The robustness of machine learning for 21cm foreground removal

ArXiv: 2311.00493

Tianyue Chen

Postdoctoral Researcher

Ecole Polytechnique Fédérale de Lausanne (EPFL)

SKA SWG, Porto, Jan 2024

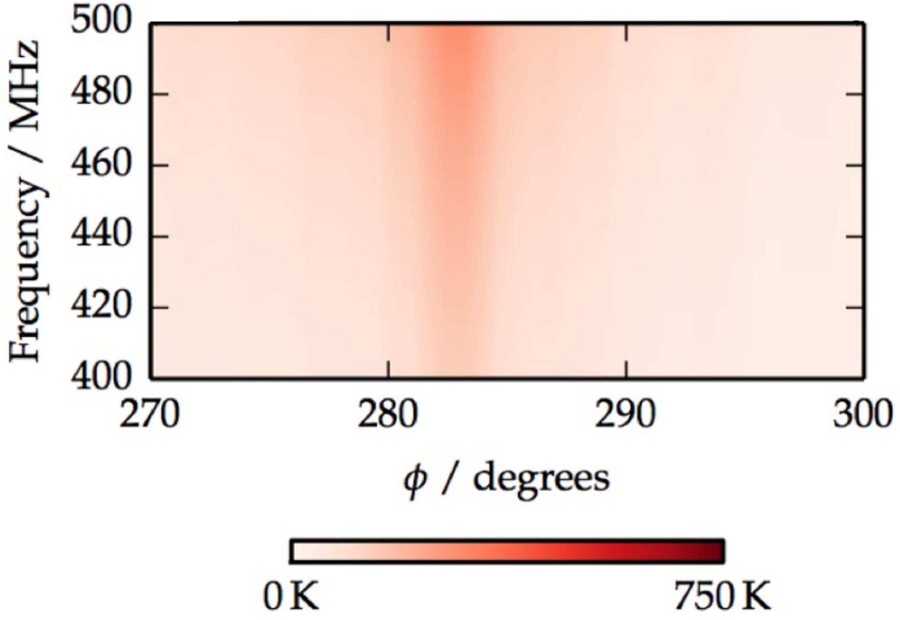
Introduction

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

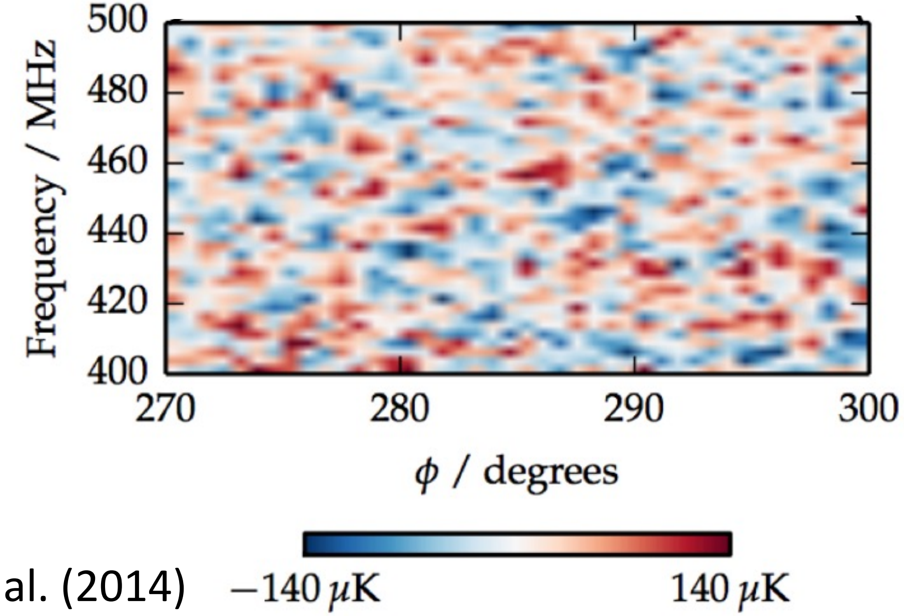
- Traditional approach:
  - Sensitive to systematics (e.g., KL filter)
  - Signal loss (e.g., PCA)

- Machine learning algorithm?
  - Comparable with mature technique?
  - Consistent under different models?
  - Robust againt systematics?

Synchrotron



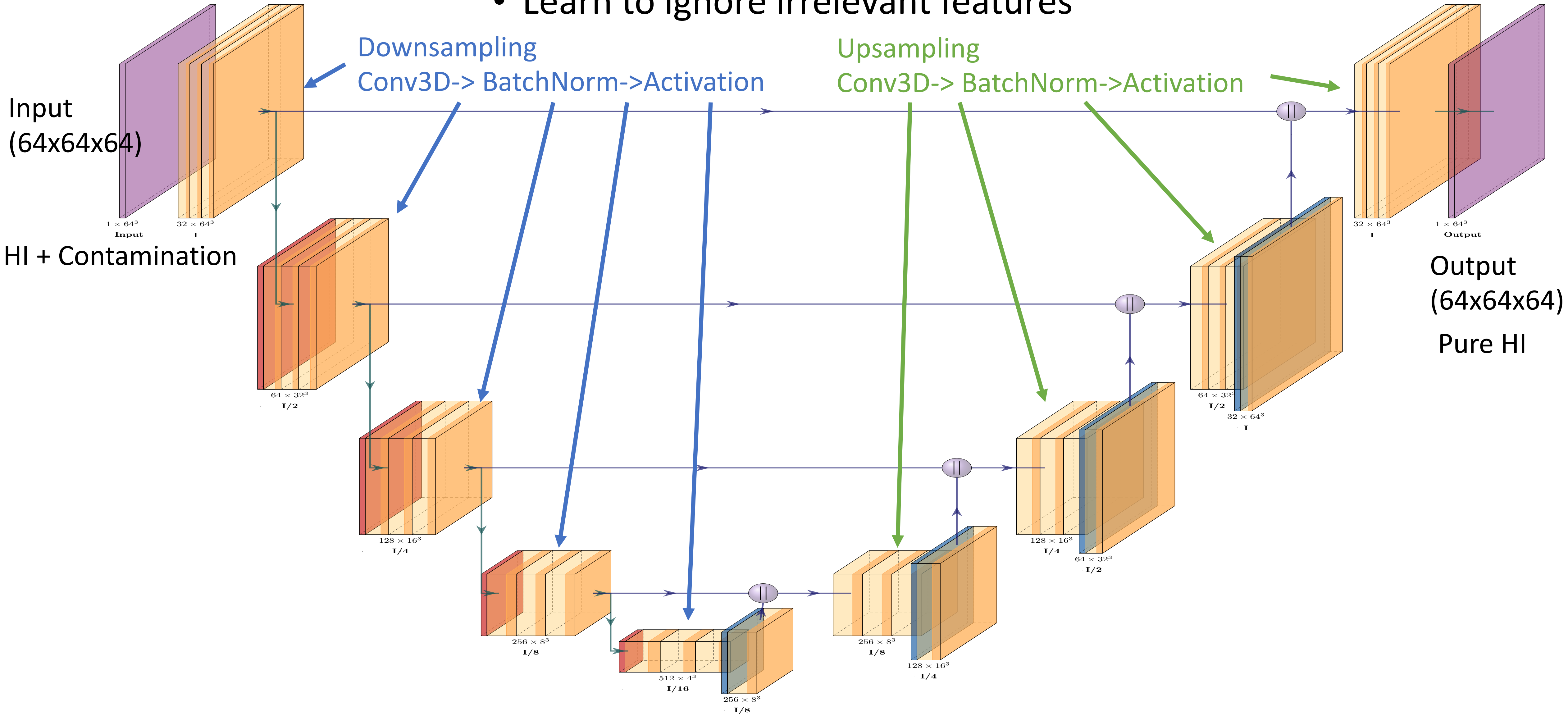
HI



Shaw et al. (2014)

U-net for IM

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



Sky models

- MS model (Gaussian model):
  - Santos et al. (2005)
  - FG: $$C_\ell(\nu_i, \nu_j) = A\left(\frac{1000}{\ell}\right)^\beta \left(\frac{\nu_{\rm ref}^2}{\nu_i \nu_j}\right)^\alpha I_\ell^{ij}$$
  - HI: Battye et al. 2013 $$\bar{T}_{\rm obs}(z) = 44\mu{\rm K}\left(\frac{\Omega_{\rm HI}h}{2.45^{-4}}\right)\frac{(1+z)^2}{E(z)} \qquad C_\ell = \frac{H_0 b^2}{c}\int {\rm d}z E(z)\left[\frac{W(z)\bar{T}(z)D(z)}{r(z)}\right]^2 P_{\rm cdm}\left(\frac{\ell+\frac{1}{2}}{r}\right)$$
- CoLoRe model (non-Gaussian HI):
  - HI: Lagrangian perturbation theory
- Planck Sky Model (non-Gaussian FG):
  - Synchrotron : Haslam 408 map;
  - Free-free : Hα template;
  - Point source: NVSS catalogue;

Instrumental systematics

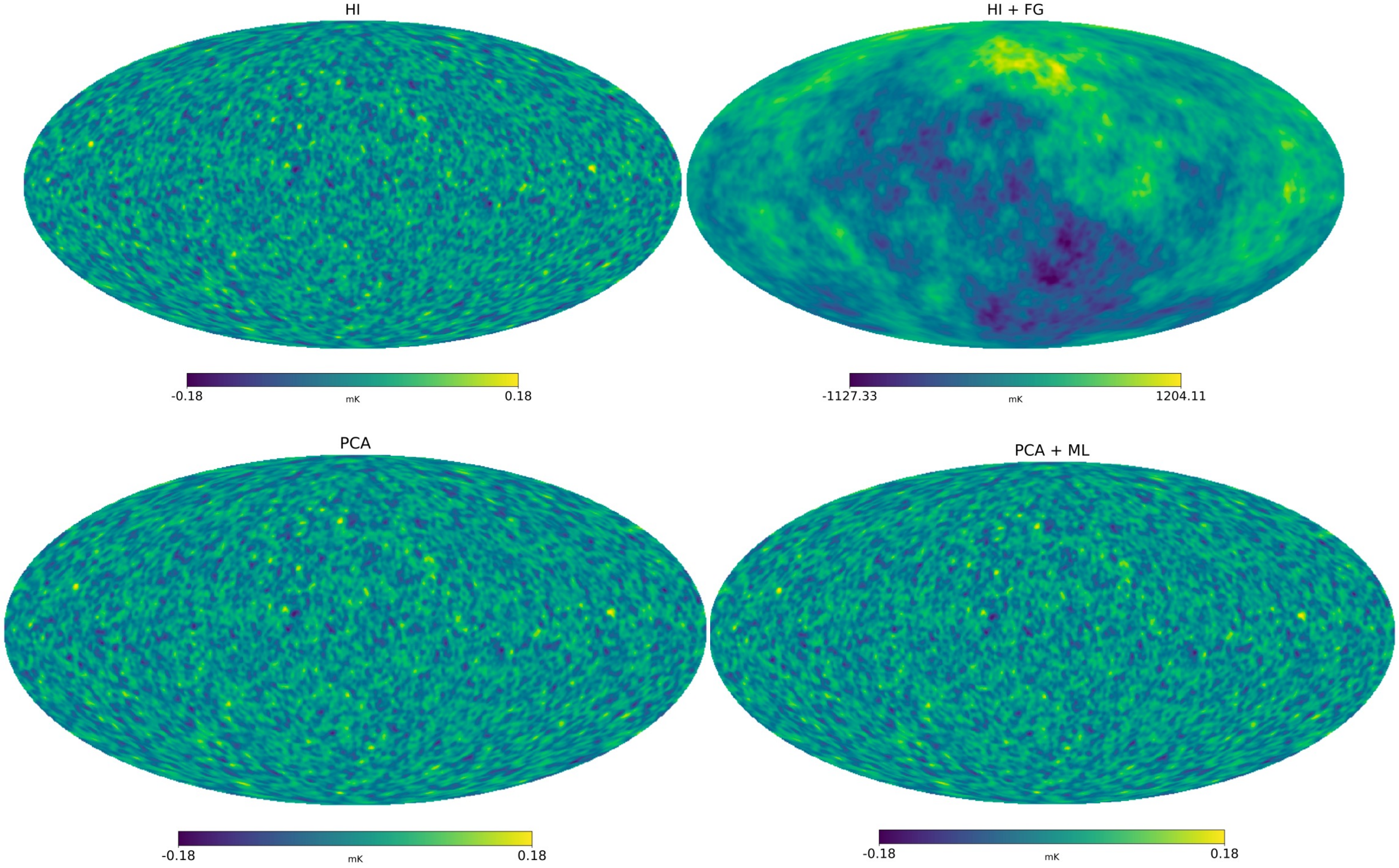
- Instrumental parameters:
  - Beam: SKA-mid single dish Gaussian beam
  - Frequency range: 700-1020 MHz, 64 channels
- Instrumental systematics:
  - Frequency-dependent beam $$\theta_{\mathrm{B}}(z_i) = \theta_{\mathrm{FWHM}}(\nu_{\mathrm{mid}})\frac{\nu_{\mathrm{mid}}}{\nu_i}$$
  - Gain drift G_v = 1+ΔG_v $$\Delta G(\nu) = G_0 \sin(G_1\nu + G_2) + 1$$
- Format:
  - Healpix full sky maps $\rightarrow$ 192 equal-size patches (64x64x64)
  - Training: 40 healpix maps (7680 samples)
  - Validation: 10 healpix maps (1920 samples)
  - Test: 10 healpix maps (1920 samples)

MS model - maps

- Network can’t handle large dynamic range
- Apply PCA to pre-process the data (mode = 2), use ML for fine tuning

HI

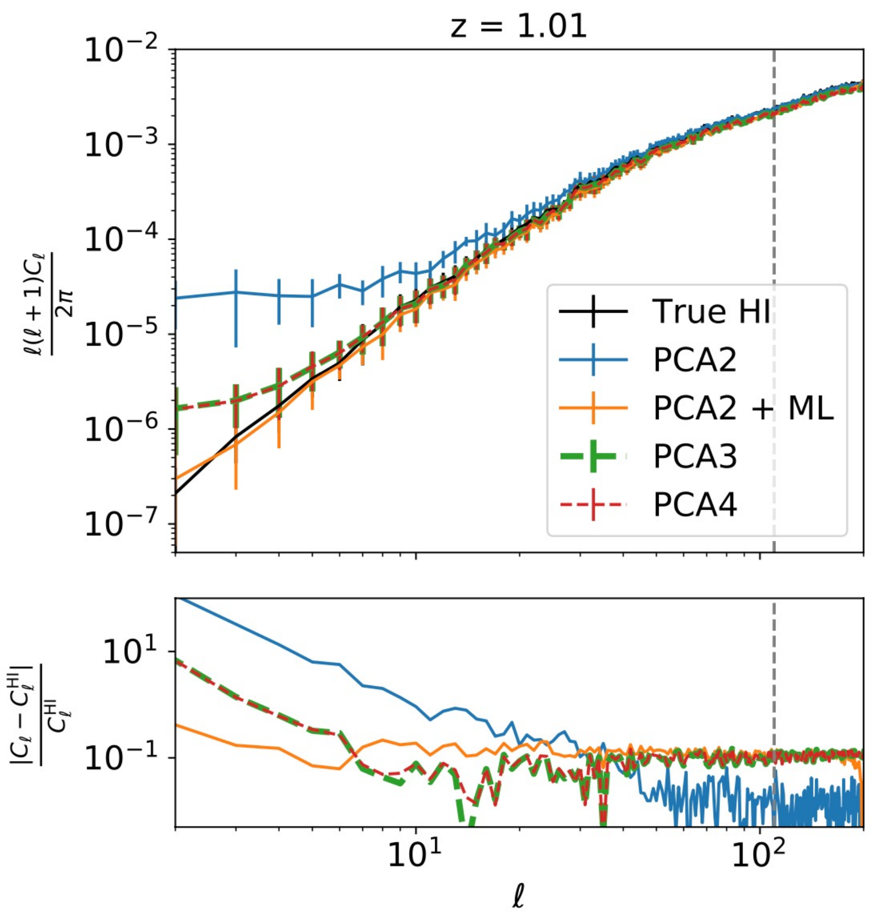
HI + FG

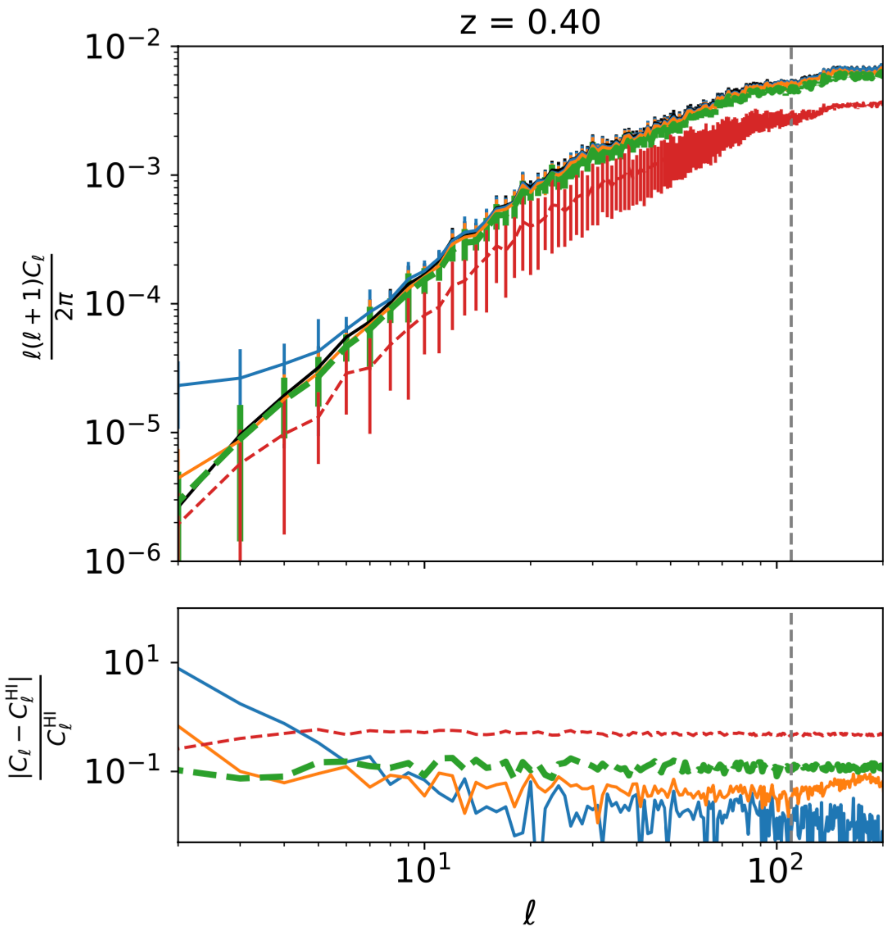


PCA

PCA + ML

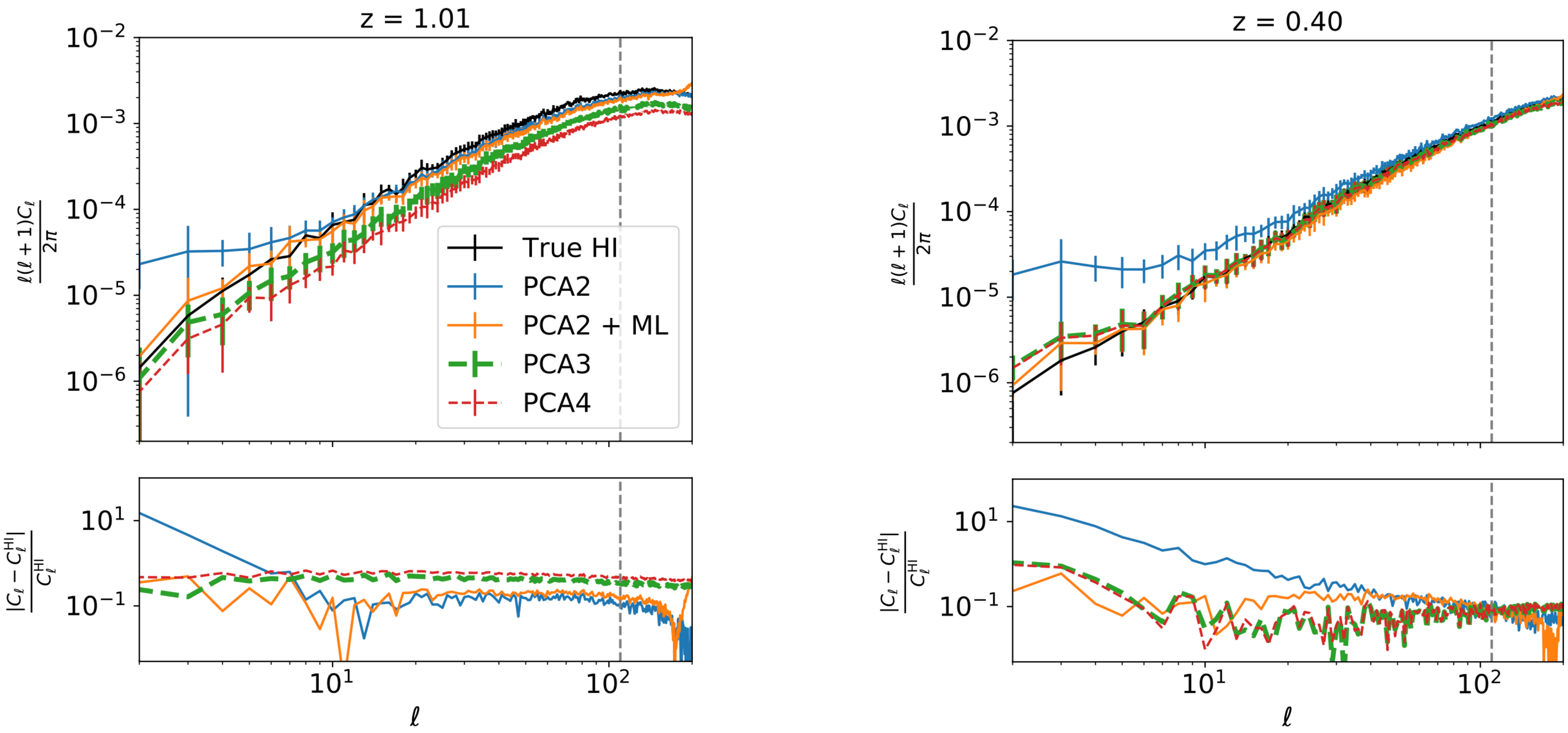
MS model – Power Spectrum





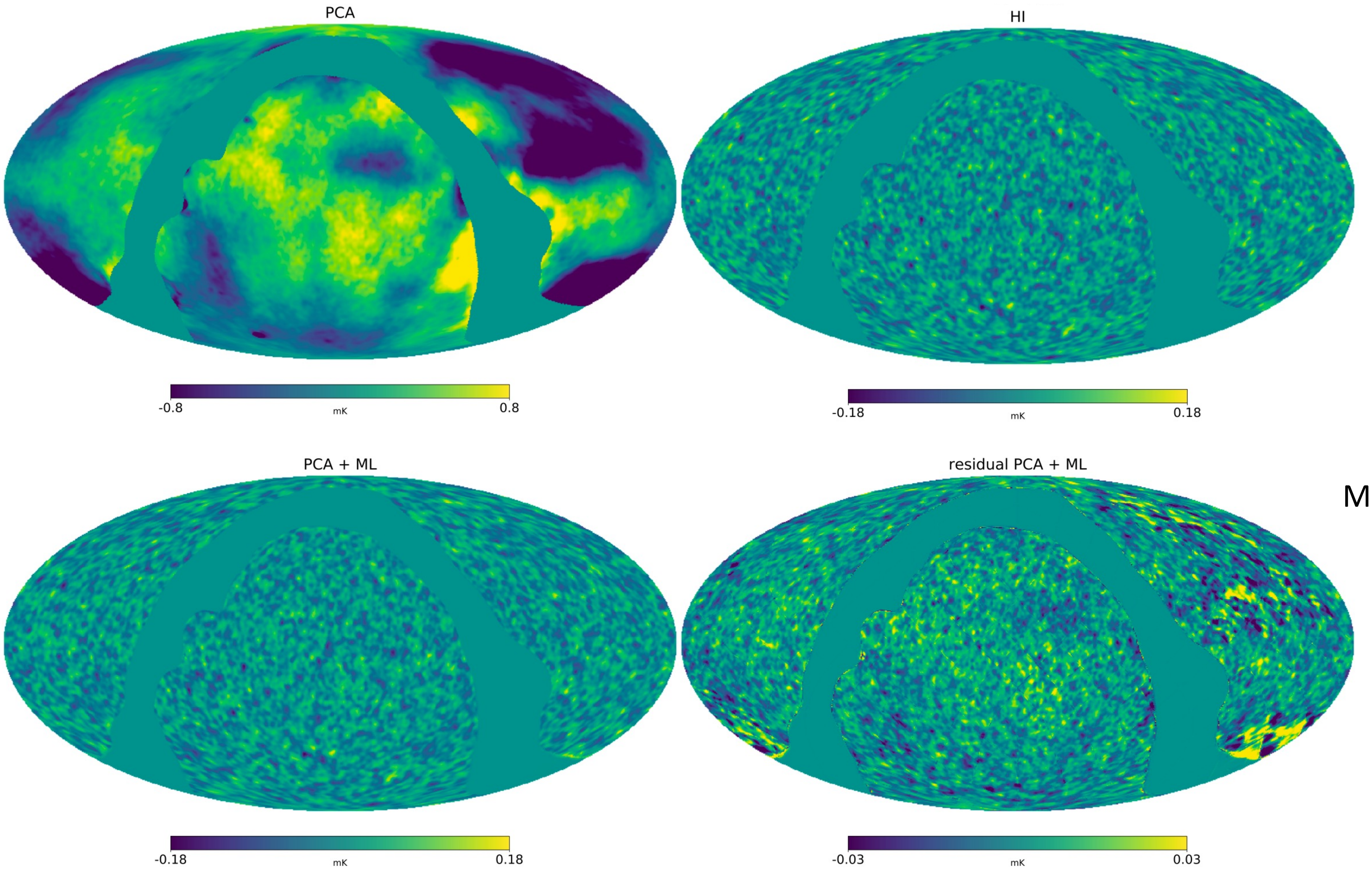
- ML particularly effective at reducing large-scale FG residuals
- ML comparable with PCA 3 alone
- ML less sensitive to redshift
- On average, fractional residual of 10% signal over all scales

CoLoRe model – Power Spectrum



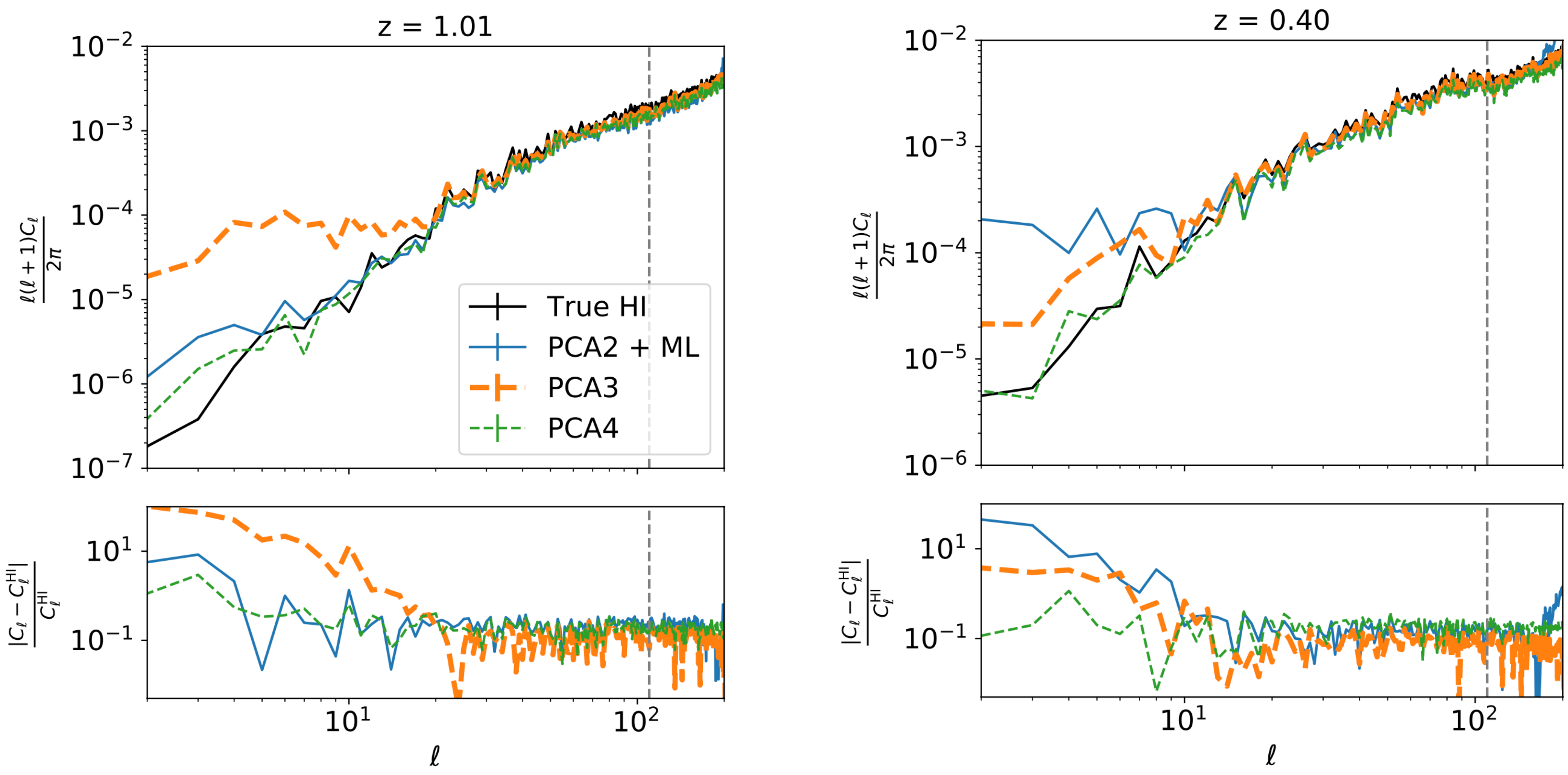
- Consistent with MS model
- Comparable with PCA 3,4 alone
- On average, fractional residual of 10% signal over all scales

PSM model – maps



ML can safely handle masks

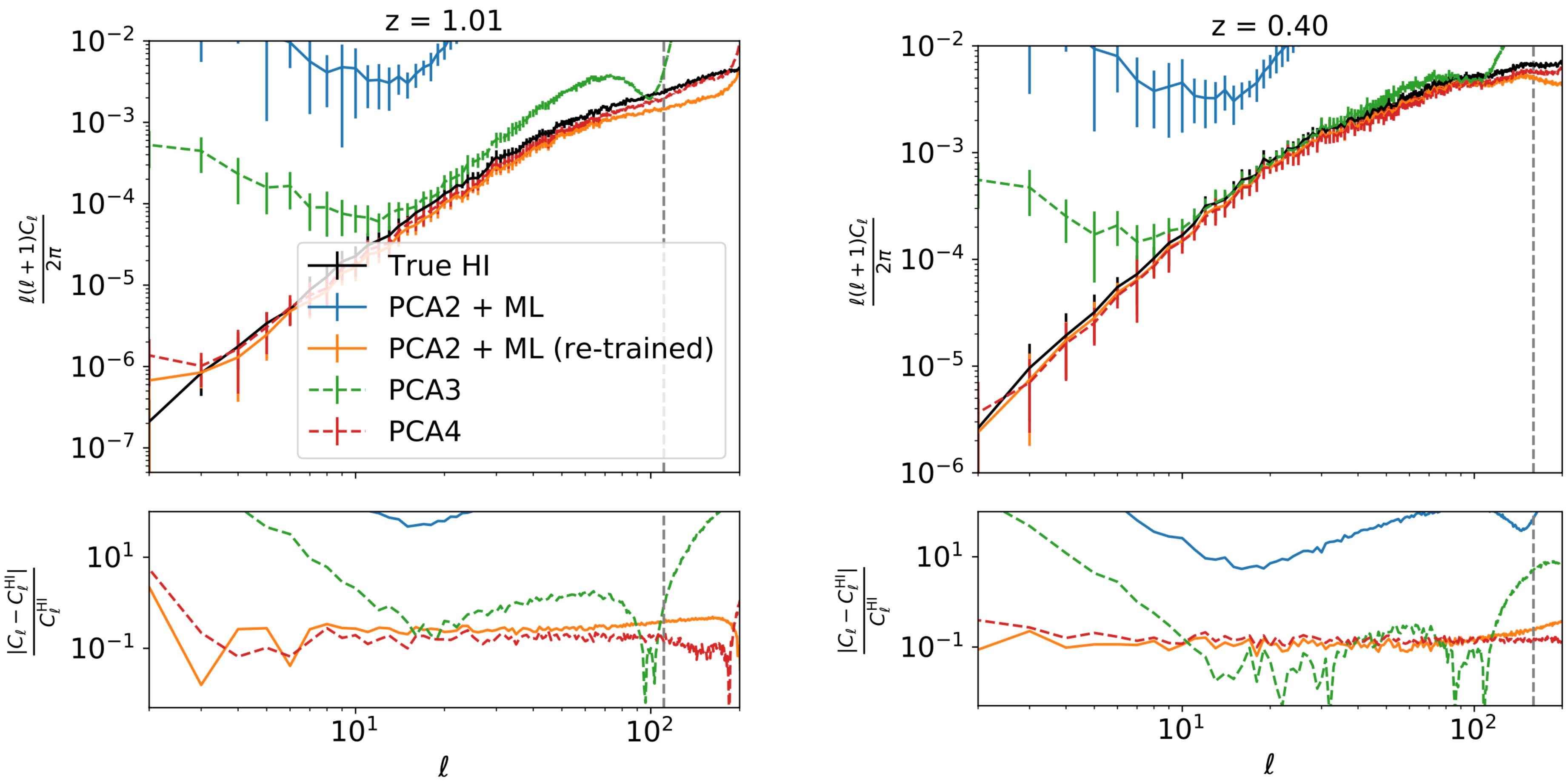
PSM model – Power Spectrum



- Affected by large residuals after PCA2 removal
- Overall comparable with MS, CoLoRe model
- Lack of signal-to-noise revolution information along redshift

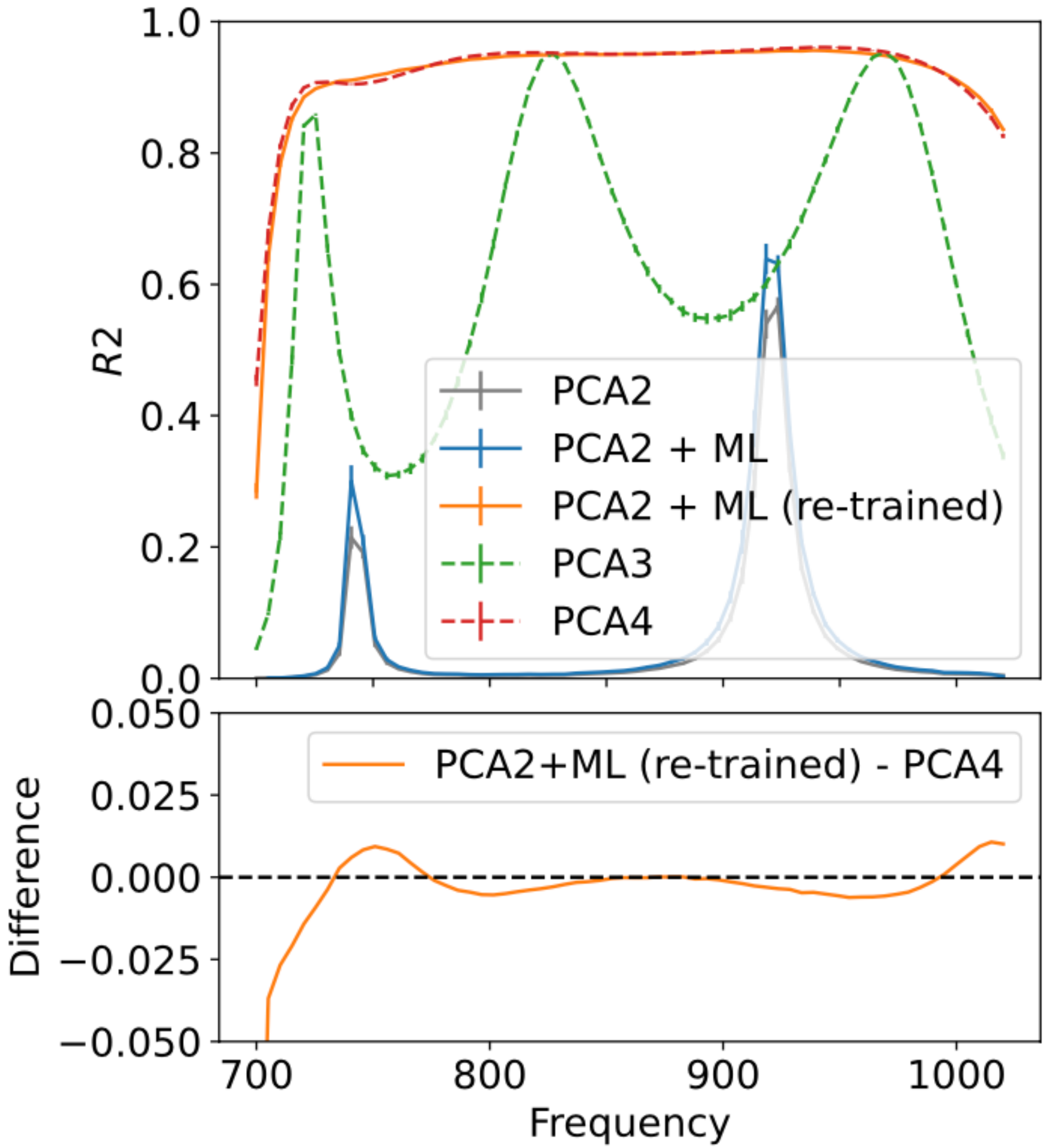
Frequency Beam – Power Spectrum

$$\theta_{\mathrm{B}}(z_i) = \theta_{\mathrm{FWHM}}(\nu_{\mathrm{mid}})\frac{\nu_{\mathrm{mid}}}{\nu_i}$$



- ML doesn't handle surprise
- Beam info during re-training is critical
- Consistent with fixed-beam after re-training (residual ~ 10% signal)
- Comparable to PCA4 alone

Frequency Beam – R^2 score



Coefficient of determination

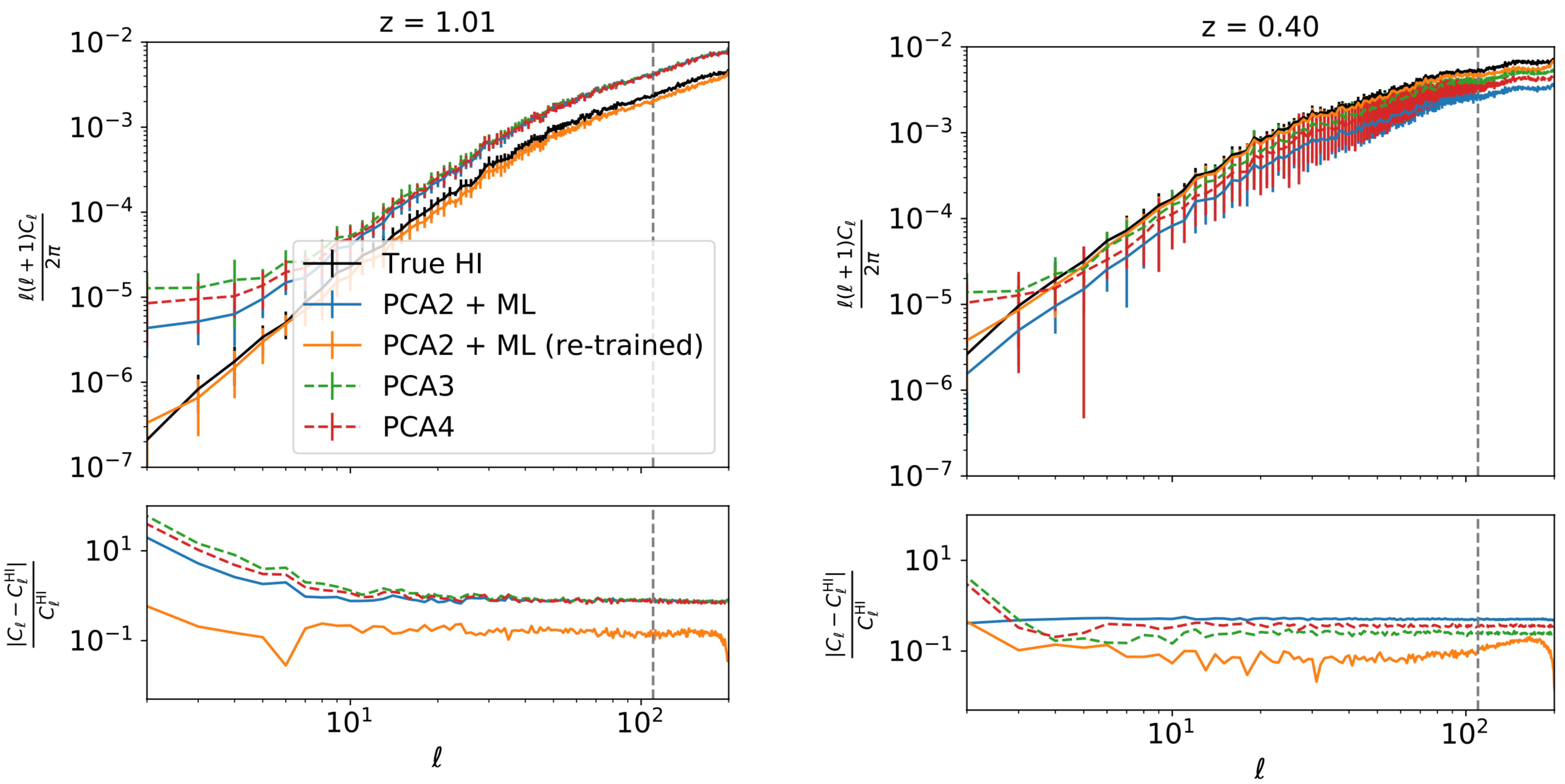
Evaluate the performance of the ML model
Accuracy measurement of predictions v.s. target

$$R^2 = 1 - \frac{\sum_i (t_i - o_i)^2}{\sum_i (t_i - \bar{t})^2}$$

- ML doesn't handle surprise
- Beam info during re-training is critical
- Comparable to PCA4 alone

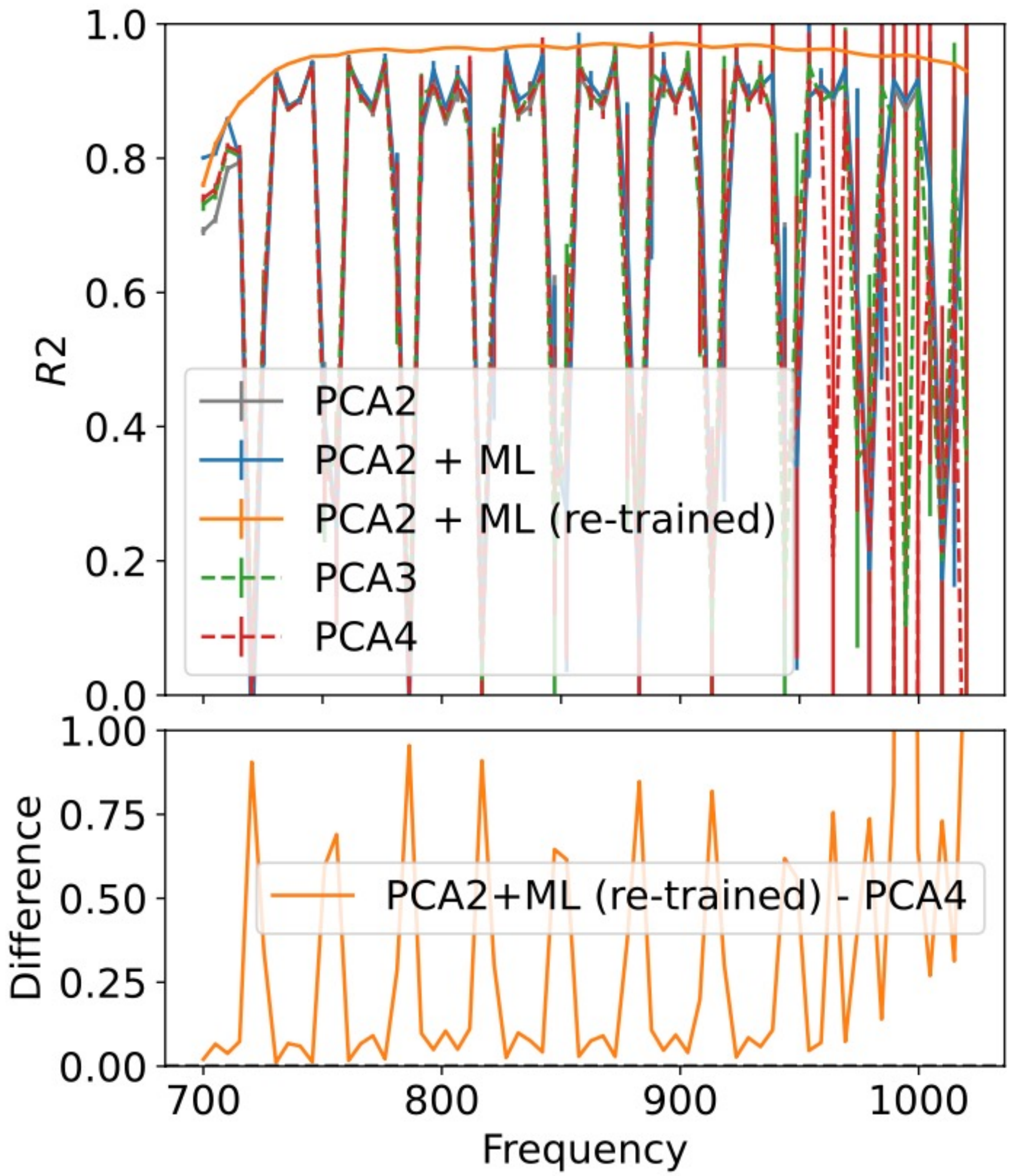
Gain drift – Power Spectrum

$$\Delta G(\nu) = G_0 \sin(G_1 \nu + G_2) + 1$$



- ML doesn't handle surprise
- Gain info during re-training is critical
- Consistent with unit gain after re-training (residual ~ 10% signal)
- Advantage over PCA alone

Gain drift – R^2 score



$$\Delta G(\nu) = G_0 \sin(G_1 \nu + G_2) + 1$$

- Beam info during re-training is critical
- PCA alone shows sinusoidal gain pattern

Conclusions

- ML has consistent performance under different simulations
- ML returns comparable results with traditional methods
- ML requires knowledge of the data – blind usage doesn’t work
- ML can’t handle well systematics without prior knowledge
- Prior systematics knowledge significantly improves ML performance
- In real data:
  - ML provides complementary method for 21cm foreground removal
  - One should estimate the potential systematics before applying ML
- Limitations:
  - Depends on pre-process
  - Lack of redshift information

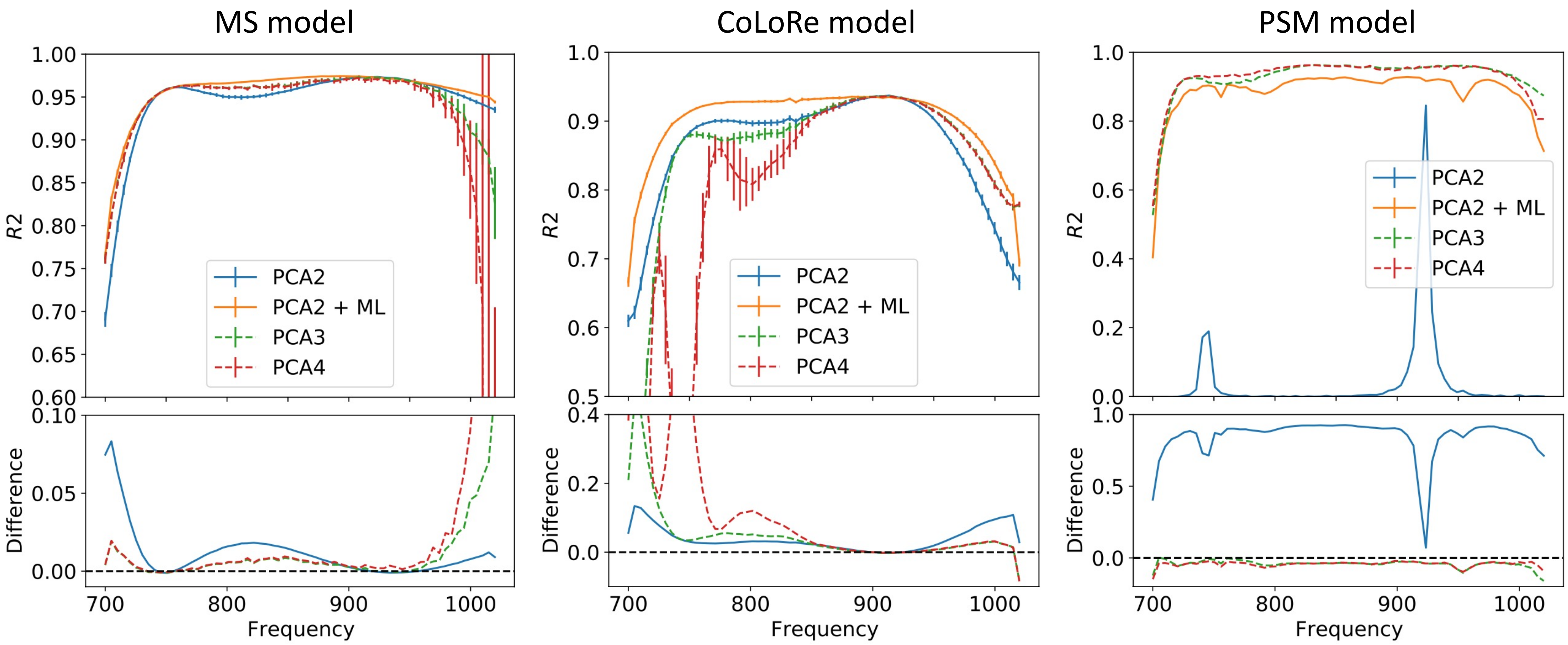
R^2 score comparison

Coefficient of determination

Evaluate the performance of the ML model

Accuracy measurement of predictions v.s. target

$$R^2 = 1 - \frac{\sum_i (t_i - o_i)^2}{\sum_i (t_i - \bar{t})^2}$$



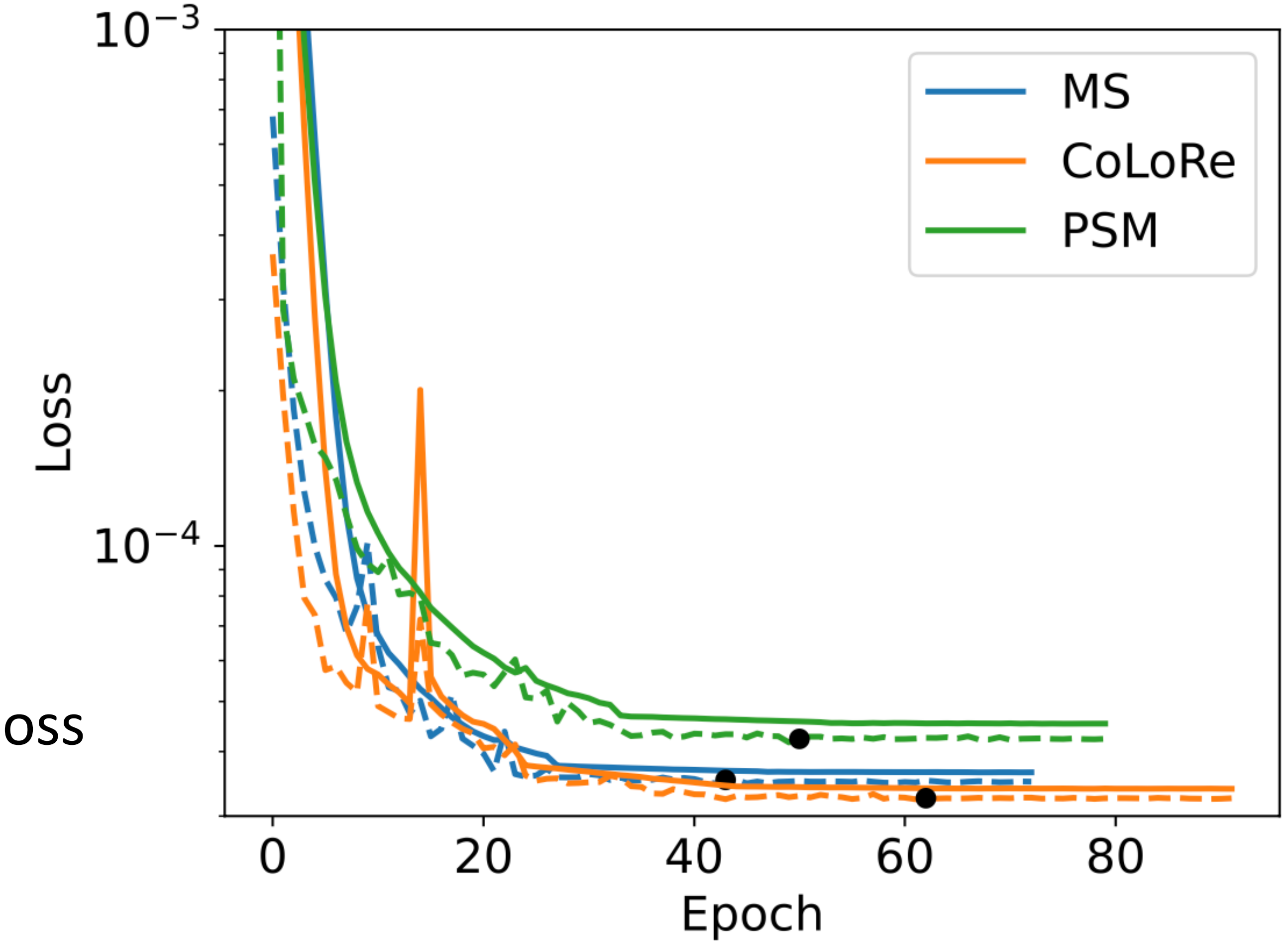
Loss function

Minimise reconstruction errors (loss)

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

Train loss: how model fits the training data
Validation loss: how model performs

- Training: 40 healpix maps (7680 samples)
- Validation: 10 healpix maps (1920 samples)
- Test: 10 healpix maps (1920 samples)



- Comparable under different models
- PSM: more complicated feature -> higher loss