DENOISING GW STRAIN DATA WITH DEEPCLEAN

Muhammed Saleem,
University of Minnesota

AcceleratingPhysicsWithML@MIT, Jan 28 - Feb 2, 2023.

Collaborators: Chia-Jui Chou, Alec Gunny, Li-Cheng Yang, Michael Coughlin, Dylan Rankin, Erik Katsavounidis, Ryan Magee, Phil Haris, Will Benoit, Deep Chatterjee, Ethan Marx

DENOISING GOALS

- The output reconstructed from an interferometer contains

$$h(t) = s(t) + n(t)$$

Possible GW signal Detector noise

- Objective: To recover $s(t)$ with best possible signal-to-noise ratio by minimising the noise $n(t)$

- Scientific objectives:
  - Signals that are below the noise (un-detectable) becomes detectable
  - Improved SNR improves parameter estimations

REMOVABLE AND NON-REMOVABLE NOISES

$$n(t) = n_{nw}(t) + n_w(t)$$

- Non-removable (fundamental noise)
- Budgeted by system design
- Eg: photon shot noise, thermal noise
- Can be reduced only with upgraded design and technology

- Source of noise witnessed by dedicated system monitors (witness sensors)
- Environmental contamination or technical noise eg: noise arising from the control of suspended optics

WITNESSED NOISE

- witness sensors or channels - the timeseries denoted by $\{w_i(t)\}$
- The noise $n_w(t)$ is the collective contribution from $\{w(t)\}$
- Schematically,

$$n_w(t) = \mathscr{T}\left(\{w_i(t)\}\right)$$

where $\mathscr{T}$ is some activation function representing non-linear or non-stationary coupling of the output of witness channels to the strain channel.

DeepClean: A neural Network to predict $n_w(t)$

$h(t)$

$\{w_i(t)\}$

$$n_w(t) = \mathscr{F}(\{w_i\}; \vec{\theta})$$

$$r(t) = h(t) - n_w(t)$$

$r(t)$

Residual strain
after noise subtraction

- $\vec{\theta} \rightarrow$ trained weights of the neural network

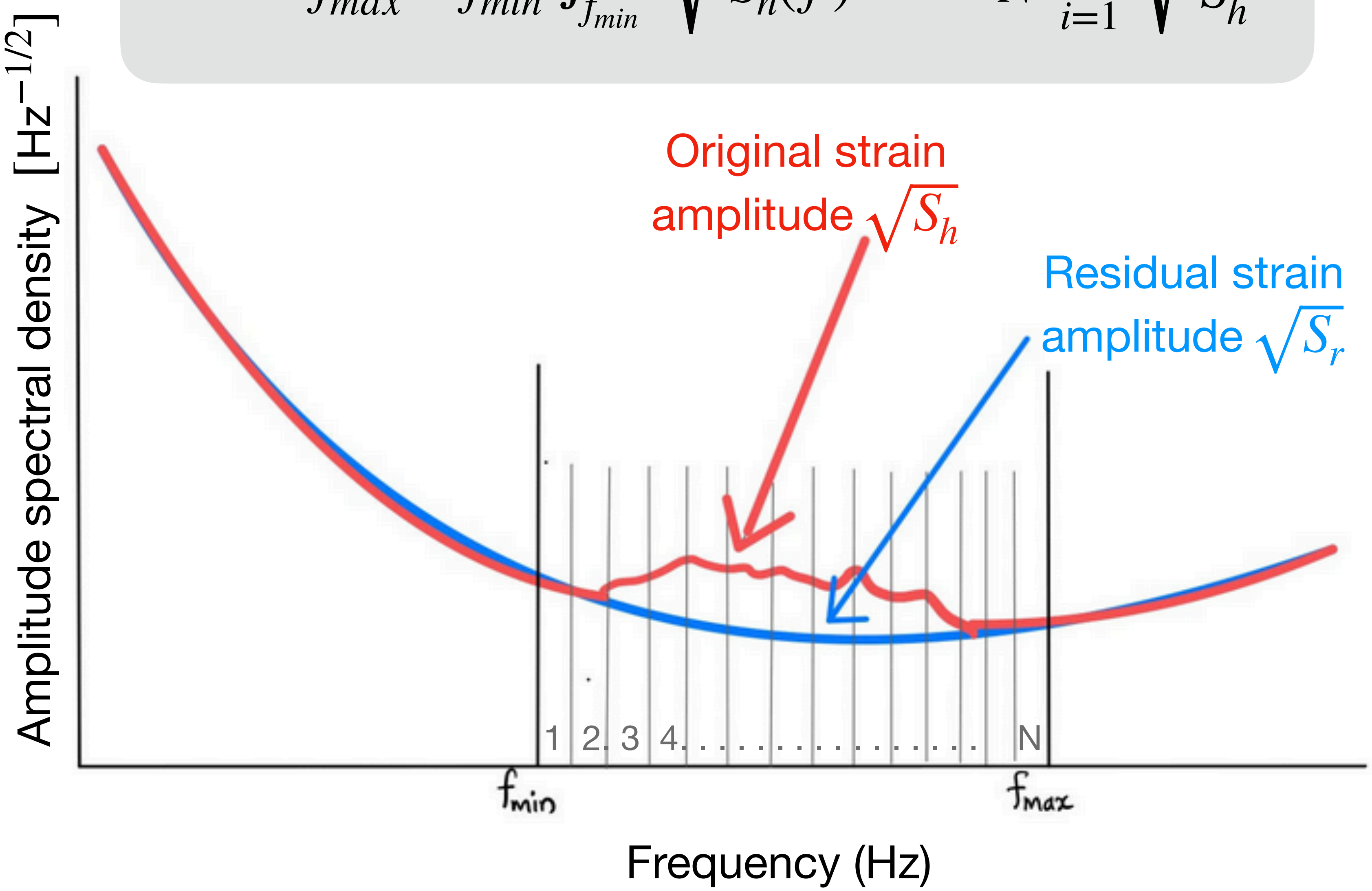
$$\vec{\theta} = \min_{\vec{\theta}'} \left[J \left(h(t), \mathscr{F}(\{w_i(t)\}; \vec{\theta}') \right) \right]$$

where J is some appropriate loss function

Ormiston et al (2020)

Loss function in terms of Amplitude Spectral Density (ASD)

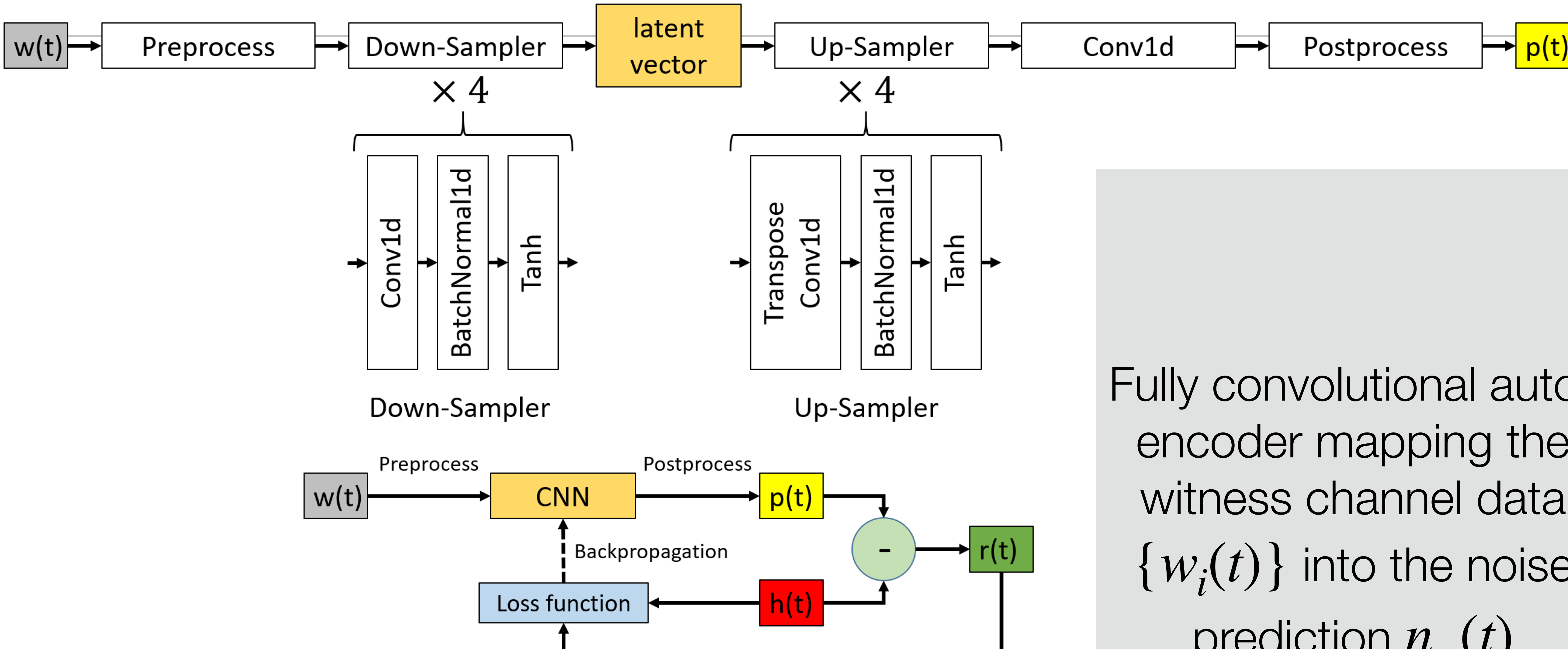
$$J_{asd} = \frac{1}{f_{max} - f_{min}} \int_{f_{min}}^{f_{max}} \sqrt{\frac{S_r(f)}{S_h(f)}}\, df = \frac{1}{N} \sum_{i=1}^{N} \sqrt{\frac{S_r^{(i)}}{S_h^{(i)}}}$$



- Minimises the ASD ratio averaged over all the bins in the desired frequency range.
- Ratio of residual ASD to the original ASD
- Involves FFT → the minimum length of the time-domain data dictated by the resolution of the features we are looking for.

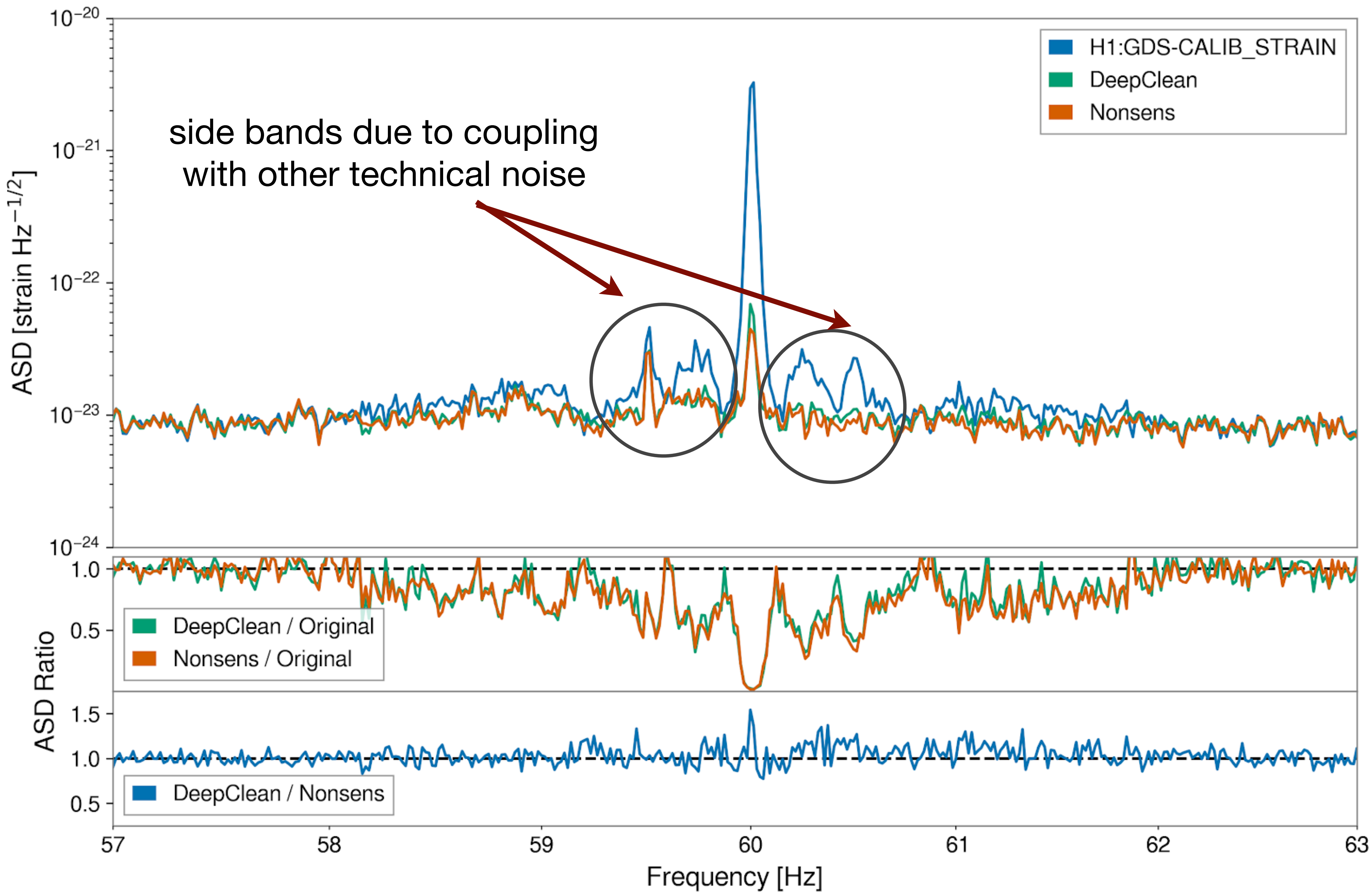
Ormiston et al (2020)

The Architecture and the Workflow



Fully convolutional auto-encoder mapping the witness channel data $\{w_i(t)\}$ into the noise prediction $n_w(t)$

EXAMPLE: SUBTRACTION OF 60 HZ POWER-LINE AND THE SIDEBANDS

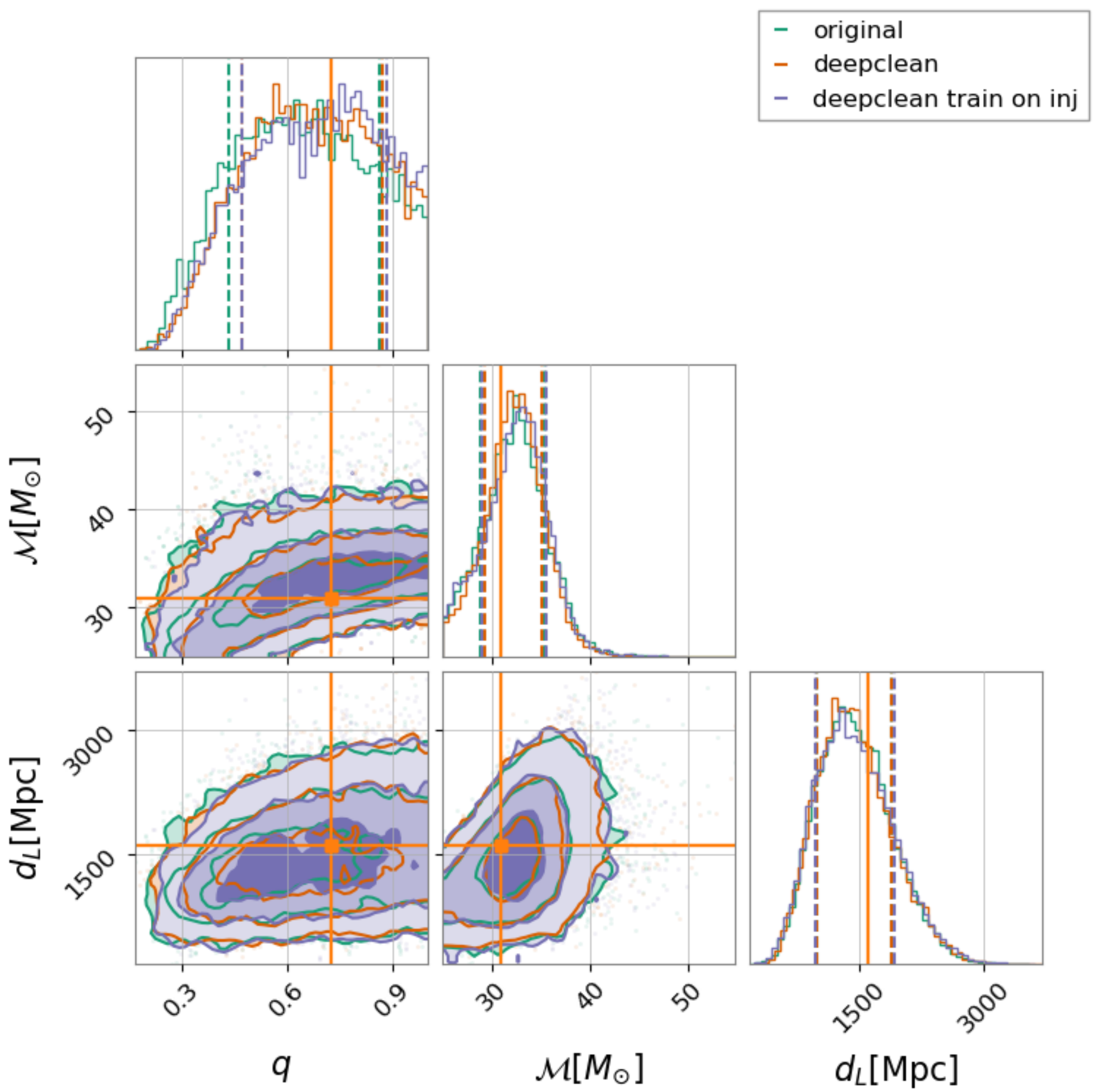


Ormiston et al (2020)

DEEPCLEAN VALIDATION TESTS WITH MOCK DATA

- Twenty days of LIGO data from O3 (between 9/1/2019 - 9/20/2019)
- Injected compact binary coalescence signals (BBH/BNS/NSBH)
    - Drawn from O3-inferred astrophysical population models
- 25000 injections in total, we analyse 266 BBH injections that have coalescence frequency between 55-70 Hz ()
- Following analysis performed on the original and cleaned data
    - Parameter estimation of the injections
    - Sensitive volume (<VT>) estimation
    - Match filter SNR of the injections

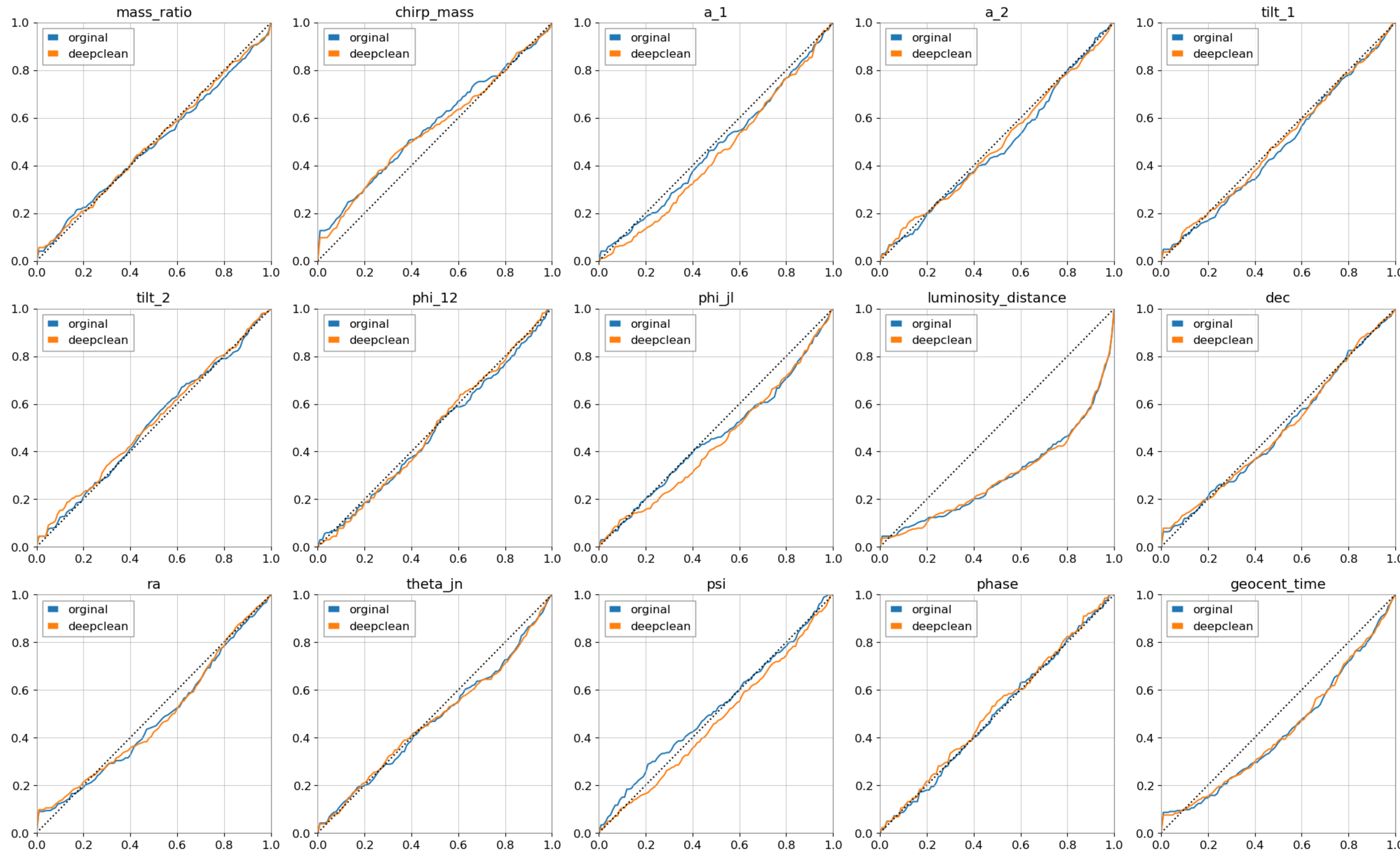
PE VALIDATION: EXAMPLE FROM A SINGLE INJECTION



- With 60 Hz subtraction, PE improvements are not prominent. This is expected as the signal spends a small fraction of time in the frequency band of 60 Hz noise
- Can we train on data that has GW signals in it?
  - Yes, they don't affect the noise prediction.

PE Validation: P-P Plots of all the parameters

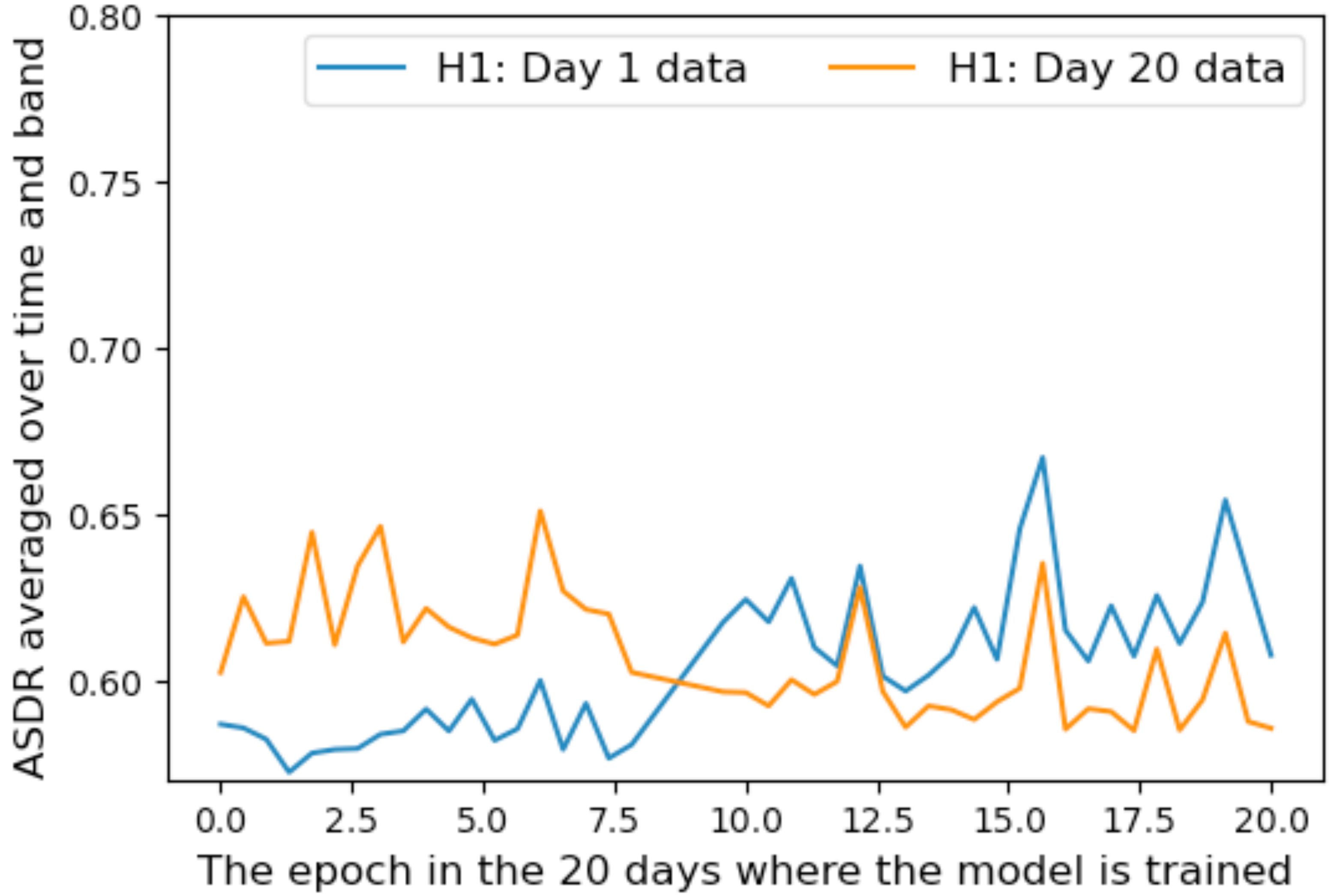
P-P plot shows that deepClean does introduce bias into the PE



DETECTION IMPROVEMENTS

- Results awaited
- GstLAL: Computes the sensitive volume improvements (how many new detections to expect due to denoising)
- Improvements in the matched filter SNR

How often to train DeepClean



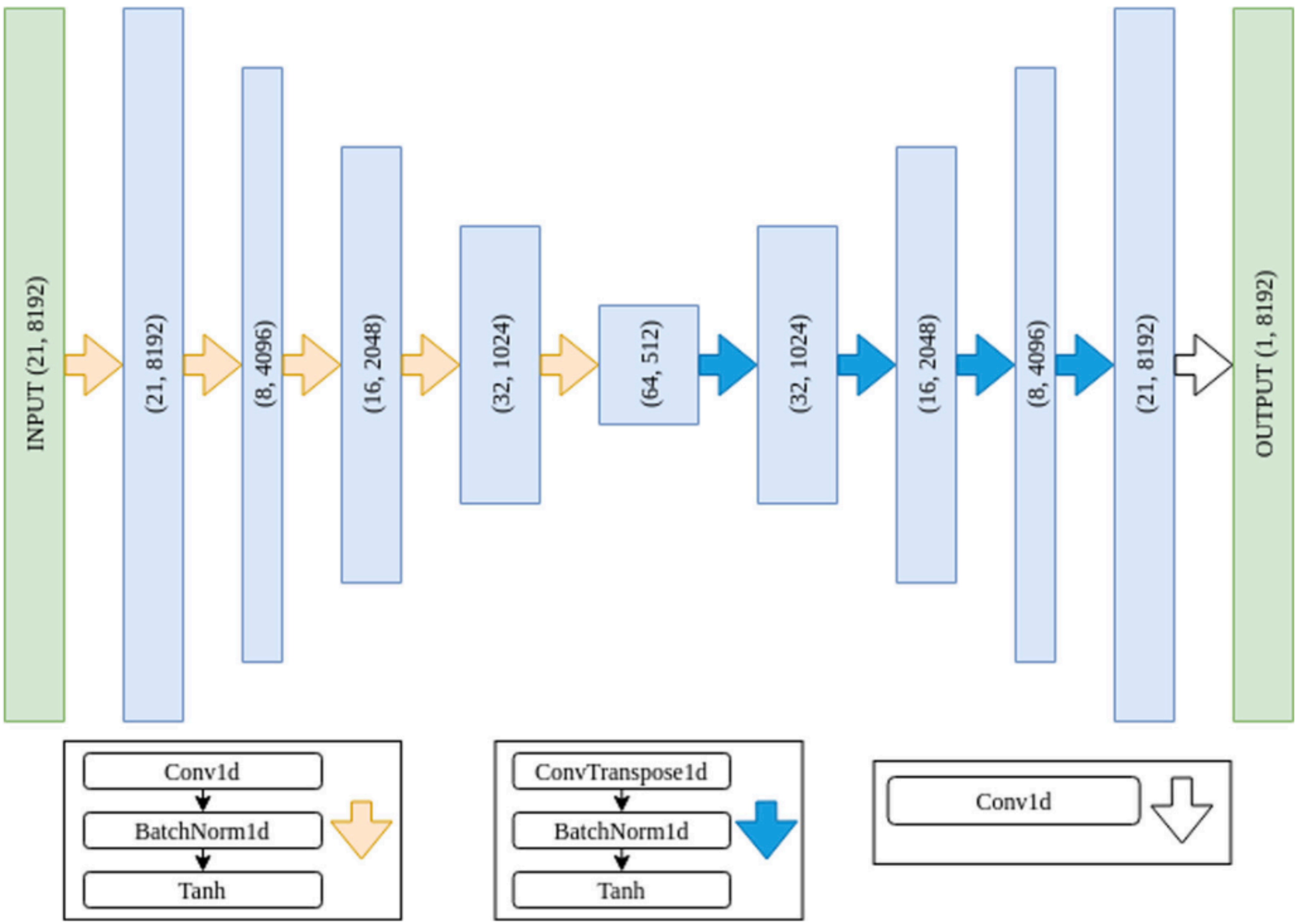
- Once over a few days is found to be enough in O3 data.
- This might change in O4 or for a different noise coupling

Ongoing and future works

- KAGRA implementation (Chia-Jui's talk up next)
- Online DeepClean on LIGO data
  - Feasibility study done (DeepClean can be performed on frames of 1 s duration and the cleaned frames can be produced in ~1 s latency)
  - Deployment with HERMES taking place (lead by Alec Gunny)
- Low frequency (10-30 Hz) broadband noise subtraction
- LIGO internal review to start soon (to be ready for O4)

ADDITIONAL SLIDES

ARCHITECTURE WITH AN EXAMPLE



Fully convolutional auto-encoder mapping the witness channel data $\{w_i(t)\}$ into the noise prediction $n_w(t)$

Input: 8 s time-series from 21 witness channels sampled at 1024 Hz

Output: Noise prediction for 8s sampled at 1024 Hz

Ormiston et al (2020)

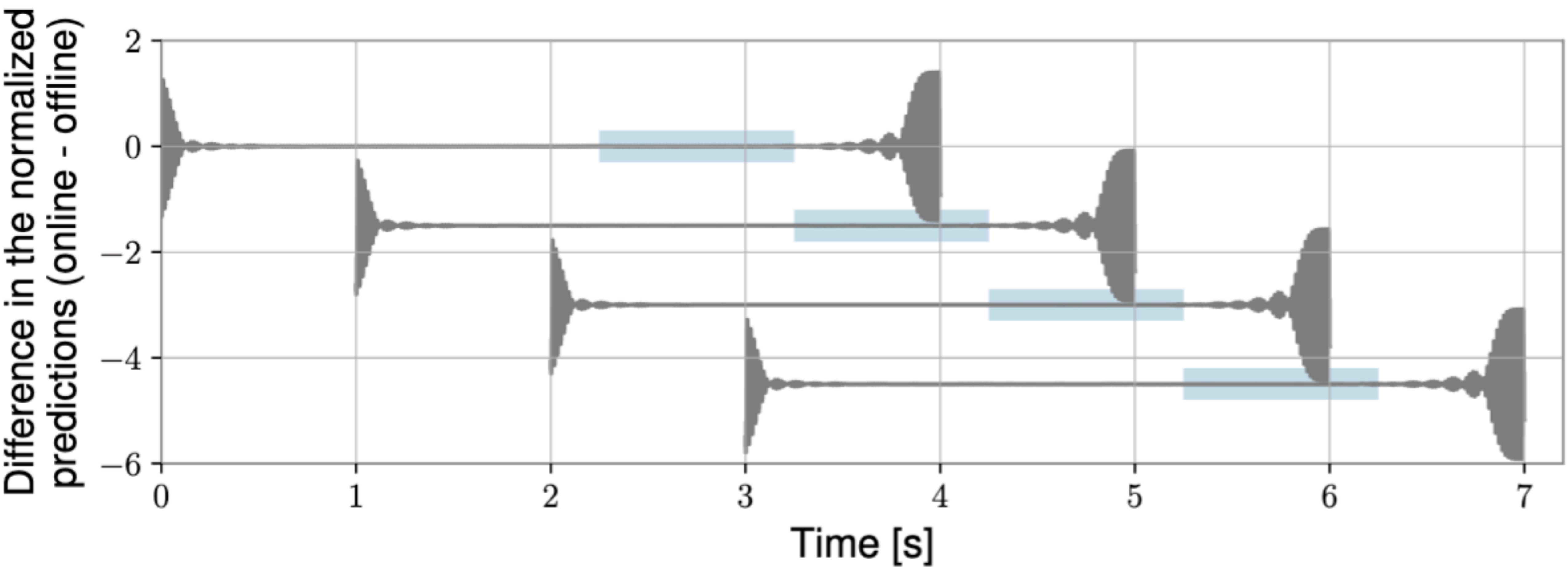
Online DeepClean

Online DeepClean

- Aim is to subtract every 1 s frame as soon as they are available
- 1 s cleaned data are noisy at the edges
  - Reason: a combination of the architecture, training and post-processing (the major part).
- Consider 4 s data for cleaning and take 1 s from somewhere in the middle excluding the edges.
  - Causes additional latency

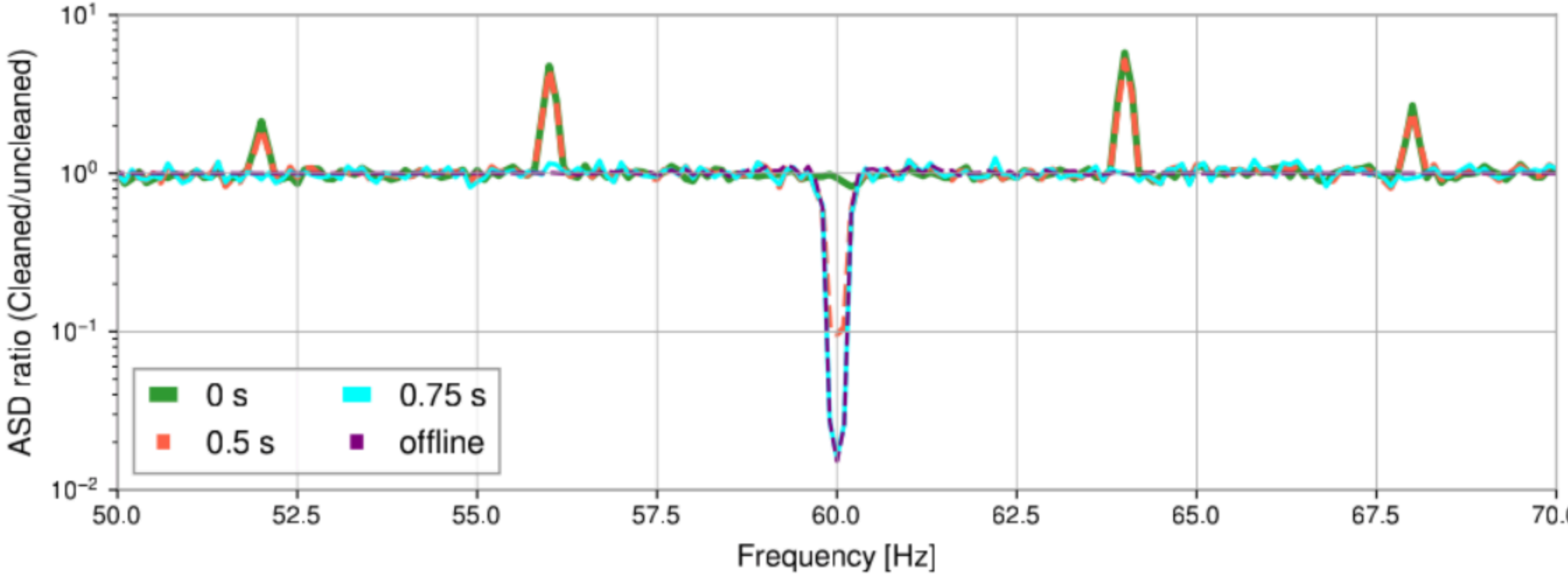
Will be discussed in other talks

Sliding Windows: Suppressing the Edge Effects



- Take unaffected 1s from the 4s cleaned segment and run cleaning on overlapping windows

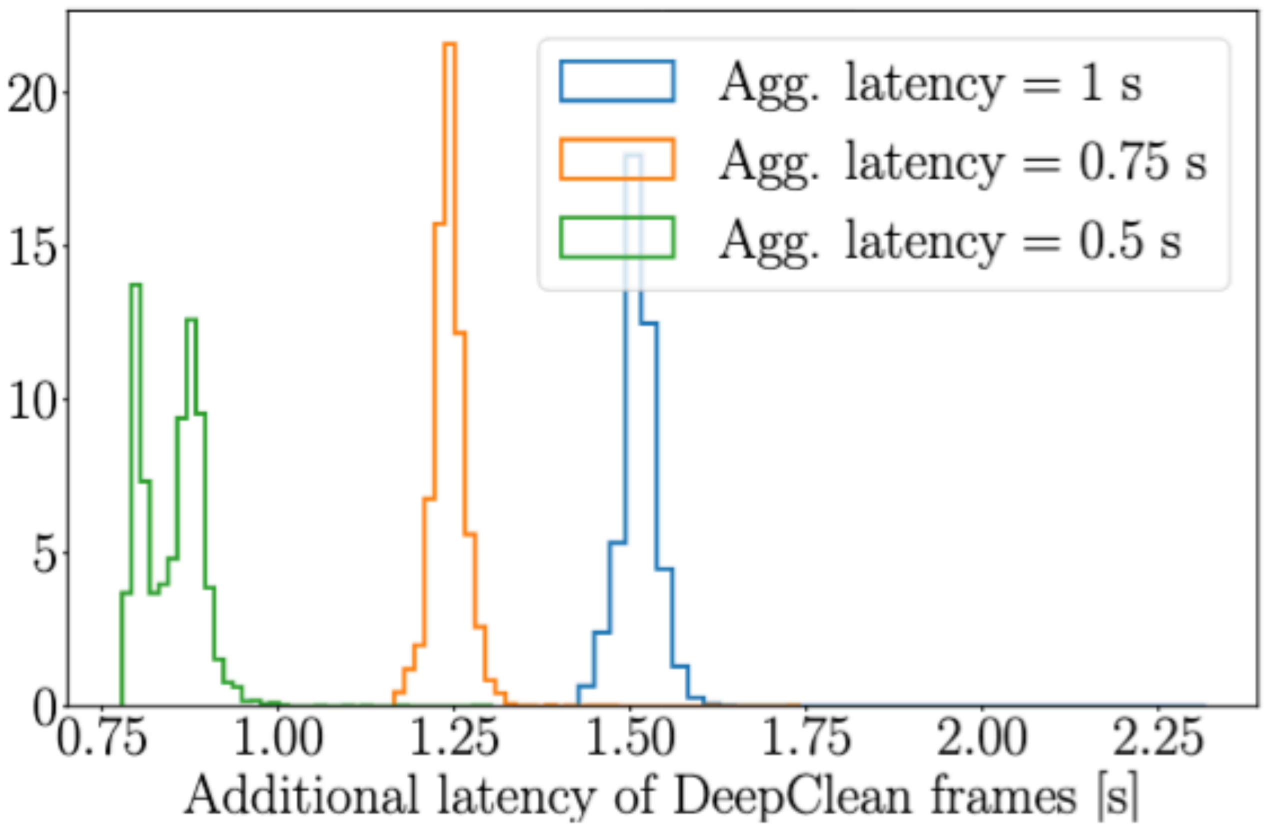
Online DeepClean: Latency vs Subtraction Quality

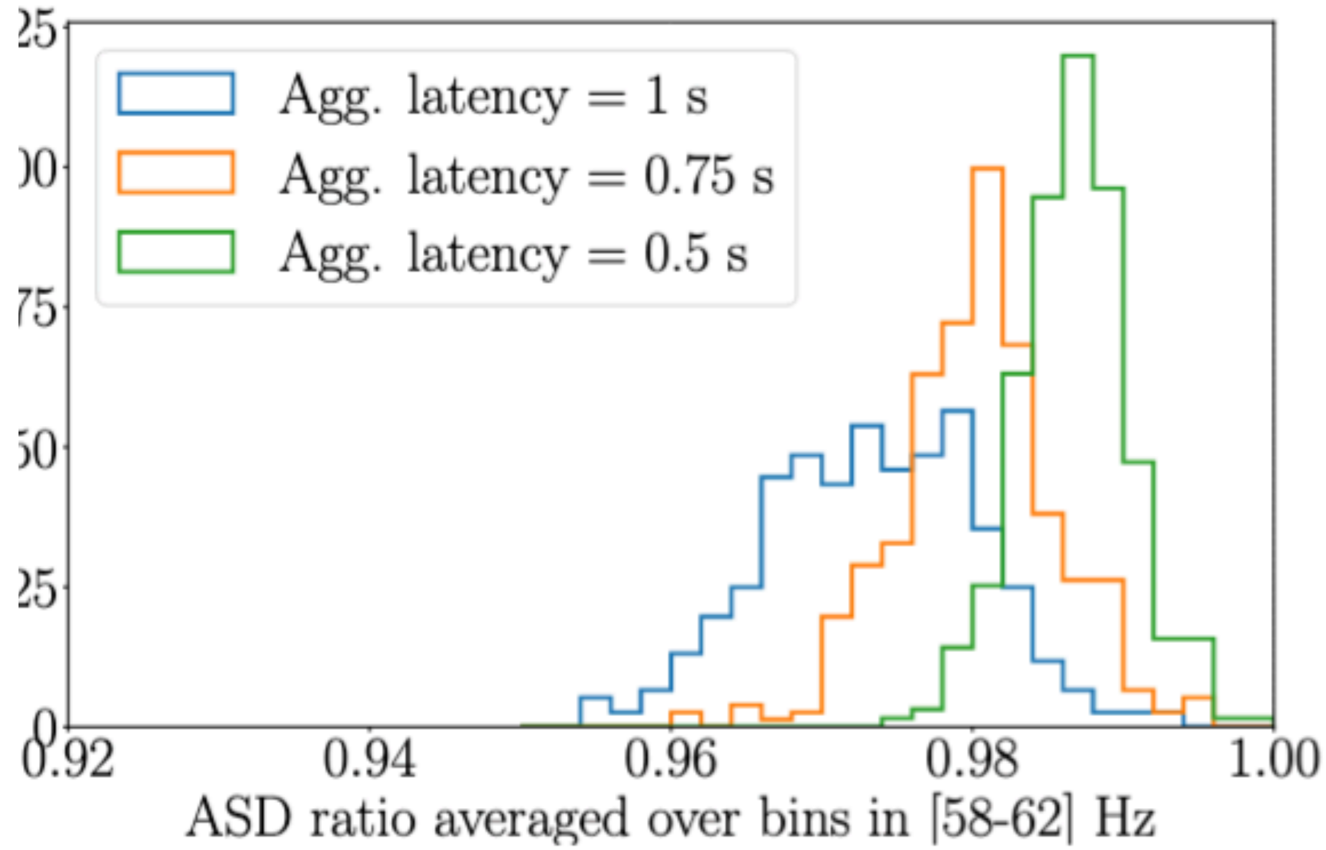


- Longer the excluded edge, better the quality of the subtraction

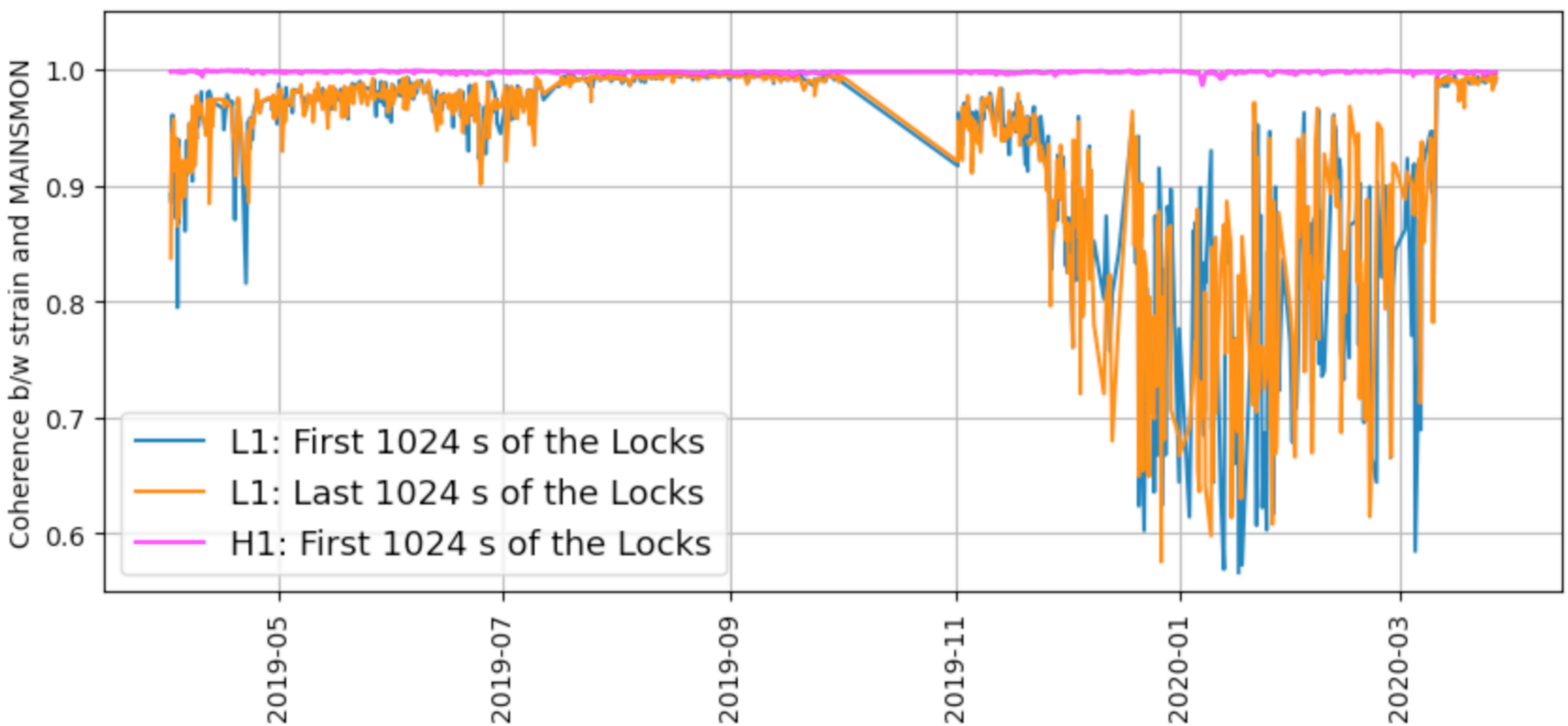
Gunny et al (2022)

ONLINE DEEPCLEAN: LATENCY VS IMPROVEMENTIN ASD RATIO

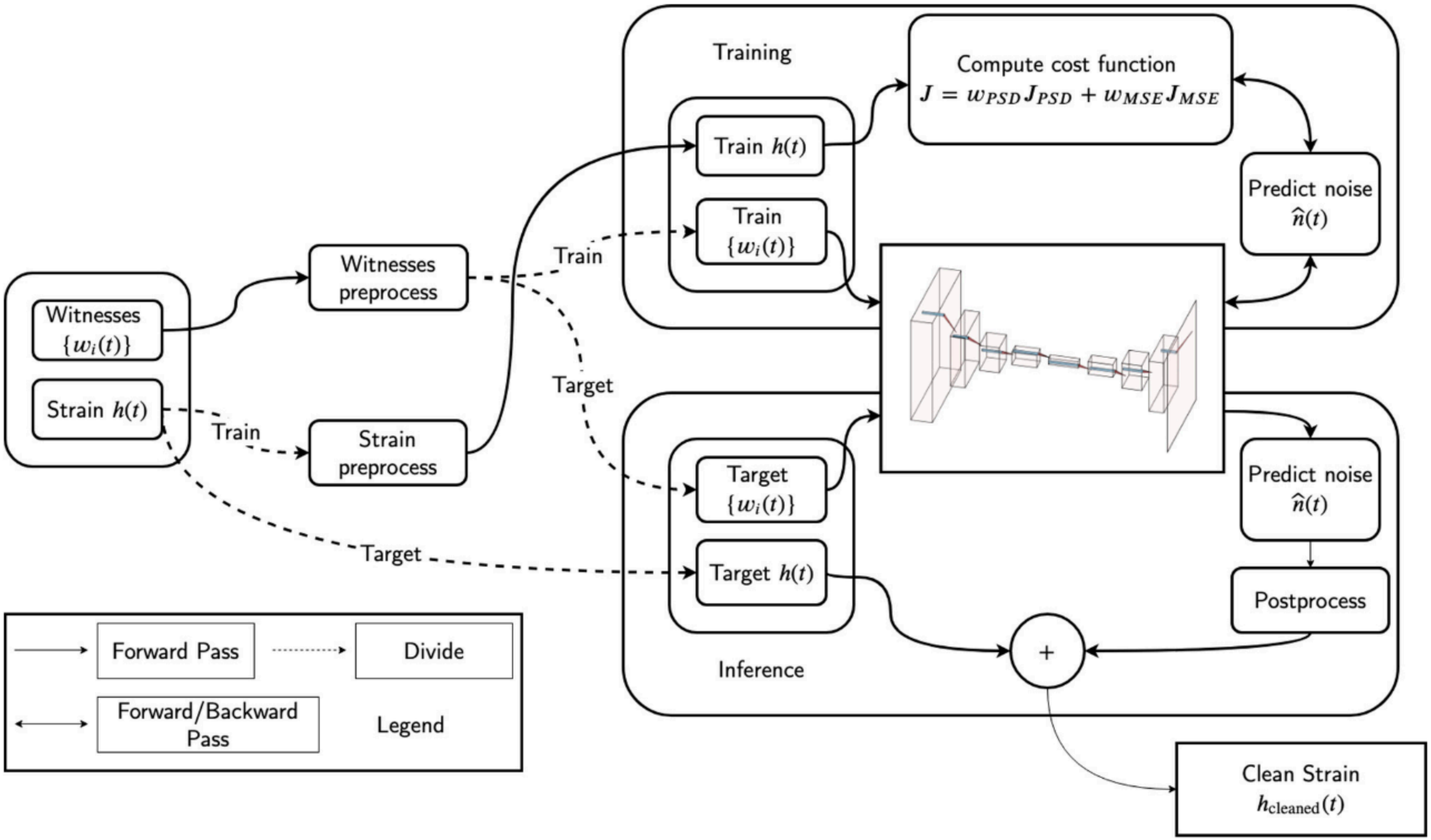




COHERENCE BETWEEN THE STRAIN AND POWER-LINE WITNESS

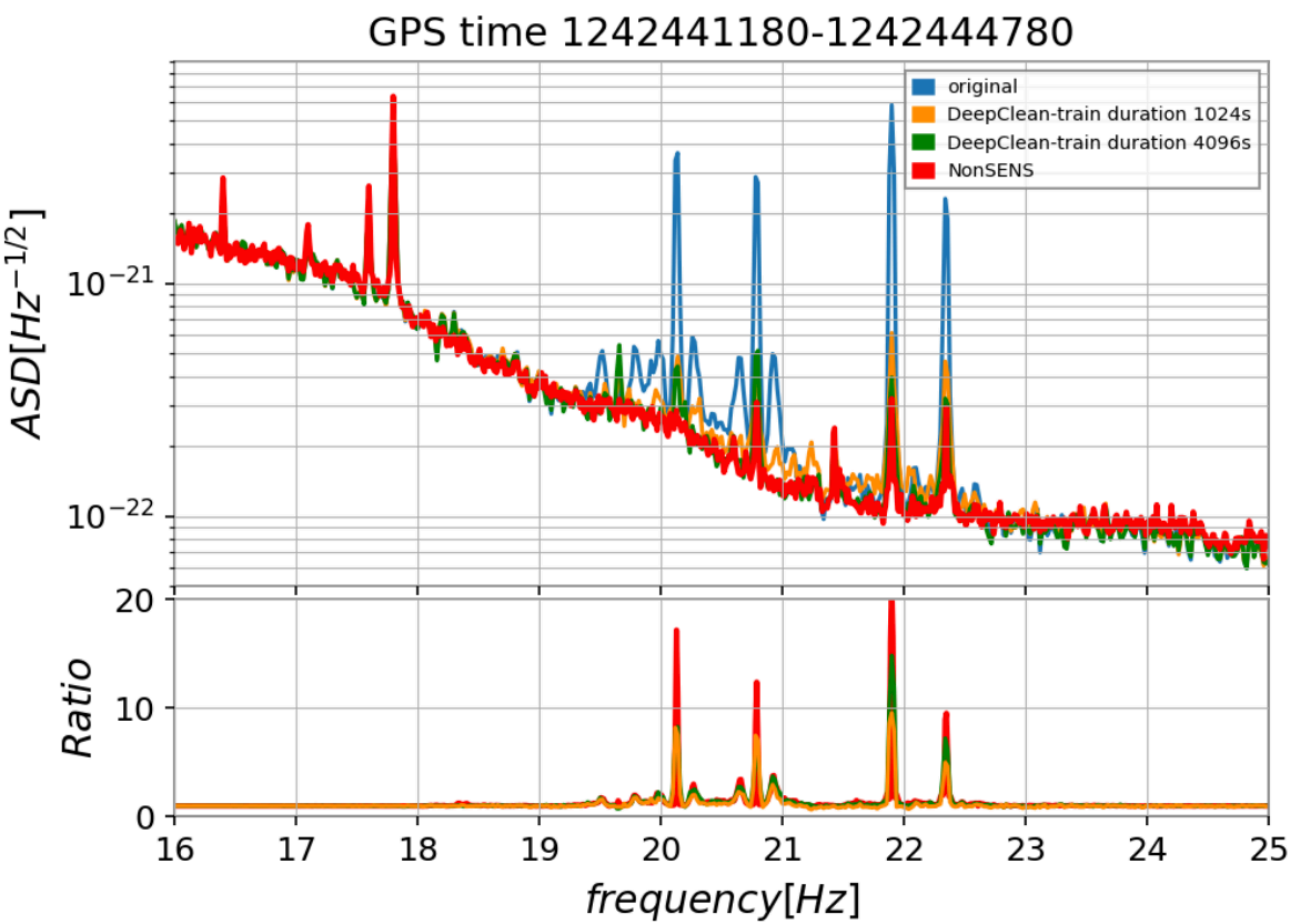


DeepClean: Full Workflow



Ormiston et al (2020)

Subtraction of low-frequency broadband noise (10-30 Hz)



A TYPICAL OFFLINE DEEPCLEAN ANALYSIS

- Training:
  - About 1000 s of data used to train
  - Sent to the CNN in batches of segments and the ASD computed by averaging over all segments.
  - Length of the segment determined by the desired resolution (for 1/4 Hz resolution, the segments should be at least 4 s longer)
- Inference:
  - No restrictions in particular on the length of the data cleaned
    - (Too shorter segments (< 10 s) not recommended, due to some edge effects)
    -

OFFLINE DEEPCLEAN: SPECIFICATIONS

- Network trained with 2000 s (or 4000 s) data from around (up to a few days before or after) the target segment.
- Target segment is chosen to be 4096 s
- Batch: 32 overlapping kernels, with each kernel being 22 x 32768 matrices
    - Assuming there are 22 channels (for 60 Hz subtraction) and one kernel is 8 s and data is sampled at 4096 Hz
    - Batch size of 16 and kernels of 4s or 2s have been found to be equally effective in cleaning, however for training, the kernels need to be at least 8 s long as the loss function is PSD-based (a mean PSD with 4 FFTs and 1/2 Hz resolution needs 8s data)

WITNESS CHANNELS USED IN 60 HZ SUBTRACTION

H1:GDS-CALIB_STRAIN

H1:PEM-CS_MAINSMON_EBAY_1_DQ

H1:ASC-INP1_P_INMON

H1:ASC-INP1_Y_INMON

H1:ASC-MICH_P_INMON

H1:ASC-MICH_Y_INMON

H1:ASC-PRC1_P_INMON

H1:ASC-PRC1_Y_INMON

H1:ASC-PRC2_P_INMON

H1:ASC-PRC2_Y_INMON

H1:ASC-SRC1_P_INMON

H1:ASC-SRC1_Y_INMON

H1:ASC-SRC2_P_INMON

H1:ASC-SRC2_Y_INMON

H1:ASC-DHARD_P_INMON

H1:ASC-DHARD_Y_INMON

H1:ASC-CHARD_P_INMON

H1:ASC-CHARD_Y_INMON

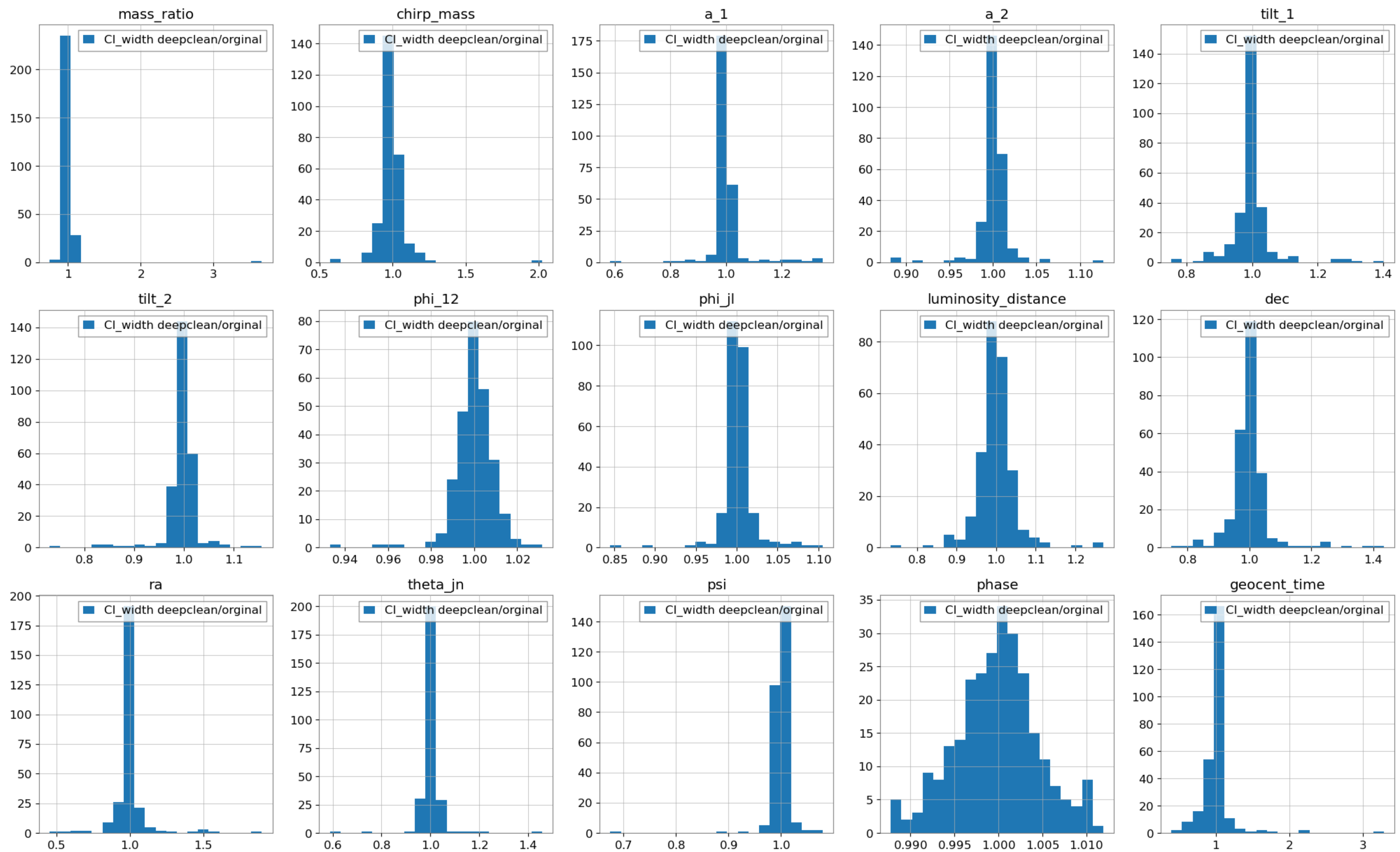
H1:ASC-DSOFT_P_INMON

H1:ASC-DSOFT_Y_INMON

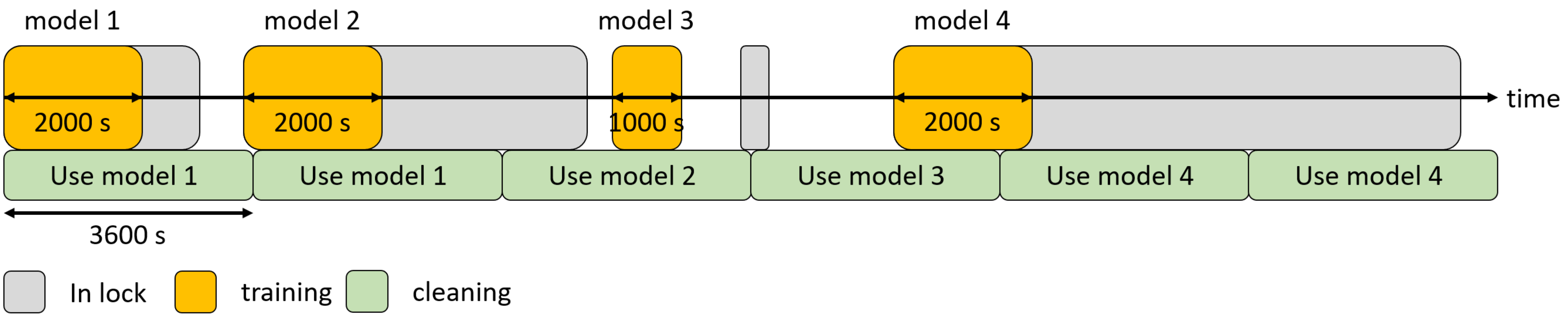
H1:ASC-CSOFT_P_INMON

H1:ASC-CSOFT_Y_INMON

PE IMPROVEMENTS: RATIO OF ERROR BARS



TRAINING STRATEGY FOLLOWED FOR MDC ANALYSIS



TRAINING DEEPCLEAN

- Preprocess: normalize (zero mean and unit variance) and bandpass to the desired frequency band.
- Data sent to the network in mini-batches with one batch typically having 32 samples
- Loss function computed for the network prediction by averaging over the mini-batch.
- Loss is minimized using ADAM (the Gradient Descent algorithm) using both Forward pass and backward pass