Fitting correlated interferometric data

Régis Lachaume

January 25, 2021

Abstract

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

1 Introduction

Some widely used long-baseline interferometric observables, like the visibility amplitudes, are usually obtained by normalisation: A set of fringe contrasts obtained on a science target are divided by the instrumental transfer function determined on calibrator stars. The uncertainty in the transfer function results in correlations in the visibilities, as Perrin (2003) thoroughly examined. In the context of experimental nuclear physics, Peelle (1987) noted that such correlated measurements could provide an estimate falling below the individual data points, a paradox known as Peelle's Pertinent Puzzle (PPP).

2 Normalisation by a transfer function

2.1 Reframing Peelle's Pertinent Puzzle

In this introduction, we rewrite and adapt the 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:

$$V_1 = \tau\nu_1 \pm\ \tau\sigma_\nu\ (\pm\tau\nu_1\sigma_\tau), \tag{1a}$$

$$V_2 = \tau\nu_2 \pm\ \tau\sigma_\nu\ (\pm\tau\nu_2\sigma_\tau). \tag{1b}$$

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

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_\nu^2\tau^2 + \sigma_\tau^2 V_1^2 & \sigma_\tau^2 V_1 V_2 \\ \sigma_\tau^2 V_1 V_2 & \sigma_\nu^2\tau^2 + \sigma_\tau^2 V_2^2 \end{pmatrix}, \tag{2}$$

when the second-order error term $(\sigma_\nu\sigma_\tau)$ is ignored. Using the matrix inverse $\boldsymbol{\Sigma}^{-1}$, the least squares estimate is

$$\begin{aligned} V &= \frac{(\boldsymbol{\Sigma}_{11}^{-1} + \boldsymbol{\Sigma}_{12}^{-1})V_1 + (\boldsymbol{\Sigma}_{22}^{-1} + \boldsymbol{\Sigma}_{12}^{-1})V_2}{(\boldsymbol{\Sigma}_{11}^{-1} + \boldsymbol{\Sigma}_{12}^{-1})\ \ + (\boldsymbol{\Sigma}_{22}^{-1} + \boldsymbol{\Sigma}_{12}^{-1})}, \\ &= \frac{V_1 + V_2}{2}\left(1 + \sigma_\tau^2\frac{(V_1 - V_2)^2}{2\tau^2\sigma_\nu^2}\right)^{-1}. \end{aligned} \tag{3}$$

It systematically falls below the average of the two values V_1 and V_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:

$$\tau = 2.000 \pm 0.100\ (5\%),$$
$$\nu_1 = 0.495 \pm 0.003\ (\approx 0.6\%),$$
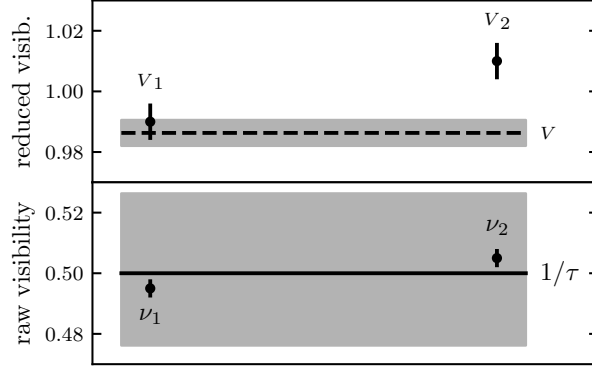$$\nu_2 = 0.505 \pm 0.003\ (\approx 0.6\%),$$

Figure 1: Original Peelle problem rewritten in the context of interferometry. *Bottom:* Two raw visibility amplitudes ν_1 and ν_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.

which yields two points 2.4 standard deviations apart

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

and the estimate

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

falls outside the data range.

See Peelle (1987); D'Agostini (1994); Burr et al. (2011); Neudecker et al. (2012, 2013, 2014); Nisius (2014).

2.2 Single value fit

We consider a set of measurements of a single normalised quantity, like the visibility amplitude, given by a vector $\boldsymbol{v} = (v_1, \cdots, v_n)^{\mathrm{T}}$. It is derived from an uncalibrated quantity like the fringe contrast,

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
$< a >$	expectancy of a
$\boldsymbol{A}^{\mathrm{T}}$	transpose of $\boldsymbol{A}$
$\boldsymbol{c} = \boldsymbol{a} \odot \boldsymbol{b}$	elementwise product of $\boldsymbol{a}$ and $\boldsymbol{b}$
$\boldsymbol{C} = \boldsymbol{a} \otimes \boldsymbol{b}$	outer product of $\boldsymbol{a}$ and $\boldsymbol{b}$
$\boldsymbol{c} = \boldsymbol{A}\boldsymbol{b}$	matrix product of $\boldsymbol{A}$ and $\boldsymbol{b}$
δ	Kronecker delta
$\boldsymbol{v}$	data
$\boldsymbol{\eta}$	error $(= \boldsymbol{v} - \overline{\boldsymbol{v}})$
σ^2	variance $(= < \eta^2 >)$
$\boldsymbol{\Sigma}$	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 $(= \boldsymbol{x}\boldsymbol{p} \approx \overline{\boldsymbol{v}})$
a_ν	a of the measurement error
a_τ	a of the normalisation error
$a^\star$	a impacted by PPP

$\boldsymbol{\nu} = (\nu_1, \cdots, \nu_n)^{\mathrm{T}}$ and a normalisation factor, like the cotransfer function, $\boldsymbol{\tau} = (\tau_1, \cdots, \tau_n)^{\mathrm{T}}$ by

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

where $\odot$ denotes the Hadamard (elementwise) product of vectors. We note $\overline{v}$, $\overline{\tau}$, and $\overline{\nu}$ the true, but unknown, values of these quantities. The error vector on $\boldsymbol{v}$

$$\boldsymbol{\eta} = \boldsymbol{v} - \overline{V}, \qquad (5)$$

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

$$\boldsymbol{\eta} = \boldsymbol{\eta}_{\boldsymbol{\nu}} + \overline{v}\boldsymbol{\eta}_{\boldsymbol{\tau}}. \qquad (6)$$

These errors are given by

$$\boldsymbol{\eta}_{\boldsymbol{\nu}} = \overline{\tau}(\boldsymbol{\nu} - \overline{\nu}), \qquad (7)$$
$$\boldsymbol{\eta}_{\boldsymbol{\tau}} = \frac{1}{\overline{\tau}}(\boldsymbol{\tau} - \overline{\tau}). \qquad (8)$$

2

We assume $\boldsymbol{\eta_\nu}$ and $\boldsymbol{\eta_\tau}$ are independent error terms of mean 0 and standard deviations are σ_ν and σ_τ, respectively. In addition, we 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 is given by

$$\boldsymbol{\Sigma} = <\boldsymbol{\eta} \otimes \boldsymbol{\eta}>, \qquad (9)$$

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

$$\Sigma_{ij} = \left[\sigma_\nu^2 + (1-\varrho)\sigma_\tau^2\overline{V}^2\right]\delta_{ij} + \varrho\sigma_\tau^2\overline{V}^2. \qquad (10)$$

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

$$\Sigma_{ij}^\star = \left[\sigma_\nu^2 + (1-\varrho)\sigma_\tau^2 V_i^2\right]\delta_{ij} + \varrho\sigma_\tau^2 V_i V_j. \qquad (11)$$

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

$$\mu^\star = \frac{\boldsymbol{x}^\mathrm{T}\boldsymbol{\Sigma}^{\star-1}\boldsymbol{V}}{\boldsymbol{x}^\mathrm{T}\boldsymbol{\Sigma}^{\star-1}\boldsymbol{x}} \qquad (12)$$

where $\boldsymbol{x} = (1, \cdots, 1)^\mathrm{T}$ is the trivial sensitivity vector. The covariance matrix is the sum of an invertible diagonal matrix and one of rank one (see Eq. (11)), so that the inverse is obtained using the Woodbury matrix identity:

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

where

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

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

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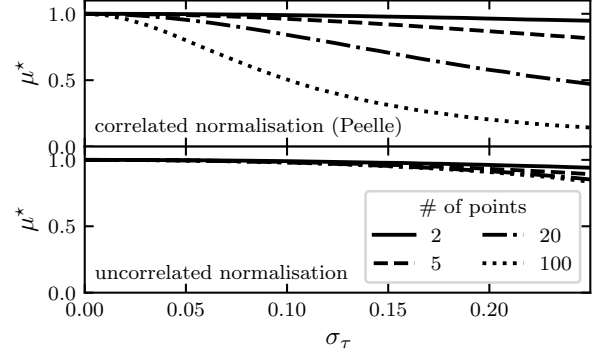

Figure 2: Fit $\mu^\star$ to unresolved visibilities ($V = 1$), as a function of the relative uncertainty on the calibration σ_τ and the number of measurements n. $2/n\times10^5$ simulations were made and averaged, assuming that η_ν and η_τ follow normal distributions. *Top:* fully correlated normalisation like in original Peelle's puzzle ($\sigma_\nu = 0.02$ and $\varrho = 1$). *Bottom:* normalisation error without correlation ($\sigma_\nu = 0.02$ and $\varrho = 0$).

For small enough errors ($\eta_i \ll \overline{V}$) the second-order Taylor development in $\eta_i = V_i - \overline{V}$ yields

$$\mu^\star \approx \overline{V} + \frac{1}{n}\sum_i \eta_i - \frac{[2 + (n-2)\varrho]\sigma_\tau^2\overline{V}^2}{2n^2\overline{\sigma}^2}\sum_{i\neq j}(\eta_i - \eta_j)^2 \qquad (16)$$

where

$$\overline{\sigma}^2 = \sigma_\nu^2 + (1-\varrho)\sigma_\tau^2\overline{V}^2 \qquad (17)$$

Since $<\eta_i> = 0$ and $<(\eta_i - \eta_j)^2> = 2\overline{\sigma}^2$, the expectancy

$$<\mu^\star> \approx \overline{V}\left[1 - \left(1 - \frac{1}{n}\right)(2 + (n-2)\varrho)\sigma_\tau^2\right] \qquad (18)$$

is biased. If the data are not correlated ($\varrho = 0$), the bias is small (σ_τ^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). Figure 2 shows a simulation of the bias as a function of the normalisation uncertainty σ_τ for various data sizes ($n = 2$ to 100).

2.3 Model fitting

We now consider a set of measurements corresponding to the linear model

$$\boldsymbol{\mu} = \boldsymbol{X}\boldsymbol{p}, \tag{19}$$

where $\boldsymbol{p}$ are the unknown parameters and $\boldsymbol{X}$ 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 that we choose to approximate by a linear one close to a solution. (u, v) is the reduced baseline. The true values $\overline{\boldsymbol{v}}$ are are impacted by errors so that the data are

$$\boldsymbol{v} = \overline{\boldsymbol{v}} + \boldsymbol{\eta} \tag{20}$$

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

$$\boldsymbol{\eta} = \boldsymbol{\eta_\nu} + \boldsymbol{\eta_\tau} \odot \overline{\boldsymbol{v}}. \tag{21}$$

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

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

We investigate four ways to determine the covariance matrix

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

2. Using the naïve estimate $\boldsymbol{\Sigma^\star} = \boldsymbol{\Sigma_\nu} + (\boldsymbol{v} \otimes \boldsymbol{v}) \odot \boldsymbol{\Sigma_\tau}$ which is known to lead to Peelle's pertinent puzzle in the trivial case of a constant model.

3. Using the data model of the fit without the normalisation error: $\boldsymbol{\Sigma_1} = \boldsymbol{\Sigma_\nu} + (\boldsymbol{\mu_0} \otimes \boldsymbol{\mu_0}) \odot \boldsymbol{\Sigma_\tau}$. The resulting least squares model is $\boldsymbol{\mu_1} = \boldsymbol{X}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{\Sigma_1}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{\Sigma_1}^{-1}\boldsymbol{v}$.
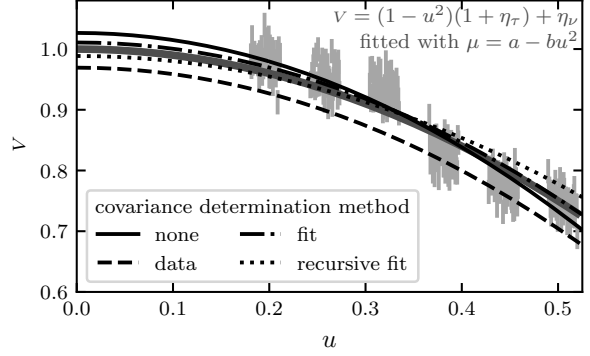

Figure 3: Example of a fit of correlated data from a four telescope interferometer with medium resolution: 6 groups of 100 data points (light gray points with the measurement error bar) with 2% uncorrelated measurement errors and 3% correlated normalisation error follow model $1 - u^2$ (thick gray line).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

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

Figure 3 shows the example of a quadratic fit to partially correlated data with the four methods above. As expected, the use of data $\boldsymbol{v}$ in the correlation matrix (method 2) leads to grossly underestimated data values, in the very same way as in the classical Peelle case (constant model). Both method 3 and 4 yield reasonable estimates.

3 Pixel-to-visibility matrix

I call I_p the luminous intensity on detector part p $(1 \cdots P)$. The scalar interferometric observable V_o for observable o $(1 \cdots O)$ is deduced using the pixel-to-visibility matrix M_{op} by

$$\boldsymbol{V} = \boldsymbol{M}\boldsymbol{I}, \tag{23}$$

where M and I can be written as

$$I = \overline{I}(1 + \boldsymbol{\eta^I}), \tag{24}$$

$$M = \overline{M}(1 + \boldsymbol{\eta^M}) \tag{25}$$

The covariance matrix Σ^V can be obtained from corralation matrices

$$\boldsymbol{\rho^I} = <\boldsymbol{\eta^M} \otimes \boldsymbol{\eta^I}> \tag{26}$$

$$\boldsymbol{\rho^M} = <\boldsymbol{\eta^M} \otimes \boldsymbol{\eta^M}> \tag{27}$$

as

$$\Sigma_{oo'}^V = \sum_{1 \leq p,p' \leq P} \left[\left(\rho_{pp'}^I + \rho_{opop'}^M - \rho_{pp'}^I \rho_{opop'}^M \right) \times \overline{I}_p \overline{I}_{p'} \overline{M}_{op} \overline{M}_{o'p'} \right], \tag{28}$$

assuming that M and I are not correlated.

Once again the true value $\overline{I}$ is unknown and replacing it by the observed value I may lead to bias.

4 Conclusion

1. 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.

2. 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.

3. 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.

4. 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

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

Neudecker D., Frühwirth R., Leeb H., 2012, Nuclear Science and Engineering, 170, 54

Neudecker D., Capote R., Leeb H., 2013, Nuclear Instruments and Methods in Physics Research A, 723, 163

Neudecker D., Frühwirth R., Kawano T., Leeb H., 2014, Nuclear Data Sheets, 118, 364

Nisius R., 2014, European Physical Journal C, 74, 3004

Peelle R. W., 1987, Informal memorandum, Peelle's Pertinent Puzzle. Oak Ridge National Laboratory
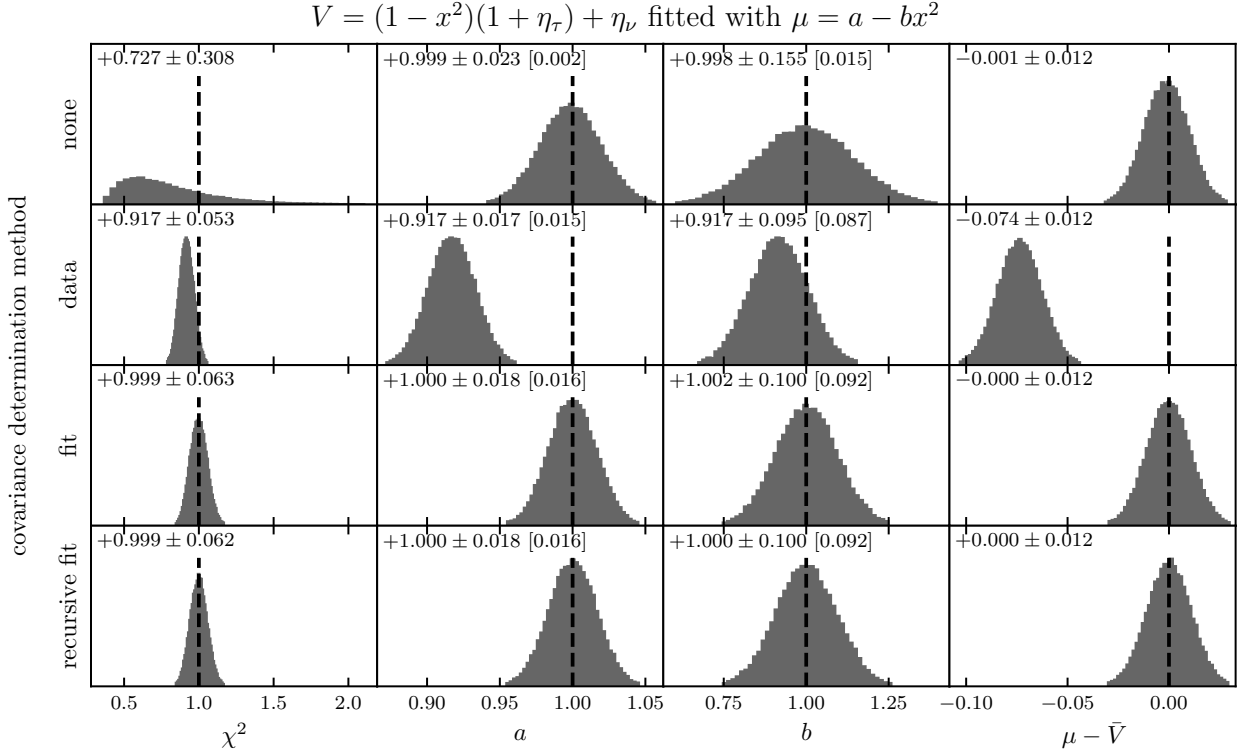
Perrin G., 2003, A&A, 400, 1173

Figure 4: Distribution of the fitted parameters and fit properties for the four covariance matrix prescriptions. 5×10^4 simulations of 6 groups of 100 correlated data points v ($\sigma_\nu = 2\%$, $\sigma_\tau = 3\%$, $\varrho = 1$, normal distributions) following model $\mu = 1 - u^2$ are performed and fitted with $a - bu^2$ using least-squares minimisation. These groups of points are typical of a single snapshot obtained with a four-telescopes interferometer with medium spectral resolution. Reported quantities include median and 1-σ 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:* quadadratic 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.