

# Avoiding bias in the model fitting of correlated interferometric data

Régis Lachaume

27 January 2021

## ABSTRACT

I show how Peelle’s pertinent puzzle can be avoided in the framework of the least squares minimisation.

## 1 INTRODUCTION

Optical and infrared long-baseline interferometry consists in measuring the fringe contrast and phase of interferences in the recombined light collected at several telescopes<sup>1</sup>. These observables hold information on the celestial object’s spatial properties, which are usually obtained through model fitting.

In spite of strong evidence of correlations in data, due to redundancy (Monnier 2007, in the case of closure phases), calibration (Perrin 2003), or atmospheric biases acting on all spectral channels in the same way (Lawson 2000), only a few authors (Perrin et al. 2004; Absil et al. 2006; Berger et al. 2006; Lachaume et al. 2019; Kammerer et al. 2020) have analysed correlated data while most assume independant errors. In particular, the only interferometric instrument I know of with a data processing software taking into account one source of correlations—calibration—is the FLUOR<sup>2</sup> instrument (at IOTA<sup>3</sup>, then CHARA<sup>4</sup> Perrin et al. 2004). None of the five first and second-generation ones at the VLTI<sup>5</sup> does (Millour et al. 2008; Hummel & Percheron 2006; Le Bouquin et al. 2011; ESO GRAVITY pipeline team 2020; ESO MATISSE pipeline team 2020). The same lack of support for correlations is present in image reconstruction programmes (e.g. MIRA, see Thiébaud 2008), model-fitting tools (e.g. Litpro, see Tallon-Bosc et al. 2008), or the still widespread first version of the Optical Interferometric FITS format (OIFITS v. 1, Pauls et al. 2005).

Unfortunately, it has been shown that ignoring uncertainties may lead to significant errors in model parameters as Lachaume et al. (2019) showed with stellar diameters as PIONIER<sup>6</sup> (Le Bouquin et al. 2011) data at the VLTI. Also Kammerer et al. (2020) established that correlations were necessary to achieve a higher contrast ratio in companion detection using GRAVITY (Eisenhauer et al. 2011) at the VLTI.

Several sources of correlated uncertainties occur in a multiplicative context, when several data points are normalised with the transfer function (Perrin 2003) or the coherent fluxes

are derived with the pixel-to-visibility matrix formalism (Tatulli et al. 2007). In both cases, the uncertainty on the multiplicative factor translates into a systematic, correlated one in the final data product. In the context of experimental nuclear physics, Peelle (1987) noted that this scenario could lead to an estimate falling below the individual data points, a paradox known as Peelle’s Pertinent Puzzle (PPP). It results from the usual and straightforward, but actually incorrect, way to propagate covariances (D’Agostini 1994; Neudecker et al. 2012). A few workarounds have been proposed (e.g. Burr et al. 2011; Becker et al. 2012; Nisius 2014) but they are not straightforward.

The issue, however, not widely known in many other fields where the problem has seldom arisen. In this paper, I present the paradox within the context of long-baseline interferometry, analyse its possible effect in model-fitting, and propose a simple, computer-efficient way to avoid it.

## 2 PEELE’S PERTINENT PUZZLE BASICS

I rewrite and adapt Peelle’s original example in the context of long-baseline interferometry. Let’s assume the inverse of the instrumental fringe contrast  $\tau \pm \tau\sigma_\tau$  has been interpolated from one or several calibrator observations. A visibility amplitude is now now estimated from two contrast measurements  $\nu_1 \pm \sigma_\nu$  and  $\nu_2 \pm \sigma_\nu$ . For each measurement, visibility amplitudes are:

$$v_1 = \tau\nu_1 \pm \tau\sigma_\nu (\pm\tau\nu_1\sigma_\tau), \quad (1a)$$

$$v_2 = \tau\nu_2 \pm \tau\sigma_\nu (\pm\tau\nu_2\sigma_\tau). \quad (1b)$$

They are normalised with the same quantity ( $\tau$ ), so they are correlated, hence the systematic uncertainty term between parentheses in Eq. (1). Error propagation gives the covariance matrix

$$\Sigma = \begin{pmatrix} \sigma_\nu^2\tau^2 + \sigma_\tau^2\nu_1^2 & \sigma_\nu^2\nu_1\nu_2 \\ \sigma_\tau^2\nu_1\nu_2 & \sigma_\nu^2\tau^2 + \sigma_\tau^2\nu_2^2 \end{pmatrix}, \quad (2)$$

when the second-order error term ( $\sigma_\nu\sigma_\tau$ ) is ignored. Using

<sup>1</sup> Recombination may be performed by software in the case of intensity interferometry or heterodyne detection.

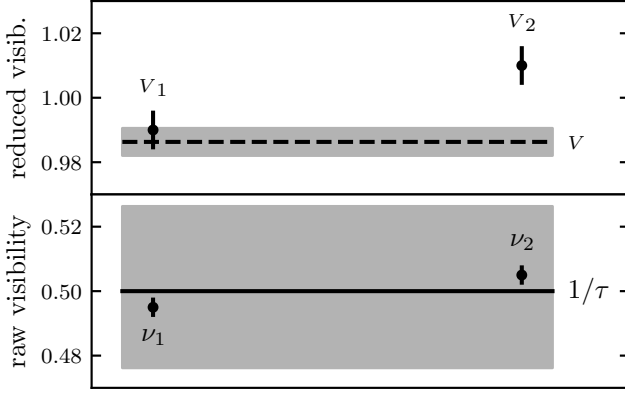
<sup>2</sup> Fiber Linked Unit for Optical Recombination

<sup>3</sup> Infrared and Optical Telescope Array

<sup>4</sup> Center for High Angular Resolution Array

<sup>5</sup> Very Large Telescope Interferometer

<sup>6</sup> Precision Integrated-Optics Near-infrared Imaging Experiment



**Figure 1.** Original Peele problem rewritten in the context of interferometry. *Bottom:* Two raw visibility amplitudes  $\nu_1$  and  $\nu_2$  (points with statistic error bars of  $\approx 0.6\%$ ) are calibrated by the transfer function  $1/\tau$  (solid line with systematic error zone of 5%). *Top:* the two calibrated visibility measurements  $V_1$  and  $V_2$  (points with statistic error bars of  $\approx 0.6\%$ ) are strongly correlated. The least-squares estimate for the visibility  $V$  (dashed line, with statistic uncertainty error zone displayed) falls outside of the data range.

the matrix inverse  $\Sigma^{-1}$ , the least squares estimate is

$$V = \frac{(\Sigma_{11}^{-1} + \Sigma_{12}^{-1})\nu_1 + (\Sigma_{22}^{-1} + \Sigma_{12}^{-1})\nu_2}{(\Sigma_{11}^{-1} + \Sigma_{12}^{-1}) + (\Sigma_{22}^{-1} + \Sigma_{12}^{-1})}, \quad (3)$$

$$= \frac{\nu_1 + \nu_2}{2} \left( 1 + \sigma_\tau^2 \frac{(\nu_1 - \nu_2)^2}{2\tau^2 \sigma_\nu^2} \right)^{-1}.$$

It systematically falls below the average of the two values  $\nu_1$  and  $\nu_2$ . If measurements differ significantly, it can even fall below the lowest value. Figure 1 gives such an example with a instrumental visibility of 50% and two measurements on an unresolved target:

$$\begin{aligned} \tau &= 2.000 \pm 0.100 \text{ (5\%)}, \\ \nu_1 &= 0.495 \pm 0.003 (\approx 0.6\%), \\ \nu_2 &= 0.505 \pm 0.003 (\approx 0.6\%), \end{aligned}$$

which yields two points 2.4 standard deviations apart

$$\begin{aligned} V_1 &= 0.990 \pm 0.006 (\pm 0.050), \\ V_2 &= 1.010 \pm 0.006 (\pm 0.050) \end{aligned}$$

and the estimate

$$V = 0.986 \pm 0.004 (\pm 0.049)$$

falls outside the data range.

### 3 SINGLE VALUE FIT

Let's consider a set of measurements of a single normalised quantity, like the visibility amplitude, given by a vector  $\mathbf{v} = (v_1, \dots, v_n)^T$ . It is derived from an uncalibrated quantity like the fringe contrast,  $\boldsymbol{\nu} = (\nu_1, \dots, \nu_n)^T$  and a normalisation factor, like the cotransfer function,  $\boldsymbol{\tau} = (\tau_1, \dots, \tau_n)^T$  by

$$\mathbf{v} = \boldsymbol{\tau} \odot \boldsymbol{\nu}, \quad (4)$$

**Table 1.** Symbols used in this paper. Lower case bold font is used for vectors and upper case bold font for matrices.

Symbol	Meaning
$\bar{a}$	true value of $a$
$\langle a \rangle$	expectancy of $a$
$\mathbf{A}^T$	transpose of $\mathbf{A}$
$\mathbf{c} = \mathbf{a} \odot \mathbf{b}$	elementwise product of $\mathbf{a}$ and $\mathbf{b}$
$\mathbf{C} = \mathbf{a} \otimes \mathbf{b}$	outer product of $\mathbf{a}$ and $\mathbf{b}$
$\mathbf{c} = \mathbf{A}\mathbf{b}$	matrix product of $\mathbf{A}$ and $\mathbf{b}$
$\delta$	Kronecker delta
$\mathbf{v}$	data
$\boldsymbol{\eta}$	error ( $= \mathbf{v} - \bar{\mathbf{v}}$ )
$\sigma^2$	variance ( $= \langle \eta^2 \rangle$ )
$\Sigma$	covariance matrix ( $= \langle \boldsymbol{\eta} \otimes \boldsymbol{\eta} \rangle$ )
$\varrho$	correlation coefficient
$\mathbf{x}, \mathbf{X}$	sensitivity vector or matrix
$\mathbf{p}$	parameters of the model
$\boldsymbol{\mu}$	model values ( $= \mathbf{X}\mathbf{p} \approx \bar{\mathbf{v}}$ )
$a_\nu$	$a$ of the measurement error
$a_\tau$	$a$ of the normalisation error
$a^*$	$a$ impacted by PPP

where  $\odot$  denotes the Hadamard (elementwise) product of vectors. If  $\bar{\mathbf{v}}$ ,  $\bar{\tau}$ , and  $\bar{\nu}$  the true, but unknown, values of these quantities, the error vector on  $\mathbf{v}$

$$\boldsymbol{\eta} = \mathbf{v} - \bar{\mathbf{v}}, \quad (5)$$

can be written as a sum of measurement and normalisation errors if one ignores a second-order term:

$$\boldsymbol{\eta} = \boldsymbol{\eta}_\nu + \bar{\mathbf{v}}\boldsymbol{\eta}_\tau. \quad (6)$$

These errors are given by

$$\boldsymbol{\eta}_\nu = \bar{\tau}(\boldsymbol{\nu} - \bar{\nu}), \quad (7)$$

$$\boldsymbol{\eta}_\tau = \frac{1}{\bar{\tau}}(\boldsymbol{\tau} - \bar{\tau}). \quad (8)$$

Let's assume  $\boldsymbol{\eta}_\nu$  and  $\boldsymbol{\eta}_\tau$  are independent error terms of mean 0 and standard deviations are  $\sigma_\nu$  and  $\sigma_\tau$ , respectively. In addition, I consider correlation of the normalisation errors, with correlation coefficient  $\varrho$ . In the case of interferometry, it can arise from the uncertainty on the calibrators' geometry. The covariance matrix is given by

$$\Sigma = \langle \boldsymbol{\eta} \otimes \boldsymbol{\eta} \rangle, \quad (9)$$

where  $\otimes$  denotes the outer product of vectors and  $\langle \rangle$  stands for the expectancy, so that

$$\Sigma_{ij} = [\sigma_\nu^2 + (1 - \varrho)\sigma_\tau^2\bar{\nu}^2] \delta_{ij} + \varrho\sigma_\tau^2\bar{\nu}^2. \quad (10)$$

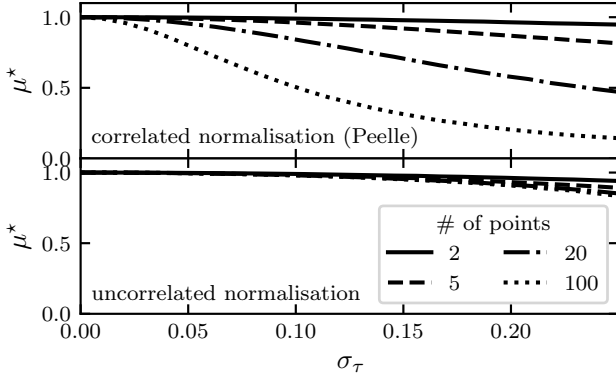
The value  $\bar{\nu}$  is yet to be determined, so the covariances are often derived using the data in the propagation:

$$\Sigma_{ij}^* = [\sigma_\nu^2 + (1 - \varrho)\sigma_\tau^2\nu_i^2] \delta_{ij} + \varrho\sigma_\tau^2\nu_i\nu_j. \quad (11)$$

The least squares estimate for  $\bar{\mathbf{v}}$  is given by

$$\boldsymbol{\mu}^* = \frac{\mathbf{x}^T \Sigma^{*-1} \mathbf{v}}{\mathbf{x}^T \Sigma^{*-1} \mathbf{x}} \quad (12)$$

where  $\mathbf{x} = (1, \dots, 1)^T$  is the trivial sensitivity vector. The covariance matrix is the sum of an invertible diagonal matrix



**Figure 2.** Fit  $\mu^*$  to unresolved visibilities ( $v = 1$ ), as a function of the relative uncertainty on the calibration  $\sigma_\tau$  and the number of measurements  $n$ .  $2/n \times 10^5$  simulations were made and averaged, assuming that  $\eta_\nu$  and  $\eta_\tau$  follow normal distributions. *Top*: fully correlated normalisation like in original Peelle’s puzzle ( $\sigma_\nu = 0.02$  and  $\varrho = 1$ ). *Bottom*: normalisation error without correlation ( $\sigma_\nu = 0.02$  and  $\varrho = 0$ ).

and one of rank one—see Eq. (11)—, so that the inverse is obtained using the Woodbury matrix identity:

$$\{\Sigma^{*-1}\}_{ij} = \frac{\delta_{ij}}{\sigma_i^2} - \frac{\varrho \sigma_\tau^2 V_i V_j}{\sigma_i^2 \sigma_j^2 \left(1 + \varrho \sigma_\tau^2 \sum_k \frac{V_k^2}{\sigma_k^2}\right)}, \quad (13)$$

where

$$\sigma_i^2 = \sigma_\nu^2 + (1 - \varrho) \sigma_\tau^2 V_i^2. \quad (14)$$

It allows me to put the least squares estimate in the relatively compact form

$$\mu^* = \frac{\sum_i \frac{V_i}{\sigma_i^2}}{\sum_i \left(1 + \varrho \sigma_\tau^2 \sum_j \frac{(V_i - V_j)^2}{\sigma_j^2}\right) \frac{1}{\sigma_i^2}}. \quad (15)$$

For small enough errors ( $\eta_i \ll \bar{v}$ ) the second-order Taylor development in  $\eta_i = v_i - \bar{v}$  yields

$$\mu^* \approx \bar{v} + \frac{1}{n} \sum_i \eta_i - \frac{[2 + (n-2)\varrho] \sigma_\tau^2 \bar{v}^2}{2n\sigma^2} \sum_{i \neq j} (\eta_i - \eta_j)^2 \quad (16)$$

where

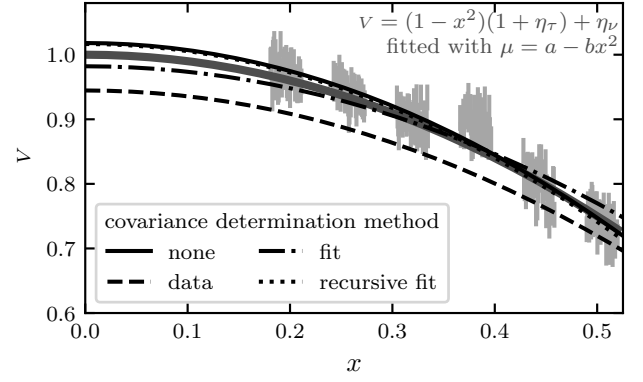
$$\bar{\sigma}^2 = \sigma_\nu^2 + (1 - \varrho) \sigma_\tau^2 \bar{v}^2 \quad (17)$$

Since  $\langle \eta_i \rangle = 0$  and  $\langle (\eta_i - \eta_j)^2 \rangle = 2\bar{\sigma}^2$ , the expectancy

$$\langle \mu^* \rangle \approx \bar{v} \left[ 1 - \left(1 - \frac{1}{n}\right) (2 + (n-2)\varrho) \sigma_\tau^2 \right] \quad (18)$$

is biased. If the data are not correlated ( $\varrho = 0$ ), the bias is small ( $\sigma_\tau^2$  to  $2\sigma_\tau^2$ ) but it becomes large for correlated data if the number of points is large ( $\sim n\sigma_\tau^2$  for fully correlated data) as D’Agostini (1994) already noted. Figure 2 shows a simulation of the bias as a function of the normalisation uncertainty  $\sigma_\tau$  for various data sizes ( $n = 2$  to 100).

For spectro-interferometric observations with 4 telescopes, the number of correlated points can be over 1,000, so even with a low correlation coefficient, the impact can be large.



**Figure 3.** Example of a fit of correlated data from a four telescope interferometer with medium resolution: 6 groups of 100 data points (light gray points with the measurement error bar) with 2% uncorrelated measurement errors and 3% correlated normalisation. Simulated data follow  $v = 1 - x^2$  (thick gray line). Least-squares model fitting  $\mu = a - bx^2$  is performed using the four prescriptions for the covariance matrix.

For instance, a single GRAVITY observation in high spectral resolution yields  $n = 6 \times 210$  visibility amplitudes. With an observed correlation of  $\rho \approx 16\%$  in the instrumental visibility amplitudes (Kammerer et al. 2020) and a typical  $\sigma_\tau = 1$ –2% normalisation error, the bias on the calibrated visibilities could be 2–8%.

## 4 MODEL FITTING

Let’s now consider a set of measurements corresponding to the linear model

$$\mu = \mathbf{X} \mathbf{p}, \quad (19)$$

where  $\mathbf{p}$  are the unknown parameters and  $\mathbf{X}$  is the known sensitivity matrix. Typically,  $x_{ik} = f_k(u_i, v_i)$  for a linear model and  $x_{ik} = \partial f / \partial p_k(u_i, v_i)$  for a non-linear model approximated by a linear one close to a solution.  $(u, v)$  is the reduced baseline. The true values  $\bar{\mathbf{v}}$  are impacted by errors so that the data are

$$\mathbf{v} = \bar{\mathbf{v}} + \boldsymbol{\eta} \quad (20)$$

with the error term  $\boldsymbol{\eta}$  again expressed as the sum of a measurement error and a normalisation one:

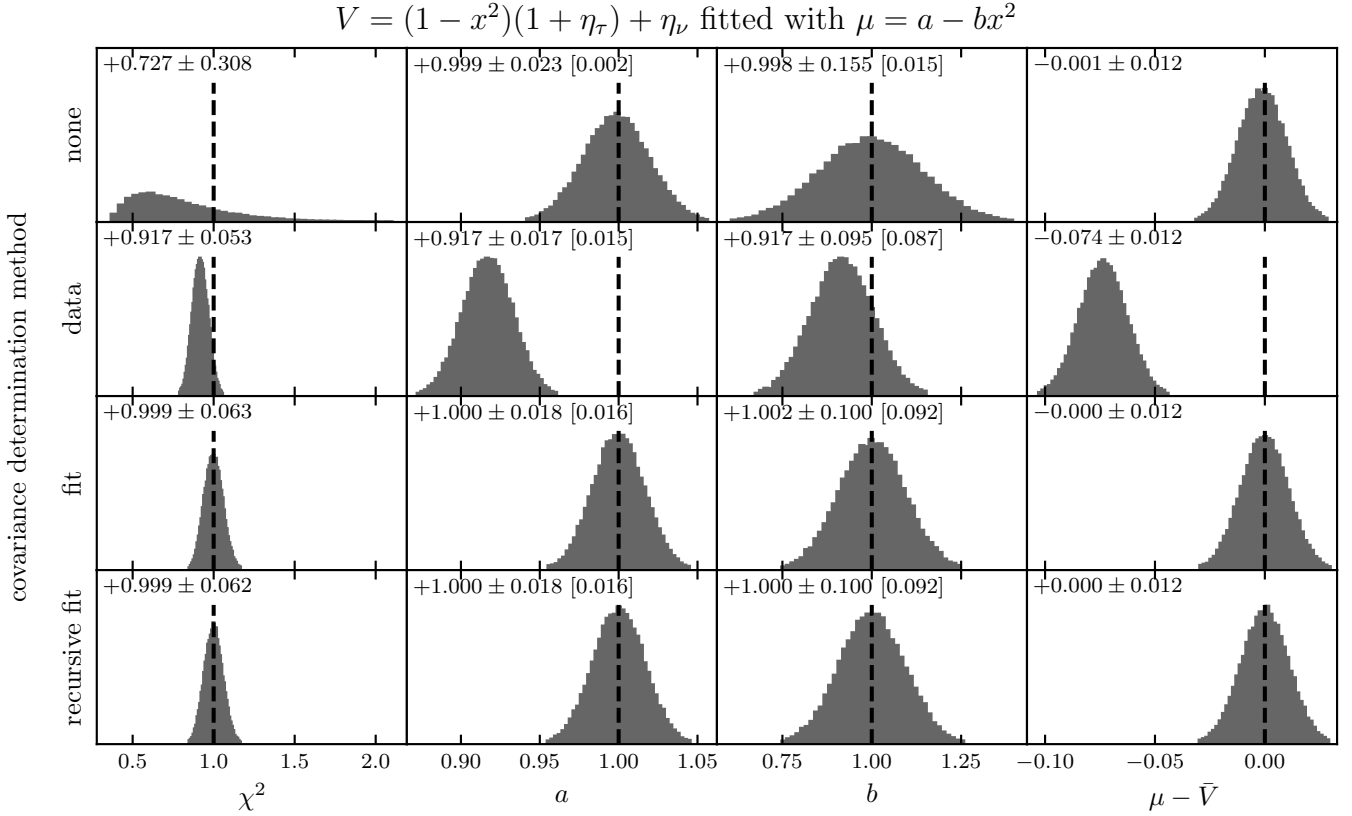
$$\boldsymbol{\eta} = \boldsymbol{\eta}_\nu + \boldsymbol{\eta}_\tau \odot \bar{\mathbf{v}}. \quad (21)$$

The measurement errors  $\boldsymbol{\eta}_\nu$  and normalisation errors  $\boldsymbol{\eta}_\tau$  following multivariate distributions of mean zero and covariance matrices  $\Sigma_\nu$  and  $\Sigma_\tau$  respectively. Given the covariance matrix  $\Sigma$  of this model, the least squares estimate is

$$\mathbf{p} = (\mathbf{X}^T \Sigma^{-1} \mathbf{X})^{-1} (\mathbf{X}^T \Sigma^{-1} \mathbf{v}) \quad (22)$$

I investigate four ways to determine the covariance matrix

(i) Ignoring the correlations in the normalisation using  $\Sigma_0 = \Sigma_\nu + (\mathbf{v} \otimes \mathbf{v}) \odot (\Sigma_\tau \odot \mathbf{I})$ . Let  $\mu_0 = \mathbf{X}(\mathbf{X}^T \Sigma_0^{-1} \mathbf{X})^{-1} \mathbf{X}^T \Sigma_0^{-1} \mathbf{v}$  the resulting model of the data.



**Figure 4.** Distribution of the fitted parameters and fit properties for the four covariance matrix prescriptions.  $5 \times 10^4$  simulations of 6 groups of 100 correlated data points  $v$  ( $\sigma_v = 2\%$ ,  $\sigma_\tau = 3\%$ ,  $\varrho = 1$ , normal distributions) following  $v = (1 - x^2)(1 + \eta) + \eta$  are performed and fitted with model  $\mu = a - bx^2$  using least-squares minimisation. Reported quantities include median and 1- $\sigma$  interval of their distribution and, within brackets, the median uncertainty reported by the least squares fit. *Leftmost column:* distribution of the least squares; *middle left:* constant coefficient  $a$ ; *middle right:* quadratic coefficient  $b$ ; *rightmost column:* mean difference between model and data. The covariance matrix prescriptions are: *top row:* correlations are ignored; *second row:* a naïve covariance matrix uses the data values; *third row:* covariance matrix uses modelled values from fit without correlations; *bottom row:* covariance matrix and model are recursively computed, with the covariance matrix of the next recursion using the modelled value of the last step.

(ii) Using the naïve estimate  $\Sigma^* = \Sigma_\nu + (v \otimes v) \odot \Sigma_\tau$  which is known to lead to Peelle’s pertinent puzzle in the trivial case of a constant model.

(iii) Using the data model of the fit without the normalisation error:  $\Sigma_1 = \Sigma_\nu + (\mu_0 \otimes \mu_0) \odot \Sigma_\tau$ . This is the generalisation of the two-variables approach by Neudecker et al. (2014). The resulting least squares model is  $\mu_1 = X(X^T \Sigma_1^{-1} X)^{-1} X^T \Sigma_1^{-1} v$ .

(iv) Recursively fitting the data by updating the data model in the covariance matrix. I derive  $\mu_k = X(X^T \Sigma_k^{-1} X)^{-1} X^T \Sigma_k^{-1} v$  using  $\Sigma_k = \Sigma_\nu + (\mu_{k-1} \otimes \mu_{k-1}) \odot \Sigma_\tau$ , starting with the estimate  $\mu_1$  ( $k = 2$ ).

Figure 3 shows the example of a typical centrosymmetric quadratic fit to under-resolved data obtained at a four-telescope facility in medium spectral resolution mode. Given the dimensionless variable  $x$  proportional to the projected baseline length ( $x \propto \sqrt{u^2 + v^2}$ ), a model fit  $\mu = a - bx^2$  is performed to simulated data following  $v \approx 1 - x^2$ , including a (correlated) systematic normalisation error  $\eta_\tau$  (3%) and independent statistical errors  $\eta_\nu$  (2%). As expected, the use of data  $v$  in the correlation matrix (method (ii)) leads to grossly underestimated data values, in the very same way as in the

classical Peelle case described in Sects. 2 & 3. Other methods, including ignoring correlations, yield reasonable fits.

Figure 4 sums up the behaviour of the same fit performed a large number of times on different simulated data sets, each following  $v \approx 1 - x^2$ . For each correlation matrix prescription, it displays the dispersion of the reduced chi squared, the model parameters  $a$  and  $b$ , and the difference between modelled value and true value. It also reports the uncertainty on model parameters reported by the least squares fit in comparison to the scatter of the distribution of the values. Both methods estimating the correlation matrix from modelled data ((iii) & (iv)) are equivalent in terms of the absence of bias, dispersion of these quantities, and correct prediction of the uncertainty on model parameters. While the model fits ignoring correlations ((i)) are not biased either, they display a higher dispersion of model parameters and grossly underestimate the uncertainty on model parameters. The correlation matrix calculated from data ((ii)) is, as expected, strongly biased.

## 5 CONCLUSION

- (i) Peelle’s pertinent puzzle primarily arises from an incorrect description of the relative uncertainties. It will occur in a fully linear problem and, to some extent, without correlations.
- (ii) For uncorrelated data, the relative bias is of the order of the square of the relative uncertainty as long as errors remain small. It increases twofold when the number of data points becomes large.
- (iii) For correlated data, the relative bias is larger by a factor of the order of the number of correlated points as long as the resulting error remain small.
- (iv) In least-squares model fits, the problem disappears if the uncertainties, or the covariance matrix, are updated with the model values instead of the measurement values.

## REFERENCES

- Absil O., et al., 2006, *A&A*, 452, 237
- Becker B., et al., 2012, *Journal of Instrumentation*, 7, P11002
- Berger D. H., et al., 2006, *ApJ*, 644, 475
- Burr T., Kawano T., Talou P., Pan F., Hengartner N., 2011, *Algorithms*, 4, 28
- D’Agostini G., 1994, *Nuclear Instruments and Methods in Physics Research A*, 346, 306
- ESO GRAVITY pipeline team 2020, GRAVITY pipeline user manual Issue 1.4
- ESO MATISSE pipeline team 2020, MATISSE pipeline user manual Issue 1.5.1
- Eisenhauer F., et al., 2011, *The Messenger*, 143, 16
- Hummel C. A., Percheron I., 2006, in *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*. p. 62683X, [doi:10.1117/12.671337](https://doi.org/10.1117/12.671337)
- Kammerer J., Mérand A., Ireland M. J., Lacour S., 2020, *A&A*, 644, A110
- Lachaume R., Rabus M., Jordán A., Brahm R., Boyajian T., von Braun K., Berger J.-P., 2019, *MNRAS*, 484, 2656
- Lawson P. R., ed. 2000, *Principles of Long Baseline Stellar Interferometry*
- Le Bouquin J.-B., et al., 2011, *A&A*, 535, A67
- Millour F., Valat B., Petrov R. G., Vannier M., 2008, in *Optical and Infrared Interferometry*. p. 701349 ([arXiv:0807.0291](https://arxiv.org/abs/0807.0291)), [doi:10.1117/12.788707](https://doi.org/10.1117/12.788707)
- Monnier J. D., 2007, *New Astronomy Reviews*, 51, 604
- Neudecker D., Frühwirth R., Leeb H., 2012, *Nuclear Science and Engineering*, 170, 54
- Neudecker D., Frühwirth R., Kawano T., Leeb H., 2014, *Nuclear Data Sheets*, 118, 364
- Nisius R., 2014, *European Physical Journal C*, 74, 3004
- Pauls T. A., Young J. S., Cotton W. D., Monnier J. D., 2005, *PASP*, 117, 1255
- Peelle R. W., 1987, *Informal memorandum, Peelle’s Pertinent Puzzle*. Oak Ridge National Laboratory
- Perrin G., 2003, *A&A*, 400, 1173
- Perrin G., Ridgway S. T., Coudé du Foresto V., Mennesson B., Traub W. A., Lacasse M. G., 2004, *A&A*, 418, 675
- Tallon-Bosc L., et al., 2008, in *Optical and Infrared Interferometry*. p. 70131J, [doi:10.1117/12.788871](https://doi.org/10.1117/12.788871)
- Tatulli E., et al., 2007, *A&A*, 464, 29
- Thiébaud E., 2008, in *Optical and Infrared Interferometry*. p. 70131I, [doi:10.1117/12.788822](https://doi.org/10.1117/12.788822)