



HAL
open science

Increasing the achievable contrast of infrared interferometry with an error correlation model

J. Kammerer, A. Mérand, M. Ireland, S. Lacour

► **To cite this version:**

J. Kammerer, A. Mérand, M. Ireland, S. Lacour. Increasing the achievable contrast of infrared interferometry with an error correlation model. *Astronomy & Astrophysics - A&A*, 2020, 644, pp.A110. 10.1051/0004-6361/202038563 . hal-03252347

HAL Id: hal-03252347

<https://hal.science/hal-03252347v1>

Submitted on 7 Jun 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Increasing the achievable contrast of infrared interferometry with an error correlation model[★]

J. Kammerer^{1,2}, A. Mérand¹, M. J. Ireland², and S. Lacour^{1,3}

¹ European Southern Observatory, Karl-Schwarzschild-Str. 2, 85748 Garching, Germany
e-mail: jens.kammerer@eso.org

² Research School of Astronomy & Astrophysics, Australian National University, Canberra, ACT 2611, Australia

³ LESIA, Observatoire de Paris, Université PSL, CNRS, Sorbonne Université, Université de Paris, 5 place Jules Janssen, 92195 Meudon, France

Received 2 June 2020 / Accepted 13 October 2020

ABSTRACT

Context. Interferometric observables are strongly correlated, yet it is common practice to ignore these correlations in the data analysis process.

Aims. We develop an empirical model for the correlations present in Very Large Telescope Interferometer GRAVITY data and show that properly accounting for them yields fainter detection limits and increases the reliability of potential detections.

Methods. We extracted the correlations of the (squared) visibility amplitudes and the closure phases directly from intermediate products of the GRAVITY data reduction pipeline and fitted our empirical models to them. Then, we performed model fitting and companion injection and recovery tests with both simulated and real GRAVITY data, which are affected by correlated noise, and compared the results when ignoring the correlations and when properly accounting for them with our empirical models.

Results. When accounting for the correlations, the faint source detection limits improve by a factor of up to ~ 2 at angular separations > 20 mas. For commonly used detection criteria based on χ^2 statistics, this mostly results in claimed detections being more reliable.

Conclusions. Ignoring the correlations present in interferometric data is a dangerous assumption which might lead to a large number of false detections. The commonly used detection criteria (e.g. in the model fitting pipeline CANDID) are only reliable when properly accounting for the correlations; furthermore, instrument teams should work on providing full covariance matrices instead of statistically independent error bars as part of the official data reduction pipelines.

Key words. techniques: interferometric – methods: statistical – planets and satellites: detection

1. Introduction

With the first detection and characterisation of an exoplanet by the Very Large Telescope Interferometer (VLTI) instrument GRAVITY (HR 8799 e, [Gravity Collaboration 2019](#)), infrared interferometry has proven to be a powerful technique for high-contrast imaging at high angular resolution. Although initially designed for observations of the galactic centre ([Bartko et al. 2009](#)), GRAVITY's dual-feed mode combined with the recently installed integrated optics beam combiner ([Perraut et al. 2018](#)) enable spectroscopy and micro-arcsecond astrometry of exoplanets with a wide range of angular separations ([Gravity Collaboration 2019](#)).

More recently, [Gravity Collaboration \(2020\)](#) have used GRAVITY observations of β Pic b in order to derive reliable estimates for the mass and the C/O ratio of the young giant planet using forward modelling and free retrieval of its atmosphere. In the future, infrared interferometry will be a promising opportunity for studying giant planet formation (e.g. with Hi-5, [Defrère et al. 2018](#)) and potentially even characterising terrestrial exoplanets from space (e.g. with a formation-flying nulling interferometer, [Léger et al. 1996](#); [Mennesson & Mariotti 1997](#); [Kammerer & Quanz 2018](#); [Quanz et al. 2018, 2019](#)). However, significant improvements are required on the technical side (e.g.

kernel nulling, [Martinache & Ireland 2018](#)), in addition to on the data reduction side in order to achieve these ambitious goals.

Because they use the dual-feed mode of GRAVITY, the aforementioned observations are not conducted anywhere close to the diffraction limit of the interferometer, but rather the diffraction limit of a single telescope. Detecting a companion within the interferometer's diffraction limit (a few λ/b_{\max}), where λ is the observing wavelength and b_{\max} is the longest baseline of the interferometer, is limited by systematic errors. While such systematic errors that are introduced by instrumental and atmospheric effects have been studied intensively (e.g. imperfect fibre coupling, [Kotani et al. 2003](#); instrument vibrations, [Le Bouquin et al. 2011](#); differential atmospheric piston; [Colavita 1999](#)), correlations are also introduced by the data reduction and the calibration. For instance, a systematic error might be introduced similarly to all complex visibilities measured on the science target if the instrumental transfer function obtained from the calibrator target is affected by unknown variability ([Perrin 2003](#)) and if the closure phases measured over telescope triplets of closing triangles are not mathematically independent ([Monnier 2007](#)). Nevertheless, most data reduction pipelines (e.g. the PIONIER data reduction pipeline, [Le Bouquin et al. 2011](#); the GRAVITY data reduction pipeline, [Lapeyriere et al. 2014](#)) and model fitting routines (e.g. LITpro¹,

[★] Based on observations made with ESO telescopes at Paranal Observatory under programme IDs 60.A-9801(U) and 0101.C-0907(B).

¹ <https://www.jmmc.fr/english/tools/data-analysis/litpro>

Tallon-Bosc et al. 2008; CANDID², Gallenne et al. 2015) assume statistically independent observables. However, in order to robustly detect faint companions, or place upper limits on their brightness, a solid understanding and description of the systematic errors is inevitable.

While Lachaume et al. (2019) proposed to use the bootstrapping method (i.e. sampling with replacement, Efron & Tibshirani 1986) in order to obtain the multivariate probability density function of the squared visibility amplitudes and the closure phases, Gravity Collaboration (2020) extracted the covariances of the complex visibilities directly from the data. Although the bootstrapping method is computationally expensive, it enables estimating the systematic errors not only between the different spectral channels, baselines and triangles, but also between different observations. This enables accounting for correlations introduced by sky rotation or the calibration method, but is only applicable at a higher level when the structure of the observing sequence is known.

In this paper, we follow a similar approach to Gravity Collaboration (2020) by extracting the correlations between the squared visibility amplitudes and the closure phases directly from the data. Then, we develop an empirical model for these correlations which can be fitted to the correlations extracted from single GRAVITY pipeline products, even if only a small number of measurements is available. This enables the attainment of a systematic error estimate for every GRAVITY data set and could ultimately be included in the GRAVITY data reduction pipeline (Lapeyre et al. 2014).

2. Methods

In Sect. 2.1 we show how we extract the correlations from individual GRAVITY pipeline products and describe their nature. In Sect. 2.2 we introduce our empirical model for these correlations and in Sect. 2.4 we present the model fitting routines with the aid of which we show the improvements that come from using our empirical correlation model. The Python code that we developed in the scope of this paper is publicly available on GitHub³.

2.1. Correlations extracted from GRAVITY data

In order to extract the correlations between the different spectral channels, baselines and triangles from GRAVITY data we use the P2VM-reduced files from the GRAVITY data reduction pipeline. These files are intermediate pipeline products which contain the individual measurements (detector read-outs) before they are averaged together. Having access to the individual measurements enables extracting the correlations from the (complex) coherent flux $\text{VIS}_{mb\lambda}$ which is stored in the P2VM-reduced file as a data cube of shape $M \times B \times \Lambda$, where $m = 1 \dots M$ is the number of individual measurements, $b = 1 \dots B$ is the number of baselines and $\lambda = 1 \dots \Lambda$ is the number of spectral channels. From the coherent flux, we compute the squared visibility amplitudes

$$\text{VIS2}_{mb\lambda} = \frac{|\text{VIS}_{mb\lambda}|^2}{\text{F1F2}_{mb\lambda}}, \quad (1)$$

where $\text{F1F2}_{mb\lambda}$ is the product of the total fluxes, and the closure phases

$$\text{T3}_{mt\lambda} = \mathbf{K} \cdot \angle \text{VIS}_{mb\lambda}, \quad (2)$$

² <https://ascl.net/1505.030>

³ <https://github.com/kammerje/InterCorr>

where $t = 1 \dots T$ is the number of triangles, \mathbf{K} is a stack of M matrices \mathbf{k} which encode how the four unique triangles can be formed from the six unique baselines of the VLTI, that is

$$\mathbf{k} = \begin{pmatrix} 1 & -1 & 0 & 1 & 0 & 0 \\ 1 & 0 & -1 & 0 & 1 & 0 \\ 0 & 1 & -1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & -1 & 1 \end{pmatrix}, \quad (3)$$

and \angle denotes the argument of a complex number (i.e. the phase).

Then, we compute the sample covariance of the squared visibility amplitudes and the closure phases according to

$$(\Sigma_X)_{ij} = \frac{1}{M-1} \sum_{m=1}^M (X_{mi} - \bar{X}_i)(X_{mj} - \bar{X}_j), \quad (4)$$

where X is VIS2/T3 and i and j run over $1 \dots B\Lambda/1 \dots T\Lambda$ so that we obtain covariance matrices of shape $(B\Lambda) \times (B\Lambda)/(T\Lambda) \times (T\Lambda)$ that contain the covariances between the different spectral channels and baselines/triangles. \bar{X} denotes the mean of X over the individual measurements, that is

$$\bar{X}_i = \frac{1}{M} \sum_{m=1}^M X_{mi}. \quad (5)$$

The correlations between the VIS2 and the T3 then follow by dividing the covariances by the standard deviation $\sigma_i = \sqrt{\Sigma_{ii}}$ of the corresponding observables, that is

$$(C_X)_{ij} = \frac{(\Sigma_X)_{ij}}{(\sigma_X)_i(\sigma_X)_j}. \quad (6)$$

The diagonal of the covariance matrix Σ equals the square of the standard deviation and that the diagonal of the correlation matrix C equals one by definition.

For developing an empirical correlation model we use data taken with GRAVITY (Gravity Collaboration 2017) at the Very Large Telescope Interferometer (VLTI) during technical time (programme 60.A-9801(U)). GRAVITY operates in the K -band (2.0–2.4 μm) and combines the light from either the four Unit Telescopes (UTs) or the four Auxiliary Telescopes (ATs) of the VLTI in order to perform interferometric imaging and astrometry by phase referencing⁴. The data used here was taken with the four UTs on the object HD 82383 (ζ Ant B) in single-field medium resolution ($R = \lambda/\Delta\lambda \approx 500$) mode. This object is relatively bright (K -band magnitude of 6.698, cf. SIMBAD⁵), hence the short integration time of 0.85 ms for the fringe tracker and 1 s for the science camera. It has a companion at an angular separation of ~ 8 arcsec (cf. WDS⁶) which is well beyond the interferometric field-of-view. By choosing a bright target with a short exposure time we make sure that there is a sufficient number of frames to compute the sample covariance (cf. Eq. (4)). The short exposure time of the fringe tracker (much less than the atmospheric coherence time t_0 , which is typically ~ 20 ms in the K -band, Kellerer & Tokovinin 2007) and spatially filtered nature of the GRAVITY beam combiner means that the fringe tracker data is less affected by systematic errors than many other beam combiners. We extract the correlations from a single P2VM-reduced file (GRAVI.2019-03-29T02-01-37.193_singlecalp2vmred.fits) in order to demonstrate the direct

⁴ <https://www.eso.org/sci/facilities/paranal/instruments/gravity.html>

⁵ <http://simbad.u-strasbg.fr/simbad/>

⁶ <http://www.astro.gsu.edu/wds/>

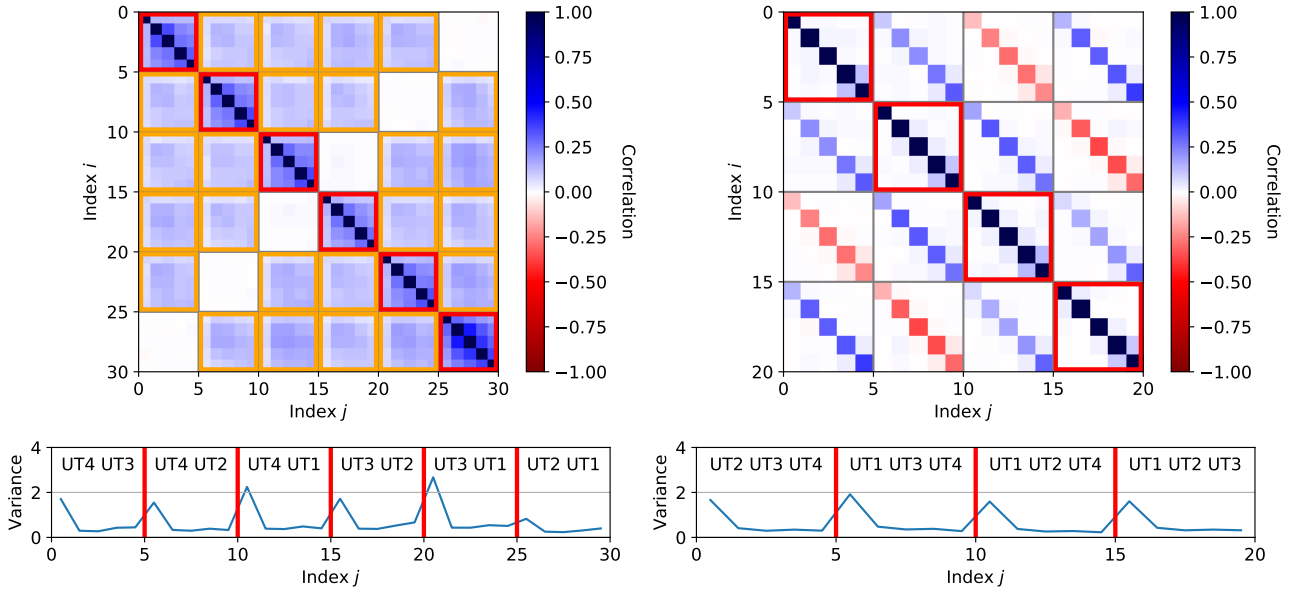


Fig. 1. Correlations of the VIS2 (*left panel*) and the T3 (*right panel*) for the GRAVITY fringe tracker, extracted from a single P2VM-reduced file. The axes run over the different baselines/triangles, with each individual baseline/triangle comprising five spectral channels. Correlations within the same baseline/triangle are highlighted with red squares and correlations between baselines having a telescope in common are highlighted with orange squares. We note that the correlations are computed from 46 592 individual measurements. Below each panel, the variance of the data and the names of the telescopes forming each baseline/triangle are shown.

applicability of our method to the GRAVITY data reduction pipeline. Correlations extracted from other P2VM-reduced files of the same program can be found in Fig. A.1.

Figure 1 shows the correlations of the VIS2 (left panel) and the T3 (right panel) for the GRAVITY fringe tracker. There are six different baselines and four different triangles with five spectral channels each, so 30 observables for the VIS2 and 20 observables for the T3 in total. Correlations within the same baseline/triangle are highlighted with red squares and correlations between baselines having a telescope in common are highlighted with orange squares.

The most dominant correlations of the VIS2 are between different spectral channels within the same baseline, with neighbouring spectral channels being affected most strongly. We suspect that these correlations are predominantly of both atmospheric or instrumental origin, since all five spectral channels follow the same optical path through the atmosphere and up to the dispersive element behind the beam combiner and before the science camera. Also, the five spectral channels do not correspond to individual pixels on the detector of the fringe tracker. In fact, the wavelengths of the five spectral channels lie somewhere between the wavelengths corresponding to the pixels on the detector of the fringe tracker, so that the values recorded by two neighbouring pixels on the detector need to be interpolated in order to find the values for the five spectral channels of the fringe tracker. This could explain the strong correlations between neighbouring spectral channels (one pixel above or below the diagonal) observed for the VIS2, but also for the T3. Furthermore, there are significant correlations between baselines having a telescope in common. Their strength is roughly half the strength of the correlations within the same baseline, which makes sense if the correlations are introduced by atmospheric or instrumental effects and affect each of the four individual beams of the interferometer separately. Also, baselines having no telescope in common are essentially uncorrelated. Hence, we conclude that most of the correlations of the VIS2 are caused by atmospheric or instrumental effects.

For the T3, we observe similar correlations between neighbouring spectral channels as for the VIS2. This makes sense since the closure phases are built from a linear combination (encoded in the matrix \mathbf{k}) of the phase of the complex visibilities, whose absolute square are the squared visibility amplitudes. Moreover, there are significant correlations of $\sim \pm 1/3$ between the same spectral channels on different triangles. These are caused by the fact that each set of two different triangles has exactly one of their three baselines in common, that is each column of the matrix \mathbf{k} has exactly two non-zero entries. If the common baseline is shared between the different triangles in parallel direction (i.e. the two entries in the corresponding column of the matrix \mathbf{k} have the same sign), the correlation is $+1/3$, otherwise it is $-1/3$. This structure with the side-diagonals being $\pm 1/3$ can also be explained by assuming uncorrelated visibility phases (i.e. a diagonal correlation matrix

$$\mathbf{C}_{\Delta\text{VIS}} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 1 \end{pmatrix} \quad (7)$$

of shape $B \times \Lambda$ for the visibility phases) and performing a basis transform

$$\mathbf{C}_{\text{T3}} = \mathbf{T} \cdot \mathbf{C}_{\Delta\text{VIS}} \cdot \mathbf{T}^T, \quad (8)$$

where \mathbf{T} represents a matrix of shape $(T\Lambda) \times (BA)$ which maps the vector of visibility phases to the vector of closure phases and can be trivially obtained from the matrix \mathbf{k} . Also, the observed correlations between neighbouring spectral channels on different triangles (pixels next to the side-diagonals) are naturally explained by this basis transform given the correlations of the VIS2 observed between neighbouring spectral channels on the same baseline.

Figure 2 shows the correlations of the VIS2 (left panel) and the T3 (right panel) for the GRAVITY science camera. There

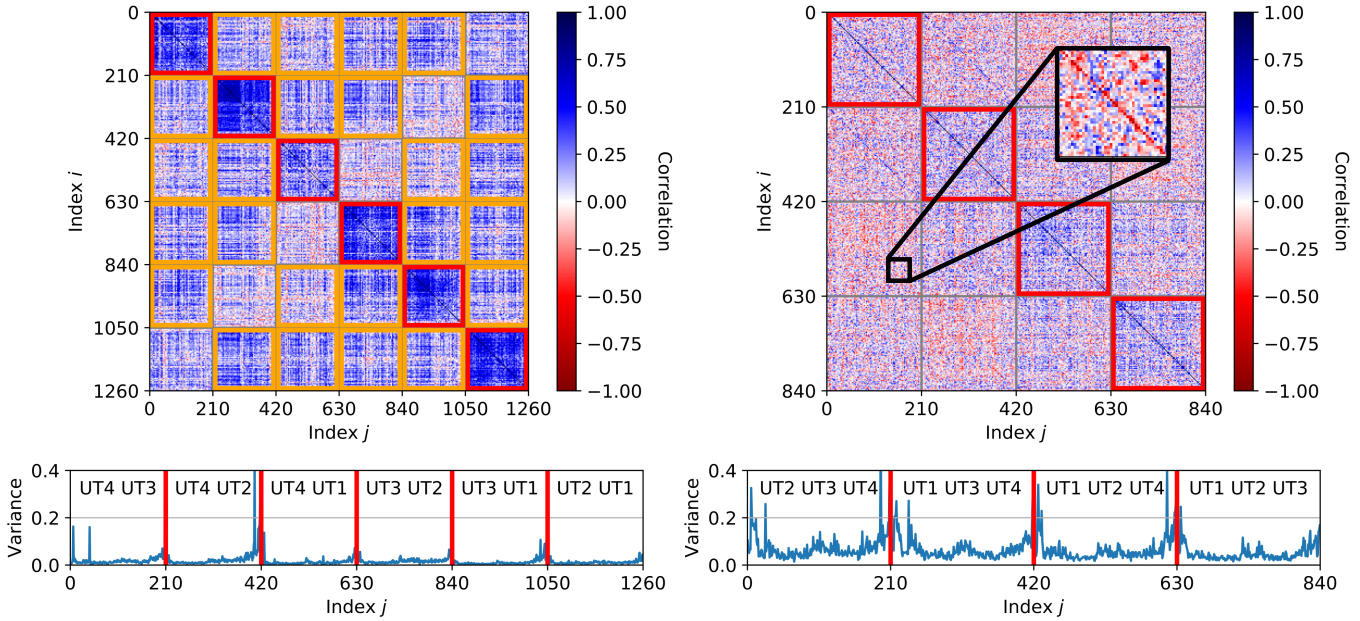


Fig. 2. Same as Fig. 1, but for the GRAVITY science camera. Each individual baseline/triangle comprises 210 spectral channels. We note that the correlations are computed from 32 individual measurements.

are six different baselines and four different triangles with 210 spectral channels each, so 1260 observables for the VIS2 and 840 observables for the T3 in total.

Due to the much smaller number of individual measurements if compared to the fringe tracker the correlations of the science camera are more dominated by noise. Nevertheless, we observe strong positive correlations between different spectral channels within the same baseline (i.e. inside the red squares) and significant positive correlations between baselines having a telescope in common (i.e. inside the orange squares) for the VIS2, similar to the correlations observed for the fringe tracker. Although the atmospheric turbulence and the optical elements (i.e. mirrors, delay lines, optical fibres, beam combiner) seen by the science camera are similar to those seen by the fringe tracker, the exposure time of the science camera is much longer than both the atmospheric coherence time t_0 and the fringe tracker inverse 3 dB bandwidth (Lacour et al. 2019), which means that the VIS2 correlations are expected to be decreased by a term proportional to the square of the fringe tracking error and the closure phase random errors are expected to be proportional to the cube of the fringe tracking error (Ireland 2013). Therefore, since there still are significant correlations for the science camera, they must be introduced by the (correlated) fringe tracker, forwarding the correlations shown in Fig. 1 to the science camera. For the T3, we again observe significant correlations of $\sim \pm 1/3$ between the same spectral channels on different triangles. On top of this, there are also weak positive correlations between different spectral channels on the same triangle (i.e. inside the red squares) and between different spectral channels on different triangles whose sign depends on whether the corresponding triangles share a baseline in parallel or anti-parallel direction. Again, these correlations are naturally explained by the basis transform T given the correlations observed for the VIS2 of the science camera.

2.2. Empirical model for the correlations

An empirical VIS2/T3 sample covariance with fewer frames than the product of the number of baselines/triangles and spectral

channels is necessarily singular. It takes a number of frames much greater than this to estimate a sample covariance matrix with a condition number approaching that of the true sample covariance. For this reason, we choose to develop an analytic model for the covariance matrix Σ of the VIS2 and the T3. This model can then be fitted to the (potentially under-conditioned) covariance matrix extracted from an arbitrary GRAVITY data set and can be used for model fitting based on log-likelihood maximisation. Most model fitting routines (e.g. LITpro, Tallon-Bosc et al. 2008; CANDID, Gallenne et al. 2015) are based on χ^2 minimisation, which is equivalent to log-likelihood maximisation, where

$$\chi^2 = R^T \cdot \Sigma^{-1} \cdot R \quad (9)$$

and $R = D - M$ is the residual between data and model (cf. Sect. 2.4).

Our approach is to model the correlation matrices C_{VIS2} of the VIS2 and C_{T3} of the T3 which have the relatively simple structure observed in Figs. 1 and 2. Moreover, the observed structure of the correlations is consistent for different data sets with different exposure times (1 s with the UTs for programme 60.A-9801(U) and 10 s with the ATs for programme 0101.C-0907(B), cf. also Figs. A.1 and A.2). Then, we compute

$$\Sigma_{ij} = C_{ij} \sigma_i \sigma_j, \quad (10)$$

where σ denotes the standard deviation of the data which can be obtained from the VIS2ERR and the T3PHIERR columns of the OIFITS files for example. We note that these standard deviations are used to build diagonal covariance matrices in LITpro and CANDID which assume uncorrelated data only. Of course, assuming uncorrelated data is a simplification and we discuss the problems that arise from this in Sect. 3.2.

A very important point is that Eq. (10) only holds if the errors on the VIS2 (σ_{VIS2}) and the T3 (σ_{T3}) are reliably estimated by the GRAVITY data reduction pipeline. The pipeline manual⁷ explains that the uncertainties are computed by bootstrapping

⁷ http://www.eso.org/sci/software/pipelines/index.html#pipelines_table

over ~ 10 independent samples, so that the final error on the mean measurement is estimated from the observed statistics at a slightly higher temporal frequency. There is no re-scaling or accounting for systematics in this process. In case there are less than five frames available, Monte-Carlo realisations of the theoretical photon and detector noise are added to the samples, which leads to less realistic uncertainties. However, our data sets consist of 32 frames exposures for programme 60.A-9801(U) and 20 frames exposures for programme 0101.C-0907(B), respectively. While we understand that the use of the pipeline uncertainties is a limitation and that an incorrect noise model can reduce the detection sensitivity or yield false positives (cf. e.g. Sect. 3 of [Delisle et al. 2020](#)), we also note that investigating and quantifying the credibility of these uncertainties is beyond the scope of this work.

Our models for the correlation matrices equal one on the diagonal according to the definition of a correlation matrix (cf. Eq. (6), that is every observable is 100% correlated with itself) and have one free parameter which can be determined by fitting the model to the correlations extracted from the P2VM-reduced files. For the correlation matrix of the VIS2 C_{VIS2} , the free parameter x represents the correlations between spectral channels within the same baseline and between baselines having a telescope in common. There are correlations of x between different spectral channels within the same baseline, correlations of $x/2$ between baselines having a telescope in common, and no correlations between baselines having no telescope in common (cf. left panel of Fig. 2), that is

$$C_{\text{VIS2}} = \begin{pmatrix} X_1 & X_2 & \cdots & \cdots & \cdots & X_2 & 0 \\ X_2 & \ddots & \ddots & & \ddots & \ddots & X_2 \\ \vdots & \ddots & \ddots & X_2 & 0 & \ddots & \vdots \\ \vdots & & X_2 & \ddots & X_2 & & \vdots \\ \vdots & \ddots & 0 & X_2 & \ddots & \ddots & \vdots \\ X_2 & \ddots & \ddots & & \ddots & \ddots & X_2 \\ 0 & X_2 & \cdots & \cdots & \cdots & X_2 & X_1 \end{pmatrix}, \quad (11)$$

$$X_1 = \begin{pmatrix} 1 & x & \cdots & x \\ x & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & x \\ x & \cdots & x & 1 \end{pmatrix}, \quad (12)$$

$$X_2 = \begin{pmatrix} x/2 & \cdots & x/2 \\ \vdots & \ddots & \vdots \\ x/2 & \cdots & x/2 \end{pmatrix}. \quad (13)$$

The correlation matrix is a block matrix consisting of $B \times B$ blocks, where each individual block is a $\Lambda \times \Lambda$ matrix. For the correlation matrix of the T3 C_{T3} , the free parameter y represents the correlations between spectral channels within the same triangle. Moreover, as illustrated by the basis transform T , this naturally leads to correlations of $\pm 1/3$ between the same spectral channel of different triangles and $\pm y/3$ between different spectral channels of different triangles (cf. right panel of Fig. 2), that is

$$C_{\text{T3}} = \begin{pmatrix} Y_1 & Y_2 & \cdots & Y_2 \\ Y_2 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & Y_2 \\ Y_2 & \cdots & Y_2 & Y_1 \end{pmatrix}, \quad (14)$$

$$Y_1 = \begin{pmatrix} 1 & y & \cdots & y \\ y & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & y \\ y & \cdots & y & 1 \end{pmatrix}, \quad (15)$$

$$Y_2 = \begin{pmatrix} \pm 1/3 & \pm y/3 & \cdots & \pm y/3 \\ \pm y/3 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \pm y/3 \\ \pm y/3 & \cdots & \pm y/3 & \pm 1/3 \end{pmatrix}. \quad (16)$$

The correlation matrix is a block matrix consisting of $T \times T$ blocks, where each individual block is a $\Lambda \times \Lambda$ matrix. The sign is positive if the two triangles share a baseline in parallel direction and negative if they share a baseline in anti-parallel direction.

We fit the previously described model to the correlations of the VIS2 and the T3 which we extracted from the single P2VM-reduced file of GRAVITY introduced in Sect. 2.1. Figure 3 shows the extracted and the model correlations (top panels) and the extracted and the model covariances (bottom panels) for the VIS2. The free parameter x takes a value of $\sim 3.2e-1$. Figure 4 shows the same for the T3 and the free parameter y takes a value of $\sim 7.4e-2$.

2.3. Simulated and real data

In order to demonstrate the improvement that comes from taking into account the correlations between the data we perform companion injection and recovery tests with simulated and real data. Therefore, we use GRAVITY data of ζ Ant B from the technical time programme 60.A-9801(U) and of HIP 78183 from the normal programme 0101.C-0907(B), PI M. J. Ireland, listed in Table 1. Both objects were observed in single-field medium resolution mode, but the former one with the four UTs and the latter one with the four ATs (medium configuration D0-G2-J3-K0).

From the file GRAVI.2019-03-29T02-01-37.193_single-calp2vmred.fits we have already extracted the covariances and correlations and fitted our empirical models to them (cf. Sect. 2.2). For the companion injection and recovery tests with simulated data, we simply use these models and the uv -tracks u and v of the files belonging to programme 60.A-9801(U) listed in Table 1 in order to obtain a realistic uv -coverage over ~ 20 min (cf. Fig. 5). We simulate the complex visibility of a uniform disc with an unresolved companion according to

$$\text{VIS}_{\text{bin}} = \frac{\text{VIS}_{\text{ud}} + f \exp\left(-2\pi i \left(\frac{\Delta_{\text{RA}} u}{\lambda} + \frac{\Delta_{\text{Dec}} v}{\lambda}\right)\right)}{1 + f}, \quad (17)$$

$$\text{VIS}_{\text{ud}} = \frac{2J_1(\pi\theta b)}{\pi\theta b}, \quad (18)$$

where $0 \leq f \leq 1$ is the relative flux of the companion, Δ_{RA} and Δ_{Dec} are the on-sky separation in the direction of the celestial north and the celestial east between the companion and its host star, λ is the observing wavelength, J_1 is the first order Bessel function of first kind, θ is the angular diameter of the uniform disc and $b = \sqrt{u^2 + v^2}$ is the distance between the two telescopes observing the object. The squared visibility amplitudes and the closure phases follow according to

$$\text{VIS}_{2\text{bin}} = |\text{VIS}_{\text{bin}}|^2, \quad (19)$$

$$\text{T3}_{\text{bin}} = k \cdot \angle \text{VIS}_{\text{bin}}. \quad (20)$$

Then, we add correlated noise to the $\text{VIS}_{2\text{bin}}$ and the T3_{bin} by drawing from a multivariate normal distribution with covariance

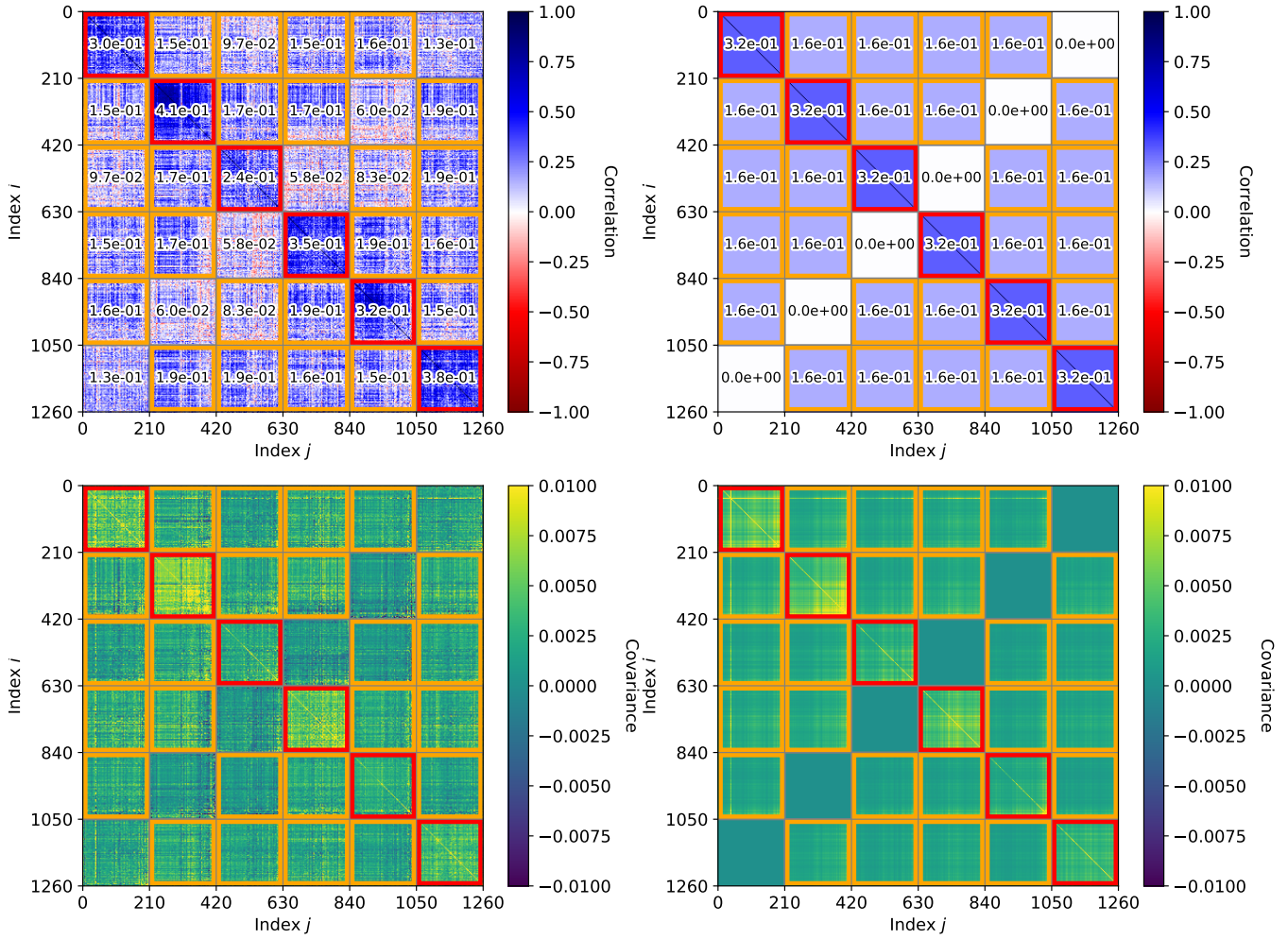


Fig. 3. Correlations of the VIS2 for the GRAVITY science camera, extracted from a single P2VM-reduced file (*upper left panel*) and our one-parameter model fitted to them (*upper right panel*). The *bottom panels* show the corresponding covariances obtained by multiplying the correlation C_{ij} with the product of the standard deviations $\sigma_i \sigma_j$. Correlations/covariances within the same baseline are highlighted with red squares and correlations/covariances between baselines having a telescope in common are highlighted with orange squares.

$\Sigma_{\text{VIS2,fit}}$ and $\Sigma_{\text{T3,fit}}$, which we obtain from our correlation model $C_{\text{VIS2,fit}}$ and $C_{\text{T3,fit}}$ (cf. Sect. 2.2) and assuming a standard deviation of $\sigma_{\text{VIS2}} = 0.01$ and $\sigma_{\text{T3}} = 1$ deg.

For the companion injection and recovery tests with real data, we extract the correlations of the visibility amplitudes VISAMP (instead of the squared visibility amplitudes VIS2) and the closure phases T3 from the P2VM-reduced files belonging to programme 0101.C-0907(B) listed in Table 1, fit our empirical models to them and compute the covariances using Eq. (10) and the errors from the corresponding final GRAVITY pipeline products (the “singlecviscalibrated” files). Using the VISAMP instead of the VIS2 can yield better results in some cases where the normalisation of the VIS2 is not done properly by the GRAVITY data reduction pipeline. From the final GRAVITY pipeline products, we also extract the VISAMP and the T3 and inject an unresolved companion according to

$$\text{VISAMP}_{\text{inj}} = \text{VISAMP} \cdot |\text{VIS}_{\text{bin}}|, \quad (21)$$

$$\text{T3}_{\text{inj}} = \text{T3} + k \cdot \angle \text{VIS}_{\text{bin}}, \quad (22)$$

where we set VIS_{ud} to one. The VISAMP are simply the square root of the VIS2, so that our correlation model and fitting routines can be equivalently applied in the high signal-to-noise regime.

2.4. Model fitting

We search for faint companions in the data by fitting the model for a uniform disc with an unresolved companion (cf. Eq. (17)) to it. We maximise the log-likelihood of the model by minimising its χ^2 in order to find the best fit parameters p_{fit} of the model, that is

$$p_{\text{fit}} = \text{argmin}_p(\chi^2) = \text{argmin}_p(R^T \cdot \Sigma^{-1} \cdot R), \quad (23)$$

where $R = D - M$ is the residual between data and model and $p = (f, \Delta_{\text{RA}}, \Delta_{\text{Dec}}, \theta)$ is the four-dimensional parameter vector of the model.

In order to find the global minimum of the χ^2 within a given range of companion separations, we first find a prior for the uniform disc diameter θ_0 by fitting the corresponding model (cf. Eq. (18)) to the data. Then, we perform a set of minimisations with priors on a $\Delta_{\text{RA}} - \Delta_{\text{Dec}}$ grid, the uniform disc diameter θ_0 , and a small relative flux $f_0 = 1e-3$. This is necessary since the χ^2 hyper-surface is bumpy (i.e. has many local extrema) if projected onto the $\Delta_{\text{RA}} - \Delta_{\text{Dec}}$ surface and the BFGS algorithm which is used to minimise the χ^2 converges on local minima. The bumpiness is a result of the sparse uv -coverage of a long-baseline optical interferometer which causes the sensitivity to vary substantially over the FOV.

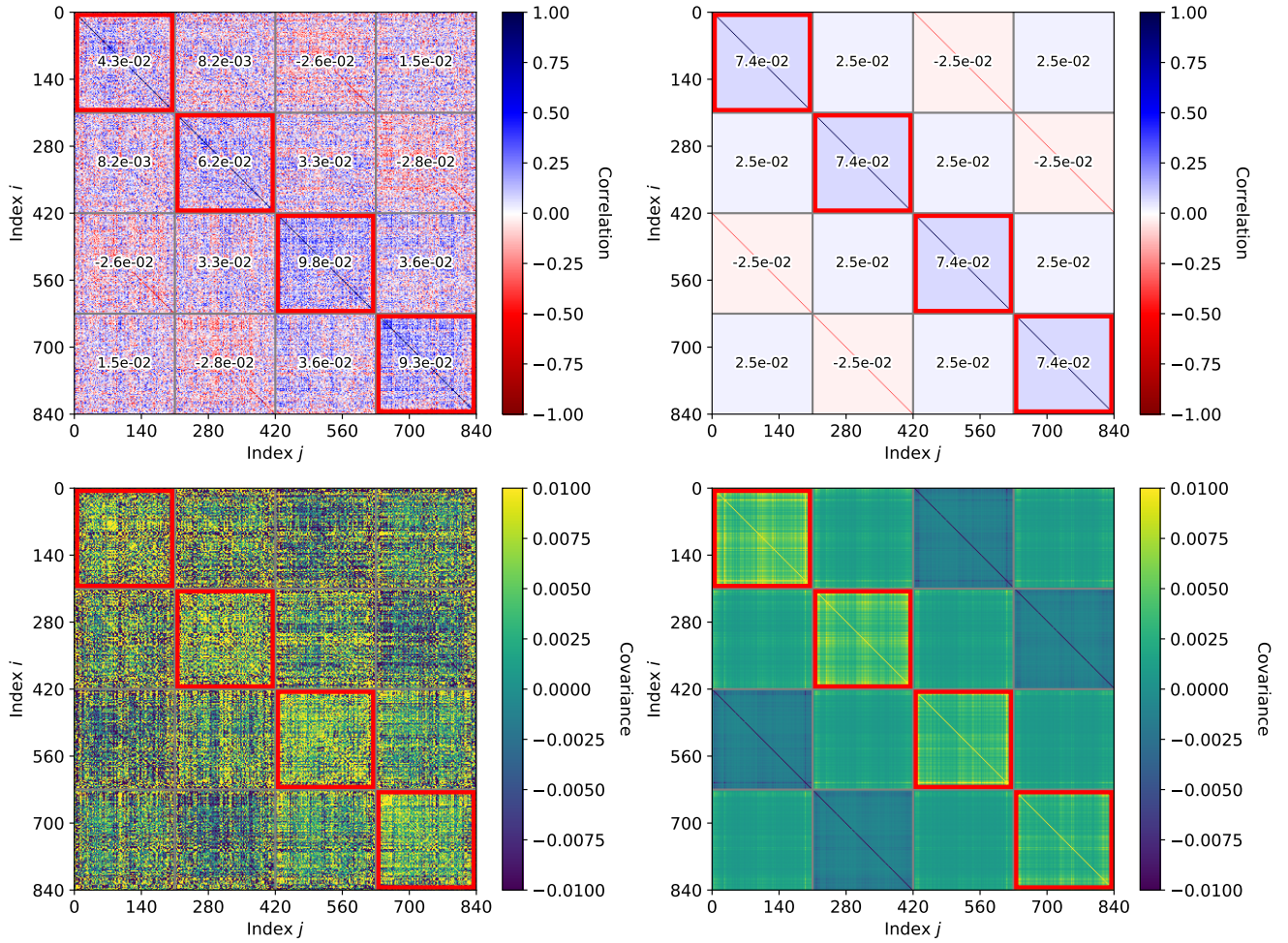


Fig. 4. Same as Fig. 3, but showing the correlations/covariances of the T3 and our one-parameter model fitted to them for the GRAVITY science camera.

Table 1. GRAVITY data used for the companion injection and recovery tests.

Programme	Filename	OB date	OB time (UT)
60.A-9801(U)	GRAVI.2019-03-29T01-42-55.145_singlecalvis.fits	2019-03-29	01:42:53
	GRAVI.2019-03-29T01-51-13.167_singlecalvis.fits	2019-03-29	01:51:09
	GRAVI.2019-03-29T02-01-37.193_singlecalvis.fits	2019-03-29	02:01:35
0101.C-0907(B)	GRAVI.2018-04-18T08-08-19.739_singlescip2vmred.fits	2018-04-18	08:08:16
	GRAVI.2018-04-18T08-12-10.749_singlescip2vmred.fits	2018-04-18	08:12:08
	GRAVI.2018-04-18T08-20-04.769_singlescip2vmred.fits	2018-04-18	08:20:02

The above method relies on the covariance matrix Σ being invertible. This is not the case for a sample covariance that is estimated from a small number of frames, which is usually singular, and is the reason why we develop an empirical covariance model. However, our empirical model for the covariances of the closure phases is also singular, since the fourth triangle can be written as a linear combination of the other three. There are multiple solutions to this problem, and for simplicity we decide to completely ignore the data recorded on the fourth triangle since it theoretically is redundant anyway⁸. There are more sophisticated methods to keep the data recorded on the

fourth triangle, such as the “jackknife” method (i.e. averaging over four model fits using data recorded on different sets of three triangles), projection into a sub-space that preserves the information in the covariance matrix (Blackburn et al. 2020), and the approach from Kulkarni (1989) which is adding a small numerical value $\epsilon \ll 1$ to the diagonal of the covariances of the closure phases, that is

$$\Sigma_{\text{T3,invertible}} = \Sigma_{\text{T3,fit}} + \epsilon \cdot \text{id}, \quad (24)$$

where id is the identity matrix, so that the covariance matrix becomes numerically invertible.

Finally, in order to determine the statistical significance of a detected companion, we compute the probability that the binary

⁸ In practice, this is not the case since the data is affected by different errors originating from different optical paths through the instrument and different detector noise.

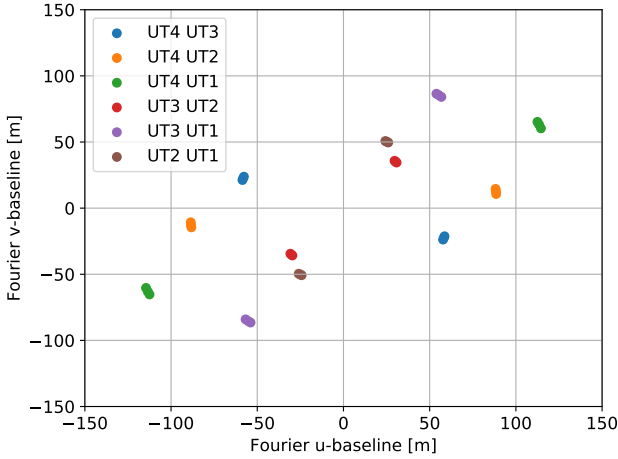


Fig. 5. Fourier u - and v -baselines of our simulated data extracted from three on-sky observations of GRAVITY using the four UTs over ~ 20 min.

model is preferred over the uniform disc model according to

$$P = 1 - \text{CDF}_{N_{\text{d.o.f}}} \left(\frac{N_{\text{d.o.f}} \chi_{\text{red,ud}}^2}{\chi_{\text{red,bin}}^2} \right), \quad (25)$$

where $\text{CDF}_{N_{\text{d.o.f}}}$ is the χ^2 cumulative distribution function with $N_{\text{d.o.f}}$ degrees of freedom, $\chi_{\text{red,ud}}^2$ is the reduced χ^2 of the best fit uniform disc model and $\chi_{\text{red,bin}}^2$ is the reduced χ^2 of the best fit binary model (cf. Gaienne et al. 2015).

If the host star is essentially unresolved (i.e. $\theta b \lambda \ll 1$) and the companion is at high contrast (i.e. $f \ll 1$) one can linearise the VIS2_{bin} and the T3_{bin} as a function of the relative flux of the companion f according to

$$\text{VIS2}_{\text{bin}} \propto 1 + f, \quad (26)$$

$$\text{T3}_{\text{bin}} \propto f. \quad (27)$$

A more detailed derivation of this relationship can be found in Appendix B. Let D be the data, M_{ref} a reference binary model which is normalised to the relative flux of the companion f_{ref} , and Σ the covariances between the data, that is

$$D = \begin{pmatrix} \text{VIS2} - 1 \\ \text{T3} \end{pmatrix}, \quad (28)$$

$$M_{\text{ref}} = \begin{pmatrix} (\text{VIS2}_{\text{bin,ref}} - 1) / f_{\text{ref}} \\ \text{T3}_{\text{bin,ref}} / f_{\text{ref}} \end{pmatrix}, \quad (29)$$

$$\Sigma = \begin{pmatrix} \Sigma_{\text{VIS2}} & \mathbf{0} \\ \mathbf{0} & \Sigma_{\text{T3}} \end{pmatrix}, \quad (30)$$

where $\text{VIS2}_{\text{bin,ref}}$ and $\text{T3}_{\text{bin,ref}}$ are the binary model VIS2 and T3 evaluated at a reference relative flux $f_{\text{ref}} = 1e-3$. Then, the best fit relative flux f_{fit} and its uncertainty $\sigma_{f_{\text{fit}}}$ follow according to

$$f_{\text{fit}} = \frac{M_{\text{ref}}^T \cdot \Sigma^{-1} \cdot D}{M_{\text{ref}}^T \cdot \Sigma^{-1} \cdot M_{\text{ref}}}, \quad (31)$$

$$\sigma_{f_{\text{fit}}} = \frac{1}{\sqrt{M_{\text{ref}}^T \cdot \Sigma^{-1} \cdot M_{\text{ref}}}}, \quad (32)$$

(cf. Le Bouquin & Absil 2012; Kammerer et al. 2019). Equation (31) can be computed on a $\Delta_{\text{RA}} - \Delta_{\text{Dec}}$ grid, and the best fit parameters p_{fit} follow from the grid position which minimises

$$\chi_{\text{red}}^2 = \frac{\chi^2}{N_{\text{d.o.f}}} = \frac{R^T \cdot \Sigma^{-1} \cdot R}{N_{\text{d.o.f}}}, \quad (33)$$

where $N_{\text{d.o.f}}$ is the number of the degrees of freedom. This grid search technique is commonly used in order to find the global minimum of the $(f, \Delta_{\text{RA}}, \Delta_{\text{Dec}}, \theta)$ parameter space and its corresponding χ_{red}^2 (e.g. Absil et al. 2011; Gaienne et al. 2015). However, the statistical structure of the grid is complex due to redundancy and periodicity in sensitivity originating from the very limited uv -coverage of a sparse interferometer such as the VLTI (cf. Fig. 5). Therefore, the detection significance is derived using Eq. (25) which yields the probability that the binary model is preferred over the uniform disc model (without any companion).

3. Results

We evaluate the impact of our full covariance model by performing model fitting and companion injection and recovery tests with both simulated and real GRAVITY data. In Sect. 3.1, we simulate data without any astronomical object, that is correlated noise only, and use model fitting to determine the fundamental detection limits when assuming uncorrelated data (i.e. a diagonal covariance) and correlated data (i.e. our full covariance model). In Sects. 3.2 and 3.3 we inject companions with different relative fluxes and separations into simulated and real GRAVITY data and try to recover them, again assuming both uncorrelated and correlated data.

3.1. Model fitting to correlated noise

In order to compare the fundamental detection limits when assuming uncorrelated and correlated data, we simulate 100 GRAVITY data sets of an unresolved host star without any companion (i.e. $\theta = f = 0$ in Eq. (17)) affected by realistic correlated noise ($\Sigma_{\text{VIS2,fit}}$ and $\Sigma_{\text{T3,fit}}$, cf. Sect. 2.3). Then, we use Eq. (31) in order to compute the best fit relative flux f_{fit} on a $\Delta_{\text{RA}} - \Delta_{\text{Dec}}$ grid for each of the 100 simulated data sets, first assuming uncorrelated data (i.e. a diagonal covariance $\text{diag}(\Sigma_{\text{VIS2,fit}})$ and $\text{diag}(\Sigma_{\text{T3,fit}})$) and then assuming correlated data (i.e. our full covariance model) in Eq. (31). Since no companion was injected into the data, the best fit flux ratios of these grids represent the fundamental contrast floor. Any companion with a higher contrast (i.e. smaller flux) would not be distinguishable from the noise. By computing an azimuthal average of these grids, we obtain a fundamental $1-\sigma$ contrast curve (i.e. best fit relative flux vs. angular separation curve).

The mean of the 100 azimuthal averages obtained for each of these two scenarios (model fitting assuming uncorrelated data in blue and correlated data in orange) is shown in the left panel of Fig. 6. The contrast floor remains roughly constant at a contrast of $\sim 6e-4$ outward of an angular separation of ~ 5 mas for the scenario assuming uncorrelated data. This is because at a contrast of $\sim 6e-4$ one is dominated by the systematic (i.e. the correlated) errors. However, for the scenario assuming correlated data, the fundamental $1-\sigma$ detection limit continues to decrease with increasing angular separation. At an angular separation of ~ 10 mas it is already a factor of four better than the limit assuming uncorrelated data (cf. dashed black curve). We note that such a behaviour has already been observed by Ireland (2013) for orthogonal kernel phases and statistically independent kernel phases, which are obtained by projecting the orthogonal kernel phases into an eigenspace with zero covariances. Its reason is that at small angular separations, the detection limits rely on the average of the VIS2 and the T3 over the spectral channels, while with increasing angular separations the VIS2 and T3 vary within the spectral bands and the impact of the correlations is

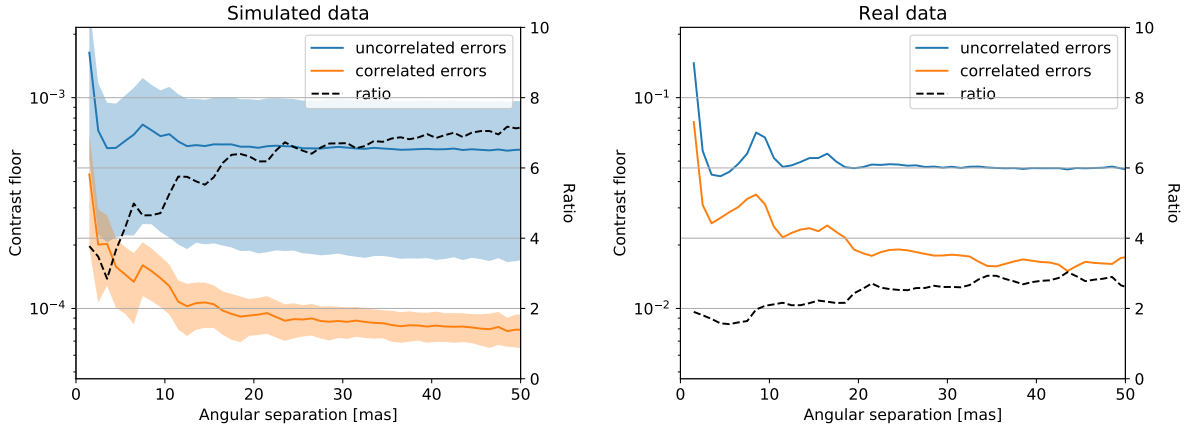


Fig. 6. *Left panel:* contrast curve (i.e. azimuthal average of the best fit relative flux) for simulated data of an unresolved host star without any companion affected by correlated errors, computed with model fitting assuming uncorrelated data (blue curve) and correlated data (orange curve). Both curves show the mean contrast curve over 100 simulated data sets and the shaded region highlights its standard deviation. The dashed black line shows the ratio of the blue and the orange curve, representing the improvement (i.e. the factor by which the detection limits improve) when using our correlated error model instead of the classical uncorrelated one. *Right panel:* same, but for the real GRAVITY data introduced in Sect. 2.3.

growing. A flat uncorrelated contrast curve (as a function of angular separation) is further consistent with previous works on interferometric observables assuming uncorrelated data (e.g. Absil et al. 2011; Gallenne et al. 2015).

Furthermore, when using our full covariance model, the contrast floor is also more stable for different representations of the noise (highlighted by the shaded regions in the left panel of Fig. 6 which show the standard deviation of the contrast curves over the 100 simulated data sets) meaning that the derived detection limits can be regarded more robust (i.e. independent of the exact representation of the noise which is a random component). Hence, if an observer is only working with a small number of data sets, they will still be able to derive universally valid detection limits.

The right panel of Fig. 6 shows the same plot, but for model fitting to the real GRAVITY data consisting of the three files belonging to programme 0101.C-0907(B) listed in Table 1. Since there is only one real GRAVITY data set we cannot compute or show any standard deviation. The plot looks similar, except for the fundamental detection limits being about two orders of magnitude worse and the ratio between the two scenarios being a lot smaller due to much weaker correlations being present in the real data if compared to the simulated data (cf. Fig. A.1).

3.2. Injection and recovery tests (simulated data)

As a next step, we perform companion injection and recovery tests with simulated data, in order to compare the empirical detection limits when assuming uncorrelated and correlated data. Therefore, we simulate GRAVITY data sets (affected by correlated noise) of a 1 mas uniform disc (the host star) and inject companions with a range of relative fluxes and at different positions around the host star, that is

$$f \in [10^{-4}, 10^{-3.75}, 10^{-3.5}, \dots, 10^{-1.5}], \quad (34)$$

$$\Delta_{\text{RA}} \in [-30, -25, -20, \dots, 30] \text{ mas}, \quad (35)$$

$$\Delta_{\text{Dec}} \in [-30, -25, -20, \dots, 30] \text{ mas}, \quad (36)$$

using Eq. (17). Then, we perform model fitting with priors on a $\Delta_{\text{RA}} - \Delta_{\text{Dec}}$ grid in order to find the global minimum of the χ^2 (cf. Eq. (23)). We note that this method is similar to how CANDID searches for companions for example. Similar to before

(cf. Sect. 3.1), we perform the model fitting once assuming uncorrelated data (i.e. a diagonal covariance $\text{diag}(\Sigma_{\text{VIS2,fit}})$ and $\text{diag}(\Sigma_{\text{T3,fit}})$) and once assuming correlated data (i.e. our full covariance model). We classify an injected companion as recovered if the best fit relative flux f_{fit} differs by no more than 10% from the injected one f_{inj} and the best fit position $(\Delta_{\text{RA,fit}}, \Delta_{\text{Dec,fit}})$ differs by no more than one resolution element of the interferometer from the injected one $(\Delta_{\text{RA,inj}}, \Delta_{\text{Dec,inj}})$, that is

$$|f_{\text{fit}} - f_{\text{inj}}|/f_{\text{inj}} < 0.1, \quad (37)$$

$$\sqrt{(\Delta_{\text{RA,fit}} - \Delta_{\text{RA,inj}})^2 + (\Delta_{\text{Dec,fit}} - \Delta_{\text{Dec,inj}})^2} < \frac{\lambda_{\text{mean}}}{2b_{\text{max}}}, \quad (38)$$

where λ_{mean} is the mean of the observed wavelength range ($\sim 2.2 \mu\text{m}$ for GRAVITY) and b_{max} is the longest baseline of the interferometer ($\sim 130 \text{ m}$ for observations with the VLTI UTs).

The left panel of Fig. 7 shows the fraction of recovered companions as a function of the relative flux of the companion for model fitting assuming uncorrelated data (blue points) and correlated data (orange points). These values are summed over all positions around the host star with $5 \text{ mas} \leq \rho \leq 45 \text{ mas}$, where $\rho = \sqrt{\Delta_{\text{RA,inj}}^2 + \Delta_{\text{Dec,inj}}^2}$ is the angular separation, so to avoid any significant influence from the 1 mas uniform disc (the host star). Although one of the findings in this paper is that the contrast curve is not flat outward a few λ/b_{max} when accounting for the data correlations (cf. Fig. 6), it is still a reasonable simplification to sum over positions with different angular separations.

The overplotted blue and orange curves are logistic growth functions

$$g(x) = \frac{L}{1 + e^{-k(x-x_0)}}, \quad (39)$$

fitted to the data points (in log-space), where $L = 1$ is the upper growth barrier, k is the logistic growth rate and x_0 is the mid-point. The fact that the orange curve is shifted towards the left compared to the blue curve means that the detection limits are fainter when assuming correlated data instead of uncorrelated data. This could be expected since the data is affected by correlated noise and correctly accounting for these correlations in the model fitting should lead to fainter detection limits. For any given fraction of detections (i.e. any point on the y-axis), the ratio

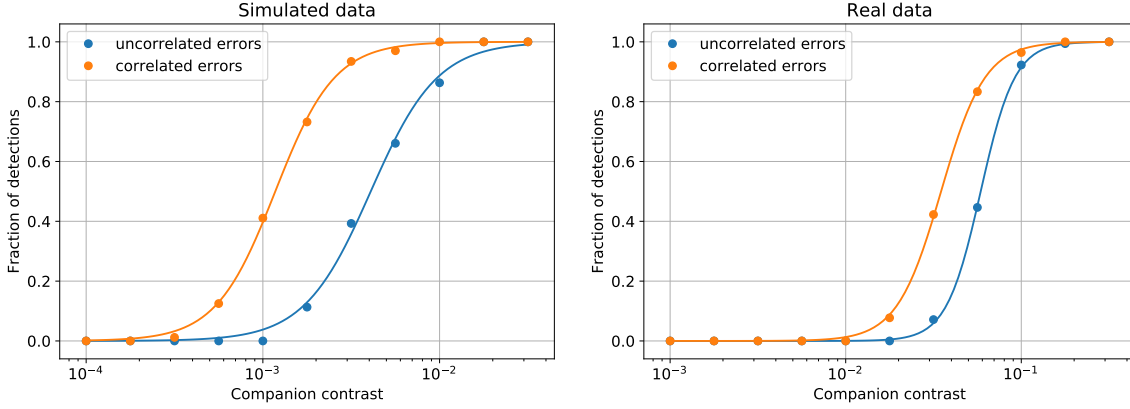


Fig. 7. *Left panel:* fraction of correctly recovered companions as a function of the relative flux of the injected companion from injection and recovery tests with simulated GRAVITY data (cf. Sect. 3.2), assuming uncorrelated data (blue points) and correlated data (orange points) for the model fitting. The blue and the orange curves are logistic growth functions fitted to the data points. *Right panel:* same, but from injection and recovery tests with real GRAVITY data (cf. Sect. 3.3).

of the contrasts (i.e. the x -values) of the blue and the orange curve gives the improvement that comes from our full covariance model over the conventional diagonal covariance model. This ratio varies between ~ 3 and ~ 4 , depending on the fraction of detections, with an average value of ~ 3.5 . This means that the detection limits improve by a factor of ~ 3.5 when assuming correlated data instead of uncorrelated data.

In order to check the validity of our detection criterion (cf. Eq. (25)) we count the number of companions in different categories, which we represent in a confusion matrix

$$\zeta = \begin{pmatrix} \# \text{ of true positives} & \# \text{ of false negatives} \\ \# \text{ of false positives} & \# \text{ of true negatives} \end{pmatrix}, \quad (40)$$

where positive/negative refers to a detection being classified as significant/insignificant according to our detection criterion. Hence, in an ideal world, there would only be true positives or true negatives and ζ would be a diagonal matrix. Obviously, this is not the case in the real world where the data is affected by noise. If we choose an optimistic detection criterion (e.g. the significance needs to be above $1-\sigma$) there will be many false positives (i.e. many detections that are classified as significant, but which are no true companions) and if we choose a pessimistic detection criterion (e.g. the significance needs to be above $5-\sigma$) there will be many false negatives (i.e. many detections that are classified as insignificant, but which are true companions).

The confusion matrices from our companion injection and recovery tests for a $3-\sigma$ detection criterion are

$$\zeta_{\text{diag}} = \begin{pmatrix} 674 & 3 \\ 517 & 654 \end{pmatrix}, \quad (41)$$

$$\zeta_{\text{full}} = \begin{pmatrix} 826 & 213 \\ 12 & 797 \end{pmatrix}, \quad (42)$$

for assuming uncorrelated data (ζ_{diag}) and correlated data (ζ_{full}). In the former case, there is a large fraction of false detections being classified as significant ($517/1171 \approx 44\%$), whereas in the latter case this fraction ($12/809 \approx 1\%$) is roughly consistent with a $3-\sigma$ result. The number of true detections is higher when assuming correlated data ($1039 = 826 + 213$) than when assuming uncorrelated data ($677 = 674 + 3$) because the detection limits are fainter, and although the number of false negatives (i.e. true companions being classified as insignificant) is a lot higher, the number of true positives is still higher when assuming correlated data. In summary, when using our full covariance model

there are less detections above $3-\sigma$ significance than when using the conventional diagonal covariance model ($838 = 826 + 12$ vs. $1191 = 674 + 517$), but the number of true positives (i.e. true companions being classified as significant) is still higher and significant detections are much more reliable since there are almost no false positives. Accounting for the correlations is therefore clearly preferred over ignoring them.

Before proceeding to the injection and recovery tests with real data we also assess the robustness of our correlation model with respect to errors in the model parameters x and y . Therefore, we repeat the injection and recovery tests with wrong correlation and covariance matrices where x and y are only 50% and 25% of their true values respectively. We find that the number of false positives or false negatives increases slightly, but not significantly. This was expected since the detection of asymmetric structure (such as a companion) is governed by the T3 whose correlations are dominated by the correlations of $\pm 1/3$ originating from shared baselines among different triangles. For scenarios where the VIS2 have a larger impact on the model (e.g. measuring stellar diameters) we expect that errors in the model parameters, especially x , have a more significant impact.

3.3. Injection and recovery tests (real data)

In the previous section it is obvious that our full covariance model would outperform the conventional diagonal covariance model, since we simulated data affected by correlated noise. Therefore, the crucial next step is to validate our methods with real GRAVITY data sets. For this purpose, we extract the correlations of the VISAMP and the T3 from the files belonging to programme 0101.C-0907(B) listed in Table 1, fit our empirical models to them and use the VISAMP and the T3 data from the corresponding final GRAVITY pipeline products as noise model.

Since the real data is affected by bright speckles arising from an imperfect calibration, for which our correlation model does not account, we subtract the theoretical VISAMP and T3 of the best fit companion from the data before performing the injection and recovery tests. This also helps us to enter the medium-contrast regime ($f_{\text{fit}} \lesssim 10\%$) where the linearisation of the binary model (see Appendix B) holds. The parameters of the subtracted best fit companion are

$$p_{\text{sub}} = (f, \Delta_{\text{RA}}, \Delta_{\text{Dec}}, \theta) = (0.0383, 0.20, -5.46, 0), \quad (43)$$

and the corresponding detection map is shown in Fig. C.1. The parameters were obtained assuming correlated errors. An extension of our correlation model to inter-observation correlations, for instance arising from the calibration process, is left for future work.

Then, we compute the covariances from the correlations, the VISAMPERR, and the T3ERR from the final GRAVITY pipeline products using Eq. (10) and inject companions with

$$f \in [10^{-3}, 10^{-2.75}, 10^{-2.5}, \dots, 10^{-0.5}], \quad (44)$$

$$\Delta_{RA} \in [-30, -25, -20, \dots, 30] \text{ mas}, \quad (45)$$

$$\Delta_{Dec} \in [-30, -25, -20, \dots, 30] \text{ mas}, \quad (46)$$

using Eqs. (21) and (22). In order to obtain empirical detection limits when assuming uncorrelated and correlated data, we then repeat the model fitting described in the previous Section.

The fraction of correctly recovered companions as a function of the relative flux of the companion for both scenarios (uncorrelated noise: blue points and correlated noise: orange points) is shown in the right panel of Fig. 7, again overplotted with logistic growth functions fitted to the data points (cf. Sect. 3.2). The plot looks similar to the one from the injection and recovery tests with simulated data and confirms the applicability of our full covariance model to real GRAVITY data. Of course, the empirical detection limits are about one to two orders of magnitude worse and the improvement that comes from our full covariance model (i.e. the lateral shift of the orange curve with respect to the blue curve) is only a factor of ~ 2 (consistent with the right panel of Fig. 6 which also shows an improvement by a factor of ~ 2) due to weaker correlations being present in the data used for the injection and recovery tests with real data. In summary, our full covariance model still brings a significant improvement over the conventional diagonal covariance model.

4. Conclusions

Correlated noise is placing fundamental detection limits on interferometric data. From on-sky VLTI/GRAVITY data, we extract and illustrate the correlations present in the data and develop an empirical model in order to describe them. This empirical model is sufficiently simple for it to be fitted to the correlations extracted from a single GRAVITY data product and could therefore be directly integrated into the GRAVITY data reduction pipeline and made available to the community as part of the OIFITS 2 file (which has a well-defined standard for providing covariance matrices Duvert et al. 2017).

Then, we evaluate the impact of our full covariance model by performing model fitting and companion injection and recovery tests with both simulated and real GRAVITY data. Our methods are based on $\chi^2 = R^T \cdot \Sigma^{-1} \cdot R$ minimisation, where we compare the scenarios assuming uncorrelated data (i.e. a diagonal covariance matrix $\text{diag}(\Sigma)$) and correlated data (i.e. a full covariance matrix Σ following from our empirical correlation model). We show that accounting for the correlations that we find to be present in GRAVITY data could yield to an improvement in the detection limits by a factor of up to ~ 3.5 over ignoring them. Moreover, the obtained detection limits (and therefore also potential detections) can be regarded more robust in the former case. We also highlight the problems which arise from ignoring the correlations, as it is done in model fitting pipelines such as LITpro (Tallon-Bosc et al. 2008) and CANDID (Gallenne et al. 2015) so far, and discuss that conventional detection criteria based on χ^2 statistics are strongly biased towards false positives (i.e. detections that are no true companions).

The empirical correlation model presented in this paper is a simple one-parameter model derived from GRAVITY data, but is arguably also applicable (with small modifications) to other instruments such as VLTI/PIONIER for example. It only treats the correlations between the different observables, but not yet between different frames or targets (such as the science and the calibrator target). We choose this approach in order to enable a simple implementation into existing data reduction and model fitting pipelines. Especially with the increasing availability of computing power, the use of full covariance matrices for describing the correlated noise in interferometric data should become a standard. Collaborations around future instruments should provide estimated data covariances as part of the official data reduction pipelines.

In the future, we aim to compare our empirical correlation model with the data covariances derived from bootstrapping (e.g. Lachaume et al. 2019) and extend our model in order to account for correlations between different frames and targets. Finally, we will re-analyse several marginal detections of companions around Cepheid stars from Gallenne et al. (2013, 2014, 2015) by properly accounting for the correlated noise.

Acknowledgements. The authors would like to thank Jean-Baptiste Le Bouquin for helpful feedback on the GRAVITY data reduction pipeline. M. J. I. acknowledges support from the ESO visitor programme. This work has been partly supported by the Australian Research Council's Discovery Projects (DPI190101477). The manuscript was also substantially improved following helpful comments from an anonymous referee.

References

- Absil, O., Le Bouquin, J. B., Berger, J. P., et al. 2011, *A&A*, **535**, A68
 Bartko, H., Perrin, G., Brandner, W., et al. 2009, *New Astron. Rev.*, **53**, 301
 Blackburn, L., Pesce, D. W., Johnson, M. D., et al. 2020, *ApJ*, **894**, 31
 Colavita, M. M. 1999, *PASP*, **111**, 111
 Defrère, D., Absil, O., Berger, J. P., et al. 2018, *Exp. Astron.*, **46**, 475
 Delisle, J. B., Hara, N., & Ségransan, D. 2020, *A&A*, **635**, A83
 Duvert, G., Young, J., & Hummel, C. A. 2017, *A&A*, **597**, A8
 Efron, B., & Tibshirani, R. 1986, *Stat. Sci.*, **1**, 54
 Gallenne, A., Monnier, J. D., Mérand, A., et al. 2013, *A&A*, **552**, A21
 Gallenne, A., Mérand, A., Kervella, P., et al. 2014, *A&A*, **561**, L3
 Gallenne, A., Mérand, A., Kervella, P., et al. 2015, *A&A*, **579**, A68
 Gravity Collaboration (Abuter, R., et al.) 2017, *A&A*, **602**, A94
 Gravity Collaboration (Lacour, S., et al.) 2019, *A&A*, **623**, L11
 Gravity Collaboration (Nowak, M., et al.) 2020, *A&A*, **633**, A110
 Ireland, M. J. 2013, *MNRAS*, **433**, 1718
 Kammerer, J., & Quanz, S. P. 2018, *A&A*, **609**, A4
 Kammerer, J., Ireland, M. J., Martinache, F., & Girard, J. H. 2019, *MNRAS*, **486**, 639
 Kellerer, A., & Tokovinin, A. 2007, *A&A*, **461**, 775
 Kotani, T., Nishikawa, J., Sato, K., et al. 2003, in *Proc. SPIE*, ed. W. A. Traub, et al., *SPIE Conf. Ser.*, **4838**, 1370
 Kulkarni, S. R. 1989, *AJ*, **98**, 1112
 Lachaume, R., Rabus, M., Jordán, A., et al. 2019, *MNRAS*, **484**, 2656
 Lacour, S., Dembet, R., Abuter, R., et al. 2019, *A&A*, **624**, A99
 Lapeyrière, V., Kervella, P., Lacour, S., et al. 2014, in *Proc. SPIE*, *SPIE Conf. Ser.*, 9146, 91462D
 Le Bouquin, J. B., & Absil, O. 2012, *A&A*, **541**, A89
 Le Bouquin, J. B., Berger, J. P., Lazareff, B., et al. 2011, *A&A*, **535**, A67
 Léger, A., Mariotti, J. M., Mennesson, B., et al. 1996, *Icarus*, **123**, 249
 Martinache, F., & Ireland, M. J. 2018, *A&A*, **619**, A87
 Mennesson, B., & Mariotti, J. M. 1997, *Icarus*, **128**, 202
 Monnier, J. D. 2007, *New Astron. Rev.*, **51**, 604
 Perraut, K., Jocou, L., Berger, J. P., et al. 2018, *A&A*, **614**, A70
 Perrin, G. 2003, *A&A*, **400**, 1173
 Quanz, S. P., Kammerer, J., Defrère, D., et al. 2018, in *Proc. SPIE*, *SPIE Conf. Ser.*, 10701, 10701I
 Quanz, S. P., Absil, O., Angerhausen, D., et al. 2019, *ArXiv e-prints* [arXiv:1908.01316]
 Tallon-Bosc, I., Tallon, M., Thiébaud, E., et al. 2008, in *Proc. SPIE*, *SPIE Conf. Ser.*, 7013, 70131J

Appendix A: Collection of correlations

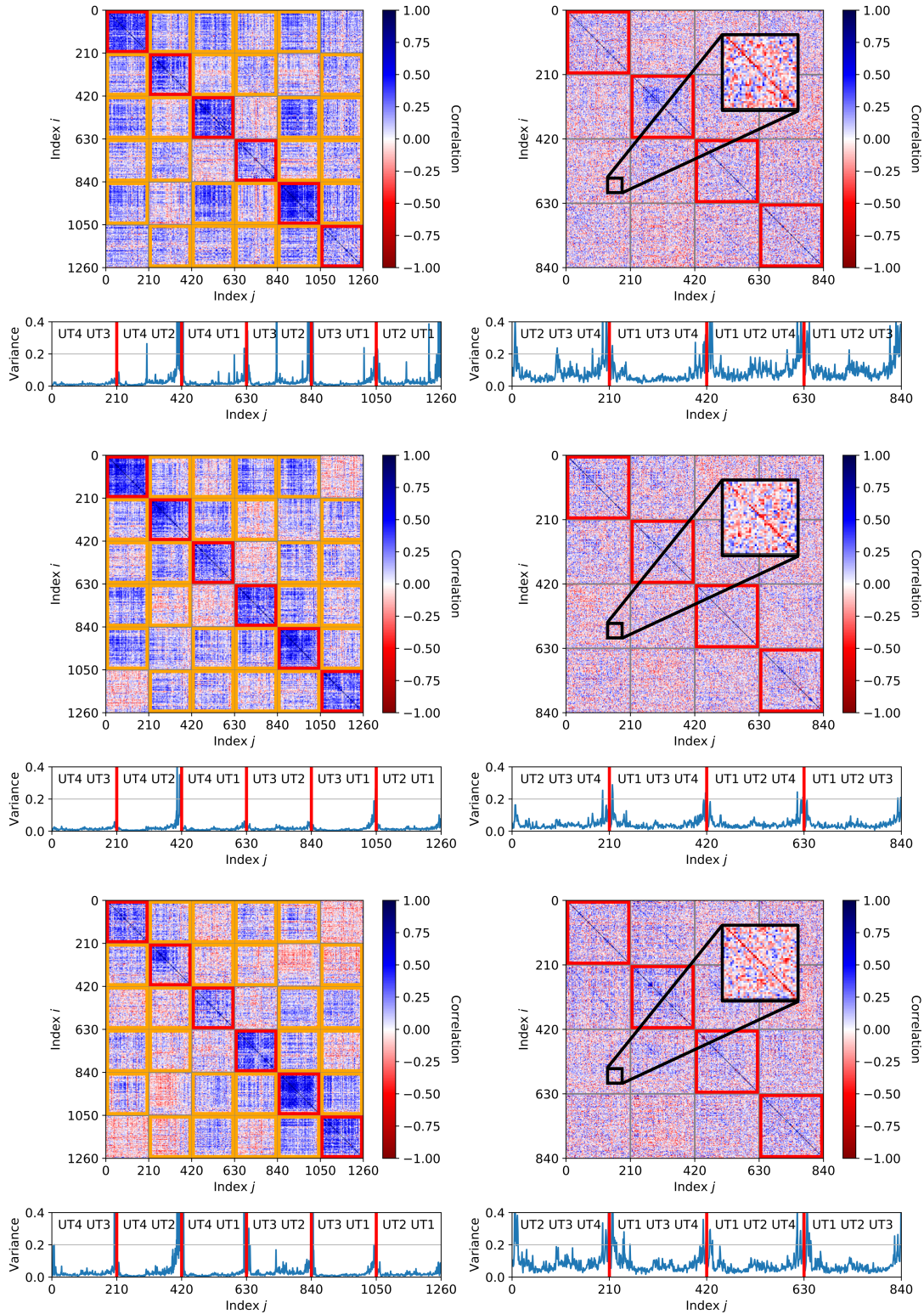


Fig. A.1. Same as Fig. 2, but extracted from three other P2VM-reduced files from programme 60.A-9801(U). *From top to bottom:* GRAVI.2019-03-29T01-46-28.155_singlecalp2vmred.fits, GRAVI.2019-03-29T01-57-13.182_singlecalp2vmred.fits, GRAVI.2019-03-29T01-59-31.188_singlecalp2vmred.fits.

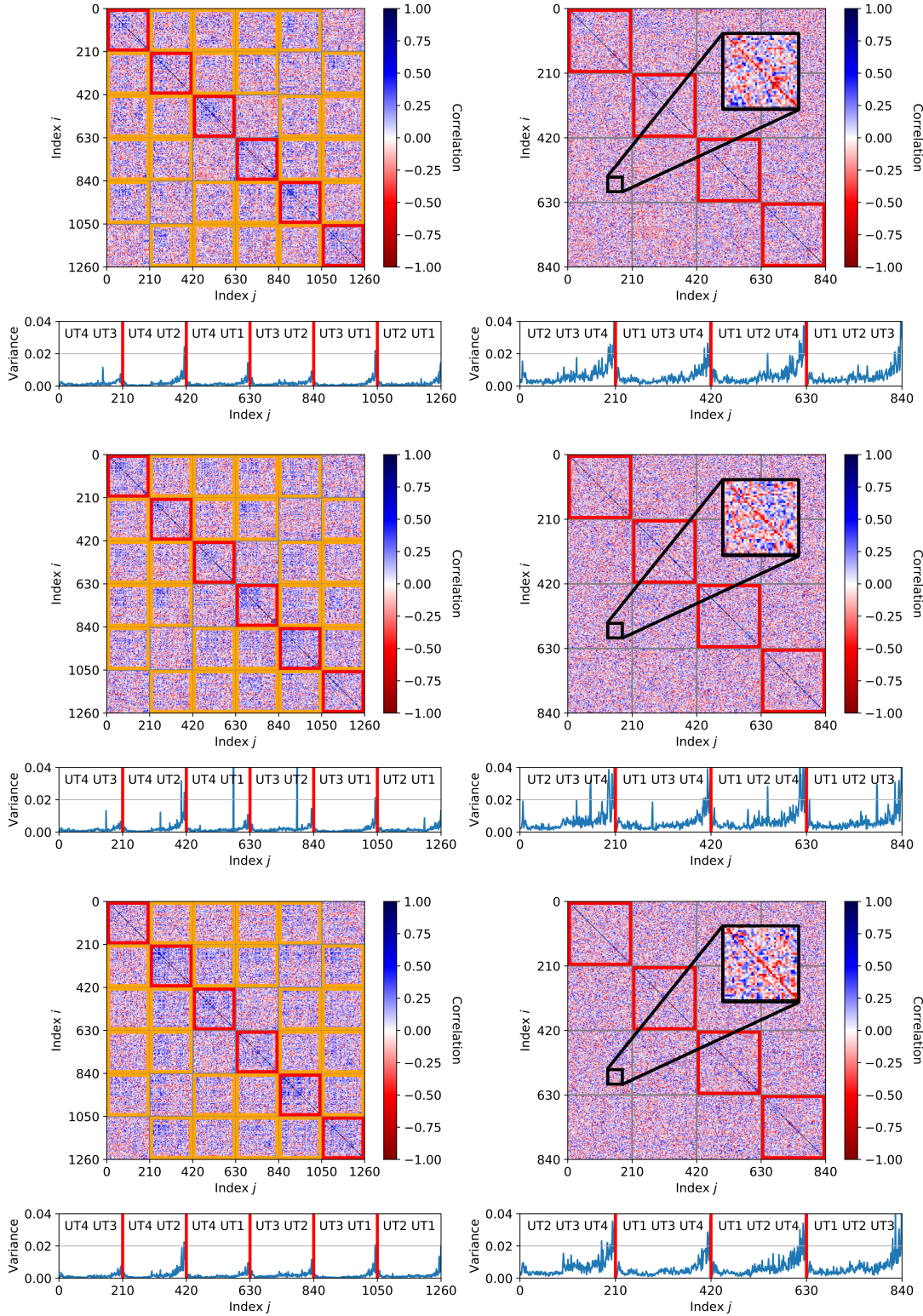


Fig. A.2. Same as Fig. 2, but showing the correlations of the VISAMP instead of the VIS2, extracted from the three P2VM-reduced files used for the injection and recovery tests with real data in Sect. 3.3 (programme 0101.C-0907(B)). From top to bottom: GRAVI.2018-04-18T08-08-19.739_singlescip2vmred.fits, GRAVI.2018-04-18T08-12-10.749_singlescip2vmred.fits, GRAVI.2018-04-18T08-20-04.769_singlescip2vmred.fits.

Appendix B: Linearised model

If the host star is essentially unresolved (i.e. $\theta b \lambda \ll 1$) and the companion is at high contrast (i.e. $f \ll 1$) one can linearise the $VIS_{2\text{bin}}$ and the $T3_{\text{bin}}$ as a function of the relative flux of the companion.

Consider the complex visibility of the binary model VIS_{bin} in the case where $\theta \rightarrow 0 \Leftrightarrow VIS_{\text{ud}} \rightarrow 1$ and $f \ll 1$, then the VIS_2 of this model is

$$|VIS_{\text{bin}}|^2 \approx \frac{1}{(1+f)^2} \left[(1+f \cos(x))^2 + f^2 \sin^2(x) \right] \quad (\text{B.1})$$

$$= \frac{1}{(1+f)^2} \left[1 + 2f \cos(x) + f^2 \right] \quad (\text{B.2})$$

$$= (1 - 2f + \mathcal{O}(f^2)) \left[1 + 2f \cos(x) + f^2 \right] \quad (\text{B.3})$$

$$= 1 + 2f \cos(x) - 2f + f^2 - 4f^2 \cos(x) + \mathcal{O}(f^3) \quad (\text{B.4})$$

$$= 1 + f(2 \cos(x) - 2) + \mathcal{O}(f^2) \quad (\text{B.5})$$

and the phase (or argument) of this model (and therefore any linear combination of phases such as the T3) is

$$\angle VIS_{\text{bin}} = \arctan \frac{\text{Im}VIS_{\text{bin}}}{\text{Re}VIS_{\text{bin}}} \quad (\text{B.6})$$

$$\approx \frac{\text{Im}VIS_{\text{bin}}}{\text{Re}VIS_{\text{bin}}} \quad (\text{B.7})$$

$$= \frac{-f \sin(x)}{1 + f \cos(x)} \quad (\text{B.8})$$

$$\approx f \frac{-\sin(x)}{1} \quad (\text{B.9})$$

for $x = -2\pi i(\Delta_{\text{RA}}u/\lambda + \Delta_{\text{Dec}}v/\lambda)$. Hence, in the high-contrast regime, one has $VIS_{2\text{bin}} \propto 1 + f$ and $T3_{\text{bin}} \propto f$.

Appendix C: Detection map for real data

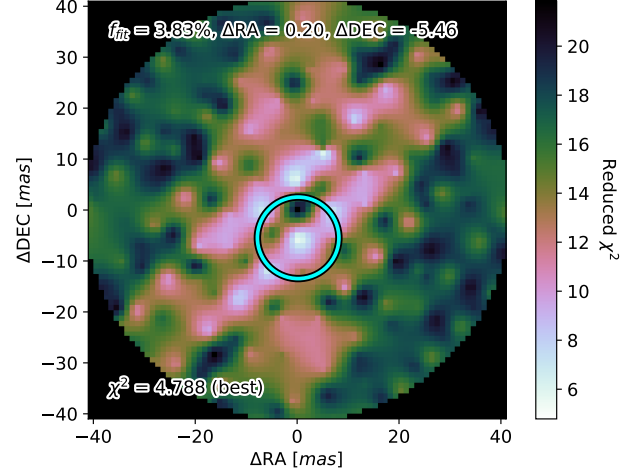


Fig. C.1. Detection map for the GRAVITY data used for the injection and recovery tests in Sect. 3.3. The host star is located in the centre of the map and the cyan circle highlights the position of the best fit companion. Its parameters and reduced chi-squared are shown at the top and the bottom of the map. North is up and east is left.