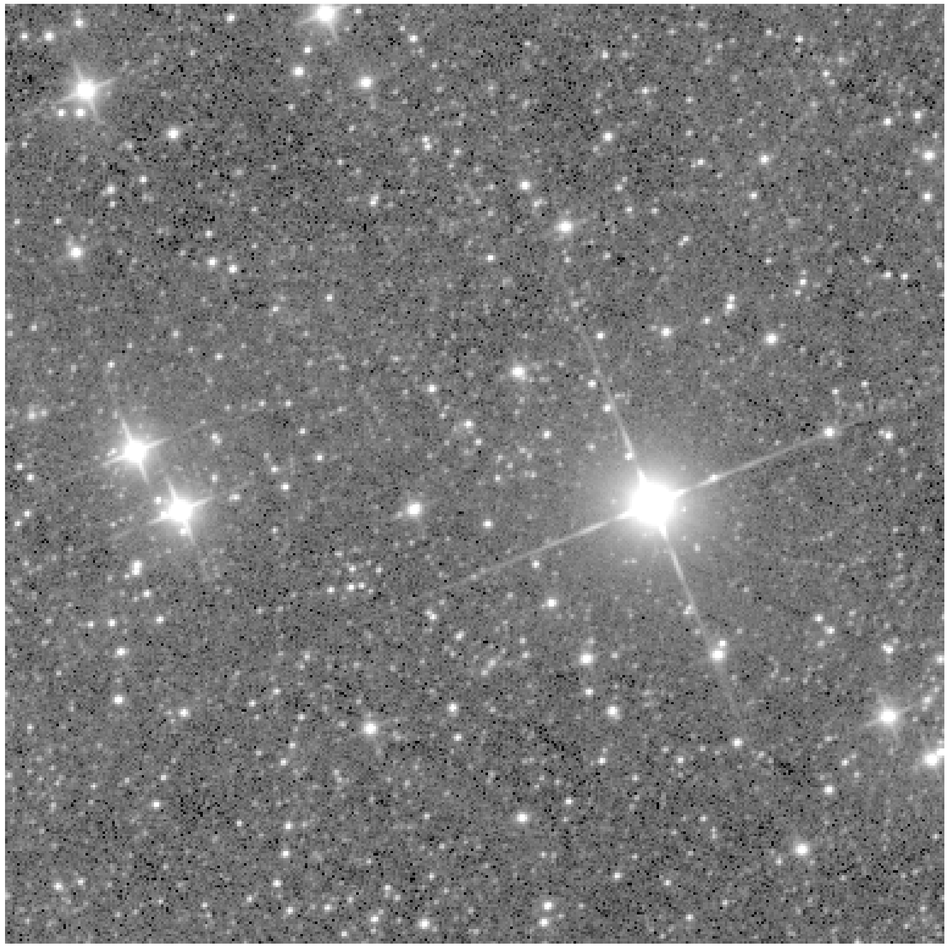
Put me here!

Towards Solving the Catalogue Cross-Match Problem

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

Alternatively — Space Is Crowded, Or: How I Learned To Start Worrying and Hate the Mess

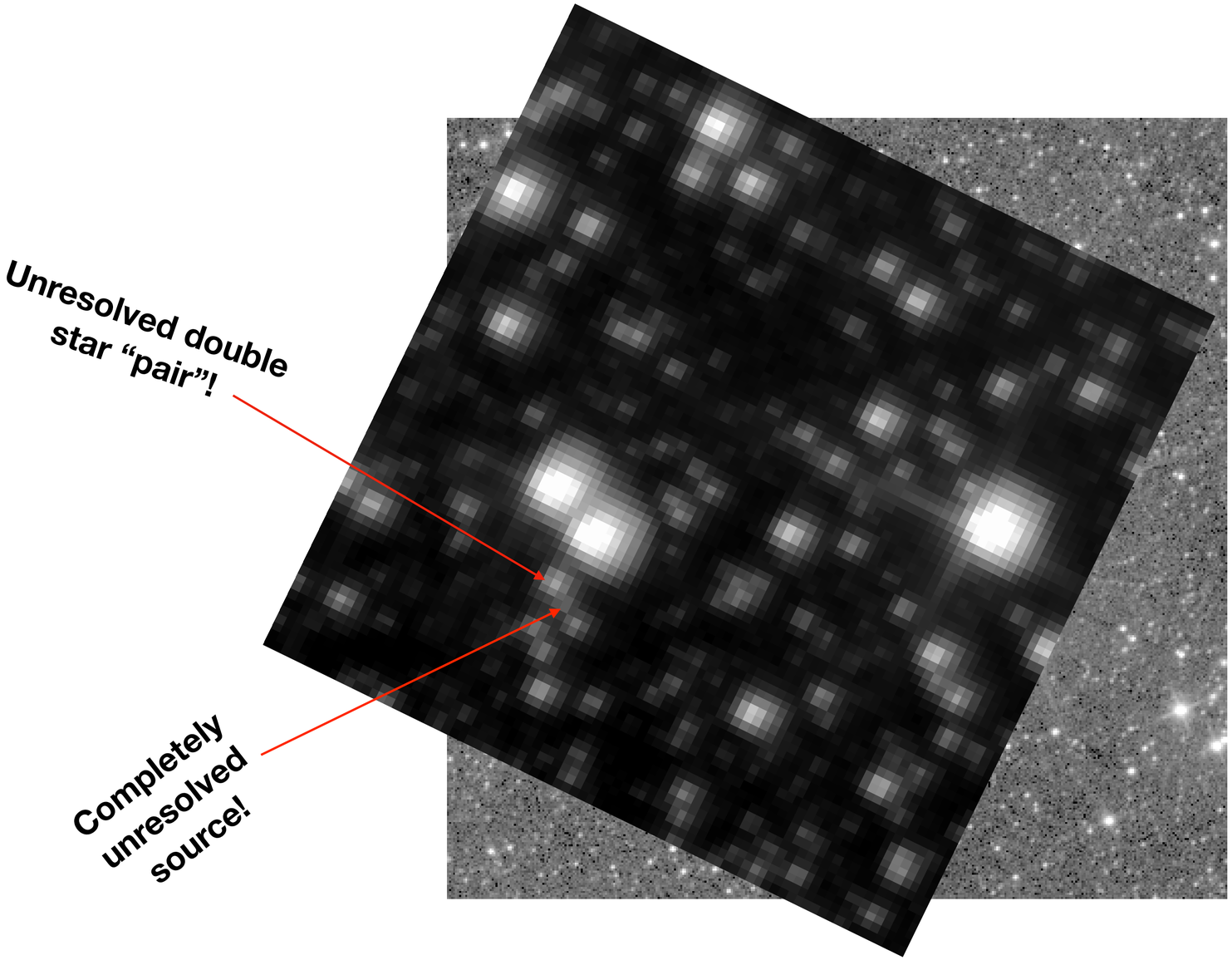
Photometric Observations



WISE - Wright et al. (2010)

WISE W1

Photometric Observations



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

TESS T

Tom J Wilson @onoddil

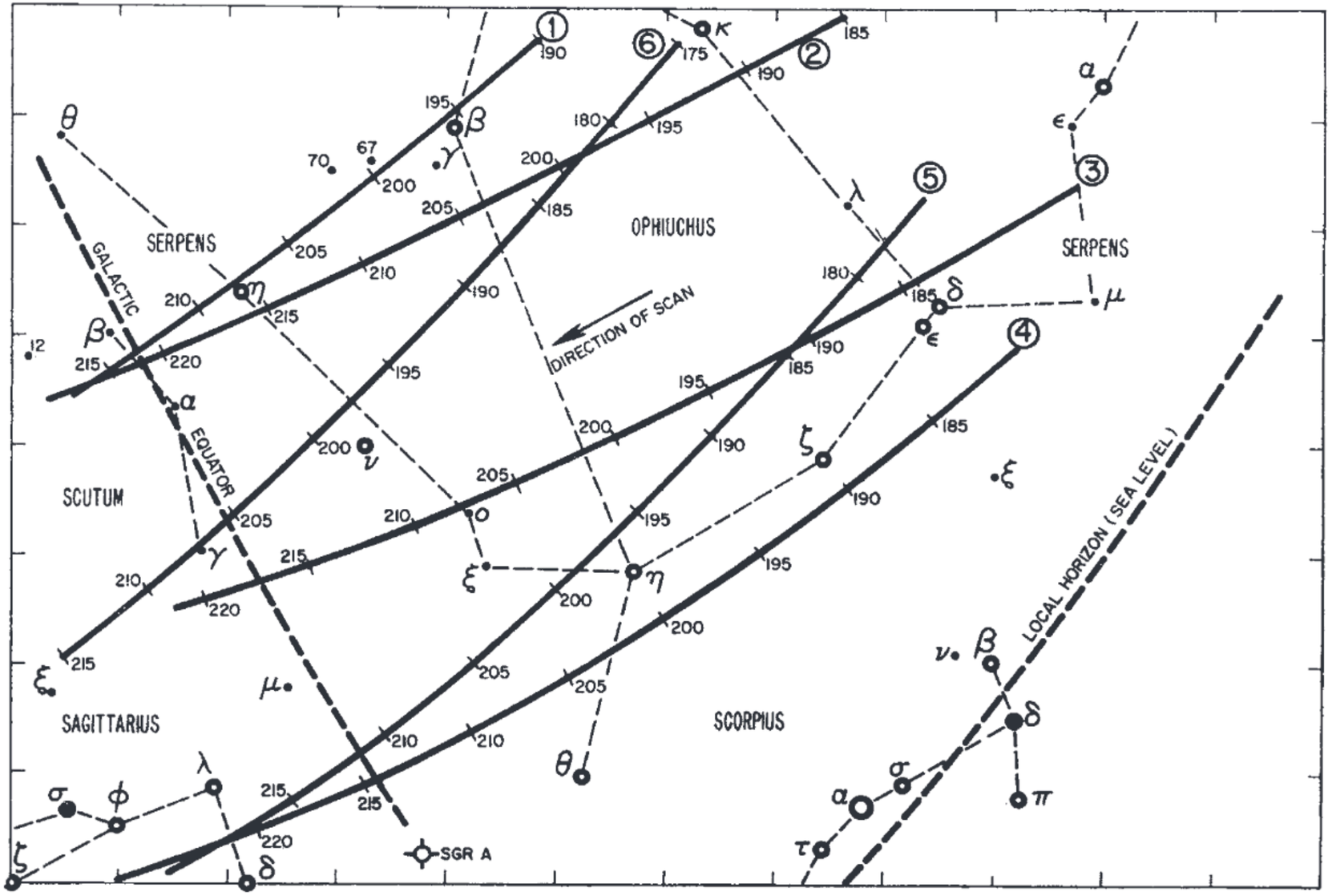
Matching Constellations

Image credit: Mouser, wikimedia

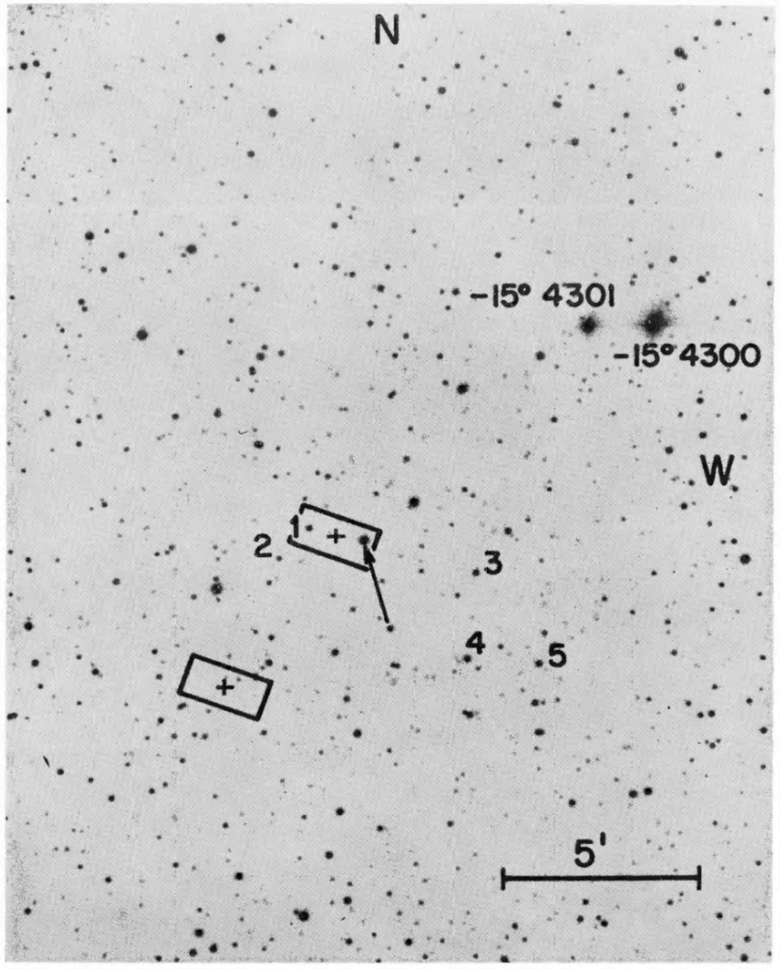
Technology Abounds

- **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



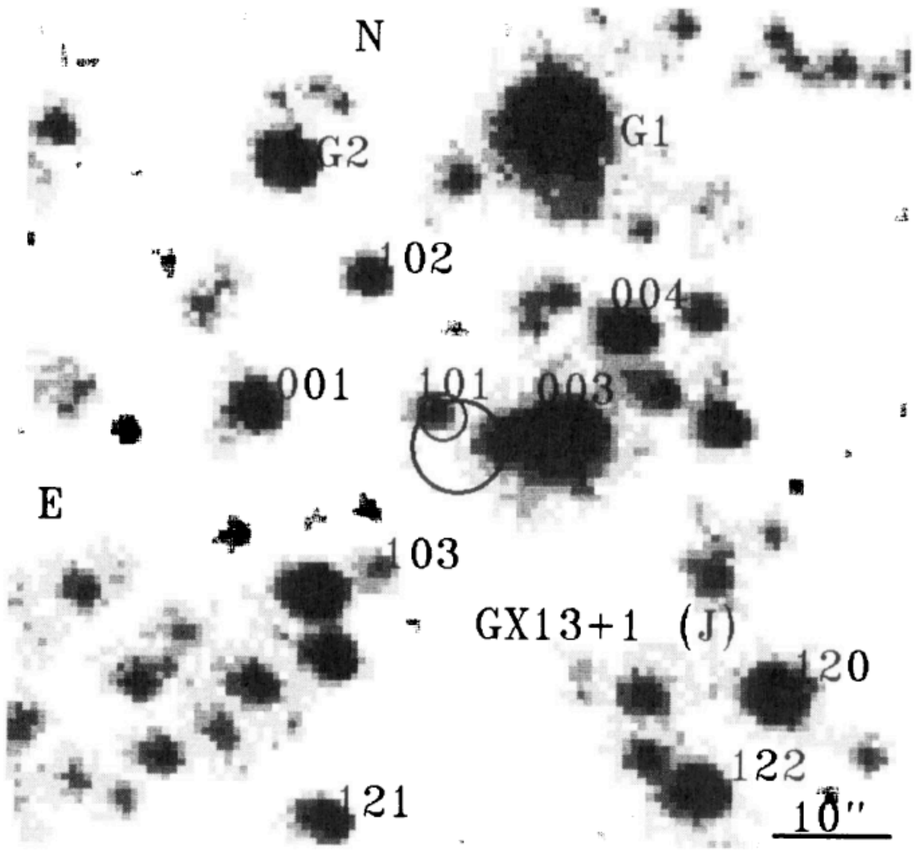
Giacconi, Gursky, & Waters (1964)



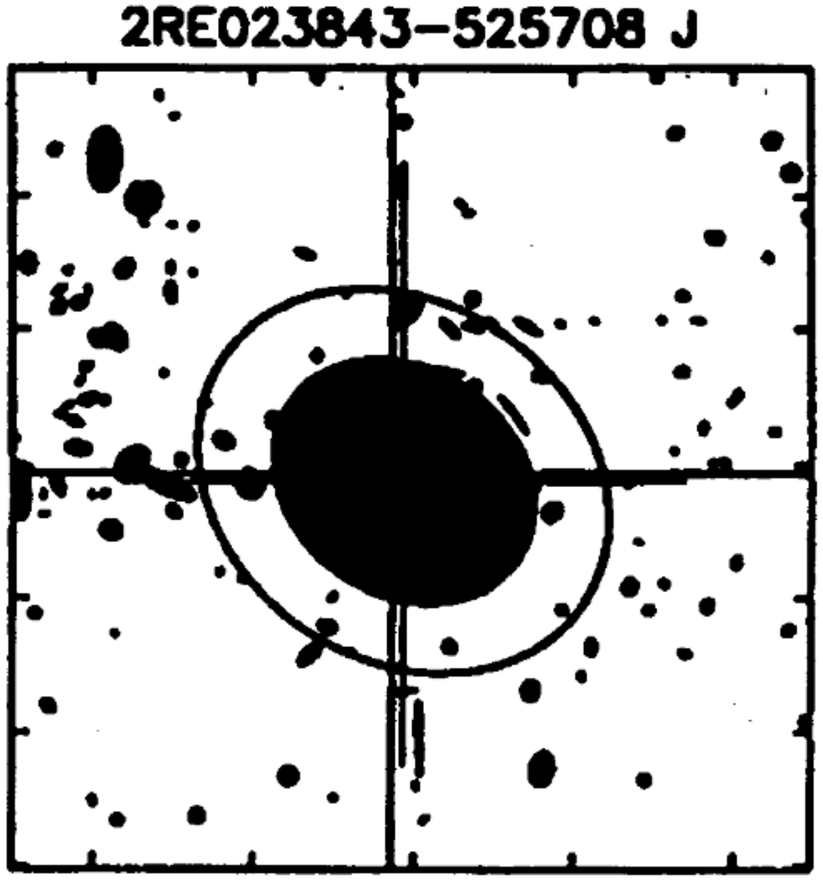
Sandage et al. (1966)

Tom J Wilson @onoddil

The Brightest Star in the Sky



Naylor, Charles, & Longmore (1991)

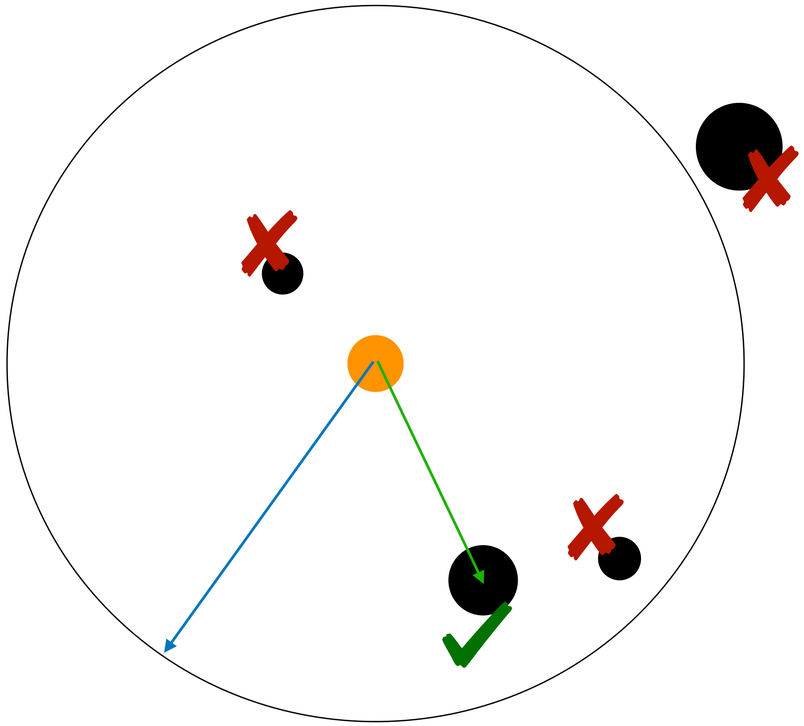


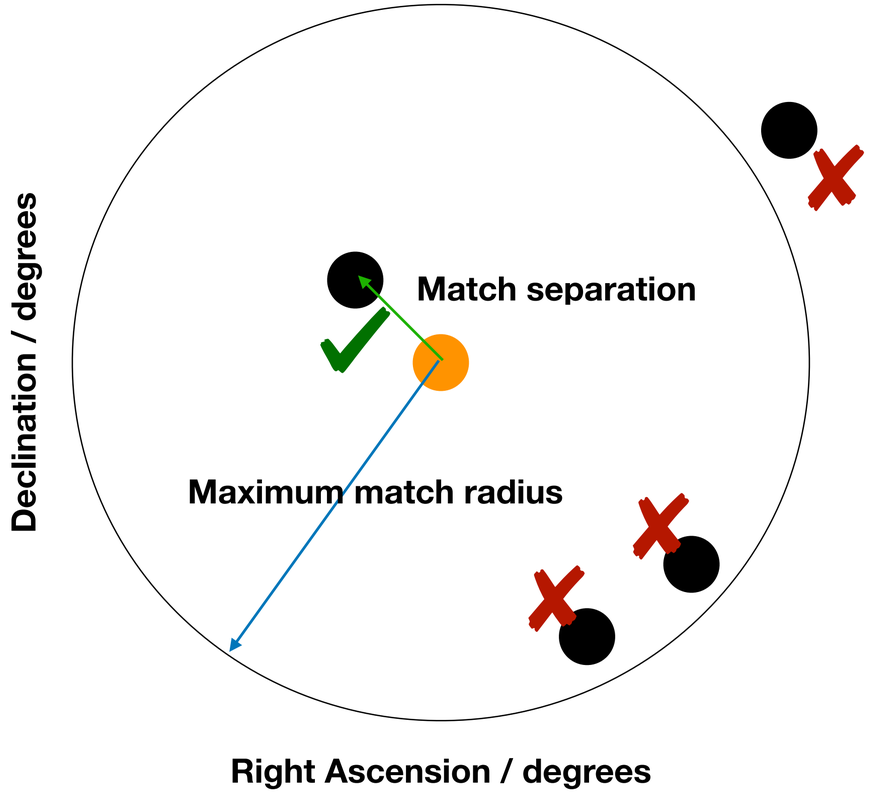
Mason et al. (1995)

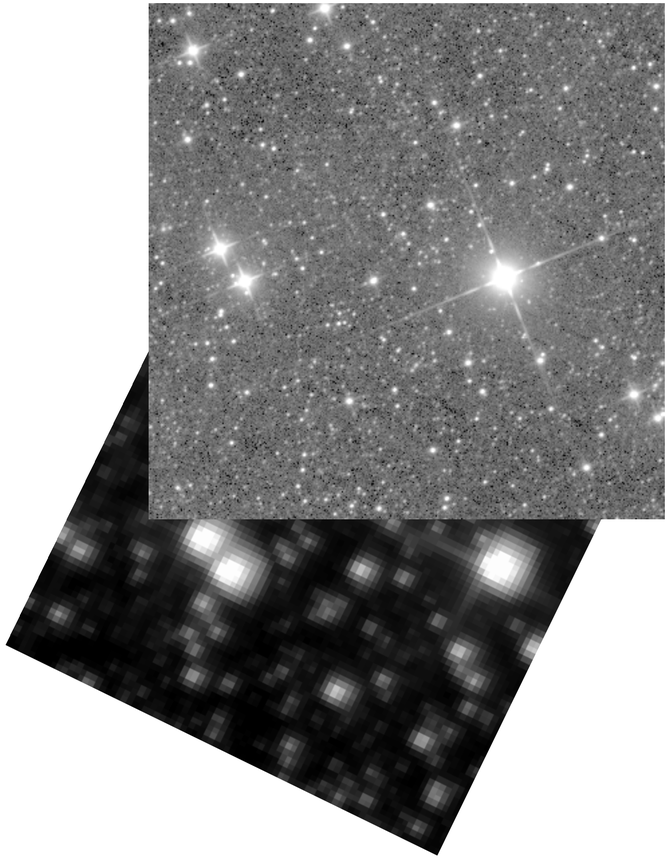
"...X-ray sources are rare events; bright optical sources are also rare events, so the observation of an X-ray source and a bright optical source in the same region of the sky is considered a non-random event"

Fotopoulou et al. (2016)

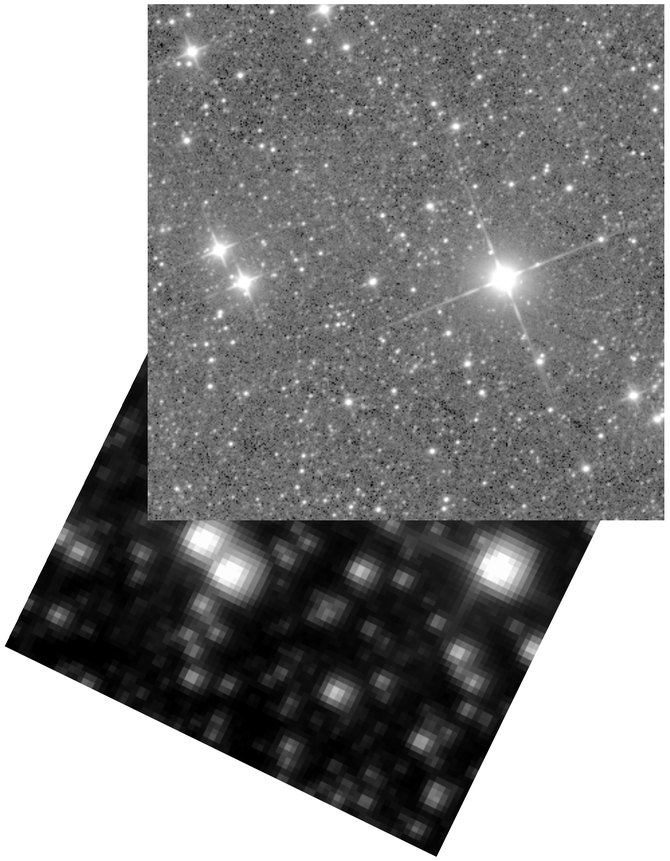
“Traditional” Cross-Matching







The Astronomy Error Function



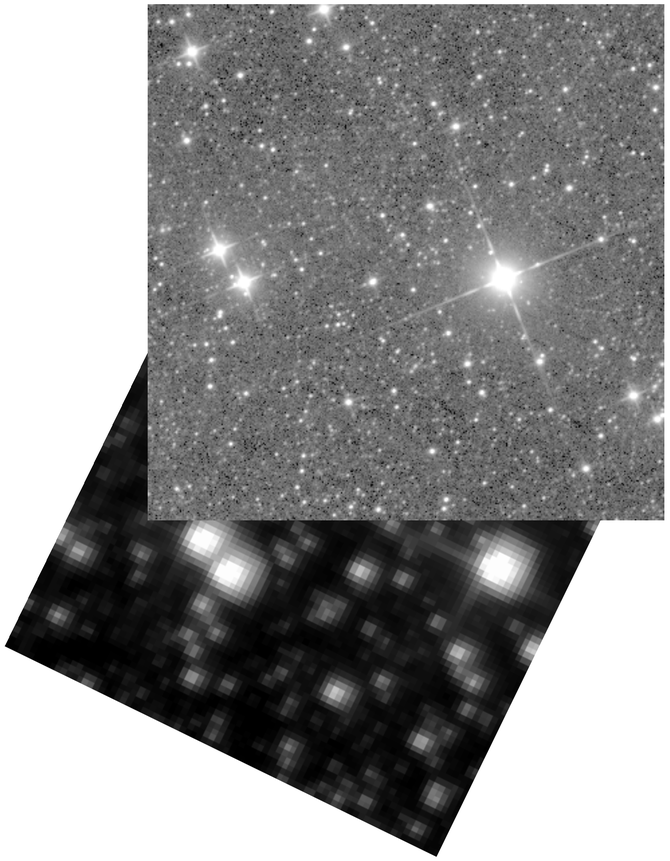
1) $p(x \text{ and } y) = p(x)p(y)$

2) $p(x)$ decreases as x increases

3) $p(x) = p(-x) \Rightarrow p(x^2)$

J. F. W. Herschel, “Quetelet on Probabilities”, 1850

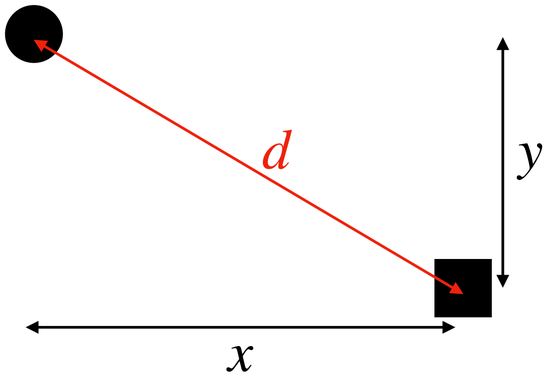
The Astronomy Error Function



1) $p(x \text{ and } y) = p(x)p(y)$

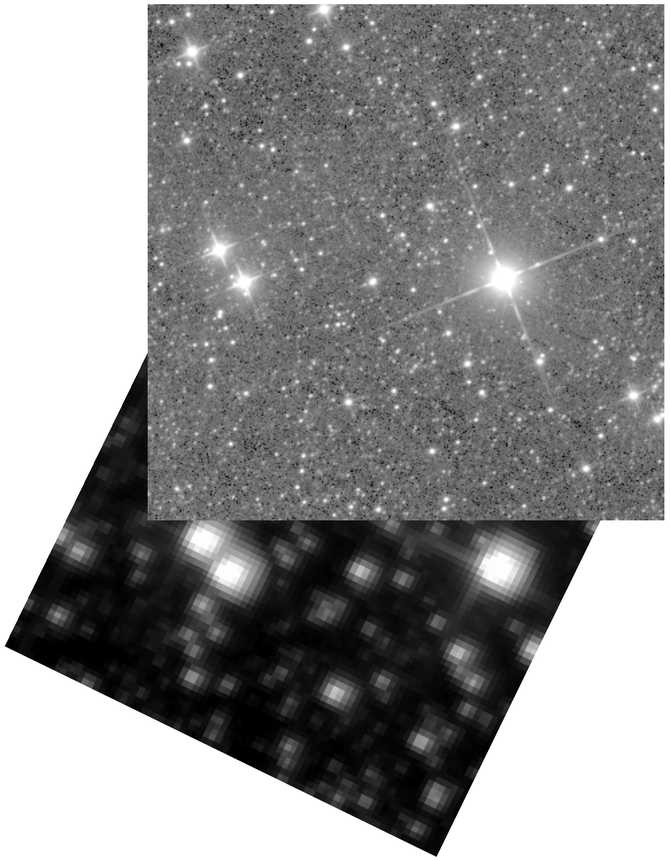
2) $p(x)$ decreases as x increases

3) $p(x) = p(-x) \Rightarrow p(x^2)$



J. F. W. Herschel, “Quetelet on Probabilities”, 1850

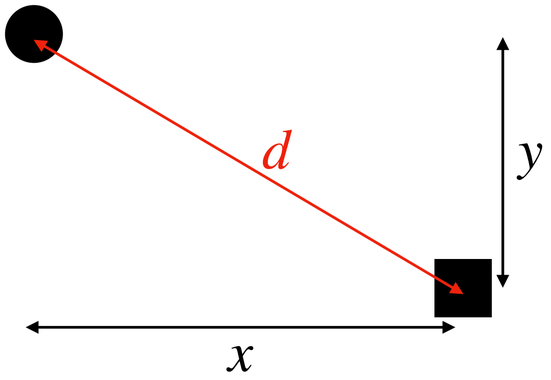
The Astronomy Error Function



1) $p(x \text{ and } y) = p(x)p(y)$

2) $p(x)$ decreases as x increases

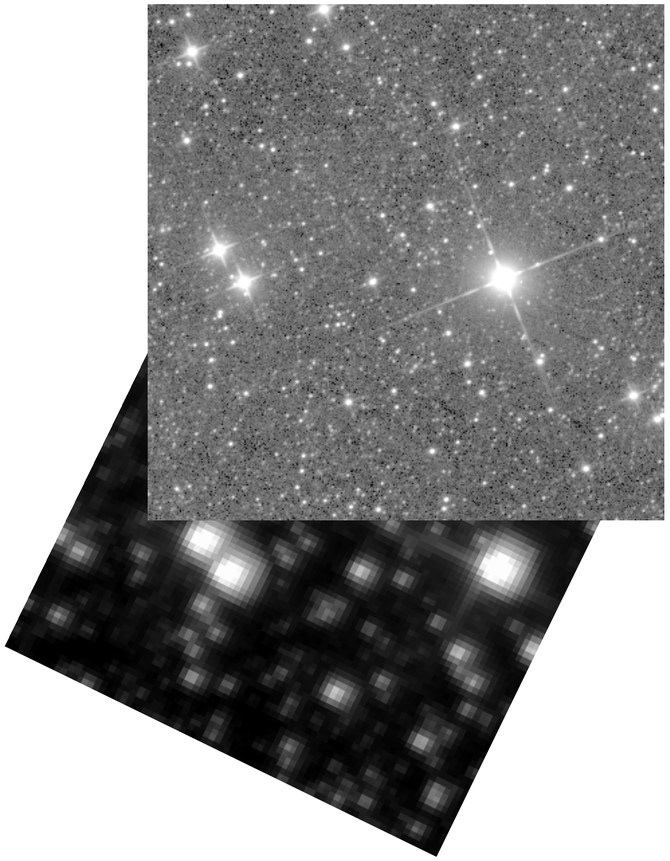
3) $p(x) = p(-x) \Rightarrow p(x^2)$



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

J. F. W. Herschel, “Quetelet on Probabilities”, 1850

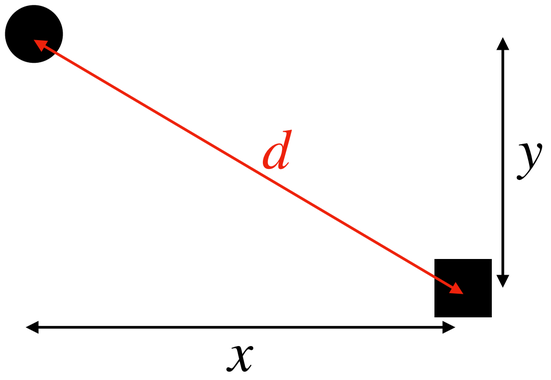
The Astronomy Error Function



1) $p(x \text{ and } y) = p(x)p(y)$

2) $p(x)$ decreases as x increases

3) $p(x) = p(-x) \Rightarrow p(x^2)$



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

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

J. F. W. Herschel, “Quetelet on Probabilities”, 1850

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)

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

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

Wolstencroft et al. (1986)

Probabilistic Cross-Matching

The Likelihood Ratio

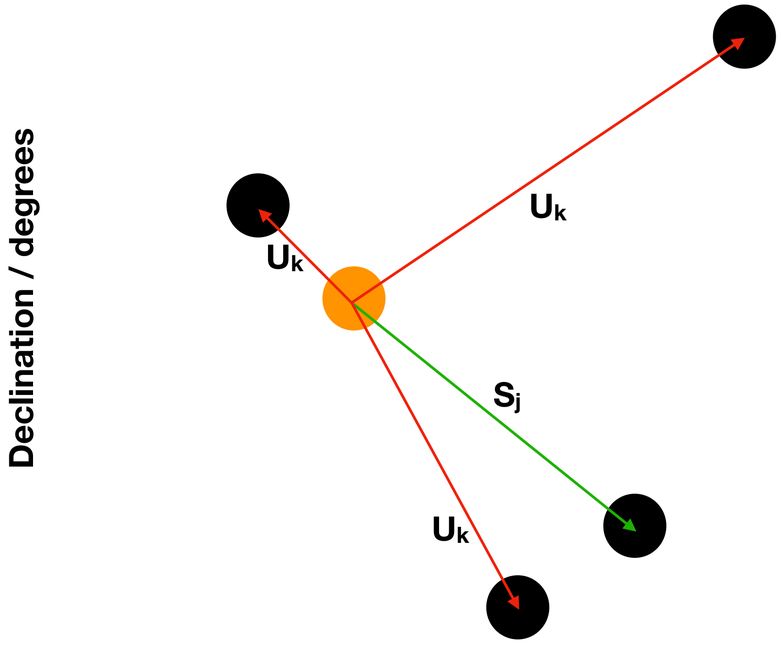
$$dp(r|id) = r \times e^{-r^2/2}\, dr.$$
$$dp(r|c) = 2\lambda r \times e^{-\lambda r^2}\, dr$$
$$LR(r) = dp(r|id)/dp(r|c) = \frac{1}{2\lambda}\exp\left\{\frac{r^2}{2}(2\lambda - 1)\right\}$$

de Ruiter, Willis, & Arp (1977)

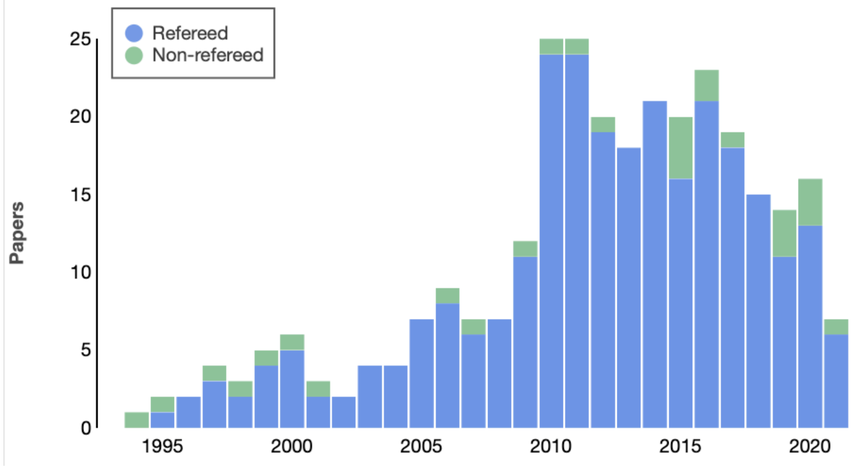
$$dp_{\mathrm{id}} = Qr\exp\left(\frac{-r^2}{2}\right)dr. \quad dp_{\mathrm{uo}} = 2\lambda r\, dr$$
$$LR(r) = \frac{dp_{\mathrm{id}}}{dp_{\mathrm{uo}}} = \frac{Q\exp\left(-r^2/2\right)}{2\lambda}$$

Wolstencroft et al. (1986)

The "Reliability" — Sutherland & Saunders (1992)



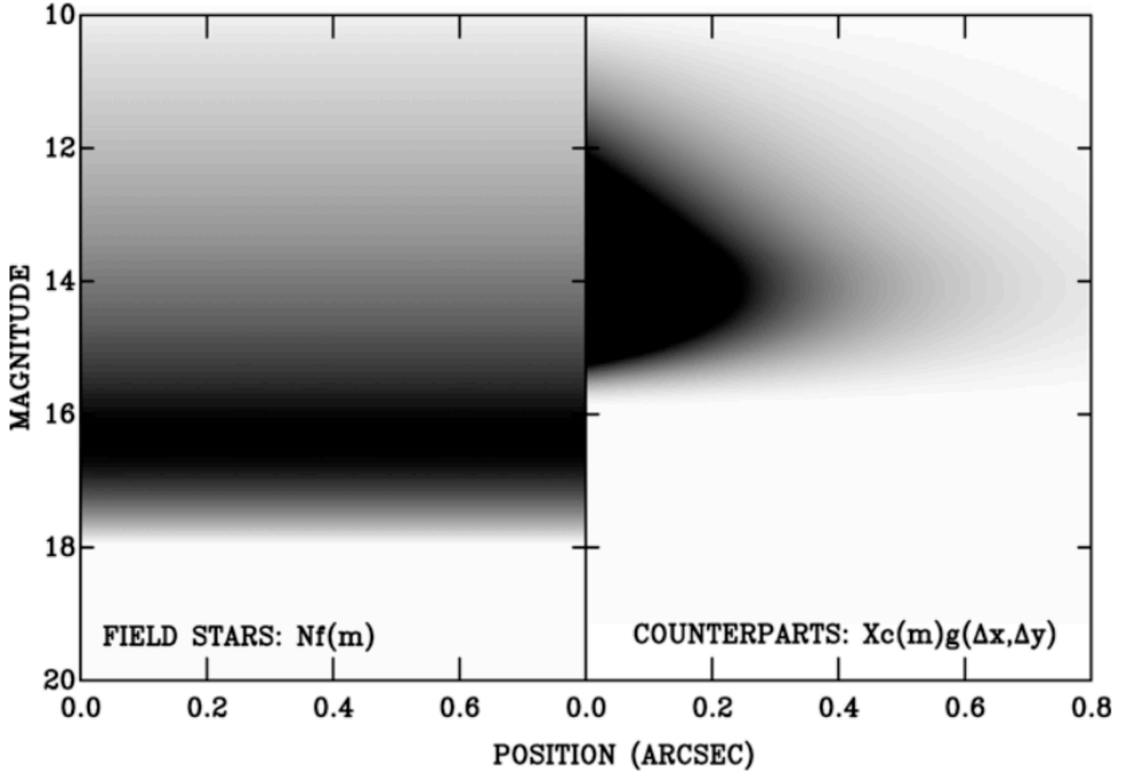
$$R_j = \frac{\Pr\left[S_j \cap \left(\bigcap_{k\neq j} U_k\right) \cap \left(\bigcap_{k'} E_{k'}\right)\right]}{\sum_i \Pr\left[S_i \cap \left(\bigcap_{k\neq i} U_k\right) \cap \left(\bigcap_{k'} E_{k'}\right)\right] + \Pr\left[(m_S > m_{\mathrm{lim}}) \cap \left(\bigcap_{k} U_k\right) \cap \left(\bigcap_{k'}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

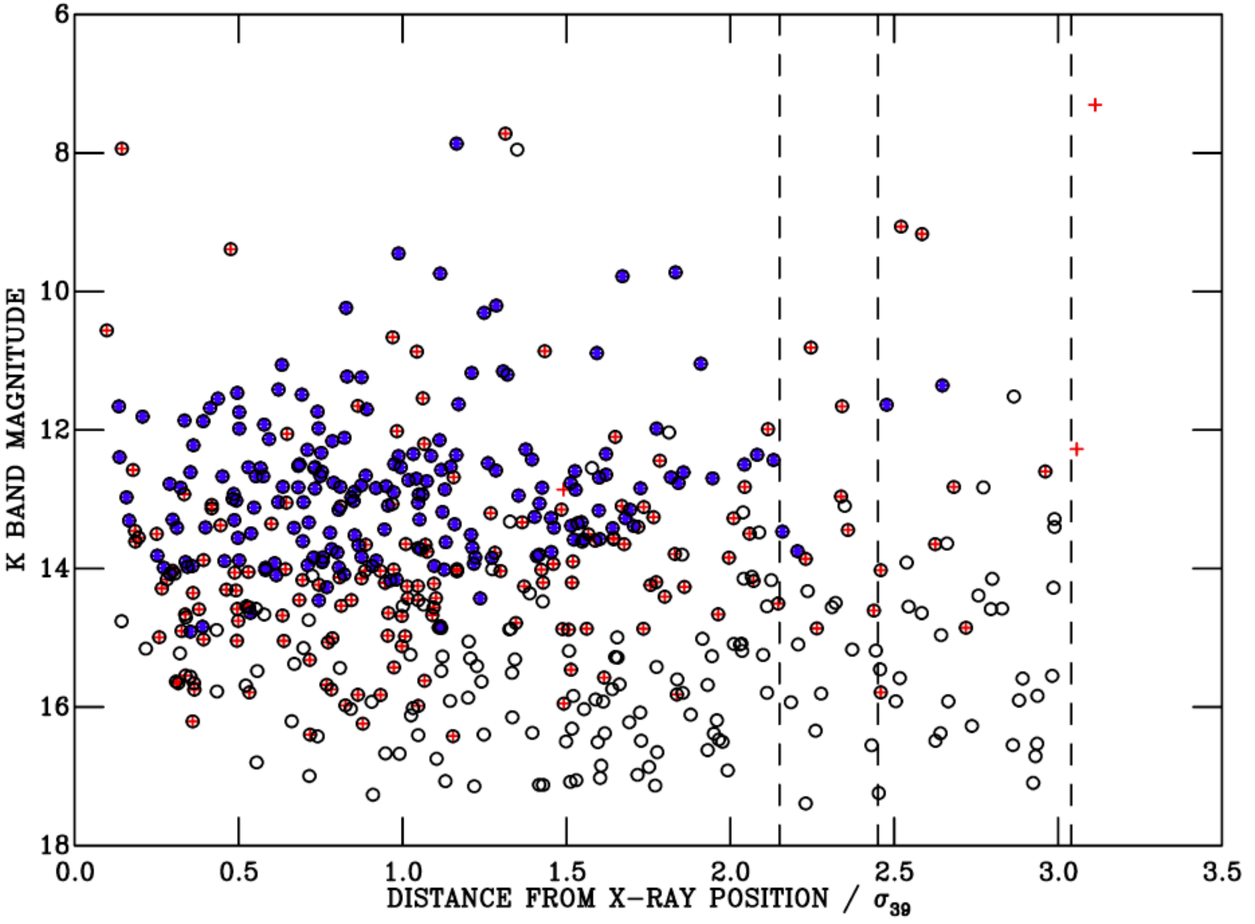


$$L = \frac{q(m,c)\, f(x,y)}{n(m,c)}$$

Probabilistic Cross-Matching

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





Naylor, Broos, & Feigelson (2013)

Tom J Wilson @onoddil

Probabilistic Cross-Matching

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

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

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

Budavári & Szalay (2008)

Includes SED model fitting to all sources

(Apologies for the lack of nice figures in this paper!)

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: positional errors of sources are Gaussian!

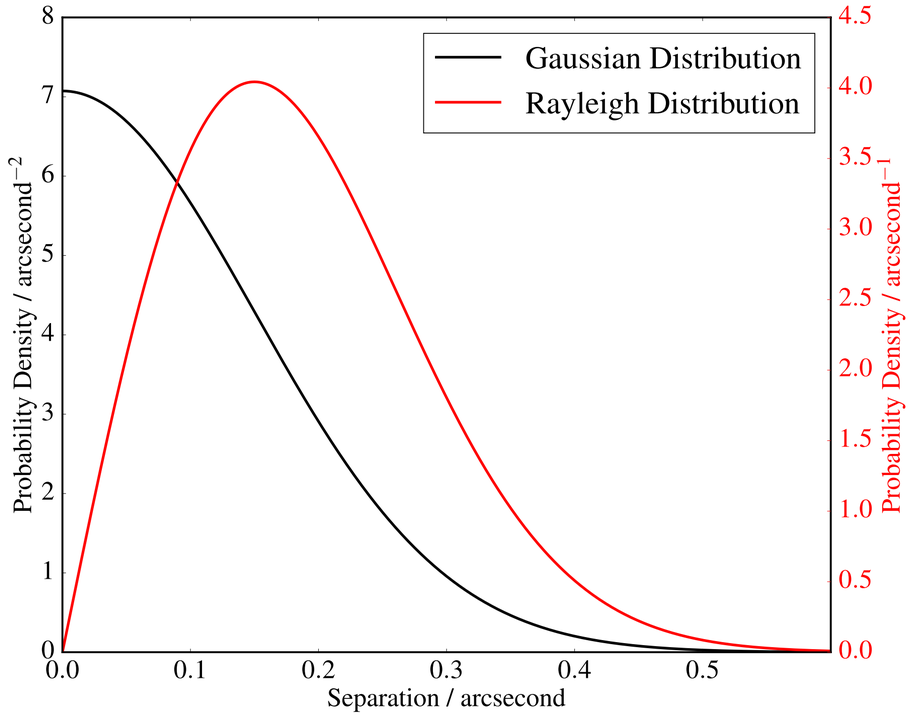
$$dp(\mathbf{r}|id) = \mathbf{r} \times e^{-\mathbf{r}^2/2}\, dr. \quad P(i) = \frac{\frac{Xc(m_i)\, g(\Delta x_i, \Delta y_i)}{Nf(m_i)}}{1 - X + \sum_j \frac{Xc(m_j)\, g(\Delta x_j, \Delta y_j)}{Nf(m_j)}} \quad p(D|H) = \int p(\boldsymbol{m}|H) \prod_{i=1}^{n} \boldsymbol{p_i}(\boldsymbol{x_i}|\boldsymbol{m},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

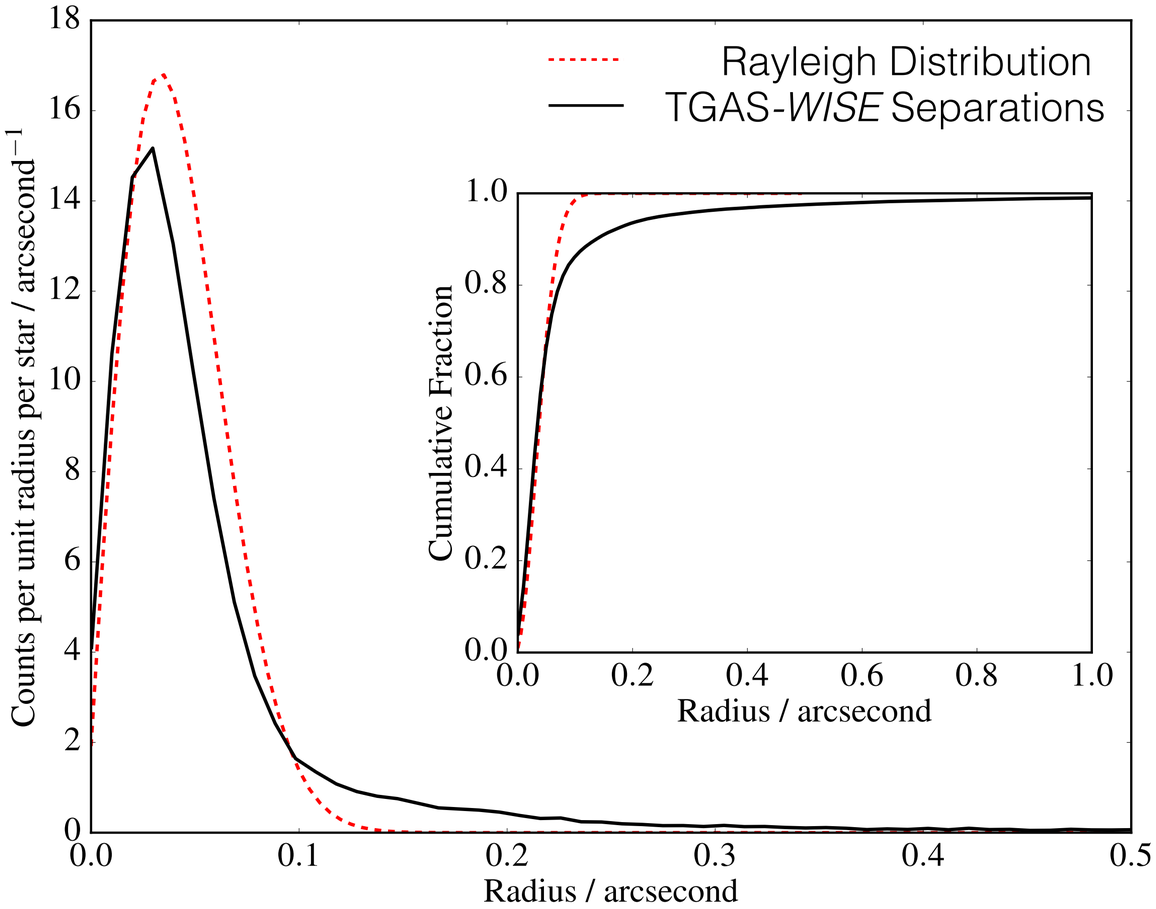
The *Astrometric Uncertainty* Function

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

$$\downarrow$$

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





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

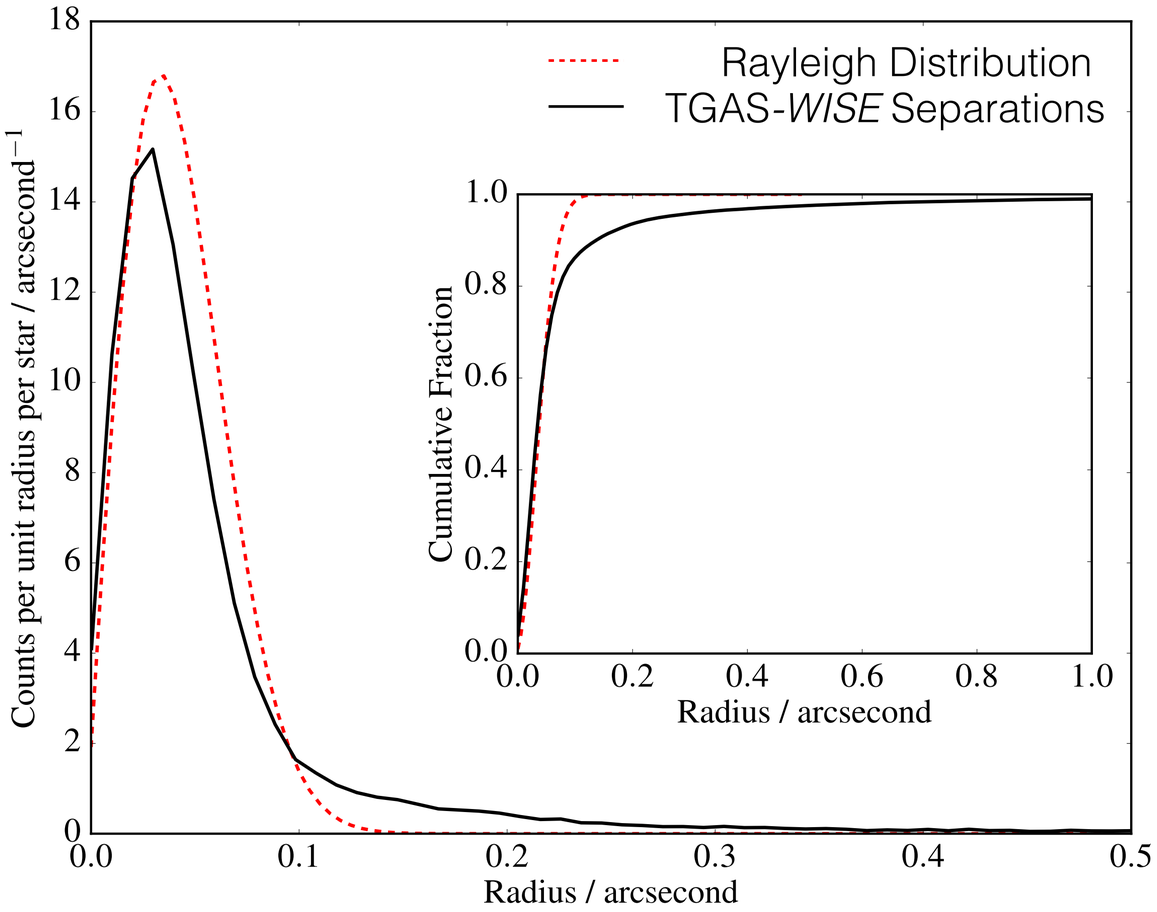
TGAS - Michalik, Lindegren, & Hobbs (2015)
Gaia - Gaia Collaboration, Brown A. G. A., et al. (2016)

Tom J Wilson @onoddil

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?

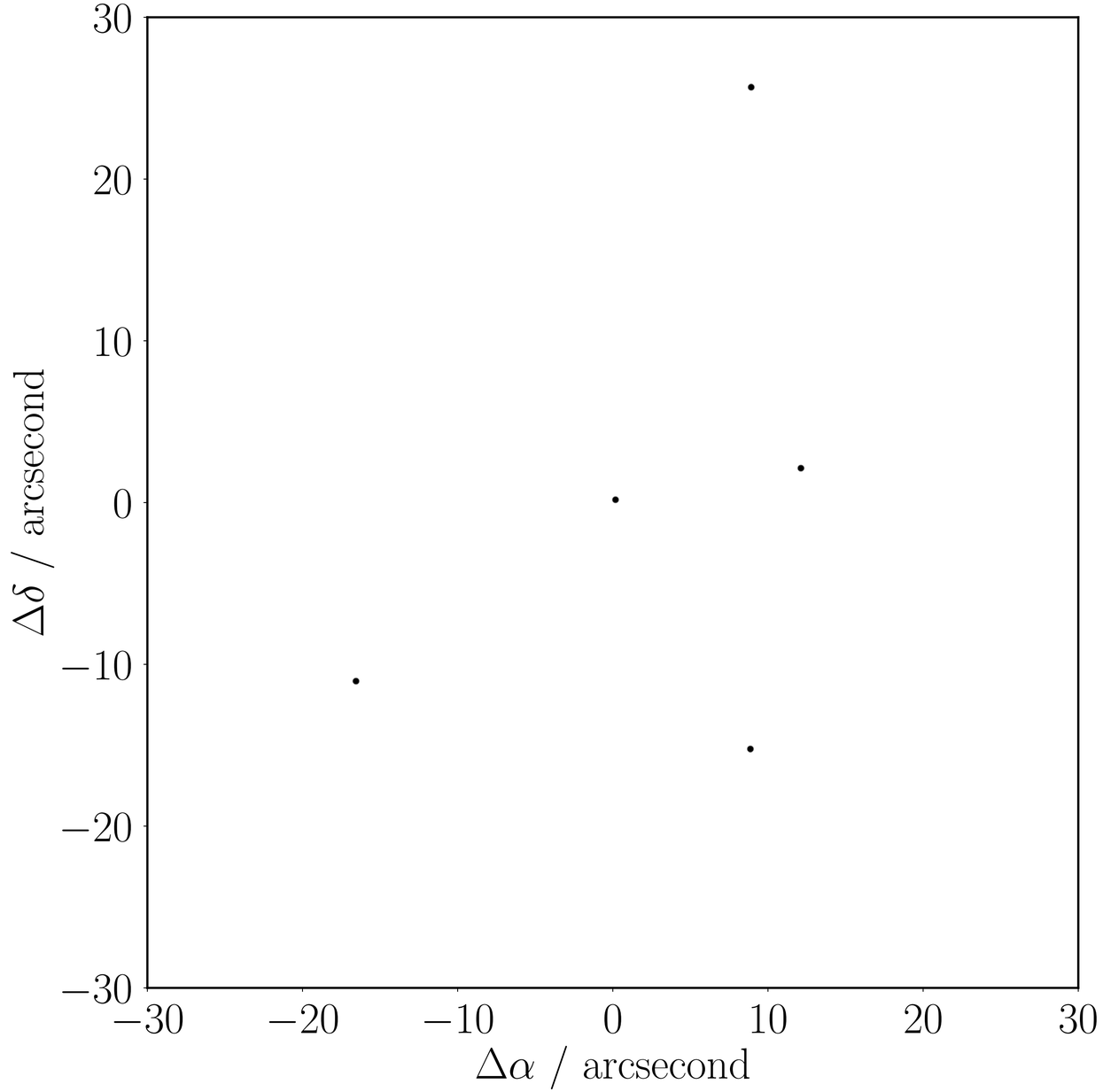


Wilson & Naylor (2017)
WISE - Wright et al. (2010)

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

Tom J Wilson @onoddil

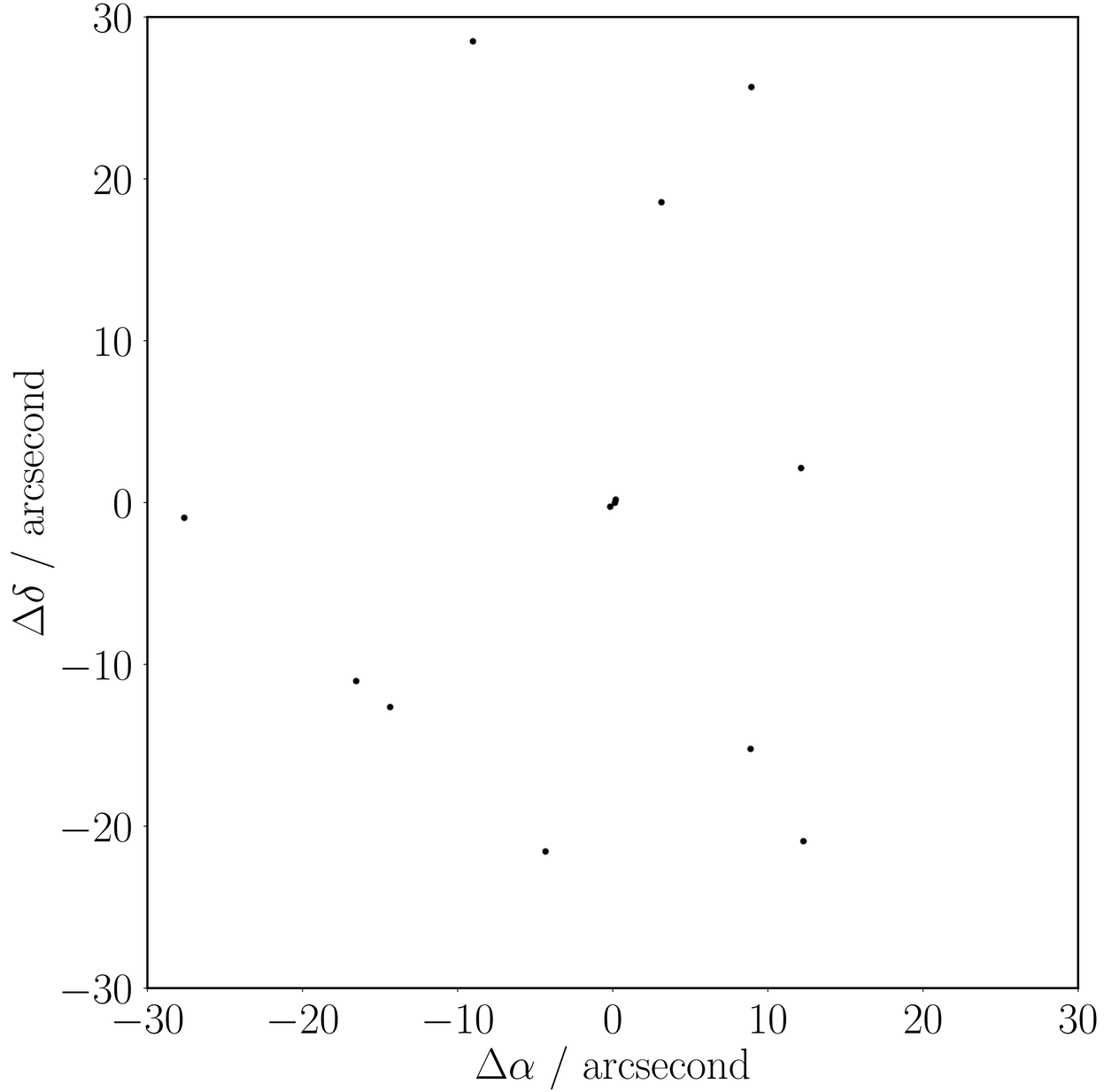
The AUF: Crowding



Wilson & Naylor (2017)

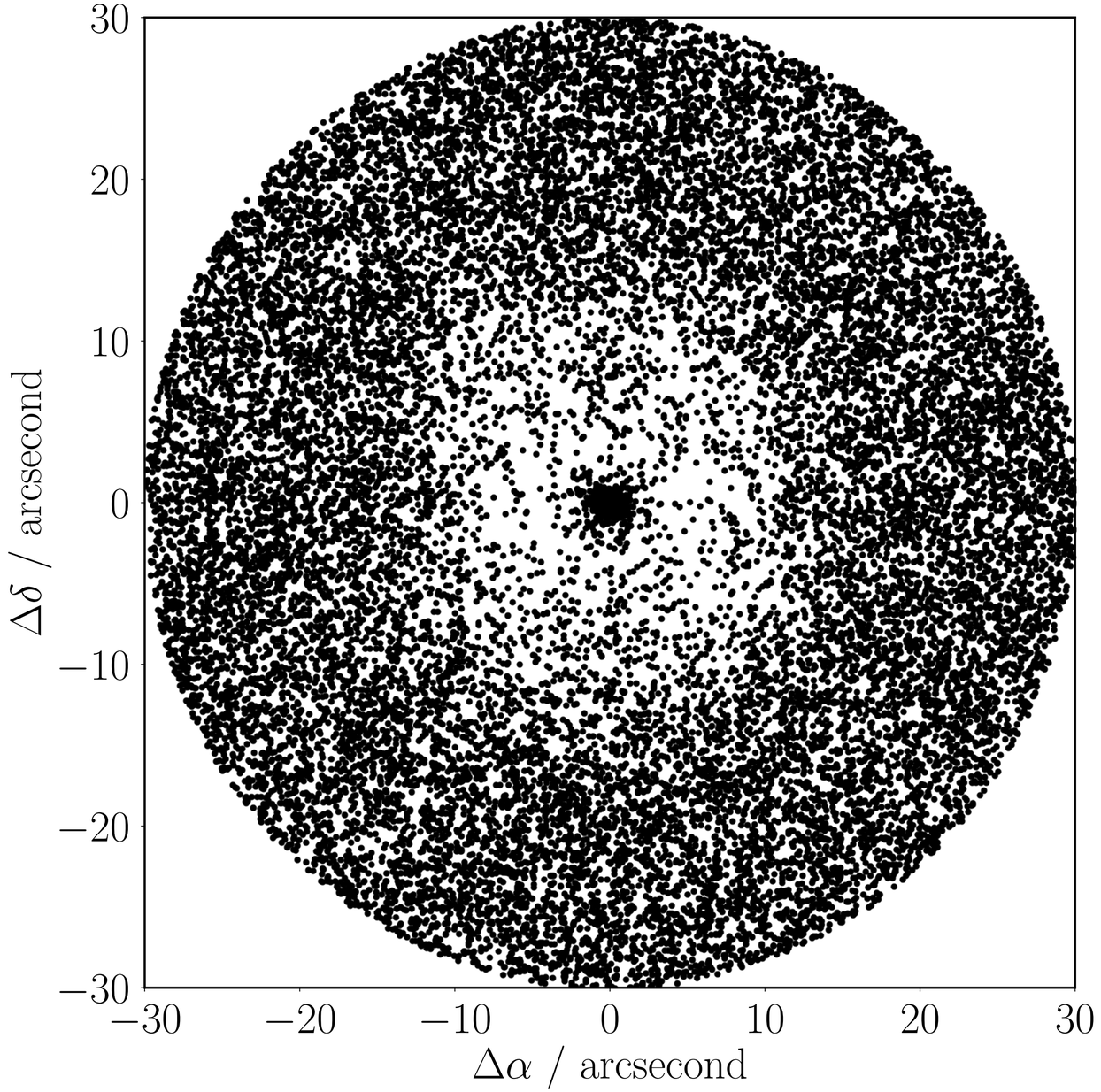
Tom J Wilson @onoddil

The AUF: Crowding



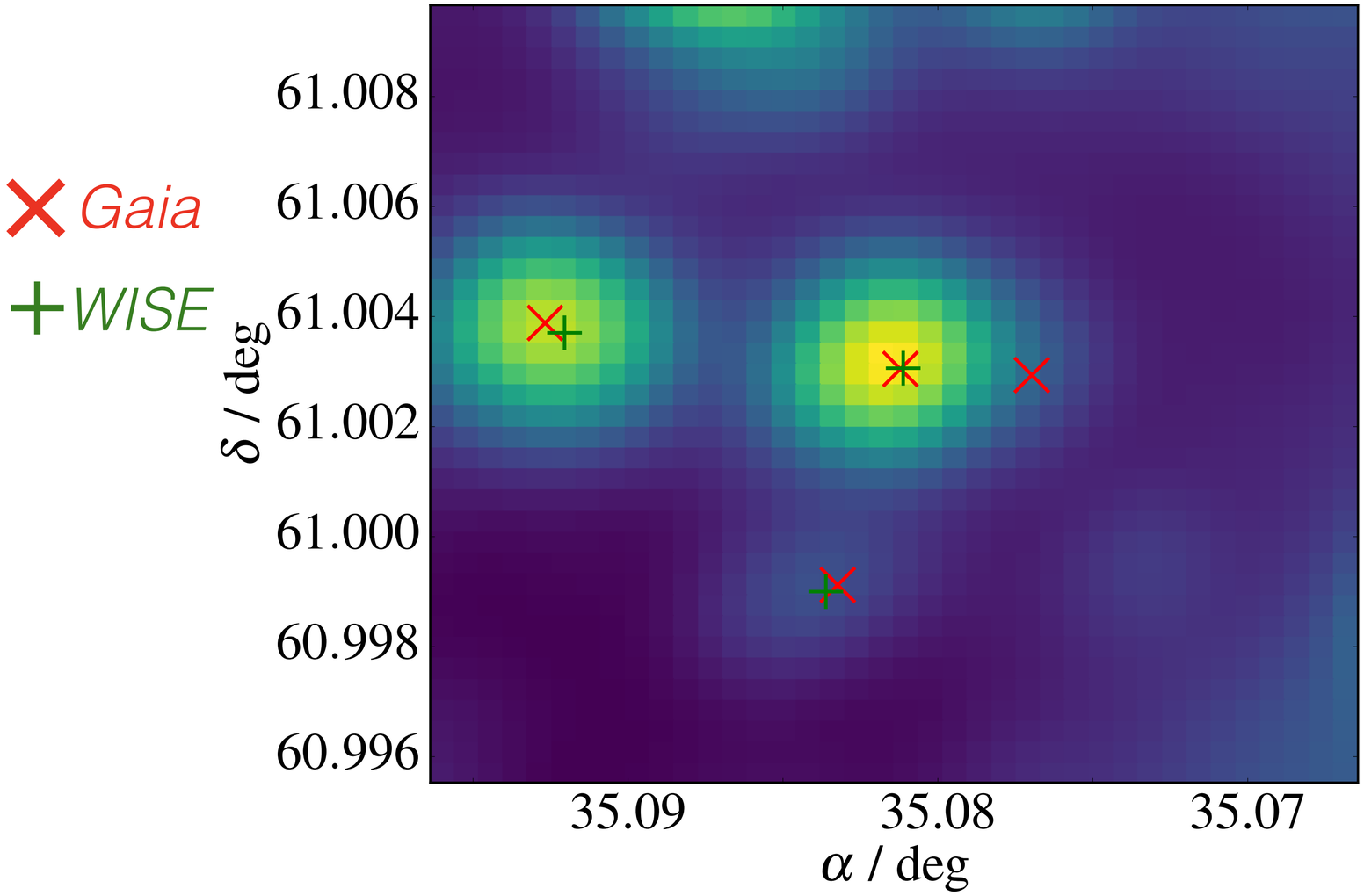
Wilson & Naylor (2017)

The AUF: Crowding



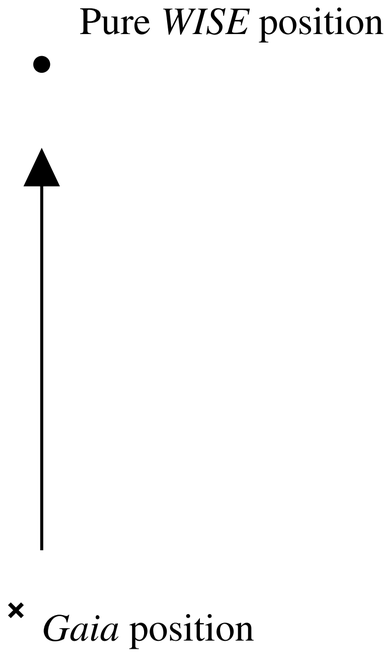
Wilson & Naylor (2017)

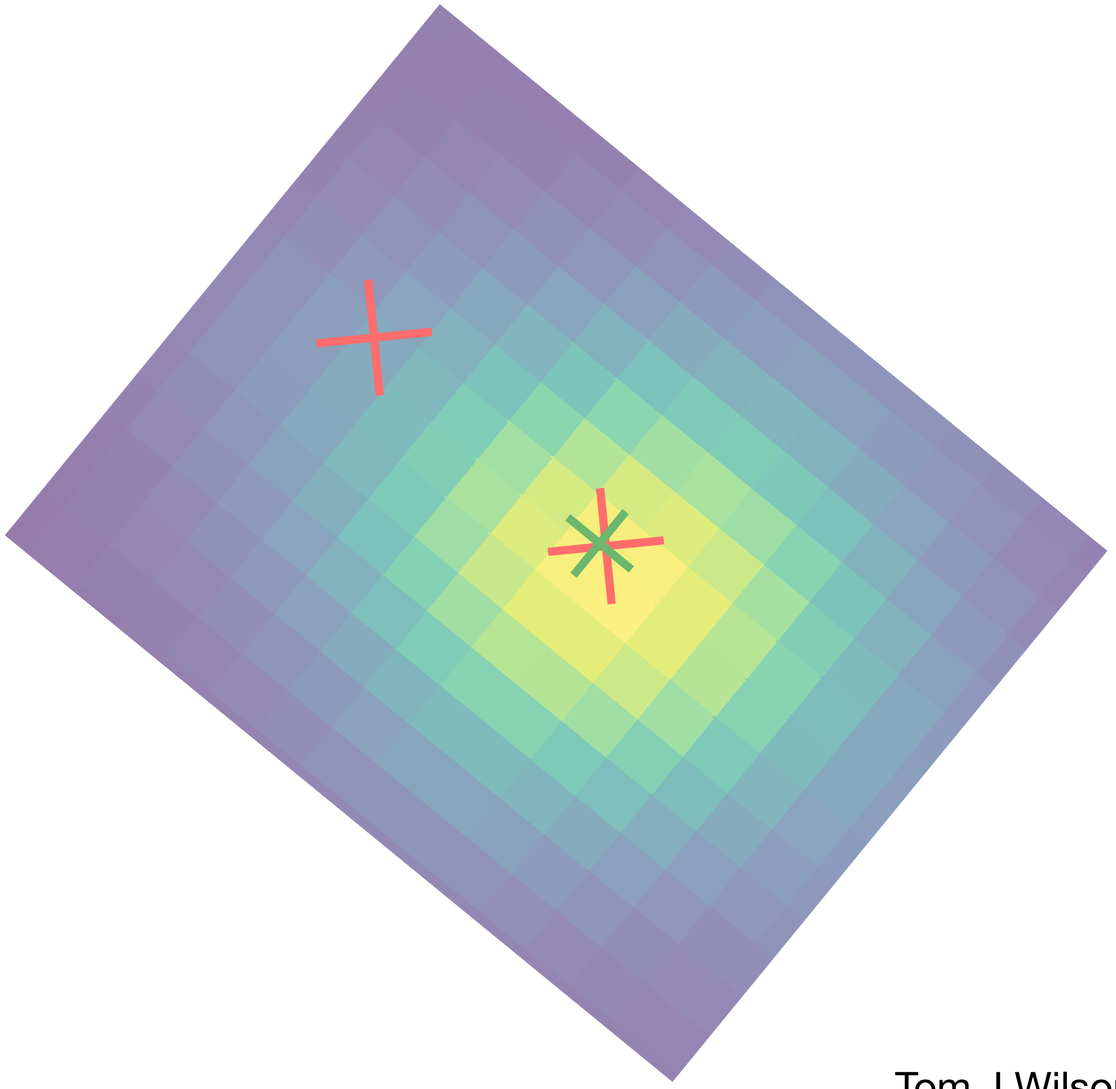
Resolving *Gaia*-*WISE* Blends



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

The AUF: Perturbation



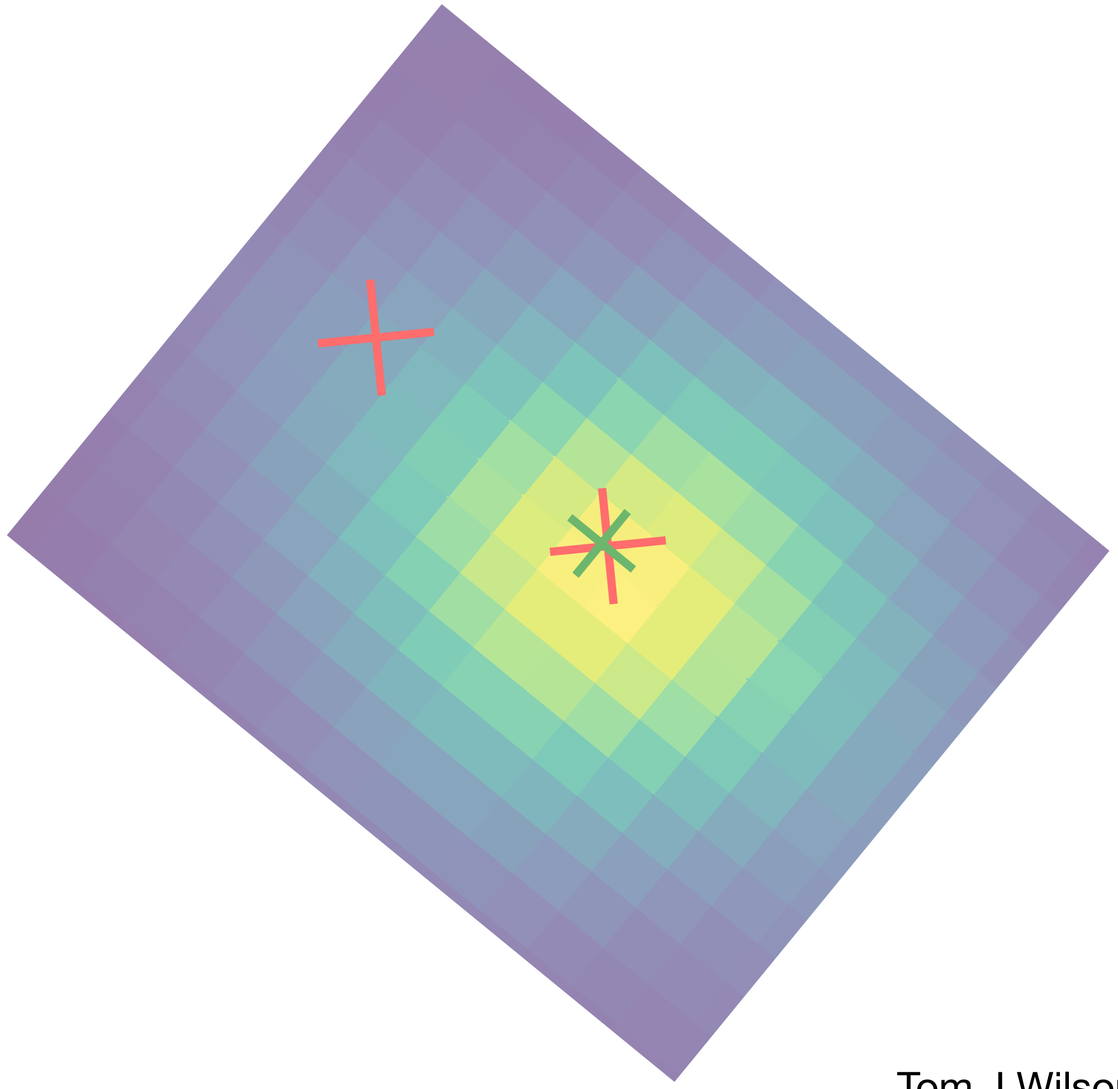


Wilson & Naylor (2017)
Wilson & Naylor (2018b)
WISE - Wright et al. (2010)
Gaia DR2 - Gaia Collaboration, Brown A. G. A., et al. (2018)

The AUF: Perturbation

To *WISE* contaminant

Pure *WISE* position

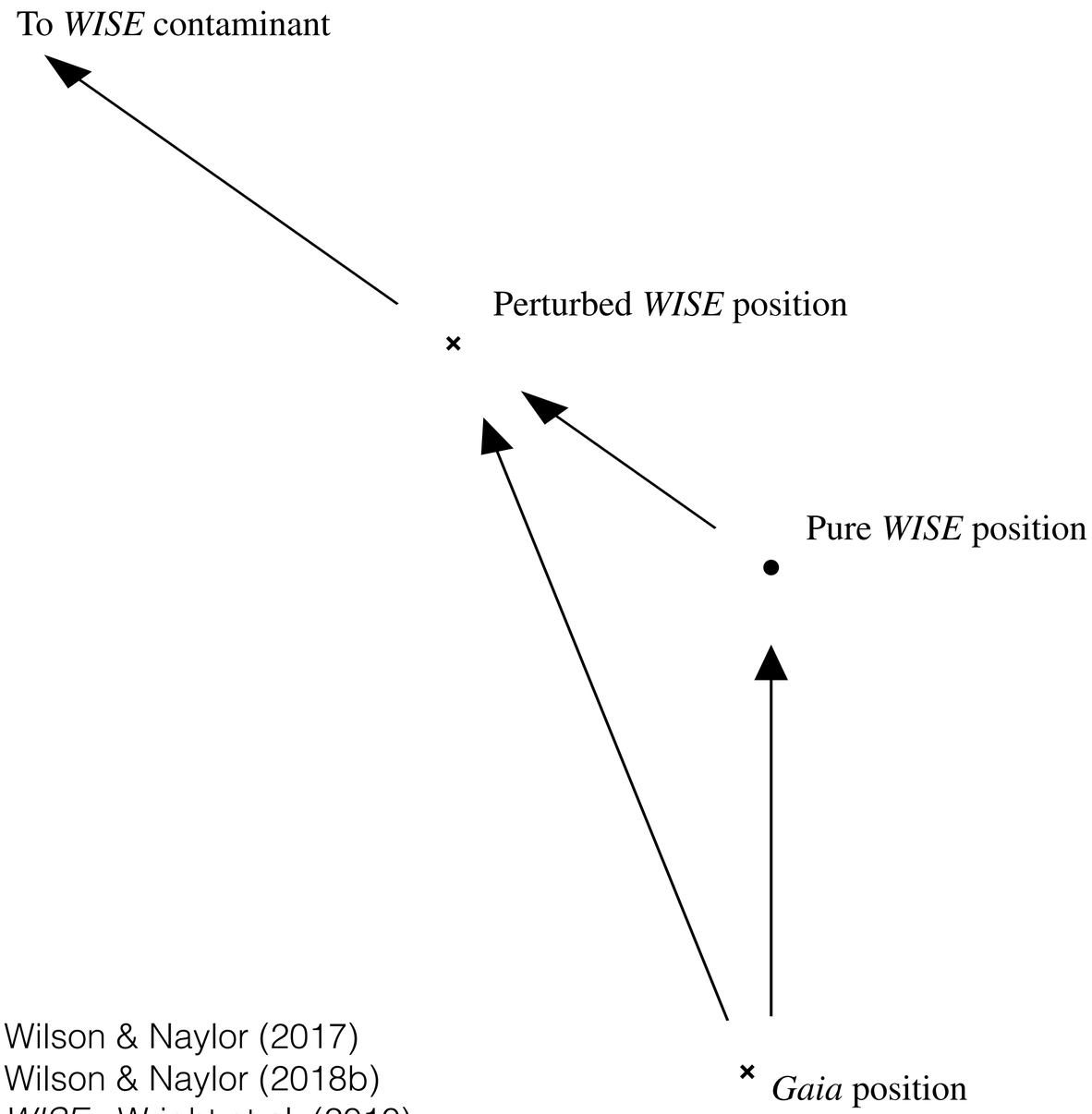


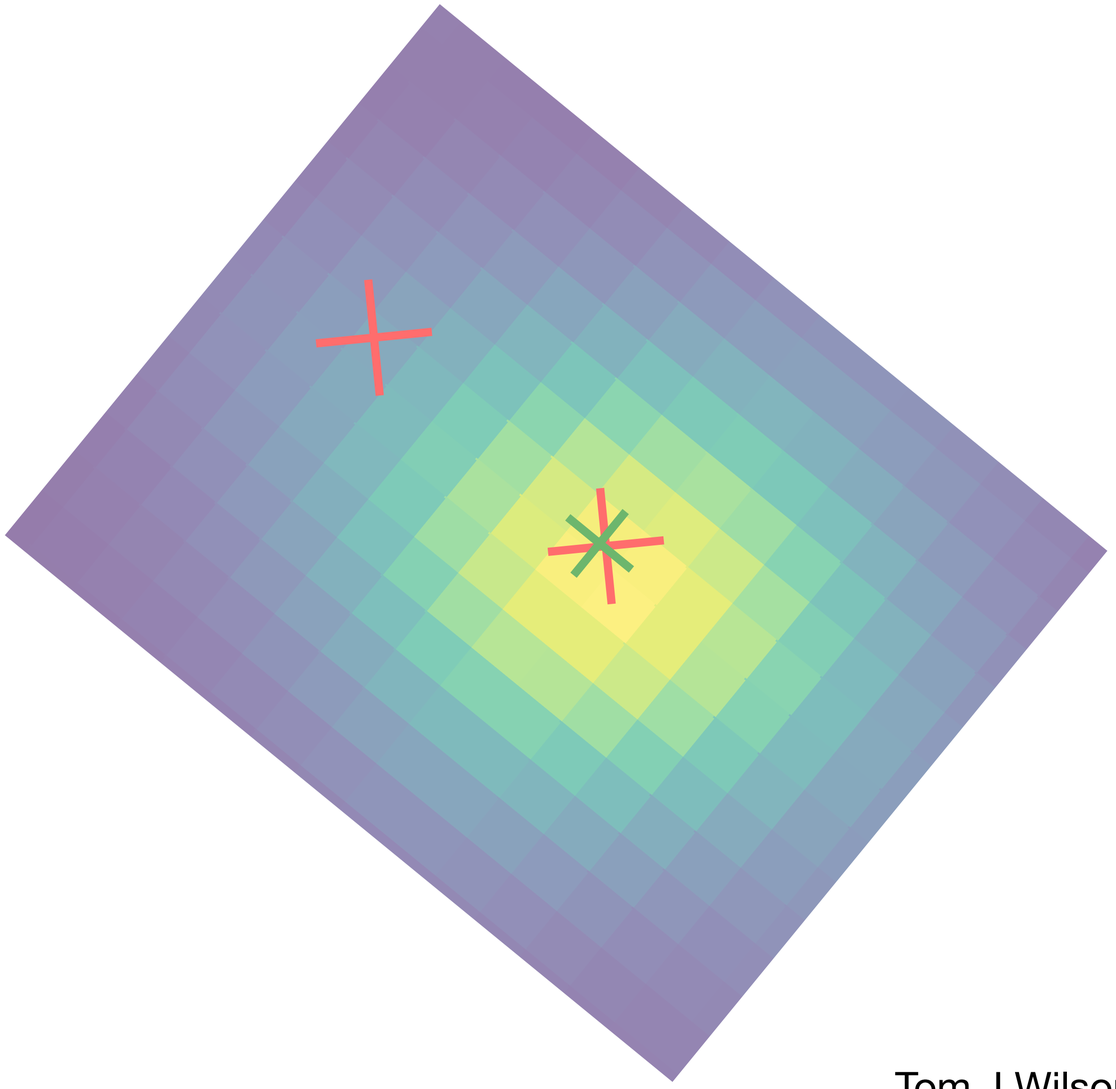
Gaia position

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

Tom J Wilson @onoddil

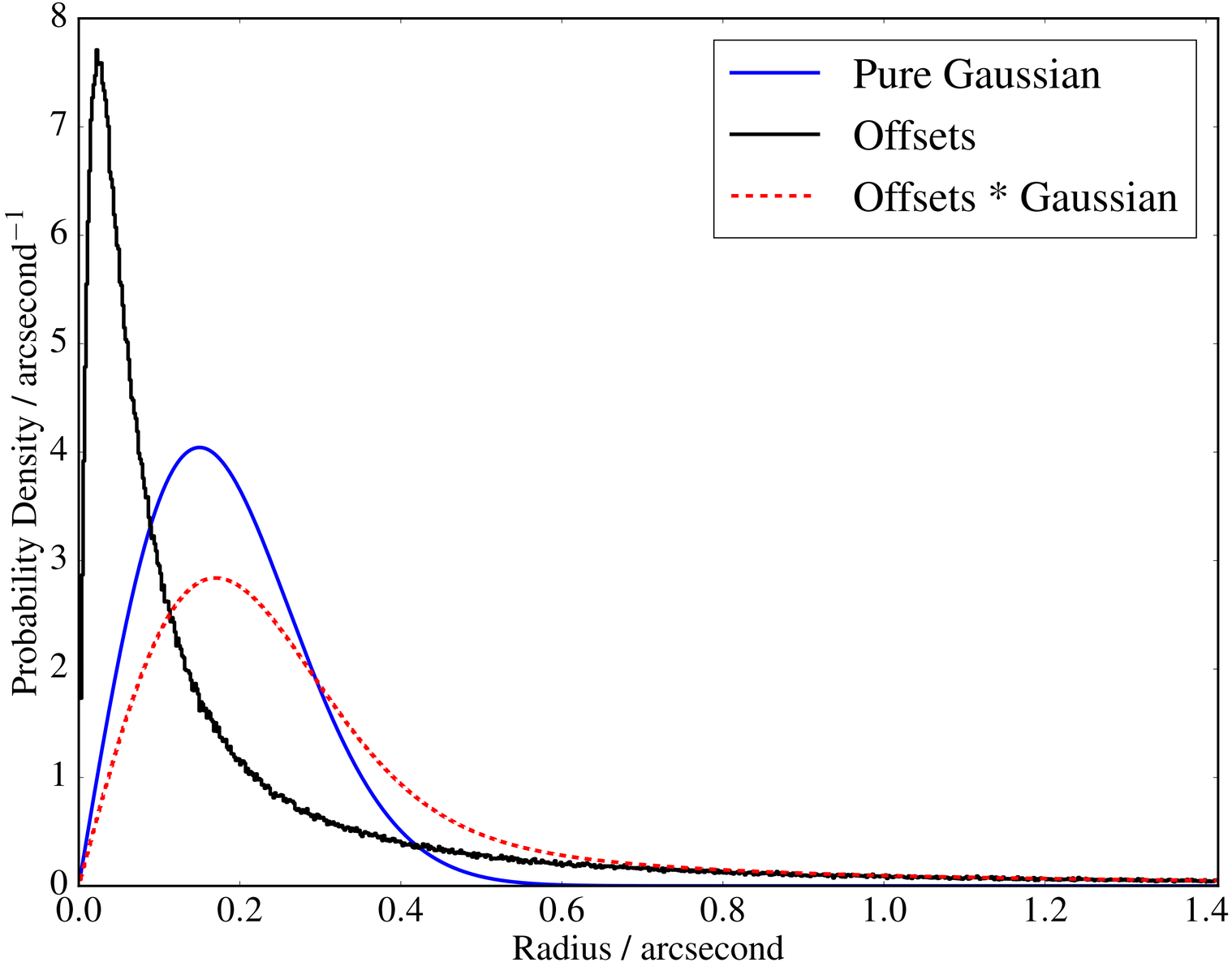
The AUF: Perturbation





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



Wilson & Naylor (2018b)

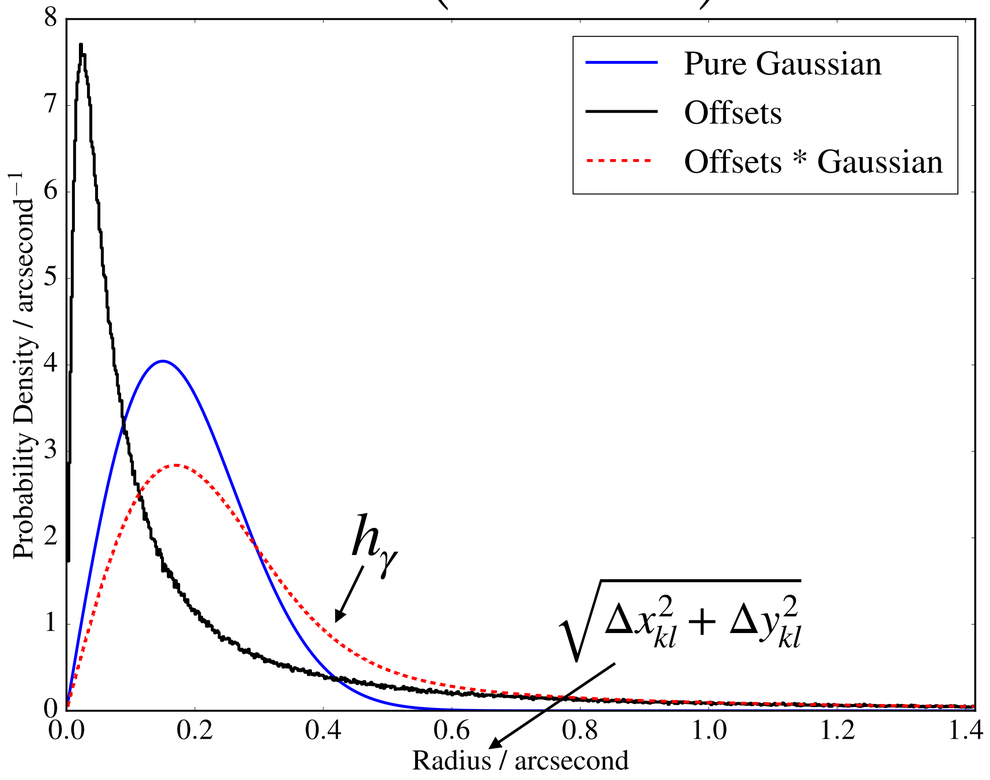
Separation Likelihood Function

$$g(x_k, y_k, x_l, y_l) = \int_{-\infty}^{+\infty}\!\!\!\iint h_\gamma(x_0 - x_k, y_0 - y_k) h_\phi(x_l - x_0, y_l - y_0) p(x_0, y_0)\,\mathrm{d}x_0\,\mathrm{d}y_0$$

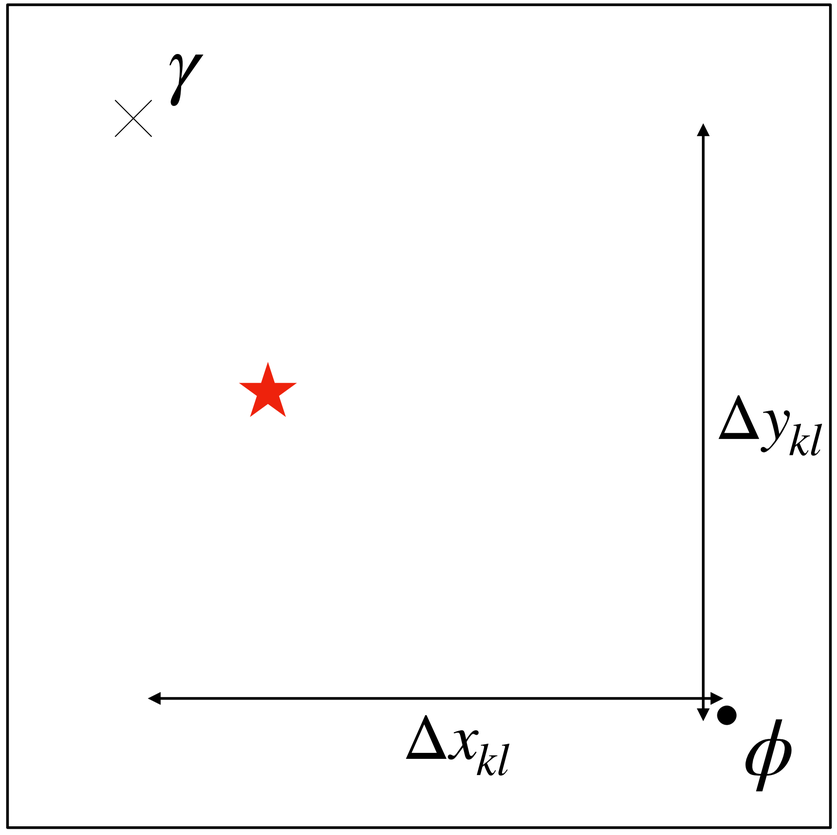
$$= N_\mathrm{c} \times (h_\gamma * h_\phi)(\Delta x_{kl}, \Delta y_{kl})$$

Wilson & Naylor (2018a)

$$g(\Delta x, \Delta y, \sigma) = (2\pi\sigma^2)^{-1}\exp\left(-\frac{1}{2}\frac{\Delta x^2 + \Delta y^2}{\sigma^2}\right) \quad \text{where} \quad \sigma^2 = \sigma_1^2 + \sigma_2^2$$



Wilson & Naylor (2018b)



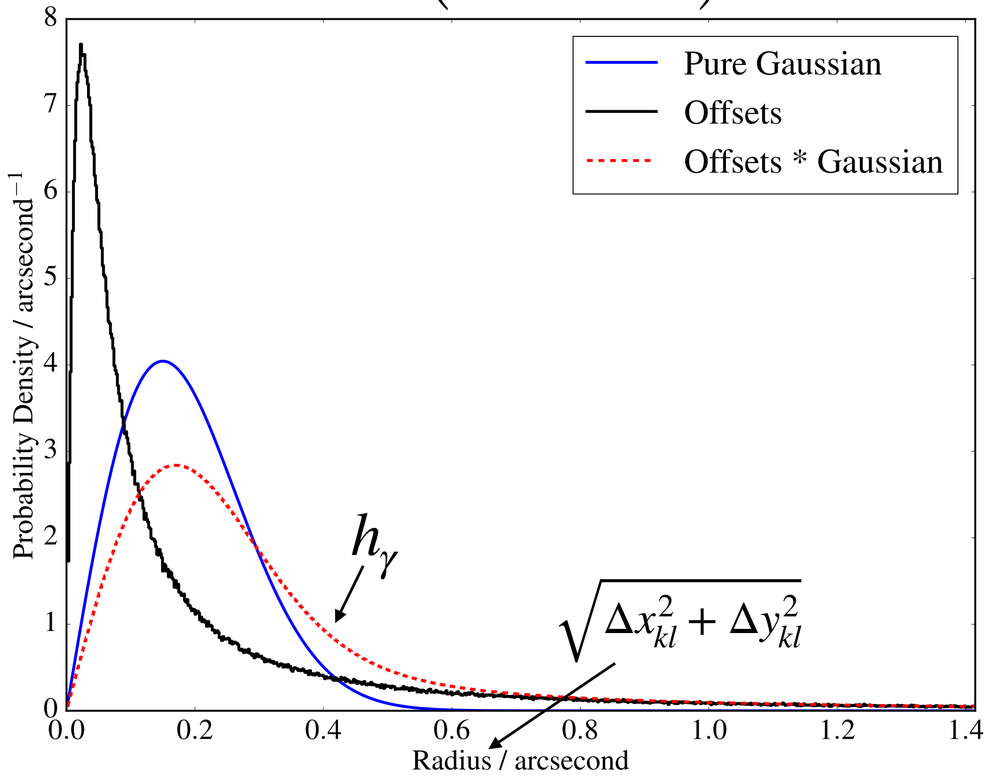
Separation Likelihood Function

$$g(x_k, y_k, x_l, y_l) = \int_{-\infty}^{+\infty}\!\!\!\iint h_\gamma(x_0 - x_k, y_0 - y_k) h_\phi(x_l - x_0, y_l - y_0) p(x_0, y_0)\,\mathrm{d}x_0\,\mathrm{d}y_0$$

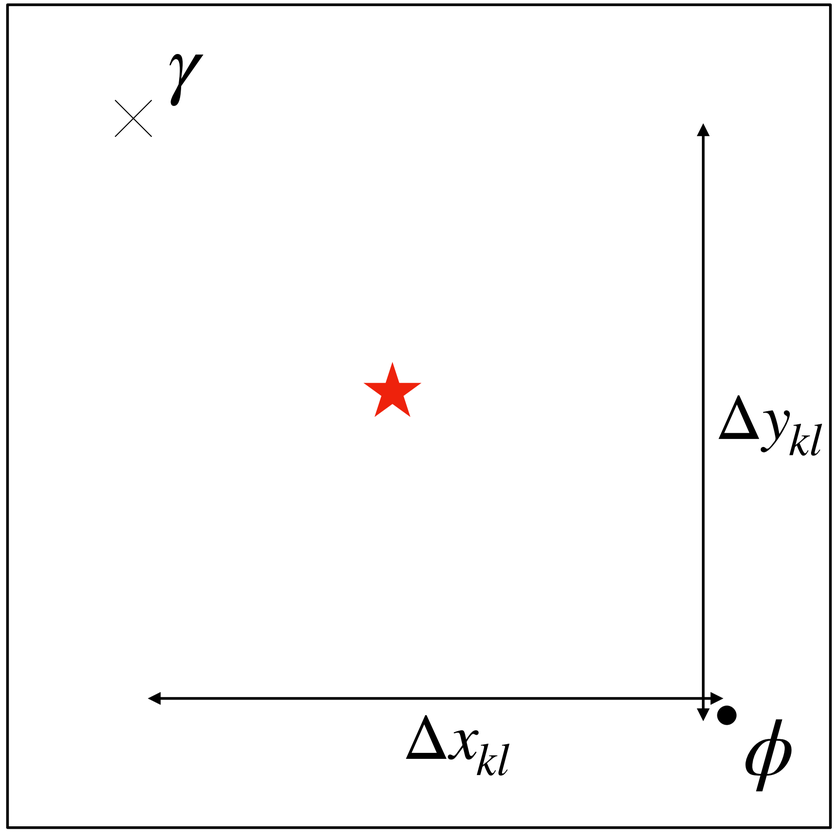
$$= N_\mathrm{c} \times (h_\gamma * h_\phi)(\Delta x_{kl}, \Delta y_{kl})$$

Wilson & Naylor (2018a)

$$g(\Delta x, \Delta y, \sigma) = (2\pi\sigma^2)^{-1}\exp\left(-\frac{1}{2}\frac{\Delta x^2 + \Delta y^2}{\sigma^2}\right) \quad \text{where} \quad \sigma^2 = \sigma_1^2 + \sigma_2^2$$



Wilson & Naylor (2018b)



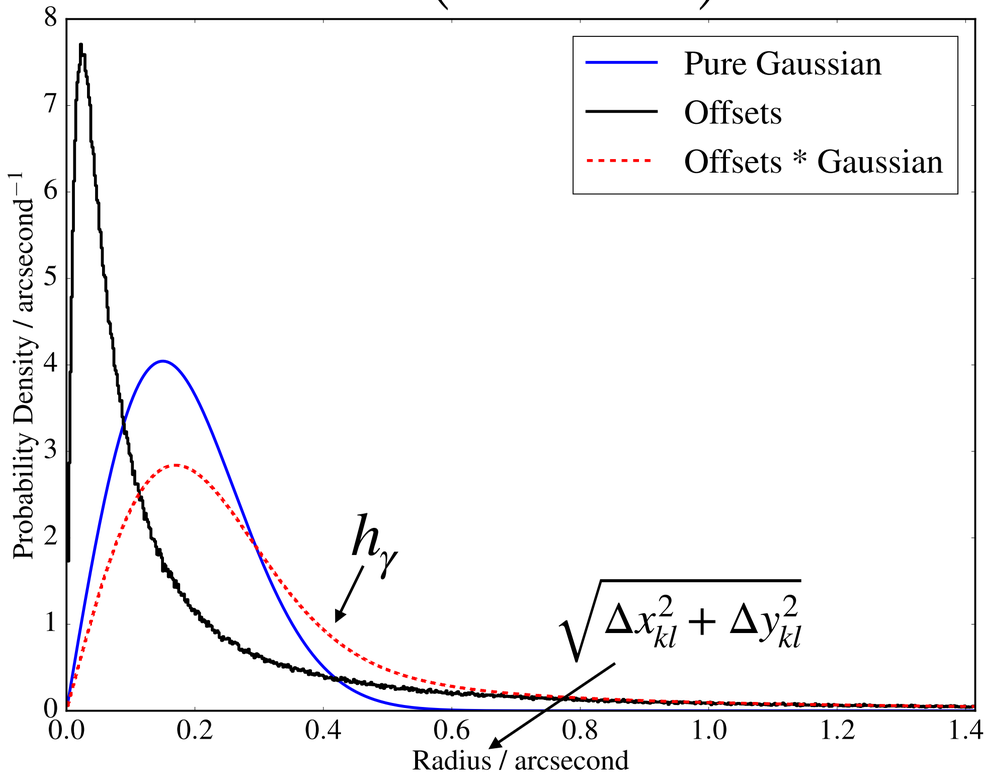
Separation Likelihood Function

$$g(x_k, y_k, x_l, y_l) = \int_{-\infty}^{+\infty}\!\!\!\iint h_\gamma(x_0 - x_k, y_0 - y_k) h_\phi(x_l - x_0, y_l - y_0) p(x_0, y_0)\,\mathrm{d}x_0\,\mathrm{d}y_0$$

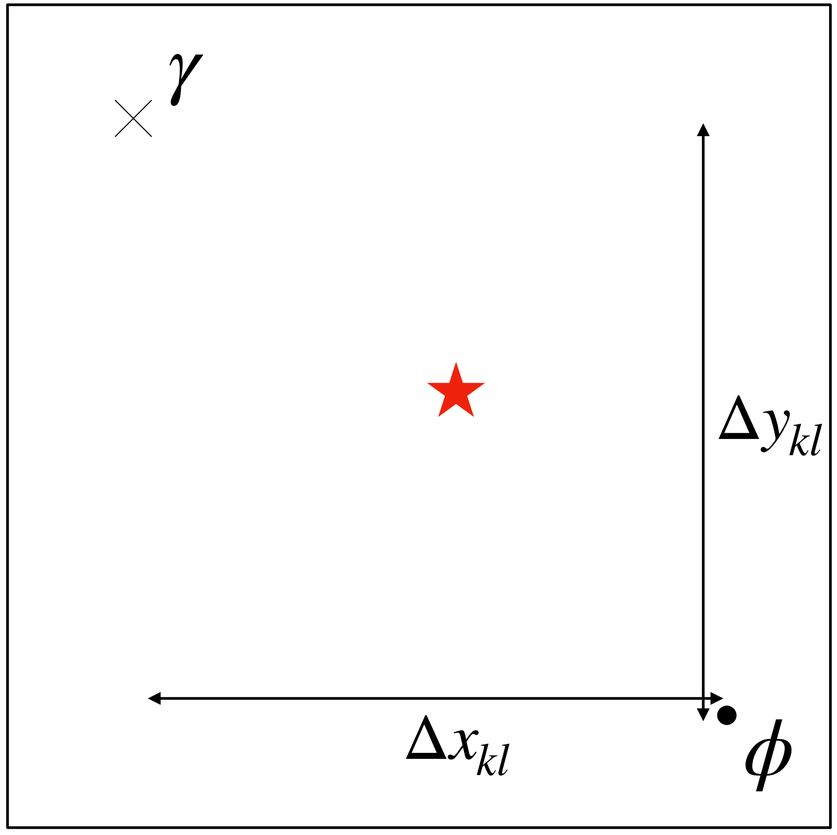
$$= N_\mathrm{c} \times (h_\gamma * h_\phi)(\Delta x_{kl}, \Delta y_{kl})$$

Wilson & Naylor (2018a)

$$g(\Delta x, \Delta y, \sigma) = (2\pi\sigma^2)^{-1}\exp\left(-\frac{1}{2}\frac{\Delta x^2 + \Delta y^2}{\sigma^2}\right) \quad \text{where} \quad \sigma^2 = \sigma_1^2 + \sigma_2^2$$



Wilson & Naylor (2018b)



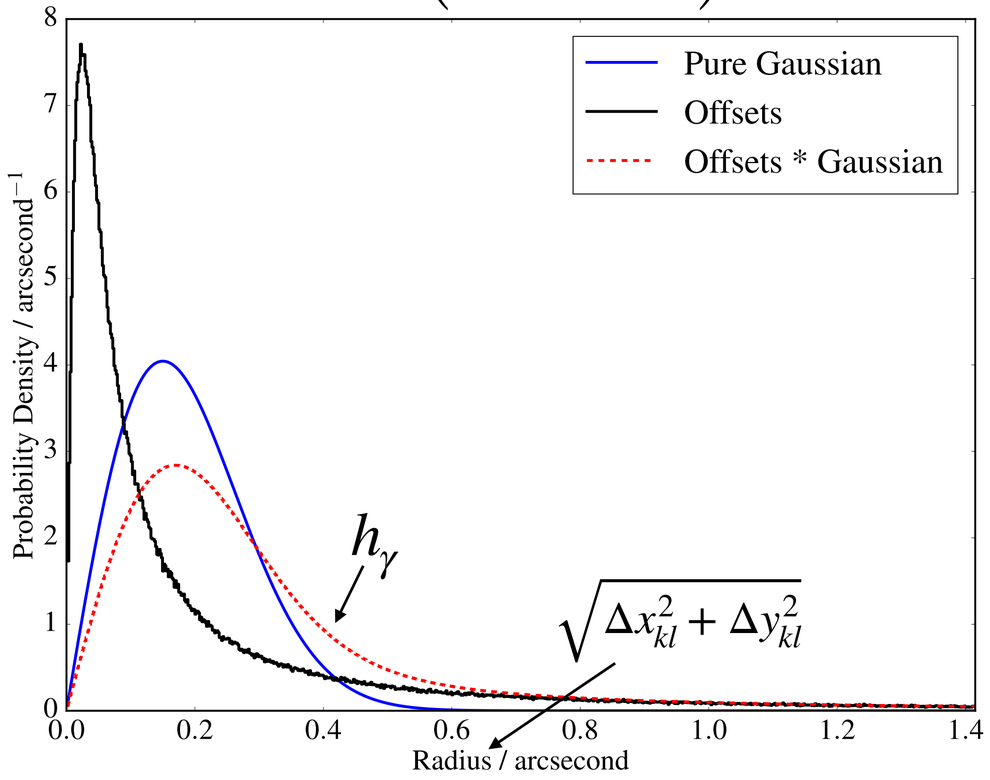
Separation Likelihood Function

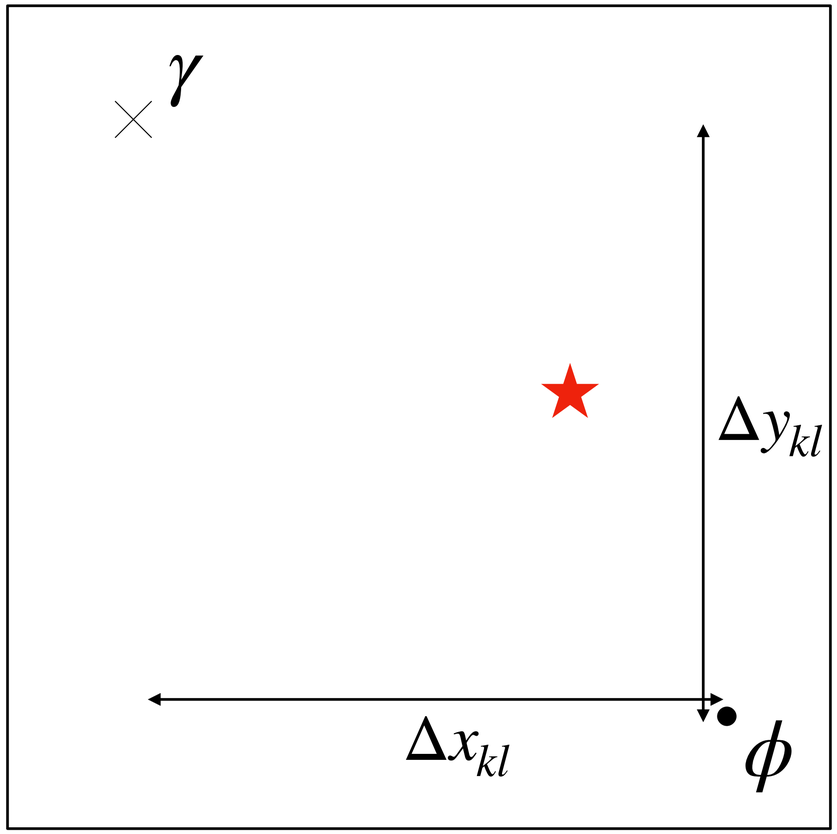
$$g(x_k, y_k, x_l, y_l) = \int_{-\infty}^{+\infty}\!\!\!\iint h_\gamma(x_0 - x_k, y_0 - y_k) h_\phi(x_l - x_0, y_l - y_0) p(x_0, y_0)\,\mathrm{d}x_0\,\mathrm{d}y_0$$

$$= N_\mathrm{c} \times (h_\gamma * h_\phi)(\Delta x_{kl}, \Delta y_{kl})$$

Wilson & Naylor (2018a)

$$g(\Delta x, \Delta y, \sigma) = (2\pi\sigma^2)^{-1}\exp\left(-\frac{1}{2}\frac{\Delta x^2 + \Delta y^2}{\sigma^2}\right) \quad \text{where} \quad \sigma^2 = \sigma_1^2 + \sigma_2^2$$





Wilson & Naylor (2018b)

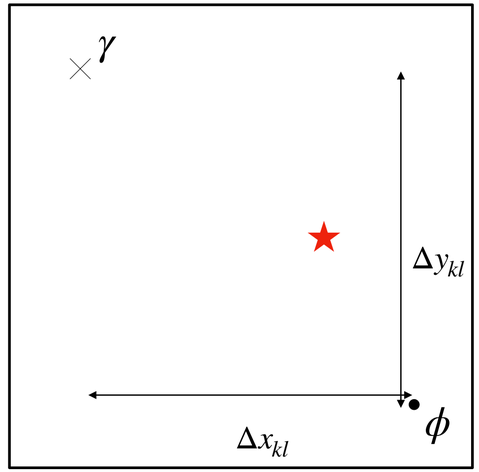
Separation Likelihood Function

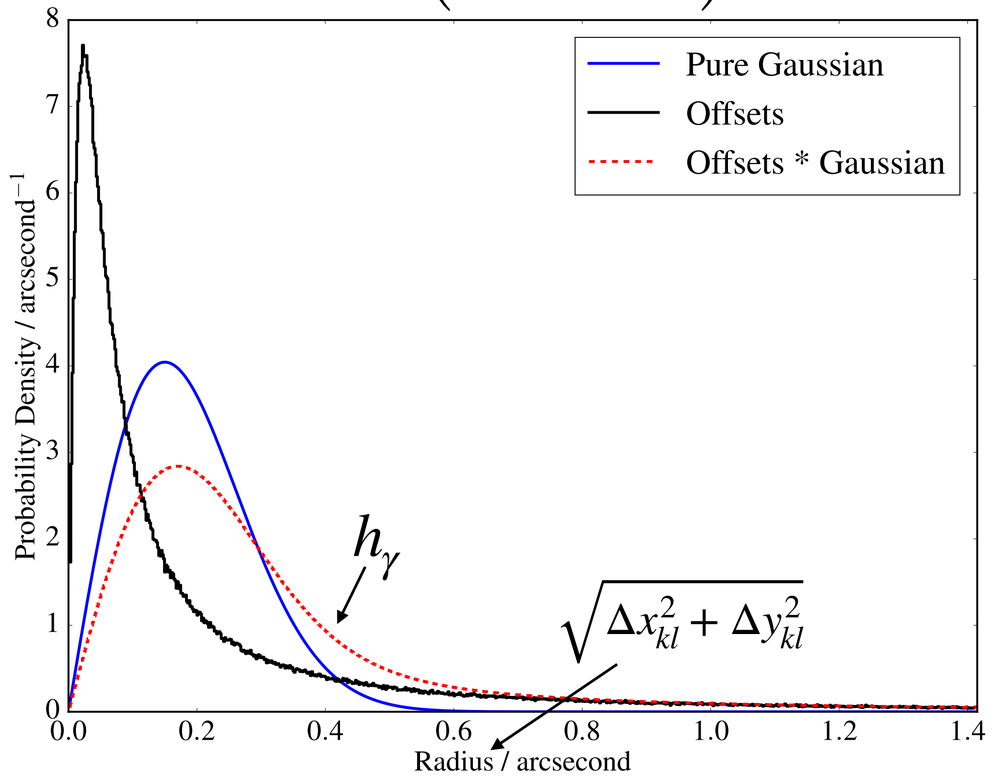
$$g(x_k, y_k, x_l, y_l) = \int_{-\infty}^{+\infty}\!\!\!\iint h_\gamma(x_0 - x_k, y_0 - y_k) h_\phi(x_l - x_0, y_l - y_0) p(x_0, y_0)\,\mathrm{d}x_0\,\mathrm{d}y_0$$

$$= N_\mathrm{c} \times (h_\gamma * h_\phi)(\Delta x_{kl}, \Delta y_{kl})$$

Wilson & Naylor (2018a)

$$g(\Delta x, \Delta y, \sigma) = (2\pi\sigma^2)^{-1}\exp\left(-\frac{1}{2}\frac{\Delta x^2 + \Delta y^2}{\sigma^2}\right) \quad \text{where} \quad \sigma^2 = \sigma_1^2 + \sigma_2^2$$

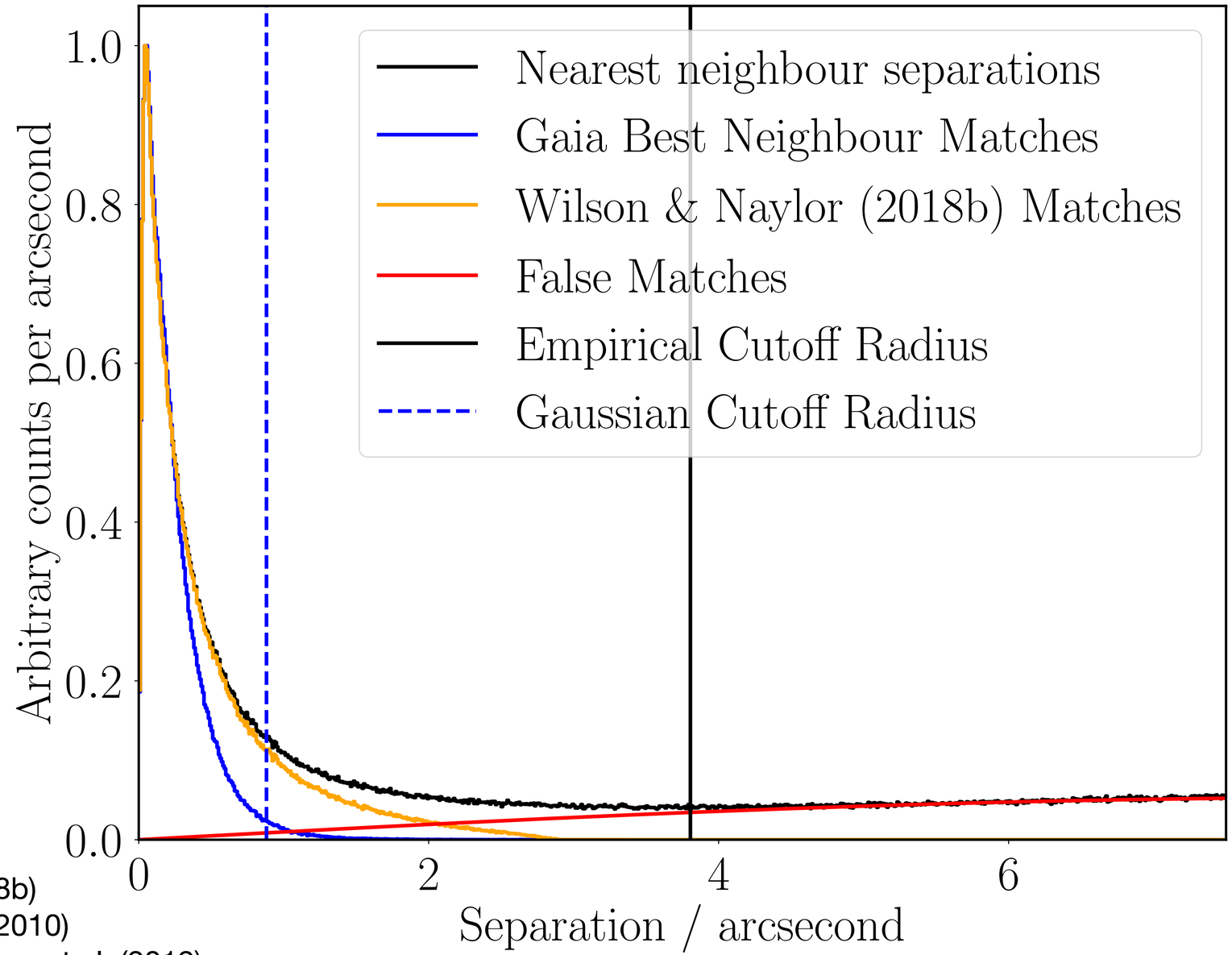




Wilson & Naylor (2018b)

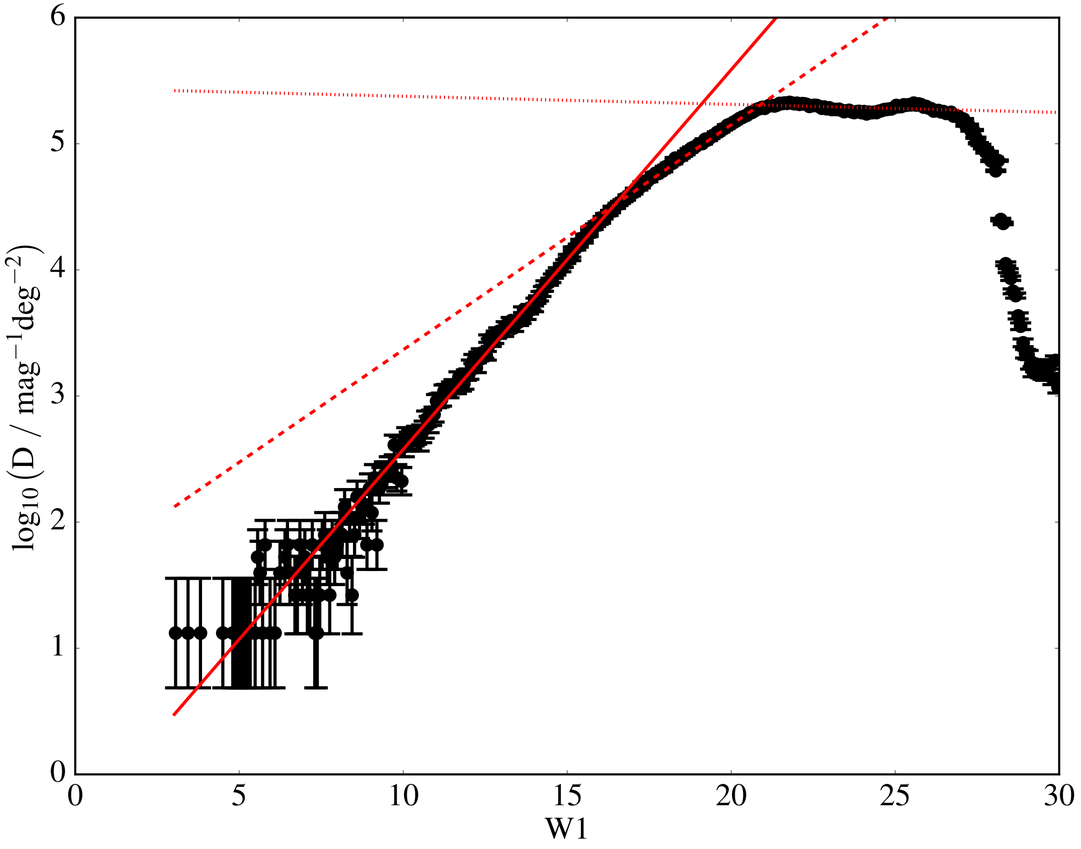
John Herschel's result, 170 years on, appears wrong, but only because the universe got in the way.
"Were the succession of stars endless... there could be absolutely no point, in all that background, at which would not exist a star."
— Edgar Allan Poe, Eureka (1848)

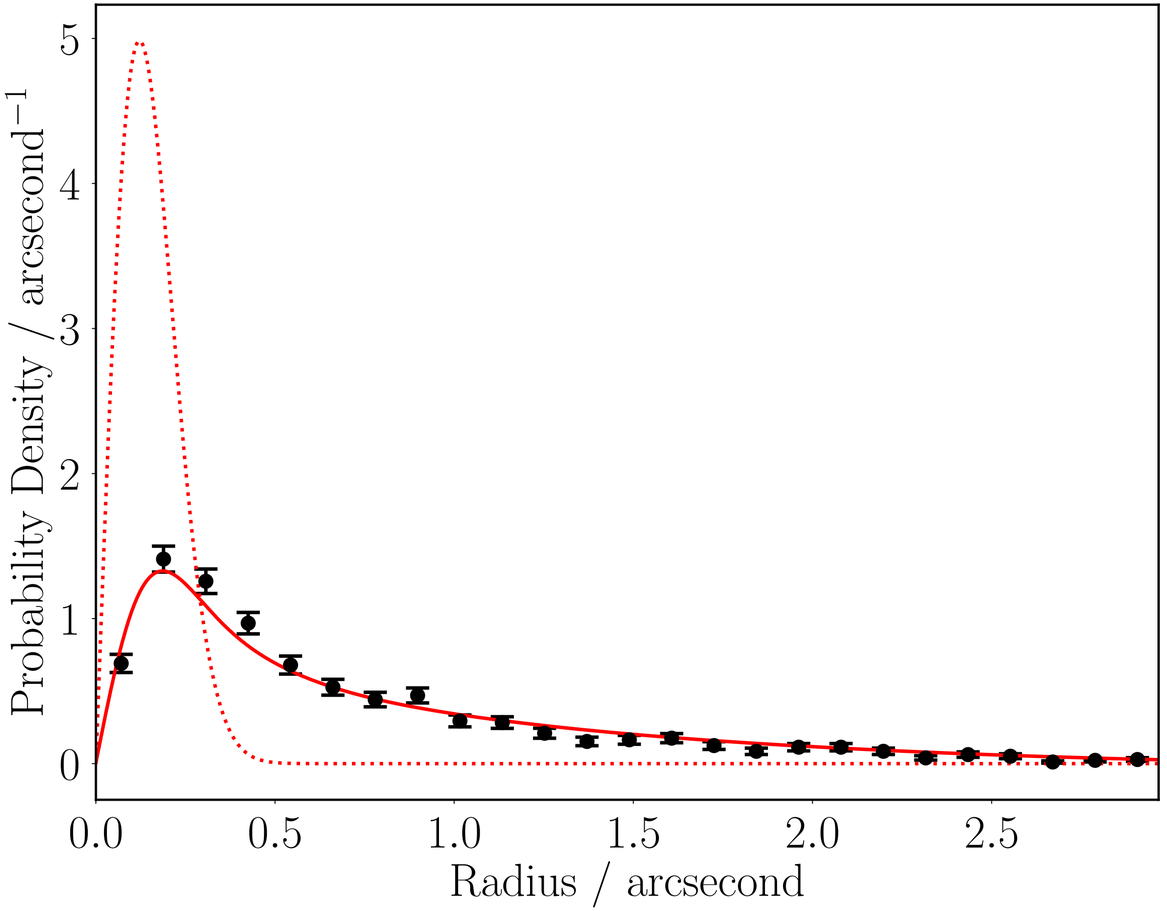
Match Separations



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)

Building Empirical AUFs

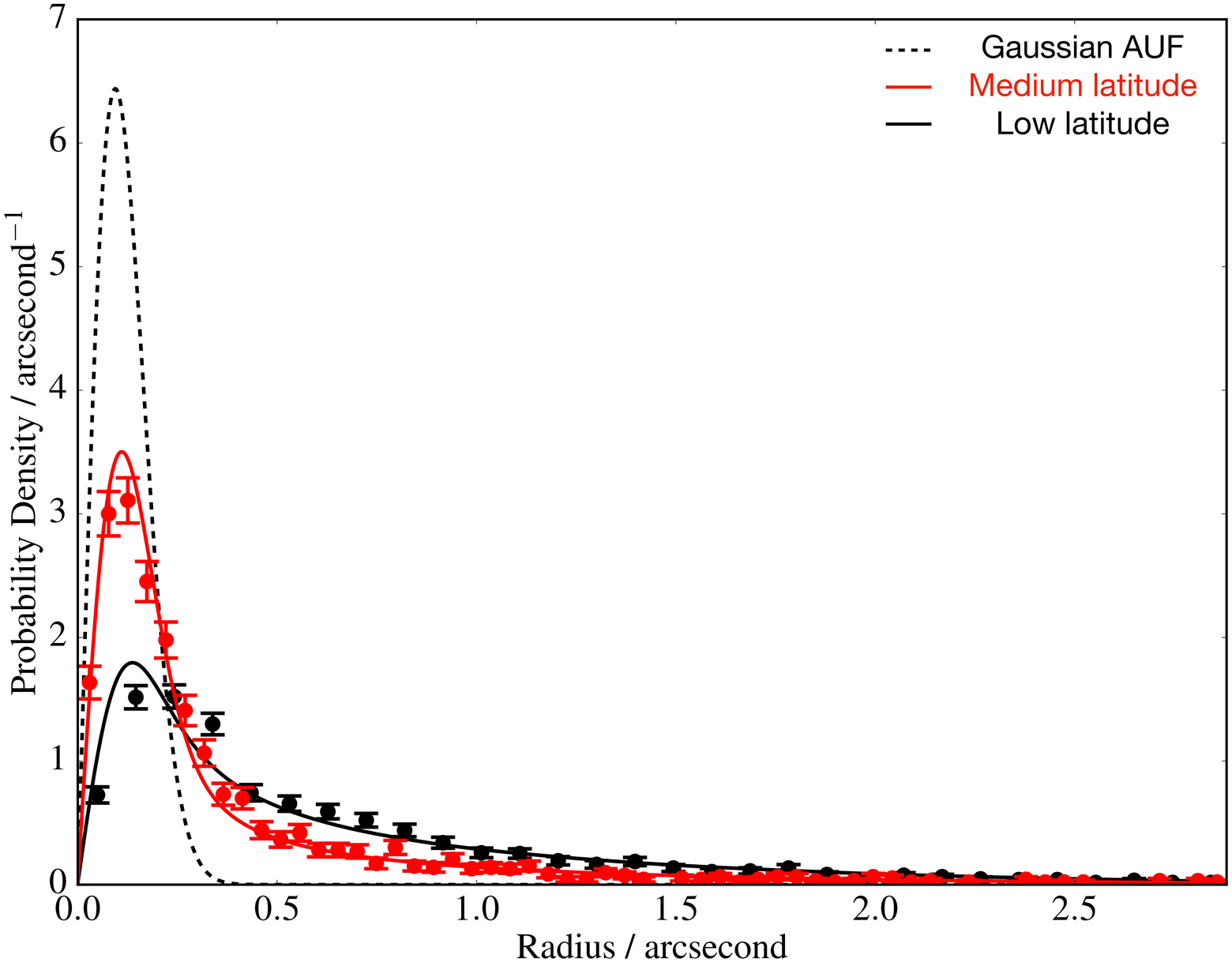




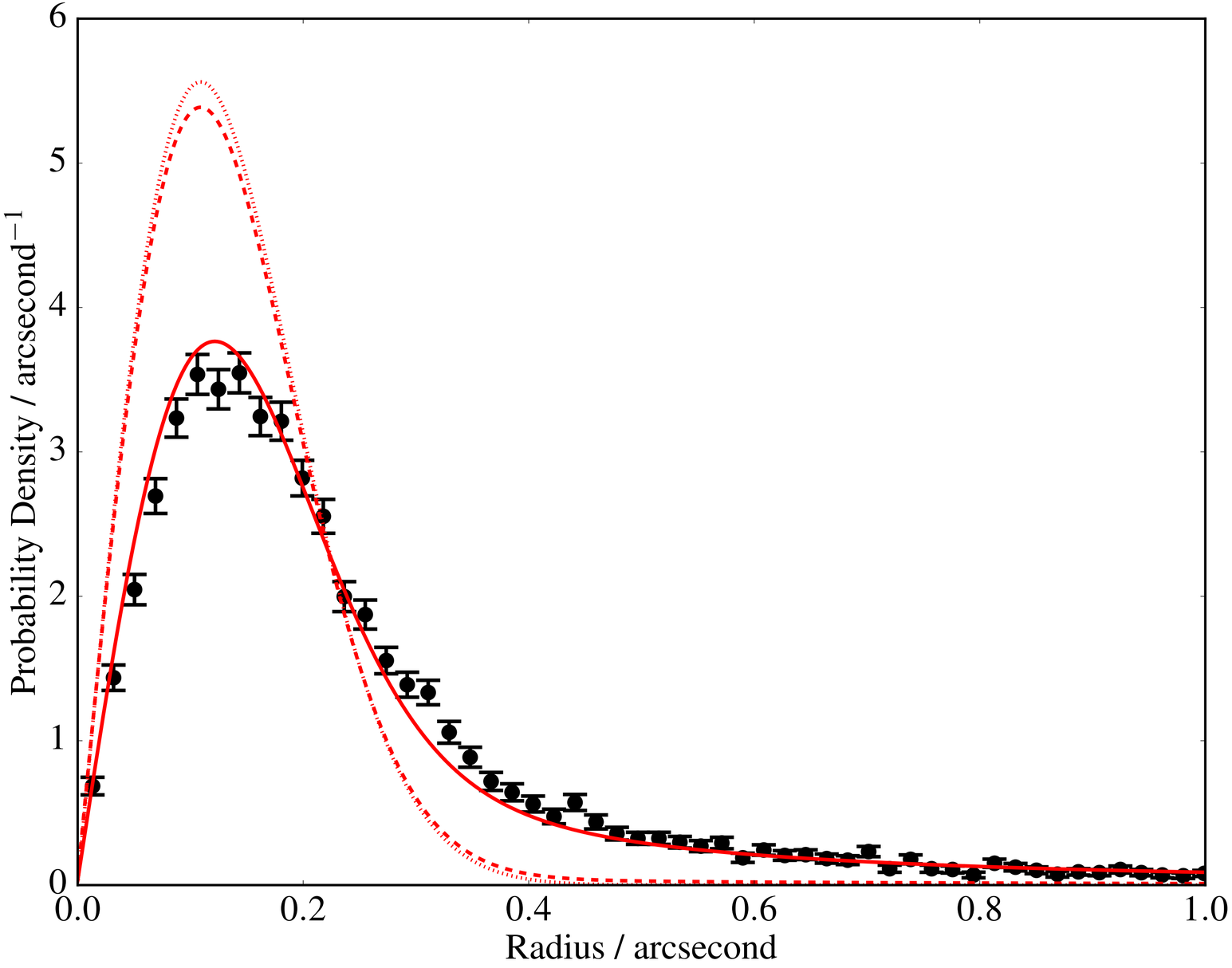
Wilson & Naylor (2018b)
TRILEGAL - Girardi et al. (2005)

Tom J Wilson @onoddil

Crowding Normalisation

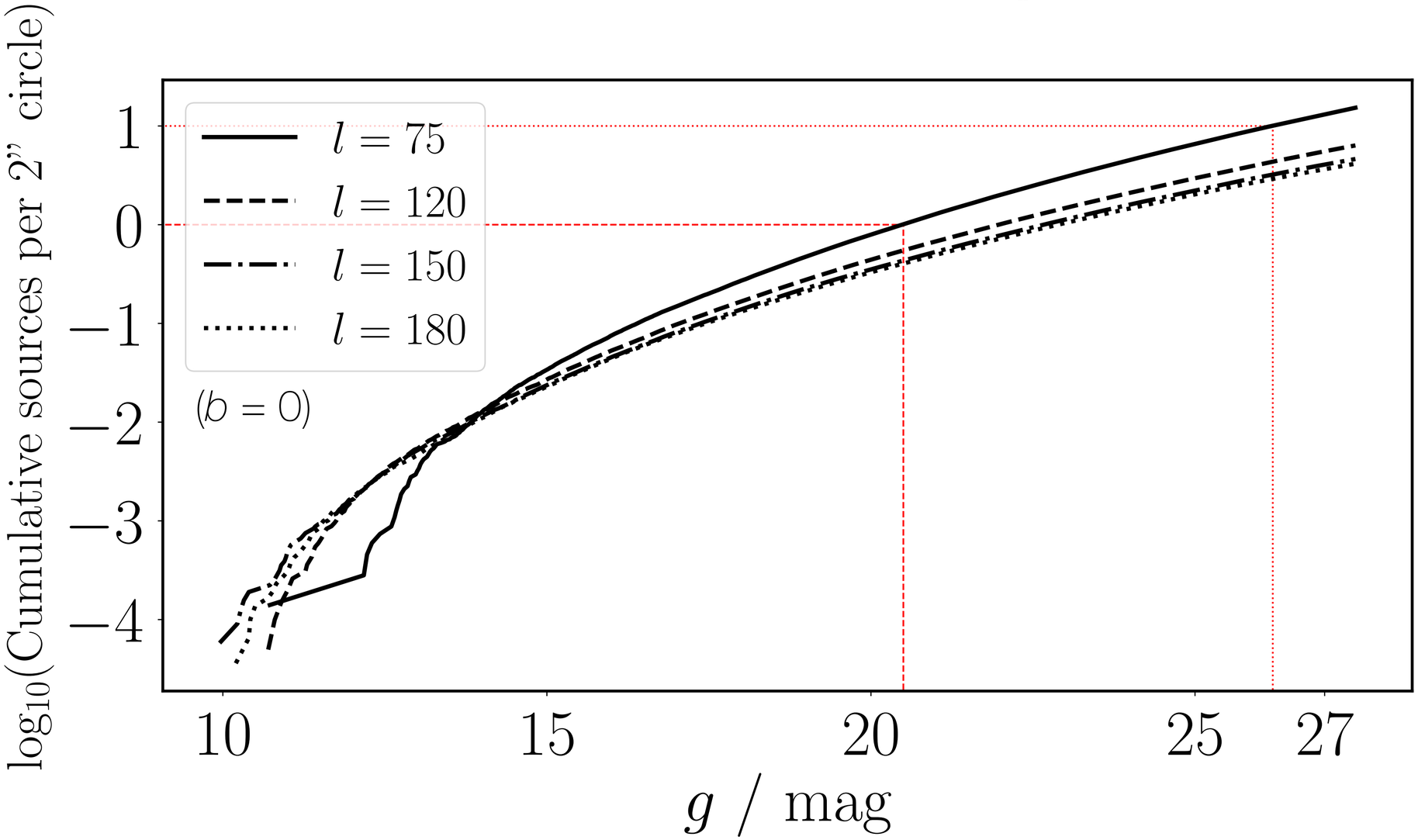


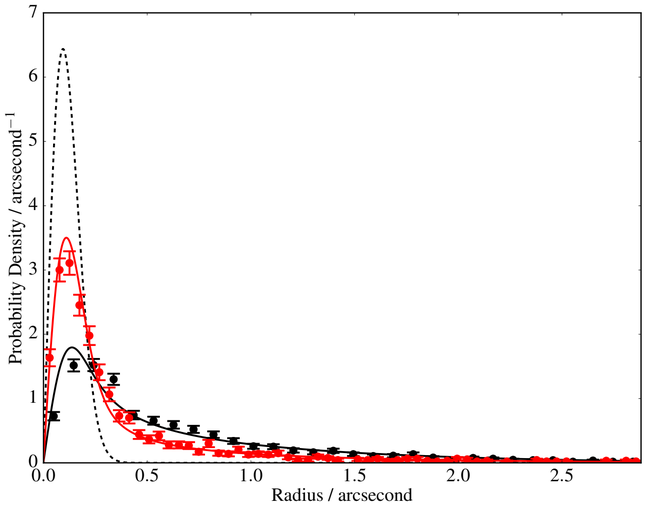
Extra-galactic Effects of Crowding



Wilson & Naylor (2018b)

Vera C. Rubin Observatory's LSST

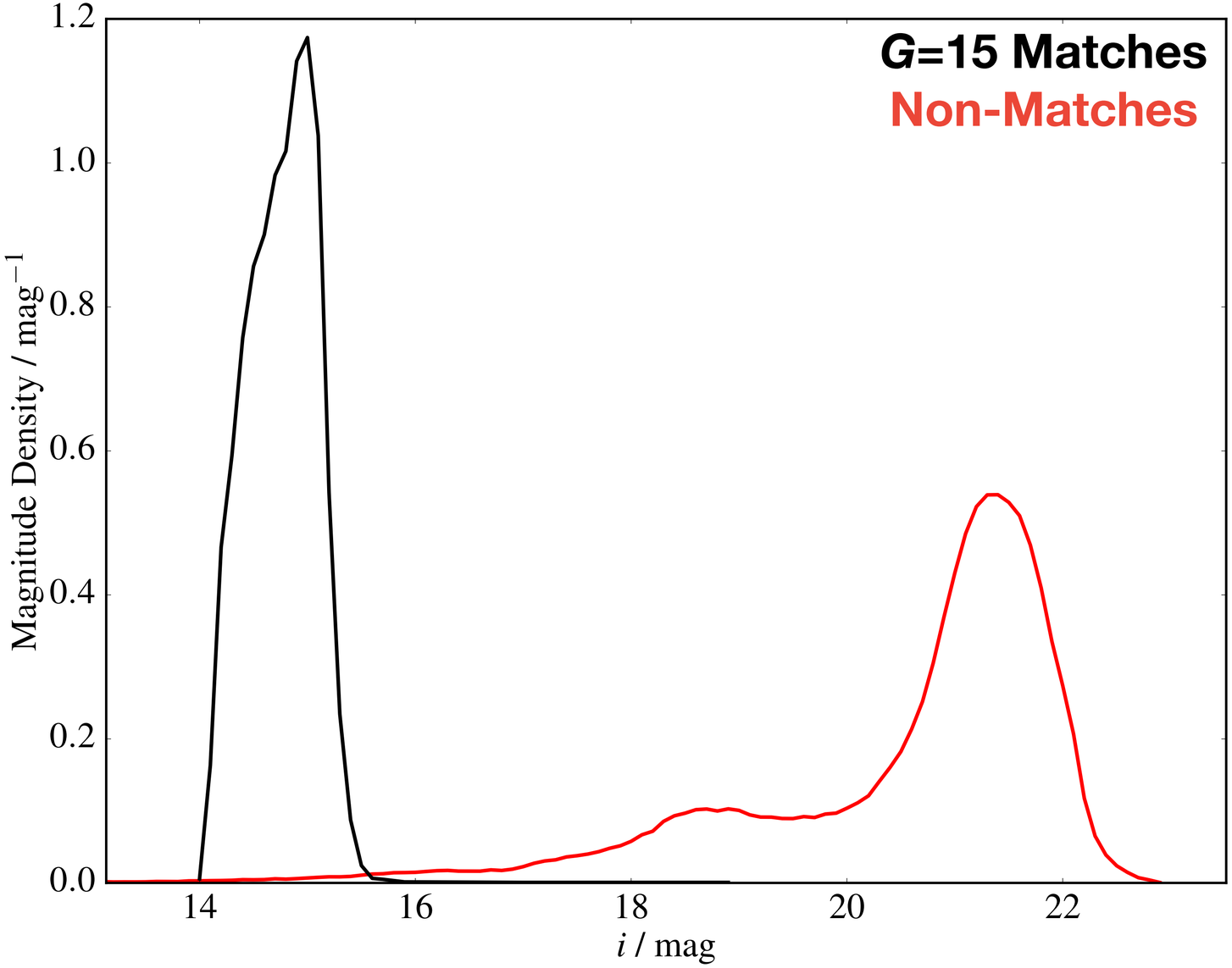


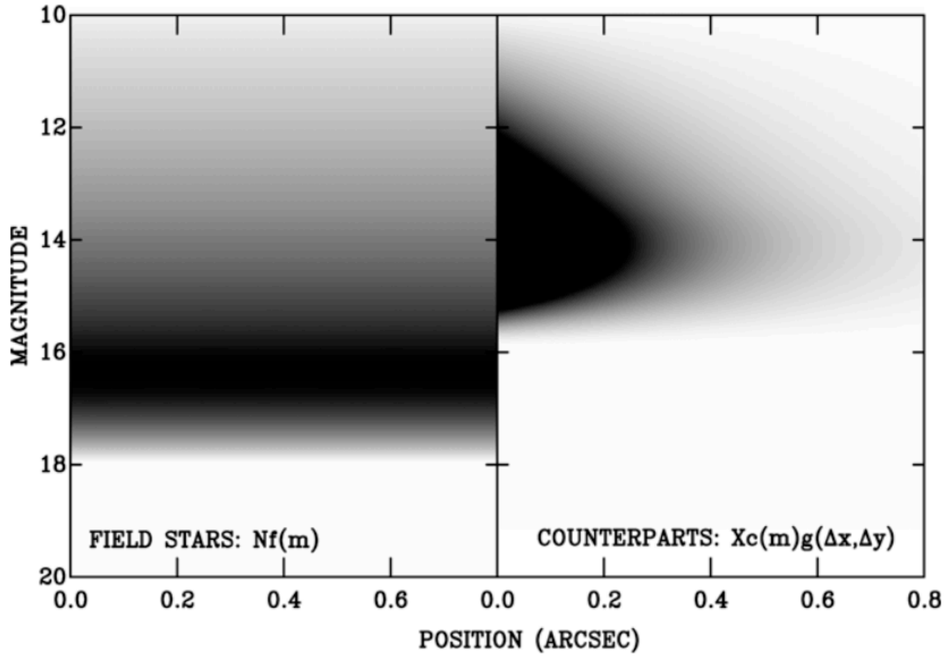


TRILEGAL - Girardi et al. (2005)

Tom J Wilson @onoddil

Including the Magnitude Information



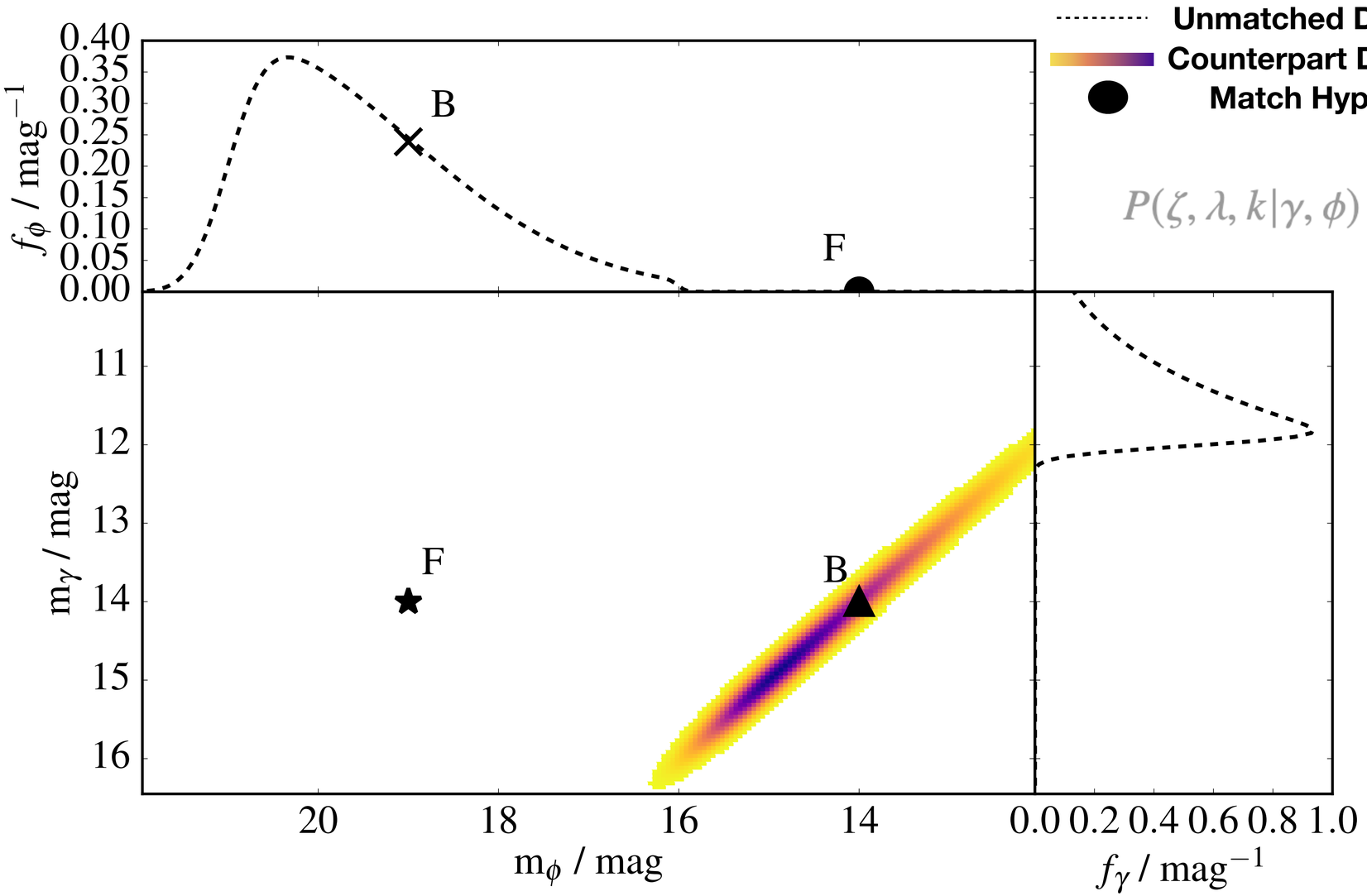


Naylor, Broos, & Feigelson (2013)

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

Tom J Wilson @onoddil

Including the Magnitude Information

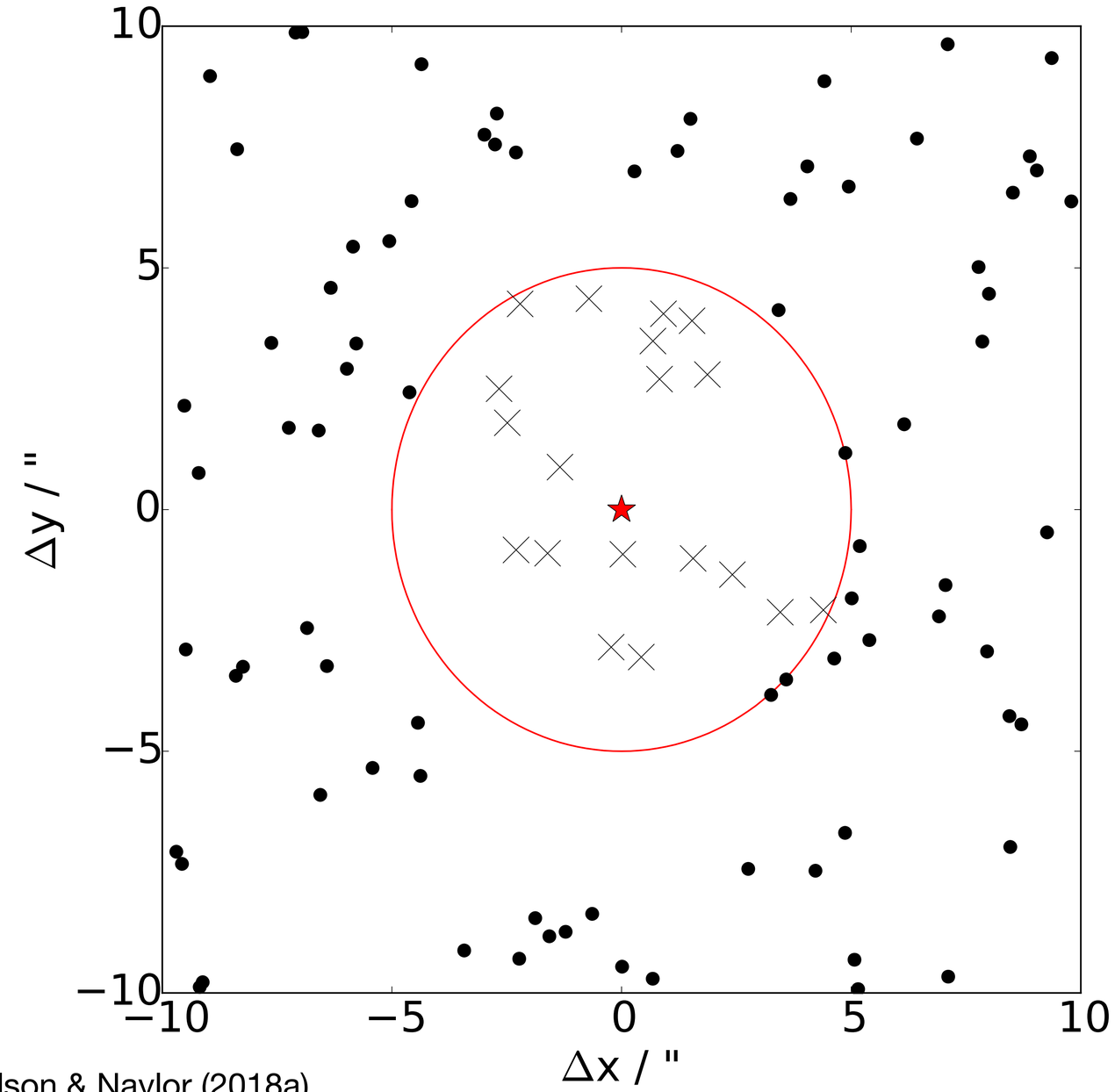


$$P(\zeta, \lambda, k|\gamma, \phi) = K^{-1} \times \prod_{\delta \notin \zeta \cap \delta \in \gamma} N_\gamma f_\gamma^\delta \prod_{\omega \notin \lambda \cap \omega \in \phi} N_\phi f_\phi^\omega \prod_{i=1}^{k} N_c G_{\gamma\phi}^{\zeta_i \lambda_i} c_{\gamma\phi}^{\zeta_i \lambda_i}$$

Unmatched **Matched**

Wilson & Naylor (2018a)

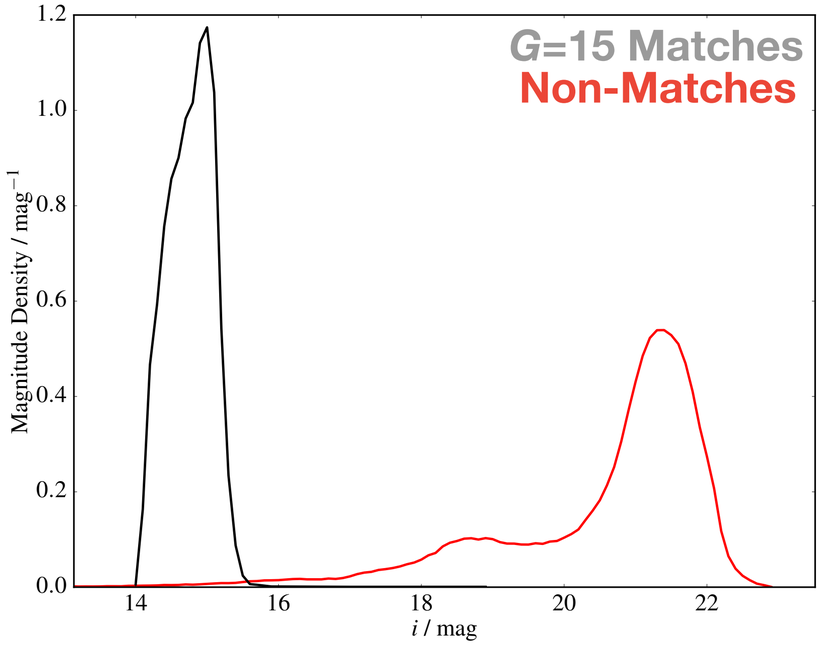
The “Field Star” Distribution



Wilson & Naylor (2018a)

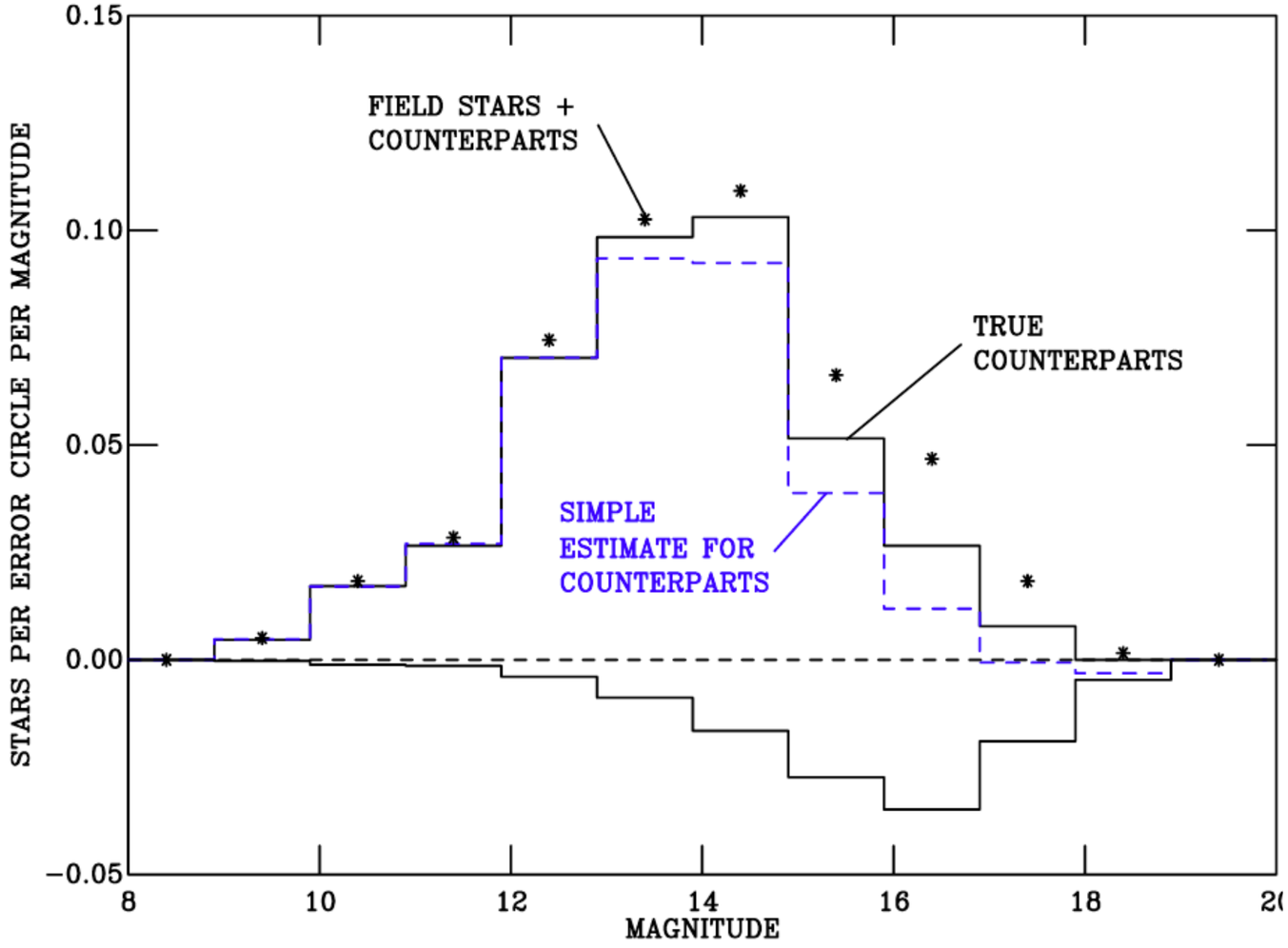
$$P(\zeta, \lambda, k|\gamma, \phi) = K^{-1} \times \prod_{\delta \notin \zeta \cap \delta \in \gamma} N_\gamma f_\gamma^\delta \prod_{\omega \notin \lambda \cap \omega \in \phi} N_\phi f_\phi^\omega \prod_{i=1}^{k} N_c G_{\gamma\phi}^{\zeta_i \lambda_i} c_{\gamma\phi}^{\zeta_i \lambda_i}$$

Unmatched **Matched**

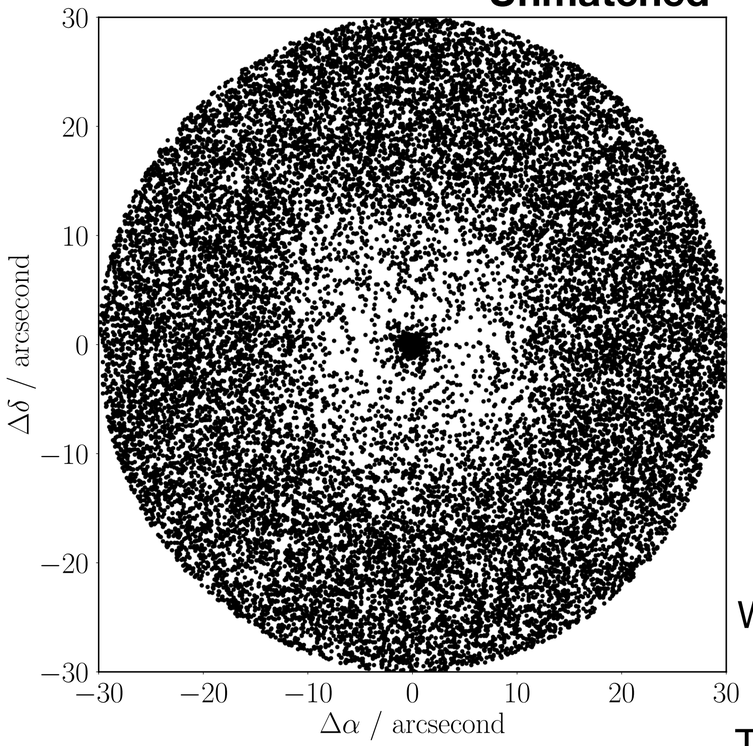


Tom J Wilson @onoddil

The Counterpart Source Distribution



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

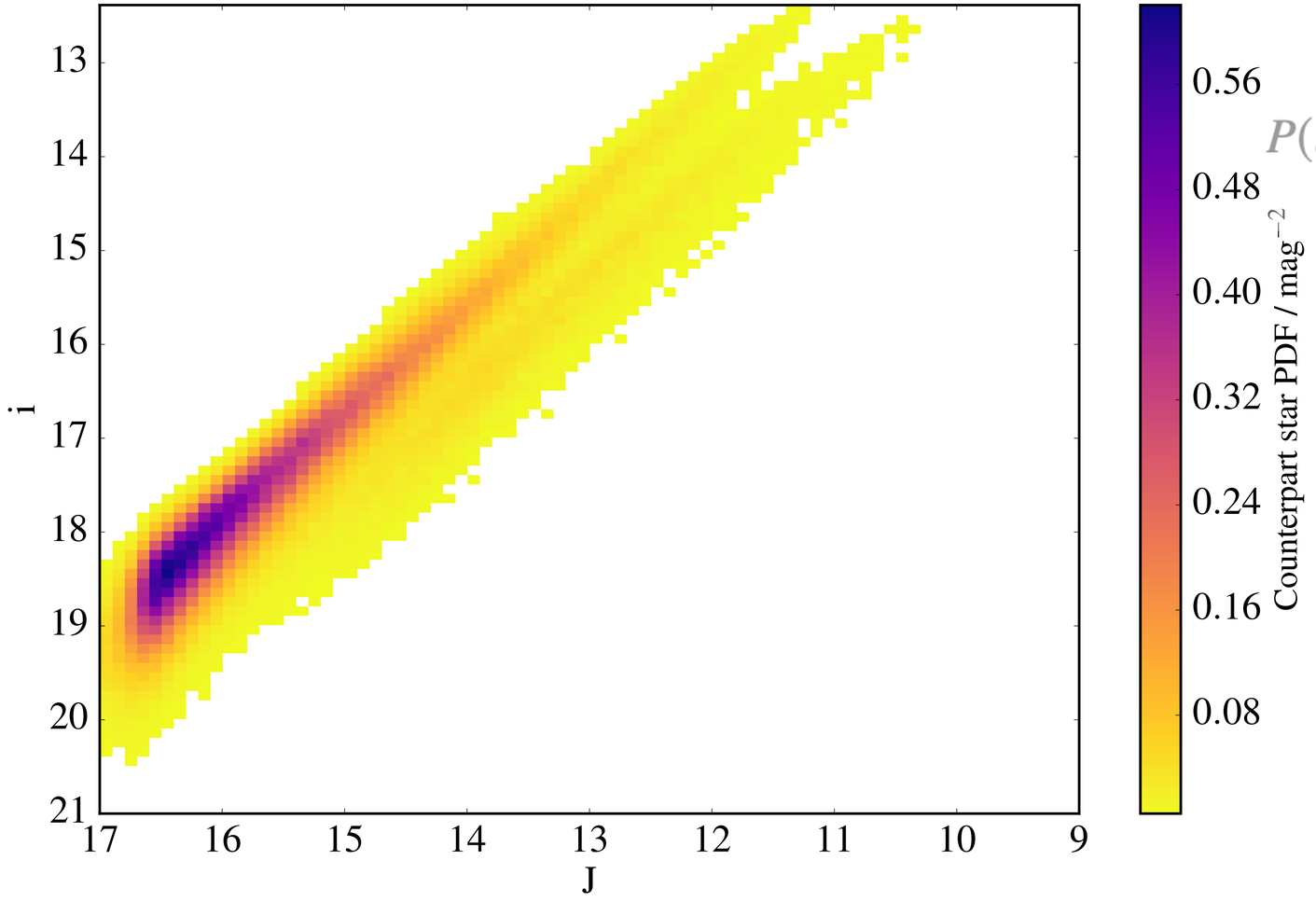


Wilson & Naylor (2017)

Naylor, Broos, & Feigelson (2013); also see Wilson & Naylor (2018a)

Tom J Wilson @onoddil

The Counterpart Source Distribution

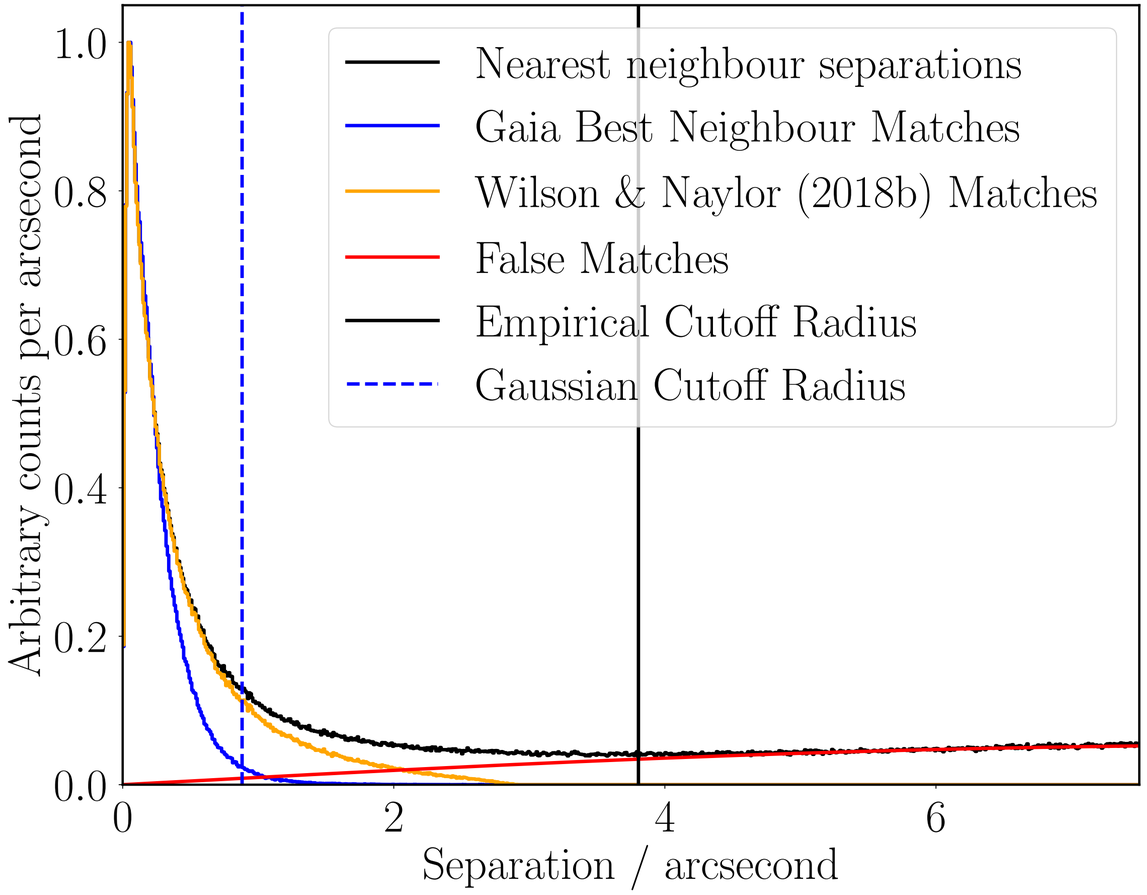


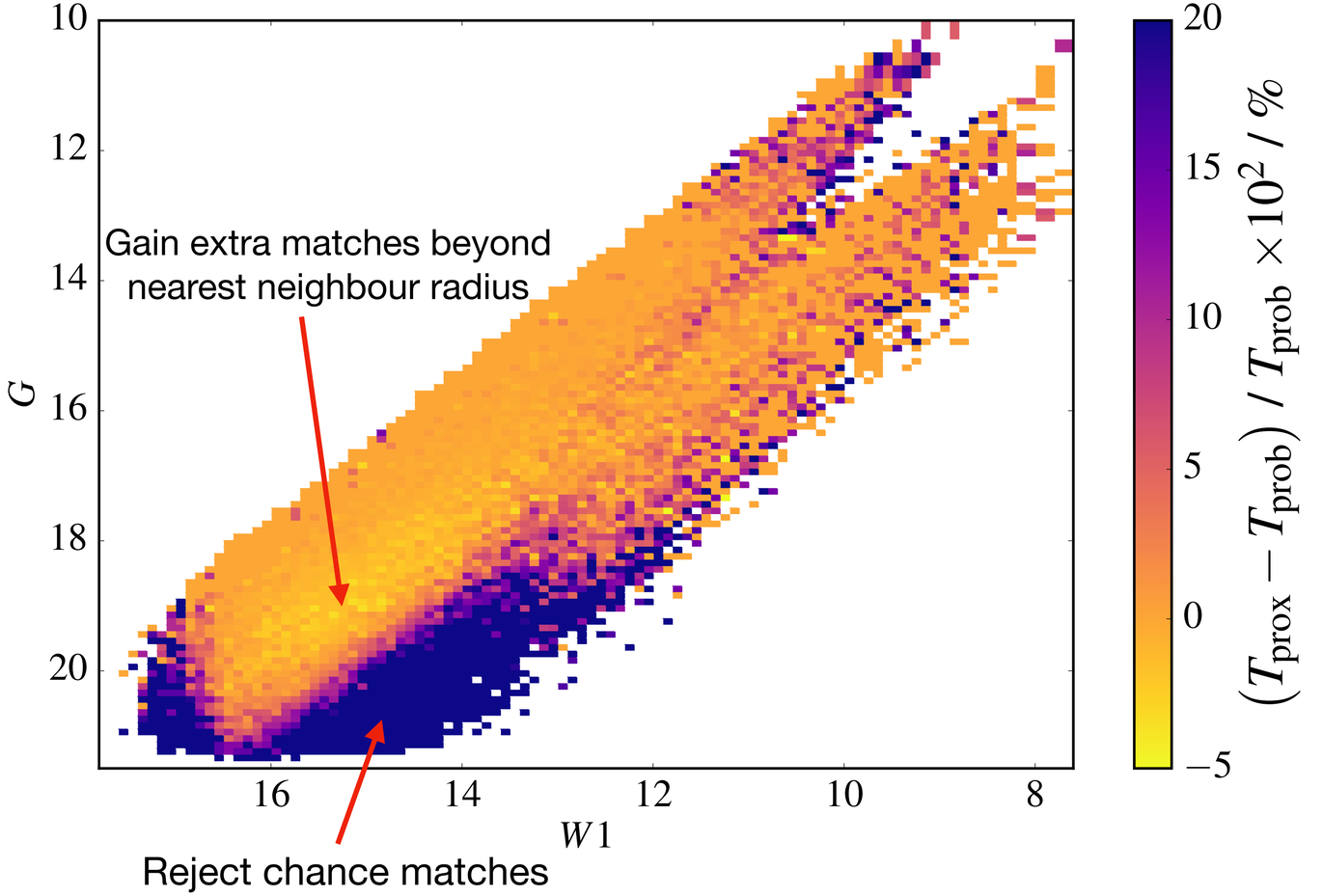
$$P(\zeta, \lambda, k|\gamma, \phi) = K^{-1} \times \prod_{\delta \notin \zeta \cap \delta \in \gamma} N_\gamma f_\gamma^\delta \prod_{\omega \notin \lambda \cap \omega \in \phi} N_\phi f_\phi^\omega \prod_{i=1}^{k} N_c G_{\gamma\phi}^{\zeta_i \lambda_i} c_{\gamma\phi}^{\zeta_i \lambda_i}$$

Unmatched **Matched**

Wilson & Naylor (2018a)

Comparing Match Distributions

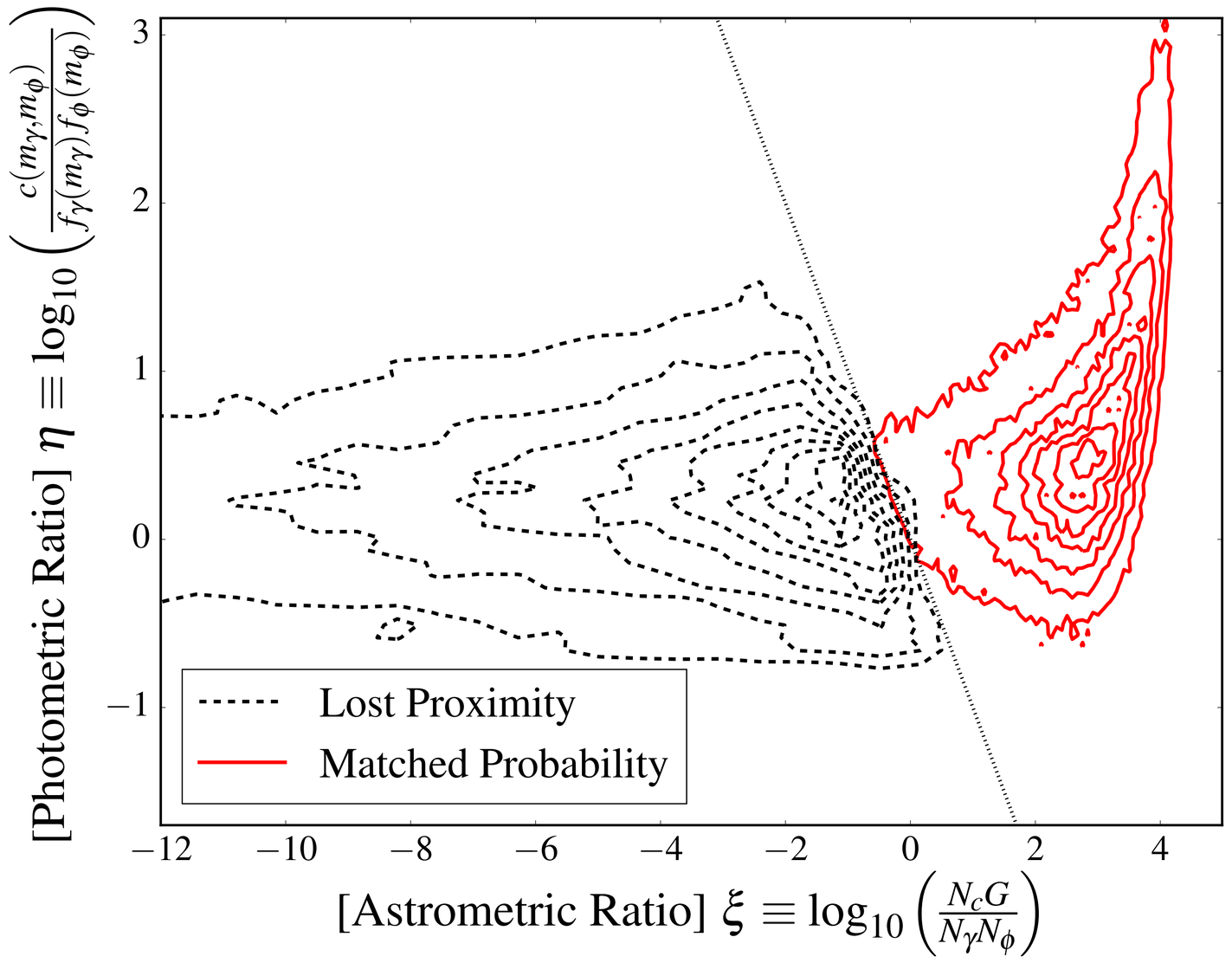




Wilson & Naylor (2018b)
WISE - Wright et al. (2010)
Gaia matches - Marrese et al. (2019)
Gaia DR2 - Gaia Collaboration, Brown A. G. A., et al. (2018)

Tom J Wilson @onoddil

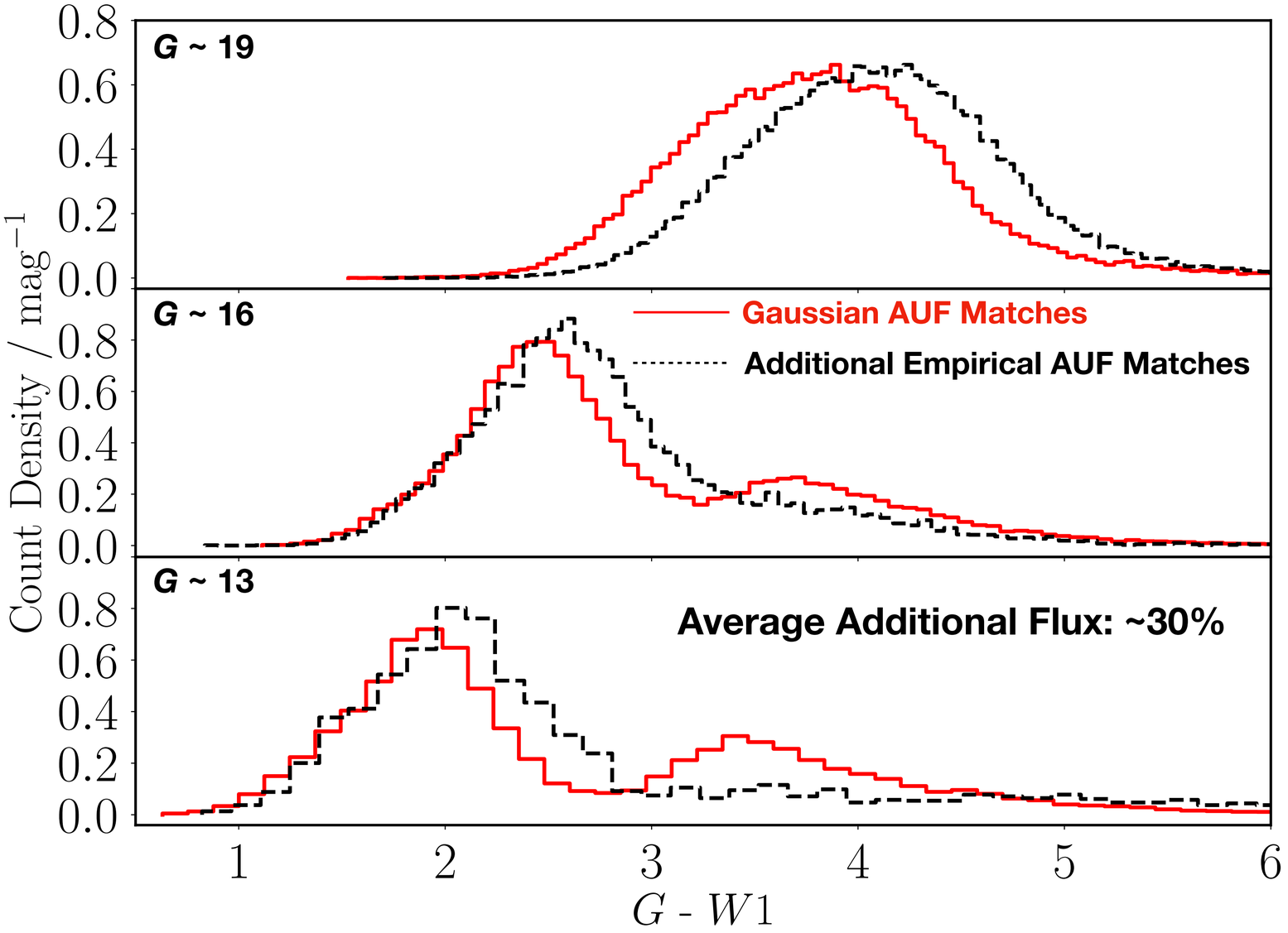
Lost Gaussian-Only Matches

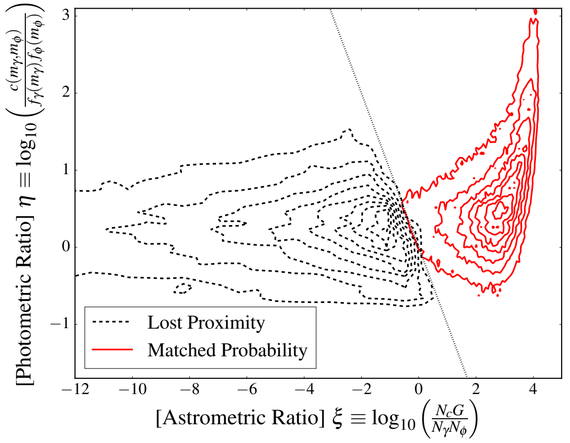


Wilson & Naylor (2018b)

Tom J Wilson @onoddil

Perturbation-Colour Correlation



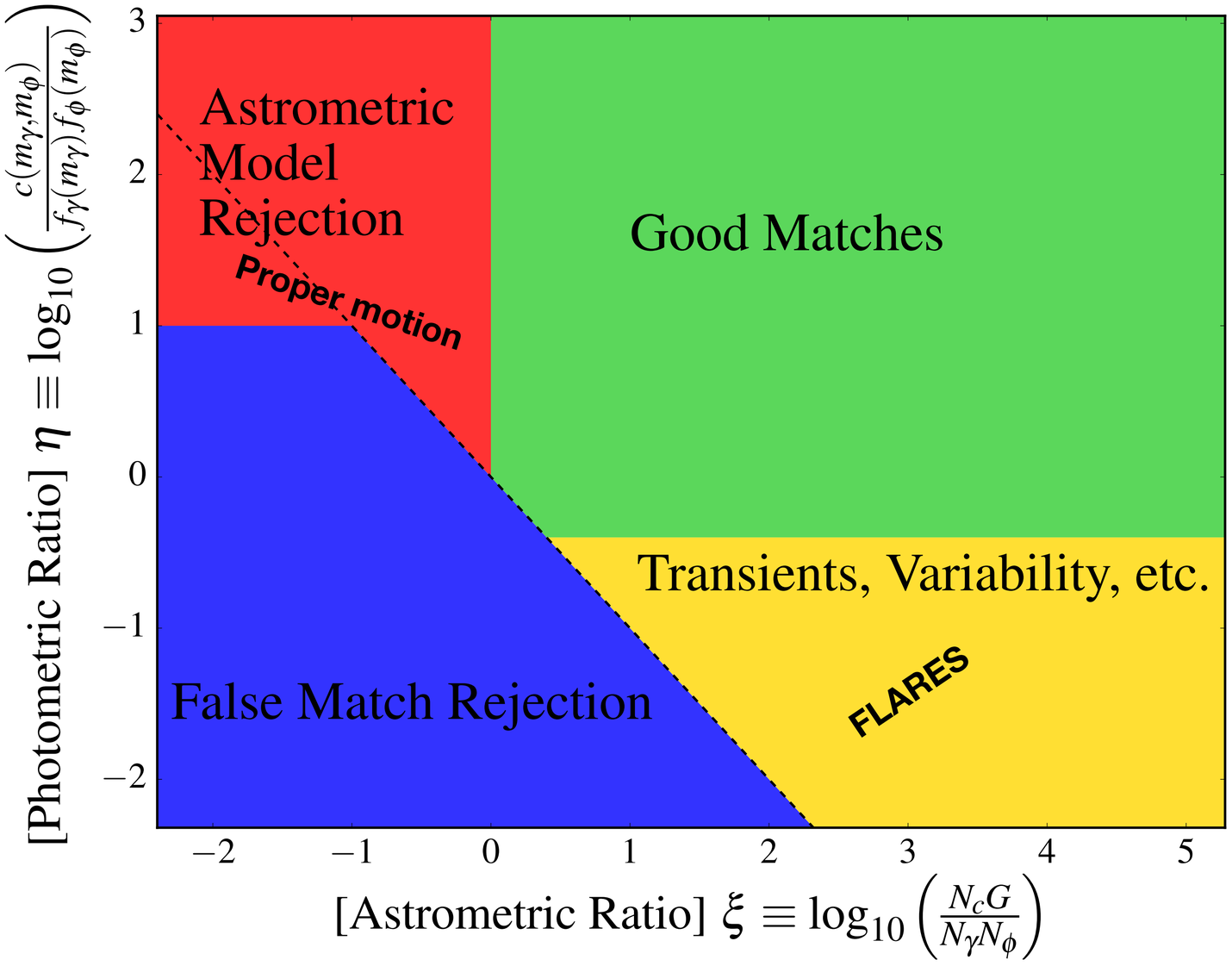


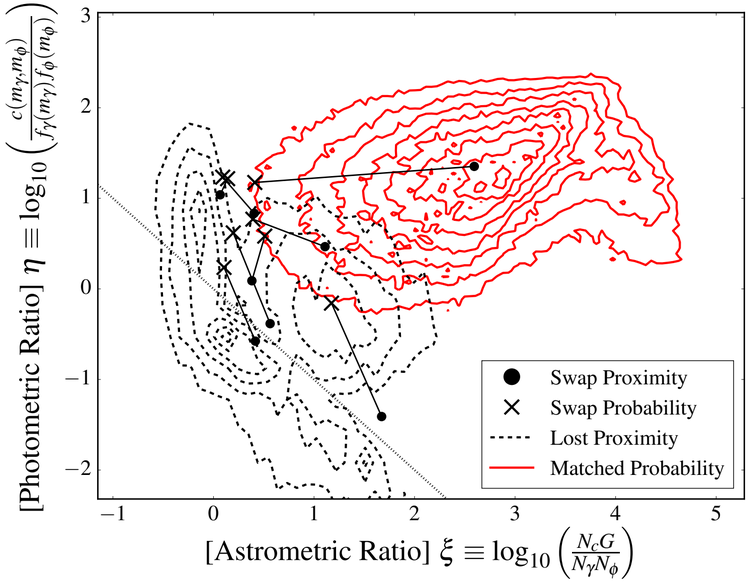
"Extra flux" has an impact on derived proper motions and parallaxes, and IR excesses!

Wilson & Naylor (2018b)

Tom J Wilson @onoddil

The Likelihood Ratio Space

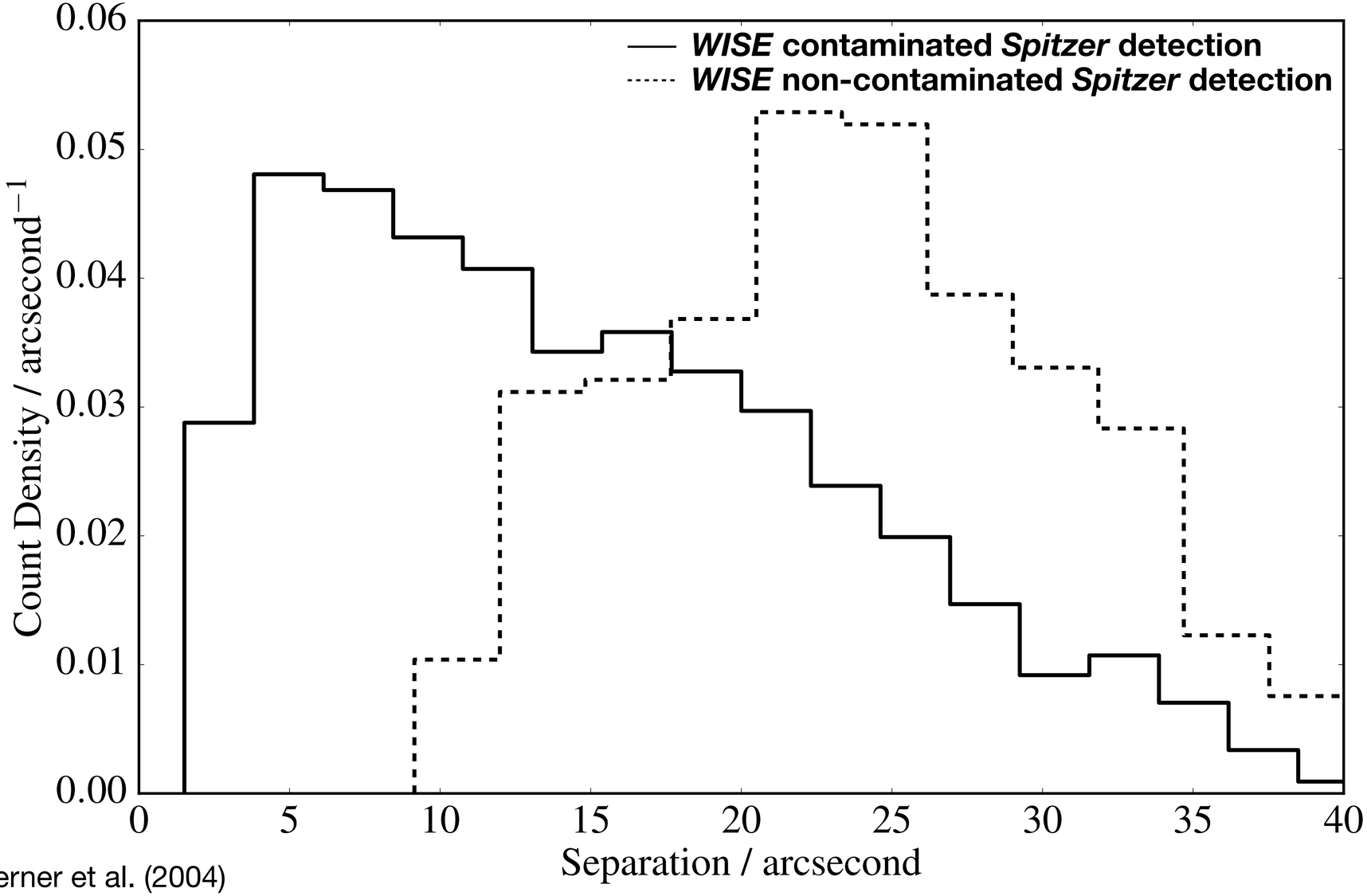




Wilson & Naylor (2018b)

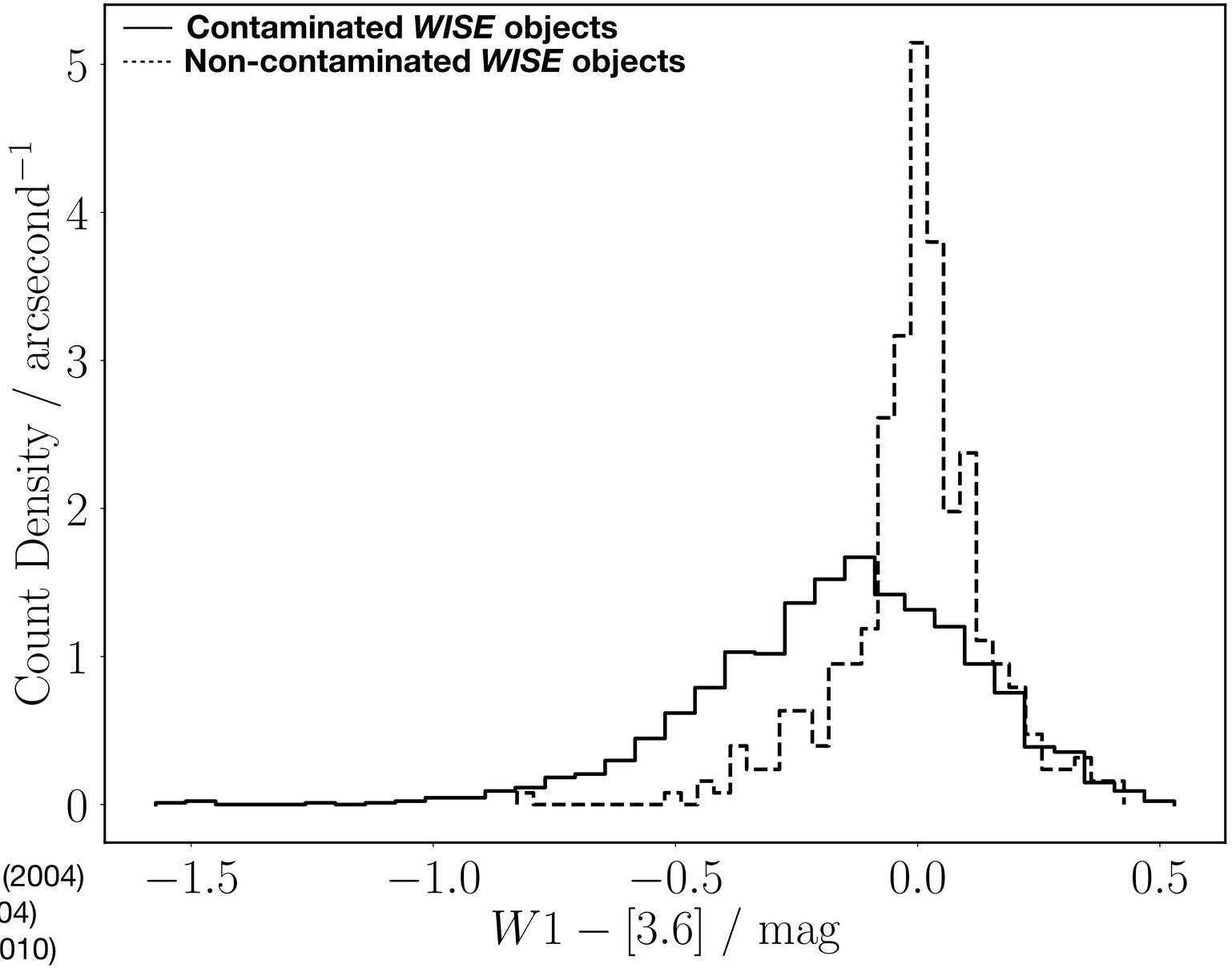
Tom J Wilson @onoddil

Resolving Contaminants



Spitzer - Werner et al. (2004)
IRAC - Fazio et al. (2004)
WISE - Wright et al. (2010)
Wilson & Naylor (2018b)

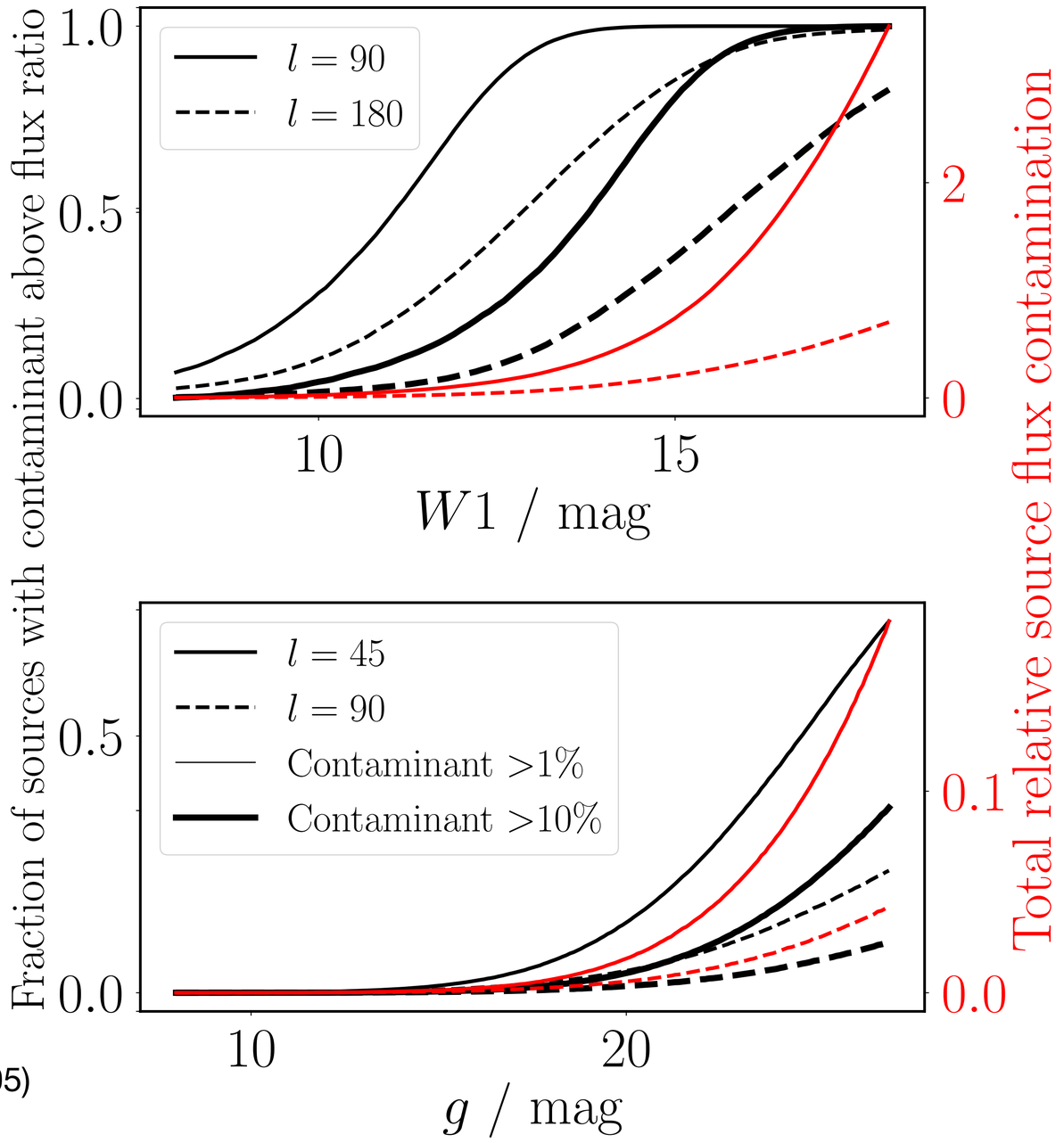
Resolving Contaminant Flux



Spitzer - Werner et al. (2004)
IRAC - Fazio et al. (2004)
WISE - Wright et al. (2010)
Wilson & Naylor (2018b)

Tom J Wilson @onoddil

Contamination Rates and Amounts



TRILEGAL - Girardi et al. (2005)
Wilson & Naylor (2018b)

Tom J Wilson @onoddil

Open Source Code: Macauff

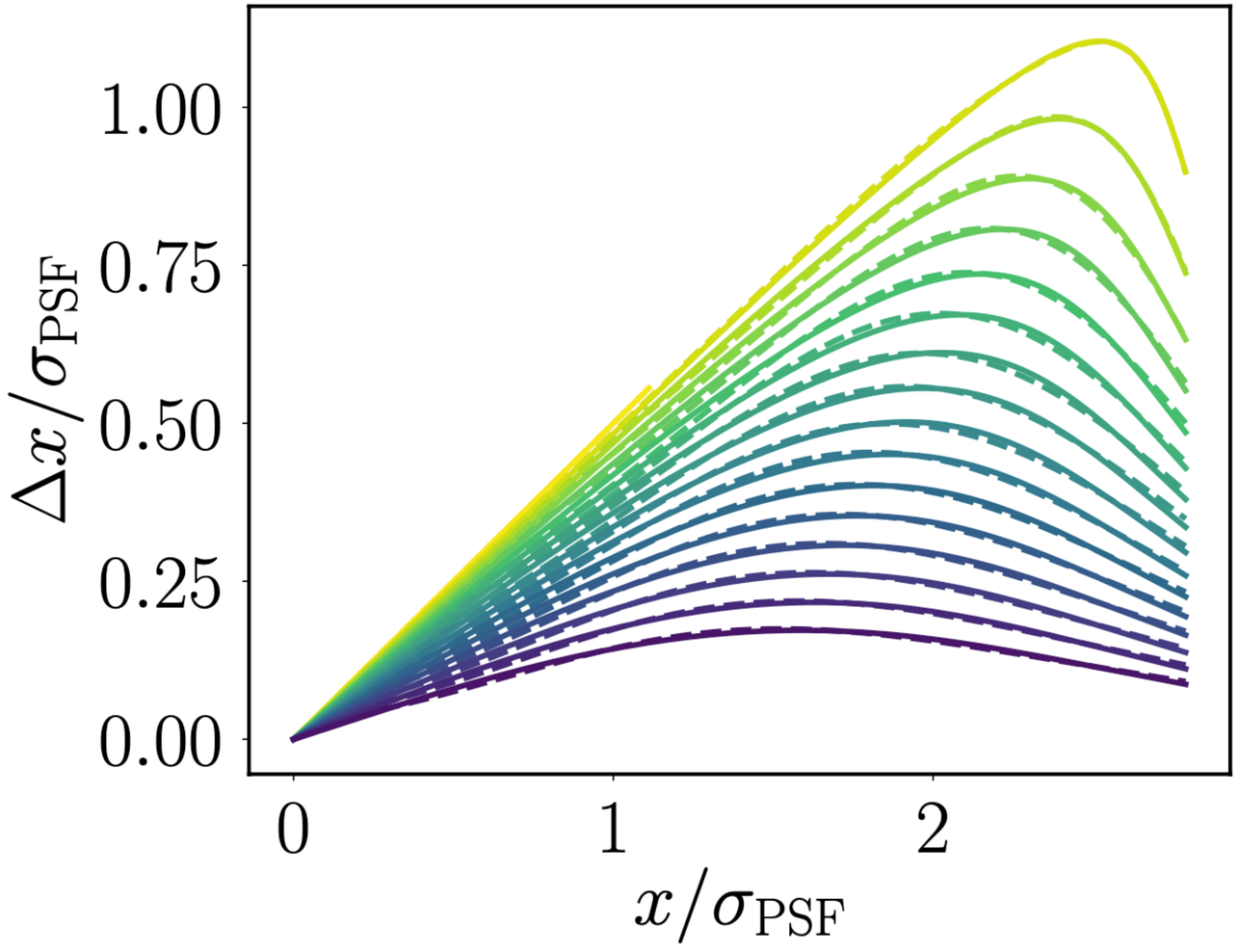
Matching Across Catalogues using the Astrometric Uncertainty Function and Flux

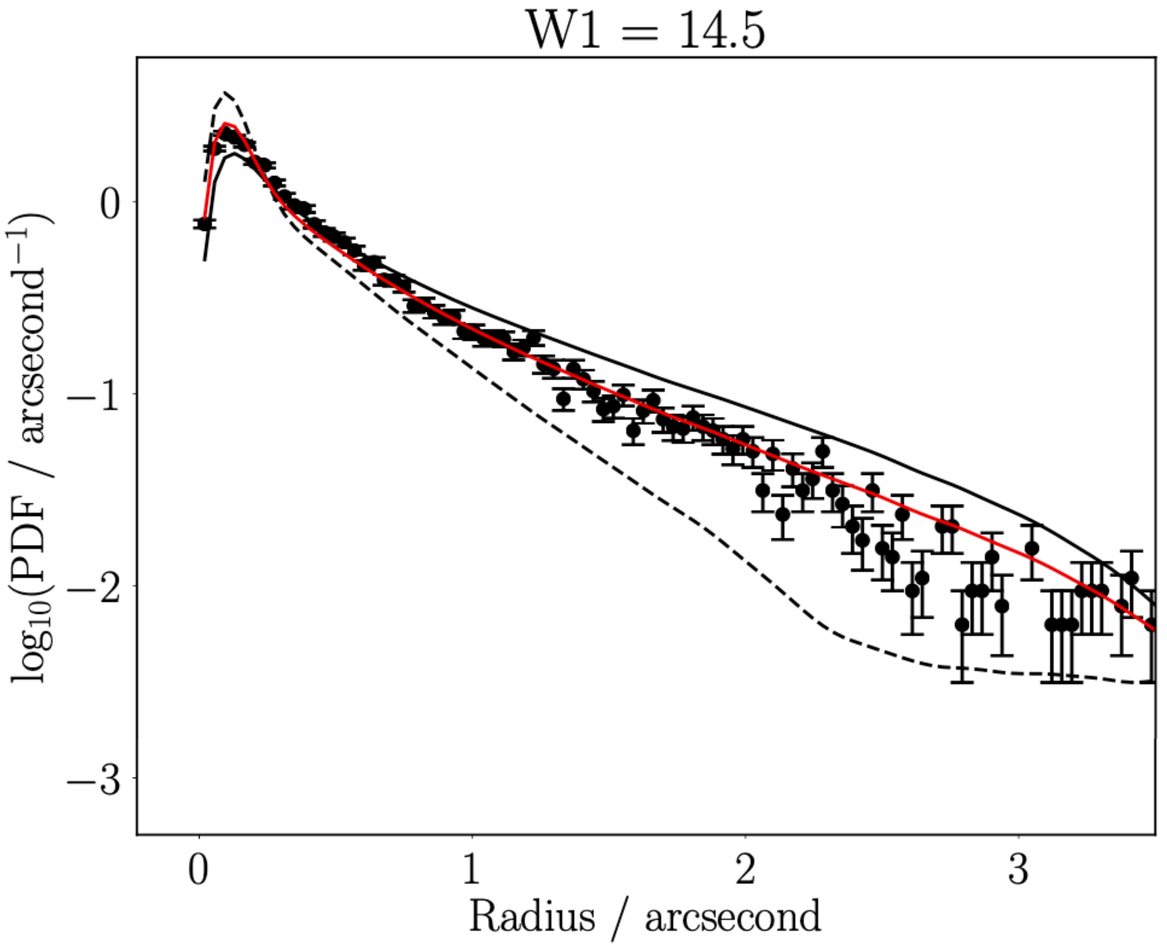
LSST:UK Consortium

(Points if you know your tartans!)

Tom J Wilson @onoddil

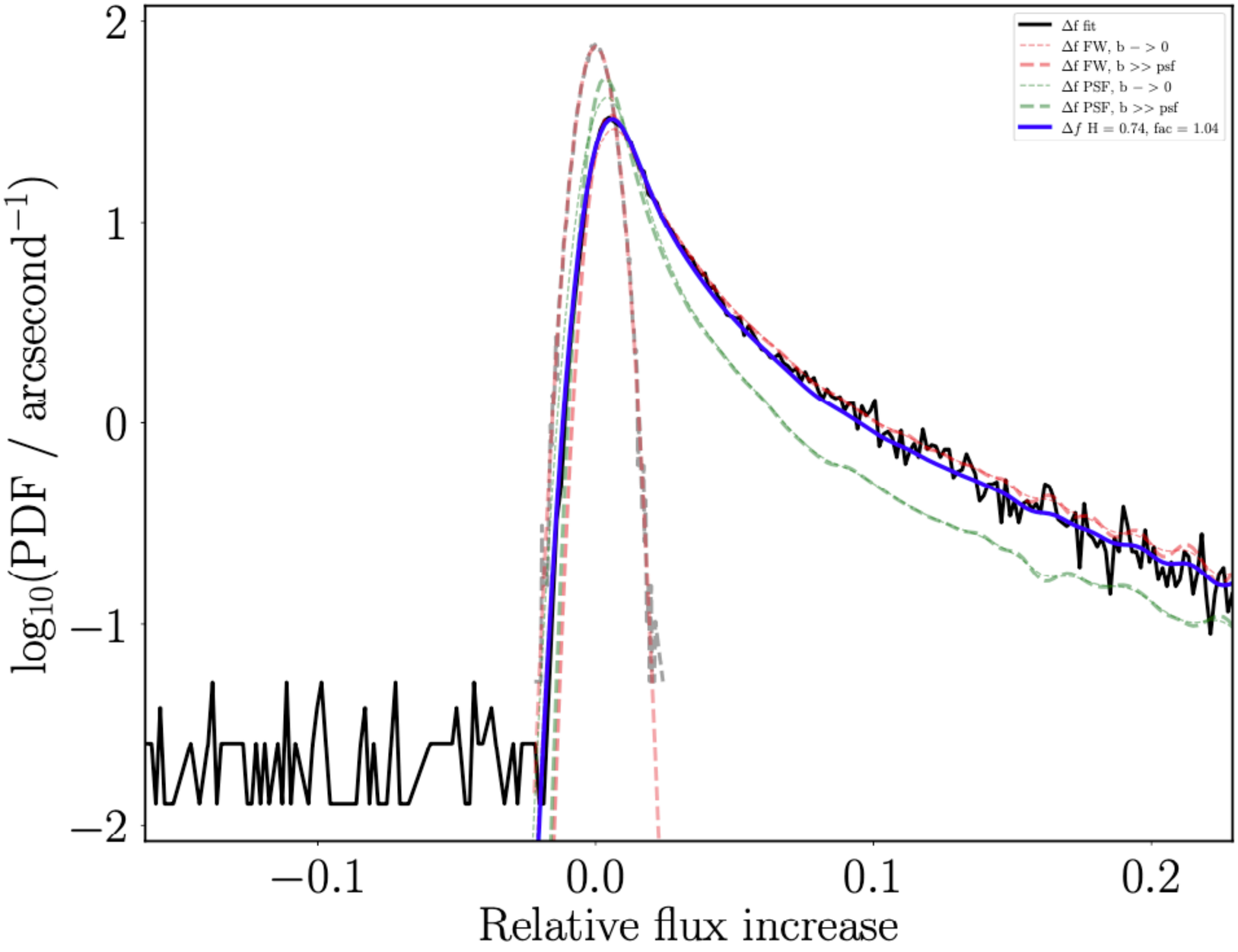
Probing the Faintest Sources





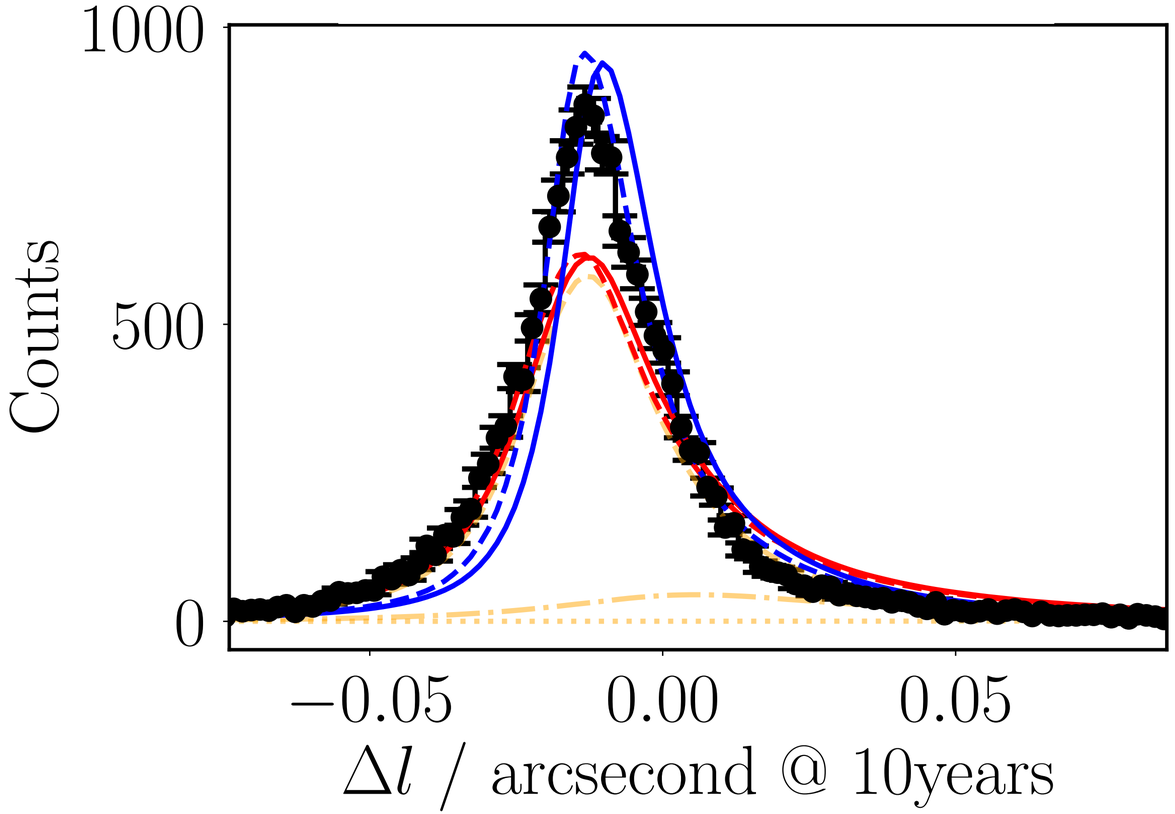
Tom J Wilson @onoddil

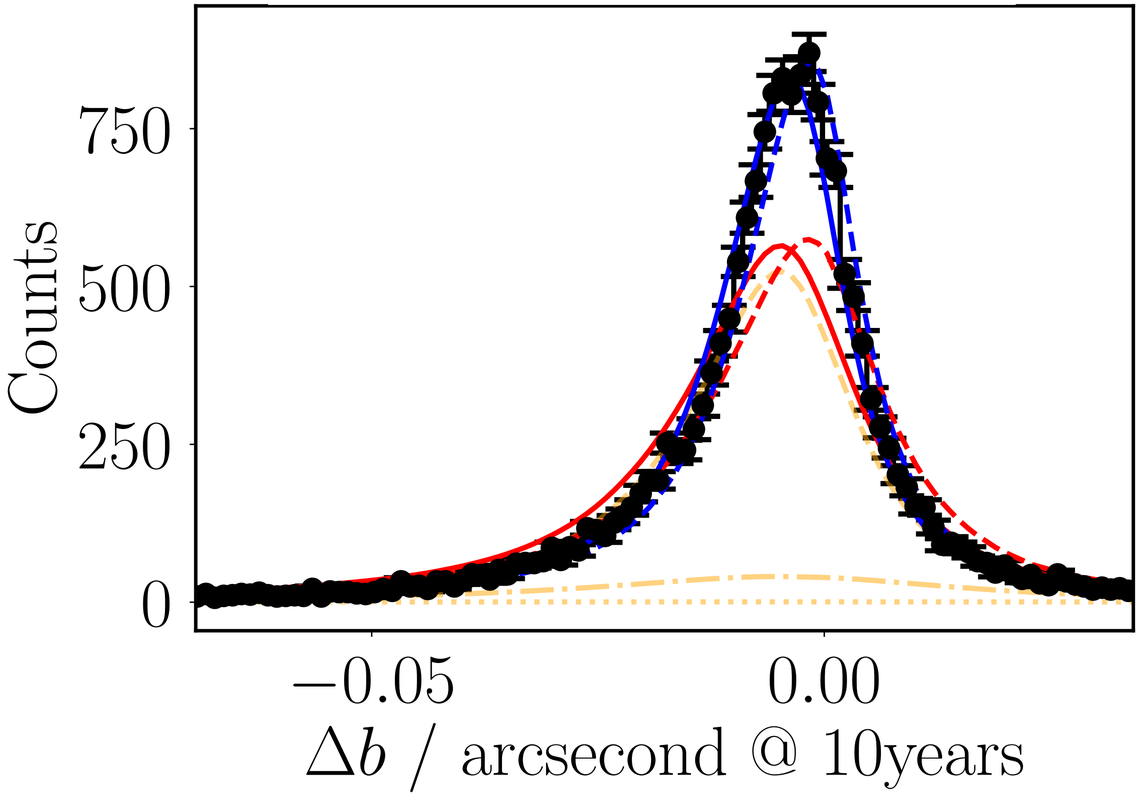
Photometric Contamination Function



Tom J Wilson @onoddil

Including Unknown Proper Motions





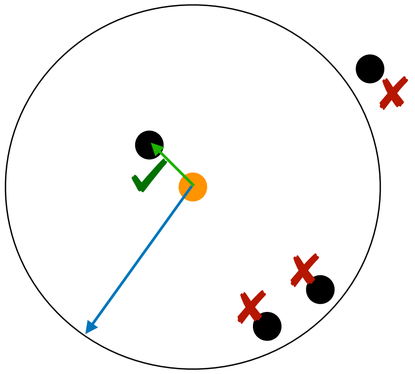
What does this mean for you?

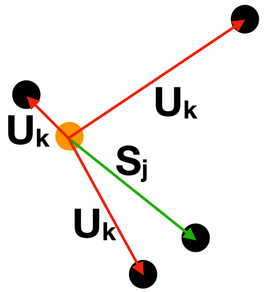
The “busy” astronomer: uses a quick and simple 2” match -> Too many matches

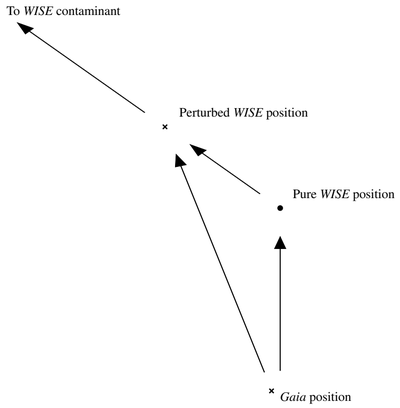
The “Bayesian” astronomer: uses astrometric centroid uncertainty to reduce match radius -> Too few matches

The “careful” astronomer: includes perturbation from blended objects in the AUF -> Correct number of matches

The “<u>smart</u>” astronomer: uses our cross-matches to get the correct number of matches *and* information on how much flux contamination is affecting their object!







you downloading your favourite cross-matches, probably

Tom J Wilson @onoddil

Conclusions

- **Blended star contamination causes positional shifts, now modelled robustly for the first time in the AUF**
- **Symmetric data-driven photometric likelihood now possible**

- ***WISE* objects are up to 30% flux contaminated**
- **LSST will suffer of order 10% flux contamination in the future**
  - **Important for extinction/distance; “1% photometry”?**
- **Can include unknown proper motions easily in AUF**
- **Modelling of statistical flux contamination allows for the recovery of “true” fluxes**
- **High dynamic range matches must account for differential crowding matching to ancillary or historic data**

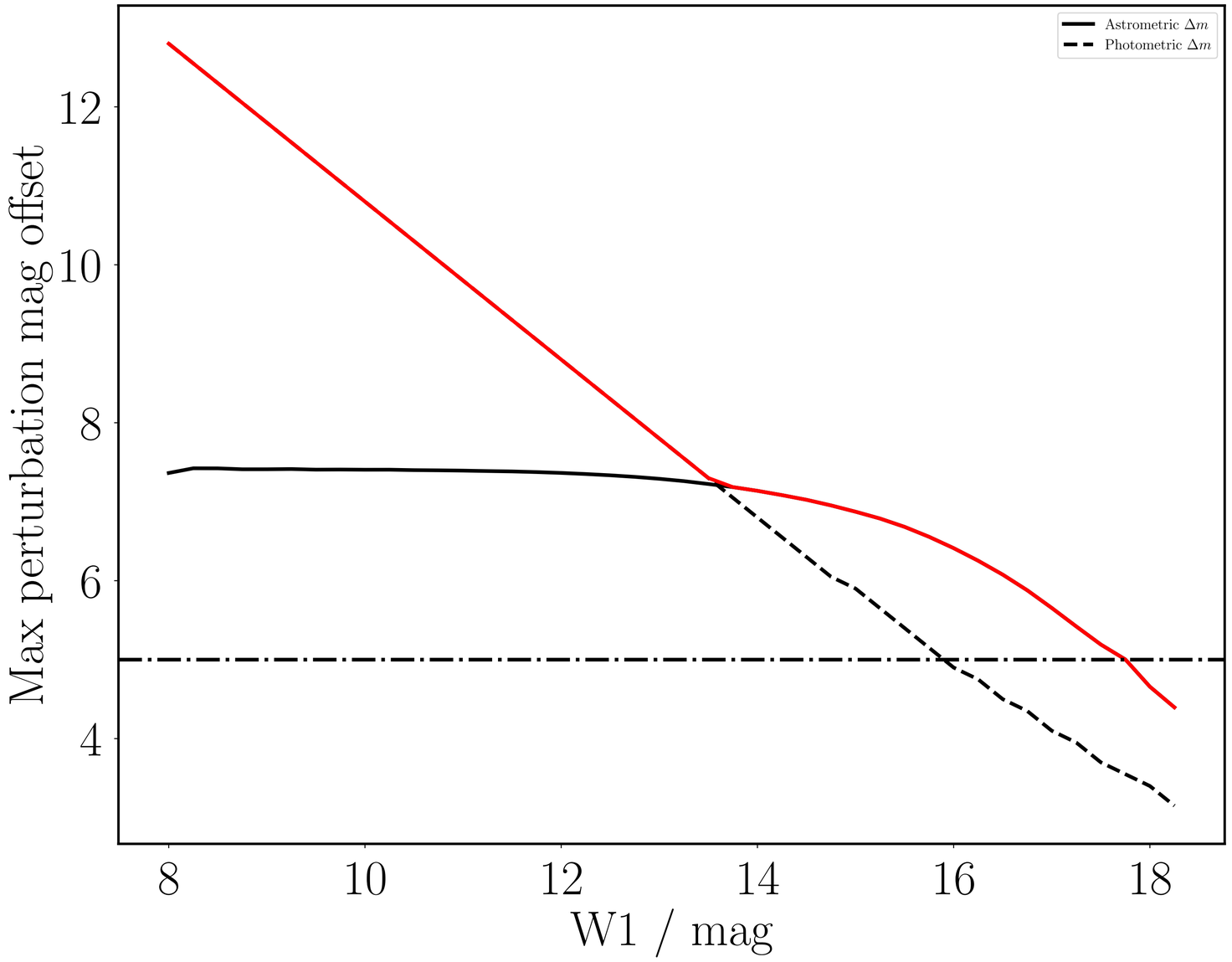
- **Upcoming LSST:UK cross-match service macauff — let me know your thoughts/needs/hopes/dreams**

Wilson & Naylor, 2017, MNRAS, 468, 2517
Wilson & Naylor, 2018a, MNRAS, 473, 5570
Wilson & Naylor, 2018b, MNRAS, 481, 2148

https://github.com/Onoddil/macauff

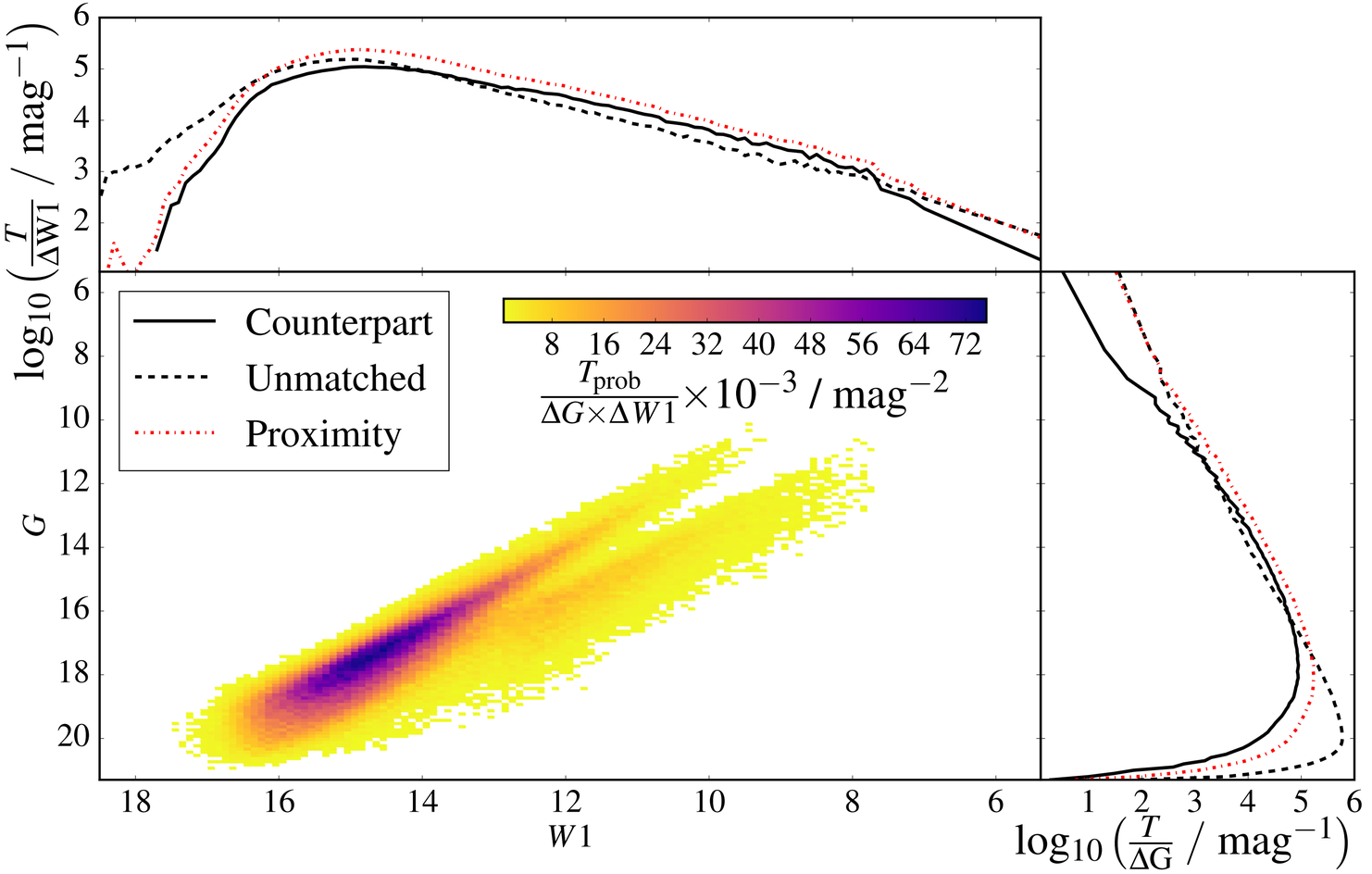
Tom J Wilson @onoddil

The Astrometric Uncertainty Function and LSST: A Crisis of Completeness Limit

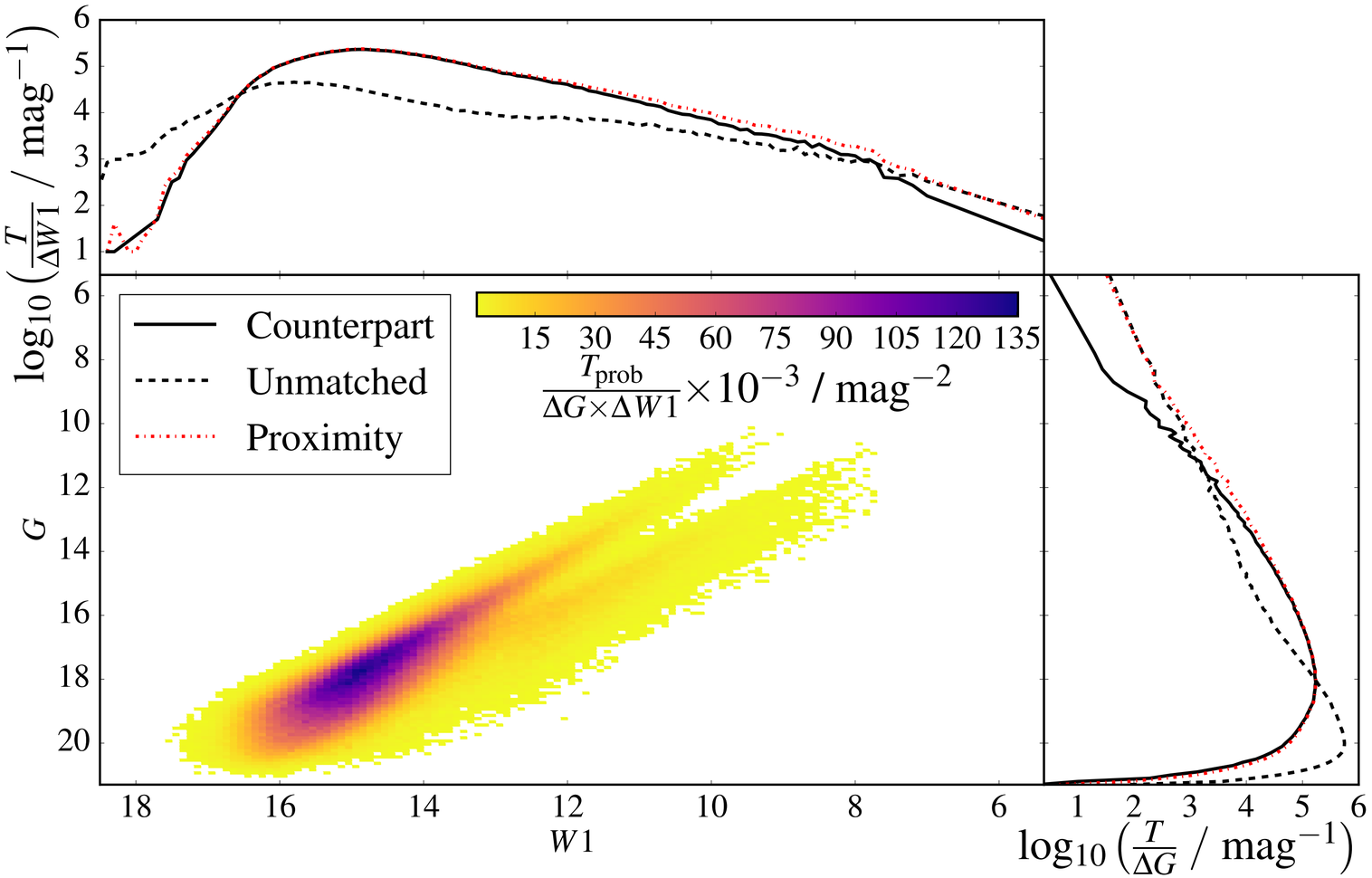


Tom J Wilson @onoddil

Contamination Effects: Gaia-WISE Gaussian Matches



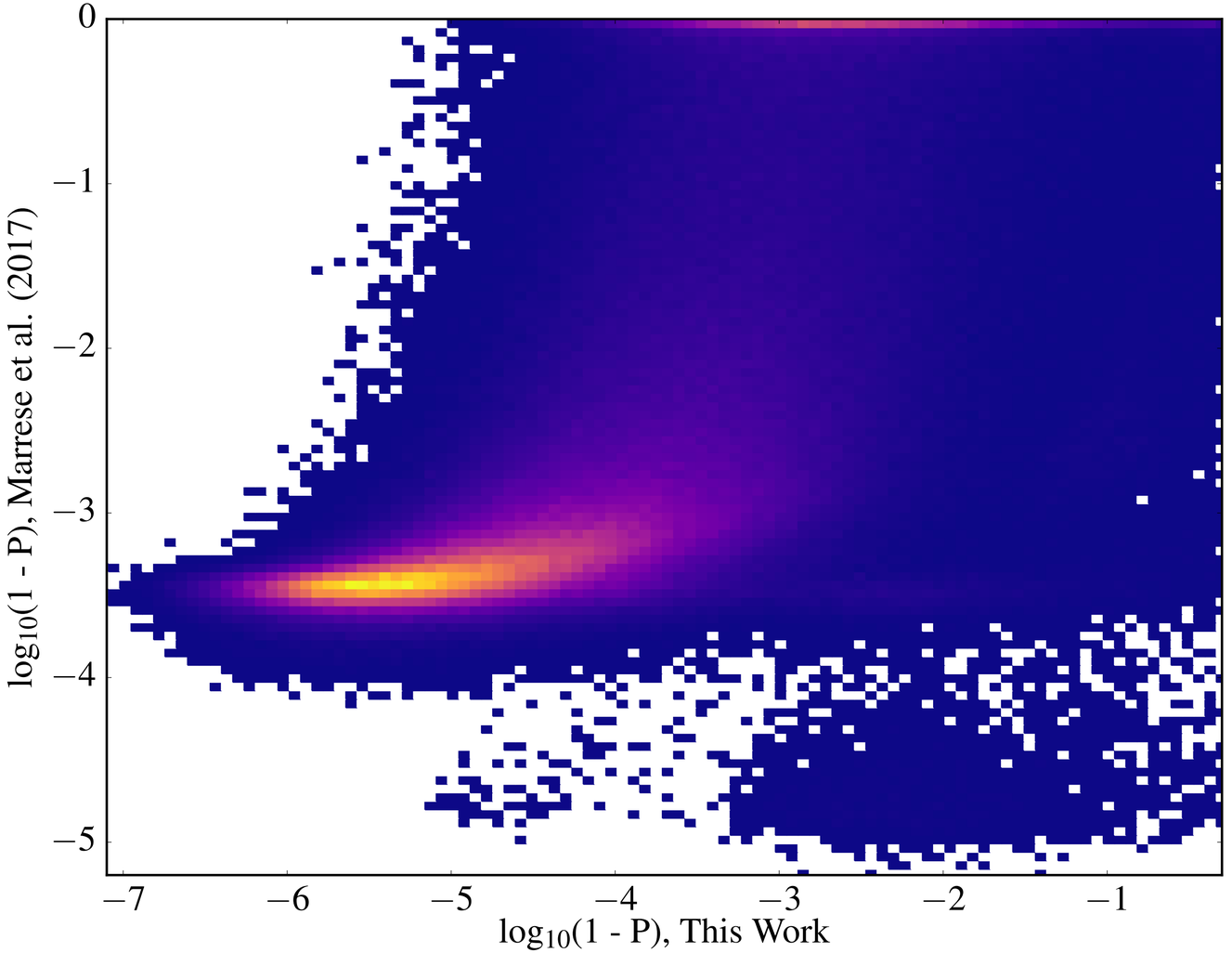
Contamination Effects: Gaia-WISE Empirical Matches



Wilson & Naylor, MNRAS, 2018b, 481, 2148

Tom J Wilson @onoddil

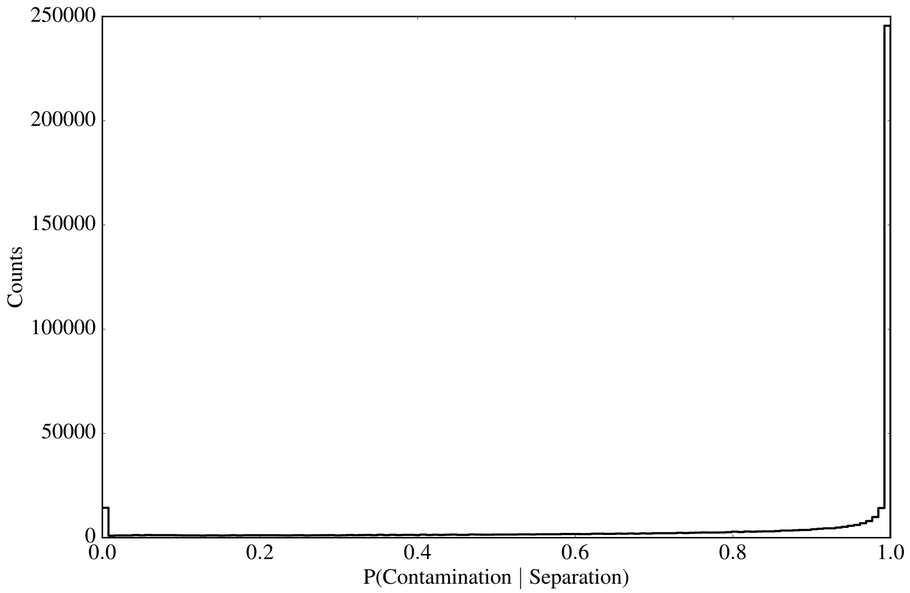
Contamination Effects: Gaia Lost Matches

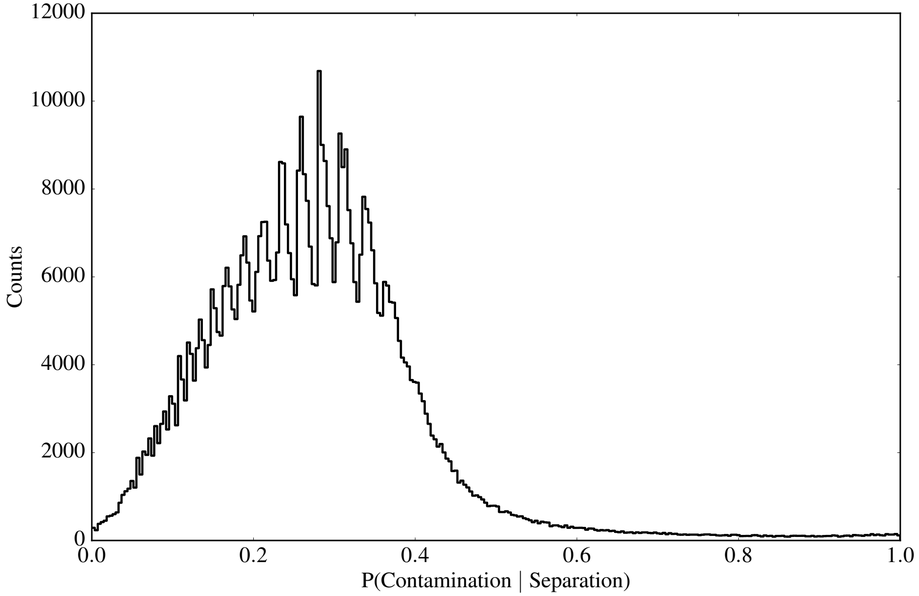


Wilson & Naylor, MNRAS, 2018b, 481, 2148

Tom J Wilson @onoddil

Photometric Contamination: WISE/Spitzer Contamination %



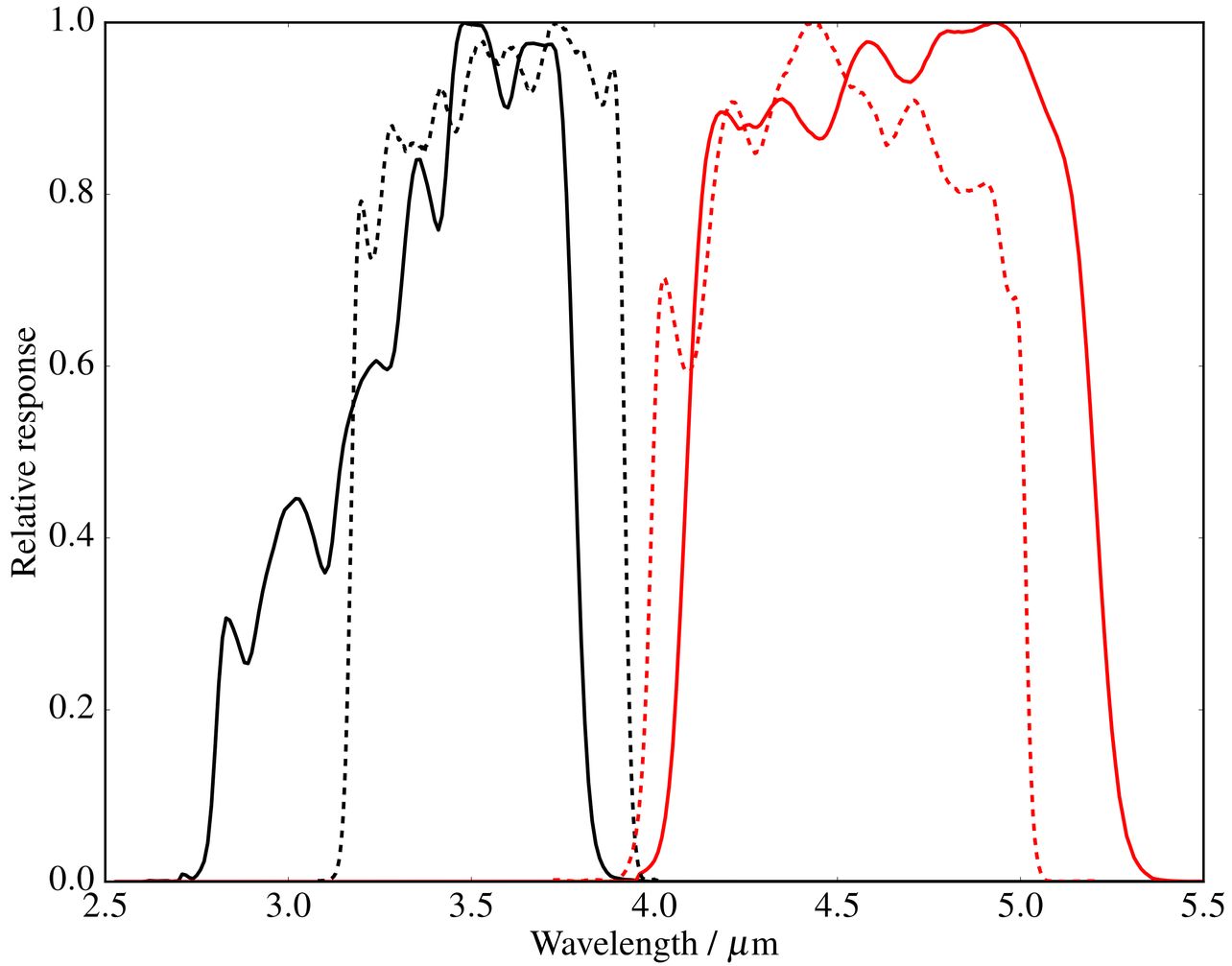


UNIVERSITY OF EXETER

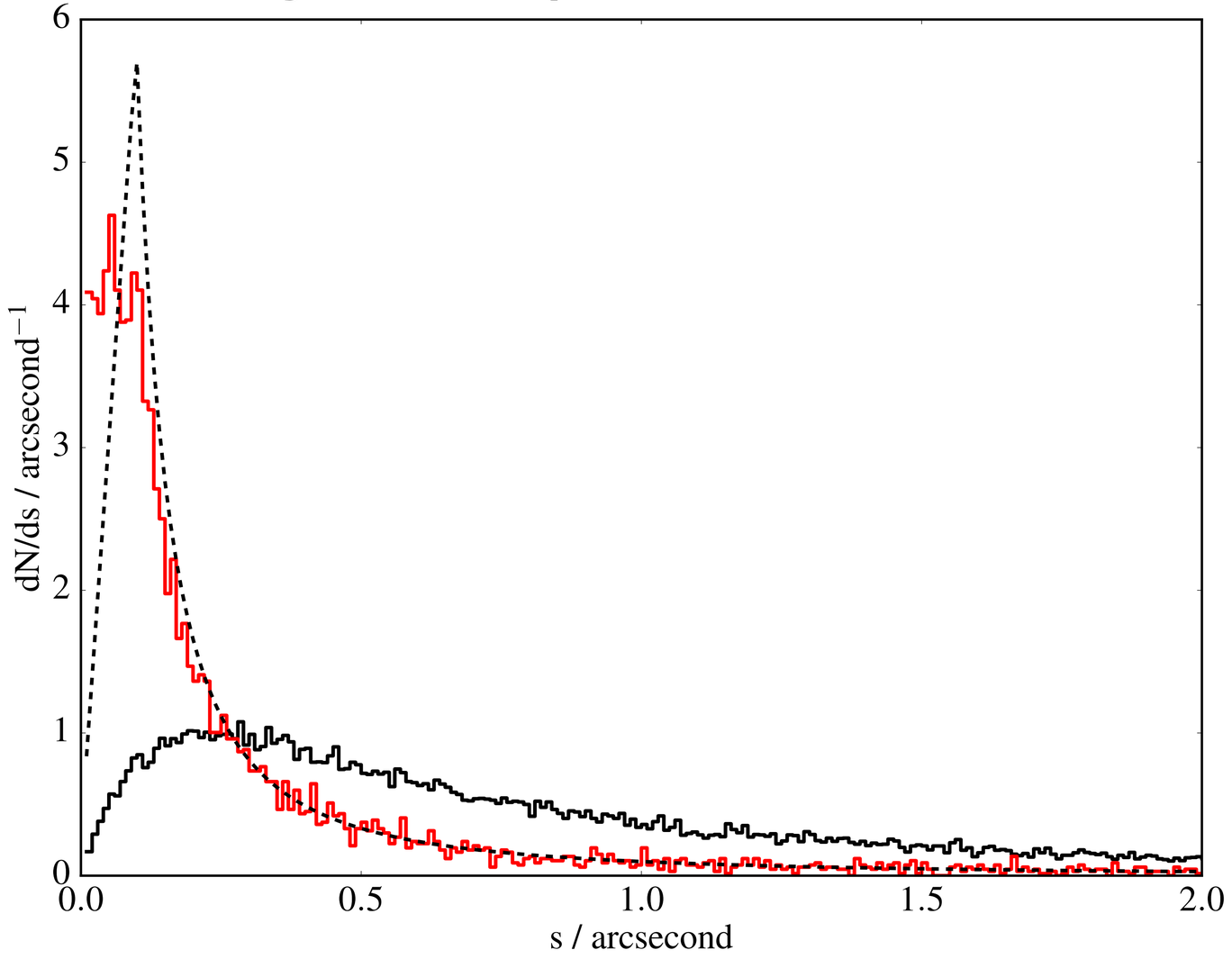
Wilson T., Naylor T., in prep…

Tom Wilson @onoastrmer

Contamination Effects: Wavelength Coverage



The Astrometric Uncertainty Function: Analytical perturbations



Wilson & Naylor, 2017, MNRAS, 468, 2517

Tom J Wilson @onoddil