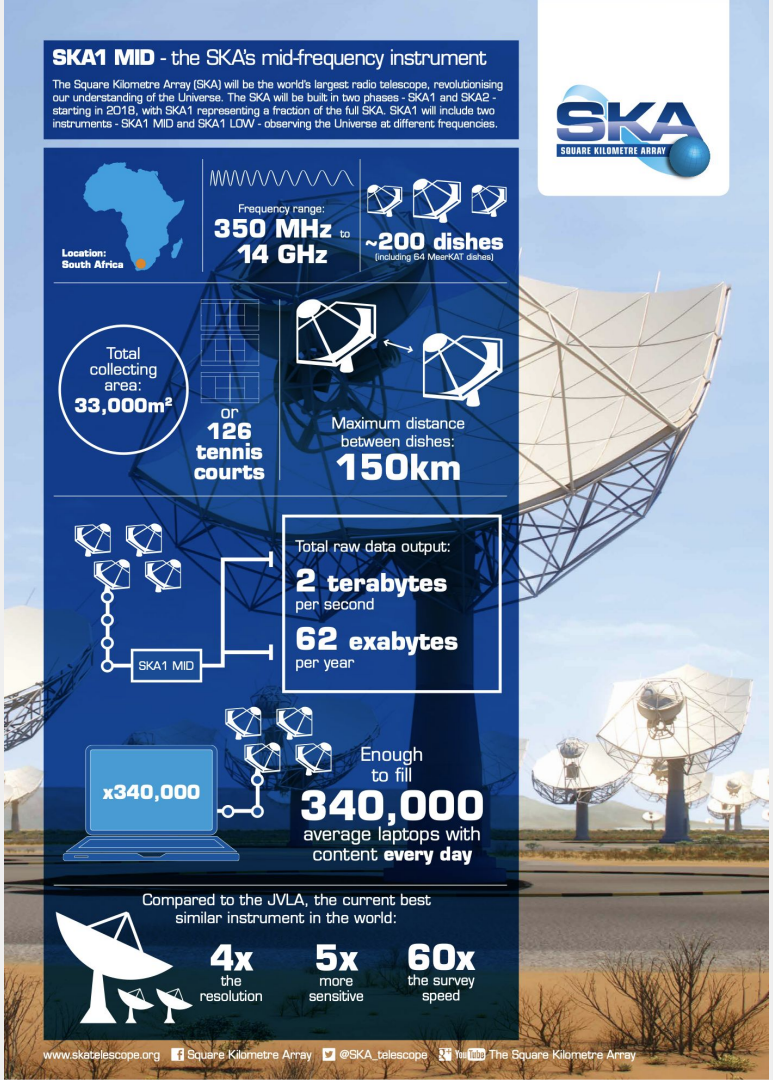
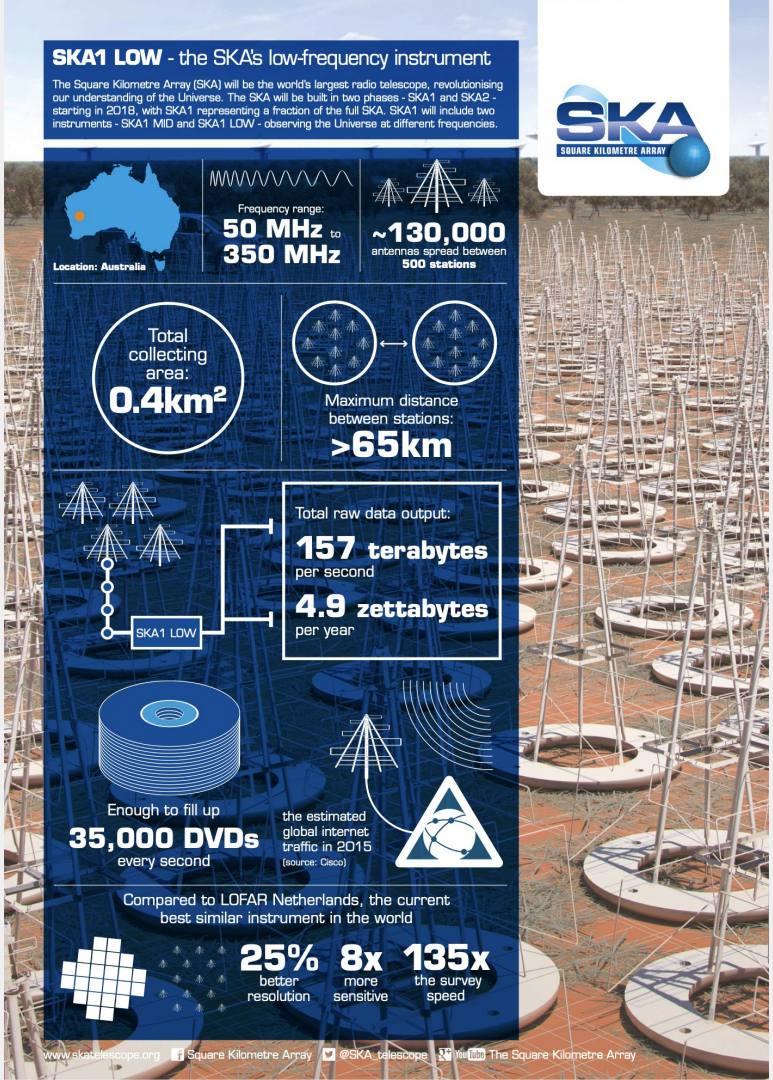
SKAO IRIS Activities

Rohini Joshi
r.joshi@skatelescope.org



Dish Array

SKA1 MID



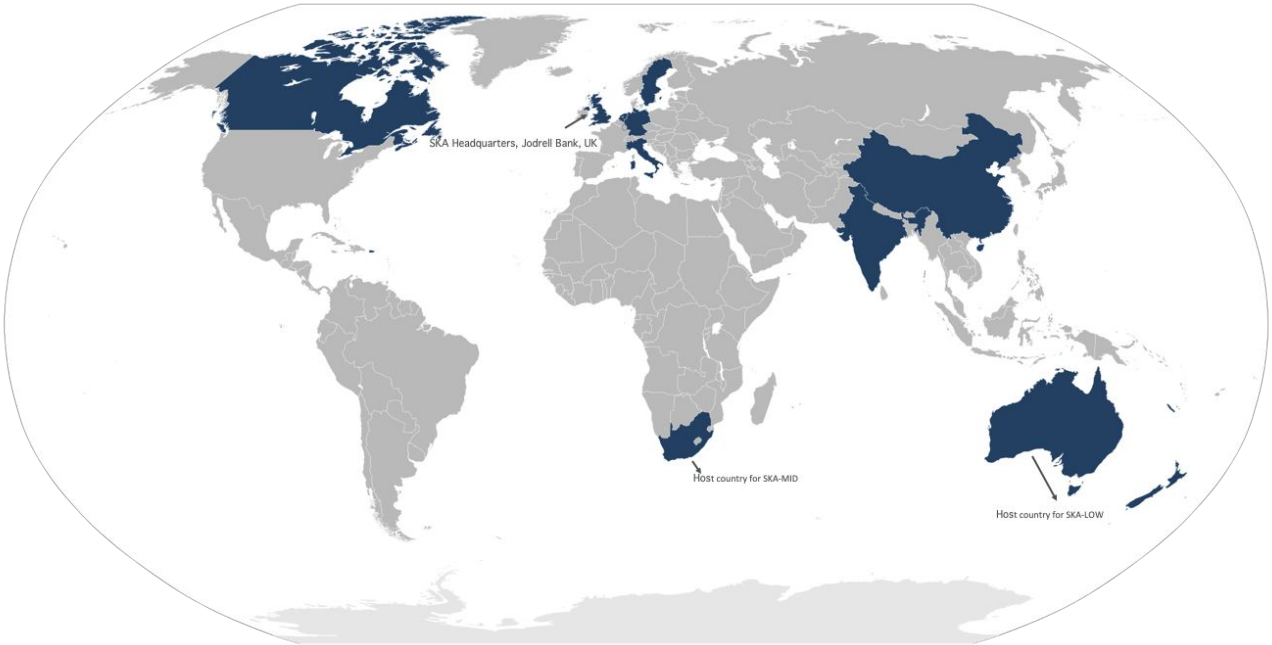
Aperture Array

SKA1 LOW

SKA Regional Centres

Science Data Processors at both the SKA sites will collect, process and churn out science data products that will be pushed out to one or more regional centres around the world

SKA Regional centres will provide a platform for data access, data distribution, post-processing, archival storage, and software development.



Our Team

SKA operates a scaled agile software development framework (SAFe), we are a part of the Data Processing Agile Release Train

ESCAPEES team: James Collinson, Rob Barnsley, Alex Clarke, Rosie Bolton

SRC related methods and technologies that aligns with responsibilities to ESCAPE

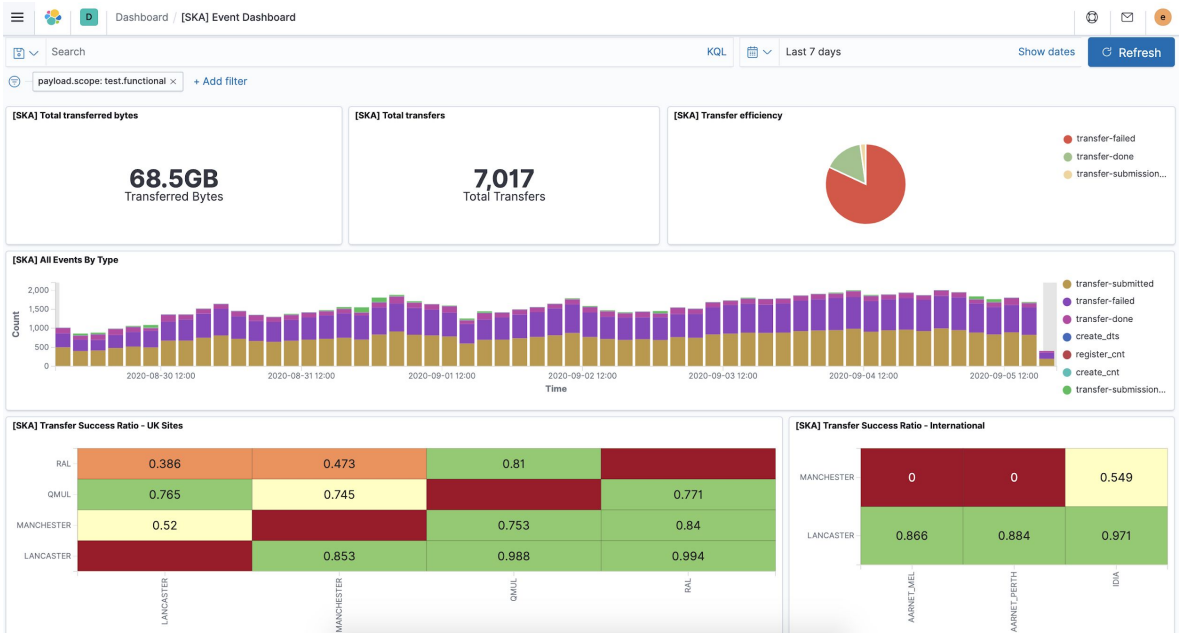
Continued updates

Rucio instance

- multi-VO upgrade
- Kibana dashboard
- Test playground
- Australian RSEs

Grid workflows, high memory node utilisation and MeerKAT pipeline work continued by Anna Scaife's group

STFC Cloud usage ...



STFC Cloud usage 1/3

- Kubernetes!
- Kubernetes for hosting JupyterHub IRIS/ESCAPE
  http://130.246.212.44/iris
  http://130.246.212.44/escape
  Deployment documented here
- For the most part, we don't know what we're doing! Happy to speak with other communities deploying similar things. User liaison meetings are great!
- Built an astronomy specific JupyterHub environments ie, RASCIL.
- Looking forward to LBaaS, and Cinder volumes

Img Credit

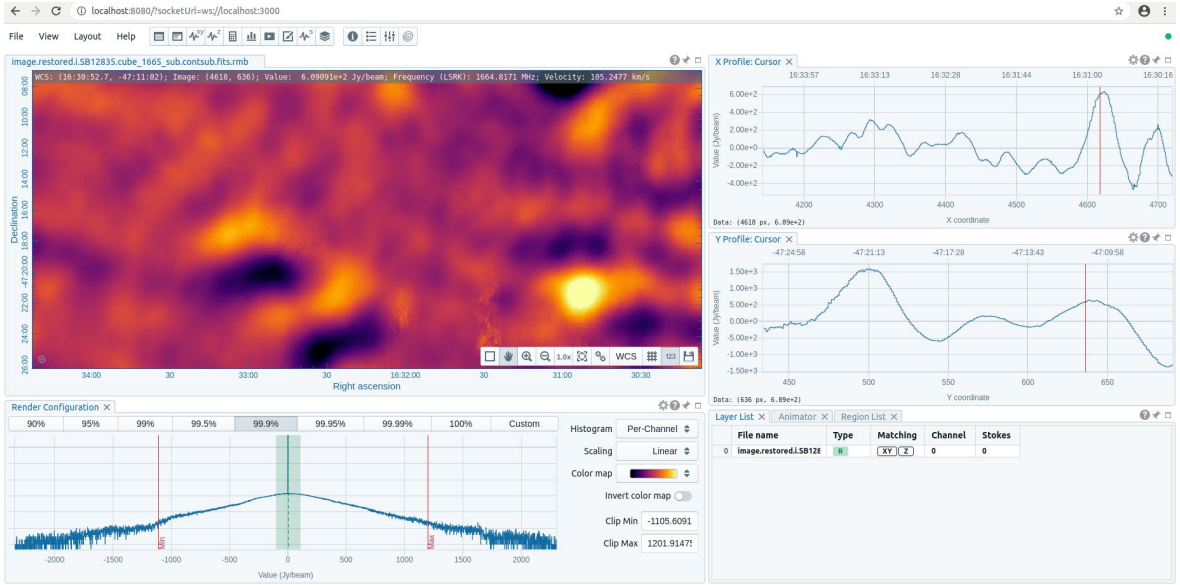
STFC Cloud Usage 2/3

SKA Science machine

- Similar specs to the in-house science server hosted at SKA
- Made available to the Science Team (via ssh pub key)

ASKAP analysis

- Visualising ASKAP data to determine suitability
- ‘Desktop’ installation of the CARTA viewer vx.x https://carta.readthedocs.io/en/latest/

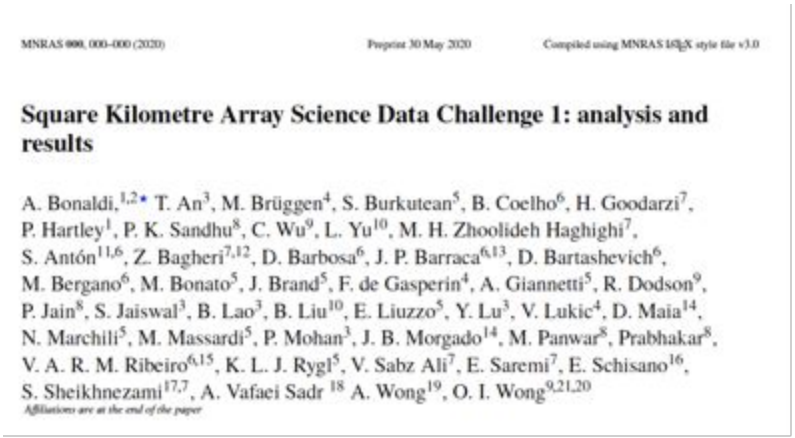


Pilot field of the 1665 MHz OH line taken for the GASKAP survey
Image Credit: Shari Breen, Rob Barnsley

Slide Credit: Philippa Hartley

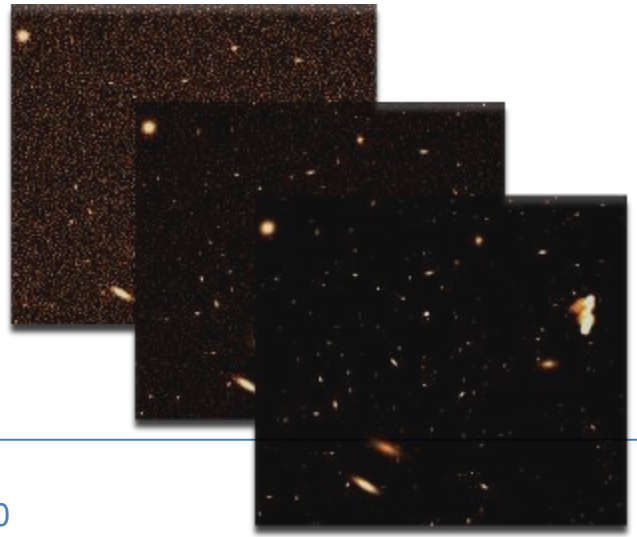
SDC1: Continuum emission

- Datasets of continuum emission images, simulating observations for SKA MID Bands 1, 2 and 5
- 32k x 32k Images populated by star forming galaxies (SFGs) and active galactic nuclei (AGN)
- Challenge description: To find and characterise sources
- Results now submitted for publication in MNRAS
- Data hosted on ownCloud. Submissions, support, troubleshooting handled over email
- Challenge was live Nov 2018 - May 2019
- https://astronomers.skatelescope.org/ska-science-data-challenge-1/


MNRAS 000, 000–000 (2020) Preprint 30 May 2020 Compiled using MNRAS LATEX style file v3.0

Square Kilometre Array Science Data Challenge 1: analysis and results

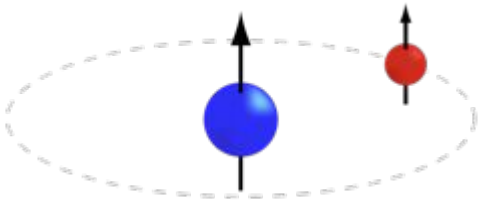
A. Bonaldi,[1,2]* T. An[3], M. Brüggen[4], S. Burkutean[5], B. Coelho[6], H. Goodarzi[7], P. Hartley[1], P. K. Sandhu[8], C. Wu[9], L. Yu[10], M. H. Zhoolideh Haghighi[7], S. Antón[11,6], Z. Bagheri[7,12], D. Barbosa[6], J. P. Barraca[6,13], D. Bartashevich[6], M. Bergano[6], M. Bonato[5], J. Brand[5], F. de Gasperin[4], A. Giannetti[5], R. Dodson[9], P. Jain[8], S. Jaiswal[3], B. Lao[3], B. Liu[10], E. Liuzzo[5], Y. Lu[3], V. Lukic[4], D. Maia[14], N. Marchili[5], M. Massardi[5], P. Mohan[3], J. B. Morgado[14], M. Panwar[8], Prabhakar[8], V. A. R. M. Ribeiro[6,15], K. L. J. Rygl[5], V. Sabz Ali[7], E. Saremi[7], E. Schisano[16], S. Sheikhnezami[17,7], A. Vafaei Sadr [18] A. Wong[19], O. I. Wong[9,21,20]
Affiliations are at the end of the paper



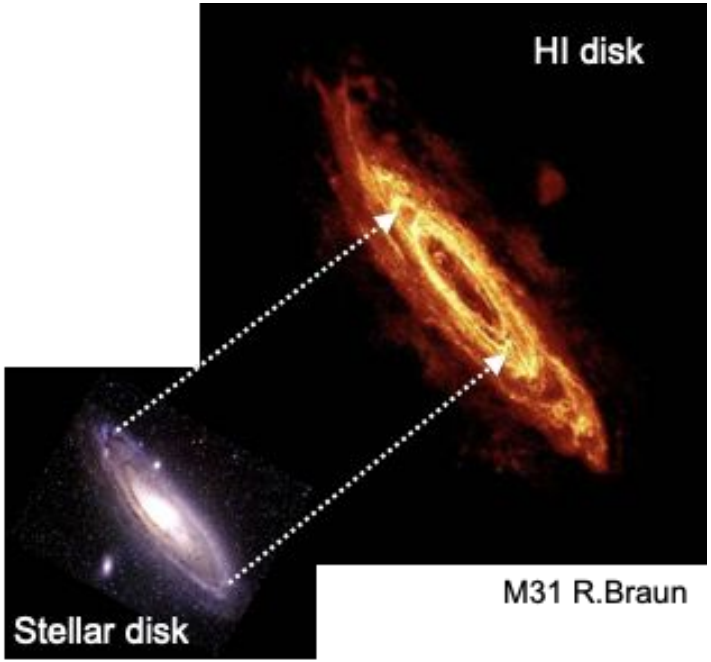
Slide Credit: Philippa Hartley

SDC2: Neutral hydrogen (HI)

- Neutral hydrogen (HI) found in star forming galaxies and in tidal interactions in the intergalactic medium
- HI emits radiation during a 'spin flip' of its single electron

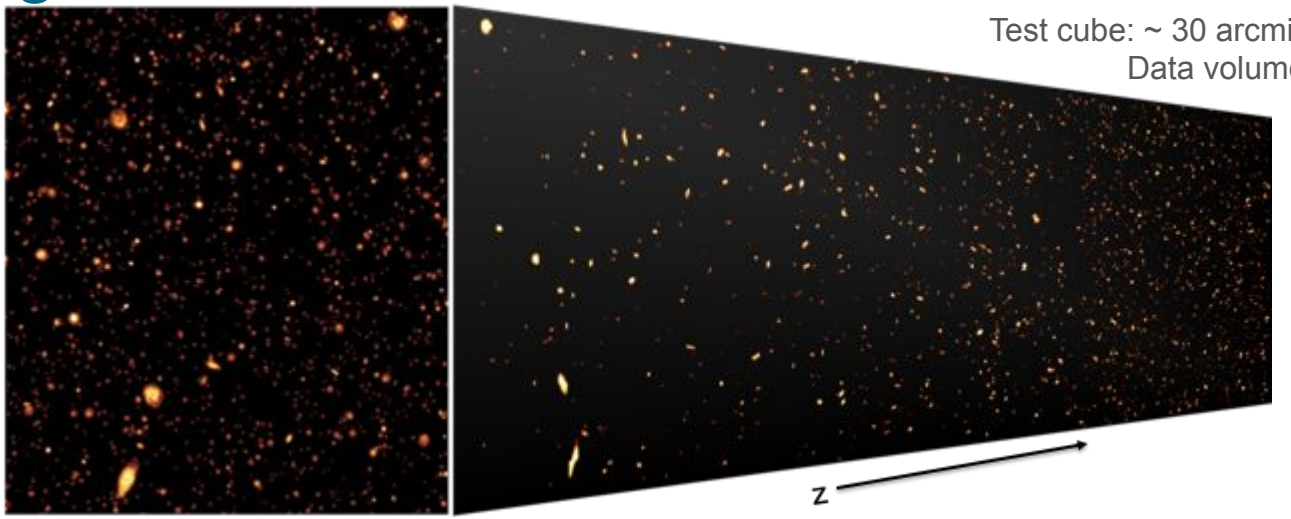


- HI spectral line at ν_0 = 1.42 GHz (λ_0=21 cm) redshifted to lower frequencies
- SKA will trace HI from the current time, all the way back to the formation of the first galaxies during 'Cosmic Dawn'



Slide Credit: Philippa Hartley

(Fairly) Big data



Test cube: ~ 30 arcmin across
Data volume: 16 GB

Integration time = 2000h
Spatial resolution = 5 arcsec
Frequency resolution = 115 kHz
RMS per channel 13-18 muJy

FoV = 20 square degrees
Frequency = 950 MHz – 1350 MHz (z = 0 to 0.5)
Data volume = 1TB

SDC2 access and processing

- The primary challenge is source finding and characterisation
- Webpage for overview of the challenge, resources available, submission templates, discussion forum, facilitate internal and external collaboration http://sdc2.astronomers.skatelescope.org, EoI
- We will be arranging access to compute facilities for participants to solve the data challenge
- Preparing the scientific community to deal not only with SKA-like data, but also future SKA/SRC practices
- Exploring the functionalities of current proto-SRCs

STFC Cloud Usage 3/3

Solving SDC1 - classifying sources in radio images

- SDC1 solution will likely form the basis for the more advanced pipeline required for SDC2 and will help test SDC2 platform hosts
- Demonstrating best-practice with respect to making our solution open, portable, and efficient.
- https://gitlab.com/ska-telescope/sdc/sdc1-solution

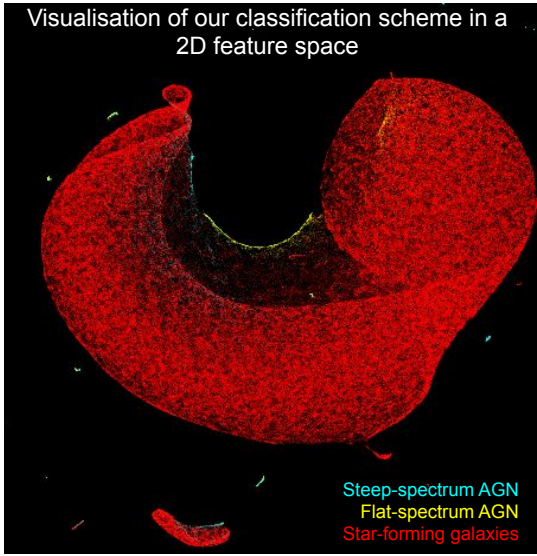
SDC2 simulation testing

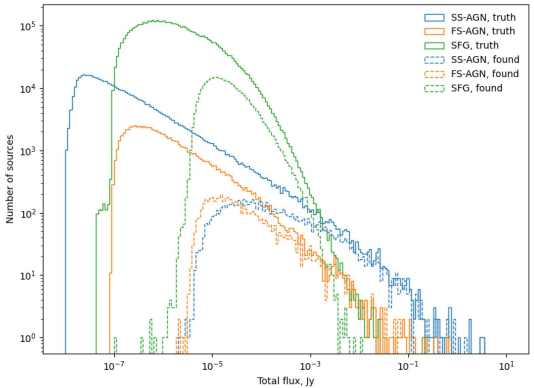
- Building and testing the portability of the SDC2 simulation pipeline
- Given the size of the input data cube, will be best to create the cube at site rather than transfer

Applied a Random Forest algorithm to classify sources based on features output by a source finder.

Visualised the results in a 2D space using dimension reduction algorithm called UMAP.

Compare results to the truth catalogue to assess performance





Finding AGN is hard because they are the minority class

SDC2 host platform requirements

- We expect ~30 multi-user teams with the challenge live for 6 months - 1 year
- Each team will likely require RAM-heavy interactive resources: at least 8 cores, 60-90 GB RAM, 100 GB+ disk with shared access to the input cube.
- Interested facilities include the Shanghai Astronomical Observatory (SHAO) HPC, EngageSKA (Portugal), INAF ICT, SKA France
- Working to establish format of IRIS resources offered and considering the operational responsibility of the same
- Looking into resource management and how to ensure fair consumption of resources among teams

In progress...

Next steps

SDC2

- Solidifying platform requirements, user models, timelines for SDC2
- Initiating conversations with STFC Cloud
- SDCSS
- Accounting dashboard consolidation

Rucio

- Deploy an ElasticSearch/Influx DB database to populate ESCAPE datalake monitoring dashboard with Rucio events
- Custom environment to interact with data in the ESCAPE data lake
- Rucio real-data transfers from IDIA to Manchester using the multi-VO instance

Thank you!