Real-time AI-based dead hot map in the ePIC detector: a self-adaptive alternative to traditional big data calibration pipelines

Prof. Dr. András Hajdu

Dr. Balazs Ujvari

Dr. Balazs Harangi

Department of Data Science and Visualization
Faculty of Informatics, University of Debrecen

4th AI4EIC - Oct 27-30, 2025
IAIFI/MIT, Boston, MA

UNIVERSITY OF DEBRECEN
FACULTY OF INFORMATICS

University of Debrecen – 13 faculties

- Agricultural and Food Sciences and Environmental Management
- Dentistry
- Economics and Business
- Education for Children and Special Educational Needs
- **Engineering**
- Health
- Humanities
- **Informatics**
- Law
- Medicine
- Music
- Pharmacy
- **Science and Technology**

UNIVERSITY OF DEBRECEN
FACULTY OF INFORMATICS

Solid background in high-energy physics and AI

- Host of the
  - CMS (CERN) group, 4 FTE
  - EPIC group, 5 FTE
- High computational capacity for AI, NVIDIA DGX (A100, H100, H200) servers
- Official industrial AI-curricula (NVIDIA DLI, MS Learn Partner, CISCO, …)
- Industrial exam centers

Background

- PHENIX detector
- direct photon analysis
- DHM@EMCal (25k towels)
- Strong energy dependence
- Patterns from HV issues

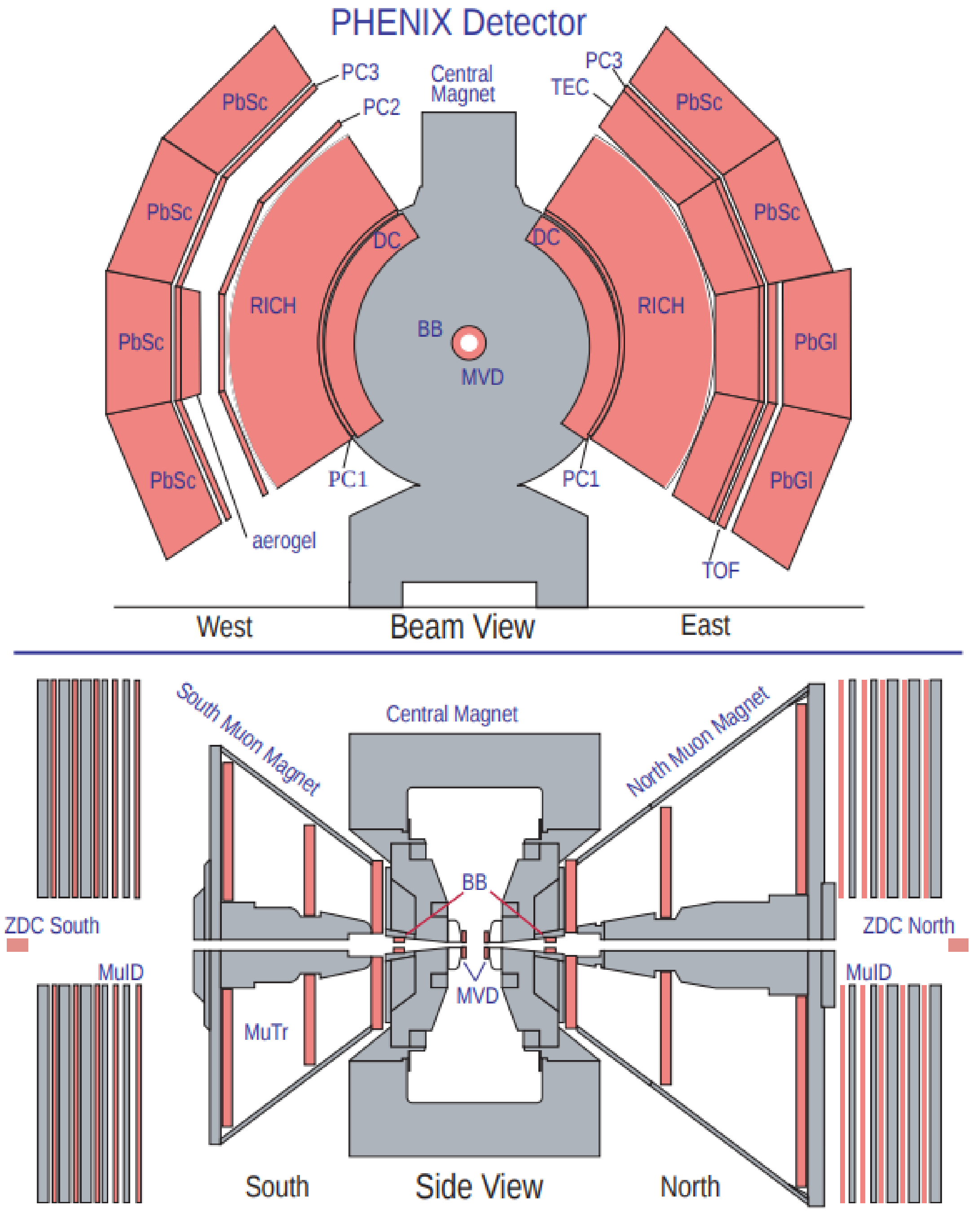
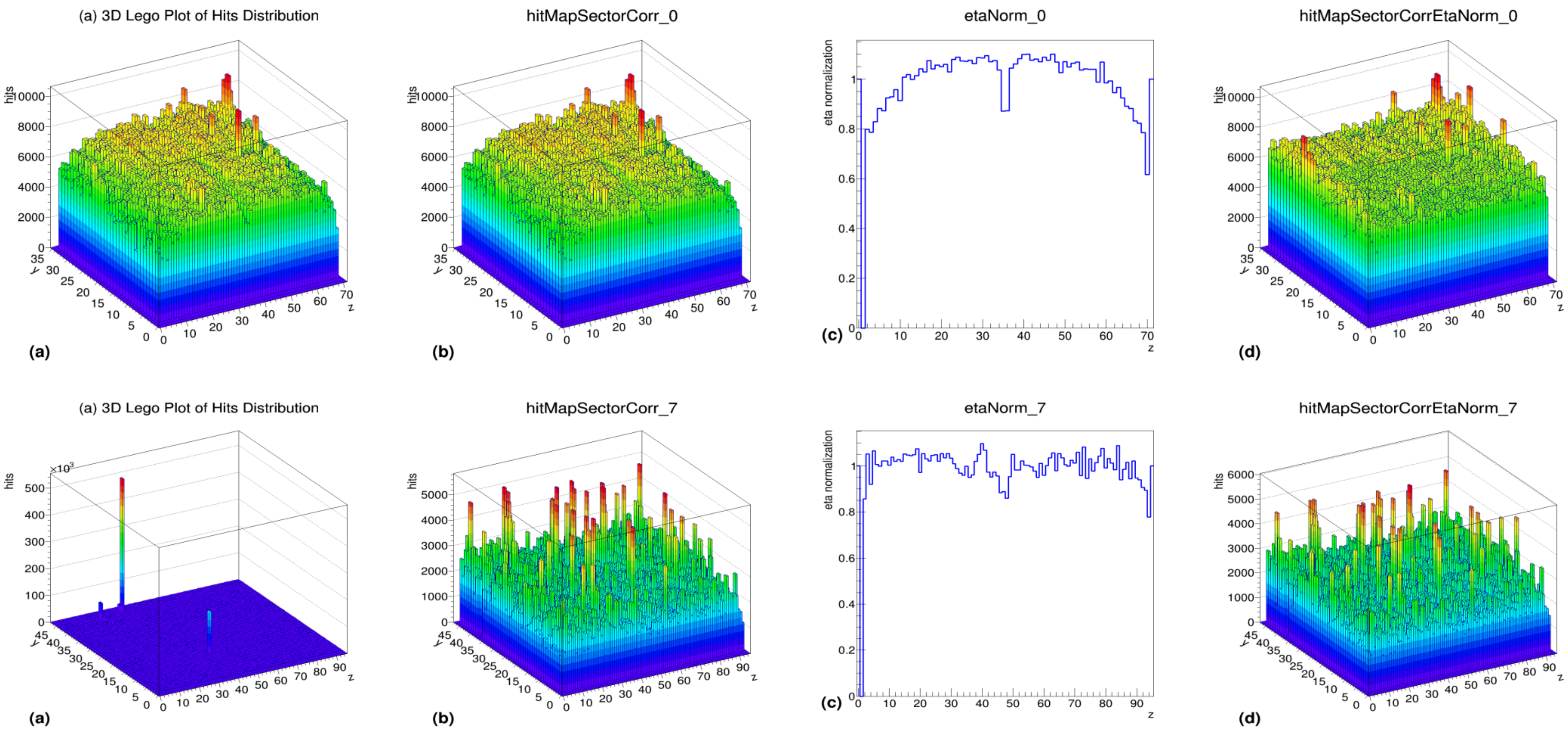


Figure 3.2: The PHENIX setup during the fourth RHIC beam period.

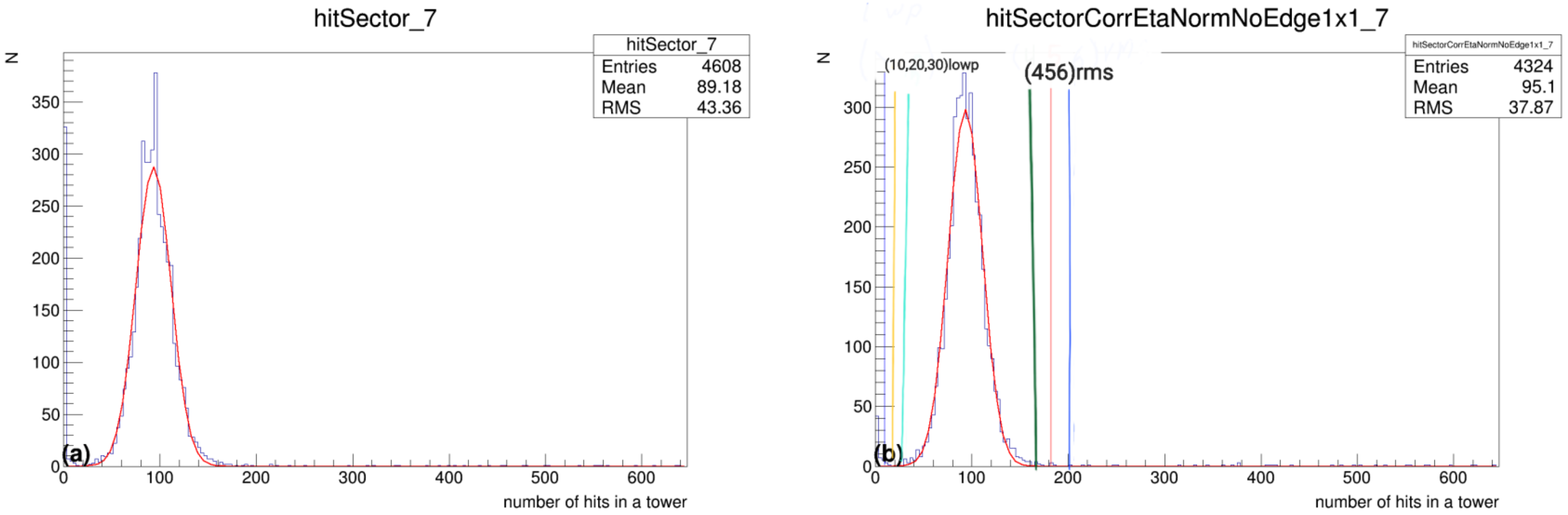
UNIVERSITY OF DEBRECEN
FACULTY OF INFORMATICS

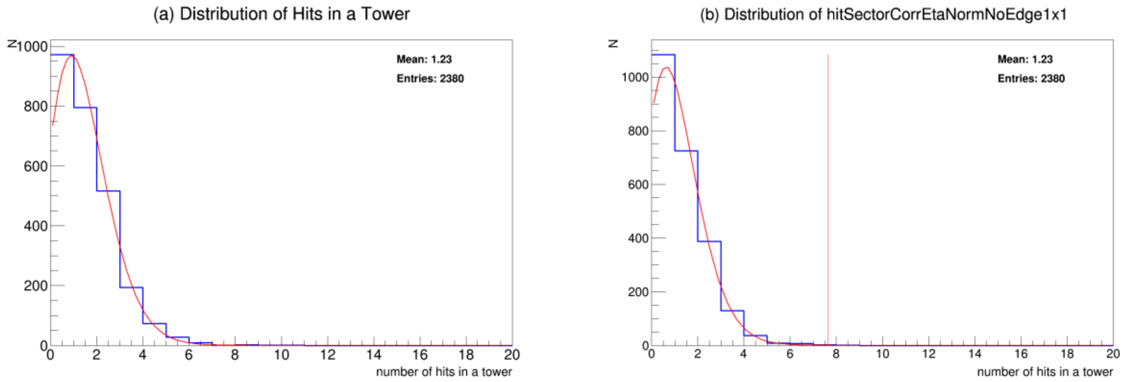
Raw maps, eta norm, 8 sectors, 40 energy bins, 1000 runs

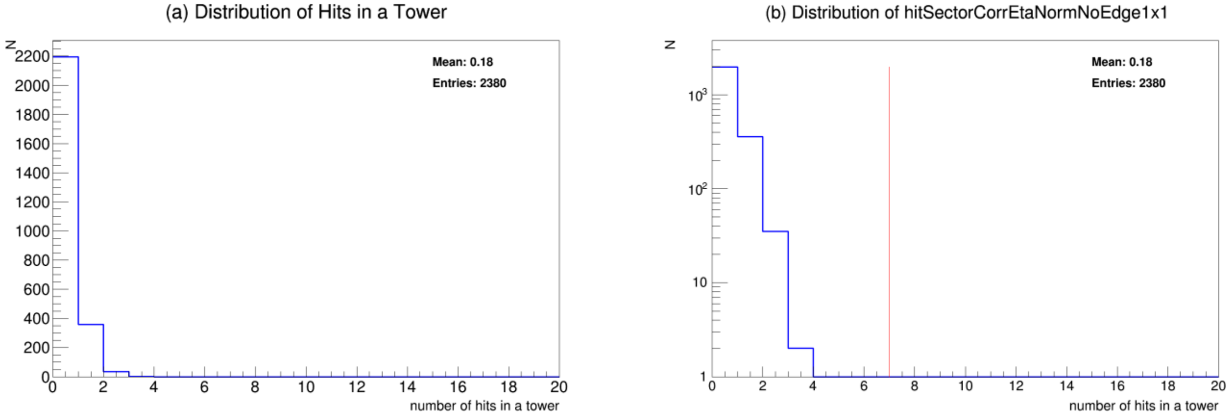


Hit distributions

- What is the normal?



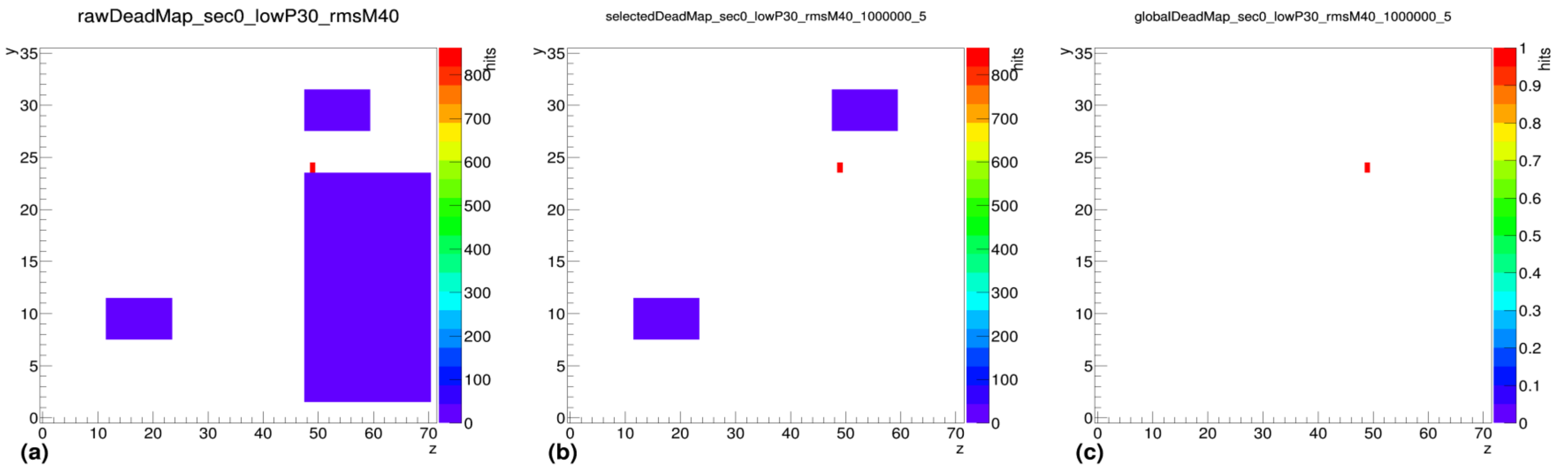




UNIVERSITY OF DEBRECEN
FACULTY OF INFORMATICS

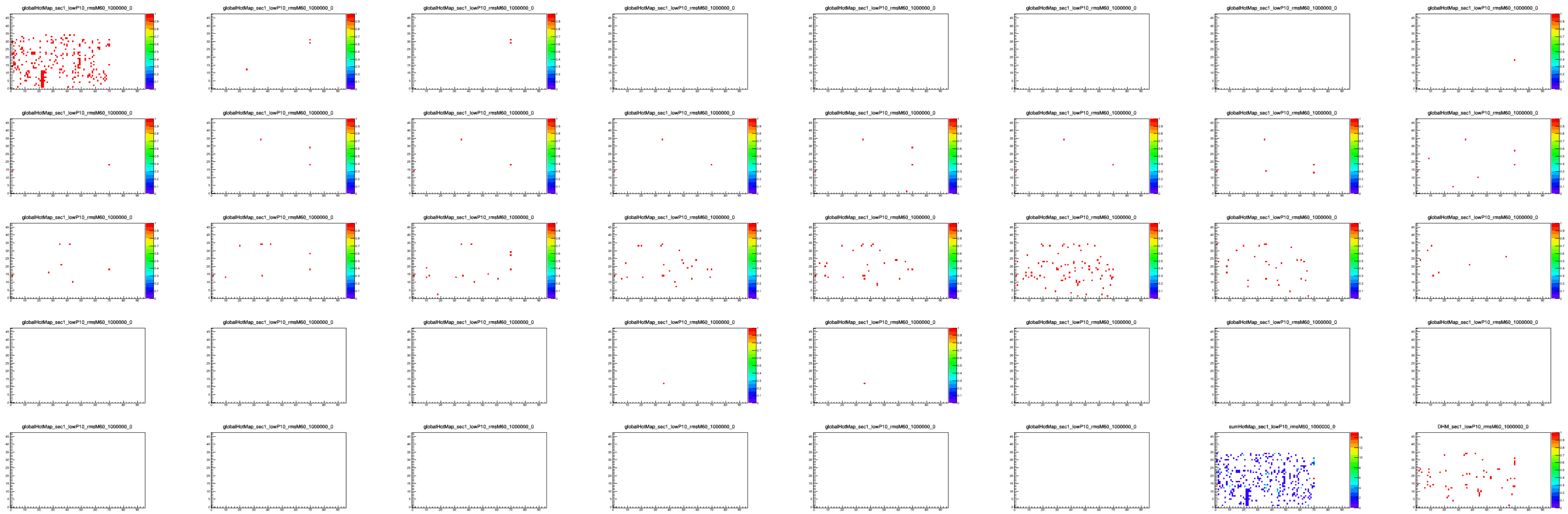
Run16 level DHM in one energy bin

- 800 good runs (low quality runs was removed from the 1000 runs)
- plot the dead towers from each run
- Keep the stable ones

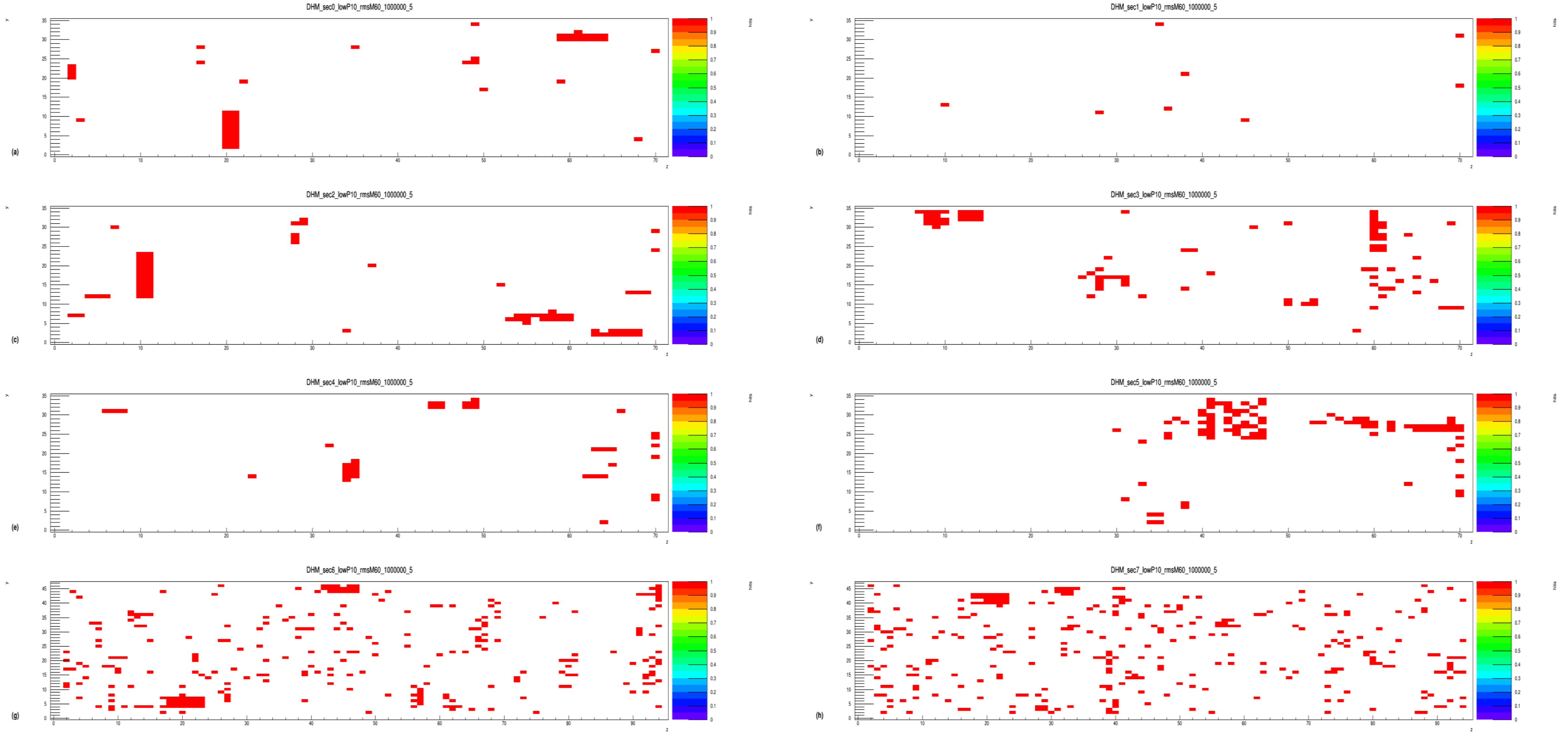


In every energy bin

- Phyisics is back, direct photon at high pT



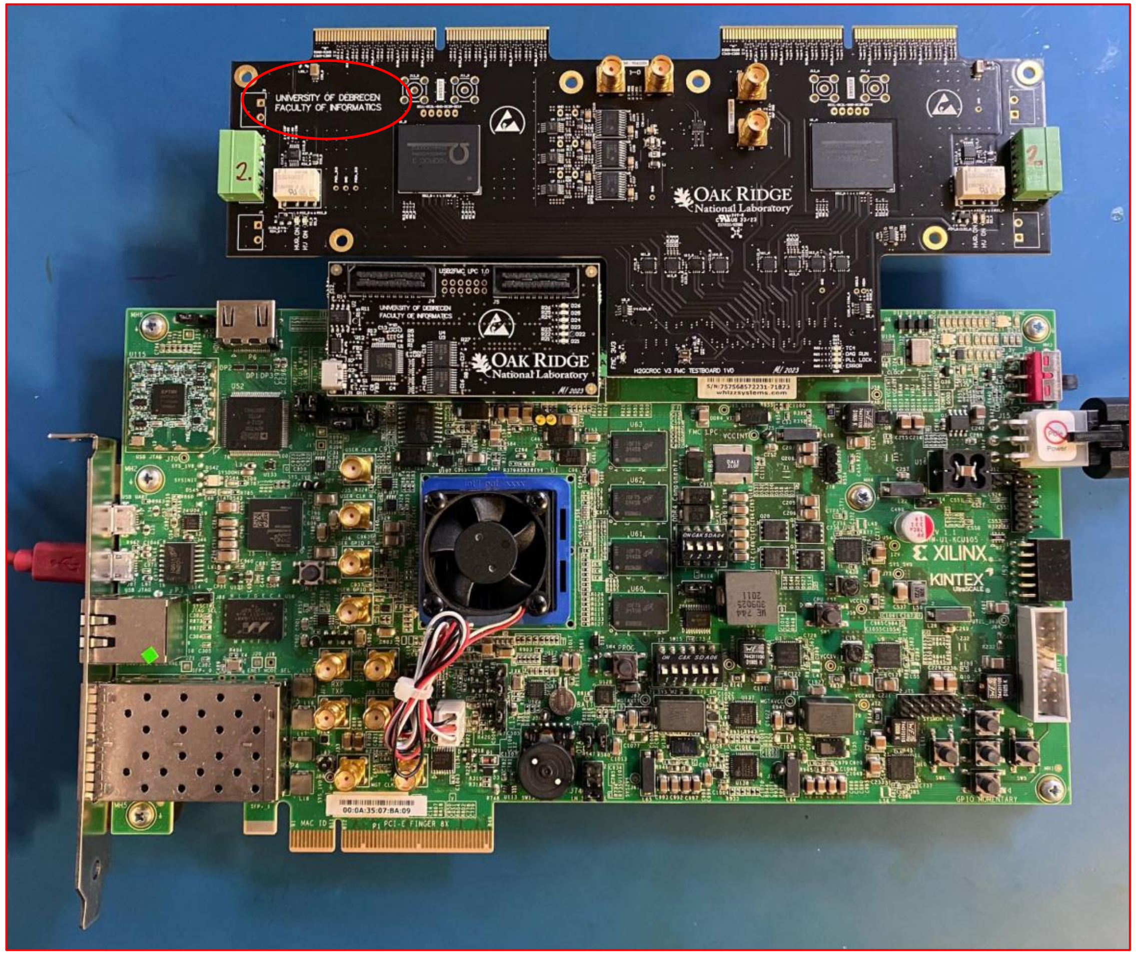
Final DHM



DHM_sec0_lowP10_rmsM60_1000000_5
DHM_sec1_lowP10_rmsM60_1000000_5
DHM_sec2_lowP10_rmsM60_1000000_5
DHM_sec3_lowP10_rmsM60_1000000_5
DHM_sec4_lowP10_rmsM60_1000000_5
DHM_sec5_lowP10_rmsM60_1000000_5
DHM_sec6_lowP10_rmsM60_1000000_5
DHM_sec7_lowP10_rmsM60_1000000_5
(a)
(b)
(c)
(d)
(e)
(f)
(g)
(h)

ePIC lfHCal

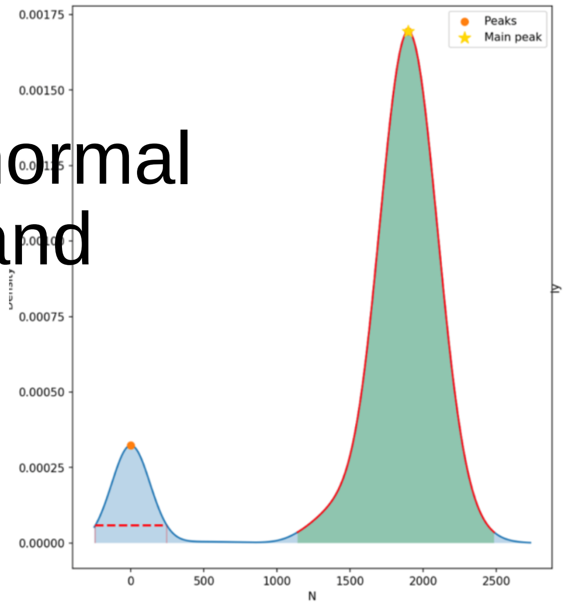
- lfHCal front-end development with ORNL
- lfHCal SiPM test
  - Prototype at Yale (100/day), final system
  - (~200k SiPM/years) at Yale/or and ORNL
  - We have to check at least 5%



UNIVERSITY OF DEBRECEN
FACULTY OF INFORMATICS

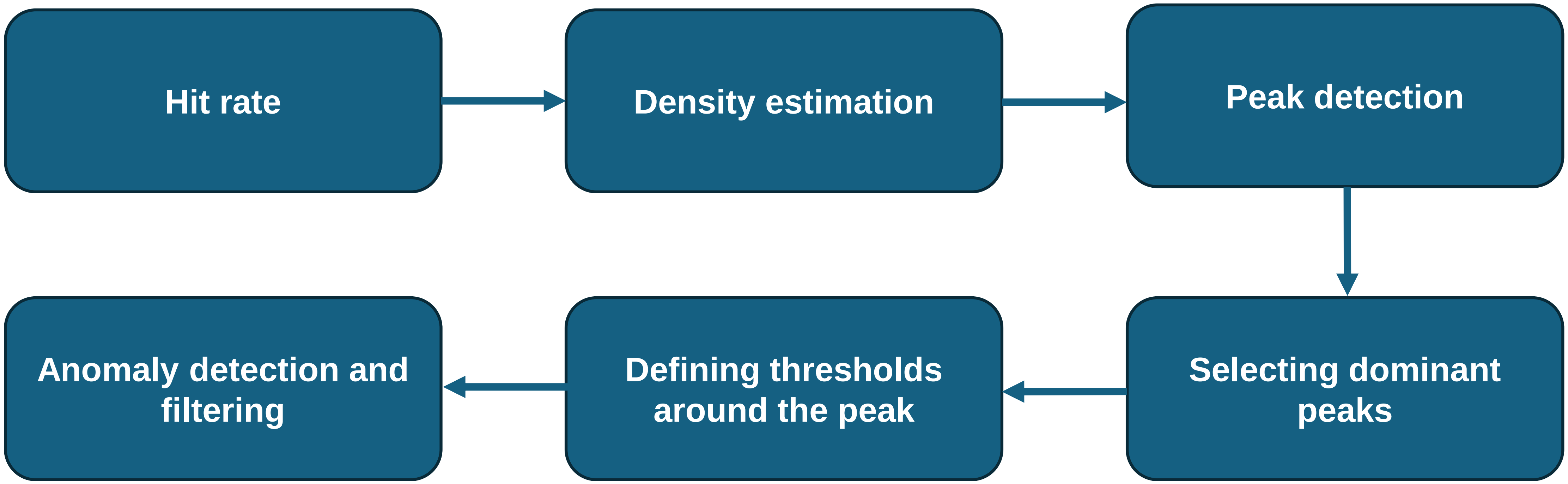
Density function

- A probability density function (PDF) is a function that shows how frequently a continuous random variable takes values in different ranges.
- The density function indicates how densely the values are distributed within an interval.
- The peak of the function corresponds to the most likely value, where observations tend to cluster.
- The method analyzes the distribution of hit rate, defines normal ranges based on the dominant peaks of the distribution, and treats any values that deviate from these as anomalies.



UNIVERSITY OF DEBRECEN
FACULTY OF INFORMATICS

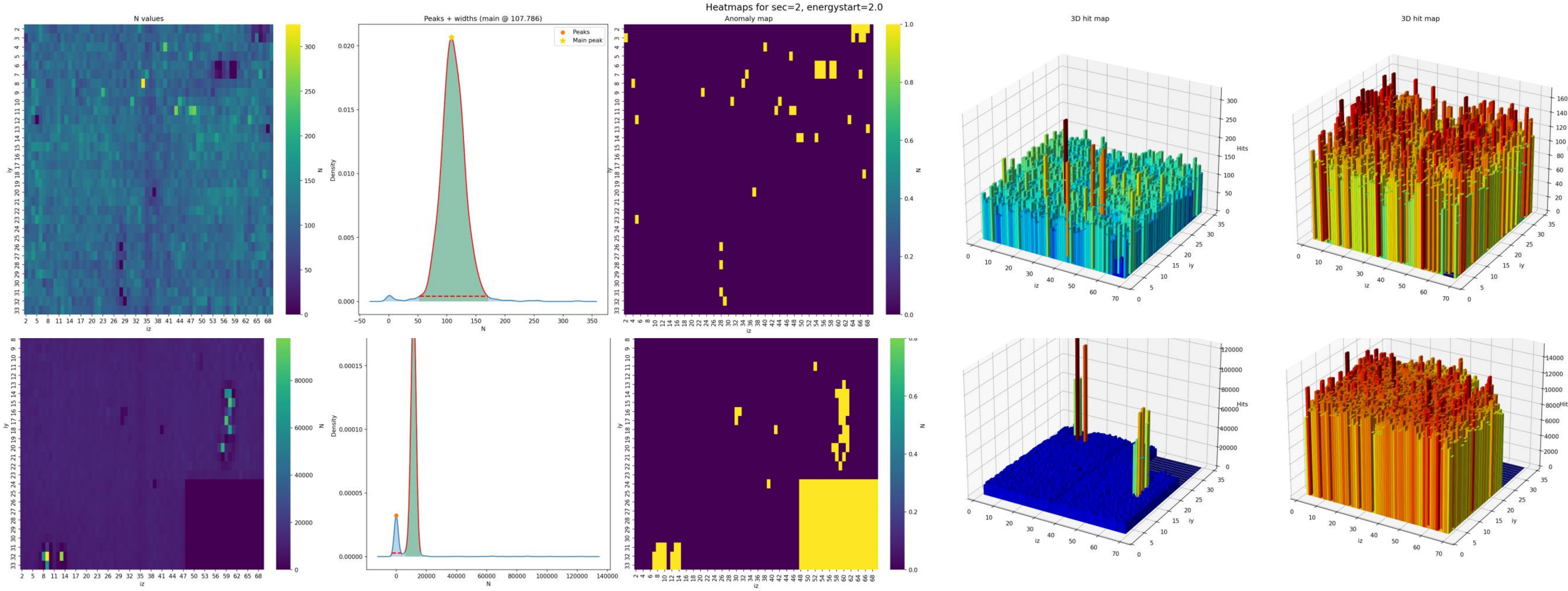
Description



Description

1. **Extracting hit rate**
2. **Density estimation:** Examining the distribution of keystroke counts using kernel density estimation (KDE), which reveals how densely values occur along the number line.
3. **Peak detection in the density function:** Identifying local maximum in the density curve, representing frequent and typical value ranges (concentration regions).
4. **Selecting dominant peaks :** Retaining only statistically significant modes. These define the system’s normal operating ranges. Currently, only the most dominant peak is used.
5. **Defining thresholds around the peak :** Determining cutoff values by analyzing the drop-off of the density curve. The tails beyond the peak boundaries represent anomaly zones.
6. **Anomaly detection and filtering**

Result

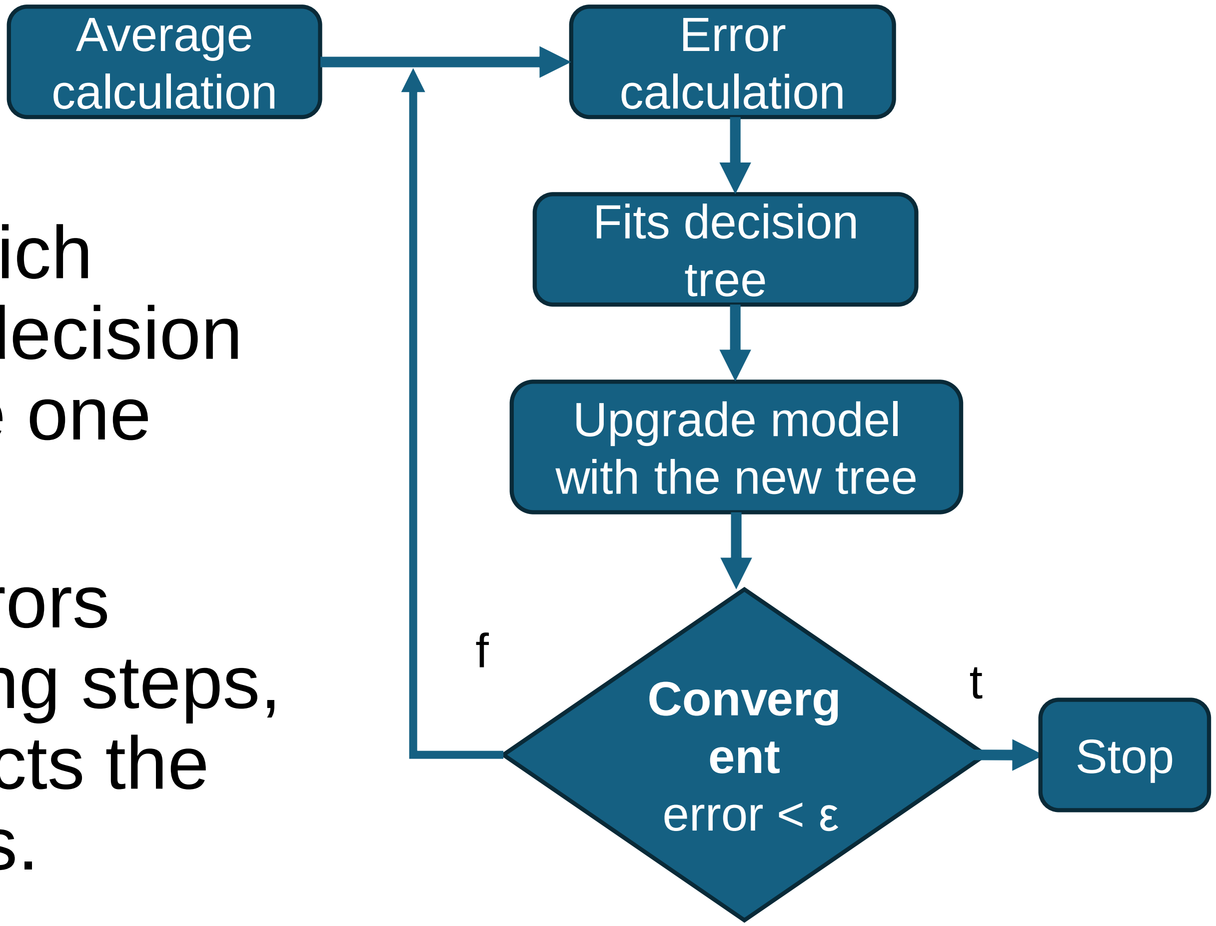


Heatmaps for sec=2, energystart=2.0
N values
Peaks + widths (main @ 107.786)
Peaks
Main peak
Anomaly map
3D hit map
3D hit map
Density
iy
iz
N
Hits

Anomaly detection by machine learning

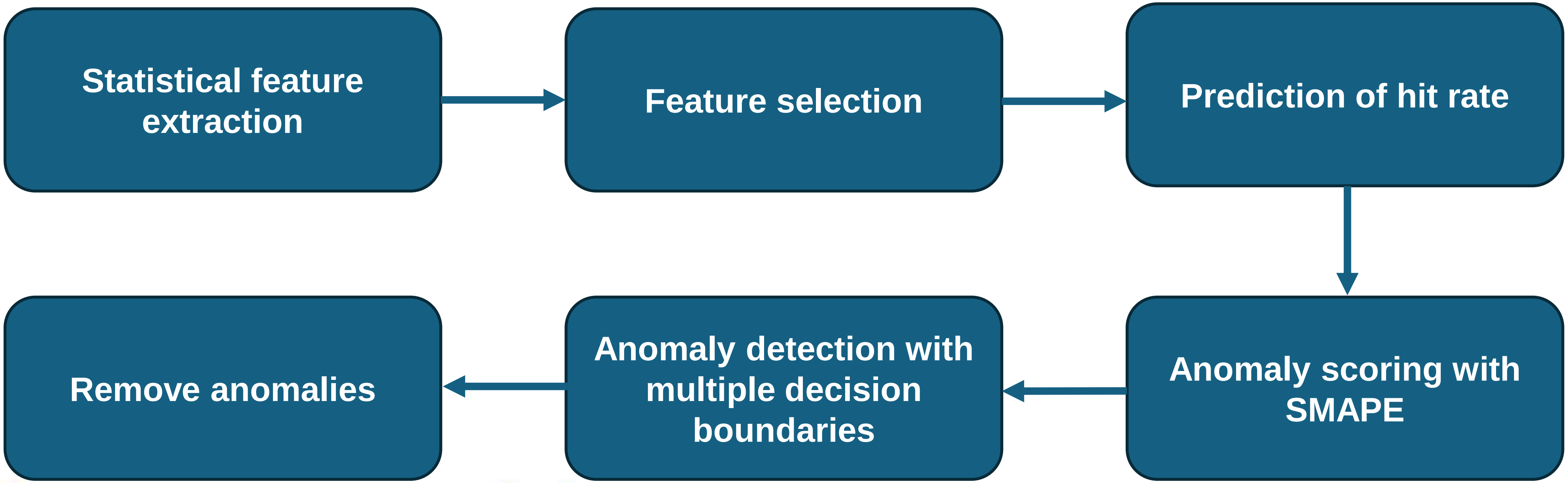
XGBoost

- eXtreme Gradient Boosting
- Boosting is a technique in which many weak models (usually decision trees) are combined to create one strong model.
- Goal: continuously reduce errors through a sequence of learning steps, where each new model corrects the mistakes of the previous ones.



UNIVERSITY OF DEBRECEN
FACULTY OF INFORMATICS

Anomaly detection by machine learning



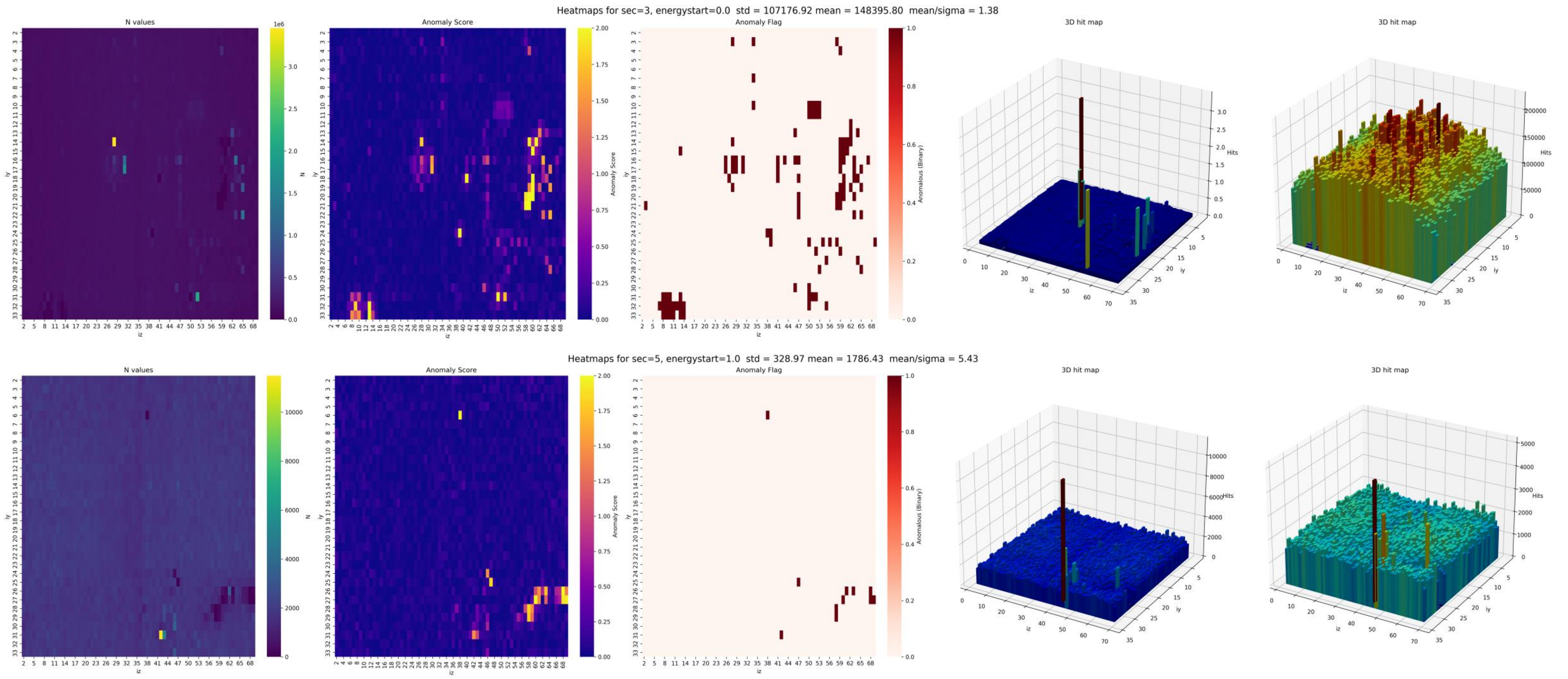
UNIVERSITY OF DEBRECEN
FACULTY OF INFORMATICS

Anomaly detection by machine learning

1. **Statistical feature extraction:**
   Computing descriptive metrics of the data: percentiles (1%, 95%, 99%), median, mean, standard deviation, skewness, kurtosis, entropy, ratio of non-zero values.
2. **Feature selection:**
   Sector, iy coordinate, iz coordinate, initial energy level, 99th percentile, median, 1st percentile, mean, standard deviation, skewness, kurtosis, entropy, ratio of non-zero values.
3. **Anomaly scoring using SMAPE (Symmetric Mean Absolute Percentage Error) metric**
   Determining the degree of anomaly based on the difference between predicted and actual hit rate using SMAPE.
4. **Anomaly detection with multiple decision boundaries**
   Applying a mean- and standard deviation-based threshold to identify actual anomalies.
5. **Removing detected anomalies from the data**

UNIVERSITY OF DEBRECEN
FACULTY OF INFORMATICS

Anomaly detection by machine learning



Anomaly detection assisted by DL

Preprocessing:

- Suspicious (anomalous) pixels are marked in the grids according to statistical rules:

  - dead: values lower than a given proportion of the grid average

  - hot: values that exceed the average by several standard deviations

  - extra hot: extremely large outliers

- We replace the marked pixels with the neighborhood average (imputation).

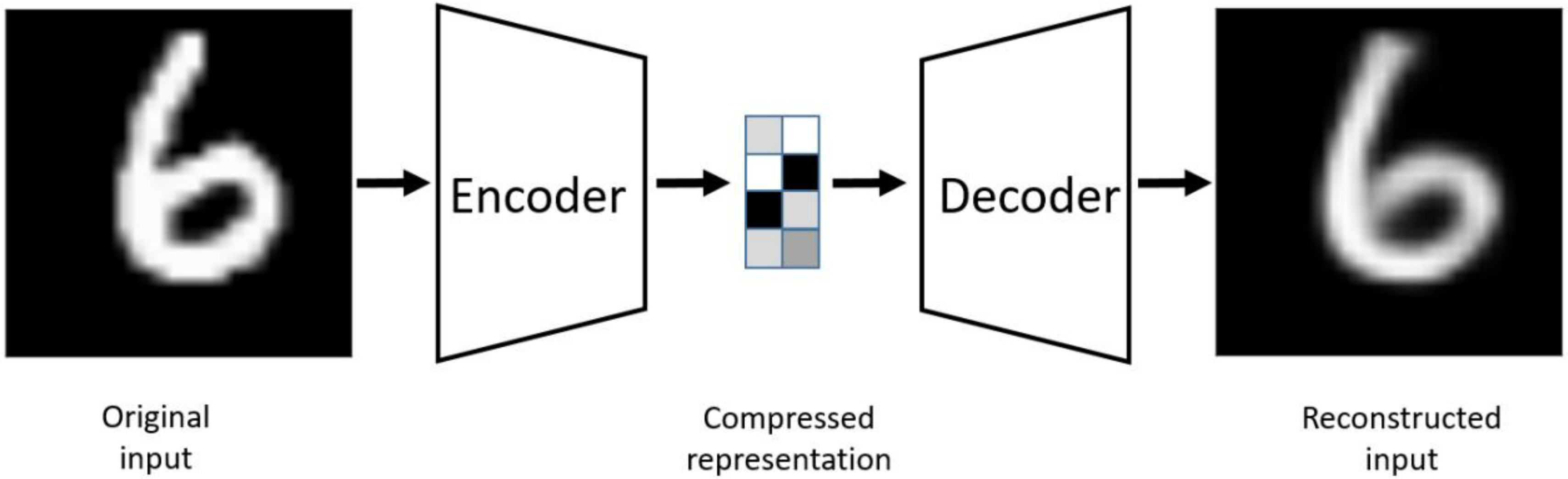
Robust scaling and normalization

- We normalize the grid reducing the impact of extreme values.
- We standardize the size of different grids (padding).

Anomaly detection assisted by DL

Convolutional Auto Encoder

- A small convolutional autoencoder:



- Goal: to learn typical, “normal” spatial patterns.

Anomaly detection assisted by DL

The training – Masked Reconstruction + Huber

- We set part of the input (mask_ratio) to zero, and the network only tries to reconstruct the omitted pixels.
- Loss function: Huber (Smooth L1) – robust to outliers.
- Early stopping: if the validation loss does not improve, we stop training.

Loss function (Huber / Smooth L1):

$$L_\delta(r) = \begin{cases} \frac{1}{2}r^2, & \text{if } |r| \le \delta \\ \delta(|r| - \frac{1}{2}\delta), & \text{otherwise} \end{cases}$$

$$r = X - \hat{X}$$

Anomaly detection assisted by DL

Inference – Residual and Thresholding

- The learned model performs reconstruction, the residual: $R = X - \hat{X}$

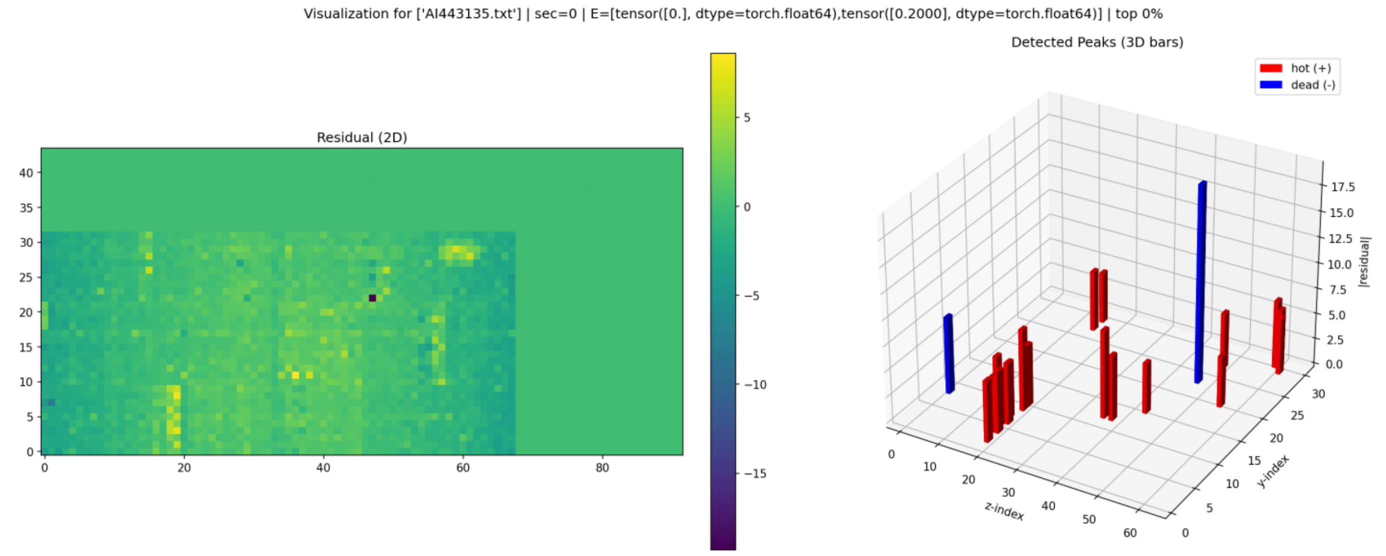
- Based on the distribution of the residual, we mark anomalies with percentile thresholds:

  - hot: above the upper percentile of the positive range

  - dead: below the upper percentile of the negative range

  - extra hot: extreme cases

- The top 1% of the largest errors is the peak.



UNIVERSITY OF DEBRECEN
FACULTY OF INFORMATICS

Thank you for your attention.

Questions?