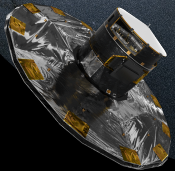
The ins and outs of Gaia DR2

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’

GAIA DR2 INFO

Information on Gaia Data Release 2 contents, completeness and limitations.

GAIA DR2 PAPERS

Titles and links to papers describing the data processing and demonstrating the science potential of Gaia Data Release 2.

GAIA DR2 DOCUMENTATION

The full documentation for the second data release, both on webpages and with a downloadable PDF-file

GAIA DR2 DATA

Gaia Data Release 2 data is now available.

GAIA DATA CREDITS

When using Gaia data, please acknowledge the work of the people involved and provide credits and necessary citations.

GAIA DR2 KNOWN ISSUES

Issues with the Gaia DR2 data important for the users to know that were discovered after the release of data and documentation

TUTORIALS AND HELP

Help is available to guide you through the process of getting the data you need. Check out the tutorials as they are very instructive!

LEARN ADQL

Gaia Data Release 2 contains a lot of data. While downloading the data will be possible, you can also bring your code to the data and access the data in a smart way. You can use ADQL queries to extract the data and then download the resulting table.

GAIA DR2 VIRTUAL REALITY RESOURCES

An overview of some visualisations and virtual reality resources available for exploring the Gaia data.

GAIA DR2 MEDIA STORIES

Here links to a selection of media stories on Gaia Data Release 2 can be found.

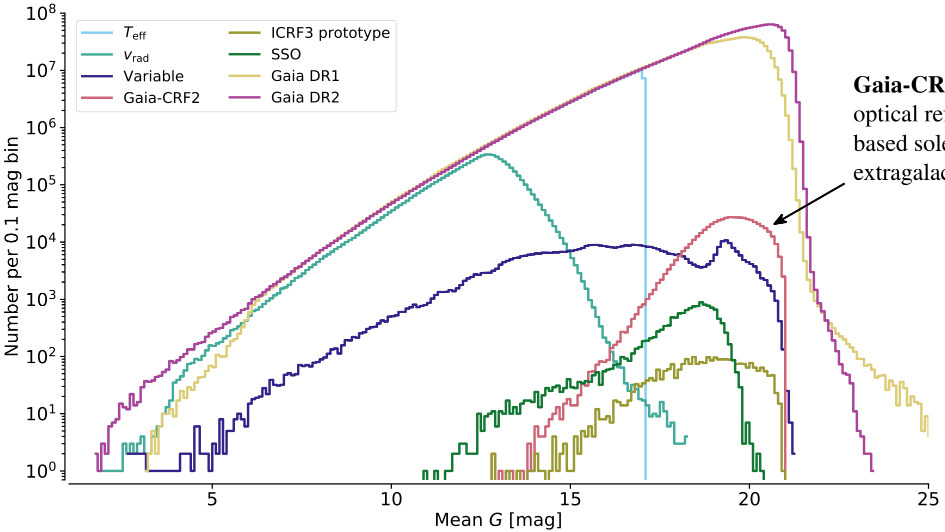
GAIA DR2 IN-DEPTH STORIES

A selection of in-depth stories on the processing towards data release 2 and the science potential of the data.

GAIA'S FAMILY PORTRAIT

An interactive visualisation of the Gaia Data Release 2 Hertzsprung-Russell diagram

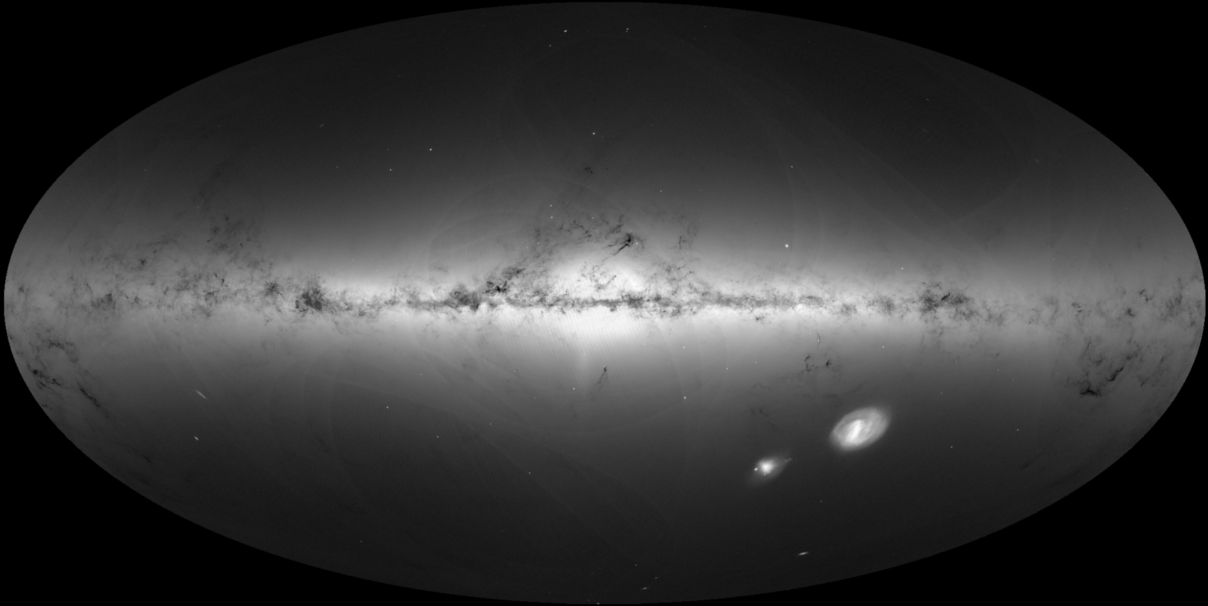
What is there besides astrometry, radial velocity, and photometry?



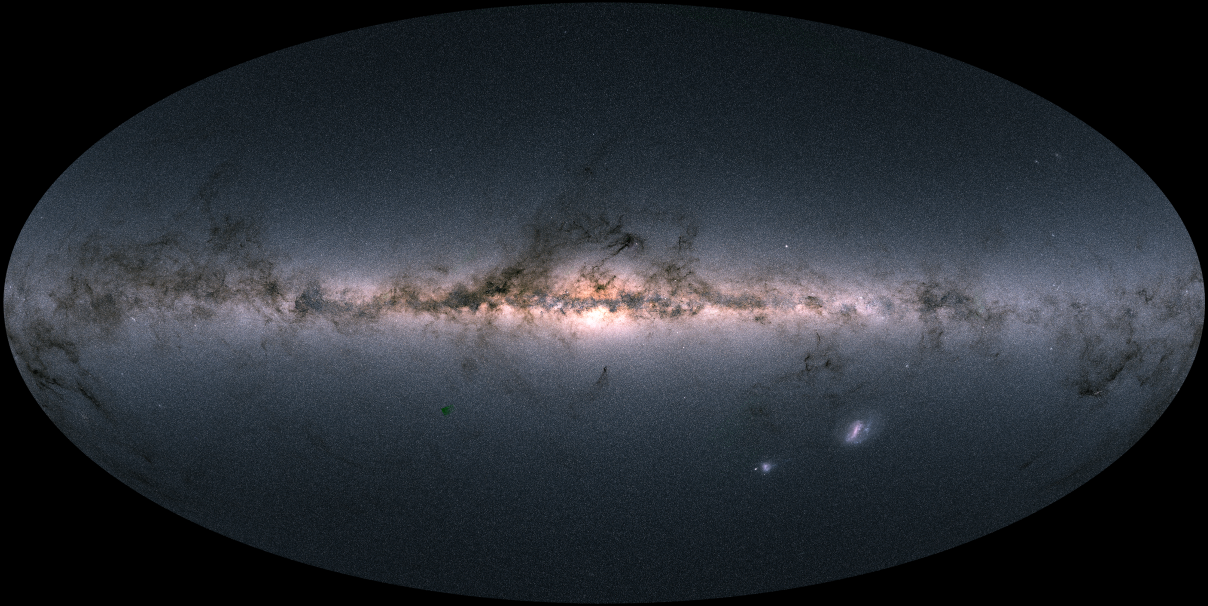
Gaia-CRF2: First optical reference frame based solely on extragalactic sources

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

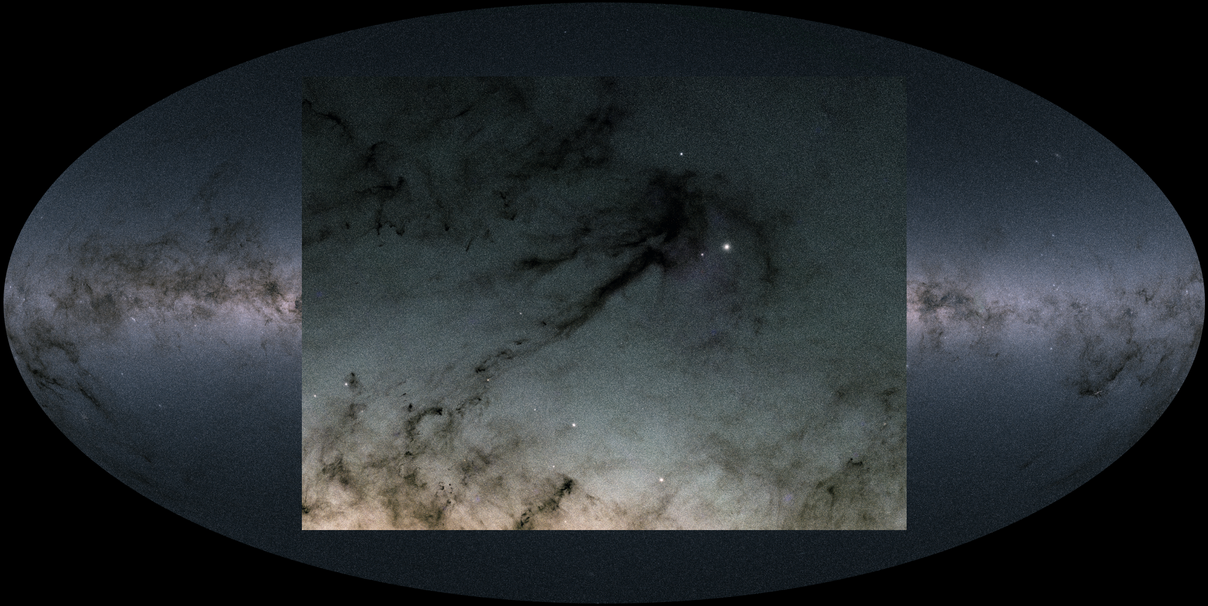
- Gaia Celestial Reference Frame
  - Materialized by $\sim$ 557 thousand QSOs identified from ALLWise
  - Aligned to ICRF-3 through subset of 2820 QSOs
- Astrophysical parameters for stars at $G \leq 17$
  - $T_{\rm eff}$, $\sim$ 161 million
  - A_G and $E(G_{\rm BP} - G_{\rm RP})$, $\sim$ 88 million
  - Radius and bolometric luminosity, $\sim$ 77 million
- Variability information
  - Photometric time series for $\sim$ 551 thousand sources identified as variable
    - NOTE: source not explicitly listed as VARIABLE in Gaia DR2 is not necessarily constant
  - 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



ESA/Gaia/DPAC

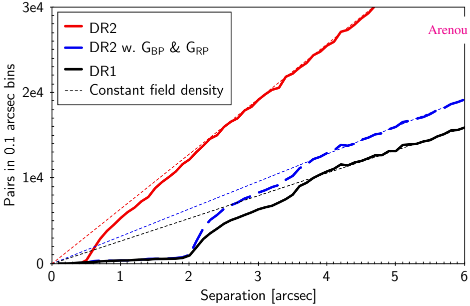


ESA/Gaia/DPAC

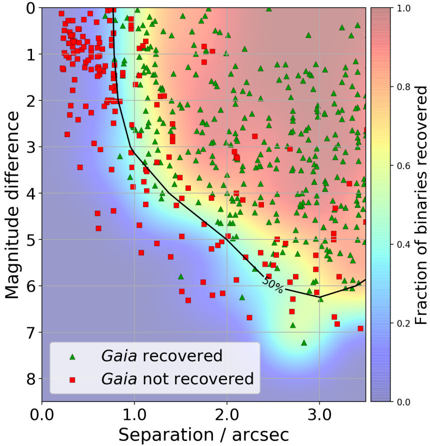


ESA/Gaia/DPAC

Gaia DR2 spatial resolution



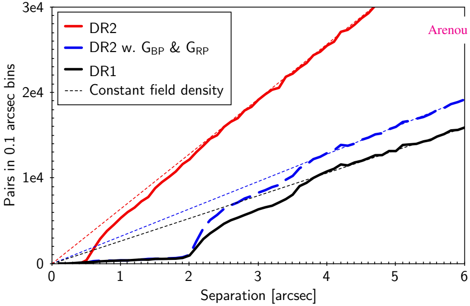
Arenou et al. 2018



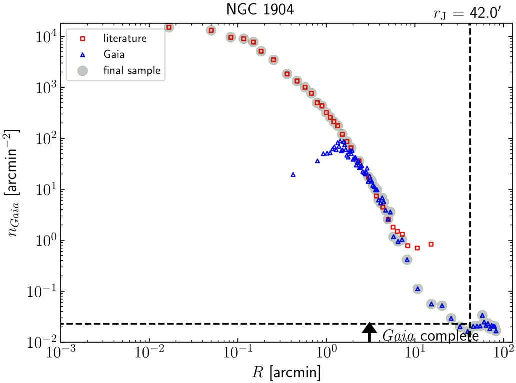
Ziegler et al. 2018

- Angular resolution limited to 0.4–0.5 arcsec
  - significantly better than all existing ground-based surveys
- Will get better in later data releases
  - currently no treatment of crowded sources, including binaries that are in principle resolved by Gaia

Gaia DR2 spatial resolution



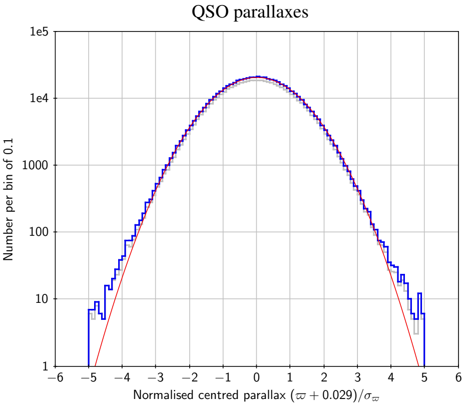
Arenou et al. 2018



de Boer et al. 2019

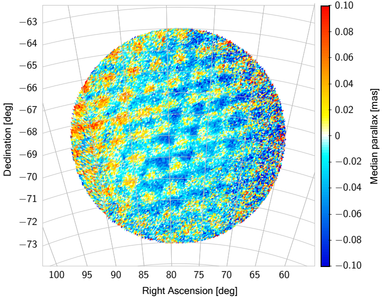
- Angular resolution limited to 0.4–0.5 arcsec
  - significantly better than all existing ground-based surveys
- Will get better in later data releases
  - currently no treatment of crowded sources, including binaries that are in principle resolved by Gaia

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



Images: Lindegren et al. (2018)

Random and systematic errors

A useful model for the total (external) error in parallax for source i is

$$\varpi_i^{\mathrm{DR2}} - \varpi_i^{\mathrm{true}} = r_i + s(\alpha, \delta, G, C, \dots) \tag{1}$$

Random error r_i:

- On average zero, uncorrelated between different sources
- Formal uncertainty σ_i is a (possibly underestimated) estimate of its standard deviation: $\sigma_r = k\sigma_i$, with correction factor $k \gtrsim 1.0$

Systematic error s:

- May depend on several variables (position, magnitude, colour, ...)
- Same for sources with sufficiently similar position, magnitude, etc
- Mean value is the parallax zero point ϖ_0
- Variance is σ_s^2

Random and systematic errors

In this model the external (total) uncertainty becomes

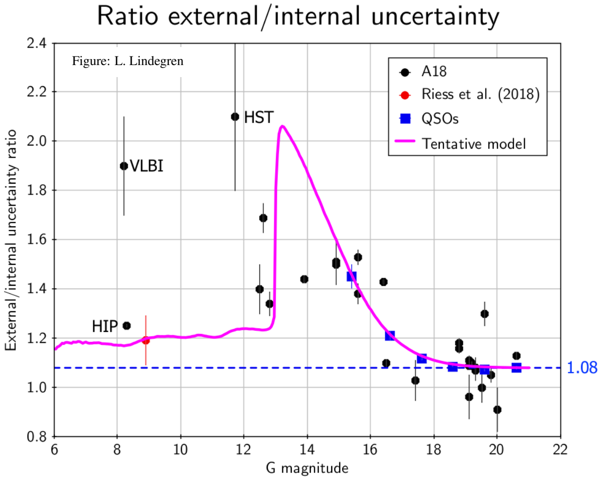
$$\sigma_{\text{ext}} = \sqrt{k^2\sigma_i^2 + \sigma_s^2} \tag{1}$$

- Astrophysical applications using likelihood or Bayesian methods require the probability density of the total error $e_i = \varpi_i^{\text{DR2}} - \varpi_i^{\text{true}}$
- Most conservative assumption:
  e_i is Gaussian with mean value ϖ_0 and standard deviation σ_{ext}

- External data must be used to 'calibrate' the model by estimating ϖ_0, k, and σ_s
  - more details in Lindegren et al. slide set
- Alternatively these parameters can be part of the forward model in a likelihood or Bayesian analysis and inferred along with the parameters of interest
  - See Sesar et al., ApJ, 2017 (arXiv:1611.07035) for a useful example

External (total) errors

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



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

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

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

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

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

Parallax zero-point (ϖ_0)

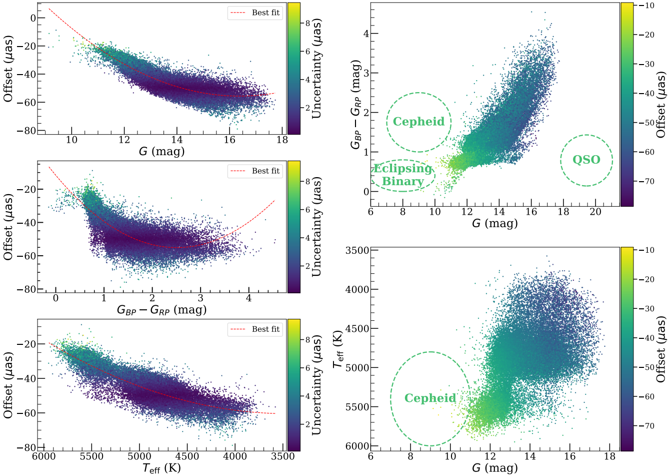
The zero-point ϖ_0 is the expected measured parallax for a source at infinity; it should thus be *subtracted* from the catalogue value.

As a global average $\varpi_0 \equiv \langle s \rangle \simeq -0.03$ mas, but

- s definitely depends on (α, δ)
- s probably depends on G
- s may depend on $C = G_{\mathrm{BP}} - G_{\mathrm{RP}}$
- 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

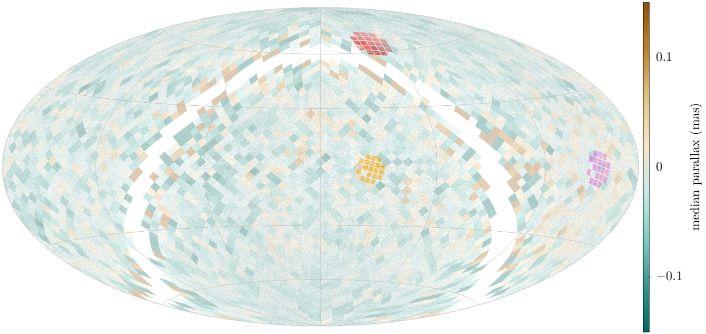
Parallax zero-point (ϖ_0)



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.

Parallax zero-point (ϖ_0)



Khan et al.: arXiv:1904.05676

- Comparison of asteroseismic and Gaia DR2 parallaxes in Kepler field and two K2 fields
  - Kepler RGB/ RC: -51.7 ± 0.8 μas/ -47.9 ± 0.9 μas
  - K2-C3 red giants: -6.4 ± 3.8 μas
  - K2-C6 red giants: -16.9 ± 2.4 μas
- Spatial variations consistent with mean QSO parallaxes

Correlated uncertainties on the astrometric parameters

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

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

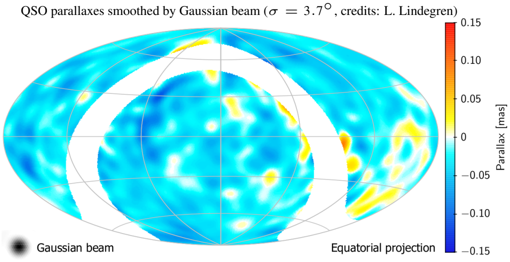
Uncertainty propagation:

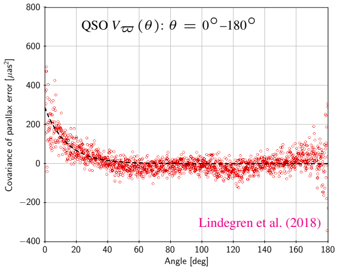
$$\boldsymbol{y} = \boldsymbol{f}(\boldsymbol{a}) \quad \longrightarrow \quad \mathsf{C}_{\boldsymbol{y}} = \mathsf{J}_f \mathsf{C}_{\boldsymbol{a}} \mathsf{J}_f' \qquad \mathsf{J}_{ij} = \frac{\partial f_i}{\partial a_j}$$

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 n sources

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

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

$\mathbf{C}$ is now the joint covariance matrix, with $\mathbf{C}_{ii} = k^2\sigma^2_{\varpi,i} + V_\varpi(0)$ 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:

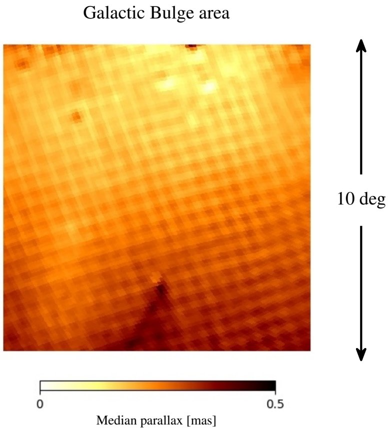
$$V_\varpi(\theta) = V_\varpi(0)\exp\left(-\theta/\tau\right) ,$$

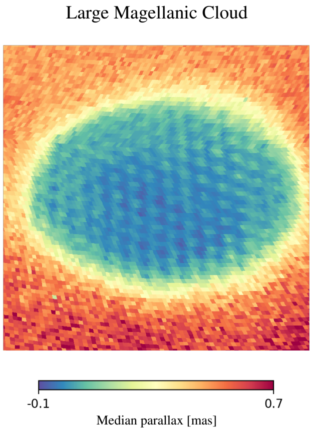
where $V_\varpi(0) = \sigma^2_{\varpi,\mathrm{s}}$

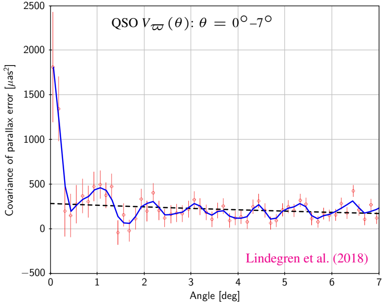
See the Lindegren et al. slide set for more details

Systematics $s(\alpha, \delta)$ on small scales

Quasi-periodic patterns imprinted by the Gaia scanning law (Arenou et al. 2018)

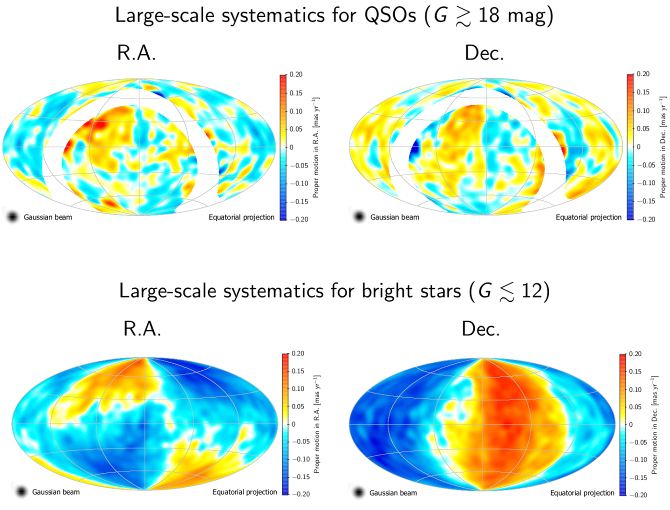






Characteristic period $\simeq$ 0.6 deg, RMS variation $\simeq$ 0.02–0.04 mas

Proper motion systematics



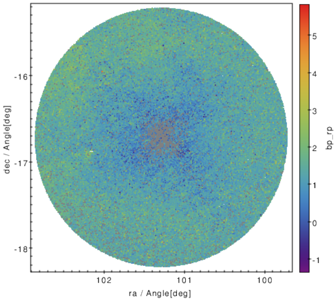
Figures: L. Lindegren

- 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

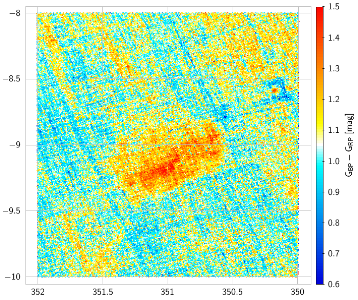
More details in Lindegren et al. slide set, including estimates of $V_\mu(\theta)$

Gaia DR2 photometry: flux excess issue

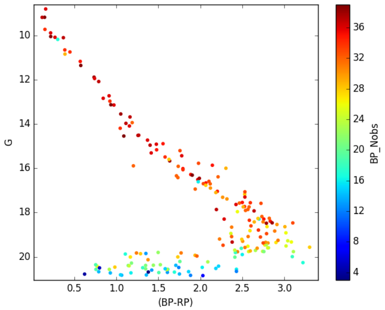
Median $(G_{\rm BP} - G_{\rm RP})$ around Sirius



Median $(G_{\rm BP} - G_{\rm RP})$ in crowded field



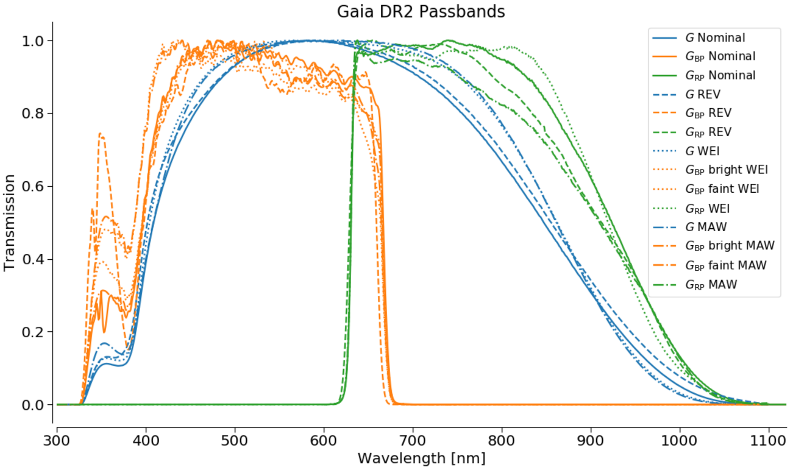
Open cluster Alessi 10



Figures from Arenou et al.

- For normal source SEDs expect: $F_{\rm BP} + F_{\rm RP} \approx F_G$
- 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

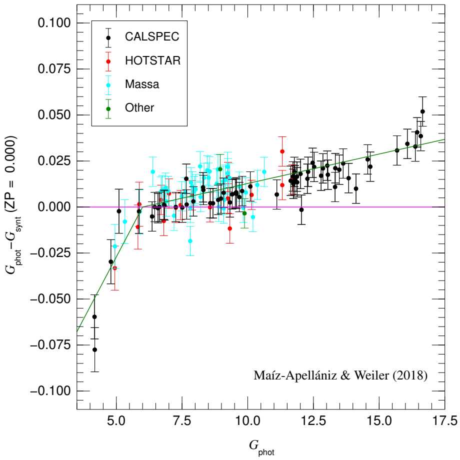


Nominal Gaia DR2 passbands and the three passband sets determined from the data

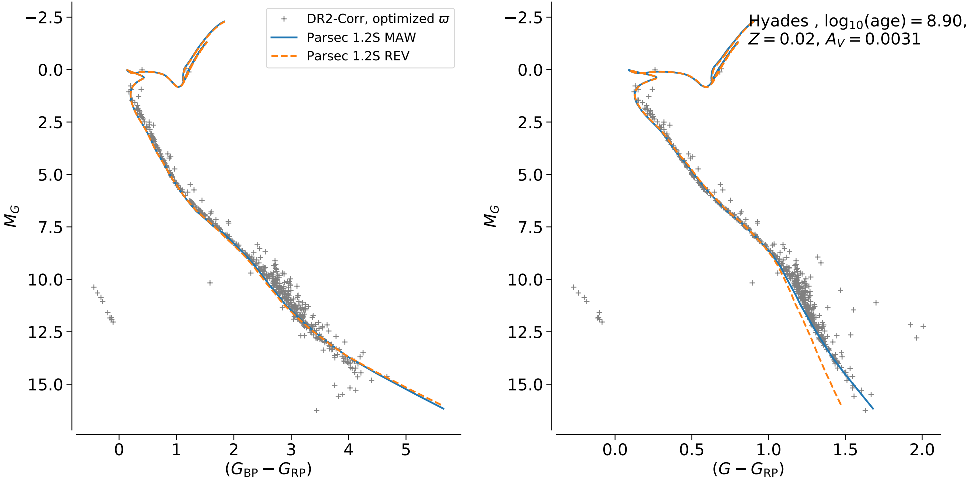
Gaia DR2 photometry: pass-bands

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

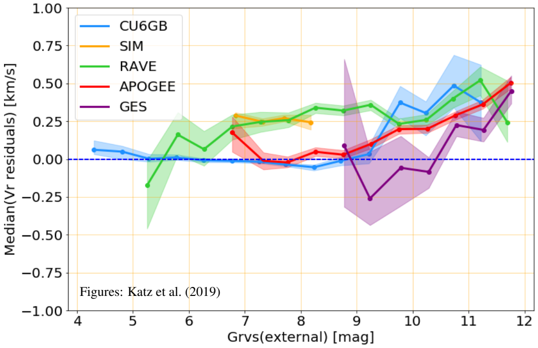


MAW: Maíz-Apellániz & Weiler pass-bands, REV: Evans et al. Gaia DR2 pass-bands.

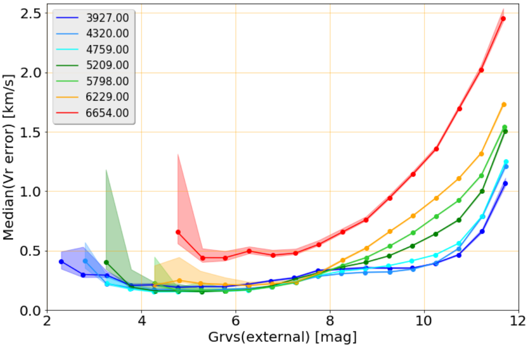
Hyades data from Gaia Collaboration, Babusiaux, et al.

Gaia DR2 radial velocities

Radial velocity accuracy

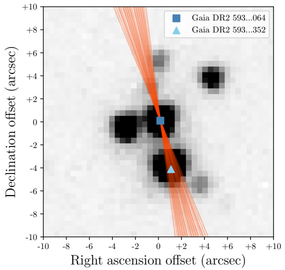


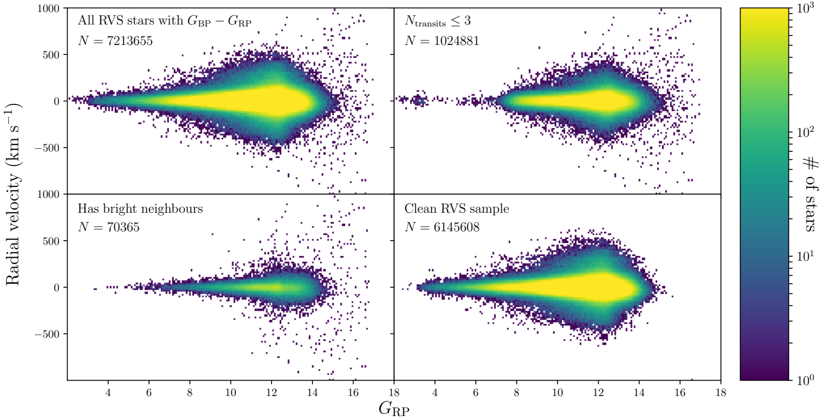
Radial velocity precision



- 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

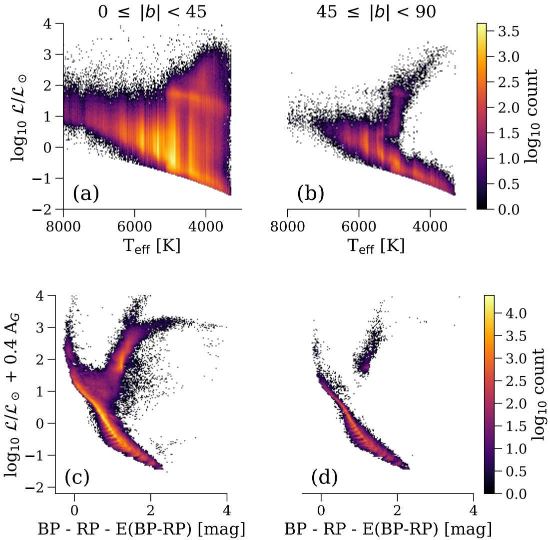




Boubert et al., 2019, arXiv:1901.10460

- 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

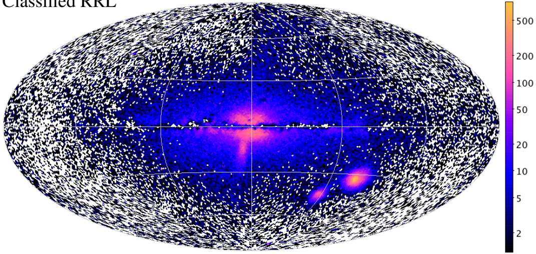


Andrae et al. (2018)

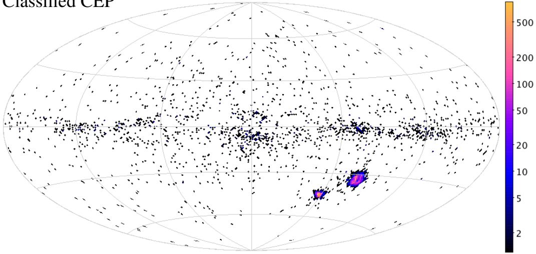
- Determination $T_{\rm eff}$, A_G, $E(G_{\rm BP} - G_{\rm RP})$, $\mathcal{L}$, $\mathcal{R}$, based *only* on G, $G_{\rm BP}$, $G_{\rm RP}$, and parallax
  - Strong $T_{\rm eff}$ - A_G degeneracy in broad-band colours necessitates strong assumptions
  - Asymmetric uncertainties, positivity constraint 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

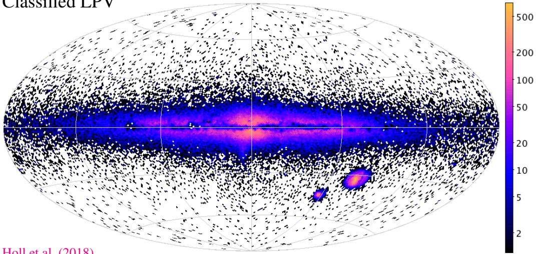
Classified RRL



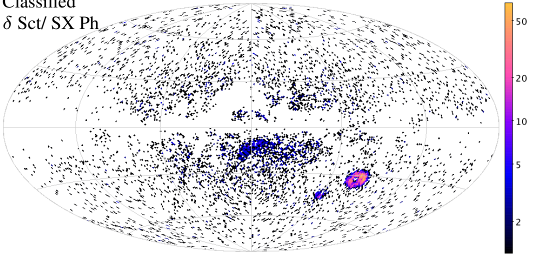
Classified CEP



Classified LPV

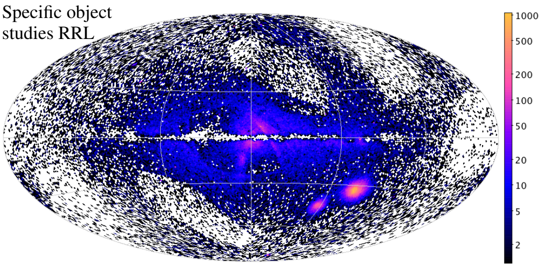


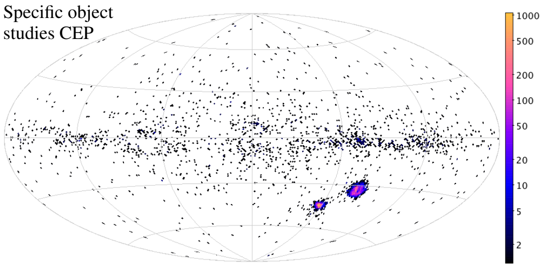
Classified
δ Sct/ SX Ph

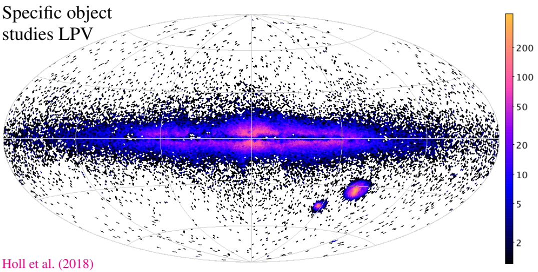


Holl et al. (2018)

Variable stars in Gaia DR2







Holl et al. (2018)

- 551 thousand variables listed in Gaia DR2
  - many more to come in future
- Subset classified by variability type
  - based on 2+ transits
- Overlapping subset studied in detail
  - based on 12+ or 20+ transits

Data quality filtering

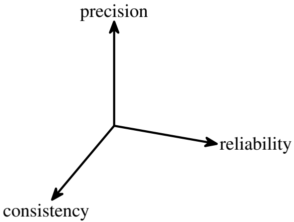
Goal: remove sources with potentially unreliable Gaia DR2 data from your sample

- All data products come with qualiy indicators
  - refer to relevant papers in the A&A Special Issue on Gaia DR2
- Further advice on data quality assessment on the Gaia DR2 Known Issues pages
  - Photometry and radial velocity issues were discussed already
  - More on astrometry quality filtering on next slides

Do not blindly apply filtering recipes. Beware the sample selection/truncation effects.

For a worked example of filtering the '6D' sub-sample of Gaia DR2 on astrometric, photometric, and radial velocity quality see `https://github.com/agabrown/gaiadr2-6dgold-example`.

Quality indicators for the astrometry



Precision `parallax_error`, `pmra_error`, `pmdec_error`, etc $\longrightarrow$ OK

Reliability `visibility_periods_used` ($\geq$ 6 for 5-parameter solutions) $\longrightarrow$ OK

Consistency (goodness of fit to the 5-parameter model)

$$\left.\begin{array}{l}\texttt{astrometric\_n\_bad\_obs\_al}\\ \texttt{astrometric\_gof\_al}\\ \texttt{astrometric\_chi2\_al}\\ \texttt{astrometric\_excess\_noise}\\ \texttt{astrometric\_excess\_noise\_sig}\end{array}\right\} \longrightarrow \text{not recommended}$$

Renormalized Unit Weight Error

- Recommded GoF indicator for Gaia DR2 astrometry
  - Not given directly in the Gaia archive
- Can be computed from the quantities:

$$
\begin{aligned}
\chi^2 &= \texttt{astrometric\_chi2\_al} \\
N &= \texttt{astrometric\_n\_good\_obs\_al} \\
G &= \texttt{phot\_g\_mean\_mag} \\
C &= \texttt{bp\_rp} \text{ (if available)}
\end{aligned}
$$

- Unit weight error $\mathrm{UWE} = \sqrt{\chi^2/(N-5)}$
- Renormalized unit weight error $\mathrm{RUWE} = \mathrm{UWE}/u_0(G, C)$
- $u_0(G, C)$ is an empirical normalization factor, provided as a lookup table on the ESA Gaia DR2 Known Issues pages
  - A separate function $u_0(G)$ is provided for sources without a known colour
- Python code to calculate RUWE: `https://github.com/agabrown/gaiadr2-ruwe-tools`

Normalization factor $u_0(G, C)$

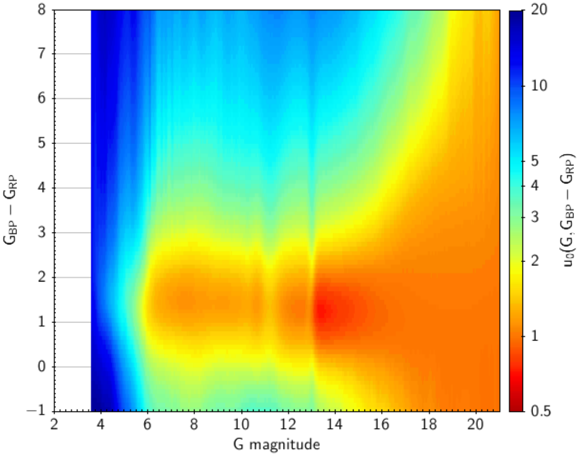
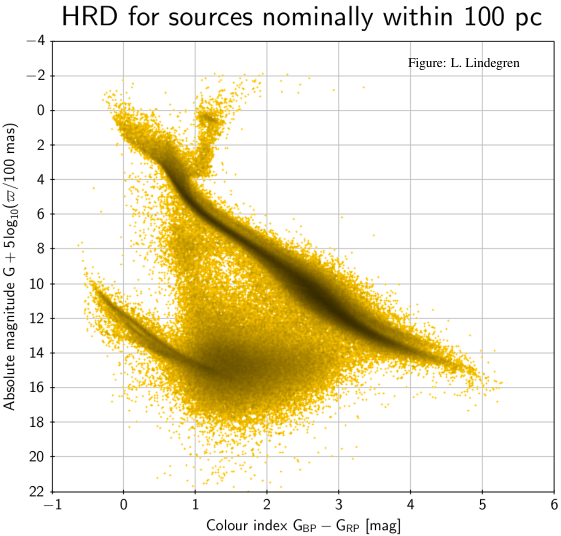


Figure: L. Lindegren

This is essentially the ‘typical’ UWE for a given magnitude and colour

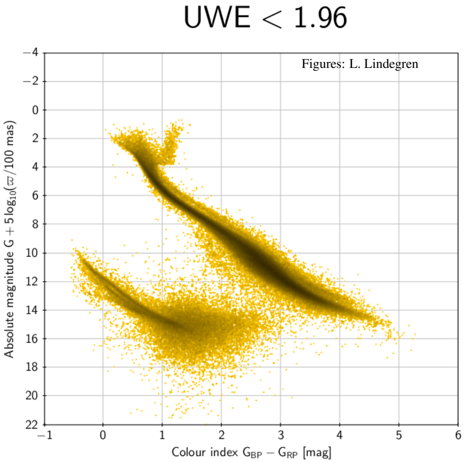
Illustration of the use of RUWE

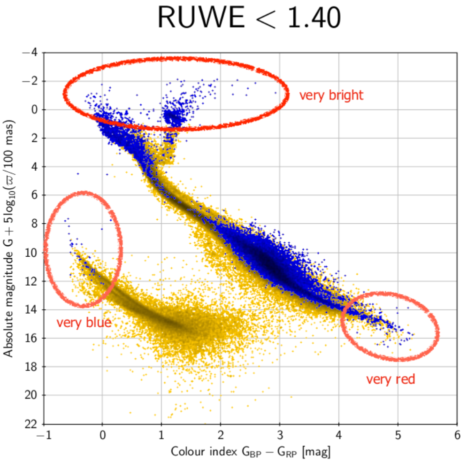


Selection:

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

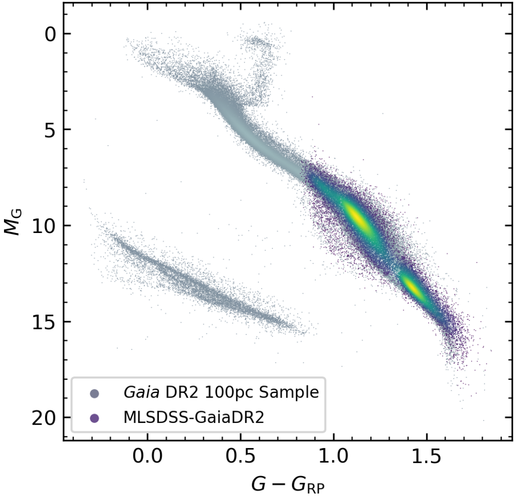
Illustration of the use of RUWE



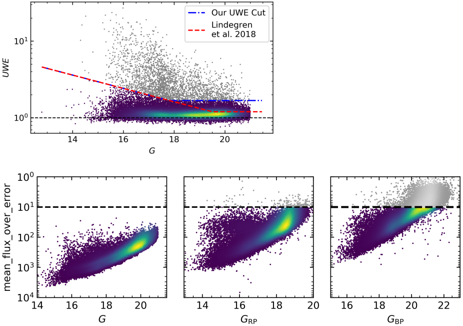


- Filtering by RUWE gives cleaner HRD
- Blue dots are sources missing in left diagram
- Experiment to decide on the limit in RUWE for your application!

Example of quality cuts customized for M/L dwarf sample

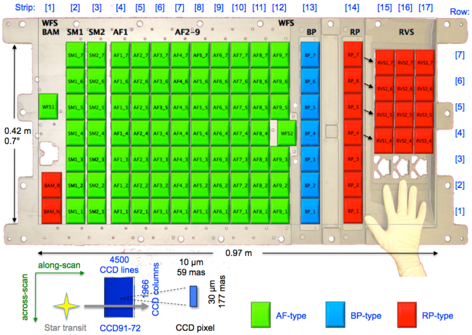


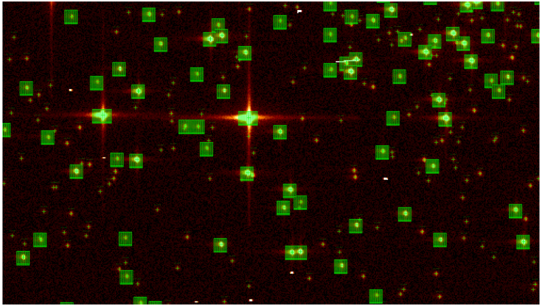
CMD with M/L dwarf sample after quality cuts described in Kiman et al. (2019, arXiv:1904.05911)



- Lindegren et al. (2018) cuts on UWE and photometric quality too conservative for M/L dwarfs
  - UWE cuts were not designed with specific source colours in mind
  - these stars are very red and thus expected to have low G_{BP} SNR

Ingredients of Gaia DR2 selection function



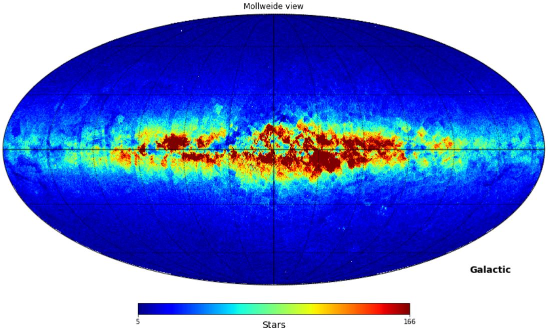


Credits: ESA/R. Kohley/J. de Bruijne

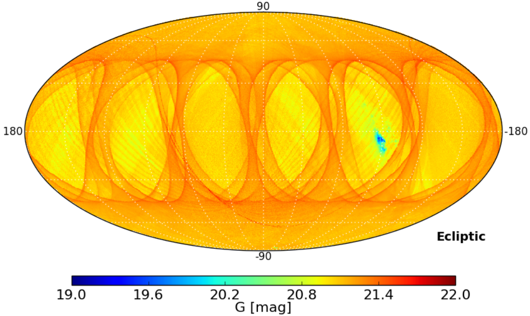
- Detection of sources in SM1/2, confirmation in AF1 (as cosmic ray/ spurious source rejection step)
- Strict flux threshold applied, but in presence of magnitude estimation errors
- Limit on source ‘size’
- Management of observation window conflicts and resource limitations affects completeness in crowded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

Ingredients of Gaia DR2 selection function

RVS Source counts (Katz et al. 2019)



99th percentile in G (Arenou et al. 2018)



- 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