Auto-Encoders for CAN Bus Anomaly Detection in Ground Vehicles

Elena I Novikova, Vu Le, Matvey Yutin, Michael Weber
BAE Systems

2021 AIAA DEFENSE Forum, 14-16 September, 2021
JHU APL Laurel, MD

CAN bus is insecure

- **CAN bus communication is known to be insecure, but is still used in a multitude of vehicles.**
- **CAN bus can be exploited to disrupt vehicle operation**
- **Many data manipulation attacks are undetectable by traditional methods due to their statistical indistinguishability from normal signal behaviors**

Auto-encoder NN is better than simple statistics

- **Autoencoder neural network applications against real-world CAN bus data**
- **Autoencoders capture and compactly encode complex, non-linear relationships between multiple signals**
- **They can catch when one signal is misbehaving in the context of a group of related signals, detecting unusual signal excursions missed by other statistical detectors.**

Not just CAN bus and not just ground vehicles

- **Our approach is not based on the details of CAN communication protocol and not based on the specifics of the signals transmitted**

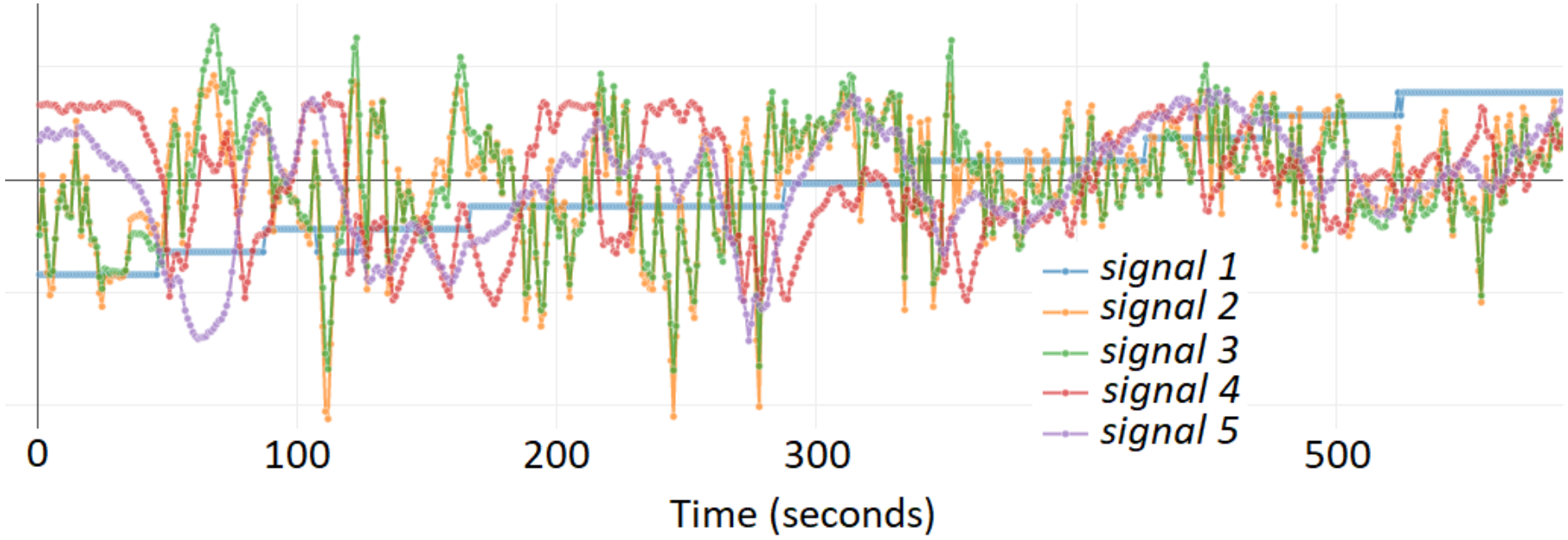
- **We use only the fact that many signals, such as signals carrying physical quantities or indications of state, are inherently correlated in groups of three or more**

What we present today

- **We show:**

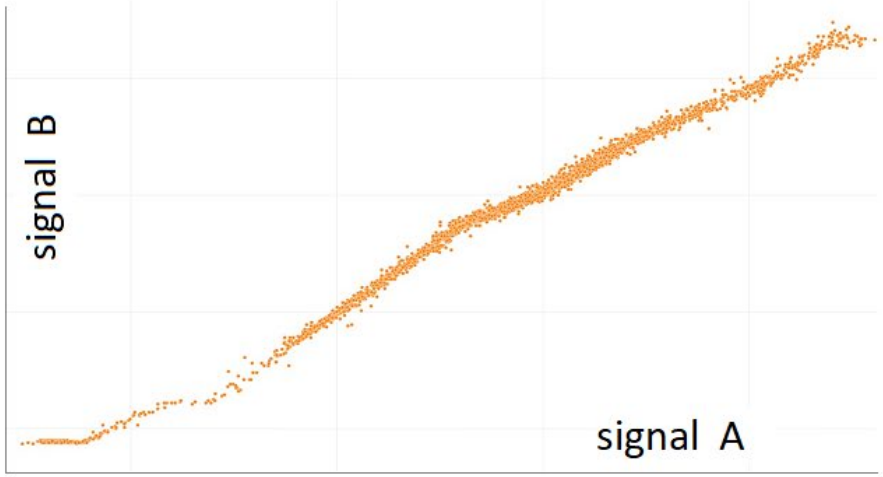
- **how anomaly detectors are built around autoencoders;**
- **how they are designed and trained (our “tricks”);**
- **how they are combined to yield excellent detection performance for signal anomalies**

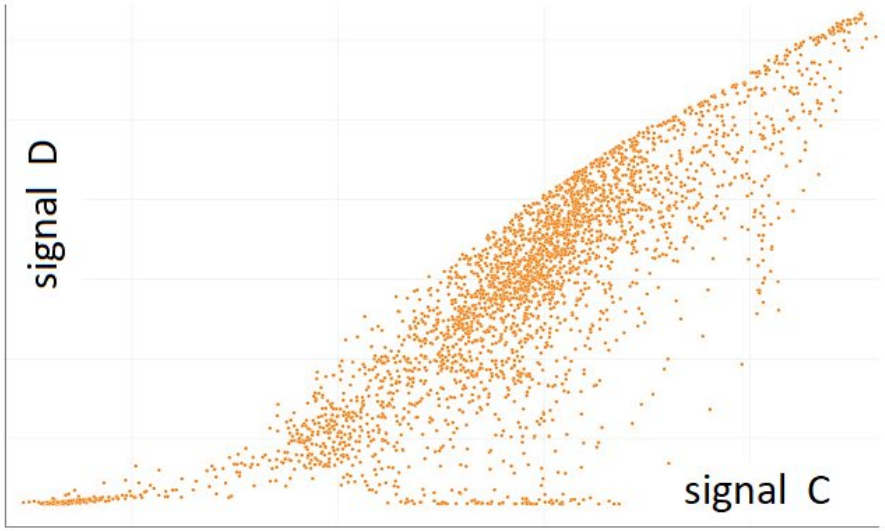
Sample of Signals



Related and unrelated signals

Non-trivial dependency between signals



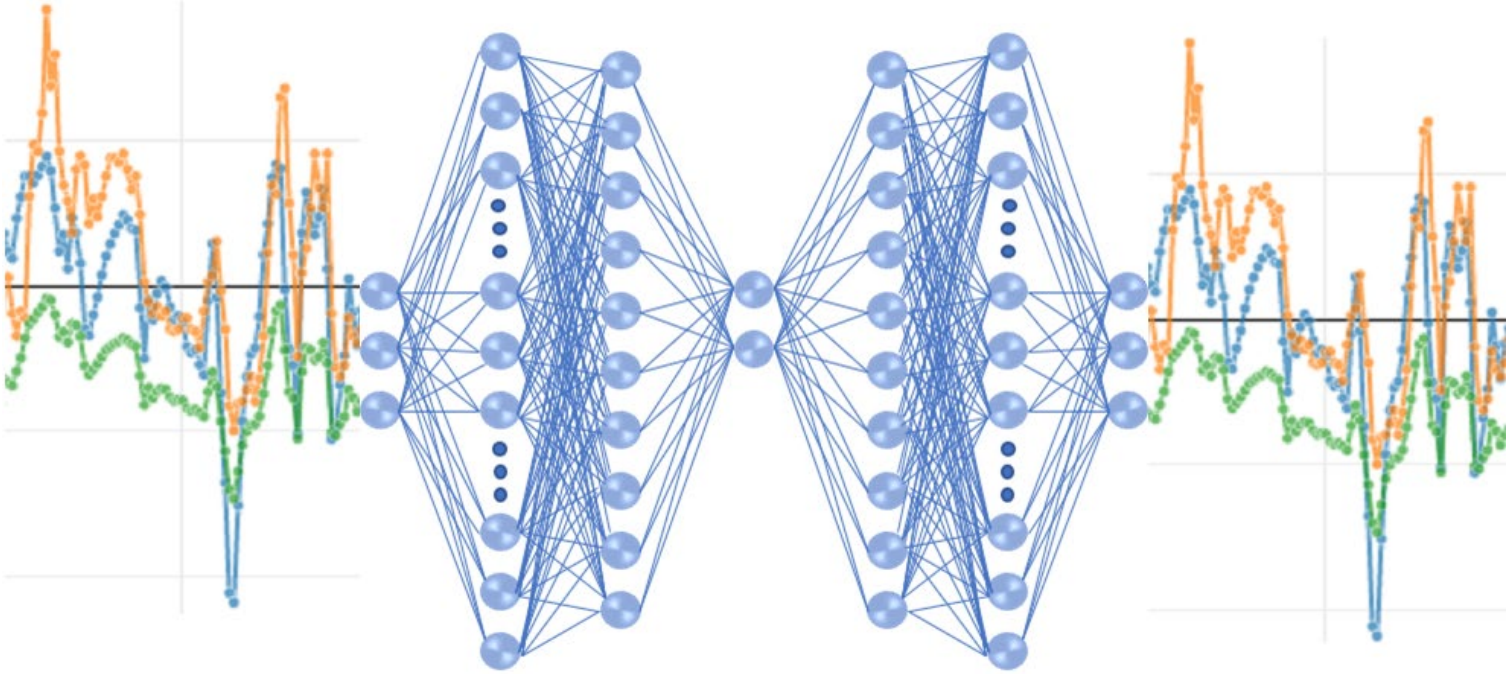


2D relationship graph for related signals

Data

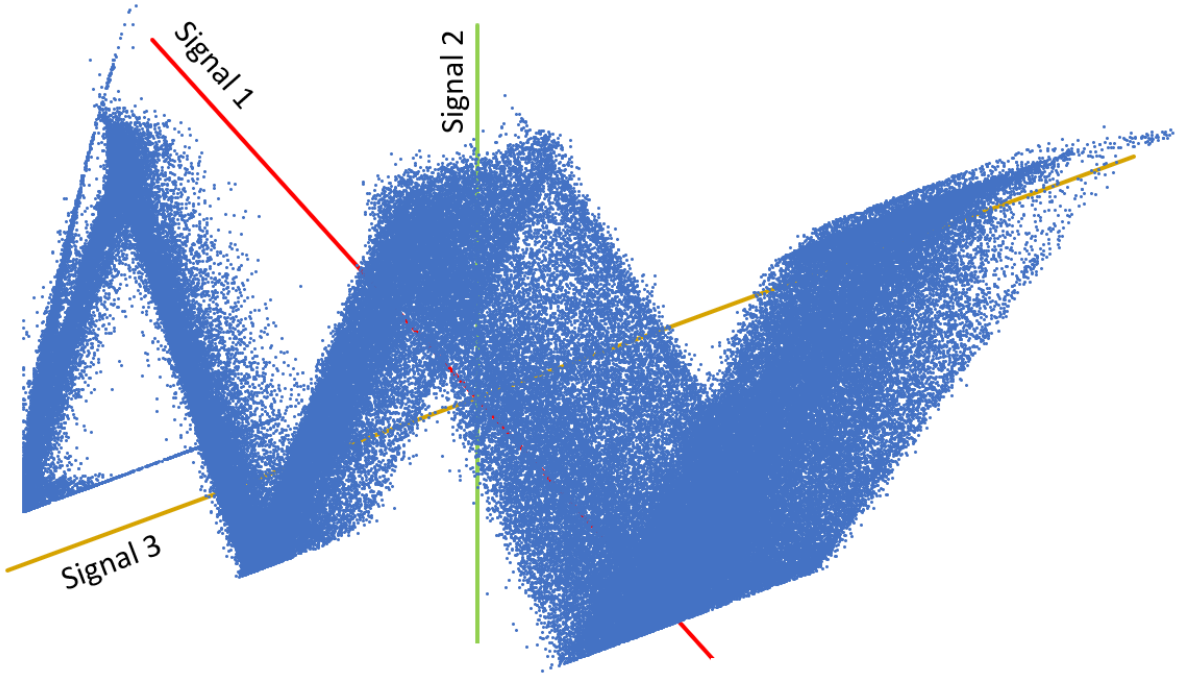
- About a dozen vehicles
- About 22 days of data
- Identified 81 physical signals
- Created 32 groups of 3-signals, ultimately used 29

Auto-encoders for anomaly detection



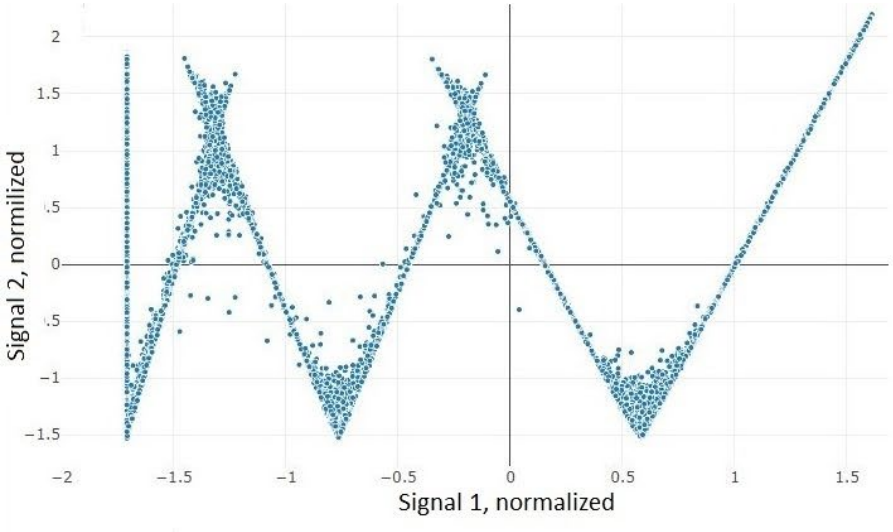
Neural Net working as an Auto-encoder

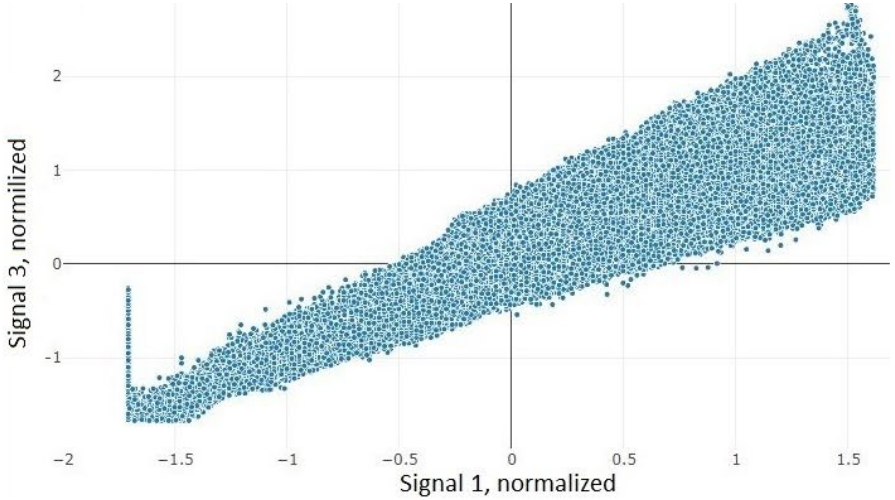
Group #1 data cloud



3D patterns of group #1

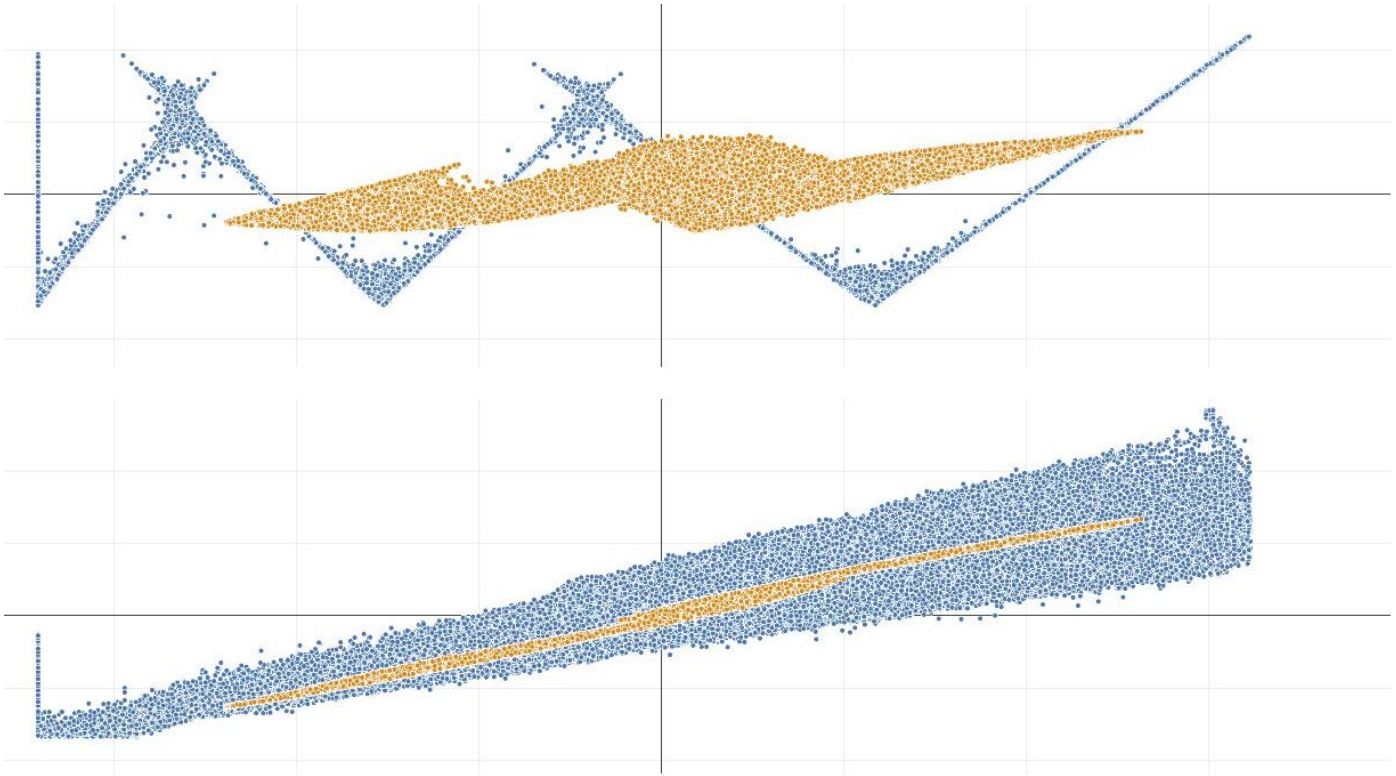
Group #1 data cloud





2D projections of the 3D patterns of group #1

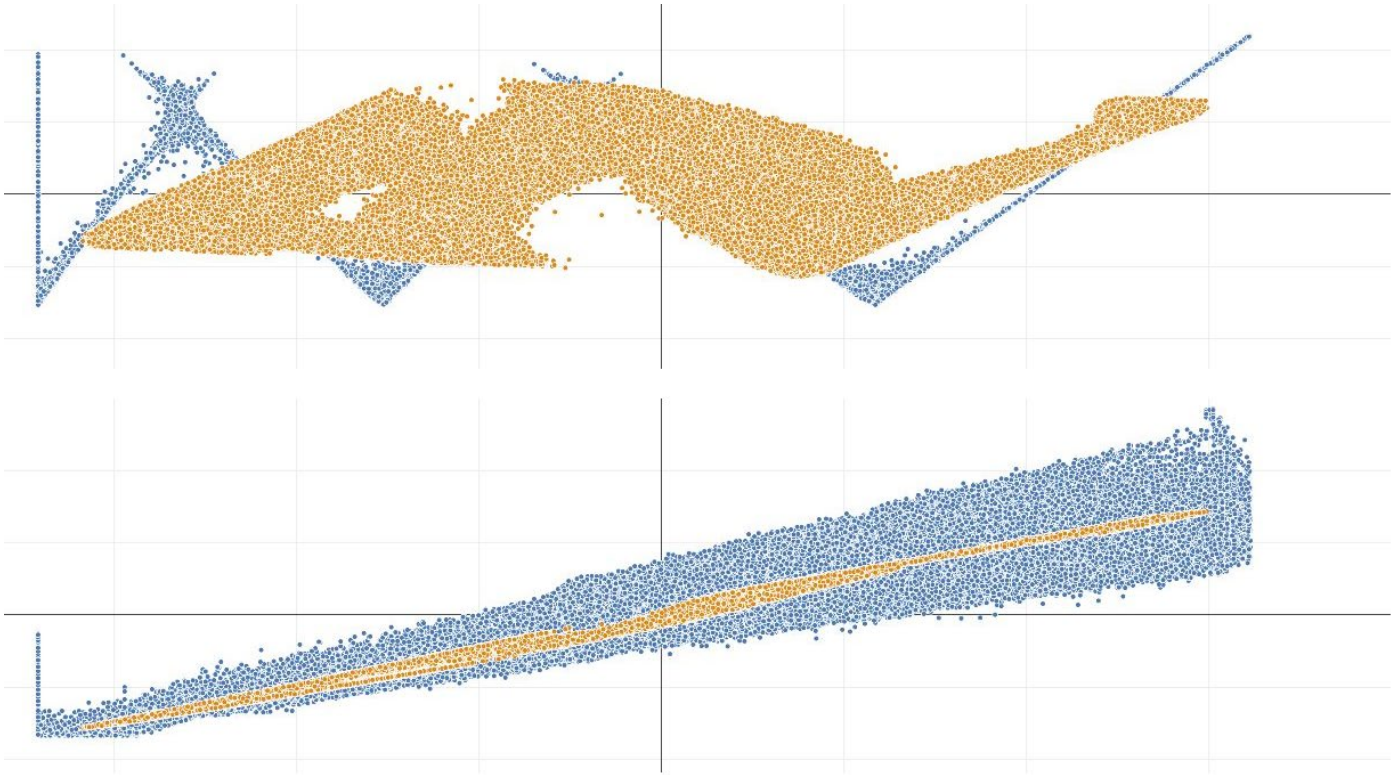
Neural Net 3-2-3: geometry of learning



Training
Neural Net
3-2-3

frame #1

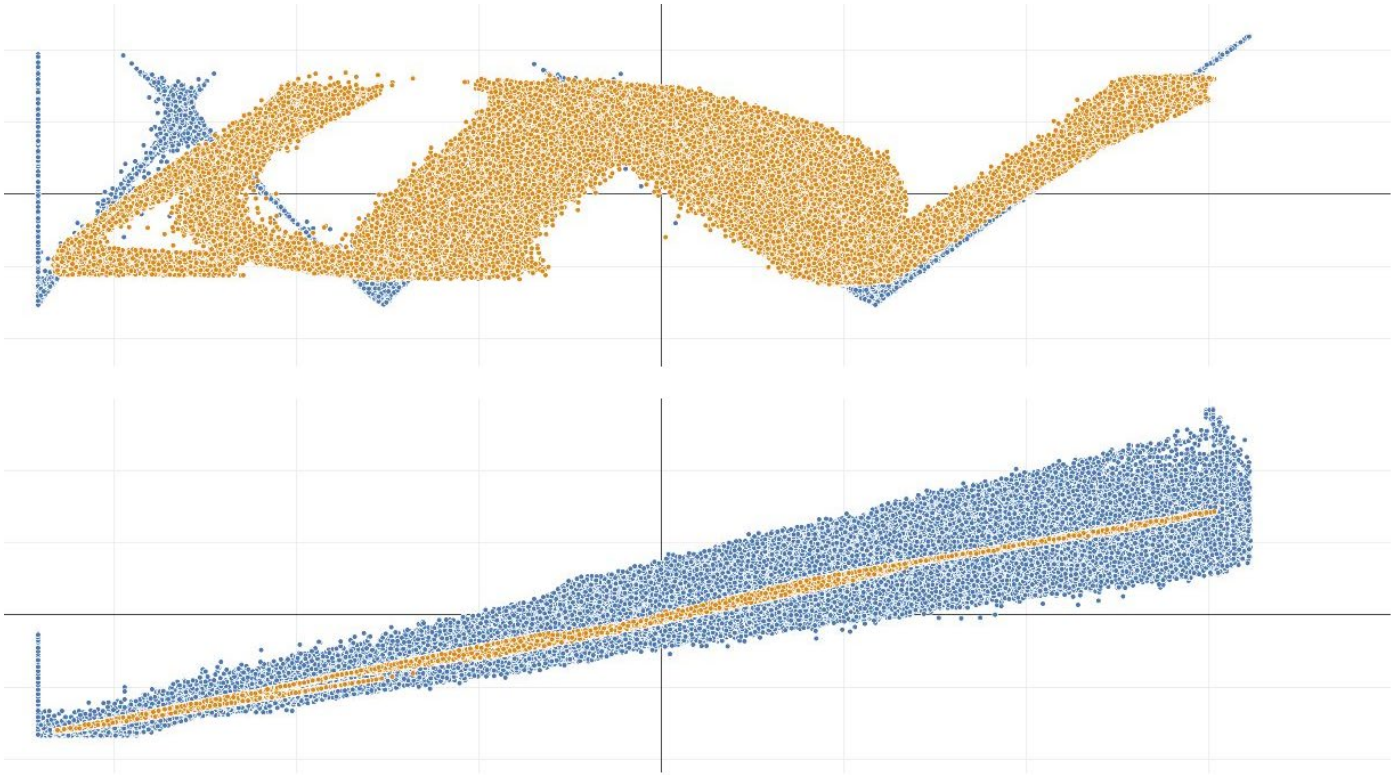
Neural Net 3-2-3: geometry of learning



Training
Neural Net
3-2-3

frame #2

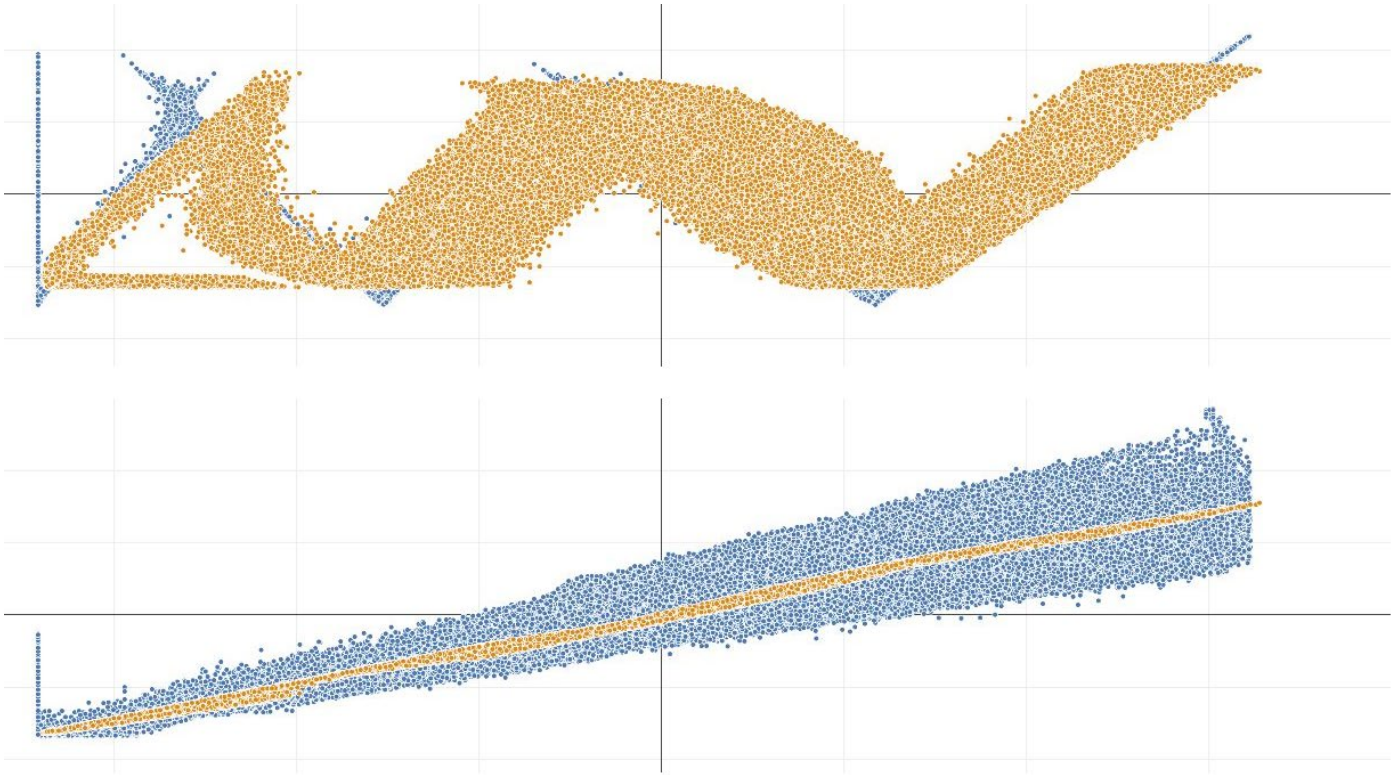
Neural Net 3-2-3: geometry of learning



Training
Neural Net
3-2-3

frame #3

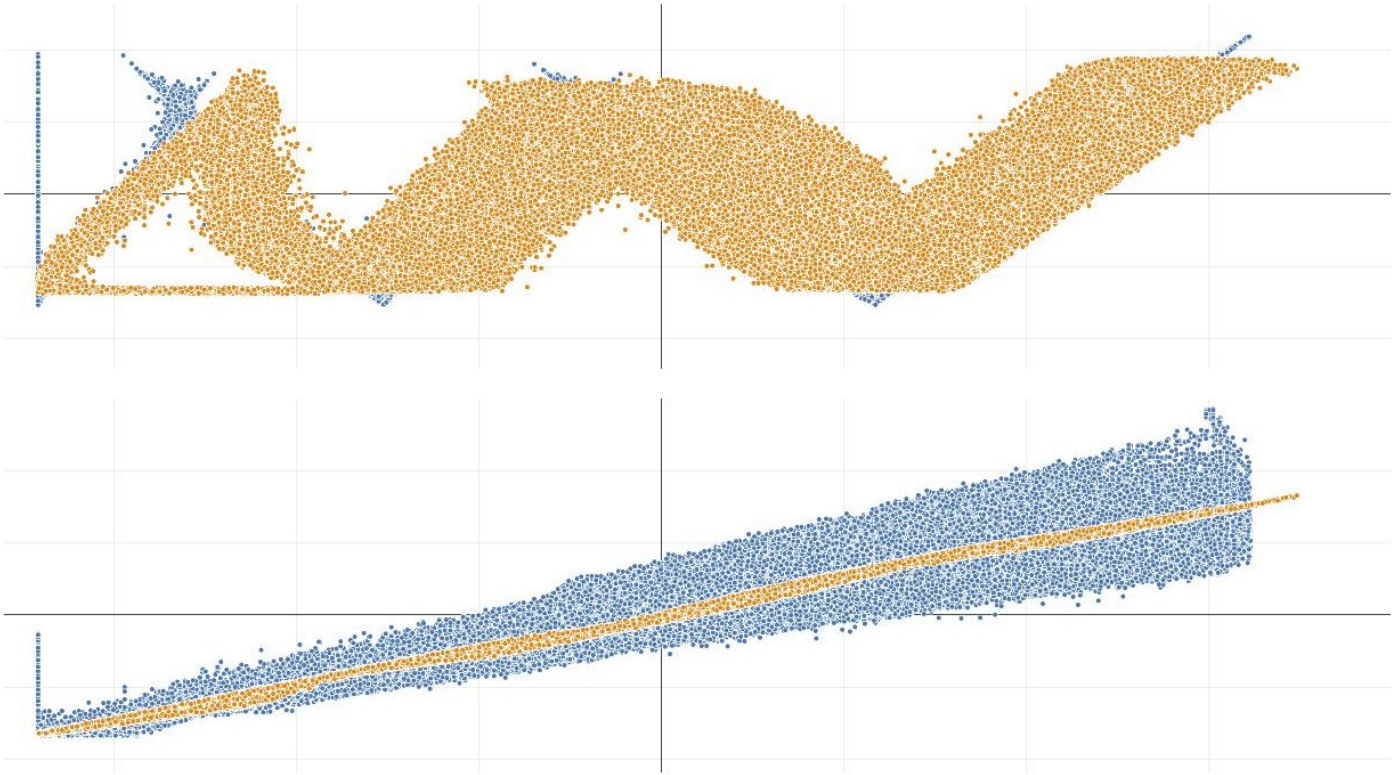
Neural Net 3-2-3: geometry of learning



Training
Neural Net
3-2-3

frame #4

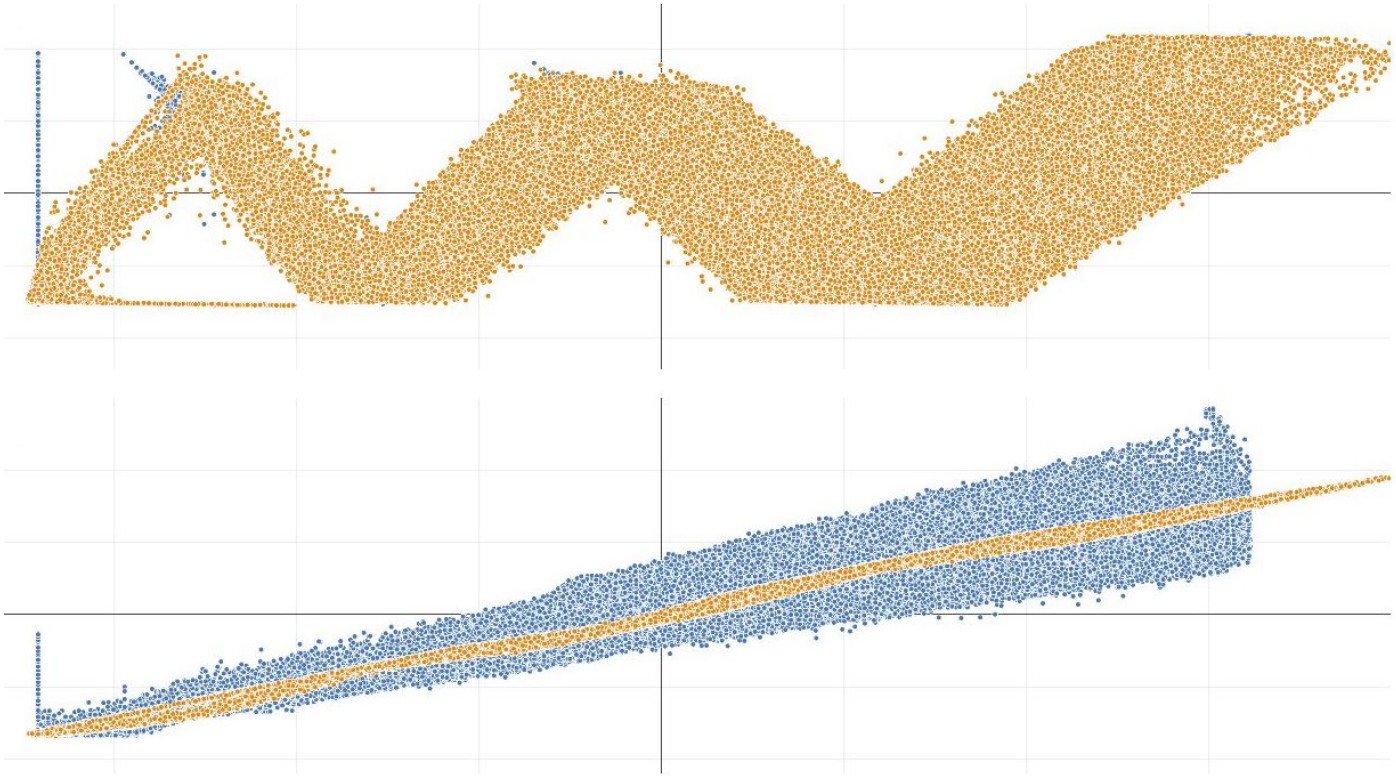
Neural Net 3-2-3: geometry of learning



Training
Neural Net
3-2-3

frame #5

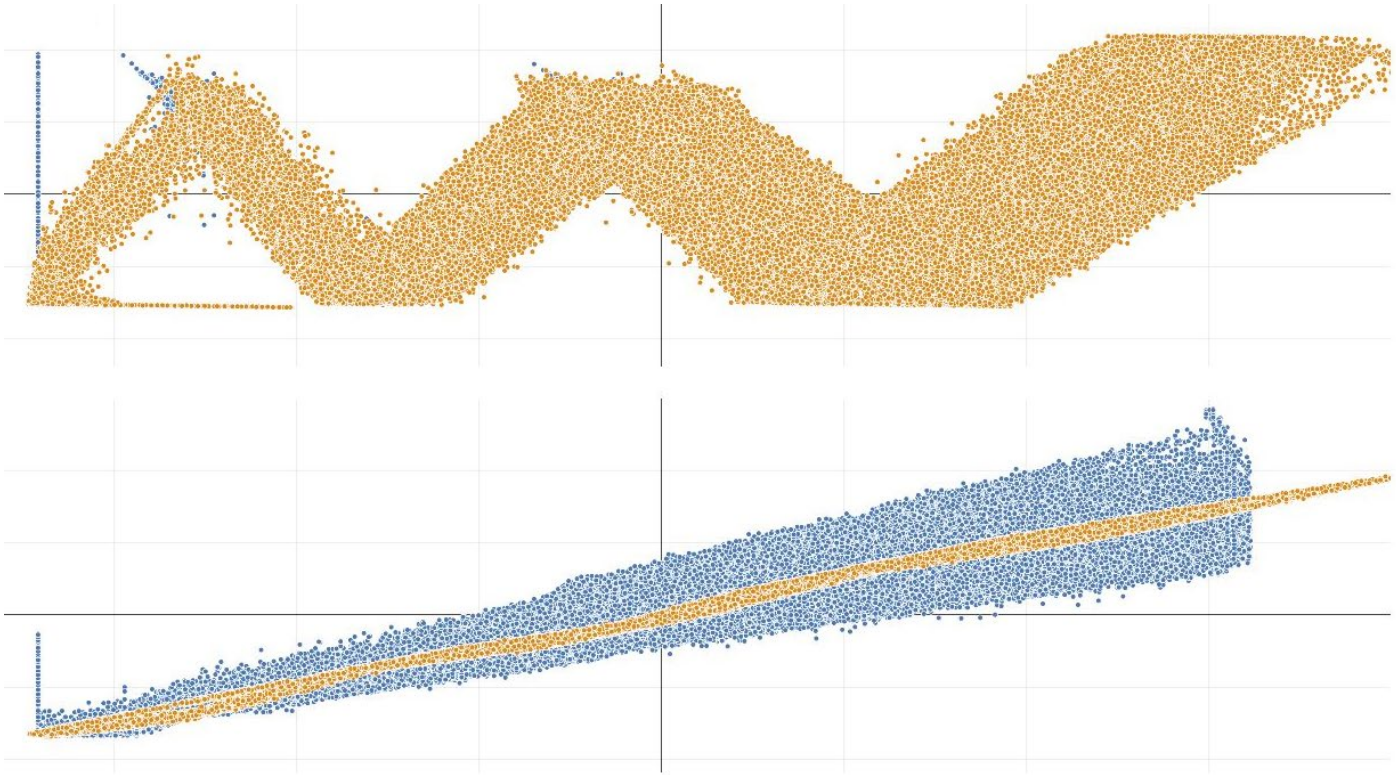
Neural Net 3-2-3: geometry of learning



Training
Neural Net
3-2-3

frame #100

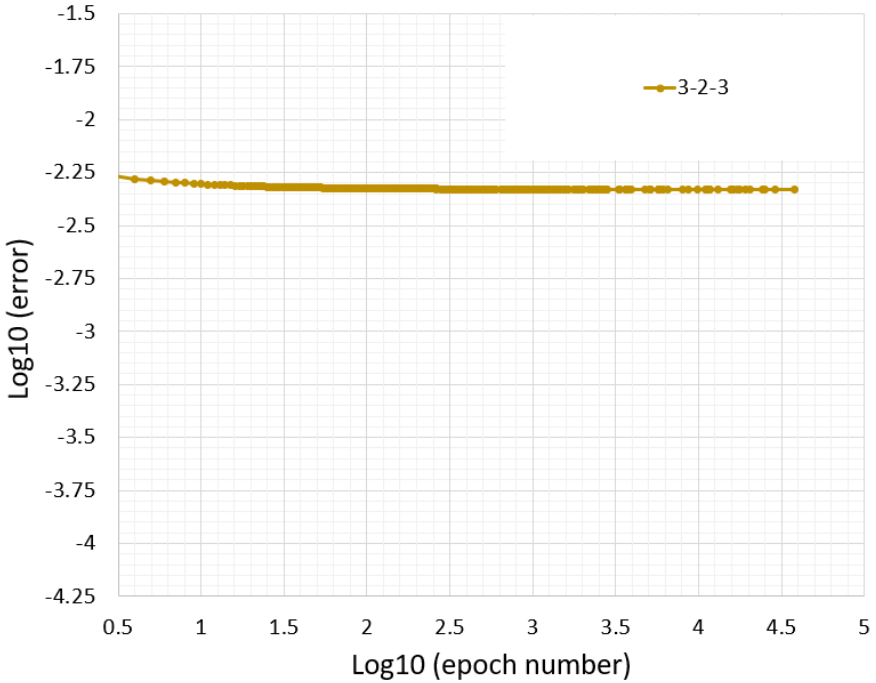
Neural Net 3-2-3: geometry of learning



Training
Neural Net
3-2-3

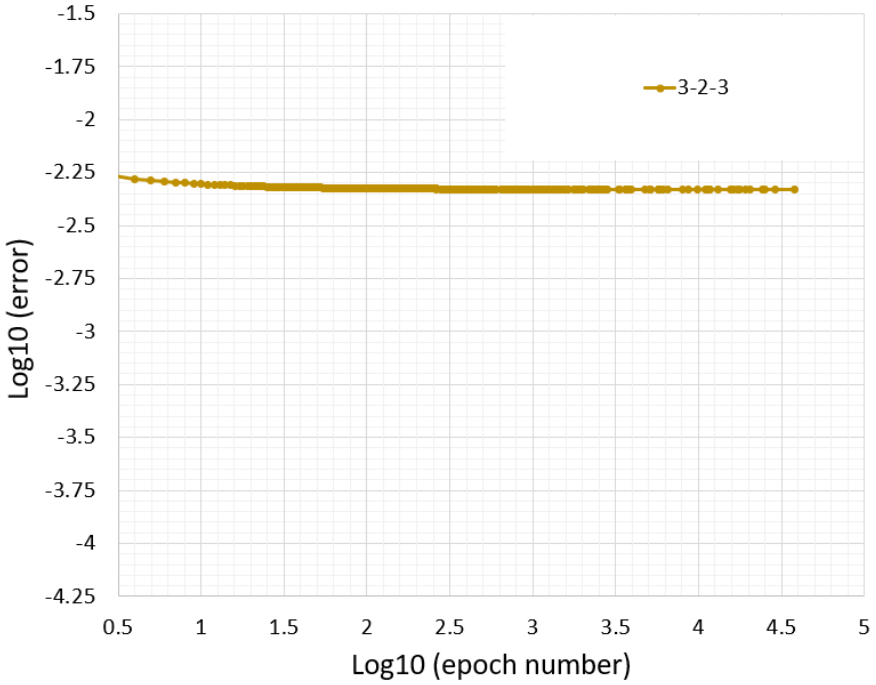
frame #1000

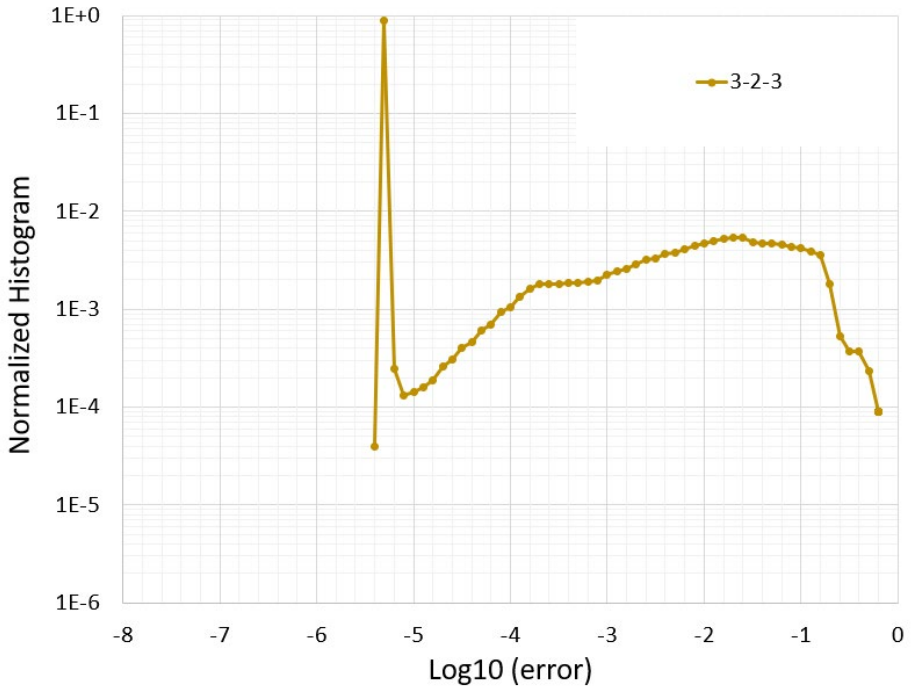
Neural Net 3-2-3



Neural Net 3-2-3: convergence

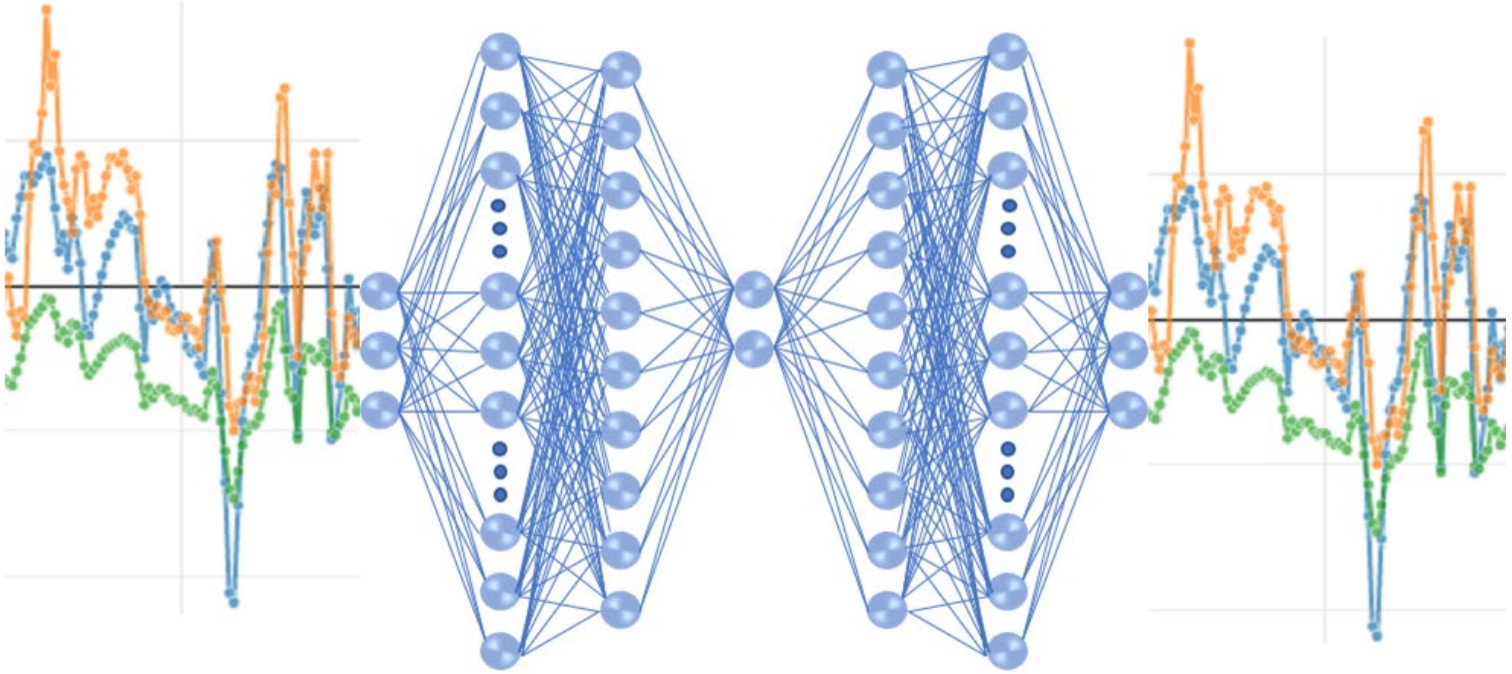
Neural Net 3-2-3





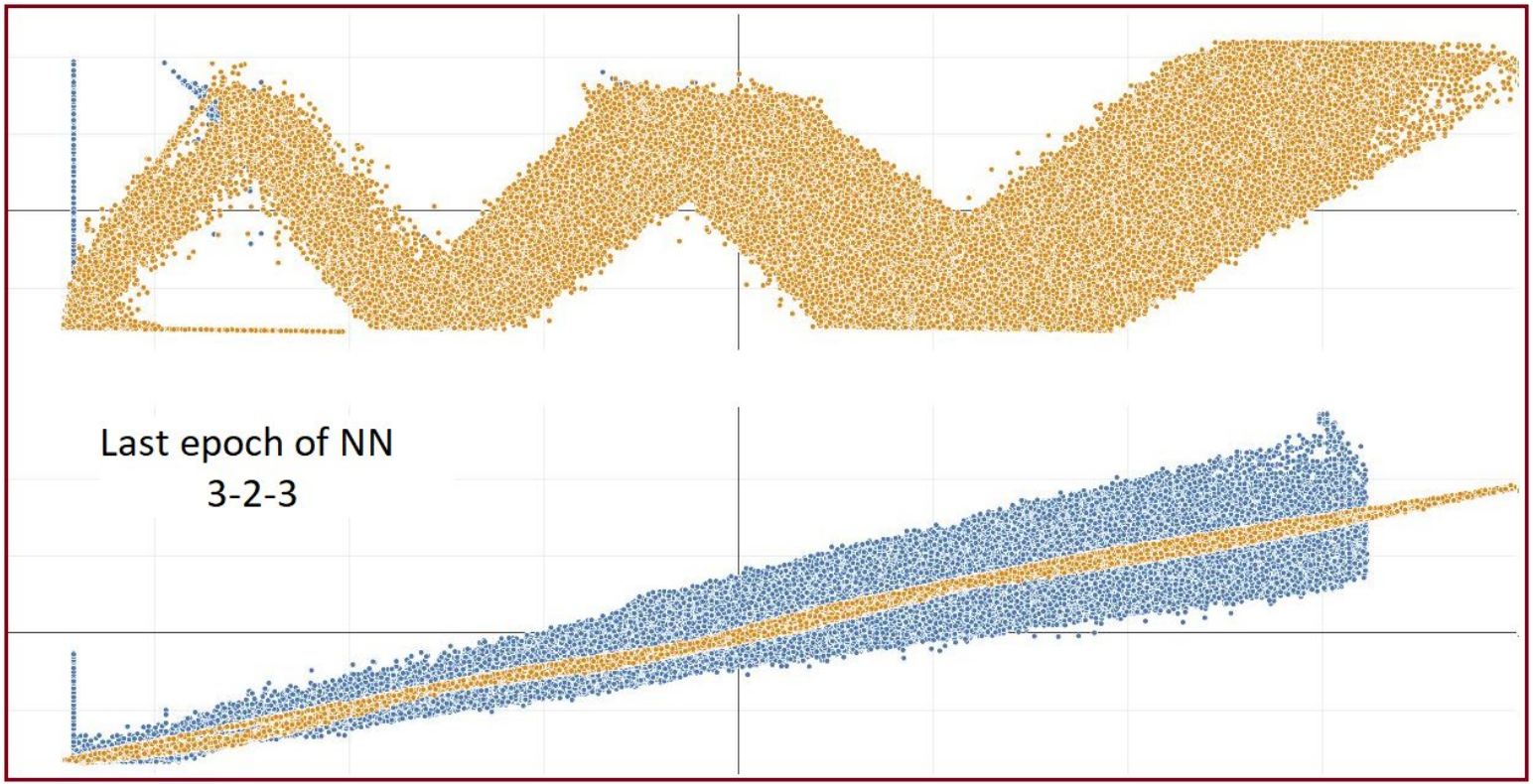
Neural Net 3-2-3: convergence; error distribution

Increasing the number of layers

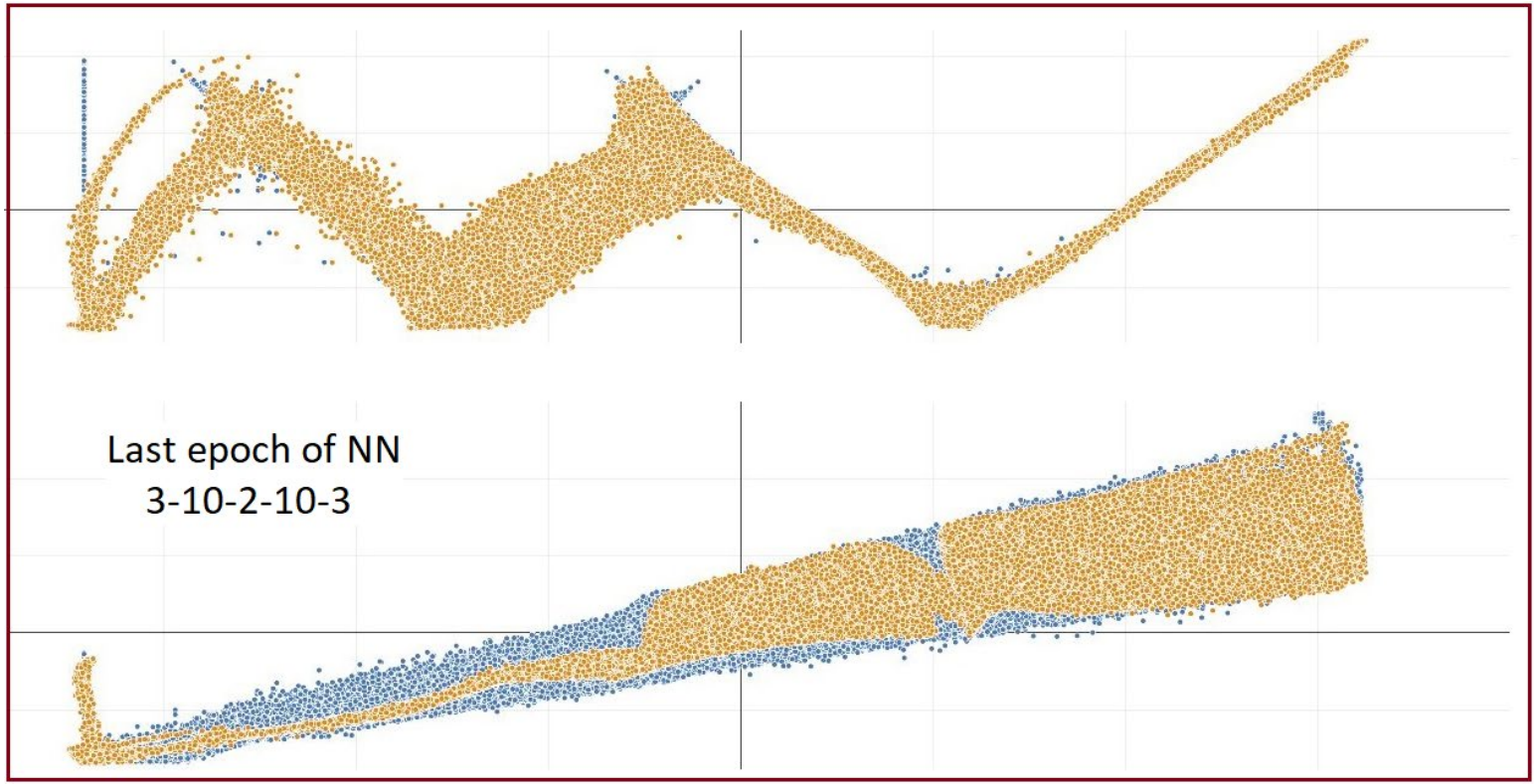


Neural Net working as an Auto-encoder

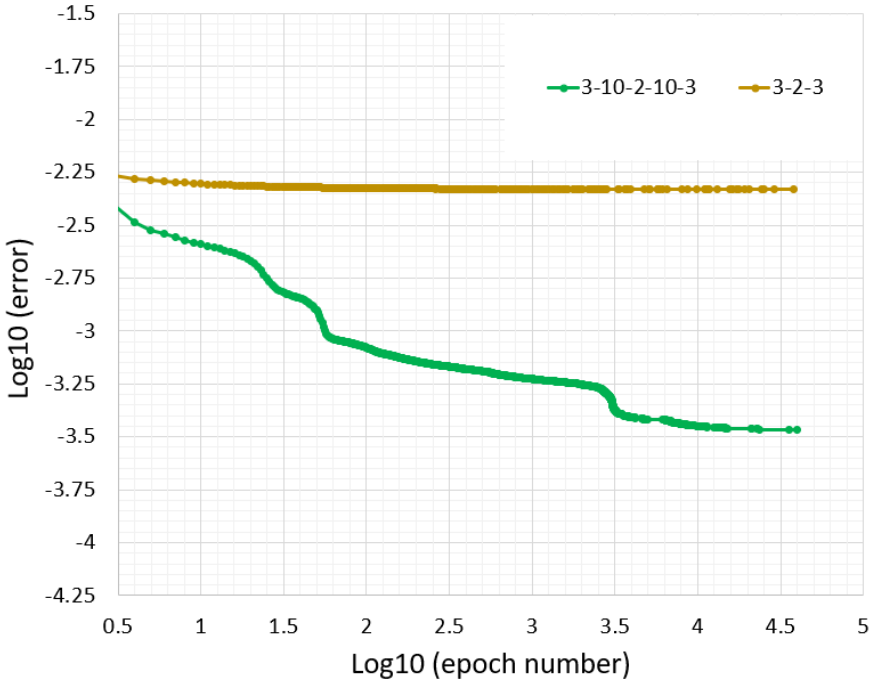
From 3-2-3 to 3-10-2-10-3: geometry of learning

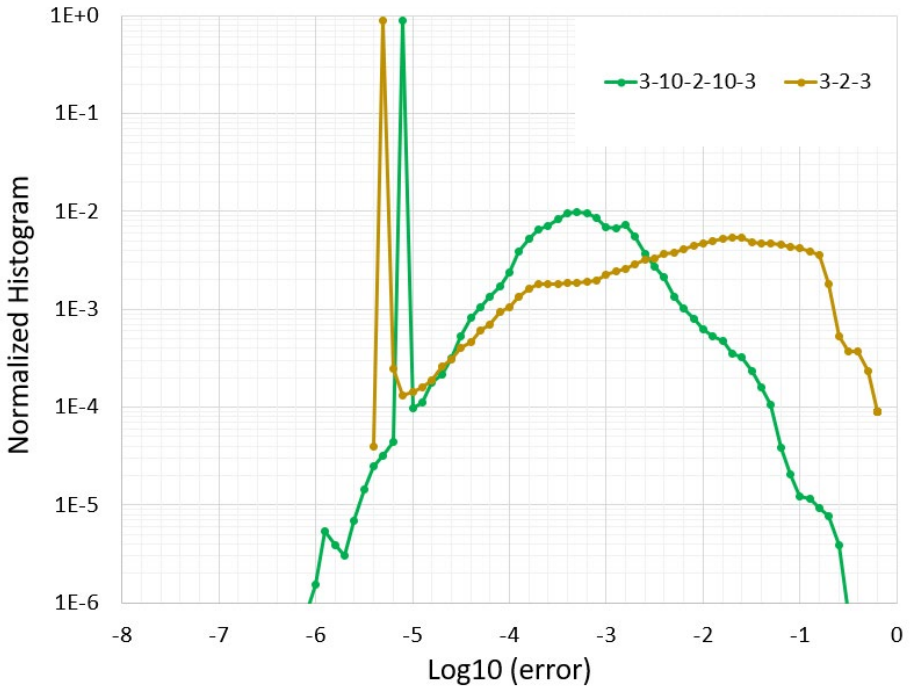


From 3-2-3 to 3-10-2-10-3: geometry of learning



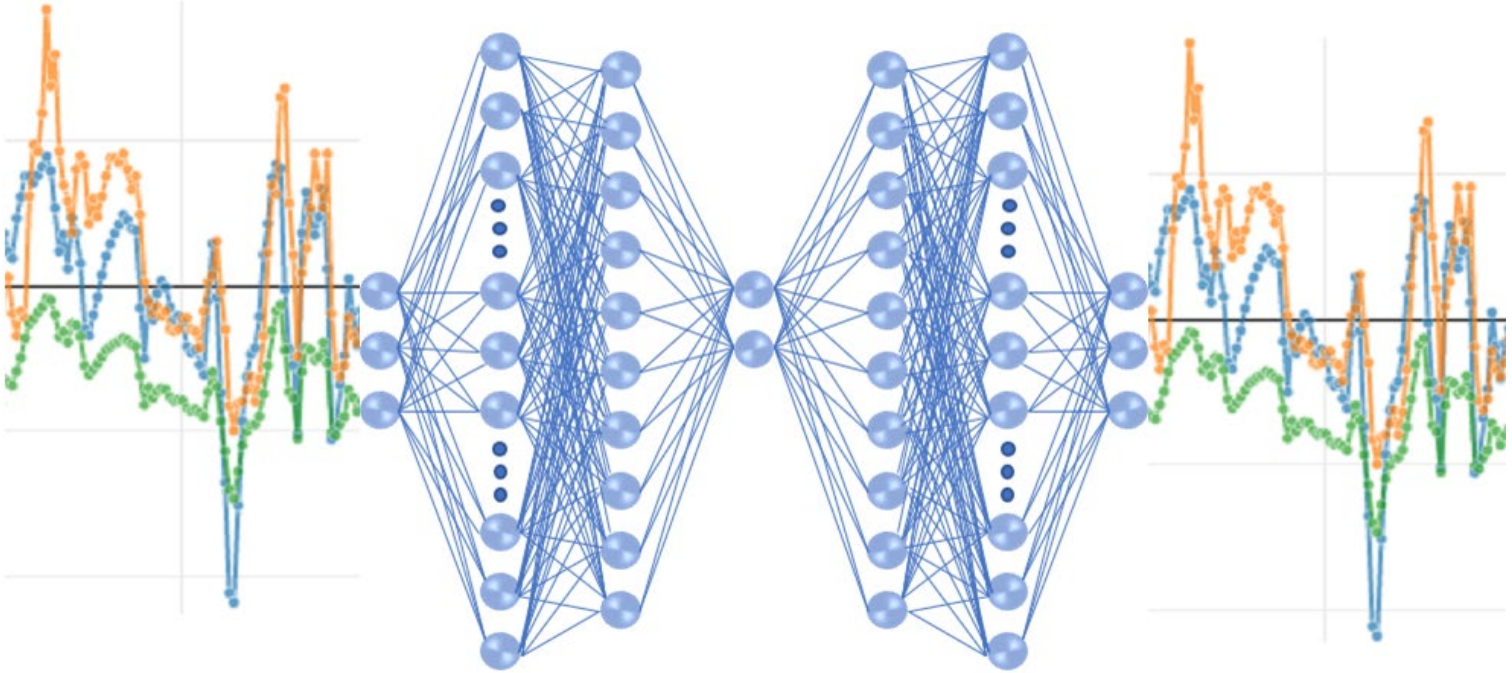
Neural Net 3-10-2-10-3





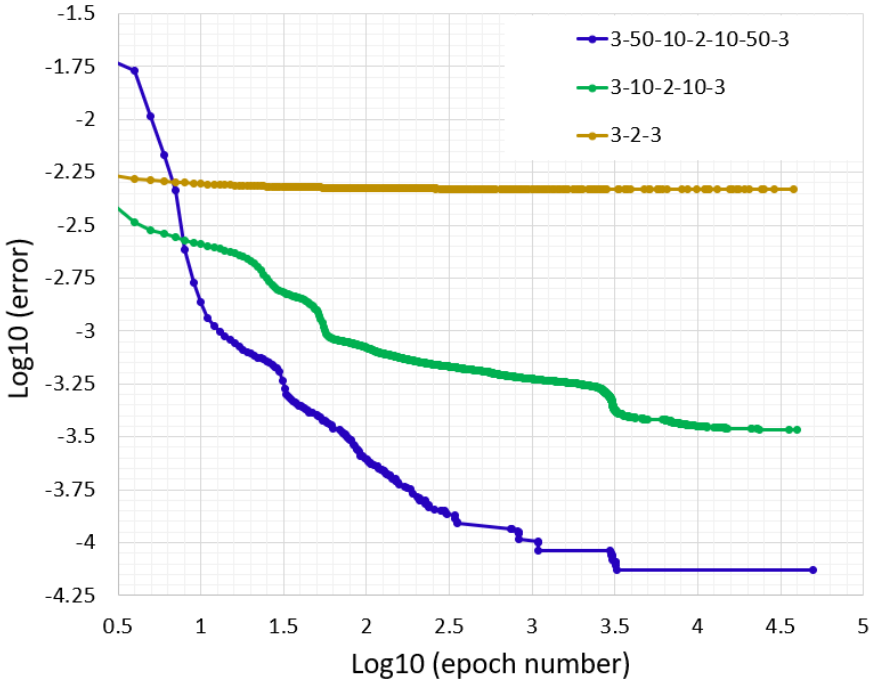
Neural Net 3-10-2-10-3: convergence; error distribution

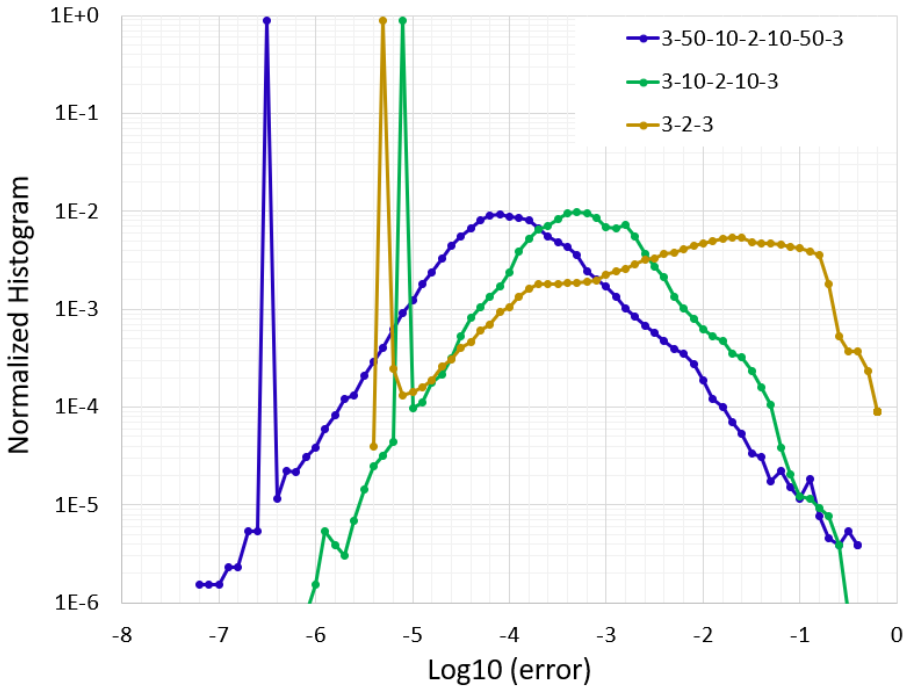
Increasing the number of layers



Neural Net working as an Auto-encoder

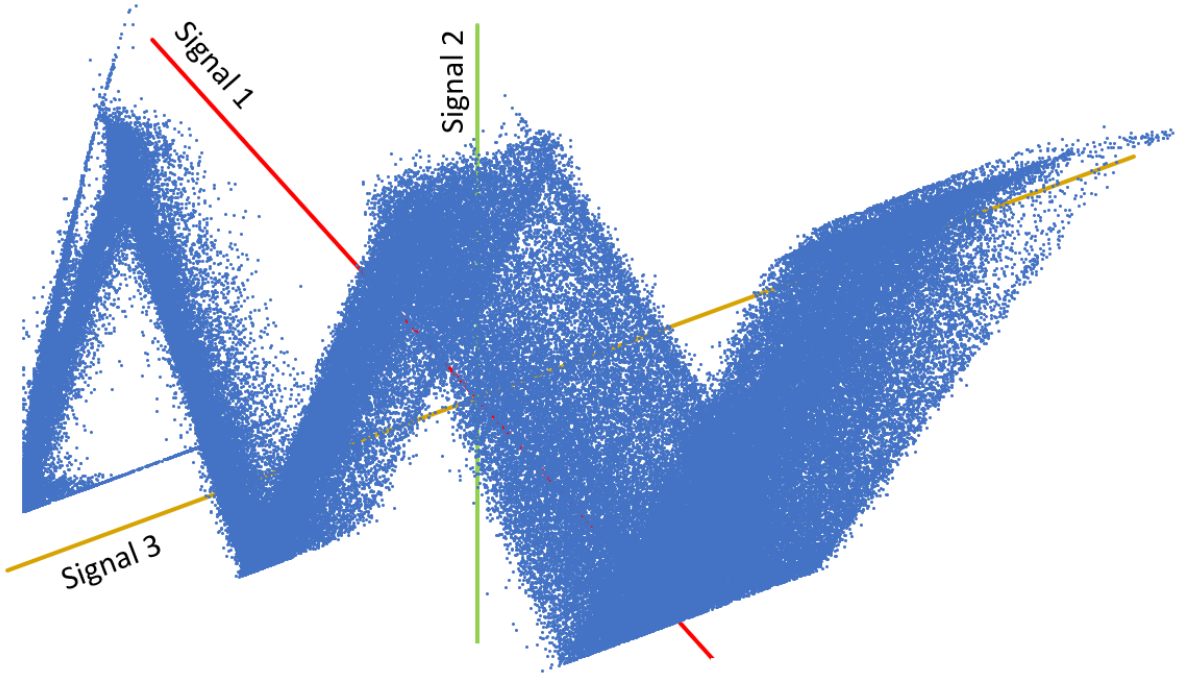
Neural Net 3-50-10-2-10-50-3





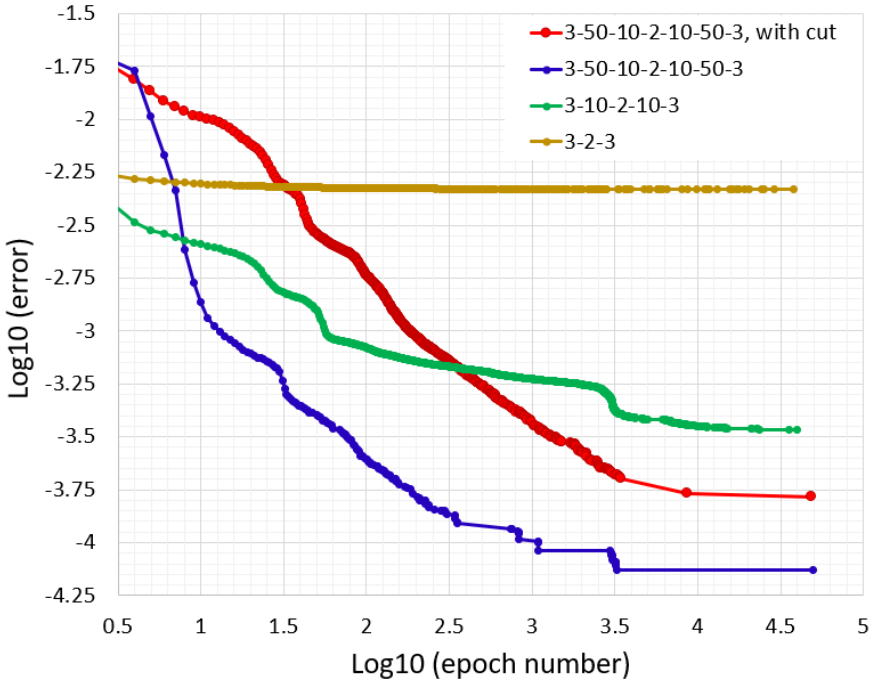
Neural Nets of three configurations

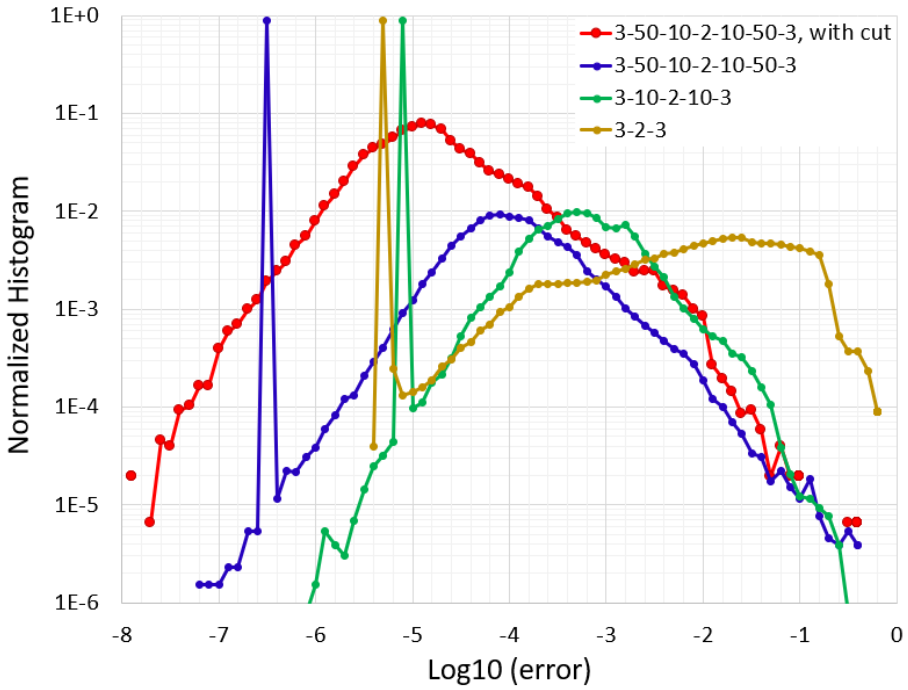
“Bin and cut” and geometry of learning



3D patterns of group #1

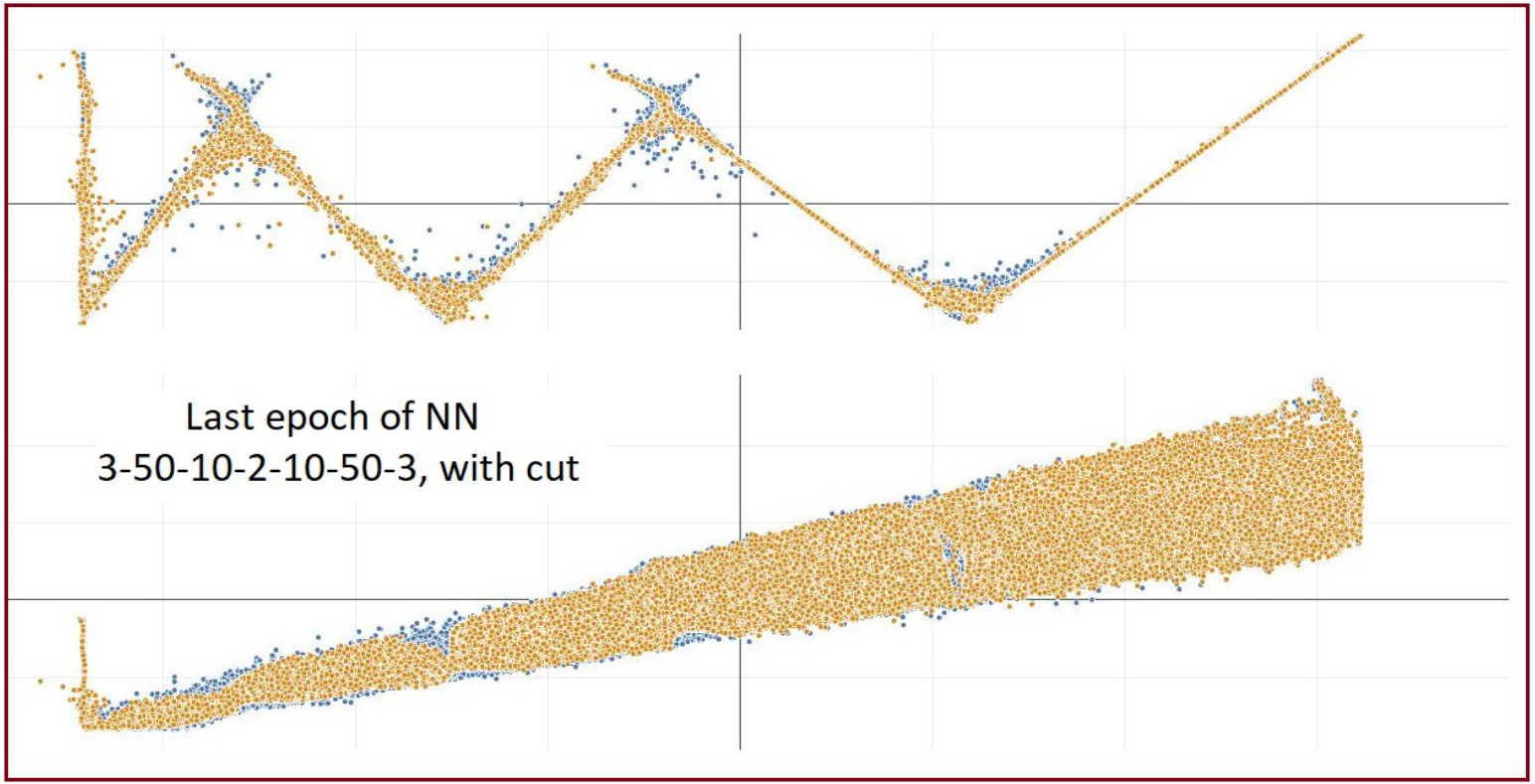
Neural Net 3-50-10-2-10-50-3 with “bin and cut”



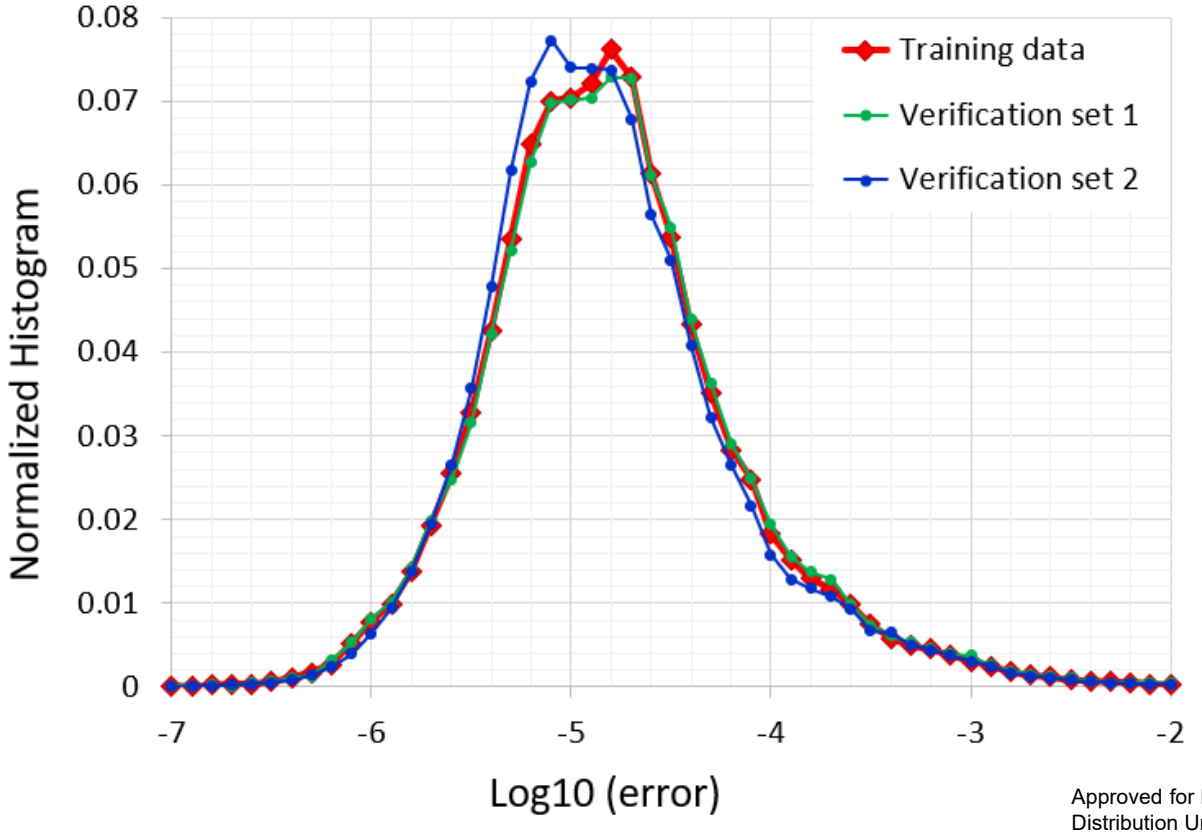


Introducing cuts

Best configuration: Geometry of Learning

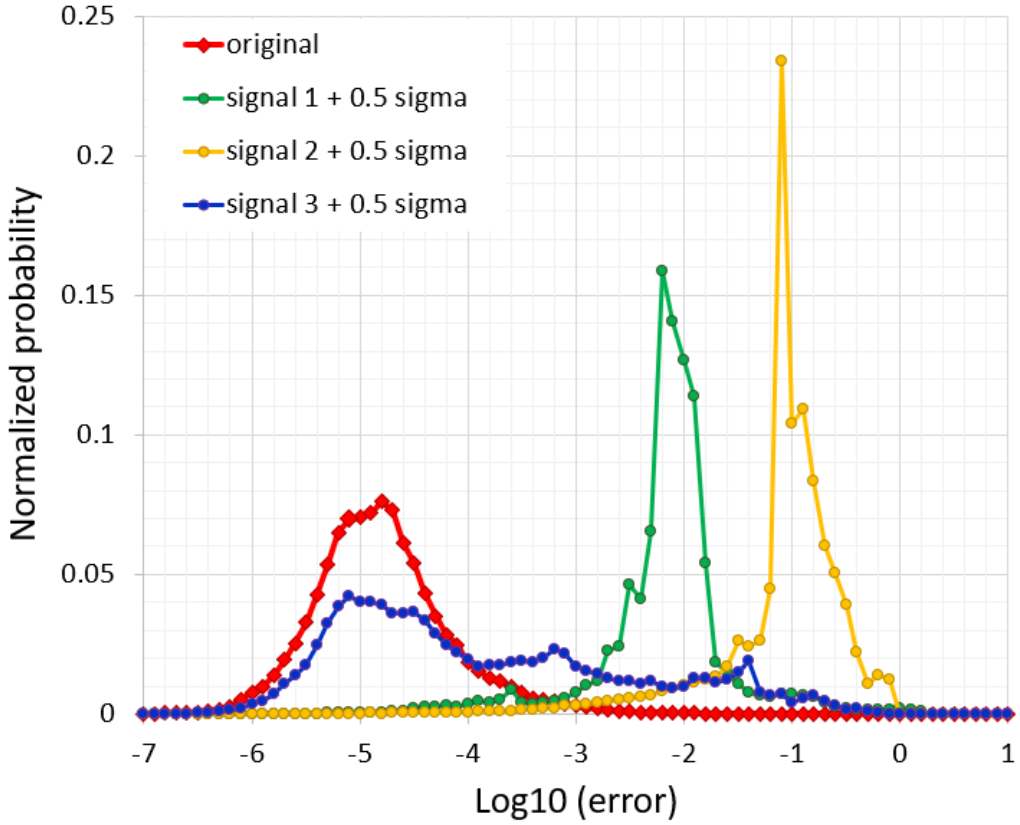


Model verification



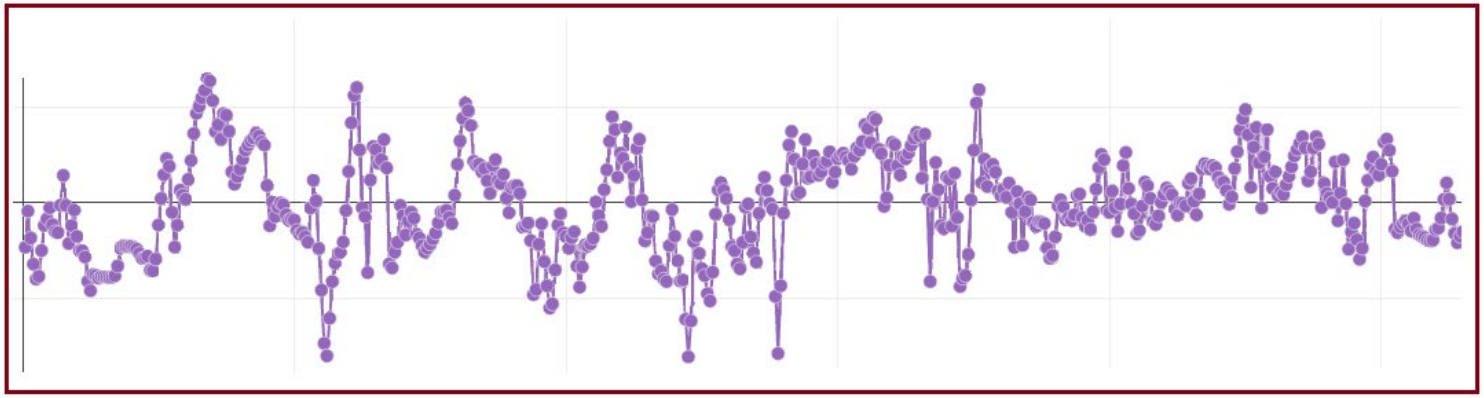
Model Verification using data that this neural net didn't see before

Need to have at least one sensitive signal



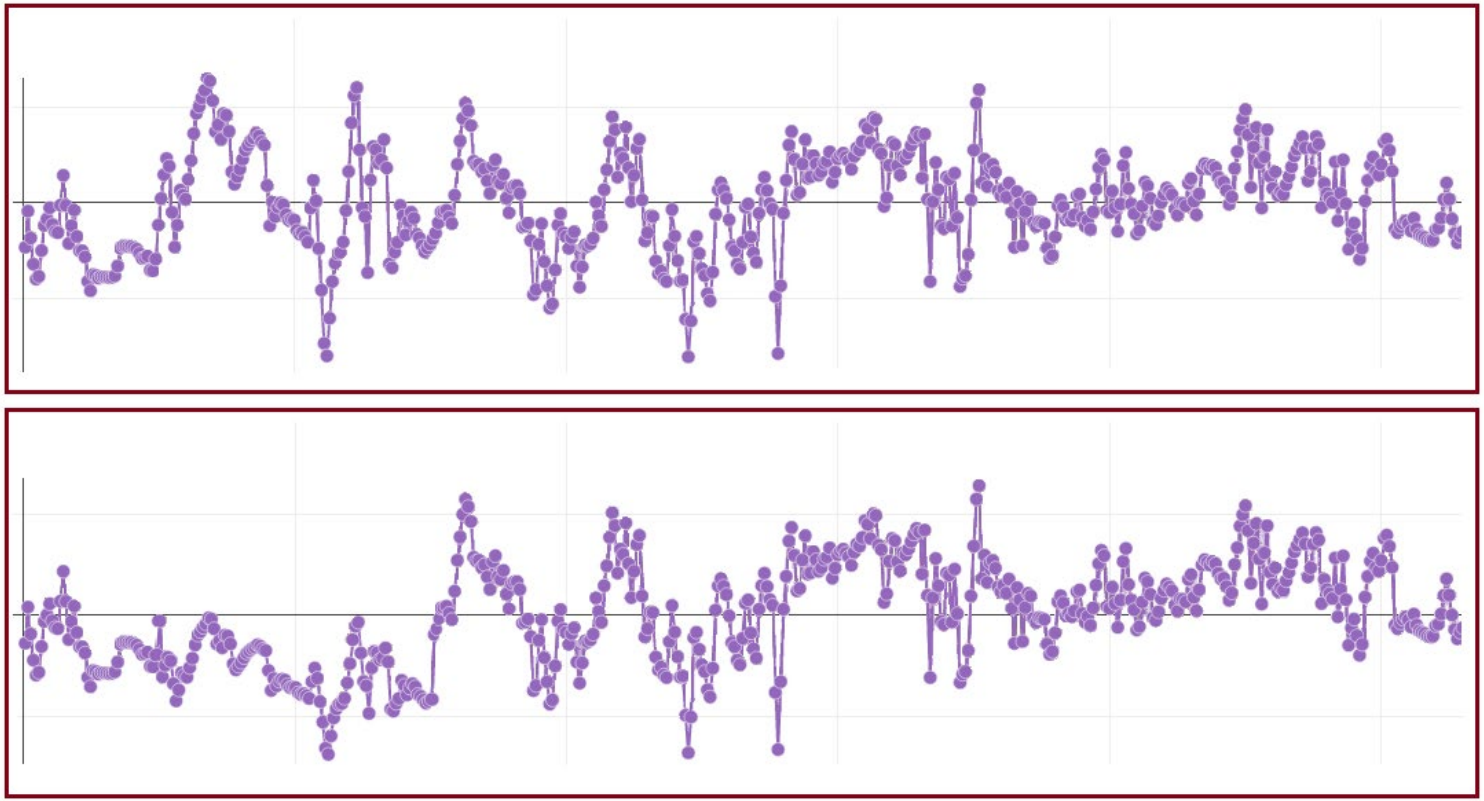
Increase in error when one of the signals is disturbed
(by one half of the standard deviation)

Example using real and modified data



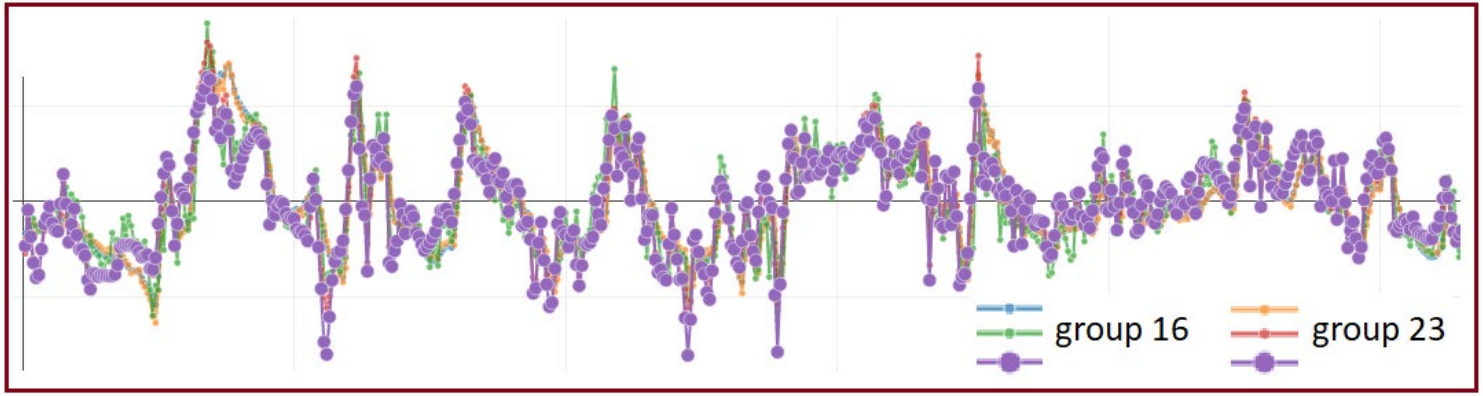
Signal recorded in the field

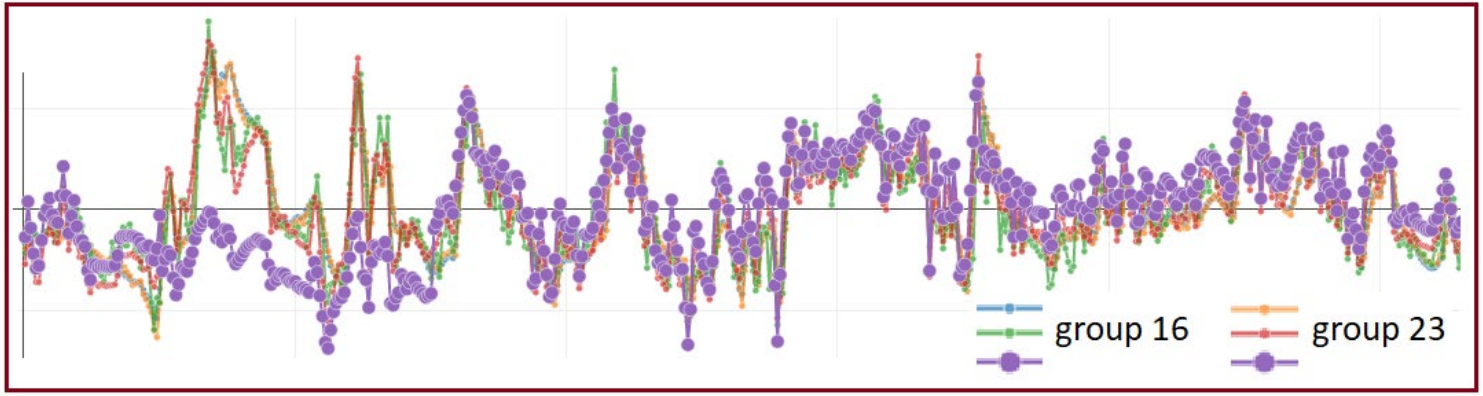
Example using real and modified data



Signal recorded in the field?

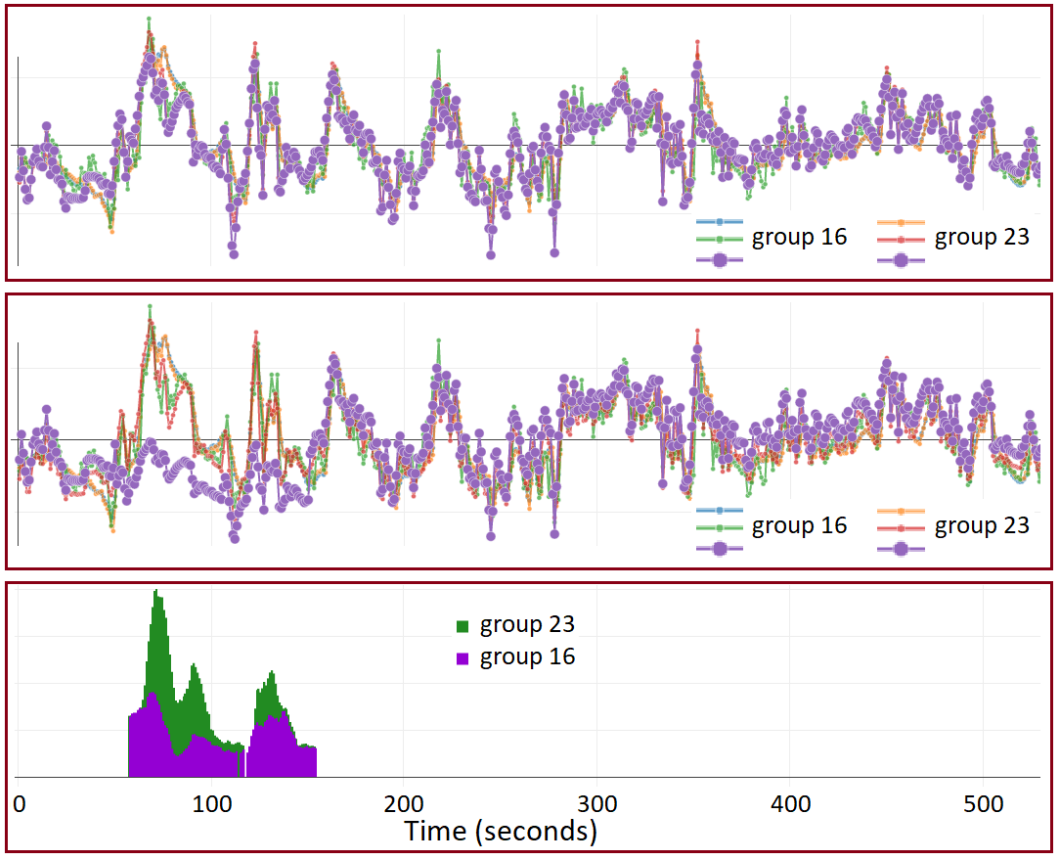
Example using real and modified data





Signals recorded in the field

Example using real and modified data



Anomaly detection
by
the auto-encoder

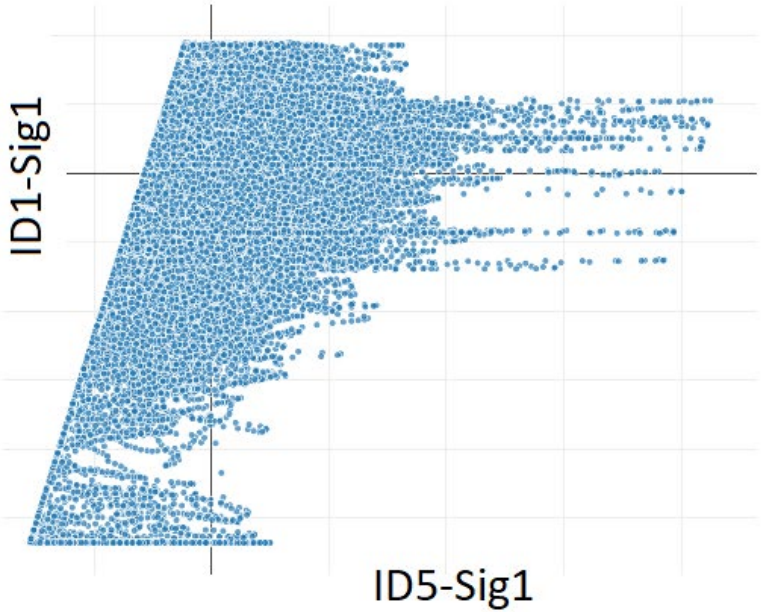
SynCAN dataset

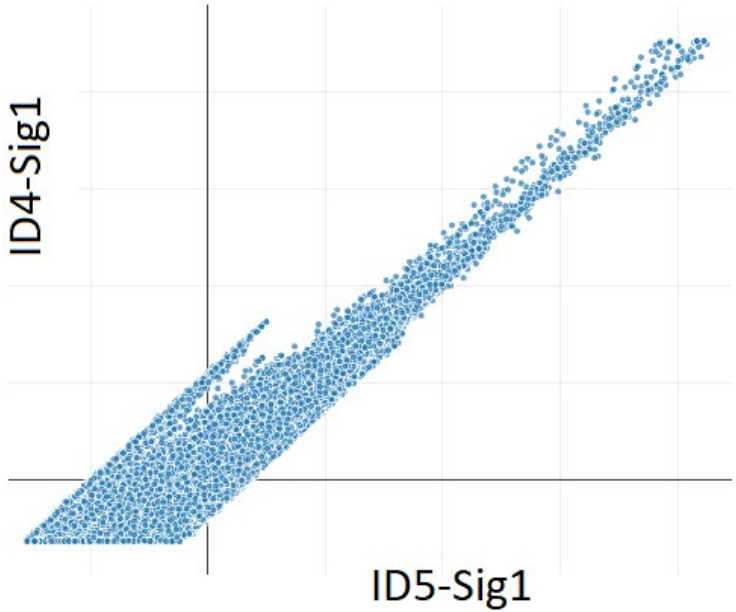
- Markus Hanselmann, Thilo Strauss, Katharina Dormann, and Holger Ulmer. 2019.

  SynCAN Dataset. https://github.com/etas/SynCAN

- We grouped 20 signals into 11 autoencoder groups

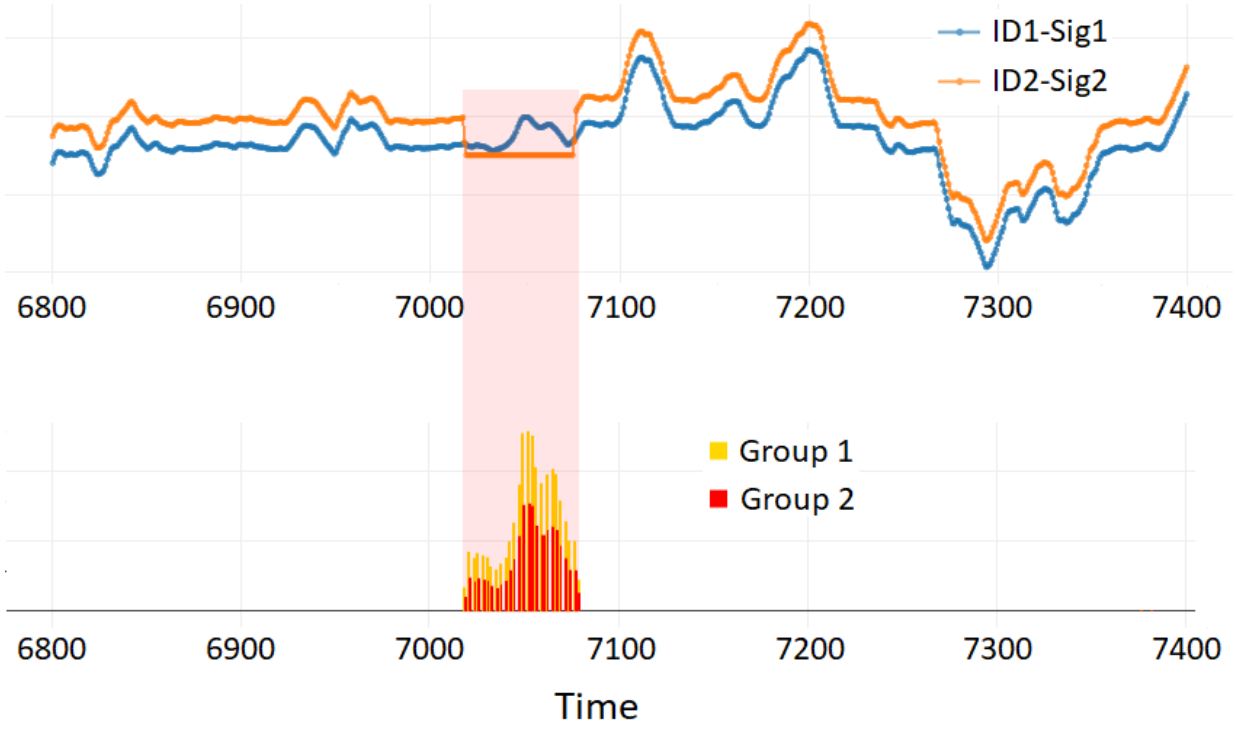
SynCAN dataset





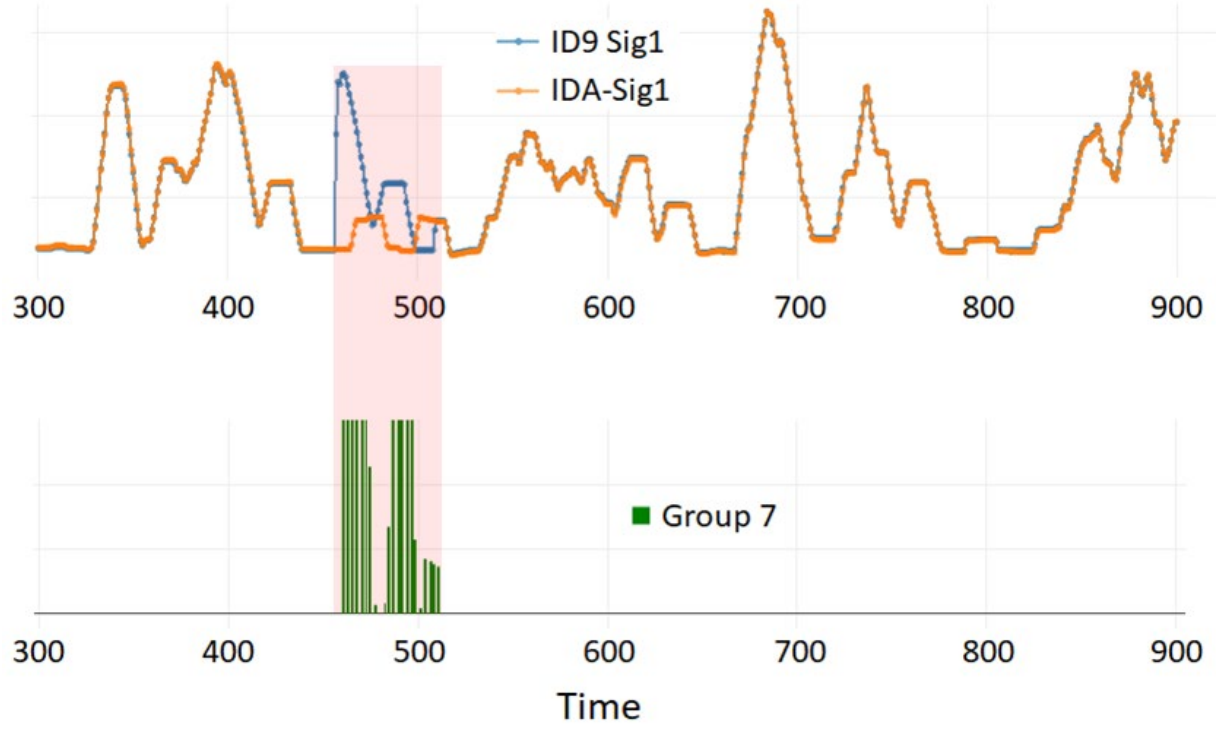
2D relationship between related signals in SynCAN dataset

SynCAN dataset



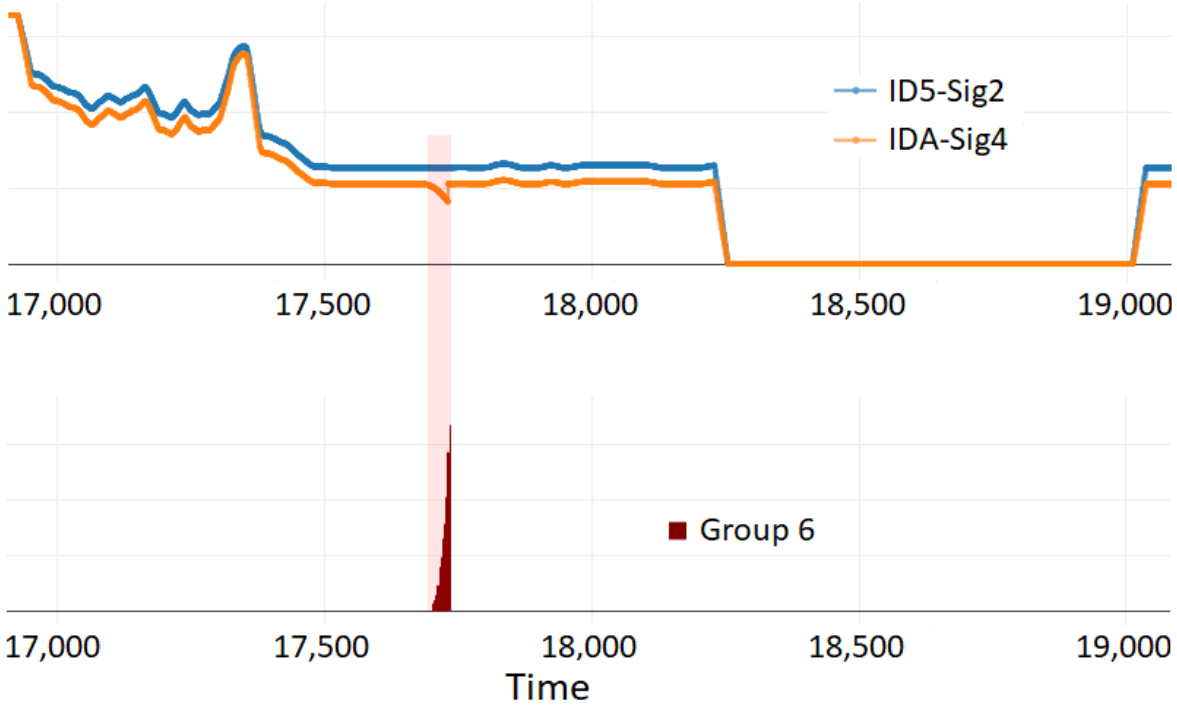
Detection of a "plateau" attack in SynCAN dataset, using auto-encoder

SynCAN dataset



Detection of a "replay" attack in SynCAN dataset, using auto-encoder

SynCAN dataset



Detection of a "continuous" attack in SynCAN dataset, using auto-encoder

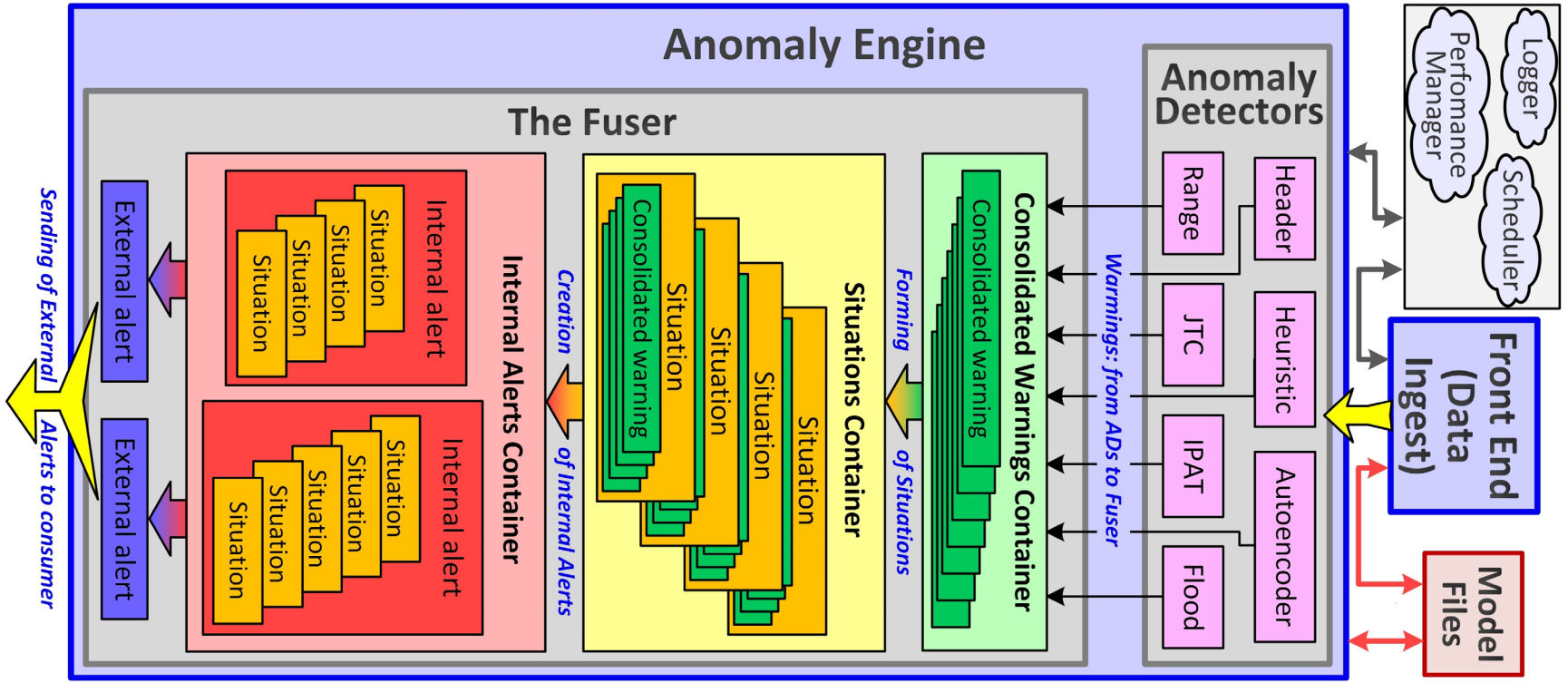
Auto-encoders for anomaly detection

Questions?

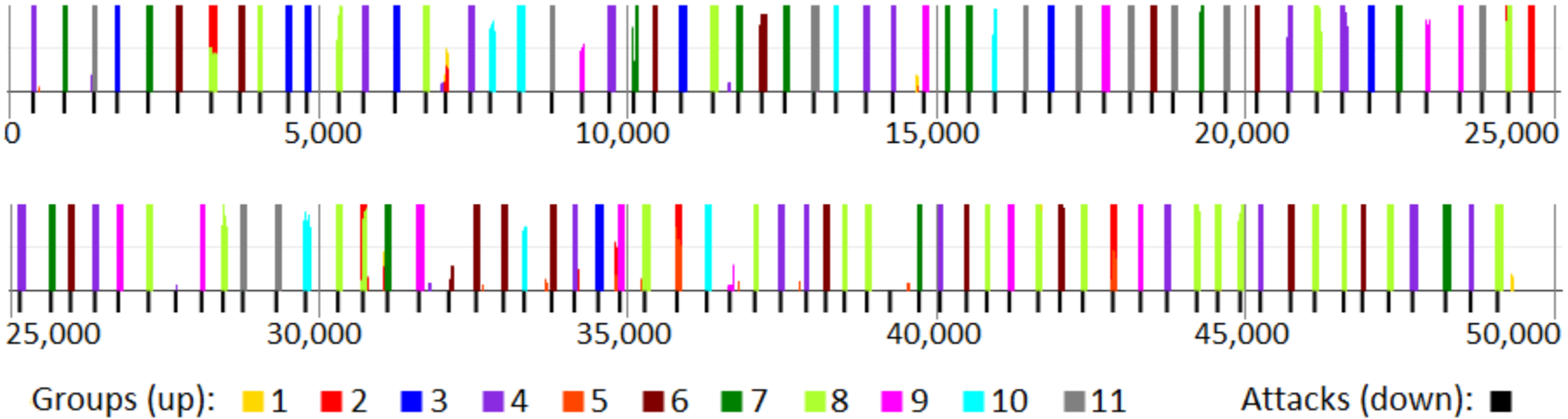
Auto-encoders for anomaly detection

Back up slides

Fuser

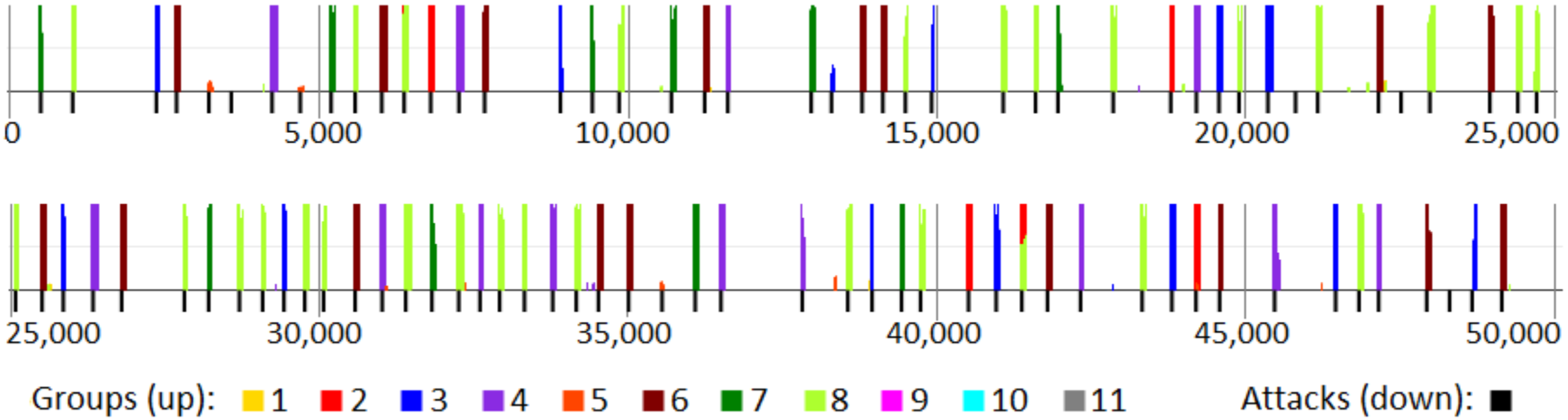


SynCAN dataset



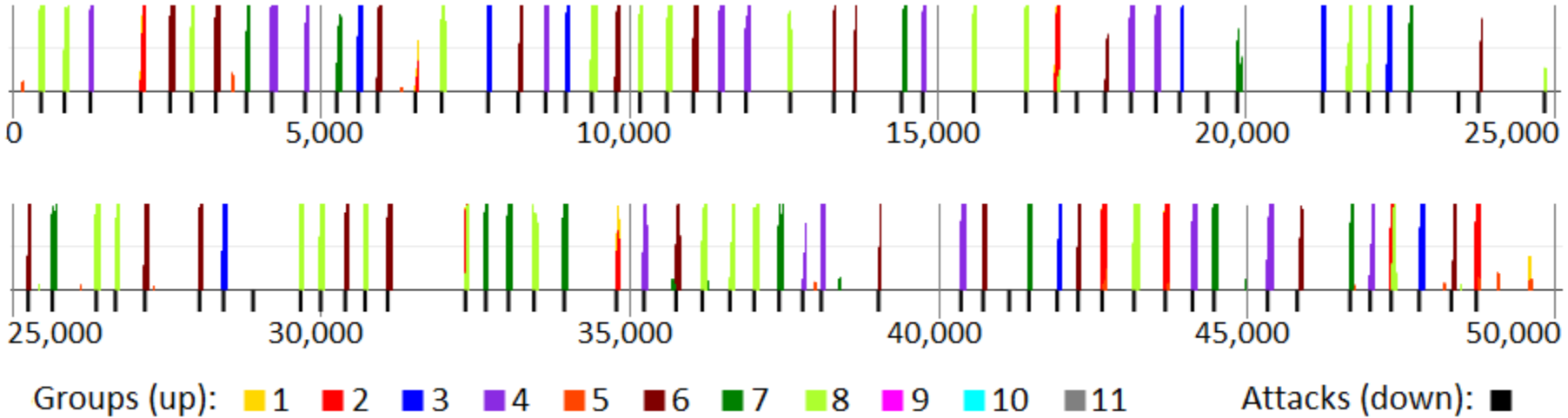
Statistics of detecting plateau attacks in SynCAN dataset

SynCAN dataset



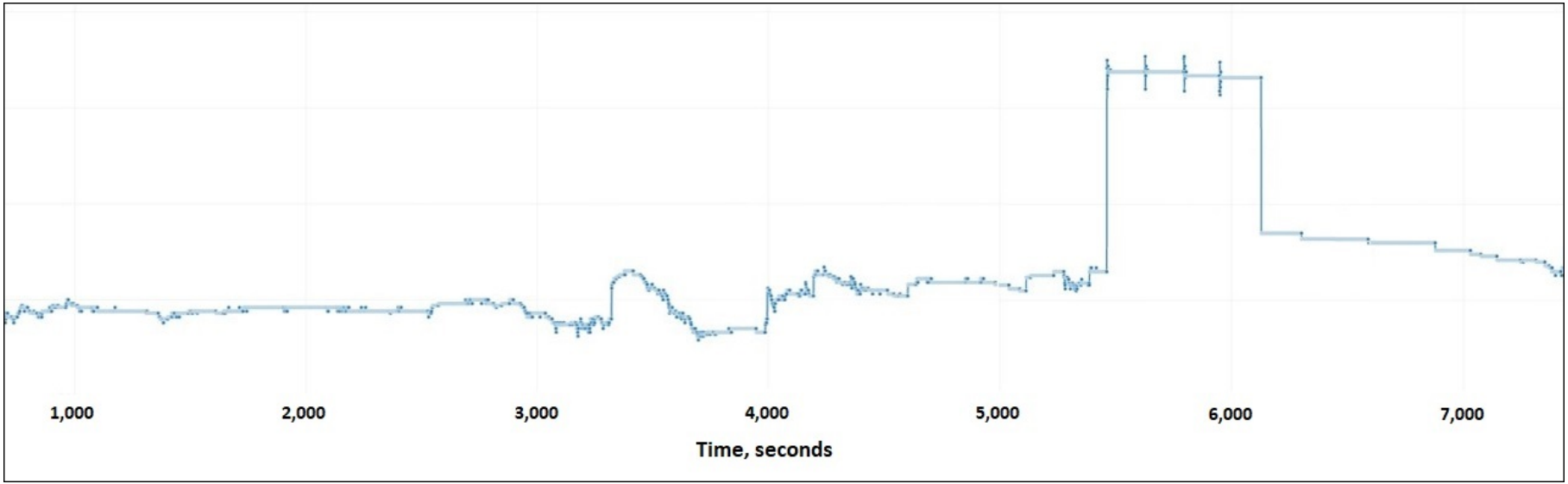
Statistics of detecting replay attacks in SynCAN dataset

SynCAN dataset



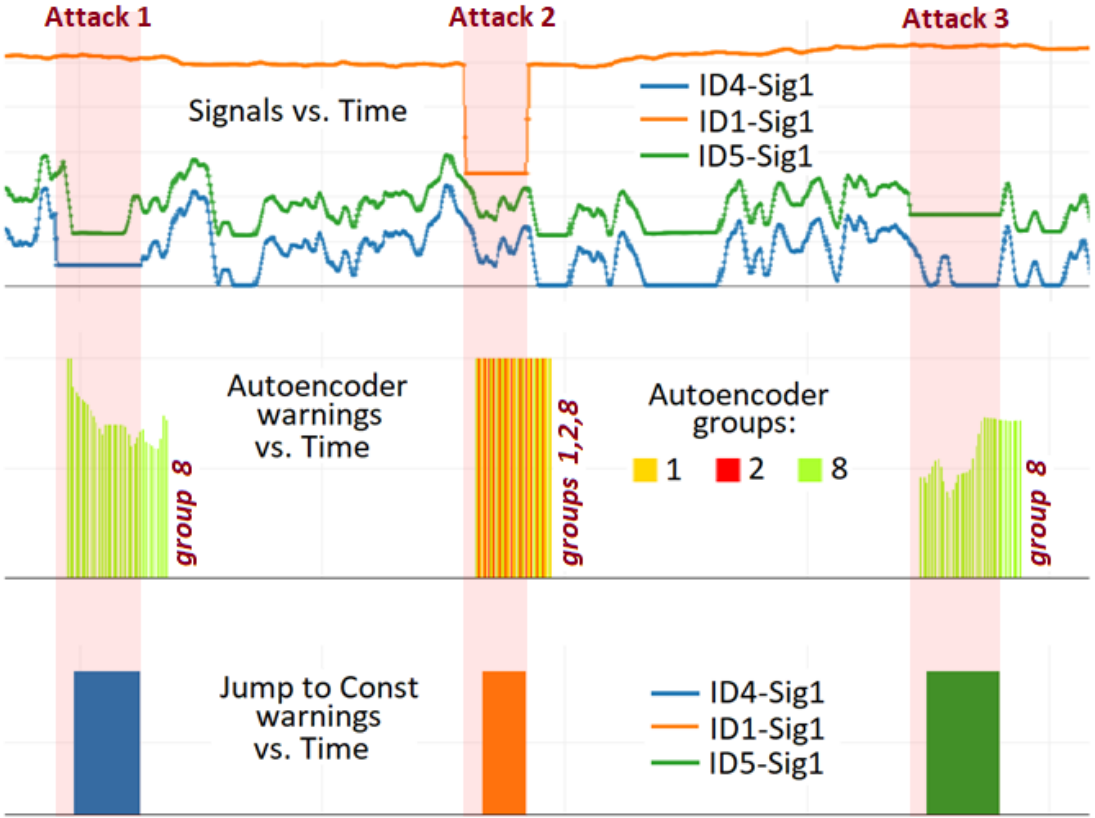
Statistics of detecting continuous attacks in SynCAN dataset

Other detectors of Fox Shield



Detection of a faulty physical sensor using JTC detector
(a real case from the field observations)

SynCAN dataset



Detection of "plateau" attacks in SynCAN dataset, using auto-encoder and Jump to Constant detector

AMERICAN INSTITUTE OF
AERONAUTICS AND ASTRONAUTICS