SMARTHEP
REAL-TIME ANALYSIS FOR SCIENCE AND INDUSTRY

Synthetic Data Generation with Lorenzetti

for Time Series Anomaly Detection in High-Energy Physics Calorimeters

Laura Boggia, Bogdan Malaescu and the Lorenzetti Team (Edmar De Souza, Juan Marin, Eduardo De Simas, Lucas Nunes and many more)

EuCAIF Conference 2025, Cagliari

SMARTHEP is funded by the European Union's Horizon 2020 research and innovation programme, call H2020-MSCA-ITN-2020, under Grant Agreement n. 956086

ATLAS EXPERIMENT IBM LPNHE PARIS MARIE CURIE ACTIONS

Artificial (synthetic) anomalies

With Lorenzetti shower simulation framework

- **Motivation:** Anomaly detection in multivariate time series is crucial for various fields
  - Healthcare, financial services, cybersecurity, manufacturing lines, **data quality monitoring at physics detectors**, etc.
  - Often serious lack of reliable labels —> **artificial anomalies**

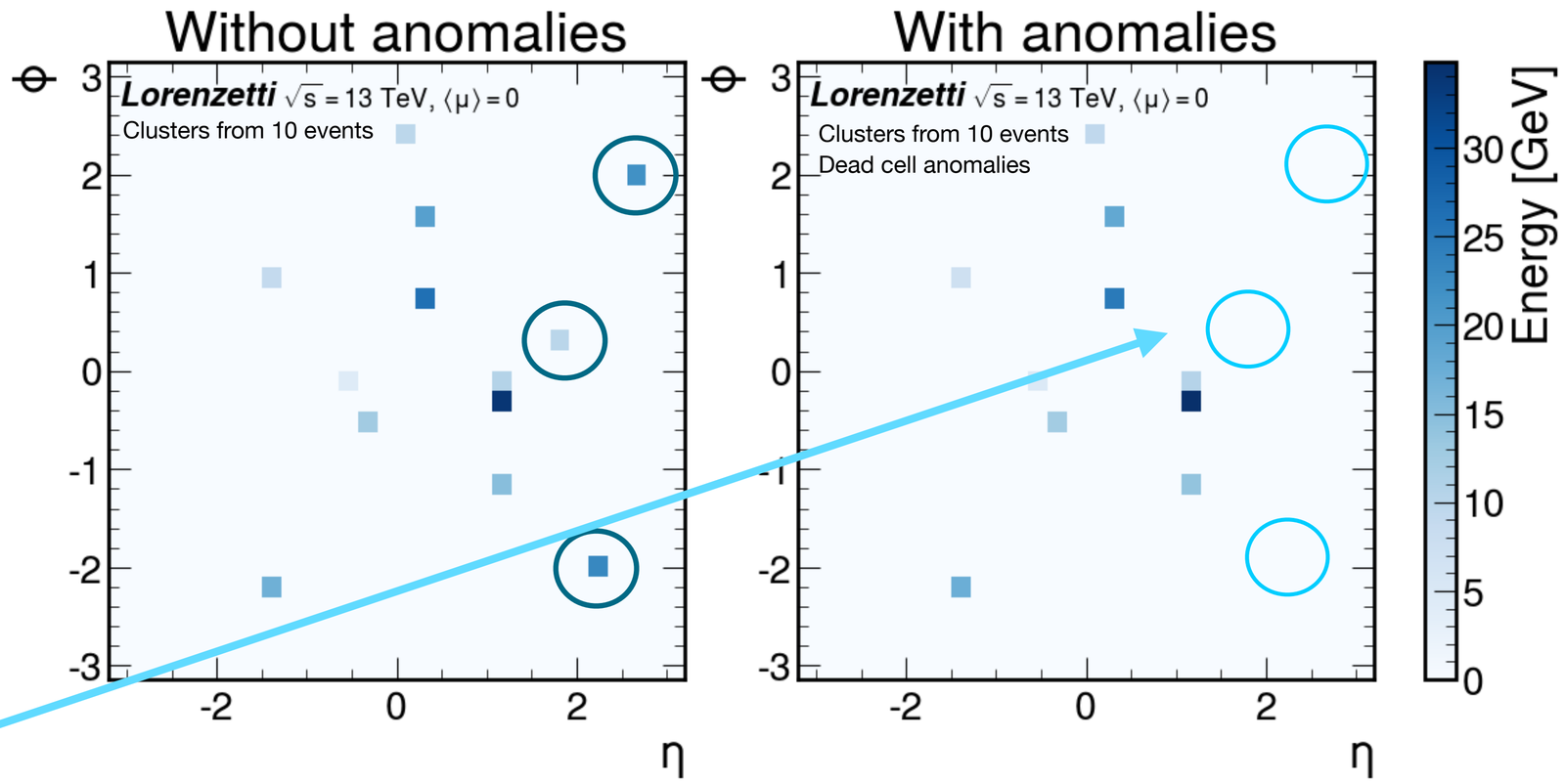
- **Project:** Lorenzetti calorimeter simulation with artificial anomalies
  1. Insert various anomaly rates and types
  2. Identify the anomalies with deep learning anomaly detection for time series

Lorenzetti code: github.com/lorenzetti-hep/

Lorenzetti

Calorimeter simulation

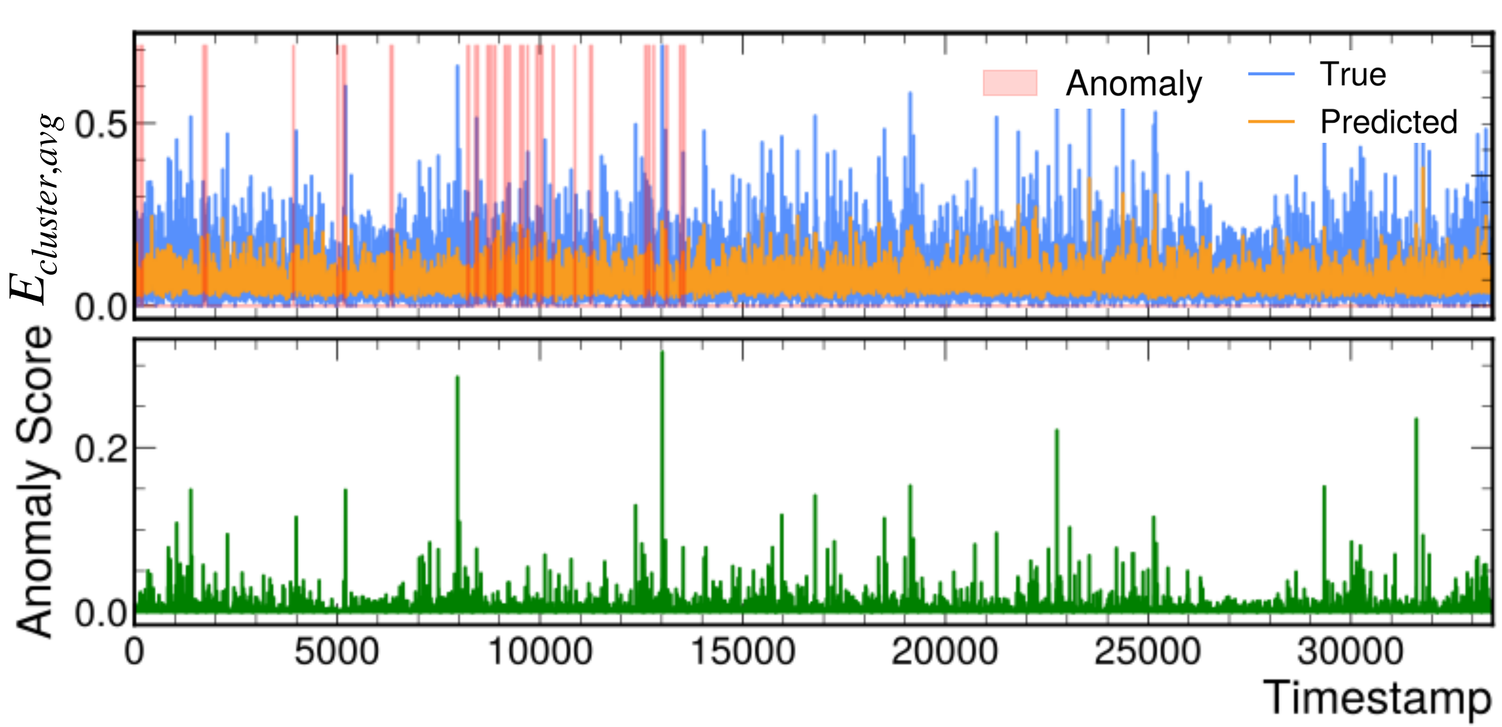
- Open source framework
- Simulation of general-purpose calorimeter, based on ATLAS
- Introduce various synthetic anomalies:
  - Increase noise for cells containing physics signal
  - Dead detector cells

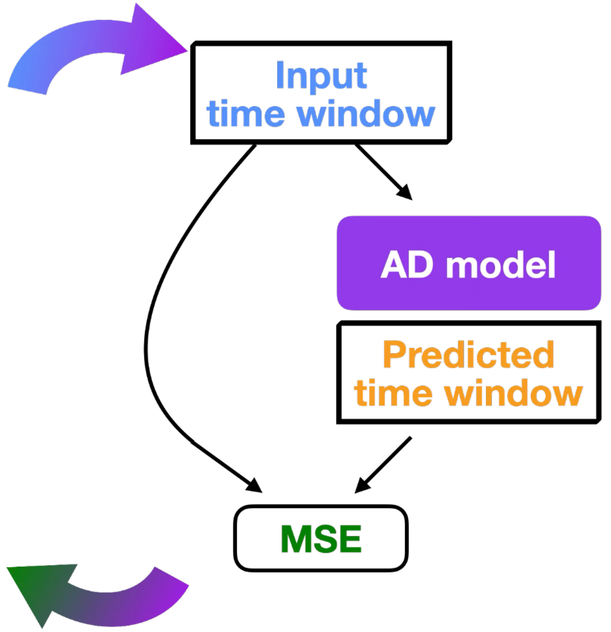


Lorenzetti paper: M.V. Araújo et al. "Lorenzetti Showers - A general-purpose framework for supporting signal reconstruction and triggering with calorimeters."

Time series anomaly detection

With reconstruction-based model





- **Anomaly score** defined as reconstruction error (computed with MSE)
- Compare 3 deep learning approaches and one unsupervised baseline

Thank you for your attention…

& see you at the Wednesday poster session!

Feel free to reach out:
laura.boggia@cern.ch

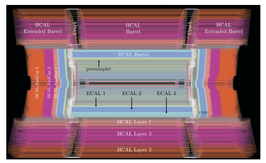
SMARTHEP is funded by the European Union's Horizon 2020 research and innovation programme, call H2020-MSCA-ITN-2020, under Grant Agreement n. 956086

Synthetic Data Generation with Lorenzetti for Time Series Anomaly Detection in High-Energy Physics Calorimeters

Laura Boggia*, Bogdan Malaescu and the Lorenzetti Team (Edmar De Souza, Juan Marin, Eduardo De Simas, Lucas Nunes et al.)

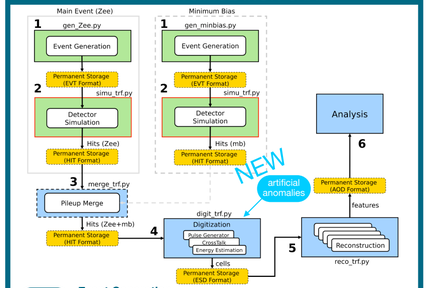
LORENZETTI SHOWER SIMULATION FRAMEWORK

The **Lorenzetti Showers** (LZT) simulation is an integrated software framework that provides complete calorimeter information [1].



- Simulates general-purpose calorimeter, though calorimeter cell granularity is based on ATLAS technical design
- Provides complete calorimeter information (no tracking), including pileup and crosstalk effects
- Open-source code can be found on github.com/lorenzetti-hep/ or scan:

LORENZETTI SIMULATION CHAIN



1 EVT — Event Generation
- Physics event generation with Pythia8 particle generator [2]
- For our studies: $pp \rightarrow$ jets

2 HIT — Detector Simulation
- Detector simulation with Geant4 [3]
- Simulates interaction of particles with detector
- Produces energy deposits (hits) in calorimeter from particles

3 MB — Pileup Merge (optional)
- Repeat 1 and 2 once for signal particles and once for background (soft QCD) events
- Merge the energy deposits in the calorimeters

4 ESD — Digitization
- Simulates electronic pulses reacting to energy deposits in calorimeter cells

NEW: insert artificial anomalies corresponding to detector defects

5 AOD — Reconstruction
- Build energy clusters from cell signals
- Construct various other high-level variables based on cell signals such as shower shape variables

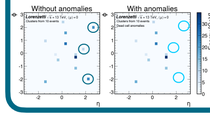
REFERENCES

[1] M.V. Araújo et al. "Lorenzetti Showers - A general-purpose framework for supporting signal reconstruction and triggering with calorimeters."
[2] T. Sjöstrand. "The Pythia Event Generator: Past, Present and Future."
[3] S. Agostinelli et al. "Geant4—a Simulation Toolkit."
[4] A. Siffer et al. "Anomaly Detection in Streams with Extreme Value Theory."
[5] Y. Liu et al. "iTransformer: Inverted Transformers Are Effective for Time Series Forecasting."
[6] S. Tuli et al. "TranAD: deep transformer networks for anomaly detection in multivariate time series data."
[7] J. Audibert et al. "USAD: UnSupervised Anomaly Detection on Multivariate Time Series."

DATA QUALITY / SYNTHETIC ANOMALIES

Anomaly detection in **multivariate time series** is crucial to ensure the **quality of data** coming from a physics experiment.

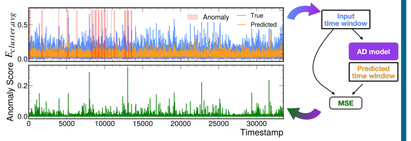
- Introduce **artificial anomalies** into detector ⟼ complete control over anomalies & labels



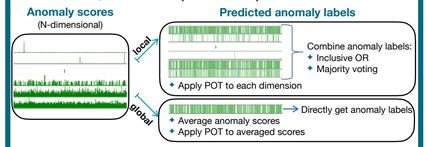
- Focus on QCD jet events
- Artificially increase noise for some cells
- Simulate **dead cells** (i.e. no detector signal)
  - Some clusters disappeared due to dead cells (see figure)

TIME SERIES ANOMALY DETECTION

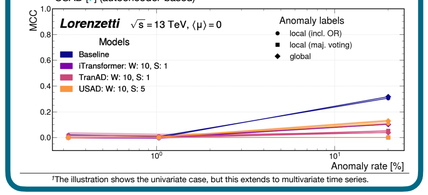
- Time series anomaly detection aims to identify anomalous time stamps in **sequential data**.
- Focusing on **reconstruction-based anomaly detection** as visualised here:†
  - **Anomaly score** defined as reconstruction error (computed with MSE)



- Dimensions correspond to cluster energy (2), position (2) and shower shapes (5) aggregated (mean and std) over each event ⟼ $2 \cdot 9 = 18$ dimensions
- Compare unsupervised deep learning models for anomaly detection on multivariate time series using **Peak-over-threshold (POT)** method [4]
- For all models, combine anomaly scores in 3 ways:



- Anomalies correspond to increased noise convoluted with physics signal
- Same train data, variable anomaly rate in test data
- For comparison: unsupervised baseline
  - Use absolute value of time series as input for POT
- Deep time series anomaly detection models:
  - iTransformer [5] and TranAD [6] (transformer-based)
  - USAD [7] (autoencoder-based)



†The illustration shows the univariate case, but this extends to multivariate time series.

OUTLOOK

- Insert noise signals that can create clusters independently of the physics signal
- More complex noise structures like larger coherent noise signals, detector blackouts etc.
- Feature selection:
  - What cluster-level variables are the most relevant to identify anomalies?
  - Can shower shape variables pick up on injected noise?

We acknowledge funding from the European Union Horizon 2020 research and innovation programme, call H2020-MSCA-ITN-2020, under Grant Agreement n. 956086

SMARTHEP *LPNHE, Sorbonne Université, CNRS/IN2P3, Paris
laura.boggia@cern.ch