21cm foreground removal with machine learning

Tianyue Chen

Postdoctoral Researcher

Ecole Polytechnique Fédérale de Lausanne (EPFL)

SKACH summer 2023

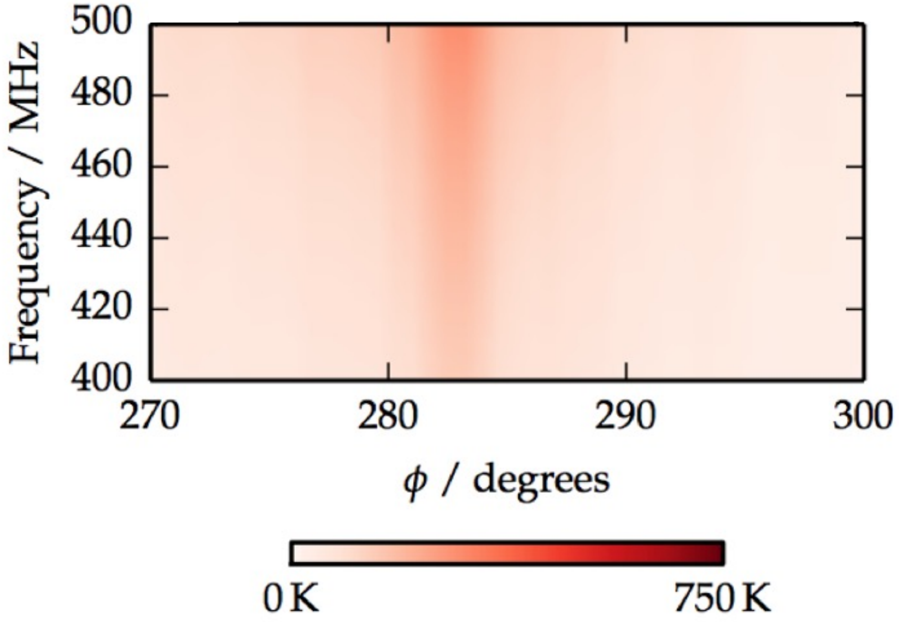
Introduction

- IM as one main goal of SKA
- Foreground critical for 21cm detection
- Large SKA dataset incoming

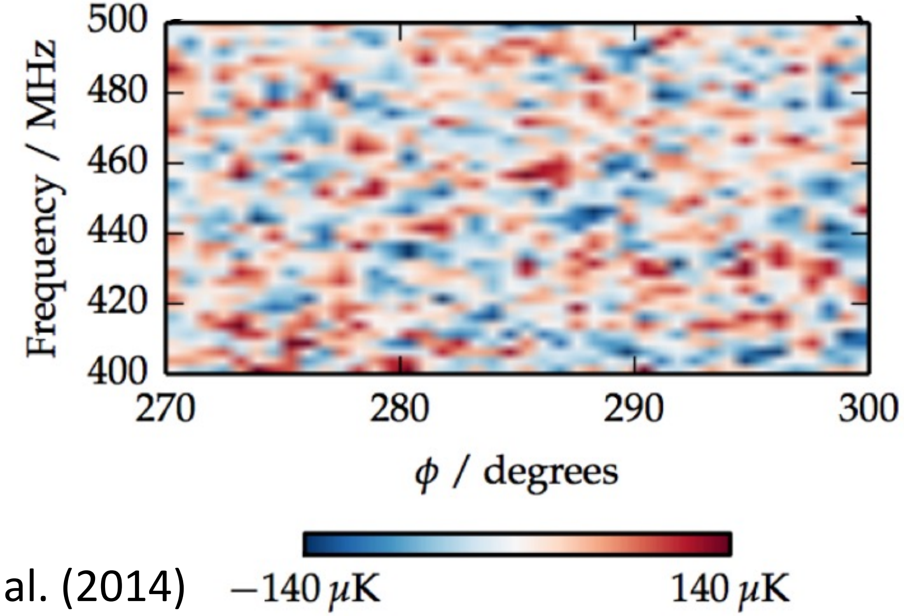
- Traditional approach:
  - Sensitive to systematics (e.g., KL filter)
  - Signal loss (e.g., PCA)

- Machine learning approach?
  - Consistent with traditional approach?
  - Consistent with different models?
  - Robust againt systematics

Synchrotron



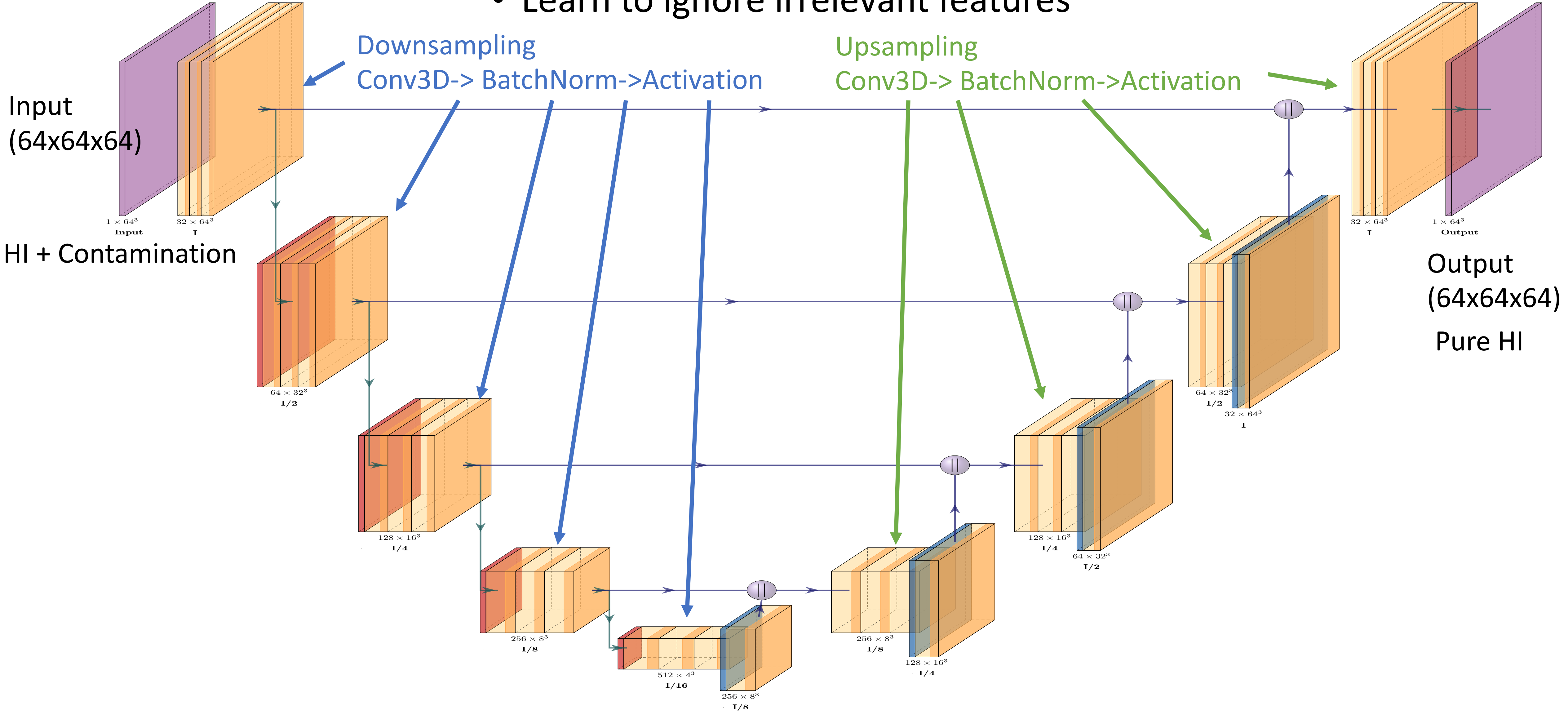
HI



Shaw et al. (2014)

U-net for IM

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



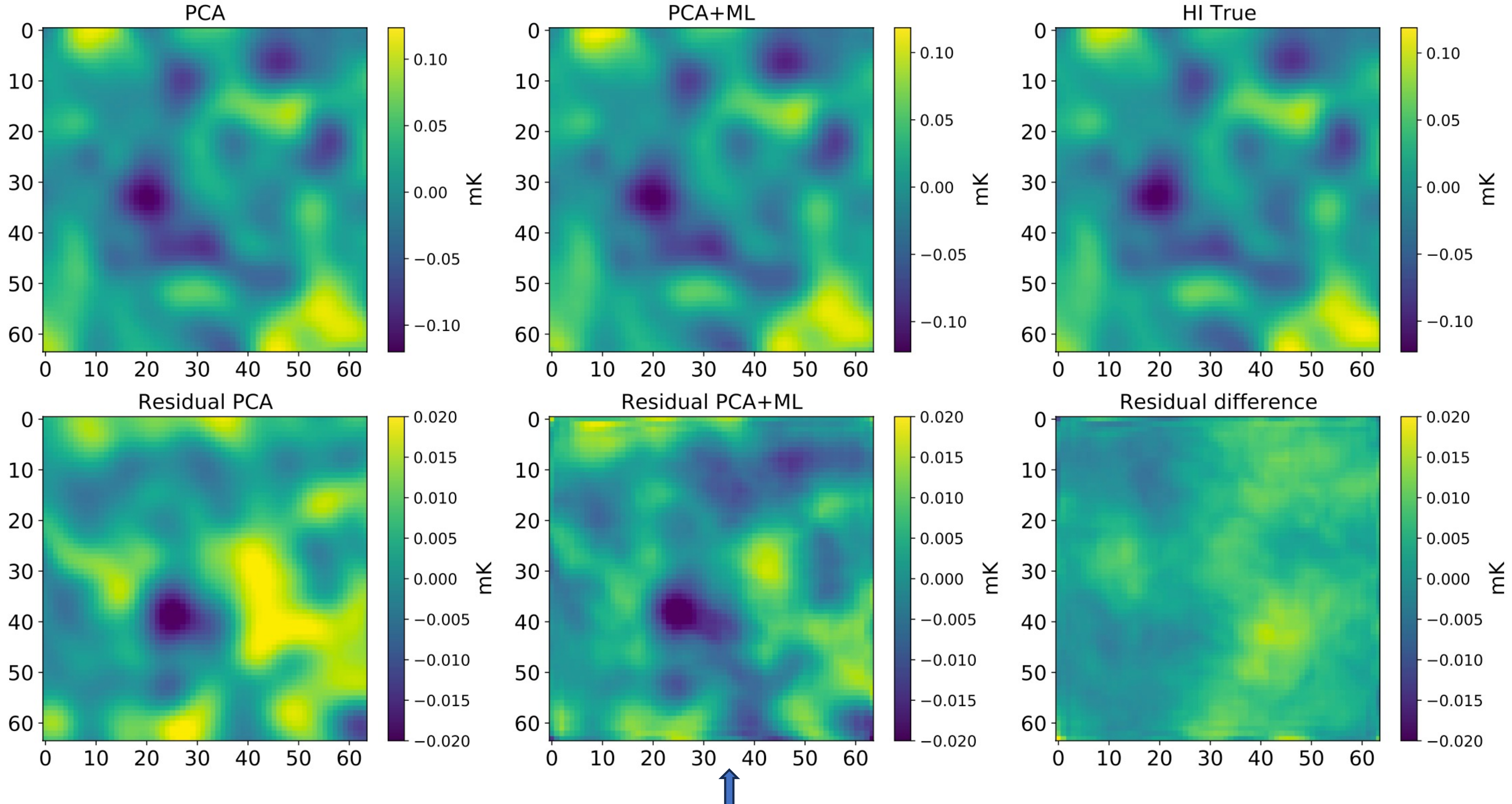
Simulation models

- **Sky models:**
  - Santos et al. (2005) – Gaussian realization according to power spectrum
  - CoLoRe – HI simulated by lognormal fields
  - Planck Sky Model – FG simulated by observed maps
- **Instrumental effect:**
  - Uniform Gaussian beam

    SKA-mid single dish beam (700-1020 MHz, 64 channels)
  - Frequency-dependent Gaussian beam ($\frac{\lambda}{D}$)
  - Sinusoidal gain drift $G(\nu) = G_0 \sin(G_1 \nu + G_2)$
- **Format:**
  - Healpix full sky maps $\rightarrow$ 192 equal-size patches (64x64x64)

Baseline results - map

- Santos model
- PCA 2 mode pre-processing for dynamic range

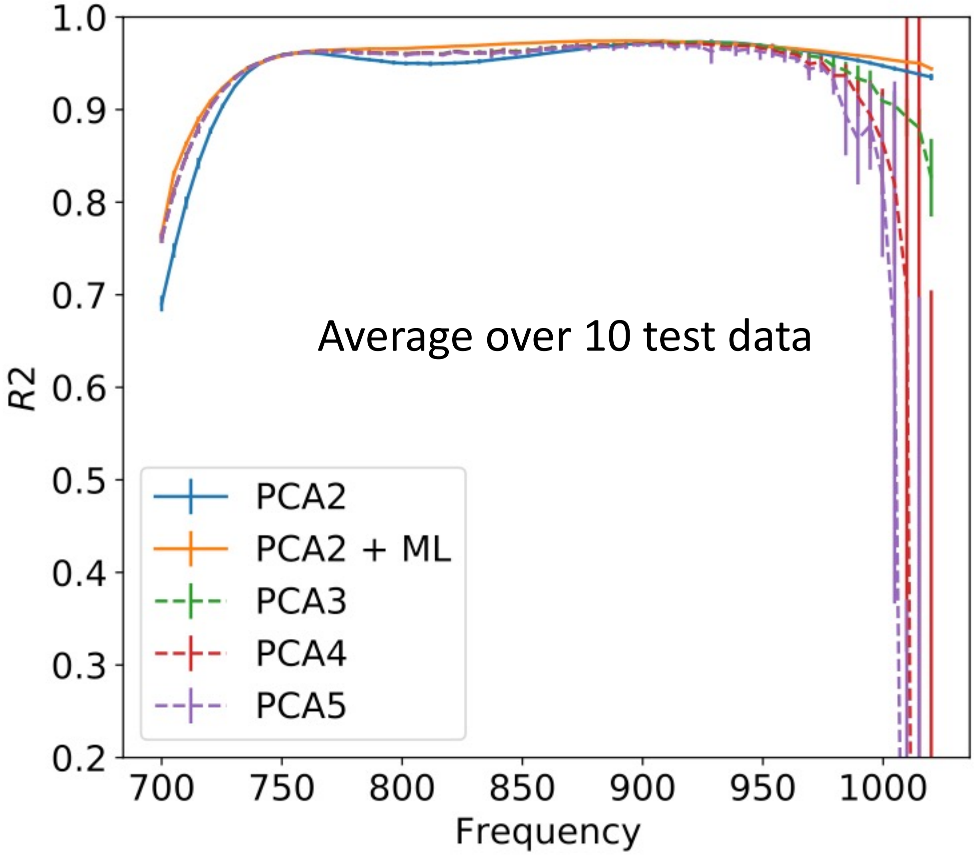


Less residual

Baseline results – R2 score

$$R^2 = 1 - \frac{SS_{\text{res}}}{SS_{\text{tot}}}$$

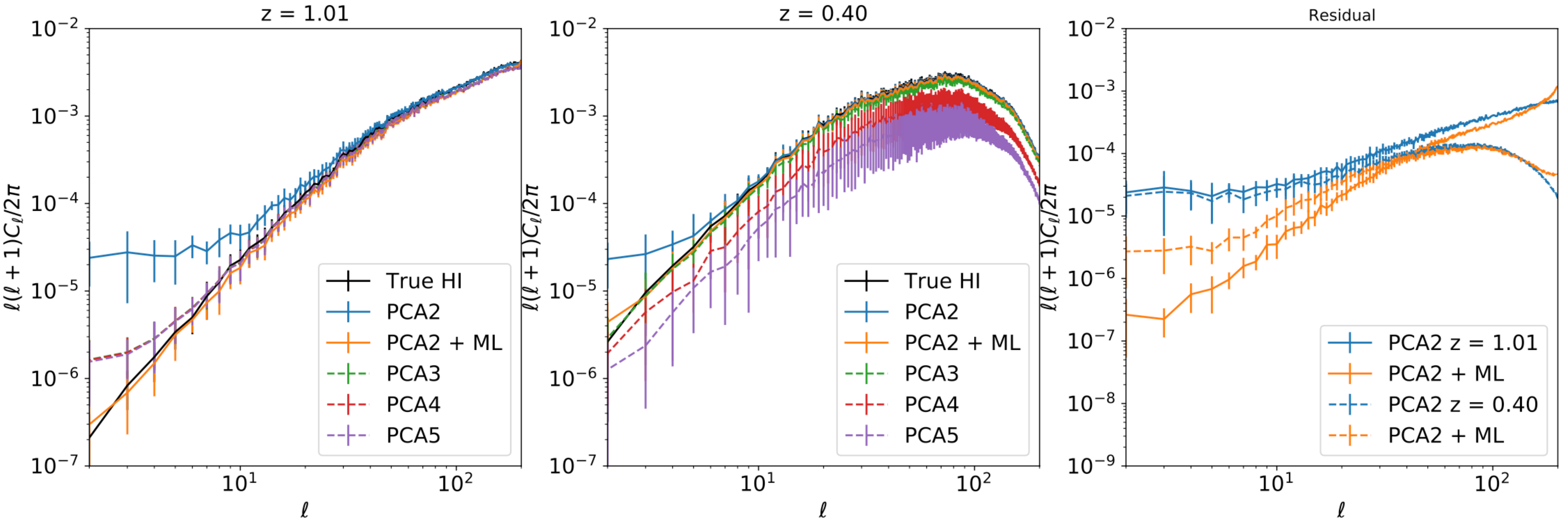
Evaluate the performance of the ML model in image space
Accuracy measurement of predictions v.s. target



Better accuracy in image space compared with PCA alone

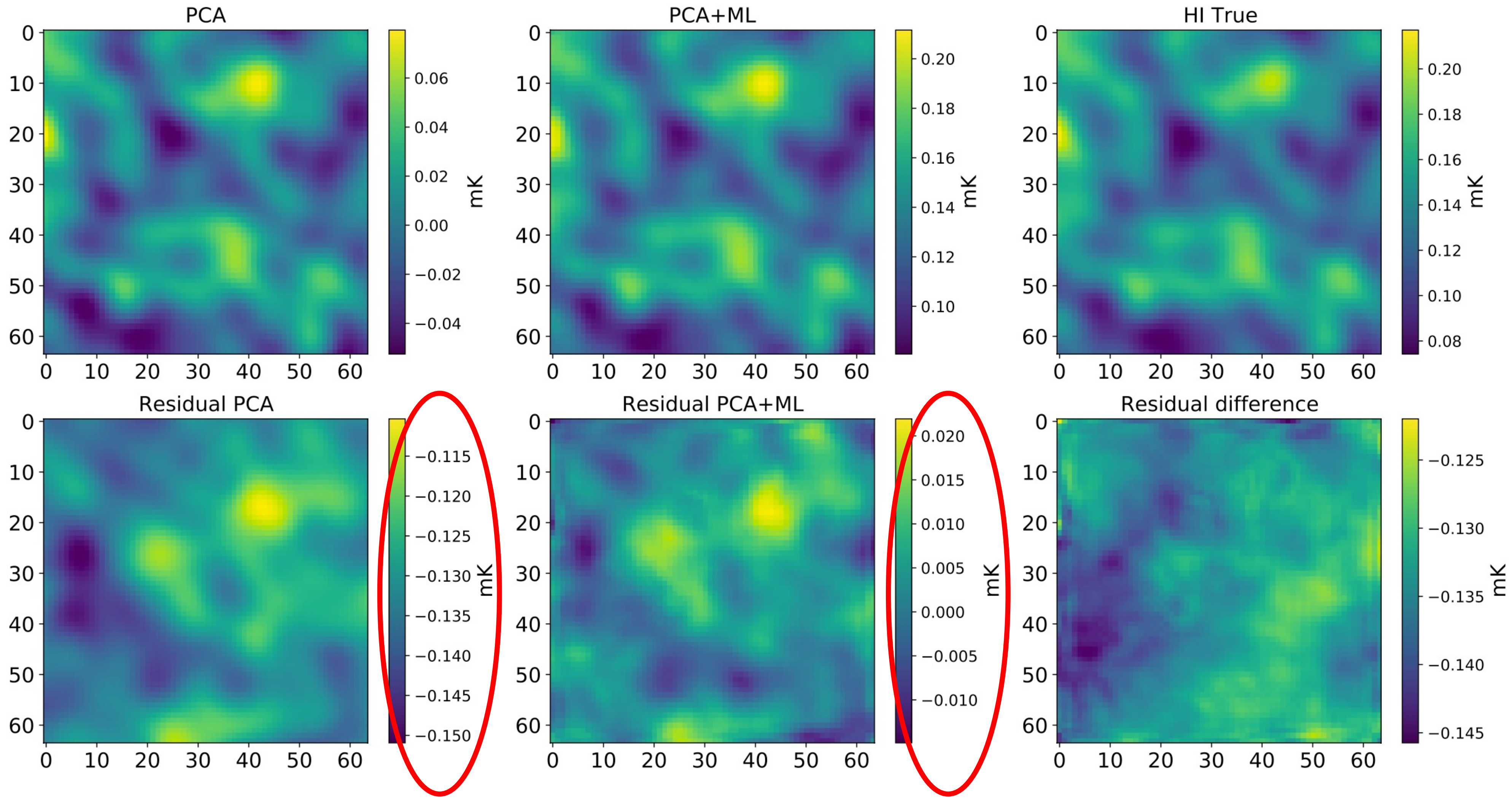
Baseline results – power spectrum

Average over 10 test data

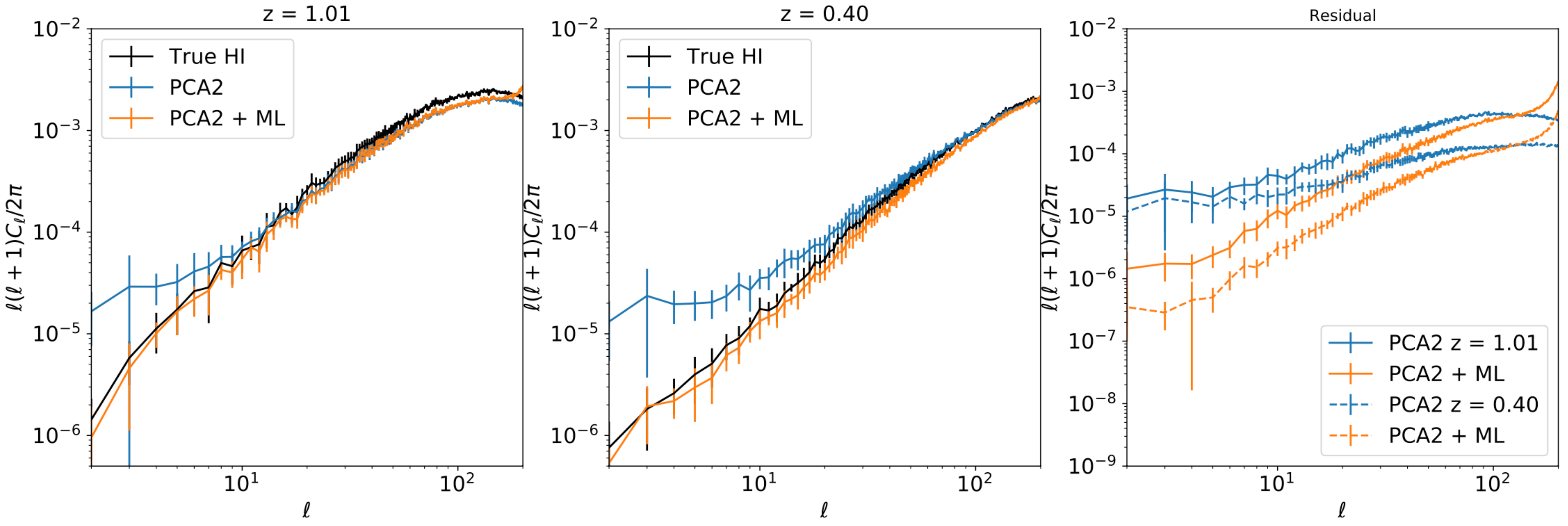


ML reduces the FG residual compared with PCA2
ML has less over-subtraction compared with PCA 3+

CoLoRe results (different HI) - map



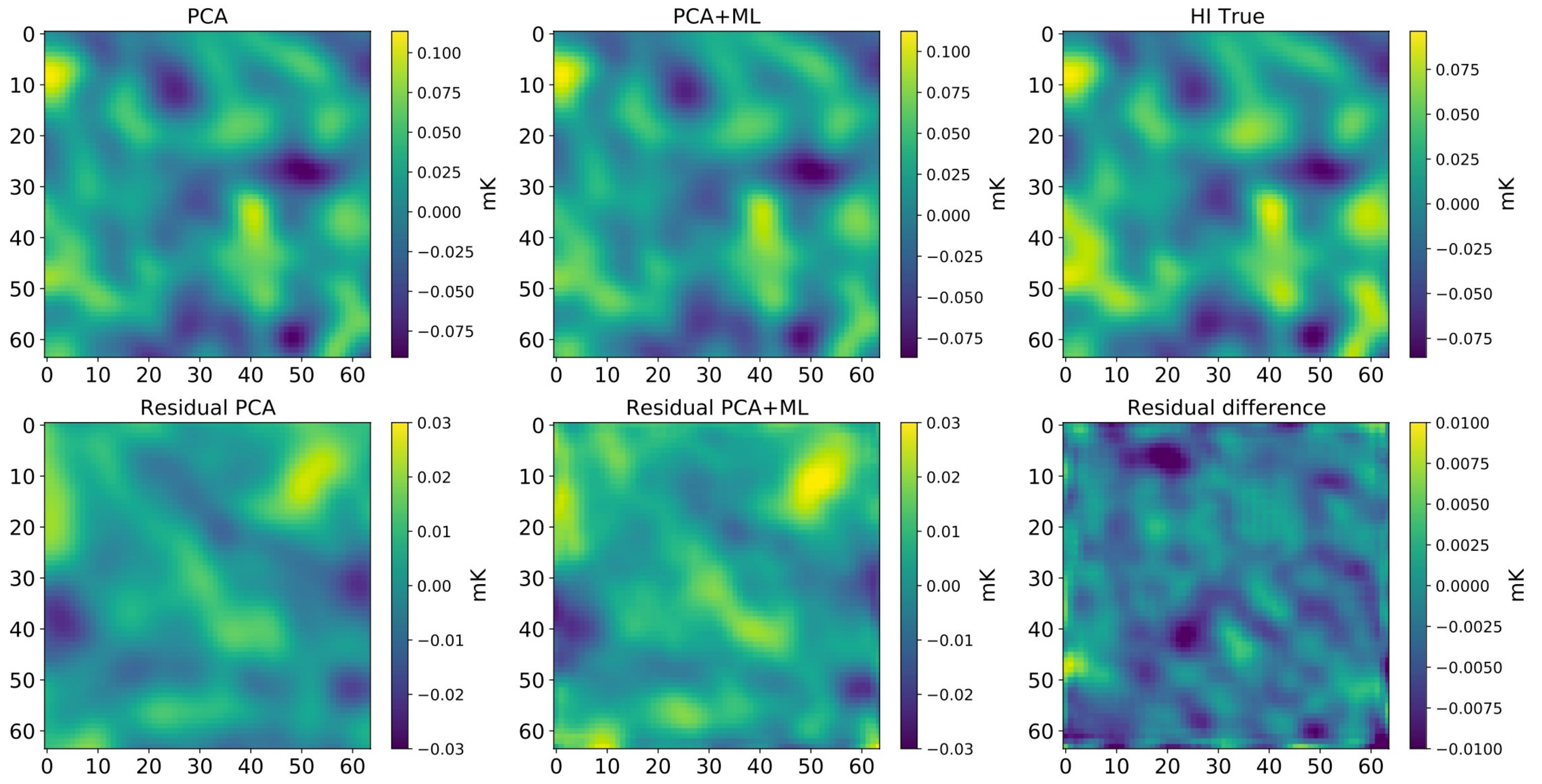
CoLoRe results (different HI) – power spectrum



Reduced residual compared with ML alone
Consistent with default Santos model

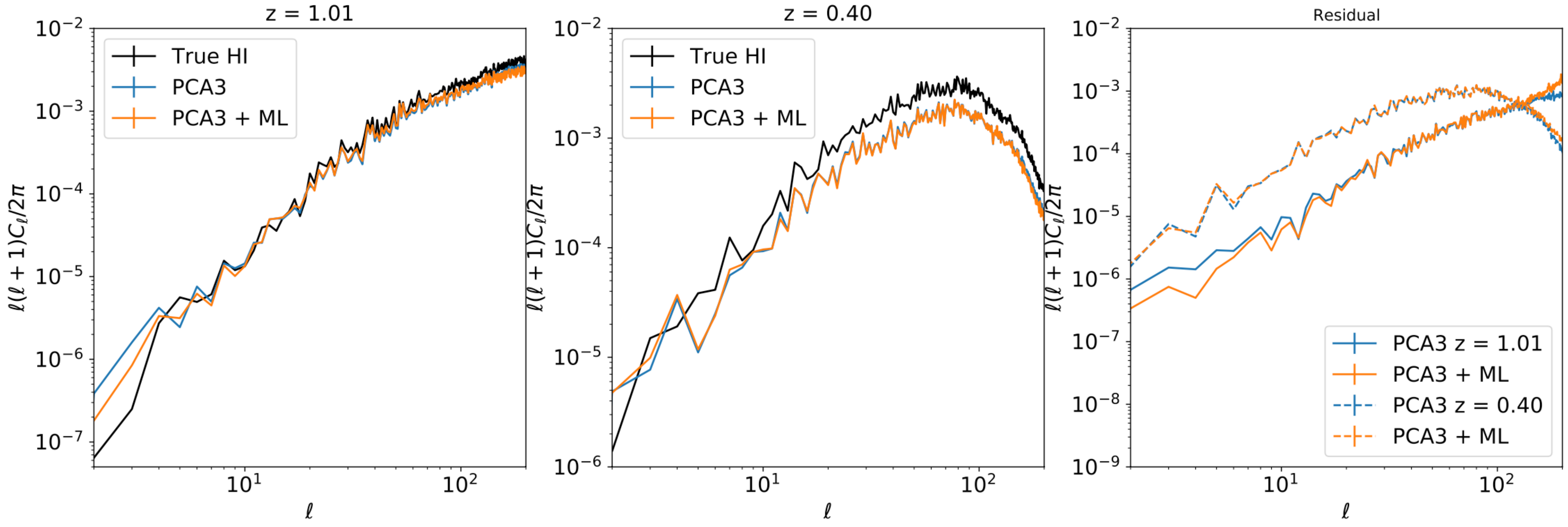
PSM results (different FG) - map

PCA 3 is applied for galactic plane



PSM results (different FG) – power spectrum

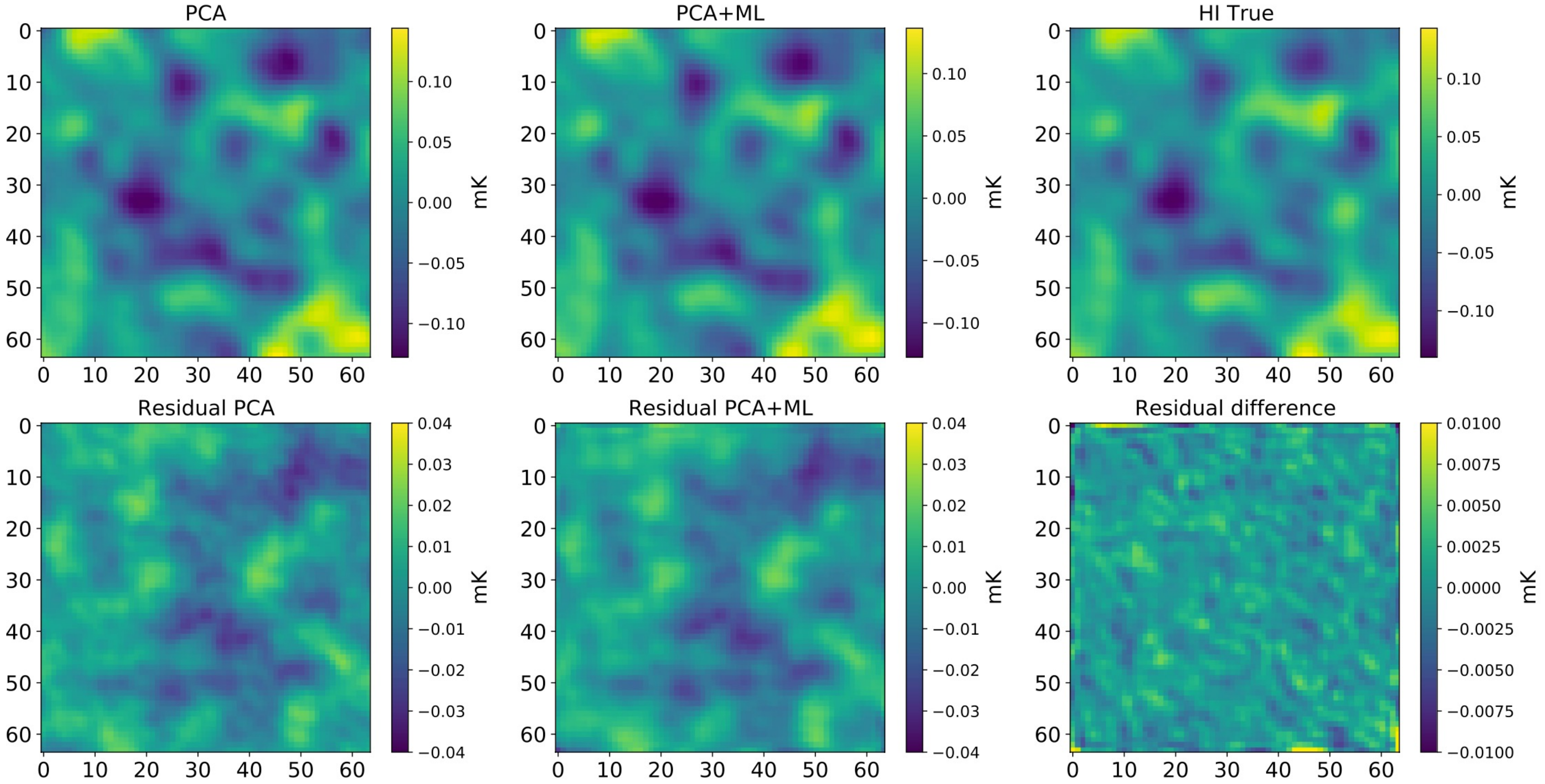
PCA 3 used for pre-processing



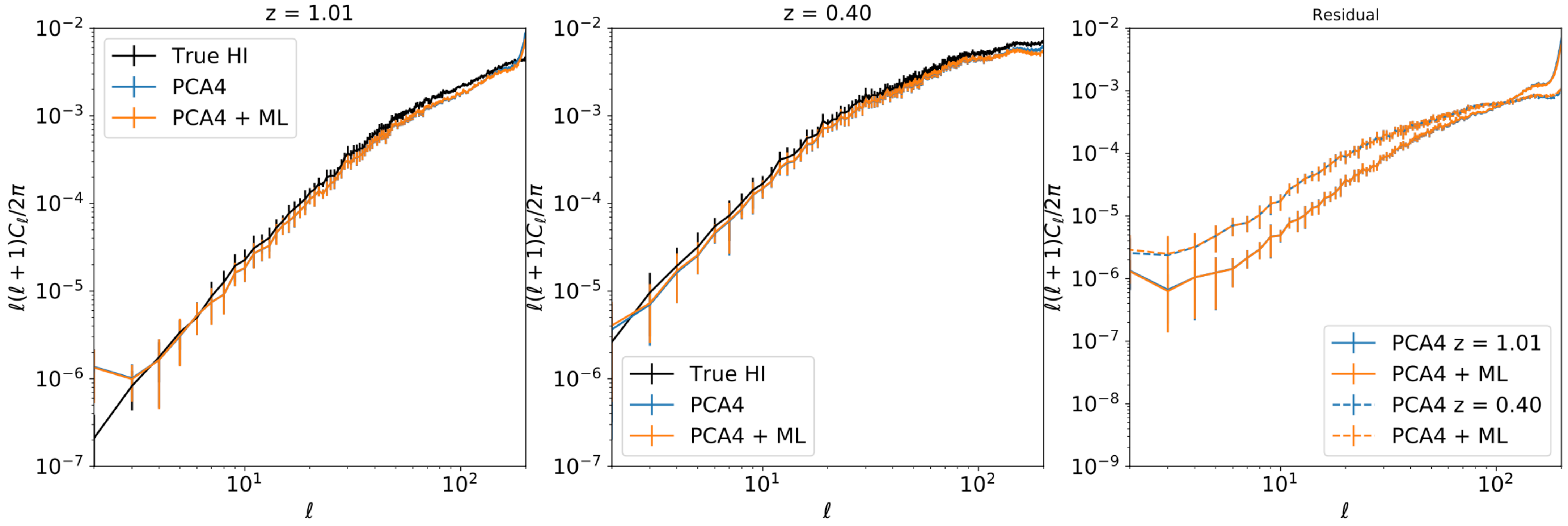
ML subject to over-subtraction from PCA 3

Frequency-beam (untrained) - map

Network same as before
Surprise with frequency-beam test data
Needs PCA4 to reduce dynamic range



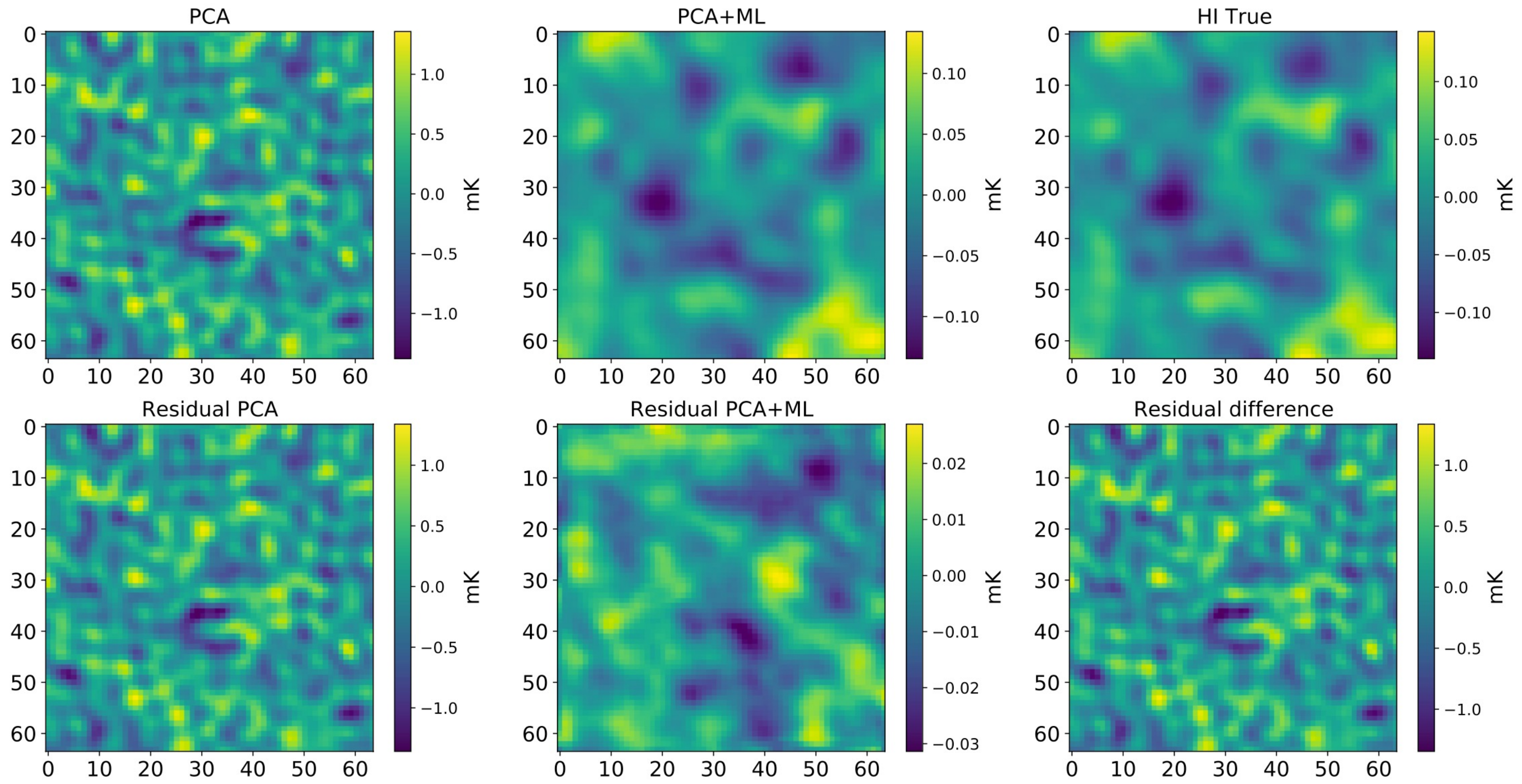
Frequency-beam (untrained) – Power Spectrum



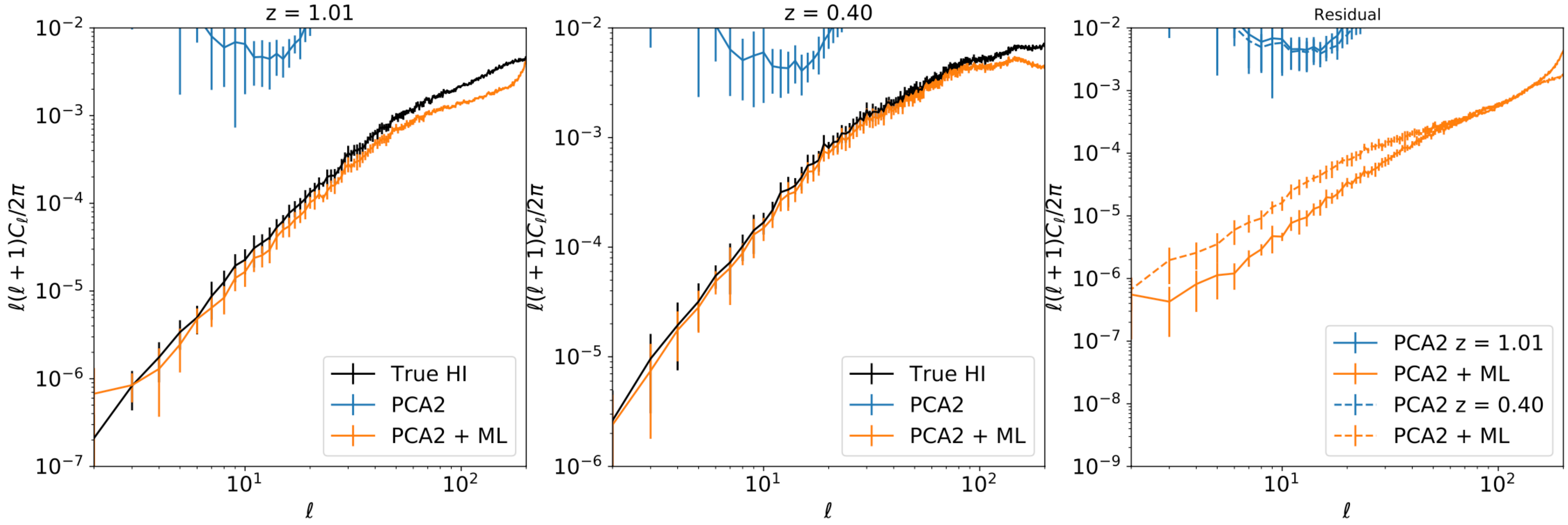
ML subject to over-subtraction from PCA 4

Frequency-beam (pre-trained) - map

Network trained with frequency-beam data
Consistent with test data
PCA2 for pre-processing



Frequency-beam (pre-trained) – power spectrum

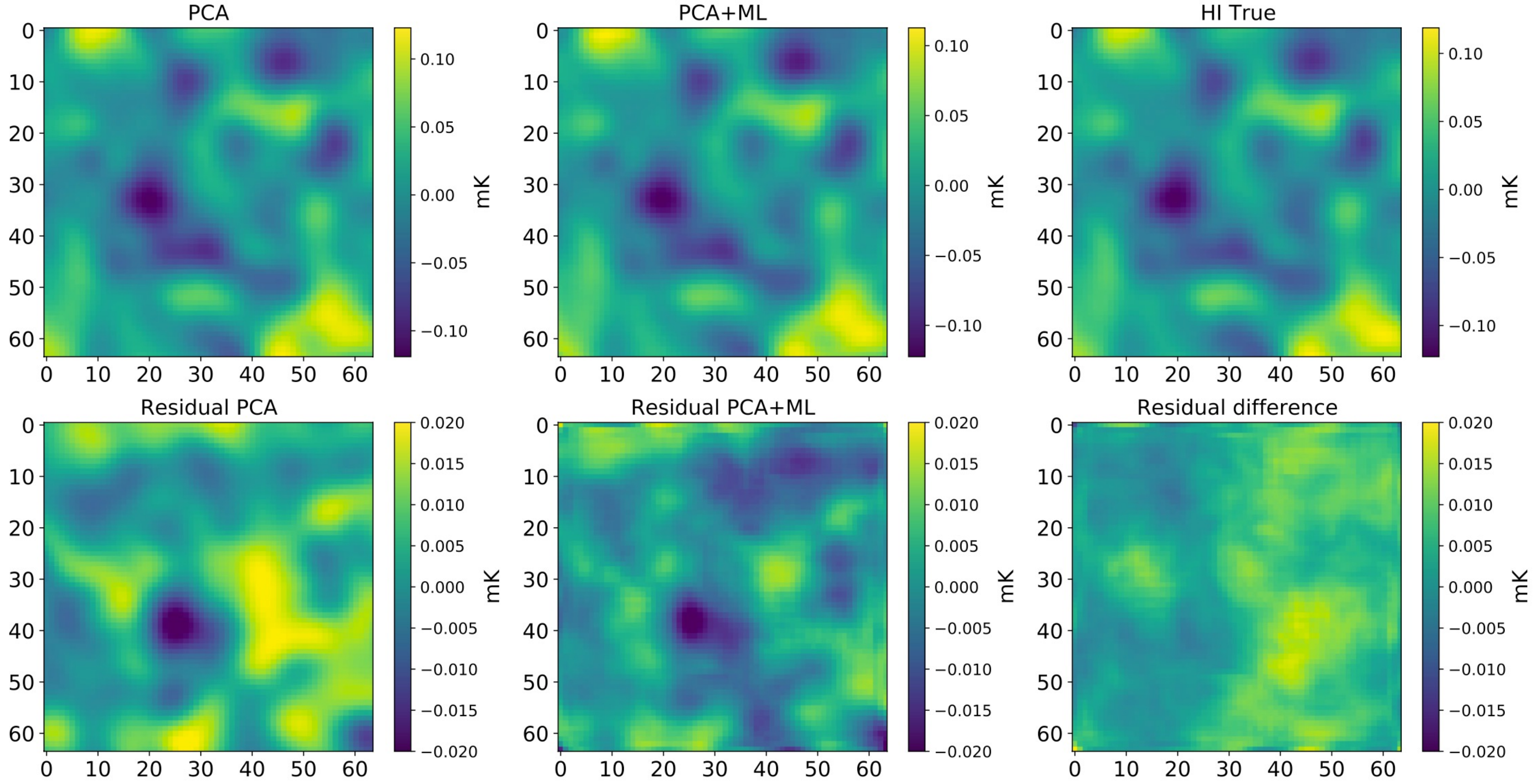


Significantly reduce beam-induced resdiduals
Extra prior information is critical for network training

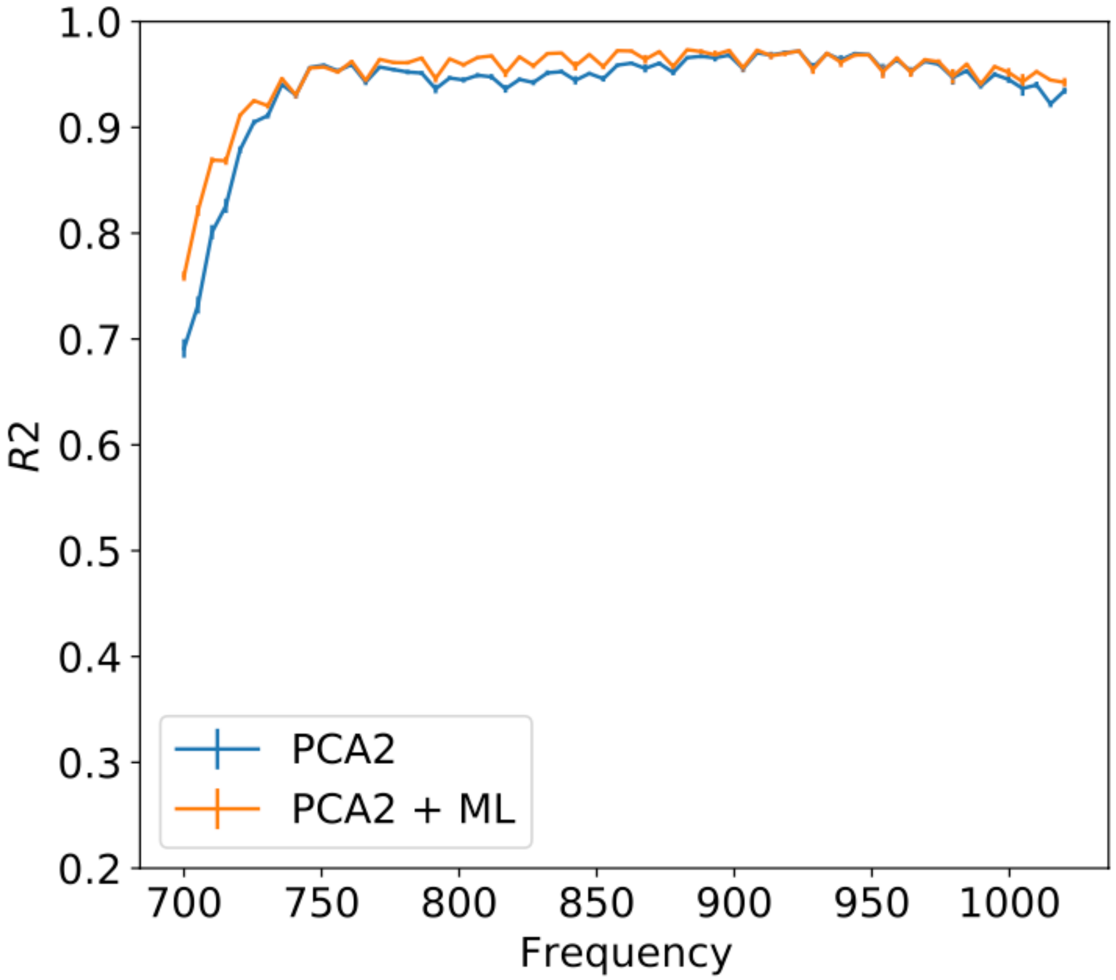
Gain drift (pre-trained) - map

$$G(\nu) = G_0 \sin(G_1 \nu + G_2)$$

$$G_0 = 0.1, G_1 = 10, G_2 = 0.5$$

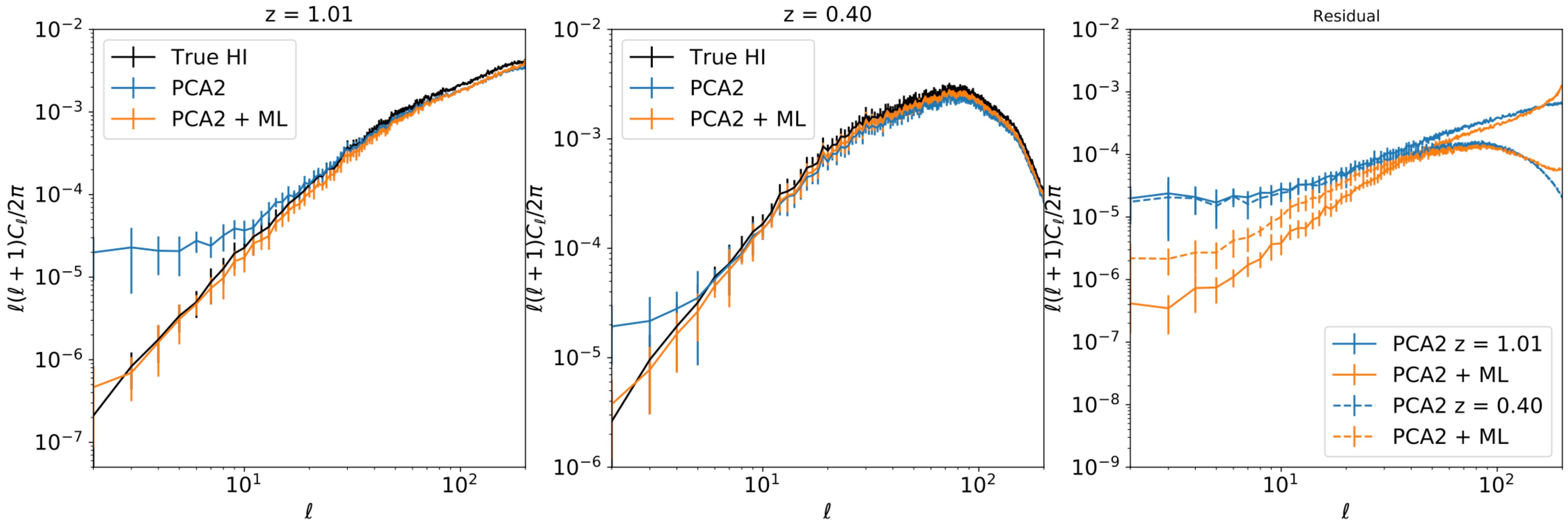


Gain drift (pre-trained) - map



Visible frequency structure due to gain drift
Comparable results to non-gain drift case
Gain drift doesn't affect ML performance

Gain drift (pre-trained) – power spectrum



Conclusions

- ML is consistent with traditional approach, with advantages at certain cases
- ML is consistent with different simulation models
- ML is limited to pre-processing and may subject to over-subtraction
- ML needs extra prior information to handle systematics