Working with astrometric data

X. Luri

Scientist's ideal case

- **Error-free data**
  - No biases
  - No random errors
  - No correlations
- **Complete sample**
  - No censorships
- **Direct measurements**
  - No transformations
  - No assumptions

Never ever available

(not even in your dreams)

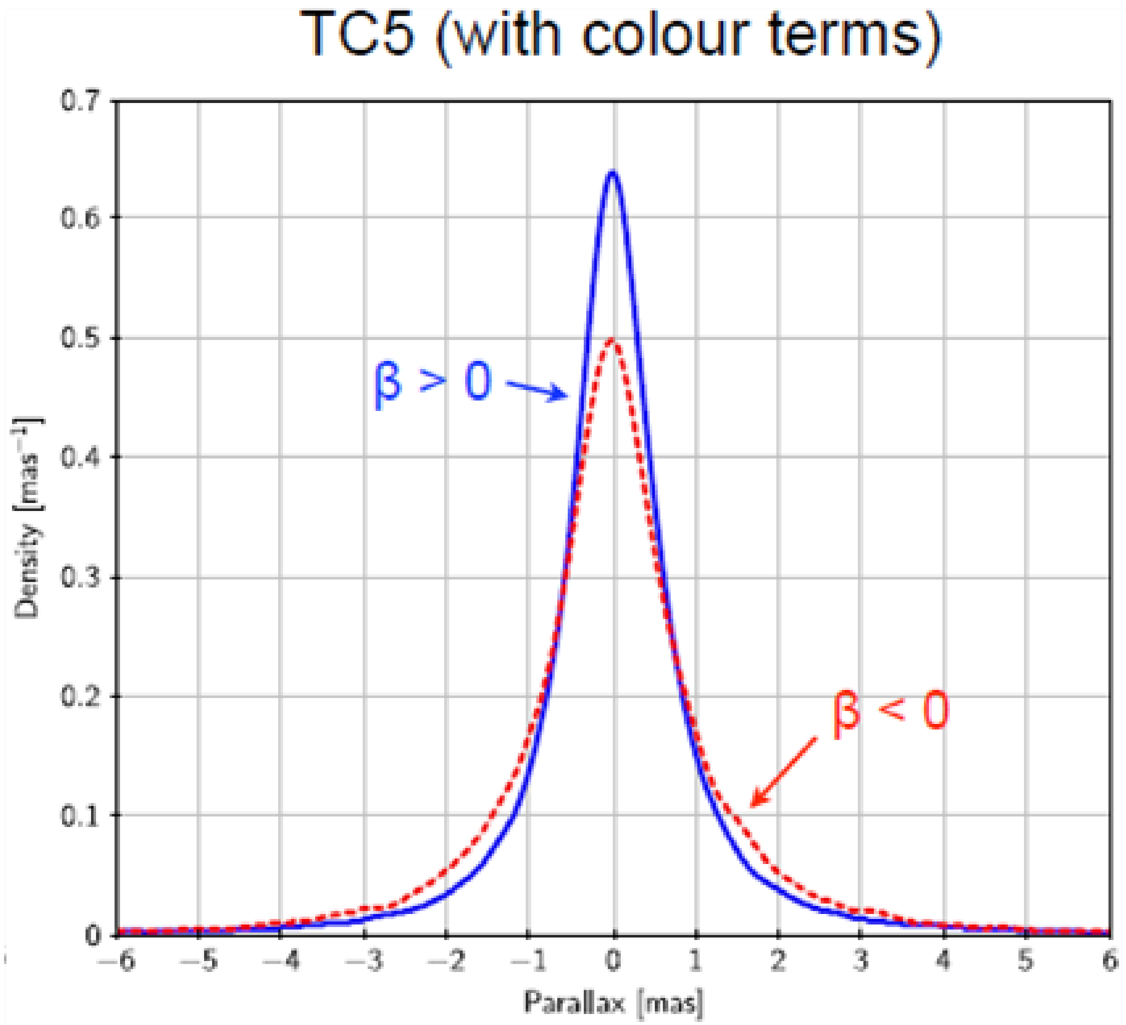
Errors 1: biases

Bias: your measurement is systematically too large or too small

For DR1 parallaxes:

- Probable global zero-point offset present; -0.04 mas found during validation
- Colour dependent and spatially correlated systematic errors at the level of 0.2 mas
- Over large spatial scales, the parallax zero-point variations reach an amplitude of 0.3 mas
- Over a few smaller areas (2 degree radius), much larger parallax biases may occur of up to 1 mas
- There may be specific problems in a few individual cases

Global zero point from QSO parallaxes

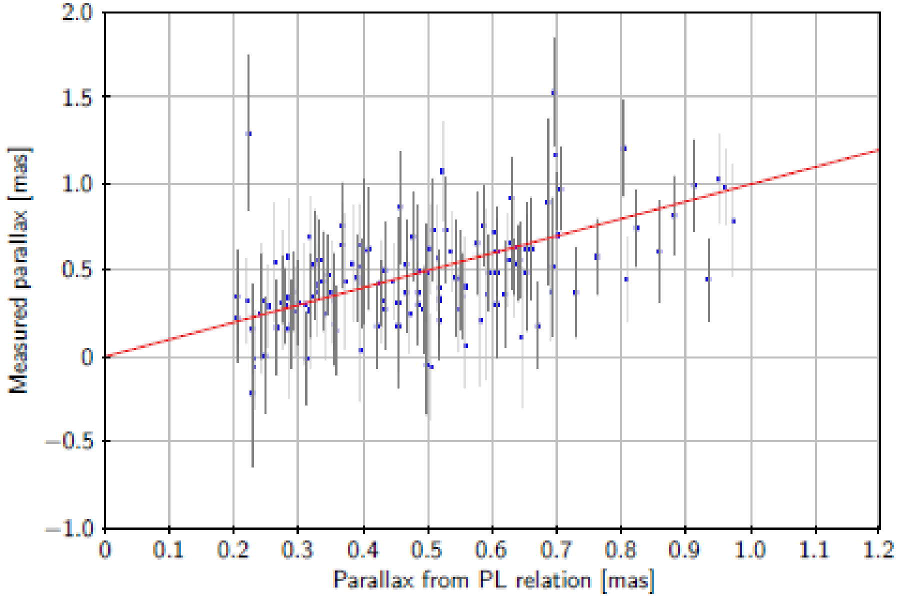


$\beta > 0$: med(ϖ) = +0.002 mas

$\beta < 0$: med(ϖ) = −0.020 mas

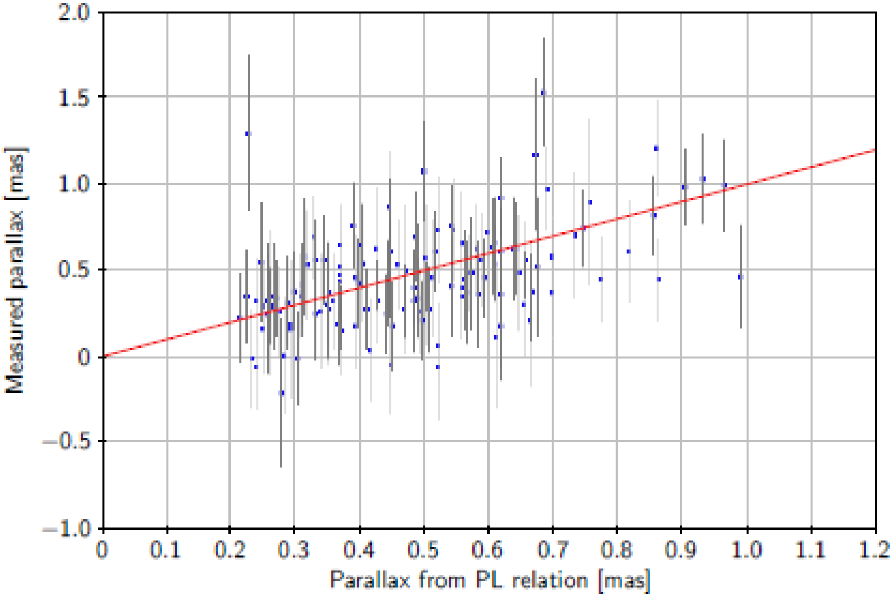
Global zero point from Cepheids

P-L relation from Tammann et al. (2003):
$M_V = -3.141 \log P - 0.820$



med(Δϖ) = −0.015 mas

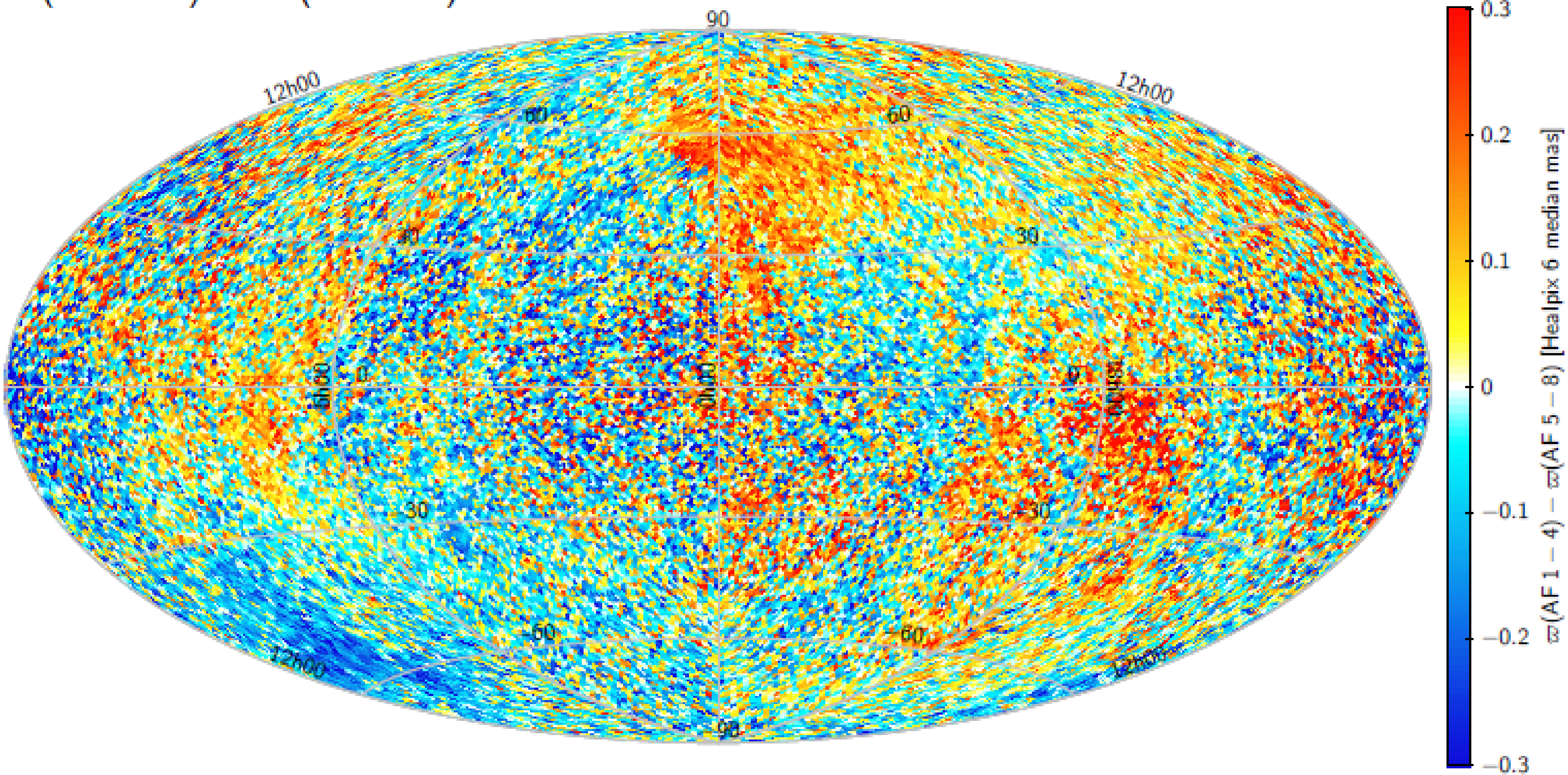
P-L relation from Fouqué et al. (2007)
$M_V = -2.678 \log P - 1.275$



med(Δϖ) = −0.017 mas

Regional effects from split FOV solutions
(equatorial coordinates)

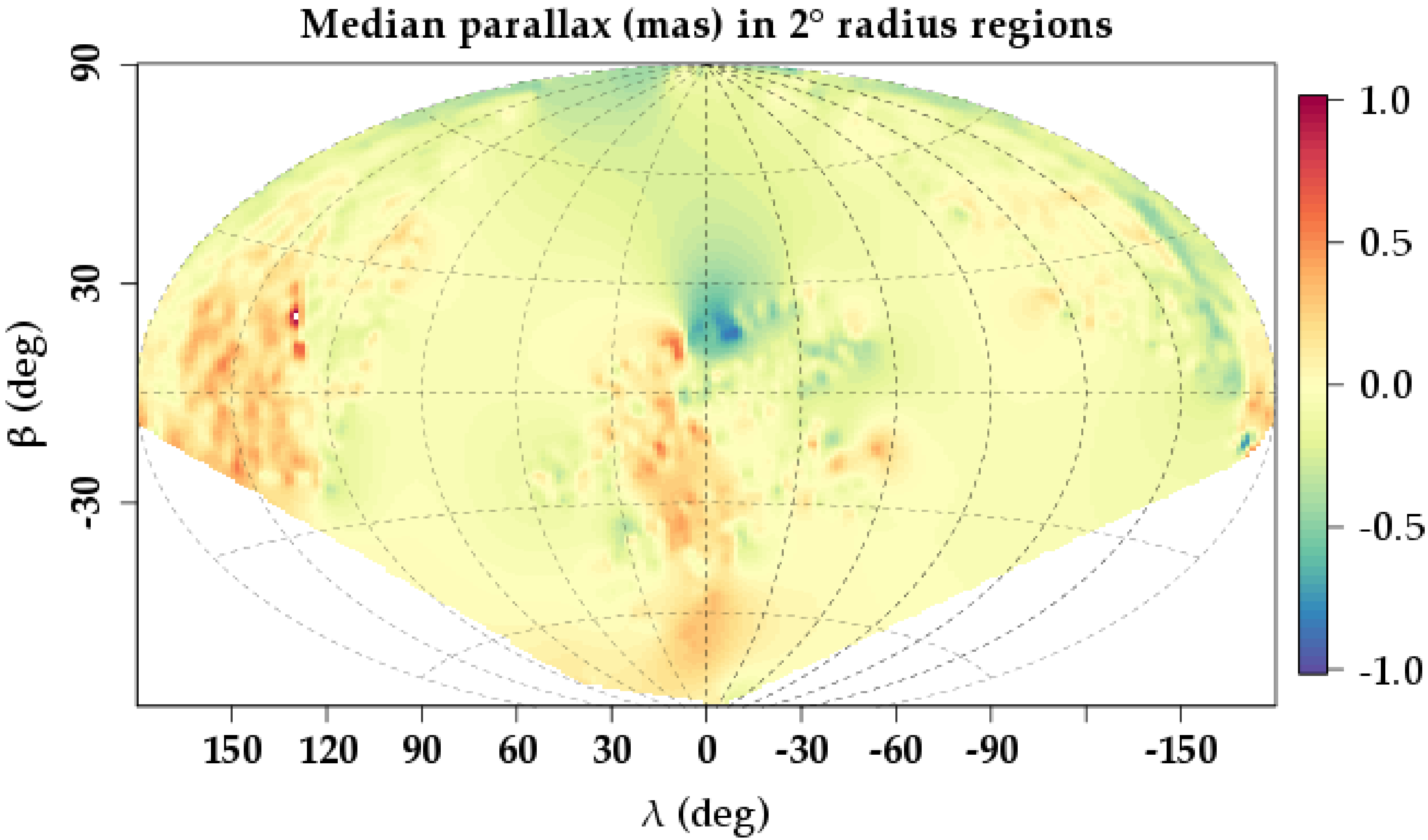
ϖ(AF1-4) − ϖ(AF5-8)



Median value per pixel (~1 deg^2)

A. Bombrun

Regional effects from QSOs
(ecliptic coordinates)



How to take this into account

- You can introduce a global zero-point offset to use the parallaxes (suggested -0.04 mas)

- **You can not correct the regional features:** if we could, we would already have corrected them. We have indications that these zero points may be present, but no more.

- For most of the sky assume at least an additional systematic error of 0.3 mas; your error can not go below this value

$$\boldsymbol{\varpi \pm \sigma_{\varpi}}\ \textbf{(random)} \pm \textbf{0.3 mas (syst.)}$$

- For a few smaller regions be very aware that the systematics can reach 1 mas

More specifically: treat separately random error and bias, but if you must combine them a **worst case** formula can be as follows

Slide updated from the presentation at ESAC

- **For individual parallaxes:** to be on the safe side add 0.3 mas to the standard uncertainty

$$\sigma_{\mathrm{Total}} \approx \mathrm{sqrt}(\sigma^2_{\mathrm{Std}}+0.3^2)$$

- **When averaging parallaxes for groups of stars:** the random error will decrease as sqrt(N) but the systematic error (0.3 mas) will not decrease

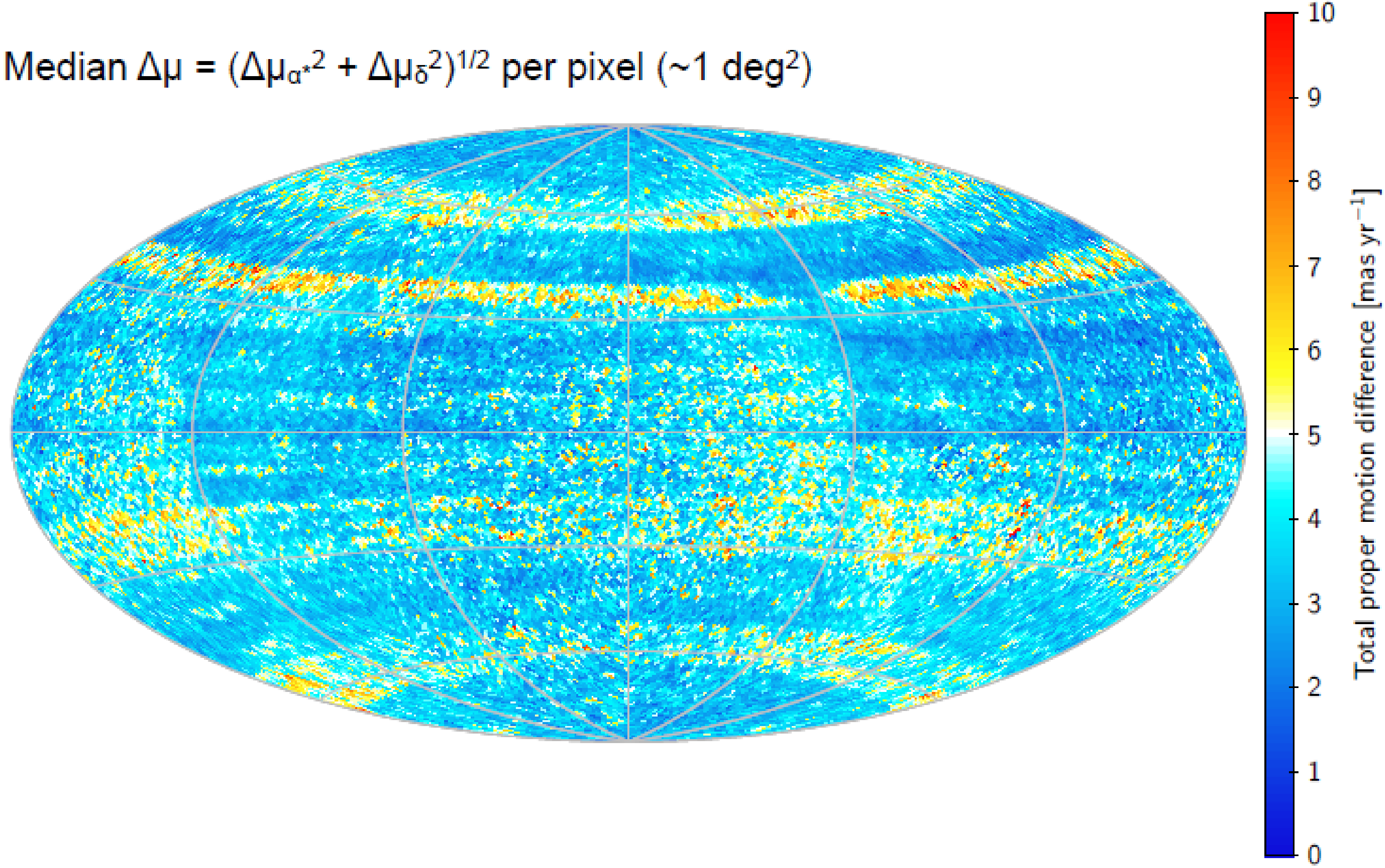
- **Don’t try to get a “zonal correction” from previous figures**, it’s too risky

For DR1 proper motions and positions:

- In this case Gaia data is the best available, by far
- We do not have means to do a check as precise as the one done for parallaxes, but there are no indications of any significant bias
- For positions remember that for comparison purposes you will likely have to convert them to another epoch. You should propagate the errors accordingly.

Slide updated from the presentation at ESAC

Comparison with Tycho-2 shows that catalogue systematics

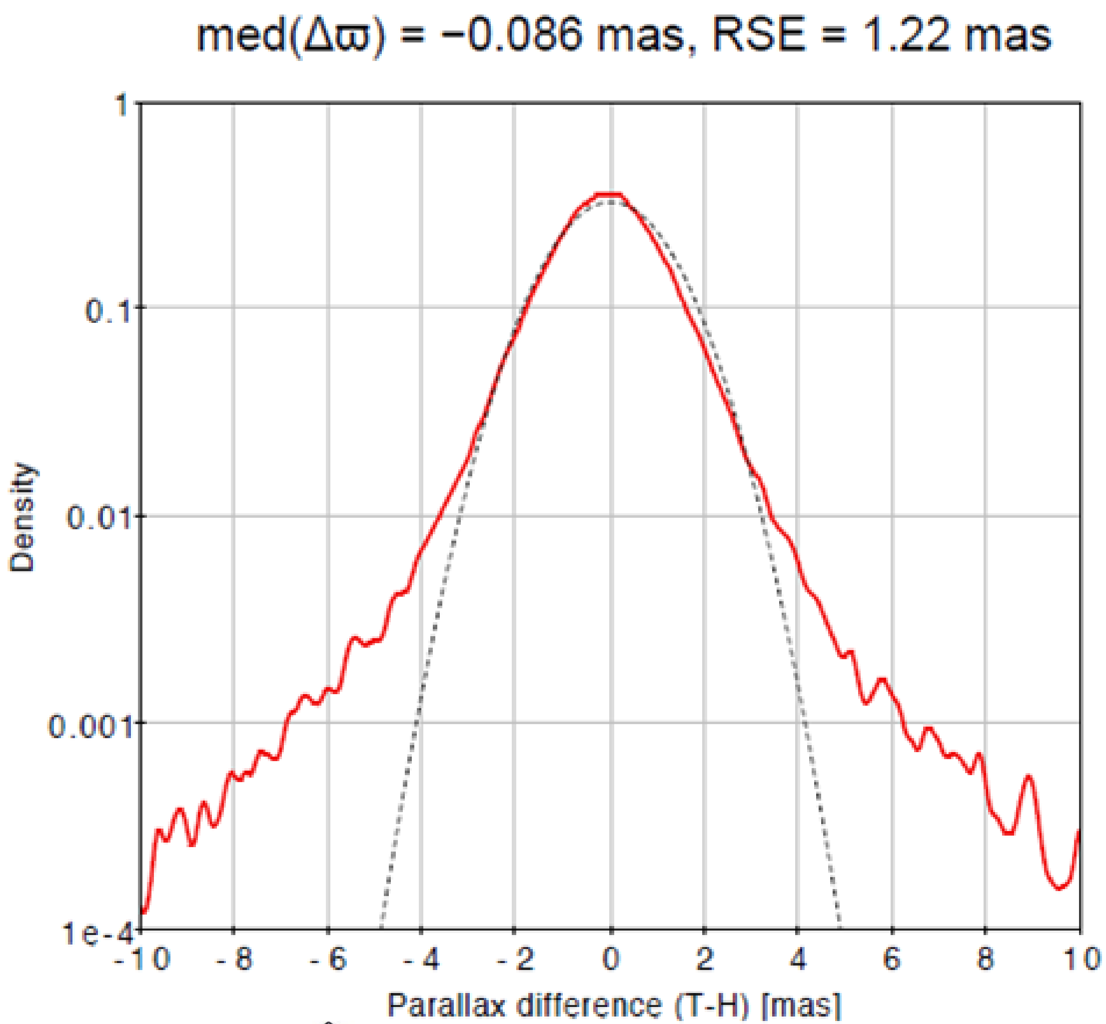


Errors 2: random errors

Random error: your measurements are randomly distributed around the true value

- Each measurement in the catalogue comes with a formal error
- Random errors in Gaia are quasi-normal. The formal error can be assimilated to the variance of a normal distribution around the true value.
- Formal errors may be slightly overestimated

Warning: comparison with Hipparcos shows deviation from normality beyond ~2σ

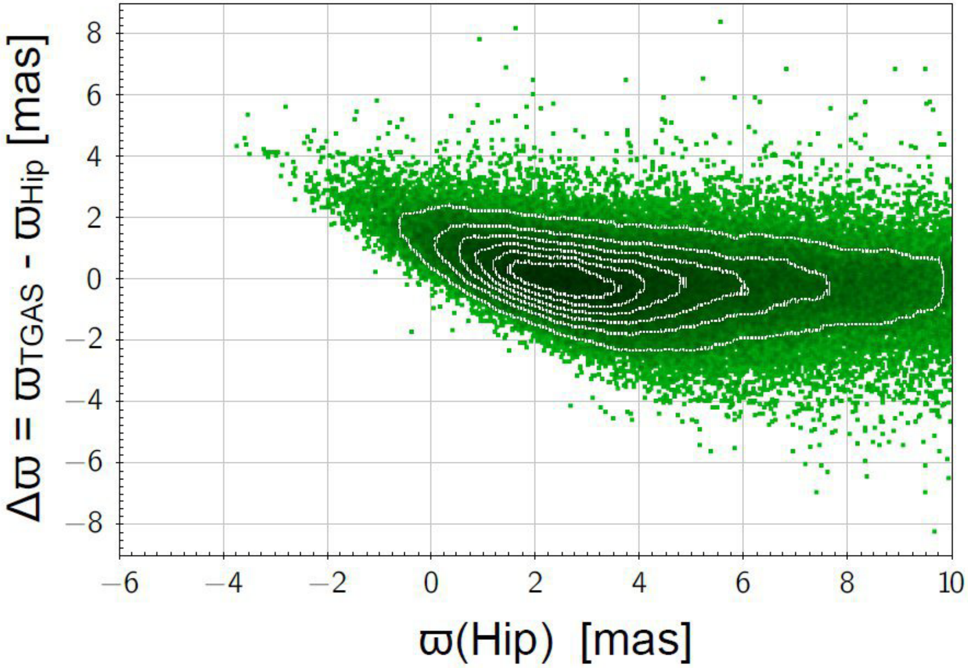


To take into account for outlier analysis

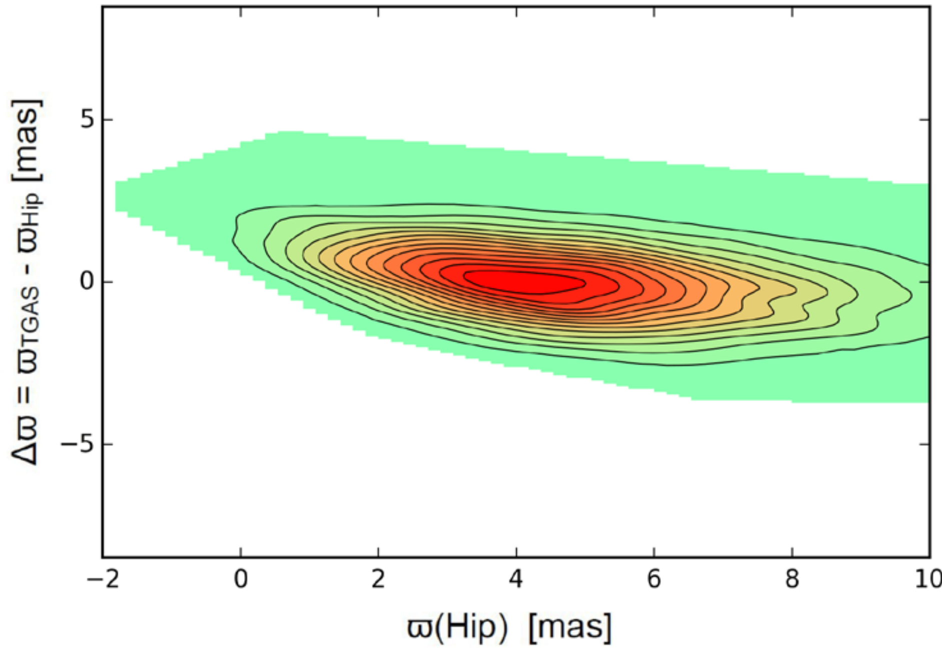
Warning: when comparing with other sources of trigonometric parallaxes take into account the properties of the error distributions

TGAS vs Hipparcos

Observations



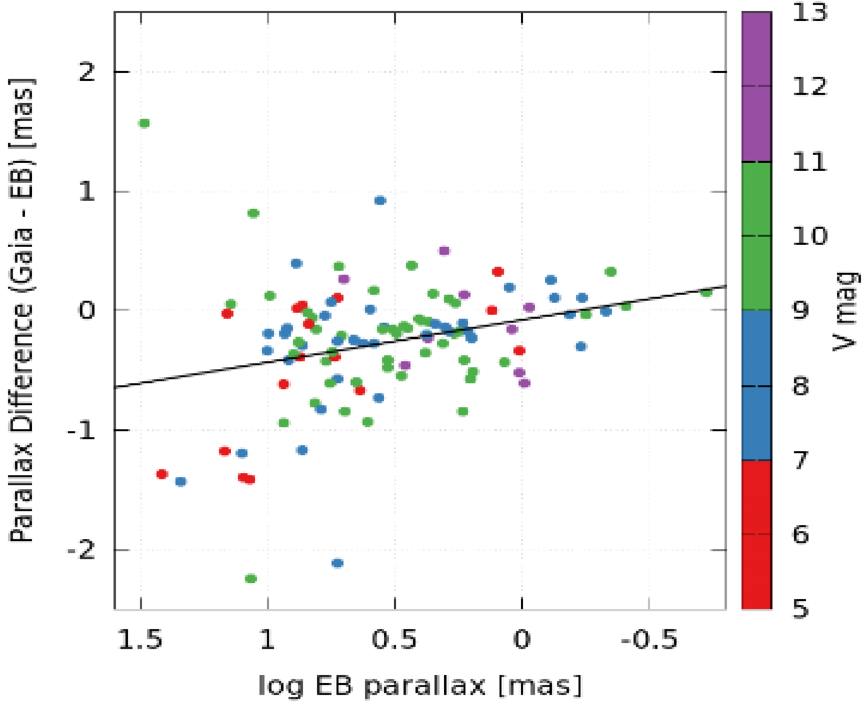
Simulations



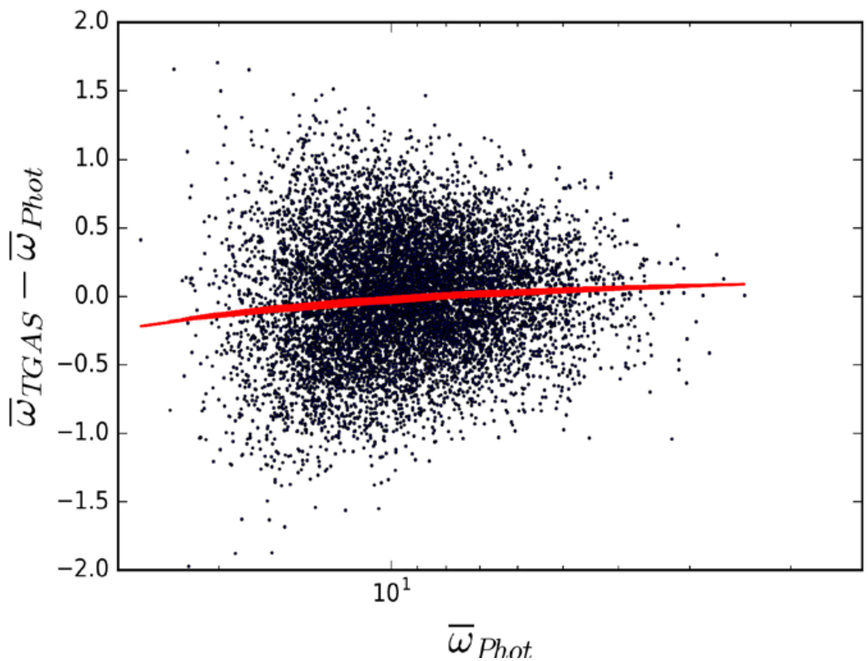
The “slope” at small parallaxes is simply due to the different magnitude of the errors in the two catalogues

Eclipsing binaries parallaxes vs TGAS

arXiv:1609.05390v3



Simulation



The overall “slope” is due to the different error distributions in parallax (lognormal for photometric, normal for trigonometric)

Errors 3: correlations

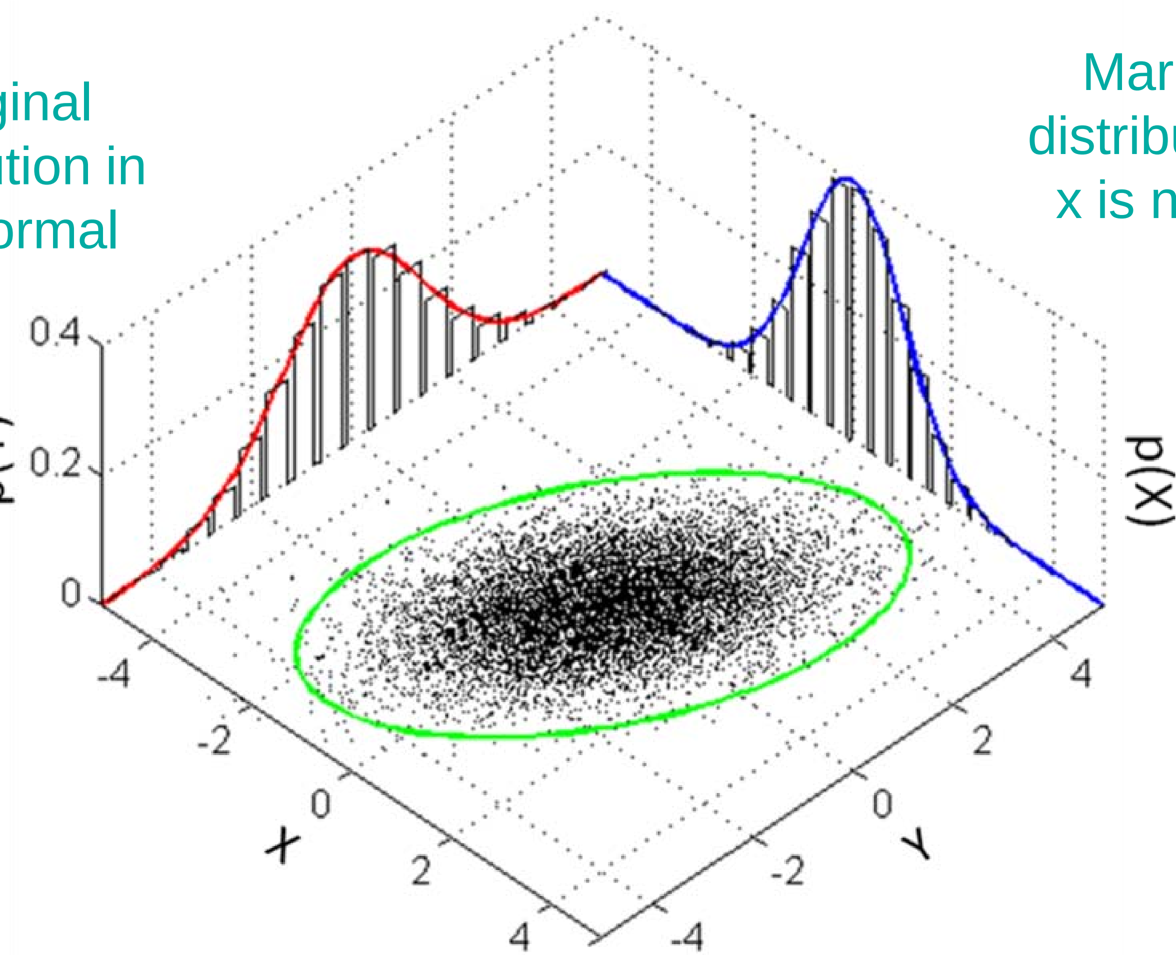
Correlation: the measurements of several quantities are not independent from each other

- The errors in the five astrometric parameters provided are not independent
- The ten correlations between these parameters are provided in the archive (correlation matrix)

Example of two correlated parameters

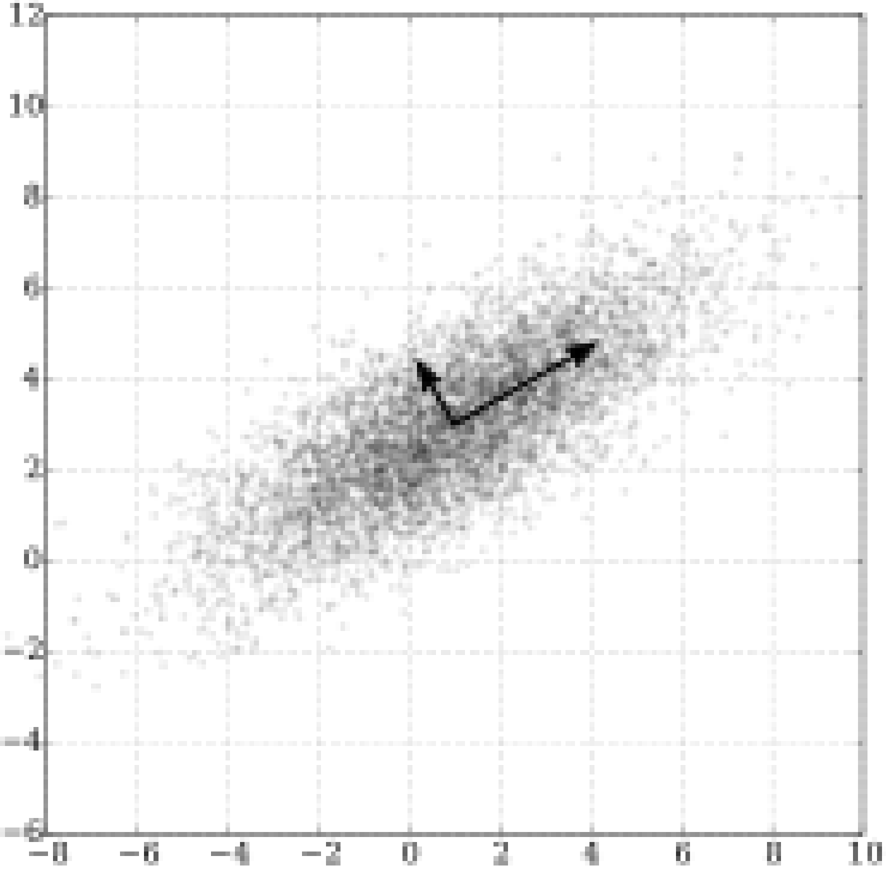
Marginal distribution in y is normal

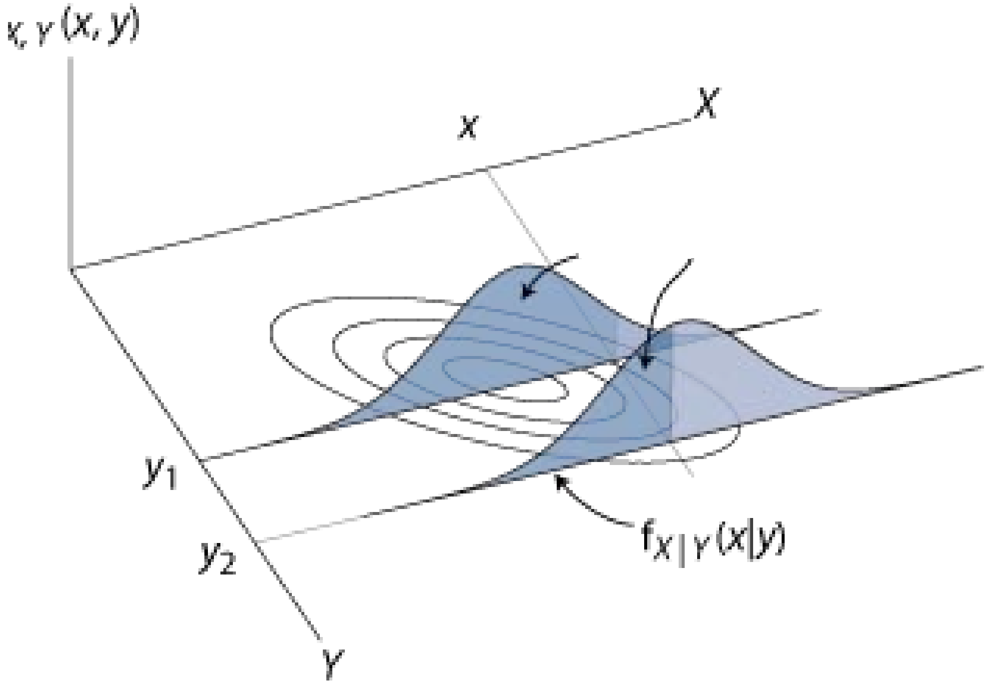
Marginal distribution in x is normal



By Bscan - Own work, CC0, https://commons.wikimedia.org/w/index.php?curid=25235145

Beware when using these quantities together

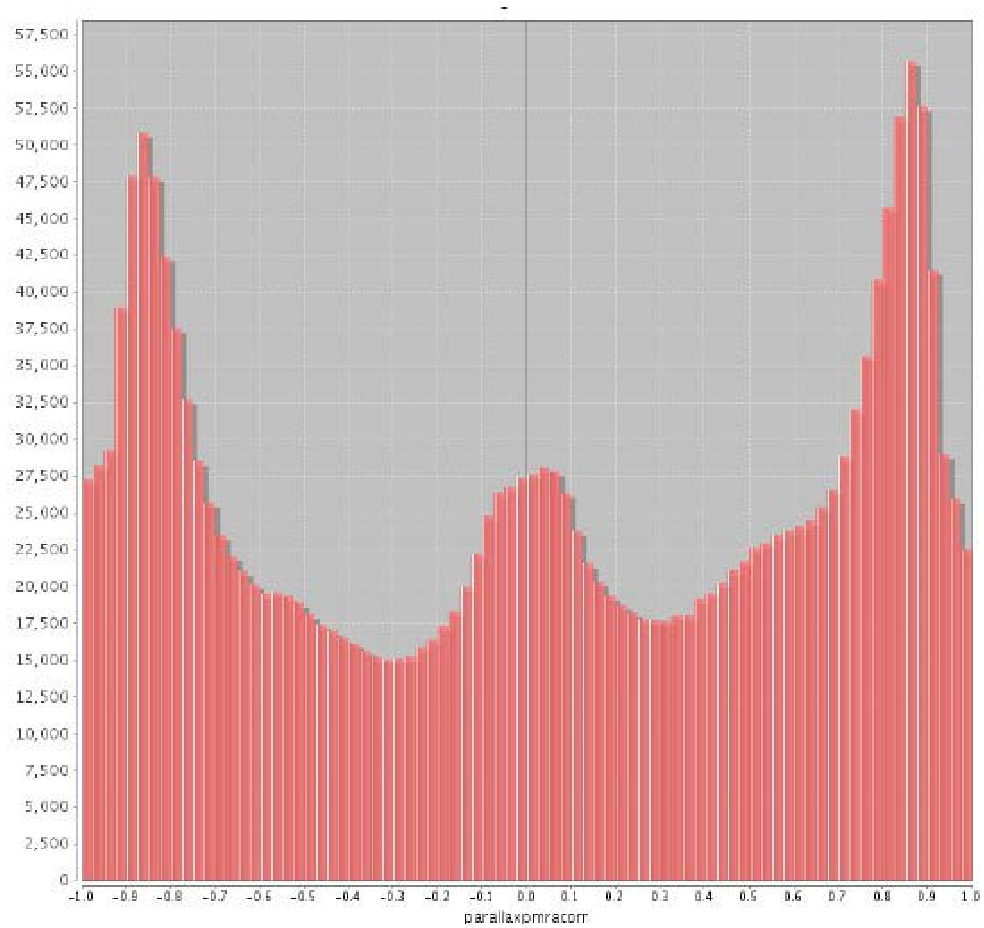


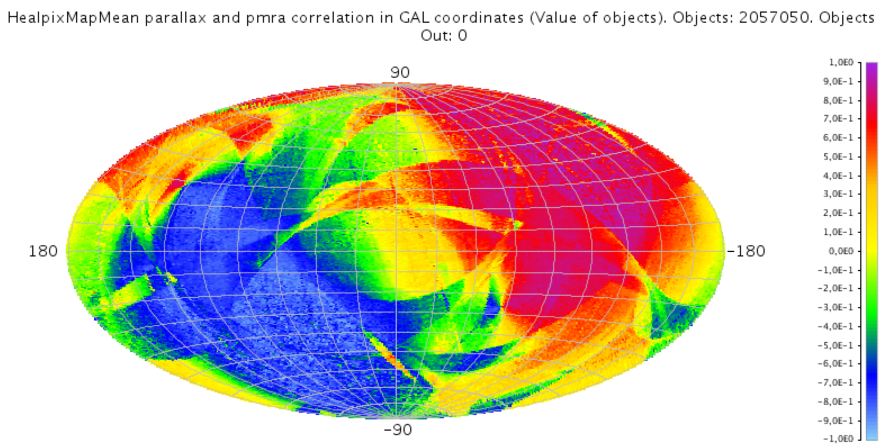


Example of problematic use:

- Calculating the transversal velocities of a set of stars
- The resulting dispersion of velocities has to be corrected from the effect of the errors in parallax and proper motion
- This correction can not be done using the parallax and proper motion errors separately, the correlations have to be taken into account

Beware: large and unevenly distributed correlations in DR1
Ex. PmRA-Paralax correlation



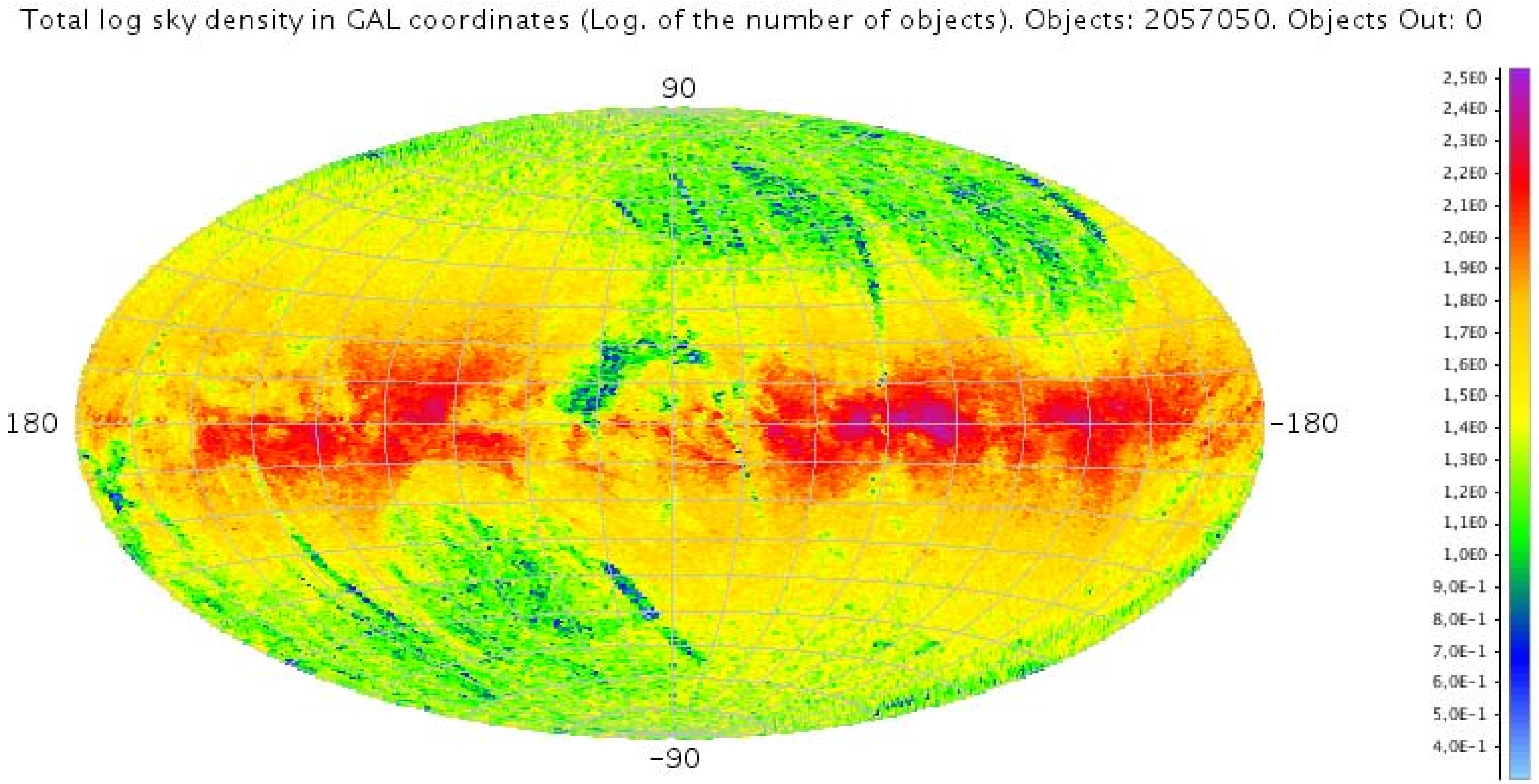


Sample censorships

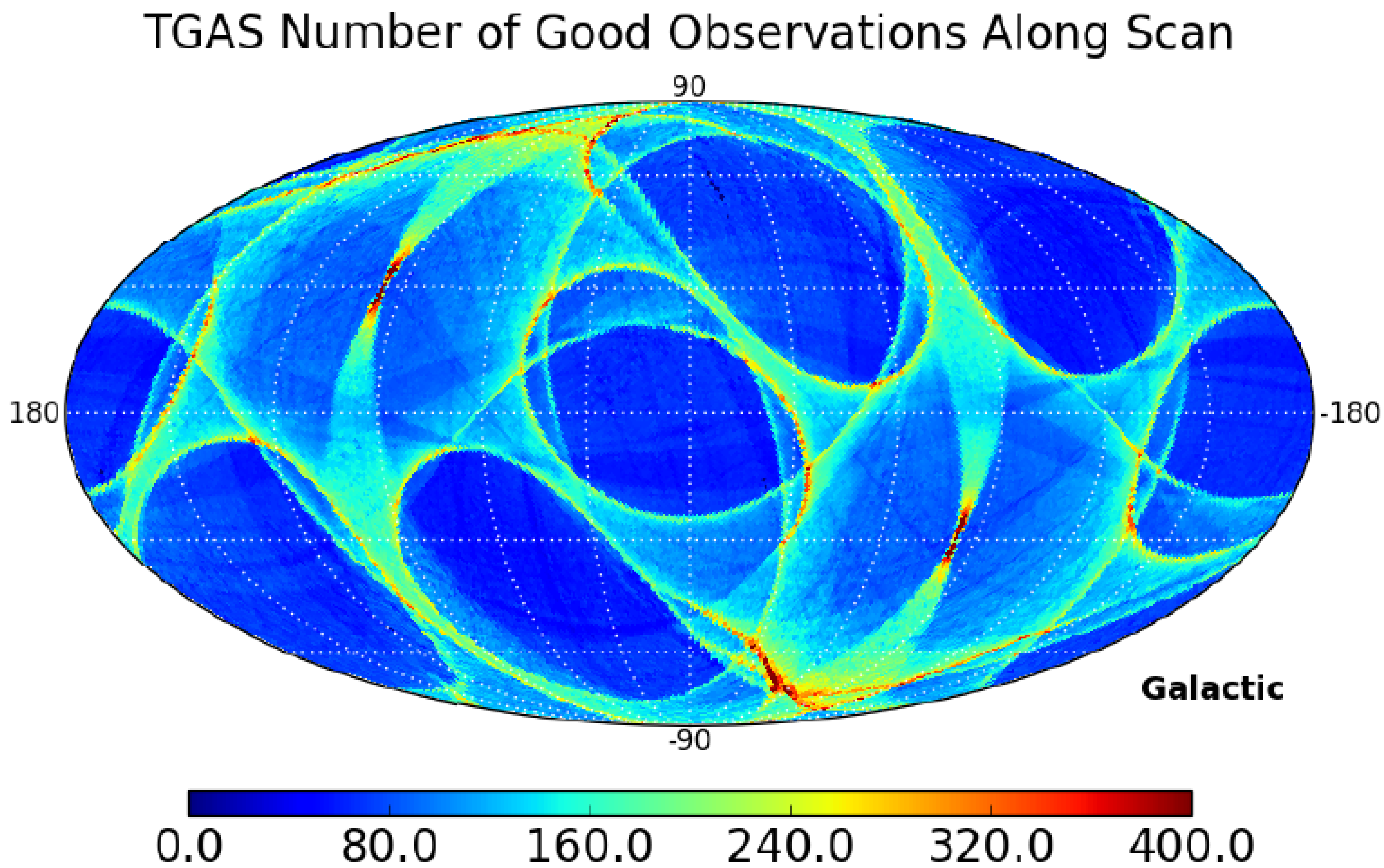
Completeness/representativeness: we have the complete population of objects or at least a subsample which is representative for a given purpose

- DR1 is a very complex dataset, its completeness or representativeness can not be guaranteed for any specific purpose

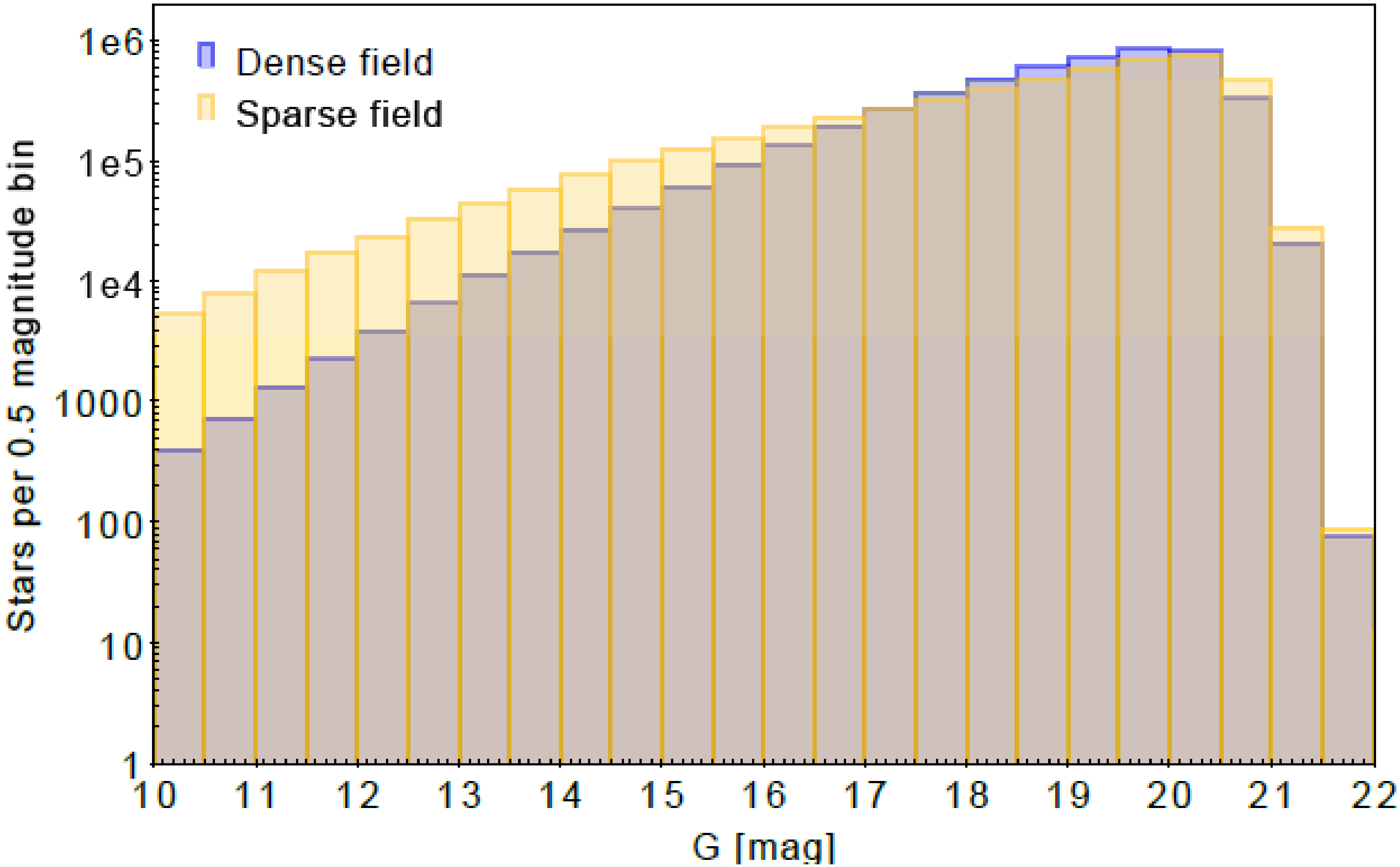
Significant completeness variations as a function of the sky position



Complex selection of astrometry (e.g. Nobs)



Not complete in magnitude or color



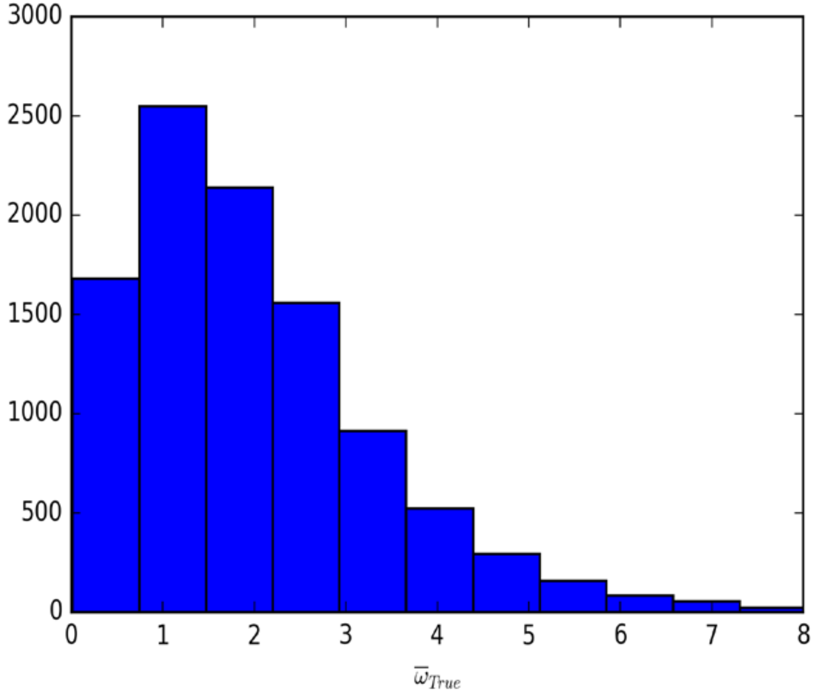
How to take this into account

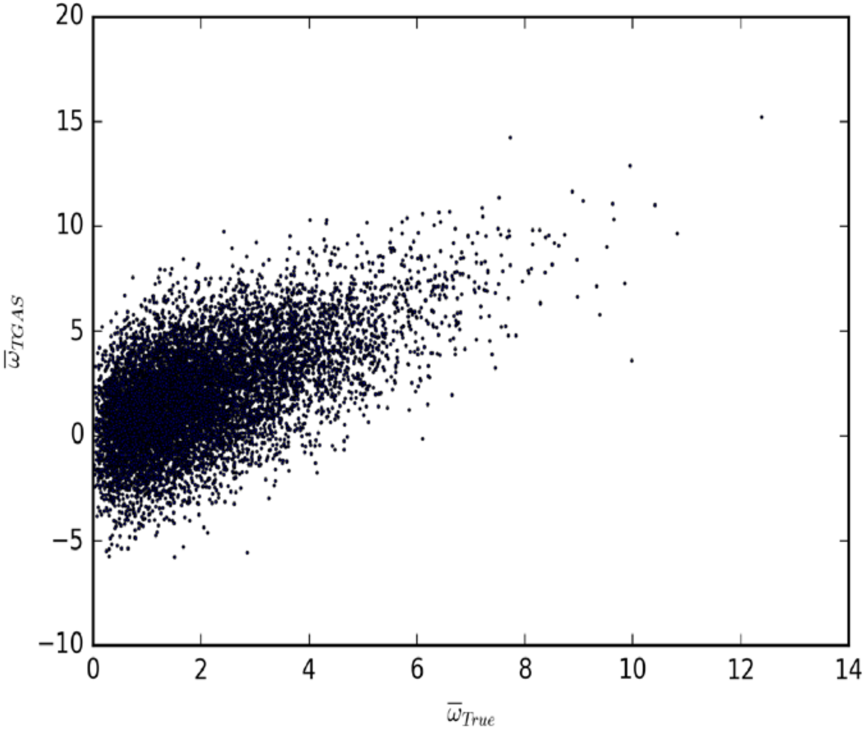
- **Very difficult,** will depend on your specific purpose
- Analyze if the problem, and try to determine if the known censorships are correlated with the parameter you are analyzing (see validation paper)
- At least do some simulations to evaluate the possible effects

IMPORTANT: do not make things worse by adding your own additional censorships

- **This is specially important for parallaxes**
- Avoid removing negative parallaxes; this removes information and biases the sample for distant stars
- Avoid selecting subsamples on parallax relative error. This also removes information and biases the sample for distant stars
- **Use instead fitting methods able to use all available data (e.g. bayesian methods) and always work on the observable space (e.g. ABL method)**

Example: Original (complete) dataset
(errors in parallax of 2mas)

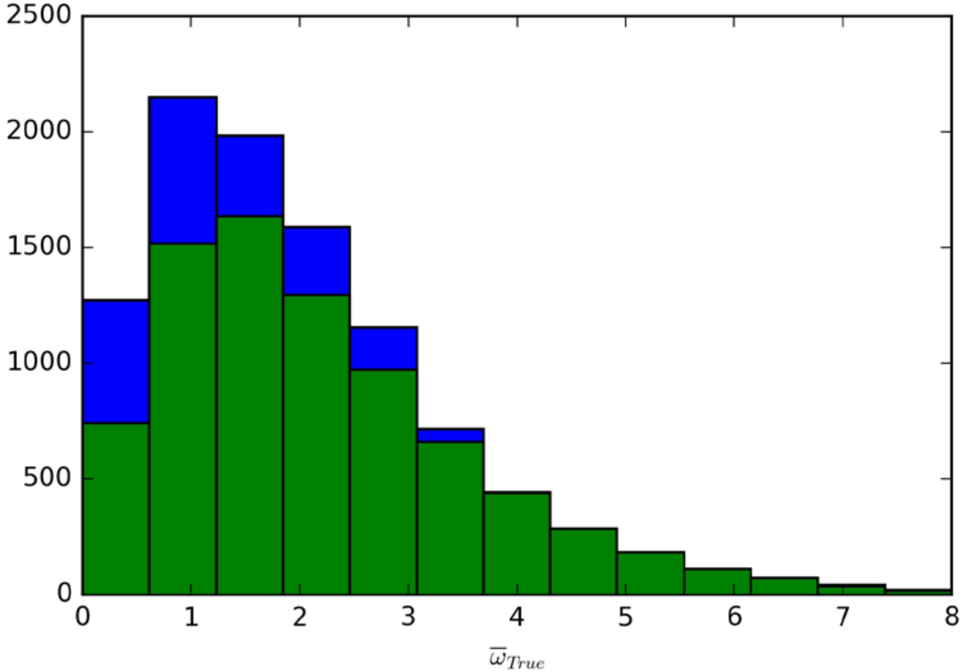


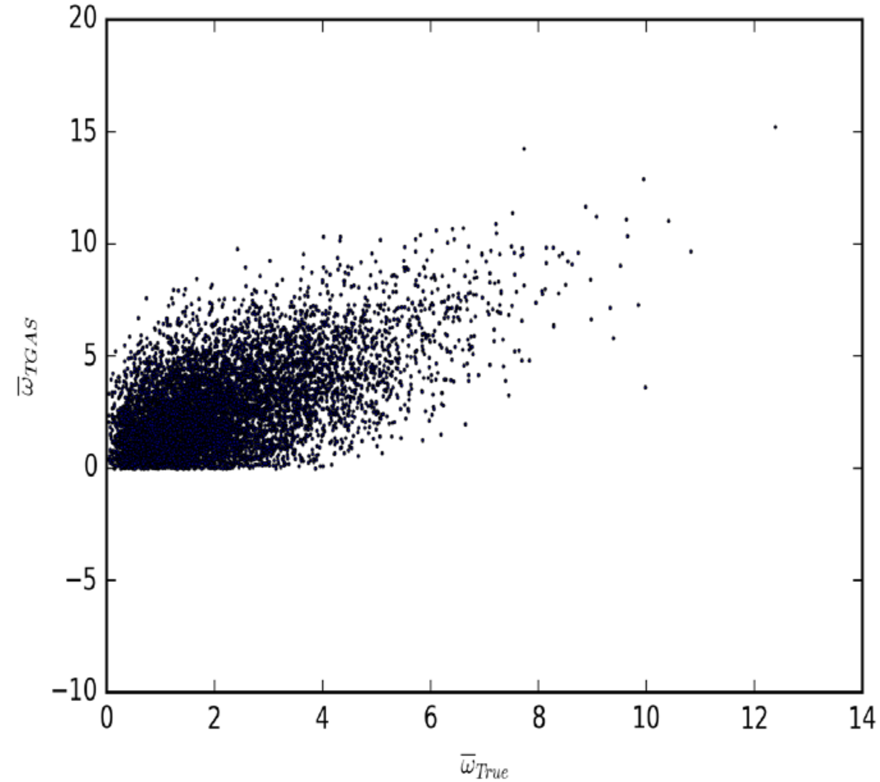


Average dif. of parallaxes = 0.002 mas

Example: removing negative parallaxes

Favours large parallaxes

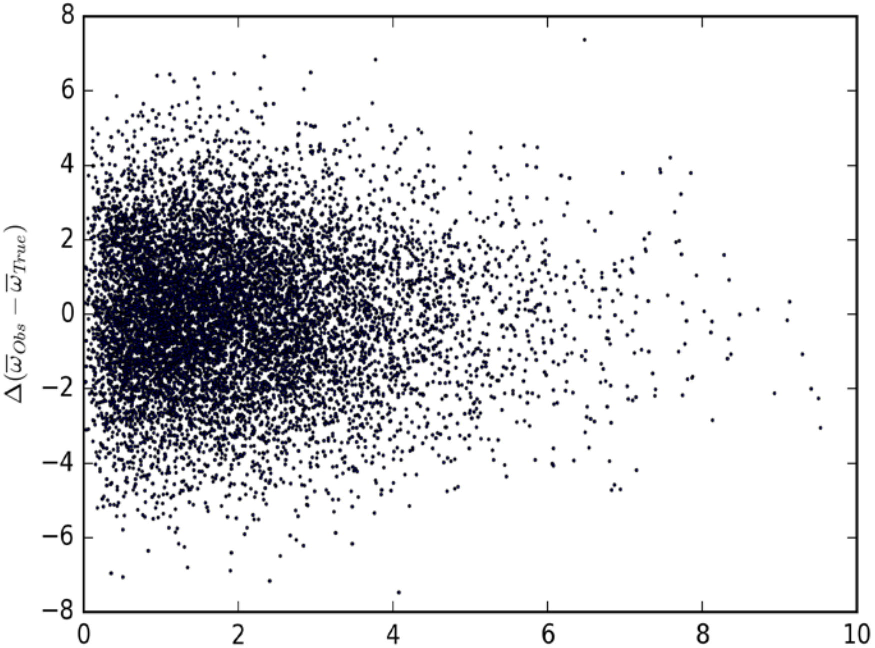


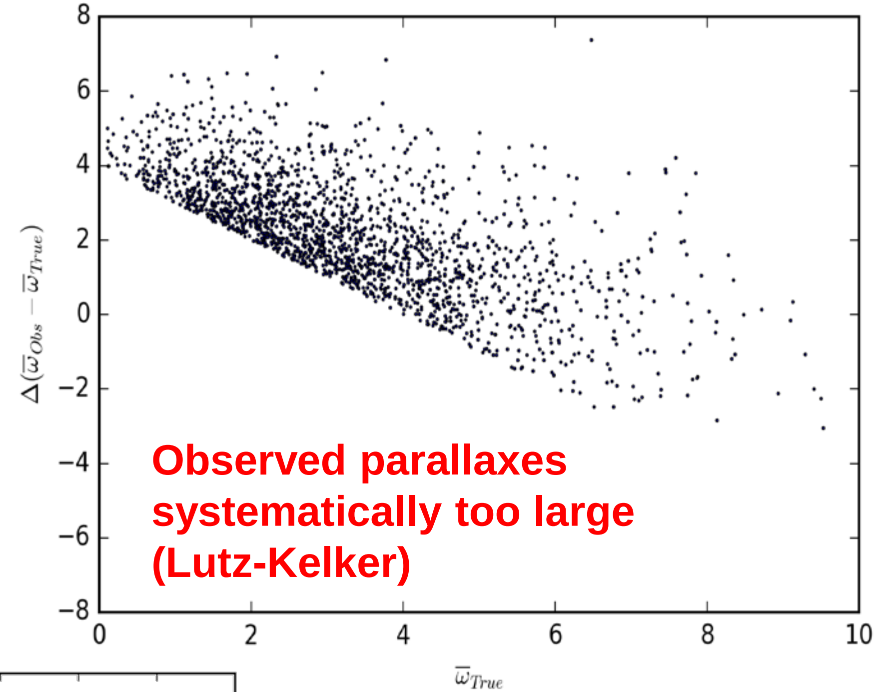


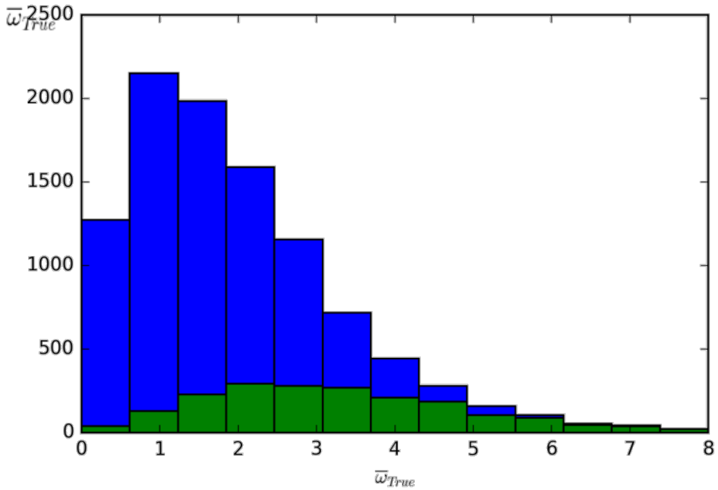
Average dif. of parallaxes = 0.65 mas

Example: removing sigmaPi/Pi > 50%

Favours errors making parallax larger



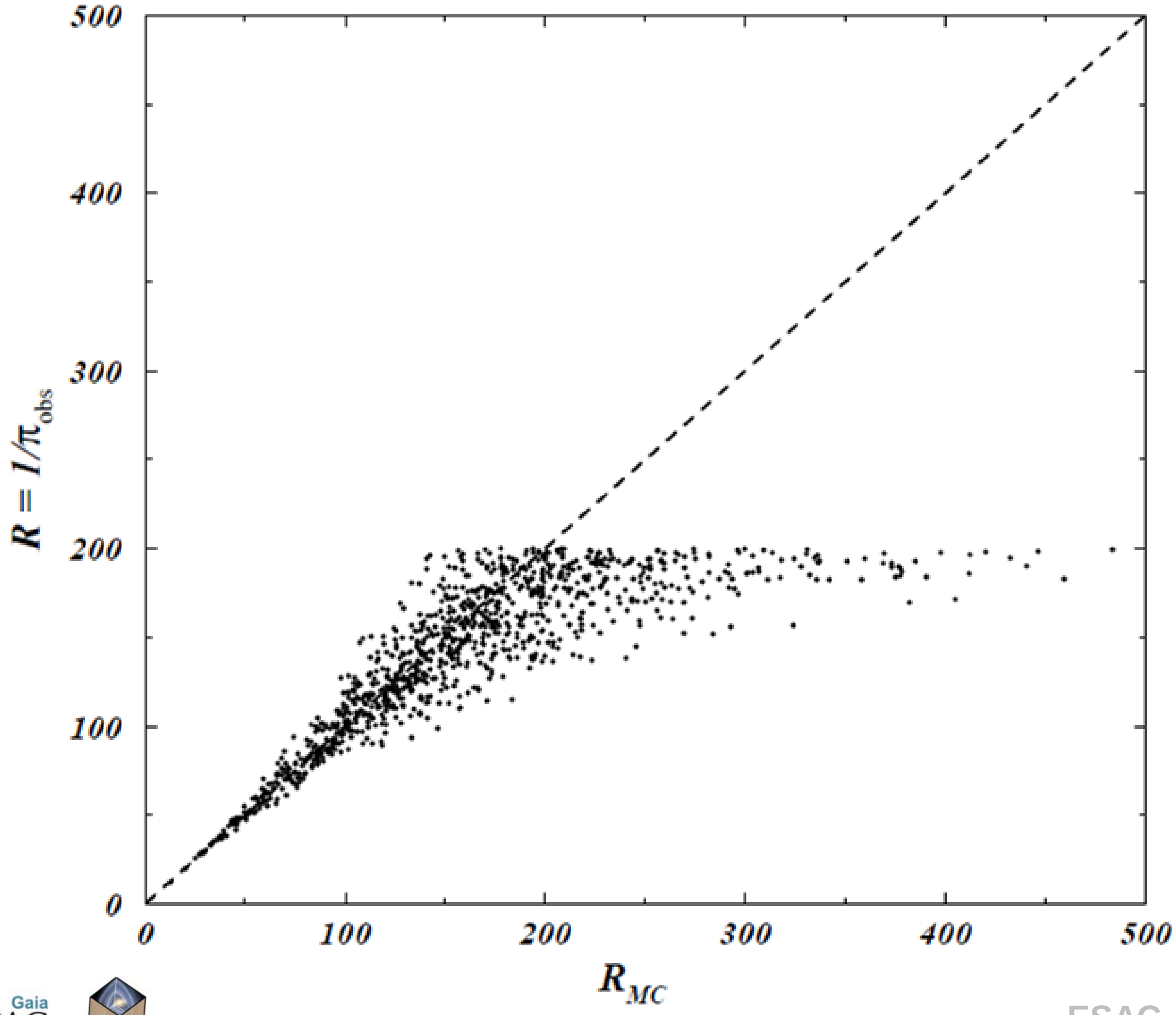




Average dif. of parallaxes = 2.2 mas

Example: truncation by observed parallax

Favours large distances



Transformations

Transformations: when the quantity you want to study is not the quantity you observe

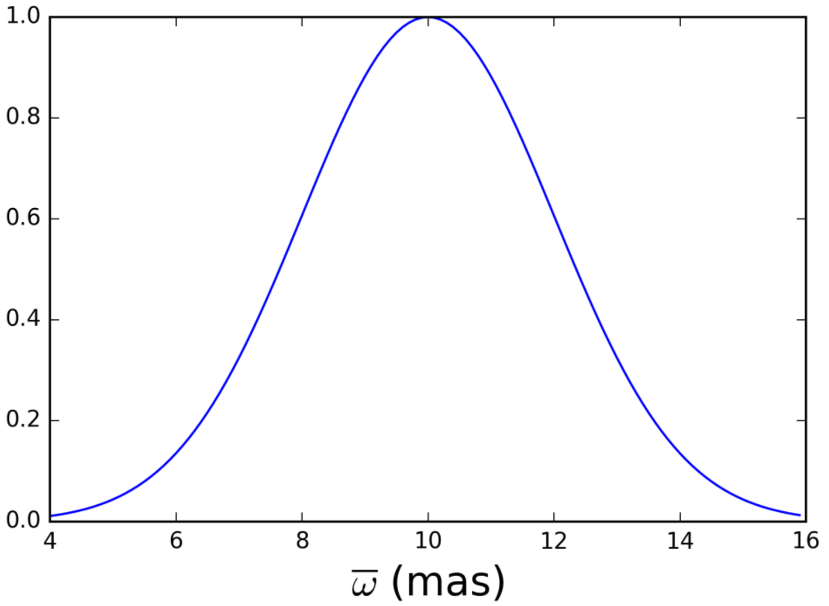
- Usually you want distances, not parallaxes
- Usually you want spatial velocities, not proper motions

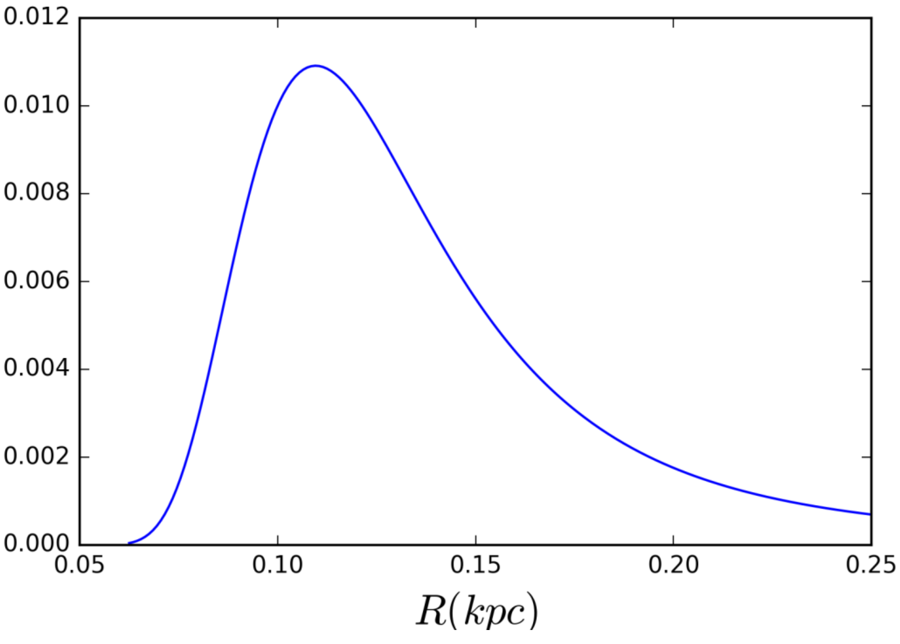
Warning: when using a transformed quantity the error distribution also is transformed

- This is specially crucial for the calculation of distances from parallaxes
- A symmetrical, well behaved error in parallax is transformed into an asymmetrical error in distance
- Can lead to negative or very large values in the error distribution for distances

Error distribution comparison: star at 100pc and parallax error 2mas

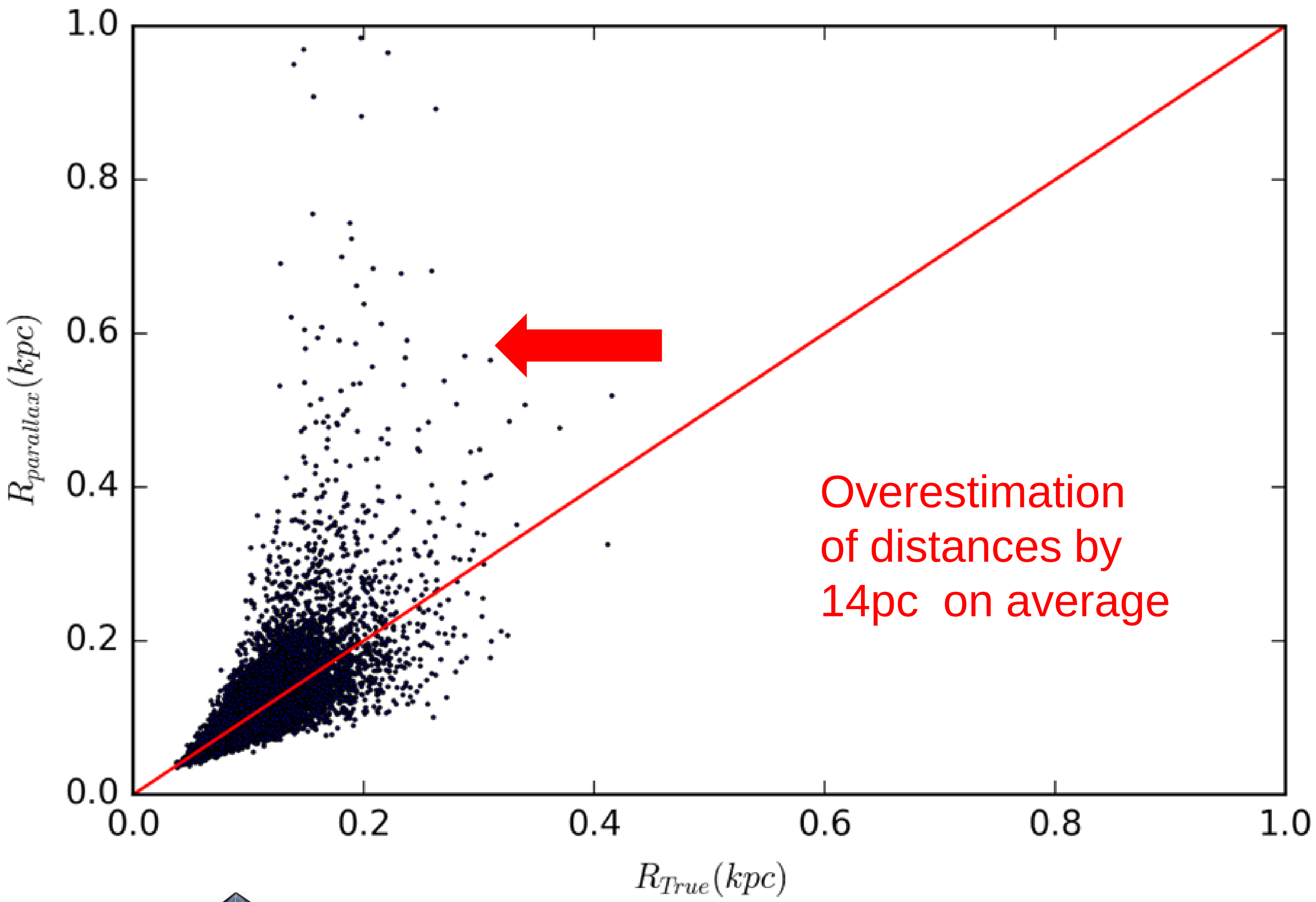
parallax and distance (non normalised)





Sample simulation with a parallax error of 2mas

True distance vs. distance from parallax



How to take this into account

- Avoid using transformations as much as possible
- If unavoidable
  - Do fits in the plane of parallaxes (e.g. PL relations using ABL method*) where errors are well behaved
  - Do any averaging in parallaxes and then do the transformation (e.g. distance to an open cluster)
  - Always estimate the remaining effect (analytically of with simulations)

*Astrometry-Based Luminosity (ABL)

$$a_V = 10^{0.2M_V} = \pi 10^{\frac{m_V+5}{5}}$$

Also beware of additional assumptions

- For instance about the absorption when calculating absolute magnitudes from parallaxes

Thank you