The robustness of machine learning for 21cm foreground removal

ArXiv: 2311.00493

Tianyue Chen

Postdoctoral Researcher

Ecole Polytechnique Fédérale de Lausanne (EPFL)

Cosmology in the Alps, Diablerets, Mar 2024

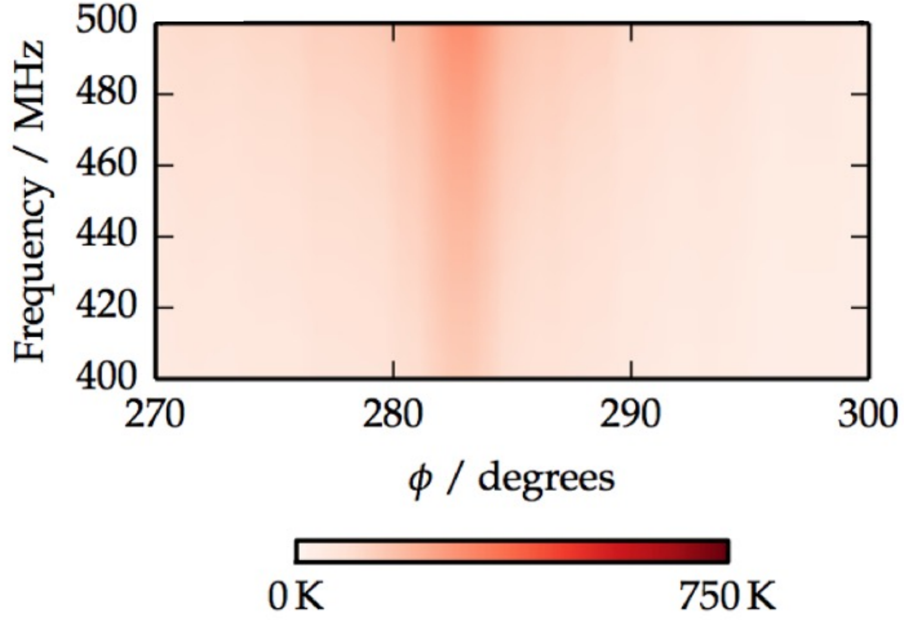
Introduction

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

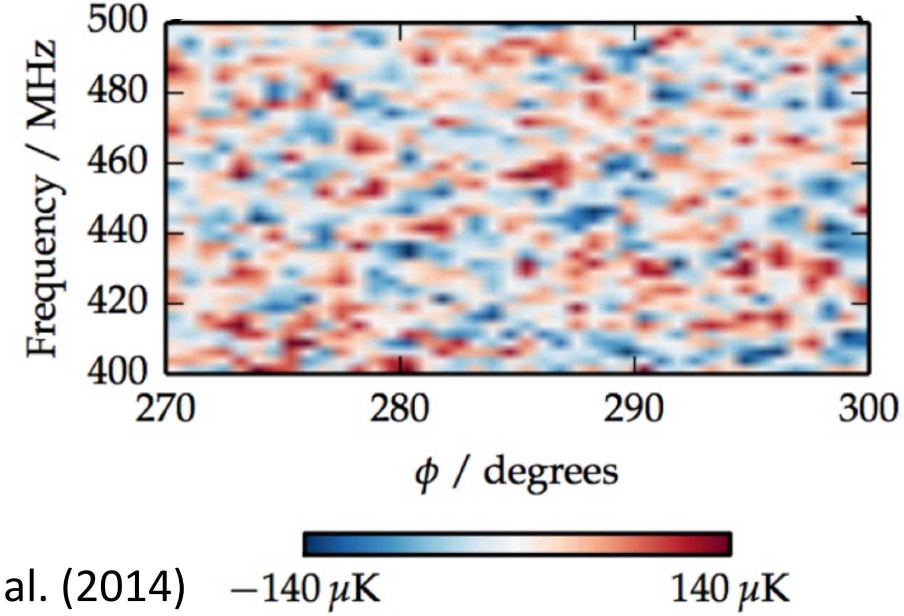
- Traditional approach:
  - Sensitive to systematics (e.g., KL filter)
  - Signal loss (e.g., PCA)

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

Synchrotron



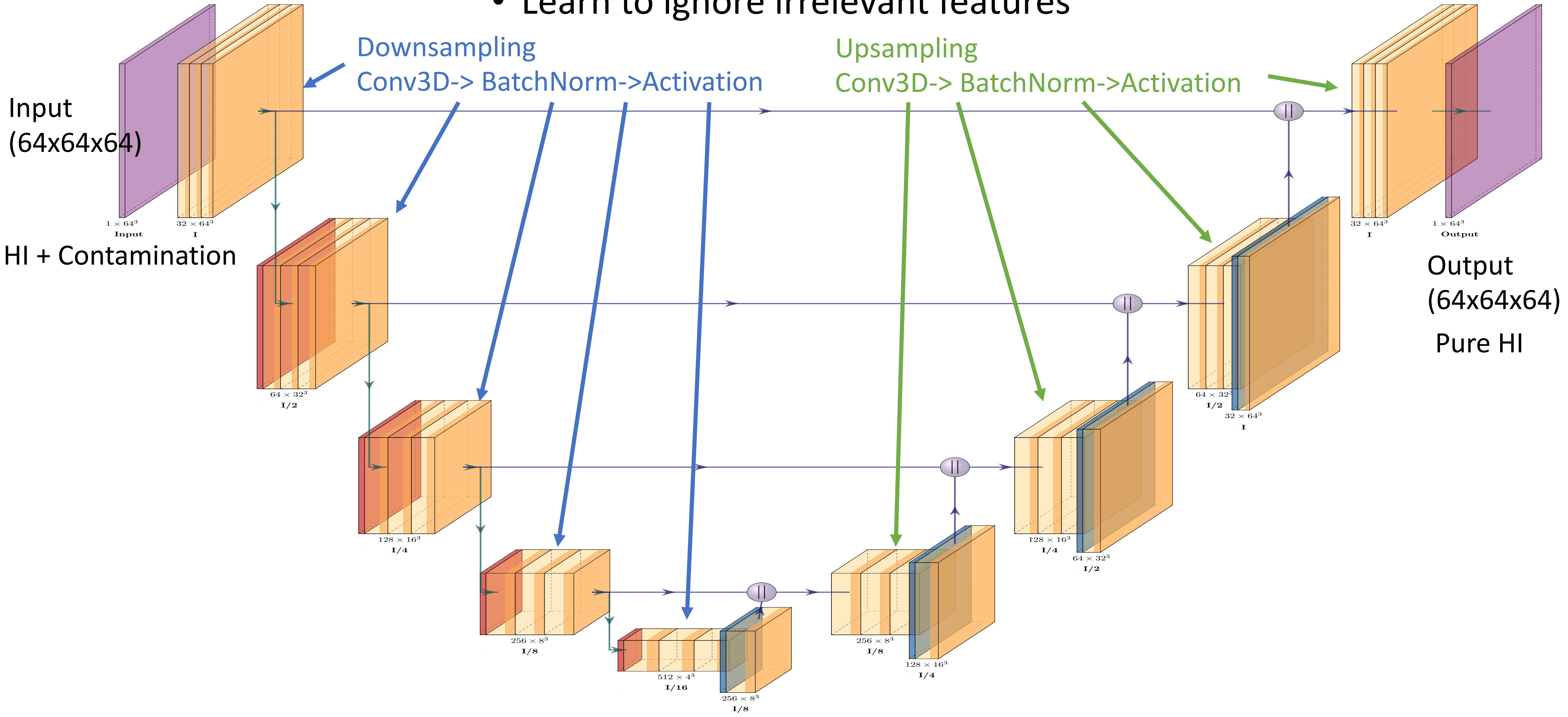
HI



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:

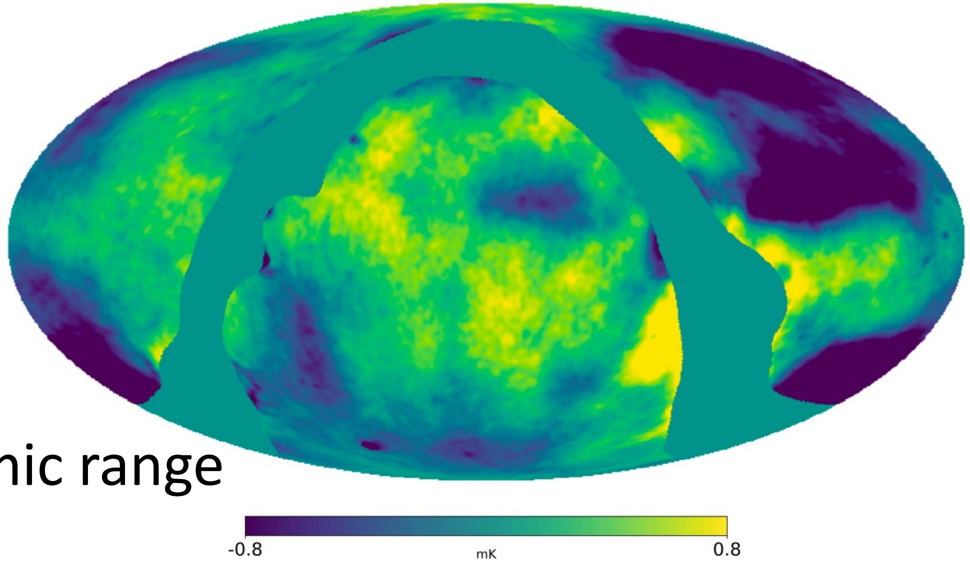
SKA-Mid Band I

Dish diameter, D (m)	15
Frequency range, $\Delta\nu$ (MHz)	[700, 1020]
No. channels, $N_{\rm bin}$	64
Channel width, $\delta\nu$ (MHz)	5
Redshift range, $[z_{\rm min}, z_{\rm max}]$	[0.4, 1.0]
Beam resolution, $[\theta_{\rm min}, \theta_{\rm max}]$ (deg)	[1.1, 1.6]

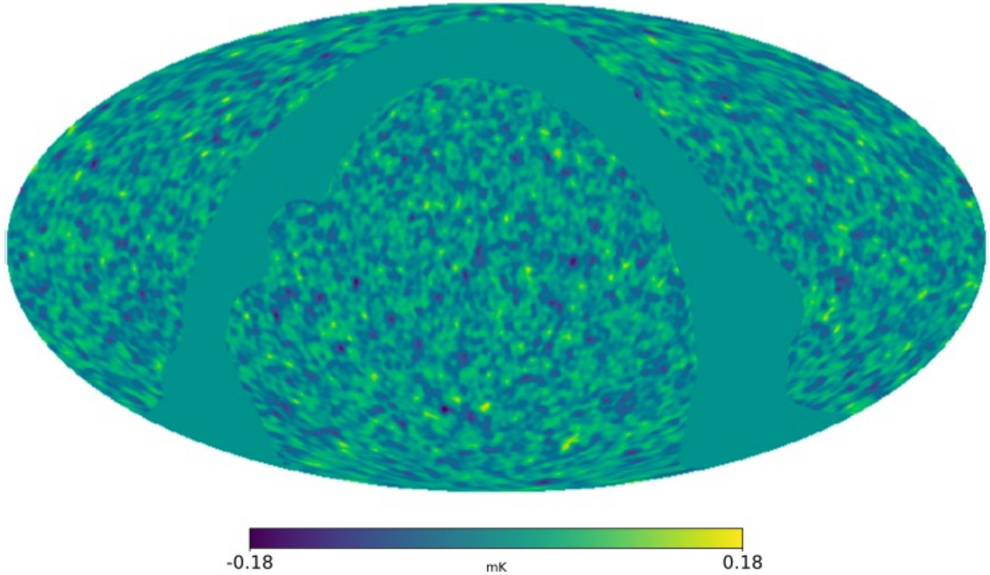
- Instrumental systematics:
  - Frequency-dependent beam $\theta_{\rm B}(z_i) = \theta_{\rm FWHM}(\nu_{\rm mid})\frac{\nu_{\rm mid}}{\nu_i}$
  - Bandpass fluctuations G_υ = 1+ΔG_υ $\Delta\ G(\nu) = G_0 \sin(G_1\nu + G_2) + 1$
- Format:
  - Healpix full sky maps →192 equal-size patches (64x64x64)
  - Training: 40 healpix maps (7680 samples)
  - Validation: 10 healpix maps (1920 samples)
  - Test: 10 healpix maps (1920 samples)

Pre-processing

PCA 2



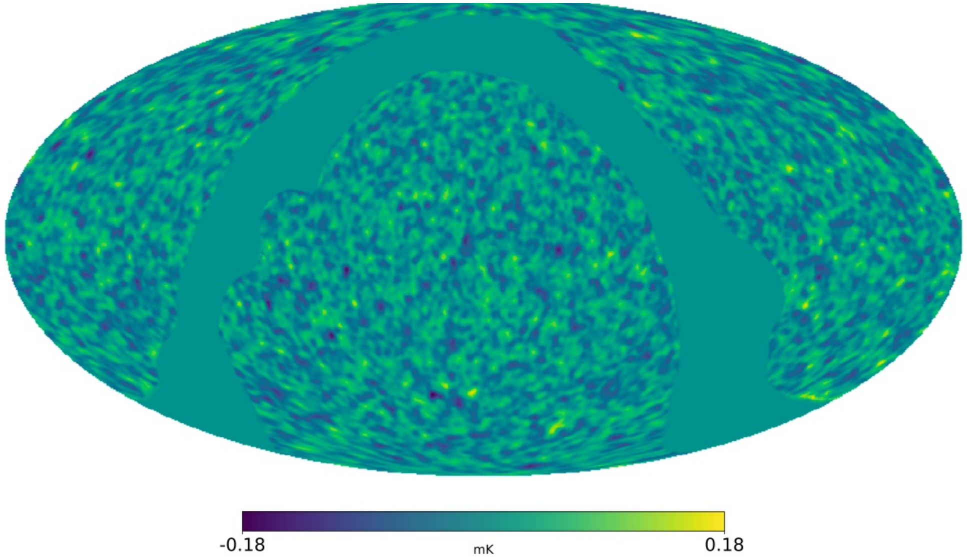
Target HI



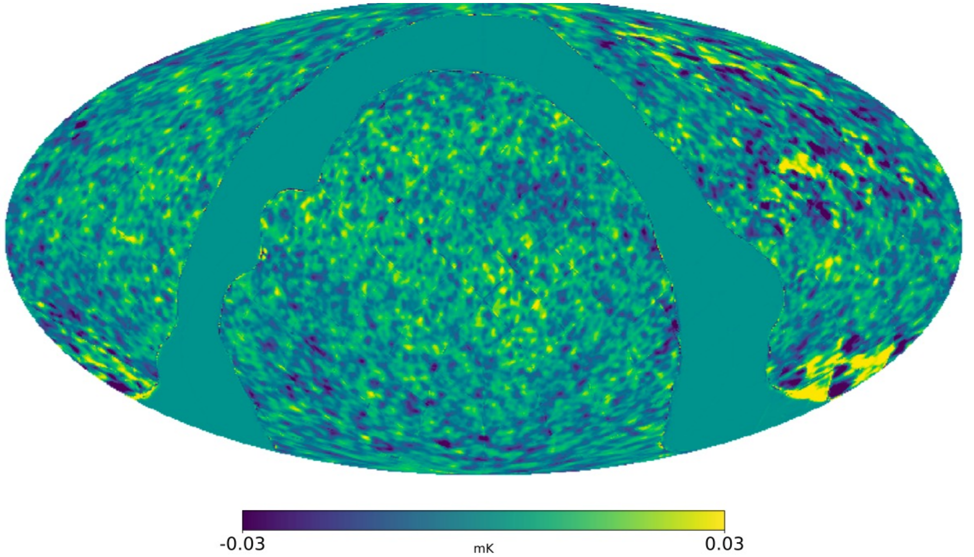
PCA2 to reduce dynamic range

ML for fine tuning

PCA 2 + ML

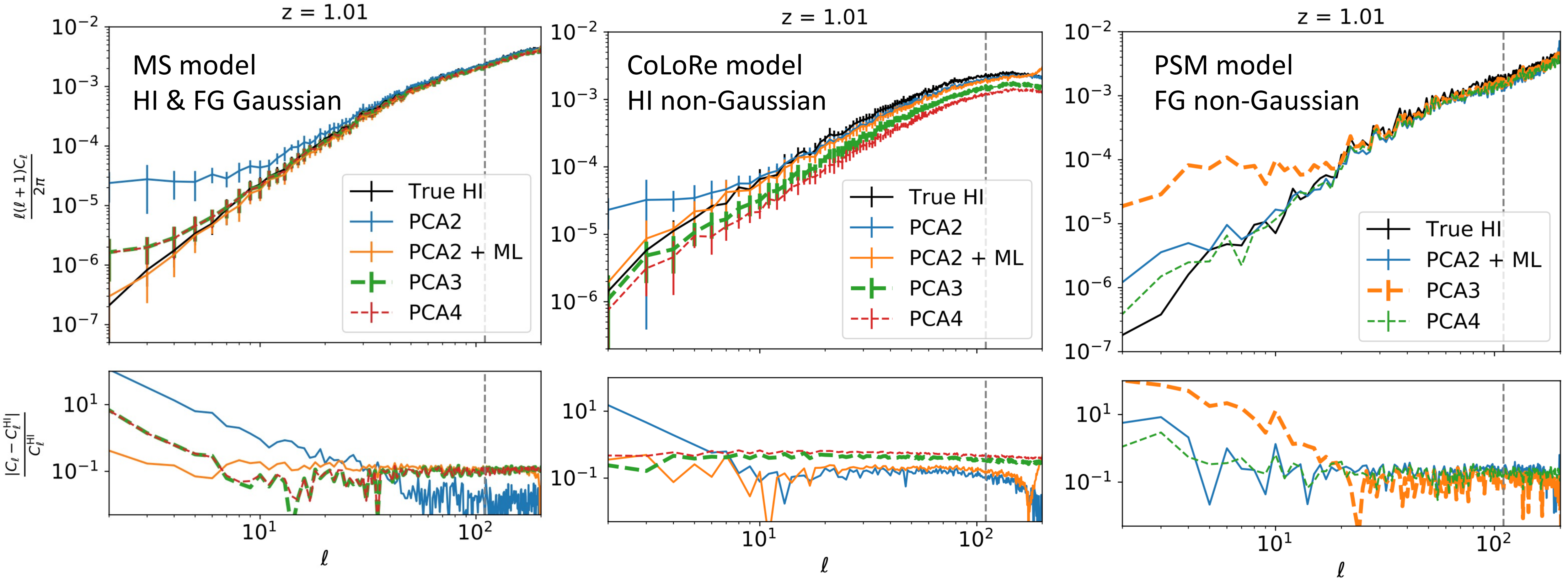


Residual PCA 2 + ML



ML can handle mask!

Sky model dependency – Power Spectrum



Training model ⟷ Testing model

- ML good at reducing large-scale FG residuals
- ML comparable with PCA alone
- Average fractional residual ~ 10% signal over all scales
- Consistent results under different models

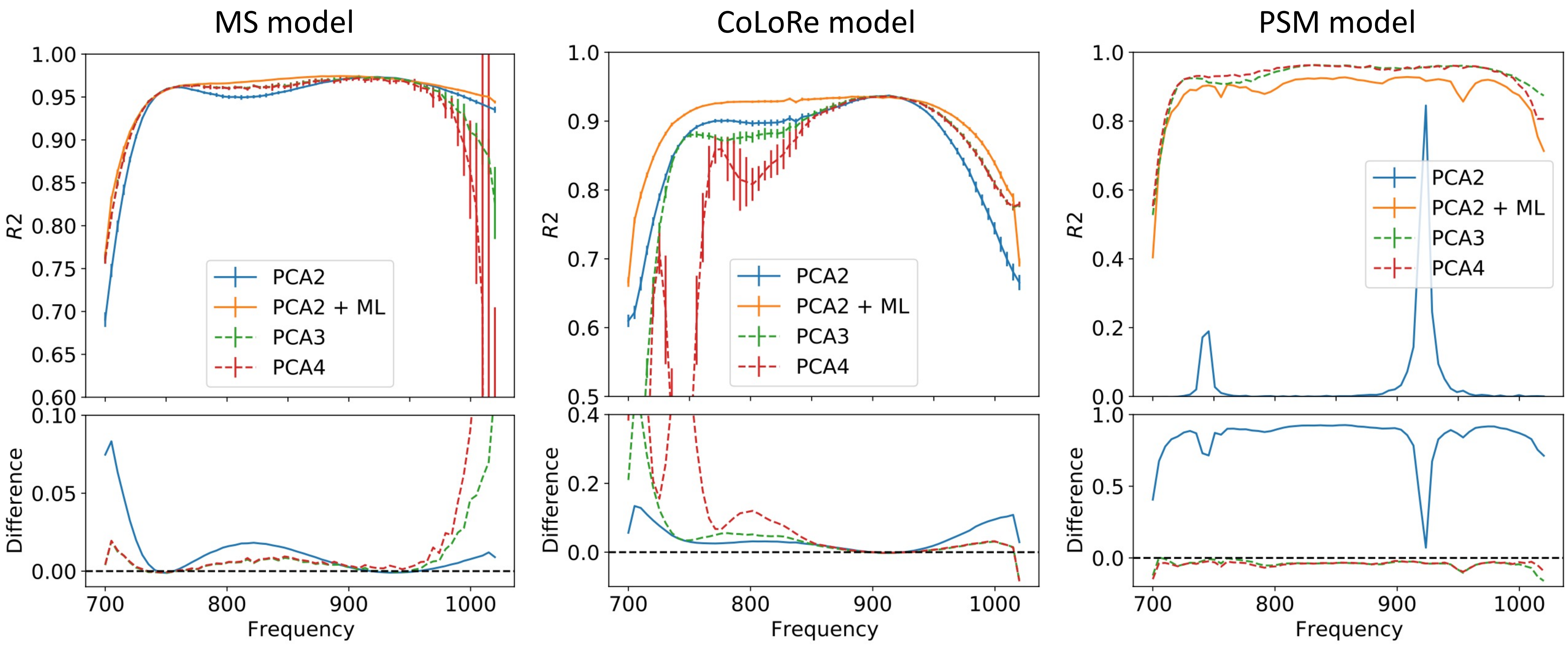
R^2 score comparison

Coefficient of determination

Evaluate the performance of the ML model

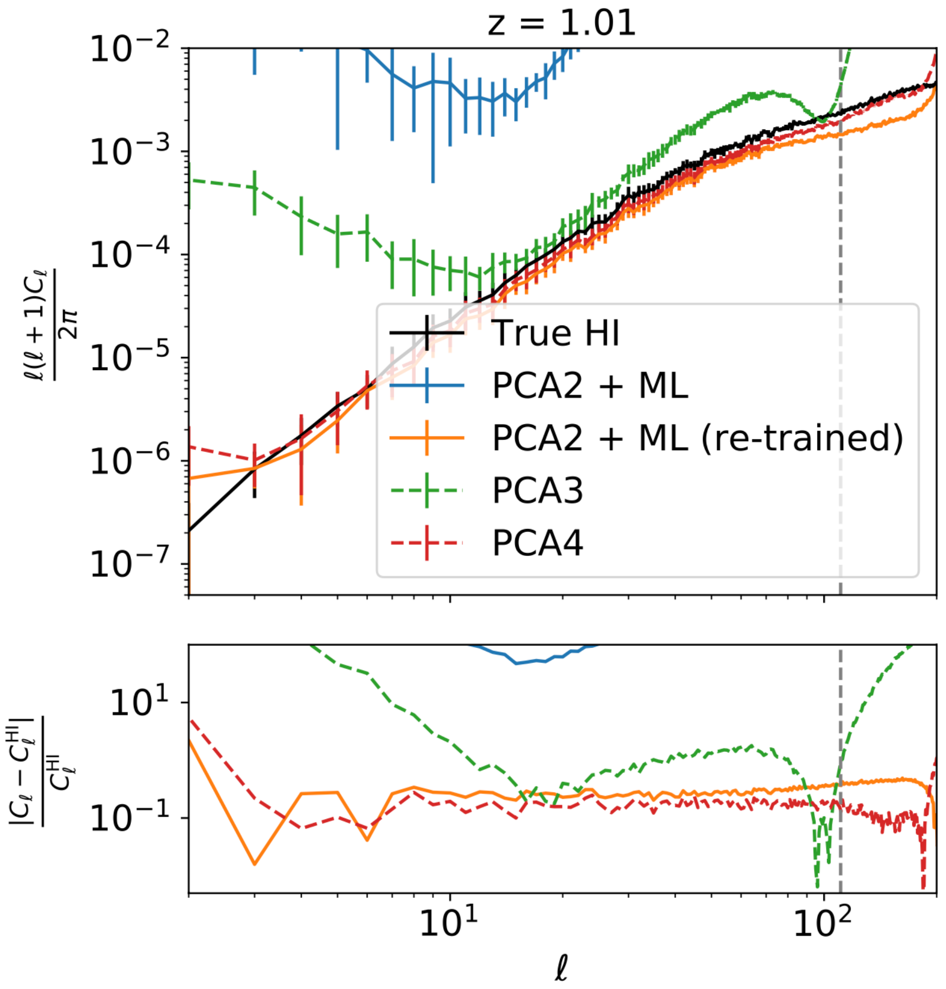
Accuracy measurement of predictions v.s. target

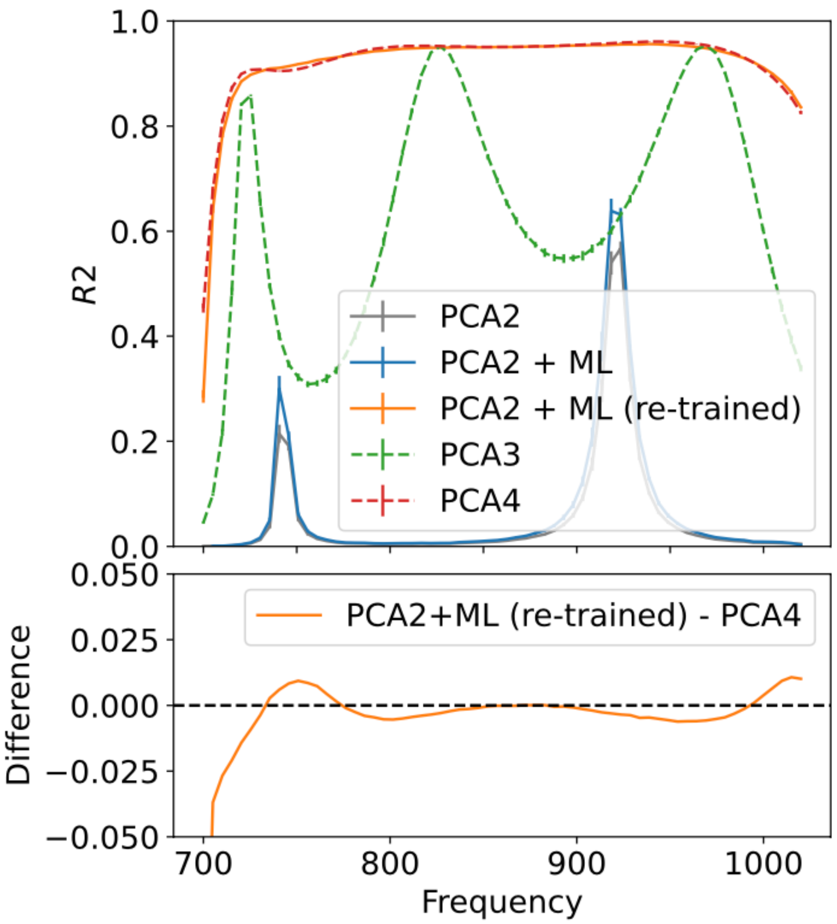
$$R^2 = 1 - \frac{\sum_i (t_i - o_i)^2}{\sum_i (t_i - \bar{t})^2}$$



Frequency Beam

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

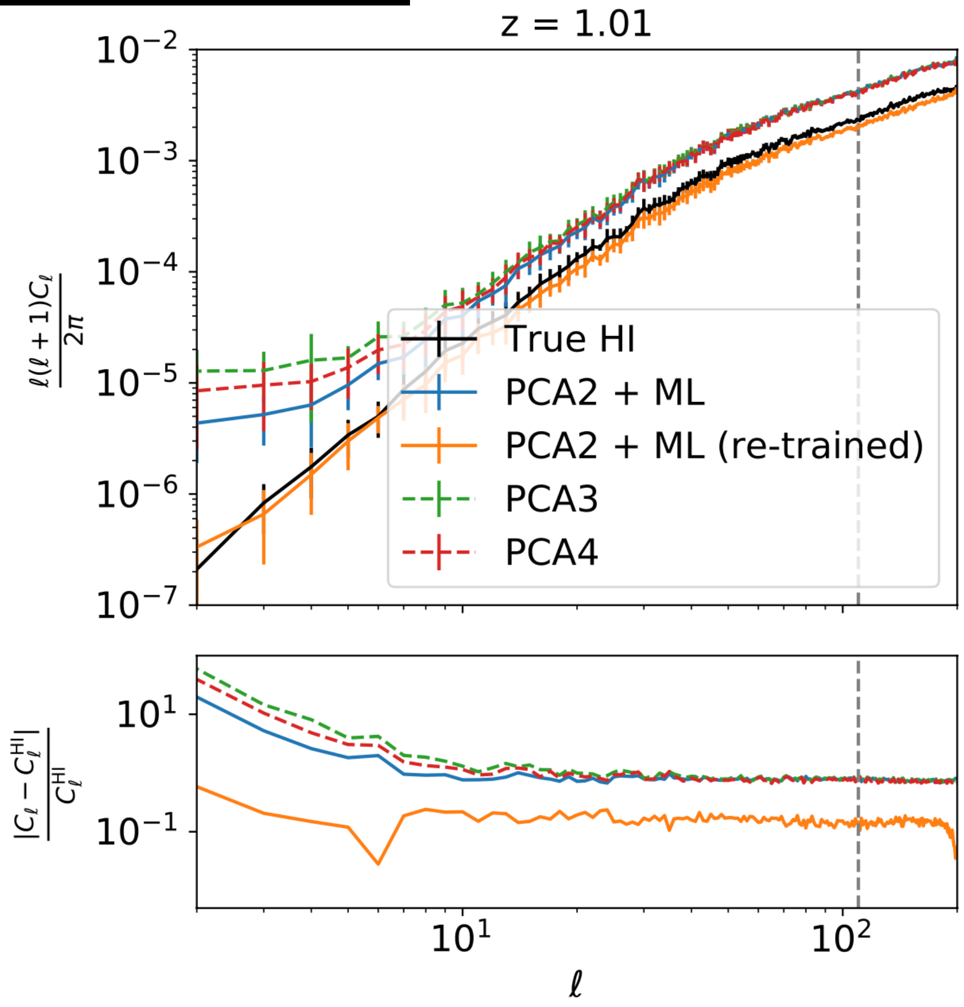


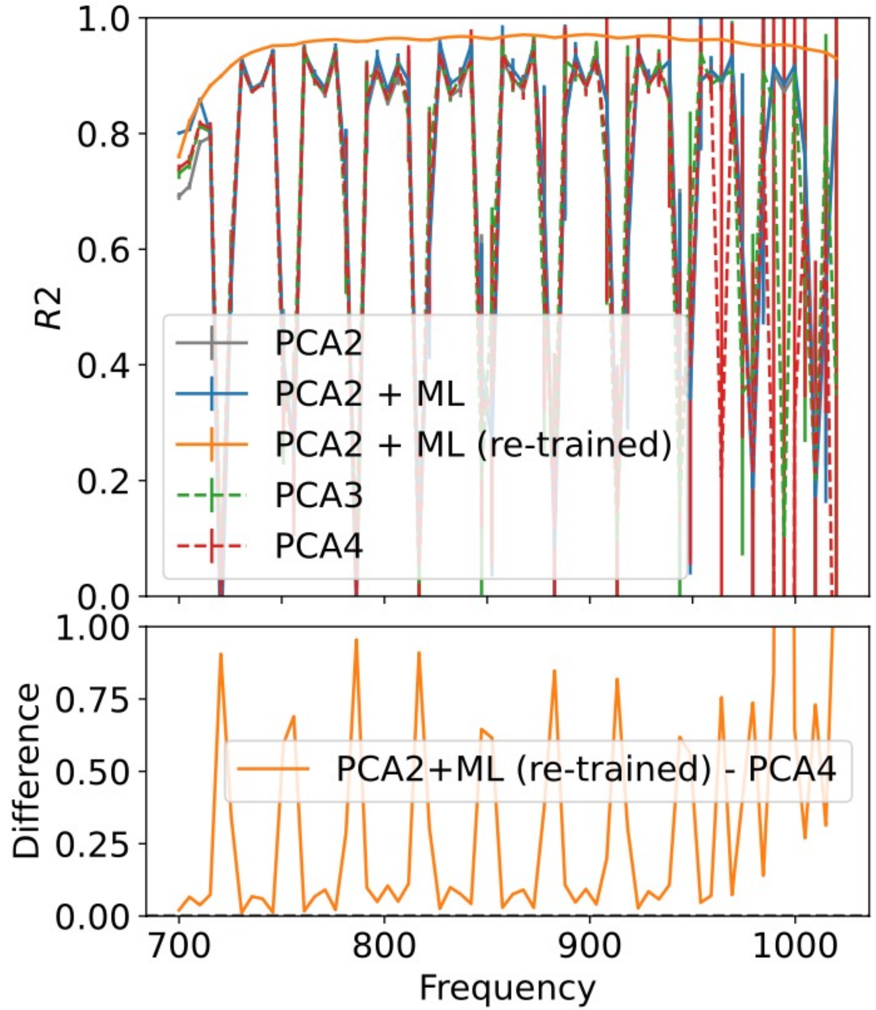


- ML doesn’t handle surprise
- Beam info during re-training is critical
- Residual back to ~ 10% after re-training
- Comparable to PCA4 alone

Bandpass

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





- Re-training is critical
- Residual back to ~ 10% after re-training
- Advantage over PCA alone

Conclusions

ArXiv: 2311.00493

- Is ML reliable?
  - ML has consistent performance under different simulations
  - ML returns comparable results with traditional methods
- What's the correct way to train ML?
  - Training data **consistent** with testing data
  - Requires **prior knowledge** of the data – No blind usage please
  - **Prior systematics** knowledge needed – No surprise please
- In case of real data:
  - ML provides complementary method for 21cm foreground removal
  - One should estimate the potential systematics before applying ML
- Limitations:
  - Pre-process
  - Lack of physics information