21cm foreground cleaning with machine learning

Tianyue Chen

Postdoctoral Researcher

Ecole Polytechnique Fédérale de Lausanne (EPFL)

HITS workshop, May 2022

Introduction

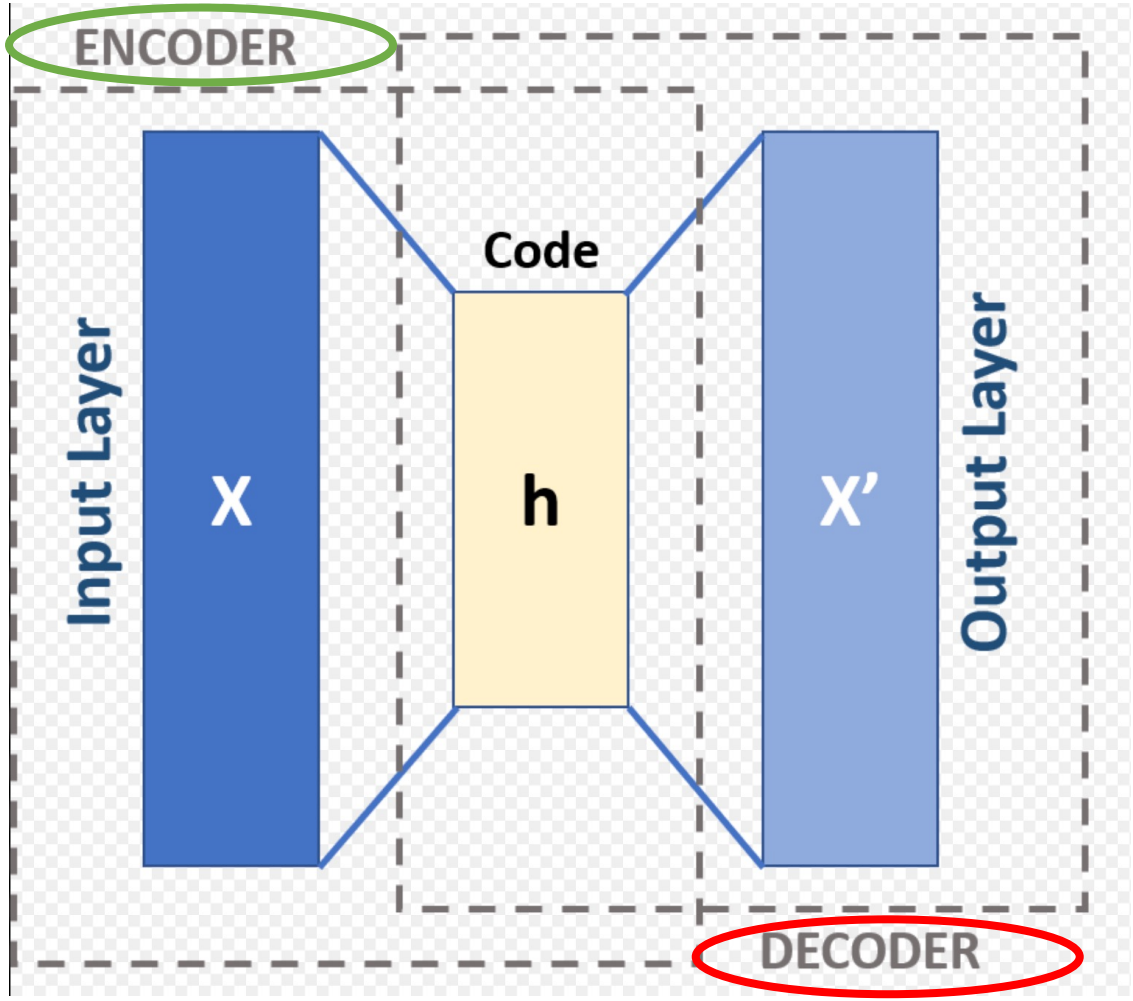
- Foreground critical for 21cm detection
- Large SKA dataset incoming

- Traditional approach:
  - Sensitive to systematics
  - Signal loss

- Can we design a machine learning algorithm?
  - Effectively remove FG
  - Robust againt systematics
  - Handle large dataset

Auto-encoder

Code input data into representations



- One type of artificial neutral network
- Learn to ignore irrelevant features

- Simple, easy architecture
- Quick to compute

Learning representations to reconstruct input

Sky models

- Flat sky 2D Gaussian maps

  HI, SYN, FREE, PS (Poisson)

- Scaled by a power spectrum
- Match amplitude to real sky signals

$$T = N(0,1) \times sqrt(a \times R^{-b})$$

T_{ps} = rand_loc + rand_flux (T_min, T_max)

- HI – independent along freq
- FG – scaled along freq

$$T_{\nu,p} = T_{0,p}(\frac{\nu}{\nu_0})^{\alpha,p}$$

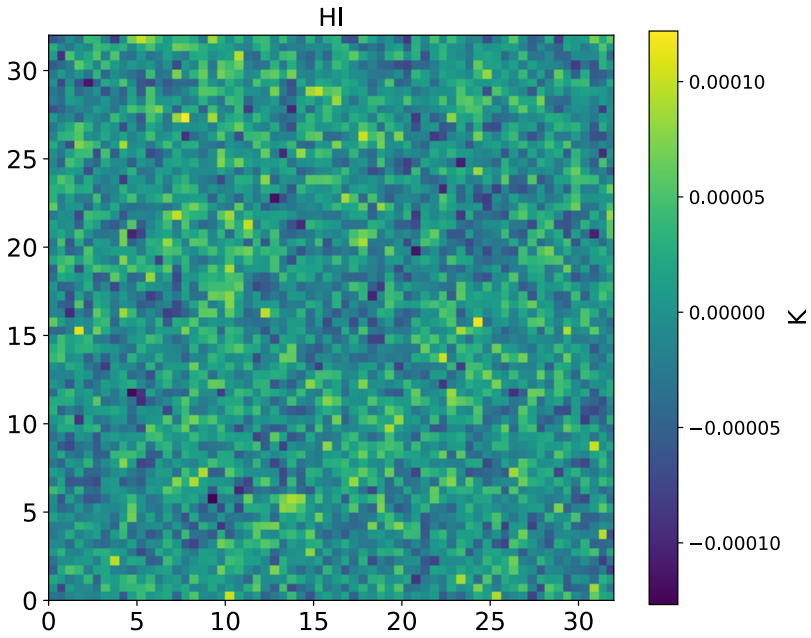
- Statistical information of sky

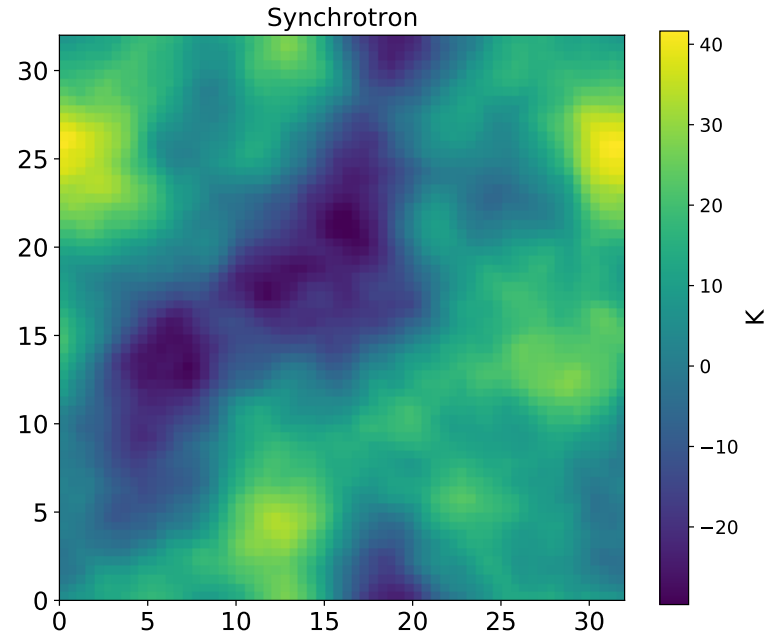
Sky maps

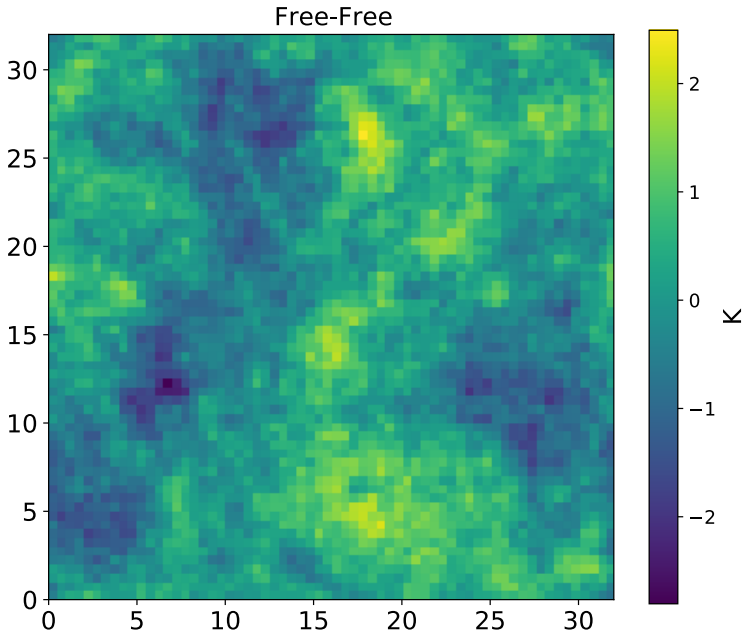
64 (npix) x 64 (npix) x 16 (nfreq)

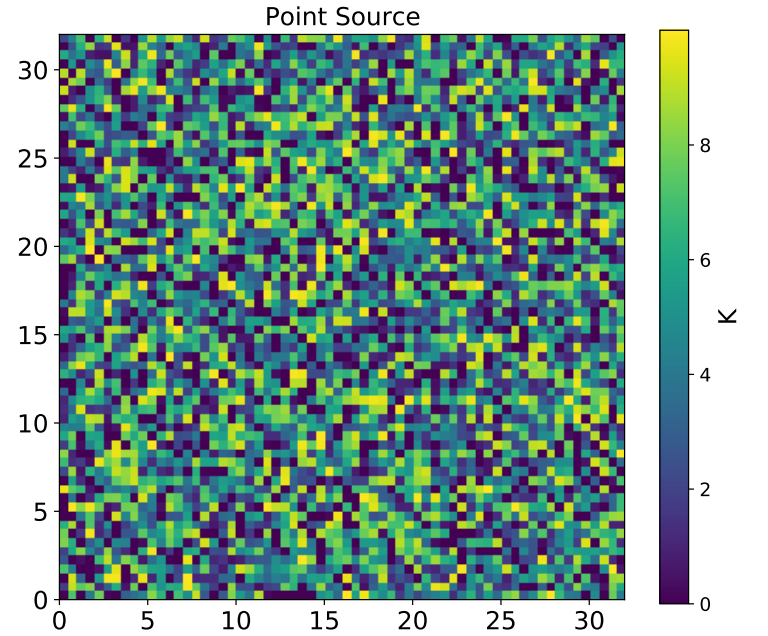
32 deg x 32 deg

400-465 MHz



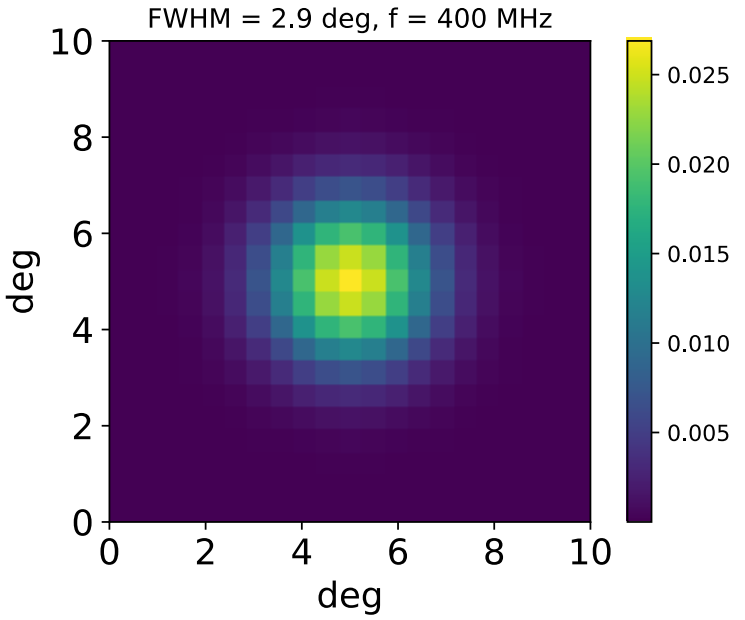


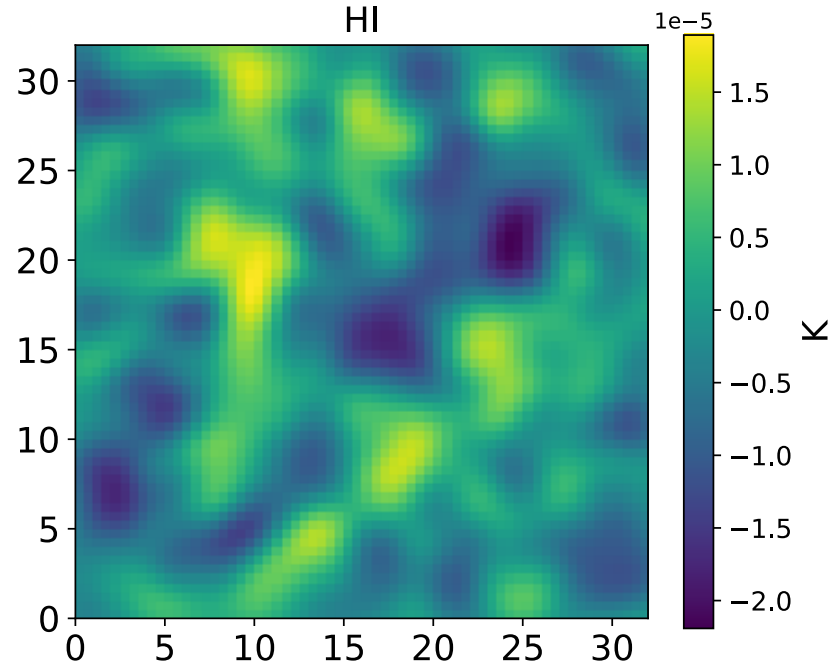


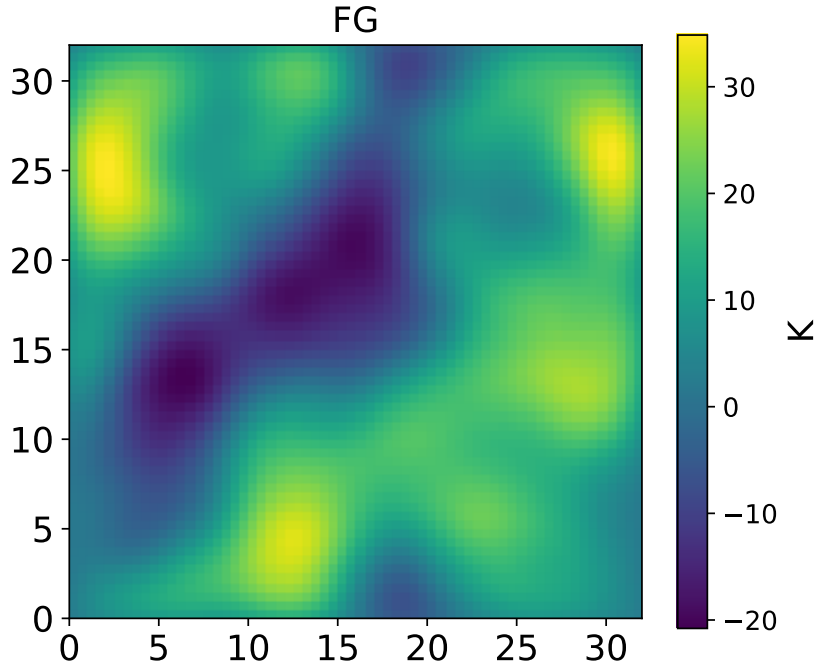


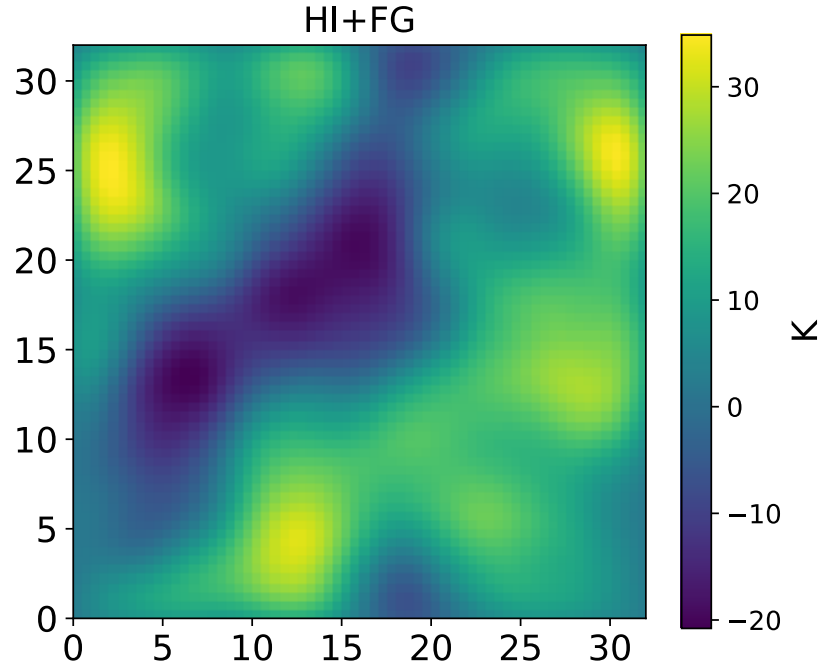
Beam

SKA-mid Band1 – like
D = 15m
FWHM ~ 2.6 deg
Gaussian





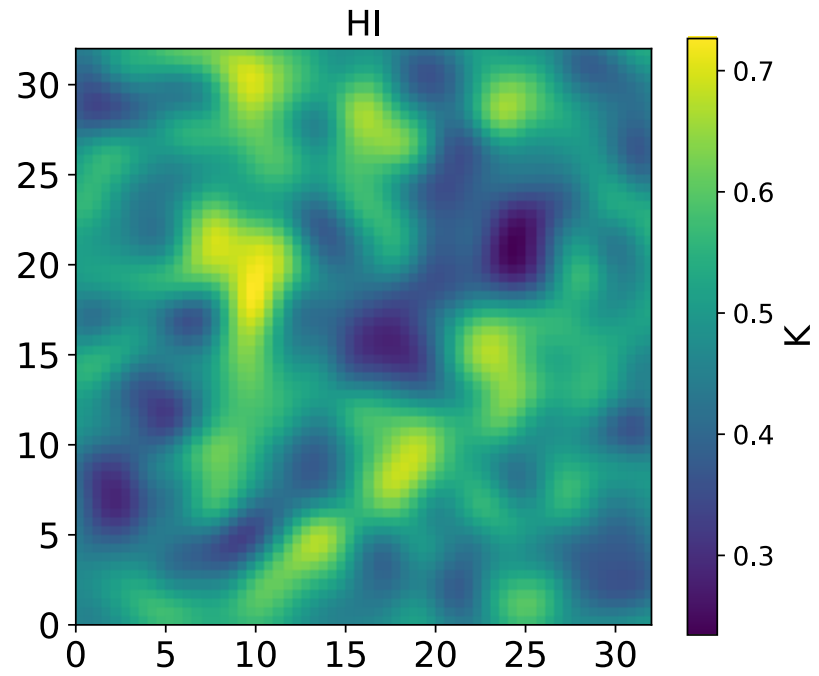


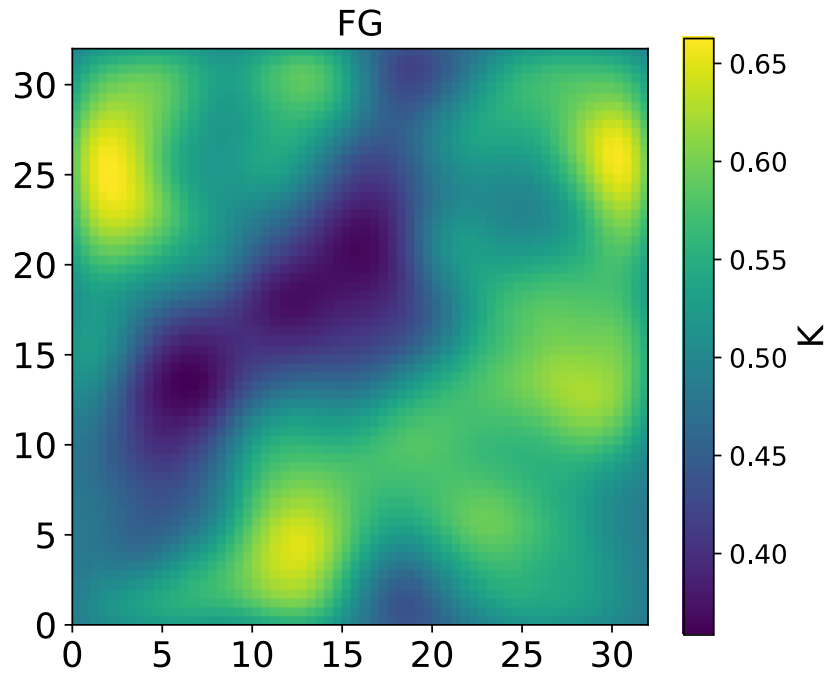


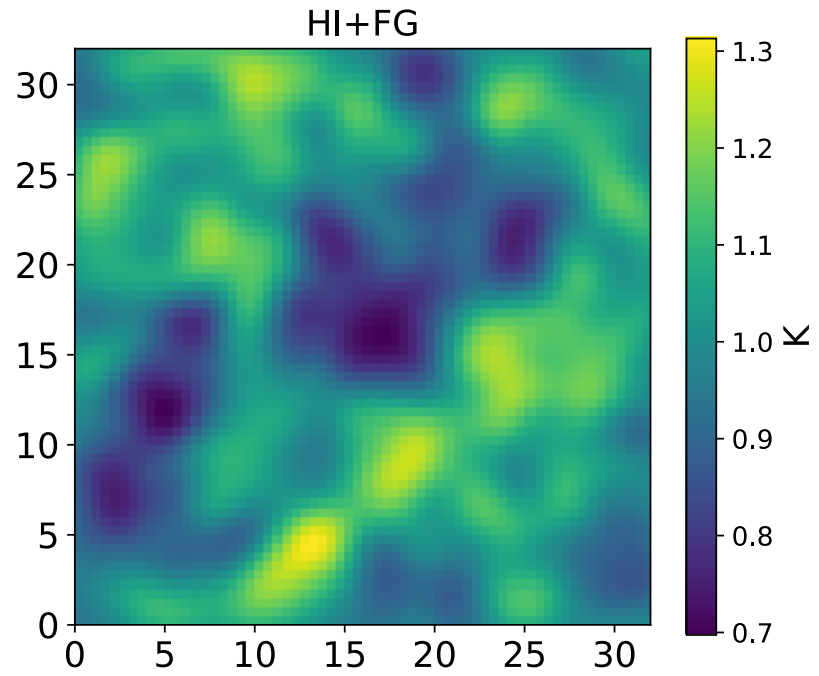
Dataset

- Training: 10000 realisations
- Validation: 5000 realisations
- Test on: 10 realisations

- Network can’t handle large dynamic range
- Artificially scale to [0,1]
- Similar to firstly apply a pre-cleaning



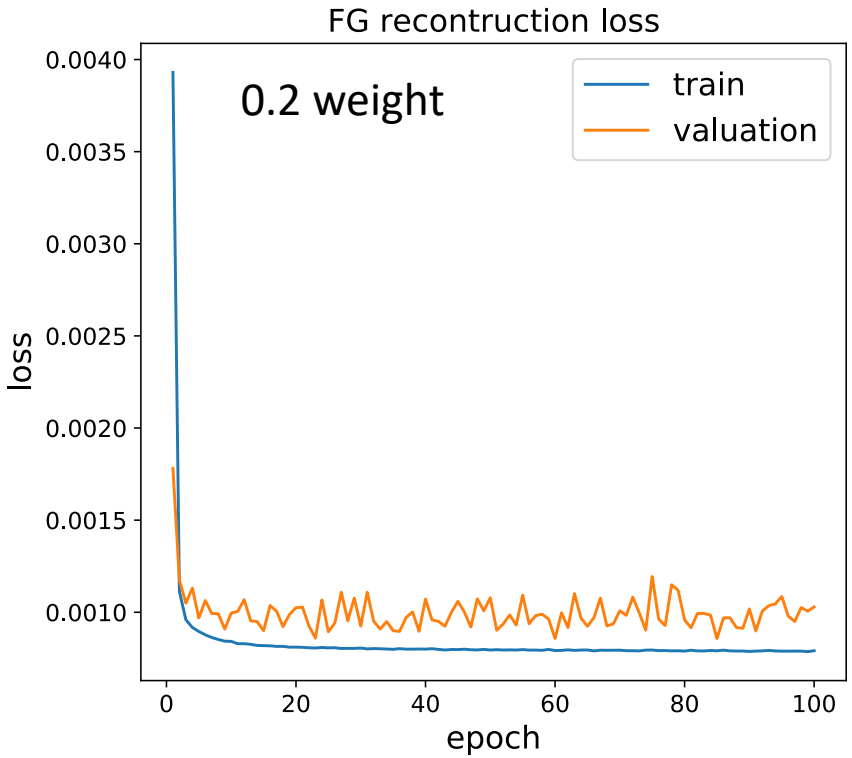


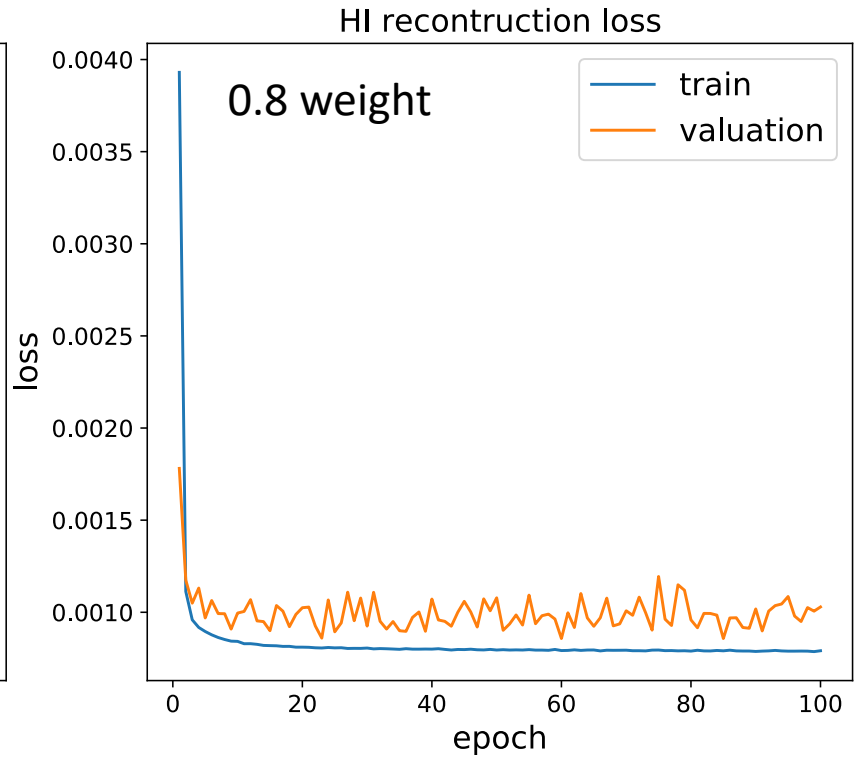


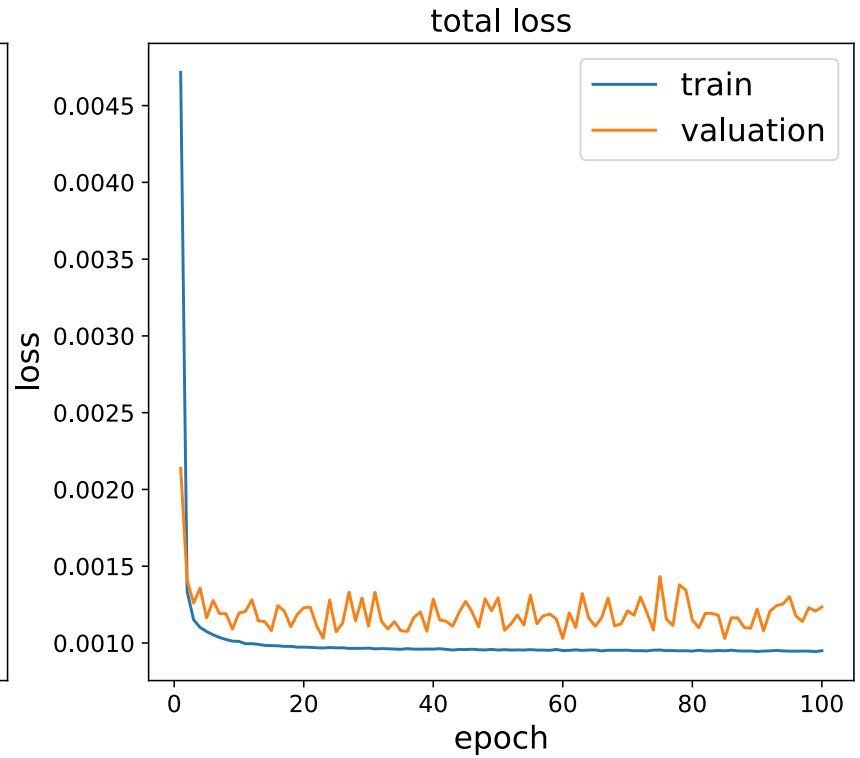
Loss function

Minimise reconstruction errors (loss)

$$\mathcal{L}(\mathbf{x}, \mathbf{x}') = \|\mathbf{x} - \mathbf{x}'\|^2$$





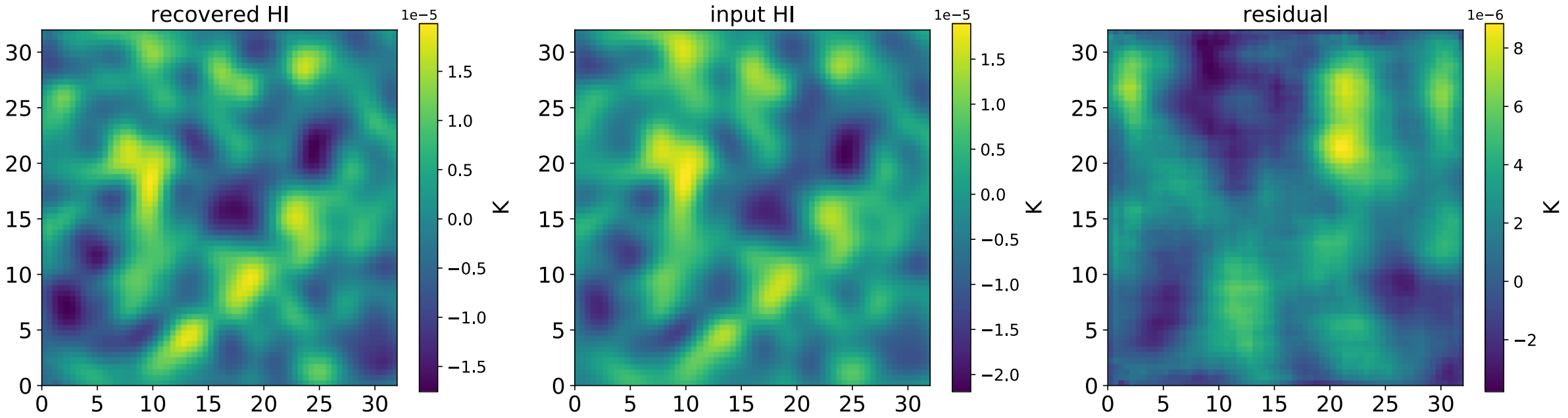


Train loss: how the model fits the training data

Validation loss: how the model performs

More layers might be needed though!

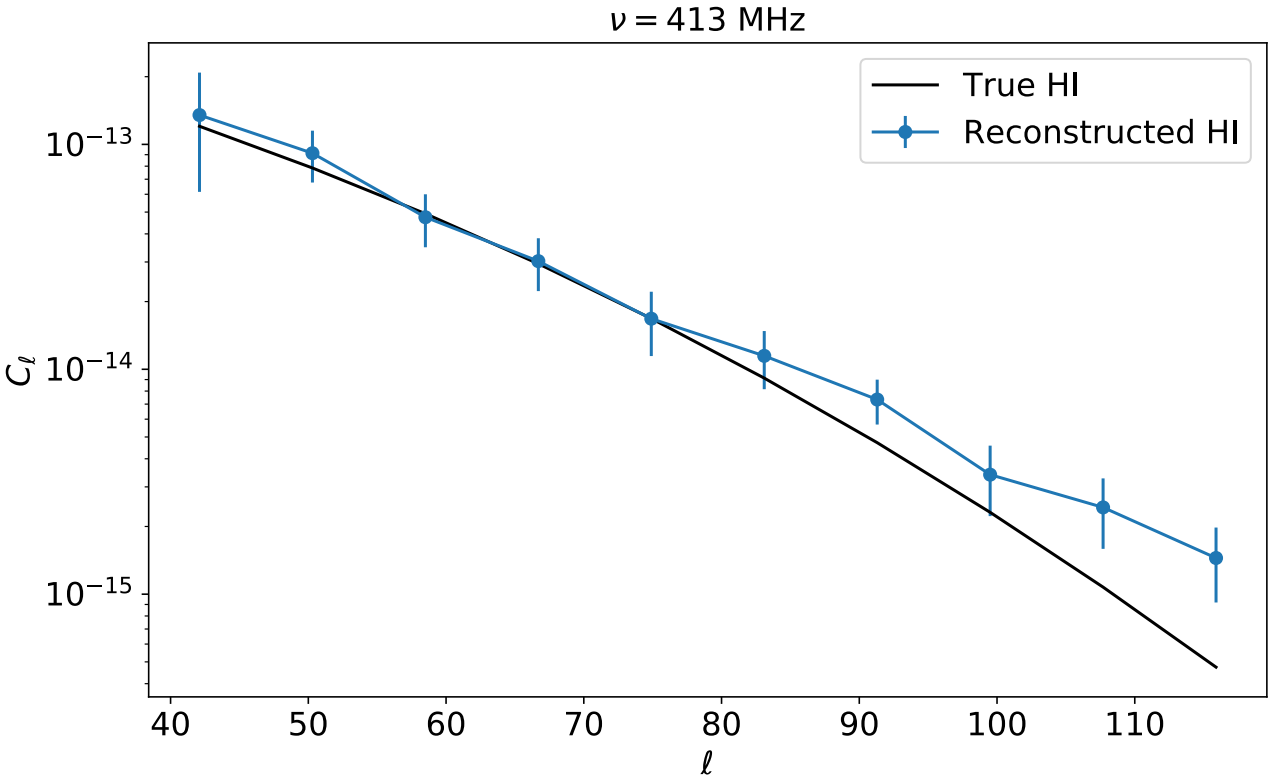
Map results



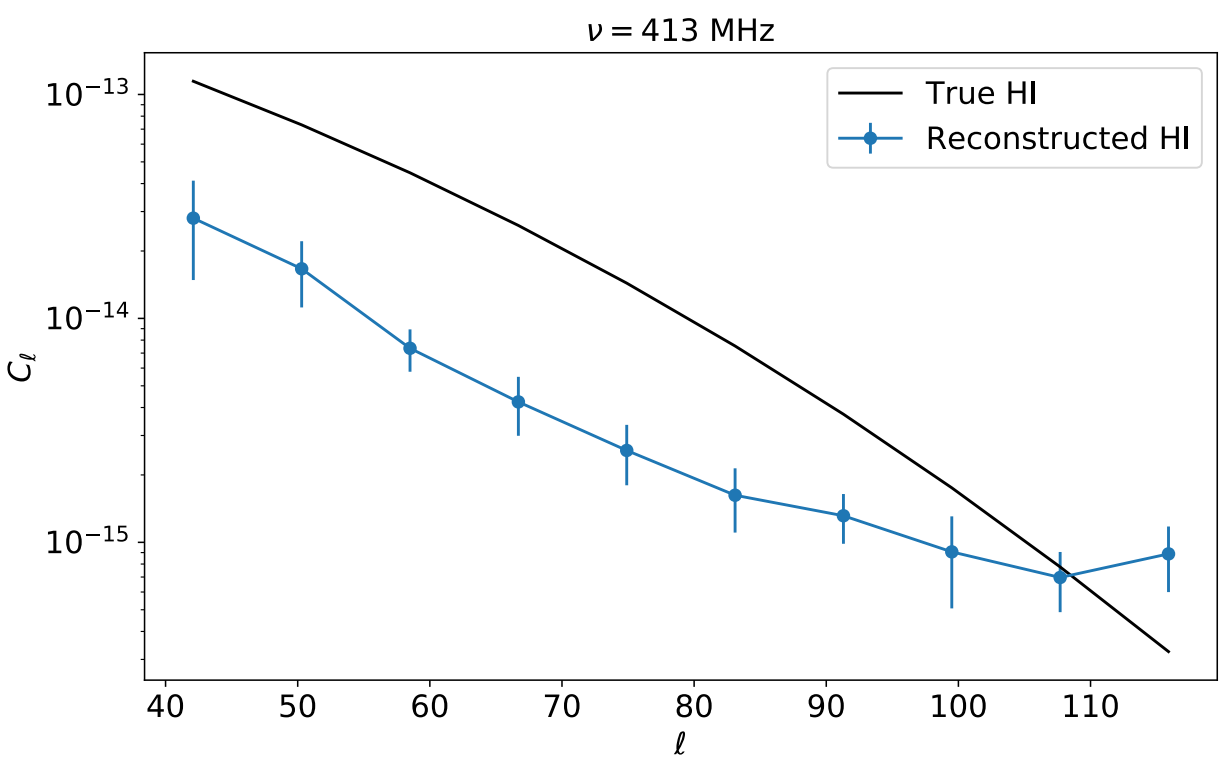
- Basically recover the HI signal
- Localised foreground residuals visible

Power spectrum comparison

Auto-encoder



PCA with 7 modes removed



- Auto-encoder is able to recover an unbiased HI spectrum
- No over-subtraction compared with PCA

Conclusions

- The simple autoencoder is able to recover majority of the HI
- Doesn’t show signal loss
- Sensitive to systematics but not the exact level

- Needs more complicated layers to
  - Improve the foreground subtraction
  - Robust against systematics

- Adapt to healpix scheme for full sky maps

- Try on the the foreground challenge data cube