

A background image of a galaxy map, showing various galaxies in shades of blue and purple against a dark background.

# **The *Herschel* Extragalactic Legacy Project (HELP) & Machine learning classifications in remote sensing**

**Steven Duivenvoorden**



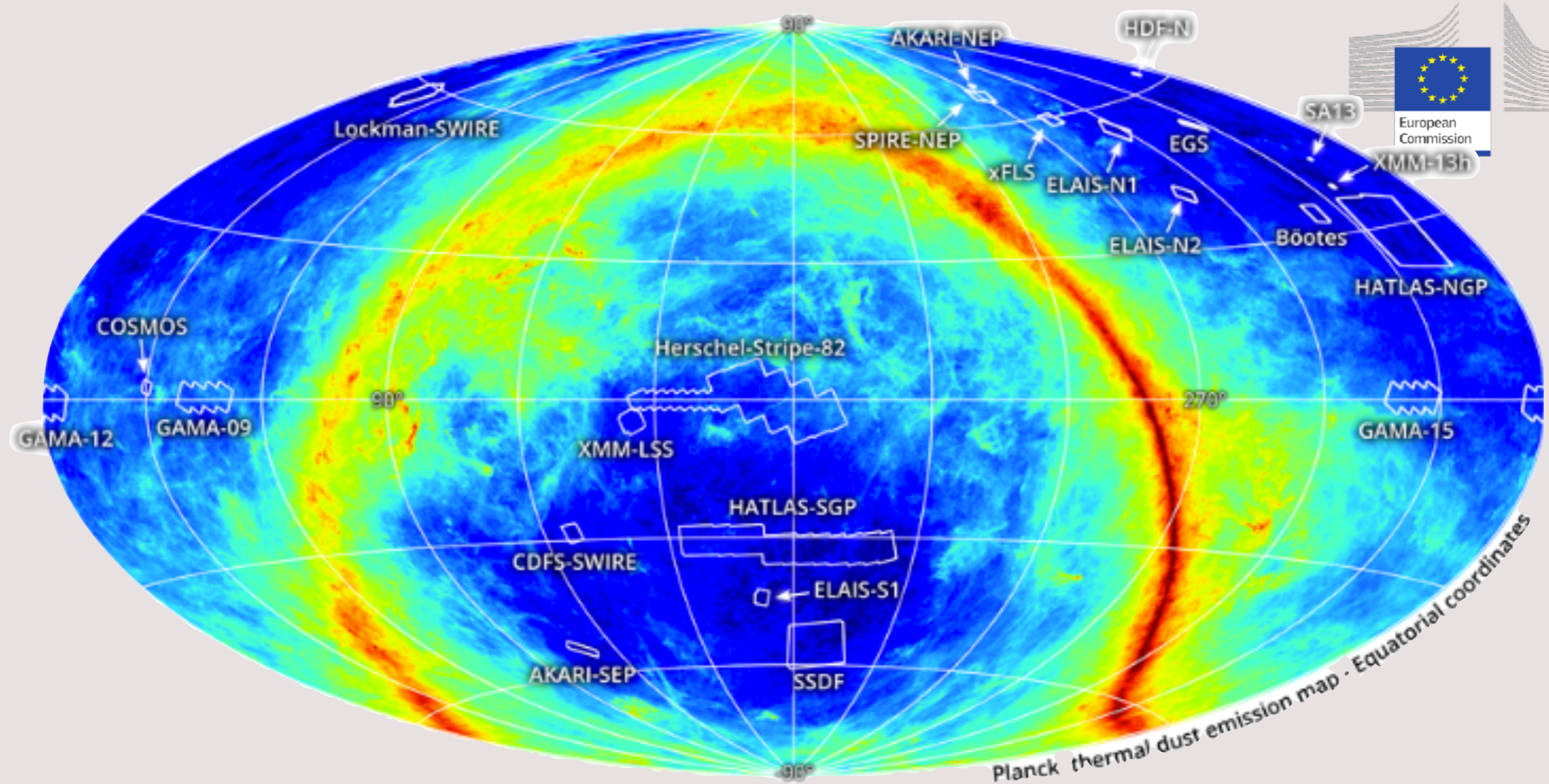


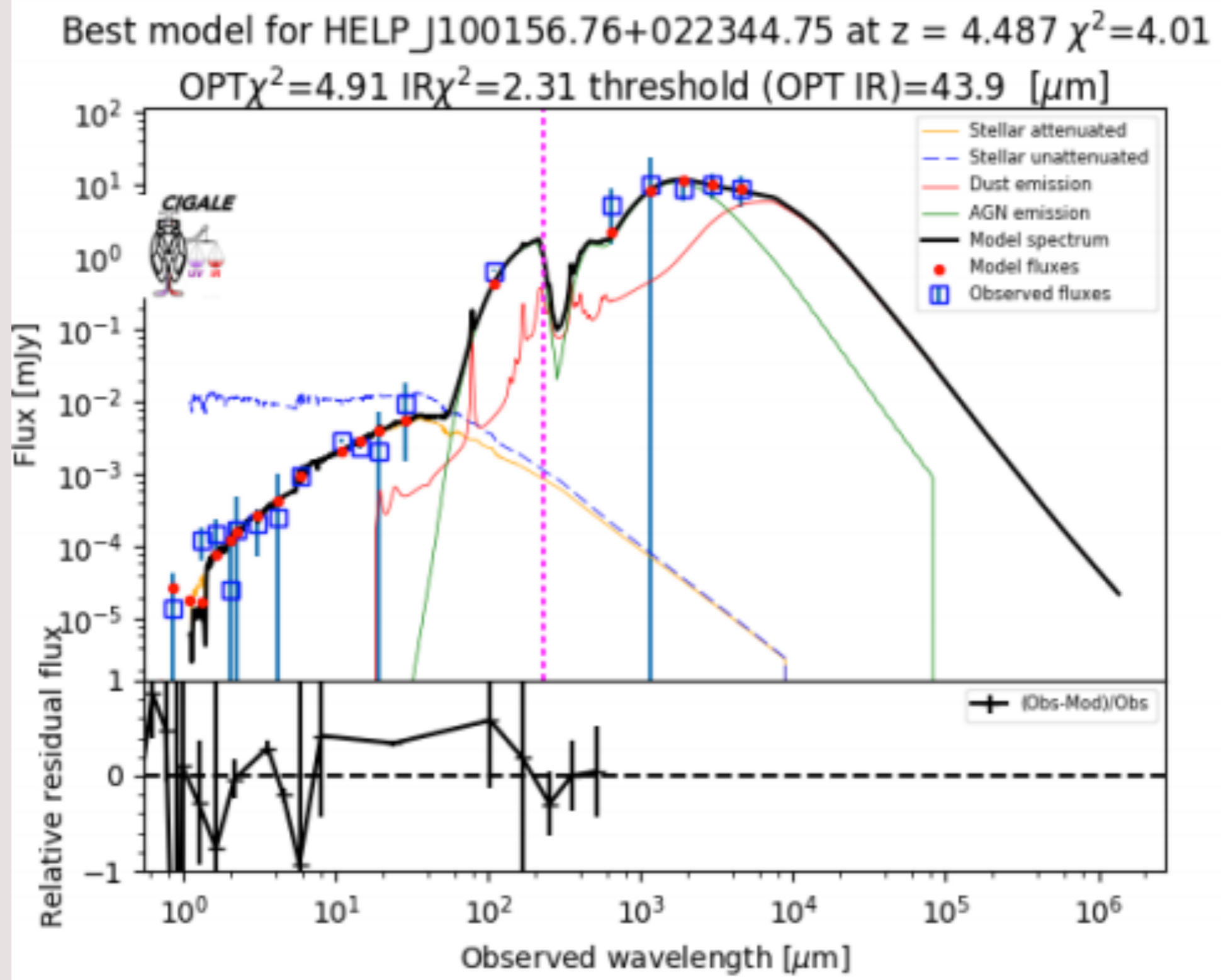
US

UNIVERSITY OF SUSSEX

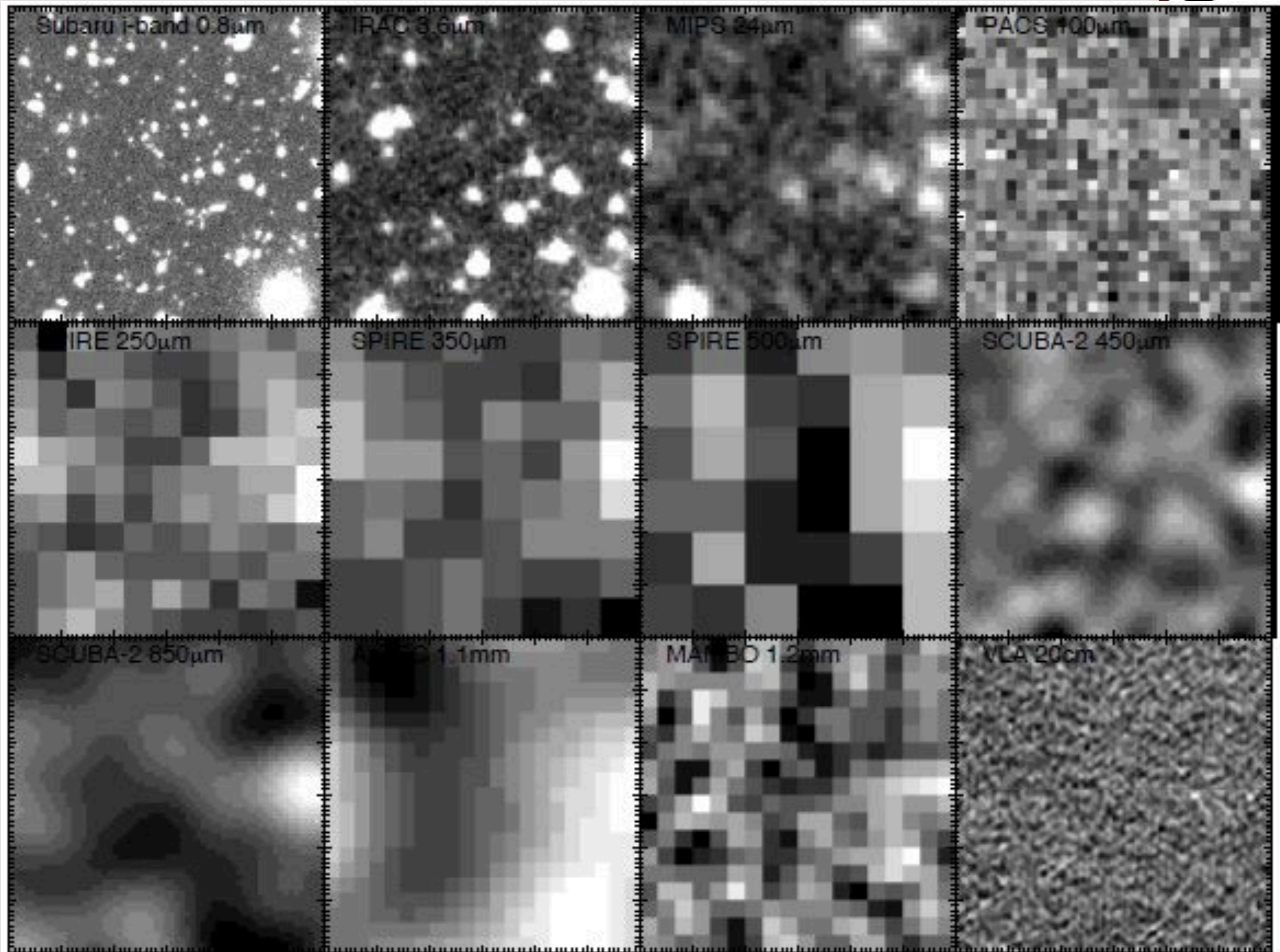
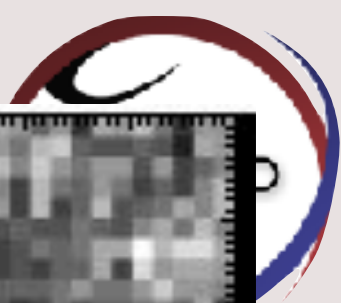
# HELP fields

- [herschel.sussex.ac.uk](http://herschel.sussex.ac.uk)

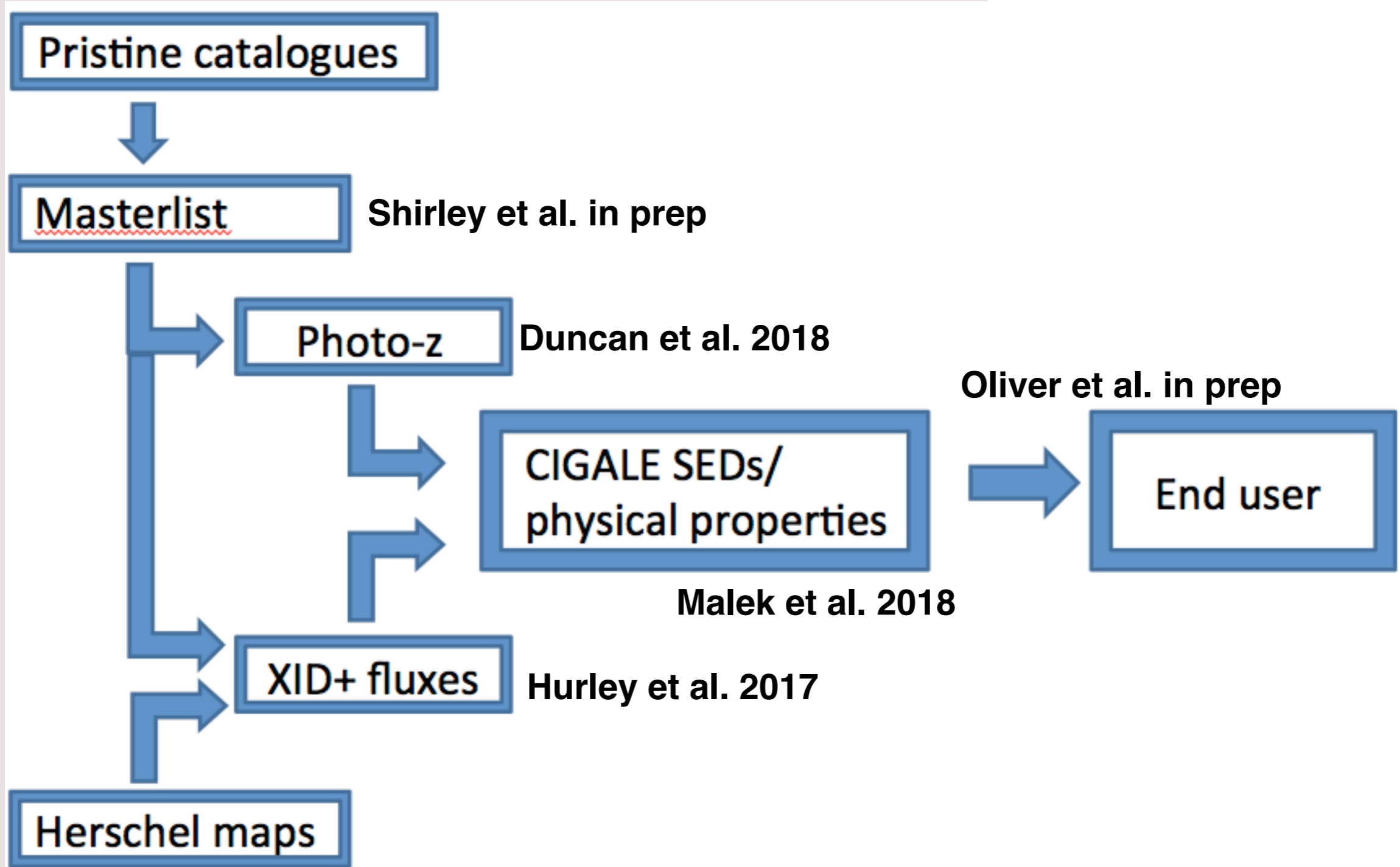








TY  
X



- [herschel.sussex.ac.uk](http://herschel.sussex.ac.uk)
- 60 Surveys
- 500 Gb catalogues
- [github.com/H-E-L-P/dmu\\_products](https://github.com/H-E-L-P/dmu_products)
- 171 570 436 Objects!
- [hedam.lam.fr/HELP](http://hedam.lam.fr/HELP)



DMU#	Responsibility	Link
0	Pristine catalogues	<a href="#">dmu0</a>
1	Masterlist data	<a href="#">dmu1</a>
2	Field definitions	<a href="#">dmu2</a>
3	Morphologies (Shapes & Sizes) of Objects	<a href="#">dmu3</a>
4	Bright Star Mask	<a href="#">dmu4</a>
5	Known Star Flag	<a href="#">dmu5</a>
6	Optical photometry validation	<a href="#">dmu6</a>
7	Optical photometry	<a href="#">dmu7</a>
8	Radio data - LOFAR & FIRST/NVSS/TGSS	<a href="#">dmu8</a>
9	Radio data - JVLA-DEEP & GMRT-DEEP	<a href="#">dmu9</a>
10	Data Fusion	<a href="#">dmu10</a>
11	Cross matching MIPS/PACS/SPIRE	<a href="#">dmu11</a>

H-E-L-P / [dmu\\_products](#)

<> Code    ⓘ Issues 190    🔗 Pull requests 0    📁 Projects 1    📖 Wiki

Branch: master ▾    [dmu\\_products](#) / [dmu1](#) /

Yannick Roehly Replace vohedamtest URLs to herschel-vos.phys.sussex.ac.uk

..

<a href="#">dmu1_ml_AKARI-NEP</a>	Correct cross-matching radius phrase
<a href="#">dmu1_ml_AKARI-SEP</a>	Correct cross-matching radius phrase
<a href="#">dmu1_ml_Bootes</a>	Correct cross-matching radius phrase
<a href="#">dmu1_ml_CDFS-SWIRE</a>	Replace vohedamtest URLs to herschel-vos.ph
<a href="#">dmu1_ml_COSMOS</a>	Correct cross-matching radius phrase
<a href="#">dmu1_ml_EGS</a>	Correct cross-matching radius phrase
<a href="#">dmu1_ml_ELAIS-N1</a>	Correct cross-matching radius phrase
<a href="#">dmu1_ml_ELAIS-N2</a>	Correct cross-matching radius phrase
<a href="#">dmu1_ml_ELAIS-S1</a>	Correct cross-matching radius phrase
<a href="#">dmu1_ml_GAMA-09</a>	Correct cross-matching radius phrase
<a href="#">dmu1_ml_GAMA-12</a>	Correct cross-matching radius phrase



# Virtual Observatory at susseX (VOX)

- Allows complex queries to leverage the richness of the data.
- VO standards

<https://herschel-vos.phys.sussex.ac.uk/>



## The Virtual Observatory at susseX (VOX)

Welcome to the Virtual Observatory at susseX (VOX). This is currently serving data from the Herschel Extragalactic Legacy Project (HELP) but will be extended to a range of astronomical datasets produced at Sussex and collaborating institutions. You can find out more about HELP [here](#).

VOX is often the quickest way to access HELP data. However you can also browse the raw data [here](#). This includes large table files and images including some that is not served on VOX. You can read about the structure of the database on [GitHub](#).

In addition to the services listed below, on this site you can access [numerous tables](#) using [TAP](#) or [form-based ADQL](#).

Please check out our [site help](#).

### Services available here

By Title    By Subject

H...

- [HELP main catalogue](#) [i](#) [q](#)  
HELP master catalogue.
- [HELP sources best fitting spectra from CIGALE](#) [i](#) [q](#)  
Best fitting spectra found by CIGALE for each HELP source
- [HELP spectroscopic redshift catalogue](#) [i](#) [q](#)  
Compendium of spectroscopic redshifts.
- [HerMES SCAT250 \(blind SPIRE 250µm extraction\)](#) [i](#) [q](#)  
SPIRE catalogues from the Herschel Extragalactic Legacy Project.
- [HerMES SCAT350 \(blind SPIRE 350µm extraction\)](#) [i](#) [q](#)  
SPIRE catalogues from the Herschel Extragalactic Legacy Project.
- [HerMES SCAT500 \(blind SPIRE 500µm extraction\)](#) [i](#) [q](#)  
SPIRE catalogues from the Herschel Extragalactic Legacy Project.
- [HerMES xID250 \(SPIRE fluxes at SPIRE 250µm positions\)](#) [i](#) [q](#)  
SPIRE catalogues from the Herschel Extragalactic Legacy Project.
- [Herschel map cutouts](#) [i](#) [q](#)  
Image maps from the Herschel Space Observatory.
- [Herschel maps](#) [i](#) [q](#)  
Image maps from the Herschel Space Observatory.

# VO server – data access

- Through the web interface
- Programmatically, e.g. in Python with pyVO.
- Through VO protocols e.g. with Topcat.



```
In [19]: #How do we establish the VO connection to our database
service = vo.dcs.TAPService("https://herchel.vo.phys.sussex.ac.uk/_system_/tap/vox/tap")

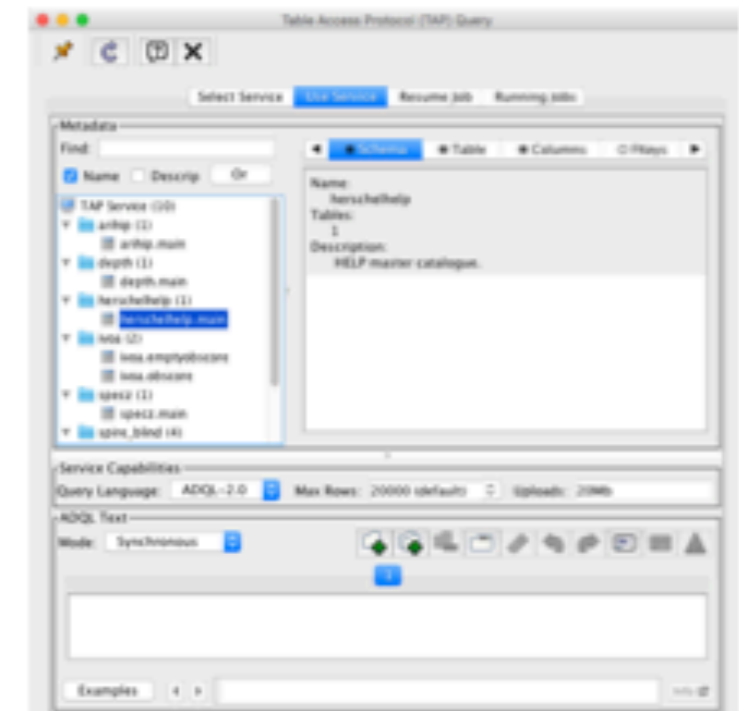
In [20]: #How do we execute the query
resultset = service.search(example_query)

WARNING: W2: Rows(1:17): W2: 00000 deprecated in MySQL 5.7 (warning-in-mysql57-row)
WARNING: W6: Rows(1:17): W6: 1ms114 DCS 'phys_dept.phys_sussex.ac.uk' follows word 'phys.de'
[warning-in-mysql57-row]

In [21]: resultset.table

Out[21]: Table metadata: length=85172
```

ra	dec	redshift	rigids_abundance
deg	deg		W
float64	float64	float64	float64
246.50262013007	55.00001000000001	4.070001	--
246.50474700007	54.99700010000001	4.870000	--
246.50700000007	54.99700140000000	4.230000	--
246.5074000100000	54.99200000000000	4.8771	--
246.50744000007	54.99800100000000	4.0401	--
246.5078400000701	55.00000000000001	5.040000	--
246.51104000100701	55.00040000000000	5.000000	--
246.513800100007	55.00041017000000	4.8802	--
246.513800100007	55.01340000000000	4.800001	--
--	--	--	--
246.5007000000701	54.94200000000000	4.870000	--

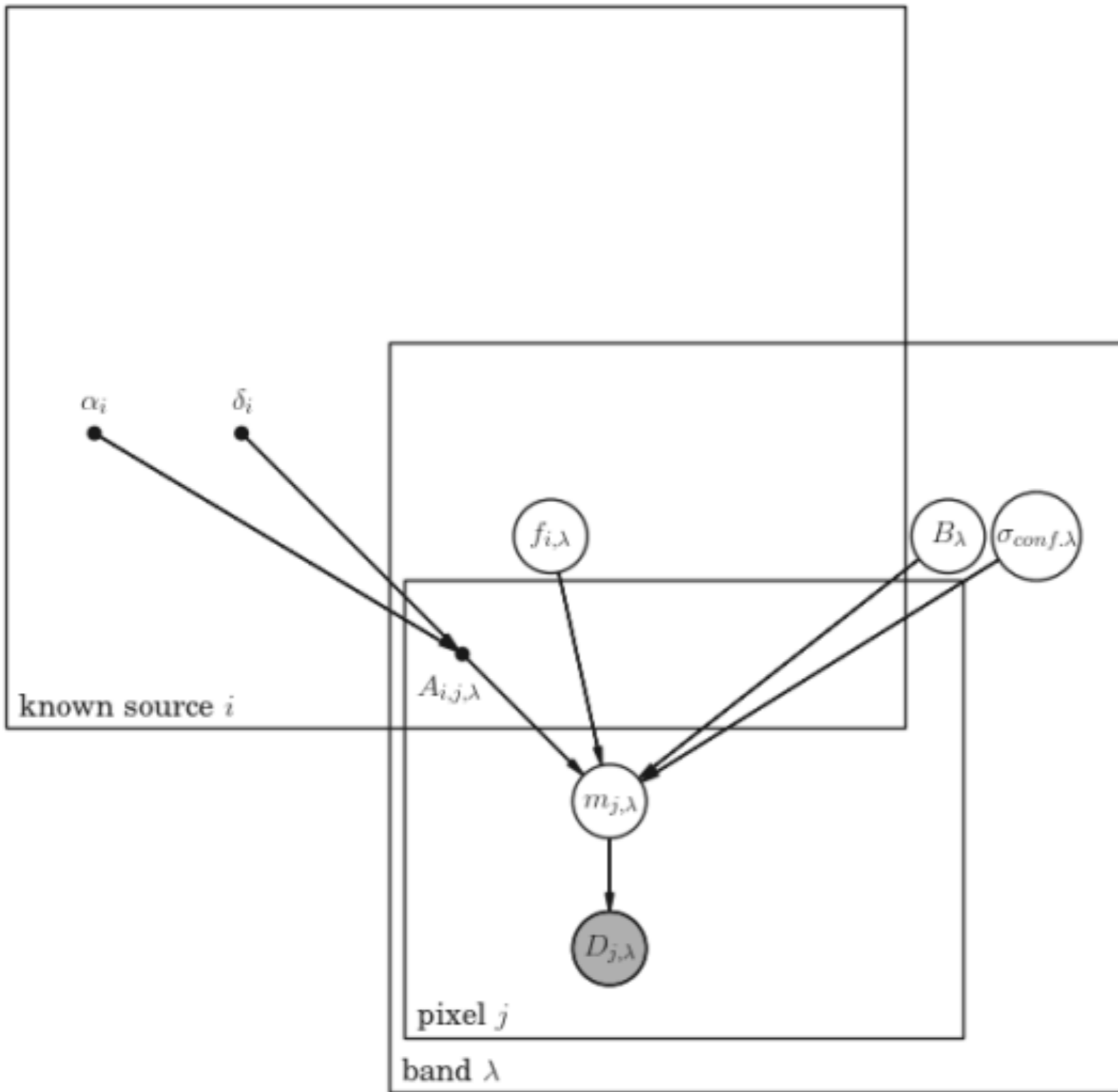




XID+: The probabilistic de-blender for confusion dominated maps, Hurley et al. 2017

->Stan

- Uses Bayesian Inference to get FULL posterior
- Provide a natural framework to introduce additional prior information
- Allows more accurate estimate of flux density errors for each source
- Provides a platform for doing science with the maps rather than catalogues

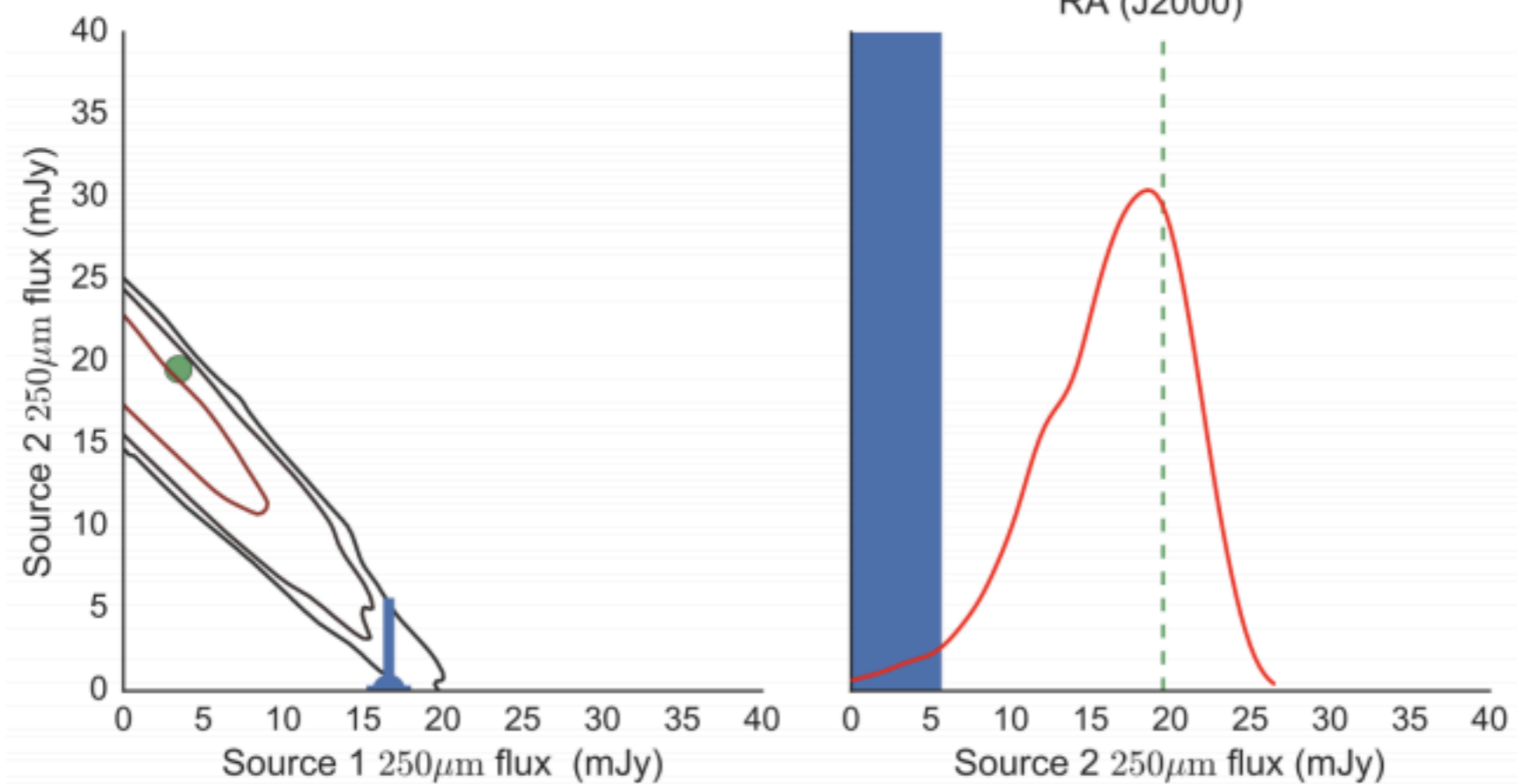
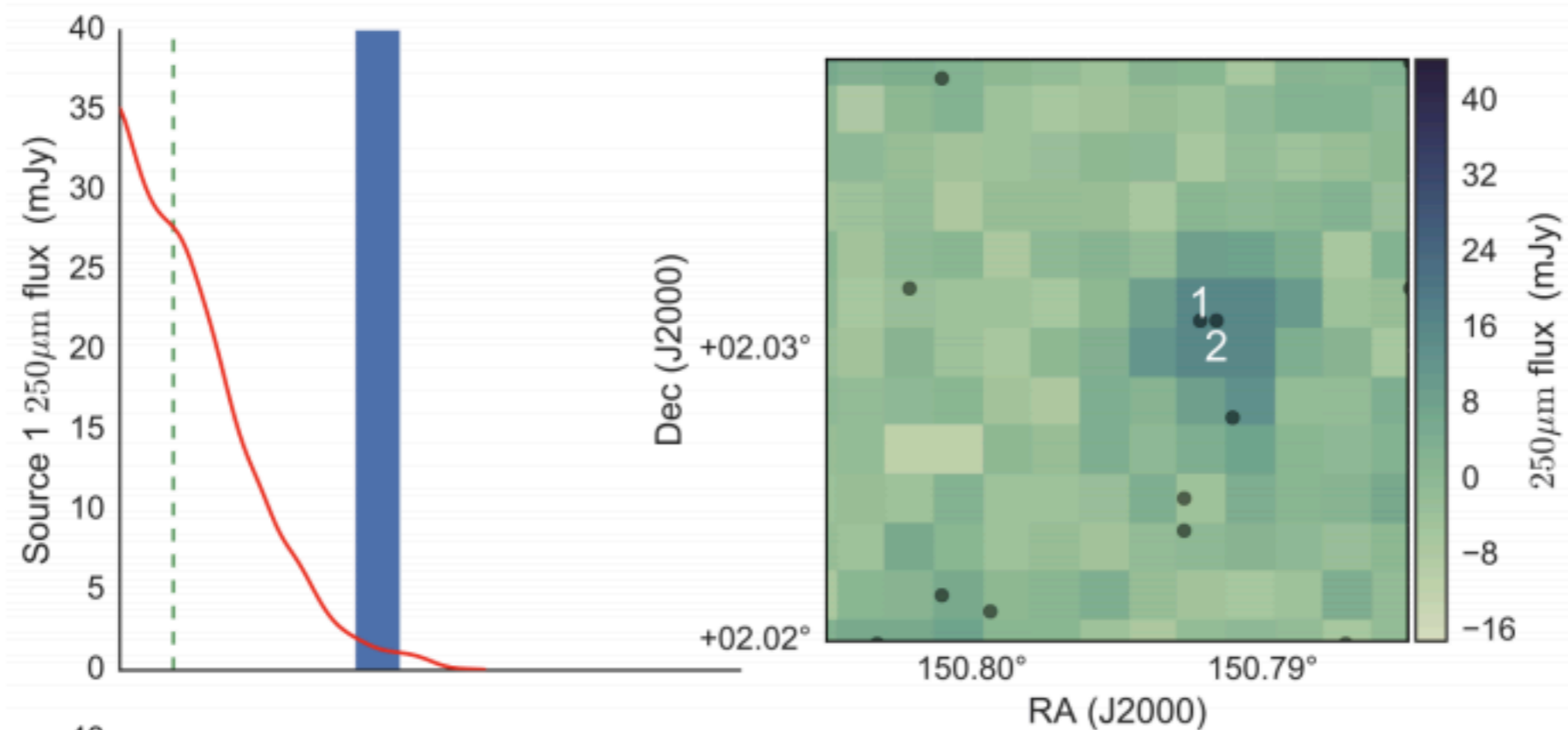


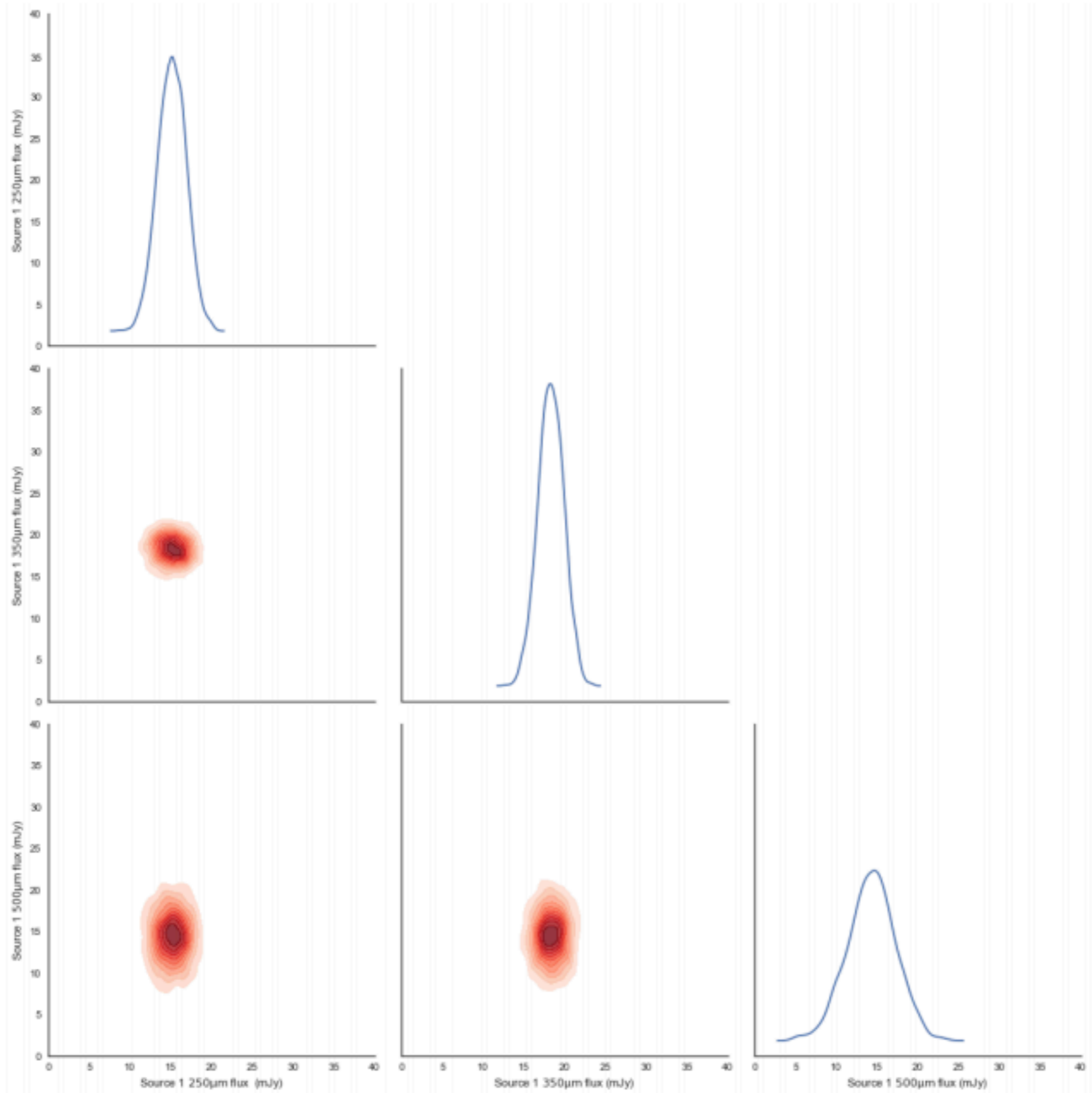
US

UNIVERSITY OF SUSSEX



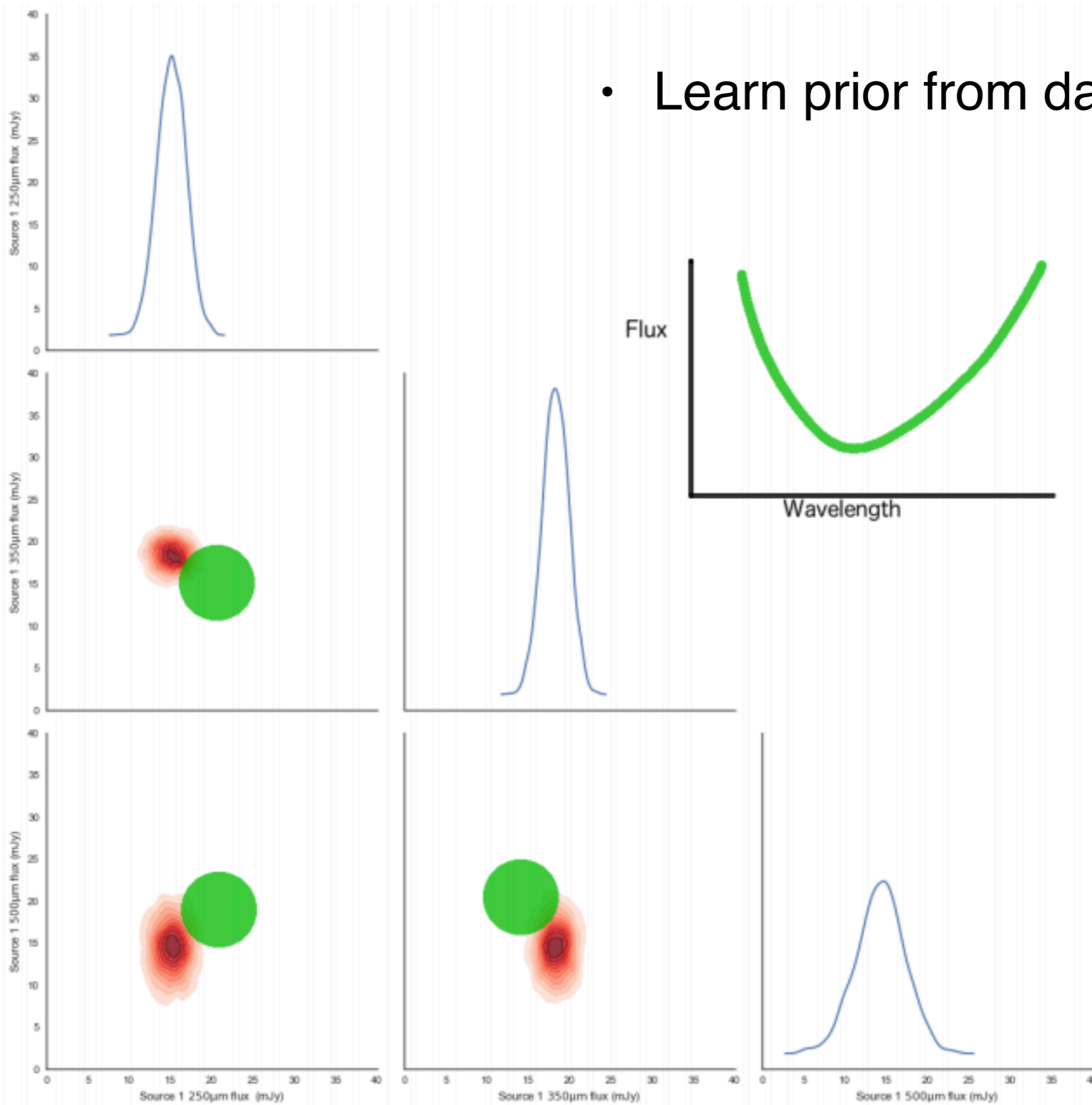








- Learn prior from data

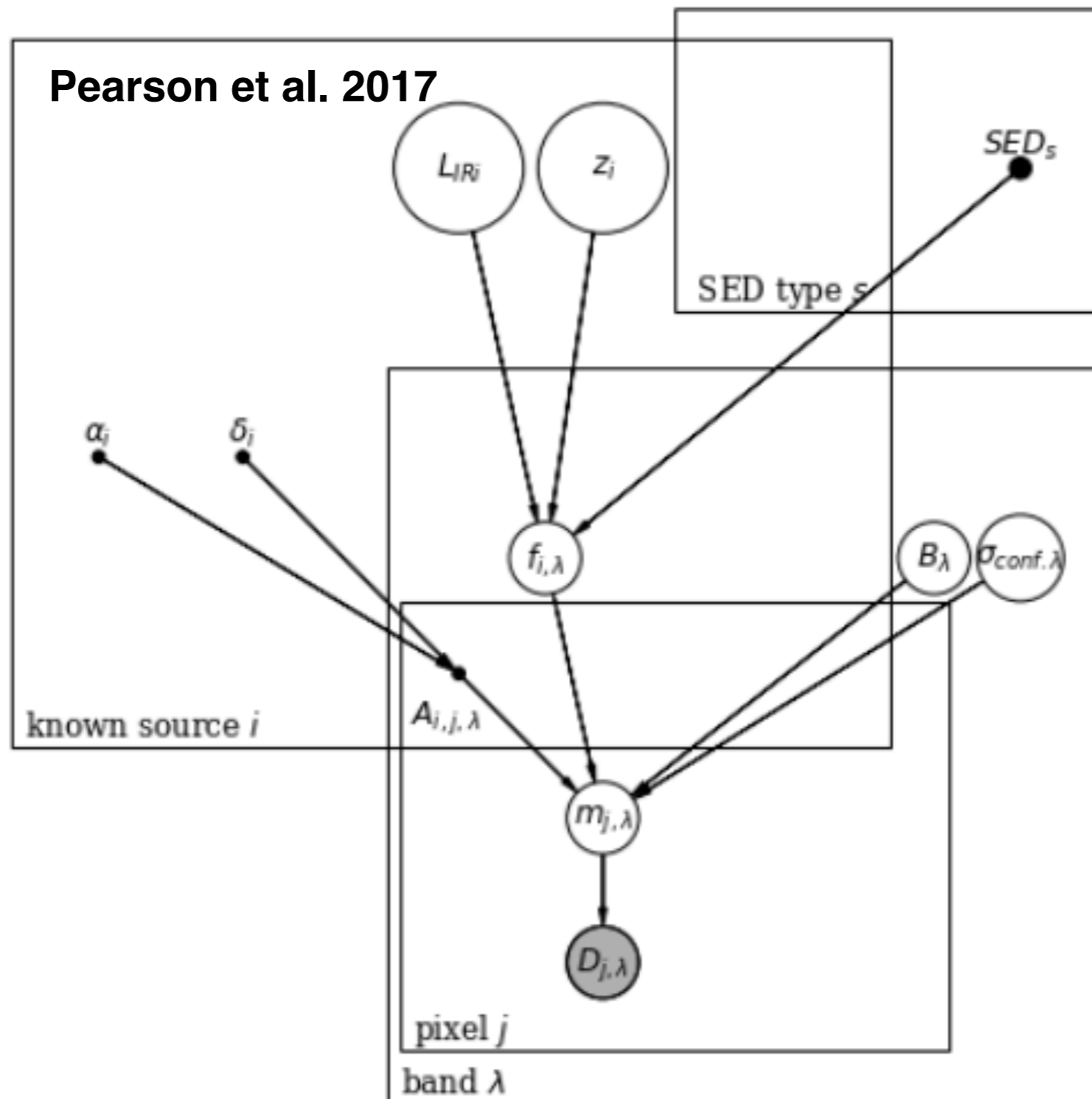


# XID+IR SED



US

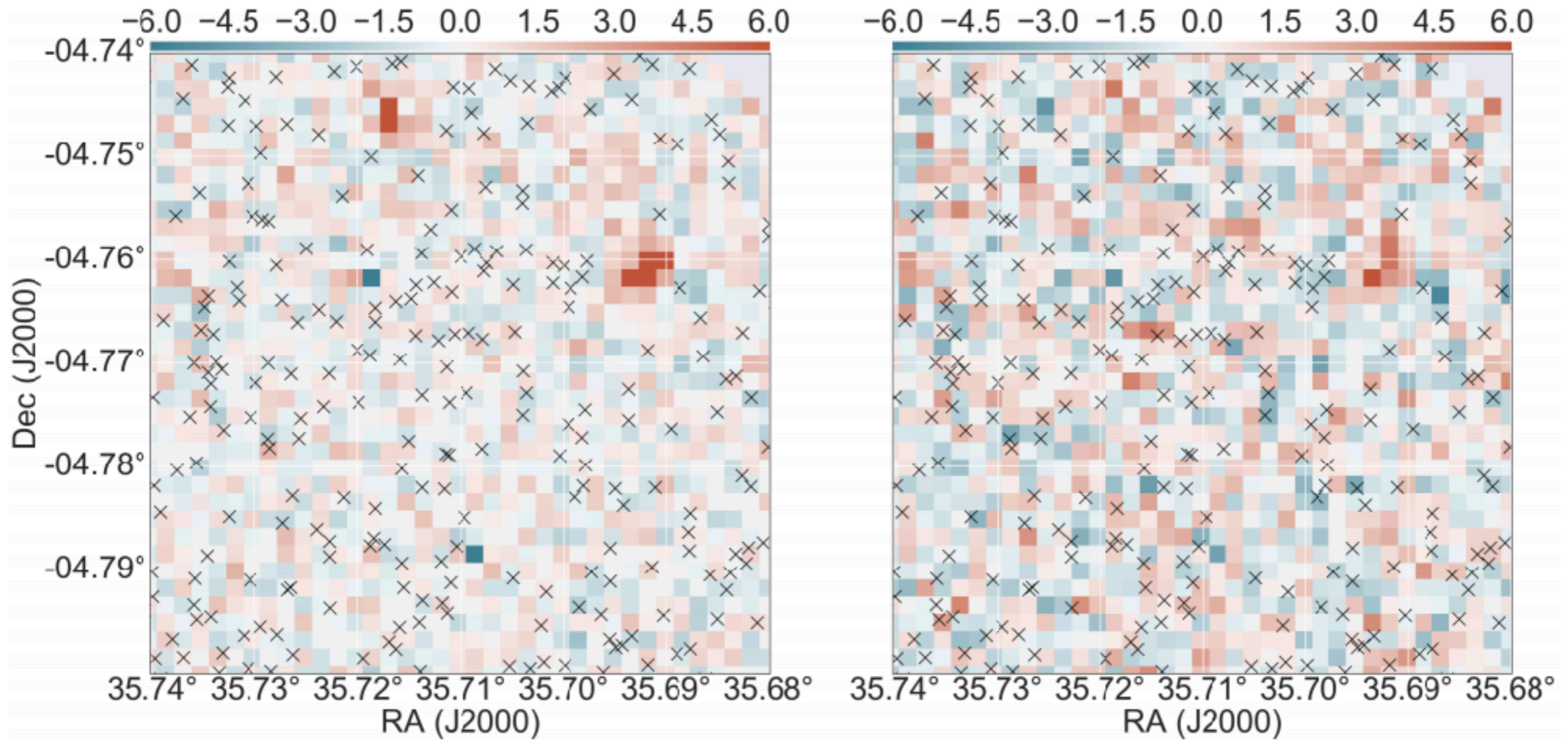
UNIVERSITY OF SUSSEX



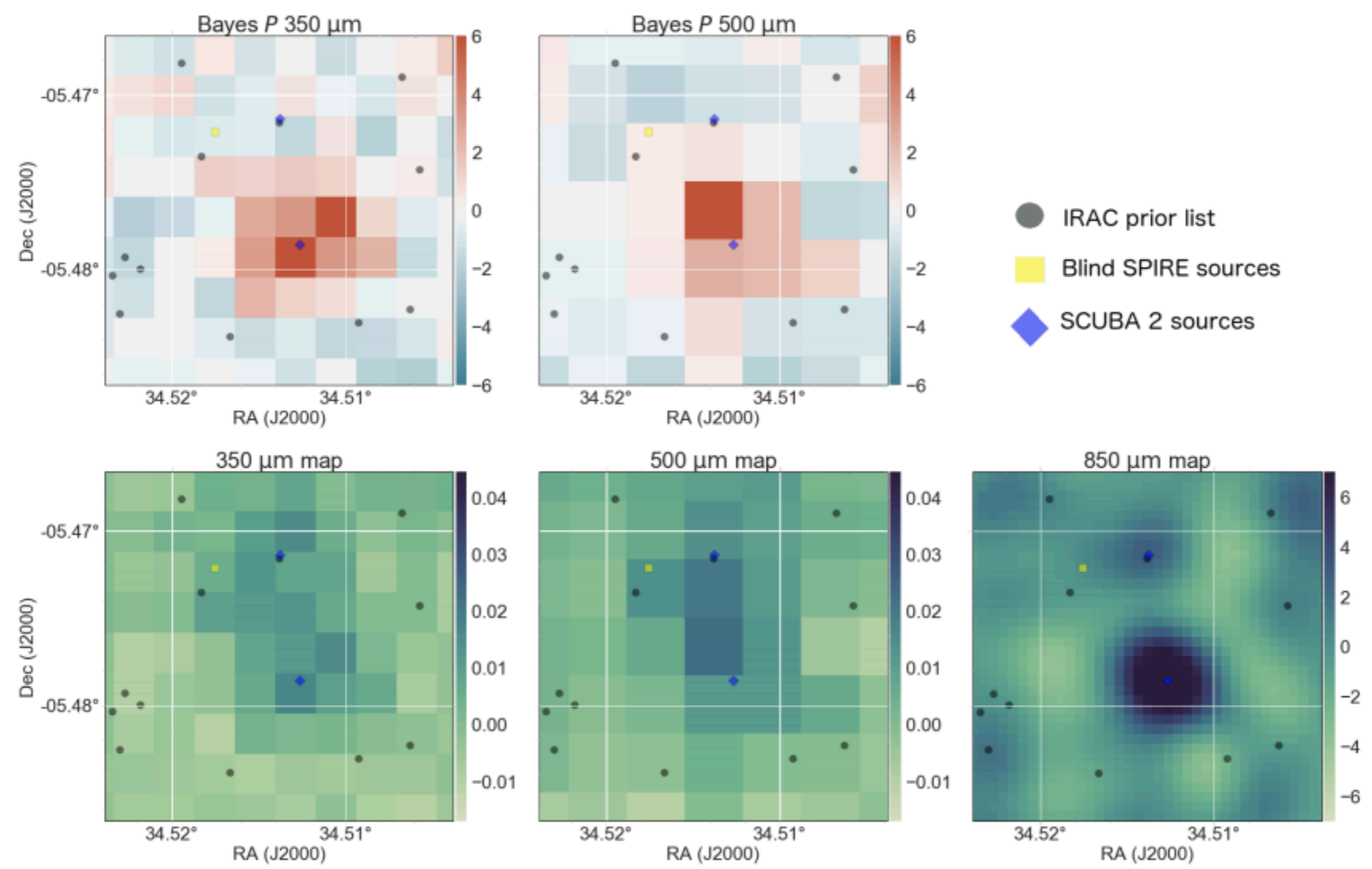




# XID+Bayesian P-value and standard MLM residual maps



# Maps reveal 'Hidden' FIR sources



- photz -> Gaussian processes
- ‘A GP is a supervised non-linear regression algorithm’
- Needs a training set
- n, number of samples, with dimensionality d
- GPz models the distribution of functions that map those inputs on to the desired output
- training (0.8), validation (0.1), and test (0.1) samples



US

UNIVERSITY  
OF SUSSEX

$$\mathbf{X} = \{\mathbf{x}_i\}_{i=1}^n \in \mathbb{R}^{n \times d}$$

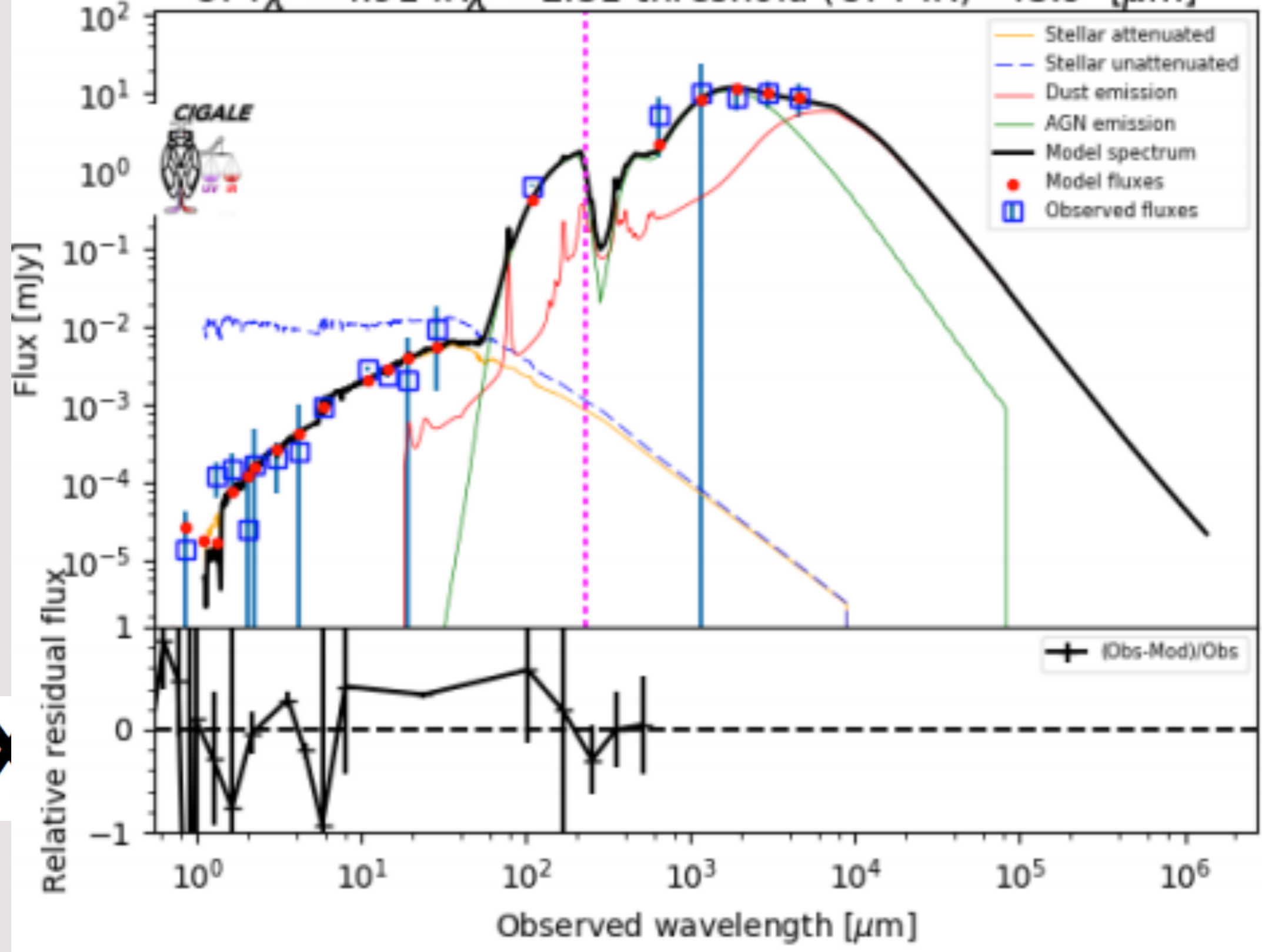
$$\mathbf{y} = \{y_i\}_{i=1}^n \in \mathbb{R}^n$$

$$y_i = f(\mathbf{x}_i) + \epsilon_i$$

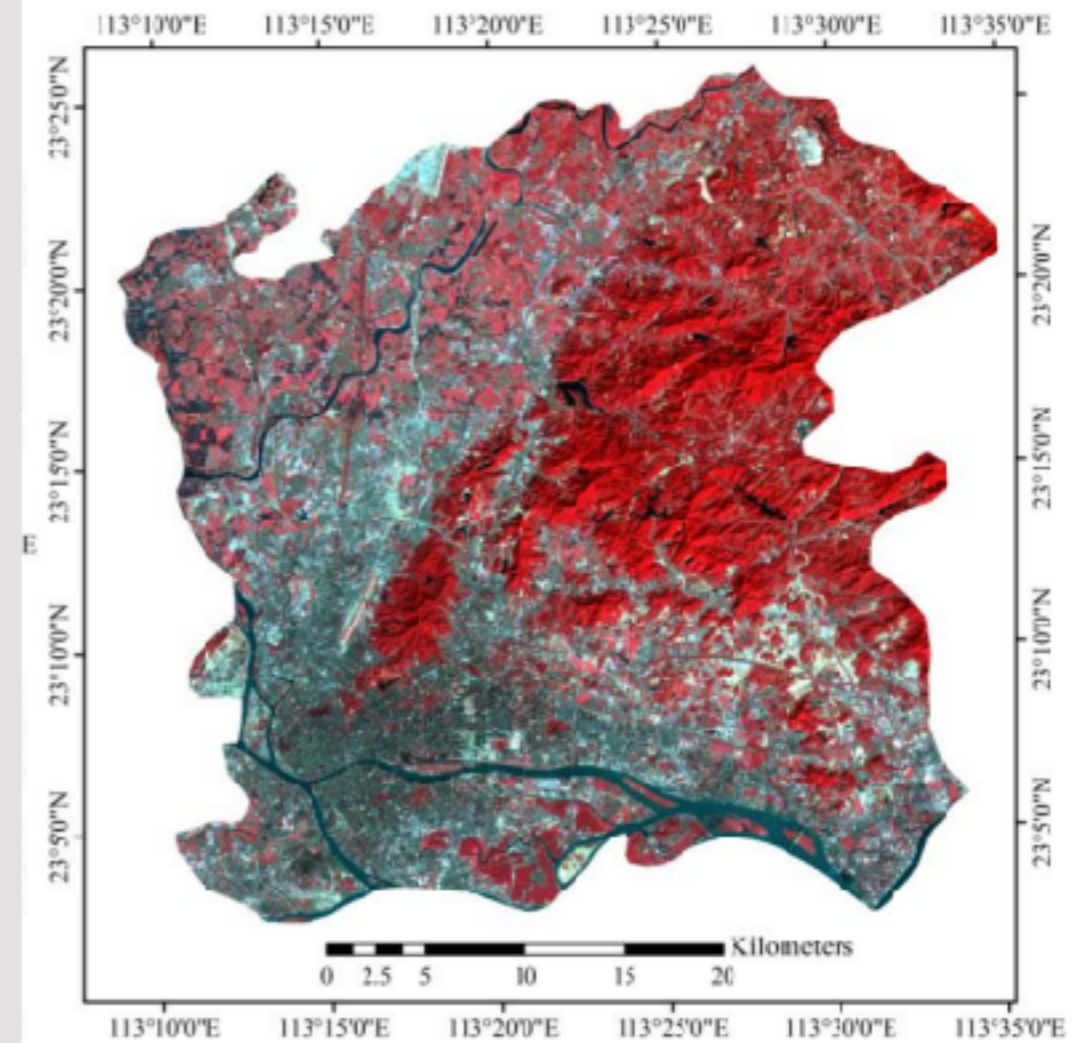


Best model for HELP\_J100156.76+022344.75 at  $z = 4.487$   $\chi^2 = 4.01$  E L P

OPT  $\chi^2 = 4.91$  IR  $\chi^2 = 2.31$  threshold (OPT IR) = 43.9  $[\mu\text{m}]$



- training sample
- 500 pixels
- mean, max, sigma for every wavelength band



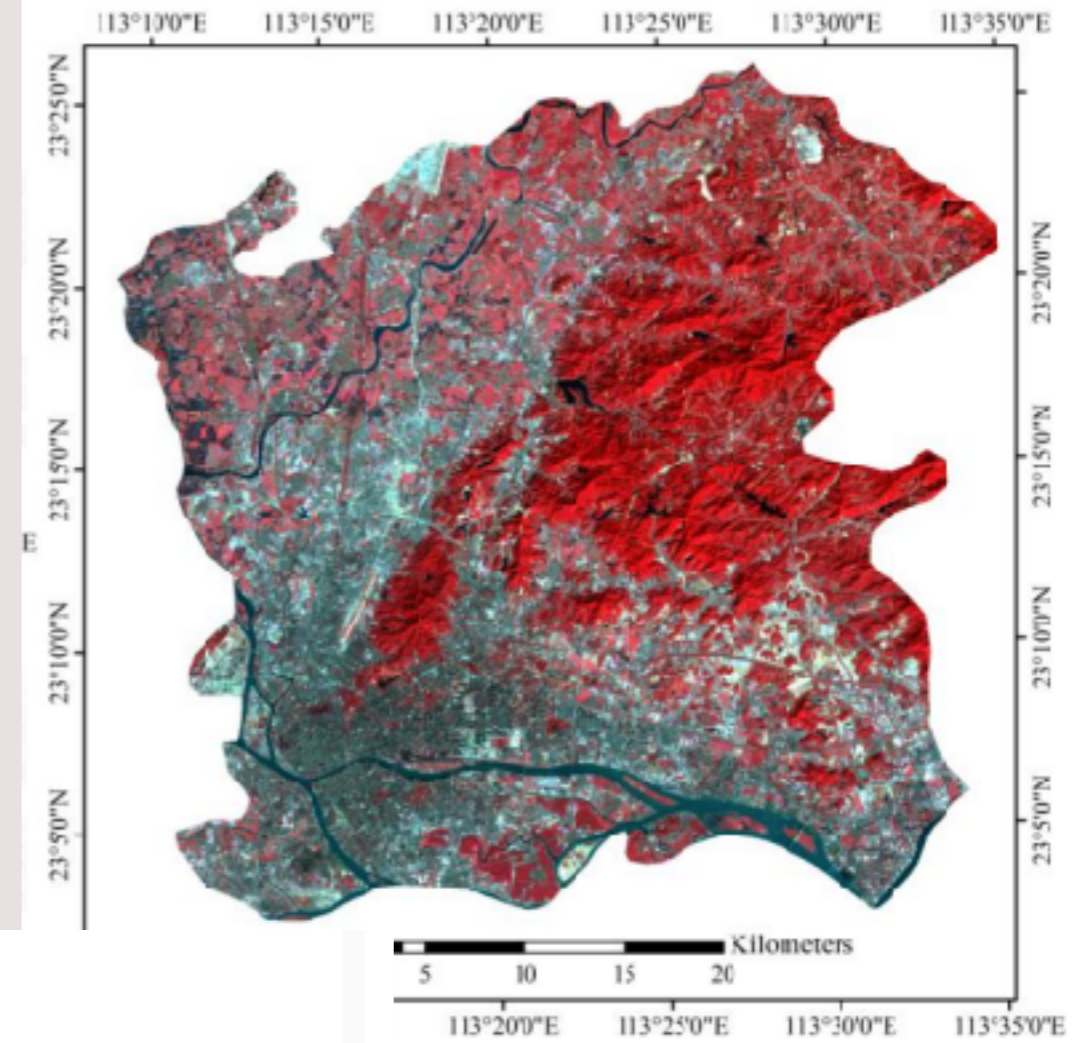
**Table 1.** Land use classification system.

Land-Use Types	Description
Water	Water bodies such as reservoirs, ponds and river
Residential area	Residential areas where driveways and roof tops dominate
Natural forest	Large area of trees
Orchard	Large area of fruit trees planted
Farmland	Fields where vegetables or crops grow
Industrial/commercial	Lands where roof tops of large buildings dominate
Cleared land/Land under construction	Lands where vegetation is denuded or where the construction is underway
Idle land	Lands where no vigorous vegetation grows

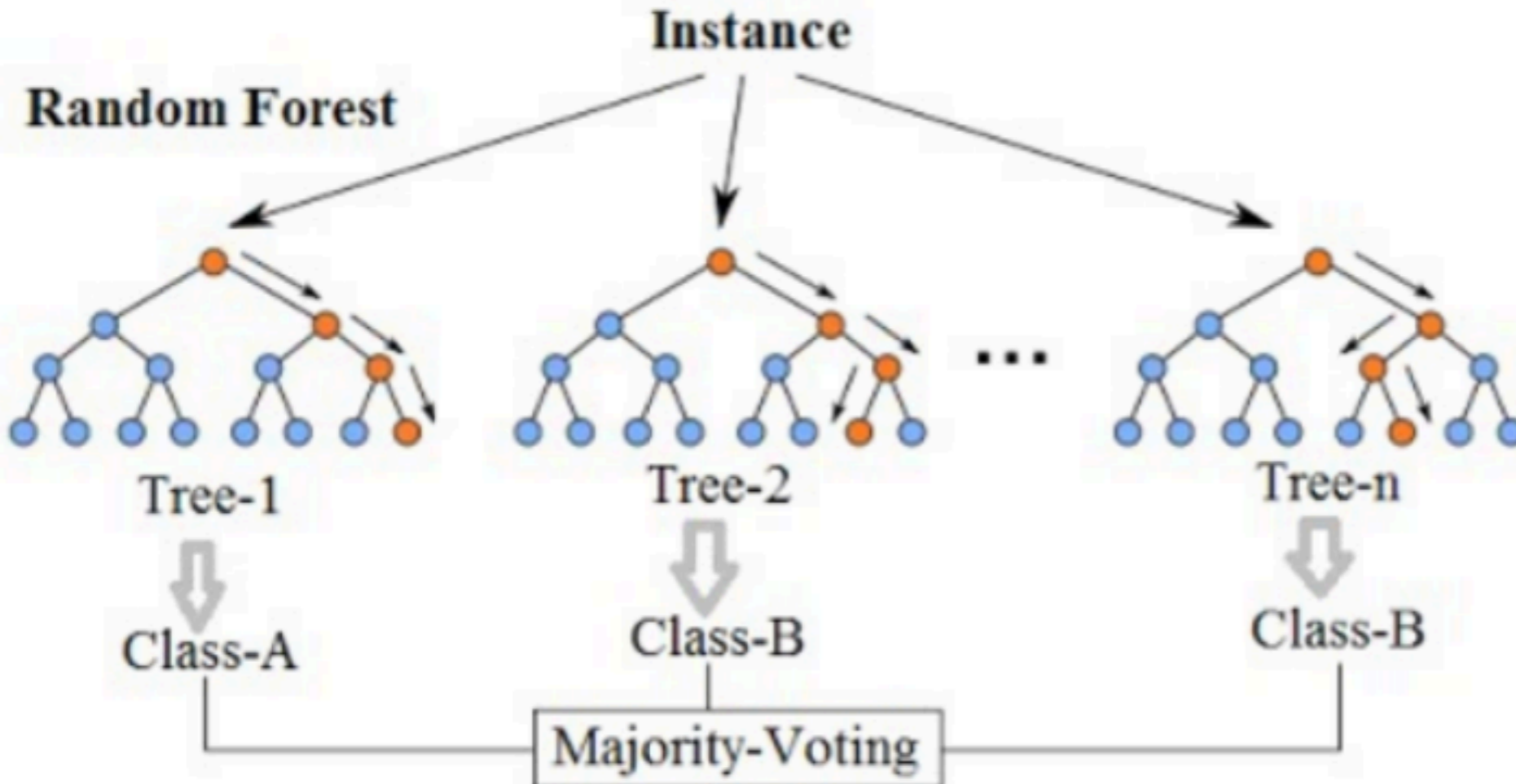
Li et al. 2014



- 85% accurate 4 bands
- 88% accurate 6 bands



### Random Forest



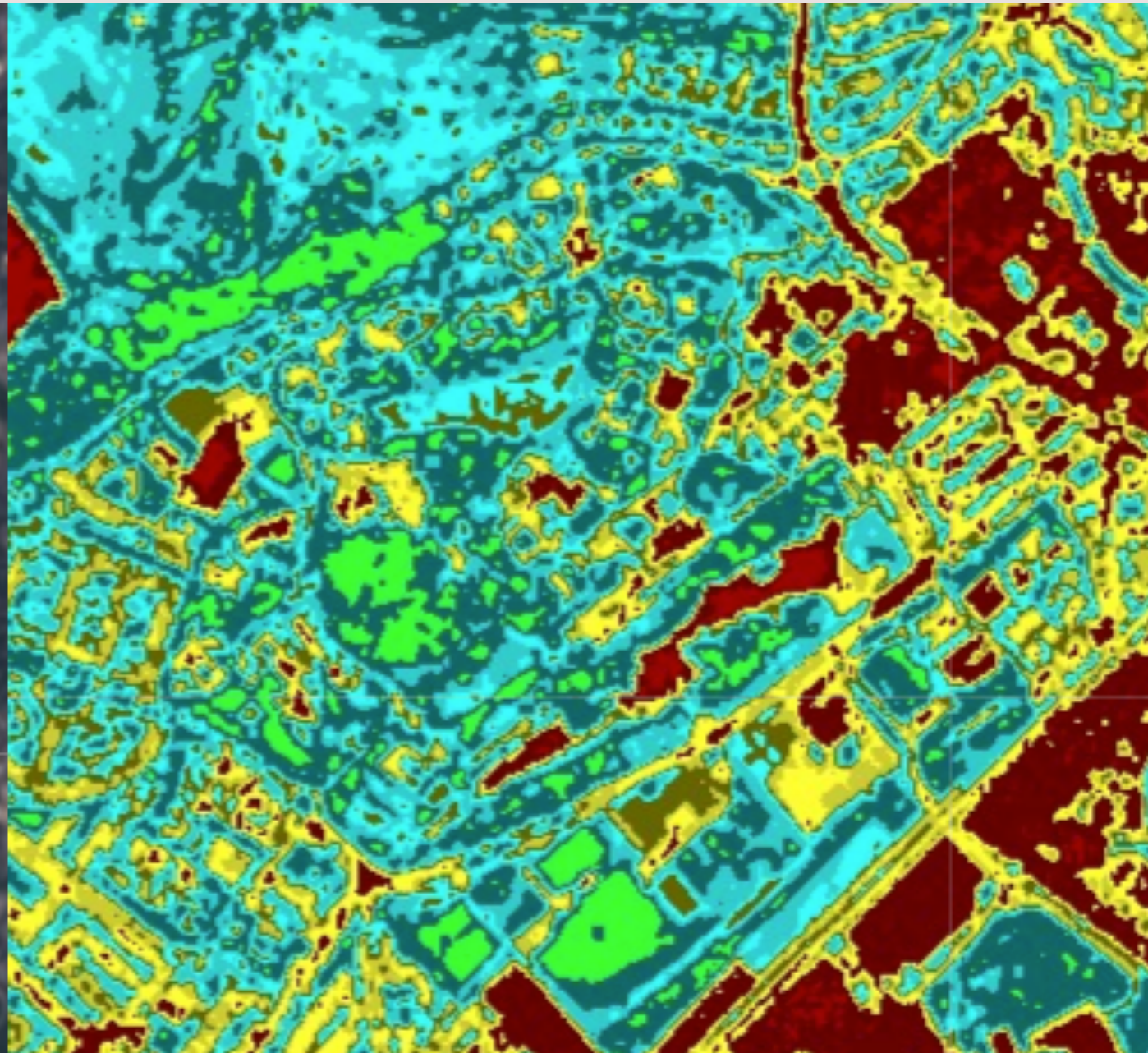
nate

struction is underway



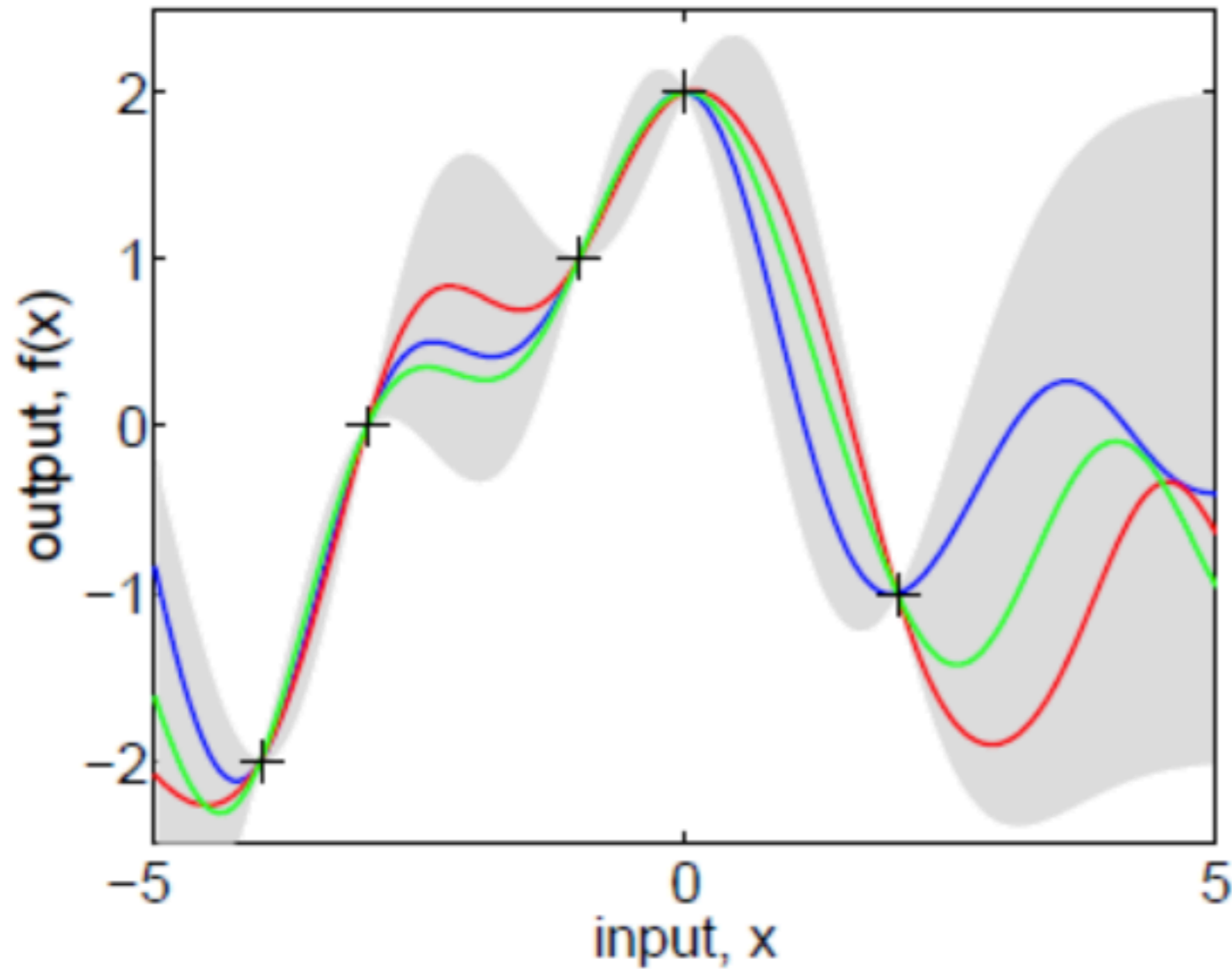


# AstroCast





# AstroCast



- GP
- scikit learn
- RBF and periodic kernel



Science & Technology  
Facilities Council

# AstroCast

```
from sklearn.gaussian_process import GaussianProcessRegressor
from sklearn.gaussian_process.kernels import RBF, WhiteKernel, ExpSineSquared
```

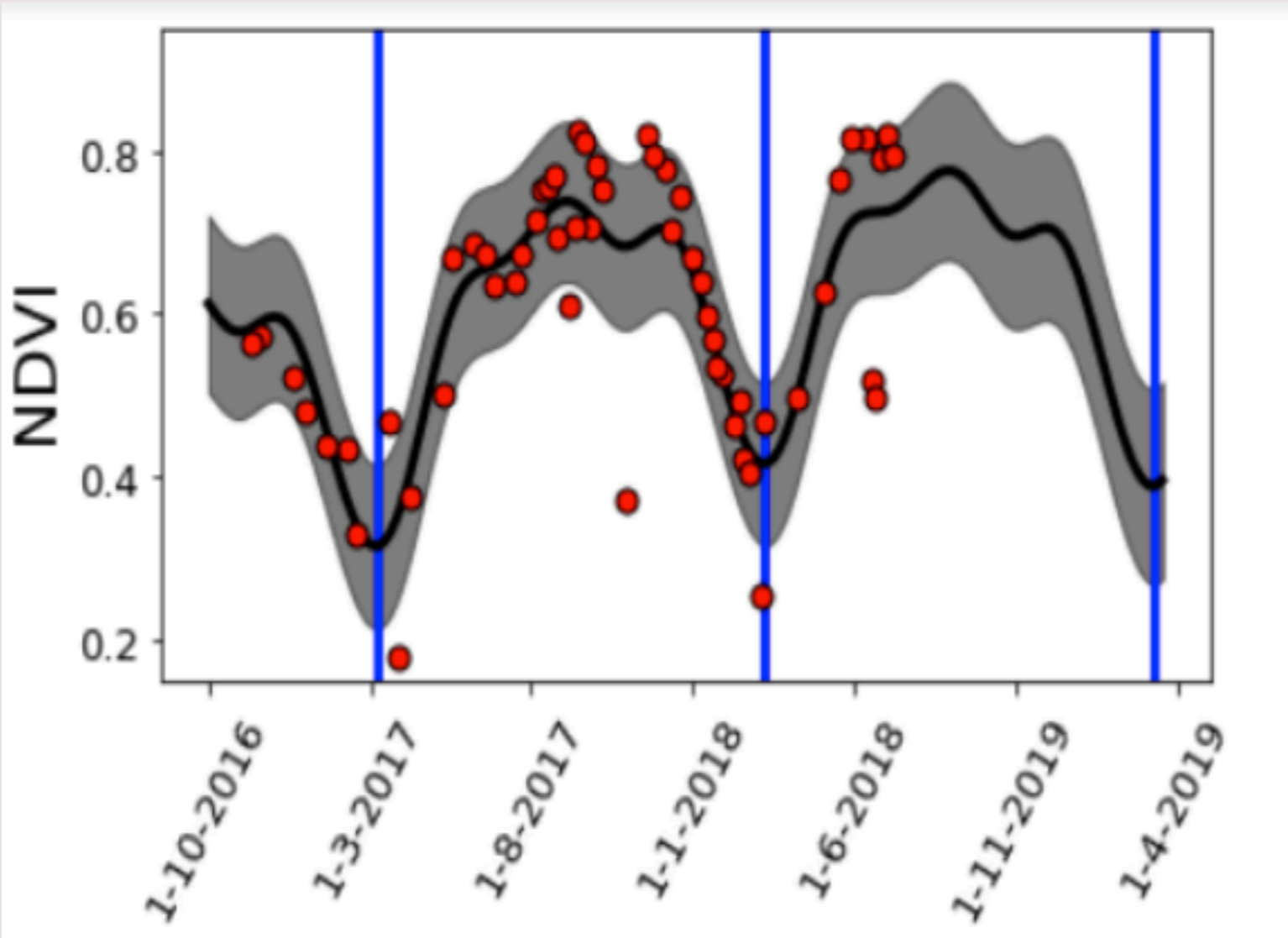
```
k3 = 2.0**2 * RBF(length_scale=300.0)*\
      ExpSineSquared(length_scale=3,length_scale_bounds=(1e-3, 10)\
                      ,periodicity=365.25, periodicity_bounds=(365.25,365.25))
k2 = 2.0**2 * RBF(length_scale=100.0, length_scale_bounds=(1, 1000))
k1 = WhiteKernel(noise_level=1, noise_level_bounds=(1e-10, 1e+2))
kernel =k1+k2+k3

gp = GaussianProcessRegressor(kernel=kernel,
                              alpha=0.0,normalize_y=True).fit(X, y)

X_ = np.linspace(0, 900, 200)
y_mean, y_cov = gp.predict(X_[ :, np.newaxis], return_cov=True)
plt.plot(X_, y_mean)
plt.fill_between(X_, y_mean - np.sqrt(np.diag(y_cov)),
                 y_mean + np.sqrt(np.diag(y_cov)),
                 alpha=0.5, color='k')
```

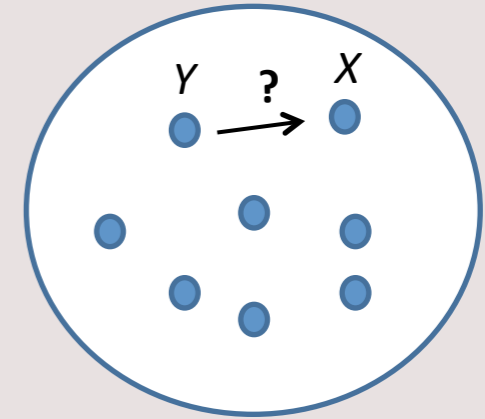


# AstroCast



- GP
- scikit learn
- RBF and periodic kernel
- Include spatial information?

# AstroCast



## *Granger causality*

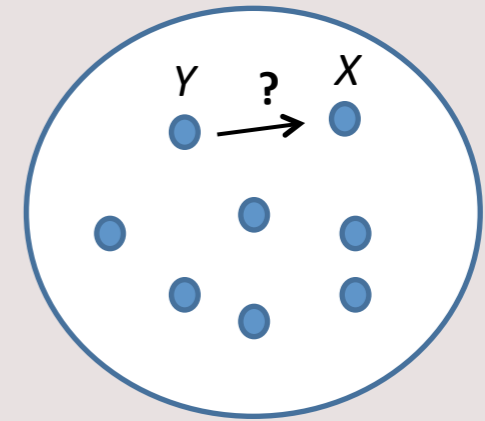
- **Causality based on prediction:** Does past of  $Y$  help predict future of  $X$ ?
- Over and above the past of  $X$  itself, and the past of other variables?
- Prediction by regression, using past observations as predictors.
- Compare prediction errors (residuals) for regressions including and excluding  $Y$

$$X_1(t) = \sum_{j=1}^p A_{11,j} X_1(t-j) + \sum_{j=1}^p A_{12,j} X_2(t-j) + E_1(t)$$

$$X_2(t) = \sum_{j=1}^p A_{21,j} X_1(t-j) + \sum_{j=1}^p A_{22,j} X_2(t-j) + E_2(t)$$



# AstroCast



## *Granger causality*

- Apply to NDVI / VCI vegetation state index time-series.
- Subtract off seasonal fluctuations and overall linear trend.
- Map dependence of GC on distance between predictor and predictee pixels.
- Compare performance of linear regression models with non-linear regression models such as random forests.
- Compare prediction of Principal Component time-series with prediction of regional mean time-series.



# AstroCast



- [herschel.sussex.ac.uk](http://herschel.sussex.ac.uk)
- [github.com/H-E-L-P/dmu\\_products](https://github.com/H-E-L-P/dmu_products)
- [hedam.lam.fr/HELP](http://hedam.lam.fr/HELP)

