Identifying correct counterparts to high-energy sources by "multiwavelength educated guesses" imbibed in a Bayesian statistic environment

a.k.a. NWAY (Salvato et al 2018, 2022)

M. Salvato (MPE)

w/ :J. Buchner, J.Wolf, T. Dwelly, A. Georgakakis, T. Budavari, and more

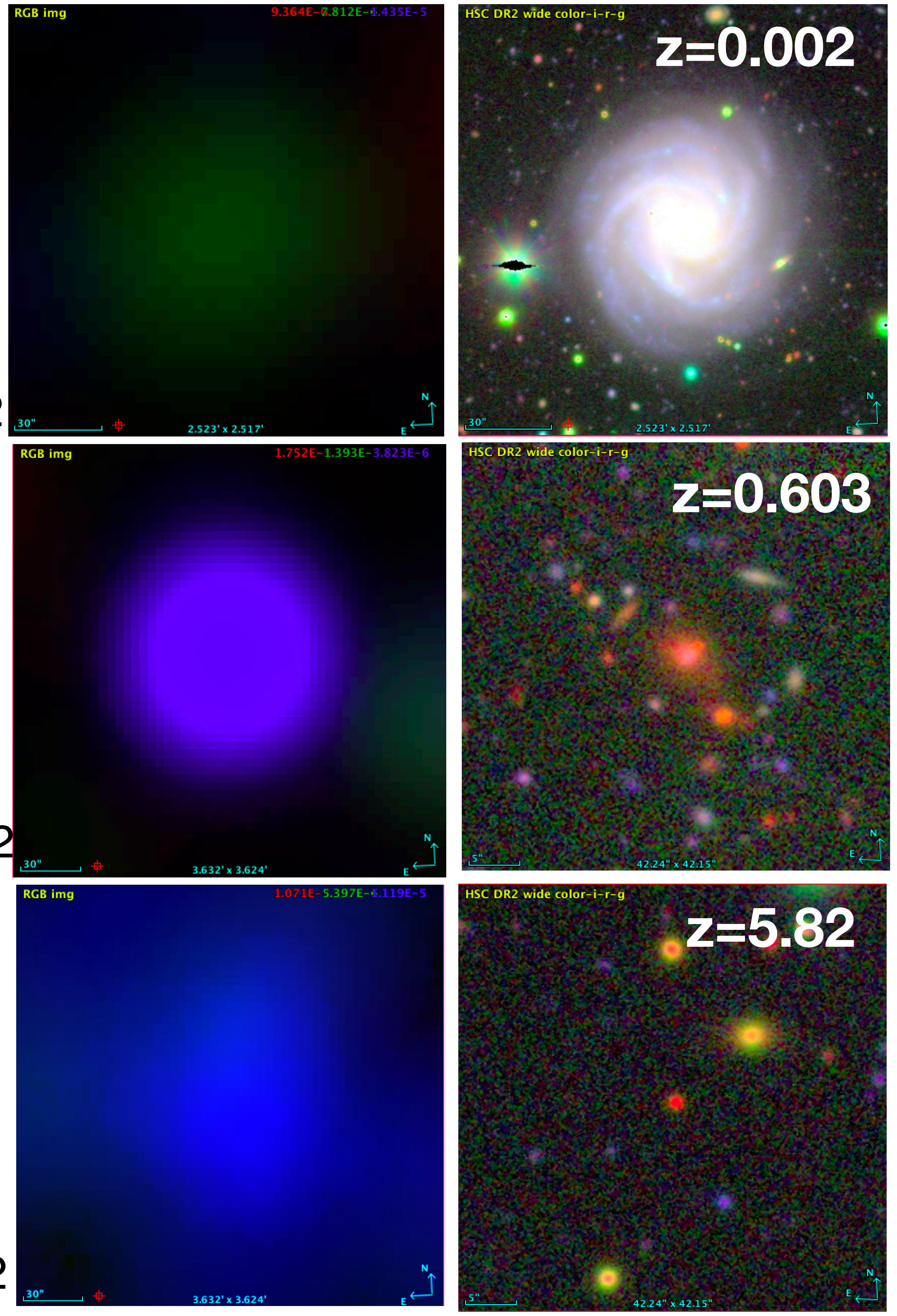
Outline

- **Motivations**
- **Issues**
- **A/The Solution**
- **Comparison between methods**
- **Identification & classification**
- **Final remarks**

CTP means “counterpart” in this talk

The challenge: find the right counterpart

	eROSITA/eFEDS	HSC
Vulic et al 2022		z=0.002
Brusa et al 2022		z=0.603
Wolf et al 2022		z=5.82



…but it could be also a Galactic source (e.g., Schneider et al 2022, Stelze et al 2022)
or an unresolved cluster (e.g., Bulbul et al 2022)

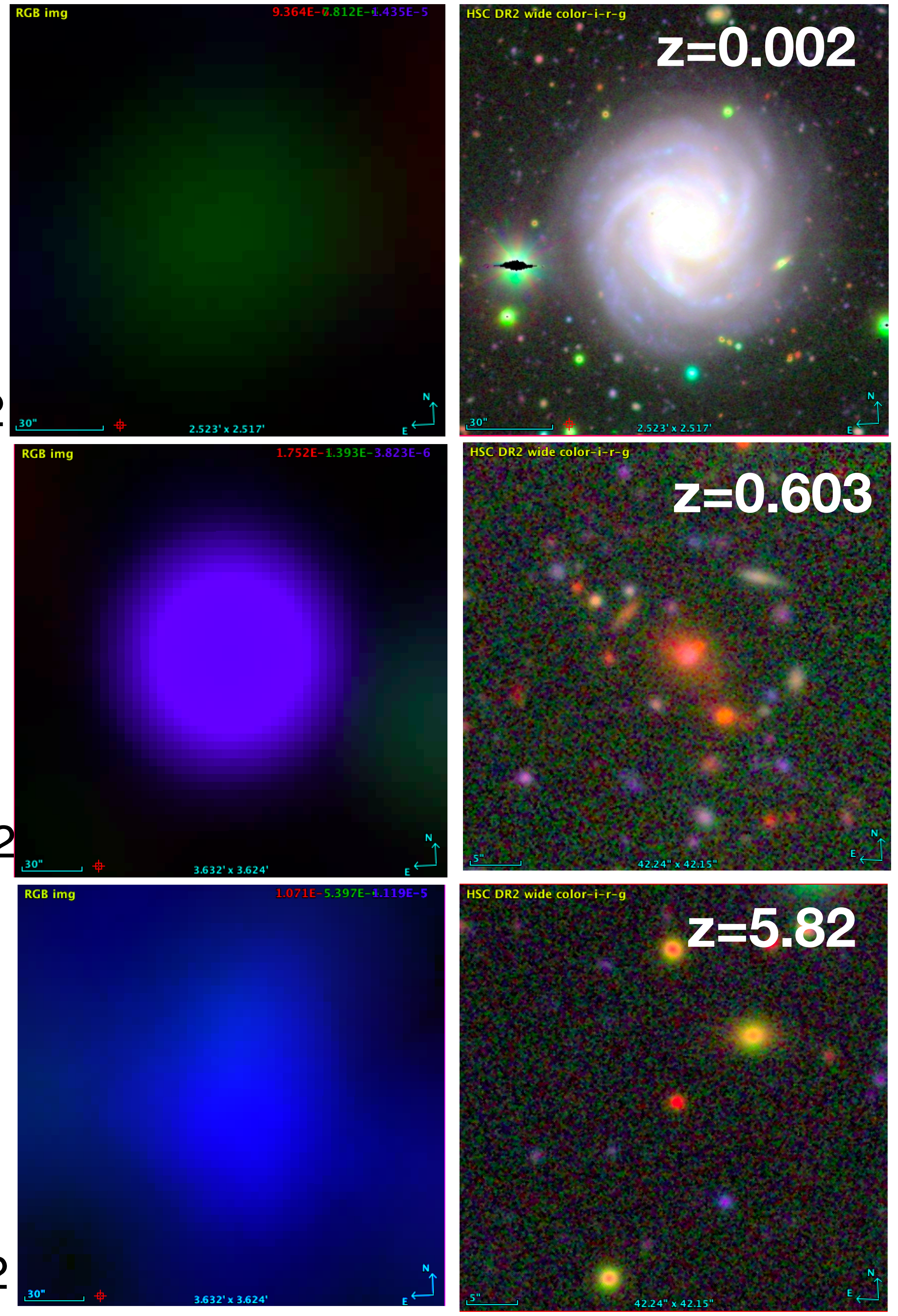
The challenge: find the right counterpart

eROSITA/eFEDS HSC

Vulic et al 2022

Brusa et al 2022

Wolf et al 2022



…but it could be also a Galactic source (e.g., Schneider et al 2022, Stelze et al 2022) or an unresolved cluster (e.g., Bulbul et al 2022)

- The size of the area search depends on the positional error which correlates with the intensity of the source.

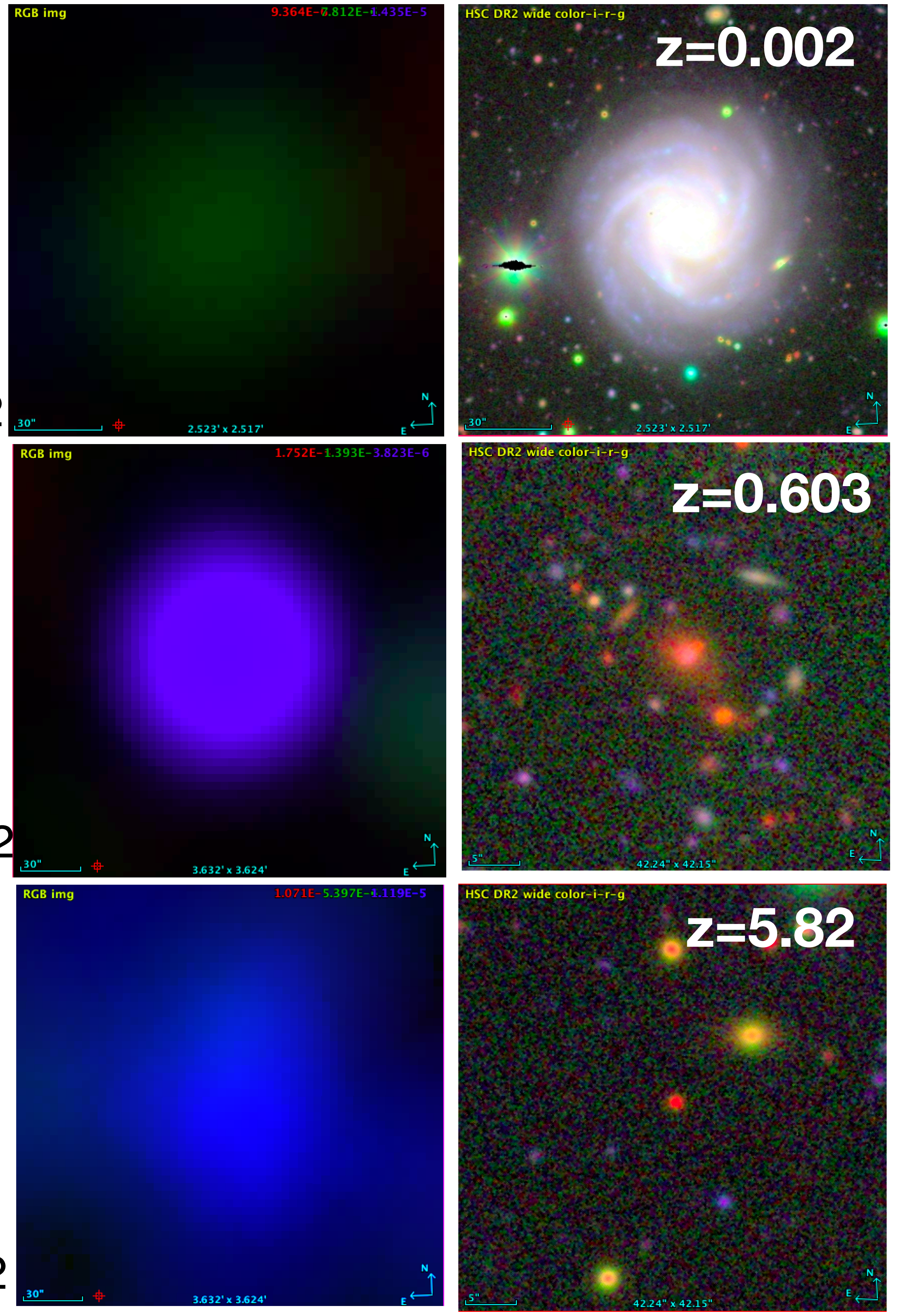
The challenge: find the right counterpart

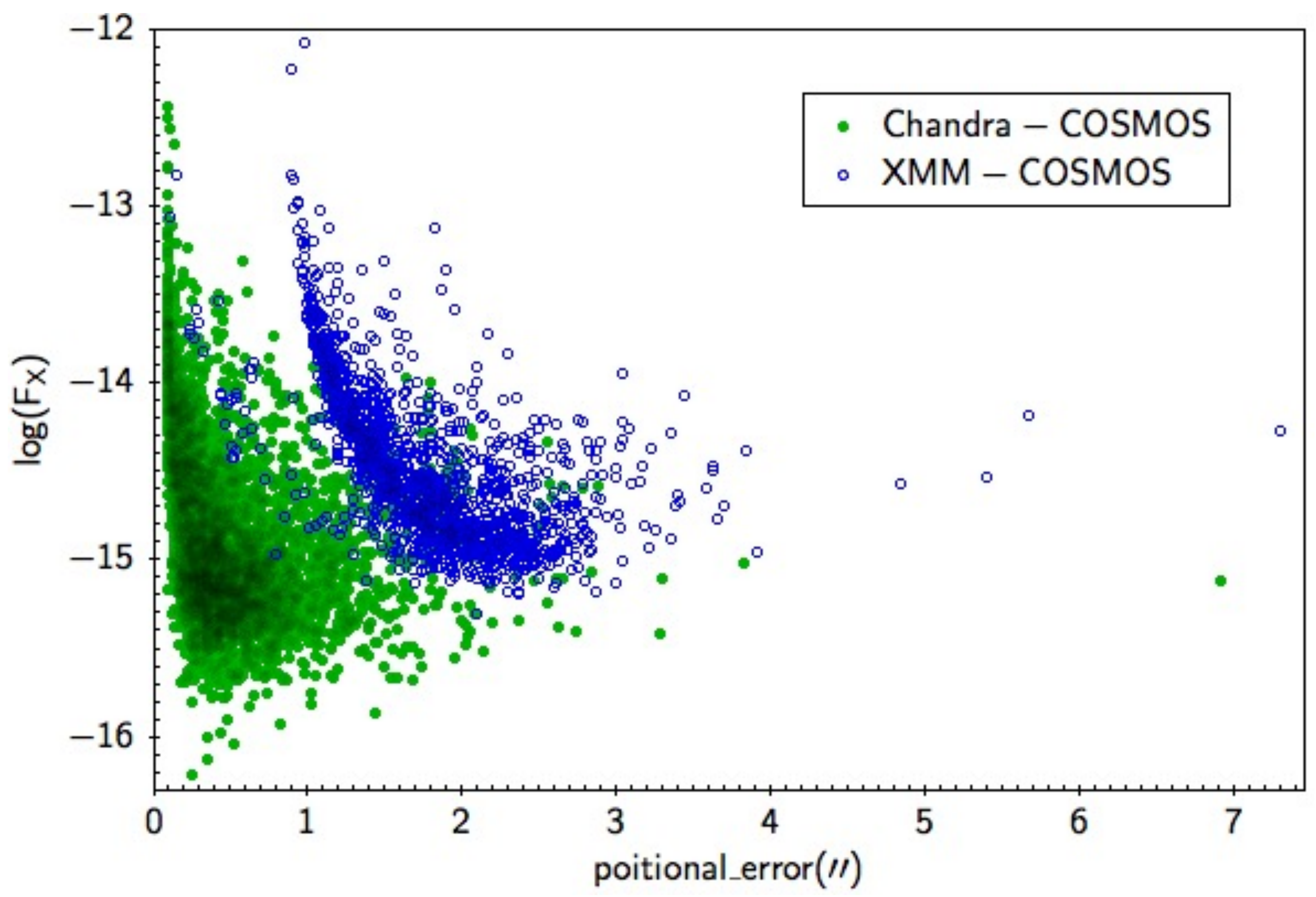
eROSITA/eFEDS HSC

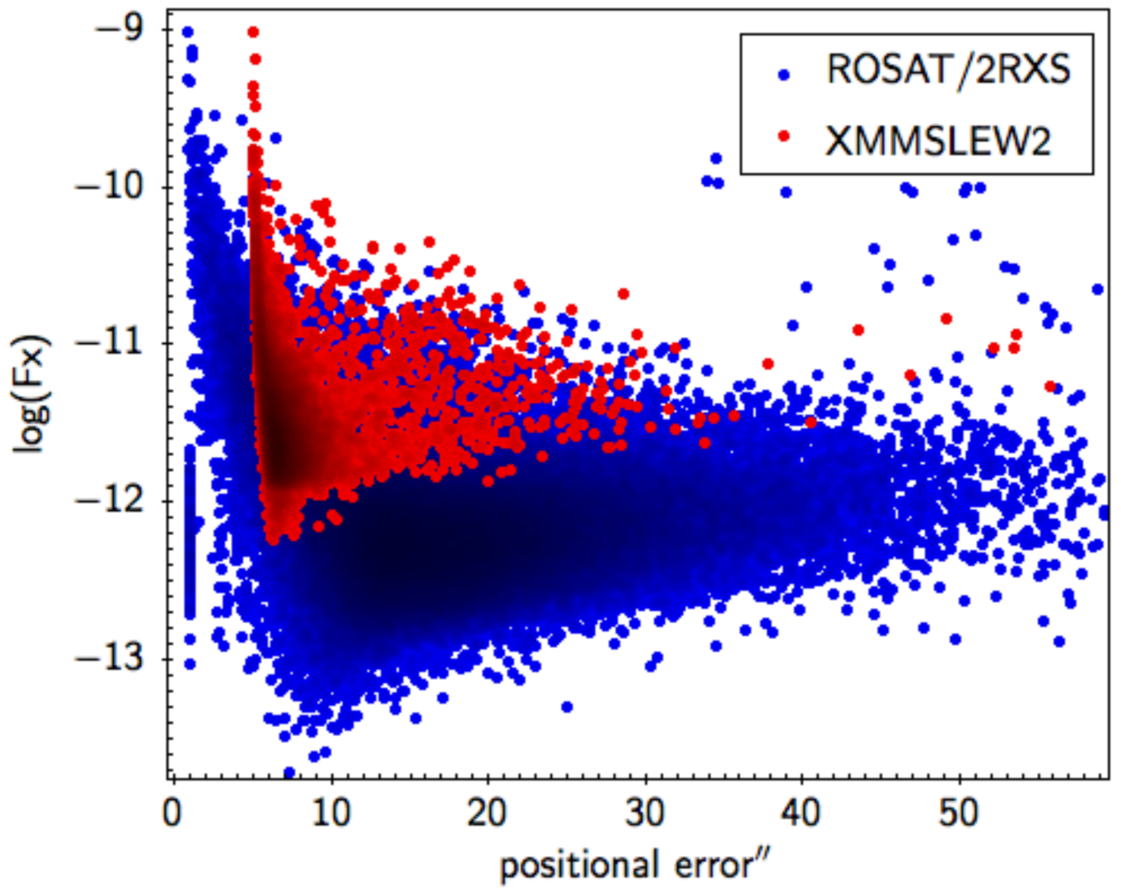
Vulic et al 2022

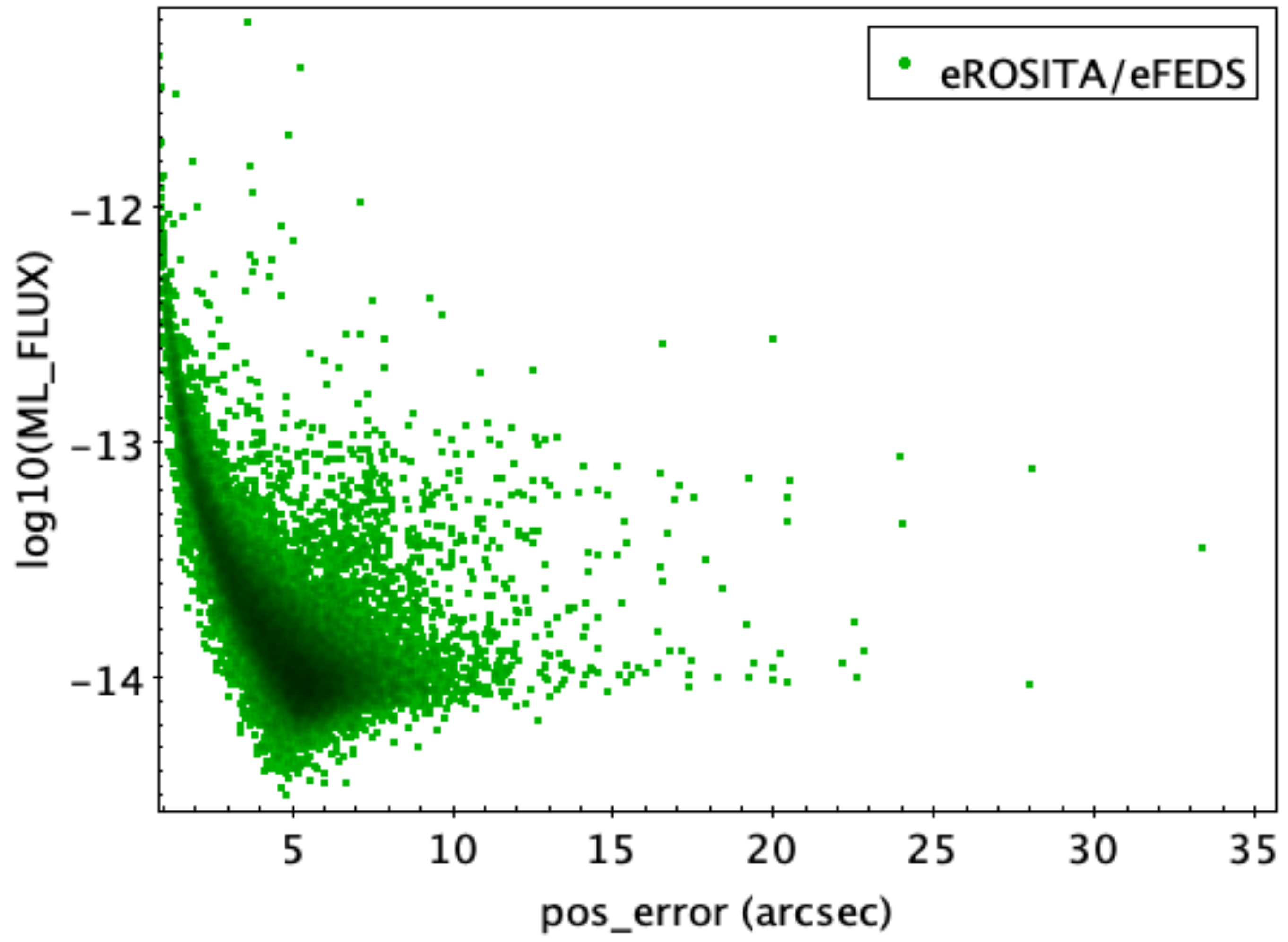
Brusa et al 2022

Wolf et al 2022









- The size of the area search depends on the positional error which correlates with the intensity of the source.

…but it could be also a Galactic source (e.g., Schneider et al 2022, Stelze et al 2022) or an unresolved cluster (e.g., Bulbul et al 2022)

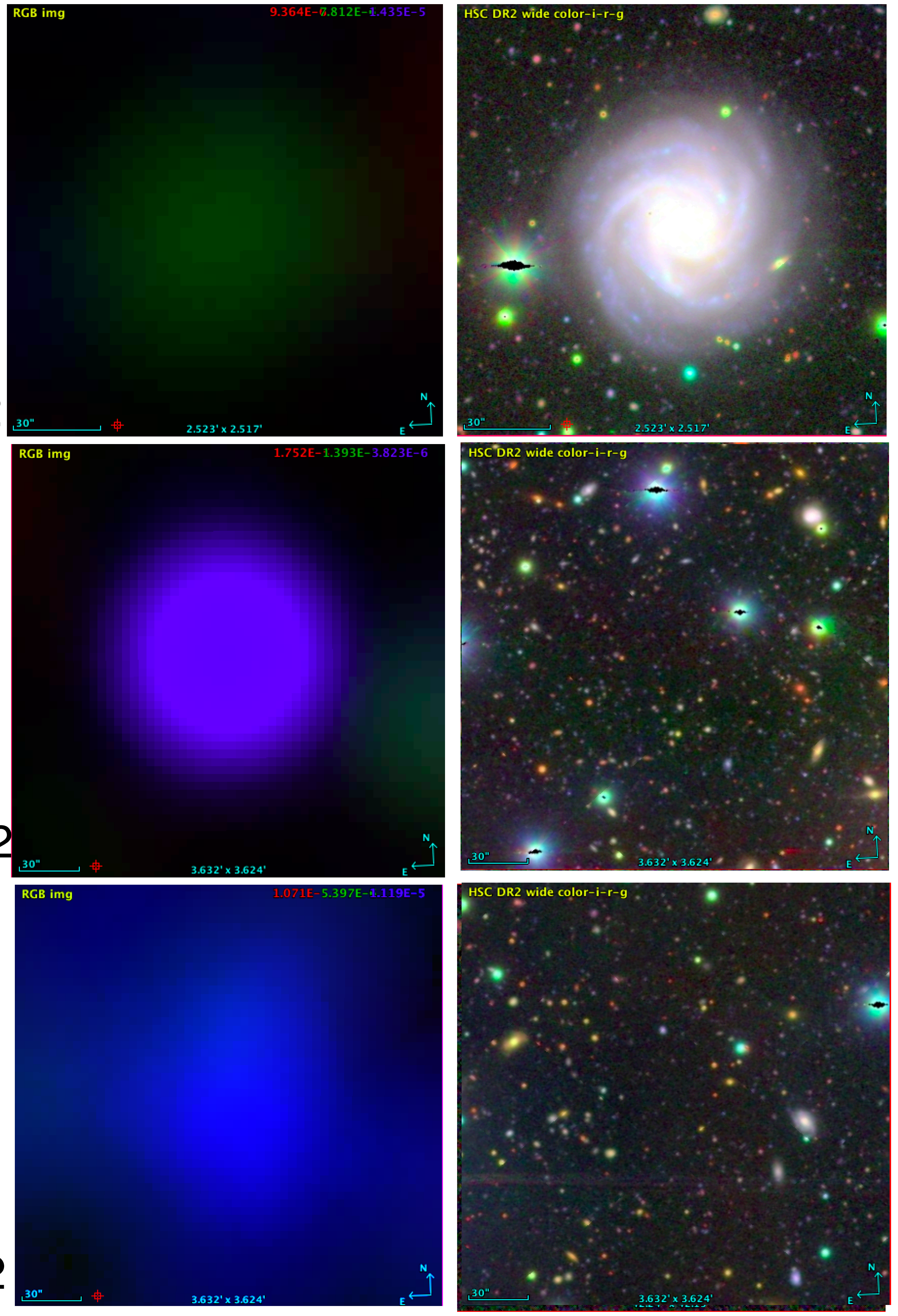
The challenge: find the right counterpart

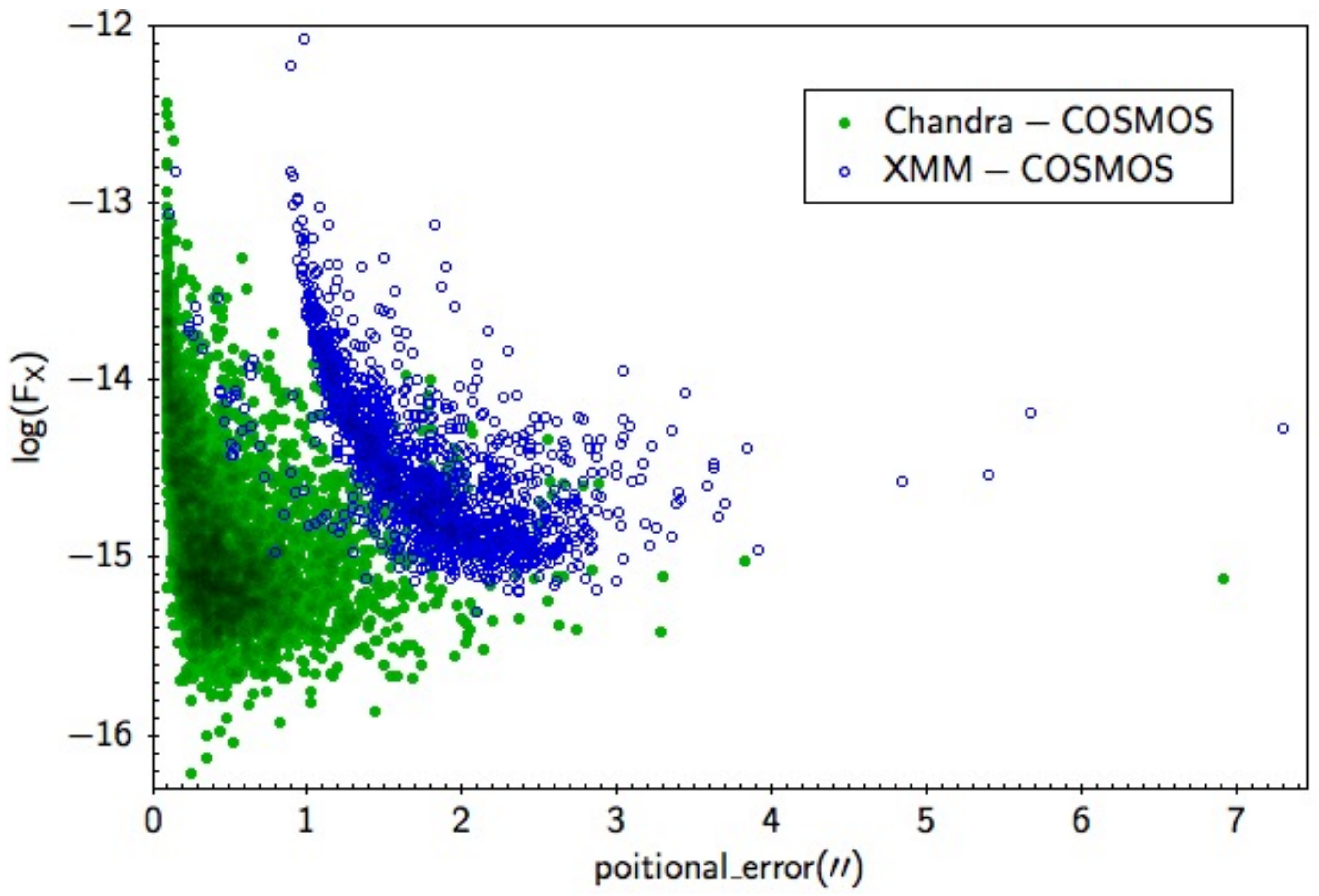
eROSITA/eFEDS | HSC

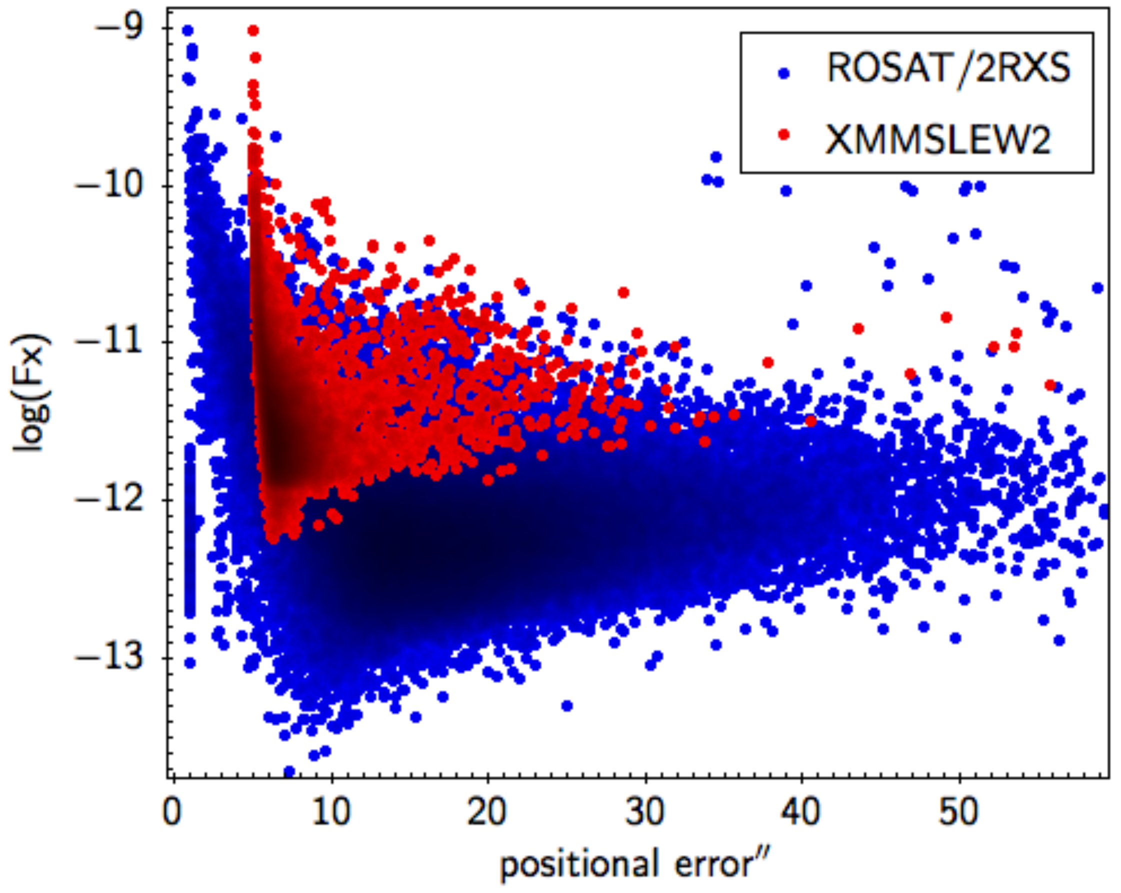
Vulic et al 2022

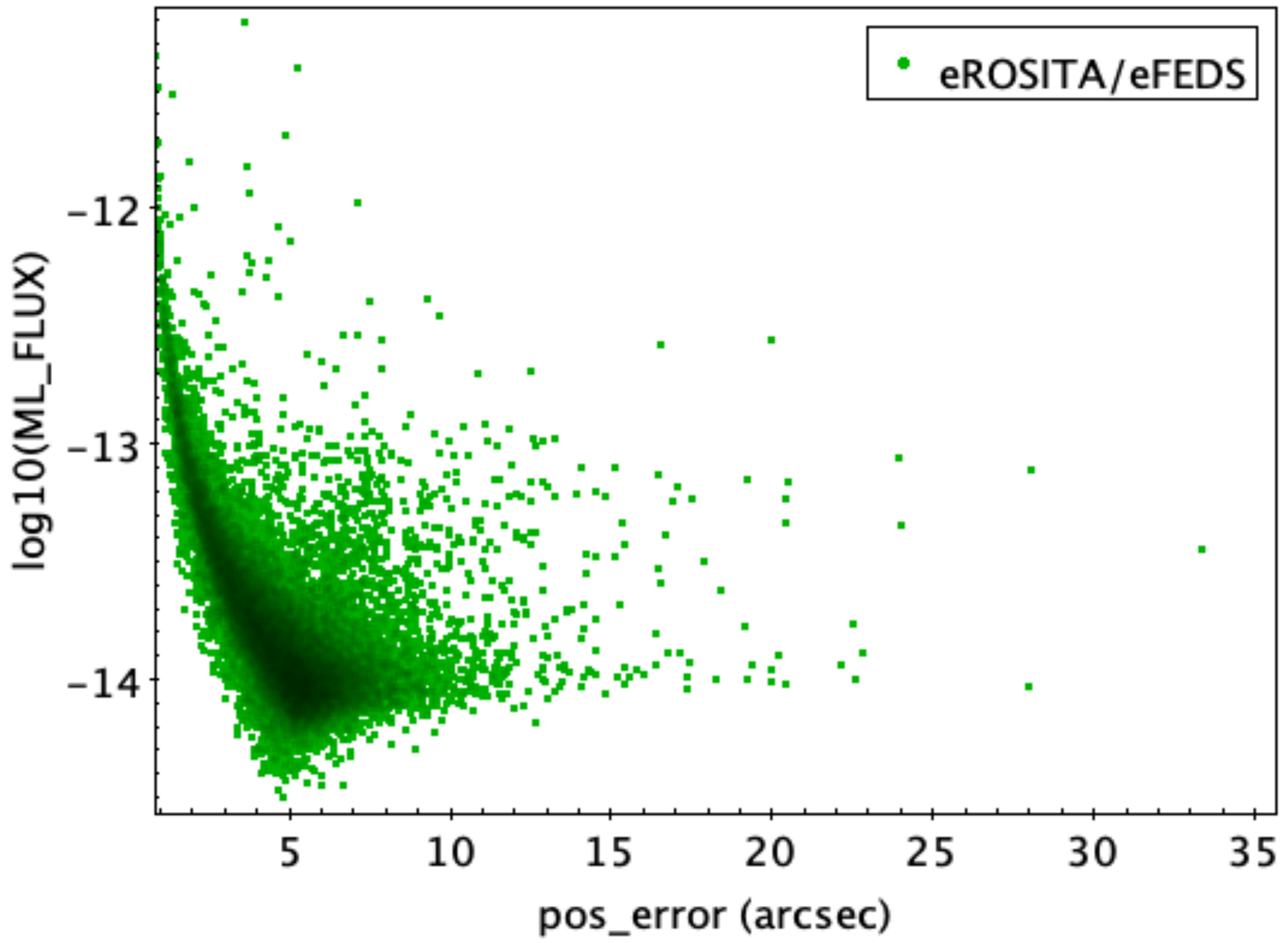
Brusa et al 2022

Wolf et al 2022









- The size of the area search depends on the positional error which correlates with the intensity of the source.

…but it could be also a Galactic source (e.g., Schneider et al 2022, Stelze et al 2022) or an unresolved cluster (e.g., Bulbul et al 2022)

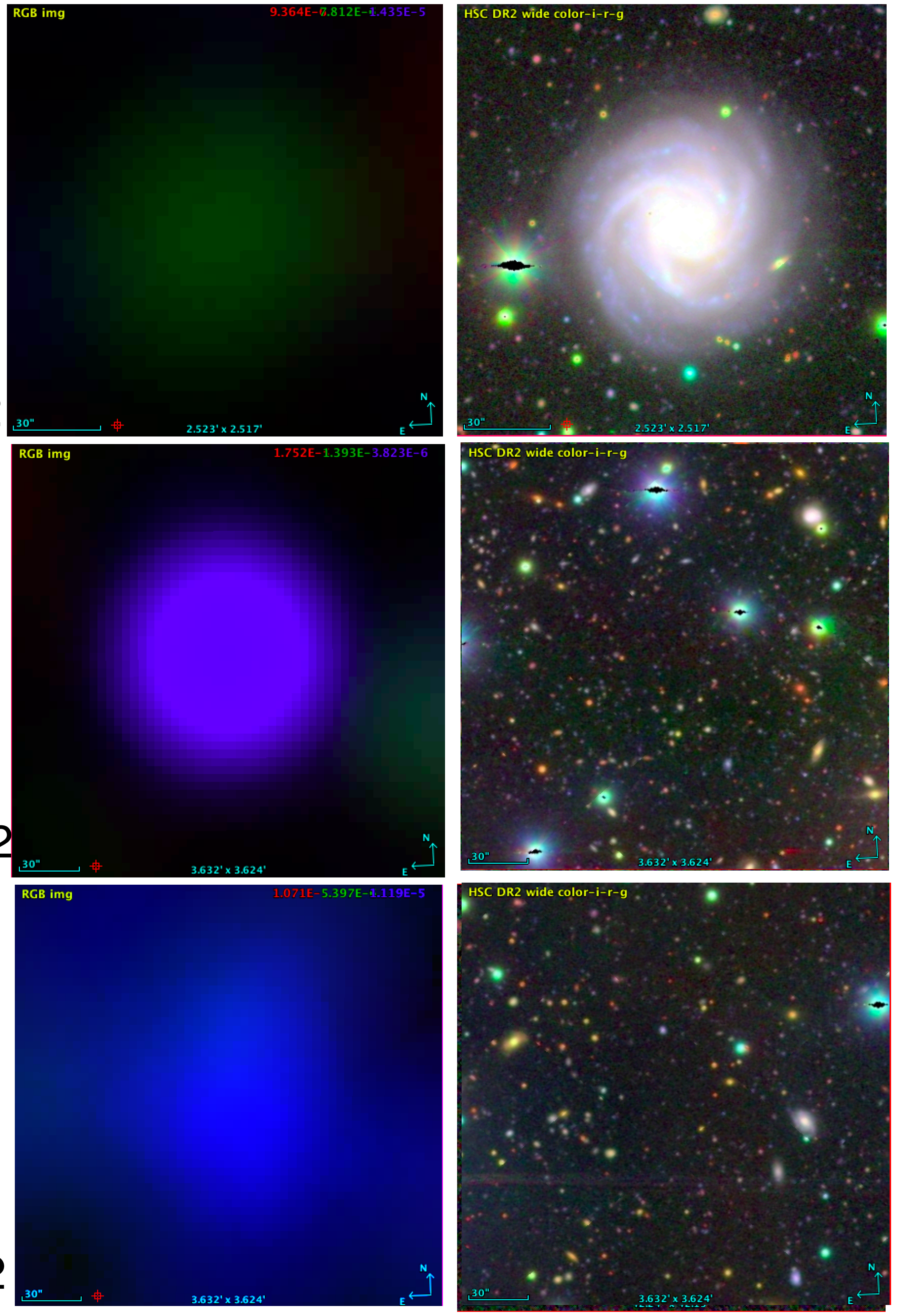
The challenge: find the right counterpart

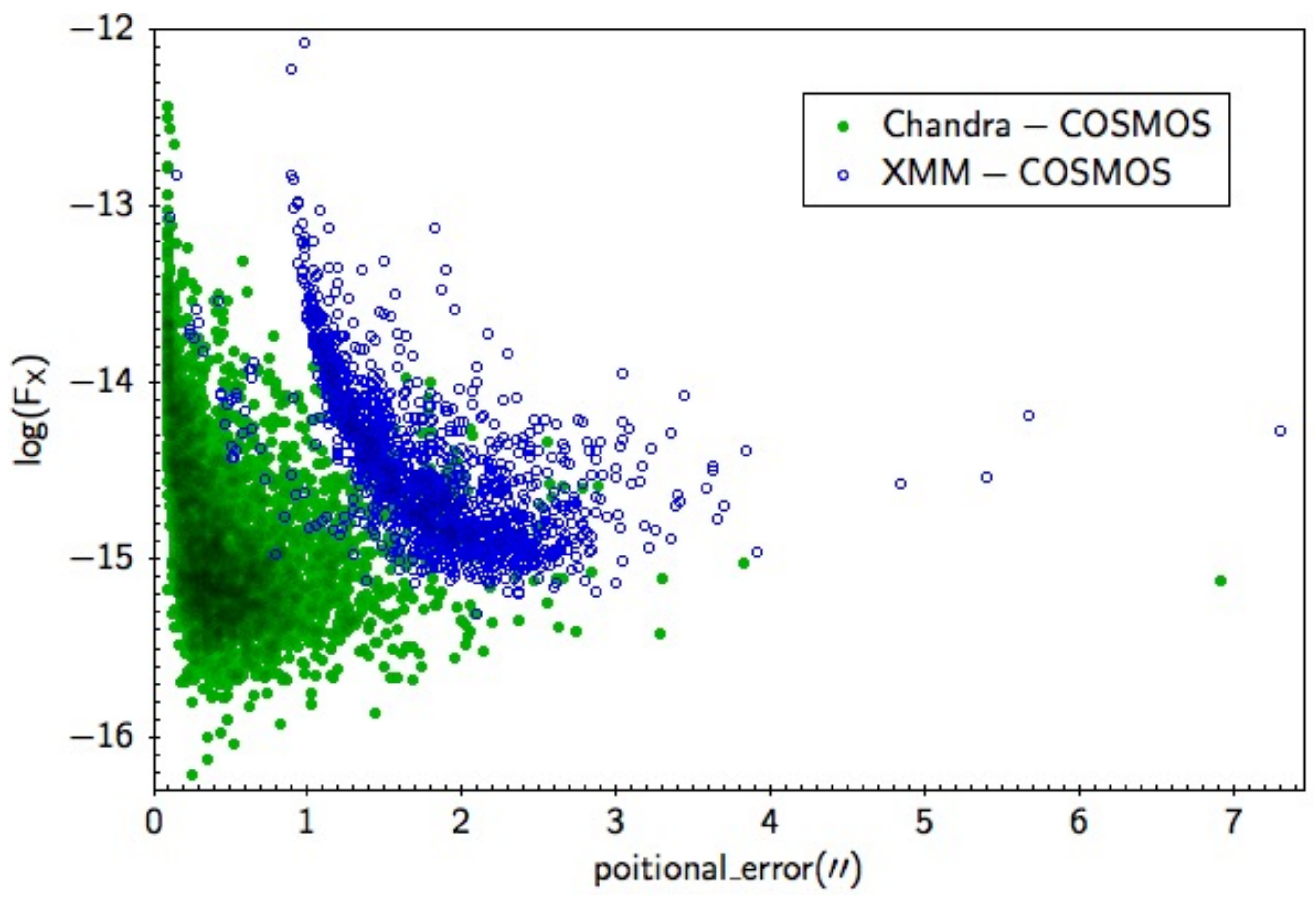
eROSITA/eFEDS HSC

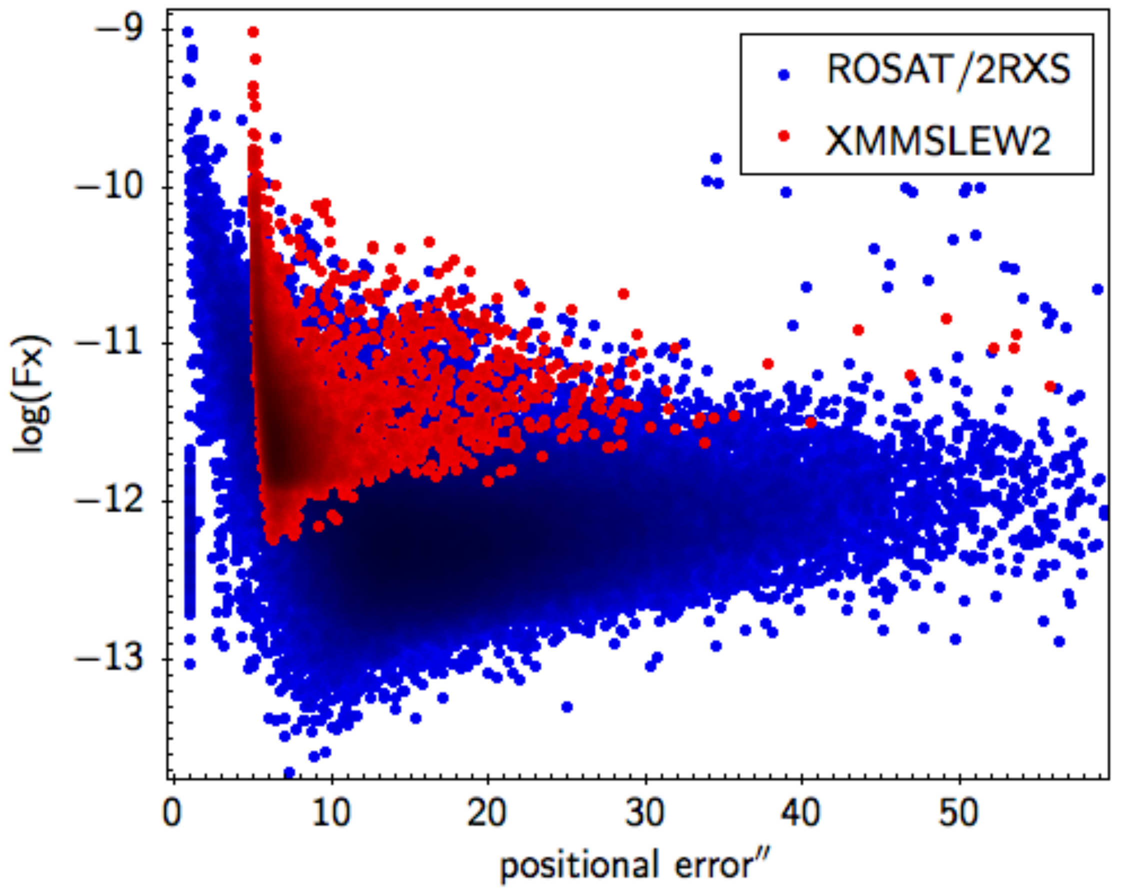
Vulic et al 2022

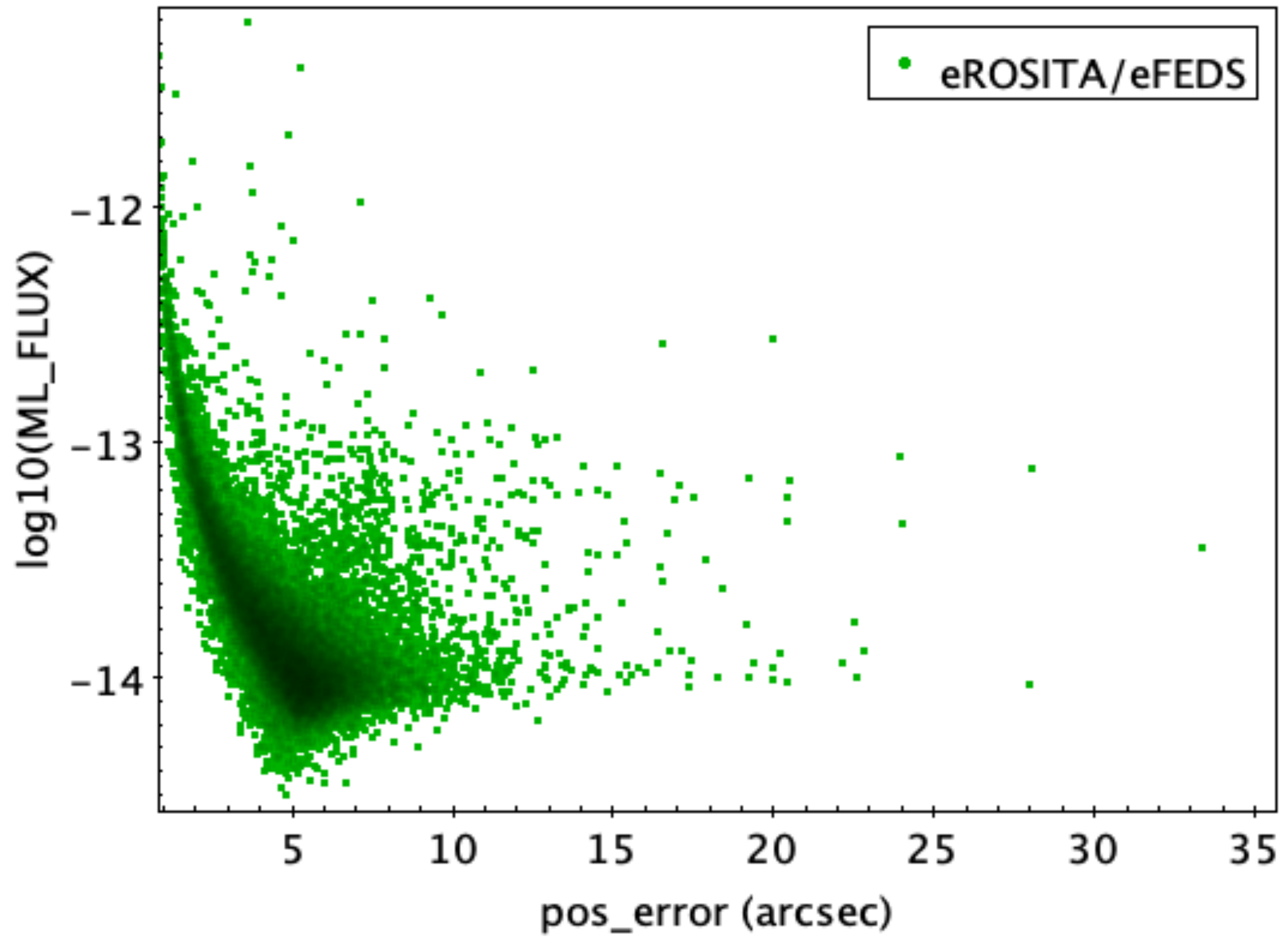
Brusa et al 2022

Wolf et al 2022









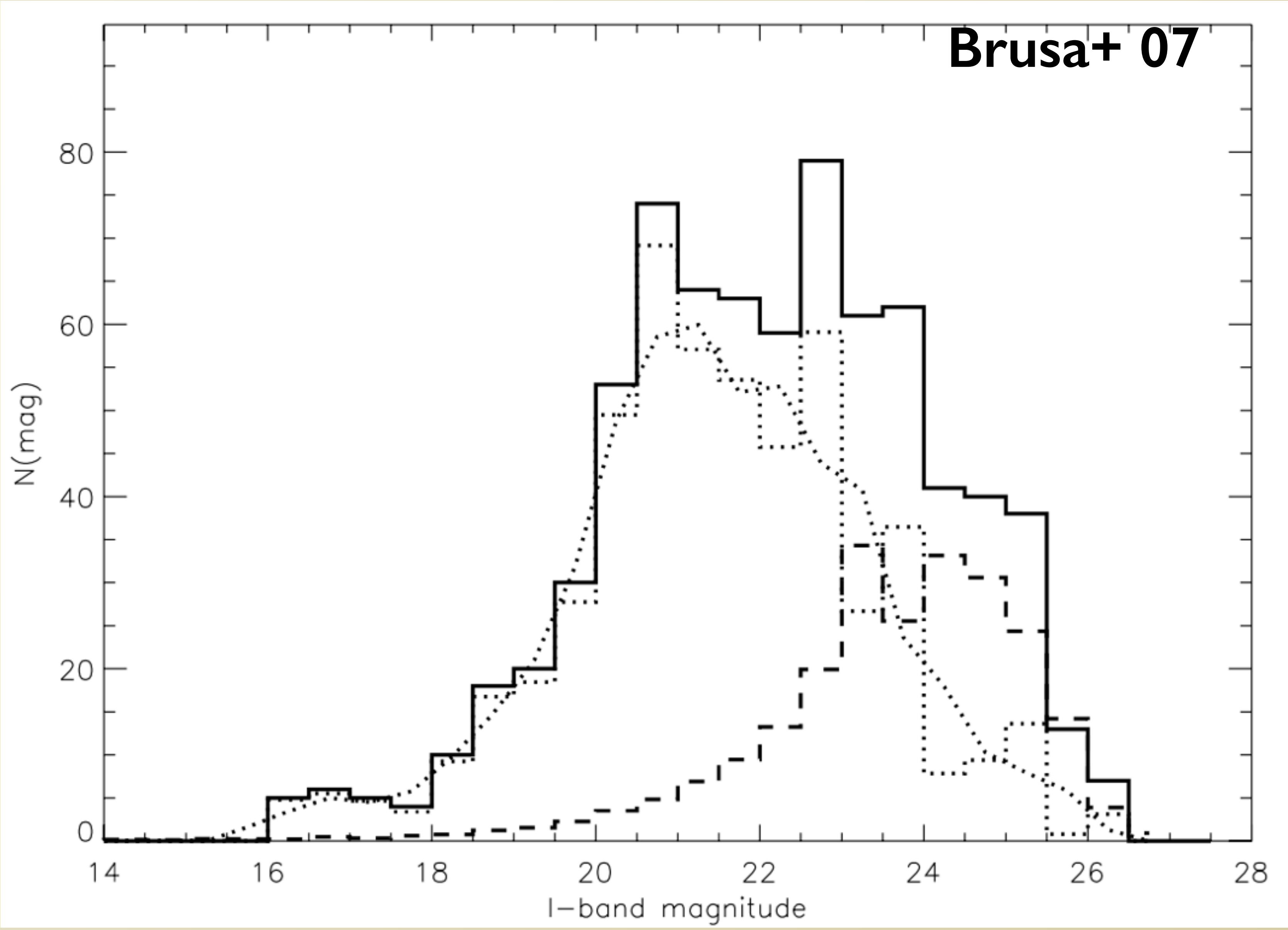
- The size of the area search depends on the positional error which correlates with the intensity of the source.

…but it could be also a Galactic source (e.g., Schneider et al 2022, Stelze et al 2022) or an unresolved cluster (e.g., Bulbul et al 2022)

- The X-ray coordinates and positional errors depend on how the X-ray data are treated (Hsu+2014)

Maximum Likelihood (ML) in a nutshell:

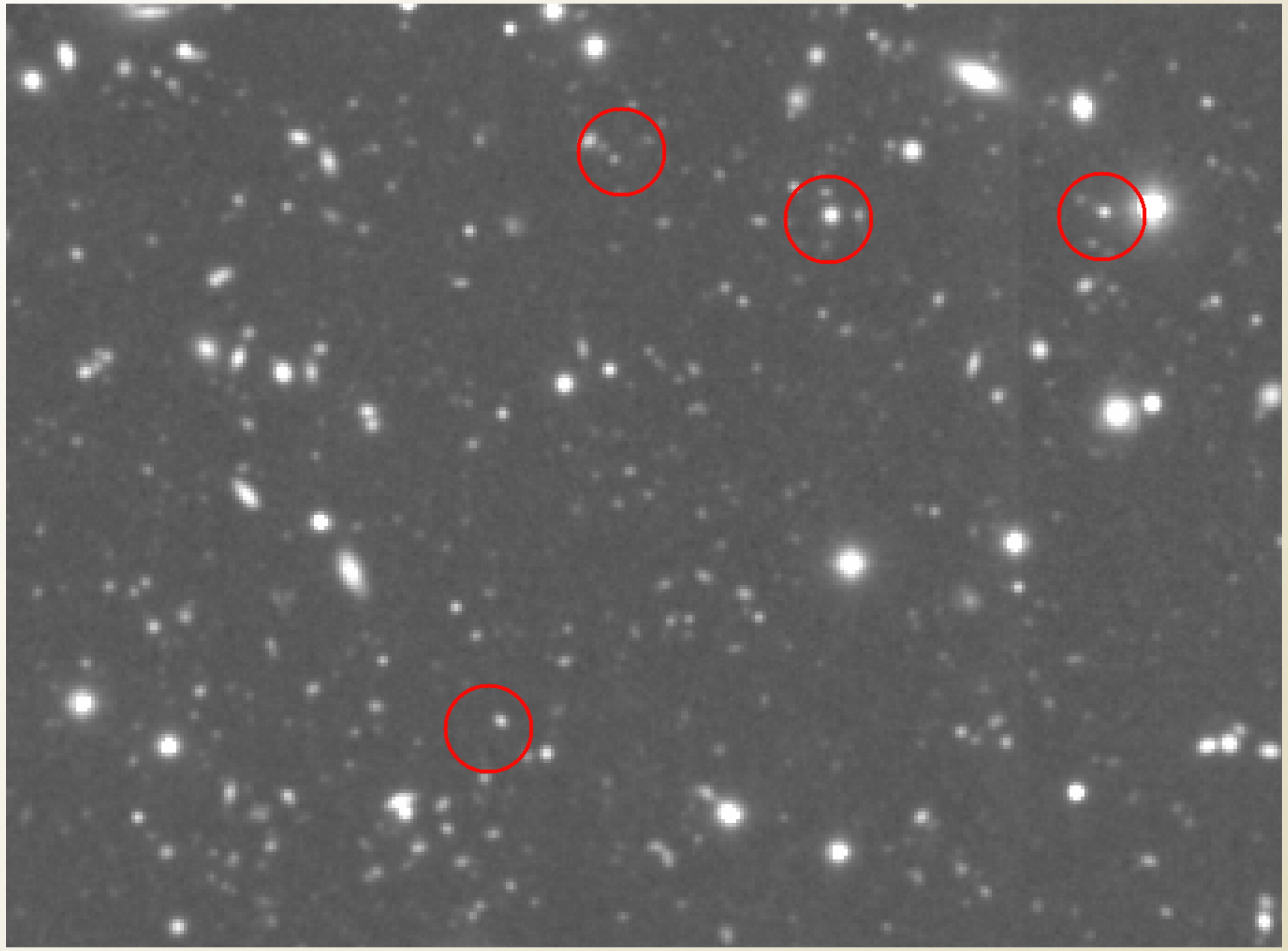
Sutherland&Saunders1992

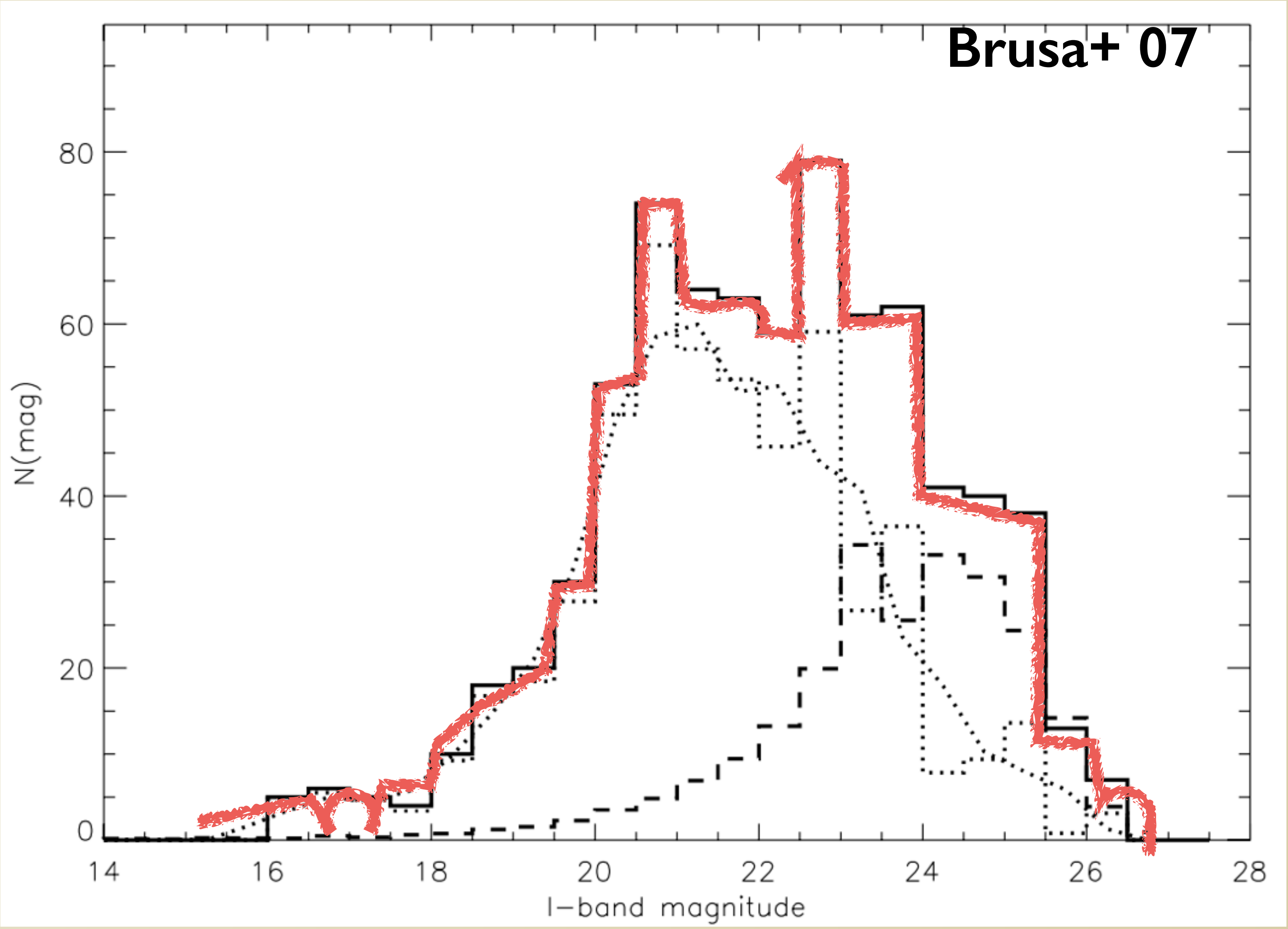


Naylor+13 for a review

Maximum Likelihood (ML) in a nutshell:

Sutherland&Saunders1992

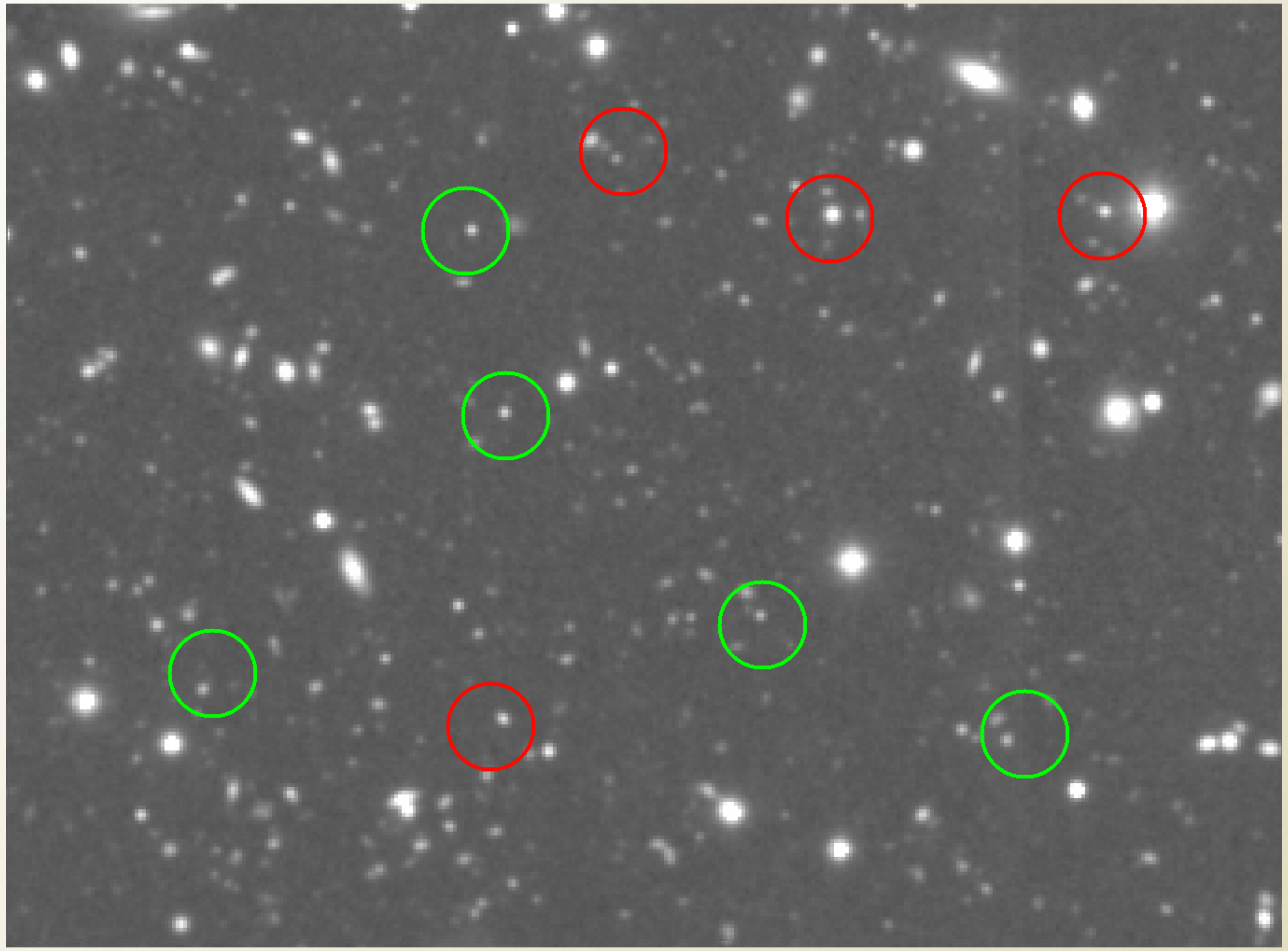


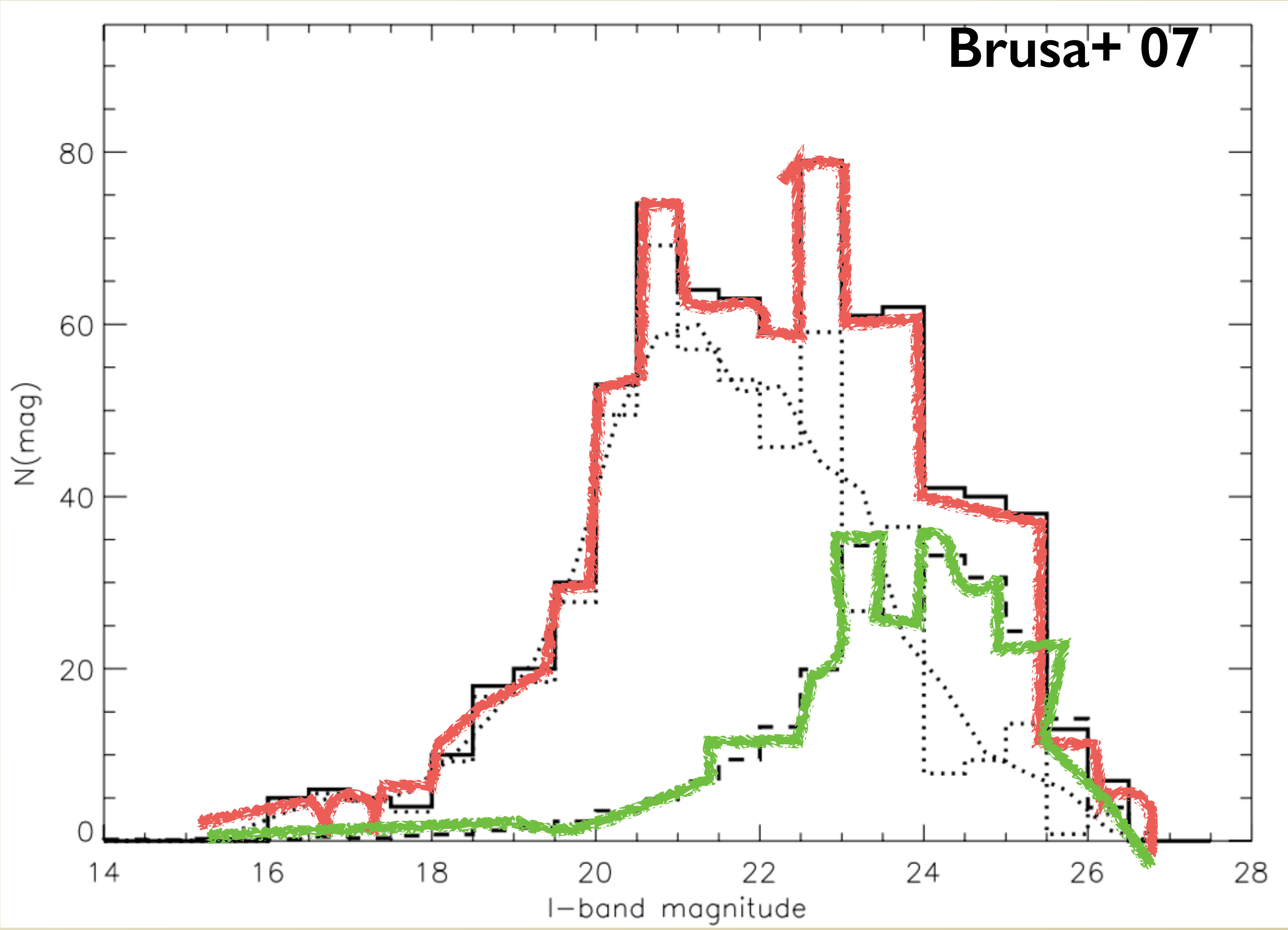


Naylor+13 for a review

Maximum Likelihood (ML) in a nutshell:

Sutherland&Saunders1992

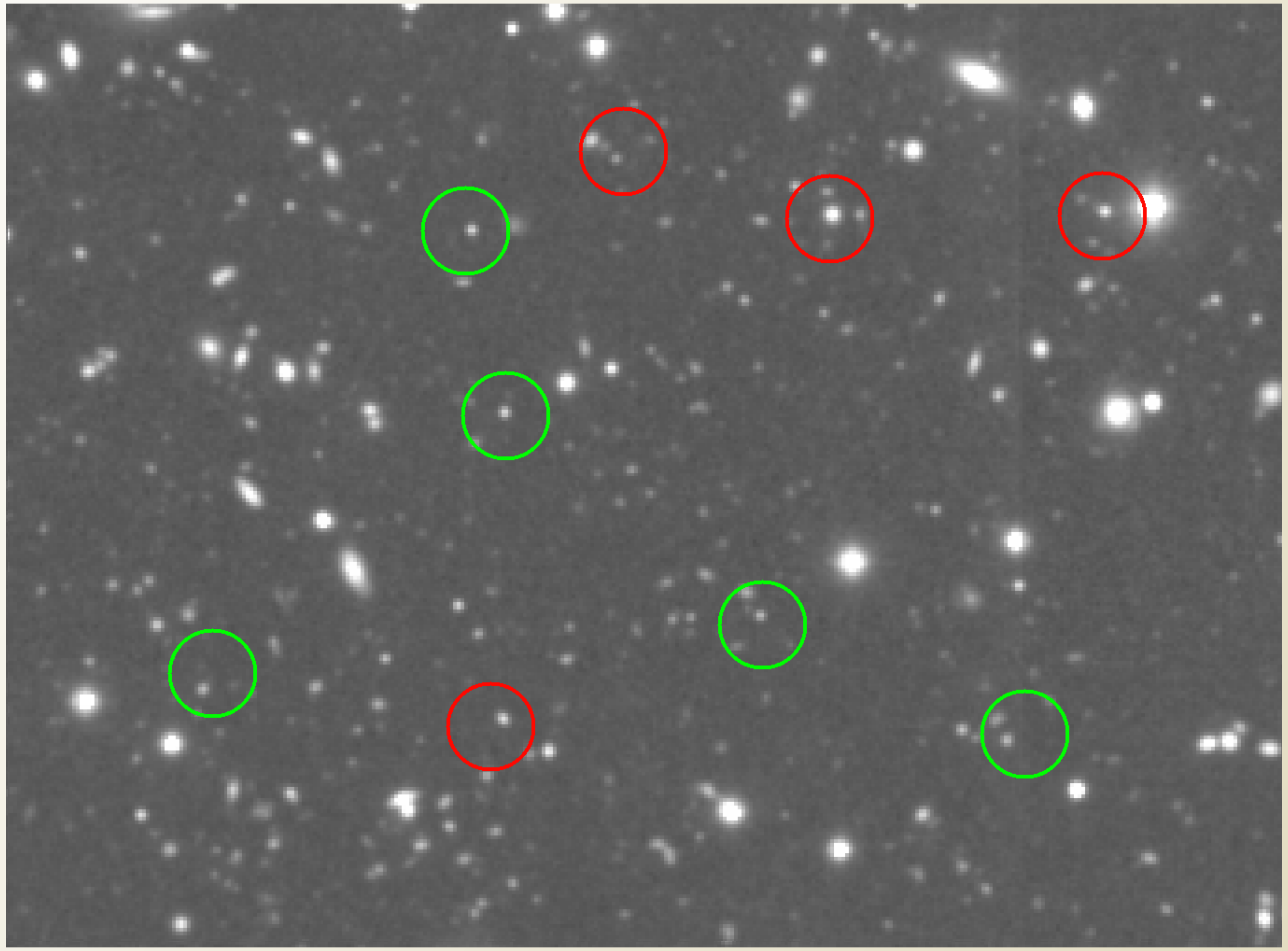


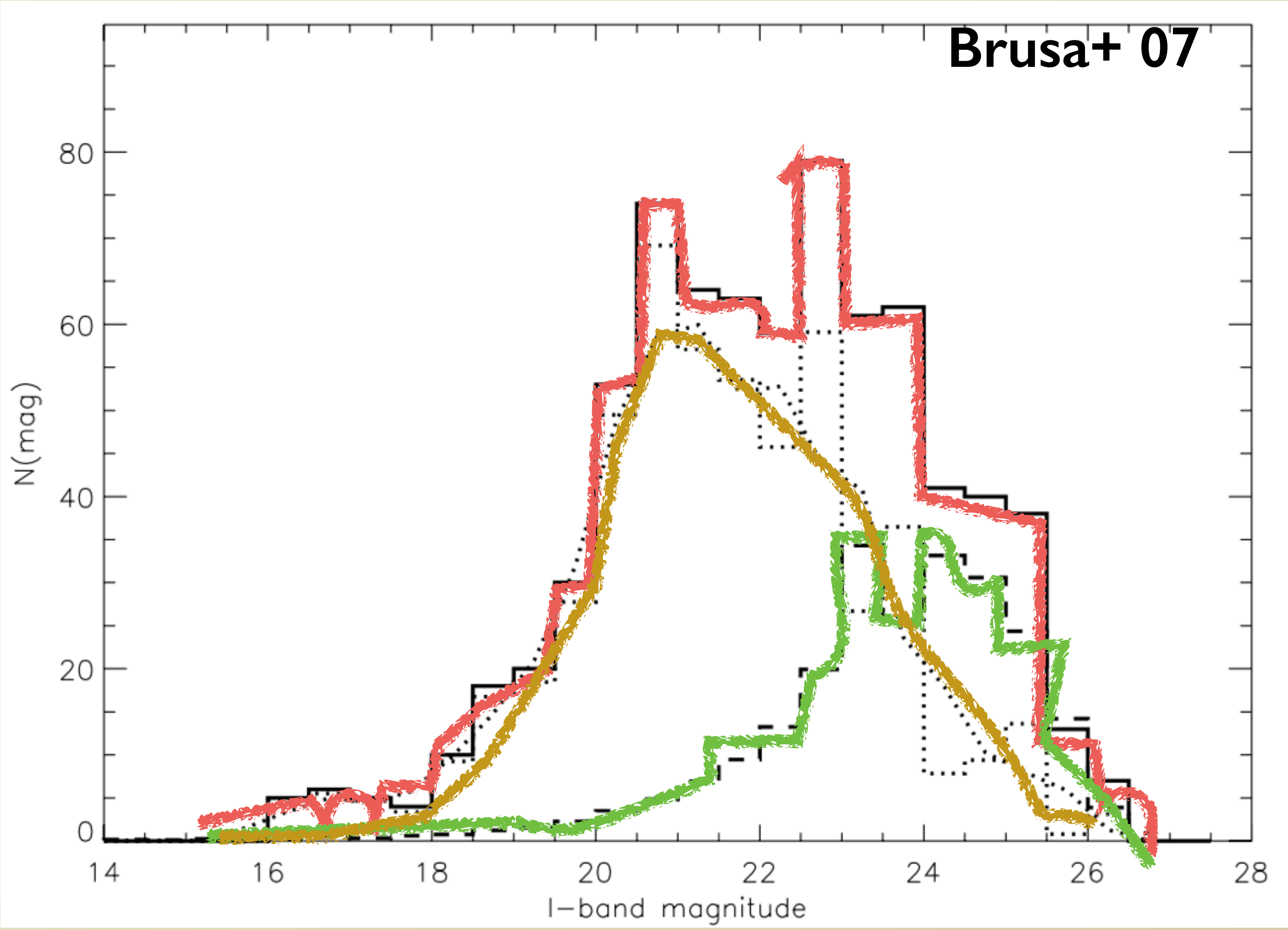


Naylor+13 for a review

Maximum Likelihood (ML) in a nutshell:

Sutherland&Saunders1992

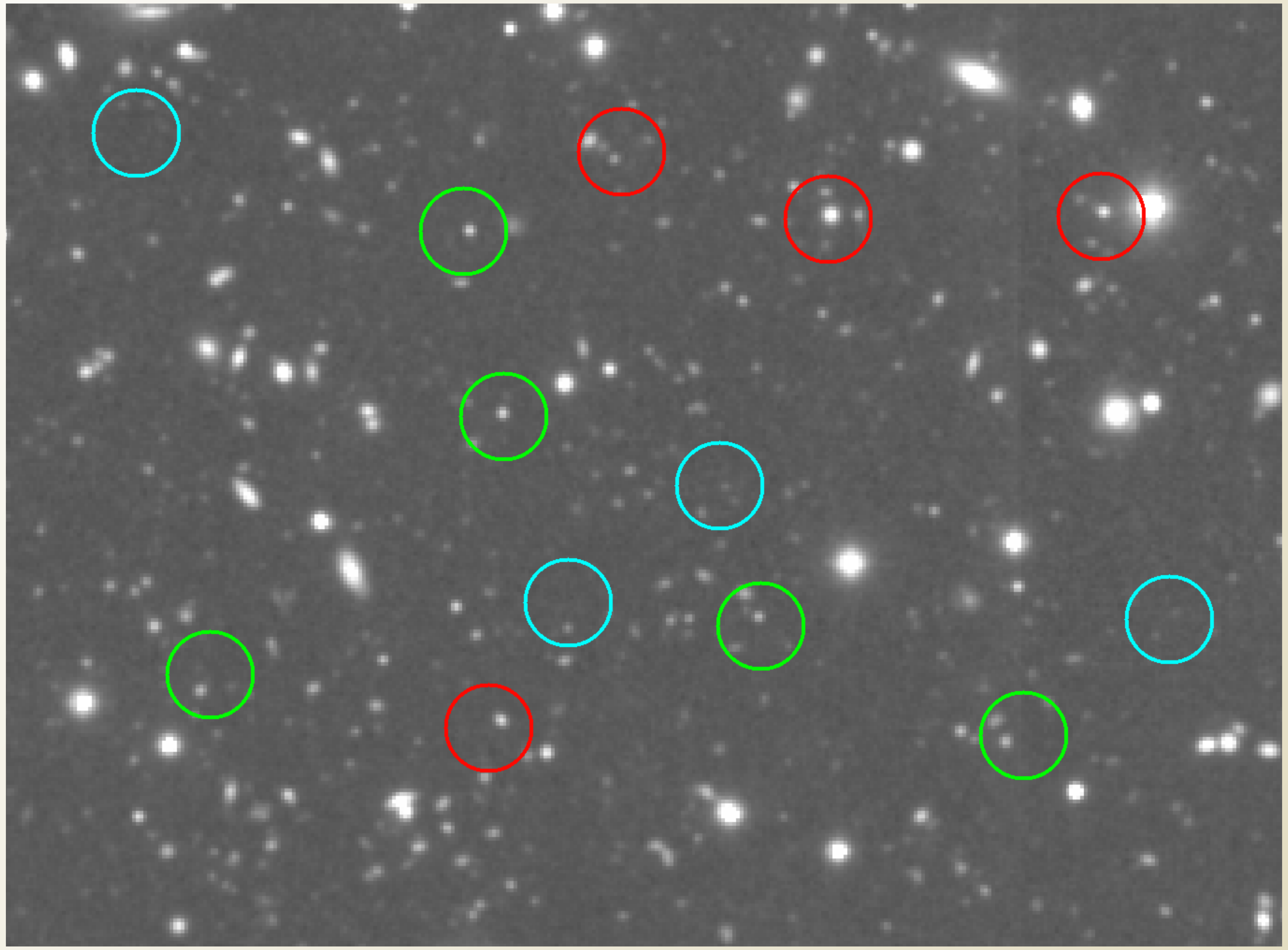


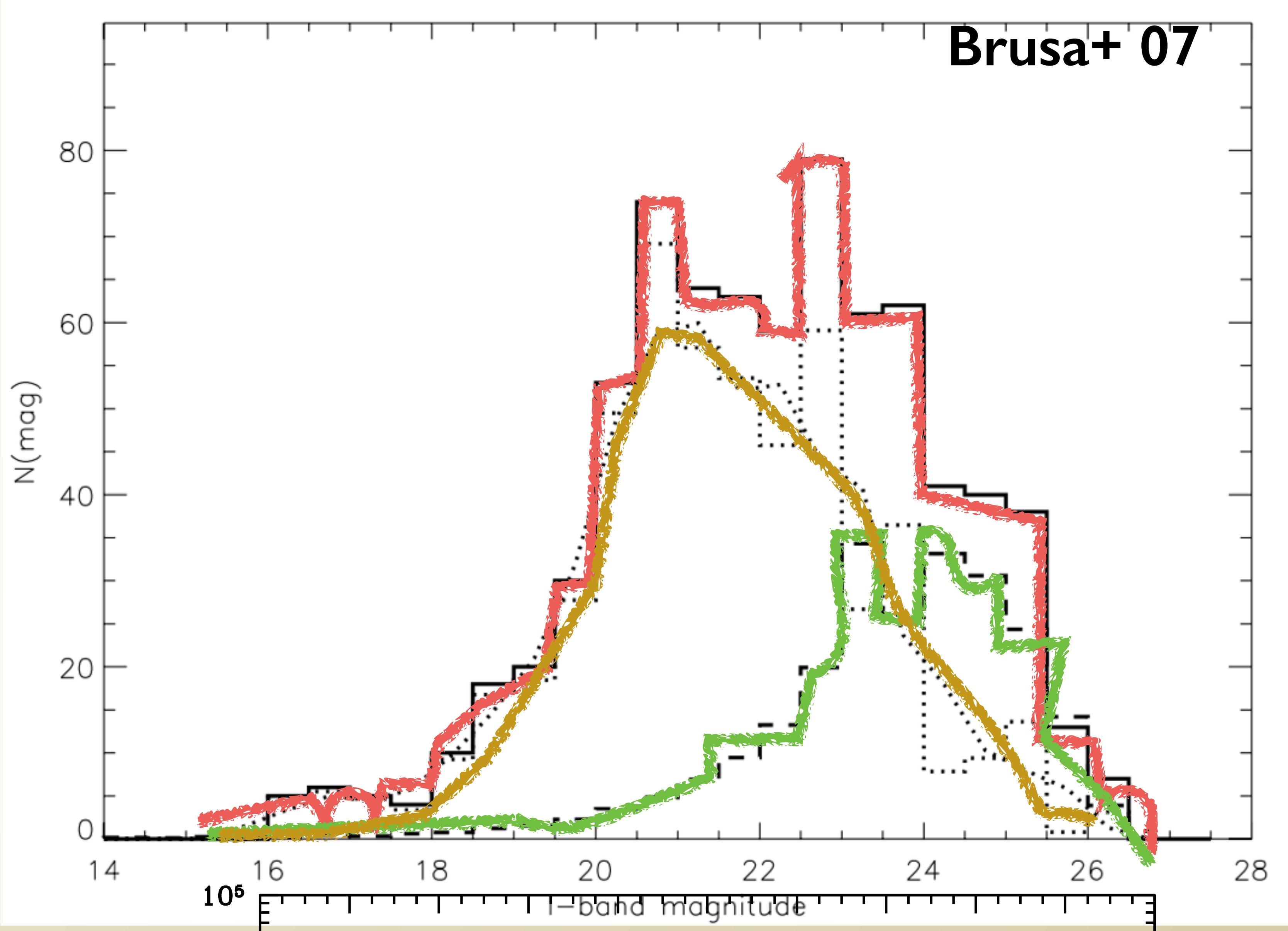


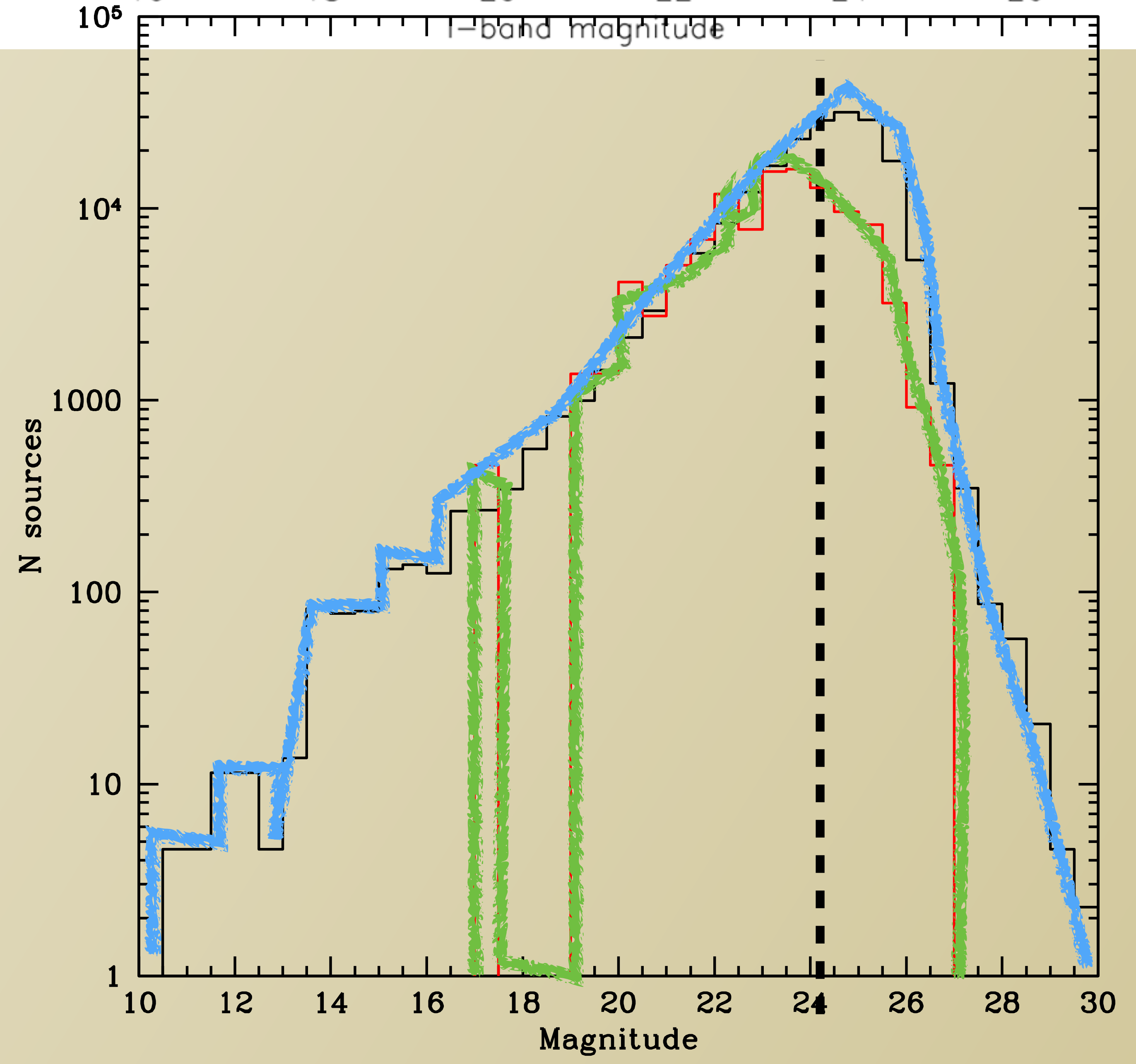
Naylor+13 for a review

Maximum Likelihood (ML) in a nutshell:

Sutherland&Saunders1992



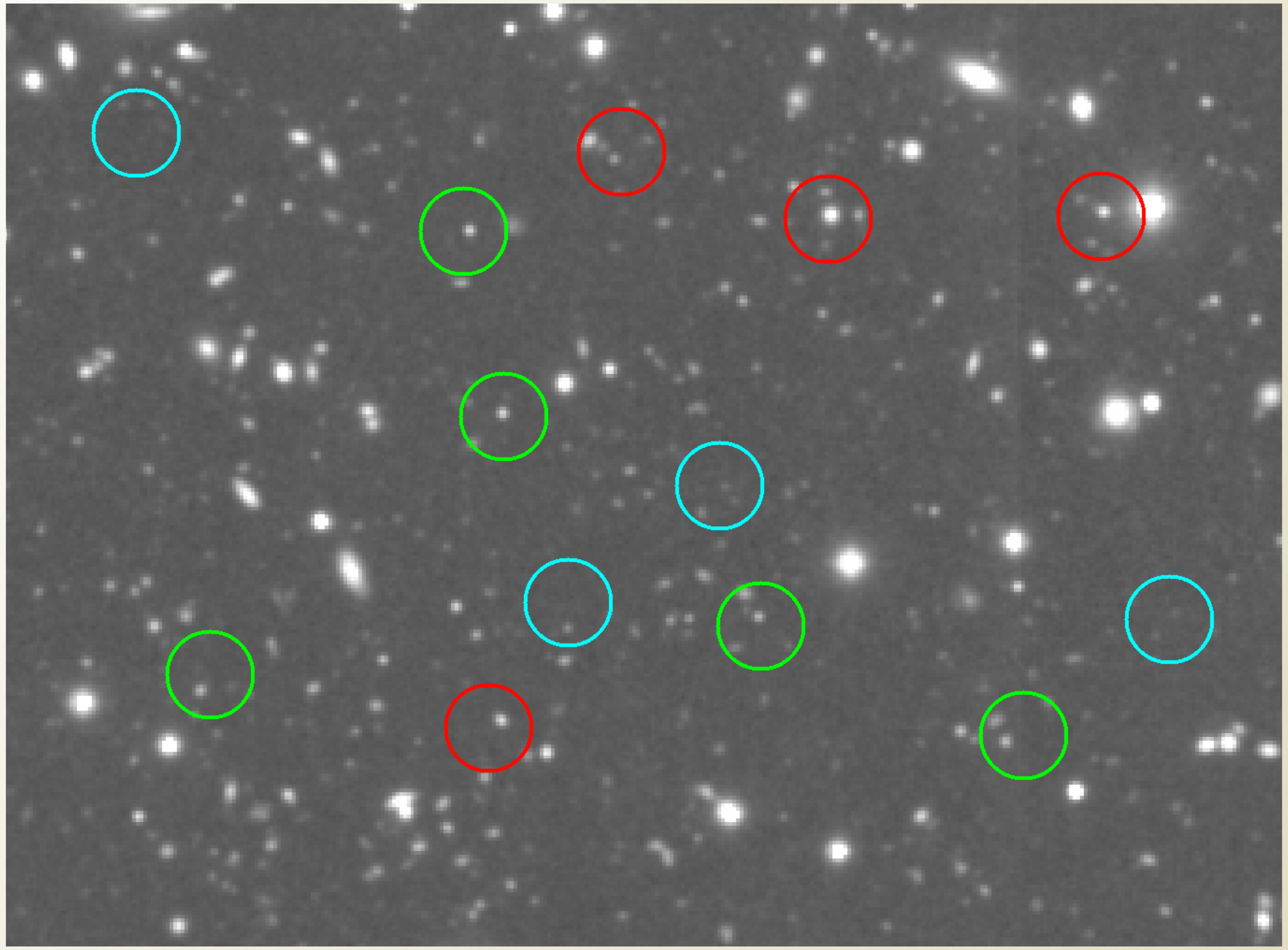


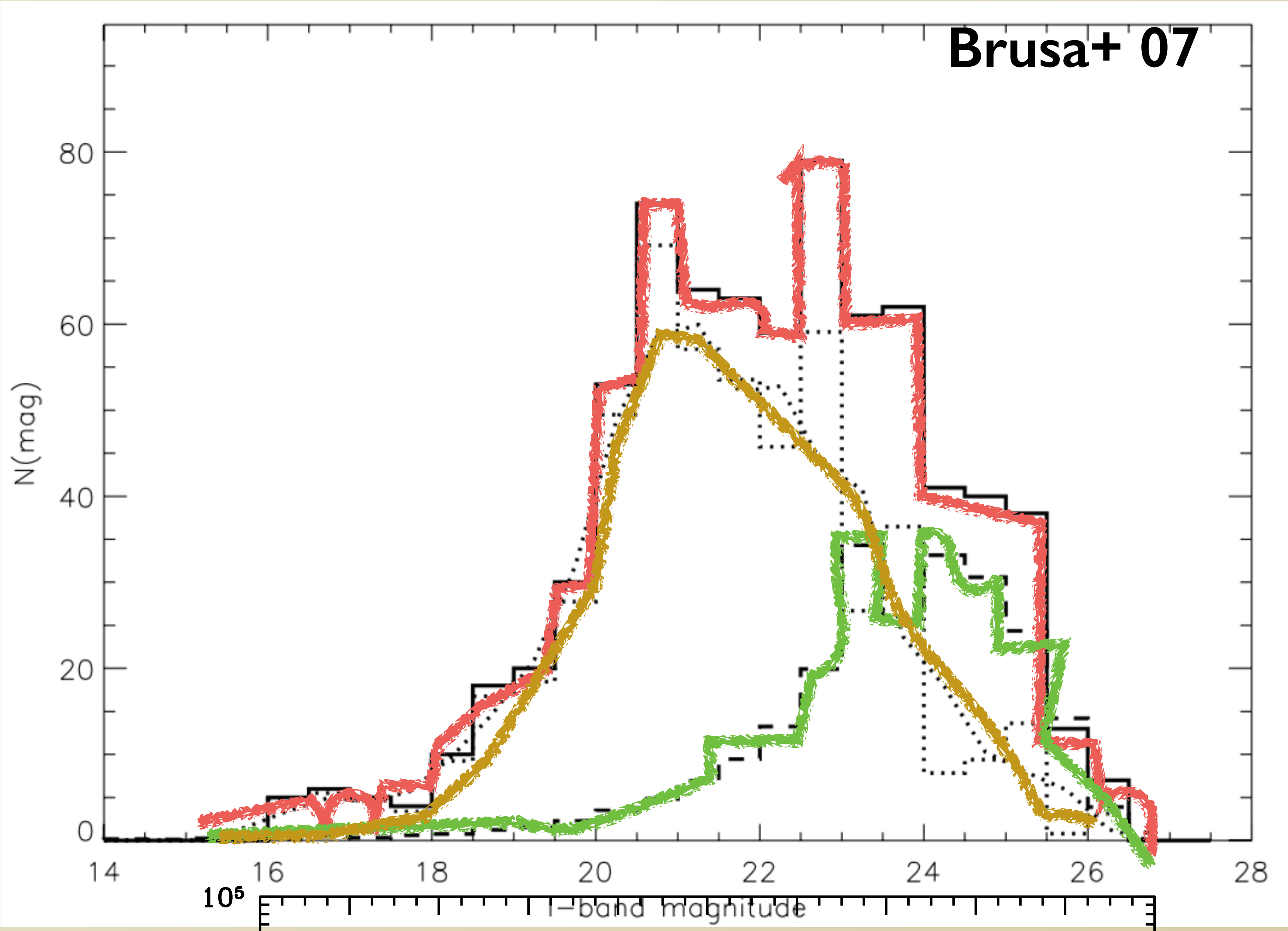


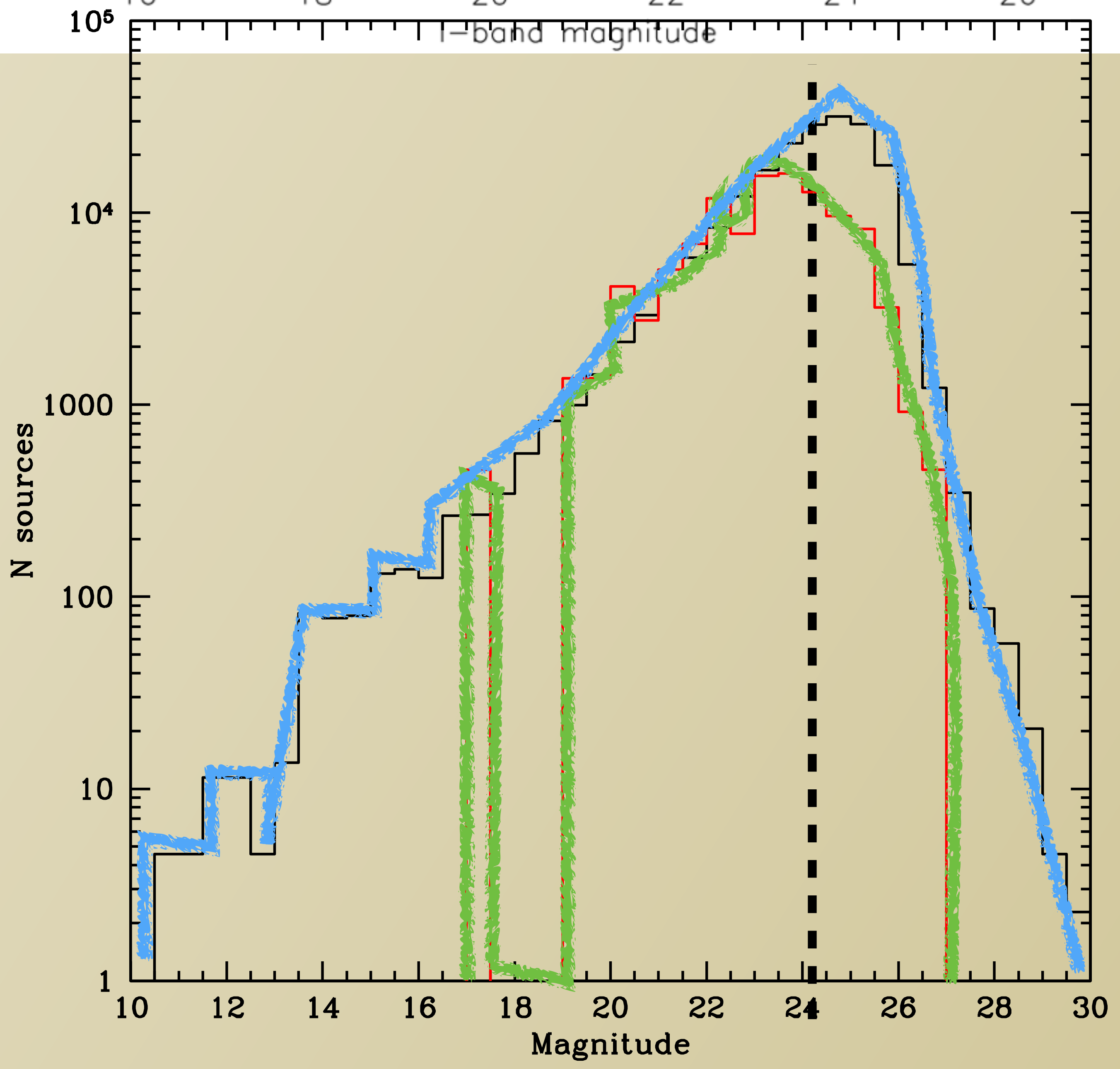
Naylor+13 for a review

Maximum Likelihood (ML) in a nutshell:

Sutherland&Saunders1992



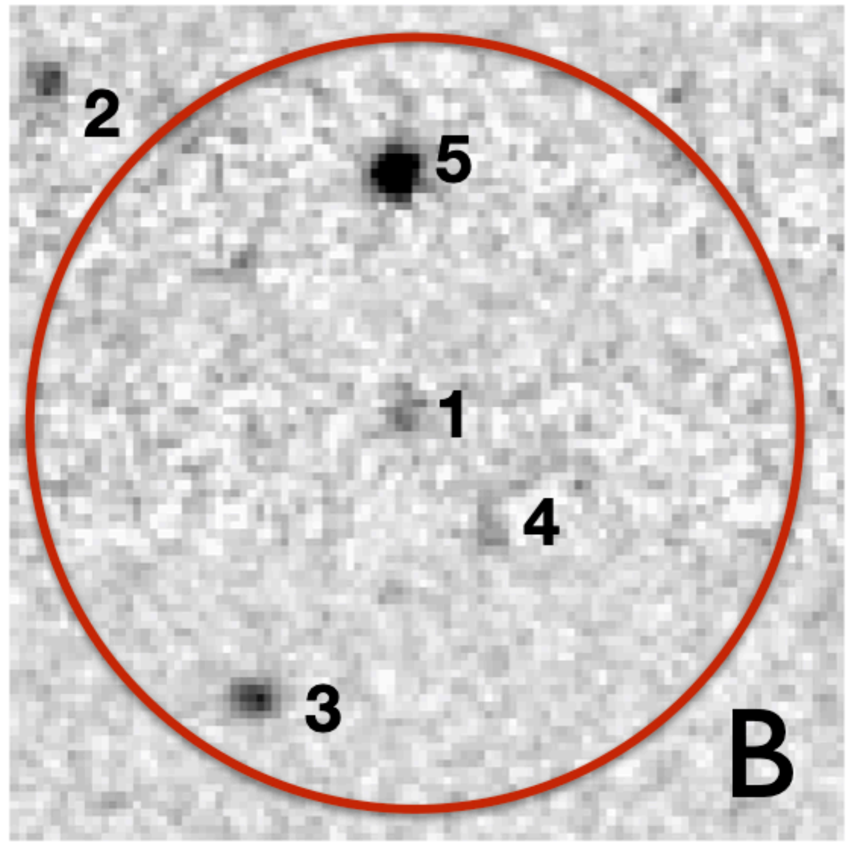




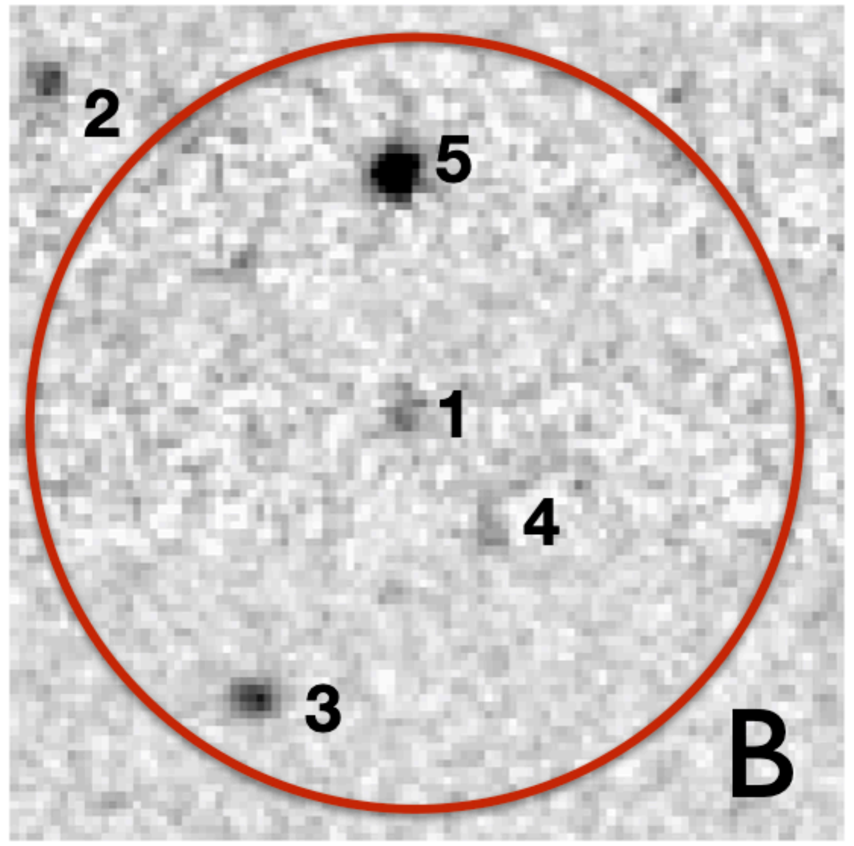
It is data driven:
problem for small data set

Naylor+13 for a review

issue: different bands provide different results

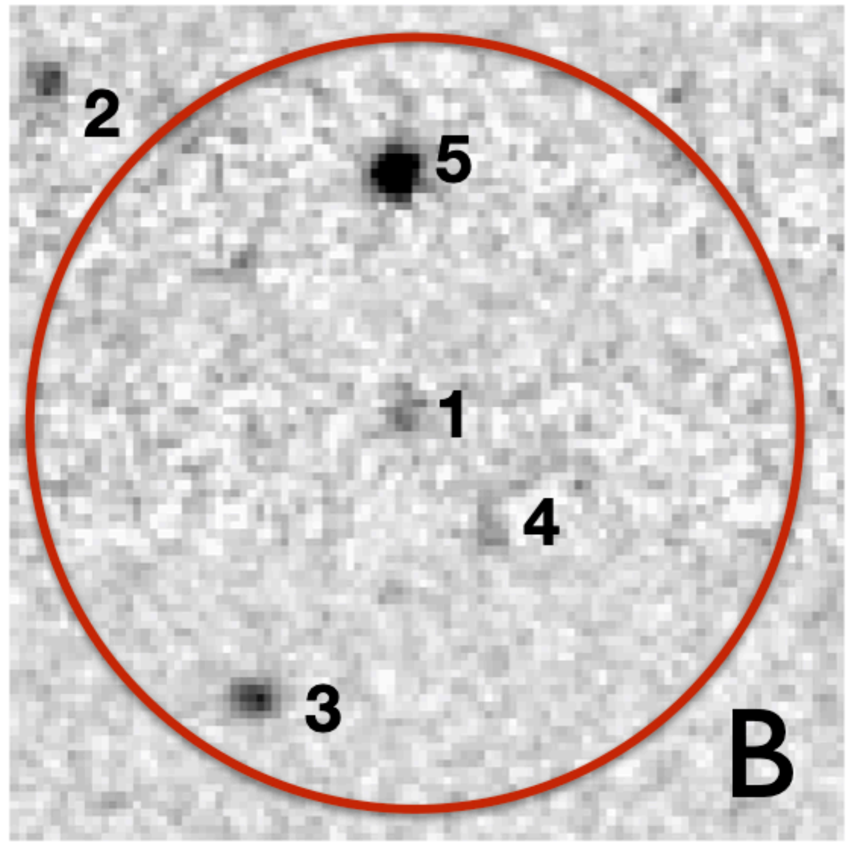


issue: different bands provide different results



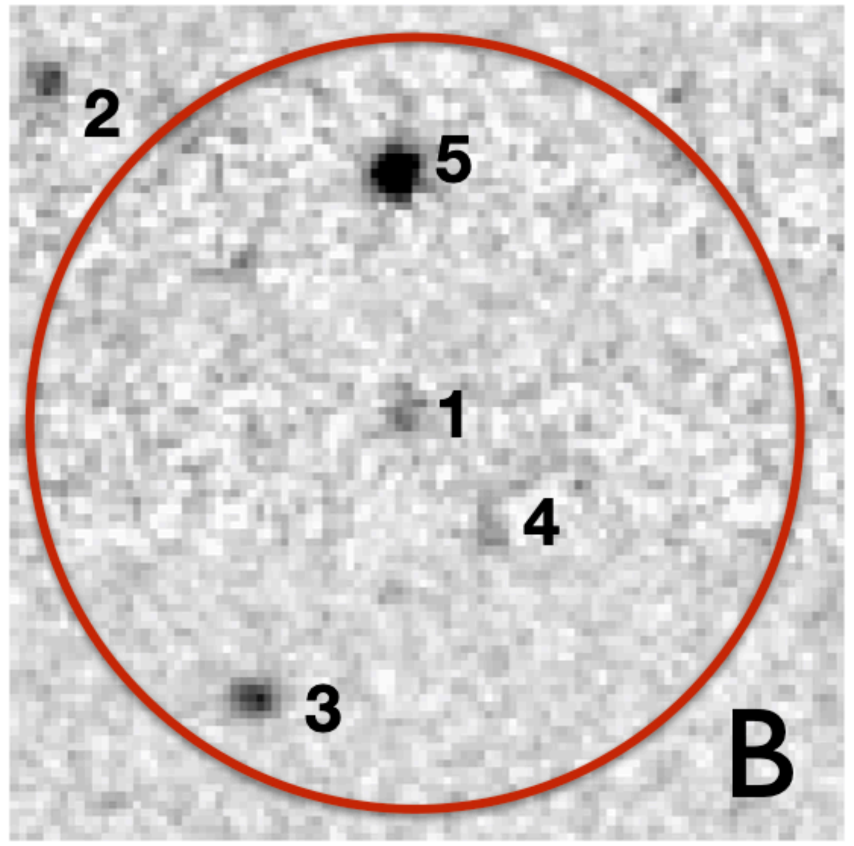
- when assigning a counterpart we consider:

issue: different bands provide different results



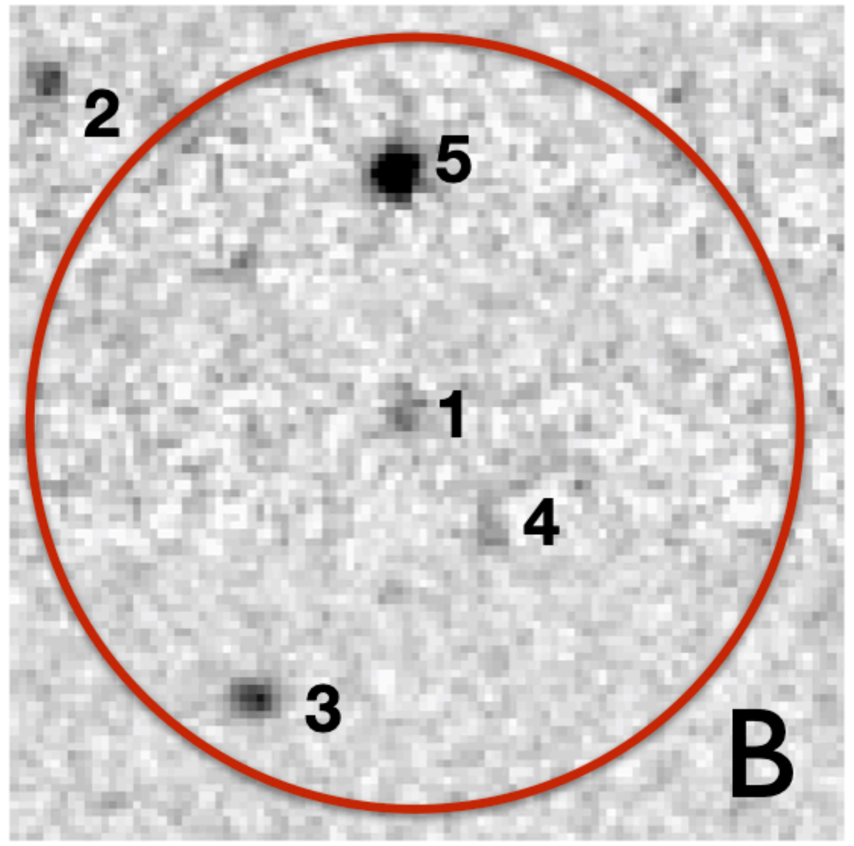
- when assigning a counterpart we consider:
  - the separation between the sources in the primary and secondary catalog

issue: different bands provide different results



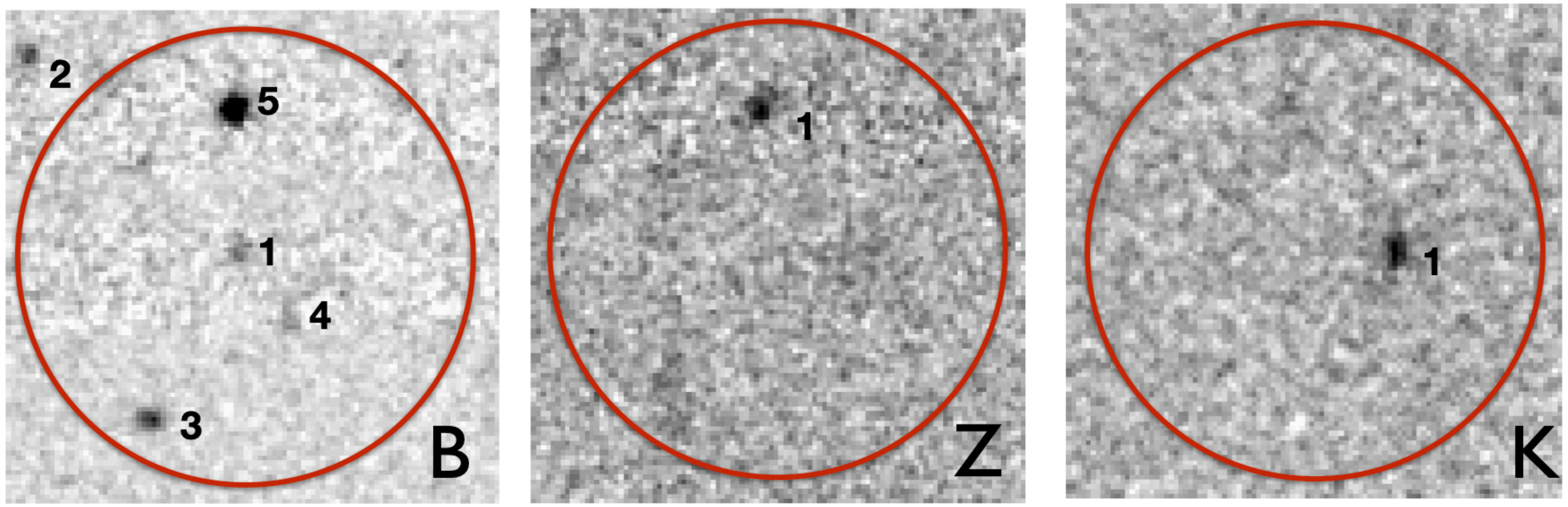
- when assigning a counterpart we consider:
  - the separation between the sources in the primary and secondary catalog
  - positional error for each source in the primary and secondary catalog

issue: different bands provide different results



- when assigning a counterpart we consider:
  - the separation between the sources in the primary and secondary catalog
  - positional error for each source in the primary and secondary catalog
  - source number density in the primary and secondary catalog

issue: different bands provide different results

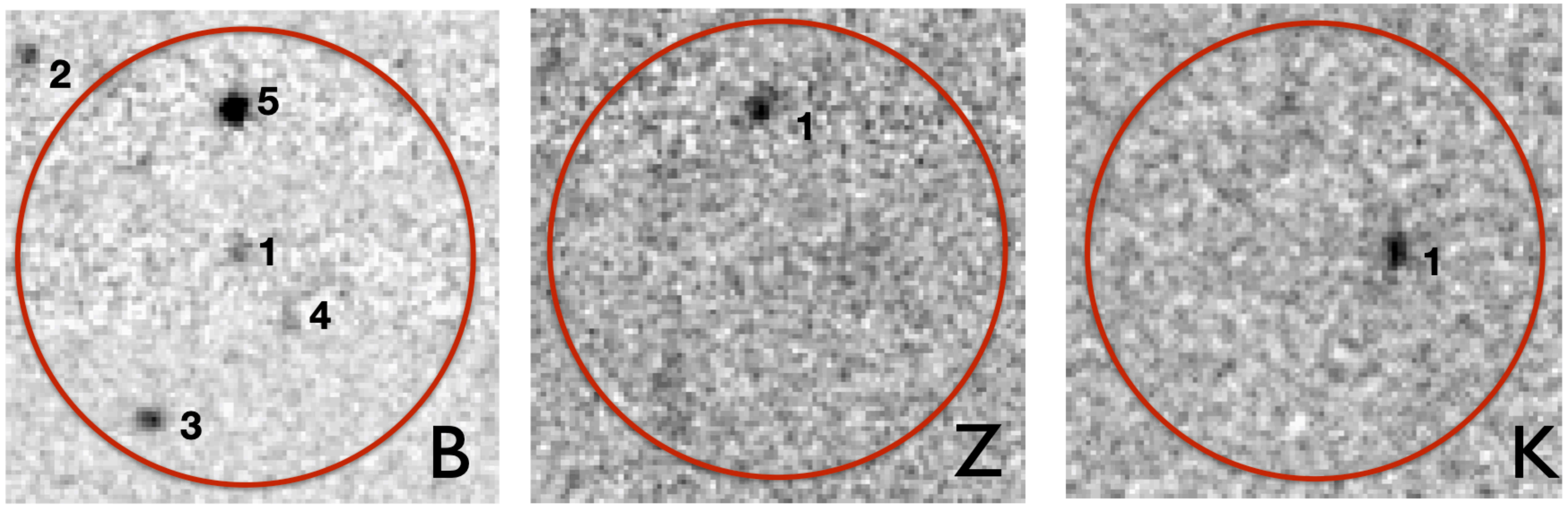


- when assigning a counterpart we consider:
  - the separation between the sources in the primary and secondary catalog
  - positional error for each source in the primary and secondary catalog
  - source number density in the primary and secondary catalog

- using only one band we do not account for the possibility that the actual counterpart is *NOT* detected in *THAT* band

- repeating the process with different bands and for each X-ray source selecting the counterpart from the band with the higher probability compensate only partially

issue: different bands provide different results



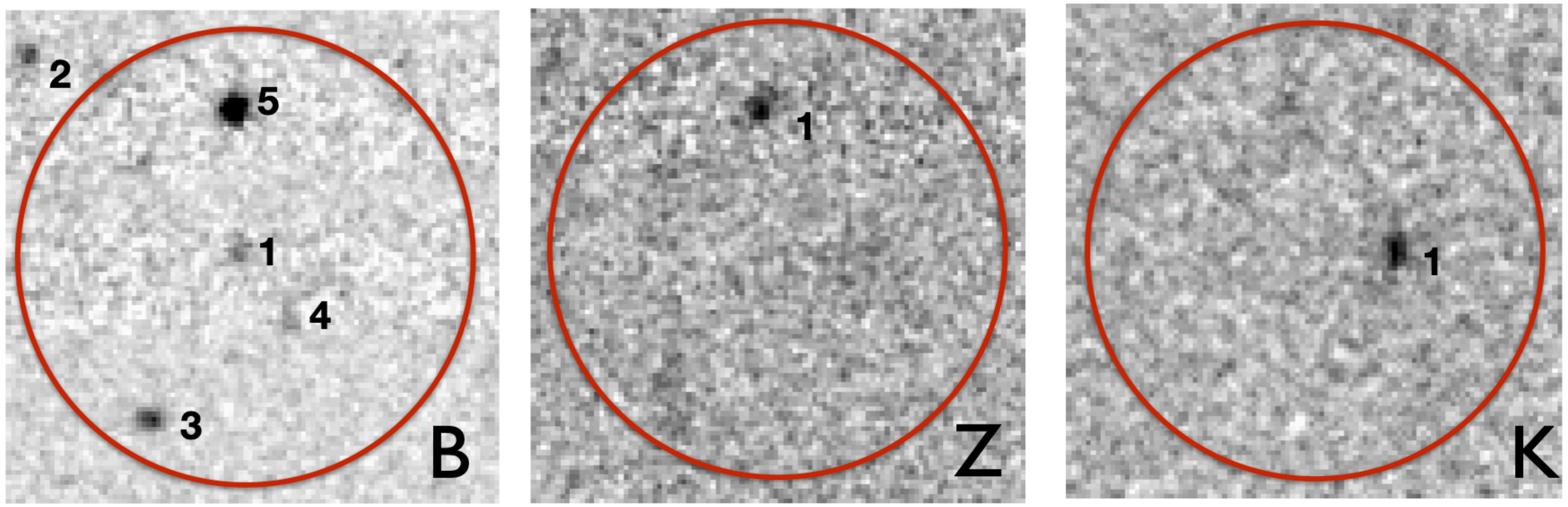
- when assigning a counterpart we consider:
  - the separation between the sources in the primary and secondary catalog
  - positional error for each source in the primary and secondary catalog
  - source number density in the primary and secondary catalog

- using only one band we do not account for the possibility that the actual counterpart is *NOT* detected in *THAT* band

- repeating the process with different bands and for each X-ray source selecting the counterpart from the band with the higher probability compensate only partially

go Bayesian: use ALL the bands at the same time
and combine the probabilities before assigning the CTP (Budavari & Szalay 2008)
But account also for missing data: Nway (Salvato+2018), Xmatch (Pineau2017)

issue: different bands provide different results



- when assigning a counterpart we consider:
  - the separation between the sources in the primary and secondary catalog
  - positional error for each source in the primary and secondary catalog
  - source number density in the primary and secondary catalog

- using only one band we do not account for the possibility that the actual counterpart is *NOT* detected in *THAT* band

- repeating the process with different bands and for each X-ray source selecting the counterpart from the band with the higher probability compensate only partially

X cat. entry	B cat. entry	Z cat. entry	K cat. entry	P (this is the correct ctp)
1	5	1	—	P1
1	—	—	1	P2
1	1	—	—	P3
1	3	—	—	...
1	...	...	...	...
2	...	...	...	...

go Bayesian: use ALL the bands at the same time
and combine the probabilities before assigning the CTP (Budavari & Szalay 2008)
But account also for missing data: Nway (Salvato+2018), Xmatch (Pineau2017)

benefit: different bands provide different information

$$P(D|H) = P(D_\phi|H) \times P(D_m|H)$$
$$= P(D_\phi|H) \times \frac{\bar{q}(m)}{\bar{n}(m)},$$

probability based purely on spatial information

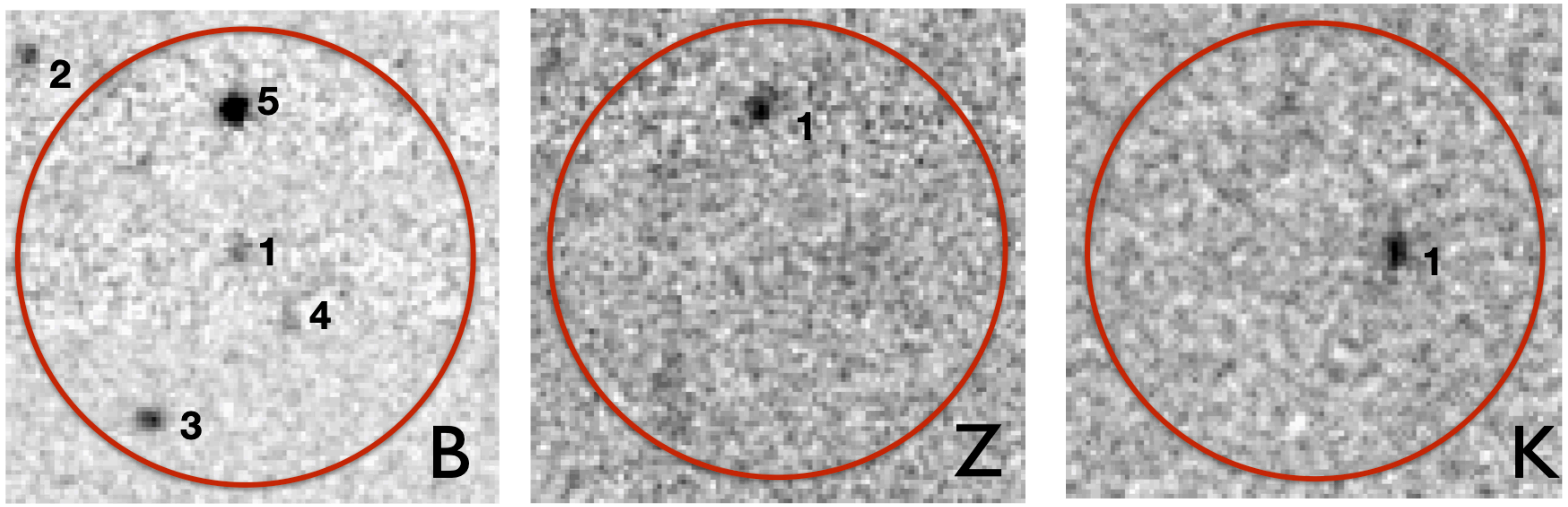
probability that a correct ctp to a X-ray source or a generic field source has a property m.

benefit: different bands provide different information

$$
\begin{aligned}
P(D|H) &= P(D_\phi|H) \times P(D_m|H) \\
&= P(D_\phi|H) \times \frac{\bar{q}(m)}{\bar{n}(m)},
\end{aligned}
$$

probability based purely on spatial information

probability that a correct ctp to a X-ray source or a generic field source has a property m.

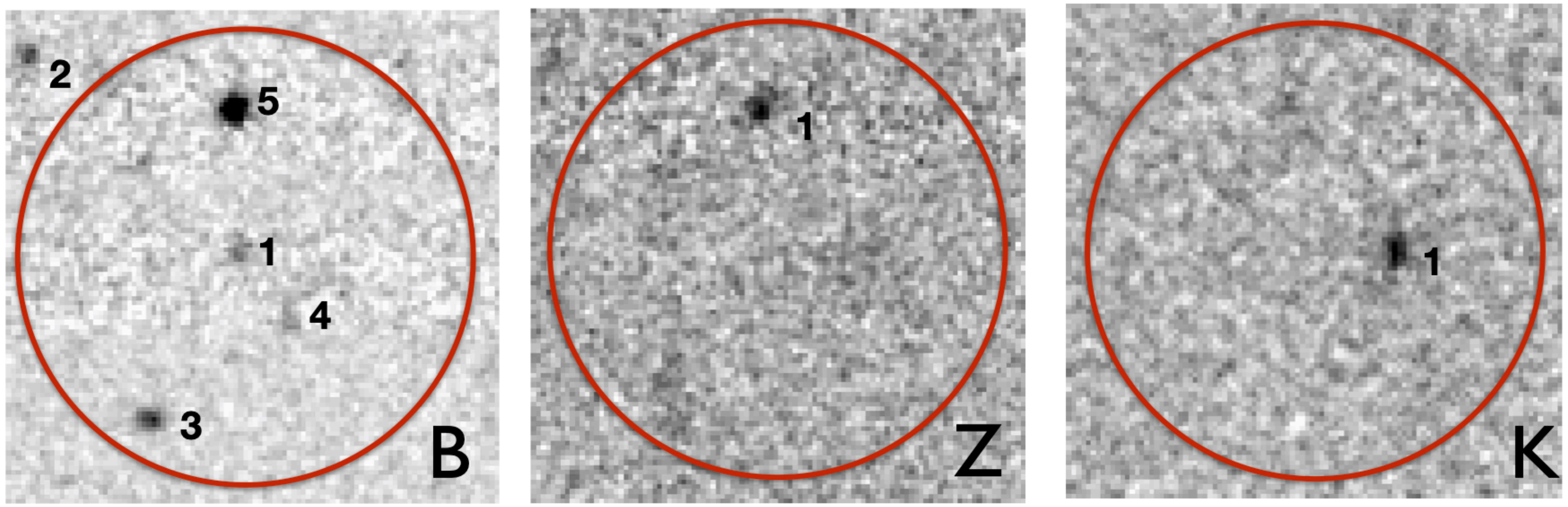


benefit: different bands provide different information

$$P(D|H) = P(D_\phi|H) \times P(D_m|H)$$
$$= P(D_\phi|H) \times \prod \frac{\int_m \bar{q}(m)\, p(m|D_m)\, dm}{\int_m \bar{n}(m)\, p(m|D_m)\, dm}$$

probability based purely on spatial information

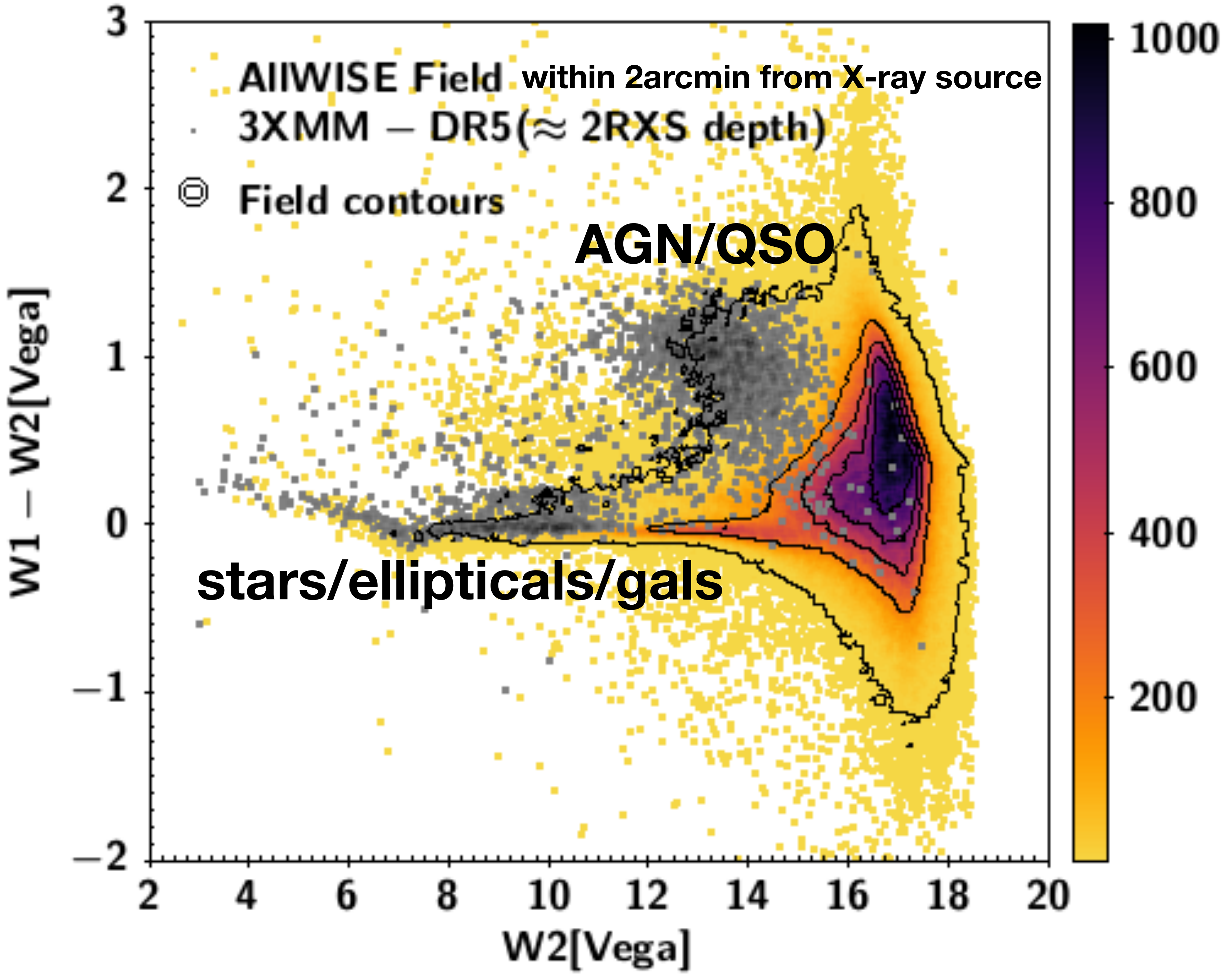
probability that a correct ctp to a X-ray source or a generic field source has a property m.



We KNOW the properties (e.g., SEDs, variability, morphology) of X-ray emitters thanks to 20 years of XMM and Chandra. Let's use that!

goal: identify features that define X-ray emitters wrt field population

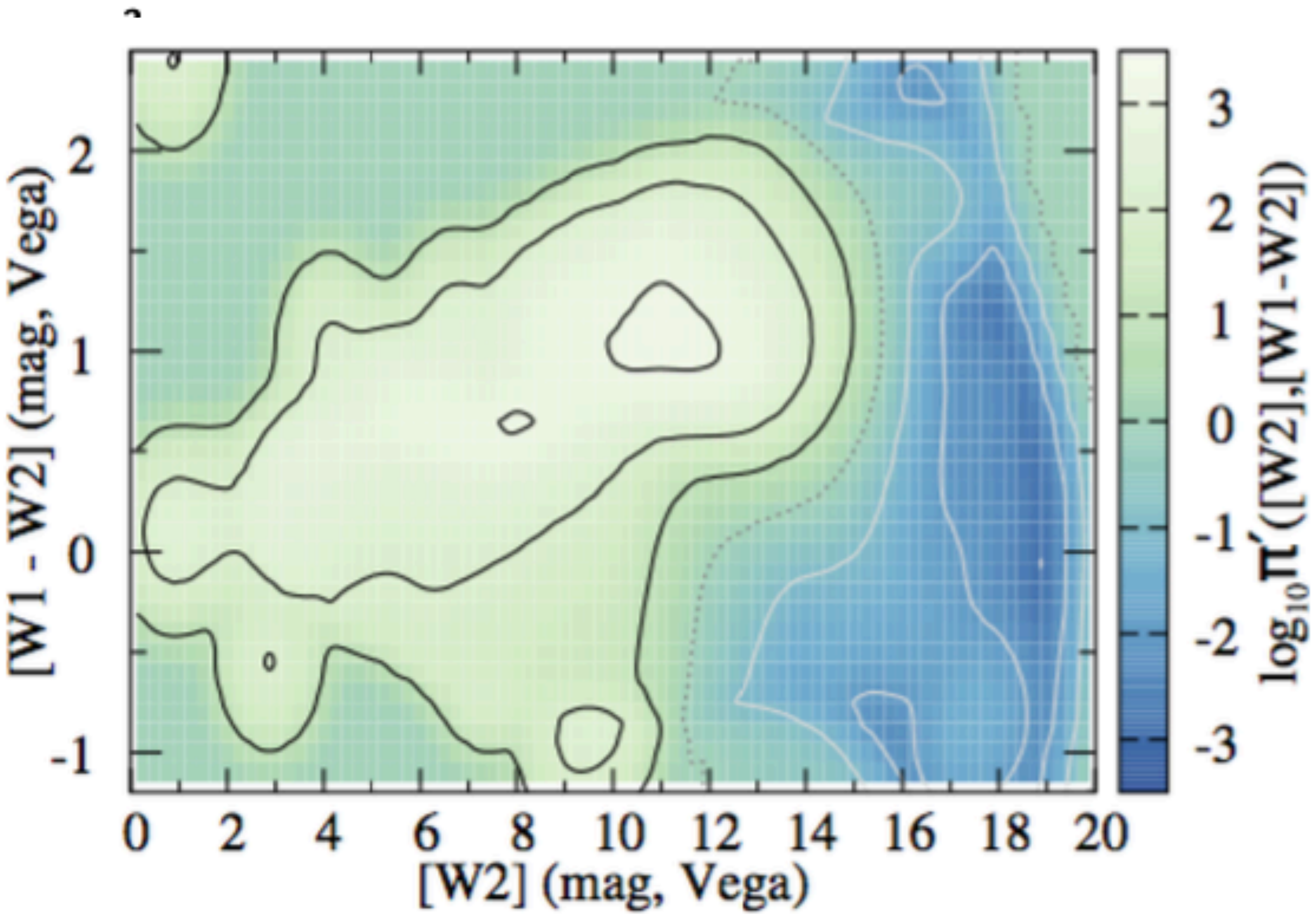
NWAY on ROSAT & XMMSLEW2 (Salvato+ 2018)



Validation: 1500 sources from 3XMM w/ secure ctp, shows 97% accuracy

goal: identify features that define X-ray emitters wrt field population

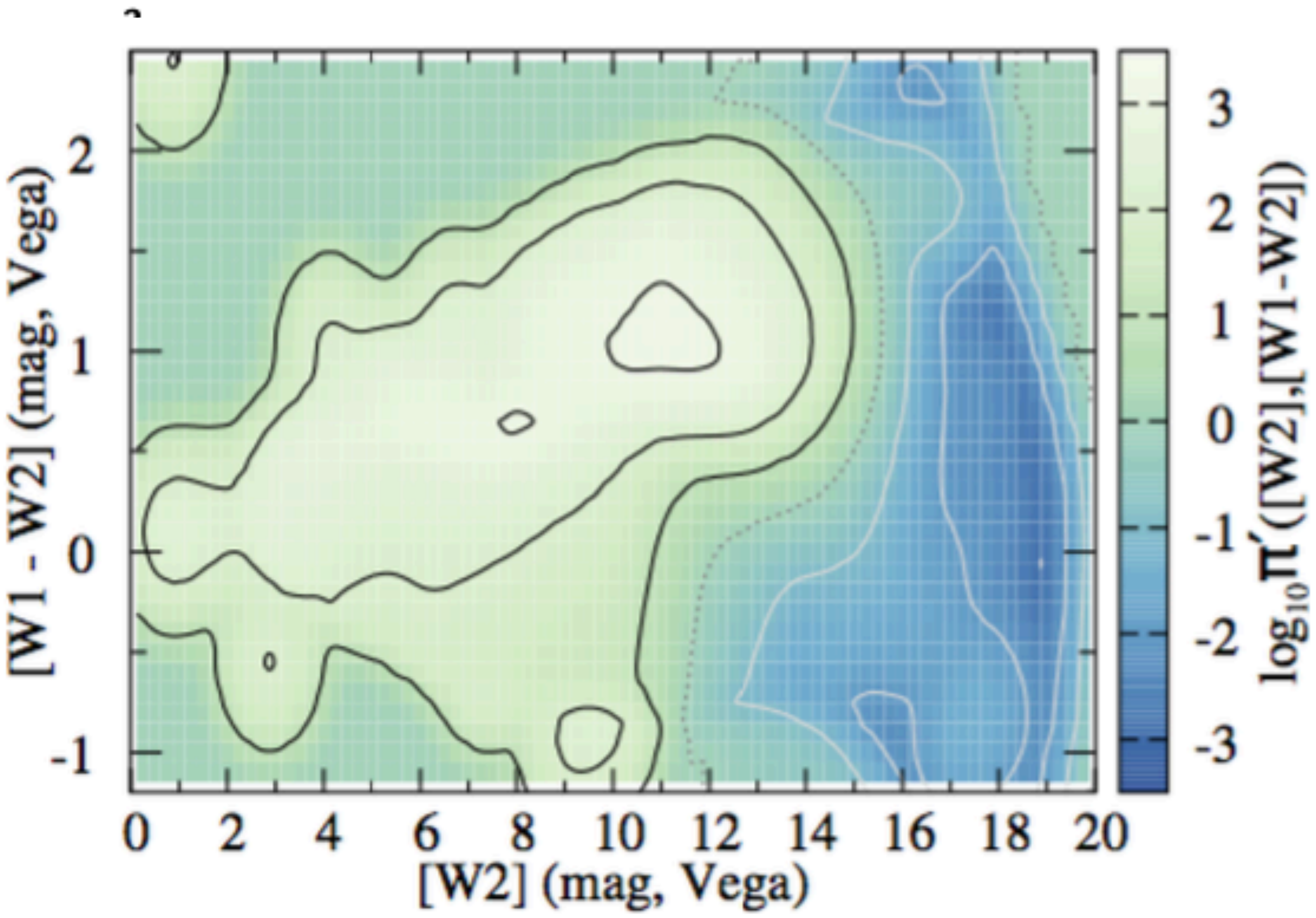
NWAY on ROSAT & XMMSLEW2 (Salvato+ 2018)



Validation: 1500 sources from 3XMM
w/ secure ctp, shows 97% accuracy

goal: identify features that define X-ray emitters wrt field population

NWAY on ROSAT & XMMSLEW2 (Salvato+ 2018)

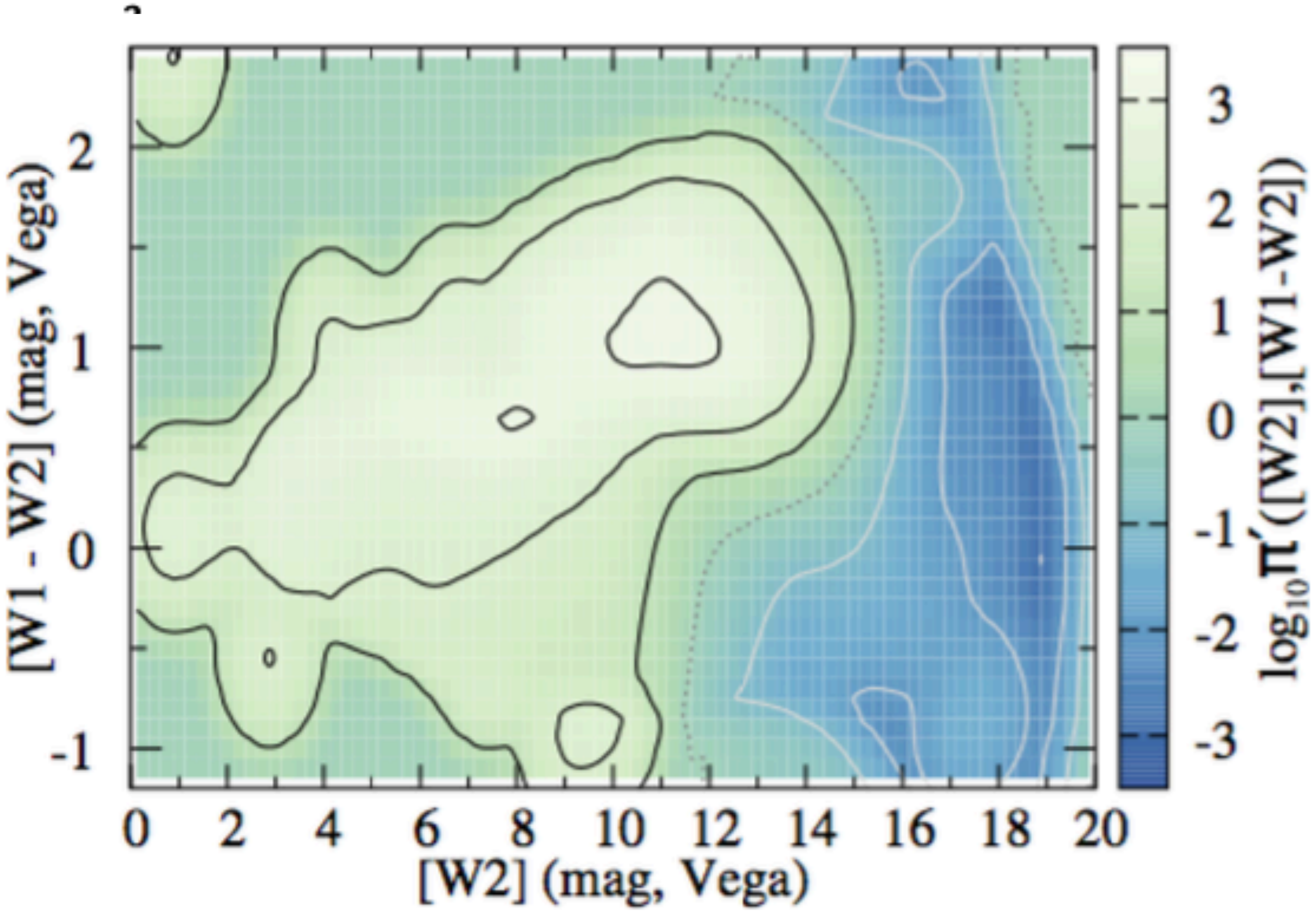


Validation: 1500 sources from 3XMM w/ secure ctp, shows 97% accuracy

XMMSL-2RXS Separation arcsec	Sources in common N	Identical AllWISE ctp. %
Sep. ≤5	1111	98.5
Sep. ≤10	3448	98.7
Sep. *le*30	7834	96.1
Sep. *le*60	8768	93.0

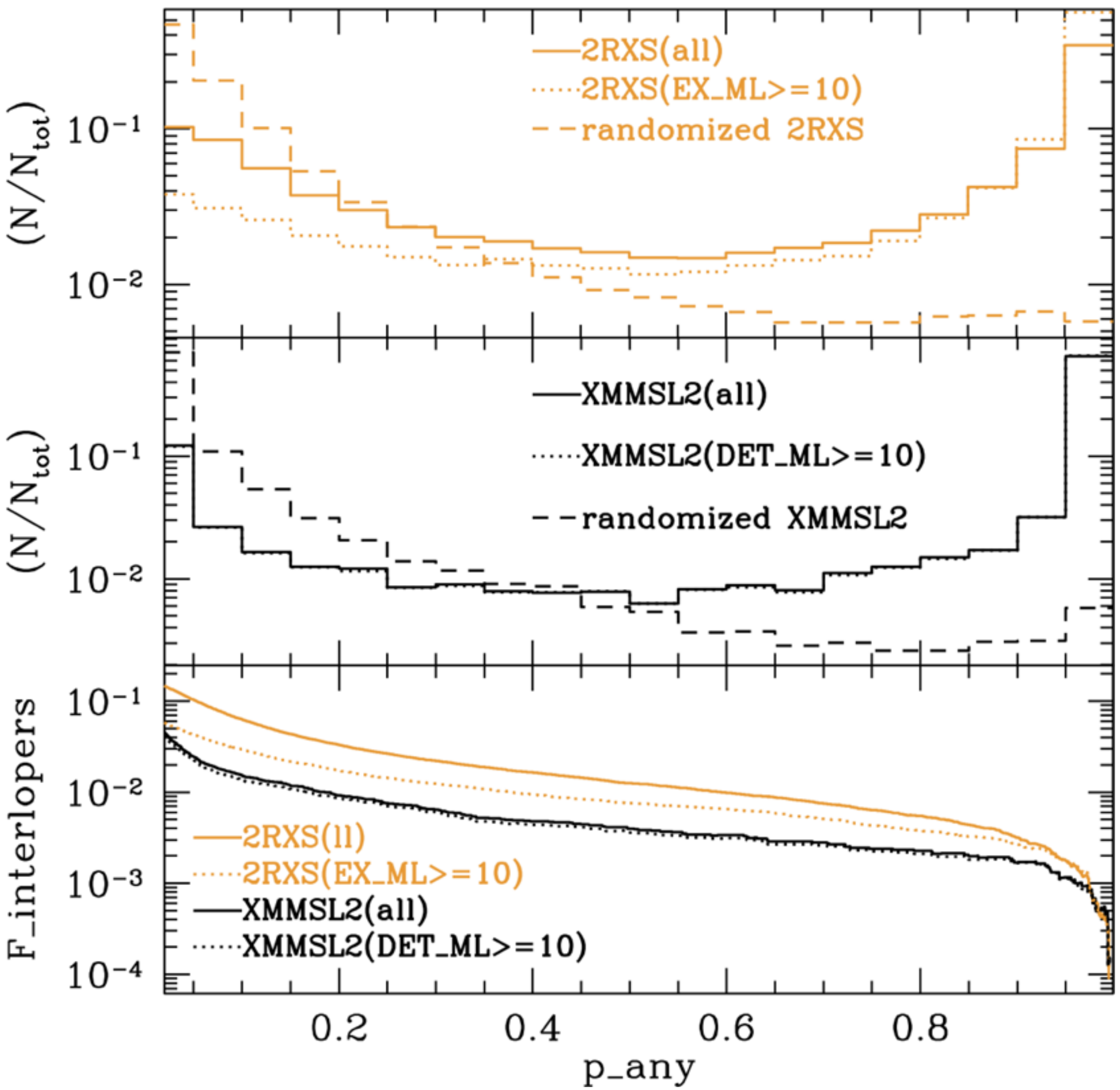
goal: identify features that define X-ray emitters wrt field population

NWAY on ROSAT & XMMSLEW2 (Salvato+ 2018)



Validation: 1500 sources from 3XMM w/ secure ctp, shows 97% accuracy

XMMSL-2RXS Separation arcsec	Sources in common N	Identical AllWISE ctp. %
Sep. ≤5	1111	98.5
Sep. ≤10	3448	98.7
Sep. *le*30	7834	96.1
Sep. *le*60	8768	93.0

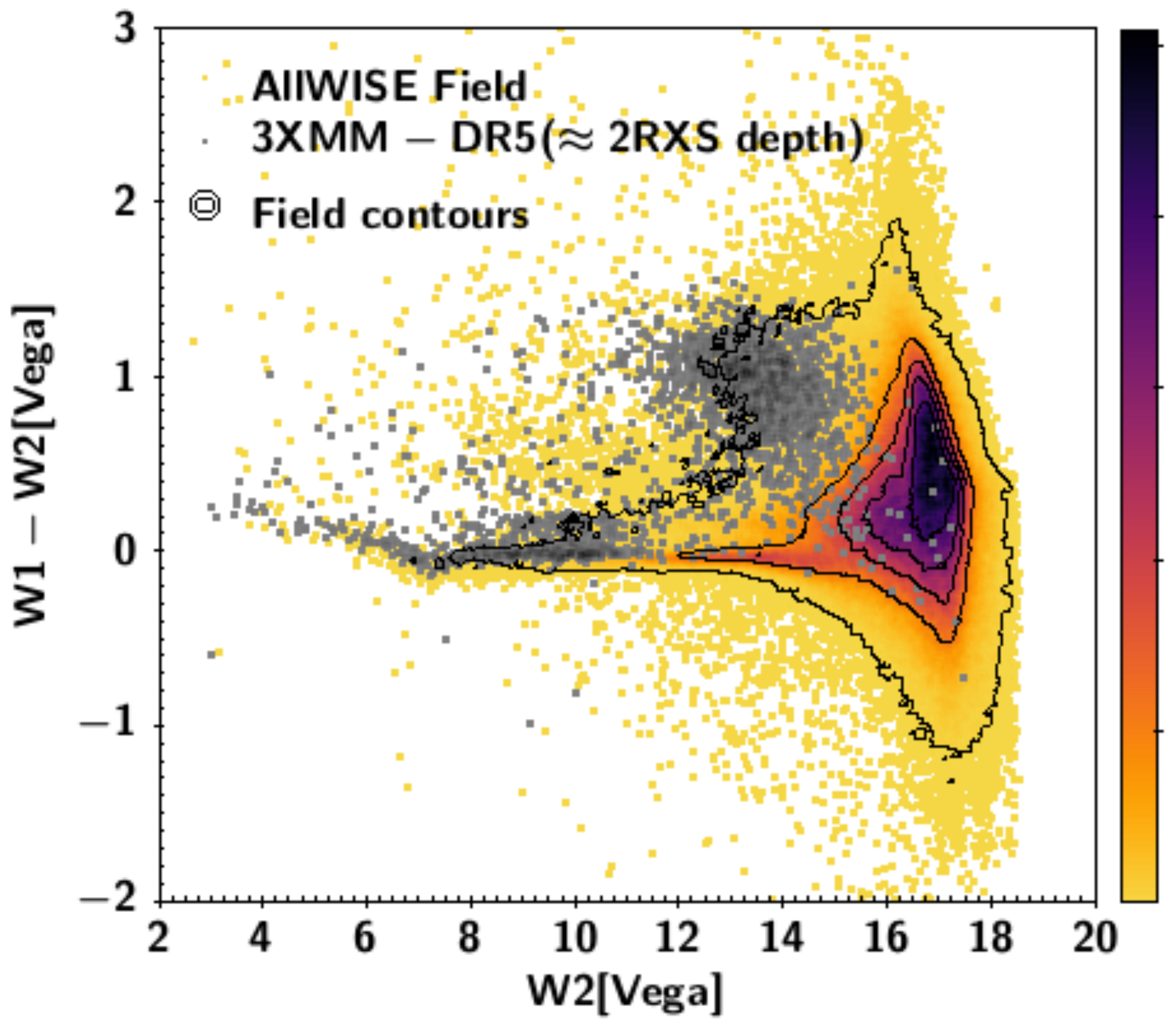


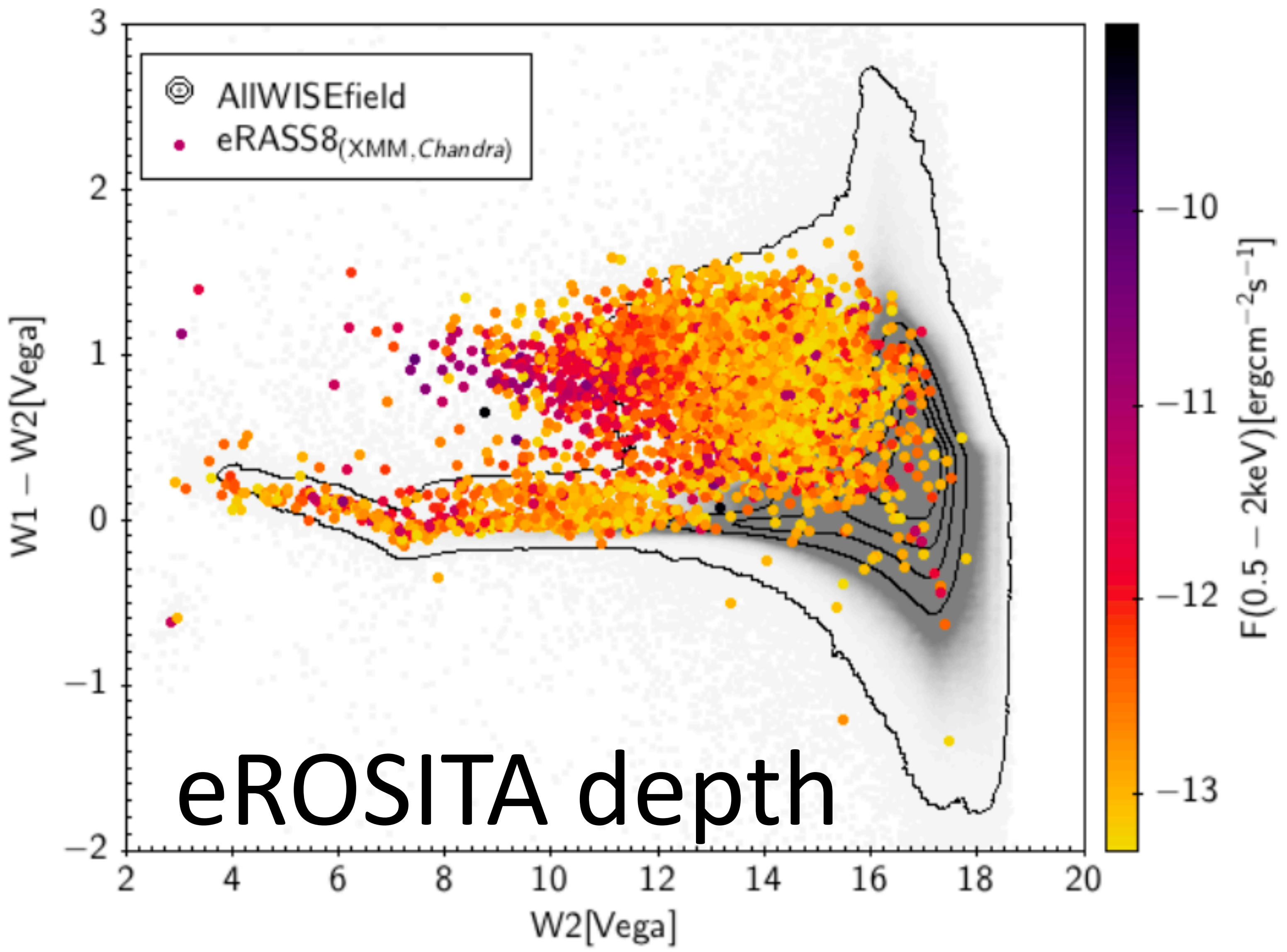
For each problem we need to find the right solution!

W2 vs W1-W2 is NOT ALWAYS the solution!
the parameter space may provide no information

For each problem we need to find the right solution!

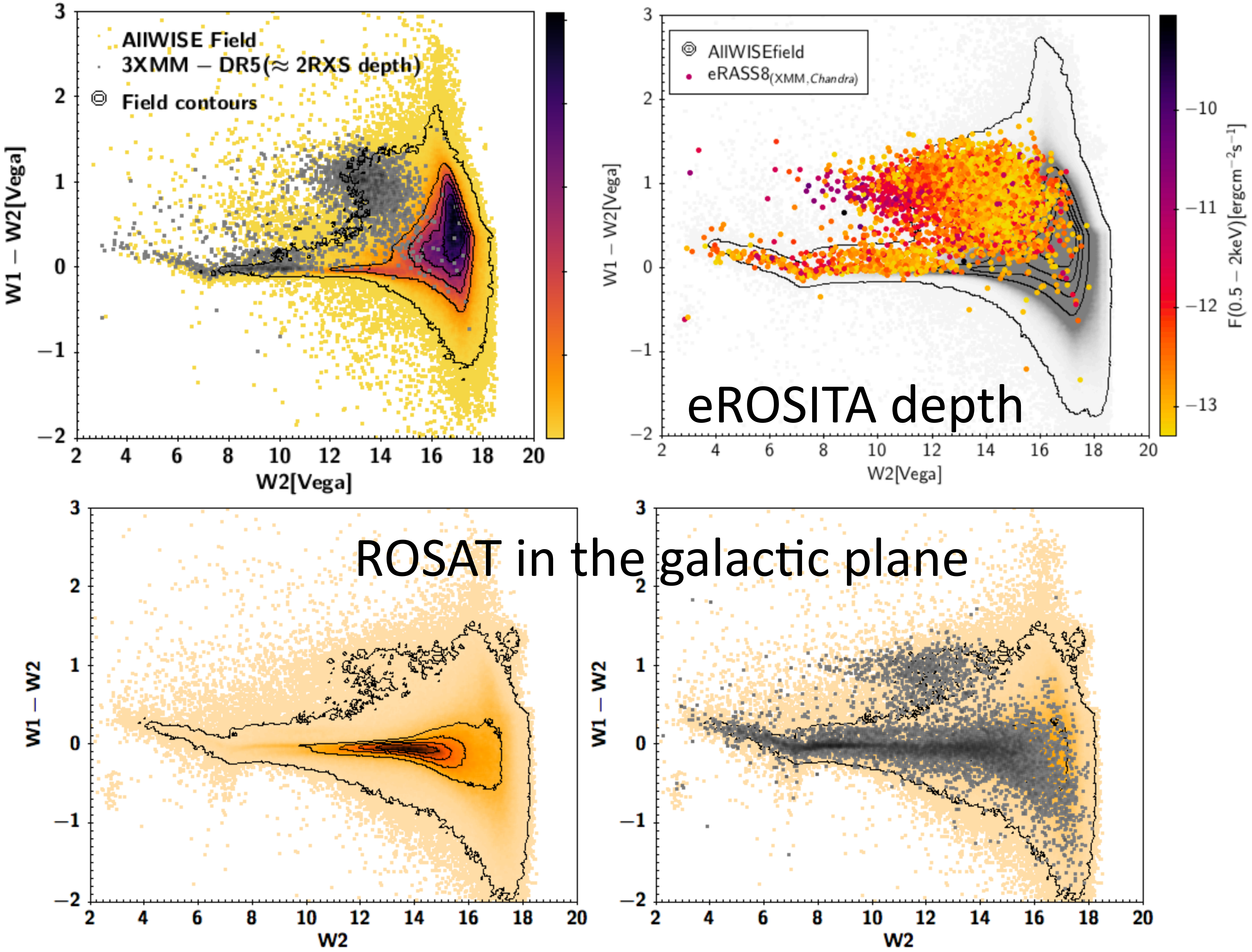
W2 vs W1-W2 is NOT ALWAYS the solution!
the parameter space may provide no information





For each problem we need to find the right solution!

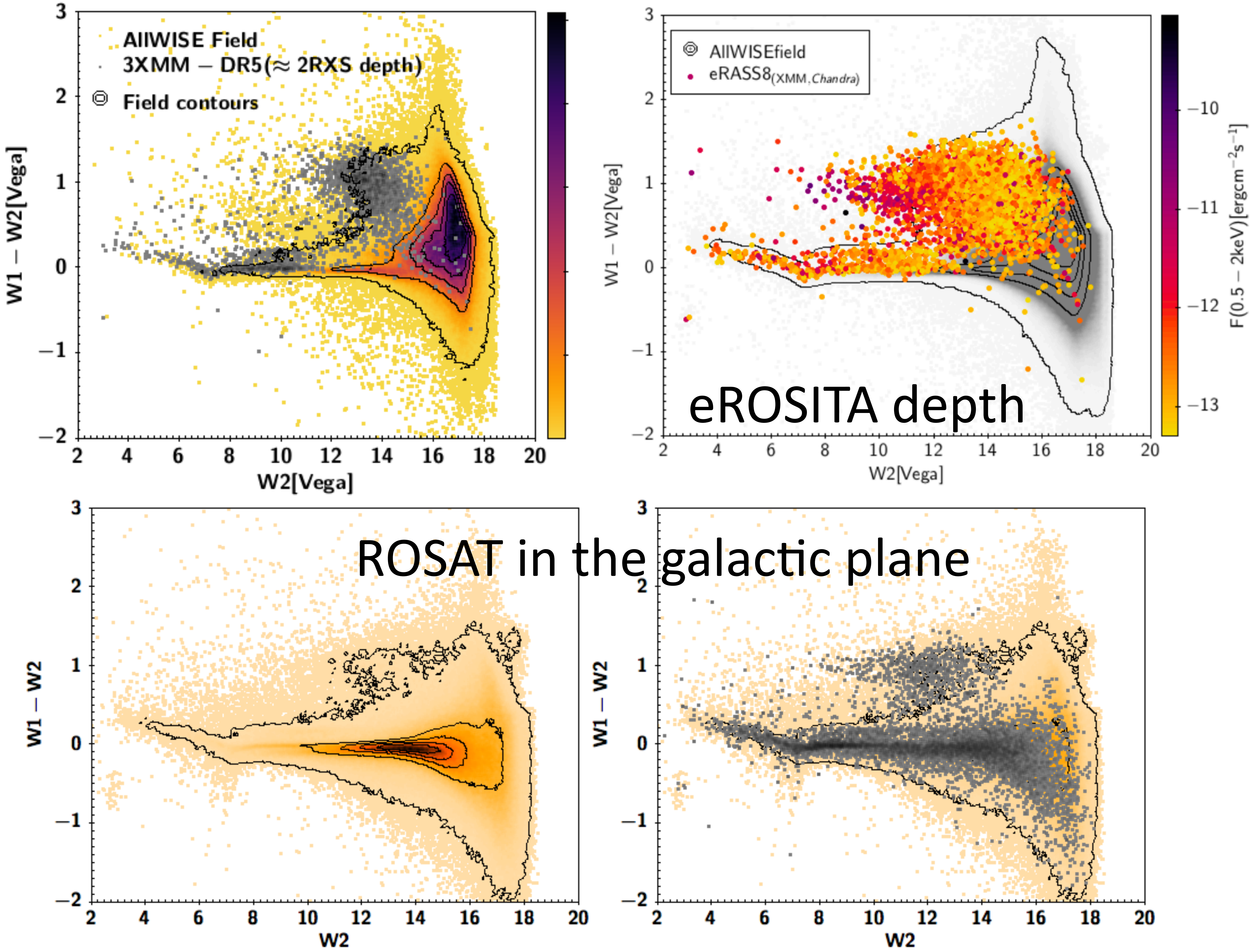
W2 vs W1-W2 is NOT ALWAYS the solution!
the parameter space may provide no information



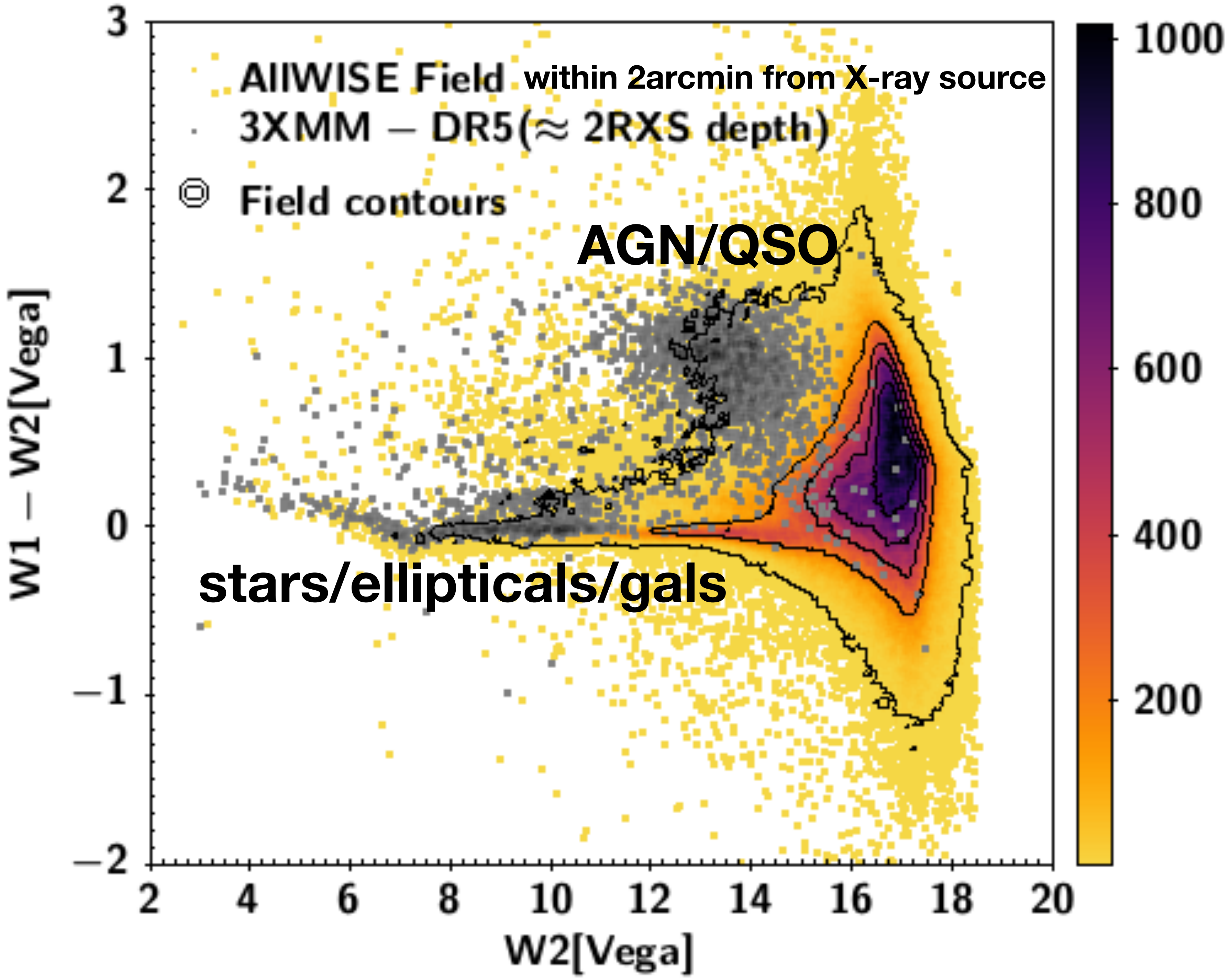
For each problem we need to find the right solution!

W2 vs W1-W2 is NOT ALWAYS the solution!
the parameter space may provide no information

In each case, we need to find the
way to disentangle the actual CTP from the field



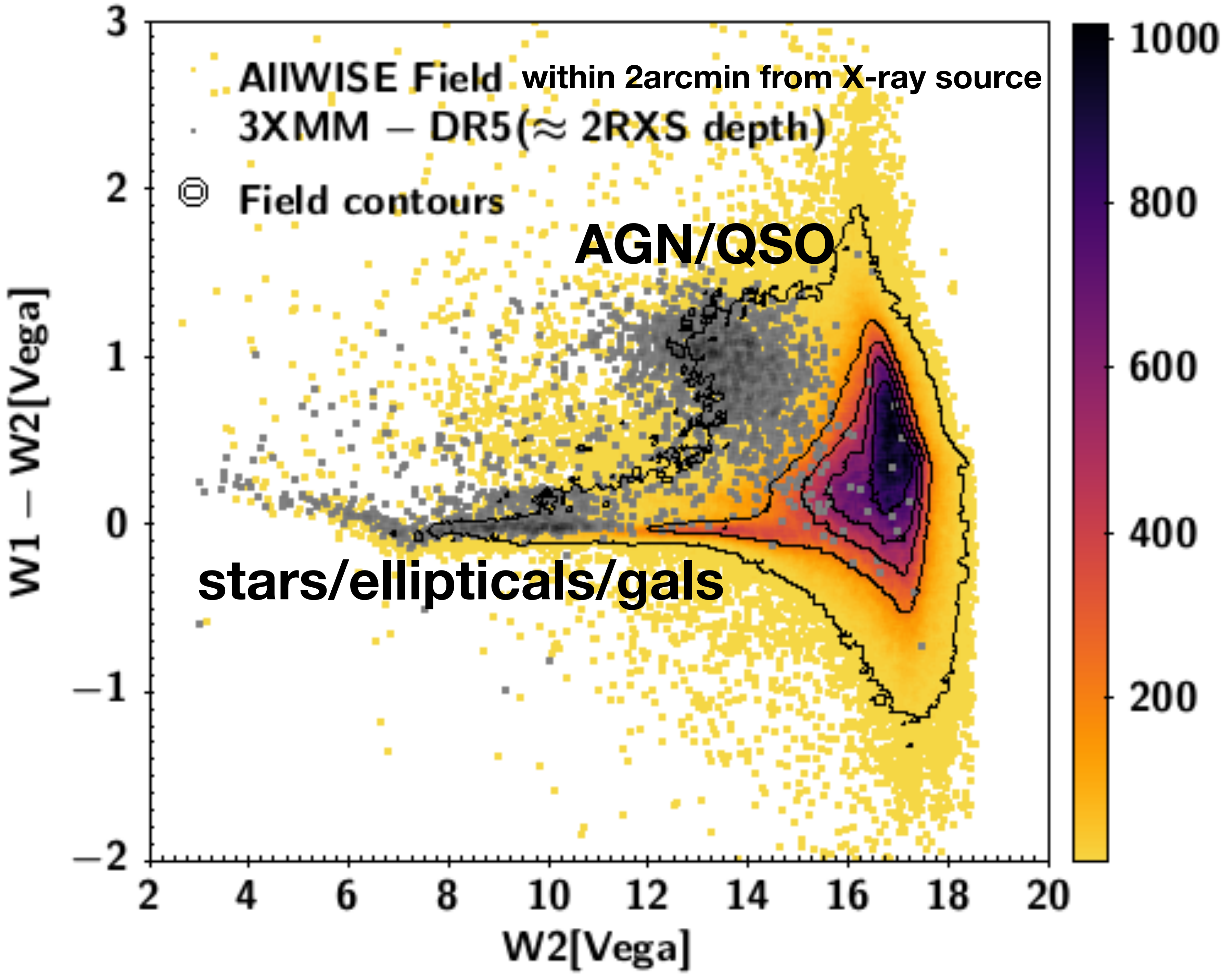
NWAY on ROSAT & XMMSLEW2 (Salvato+ 2018)



NWAY on eROSITA/eFEDS (Salvato+ 2022)

From Brain Learning to Machine Learning

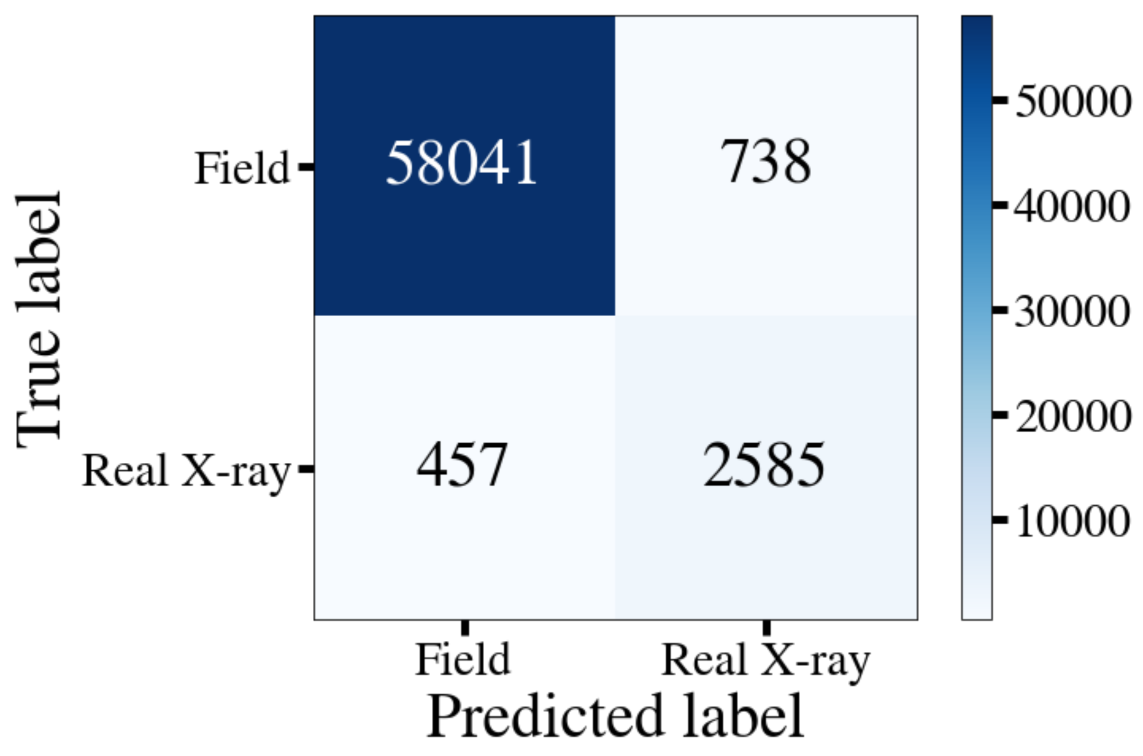
NWAY on ROSAT & XMMSLEW2 (Salvato+ 2018)



NWAY on eROSITA/eFEDS (Salvato+ 2022)

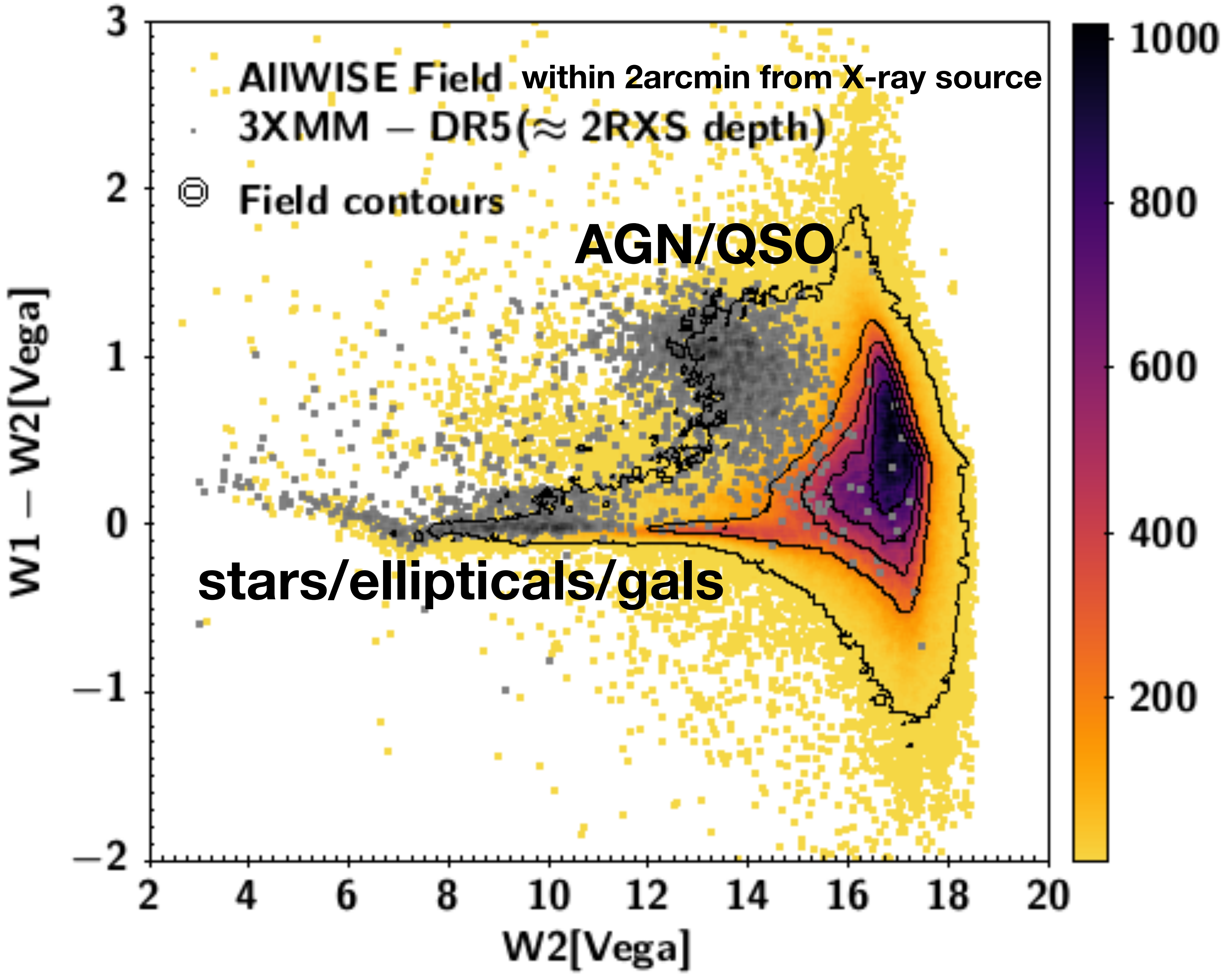
- Random Forest on :
  - training sample: 23K XMM sources with depth comparable to eFEDS and w/ secure CTP in Legacy Survey DR8 (Dey+2019).
  - control sample: the rest of the sources within 30” from each X-ray position

Feature	Description
flux_*/mw_transmission_*	deredenned flux in $g,r,z,W1,W2$
gaia_phot_*_mean_mag	original GAIA phot. in G, G_{bp}, G_{rp}
snr_*	S/N for $g,r,z,W1,W2,G,G_{bp},G_{rp}$
$\sqrt{pmra^2 + pmdec^2}$	Gaia proper motion
parallax	Gaia paralllax
g-r, r-z, z-$W1$, r-$W2$	dereddened colors



From Brain Learning to Machine Learning

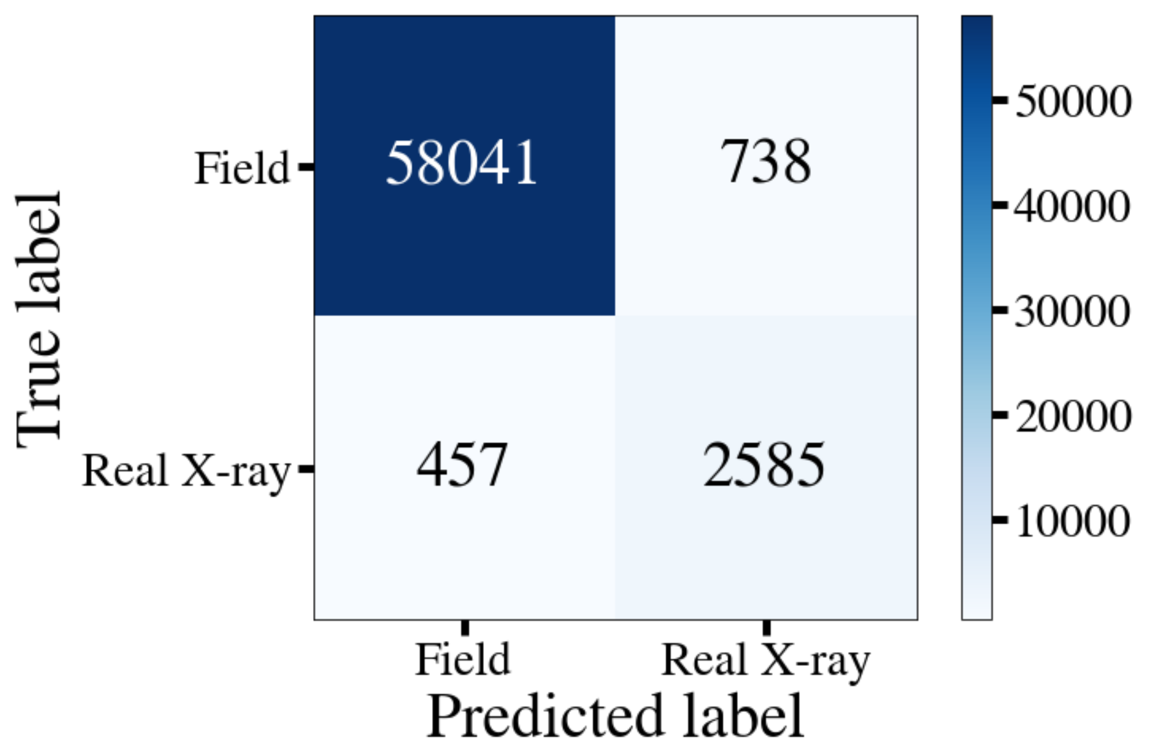
NWAY on ROSAT & XMMSLEW2 (Salvato+ 2018)



NWAY on eROSITA/eFEDS (Salvato+ 2022)

- Random Forest on :
  - training sample: 23K XMM sources with depth comparable to eFEDS and w/ secure CTP in Legacy Survey DR8 (Dey+2019).
  - control sample: the rest of the sources within 30” from each X-ray position

Feature	Description
flux_*/mw_transmission_*	deredenned flux in $g,r,z,W1,W2$
gaia_phot_*_mean_mag	original GAIA phot. in G, G_{bp}, G_{rp}
snr_*	S/N for $g,r,z,W1,W2,G,G_{bp},G_{rp}$
$\sqrt{pmra^2 + pmdec^2}$	Gaia proper motion
parallax	Gaia paralllax
g-r, r-z, z-$W1$, r-$W2$	dereddened colors

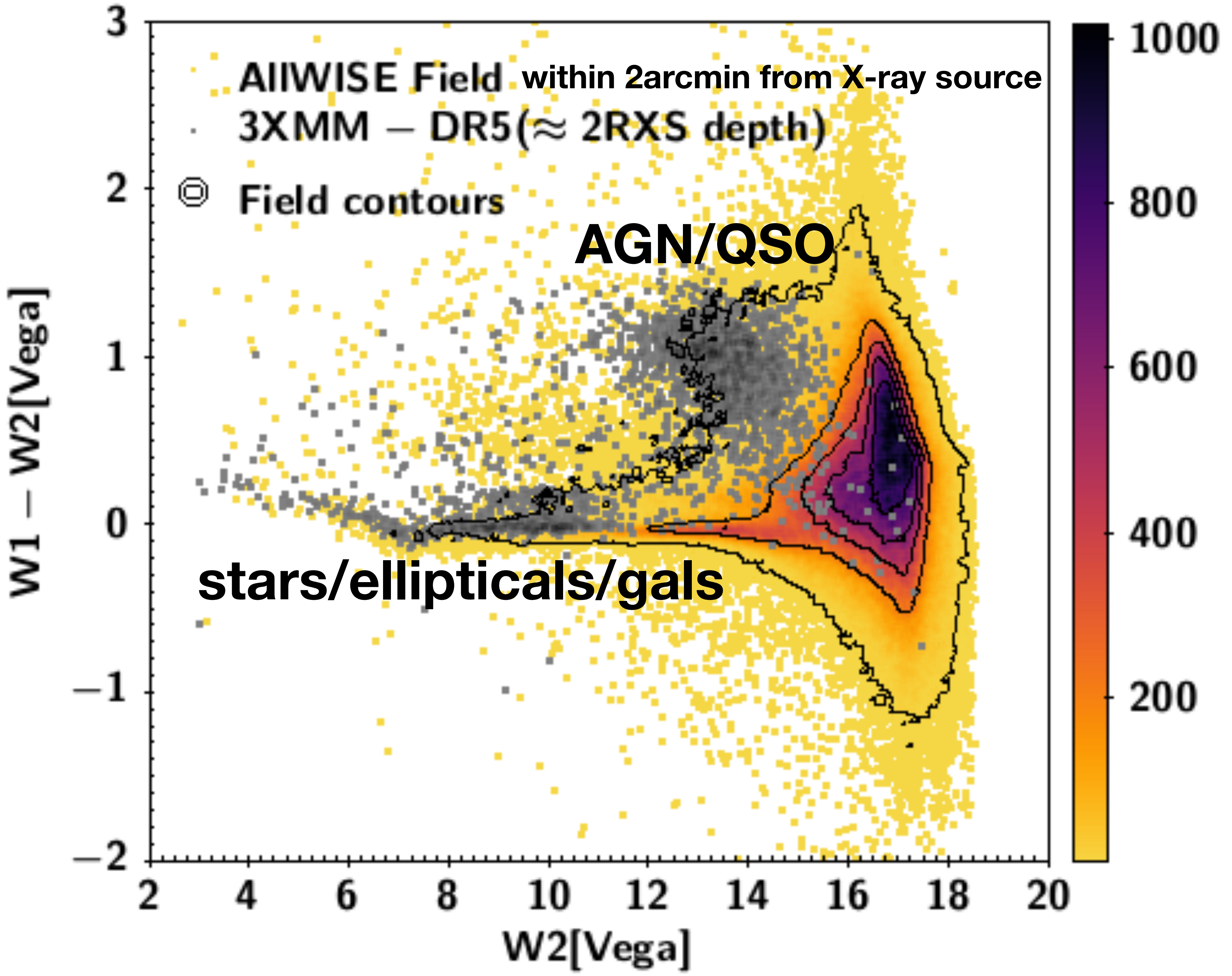


- validation sample: 3500 Chandra sources with depth comparable to eFEDS, **made eROSITA-like**

From Brain Learning to Machine Learning

eROSITA

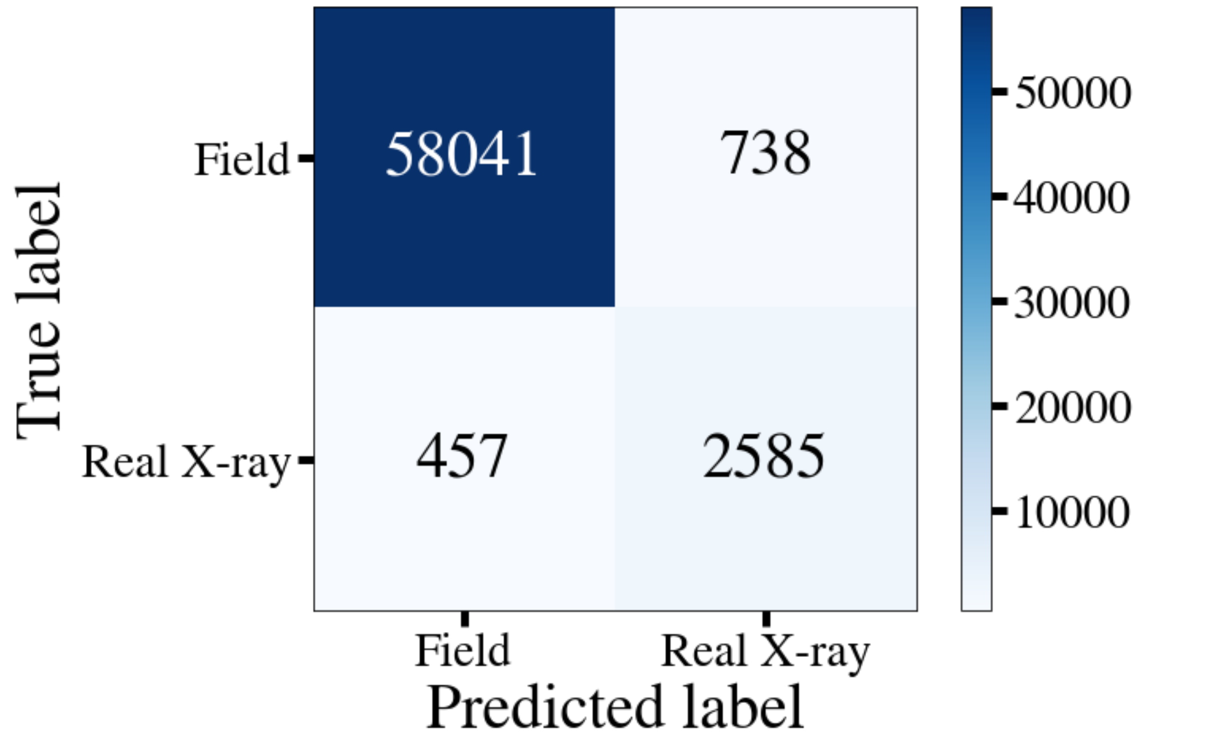
NWAY on ROSAT & XMMSLEW2 (Salvato+ 2018)



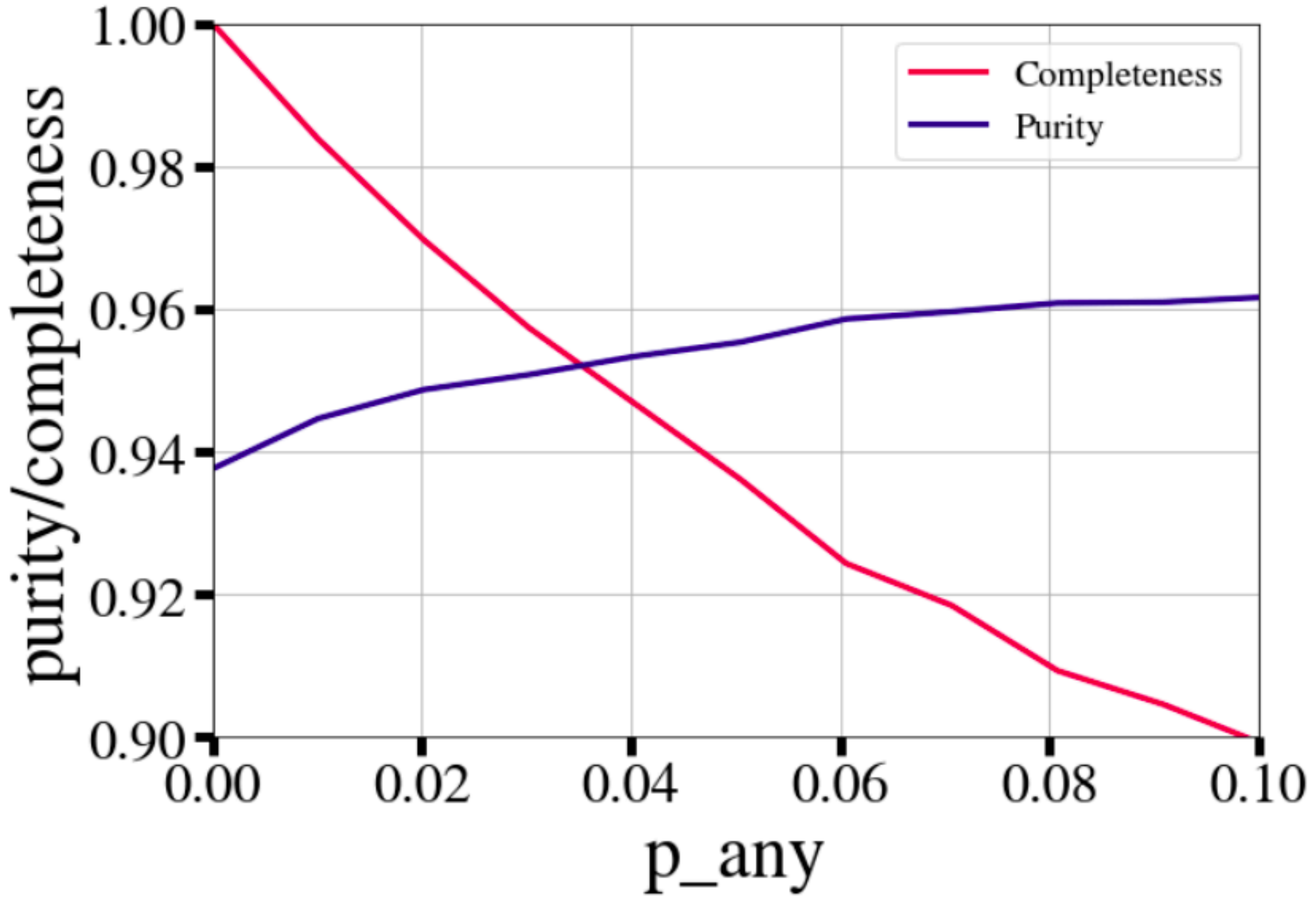
NWAY on eROSITA/eFEDS (Salvato+ 2022)

- Random Forest on :
  - training sample: 23K XMM sources with depth comparable to eFEDS and w/ secure CTP in Legacy Survey DR8 (Dey+2019).
  - control sample: the rest of the sources within 30" from each X-ray position

Feature	Description
flux_*/mw_transmission_*	deredenned flux in $g, r, z, W1, W2$
gaia_phot_*_mean_mag	original GAIA phot. in G, G_{bp}, G_{rp}
snr_*	S/N for $g, r, z, W1, W2, G, G_{bp}, G_{rp}$
$\sqrt{pmra^2 + pmdec^2}$	Gaia proper motion
parallax	Gaia paralllax
g-r, r-z, z-$W1$, r-$W2$	dereddened colors



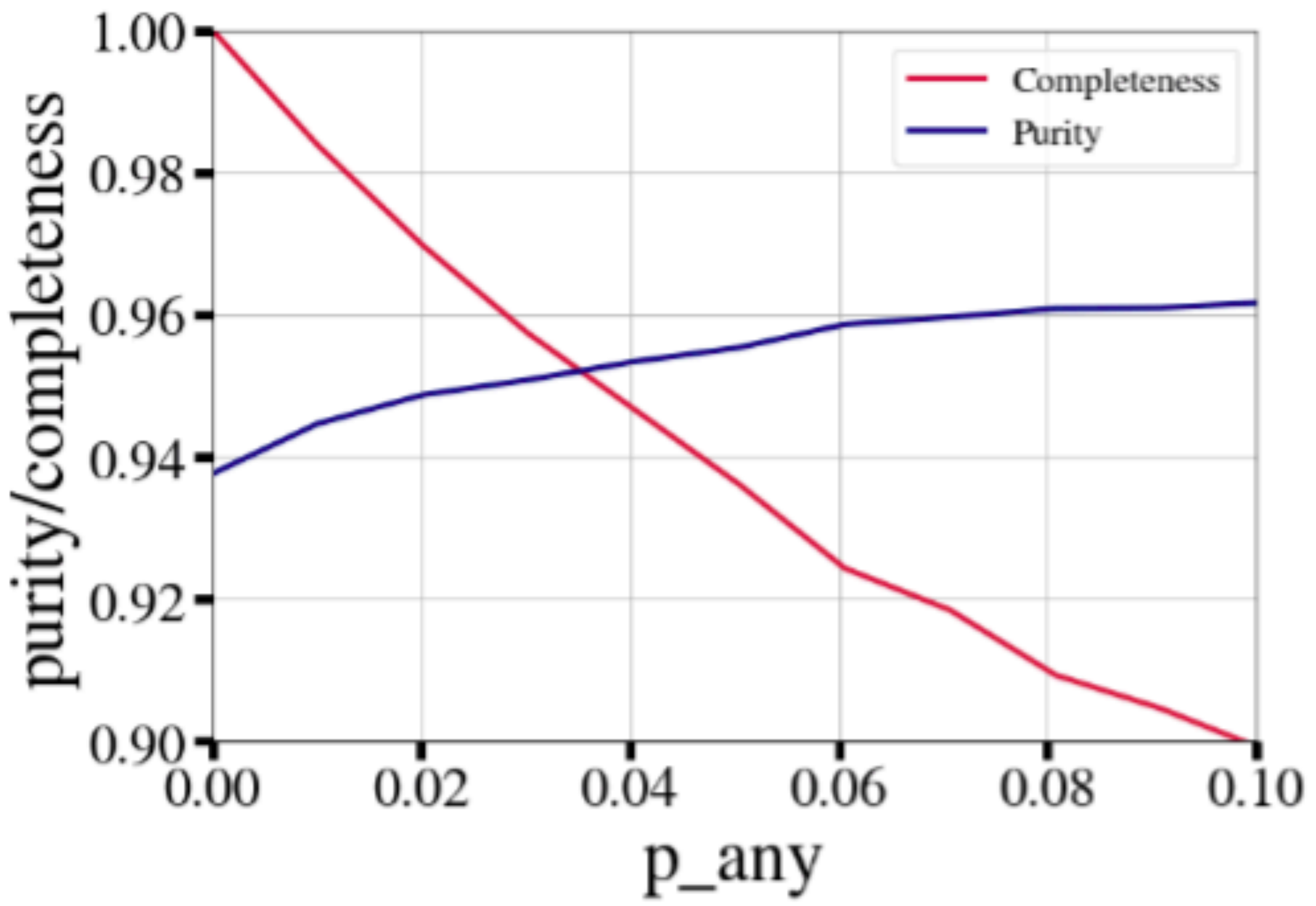
- validation sample: 3500 Chandra sources with depth comparable to eFEDS, **made eROSITA-like**



Comparison between methods

NWAY

Feature	Description
flux_*/mw_transmission_*	deredenned flux in $g,r,z,W1,W2$
gaia_phot_*_mean_mag	original GAIA phot. in G, G_{bp}, G_{rp}
snr_*	S/N for $g,r,z,W1,W2,G,G_{bp},G_{rp}$
$\sqrt{pmra^2 + pmdec^2}$	Gaia proper motion
parallax	Gaia paralllax
g-r, r-z, z-$W1$, r-$W2$	dereddened colors



95% correct ctp

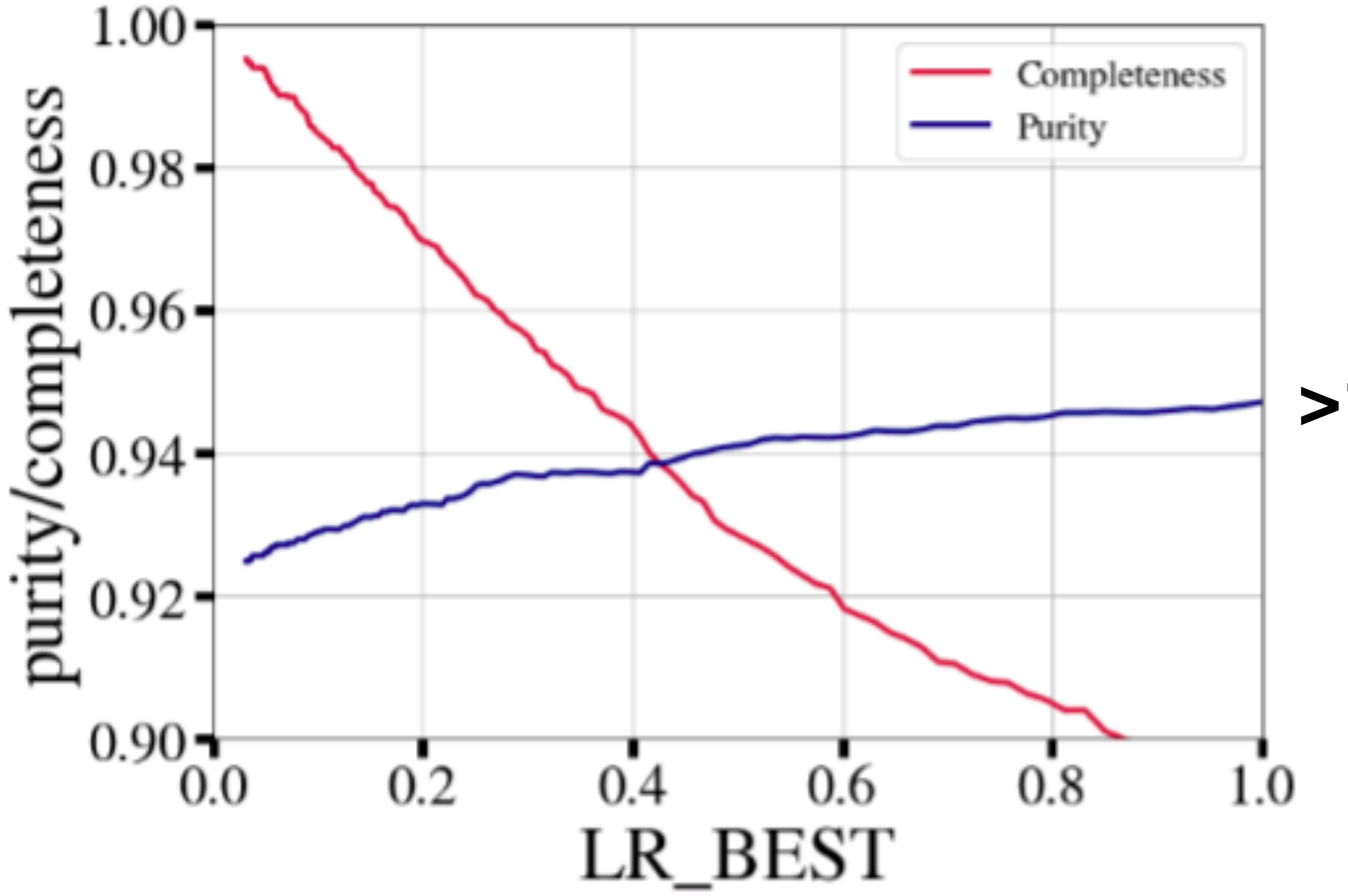
2% w/ second possible CTP

MLR in Astromatch (Ruiz+2018)

Using the same training, control and validation samples

- 3-D distribution W2, W1-W2, TYPE
- 3-D distribution r-W2, g, TYPE
- g band

then select the CTP with higher LR from one of the 3 methods



89% correct ctp

>10% w/ second possible CTP

88% agreement, with fraction of disagrement increasing with the positional error

(details in Salvato+2022)

Comparison between assumptions

Searching for a specific type of emitters: the case of coronal stars

Schneider et al 2022

- Gaia EDR3 that
- are brighter than 19th mag in G band
- have accurate magnitudes in all three Gaia photometric bands
- have a parallax/parallax_err >3

NWAY (no assumption on the type of counterparts)

Salvato et al 2022

Feature	Description
flux_*/mw_transmission_*	deredenned flux in $g, r, z, W1, W2$
gaia_phot_*_mean_mag	original GAIA phot. in G, G_{bp}, G_{rp}
snr_*	S/N for $g, r, z, W1, W2, G, G_{bp}, G_{rp}$
$\sqrt{pmra^2 + pmdec^2}$	Gaia proper motion
parallax	Gaia paralllax
g-r, r-z, z-$W1$, r-$W2$	dereddened colors

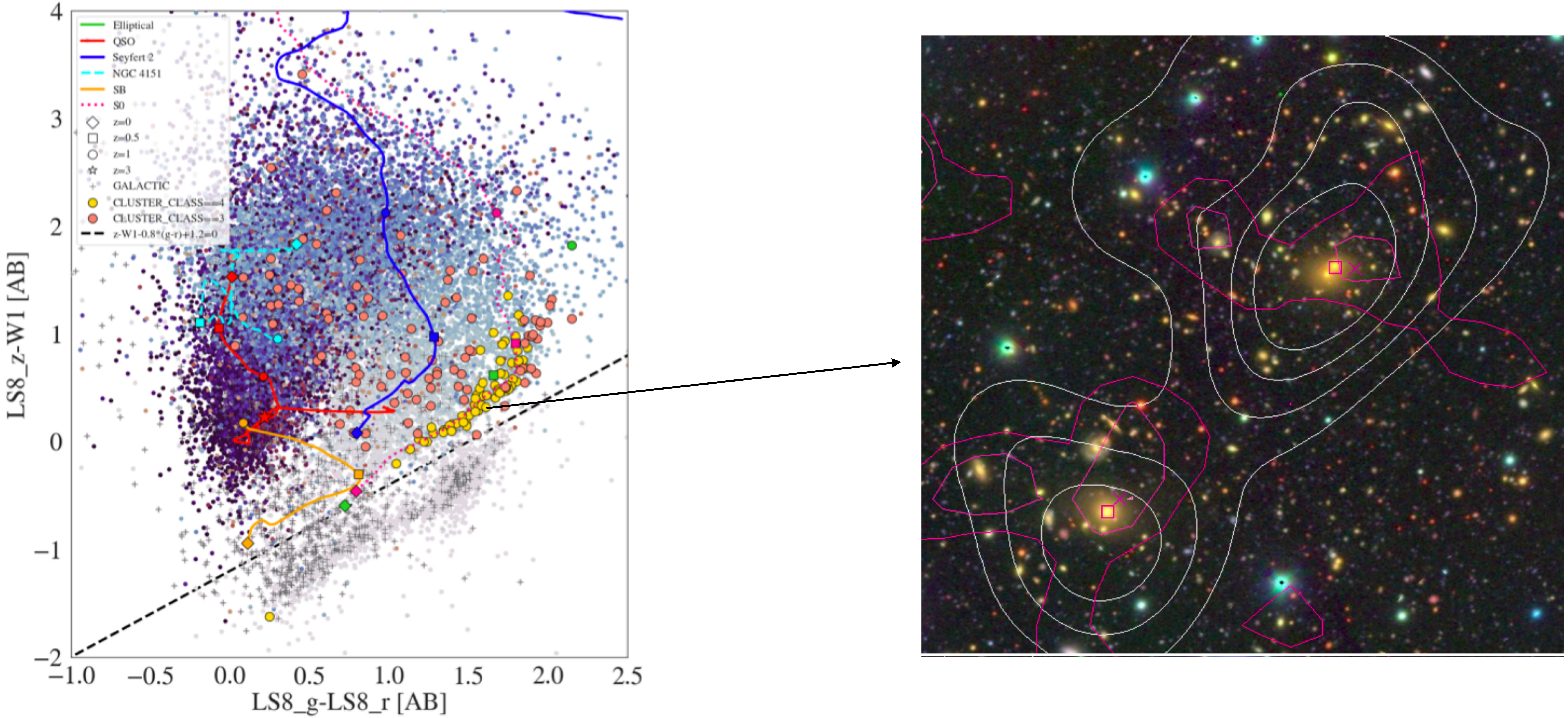
Comparing the Results:
2060 coronal stars
1912 also found by NWAY (~92%)
177/2060 could also be an AGN

For 10% of the stellar sources, probably the X-ray flux is the sum of two contributions (galactic+extragalactic)

(details in Salvato+2022)

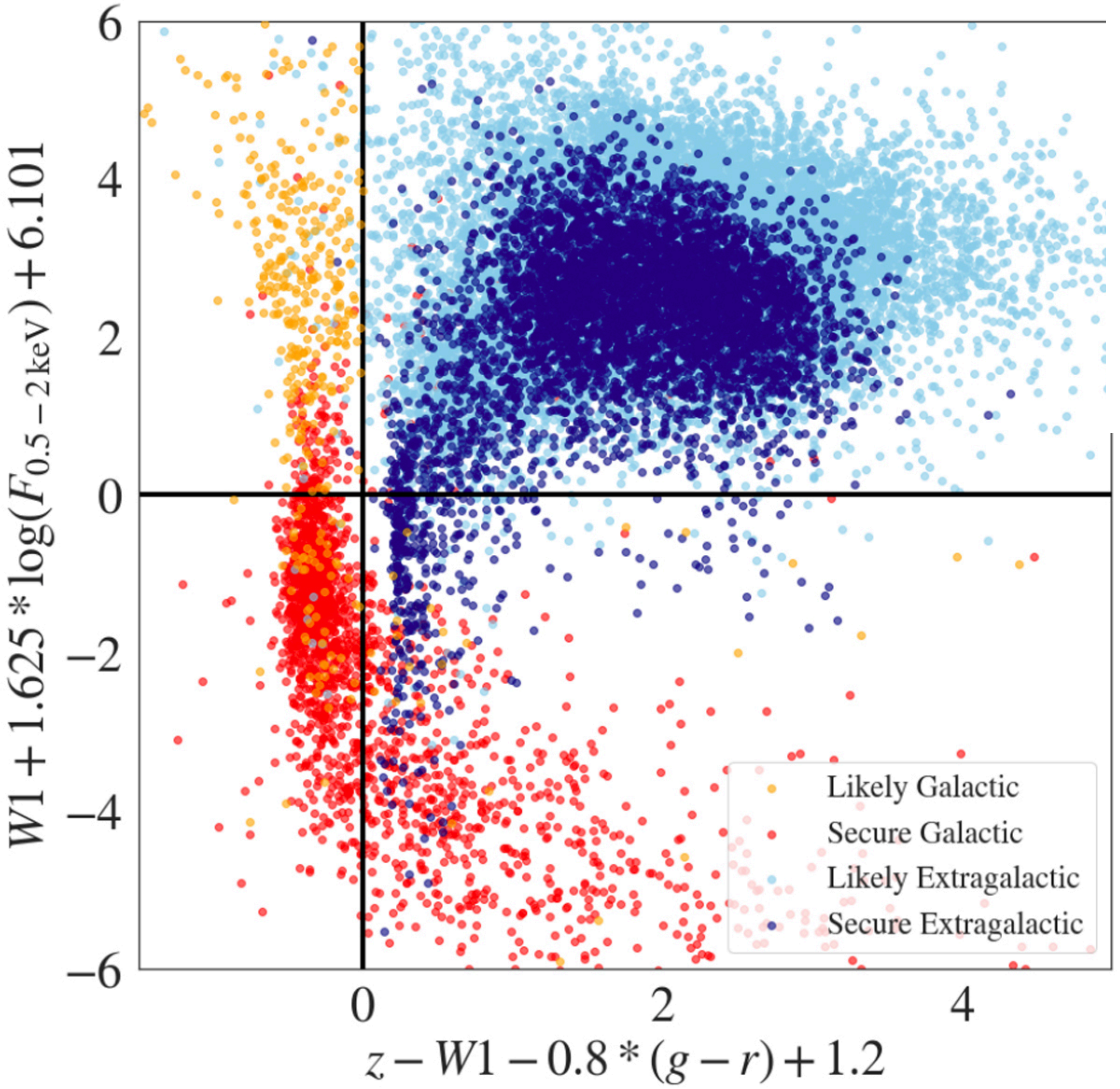
Be open to surprises

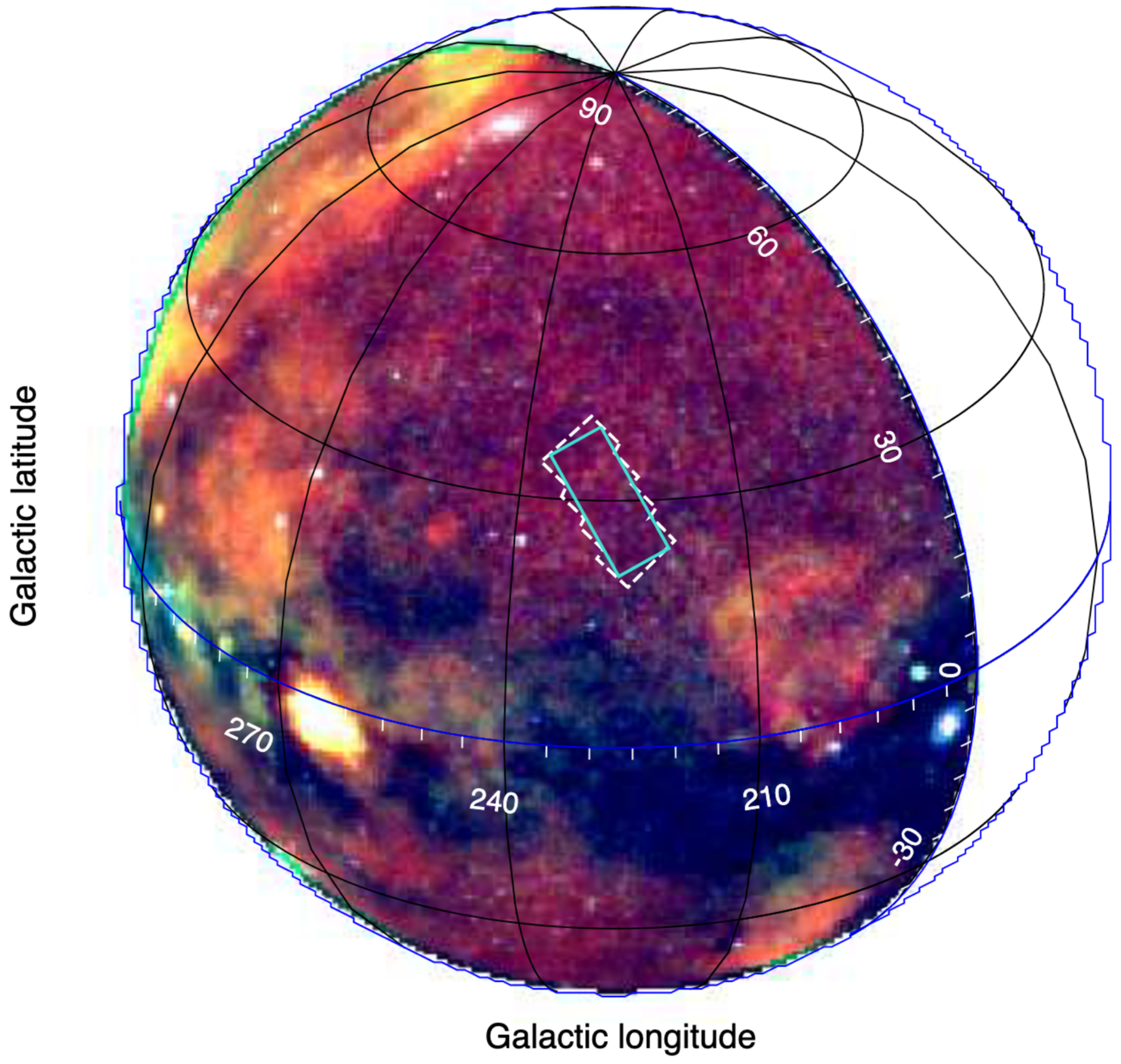
- Making first the CTPs identification and then their classification, allow pinpointing interesting populations (e.g., unresolved clusters)



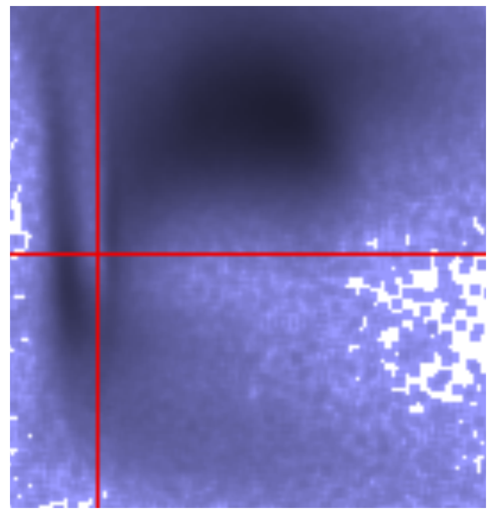
It is always a work in progress

22k eROSITA/eFEDS



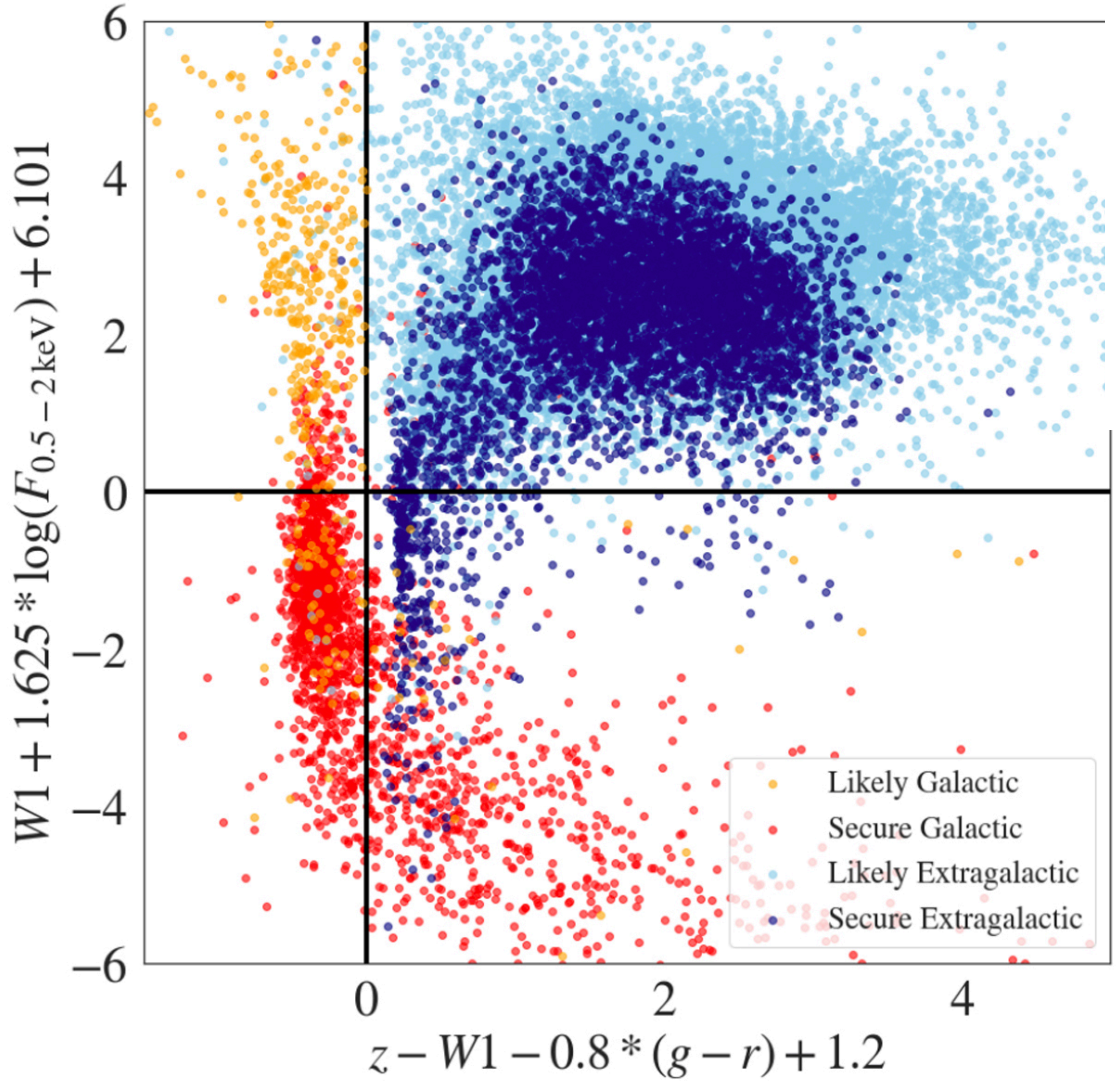


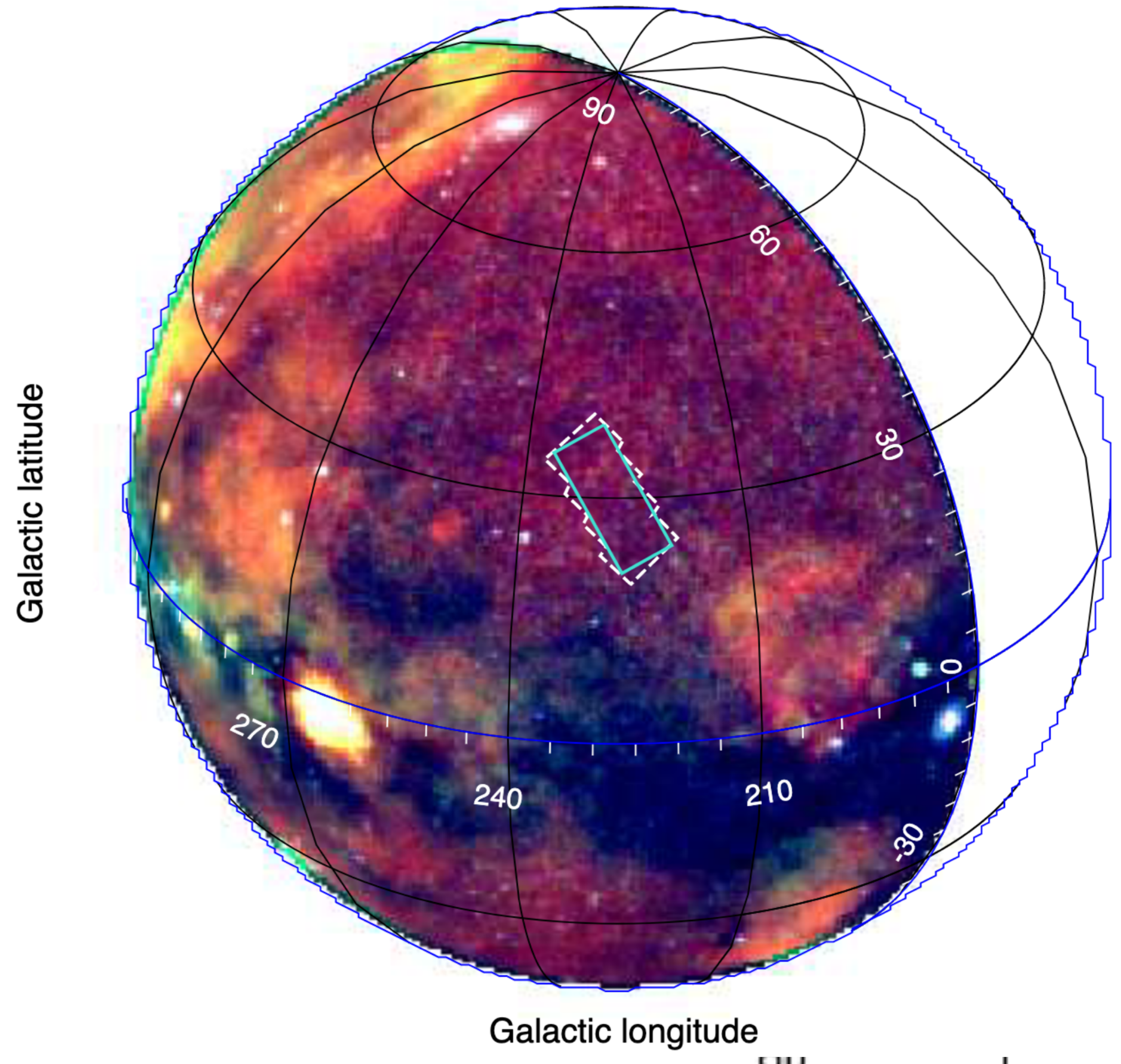
730k eRASS1 sources at DET_LIKE>7



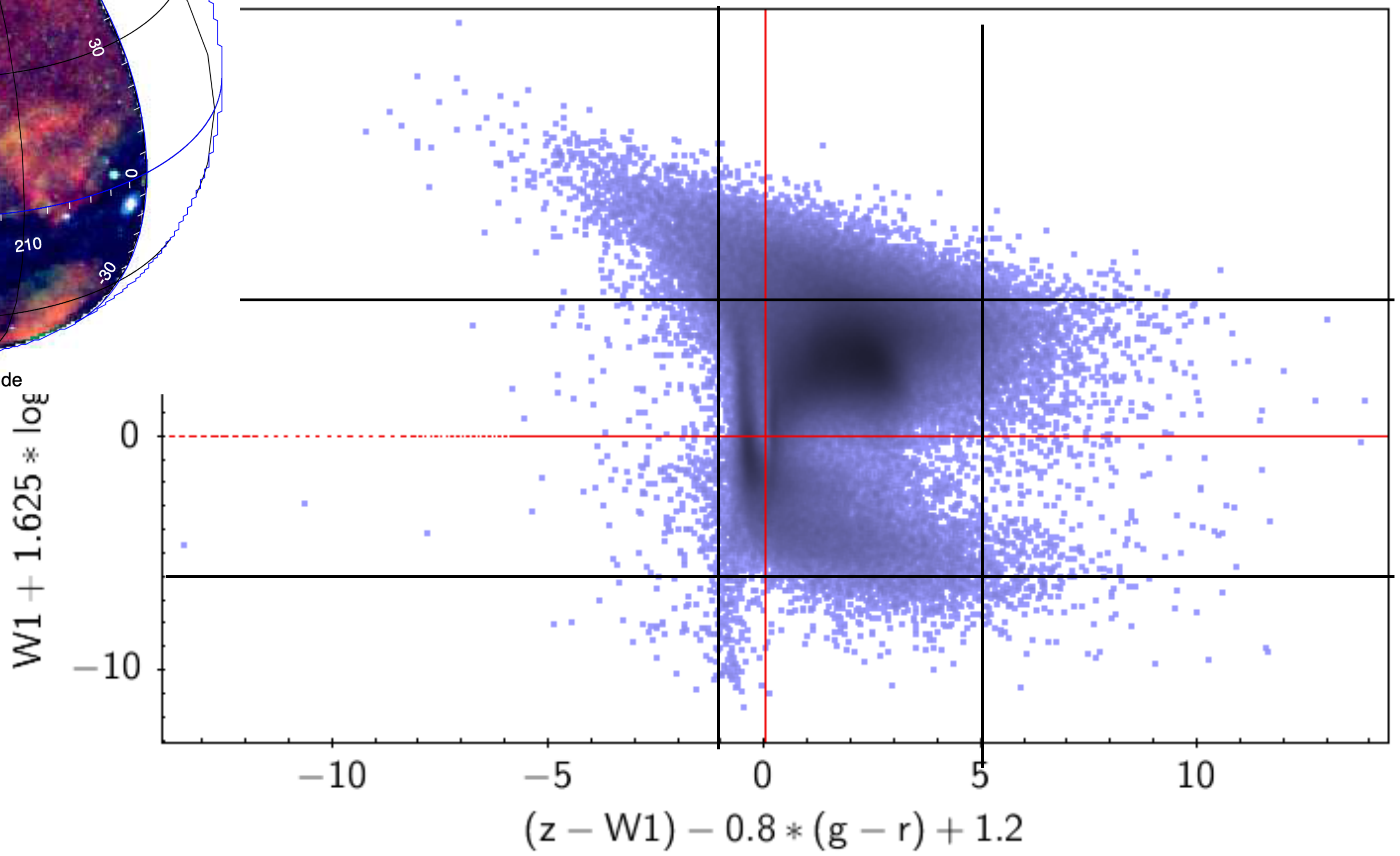
It is always a work in progress

22k eROSITA/eFEDS



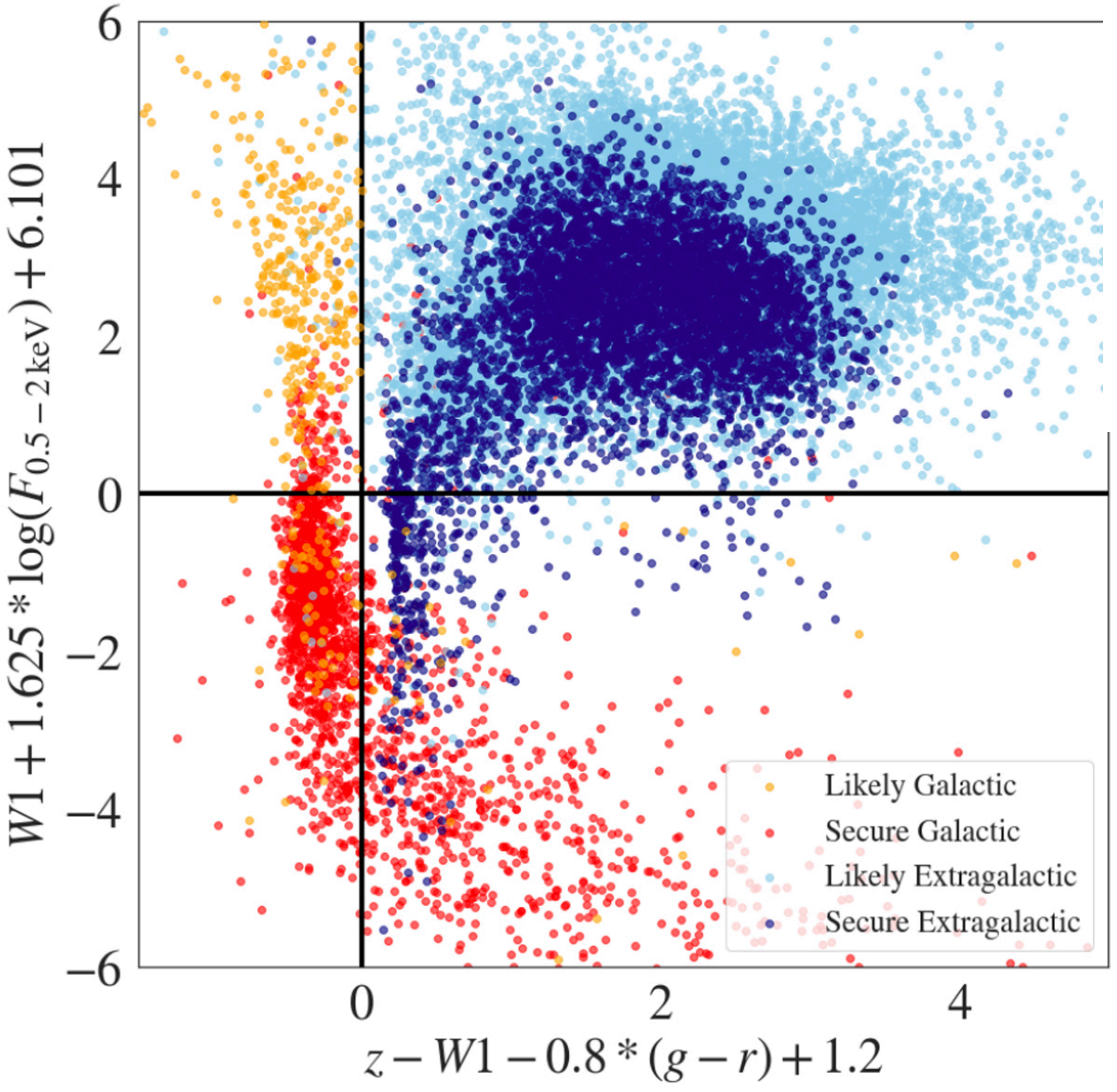


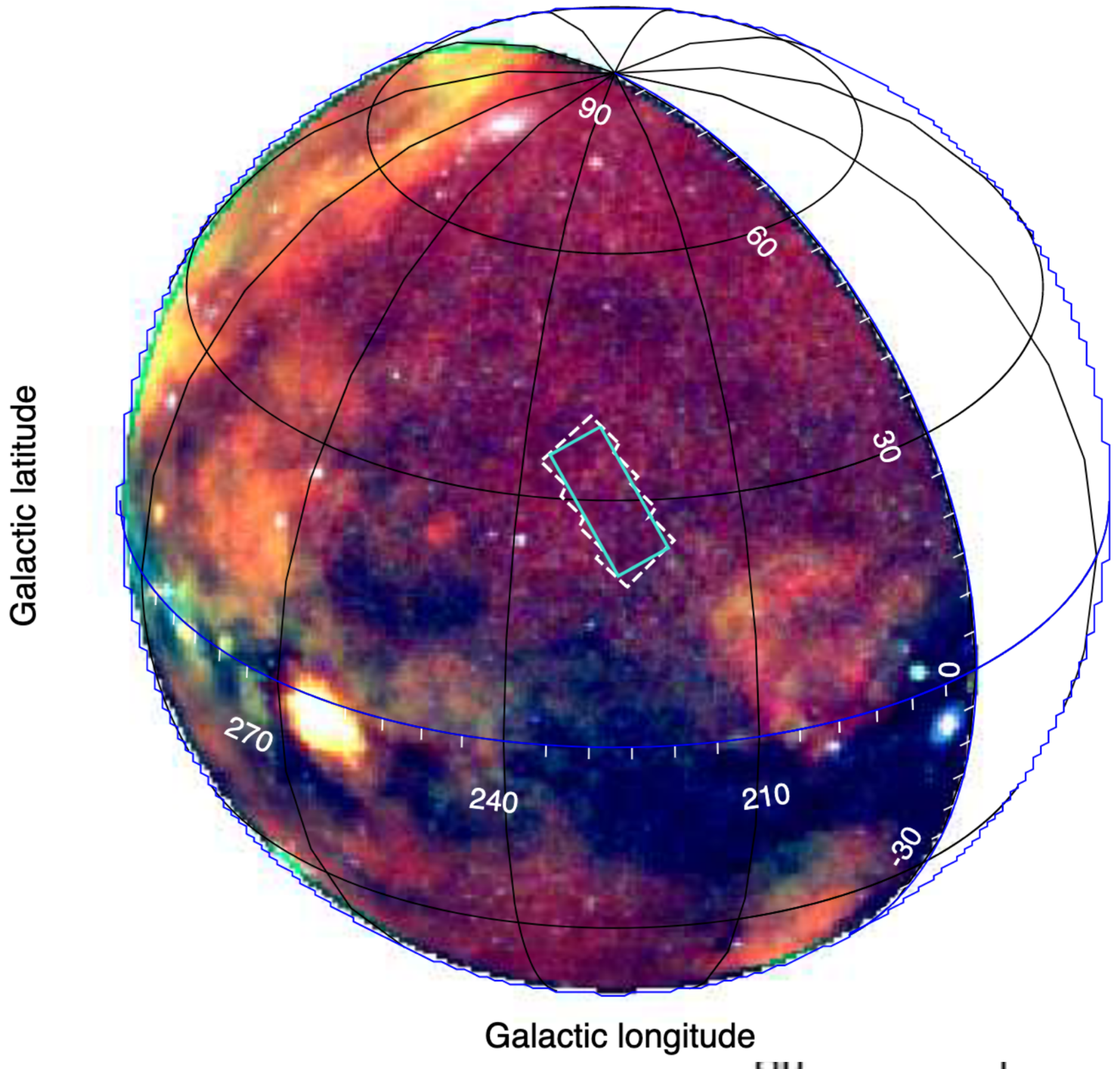
730k eRASS1 sources at DET_LIKE>7



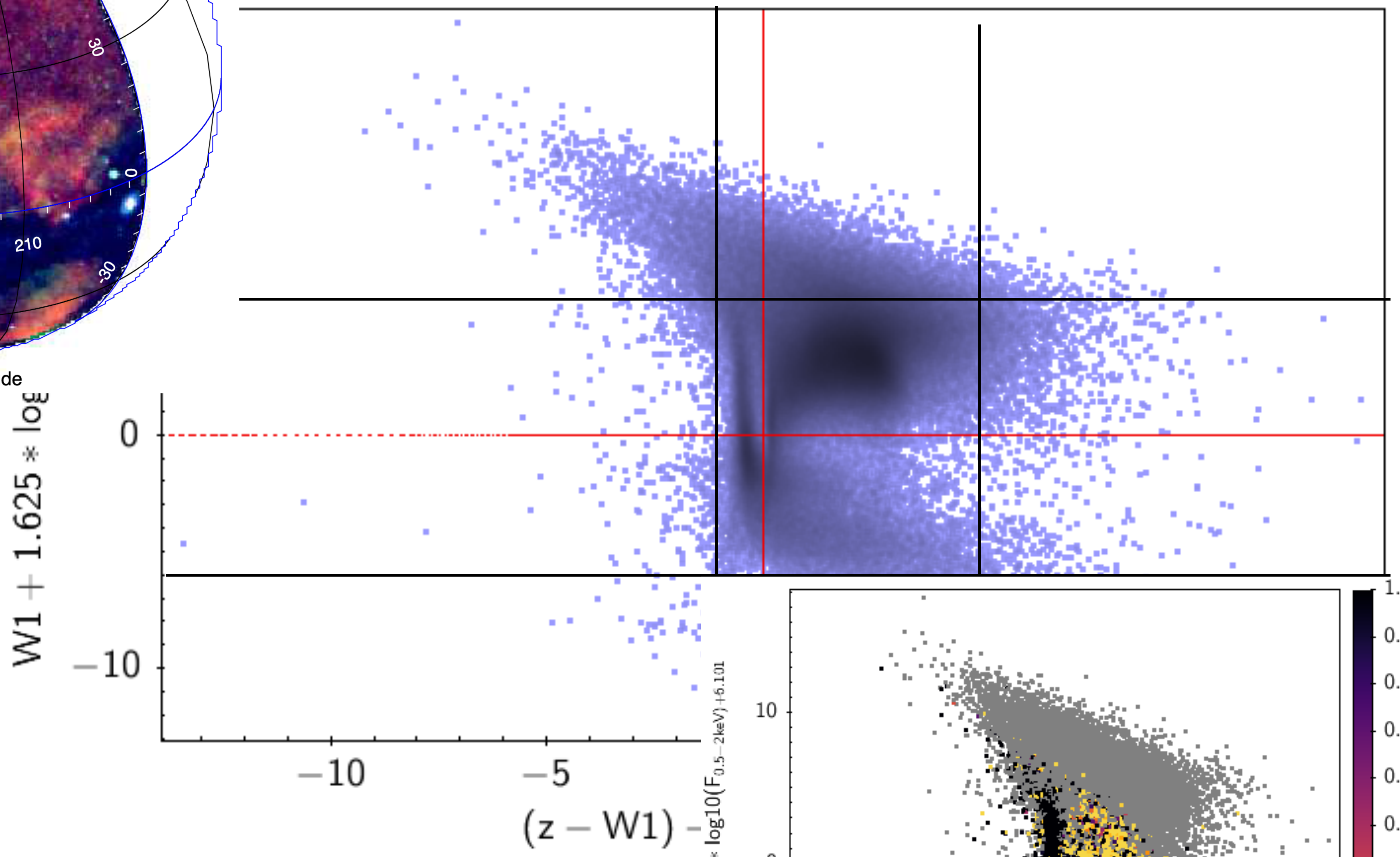
eROSITA

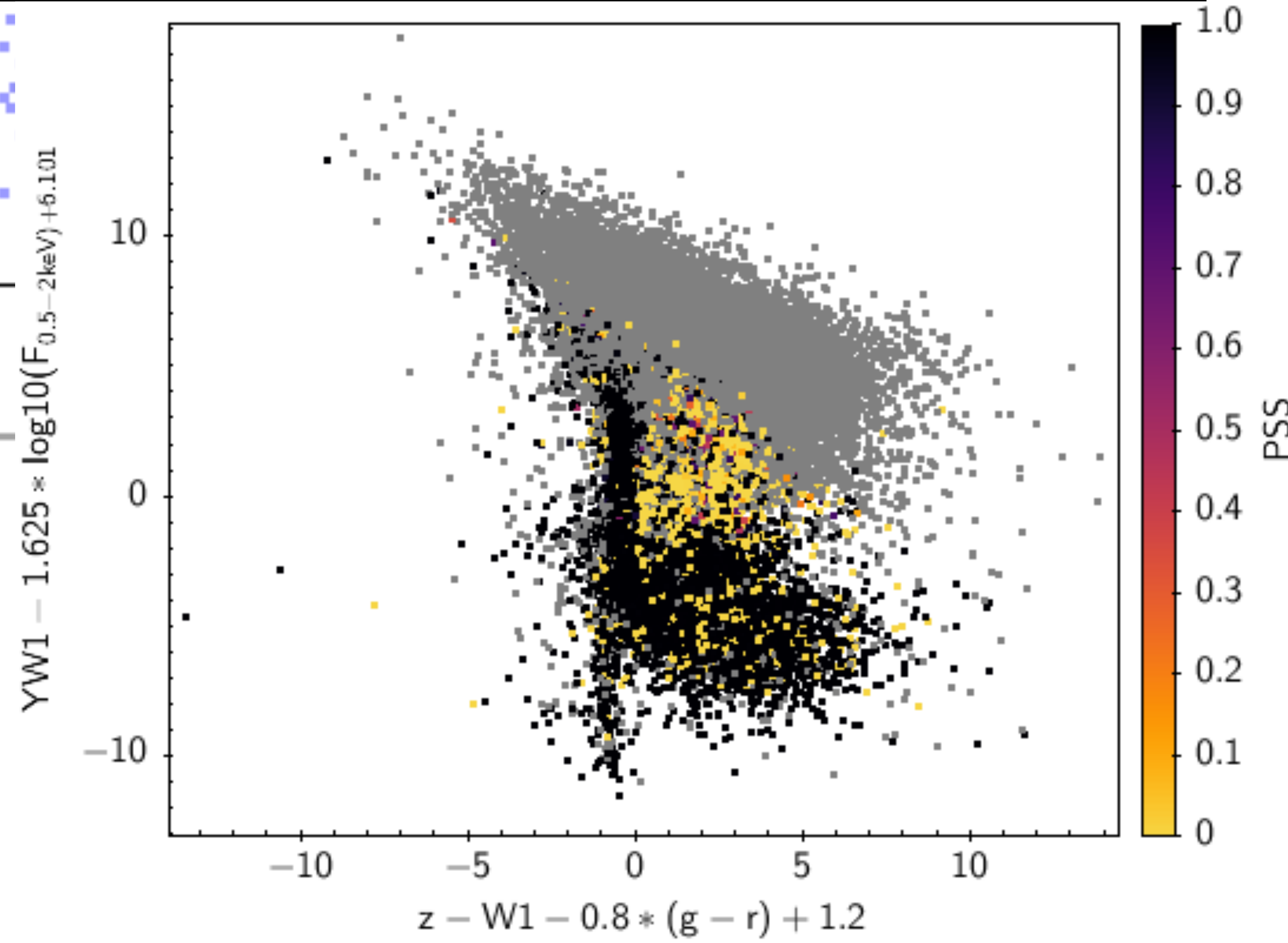
22k eROSITA/eFEDS



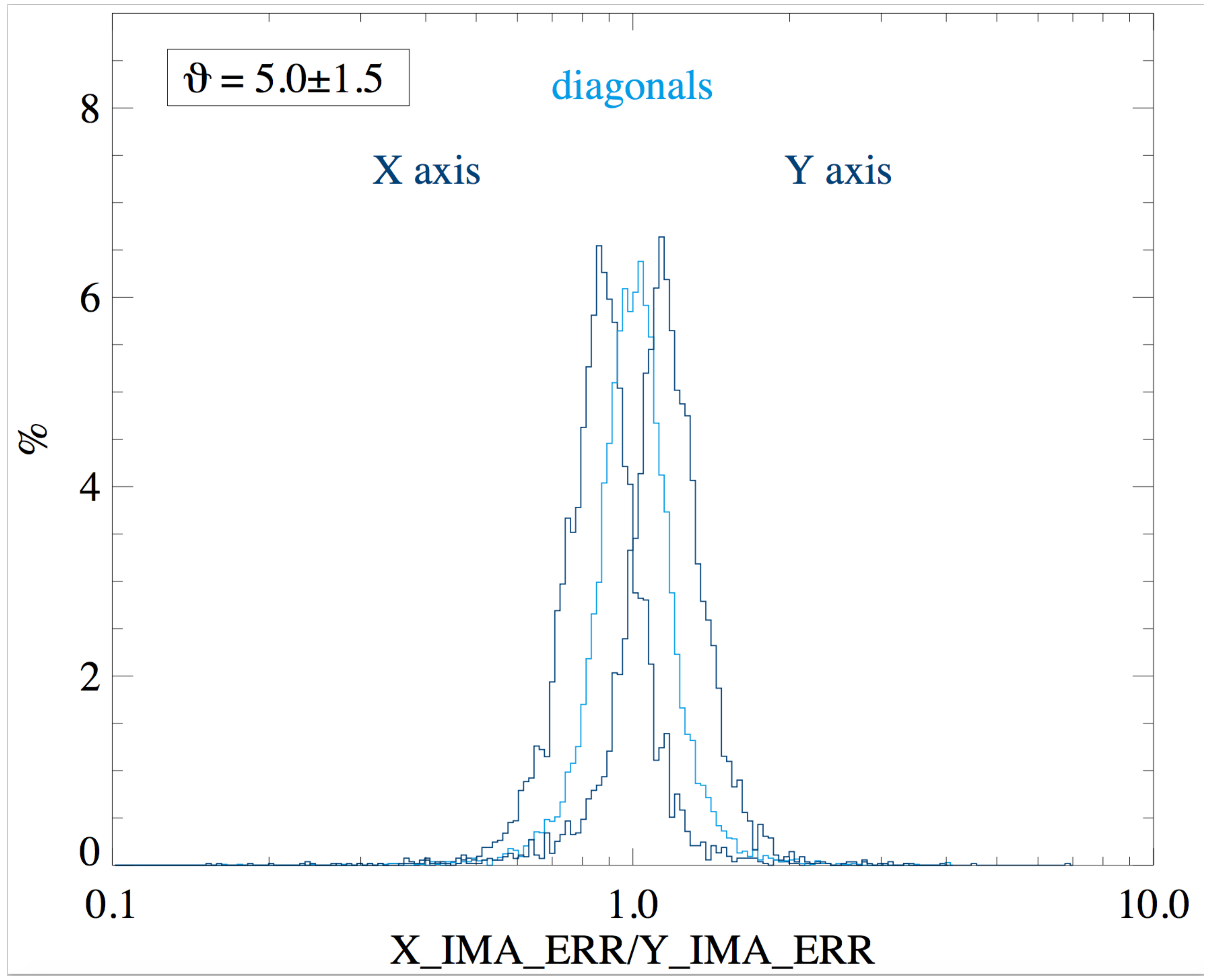


730k eRASS1 sources at DET_LIKE>7



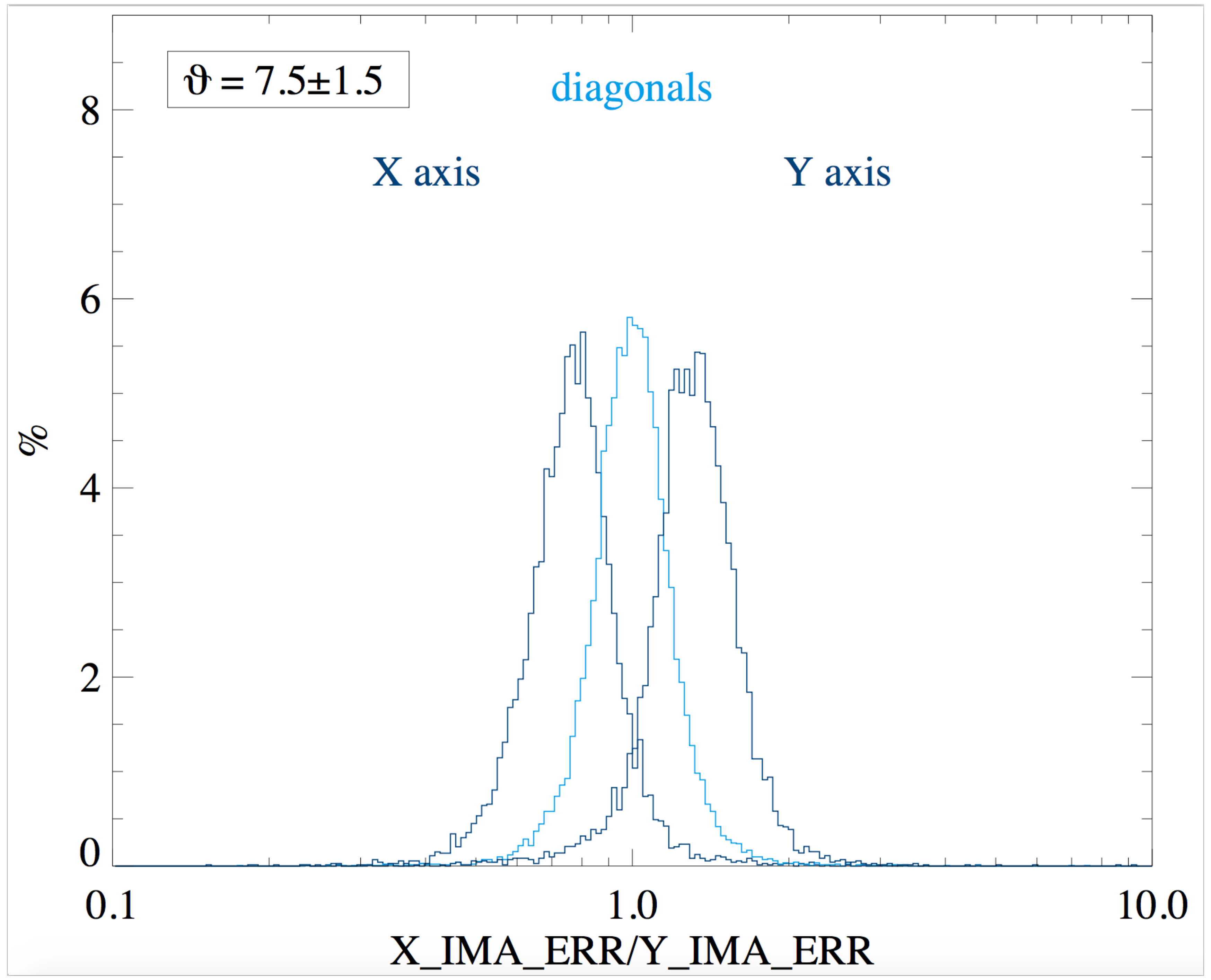


Don't look for the CTP where it is not



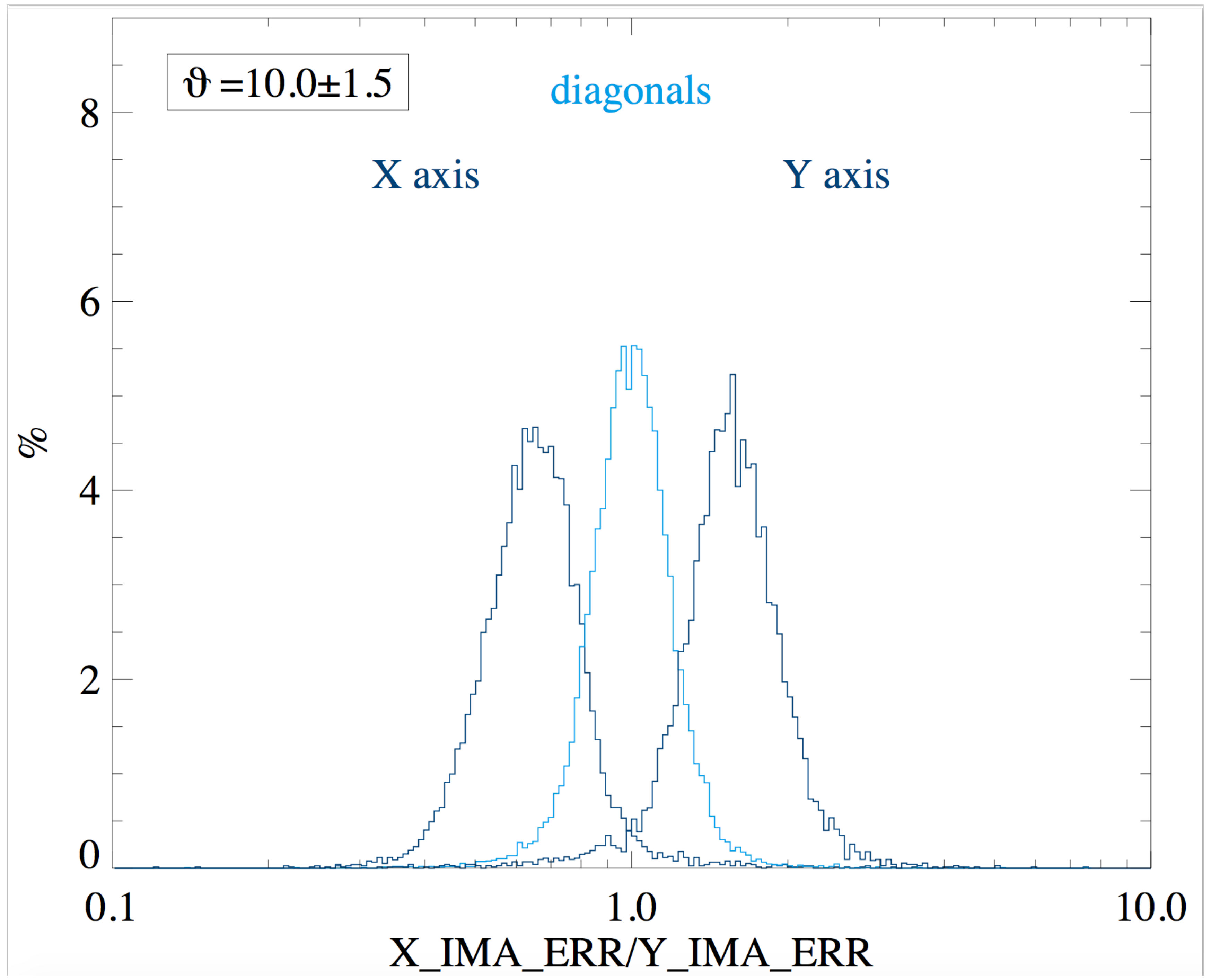
Courtesy: Iris Traulsen

Don't look for the CTP where it is not



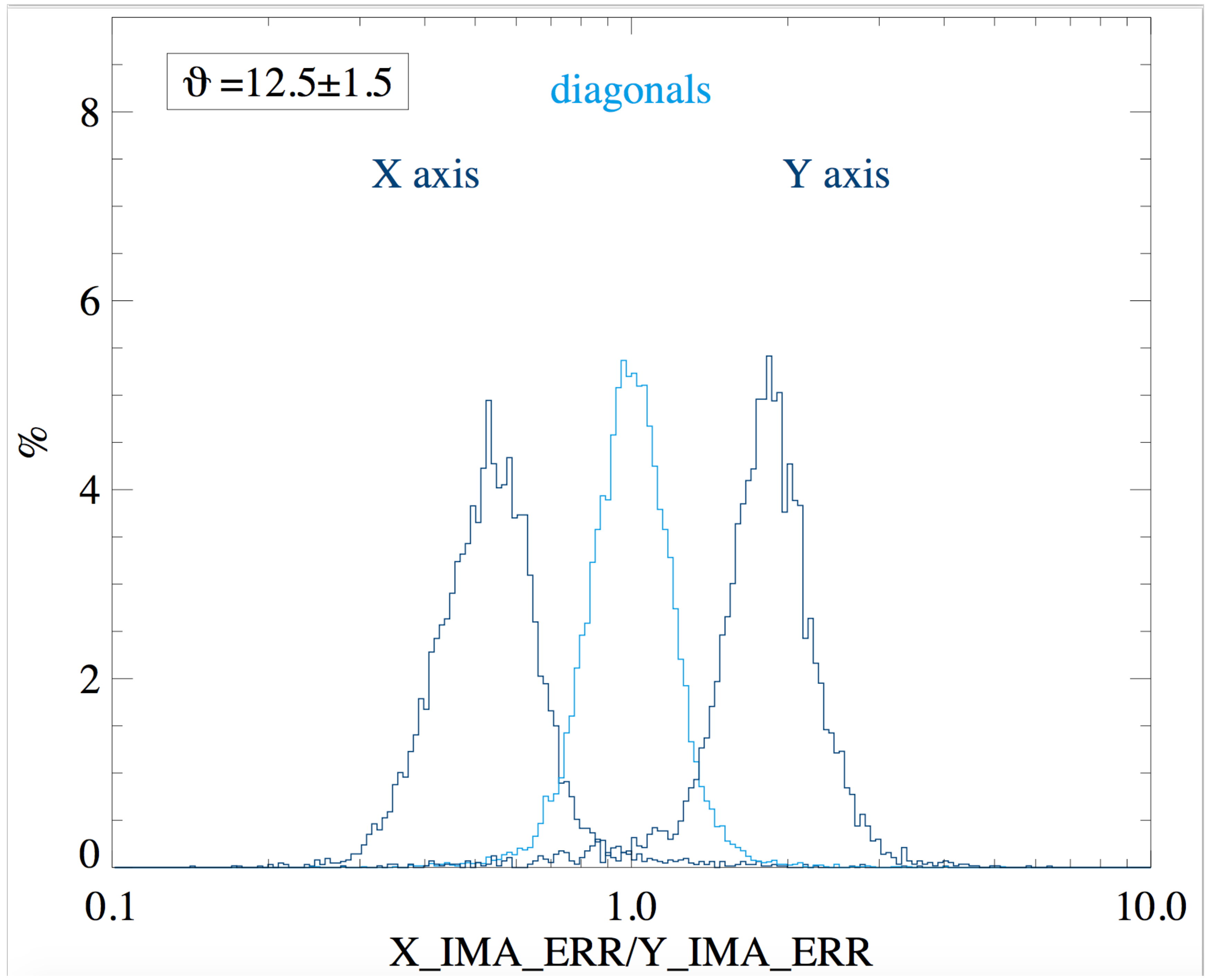
Courtesy: Iris Traulsen

Don't look for the CTP where it is not



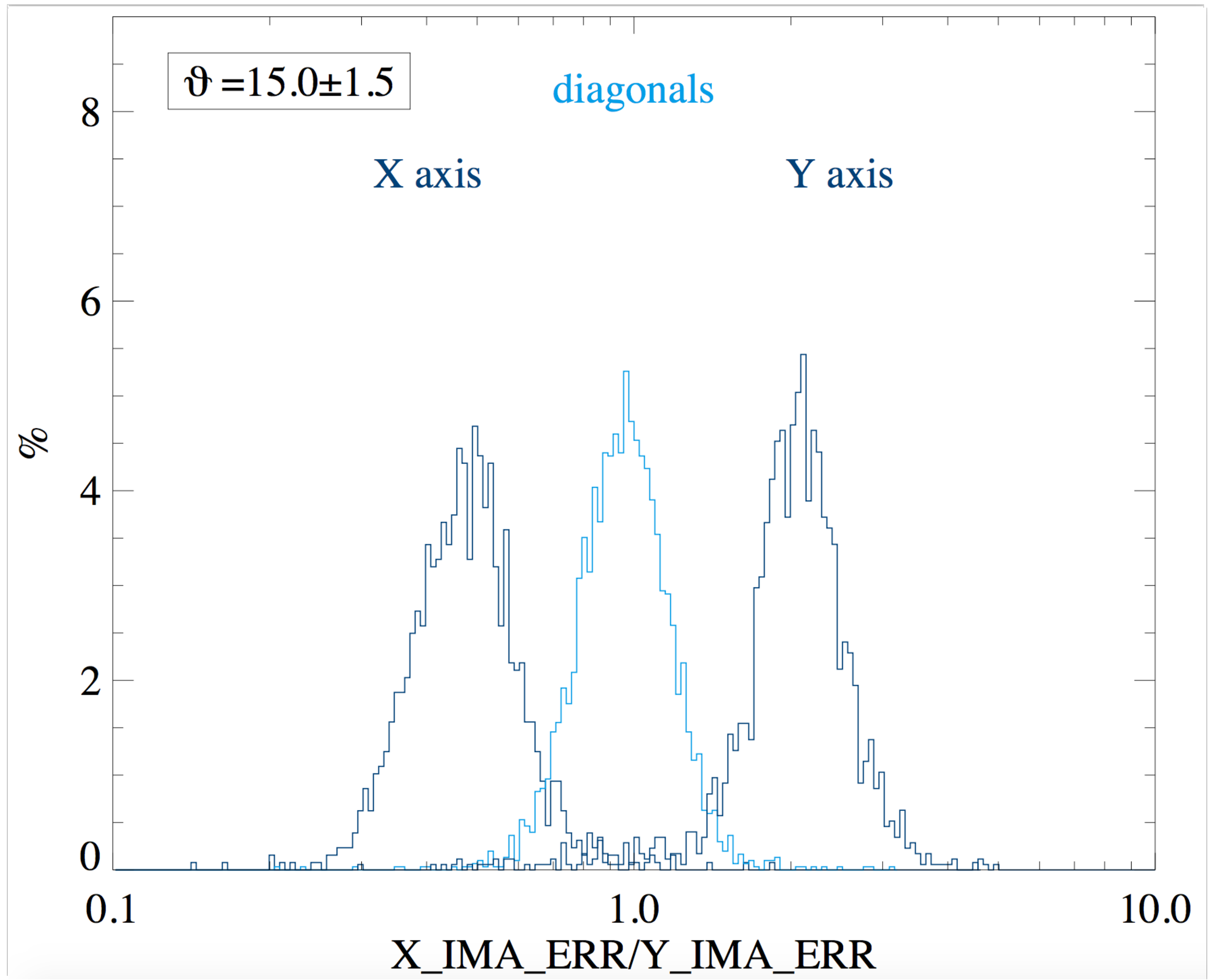
Courtesy: Iris Traulsen

Don't look for the CTP where it is not



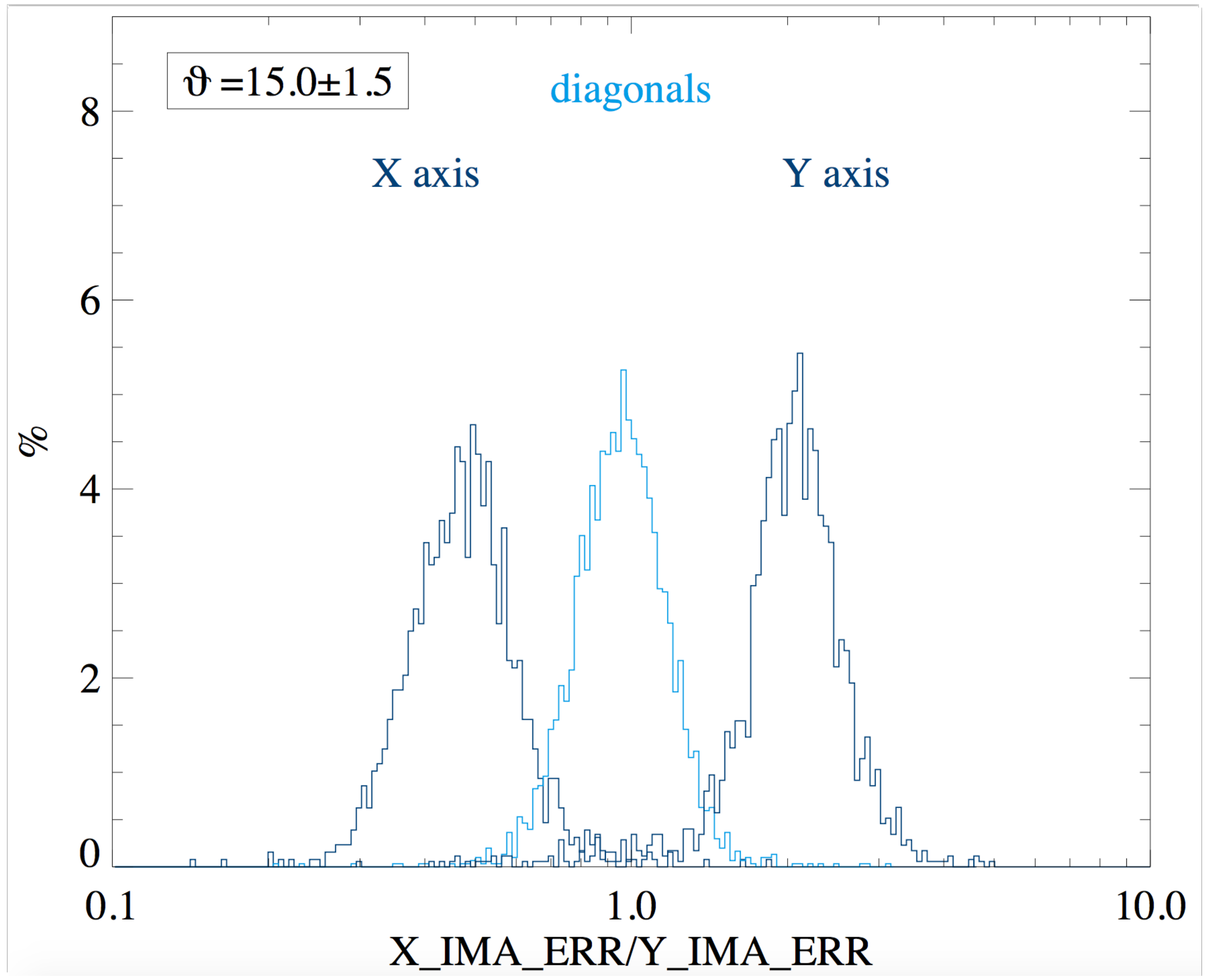
Courtesy: Iris Traulsen

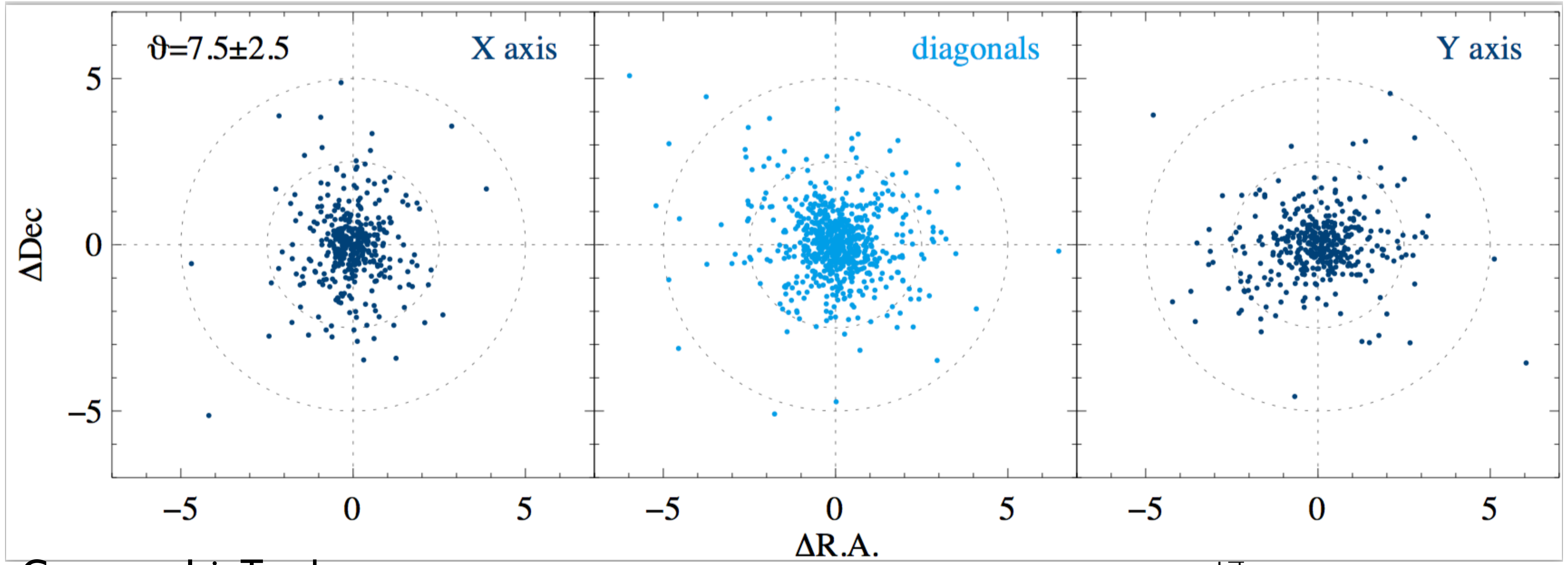
Don't look for the CTP where it is not



Courtesy: Iris Traulsen

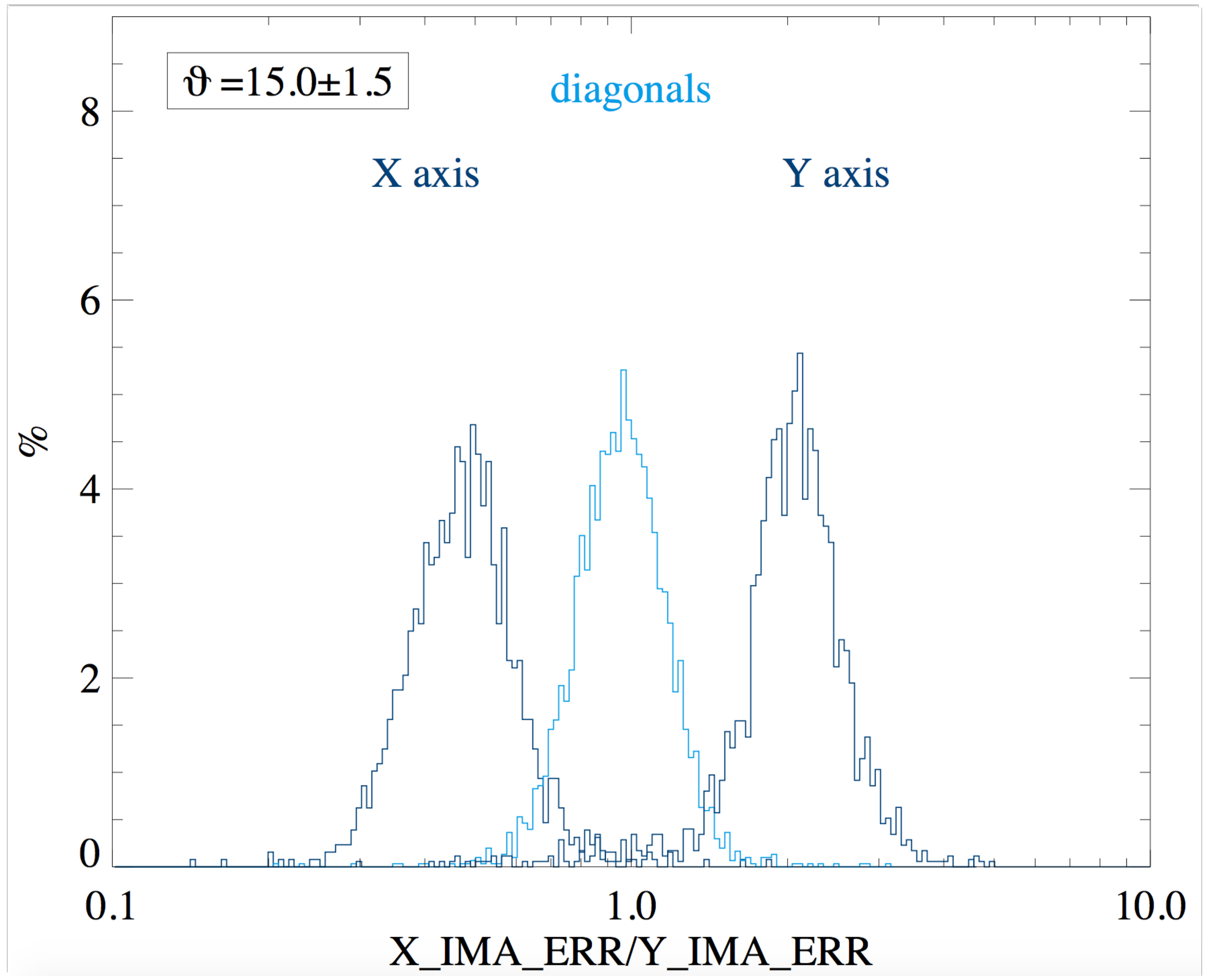
Don’t look for the CTP where it is not

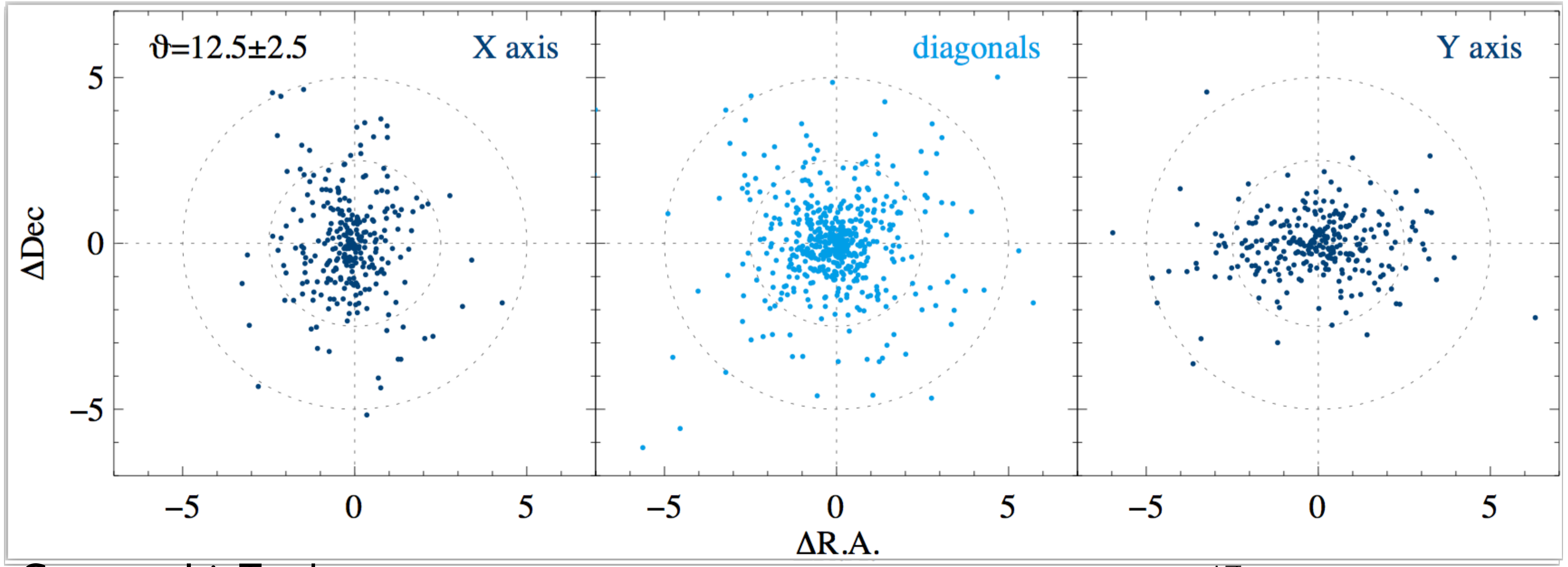




Courtesy: Iris Traulsen

Don't look for the CTP where it is not



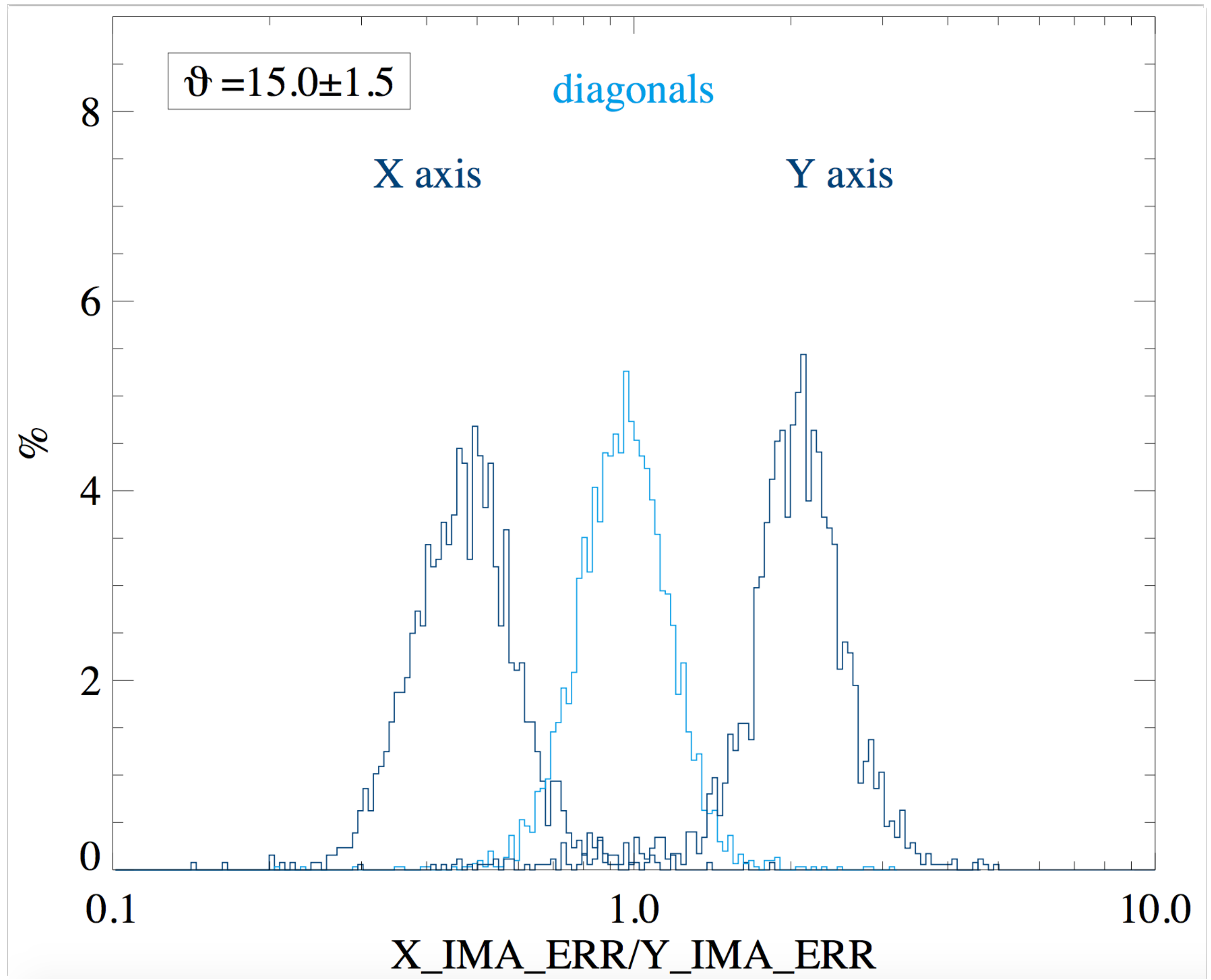


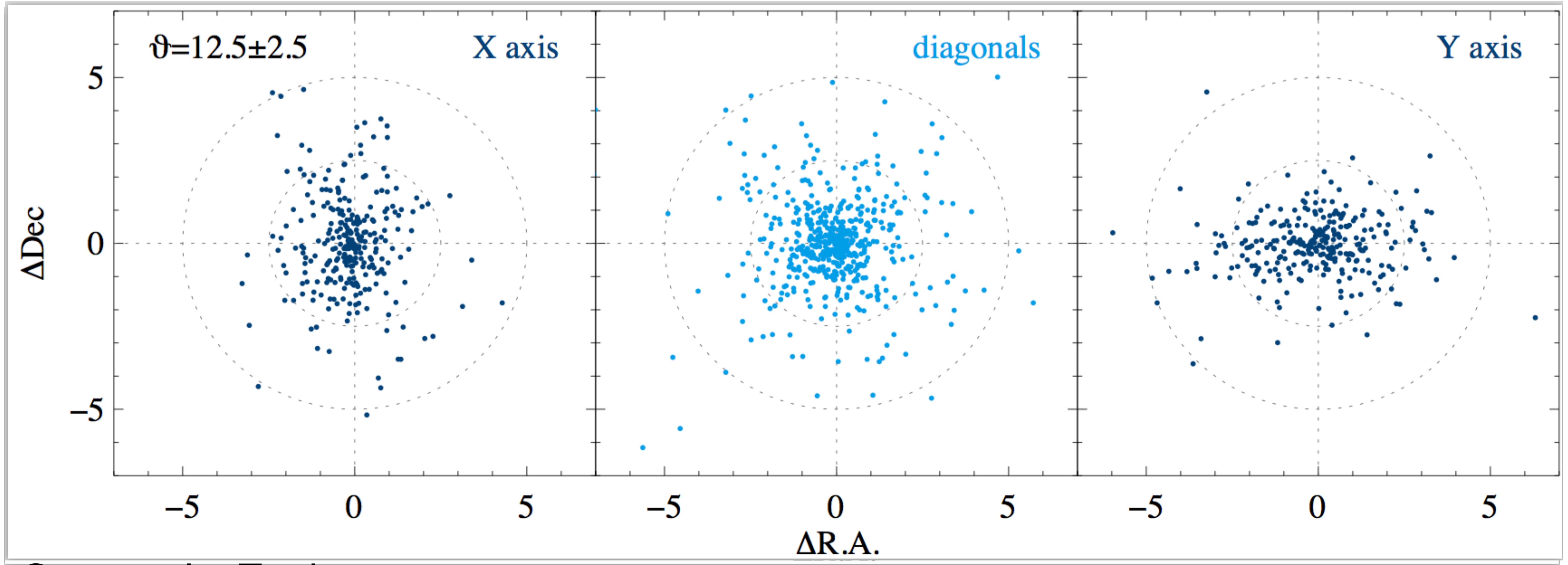
Courtesy: Iris Traulsen

Don’t look for the CTP where it is not

- The actual search of a ctp should be within an ellipse along the scanning direction

- Nway is already set for that



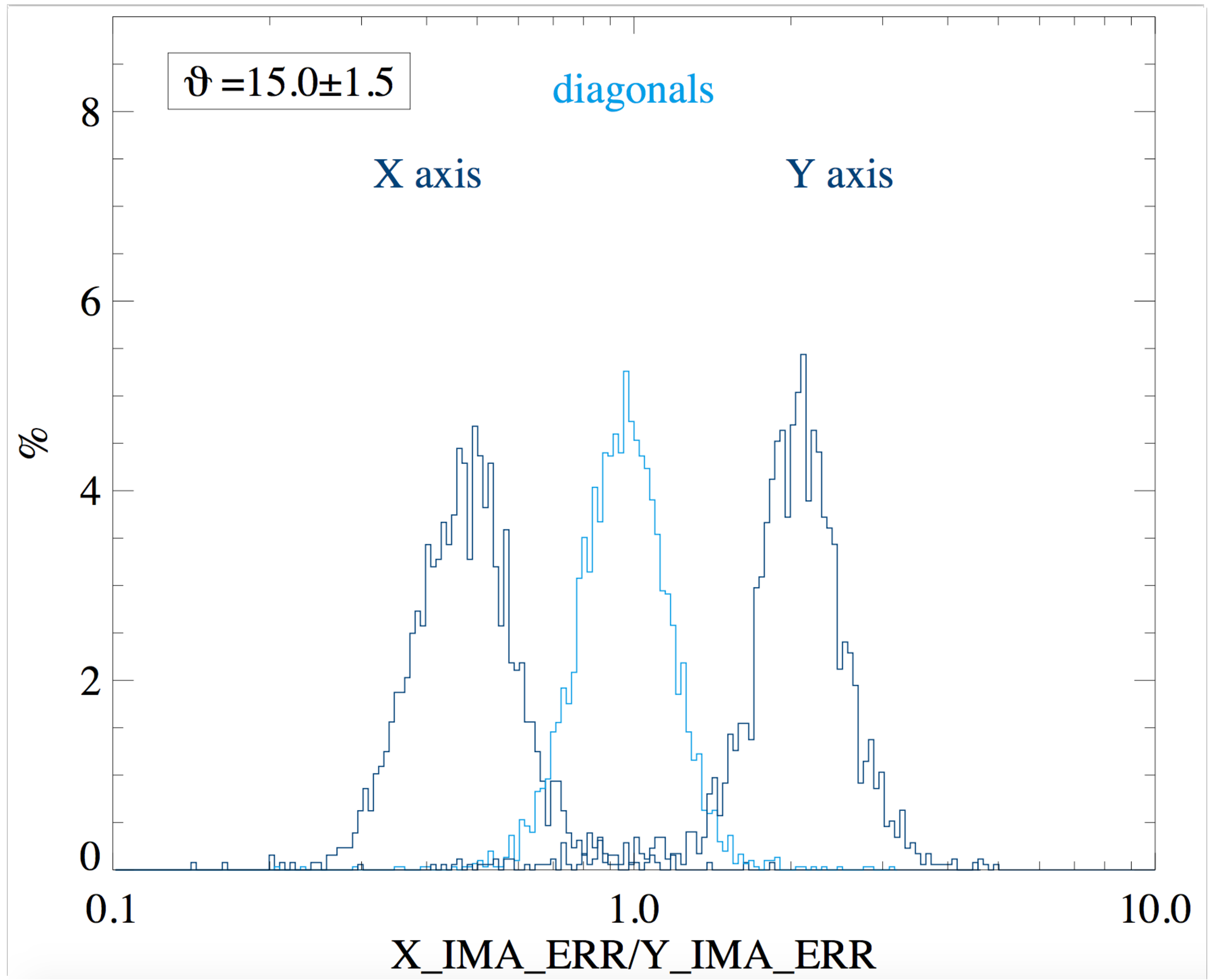


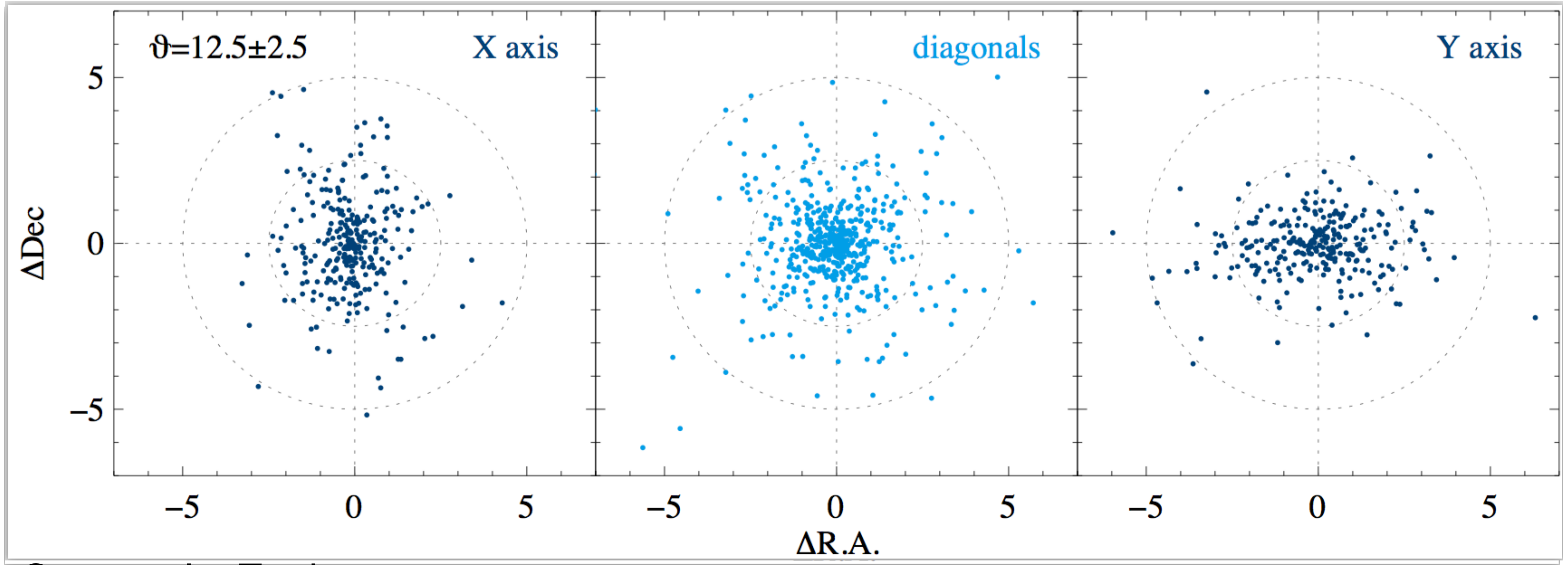
Courtesy: Iris Traulsen

Don't look for the CTP where it is not

- The actual search of a ctp
should be within an ellipse along the scanning direction

- Nway is already set for that



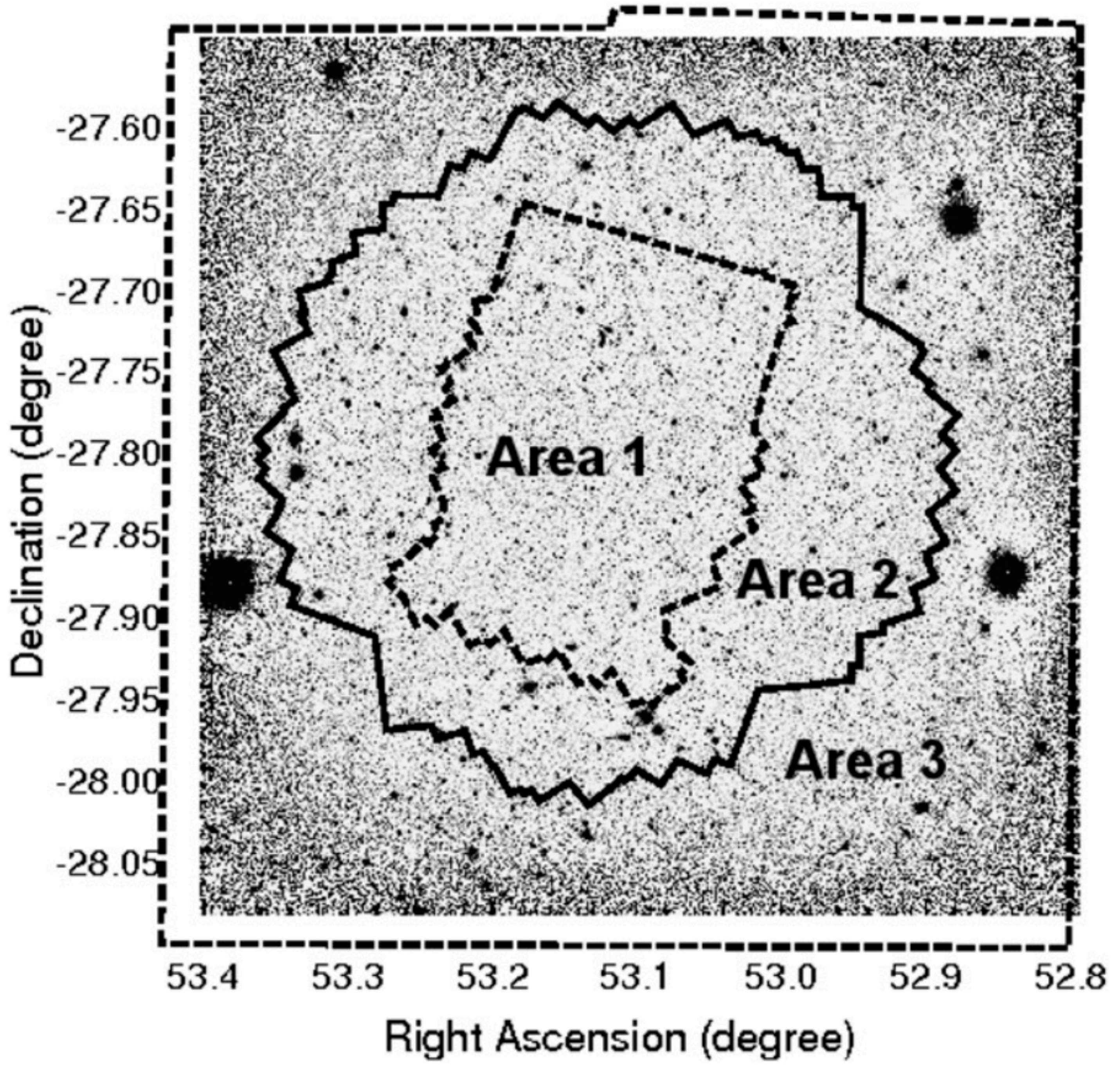


Courtesy: Iris Traulsen

Final Remarks

- Cross-matches are rarely trivial (and I did not mention blending, variability, proper motion, etc)
- Using a single magnitude/band we need to account for the possibility that the CTP is absent in that specific band (NWAY does it)
- We should not spend time searching for the ctp in the wrong place (use elliptical errors when possible)
- When using prior knowledge the model must be built using training/control/validating samples that are representative of the survey that we interested on
- Our work is catalogs-based but we are not spending enough time understanding the caveats that accompany the catalogs (purpose of the catalog, depth, flag system)
- Specific to X-ray surveys: emitters are Galactic sources, resolved sources in nearby galaxies, AGN, QSO, and unresolved clusters: it is dangerous to focus from the beginning on a specific type of source only.
- Making first the CTPs identification and then their classification, allow to pinpoint interesting populations (e.g., unresolved clusters)
- The problems mentioned are common to all cross-matches: you can tailor NWAY (Salvato et al 2018) to your needs. We are happy to help you with that.

Comparison between assumptions



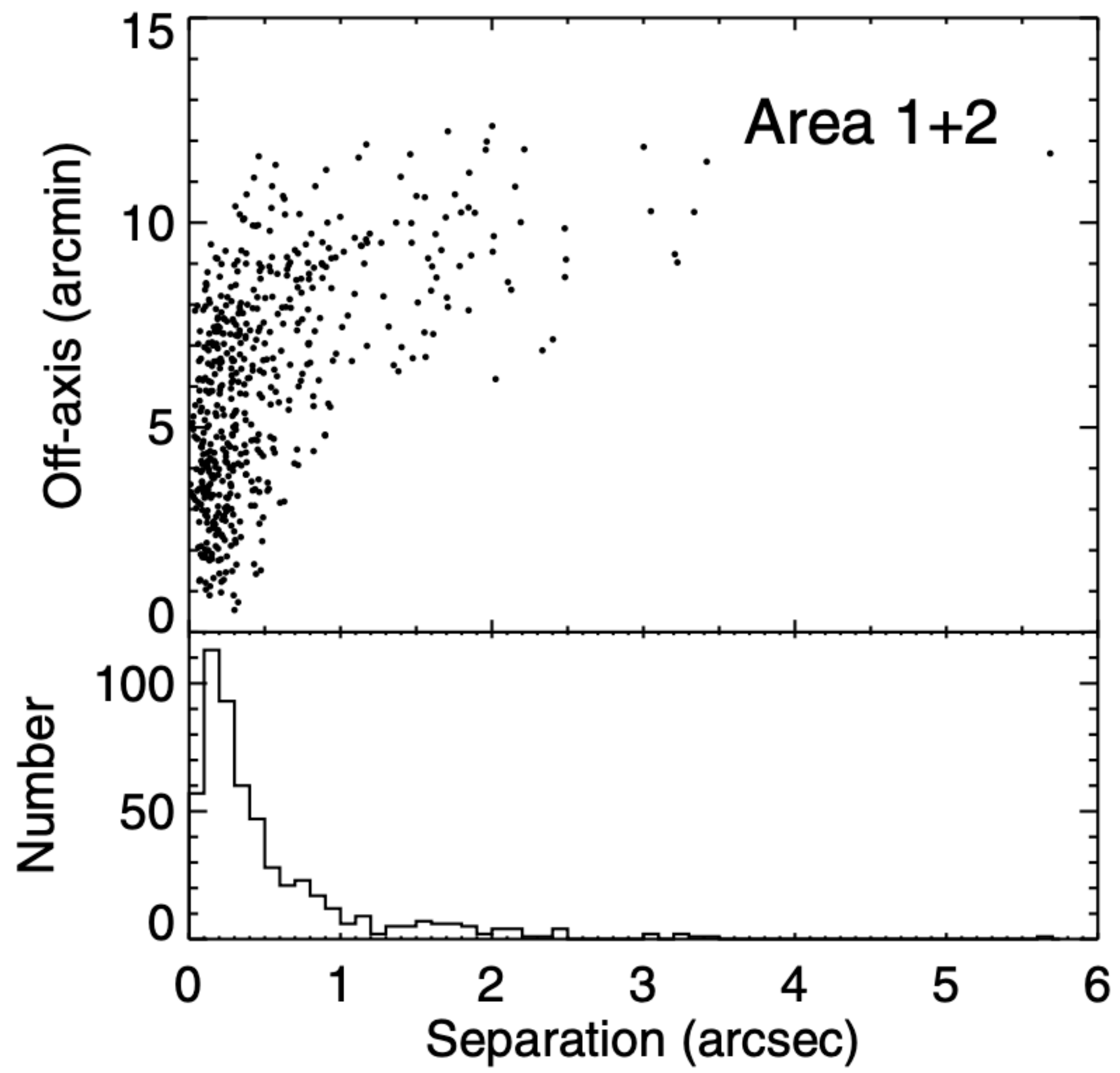


Figure 3. Coordinate differences between the X11 and R13 X-ray catalogs. The lower panel shows a histogram of offsets for the 545 sources that Areas 1 and 2 have in common in the two catalogs. The upper panel shows the off-axis angle from the *Chandra* aim point as a function of the angular offset.

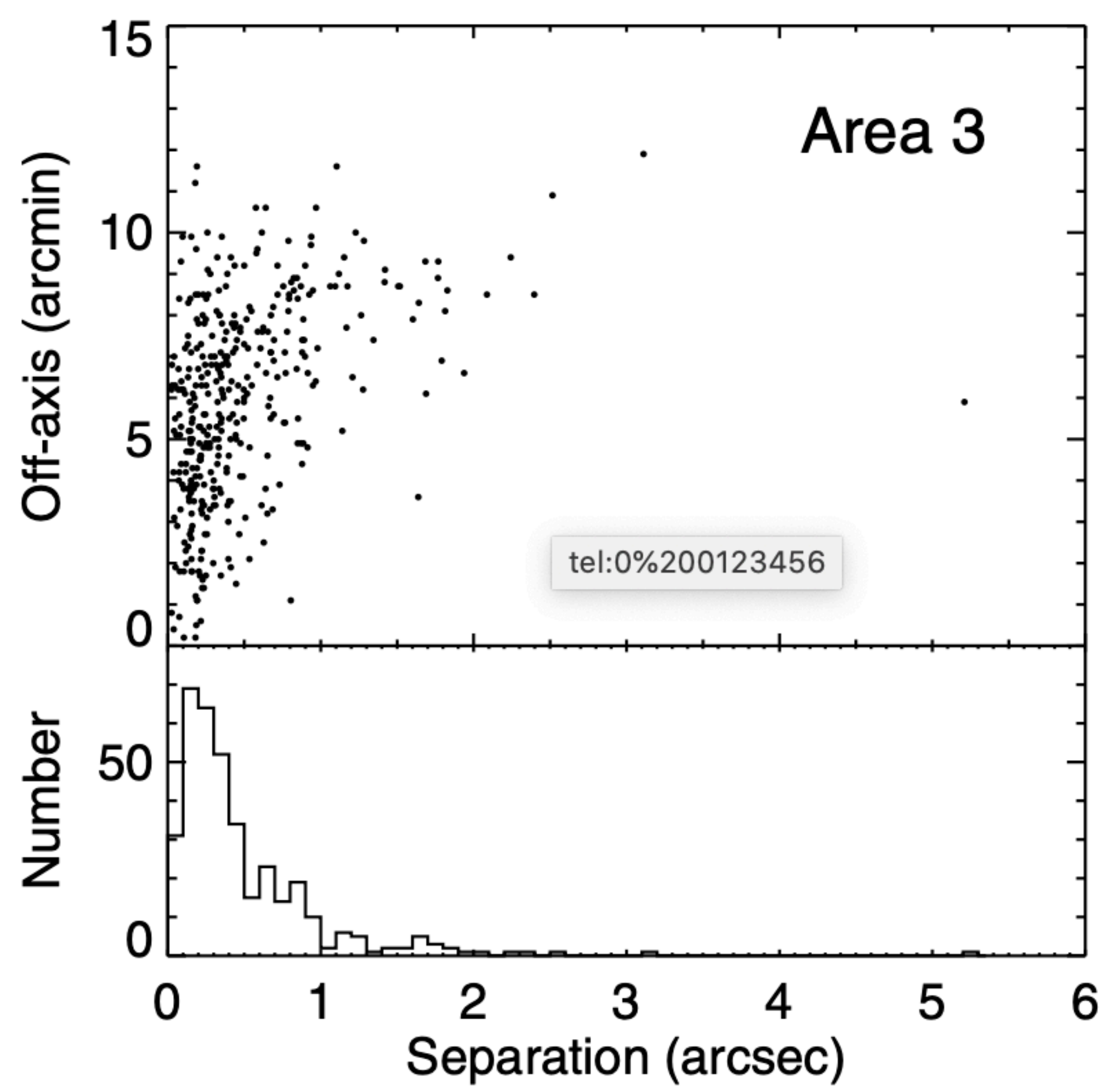


Figure 4. Coordinate differences between L05 and V06 X-ray catalogs. The lower panel shows a histogram of offsets for the 495 sources in Area 3 that are in common in the two catalogs. The upper panel shows the off-axis angle from the *Chandra* aim point as a function of the angular offset.