The Radio Universe in Full Color

The IDIA Cloud and the HIPPO Project

Mattia Vaccari

www.mattiavaccari.net

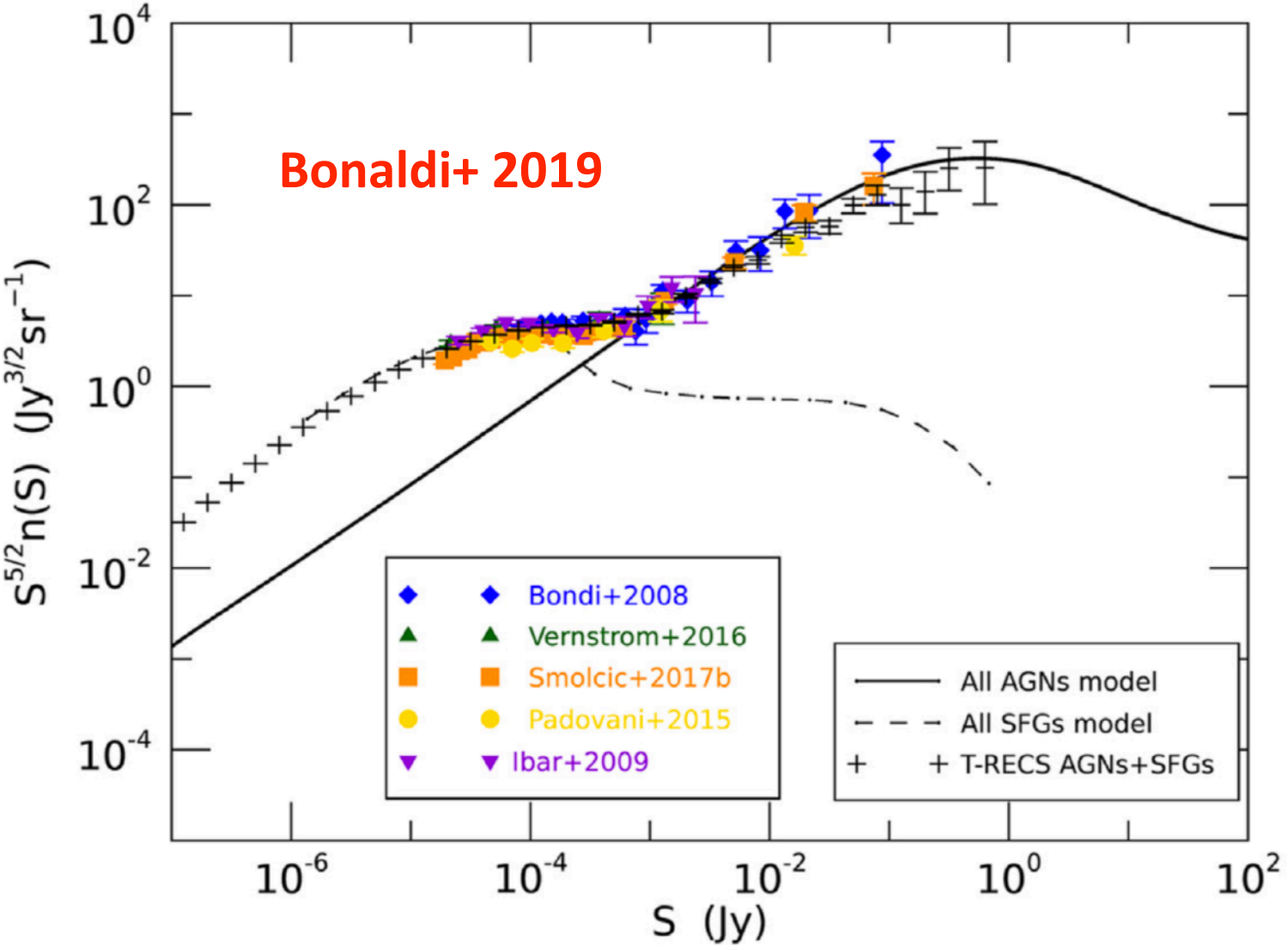
Institute for Data Intensive Astronomy (IDIA)

University of the Western Cape (UWC)

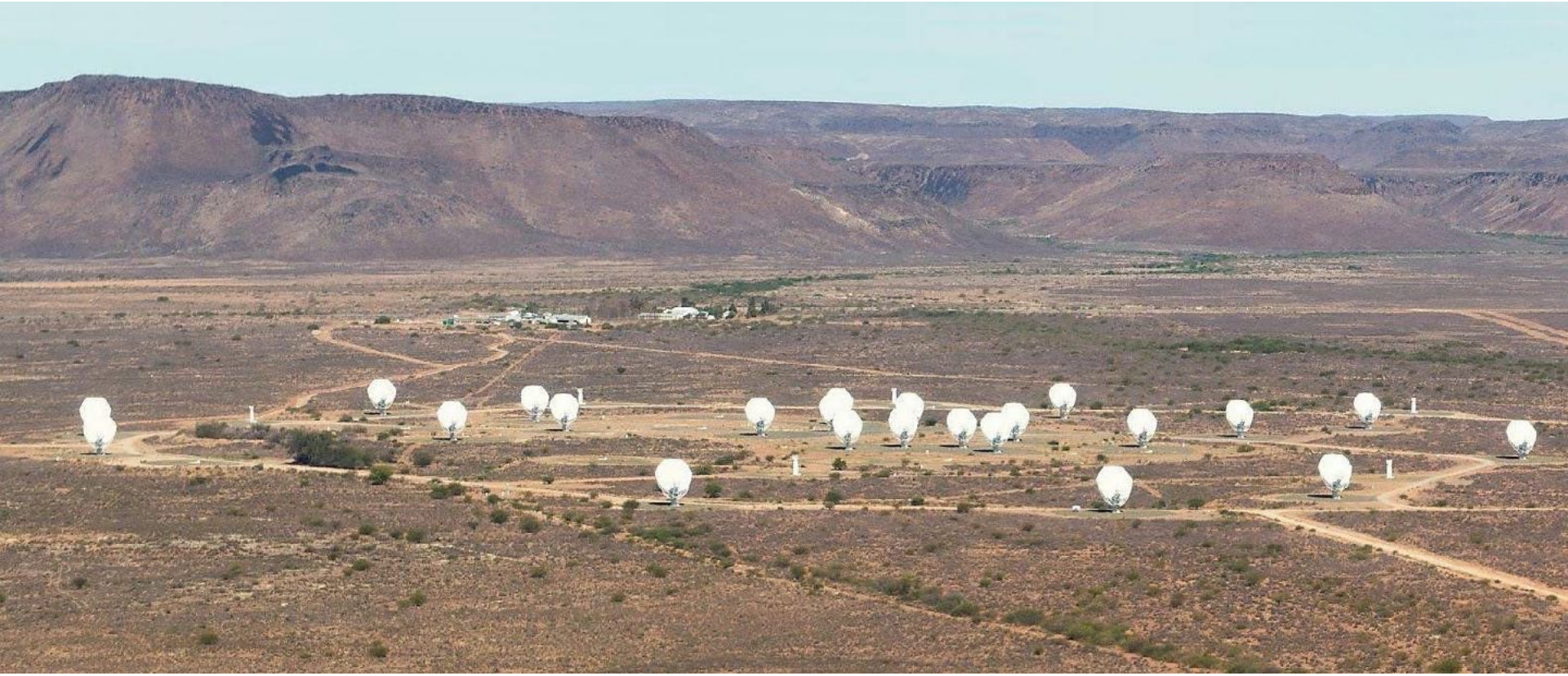
Cape Town, South Africa

www.idia.ac.za

The Faint Radio Sky

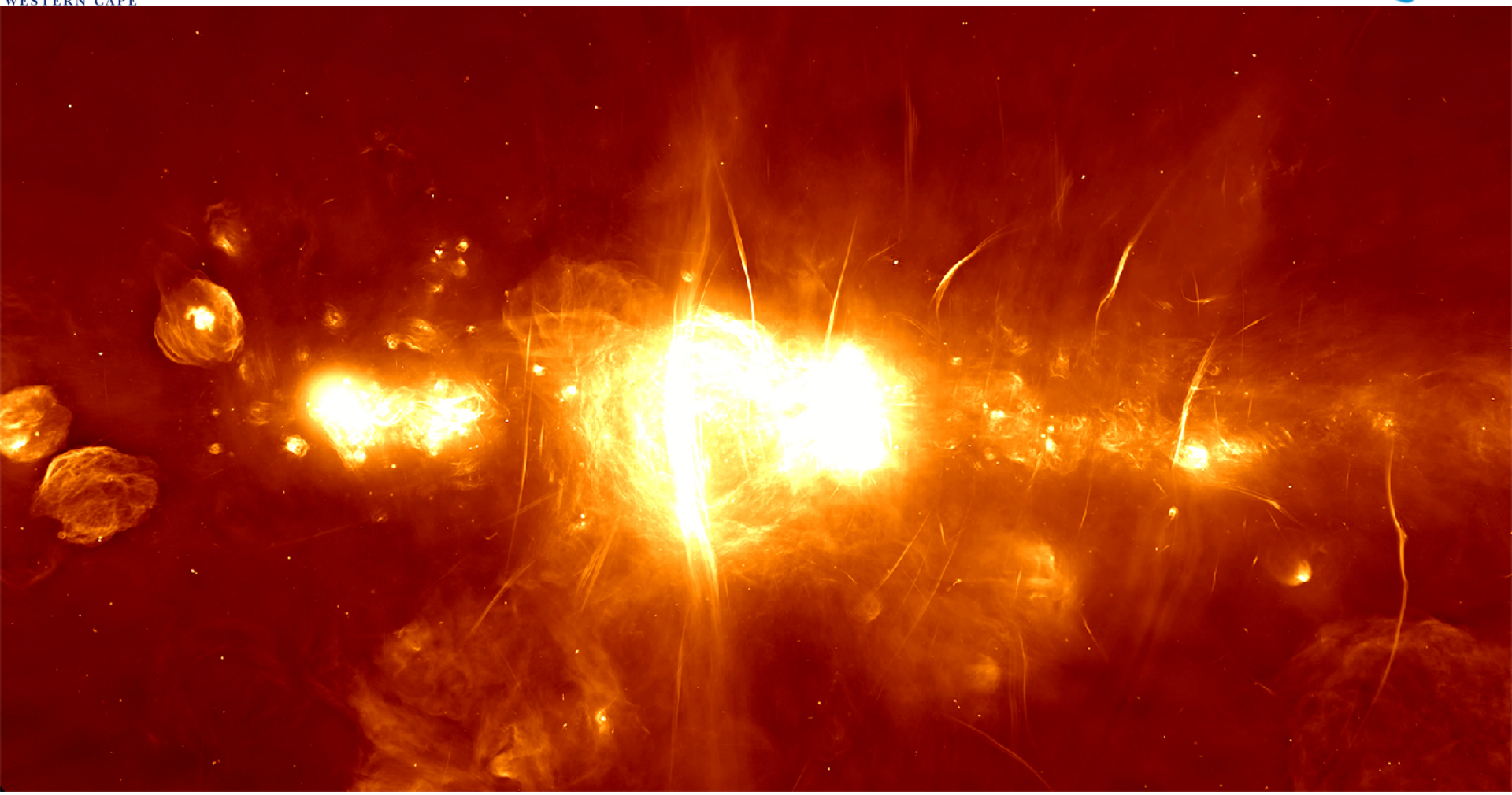


MeerKAT : SA's SKA Precursor



- MeerKAT is SA's SKA Precursor (or 'Phase 0' of SKA-Mid)
- Completed (on schedule and within budget) in Mid-2018
- Delivering Transformational Science from Day One
- Will be owned and operated by South Africa for 5 years

MeerKAT : SA's SKA Precursor



- Completed (on schedule and within budget) in Mid-2018
- Delivering Transformational Science from Day One
- Will be owned and operated by South Africa for 5 years

MeerKAT : SA's SKA Precursor

MeerKAT Large Survey Projects

imaging

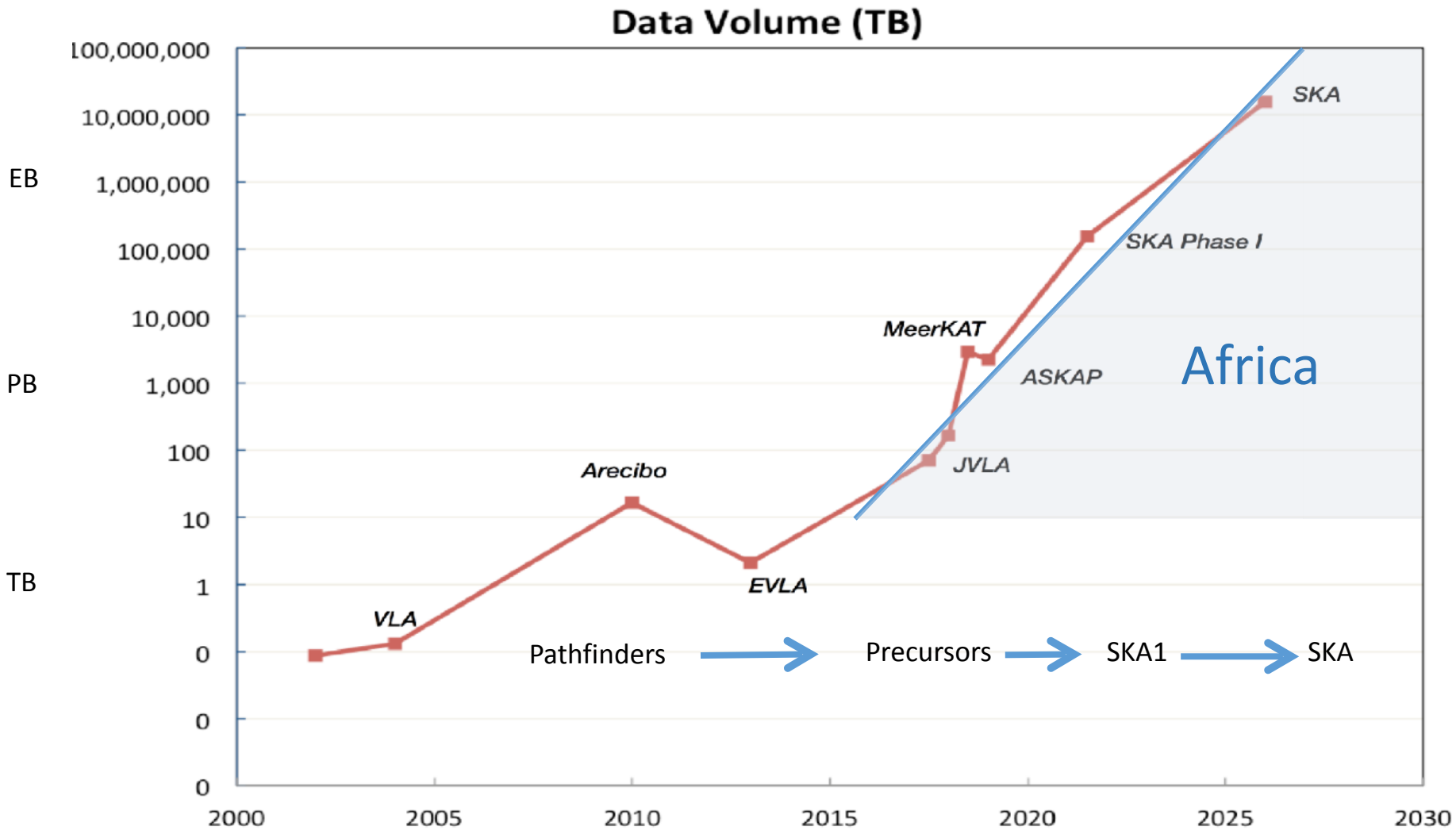
- LADUMA (Ultra-deep atomic hydrogen)
- MIGHTEE (Deep continuum imaging of the early universe)
- Fornax (Deep HI Survey of the Fornax cluster)
- MHONGOOSE (targeted nearby galaxies HI)
- MeerKAT Absorption Line Survey (extragalactic HI absorption)

Time domain

- ThunderKAT (exotic phenomena, variables and transients)
- TRAPUM (pulsar search)
- MeerTime (pulsar timing)
- MESMER (High-z CO)
- MeerGAL (Galactic Plane Survey)

http://public.ska.ac.za/meerkat/meerkat-large-survey-projects

Radio Astronomy Data Glut



- Plus similar challenges in Bioinformatics and Earth Observations

- The “Inter-University Institute for Data Intensive Astronomy” - IDIA - was launched in Sep 2015
- Driven by SKA Data Delivery, Processing & Mining challenges (plus MeerKAT science exploitation)

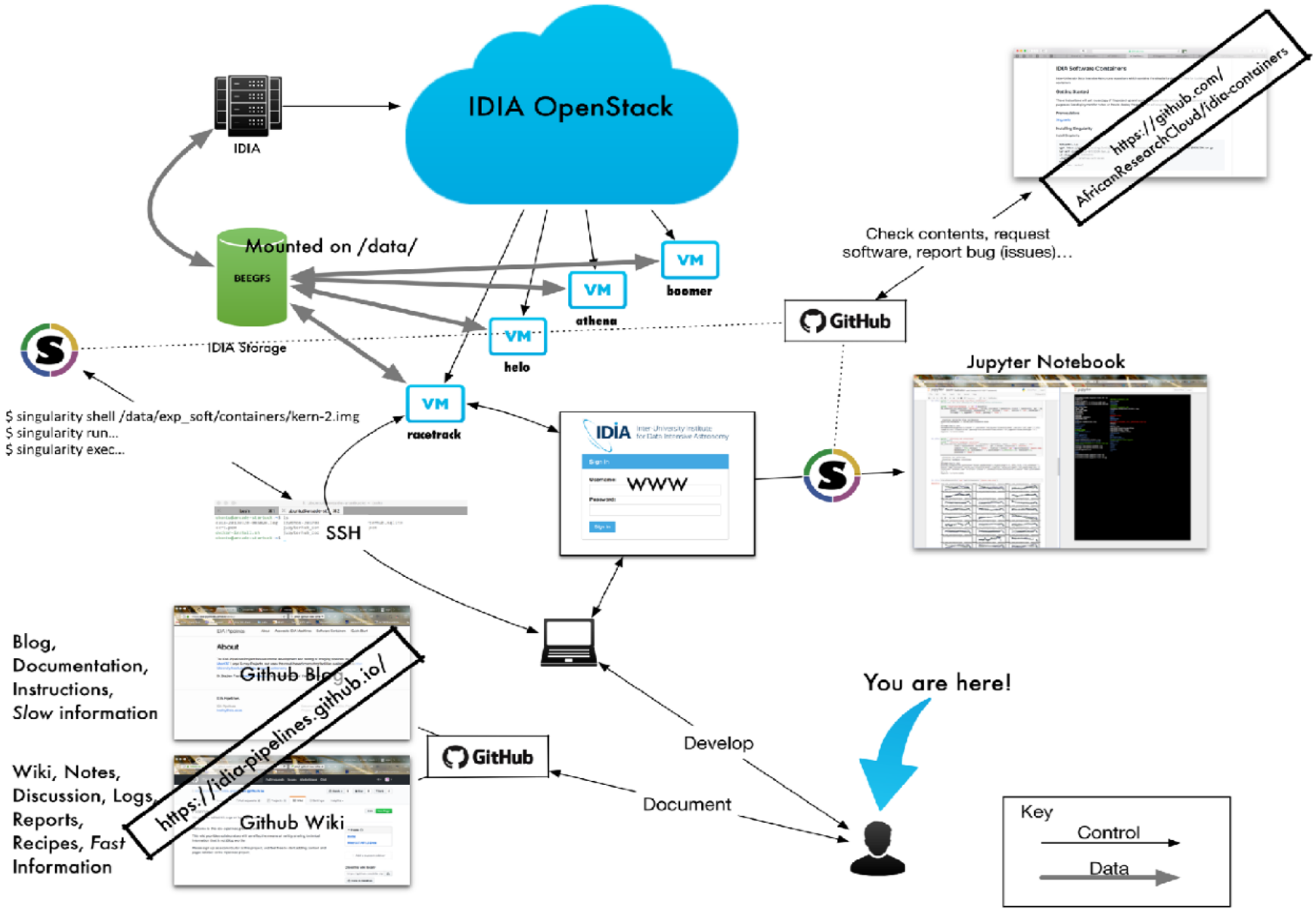
IDIA's Data Intensive Research Cloud

- IDIA is a University Consortium (UCT/UWC/UP)

- DIRC Phase I operational August 2017
- Data-centric architecture for cloud-based, data-intensive research
- 40 Compute nodes
  - 2.6GHz Xeon Processors
  - 32 cores, 256 GB RAM / node
- 4 nodes have 2x NVIDIA K80 GPUS
- 4 x Storage Targets to provide POSIX volumes that add to the block and object storage
- Currently 500TB usable
- 50Gb/s Ethernet core, attached to SA DIRC
- 10Gb/s Access network connected to SANReN

Phase II : + 80 nodes + 2PB (Q4 2018) + 4 PB (Q4 2019)

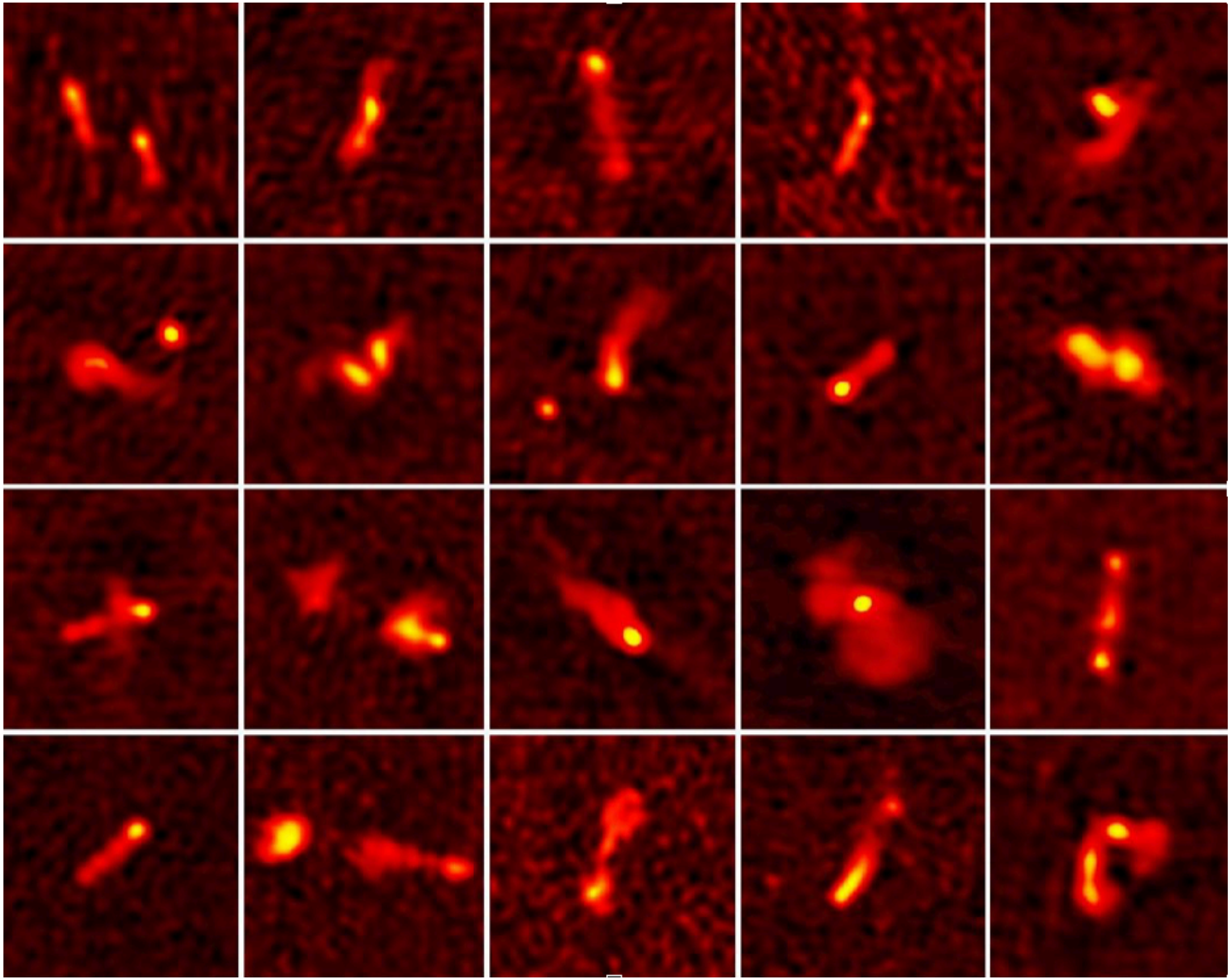
IDIA's Data Intensive Astronomy Research Cloud
User Collaboration in Processing and Analytics



UNIVERSITY of the WESTERN CAPE

IDIA

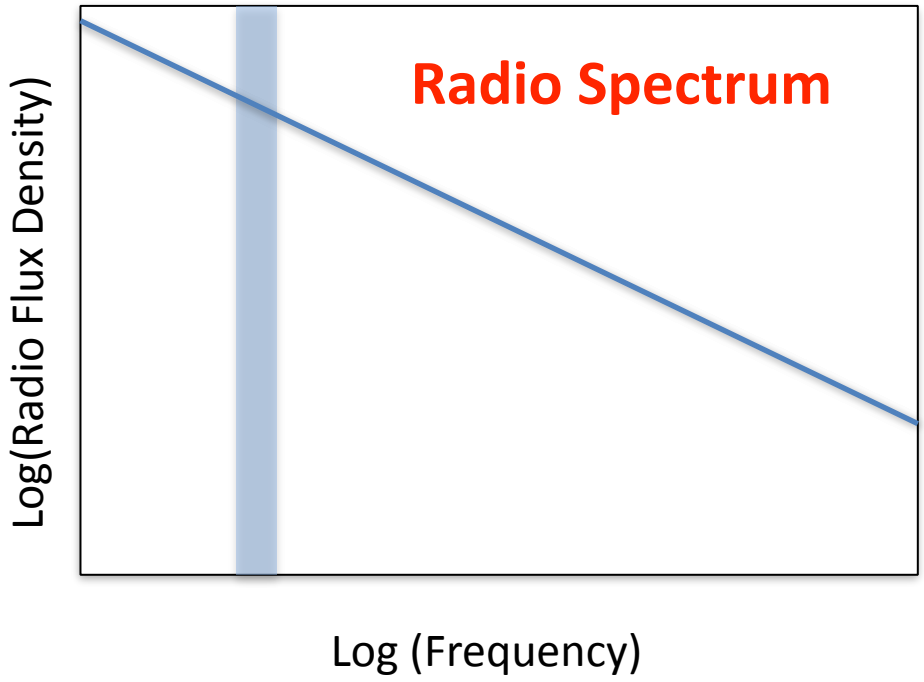
Radio Galaxies



VLSS - NRAO

But of course…

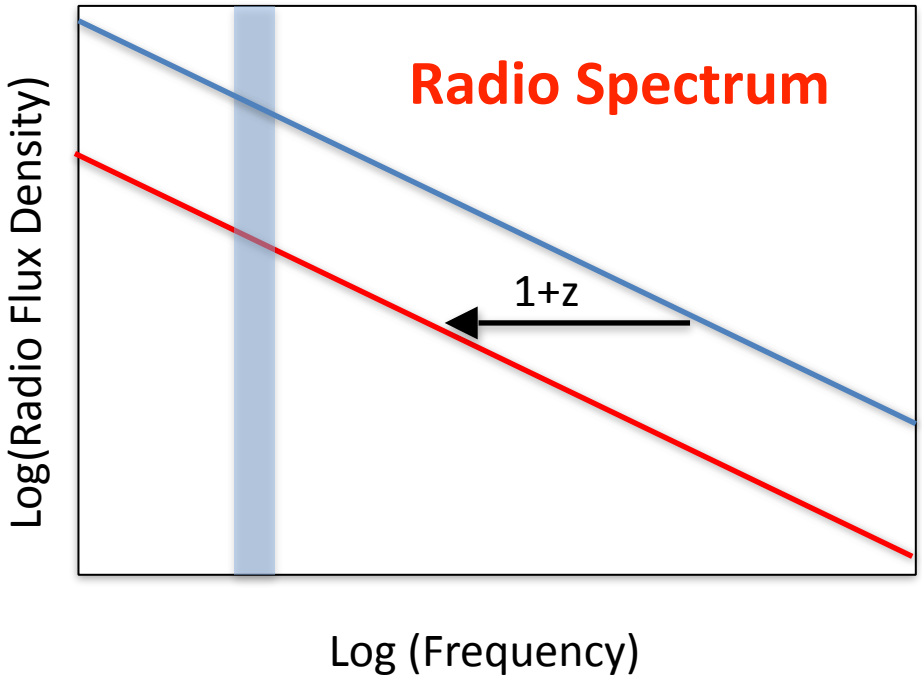
There's nothing as useless as a radio source (Jim Condon)



- Radio provides no (or very little?) redshift information
- Optical photometry provides much stronger constraints

But of course…

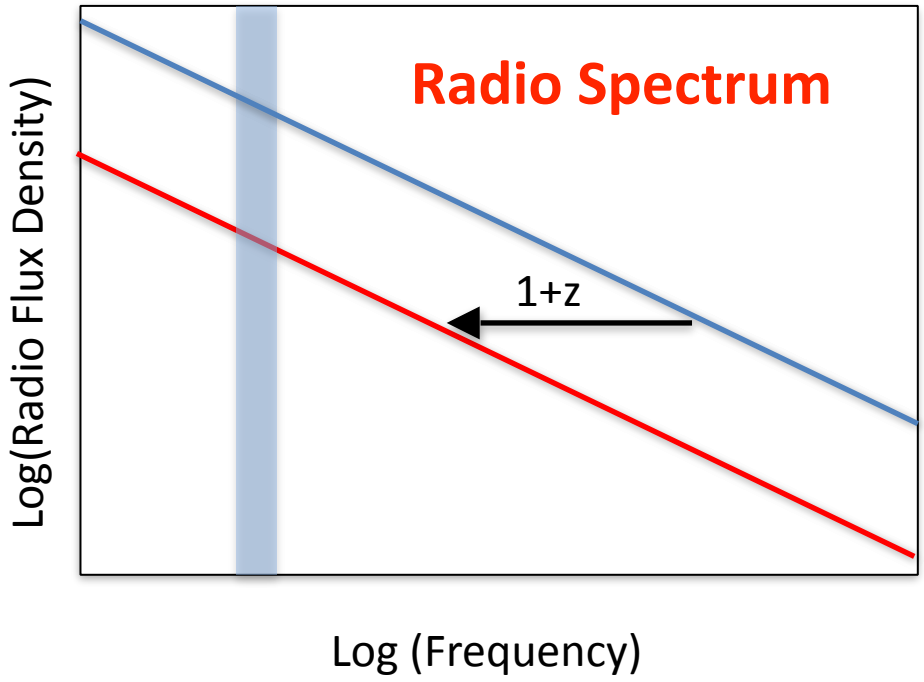
There's nothing as useless as a radio source (Jim Condon)

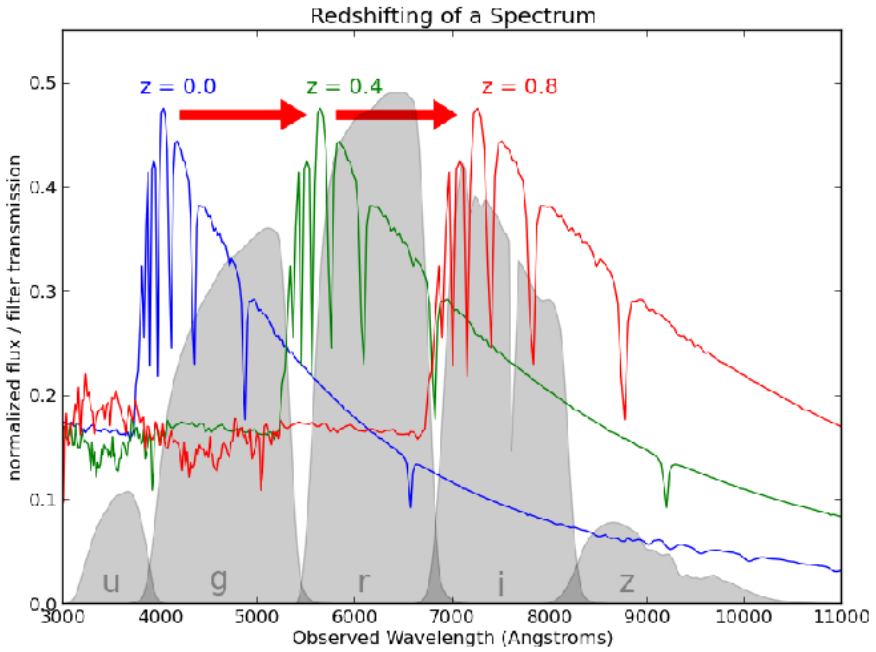


- Radio provides no (or very little?) redshift information
- Optical photometry provides much stronger constraints

But of course…

There's nothing as useless as a radio source (Jim Condon)





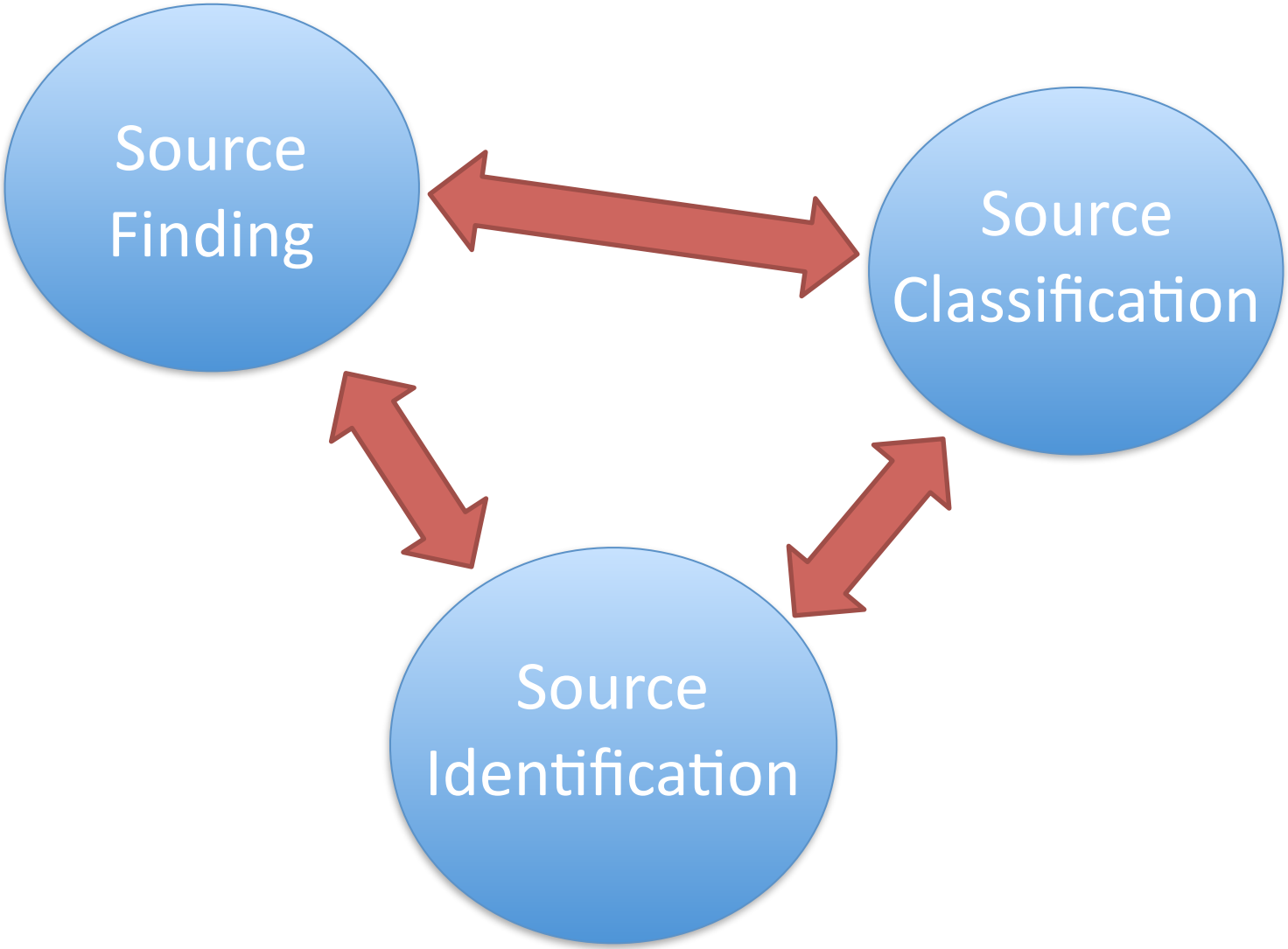
Optical Spectrum

- Radio provides no (or very little?) redshift information
- Optical photometry provides much stronger constraints

UNIVERSITY of the WESTERN CAPE

IDIA

Radio Source Characterisation



HELP Overview

- HELP = Herschel Extragalactic Legacy Project

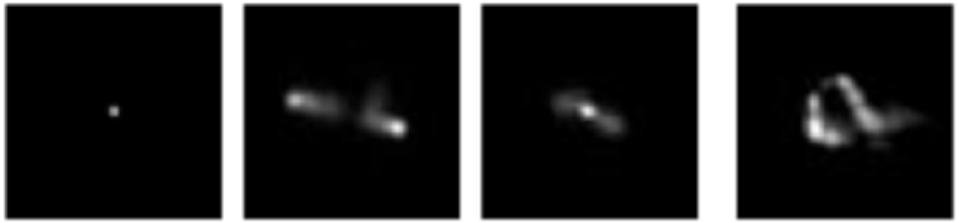
- European Commission project funded (2014-18) to:
  - Bring together multi-λ surveys over more than 1000 deg^2
  - Lower the barriers to multi-λ statistical survey science
  - Provide a resource for astronomers to study the high redshift Universe akin to SDSS (also) using Herschel
  - Provide tools to make Herschel surveys easy to use

Upcoming Data Release at http://hedam.lam.fr/HELP/

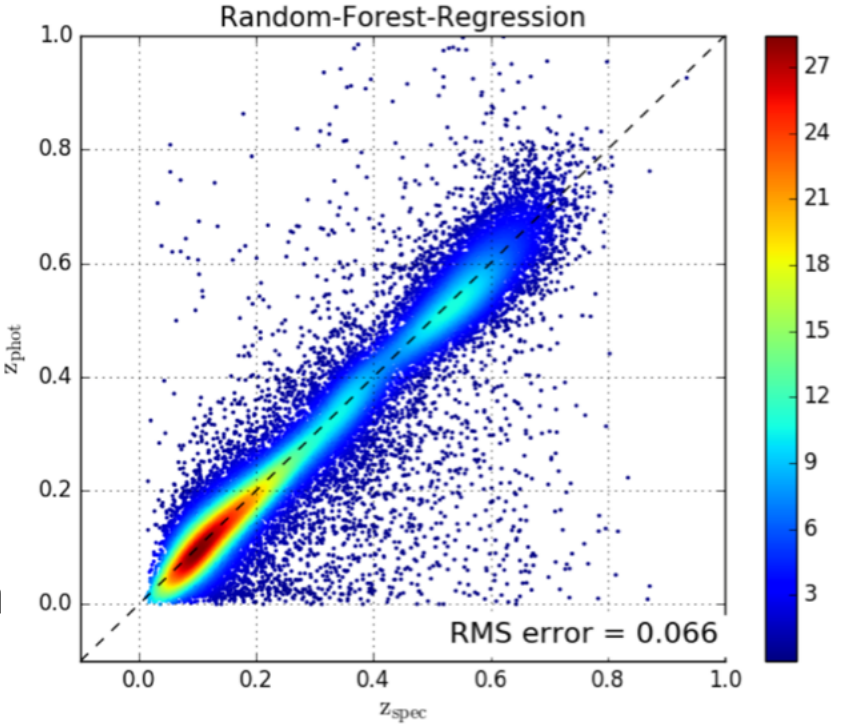
HIPPO : an IDIA Key Project

HIPPO : The HELP - IDIA Panchromatic PrOject

A Cloud-Based Environment for the Science Exploitation of Radio Surveys



Working with IDIA programmers to create a cloud-based environment where scientists can exploit MeerKAT in the context of multi-wavelength data



UNIVERSITY of the WESTERN CAPE

IDIA

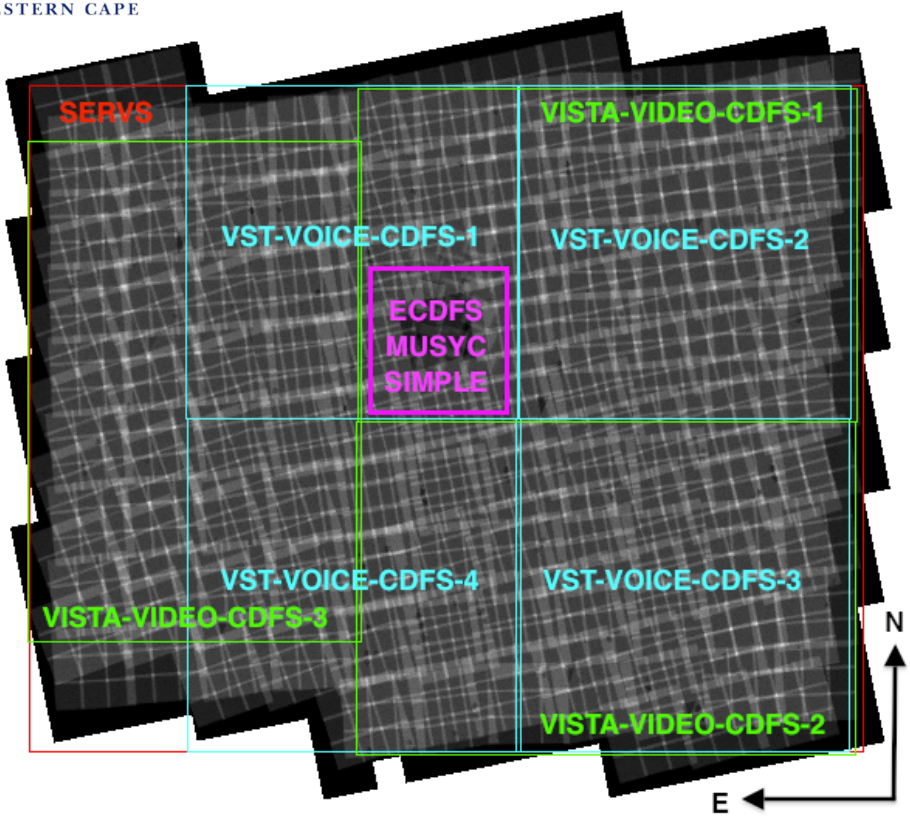
HIPPO's First Steps

- Create a python-centric 'Software Container' for Source Characterization on the IDIA Cloud.
- Assemble & Homogenize software tools to create cutouts and contours/overlays from most surveys
- Simple Visualization and Annotation Software
- Source Morphological Classification Software
- Source Spectro-Photometric Classification Software
- Extend Multi-Wavelength Ancillary Data (post-HELP)
- Scientific Exploitation of Existing Radio Surveys

‘Source Finding’ Container

- Originally aimed to Radio Source Detection
- Now aimed to (Radio) Source Characterization
  - Source Finding, Identification & Classification
  - Multi-Wavelength Catalogues & Images
    - Remote Database Queries & Local Storage
    - Mosaicing & Cross-Calibrating
    - Visualization (Contours, Overlays, Color Bars)
    - Photometric Redshifts and SED (Physical) Modeling
    - Machine Learning (SciKit, TensorFlow, Spark, etc)
    - Numerical Simulations?
    - Data Cubes?

VST Optical Imaging of MeerKAT Fields



VOICE (VST INAF GT Survey)

CDFS - 4 deg² - m_{AB}~26 in ugri

http://www.mattiavaccari.net/voice/

(PIs : Covone & Vaccari)

1) image/site quality

2) u-band sensitivity

3) multi-wavelength

Vaccari et al. 2016

- gri Multi-Epoch for SN Search & AGN Variability Studies
  (Cappellaro+ 2015, De Cicco+ 2015, Falocco et al. 2015, Botticella+ 2017)
- ugri Deep Stacks to be combined with ZYJHK & IRAC12
- Enabled improved selection / targeting for spec-z runs

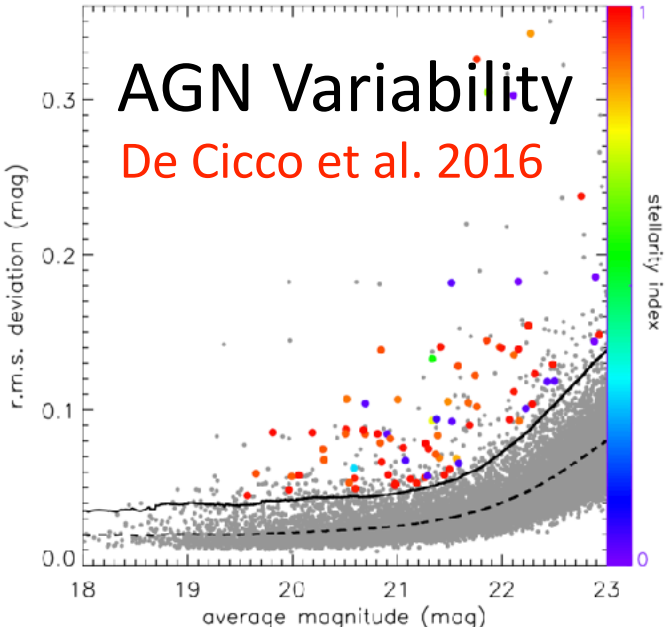
Time-Domain Astronomy

UNIVERSITY of the WESTERN CAPE

IDIA

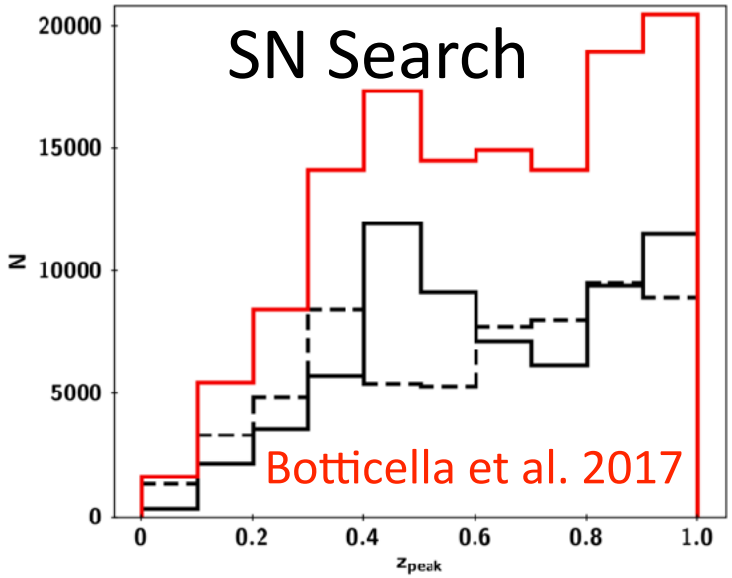
AGN Variability

De Cicco et al. 2016



SN Search

Botticella et al. 2017



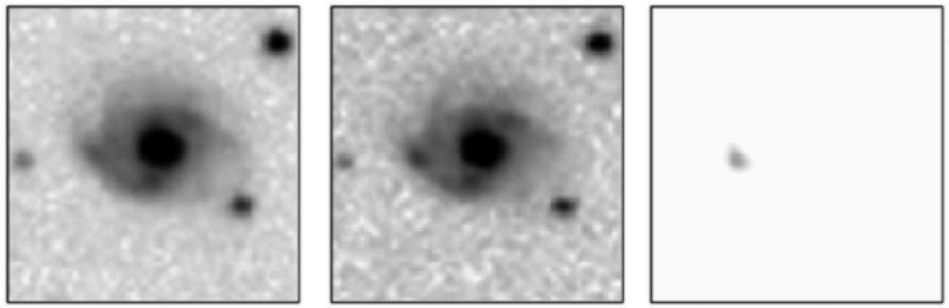


Figure 2. Examples of the reference (left) and science (centre) images. The image on the right is the ground-truth output defined for this image pair. It contains the image of a single transient, completely devoid of background and noise. The profile of the transient is the best match to reality our model can produce.

Sedaghat & Mahabal 2018

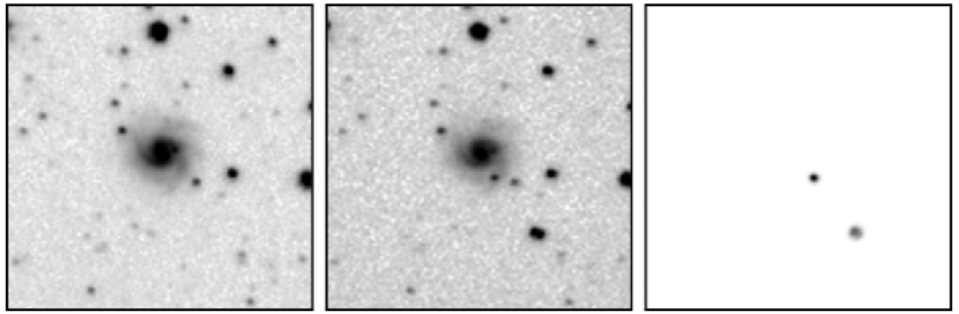
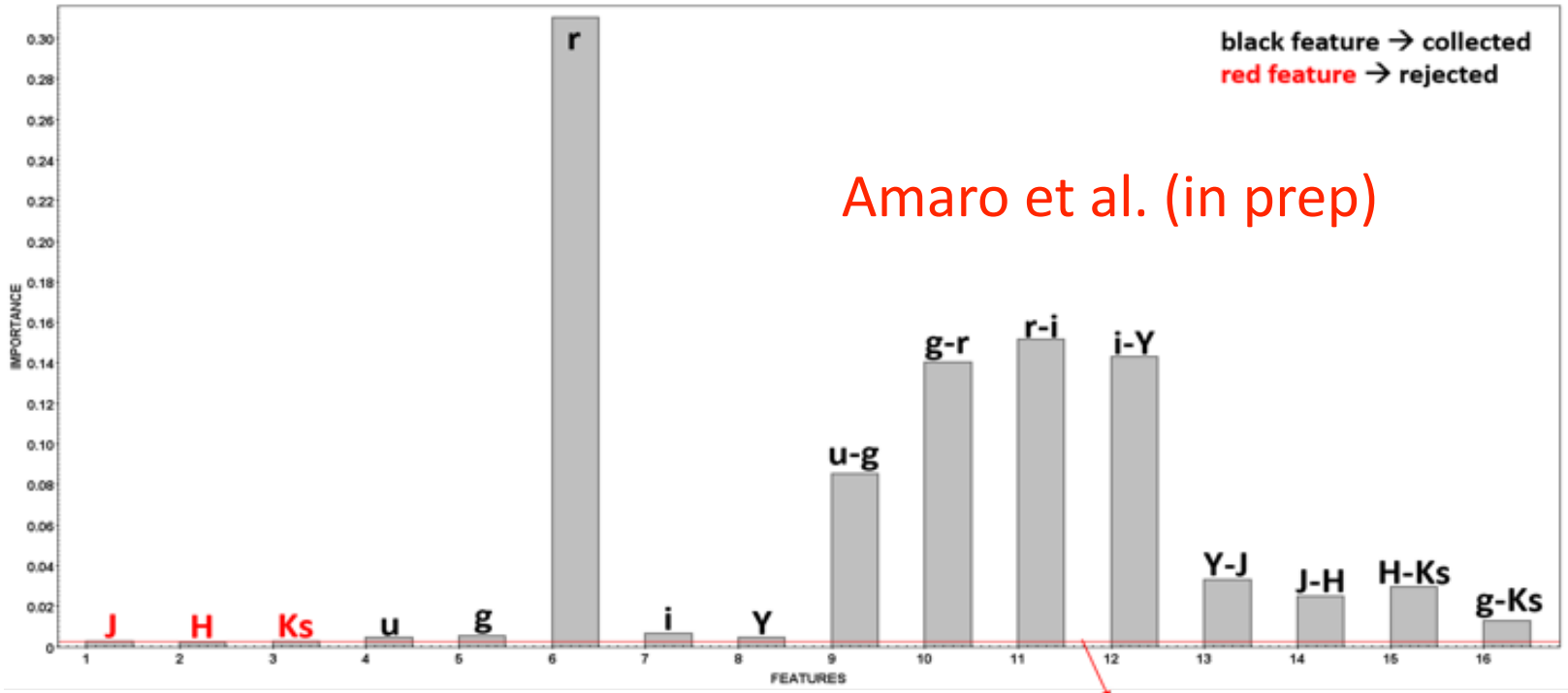


Figure 8. An exemplar multi-transient case from the CRTS SN Hunt data set. The science image (middle) has two transients and the network prediction (right) finds them both, though it was never trained explicitly to look for multiple transients.

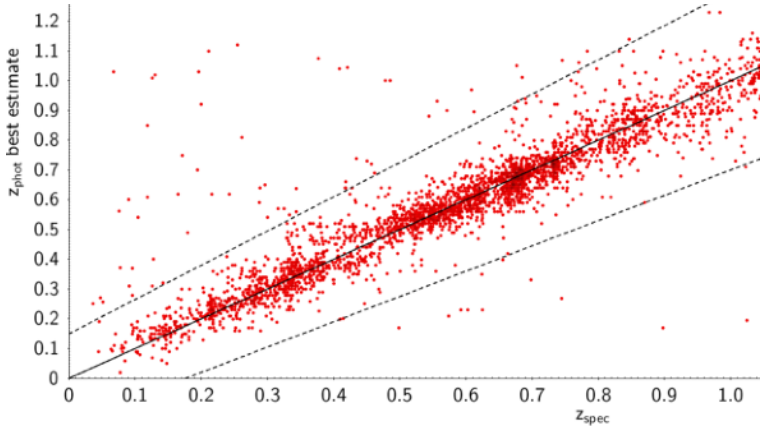
Testing ground for ML Work?

ML Photometric Redshifts

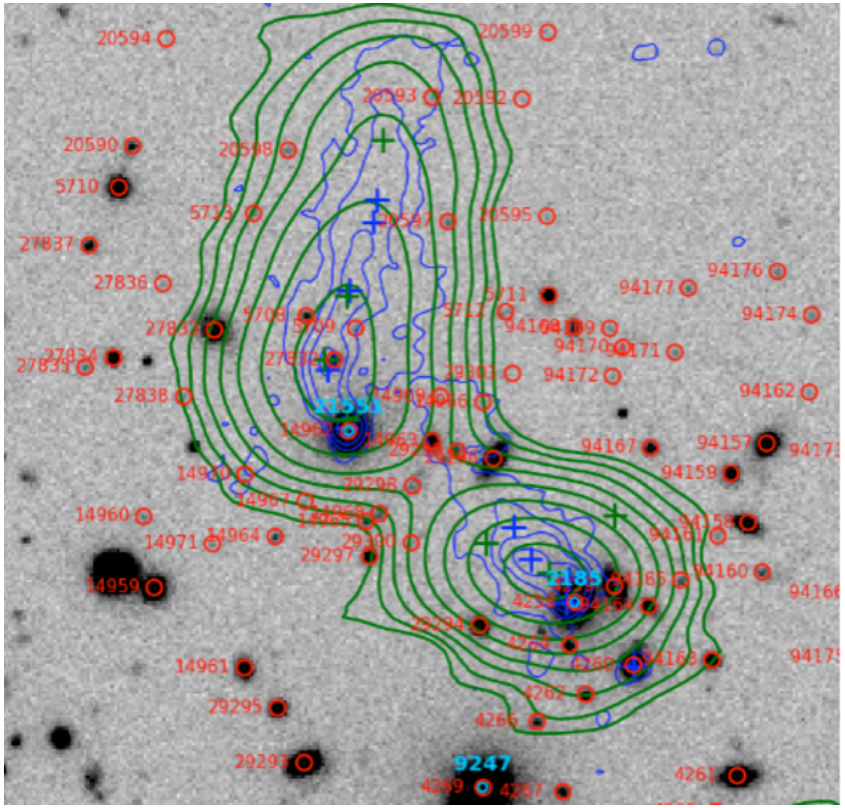


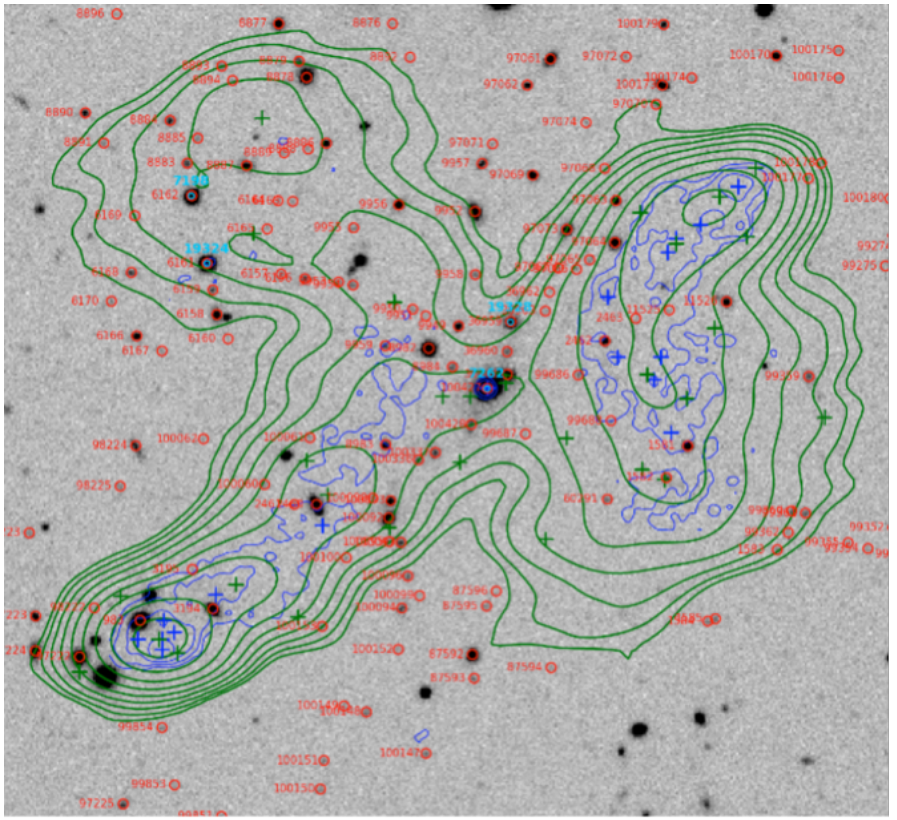
PHiLAB all-relevant feature selection on VOICE KB2

Estimator	Exp #3 no FS	Exp #4 FS
|bias|	0.005	0.004
NMAD	0.028	0.027
% out>0.15	4.32%	3.87%



Radio XID & Classification Tools

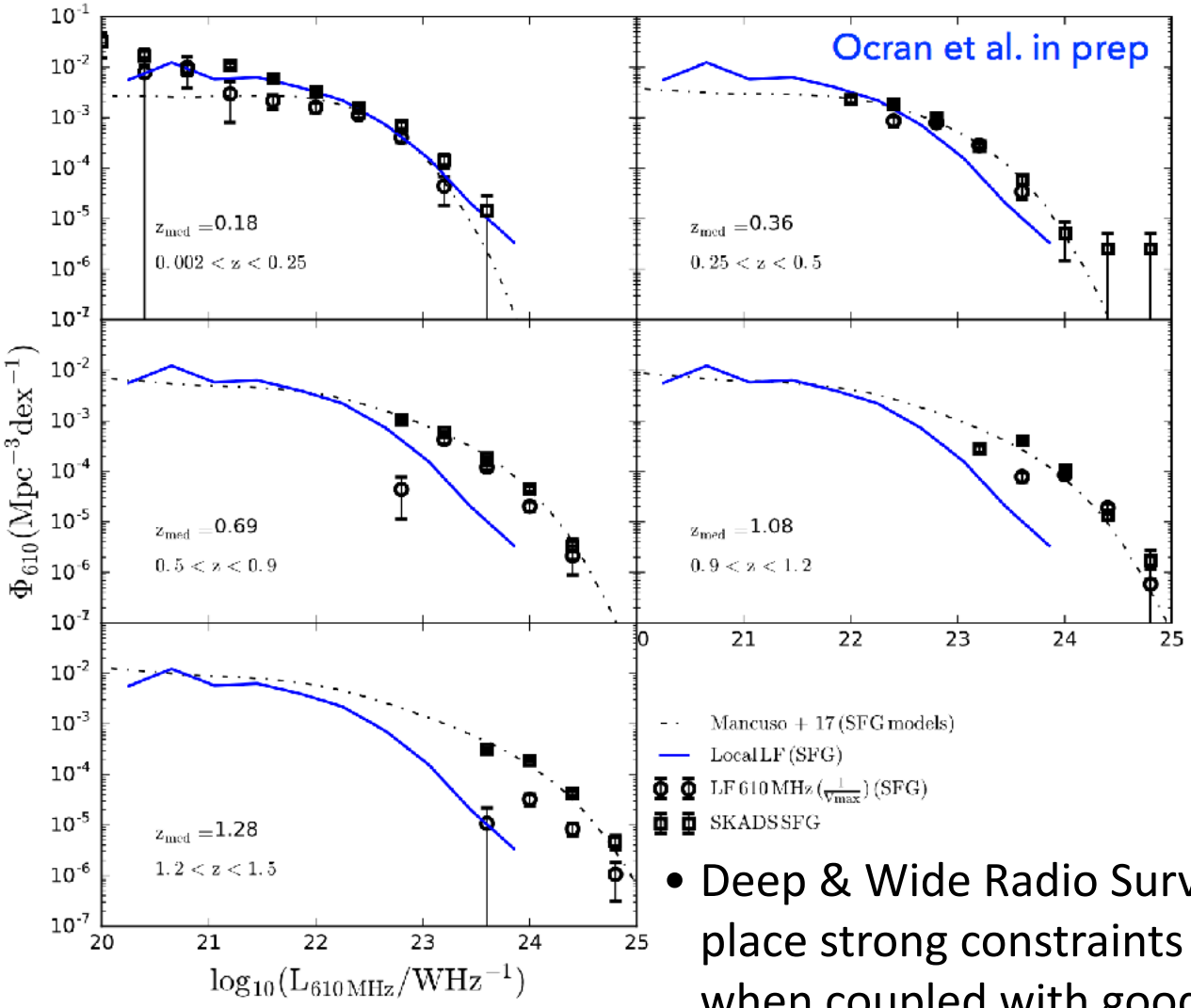


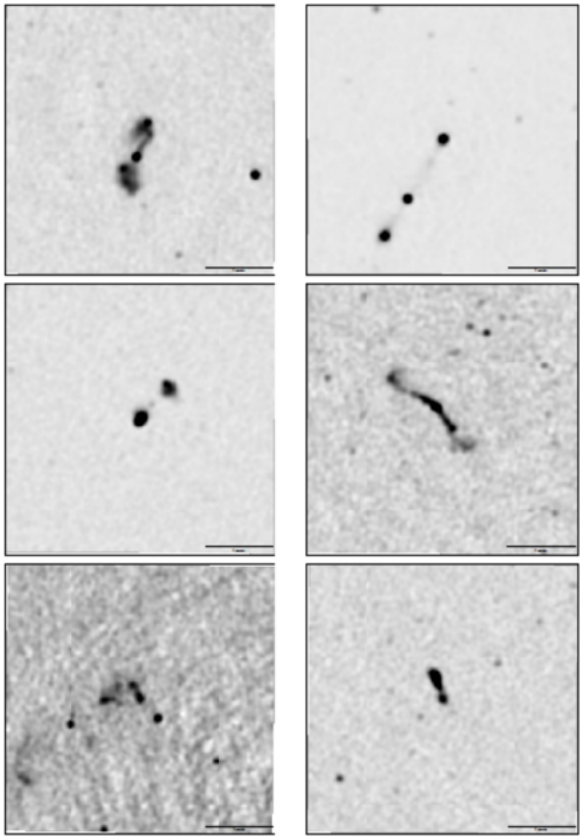


Prescott et al. 2018

- Heywood et al. 2016 - JVLA BnC Survey of SDSS Stripe 82
- Several Astronomers to annotate Radio Contour Plots
- Checking Nearest-Neighbour / Likelihood Ratio Matches

The SFG Radio LF beyond z~1





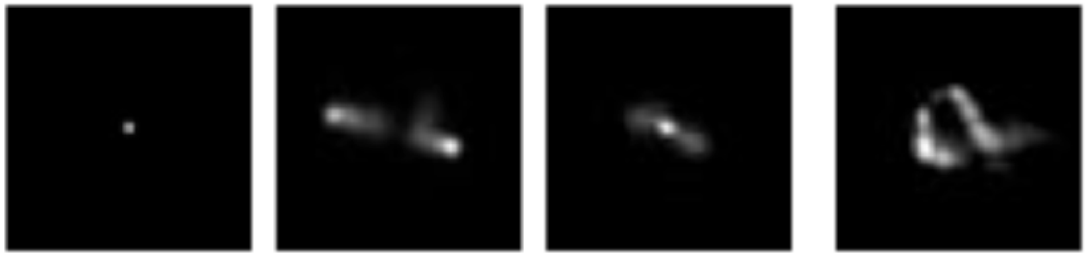
- Deep & Wide Radio Surveys (GMRT 610 MHz here) place strong constraints on galaxy evolution models when coupled with good optical spectrophotometry

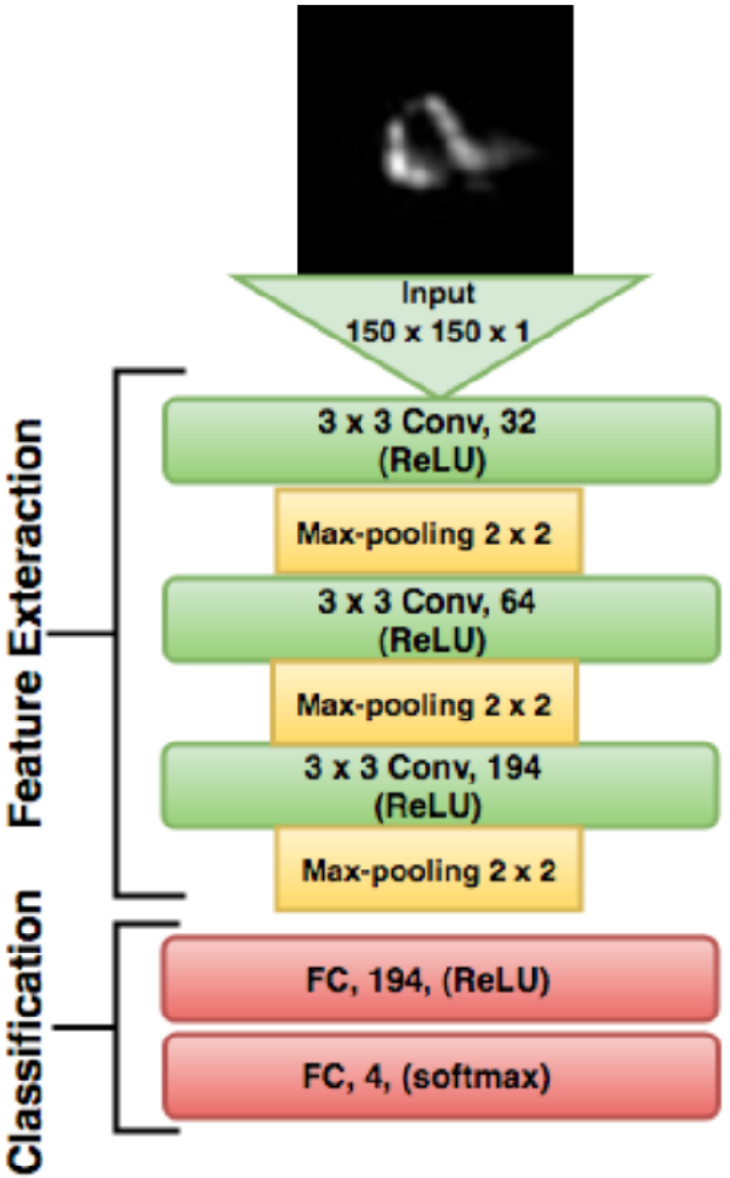
Ocran, Taylor, Vaccari, Mancuso, Prandoni et al. (in prep)

Radio Source Classification

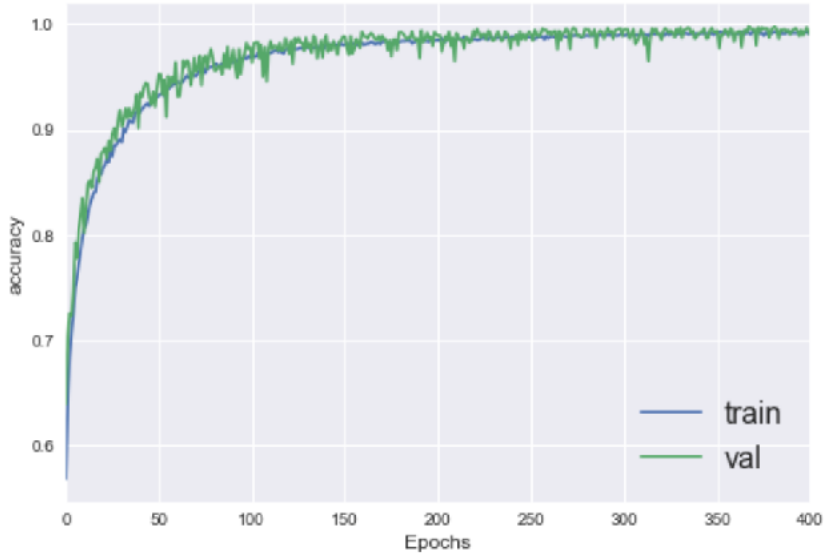
Alhassan, Taylor & Vaccari 2018
https://github.com/wathela/FIRST-CLASSIFIER

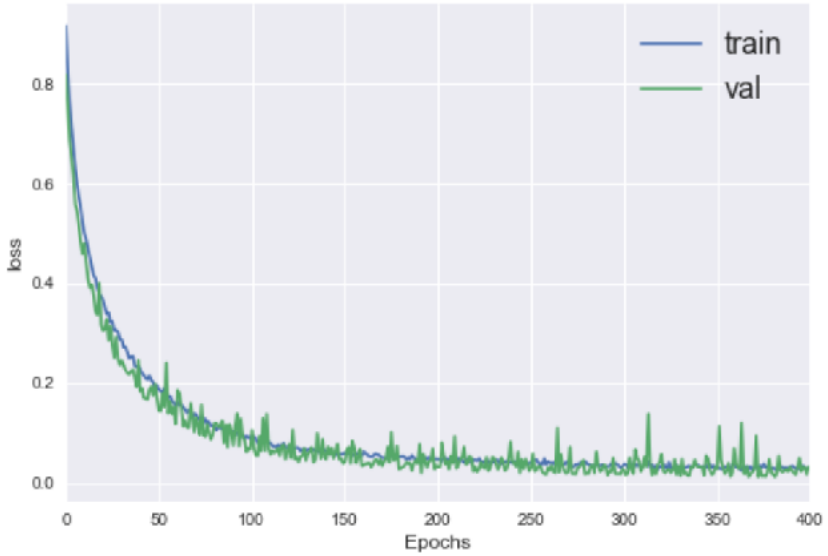
Type	Original Sample
COMP	121
FRI	201
FRII	338
BENT	177
Total	837

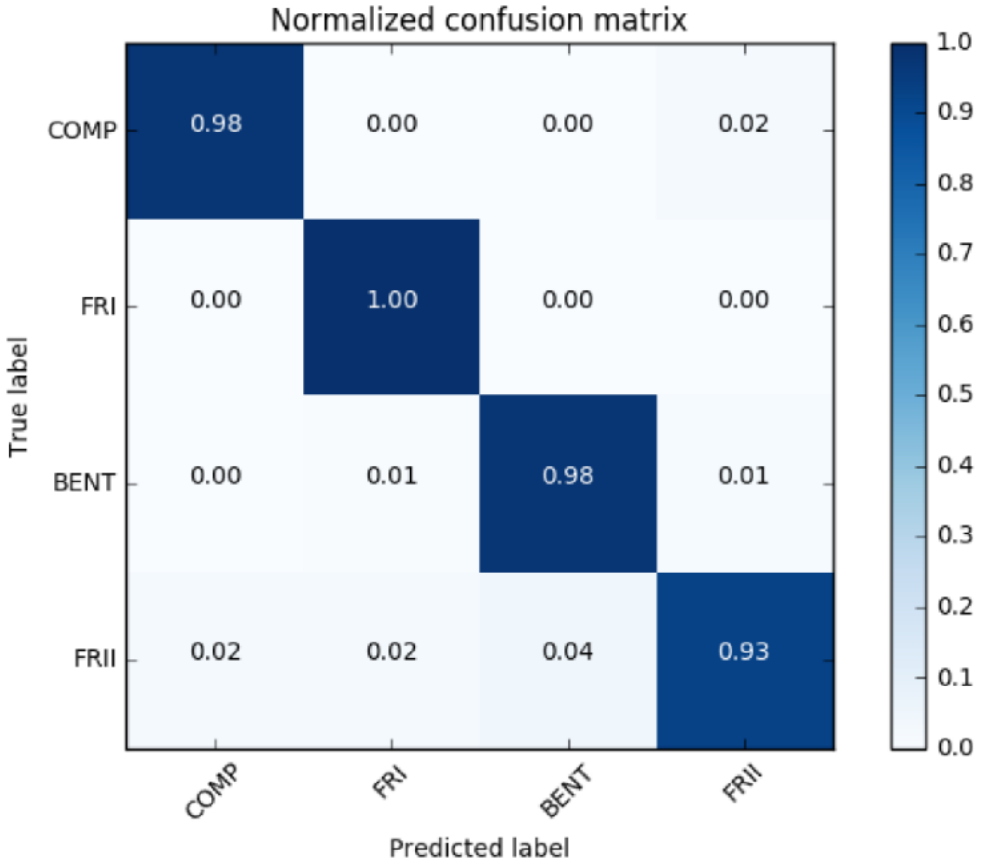




DL for Radio Source Classification



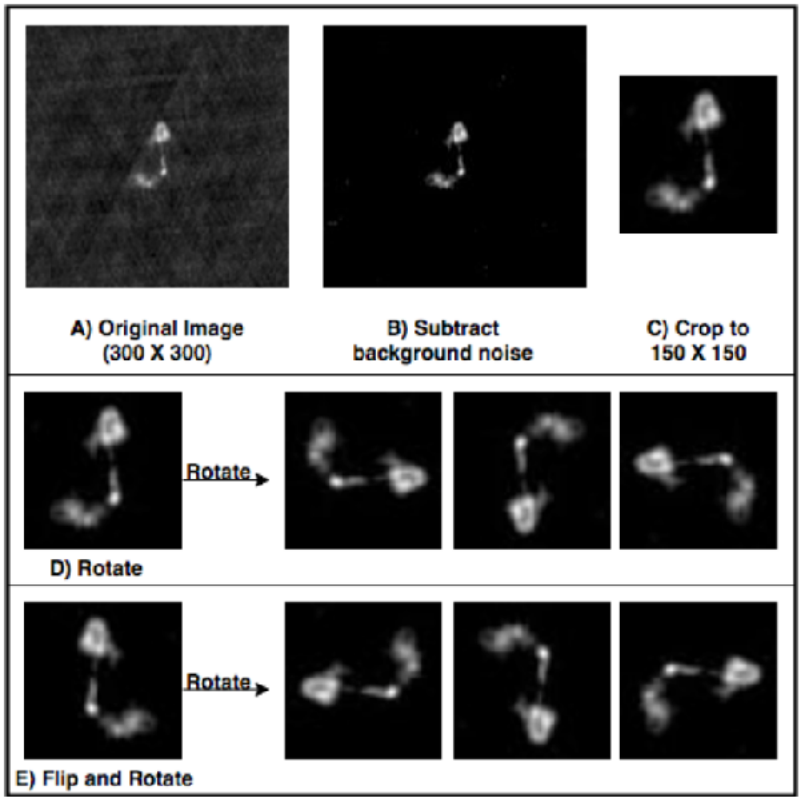


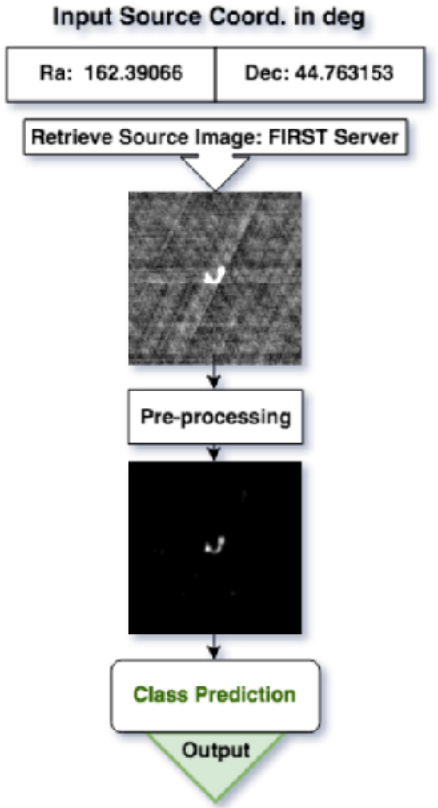


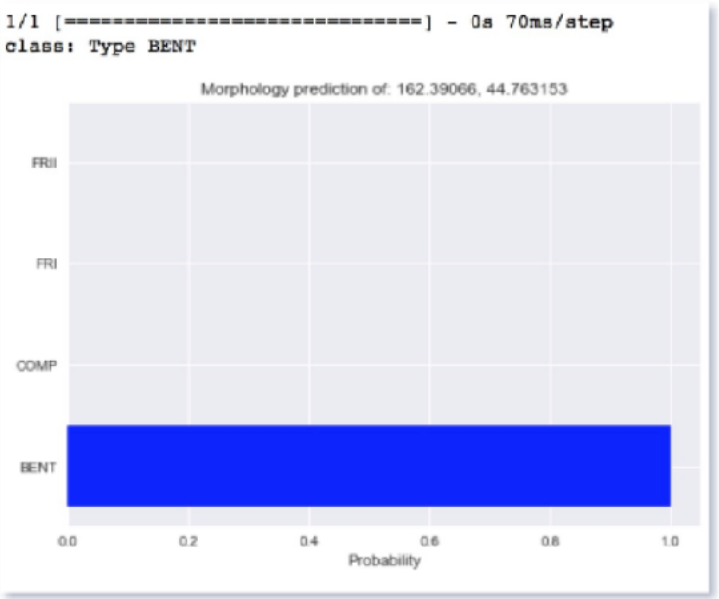
Type	precision	recall	f1-score
COMP	0.98	0.98	0.98
FRI	0.98	1.00	0.99
BENT	0.96	0.98	0.97
FRII	0.96	0.93	0.95
avg/ total	0.97	0.97	0.97

What's Next?

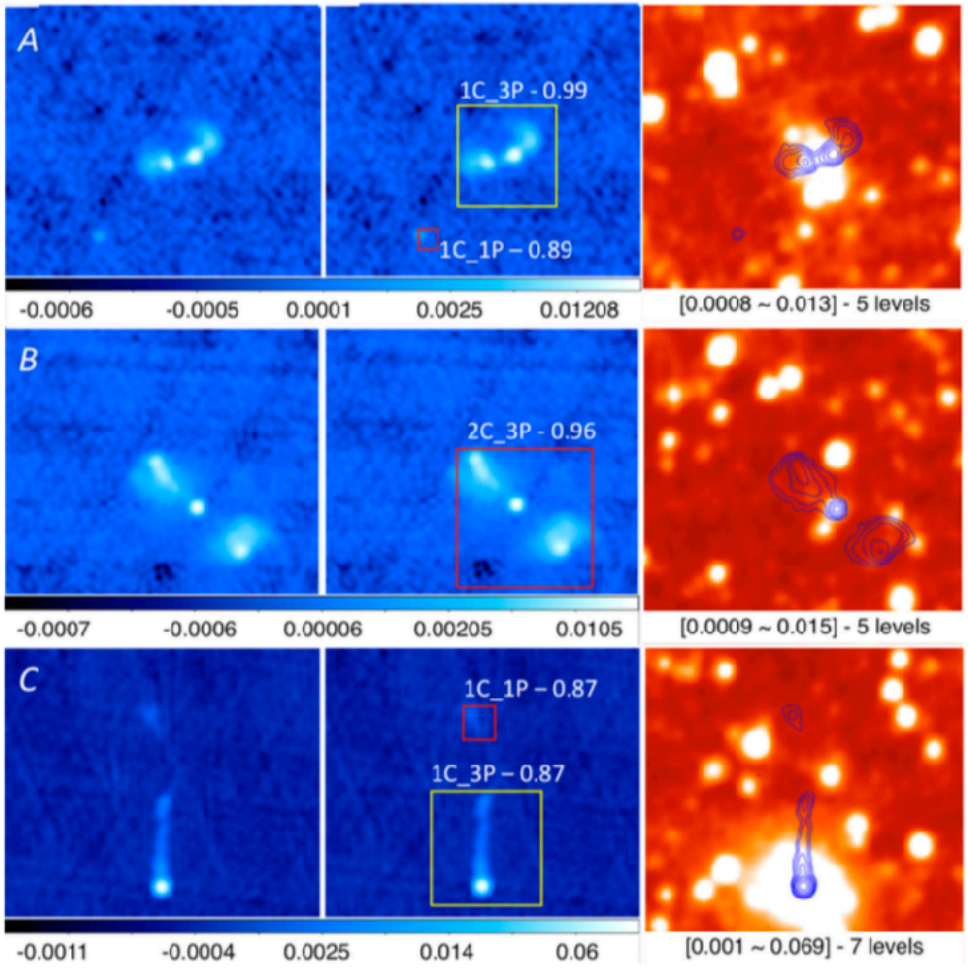
- Refine Classification Scheme
- Evaluate Classification PDF
- Incorporate Optical/Infrared Imaging
- Participate in Radio Galaxy Zoo 2





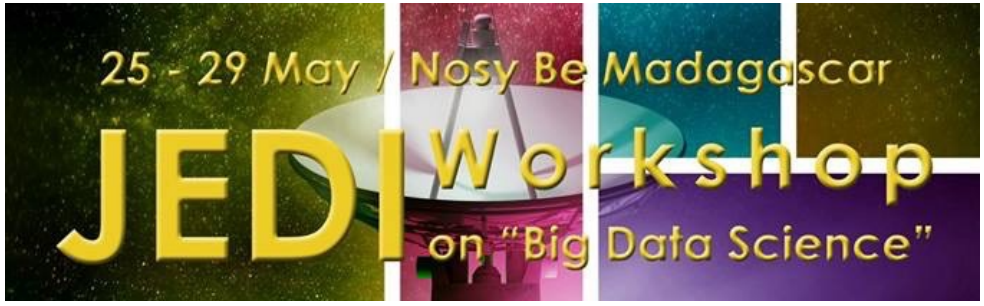


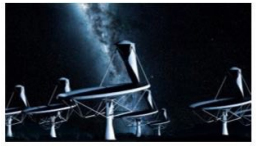
Faster Region-Based CNNs with ClaRAN for the Identification of Radio Sources



Wu et al. 2019 - https://github.com/chenwuperth/rgz_rcnn/
Mofokeng et al. (in prep) - Applied to GMRT/MeerKAT

Big Data Science Training





2nd Big Data Africa School

Cape Town, 10 – 17 September 2018

The Big Data Africa School aims to introduce fundamental data science tools & techniques to talented young science graduates across a range of disciplines, who have an interest to develop their skills and knowledge in working efficiently on extremely large datasets in any research environment.

A UK-South Africa Newton Fund initiative

NRF National Research Foundation | SARAO South African Radio Astronomy Observatory

DARA BIG DATA

IDIA Inter-University Institute for Data Intensive Astronomy — *from big data to big ideas*

SKA AFRICA SQUARE KILOMETRE ARRAY

DARA DEVELOPMENT IN AFRICA WITH RADIO ASTRONOMY

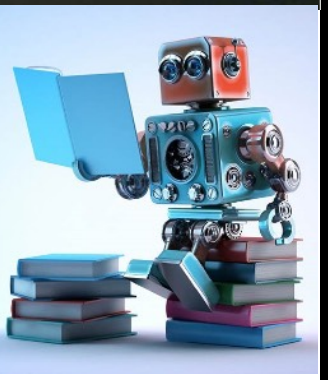
Newton Fund

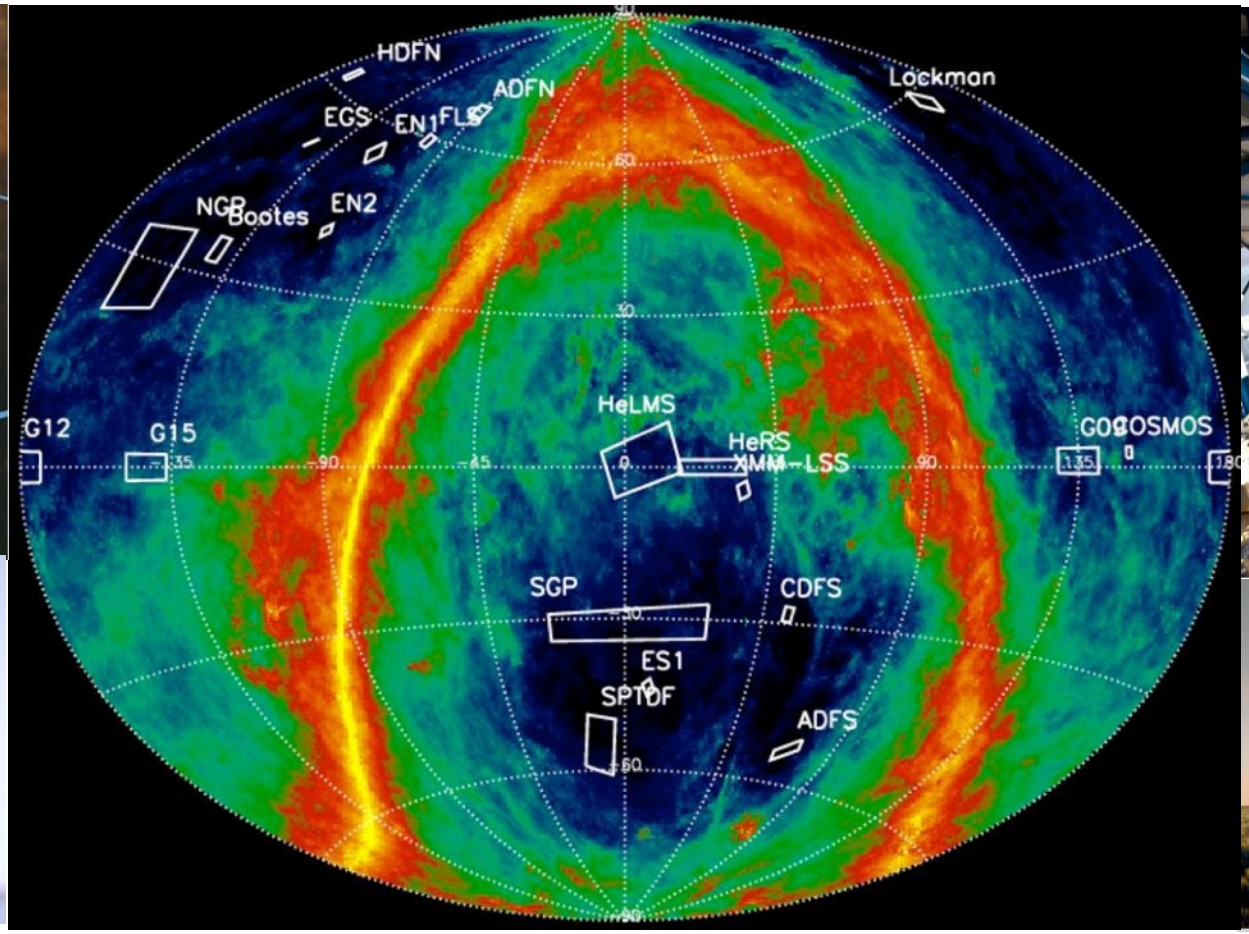
Get Involved!

UNIVERSITY of the WESTERN CAPE

IDIA

Thanks!





http://www.idia.ac.za

http://www.mattiavaccari.net