SKA Data Challenge 3a

results & goals

SKACH

Michele Bianco
Tianyue Chen
Shreyam Krishna

Rohit Sharma

Sambit Kumar Giri

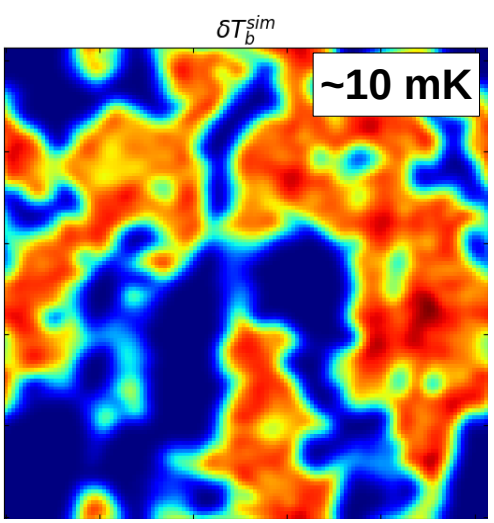
Philipp Denzel

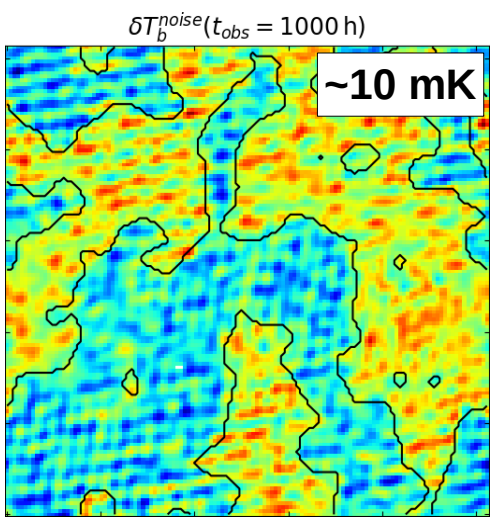
Hatem Ghorbel
Massimo De Sanctis

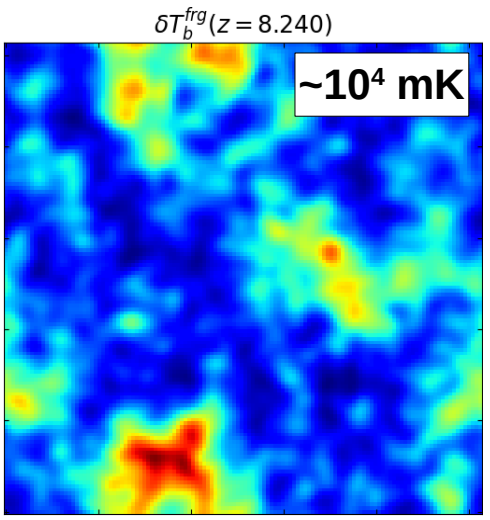
Viraj Nistane
Christopher Finlay

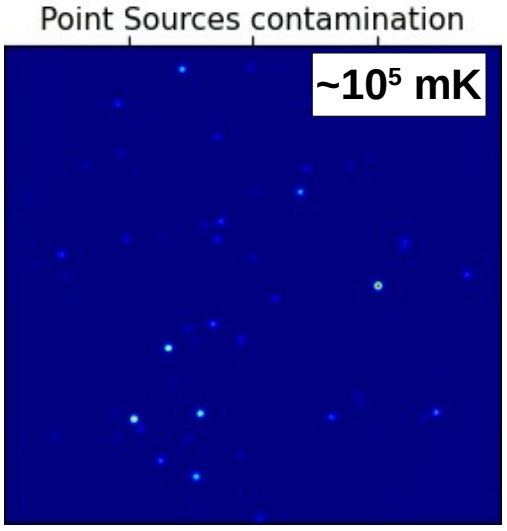
Real Space: 21-cm Observations

Currently we can create EoR mock observation with a combination of numerical models for 21-cm, systematic noise and foregrounds.







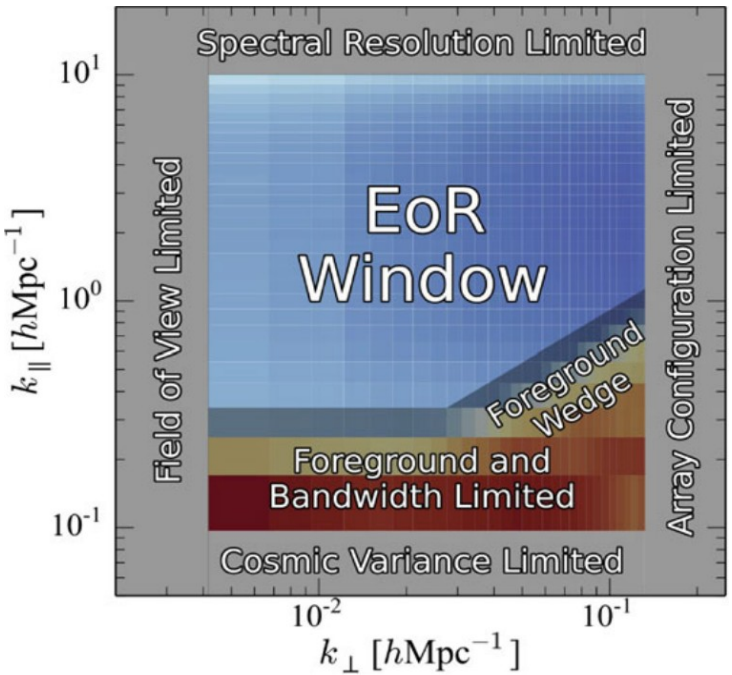


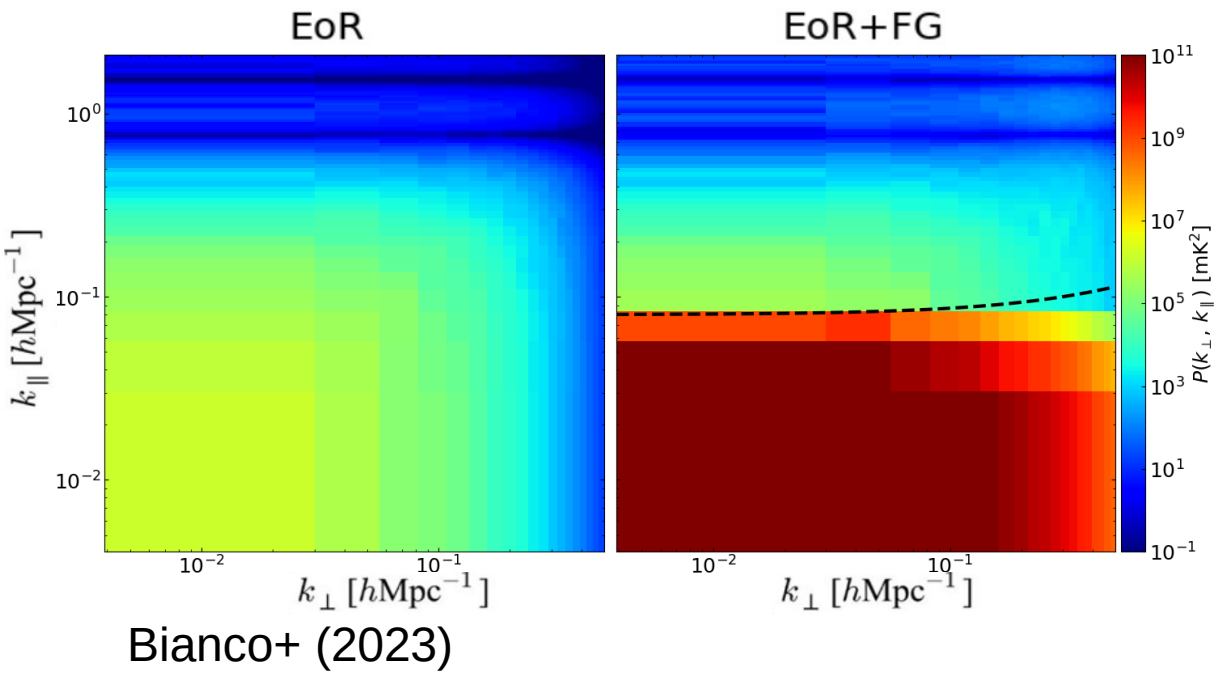
<u>Goal:</u>

Recover 21-cm differential brightness and the distribution of neutral hydrogen from SKA-Low mock observations.

Fourier Space: 21-cm 2D Power Spectra

Fourier transformation of the 21-cm lightcone tomographic dataset provide the 2D power spectra.





Bianco+ (2023)

Remove k-modes contaminated by foreground with **avoidance** technique model for **substract** the foreground contribuiton

SDC3 Data

Data stored on Piz Daint, stored at: **/store/ska/sk01**

Image Products/Files

- ZW2.msw_image.fits → uniform weighting data cube
- ZW2.msw_psf.fits → uniform weighting psf
- ZW2.msn_image.fits → natural weighting data cube
- ZW2.msn_psf.fits → natural weighting psf
- station_beam.fits → time-averaged station beam cube
- ZW2.%s_image.md5 → checksum file for ZW2.%s_image.fits

15 GB

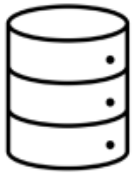


Image cube
Data integrated over time

7.5 TB

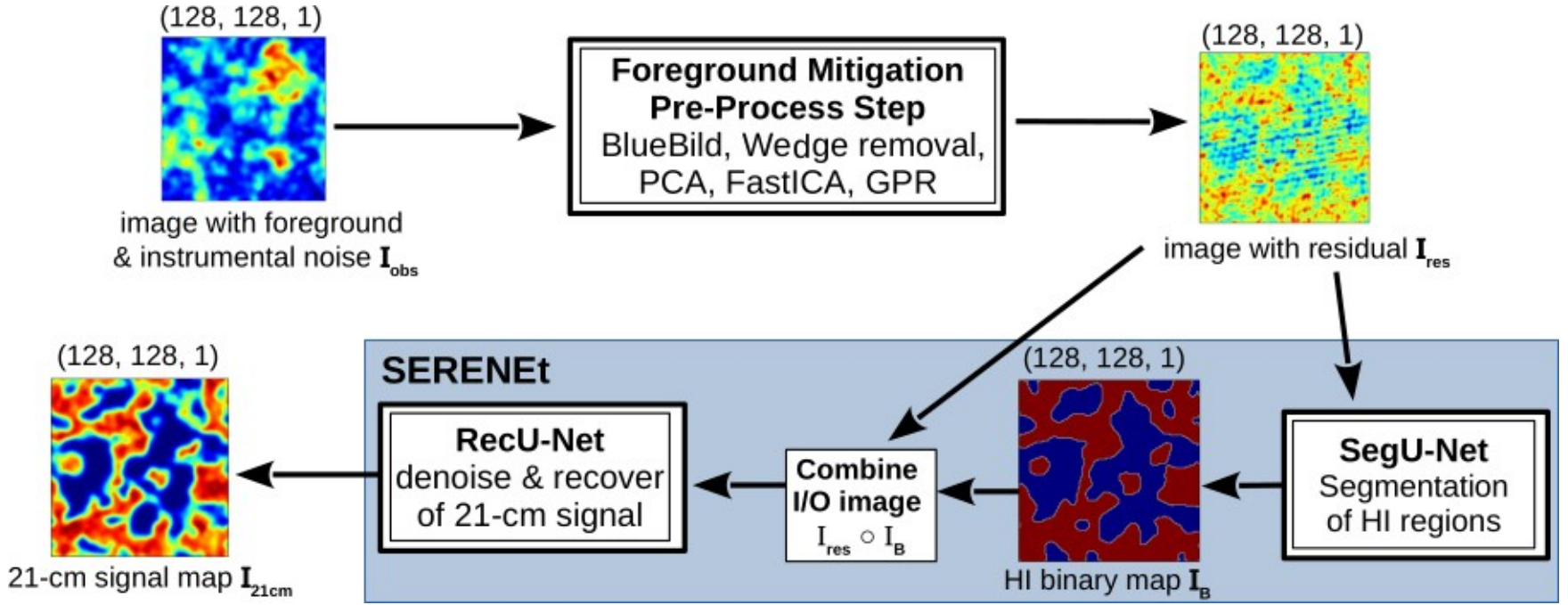
Measurement sets visibilities

- ZW2_IFRQ_**%04d**.ms → 900 MS at frequency channel

Calibrated visibilities
Data in 1400 "timesteps"

SERENEt
SEgmentation and **RE**gression **NEt**work

Combine the predicted binary maps of **SegU-Net** as additional input of **Rec-Unet** training step in order to include prior in the network training.



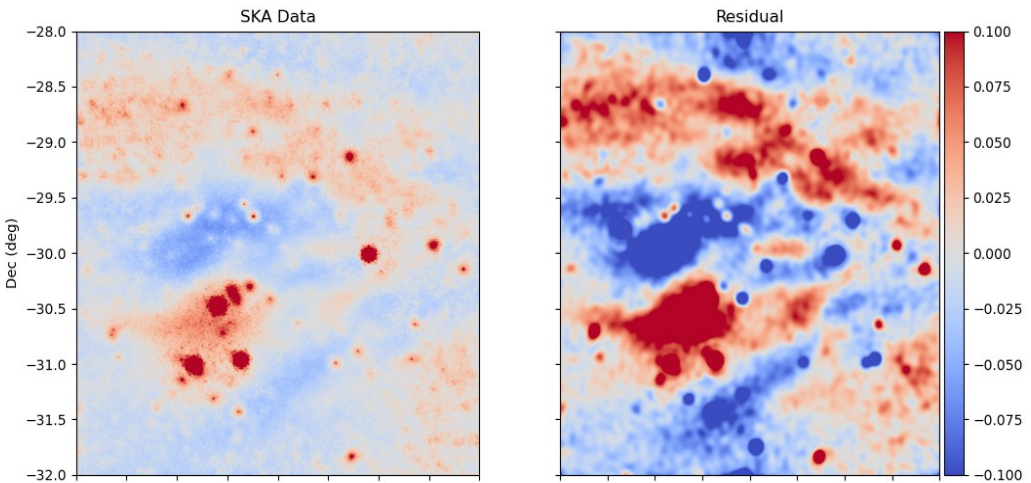
SKACH **840k core-h** and **GPU-h** allocation projects at Pitz Daint @ CSCS

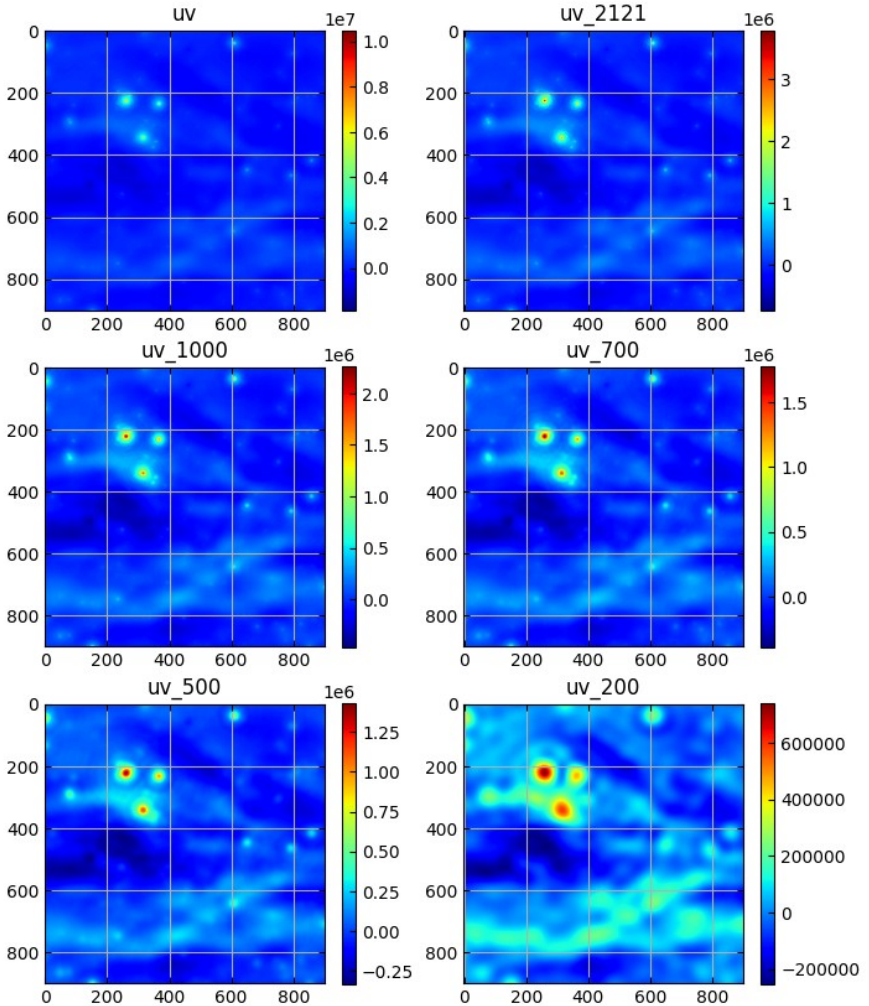
SERENEt Pre-Process Step

Application of pre-process to tomographic data images for 166–181 MHz.

- Point source subtraction:

$$V_{21\text{-cm}} \approx V_{21\text{cm+point}} - V_{\text{point}}$$

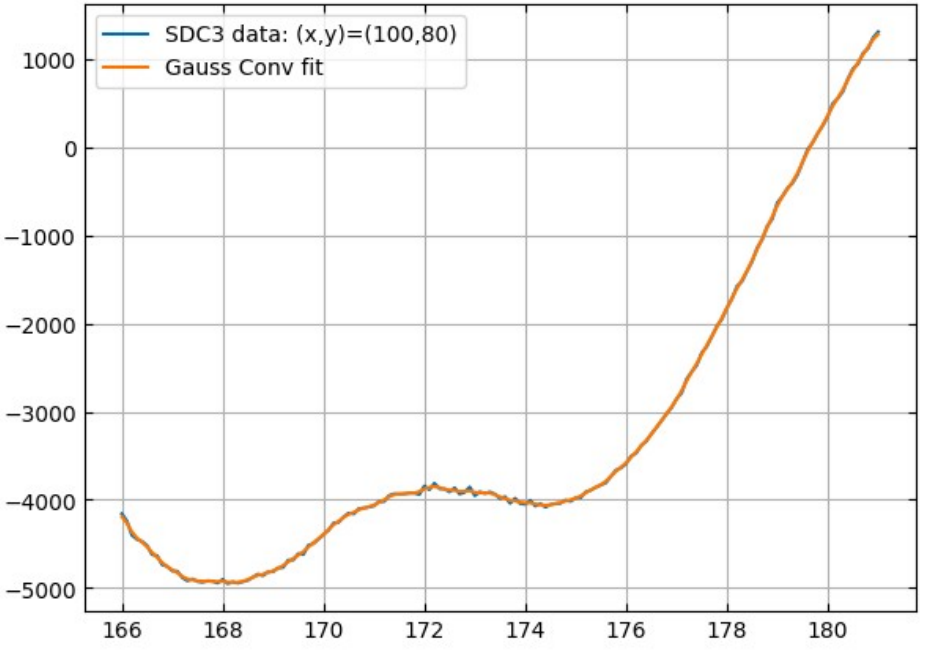


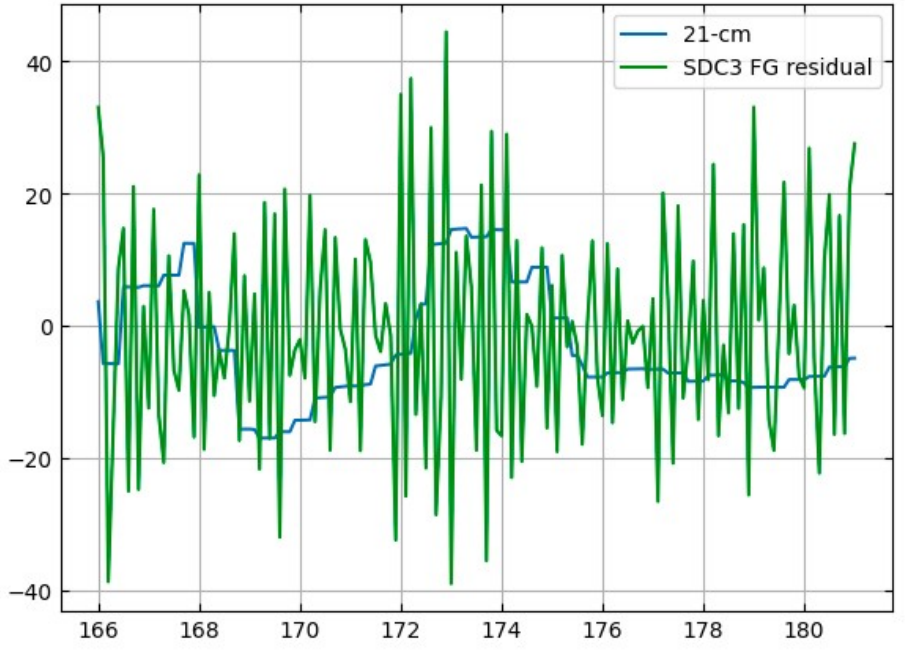


SERENEt Pre-Process Step

Application of pre-process to the SKACH simulated data for the SDC3 tomographic data images for 166–181 MHz.

- Polynomial fitting of the signal in the frequency direction (LoS)

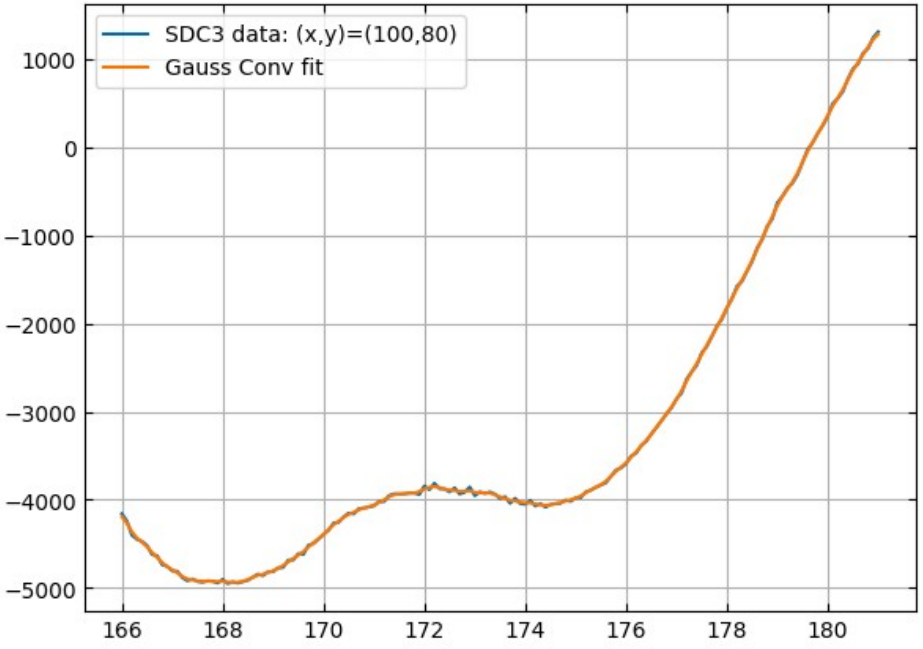


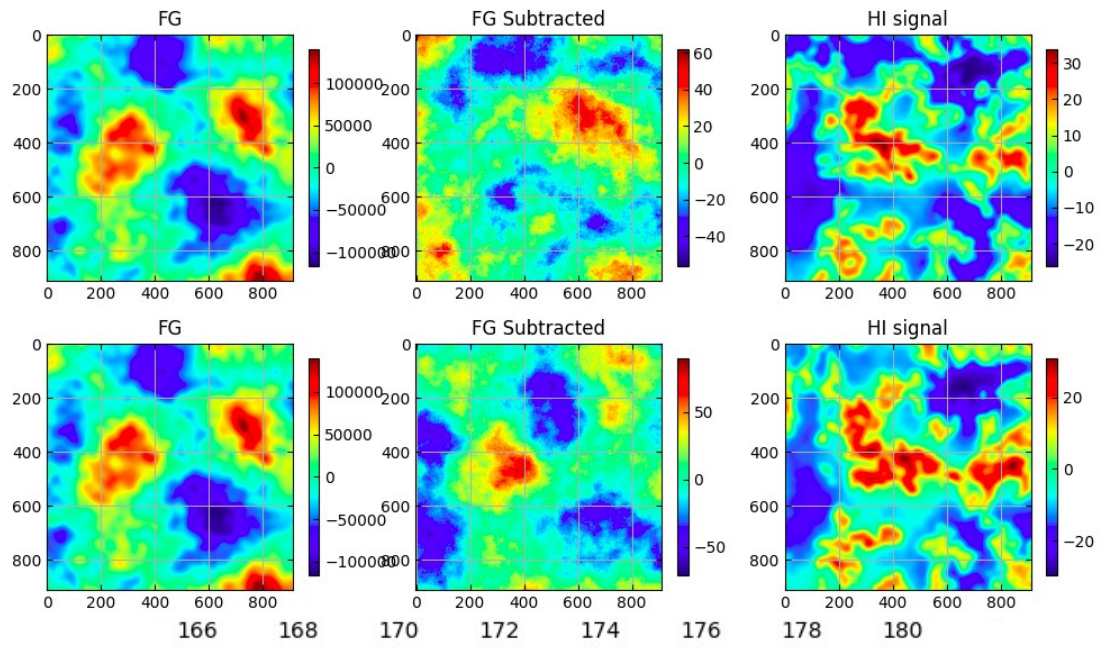


SERENEt Pre-Process Step

Application of pre-process to the SKACH simulated data for the SDC3 tomographic data images for 166–181 MHz.

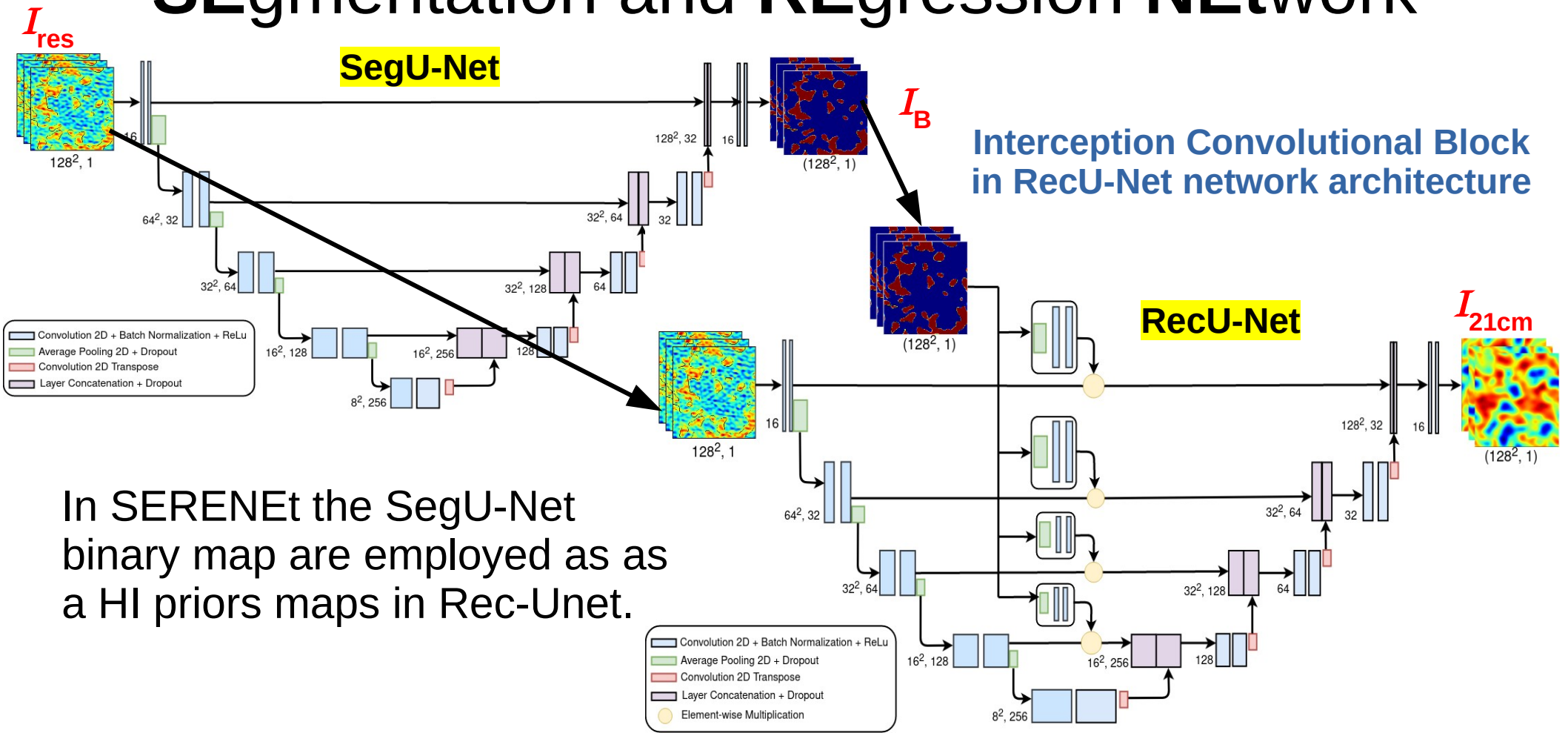
- Polynomial fitting of the signal in the frequency direction (LoS)





SERENEt

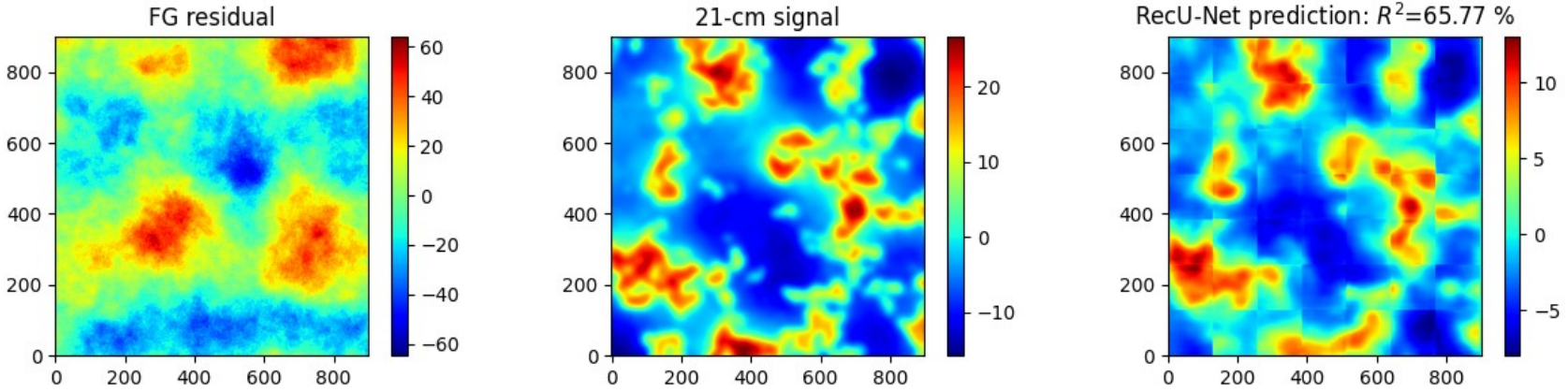
SEgmentation and REgression NEtwork



In SERENEt the SegU-Net binary map are employed as as a HI priors maps in Rec-Unet.

SERENEt prediction

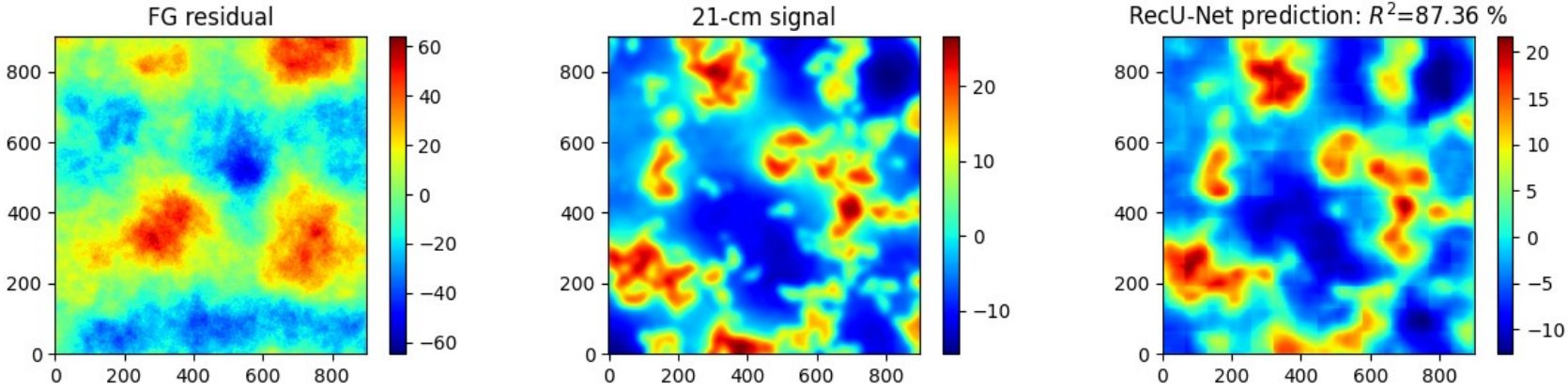
Prediction on the SKACH simulated data at 166–181 MHz.



Iterative prediction with trained SERENEt neural network and for overlapping regions on the same 21-cm image at a given frequency.

SERENEt prediction

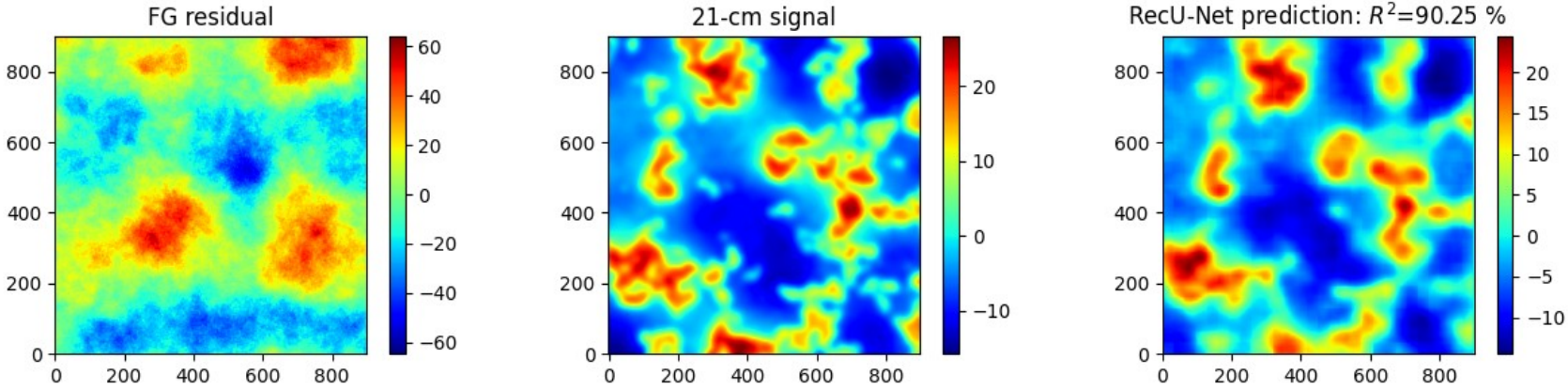
Prediction on the SKACH simulated data at 166–181 MHz.



Iterative prediction with trained SERENEt neural network and for overlapping regions on the same 21-cm image at a given frequency.

SERENEt prediction

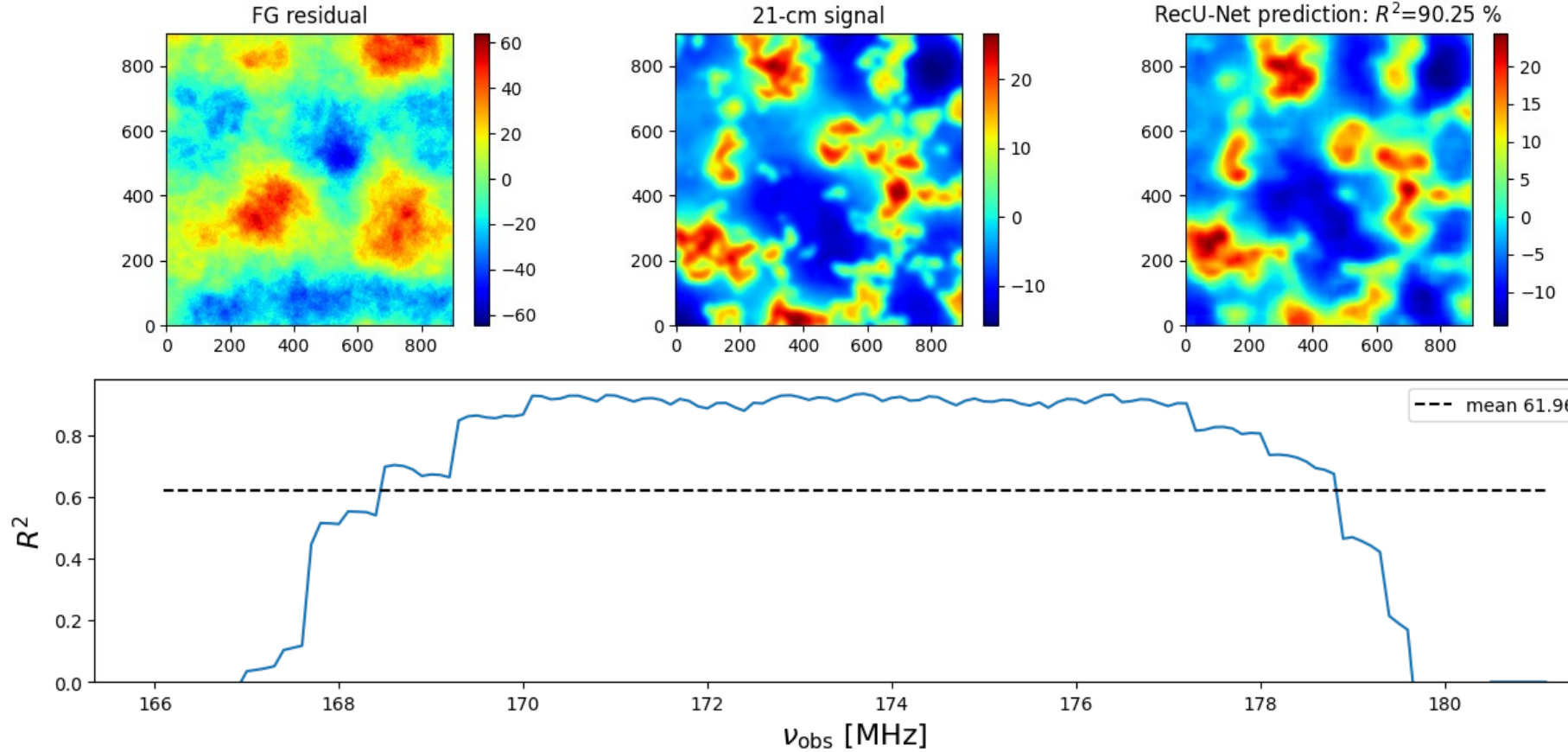
Prediction on the SKACH simulated data at 166–181 MHz.



Iterative prediction with trained SERENEt neural network and for overlapping regions on the same 21-cm image at a given frequency.

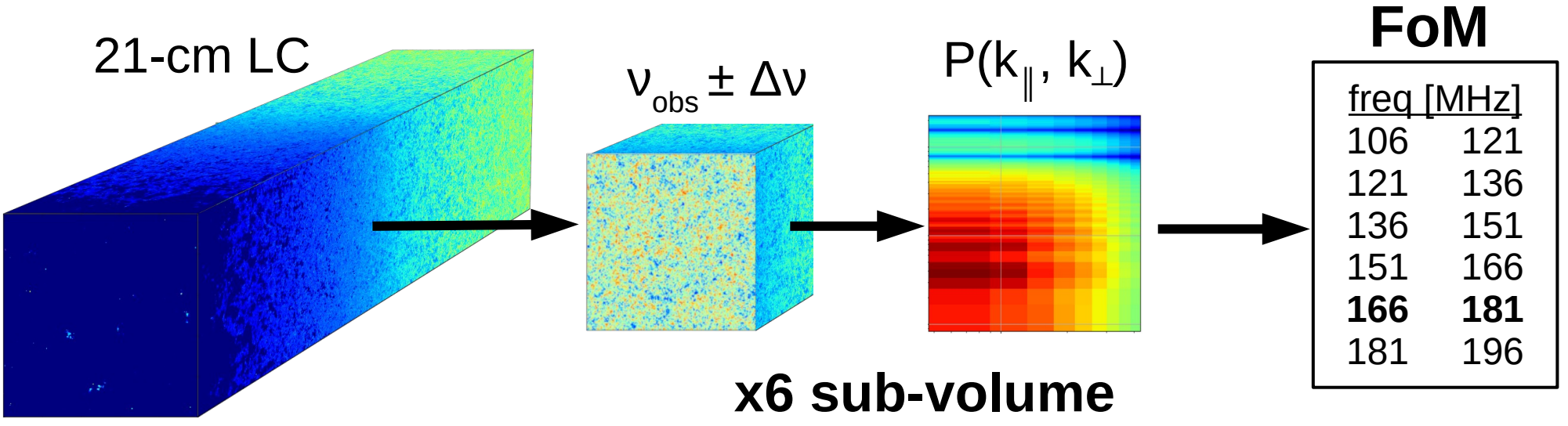
SERENEt prediction

Prediction on the SKACH simulated data at 166–181 MHz.



SDC3 Scoring System

The Figure of Merit (FoM) based on *"metric distance"* between the true and recovered spectra for **6 sub-volume** at a central frequency ν_{obs} and width $\Delta\nu$



Science Data Challenge 3

Overview

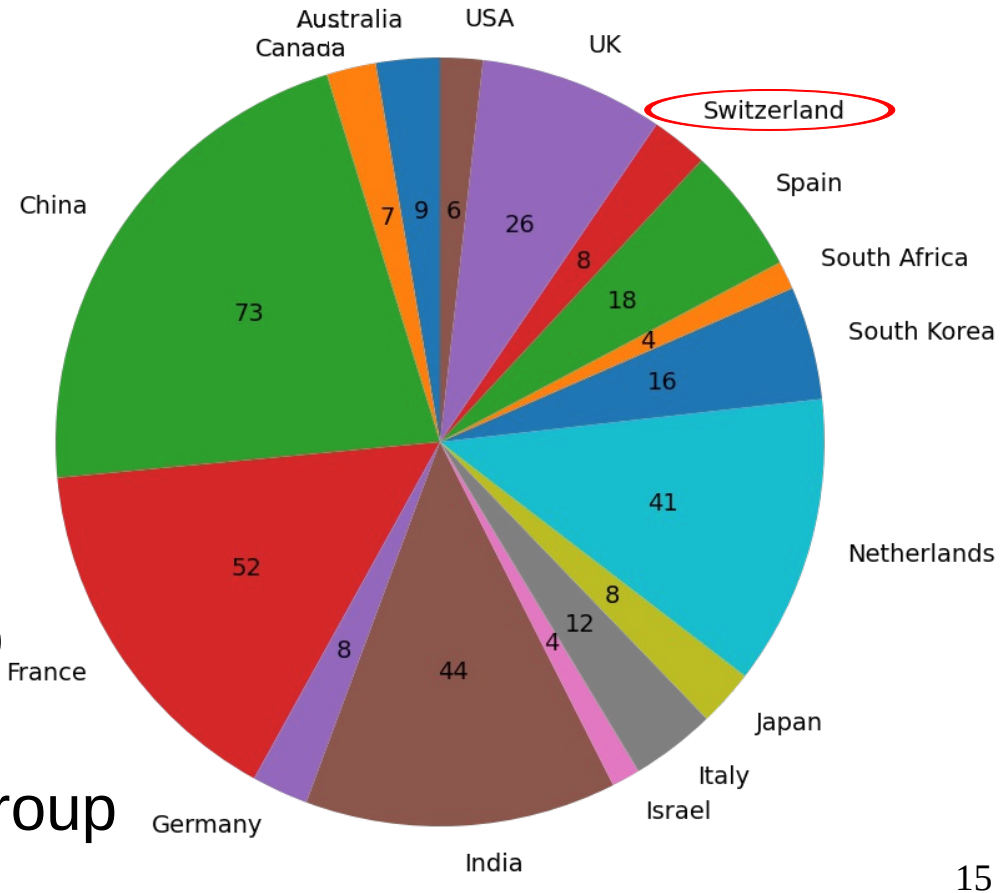
A total of 33 teams partecipated

for a total of 336 partecipants

from 16 different nations.

Majority of the contestants from:

- China (73)
- France (52): LoTSS LOFAR group
- India (44): GMRT group in Pune
- Netherland (41): LoTSS LOFAR group



Science Data Challenge 3

Overview

SKACH team representing Switzerland SKA consortium

Gain Group

Tianyue Chen
Rohit Sharma
Michele Bianco
Shreyam Krishna
Chris Finlay

$$g_i \equiv 1 + q_{i\nu}$$
$$I^{noise}(\sigma_{rms},\ \nu)$$

Points Group

Michele Bianco
Philipp Denzel
Viraj Nistane

$$I^{points}(\hat{\mathbf{n}},\ \nu)$$

ML Group

Michele Bianco
Tianyue Chen
Hatem Ghorbel
Massimo De Sanctis

$$\mathcal{L}(y_{true},\ y_{pred})$$

Analysis Group

Michele Bianco
Sambit Giri
Rohit Sharma

Rank	Team	Score
1	HIMALAYA	74758
2	DOTSS-21cm_ML-GPR	71573
3	DOTSS-21cm_Advanced_ML-GPR	71135
4	ERWA	63670
5	DOTSS-21cm_Avoidance	51889
6	Shuimu-Tianlai	43422
7	Wizards_of_Oz_3D	33295
8	Akashganga	31864
9	REACTOR	21888
10	SKACH	12103
11	KUSANAGI	
12	Cantabrigians	
13	Hausos	
14	KUSANAGIb	
15	Nottingham-Imperial	
16	Pisano_Galaxy_Moppers	
17	HAMSTER	
18	Foregrounds-FRIENDS	
19	KORSDC	
20	SROT	

Science Data Challenge 3

Overview

China – Shingai Group

France, India & Netherland – LoTSS LOFAR

Australia – MWA group

India – GMRT Pune group(?)

SKACH representing Switzerland

UK – Cavendish Radio Cosmology group

UK – Notthingham & London Imperial group

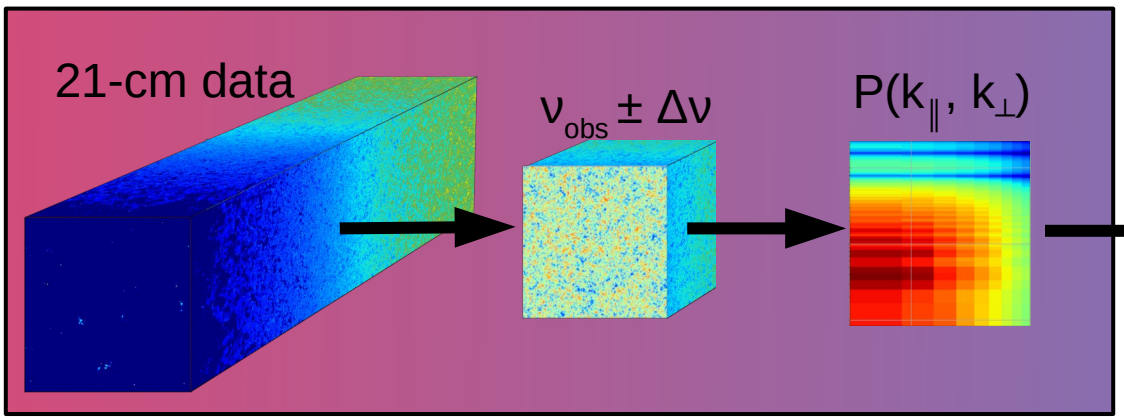
Italy & UK – Scuola Normale Superiore and Imperial group

1	HIMALAYA	74758
2	DOTSS-21cm_ML-GPR	71573
3	DOTSS-21cm_Advanced_ML-GPR	71135
4	ERWA	63670
5	DOTSS-21cm_Avoidance	51889
6	Shuimu-Tianlai	43422
7	Wizards_of_Oz_3D	33295
8	Akashganga	31864
9	REACTOR	21888
10	SKACH	12103
11	KUSANAGI	
12	Cantabrigians	
13	Hausos	
14	KUSANAGIb	
15	Nottingham-Imperial	
16	Pisano_Galaxy_Moppers	
17	HAMSTER	
18	Foregrounds-FRIENDS	
19	KORSDC	
20	SROT	

SDC3 – Foreground & Inference

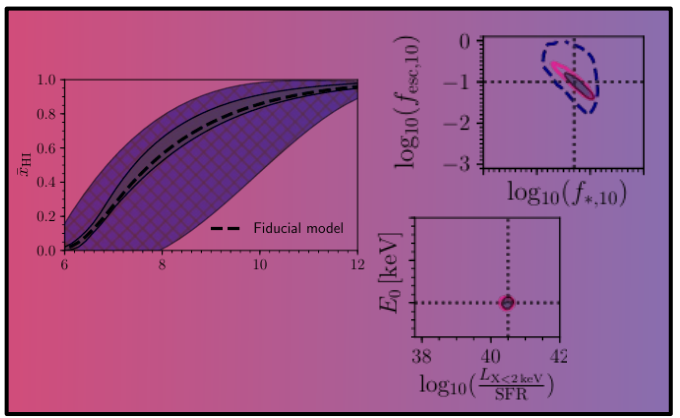
The goal of the SDC3 is to simulate the full pipeline of SKA-Low data calibration and cosmological & astrophysical inference.

SDC3a - Foreground



- Took place between Mar – Set 2023
- Ongoing: writing of a SDC summary results
- (SKACH team separated paper publication)

SDC3b - Inference



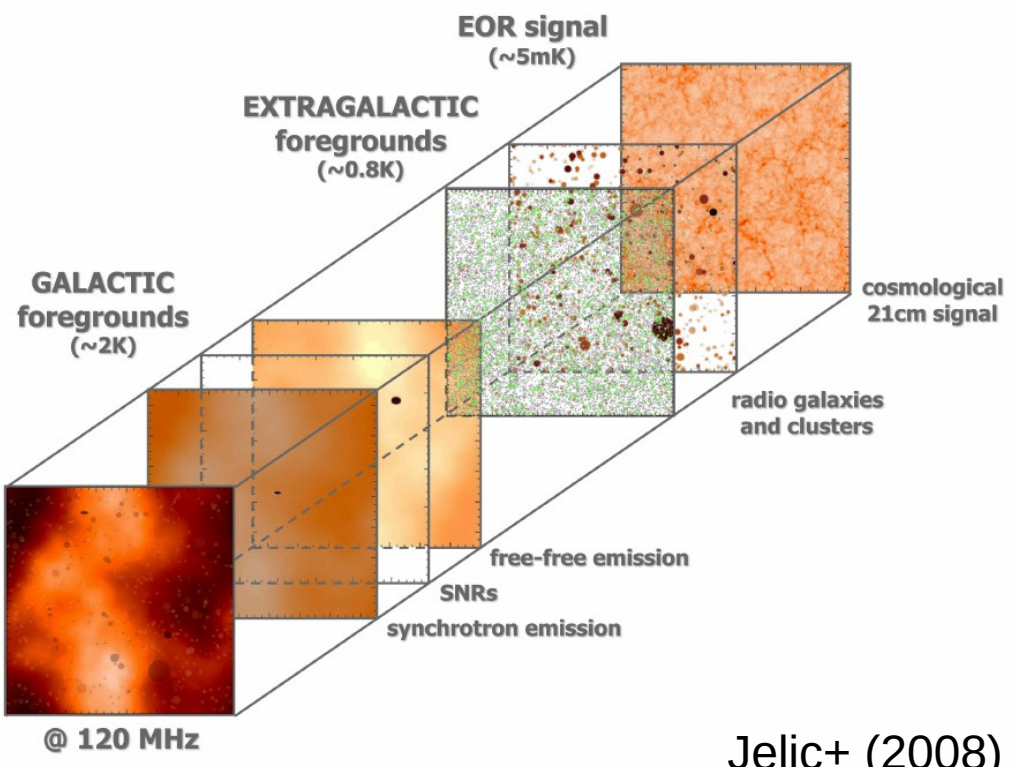
- Registration expected 2024(?)
- SDC3b start in 2025(?)
- Ongoing discussion: data sintetic

Support Slides

Tomographic Imaging of the 21-cm signal

SKA1-Low tomographic images of redshifted 21-cm signal challenges:

- Instrumental noise (signal ~ 5 K)
- Foreground emission (signal ~ 1 - 1000 K)
- Antennas gain errors
- Ionospheric refraction effects
- Radio frequency interference
- And more ...



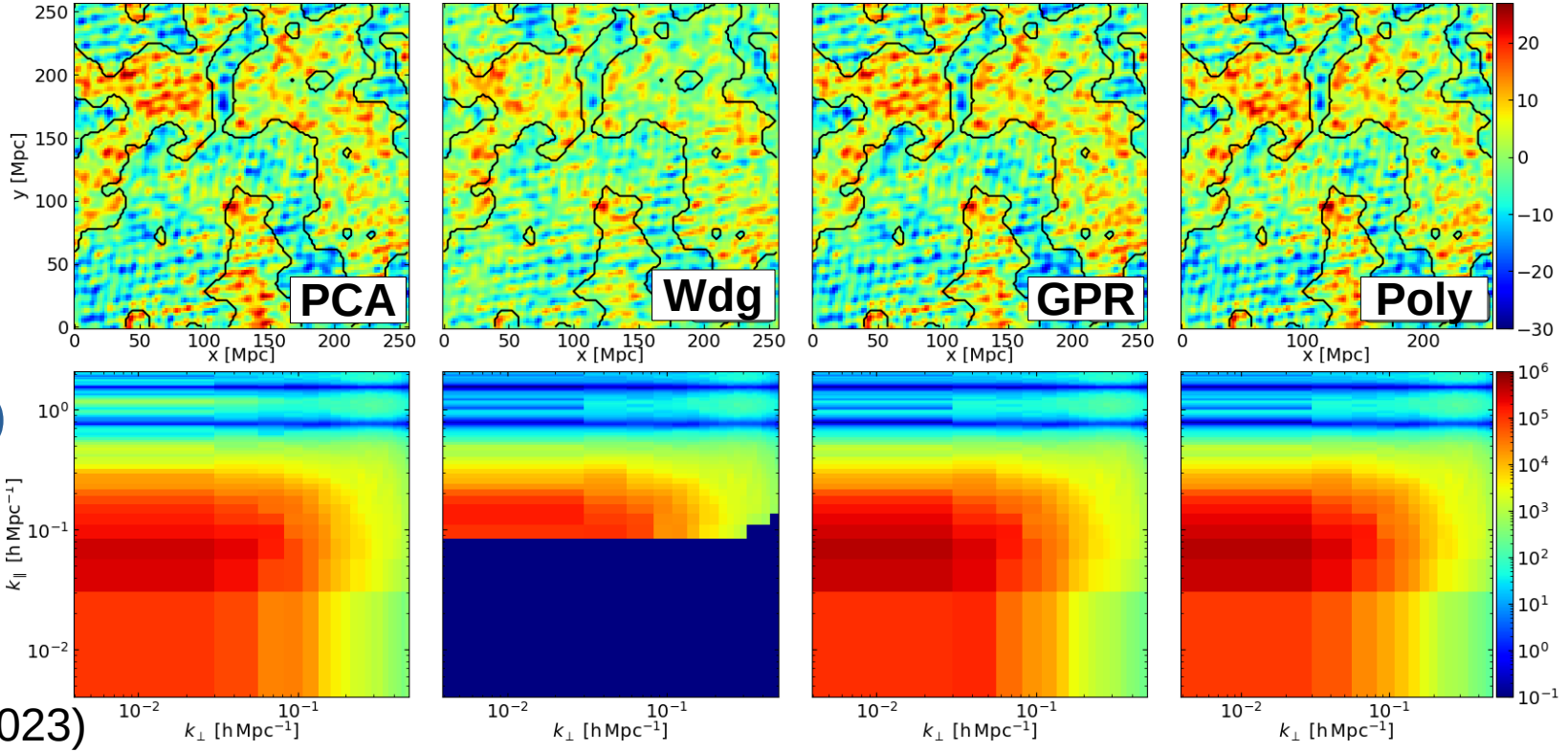
Jelic+ (2008)

Pre-process: Foreground Mitigation & Avoidance

Residual image and 2D P_k tomographic sub-volume centered at redshift **z = 8.25** (x_{HI}~0.5) and **Δν = 20 MHz**

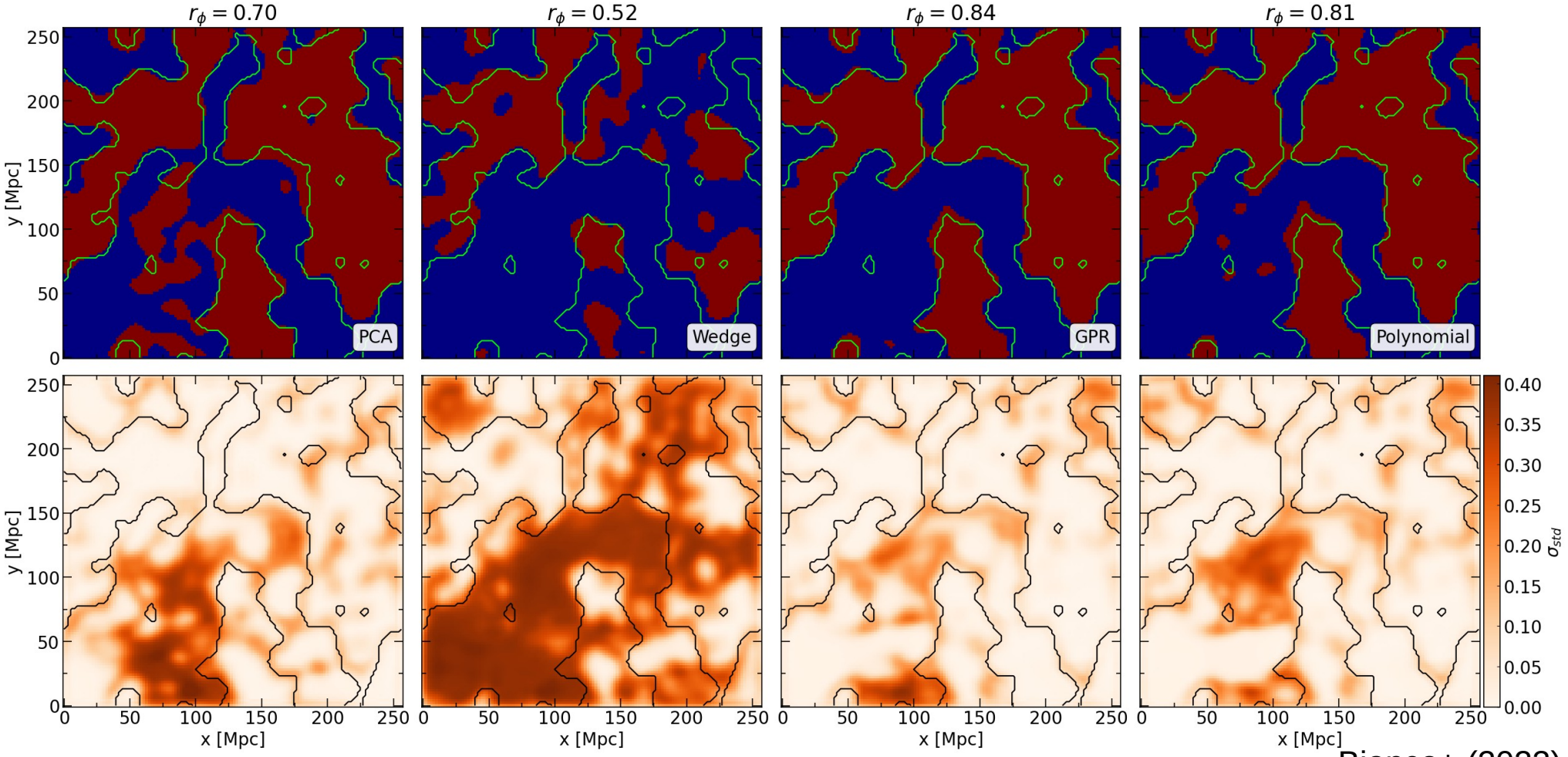
I_{res}

$P(k_\perp, k_\parallel)$



Bianco+ (2023)

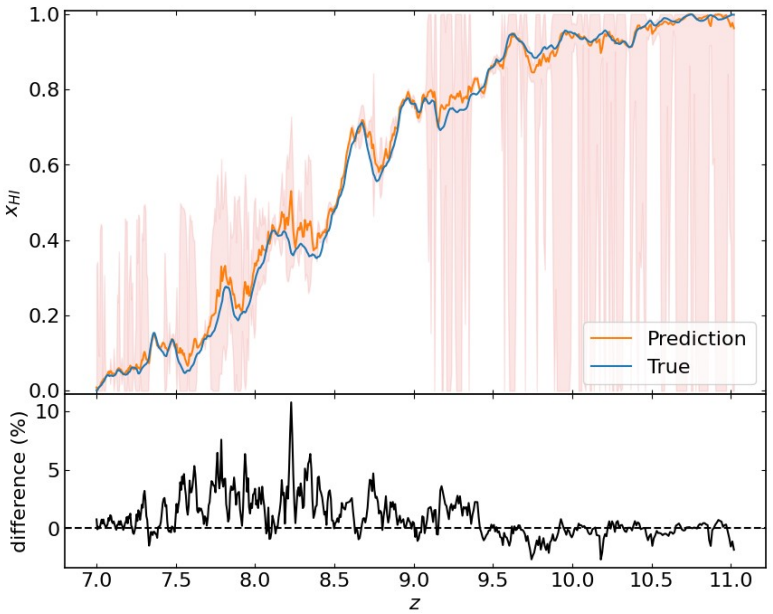
SegU-Net: Visual Comparison

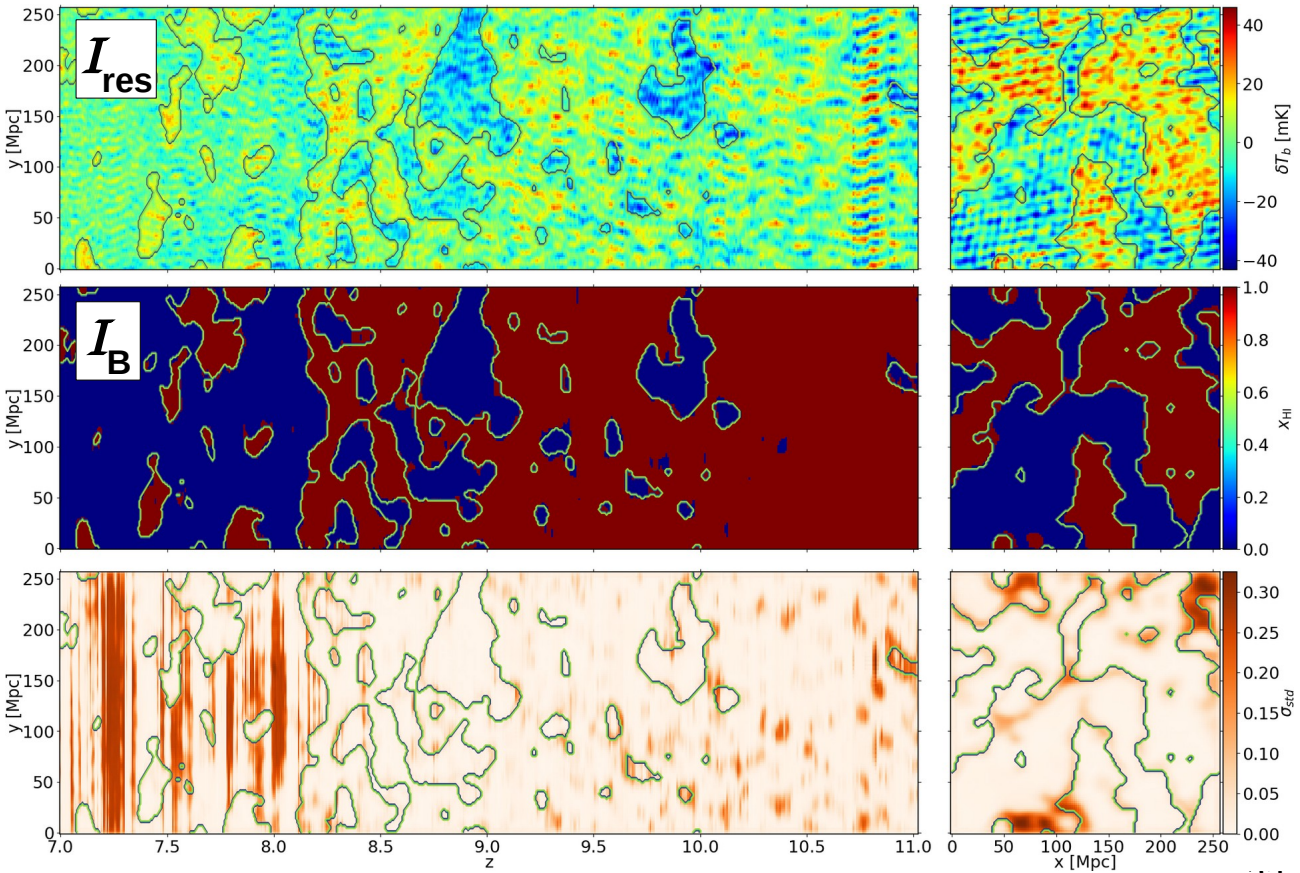


Bianco+ (2023)

SegU-Net: Tomographic Data & Reionization History

z_c	pre-process	$\overline{x}_{\rm HI}$
8.24	Ground Truth	**0.45**
	all z PCA	0.48 ± 0.07
	PCA	0.49 ± 0.11
	Wedge	0.16 ± 0.13
	GPR	0.48 ± 0.09
	Polynomial	0.49 ± 0.10





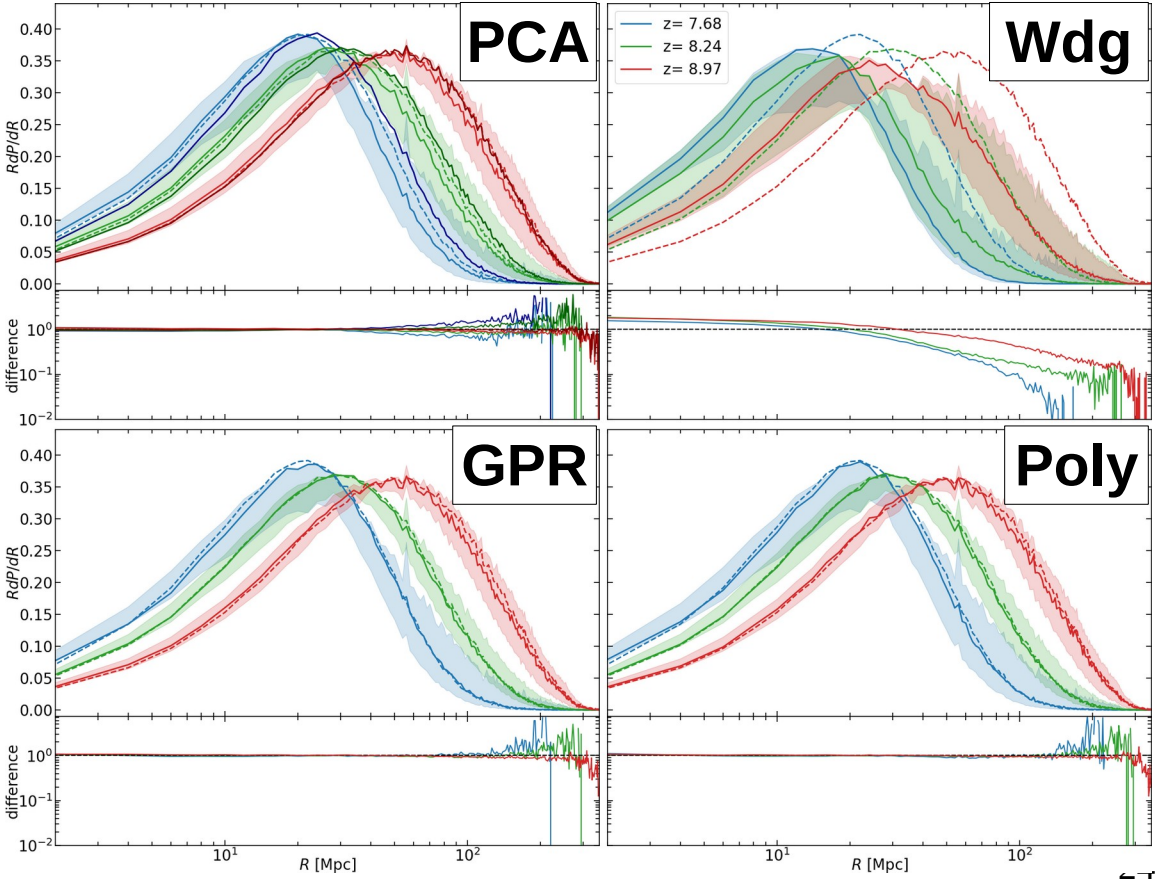
SegU-Net: HI size distribution

The Island Size Distribution (ISD): statistical distribution of HI regions during EoR (Giri+ 2018)

$$\overline{R}_C(z) = \int_{R_{\min}}^{\infty} R \, \frac{dP}{dR}(z) \, dR$$

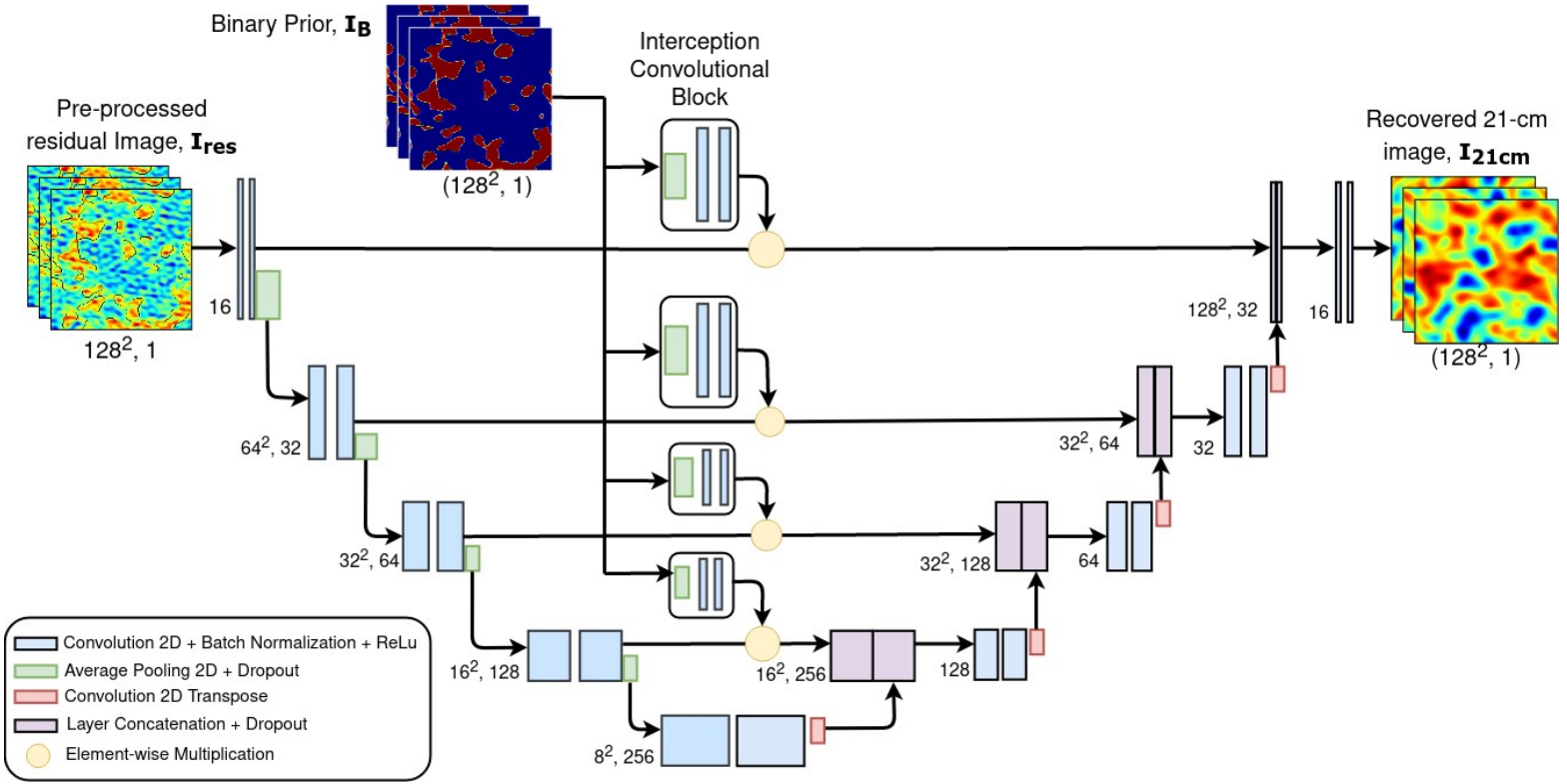
SegU-Net results:

z_c	pre-process	$\overline{R}_C$ [cMpc]
8.24	Ground Truth	**29.54**
	all z PCA	$31.37^{+3.09}_{-3.93}$
	PCA	$27.65^{+9.13}_{-6.12}$
	Wedge	$15.20^{+24.13}_{-6.18}$
	GPR	$29.14^{+5.26}_{-4.89}$
	Polynomial	$29.21^{+5.83}_{-5.21}$



RecU-Net: Recover 21-cm with U-Net

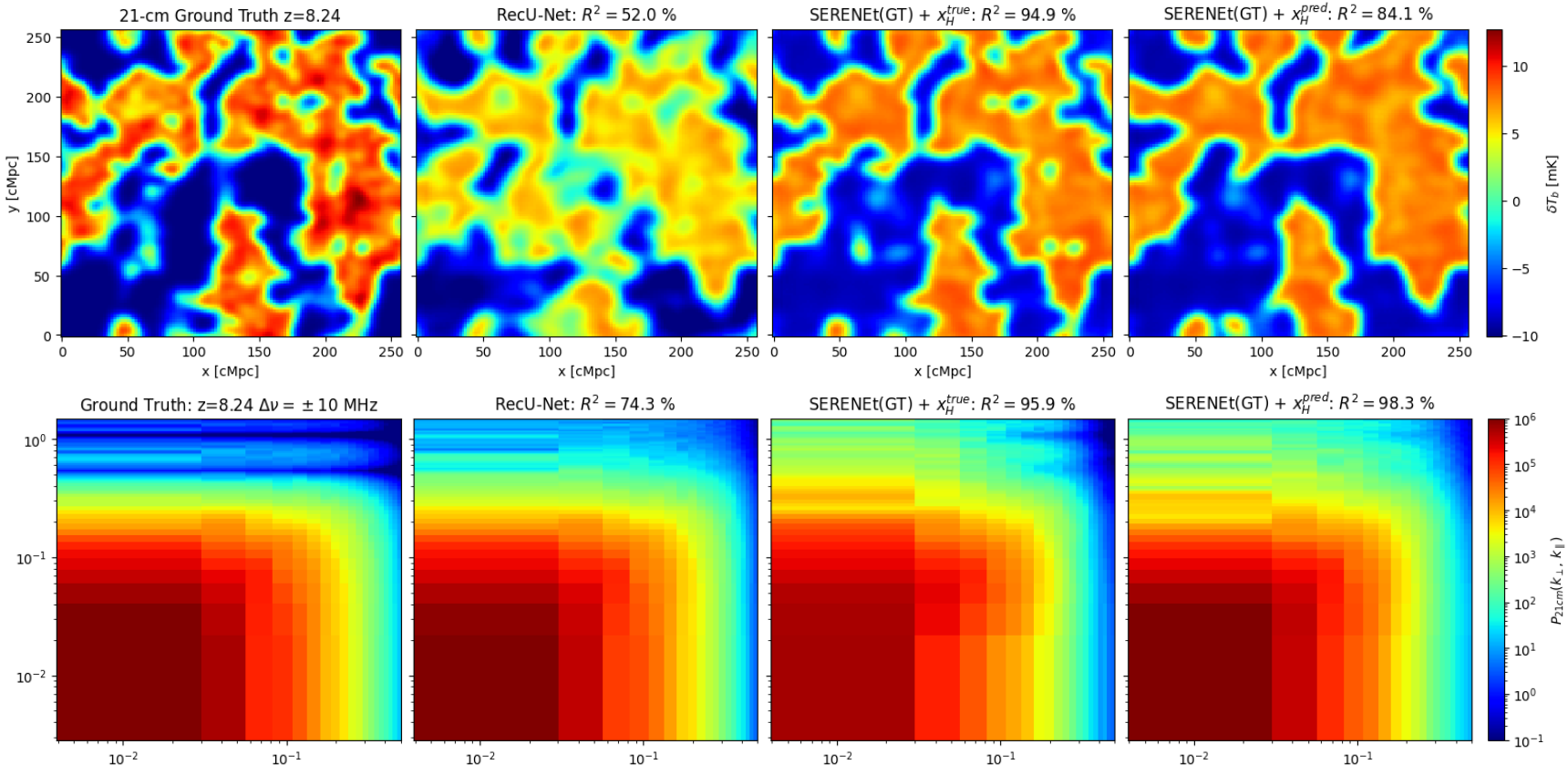
(Bianco+ in prep.)



U-Net architecture with intercepting convolution block to process the binary prior map from SegU-Net

SERENEt: Recover of 21-cm Signal

(Bianco+ in prep.)



Recovered 21-cm signal for EoR for lightcone subvolume centered at redshift **z = 8.25** (x_{HI}~0.5) and **Δν = 20 MHz** on PCA pre-process images

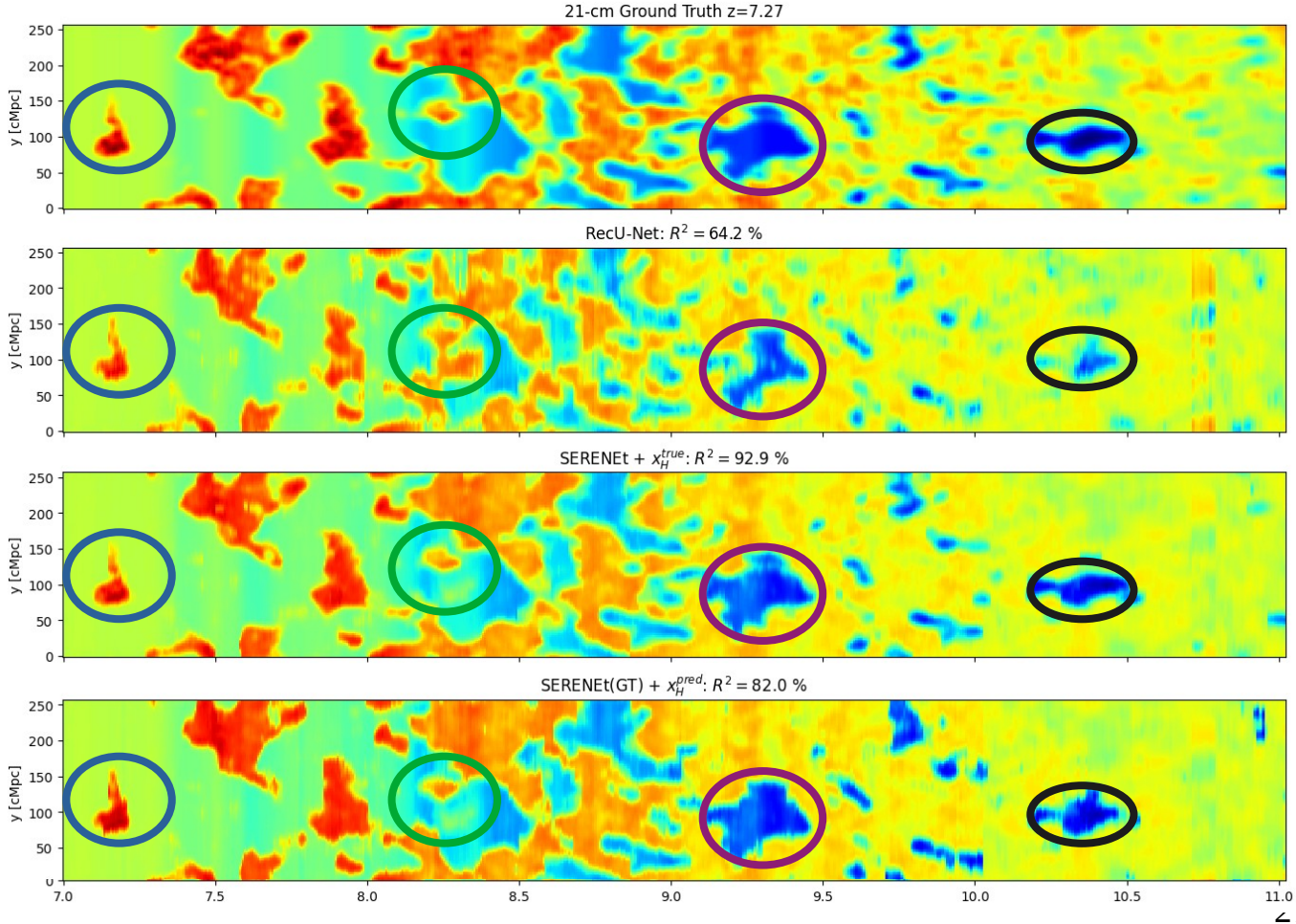
SERENEt: Recover of 21-cm Signal

Ground Truth
at z = 7.27

RecU-Net (no prior)
R^2 = 64.2 %

SERENEt**(GT)** + $x_H^{\mathbf{TRUE}}$
R^2 = 92.9 %

SERENEt**(GT)** + $x_H^{\mathbf{PRED}}$
R^2 = 82.9 %



SERENEt: Recover of 21-cm Signal

Coefficient of determination: **R^2 score VS redshift evolution** to quantify the accuracy of SERENEt foreground mitigation pipeline:

- **SERENEt(GT) + x_H^{TRUE}**
  upper limit based on best prior binary map (ground truth)

- **SERENEt(GT) + x_H^{PRED}**
  next best results when compared to **RecU-Net**