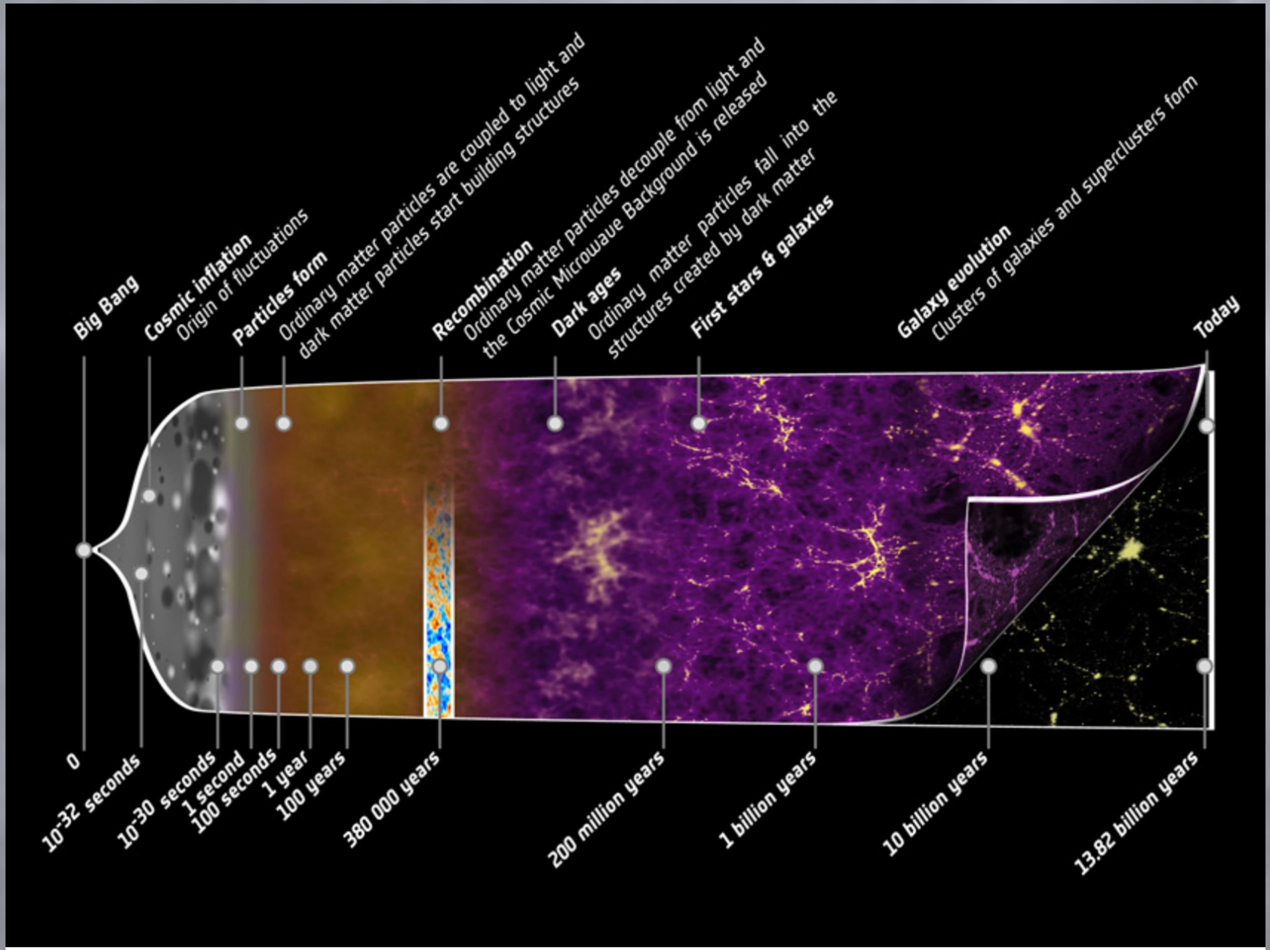
Neural Networks for Cosmic Microwave Background Radiation Recovery

Laura Bonavera

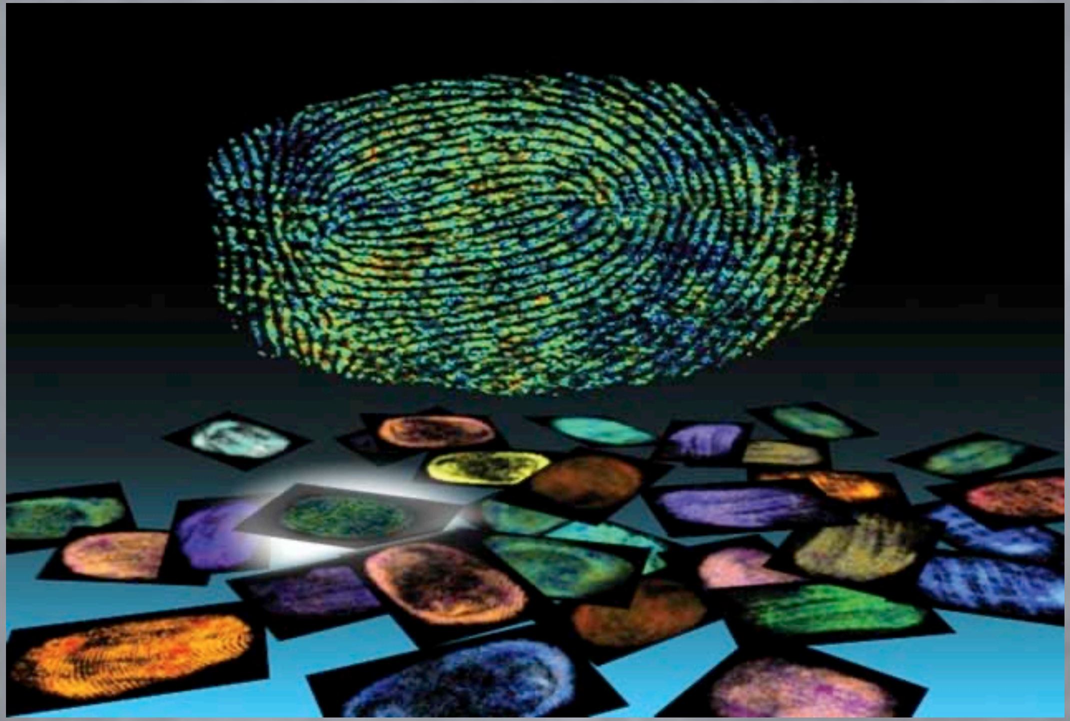
ICTEA – University of Oviedo

Cosmic Microwave Background

The remnant electromagnetic radiation from the Big Bang, which we can still see 13.8 billion years later



The history of the Universe, from the Big Bang to today. *Image credit: ESA*



Cosmic Microwave Background

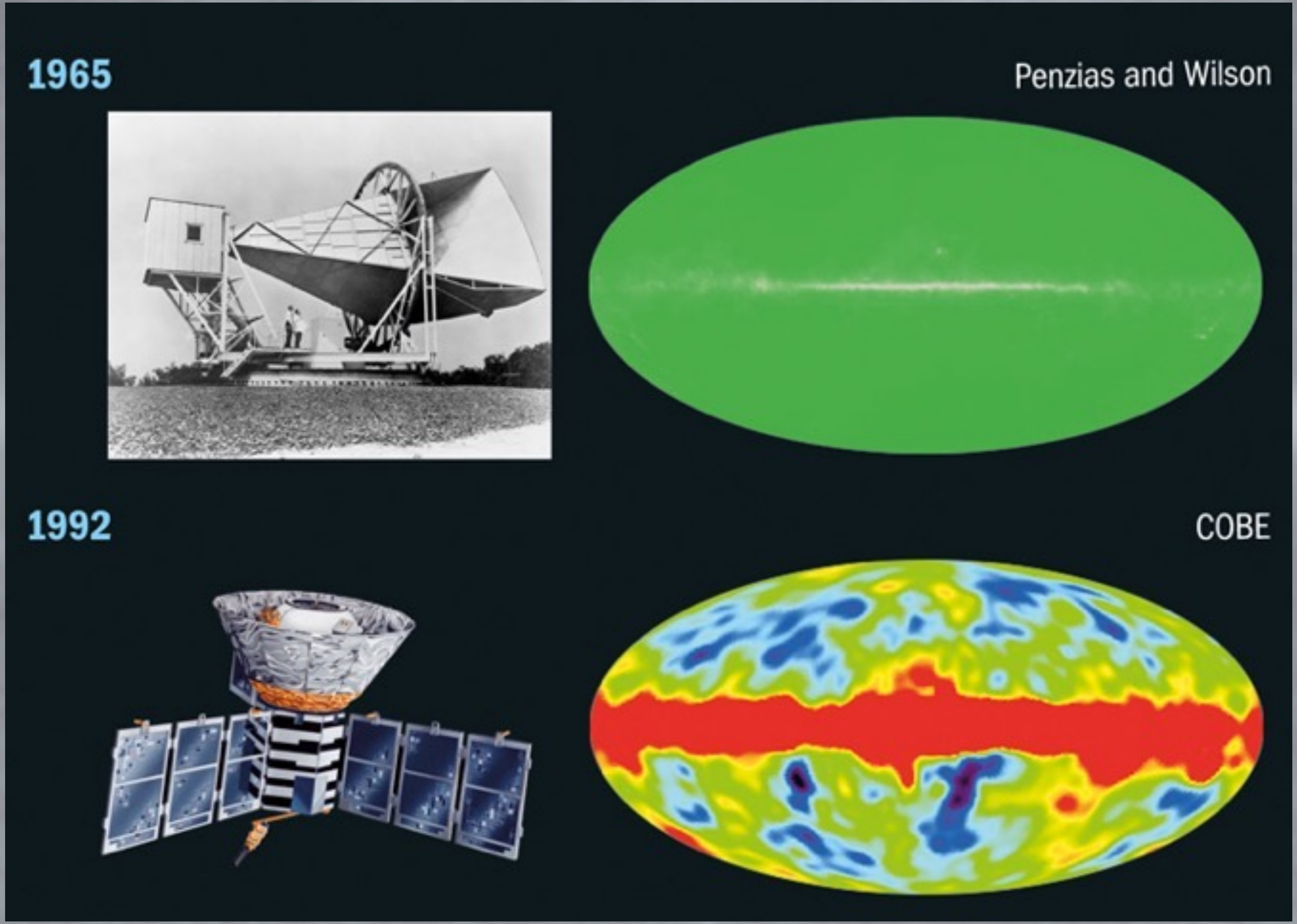
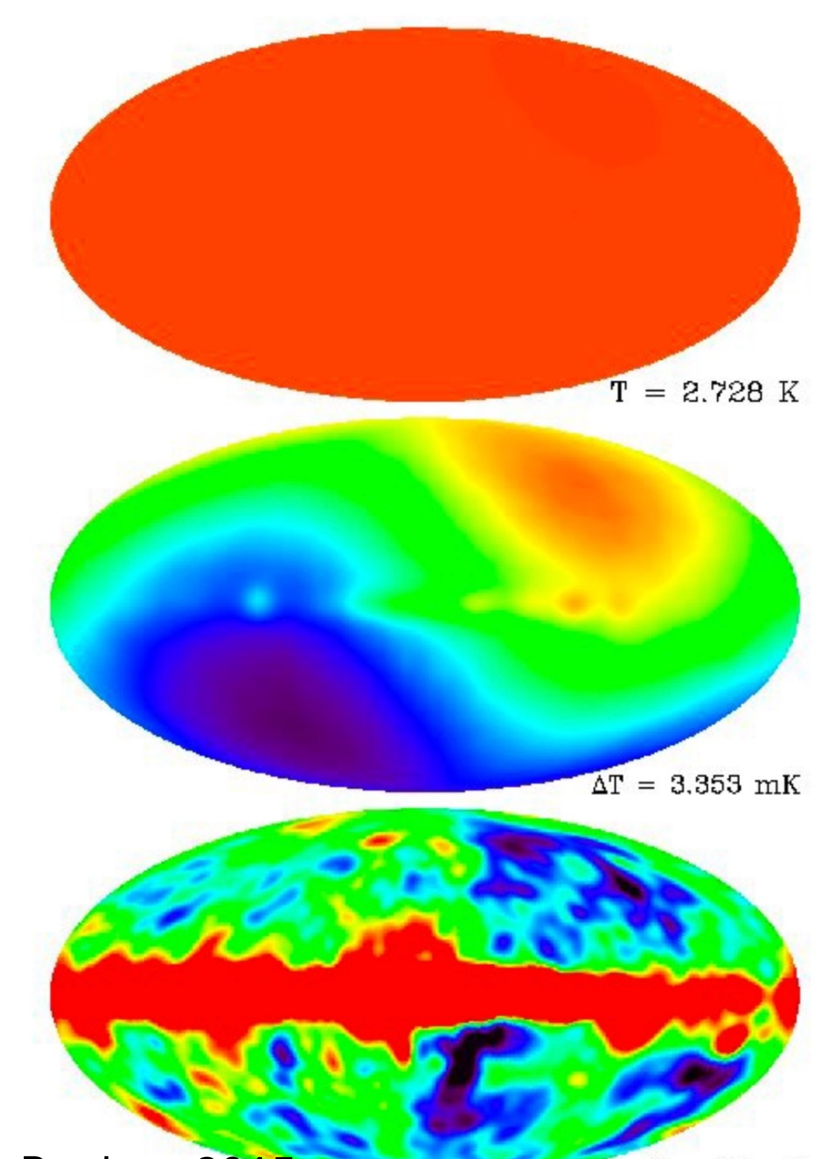
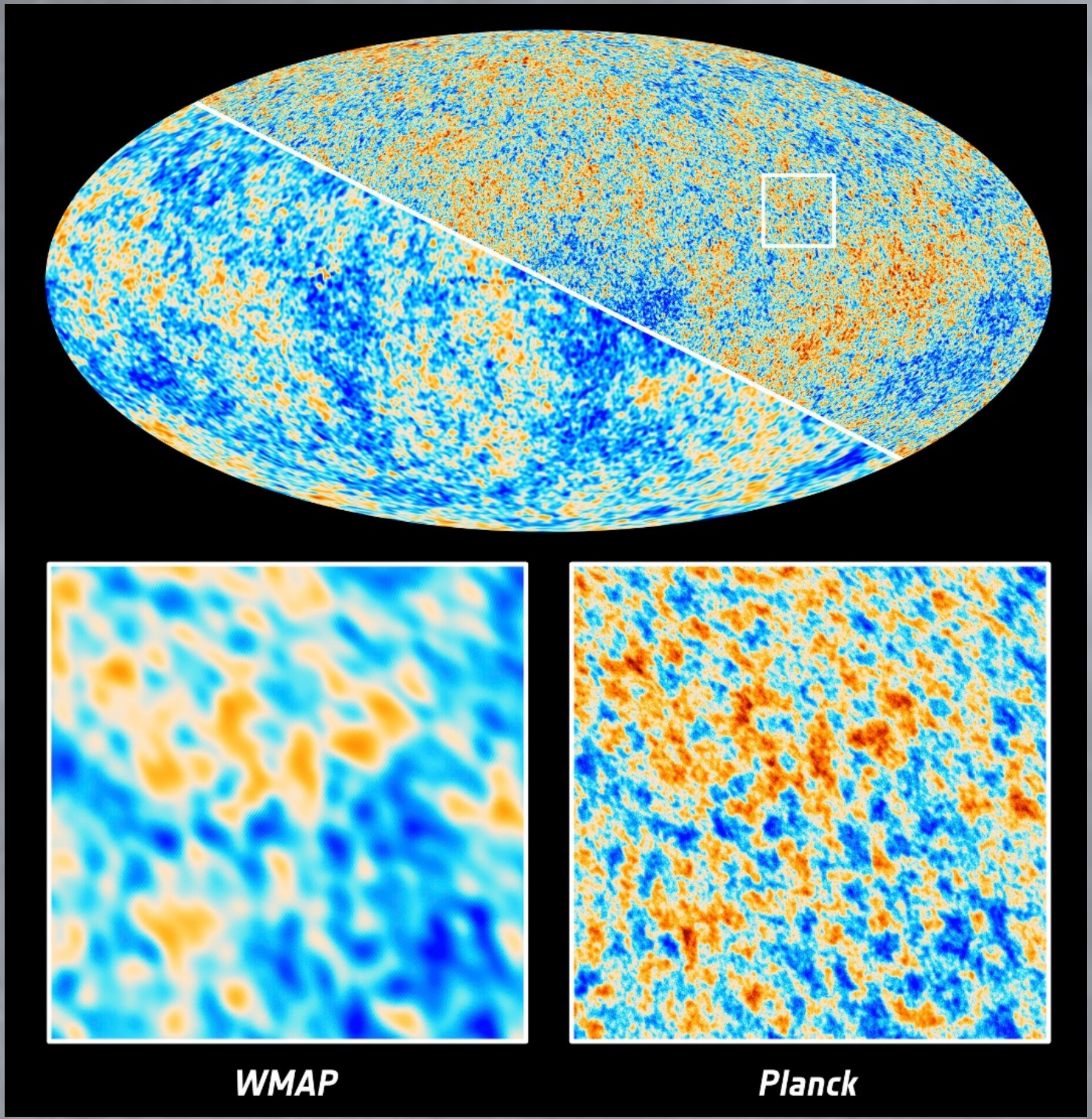


Image credit: NASA



Bucher+2015

Cosmic Microwave Background



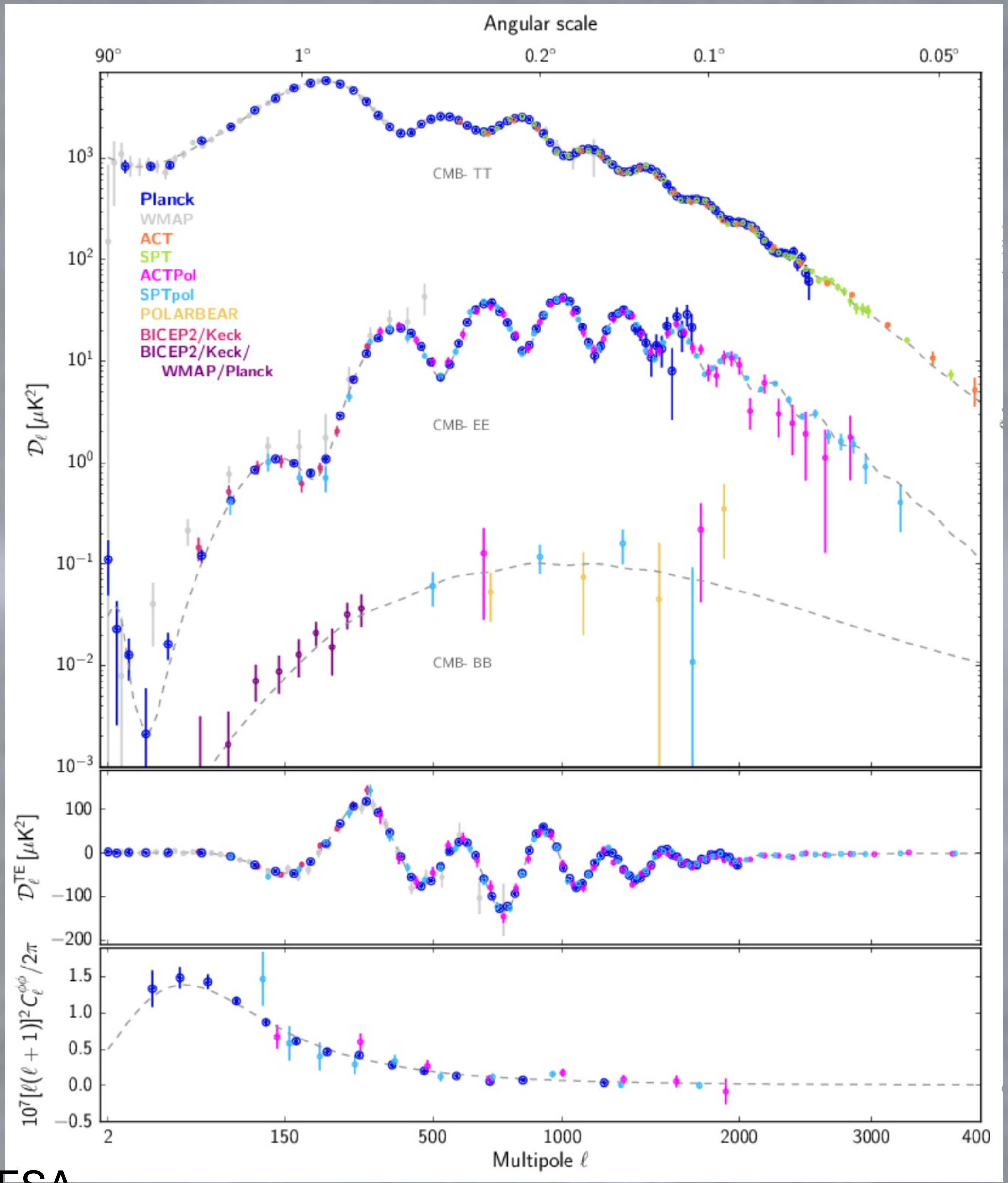


Image credit: ESA

Cosmic Microwave Background

Anisotropy Formalism

- Fluctuations of CMB temperature in different directions on the sky:

$$\frac{(\Delta T)}{T}(\theta,\varphi) = \frac{(T(\theta,\varphi) - T_0)}{T_0}$$

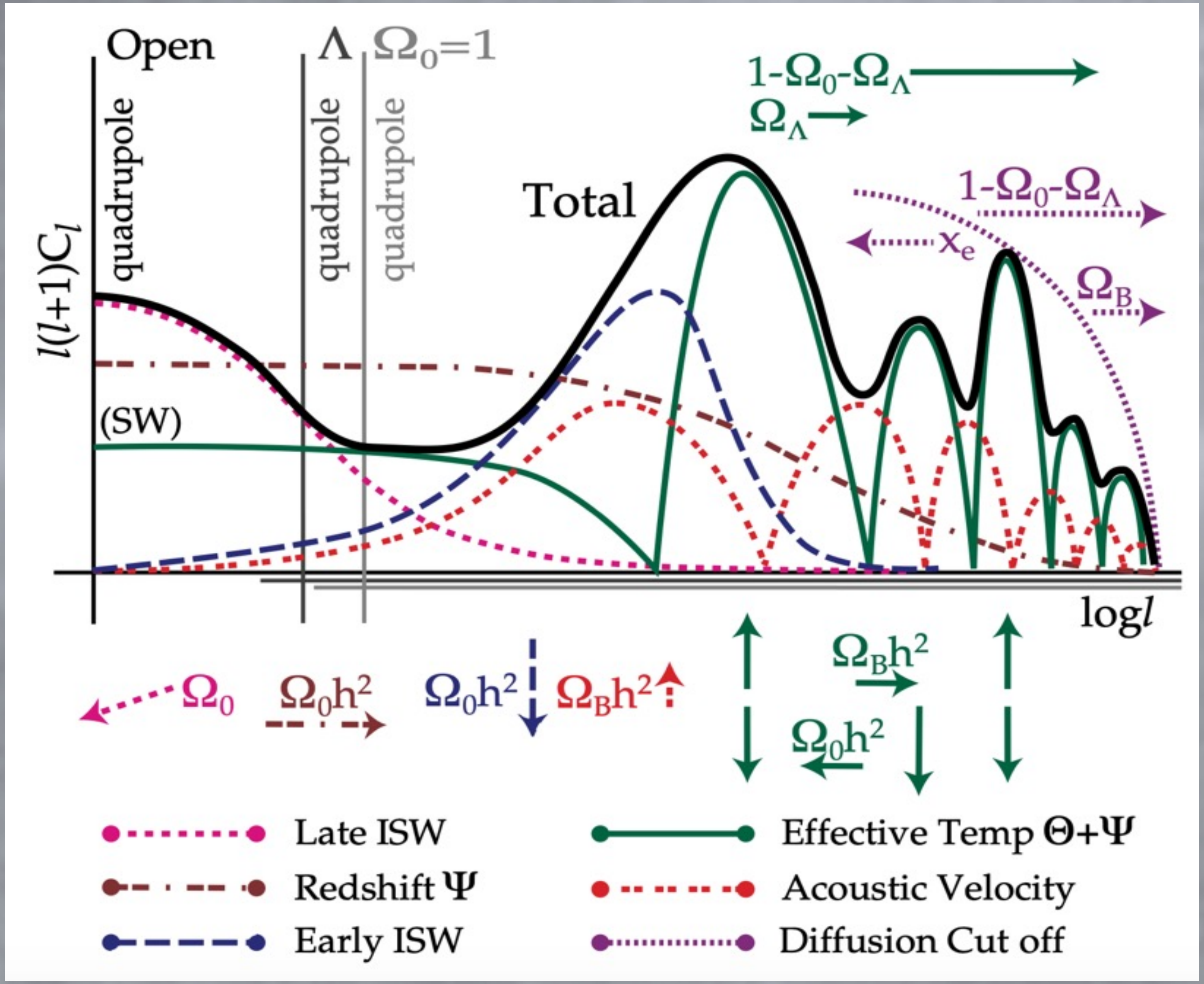
- Spherical harmonic decomposition:

$$\frac{(\Delta T)}{T}(\theta,\varphi) = \sum_{l,m} a_{lm} Y_{lm}(\theta,\varphi) \qquad a_{lm} = \int_{\Omega} \frac{(\Delta T)}{T} \hat{Y}_{lm}(\theta,\varphi) d\vec{\Omega}$$

- We define the angular power spectrum as:

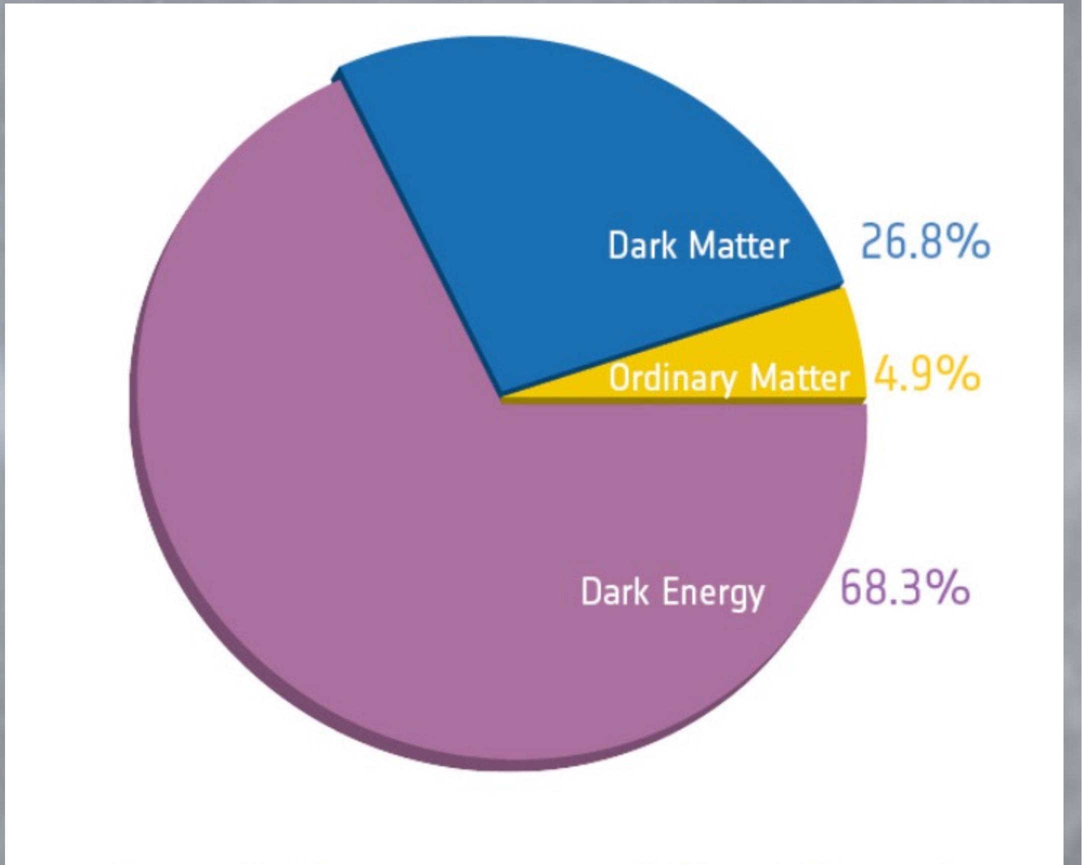
$$C_l \equiv a_l^2 = \langle | \, a_{lm}^2 \, | \rangle$$

Cosmic Microwave Background



Hu, Sugiyama & Silk (1995)

The universe today

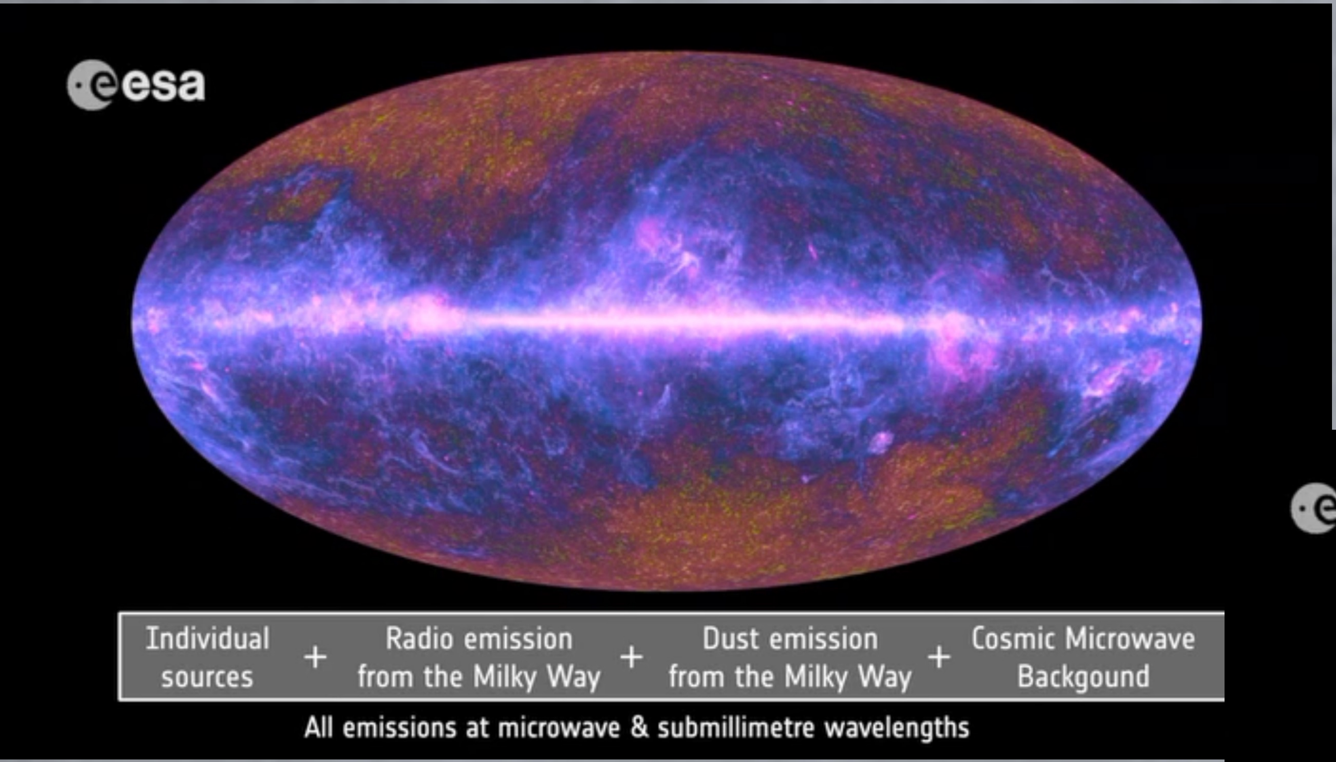


The relative amounts of the different constituents of the Universe. *Image credit: ESA/Planck*

Component Separation

- CMB importance in understanding the Universe
- Precision in CMB anisotropy measurements, especially in polarization
- Characterize and subtract contaminants that “hide” the CMB signal:
    - Diffuse emissions in the sky
    - Extragalactic point sources (PS)
- Develop new methods looking for better performance

Component Separation



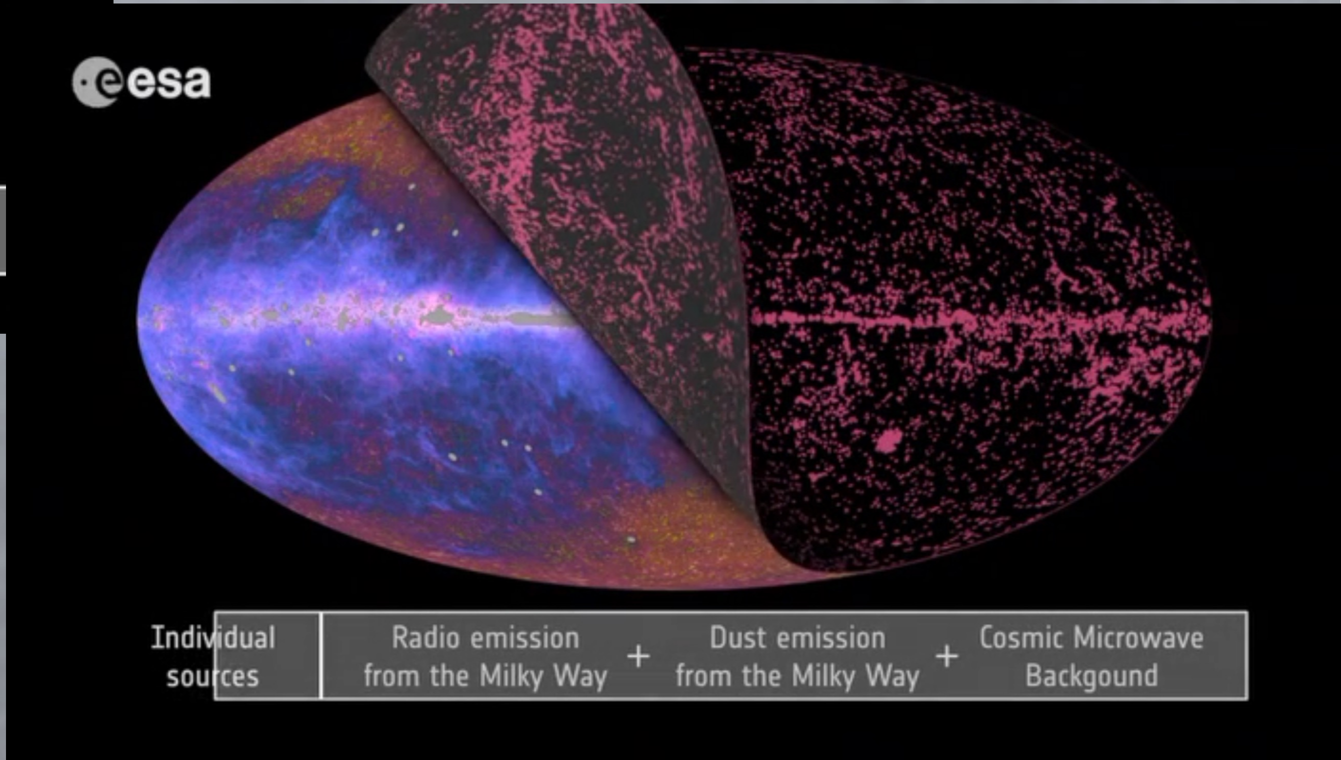
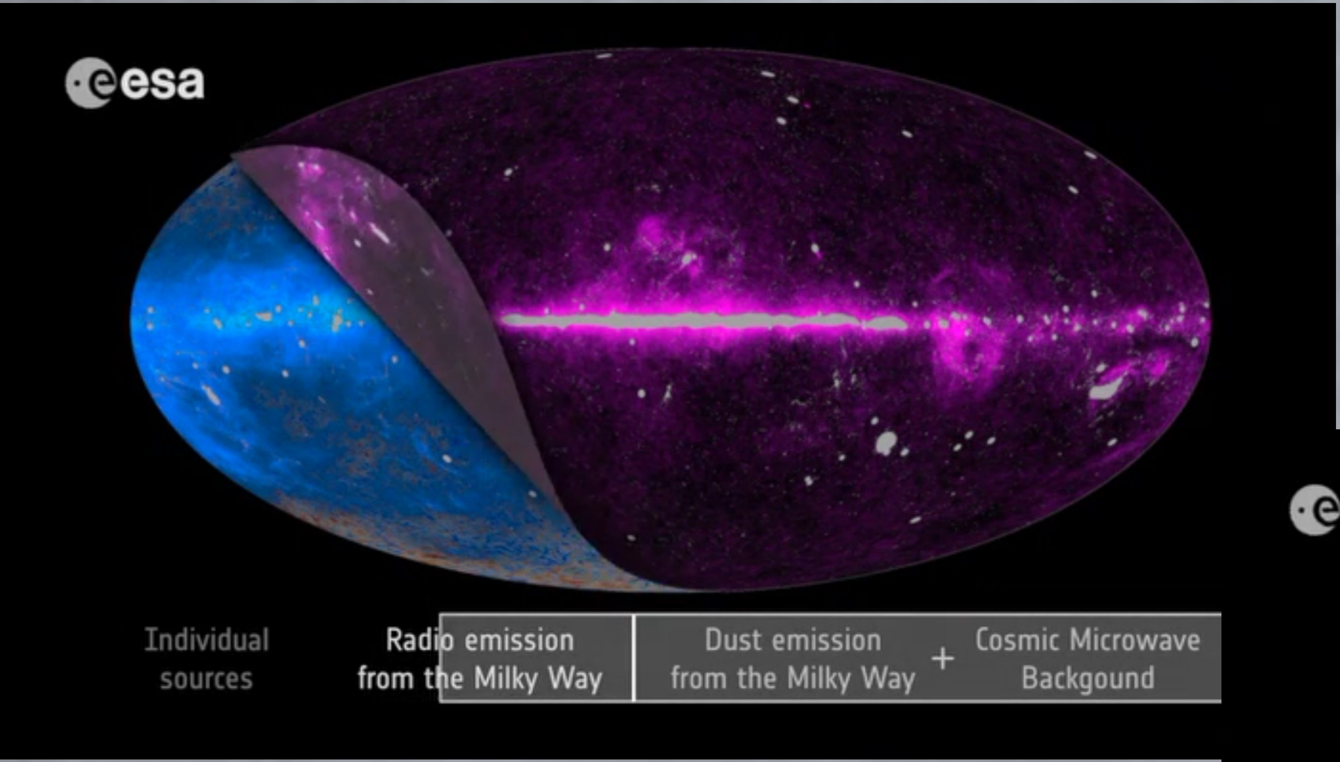


Image credit: ESA

Component Separation



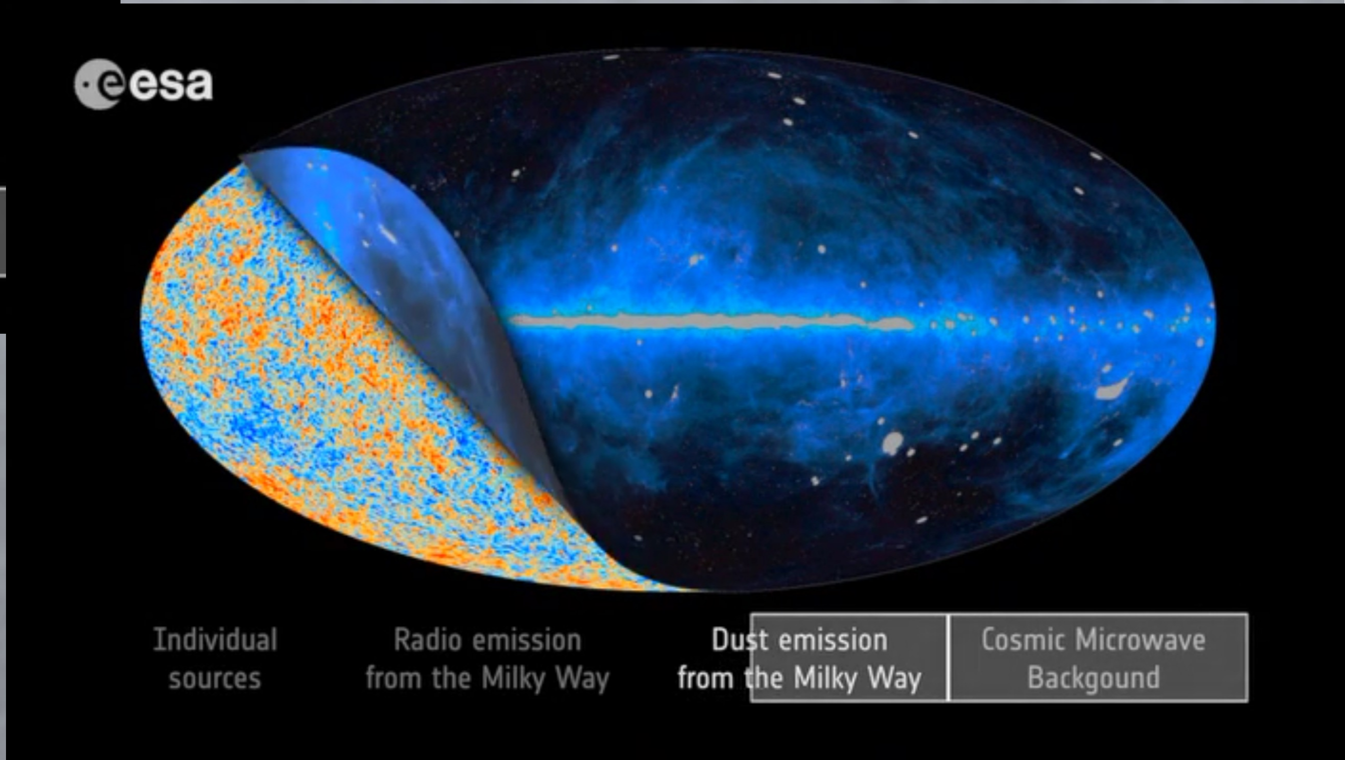
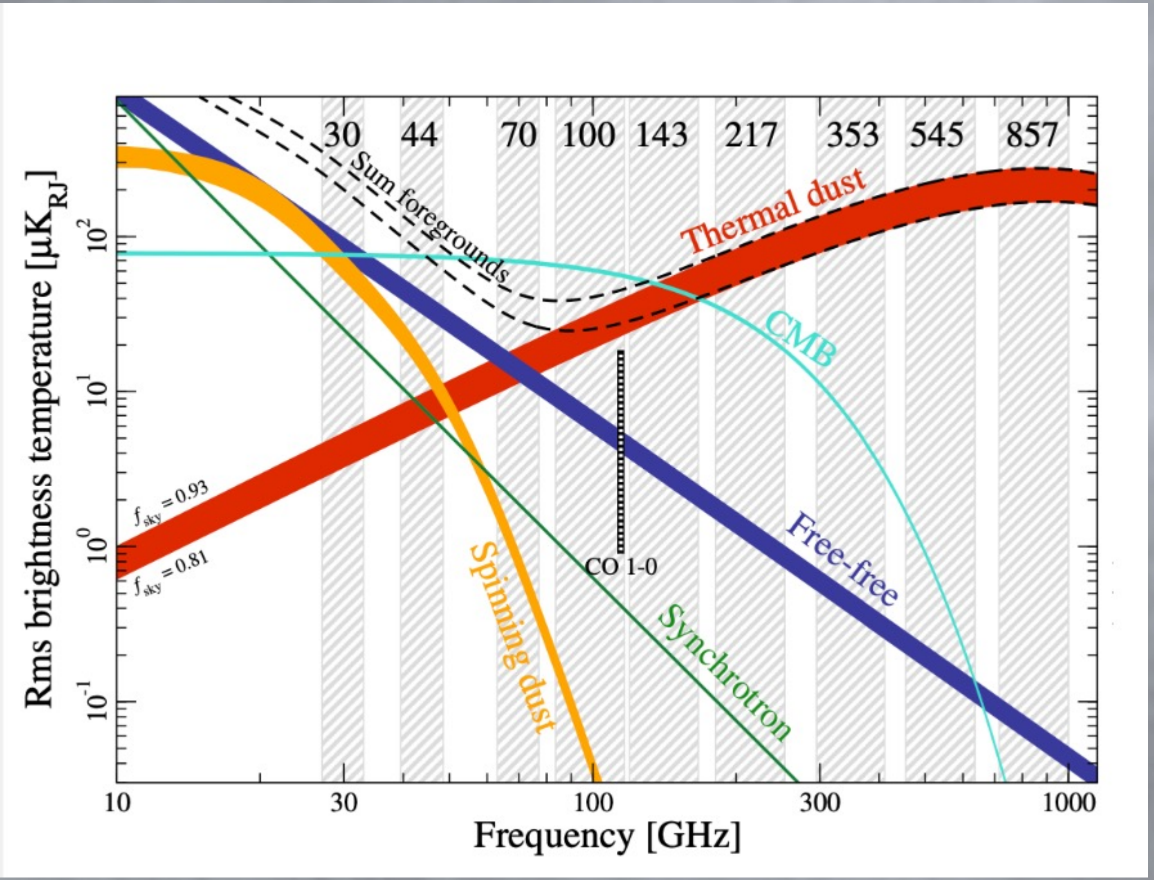
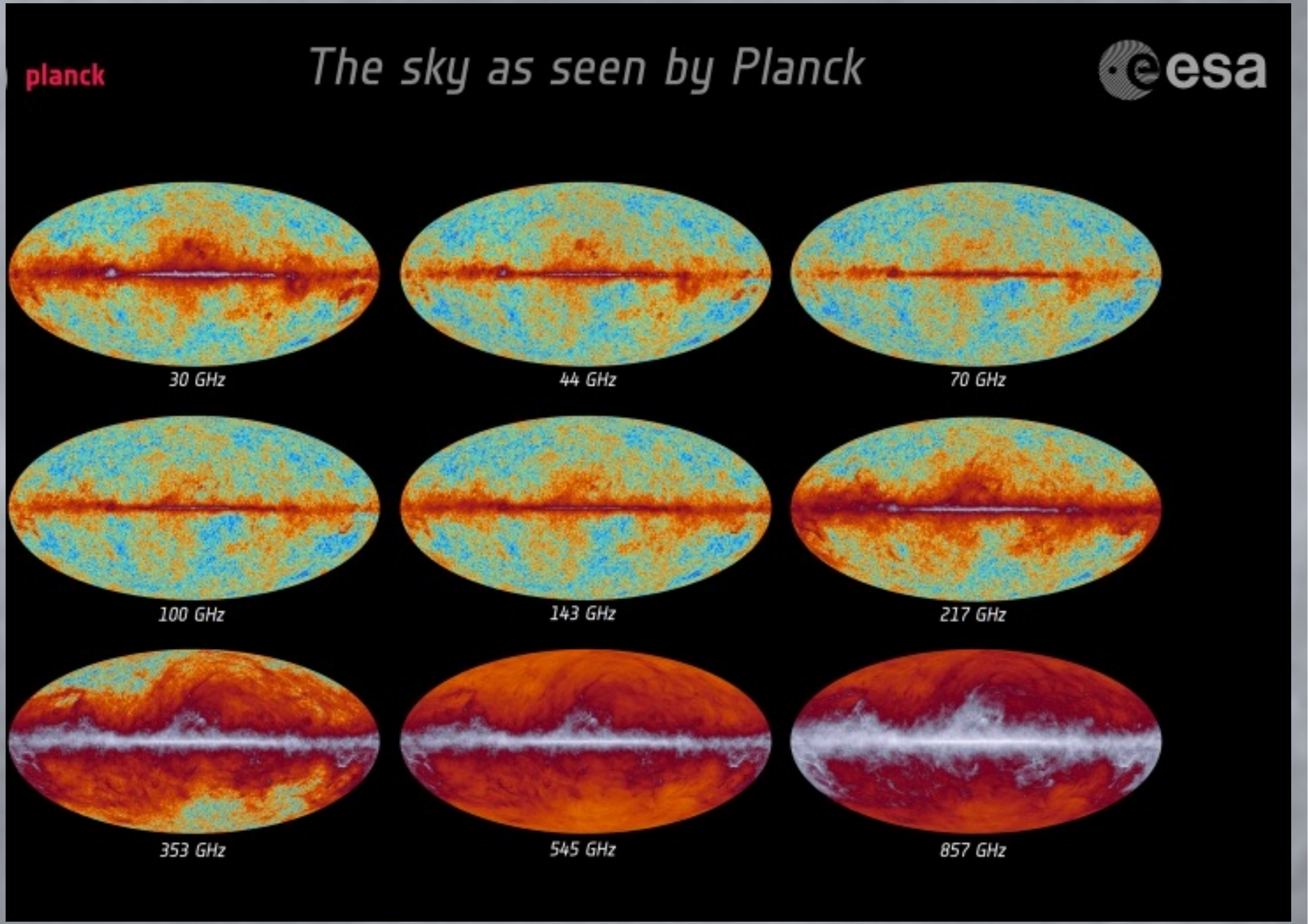


Image credit: ESA

Component Separation

Total intensity

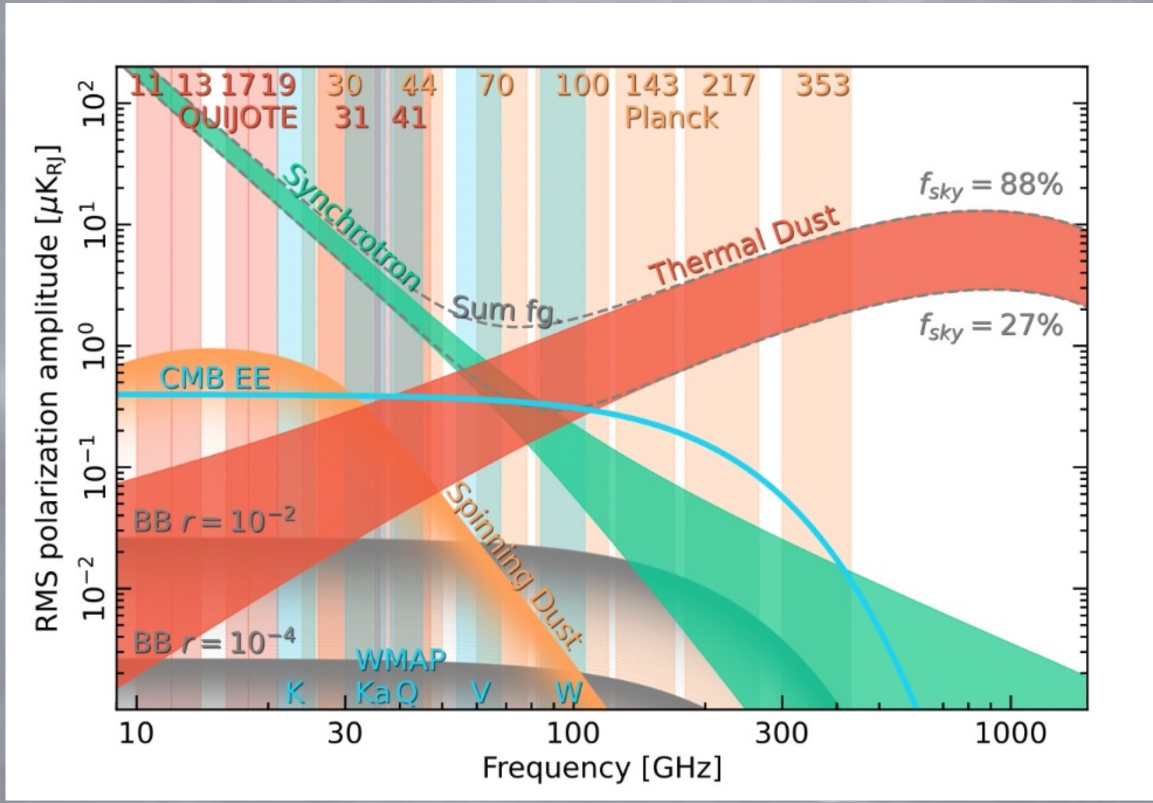




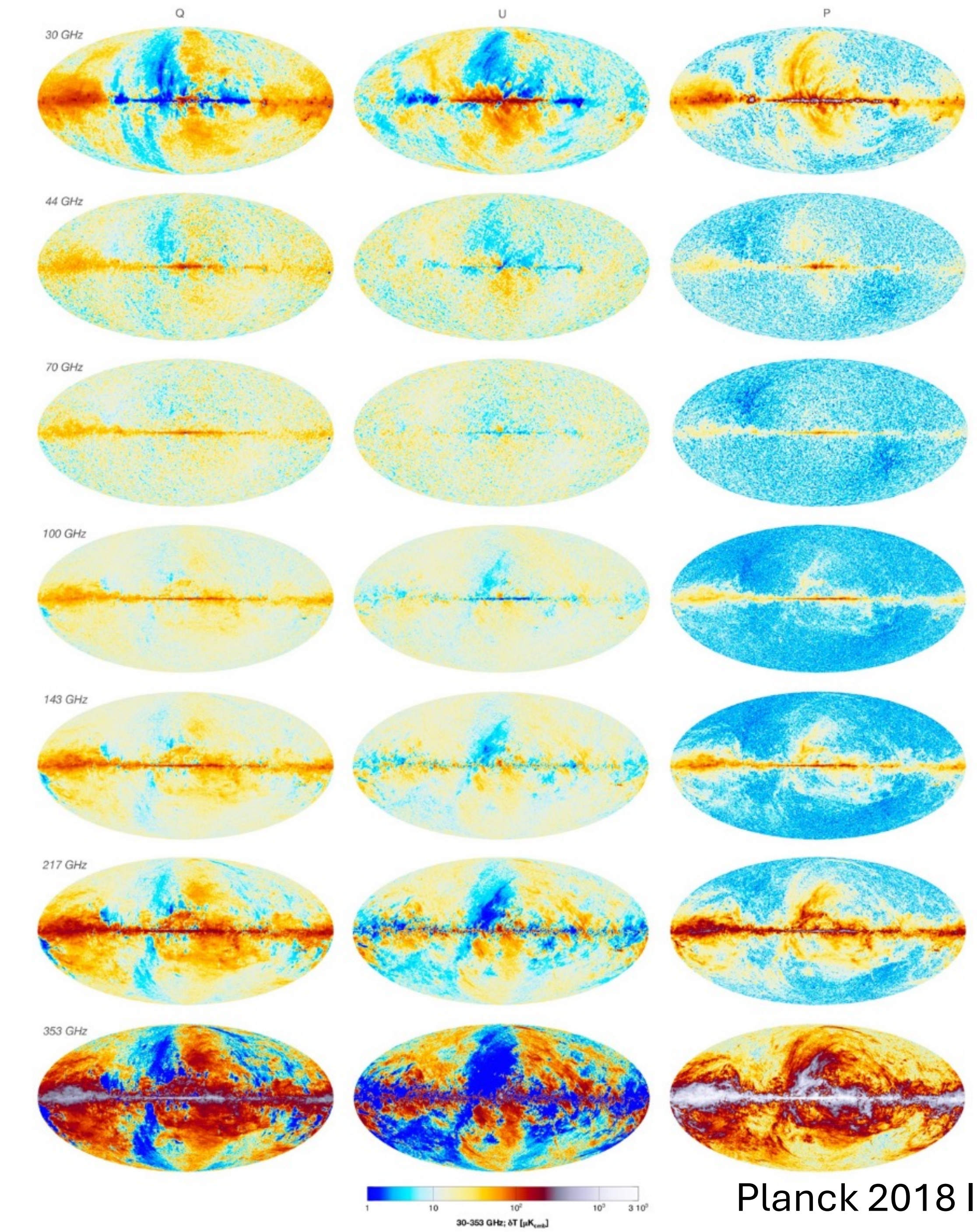
Planck 2018 I

Component Separation

Polarization



Beyond Planck 2023 XIV

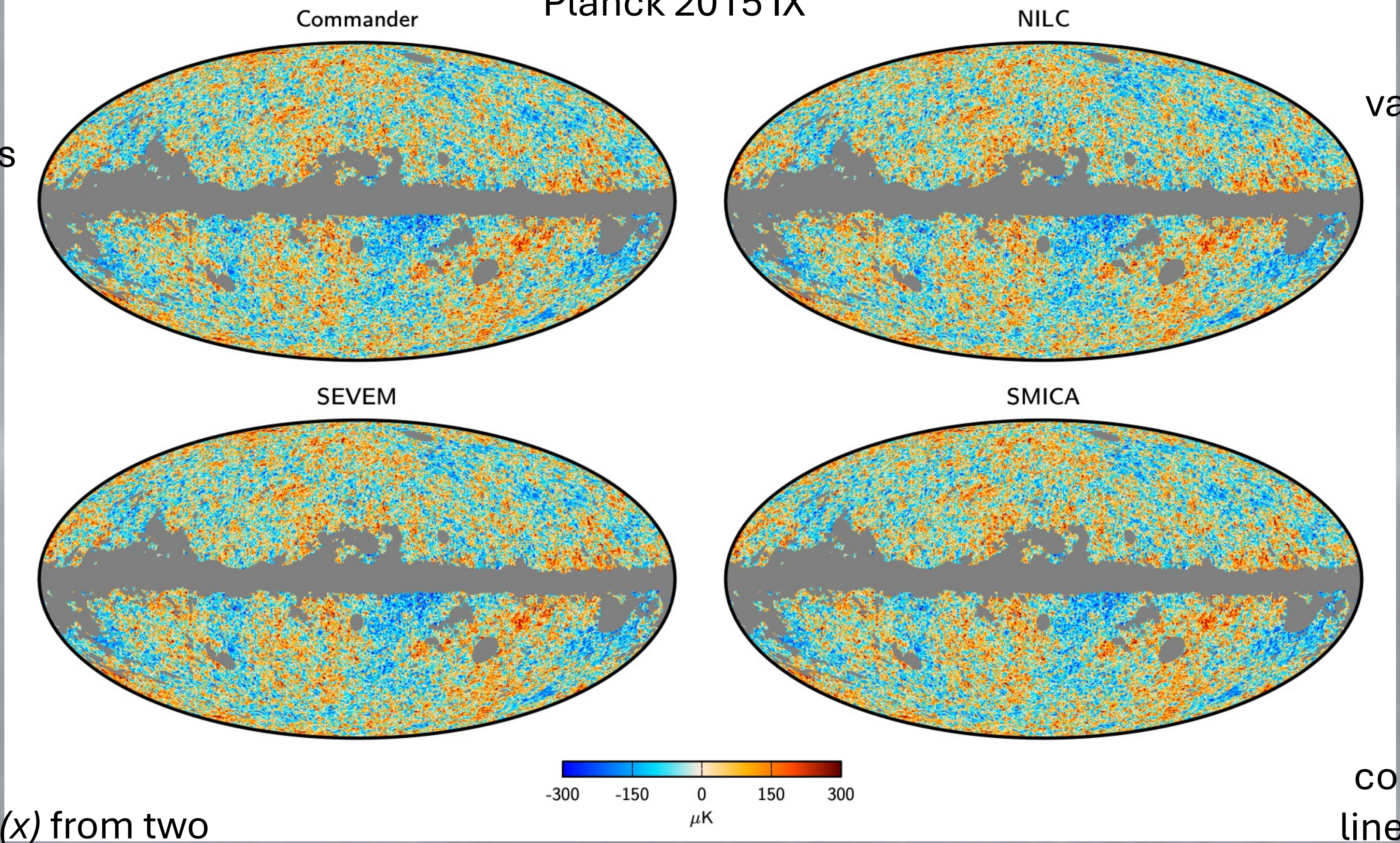


Planck 2018 I

Component Separation

Fitting model to set of obs w/i standard Bayesian parametric framework (parameters and priors)

Planck 2015 IX



Linear combination of maps w/ minimum variance using a basis of spherical wavelts (needlets)

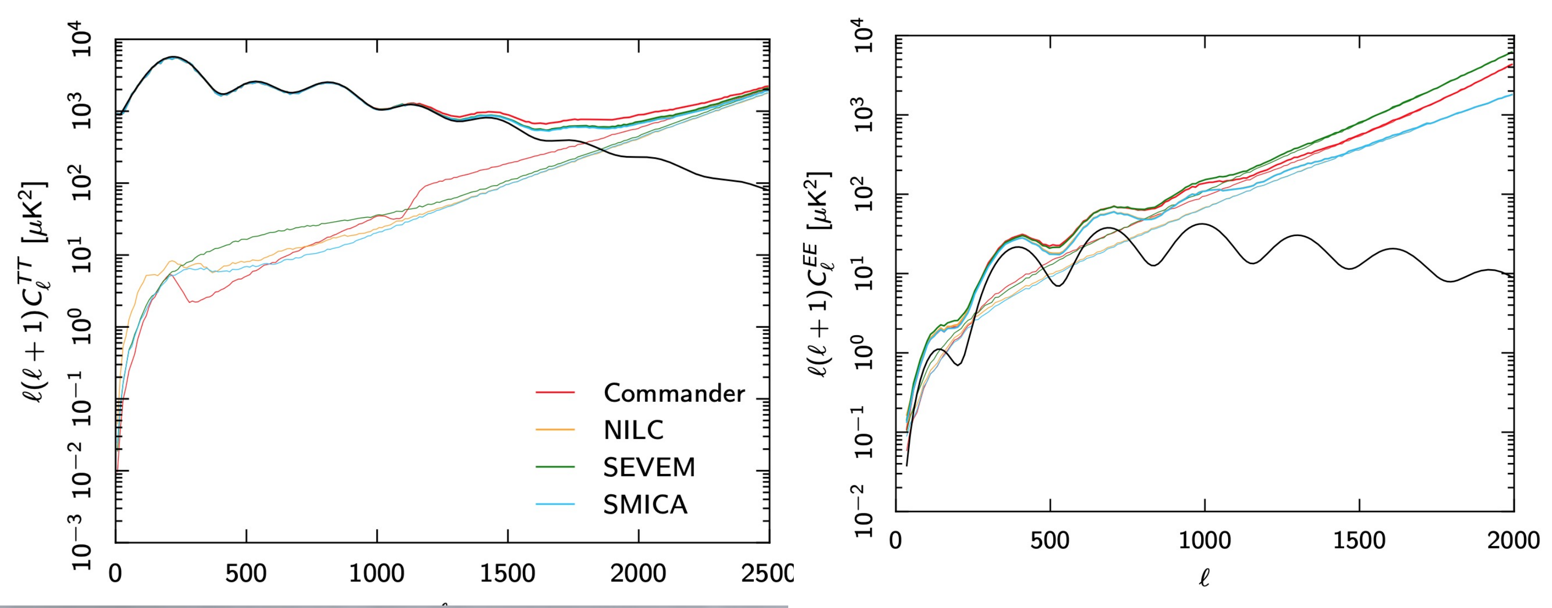
Construct templates $t_j(x)$ from two close channels, α_j by minimizing the variance outside mask

$$T_c(x,v) = d(x,v) - \sum_{j=1}^{n_t} \alpha_j \, t_j\,(x)$$

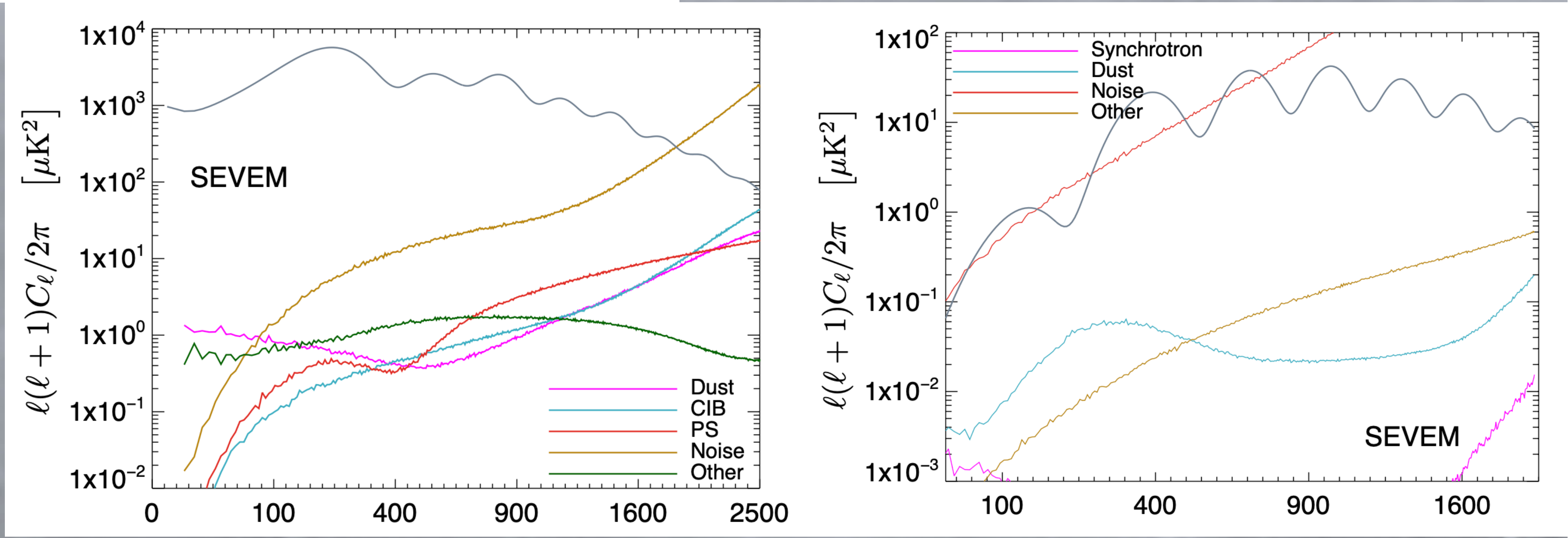
CMB harmonic coefficients by weighted linear combination of the input maps harmonic coefficients

$$s_{lm} = w_l^T x_{lm}$$

Component Separation



FFP8 simulations

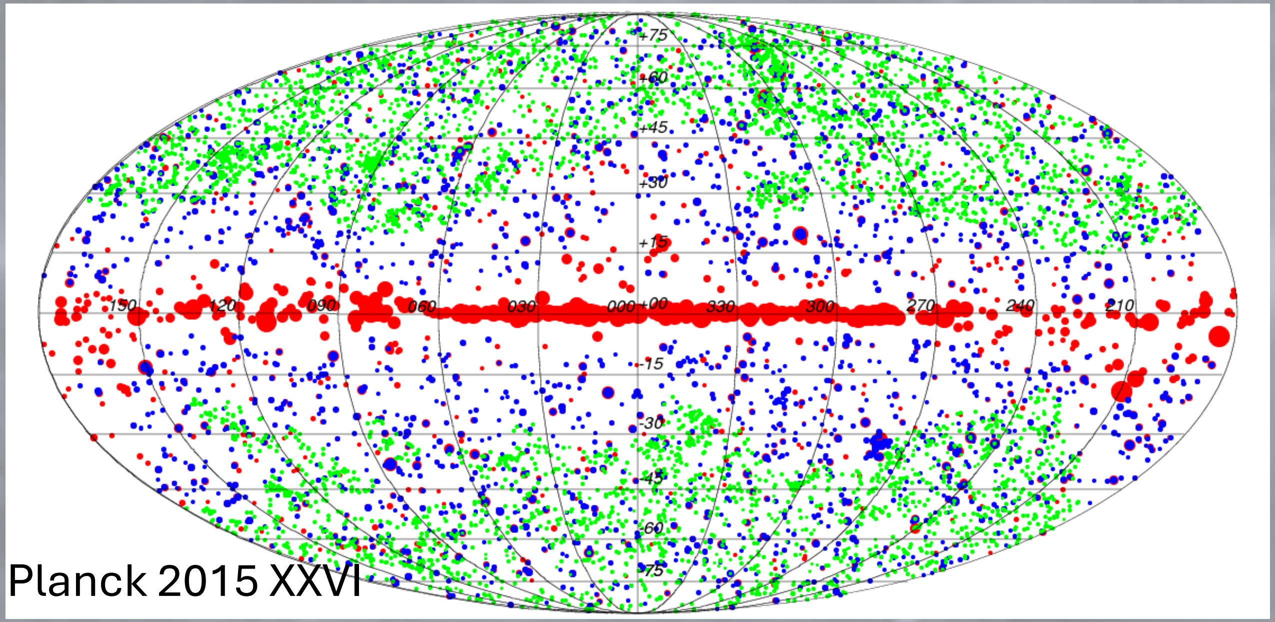


Planck 2015 IX

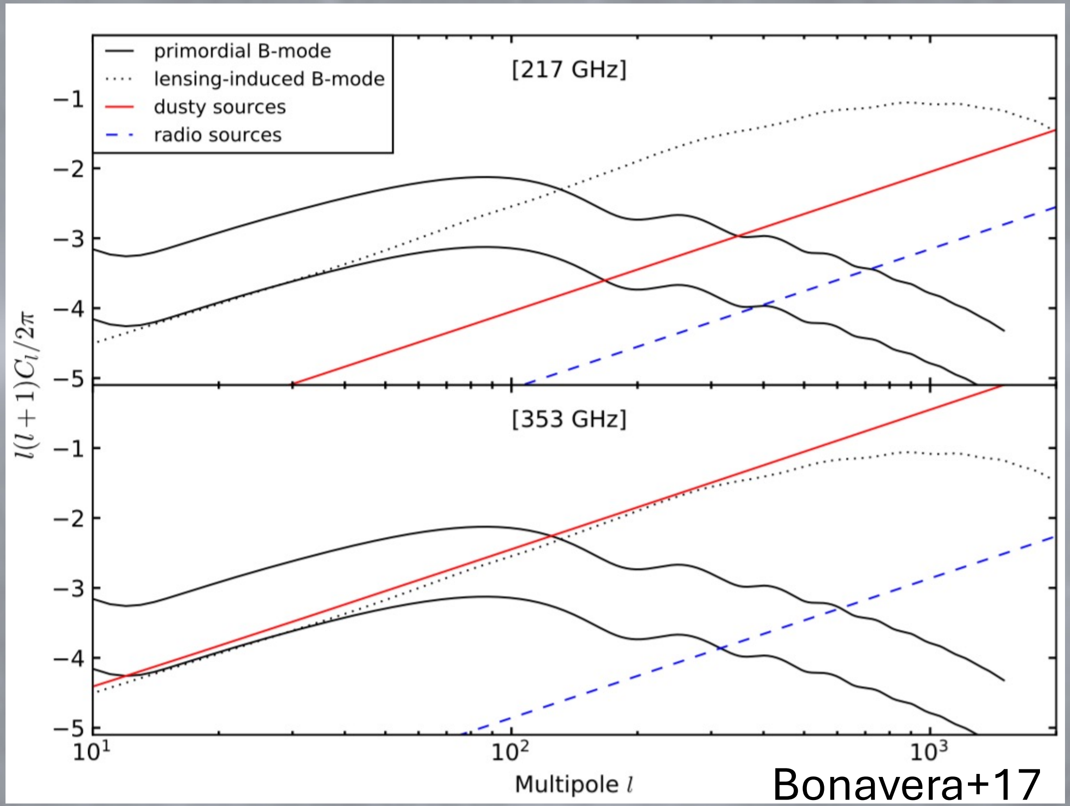
Point Sources

- …Unfortunately, many of the component separation techniques that are generally used to separate diffuse Galactic foregrounds are not well suited to deal with PS…
- PS: Distant galaxies seen as point-like objects through the observational beam
- Albeit “clustered”, their distribution is isotropic on very large scales ($\gtrsim$100 Mpc)
- Their contribution reduced by detecting and removing them from the maps
- An alternative is to MASK them

Point Sources



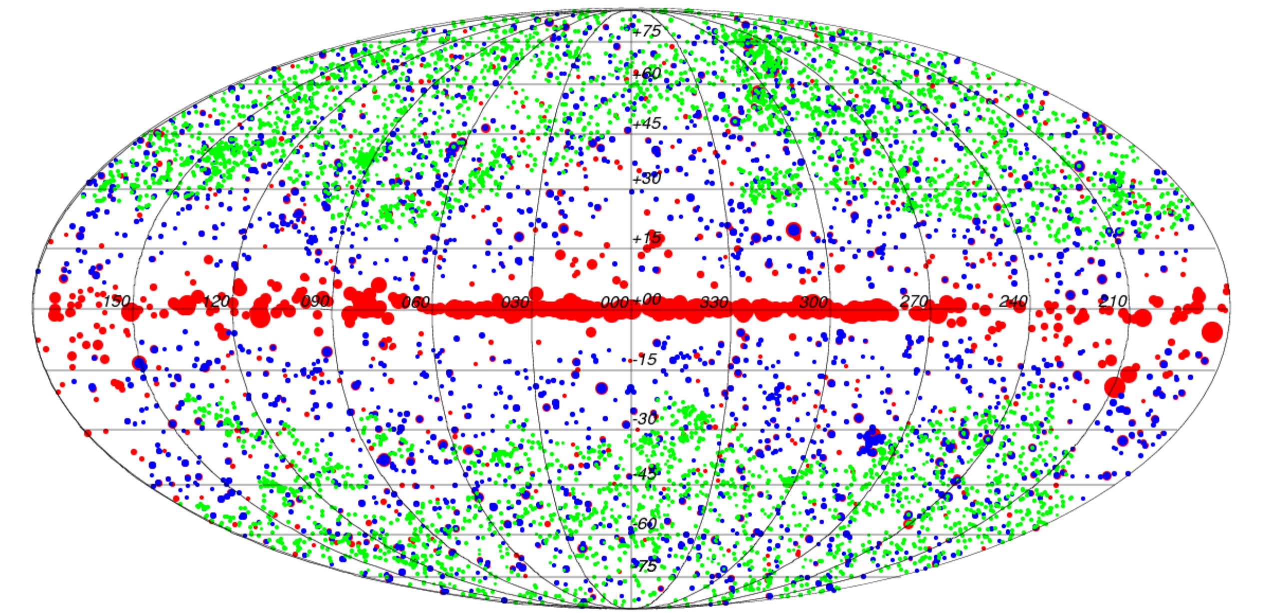
Planck 2015 XXVI



Bonavera+17

Point Sources

The Second Planck Catalogue of Compact Sources



Channel	Flux density 90% completeness [mJy]	No. of sources PCCS2	No. of sources PCCS2E	Polarized sources PCCS2	Polarized sources PCCS2E
30	427	1560	...	122	...
44	692	934	...	30	...
70	501	1296	...	34	...
100	269	1742	2487	20	43
143	177	2160	4139	25	111
217	152	2135	16842	11	325
353	304	1344	22665	1	666
545	555	1694	31068	...	...
857	791	4891	43290	...	...

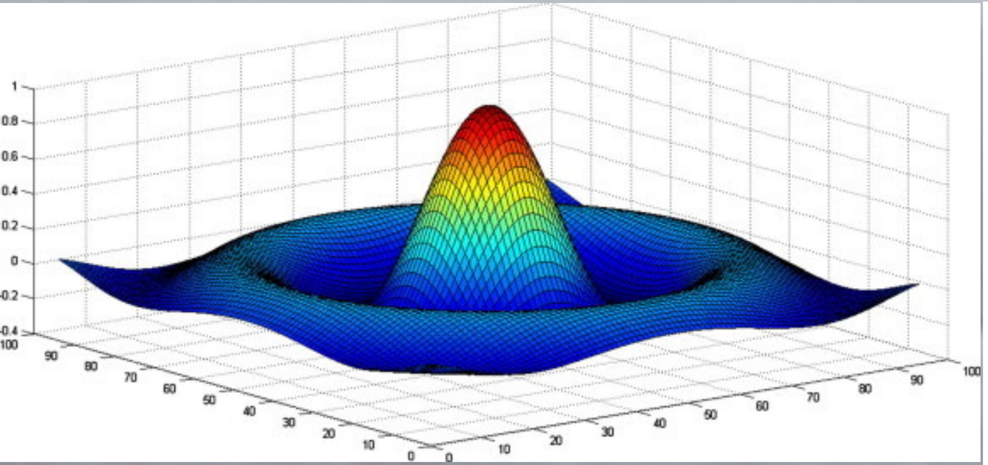
Planck 2015 XXVI

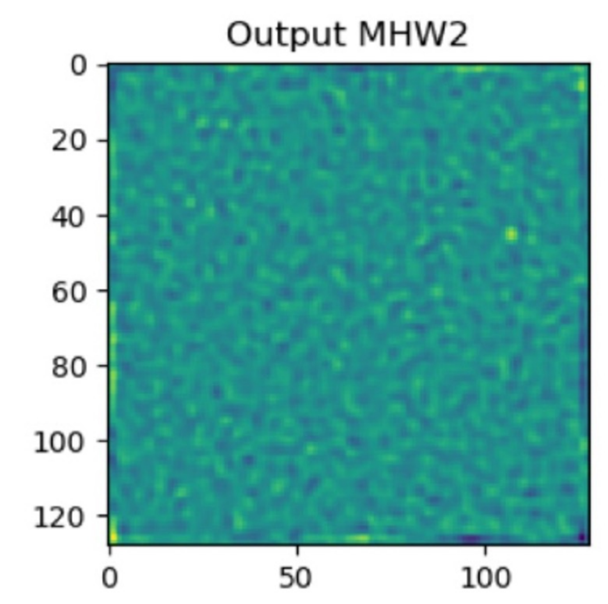
Point Sources

Detection

MHW family

$$\hat{\psi}_n(k) = \frac{k^{2n} e^{-\frac{k^2}{2}}}{2^n n!}$$



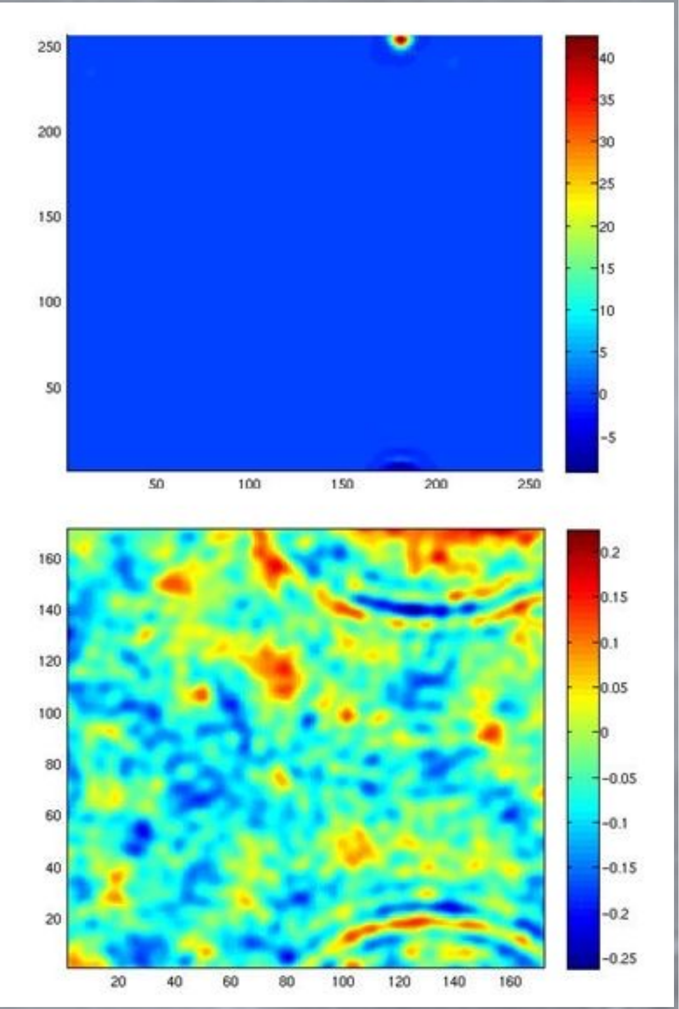


Matched Filter

Fourier transform of src profile(beam)

$$\psi_{MF} = \frac{1}{a} \frac{\tau(q)}{P(q)},$$

Bk power spectrum

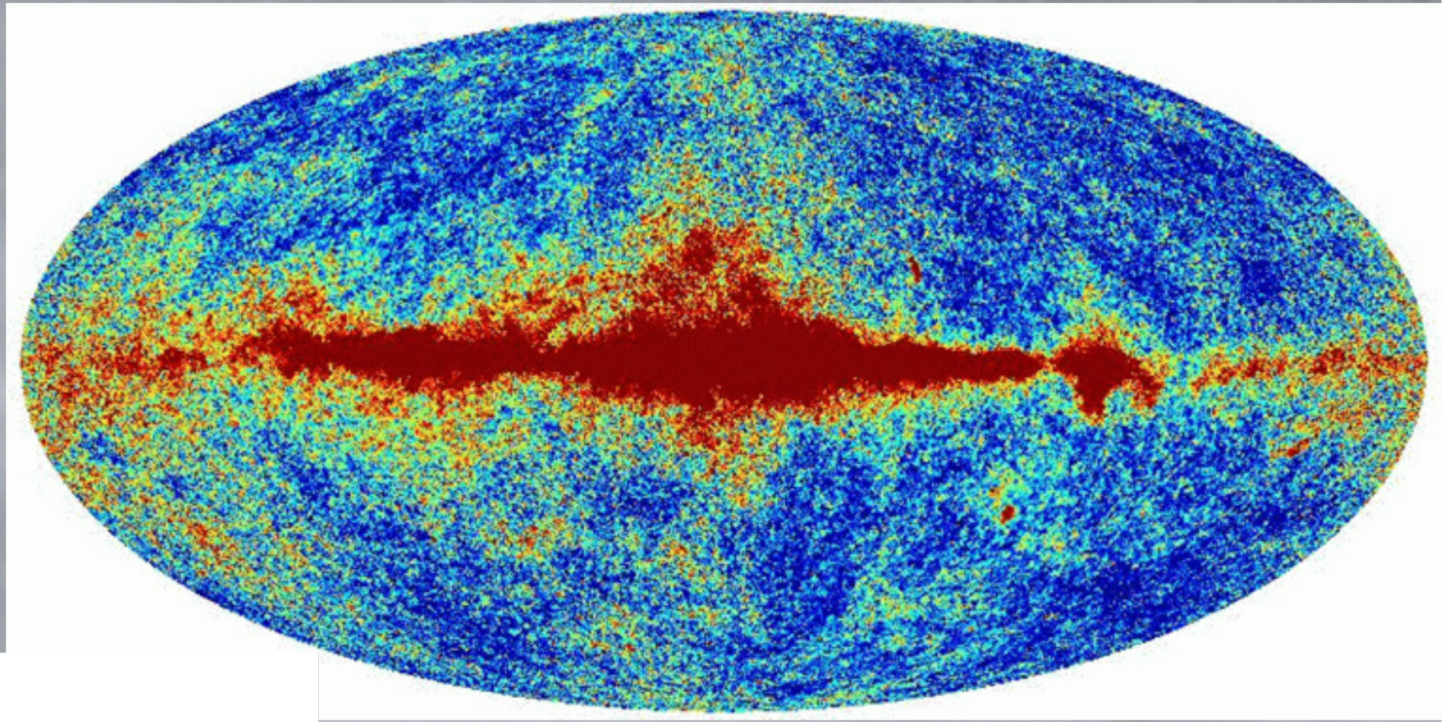


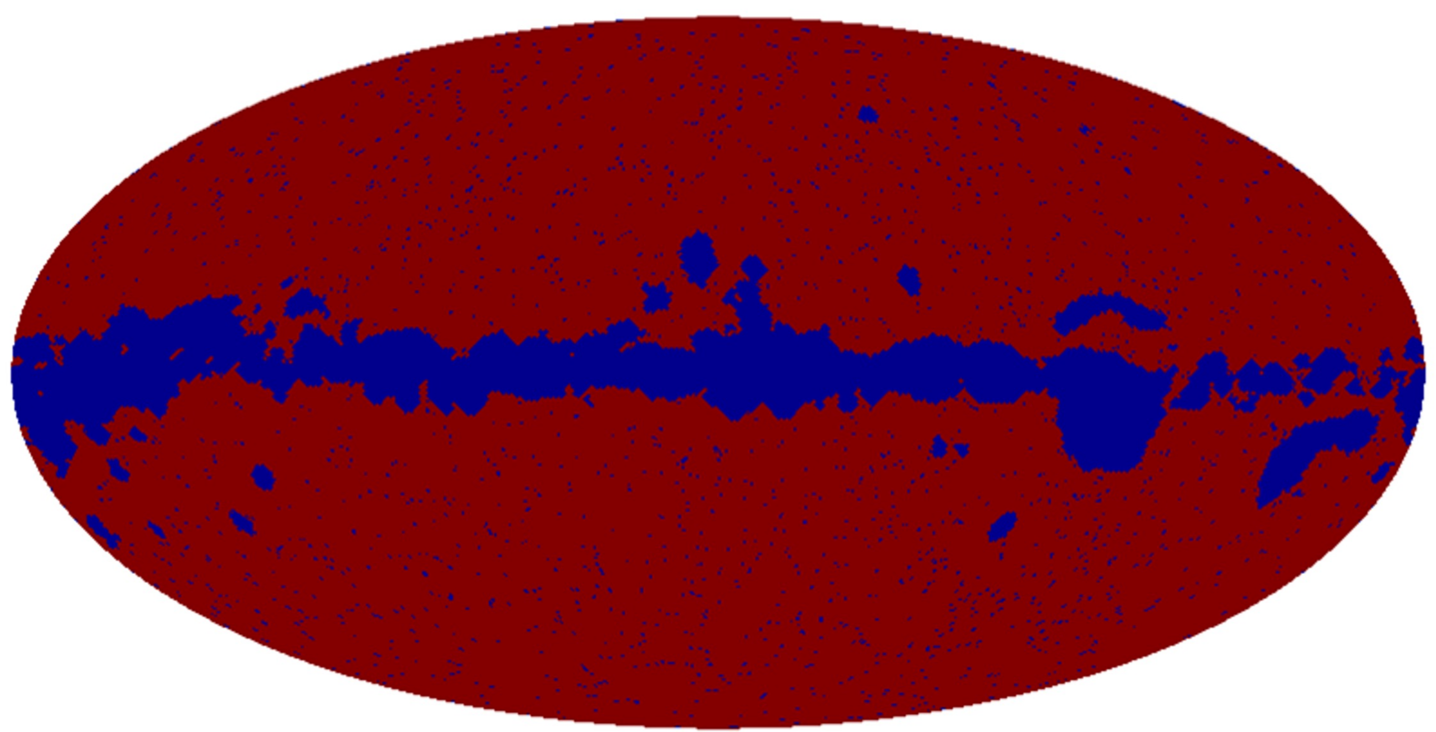
Gonzalez-Nuevo+2006

Lopez-Caniego+2006

Point Sources

Masking





Neural Network approach

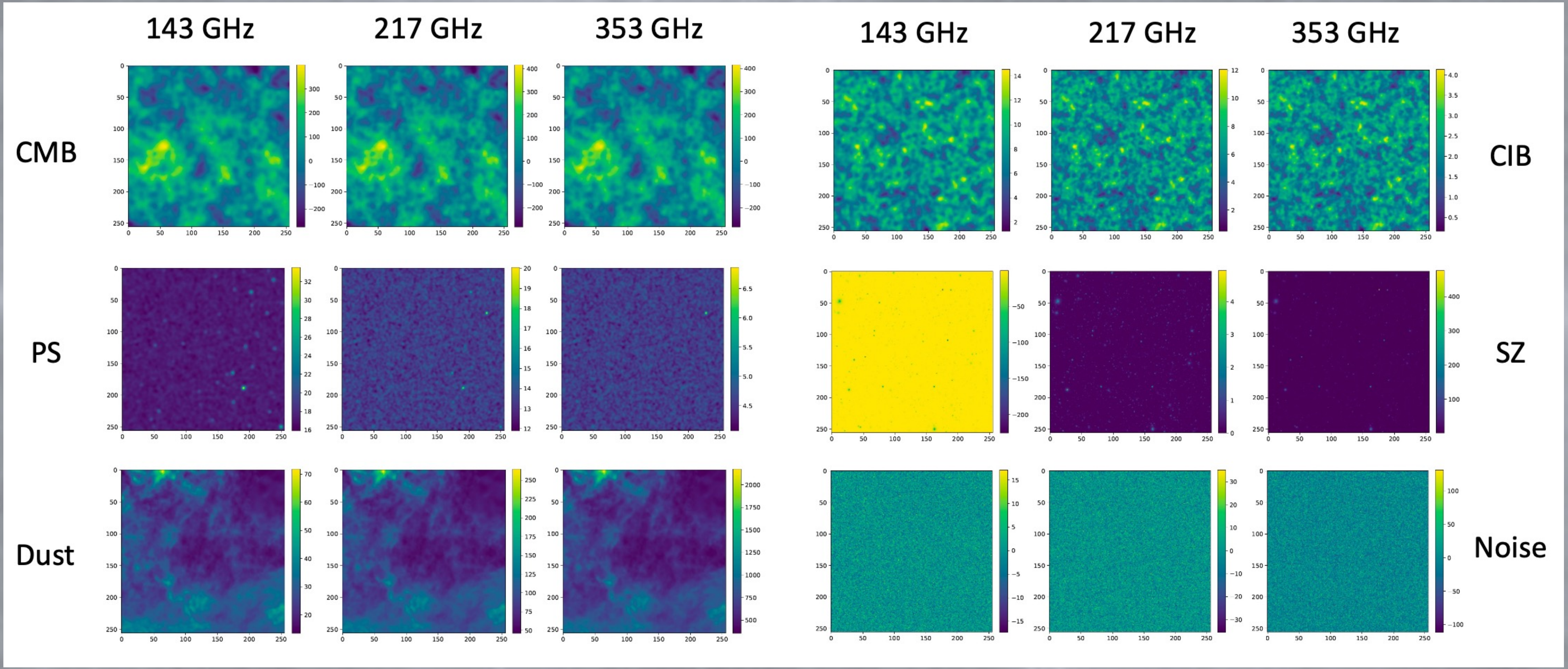
Looking for:

- Better performance
- No ringing
- No border effect
- No mask needed
- No bk power spectrum estimation
- More flexible and authomatic

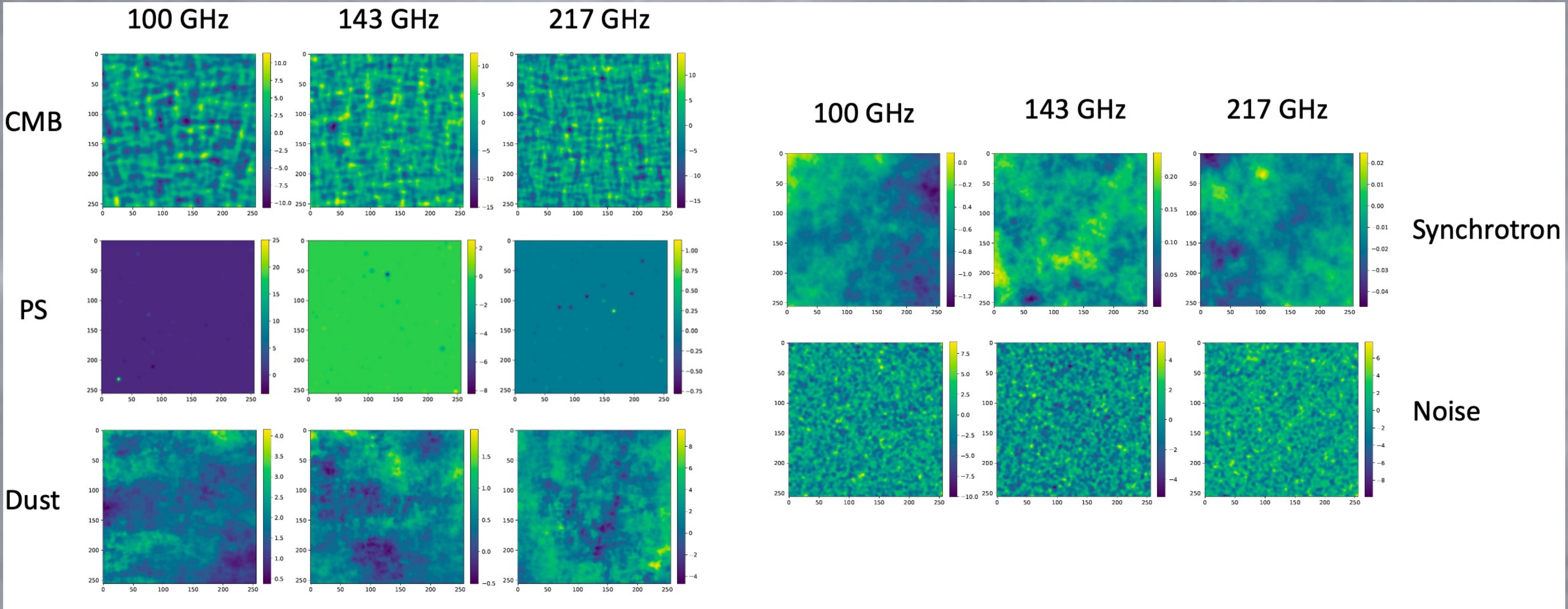
Realistic simulations needed:

- Patches of the sky
- CMB signal (label)
- Galactic thermal dust and synchrotron emission
- PS radio (label) and IR background
- Instrumental white noise

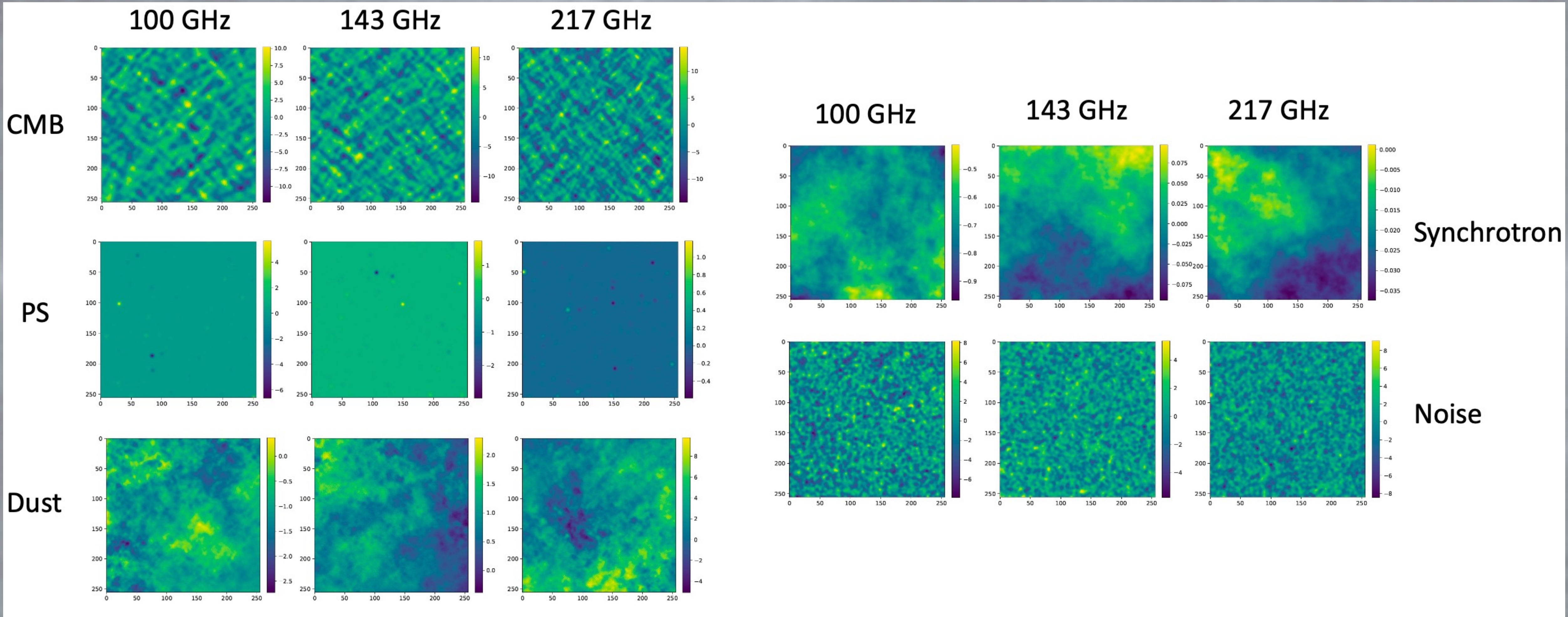
Simulations - T



Simulations - Q

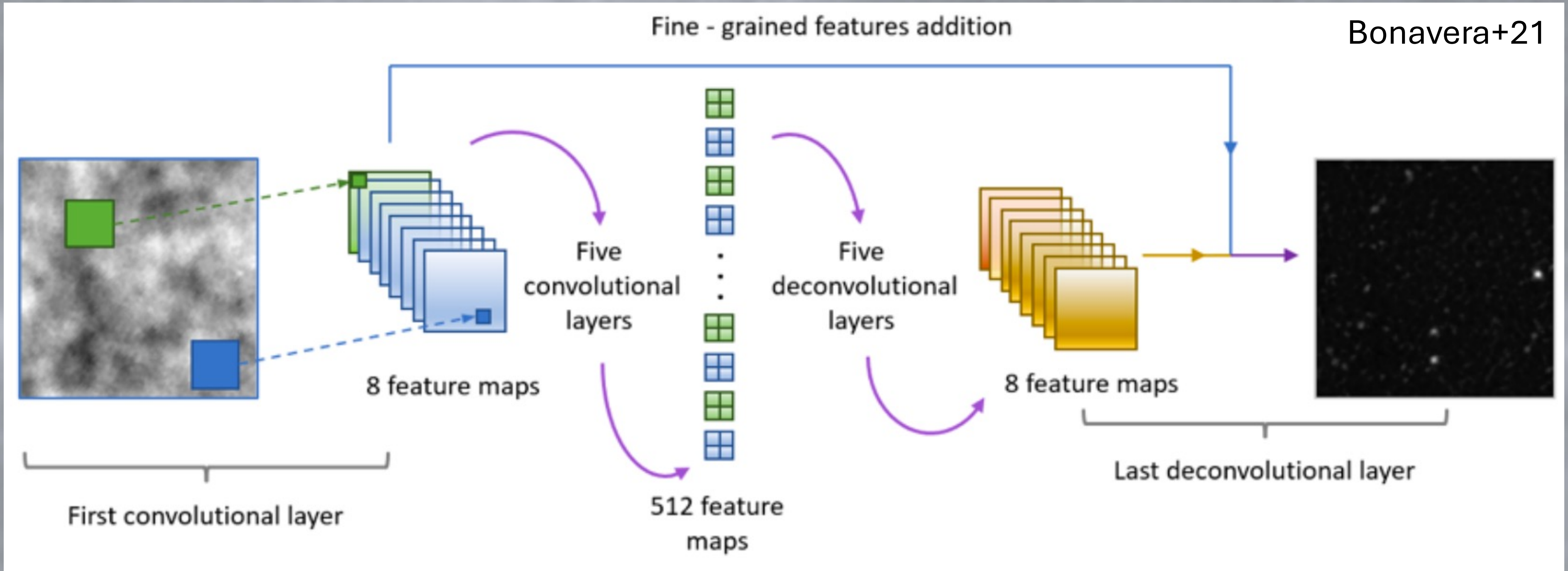


Simulations - U



PoSeIDoN

Point **S**ourc**e** **I**mage **D**etecti**o**n **N**etwork

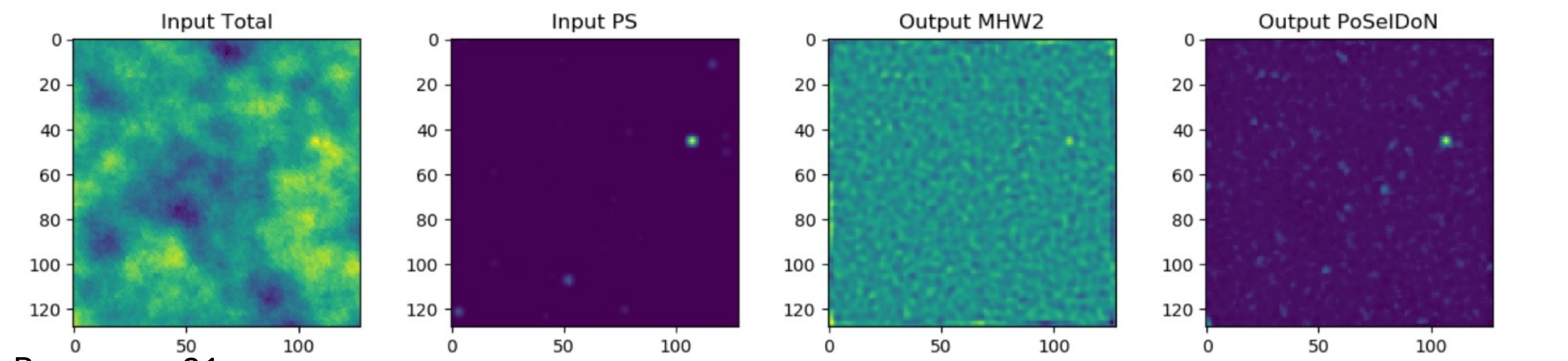


Convolutional block: 6 layers
8-16-64-128-256-512 feature maps

Paddding Same
Leaky ReLU
MSE loss function
50 epochs

Deconvolutional block: 6 layer
256-128-64-16-8-1 feature maps

PoSeIDoN



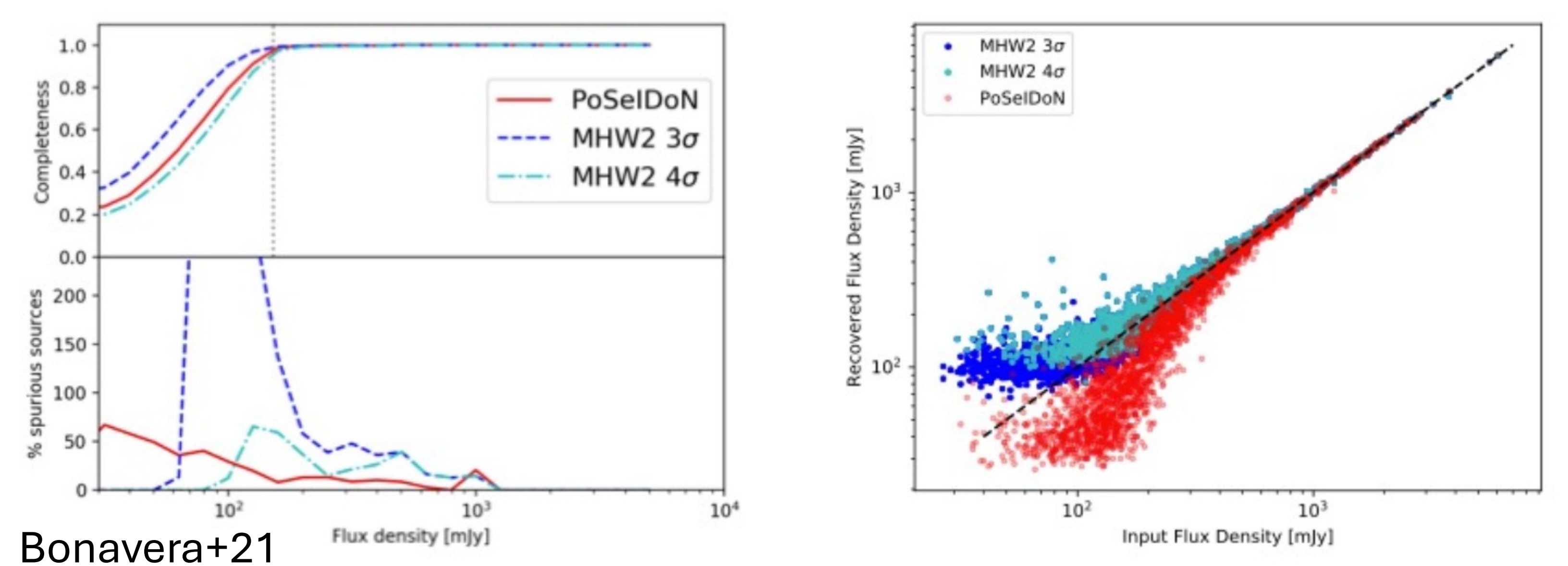
Bonavera+21

Simulations @217 GHz:
Patch of 32 x 32 pixels
50 000 training set (total & PS)
5 000 validation set

Catalogue:
searching peaks (i.e. local maxima)

- above σ_{MHW2} intensity threshold (PoSeIDoN)
- above 4 σ_{MHW2} (MHW2)

PoSeIDoN



Bonavera+21

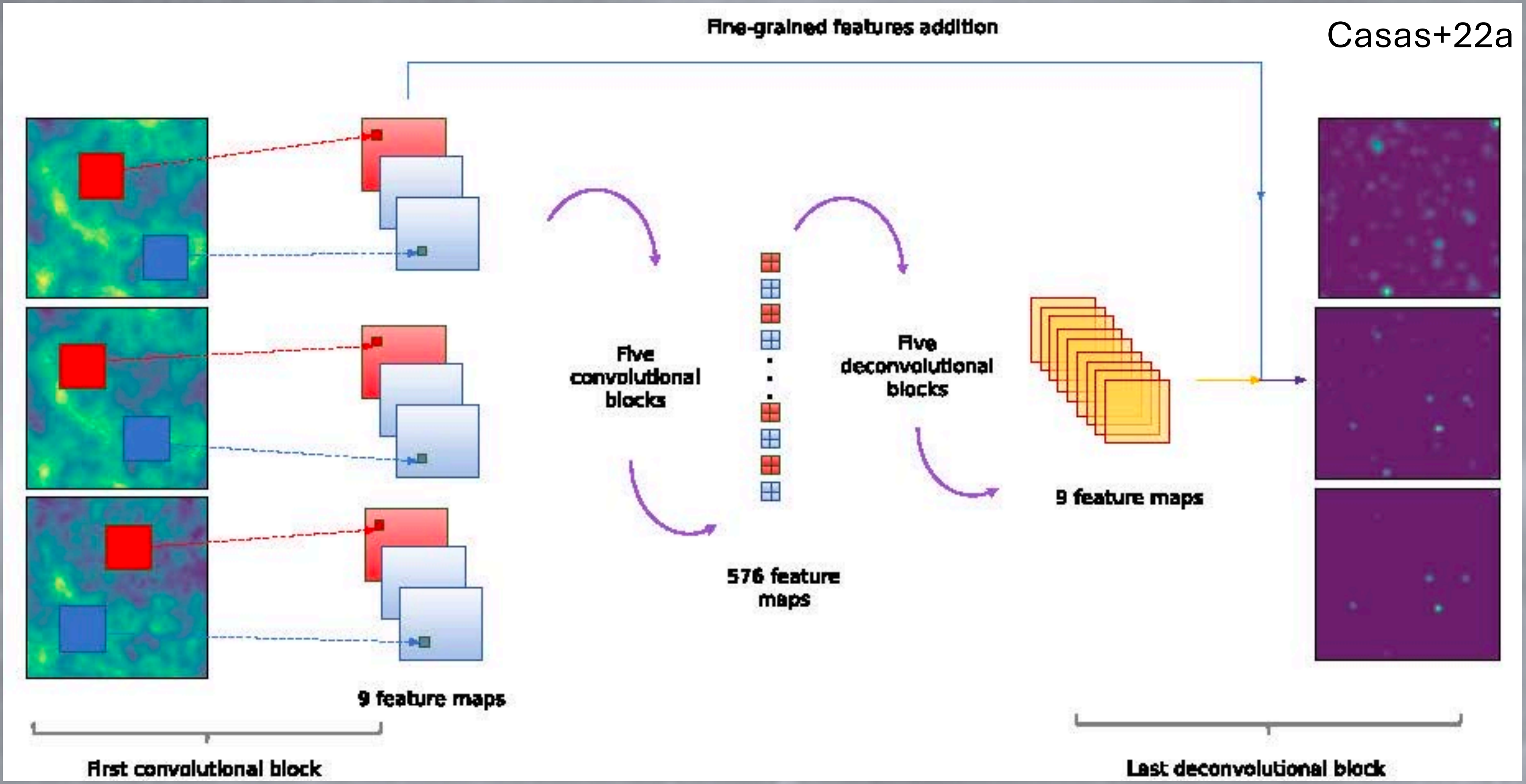
$$C(> S_0) = \frac{N_{true\ detected}(> S_0)}{N_{input}(> S_0)}$$

$$R(S_0) = \frac{N_{output}(> S_0) - N_{true\ detected}(> S_0)}{N_{input}(> S_0)}$$

- ➢ PoSeIDoN provides more reliable results (i.e. a lower number of spurious sources)
- ➢ PoSeIDoN does not have border effects like any filtering approach
- ➢ good PoSeIDoN performance even at the freq.s where it was not trained

❖ Flux density estimation is not optimal WRT the MHW2, but best option for blind detection

MultiPoSeIDoN

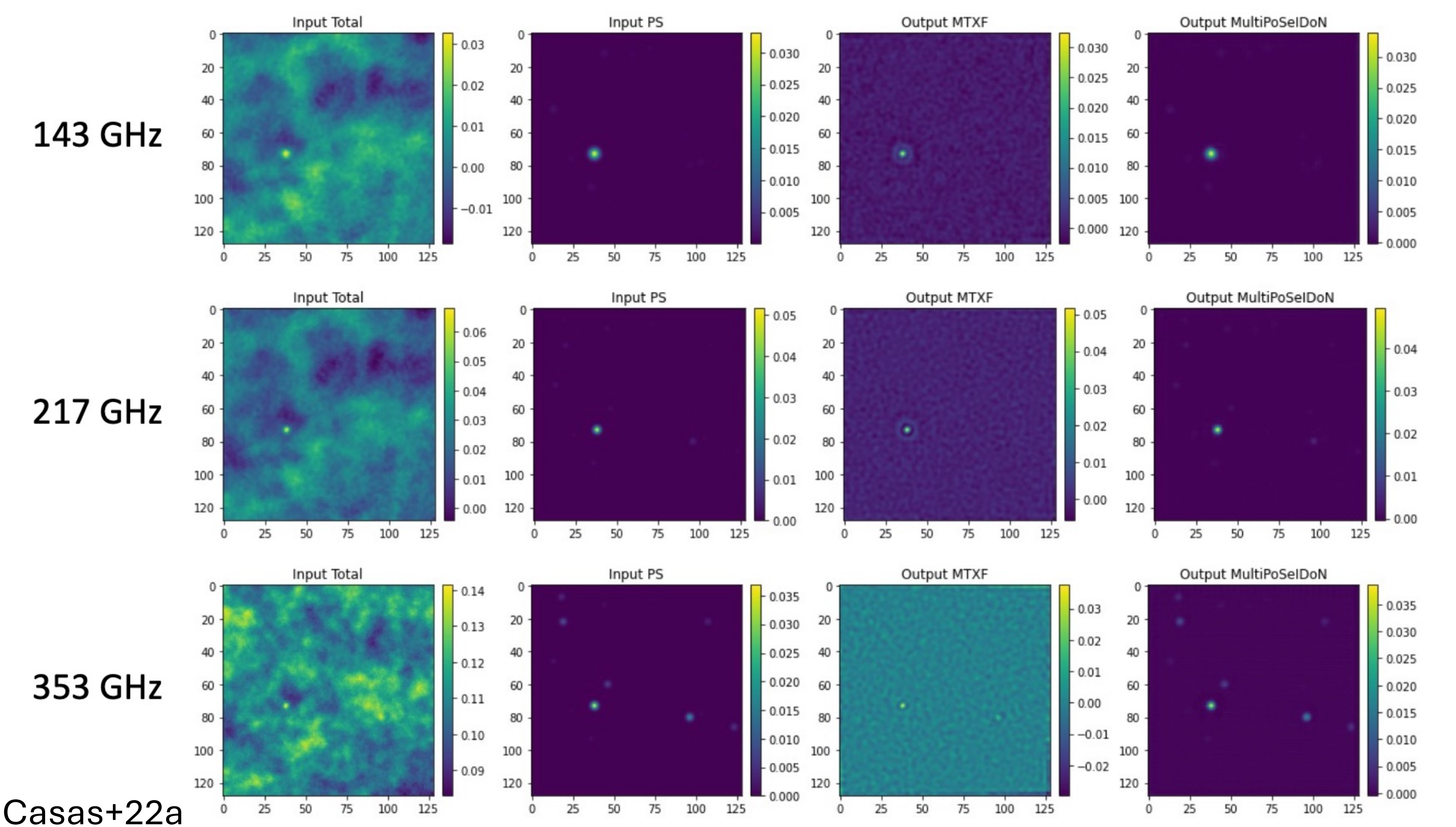


Sub-sampling factor of 2
Paddding Same
Leaky ReLU
MSE loss function
AdaGrad optimizer
Minibatch of 32 sample
500 epochs

6 convolutional & pooling layers
9-18-72-144-288-576 feature maps
Learning the PS WRT the background

6 deconvolutional & pooling layers
288-144-72-18-9-3 feature maps
PS segmantation from the total map

MultiPoSeIDoN



Casas+22a

Simulations:
143, 217 & 353 GHz
PS flux density scaling w/ freq
Patch of 128 x 128 pixels (90”)
50 000 training set (total & PS)
5 000 validation set

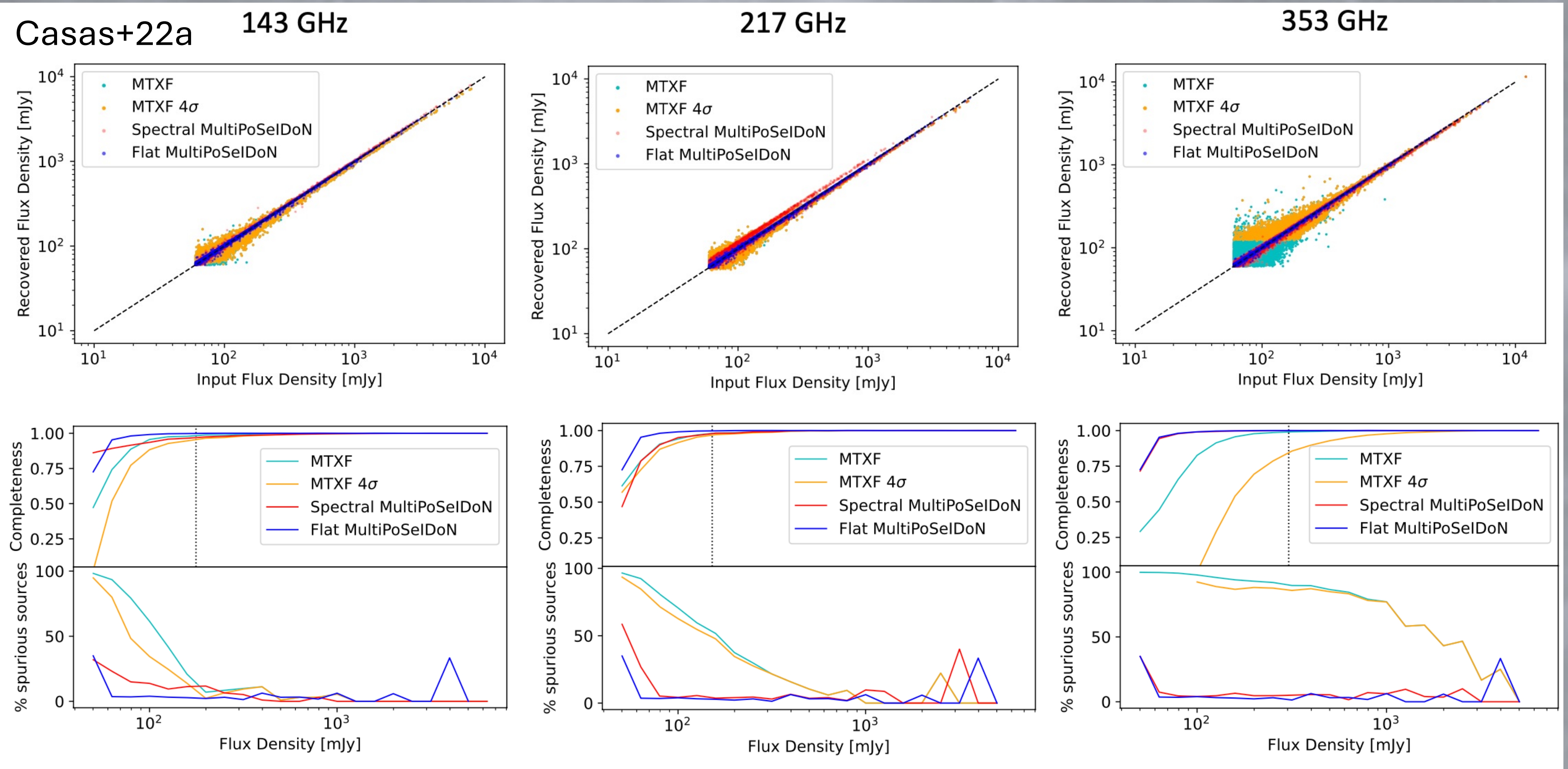
Catalogue:
searching peaks
(i.e. local maxima)

- NN above 60 mJy threshold
- MTXF 4σ

MultiPoSeIDoN

MultiPoSeIDoN performs better than the MTXFs

- Similar completeness @ 143 & 217 GHz
- Better completeness @ 353 GHz
- Better in number of spurious sources



@143-217-353 GHz

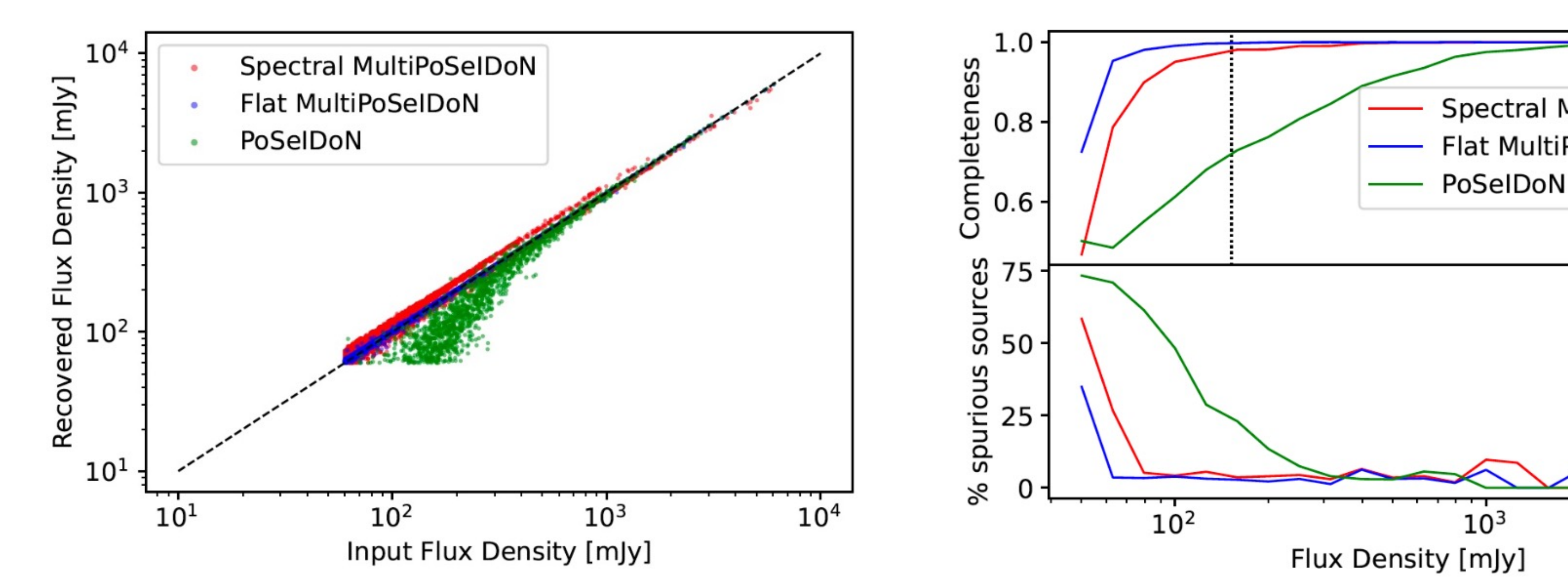
90% completeness level

- NN 79-71-60 mJy
- MTXF 84-79-123 mJy

Spurious source

- NN ~20% S<100 mJy
- MTXF > 20% S< 180-400-1200 mJy

MultiPoSeIDoN

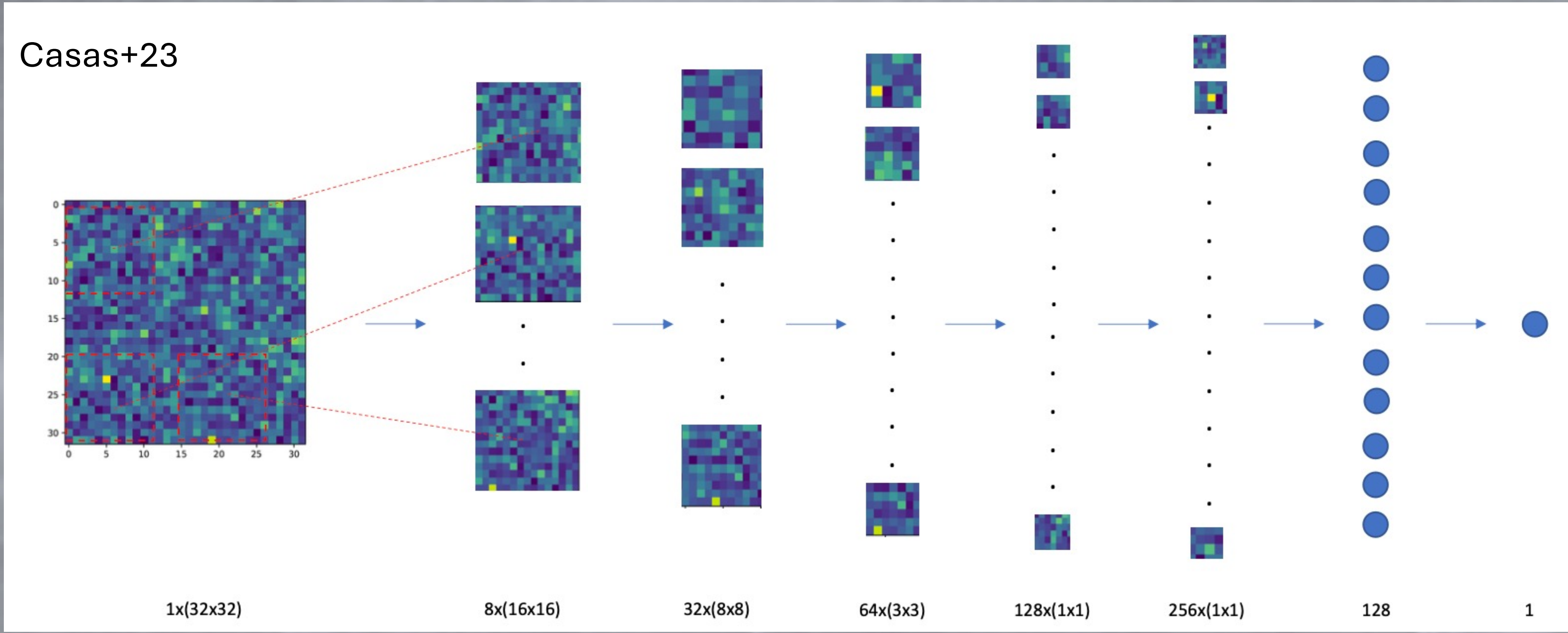


Casas+22a

- MultiPoSeIDoN better than PoSeIDoN, recovering flux density of fainter PS w/ lower relative error
- Thanks to the increasing of the training information, it learns the different correlations between the elements in the simulations due to their spectral behaviors

POSPEN

POint **S**ource **P**olarization **E**stimation **N**etwork



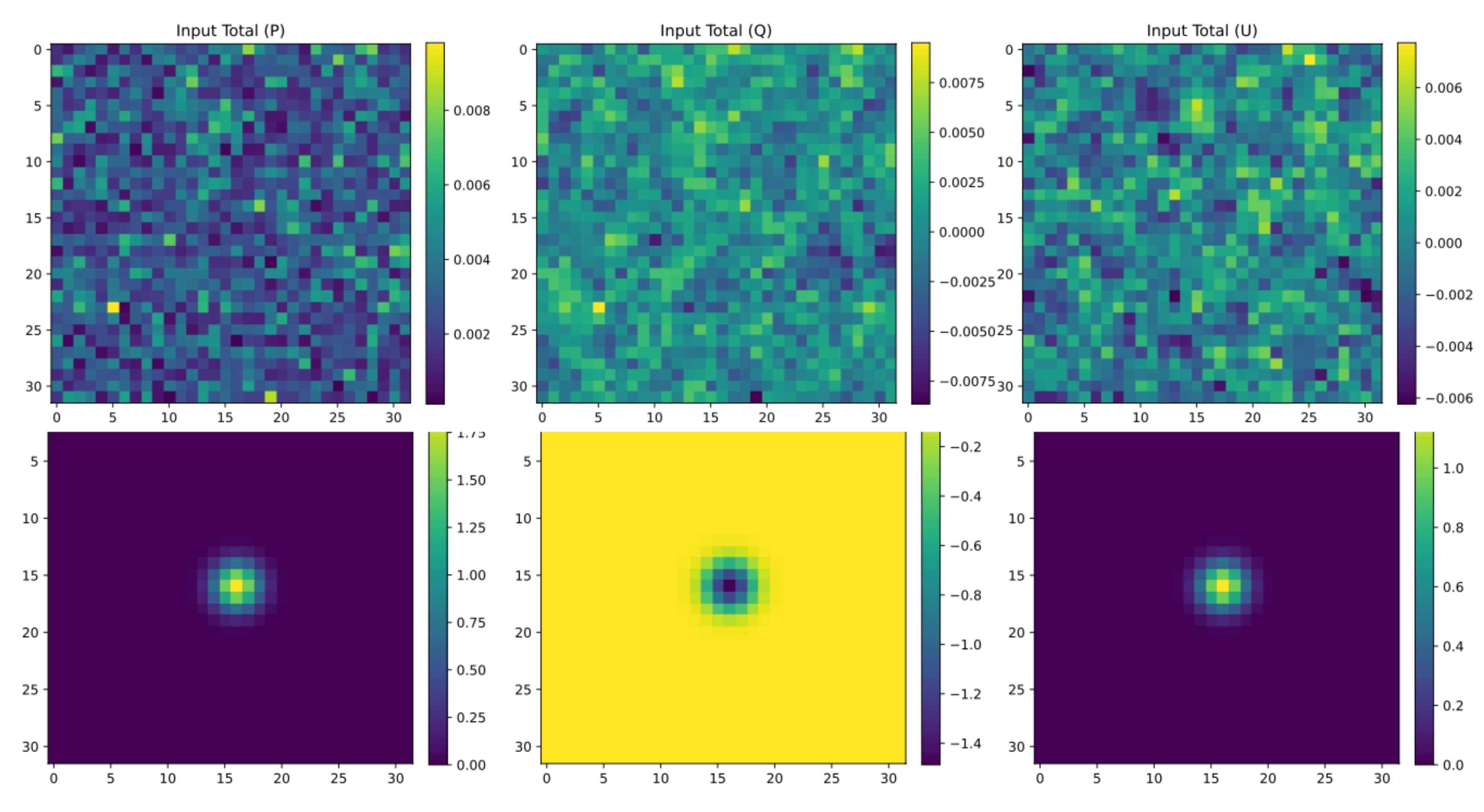
leaky ReLU
AdaGrad optimiser
500 epochs
batch size of 16
MSE loss function

1st block read the input 32x32 patch
Five convolutional blocks, formed by
8-32-64-128-256 filters

\+

two layers of 128 and 1 neurons
converting info to numerical
values

POSPEN

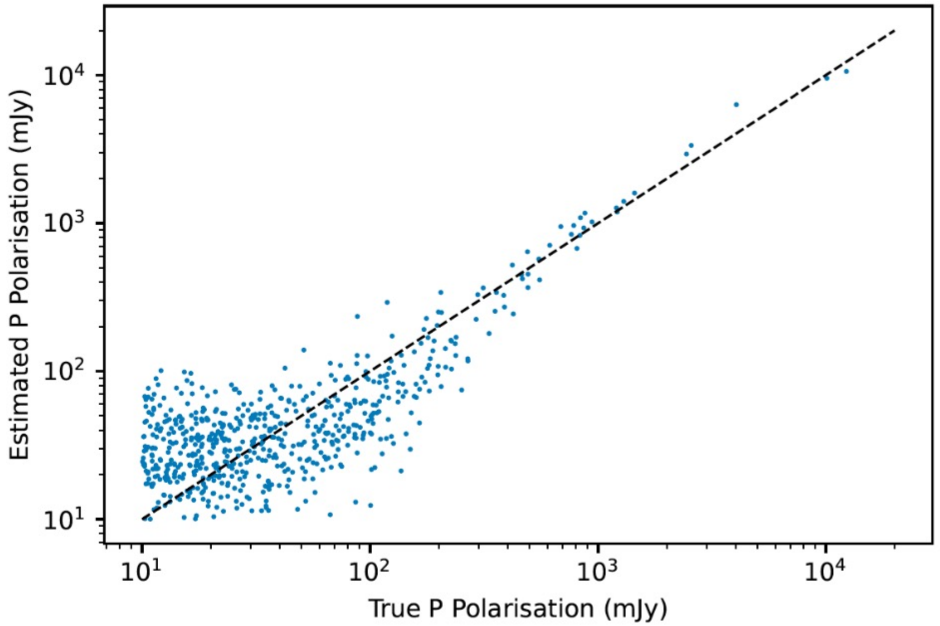


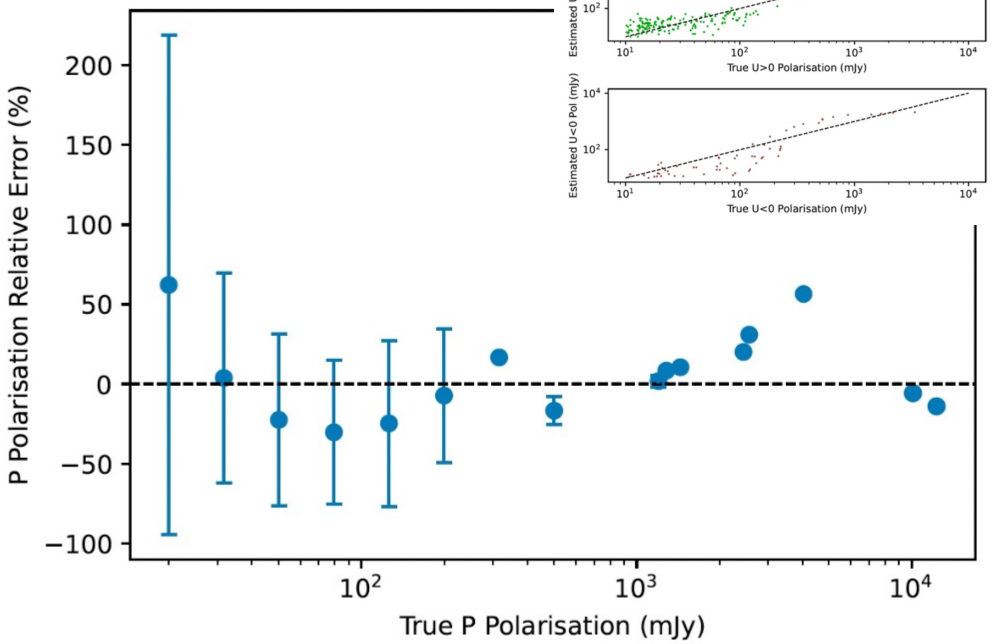
Polarization simulations (P Q U)
@217 GHz
32 × 32 pixels of 90”
a central injected PS
(non-blind method)
+ contaminants & CMB

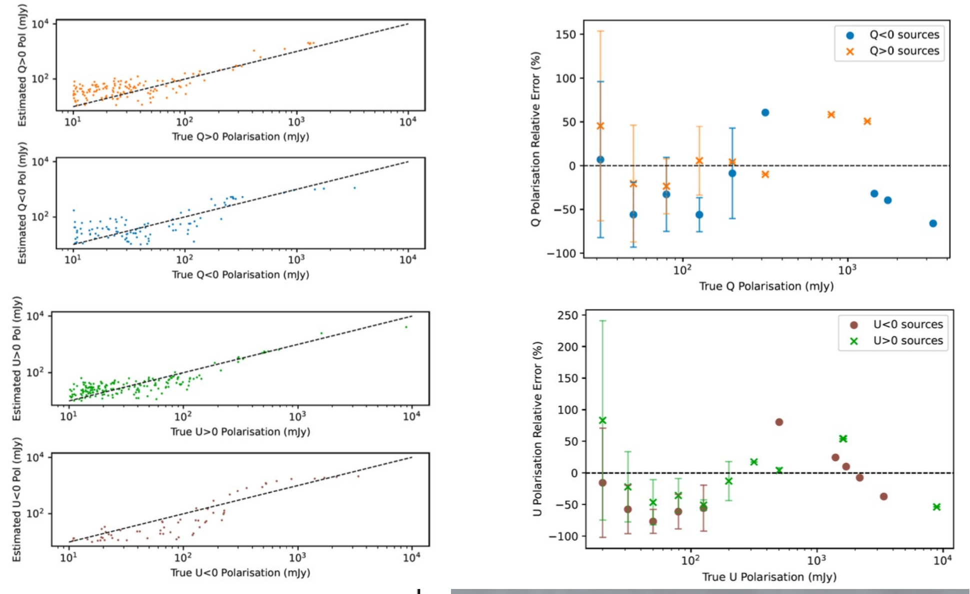
10 000 training set (label PS flux)
1 000 validation set

Casas+23

POSPEN







POSPEN appears to be promising for estimating **polarization flux density** (non-blind way)

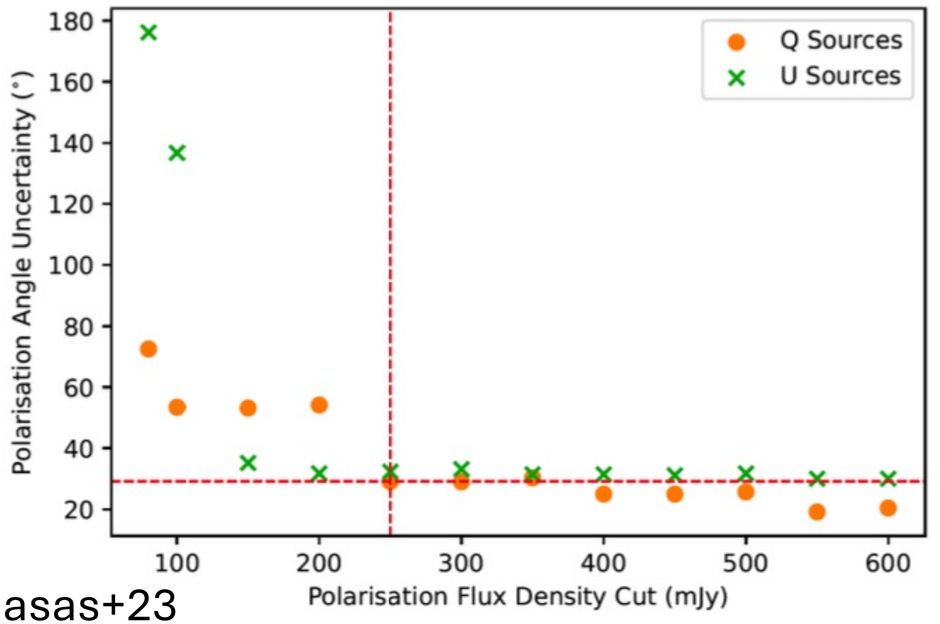
- ➢ It well recovers the polarization flux density of sources above 80 mJy
- ➢ Relative error of 30% in most of the flux-density intervals

Casas+23

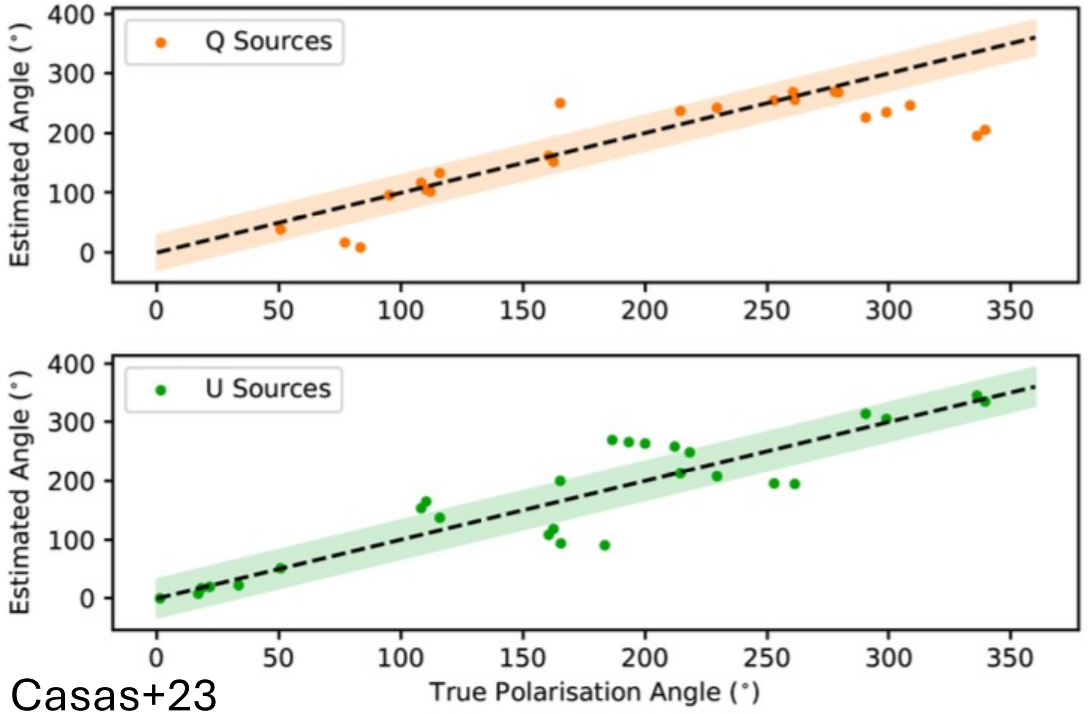
POSPEN

$\psi = \tan^{-1}(U/Q)$
$Q = P\cos\psi$
$U = P\sin\psi$

The **polarization angle** (ψ) can be estimated even when Q is well estimated but not its corresponding U, or vice versa



Casas+23



Casas+23

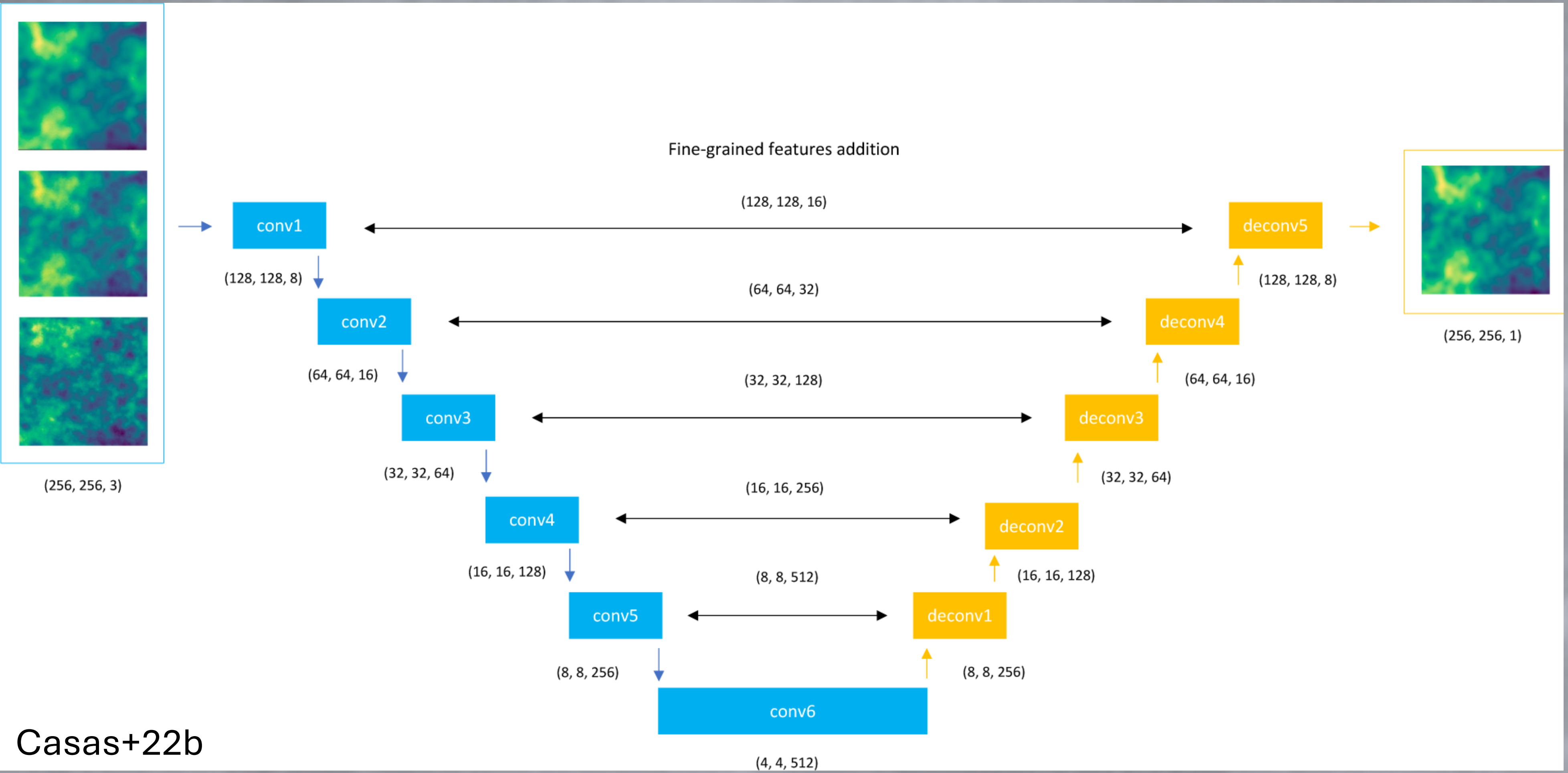
POSPEN appears to be promising for estimating polarization flux density and angle in a non-blind way

NN4CMB

- NN better performance in PS WRT filters
- Same expected in CMB recovery WRT "classical" methods
- In particular:
    - Better background removal expected
    - Better noise removal expected

CENN - T

CMB **E**xtraction **N**eural **N**etwork



Casas+22b

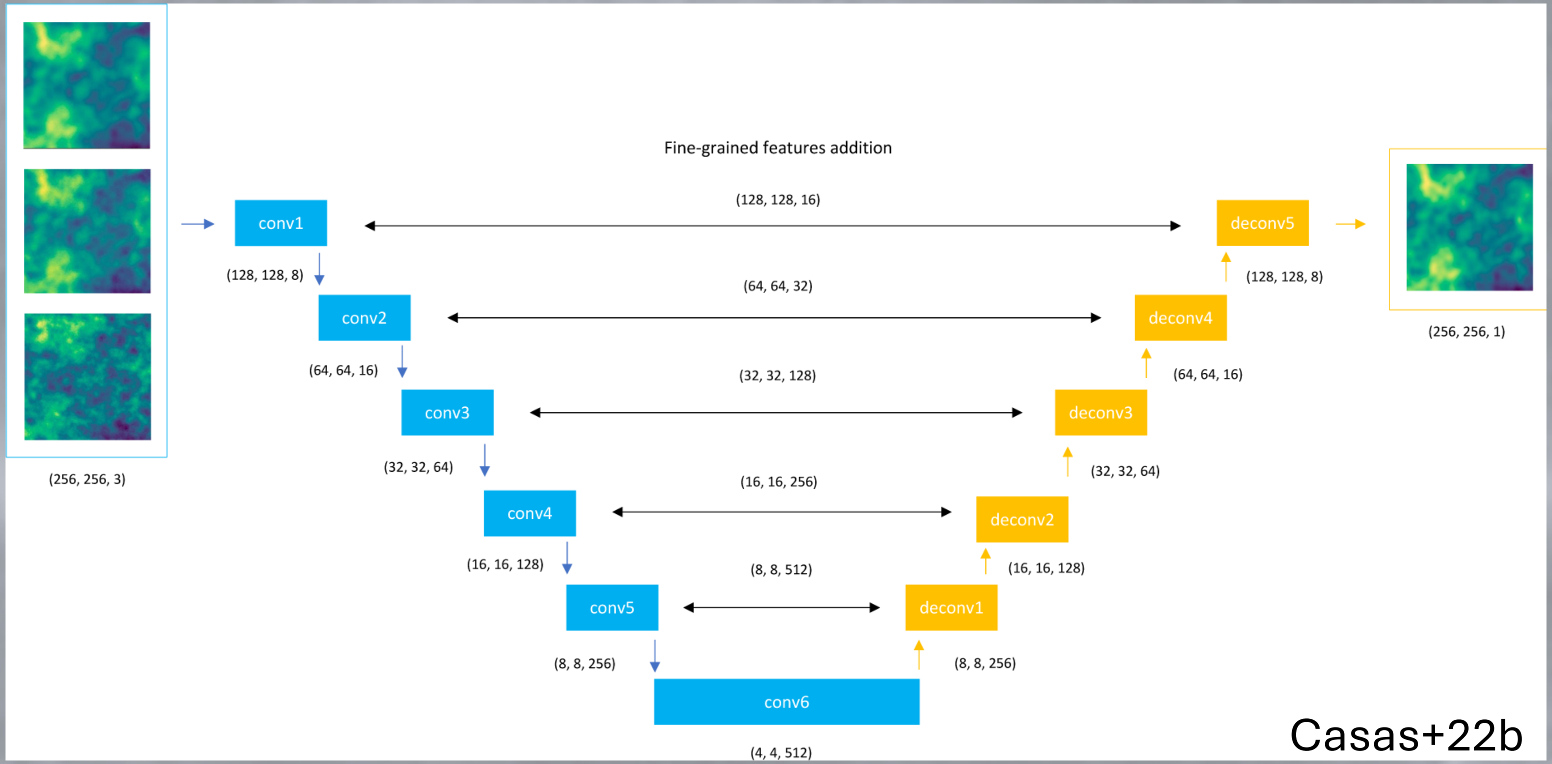
In: 3 patches
Out: 1 patch w/ clan CMB

6 convolutional blocks:
layers w/
8-16-64-128-256-512 #filters

6 deconvolutional blocks:
layers w/
256-128-64-16-8-1 #filters

500 epochs
Mini-batch 32
MSE loss function
Subsampling factor of 2
Padding type Same
leaky ReLU

CENN - T



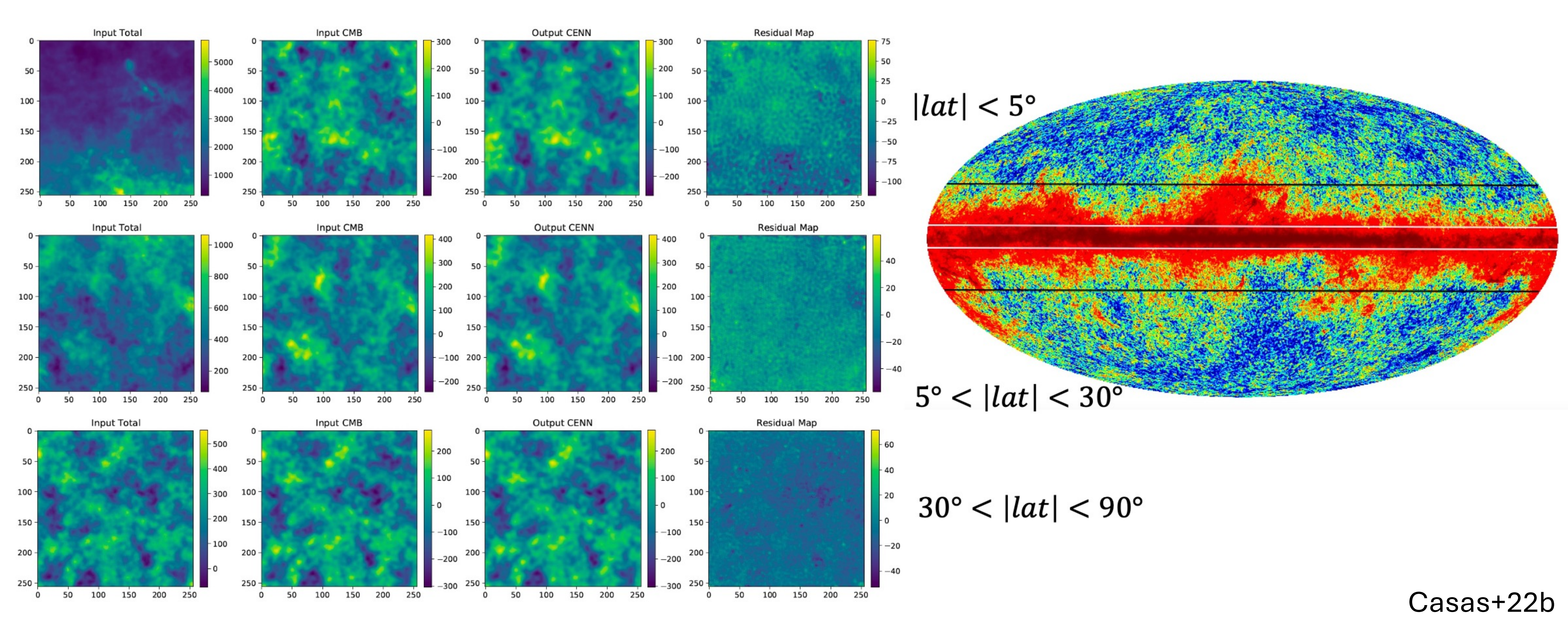
- layers to connect the convolutional and deconvolutional blocks, doubling the space of feature maps before each deconvolutional block
- these layers help to predict low-level features with the deconvolutional blocks by taking into account high-level features inferred by the convolutional blocks
- the addition of these layers is related to the task of predicting small-scale regions of the CMB signal by considering already inferred large-scale structures

CENN - T

Simulations @ 143, 217 & 353 GHz
PS flux density scaling w/ freq
256 x 256 pixels 90”

60 000 training set (labels CMB @ 217 GHz)
validation set:

- 6 000 all sky
- 2 000 x 3 regions



CENN - T

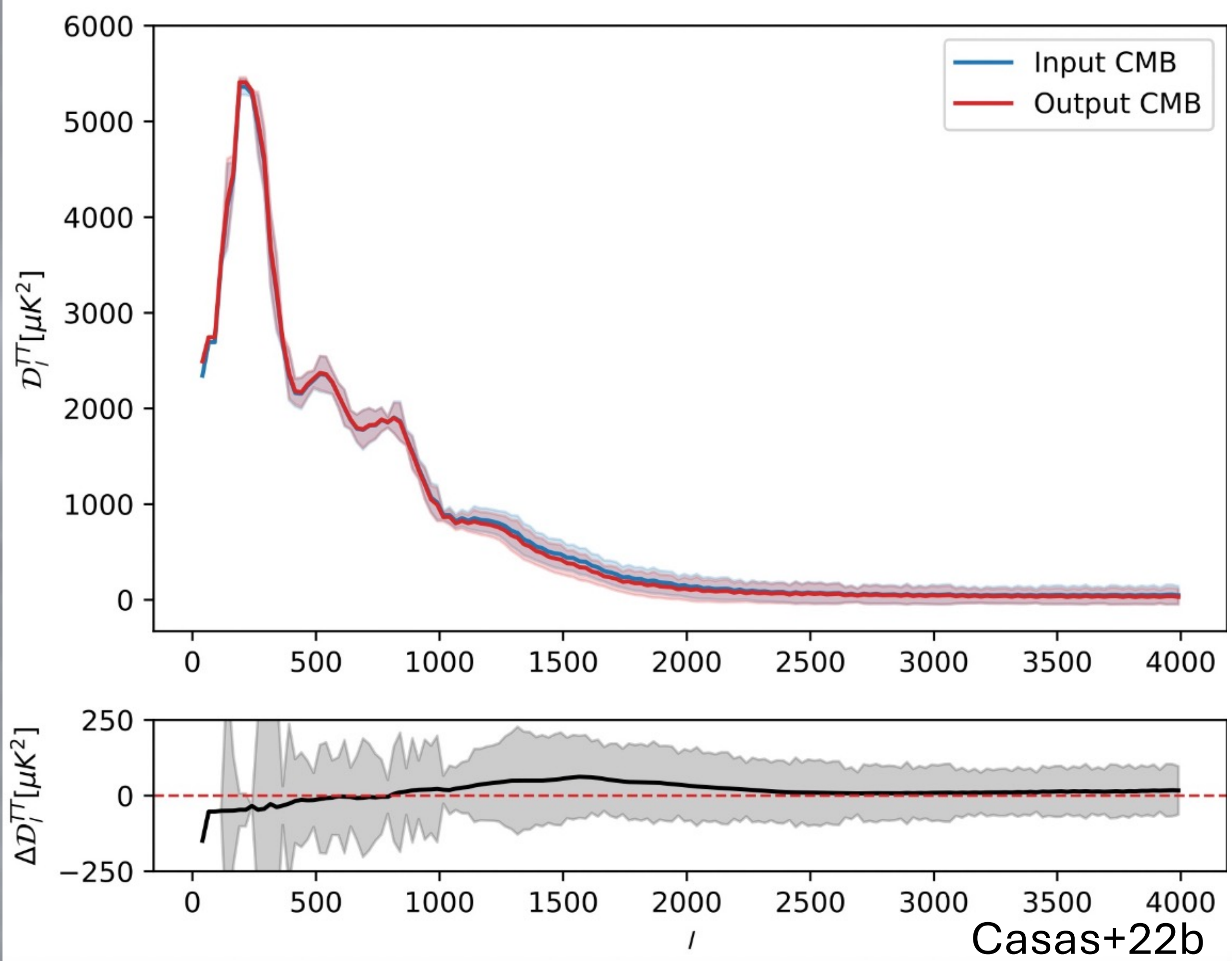
Mean power spectra of the input and output CMB

Difference of 13±113 μ K^2 for l ≤4000

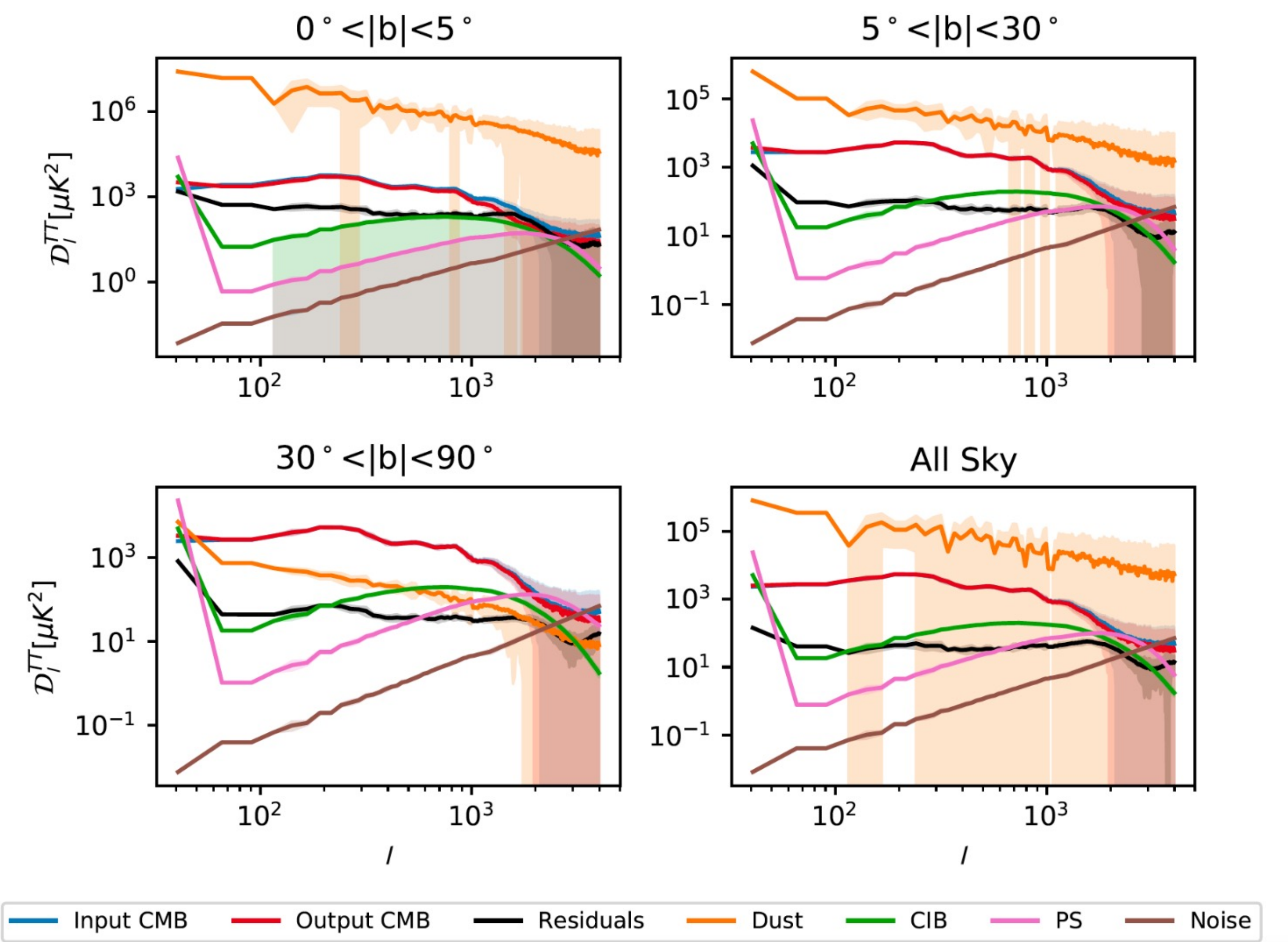
- ➢ CENN reliable also @ l> 2500
- ➢ PS contamination very small, only affects l~2000

Mean power spectrum of the residuals + foregrounds (difference between input and output CMB) for each region and for the whole sky

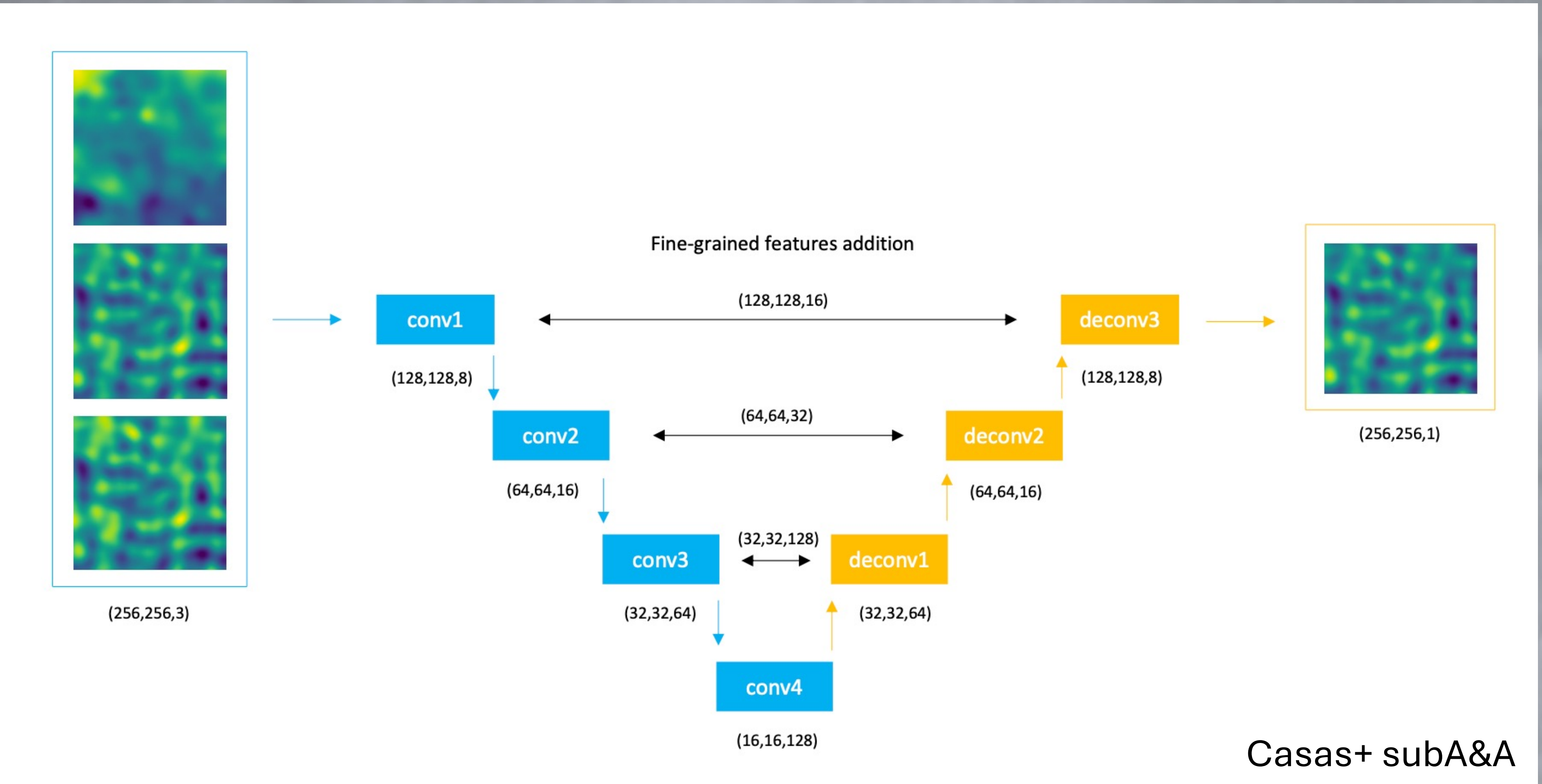
- ➢ reasonable residuals also in contaminated regions



Casas+22b



CENN - Pol

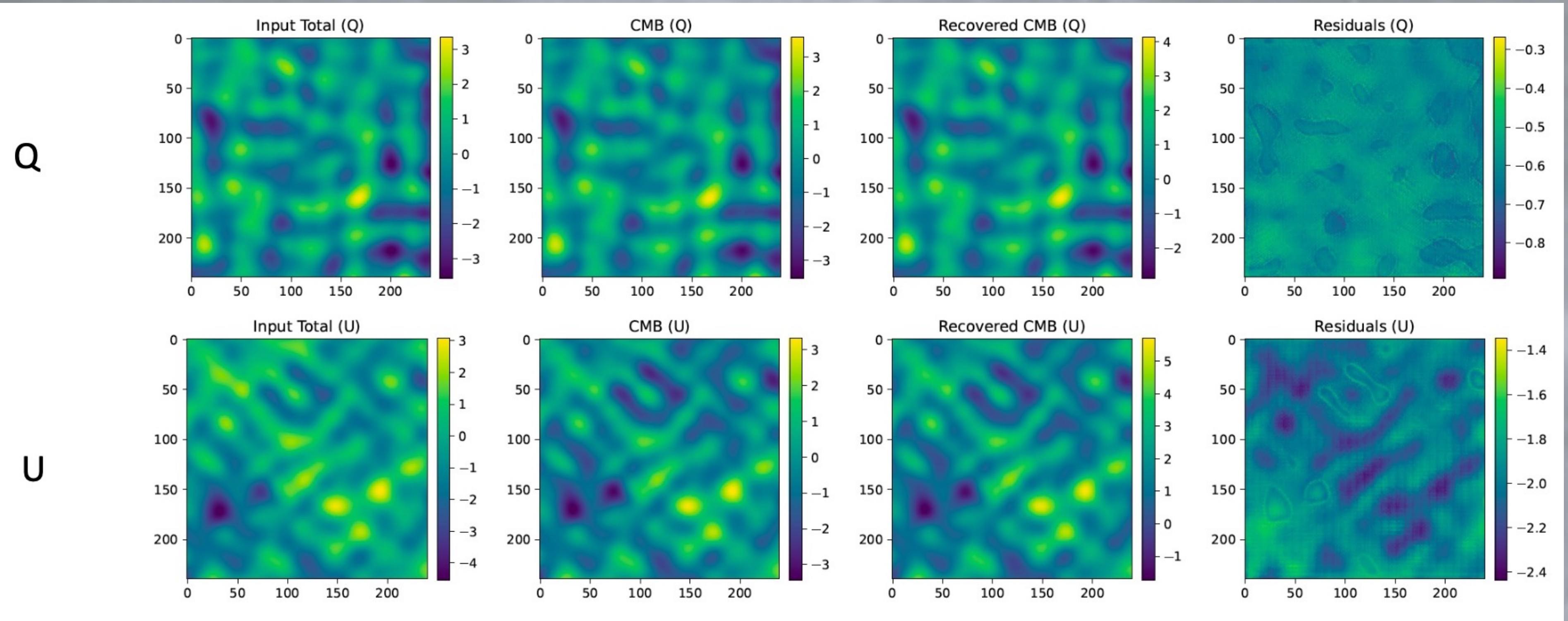


4 convolutional blocks:
8-16-64-128 #filters

4 deconvolutional blocks
64-16-8-1 #filters

Convol. connected to deconv.
to add fine-grained features
padding type Same
activation function leaky ReLU
Final layer MSE loss function
AdaGrad optimizer
500 epochs

CENN - Pol

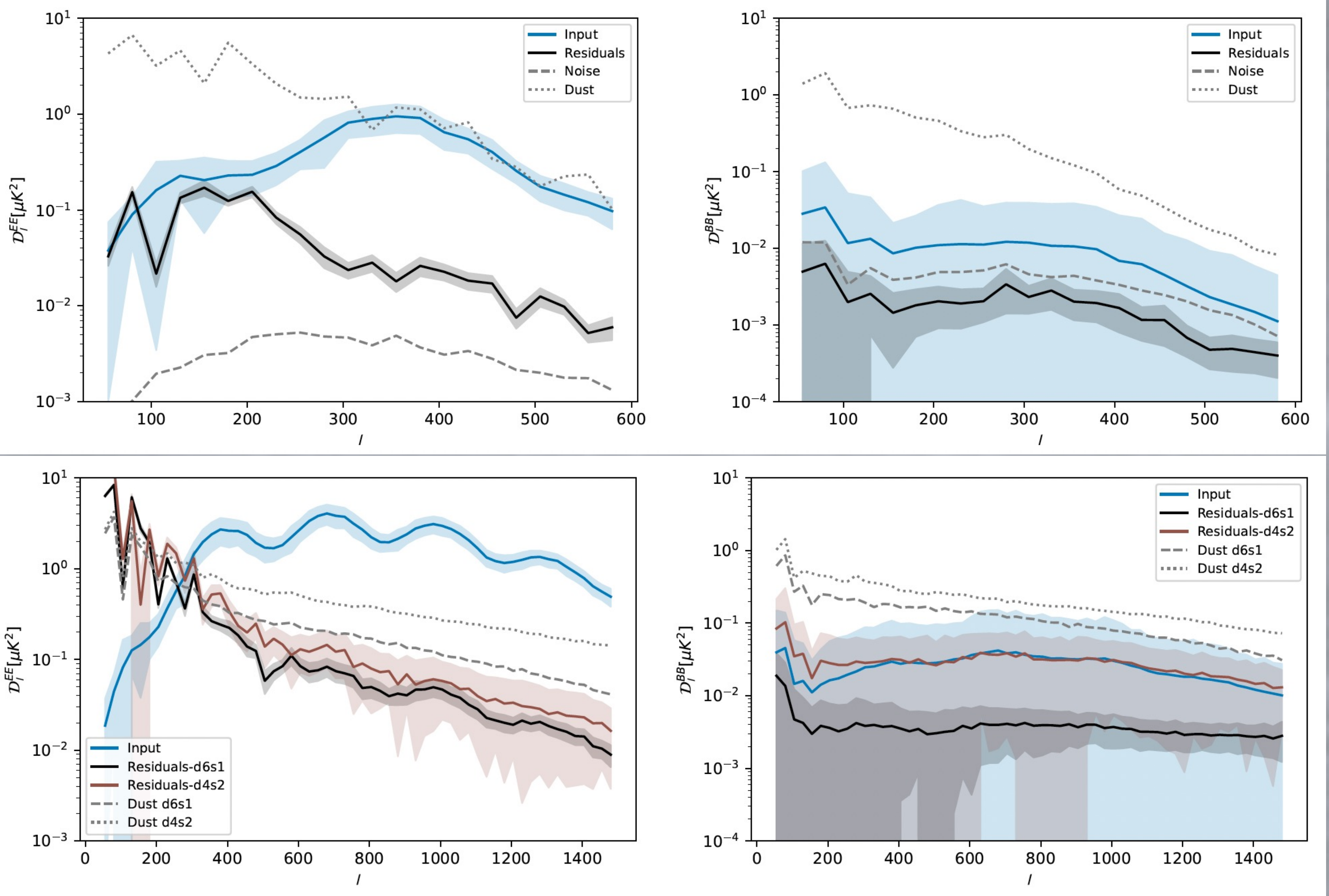


Simulations in Q and U
@ 100, 143 & 217 GHz
256×256 pixels & 90” pixel size

10 000 training set
(labels CMB @ 217 GHz)
1 000 validating set

Casas+ subA&A

CENN - Pol



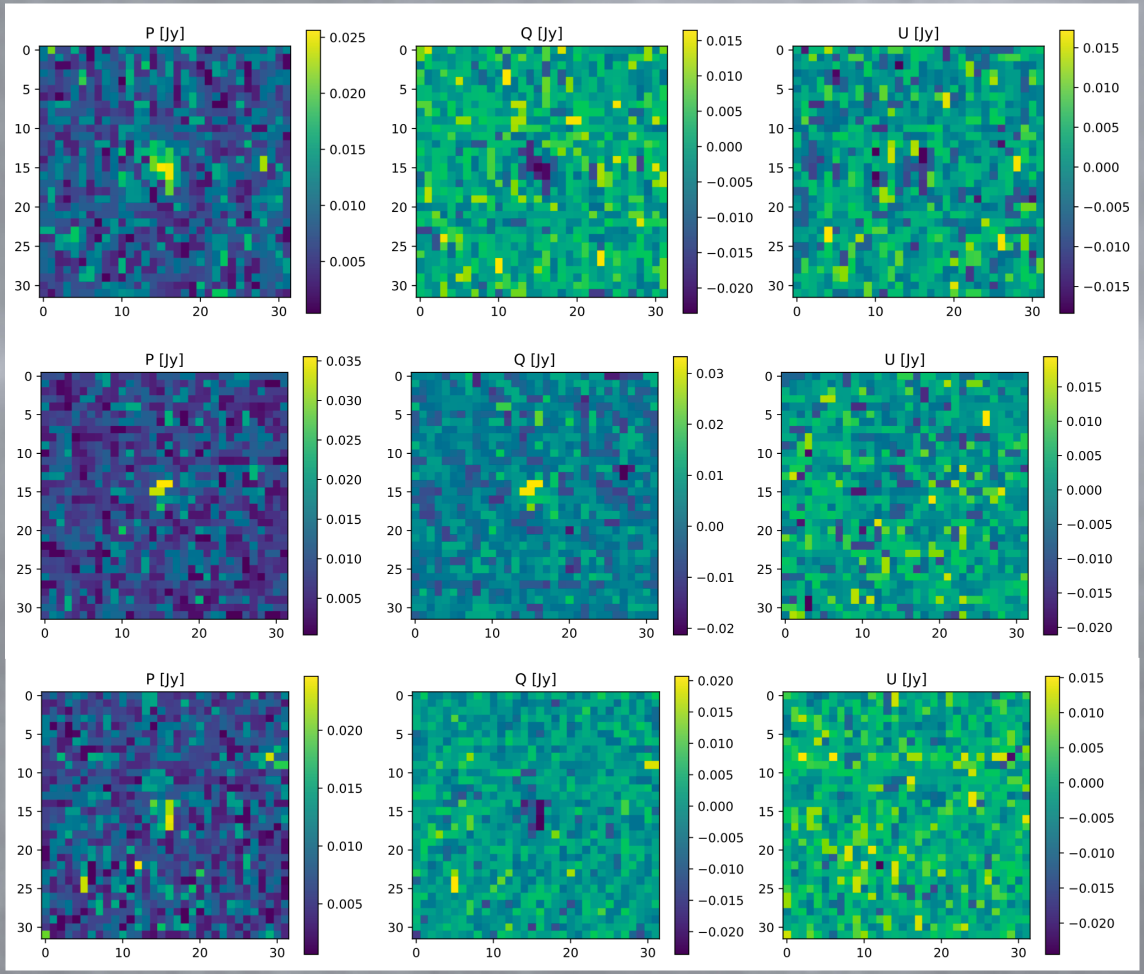
Reasonable residuals:
E recovery w/
10^{-1}-10^{-2} μK^2
B recovery w/
2×10^{-3} μK^2 at $l < 400$
5×10^{-4} μK^2 at $l > 400$

B-mode recovery
sensitive to the use of
a different foreground model

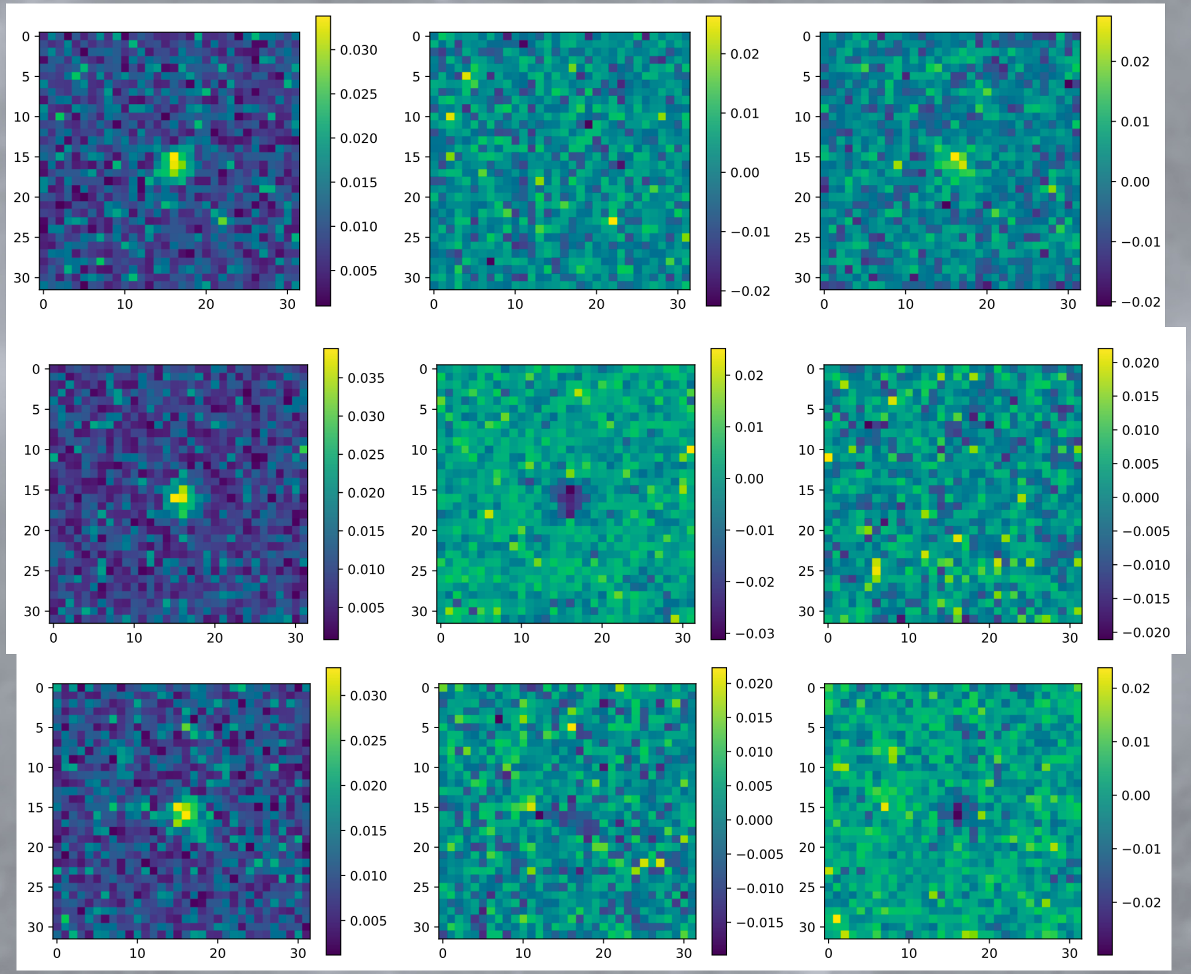
Casas+ subA&A

POSPEN to real data

Example of 3 PCCS2 srcs @ 217 GHz in polarization

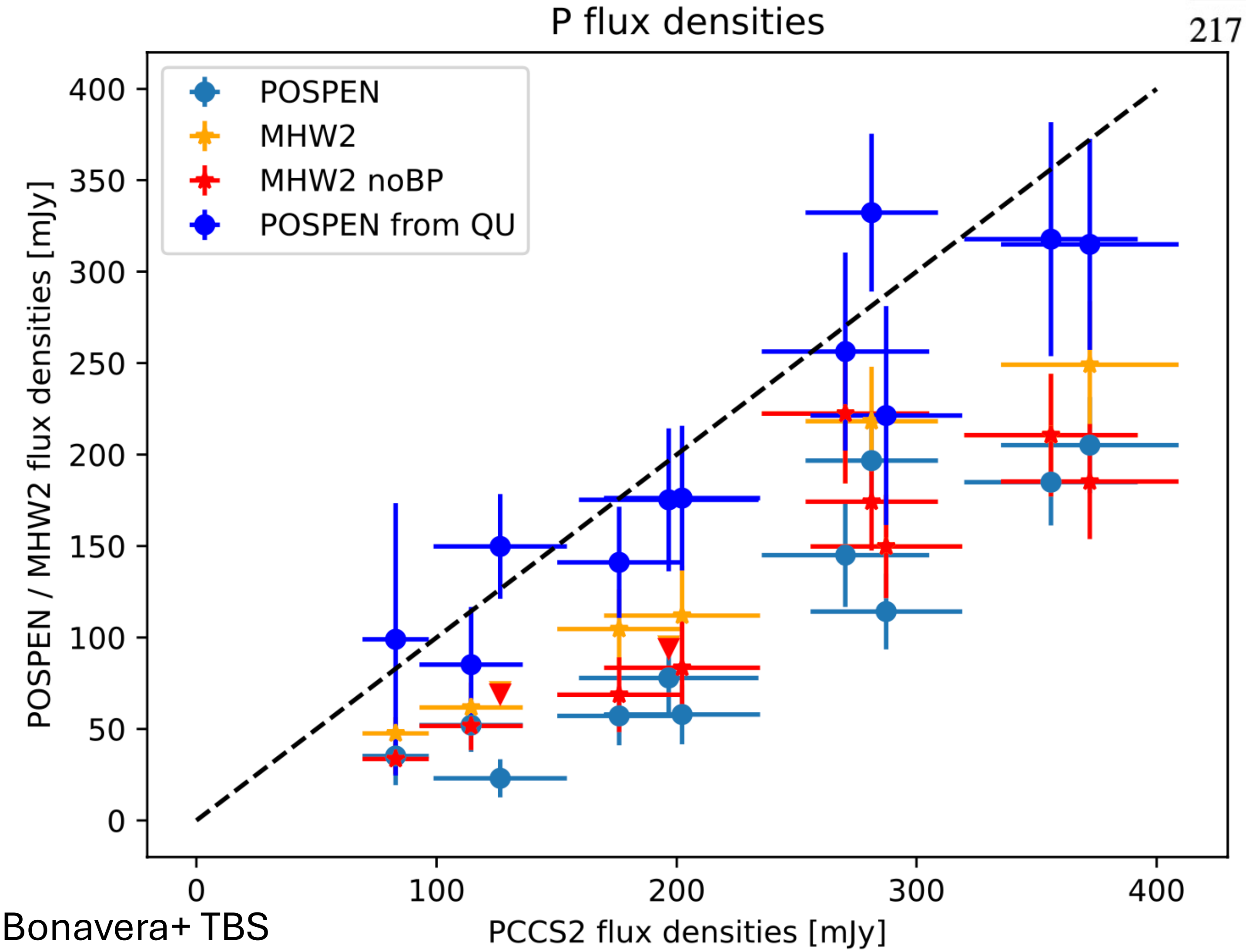


Example of 3 simulated srcs @ 217 GHz in polarization



POSPEN to real data

Channel	Flux density 90% completeness [mJy]	No. of sources		Polarized sources	
		PCCS2	PCCS2E	PCCS2	PCCS2E
217	152	2135	16842	11	325

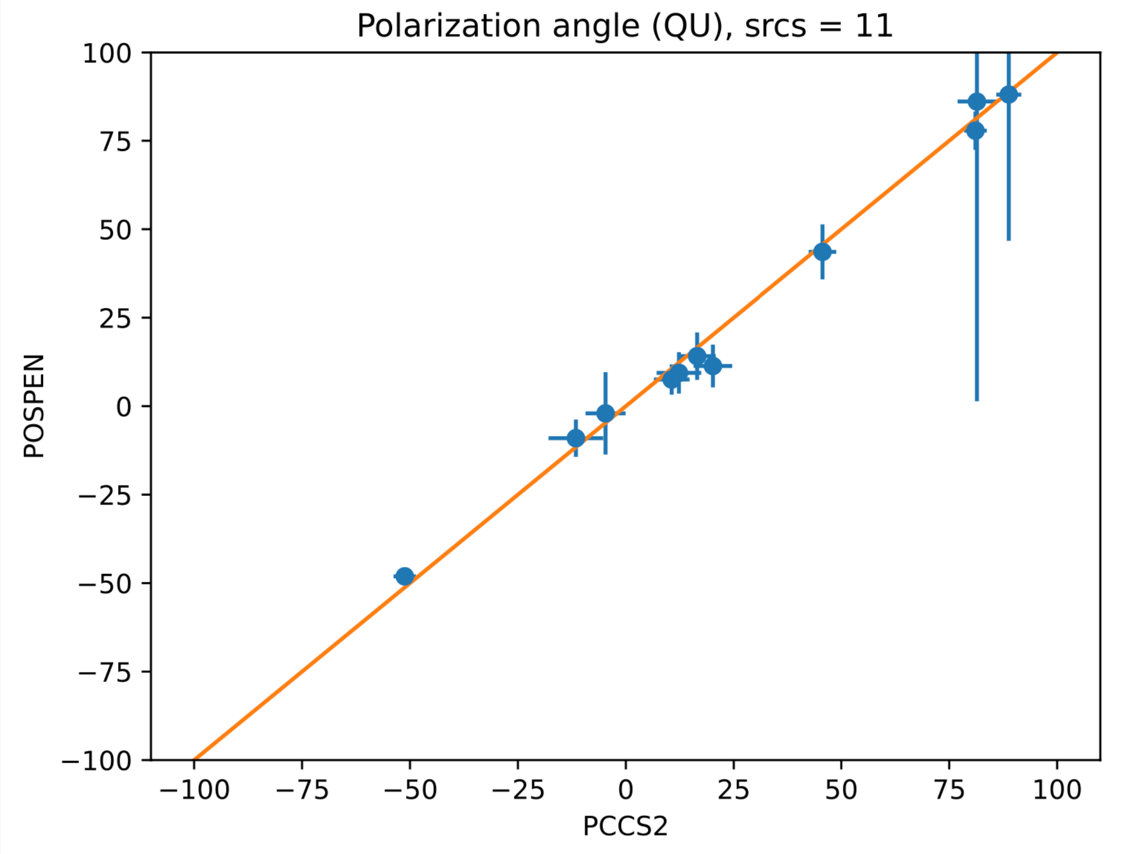


Bonavera+ TBS

POSPEN applied to the 11 srcs in the PCCS2

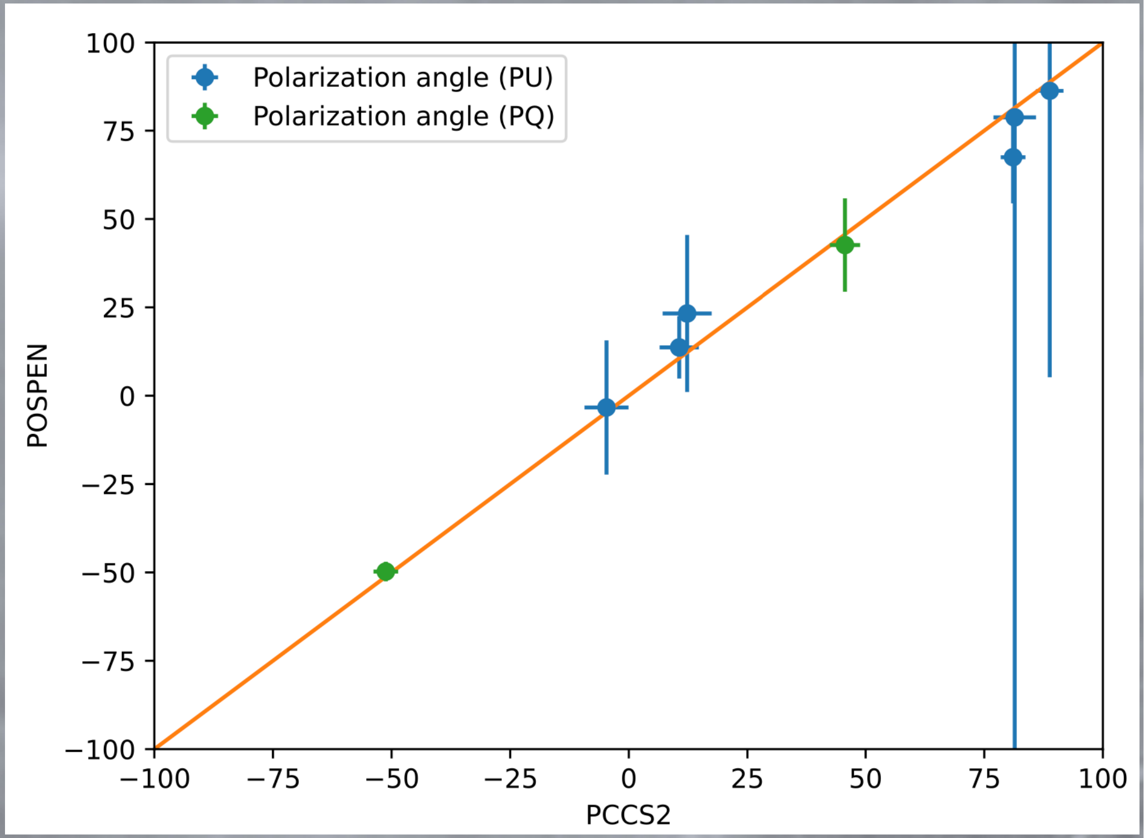
- Trained in P
- Trained in Q & U and $P_{QU} = \sqrt{Q^2 + U^2}$

POSPEN to real data



$$Q = P\cos 2\psi \qquad \psi_{PQ} = \frac{1}{2}\cos^{-1}(Q/P)$$

$$U = -P\sin 2\psi \qquad \psi_{PU} = \frac{1}{2}\sin^{-1}(-U/P)$$



IAU convention

$$\psi_{QU} = \frac{1}{2}\tan^{-1}(-U/Q)$$

Bonavera+ TBS

POSPEN to real data

Relative errors to unbias the results
Same for Q and U

Bin [Jy]	# PS	P_{QU}		P	
		μ	σ	μ	σ
>0.5	16	1.72	9.60	1.18	2.43
0.3-0.5	27	-7.38	11.68	-2.22	6.03
0.2-0.3	32	-12.54	17.50	-3.35	12.07
0.15-0.2	34	-21.78	27.89	-5.55	13.51
0.1-0.15	79	-29.95	31.03	-5.23	20.54
0.08-0.1	51	-33.79	35.65	-15.03	20.84
0.06-0.08	89	-31.49	45.02	-12.51	29.21
0.04-0.06	123	-26.20	44.59	-12.30	31.59
0.02-0.04	255	-6.78	57.44	6.79	42.14
0.0-0.02	294	136.23	215.39	109.42	127.99

Bonavera+ TBS

POSPEN to real data

Relative errors to unbias the results
Same for Q and U

Bin [Jy]	# PS	P_{QU}		P	
		μ	σ	μ	σ
>0.5	16	1.72	9.60	1.18	2.43
0.3-0.5	27	-7.38	11.68	-2.22	6.03
0.2-0.3	32	-12.54	17.50	-3.35	12.07
0.15-0.2	34	-21.78	27.89	-5.55	13.51
0.1-0.15	79	-29.95	31.03	-5.23	20.54
0.08-0.1	51	-33.79	35.65	-15.03	20.84
0.06-0.08	89	-31.49	45.02	-12.51	29.21
0.04-0.06	123	-26.20	44.59	-12.30	31.59
0.02-0.04	255	-6.78	57.44	6.79	42.14
0.0-0.02	294	136.23	215.39	109.42	127.99

Bonavera+ TBS

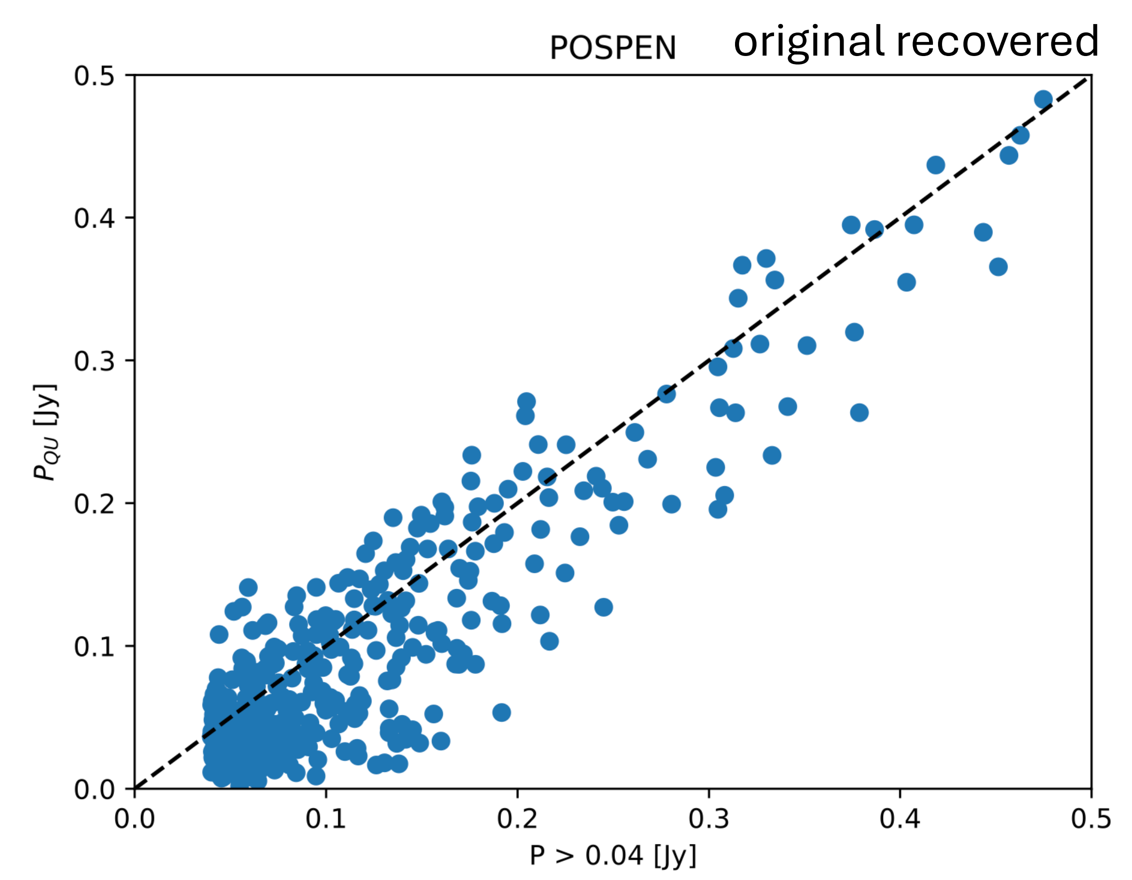
POSPEN to real data

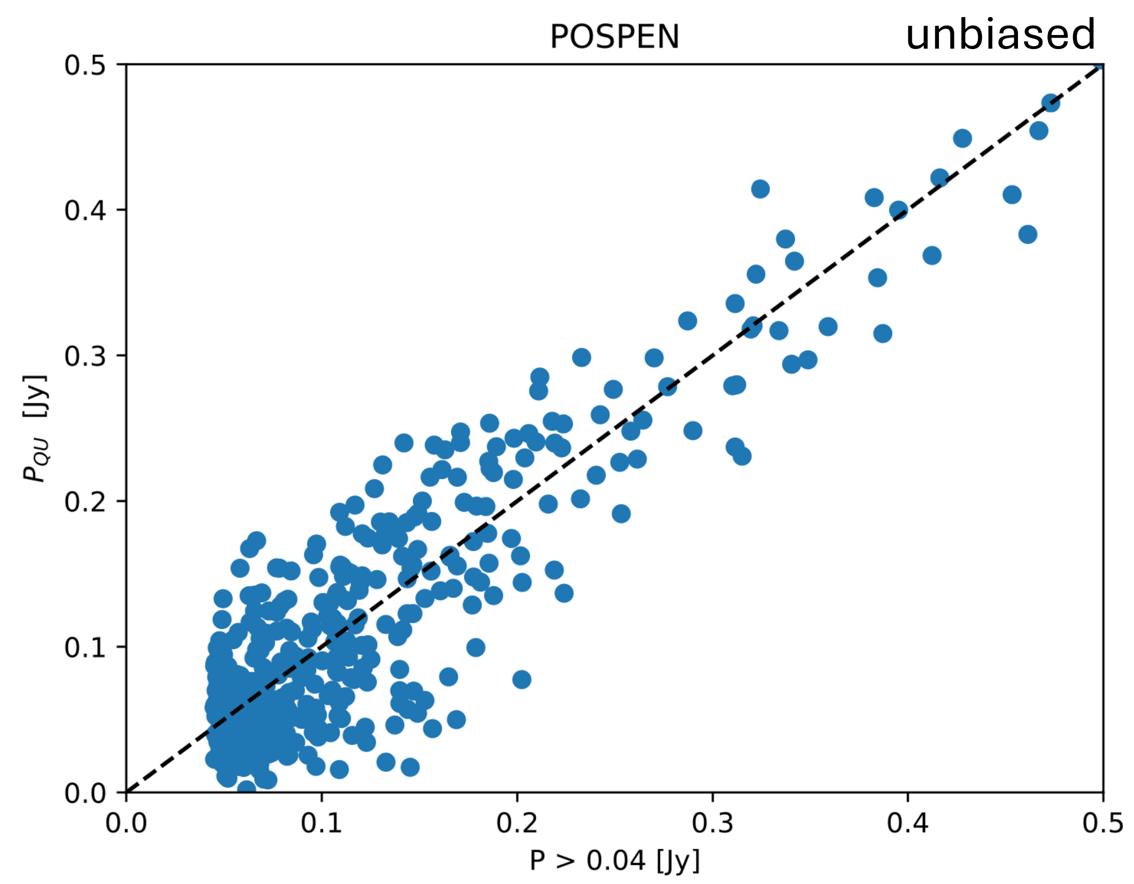
Relative errors to unbias the results
Same for Q and U

Bin [Jy]	# PS	P_{QU}		P	
		μ	σ	μ	σ
>0.5	16	1.72	9.60	1.18	2.43
0.3-0.5	27	-7.38	11.68	-2.22	6.03
0.2-0.3	32	-12.54	17.50	-3.35	12.07
0.15-0.2	34	-21.78	27.89	-5.55	13.51
0.1-0.15	79	-29.95	31.03	-5.23	20.54
0.08-0.1	51	-33.79	35.65	-15.03	20.84
0.06-0.08	89	-31.49	45.02	-12.51	29.21
0.04-0.06	123	-26.20	44.59	-12.30	31.59
0.02-0.04	255	-6.78	57.44	6.79	42.14
0.0-0.02	294	136.23	215.39	109.42	127.99

Bonavera+ TBS

POSPEN to real data





P vs P_{QU} estimations (srcs w/ P>40 mJy)

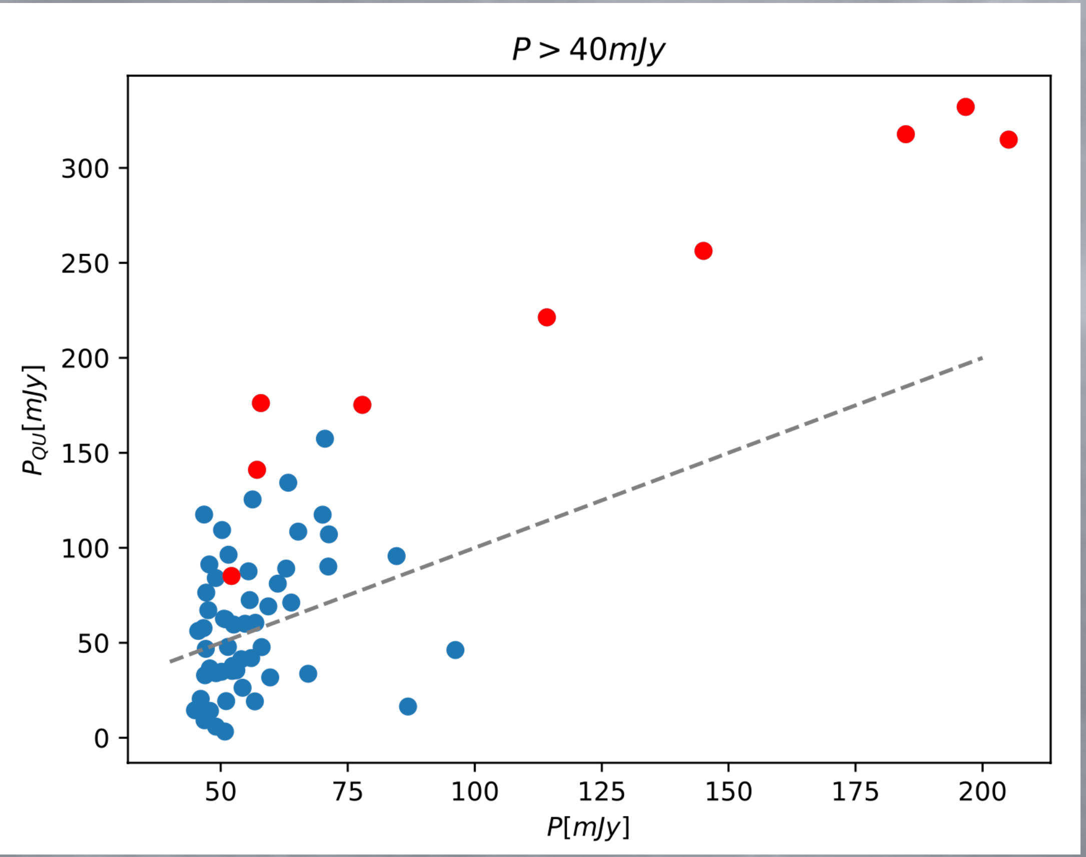
Bonavera+ TBS

POSPEN to real data

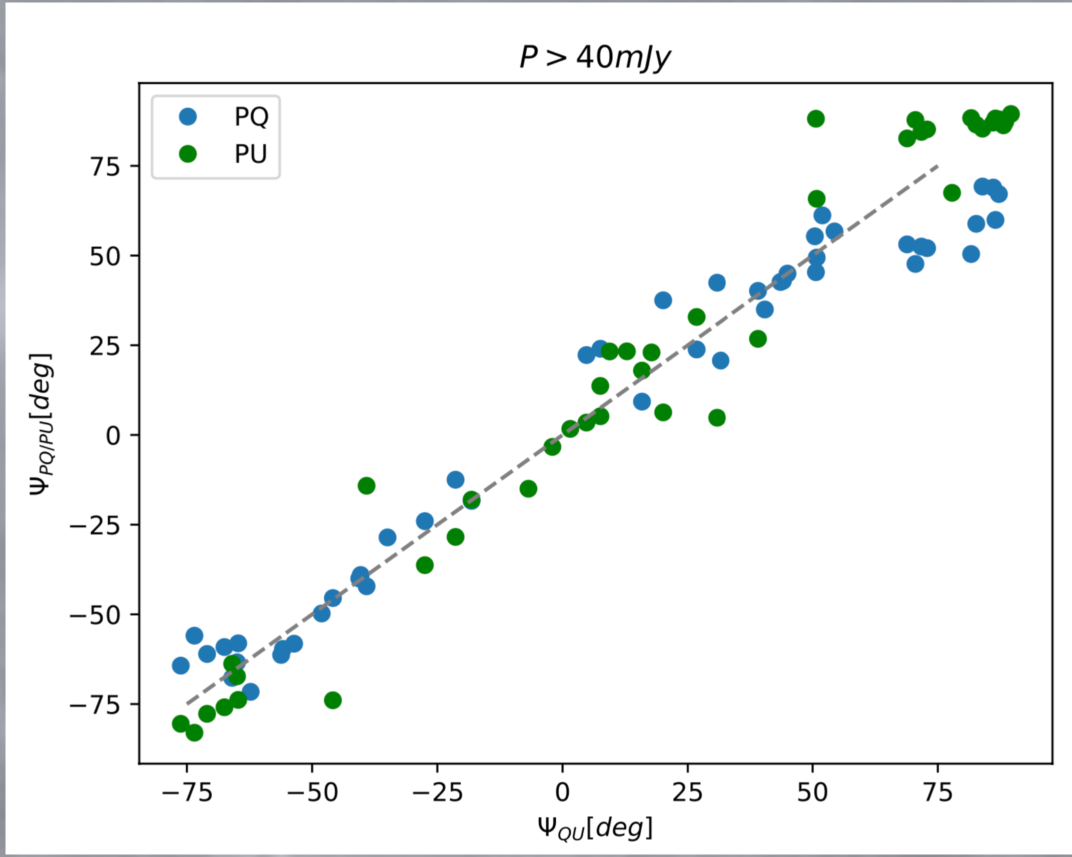
Channel	Flux density 90% completeness [mJy]	No. of sources PCCS2	No. of sources PCCS2E	Polarized sources PCCS2	Polarized sources PCCS2E
217	152	2135	16842	11	325

POSPEN applied to PCCS2 positions

Selection P_{raw} > 40 mJy: 63 sources (>5 times PCCS2 srcs)



P estimations comparisons



Bonavera+ TBS

Pol. angle estimations comparison

Conclusions

- NN reliable methods for PS detection and CMB recovery in T and P
- NN reliable methods also for foreground characterization in T and P
- More flexible and automatic methods
- Very suitable for future experiments providing larger amounts of data
- Not a “filter” (no Fourier space), then no ringing or border effects

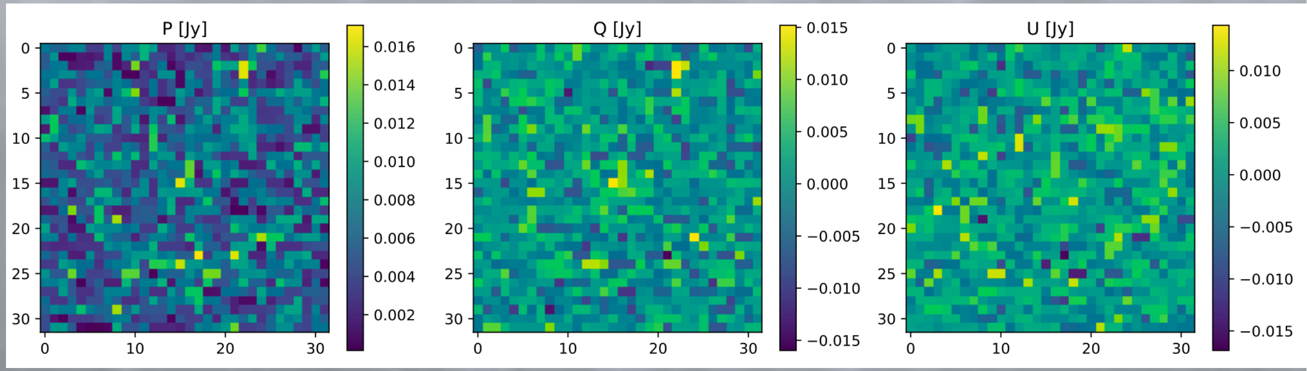
Conclusions – PS

Simulations

- Lower flux densities limit: larger number of detected PS
- Smaller number of spurious detections
- Multi-frequency methodology, very important for spectral characterization of galaxies
- Estimation of polarization angle for even not so bright PS

POSPEN 4 Planck

- Not stightforward application
- Discrepancies bw P & P_{QU} ?
- (hot pixels issue?)



23 mJy 142 mJy 47 mJy

Conclusions - CMB

- No PS mask needed
- No mask needed to avoid strong Galactic contamination regions
- Better performance at small scales
- Best performance when trained w/ lower noise
- Possible bias when trained w/ no accurate simulated sky components (also for traditional methods):
    - ➢ Train various NN w/ different simulated diffuse components and use Ensemble Learning