



Machine Learning for Data Quality Management

Monitoring Trigger Rates with Variational Autoencoders

Madeline Hagen



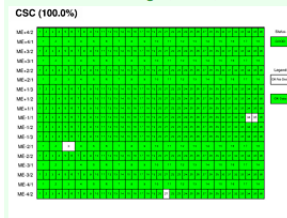
TEXAS TECH
UNIVERSITY.

M UNIVERSITY OF
MICHIGAN

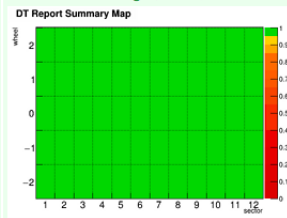
Data Quality Monitoring

The goal of the HCAL ML group is to automate the monitoring system for early warning of detector issues

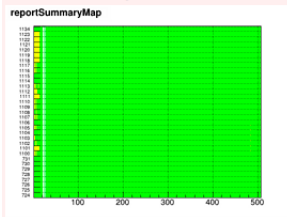
CSC - 100.0% - Aug 9, 2017, 09:56.19



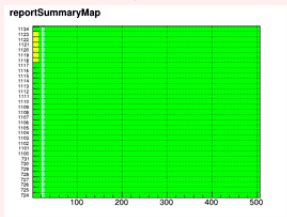
DT - 100.0% - Aug 9, 2017, 09:56.19



Hcal - 0.0% - Aug 9, 2017, 09:56.11



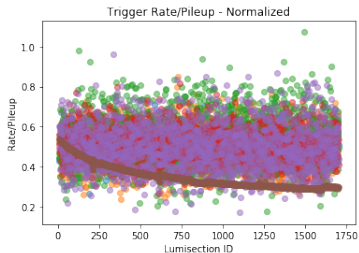
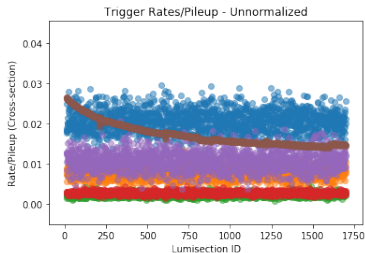
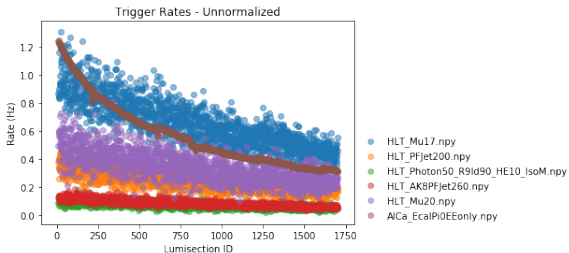
Hcal2 - 0.0% - Aug 9, 2017, 09:56.06



Trigger Rates as a Monitor

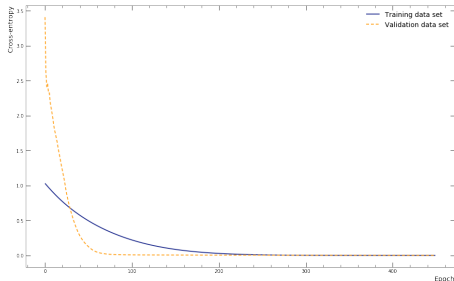
- Changes in the rate of data collection is a signal poor detector health
- Trained an autoencoder to learn and reproduce rates. The distance between the input and output is used to decipher between good and bad rates.

Inputs

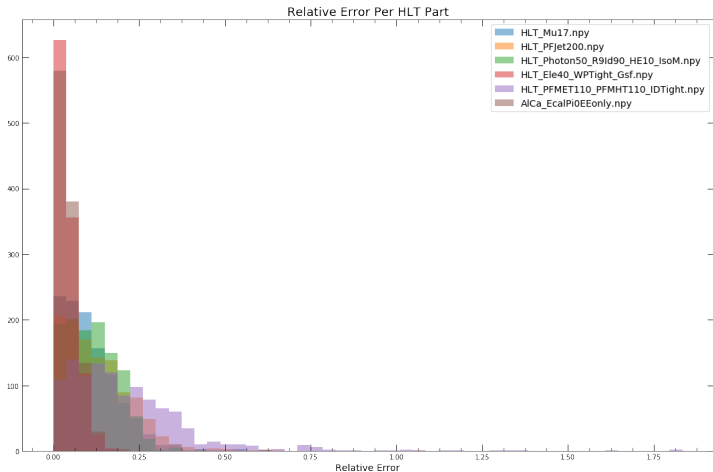


Architecture

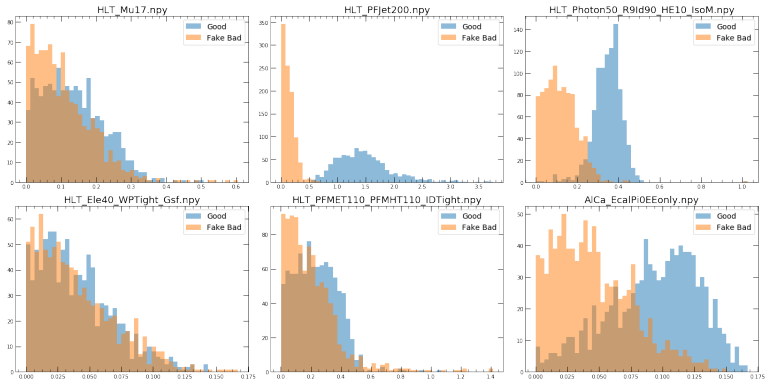
Layer (type)	Output Shape	Param #
input (InputLayer)	(None, 6)	0
middle (Dense)	(None, 5)	35
batch_normalization_33 (Batch Normalization)	(None, 5)	20
encoded (Dense)	(None, 3)	18
batch_normalization_34 (Batch Normalization)	(None, 3)	12
middle2 (Dense)	(None, 5)	20
batch_normalization_35 (Batch Normalization)	(None, 5)	20
reconstructed (Dense)	(None, 6)	36
batch_normalization_36 (Batch Normalization)	(None, 6)	24
Total params: 185		
Trainable params: 147		
Non-trainable params: 38		



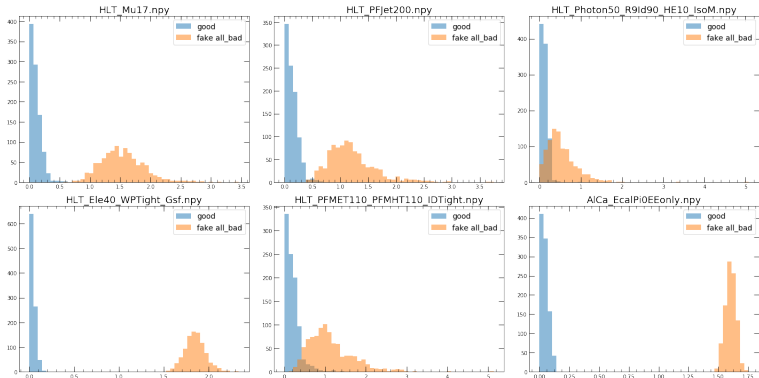
Training Data - Error



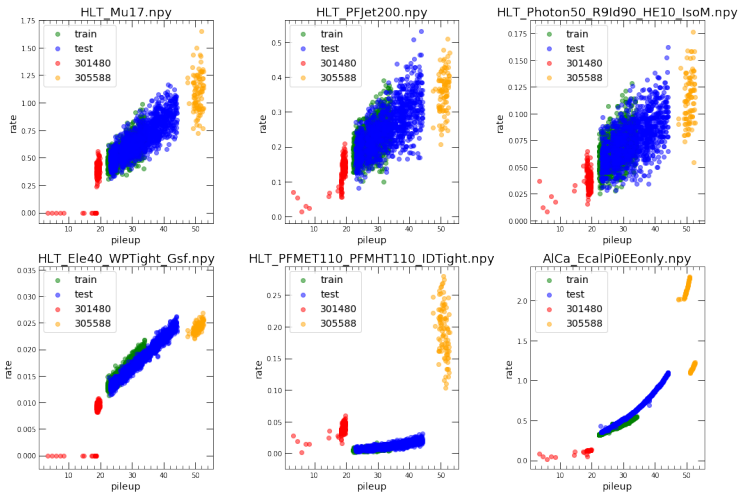
Fake Bad Data - One Rate



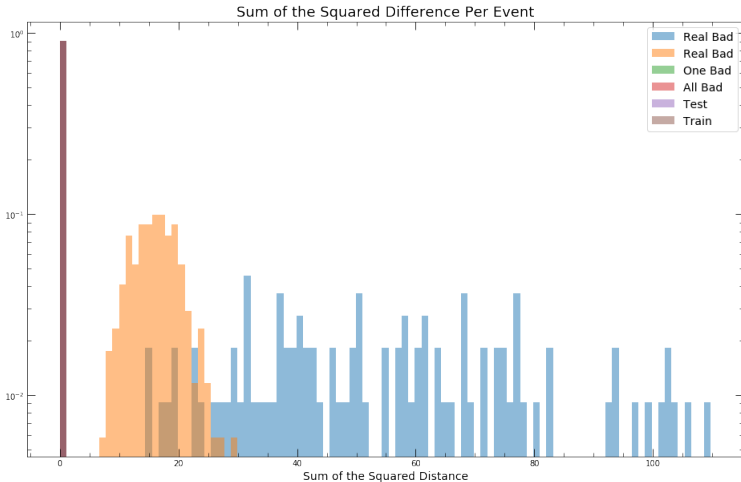
Fake Bad Data - All Rates



Real Bad Data



Real Bad Data



Roadblocks

- Normalization - Tried 7 different methods
- Architecture - For each normalization, 5-10 architectures were tried
- Used these to narrow in on current method

Next Steps

- Challenge the autoencoder - more HLT parts, varying input shapes
- Prediction - Can the autoencoder predict the next few rates to raise alarm of mid-run issues







home.cern