XMM2ATHENA

Classifying the XMM-Newton Optical Monitor Serendipitous UV Source Survey sources.

Mat Page

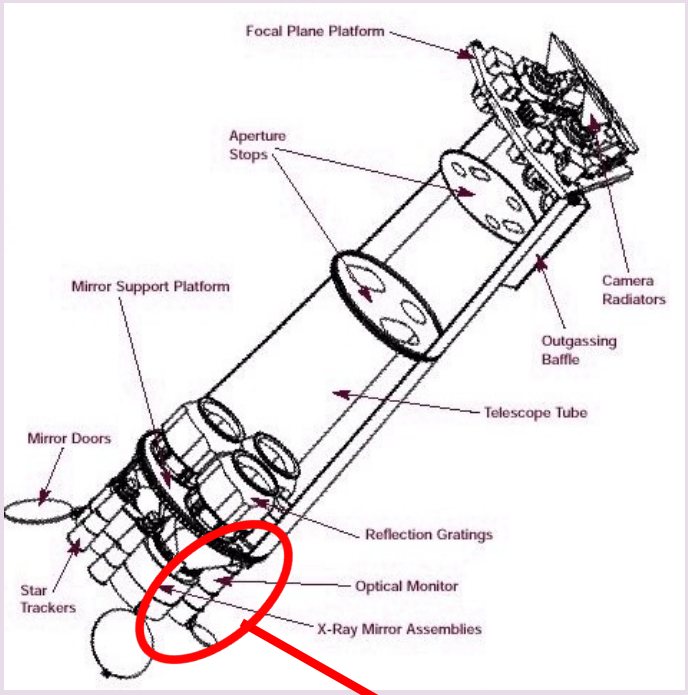
On behalf of colleagues at MSSL, IFCA, Leicester, NOA, AIP, Strasbourg

Particularly: Paul Kuin, Keir Burchell, George Mountrichas, Francisco Carrera, Angel Ruiz, Ioannis Georgantopoulos, Ektoras Pouliasis, Thanassis Akylas, Ada Nebot

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101004168

INTRODUCTION

- XMM-Newton has been flying for more than 20 years.
- 20 years of the XMM-SSC and others working to characterise the X-ray source population
- Not much effort put into characterising the UV/optical source population imaged with XMM-OM.
- But there are 6 million sources and counting!
- UV data are unique.
- Simultaneity with X-ray data are unique.
- These data are really missing out on a classification scheme.



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101004168

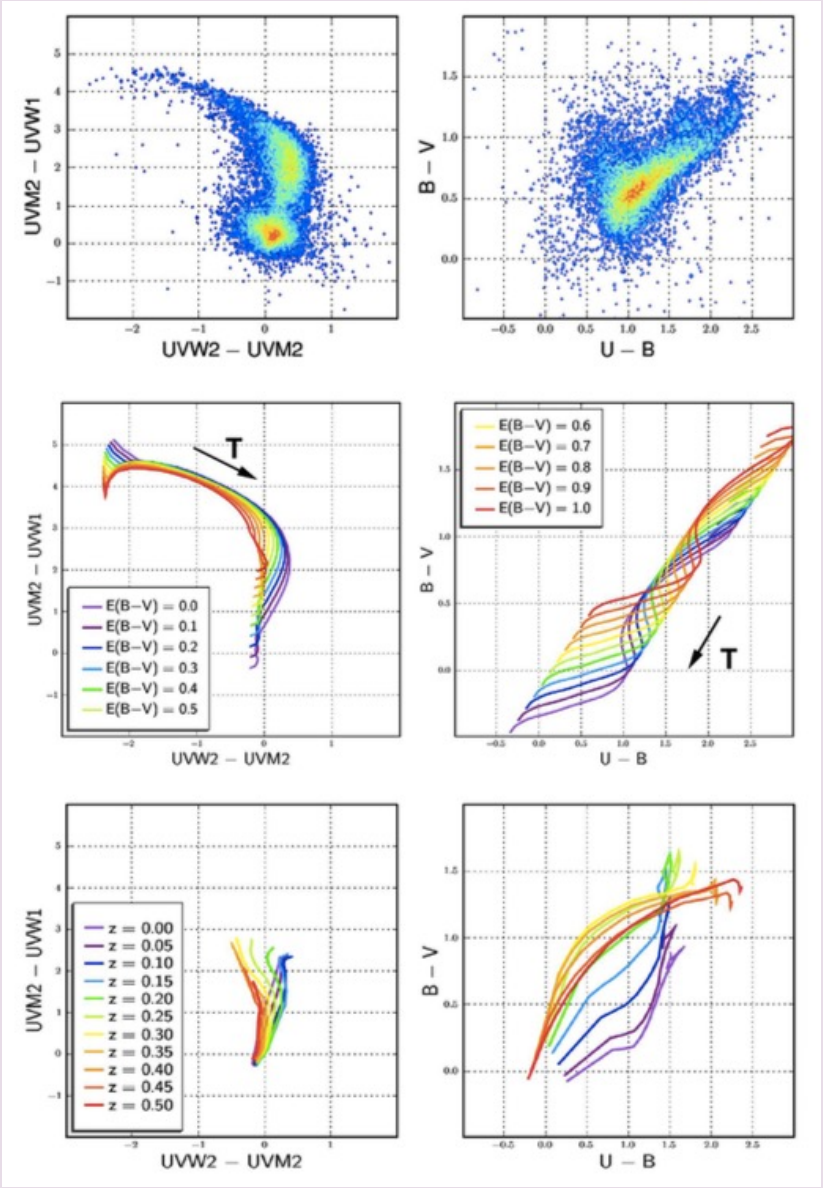
OBJECTIVES

- **Classify the sources in the XMM-Newton Serendipitous UV Source Survey (XMM-SUSS, the XMM-OM catalogue) into Galactic and extragalactic source types.**
- **We want broad applicability, i.e. the majority of the XMM-SUSS 5.0 sources.**
- **Add photometric redshifts for the extragalactic sources.**

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101004168

UNDERSTANDING PRE-XMM2ATHENA

- We already knew that we can classify the sources to some extent from the SUSS data when we have UV colours.
- Models for star and galaxy colours show that we can at least discriminate between galaxies and stars.
- Morphology also helps for brighter sources.
- Much harder if we just have the optical colours.
- And (of course) often impossible if only one band observed.



Page et al. 2012, MNRAS, 426, 903

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101004168

UV/OPTICAL SOURCE CLASSIFICATION

Why have we gone with CLAXBOL (Tranin et al. 2021) approach?

- **If we had a homogeneous set of data we know we could get very reliable classification from more "classical" approaches.**
  - We've demonstrated this with PCA on a well-observed UV subsample of the data.
- **But our catalogue is highly inhomogeneous in data coverage.**
  - Wide variety of missing data in terms of photometric coverage.
- **Experiments have taught us how easy it is to achieve apparently superlative classification that's actually spurious.**
- **Claxbol is reasonably understandable, and we could diagnose what was going on.**
  - Rubbish in, rubbish out.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101004168

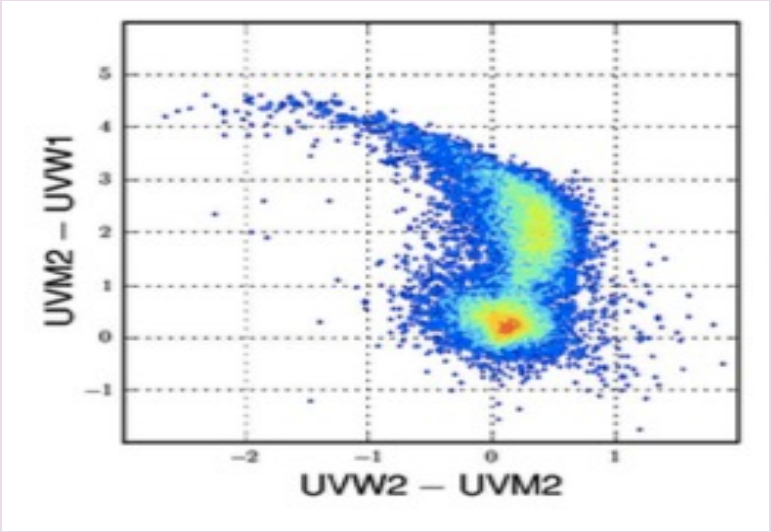
THE TRAINING SET

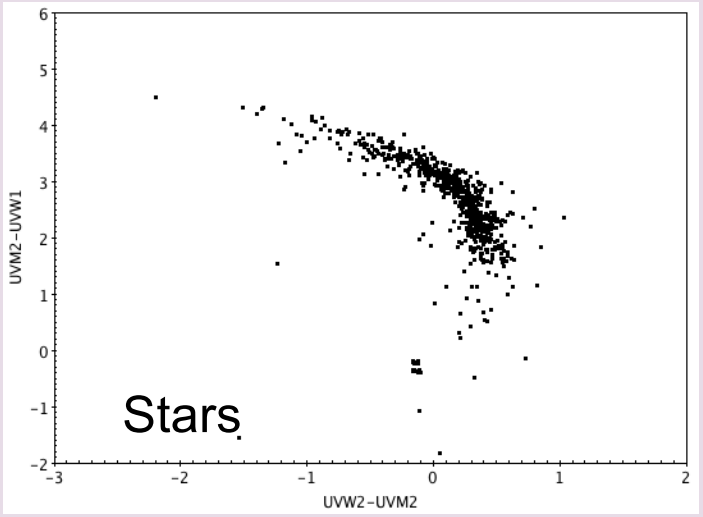
- Stars selected on the basis of significant parallax in Gaia. Very simple criterion. We know they are stars in our Galaxy because we can measure their parallax.
- Galaxies selected from SDSS. We require them to have a spectroscopic classification as a galaxy* from SDSS and to have a measured spectroscopic redshift.
- QSOs are selected from SDSS. We require them to have a spectroscopic classification as a QSO from SDSS and to have a measured spectroscopic redshift.

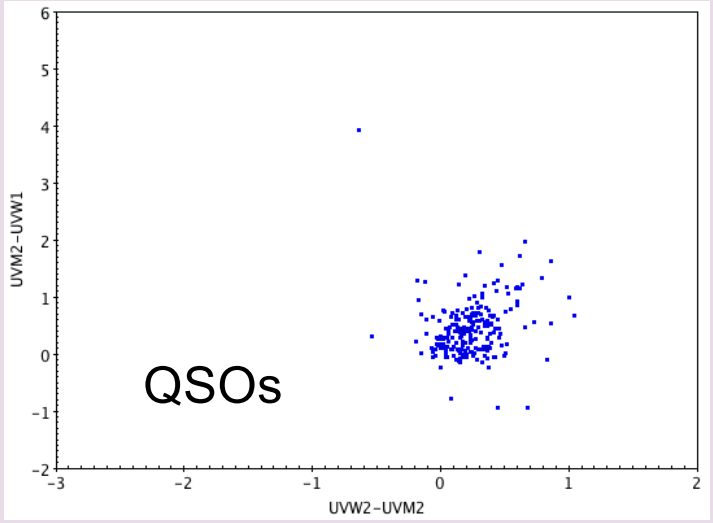
* This is to stop broad-line AGN contaminating the galaxy subsample, which would mean that even in the training sample the QSO and galaxy categories would be inherently blurred

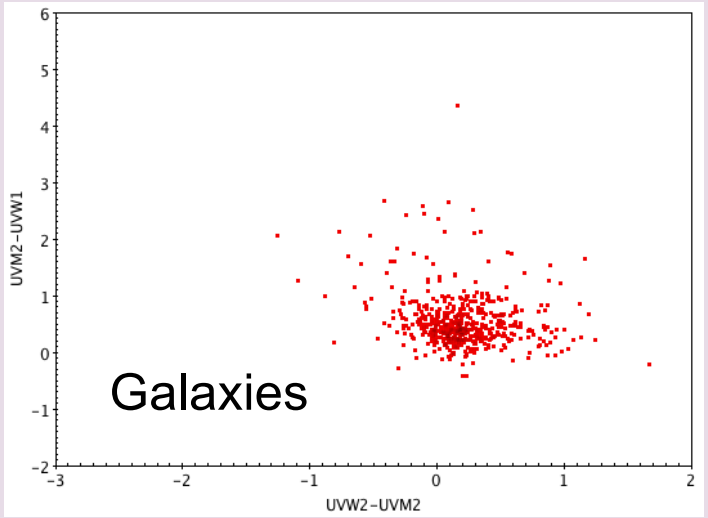
This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101004168

THE TRAINING SET









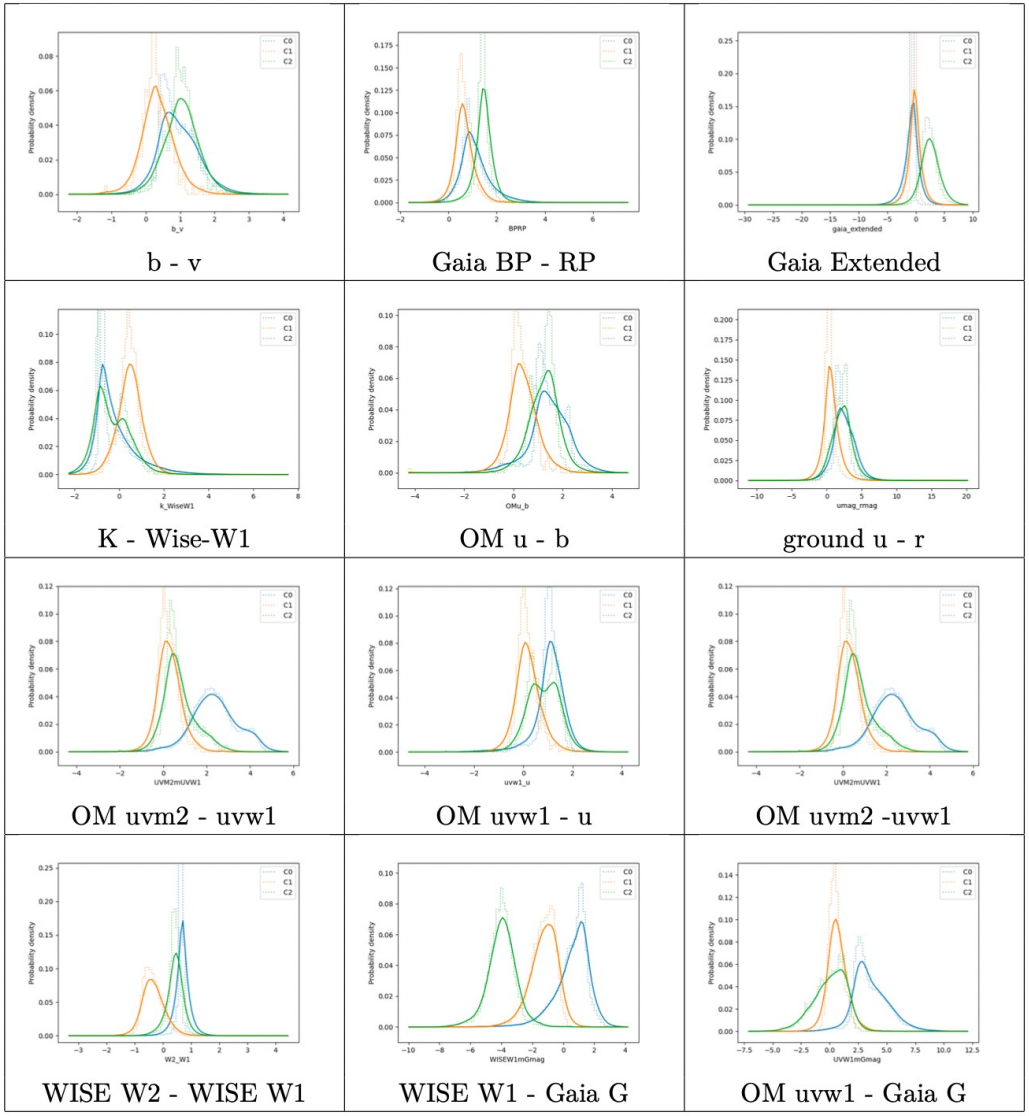
➢ **Training set confirms and quantifies UV colour diagnostics**

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101004168

CARRY OUT THE AUTO-CLASSIFICATION

- We use the CLAXBOL method of Tranin et al. (2021).
- 12 Parameters are colours, extent, and colour-extent hybrids.
- We don't use location, flux or magnitude, parallax, proper motion.

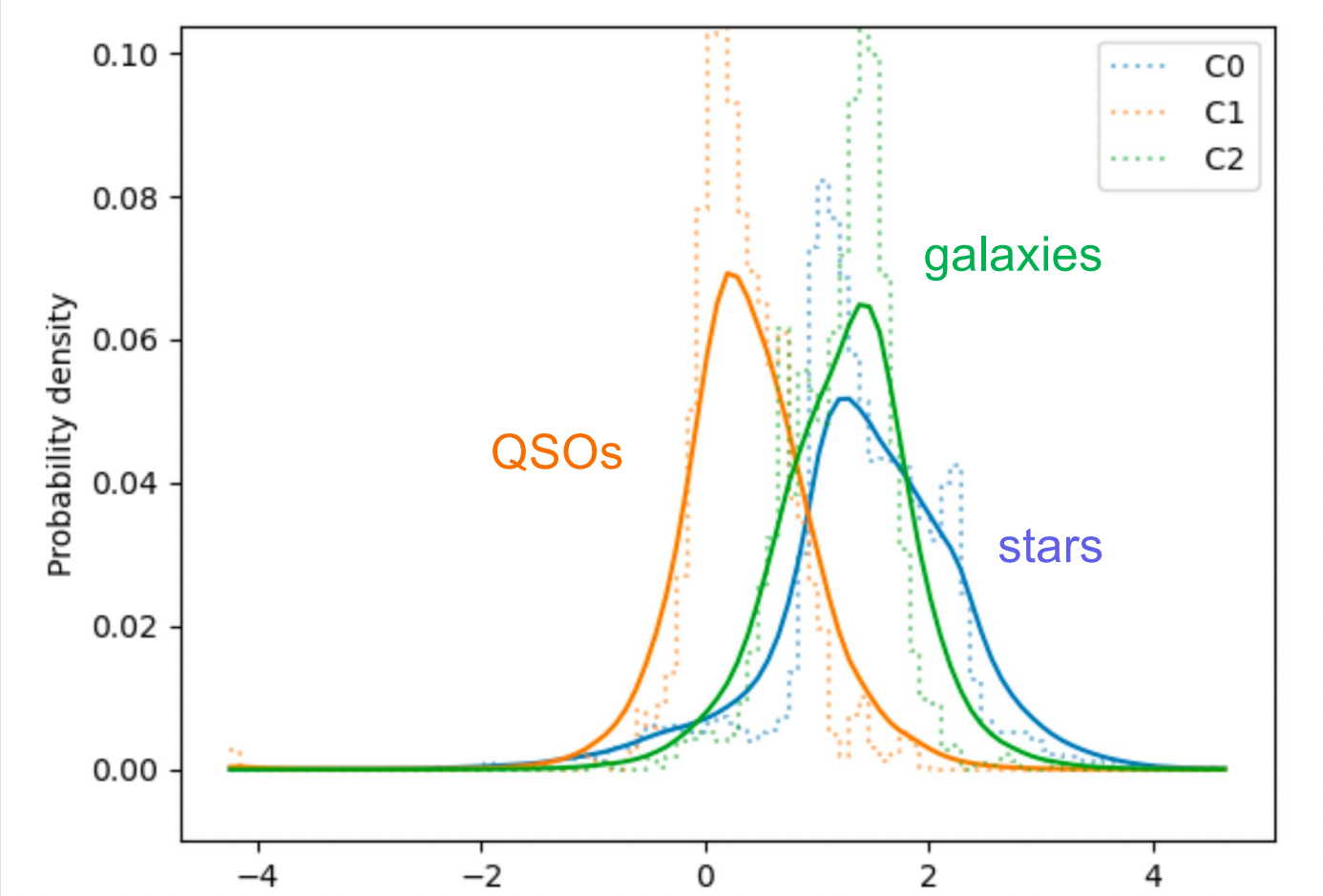
QSOs
galaxies
stars



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101004168

COLOUR

- **Colour is an age-old means to discriminate different source types.**
- **U-B is an example that discriminates QSOs from other sources, but doesn't discriminate between stars and galaxies.**
- **Most of our parameters are colours. Large variation in how many colours are available for different locations in the sky.**

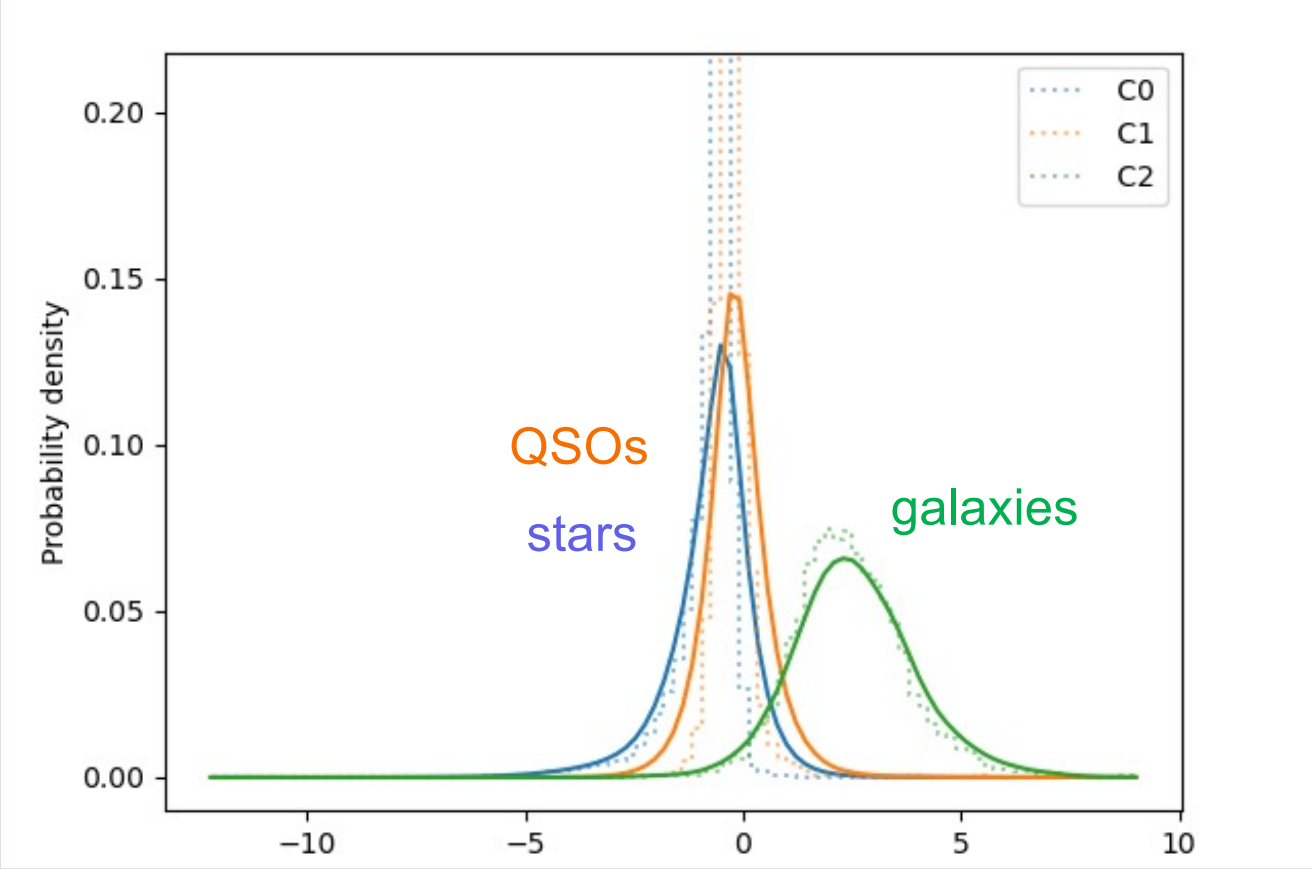


OM U-B

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101004168

EXTENT

- Whether the source is pointlike or not.
- Information frequently provided in some sense by ground based surveys, but highly inhomogeneous.
- Best measure we've found is to contrast Gaia and ground based G magnitude.

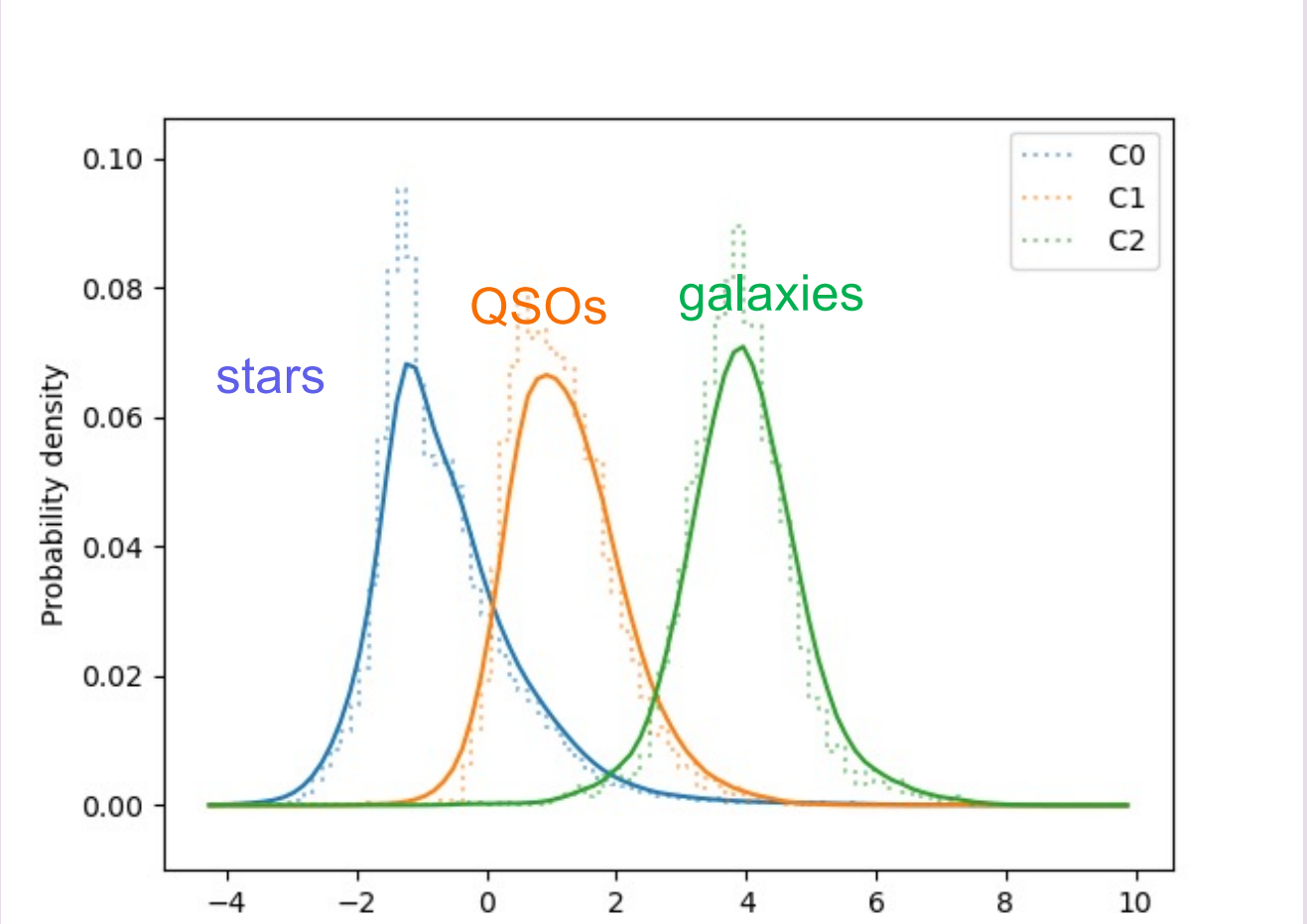


Gaia G – ground G

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101004168

COMPOSITE EXTENT AND COLOUR

- In colour terms, galaxies and QSOs are redder in G-W1 colour than stars. But galaxies and QSOs are indistinguishable.
- When we use Gaia for the G magnitude, the extent of galaxies makes them much fainter. So in a Gaia G – WISE W1 colour, galaxies now are easily distinguished from stars, and even quite easily distinguished from QSOs.



Gaia G – Wise W1

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101004168

HOW WELL DOES IT WORK (1)

	Real star	Real QSO	Real galaxy	Total classified	Precision %
Appears to be star	62647	67	40	62754	99.5
Appears to be QSO	795	3476	256	4537	86.2
Appears to be galaxy	251	636	5541	6428	88.4
Total	63693	4179	5837		
Recall %	98.4	83.2	94.9		

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101004168

HOW WELL DOES IT WORK (2)

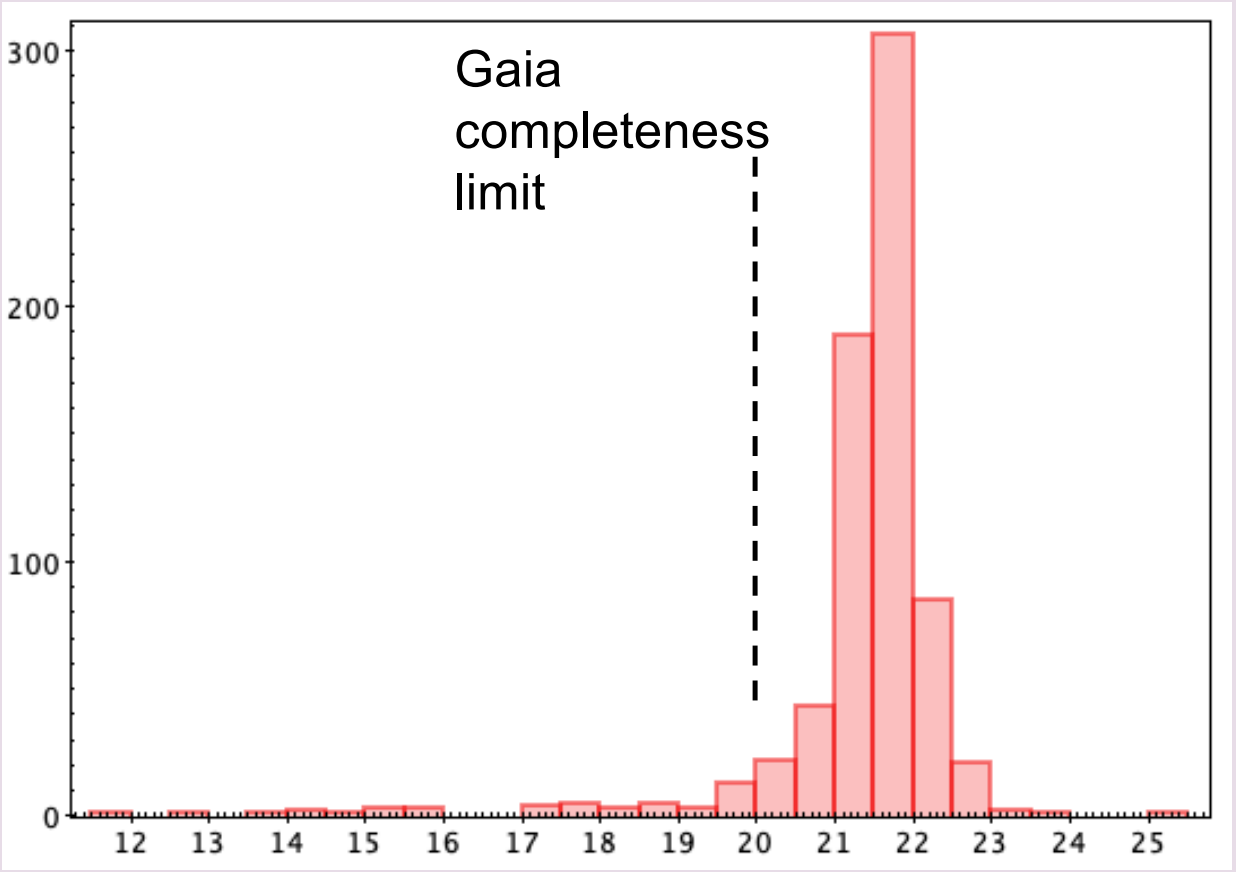
- We need to be cautious about using the training set to assess how well we do.
- Comparison made to Veron et al. 2010 QSOs and AGN catalogue 13th edition (A&A 518, A10)
- Matched to SUSS 5.
- We classify 89% of 2562 VV10 QSOs as QSOs. If we exclude sources in common with our training set, we still get 89%
- Similarly, matching QSOs classified in SIMBAD with SUSS 5, we classify 84% of 8905 SIMBAD QSOs as QSOs.
- Independent checks reasonably close (slightly better) than the 83% "recall" for QSOs derived from the training set.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101004168

WHAT ARE THE PROBLEM SOURCES?

☹ ☹ sources

- These are the quasars classified as galaxies.
- Key issue: they're faint.
- Too faint for Gaia, so too faint to distinguish extended from point like sources



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101004168

TRAINING SET PITFALLS

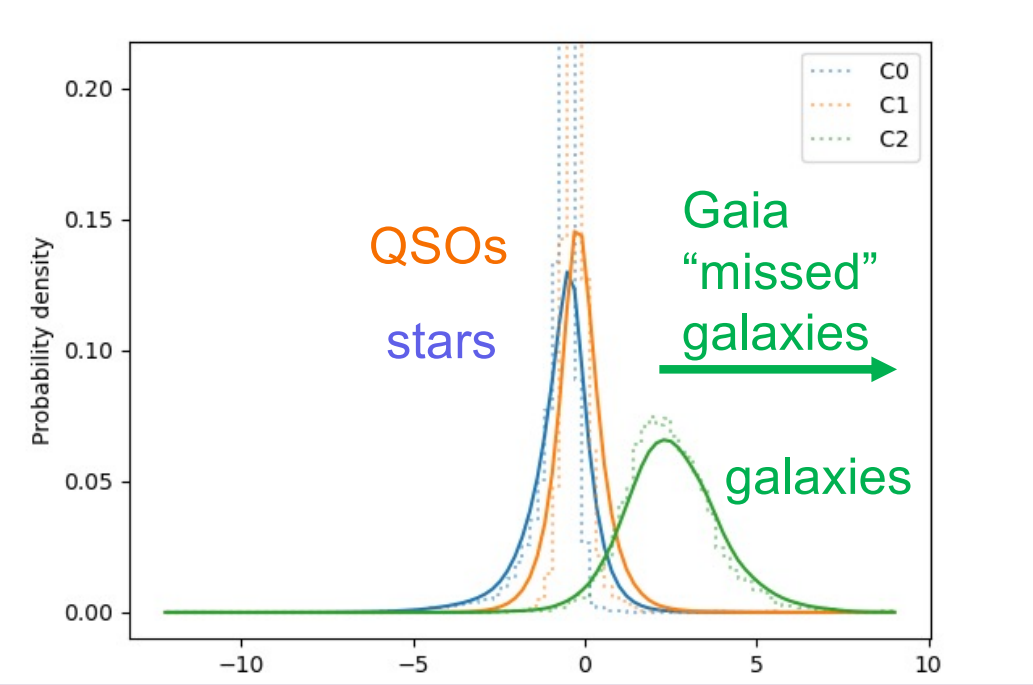
- The way that the training sets are selected has significant impact on which parameters can be used for classification.
- An example is Gaia parallax. Parallax was used to select the stars training set. So all of the stars in the training set have significant parallax.
- Not surprisingly, parallax is not significant for the QSOs and galaxies used in the training sets.
- If we try to use parallax in the auto-classification it will discriminate stars from extragalactic sources almost perfectly in the training set, and it will look wonderful in the truth table. But if we then unleash the auto-classification on the SUSS, all distant stars will be classified as QSOs and galaxies. 
- Similarly, as we use SDSS to define QSO and galaxy training sets, the footprint of SDSS on the sky is encoded in the training sets. If we use Galactic latitude as a parameter in the training, we will get excellent discrimination between stars and extragalactic objects with the training sets. But because SDSS doesn't cover the Southern Galactic cap, all QSOs and galaxies in that part of the sky will be confidently classified as a Galactic stars. 

- It's a shame, because parallax and Galactic latitude are powerful classifiers

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101004168

UPPER / LOWER LIMITS

- Upper limits turn out to be important.
- Remember this measure of extent using Gaia:
- Galaxies are systematically faint in Gaia catalogue. In fact many galaxies are so faint that they have no Gaia measurement at all.
- Since Gaia covers the whole sky, absence from the Gaia catalogue is meaningful. Using the Gaia magnitude completeness limit we can derive lower limits on Gaia G – Ground G
- No obvious, rigorous way to treat upper limits in CLAXBOL.
- If galaxies have large enough lower limits on Gaia G – Ground G, we treat them as though they are detected at the Gaia limit.



Gaia G – ground G

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101004168

WHAT'S IN THE AUTO-CLASSIFIED SUSS?

- 4,679,099 stars
- 232,038 QSOs
- 1,054,104 galaxies

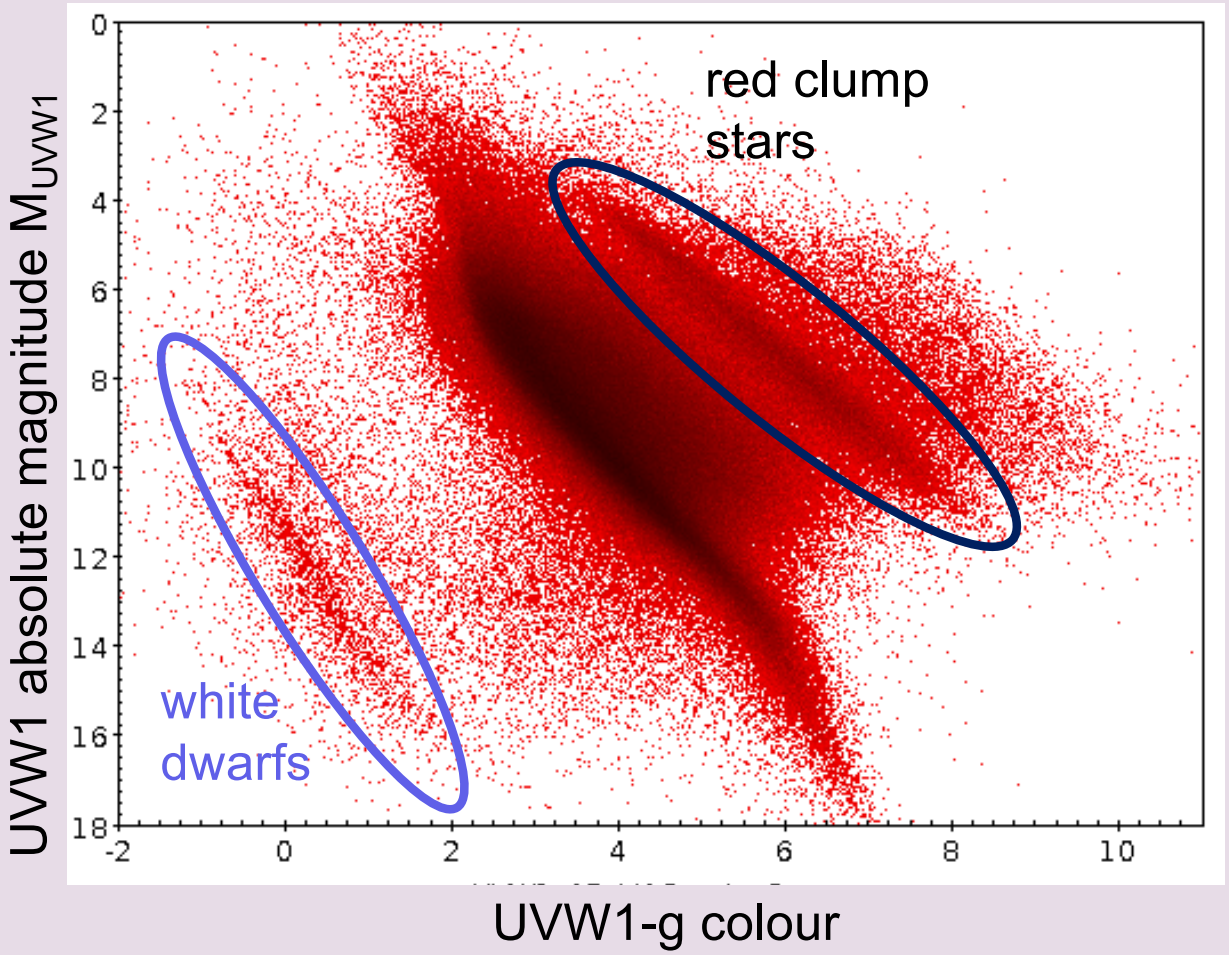
- Most of the 'leakage' is between QSOs and galaxies.
- Stars are quite 'clean'.
- So Galactic/extragalactic split is quite 'clean'

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101004168

WHAT SORT OF SCIENCE CAN WE DO? - 1

- We can select samples of stars, galaxies and QSOs in the UV.

- We can make white dwarf luminosity functions in the UV
- We can use red-clump stars to examine the variation in the 2175A bump in the Galaxy



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101004168

WHAT SORT OF SCIENCE CAN WE DO? - 2

- **Population studies of UV-selected QSOs**
- **Studies of Galactic extinction in the UV using QSOs as back-lights**
  - **The XMM-OM QSO sample isn't deliberately restricted to high Galactic latitudes**
- **Population studies of UV-selected galaxies**
- **Selection of X-ray-weak QSOs**
  - **Simultaneity of X-ray and optical advantageous**
  - **XMM-Newton pointings more sensitive than eRosita**

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101004168

BROADER BENEFITS IN XMM2ATHENA CONTEXT

- Classification type of OM sources can assist in correct association of X-ray and optical counterparts
- So can help in the classification of the X-ray sources

- XMM-OM QSO sample could help in aspect correction of EPIC images when the standard position correction fails
- QSOs are ‘reliable’ X-ray emitters to cross-correlate with X-ray source lists.
- XMM-OM covers highest sensitivity, best PSF parts of EPIC images.

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 101004168

CONCLUSIONS

- We've used CLAXBOL to classify the sources in the XMM-SUSS 5.0 catalogue.
- 4.7 million stars, 1 million galaxies, 230k QSOs.
- Good discrimination between Galactic and extragalactic sources
- Some leakage between QSOs and galaxies.

- Get the classifications at:
- http://xmm-ssc.irap.omp.eu/xmm2athena/catalogues/

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101004168