



NOTED and NL

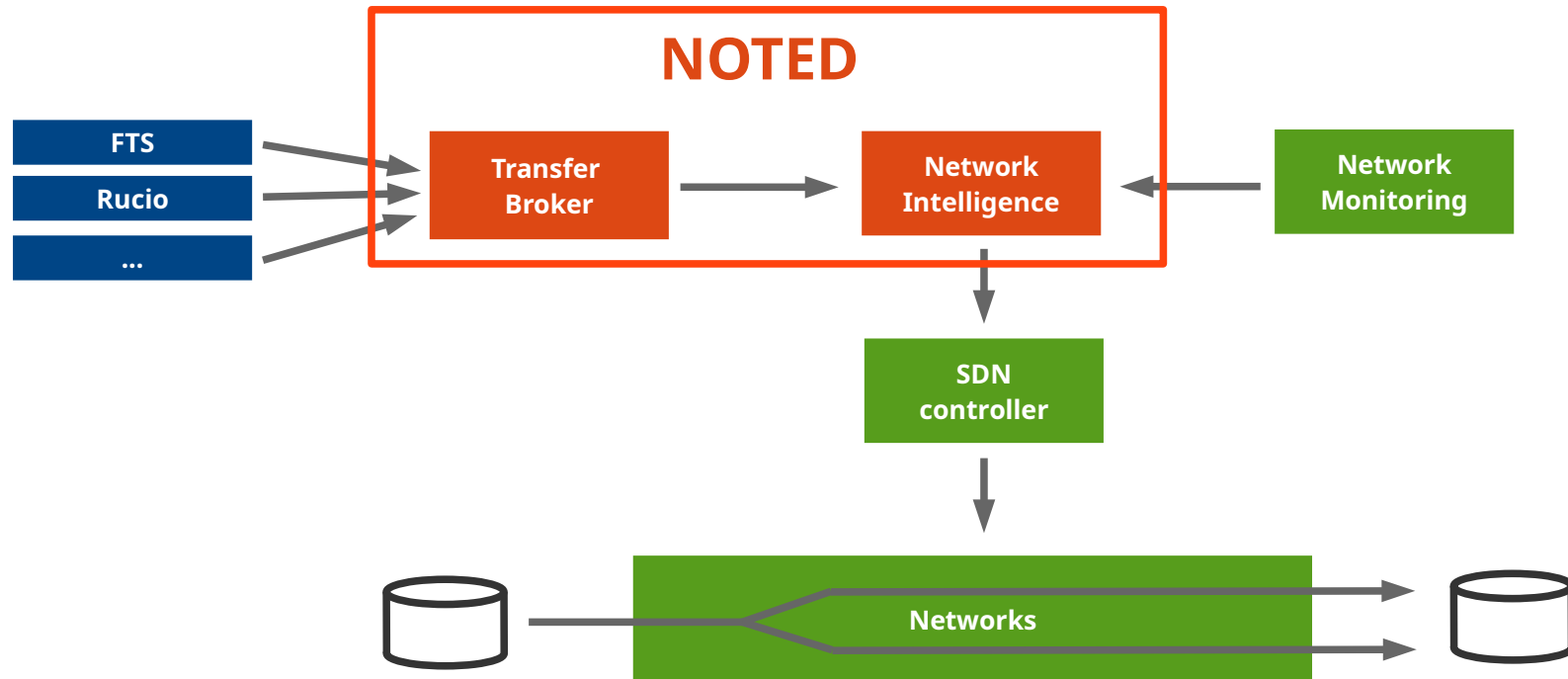
CERN IT Machine Learning Infrastructure Workshop

10th March 2023

Carmen Misa Moreira and Edoardo Martelli

Network Optimized Transfers of Experimental Data

NOTED: framework that dynamically improves network performances for **large, on-going, long-lasting** data transfers



Data Transfers

- The current NOTED implementation works only with FTS
- NOTED queries FTS via the CERN MONIT Infrastructure
- Relevant parameters collected:
 - **{source se, dest se}**: source and destination endpoints involved in the transfer
 - **{throughput, filesize avg}**: throughput [bytes/s] and filesize [bytes] of the transfer
 - **{active count, success rate}**: number of TCP parallel flows and successful rate of the transfer
 - **{submitted count, connections}**: number of transfers in the queue and maximum number of transfers that can be held

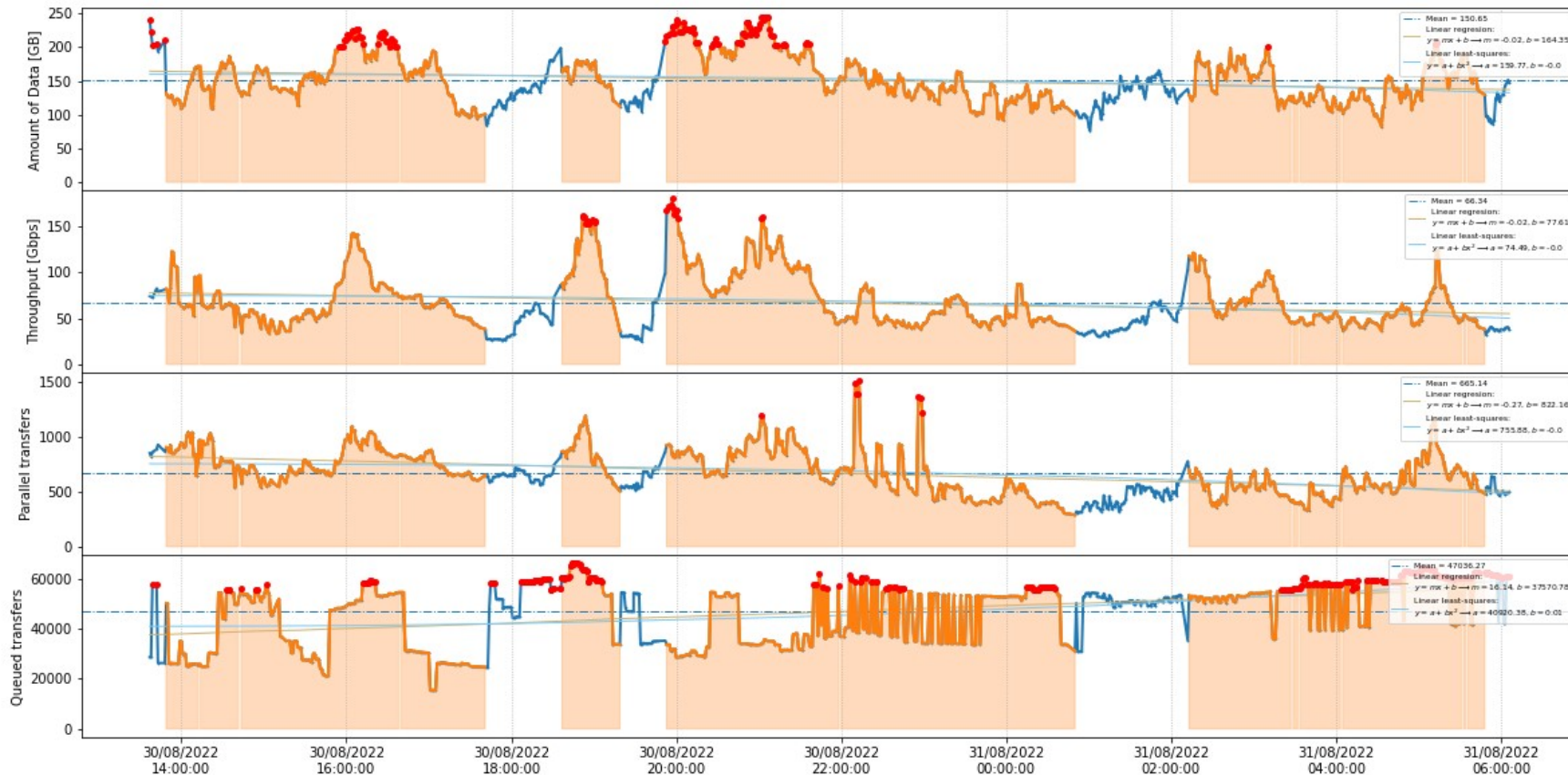
Machine learning

Machine Learning LSTM has been tested to better estimate the duration and the size of the transfers

Work in progress

“Plain” NOTED in actions

LHCONE 31th of August 2022

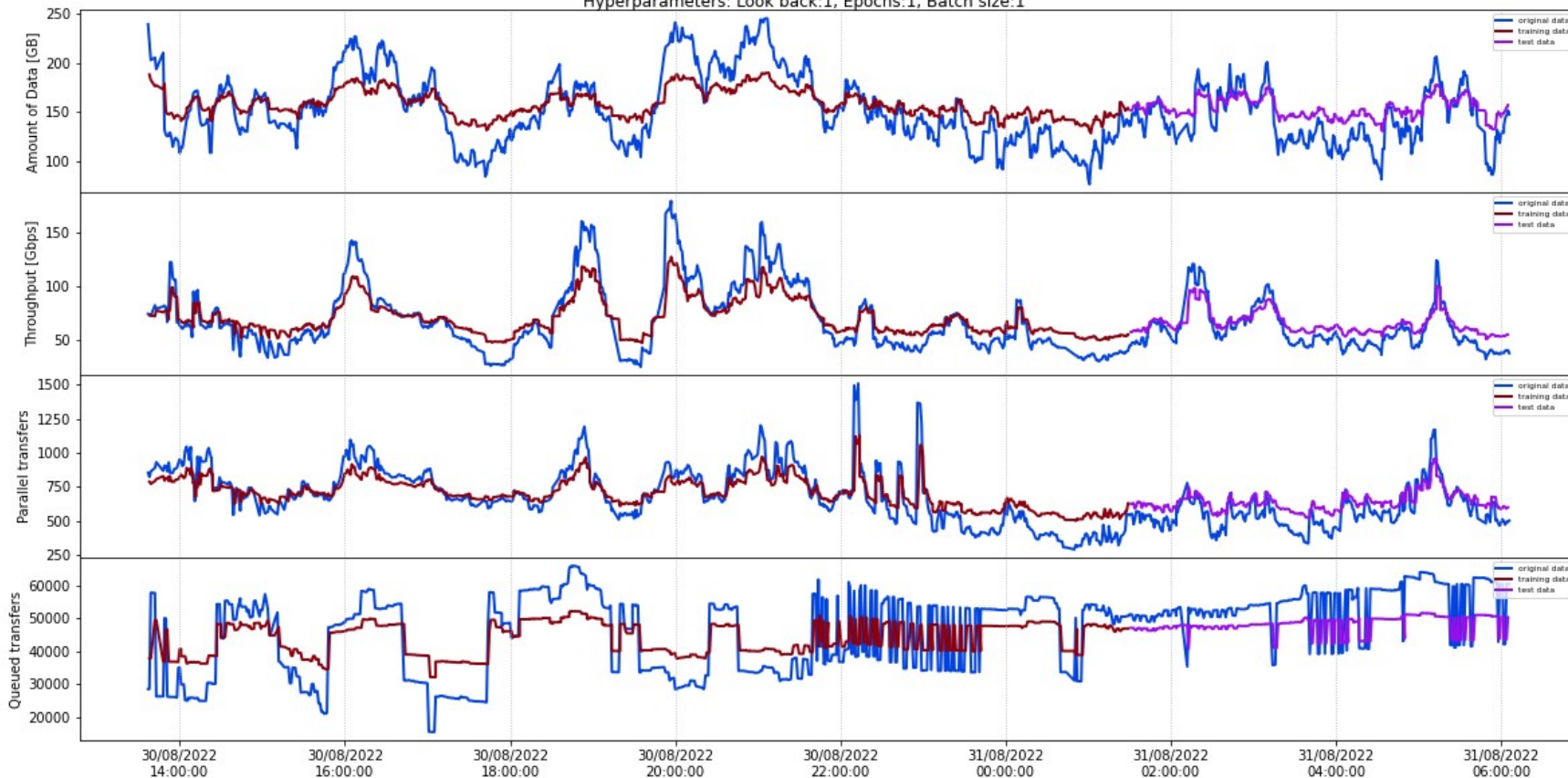


Orange area: NOTED triggered network action

Traffic forecast with LSTM

Long-Short Term Memory Machine Learning Algorithm
Traffic Forecasting
LHCONE 31th of August 2022

Hyperparameters: Look back:1, Epochs:1, Batch size:1



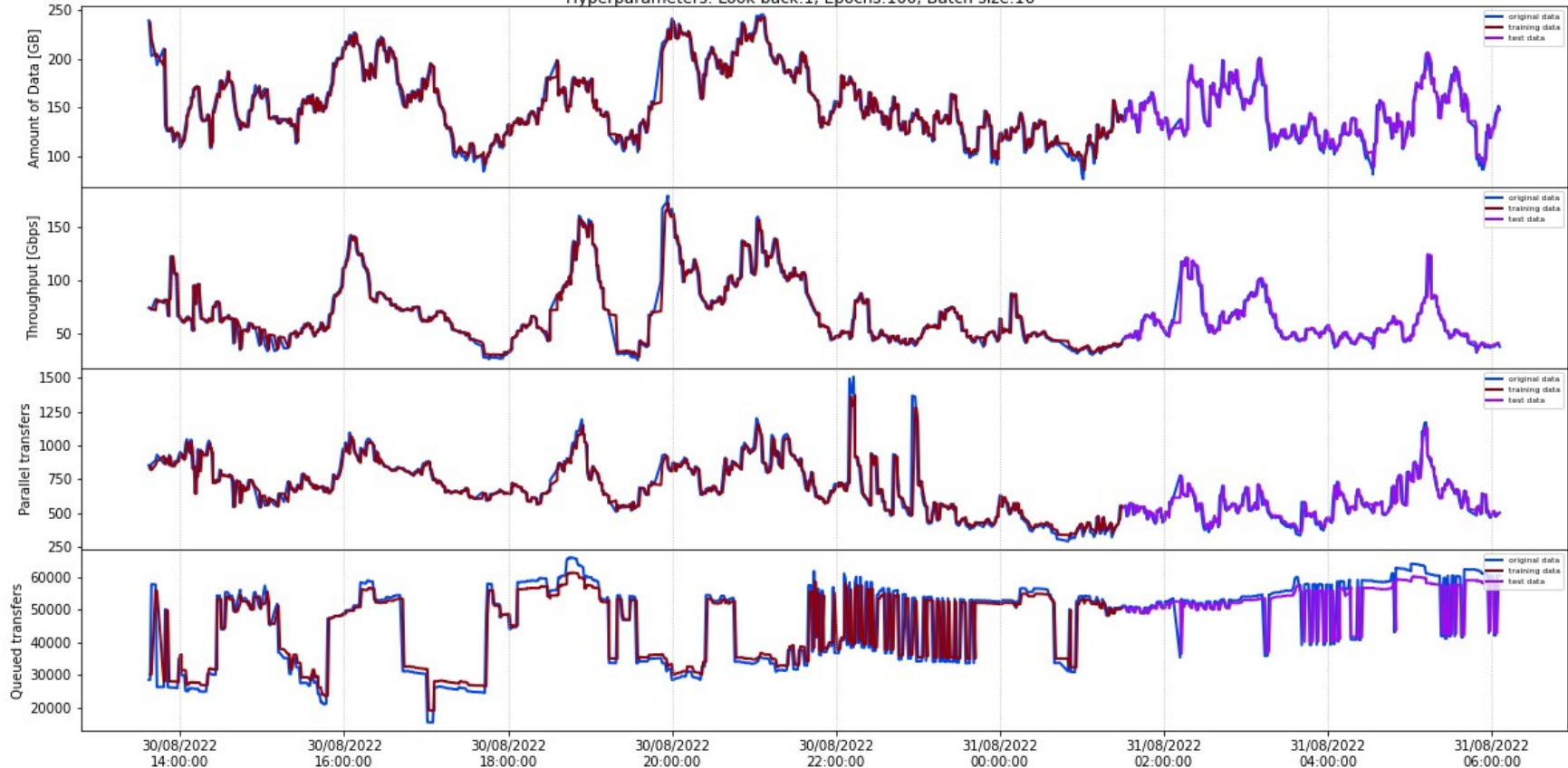
This model is not well fitted to the data

Real data
Training data
Predicted data

Traffic forecast with LSTM

Long-Short Term Memory Machine Learning Algorithm
Traffic Forecasting
LHCONE 31th of August 2022

Hyperparameters: Look back:1, Epochs:100, Batch size:16



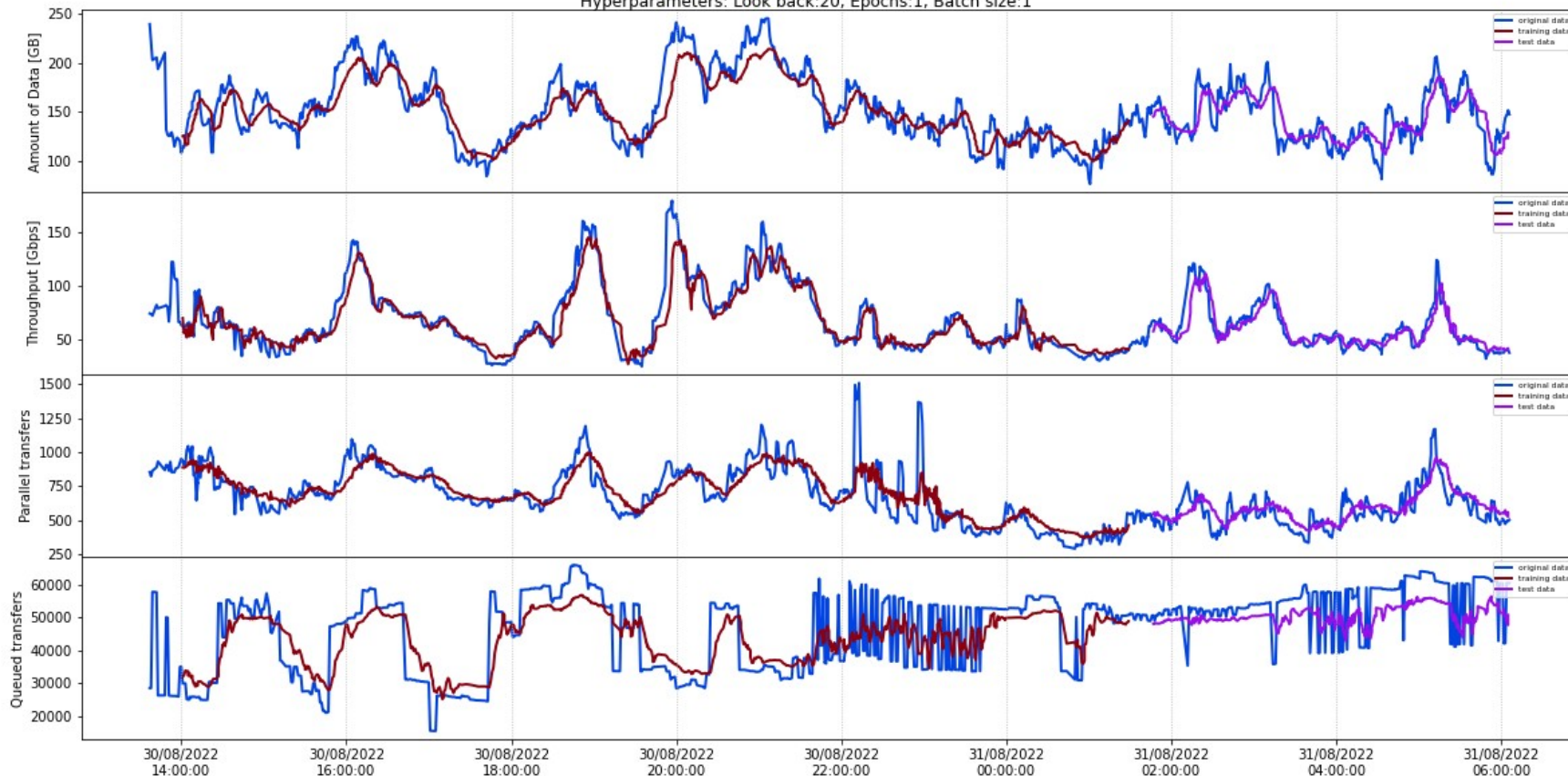
With increased Epoch and Batch size, the model fits very well

Real data
Training data
Predicted data

Traffic forecast with LSTM

Long-Short Term Memory Machine Learning Algorithm
Traffic Forecasting
LHCONE 31th of August 2022

Hyperparameters: Look back:20, Epochs:1, Batch size:1



With increased loopback (20), the model fits well even with epoch 1 and batch 1

Real data
Training data
Predicted data

Layers of the LSTM network

The LSTM network has:

- a visible layer with 1 input,
- a hidden layer with 4 LSTM blocks or neurons,
- an output layer that makes a single value prediction

The sigmoid activation function is used for the LSTM blocks

Execution details

Look back: 1 Epochs: 1 Batch size: 1

CPU times: user 3.45 s, sys: 120 ms, **total: 3.57 s**
Peak memory: 711.70 MiB
Train Score: 15.77 RMSE
Test Score: 11.37 RMSE
Length of train dataset: 821
Length of test dataset: 353

Look back: 1 Epochs: 100 Batch size: 16

CPU times: user 16.3 s, sys: 578 ms, **total: 16.8 s**
Peak memory: 816.57 MiB
Train Score: 6.96 RMSE
Test Score: 5.59 RMSE
Length of train dataset: 821
Length of test dataset: 353

Look back: 20 Epochs: 1 Batch size: 1

CPU times: user 5.57 s, sys: 138 ms, **total: 5.7 s**
Peak memory: 823.27 MiB
Train Score: 11.83 RMSE
Test Score: 9.22 RMSE
Length of train dataset: 821
Length of test dataset: 353

Future research

Use autoencoders and transformers

Make predictions in real time

Questions?

edoardo.martelli@cern.ch
carmen.misa@cern.ch

