MSE per sample on the test set using zero-shot model, fine-tuned and post discarding unreliable predictions

Legend:
- Zero-shot Model
- Fine-tuned model
- After discarded predictions (using ZS)
- After discarded predictions (using FFT)