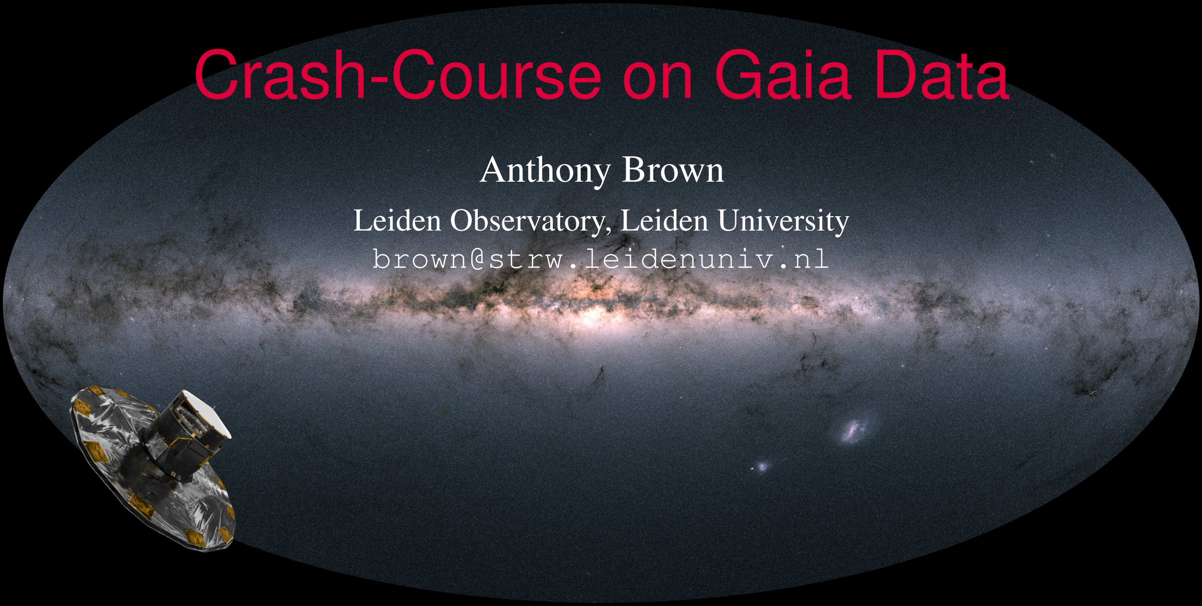
Crash-Course on Gaia Data

Anthony Brown

Leiden Observatory, Leiden University

brown@strw.leidenuniv.nl

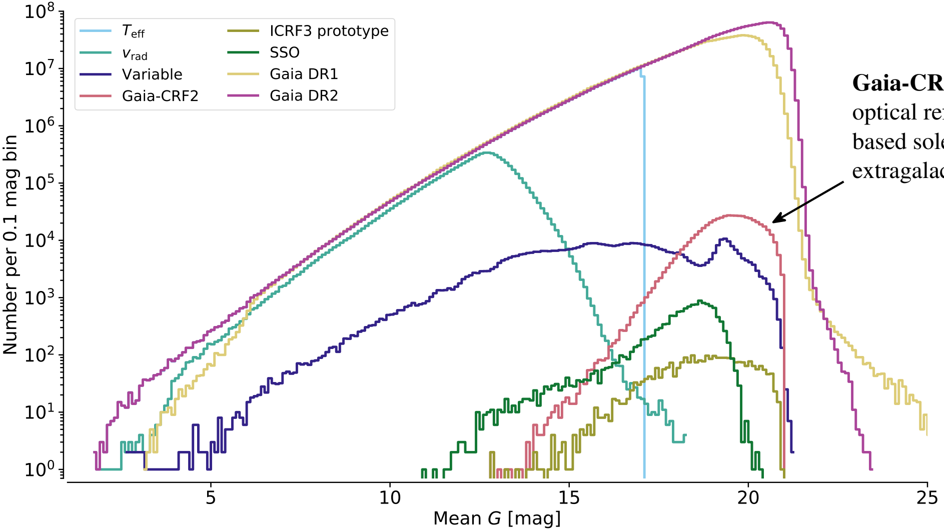


ESA/Gaia/DPAC

Credits

Most of the material in these slides is from the papers in the
Gaia DR2 A&A special issue
and from the
Lindegren et al. slide set 'Gaia DR2 astrometry'

What is there besides astrometry, radial velocity, and photometry?

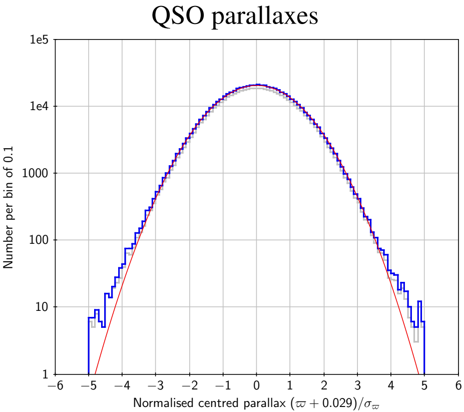


Gaia-CRF2: First optical reference frame based solely on extragalactic sources

What is there besides astrometry, radial velocity, and photometry?

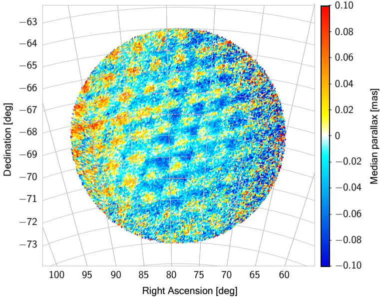
- Gaia Celestial Reference Frame
  - Materialized by $\sim$ 557 thousand QSOs identified from ALLWise
  - Aligned to ICRF-3 through subset of 2820 QSOs
- Astrophysical parameters for stars at $G \leq 17$
  - T_{eff}, $\sim$ 161 million
  - A_G and $E(G_{\mathrm{BP}} - G_{\mathrm{RP}})$, $\sim$ 88 million
  - Radius and bolometric luminosity, $\sim$ 77 million
- Variability information
  - Photometric time series for $\sim$ 551 thousand sources identified as variable
  - Classification for $\sim$ 364 thousand sources
    - RRL, LPV, Cep, δ Sct, SX Phe
  - Detailed characterization for $\sim$ 391 thousand sources
    - RRL, Cep, LPV, rotation modulation variables, short time scale variables
- Astrometric and photometric time series for $\sim$ 14 thousand minor planets

Gaia DR2 astrometry: uncertainties and systematic errors



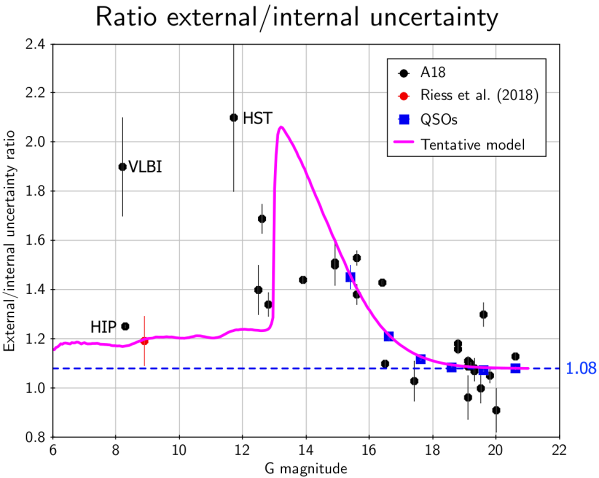
- Uncertainties are nearly Gaussian
  - NOTE: uncertainties on the astrometric parameters are correlated
- Dependencies on celestial position, magnitude, colour
- Systematic errors are present
  - non-zero mean of Gaussian uncertainty
  - dependencies on celestial position, magnitude, colour
  - spatially correlated

Median parallax LMC region



External (total) errors

Tentative calibration of external errors suggested in Lindegren et al. slide set:



$$\sigma_{\mathrm{ext}} = \sqrt{k^2\sigma_i^2 + \sigma_{\mathrm{s}}^2}$$

$k\sigma_i$: standard deviation of random error (formal estimate inflated by factor k)
σ_{s}: standard deviation of systematic error

Faint ($G \gtrsim 13$): $k = 1.08$, $\sigma_{\mathrm{s}} = 0.043$ mas

Bright ($G \lesssim 13$): $k = 1.08$, $\sigma_{\mathrm{s}} = 0.021$ mas

The model may be too pessimistic for $G \simeq 13$ to 15

Parallax zero-point (ϖ_0)

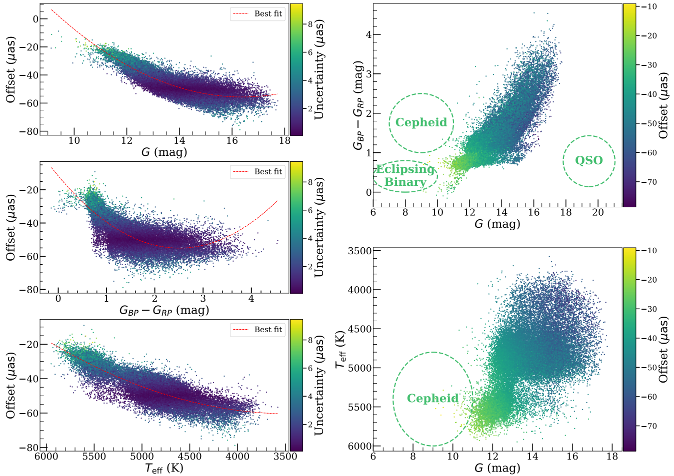
The zero-point ϖ_0 is the expected measured parallax for a source at infinity; it should thus be *subtracted* from the catalogue value.

As a global average $\varpi \equiv \langle s \rangle \simeq -0.03$ mas, but

- s definitely depends on (α, δ)
- s probably depends on G
- s may depend on $C = G_{\rm BP} - G_{\rm RP}$
- the dependence is probably multi-variate, $s(\alpha, \delta, G, C, \dots)$

No general recipe can be given for the correction of the zero-point

Parallax zero-point (ϖ_0)



Leung & Bovy: arXiv:1902.08634

- Simultaneous calibration of spectro-photometric distances and the Gaia DR2 parallax zero-point
- Illustrates variation with apparent brightness and colour
- Shows the importance of investigating the zero-point specifically for the sample of sources your are interested in
- See also Arenou et al.

Correlated uncertainties on the astrometric parameters

Distribution of measurements $\boldsymbol{a}$ for a given source is approximately multi-variate normal around mean $\boldsymbol{m}$:

$$p(\boldsymbol{a}|\boldsymbol{m},\mathbf{C}) = \mathcal{N}_k(\boldsymbol{m},\mathbf{C}) = \frac{1}{\sqrt{(2\pi)^k \det(\mathbf{C})}} \exp\left(-\frac{1}{2}(\boldsymbol{a}-\boldsymbol{m})'\mathbf{C}^{-1}(\boldsymbol{a}-\boldsymbol{m})\right)$$

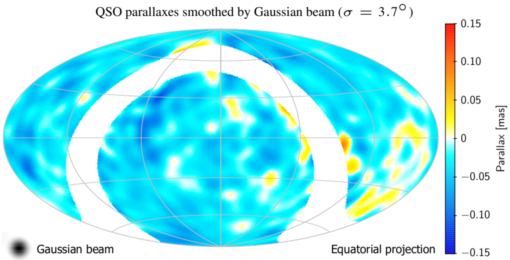
Uncertainty propagation:

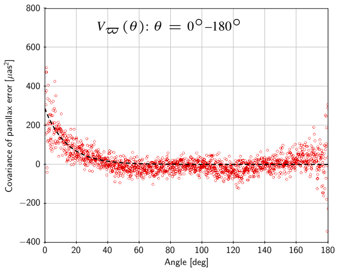
$$\boldsymbol{y} = \boldsymbol{f}(\boldsymbol{a}) \quad \longrightarrow \quad \mathbf{C}_{\boldsymbol{y}} = \mathbf{J}_f \mathbf{C}_{\boldsymbol{a}} \mathbf{J}_f' \qquad \mathbf{J}_{ij} = \frac{\partial f_i}{\partial a_j}$$

Account for covariances in your data analysis when:

- propagating uncertainties on subsets and/or linear combinations of astrometric parameters
- estimating model parameters: χ^2-fitting, maximum likelihood, Bayesian inference, etc
- sampling the astrometric uncertainties in some Monte Carlo procedure
  - usually better to sample in the astrometric parameters before transforming to, e.g., phase space quantities

Spatially correlated systematic errors





Example: QSO parallaxes $\{\varpi_i\} = \boldsymbol{\varpi}$ described by joint distribution for collection of k sources

$$p(\boldsymbol{\varpi}|\varpi_0, \mathbf{C}) = \mathcal{N}_k(\varpi_0, \mathbf{C}) =$$

$$\frac{1}{\sqrt{(2\pi)^k \det(\mathbf{C})}} \exp\left(-\frac{1}{2}(\boldsymbol{\varpi} - \varpi_0)'\mathbf{C}^{-1}(\boldsymbol{\varpi} - \varpi_0)\right)$$

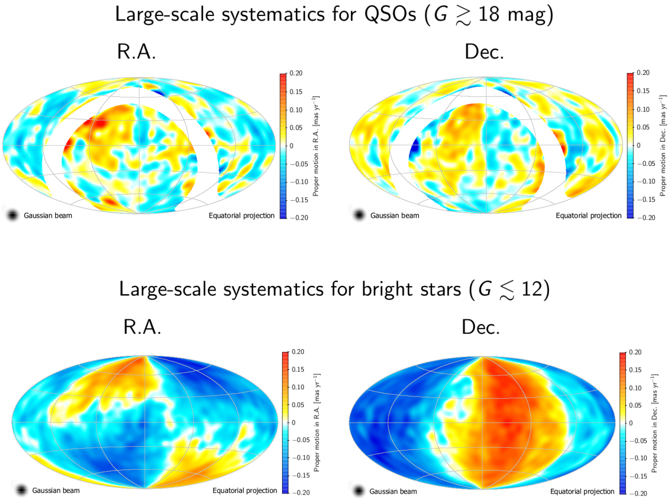
$\mathbf{C}$ is now the joint covariance matrix, with $\mathbf{C}_{ii} = \sigma^2_{\varpi,i} + V_\varpi(0)$ and $\mathbf{C}_{ij} = V_\varpi(\theta_{ij})$ ($i \neq j$), where one choice for modelling the spatial covariance function $V_\varpi(\theta)$ could be:

$$V_\varpi(\theta) = V_\varpi(0) \exp\left(-\theta/\tau\right) \ ,$$

where $V_\varpi(0) = \sigma^2_{\varpi,\mathrm{s}}$

See the Lindegren et al. slide set for more details

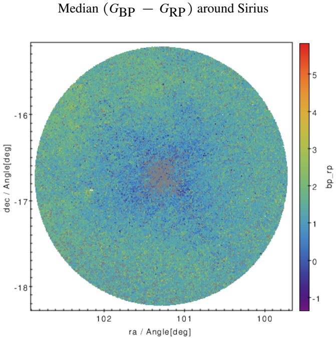
Proper motion systematics

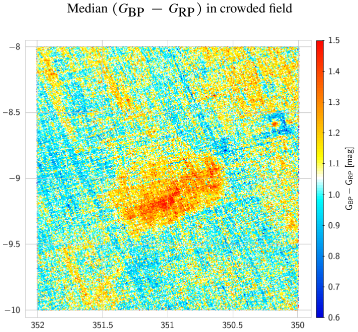


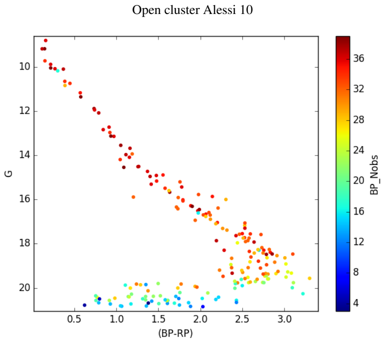
- Bright star systematics from comparison to Hipparcos-Gaia proper motions
- Note global rotation pattern of $\simeq 0.15$ mas yr^{-1}
- See Lindegren et al. slide set for suggested correction
- Applies *only* to bright sources, no net rotation at faint end

More details in Lindegren et al. slide set, including estimates of $V_{\mu}(\theta)$

Gaia DR2 photometry: flux excess issue







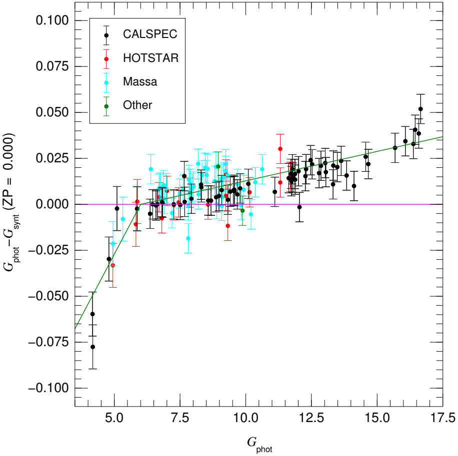
Figures from Arenou et al.

- For normal source SEDs expect: $F_{BP} + F_{RP} \approx F_G$
- Colours suffer from insufficiently accurate background characterization
  - crowded regions, near bright stars, faint sources at $G > 19$
  - use `phot_bp_rp_excess_factor` for photometric quality filtering
  - examples in Gaia Collaboration, Babusiaux, et al. and Lindegren et al.

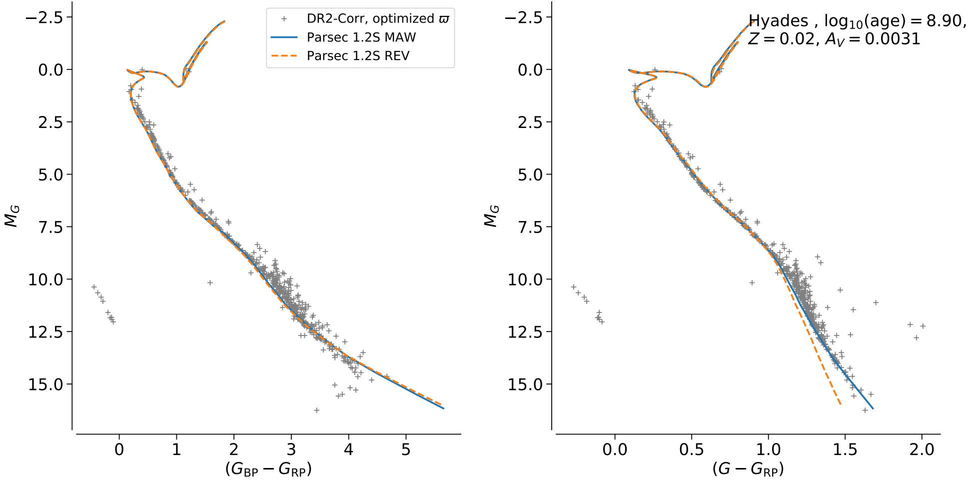
Gaia DR2 photometry: pass-bands

See the Gaia known issues pages for details

- Recommended pass-bands for synthetic photometry are those from Maíz-Apellániz & Weiler
- Use these with a slightly corrected version G' of the catalogue G
- See the above link for details
- When using stellar tracks/isochrones check carefully which passbands were implemented to predict Gaia DR2 photometry
- NOTE: there are two BP pass-bands defined in Maíz-Apellániz & Weiler, for $G < 10.87$ and $G > 10.87$
- See also Arenou et al.



Gaia DR2 photometry: pass-bands

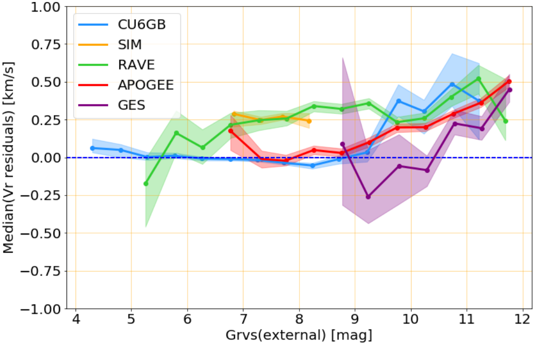


MAW: Maíz-Apellániz & Weiler pass-bands, REV: Evans et al. Gaia DR2 pass-bands.

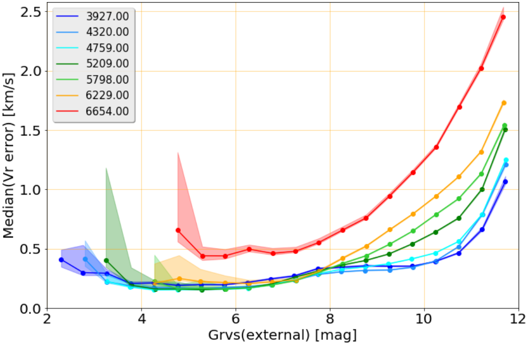
Hyades data from Gaia Collaboration, Babusiaux, et al.

Gaia DR2 radial velocities

Radial velocity accuracy

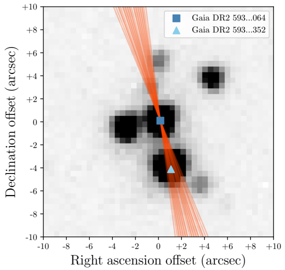


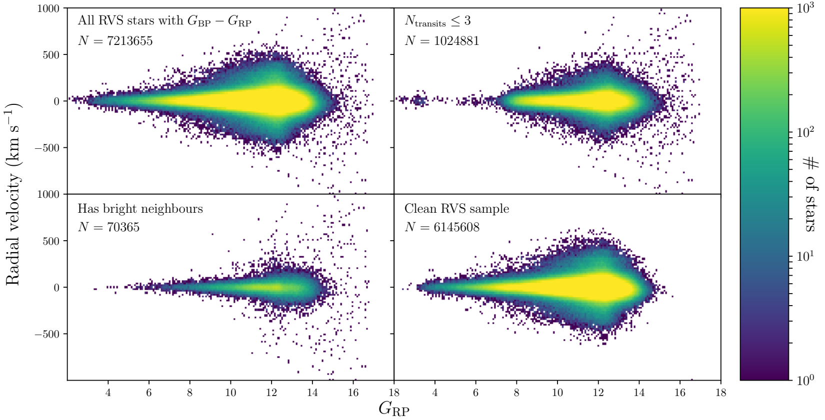
Radial velocity precision



- Radial velocity residuals with respect to other surveys reflect a magnitude term in RVS results as well as systematic errors in the other surveys
- Radial velocities only for sources at $3550 \lesssim T_{\rm eff} \lesssim 6900$ K (this is DR2-specific!)
- Details: Katz et al., Sartoretti et al., Soubiran et al.

Gaia DR2 radial velocities

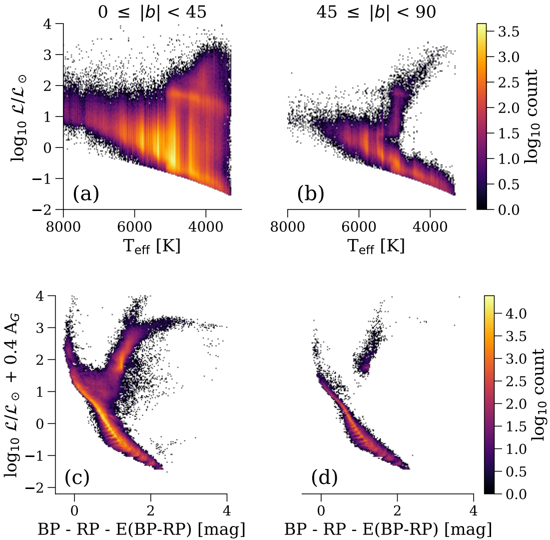




Boubert et al., 2019, arXiv:1901.10460

- See known issues pages for details on potentially spurious radial velocities
- Be careful when examining tails of velocity distributions
- For your favourite star, do not blindly apply Boubert et al. filters, but examine the case in detail

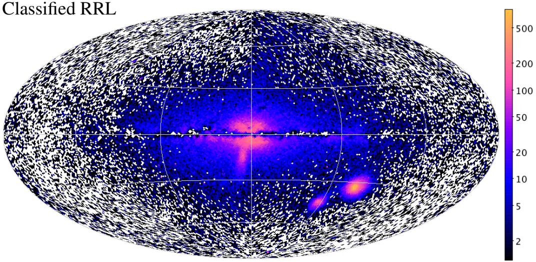
Gaia DR2 Astrophysical parameters



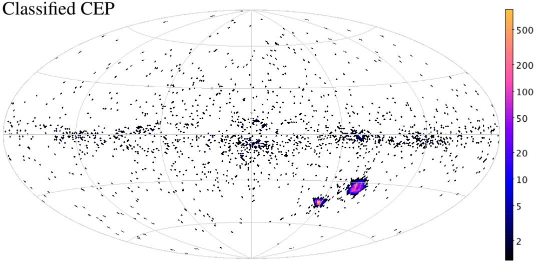
- Determination $T_{\rm eff}$, A_G, $E(G_{\rm BP} - G_{\rm RP})$, $\mathcal{L}$, $\mathcal{R}$, based *only* on G, $G_{\rm BP}$, $G_{\rm RP}$, and parallax
  - Strong $T_{\rm eff}$ - A_G degeneracy in broad-band colours necessitates strong assumptions
  - Asymmetric uncertainties, positivity constraint on A_G
  - $T_{\rm eff}$ estimates constrained to 3300–8000 K
  - Radius/luminosity estimation assumes $A_G = 0$ (correction to non-zero A_G possible)
  - Results to be interpreted with care
- See Andrae et al. and online documentation

Variable stars in Gaia DR2

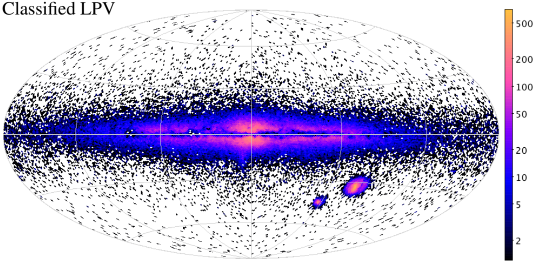
Classified RRL



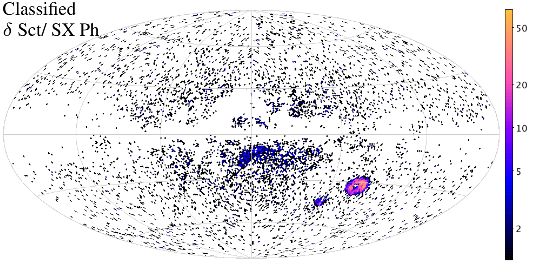
Classified CEP



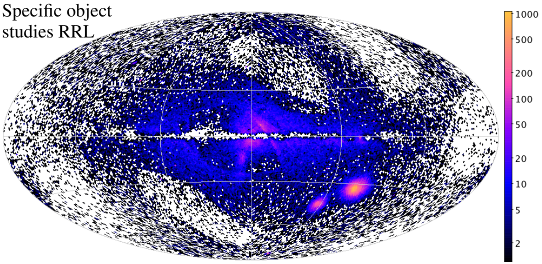
Classified LPV

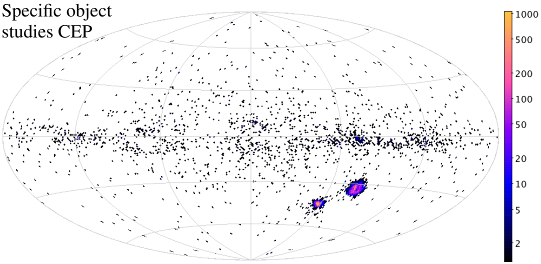


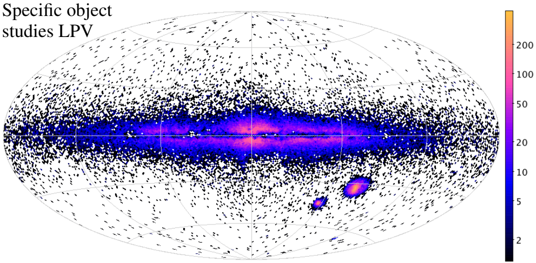
Classified
δ Sct/ SX Ph



Variable stars in Gaia DR2







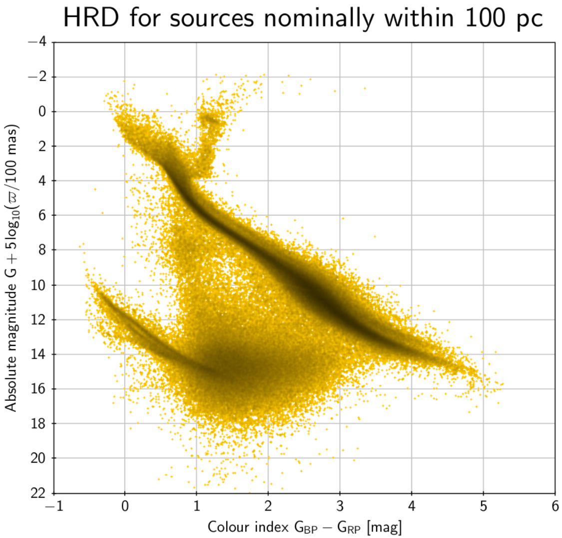
- 551 thousand variables identified
  - many more to come in future
- Subset classified by variability type
  - based on 2+ transits
- Overlapping subset studied in detail
  - based on 12+ transits

Data quality filtering: astrometry and photometry

Goal: get rid of sources with unreliable astrometry (parallaxes) and/or photometry

- Use combination of filtering on Re-normalized Unit Weight Error (RUWE), visibility periods, photometry signal to noise, and flux excess factor
- Astrometric goodness of fit indicators in catalogue (`astrometric_nobs_al`, `astrometric_gof_al`, `astrometric_chi2_al`, `astrometric_excess_noise`, `astrometric_excess_noise_sig`) are not recommended
- RUWE can be calculated from catalogue quantities
  - supersedes UWE filter from Lindegren et al. paper on Gaia DR2 astrometry
  - details in Lindegren et al. slide set
  - data tables available from known issues pages

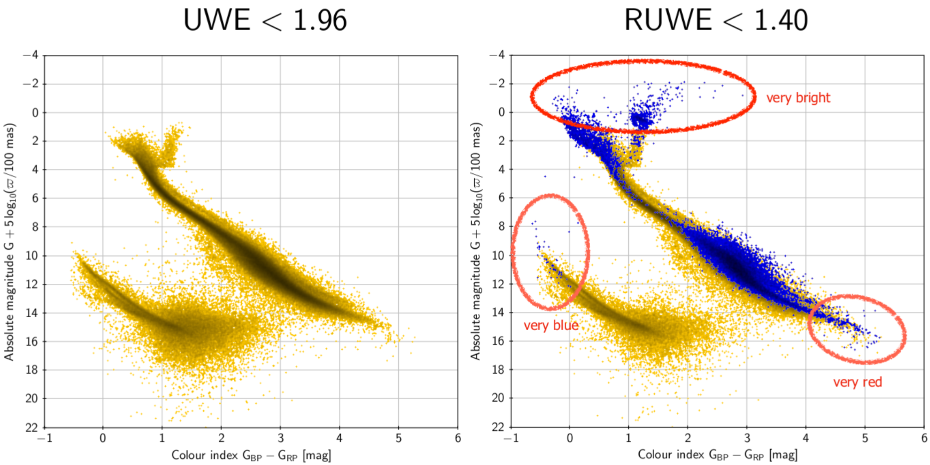
Illustration of the use of RUWE



Selection:

- $\varpi > 10$ mas
- $\varpi/\sigma_\varpi > 10$
- Signal to noise in BP and RP larger than 10
- No filtering on goodness of fit indicators

Illustration of the use of RUWE



- Filtering by RUWE gives cleaner HRD
- Blue dots are sources missing in left diagram
- Experiment to decide on the limit in RUWE for you application!

Responsible use of parallaxes

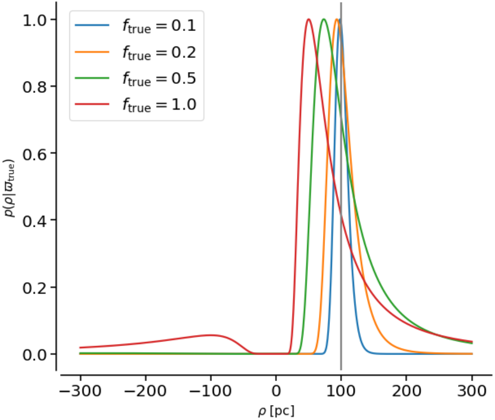
See paper by Luri et al. for the details and examples worked out as Python or R notebooks, available at
`https://github.com/agabrown/astrometry-inference-tutorials`

Why can't I invert the parallax?

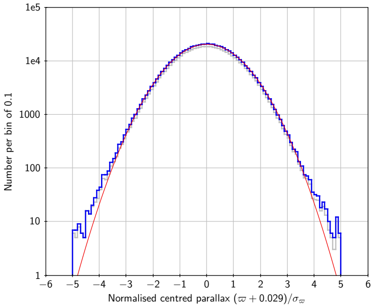
Naive estimate for distance $\rho = 1/\varpi$

$$p(\rho \mid \varpi_{\text{true}}) = \frac{1}{\rho^2 \sigma_\varpi \sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{1/\rho - \varpi_{\text{true}}}{\sigma_\varpi}\right)^2\right)$$

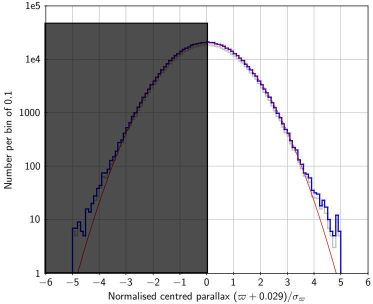
- PDF of ρ has nonphysical negative tail
- Mode moves away from true value of parallax as $f_{\text{true}} = \sigma_\varpi/\varpi_{\text{true}}$ increases
- Expectation value and variance are undefined
- PDF expressed in terms of *unknown* value of ϖ_{true}
- Statements above also hold for small relative uncertainties



Okay, so I keep only positive parallaxes with small uncertainties?

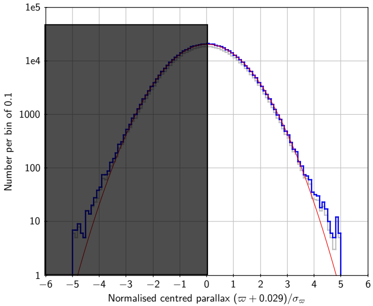


Okay, so I keep only positive parallaxes with small uncertainties?

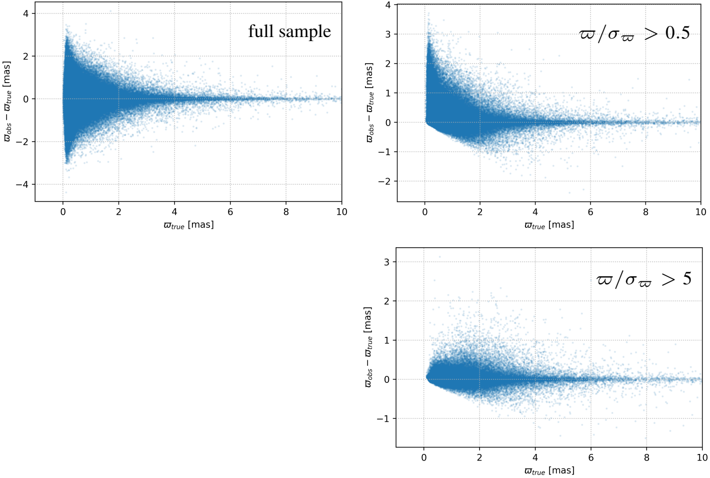


After discarding negative parallaxes
average QSO parallax is 0.8 mas (!)

Okay, so I keep only positive parallaxes with small uncertainties?

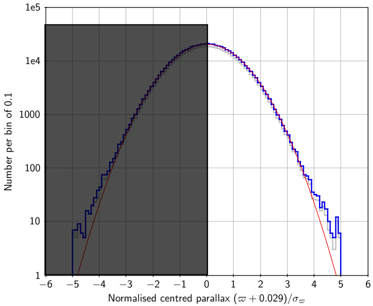


After discarding negative parallaxes average QSO parallax is 0.8 mas (!)

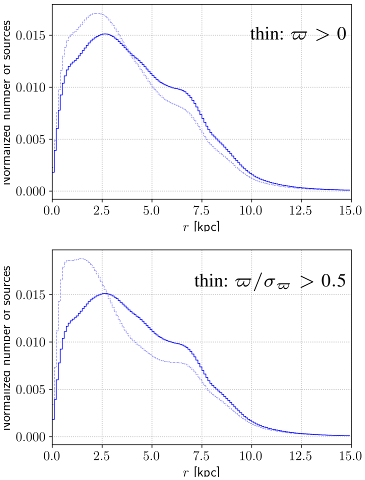


Truncation on the data values distorts the underlying sample and will bias the interpretation

Okay, so I keep only positive parallaxes with small uncertainties?

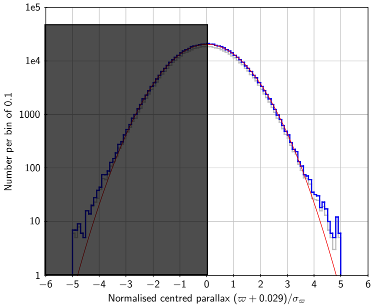


After discarding negative parallaxes average QSO parallax is 0.8 mas (!)

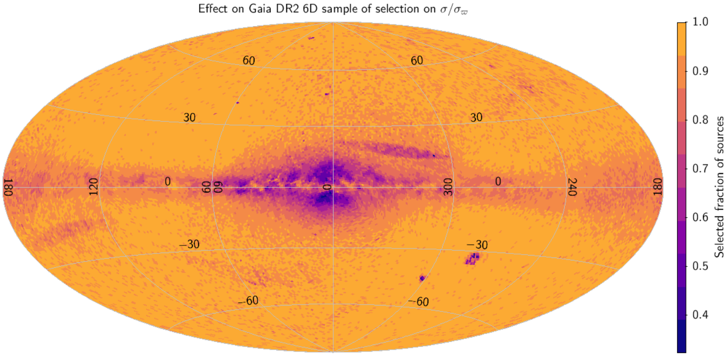


Truncation on the data values distorts the underlying sample and will bias the interpretation

Okay, so I keep only positive parallaxes with small uncertainties?



After discarding negative parallaxes average QSO parallax is 0.8 mas (!)

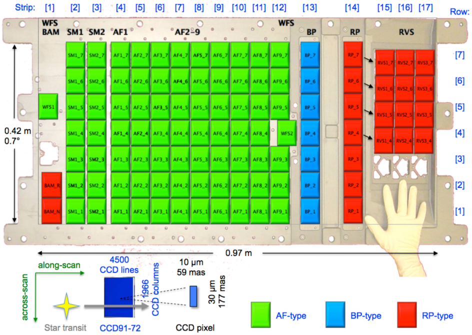


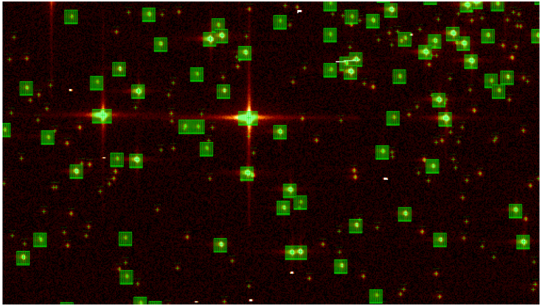
Truncation on the data values distorts the underlying sample and will bias the interpretation

So what should I do?

- Treat the derivation of quantities or model parameters from the astrometric data as an inference problem
- Where possible formulate the problem in the data space
  - data uncertainties well understood
  - easier handling of covariances in measured quantities
  - quantities to be inferred are parameters in ‘forward model’
- Use all relevant information
  - proper motions, magnitudes, colours, all contain distance information
- Account for data selection, survey completeness
- Bayesian analysis naturally fits above points
  - Use proper priors (such that posterior is normalized) that represent the information you already have
- Maximum likelihood as alternative is fine when you have large amounts of data or very precise measurements
- For *initial exploration* of the problem it is okay to select the ‘best data’
  - beware sample truncation effects!
- Is the *distance* really of interest to the question you are trying to answer?

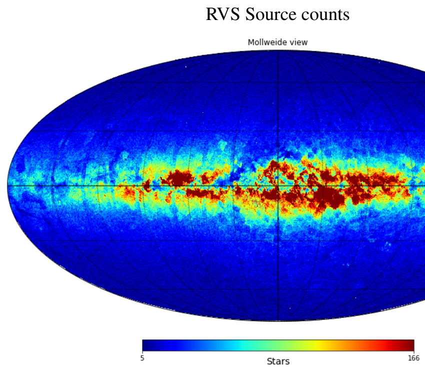
Ingredients of Gaia DR2 selection function

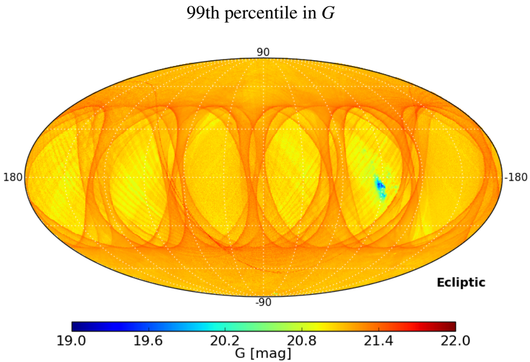




- Detection of sources in SM1/2, confirmation in AF1 (as cosmic ray/ spurious source rejection step)
- Strict flux threshold applied, but in presence of magnitude estimation errors
- Limit on source ‘size’
- Management of observation window conflicts and resource limitations affects completeness in crowded fields (above few 100k stars/degree^2)
- Details in Gaia Collaboration, Prusti et al., de Bruijne et al.

Ingredients of Gaia DR2 selection function





- Imprints from Initial Gaia Source List in RVS counts (will disappear in future)
- Brighter magnitude limit in crowded fields
- Imprint of combination of scan law pattern and data quality filtering
- Knock-on effects of data quality filtering during data processing (also follows scan law)

Epoch propagation and cross-matching to other catalogues

- Gaia DR2 has high spatial resolution, down to 0.4–0.5 arcsec (PSF is ~ 0.1 arcsec)
  - most other catalogues are of lower resolution so beware of blended sources
- Gaia DR2 reference epoch is 2015.5
  - propagate Gaia positions to epoch of other catalogue before doing positional match
  - requires proper motions (and parallax and radial velocity for rigorous treatment)
    - if not available use proper motion dispersion for the source population being matched to estimate positional uncertainties at other epoch
  - reference system for modern catalogues is ICRS; no need to worry about precession, nutation, etc
  - all maths, including propagation of covariance matrix, in Gaia DR2 online documentation
  - see `astropy.coordinates.SkyCoord.apply_space_motion()`
- Pre-computed cross-matches to large catalogues available from Gaia archive
  - these are carefully done positional matches, not necessarily complete
  - details in Marrese et al.
- Convenient tools offered by Topcat and CDS x-match service

Entry points to Gaia literature

<u>Mission, spacecraft, payload, data processing and validation</u>

Gaia DR1 A&A special issue Data processing and validation
https://www.aanda.org/component/toc/?task=topic&id=641

Gaia DR2 A&A special issue Data processing and validation
https://www.aanda.org/component/toc/?task=topic&id=922

Mission, instruments, and data processing overview
https://doi.org/10.1051/0004-6361/201629272

RVS detailed description https://doi.org/10.1051/0004-6361/201832763

On-board detection capabilities https://doi.org/10.1051/0004-6361/201424018

In-orbit CCD performance https://doi.org/10.1051/0004-6361/201628990

Documentation http://gea.esac.esa.int/archive/documentation/index.html

Entry points to Gaia literature

Description of Gaia data products, mostly pre-launch

Gaia broad-band photometry `https://doi.org/10.1051/0004-6361/201015441`

Astrophysical parameters `https://doi.org/10.1051/0004-6361/201322344`

Astrophysics from RVS `https://doi.org/10.1051/0004-6361/201425030`

Double and multiple stars `http://dx.doi.org/10.1063/1.3597594`

Variable stars `https://doi.org/10.1051/eas/1567012`

Solar system `https://doi.org/10.1016/j.pss.2012.03.007` and
`https://doi.org/10.1016/j.pss.2015.11.009`

Galaxy morphology with Gaia `https://doi.org/10.1051/0004-6361/201219697`

Source environment analysis `https://doi.org/10.1007/s10686-011-9240-7`

Transient astronomy `https://doi.org/10.1098/rsta.2012.0239`

Simulated Gaia data `https://doi.org/10.1051/0004-6361/201118646` and
`https://doi.org/10.1051/0004-6361/201423636`

Entry points to Gaia literature

Astrometry with Gaia

Astrometric Global Iterative Solution https://www.aanda.org/articles/aa/abs/2012/02/aa17905-11/aa17905-11.html

Tycho-Gaia Astrometric Solution https://doi.org/10.1051/0004-6361/201425310