Statistical techniques for interferometric signal analysis

Donata Bonino[1], Mario Gai[1], Laura Sacerdote[2]

[1] INAF - Osservatorio Astronomico di Torino
[2] Università degli Studi di Torino, Dip. di matematica

7 maggio 2009

Summary

- **Detailed interferometric models**
  - algorithms needed for *scientific data reduction*, but also for *fringe tracking*: based on interferometric models
  - models derived from theory, but adapted with real data measurements need for methods of data analysis

- **Statistical approach: time series analysis**
  - factorization of signal structures for photometry and interferometry

- **Statistical approach: noise identification**
  - identification of appropriate tools: *regression analysis*
  - noise identified as the variability on interferometric outputs not explained by photometric inputs
  - model validated with the residuals analysis

VLTI Fringe Trackers: FINITO e PRIMA FSU

ESO VLTI (European Southern Observatory
Very Large Telescope Interferometer)
Cerro Paranal, Chile

FINITO (Fringe-tracking Instrument of Nice and TOrino)

PRIMA FSU (Phase Referenced Imaging and Micro-arcsecond Astrometry Fringe Sensor Unit)

Collaboration between **ESO** and the **Astronomical Observatory of Turin**

Fringe sensing: goals and algorithms

Measurement of fringe parameters: differential optical path (OPD), the group delay (GD), visibility

FINITO (demodulation, correlation with a template, ABCD):

signal model: $$y(OPD) = \frac{1}{2}(I_1 + I_2) \cdot [1 + V \cdot \sin(\frac{2\pi}{\lambda} \cdot OPD + \phi)]$$

λ	wavelength	V	visibility
I_1, I_2	source intensities	ϕ	phase

PRIMA (least squares method, ABCD):

signal model: $$y(OPD) = \int s(\lambda, OPD)\, d\lambda$$

$$y(OPD) = 2 \cdot A \int F(\lambda) \cdot \tau(\lambda) \cdot QE(\lambda) \cdot T \cdot [1 + V(\lambda) \cdot \sin(\frac{2\pi}{\lambda}[n \cdot OPD + (n - n_0) \cdot p] + \phi)]\, d\lambda$$

λ	wavelength	T	exposure time
$F(\lambda)$	source flux	$V(\lambda)$	source visibility
A	area of each single aperture	$\varphi(\lambda)$	instrumental phase
$\tau(\lambda)$	transmission factor	$n(\lambda)$	air refractive index, $n_0 = n(\lambda_0)$
$QE(\lambda)$	detector efficiency	p	air path

Model's parameters definition: calibration

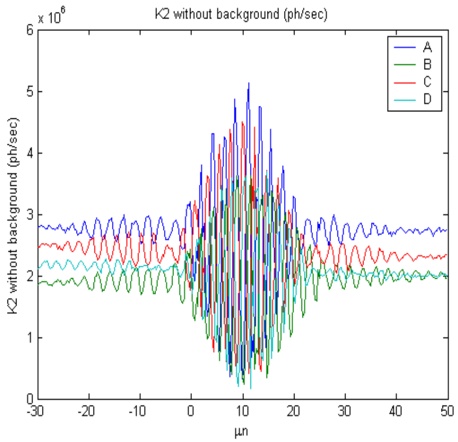
Essential parameters of the model have to be estimated from *calibration* measurements

Example of calibration procedure (to be done on a bright reference star):
PRIMA FSU calibration algorithm

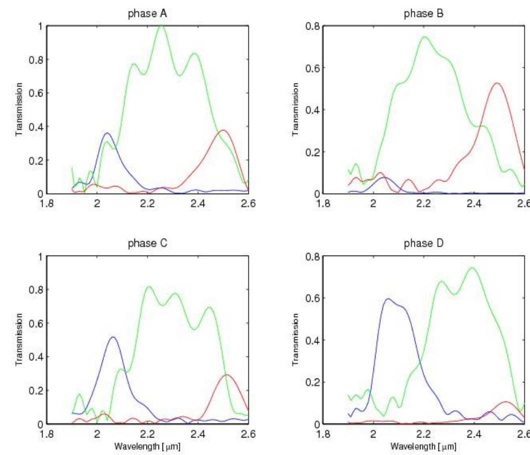
- Definition of the OPD parameters (step, modulation amplitude and so on)
- Fourier Transform of the modulated component of the measured signal
- Subtraction of the modelled source spectrum
- Estimation of the effective wavelength
- Estimation of the effective source magnitude
- Estimation of the overall visibility
- Estimation of the noise on the intensity
- Template construction

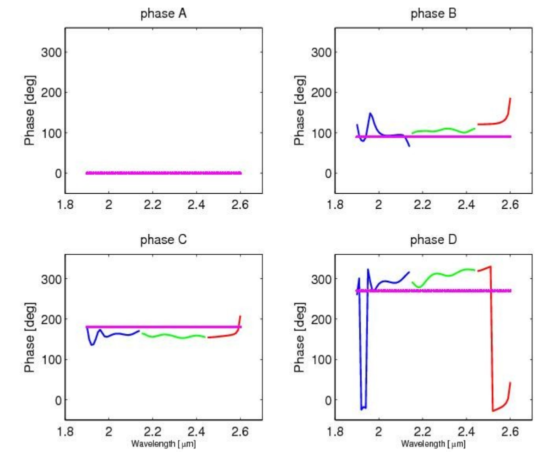
Model's parameters definition: calibration

1) Raw data
(PRIMA K2 in figure)

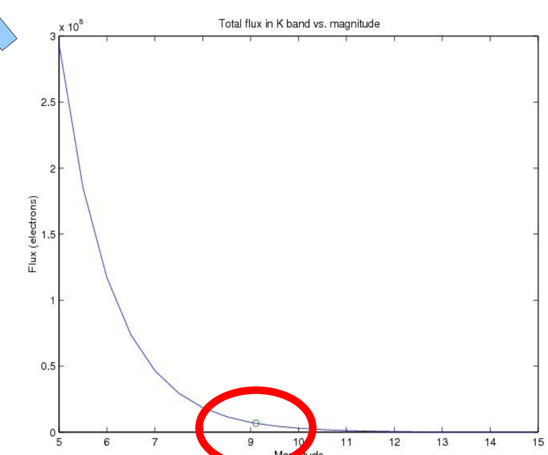


2) Instrumental transmission and phase, with subtraction of the source spectrum





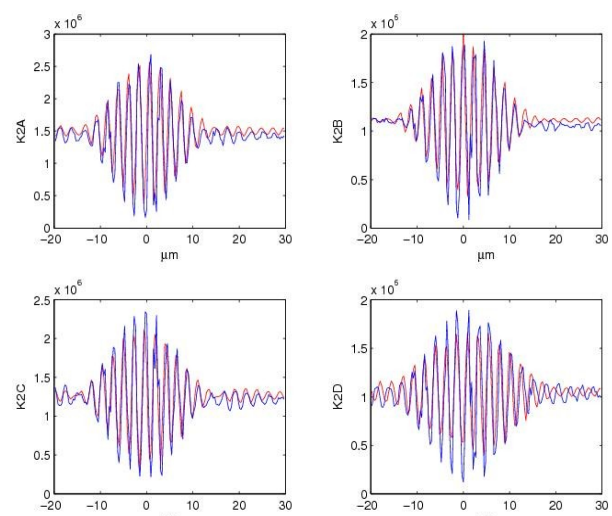
3) Magnitude estimation

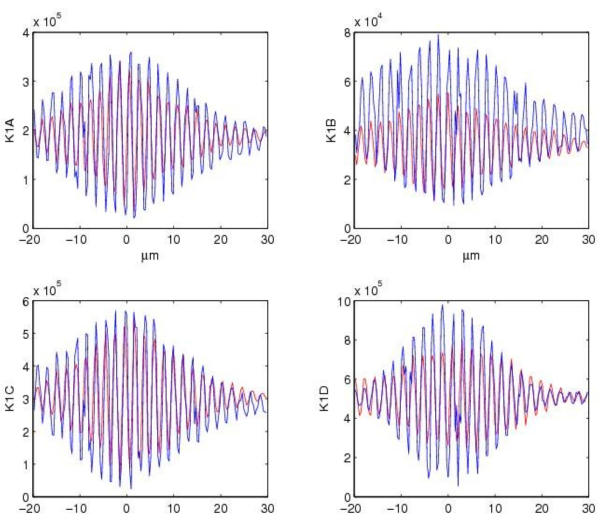


	phase A	phase B	phase C	phase D
K1	0.891	0.789	0.920	0.895
K2	0.884	0.916	0.829	0.879
K3	0.794	0.840	0.740	0.640

TABLE 5: Michelson visibilities from calibration data

4) Visibility estimation





5) Template construction, noise added
(blue: raw signal
red: reconstructed)

a 2009

Technological to mathematical link

The **spectral distribution** reconstruction is good:

effective central wavelength ⟷ fringe period

spectral band ⟷ coherence length

There are still some difficulties in reproducing the **intensity levels**, especially for low fluxes (e.g. narrow spectral bands)

⇒ There are phenomena on the signal that our model does not includes, at least not to full satisfaction, and our procedures do not completely describe.

What kind of analysis can we use for their characterisation?

STATISTICAL TECHNIQUES: inference approaches can take advantage of the availability of interferometric data, even if they are not homogeneous.

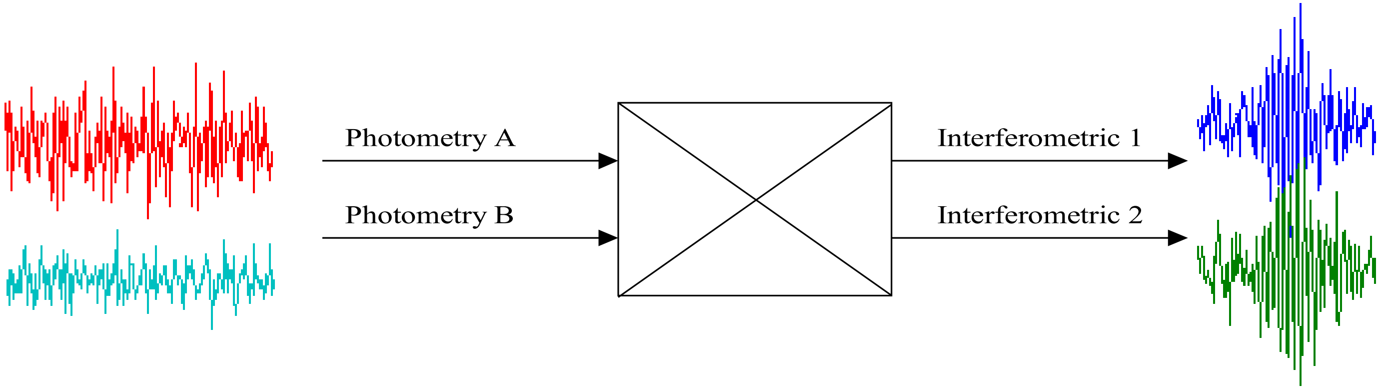
Statistical analysis

Signal analysis

time series techniques → signal components (trends, residual processes...)

Interferometric process: inputs and outputs analysis

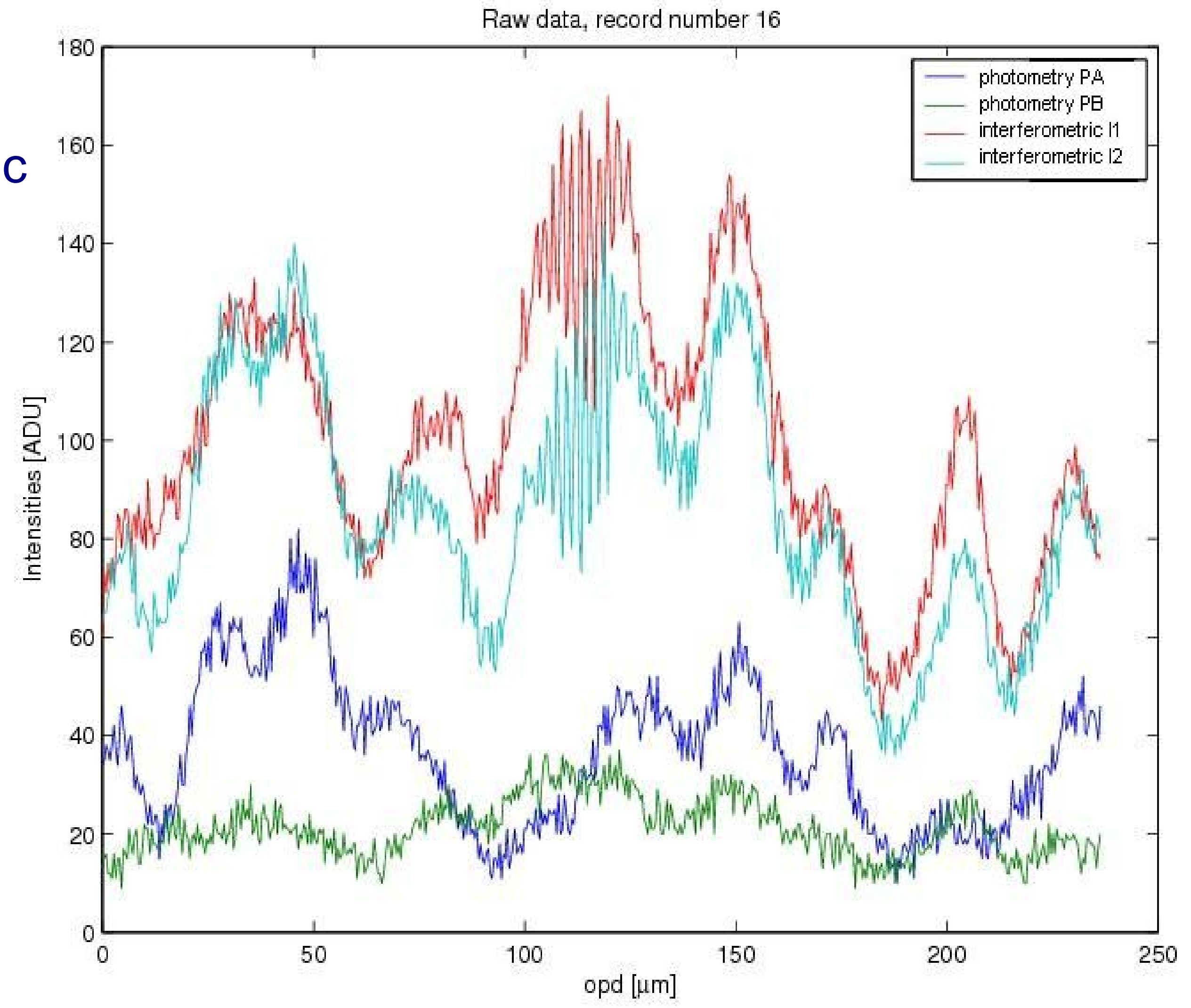
regression technique → regression coefficients, unexplained variability, residuals analysis



Statistical approach

DATA: interferometric instrument: VLTI VINCI

✗ Synchronous recording of two photometric and two interferometric channels
✗ K band
✗ No real-time OPD correction
✗ Observational data
✗ Adequate sample dimension



Data organization:
500 observational records
300 calibration records
Each record contains 526 samples

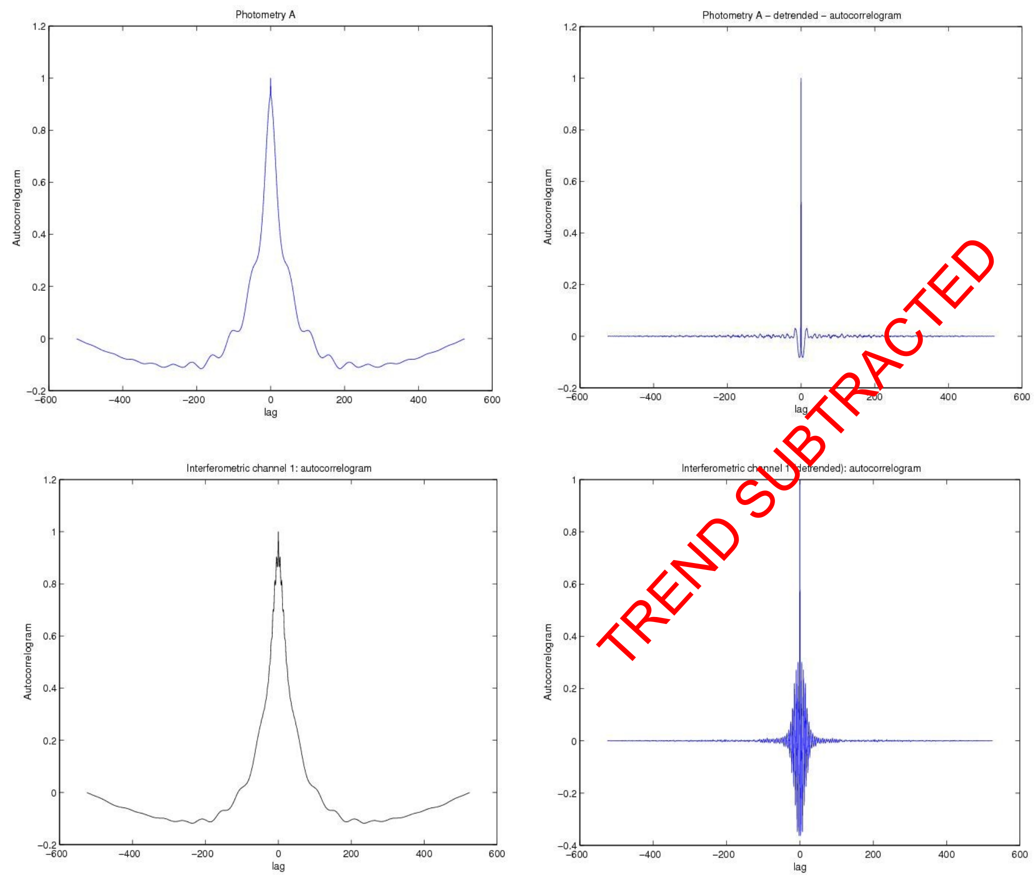
Statistical analysis in time domain

Data show a trend, changing in time >> phenomenon to be carefully handled in the analysis

Sample autocorrelation function (correlogram) before and after the trend subtraction

photometric channels

interferometric channels

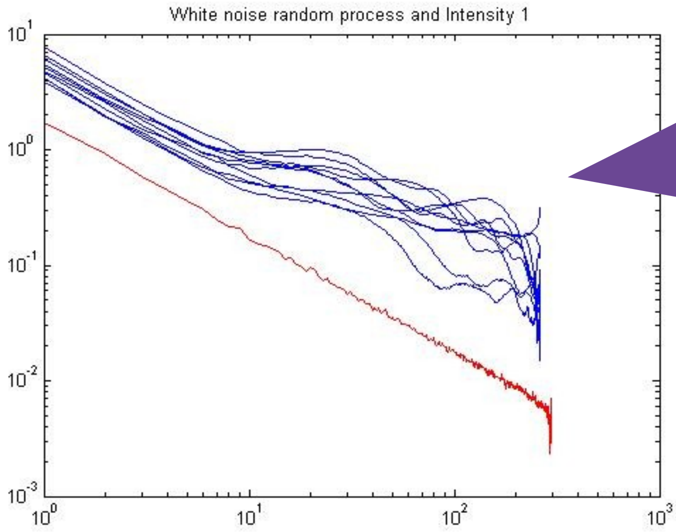


SIGNAL = TREND + UNCORRELATED PROCESS + MODULATION FUNCTION

Frequency domain analysis confirms the result

Statistical analysis in the frequency domain

Allan variance tool

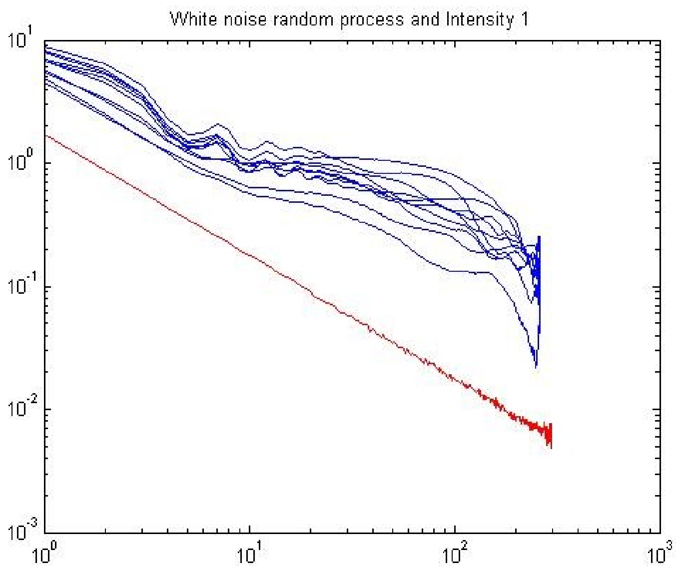


Photometric channels

Two superposed structures:
local behaviour as *white noise* (over 10 samples, i.e. 6.9 msec, equivalently 4.5 μm, ~ 2 fringes), then *trend* dominates.

Interferometric channels

No evidence of sample stability.



Work to be done:

✗ Identification of additional noise sources
✗ Investigation of other tools, like dynamic variances, that could identify *when* specific patterns are dominant

Least squares regression requirements

Discrete sampling... $$X_i = b_0 + b_j R_{i,j} + \varepsilon_i\,, \quad j = 1 \ldots p$$

Under the assumptions:

✓ $\varepsilon_i \sim N(0, \sigma^2)$

✓ regressors and dependent variables measured without errors

⇒ the **least squares estimators of the regression coefficients are the best among all possible unbiased estimators**, in the minimum variance sense.

Otherwise...

✗ Every violation of assumptions causes loss of estimators goodness (variance, bias)

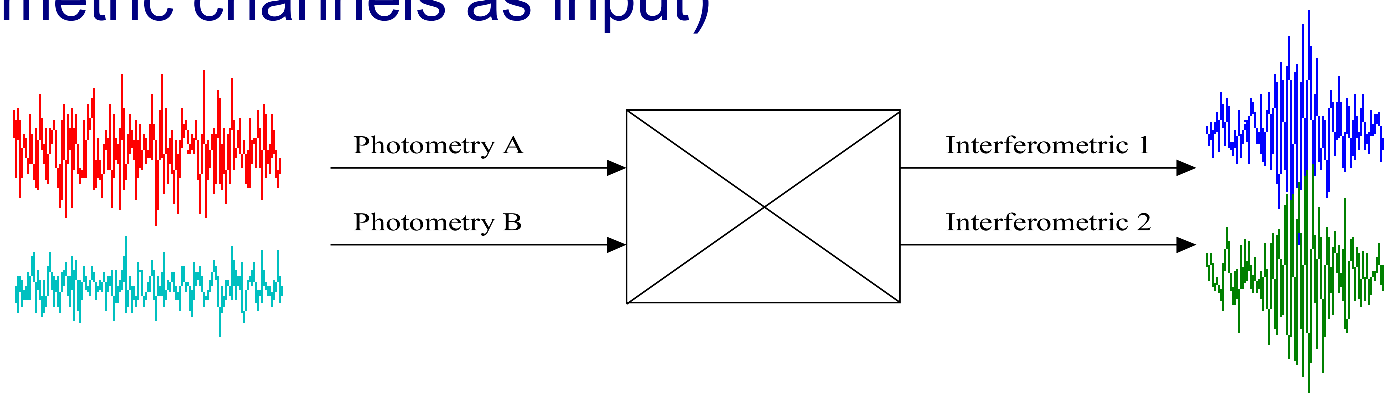
QUANTITIES OF INTEREST

regression coefficients values
multiple correlation coefficient (R^2)
residuals
statistical tests for validation

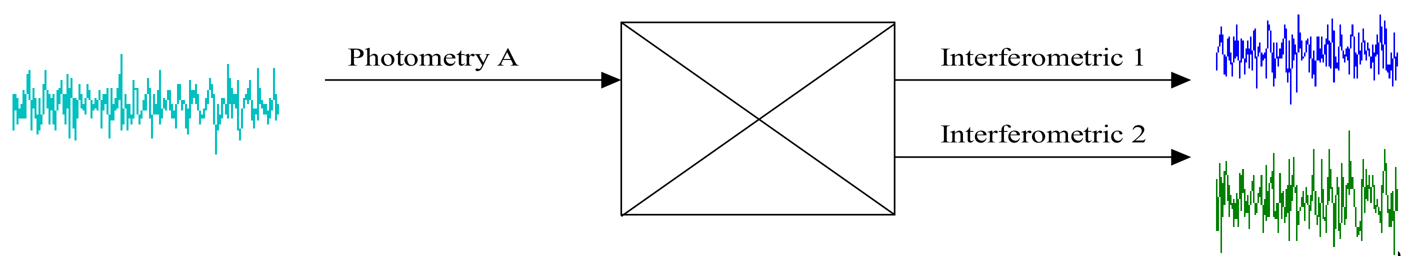
Case of homogeneous data: regression analysis steps

1) search for data with homogeneous variance (outside the coherence length)

2) regression analysis of observational data – outside the coherence length (two photometric channels as input)



3) regression analysis of calibration data (one photometric channel as input)



Test for variance homogeneity

Test di Levene di Omogeneità delle Varian
Effetto: nessuno/a
Gradi di libertà per tutte le F: 40, 366

	MS Effetto	MS Errore	F	p
Rec1	5,48930	4,002667	1,371410	0,072270
Rec2	4,18425	3,255169	1,285418	0,122354
Rec3	3,30886	2,915763	1,134817	0,271559
Rec4	6,27469	3,833693	1,636721	0,010928
Rec5	6,04472	4,354737	1,388080	0,064906
Rec6	8,10095	6,606541	1,226202	0,170798
Rec7	4,16468	2,668749	1,560535	0,019491
Rec8	2,27471	1,606271	1,416142	0,053963
Rec9	1,72487	1,466091	1,176510	0,221547
Rec10	2,31144	1,886228	1,225432	0,171512
Rec11	7,46742	6,162318	1,211788	0,184537
Rec12	7,01998	5,345805	1,313174	0,103775
Rec13	4,32227	4,711770	0,917335	0,617238
Rec14	3,76951	3,252774	1,158860	0,241890
Rec15	7,27700	4,032547	1,804566	0,002806
Rec16	5,56673	3,352169	1,660636	0,009065
Rec17	2,41027	2,013528	1,197038	0,199423
Rec18	2,19389	2,663221	0,823772	0,769513
Rec19	3,84950	3,733527	1,031062	0,423376
Rec20	2,73215	4,189967	0,652070	0,950386
Rec21	3,75907	4,284449	0,877375	0,684928
Rec22	6,66293	4,690627	1,420477	0,052424
Rec23	7,30231	5,829225	1,252707	0,147562
Rec24	8,01466	3,405601	2,353375	0,000018
Rec25	3,27652	2,160827	1,516329	0,026929
Rec26	1,59601	1,641158	0,972488	0,521780
Rec27	2,03408	1,824645	1,114780	0,297977

Levene Test

... handling of complex information!

Channel	MSEffect	MSError	F	p
I1	654,1682	11,95693	54,71037	0,00
I2	700,2810	12,88420	54,35190	0,00
PA	186,1260	4,29007	43,38531	0,00
PB	21,6968	1,38846	15,62652	0,00

TABLE 4: Levene test for Homogeneity of Variances - case 4, channel A, B with flux, record from 1 to 300

The interferometric process "smooths" the non-homogeneity of inputs channels... consistent with instrument concept

Case	PA and PB	I1 and I2
1	40%	90%
2	2%	16%
3	66%	28%
4	5.2% (26/500)	65.8% (329/500)

TABLE 6: Levene test for Homogeneity of Variances in two synchronous channels

Regression analysis: models

LINEAR MODEL: $I_i = \beta_{i,A} PA + \beta_{i,B} PB, \quad i = 1,2$

dependent var: interferometric channels

regressor: photometric channels

MIXED LINEAR MODEL: $I_i = \beta_{i,A} PA + \beta_{i,B} PB + \beta_{i,AB} PA \cdot PB, \quad i = 1,2$

GOOD EXPLAINED VARIANCE (~99%)

- ➔ **Quasi-normal residuals but linear model has tails**
- ➔ **Magnitude of residuals of linear model has more variability than mixed model**

➔ The mixed terms intercept most of the residuals variability: validation of the mixed linear model.

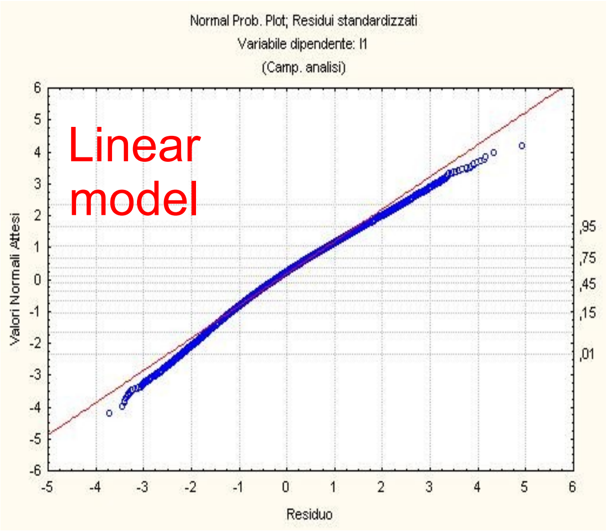
The mixed terms are statistically significant even with homogeneous data!

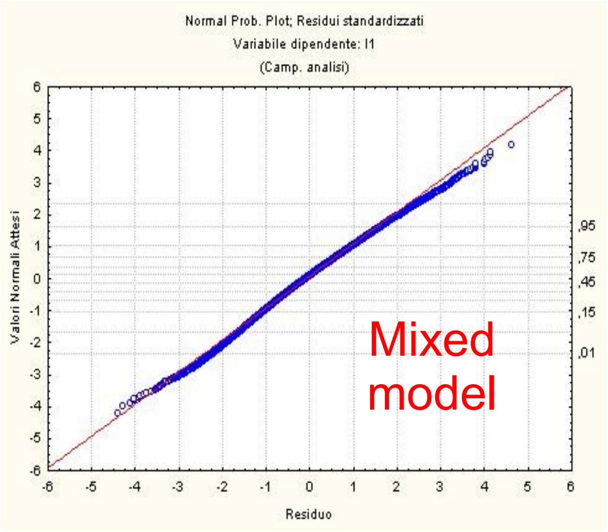
Possible physical explanation: residuals from the modulation

FRINGES DETECTABLE ALSO IN LOW VISIBILITY REGION!

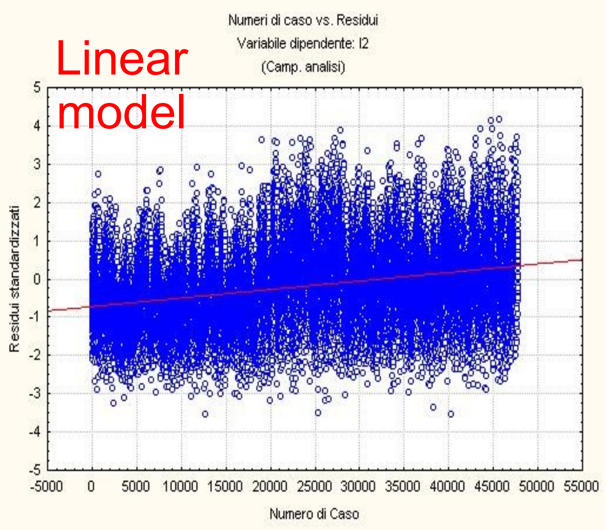
Regression analysis: models

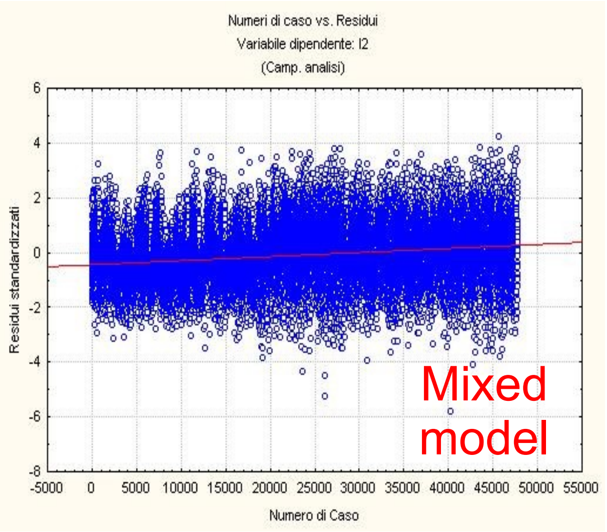
Residuals distribution: quasi normal





Residuals as function of time



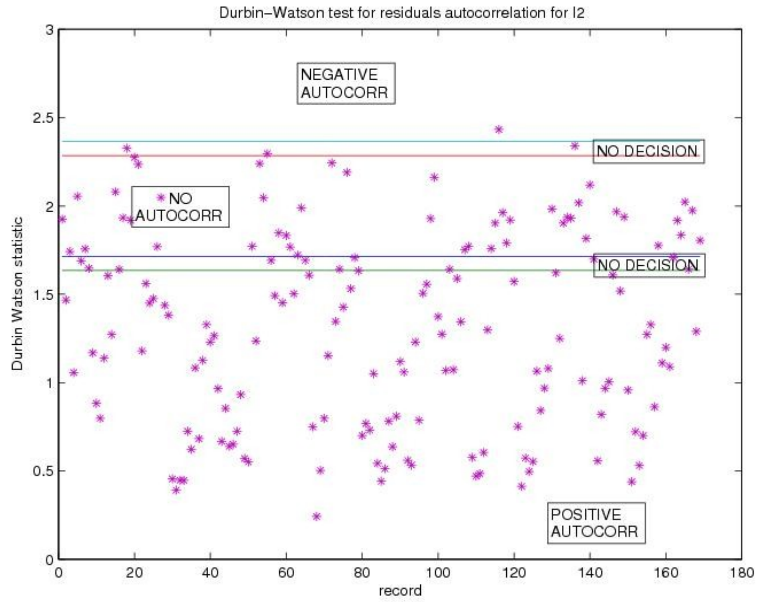


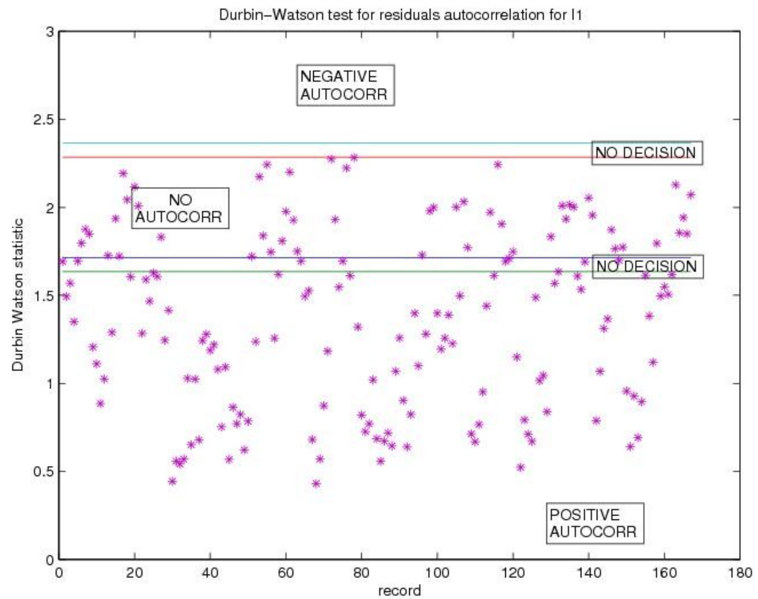
Mixed model: autocorrelation of residuals

Durbin-Watson test

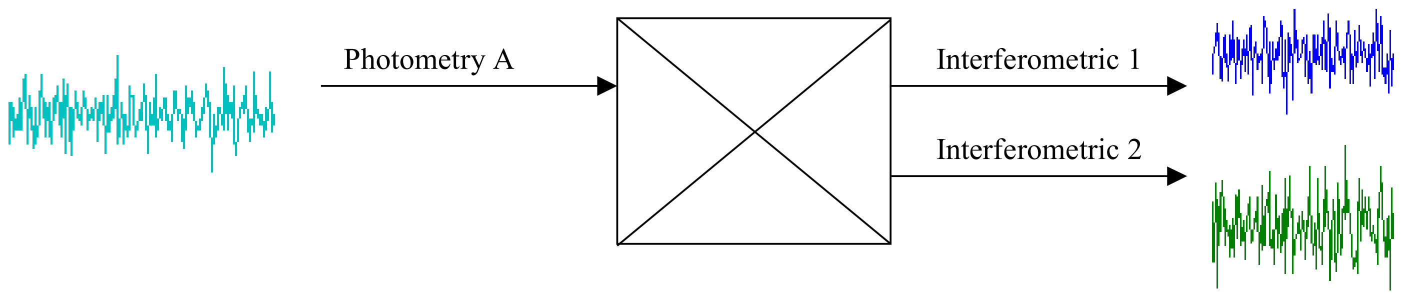
The residuals of both the linear and the mixed linear model shows in many cases the evidence of positive correlation!

Effects on the coefficients variance (not minimum anymore)





Case of homogeneous data: instrument calibration data



LINEAR MODEL: $I_i = \beta_{(i,\,A)} PA, \quad i = 1,2$

dependent var: interferometric channels

regressor: photometric channels

- **Regression coefficients** give information on weight of the photometric inputs on interferometric outputs
- **Residuals autocorrelation analysis** show some lack of homogeneity in the measurement condition

Conclusions

✗ We could isolate two components of the signals: a *trend*, changing with time, and a *uncorrelated process*, stationary over small time intervals.

➡ **Need for adaptive modelling?**

✗ We could estimate a residual noise, probably due to the interferometric process, not explained by the variability of the inputs.

✗ We could validate the mixed linear model, i.e. isolate the non-linear term within the noise component

✗ We propose the use of several statistical techniques for analysis of data quality and for instrument performance assessment (Levene test, residuals analysis, Allan variance)

Selected bibliography

- J. W. Goodman, ***Statistical Optics***, Wiley Classics Library, 1987
- Gai M., Casertano S., Carollo D., Lattanzi M.G.: ***Location estimators for interferometric fringes***, PASP, vol 110, pg 848-862, 1998
- Gai M., Menardi S., Cesare S., Bauvir B., Bonino D. et al: ***The VLTI Fringe Sensors: FINITO and PRIMA FSU***, SPIE, 5491-61, pg. 528, 2004
- Bonino D., Gai M., Corcione L.,Massone G.: ***Models for VLTI Fringe Sensors: FINITO and PRIMA FSU***, SPIE, 5491-168, pg. 1463, 2004
- Bonino, D., Gai, M., Corcione, L.: ***Fringe tracking with noisy interferometric data***, proc. of the *SPIE* workshop “Fluctuations and Noise”, ref. 6603-96, Firenze, 2007
- M. B. Priestley: ***Spectral analysis and time series***, Probability and mathematical statistics, Academic Press, 1981
- Allan D.W.: ***Time and frequency characterization, estimation and prediction of precision clocks and oscillators,*** IEEE, UFFC-34:647-654,1987
- J. O. Rawlings, S. G. Pantula, D. A. Dickey: ***Applied Regression Analysis: a research tool,*** Springer Texts in Statistics, 1998

- D. Bonino: ***Analysis of measurement algorithms and modelling of interferometric signals for infrared astronomy,*** Ph.D. Thesis, University of Turin, Dep. Mathematics, 2008; Advisors: M. Gai, L. Sacerdote.

in the framework of the project PRIN INAF 2007 no. 6