (SPIE), 84451U
- Por, E. H., Haffert, S. Y., Radhakrishnan, V. M., et al. 2018, in *Adaptive Optics Systems VI*, 10703, eds. L. M. Close, L. Schreiber, & D. Schmidt, International Society for Optics and Photonics (SPIE), 1070342
- Pourré, N., Le Bouquin, J.-B., Milli, J., et al. 2022, *A&A*, **665**, A158
- Querel, R. R., & Kerber, F. 2014, in *Ground-based and Airborne Instrumentation for Astronomy V*, 9147, eds. S. K. Ramsay, I. S. McLean, & H. Takami, International Society for Optics and Photonics (SPIE), 914792
- Ravet, M., Bonnefoy, M., Chauvin, G., et al. 2025, *A&A*, **704**, A325
- Royer, F., Zorec, J., & Gómez, A. E. 2007, *A&A*, **463**, 671
- Scheithauer, S., Brandner, W., Deen, C., et al. 2016, in *Adaptive Optics Systems V*, 9909, eds. E. Marchetti, L. M. Close, & J.-P. Véran, International Society for Optics and Photonics (SPIE), 99092L
- Tatulli, E., Millour, F., Chelli, A., et al. 2007, *A&A*, **464**, 29
- Van Leeuwen, F., ed. 2007, *Astrophysics and Space Science Library*, 350, Hipparcos, the New Reduction of the Raw Data (Dordrecht: Springer Netherlands)
- Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, *Nat. Methods*, **17**, 261
- Waskom, M. 2021, *JOSS*, **6**, 3021
- Young, A. T. 2004, *AJ*, **127**, 3622

Appendix A: Baseline-dependent systematic throughput variations

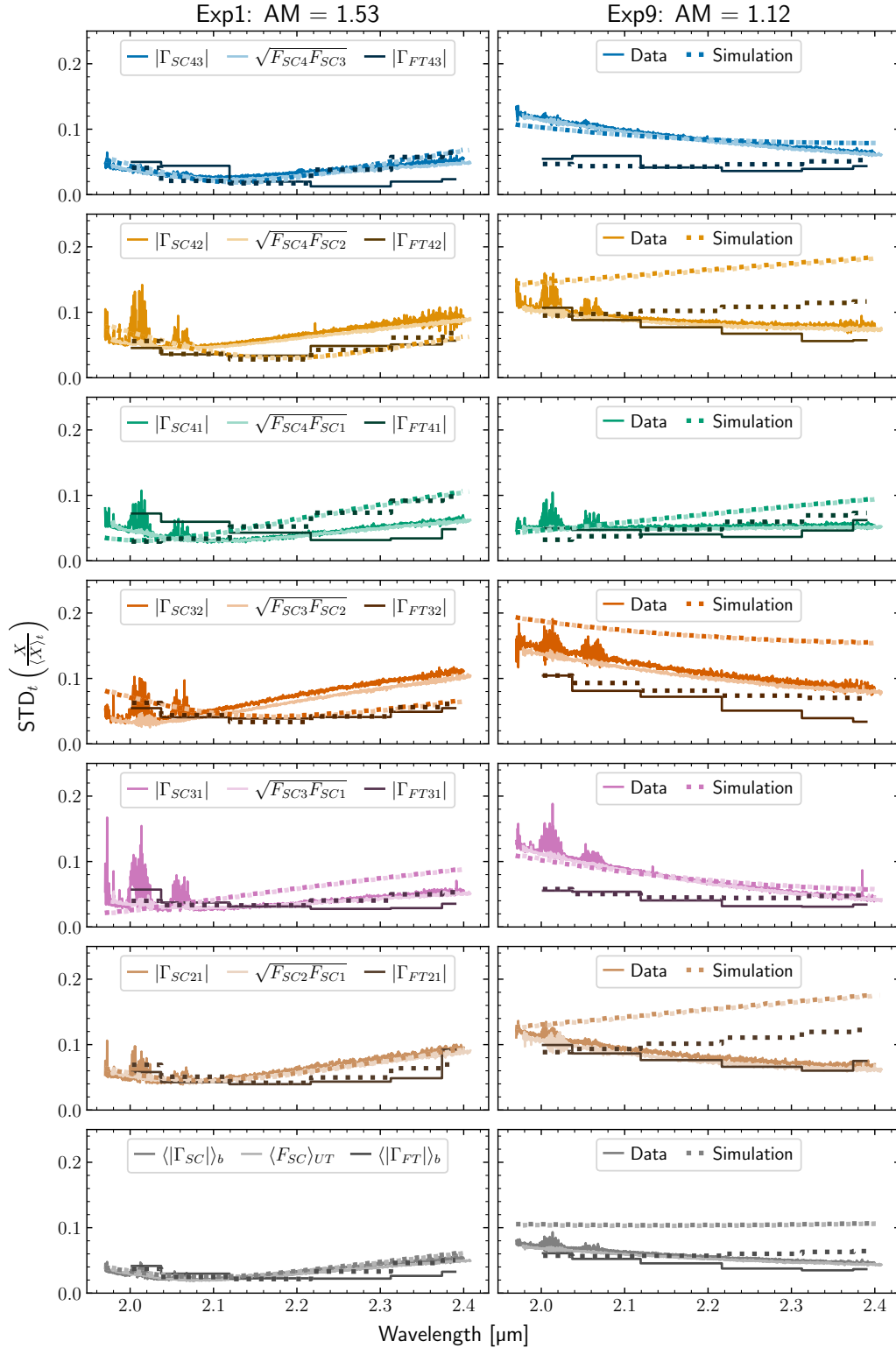


Fig. A.1. *Left:* Standard deviation (STD) of the measured (solid lines) and simulated (dotted lines) relative SC (1628 spectral bins) absolute coherent fluxes $|\Gamma_{SC43}|$, ..., $|\Gamma_{SC21}|$, combined SC fluxes $\sqrt{F_{SC4}F_{SC3}}$, ..., $\sqrt{F_{SC2}F_{SC1}}$, and FT (6 spectral bins) absolute coherent fluxes $|\Gamma_{FT43}|$, ..., $|\Gamma_{FT21}|$ (different shades) for an example exposure at high airmass, shown for the six baselines (43, ..., 21) and their respective baseline averages $\langle \cdot \rangle_b$ (different colors from top to bottom). *Right:* Same as the left panels but for an exposure at low airmass.