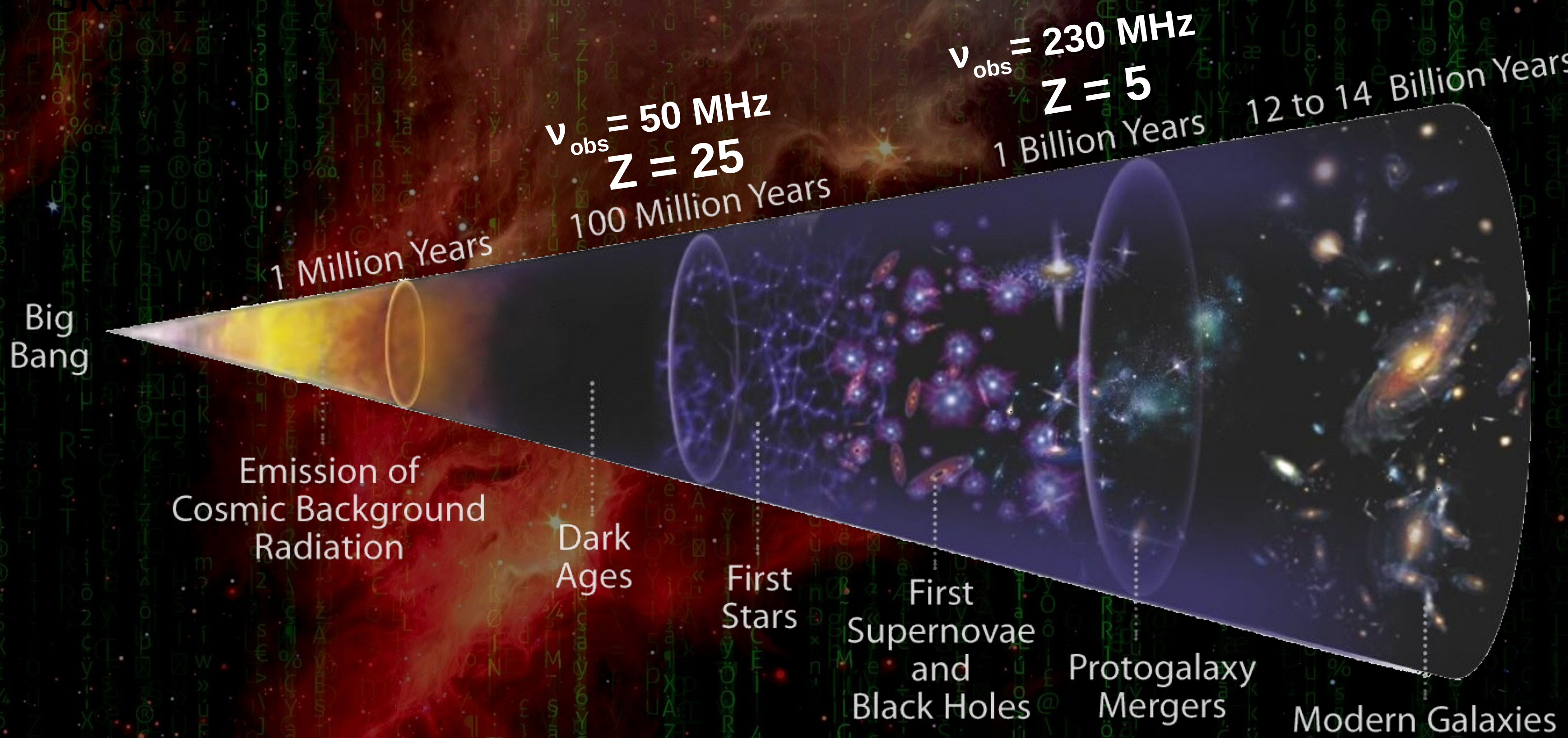
21cm Cosmology Workshop 2022

# Deep Learning approach for HI regions identification and 21-cm signal recover from SKA-Low observations


Michele Bianco

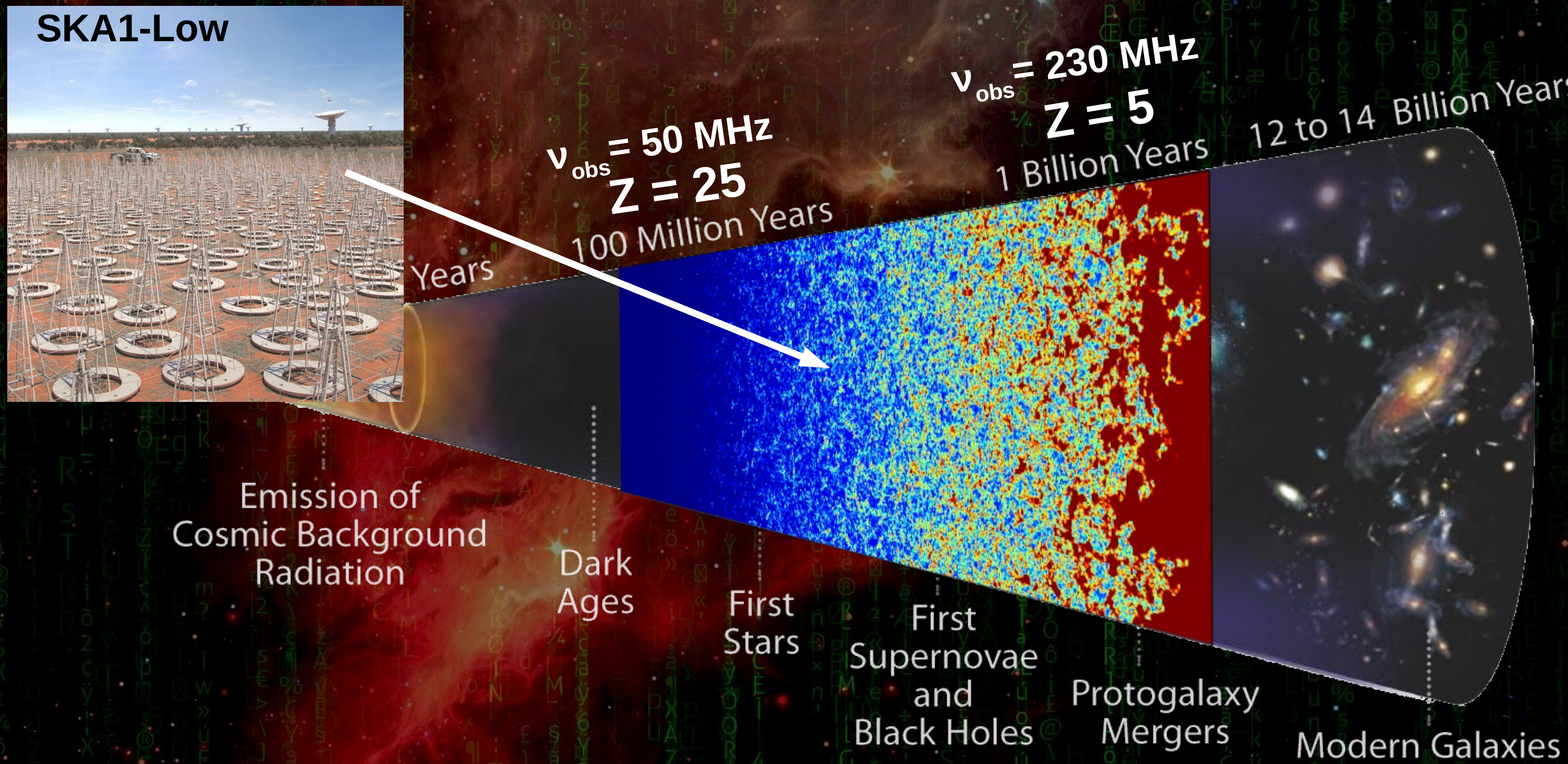
David Prelogović (SNS Pisa), Tianyue Cheng (EPFL), Sambit K. Giri (University Zurich), Emma Tolley (EPFL), Andrei Mesinger (SNS Pisa)

# The Epoch of Cosmic Reionisation




- Transition the Universe from a cold, neutral state to hot and ionised
- Detect 21-cm signal from hydrogen in the Intergalactic Medium (IGM)

# The Epoch of Cosmic Reionisation




- Transition the Universe from a cold, neutral state to hot and ionised
- Detect 21-cm signal from hydrogen in the Intergalactic Medium (IGM)

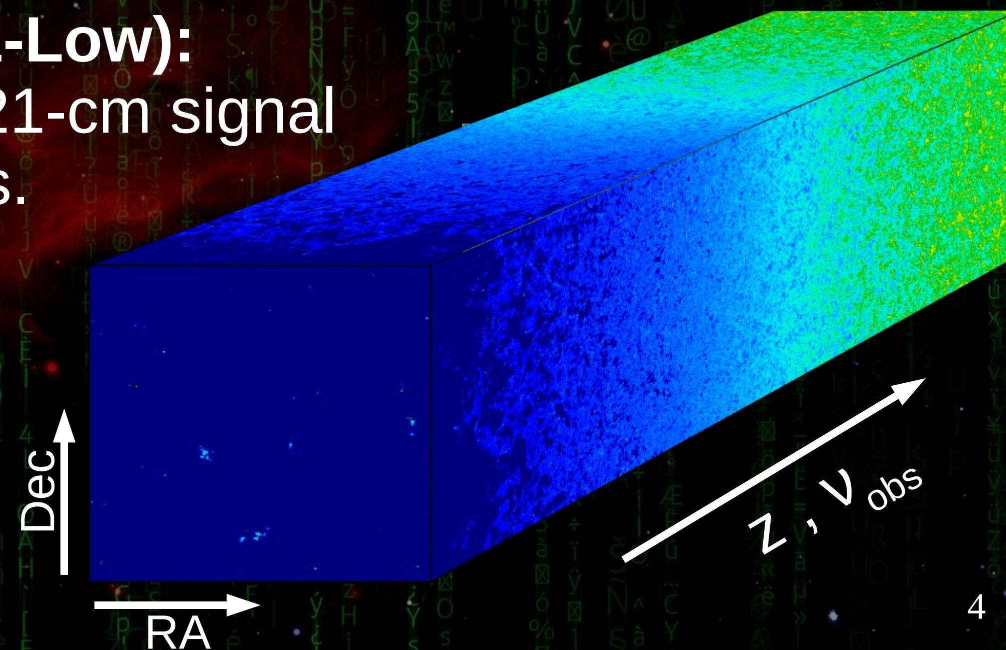
# Tomographic imaging of the 21-cm signal

Probe reionization process by observing the redshifted 21-cm signal

$$\delta T_b(z) \propto x_{HI}(z)$$

**Square Kilometre Array (SKA1-Low):**
Images sequence of redshifted 21-cm signal at different observed frequencies.

3D tomographic dataset or
a.k.a. 21-cm lightcones



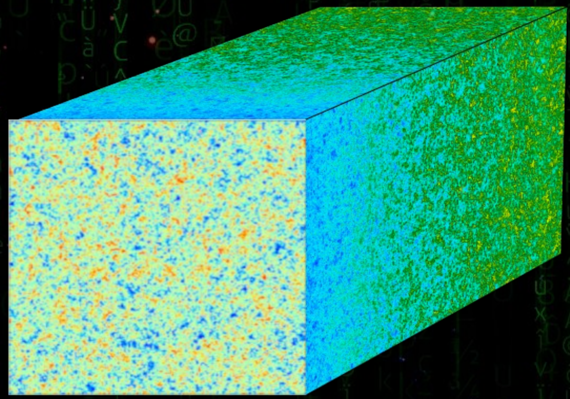
# Tomographic imaging of the 21-cm signal

Probe reionization process by observing the redshifted 21-cm signal

$$\delta T_b(z) \propto x_{HI}(z)$$

**Square Kilometre Array (SKA1-Low):**
Images sequence of redshifted 21-cm signal
at different observed frequencies.

3D tomographic dataset or
a.k.a. 21-cm lightcones

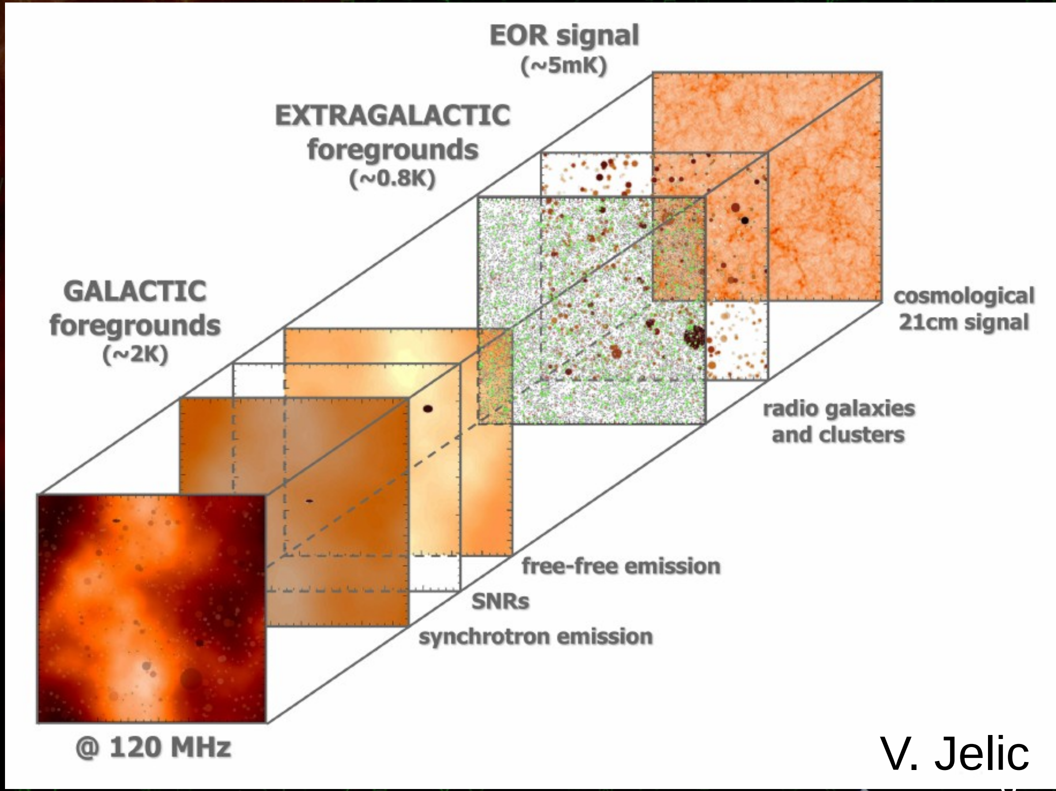


$z = 13.2$
$\nu_{obs} = 100$ MHz

# Tomographic imaging of the 21-cm signal

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

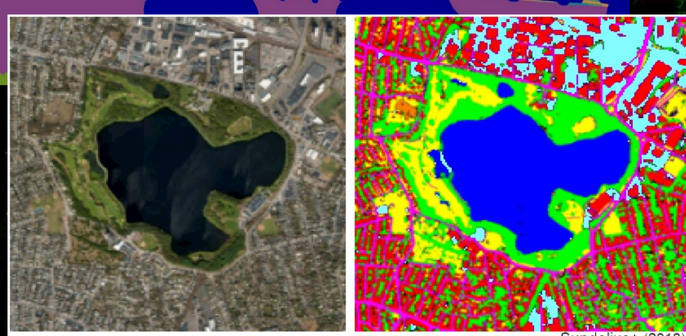
- Instrumental noise
  (signal ~ 5 K)
- Foreground emission
  (signal ~ 1 - 1000 K)
- Antennas gain errors
- Ionospheric refraction effects
- Radio frequency interference
- And more ...



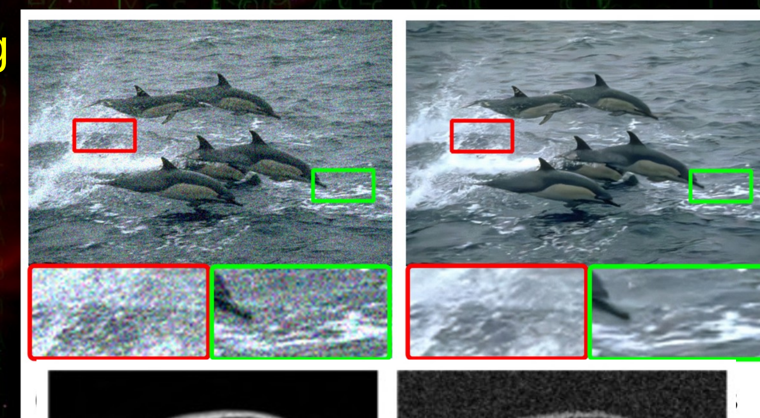
# Deep Learning algorithm with Convolutional Neural Networks

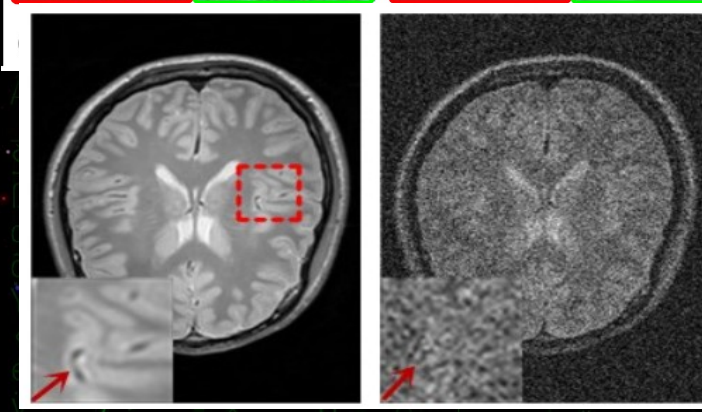
Modern Computer Vision technology based on AI and deep learning methods are able to **identify object** and/or **de-noise** images with great precision. (e.g.: self-driving cars, image satellites, medical image, etc...)

Image segmentation




de-noising images

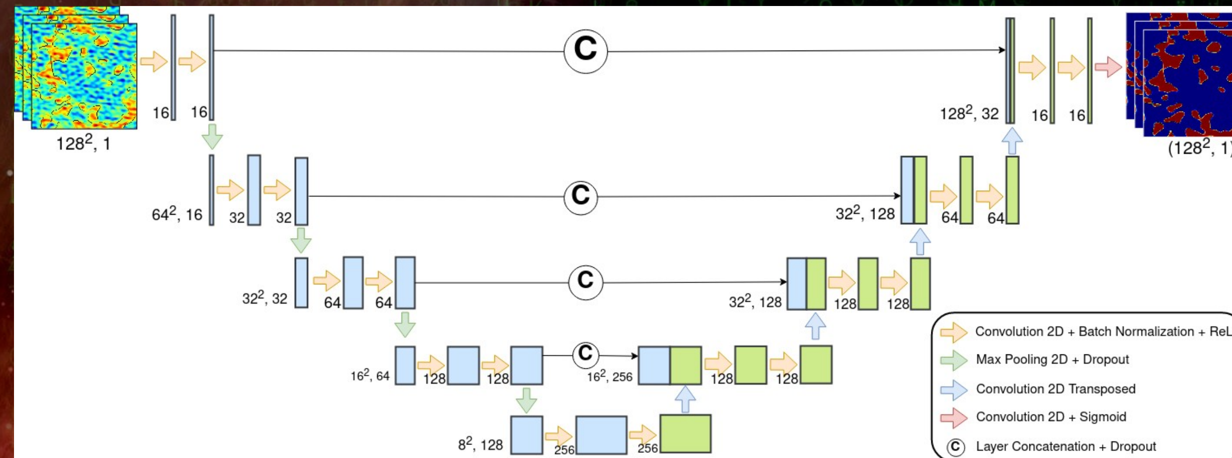


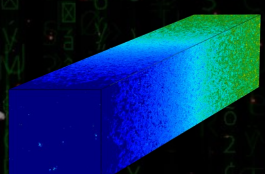


# SegU-Net: Segmentation with U-Net for EoR

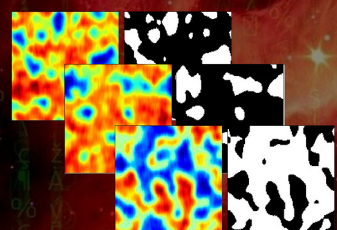
## (Bianco+ 2021) arXiv:2102.06713

- U-Net:
  Network with interconnected encoder/decoder layers
- Convolutional layers on 2D slice of tomographic dataset (rolling procedure along z-axis)



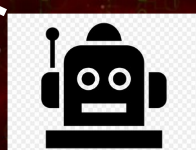
21cm tomography dataset



$(z, \nu_{obs})$

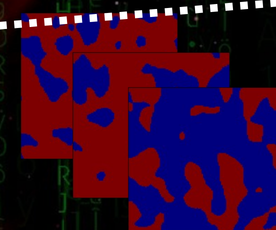
21cm signal = image

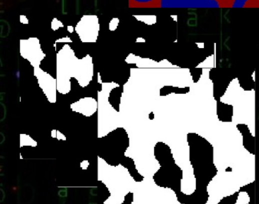
$x_{HI}$ field = mask



SegU-Net

(2.5M trainable params.)





Compare with ground truth

$$\mathcal{L}(y, \hat{y})$$

Calculate Loss

# Mock Data for 21-cm Observations

EoR semi-numerical simulations:

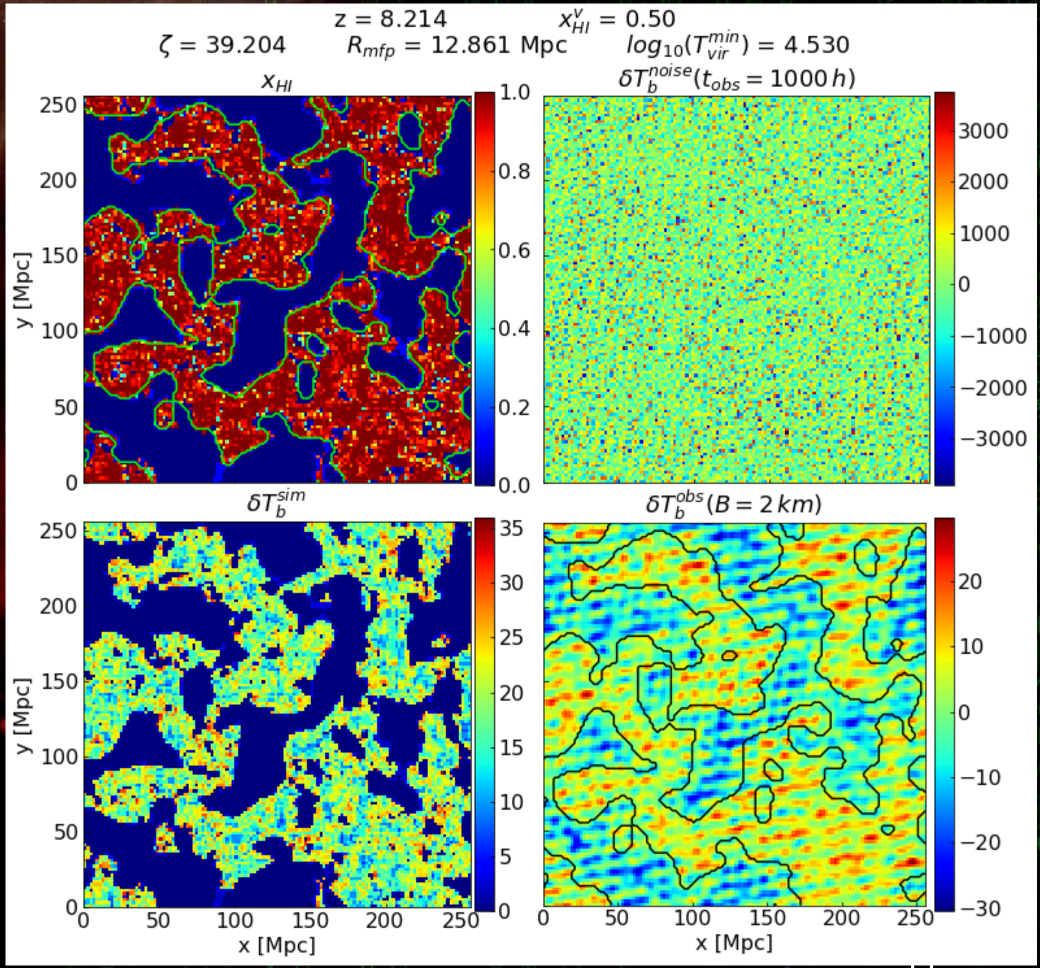
- 10k **21cmFAST** lightcone simulation
  - Astrophysical parameters
  - Redshift range: 7 – 9
  - Heating approximation: $\delta T_b \sim n_{HI}(z)$

Noise:

- SKA1-Low instrumental noise model (Giri+ 2018b)
- $t_{obs}$ = 1000h of integration time

Interferometric Smoothing scale:

- Gaussian kernel, B=2km



# Mock Data for 21cm Observations

EoR semi-numerical simulations:

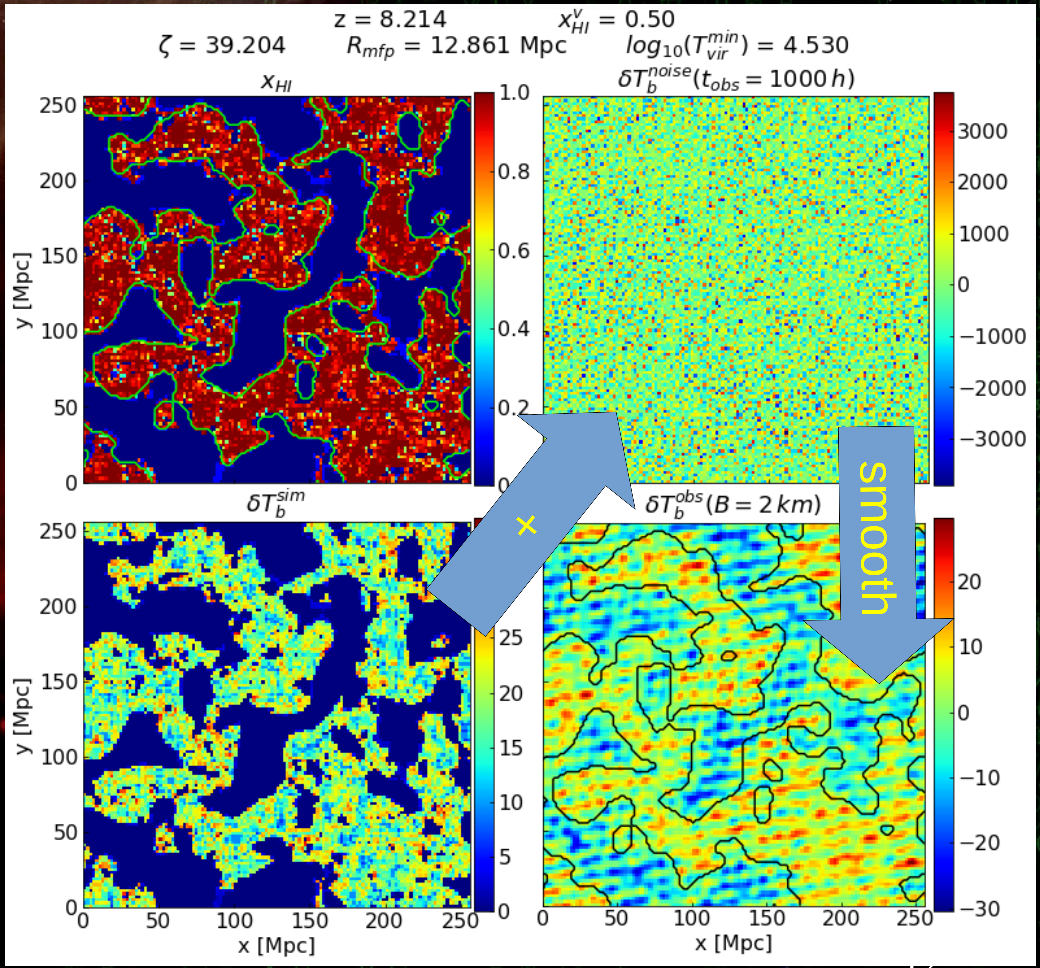
- 10k **21cmFAST** lightcone simulation
  - Astrophysical parameters
  - Redshift range: 7 – 9
  - Heating approximation: $\delta T_b \sim n_{HI}(z)$

Noise:

- SKA1-Low instrumental noise model (Giri+ 2018b)
- $t_{obs}$ = 1000h of integration time

Interferometric Smoothing scale:

- Gaussian kernel, B=2km



# Mock Data for 21cm Observations

EoR semi-numerical simulations:

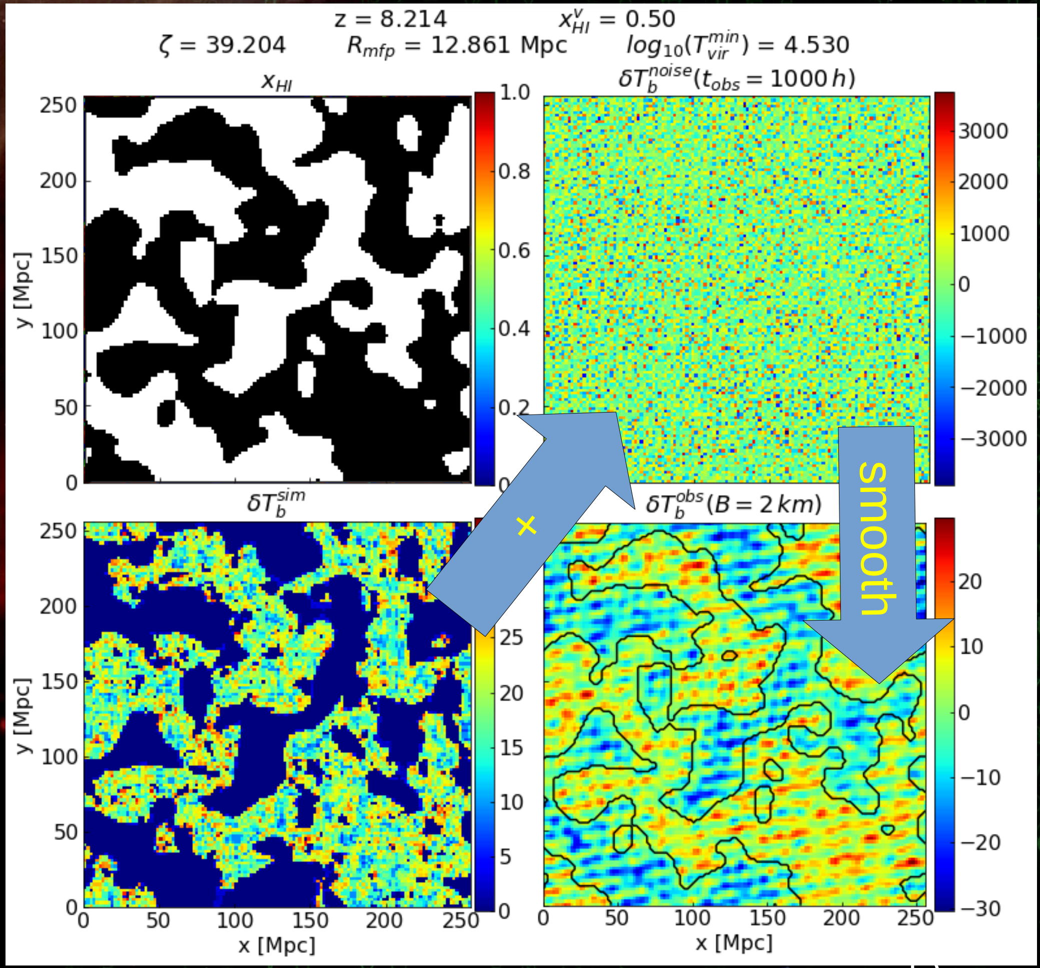
- 10k **21cmFAST** lightcone simulation
  - ➔ Astrophysical parameters
  - ➔ Redshift range: 7 – 9
  - ➔ Heating approximation: $\delta T_b \sim n_{HI}(z)$

Noise:

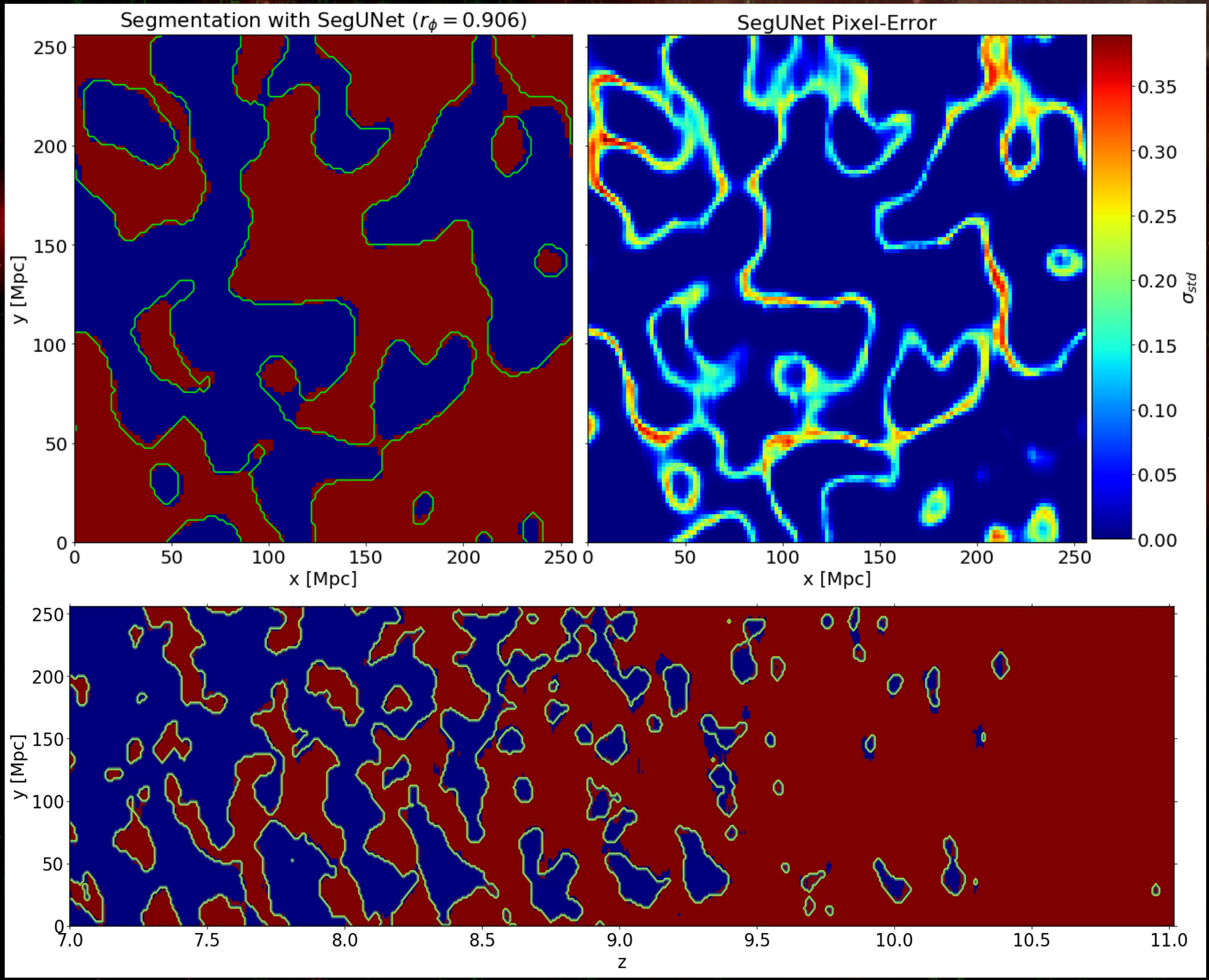
- SKA1-Low instrumental noise model (Giri+ 2018b)
- $t_{obs}$ = 1000h of integration time

Interferometric Smoothing scale:

- Gaussian kernel, B=2km

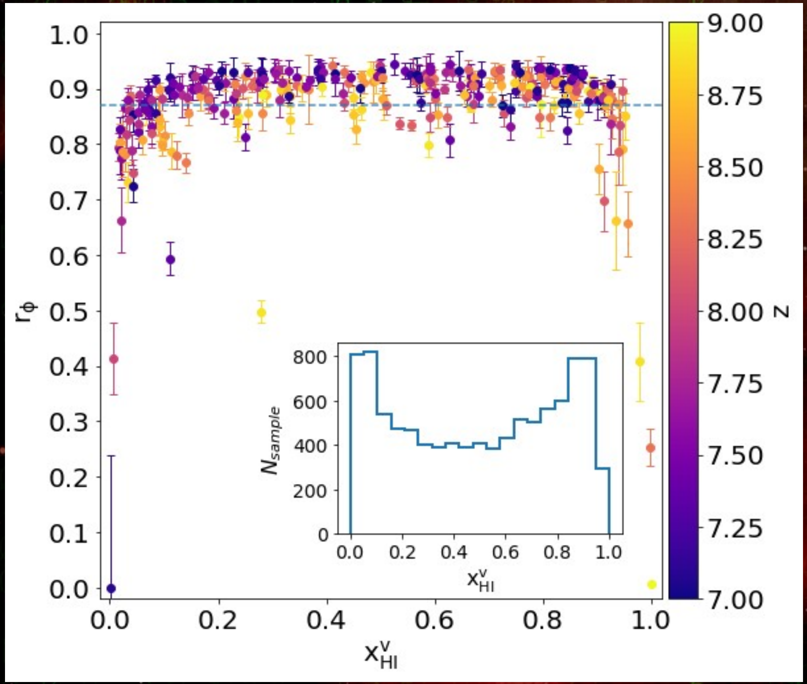


# SegU-Net Results: Visual Evaluation & Uncertainty-map

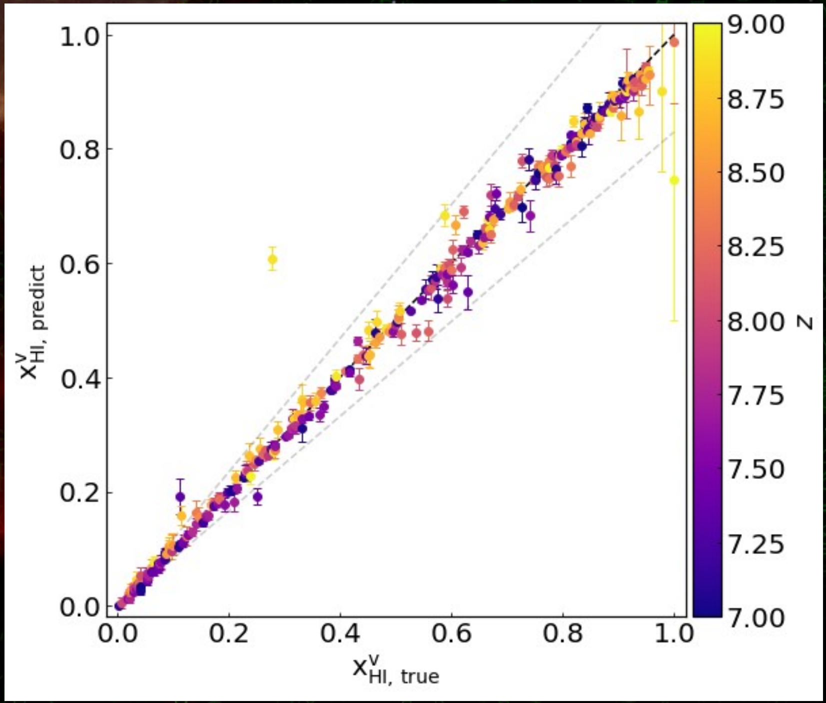



- Network binary field recovers with “confidence” large interconnected ionised/neutral regions
- Higher uncertainty at Bottleneck and regions with low-dynamic range

# SegU-Net Results : Correlation Coefficient $r_\Phi$ and Reionisation History $x_{HI}$




Average accuracy: 85% better than state-of-the-art algorithm for segmentation

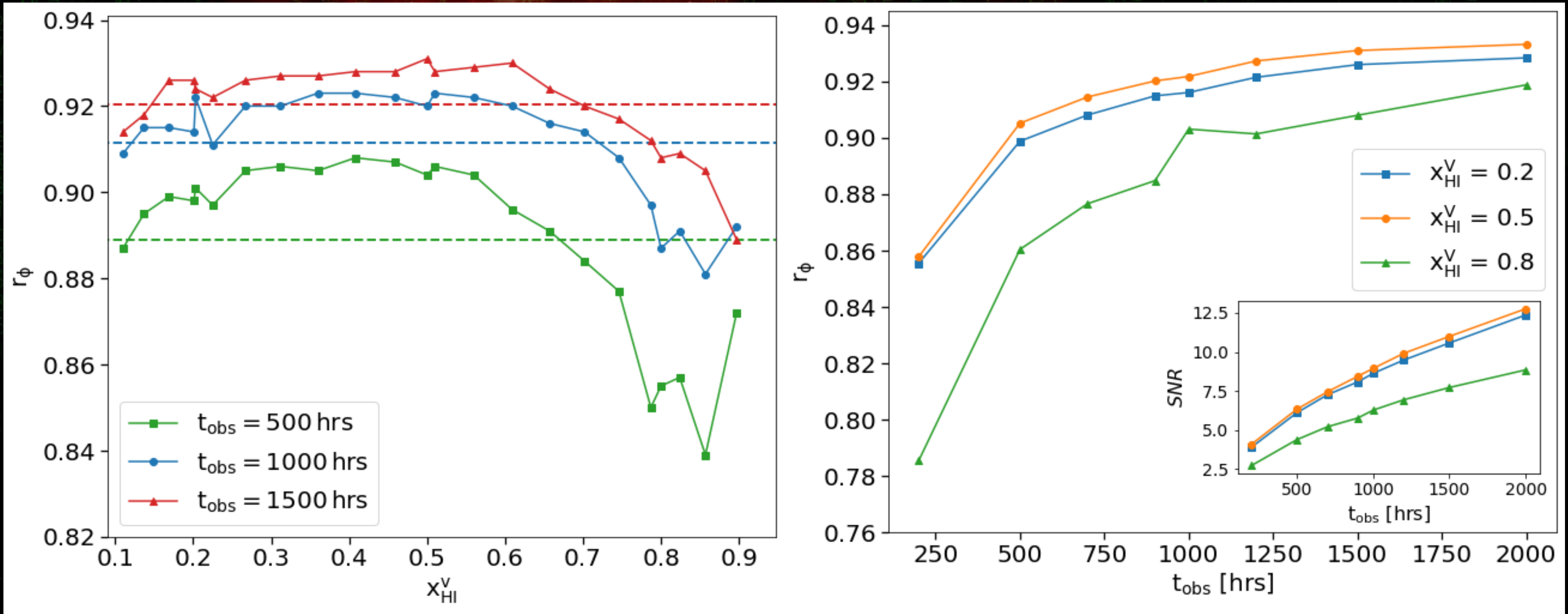



Recovered EoR history from the network binary field within ~0.5 σ difference

# SegU-Net Results: Response to Noise level

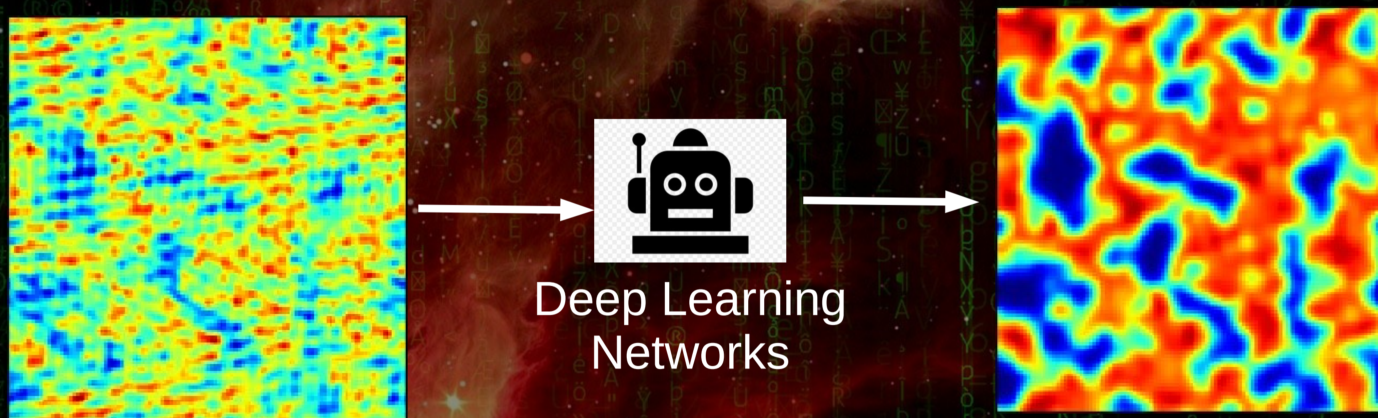
Test on different instrumental noise level: under- or over-estimate

- Predictions on un-trained data with $\mathbf{t_{obs}}$ **= 500 - 2000 hrs**
- $\mathbf{t_{obs}}$ > 500 hrs (SNR>3) same level of accuracy (~85%) as in the training
- Network accuracy affected by the **dynamic range** in the images



# The Next Goal of the Project

Deep learning approach for HI regions identification….
…. and 21-cm signal recover from SKA-Low observations




1) Use accurate modelling of the foreground and interferometry instrumental response (incomplete uv-coverage)

2) Data pre-process for foreground avoidance (wedge removal) and/or foreground mitigation techniques (PCA)

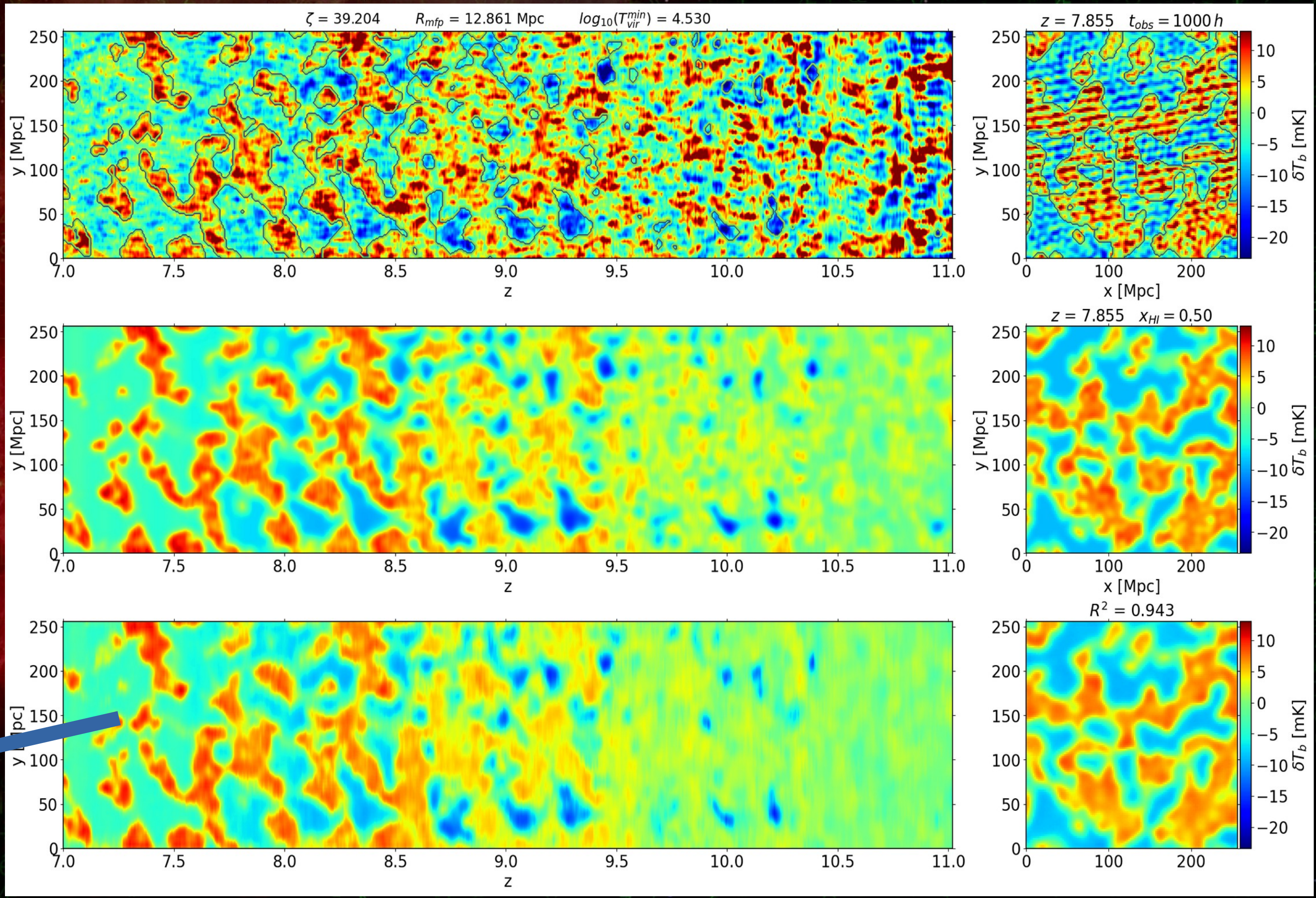
3) Use identified HI regions as prior information in the training process of the 21-cm recover network (RecU-Net)

# RecU-Net Results:
# 21-cm Visual comparison

A first test:

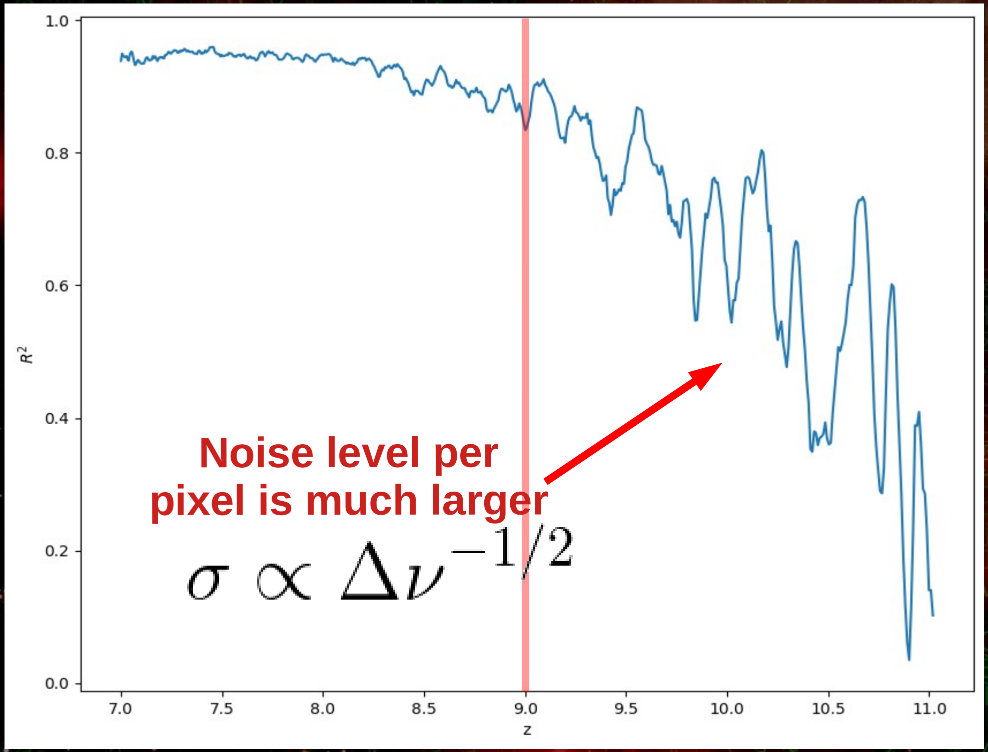
- RecU-Net≈SegU-Net
  1) final activation
  2) changed target
  3) different loss

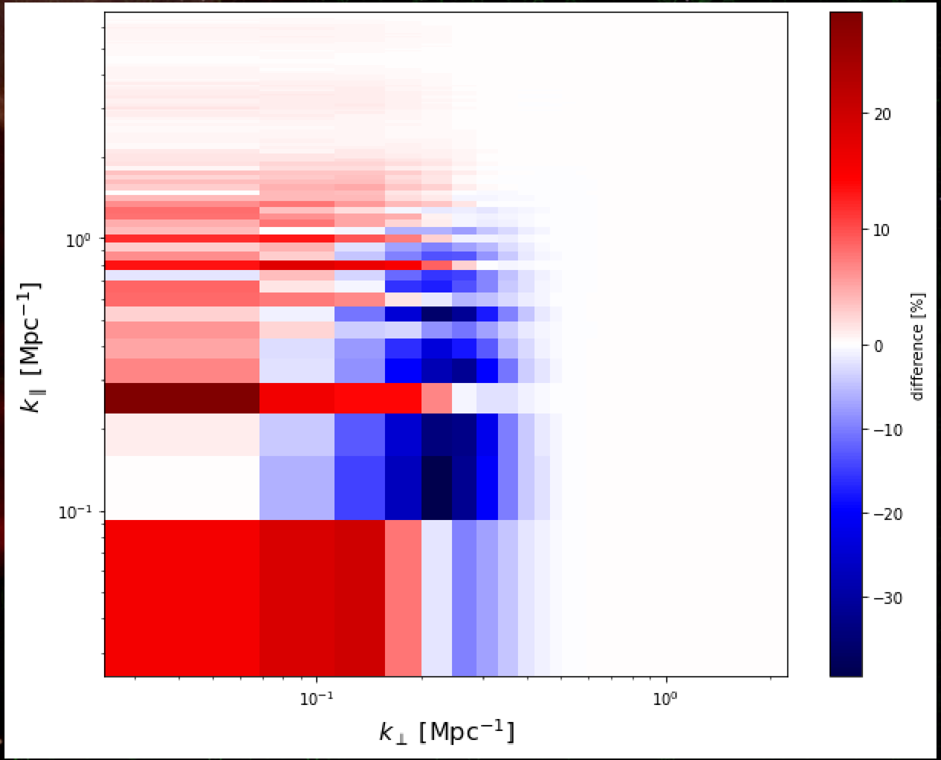
- Recover 21cm signal from images with SKA-Low instrumental noise




This is RecU-Net prediction

# RecU-Net Results: $r_\Phi$ and $x_{HI}$ on entire Tomographic data




RecU-Net is extremely accurate ($R^2 \sim 92\%$) redshift range **7 < z < 9**

21-cm Power spectra (top) recover within ~5% small scale correlation, $k > 0.2$ Mpc$^{-1}$ and redshift $z < 9$

# Foreground Wedge Removal for SKA-Low

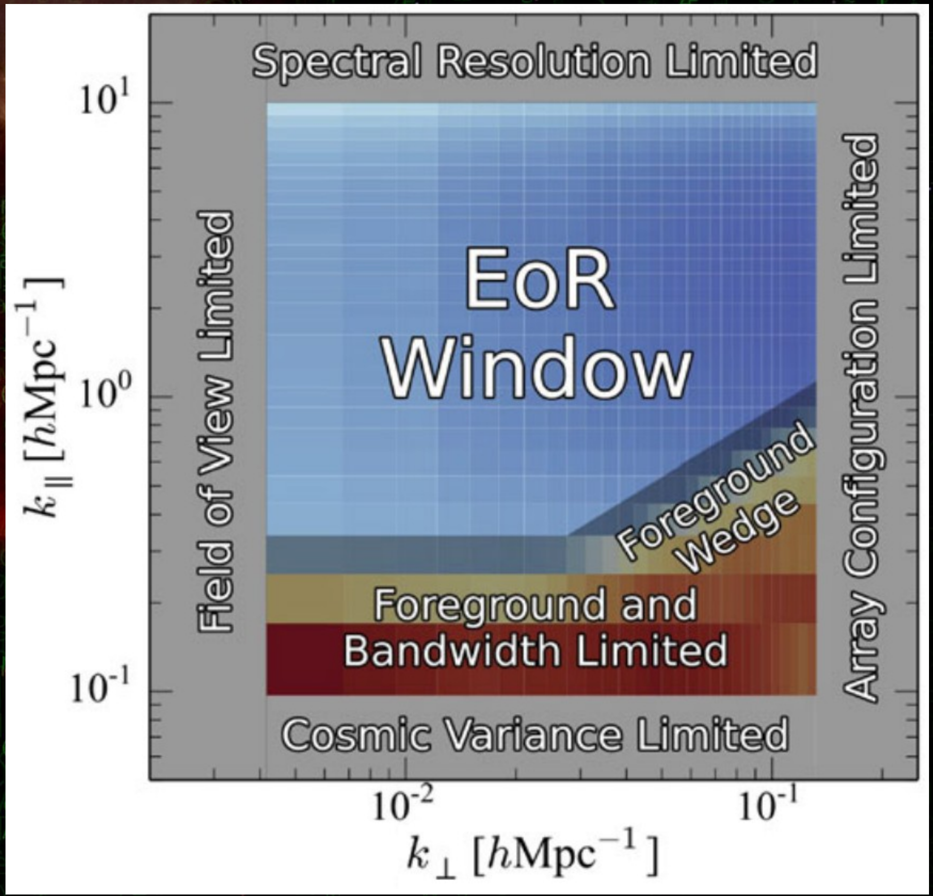
Advantage of SKA-Low long baselines: power spectra in the cylindrical coordinates ($k_\perp$, $k_\parallel$)

Removing k-mode contaminated by foreground (Foreground Wedge) as a **avoidance technique**

$$k_\parallel \leq |k_\perp| \frac{H(z)}{1+z} \int_0^{z'} \frac{dz'}{H(z)} \cdot sin\,\theta + b$$

Horizon limit angle: **θ**

(Pregolević+ 2021)



# (Preliminary) 21-cm images Pre-process

Master student project @ EPFL:

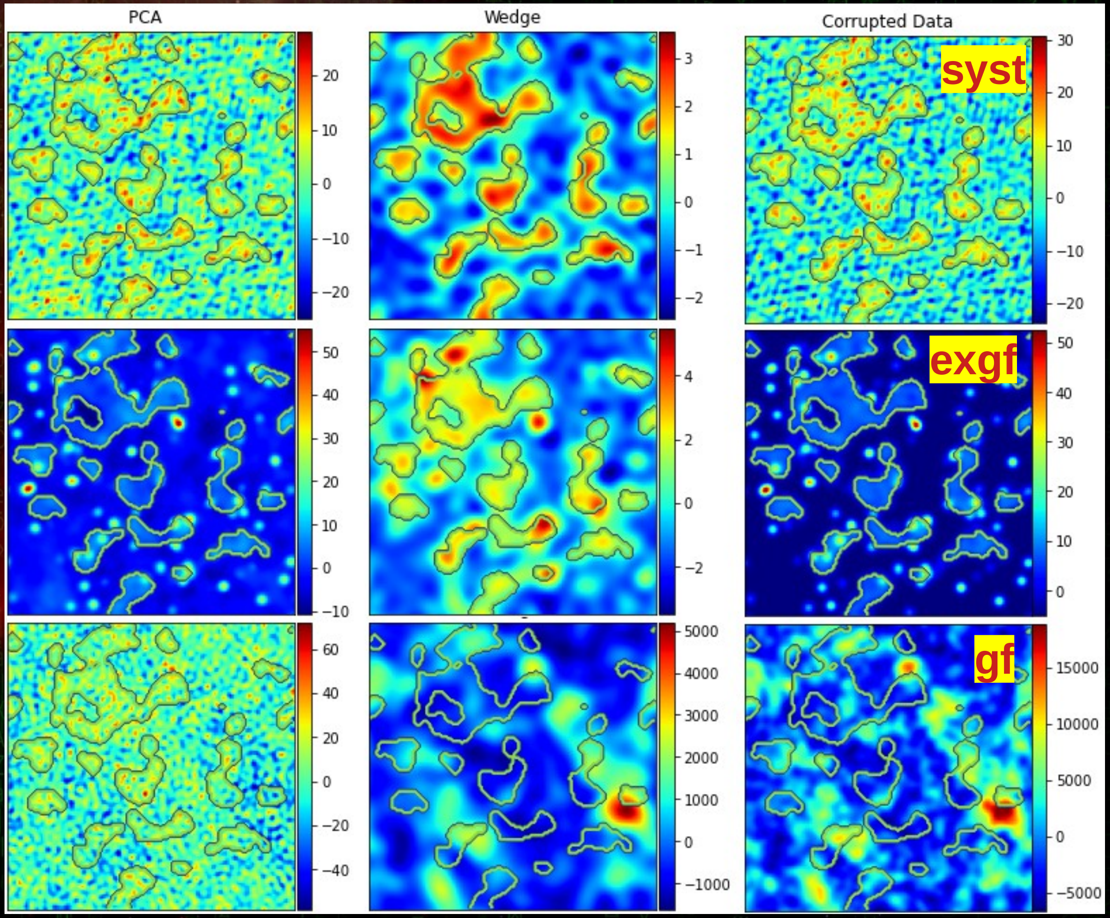
Investigate the efficiency of pre-processing

- Foreground Wedge Removal
- Principle Component Analysis (PCA) decomposition

Wedge Removal works better on extragalactic point sources

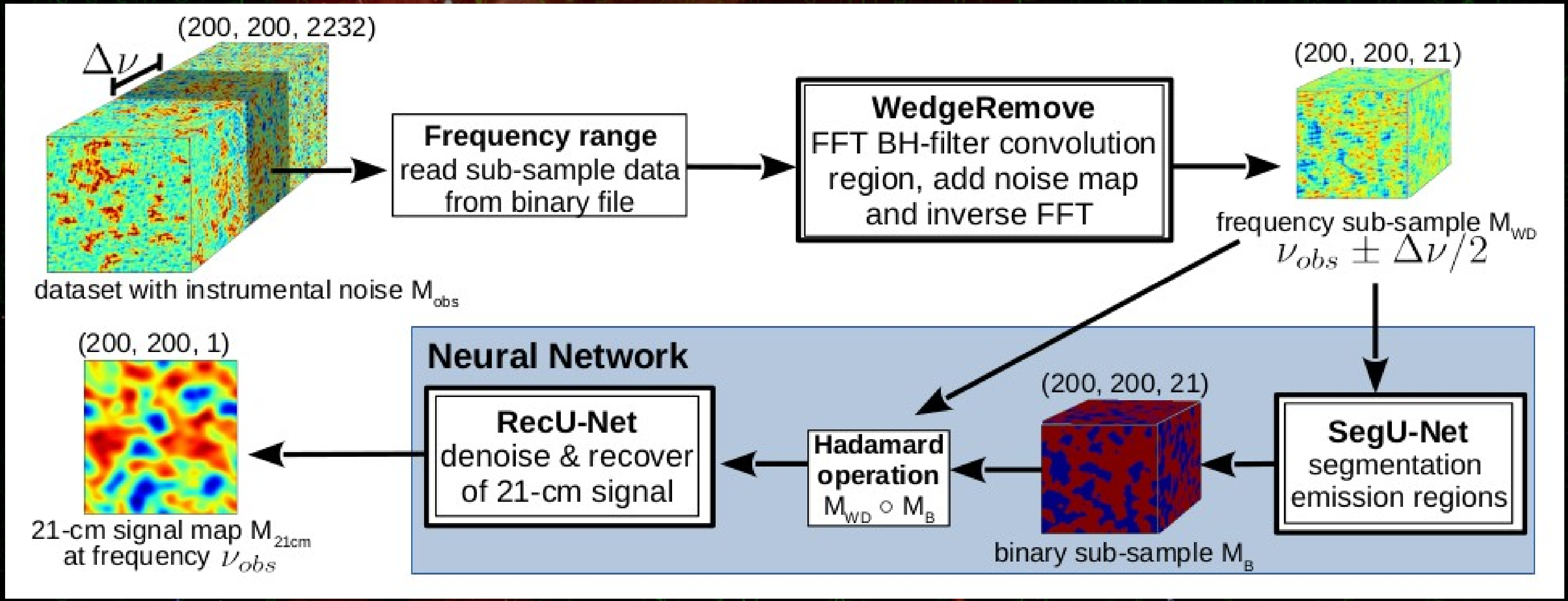
PCA works well on synchrotron galactic foreground

Both fails at higher redshift (z>9)



# SERENEt
# Segmentation and Regression NEtwork

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




Proposal accepted for SKACH large HPC allocation projects at Pitz Daint @ CSCS (start 1th of August)

# Summary & Future Work

Work done so far:

- U-Net are powerful tools for segmentation on noisy 21-cm images.
- U-Net can also be use to recover 21-cm signal from noisy data.
- Existing technique can be employed to pre-process data before training. to reduce the dynamic range of foreground contamination.

Open challenges list:

- Implement foreground contamination in the training dataset .
- Employ OSKAR simulation for radio interferometery effects.
- Feed prior information in training (combination of Seg and RecU-Net).
- Impact of the foreground pre-process step on the training and prediction.
- Combine foreground subtraction algorithm (PCA or Wedge + U-Net).