Anomaly Detection in Particle Accelerators using Autoencoders

Jonathan Edelen*, Nathan Cook, Callie Federer, Christopher Hall

*jedelen@radiasoft.net

September 11, 2020

SLAC Machine Learning Seminar

This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of High Energy Physics, under Award Number DE-SC0019979

Acknowledgements

- Institutional Support
  - Many thanks to Michael Borland at the APS for the magnet fault data
  - Data from Fermilab acquired between Oct 2019 – March 2020

Anomaly Detection in Particle Accelerators

- Supervised Learning
  - you know the what the faults are
- Unsupervised Learning
  - you don't know what the faults are

[1] A. S. Nawaz, S. Pfeiffer, G. Lichtenberg, and H. Schlarb, "Self-organzied critical control for the european xfel using black box parameter identification for the quench detection system," in 2016 3rd Conference on Control and Fault-Tolerant Systems (SysTol), Sep. 2016, pp. 196–201.

[2] A. Nawaz, S. Pfeiffer, G. Lichtenberg, and P. Rostalski, "Anomaly detection for the european xfel using a nonlinear parity space method," IFAC-PapersOnLine, vol. 51, no. 24, pp. 1379 – 1386, 2018, 10th IFAC Symposium on Fault Detection, Supervision and Safety for Technical Processes SAFEPROCESS 2018.

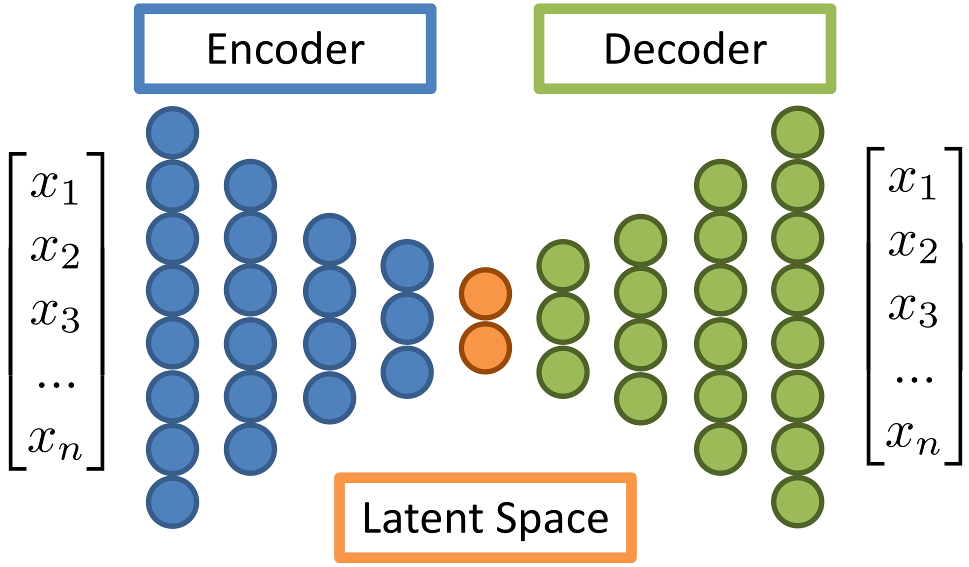
[3] M. Wielgosz, A. Skoczea, and M. Mertik, "Using lstm recurrent neural networks for monitoring the lhc superconducting magnets," Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, vol. 867, pp. 40 –50, 2017.

[4] Elena Fol, "Evaluation of Machine Learning Methods for LHC Optics Measurements and Corrections Software" CERN-THESIS-2017-336, Aug 2017

[5] G. Valentino, R. Bruce, S. Redaelli, R. Rossi, P. Theodoropoulos, and S. Jaster-Merz, "Anomaly detection for beam loss maps in the large hadron collider," Journal of Physics: Conference Series, vol. 874, p. 012002, jul 2017.

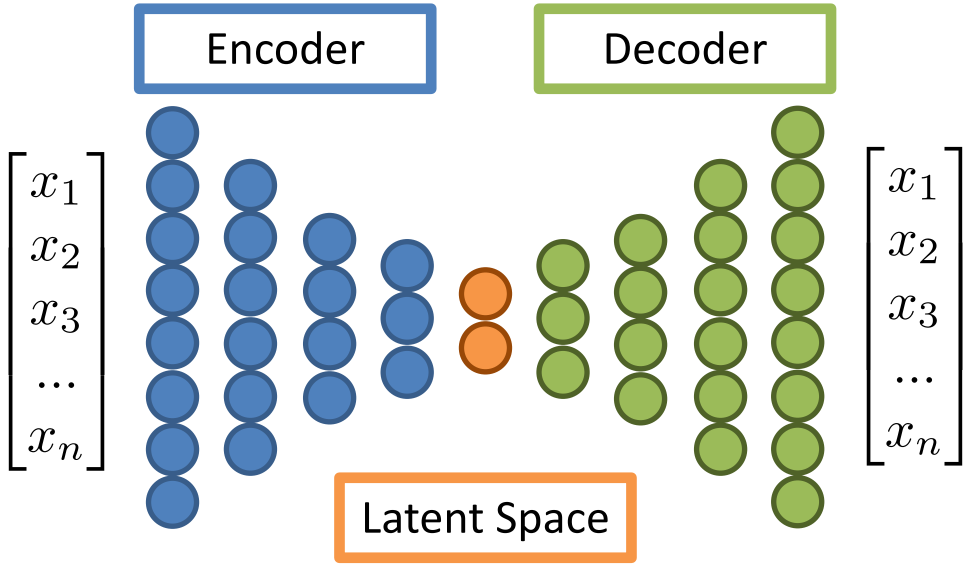
What is an autoencoder?

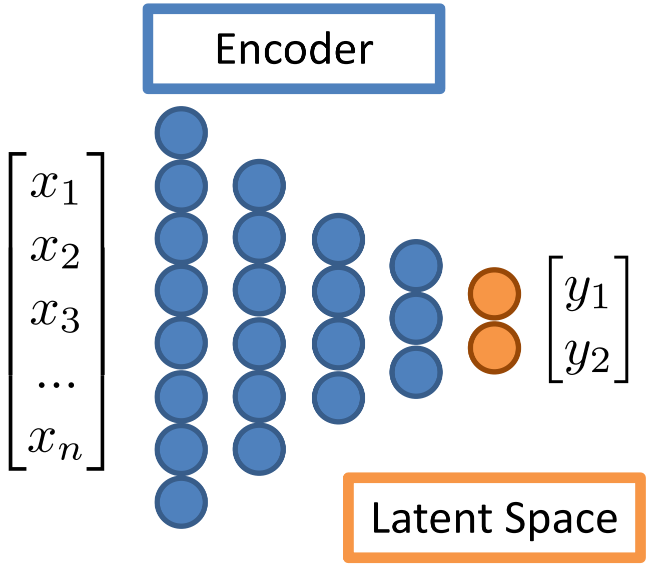
- Neural networks structures
  - The encoder reduces the input dimension to some minimum number of nodes
  - The decoder reconstructs the input data
  - A trained autoencoder is essentially an identity transformation for the input data



Direct analysis of latent space

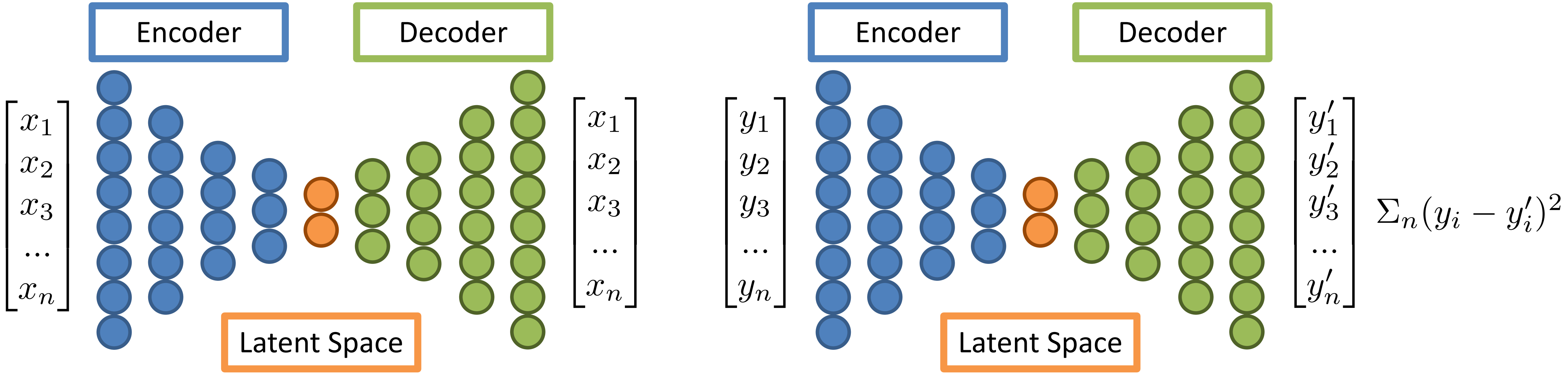
- Use output of encoded dimension for given inputs
  - Compressed representation of the input dataset
  - Clustering can provide information about anomalies
  - Use as input to neural network for model development





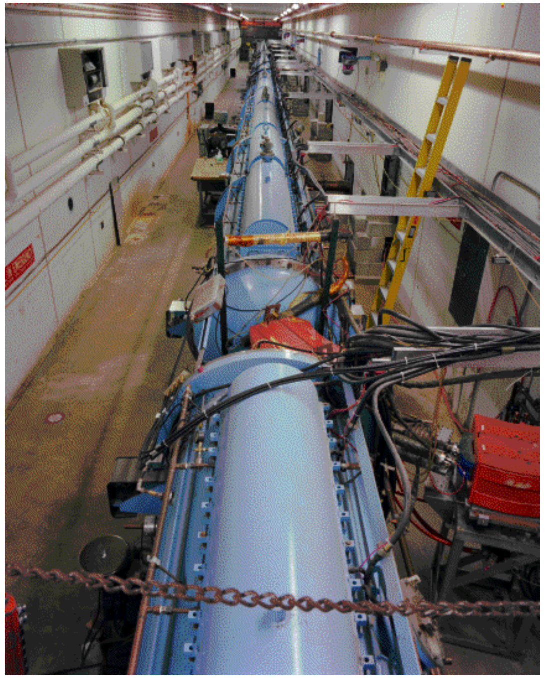
Reconstruction Tests

- Reconstruct unknown data using an autoencoder
  - Train and validate the autoencoder on known good datasets
  - Test on unknown data (may be good or bad)
  - Measure the degree to which the autoencoder successfully reconstructs the unknown data



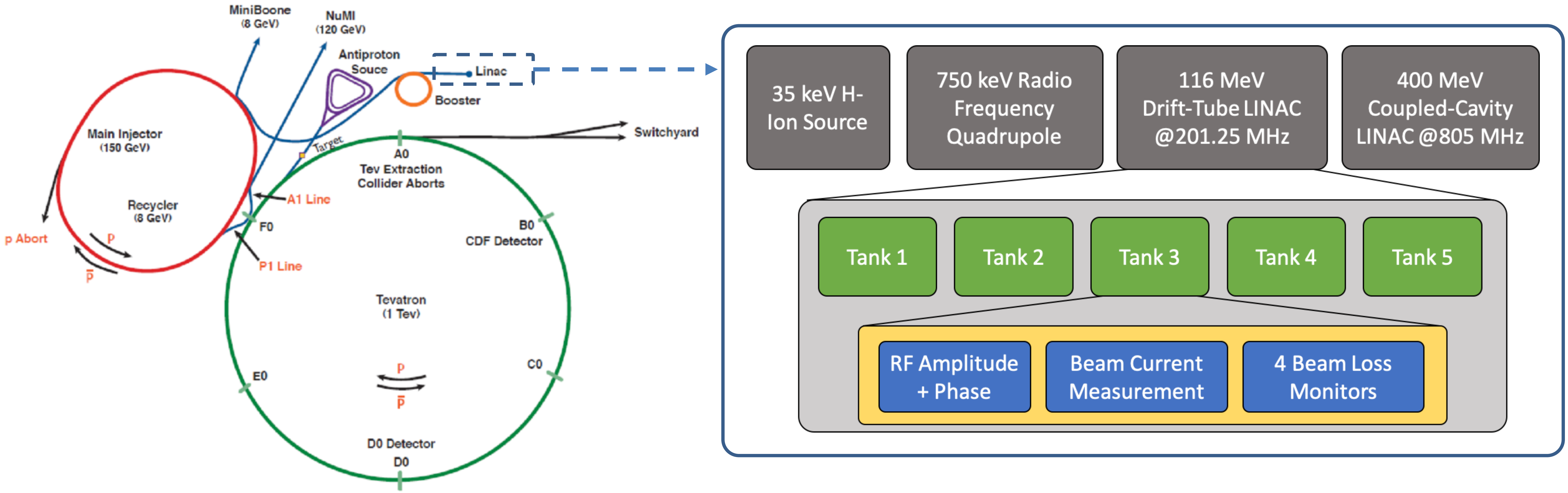
Two Case Studies

- Fermilab Low Energy LINAC
  - Identify root cause of changes in the output current
  - Study both latent space analysis and reconstruction tests
  - 10 input parameters



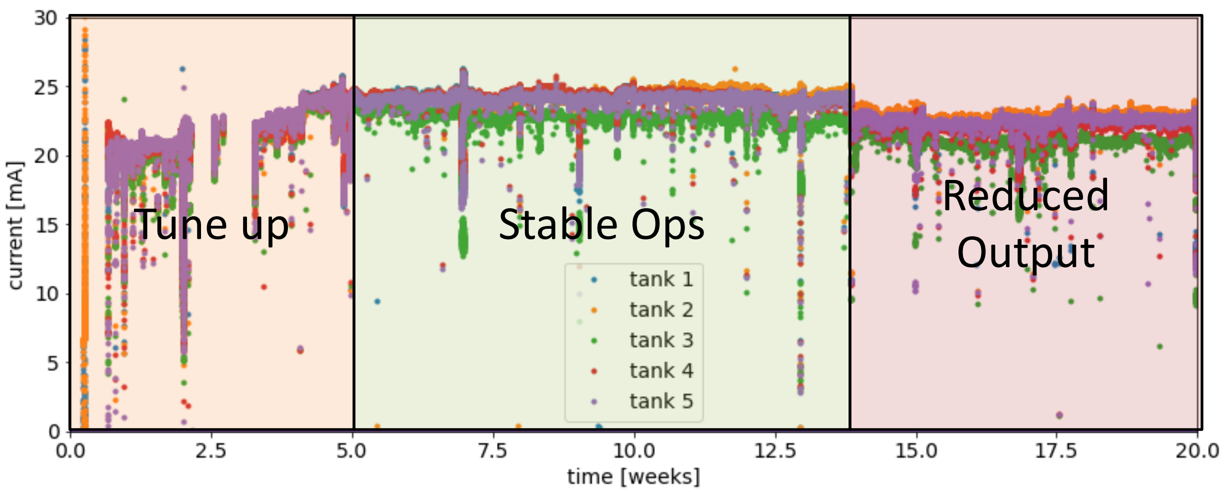
- APS Storage Ring
  - Identify faulty magnets
  - Study reconstruction tests
  - Up to 1320 input parameters

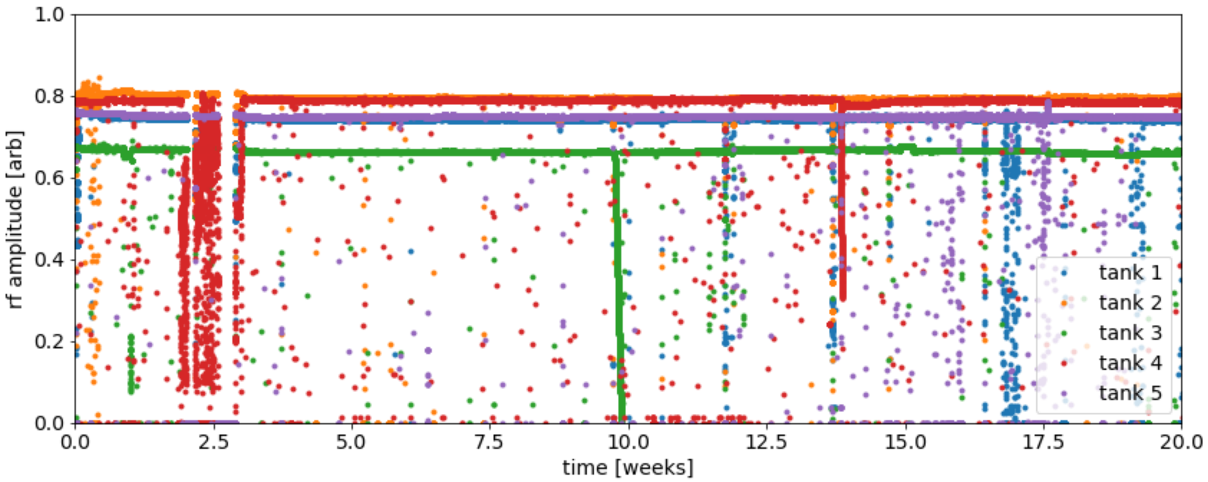
The Fermilab Low Energy LINAC

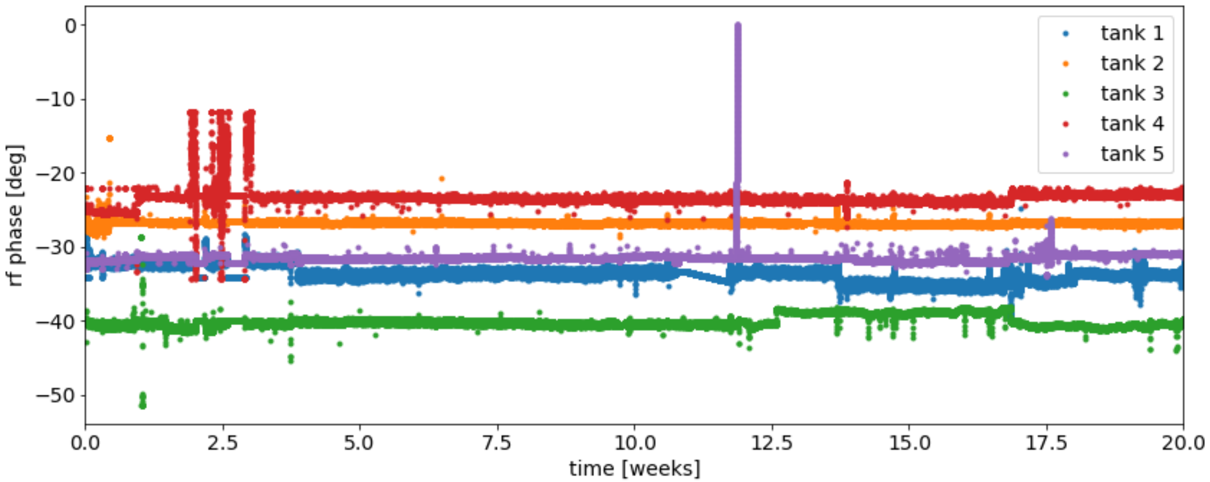


Diagnostic data from the DTL sections

- Data taken over 20 week period
  - RF amplitude and Phase data, and toroid data
- Three operational modes
  - Weeks 1-4 are largely tune up
  - Weeks 5-13 operations are relatively stable
  - Change in current around week 13

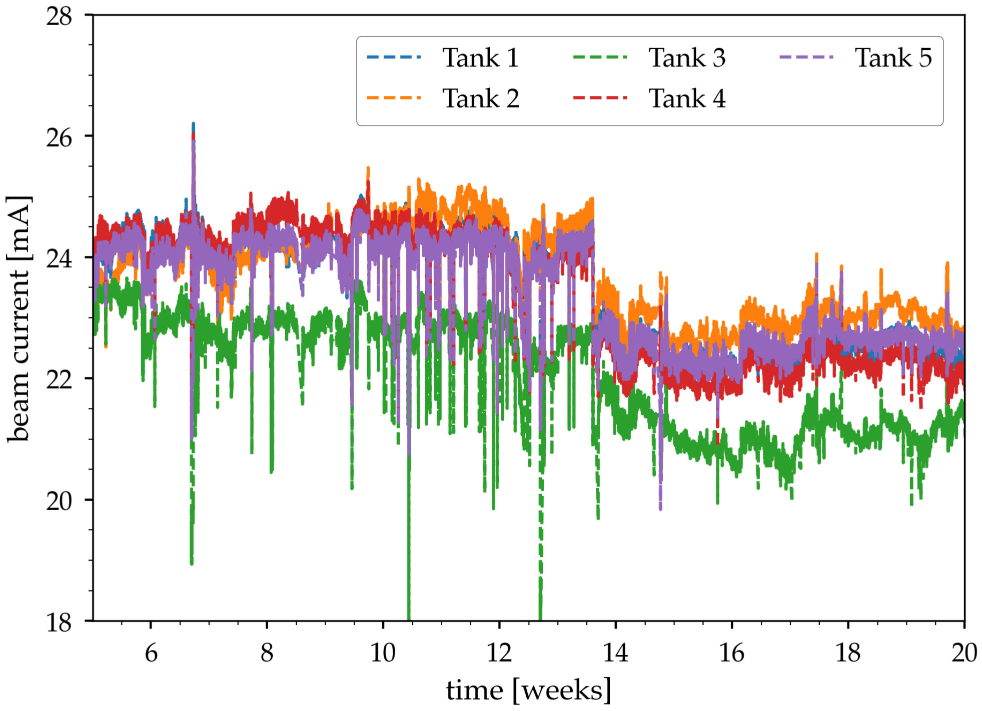




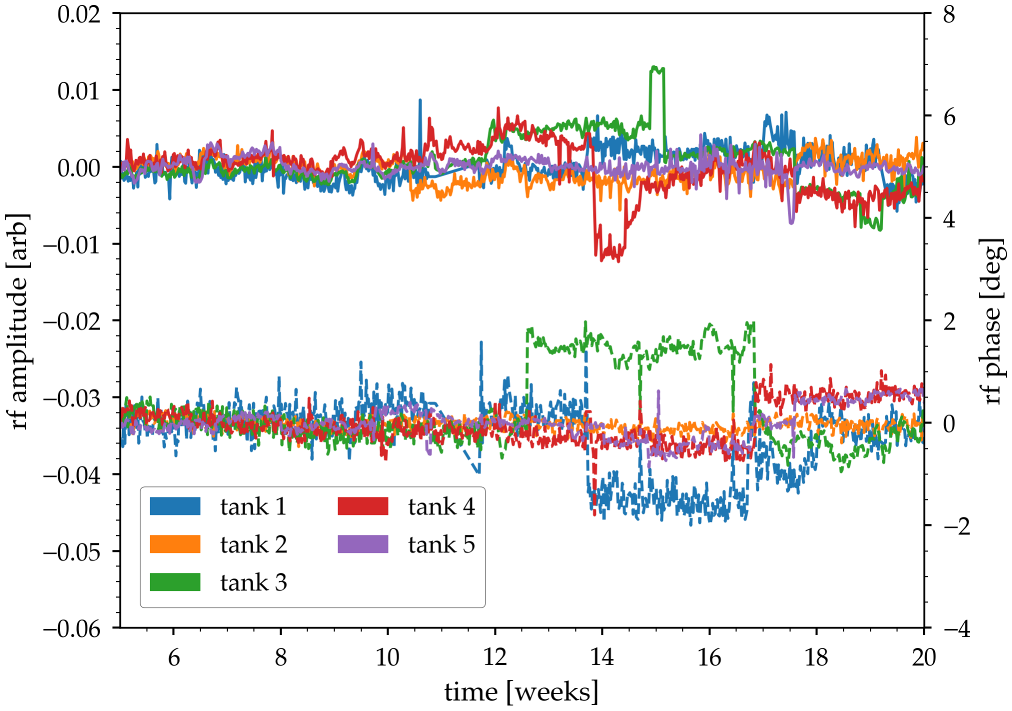


Can we understand where the change in current comes from?

- Zooming in on Weeks 5-20
  - Weeks 5-13 operations are relatively stable
  - Abrupt Change in current around week 13
  - Slow change settles out around week 16

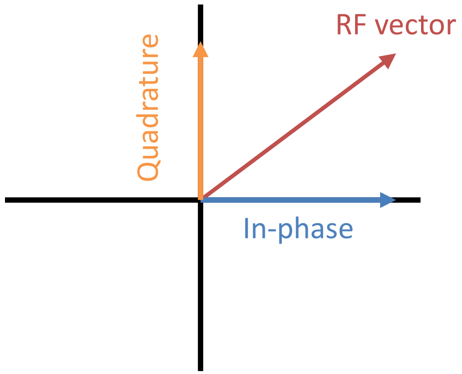


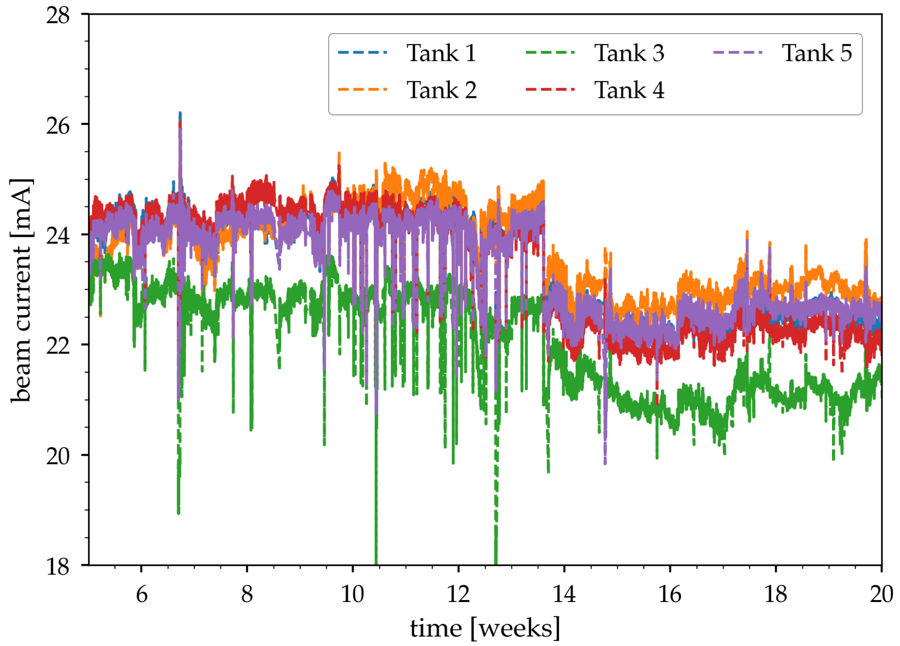
- RF Data from the five DTL tanks
  - Amplitude and phase information recorded at 1 min intervals
  - Median values for each parameter removed to show change with time

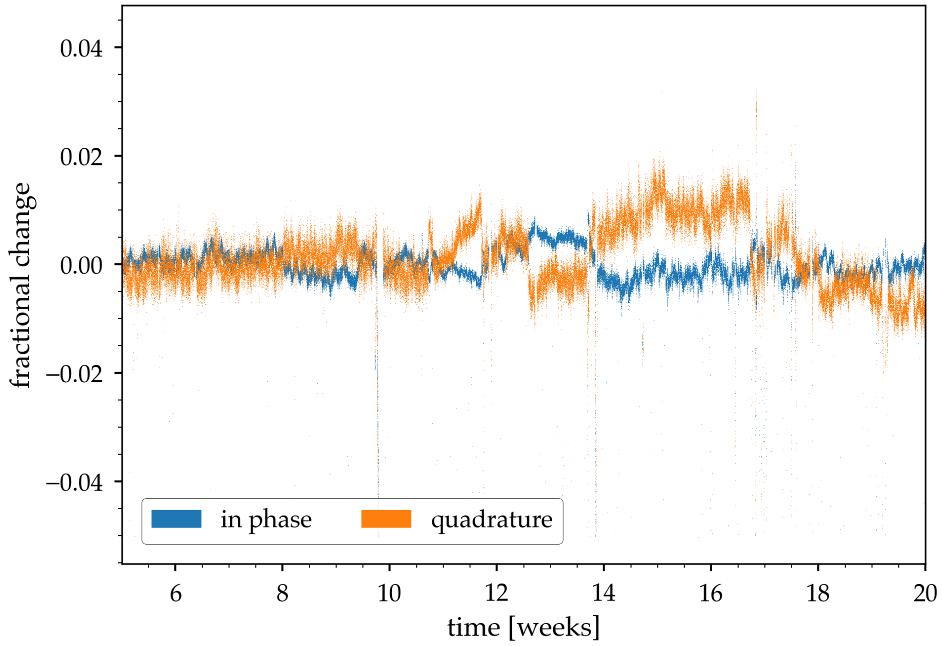


Examine vector components of RF data

- Vector components
  - Decompose each RF signal into in-phase and quadrature components
  - Sum in-phase and quadrature components
- I/Q components of the low energy LINAC between weeks 5 and 20
  - Fractional change is relative to the median value of each signal.
  - Changes on the order of 1%

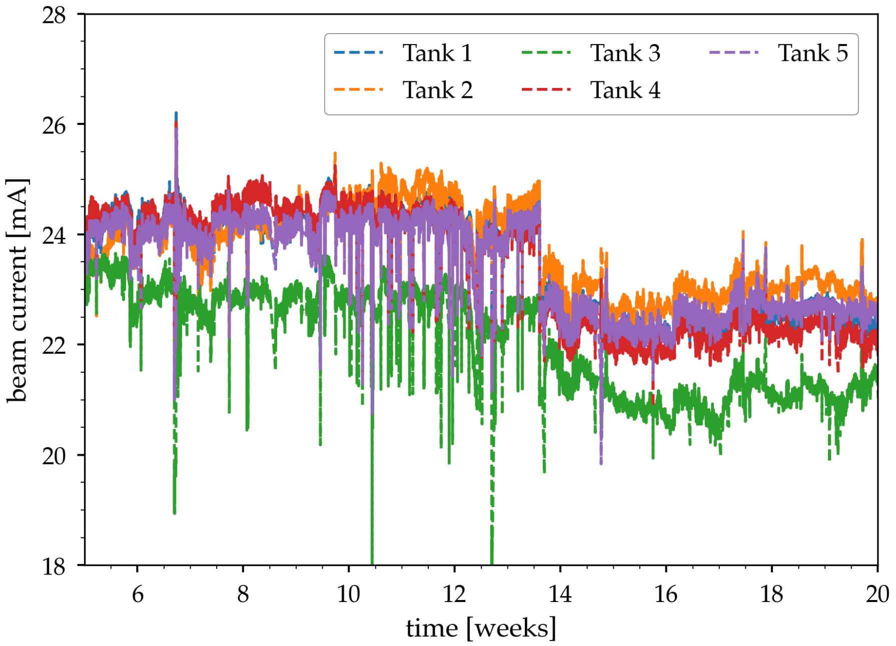


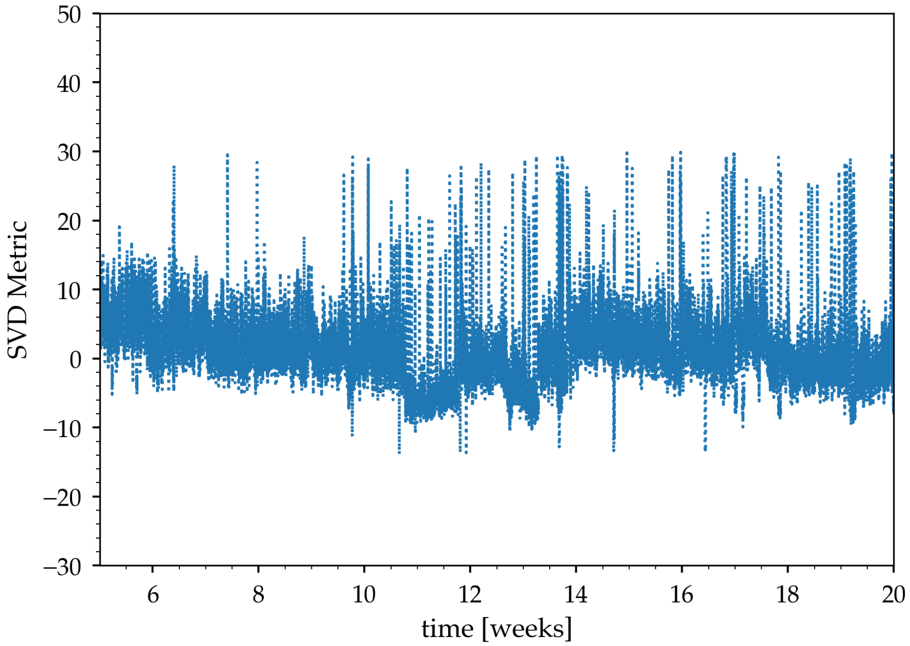




Singular Value Decomposition of RF data

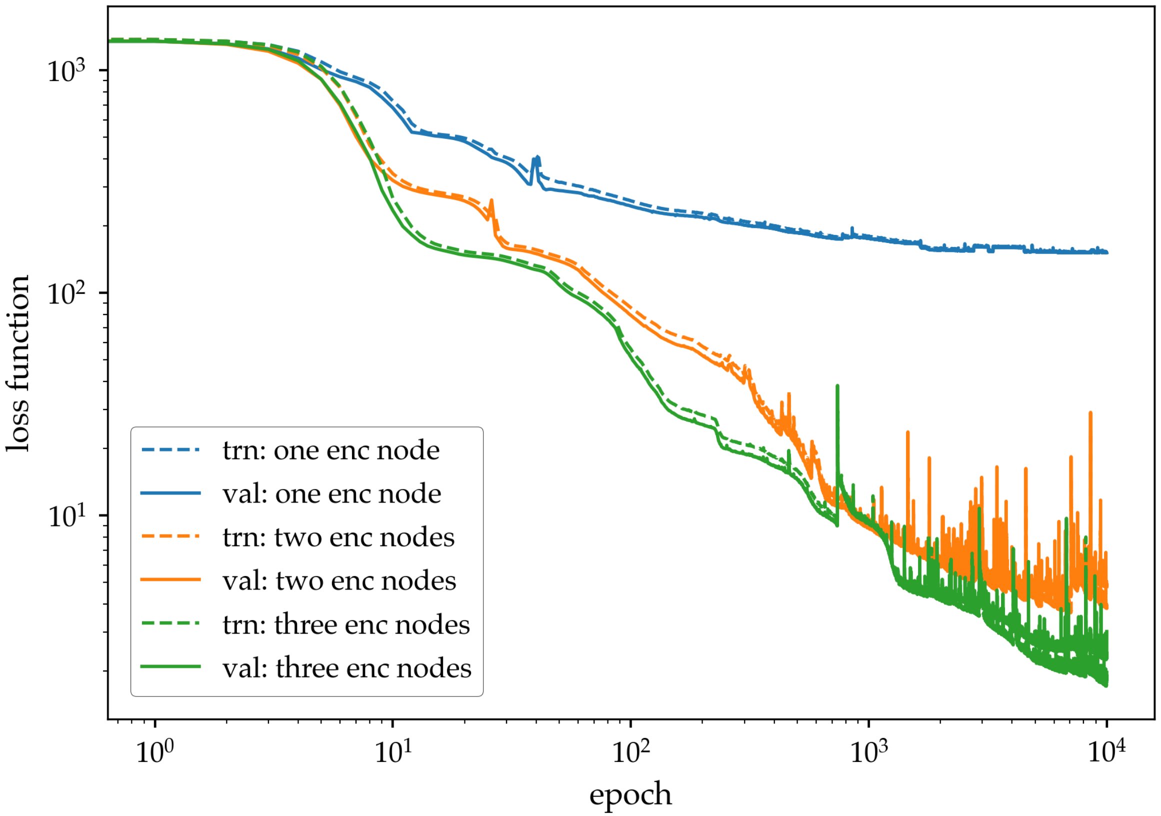
- Linear SVD of RF data over time
  - In increments of 10 samples perform SVD of the RF data
  - Scale each singular value relative to its median
  - Sum over the singular values
- SVD metric between weeks 5 and 20
  - Change over time does not provide clear indicators of a relationship with the change in output current





Autoencoder Parameter/Architecture Selection

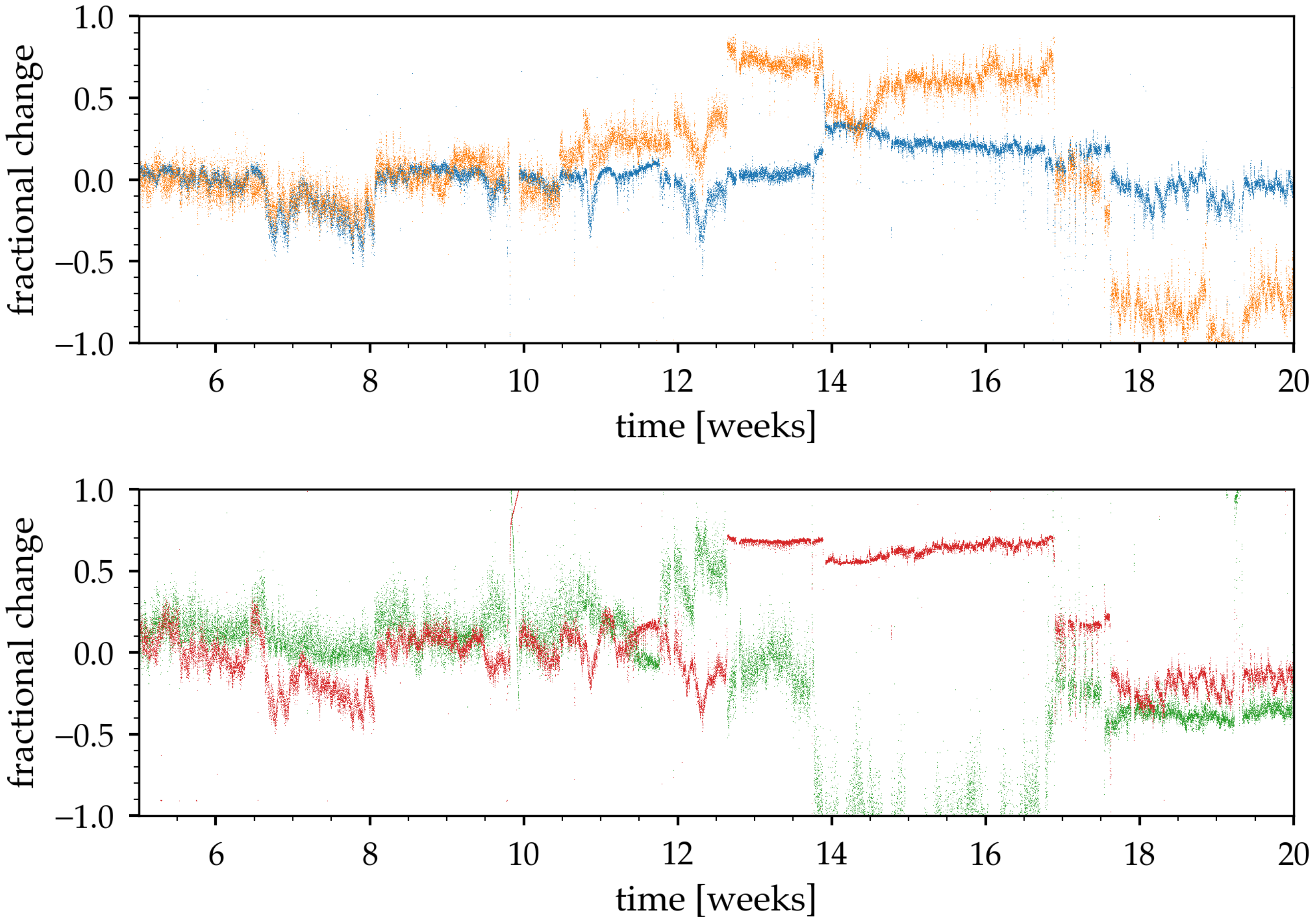
- Autoencoders perform dimensionality reduction for latent space analysis
  - 10 inputs and outputs (5 amplitude and 5 phase measurements)
  - Test architectures with 1, 2, and 3 encoded nodes
  - Right: Results averaged over 10 random initializations
- Training parameters:
  - 140k samples for training
  - 60k samples for validation.
  - Batch size was 20k
  - RELU activation functions
  - 30% Gaussian noise
- Architecture
  - 30-20-10-6-4-enc-4-6-10-20-30



Autoencoders provide clear indicators of a state change in the RF

- Latent space of RF parameters as a function of time
  - Networks initialized with different random weights
  - Median value is subtracted
  - Normalized to the median

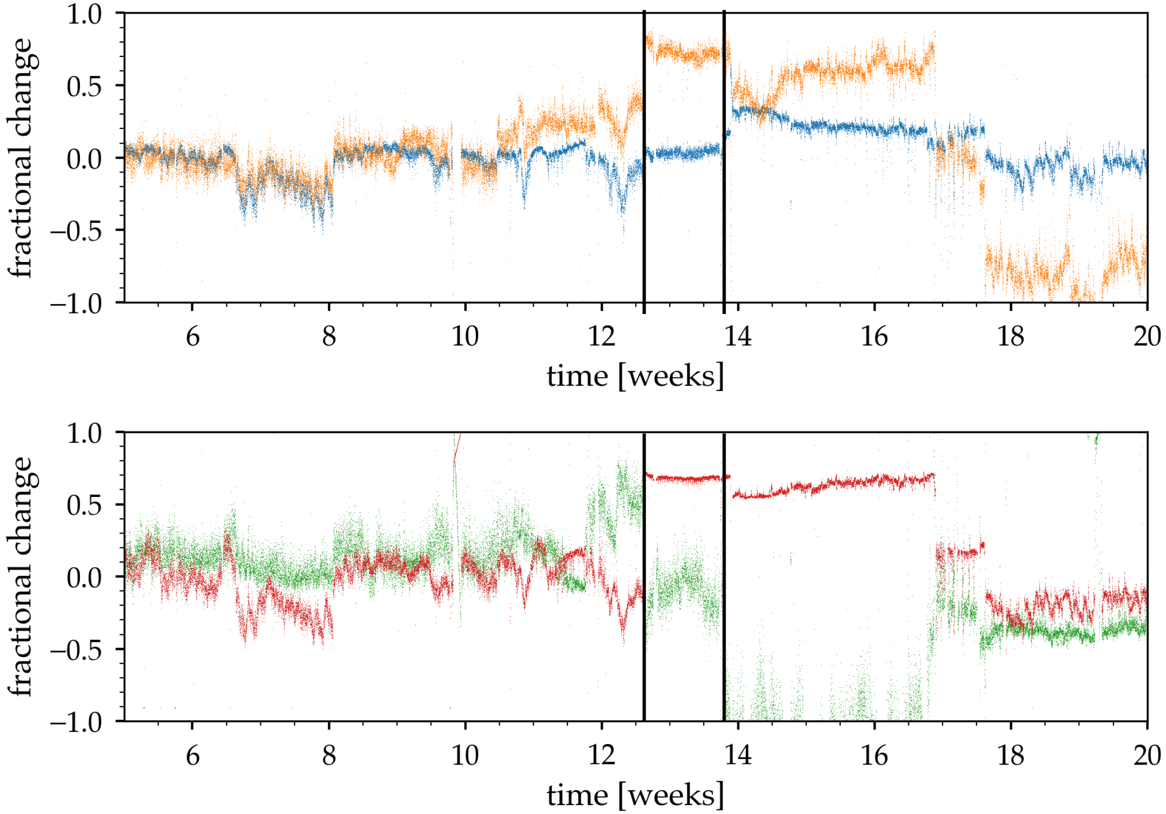
Two examples of latent space analysis

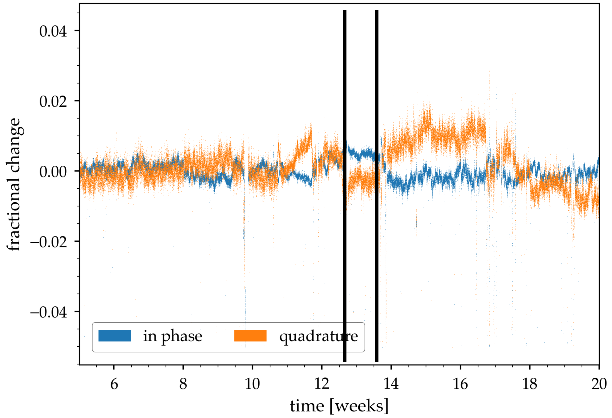


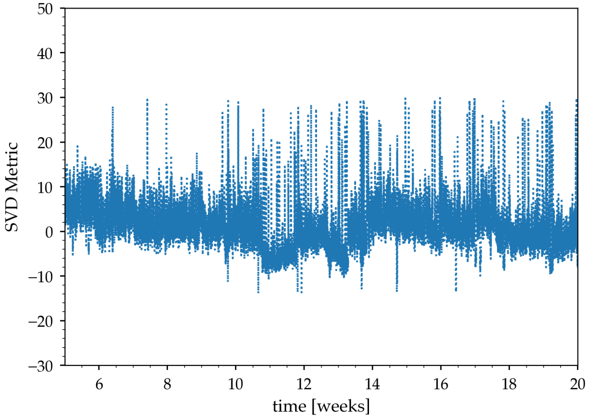
Autoencoders provide clear indicators of a state change in the RF

- Latent space of RF parameters as a function of time
  - Magnitude of changes better reflected by latent space analysis.
  - Change at week 13 significant in latent space but not significant in I/Q or SVD analysis

Two examples of latent space analysis

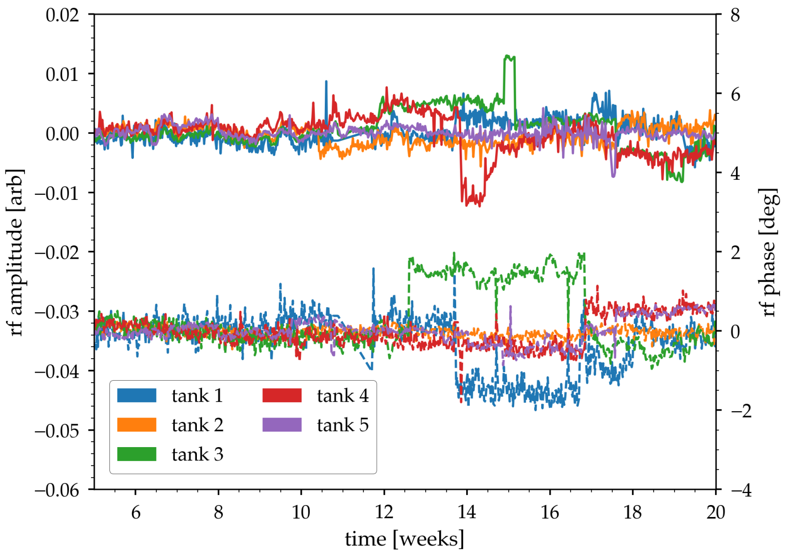






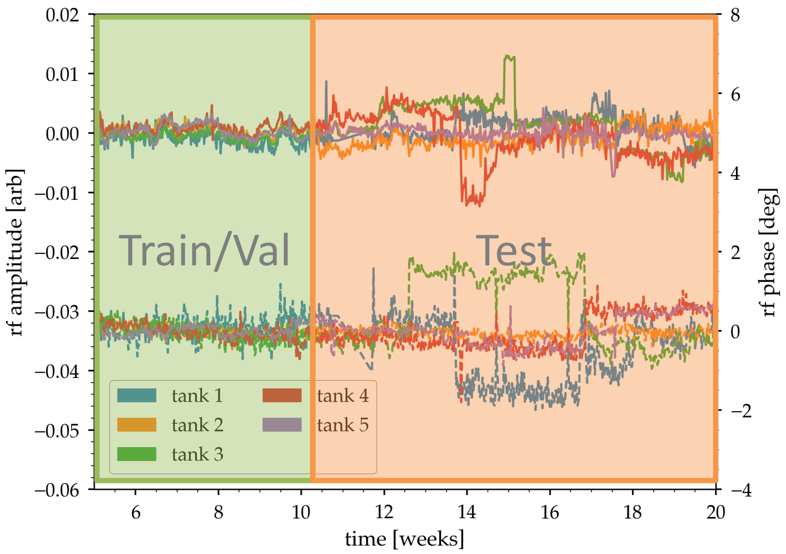
Autoencoder reconstruction study

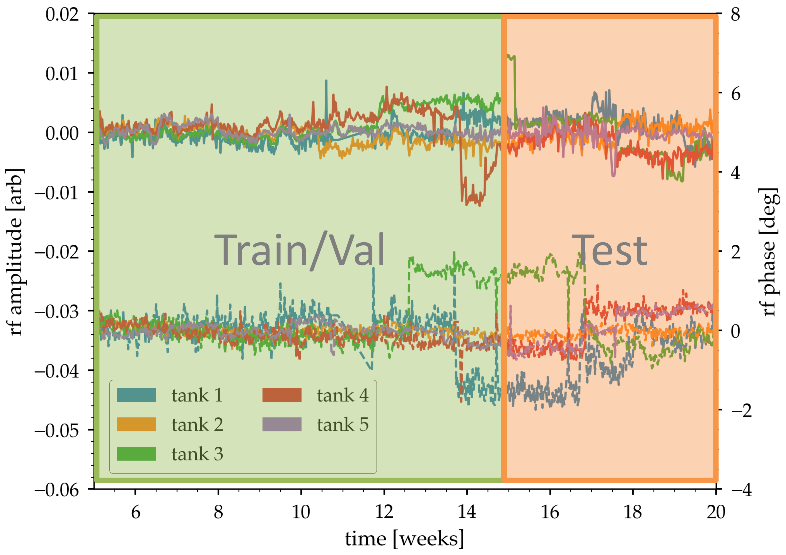
- Two cases
  - Train and validate on 5 weeks of RF data – test on 10 weeks
  - Train and validate on 10 weeks of RF data – test on 5 weeks



Autoencoder reconstruction study

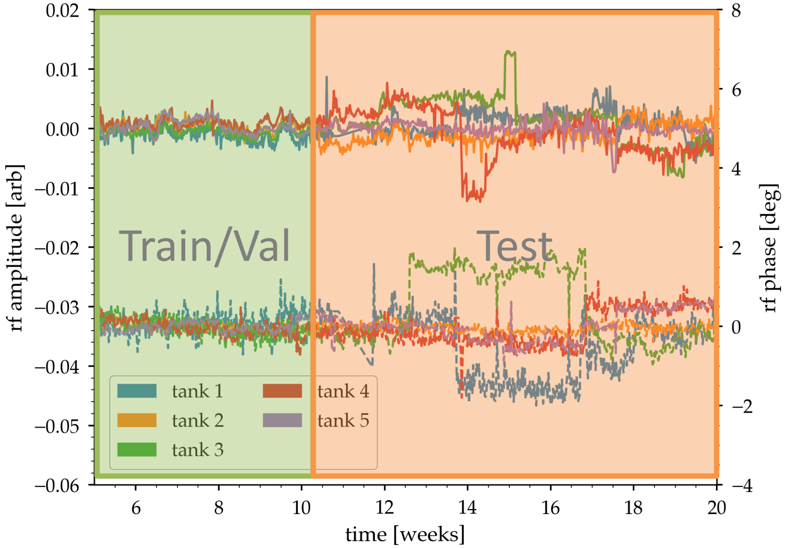
- Two cases
  - Top: Train and validate on 5 weeks of RF data – test on 10 weeks
  - Bottom: Train and validate on 10 weeks of RF data – test on 5 weeks

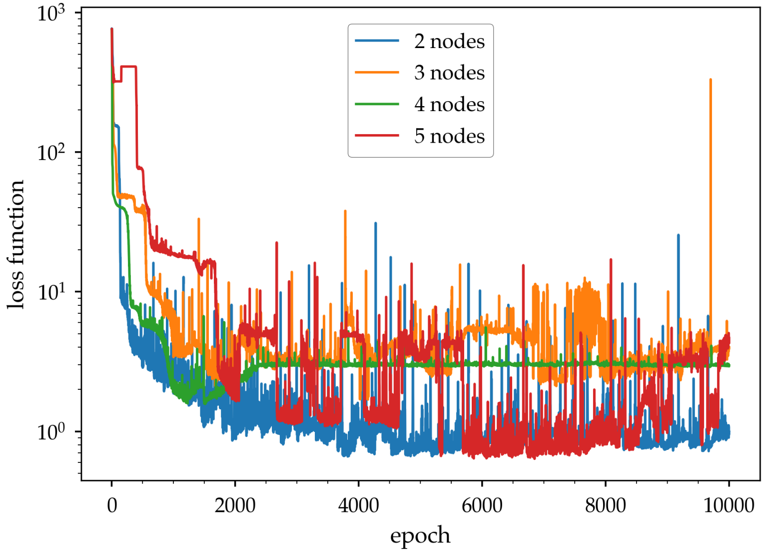


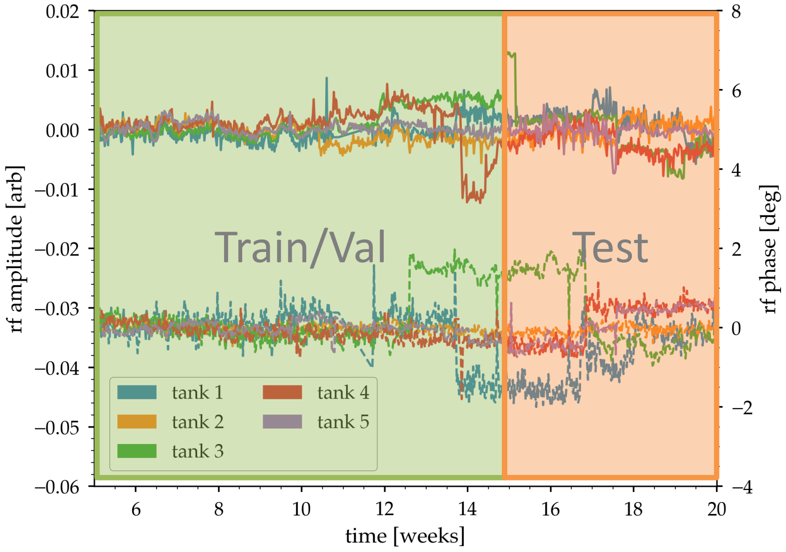


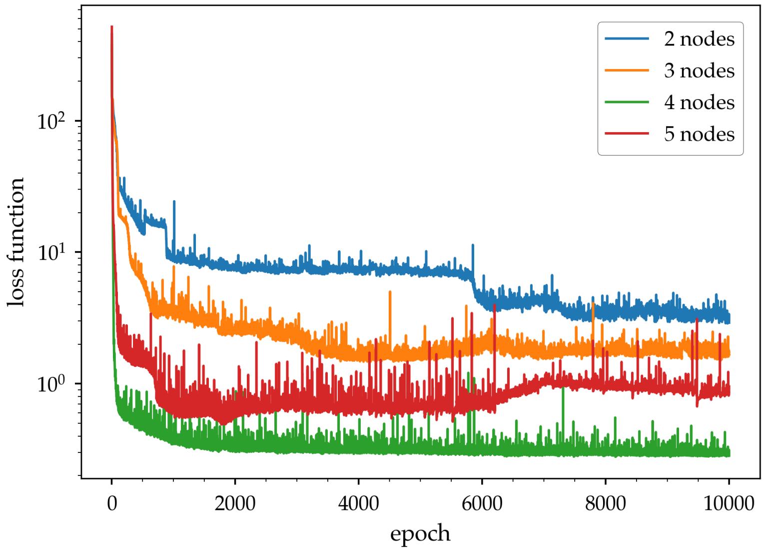
Autoencoder reconstruction study

- Two cases
  - Top: Train and validate on 5 weeks of RF data – test on 10 weeks
  - Bottom: Train and validate on 10 weeks of RF data – test on 5 weeks
- Right: convergence for different number of encoded dimensions
  - Performance metric is RMS reconstruction error
  - Not dependent on the number of encoded nodes



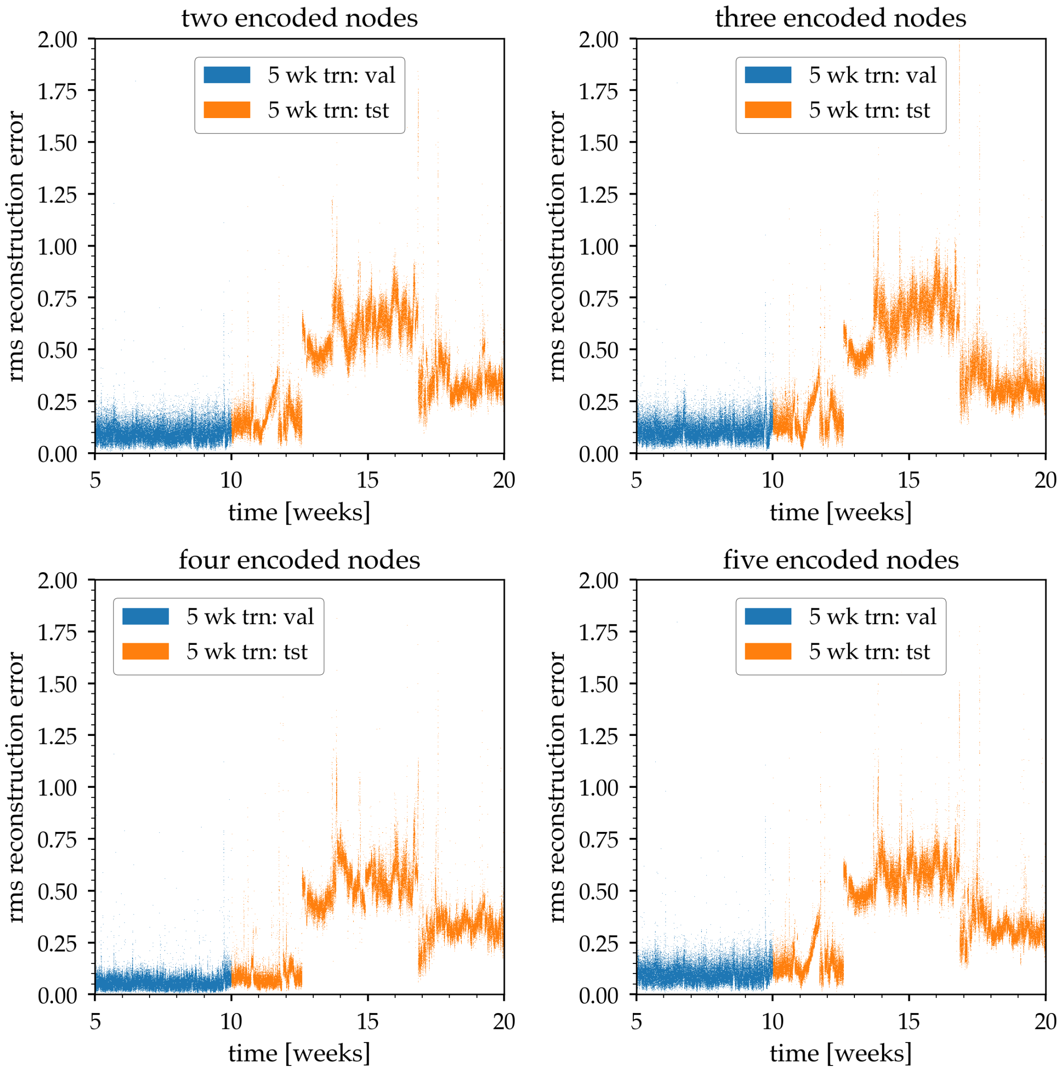






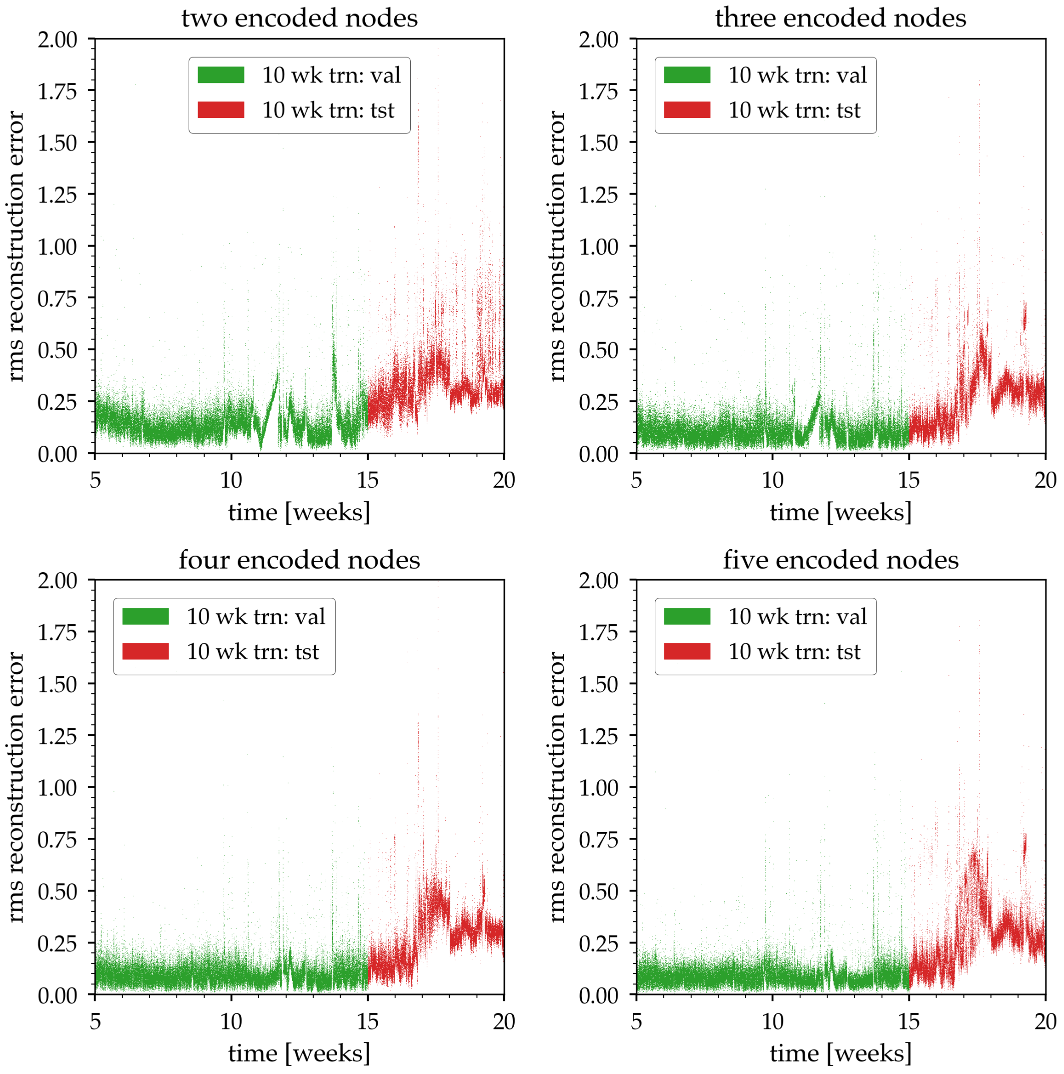
Autoencoder reconstruction study

- Case 1
  - Train and validate on 5 weeks of RF data (blue)
  - Test on 10 weeks (orange)
- RMS Reconstruction error as a function of time
  - Variance in the error goes down as the number of encoded nodes is increased up to four encoded nodes.
  - Difference in the reconstruction error between weeks 5 and 19 remains constant



Autoencoder reconstruction study

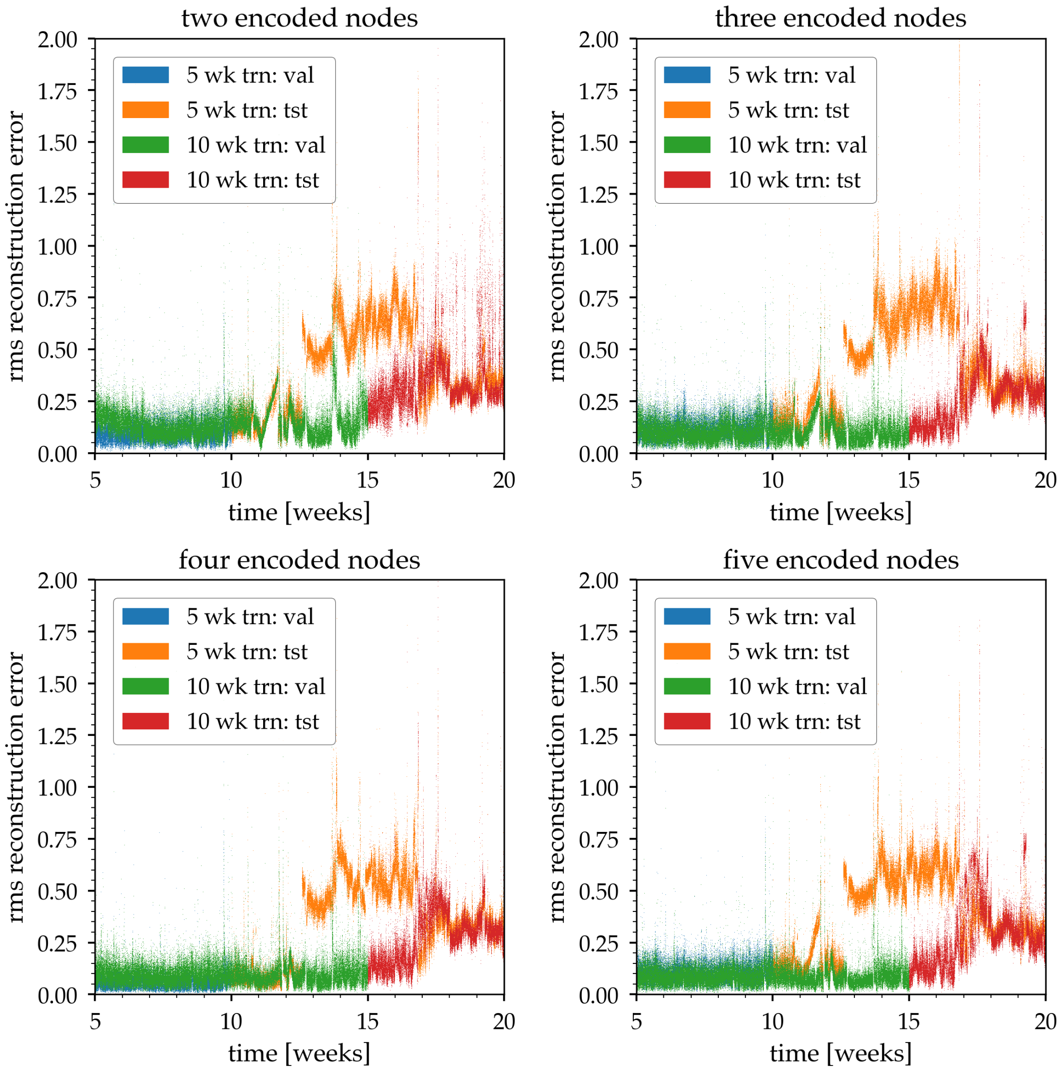
- Case 2
  - Train and validate on 10 weeks of RF data (green)
  - Test on 5 weeks (red)
- RMS Reconstruction error as a function of time
  - Variance in the error goes down as the number of encoded nodes is increased up to four encoded nodes.
  - Difference in the reconstruction error between weeks 5 and 19 remains constant



Autoencoder reconstruction study

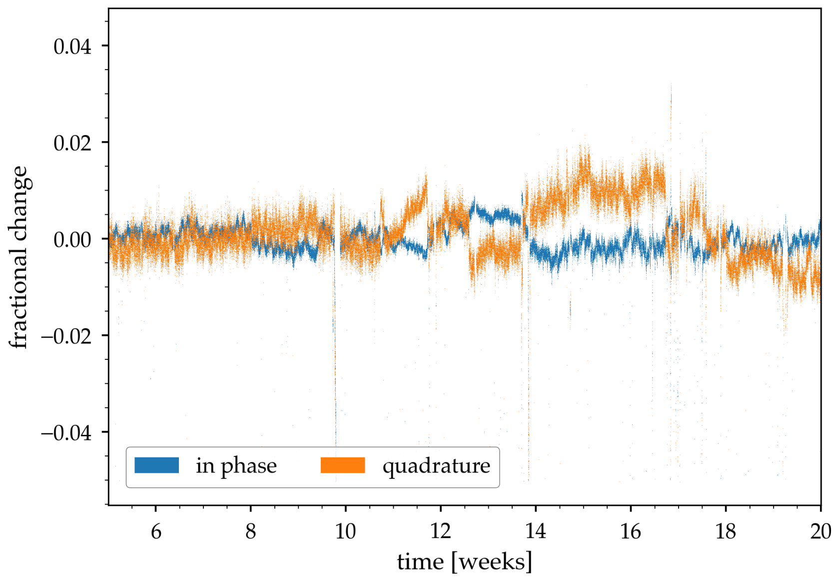
- Comparison of two training paradigms
  - Case 1 (blue/orange)
  - Case 2 (green/red)

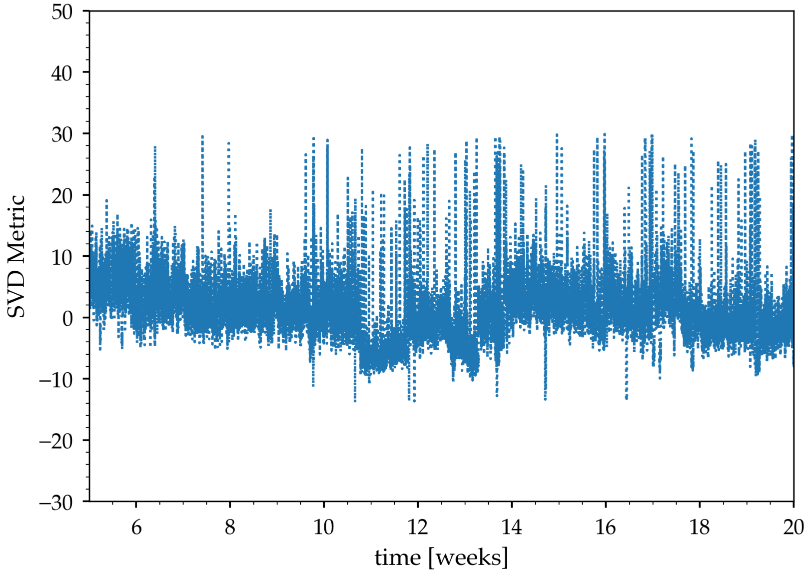
- RMS Reconstruction error as a function of time
  - Variance in the error goes down as the number of encoded nodes is increased up to four encoded nodes.
  - Difference in the reconstruction error between weeks 5 ands 19 is consistent across all studies

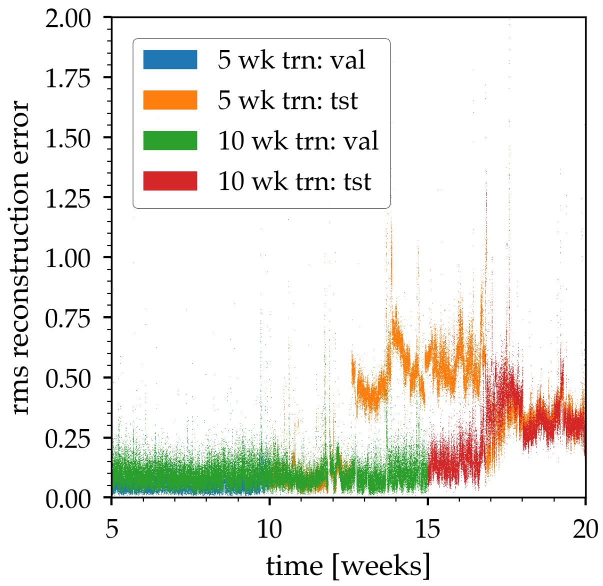


Summary of reconstruction analysis

- Reconstruction error highlights significant change in the RF state between weeks 15 and 20
- Fractional change in I/Q components show a change, but it is comparatively small
- SVD metric does not show clear picture
- Suggestive that change in output current caused by RF

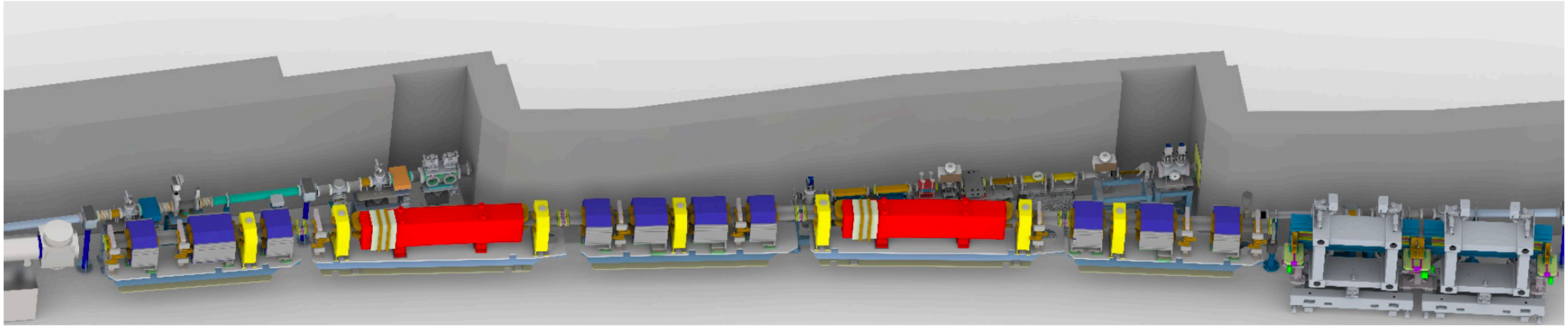






Detecting faulty magnet power supplies in the APS

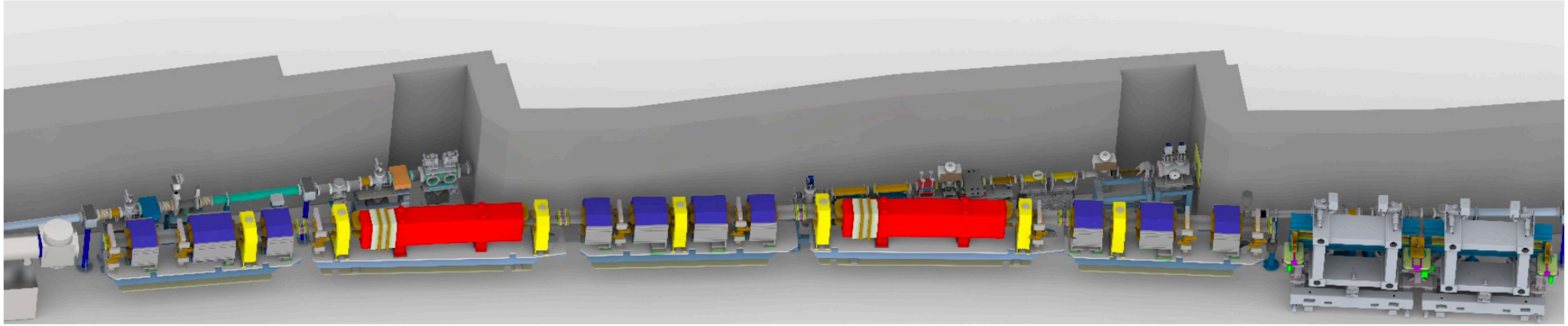
- Can we predict if a fault will occur?
  - If yes, can we predict which magnet will fault



https://www.energy.gov/sites/prod/files/2019/04/f62/Advanced-Photon-Source-Upgrade-Project.pdf

Detecting faulty magnet power supplies in the APS

- Can we predict if a fault will occur?
  - If yes, can we predict which magnet will fault
- Components of interest
  - 1320 magnet power supplies
  - 40 sectors (each has A and B sections)
  - A section: 4 horizontal correctors, 4 vertical correctors, 5 quads, and 4 sextupoles
  - B type: 4 horizontal correctors, 4 vertical correctors, 5 quads, and 3 sextupoles

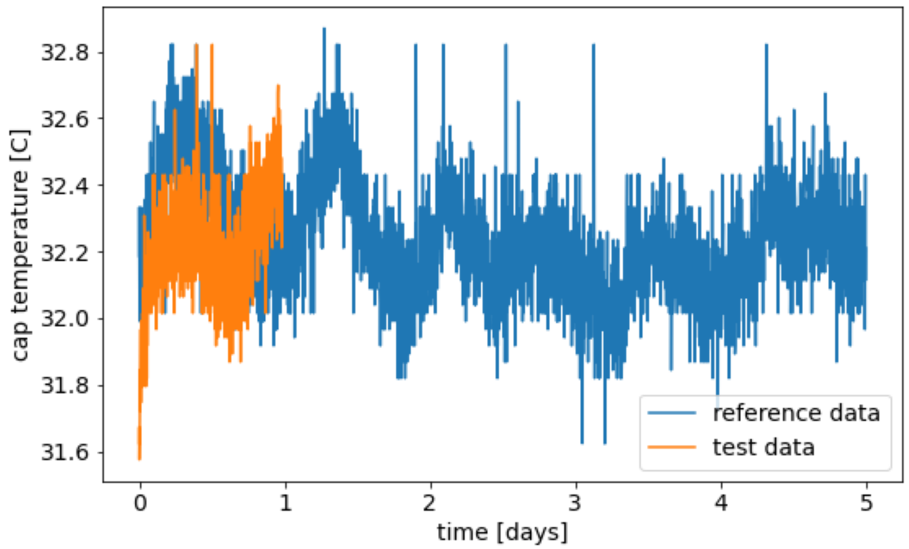


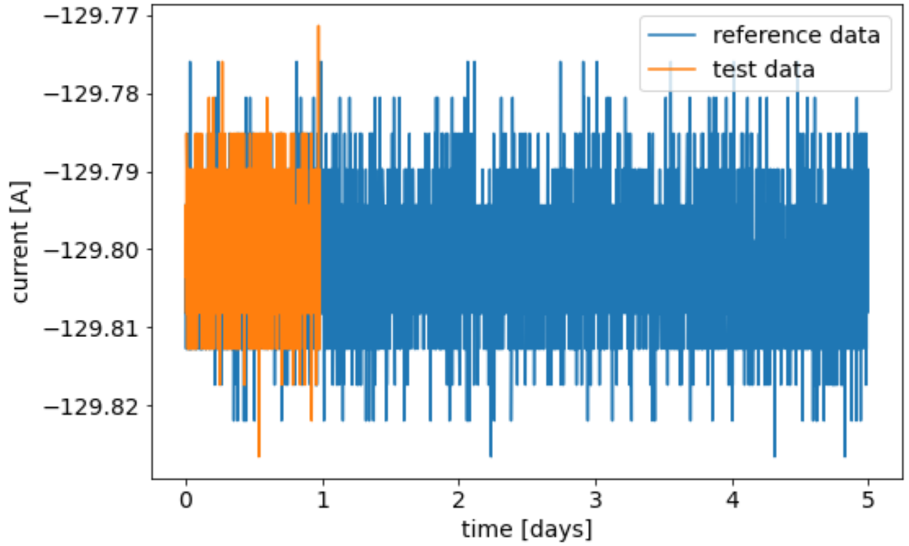
https://www.energy.gov/sites/prod/files/2019/04/f62/Advanced-Photon-Source-Upgrade-Project.pdf

Detecting faulty magnet power supplies in the APS

- Time series data for 1320 magnets
  - Power supply cap temperature
  - Current
  - Magnet temperature

- Reference data (blue)
  - No fault occurs in vicinity, normal operations
- Test data (orange)
  - Magnet failure occurs
  - Data is clipped and does not include final minutes prior to magnet fault

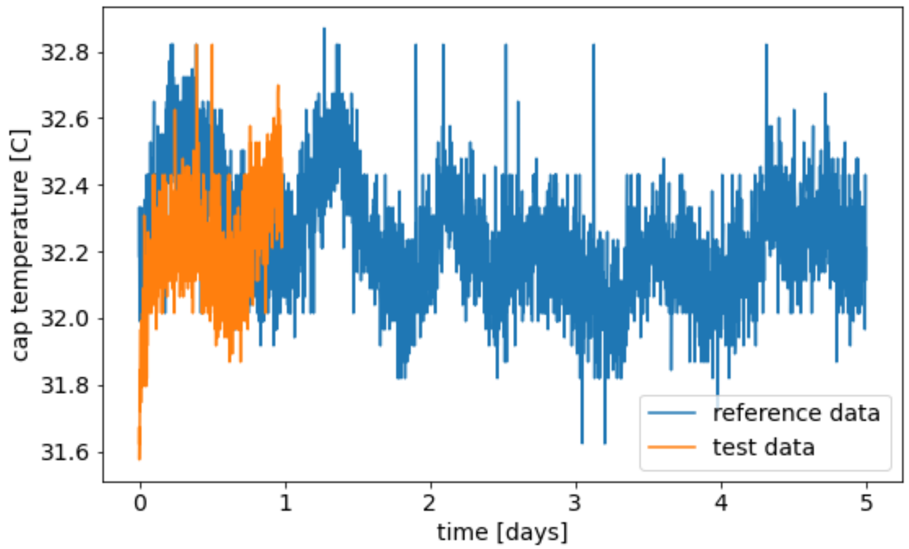


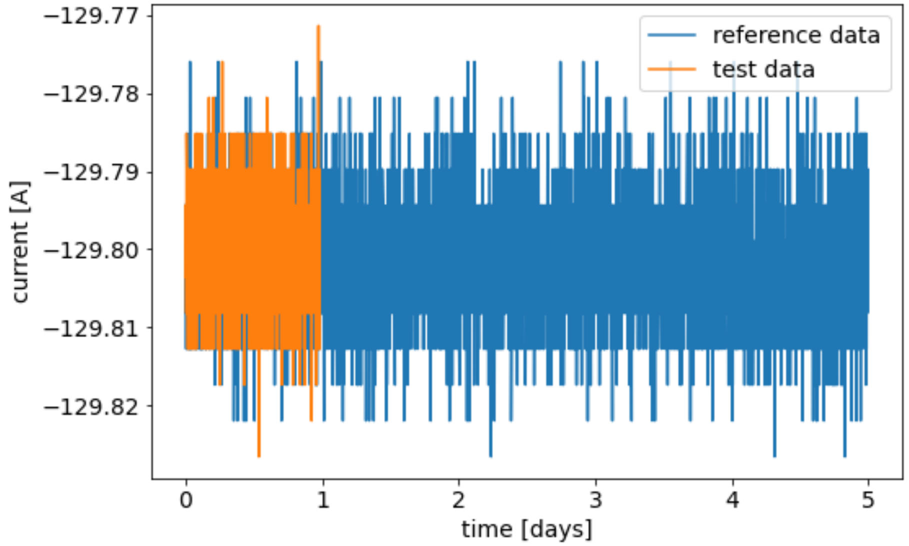


Detecting faulty magnet power supplies in the APS

- Time series data for 1320 magnets
  - Power supply cap temperature
  - Current
  - Magnet temperature

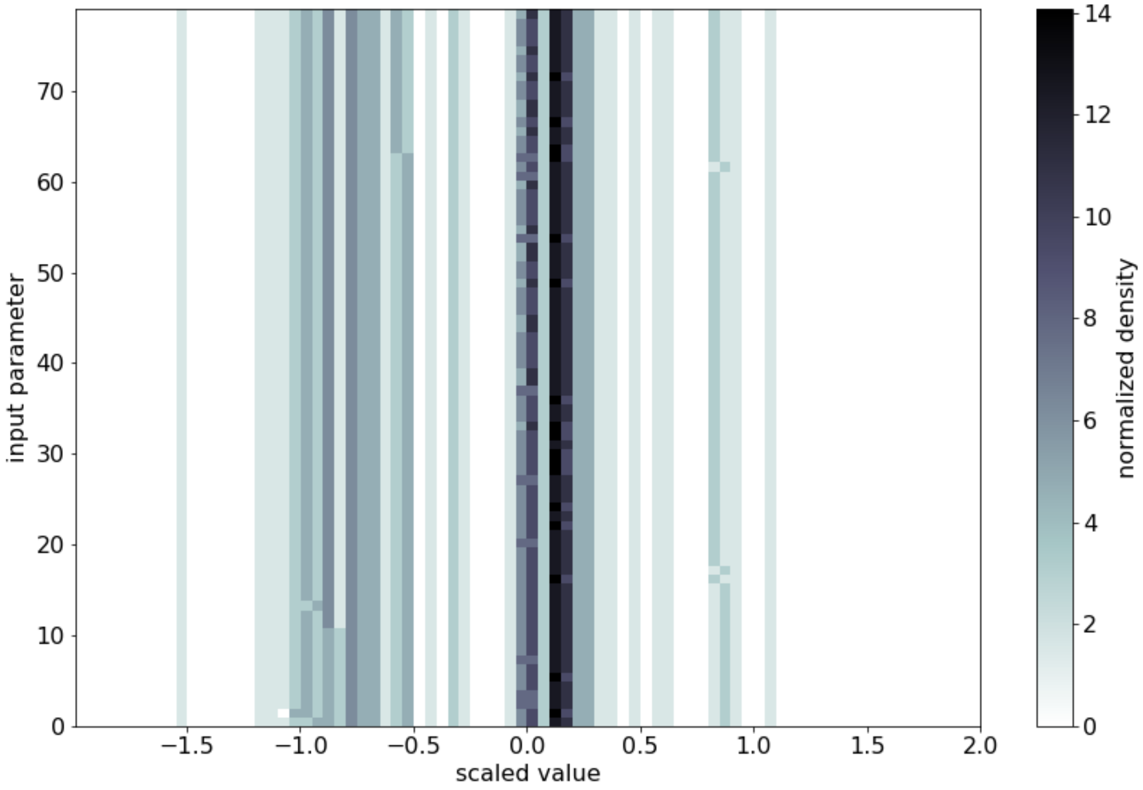
- Simplifications
  - Aggregate by sector: sum current across magnets in a sector (80 inputs/outputs)
  - Aggregate by magnet type in sectors: sum current across each magnet type in a sector (320 inputs/outputs)
  - Consider magnet current or temperature

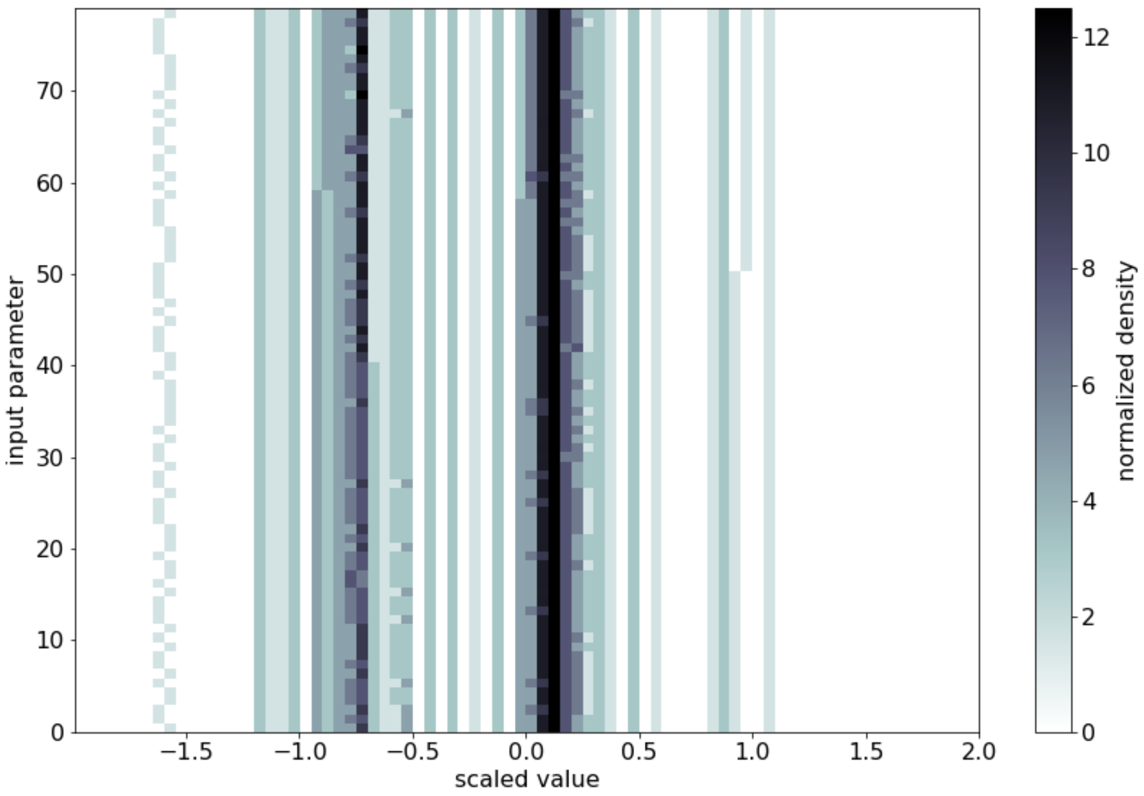




Aggregating by Sector

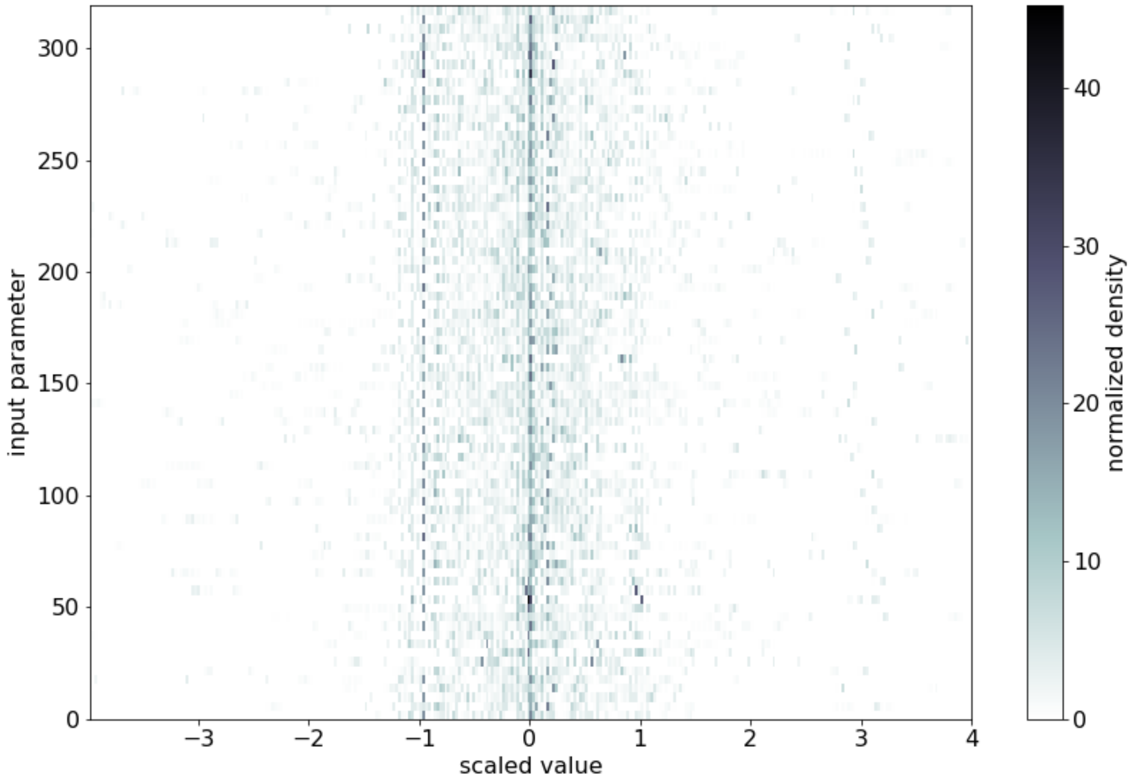
- Reference data (left) used for training and validation and test data (right) with known and unknown anomalies
- Some clear visual differences but datasets are qualitatively similar
- Differences between sectors are subtle at best, may be difficult to isolate anomalous sector

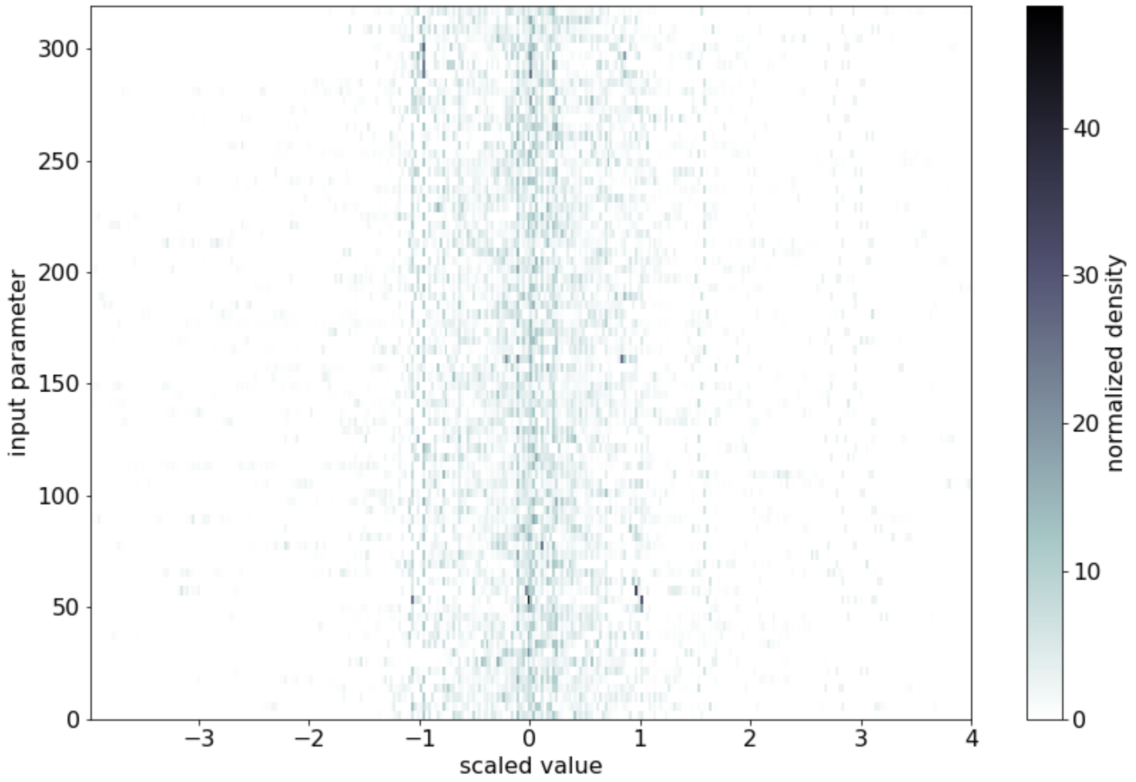




Aggregating by Magnet Type in Sector

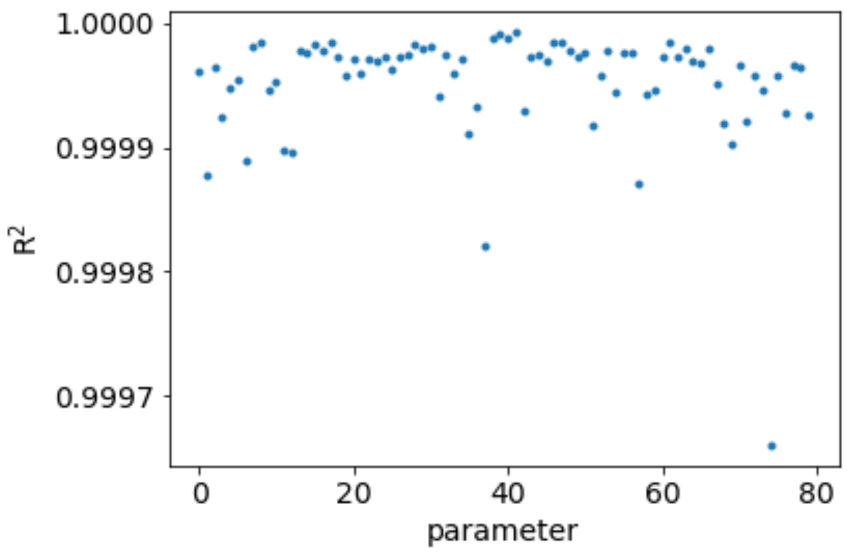
- Reference data (left) used for training and validation and test data (right) with known and unknown anomalies
- Some clear visual differences but datasets are qualitatively similar
- Differences between magnet types and sectors are much more clear



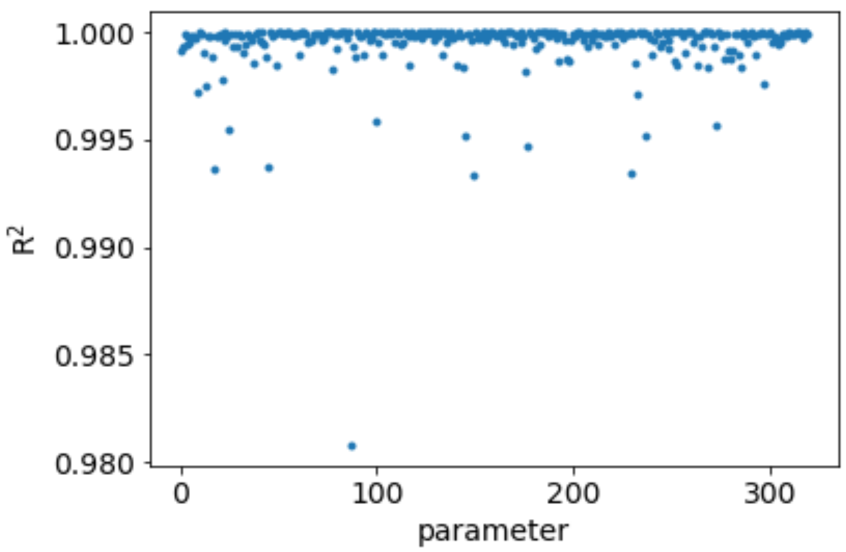


Autoencoder performance for two studies

- Prediction results for the validation data
  - Data aggregated by sector
  - Layers: 60, 40, 20, 10, 8, 2, 8, 10, 20, 40, 60
  - relu activation functions
  - Gaussian noise for regularization

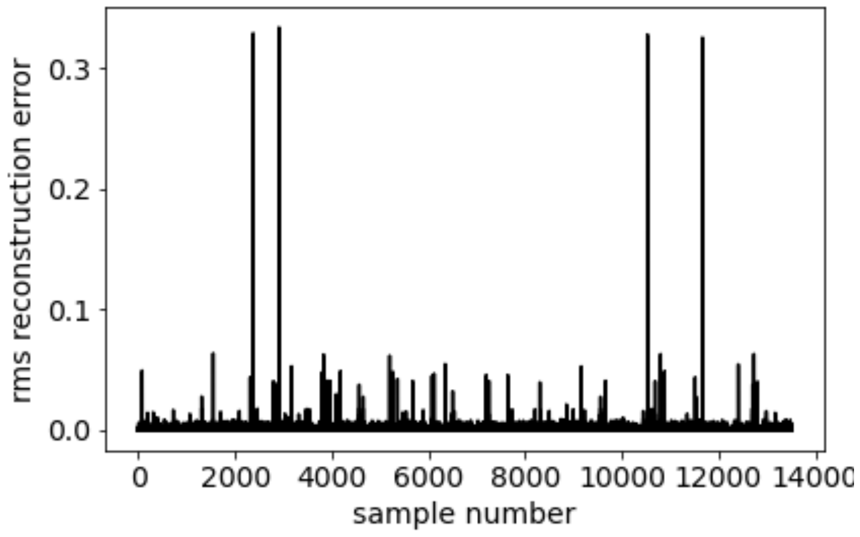


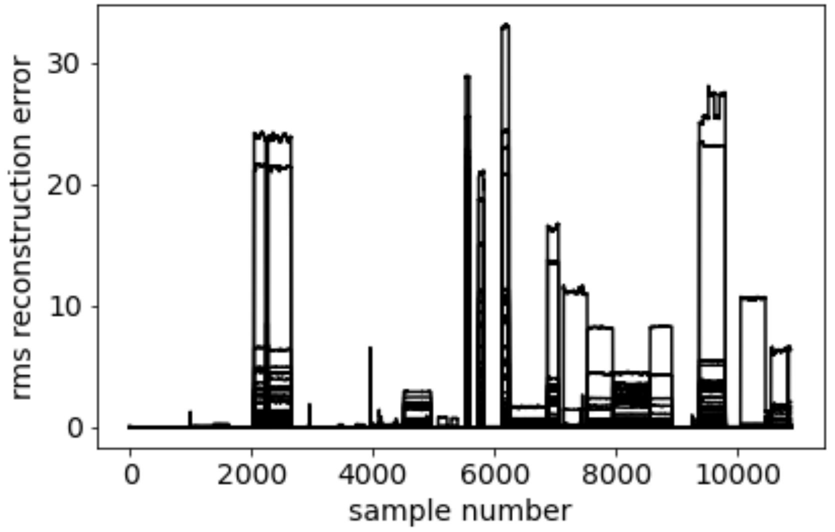
- Prediction results for the validation data
  - Data aggregated by magnet type in a sector
  - Layers : 160, 120, 80, 60, 40, 30, 20, 10, 20, 30, 40, 60, 80, 120, 160
  - relu activation functions
  - Gaussian noise for regularization



Reconstruction of reference data and test data (by sector)

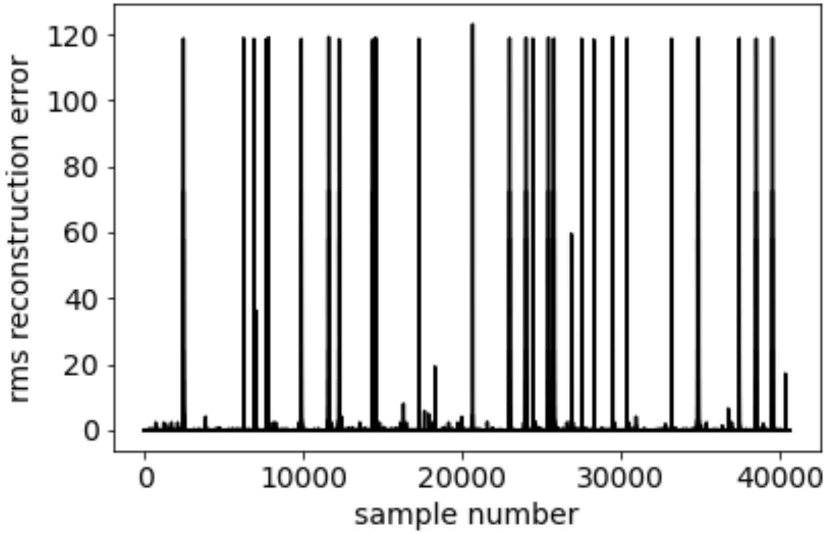
- Train autoencoder on reference data (no faults)
  - Set reconstruction error threshold based on reference data performance
  - Reconstruction threshold of 0.01 results in 0.05% of the reference data being flagged as faulty
- Test autoencoder on fault data
  - Note the actual time when the magnet faults is not in the dataset
  - Reconstruction threshold of 0.01 results in 41% of test data being flagged as faulty

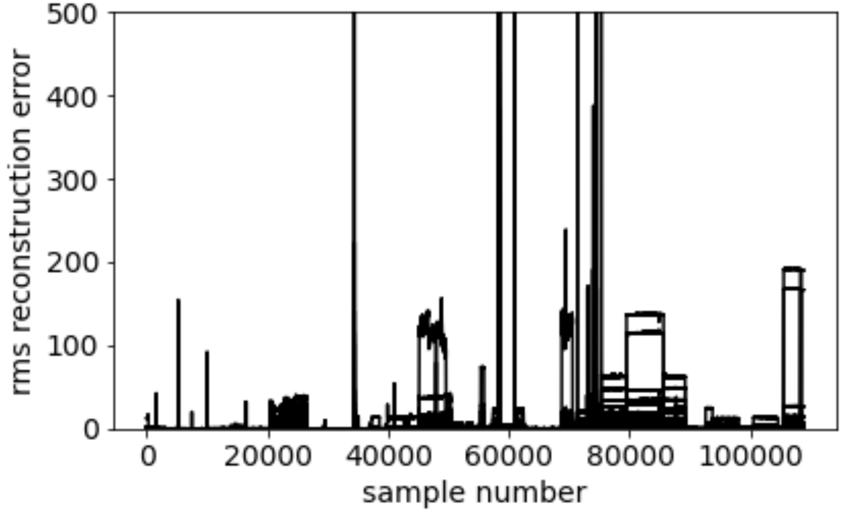




Reconstruction of reference data and test data (by type in sector)

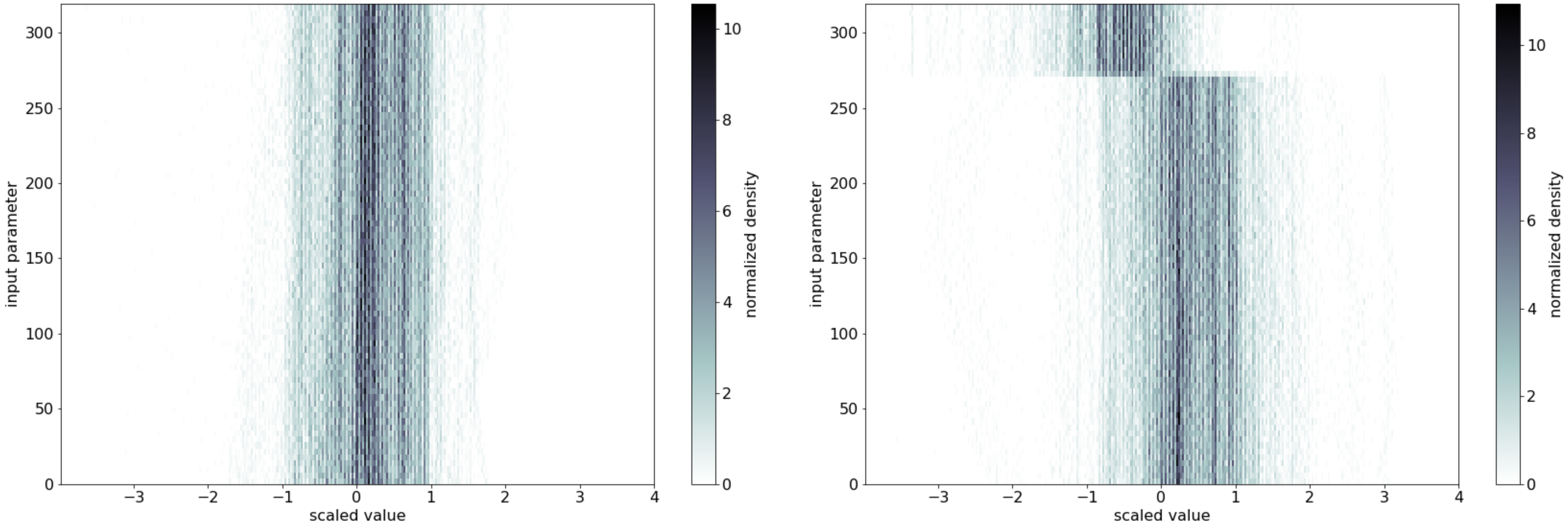
- Train autoencoder on reference data (no faults)
  - Set reconstruction error threshold based on reference data performance
  - Reconstruction threshold of 0.1 results in 0.04% of the reference data being flagged as faulty
- Test autoencoder on fault data
  - Note the actual time when the magnet faults is not in the dataset
  - Reconstruction threshold of 0.1 results in 30% of test data being flagged as faulty





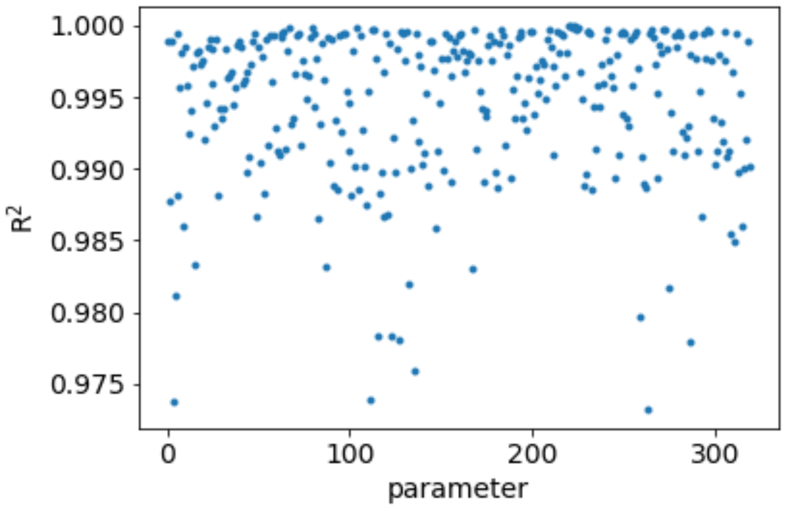
Reconstruction tests for magnet temperature

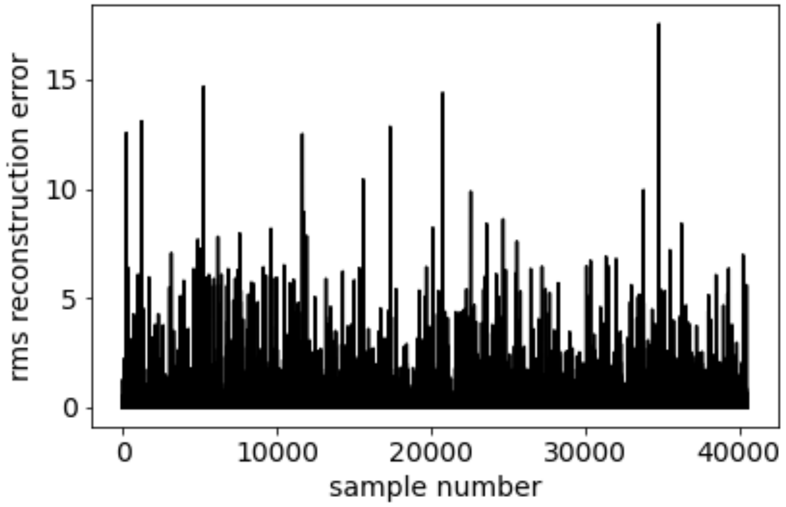
- Reference data (left) used for training and validation and test data (right) with known and unknown anomalies
- Some clear visual differences but datasets are qualitatively similar



Reconstruction of reference data and test data (by type in sector)

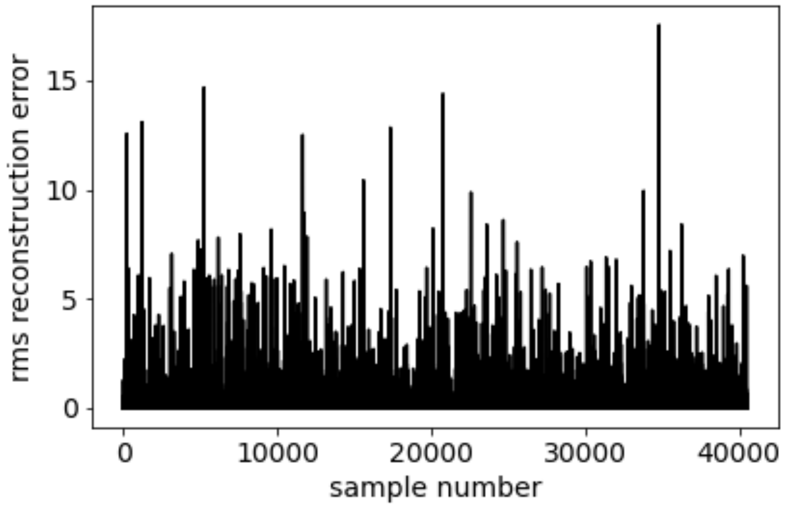
- Train autoencoder on reference data (no faults)
  - Set reconstruction error threshold based on reference data performance
  - Reconstruction threshold of 0.1 results in 0.2% of the reference data being flagged as faulty

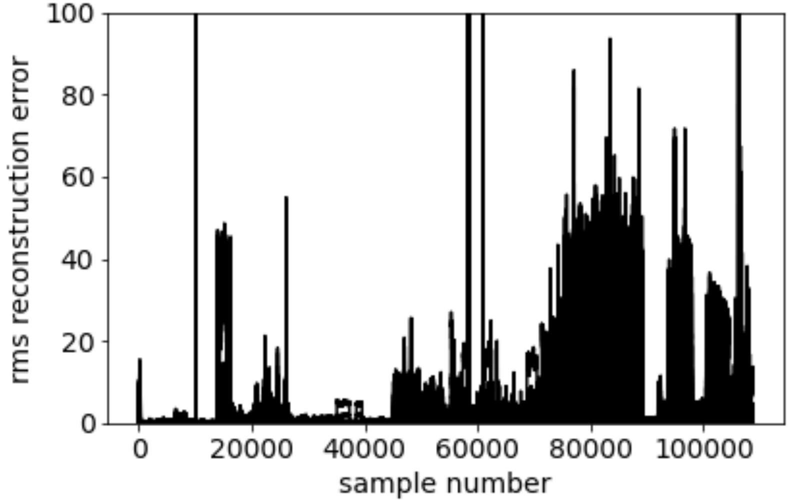




Reconstruction of reference data and test data (by type in sector)

- Train autoencoder on reference data (no faults)
  - Set reconstruction error threshold based on reference data performance
  - Reconstruction threshold of 0.1 results in 0.2% of the reference data being flagged as faulty
- Test autoencoder on fault data
  - Note the actual time when the magnet faults is not in the dataset
  - Reconstruction threshold of 0.1 results in 27% of test data being flagged as faulty





Conclusions

- Fermilab LINAC:
  - Latent space analysis suggestive that RF parameters are correlated with change in current
  - Reconstruction analysis also suggestive that RF parameters are correlated with change in current
  - Latent space and reconstruction analysis outperform conventional methods for analyzing RF data
  - Reconstruction analysis and latent space analysis are robust to perturbations in architecture and initial weights

- APS Storage Ring
  - Autoencoder successfully identified anomalous data on the whole
    - Sector current (41% / 0.05%)
    - Magnet type in sector current (30% / 0.04%)
    - Magnet type in sector temperature (25% / 0.2%)
  - Identifying faulty sector from data aggregated by sector was unsuccessful
  - Continue to explore temperature data and more detailed current data
  - Combine approaches

Disclaimer

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.