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

Tom J. Wilson (he/him) and Tim Naylor
t.j.wilson@exeter.ac.uk

**LSST:UK All-Hands Meeting, 14/Apr/26**

Science and Technology Facilities Council

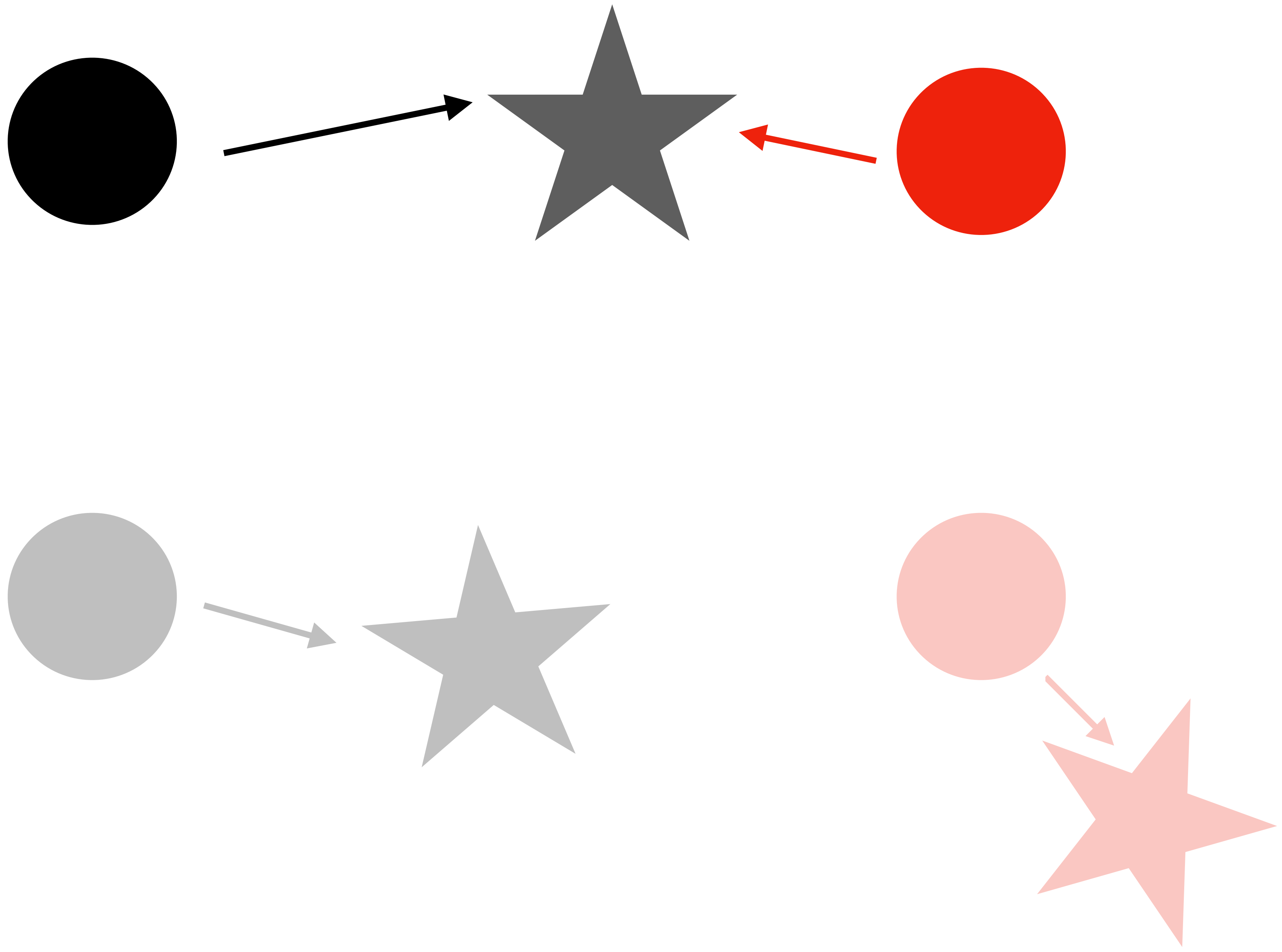
University of Exeter

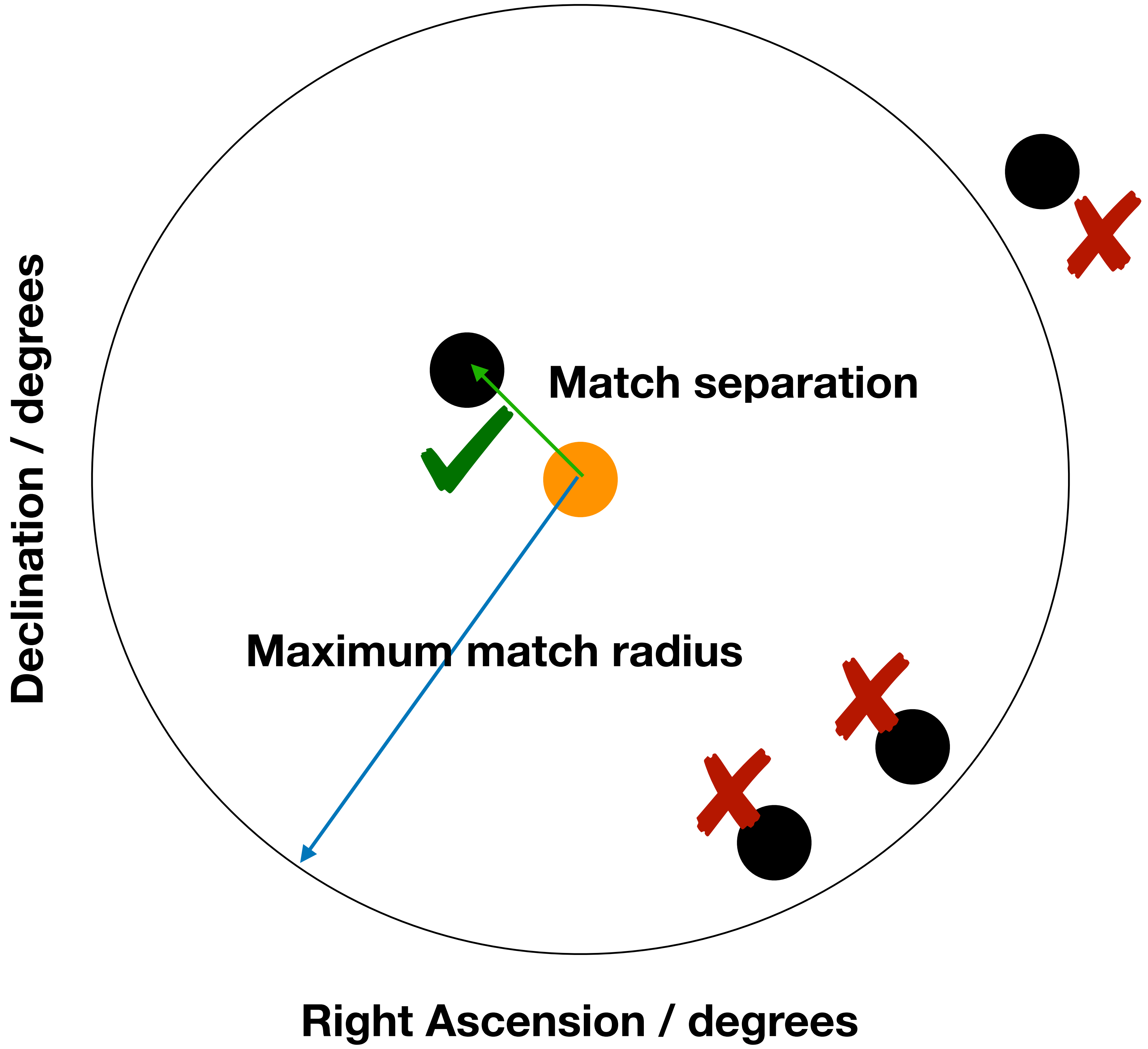
LSST:UK Consortium

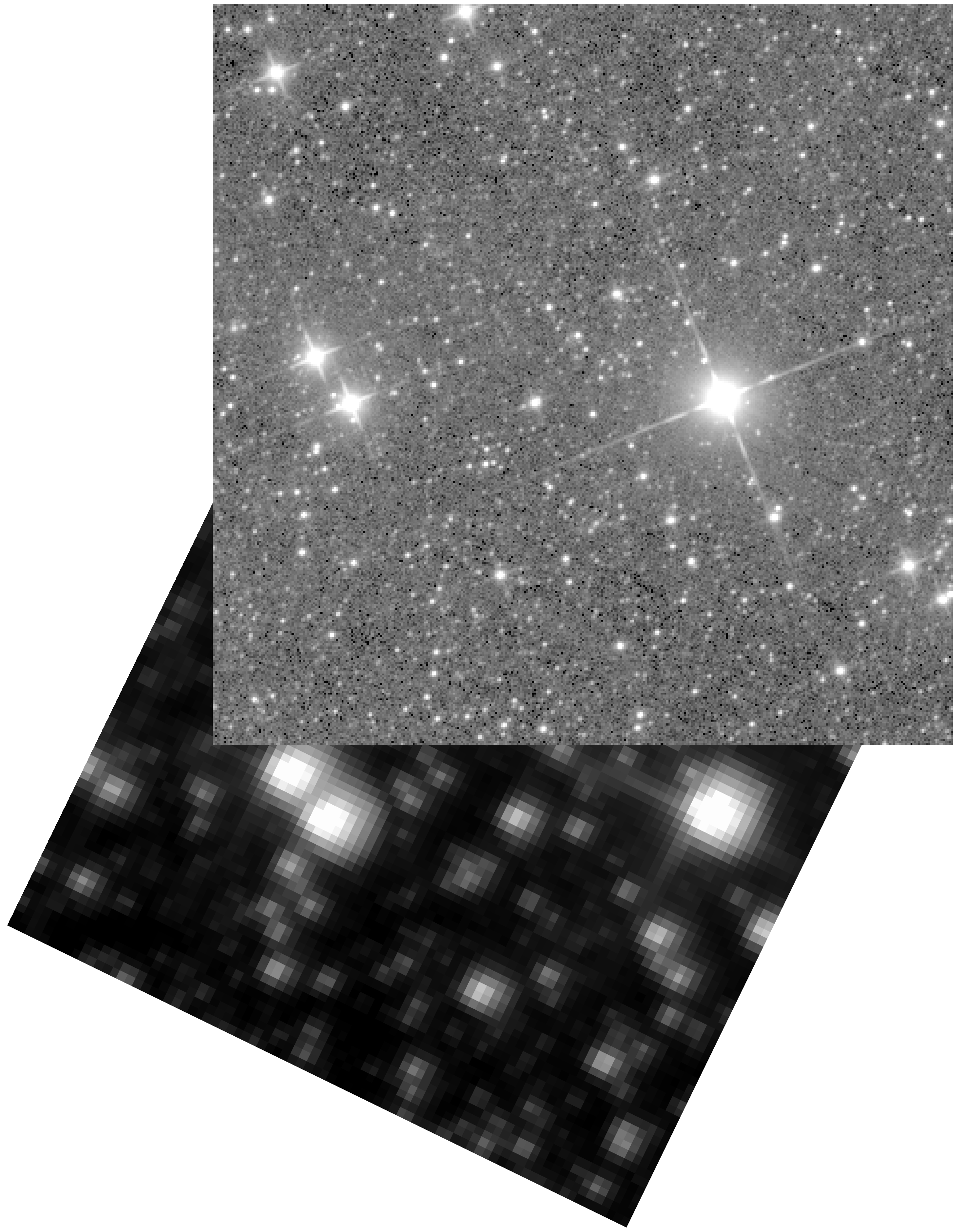
VERA C. RUBIN OBSERVATORY

@Onoddil @pm.me
.github.io

# “Simple” Cross-Matching

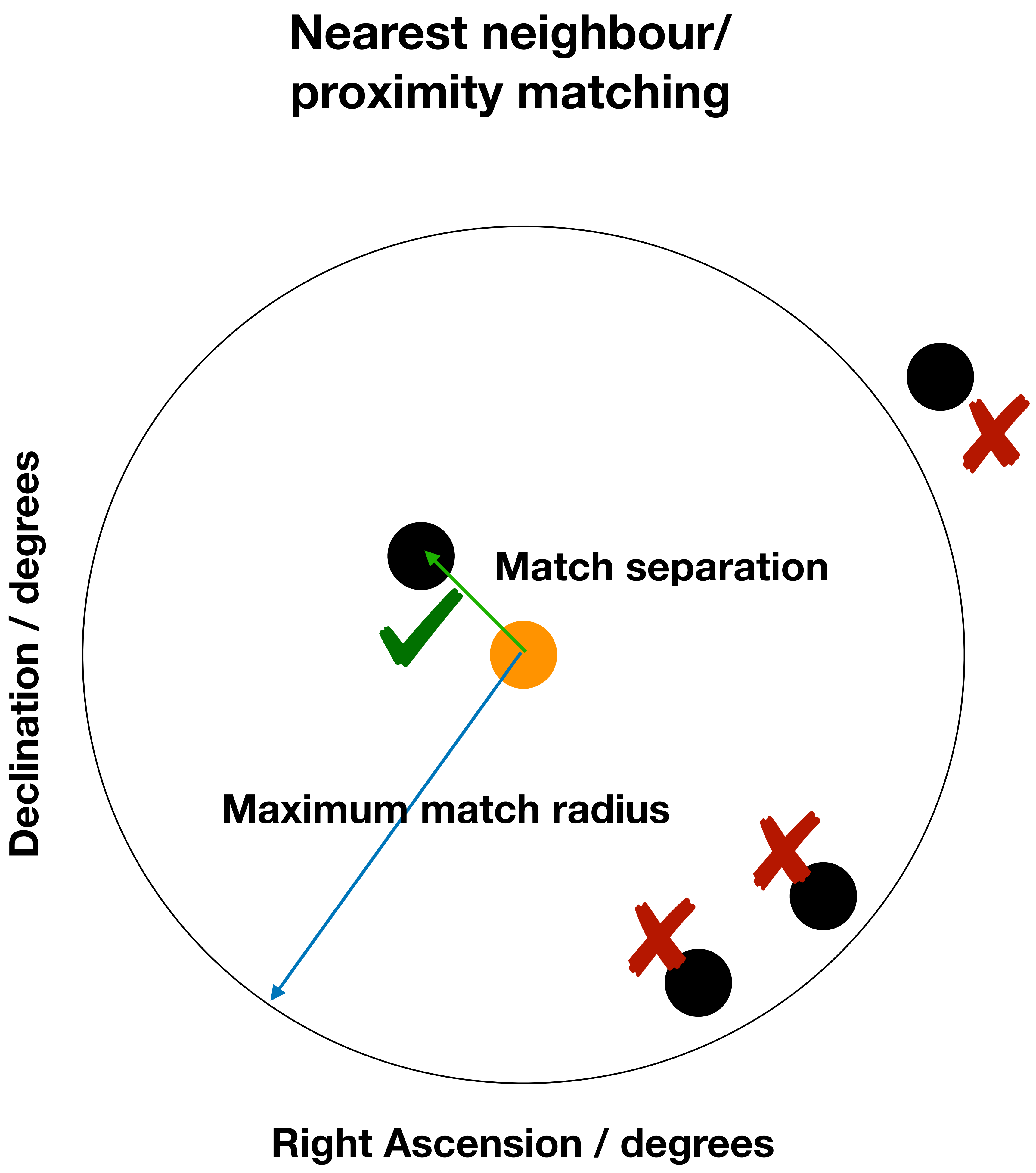


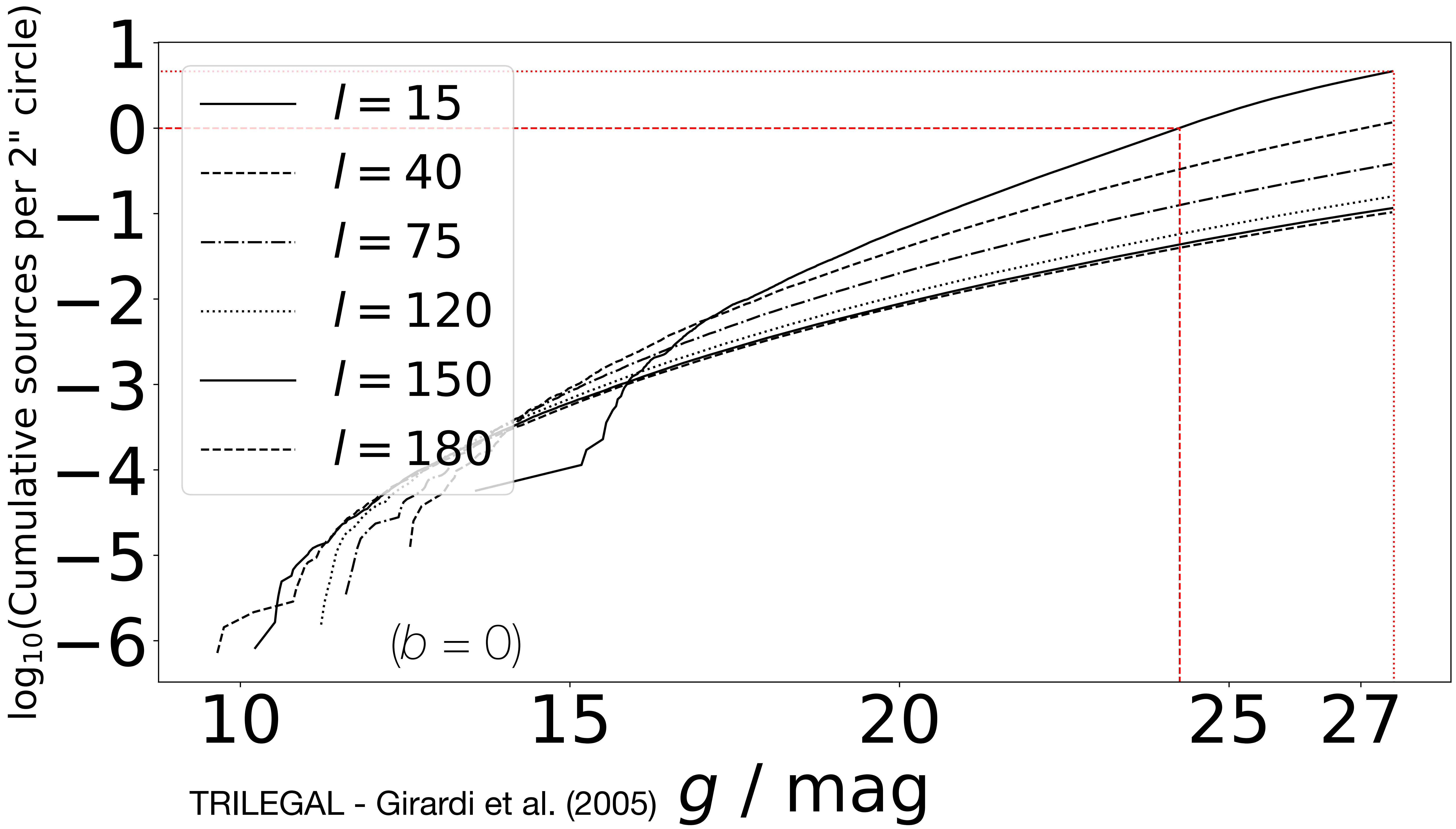


Tom J Wilson @onoddil

# The Problem With LSST

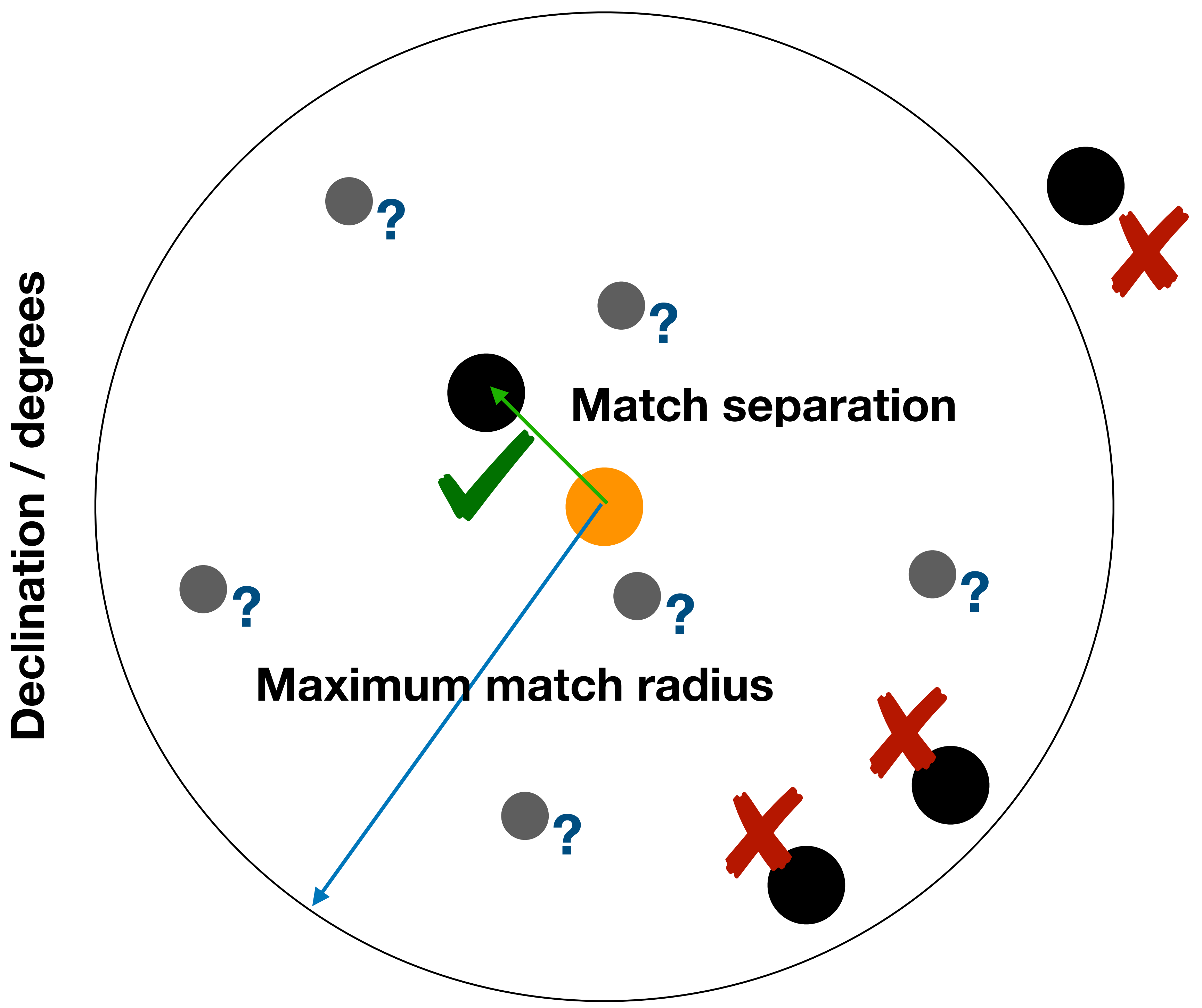


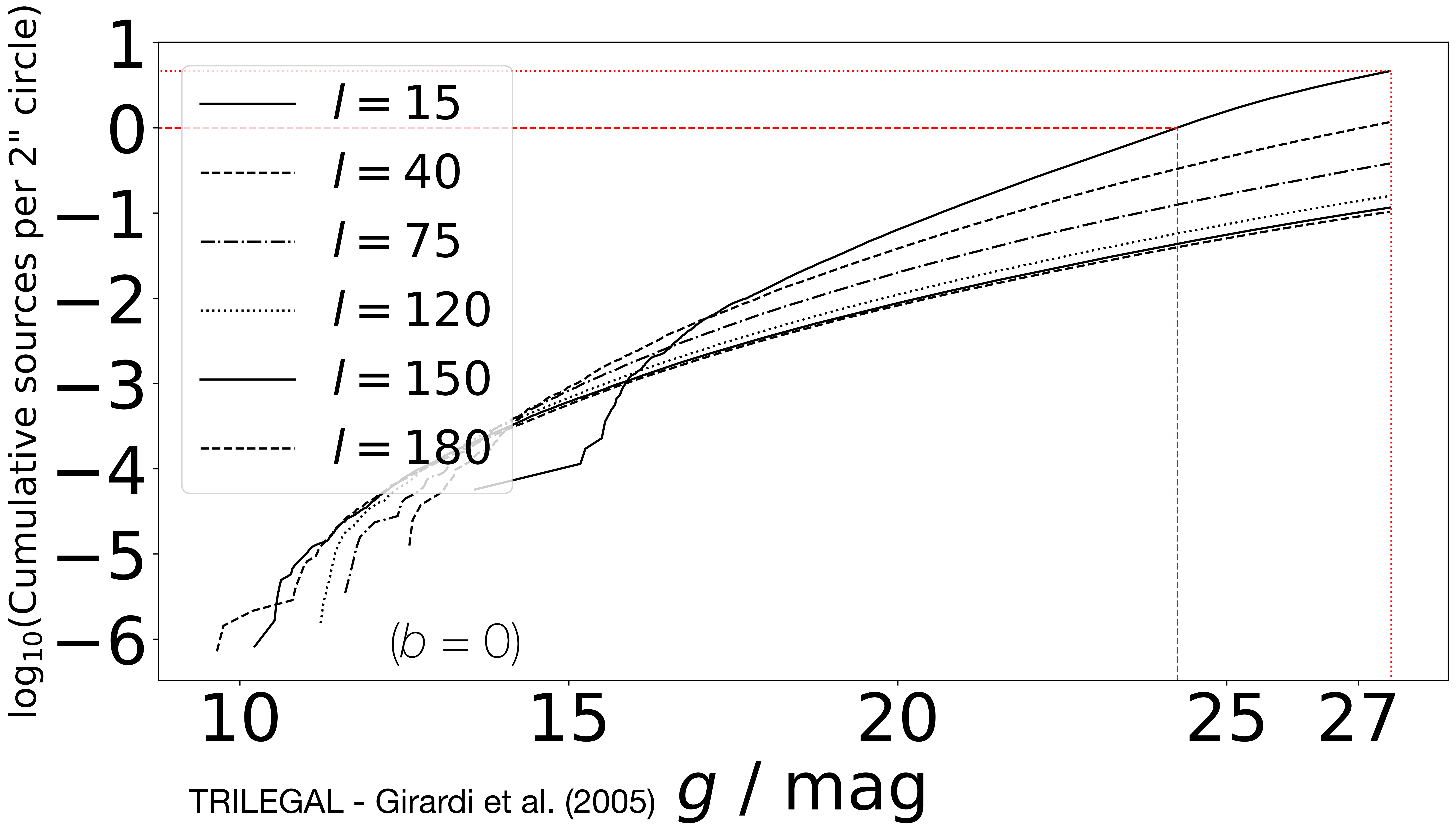

Tom J Wilson @onoddil

# The Problem With LSST

**Nearest neighbour/ proximity matching**



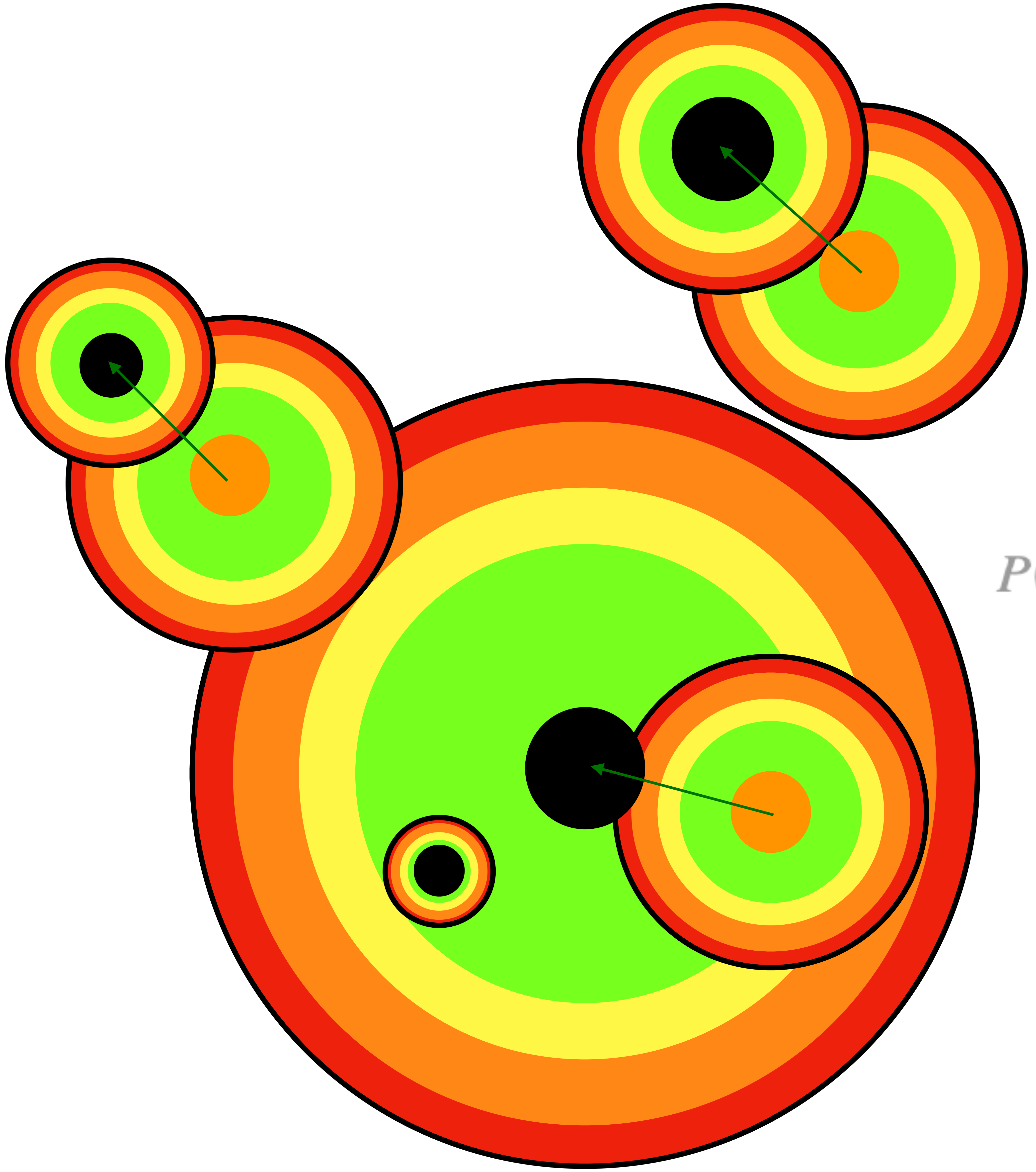

**(It's still a few randomly placed objects in every match radius at high Galactic latitudes)**

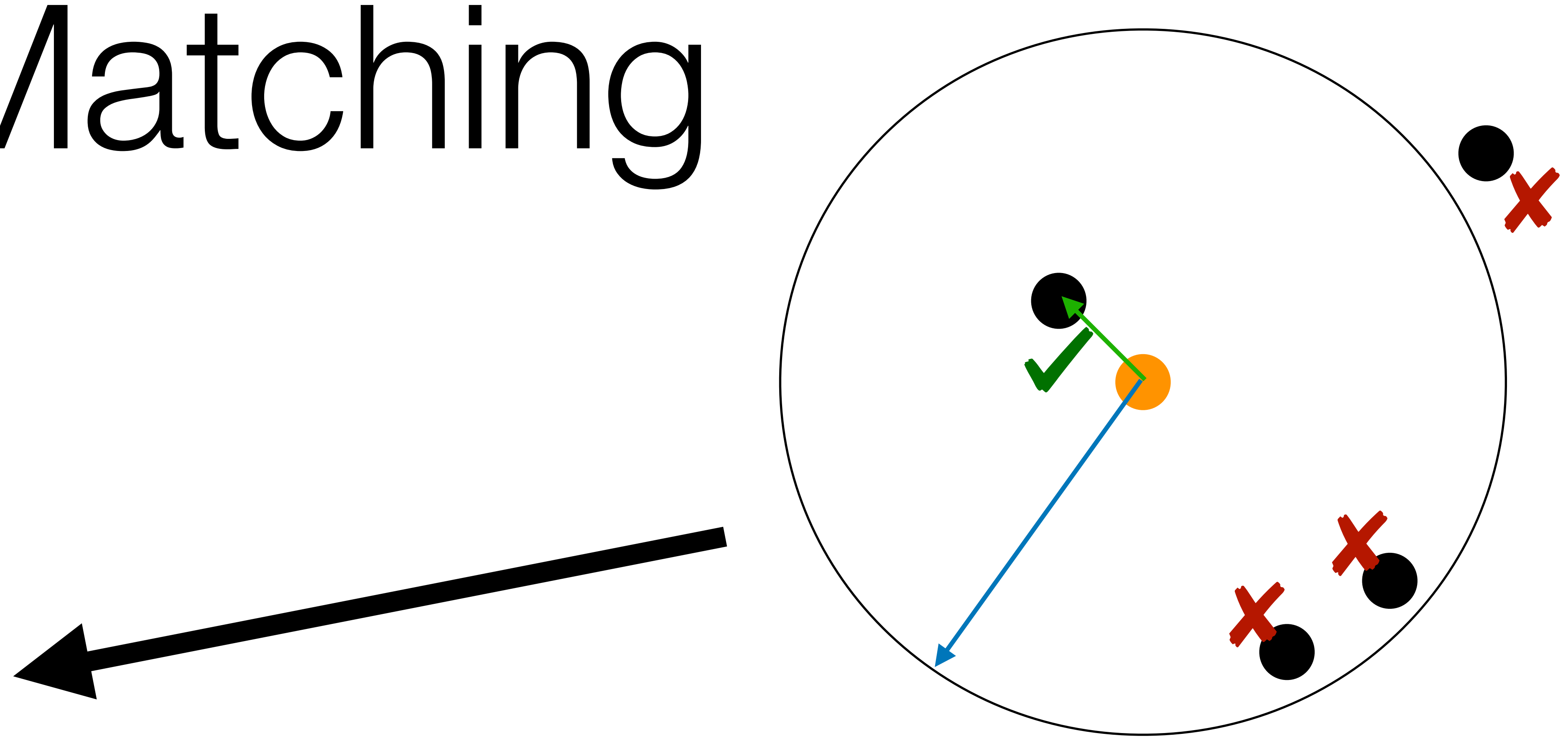



**Nearest-neighbour matching *will not* work in the era of Rubin!**

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")

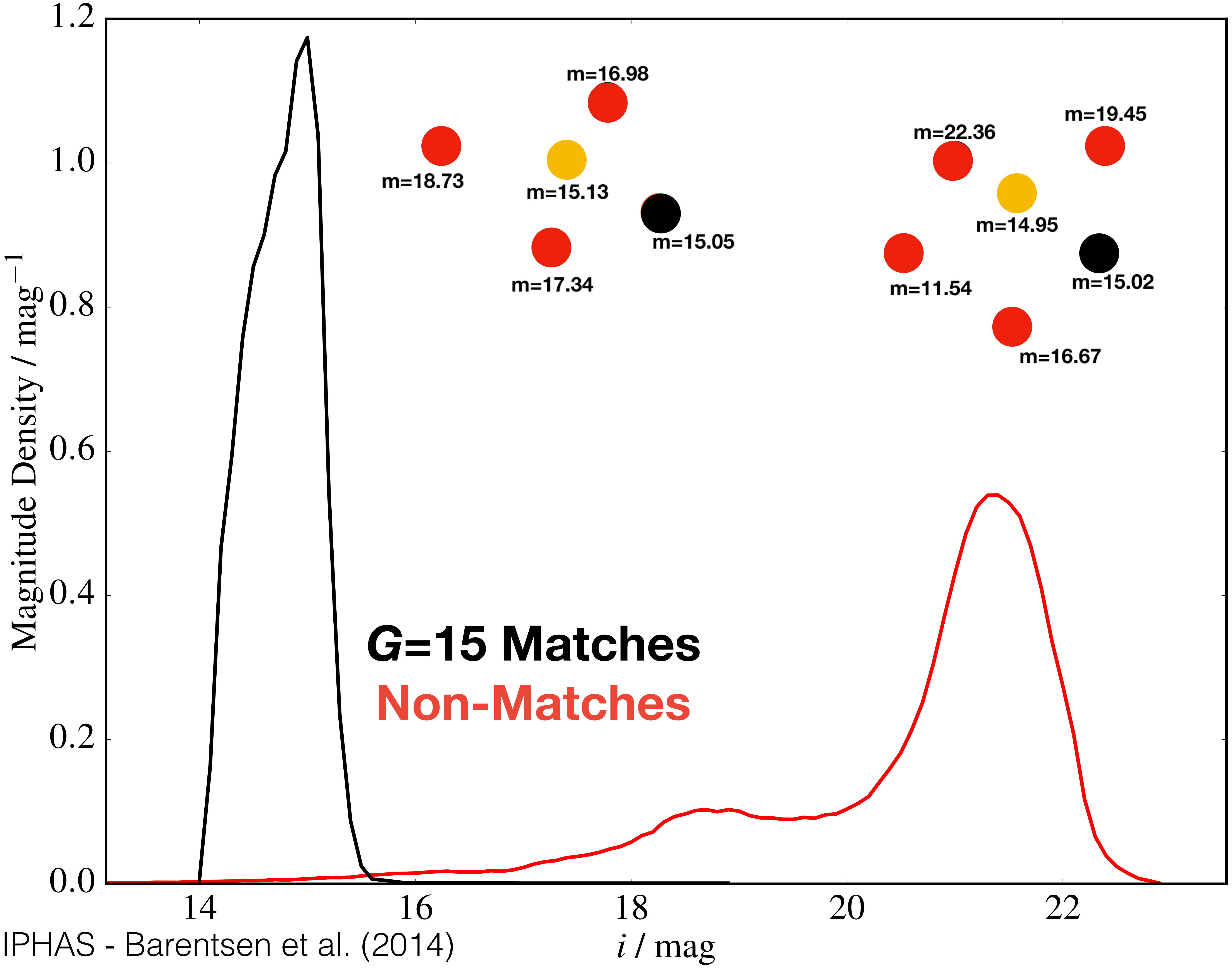
Probability of sources having their <u>brightnesses</u> given they are counterparts

Wilson & Naylor (2018a)

Tom J Wilson @onoddil

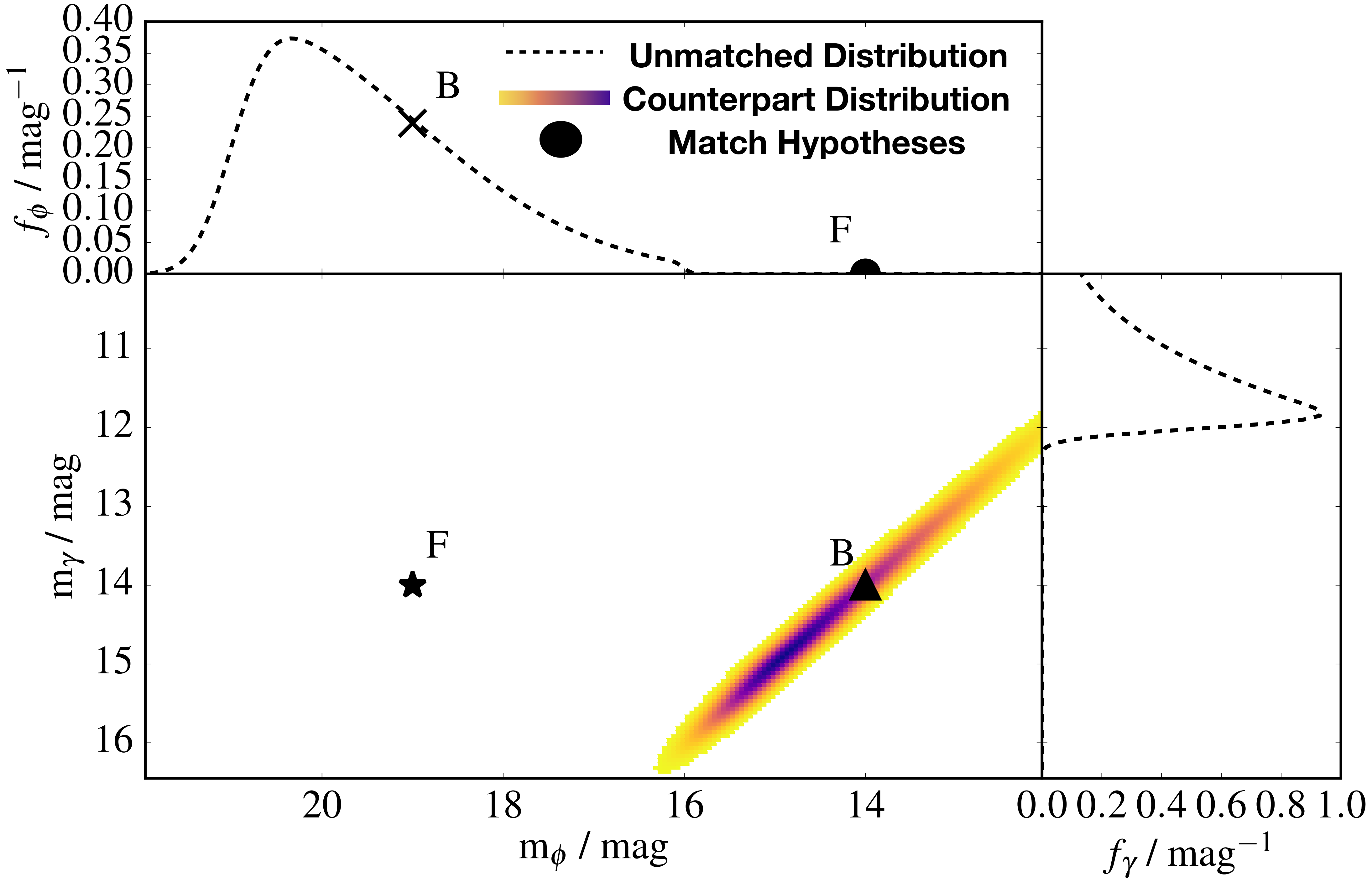
# Photometry: Rejecting False Positives

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




IPHAS - Barentsen et al. (2014)

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

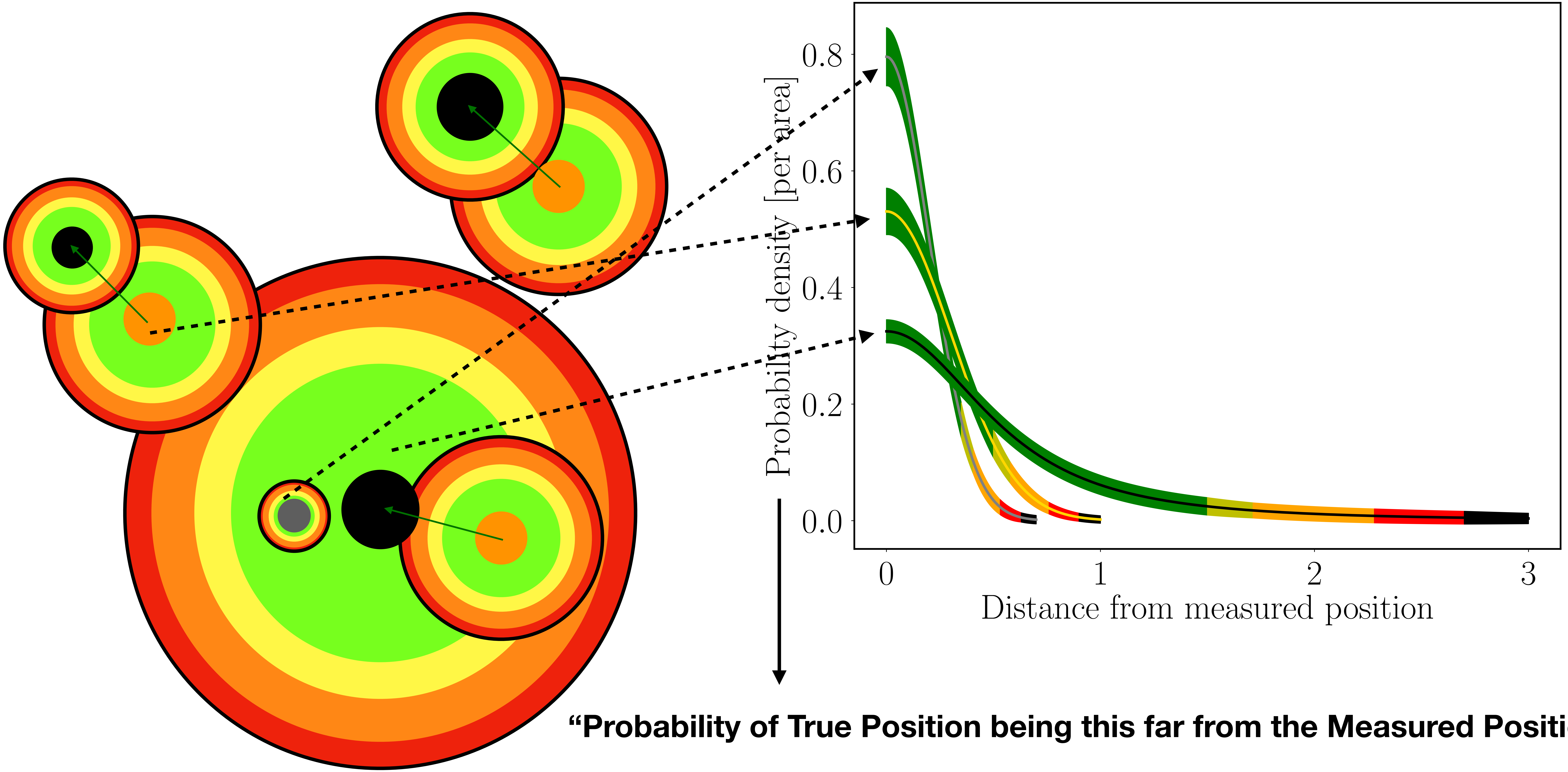



Wilson & Naylor (2018a)

**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

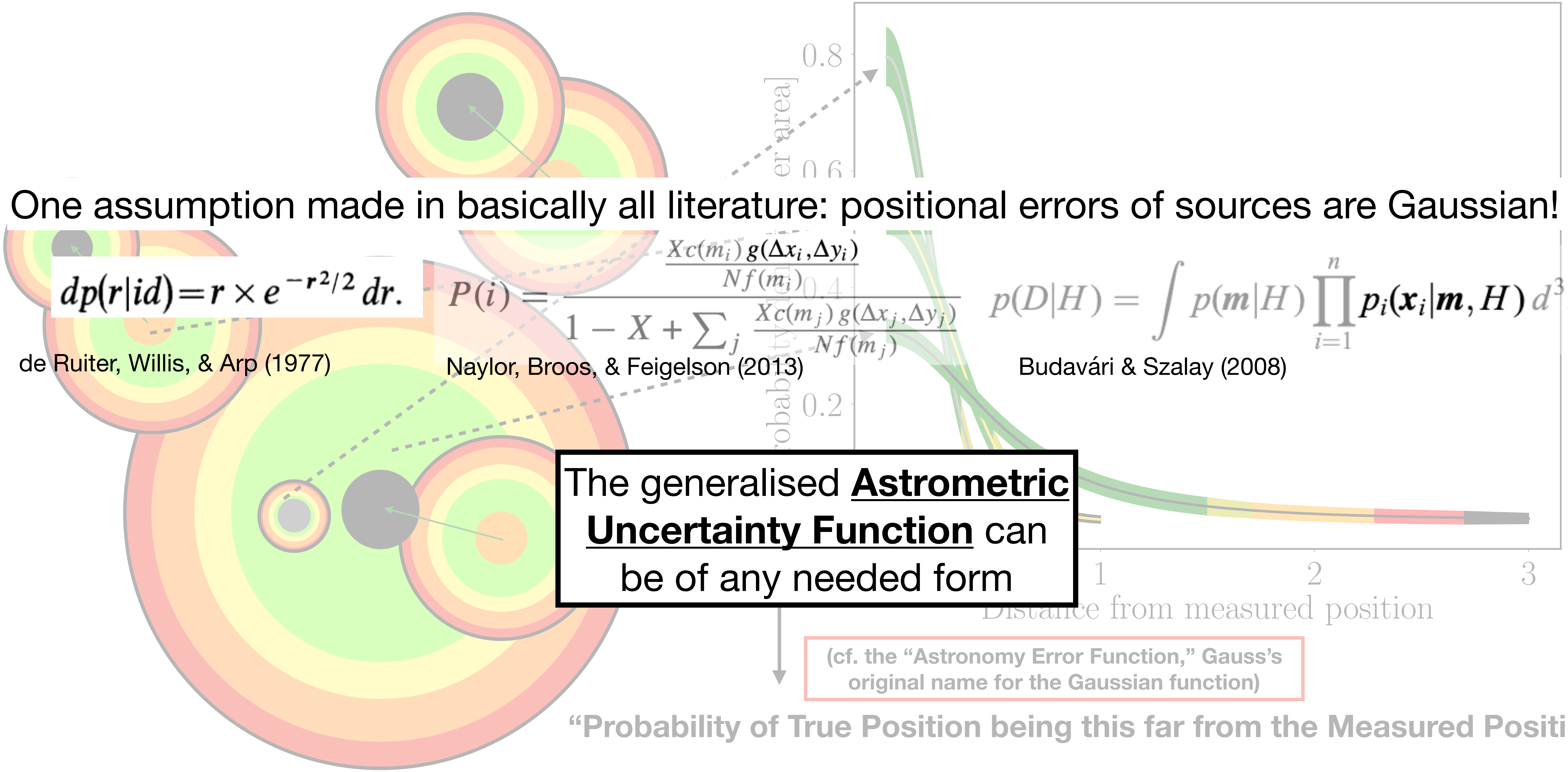
# The "Astrometric Uncertainty Function"




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

Tom J Wilson @onoddil

# The "Astrometric Uncertainty 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)

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

Budavári & Szalay (2008)

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

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

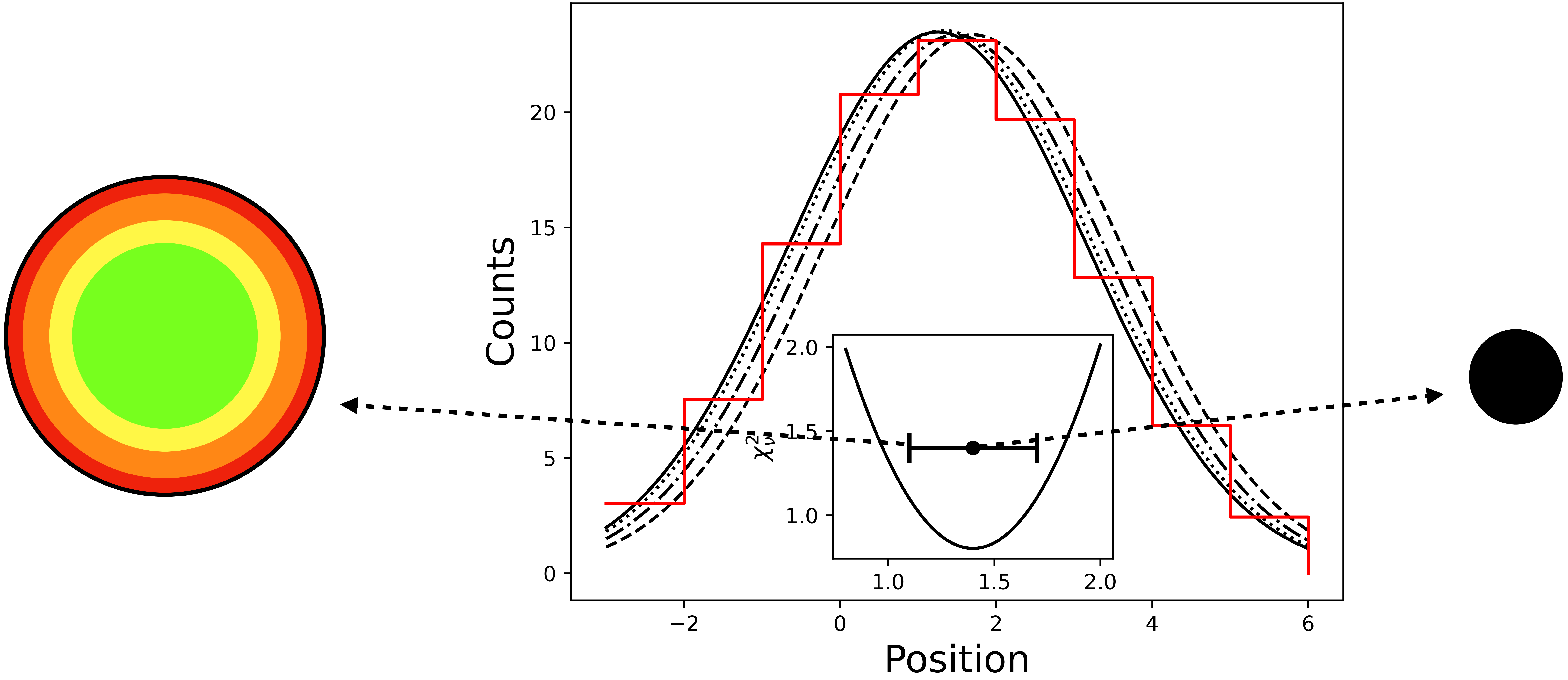
"Probability of True Position being this far from the Measured Position"

Tom J Wilson @onoddil

# The “Astrometric Uncertainty Function”

(cf. the “Astronomy Error Function,” Gauss’s original name for the Gaussian function)

## The *Centroid* Component of the AUF




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)

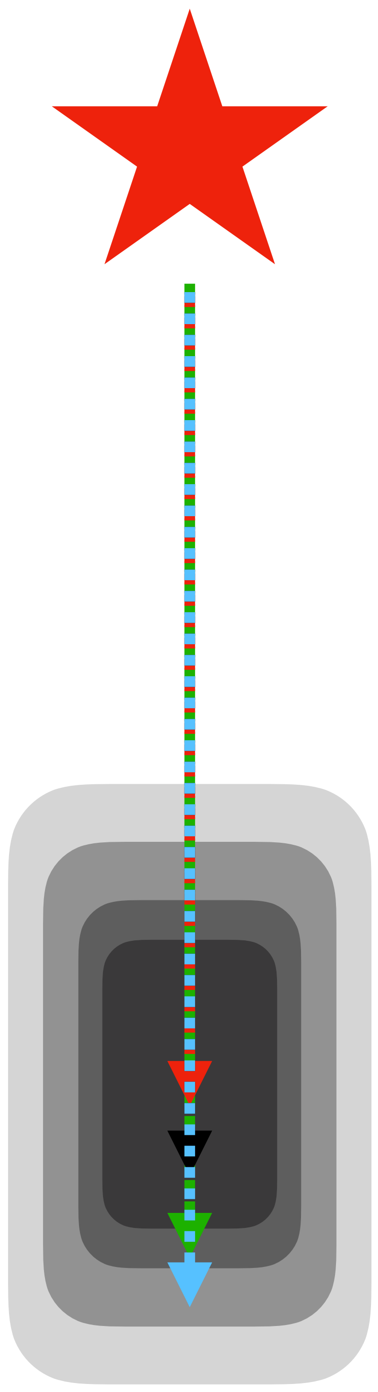
$$p(D|H) = \int p(\boldsymbol{m}|H) \prod_{i=1}^{n} p_i(\boldsymbol{x}_i|\boldsymbol{m}, H)\, d^3m$$

Budavári & Szalay (2008)

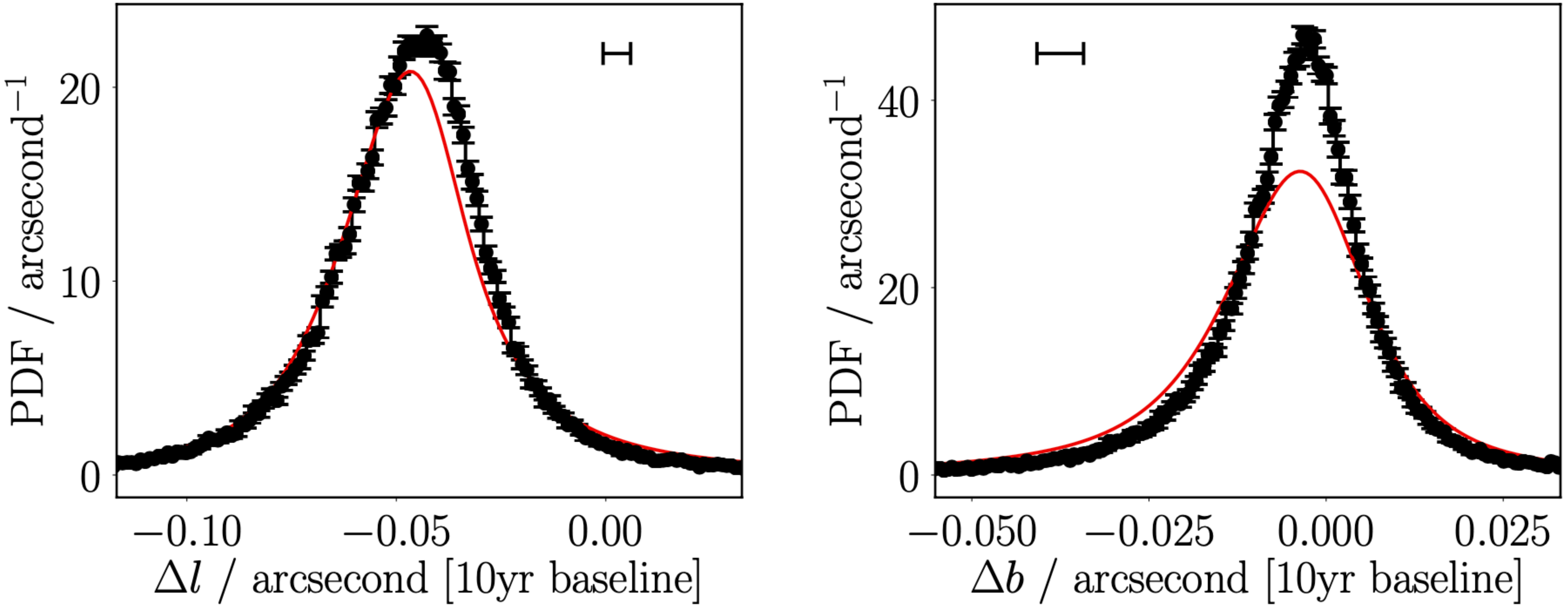
Tom J Wilson @onoddil

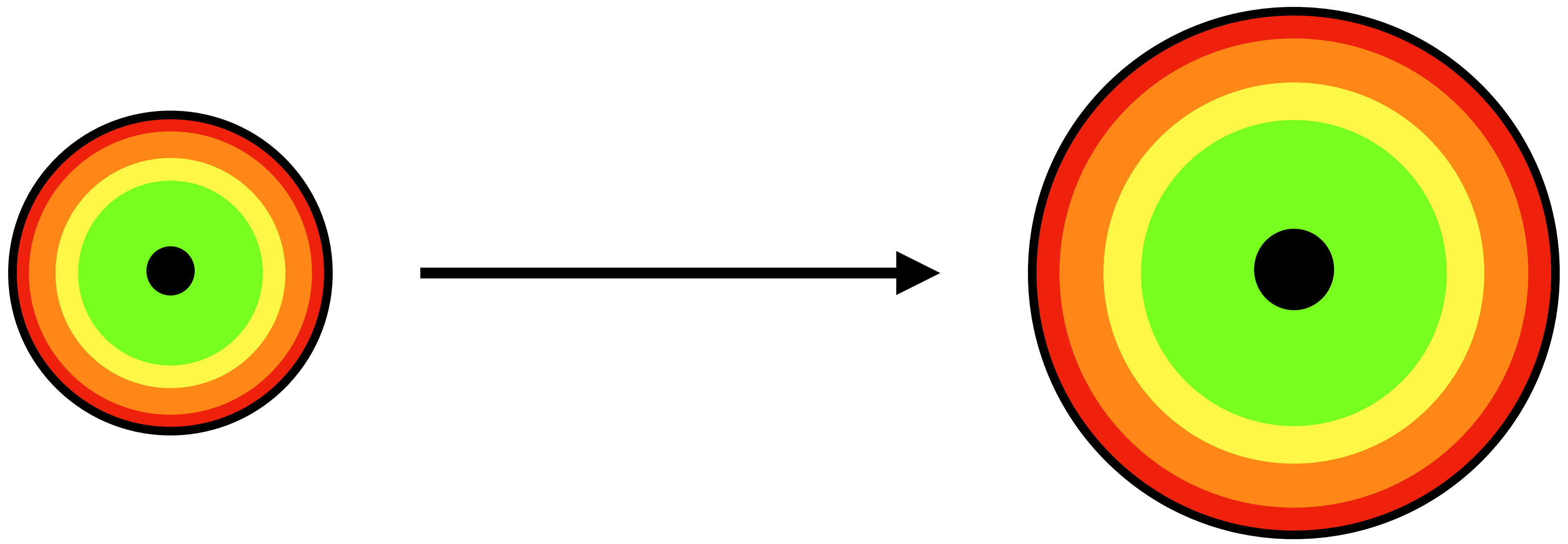
# AUF Components: Unknown Proper Motions

**Object in 2015**



**Projected to 2025**

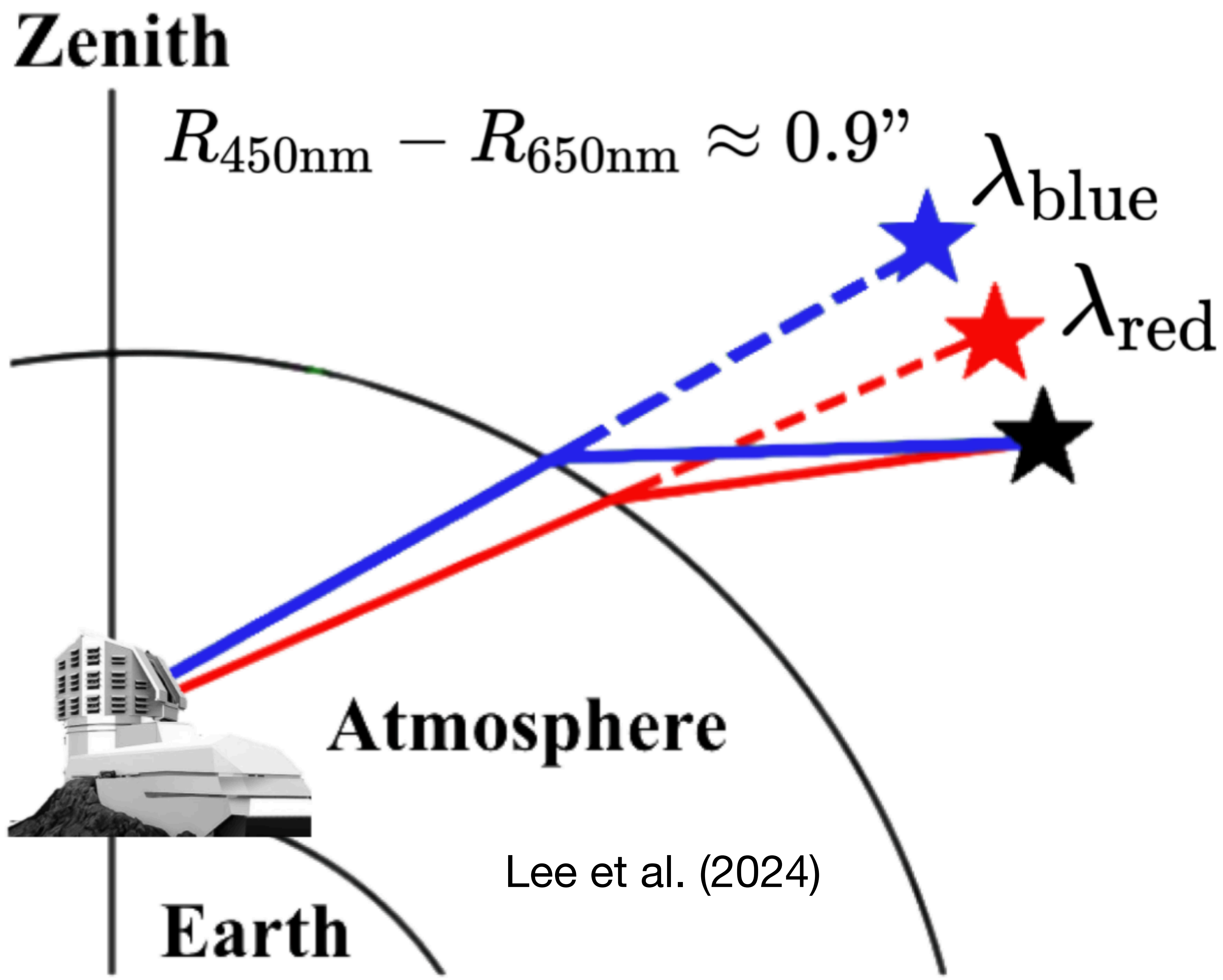


(This also applies to *uncertain* proper motions, where we can incorporate the covariance matrix of weakly-constrained proper motions, e.g. just above the single-visit LSST limit)

Wilson (2023, RASTI, 2, 1)
*Gaia* eDR3 - Gaia Collaboration et al. (2021, A&A, 649, A1)

# AUF Components: DCR



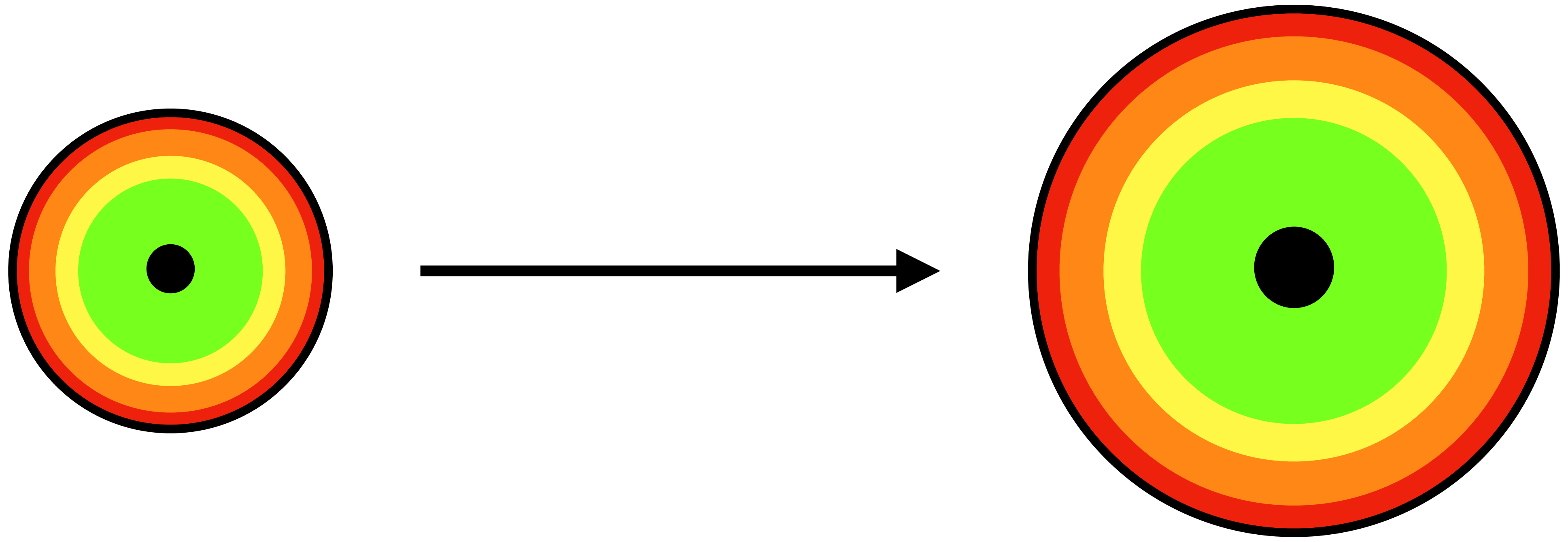

Lee et al. (2024)

e.g. gbdes, Bernstein et al. (2017)

$$\Delta\mathbf{x}^{w} = K_b c \tan z\, \hat{\mathbf{p}}$$

Unknown/uncertain per-band (*b*) scaling factor

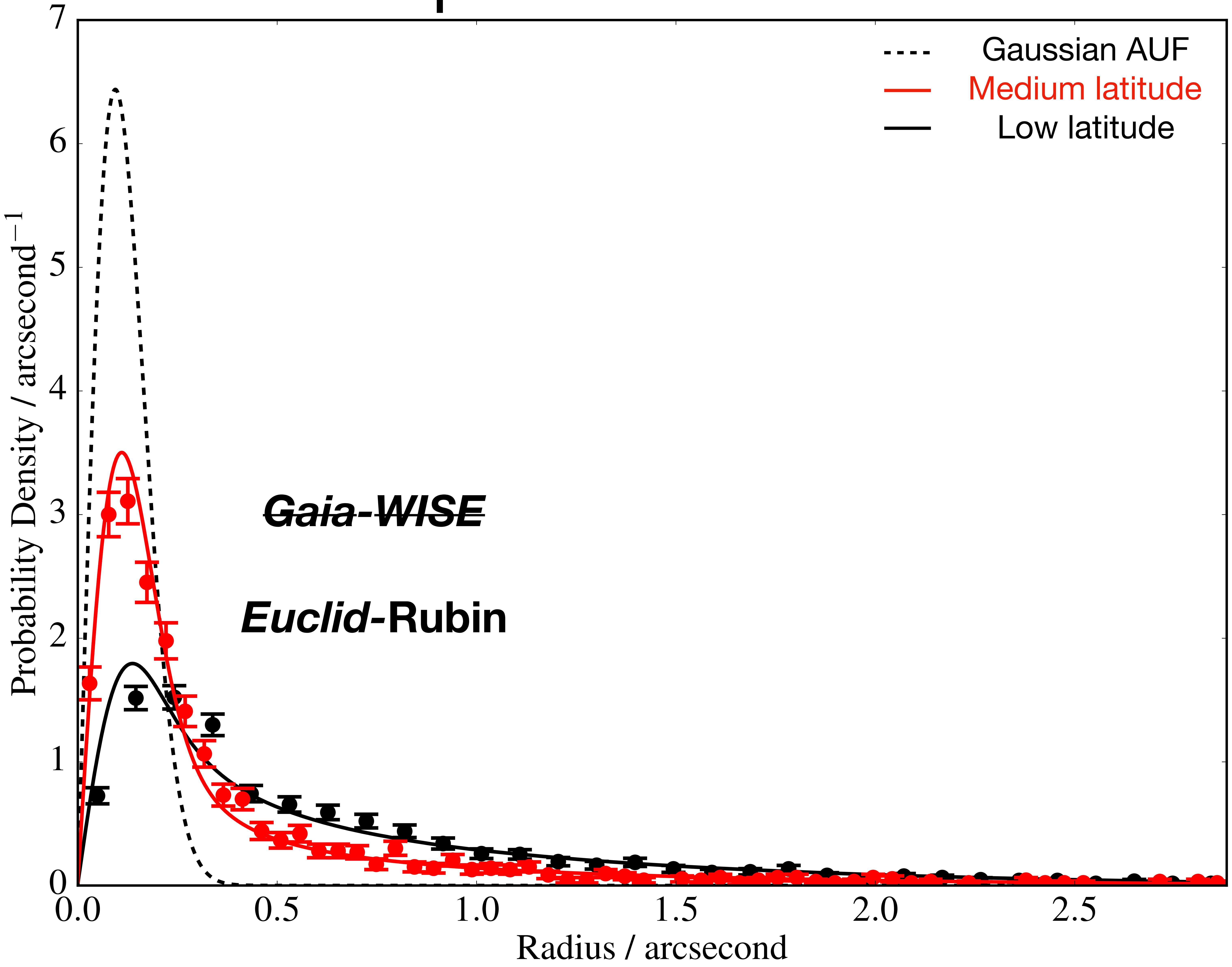
Unknown/uncertain photometric colour *c*



Tom J Wilson @onoddil

# AUF Components: Perturbation

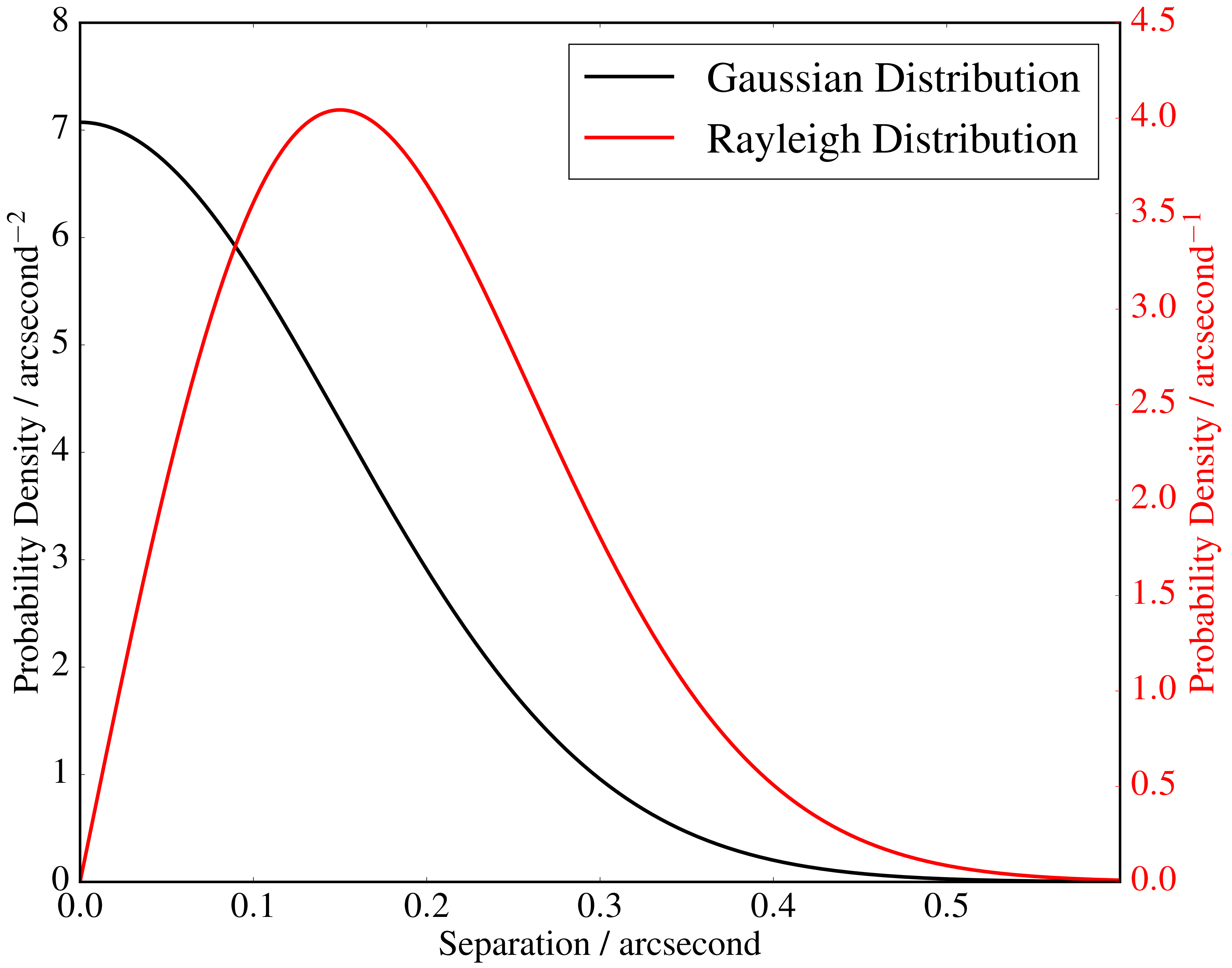
$$P(\zeta, \lambda, k|\gamma, \phi) = \frac{1}{K} \times \prod_{\delta \notin \zeta \cap \delta \in \gamma} N_\gamma f_\gamma^\delta \prod_{\omega \notin \lambda \cap \omega \in \phi} N_\phi f_\phi^\omega \prod_{i=1}^{k} N_c \boldsymbol{G}_{\gamma\phi}^{\zeta_i \lambda_i} c_{\gamma\phi}^{\zeta_i \lambda_i}$$




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

↓

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

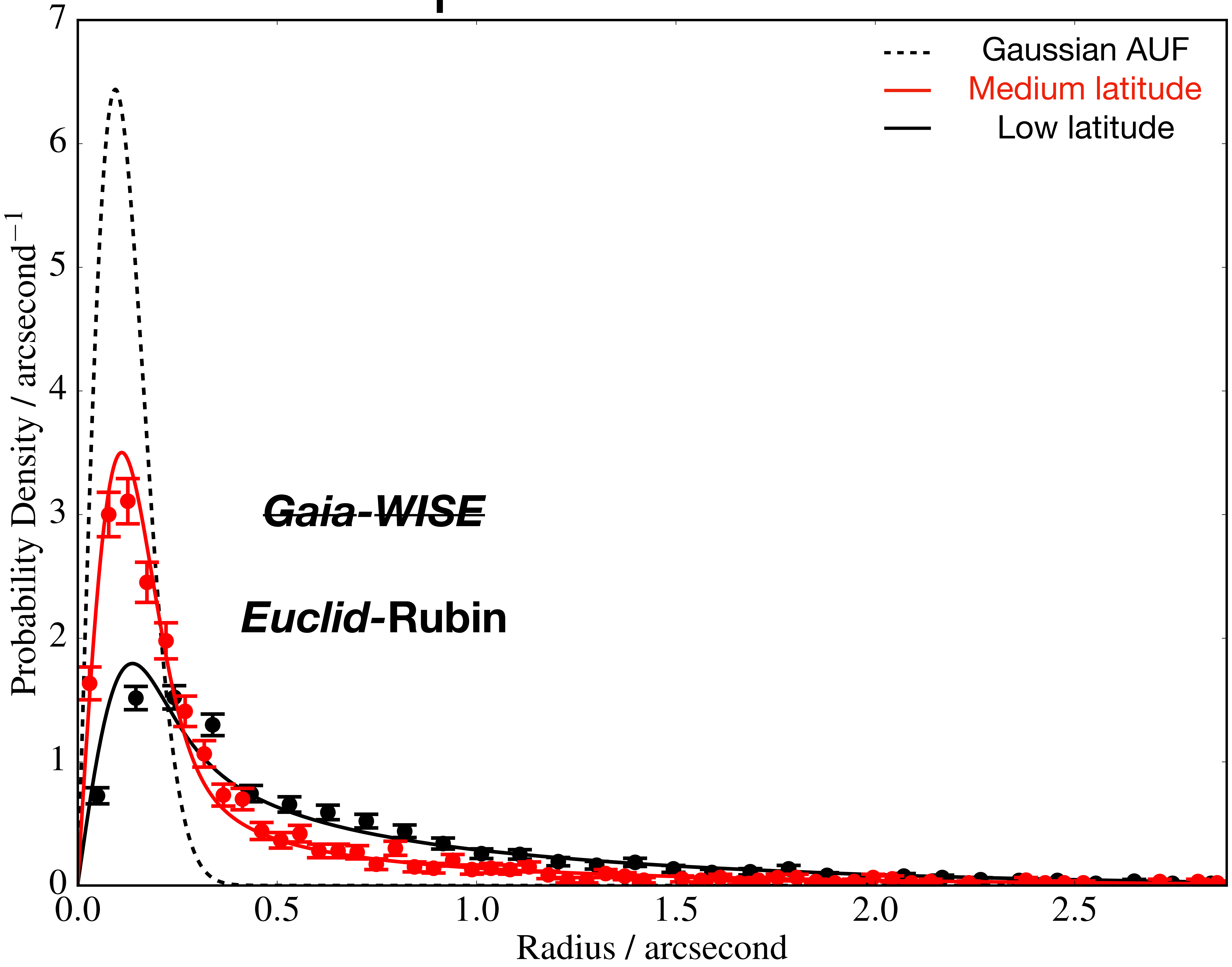


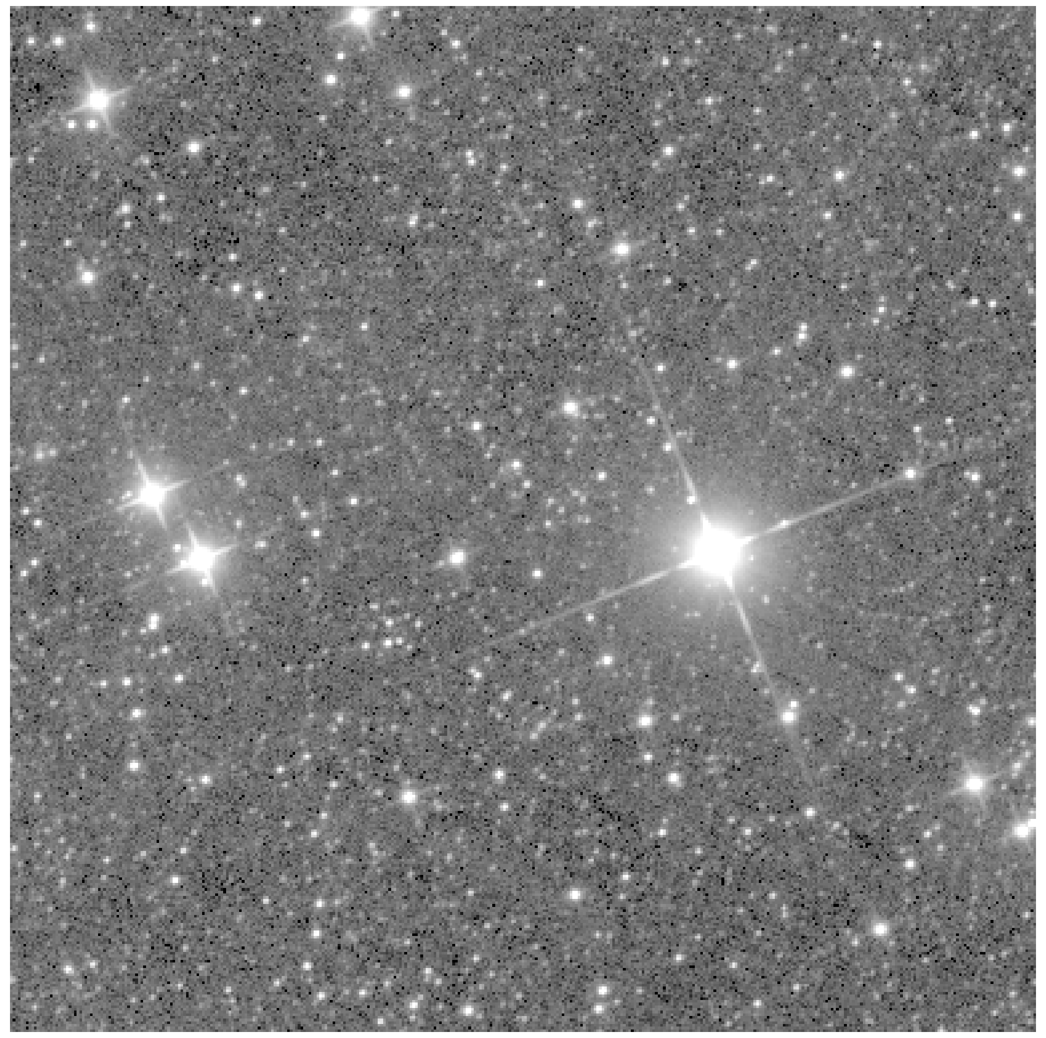

*WISE* - Wright et al. (2010)
*Gaia* DR2 - Gaia Collaboration, Brown A. G. A., et al. (2018)

Tom J Wilson @onoddil

# AUF Components: Perturbation

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

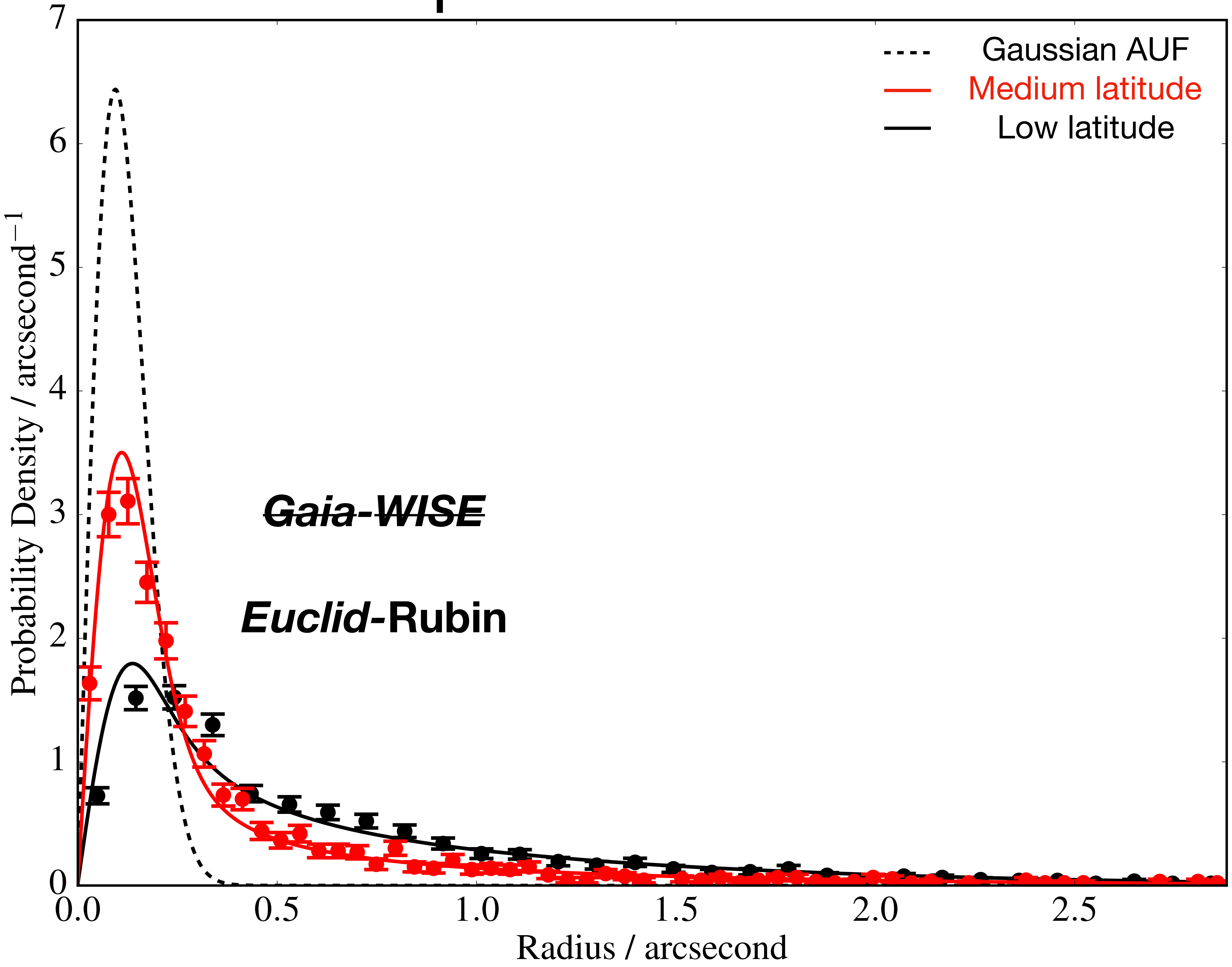


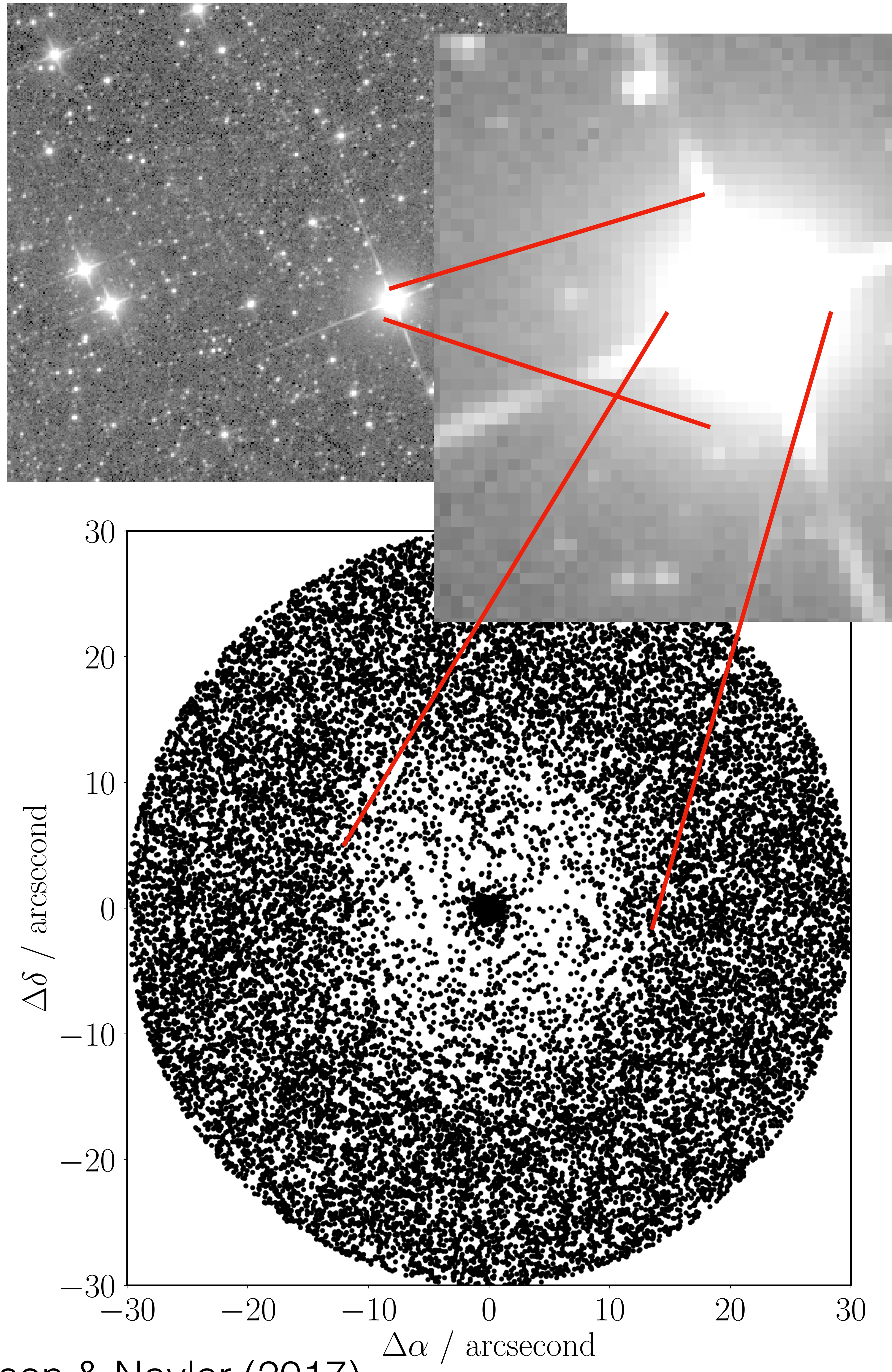



*WISE* - Wright et al. (2010)
*Gaia* DR2 - Gaia Collaboration, Brown A. G. A., et al. (2018)

# AUF Components: Perturbation

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



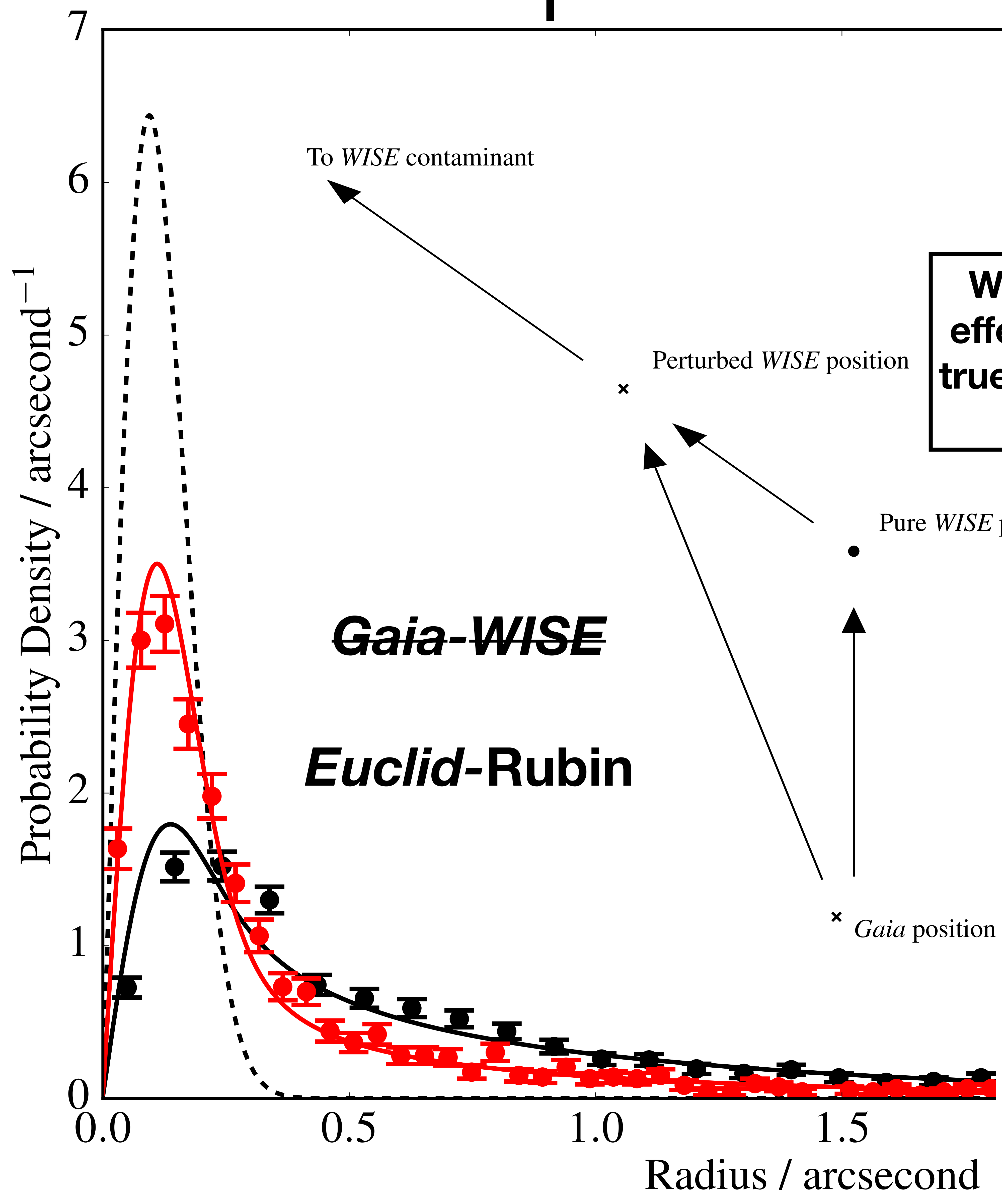

Wilson & Naylor (2017)

*WISE* - Wright et al. (2010)

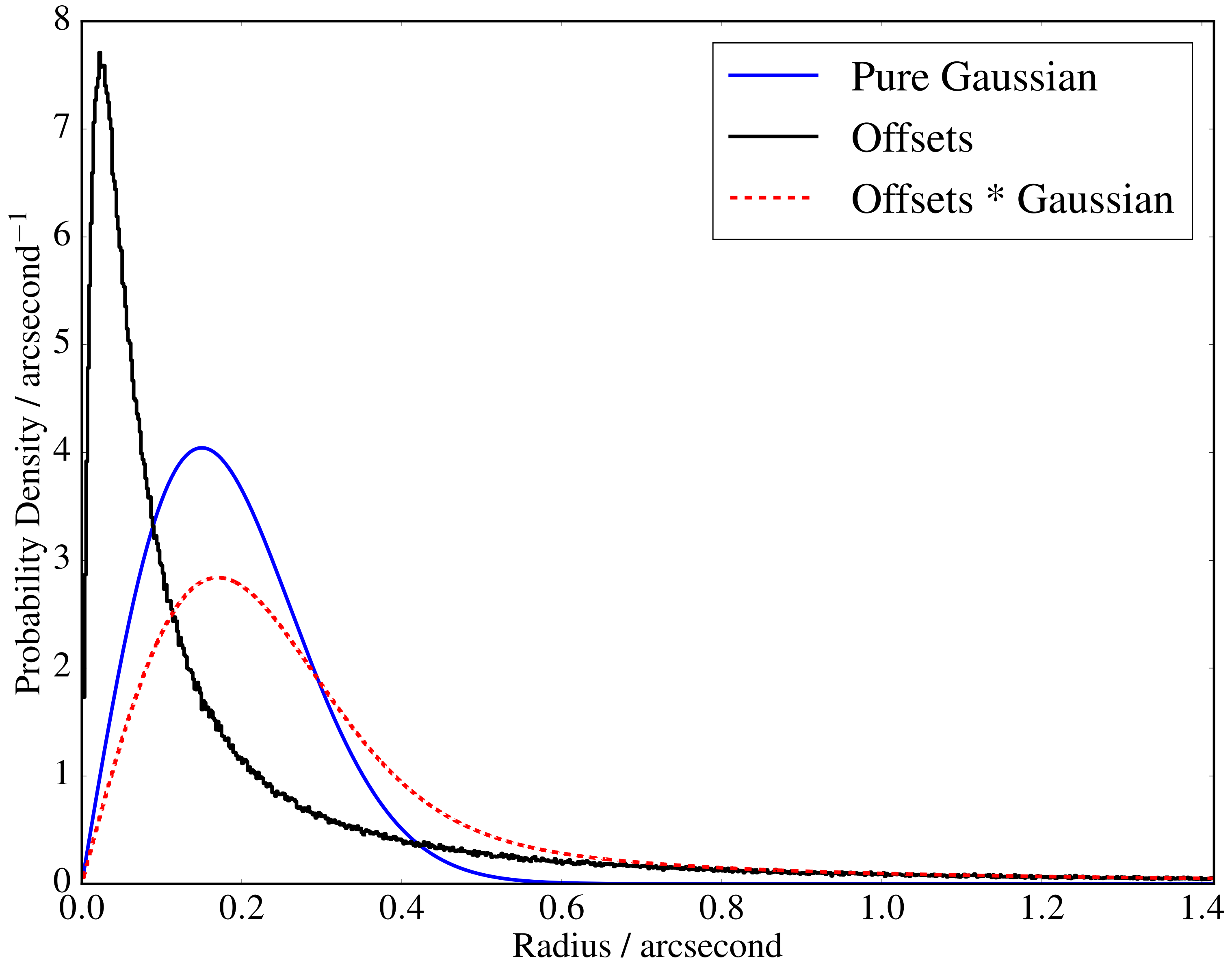
*Gaia* DR2 - Gaia Collaboration, Brown A. G. A., et al. (2018)

# AUF Components: Perturbation

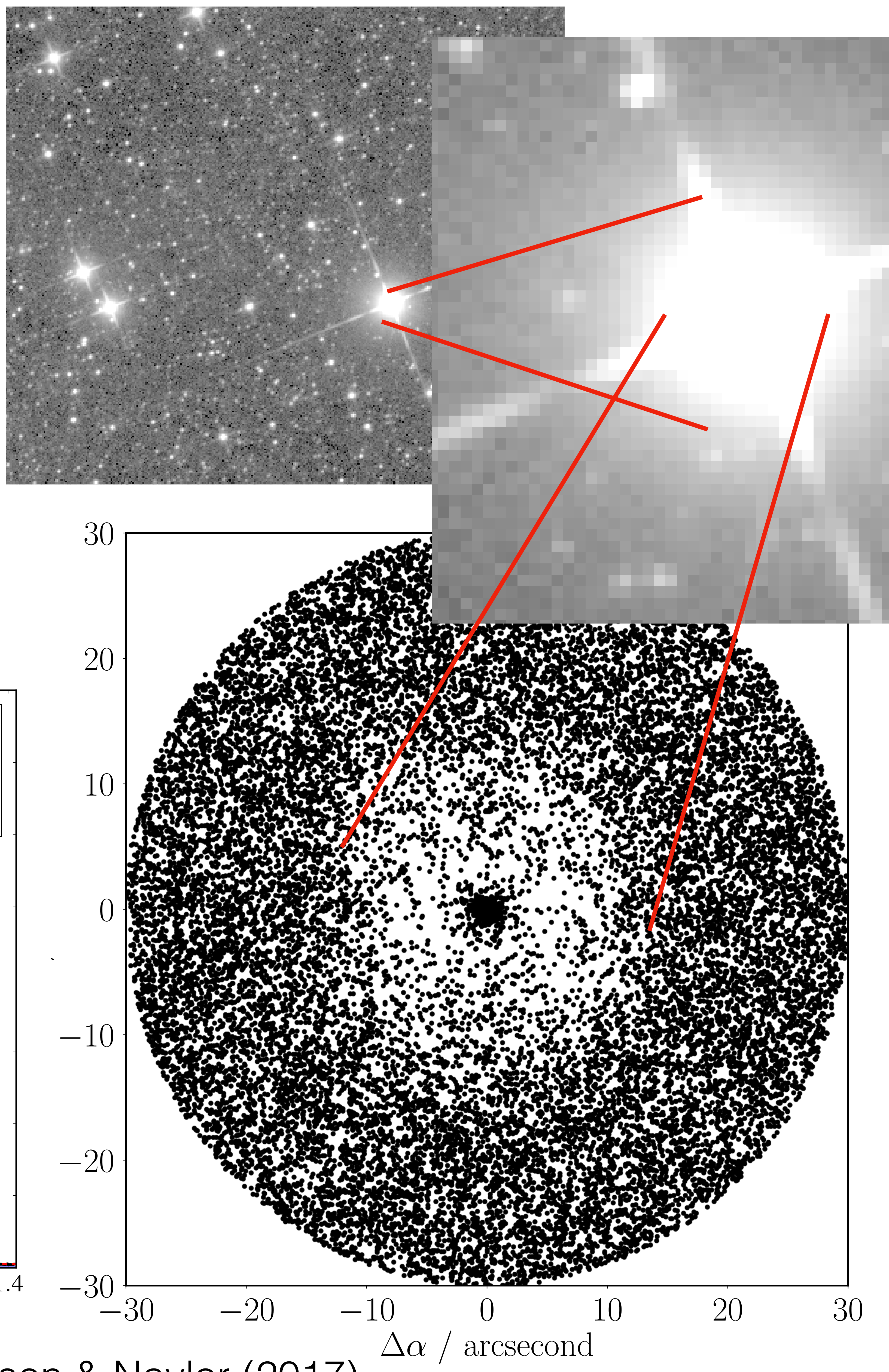
$$P(\zeta, \lambda, k|\gamma, \phi) = \frac{1}{K} \times \prod_{\delta \notin \zeta \cap \delta \in \gamma} N_\gamma f_\gamma^\delta \prod_{\omega \notin \lambda \cap \omega \in \phi} N_\phi f_\phi^\omega \prod_{i=1}^{k} N_c \boldsymbol{G}_{\boldsymbol{\gamma\phi}}^{\boldsymbol{\zeta_i \lambda_i}} c_{\gamma\phi}^{\zeta_i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




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




Wilson & Naylor (2018b)

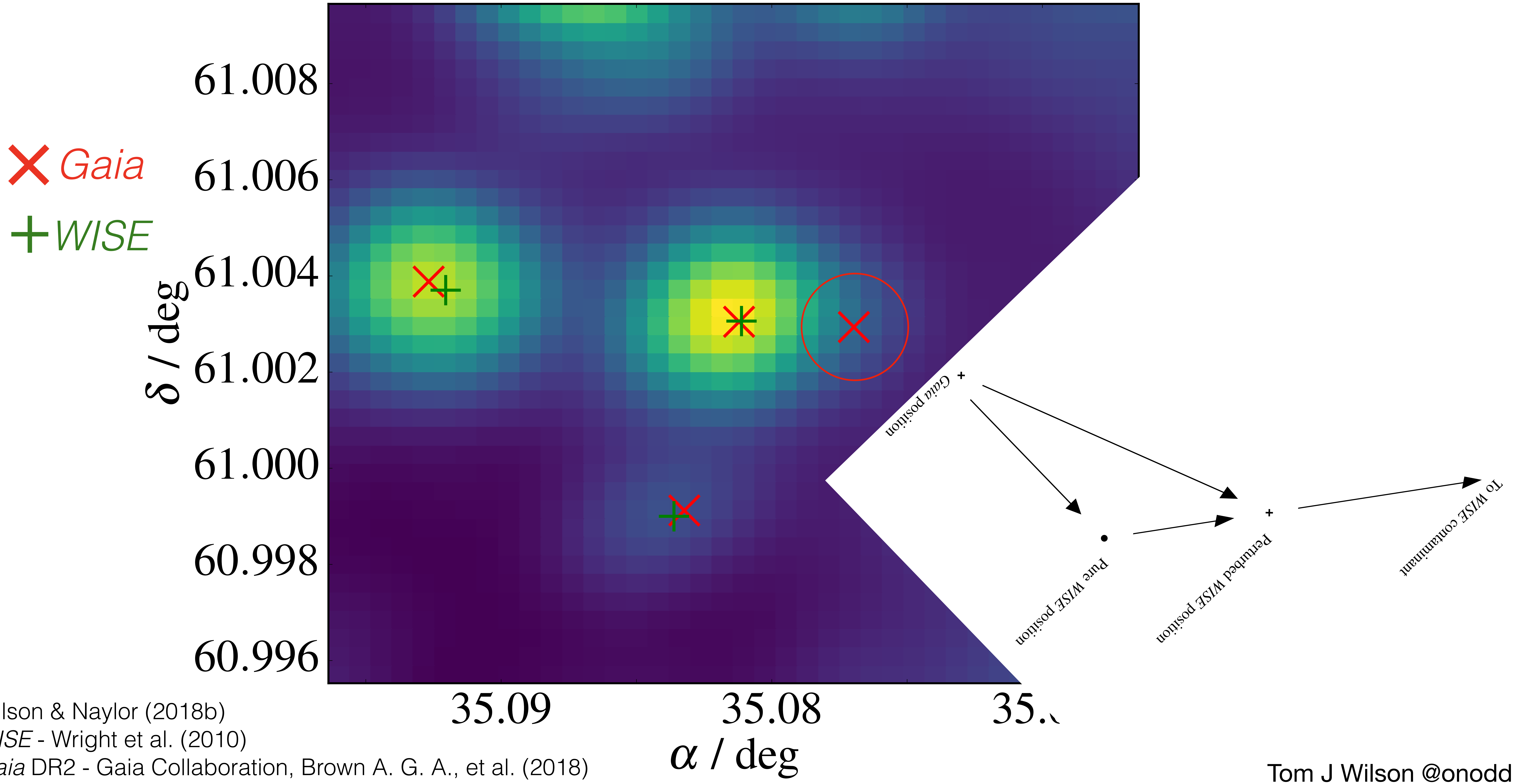



Wilson & Naylor (2017)

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

Tom J Wilson @onoddil

# Unresolved, Hidden Contaminant Objects

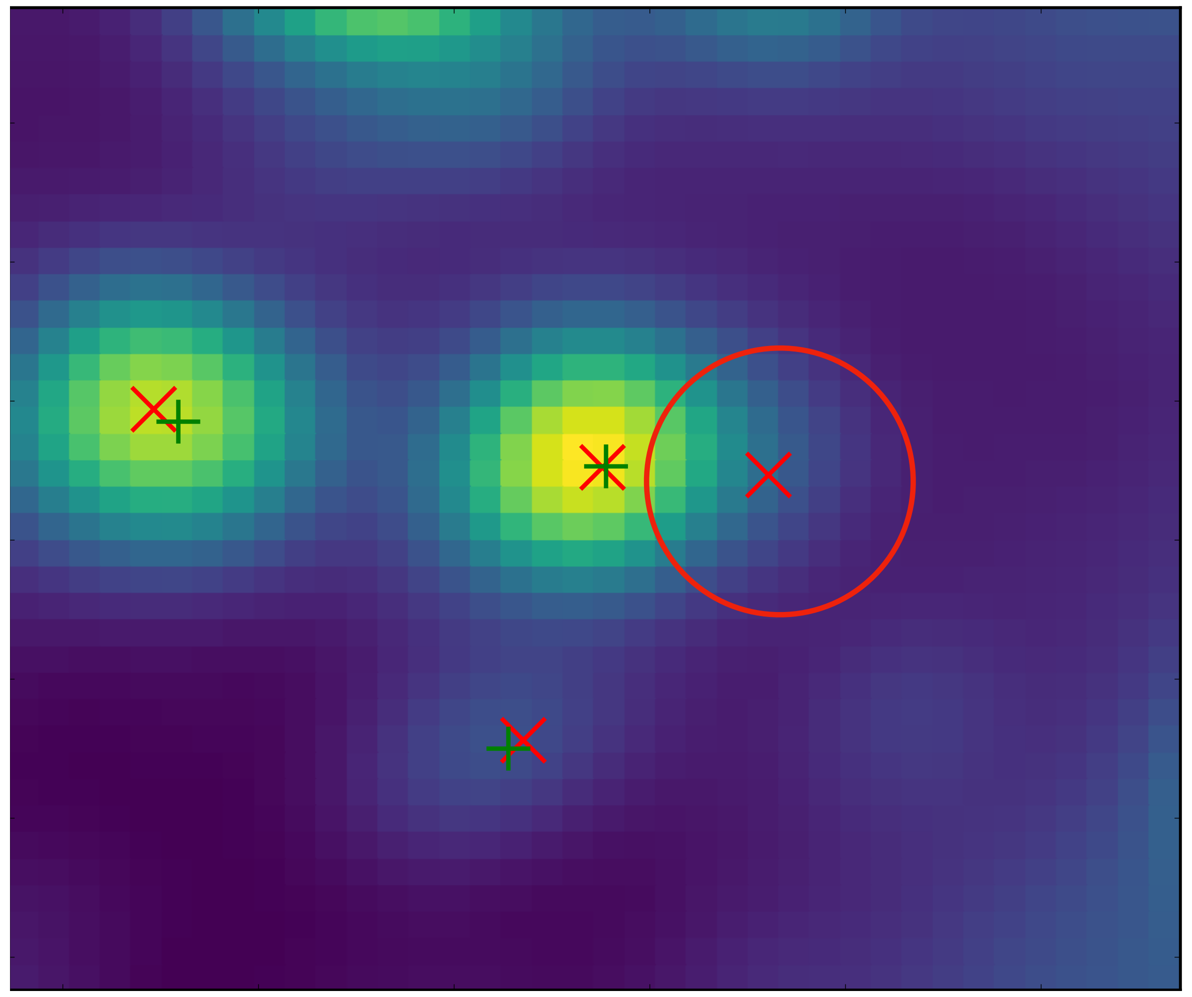


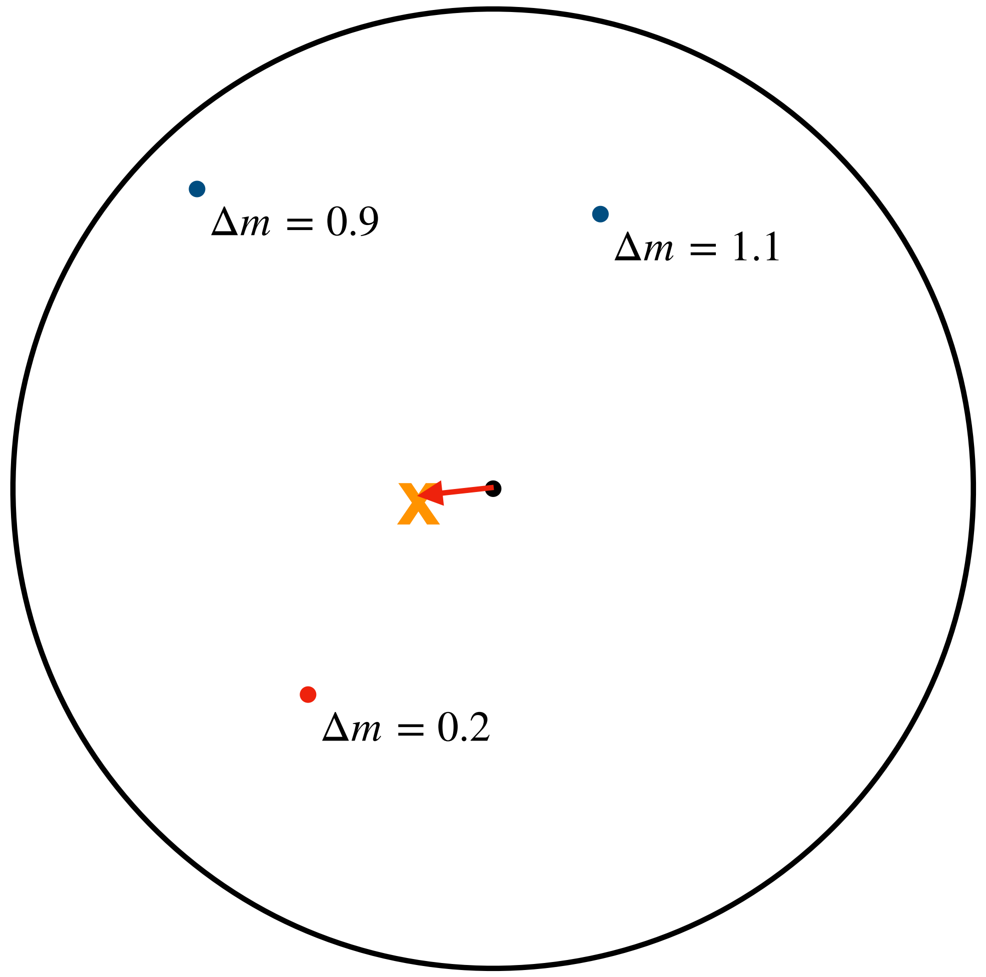

Wilson & Naylor (2018b)
*WISE* - Wright et al. (2010)
*Gaia* DR2 - Gaia Collaboration, Brown A. G. A., et al. (2018)

Tom J Wilson @onoddil

# Perturbation AUF Component: Flux Brightening

**High SNR PSF or Aperture Photometry**



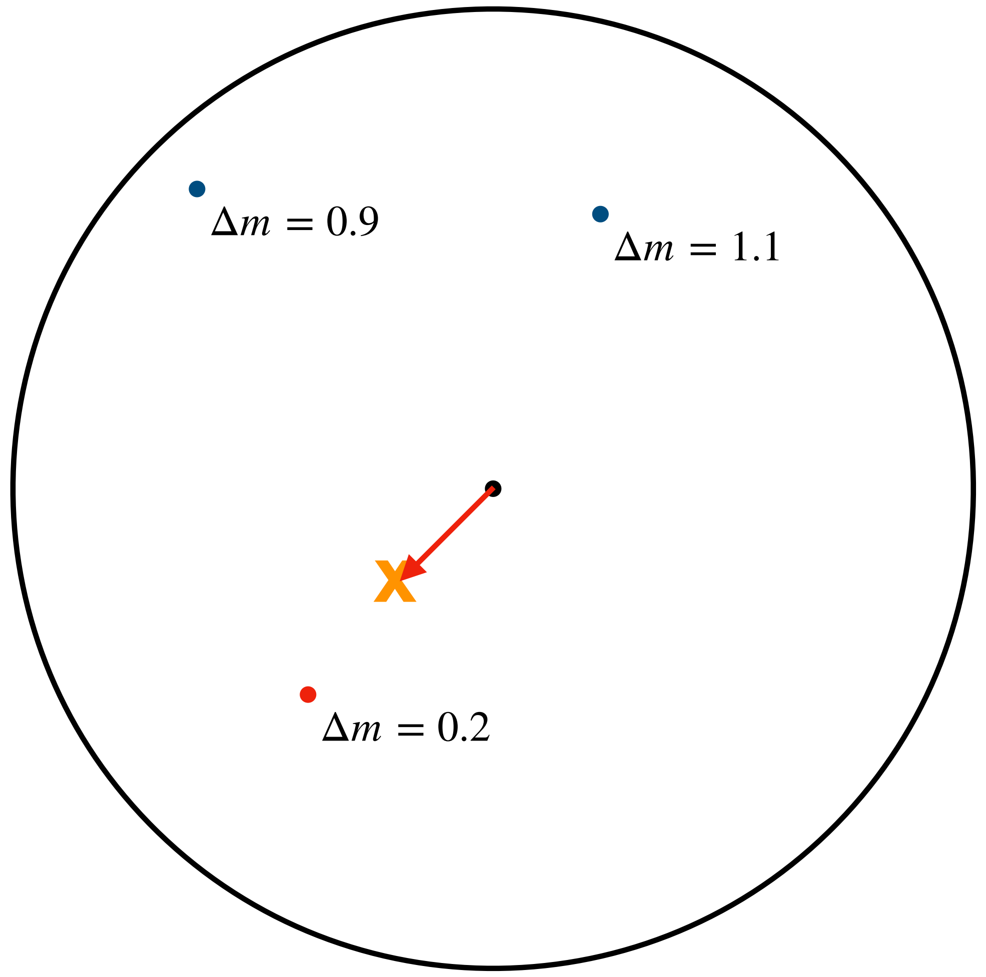


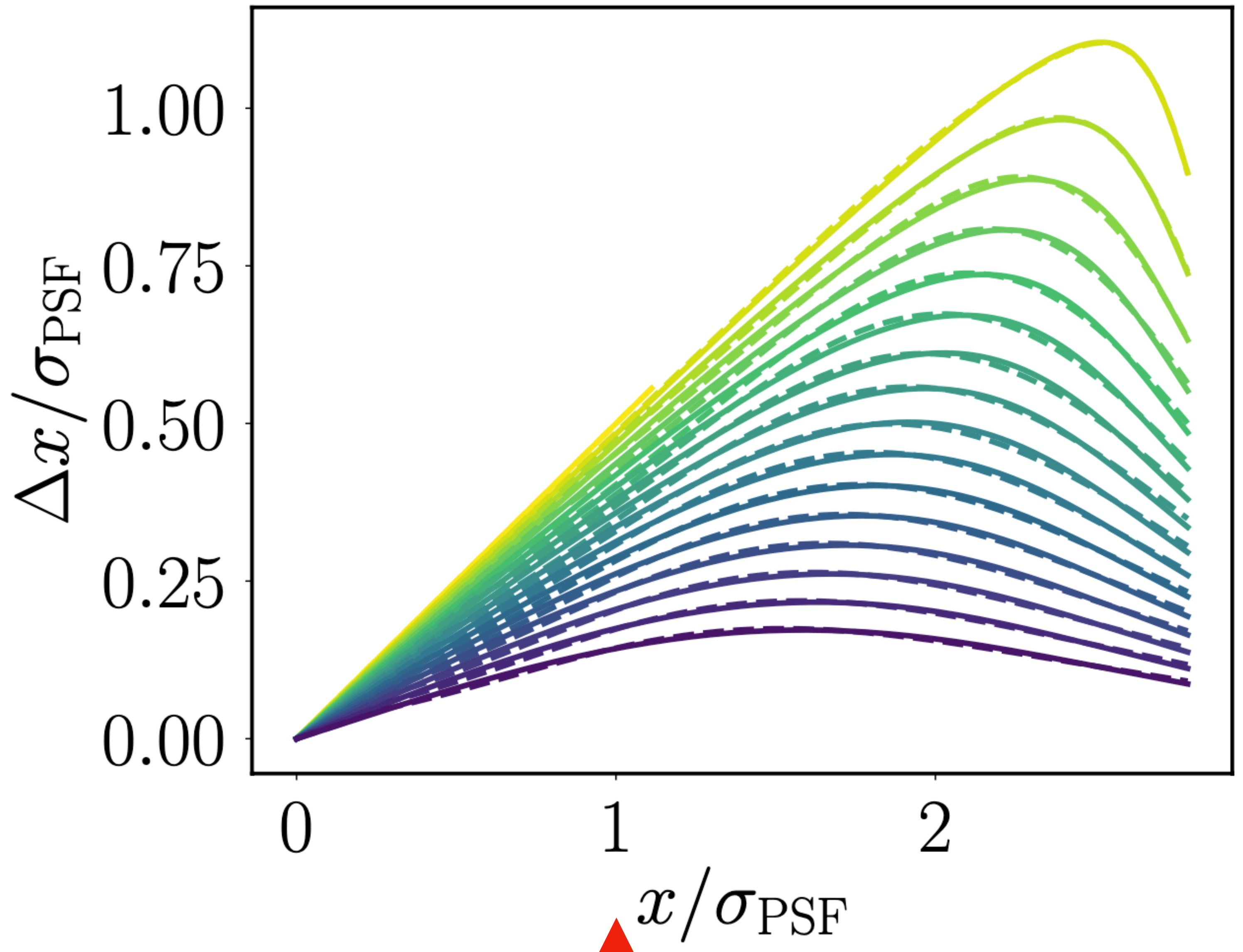

$$\Delta x = \frac{\sum_i f_i x_i}{1 + \sum_i f_i}$$

$$\Delta f = \sum_i f_i$$

(This raises questions about the validity of quoting photometric statistical precisions if objects are systematically biased)

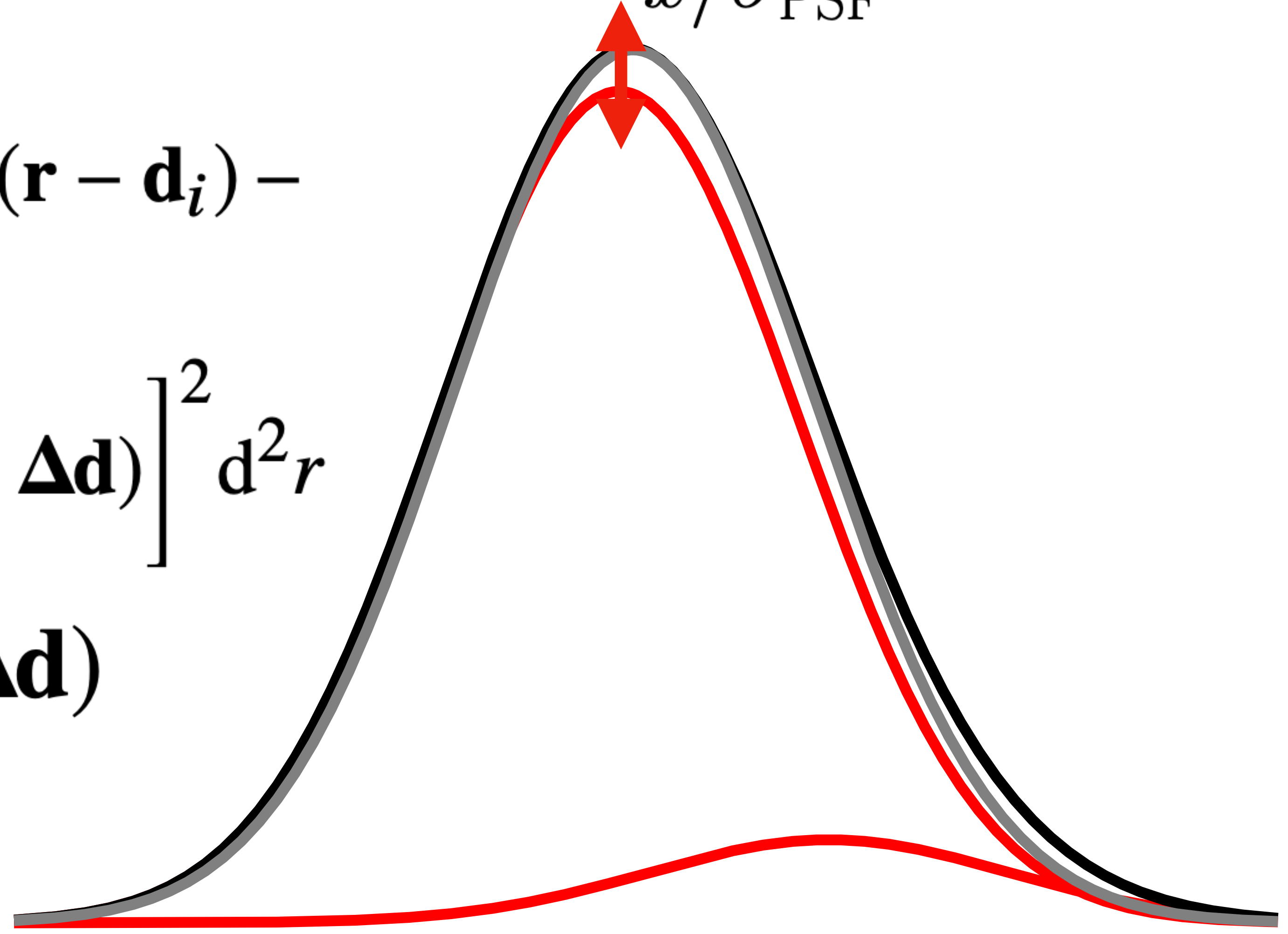
**Low SNR PSF Photometry**




$$\log \mathcal{L} = -\frac{1}{2} \times L \int_{-\infty}^{\infty} \left[ \phi(\mathbf{r}) + \sum_i f_i \phi(\mathbf{r} - \mathbf{d}_i) - (1 + \Delta f)\phi(\mathbf{r} - \mathbf{\Delta d}) \right]^2 \mathrm{d}^2 r$$

$$\Delta f = \psi'(\mathbf{\Delta d}) - 1 + \sum_i f_i \psi'(\mathbf{d}_i - \mathbf{\Delta d})$$



Wilson & Naylor (2018b, in prep.)
Plewa & Sari (2018)

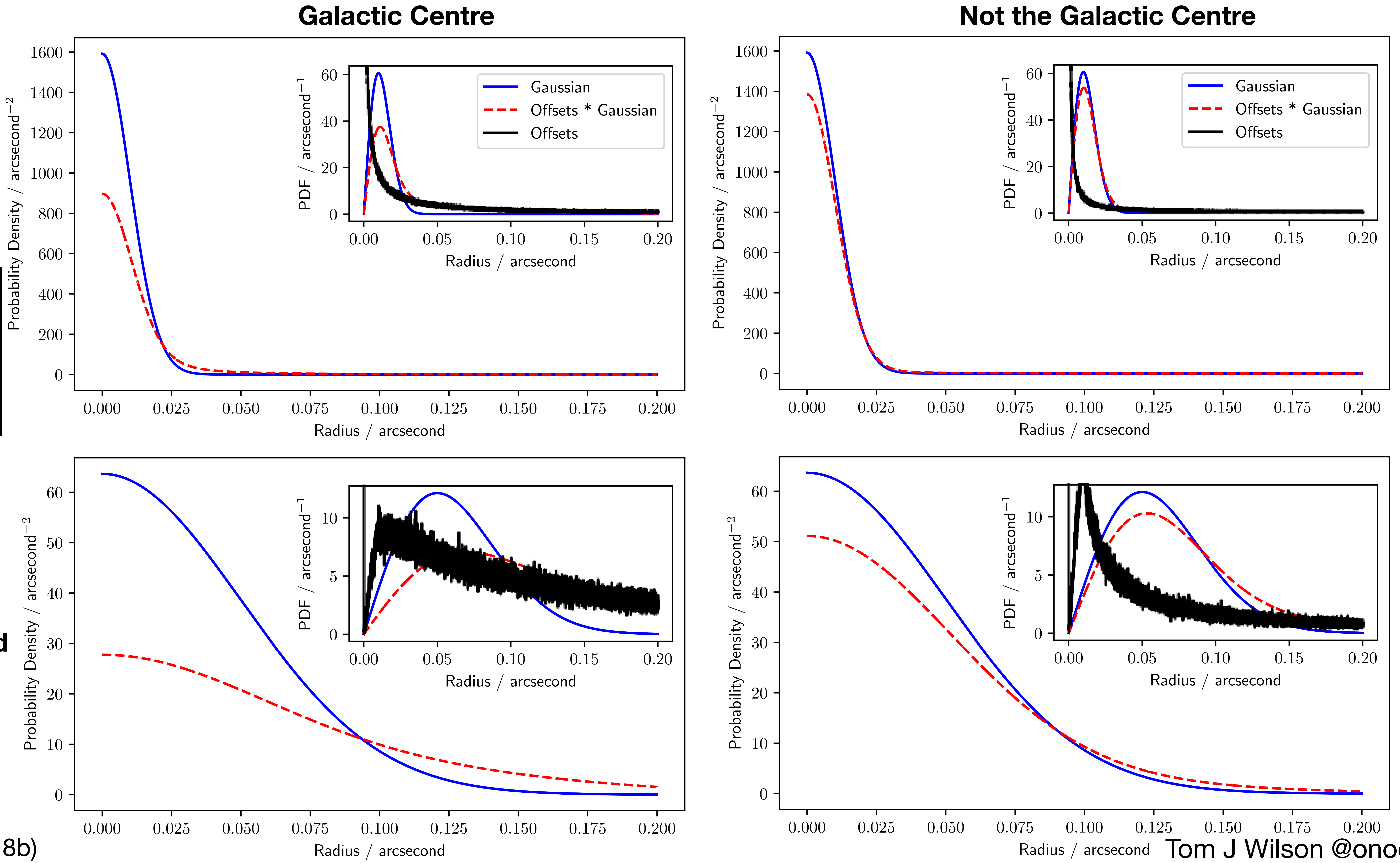
Tom J Wilson @onoddil

# The Rubin AUF: Galactic Plane

**Single-visit**

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

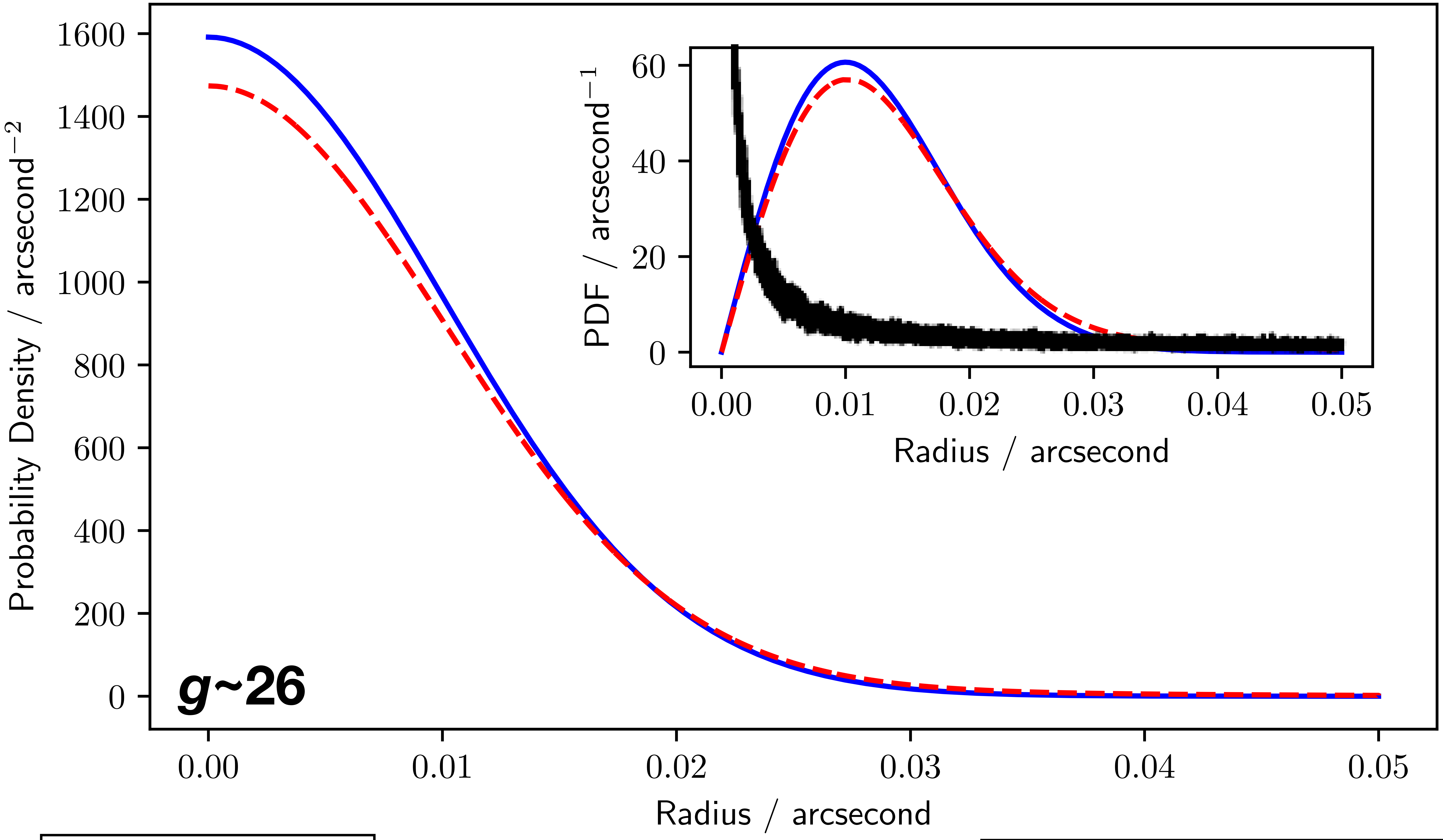
**Co-add**

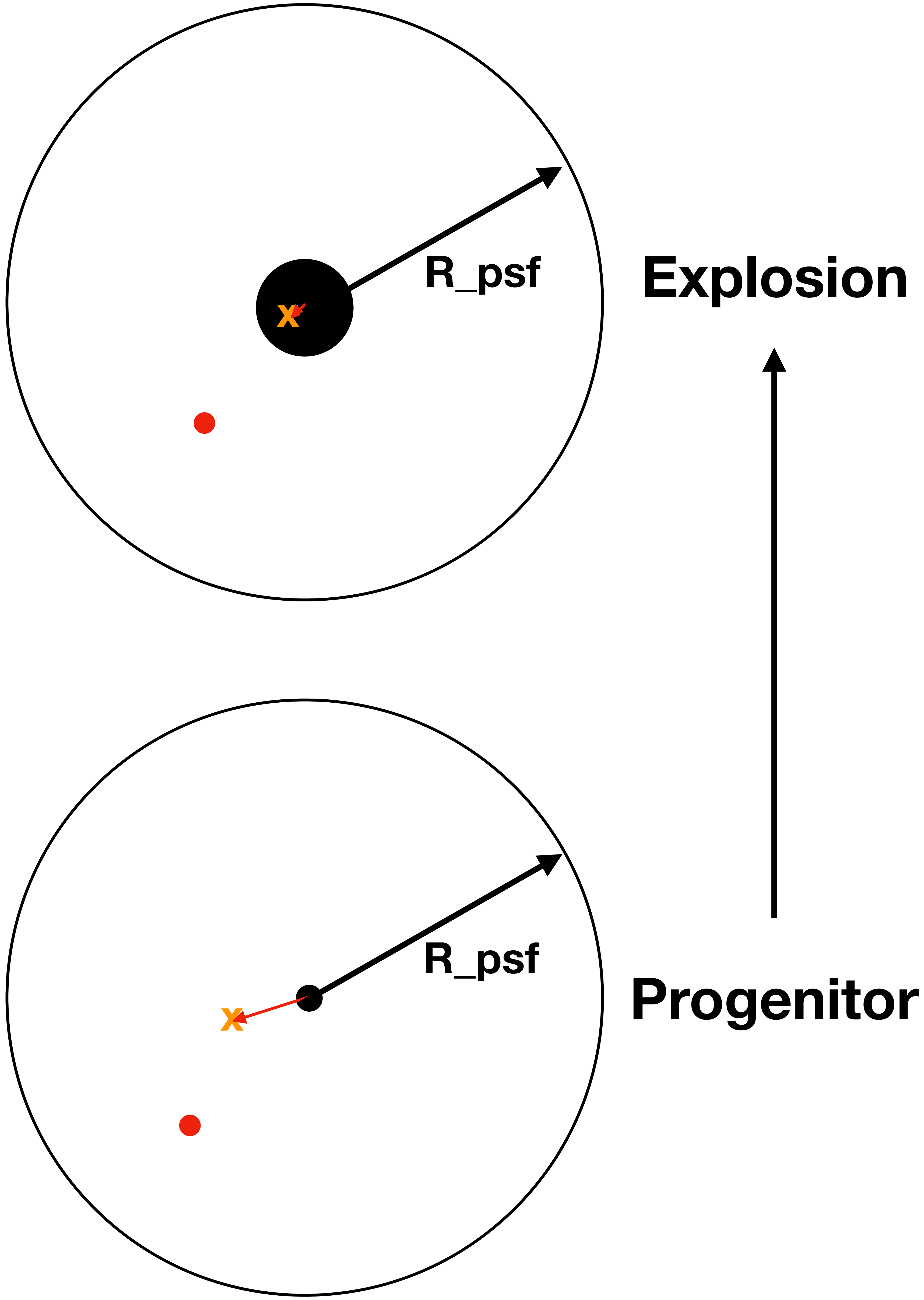



Wilson & Naylor (2018b)

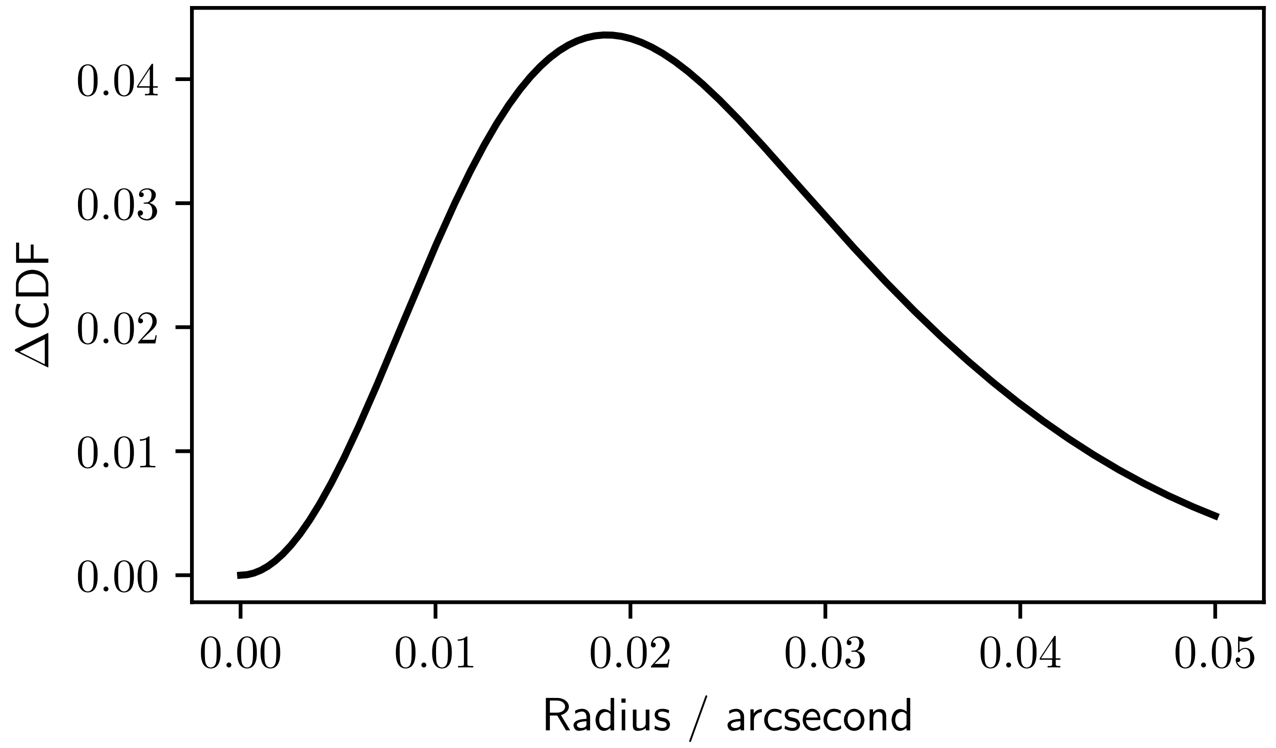
Tom J Wilson @onoddil

# The Rubin AUF: Extra-Galactic, Transients



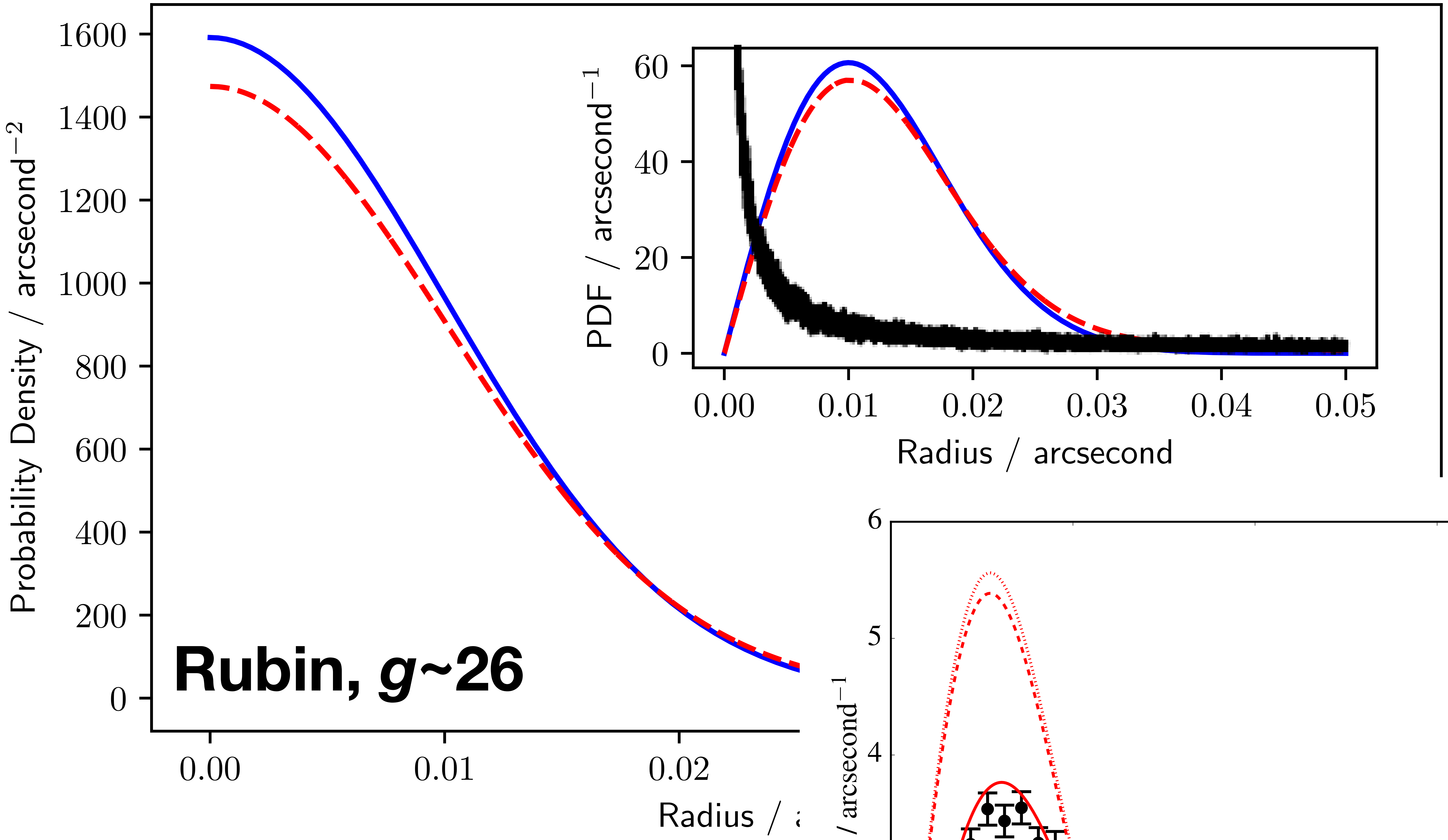

**Without modelling this extra effect, we fail to recover many true pairings, with an artificially high false negative rate!**

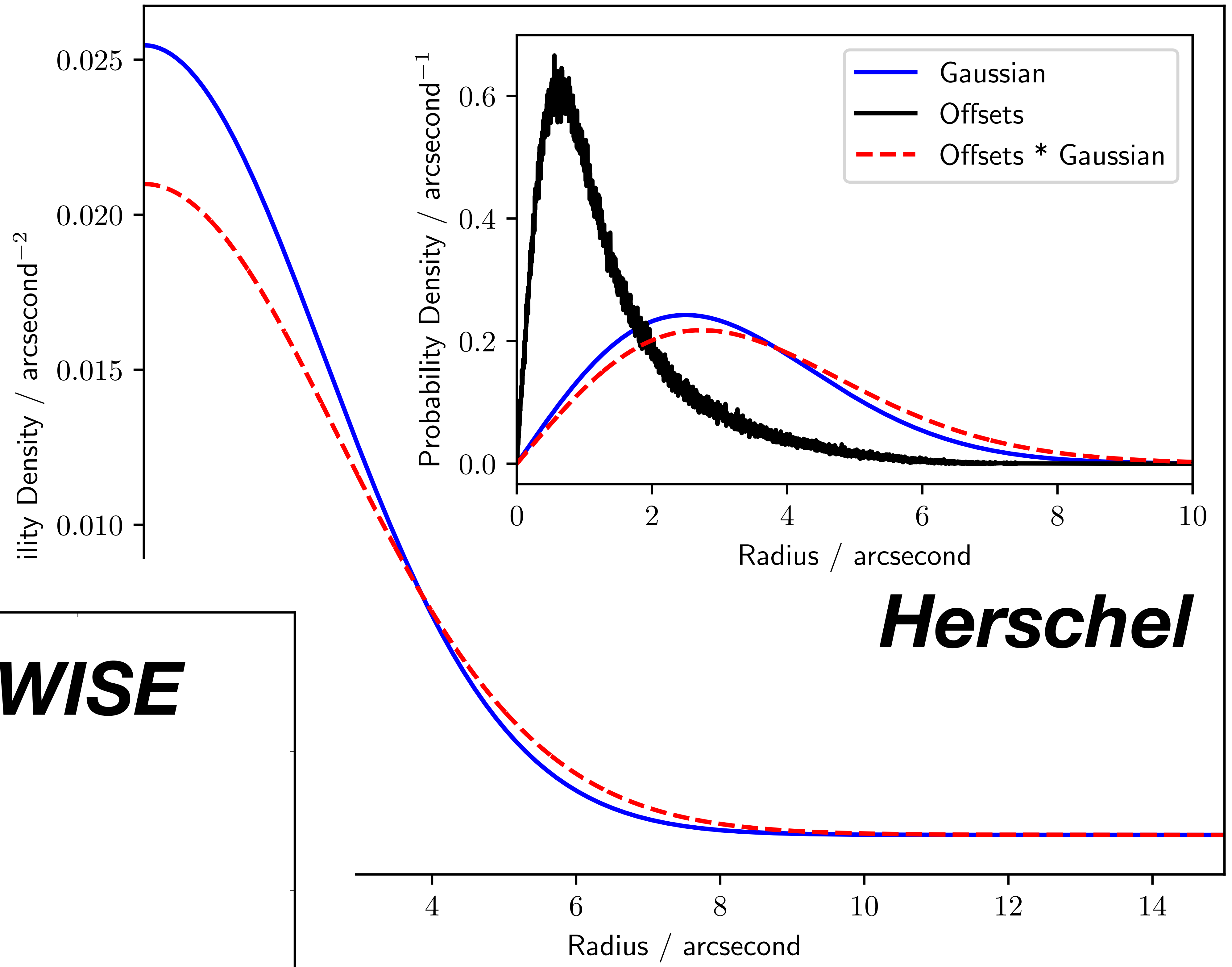


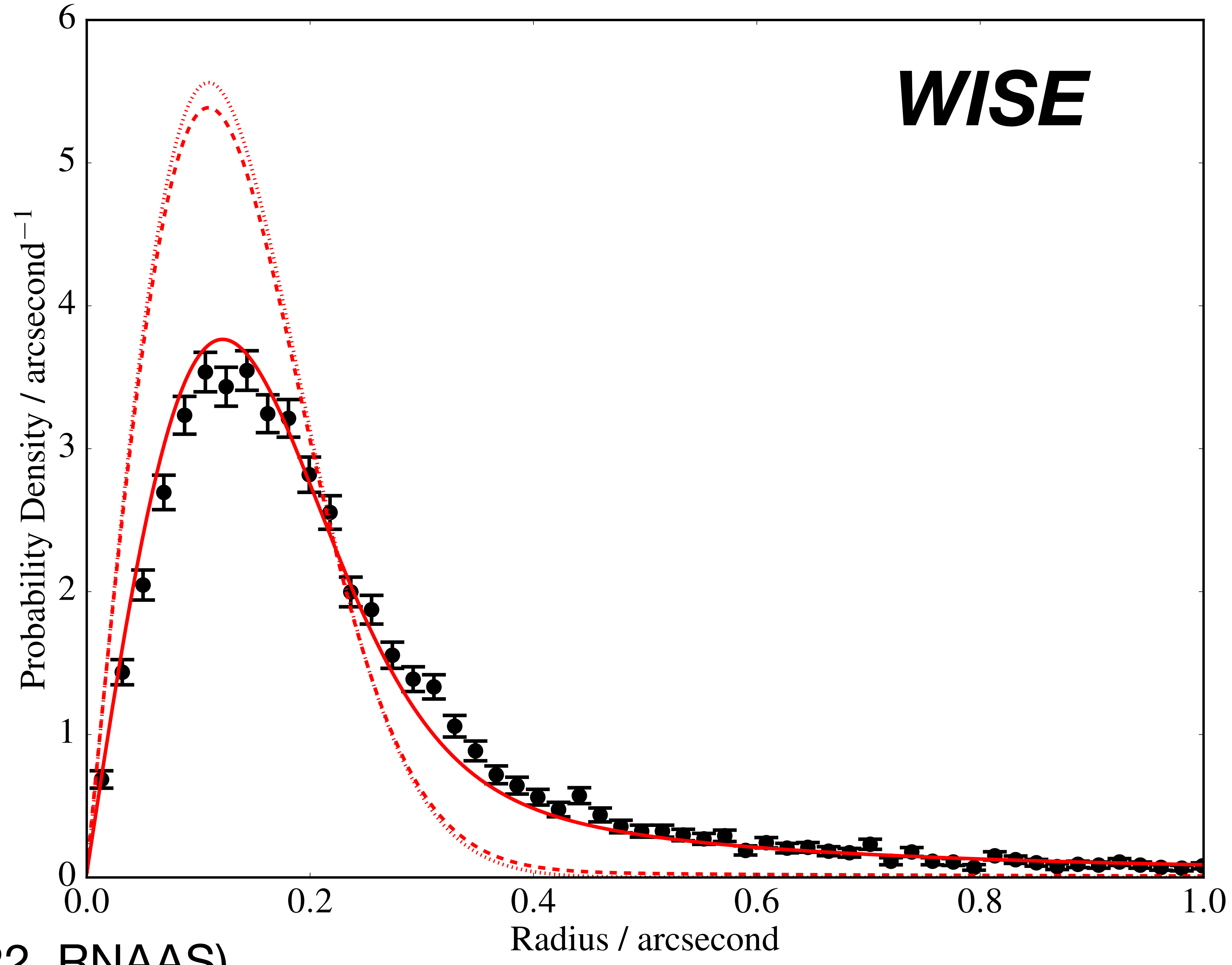

Wilson & Naylor (2018b); also see Wilson (2022, RNAAS)

Tom J Wilson @onoddil

# 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

# Matching Across Catalogues using the Astrometric Uncertainty Function and Flux

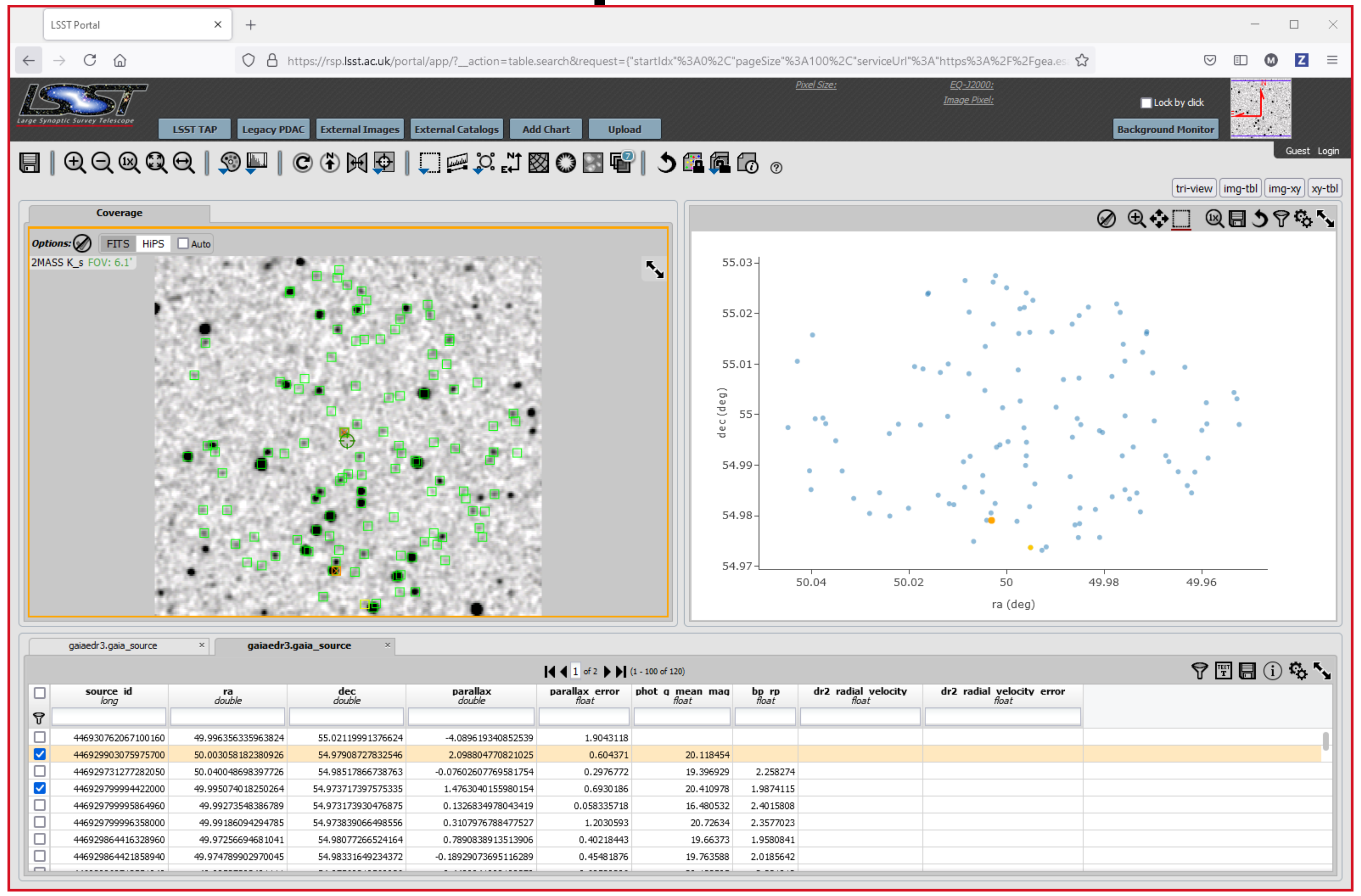
**https://github.com/macauff/macauff**

Tom J Wilson @onoddil

# How To Use Our Cross-Matches and Super-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 eq. 26, Wilson & Naylor 2018a)**
- **Eta - Photometric likelihood ratio (counterpart vs non-match probability, just for brightnesses; see equation 37 of Wilson & Naylor 2018a)**
- **Xi - Astrometric likelihood ratio (just position match/non-match comparison; see eq. 38, 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**

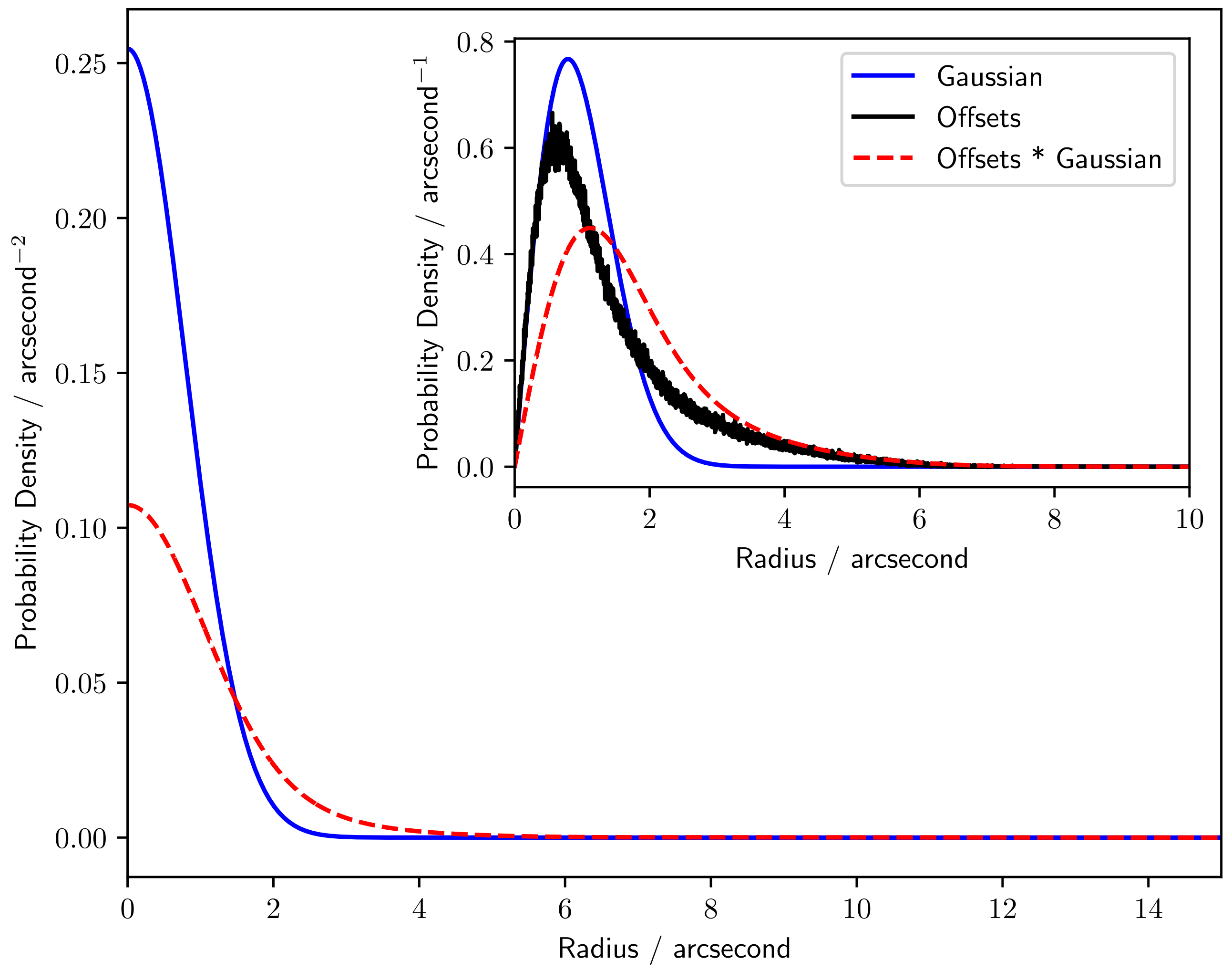
**Example columns:**

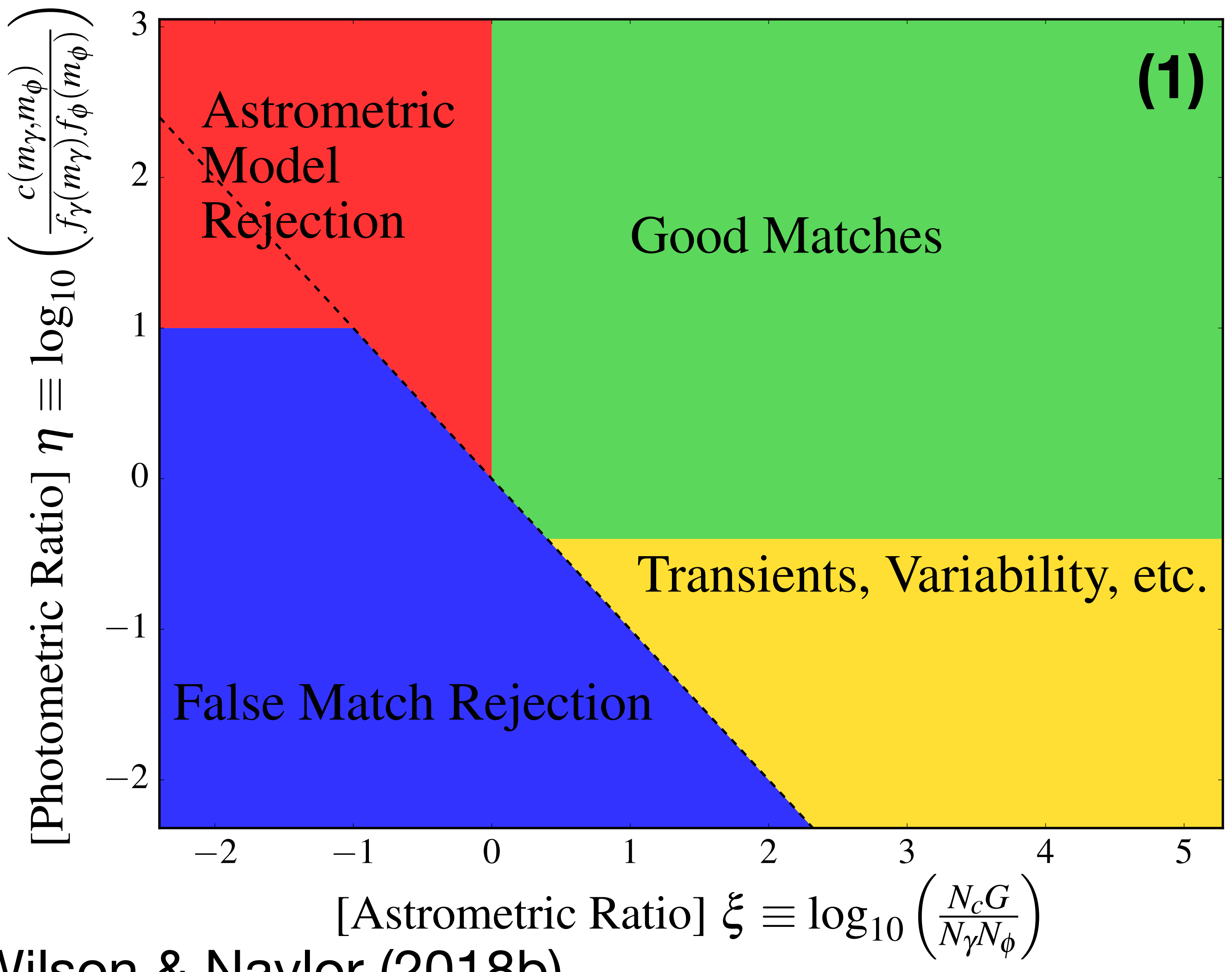
- **Designations of *N* sources (e.g., WISE J…, *Gaia* EDR3…, 2MASS J…)**
- **Super-match probability — probability of the given permutation**

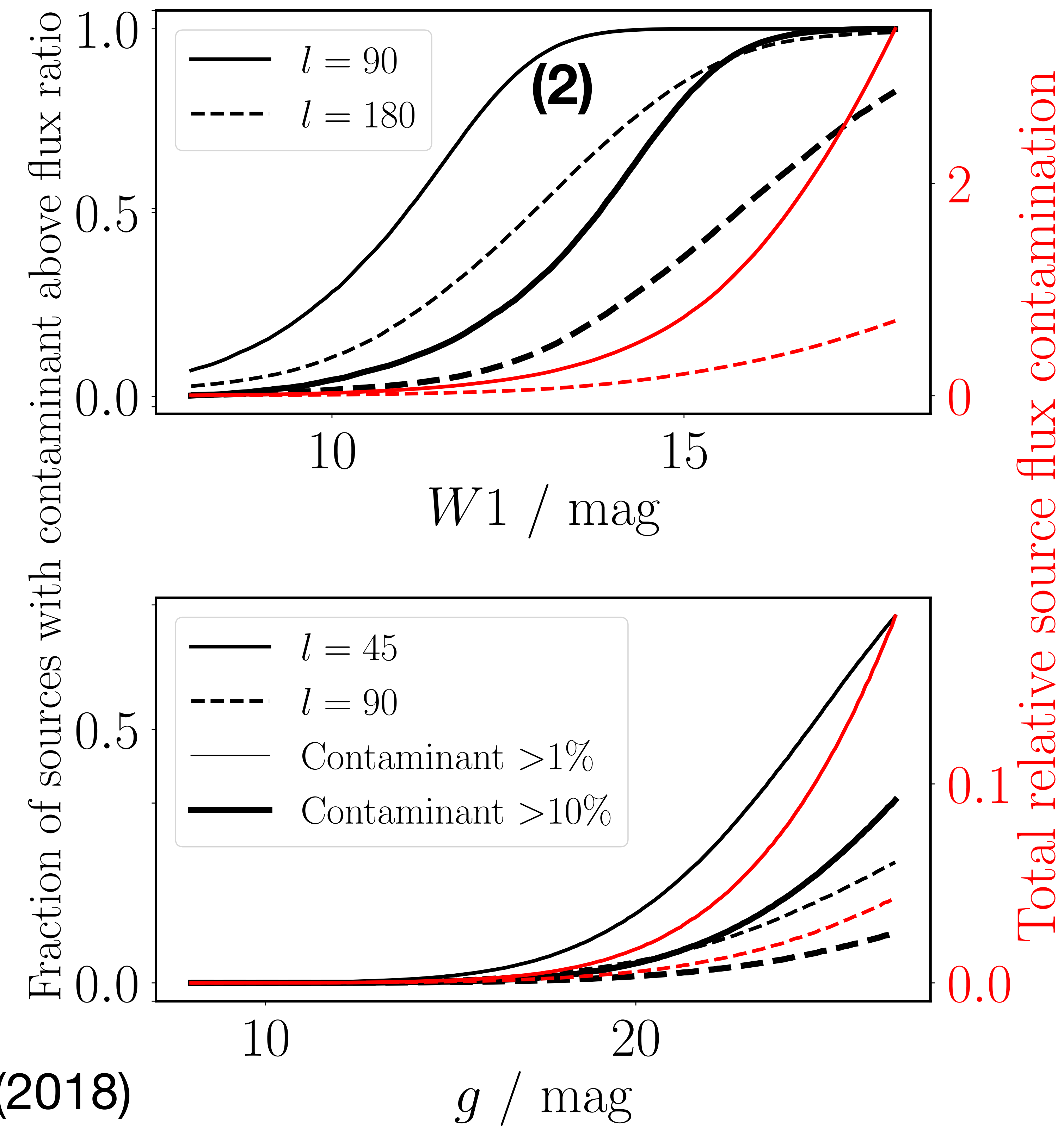
Tom J Wilson @onoddil

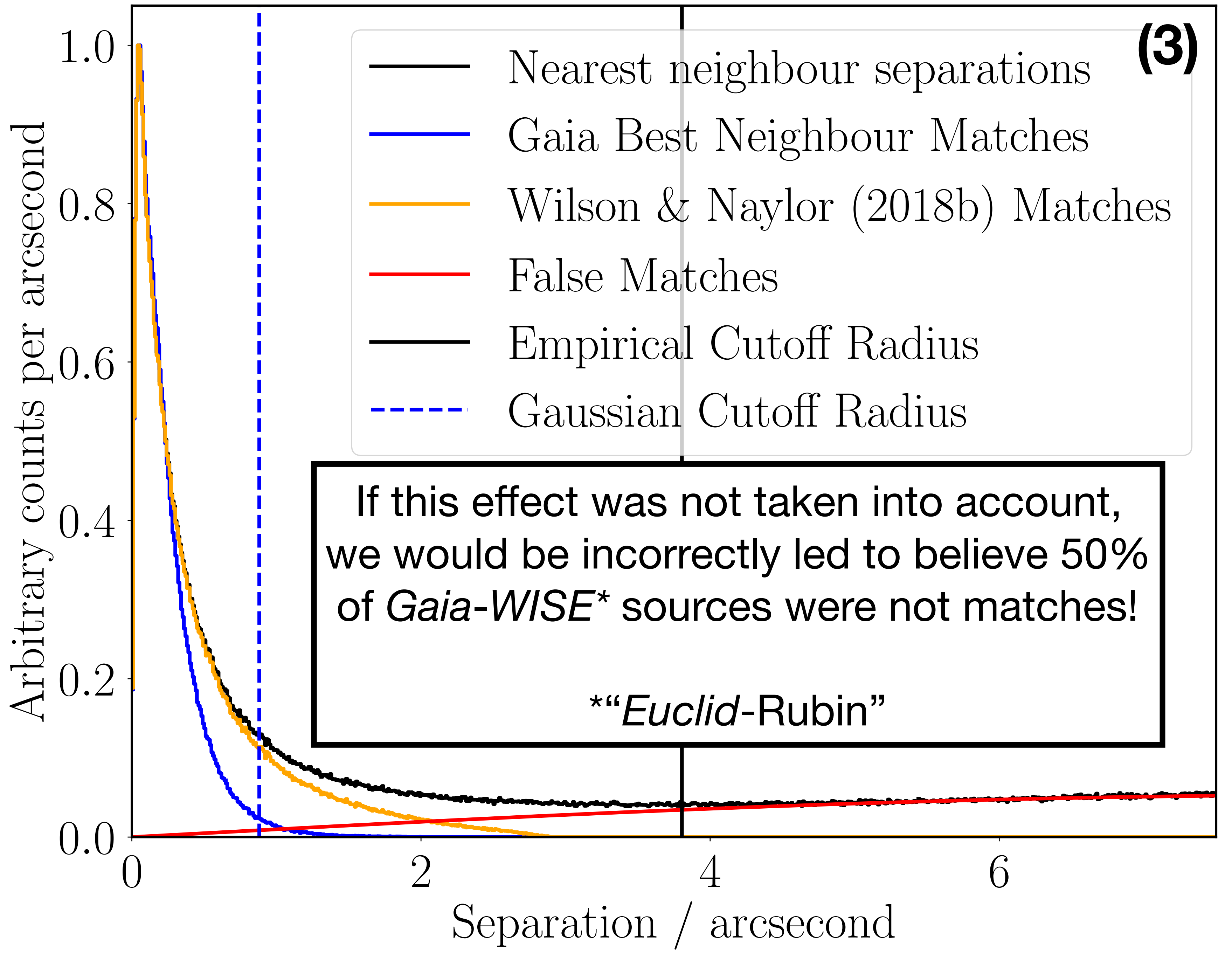
# Why Use Our Cross-Matches (and Extensions)?

**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

- 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

- **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!>**

University of Exeter

@Onoddil @pm.me
@Onoddil .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

# Conclusions

- **Cross-Matches to LSST Data Releases with macauff – Form**

- **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!>**

University of Exeter

@Onoddil@pm.me

@Onoddil.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