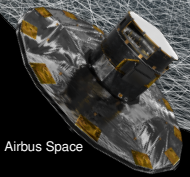
The GaiaUnlimited Project

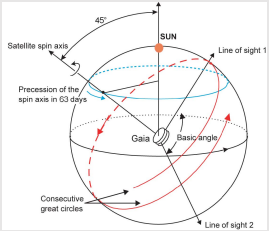
Anthony Brown & Alfred Castro Ginard
Leiden University, Universitat de Barcelona
brown@strw.leidenuniv.nl

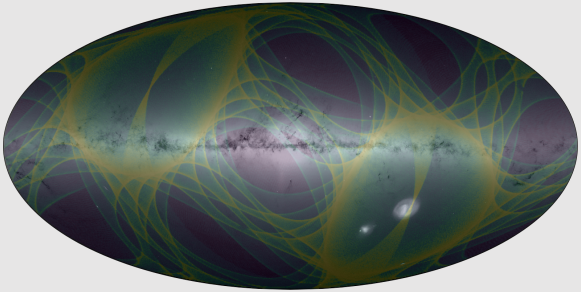


Airbus Space

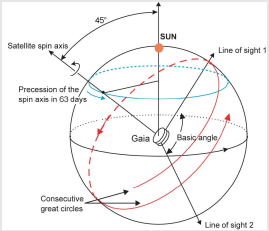
ESA/Gaia/DPAC

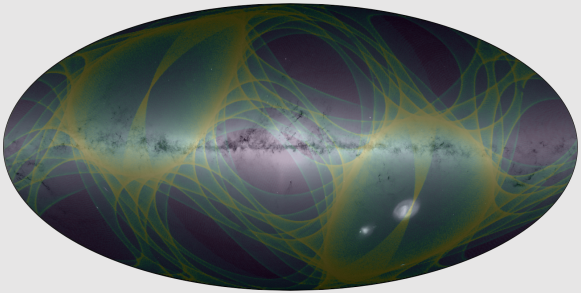
Gaia selection function elements

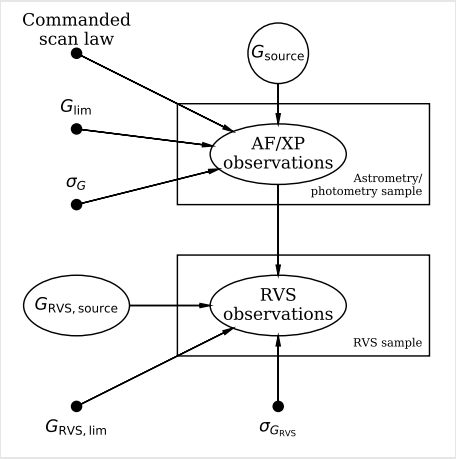




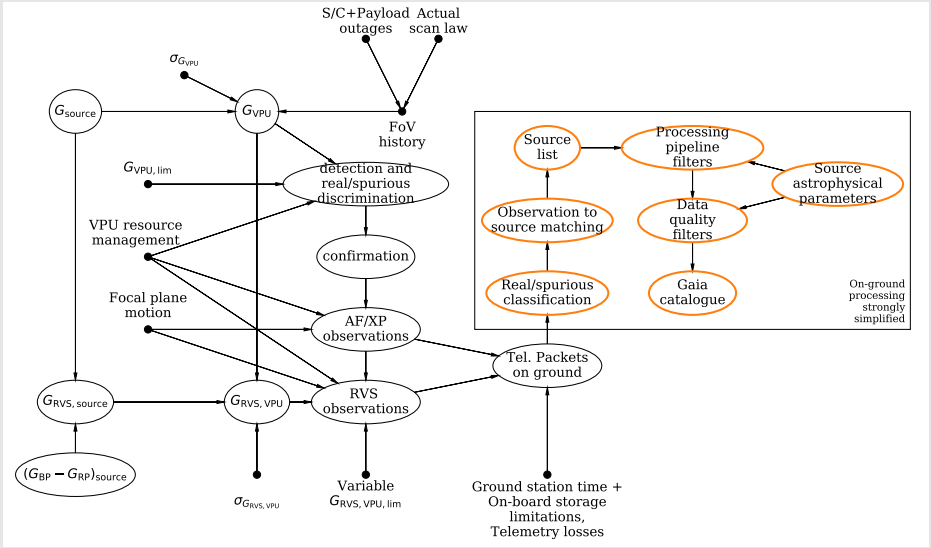
Gaia selection function elements



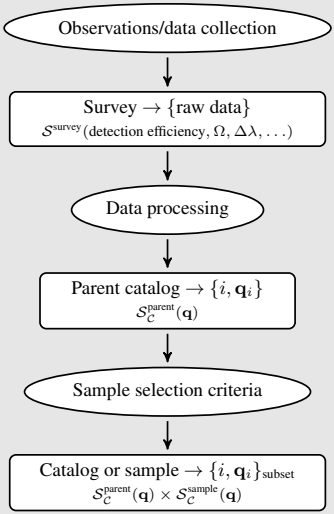




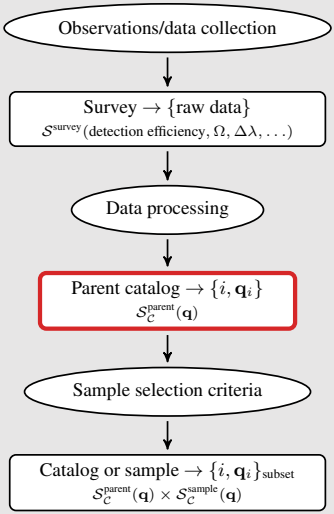
Gaia selection function elements



GaiaUnlimited approach to selection function



GaiaUnlimited approach to selection function



$\mathcal{S}_{\mathcal{C}}(\boldsymbol{q}_i)$: probability that source i with attributes $\boldsymbol{q}_i$ is contained in catalogue or sample $\mathcal{C}$

GaiaUnlimited provides:

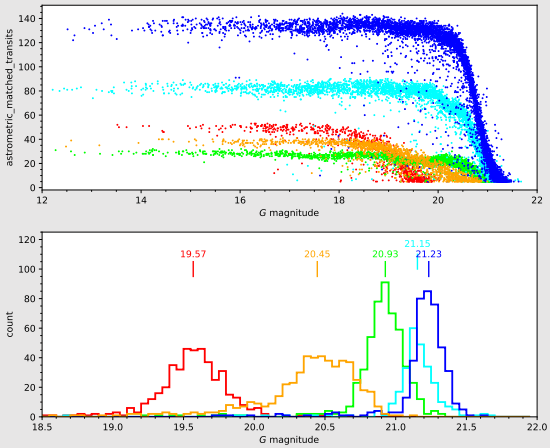
- Selection function of parent Gaia catalogue
  - all sources with $\boldsymbol{q}_i = \{\alpha_i, \delta_i, G_i, \ldots\}$
- Selection function for subsets
  - 5-parameter astrometry, full photometry, radial velocity, RUWE $< x$, ...
- Selection functions for Gaia combined with other surveys
  - Photometric and spectroscopic surveys (e.g. AllWISE, APOGEE)
- Examples of specific Gaia selection functions
  - For example: binaries, Red Clump stars, metal poor Red Giants
- Basic principles in Rix et al. (2022), arXiv:2106.07653

Availability of tools and data

- Website: https:/gaia-unlimited.org
- Github repository: https://github.com/gaia-unlimited
  - Code and notebooks with worked examples
- Documentation: https://gaiaunlimited.readthedocs.io
  - Including tutorials and links to worked examples

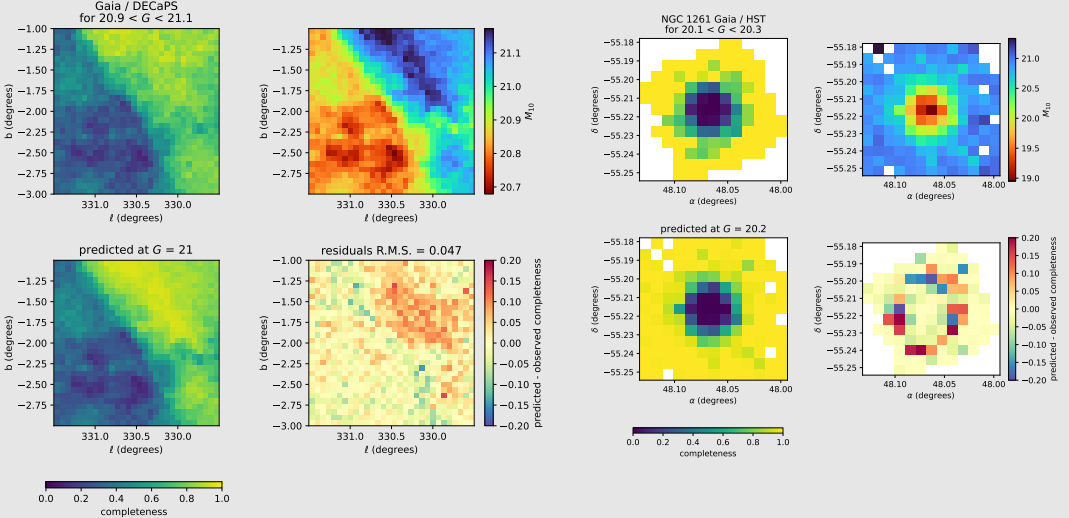
This project was funded by the European Union's Horizon 2020 research and innovation program under grant agreement No 101004110. 2021.0–2024.5

Empirical model of parent survey selection function

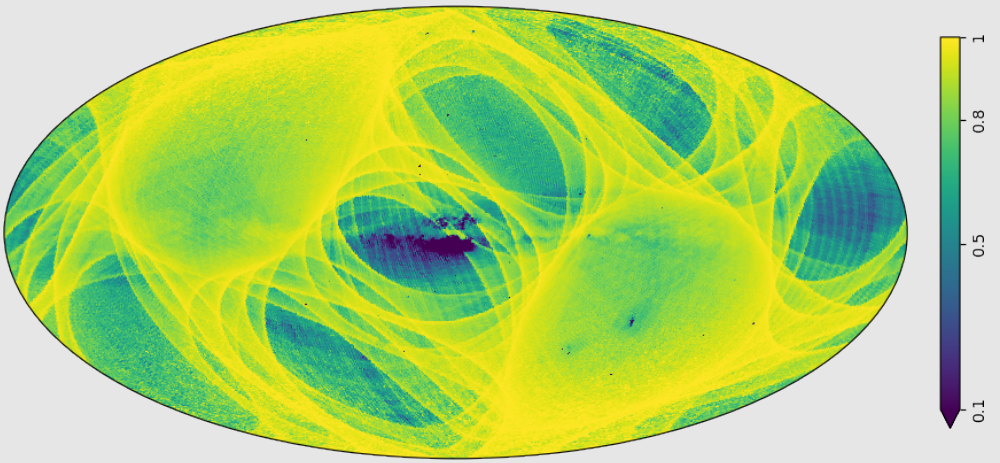


- M10: median magnitude of sources with less than 10 `astrometric_matched_transits`
- Correlates well with survey completeness level
- Basis for emprical model of selection function $\mathcal{S}(\ell, b, G)$
- Calibrated on DECaPS survey of Galactic plane
- Cantat-Gaudin et al. (2023), arXiv:2208.09335

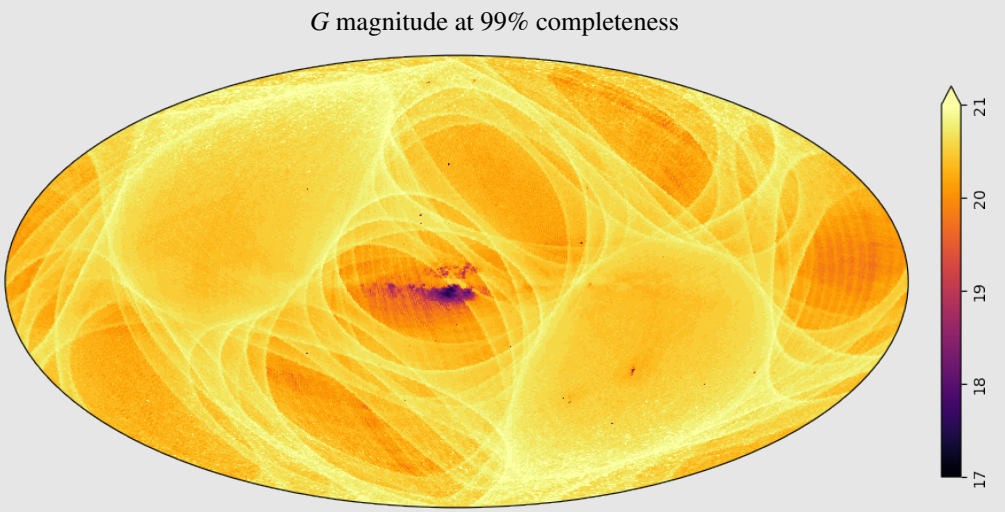
Empirical model of parent survey selection function



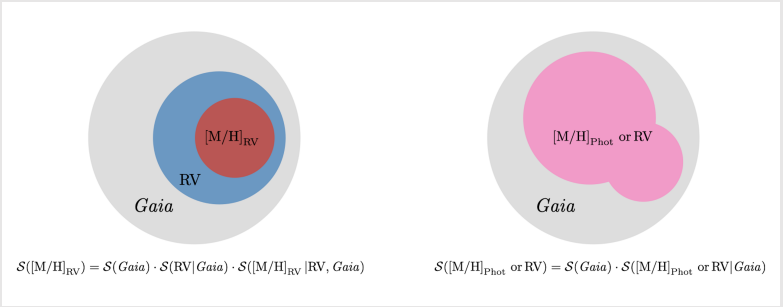
Empirical model of parent survey selection function



Empirical model of parent survey selection function



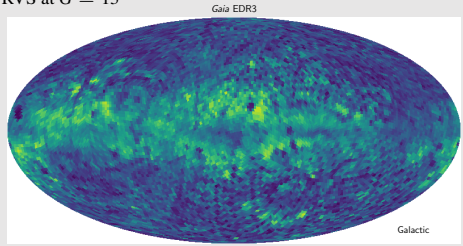
Selection functions for subsets of the Gaia catalogue



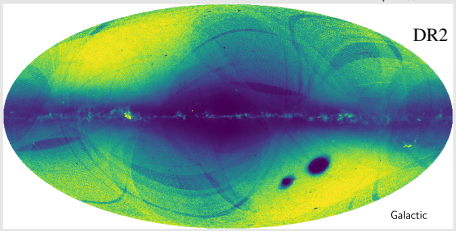
- $\mathcal{S}_{\mathcal{C}}^{\text{subset}}(\boldsymbol{q}) = \mathcal{S}_{\mathcal{C}}(\boldsymbol{q} \mid \boldsymbol{q} \text{ in parent}) \times \mathcal{S}_{\mathcal{C}}^{\text{parent}}(\boldsymbol{q})$
- Model $\mathcal{S}_{\mathcal{C}}(\boldsymbol{q} \mid \boldsymbol{q} \text{ in parent})$ as binomial process with probability p
- Estimate p as $(k+1)/(n+2)$: k number of sources with $\boldsymbol{q}$ in given bin (sky, magnitude, colour, etc) and n total number of sources in parent catalogue
- Details in Castro-Ginard et al. (2023), arXiv:2303.17738

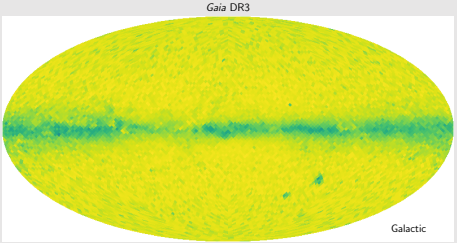
Selection functions for subsets of the Gaia catalogue

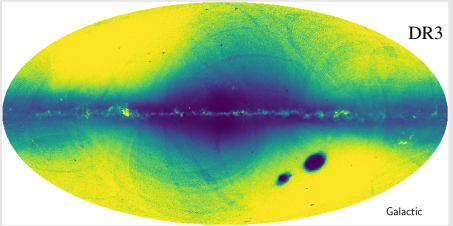
RVS at $G = 13$



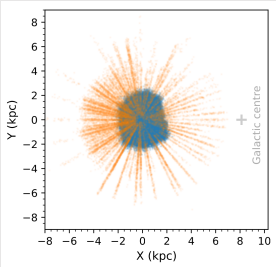
$\varpi > 0.1$ and $\varpi/\sigma_\varpi > 5$

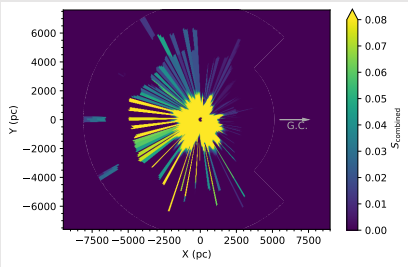


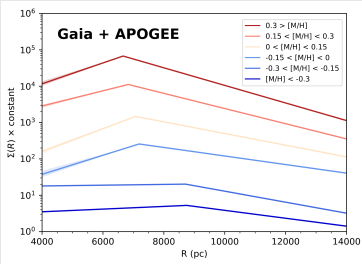


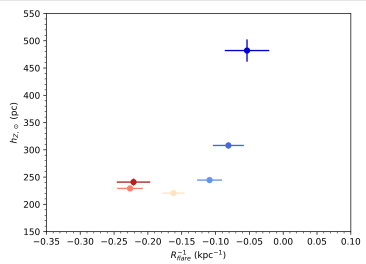


Milky Way disk with Gaia-GSP-Spec/APOGEE combination





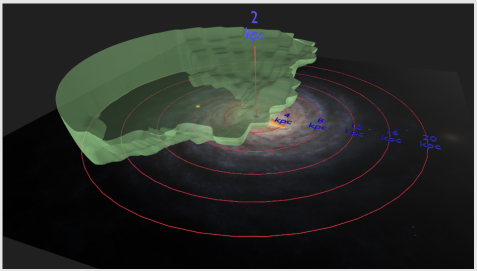


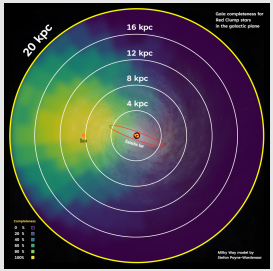


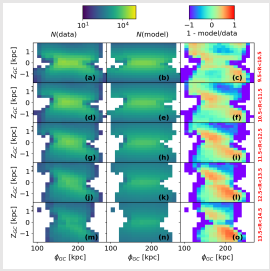
- Selection functions constructed for Red Clump stars: APOGEE DR17, GSP-Spec, and combination
- Joint sample used to study density profile of mono-abundance populations in different metallicity ranges
- Density profiles steepen with metallicity
- Metallicity-dependent flaring of $[\alpha/\mathrm{Fe}] < 0.1$ disk

Cantat-Gaudin et al,
arXiv:2401.05023

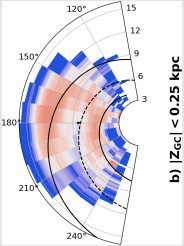
The disk in Red Clump stars from Gaia+AllWISE





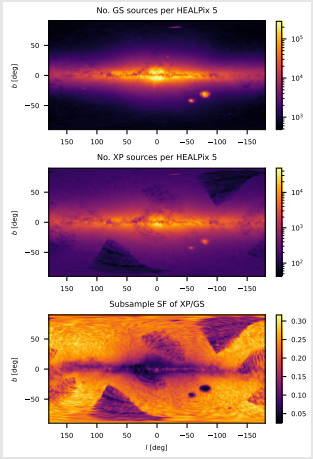


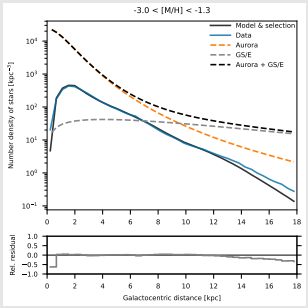
- Construction of combined Gaia+AllWISE selection function for Red Clump stars
- Disk structure for this sample: flared thin disk + shorter/thicker disk with no flare
- Outer disk shows clear signs of the warp
- Model + selection function residuals near disk plane show hints of spiral arms

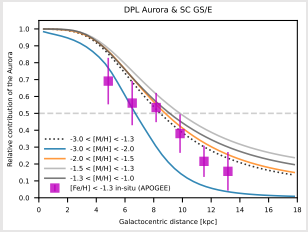


Khanna et al,
arXiv:2410.22036

Aurora and Gaia-Enceladus from XP-based RGB sample



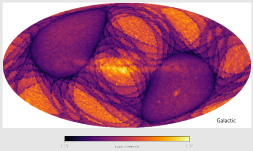




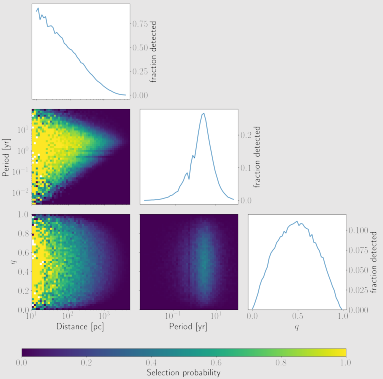
Kurbatov et al,
arXiv:2410.22250

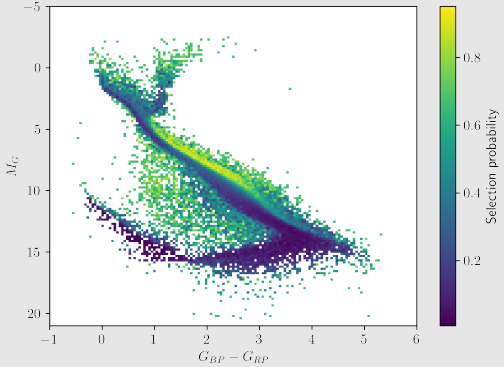
- Metal poor RGB stars, selection function: $G < 16$, $\varpi > 4\sigma_{\varpi}$, metallicities from XP spectra
- Density + luminosity function + extinction model for Aurora and Gaia-Enceladus populations
- Model predicts observables, accounting for selection function and parallax uncertainties
- Fairly complex model available in Python/Jupyter from GaiaUnlimited repository

Binaries selected on the basis of RUWE



- Sky-dependent RUWE threshold for selecting candidate binary stars
- Translation to selection effects on underlying binary orbital parameters
- Tools available to simulate selection effects for your favourite binary population





Castro-Ginard et al,
arXiv:2404.14127

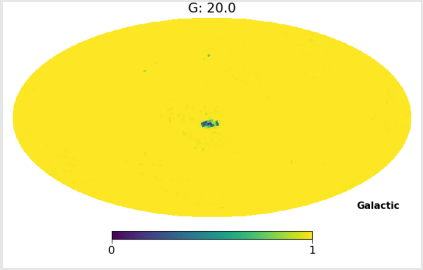
Statistical model of parent selection function for Gaia DR4

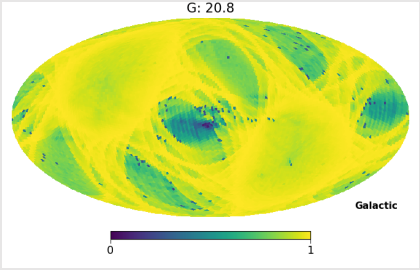
- Detection of sources by Gaia is considered a coin-flipping experiment with effective detection probability θ which varies with (ℓ, b, G) (Boubert & Everall 2020)
- Probability of k detections out of n field of view transits across a source:

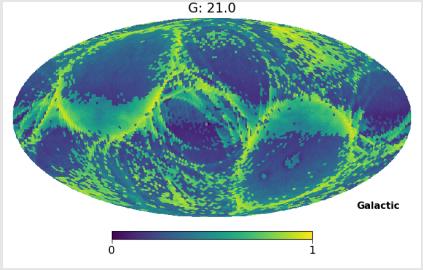
$$p(k \mid n, \theta) = \binom{n}{k} \theta^k (1-\theta)^{n-k}$$

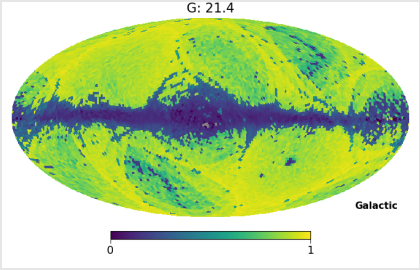
- k is the number of times a source is detected (in either Gaia field of view)
  - `astrometric_matched_transits`
  - Implicitly accounts for observation gaps, discarded data, and crowded fields
- n is the number of times that source crossed one of Gaia's fields of view
  - as predicted by Gaia Observation Scheduling Tool (GOST)
- Infer $\theta(\ell, b, G)$ from values of k and n per HEALPix and magnitude bin (and take into account $k \geq 5$)
- Following preliminary results **for DR3** are by Alfred Castro Ginard

Statistical model: preliminary results for Gaia DR3



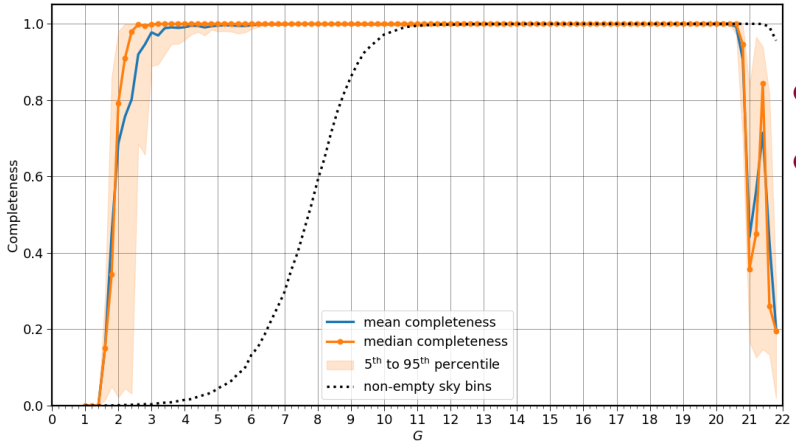






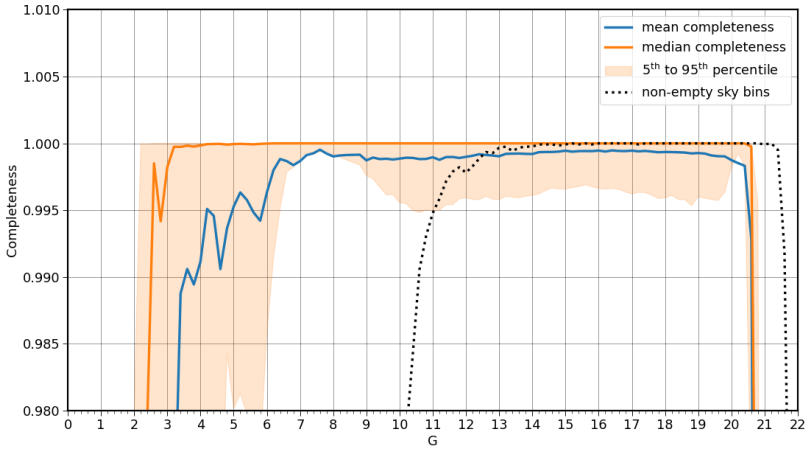
?

Statistical model: preliminary results for Gaia DR3



- Unexpected dip/ peak at $G > 21$
- Due to galaxies or QSO/AGN with too faint flux estimates
  - Mismatch PSF vs. galaxy image
  - Missing flux outside observation window

Statistical model: preliminary results for Gaia DR3



To do for Gaia DR4(+)

- Filter out extended galaxies and QSO/AGN or limit selection function to $G \leq 21$
- Improve statistical model to account for differences between GOST and the actual Gaia scanning law
  - Tests are underway with n as latent variable with prior informed by GOST
  - Speed-up of inference process
- Determine selection function for Gaia DR4

SPACIOUS: Computational opportunities and challenges

- Extensive use of the SPACIOUS/Spark platform
  - Automated task scheduling for parallelization
  - Efficient access to data (k for 2 billion sources, n for $\sim 2 \times 10^5$ sky pixels, ~ 100 magnitude bins)
- Numpyro + JAX for speed up of MCMC does not work on Spark
  - Use Dask instead?
- Time consuming step is a specific integral in likelihood calculation
  - Currently trying out evaluation of integral by neural network trained on pre-computed samples

GaiaUnlimited team

GaiaUnlimited postdocs

- Tristan Cantat-Gaudin (MPIA Heidelberg)
- Alfred Castro-Ginard (Leiden)
- Shourya Khanna (INAF-Torino)
- Evgeny Kurbatov (Cambridge)

Past members/GaiaVerse

- Douglas Boubert (Oxford)
- Andy Everall (Cambridge)
- Semyeong Oh (IoA Cambridge)
- Zuzanna Kostrzewa-Rutkowska (Leiden)

- Anthony Brown (Leiden, PI)
- Vasily Belokurov (IoA Cambridge)
- Andy Casey (Monash)
- Ron Drimmel (INAF-Torino)
- Morgan Fouesneau (MPIA Heidelberg)
- David Hogg (NYU/ Flatiron/ MPIA Heidelberg)
- Adrian Price-Whelan (Flatiron)
- Eloisa Poggio (Nice)
- Elena Sellentin (Leiden)
- Richard Smart (Torino)
- Alessandro Spagna (Torino)
- Hans-Walter Rix (MPIA Heidelberg)