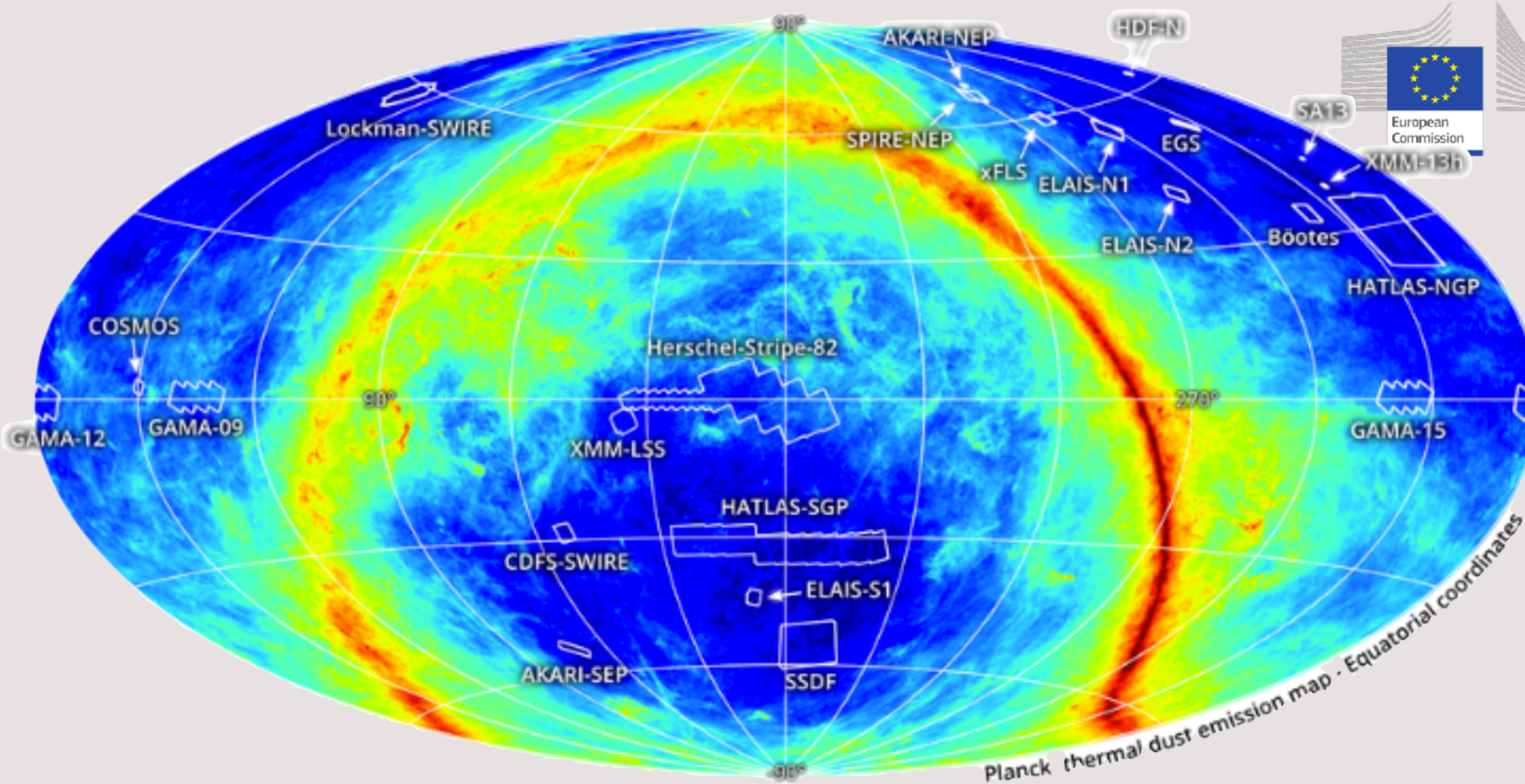
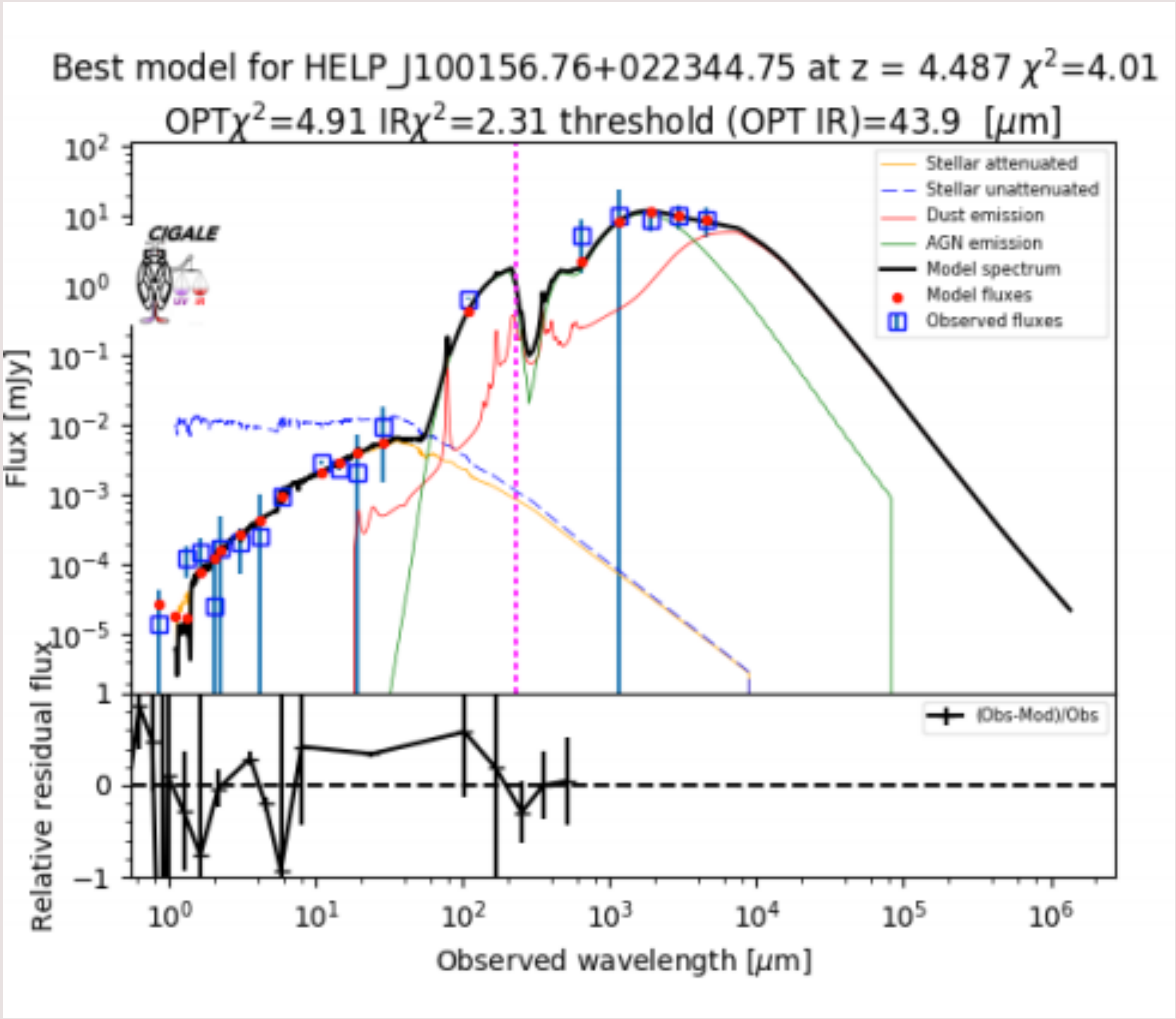
The *Herschel* Extragalactic Legacy Project (HELP) & Machine learning classifications in remote sensing

Steven Duivenvoorden

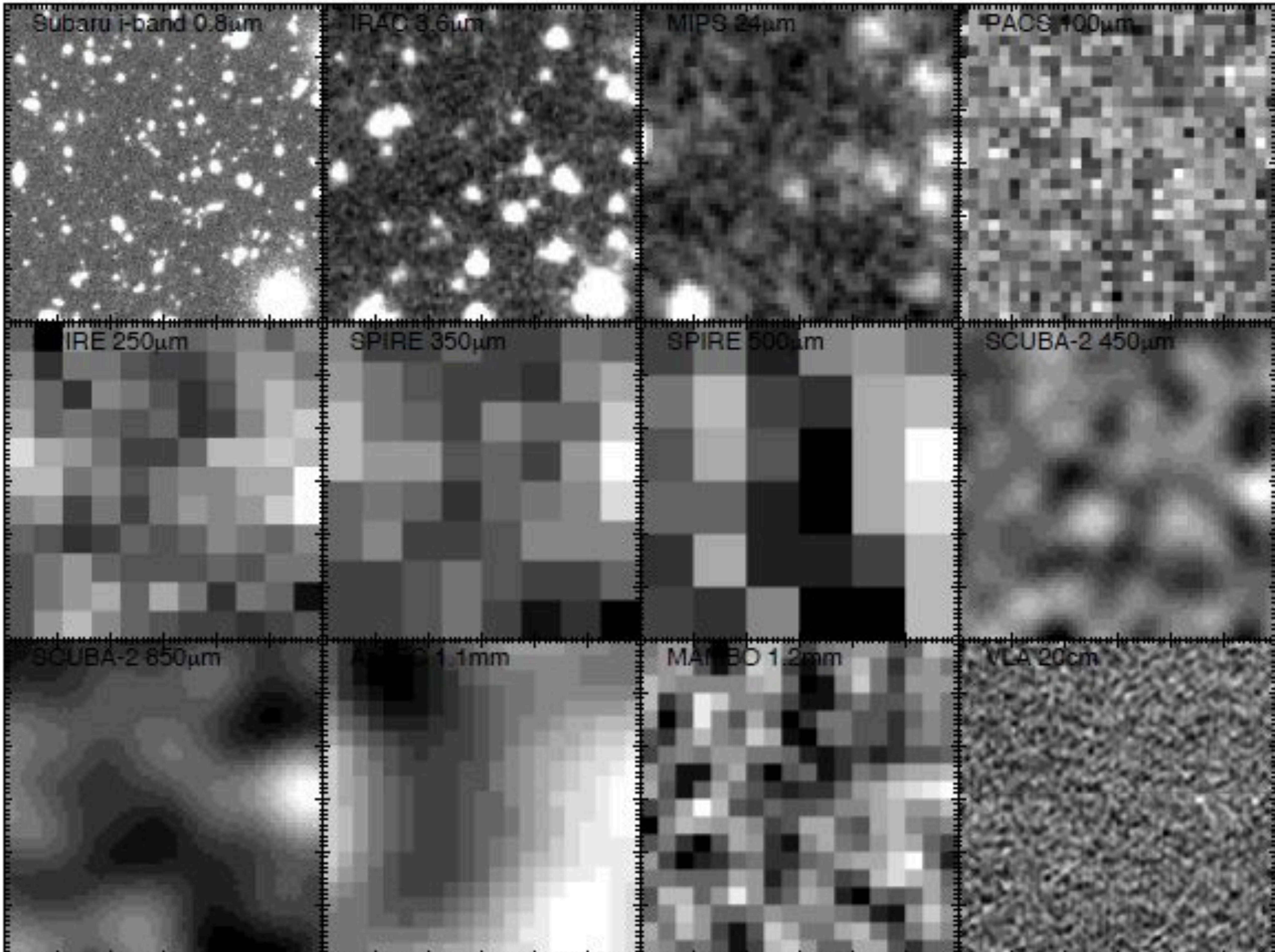
HELP fields

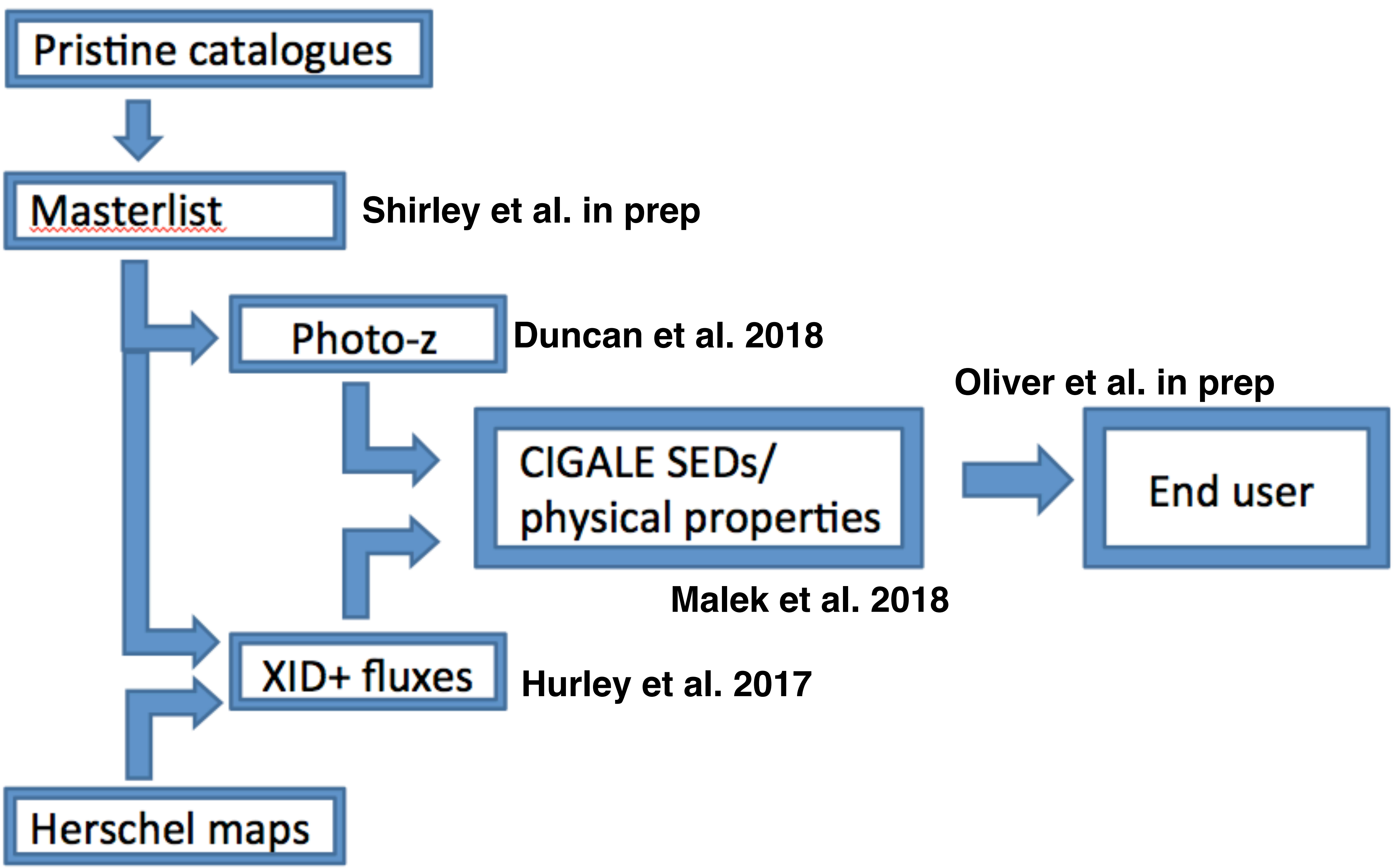
- herschel.sussex.ac.uk





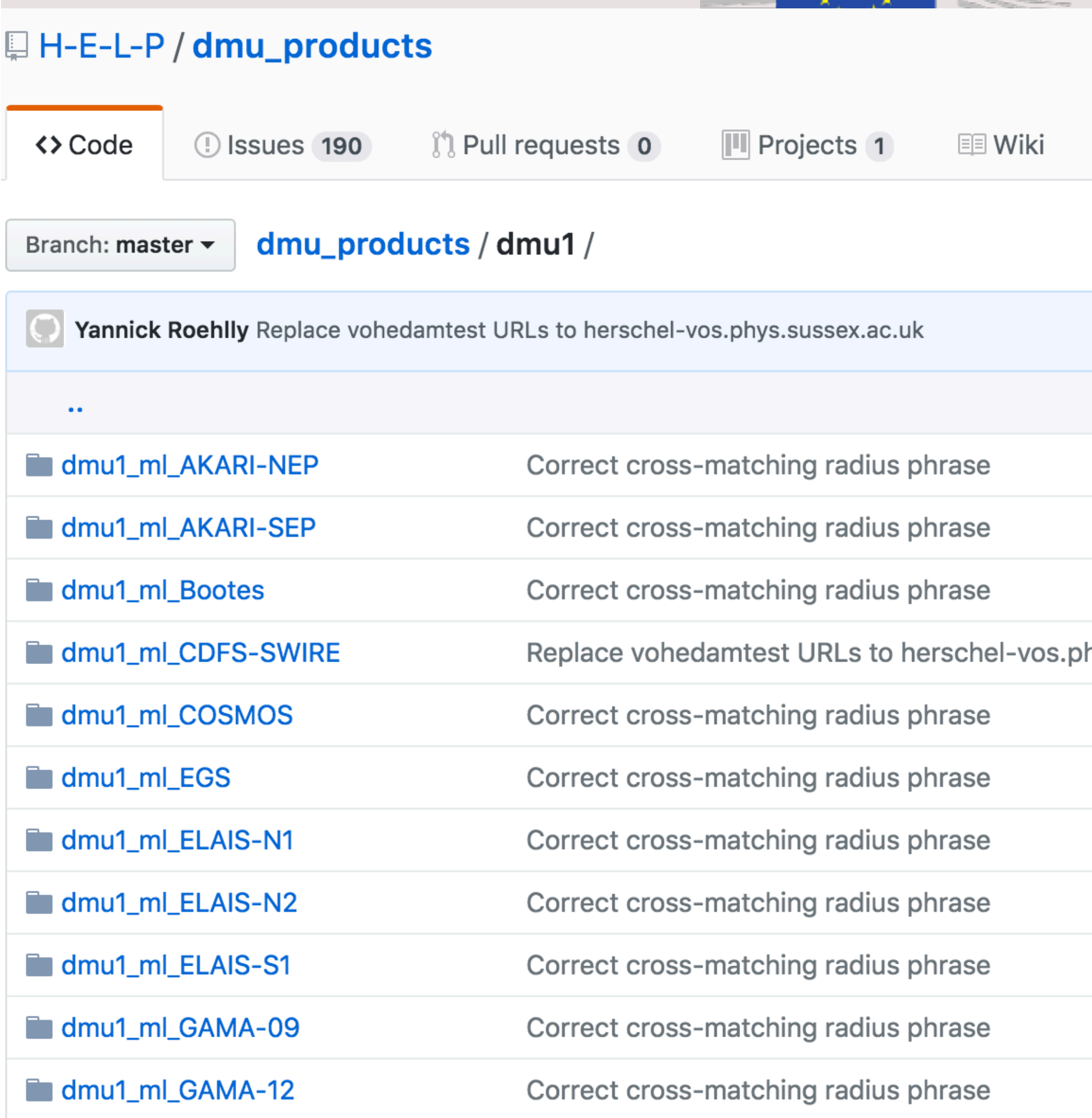
Best model for HELP_J100156.76+022344.75 at z = 4.487 χ²=4.01
OPTχ²=4.91 IRχ²=2.31 threshold (OPT IR)=43.9 [μm]
Flux [mJy]
Relative residual flux
Observed wavelength [μm]
Stellar attenuated
Stellar unattenuated
Dust emission
AGN emission
Model spectrum
Model fluxes
Observed fluxes
(Obs-Mod)/Obs
CIGALE


Subaru i-band 0.8μm
IRAC 3.6μm
MIPS 24μm
PACS 100μm
250μm
SPIRE 350μm
SPIRE 500μm
SCUBA-2 450μm
SCUBA-2 850μm
1.1mm
1.2mm
20cm


Pristine catalogues
Masterlist
Shirley et al. in prep
Photo-z
Duncan et al. 2018
Oliver et al. in prep
CIGALE SEDs/
physical properties
End user
Malek et al. 2018
XID+ fluxes
Hurley et al. 2017
Herschel maps

- herschel.sussex.ac.uk
- 60 Surveys
- 500 Gb catalogues
- github.com/H-E-L-P/dmu_products
- 171 570 436 Objects!
- hedam.lam.fr/HELP

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



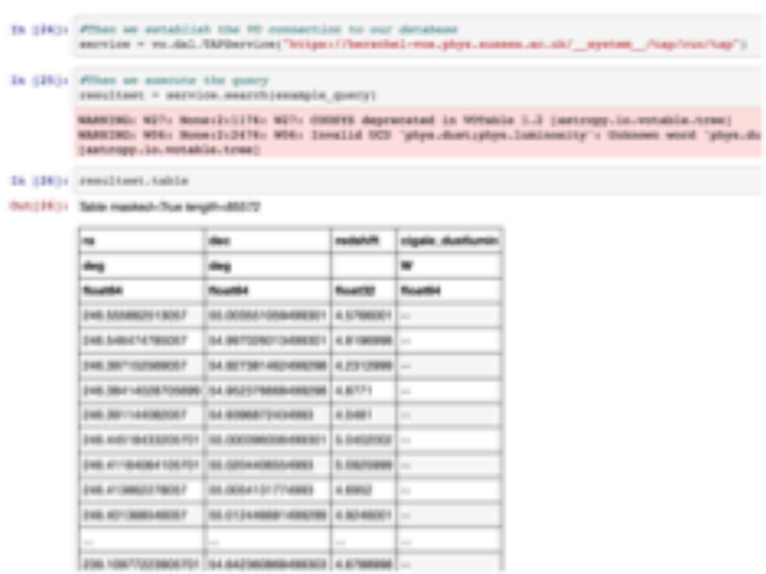
Virtual Observatory at susseX (VOX)

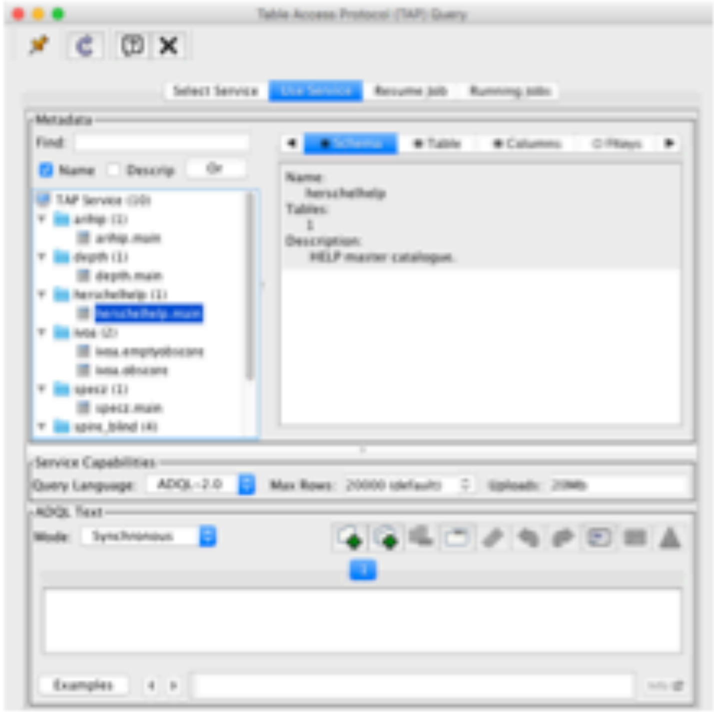
- Allows complex queries to leverage the richness of the data.
- VO standards

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

VO server – data access

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

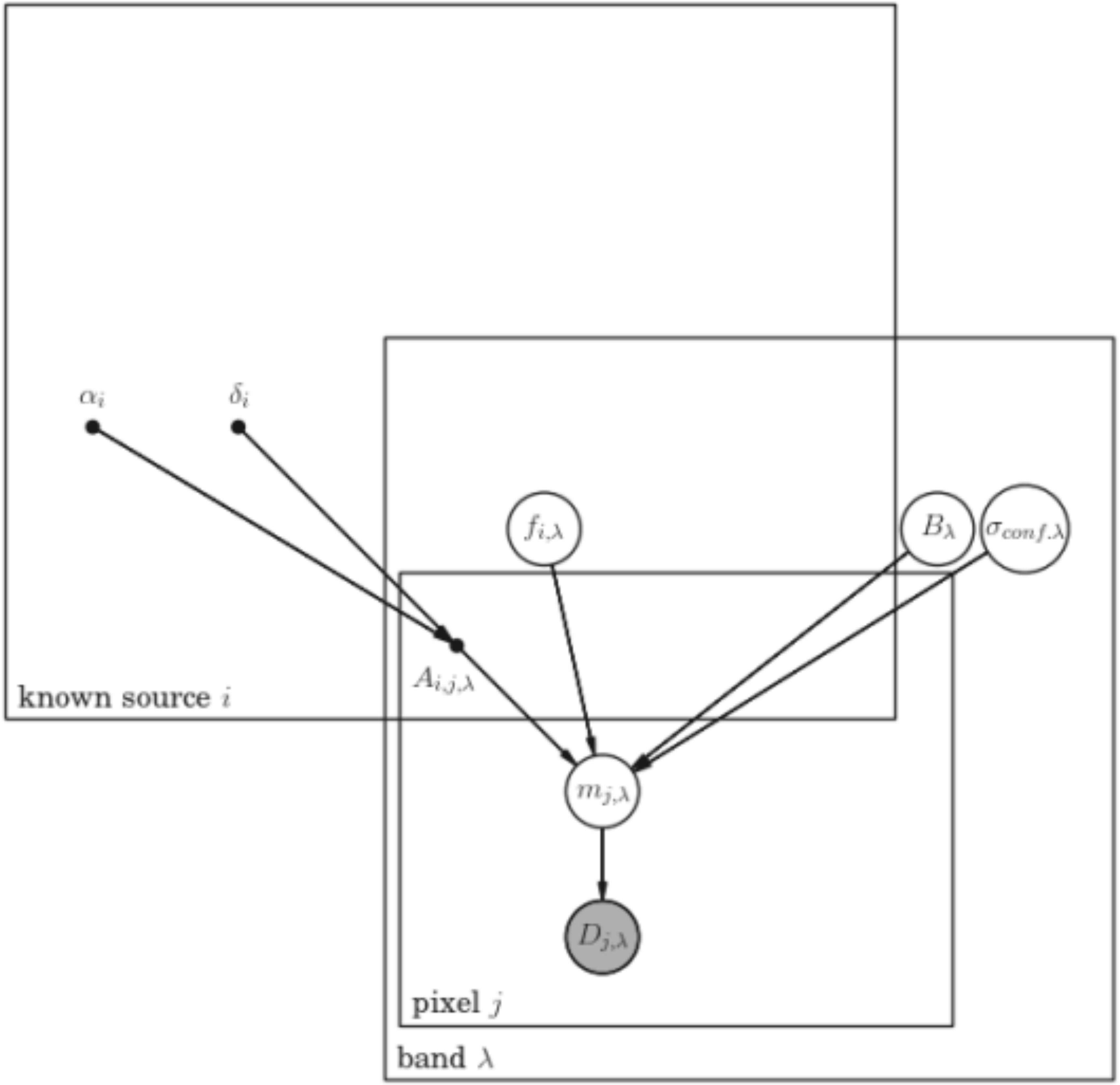




XID+: The probabilistic de-blender for confusion dominated maps, Hurley et al. 2017

->Stan

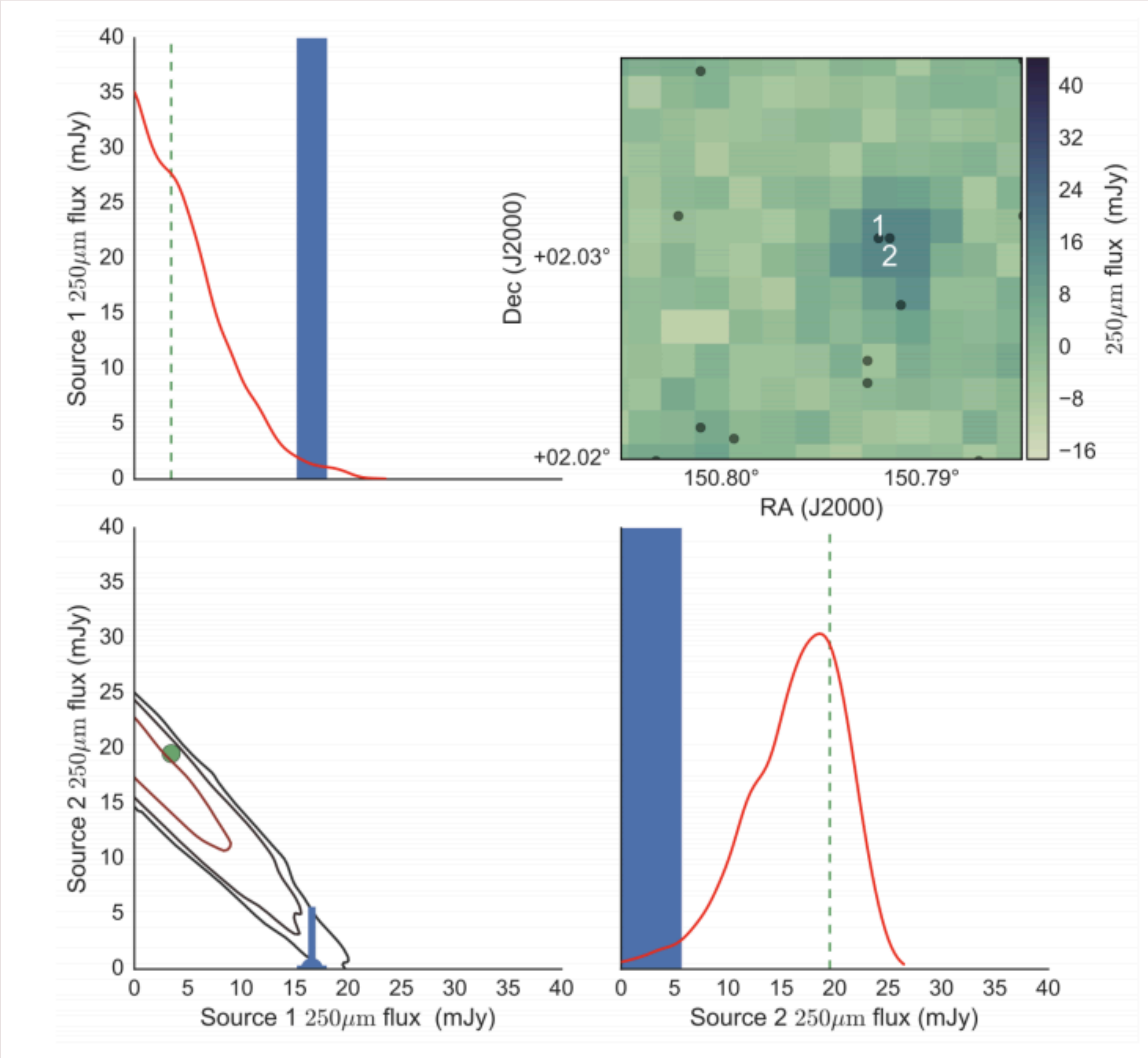
- Uses Bayesian Inference to get FULL posterior
- Provide a natural framework to introduce additional prior information
- Allows more accurate estimate of flux density errors for each source
- Provides a platform for doing science with the maps rather than catalogues



HELP

UNIVERSITY OF SUSSEX

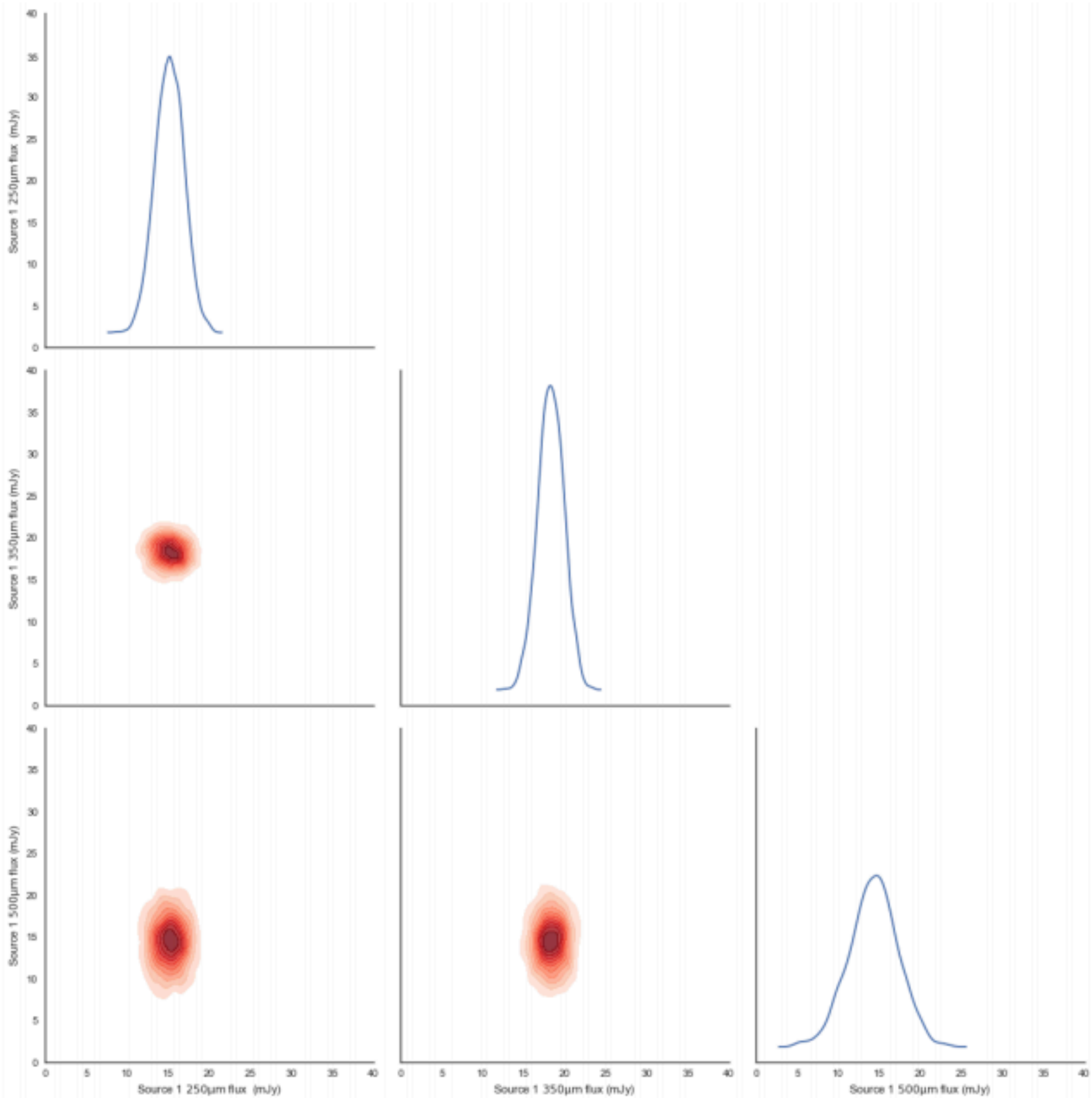
European Commission



HELP

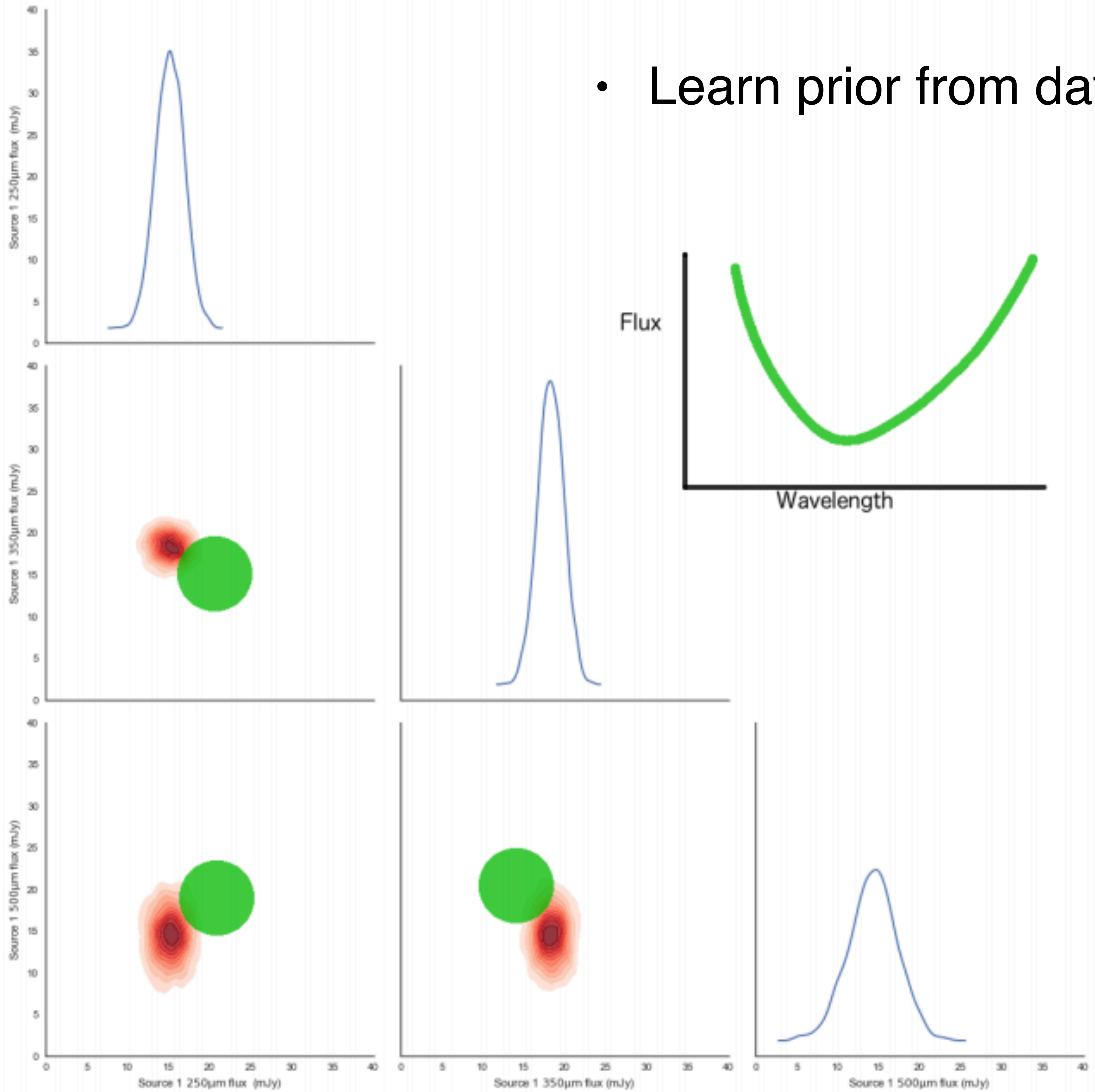
UNIVERSITY OF SUSSEX

European Commission

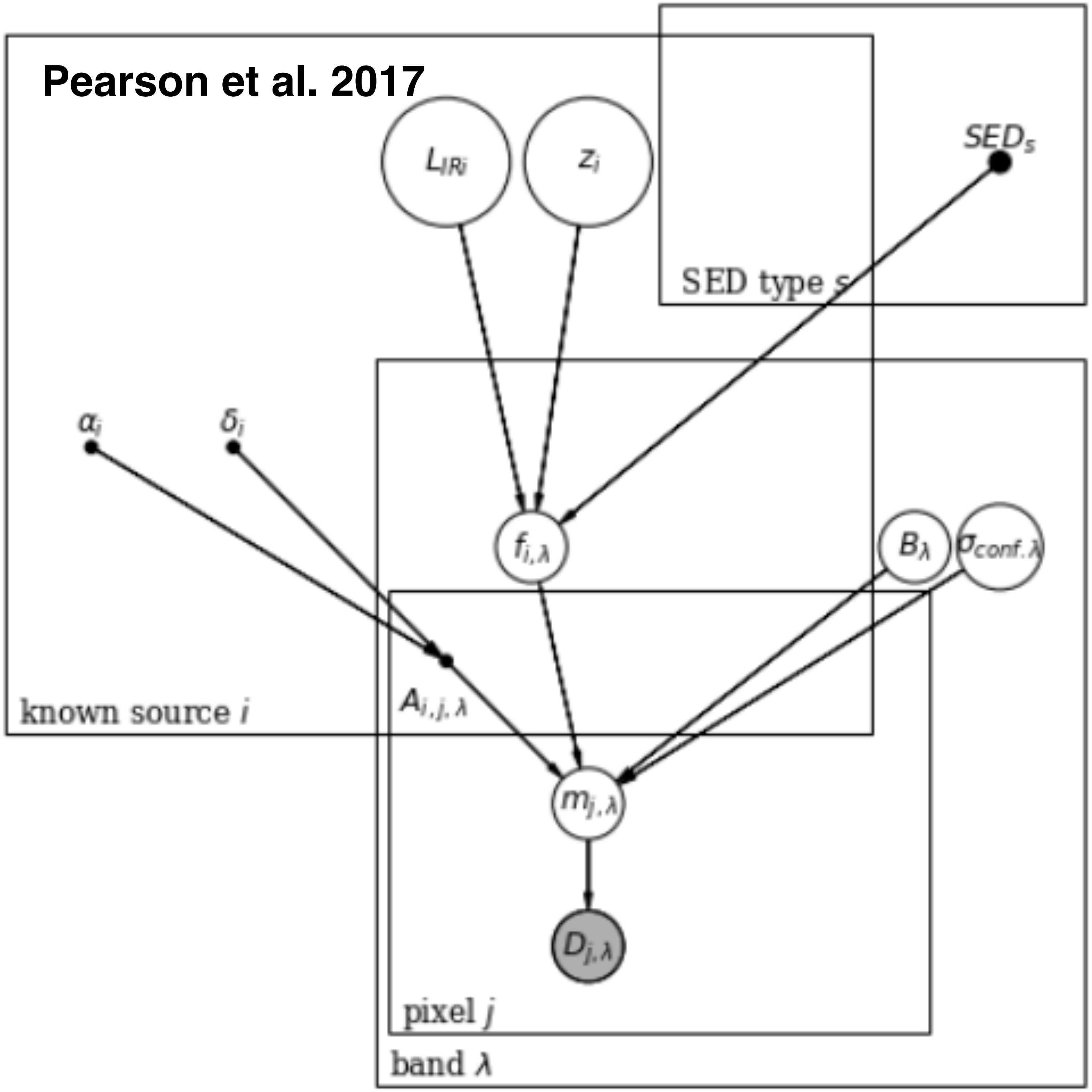


Source 1 250µm flux (mJy)
Source 1 350µm flux (mJy)
Source 1 500µm flux (mJy)
Source 1 250µm flux (mJy)
Source 1 350µm flux (mJy)
Source 1 500µm flux (mJy)

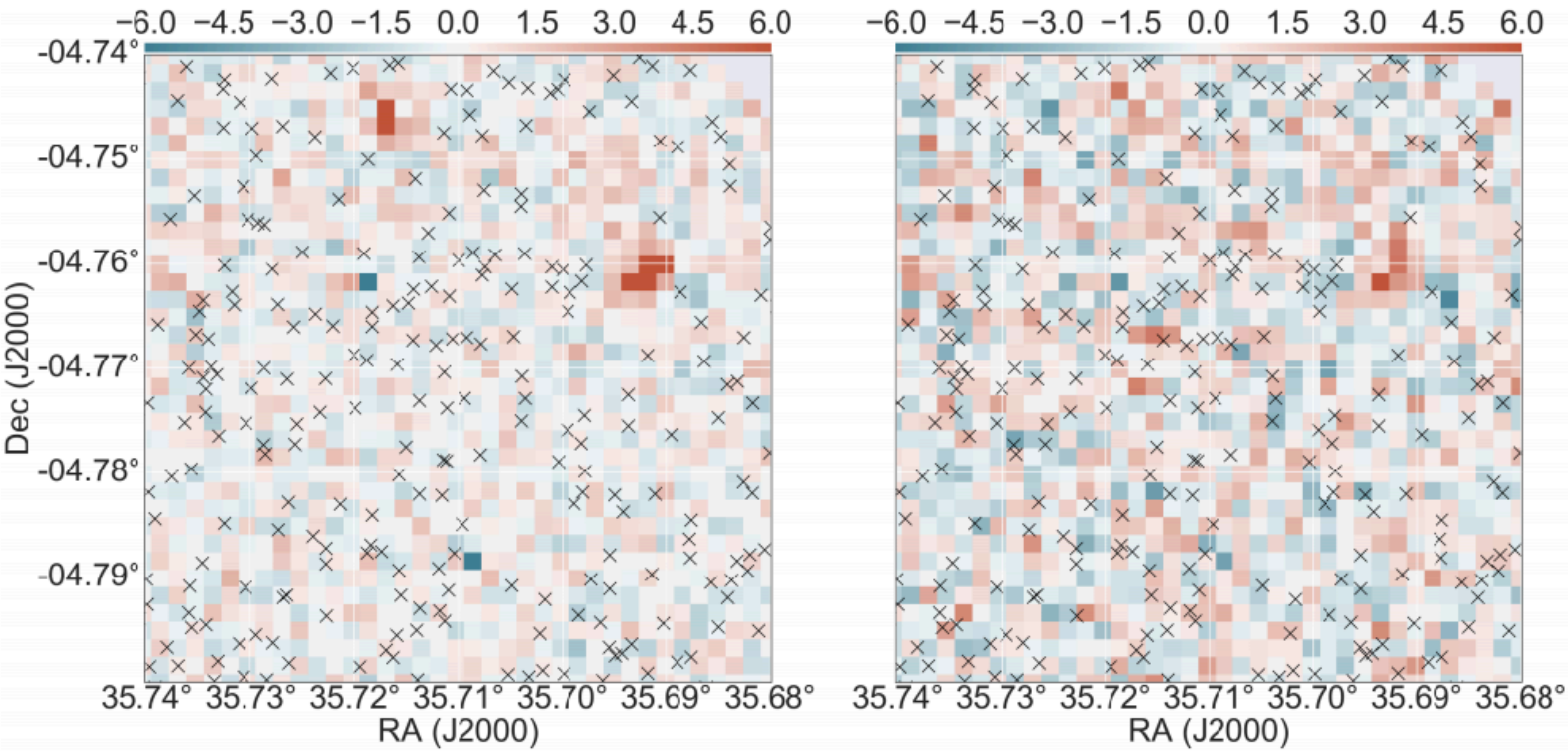
- Learn prior from data



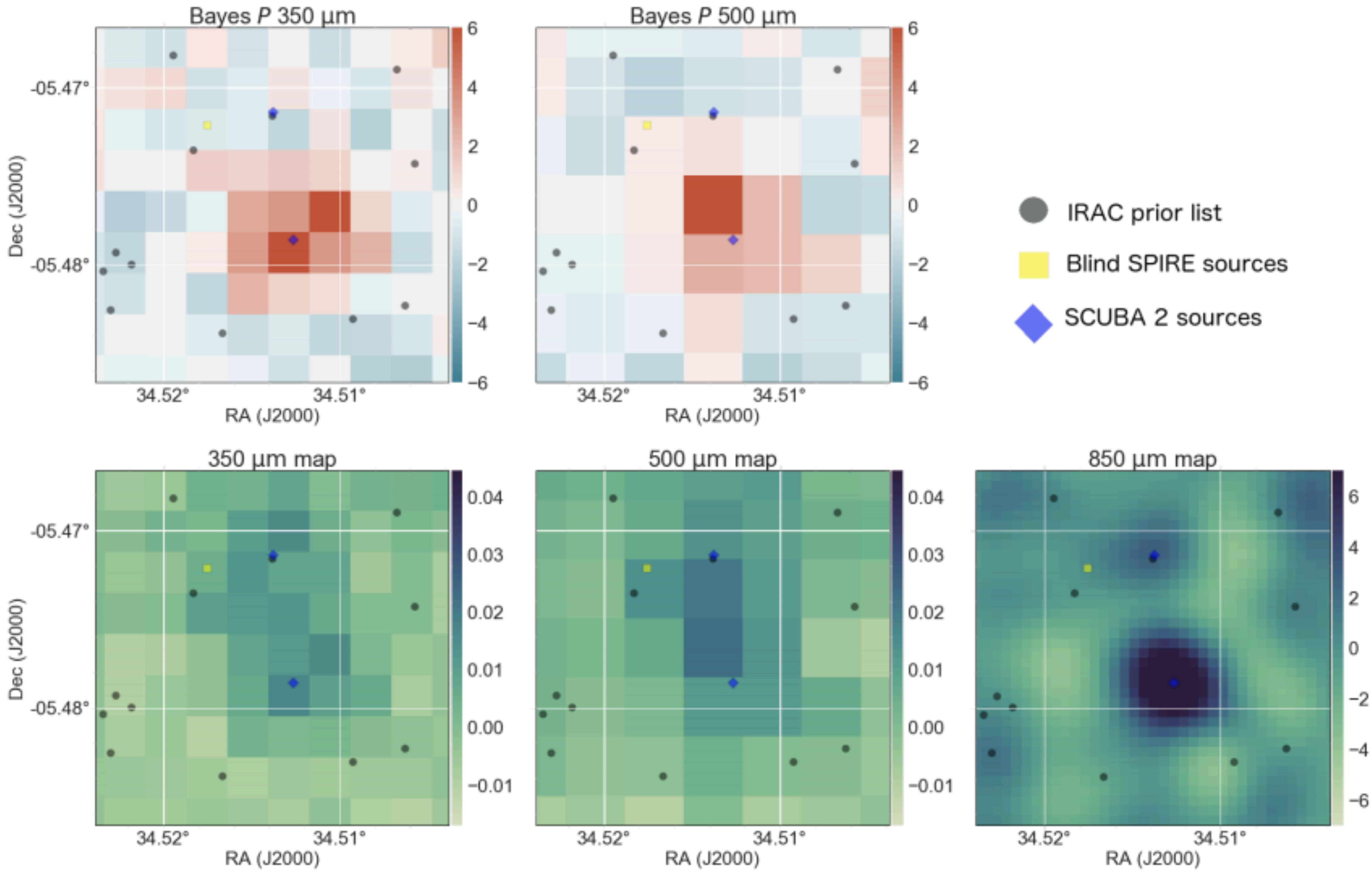
XID+IR SED


Pearson et al. 2017
L_{IRi}
z_i
SED_s
SED type s
α_i
δ_i
$f_{i,\lambda}$
B_λ
$\sigma_{conf.\lambda}$
$A_{i,j,\lambda}$
known source i
$m_{j,\lambda}$
$D_{j,\lambda}$
pixel j
band λ

XID+Bayesian P-value and standard MLM residual maps



Maps reveal 'Hidden' FIR sources



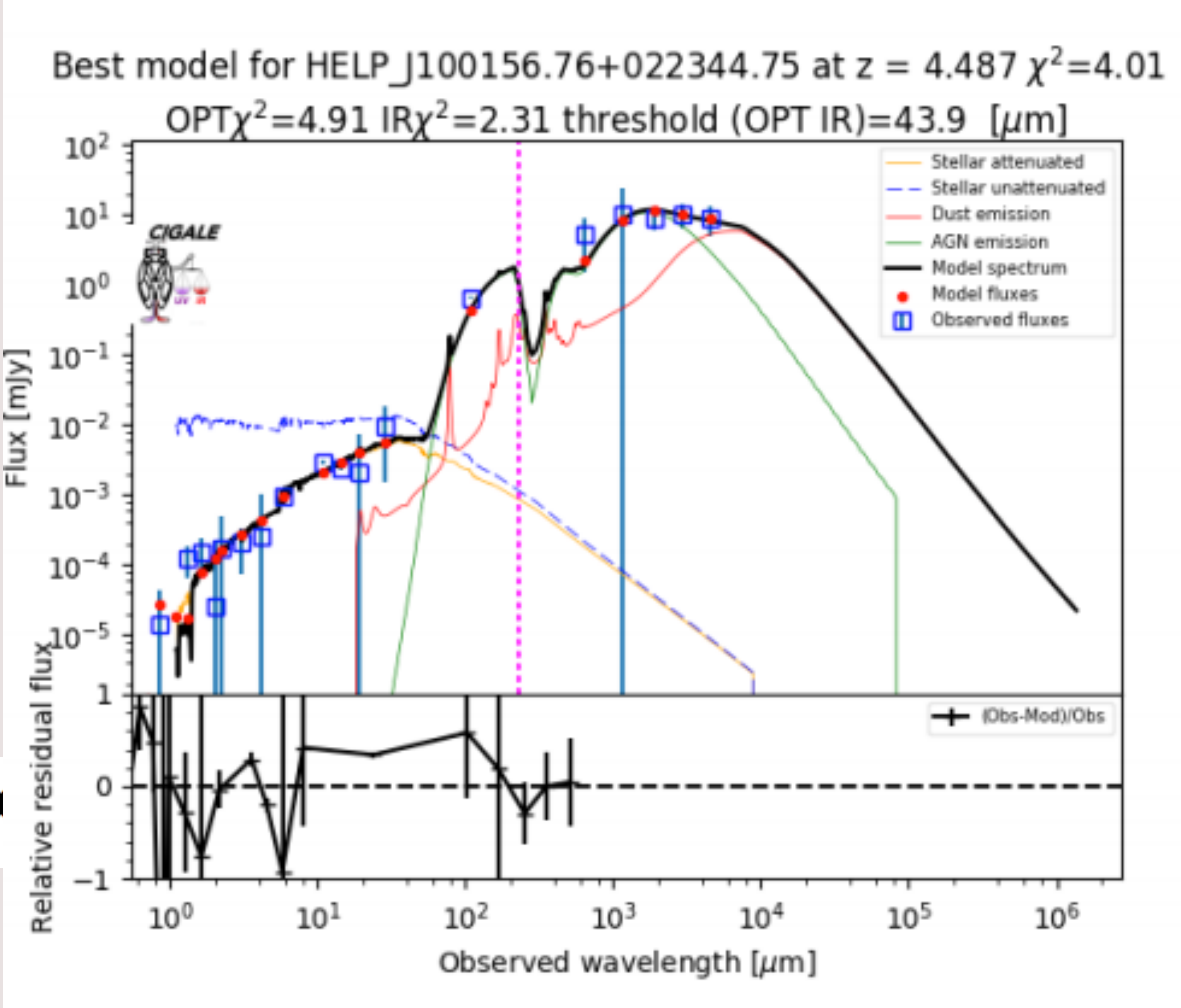
Duncan et al. 2018
Almosallam et al. 2016

- 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

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

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

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



Best model for HELP_J100156.76+022344.75 at z = 4.487 χ²=4.01
OPTχ²=4.91 IRχ²=2.31 threshold (OPT IR)=43.9 [μm]
CIGALE
Stellar attenuated
Stellar unattenuated
Dust emission
AGN emission
Model spectrum
Model fluxes
Observed fluxes
(Obs-Mod)/Obs
Flux [mJy]
Relative residual flux
Observed wavelength [μm]

AstroCast

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

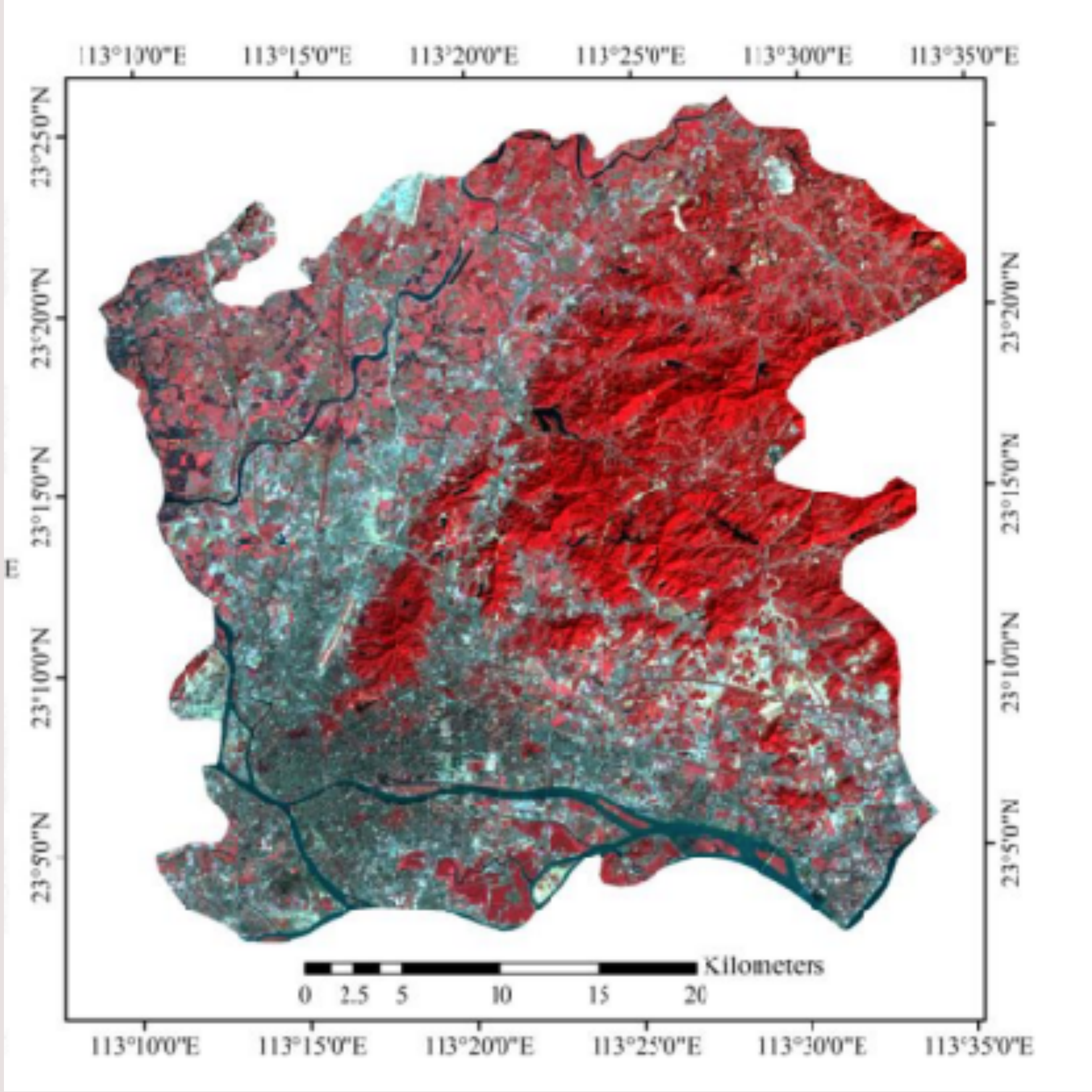


Table 1. Land use classification system.

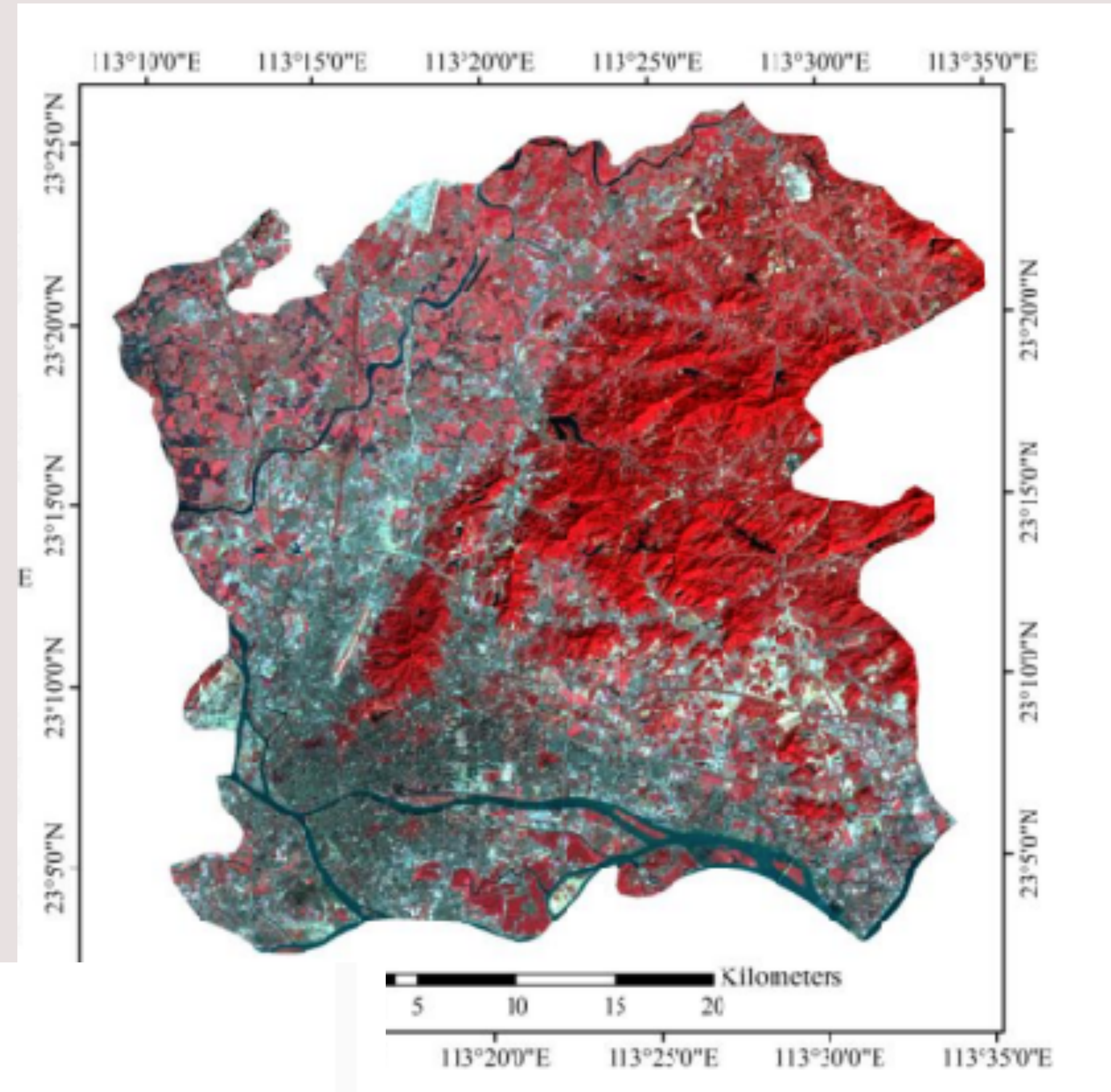
Land-Use Types	Description
Water	Water bodies such as reservoirs, ponds and river
Residential area	Residential areas where driveways and roof tops dominate
Natural forest	Large area of trees
Orchard	Large area of fruit trees planted
Farmland	Fields where vegetables or crops grow
Industrial/commercial	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

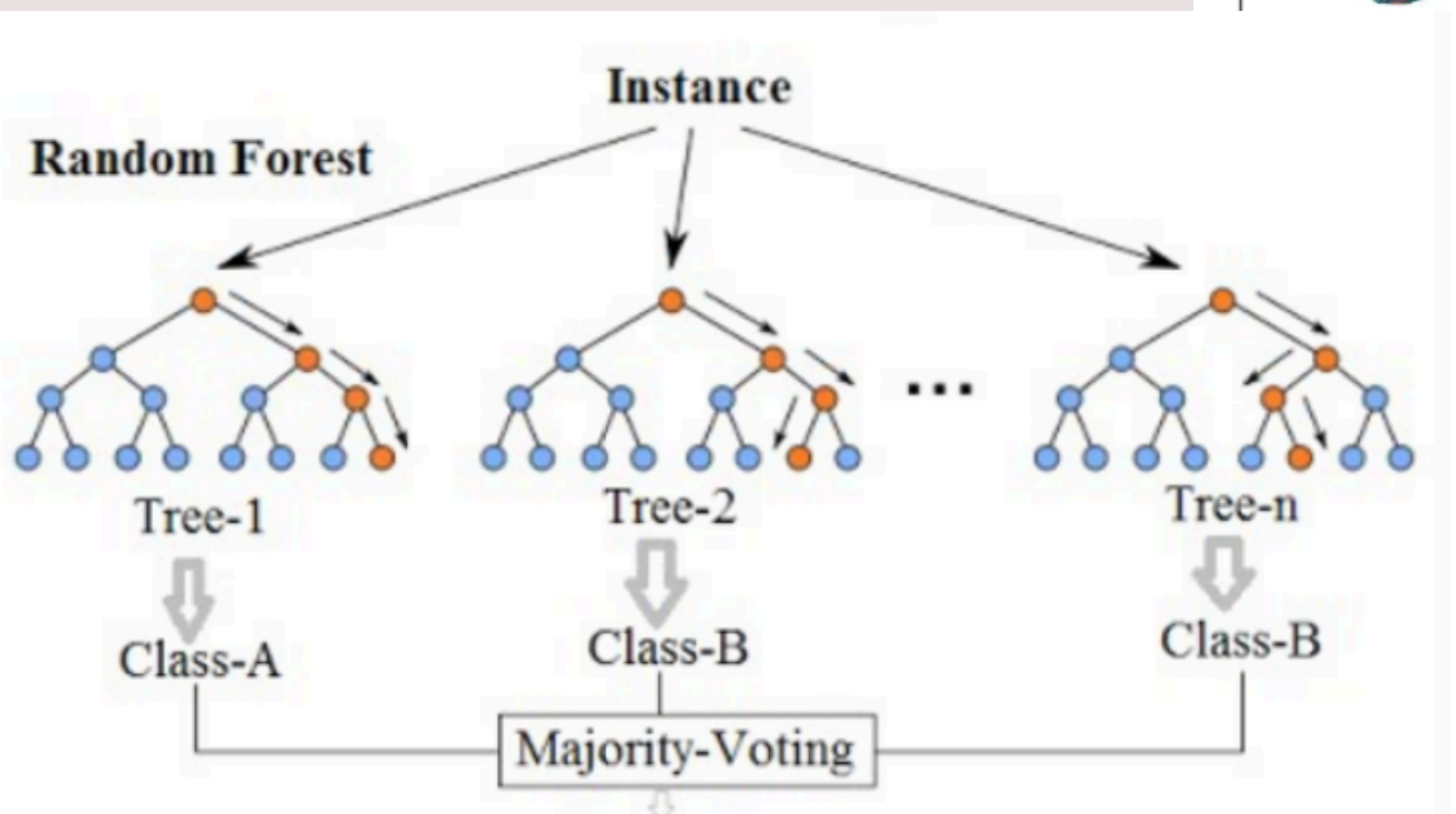
Li et al. 2014

GCRF
Global Challenges Research Fund

Science & Technology Facilities Council

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

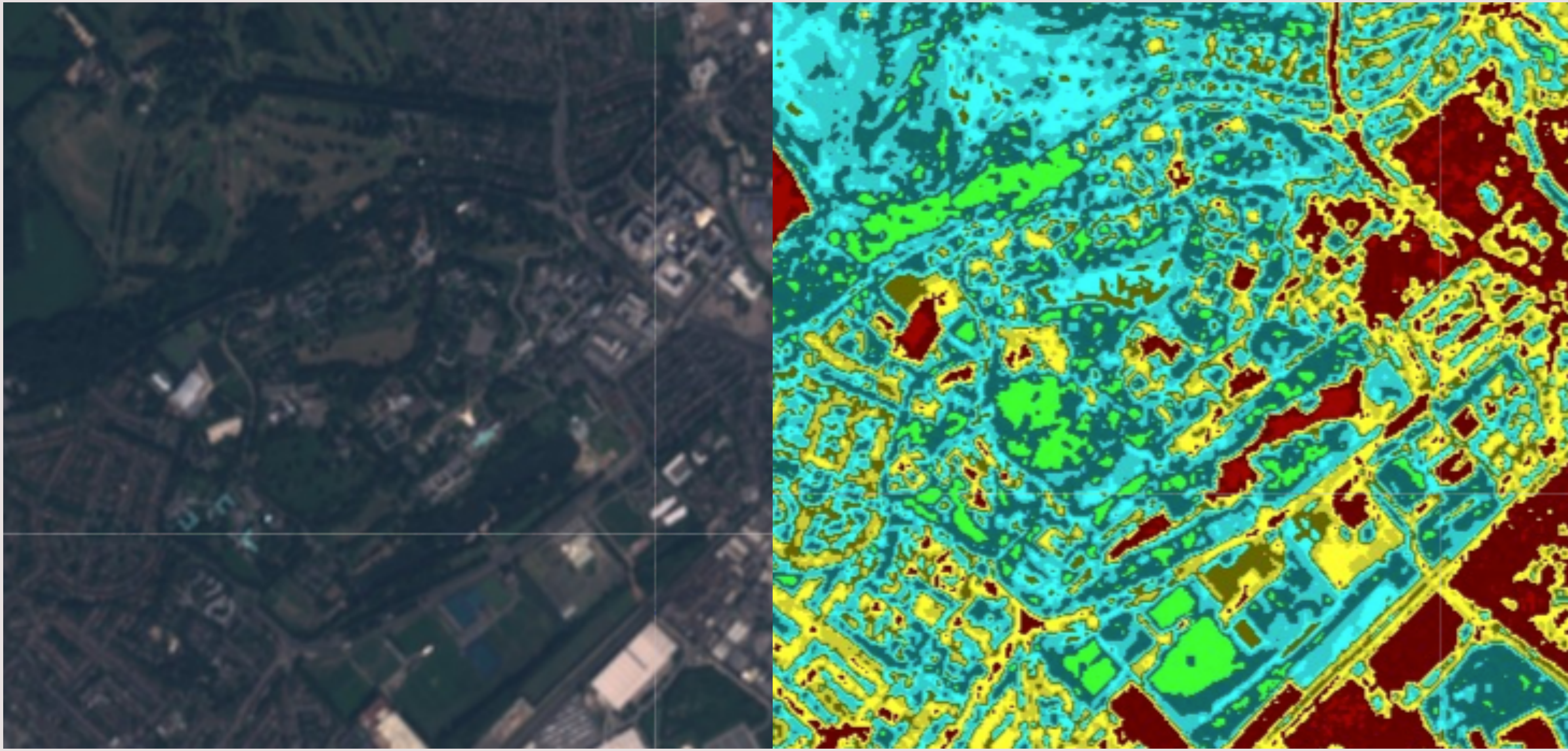




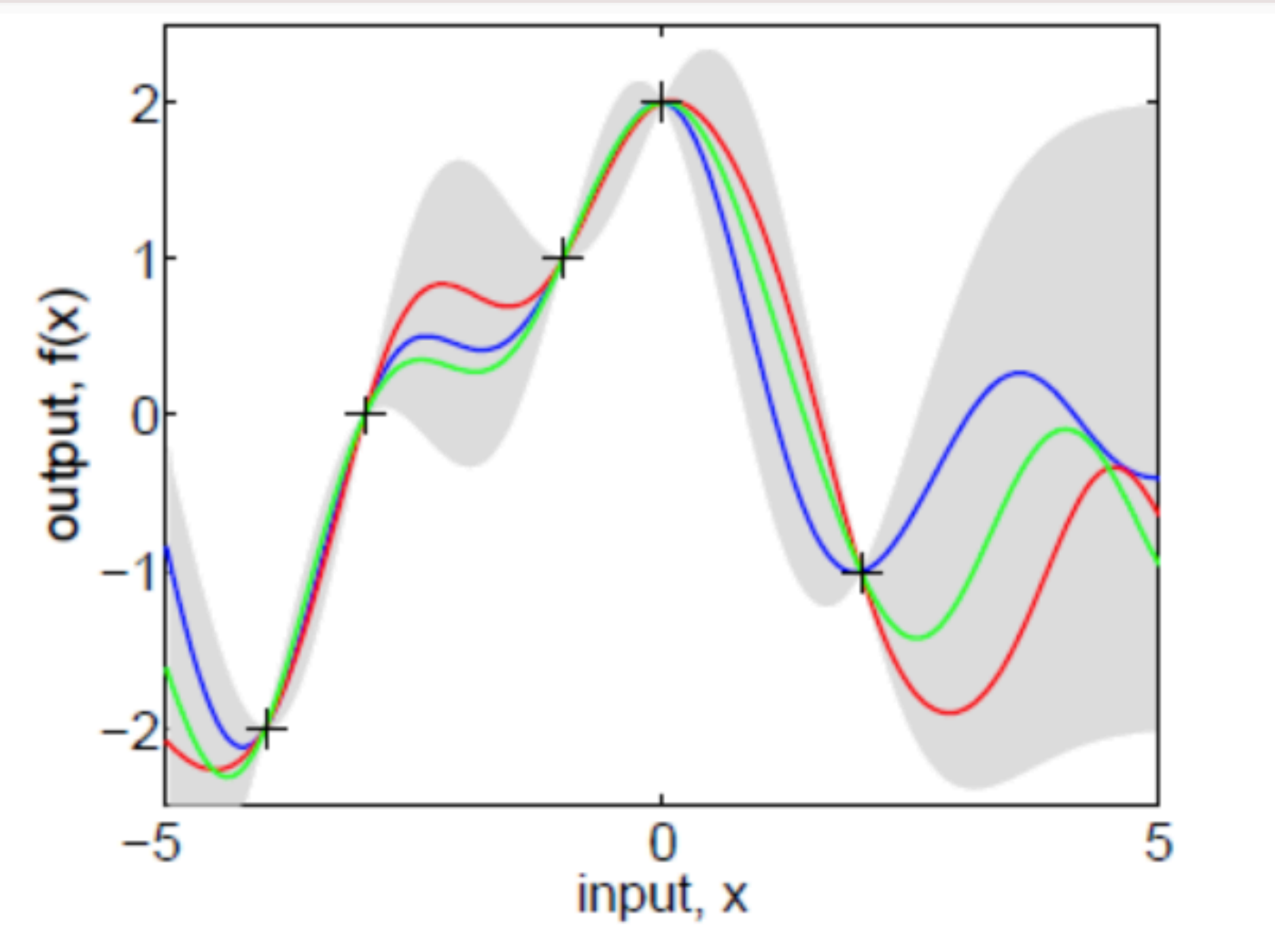
nate

ruction is underway

AstroCast



AstroCast



- GP
- scikit learn
- RBF and periodic kernel

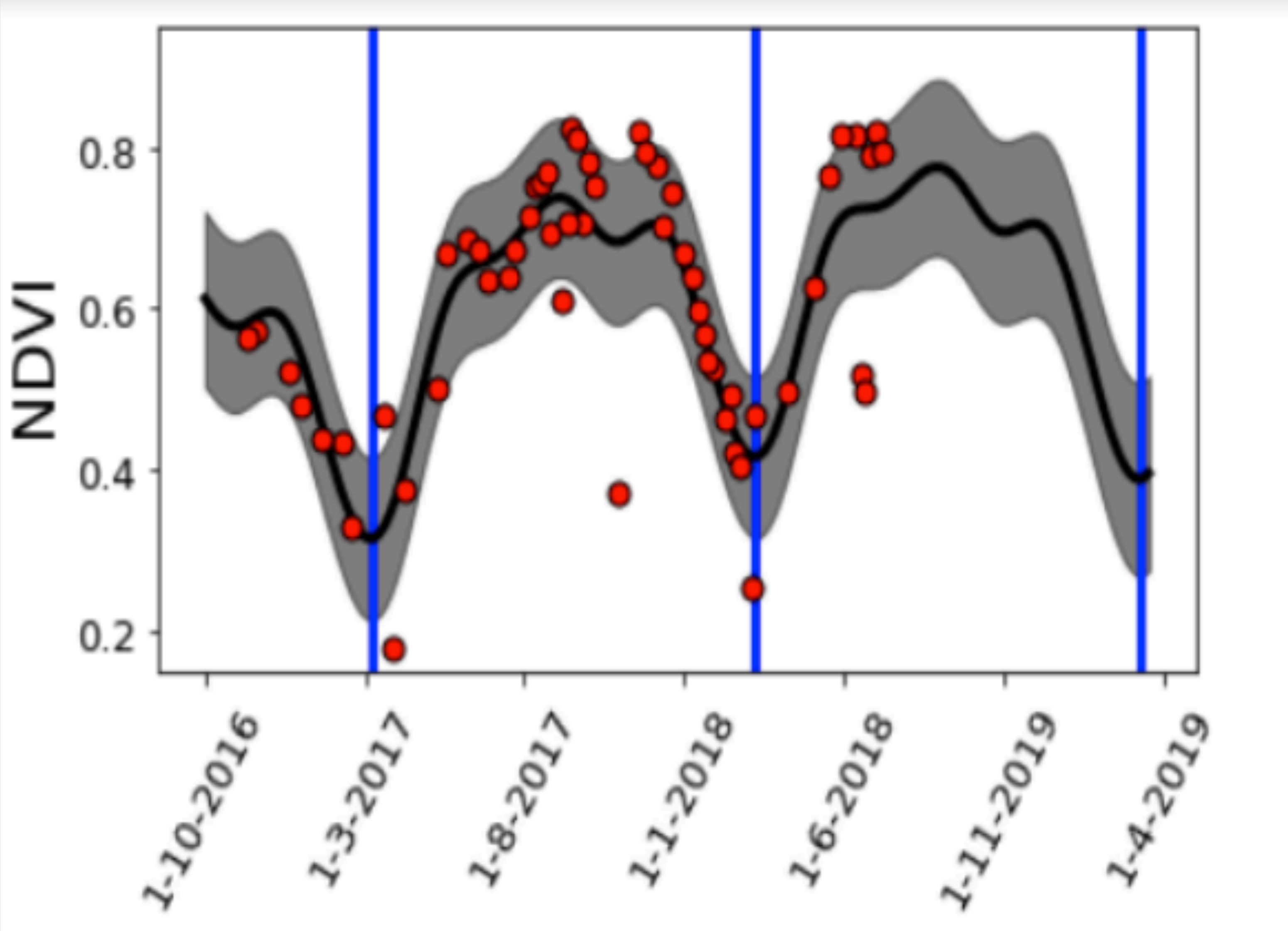
AstroCast

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

```python
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



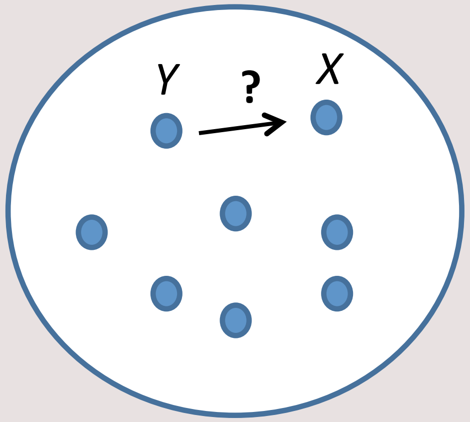
- GP
- scikit learn
- RBF and periodic kernel

- Include spatial information?

Global Challenges Research Fund

Science & Technology Facilities Council

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*

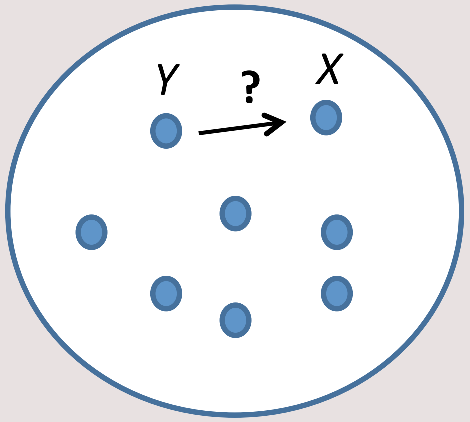
$$X_1(t) = \sum_{j=1}^{p} A_{11,j} X_1(t-j) + \sum_{j=1}^{p} A_{12,j} X_2(t-j) + E_1(t)$$

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

GCRF Global Challenges Research Fund

Science & Technology Facilities Council

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
- github.com/H-E-L-P/dmu_products
- hedam.lam.fr/HELP