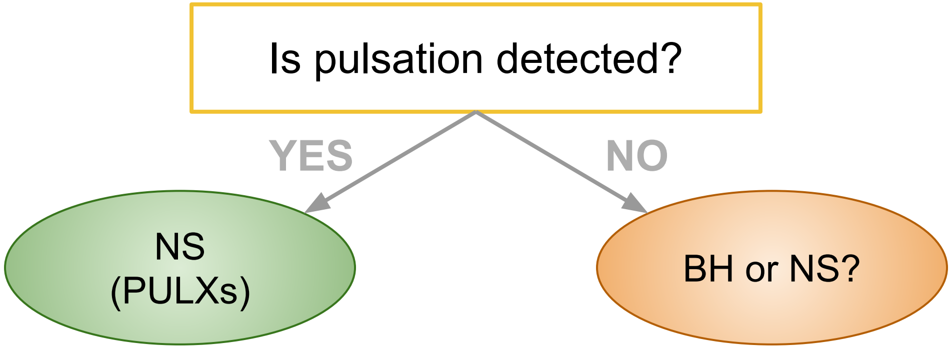
Investigating the ULX population with machine learning techniques

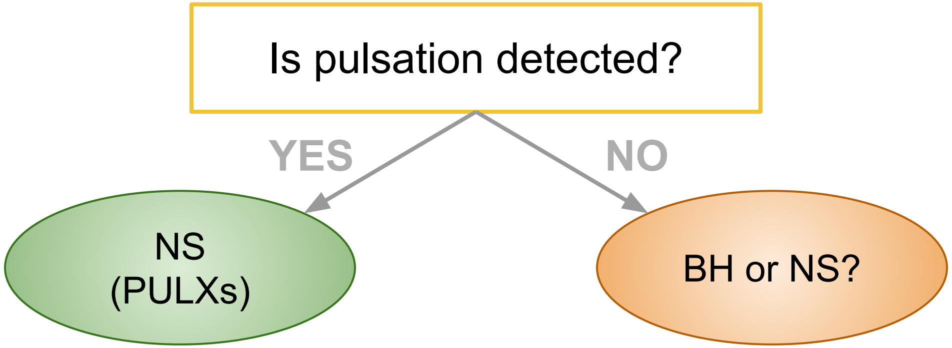
R. Amato [1], N. O. Pinciroli Vago [1,2], M. Imbrogno [1,3], G. L. Israel [1]

[1]INAF Osservatorio Astronomico di Roma
[2]Politecnico di Milano
[3]Università degli Studi di Roma "Tor Vergata"

XMM-NEWTON SCIENCE WORKSHOP
5-7 June 2024, ESAC, Spain

On the nature of the compact object



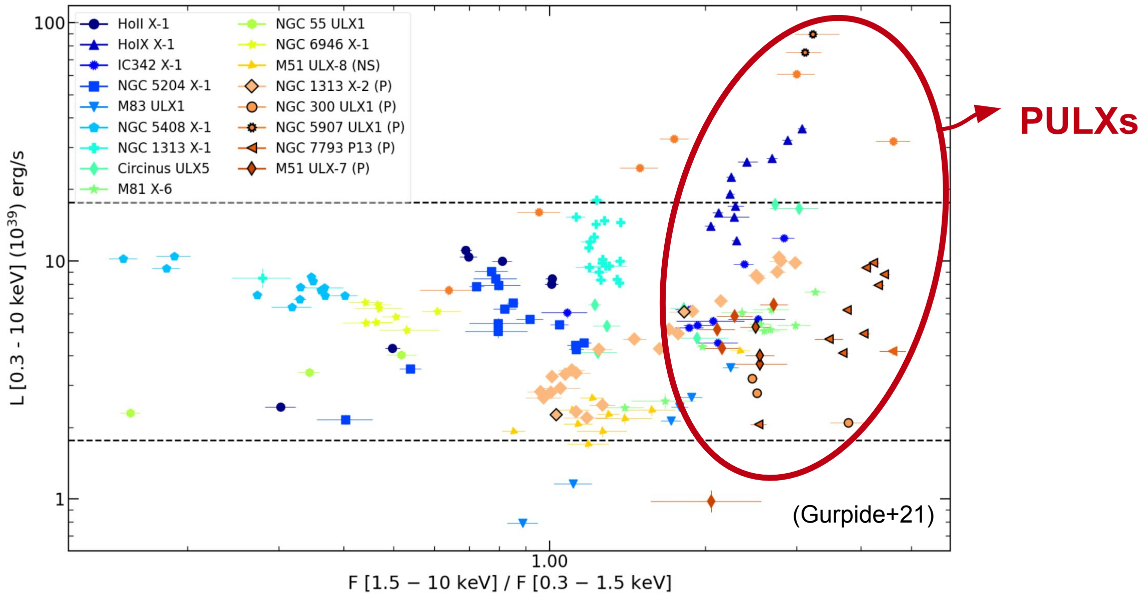


ISSUES:

- ❑ Pulsation searches algorithms need high statistics (~10000 cts)
  - ❑ High exposure times required
- ❑ Pulsations are transitory (even within the same observation)
- ❑ Strong Doppler effect and extreme Pdot
- ❑ PULXs can go below 10^{39} erg s^{-1} (less statistics)

On the nature of the compact object

- ❑ PULXs have typically **harder spectra**
- ❑ PULXs reach **higher lum.** than other ULXs (e.g. NGC 5907 X-1 peaked at $1000L_{Edd}$)
- ❑ Pulsations have been found in **~25%** of data with high statistics (**BH/NS population ratio?**)



Can we find alternative ways to discern BH-/NS-ULXs?

Can we find alternative ways to discern BH-/NS-ULXs?

↓

MACHINE LEARNING TECHNIQUES

Can we find alternative ways to discern BH-/NS-ULXs?

↓

MACHINE LEARNING TECHNIQUES

↓

CLUSTERING

CLUSTERING: a branch of unsupervised machine learning that is concerned with organised the data in groups whose members share some measure of similarities. In general, the goal is to highlight similarities and differences in the data (ML4PA, Acquaviva V., 2023)

GAUSSIAN MIXTURE MODELS (GMM): a probabilistic model that assumes all the data points are generated from a mixture of a finite number of Gaussian distributions with unknown parameters (ref). MMs in general provide a *probabilistic interpretation* of cluster membership.

The probability density function is (Reynolds 2009):

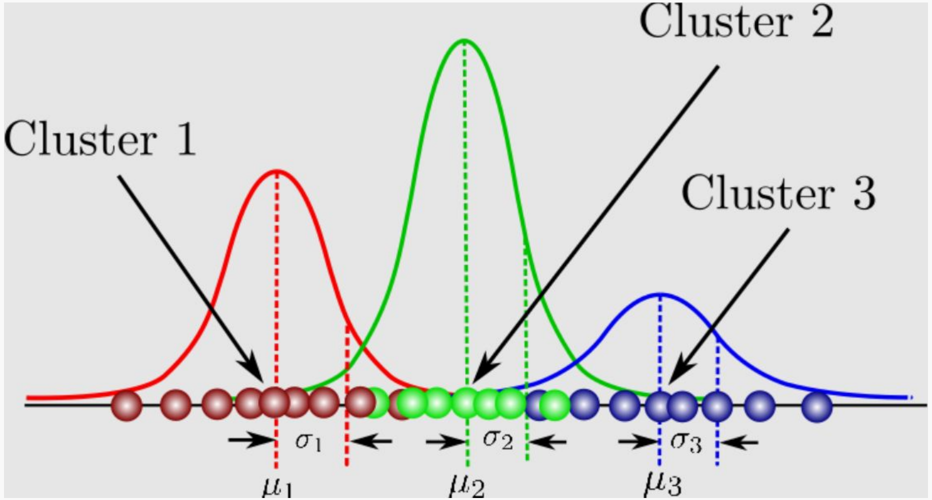
WEIGHT GAUSSIAN

$$p(x|\lambda) = \sum_{c=1}^{M} w_c \, g(x|\mu_c, \Sigma_c)$$

Mean

Covariance matrix

$$\sum_{c=1}^{M} w_c = 1$$



https://vitalflux.com/gaussian-mixture-models-what-are-they-when-to-use/

ULX catalogue

We used the catalogue of **Walton+22** (~1800 ULX candidates):

(See also ULX cat. Tranin+24)

Table 2. The final sample of ULX candidates compiled from the 4XMM-DR10, 2SXPS and CSC2 catalogues

	4XMM-DR10	2SXPS	CSC2	Combined Sample
Number of ULX Candidates	641	501	1031	1843
(with multiple ULX detections in the parent catalogue)	177	291	246	702
(seen as a ULX by multiple observatories)	241	173	209	293
(HLX candidates)	22	36	17	71
Host Galaxies	403	269	548	951
(average distance, Mpc)	62.3 ± 3.5	34.8 ± 2.7	83.8 ± 3.8	74.7 ± 2.7
(containing multiple ULX candidates)	130	89	190	333

- The catalogue is a collection of all entries of X-ray catalogues + Fpeak, Lpeak, host galaxy, distance, and cross-correlation references from previous catalogues

Building the ULX sample

- Only XMM-Newton (better data quality)
- Cross-correlation with the latest release 4XMM-DR13
- Parameters:
  - SRCID, OBSID, Host galaxy, distance
  - **Fluxes in different bands, hardness ratios, counts and rates (band 8), FVAR flag**
  - Labels PULX, Pulse, QPO (manually added)
- M82 X-1 and X-2 were removed due to source confusion and NGC 1313 X-2 was manually added
- Final numbers: **1769 instances** for 640 ULXs, **95 PULX instances** (for 7 known PULX), 32 of which show pulsations

Clustering ULXs

- Only two classes: PULXs and ULX (non-pulsating)
- Several attempts (Fpeak, Lpeak, none, or both)
- We define the metrics:
  - PULX Ratio:

$$\mathrm{PR} = \frac{\max_j(|c_j \cap P|)}{|P|}$$

The number of PULXs correctly placed in the PULX cluster over the total number of PULXs

  - Uncertain Ratio:

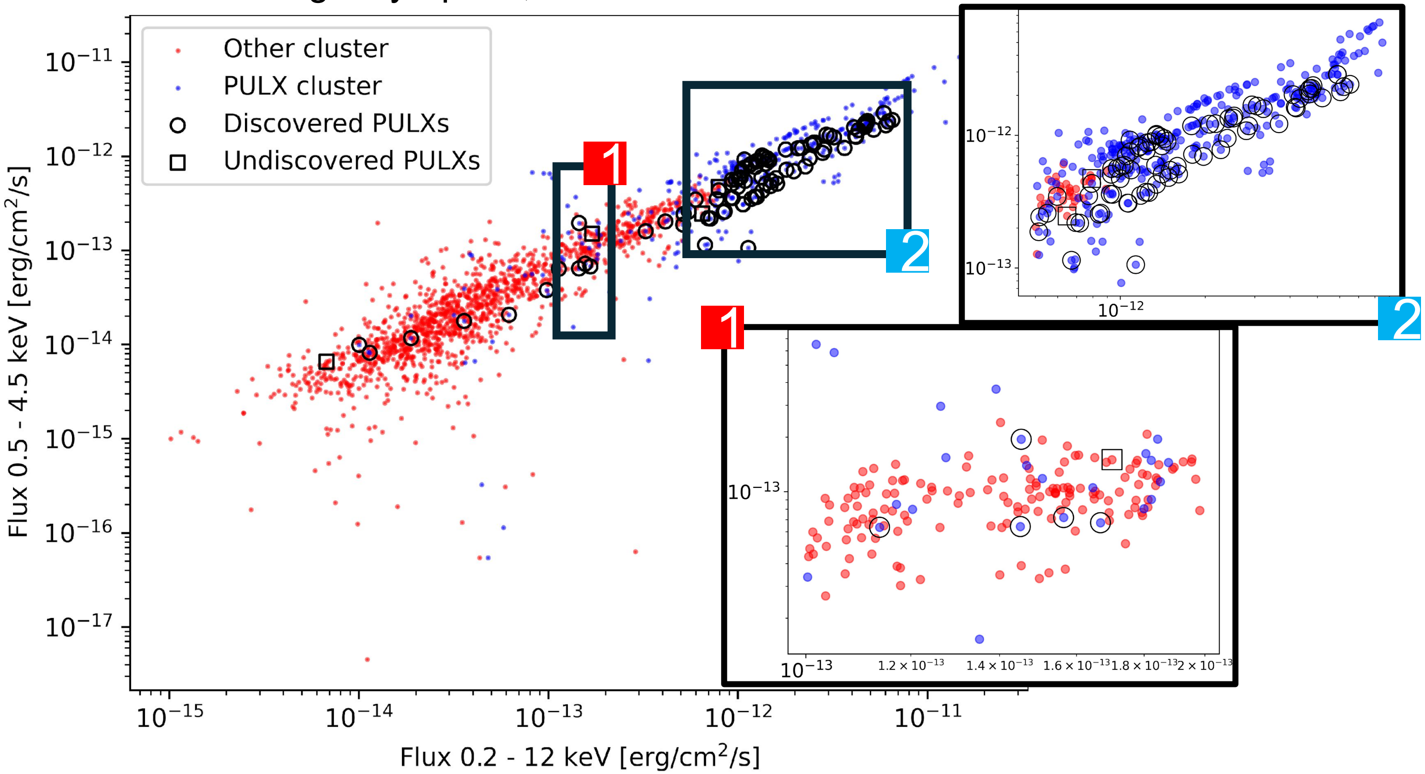
$$\mathrm{UR} = 1 - \frac{(|c_P - P|)}{|U|}$$

The number of non-PULXs correctly placed in the non-PULX cluster over the total number of non-PULXs

- We produce decision trees (DTs) for explainability

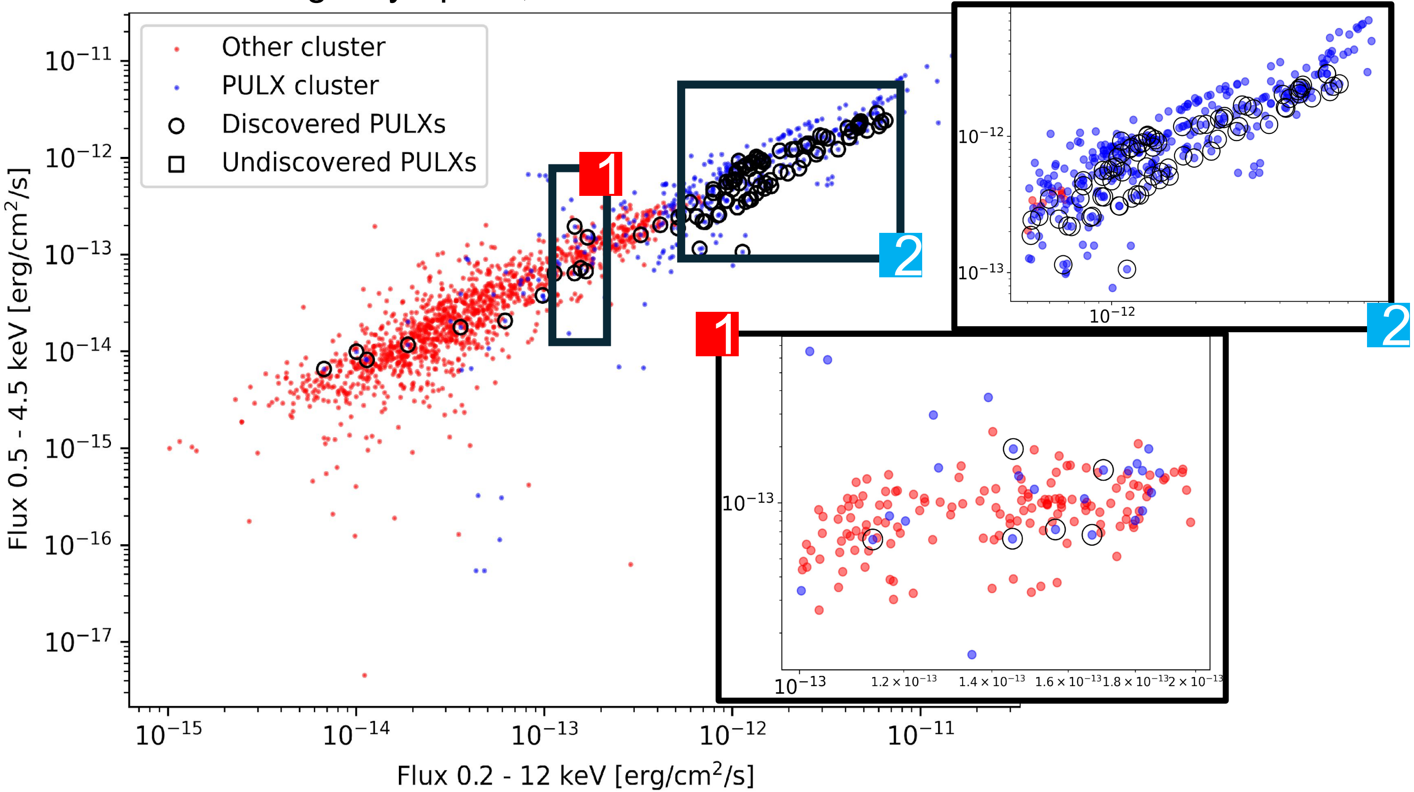
Clustering ULXs - Preliminary results

Considering only Fpeak, PR ⩾90%

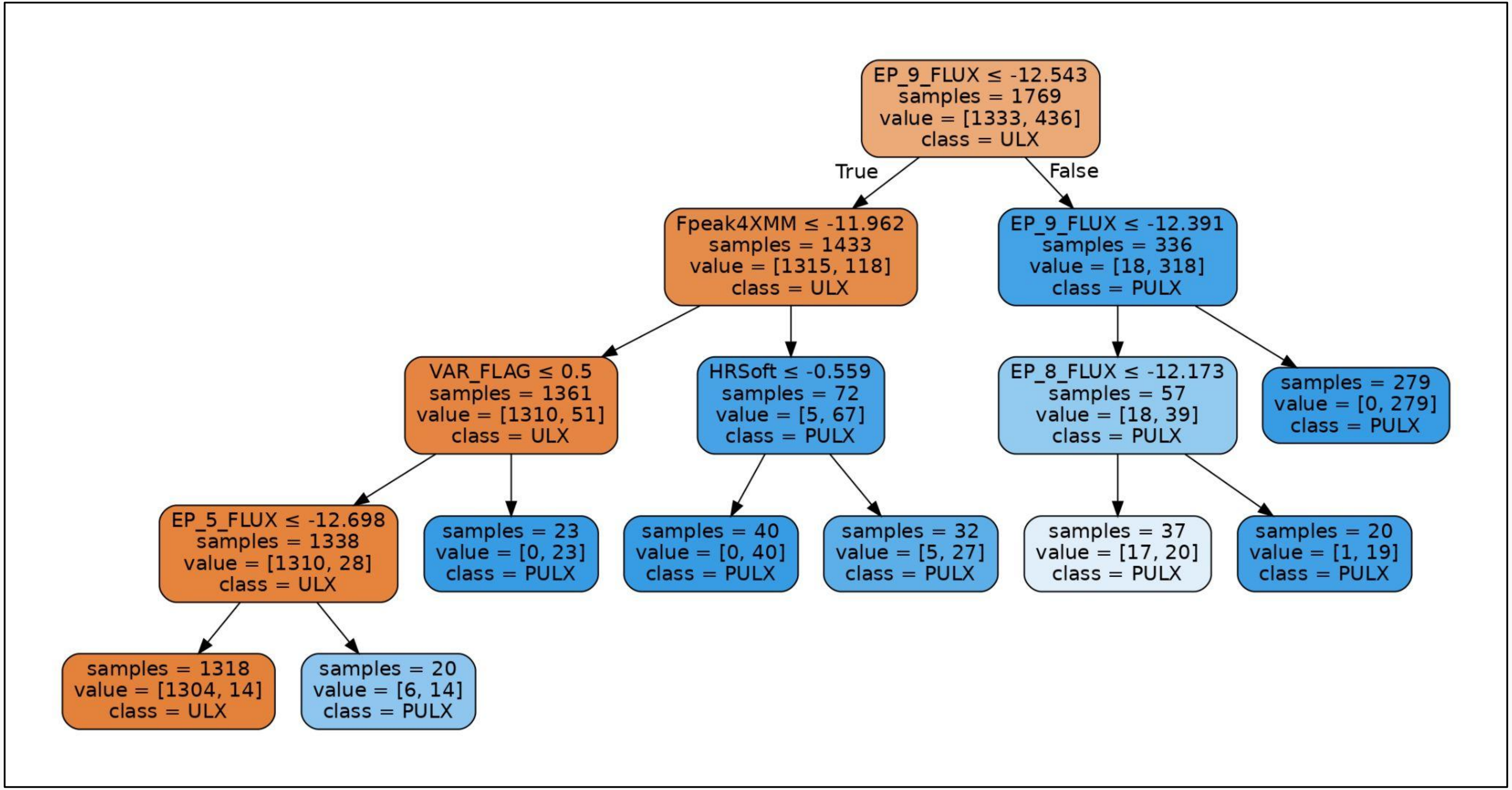


Clustering ULXs - Preliminary results

Considering only Fpeak, PR ⩾99%



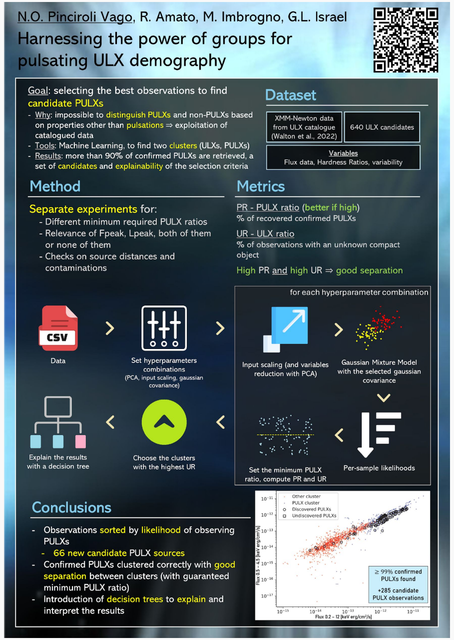
Clustering ULXs - Preliminary results



Clustering ULXs - Preliminary results

Possible BIASES and SOLUTIONS:

- ❑ The 4XMM-DR13 fluxes are obtained by fitting the spectra with the same absorbed power law with fixed nH (3e20 cm^{-2}) and photon index (1.7) →**Fit all spectra with proper models**
- ❑ More distant sources might be more contaminated (misclassification with AGN) →**Only few sources at d>100 Mpc**
- ❑ Uncertainties are not taken into account
- ❑ Increasing the number of data points will increase the performances of the algorithm →**More data are needed**



See poster by N.O. Pinciroli Vago

BACKUP

SLIDES

Pulsating ULXs (PULXs)

Name	L_X [10^{39} erg s^{-1}]	P_S [s]	$\dot{\nu}$ [10^{-10} s^{-2}]	P_{orb} [d]	M_2 [$M_\odot$]	References
M82 X-2	20	1.37	~1	2.51 (?)	≳5.2	Bachetti et al. (2014)
NGC 7793 P13	5	0.42	2	63.9	18–23 (B9I)	Fürst et al. (2016,2018), Israel et al. (2017), Motch et al. (2014)
NGC 5907 X-1	~100	1.13	38	5.3		Fürst et al. (2017), Israel et al. (2017)
NGC 300 X-1	4.7	~31.5	5.6		8–10 (?) (RSB)	Carpano et al. (2018), Heida et al. (2019)
M51 ULX-7	7	2.8	2.8 (0.31)	~2	>8	Rodrigues Castillo et al. (2020), Vasilopoulos et al. (2019)
NGC 1313 X-2	20	1.5	1.2	?		Sathyaprakash et al. (2019)
SMC X-3*	2.5	~7.7	0.69	45.04	>3.7 (Be?)	Townsend et al. (2017), Tsygankov et al. (2017)
NGC 2403 ULX	1.2	~18	3.4	60–100 (?)	(Be?)	Trudolyubov et al. (2017)
Swift J0243.6+6124*	≳1.5 (?)	9.86	2.2	28.3	(Be?)	Doroshenko et al. (2018)
RXJ0209.6-7427*	1–2	9.3	1.165	>50 (?)	Be	Vasilopoulos et al. (2020)
M51 X-8**	2	–	–	8–400 (?)		Brightman et al. (2018)

* Galactic or in the Magellanic clouds
** Cyclotron line detected (no pulsations)

AND SOME MORE…

Building the ULX sample

Parameters:

- ➢ Name, Dist, 4XMMName, 4XMMSRCID, RAXdeg, DEXdeg, ePosX, OBS_ID
- ➢ Fpeak4XMM, e_Fpeak4XMM, Lpeak4XMM, e_Lpeak4XMM
- ➢ EP_#_FLUX and EP_#_FLUX_ERR →
- ➢ EP_8_RATE, EP_8_RATE_ERR, EP_9_RATE, EP_9_RATE_ERR
- ➢ EP_8_CTS, EP_8_CTS_ERR,
- ➢ EP_HR#, EP_HR#_ERR →
- ➢ FVAR and FVARERR for PN, M1, M2, VAR_FLAG
- ➢ HRSoft, HRSoft_err, HRHard, HRHard_err
- ➢ Labels: PULX, Pulse, QPO

[1] 0.2–0.5 keV
[2] 0.5–1.0 keV
[3] 1.0–2.0 keV
[4] 2.0–4.5 keV
[5] 4.5–12 keV
[8] 0.2–12 keV (broadband)
[9] 0.5–4.5 keV

HR1: (0.2–0.5 keV)/(0.5–1 keV)
HR2: (0.5–1 keV)/(1–2 keV)
HR3: (1–2 keV)/(2–4.5 keV)
HR4: (2–4.5 keV)/(4.5–12 keV)

HRhard: 2–12 keV/0.5-12 keV
HRsoft: 0.5-2.0 keV/0.5-12 keV

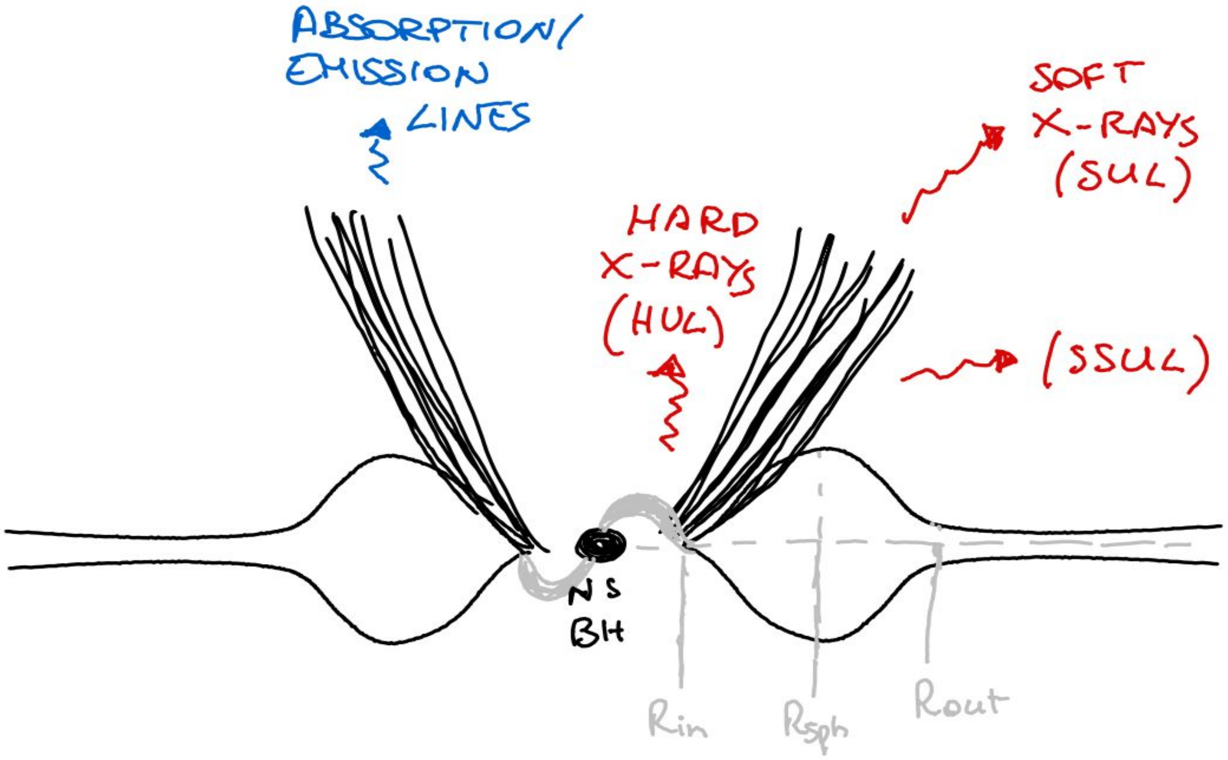
Ultraluminous X-ray sources (ULXs)

- ❑ Point-like
- ❑ Extra-galactic
- ❑ In outskirts of galaxies (off-nuclear)
- ❑ **$L_X \geq 10^{39}$ erg s^{-1}**

$$L_{\rm Edd} \approx 1.3 \times 10^{38} \left(\frac{M}{M_\odot}\right) \text{erg s}^{-1}$$

- ➢ For a NS of 1.4$M_\odot$: $L_{\rm Edd} \sim 2 \times 10^{38}$ erg s^{-1}
- ➢ For a 10$M_\odot$ BH: $L_{\rm Edd} \sim 1 \times 10^{39}$ erg s^{-1}

For $M \gtrsim 10\, M_\odot \Rightarrow L_{\rm Edd} \gtrsim 10^{39}$ erg s^{-1}


ABSORPTION/
EMISSION
LINES
SOFT
X-RAYS
(SUL)
HARD
X-RAYS
(HUL)
(SSUL)
NS
BH
Rin
Rsph
Rout

On the nature of the compact object

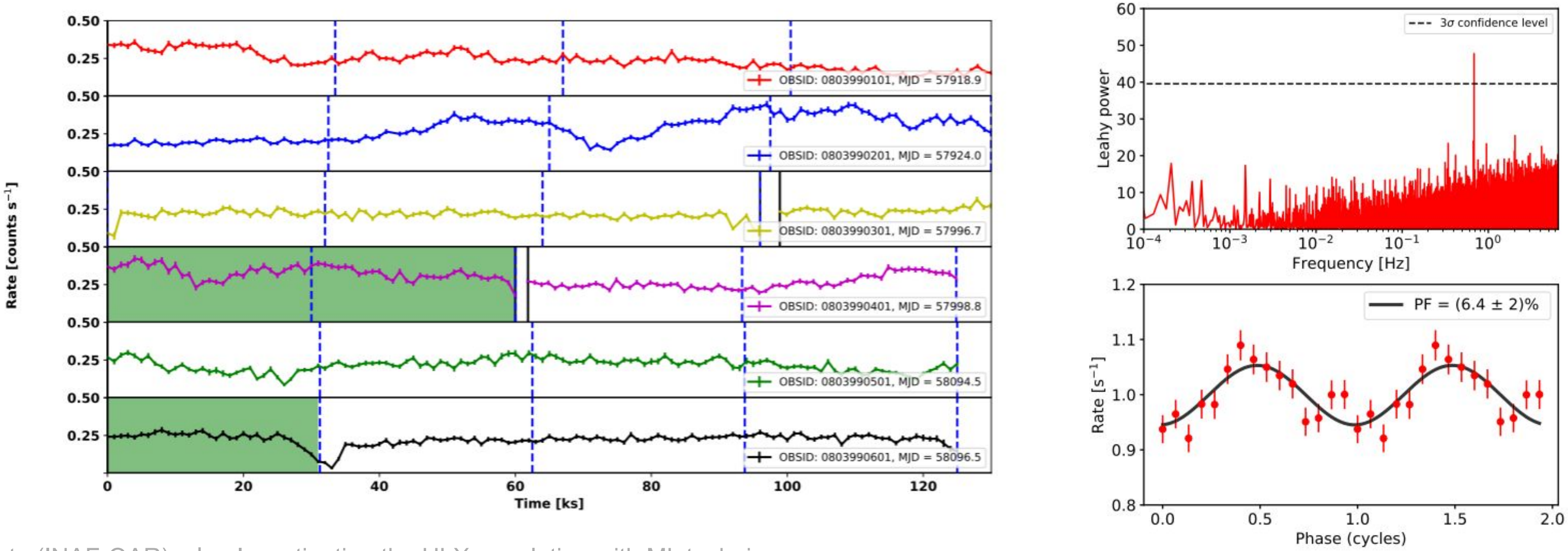
OTHER PIECES OF EVIDENCE

- ❏ PULXs have typically **harder spectra**
- ❏ Pulsations have been found in **~25%** of data with high statistics (**BH/NS population ratio?**)
- ❏ PULXs reach **higher lum.** than other ULXs (e.g. NGC 5907 X-1 peaked at 1000L_{Edd})
- ❏ One ULX (M51 ULX8, Brightman+18) shows a **cyclotron line** (**NS-ULX?**). Other candidates cyclotron lines have been found in NGC 300 ULX-1 (Walton+18), NGC 4045 X-1 (Brightman+22)
- ❏ The Galactic pulsar Swift J0243.6+6124 (e.g. Wilson-Hodge+18) and the pulsars SMC X-3 (Tsygankov+17) and RX J0209.6-7427 (Vasilopoulos+20) reached **super-Eddington luminosities during outburst** (**same physics?**)

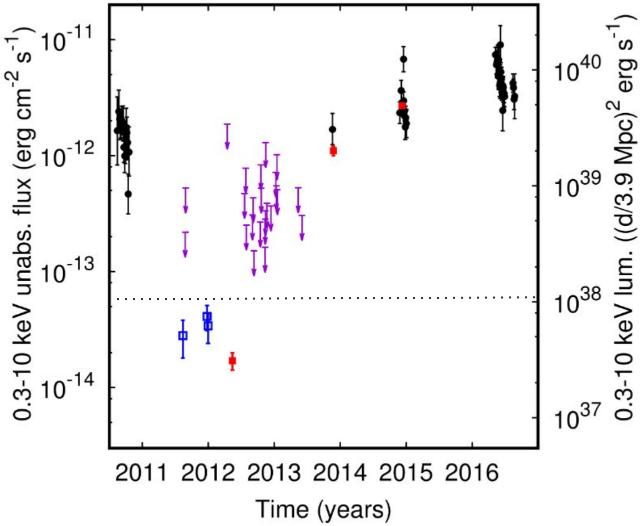
Latest PULX discovered
P = 1.5s
PF ~ 5%
Signal duty cycle ~14%

TRANSIENT PULSATIONS

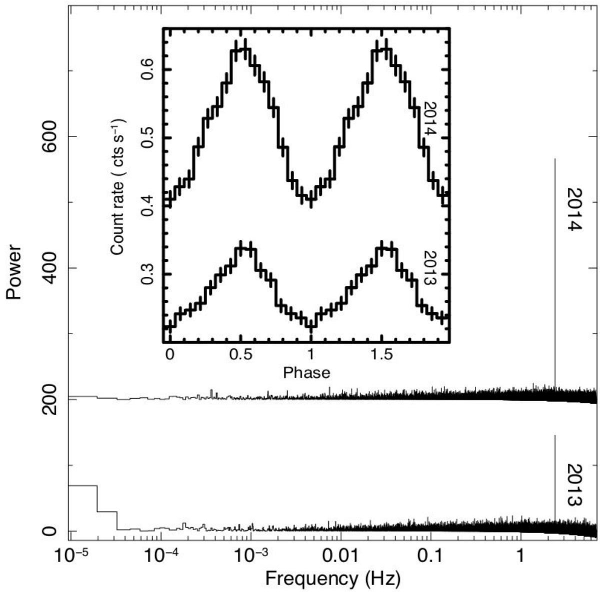
(Sathyaprakash +19)



The fastest PULX with P = 0.42s discovered in two XMM-Newton obs
PF ~ 10% – 40%
Porb=66d
Lx~10^{37} – 10^{40} erg s^{-1}

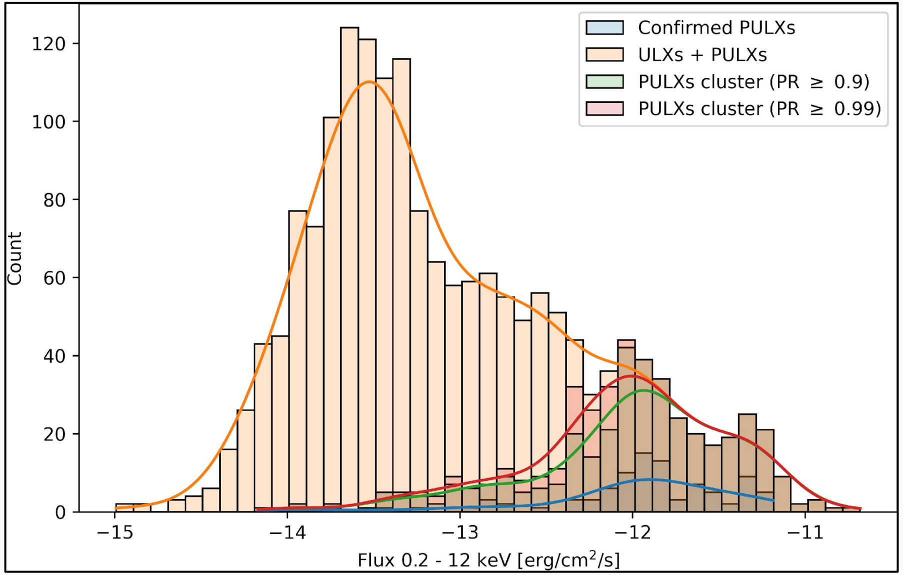


(Israel+17)



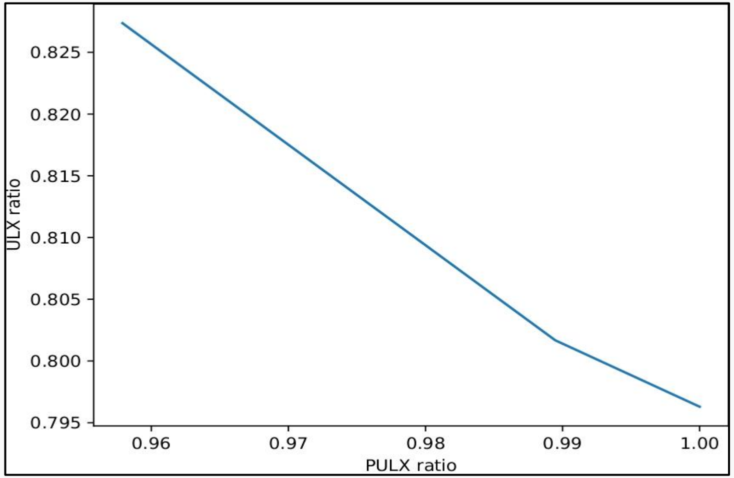
TRANSIENT PULX
TRANSIENT PULSATIONS

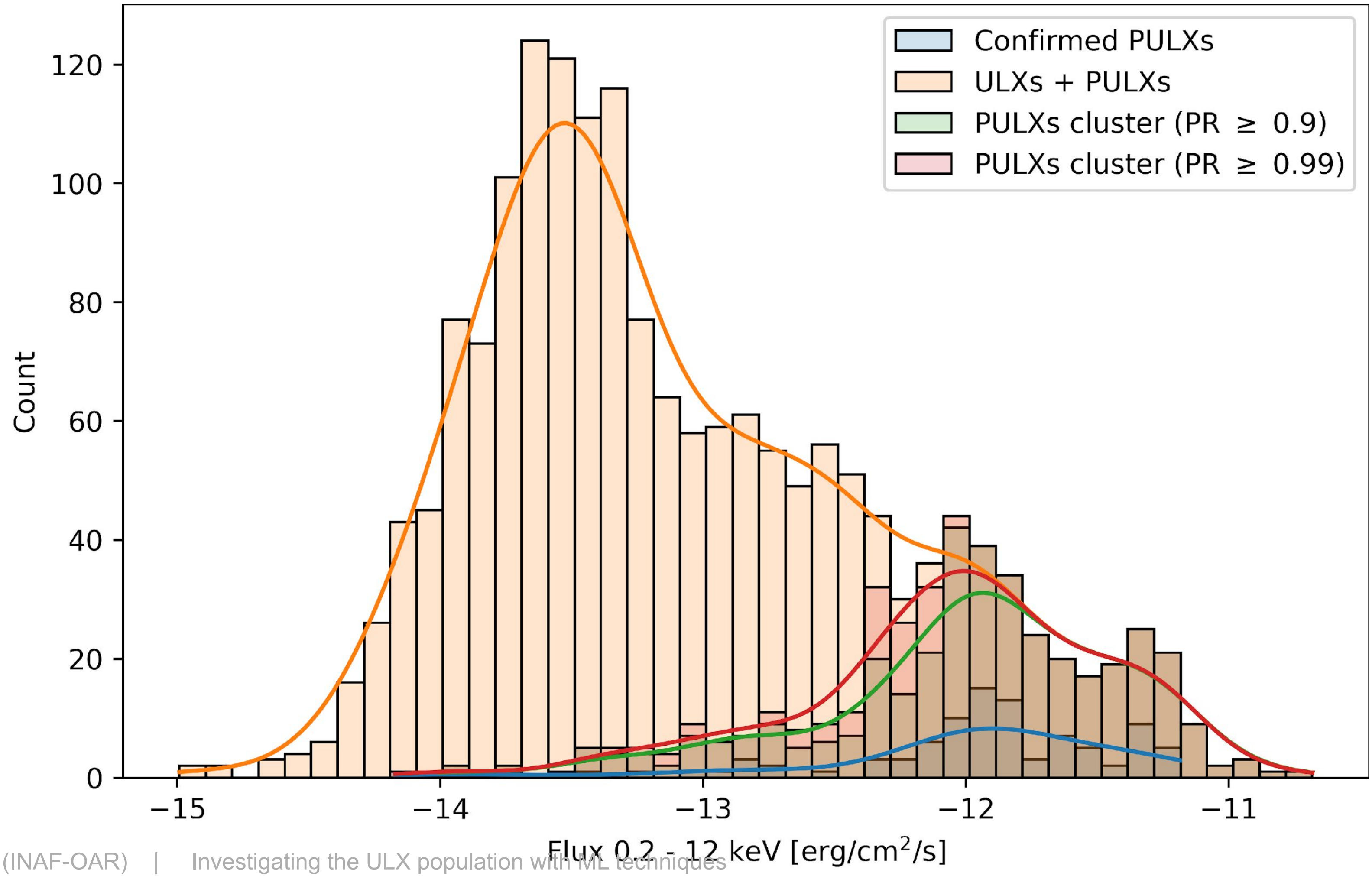
Clustering ULXs - Preliminary results

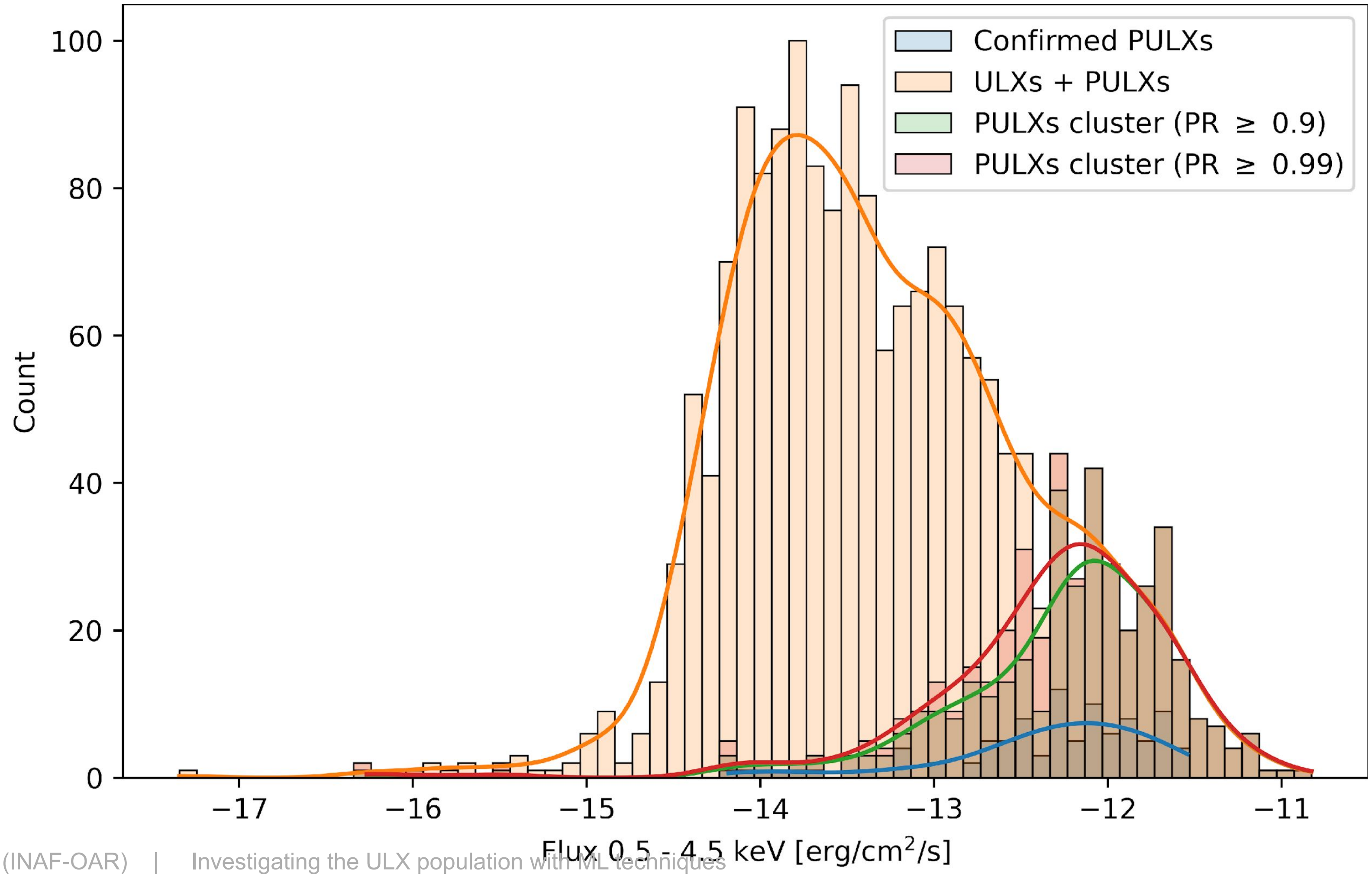


PR ≥90%:
- Correctly identified PULXs: 91/95
- non-PULX in PULX cluster: 289/1674

PR ≥99%:
- Correctly identified PULXs: 95/95
- non-PULX in PULX cluster: 341/1674




Confirmed PULXs
ULXs + PULXs
PULXs cluster (PR ≥ 0.9)
PULXs cluster (PR ≥ 0.99)
Count
0
20
40
60
80
100
120
−15
−14
−13
−12
−11
Flux 0.2 - 12 keV [erg/cm^2/s]


Confirmed PULXs
ULXs + PULXs
PULXs cluster (PR ≥ 0.9)
PULXs cluster (PR ≥ 0.99)
Count
100
80
60
40
20
0
−17
−16
−15
−14
−13
−12
−11
Flux 0.5 - 4.5 keV [erg/cm^2/s]

CLUSTERING: a branch of unsupervised machine learning that is concerned with organised the data in groups whose members share some measure of similarities. In general, the goal is to highlight similarities and differences in the data (ML4PA, Acquaviva V., 2023)