Enabling Rubin Science with Robust Cross-Matches in the Crowded LSST Sky

Tom J Wilson (he/him) and Tim Naylor
with the LSST:UK DAC team
t.j.wilson@exeter.ac.uk
University of Exeter

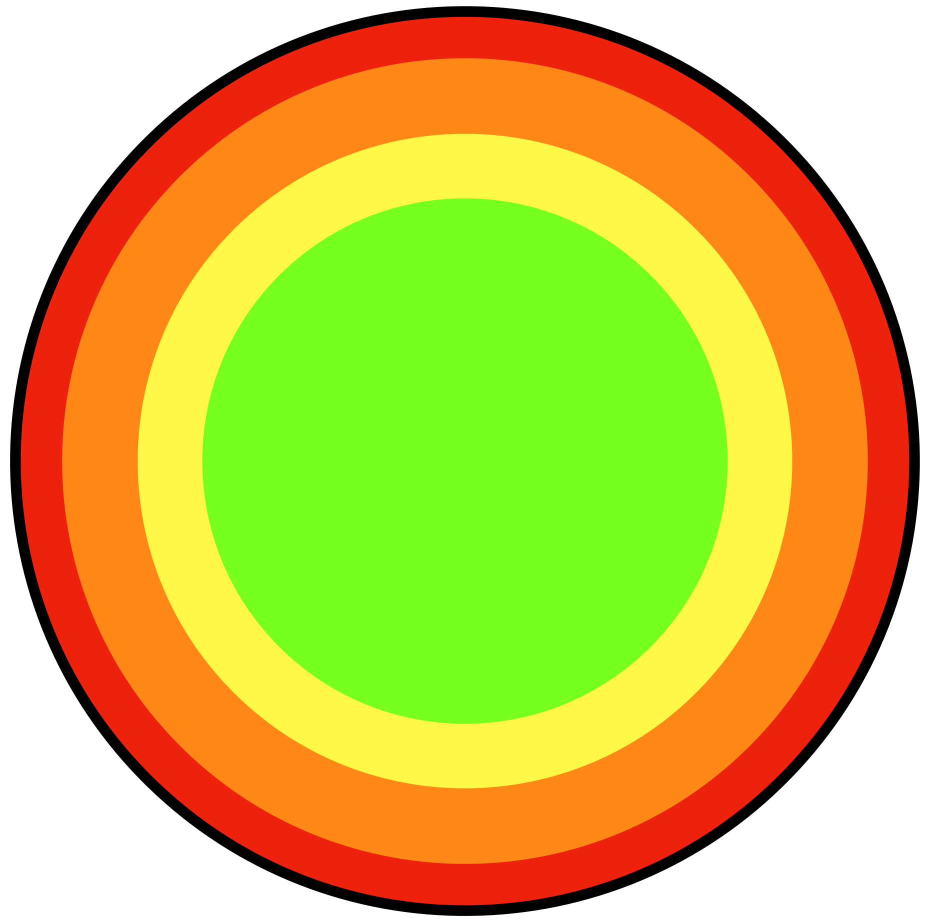
National Astronomy Meeting 2024 — Preparing for UK involvement in early science with the Rubin LSST, 15/Jul/24

Science and Technology Facilities Council

@Onoddil@pm.me
@Onoddil.github.io

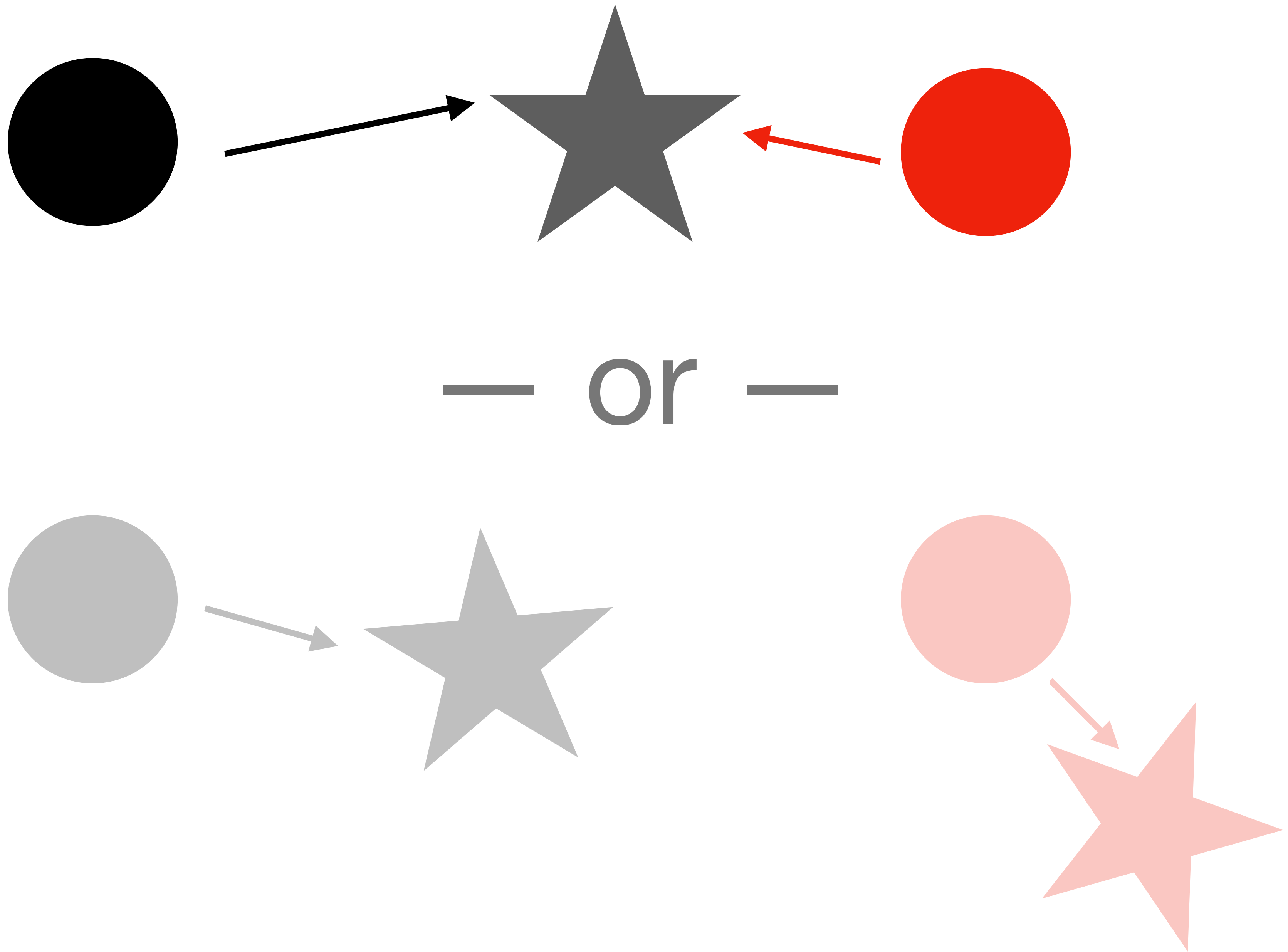
LSST:UK Consortium

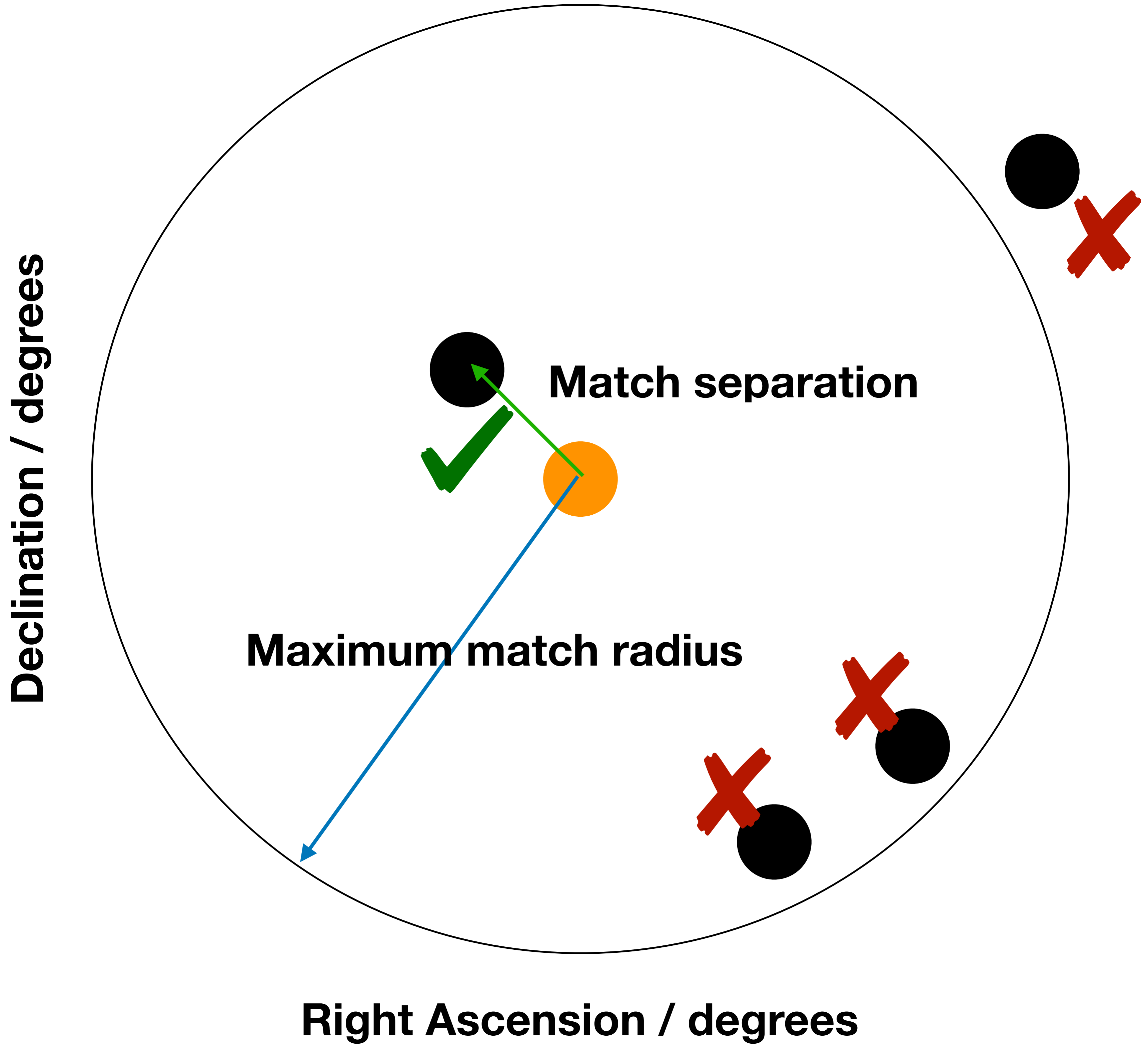
Cross-Match Science, Methodology, Background

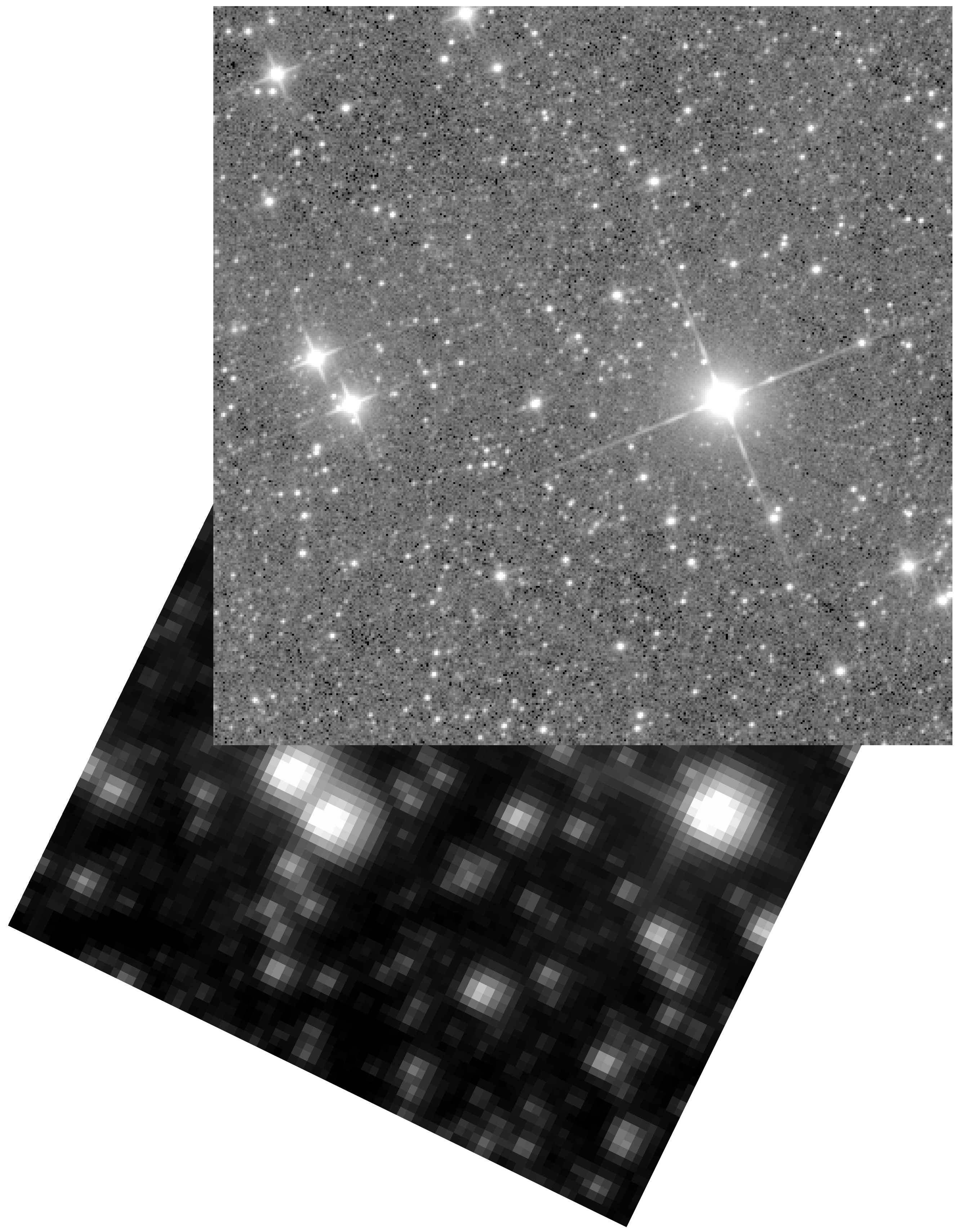


Tom J Wilson @onoddil

“Simple” Cross-Matching

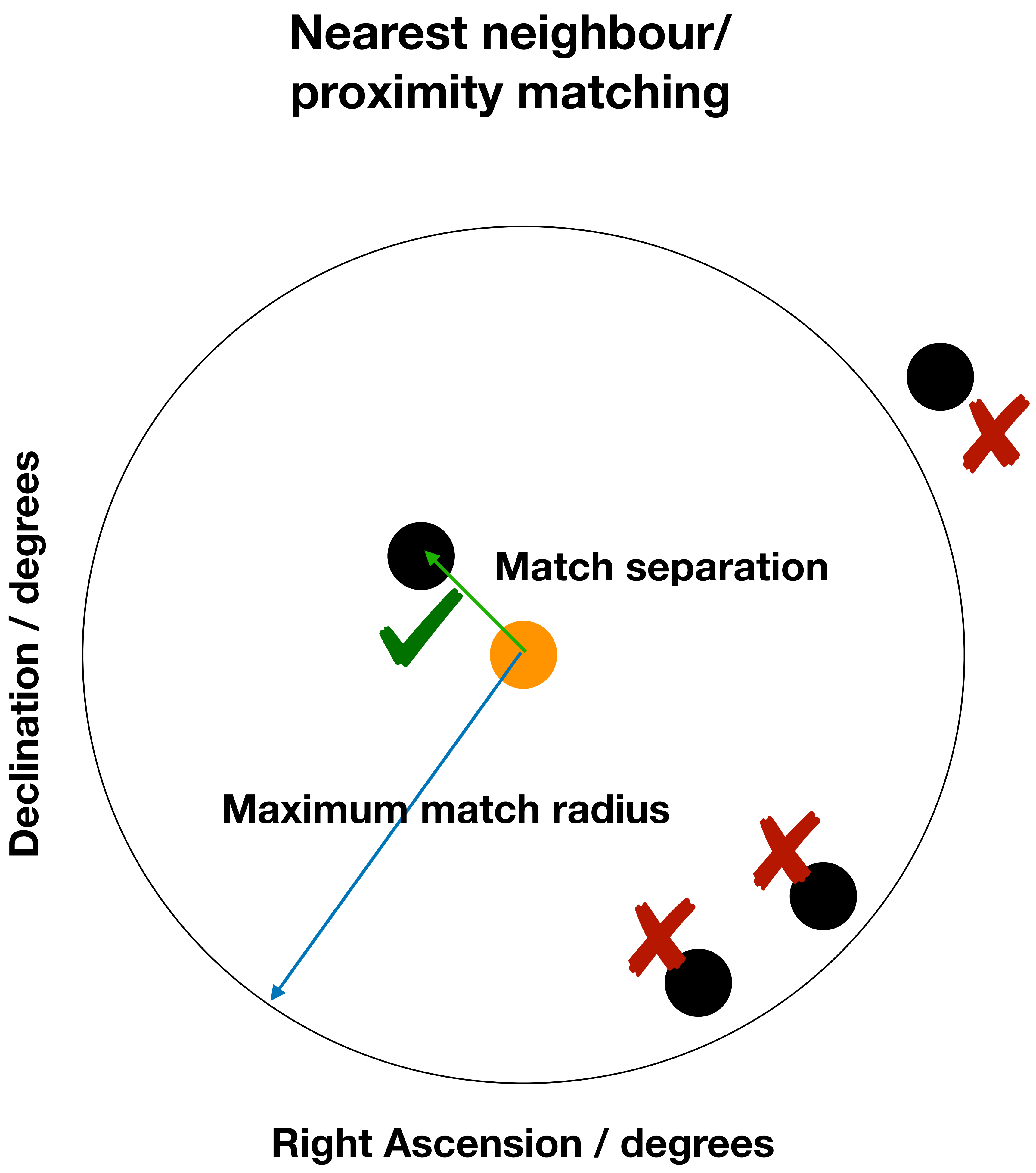


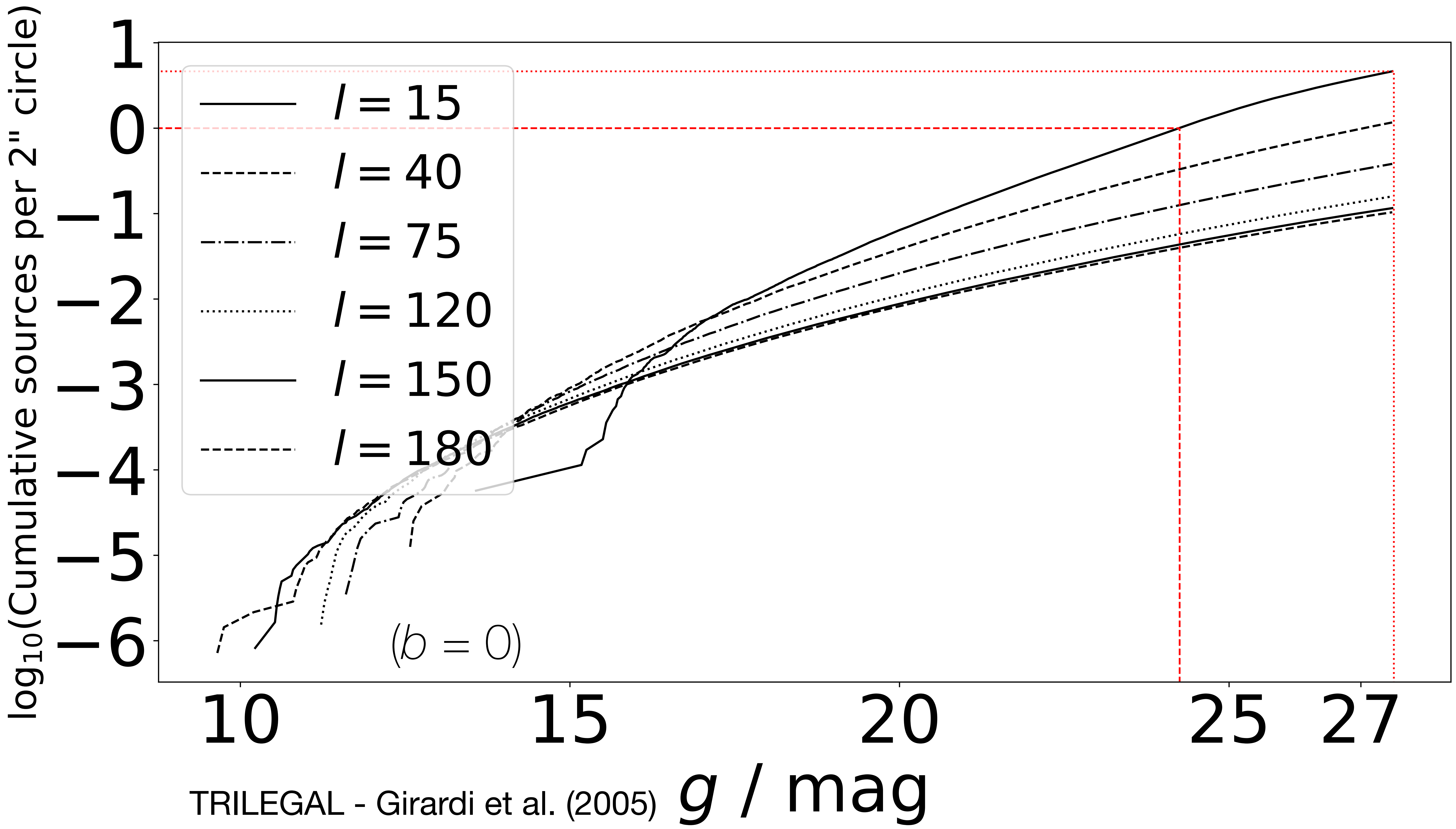




Tom J Wilson @onoddil

The Problem With LSST



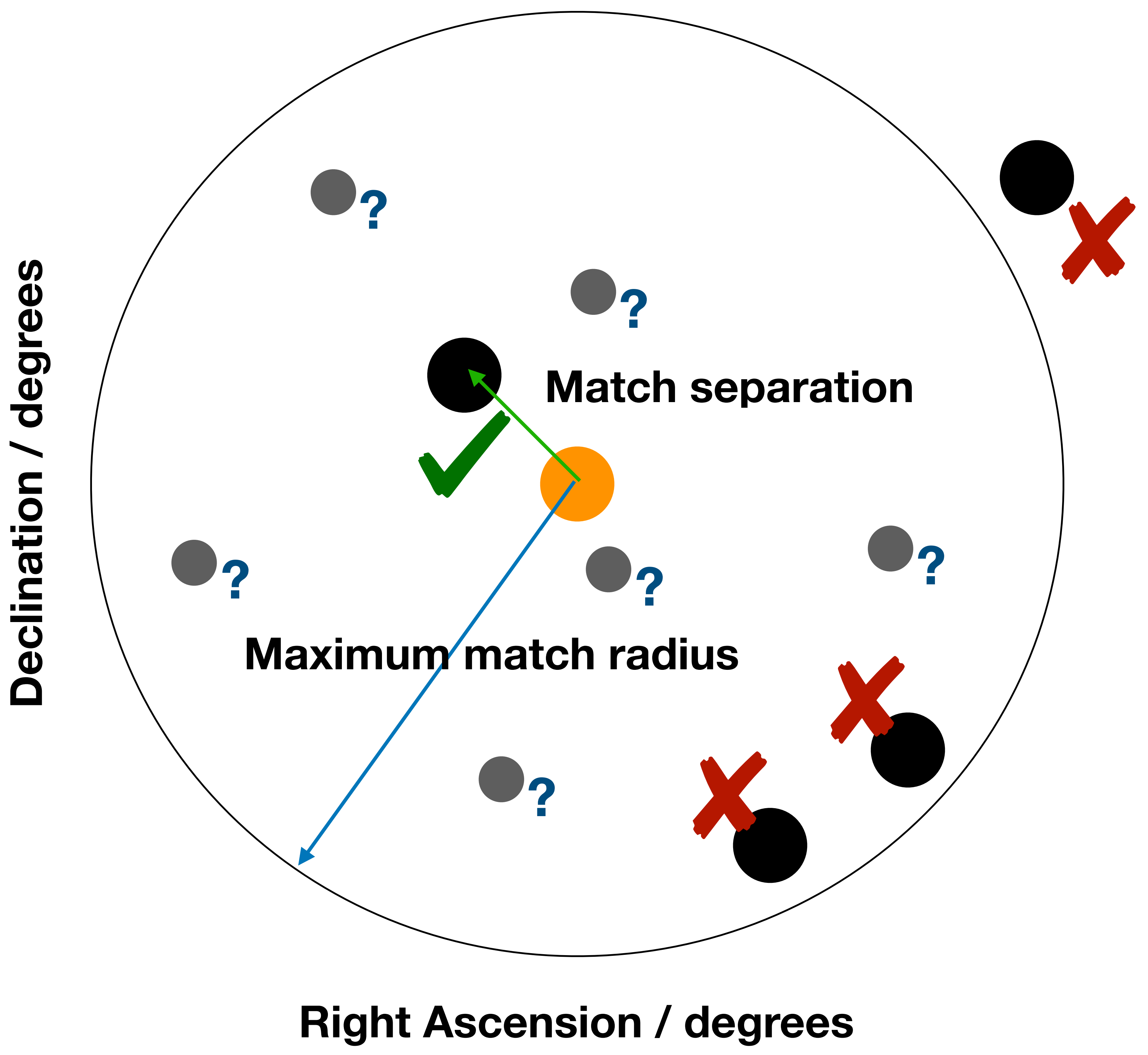


TRILEGAL - Girardi et al. (2005)

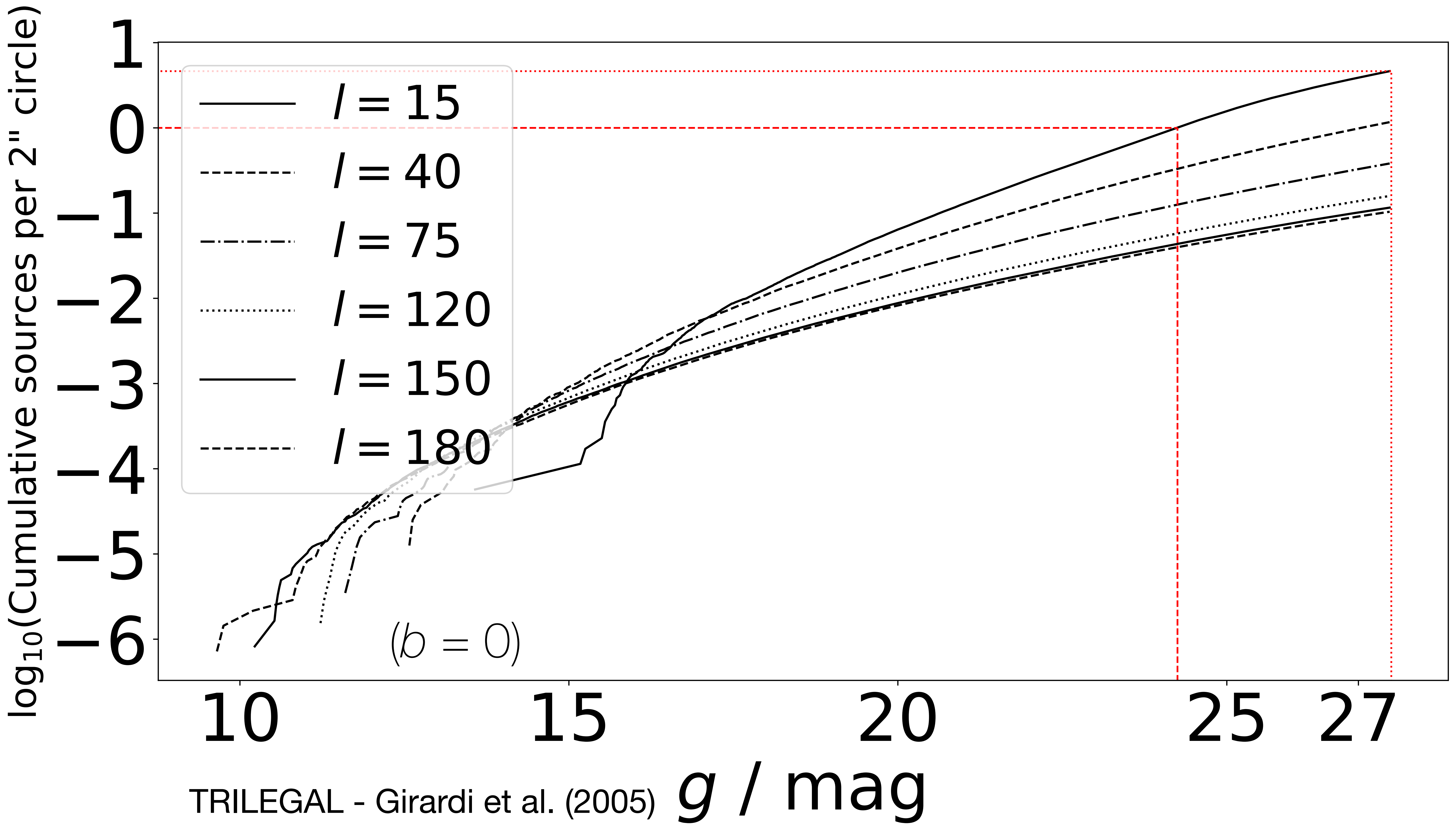
Tom J Wilson @onoddil

The Problem With LSST

Nearest neighbour/ proximity matching



(It's still a few randomly placed objects in every match radius at high Galactic latitudes)

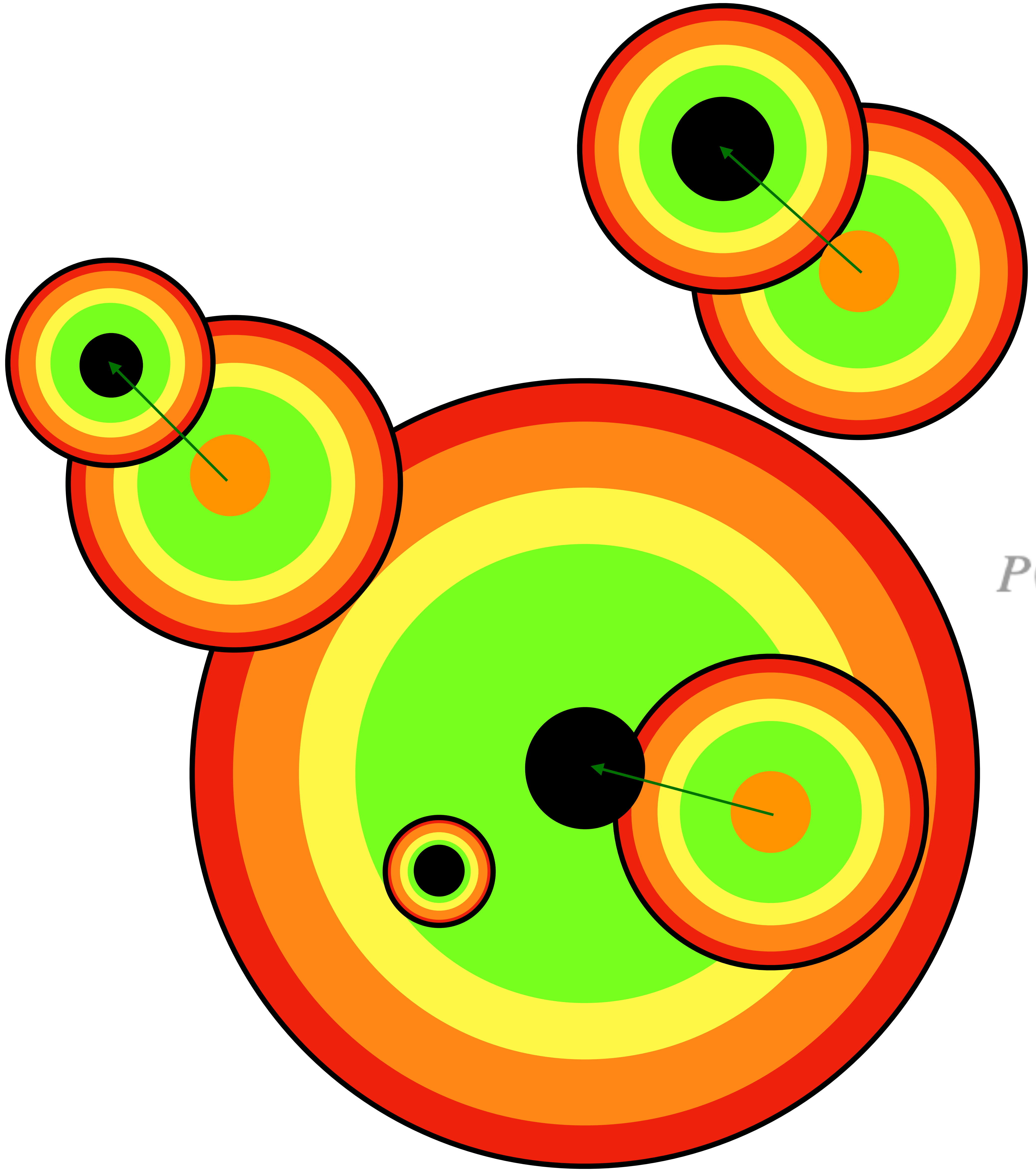


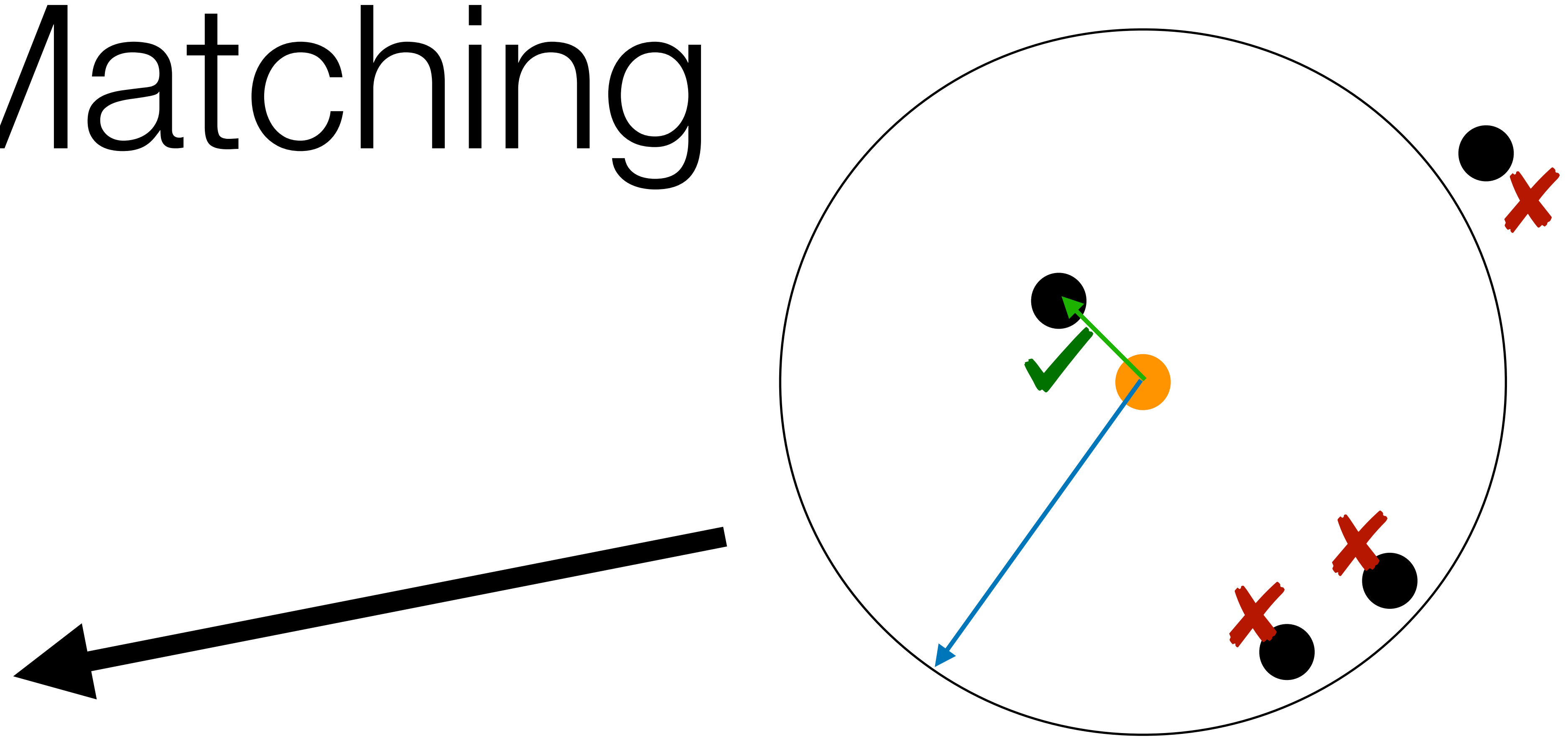
TRILEGAL - Girardi et al. (2005)

Nearest-neighbour matching *will not* work in the era of Rubin!

Tom J Wilson @onoddil

Probabilistic Cross-Matching





Probability of two sources having their <u>on-sky separation</u> given the hypothesis they are counterparts

$$P(\zeta, \lambda, k|\gamma, \phi) = \frac{1}{K} \times \prod_{\delta \notin \zeta \cap \delta \in \gamma} N_\gamma f_\gamma^\delta \prod_{\omega \notin \lambda \cap \omega \in \phi} N_\phi f_\phi^\omega \prod_{i=1}^{k} N_c G_{\gamma\phi}^{\zeta_i \lambda_i} c_{\gamma\phi}^{\zeta_i \lambda_i}$$

Probability of sources having their <u>brightnesses</u> given they are unrelated to one another (“field stars”)

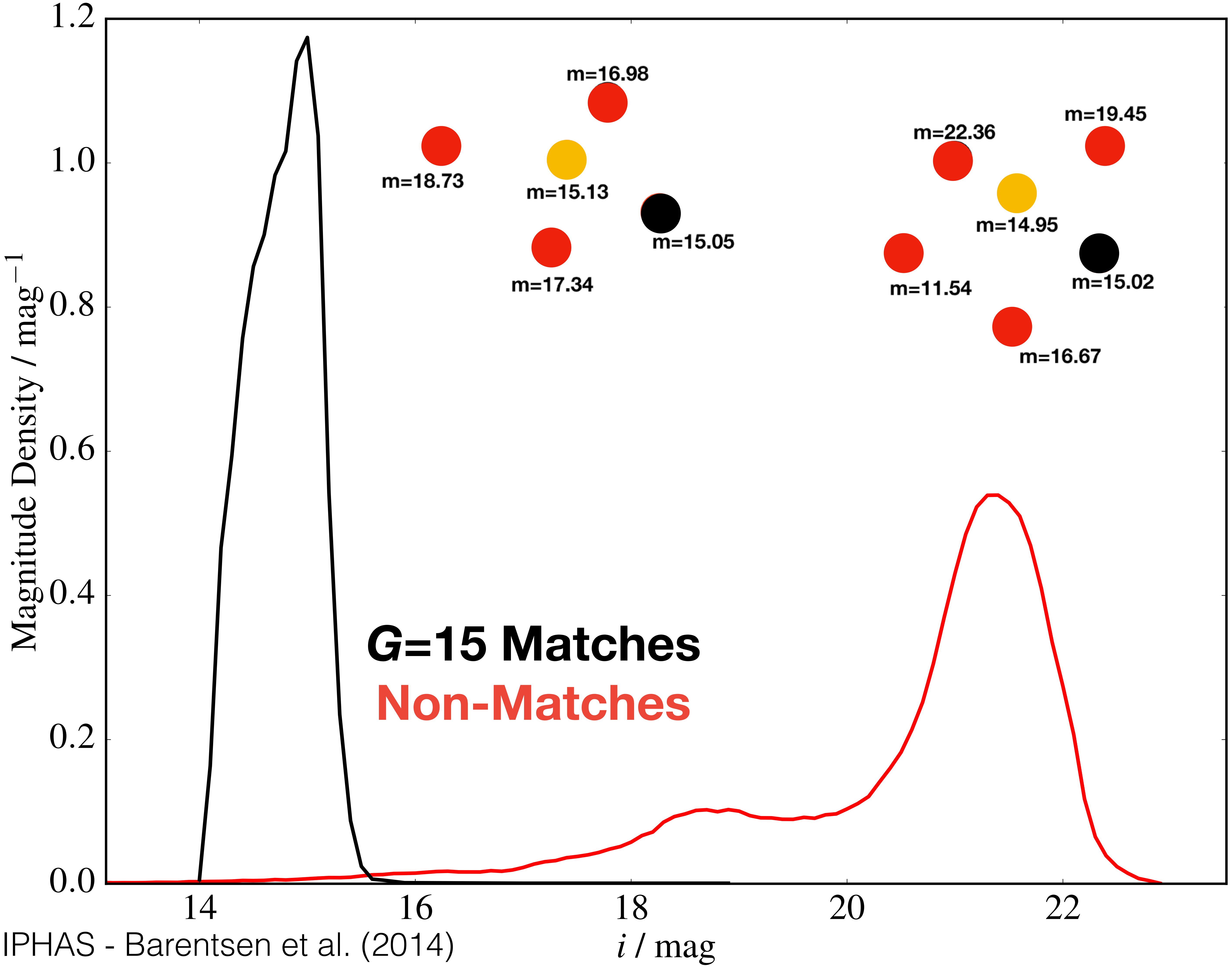
Probability of sources having their <u>brightnesses</u> given they are counterparts

Wilson & Naylor (2018a)

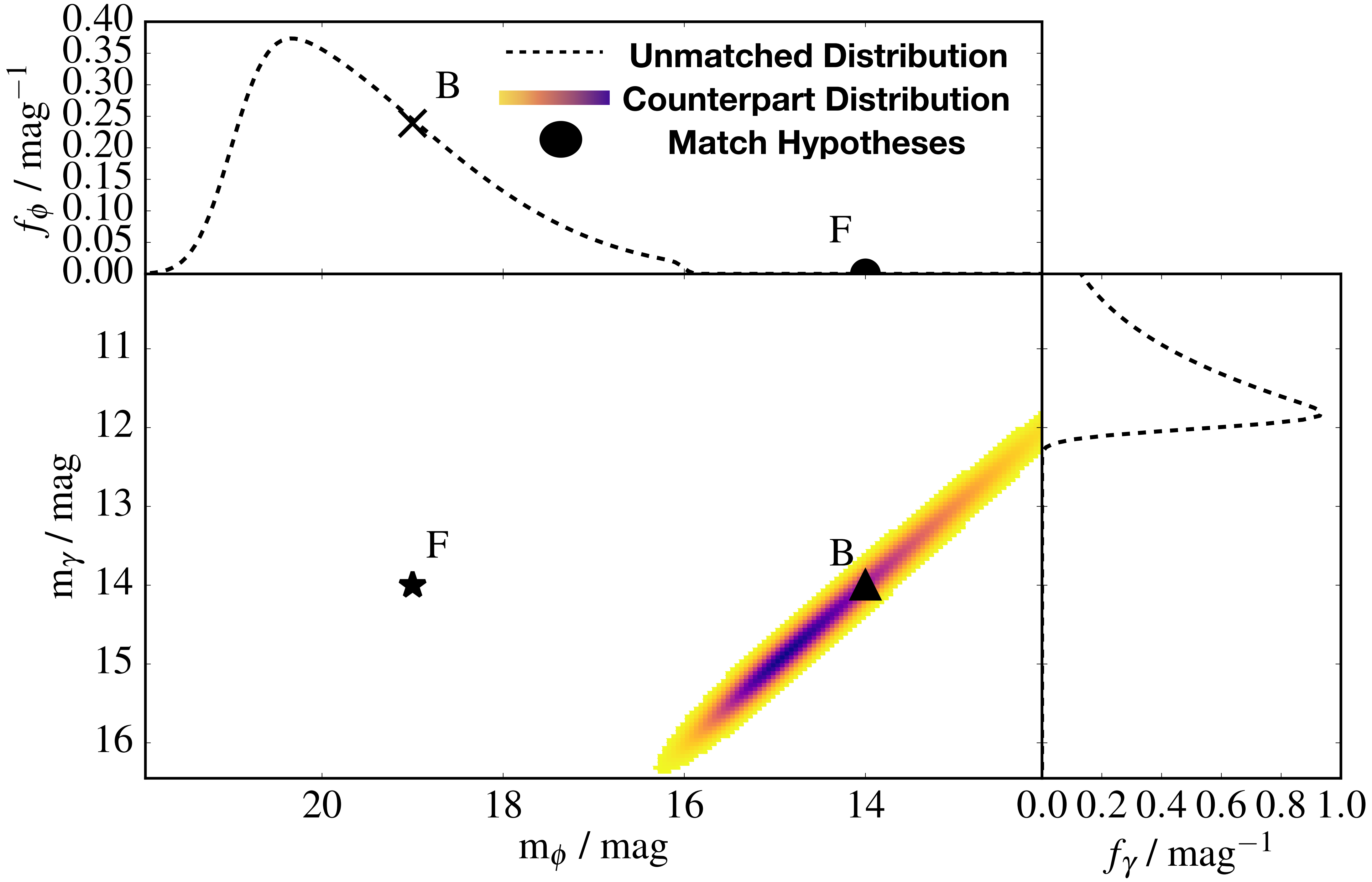
Tom J Wilson @onoddil

Photometry: Rejecting False Positives

$$P(\zeta, \lambda, k|\gamma, \phi) = \frac{1}{K} \times \prod_{\delta \notin \zeta \cap \delta \in \gamma} N_\gamma \boldsymbol{f_\gamma^\delta} \prod_{\omega \notin \lambda \cap \omega \in \phi} N_\phi \boldsymbol{f_\phi^\omega} \prod_{i=1}^{k} N_c G_{\gamma\phi}^{\zeta_i \lambda_i} \boldsymbol{c_{\gamma\phi}^{\zeta_i \lambda_i}}$$



IPHAS - Barentsen et al. (2014)
Gaia DR2 - Gaia Collaboration, Brown A. G. A., et al. (2018)

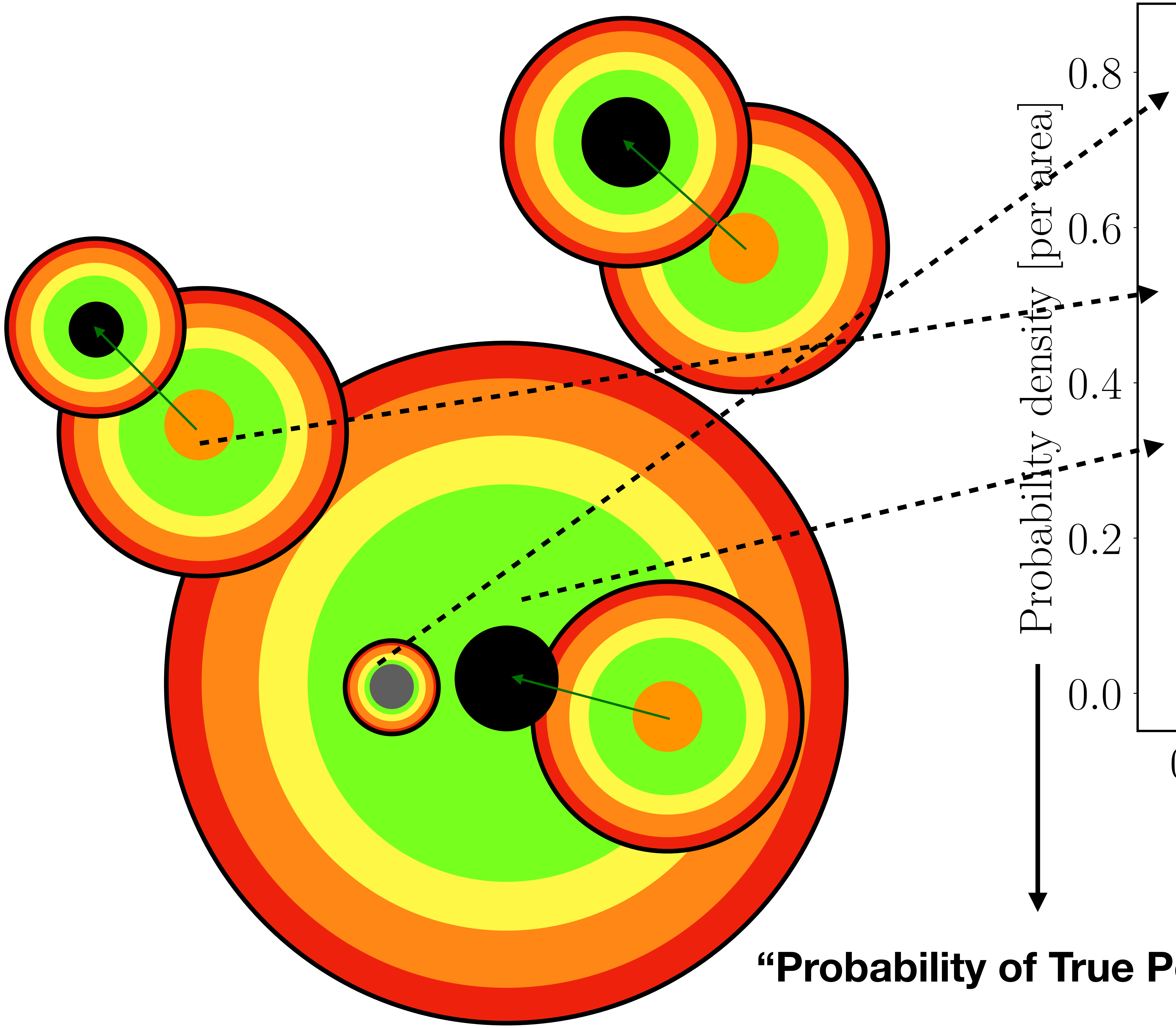


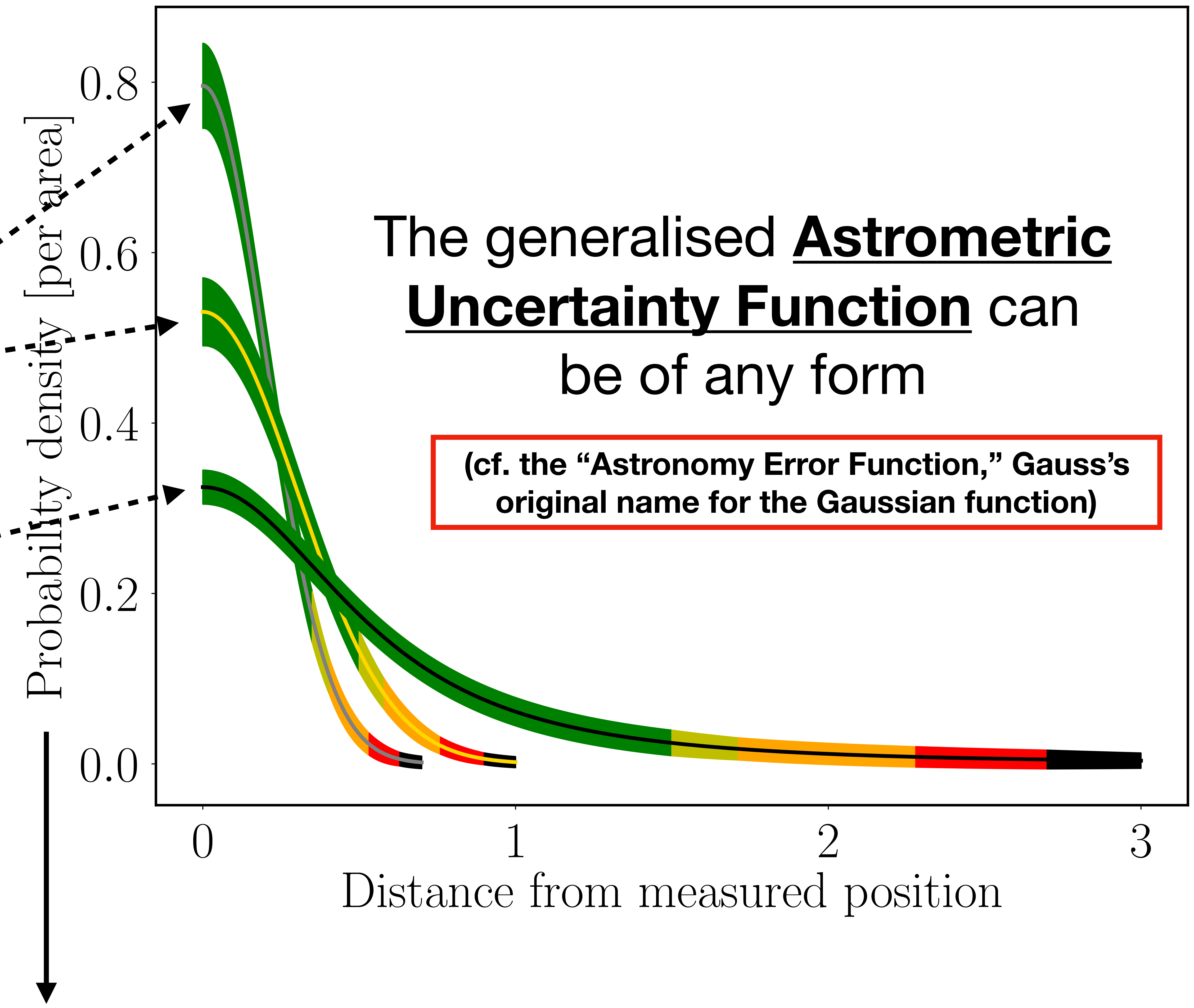
Wilson & Naylor (2018a)

The photometry-based likelihoods (c and f) allow us to mitigate high false positive rate in crowded fields, but now we need the position-based likelihood G

Tom J Wilson @onoddil

Probabilistic Cross-Matching: the AUF





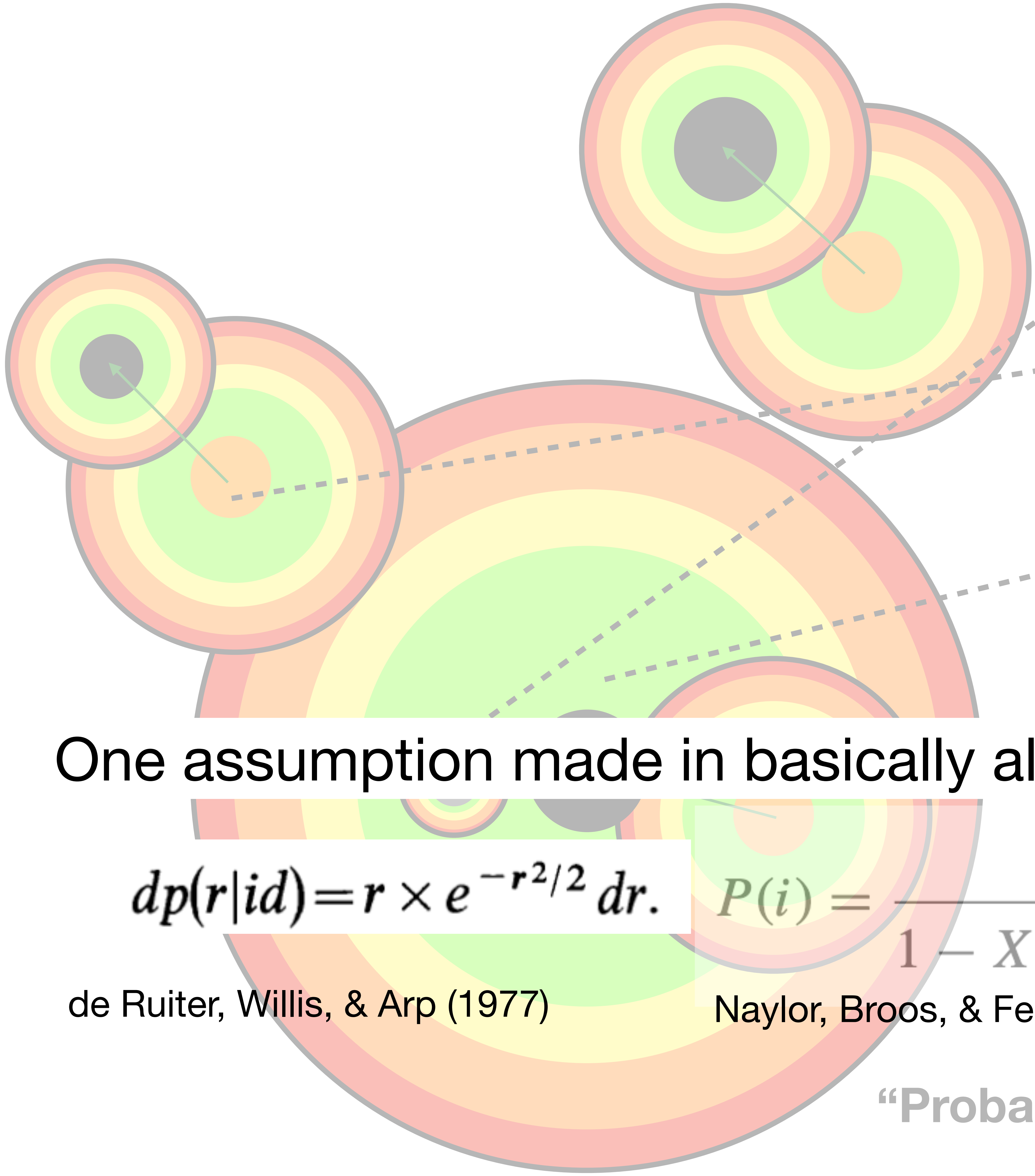
The generalised **Astrometric Uncertainty Function** can be of any form

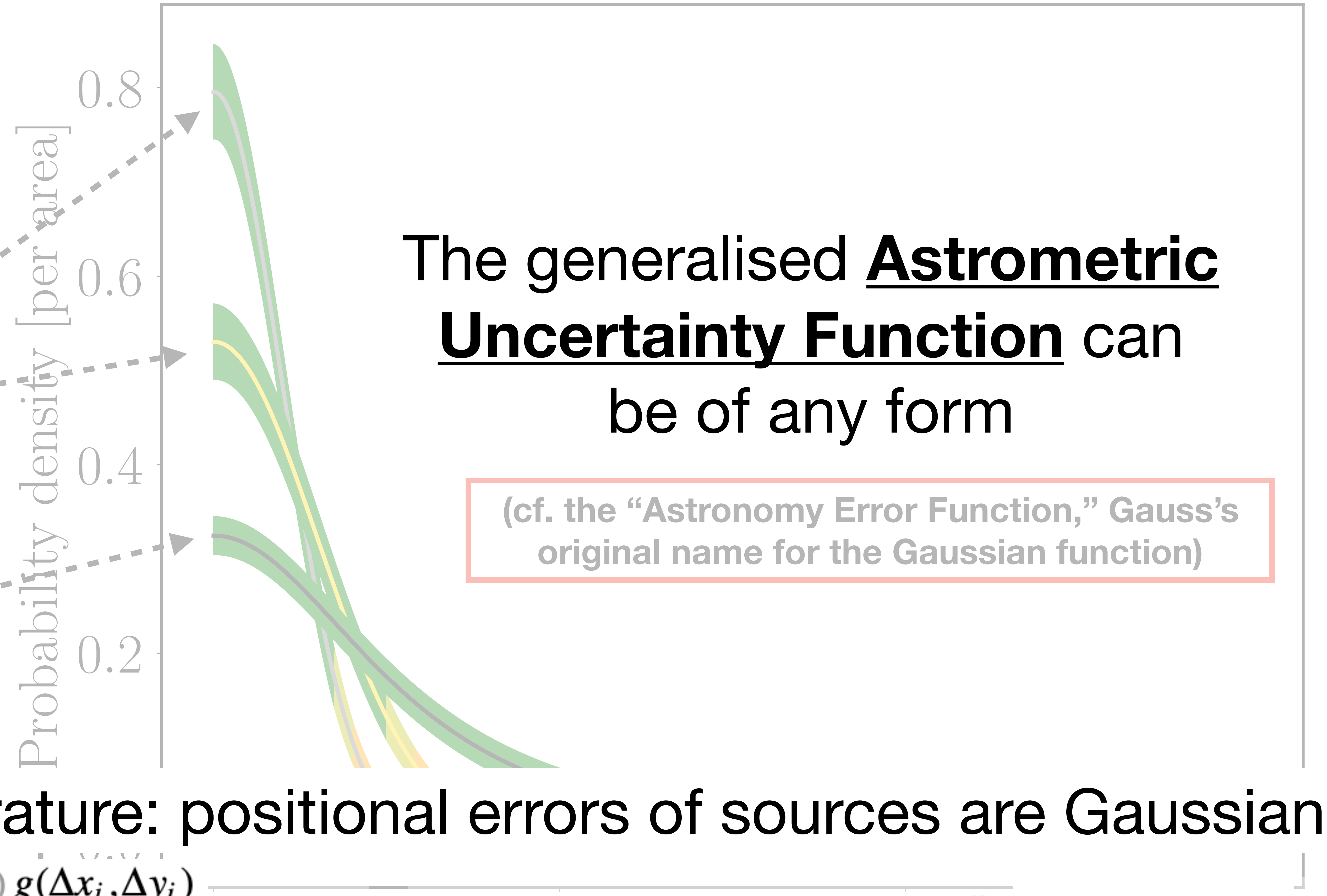
(cf. the "Astronomy Error Function," Gauss's original name for the Gaussian function)

"Probability of True Position being this far from the Measured Position"

Tom J Wilson @onoddil

Probabilistic Cross-Matching: the AUF





The generalised **Astrometric Uncertainty Function** can be of any form

(cf. the "Astronomy Error Function," Gauss's original name for the Gaussian function)

One assumption made in basically all literature: positional errors of sources are Gaussian!

$$dp(r|id) = r \times e^{-r^2/2}\, dr.$$

de Ruiter, Willis, & Arp (1977)

$$P(i) = \frac{\frac{Xc(m_i)\,g(\Delta x_i, \Delta y_i)}{Nf(m_i)}}{1 - X + \sum_j \frac{Xc(m_j)\,g(\Delta x_j, \Delta y_j)}{Nf(m_j)}}$$

Naylor, Broos, & Feigelson (2013)

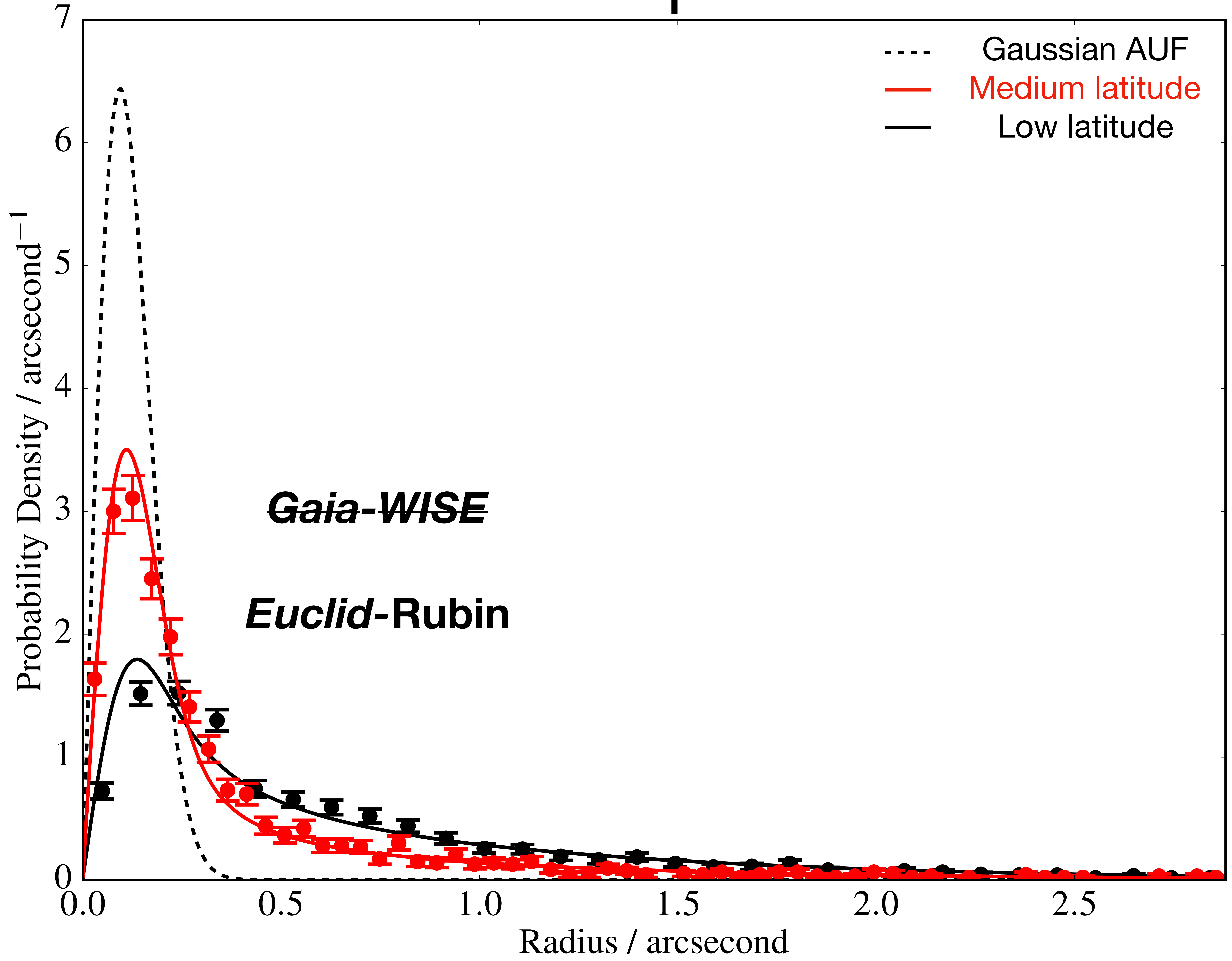
$$p(D|H) = \int p(m|H) \prod_{i=1}^{n} p_i(\mathbf{x}_i|\mathbf{m}, H)\, d^3m$$

Budavári & Szalay (2008)

"Probability of True Position being this far from the Measured Position"

Additional Components of the AUF

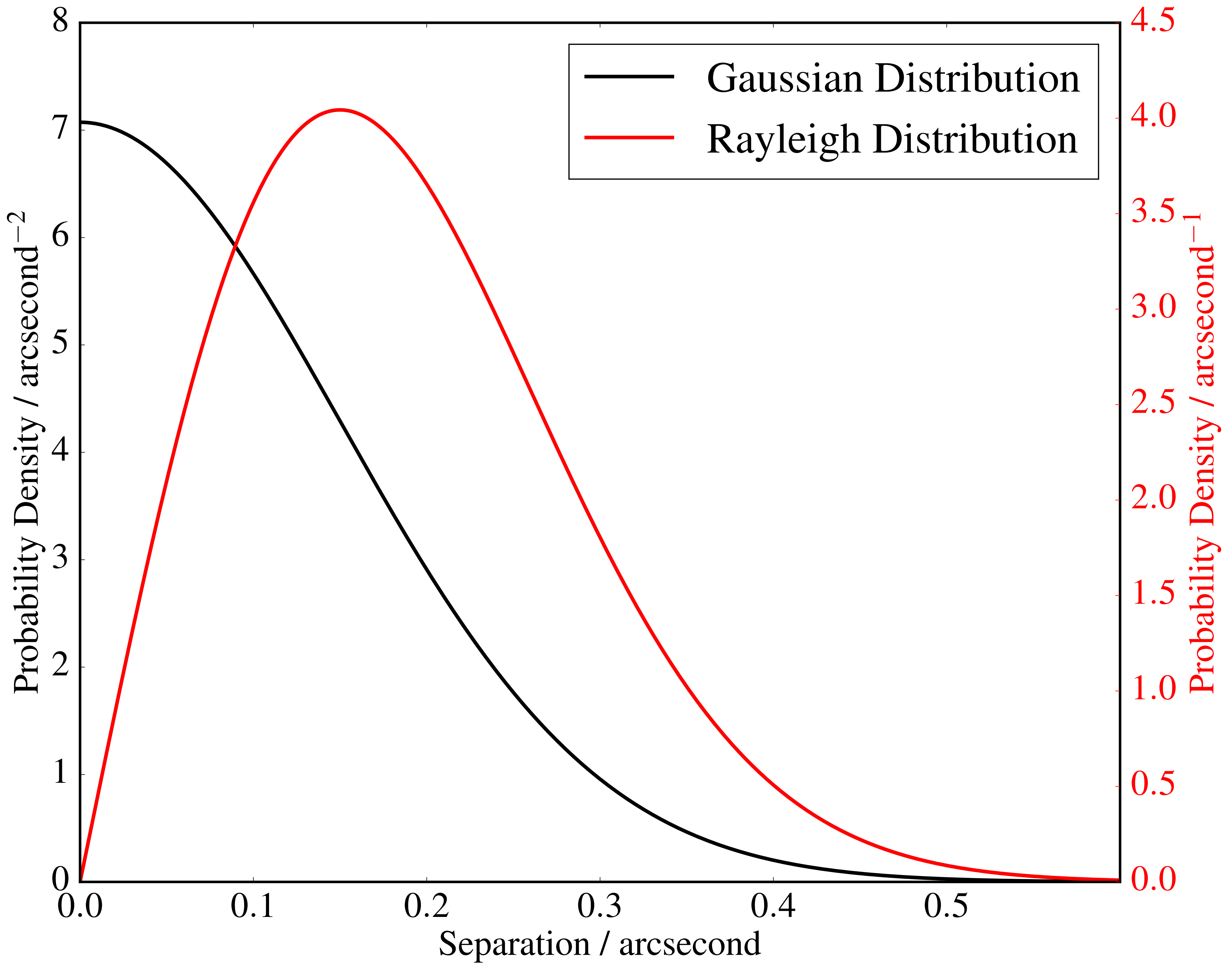
$$P(\zeta, \lambda, k|\gamma, \phi) = \frac{1}{K} \times \prod_{\delta \notin \zeta \cap \delta \in \gamma} N_\gamma f_\gamma^\delta \prod_{\omega \notin \lambda \cap \omega \in \phi} N_\phi f_\phi^\omega \prod_{i=1}^{k} N_c \boldsymbol{G}_{\boldsymbol{\gamma\phi}}^{\boldsymbol{\zeta_i \lambda_i}} c_{\gamma\phi}^{\zeta_i \lambda_i}$$



$$g(x, y, \sigma) = (2\pi\sigma^2)^{-1} \exp\left(-\frac{1}{2}\frac{x^2 + y^2}{\sigma^2}\right)$$

↓

$$g(r, \sigma) = \frac{r}{\sigma^2} \exp\left(-\frac{1}{2}\frac{r^2}{\sigma^2}\right)$$

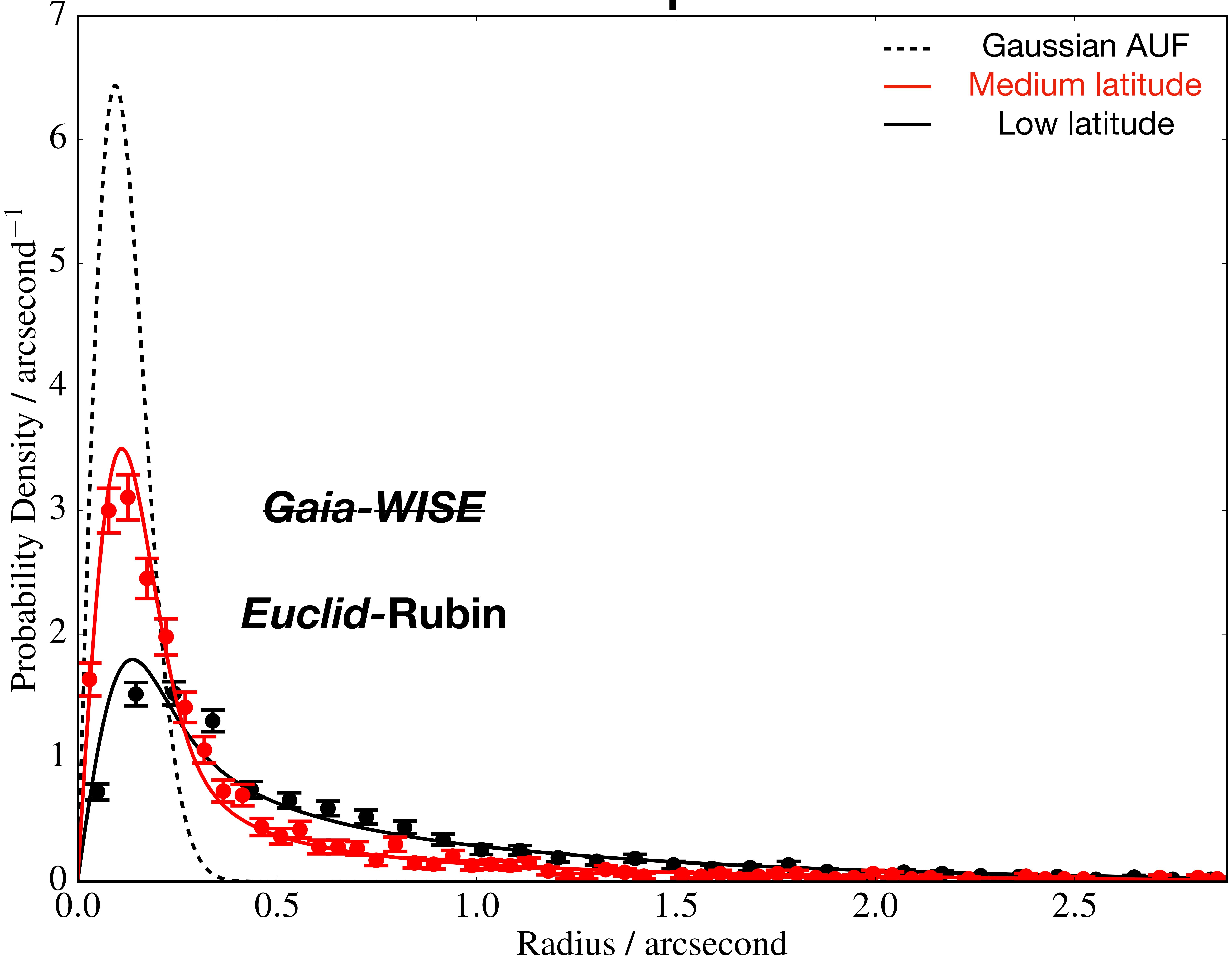


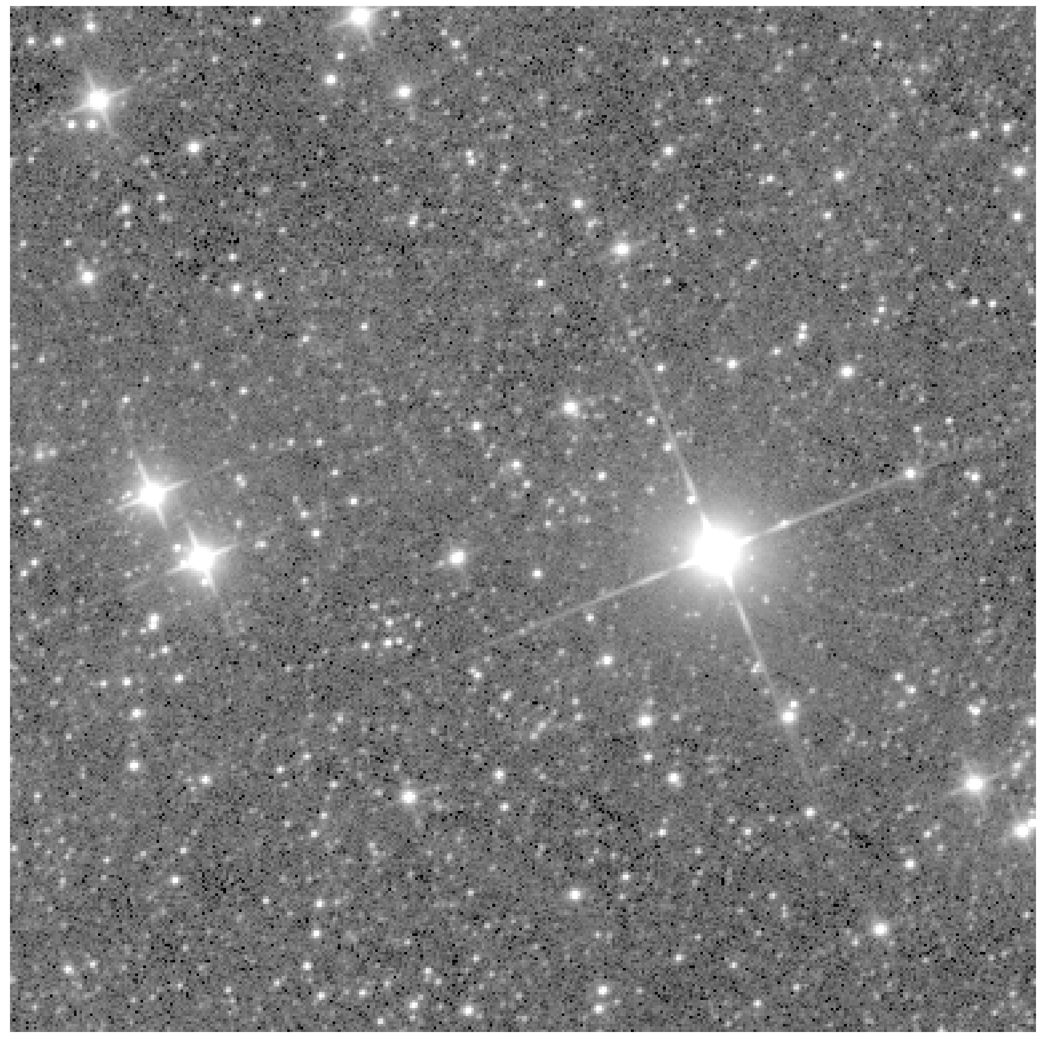
WISE - Wright et al. (2010)
Gaia DR2 - Gaia Collaboration, Brown A. G. A., et al. (2018)

Tom J Wilson @onoddil

Additional Components of the AUF

$$P(\zeta, \lambda, k | \gamma, \phi) = \frac{1}{K} \times \prod_{\delta \notin \zeta \cap \delta \in \gamma} N_\gamma f_\gamma^\delta \prod_{\omega \notin \lambda \cap \omega \in \phi} N_\phi f_\phi^\omega \prod_{i=1}^{k} N_\mathrm{c} \boldsymbol{G}_{\boldsymbol{\gamma\phi}}^{\boldsymbol{\zeta_i \lambda_i}} c_{\gamma\phi}^{\zeta_i \lambda_i}$$



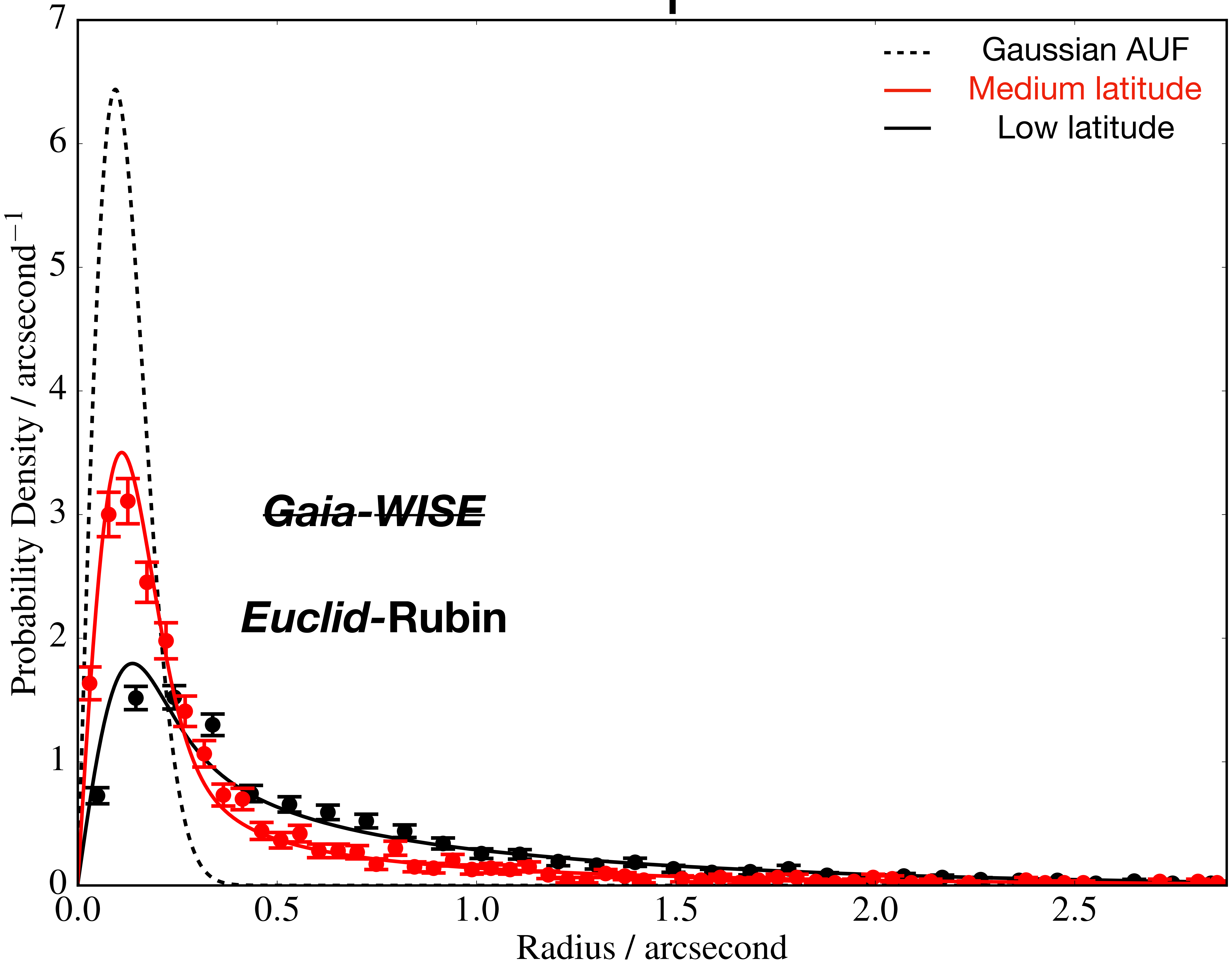


WISE - Wright et al. (2010)
Gaia DR2 - Gaia Collaboration, Brown A. G. A., et al. (2018)

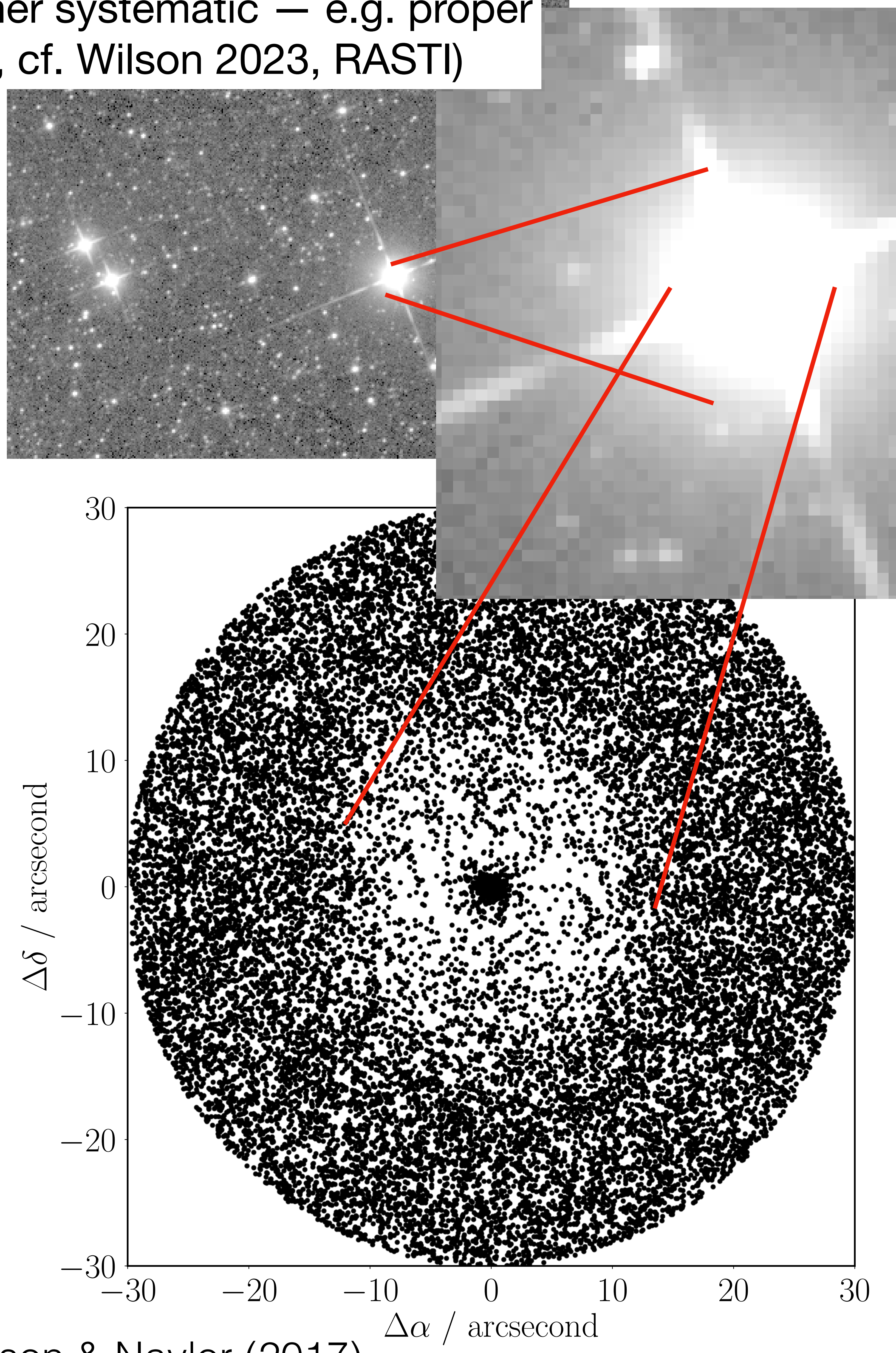
Tom J Wilson @onoddil

Additional Components of the AUF

$$P(\zeta, \lambda, k|\gamma, \phi) = \frac{1}{K} \times \prod_{\delta \notin \zeta \cap \delta \in \gamma} N_\gamma f_\gamma^\delta \prod_{\omega \notin \lambda \cap \omega \in \phi} N_\phi f_\phi^\omega \prod_{i=1}^{k} N_c \boldsymbol{G}_{\boldsymbol{\gamma\phi}}^{\boldsymbol{\zeta_i \lambda_i}} c_{\gamma\phi}^{\zeta_i \lambda_i}$$



(and any other systematic — e.g. proper motions, cf. Wilson 2023, RASTI)



Wilson & Naylor (2017)

WISE - Wright et al. (2010)

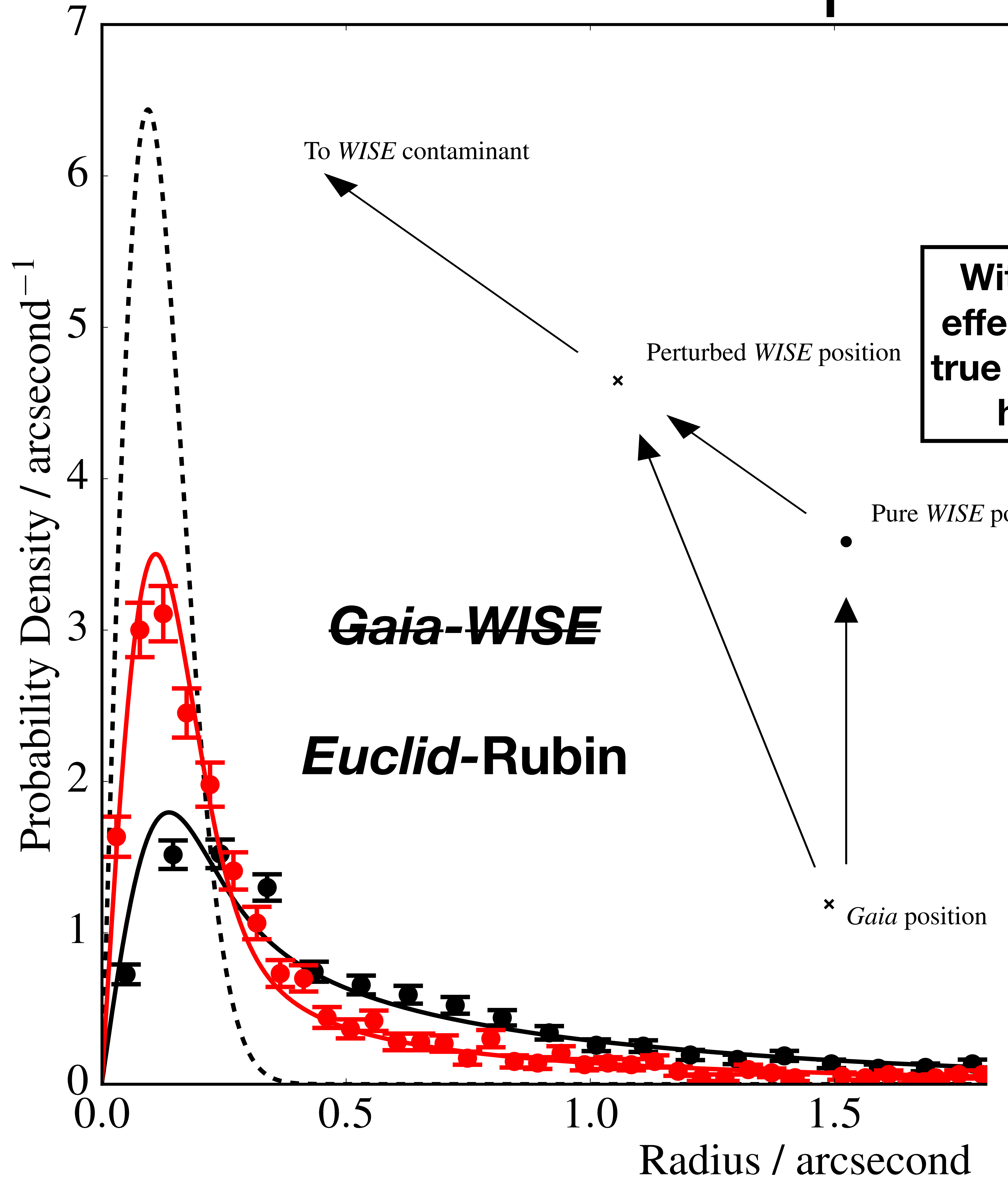
Gaia DR2 - Gaia Collaboration, Brown A. G. A., et al. (2018)

Tom J Wilson @onoddil

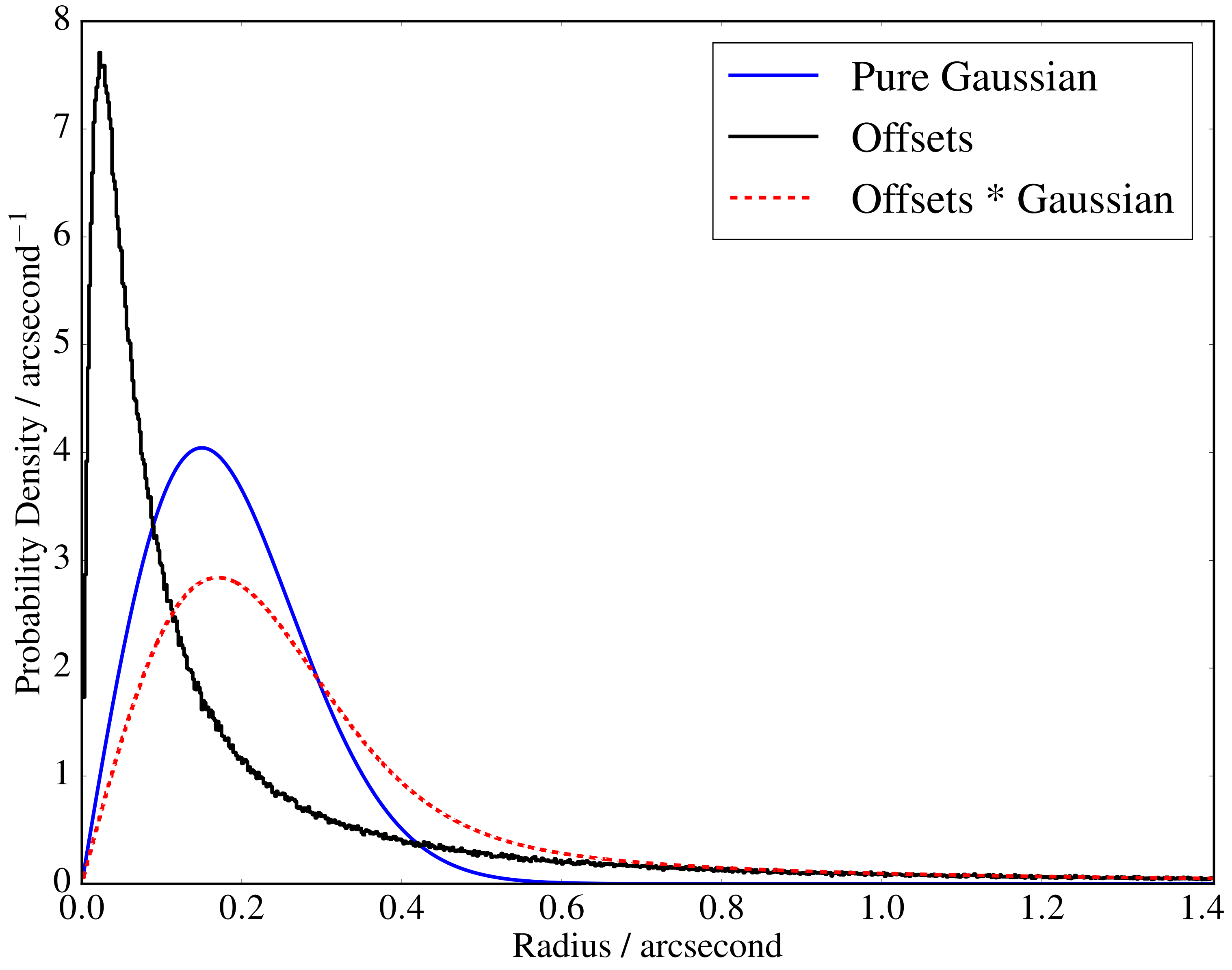
Additional Components of the AUF

$$P(\zeta,\lambda,k|\gamma,\phi) = \frac{1}{K} \times \prod_{\delta\notin\zeta\cap\delta\in\gamma} N_\gamma f_\gamma^\delta \prod_{\omega\notin\lambda\cap\omega\in\phi} N_\phi f_\phi^\omega \prod_{i=1}^{k} N_c \boldsymbol{G}_{\gamma\phi}^{\zeta_i\lambda_i} c_{\gamma\phi}^{\zeta_i\lambda_i}$$

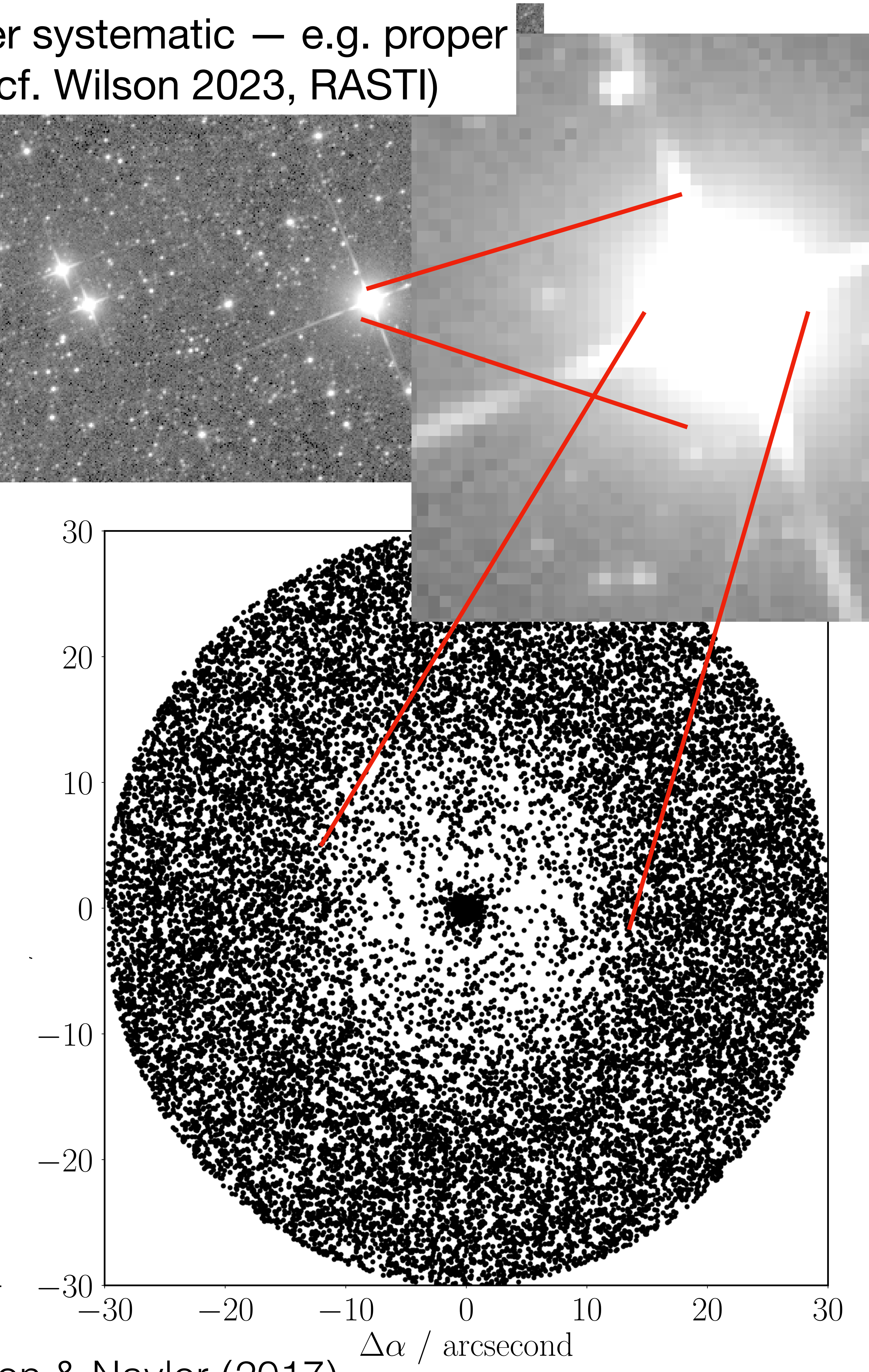
(and any other systematic — e.g. proper motions, cf. Wilson 2023, RASTI)



Without modelling this extra effect, we fail to recover many true pairings, with an artificially high false negative rate!



Wilson & Naylor (2018b)



Wilson & Naylor (2017)

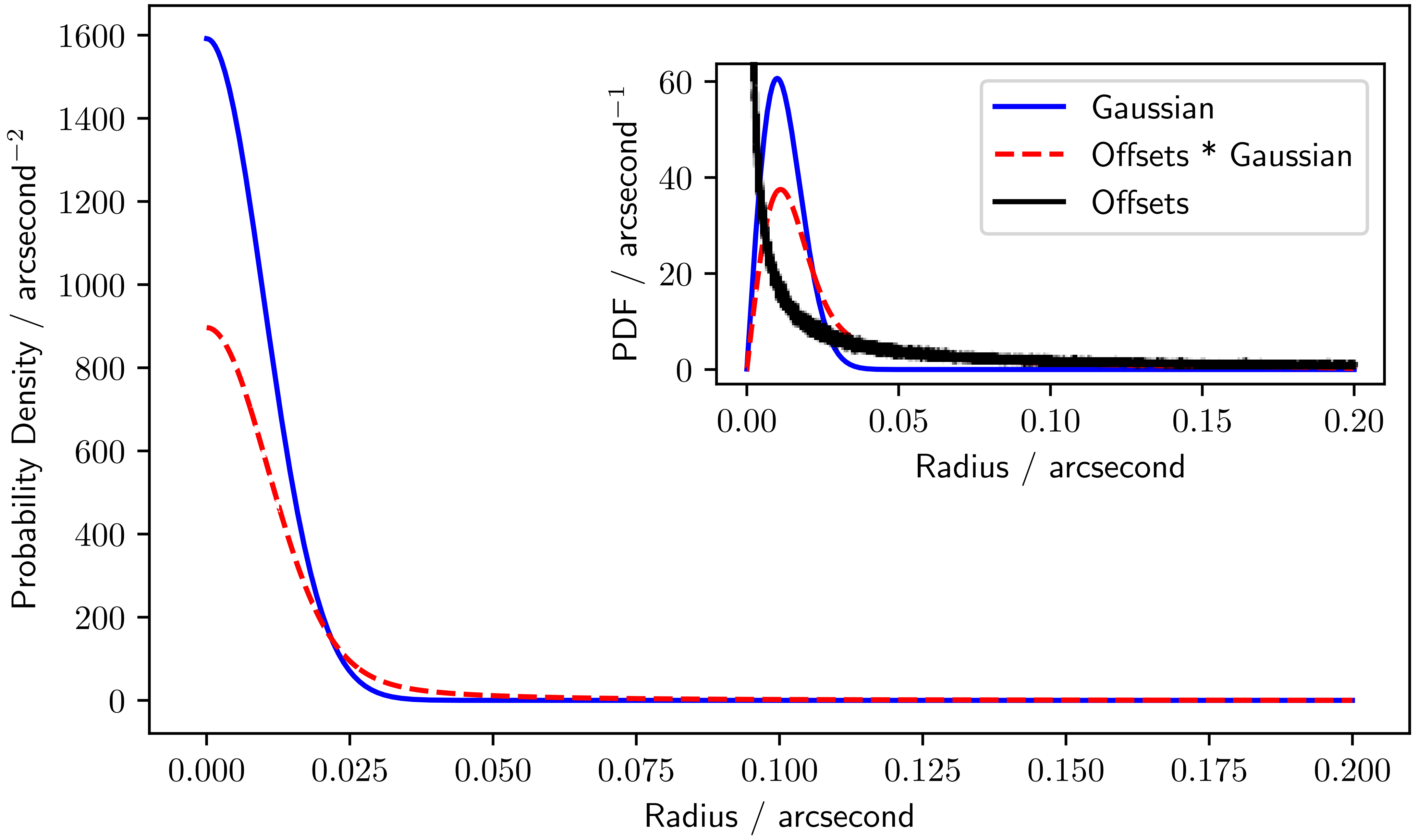
WISE - Wright et al. (2010)
Gaia DR2 - Gaia Collaboration, Brown A. G. A., et al. (2018)

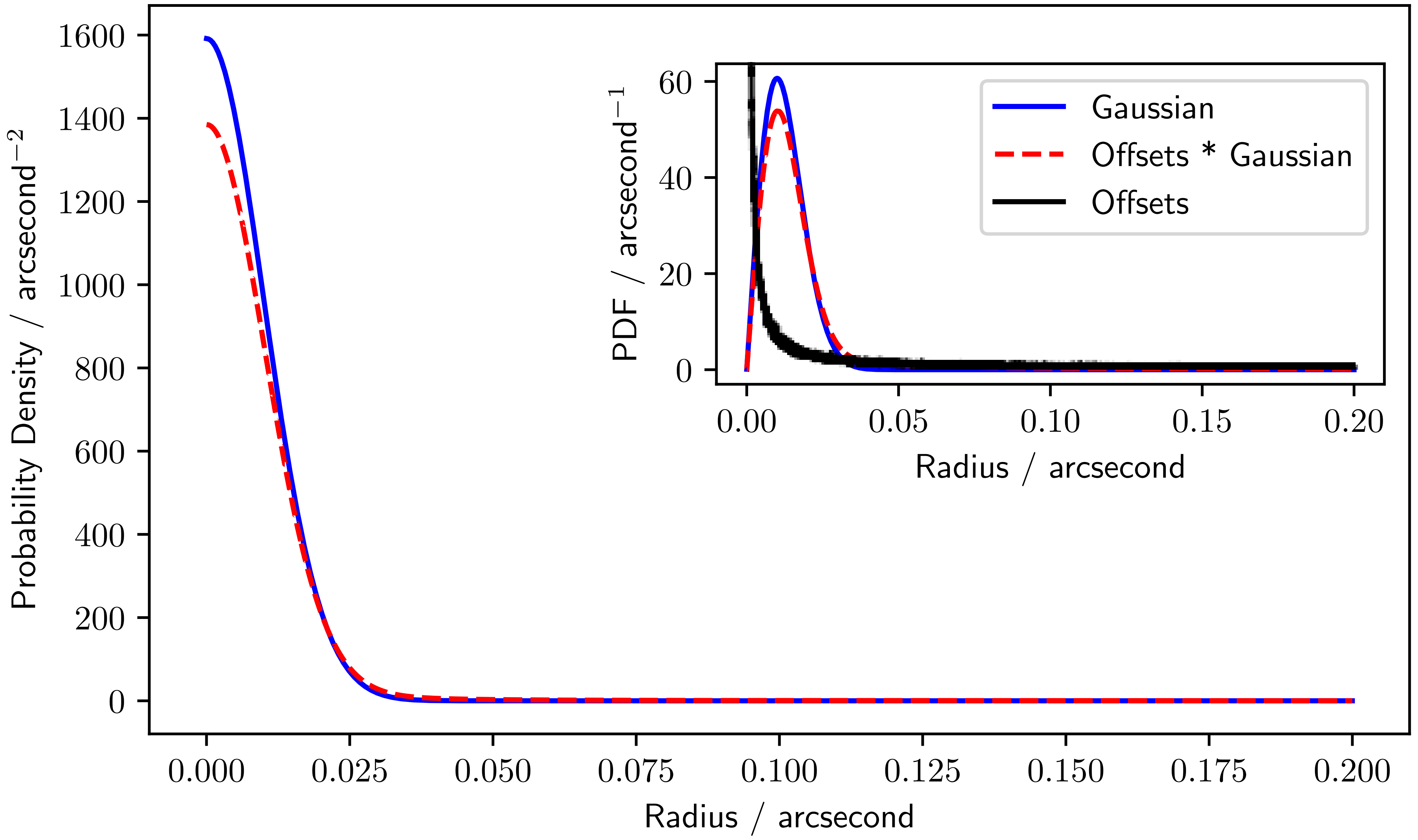
Tom J Wilson @onoddil

The Rubin AUF: Galactic Plane

Galactic Centre | **Not the Galactic Centre**

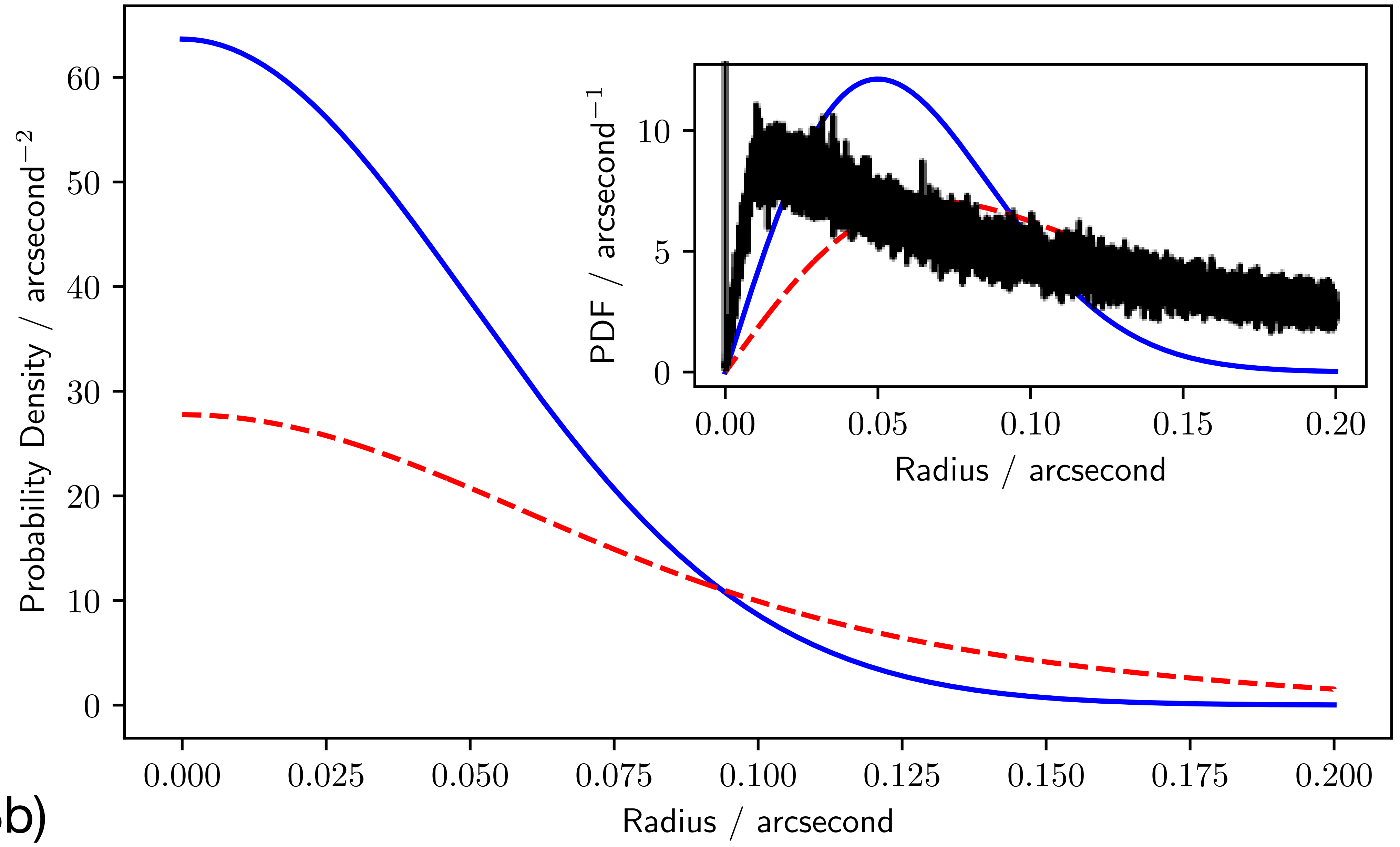
Single-visit

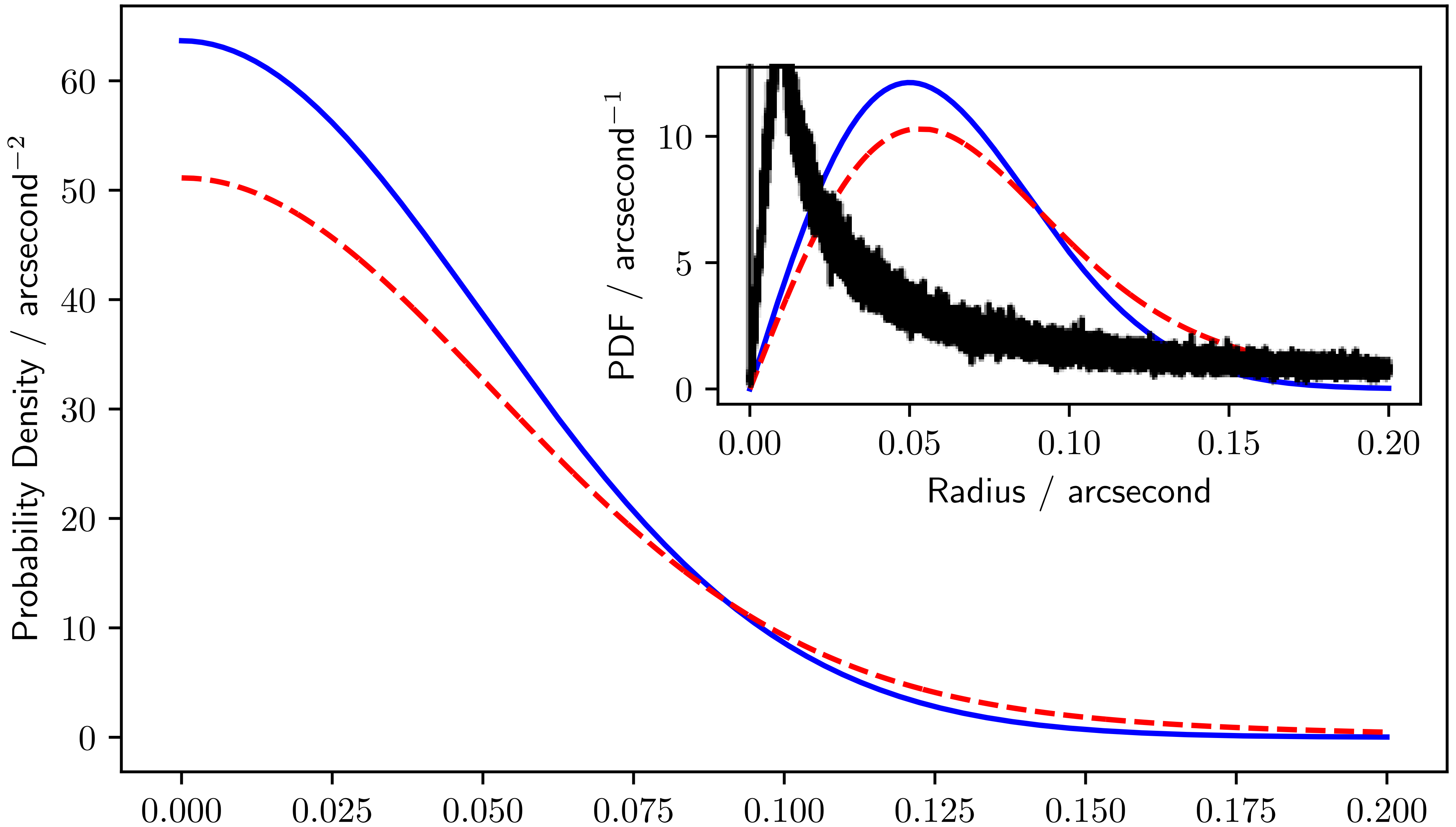




Without modelling this extra effect, we fail to recover many true pairings, with an artificially high false negative rate!

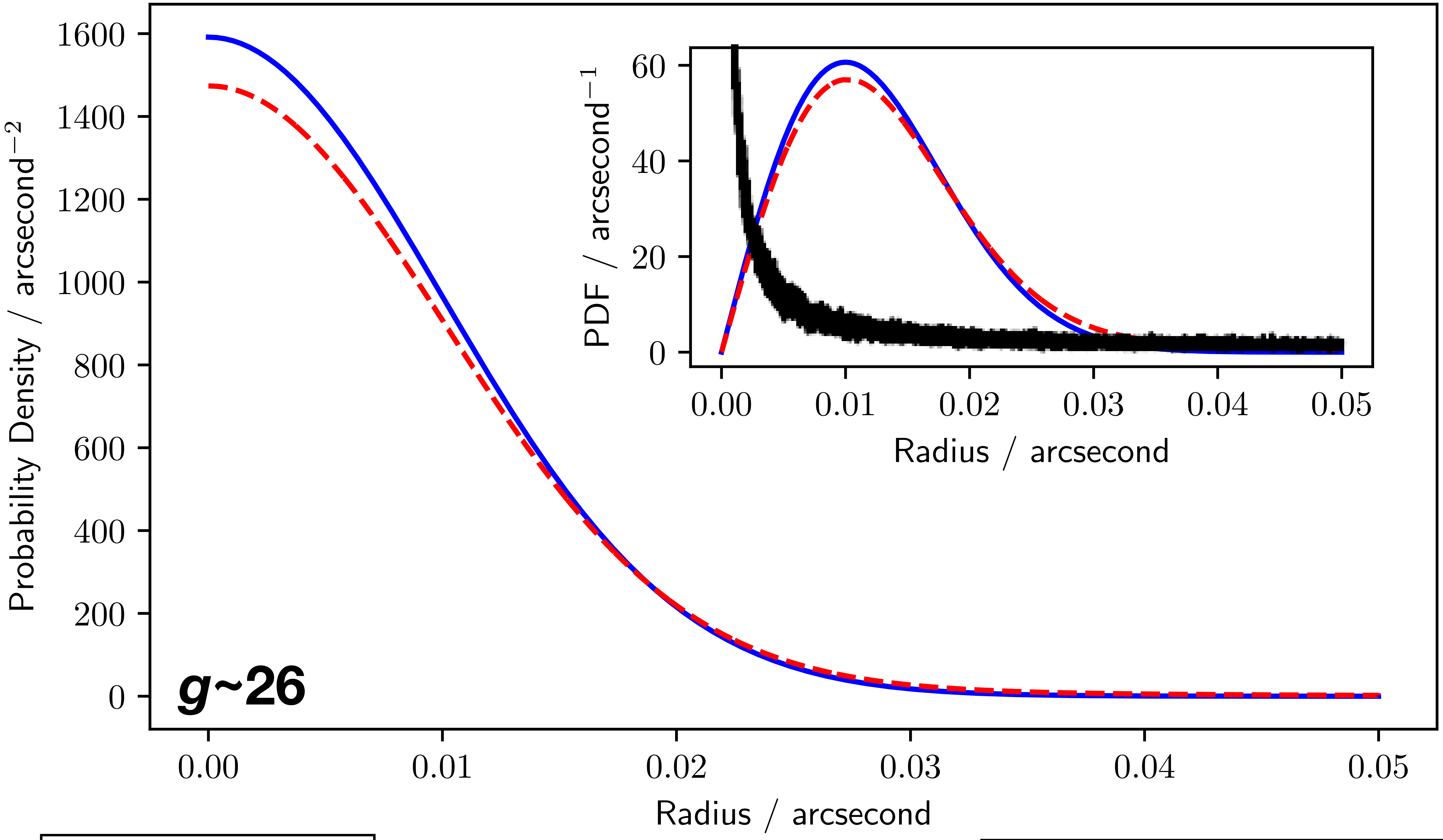
Co-add

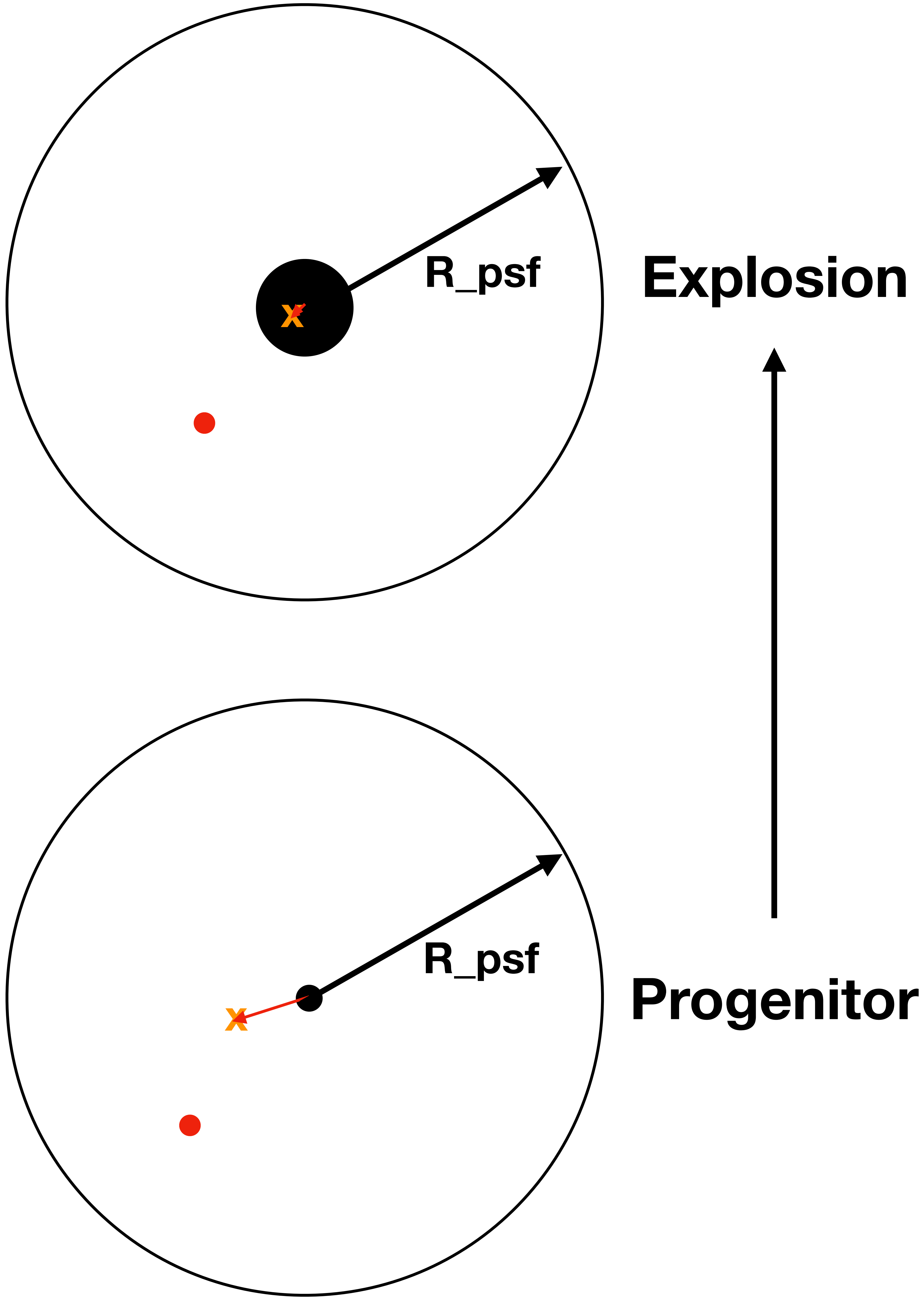




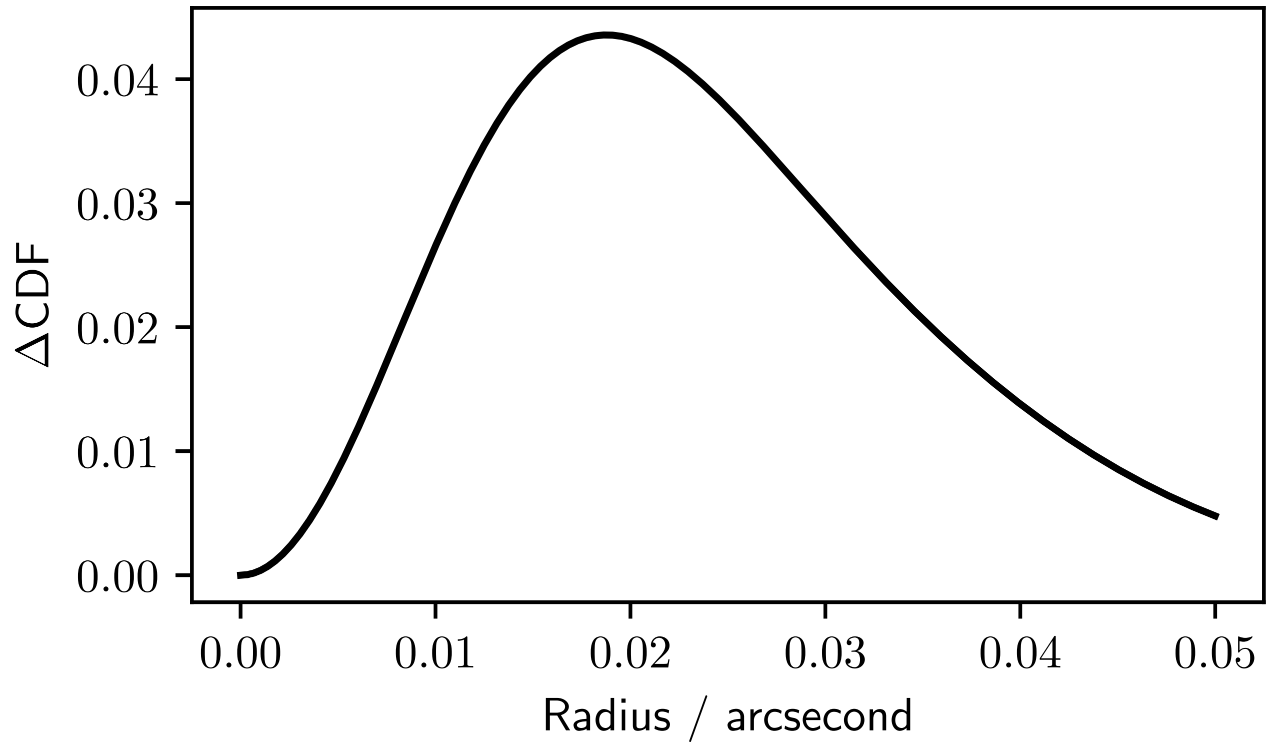
Wilson & Naylor (2018b)

The Rubin AUF: Extra-Galactic, Transients





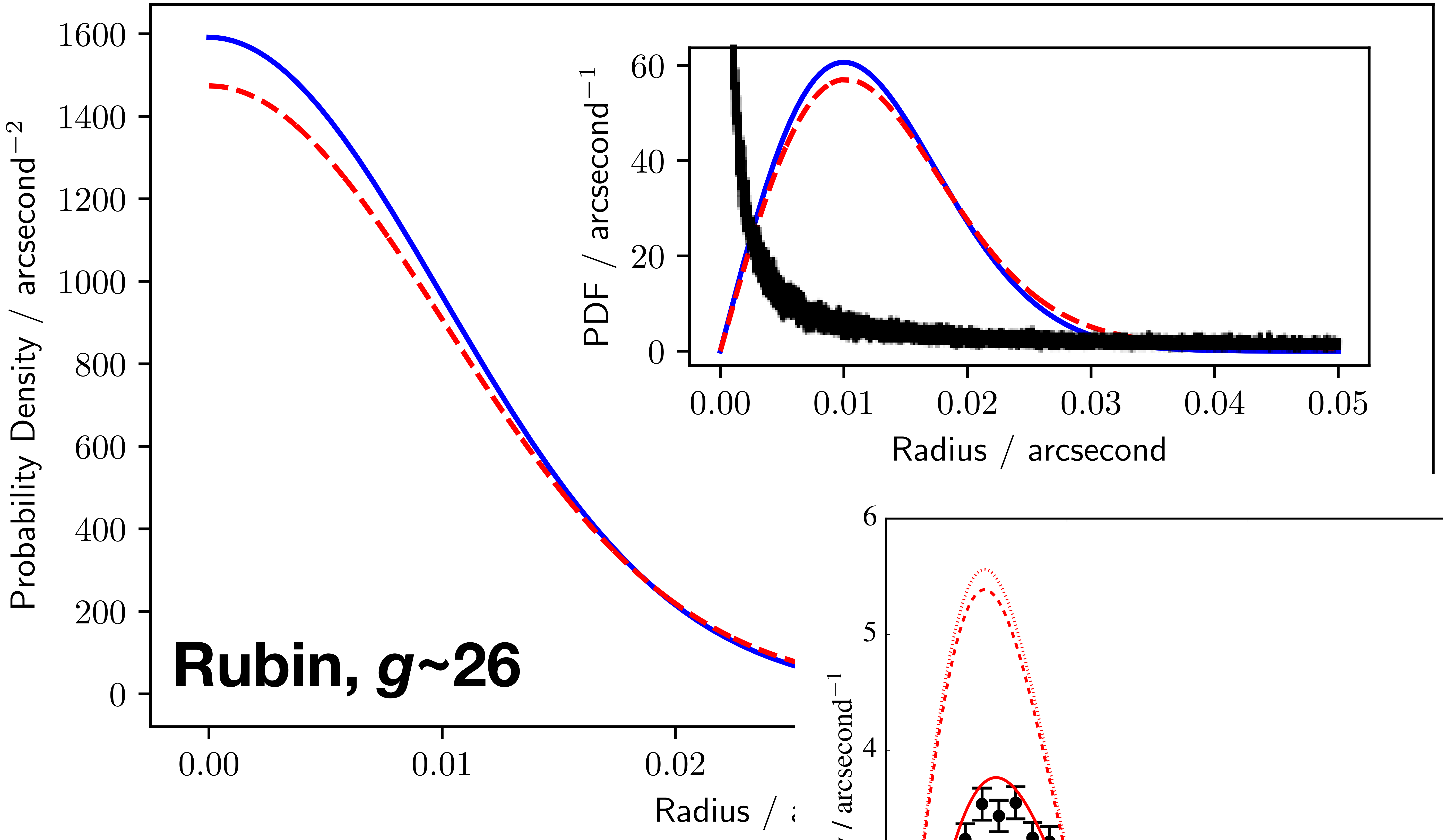
Without modelling this extra effect, we fail to recover many true pairings, with an artificially high false negative rate!

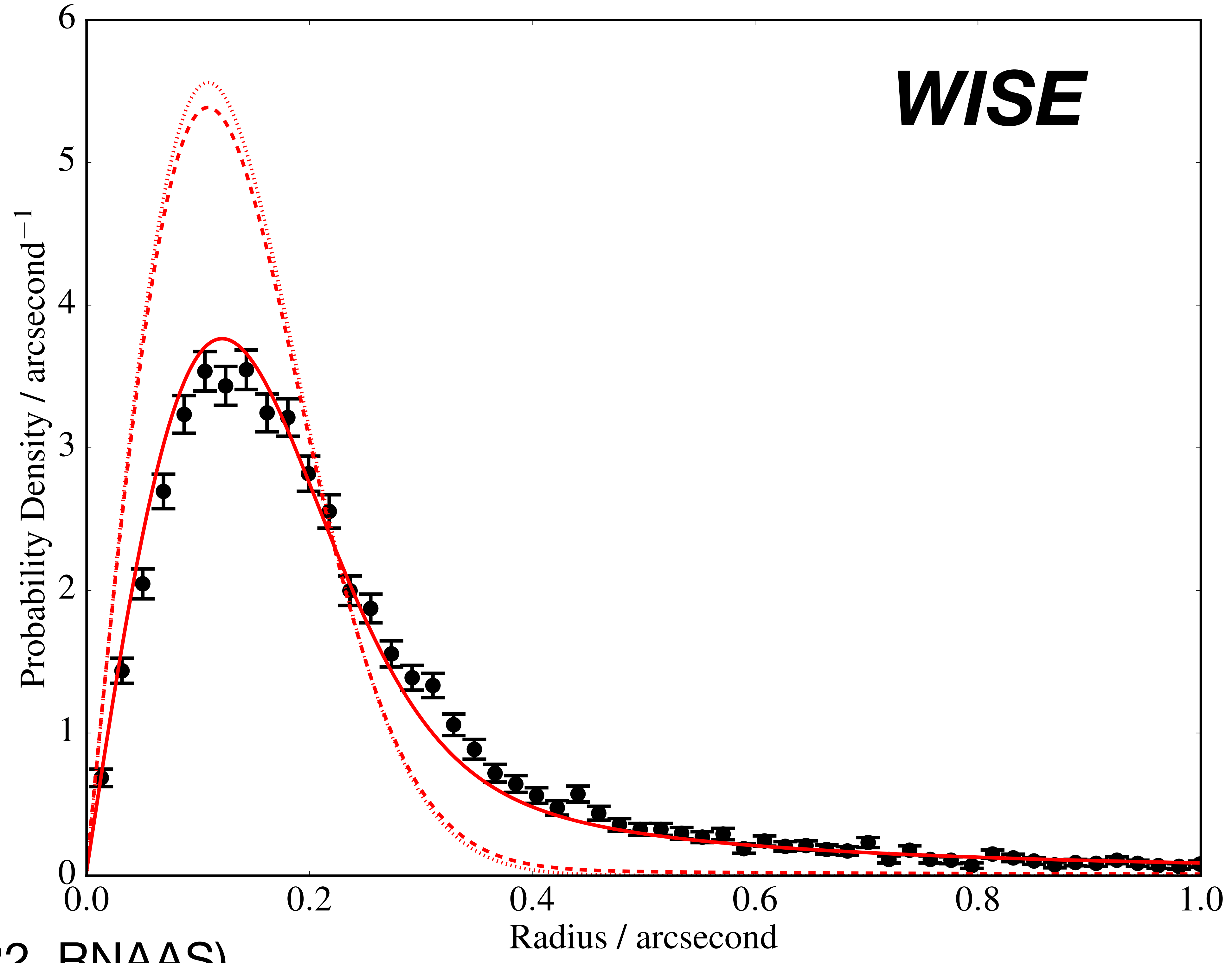


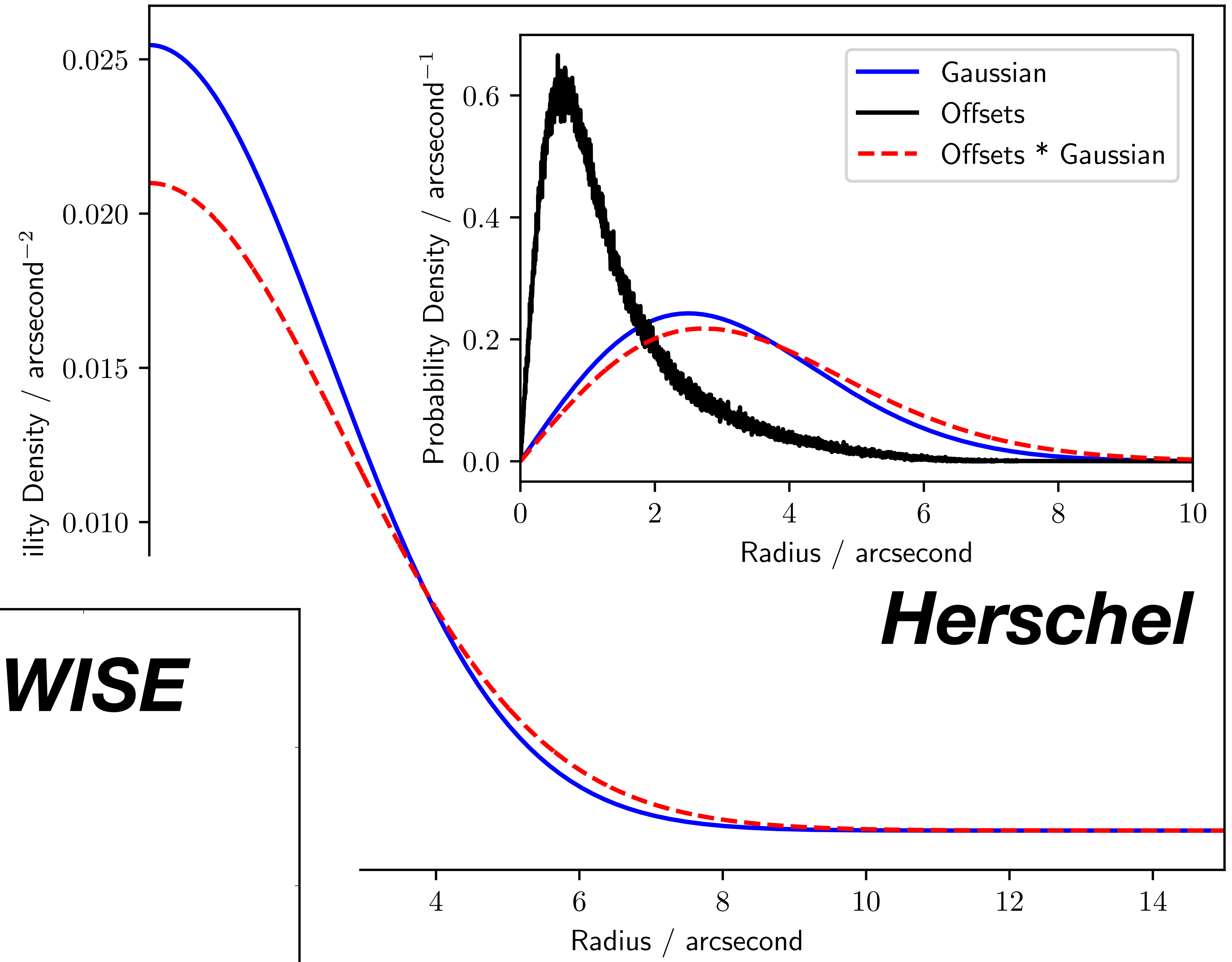
Wilson & Naylor (2018b); also see Wilson (2022, RNAAS)

Tom J Wilson @onoddil

The Rubin AUF: Extra-Galactic, Transients







Without modelling this extra effect, we fail to recover many true pairings, with an artificially high false negative rate!

Wilson & Naylor (2018b); also see Wilson (2022, RNAAS)

Tom J Wilson @onoddil

Cross-Match Tools, Framework, Usage

Tom J Wilson @onoddil

Matching Across Catalogues using the Astrometric Uncertainty Function and Flux

https://github.com/macauff/macauff

Tom J Wilson @onoddil

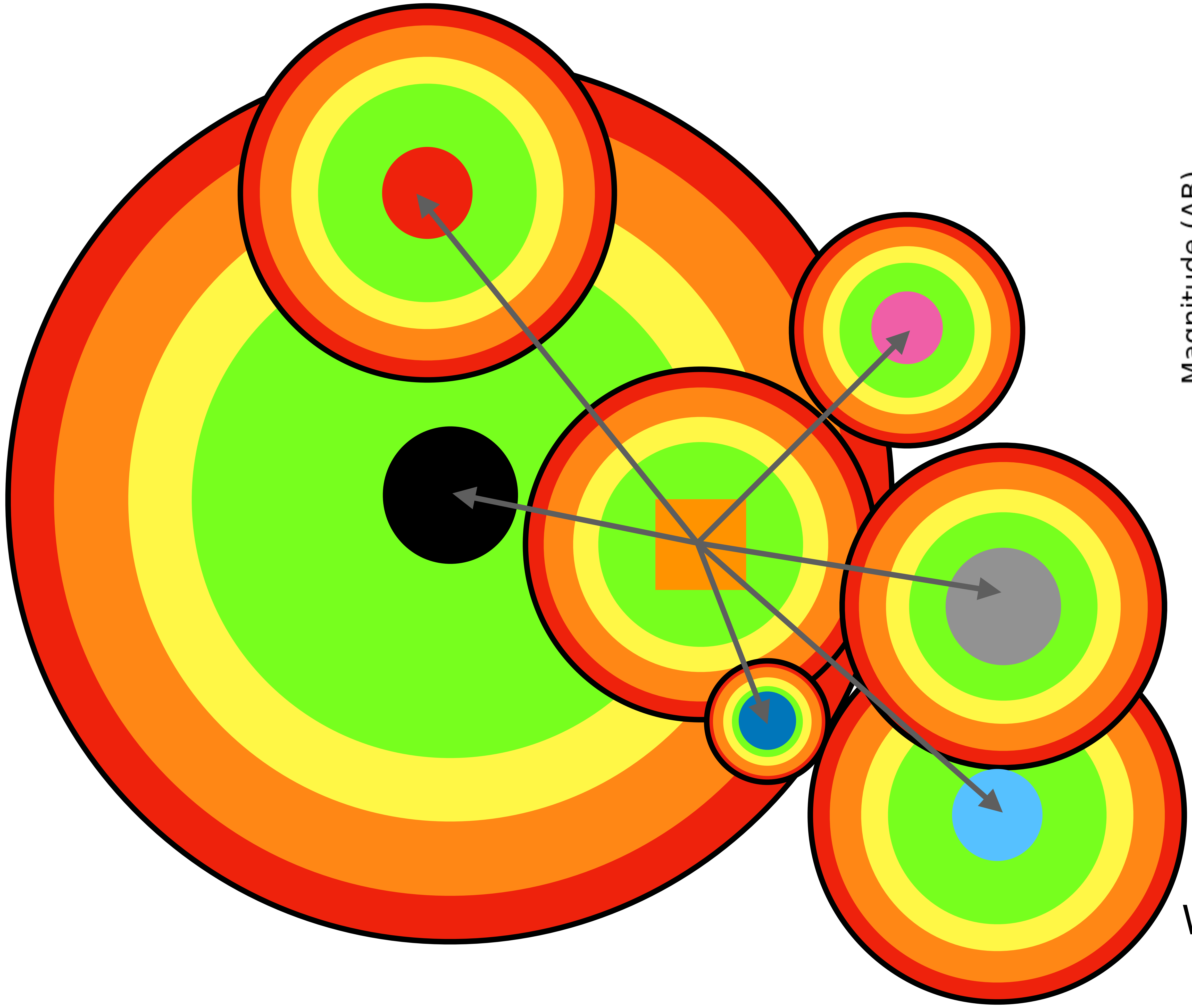
The Rubin "Super-Match"

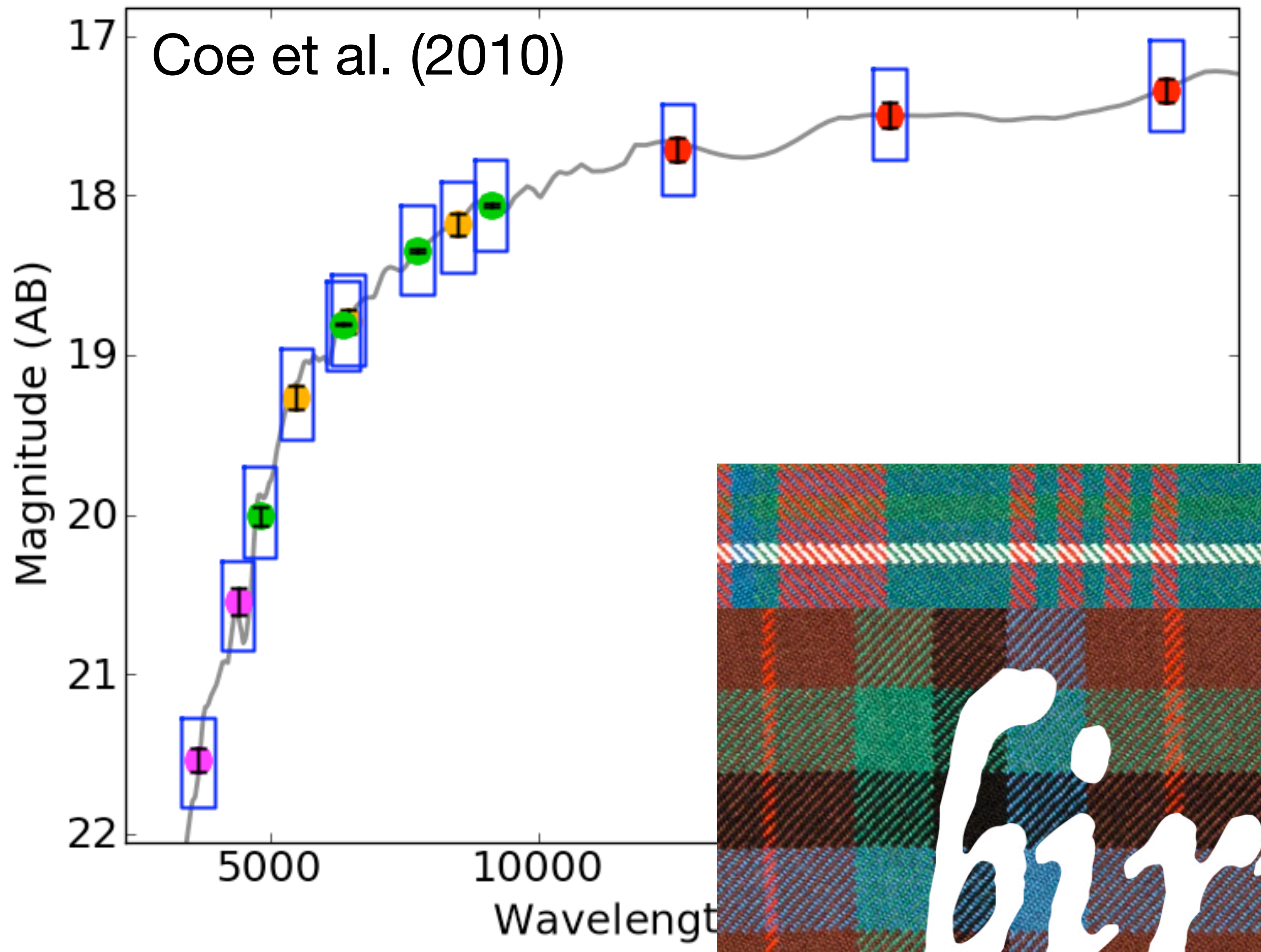
Bringing Independent Results together to Notify of Associations across Multiple catalogues

LSST -> *Gaia*, *WISE*, VISTA, *Euclid*, SDSS, … matches

Quick and easy construction of spectral energy distributions for each LSST source

Includes SED probabilities, individual match reliability, contamination statistics etc.





https://github.com/macauff/macauff
https://github.com/macauff/birnam

Wilson & Naylor (in prep.)
Pineau et al. (2017)

Tom J Wilson @onoddil

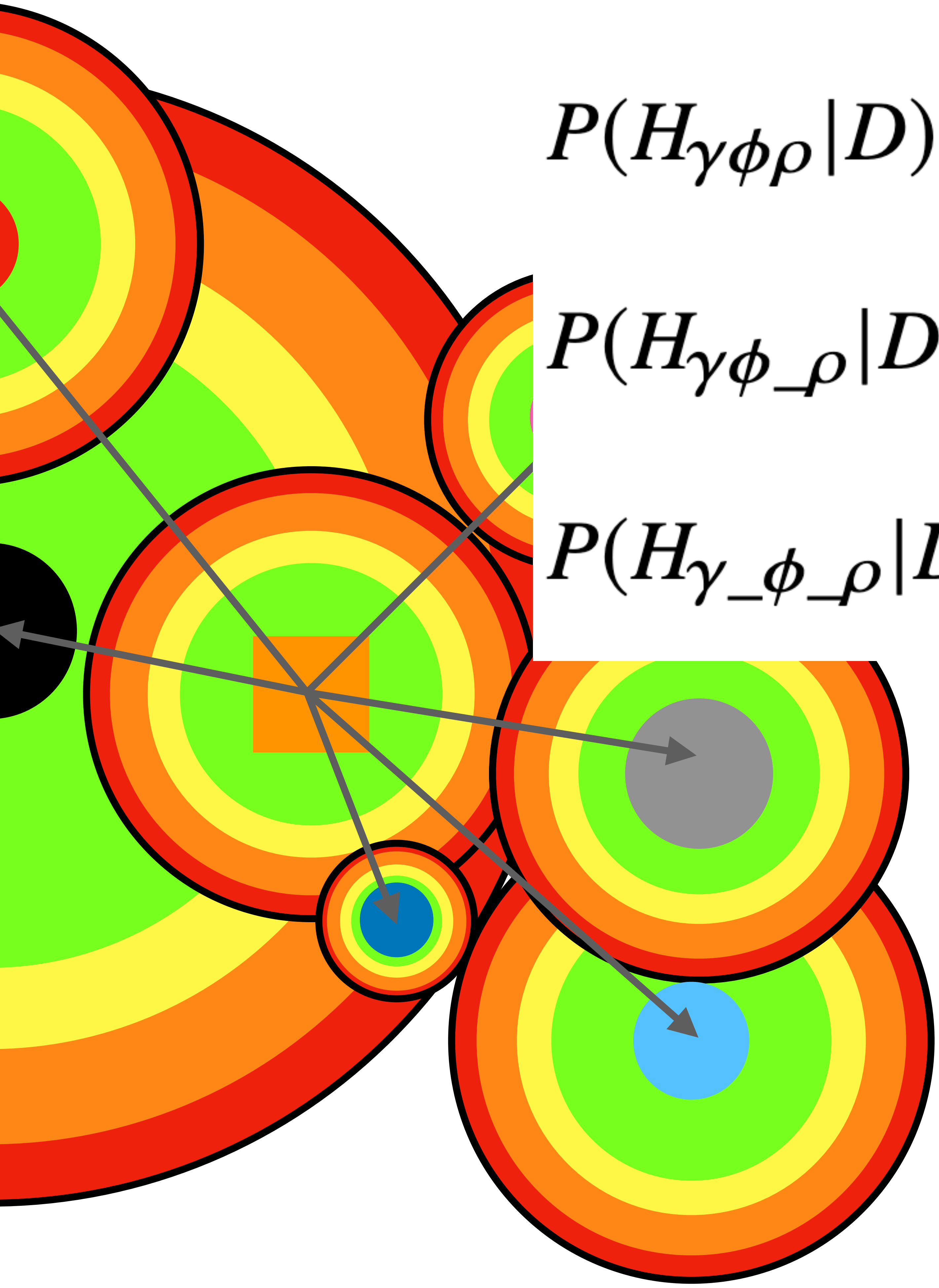
The Rubin “Super-Match”

Bringing Independent Results together to Notify of Associations across Multiple catalogues

LSST -> *Gaia*, *WISE*, VISTA, *Euclid*, SDSS, … matches

Quick and easy construction of spectral energy distributions for each LSST source

Includes SED probabilities, individual match reliability, contamination statistics etc.



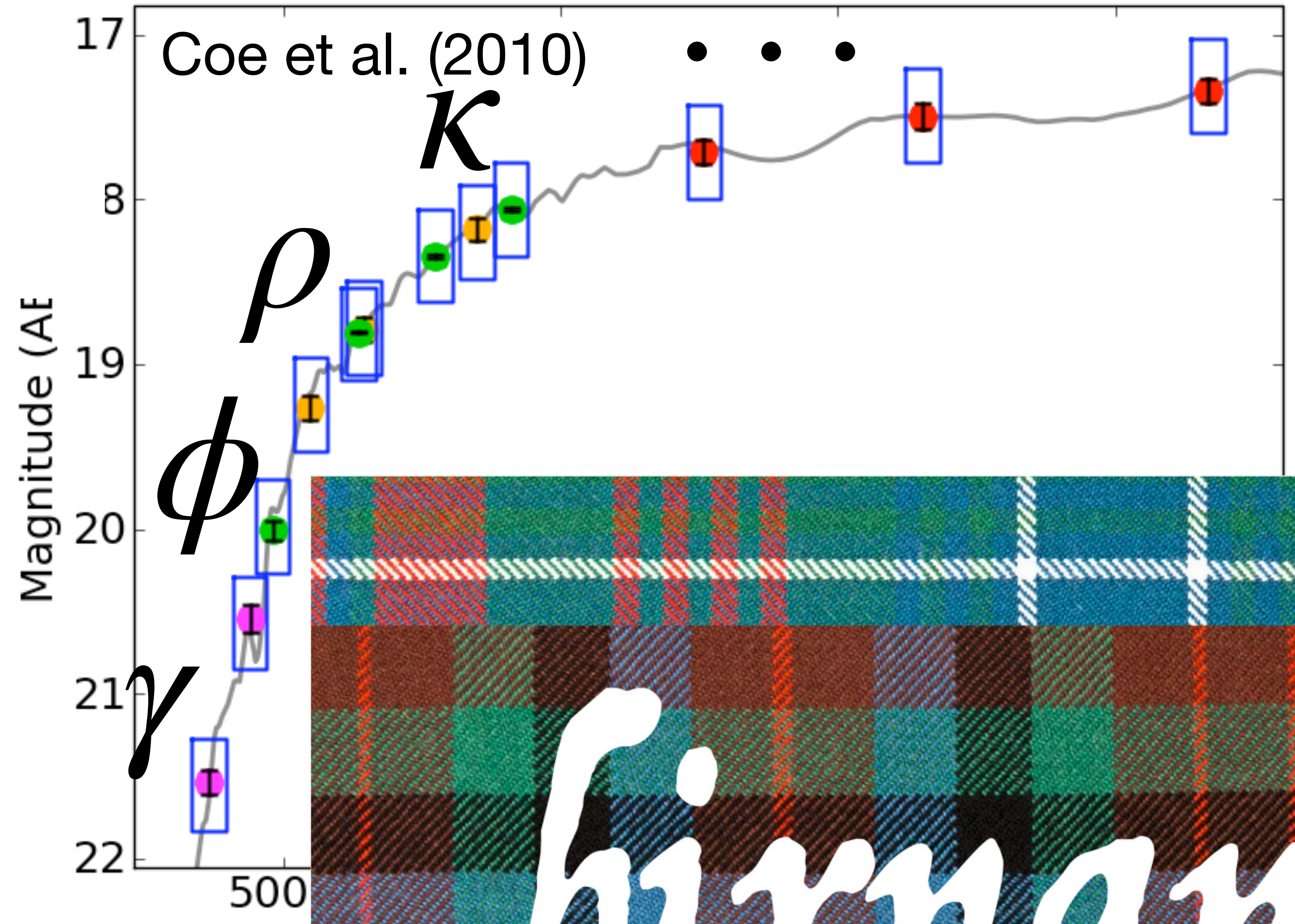
$$P(H_{\gamma\phi\rho}|D) = \frac{1}{K} N_{\gamma\phi} G_{\gamma\phi} N_{\gamma\rho} G_{\gamma\rho} = A_{\gamma\phi} A_{\gamma\rho},$$

$$P(H_{\gamma\phi\_\rho}|D) = \frac{1}{K} N_{\gamma\phi} G_{\gamma\phi} N_{\gamma} N_{\rho} = B_{\gamma\phi} A_{\gamma\rho},$$

$$P(H_{\gamma\_\phi\_\rho}|D) = \frac{1}{K} N_{\gamma} N_{\phi} N_{\gamma} N_{\rho} = B_{\gamma\phi} B_{\gamma\rho},$$

$$A_{\gamma\phi} = \frac{N_{\gamma\phi} G_{\gamma\phi}}{N_{\gamma} N_{\phi} + N_{\gamma\phi} G_{\gamma\phi}}$$

$$B_{\gamma\phi} = \frac{N_{\gamma} N_{\phi}}{N_{\gamma} N_{\phi} + N_{\gamma\phi} G_{\gamma\phi}}$$



Wilson & Naylor (2018a)

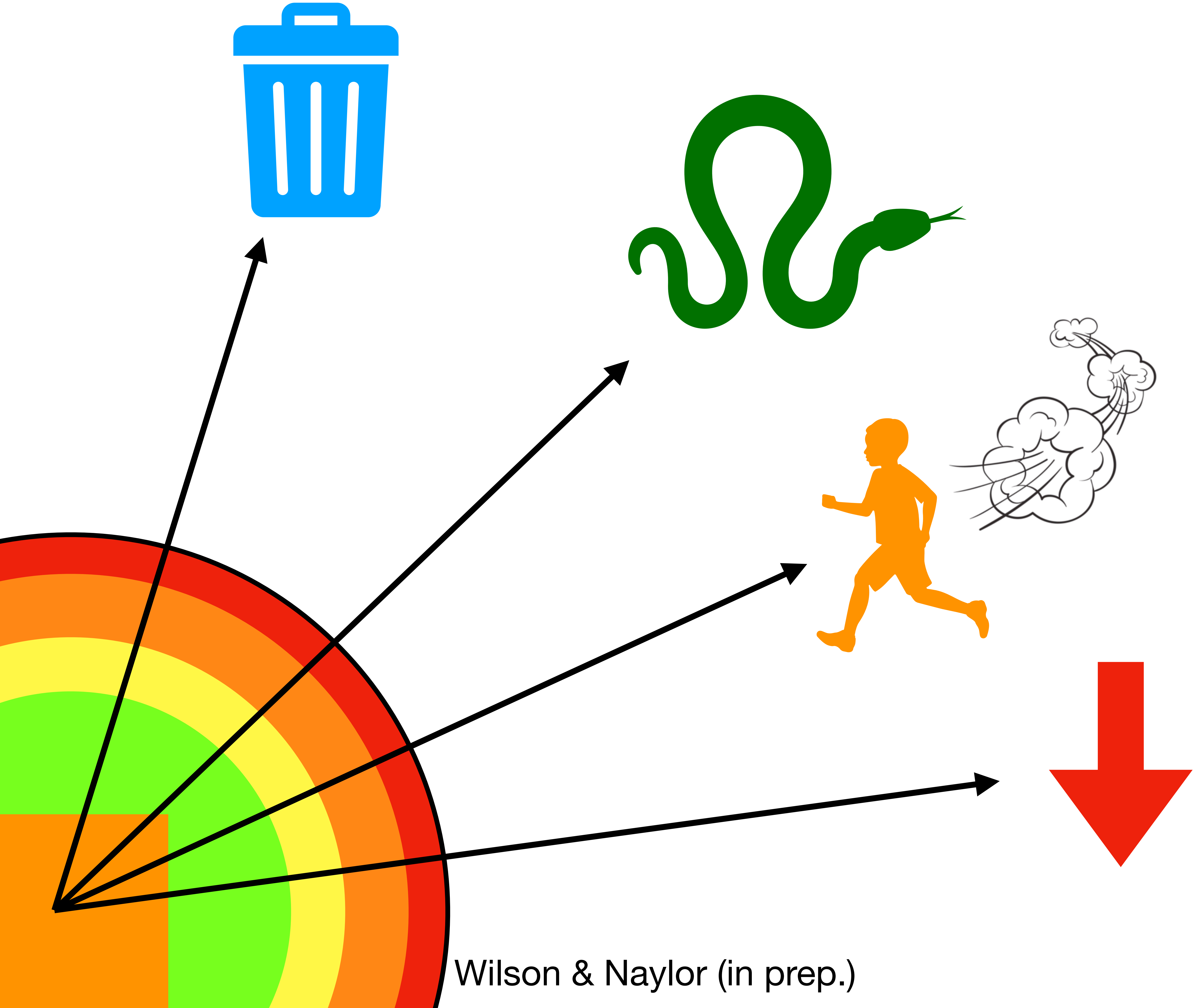
Wilson & Naylor (in prep.)
Pineau et al. (2017)

https://github.com/macauff/macauff
https://github.com/macauff/birnam

Tom J Wilson @onoddil

Confirming Lonely Rubin Sources

Blanks And Near-misses, Questionable sources, Upper-limits, and Objects of varying brightness



Most LSST sources will be “lonely” with 15x as many sources as the next dataset. We will follow up all non-matches, and confirm whether these objects are:

- Image artefacts
- Astrophysically variable objects
- High proper motion sources
- Regular objects that are simply too faint to be seen in the opposing catalogue

Wilson & Naylor (in prep.)

https://github.com/macauff/macauff

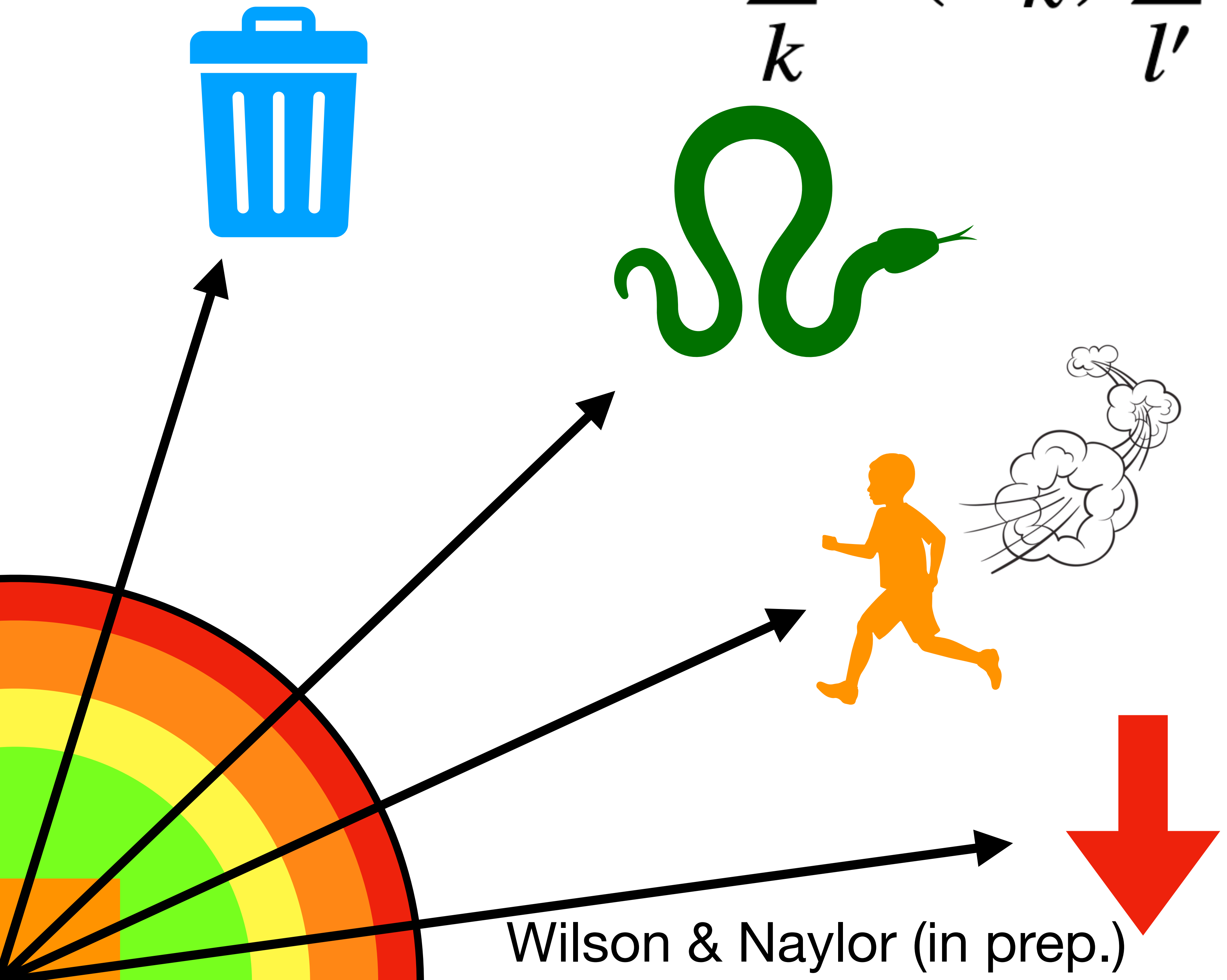
Tom J Wilson @onoddil

Confirming Lonely Rubin Sources

Blanks And Near-misses, Questionable sources, Upper-limits, and Objects of varying brightness

$$P(H_i, M_j|\mathbf{D}) = \frac{P(H_i)P(M_j|H_i)p(\mathbf{D}|H_i, M_j)}{p(\mathbf{D})}$$

$$P(H_i, M_j|\mathbf{D}) = \frac{P(H_i)L_j}{\sum_k P(H_k)\sum_{l'} L_{l'} + \sum_{k'} P(H_{k'})}$$



$$L = \frac{Gc}{Nf}$$

Wilson & Naylor (in prep.)

Most LSST sources will be "lonely" with 15x as many sources as the next dataset. We will follow up all non-matches, and confirm whether these objects are:

- Image artefacts
- Astrophysically variable objects
- High proper motion sources
- Regular objects that are simply too faint to be seen in the opposing catalogue

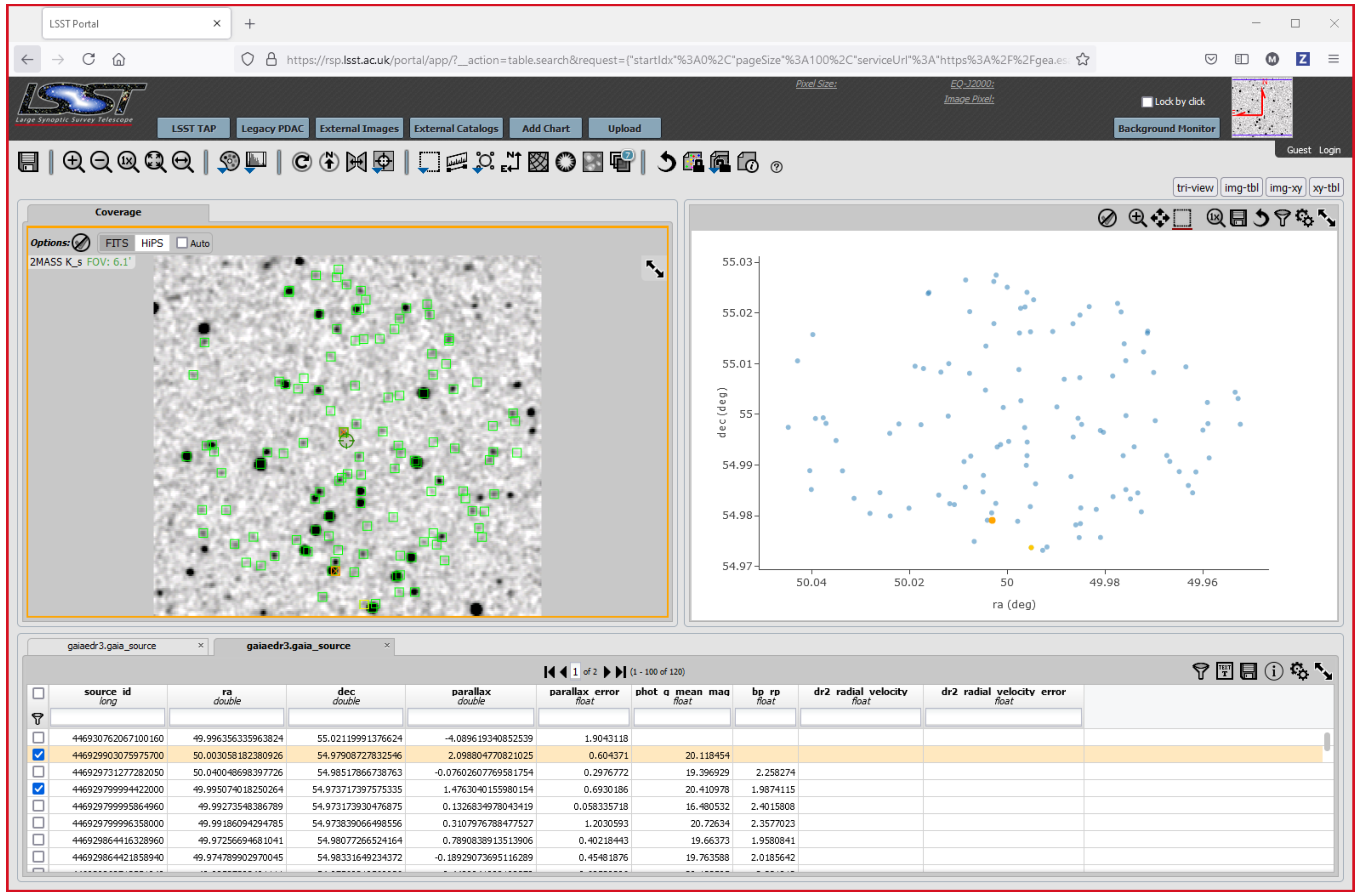
banquo

https://github.com/macauff/macauff

Tom J Wilson @onoddil

How To Use Our Cross-Matches

(Or, how this impacts you on a day-to-day basis)



Three tables per cross-match: merged catalogue dataset, and 2x non-match dataset (one per catalogue)

Example columns:

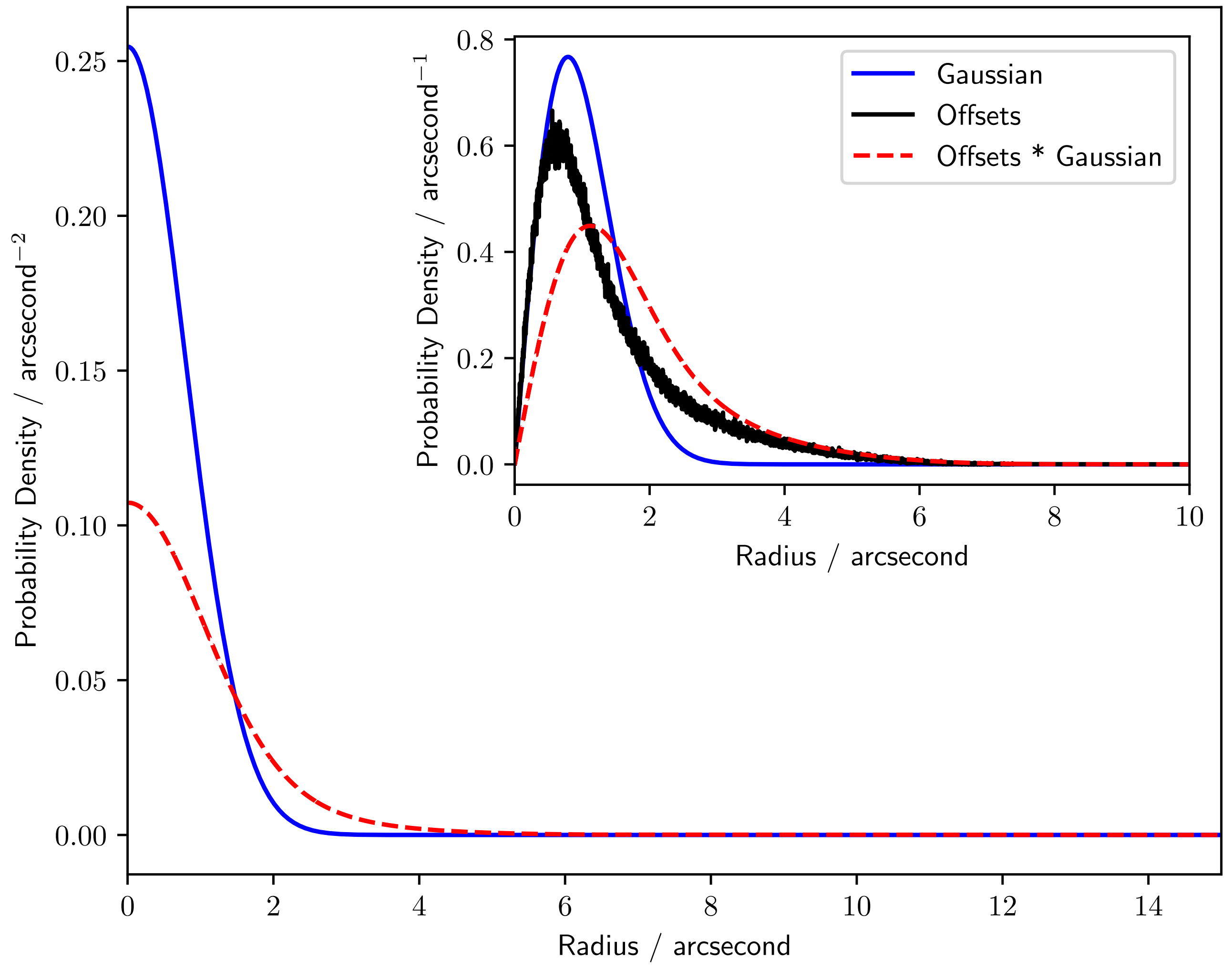
- **Designations of the two sources (e.g., WISE J… and *Gaia* EDR3…)**
- **RA and Dec (or Galactic l/b) of the two sources**
- **Magnitudes (corrected for necessary effects, such as e.g. *Gaia*) in all bandpasses for both objects**
- **Match probability — probability of the most likely permutation (see equation 26 of Wilson & Naylor 2018a)**
- **Eta - Photometric likelihood ratio (counterpart vs non-match probability, just for brightnesses; see eq37 of WN18a)**
- **Xi - Astrometric likelihood ratio (just position match/non-match comparison; see eq38 of WN18a)**
- **Average contamination - simulated mean (percentile) brightening of the two sources, based on number density of catalogue**
- **Probability of sources having blended contaminant above e.g. 1% relative flux**

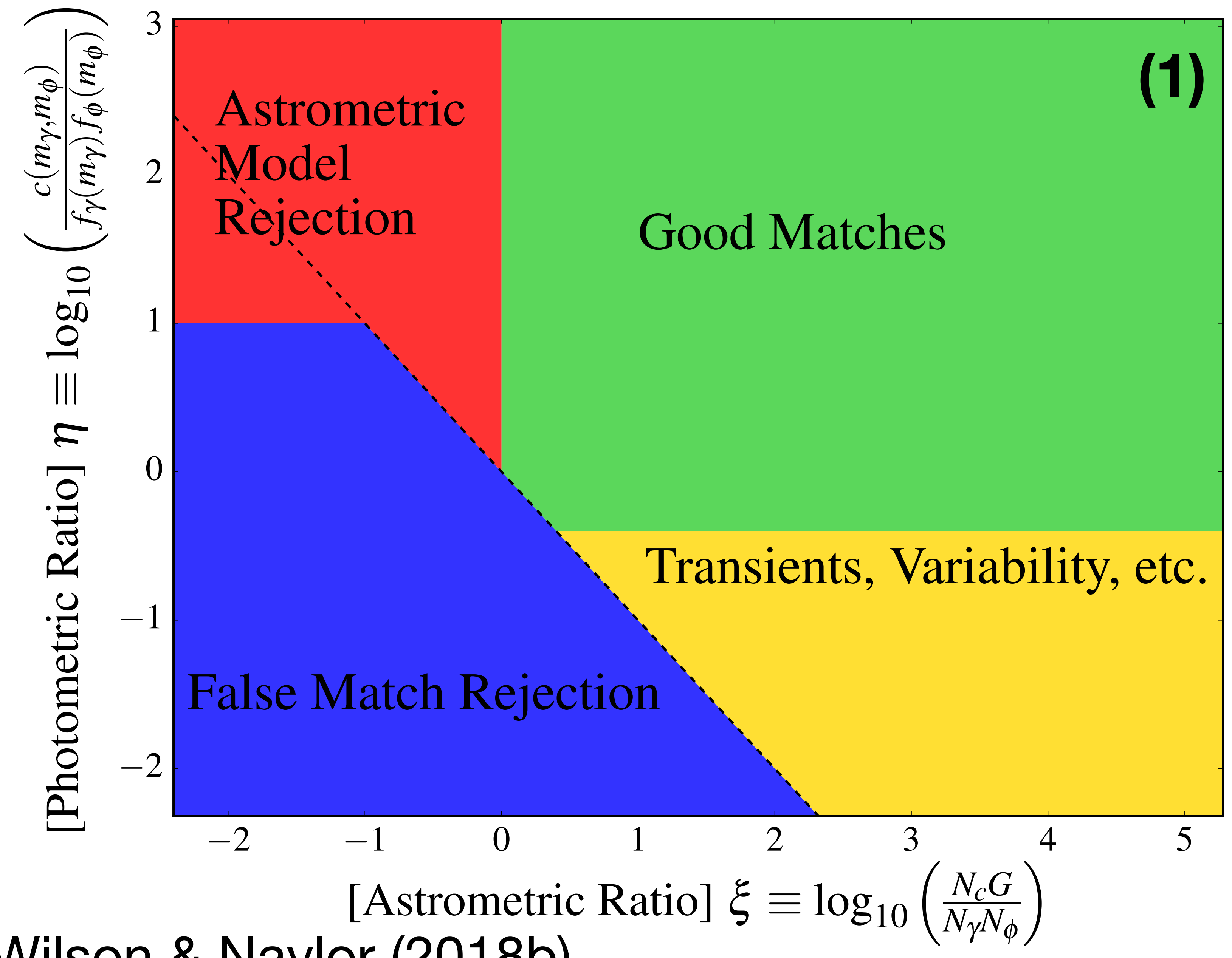
We will provide two match runs per catalogue pair match: one with, and one without, the photometry considered, to allow for the recovery of sources with “weird” colours but otherwise agreeable astrometry

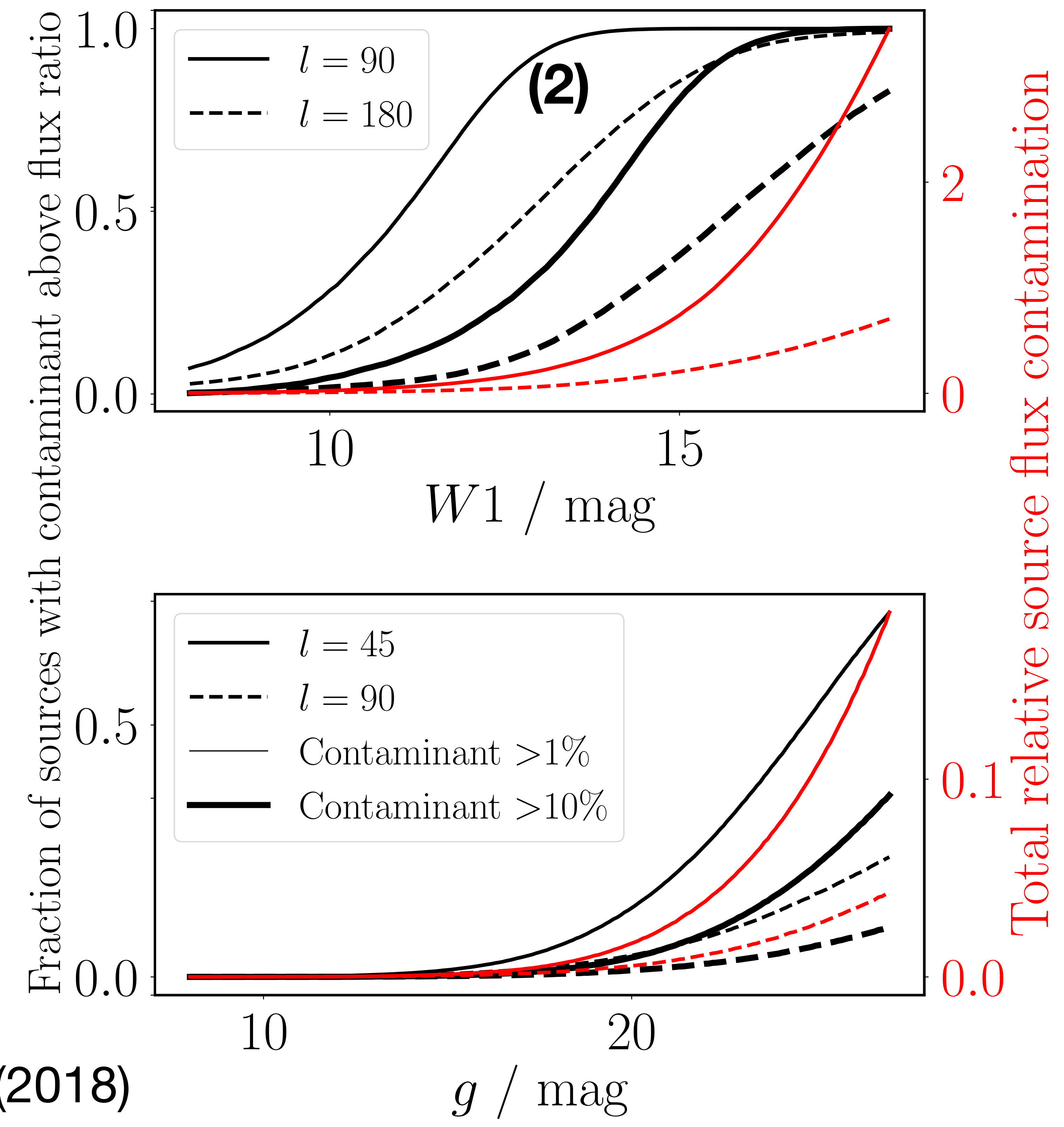
Tom J Wilson @onoddil

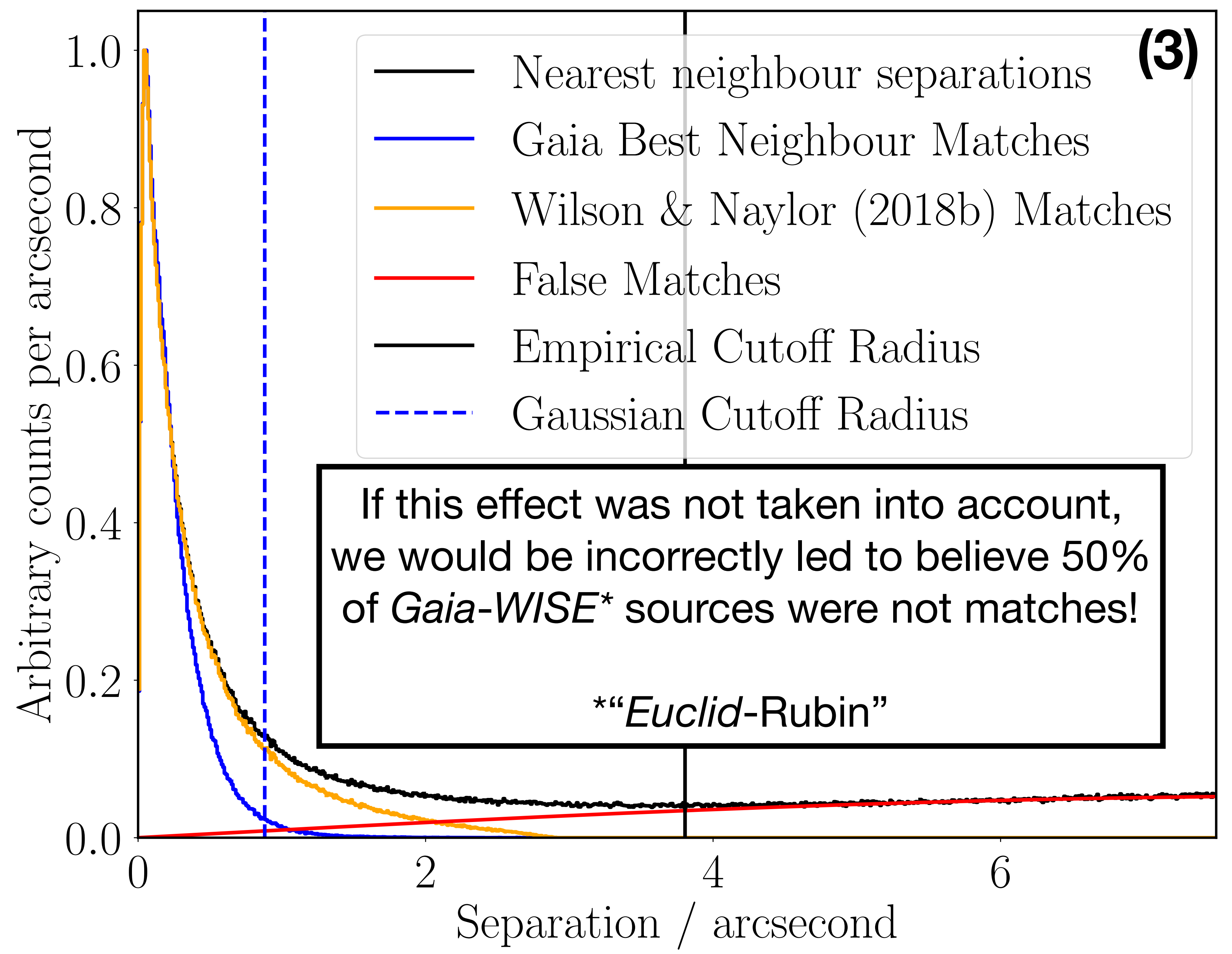
Why Use Our Cross-Matches?

0) Getting cross-matches, even for “well behaved” fields
1) Finding “odd” objects, either using the inclusion vs non-inclusion of the photometry in the two match runs, or via the likelihood ratio space — separately-planned “real time” matching service for transient objects
2) Removing e.g. IR excess or correcting for extinction-like crowding brightening, through Average Contamination; crucial for “1% photometry” in both precision *and* accuracy
3) Recovering additional sources missed by other match services — either in crowded fields (we recover up to twice as many *Gaia*-*WISE* matches than the *Gaia* best neighbour matches), or with our extension to unknown proper motion modelling as an extra systematic









Wilson & Naylor (2018b)
WISE - Wright et al. (2010)
Gaia matches - Marrese et al. (2019)
Gaia DR2 - Gaia Collaboration, Brown A. G. A., et al. (2018)

Tom J Wilson @onoddil

Conclusions

- **Upcoming LSST:UK cross-match service macauff — let me know your thoughts/needs/hopes/dreams**
    - **Provide tables of cross-matches between LSST and <your favourite catalogue here!>**

- **Our cross-matches include two key elements for avoiding issues with the crowded LSST sky**
    - **A generalised approach to the Astrometric Uncertainty Function allows for the inclusion of the effects of perturbation due to blended sources and unknown proper motions — reduce false -ves!**
    - **Use of the photometry of sources allows for the rejection of false matches (with >1 "extra" source per 2 arcsecond circle in most of the LSST sky) — reduce false +ves!**
- **Will include additional information on the crowding of sources, allowing for selection of uncontaminated objects, or modelling of excess flux — crucial for removal of red excess in SEDs**
    - **LSST will suffer of order 10% flux contamination, which could be confused with extinction**

- **macauff cross-match tools are being extended currently**
    - **We will provide an easy-to-use "SED grabber" tool for each LSST source**
    - **And follow up the $\gtrsim$93% of non-matched Rubin objects to confirm flux upper limits in other surveys**

University of Exeter

@Onoddil @pm.me .github.io

LSST:UK Consortium

UKRI Science and Technology Facilities Council

Vera C. Rubin Observatory

Wilson & Naylor, 2017, MNRAS, 468, 2517
Wilson & Naylor, 2018a, MNRAS, 473, 5570
Wilson & Naylor, 2018b, MNRAS, 481, 2148
Wilson, 2022, RNAAS, 6, 60
Wilson, 2023, RASTI, 2, 1

https://github.com/macauff/macauff

Macauff

Tom J Wilson @onoddil