

Solving the Catalogue Cross-Match Problem in the Era of LSST

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

Southampton, 15/10/24

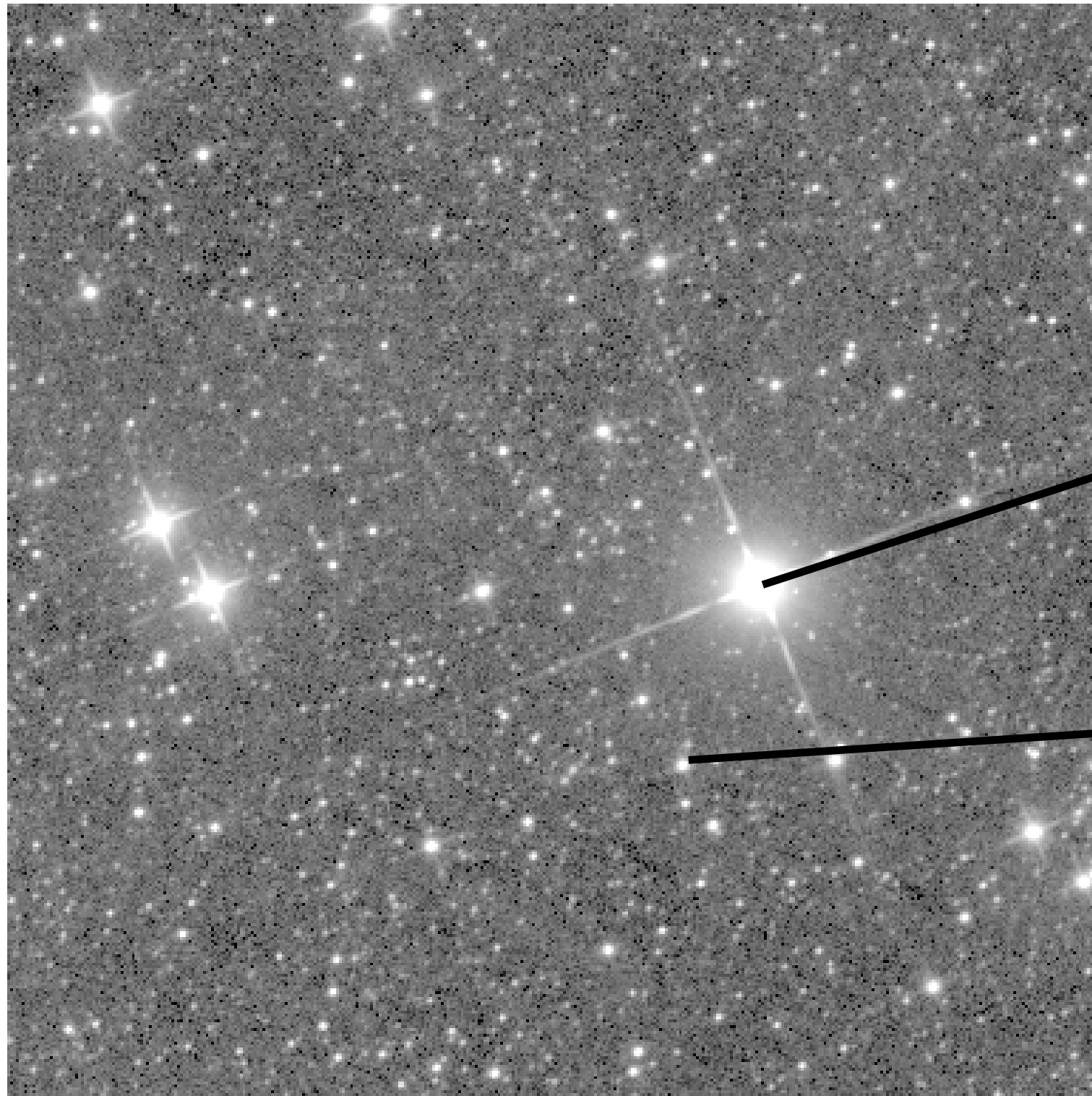


 @Onoddil  @pm.me  .github.io

Tom J Wilson @onoddil

What's In A Photometric Catalogue?

(Ironically, it's half astrometry!)

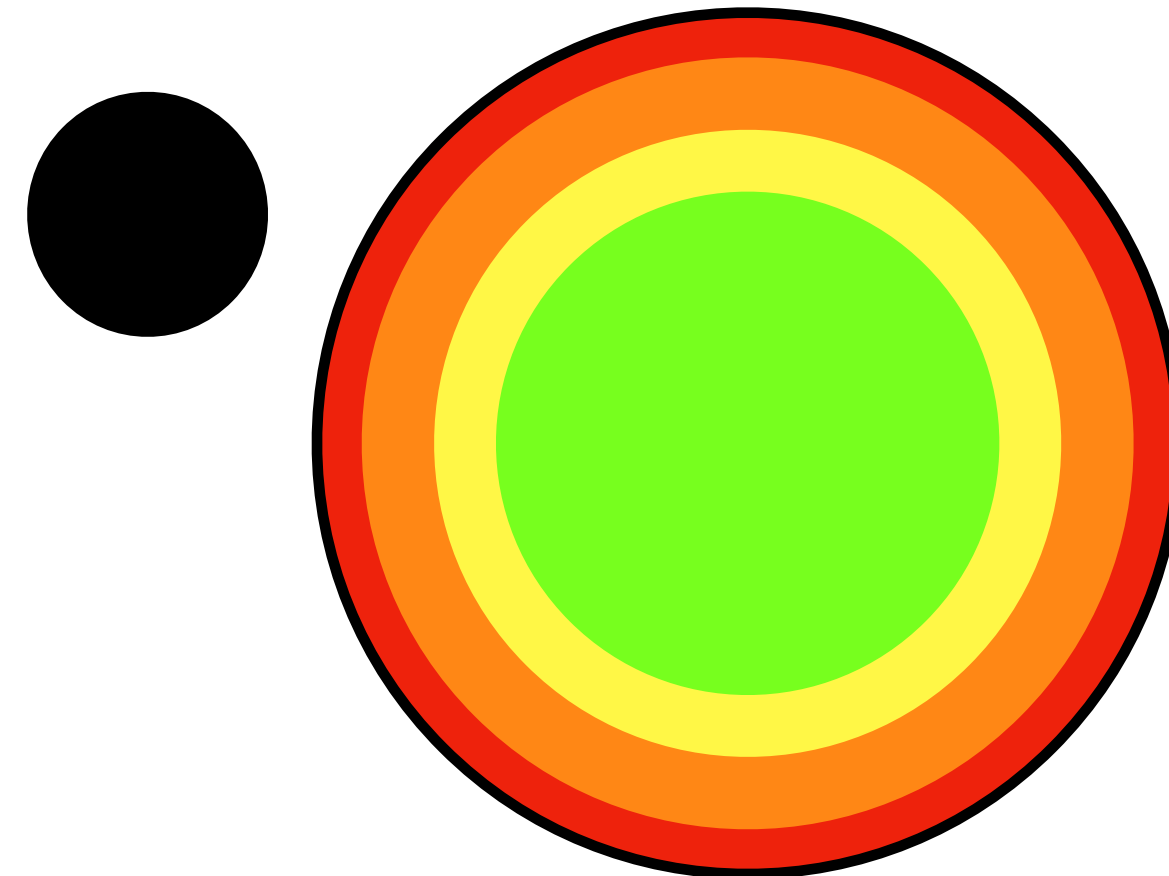


WISE - Wright et al. (2010)

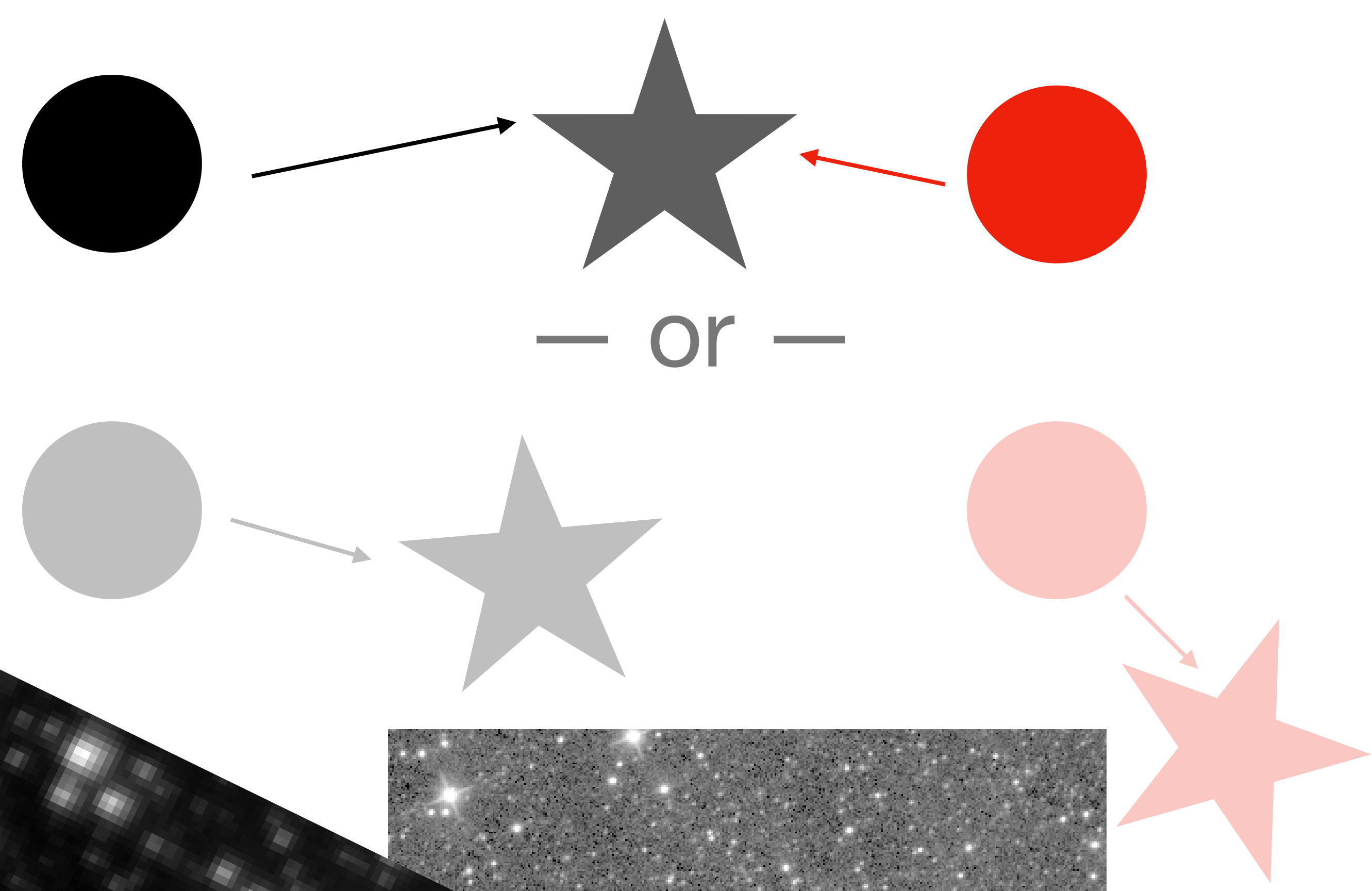
Source ID	Position (deg)	Uncertainty (arcsecond)	Brightness (mag)	Uncertainty (mag)
-----------	----------------	-------------------------	------------------	-------------------

1	218.4763	0.073	14.94	0.04
---	----------	-------	-------	------

2	218.3951	0.217	20.32	0.15
---	----------	-------	-------	------

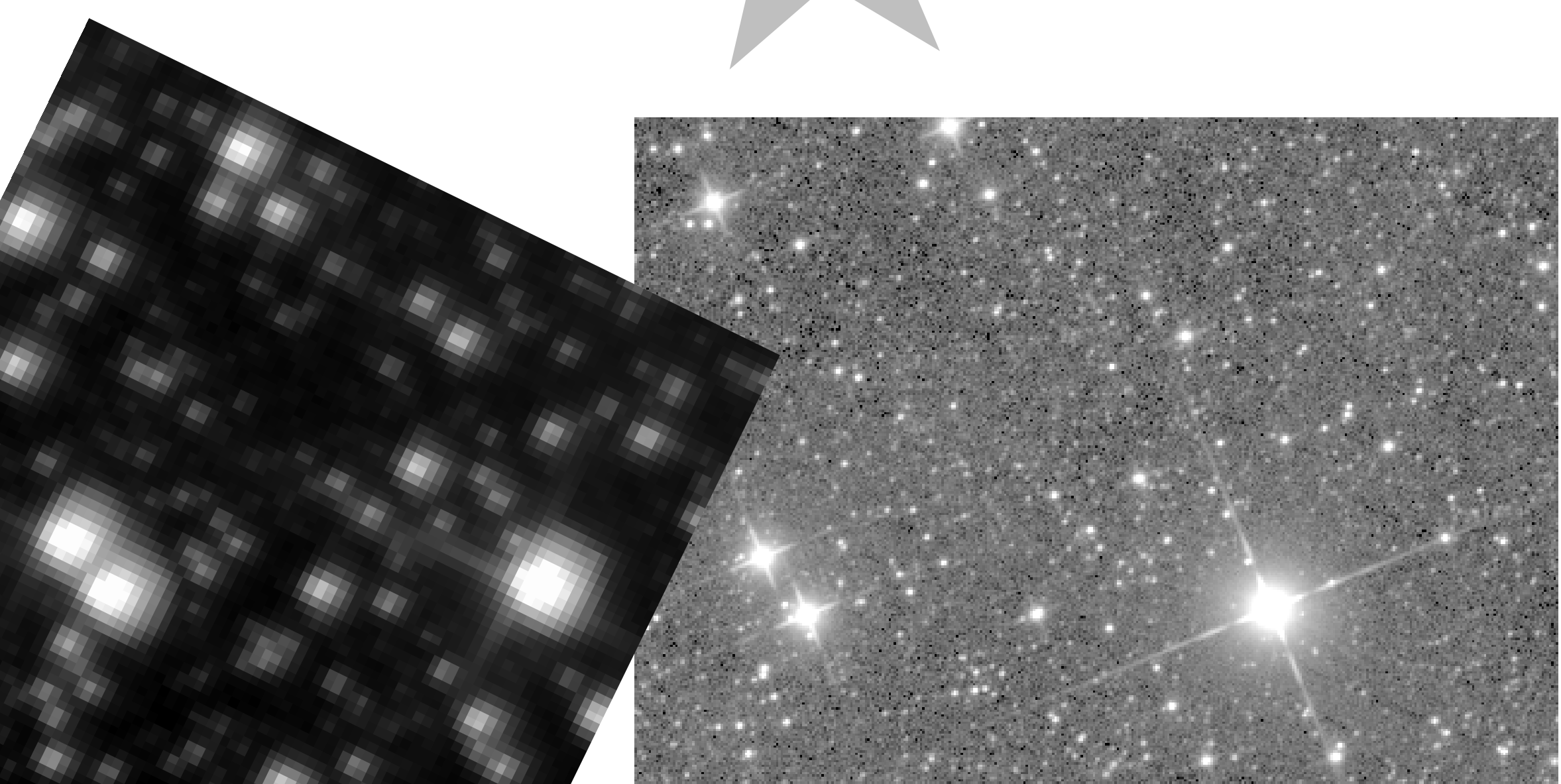


Cross-Matching and Counterpart Assignment



ID A	ID 1	A magnitude	Magnitude 1
A J...	CAT1 1	14.94	17.53
...

ID A	ID 1	A magnitude	Magnitude 1
A J...	NULL	14.94	NULL
NULL	CAT1 1	NULL	17.53
...

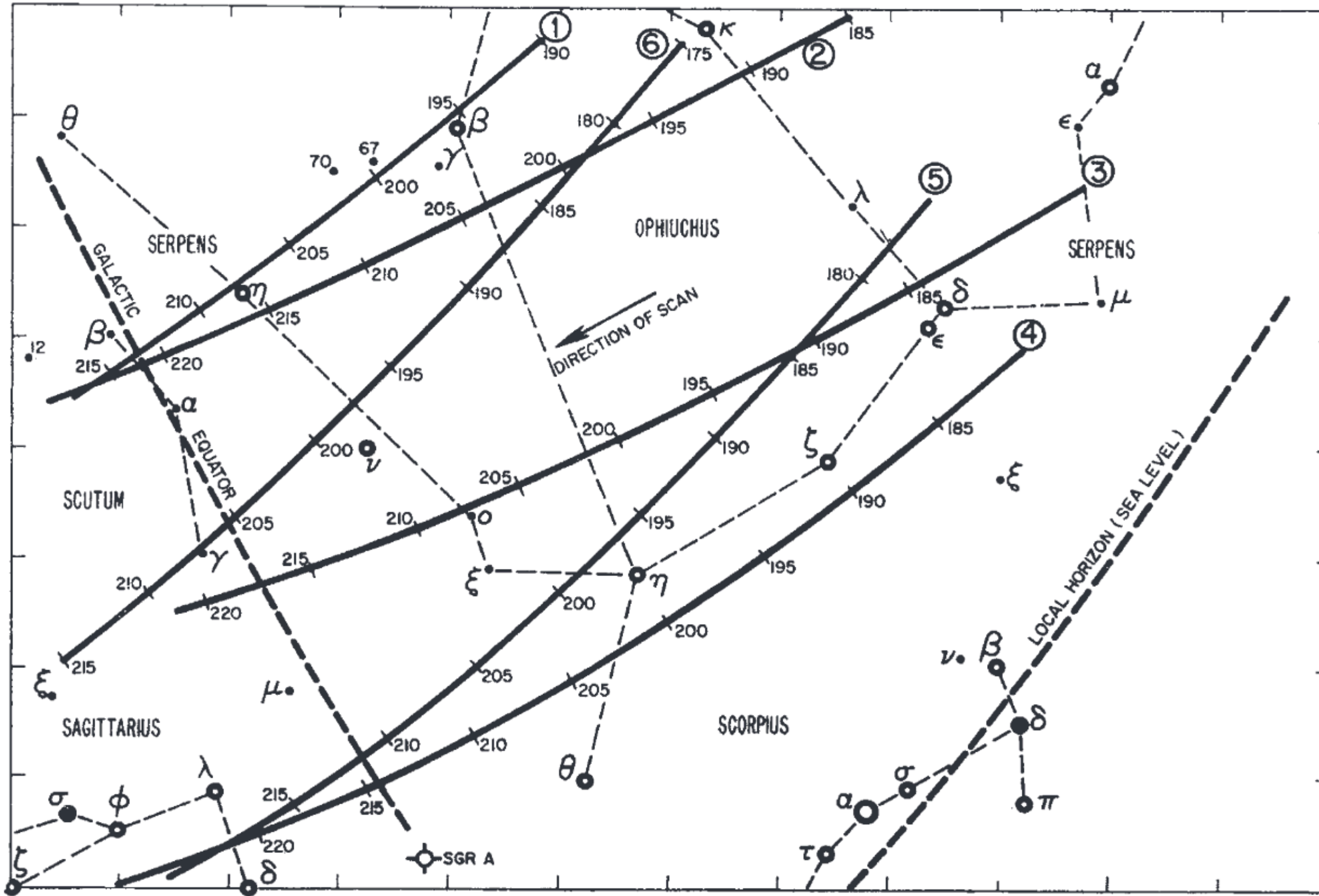


WISE - Wright et al. (2010)
TESS - Ricker et al. (2015)

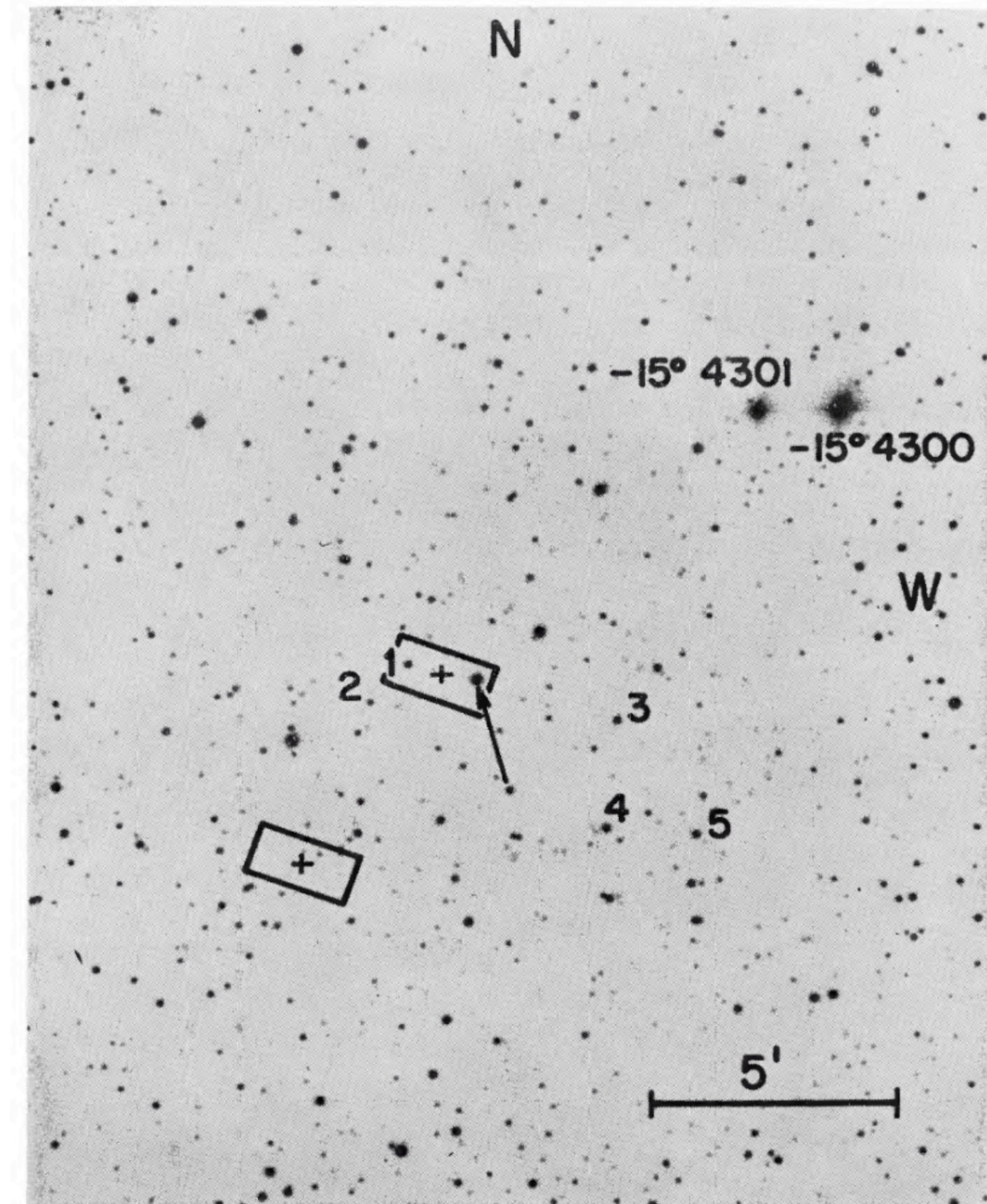
Technology Abounds

- **Ancient lists of stars (Ptolemy, 150; Brahe, 1598)**
- **Galileo invents the telescope (1610)**
- **Greenwich Observatory catalogues (e.g. Bradley, 1798)**
- **Astrophotography invented (Bond & Whipple, 1850)**
- **Harvard Observatory surveys (8th magnitude, 1882-1886)**
- **Astrographic Chart (11th magnitude; 1887-1962)**
- **Carte Du Ciel (14th magnitude; 1880s-never finished)**
- **Invention of the CCD (Boyle & Smith, 1970)**
- **InfraRed detector invented (Forrest et al. 1985)**
- **4- and 5-m class telescopes (1970s-1980s; e.g. LAT, MMT, UKIRT, CFHT, WHT)**
- **Space Telescopes (1980s-2010s; e.g. IRAS, ISO, AKARI, *WISE*, *Spitzer*)**
- **All-sky ground-based surveys (e.g. 2MASS, 1997-2001; SDSS, 2000-; Pan-STARRS, 2010-).**

X-ray Detections: Hunting for Sco X-1

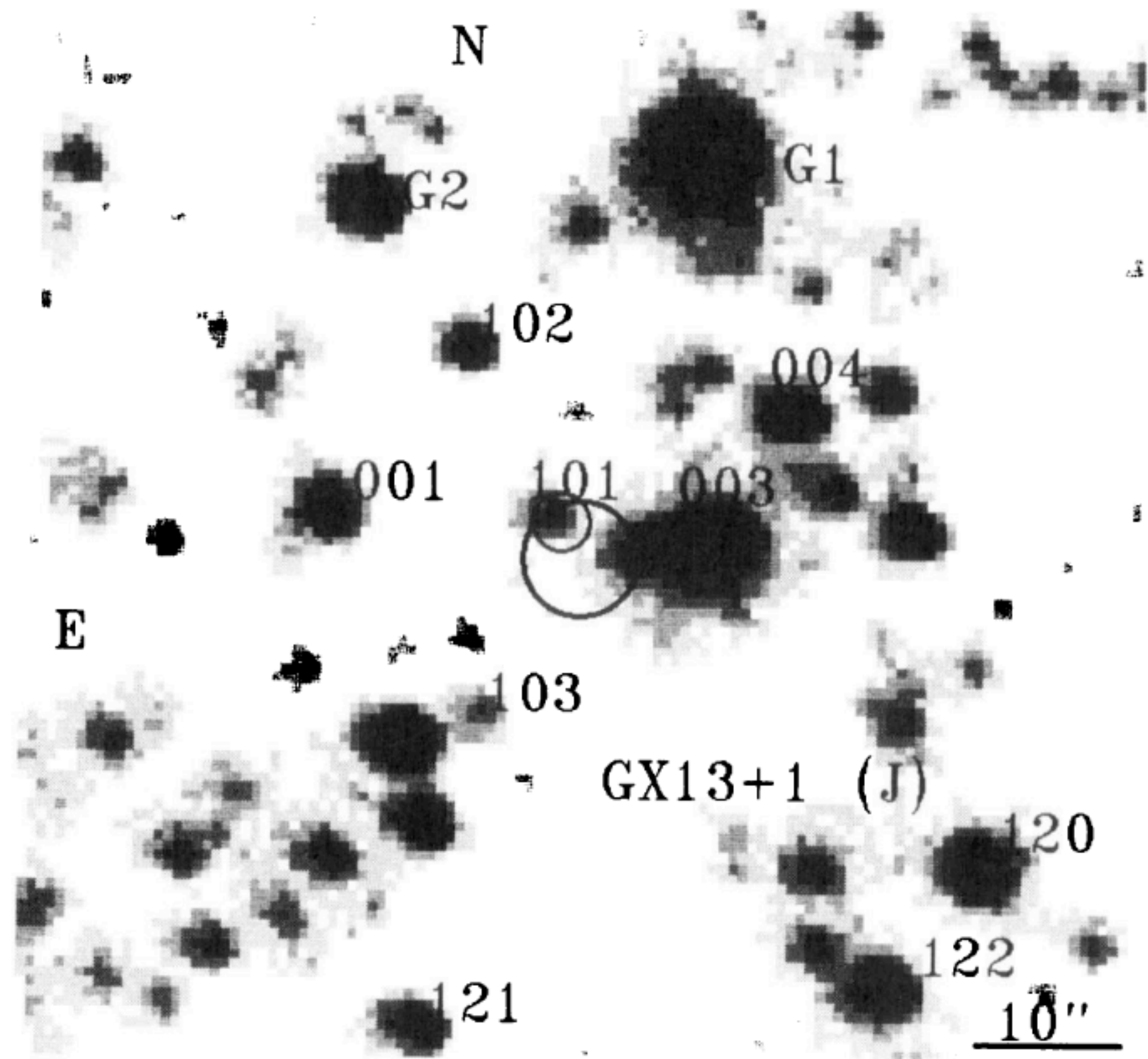


Giacconi, Gursky, & Waters (1964)

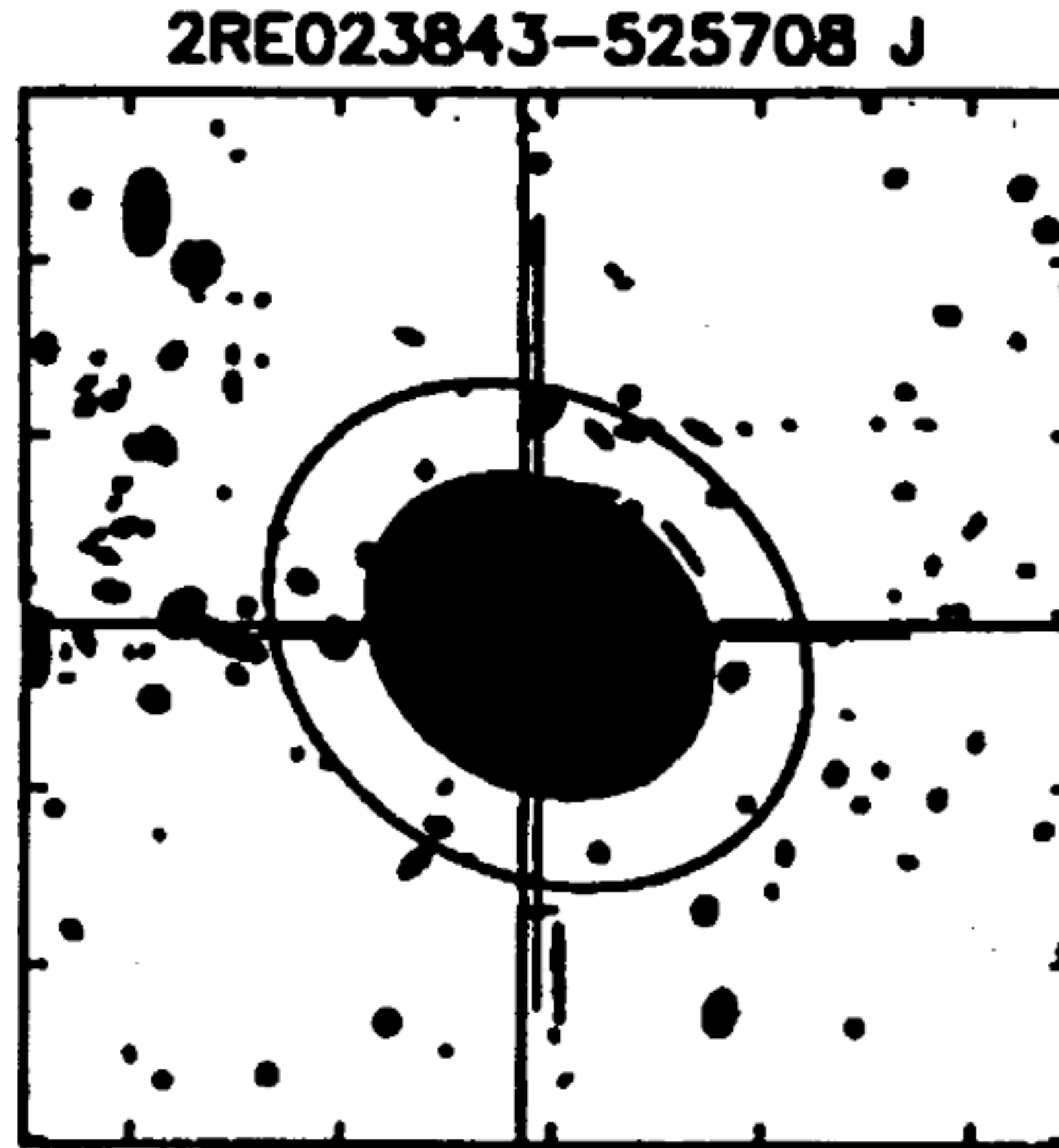


Sandage et al. (1966)

The Brightest Star in the Sky



Naylor, Charles, & Longmore (1991)

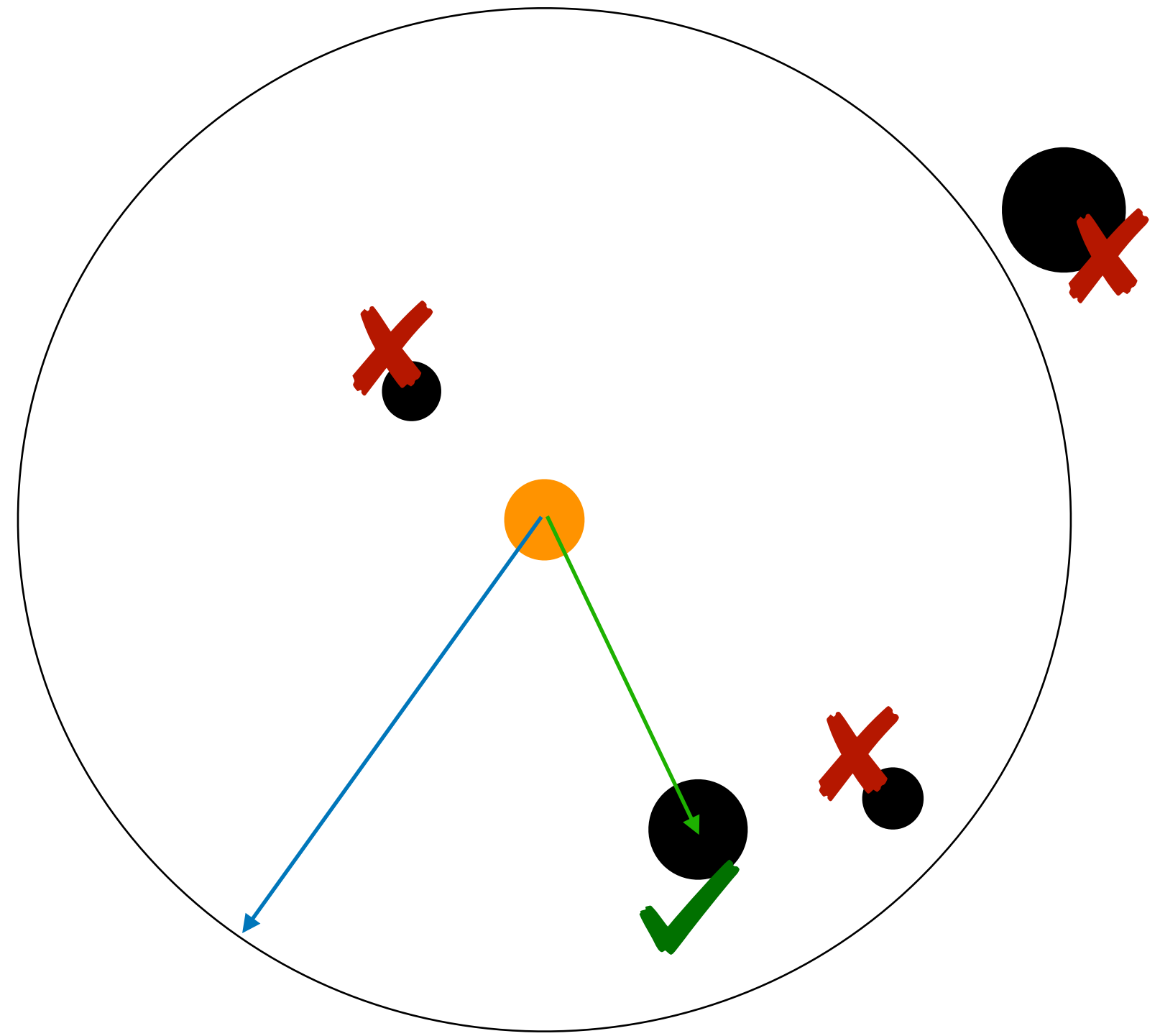


Mason et al. (1995)

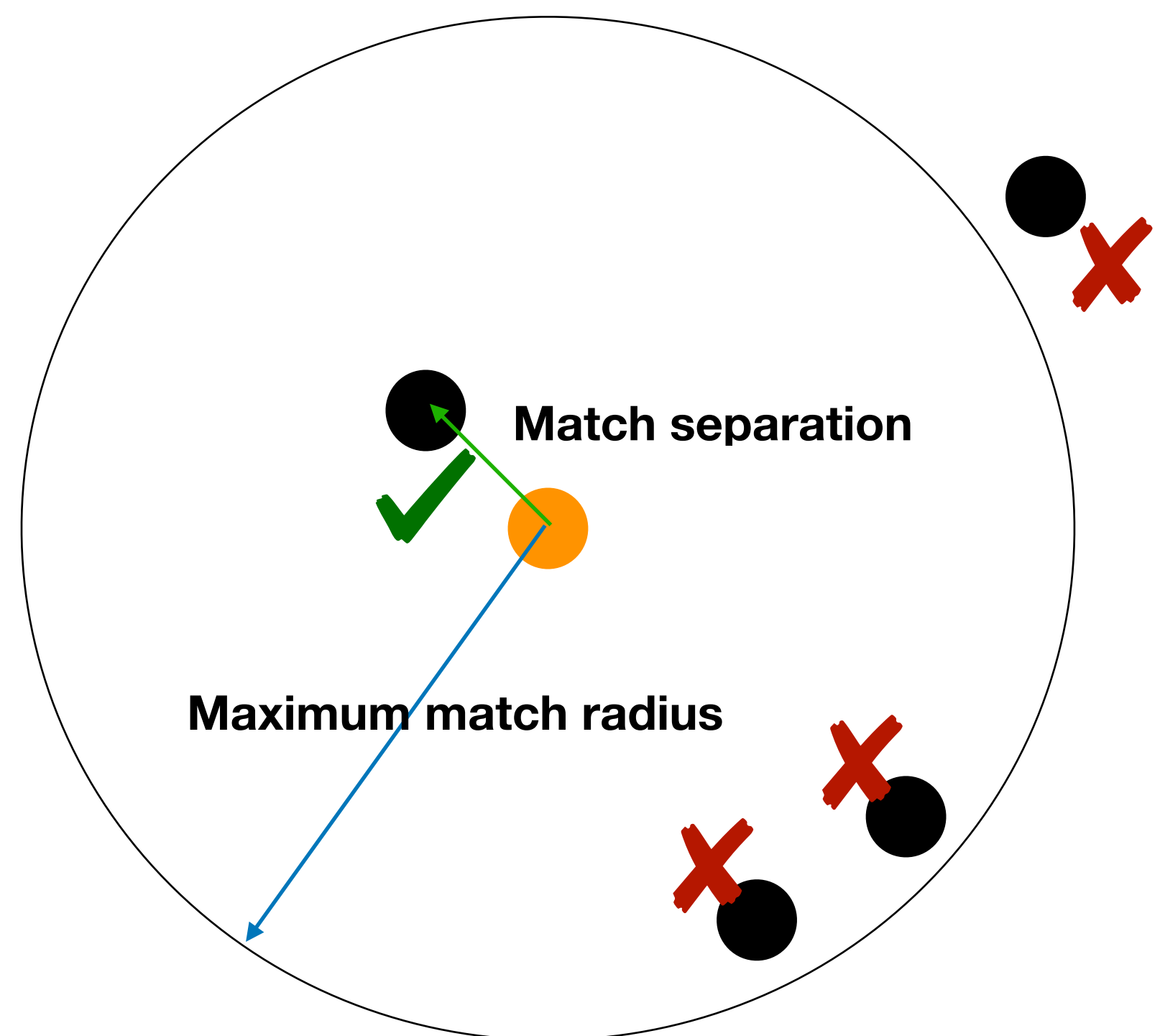
“...X-ray sources are rare events; bright optical sources are also rare events, so the observation of an X-ray source and a bright optical source in the same region of the sky is considered a non-random event”

Fotopoulou et al. (2016)

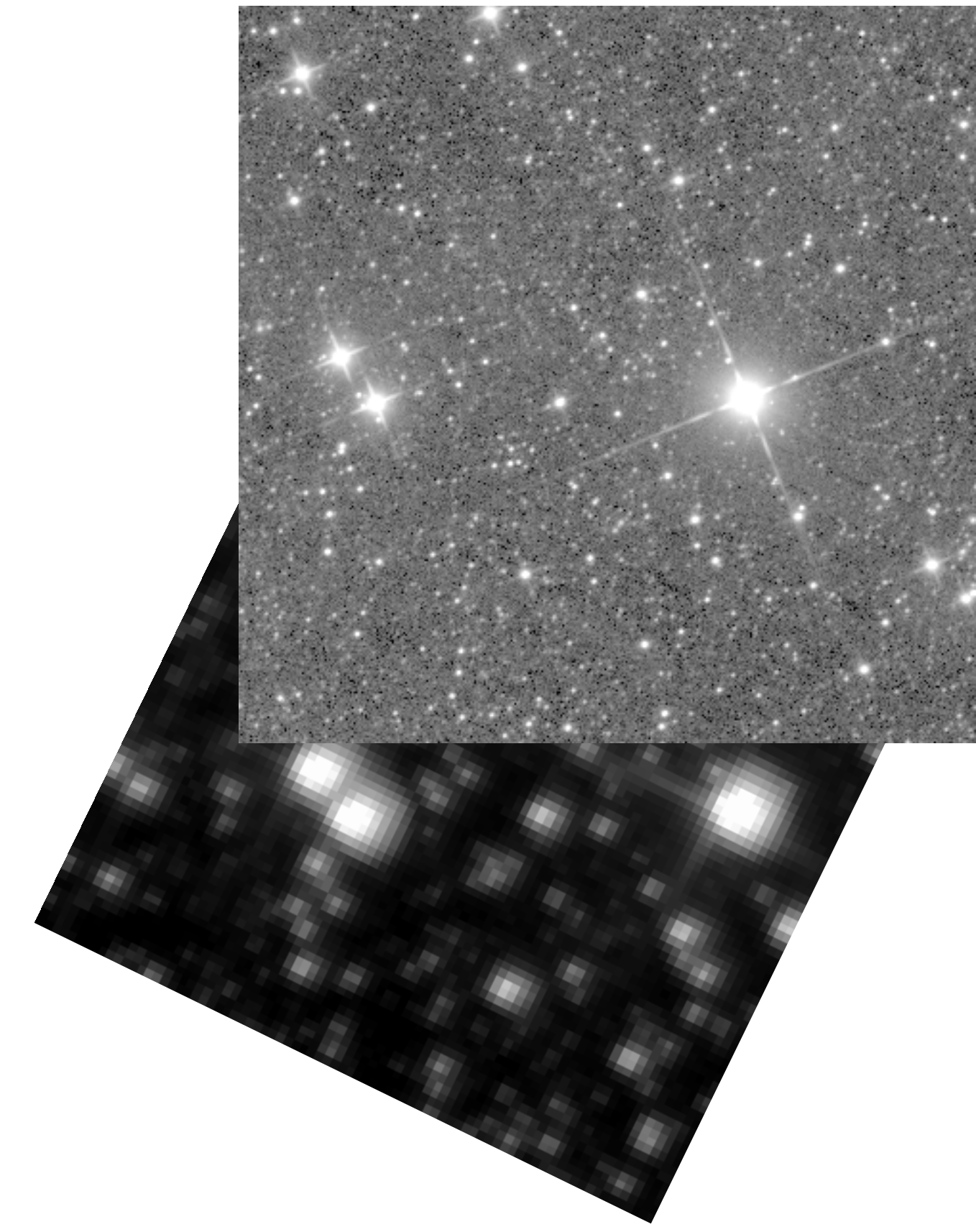
“Traditional” Cross-Matching



Declination / degrees



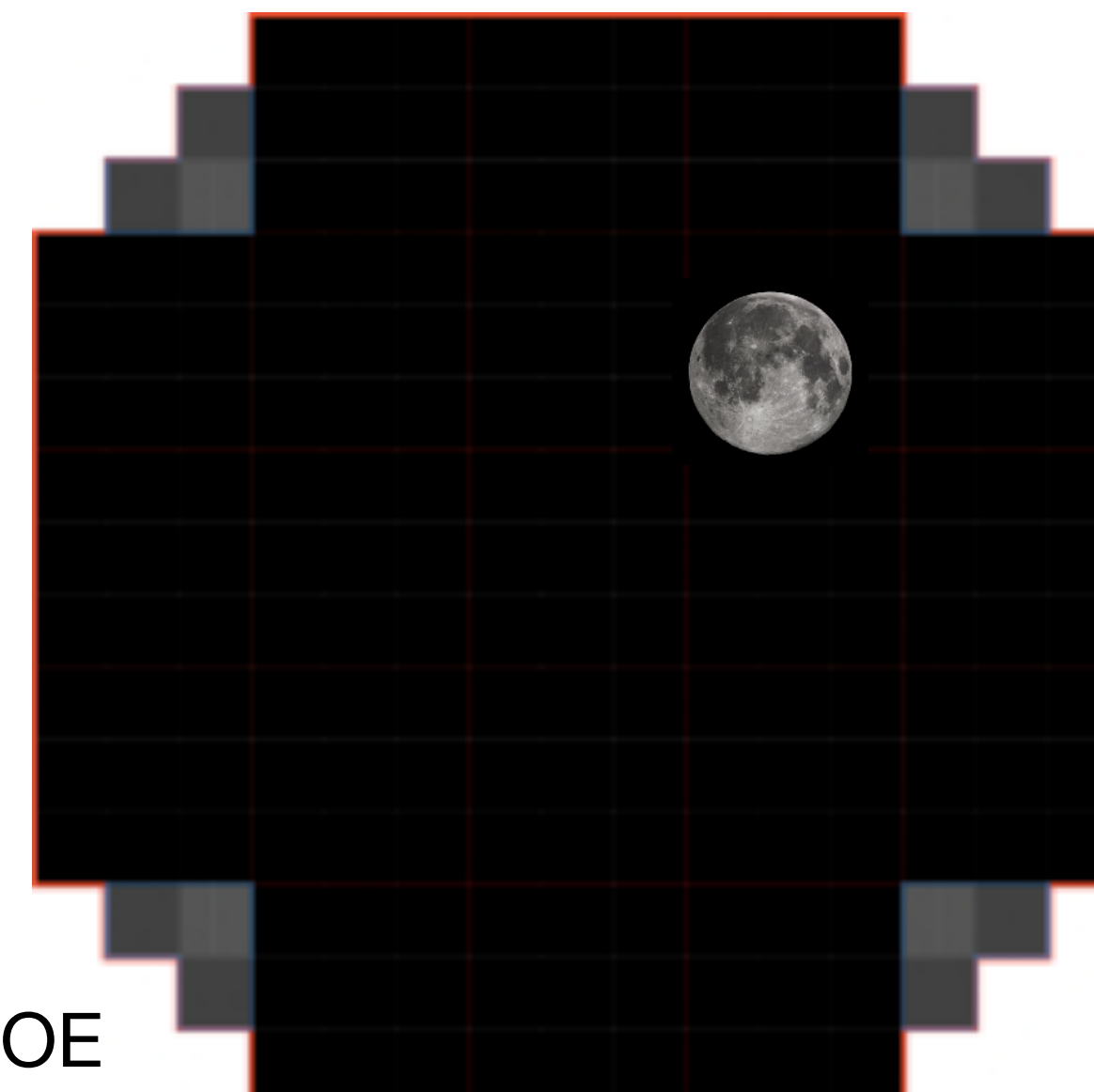
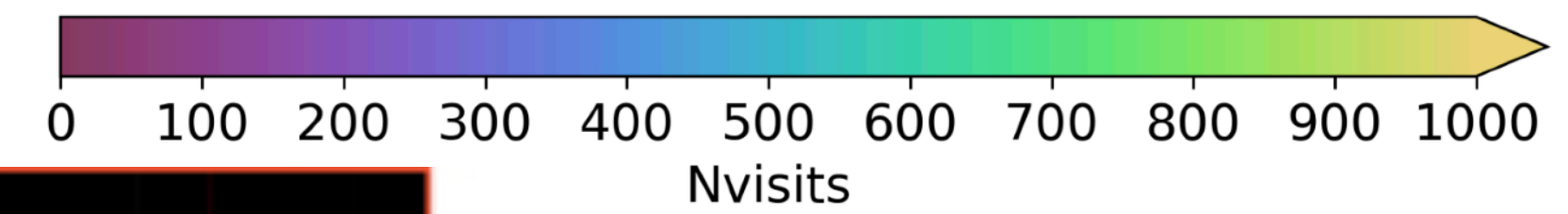
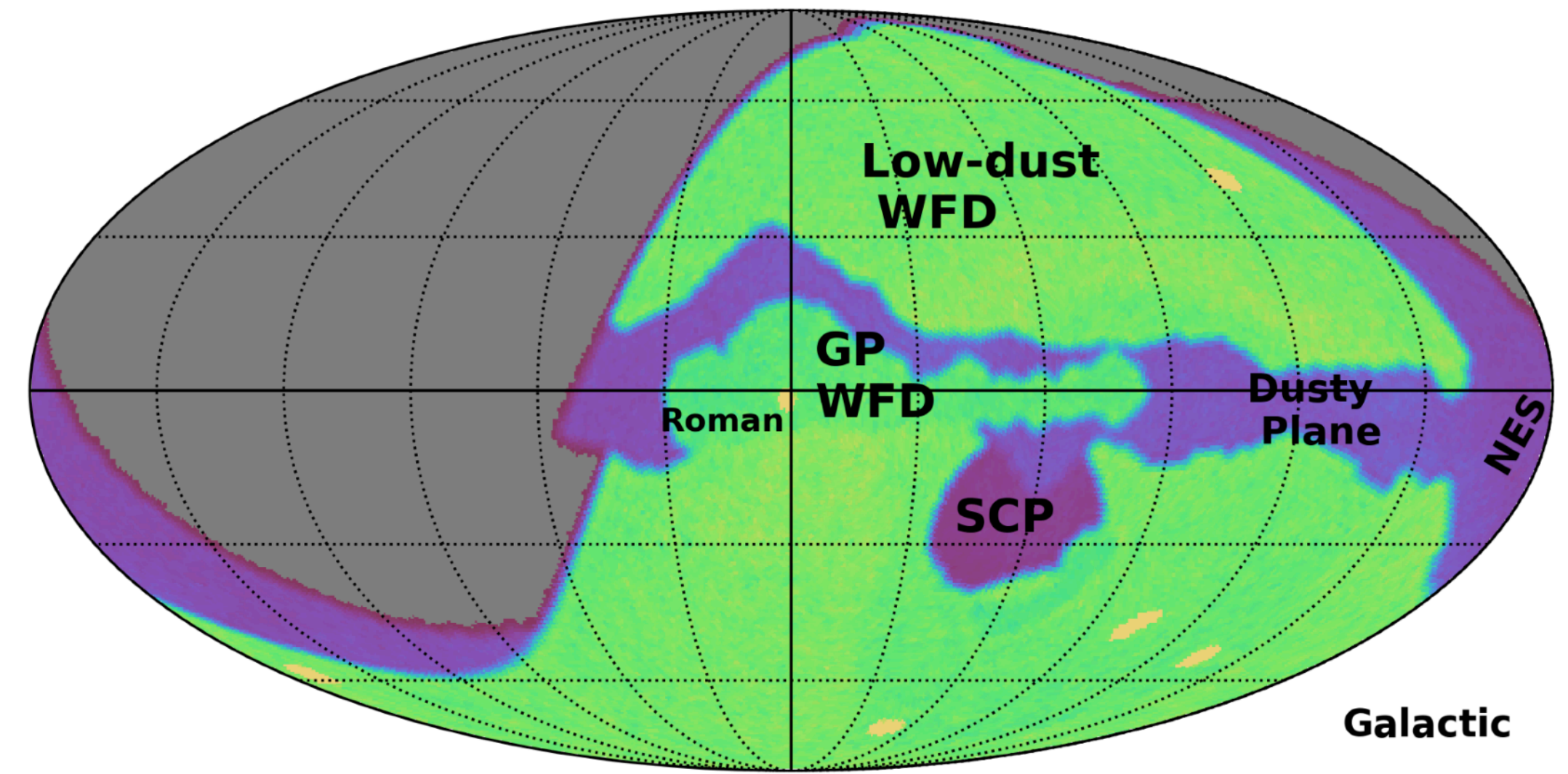
Right Ascension / degrees



The Vera C. Rubin Observatory's LSST



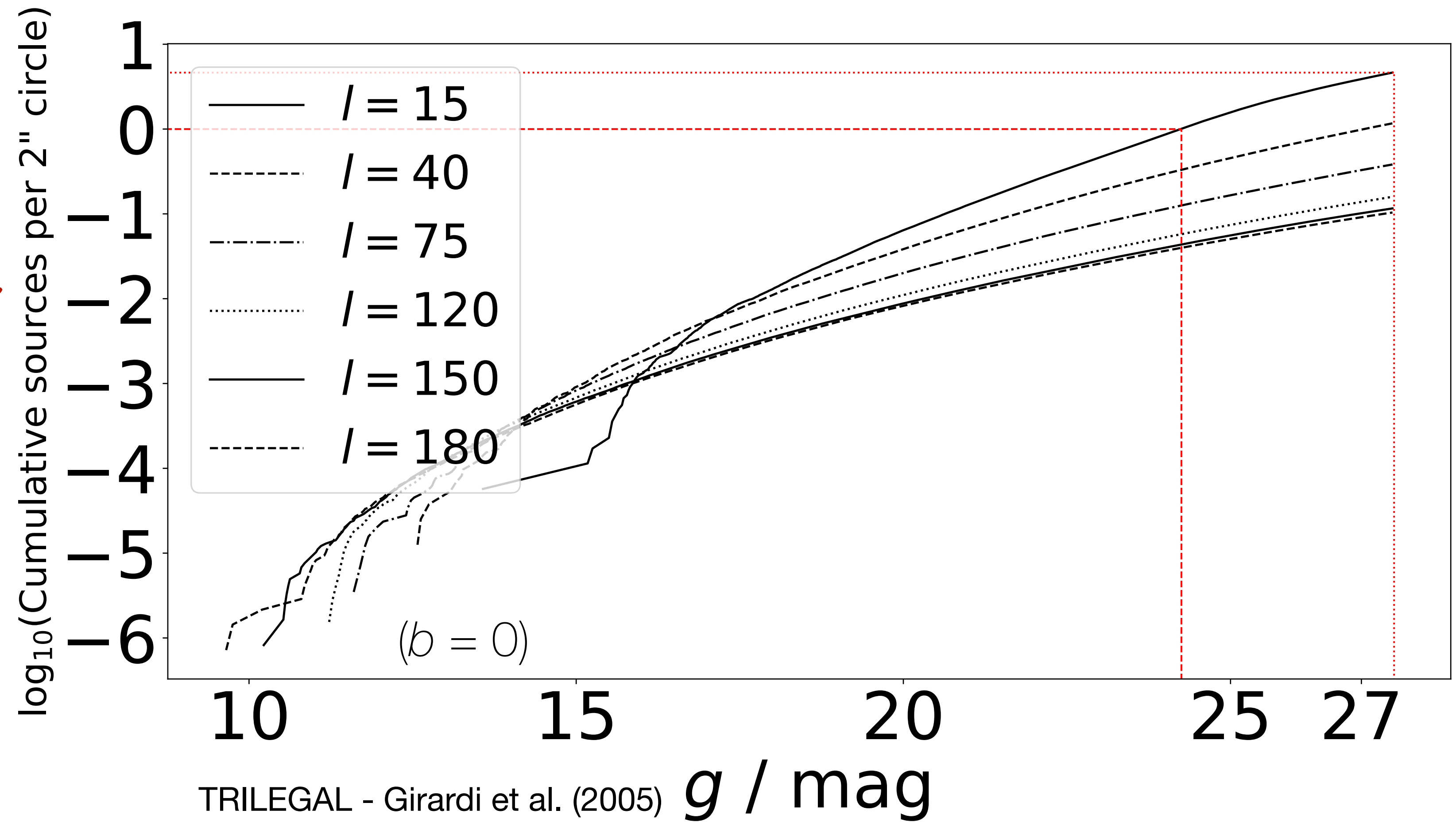
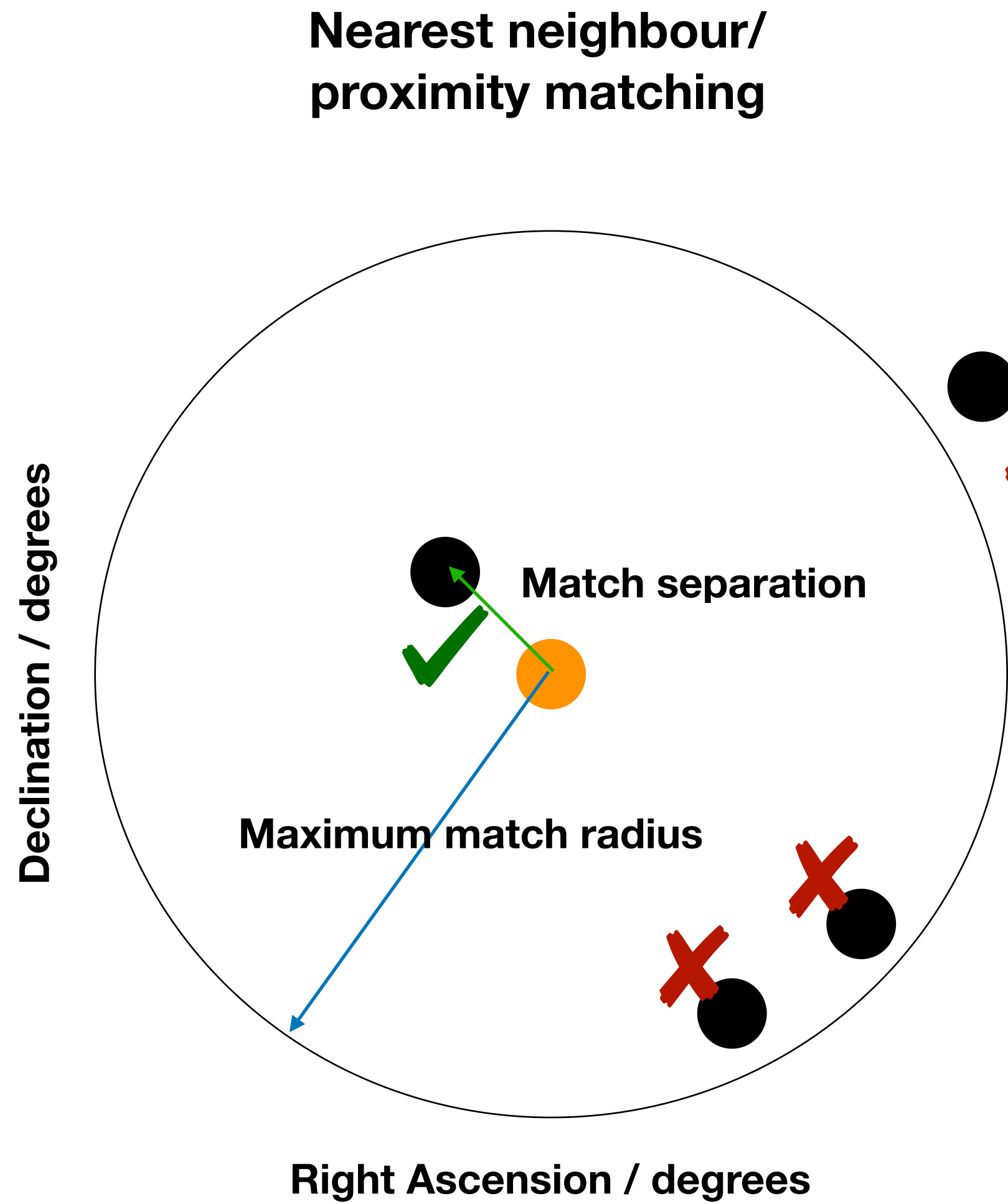
NOIRLab/NSF/AURA/F. Bruno



LSST/DOE

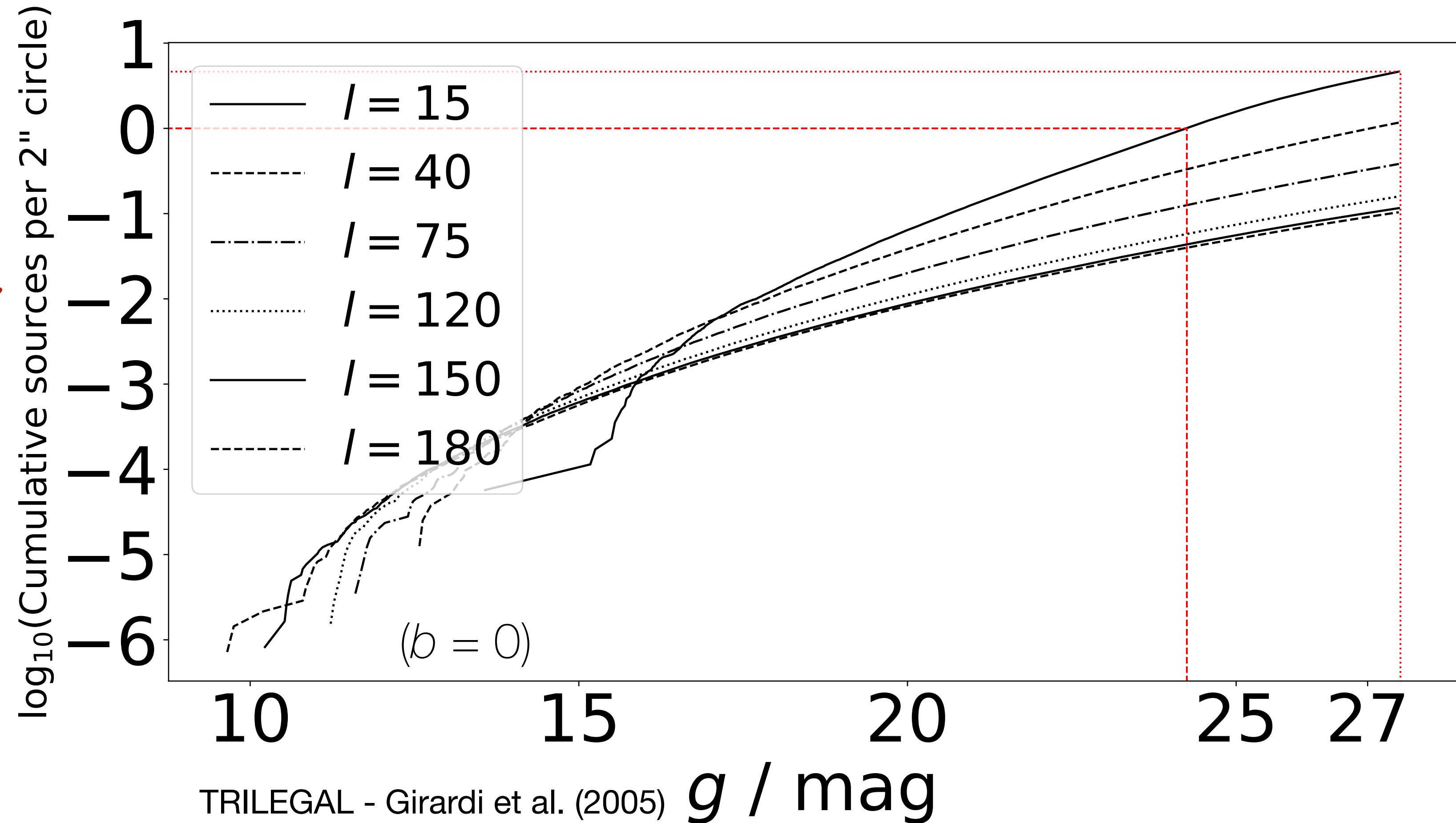
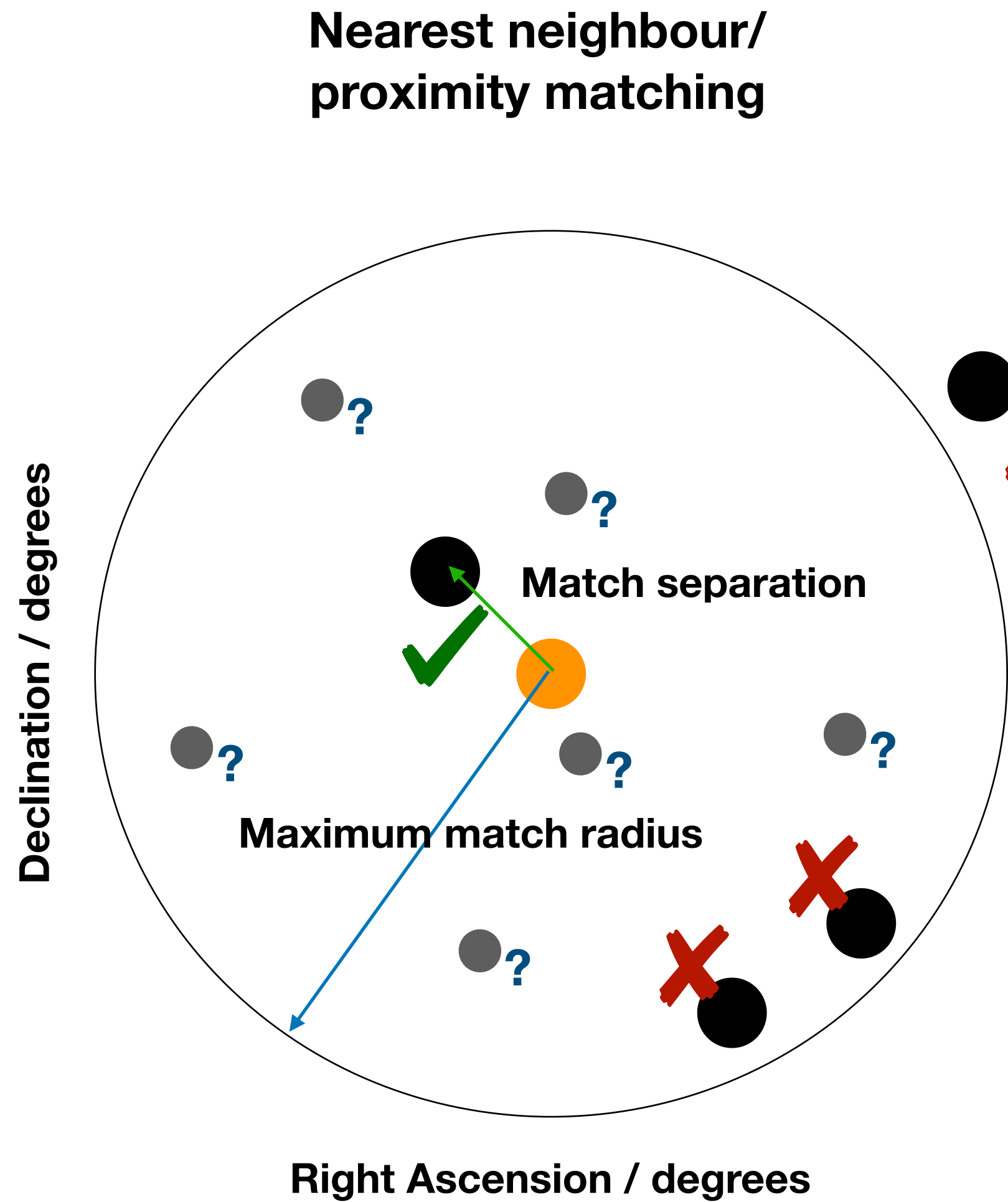
Tom J Wilson @onoddil

The Looming Problem With LSST



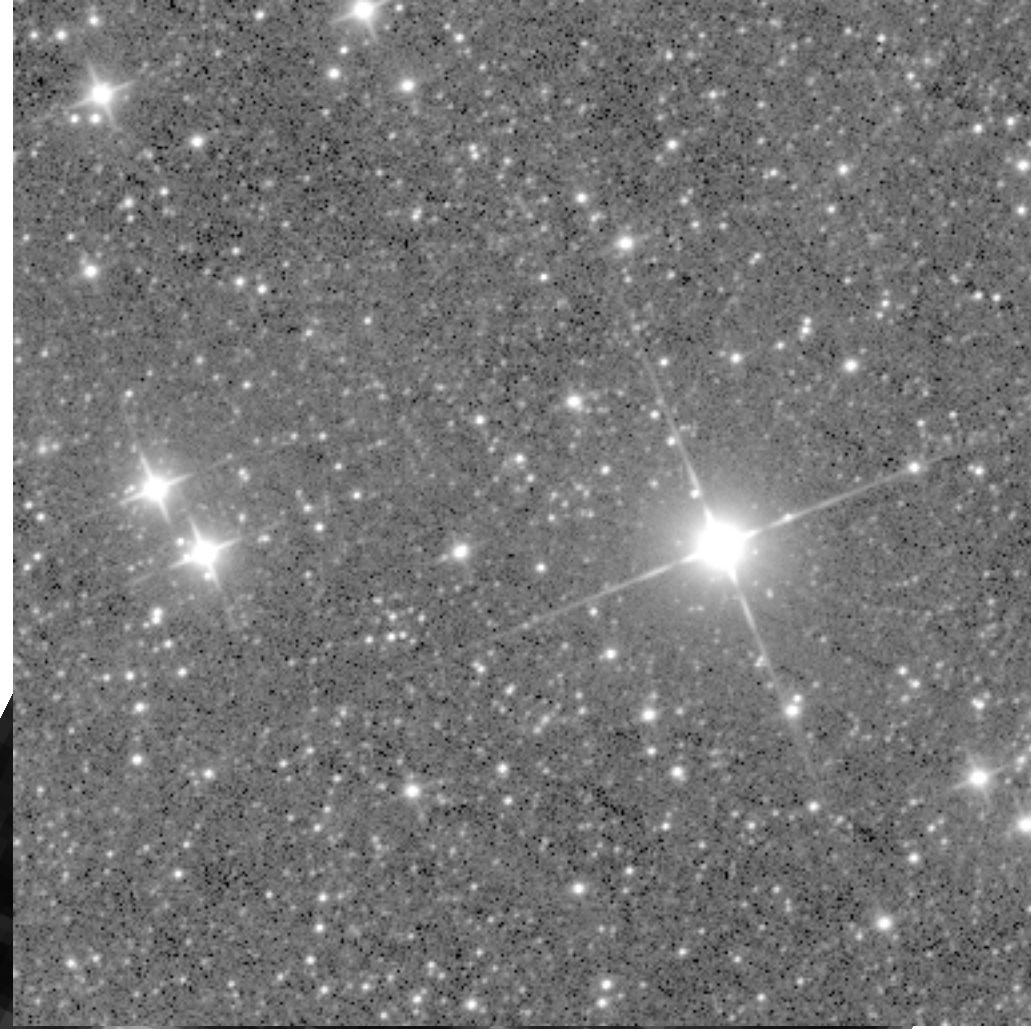
The Looming Problem With LSST

(It's 2-4 randomly placed objects in every match radius at high Galactic latitudes)



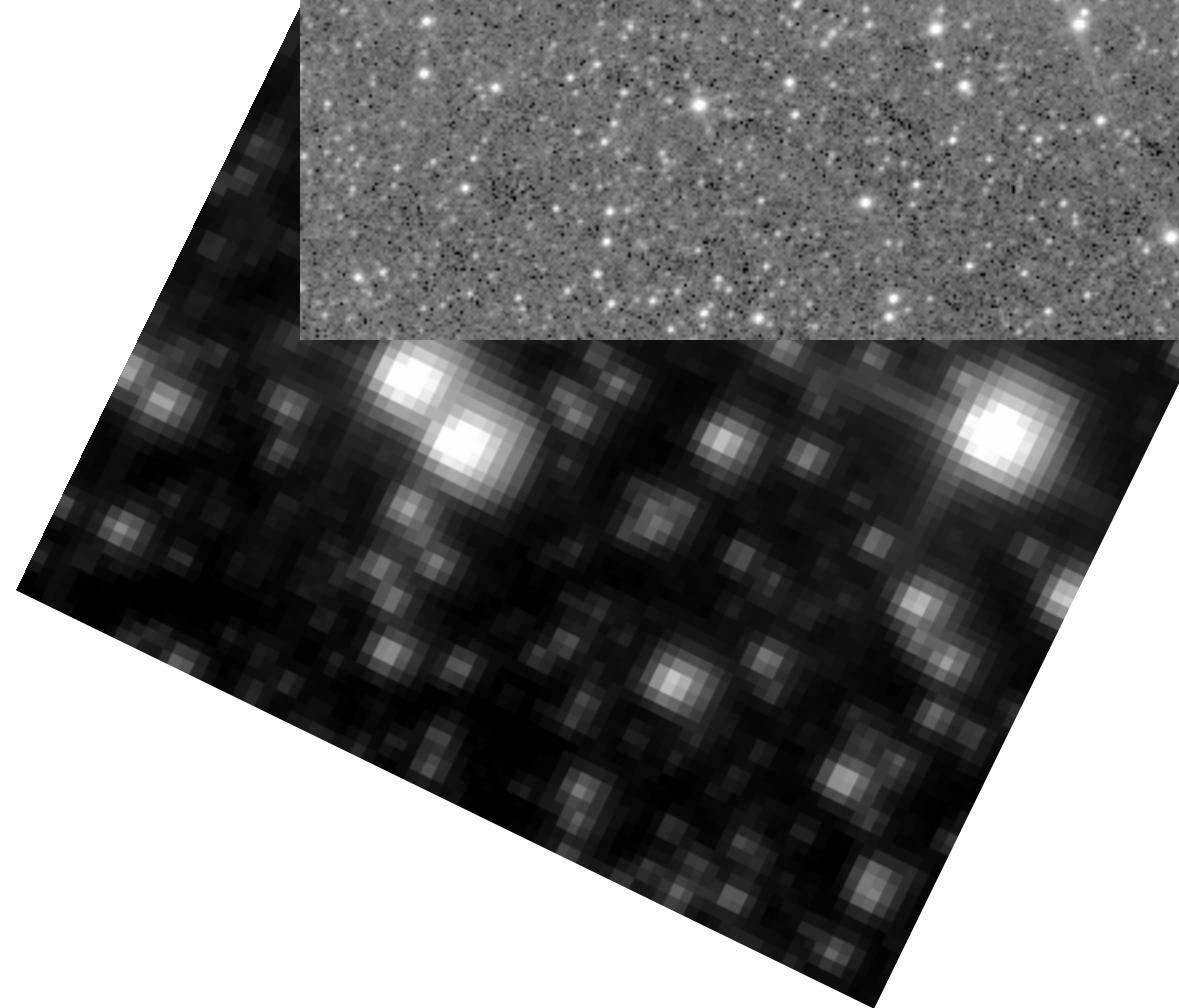
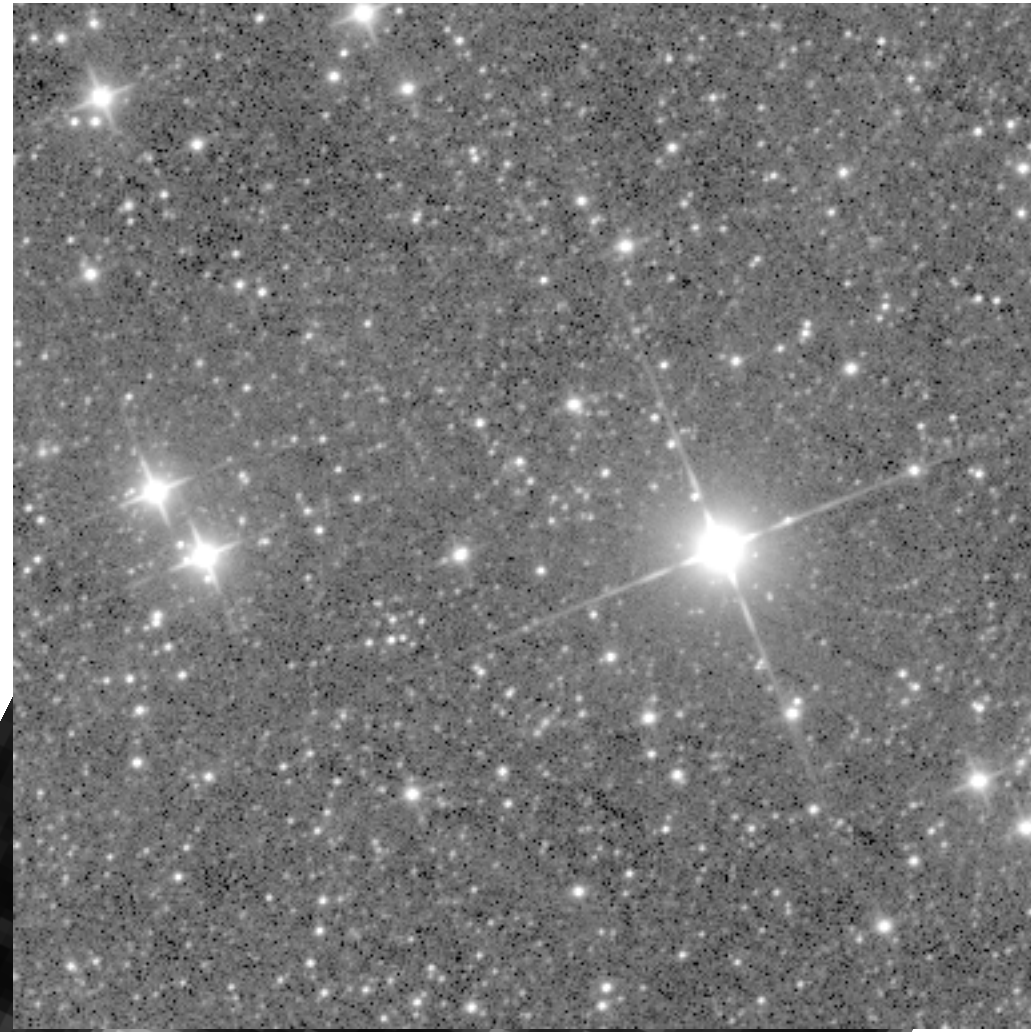
Nearest-neighbour matching *will not* work in the era of Rubin!

The Astronomy Error Function

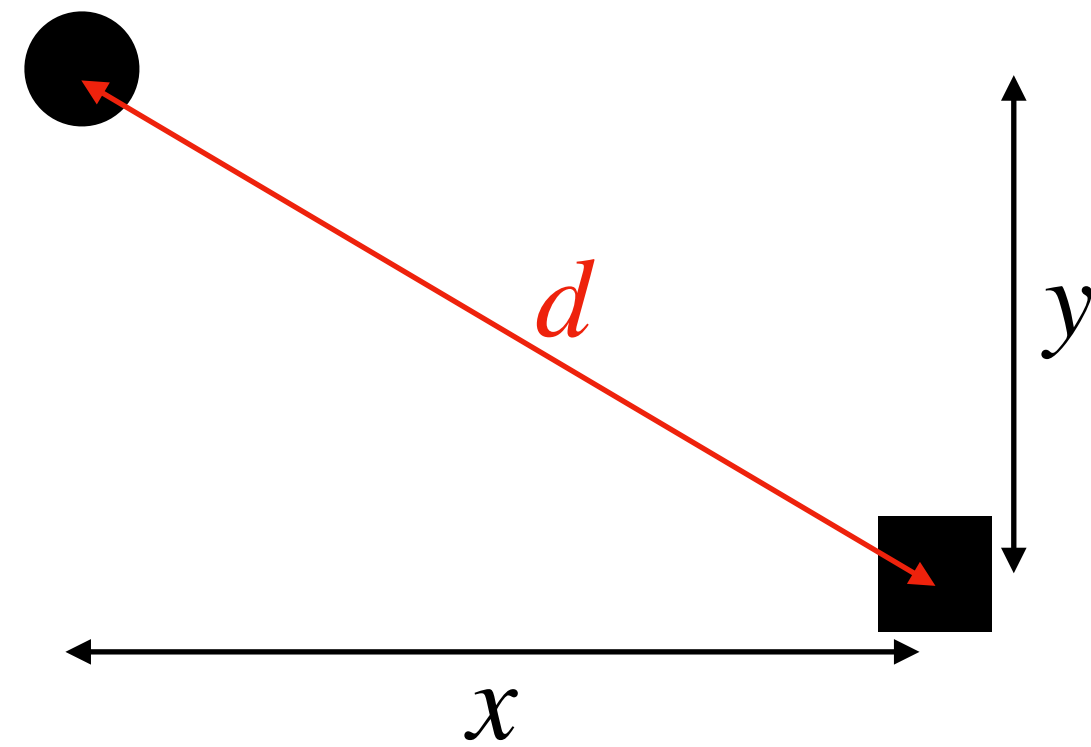


- 1) $p(x)$ decreases as x increases
- 2) $p(x \text{ and } y) = p(x)p(y)$
- 3) $p(x) = p(-x) \Rightarrow p(x^2)$

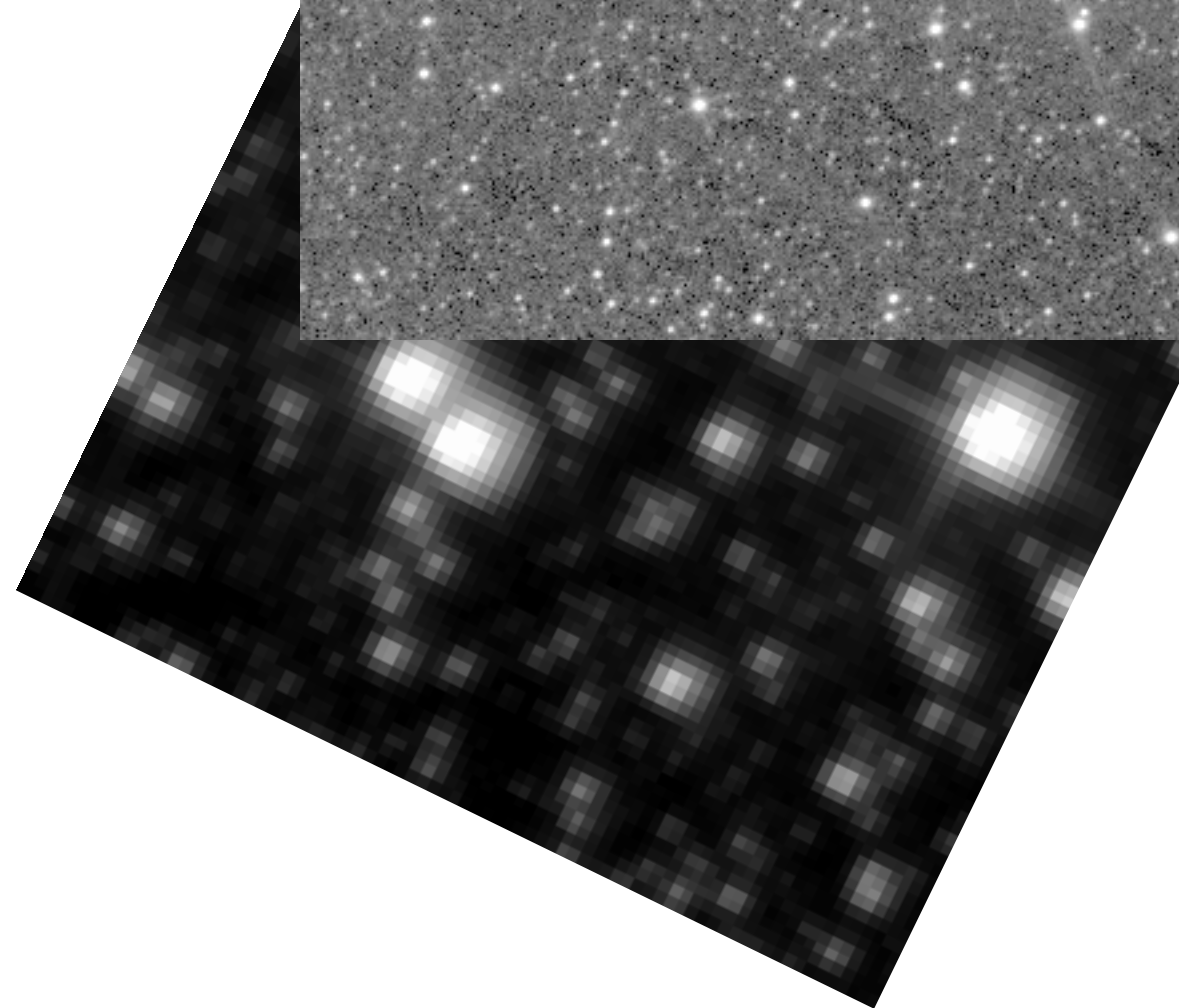
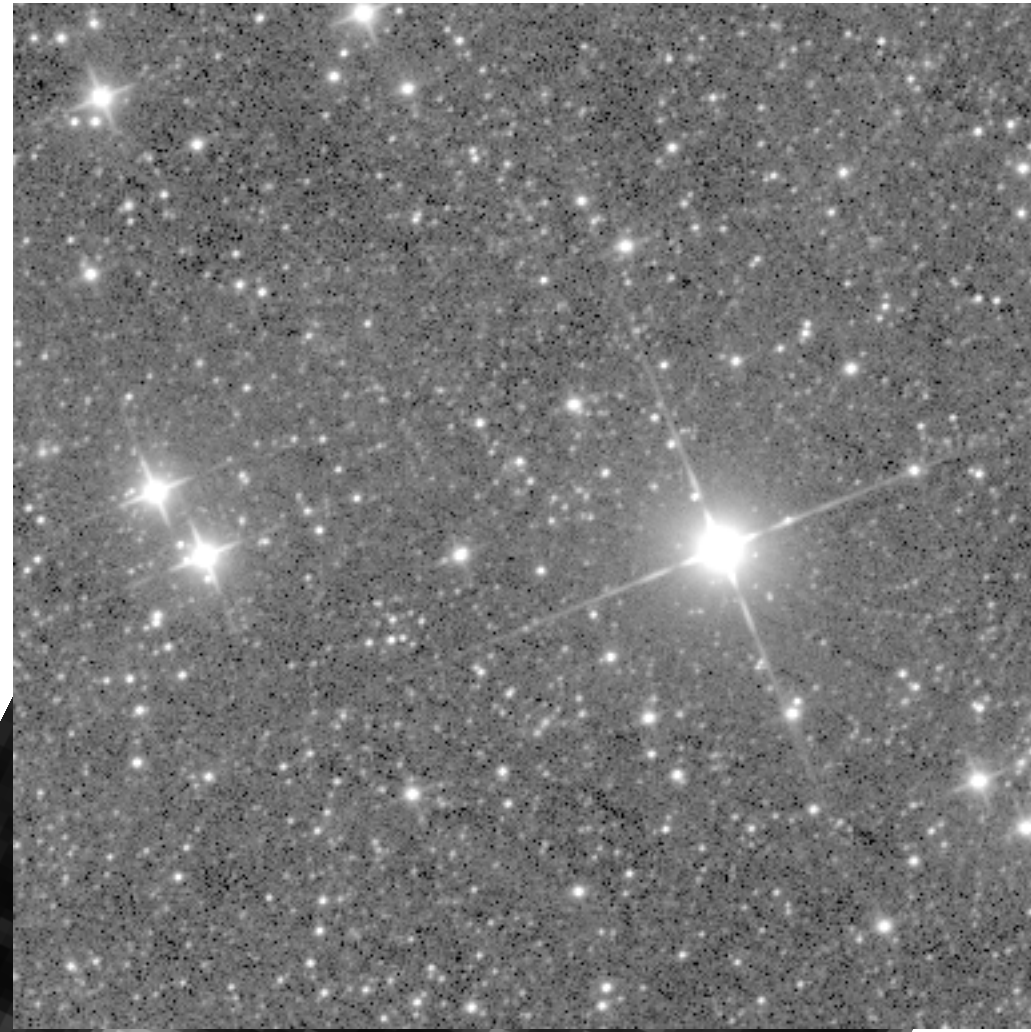
The Astronomy Error Function



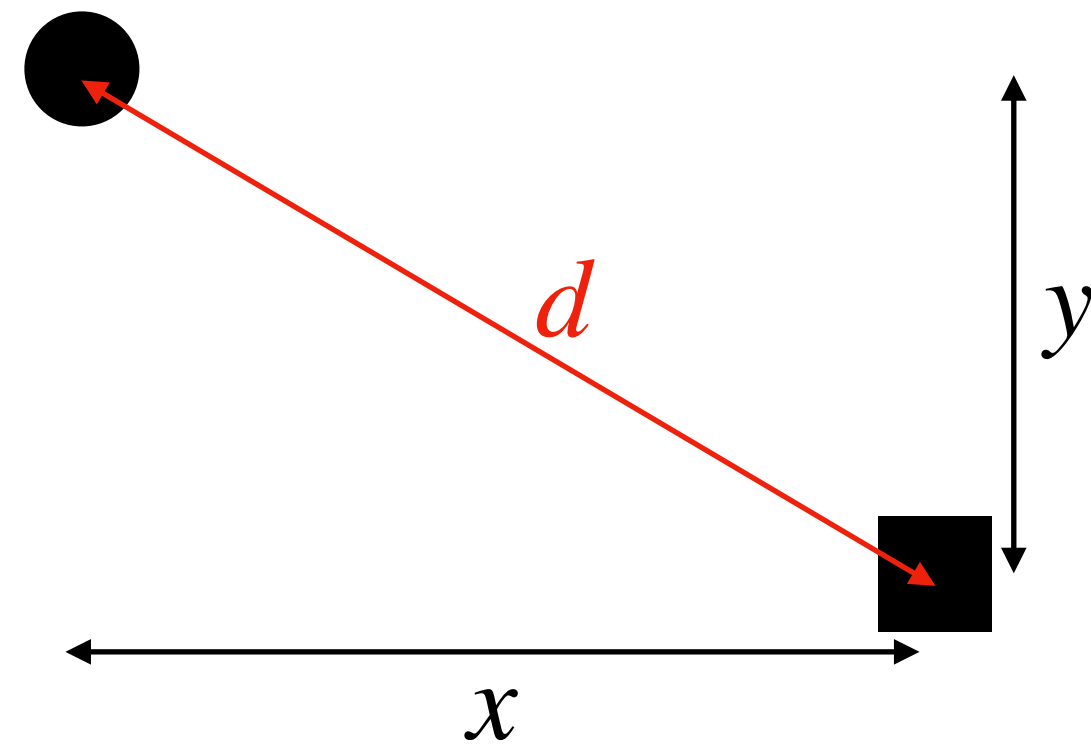
- 1) $p(x)$ decreases as x increases
- 2) $p(x \text{ and } y) = p(x)p(y)$
- 3) $p(x) = p(-x) \Rightarrow p(x^2)$



The Astronomy Error Function

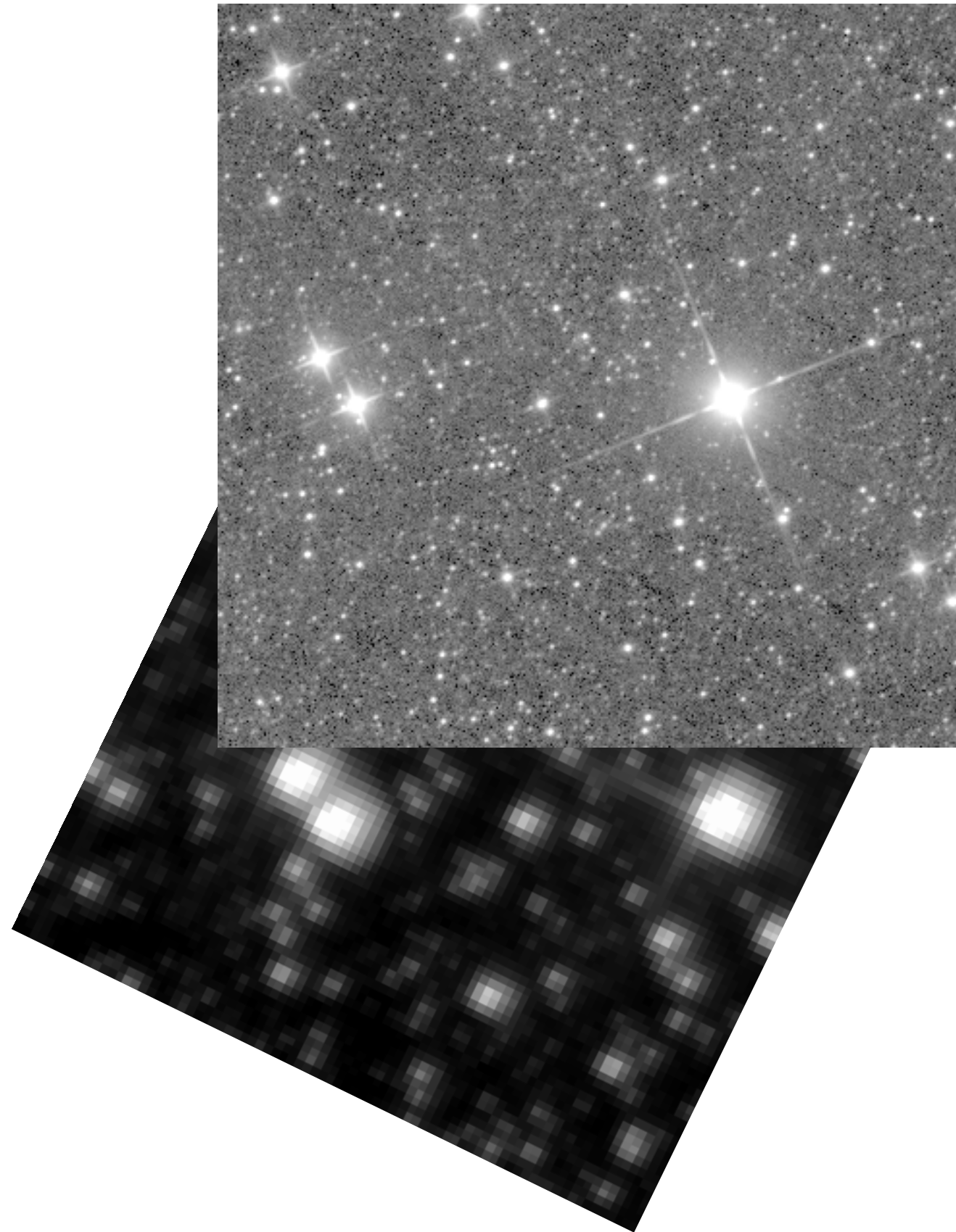


- 1) $p(x)$ decreases as x increases
- 2) $p(x \text{ and } y) = p(x)p(y)$
- 3) $p(x) = p(-x) \Rightarrow p(x^2)$

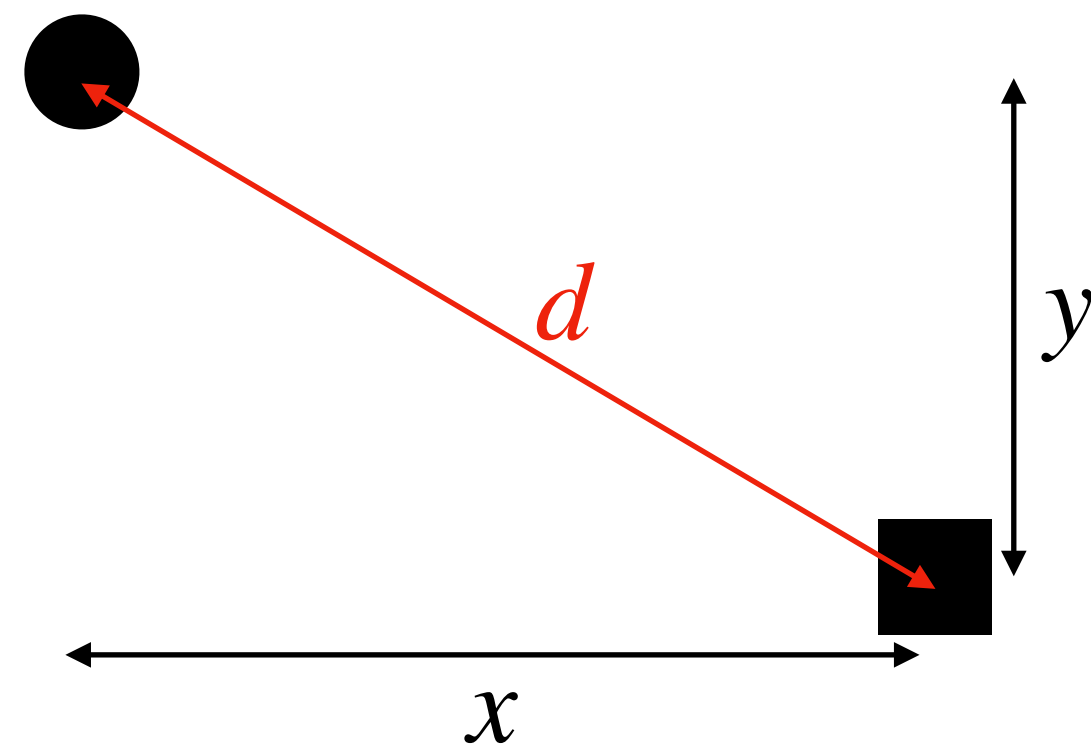


$$p(d^2) = p(x^2 + y^2) = p(x^2) p(y^2)$$

The Astronomy Error Function



- 1) $p(x)$ decreases as x increases
- 2) $p(x \text{ and } y) = p(x)p(y)$
- 3) $p(x) = p(-x) \Rightarrow p(x^2)$

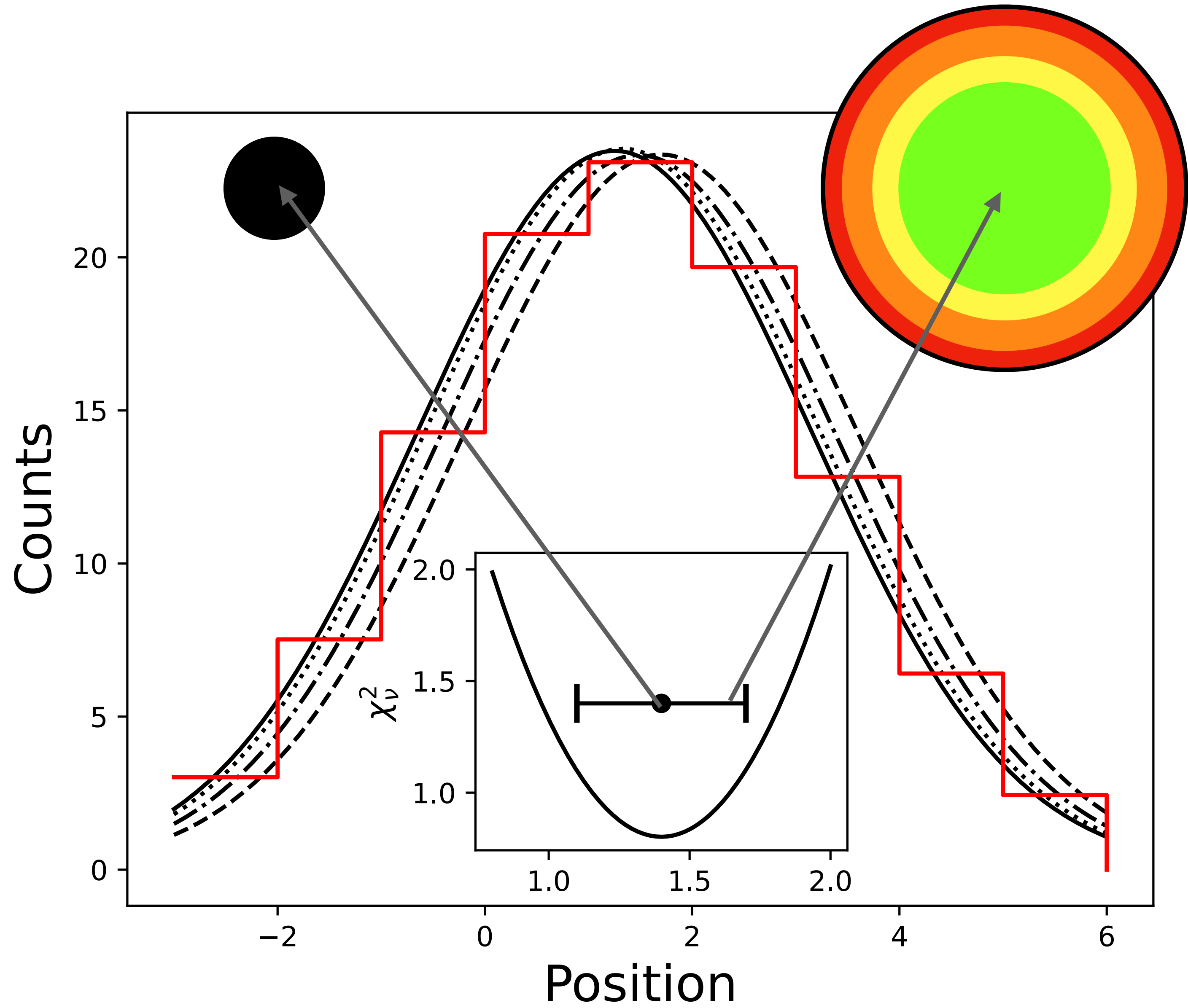


$$p(d^2) = p(x^2 + y^2) = p(x^2) p(y^2)$$

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

Nearest-neighbour match radius: $\sim 2''$; typical precision σ on source position: $\lesssim 0.2''$

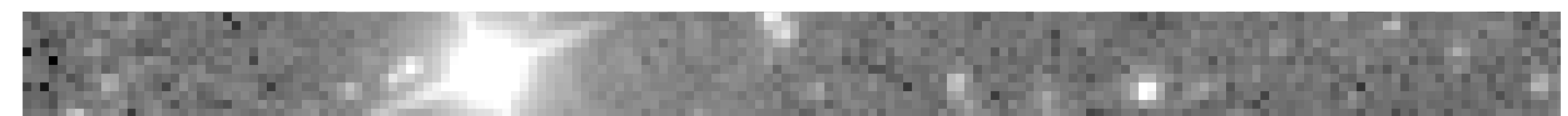
Centroid Positions and Uncertainties



$$p(D|M) \propto \frac{\exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})\right)}{\sqrt{(2\pi)^k |\boldsymbol{\Sigma}|}}$$

$$\mathbf{x} = \begin{pmatrix} x \\ y \end{pmatrix}, \boldsymbol{\mu} = \begin{pmatrix} \mu_x \\ \mu_y \end{pmatrix}, \boldsymbol{\Sigma} = \begin{pmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{pmatrix}$$

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



Probabilistic Cross-Matching

The Likelihood Ratio

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

$$dp(r|c) = 2\lambda r \times e^{-\lambda r^2} dr$$

$$LR(r) = dp(r|id)/dp(r|c) = \frac{1}{2\lambda} \exp\left\{\frac{r^2}{2}(2\lambda - 1)\right\}$$

de Ruiter, Willis, & Arp (1977)

$$dp_{id} = Qr \exp\left(\frac{-r^2}{2}\right) dr. \quad dp_{uo} = 2\lambda r dr$$

$$LR(r) = \frac{dp_{id}}{dp_{uo}} = \frac{Q \exp(-r^2/2)}{2\lambda}$$

Wolstencroft et al. (1986)

Probabilistic Cross-Matching

The Likelihood Ratio

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

$$dp(r|c) = 2\lambda r \times e^{-\lambda r^2} dr$$

$$LR(r) = dp(r|id)/dp(r|c) = \frac{1}{2\lambda} \exp\left\{\frac{r^2}{2}(2\lambda - 1)\right\}$$

de Ruiter, Willis, & Arp (1977)

$$dp_{id} = Qr \exp\left(\frac{-r^2}{2}\right) dr. \quad dp_{uo} = 2\lambda r dr$$

$$LR(r) = \frac{dp_{id}}{dp_{uo}} = \frac{Q \exp(-r^2/2)}{2\lambda}$$

Wolstencroft et al. (1986)

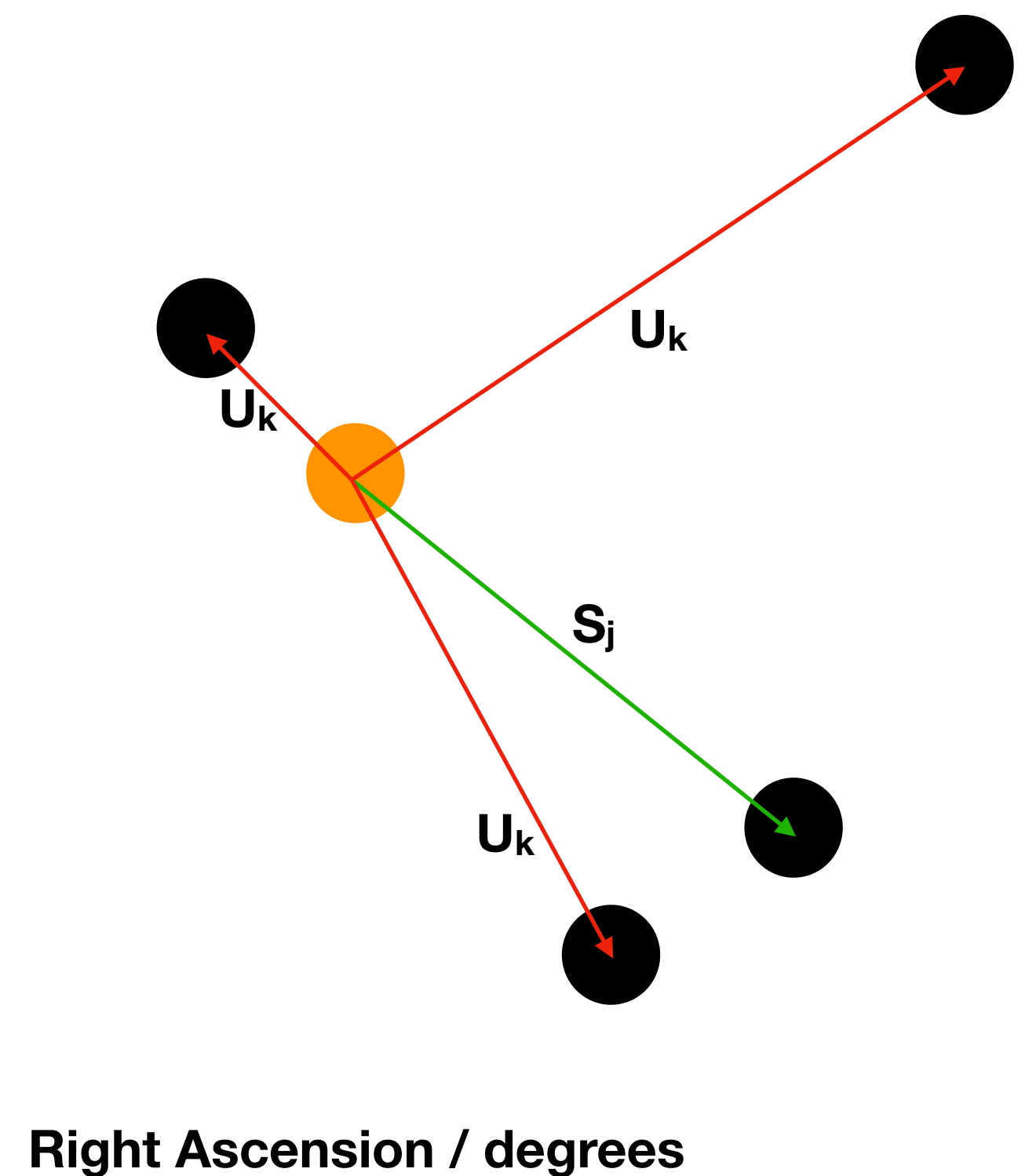
The "Reliability" – Sutherland & Saunders (1992)

$$\Pr\left[S_j \cap \left(\bigcap_{k \neq j} U_k\right) \cap \left(\bigcap_{k'} E_{k'}\right)\right]$$

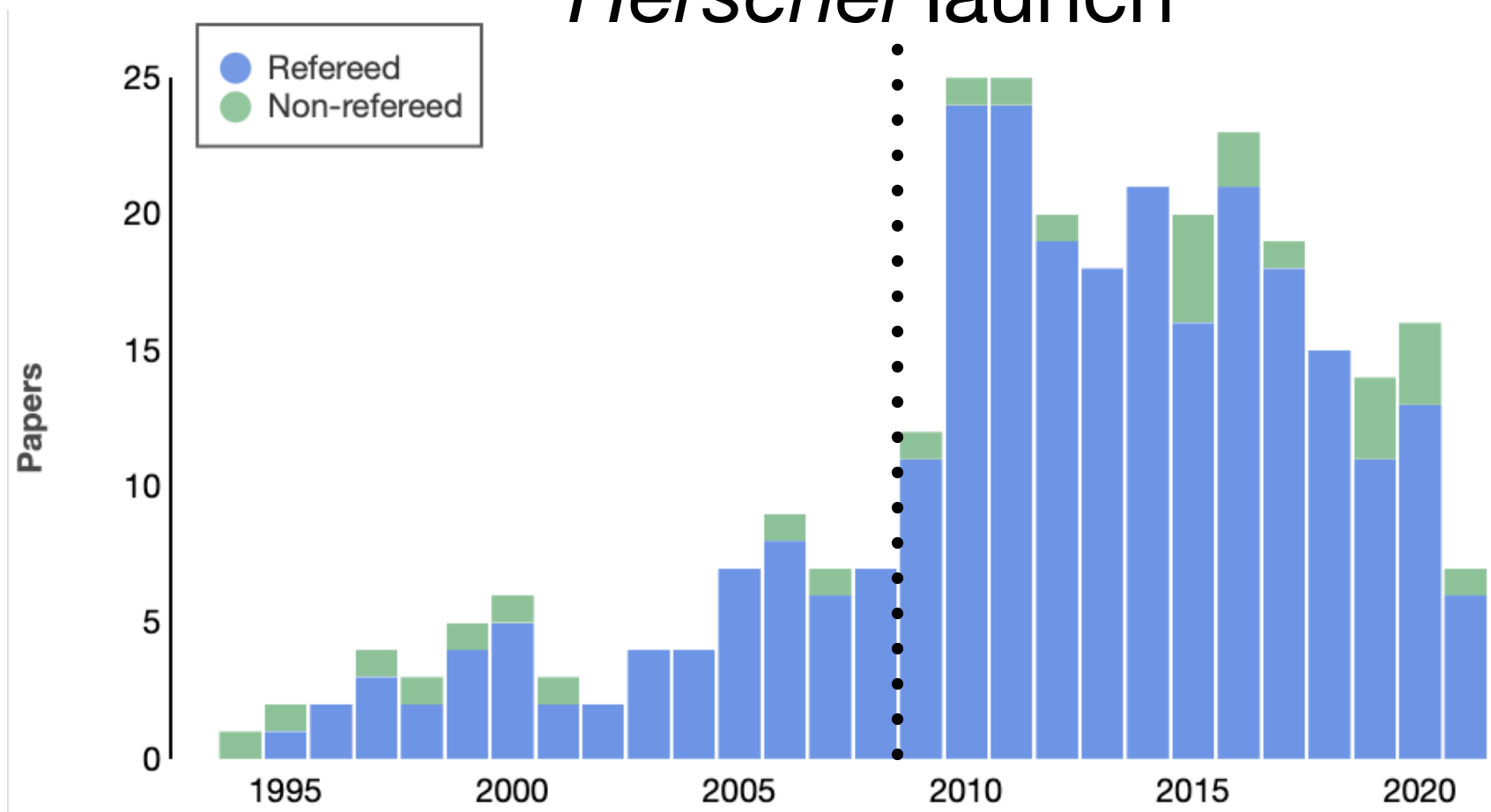
$$R_j = \frac{\Pr\left[S_j \cap \left(\bigcap_{k \neq j} U_k\right) \cap \left(\bigcap_{k'} E_{k'}\right)\right]}{\sum_i \Pr\left[S_i \cap \left(\bigcap_{k \neq i} U_k\right) \cap \left(\bigcap_{k'} E_{k'}\right)\right] + \Pr\left[(m_S > m_{lim}) \cap \left(\bigcap_k U_k\right) \cap \left(\bigcap_{k'} E_{k'}\right)\right]} = \frac{L_j}{\sum_i L_i + (1-Q)}$$

$$L = \frac{q(m, c) f(x, y)}{n(m, c)}$$

Declination / degrees

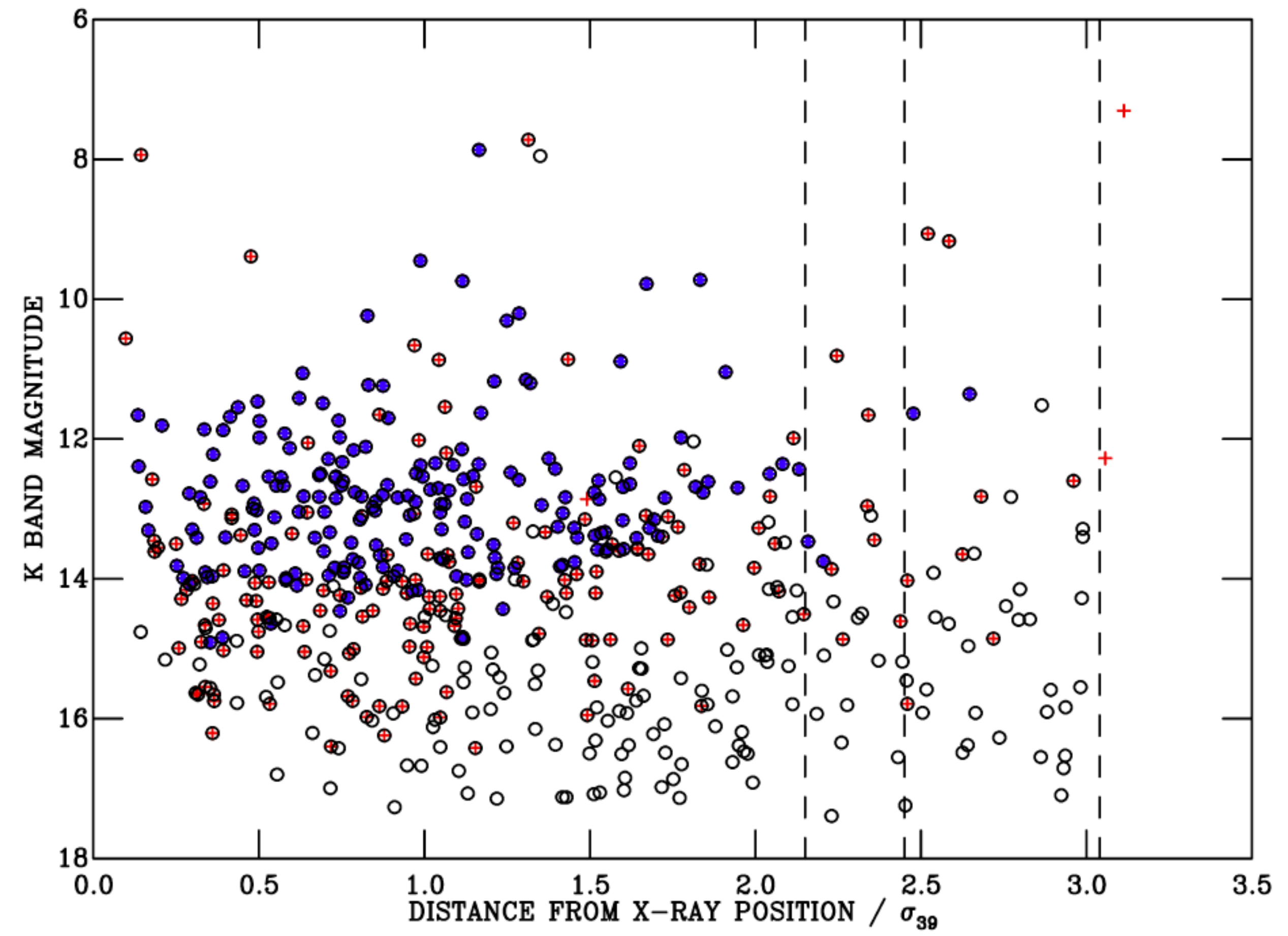
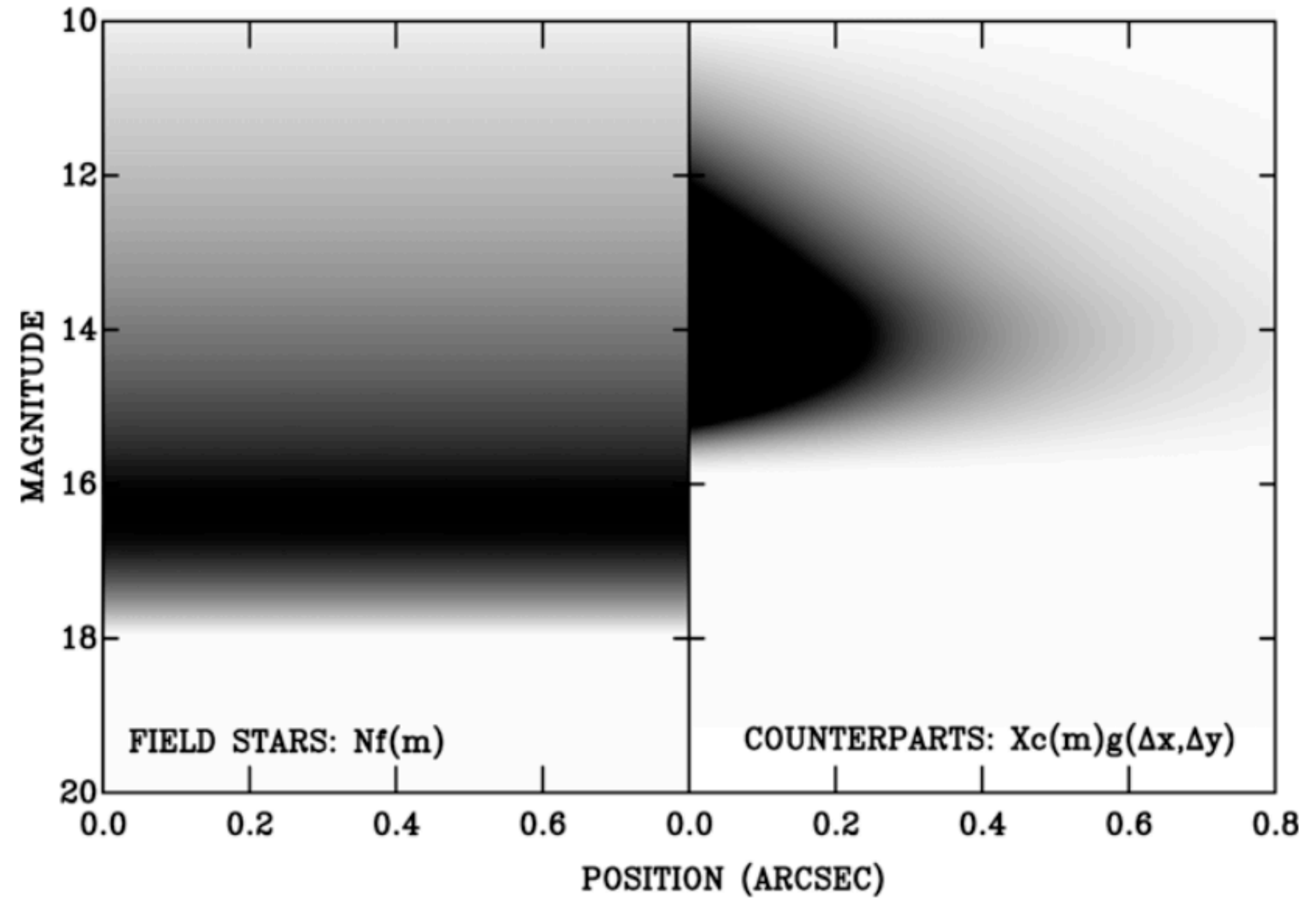


Herschel launch



Probabilistic Cross-Matching

$$P(0) = \frac{1 - X}{1 - X + \sum_j \frac{Xc(m_j)g(\Delta x_j, \Delta y_j)}{Nf(m_j)}} \quad P(i) = \frac{\frac{Xc(m_i)g(\Delta x_i, \Delta y_i)}{Nf(m_i)}}{1 - X + \sum_j \frac{Xc(m_j)g(\Delta x_j, \Delta y_j)}{Nf(m_j)}}$$



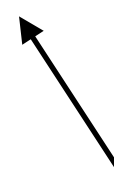
Probabilistic Cross-Matching

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

$$p(D|K) = \prod_{i=1}^n \left[\int p(\mathbf{m}_i|K) p_i(\mathbf{x}_i|\mathbf{m}_i, K) d^3 m_i \right]$$

$$B(H, K|D') = \frac{\int p(\boldsymbol{\eta}|H) \prod_{i=1}^n p_i(\mathbf{g}_i|\boldsymbol{\eta}, H) d^r \boldsymbol{\eta}}{\prod_{i=1}^n \left[\int p(\boldsymbol{\eta}_i|K) p_i(\mathbf{g}_i|\boldsymbol{\eta}_i, K) d^r \boldsymbol{\eta}_i \right]}$$

Budavári & Szalay (2008)



Includes SED model fitting to all sources

Probabilistic Cross-Matching

Nearest neighbour or brightest neighbour: one-to-one, either astrometry OR photometry

Likelihood ratio: one-to-one matches, mostly just astrometry (e.g., Wolstencroft et al. 1986)

Reliability: One-to-many matches, uses photometry from one dataset (e.g. Naylor et al. 2013)

Budavári & Szalay (2008): one-to-one-to-one-to... matches, include SED fitting

e.g. Pineau et al. (2017): many-to-many-to-many-to... matches, no photometry implemented

Probabilistic Cross-Matching

Nearest neighbour or brightest neighbour: one-to-one, either astrometry OR photometry

Likelihood ratio: one-to-one matches, mostly just astrometry (e.g., Wolstencroft et al. 1986)

Reliability: One-to-many matches, uses photometry from one dataset (e.g. Naylor et al. 2013)

Budavári & Szalay (2008): one-to-one-to-one-to... matches, include SED fitting

e.g. Pineau et al. (2017): many-to-many-to-many-to... matches, no photometry implemented

One assumption made in all of these works: source positions uncertainties are Gaussian!

$$dp(r|id) = r \times e^{-r^2/2} dr. \quad P(i) = \frac{\frac{Xc(m_i) g(\Delta x_i, \Delta y_i)}{Nf(m_i)}}{1 - X + \sum_j \frac{Xc(m_j) g(\Delta x_j, \Delta y_j)}{Nf(m_j)}} \quad p(D|H) = \int p(\mathbf{m}|H) \prod_{i=1}^n p_i(\mathbf{x}_i|\mathbf{m}, H) d^3 m$$

Probabilistic Cross-Matching

Nearest neighbour or brightest neighbour: one-to-one, either astrometry OR photometry

Likelihood ratio: one-to-one matches, mostly just astrometry (e.g., Wolstencroft et al. 1986)

Reliability: One-to-many matches, uses photometry from one dataset (e.g. Naylor et al. 2013)

Budavári & Szalay (2008): one-to-one-to-one-to... matches, include SED fitting

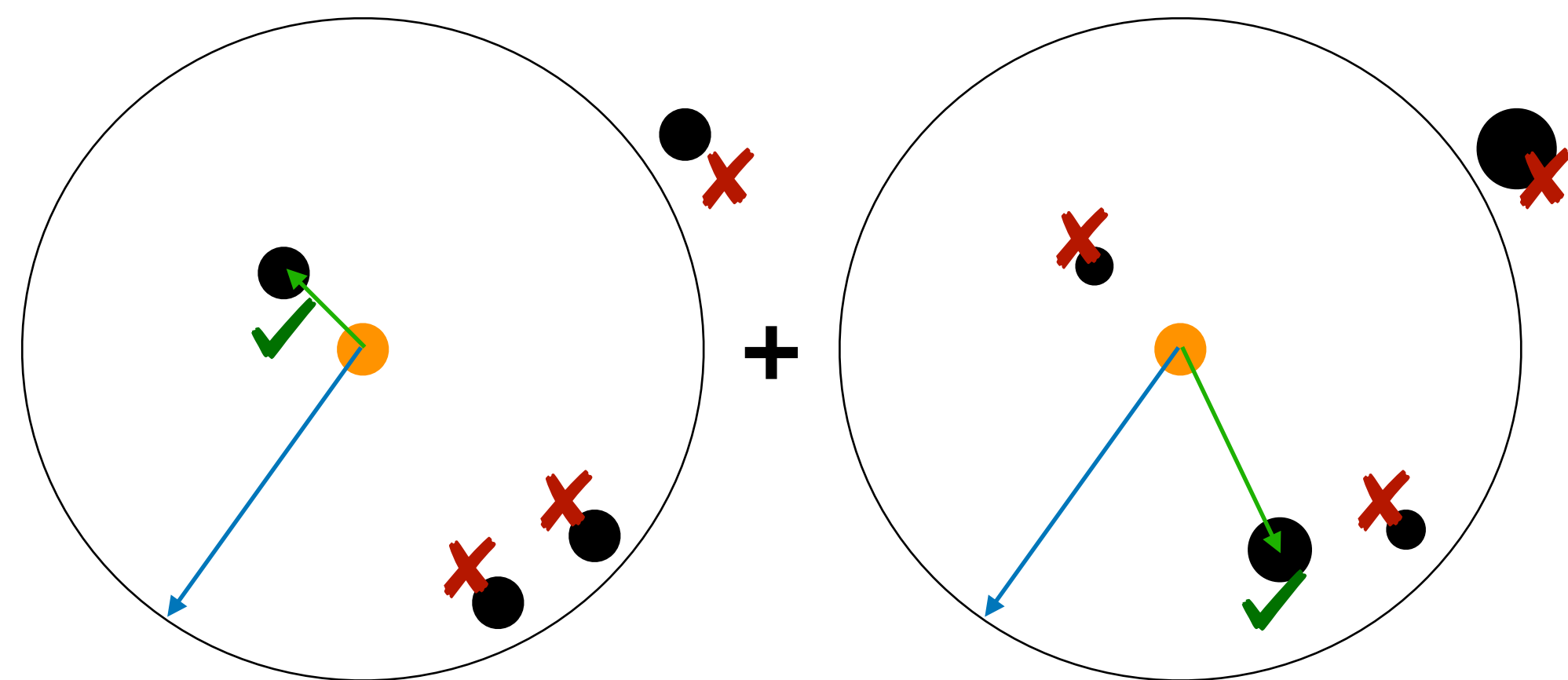
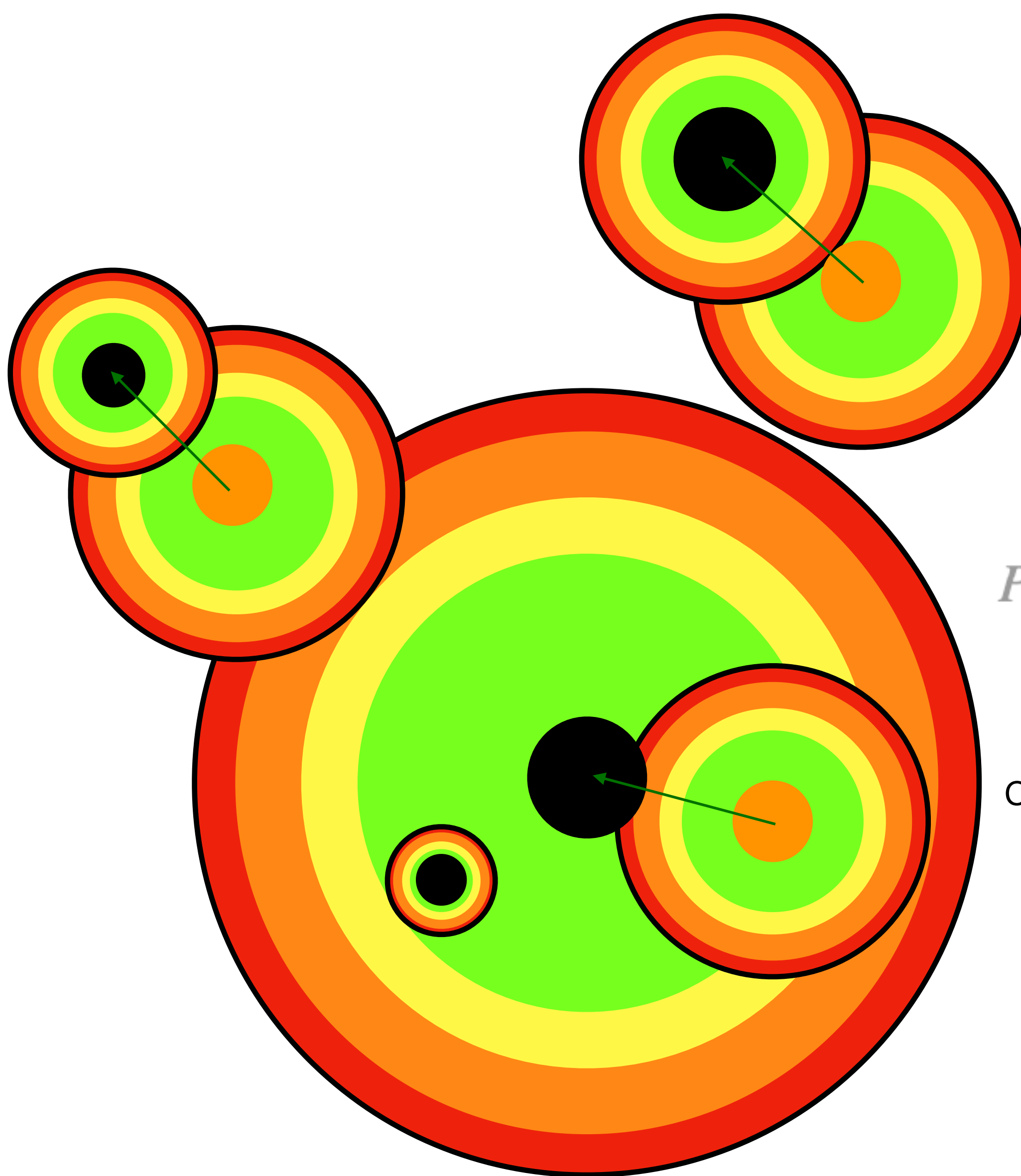
e.g. Pineau et al. (2017): many-to-many-to-many-to... matches, no photometry implemented

Wilson & Naylor (2017, 2018a, 2018b), Wilson (2022, 2023), and so on:

a many-to-many match with data-driven photometric probabilities

also describes a flexible approach to positional “errors,” vital for future surveys

Probabilistic Cross-Matching



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

G includes information on position (un)certainty

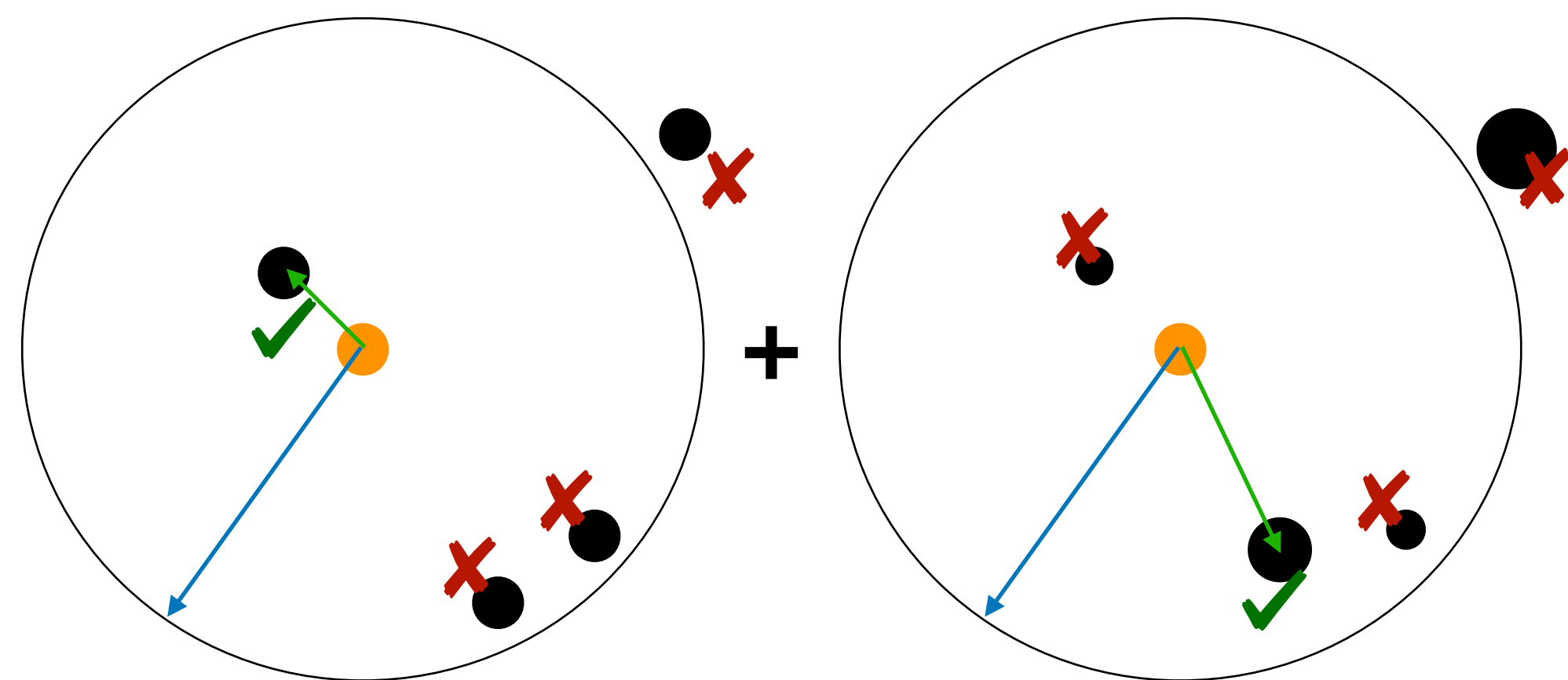
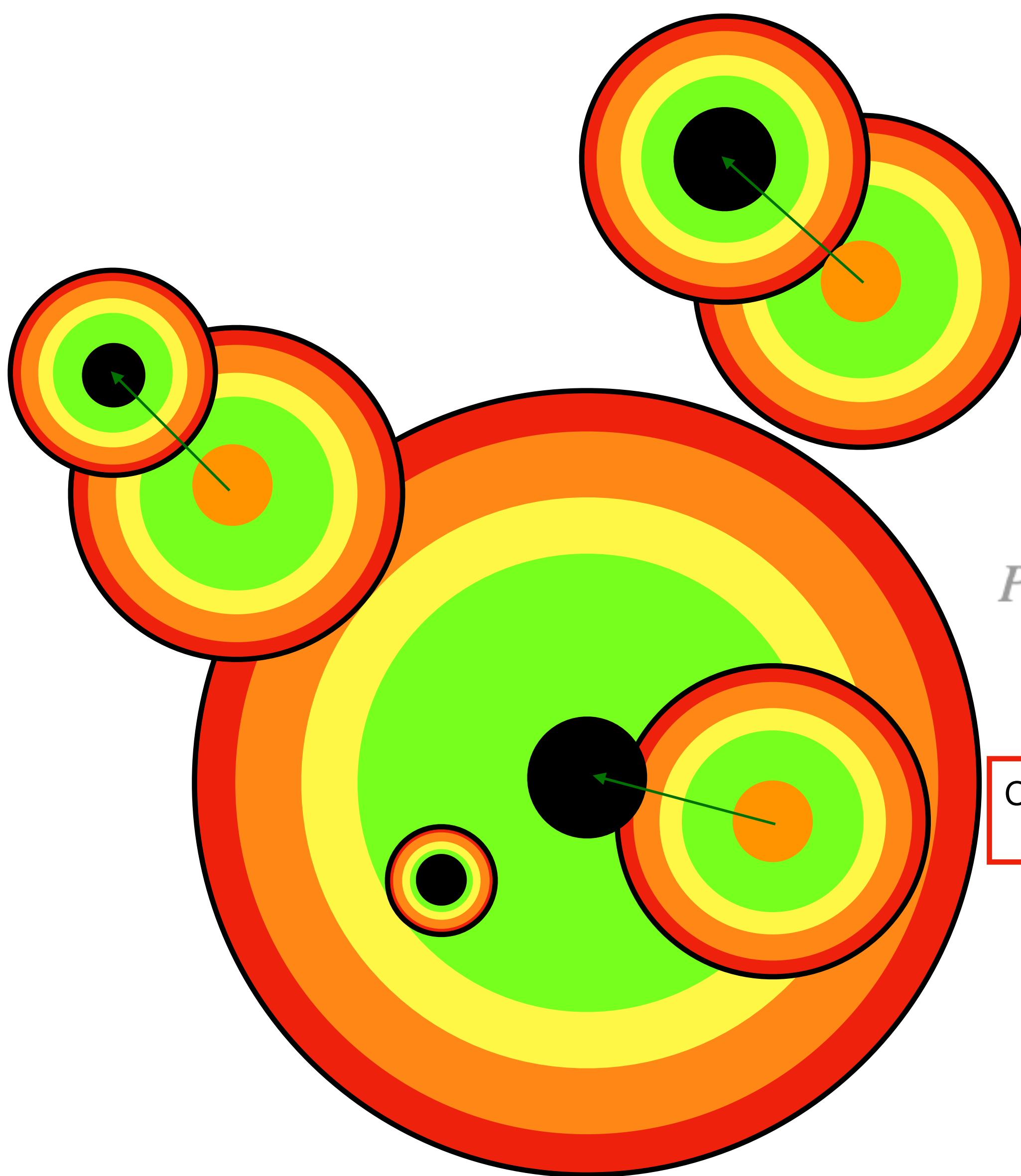
$$P(\zeta, \lambda, k | \gamma, \phi) = \frac{1}{K} \times \prod_{\delta \notin \zeta \cap \delta \in \gamma} N_\gamma f_\gamma^\delta \prod_{\omega \notin \lambda \cap \omega \in \phi} N_\phi f_\phi^\omega \prod_{i=1}^k N_c G_{\gamma\phi}^{\zeta_i \lambda_i} c_{\gamma\phi}^{\zeta_i \lambda_i}$$

Consider multiple objects in both catalogues simultaneously

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

Probability of sources having their brightnesses given they are counterparts

Probabilistic Cross-Matching



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

G includes information on position (un)certainty

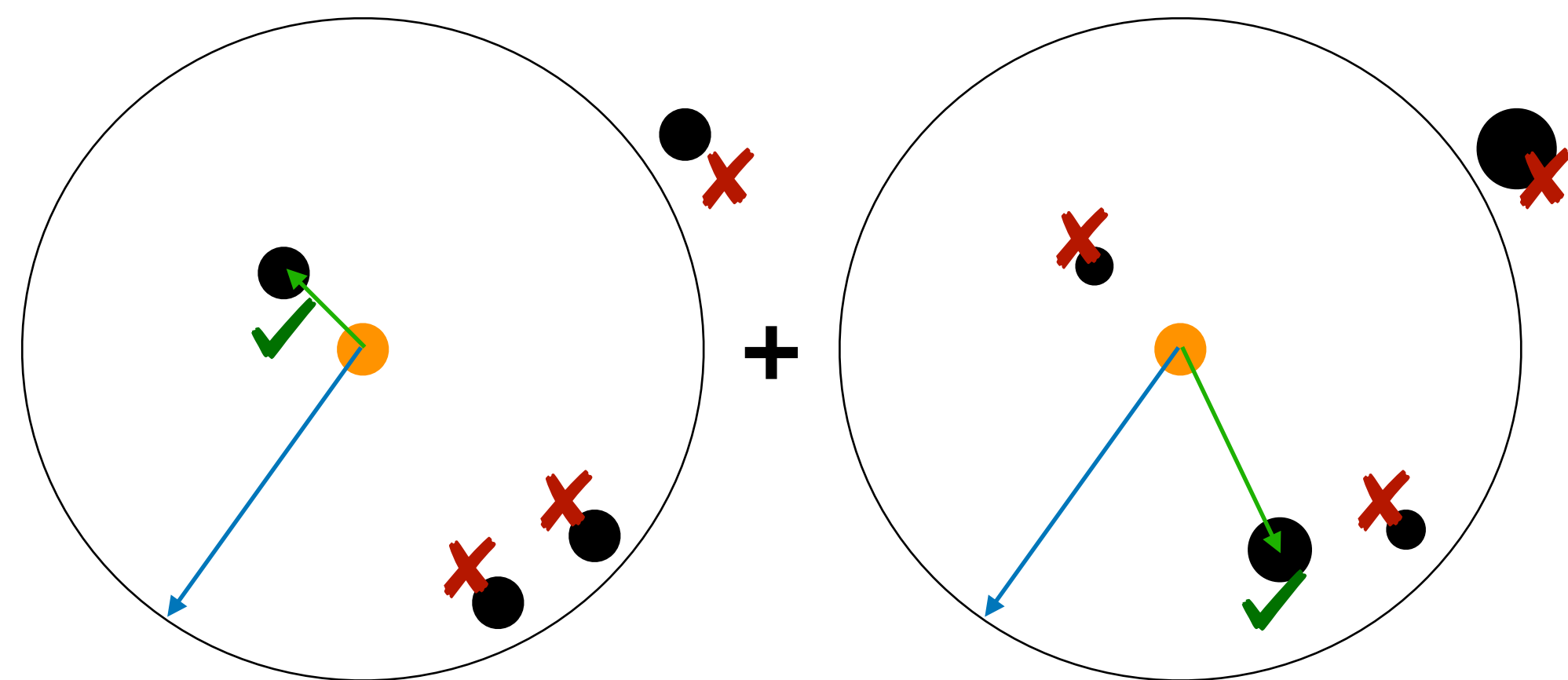
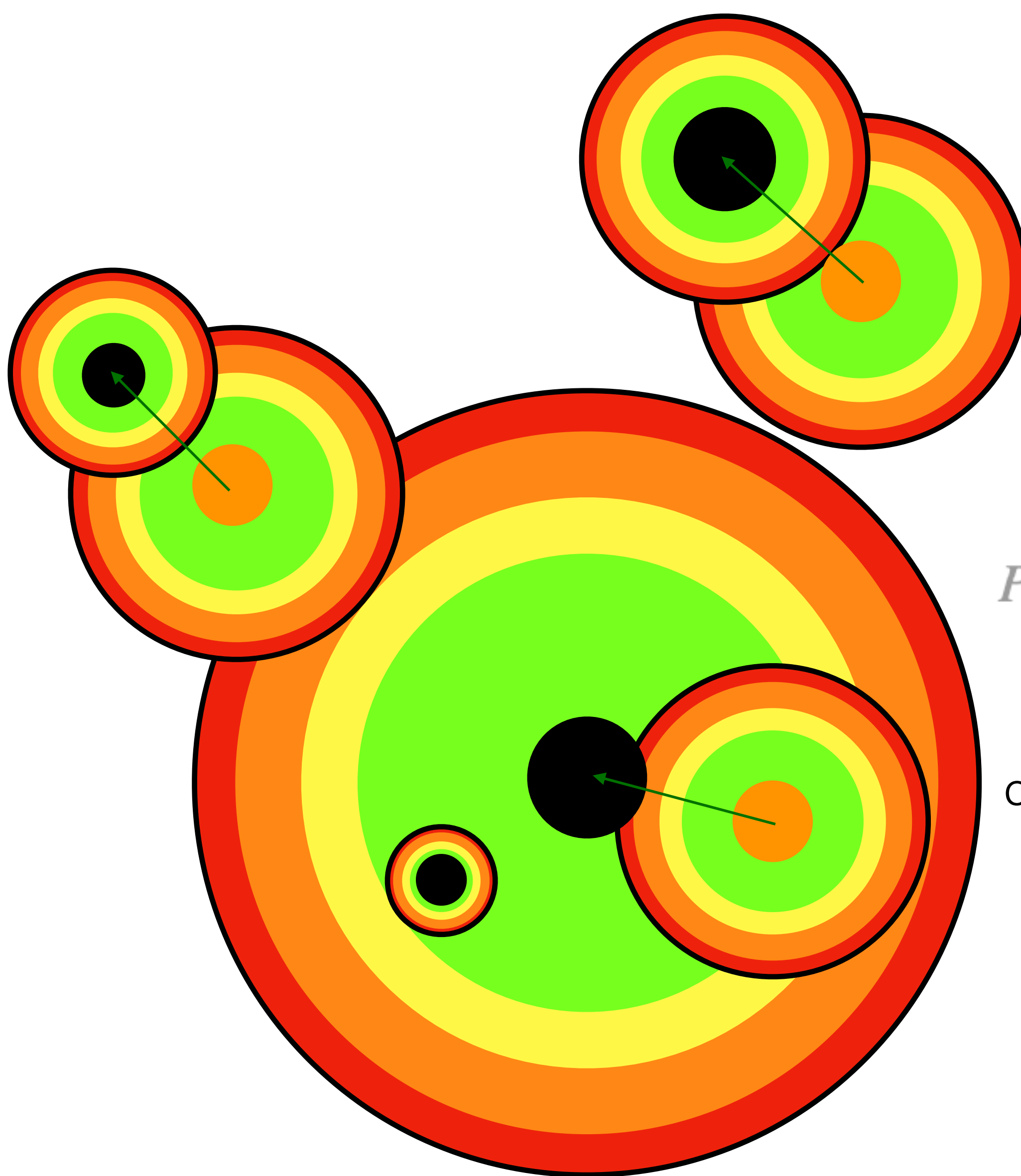
$$P(\zeta, \lambda, k | \gamma, \phi) = \frac{1}{K} \times \prod_{\delta \notin \zeta \cap \delta \in \gamma} N_{\gamma} f_{\gamma}^{\delta} \prod_{\omega \notin \lambda \cap \omega \in \phi} N_{\phi} f_{\phi}^{\omega} \prod_{i=1}^k N_c G_{\gamma\phi}^{\zeta_i \lambda_i} c_{\gamma\phi}^{\zeta_i \lambda_i}$$

Consider multiple objects in both catalogues simultaneously

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

Probability of sources having their brightnesses given they are counterparts

Probabilistic Cross-Matching



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

G includes information on position (un)certainty

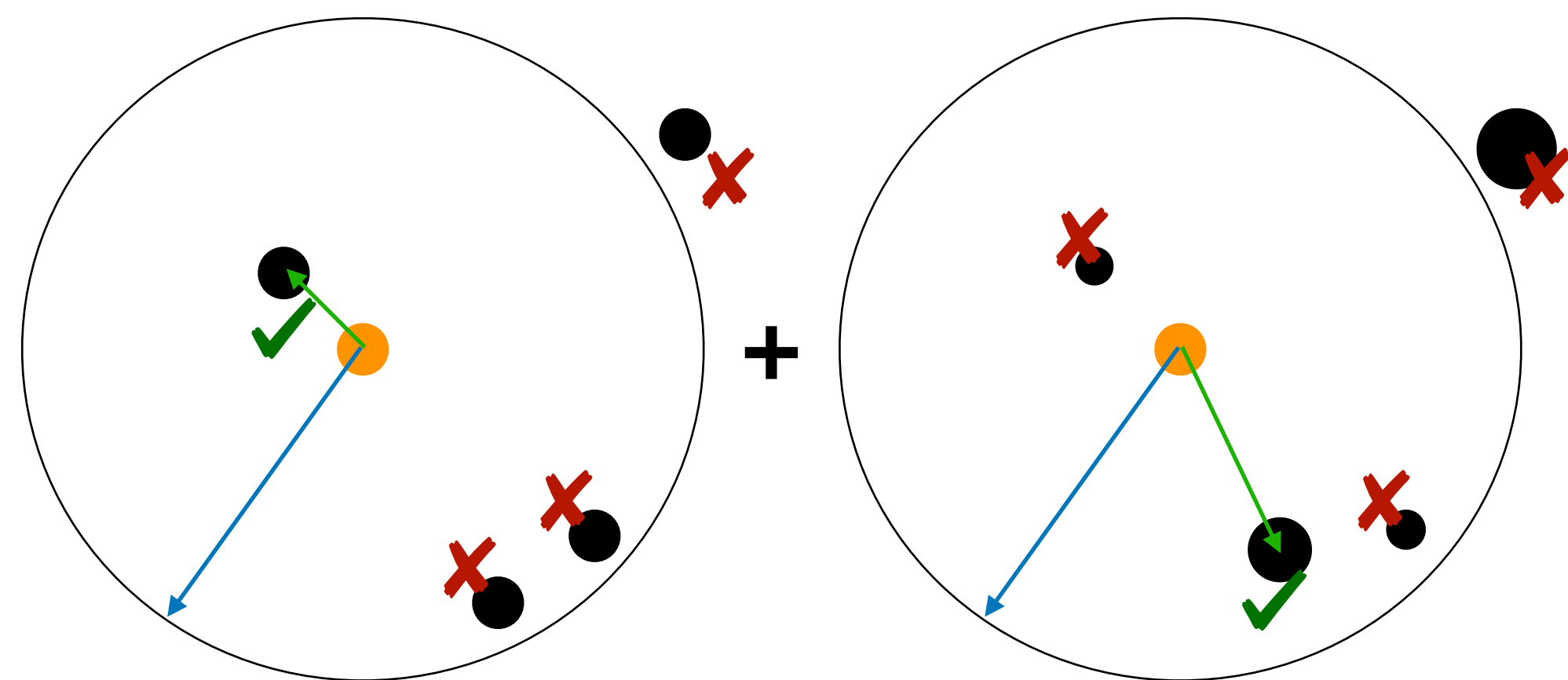
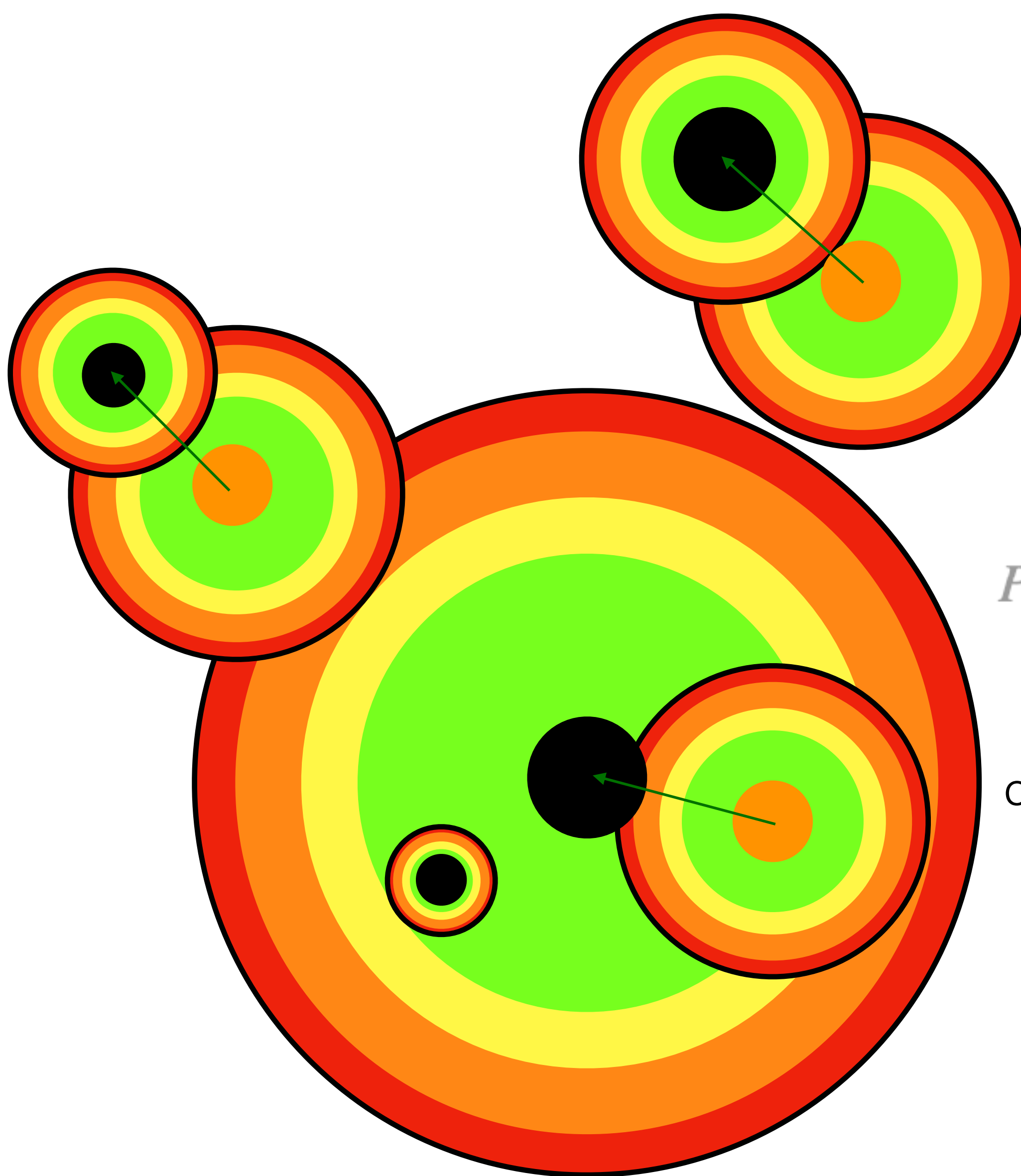
$$P(\zeta, \lambda, k | \gamma, \phi) = \frac{1}{K} \times \prod_{\delta \notin \zeta \cap \delta \in \gamma} N_\gamma f_\gamma^\delta \prod_{\omega \notin \lambda \cap \omega \in \phi} N_\phi f_\phi^\omega \prod_{i=1}^k N_c G_{\gamma\phi}^{\zeta_i \lambda_i} c_{\gamma\phi}^{\zeta_i \lambda_i}$$

Consider multiple objects in both catalogues simultaneously

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

Probability of sources having their brightnesses given they are counterparts

Probabilistic Cross-Matching



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

G includes information on position (un)certainty

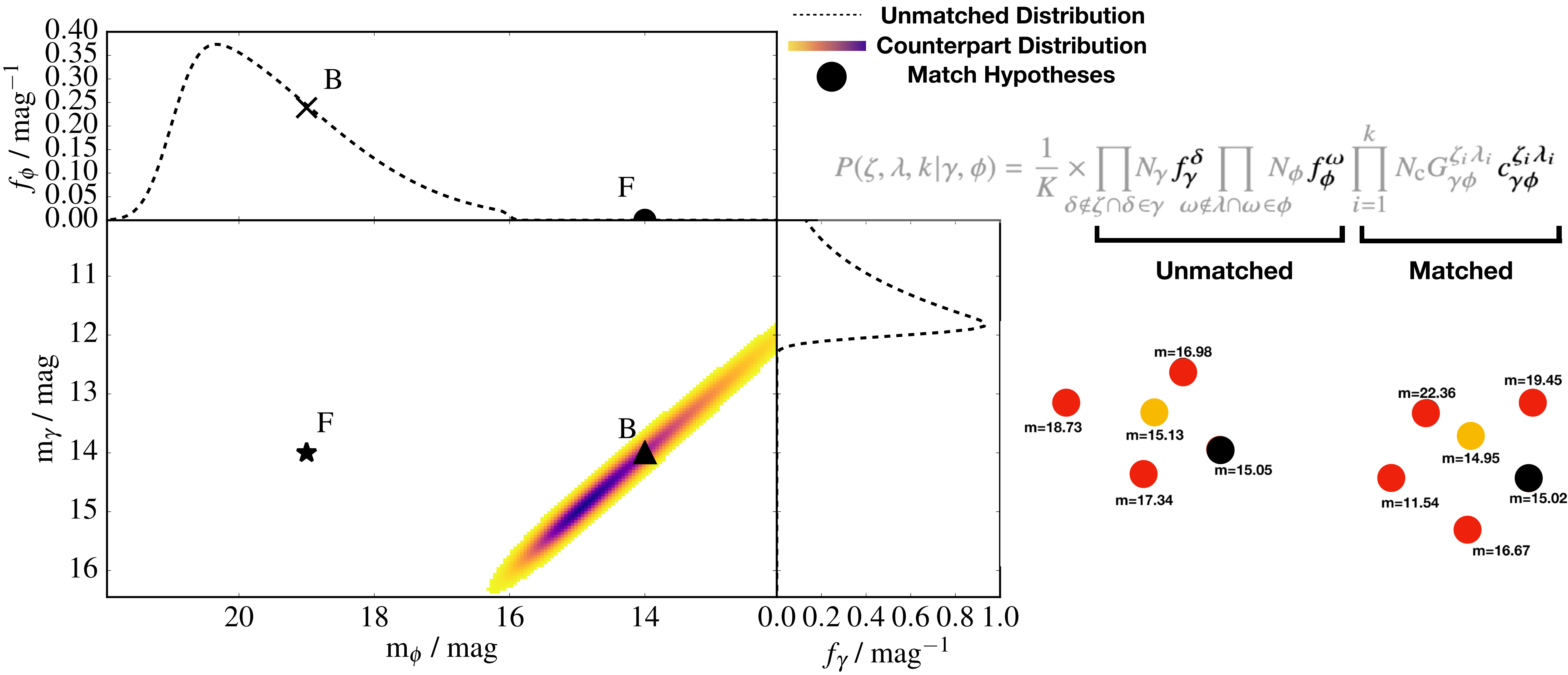
$$P(\zeta, \lambda, k | \gamma, \phi) = \frac{1}{K} \times \prod_{\delta \notin \zeta \cap \delta \in \gamma} N_\gamma f_\gamma^\delta \prod_{\omega \notin \lambda \cap \omega \in \phi} N_\phi f_\phi^\omega \prod_{i=1}^k N_c G_{\gamma\phi}^{\zeta_i \lambda_i} c_{\gamma\phi}^{\zeta_i \lambda_i}$$

Consider multiple objects in both catalogues simultaneously

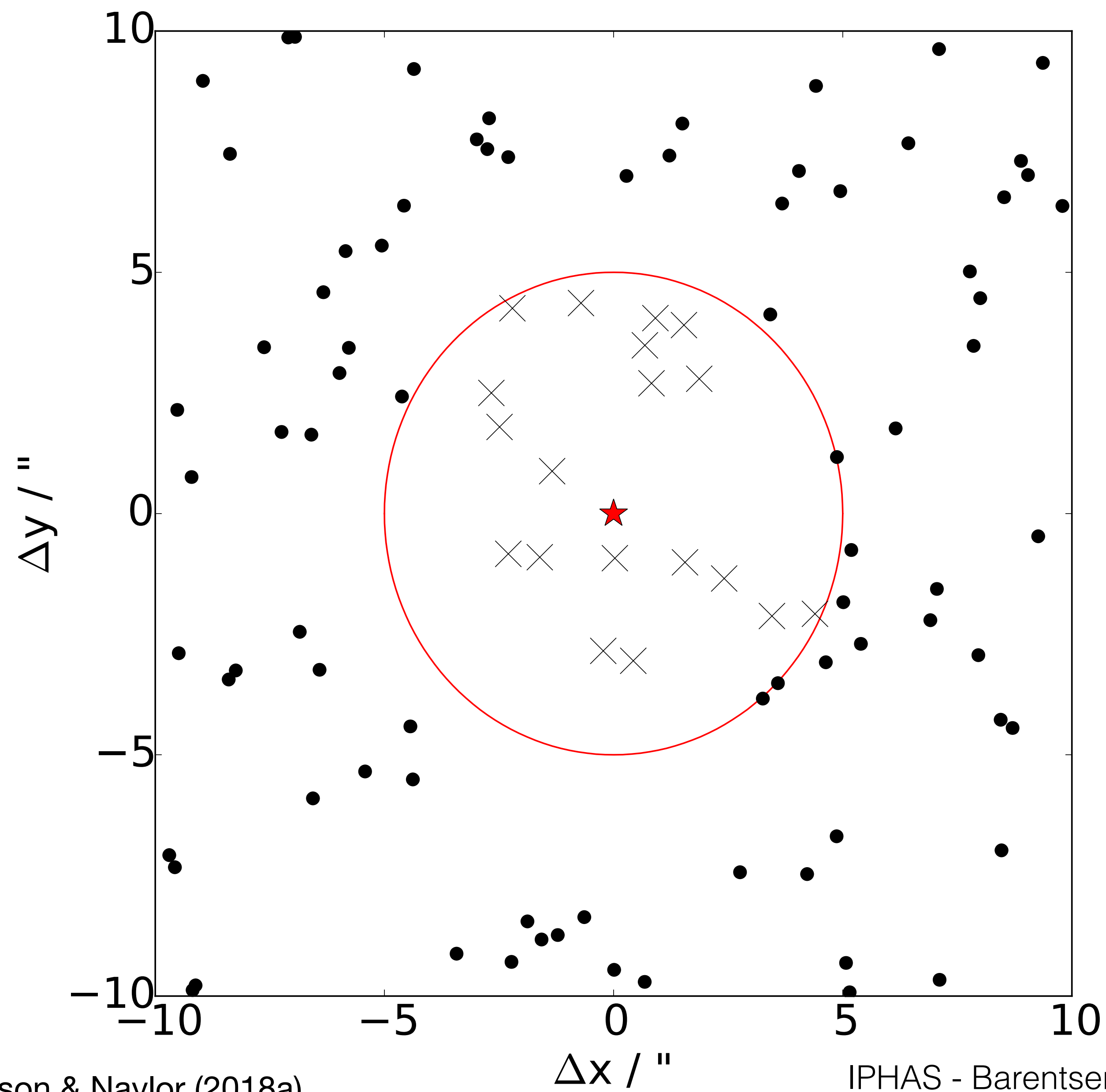
Probability of sources having their brightnesses given they are unrelated to one another (“field stars”)

Probability of sources having their brightnesses given they are counterparts

Photometry: Rejecting False Positives

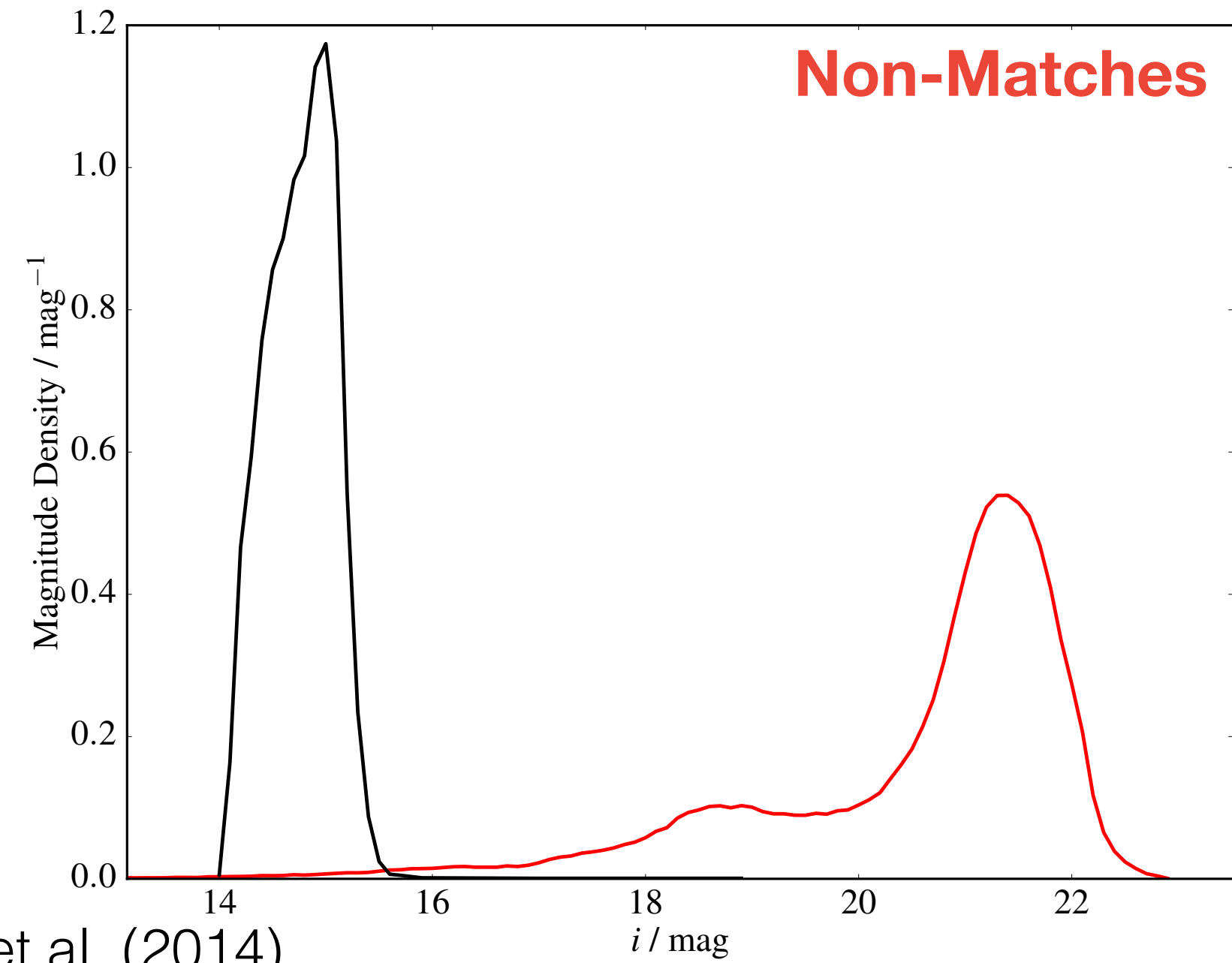


Photometry: The “Field Star” Distribution



Wilson & Naylor (2018a)

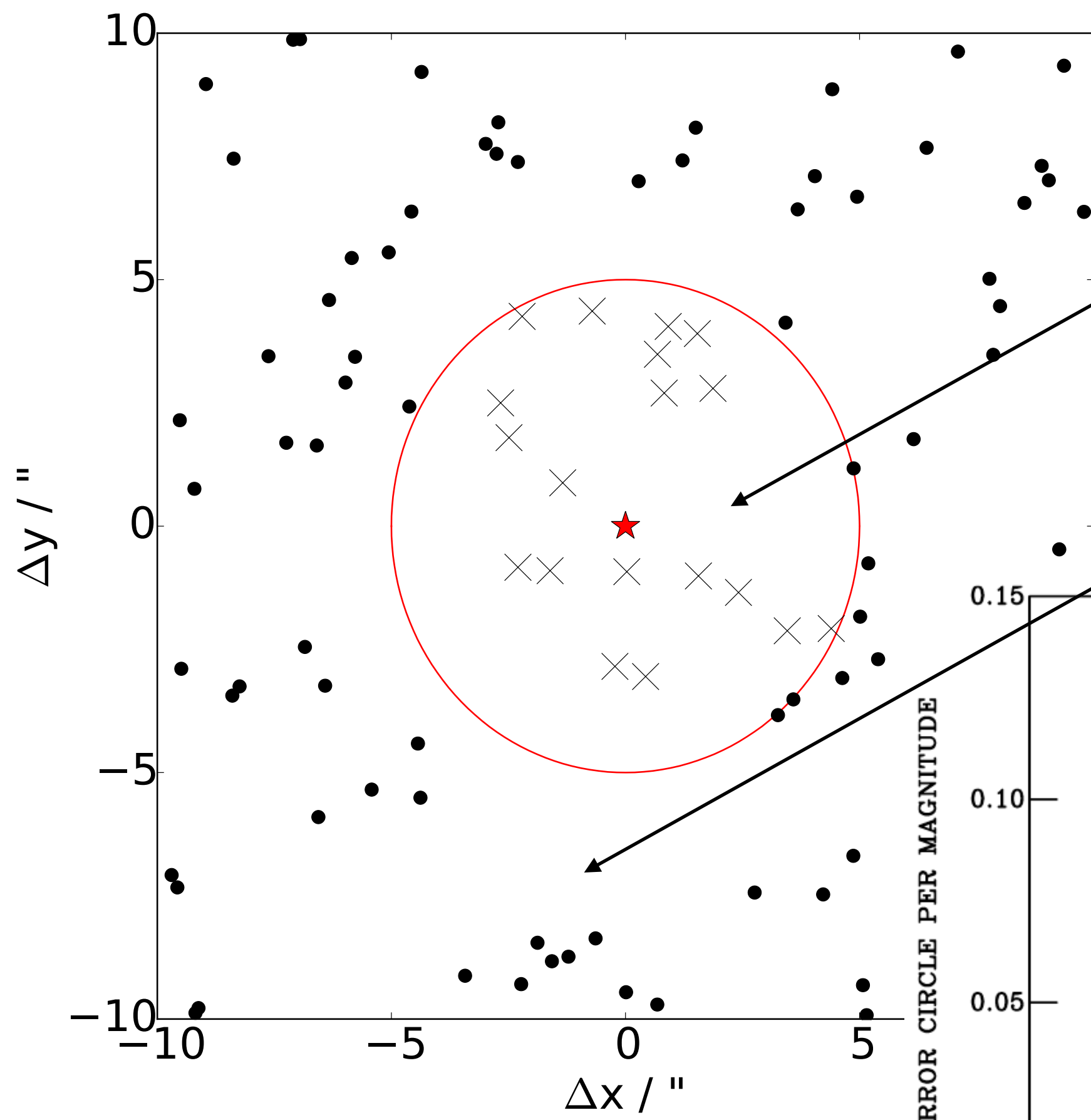
$$P(\zeta, \lambda, k | \gamma, \phi) = \frac{1}{K} \times \underbrace{\prod_{\delta \notin \zeta \cap \delta \in \gamma} N_{\gamma} f_{\gamma}^{\delta}}_{\text{Unmatched}} \underbrace{\prod_{\omega \notin \lambda \cap \omega \in \phi} N_{\phi} f_{\phi}^{\omega}}_{\text{Matched}} \prod_{i=1}^k N_c G_{\gamma\phi}^{\zeta_i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



IPHAS - Barentsen et al. (2014)

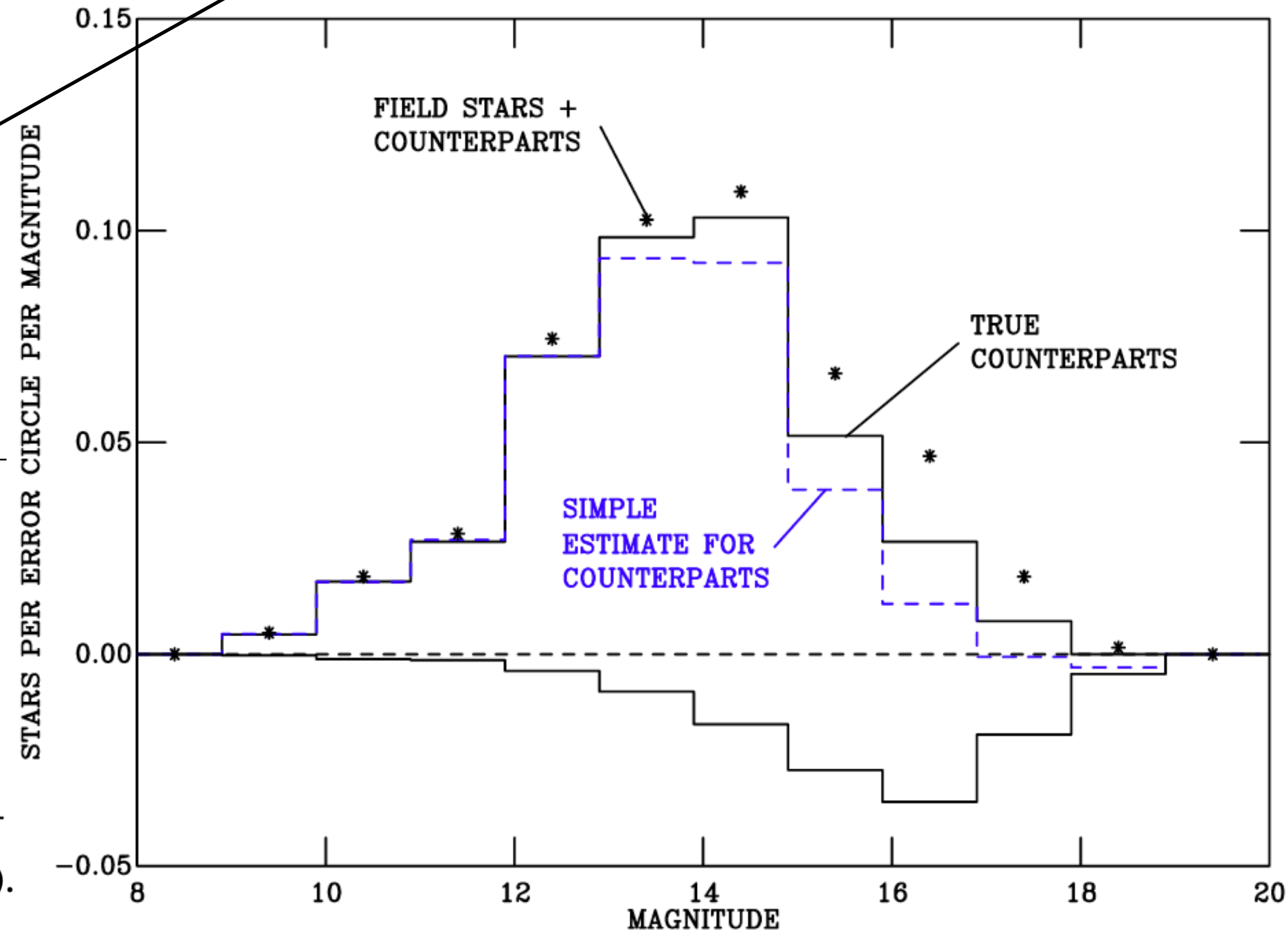
Gaia DR2 - Gaia Collaboration, Brown A. G. A., et al. (2018)

Photometry: The Counterpart Distribution

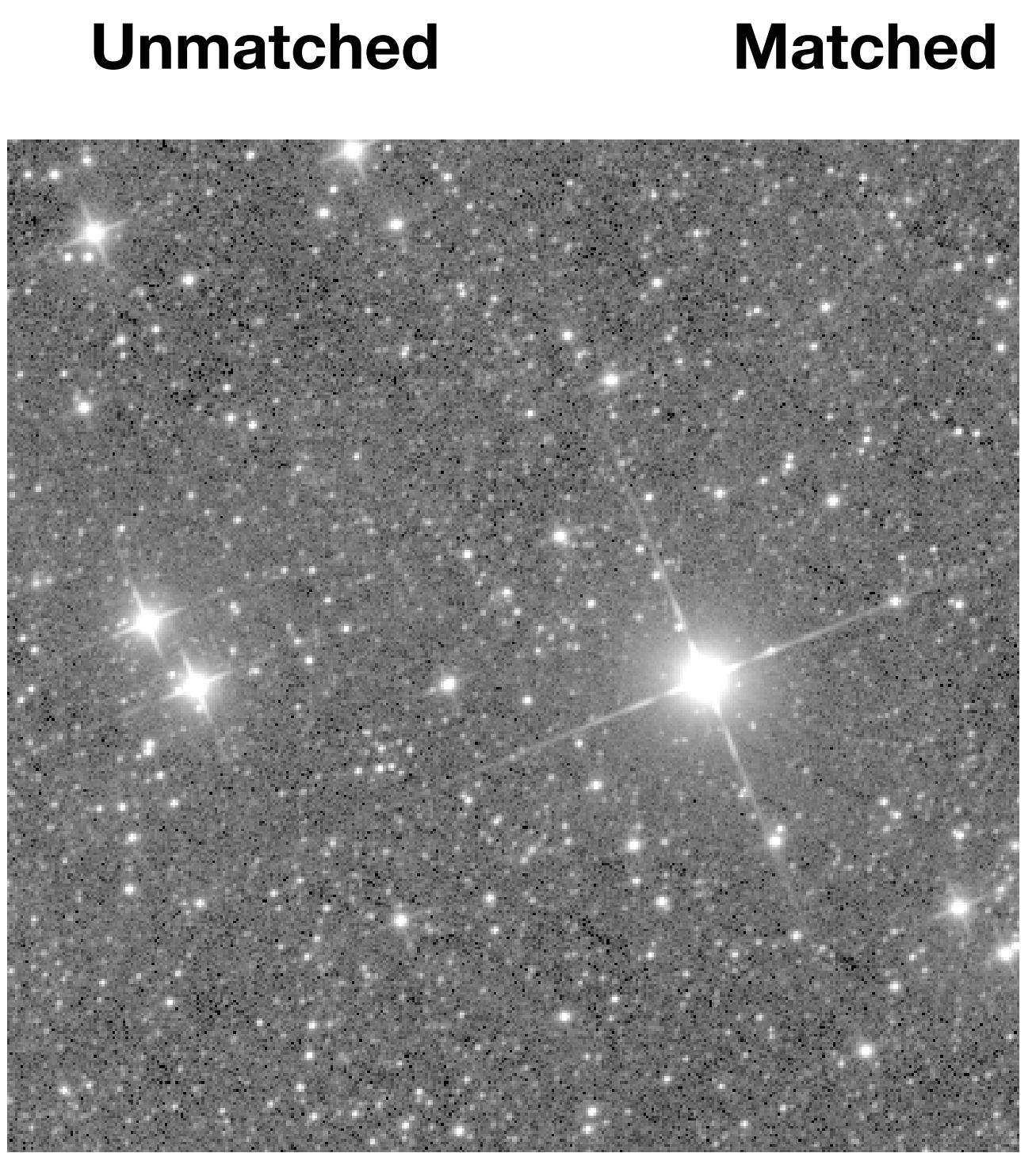


Counterparts =
(Counterparts + field) - field

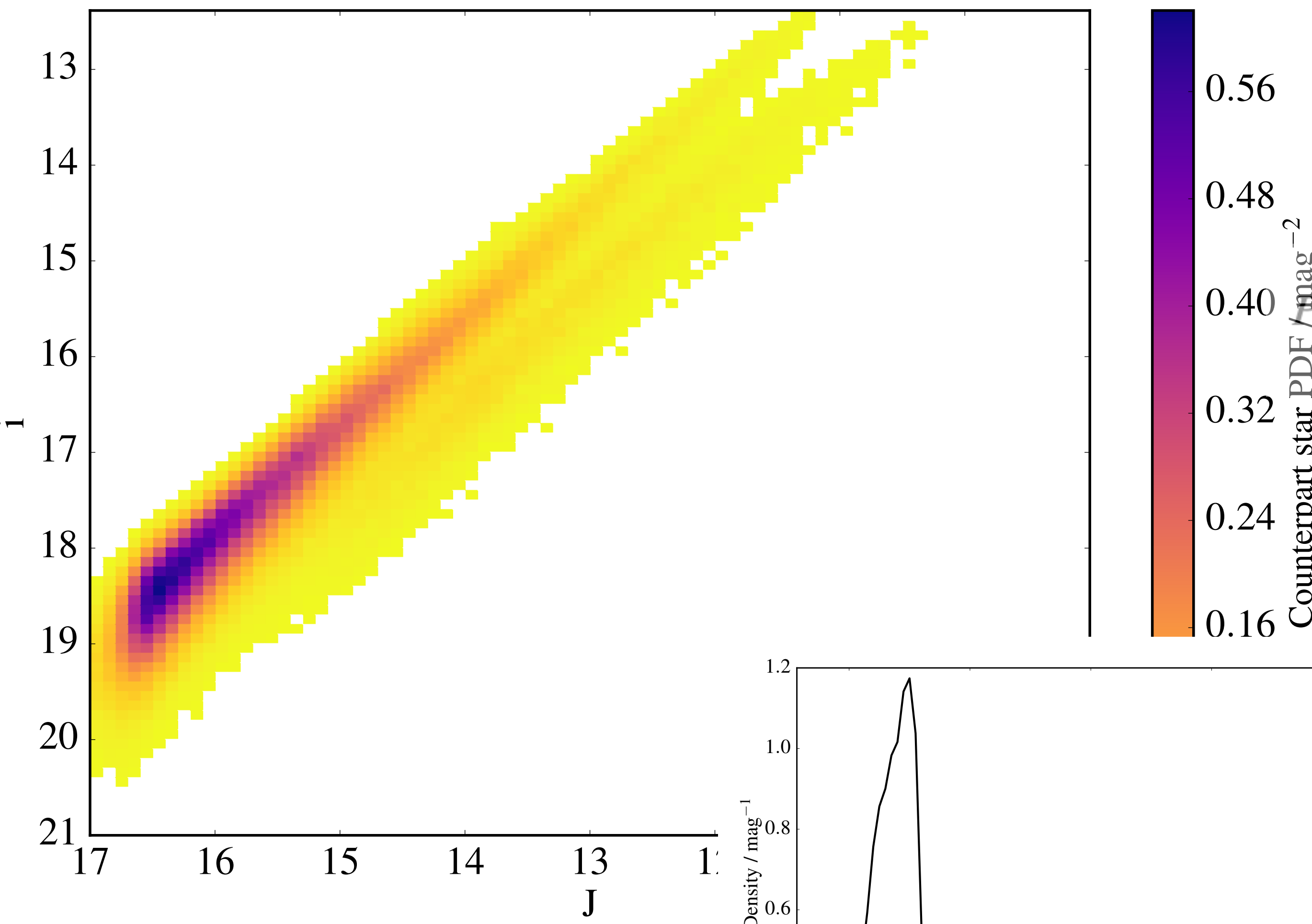
$$P(\zeta, \lambda, k | \gamma, \phi) = \frac{1}{K} \times \underbrace{\prod_{\delta \notin \zeta \cap \delta \in \gamma} N_\gamma f_\gamma^\delta}_{\text{Unmatched}} \prod_{\omega \notin \lambda \cap \omega \in \phi} N_\phi f_\phi^\omega \underbrace{\prod_{i=1}^k N_c G_{\gamma\phi}^{\zeta_i \lambda_i}}_{\text{Matched}}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



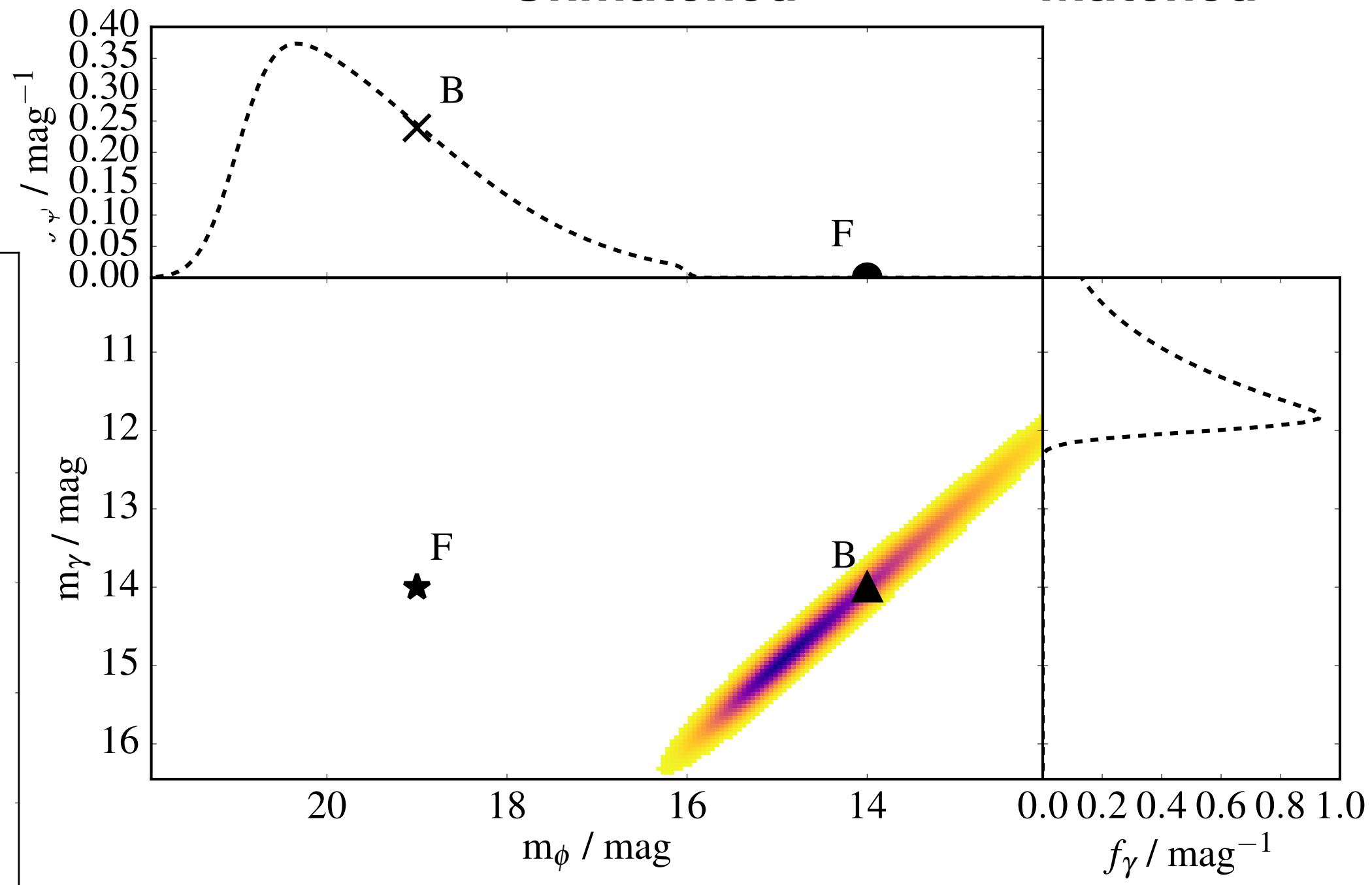
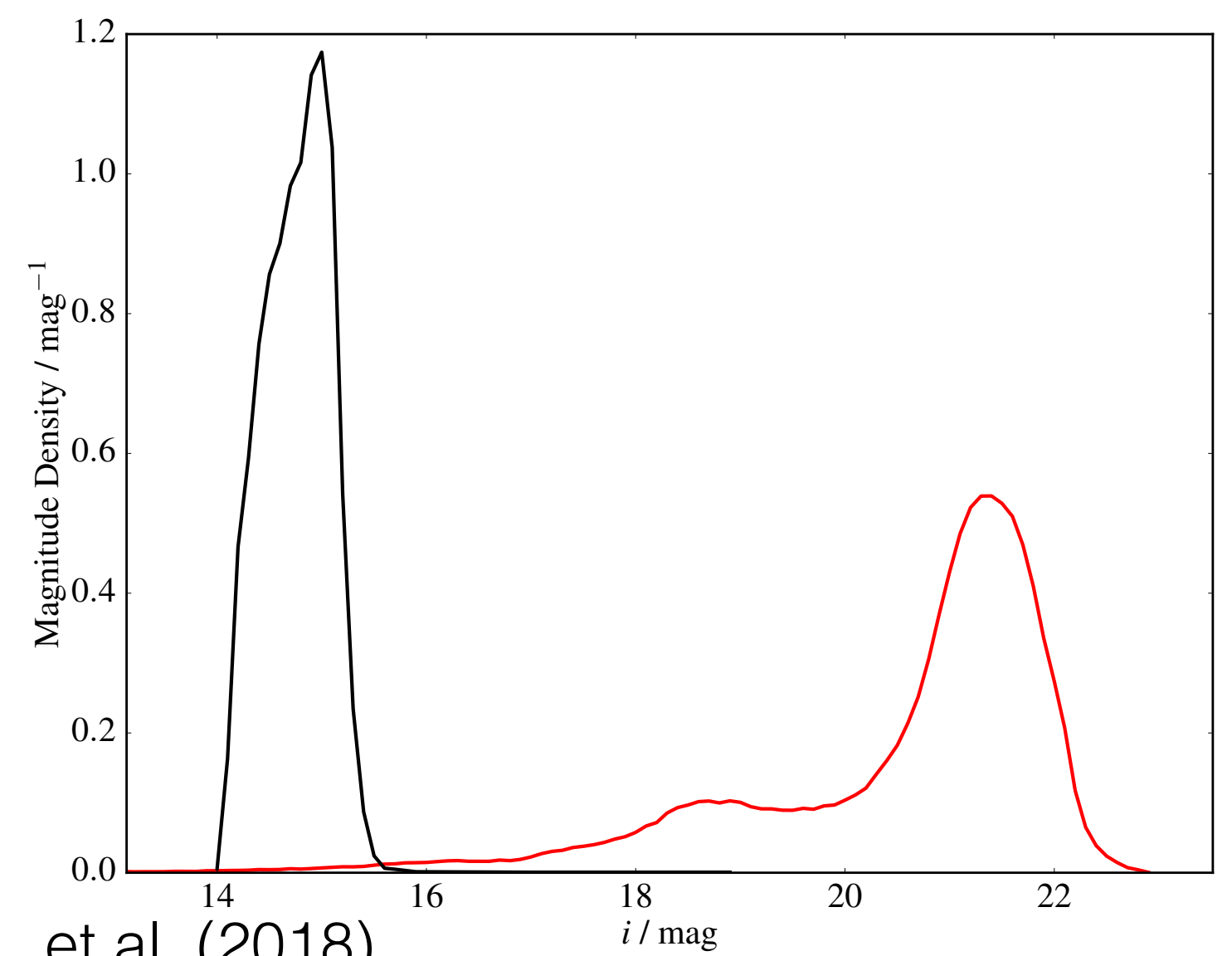
$$Z_{c\gamma} \cdot c_\gamma(m_\phi | m_\gamma) = Z_\gamma b_\gamma(m_\phi | m_\gamma) \exp(A_\gamma N_\phi F_\phi(m_\phi)) - (1 - Z_{c\gamma} C_\gamma(m_\phi | m_\gamma)) A_\gamma N_\phi f_\phi(m_\phi).$$



Photometry: The Counterpart Distribution



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



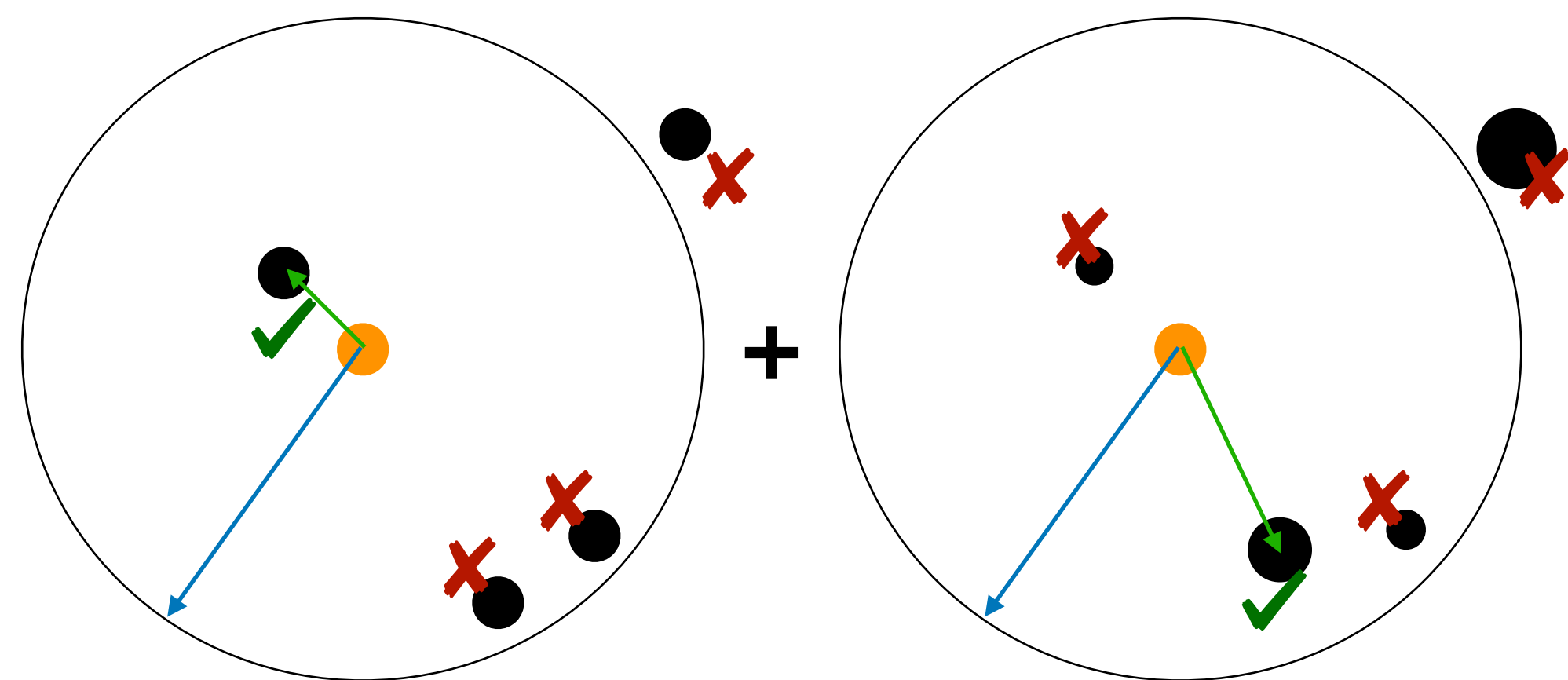
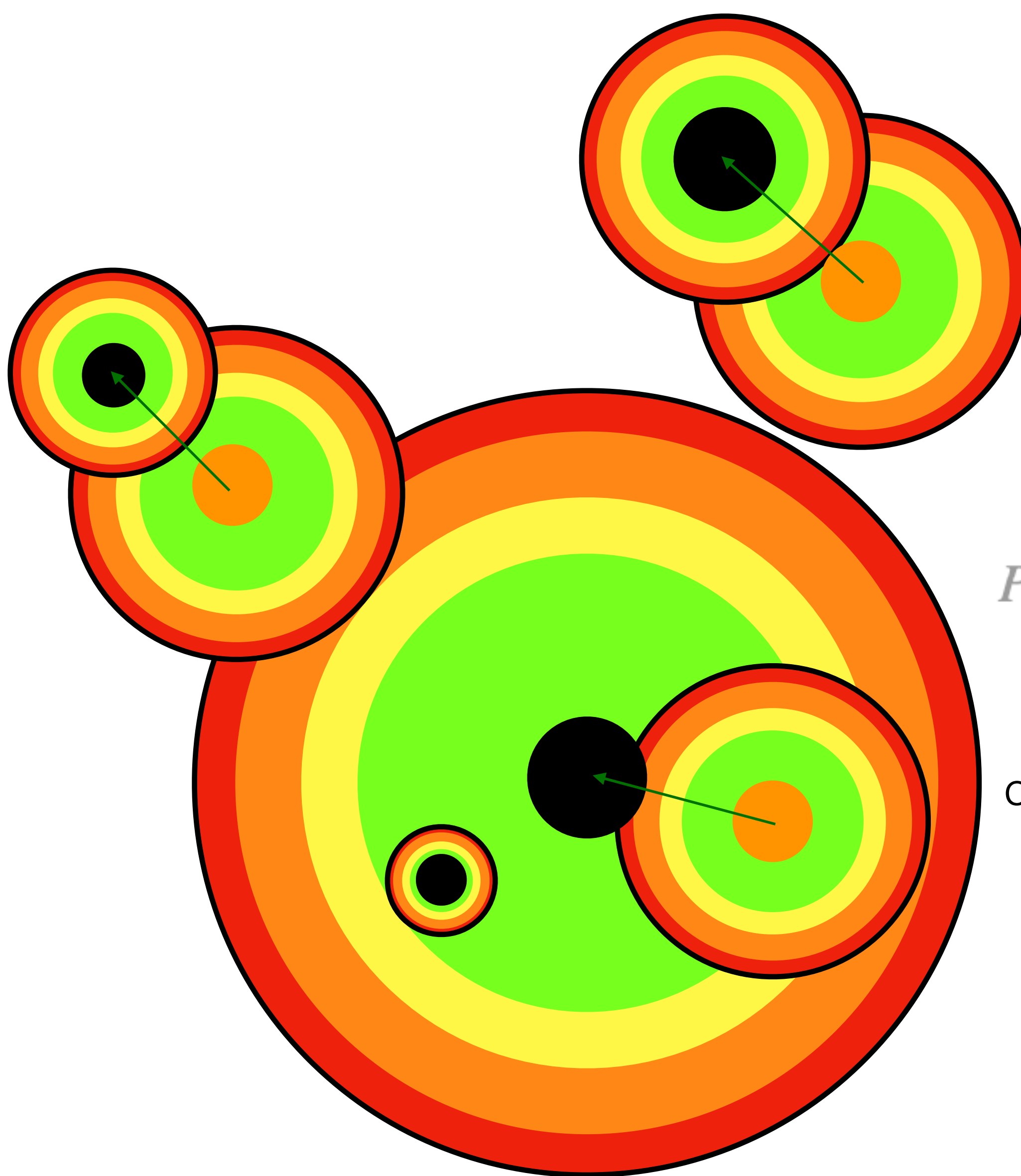
Wilson & Naylor (2018a)

IPHAS - Barentsen et al. (2014)

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

Tom J Wilson @onoddil

Probabilistic Cross-Matching



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

G includes information on position (un)certainty

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

Consider multiple objects in both catalogues simultaneously

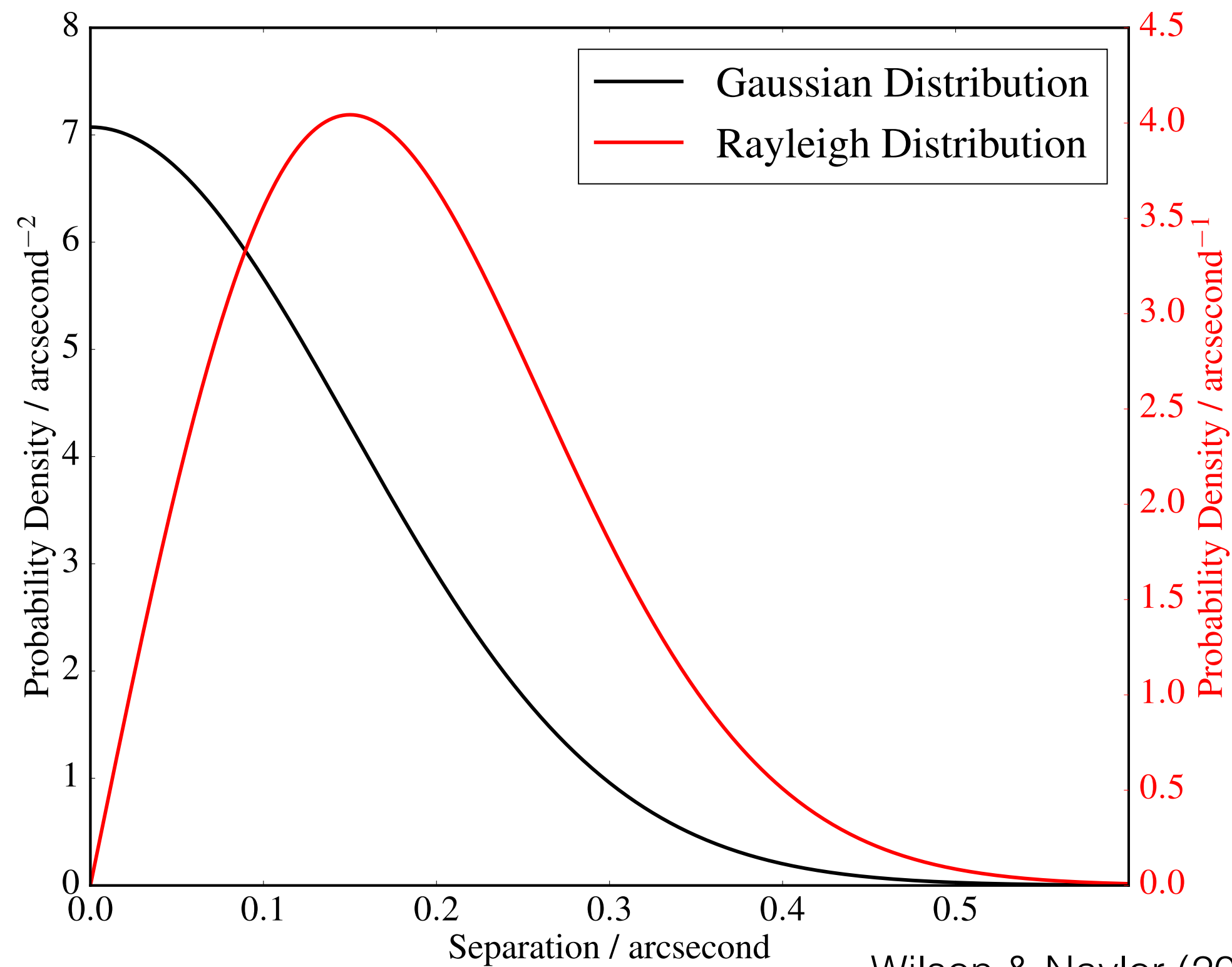
Probability of sources having their brightnesses given they are unrelated to one another (“field stars”)

Probability of sources having their brightnesses given they are counterparts

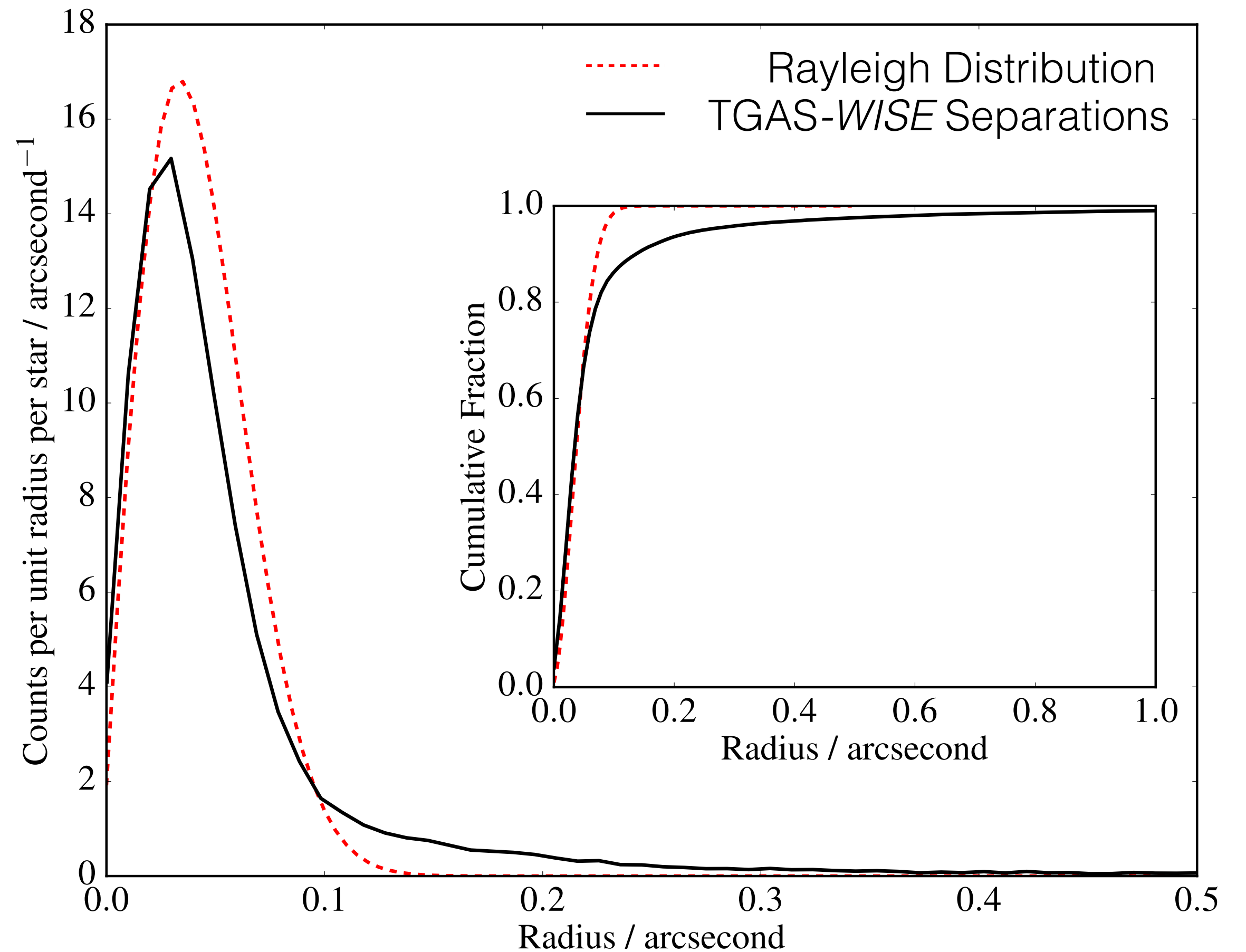
The *Astrometric Uncertainty* Function

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

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



Wilson & Naylor (2017)
WISE - Wright et al. (2010)

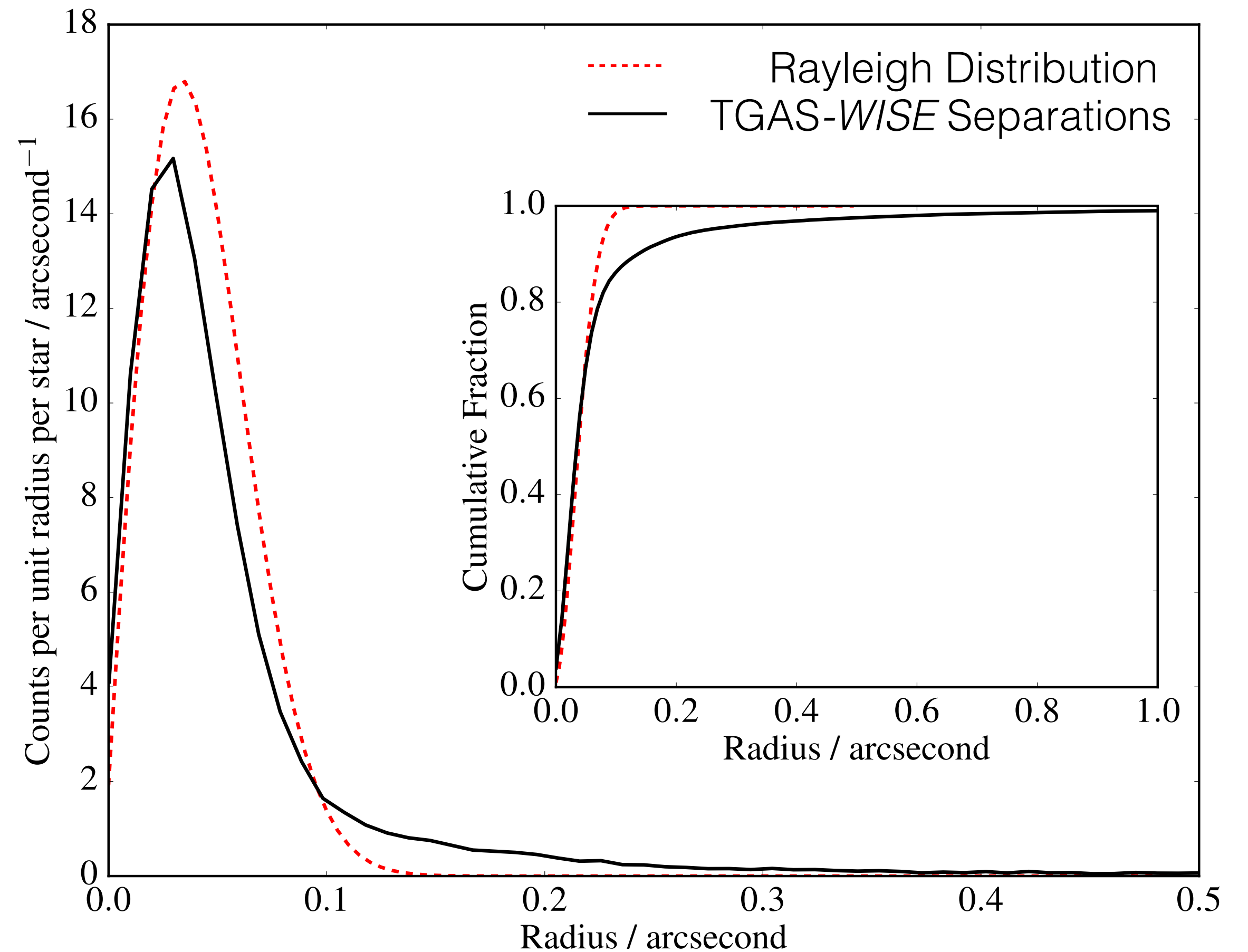


TGAS - Michalik, Lindegren, & Hobbs (2015)
Gaia - Gaia Collaboration, Brown A. G. A., et al. (2016)

The *Astrometric Uncertainty* Function

Reasons for large separations:

- 1) proper motions (e.g. AllWISE Supplement 6.4, Cutri et al. 2012) — no, TGAS provided for all sources
- 2) false matches — no, 0.1% chance of random match within 0.5 arcseconds
- 3) What else could it be?

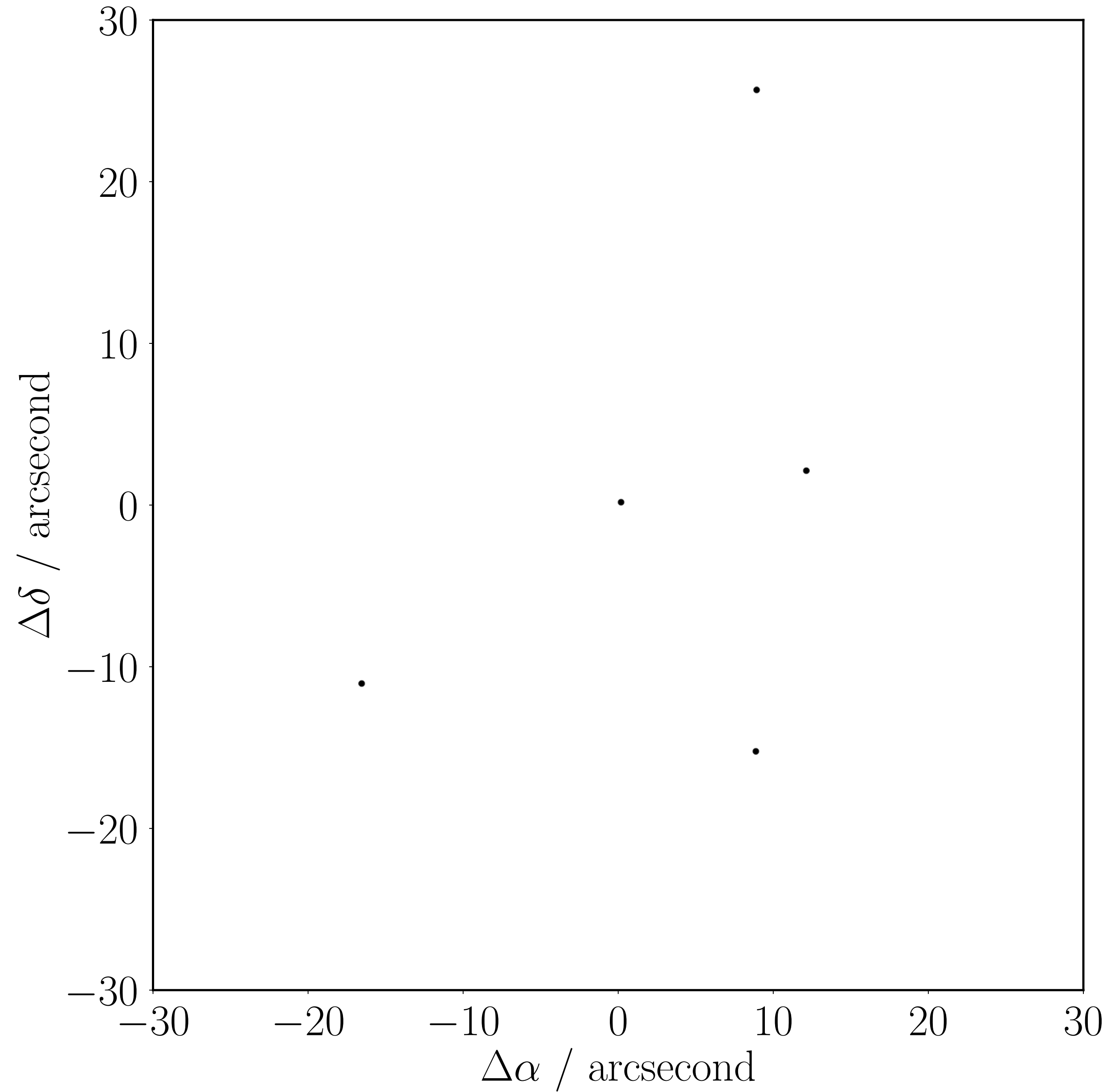


Wilson & Naylor (2017)
WISE - Wright et al. (2010)

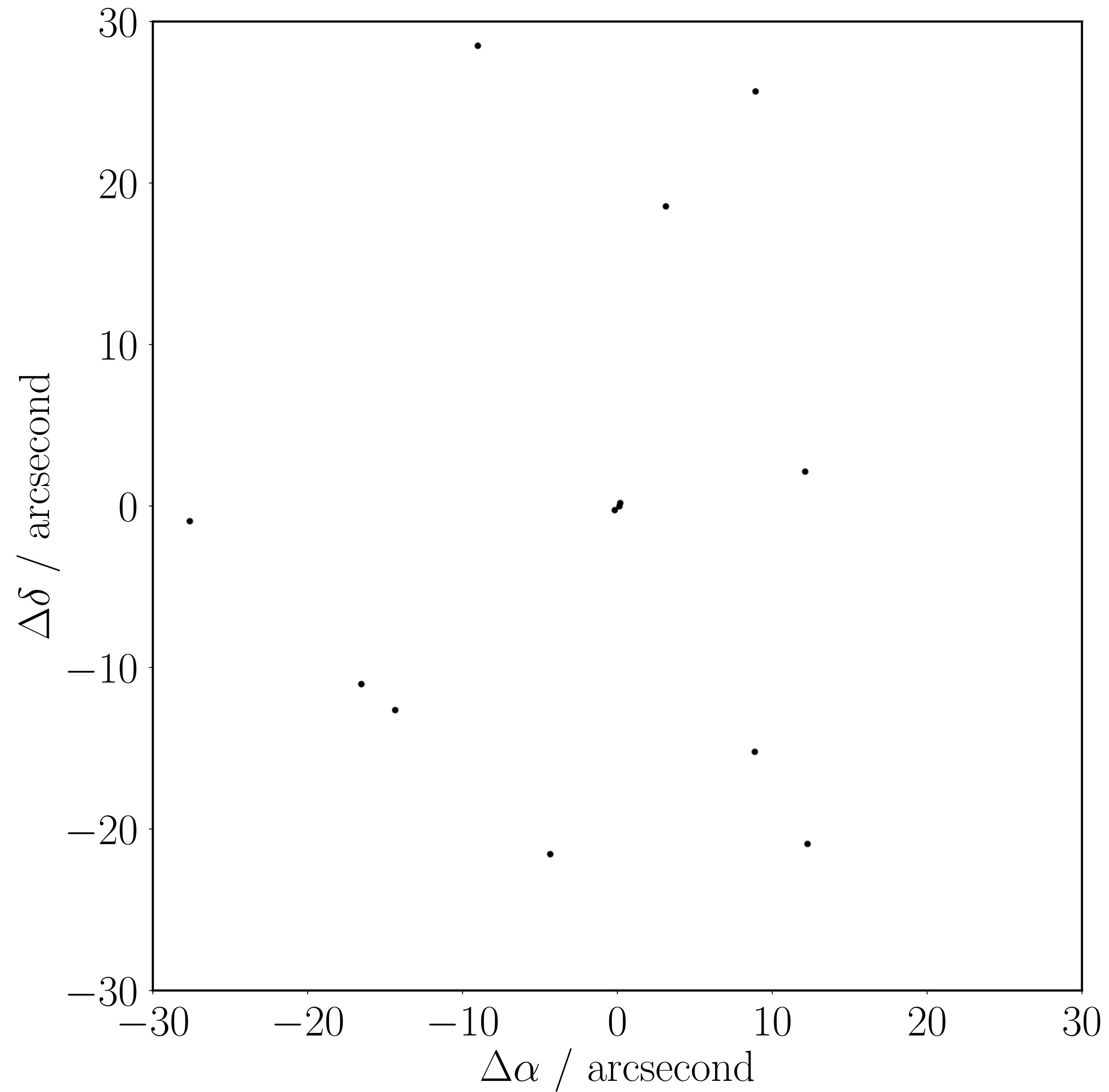
TGAS - Michalik, Lindegren, & Hobbs (2015)
Gaia - Gaia Collaboration, Brown A. G. A., et al. (2016)

Tom J Wilson @onoddil

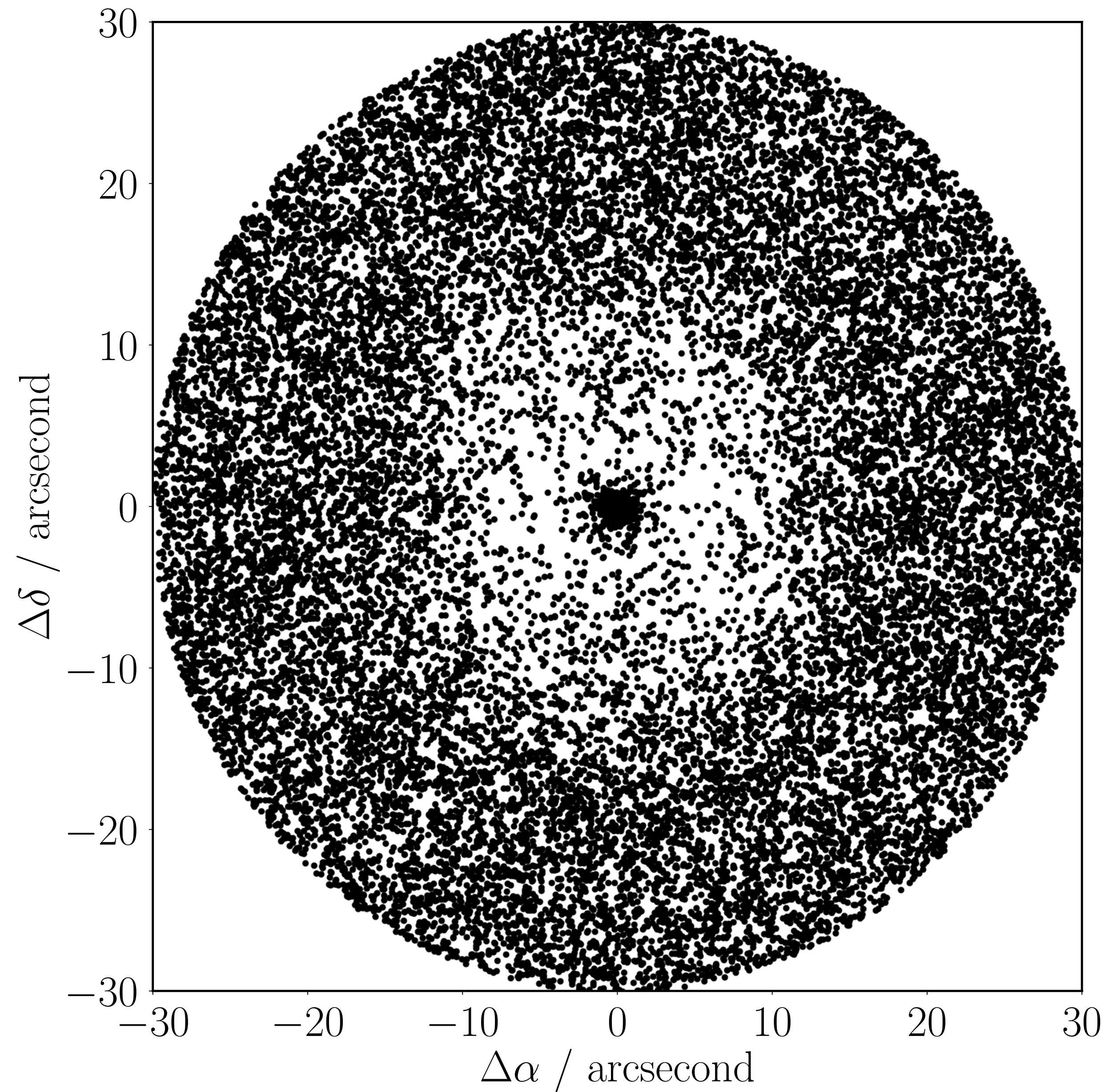
The AUF: Crowding



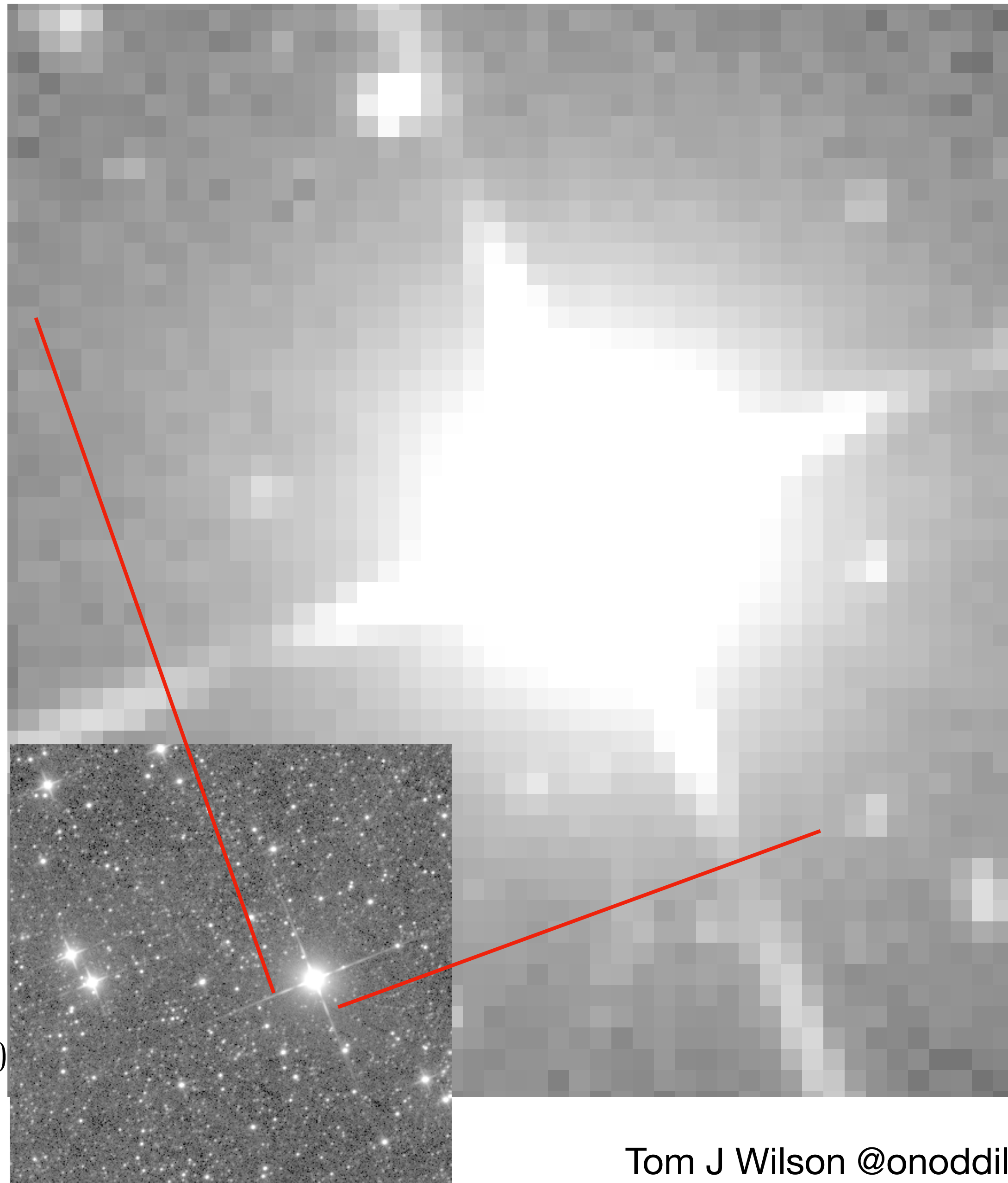
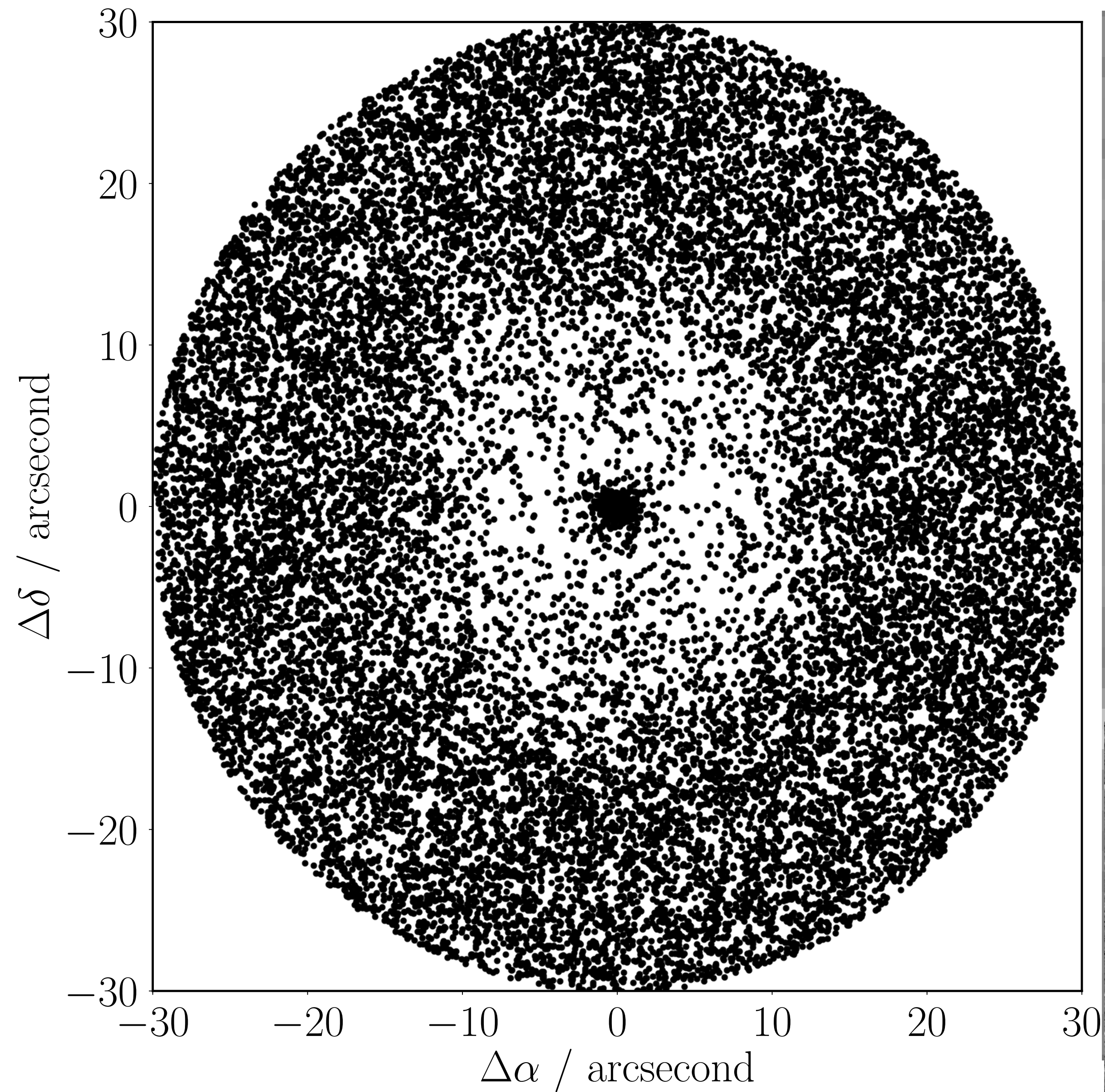
The AUF: Crowding



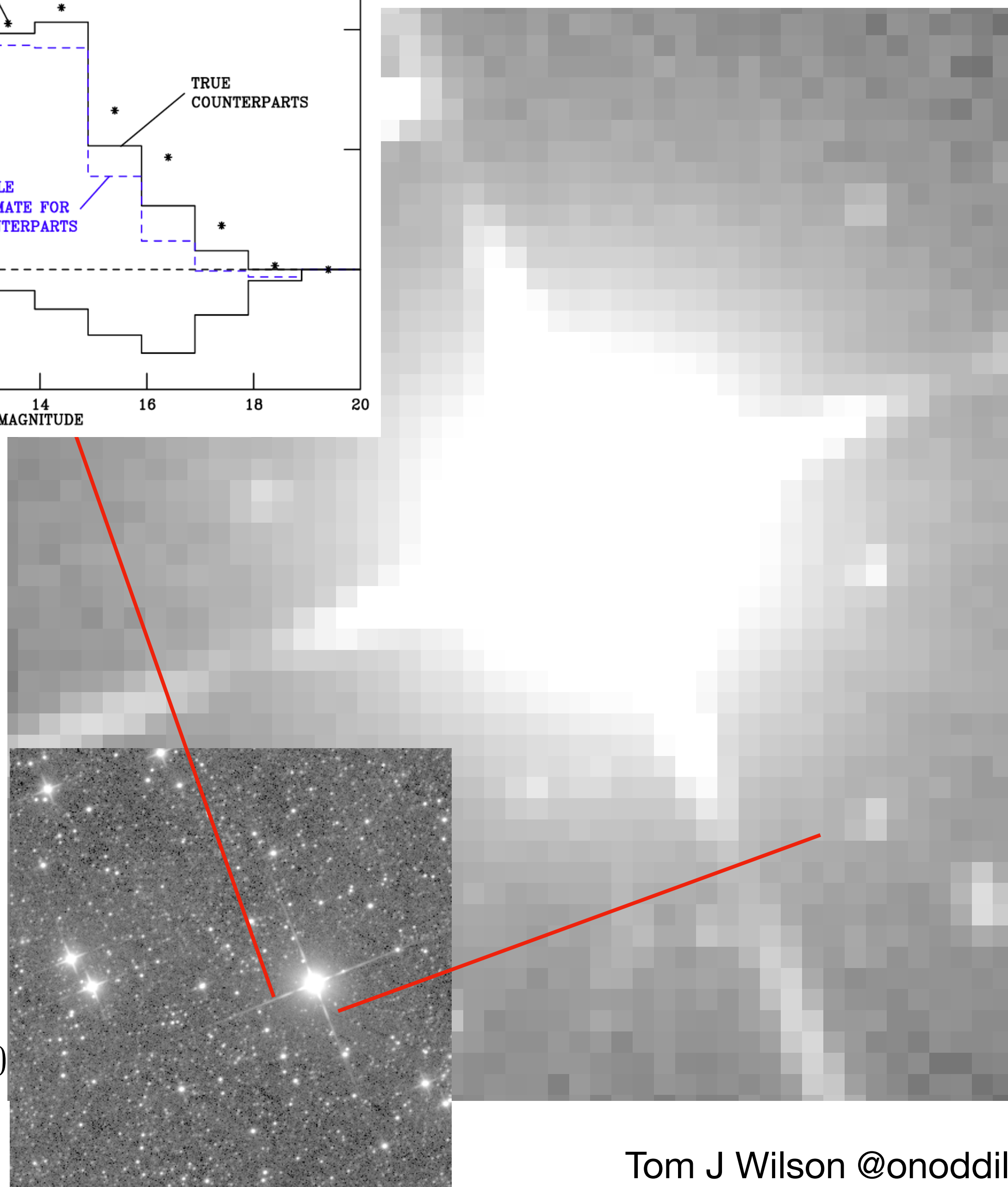
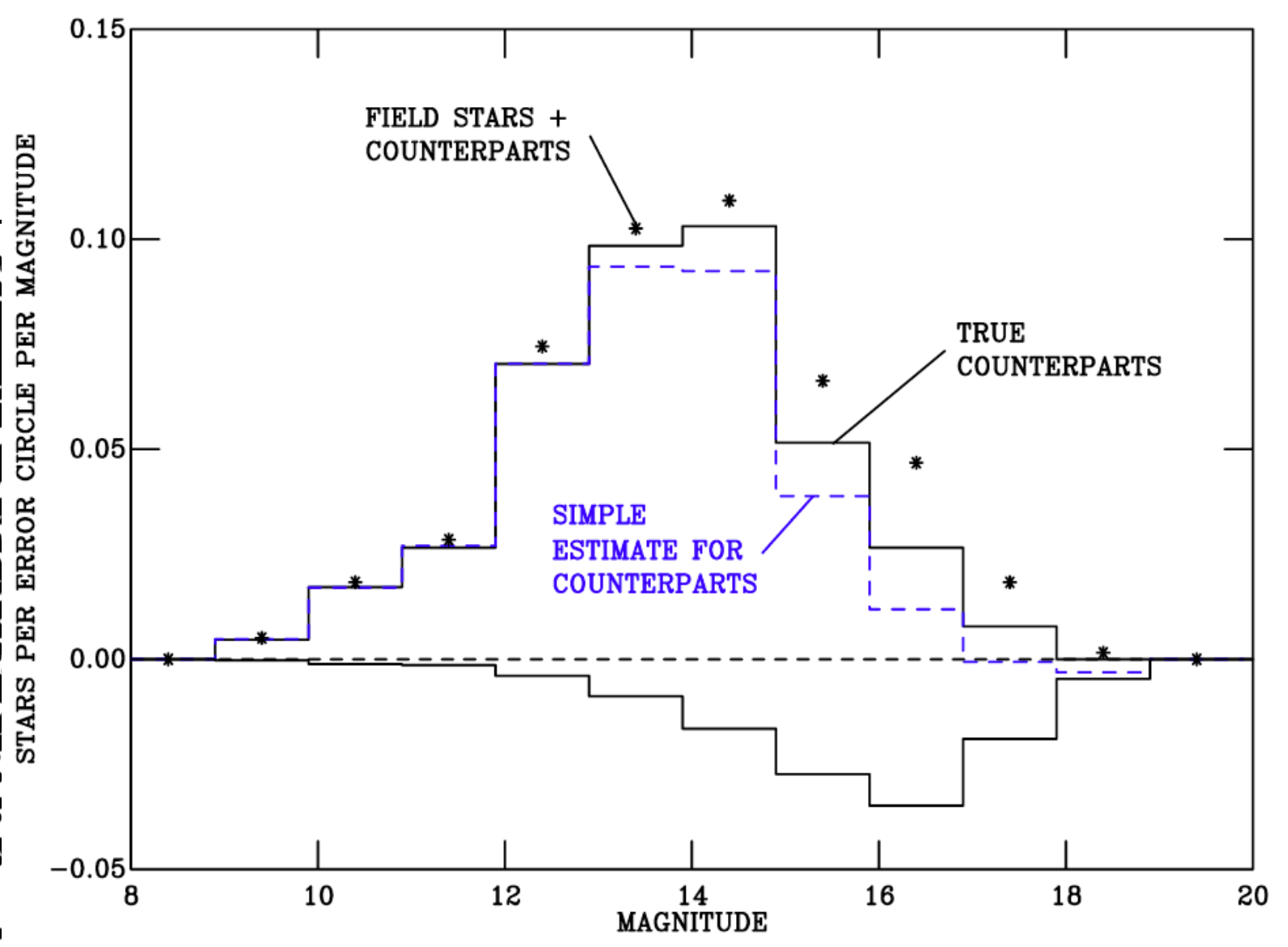
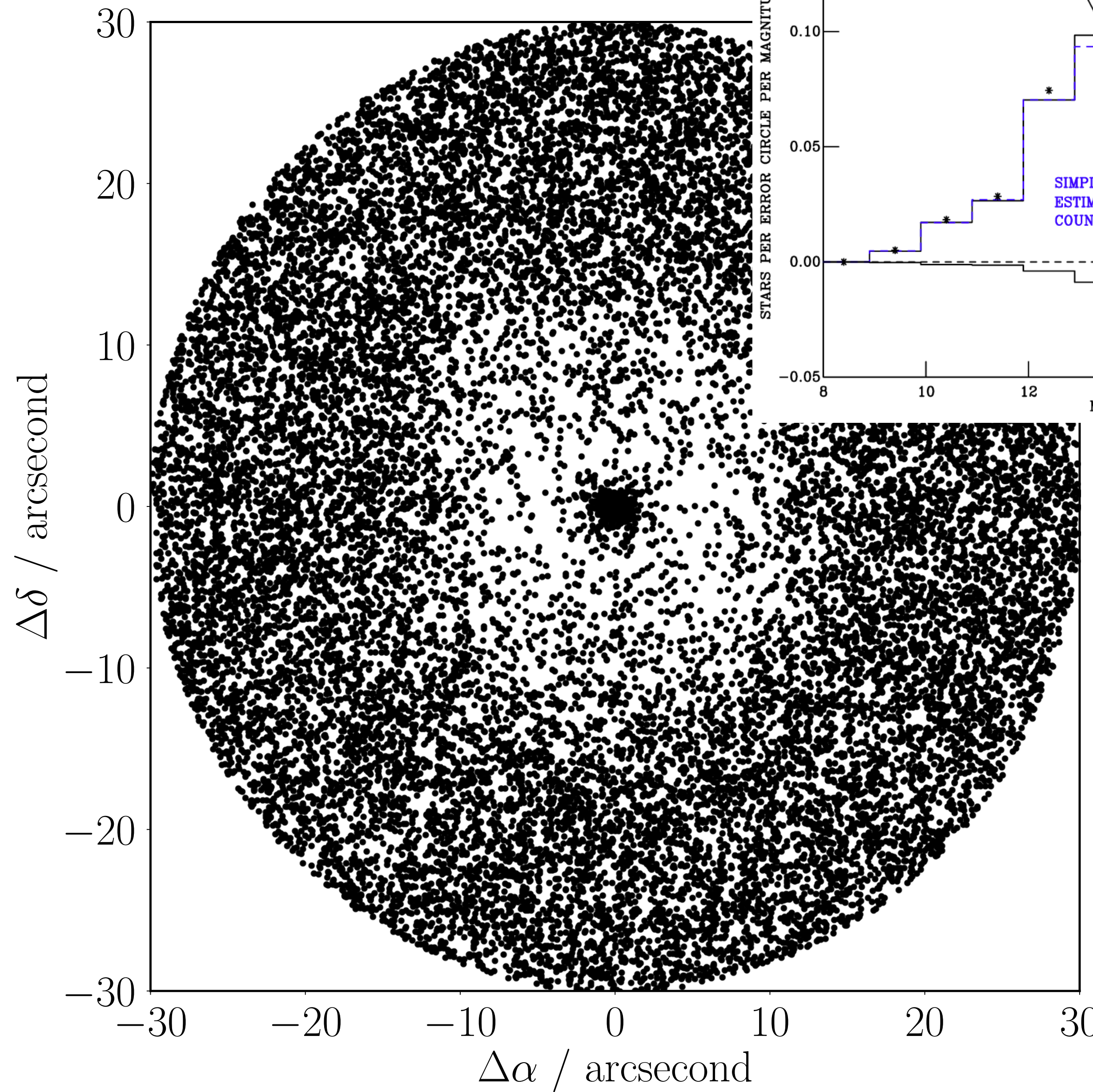
The AUF: Crowding



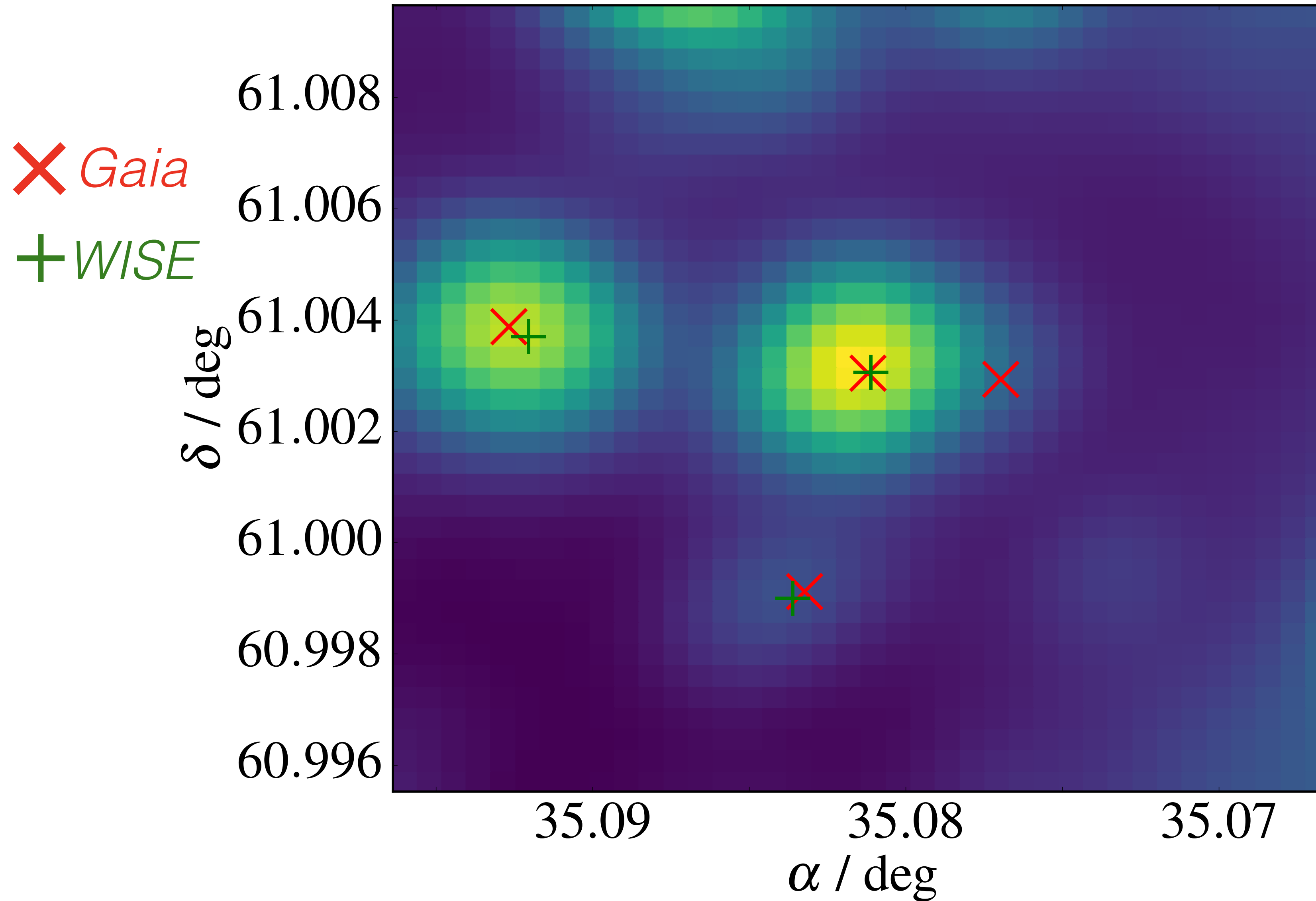
The AUF: Crowding



The AUF: Crowding



Resolving *Gaia*-*WISE* Blends



“Were the succession of stars endless... there could be absolutely no point, in all that background, at which would not exist a star.”

— Edgar Allan Poe, *Eureka* (1848)

Tom J Wilson @onoddil

Wilson & Naylor (2018b)

WISE - Wright et al. (2010)

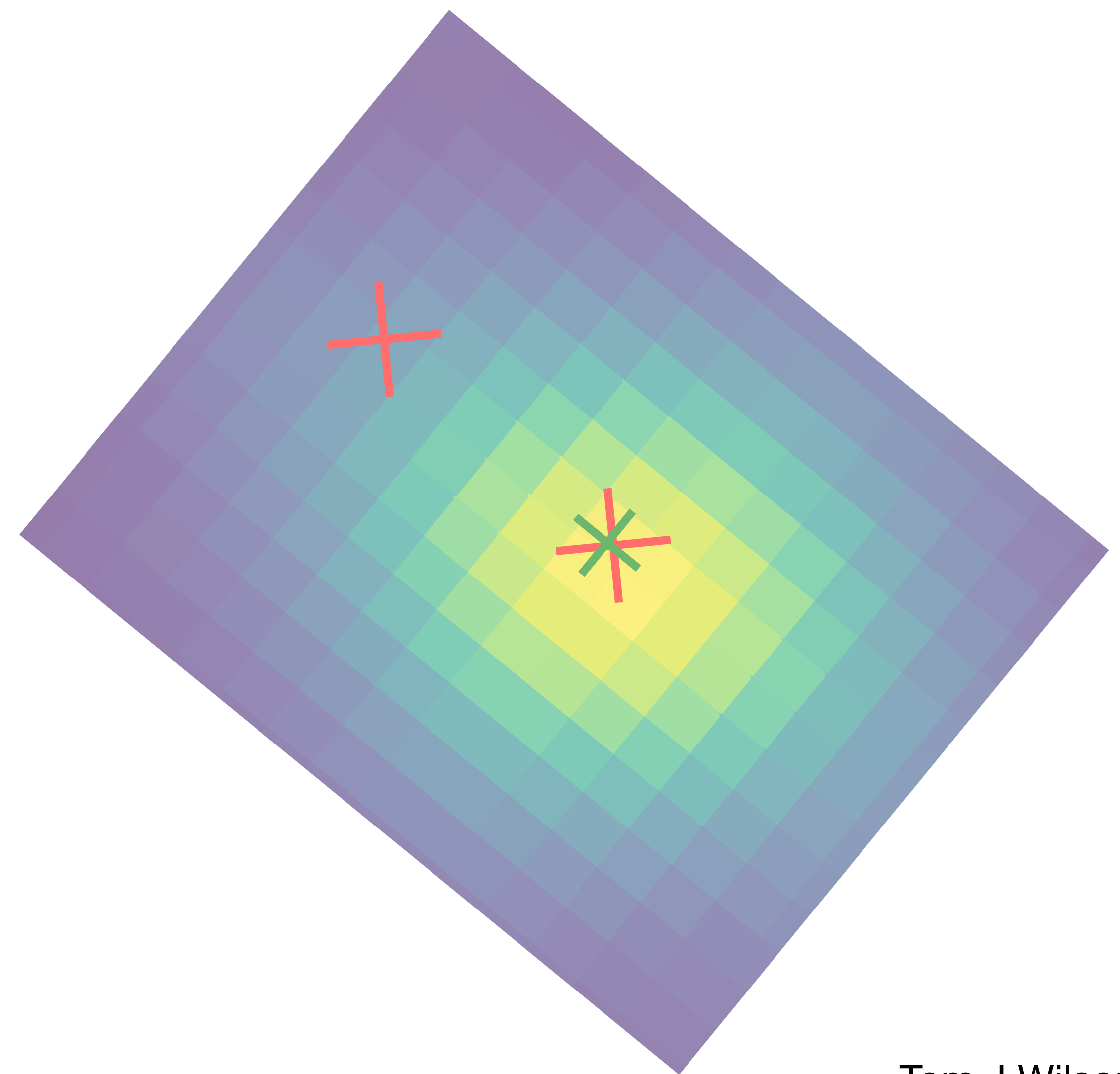
Gaia DR2 - Gaia Collaboration, Brown A. G. A., et al. (2018)

The AUF: Perturbation

● Pure *WISE* position



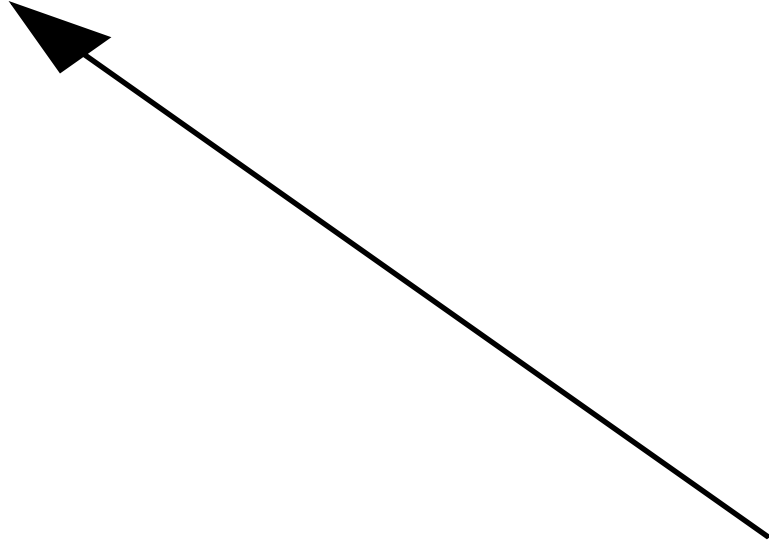
✕ *Gaia* position



Wilson & Naylor (2017)
Wilson & Naylor (2018b)
WISE - Wright et al. (2010)
Gaia DR2 - Gaia Collaboration, Brown A. G. A., et al. (2018)

The AUF: Perturbation

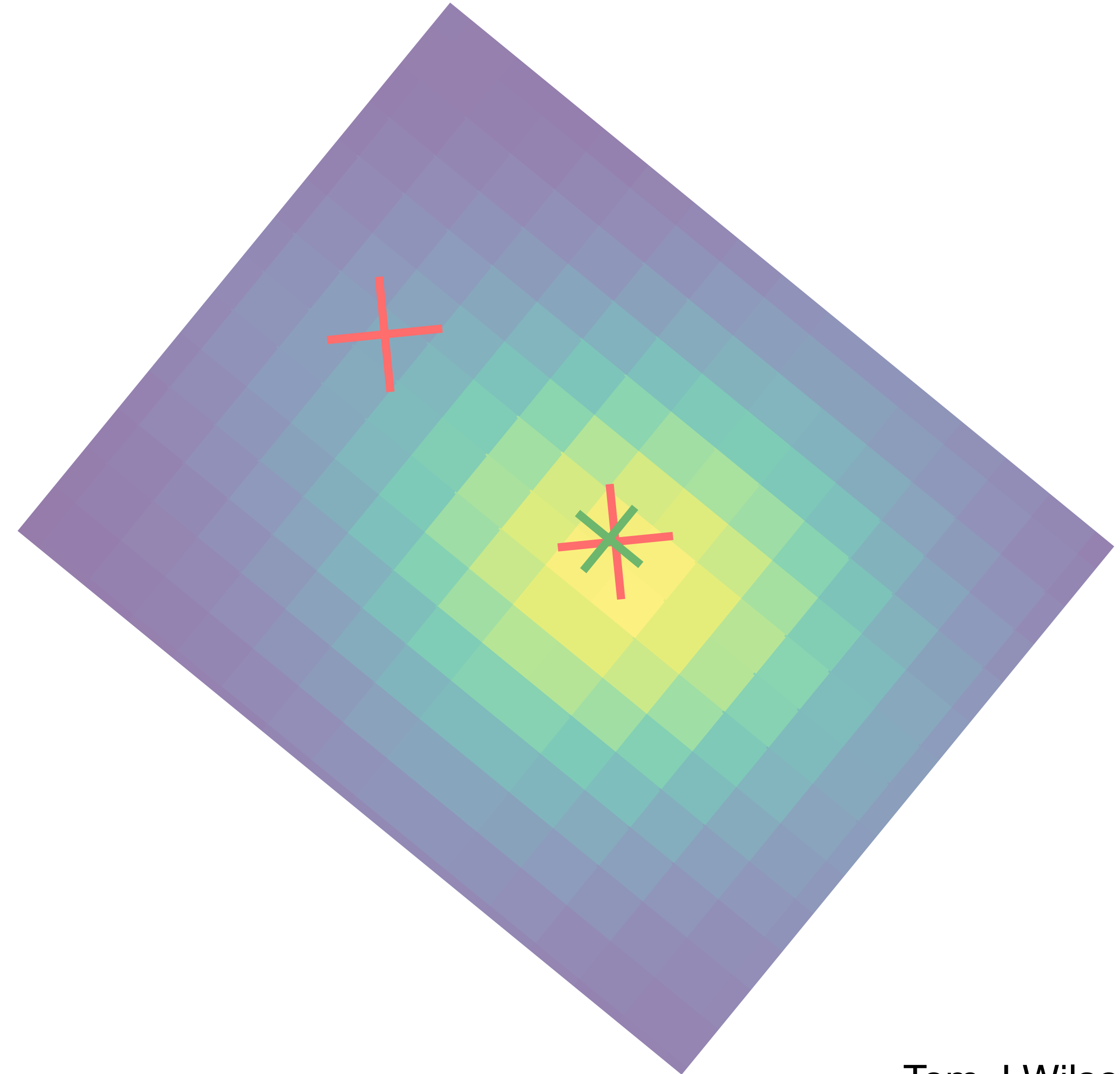
To *WISE* contaminant



Pure *WISE* position



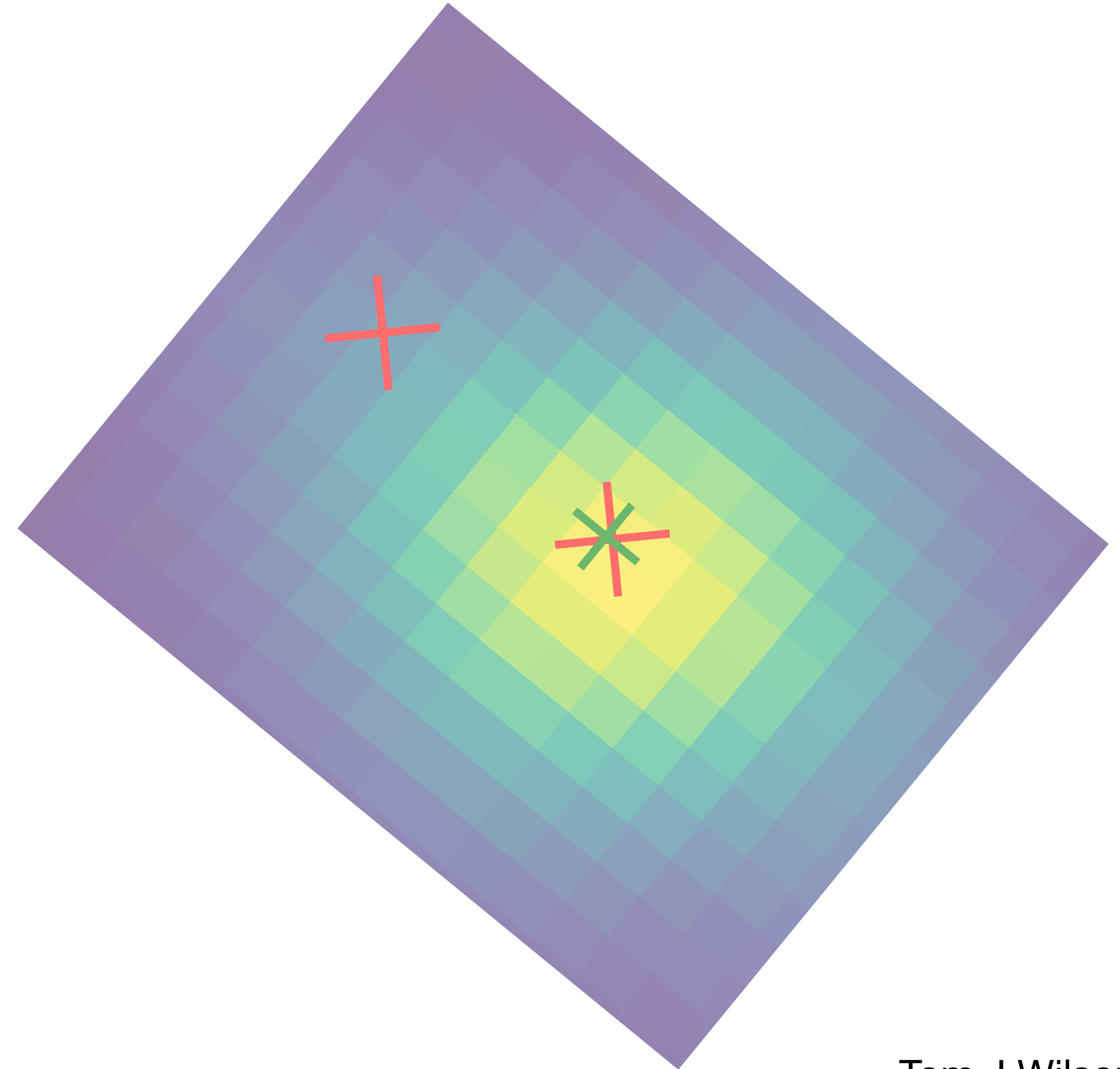
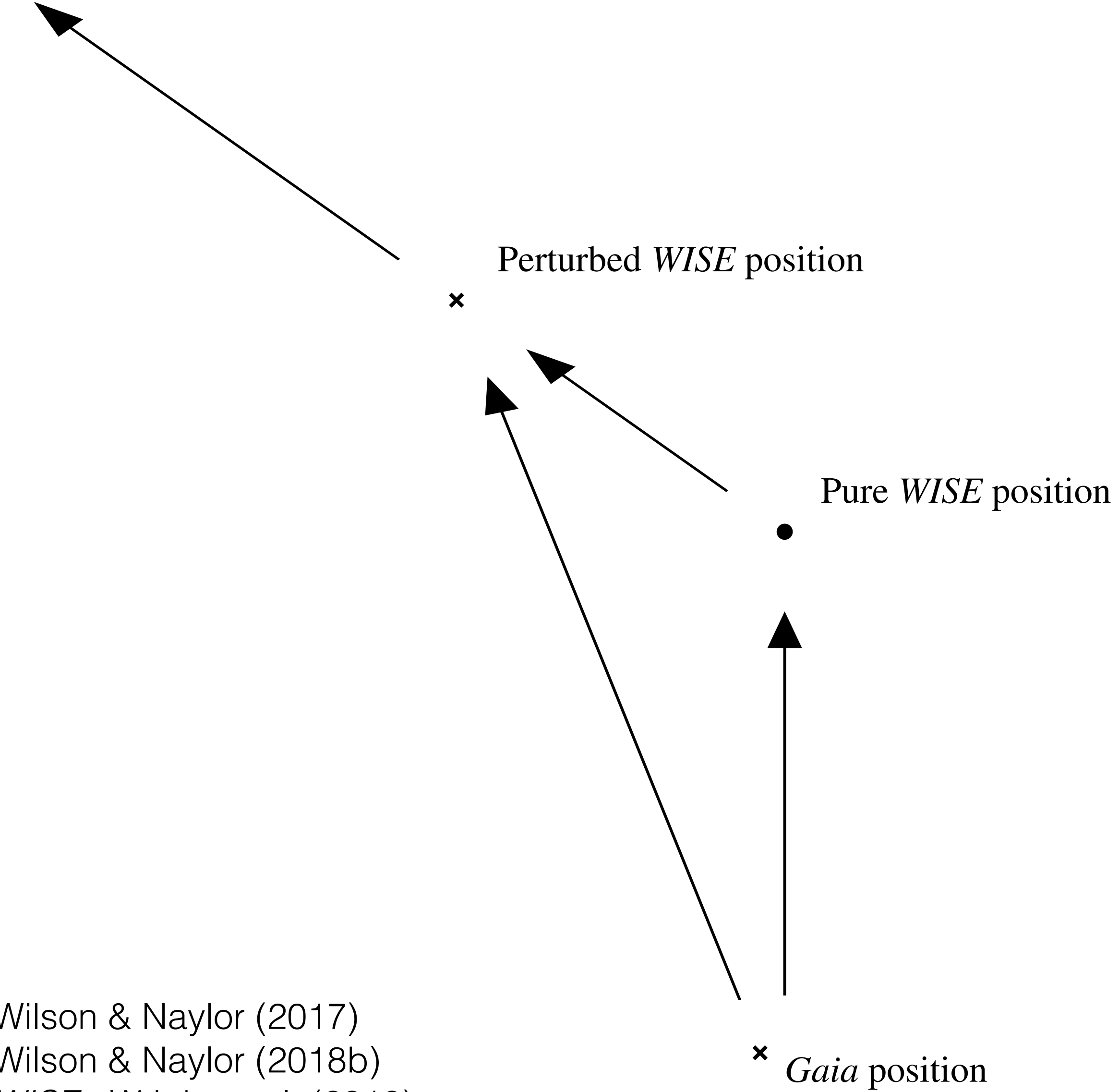
× *Gaia* position



Wilson & Naylor (2017)
Wilson & Naylor (2018b)
WISE - Wright et al. (2010)
Gaia DR2 - Gaia Collaboration, Brown A. G. A., et al. (2018)

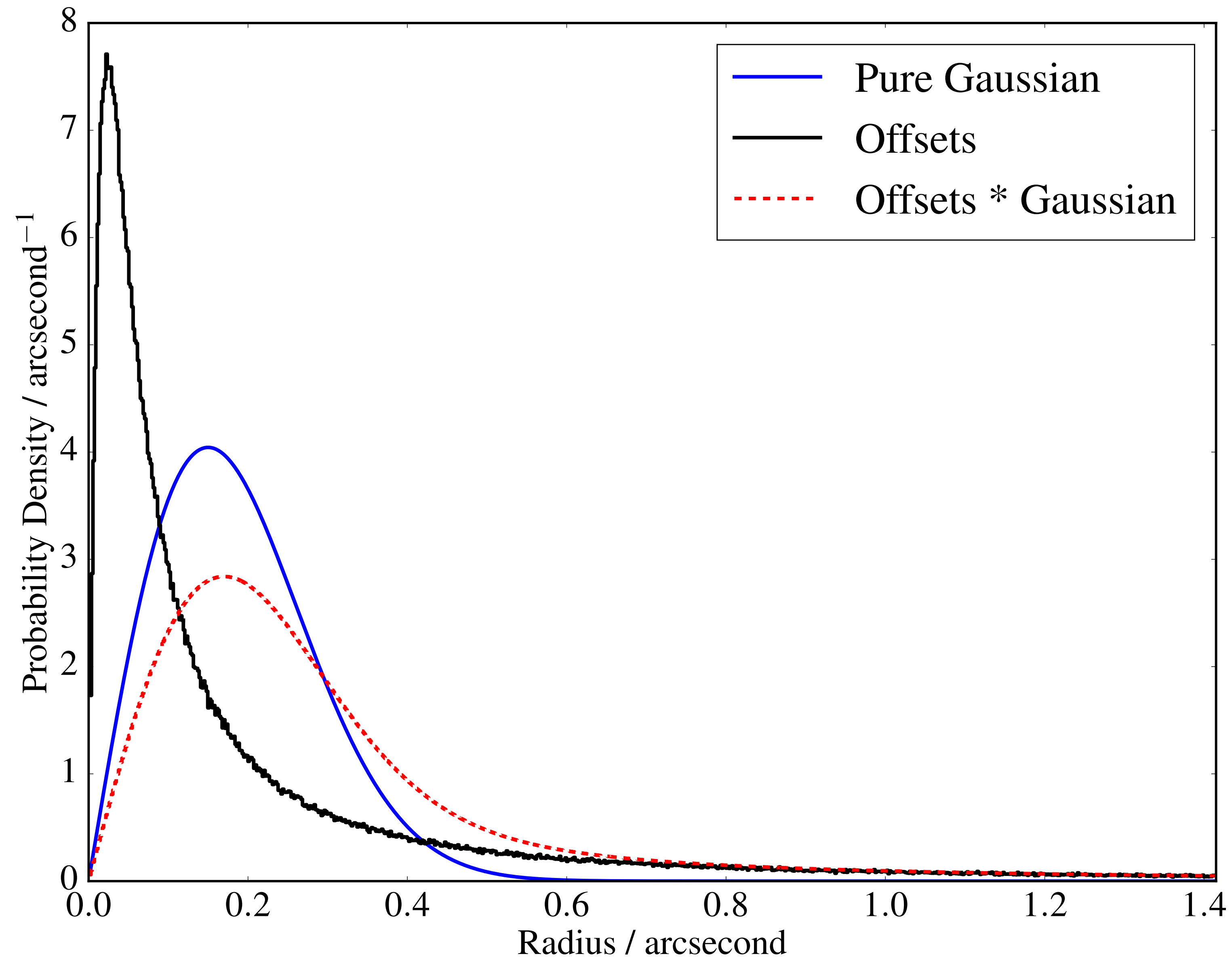
The AUF: Perturbation

To *WISE* contaminant

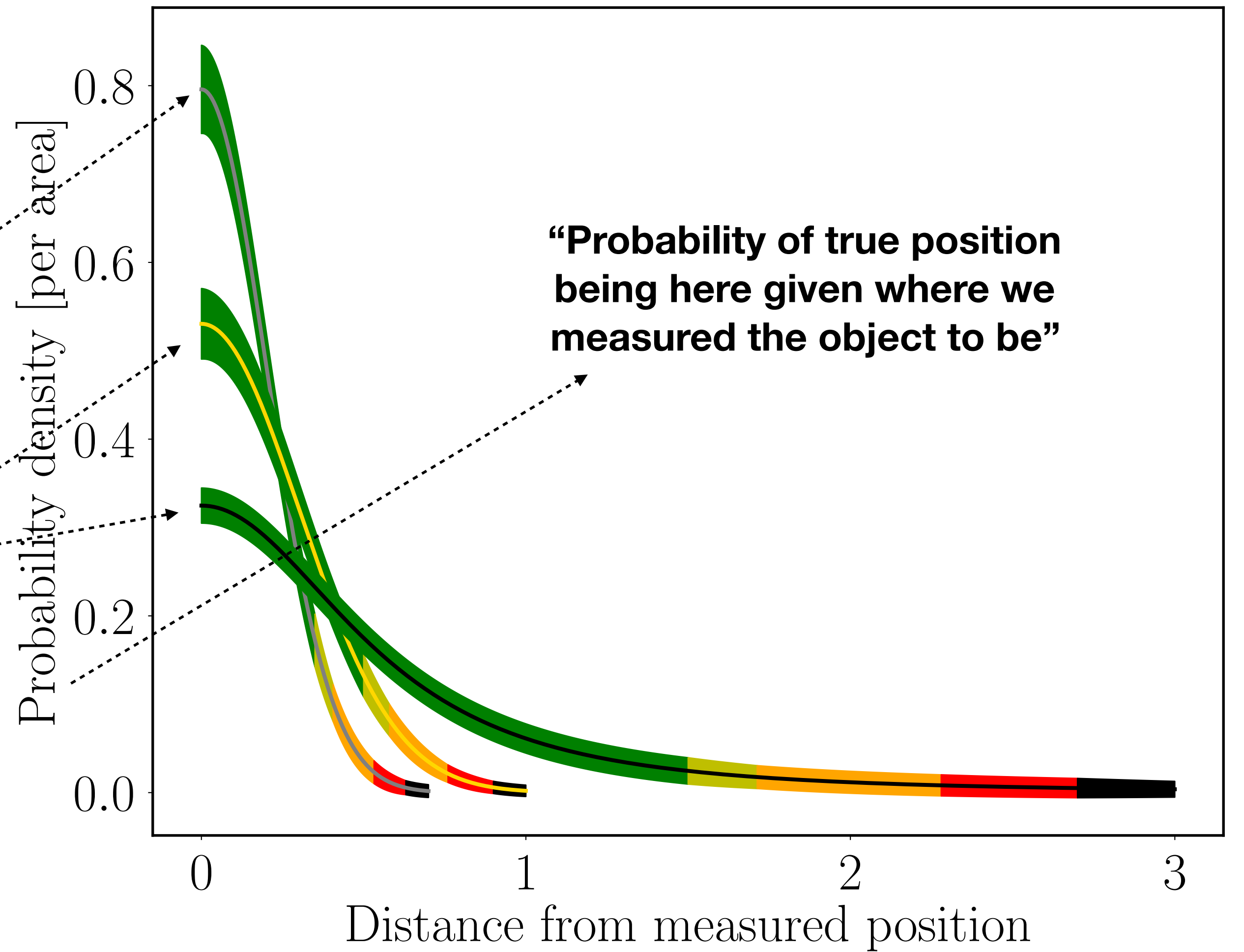
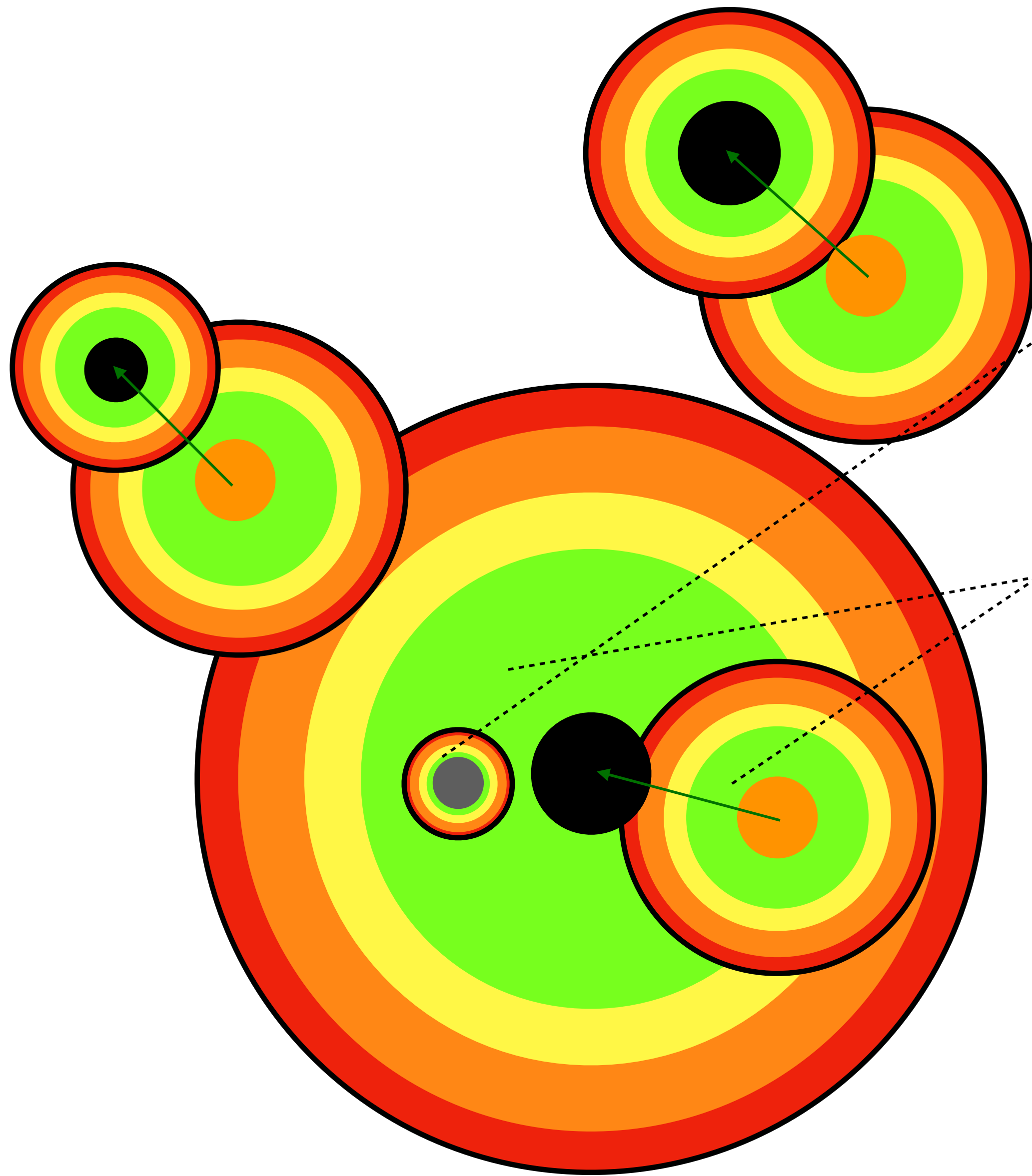


Wilson & Naylor (2017)
Wilson & Naylor (2018b)
WISE - Wright et al. (2010)
Gaia DR2 - Gaia Collaboration, Brown A. G. A., et al. (2018)

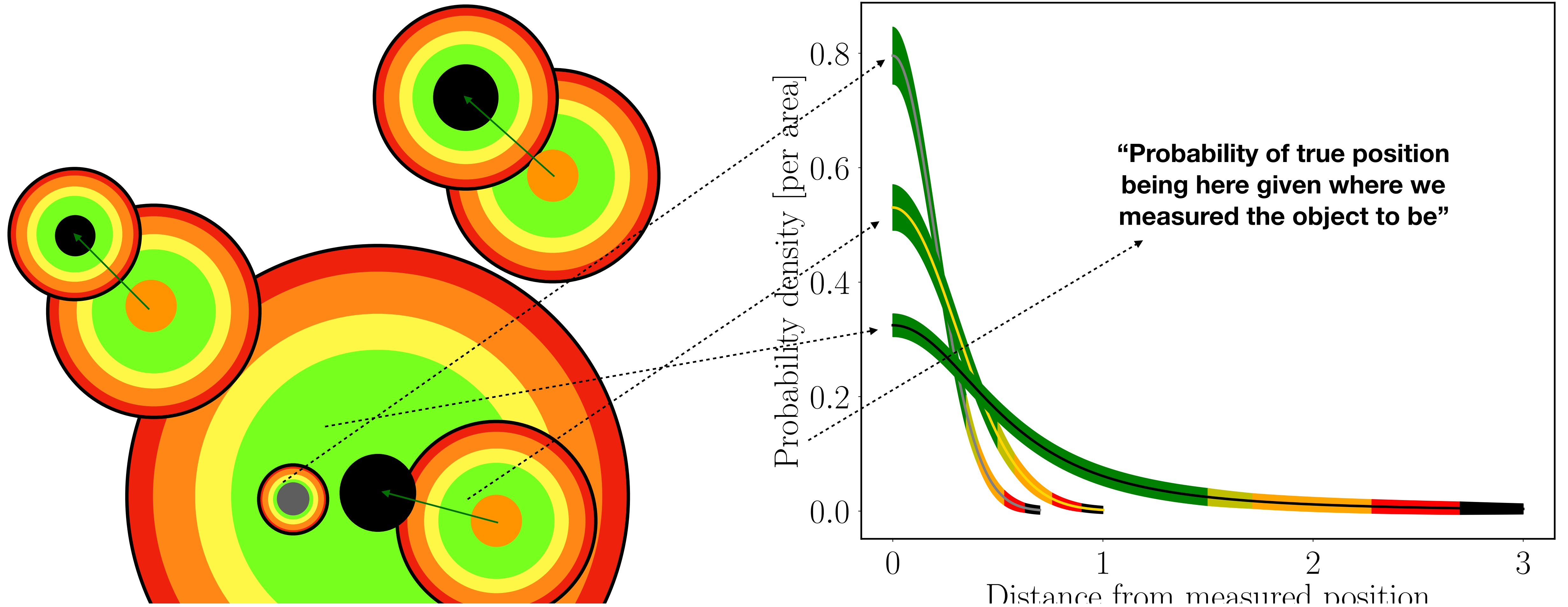
The AUF: Perturbation



The AUF: Position (Un)Certainty



The AUF: Position (Un)Certainty



One assumption made in all of previous works: source positions uncertainties are Gaussian!

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

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

de Ruiter, Willis, & Arp (1977)

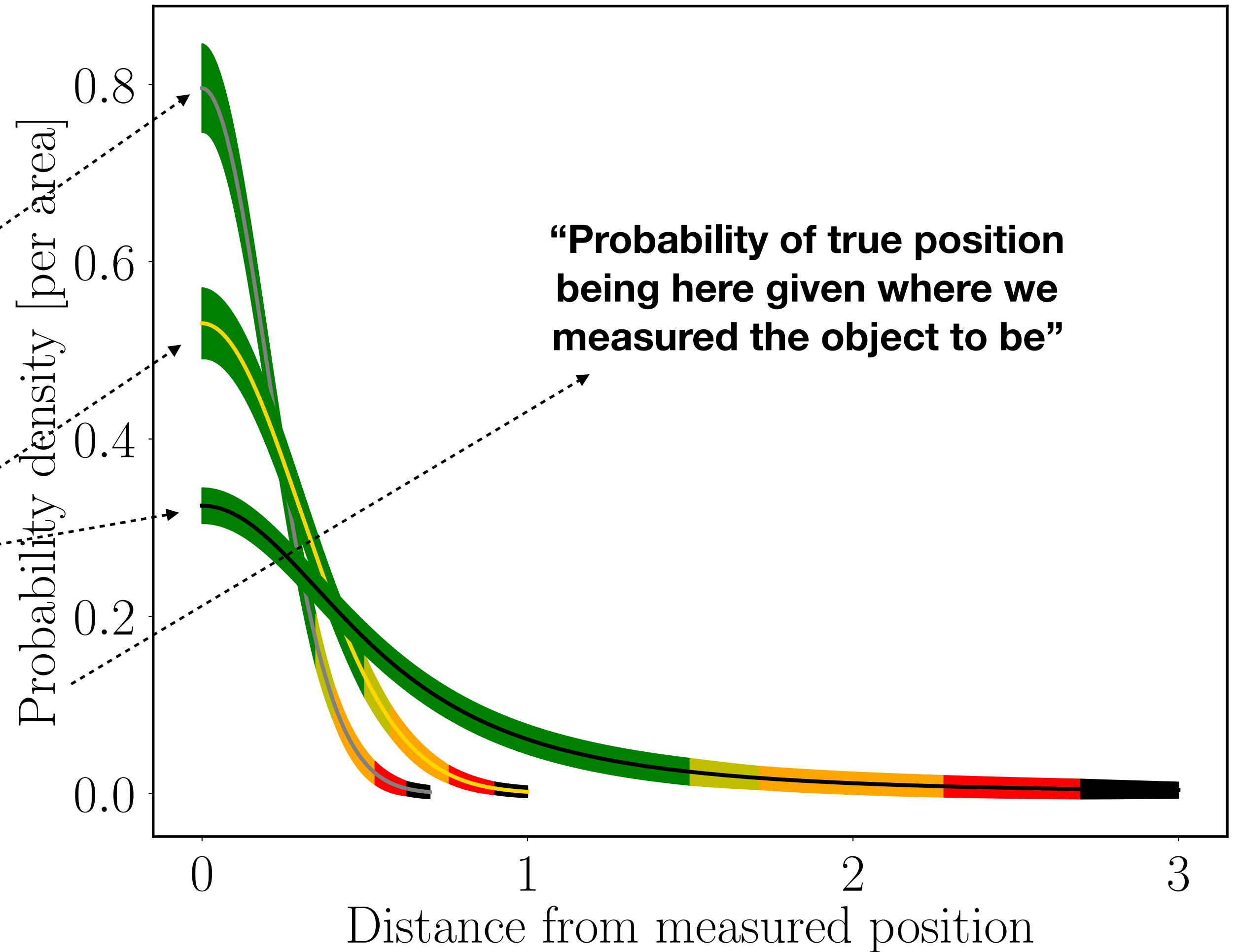
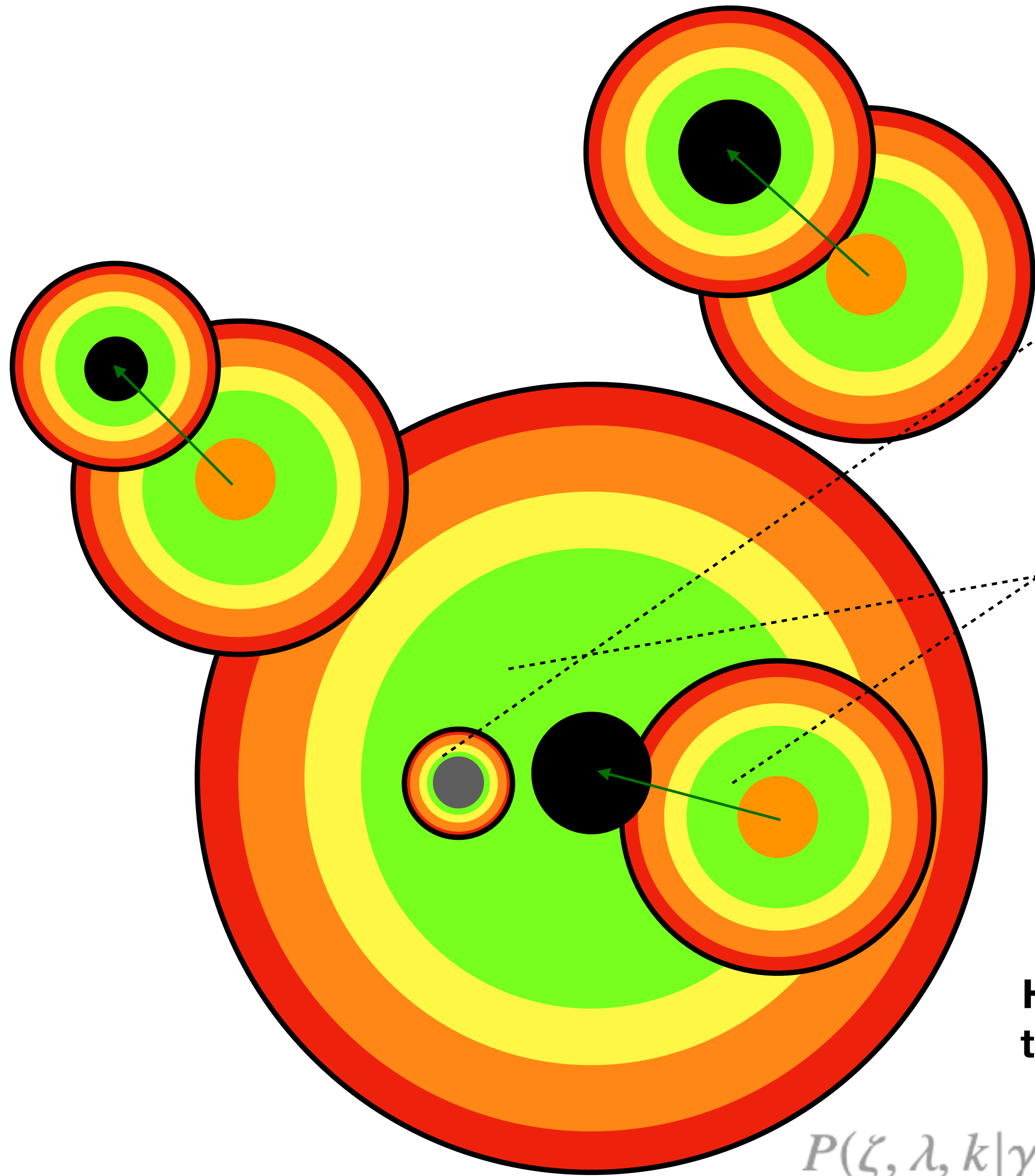
Naylor, Broos, & Feigelson (2013)

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

Budavári & Szalay (2008)

Tom J Wilson @onoddil

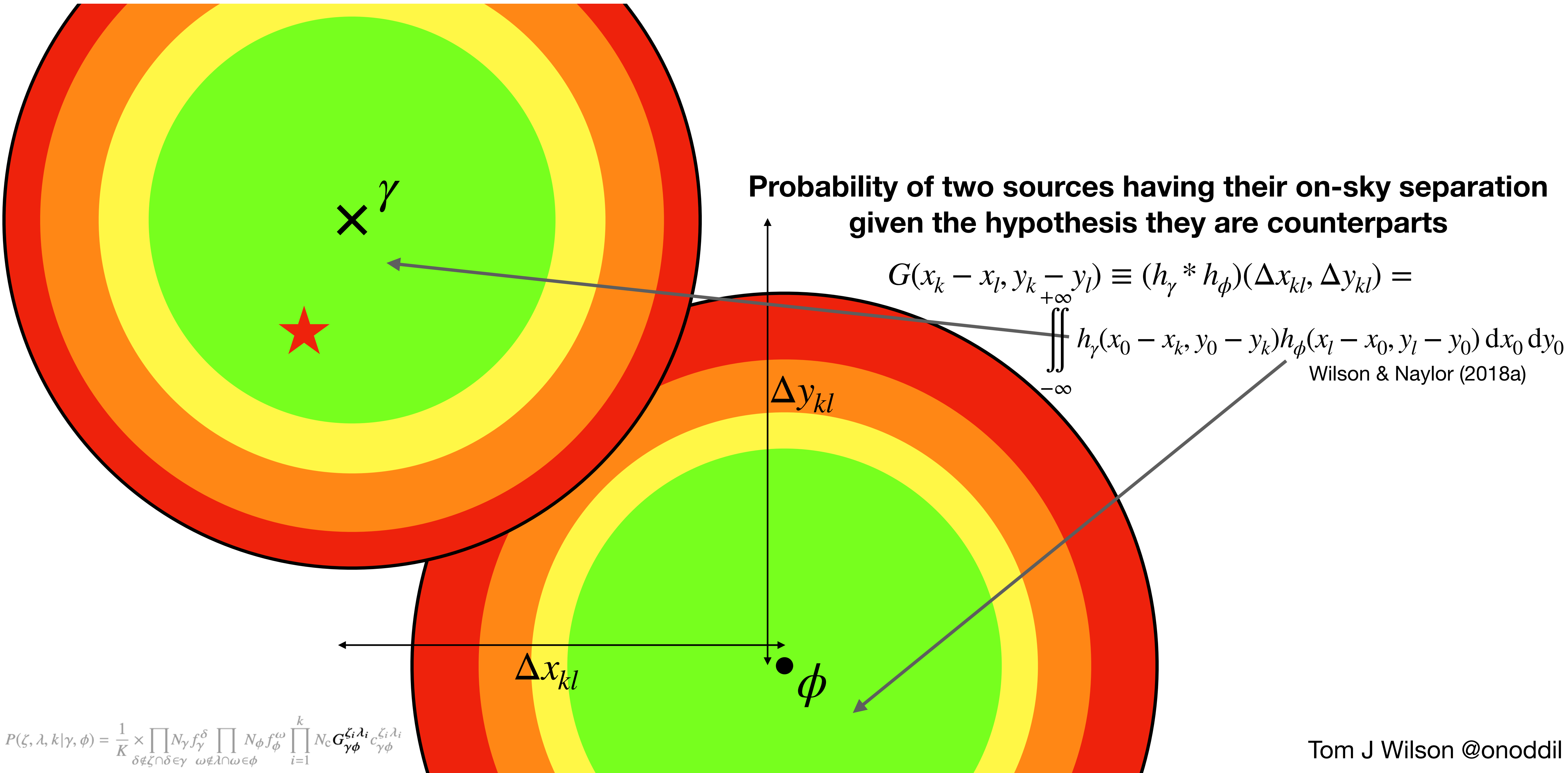
The AUF: Position (Un)Certainty



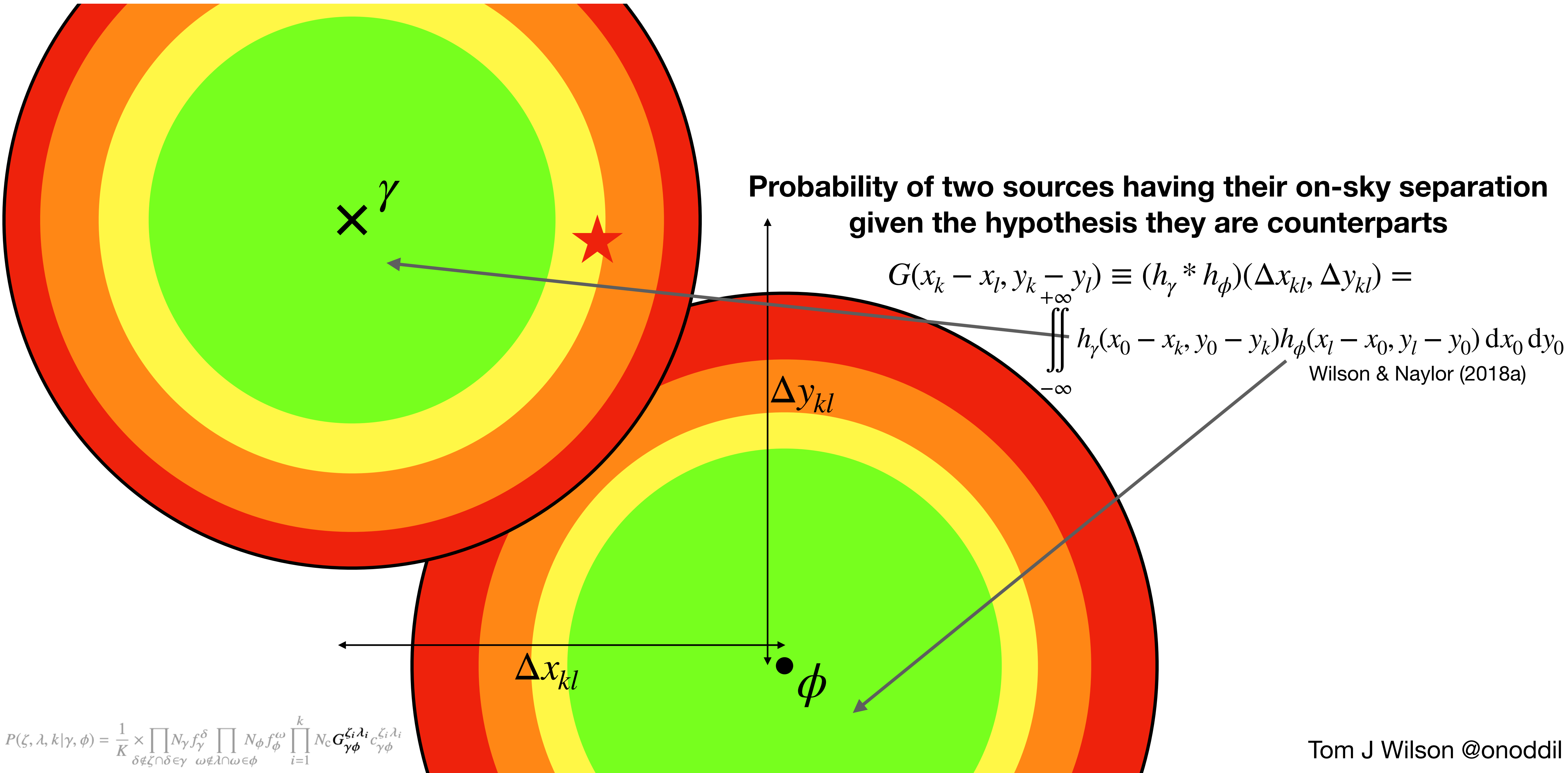
How can we calculate the *general* probability of two sources having their on-sky separation given the hypothesis they are counterparts?

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

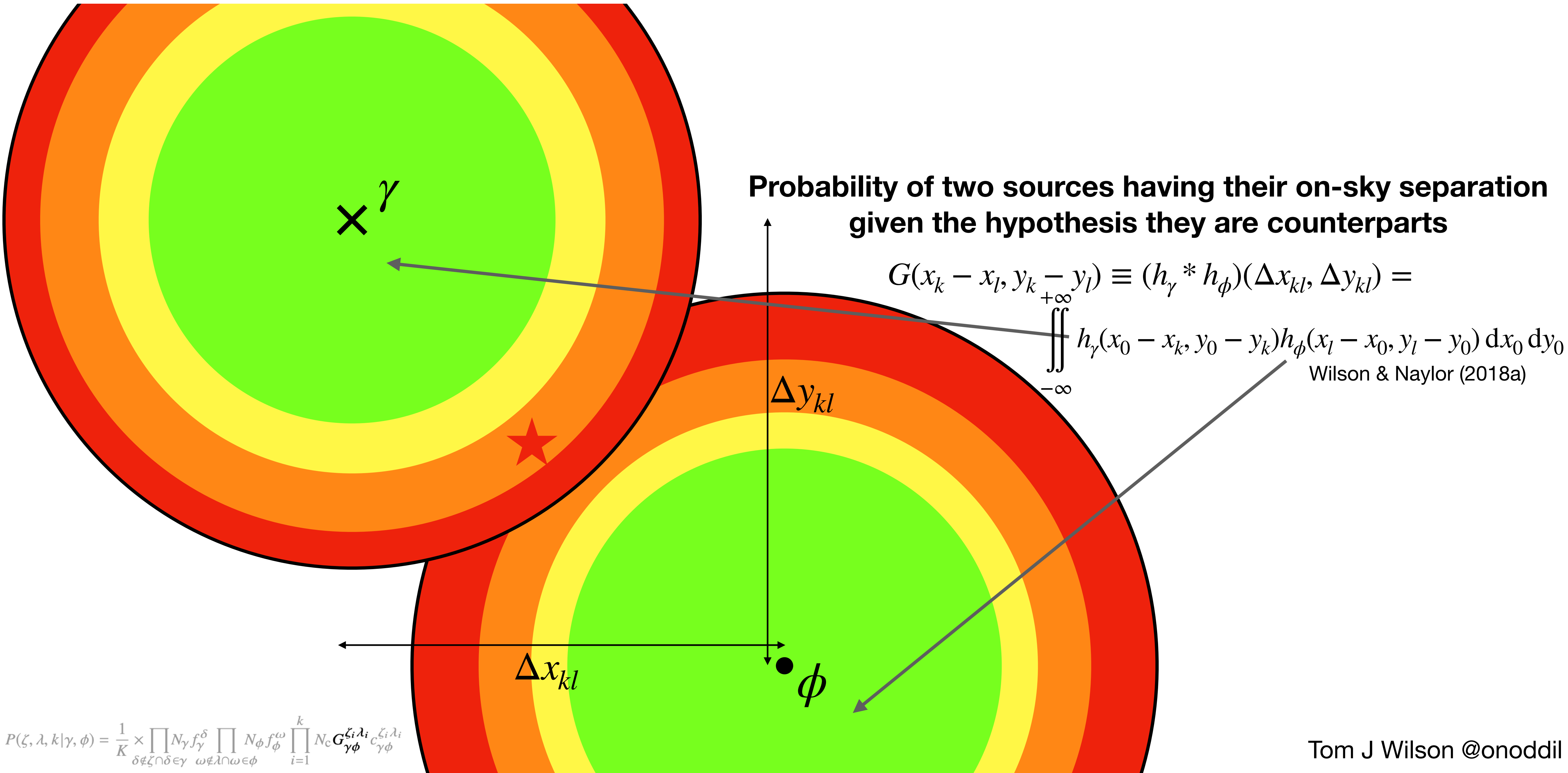
The AUF: Match Separation Probability



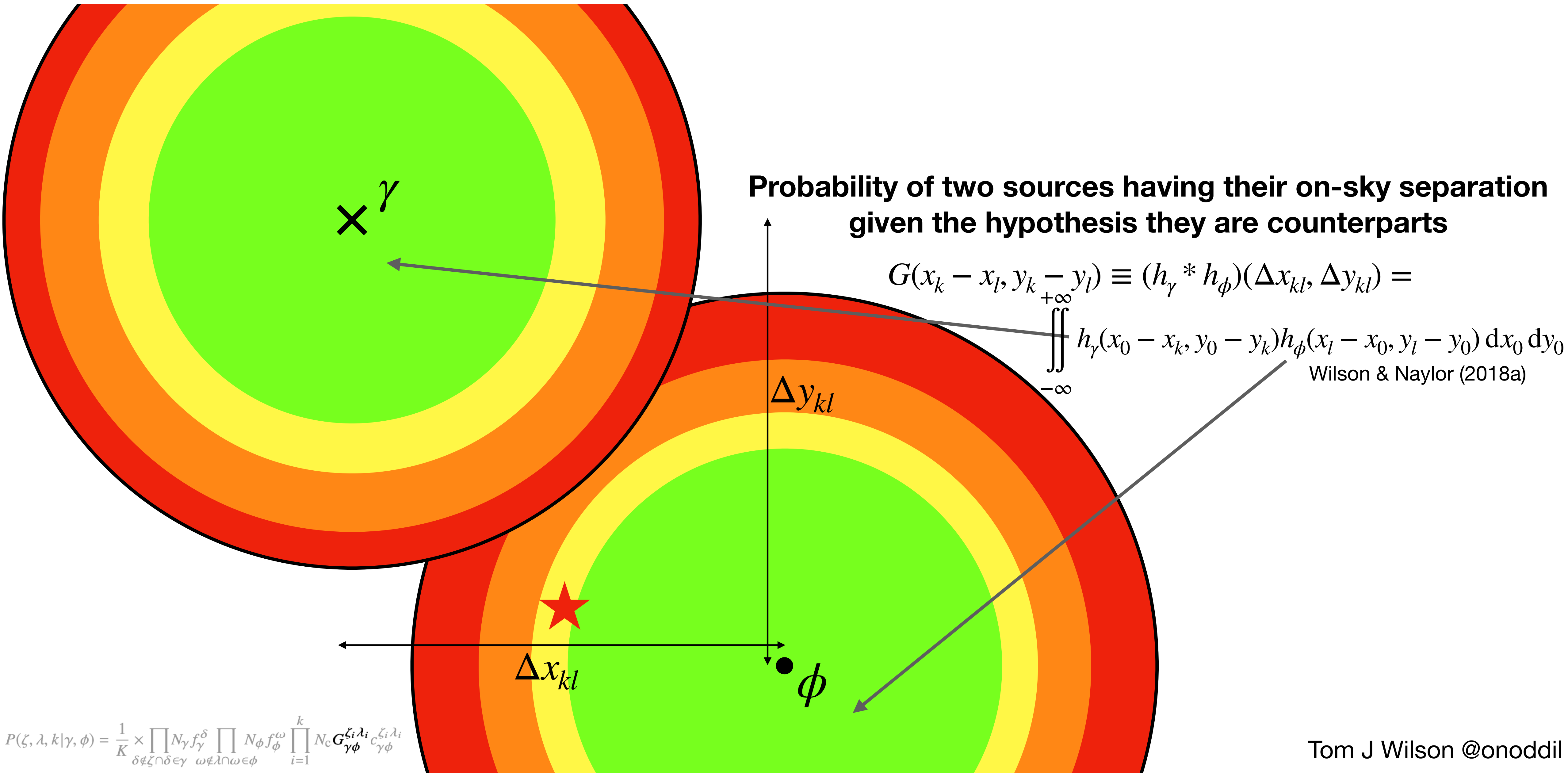
The AUF: Match Separation Probability



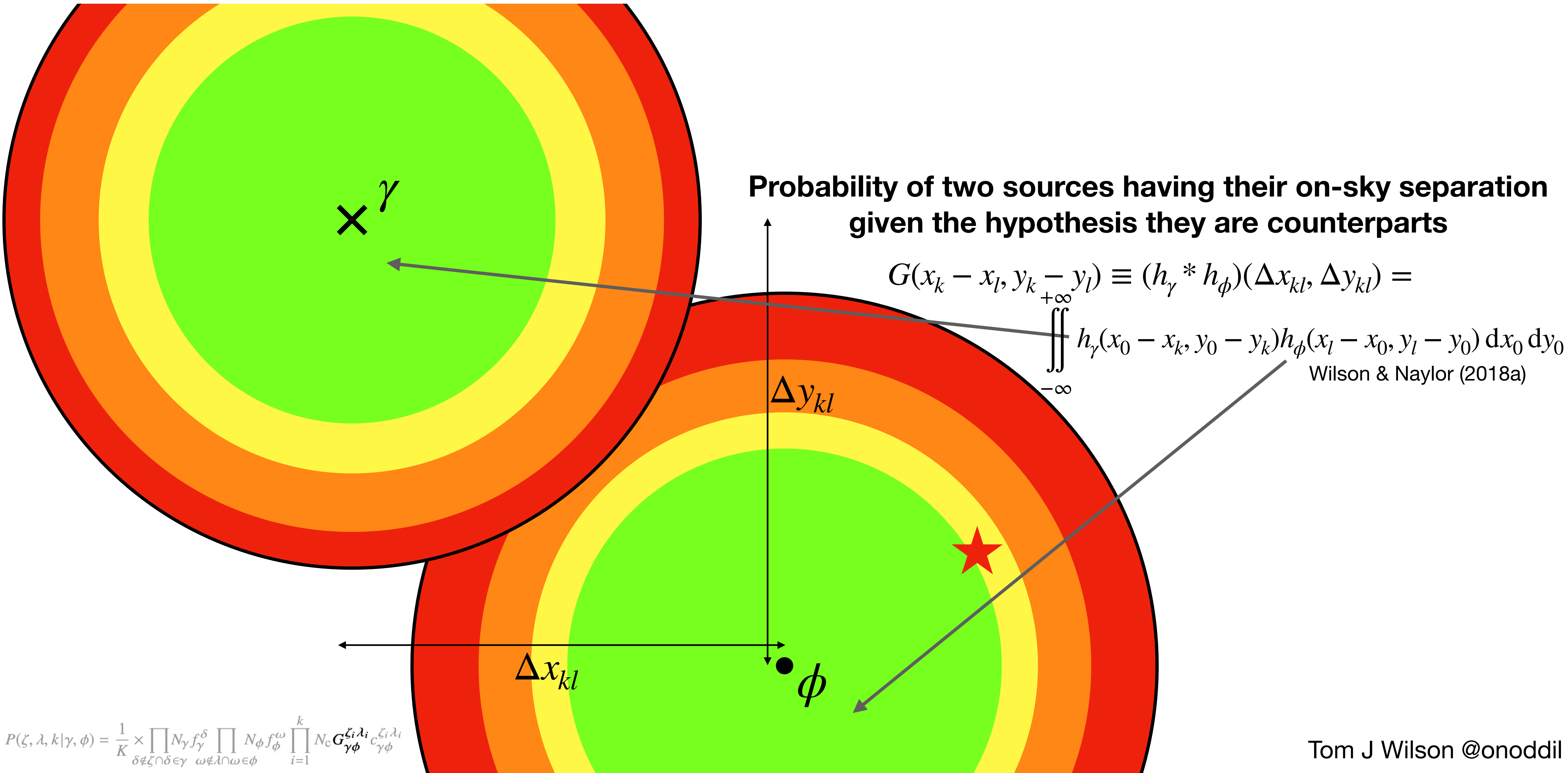
The AUF: Match Separation Probability



The AUF: Match Separation Probability

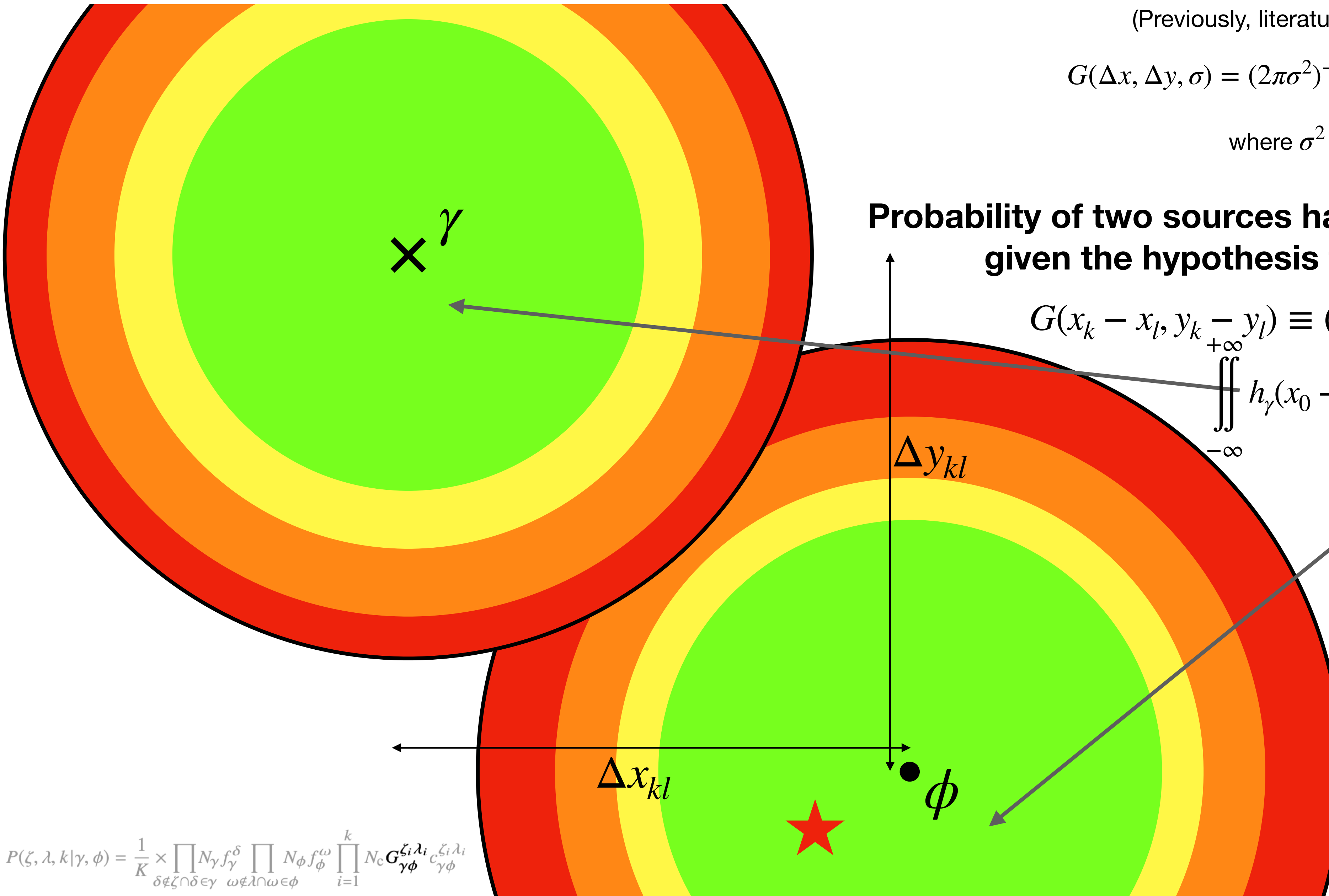


The AUF: Match Separation Probability



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

The AUF: Match Separation Probability



(Previously, literature assumed that e.g.

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

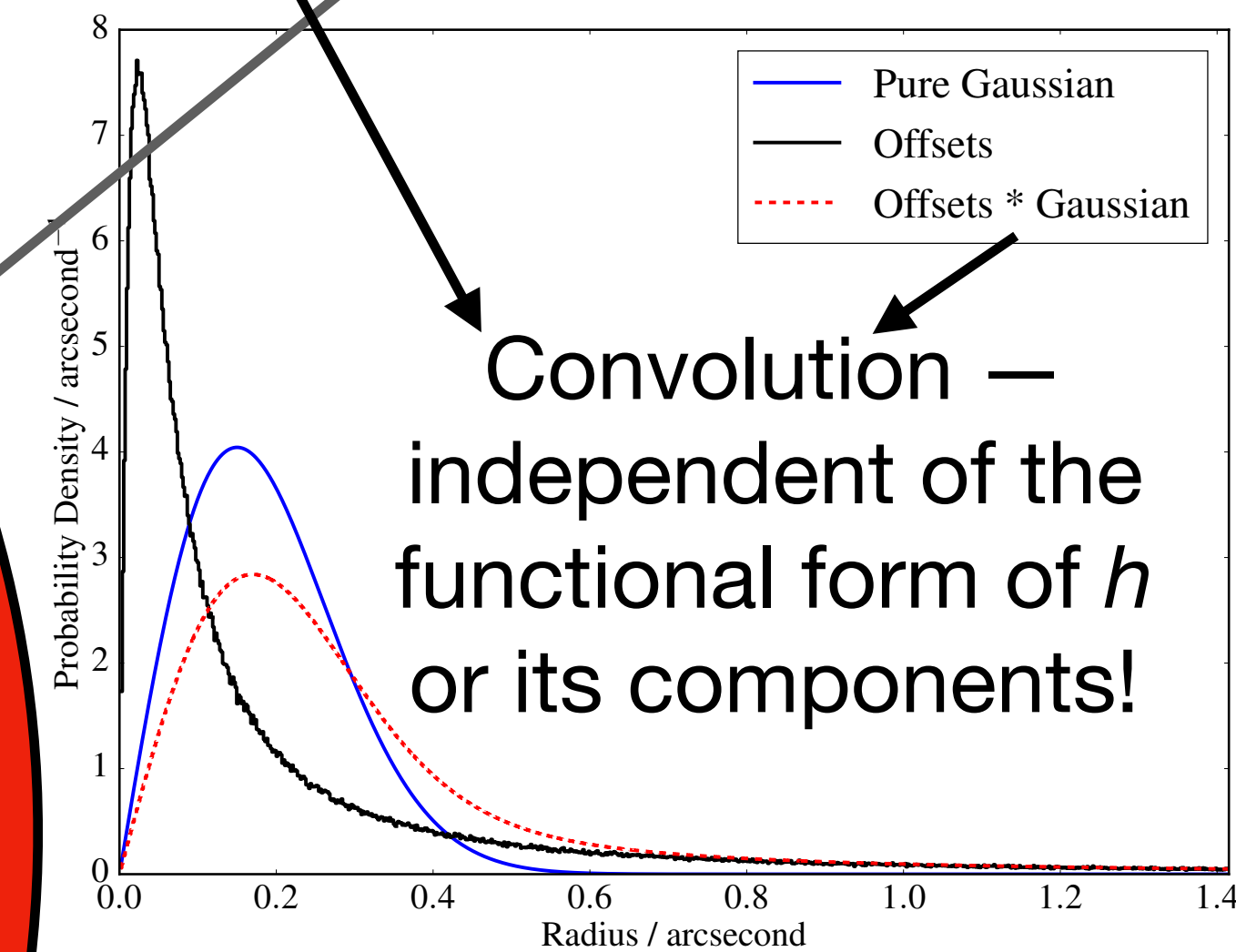
where $\sigma^2 = \sigma_\gamma^2 + \sigma_\phi^2$)

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

$$G(x_k - x_l, y_k - y_l) \equiv (h_\gamma * h_\phi)(\Delta x_{kl}, \Delta y_{kl}) =$$

$$\iint_{-\infty}^{+\infty} h_\gamma(x_0 - x_k, y_0 - y_k) h_\phi(x_l - x_0, y_l - y_0) dx_0 dy_0$$

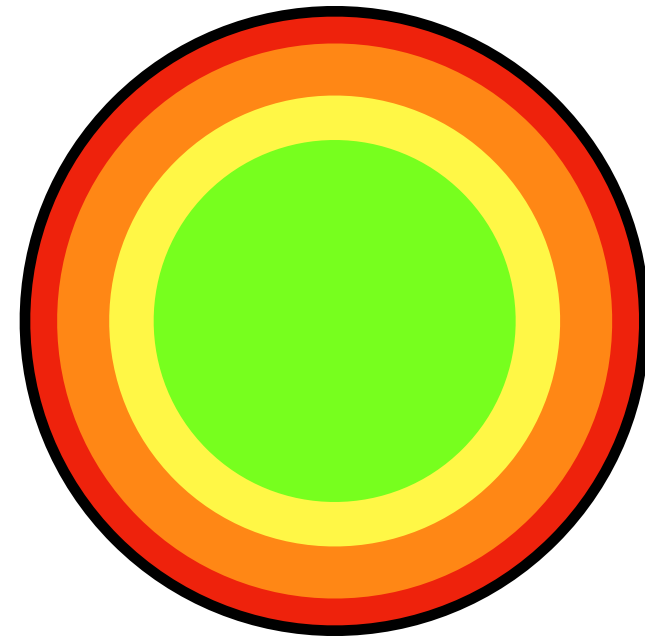
Wilson & Naylor (2018a)



Convolution — independent of the functional form of h or its components!

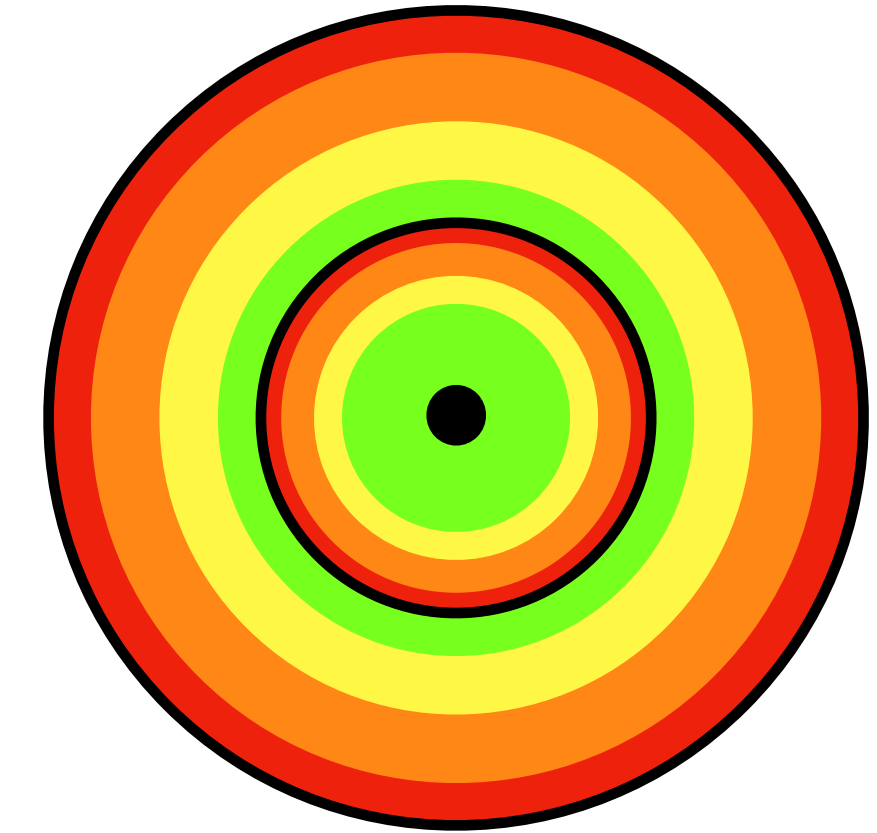
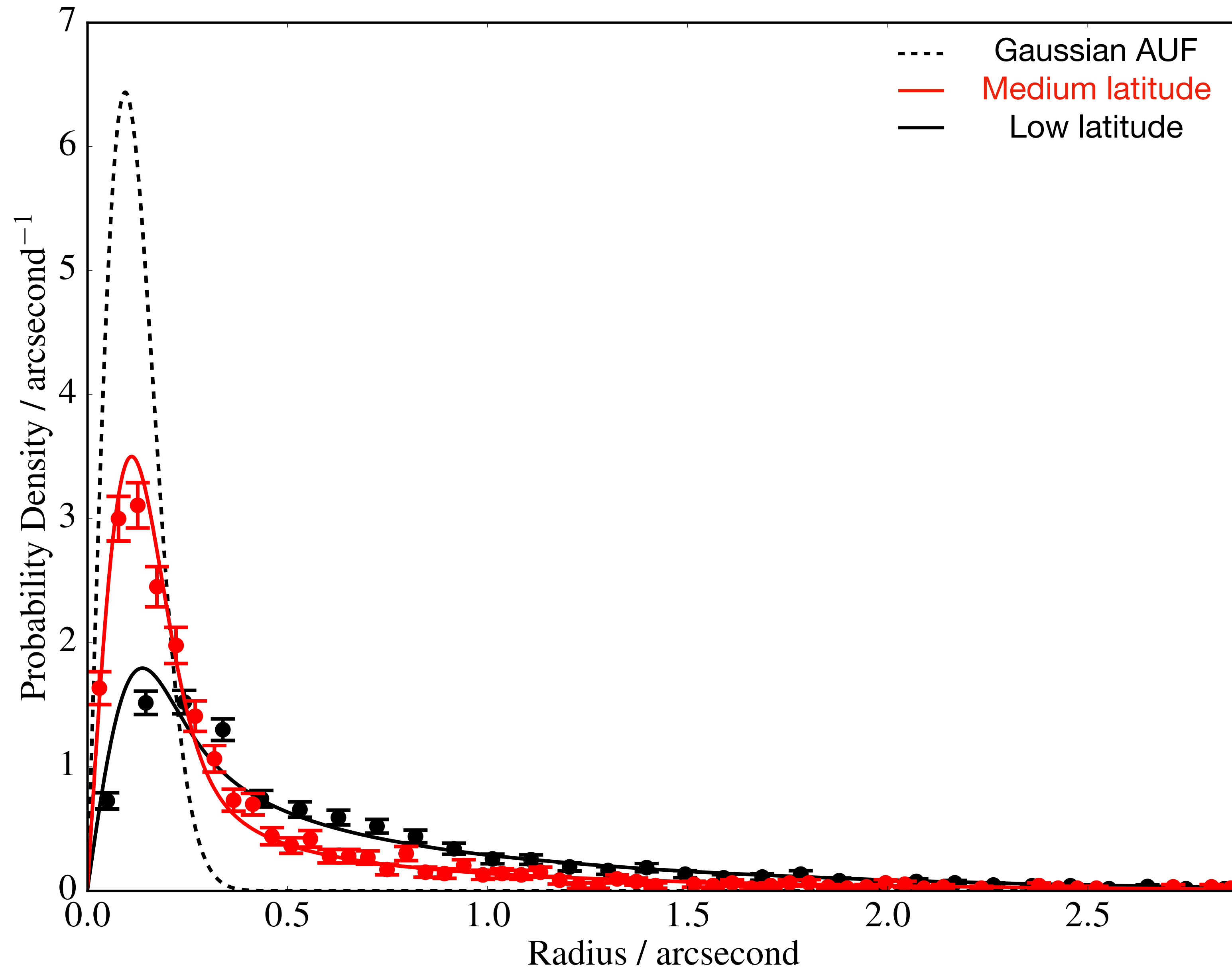
$$P(\zeta, \lambda, k | \gamma, \phi) = \frac{1}{K} \times \prod_{\delta \neq \zeta \cap \delta \in \gamma} N_\gamma f_\gamma^\delta \prod_{\omega \neq \lambda \cap \omega \in \phi} N_\phi f_\phi^\omega \prod_{i=1}^k N_c G_{\gamma\phi}^{\zeta_i \lambda_i} c_{\gamma\phi}^{\zeta_i \lambda_i}$$

The Astrometric Uncertainty Function

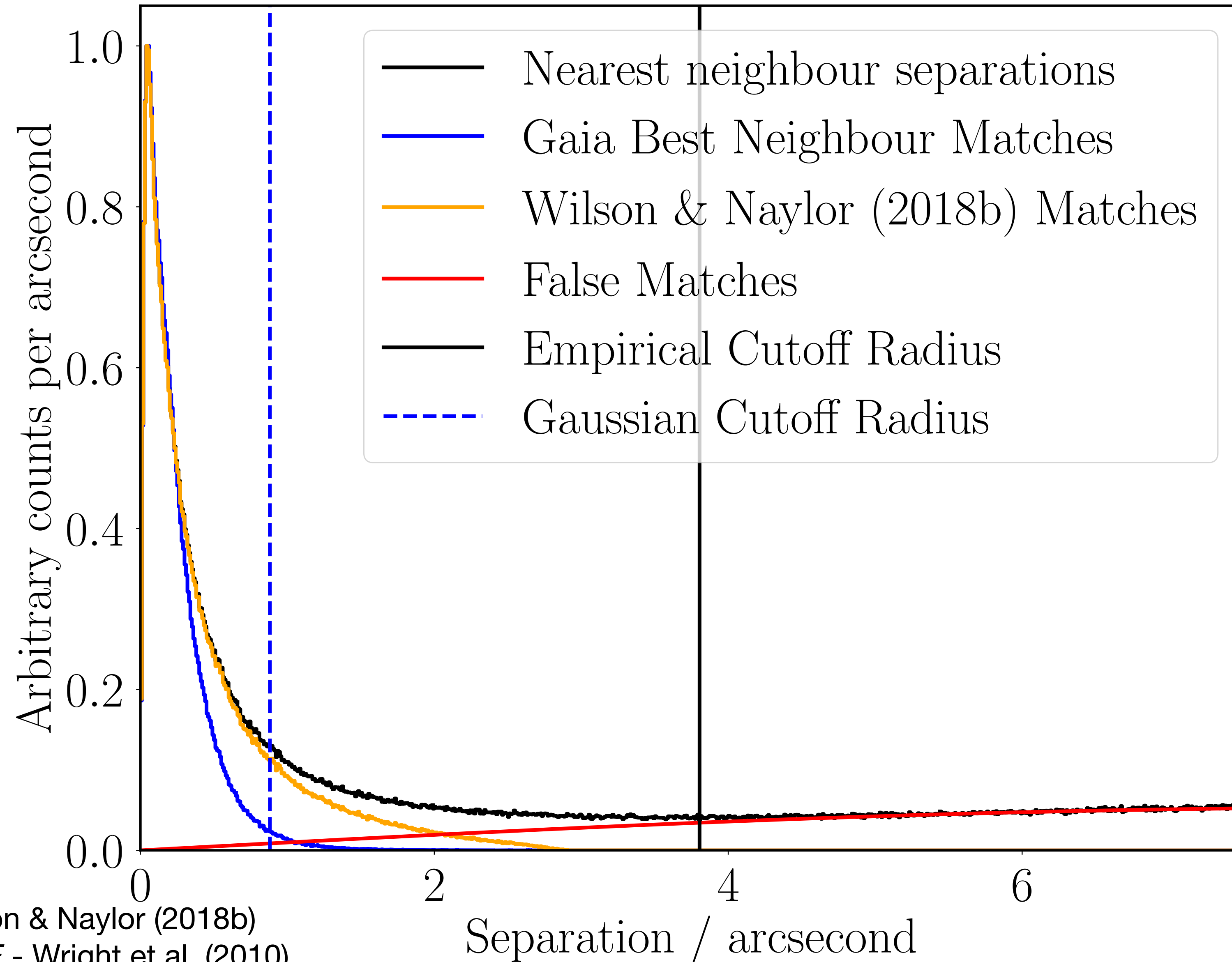


The AUF does not need to, and in fact quite often should not, be Gaussian!

The Effect of Crowding on *Gaia*-*WISE* Matches



The Effect of Crowding on *Gaia-WISE* Matches



If this effect was not taken into account, we would be incorrectly led to believe 50% of *Gaia-WISE** sources were not matches!

*“*Euclid*-Rubin”

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)

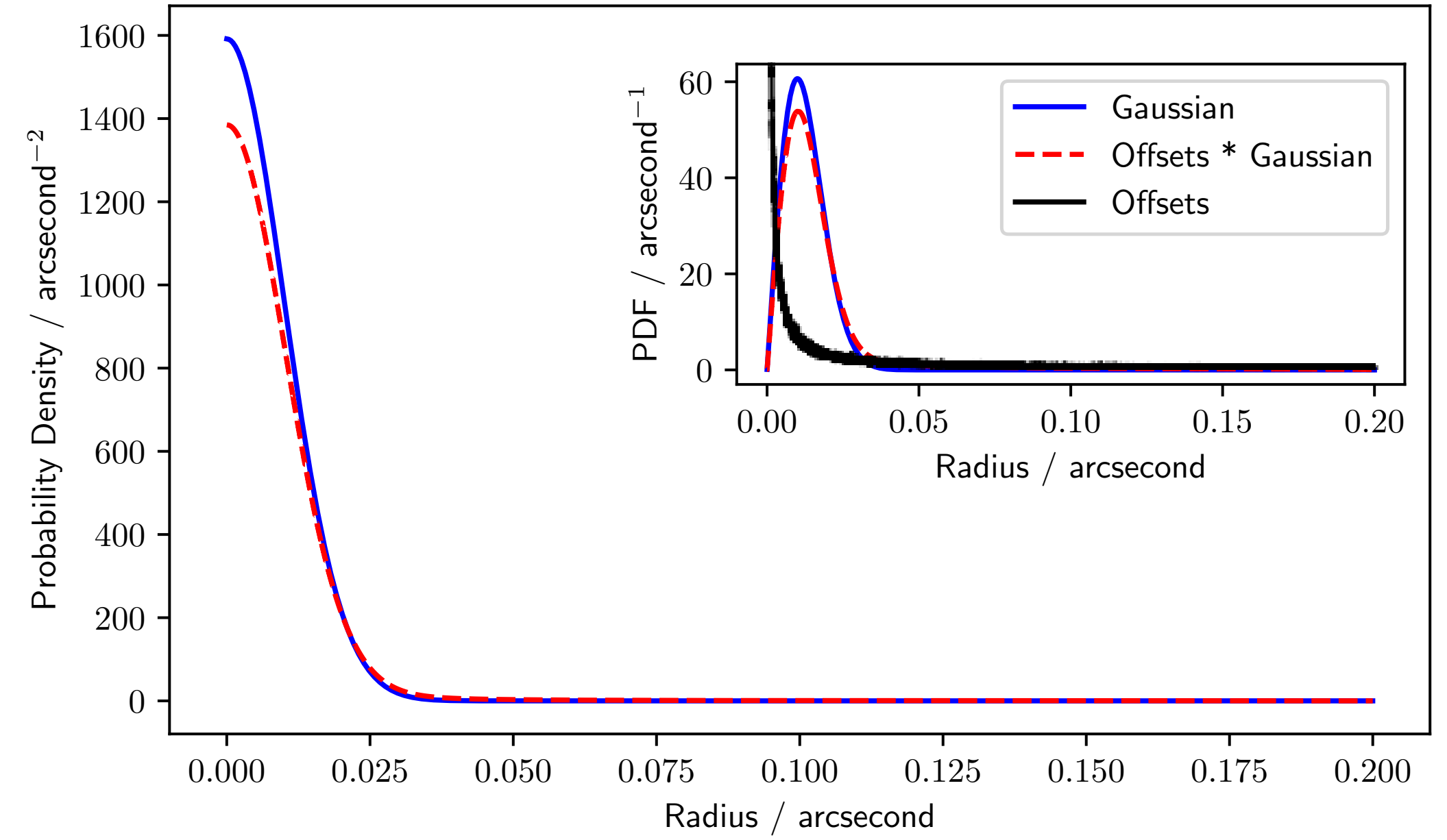
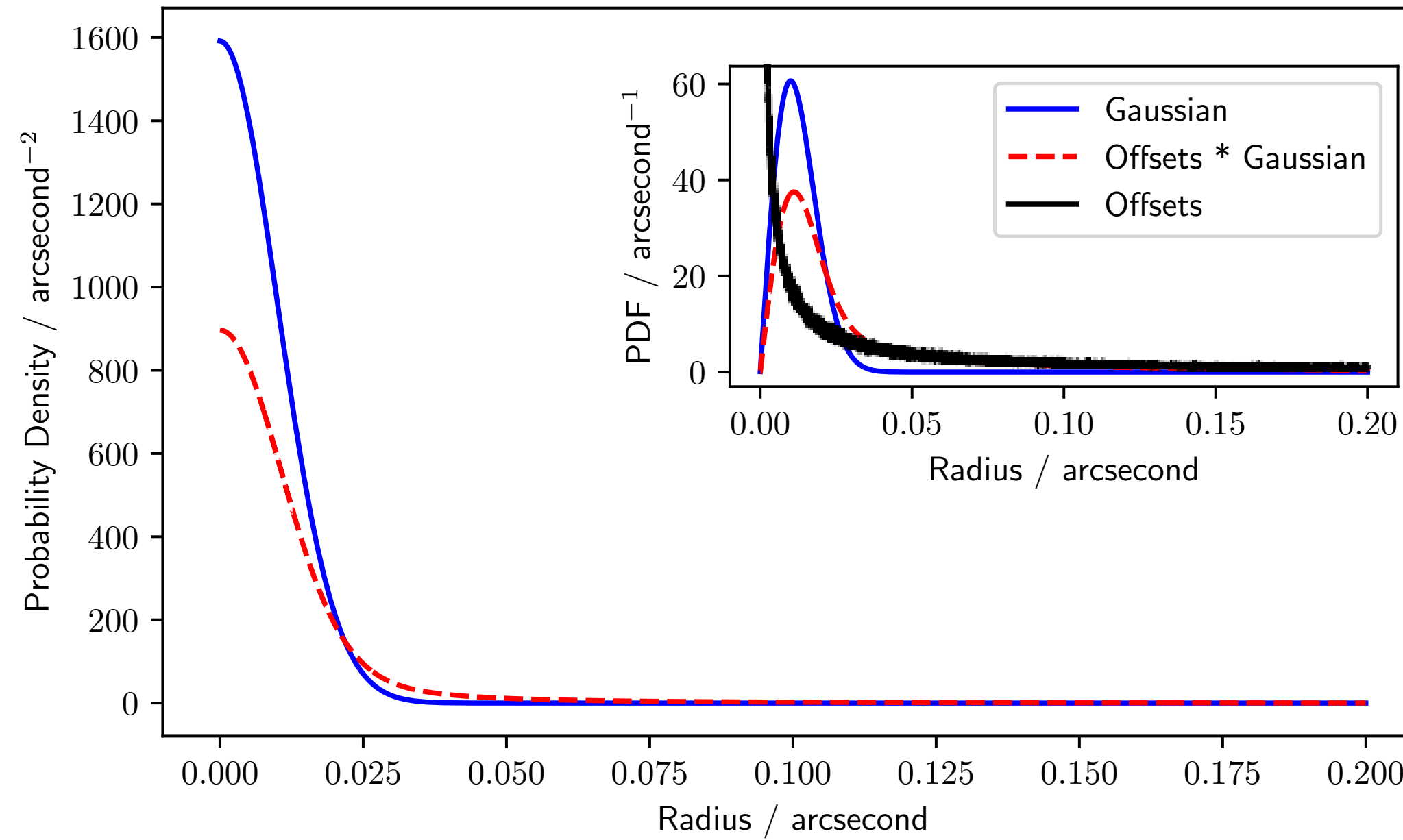
Tom J Wilson @onoddil

The Rubin AUF: Galactic Plane

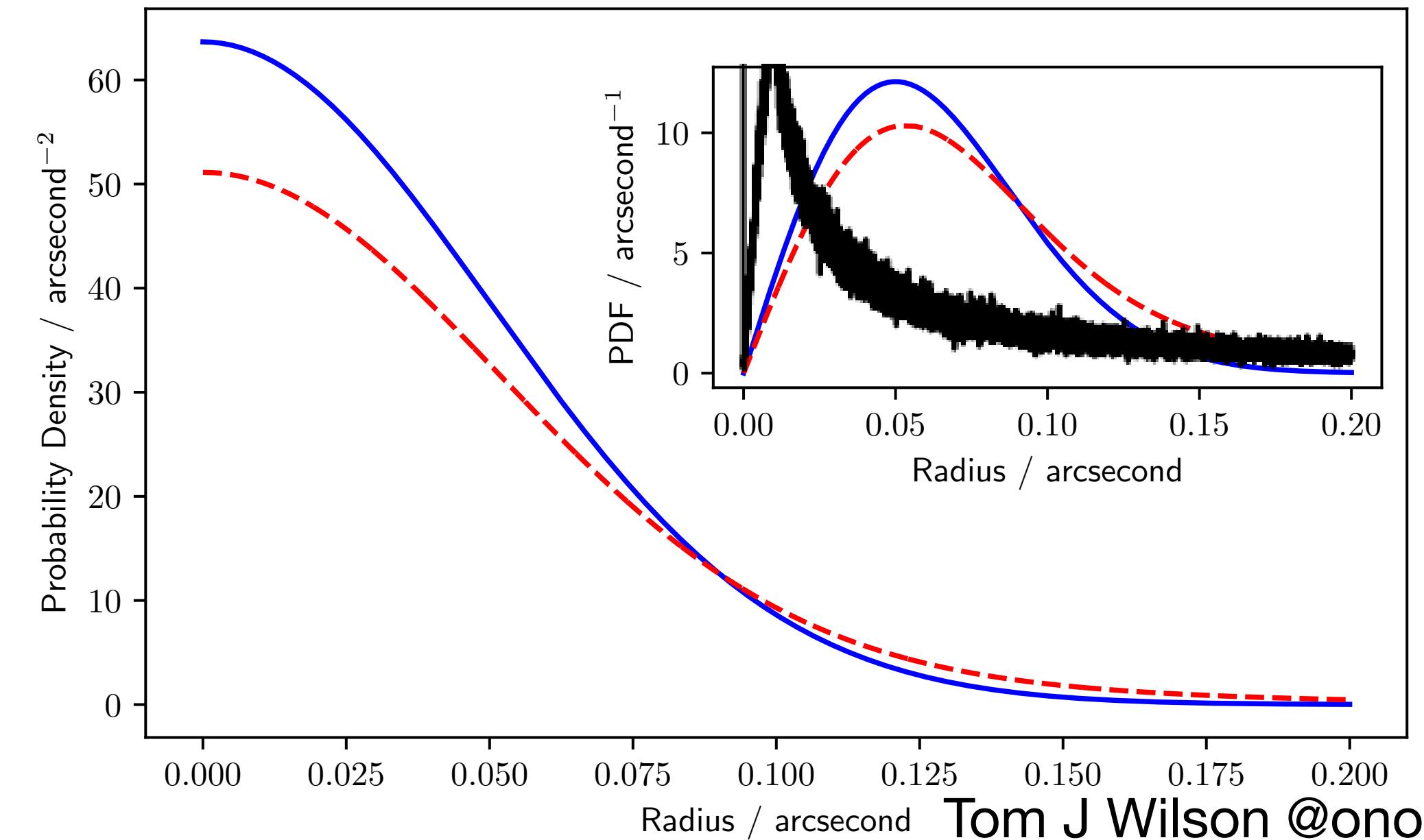
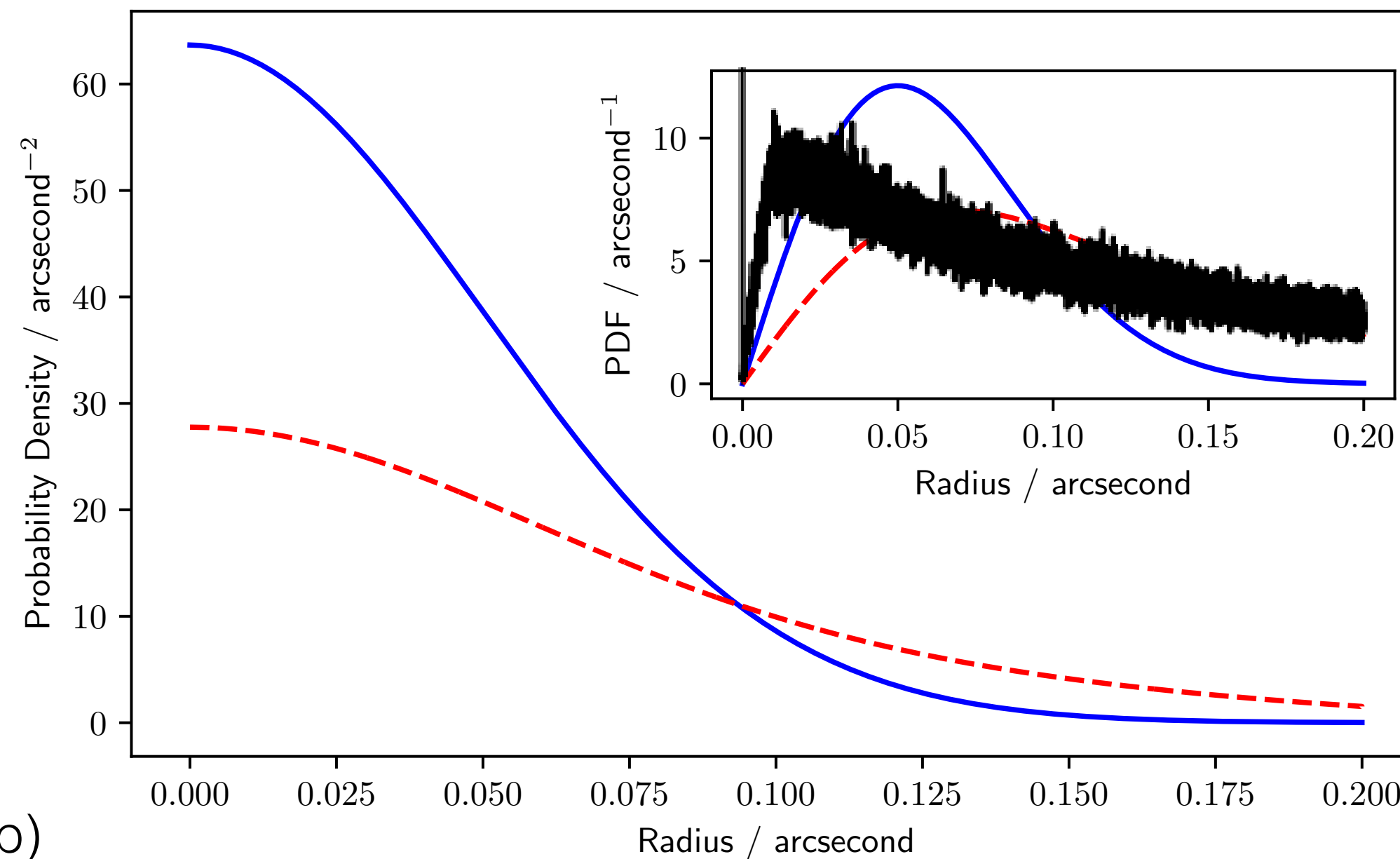
Galactic Centre

Not the Galactic Centre

Single-visit



Co-add

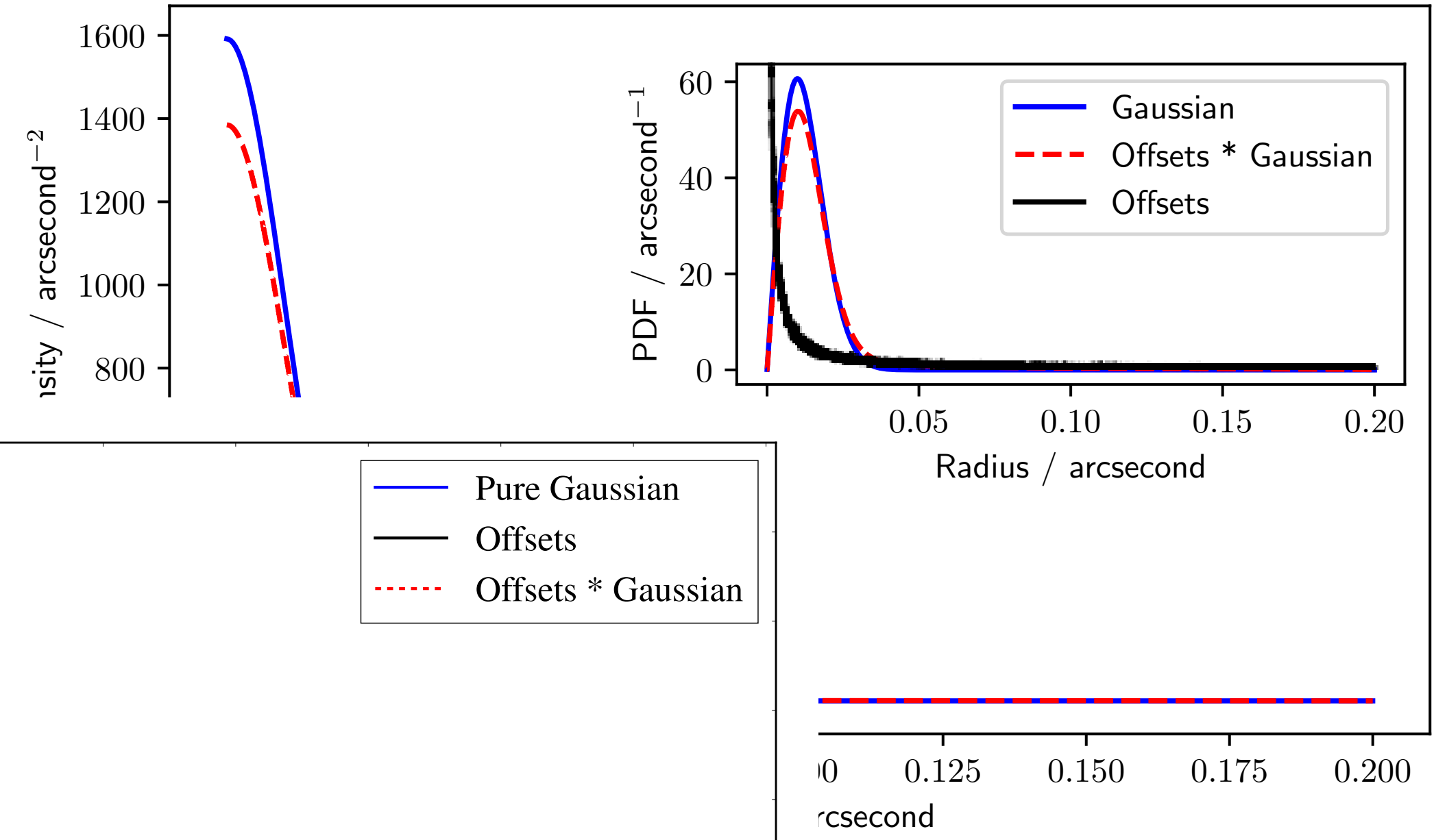
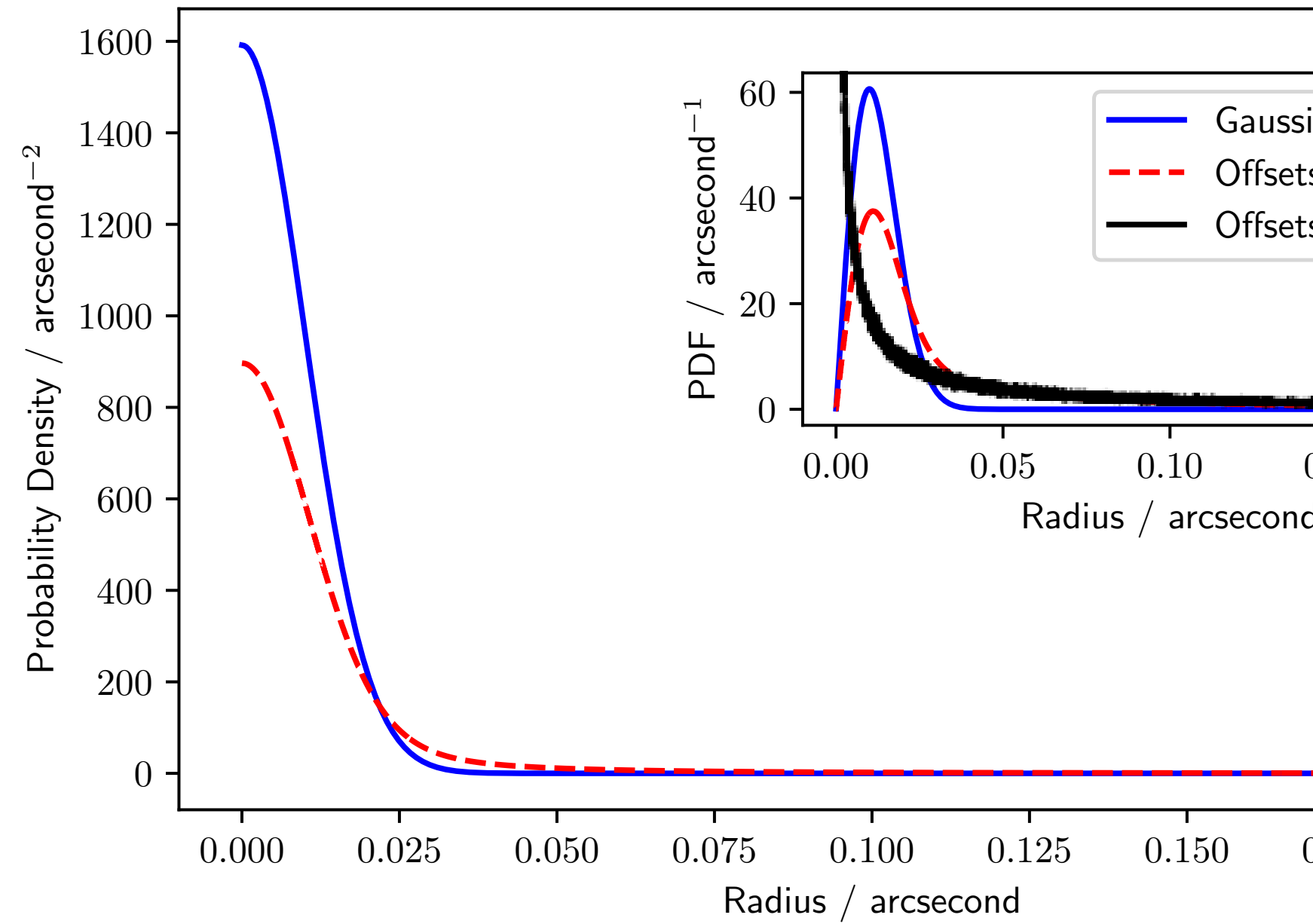


The Rubin AUF: Galactic Plane

Galactic Centre

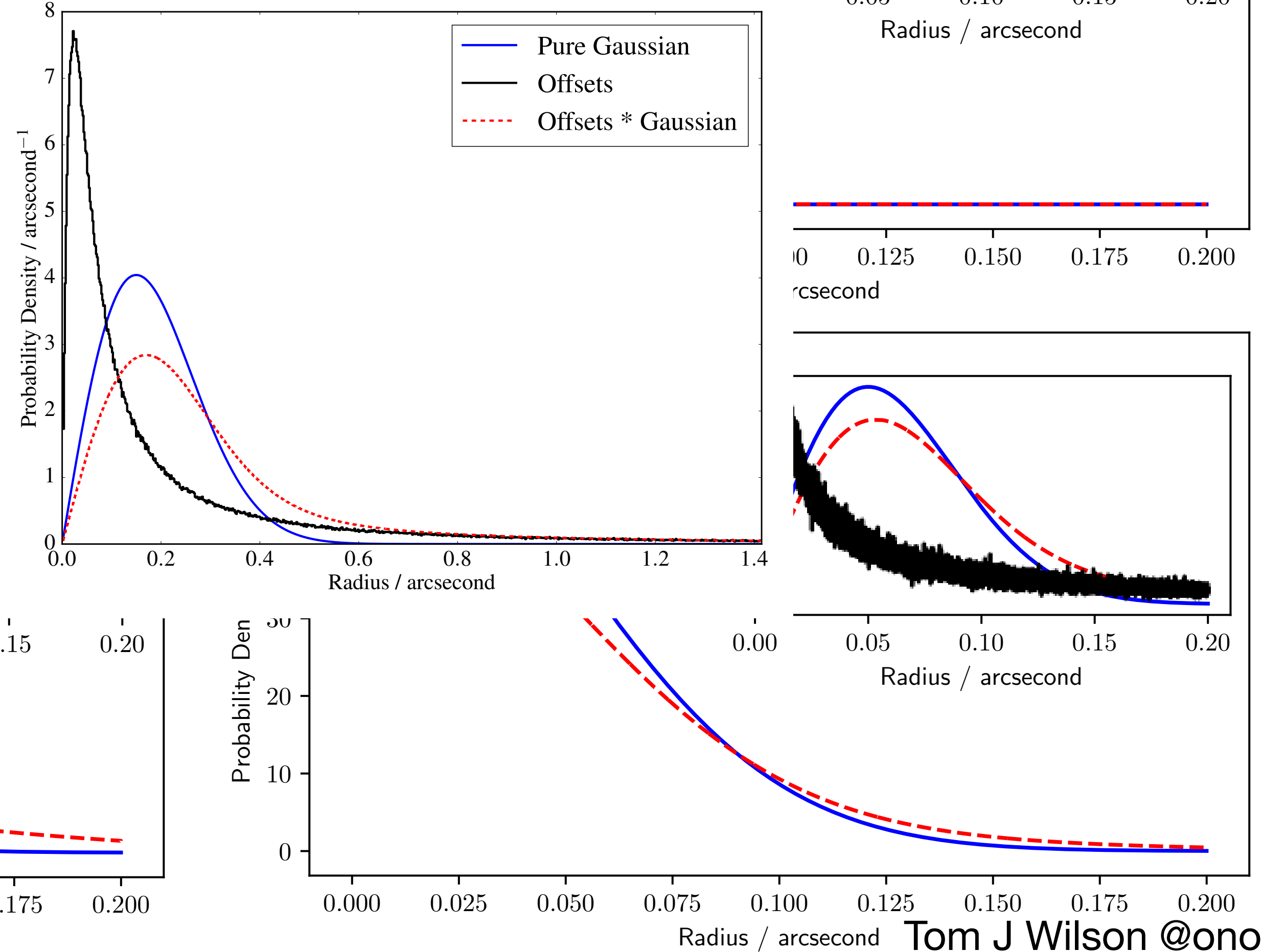
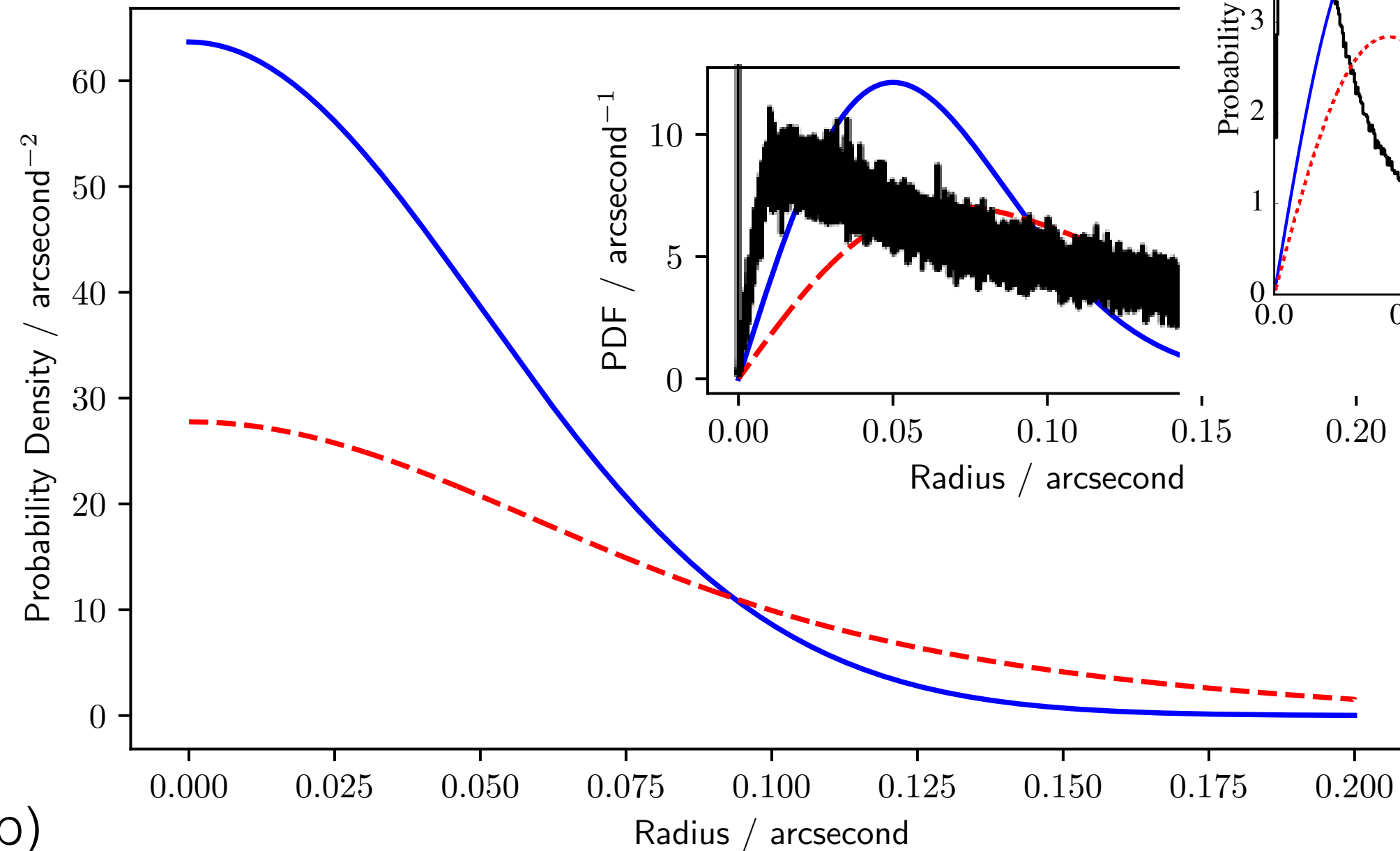
Not the Galactic Centre

Single-visit

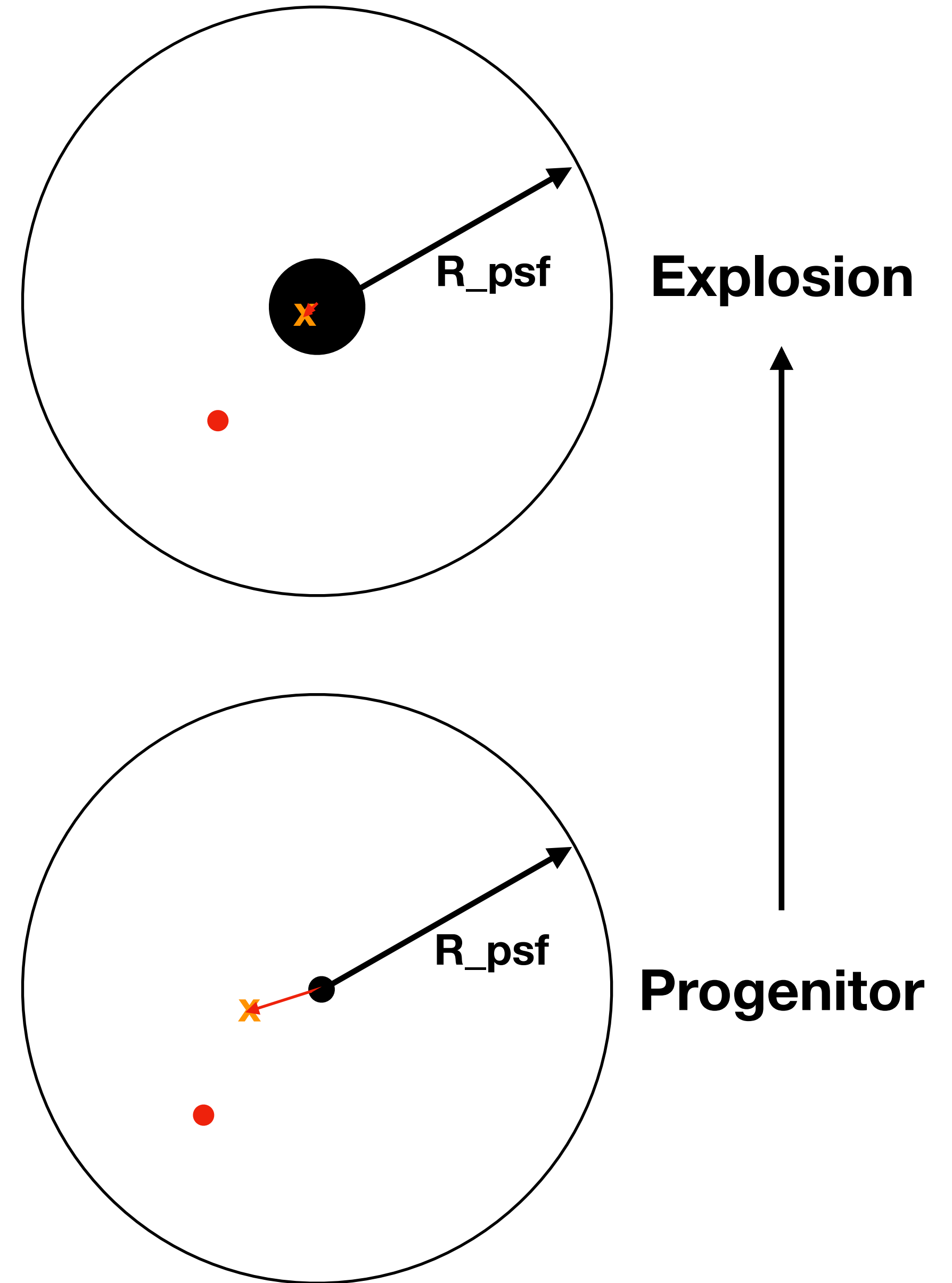
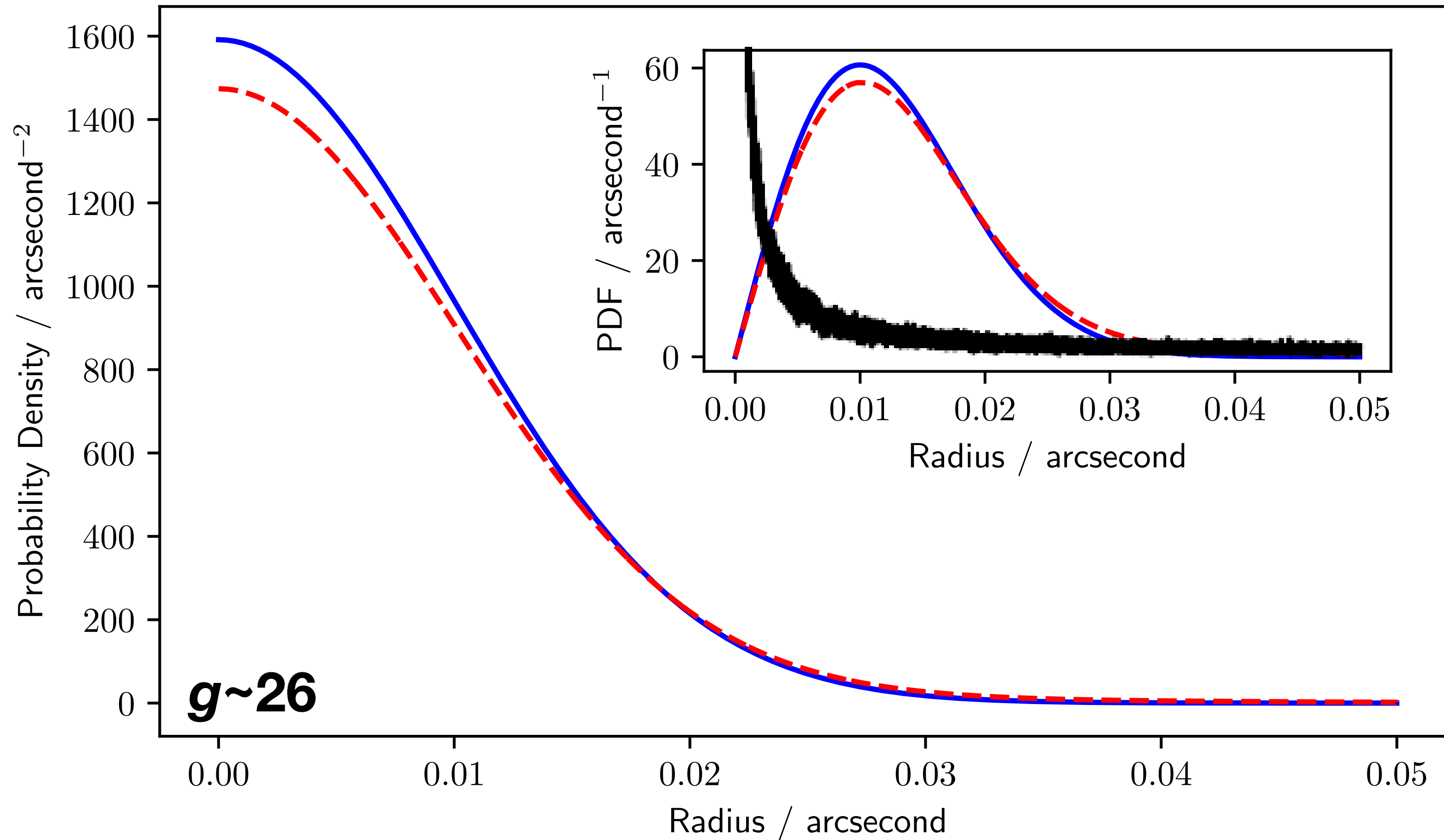


LSST will suffer approximately the same number of unresolved contaminants per PSF area as *WISE*! The Perturbation component of the AUF will overwhelm the Centroid component for a large % of the Galactic Plane.

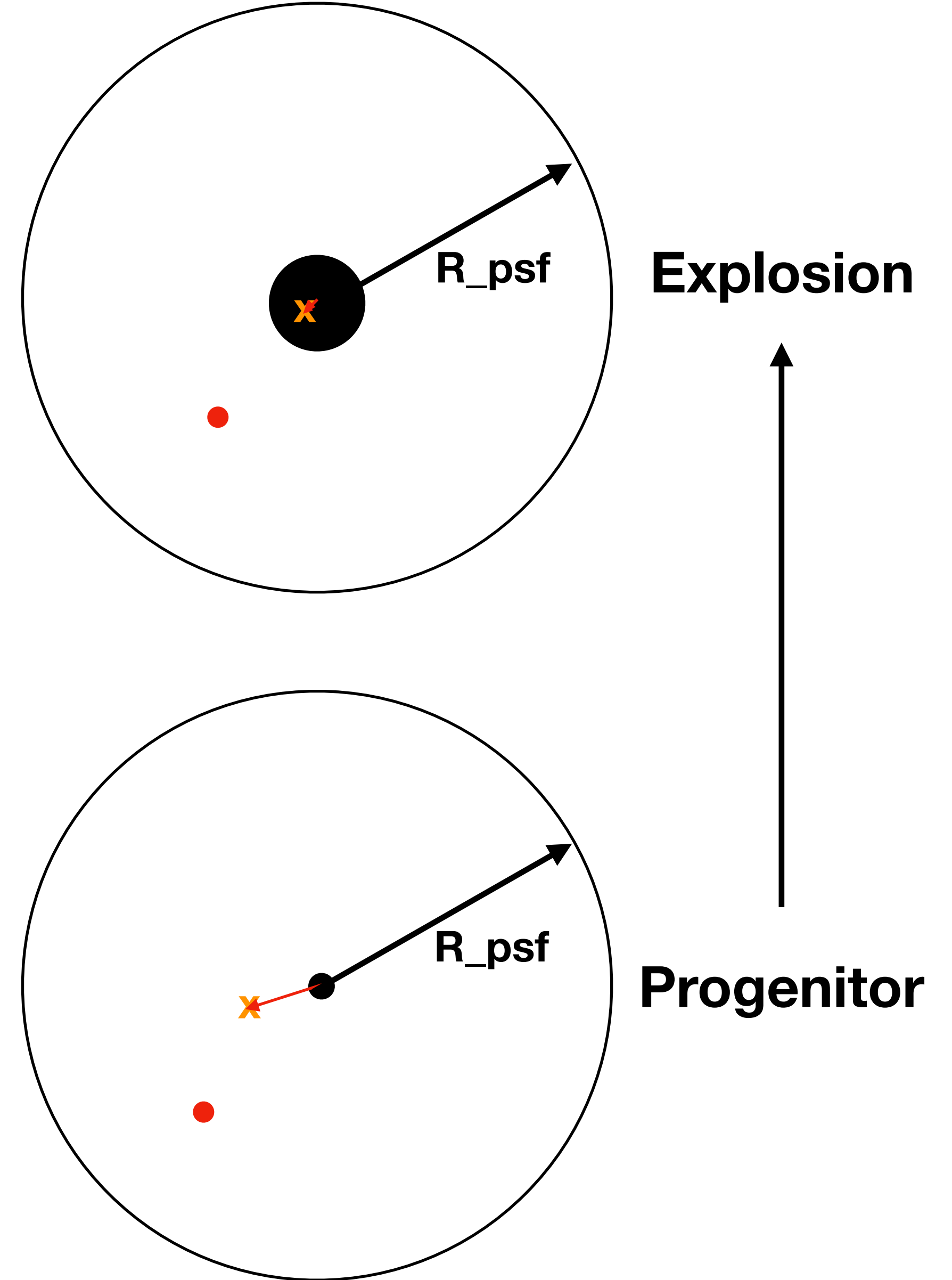
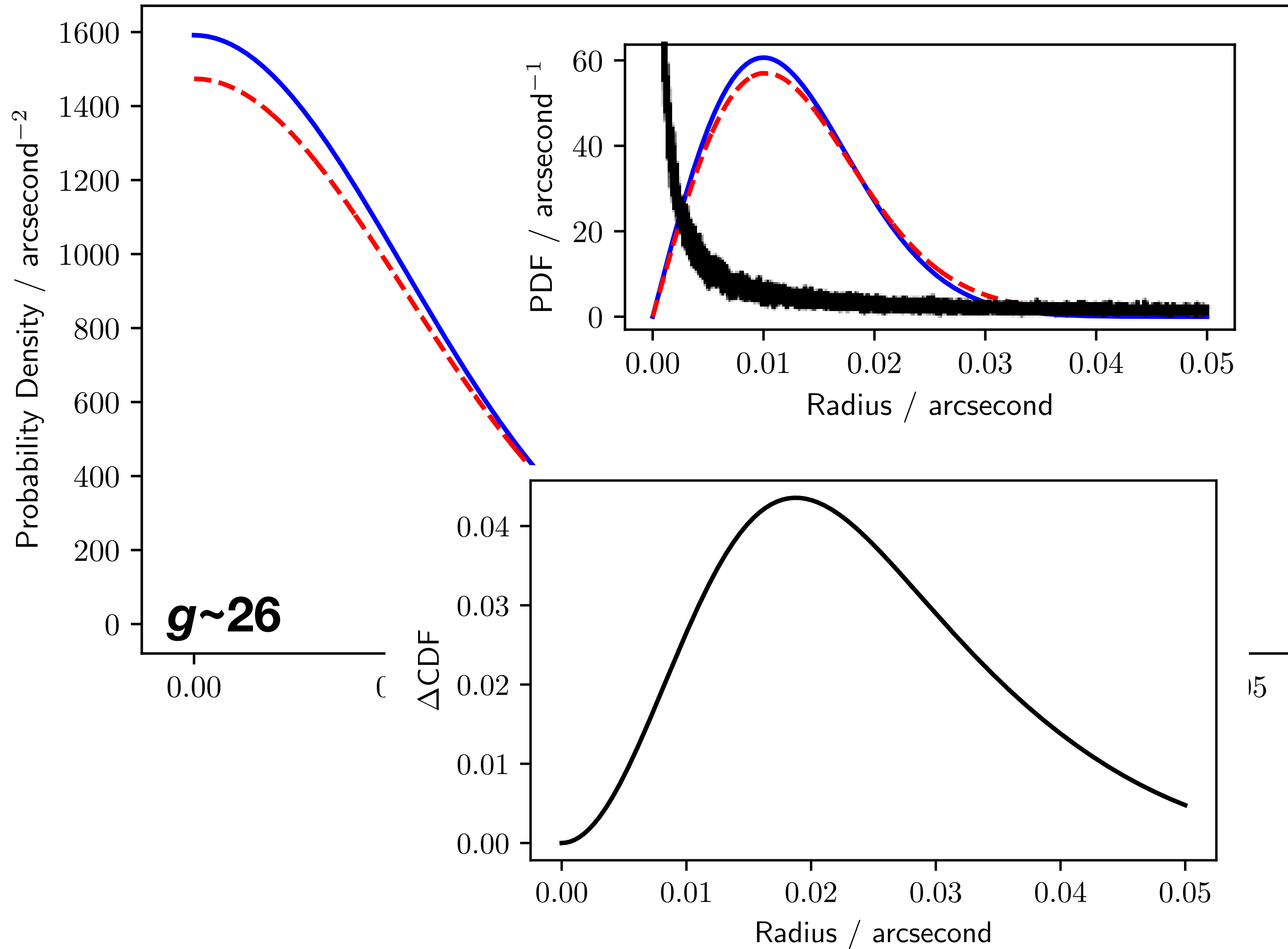
Co-add



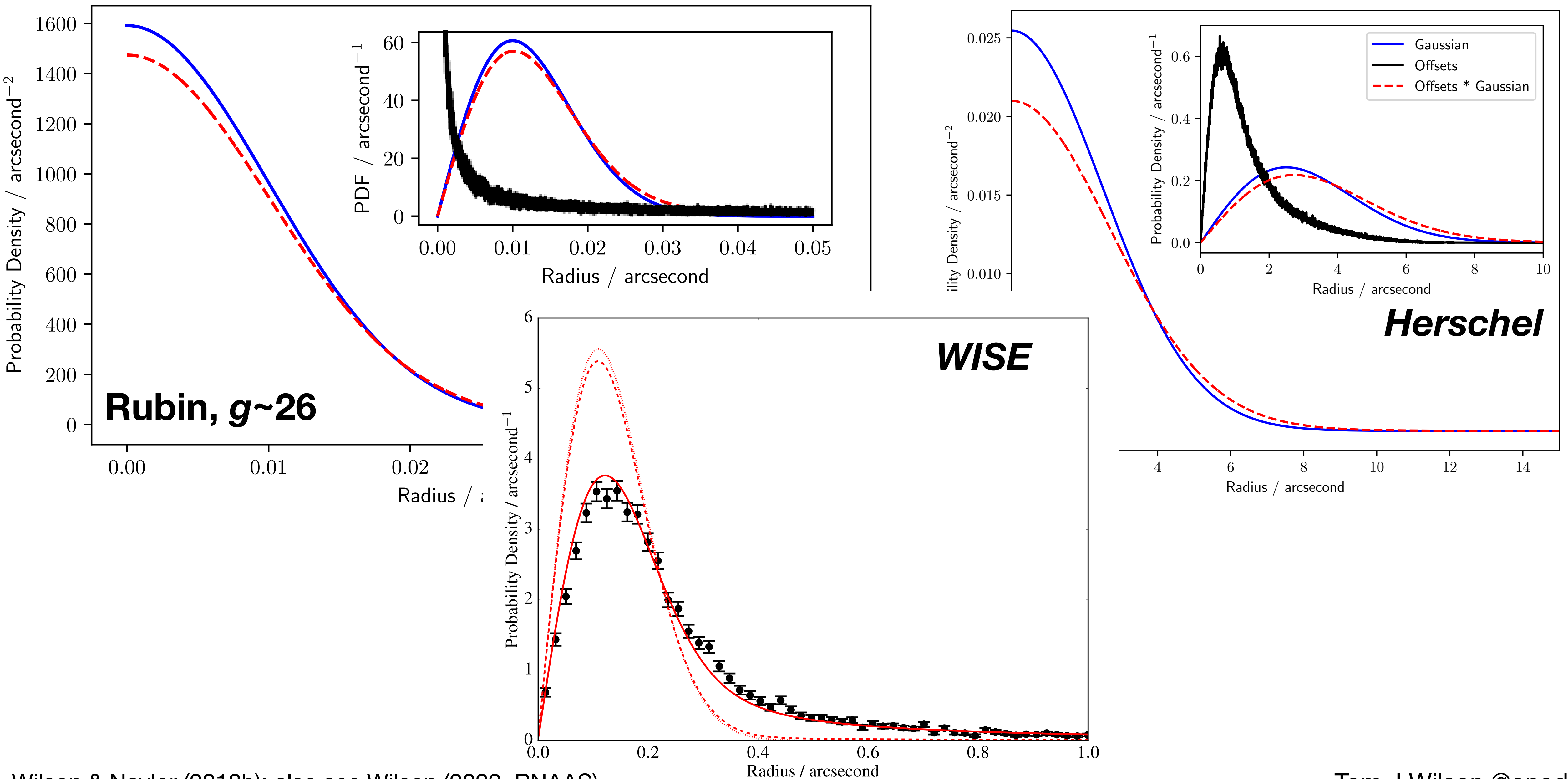
The Rubin AUF: Extra-Galactic, Transients



The Rubin AUF: Extra-Galactic, Transients

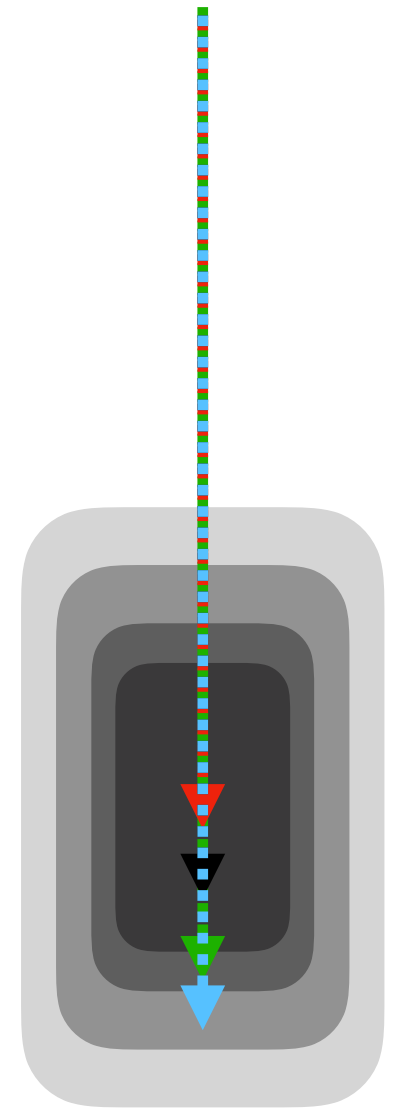


The Rubin AUF: Extra-Galactic, Transients

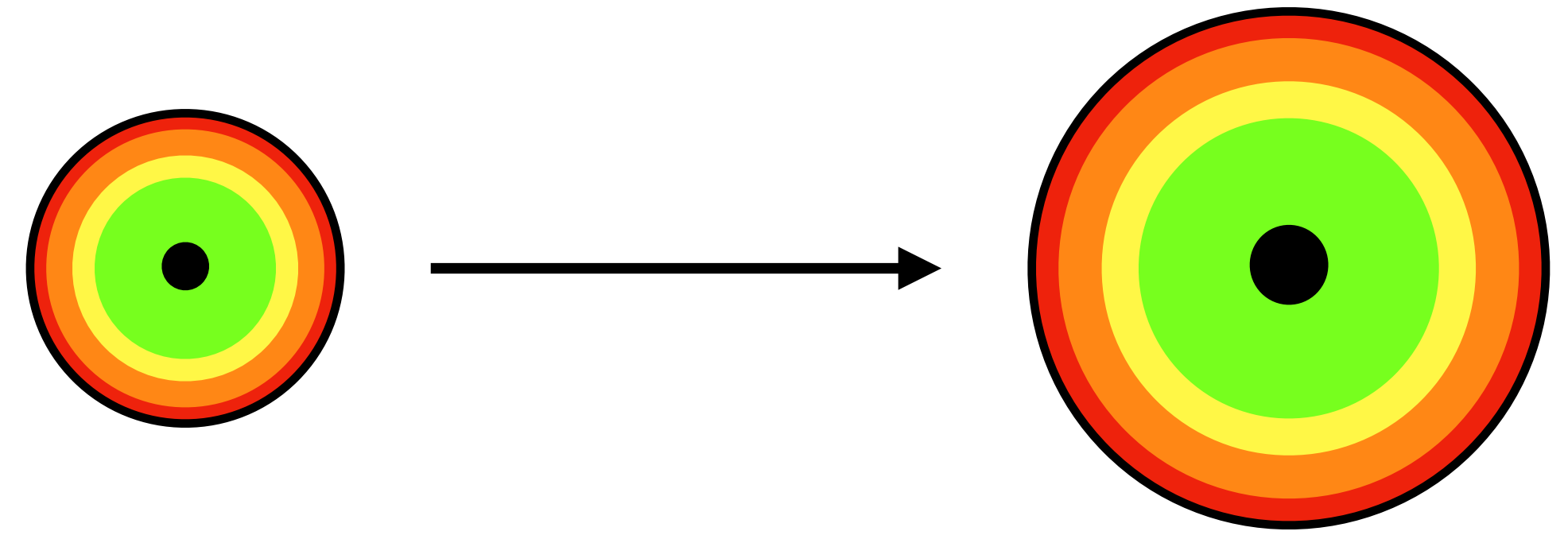
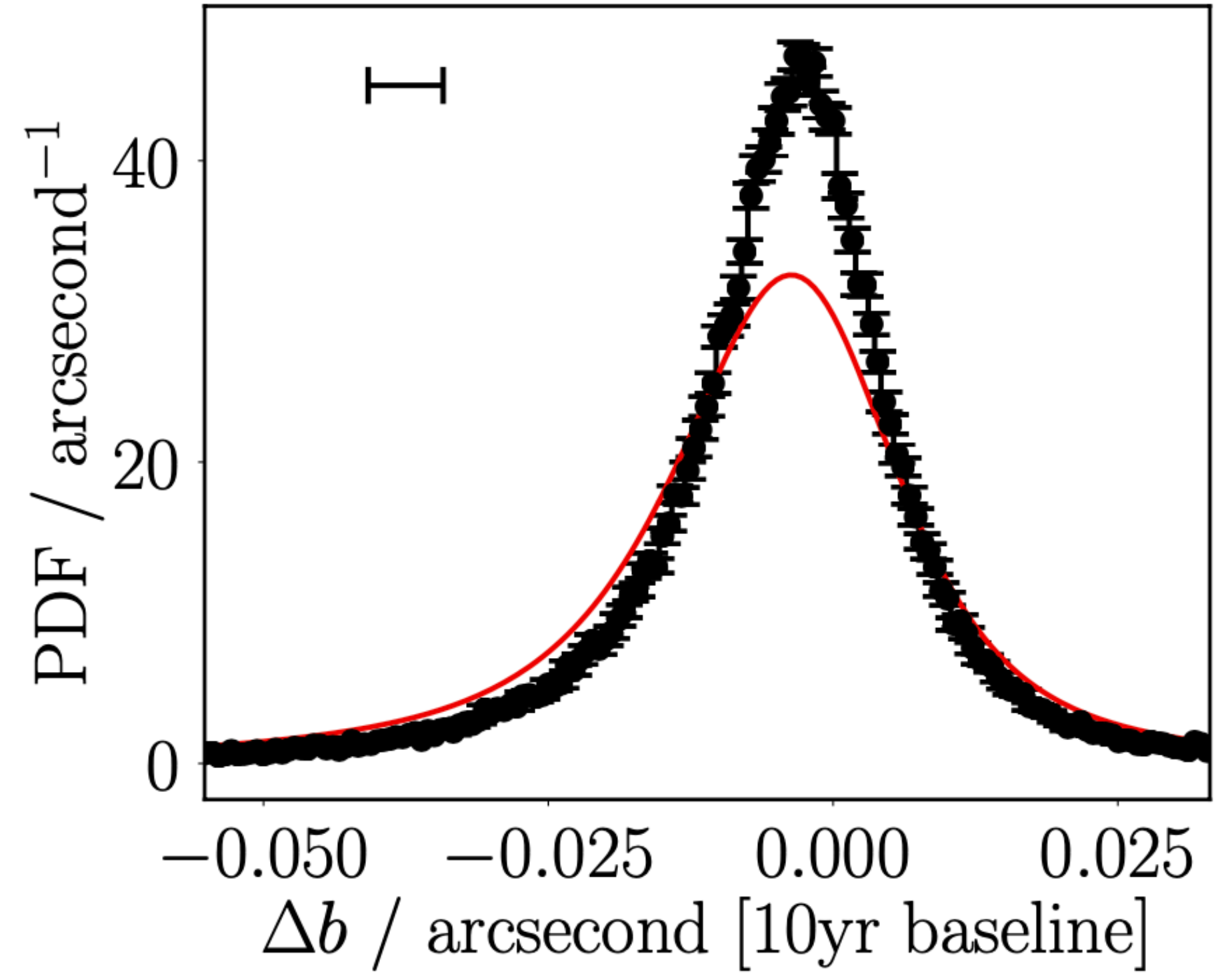
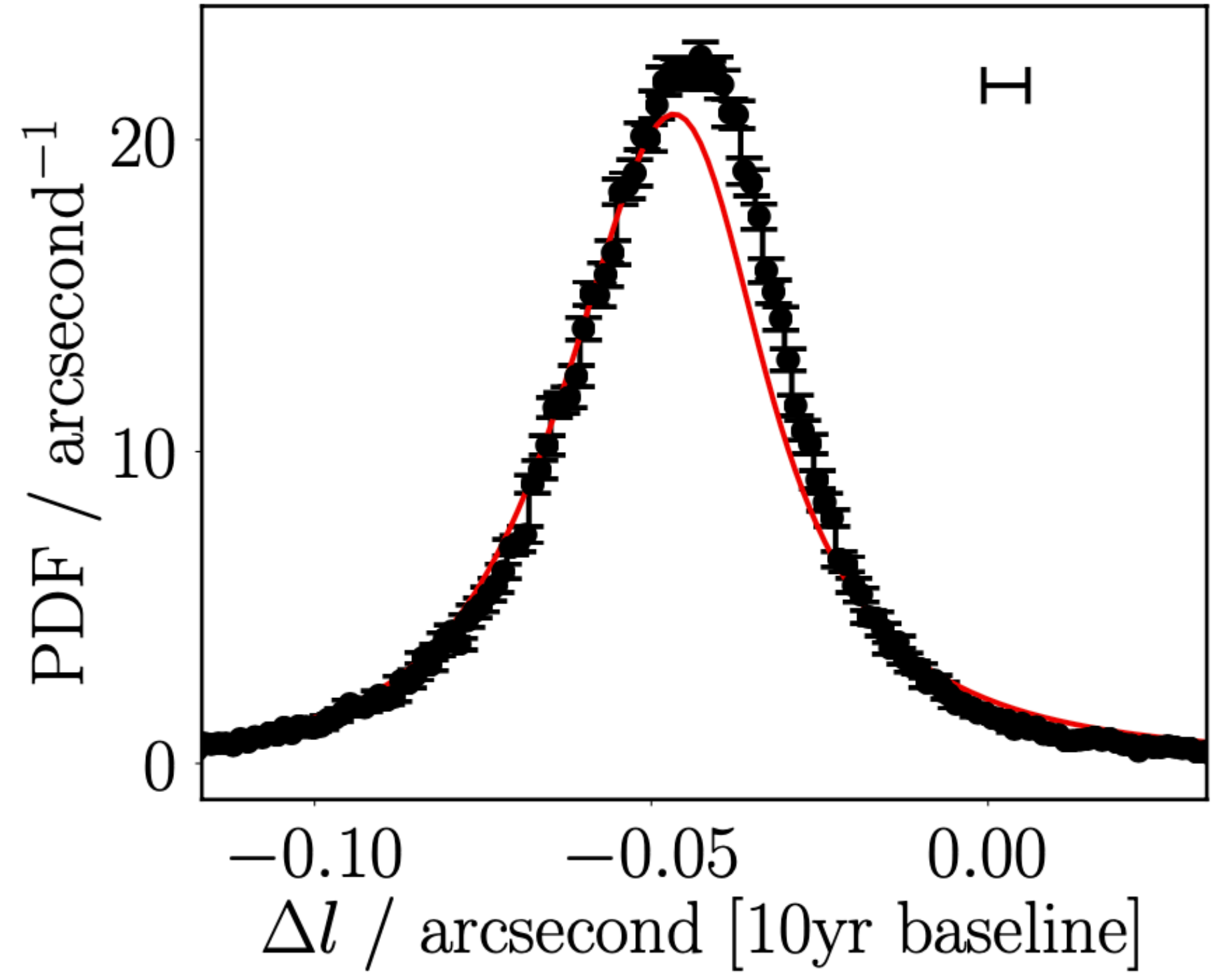


Unknown Proper Motions

Object in 2015

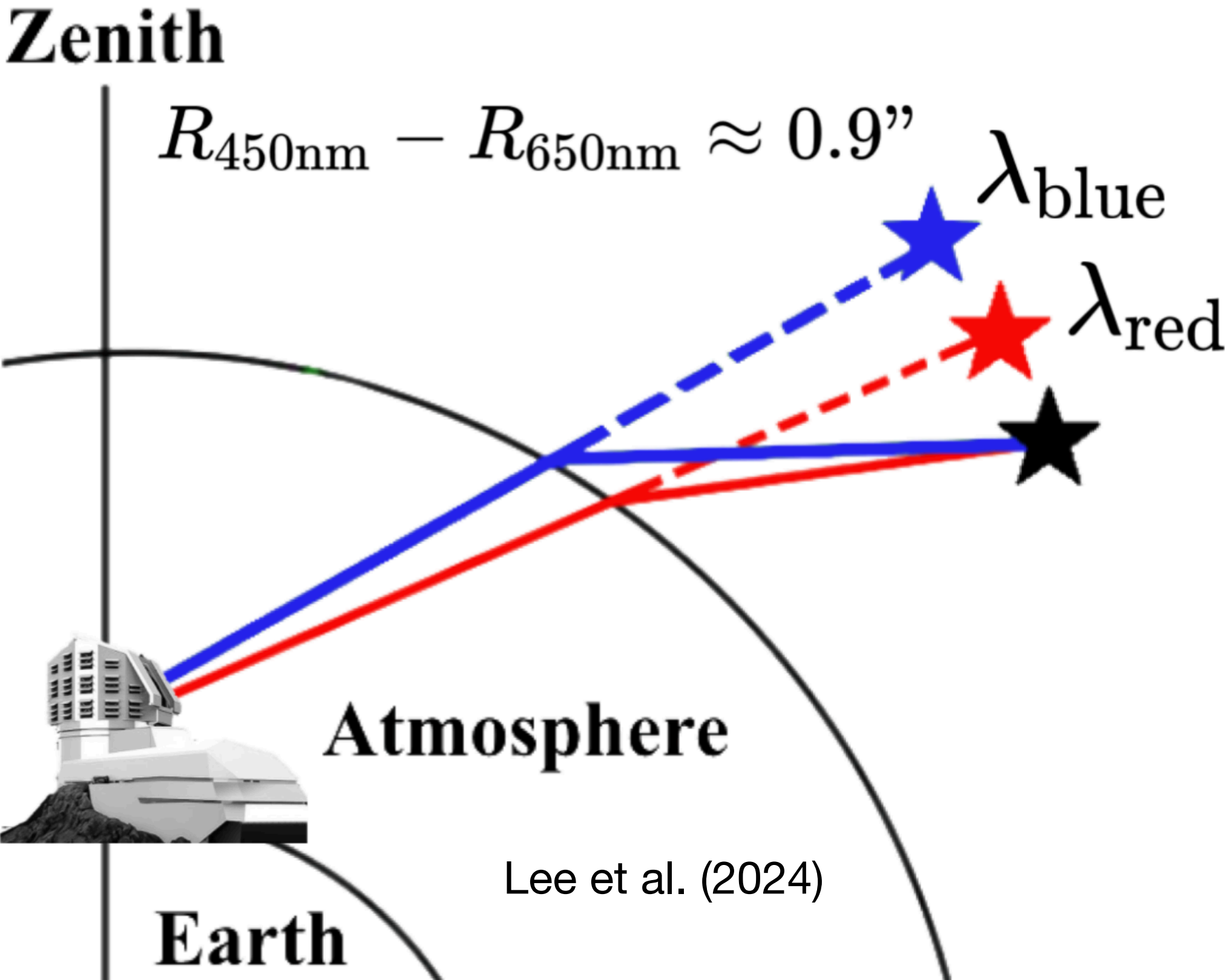


Projected to 2025



(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)

Differential Chromatic Refraction

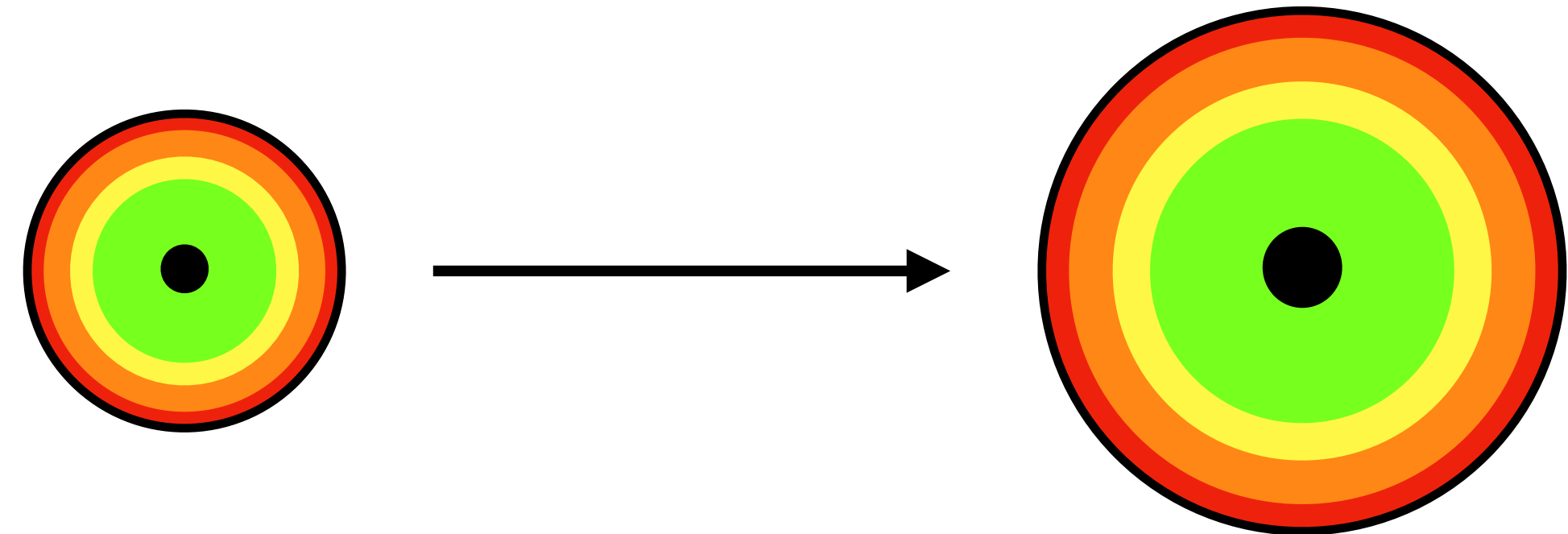


e.g. gbdes, Bernstein et al. (2017)

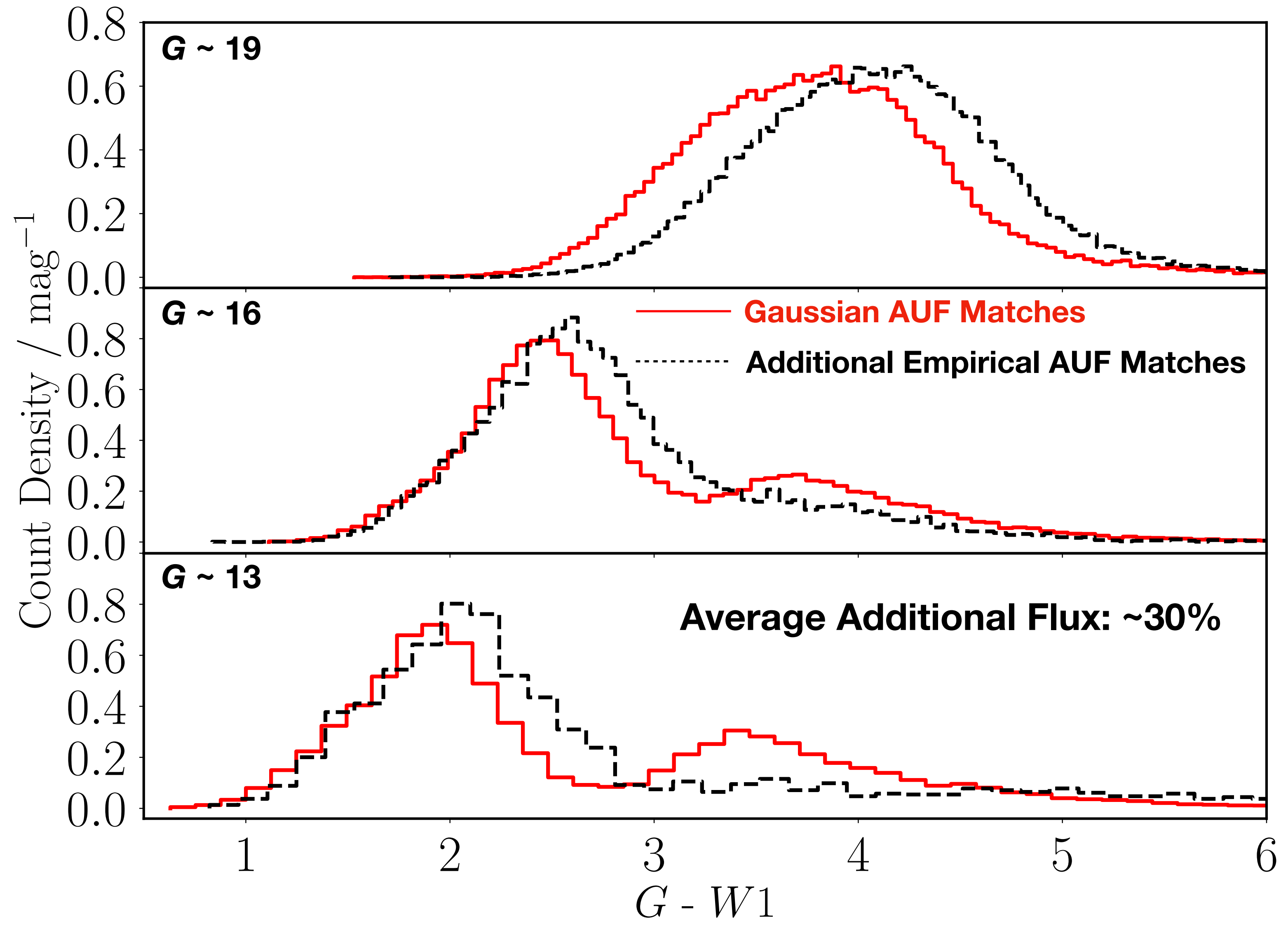
$$\Delta \mathbf{x}^w = K_b c \tan z \hat{\mathbf{p}}$$

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

Unknown/uncertain
photometric colour *c*

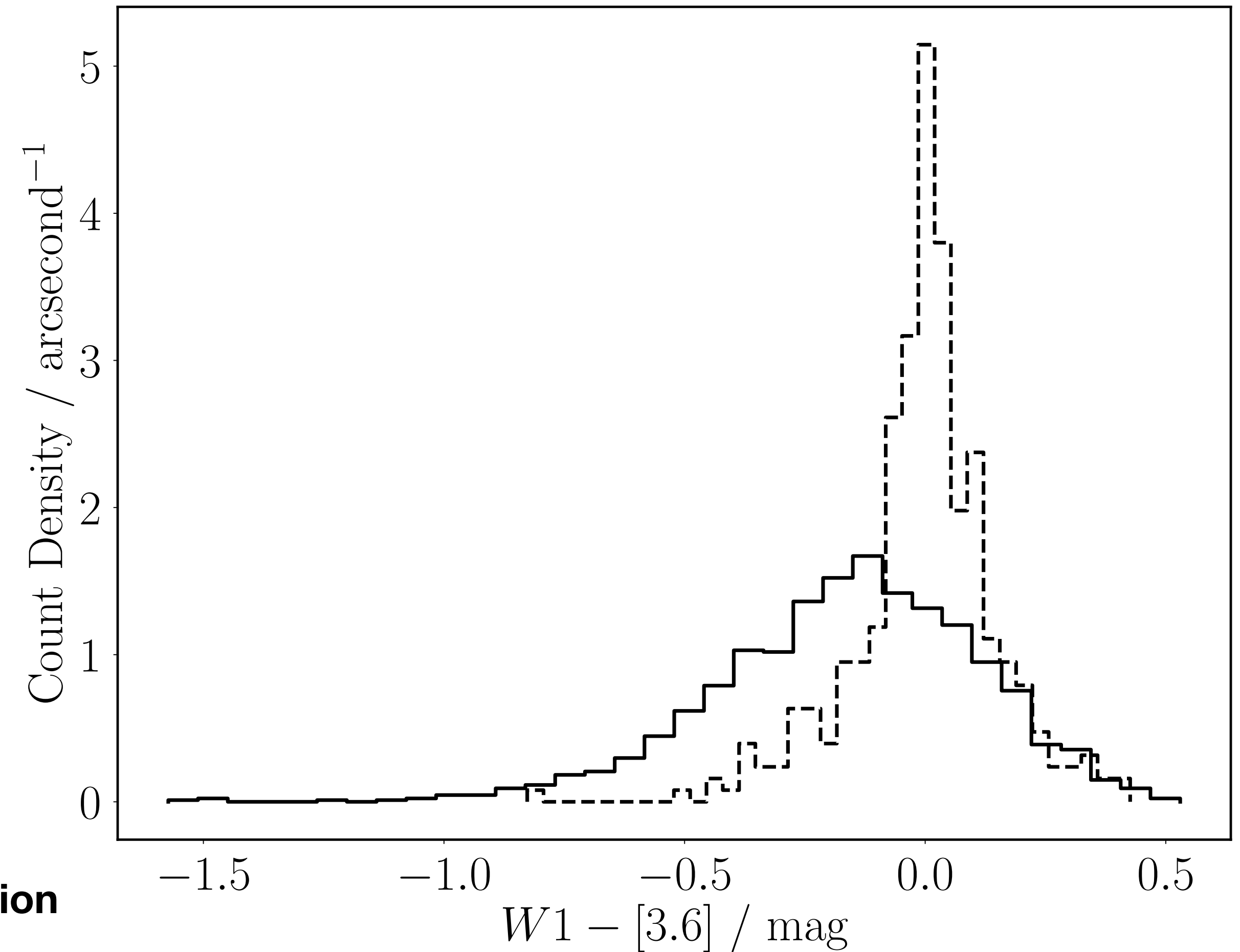
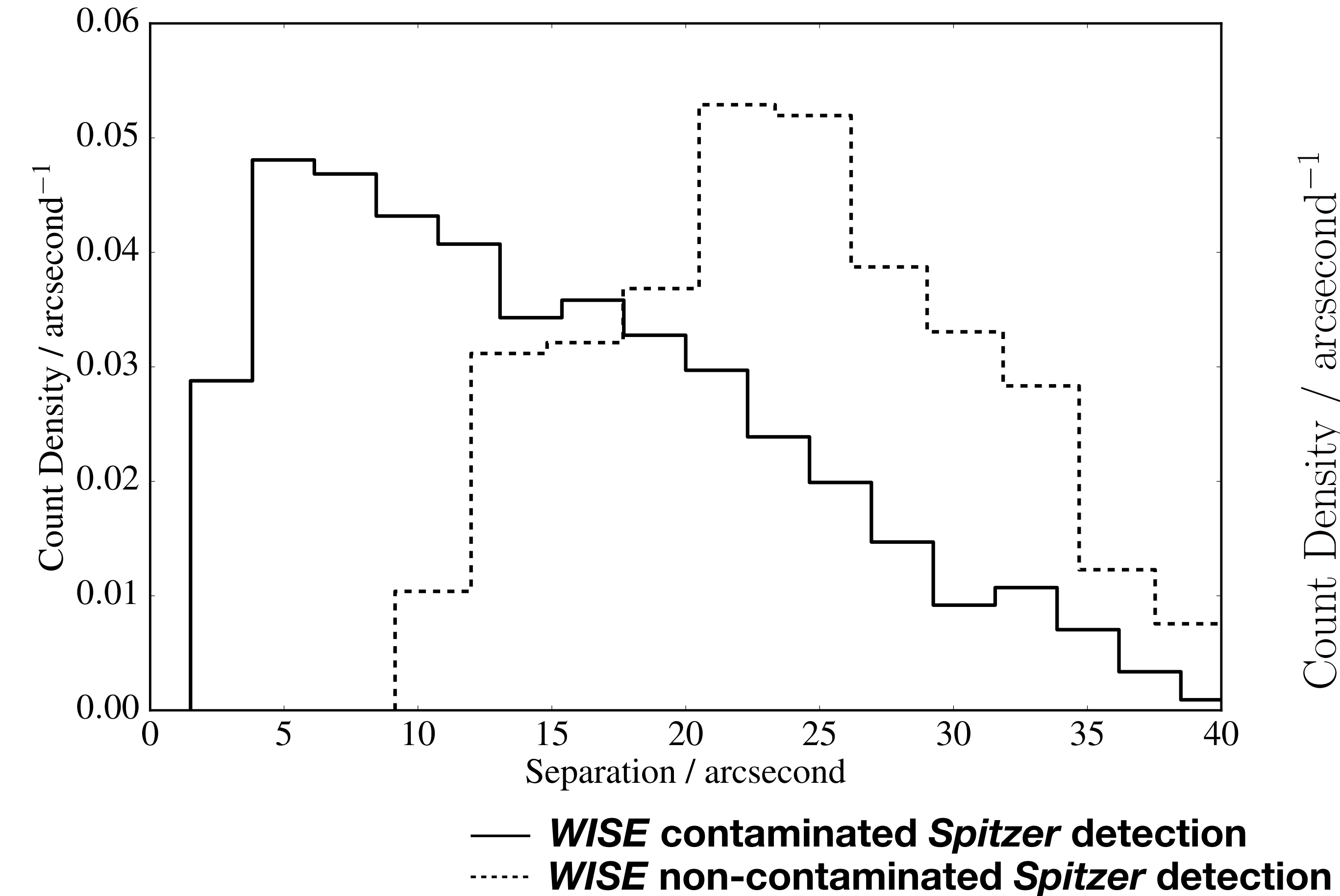


Photometric Effects of Crowding



“Extra flux” has an impact on derived proper motions and parallaxes, and IR excesses!

Resolving Contaminants



Spitzer - Werner et al. (2004)

IRAC - Fazio et al. (2004)

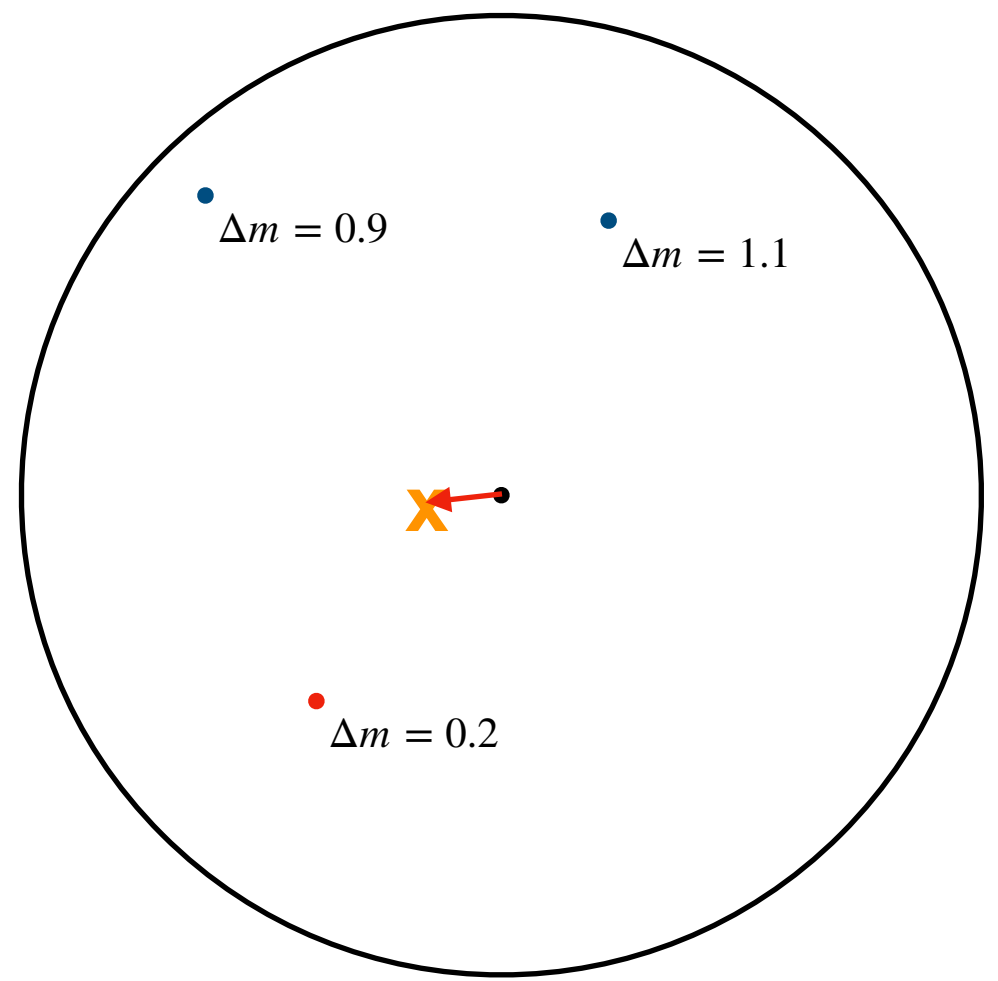
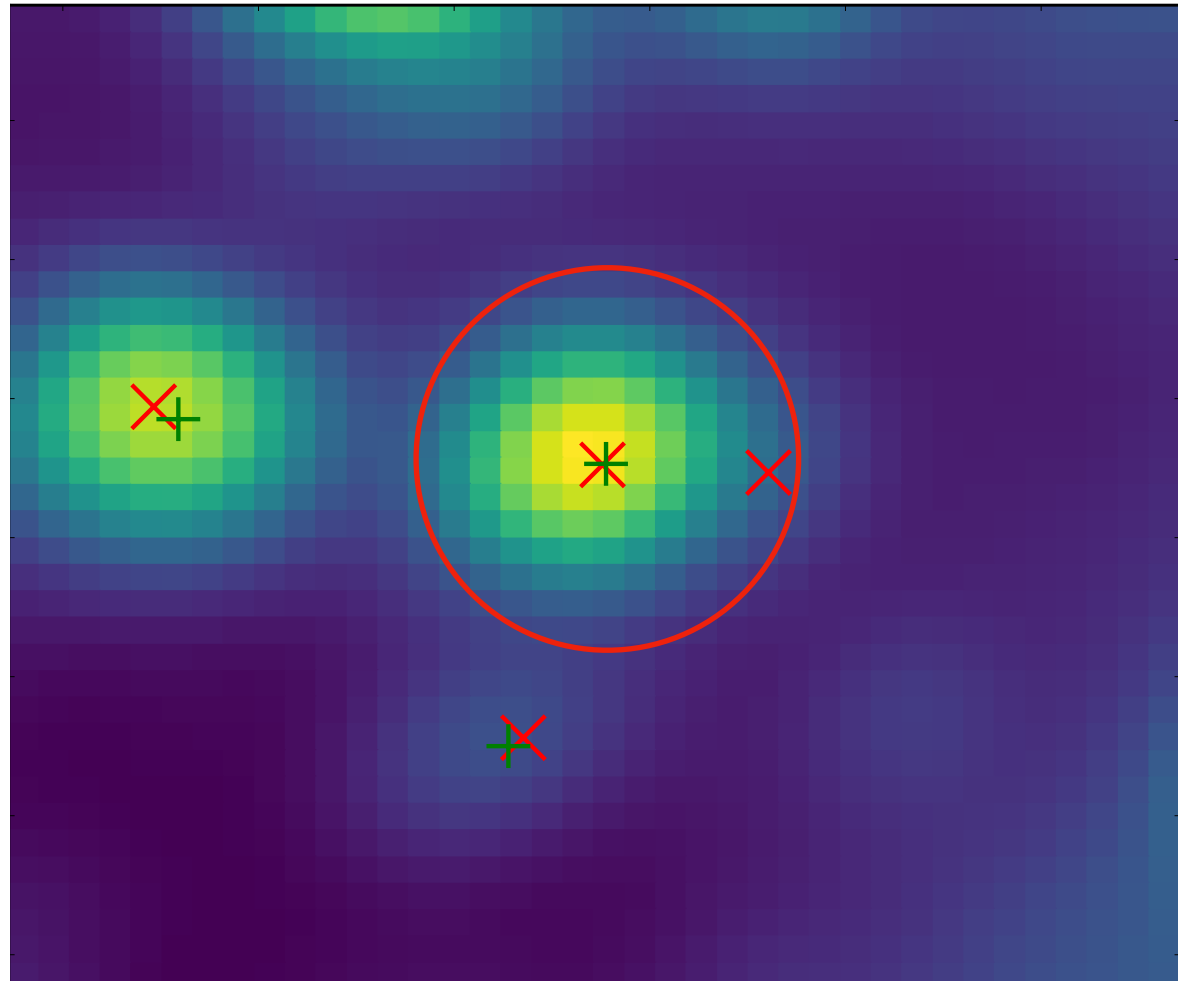
WISE - Wright et al. (2010)

Wilson & Naylor (2018b)

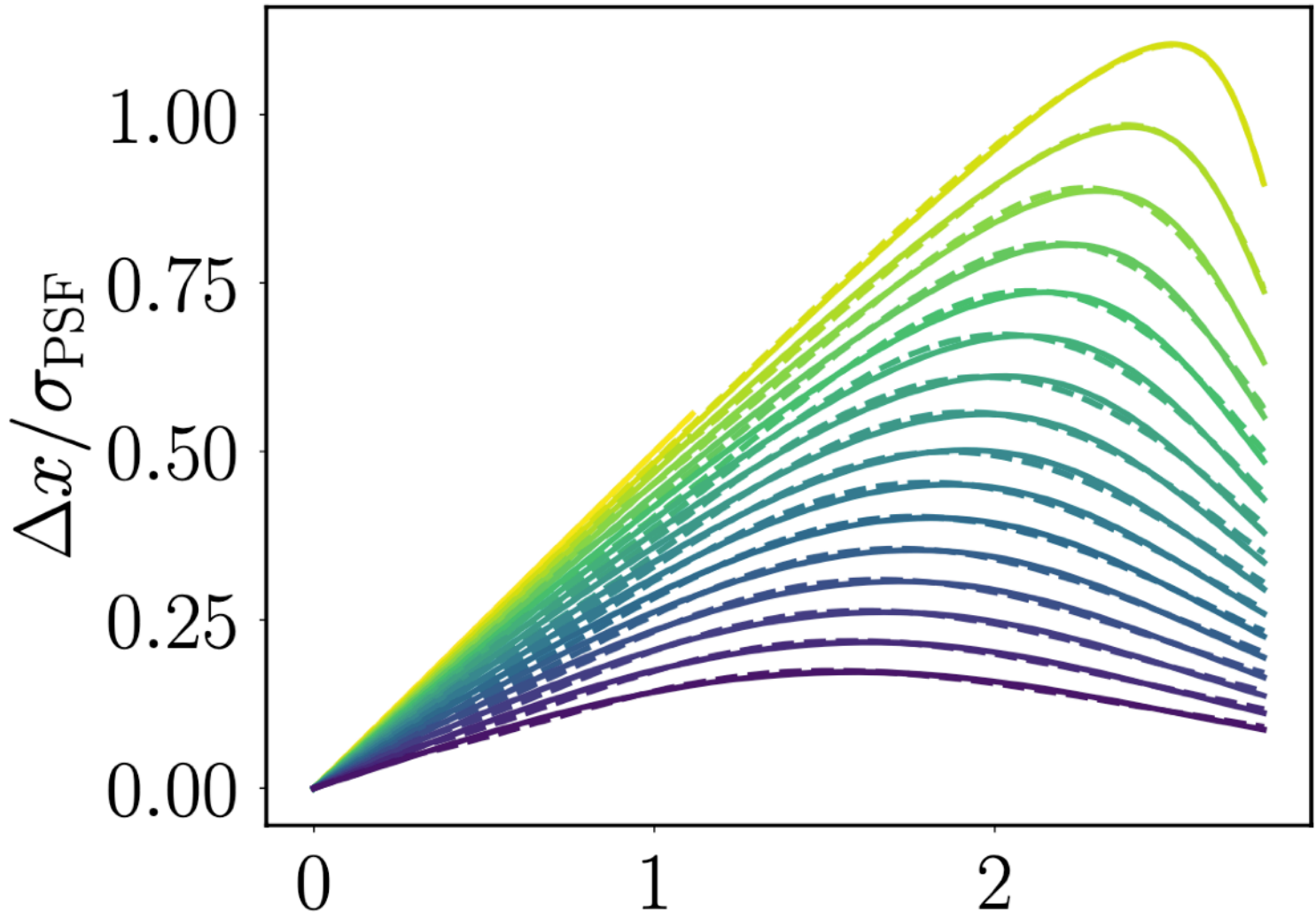
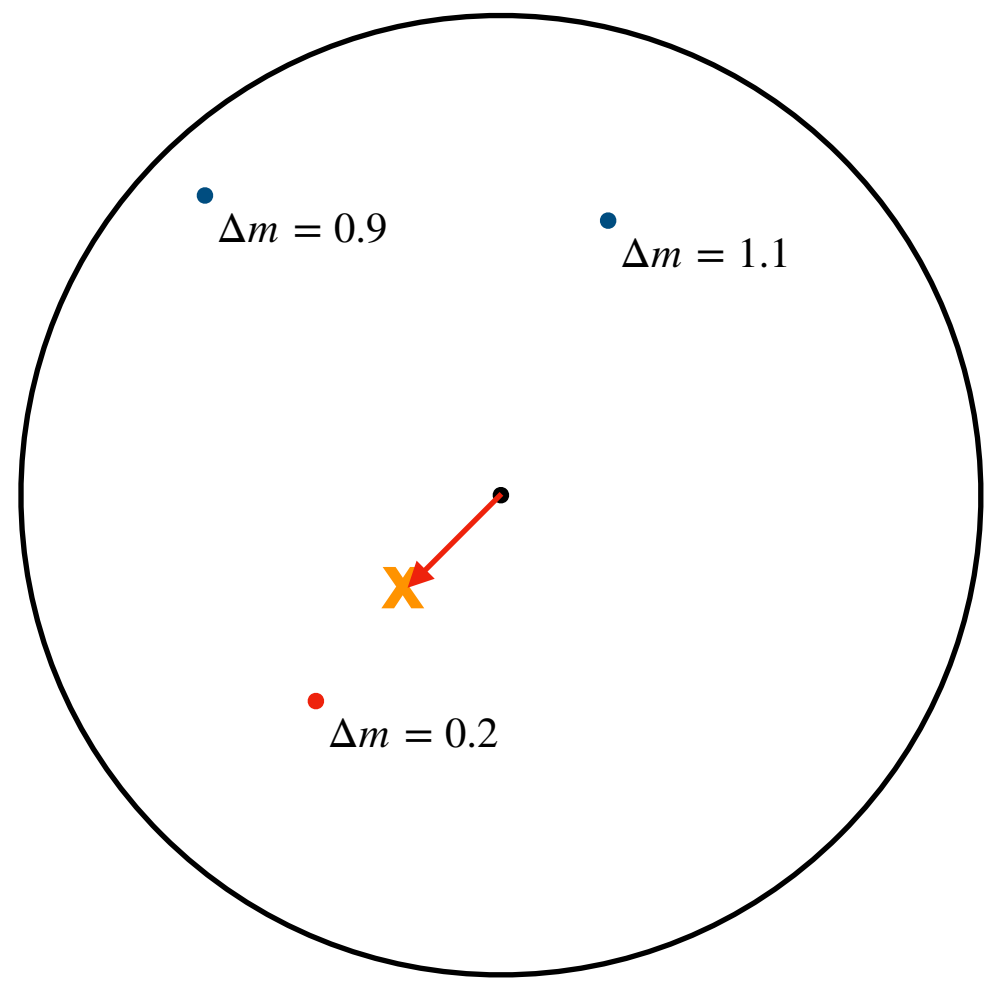
Tom J Wilson @onoddil

Modelling Crowded-Field Flux Brightening

High SNR PSF or Aperture Photometry



Low SNR PSF Photometry



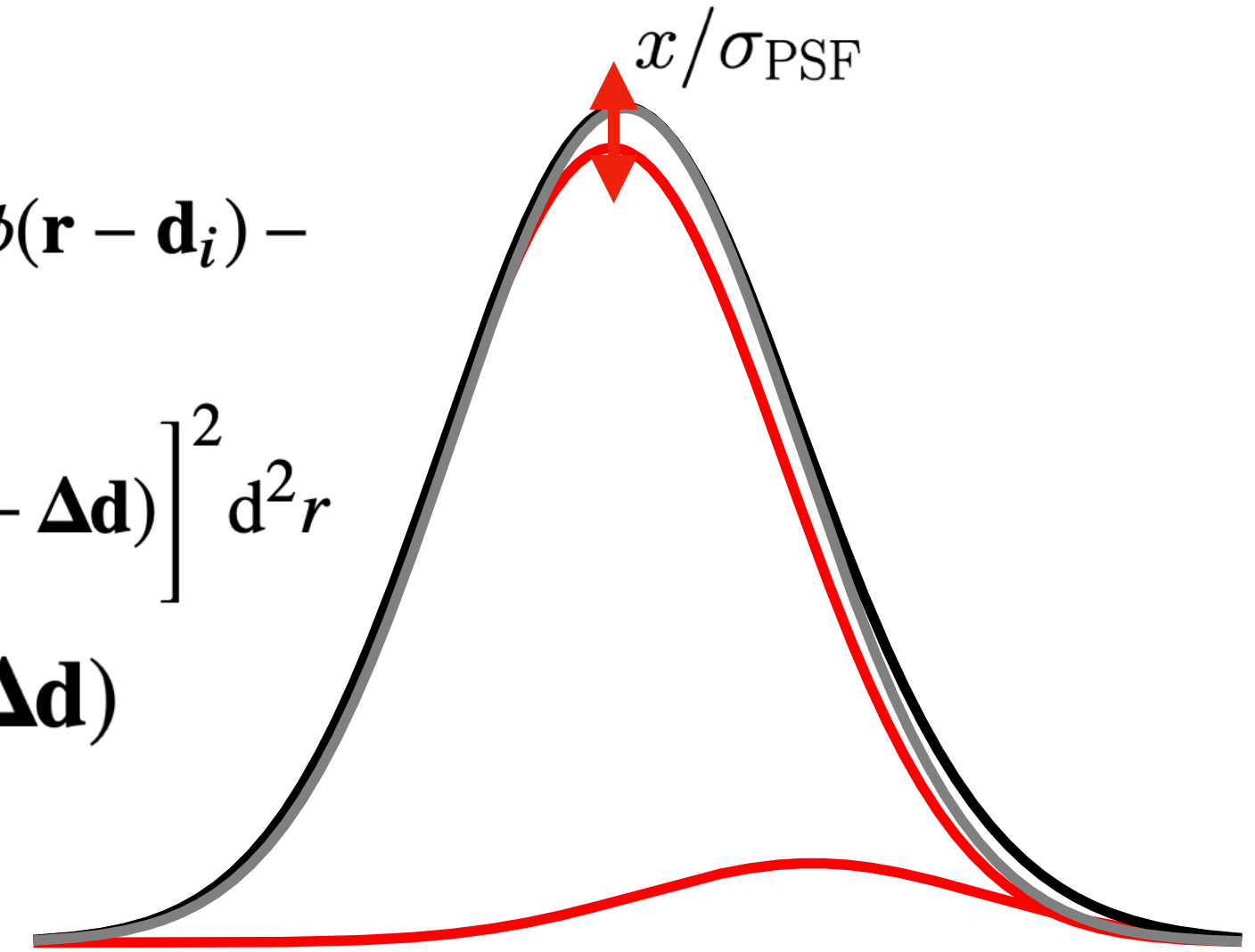
(This raises questions about the validity of quoting photometric statistical precisions if objects are systematically biased, and SED fitting in general in crowded fields)

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

$$\Delta f = \sum_i f_i$$

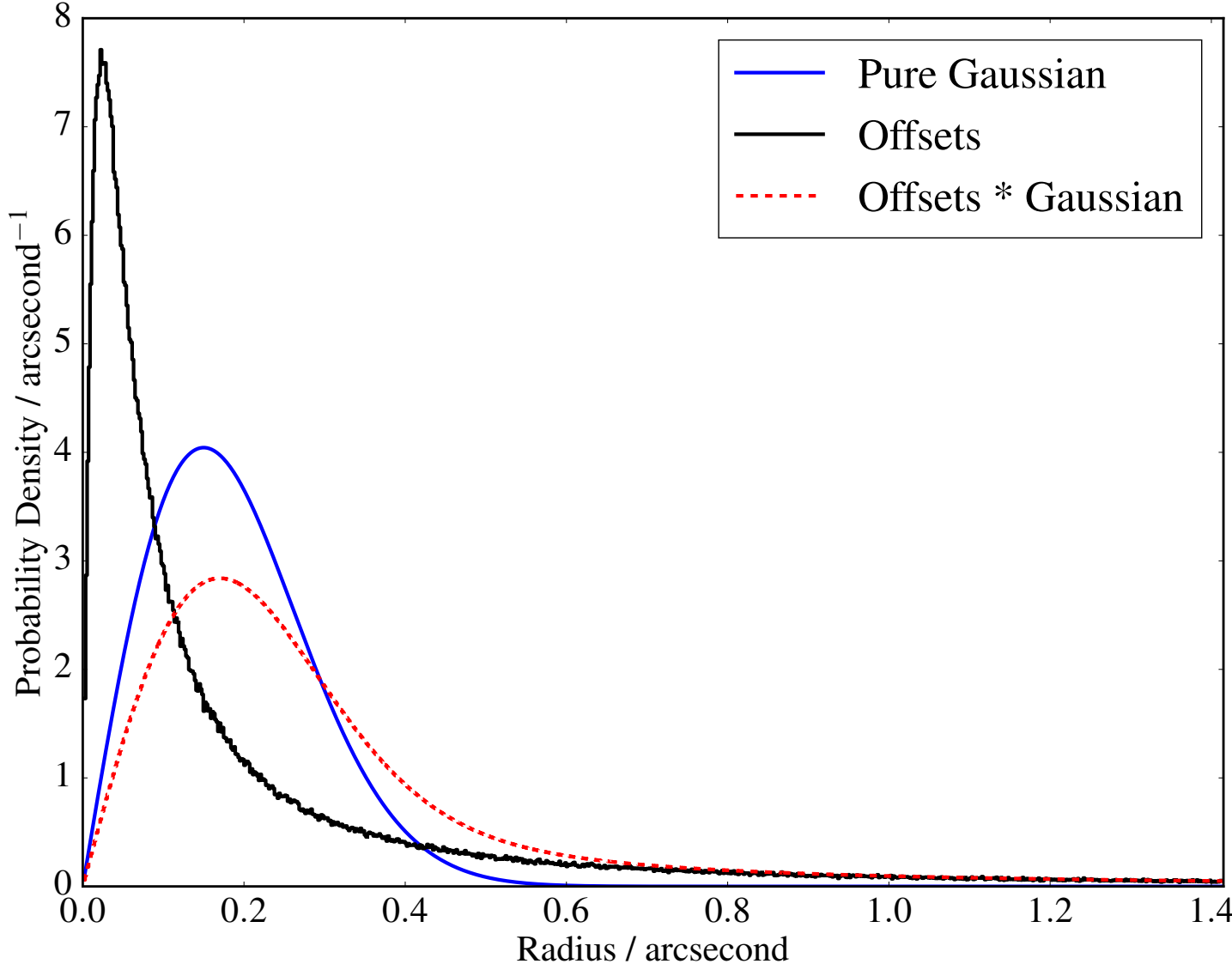
$$\log \mathcal{L} = -\frac{1}{2} \times L \int_{-\infty}^{\infty} \left[\phi(\mathbf{r}) + \sum_i f_i \phi(\mathbf{r} - \mathbf{d}_i) - (1 + \Delta f) \phi(\mathbf{r} - \Delta \mathbf{d}) \right]^2 d^2 r$$

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



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

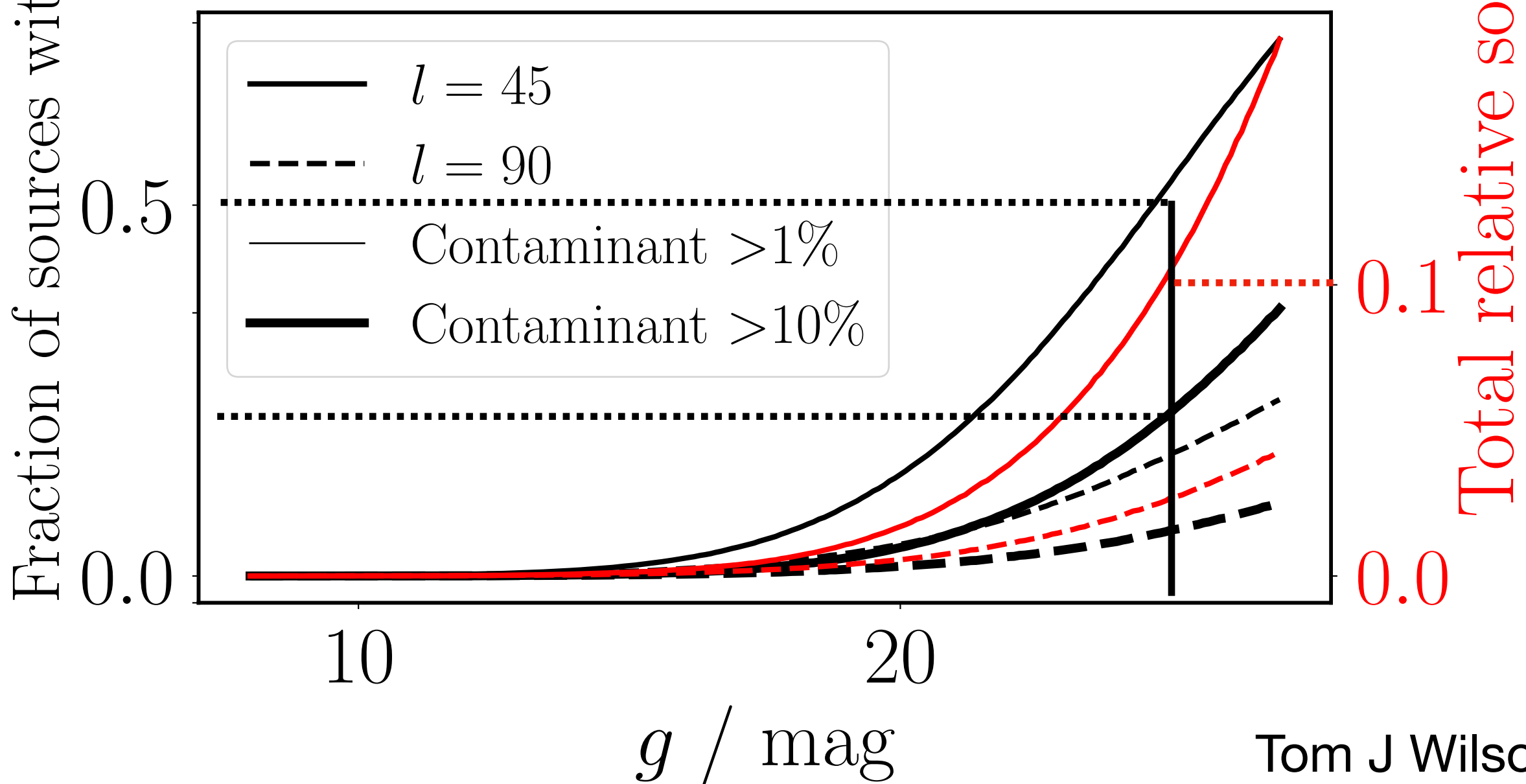
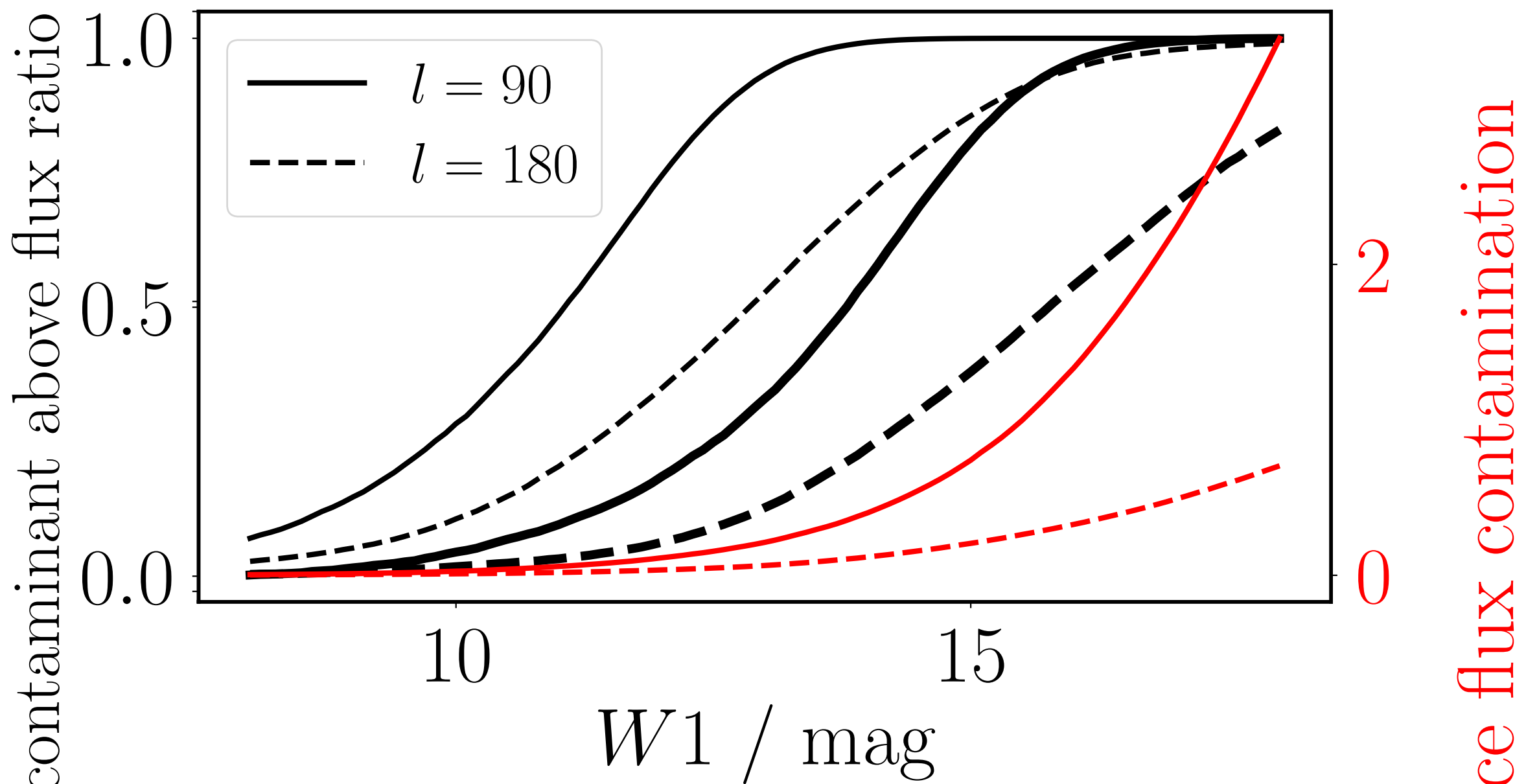
Photometry: Contamination Rates and Amounts



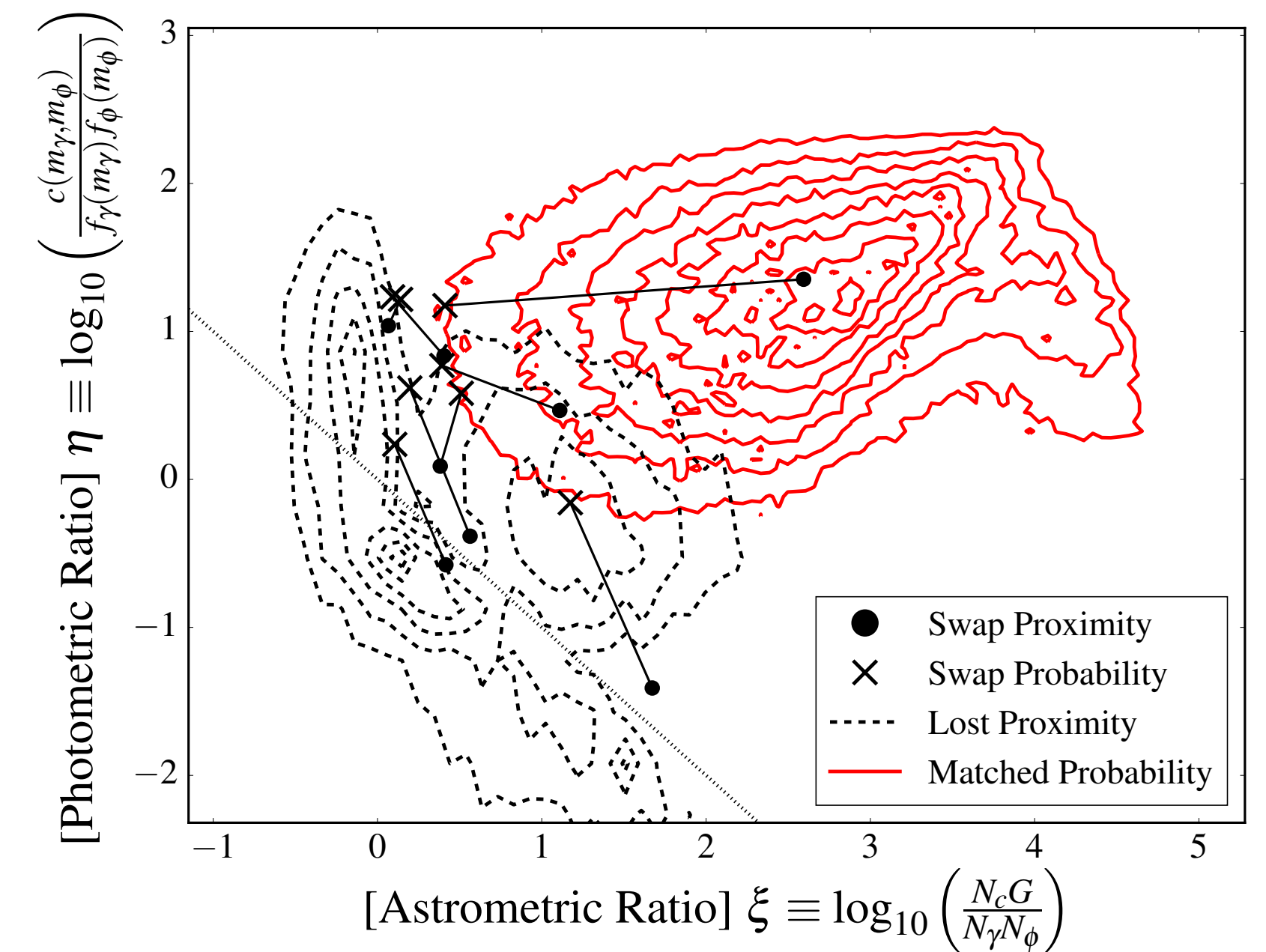
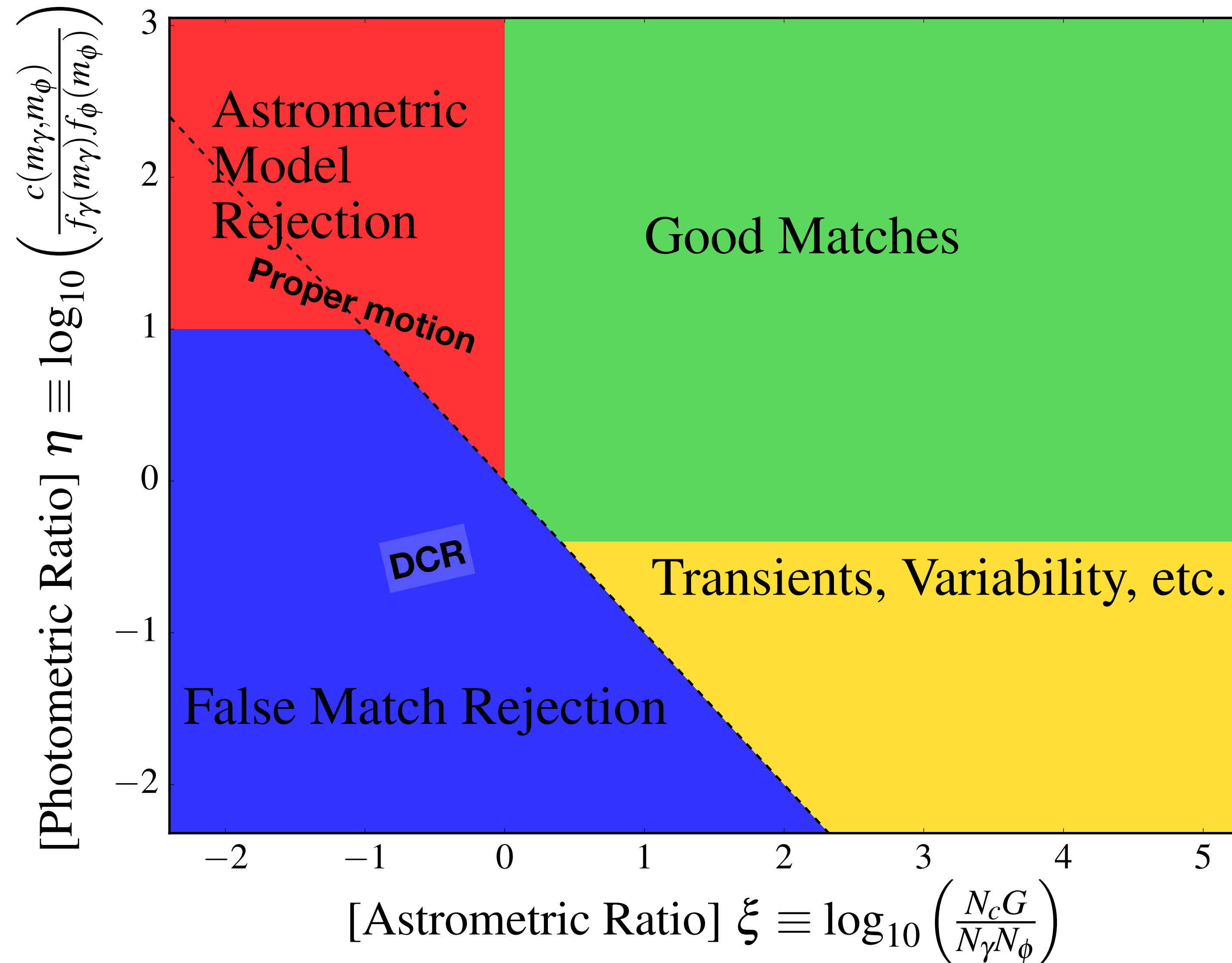
Typical, single visit images in near-Bulge regions of the Plane will have:

- 50% of objects with at least one >1% flux object in their PSF
- 20% of objects with a >10% relative flux object contaminating them
- an average 10% total “extra” flux

(the Bulge will be much more crowded! Nearest-neighbour matching won't work there, but neither will probabilistic matching without taking this effect into account...)



The Likelihood Ratio Space



Open Source Code: macauff

Matching Across Catalogues using the Astrometric Uncertainty Function and Flux



<https://github.com/macauff/macauff>



(Points if you know your tartans!)

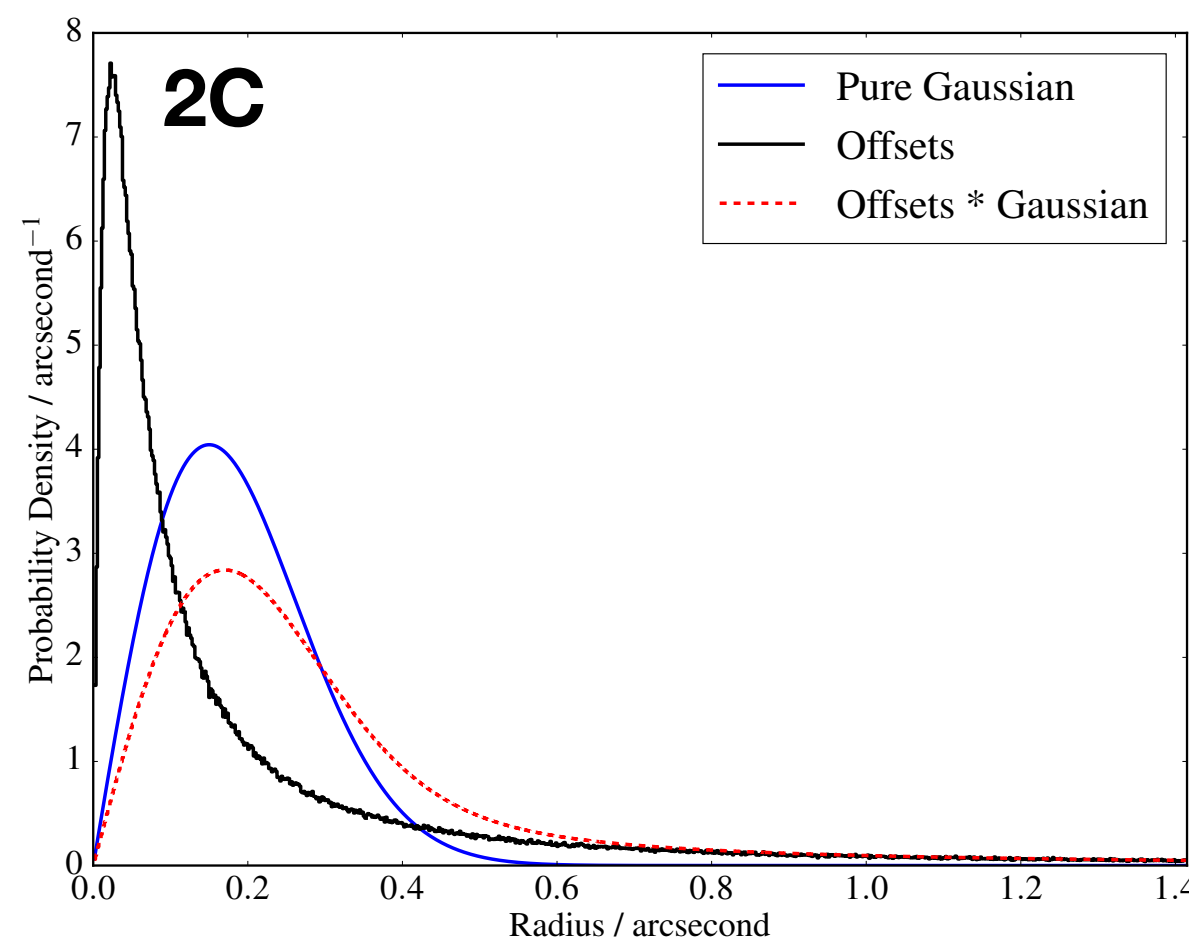
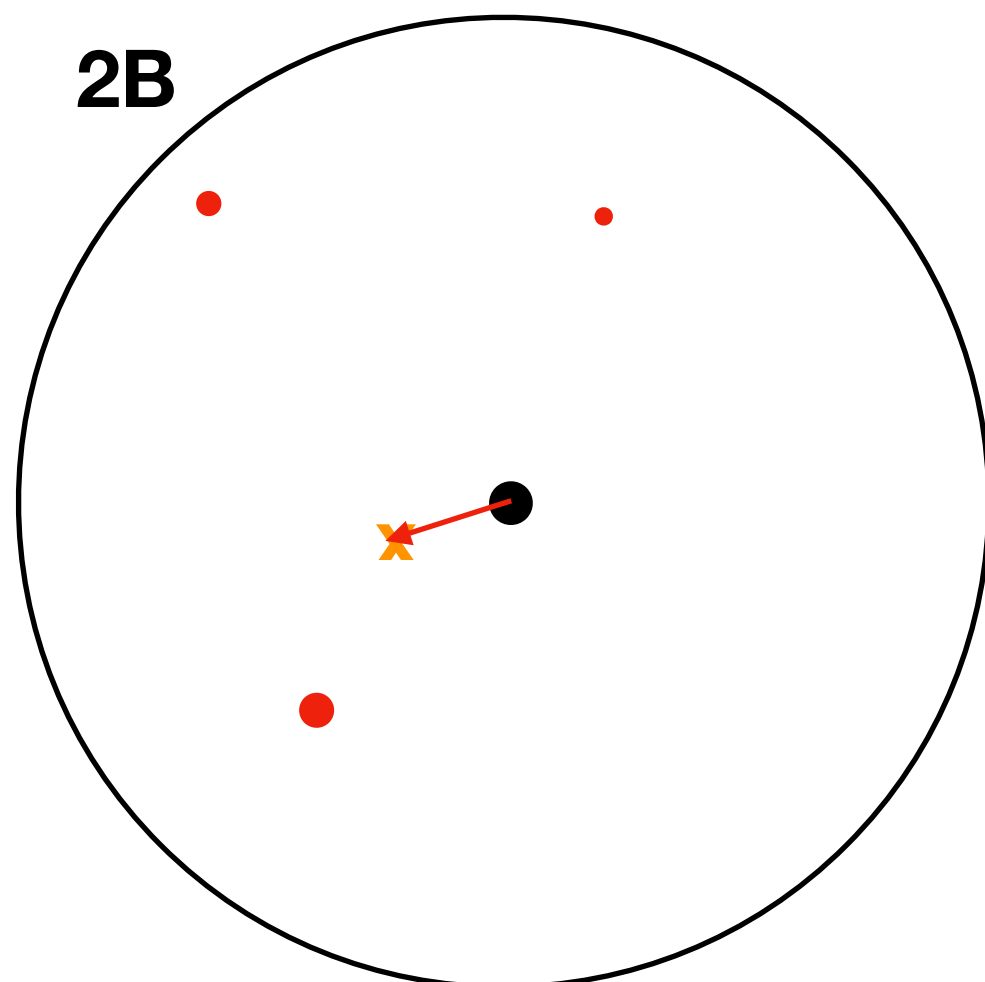
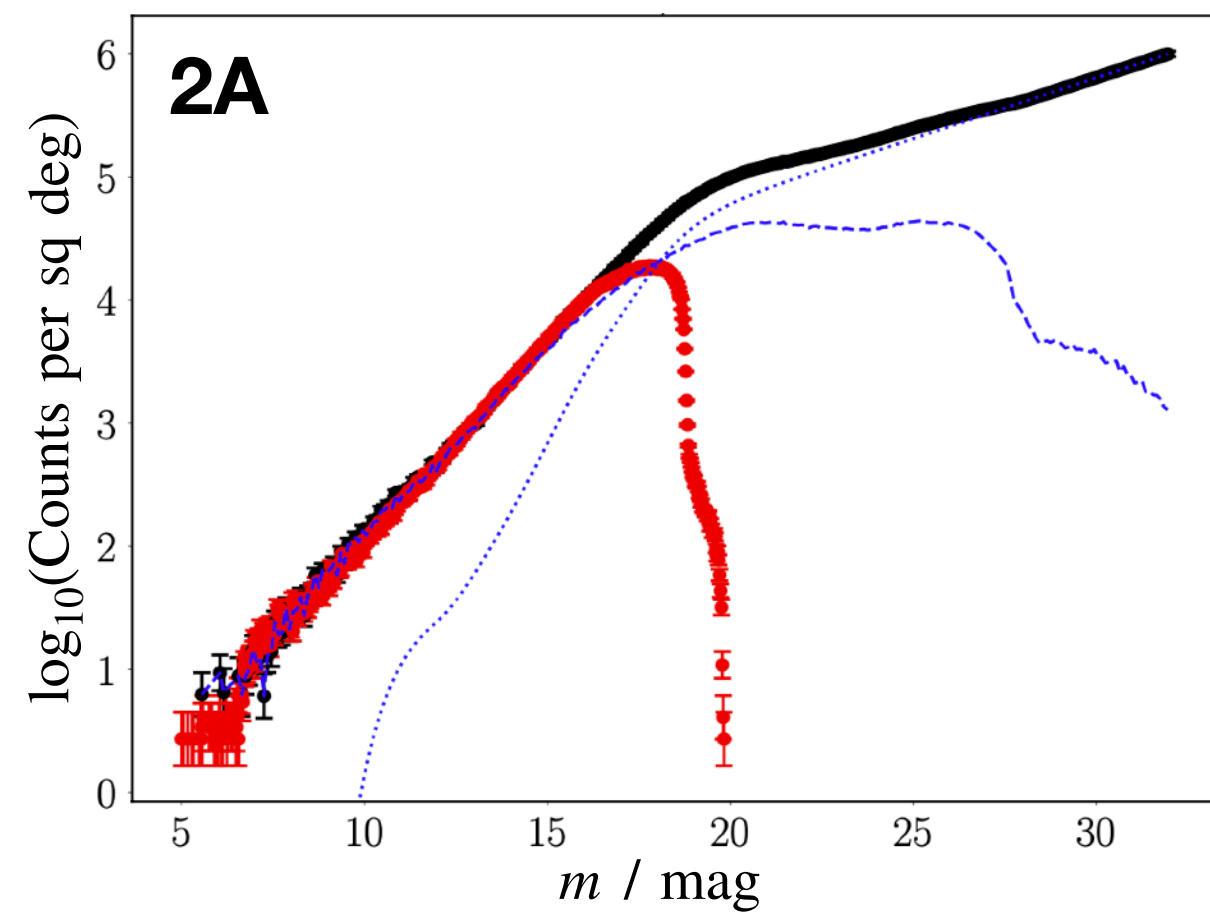
Tom J Wilson @onoddil

Verifying Astrometry: Accounting For Systematics

In each sightline (10s of sq deg for good bright source counting N):

1. Cross-match your high angular resolution, high astrometric precision data to LSST to obtain separation distributions
2. **Create systematics model for all non-centroid astrometric components of uncertainty**
3. Fit full AUF to data, allowing centroid Gaussian uncertainty to be fit
4. Repeat for each brightness (and effectively different astrometric uncertainty)
5. Derive fit-quoted astrometric uncertainty relations

- A. Create crowding-caused perturbation model, for example:
- B. Verify model source count densities match observed data
- C. Randomly draw perturbing sources within your PSF (“darts at a dartboard”)
- D. Repeat lots of times to get a distribution of perturbation offsets
- E. Repeat however many times you have different perturbation algorithms
- F. Combine your perturbation algorithms



$$\bar{x} = \frac{1 \times 0 + \sum_i f_i x_i}{1 + \sum_i f_i}$$

- or -

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

$$\Delta x(x, y, f) = \begin{cases} f x \exp\left(-\frac{1}{4} \frac{x^2 + y^2}{\sigma_\psi^2}\right) & f < 0.15 \\ \Omega(x, f) & f \geq 0.15, \end{cases}$$

where

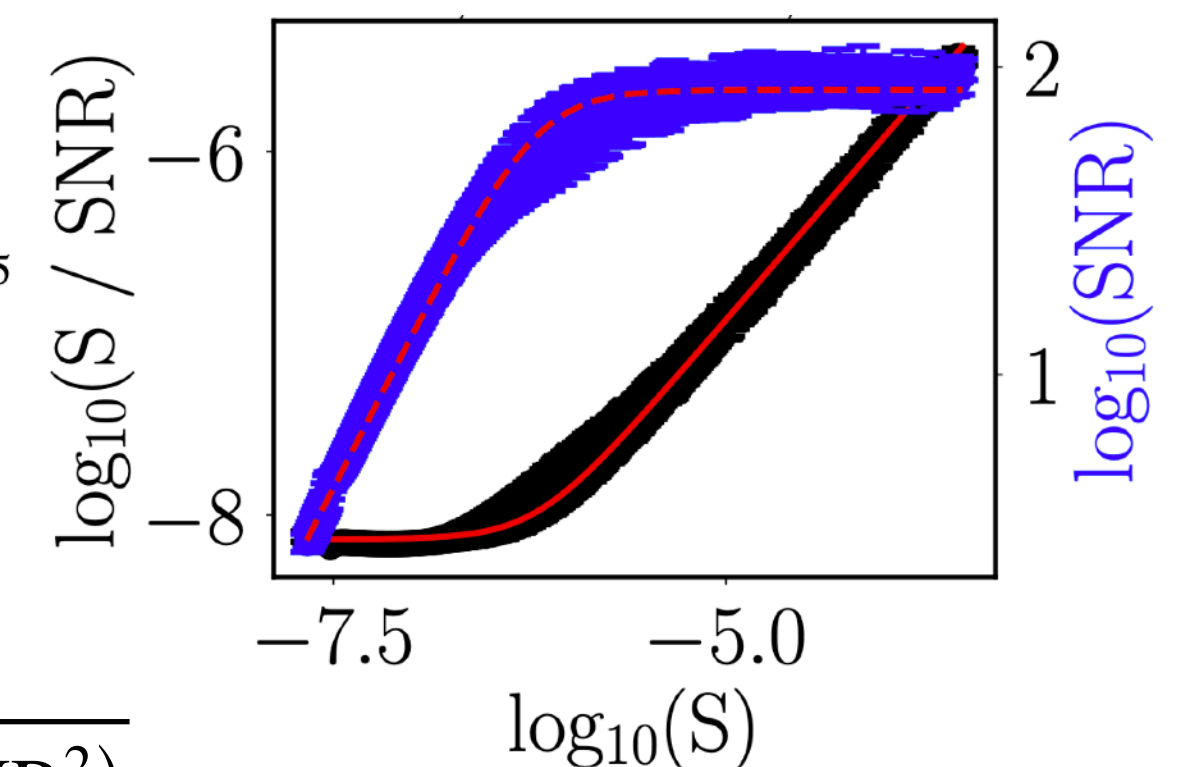
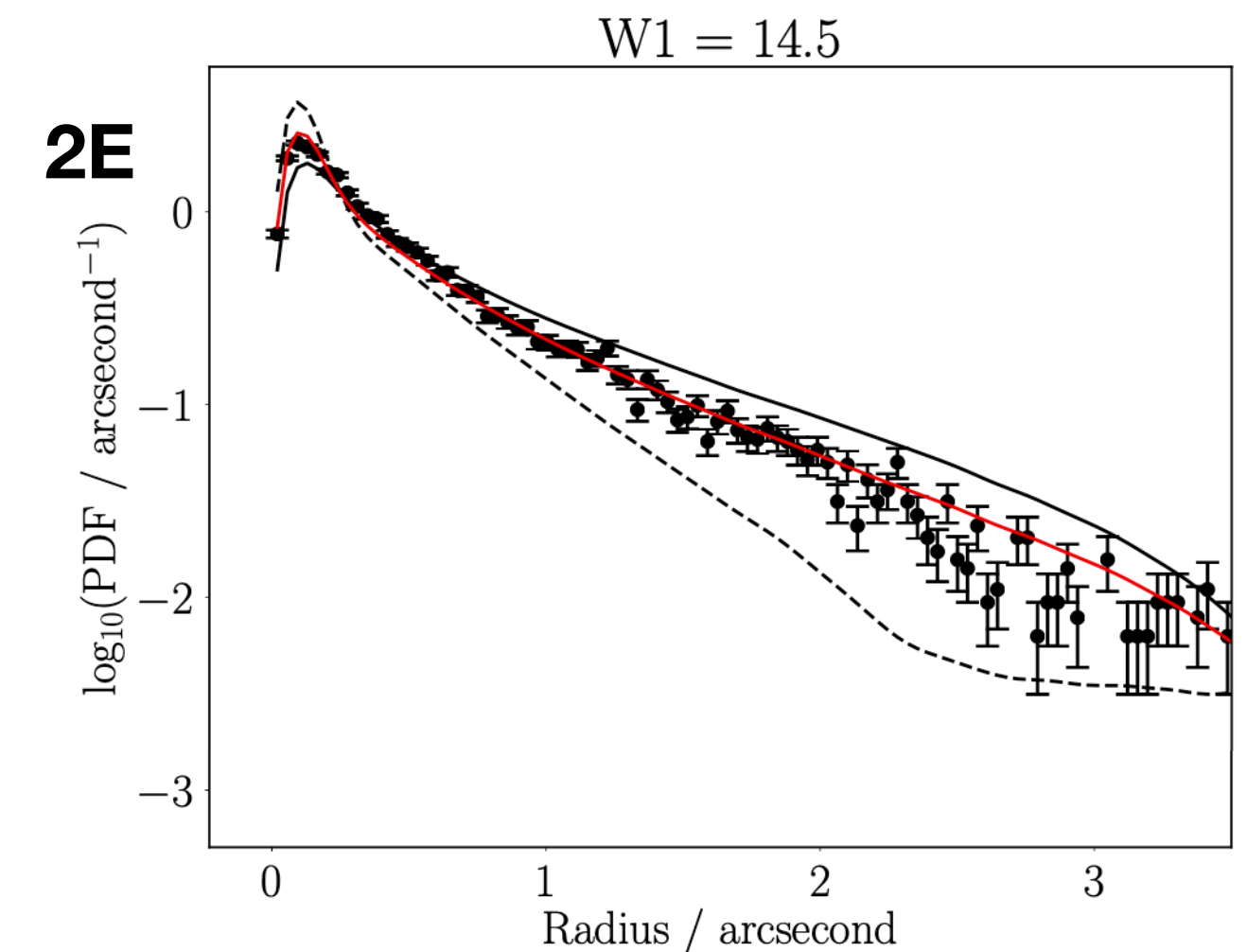
$$\Omega(x, f) = \Omega(x, f, \sigma, \mu, \alpha, T, r_c)$$

$$= \begin{cases} x f / (1 + f) & x < r_c \text{ or } f > 0.9975 \\ 2 f \frac{T}{\sigma} \lambda \left(\frac{x - \mu}{\sigma}\right) \Lambda\left(\alpha \frac{x - \mu}{\sigma}\right) & x > r_c \text{ and } f \leq 0.9975 \end{cases}$$

Wilson & Naylor (in prep.)
cf. Plewa & Sari (2018)

$$\text{SNR} = \frac{S}{\sqrt{c \times S + b + (a \times S)^2}}$$

$$H = 1 - \sqrt{1 - \min(1, a \times \text{SNR}^2)}$$



Verifying Astrometry: Accounting For Systematics

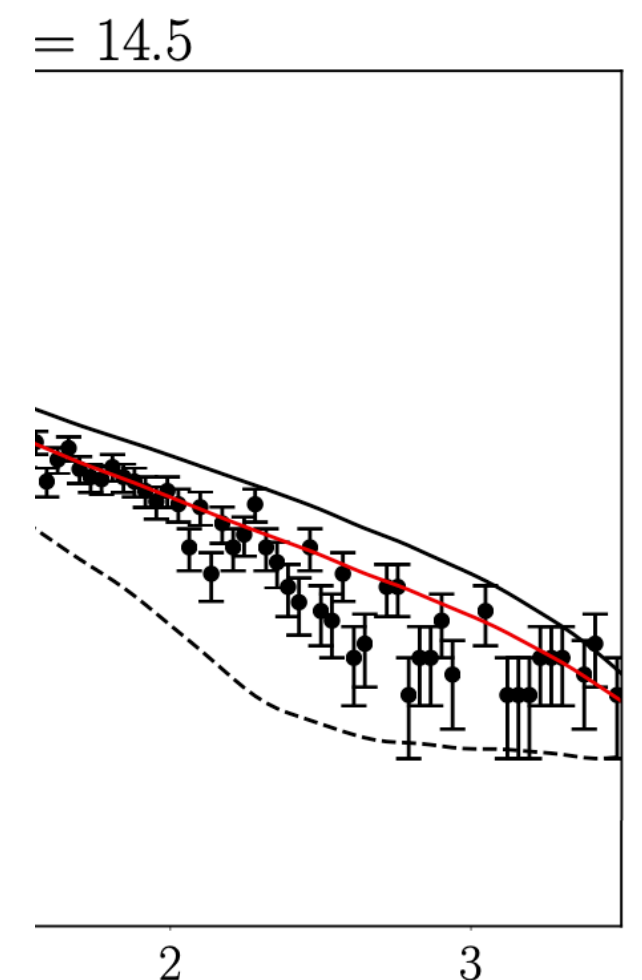
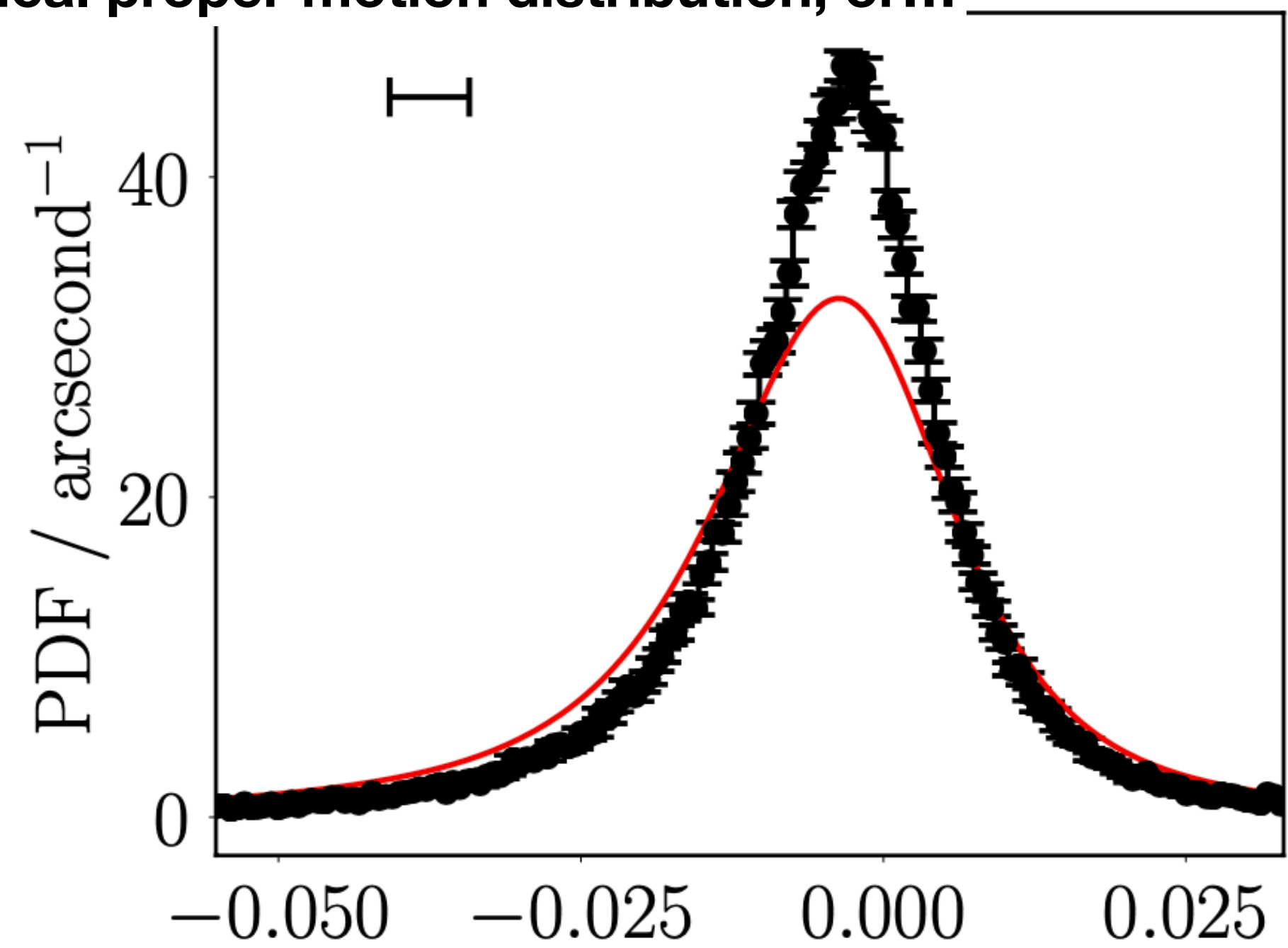
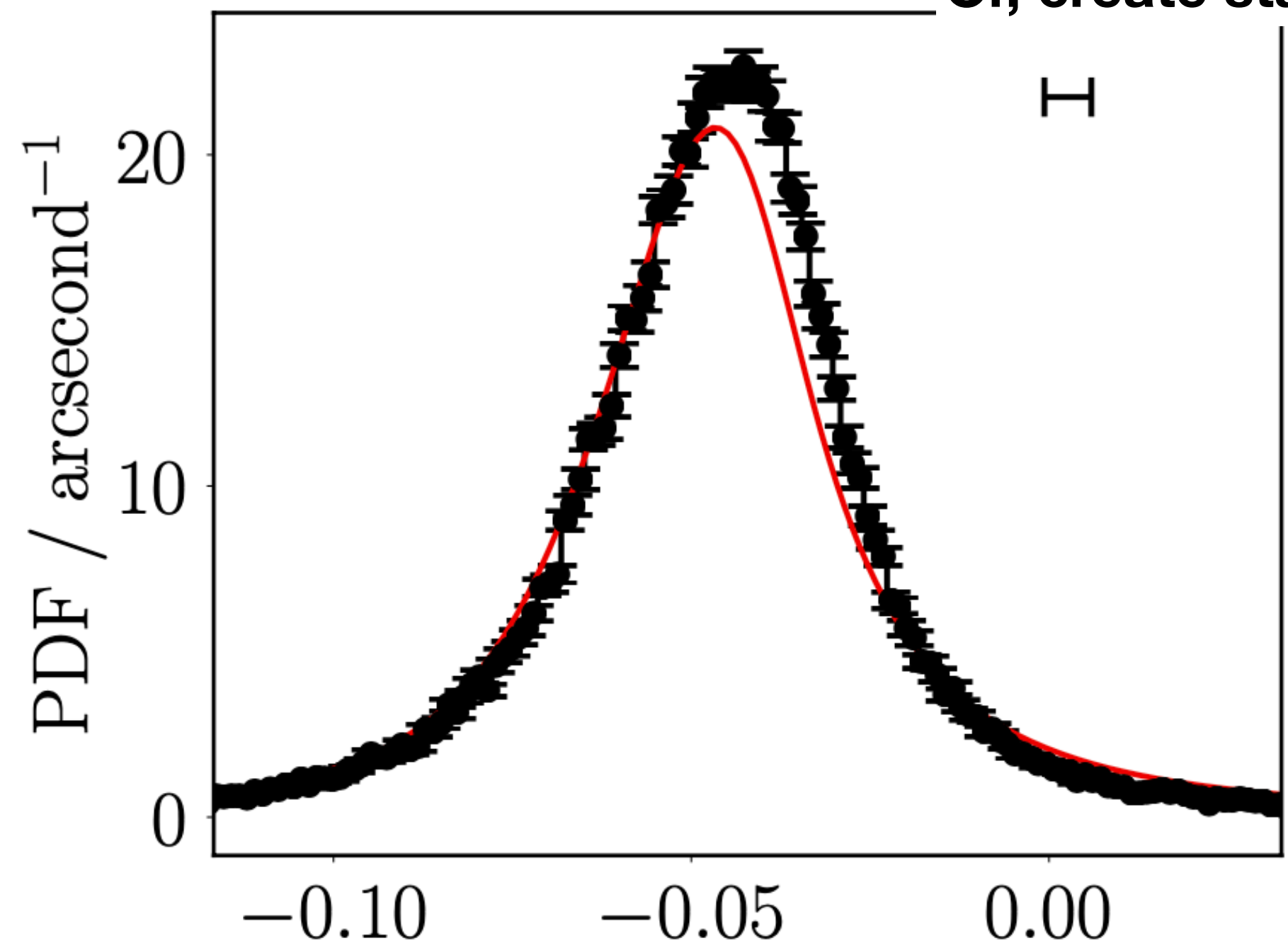
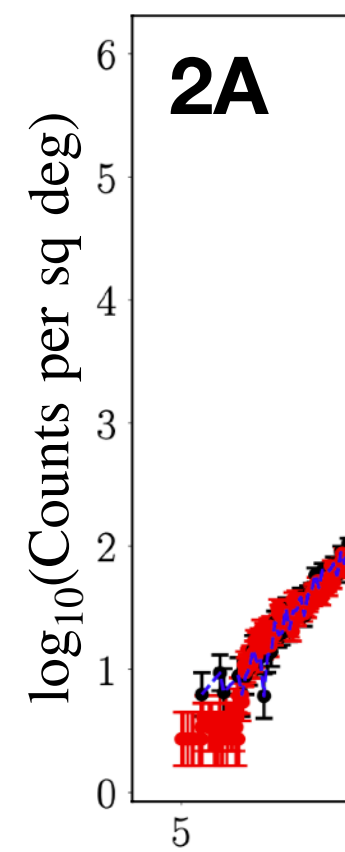
In each sightline (10s of sq deg for good bright source counting N):

1. Cross-match your high angular resolution, high astrometric precision data to LSST to obtain separation distributions
2. **Create systematics model for all non-centroid astrometric components of uncertainty**
3. Fit full AUF to data, allowing centroid Gaussian uncertainty to be fit
4. Repeat for each brightness (and effectively different astrometric uncertainty)
5. Derive fit-quoted astrometric uncertainty relations

Create crowding-caused perturbation model, for example:

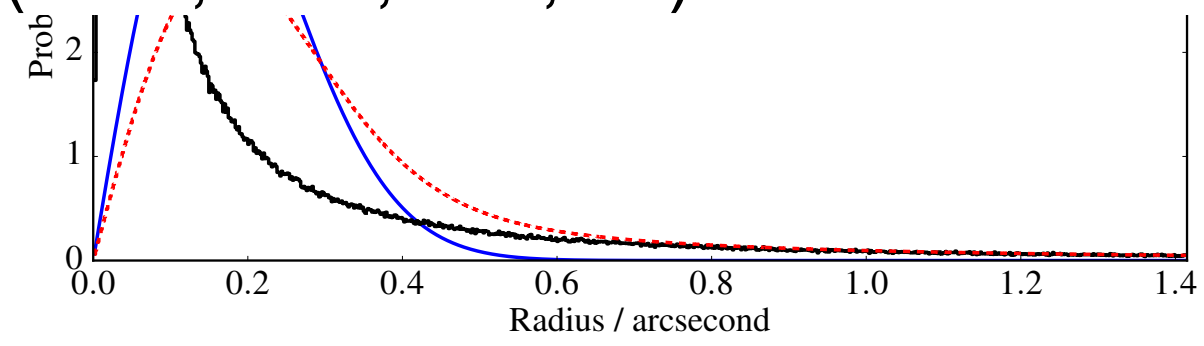
- A. Verify model source count densities match observed data
- B. Randomly draw perturbing sources within your PSF (“darts at a dartboard”)
- C. Repeat lots of times to get a distribution of perturbation offsets
- D. Repeat however many times you have different perturbation algorithms
- E. Combine your perturbation algorithms

Or, create statistical proper motion distribution, or...



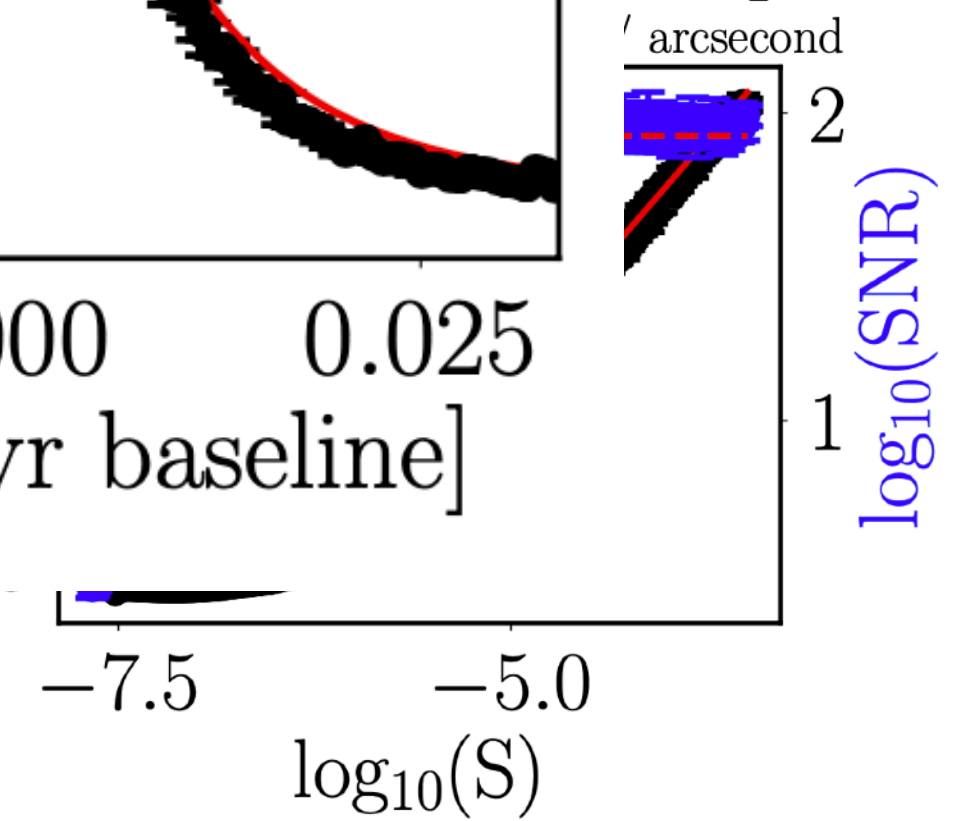
2B

Wilson (2023, RASTI, 2, 1) Δl / arcsecond [10yr baseline]
 Gaia eDR3 - Gaia Collaboration et al. (2021, A&A, 649, A1)



$$SNR = \frac{1}{\sqrt{c \times S + b + (a \times S)^2}}$$

$$H = 1 - \sqrt{1 - \min(1, a \times SNR^2)}$$



Verifying Astrometry: Fitting Centroid Uncertainty

In each sightline (10s of sq deg for good bright source counting N):

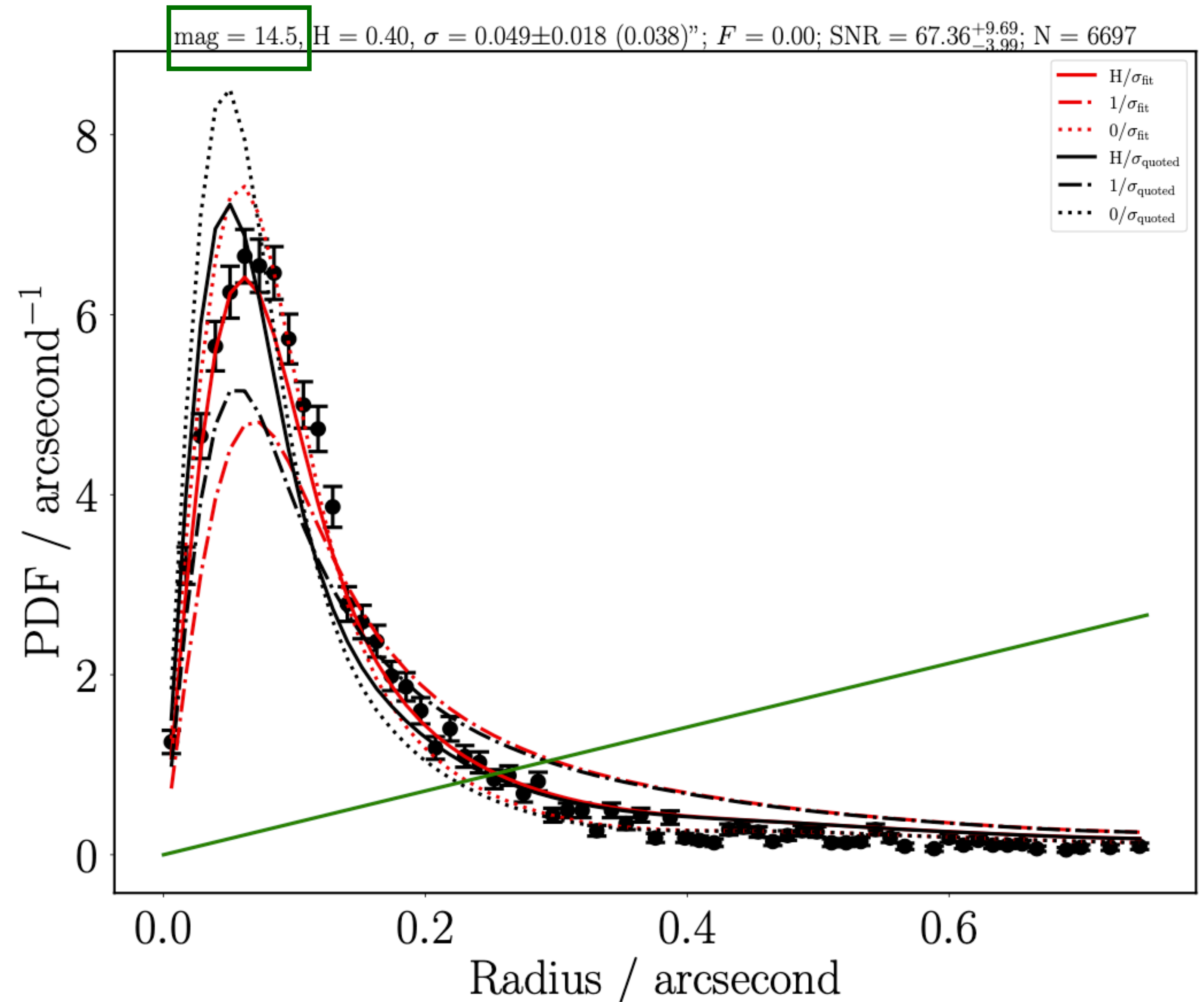
1. Cross-match your high angular resolution, high astrometric precision data to LSST to obtain separation distributions
2. Create systematics model for all non-centroid astrometric components of uncertainty
3. Fit full AUF to data, allowing centroid Gaussian uncertainty to be fit
4. Repeat for each brightness (and effectively different astrometric uncertainty)
5. Derive fit-quoted astrometric uncertainty relations

For each magnitude (uncertainty) slice in a given sightline, combine centroid uncertainty (Gaussian) and other AUF components (empirical) and fit for best-fitting sigma-value.

$$h_\gamma = h_{\gamma, \text{centroiding}} * h_{\gamma, \text{perturbation}} * \dots$$

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

Also include false positive match rate (F) in case simple match case was not perfect



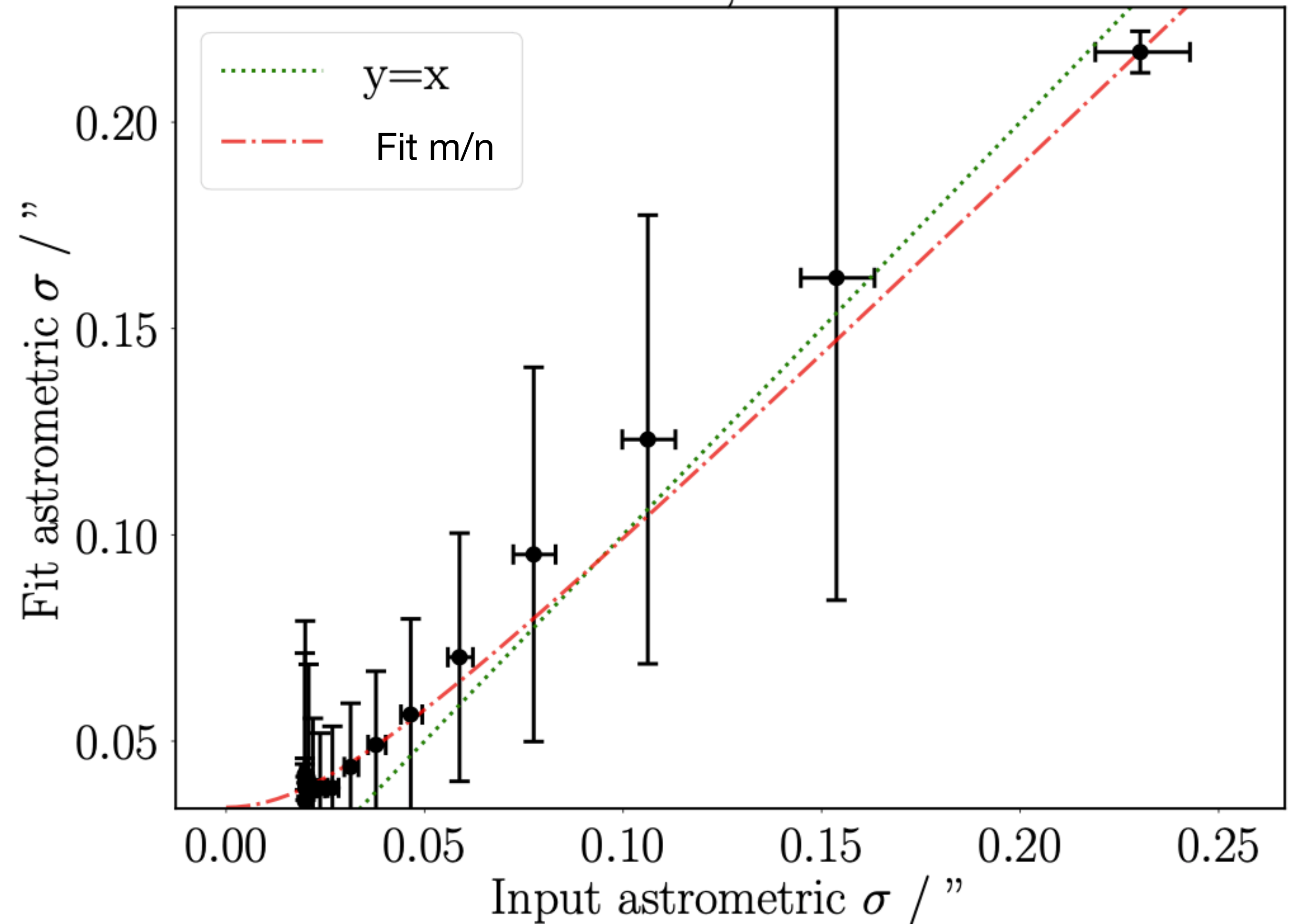
Verifying Astrometry: Characterisation

In each sightline (10s of sq deg for good bright source counting N):

1. Cross-match your high angular resolution, high astrometric precision data to LSST to obtain separation distributions
2. Create systematics model for all non-centroid astrometric components of uncertainty
3. Fit full AUF to data, allowing centroid Gaussian uncertainty to be fit
4. Repeat for each brightness (and effectively different astrometric uncertainty)
5. **Derive fit-quoted astrometric uncertainty relations**

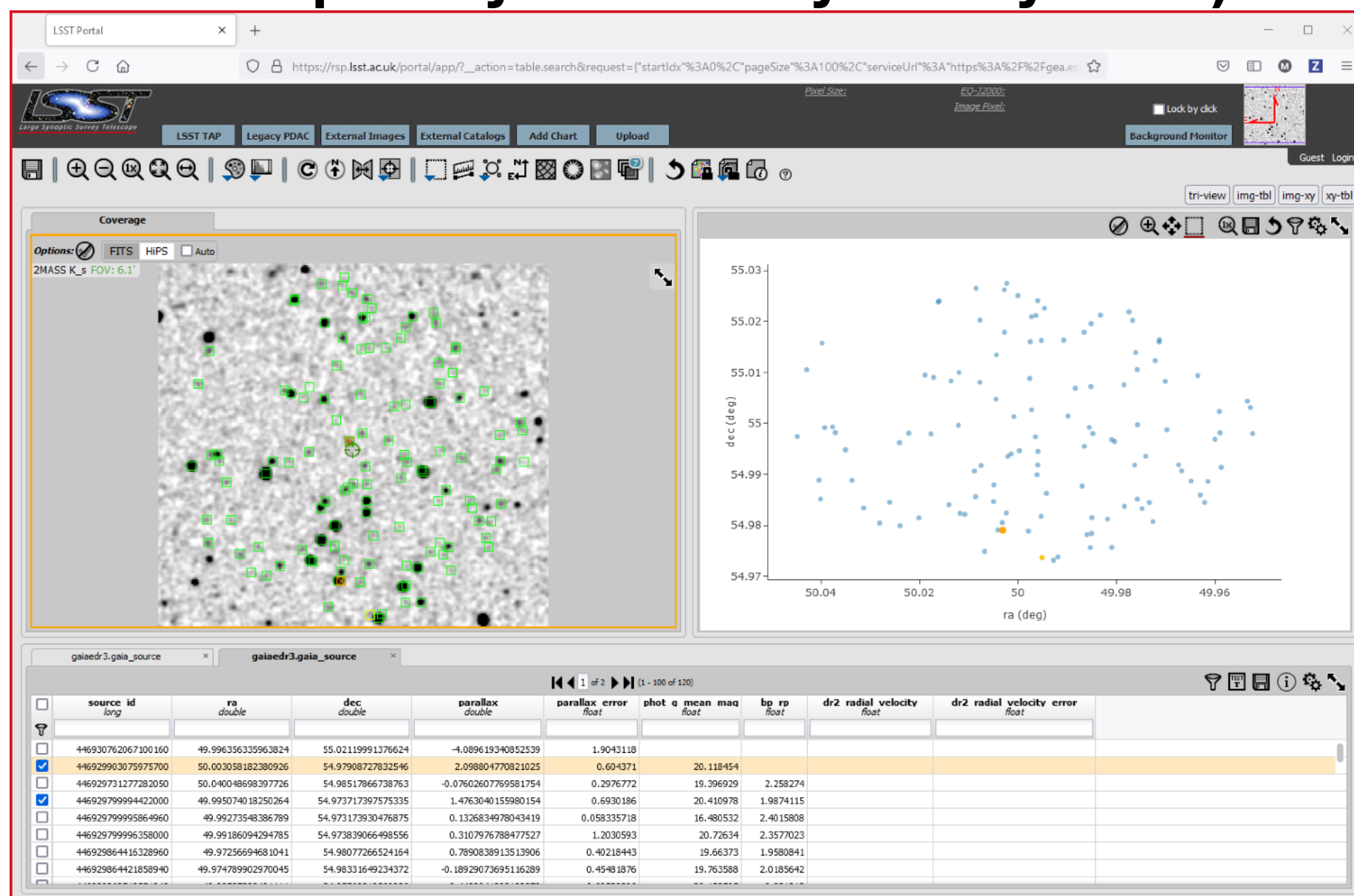
Fit for $y = \sqrt{(mx)^2 + n^2}$ (or, optionally, $y = mx + n$) to account for simple systematic bias n missing and compensating scaling factor m at lower SNR data

$$l = 130.0, b = -10.0$$
$$m = 0.93, n = 0.03$$



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)

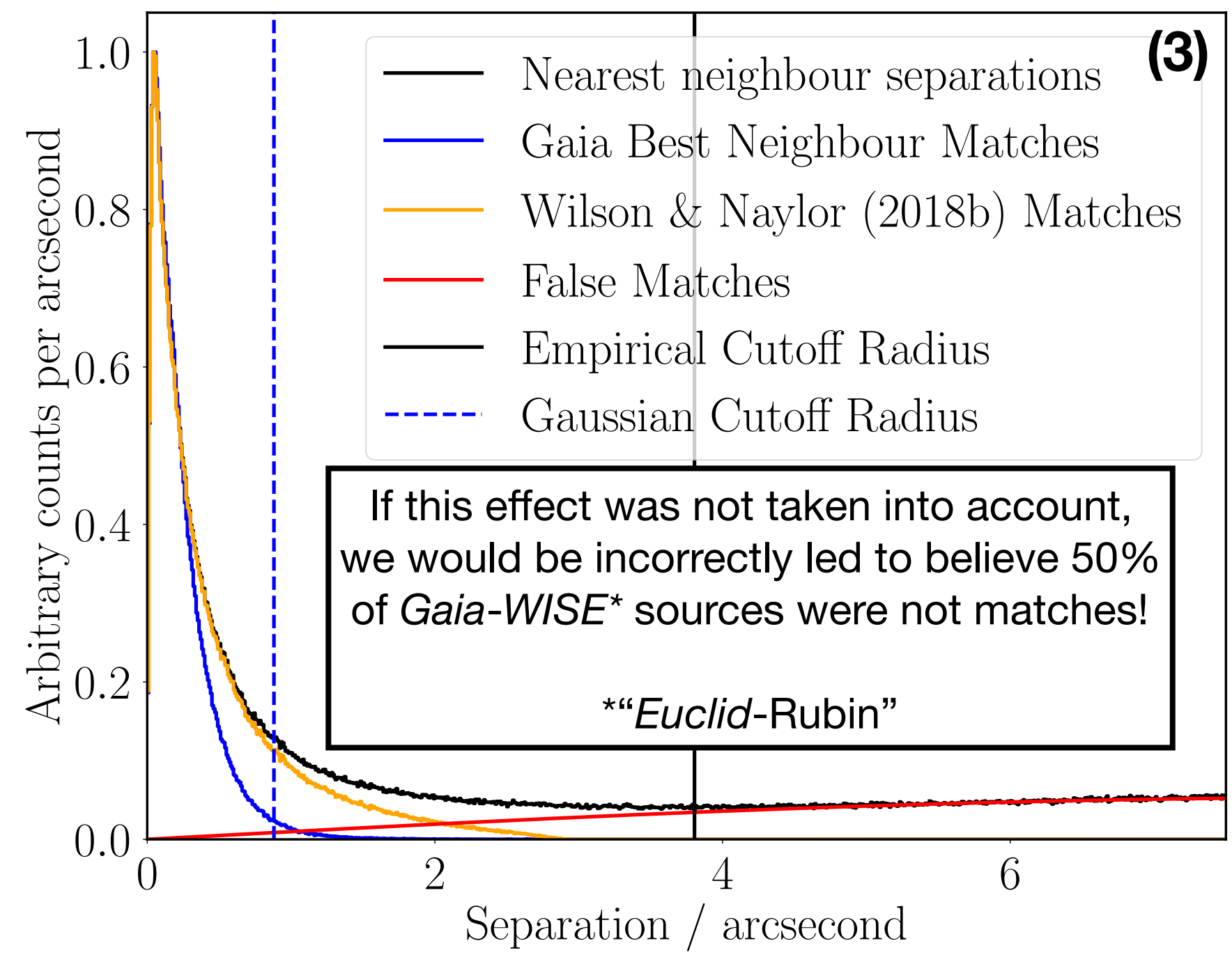
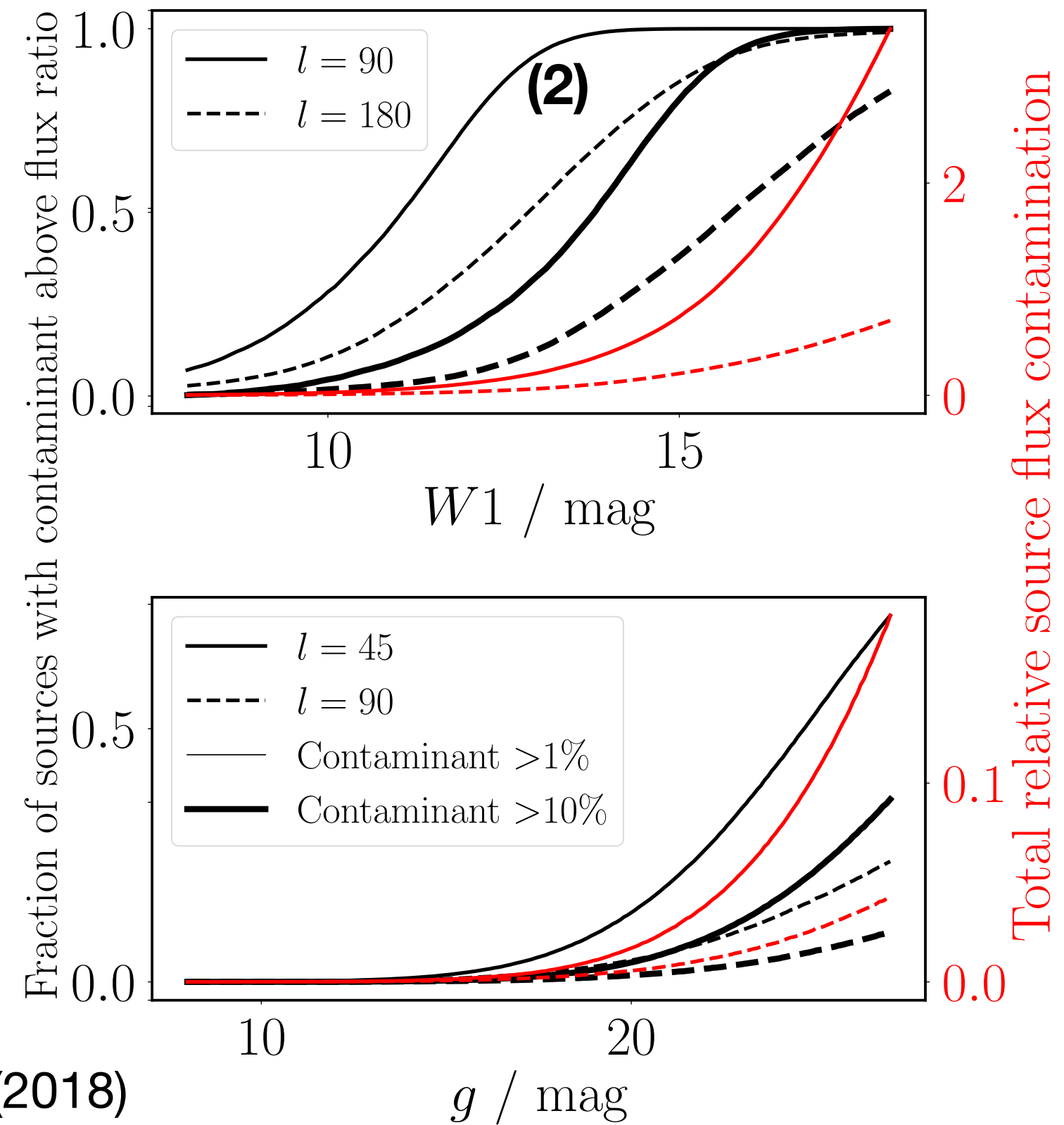
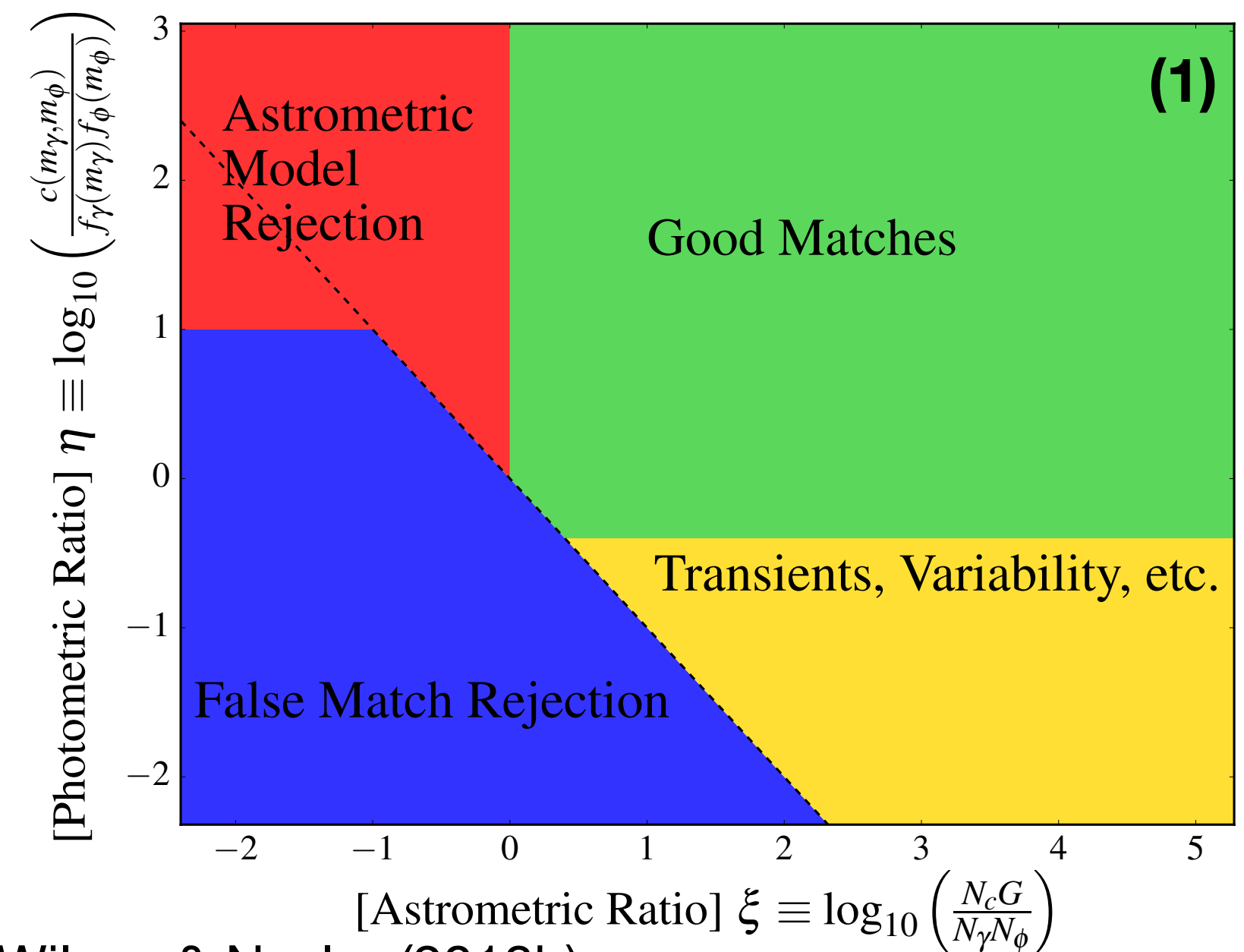
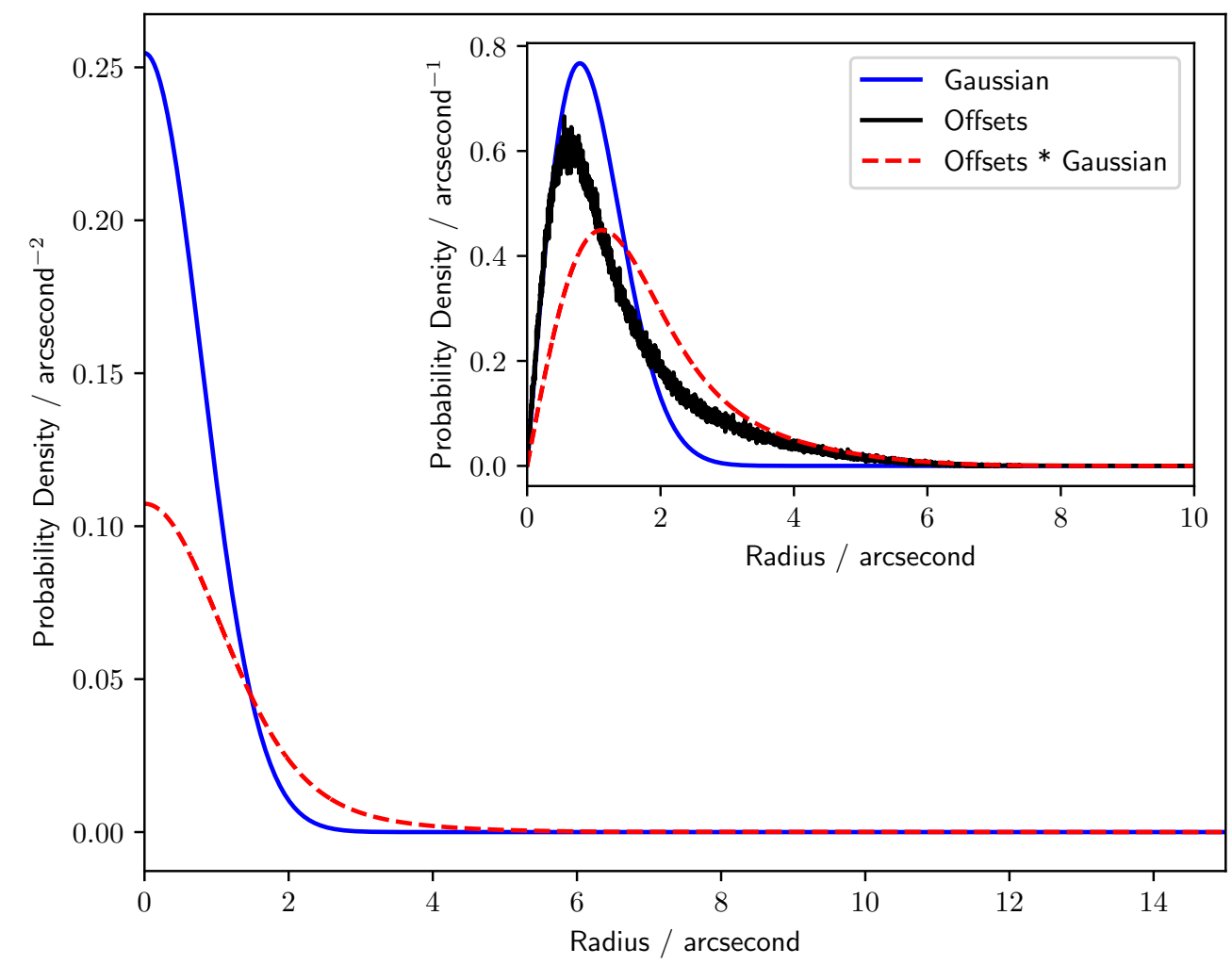
Example columns:

- 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 a 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

Why Use Macauff's 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)



Conclusions

- **Blended star contamination causes positional shifts, now modelled robustly for the first time in the AUF**
- **Symmetric data-driven photometric likelihood now possible**
- **LSST will suffer of order 10% flux contamination in the future**
 - Important for extinction/distance; “1% photometry”?
 - Modelling of statistical flux contamination allows for the recovery of “true” fluxes
- **LSST will suffer at least one extra source (possibly up to 10!) in each 2” matching circle**
 - Need to use astrometric uncertainty to reduce length scale over which matches are considered
 - Can use photometry in catalogues to break these false match degeneracies
- **Can include other effects, like unknown proper motions or DCR, easily within AUF match framework**
- **High dynamic range matches must account for differential crowding matching to ancillary or historic data**
- **Accounting for these systematics, we can *confirm* quoted catalogue astrometric uncertainties in more extremes than would otherwise be possible, avoiding mistakenly thinking the pipeline values are “wrong”**
- **Upcoming LSST:UK cross-match service macauff – let me know your thoughts/needs/hopes/dreams**

Wilson & Naylor (2017, MNRAS, 468, 2517); Wilson & Naylor (2018a, MNRAS, 473, 5570);
Wilson & Naylor (2018b, MNRAS, 481, 2148); Wilson (2022, RNAAS); Wilson (2022, RASTI, 2, 1);
Wilson & Naylor (in prep.) – more AUF-related improvements!



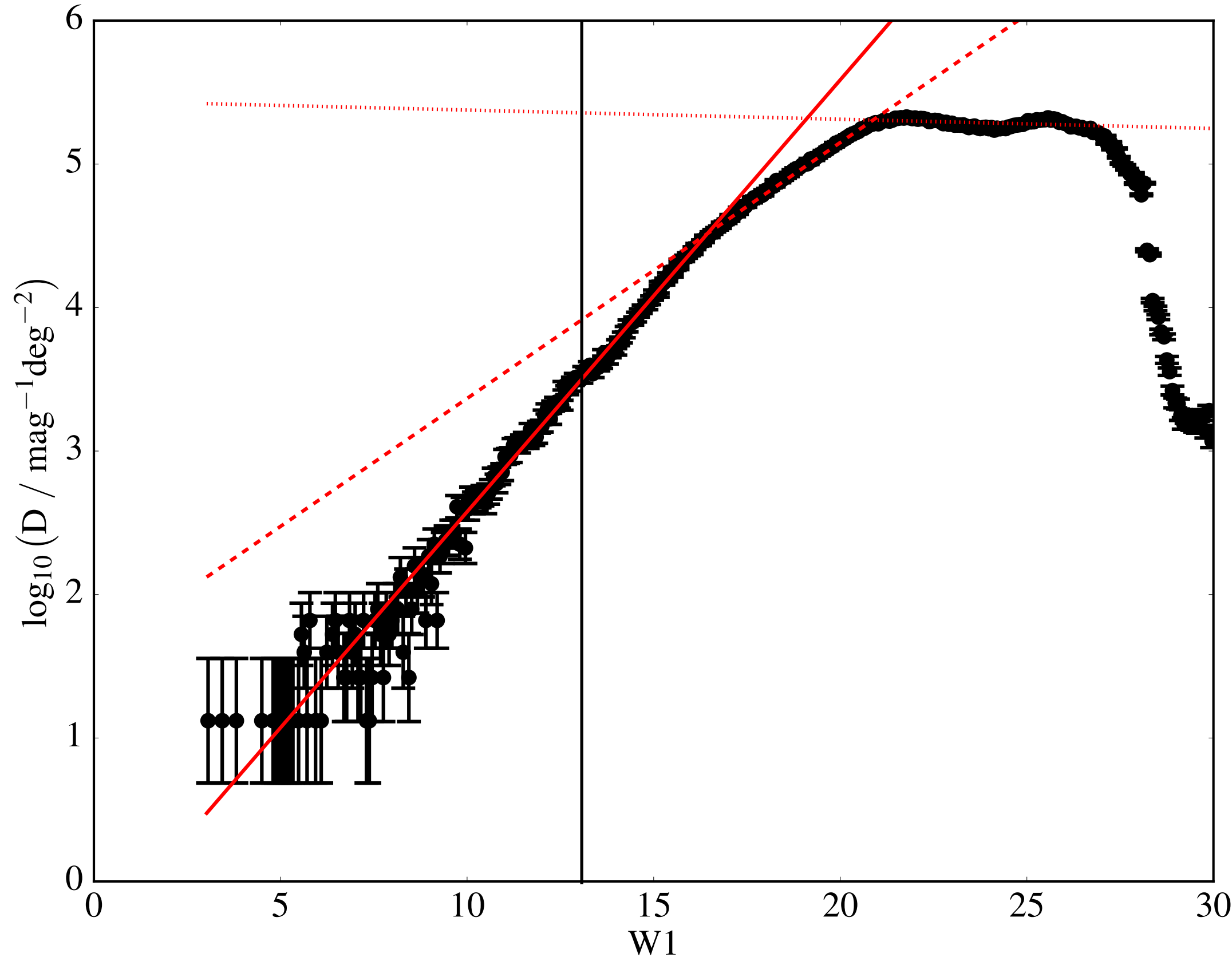
Science and
Technology
Facilities Council



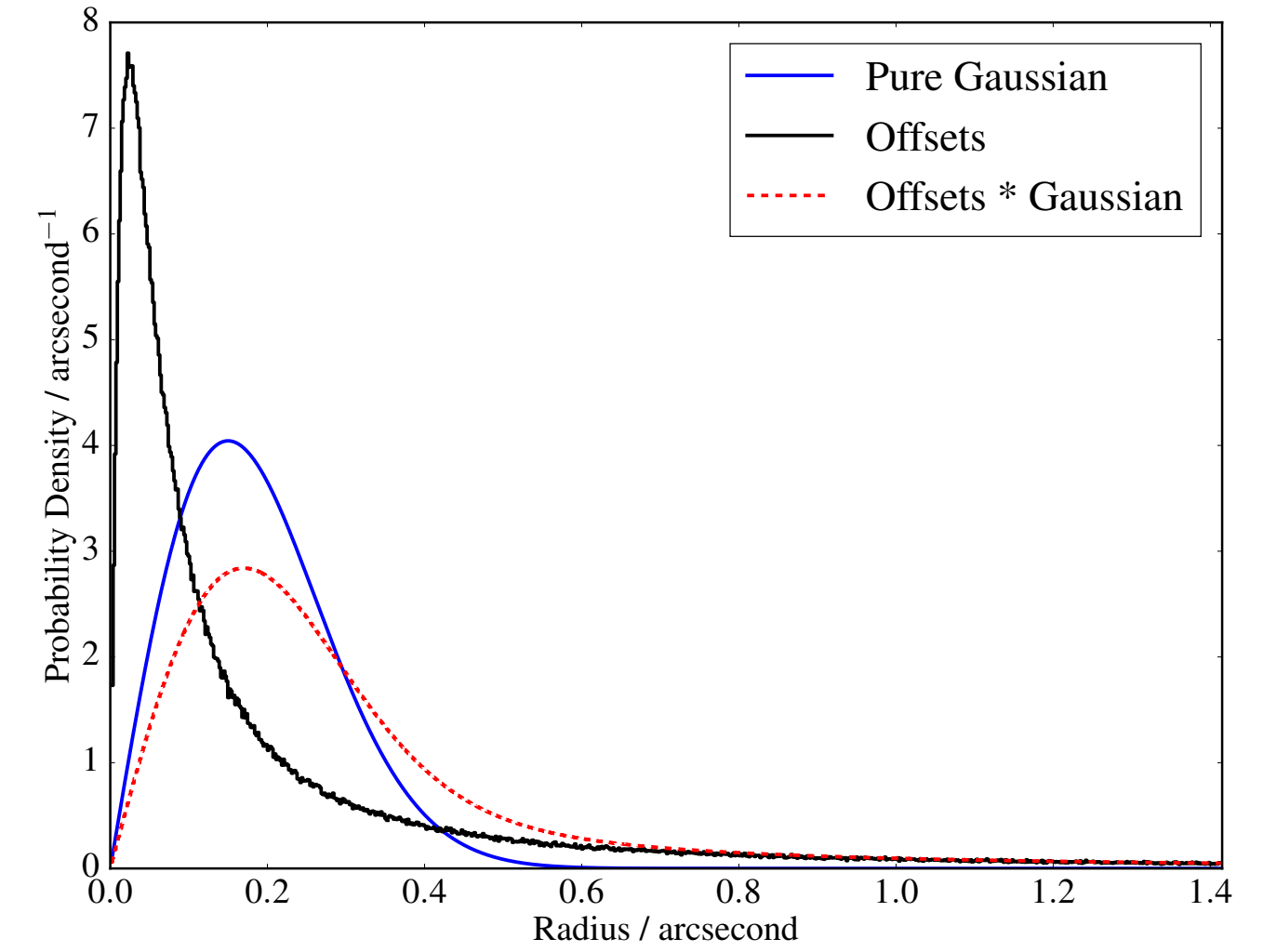
@Onoddil @pm.me
.github.io

Tom J Wilson @onoddil

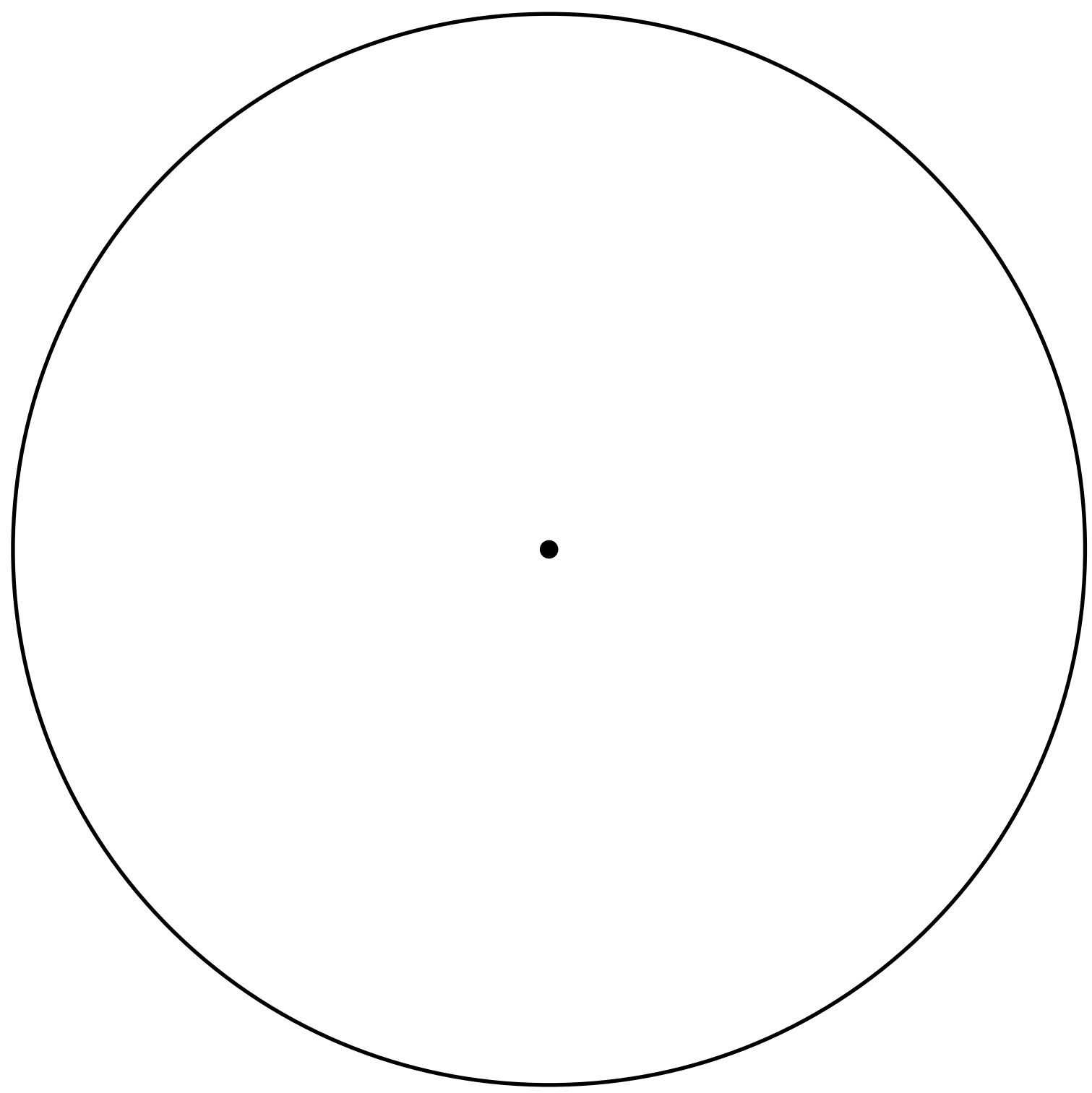
Building Empirical AUFs



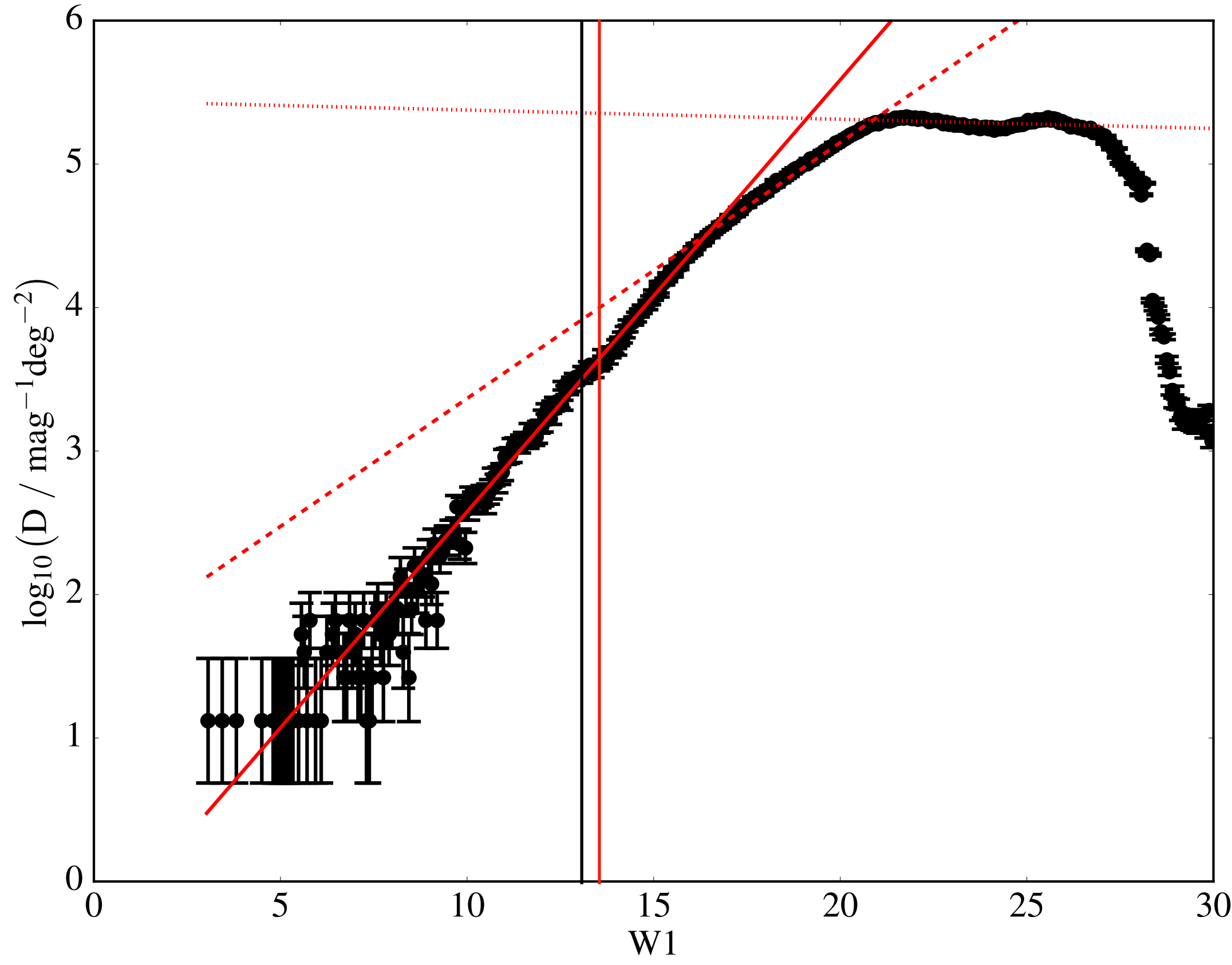
(sources per PSF circle $\sim 10^{-6}$ sources per mag per sq deg)



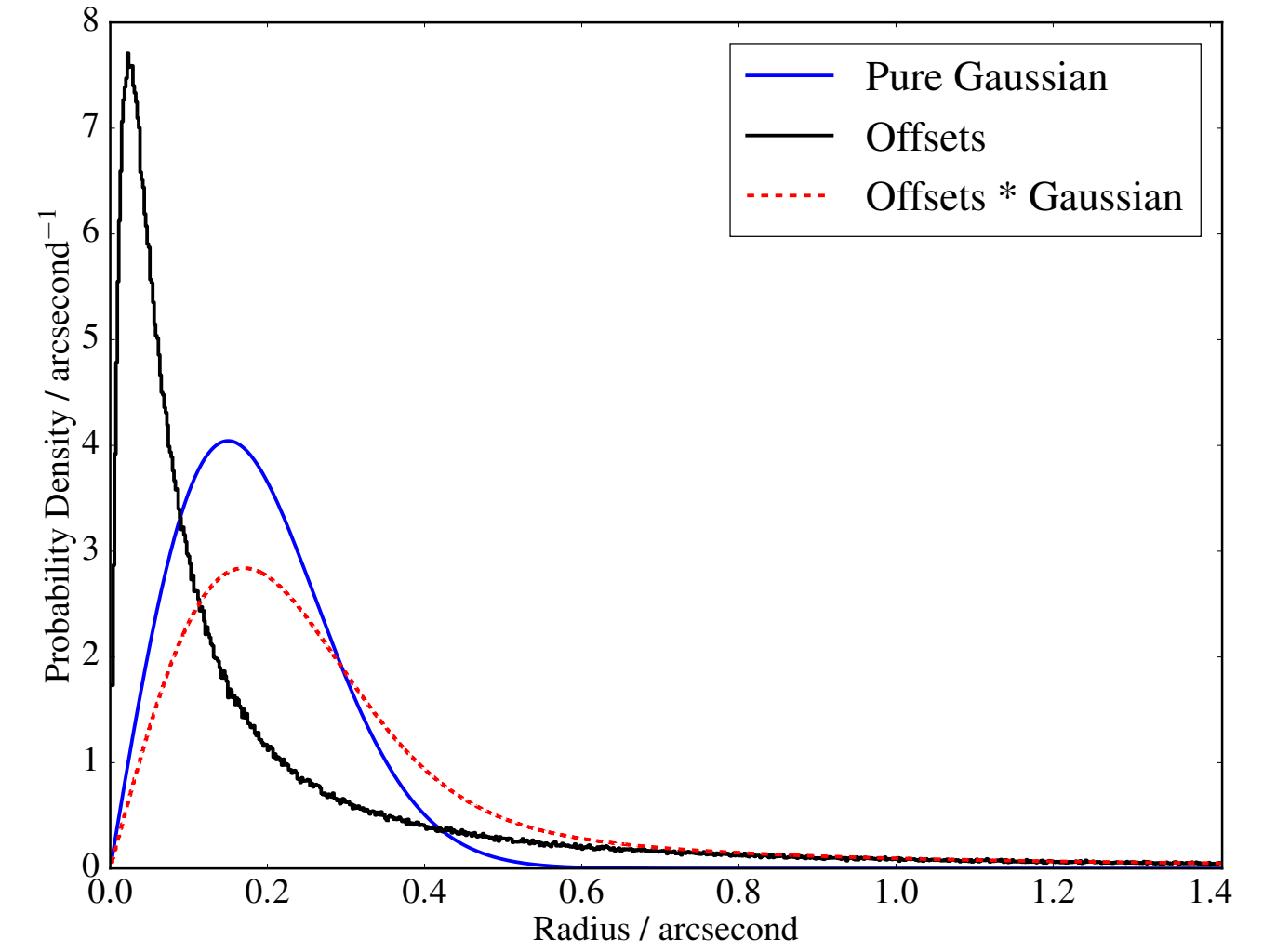
PSF radius ~ 1.2 FWHM (Rayleigh criterion)



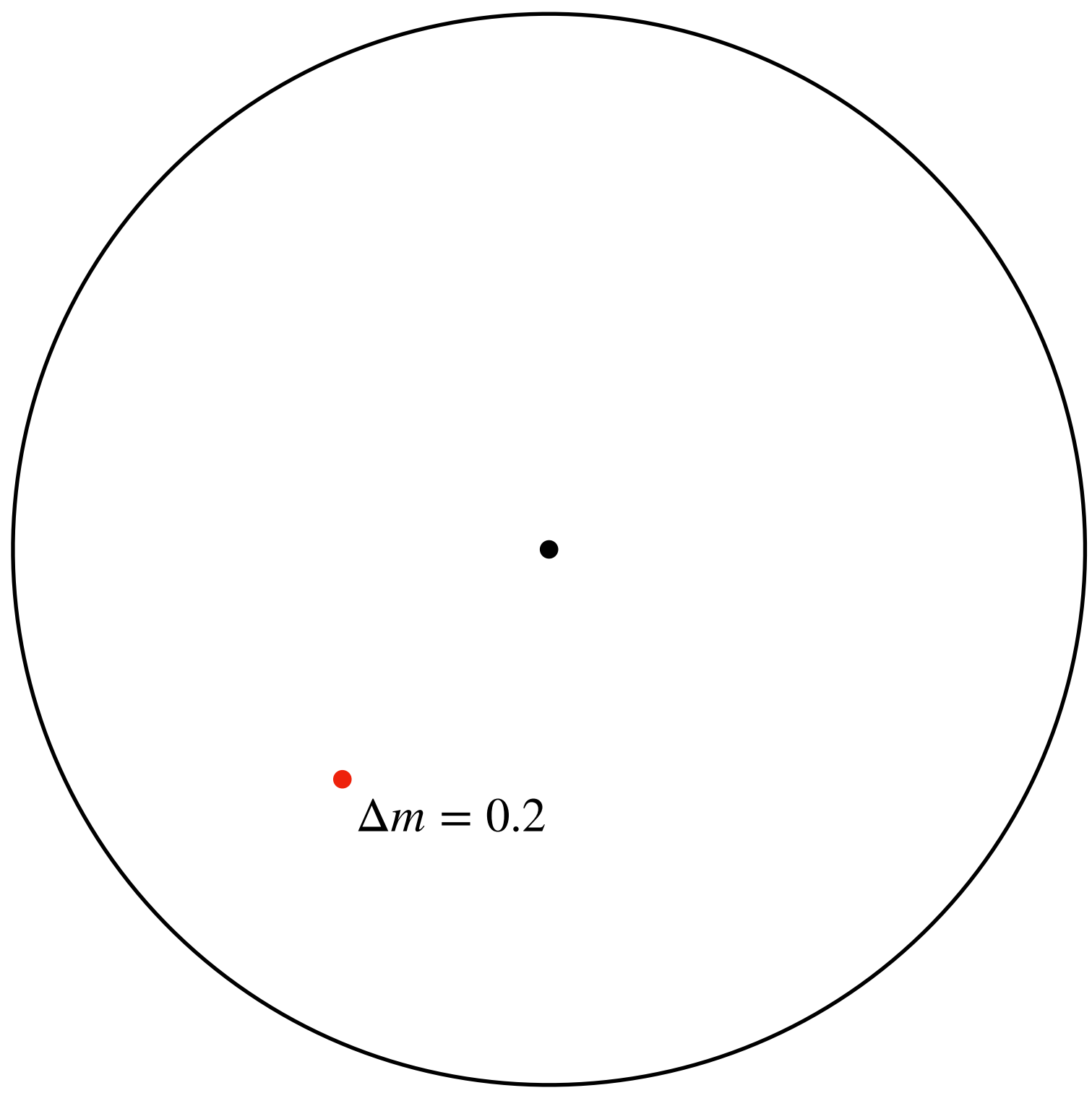
Building Empirical AUFs



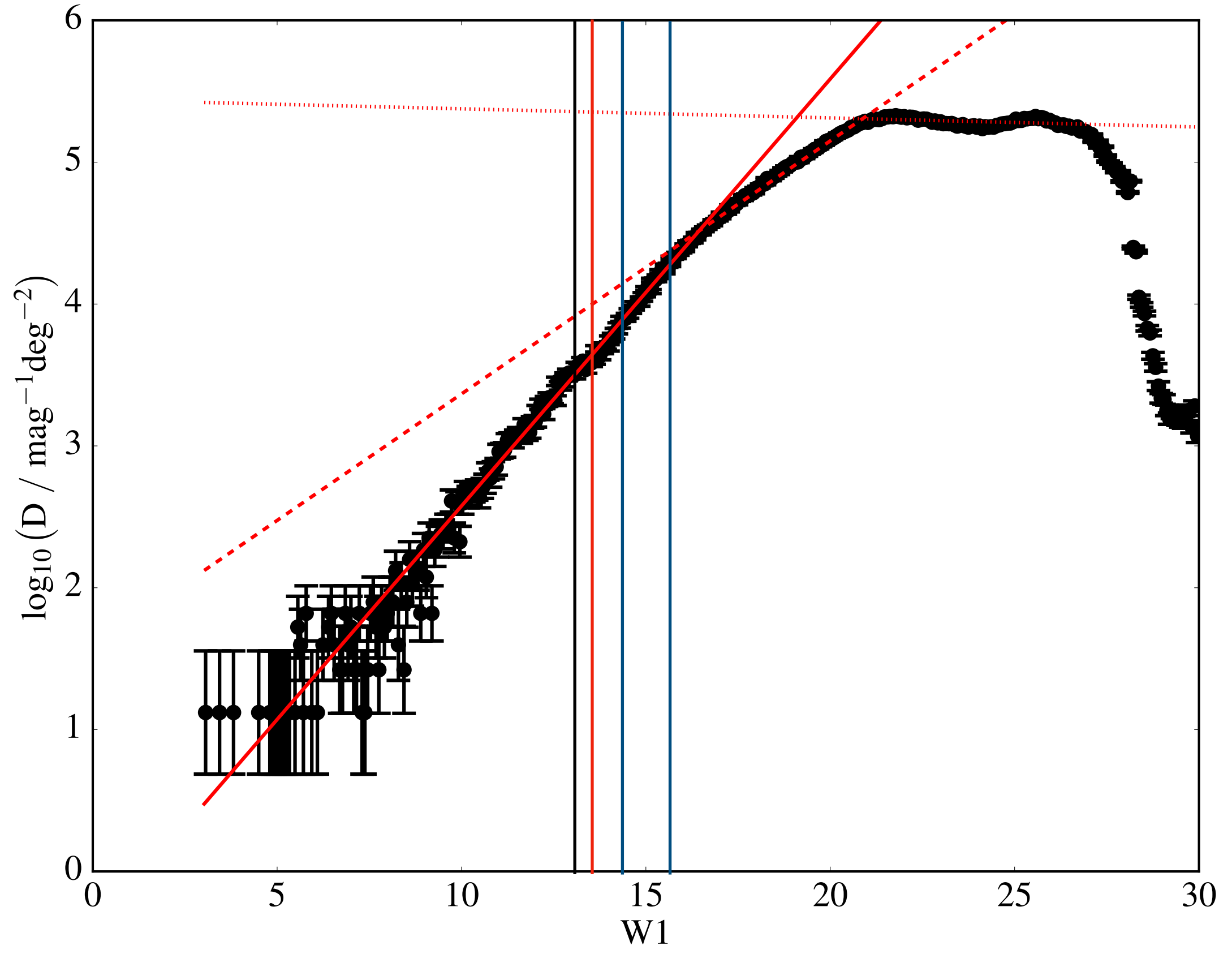
(sources per PSF circle $\sim 10^{-6}$ sources per mag per sq deg)



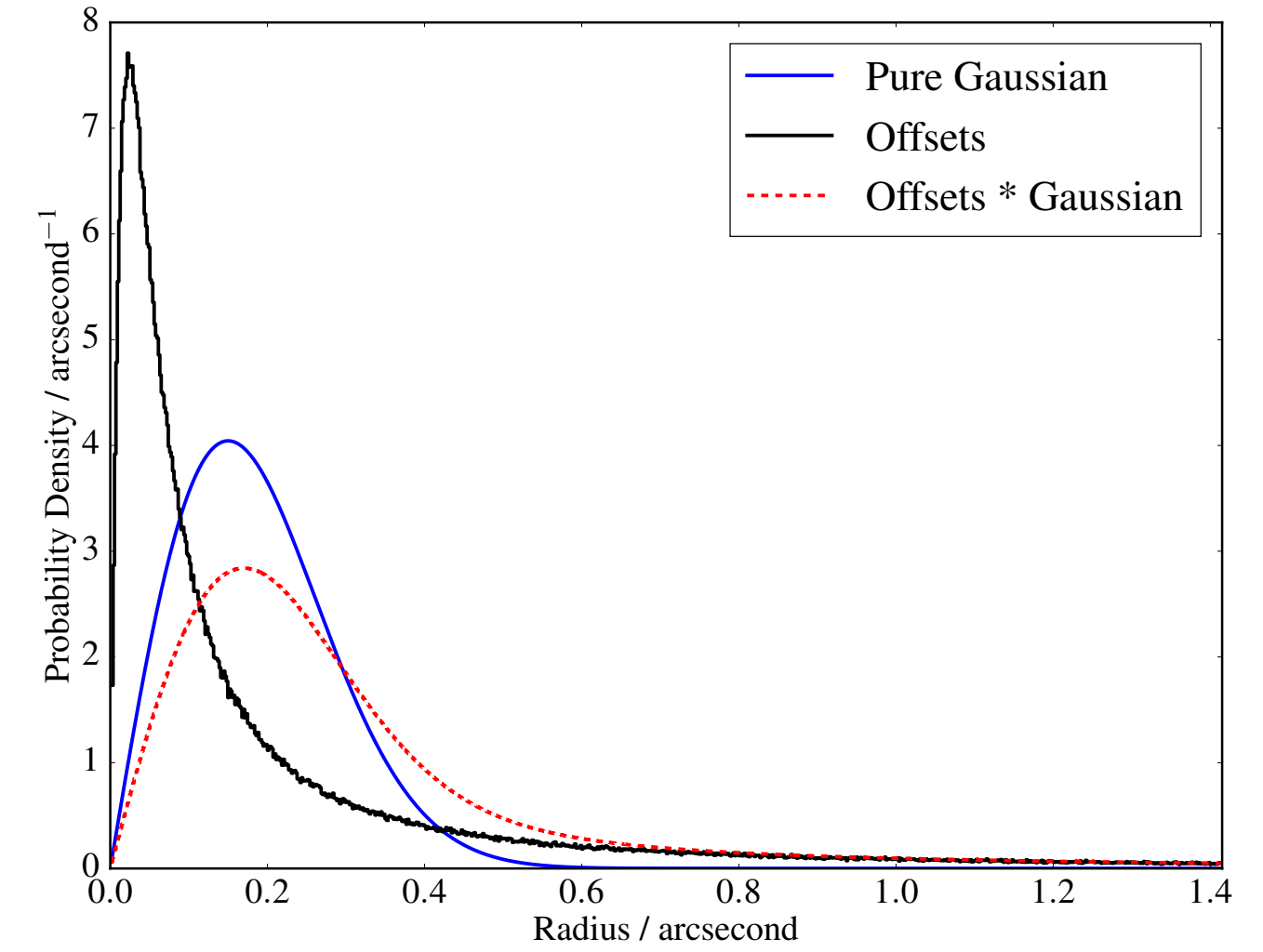
PSF radius ~ 1.2 FWHM (Rayleigh criterion)



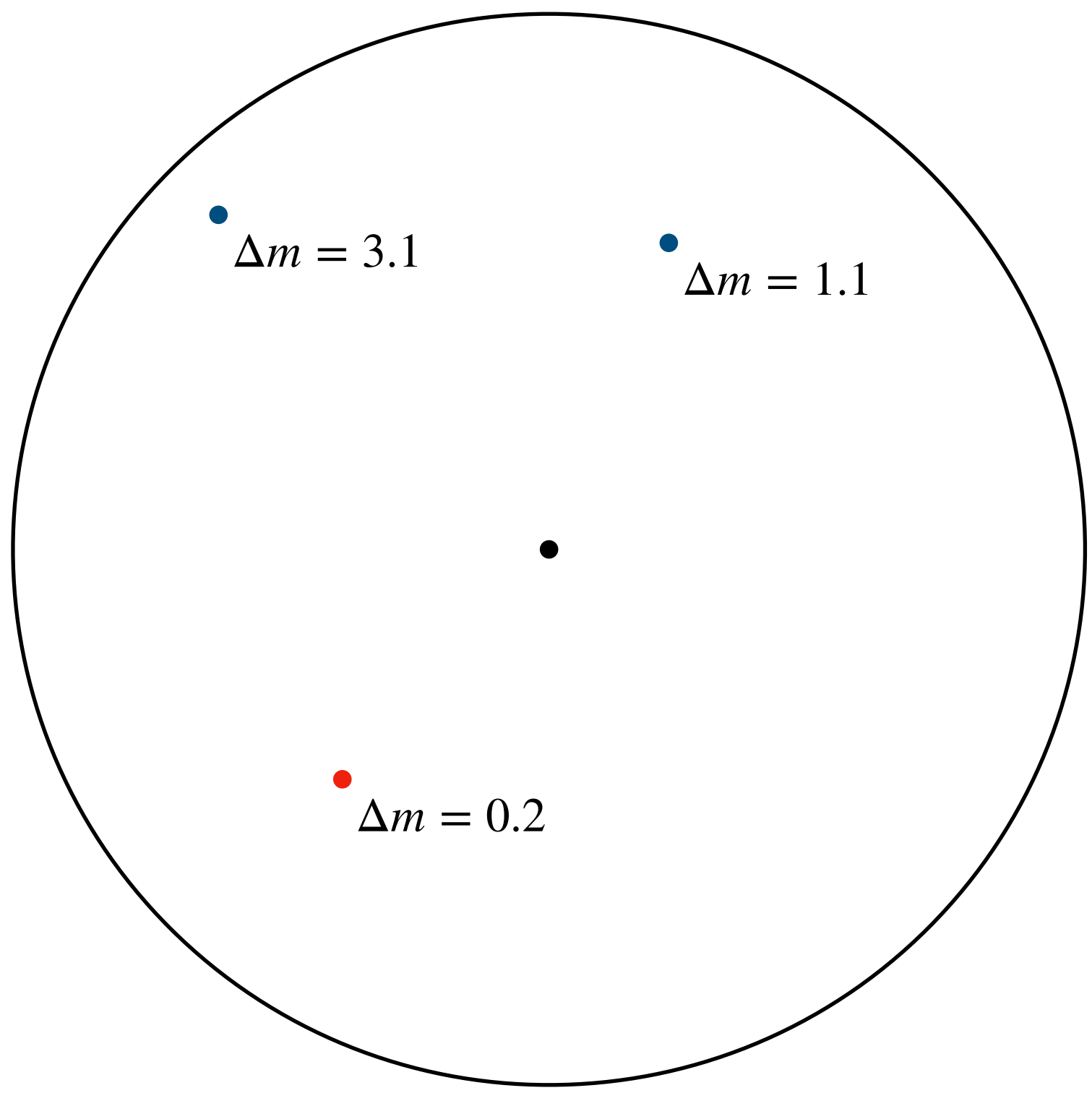
Building Empirical AUFs



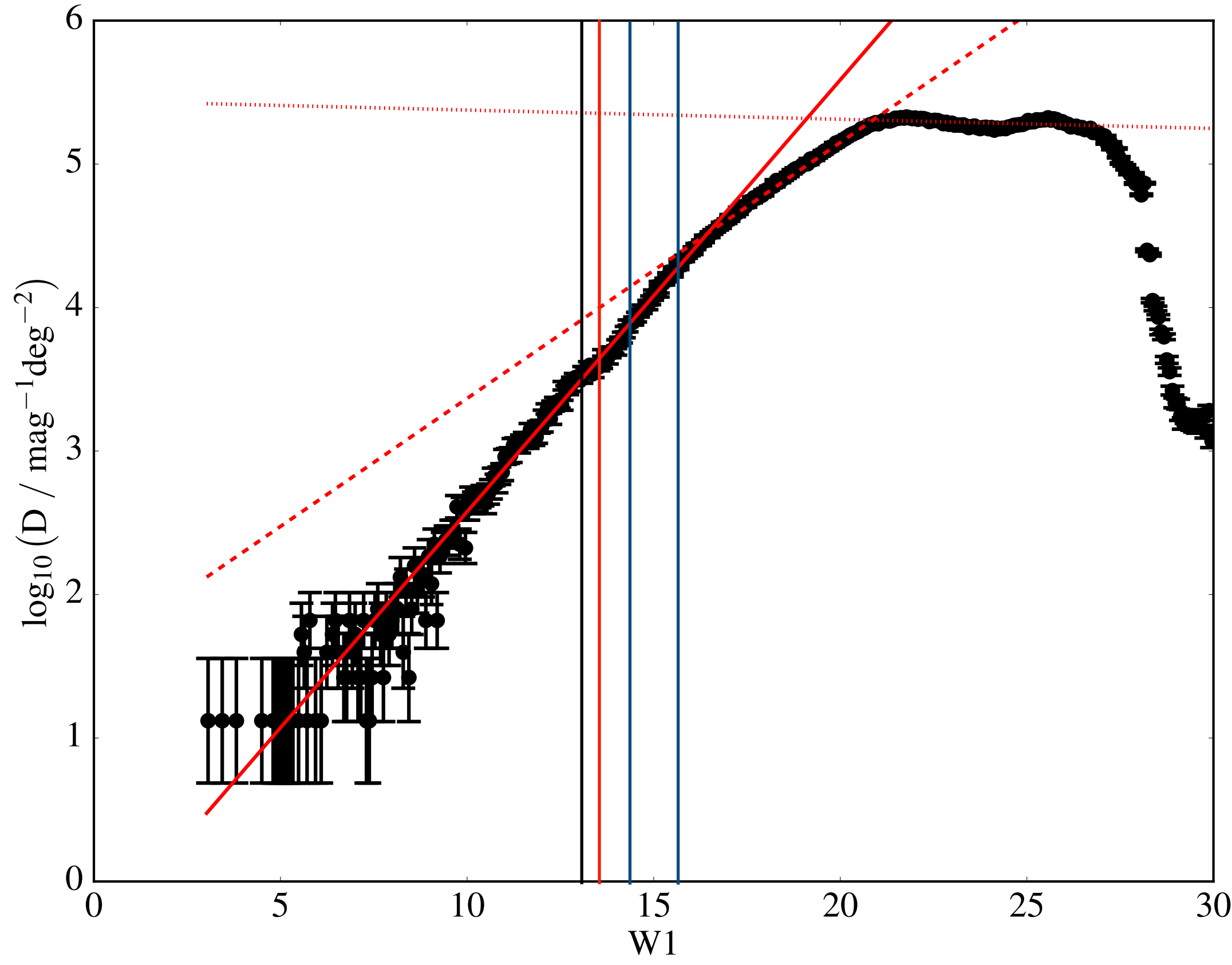
(sources per PSF circle $\sim 10^{-6}$ sources per mag per sq deg)



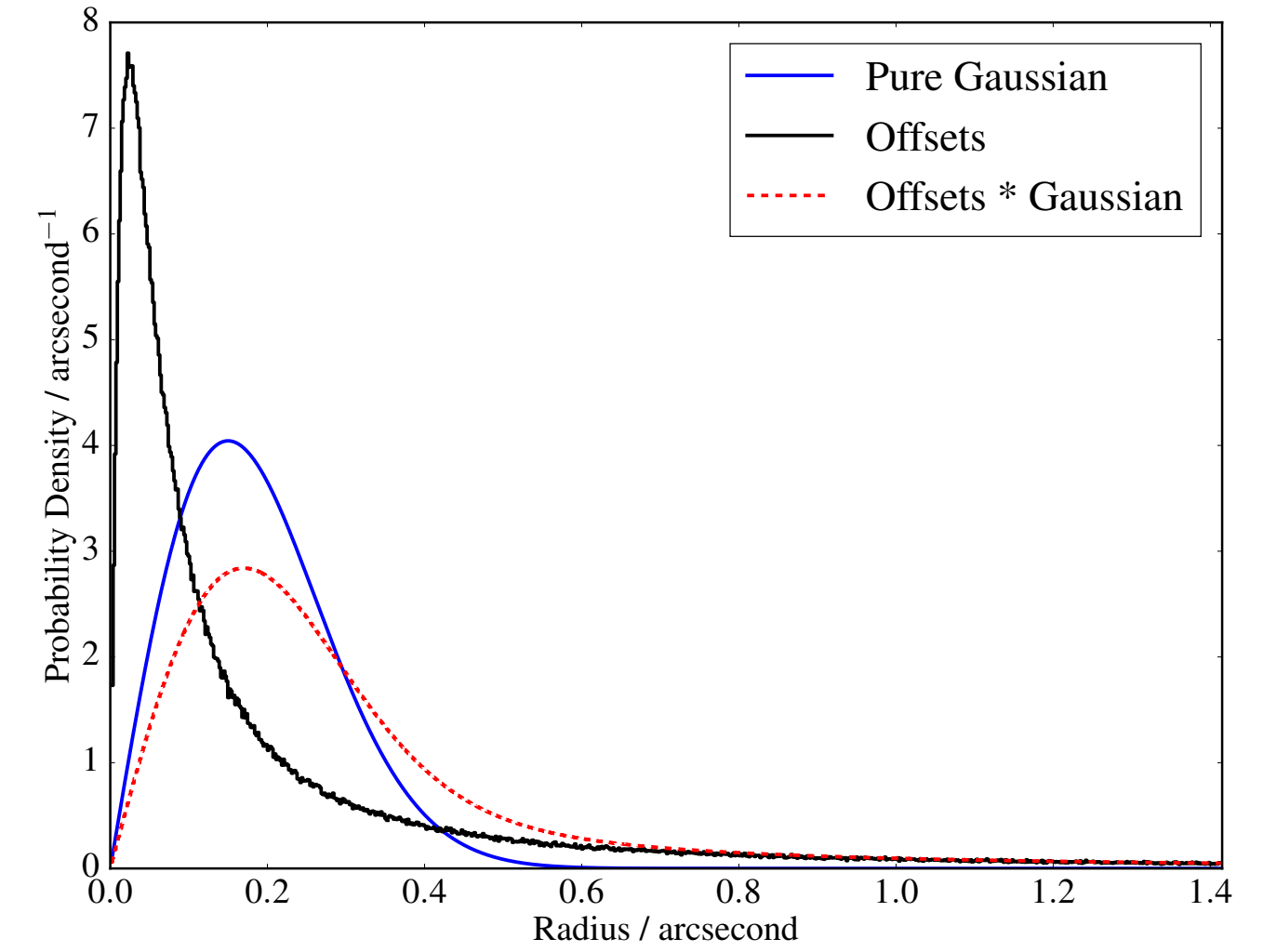
PSF radius ~ 1.2 FWHM (Rayleigh criterion)



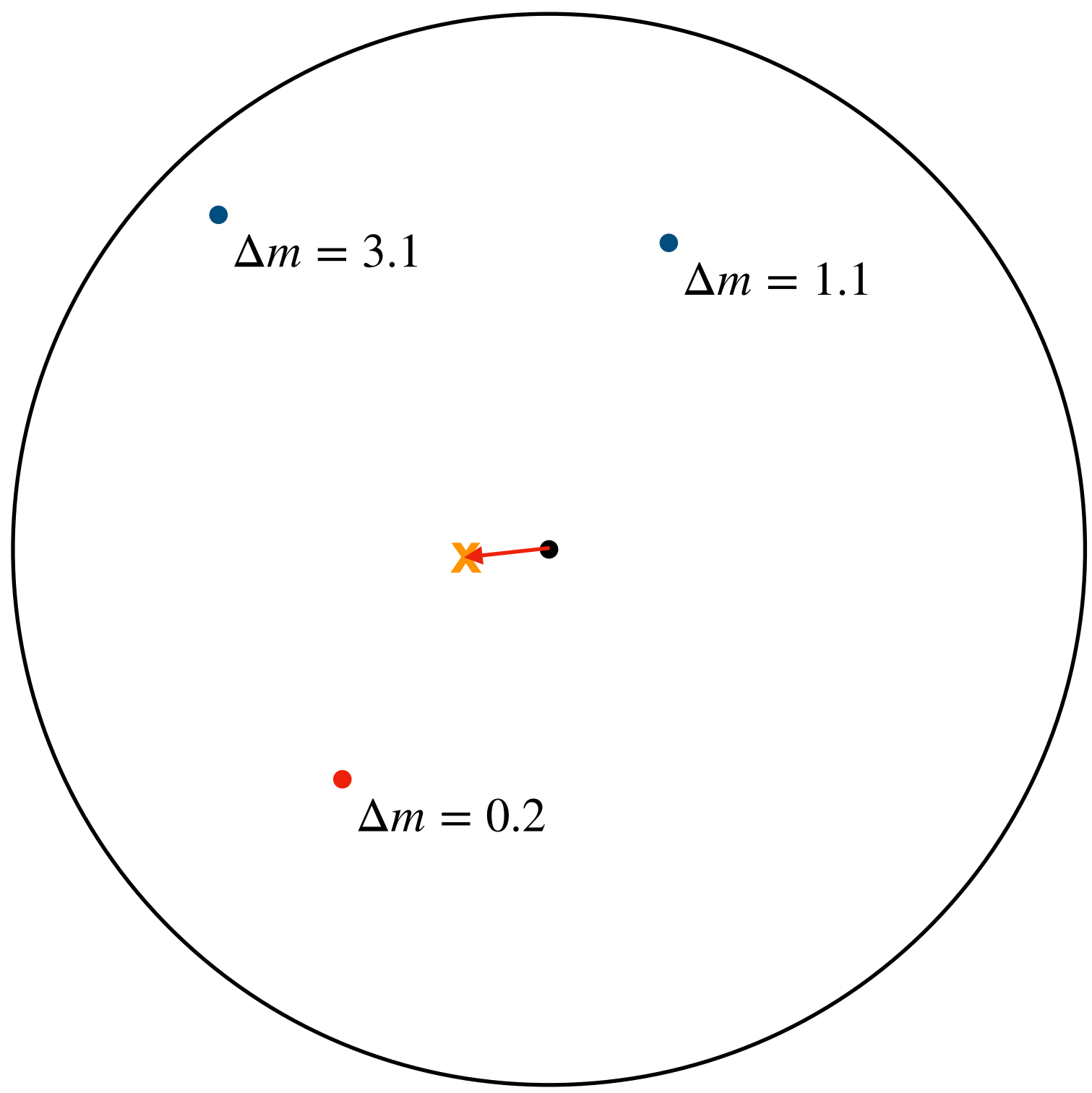
Building Empirical AUFs



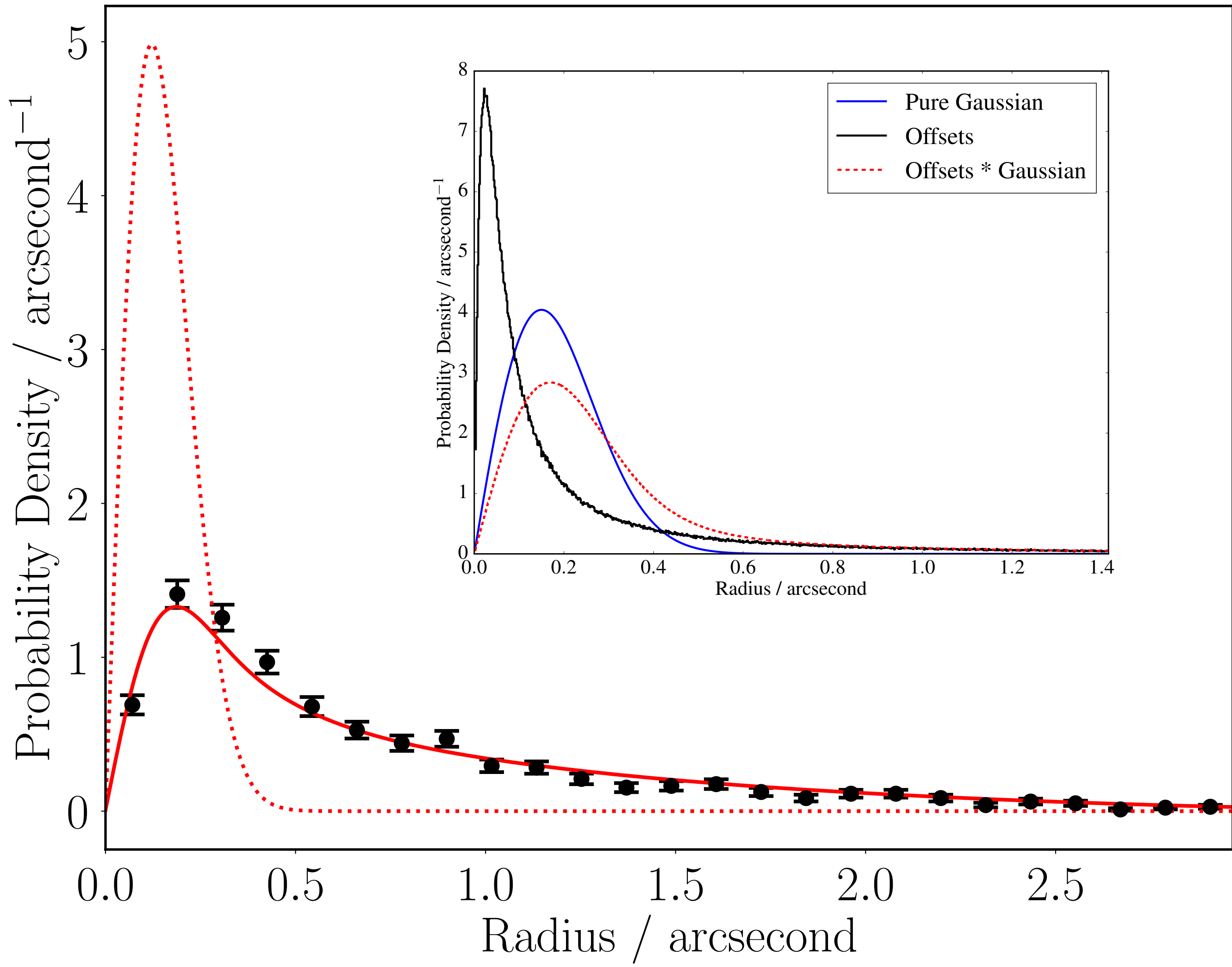
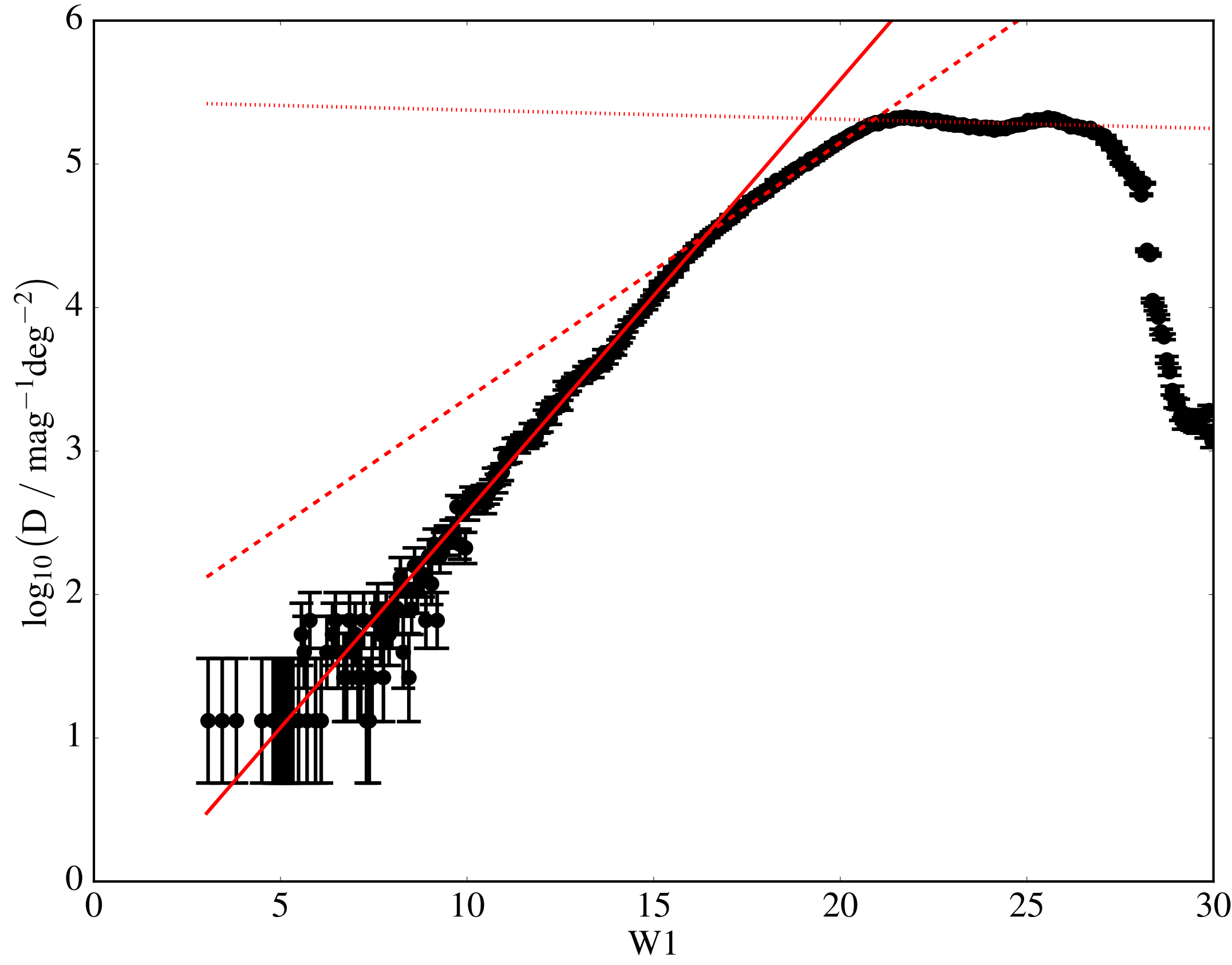
(sources per PSF circle $\sim 10^{-6}$ sources per mag per sq deg)



PSF radius ~ 1.2 FWHM (Rayleigh criterion)



Building Empirical AUFs



WISE - Wright et al. (2010)
Gaia DR2 - Gaia Collaboration, Brown A. G. A., et al. (2018)
Wilson & Naylor (2018b)
TRILEGAL - Girardi et al. (2005)