CrossMark

CAMBRIDGE
UNIVERSITY PRESS

Research Paper

Bias-free model fitting of correlated data in interferometry

Régis Lachaume [ORCID]

Instituto de Astronomía and Centro de Astroingeniería, Facultad de Física, Pontificia Universidad Católica de Chile, casilla 306, Santiago 22, Chile and
Max-Planck-Institut für Astronomie, Königstuhl 17, D-69117 Heidelberg, Germany

Abstract

In optical and infrared long-baseline interferometry, data often display significant correlated errors because of uncertain multiplicative factors such as the instrumental transfer function or the pixel-to-visibility matrix. In the context of model fitting, this situation often leads to a significant bias in the model parameters. In the most severe cases, this can can result in a fit lying outside of the range of measurement values. This is known in nuclear physics as Peelle's Pertinent Puzzle. I show how this arises in the context of interferometry and determine that the relative bias is of the order of the square root of the correlated component of the relative uncertainty times the number of measurements. It impacts preferentially large datasets, such as those obtained in medium to high spectral resolution. I then give a conceptually simple and computationally cheap way to avoid the issue: model the data without covariances, estimate the covariance matrix by error propagation using the modelled data instead of the actual data, and perform the model fitting using the covariance matrix. I also show that a more imprecise but also unbiased result can be obtained from ignoring correlations in the model fitting.

Keywords: techniques: interferometric – methods: statistical – methods: data analysis

(Received 2 February 2021; revised 28 April 2021; accepted 28 April 2021)

1. Introduction

Optical and infrared long-baseline interferometry consists in measuring the fringe contrast and phase of interference fringes in the recombined light collected at several telescopes.[a] These observables hold information on the celestial object's spatial properties, often obtained through model fitting.

In spite of strong evidence of correlations in the 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 accounted for these correlations while most assumed statistically independent errors. In particular, the only interferometric instrument I know of with a data processing software taking into account one source of correlations—calibration—is FLUOR[b] (at IOTA,[c] then CHARA,[d] Perrin et al. 2004). None of the five first and second-generation ones at the VLTI[e] 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 Thiébaut 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, ignoring correlations may lead to significant errors in model parameters as Lachaume et al. (2019) evidenced with stellar diameters using PIONIER[f] (Le Bouquin et al. 2011) data at the VLTI. Also Kammerer et al. (2020) established that accounting for correlations is 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, but actually incorrect, way to propagate covariances, in which the measured values are used in the calculations (D'Agostini 1994; Neudecker, Frühwirth, & Leeb 2012). A few workarounds have been proposed but they are either computationally expensive (e.g. sampling of the posterior probability distribution for Bayesian analysis, see Neudecker et al. 2012) or require a conceptually difficult implementation (Becker et al. 2012; Nisius 2014).

The issue, however, is 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 (Section 2), derive the order of magnitude of its effect using the

[a]Recombination may be performed by software in the case of intensity interferometry or heterodyne detection.
[b]Fiber Linked Unit for Optical Recombination.
[c]Infrared and Optical Telescope Array.
[d]Center for High Angular Resolution Array.
[e]Very Large Telescope Interferometer.
[f]Precision Integrated-Optics Near-infrared Imaging ExpeRiment.

modelling of a single value (Section 3), analyse in detail its effect in least squares model fitting (Section 4), and propose a simple, computer-efficient way to avoid it (Section 5).

2. Peelle's pertinent puzzle

I rewrite and adapt Peelle's original example in the context of long-baseline interferometry (see Neudecker et al. 2012, Sections 1 and 2.1). The mathematical symbols used in this article are summarised in Table 1. One or several calibrator observations yield the inverse of the instrumental fringe contrast $\tau \pm \tau \varsigma_\tau$. I use the relative uncertainty ς_τ on the transfer function as it is often referred to in percentage terms. A visibility amplitude is now estimated from two contrast measurements $v_1 \pm \sigma_v$ and $v_2 \pm \sigma_v$. For each measurement, the visibility amplitudes are

$$v_1 = \tau v_1 \pm \tau \sigma_v \, (\pm \tau v_1 \varsigma_\tau), \tag{1a}$$
$$v_2 = \tau v_2 \pm \tau \sigma_v \, (\pm \tau v_2 \varsigma_\tau), \tag{1b}$$

where the second-order error term $\tau \varsigma_\tau \sigma_v$ has been ignored.

They are normalised with the same quantity (τ), so they are correlated, hence the systematic uncertainty term between parentheses in Eq. (1). Error propagation yields the covariance matrix:

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_v^2 \tau^2 + \varsigma_\tau^2 v_1^2 & \varsigma_\tau^2 v_1 v_2 \\ \varsigma_\tau^2 v_1 v_2 & \sigma_v^2 \tau^2 + \varsigma_\tau^2 v_2^2 \end{pmatrix}. \tag{2}$$

Under the hypothesis of Gaussian errors, I obtain the least squares estimate using the weight matrix $\boldsymbol{W} = \boldsymbol{\Sigma}^{-1}$:

$$v = \frac{v_1 W_{11} + (v_1 + v_2)W_{12} + v_2 W_{22}}{W_{11} + 2W_{12} + W_{22}},$$
$$= \frac{v_1 + v_2}{2}\left(1 + \varsigma_\tau^2 \frac{(v_1 - v_2)^2}{2\tau^2 \sigma_v^2}\right)^{-1}, \tag{3}$$

with the uncertainty:

$$\sigma_V^2 = \frac{1}{W_{11} + 2W_{12} + W_{22}},$$
$$= \left[\left(\frac{\tau \sigma_v}{\sqrt{2}}\right)^2 + \frac{v_1^2 + v_2^2}{2}\varsigma_\tau^2\right]\frac{2v}{v_1 + v_2}. \tag{4}$$

The visibility estimate v systematically falls below the average of the two values v_1 and v_2. If the measurements differ significantly, it can even fall below the lowest value. Figure 1 gives such an example with an instrumental visibility of 50% and two measurements on an unresolved target:

$$\tau = 2.000 \pm 0.100 \, (\varsigma_\tau = 5\%),$$
$$v_1 = 0.495 \pm 0.003 \, (\sigma_v/v_1 \approx 0.6\%),$$
$$v_2 = 0.505 \pm 0.003 \, (\sigma_v/v_2 \approx 0.6\%),$$

which yields two points 2.4 standard deviations apart

$$v_1 = 0.990 \pm 0.006 \, (\pm 0.050),$$
$$v_2 = 1.010 \pm 0.006 \, (\pm 0.051)$$

and the visibility amplitude estimate

$$v = 0.986 \pm 0.004 \, (\pm 0.050)$$

falls outside the data range. The uncertainties quoted for v correspond to the first and second terms within the square brackets of Eq. (4).

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

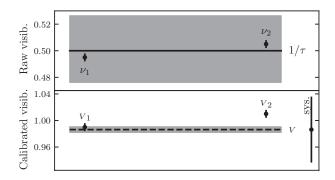| Symbol	Meaning
$\overline{a}$	True value of a
$<a>$	Expected value of a
$\boldsymbol{A}^\mathsf{T}$	Transpose of $\boldsymbol{A}$
$\boldsymbol{c} = \boldsymbol{a} \odot \boldsymbol{b}$	Element-wise product of $\boldsymbol{a}$ and $\boldsymbol{b}$
$\boldsymbol{C} = \boldsymbol{a} \otimes \boldsymbol{b}$	Outer product of $\boldsymbol{a}$ and $\boldsymbol{b}$
$\boldsymbol{c} = \boldsymbol{Ab}$	Matrix product of $\boldsymbol{A}$ and $\boldsymbol{b}$
δ	Kronecker delta
$\boldsymbol{v}$	Data
$\boldsymbol{\eta}$	Error $(= \boldsymbol{v} - \overline{\boldsymbol{v}})$
ε	Relative error $(= \eta/\overline{\boldsymbol{v}})$
σ	Deviation $(= \sqrt{<\eta^2>})$, uncertainty
ς	Relative uncertainty $(= \sqrt{<\varepsilon^2>})$
$\boldsymbol{\Sigma}$	Covariance matrix $(= <\boldsymbol{\eta} \otimes \boldsymbol{\eta}>)$
ϱ	Correlation coefficient
$\boldsymbol{x}, \boldsymbol{X}$	Sensitivity vector or matrix
$\boldsymbol{p}$	Parameters of the model
$\boldsymbol{\mu}$	Model values $(= \boldsymbol{Xp} \approx \overline{\boldsymbol{v}})$
a_v	a of the measurement error
a_τ	a of the normalisation error
$a^\star$	a impacted by PPP
$\varsigma_{\text{sys.}}$	Relative systematic uncertainty
$\varsigma_{\text{stat.}}$	Relative statistical uncertainty

Figure 1. Original Peelle problem rewritten in the context of interferometry. *Top:* Two raw visibility amplitudes v_1 and v_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%). *Bottom:* 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 systematic error on v, v_1, and v_2 is shown on the right.

3. Fit by a constant

I now generalise the results of the last section to an arbitrary number of measurements of a single normalised quantity, such as the visibility amplitude of an unresolved source, which is expected to be constantly one for all interferometric baselines. Let the column vector $v = (v_1, \cdots, v_n)^\mathsf{T}$ contain the n visibility amplitudes. It is derived from an uncalibrated quantity like the fringe contrast, $\boldsymbol{v} = (v_1, \cdots, v_n)^\mathsf{T}$ and a normalisation factor, like the cotransfer function, $\boldsymbol{\tau} = (\tau_1, \cdots, \tau_n)^\mathsf{T}$ by

$$v = \boldsymbol{\tau} \odot \boldsymbol{v}, \tag{5}$$

where $\odot$ denotes the Hadamard (element-wise) product of vectors. With $\overline{v}$, $\overline{\tau}$, and $\overline{\nu}$ the true, but unknown, values of these quantities, the error vector on v:

$$\boldsymbol{\eta} = v - \overline{v}, \tag{6}$$

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

$$\boldsymbol{\eta} = (\boldsymbol{\varepsilon}_\nu + \boldsymbol{\varepsilon}_\tau)\overline{v}. \tag{7}$$

These errors are given by:

$$\boldsymbol{\varepsilon}_\nu = \frac{1}{\overline{\nu}}(\nu - \overline{\nu}), \tag{8}$$

$$\boldsymbol{\varepsilon}_\tau = \frac{1}{\overline{\tau}}(\tau - \overline{\tau}). \tag{9}$$

I assume $\boldsymbol{\varepsilon}_\nu$ and $\boldsymbol{\varepsilon}_\tau$ are independent, of mean 0, and have standard deviations ς_ν and ς_τ, respectively. In addition, I consider correlation of the normalisation errors, with correlation coefficient ϱ. In the case of interferometry, it can arise from the uncertainty on the calibrators' geometry. The covariance matrix of the visibility amplitudes is given by:

$$\boldsymbol{\Sigma} = <\boldsymbol{\eta} \otimes \boldsymbol{\eta}>, \tag{10}$$

where $\otimes$ denotes the outer product of vectors and $<>$ stands for the expected value, so that

$$\Sigma_{ij} = \underbrace{\left[\varsigma_\nu^2 + (1-\varrho)\varsigma_\tau^2\right]}_{\varsigma_{\text{stat.}}^2}\overline{v}^2\delta_{ij} + \underbrace{\varrho\varsigma_\tau^2}_{\varsigma_{\text{sys.}}^2}\overline{v}^2. \tag{11}$$

The non-diagonal diagonal elements of the matrix feature the systematic relative uncertainty $\varsigma_{\text{sys.}}$, that is, the correlated component of the uncertainties. In the case of a fully correlated transfer function ($\varrho = 1$), it is equal its uncertainty ($\varsigma_{\text{sys.}} = \varsigma_\tau$). The diagonal term of the matrix additionally includes the statistical relative uncertainty $\varsigma_{\text{stat.}}$, that is, the uncorrelated component of the uncertainties. In the case of a fully correlated transfer function, it is equal to the uncertainty of the uncalibrated visibility ($\varsigma_{\text{stat.}} = \varsigma_\nu$).

The value $\overline{v}$ is yet to be determined, so the covariances are often derived using the measurements v in the propagation:

$$\Sigma_{ij}^\star = \varsigma_{\text{stat.}}^2 v_i^2 \delta_{ij} + \varsigma_{\text{sys.}}^2 v_i v_j. \tag{12}$$

The least squares estimate for $\overline{v}$ is given by:

$$\mu^\star = \frac{\boldsymbol{x}^{\text{T}}\boldsymbol{\Sigma}^{\star-1}\boldsymbol{v}}{\boldsymbol{x}^{\text{T}}\boldsymbol{\Sigma}^{\star-1}\boldsymbol{x}} \tag{13}$$

where $\boldsymbol{x} = (1, \cdots, 1)^{\text{T}}$ is the trivial sensitivity vector. The covariance matrix is the sum of an invertible diagonal matrix and one of rank one—see (12)—, so that the inverse is obtained using the Woodbury matrix identity:

$$\{\boldsymbol{\Sigma}^{\star-1}\}_{ij} = \frac{\delta_{ij}}{\sigma_i^2} - \frac{\varsigma_{\text{sys.}}^2 v_i v_j}{\sigma_i^2 \sigma_j^2 \left(1 + \varsigma_{\text{sys.}}^2 \sum_k \frac{v_k^2}{\sigma_k^2}\right)}, \tag{14}$$

where we have introduced the statistical (uncorrelated) component of the uncertainty on the calibrated visibilities:

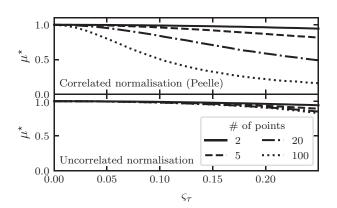$$\sigma_i = \varsigma_{\text{stat.}} v_i. \tag{15}$$

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

Appendix A.1 shows the analytical derivation for the least squares estimate $\mu^\star$ using the previous formulae. I write it in a way that highlights the generalisation of Eq. (3) of the previous section:

$$\mu^\star = \frac{\sum_i \frac{V_i}{\sigma_i^2}}{\sum_i \frac{1}{\sigma_i^2}}\left(1 + \varsigma_{\text{sys.}}^2 \frac{\sum_{i<j}\frac{(V_i - V_j)^2}{\sigma_i^2 \sigma_j^2}}{\sum_i \frac{1}{\sigma_i^2}}\right)^{-1}. \tag{16}$$

For small enough errors ($\eta_i \ll \overline{v}$), the second-order Taylor development in $\eta_i = v_i - \overline{v}$ yields (see Appendix A.2):

$$\mu^\star \approx \overline{v} + \frac{1}{n}\sum_i \eta_i - \frac{1}{\overline{v}}\left(\frac{1}{2n}\frac{\varsigma_{\text{sys.}}^2}{\varsigma_{\text{stat.}}^2} + \frac{1}{n^2}\right)\sum_{i \neq j}(\eta_i - \eta_j)^2. \tag{17}$$

Since $<\eta_i> = 0$ and $<(\eta_i - \eta_j)^2> = 2\varsigma_{\text{stat.}}^2\overline{v}^2$, the expected value:

$$<\mu^\star> \approx \overline{v}\left[1 - \left(1 - \frac{1}{n}\right)\left(2\varsigma_{\text{stat.}}^2 + n\varsigma_{\text{sys.}}^2\right)\right] \tag{18}$$

is biased. If the data are not correlated ($\varsigma_{\text{sys.}} = 0$), the bias is small ($\varsigma_{\text{stat.}}^2$ to $2\varsigma_{\text{stat.}}^2$) but it becomes larger for correlated data if the number of points is large ($n/2$ to $n \times \varsigma_\tau^2$ for fully correlated data) as D'Agostini (1994) already noted. This analytical derivation confirms the numerical simulation by (Neudecker et al. 2012, see their Figure 4). For visualisation purposes, Figure 2 shows a similar simulation of the bias as a function of the normalisation uncertainty ς_τ for various data sizes ($n = 2$ to 100). I have verified that it reproduces the quadratic behaviour of Eq. (18) for small values of $\varsigma_{\text{stat.}}$ and $\varsigma_{\text{sys.}}$ (bias inferior to 10–20% of $\overline{v}$).

The bias from PPP arises, intuitively, because the modelled uncertainty is a non-constant function of the measured value. In the present case, data that fall below the average are given a lower uncertainty and, thus, a higher weight in the least squares fit. Conversely, data that fall above the average have a higher uncertainty and a lower weight. This fundamentally biases the estimate towards lower values. The effect is much stronger with correlations because it impacts the $\sim n^2/2$ independent elements of the covariance matrix instead of being restricted to the n diagonal ones. In the literature, the puzzle is generally discussed as arising from a normalisation, as it it where it has been first identified. However, I show in Appendix A.3 that it is not necessary and determine the bias in the case of correlated photon noise.

For spectro-interferometric observations with four telescopes, the number of correlated points can be over 1 000, so even with a low correlation coefficient, the bias can be significant. For instance, a single GRAVITY observation in medium spectral resolution yields $n = 6 \times 210$ visibility amplitudes. With an observed correlation of $\varrho \approx 16\%$ in the instrumental visibility amplitudes (Kammerer et al. 2020) and a typical $\varsigma_\tau = 1\text{--}2\%$ normalisation error, the bias on the calibrated visibilities could be 2–8%.

4. General model fitting

I now consider a set of measurements corresponding to the linear model:

$$\boldsymbol{\mu} = X\boldsymbol{p}, \qquad (19)$$

where $\boldsymbol{p}$ are the unknown parameters and 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 projected on to the sky, that is, $u = B_u / \lambda$ if B is the baseline and λ, the wavelength. The true values $\overline{v}$ are impacted by errors so that the data are

$$v = \overline{v} + \boldsymbol{\eta} \qquad (20)$$

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

$$\boldsymbol{\eta} = \boldsymbol{\eta}_v + \boldsymbol{\varepsilon}_\tau \odot \overline{v}. \qquad (21)$$

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

$$\boldsymbol{p} = (X^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} X)^{-1} (X^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} v) \qquad (22)$$

I investigate four ways to determine the covariance matrix $\boldsymbol{\Sigma}$

1. Ignoring the correlations in the normalisation using $\boldsymbol{\Sigma}_0 = \boldsymbol{\Sigma}_v + (v \otimes v) \odot (\boldsymbol{\Sigma}_\tau \odot I)$. Let $\boldsymbol{\mu}_0 = X(X^{\mathrm{T}} \boldsymbol{\Sigma}_0^{-1} X)^{-1} X^{\mathrm{T}} \boldsymbol{\Sigma}_0^{-1} v$ the resulting model of the data.

2. Using the naïve estimate $\boldsymbol{\Sigma}^\star = \boldsymbol{\Sigma}_v + (v \otimes v) \odot \boldsymbol{\Sigma}_\tau$ which is known to lead to PPP in the trivial case of a constant model.

3. Using the data model of the fit without the normalisation error: $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_v + (\boldsymbol{\mu}_0 \otimes \boldsymbol{\mu}_0) \odot \boldsymbol{\Sigma}_\tau$. This is the generalisation of the two-variable approach by Neudecker et al. (2014). The resulting least squares model is $\boldsymbol{\mu}_1 = X(X^{\mathrm{T}} \boldsymbol{\Sigma}_1^{-1} X)^{-1} X^{\mathrm{T}} \boldsymbol{\Sigma}_1^{-1} v$.

4. Recursively fitting the data by updating the data model in the covariance matrix. I derive $\boldsymbol{\mu}_k = X(X^{\mathrm{T}} \boldsymbol{\Sigma}_k^{-1} X)^{-1} X^{\mathrm{T}} \boldsymbol{\Sigma}_k^{-1} v$ using $\boldsymbol{\Sigma}_k = \boldsymbol{\Sigma}_v + (\boldsymbol{\mu}_{k-1} \otimes \boldsymbol{\mu}_{k-1}) \odot \boldsymbol{\Sigma}_\tau$, starting with the estimate $\boldsymbol{\mu}_1$ ($k = 2$).

In order to compare these covariance matrix prescriptions, I will use the typical example of an under-resolved centro-symmetric source observed at a four-telescope facility in medium spectral resolution. It is close to the context under which I serendipitously noticed the effect while modelling stellar diameters (see Lachaume et al. 2019). The python code to produce the results (figures in this paper) is available on github.[g] In the under-resolved case all models—Gaussian, uniform disc, or limb-darkened disc—are equivalent (Lachaume 2003), so I will use

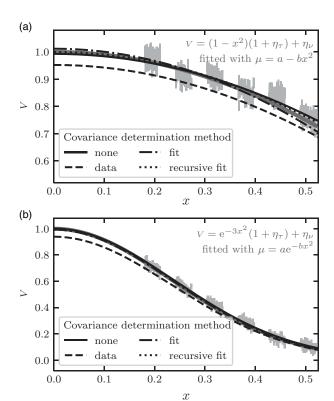[g] https://github.com/loqueelvientoajuarez/peelles-pertinent-puzzle.

Figure 3. Model fitting to simulated correlated data from a four-telescope interferometer (6 baselines) with medium spectral resolution (R = 100) with 2% uncorrelated measurement error and 3% correlated normalisation error (light grey points with the measurement error bar). *Top:* Simulated under-resolved data $v = 1 - x^2$ (thick grey line) are fitted with linear least squares model $\mu = a - bx^2$ using the four prescriptions for the covariance matrix. *Bottom:* The same for well-resolved data $v = \exp{-3x^2}$ and non-linear least squares with model $\mu = a \exp{-bx^2}$. (a) Under-resolved, linear least squares. (b) Well resolved, non-linear least squares.

instead a linear least squares fit $\mu = a - bx^2$ to $v \approx 1 - x^2$ where x is dimensionless variable proportional to the projected baseline length $\sqrt{u^2 + v^2}$. This fit corresponds to the second-order Taylor development of any of the aforementioned models. Figure 3(a) shows the example of such a fit performed for each covariance matrix prescription. Data have been simulated using $v = (1 - x^2)(1 + \varepsilon_\tau) + \eta_v$ where ε_τ is a fully correlated normalisation error (3%) and η_v are uncorrelated statistical errors (2%). As expected, the use of data v in the correlation matrix, method 2, leads to grossly underestimated data values, in the very same way as in the classical Peelle case described in Sections 2 and 3. Other methods, including ignoring correlations, yield reasonable parameter estimates.

Figure 4(a) sums up the behaviour of the same fit performed a large number of times on different simulated datasets, 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 given by the least squares optimisation routine in comparison to the scatter of the distribution of the values. While the model fitting ignoring correlations (method 1) does not show any bias on the parameter estimates, it displays a higher dispersion of model parameters, grossly underestimates the uncertainty on model parameters and has a biased chi-squared. The correlation matrix calculated from data (method 2) is, as expected, strongly biased. Both methods

(a)

(b)

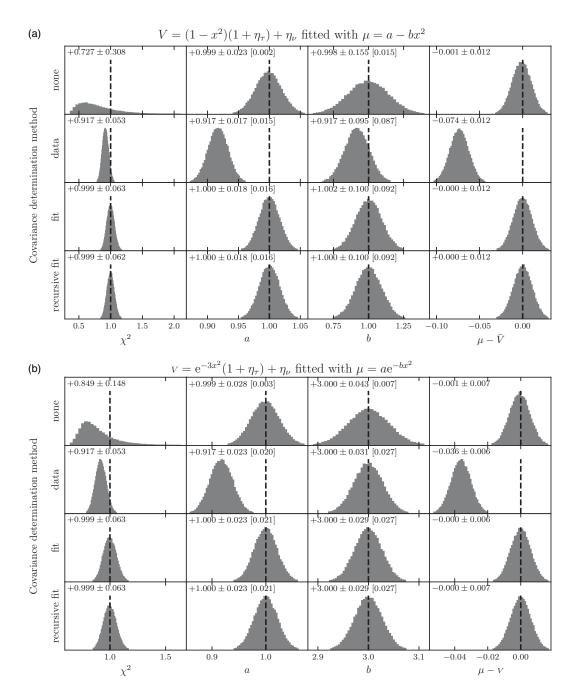**Figure 4.** Distribution of the fitted parameters and fit properties for the four covariance matrix prescriptions analysed in Sect. 4. 5×10^4 simulations of 6 groups of 100 correlated data points v (measurement error $\varsigma_v = 2\%$ and normalisation error $\varsigma_\tau = 3\%$, correlation of the latter $\varrho = 1$, normal distributions) are performed and fitted with a model using least squares minimisation. *Top graph:* under-resolved data follow $v = 1 - x^2$ and are fitted with linear least squares $\mu = a - bx^2$. *Bottom graph:* well resolved data follow $v = \exp{-3x^2}$ and are fitted with non-linear least squares $\mu = a \exp{-bx^2}$. Reported quantities include median and 1-σ interval of their distribution and, within brackets, the median uncertainty reported by the least squares fit. 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. (a) Under-resolved data fitted with a linear least squares model. (b) Well-resolved data fitted with a non-linear least squares model.

estimating the correlation matrix from modelled data (methods 3 & 4) are equivalent in terms of the absence of bias, dispersion of these quantities, and correct prediction of the uncertainty on model parameters.

Given that fitting recursively the covariance matrix does not yield additional benefits for the modelling, I would suggest to use

method 3. One would expect this to hold for any smooth enough model, as the update in the covariance matrix is expected to be a small effect. Indeed, I have checked that the result holds for a fully resolved Gaussian disc $v \approx \exp{-3x^2}$ fit by $\mu = a \exp{-bx^2}$ (see Figures 3(b) and 4(b)) a well-resolved binary, with methods 3 & 4 providing unbiased estimates and similar uncertainties. If,

for some other application, the model $\boldsymbol{\mu}_1$ obtained with method 3 were to differ significantly from the starting guess $\boldsymbol{\mu}_0$ (method 1), it would certainly make sense to examine whether recursive fitting (method 4) is needed. However, while it converged for the smooth models I tested, I have not proven that it will necessarily do so, in particular for less smooth models that may require it.

5. Conclusion

The standard covariance propagation, using the measurement values in the calculation, can result in a bias in the model parameters of a least squares fit taking correlations into account. It will occur as soon as the error bars and covariances depend on the measured values, in particular when a normalisation factor, such as the instrumental transfer function of an interferometer, is obtained experimentally. Some bias will even occur without correlations, but the effect is strongest when a large set of correlated data is modelled. This is precisely the case in optical and infrared long-baseline interferometry, where the calibration of spectrally dispersed fringes easily yields 10^2–10^3 correlated data points.

While solutions exist that are either numerically expensive or require some care to be implemented (Burr et al. 2011; Becker et al. 2012; Nisius 2014), I have shown with a simple example that there is an easy and cheap way to solve the issue. First an uncorrelated fit is performed to estimate the true values corresponding to the data. Secondly, these estimates are used to determine the covariance matrix by error propagation. At last, this covariance matrix is used to perform a least squares model fit.

Alternatively, it is possible to obtain an (almost) unbiased estimate for the model parameters by ignoring correlations altogether, with the cost of a larger imprecision, underestimated uncertainties, and a biased chi-squared. It is, at the moment, the approach taken in the vast majority of published studies in optical interferometry, as data processing pipelines of most instrument do not determine covariances. To my knowledge, Lachaume et al. (2019) is the only work where PPP has been explicitly taken care of in optical interferometry.

References

Absil, O., et al. 2006, A&A, 452, 237
Becker, B., et al. 2012, JI, 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, NIMPRA, 346, 306
Eisenhauer, F., et al. 2011, Msngr, 143, 16
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
Hummel, C. A., & Percheron, I. 2006, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series. p. 62683X, 10.1117/12.671337
Kammerer, J., Mérand, A., Ireland, M. J., & Lacour, S. 2020, A&A, 644, A110
Lachaume, R. 2003, A&A, 400, 795
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
Mayer, P. M., Rana, F., & Ram, R. J. 2003, APL, 82, 689
Millour, F., Valat, B., Petrov, R. G., & Vannier, M. 2008, in Optical and Infrared Interferometry. p. 701349 (arXiv:0807.0291), 10.1117/12.788707
Monnier, J. D. 2007, NewAR, 51, 604
Neudecker, D., Frühwirth, R., & Leeb, H. 2012, NSE, 170, 54
Neudecker, D., Frühwirth, R., Kawano, T., & Leeb, H. 2014, NDS, 118, 364
Nisius, R. 2014, EPJC, 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, I., et al. 2008, in Optical and Infrared Interferometry. p. 70131J, 10.1117/12.788871
Tatulli, E., et al. 2007, A&A, 464, 29
Thiébaut, E. 2008, in Optical and Infrared Interferometry. p. 70131I, 10.1117/12.788822

A. Analytical derivation

To shorten summations in the derivation, I introduce

$$m_l = \sum_i \frac{v_i^l}{\sigma_i^2}, \quad \text{(moments)}$$

$$d_{lk} = \sum_{i \neq j} \frac{(v_i^k - v_j^k)^l}{\sigma_i^2 \sigma_j^2}, \quad \text{(moments of differences)},$$

$$e_l = \sum_i \eta_i^l, \quad \text{(moments of errors)}$$

$$f_{lk} = \sum_{i,j} (\eta_i^k - \eta_j^k)^l \quad \text{(moment of error differences)}$$

and note that the differences can be developed as:

$$d_{21} = 2m_0 m_2 - 2m_1^2,$$
$$f_{21} = 2n e_2 - 2e_1^2,$$
$$d_{41} = 2m_0 m_4 - 8m_1 m_3 + 6m_2^2,$$
$$d_{22} = 2m_0 m_4 - 2m_2^2.$$

A.1. Equation (16)

I rewrite Equation (14) as:

$$\{\boldsymbol{\Sigma}^{\star-1}\}_{ij} = \frac{\delta_{ij}}{\sigma_i^2} - \frac{\varrho \varsigma_\tau^2 V_i V_j}{\sigma_i^2 \sigma_j^2 (1 + \varrho \varsigma_\tau^2 m_2)}.$$

I then proceed from Equation (13):

$$\mu^\star = \frac{\sum\limits_{i,j} \{\boldsymbol{\Sigma}^{\star-1}\}_{ij} V_i}{\sum\limits_{i,j} \{\boldsymbol{\Sigma}^{\star-1}\}_{ij}}$$

$$= \frac{\sum\limits_i \frac{V_i}{\sigma_i^2} - \frac{\varrho \varsigma_\tau^2}{1 + \varrho \varsigma_\tau^2 m_2} \sum\limits_{i,j} \frac{V_i^2 V_j}{\sigma_i^2 \sigma_j^2}}{\sum\limits_i \frac{1}{\sigma_i^2} - \frac{\varrho \varsigma_\tau^2}{1 + \varrho \varsigma_\tau^2 m_2} \sum\limits_{i,j} \frac{V_i V_j}{\sigma_i^2 \sigma_j^2}}$$

so that, by separating summations in i and j,

$$= \frac{m_1 - \frac{\varrho\varsigma_\tau^2 m_1 m_2}{1+\varrho\varsigma_\tau^2 m_2}}{m_0 - \frac{\varrho\varsigma_\tau^2 m_1^2}{1+\varrho\varsigma_\tau^2 m_2}}$$

$$= \frac{m_1}{m_0 + \varrho\varsigma_\tau^2(m_0 m_2 - m_1^2)}$$

$$= \frac{m_1}{m_0}\left(1 + \varrho\varsigma_\tau^2 \frac{d_{21}}{2m_0}\right)^{-1}.$$

Equation (16) is obtained by noting that summation over $i < j$ is half of that over i, j.

A.2. Equation (17)

Since d_{12} is expressed in terms of $(v_i - v_j)^2 = (\eta_i - \eta_j)^2$, the previous equation can be simplified in the second order as:

$$\mu^\star \approx \frac{m_1}{m_0} - \varrho\varsigma_\tau^2 \left.\frac{m_1}{2m_0^2}\right|_{\eta=0} d_{21}\big|_{\sigma_i=\overline\sigma} + o(||\eta||^2)$$

$$= \frac{m_1}{m_0} - \frac{\varrho\varsigma_\tau^2}{2n\varsigma_{\text{stat.}}^2 \overline{v}} f_{21} + o(||\eta||^2).$$

I now determine the Taylor development for

$$m_l = \frac{\overline{v}^{l-2}}{\varsigma_{\text{stat.}}^2} \sum_i (1 + \eta_i/\overline{v})^{l-2}$$

$$m_l \approx \frac{n\overline{v}^{l-2}}{\varsigma_{\text{stat.}}^2}\left[1 + \frac{l-2}{n\overline{v}}\sum_i \eta_i + \frac{(l-2)(l-3)}{2n\overline{v}^2}\sum_i \eta_i^2\right]$$

$$m_l \approx \frac{n\overline{v}^{l-2}}{\varsigma_{\text{stat.}}^2}\left[1 + \frac{l-2}{n\overline{v}}e_1 + \frac{(l-2)(l-3)}{2n^2\overline{v}}e_2\right],$$

The Taylor developments for m_0 and m_1 yield

$$\frac{m_1}{m_0} \approx \overline{v}\left[1 + \frac{e_1}{n\overline{v}} + \frac{2e_1^2 - 2ne_2}{n^2\overline{v}^2}\right],$$

$$\frac{m_1}{m_0} \approx \overline{v}\left[1 + \frac{e_1}{n\overline{v}} - \frac{f_{21}}{n^2\overline{v}^2}\right],$$

so that

$$\mu^\star = \overline{v} + \frac{e_1}{n} - \frac{1}{\overline{v}}\left(\frac{1}{2n}\frac{\varrho\varsigma_\tau^2}{\varsigma_{\text{stat.}}^2} + \frac{1}{n^2}\right)f_{21}.$$

A.3. PPP with photon detection

I model n photon measurements N_i of expected value $\overline{N}$ with noise uncertainty $y_i = \sqrt{N_i}$, showing correlation ϱ[h] under the assumption of Gaussian errors ($\overline{N} \gg 1$). The statistical component of the uncertainty is given by $\sigma_i^2 = (1-\varrho)N_i$. The correlation matrix is

$$\Sigma_{ij}^\star = \sigma_i^2\delta_{ij} + \varrho y_i y_j$$

and its (Woodbury) inverse:

$$\{\Sigma^{\star-1}\}_{ij} = \frac{\delta_{ij}}{\sigma_i^2} - \frac{\varrho y_i y_j}{\sigma_i^2\sigma_j^2(1+\varrho m_2')},$$

---

[h] Correlation in the photon noise is a quantum effect detected in particular experimental setting such as coupled lasers (e.g. Mayer, Rana, & Ram 2003). In astronomy, intensity interferometry makes use of these correlations.

with the moments m_l', d_{lk}', etc., defined with respect to y_i, while m_l, d_{lk}, etc., are defined with respect to N_i.

The least squares estimate for the number of photons is given by:

$$\mu^\star = \frac{x^{\mathrm{T}}\Sigma^{\star-1}N}{x^{\mathrm{T}}\Sigma^{\star-1}x}.$$

and, by using $N_i = y_i^2$,

$$\mu^\star = \frac{\sum_{i,j}\{\Sigma^{\star-1}\}_{ij}y_i^2}{\sum_{i,j}\{\Sigma^{\star-1}\}_{ij}}$$

$$\mu^\star = \frac{\sum_i \frac{y_i^2}{\sigma_i^2} - \frac{\varrho}{1+\varrho m_2'}\sum_{i,j}\frac{y_i^3 y_j}{\sigma_i^2\sigma_j^2}}{\sum_i \frac{1}{\sigma_i^2} - \frac{\varrho}{1+\varrho m_2'}\sum_{i,j}\frac{y_i y_j}{\sigma_i^2\sigma_j^2}}$$

so that, separating summations along i and j,

$$\mu^\star = \frac{m_2' - \frac{\varrho m_1' m_3'}{1+\varrho m_2'}}{m_0' - \frac{\varrho m_1'^2}{1+\varrho m_2'}},$$

$$\mu^\star = \frac{m_2' + \varrho(m_2'^2 - m_1' m_3')}{m_0' + \varrho(m_0' m_2' - m_1'^2)},$$

$$\mu^\star = \frac{m_2'}{m_0'}\left(1 - \varrho\frac{d_{22}' - d_{41}'}{8m_2'}\right)\left(1 + \varrho\frac{d_{21}'}{2m_0'}\right)^{-1}$$

or, more explicitly,

$$\mu^\star = \frac{n}{\sum_i N_i^{-1}}\frac{1 - \frac{\varrho}{2n(1-\varrho)}\sum_{i,j}\frac{(\sqrt{N_i}-\sqrt{N_j})^2}{\sqrt{N_i N_j}}}{1 + \frac{\varrho}{2(1-\varrho)\sum_i N_i^{-1}}\sum_{i,j}\frac{(\sqrt{N_i}-\sqrt{N_j})^2}{N_i N_j}}.$$

In the second order in η, it can be simplified to

$$\mu^\star = \frac{n}{\sum_i N_i^{-1}}\left(1 - \frac{\varrho}{2n\overline{N}(1-\varrho)}\sum_{i,j}(\sqrt{N_i} - \sqrt{N_j})^2\right)$$

and, by noting that $\sqrt{N_i} - \sqrt{N_j} = (N_i - Nj)/(\sqrt{N_i} + \sqrt{N_j}) \approx (\eta_i - \eta_j)/(2\sqrt{N})$,

$$\mu^\star = \frac{n}{\sum_i N_i^{-1}}\left(1 - \frac{\varrho}{4n\overline{N}^2(1-\varrho)}\sum_{i,j}(\eta_i - \eta_j)^2\right).$$

The leading factor can be approximated in the second order using the Taylor series:

$$\sum_i N_i^{-1} \approx \frac{n}{\overline{N}} - \frac{e_1}{\overline{N}^2} + \frac{e_2}{\overline{N}^3}$$

so that

$$\frac{n}{\sum_i N_i^{-1}} \approx \overline{N} + \frac{e_1}{n} + \frac{e_1^2 - ne_2}{n^2\overline{N}},$$

$$\frac{n}{\sum_i N_i^{-1}} \approx \overline{N} + \frac{e_1}{n} - \frac{f_{21}}{2n^2\overline{N}}.$$

Finally,

$$\mu^{\star} \approx \overline{N} + \frac{\sum_i \eta_i}{n} - \left(1 + \frac{n\varrho}{2(1-\varrho)}\right) \frac{\sum_{i,j} (\eta_i - \eta_j)^2}{2Nn^2}.$$

With $< (\eta_i - \eta_j)^2 >= 2(1-\varrho)\overline{N}$ and $< \eta_i >= 0$, the bias of the best-fit estimate for the average number of photons is

$$< \mu^{\star} > \approx \overline{N} - \left(1 - \frac{1}{n}\right) \left((1-\varrho) + \frac{n\varrho}{2}\right)$$

or, with the relative statistical and systematic uncertainties $\varsigma^2_{\text{stat.}} = (1-\varrho)/\overline{N}$ and $\varsigma^2_{\text{sys.}} = \varrho/\overline{N}$,

$$< \mu^{\star} > \approx \overline{N} \left[1 - \left(1 - \frac{1}{n}\right) \left(\varsigma^2_{\text{stat.}} + \frac{n}{2} \varsigma^2_{\text{sys.}}\right)\right].$$

The bias from PPP is exactly half of that determined for normalisation errors in the main part of the paper. It shows that the effect does not necessarily arise from a normalisation.