

# Machine Learning Pipeline for Secondary Particle Classification for the LAGO Water Cherenkov Detectors

Modeling and Validation with Monte Carlo  
Simulation Datasets

**T. Torres Peralta (a,b,h), M. G. Molina (a,b,c,h), H. Asorey (d), I. Sidelnik (h,e), A. J. Rubio-Montero (g), R. Mayo-Garcia (g), A. Taboada (f), S. Dasso (i,j), L. Otiniano (k), for the LAGO Collaboration (l)**

(a) Tucuman Space Weather Center (TSWC), Argentina. (b) Facultad de Ciencias Exactas y Tecnología (FACET-UNT), Argentina; (c) Instituto Nazionale di Geofisica e Vulcanologia (INGV), Italy. (d) Medical Physics Department, Centro Atómico Bariloche, Comisión Nacional de Energía Atómica (CNEA), Argentina. (e) Dpto. de física de neutrones, Centro Atómico Bariloche, CNEA, Argentina. (f) Instituto de Tecnologías en Detección y Astropartículas (ITeDA), Argentina. (g) Centro de Investigaciones Energéticas Medioambientales y Tecnológicas (CIEMAT), Spain. (h) Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Argentina. (i) Laboratorio Argentino de Meteorología del Espacio (LAMP), Argentina. (j) Instituto de Astronomía y Física del Espacio (IAFE, UBA/CONICET) & Departamento Cs Atmósfera y Océanos (DCAO, FCEN, UBA) Argentina. (k) Comisión Nacional de Investigación y Desarrollo Aeroespacial (CONIDA), Peru. (l) The LAGO Collaboration, see the complete list of authors and institutions at <https://lagoproject.net/collab.html>  
<https://lagoproject.net/collab.htm>

CONTACT

[ttorres@herrera.unt.edu.ar](mailto:ttorres@herrera.unt.edu.ar)

- Single (in most of the sites) WCDs deployed at different lat and altitudes
- WCDs: different geometry and materials. Autonomous, reliable, simple, smart, and inexpensive. Own designed electronics
- Own software suite: data analysis (at different levels), data sharing, integrated simulations (ARTI+CORSIKA+MEIGA-Geant4), ML, etc.



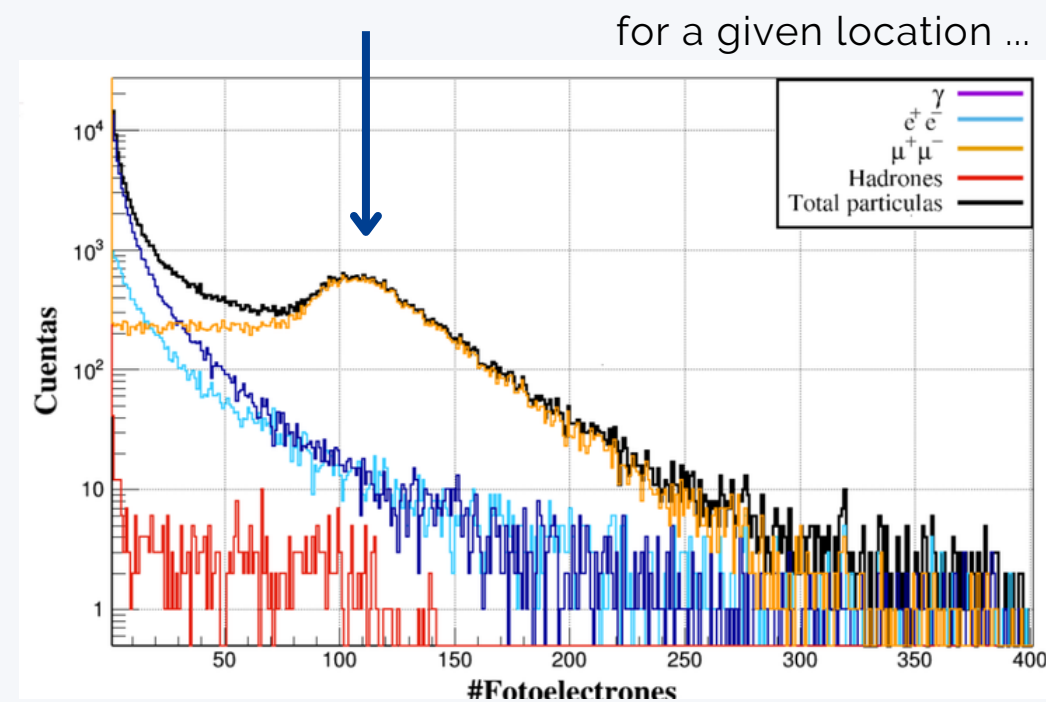
<http://lagoproject.net>

Extended astroparticle observatory at continental scale: from Mexico to Antarctica

Scientific goals:

- Astroparticle physics (extreme universe)
- Transient and long term space weather phenomena trough Solar modulation of CR
- Measurement of background radiation at ground level

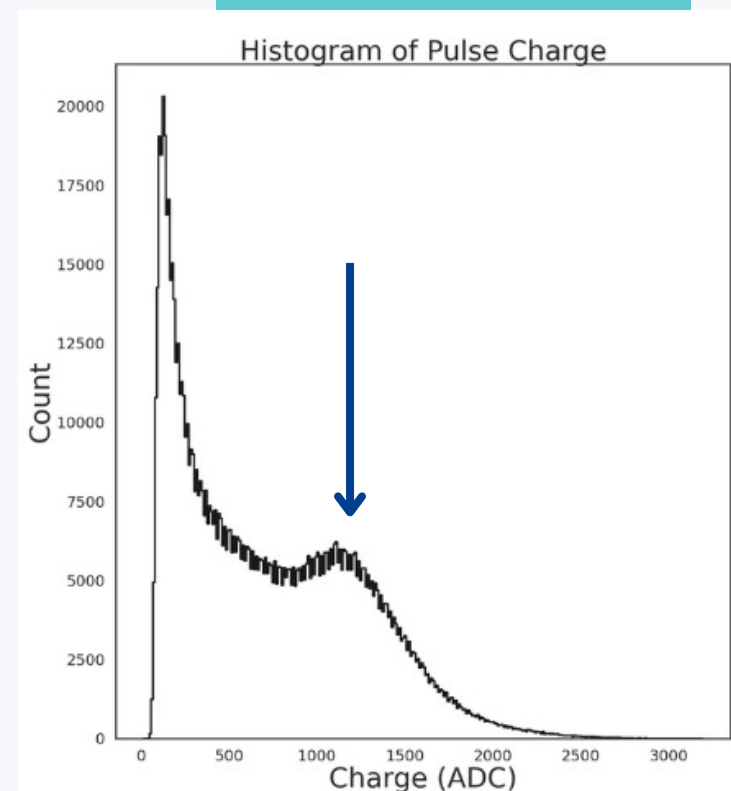
+ Academic goals



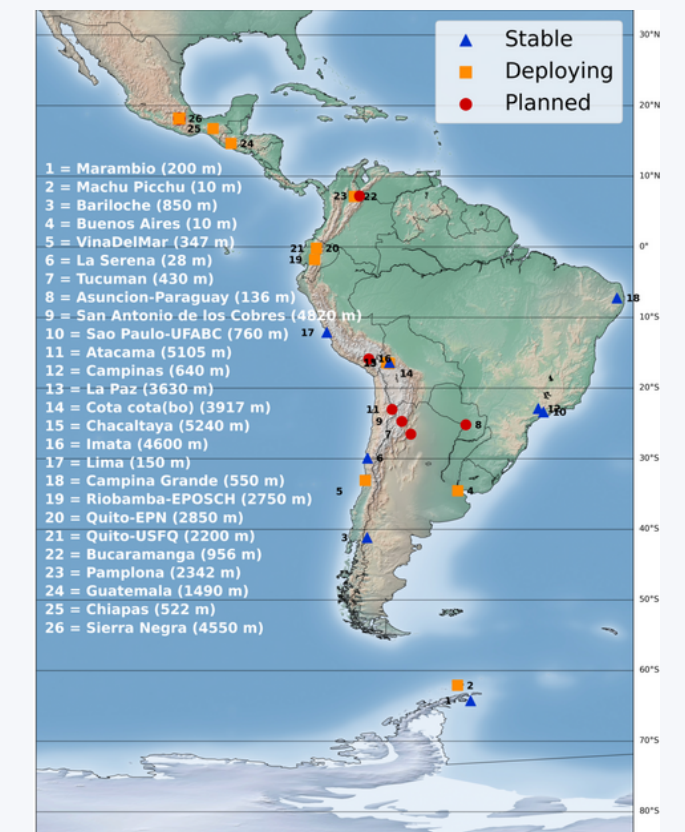
**Simulation**

from ARTI+  
CORSIKA

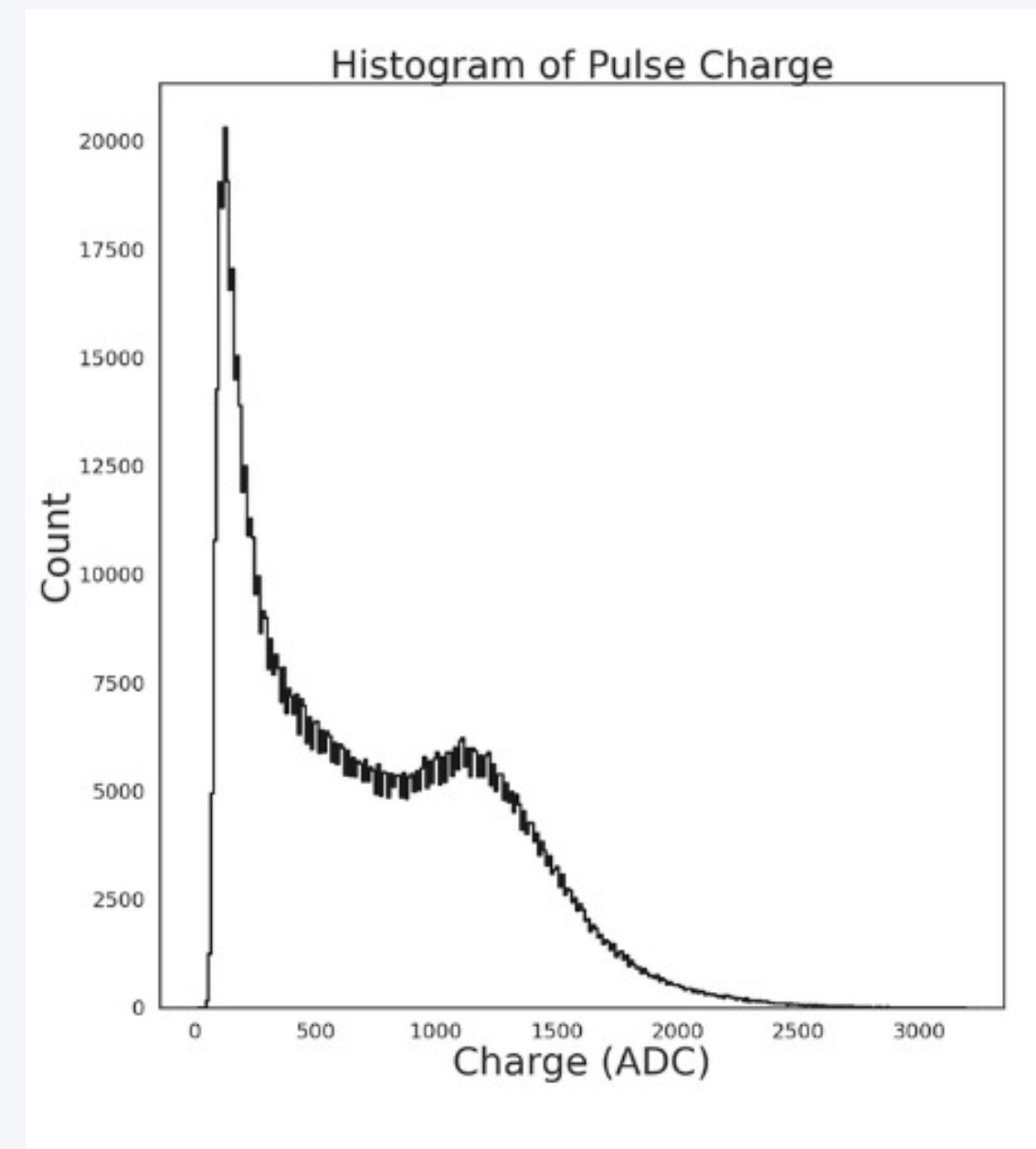
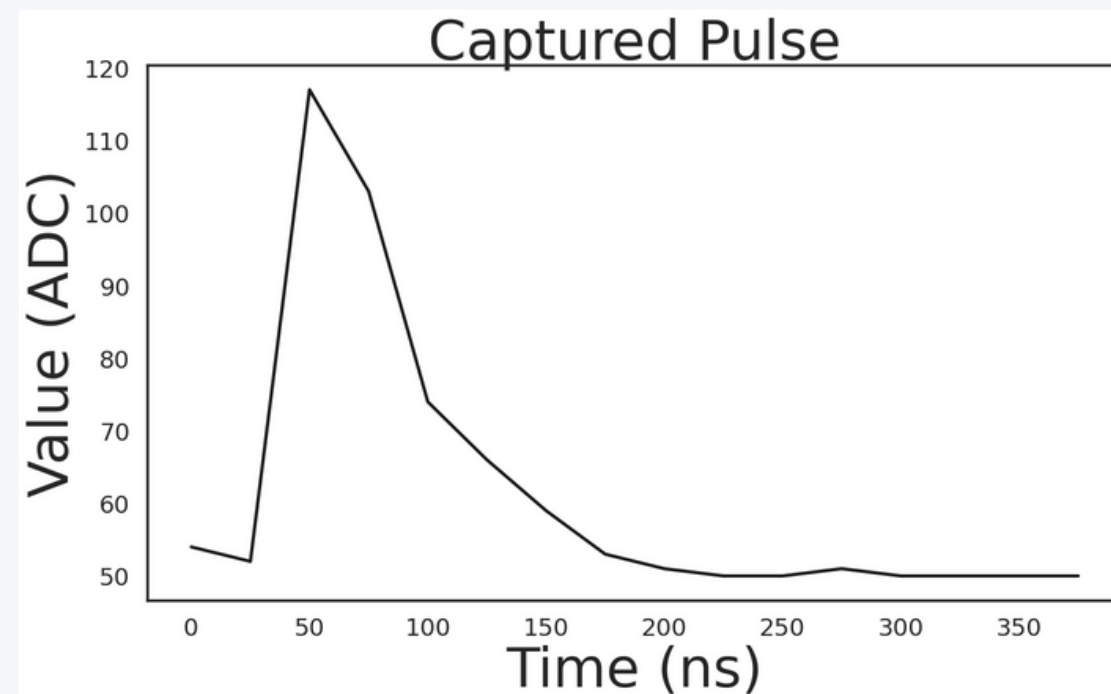
**Actual data**



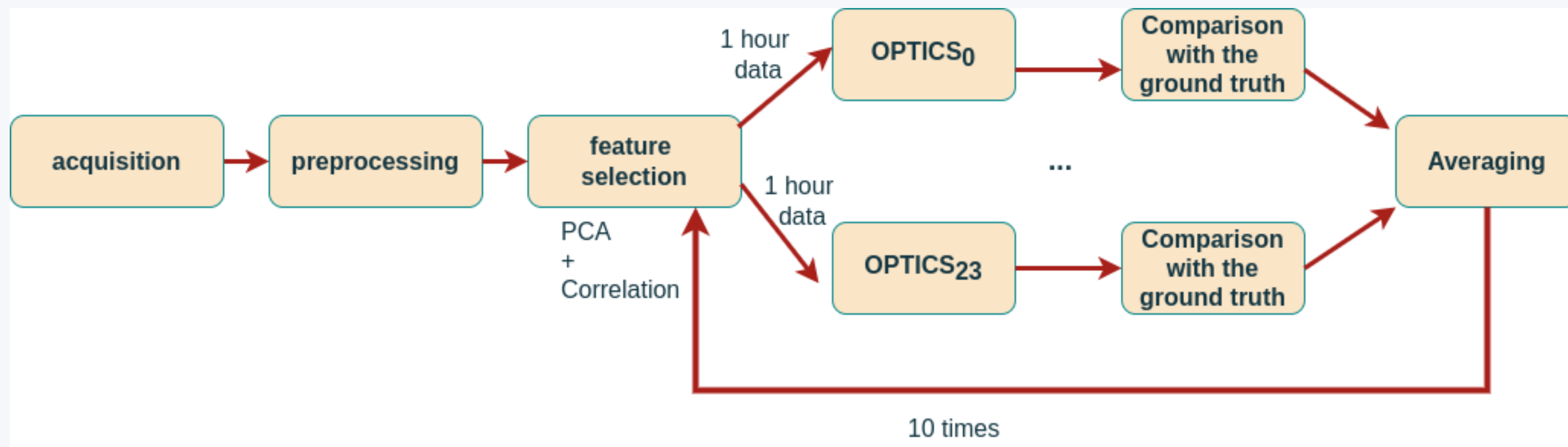
from Anna (LAGO's data  
analysis toolbox)



- All observations are transient pulses produced by photonic emissions from the secondary particles passing through the tank, and captured by the photomultiplier and electronics.
- The captured pulses by WCD has limited resolution and provide no direct way to discriminate between secondary particles.



- We proposed a data science approach where we use machine learning (ML) to implement a data-driven model and processing pipeline. The main protagonist of this pipeline is a hierarchical density-based unsupervised machine learning method for clustering pulses based on similarity patterns, called OPTICS (Ordering Points to Identify the Clustering Structure).

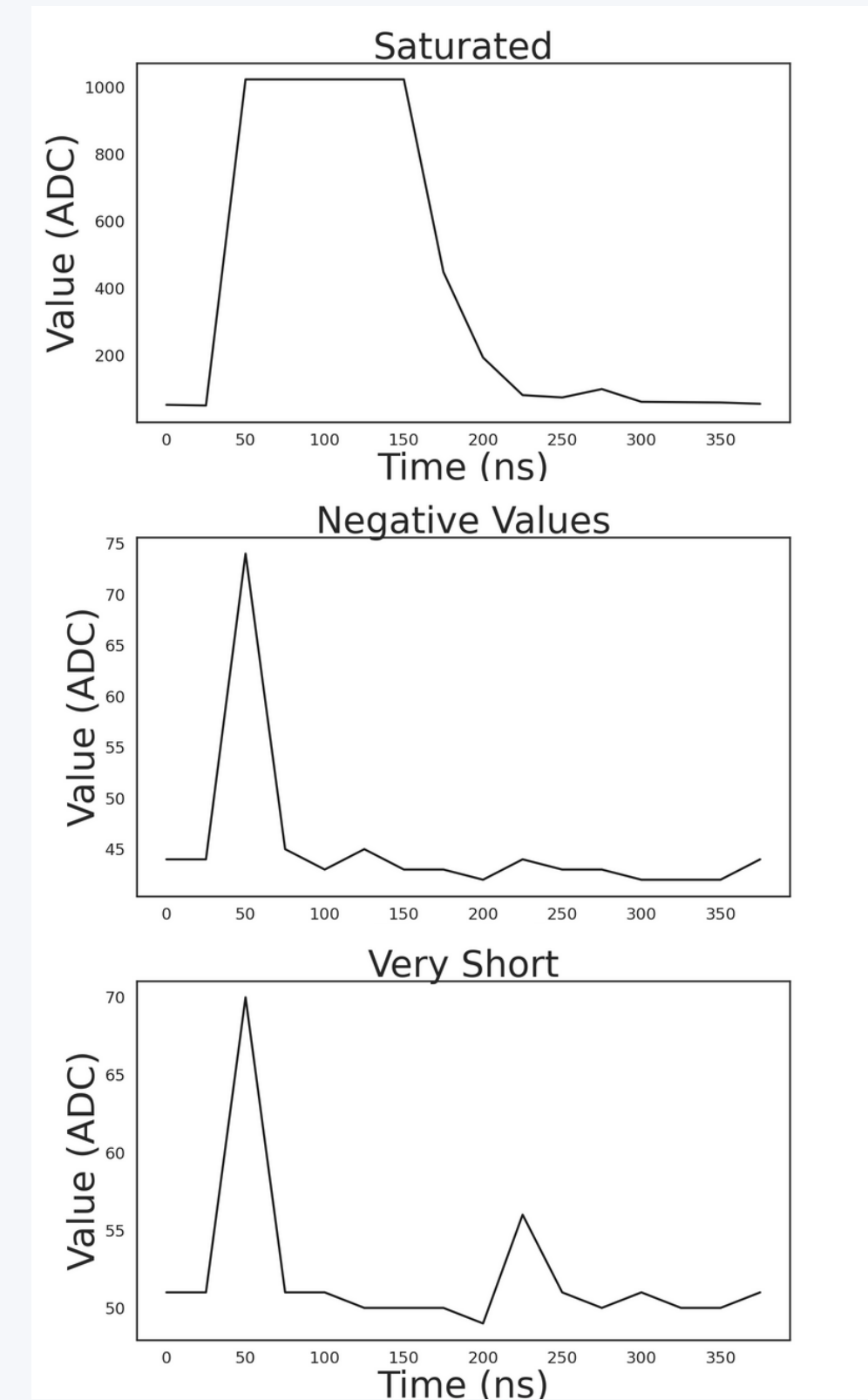


### Real Data

- Raw data from “Nahuelito” WCD site at Bariloche, Argentina. Total of 24hs of data between 13:00, 01 of March of 2012 and 12:00, 2 of March of 2012.
- Pulses have 10 bits – 25 ns resolution.
- ~39 millions preserved after preprocessing (~ 40%). Each hour contained ~1.6 million events.
- A random sample of 1 million was used for the machine learning stage.

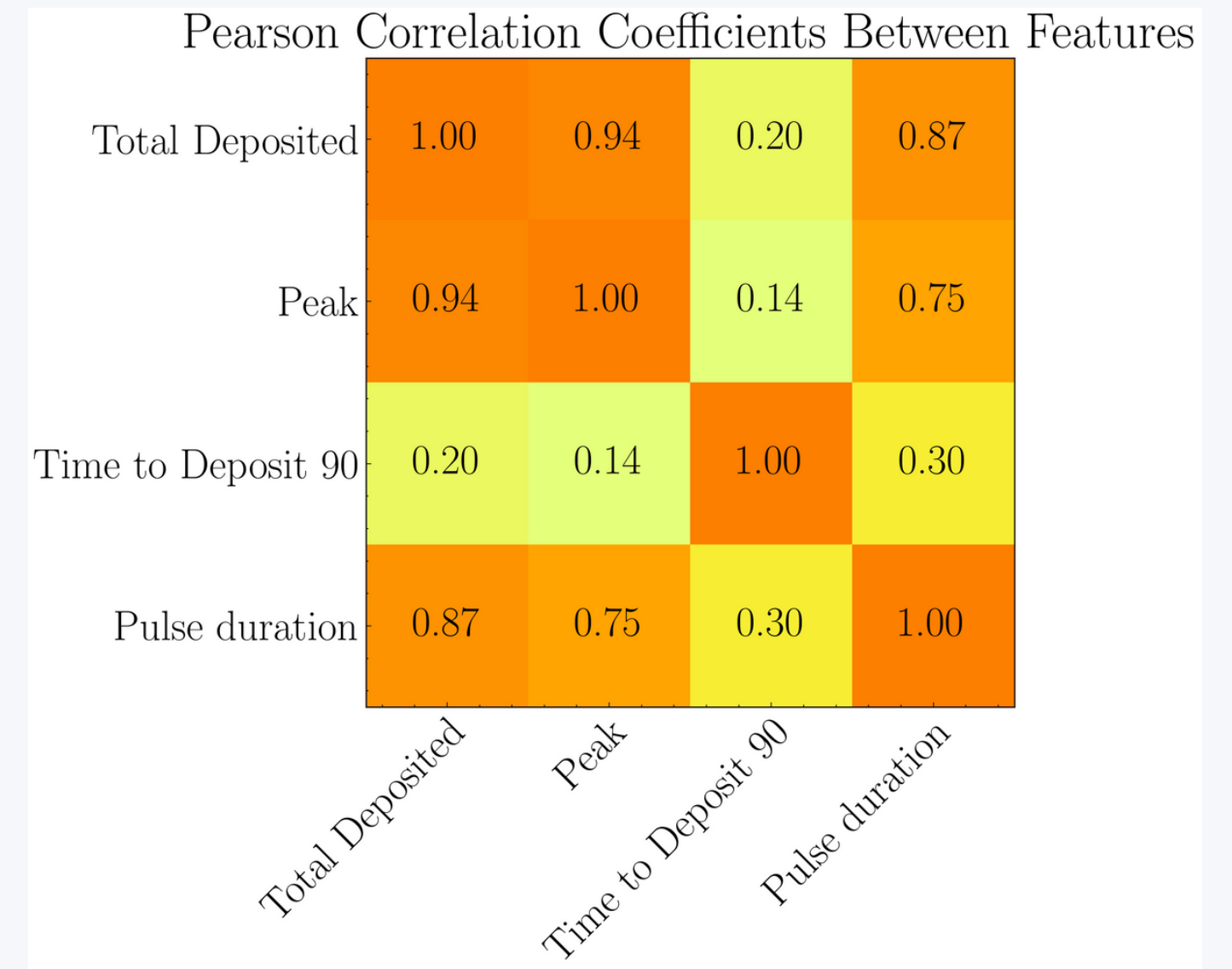
### Simulated Data

- Monte Carlo simulations consist of 24 hours of synthetic data for similar conditions as the actual data: March of 2012 at “Nahuelito” WCD.
- Not all simulated events produced a pulse at the WCD.
- Events with no pulse were eliminated from dataset, resulting in ~24 million events after preprocessing.

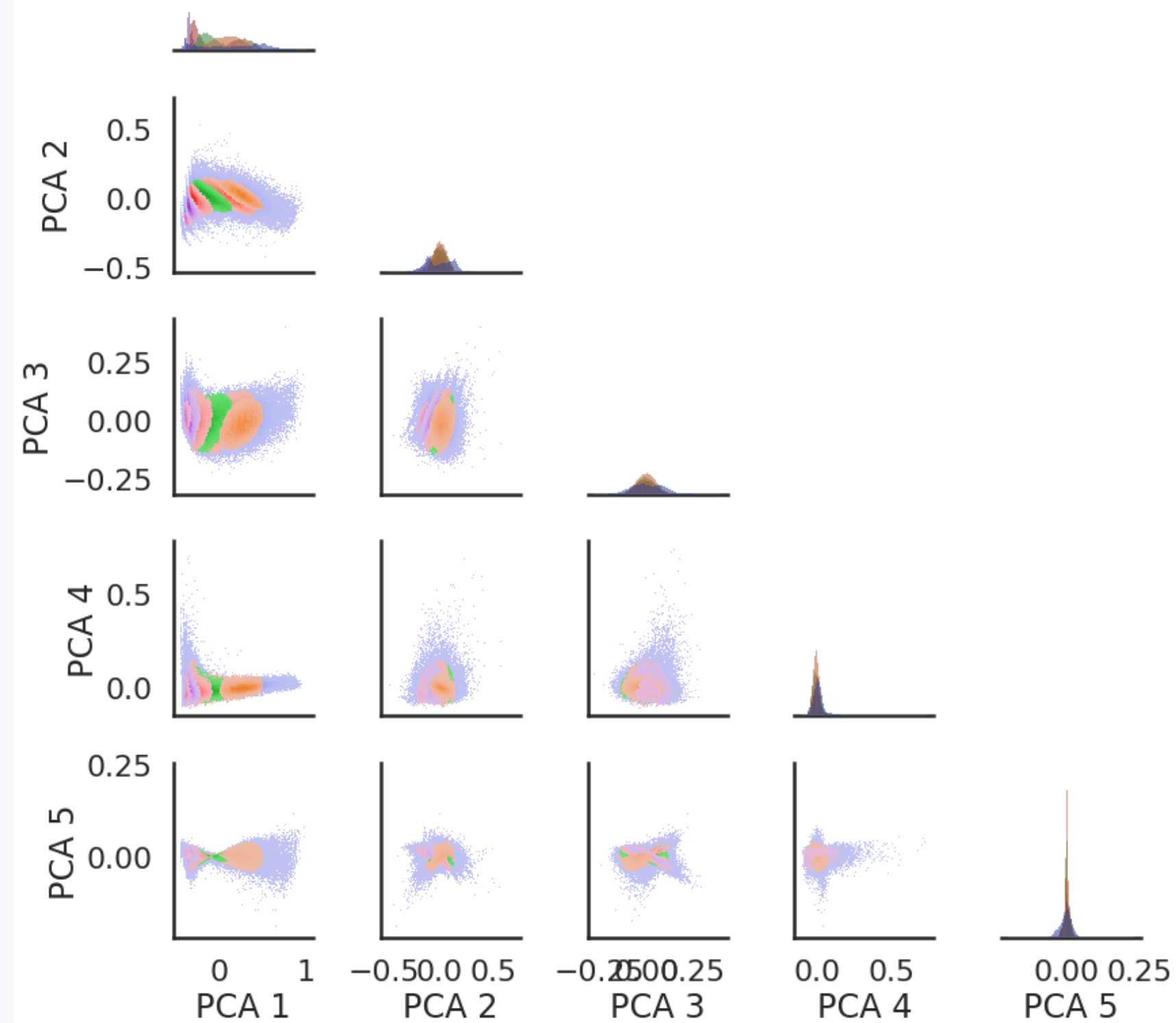
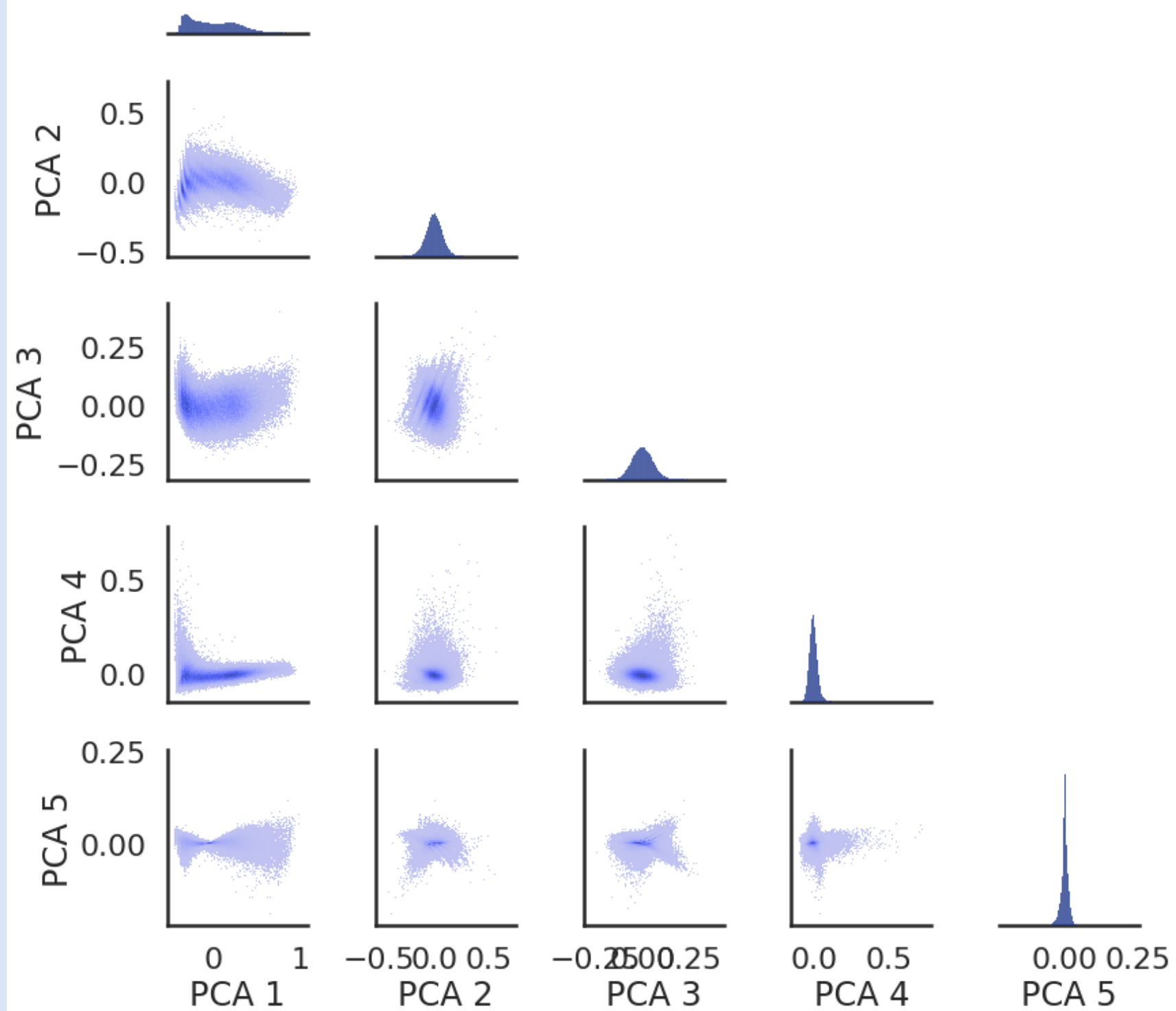


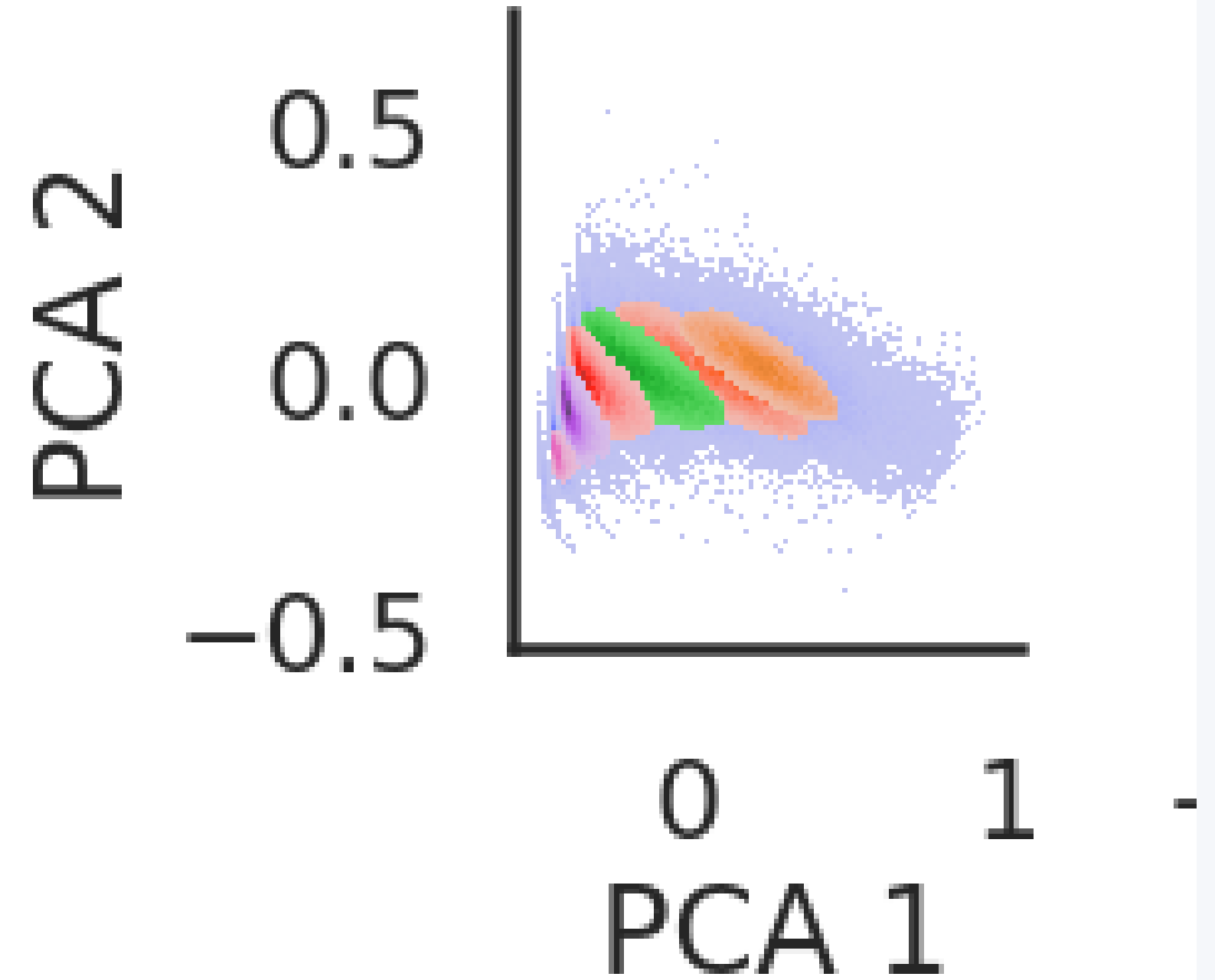
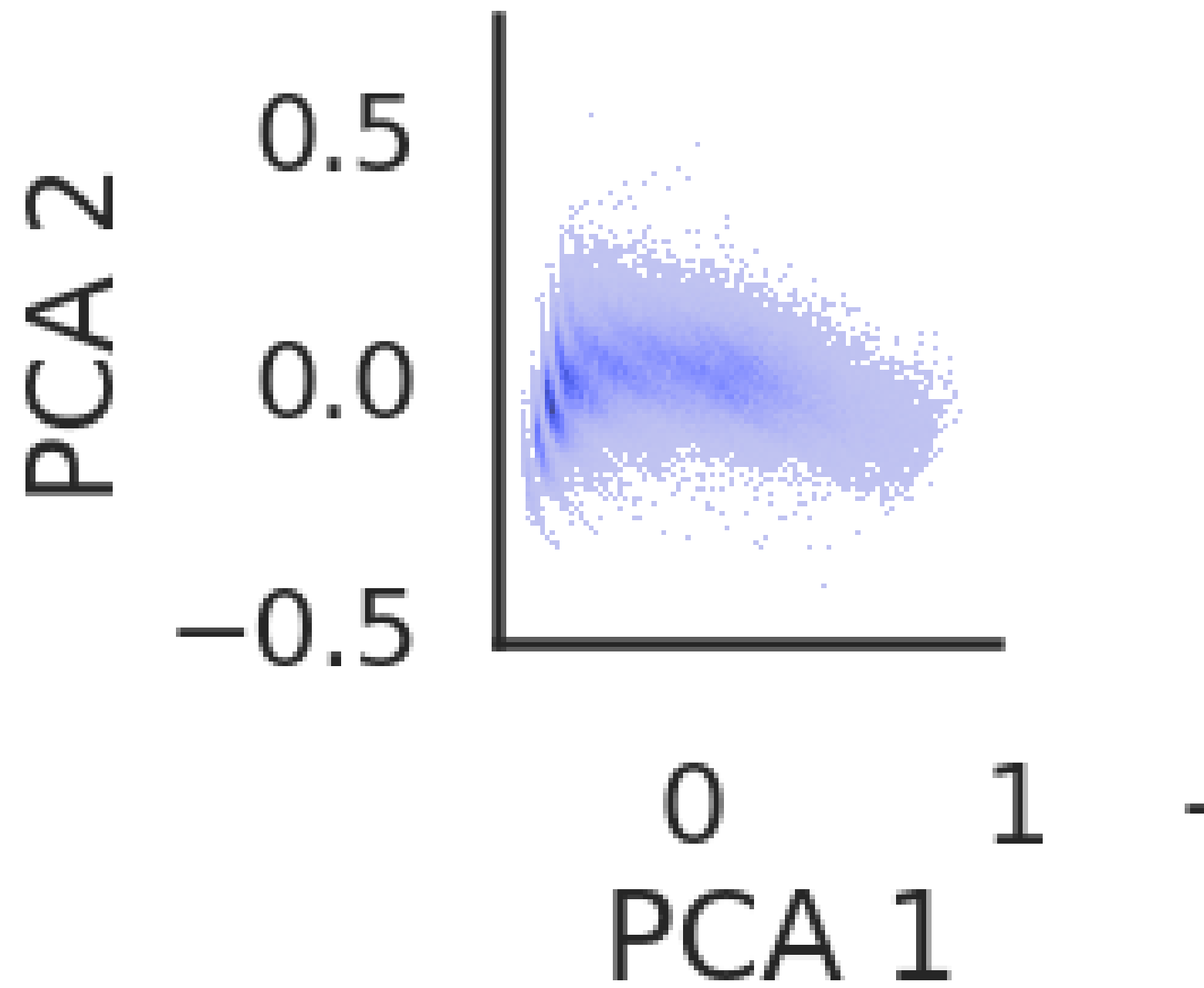
- “Peak” feature was eliminated due to very high correlations with “Total Deposited”.
- Final feature list: “Total Deposited”, “Time to Deposit 90%”, “Pulse Duration”

| Feature                | Description                                                                        |
|------------------------|------------------------------------------------------------------------------------|
| Total Deposited Energy | Total Photoelectrons (PE) that where deposited by the pulse, in total count of PE. |
| Time to Deposit        | Time the pulse took to depoite 90% of its PEs, in ns.                              |
| Pulse Duration         | Duration of the pulse, in ns.                                                      |

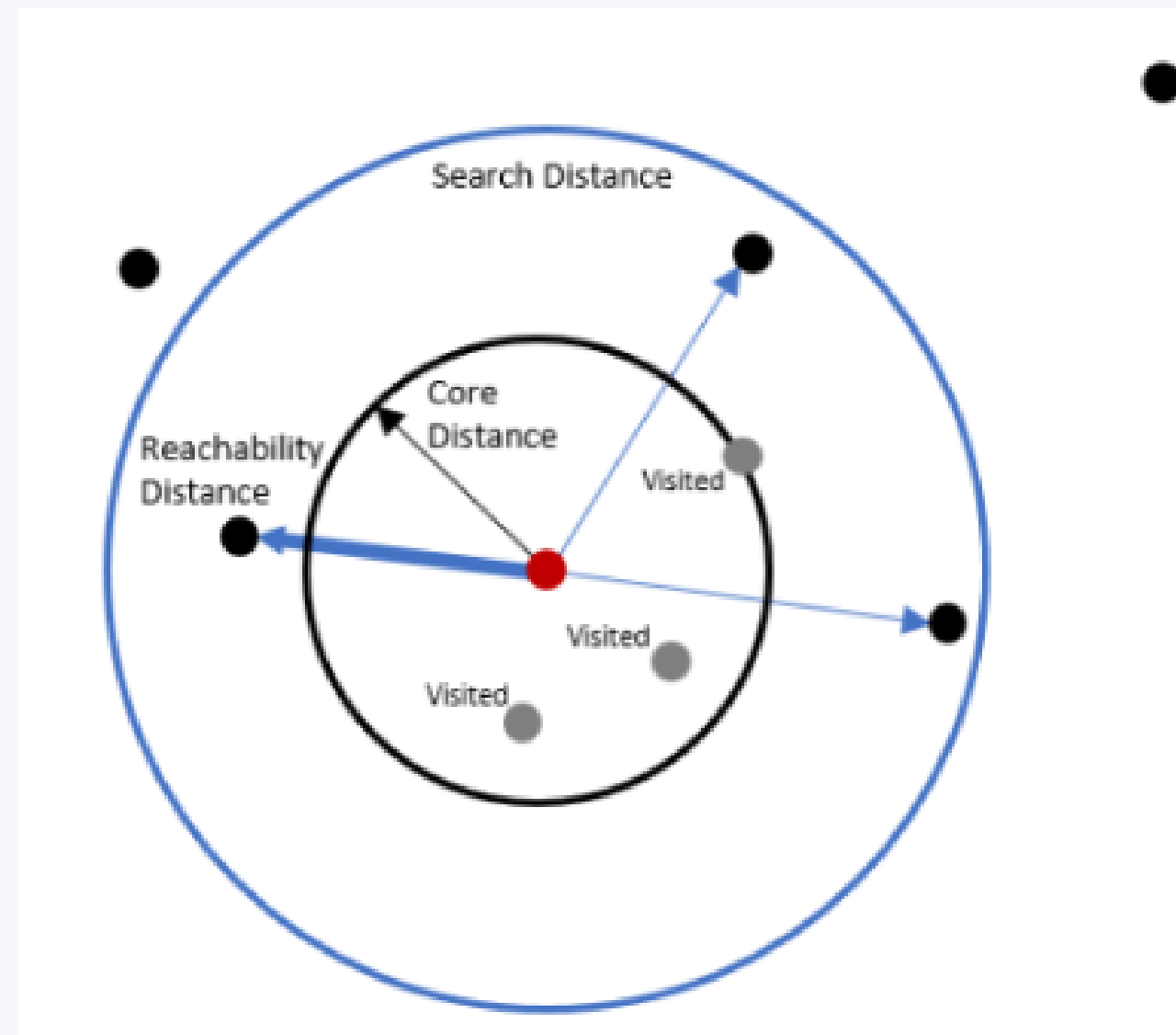


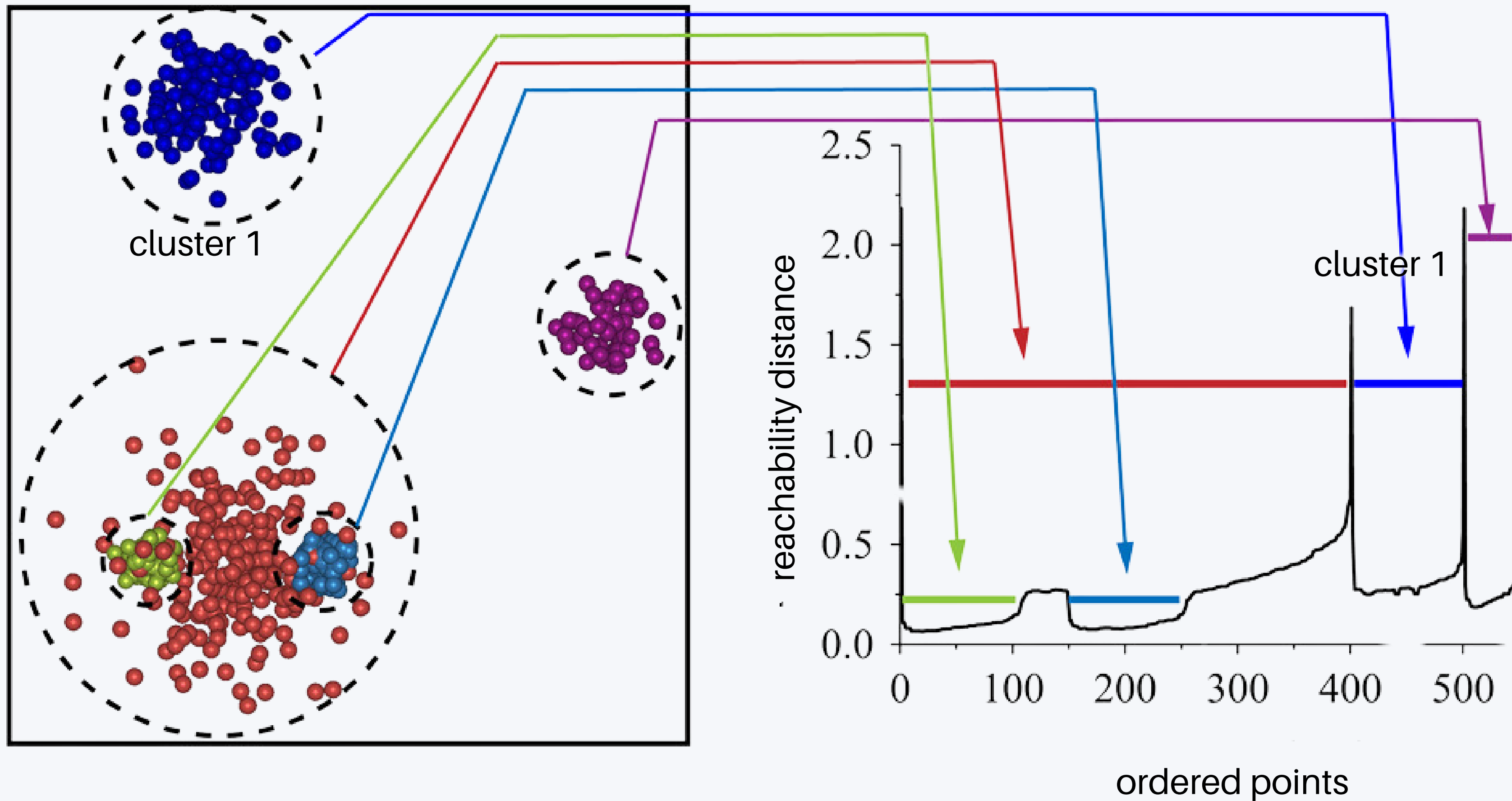
- Original Features used in exploration (Real Data):
  - Pulse charge
  - Pulse peak
  - Charge deposit time
  - Pulse width
  - Time delta with next pulse
  - Time delta with previous pulse
- Multidimensional feature space was very noisy. Applied Normalization and Principal Component Analysis to improve separability of feature space.



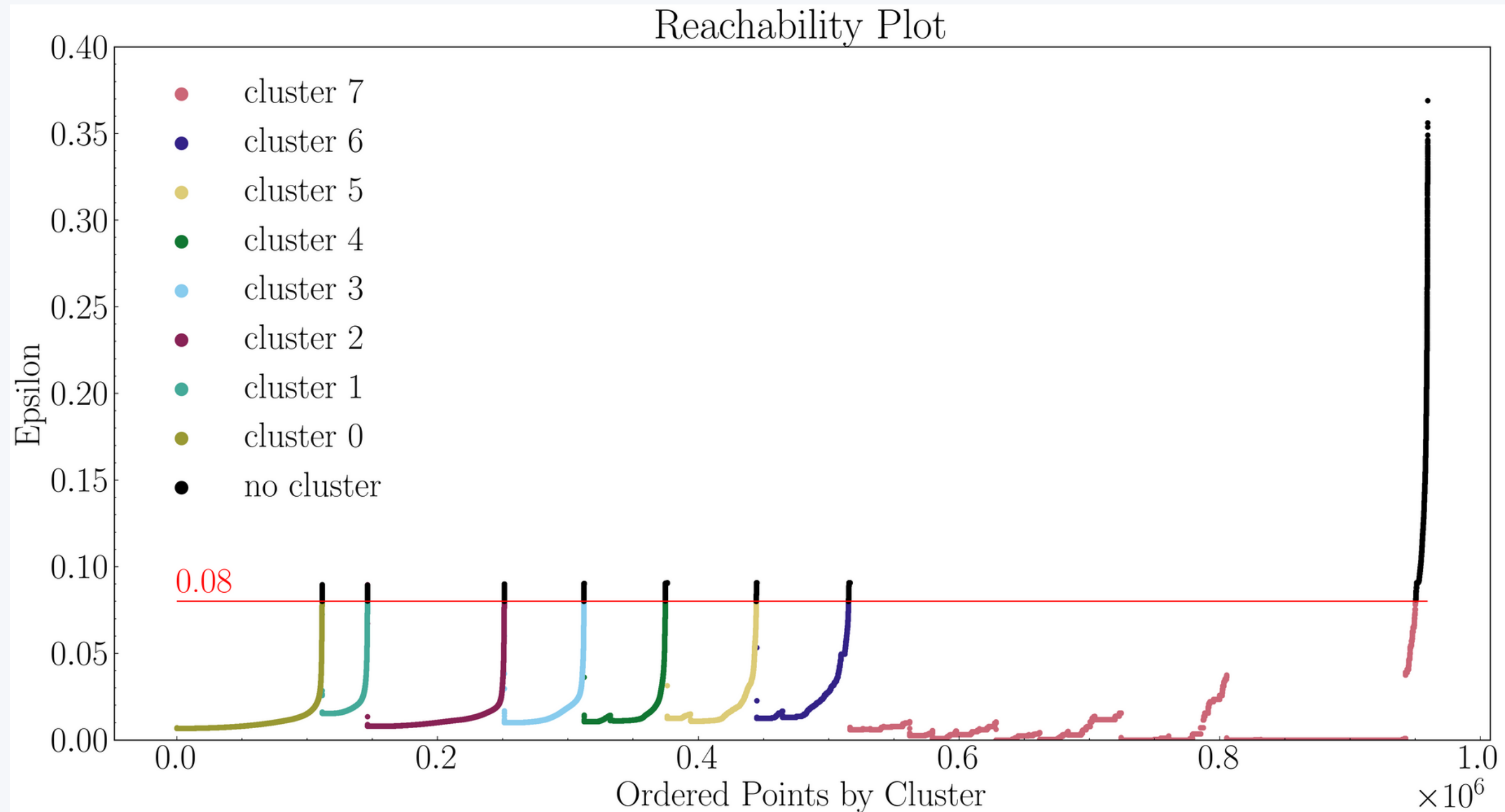


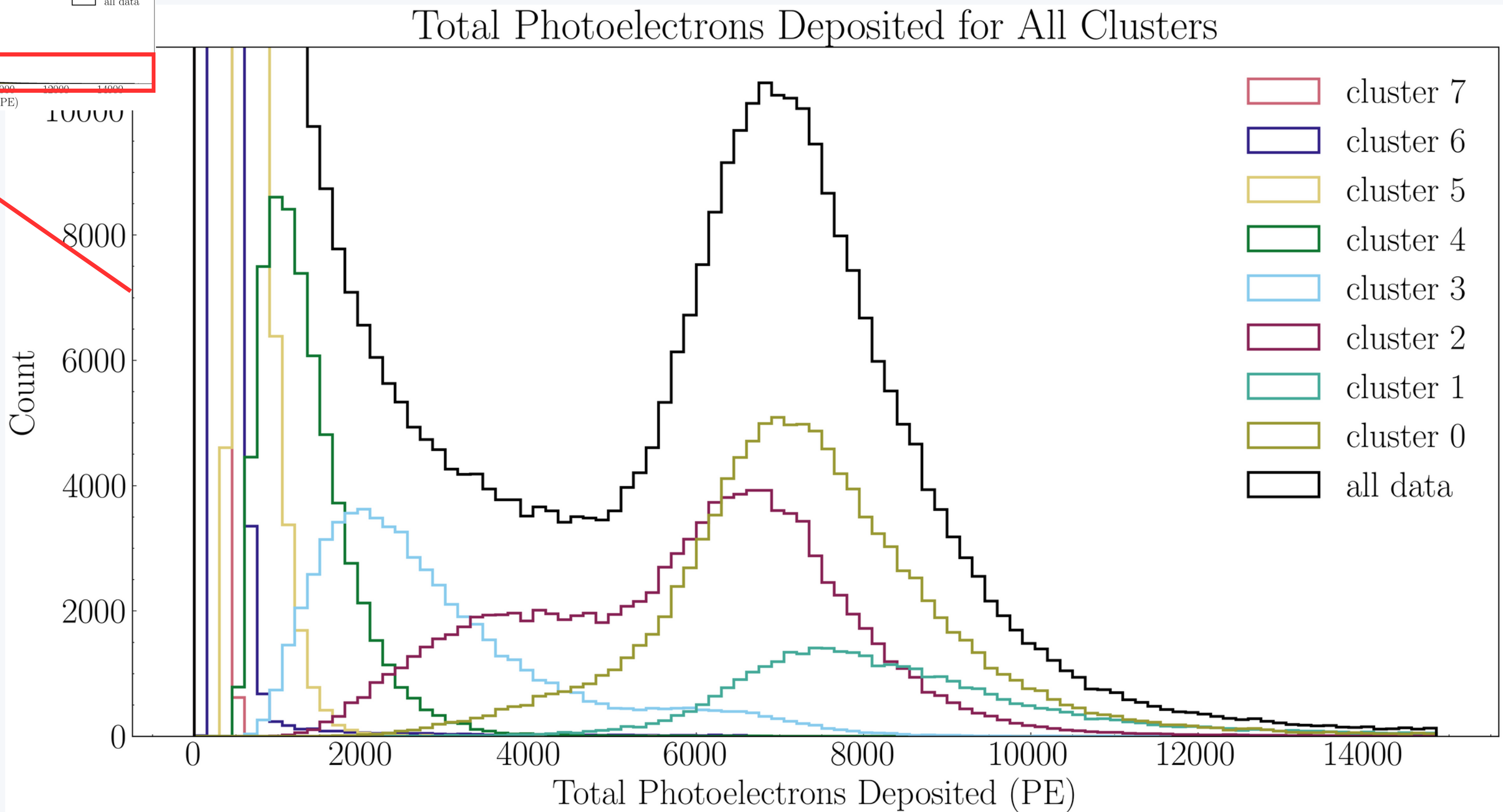
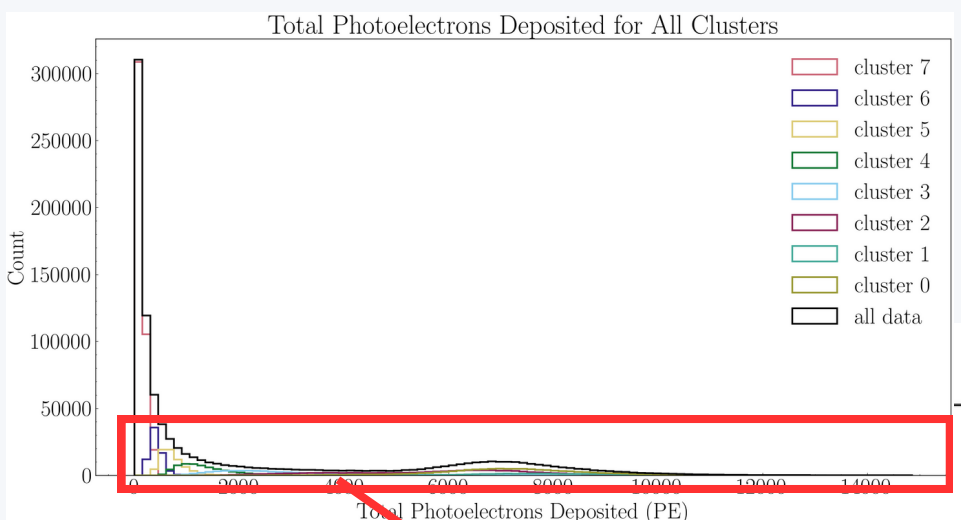
- Hierarchical density-based clustering algorithm Ordering Points To Identify Clustering Structure (OPTICS).
- It defines the reachability-distance, a minimum distance that helps describe cluster structure, and creates a reachability plot.

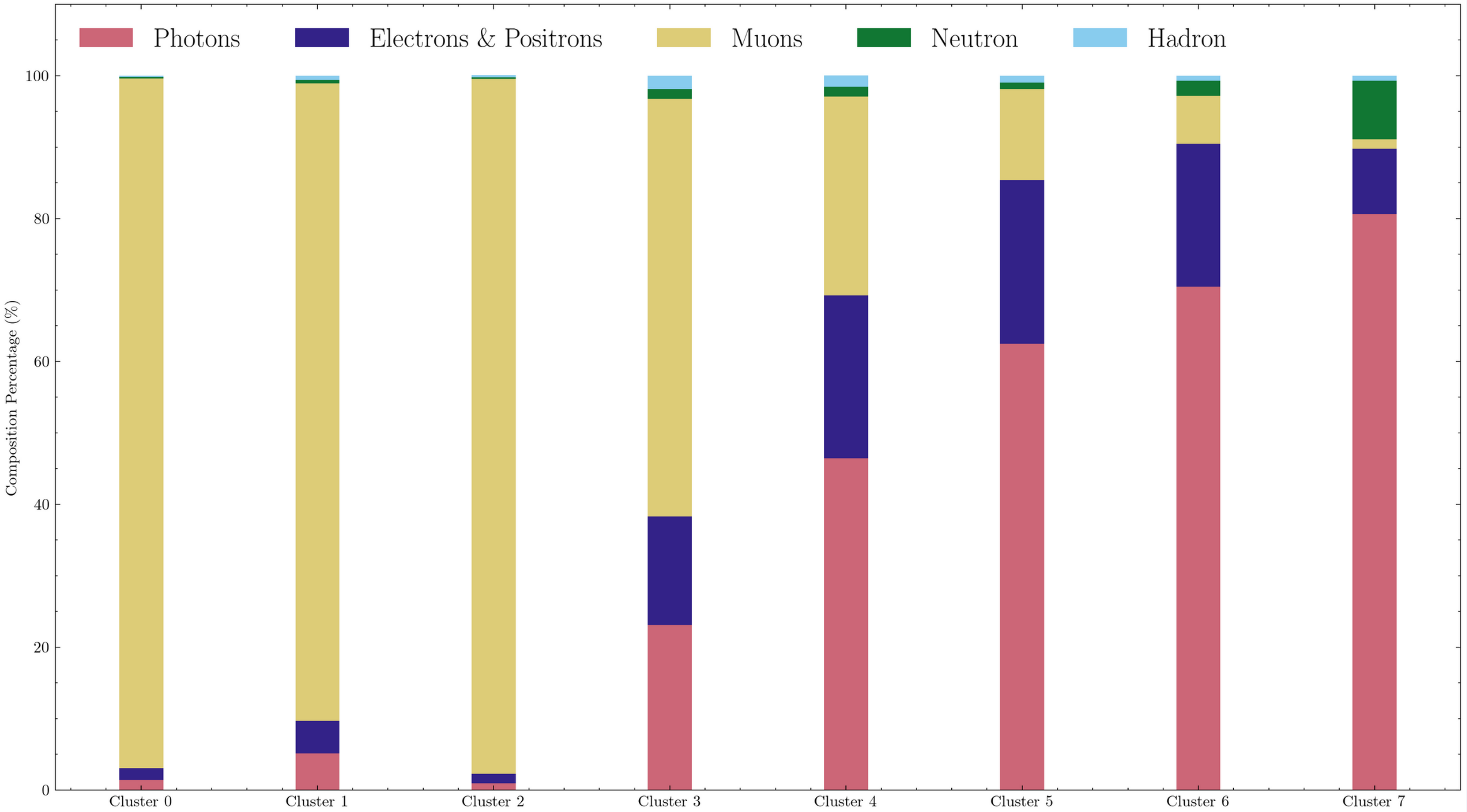


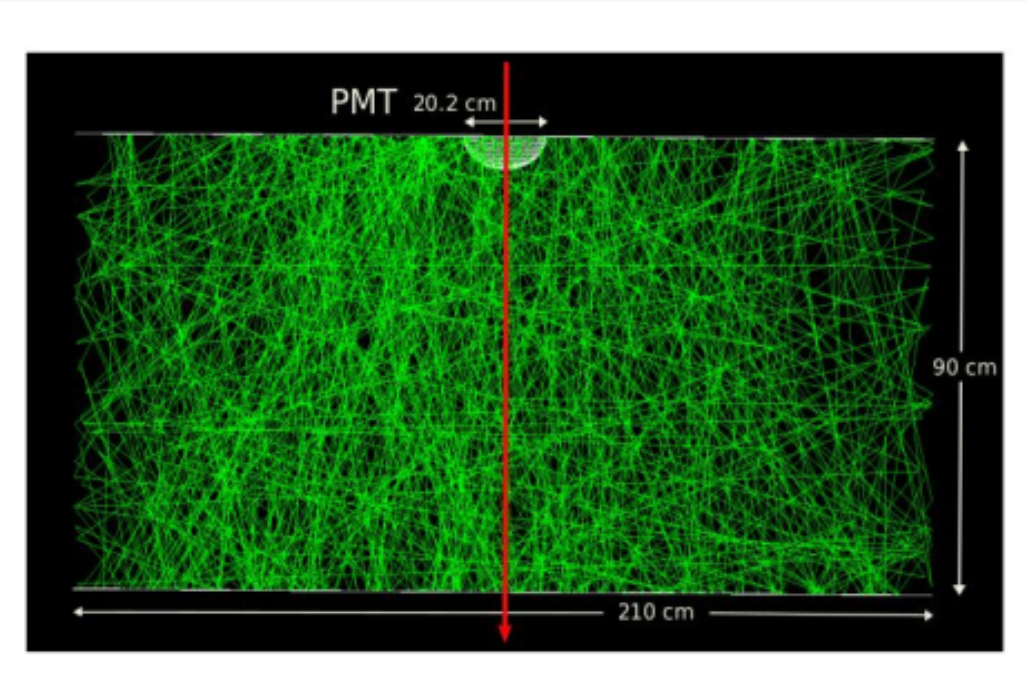
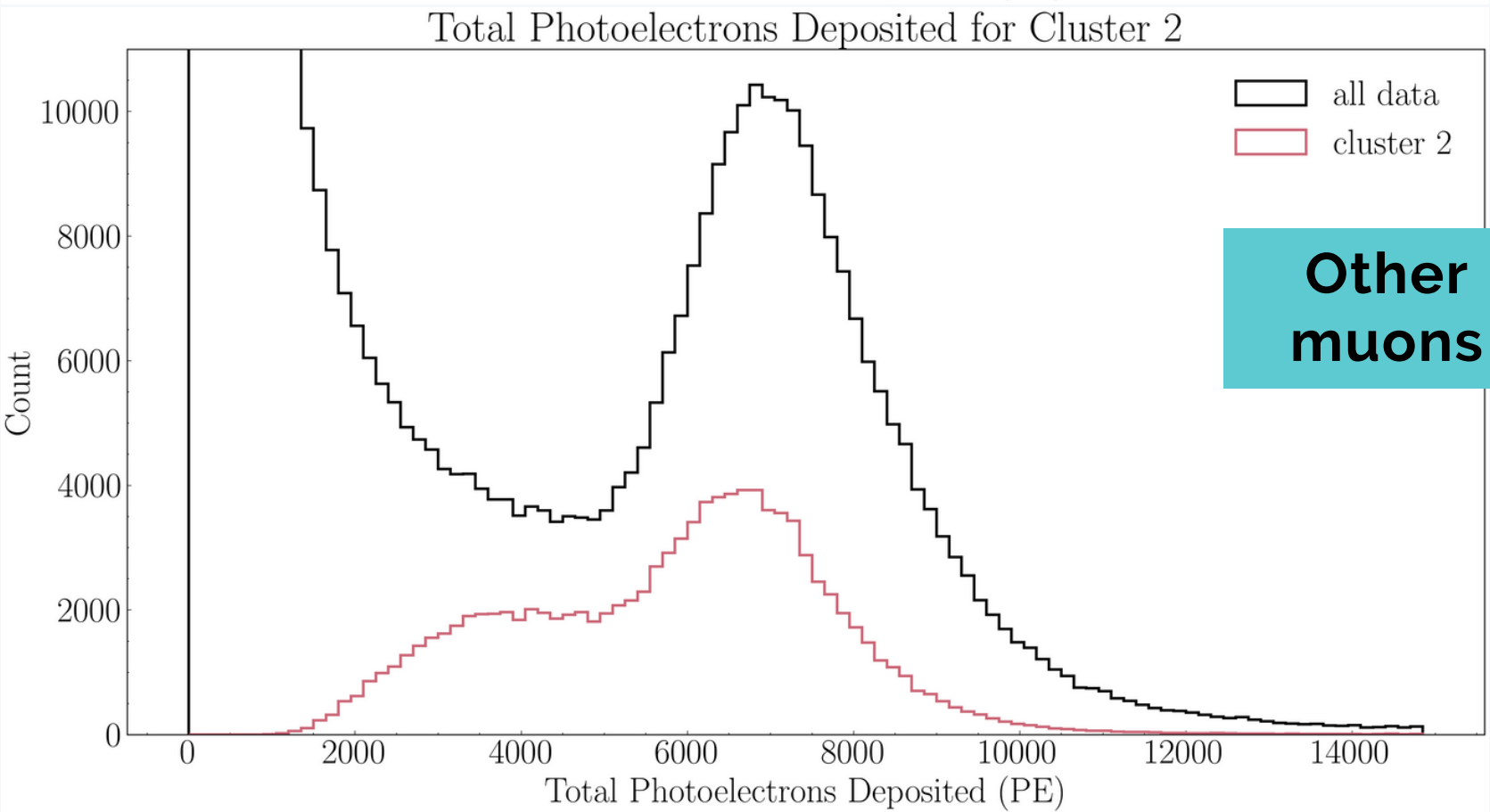
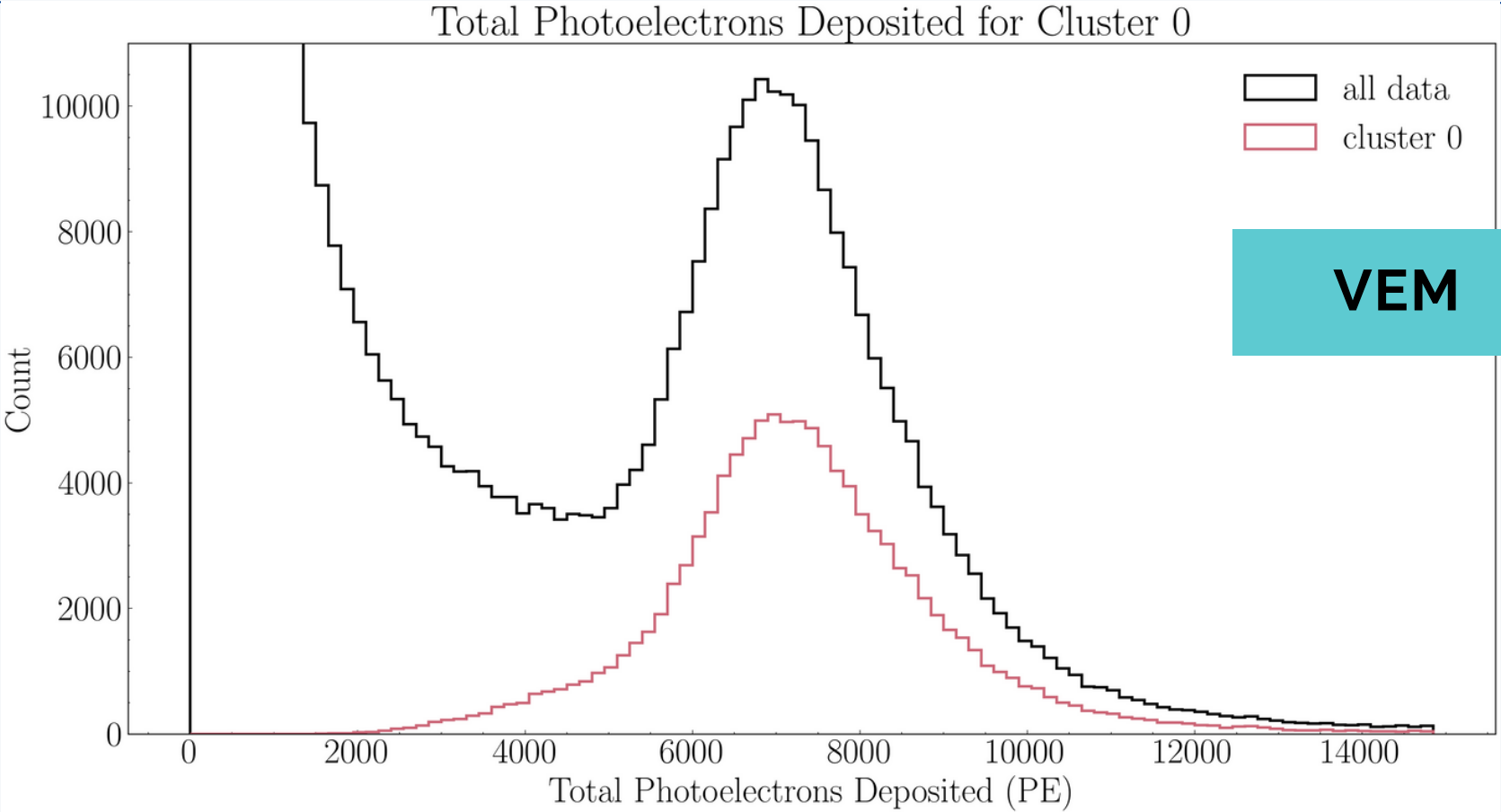
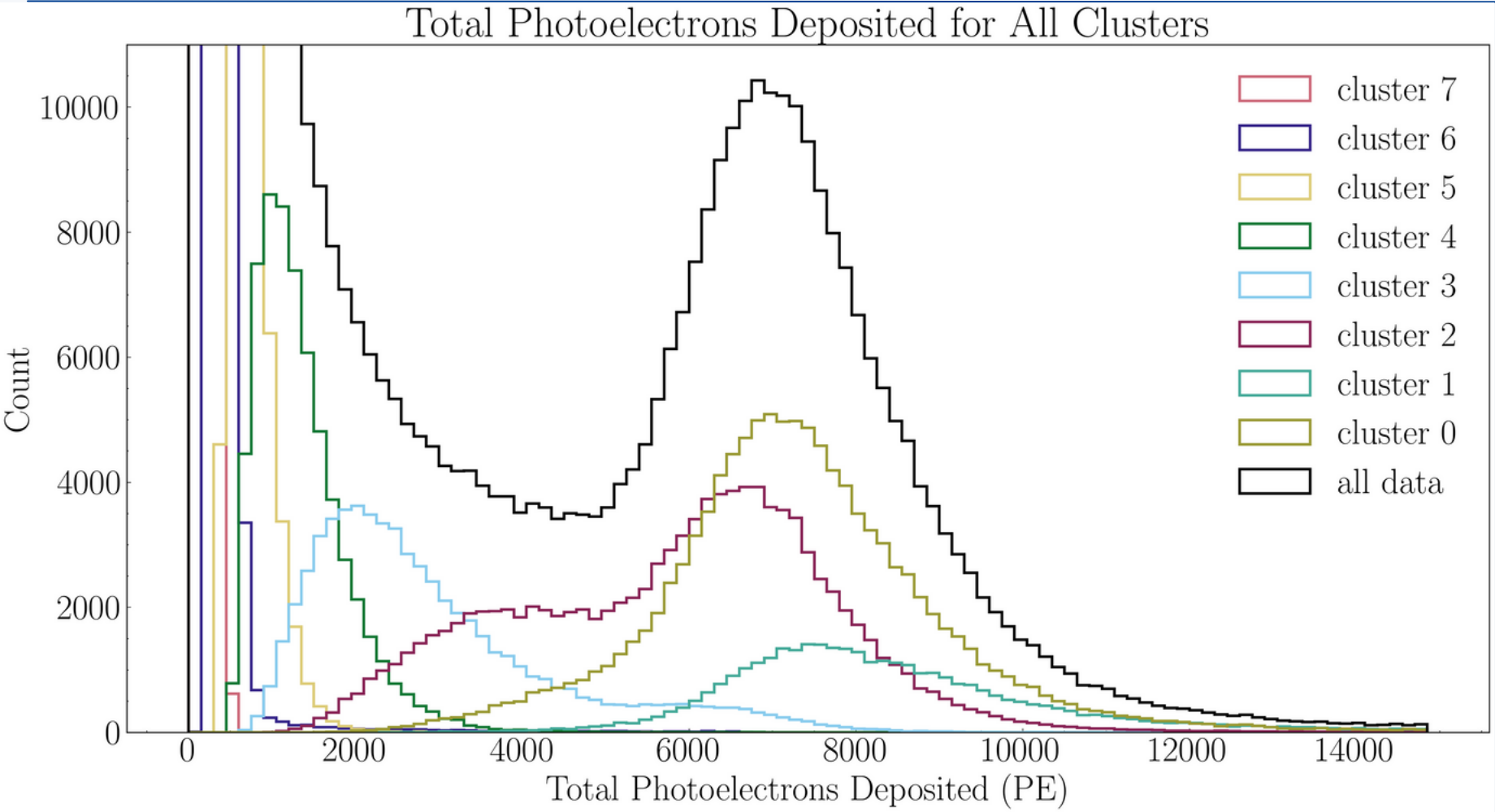


- A threshold through visual inspection can be defined on the reachability plot to provide a cutoff for cluster membership. A value of 0.08 was chosen (after a basic hyperparameter tuning).

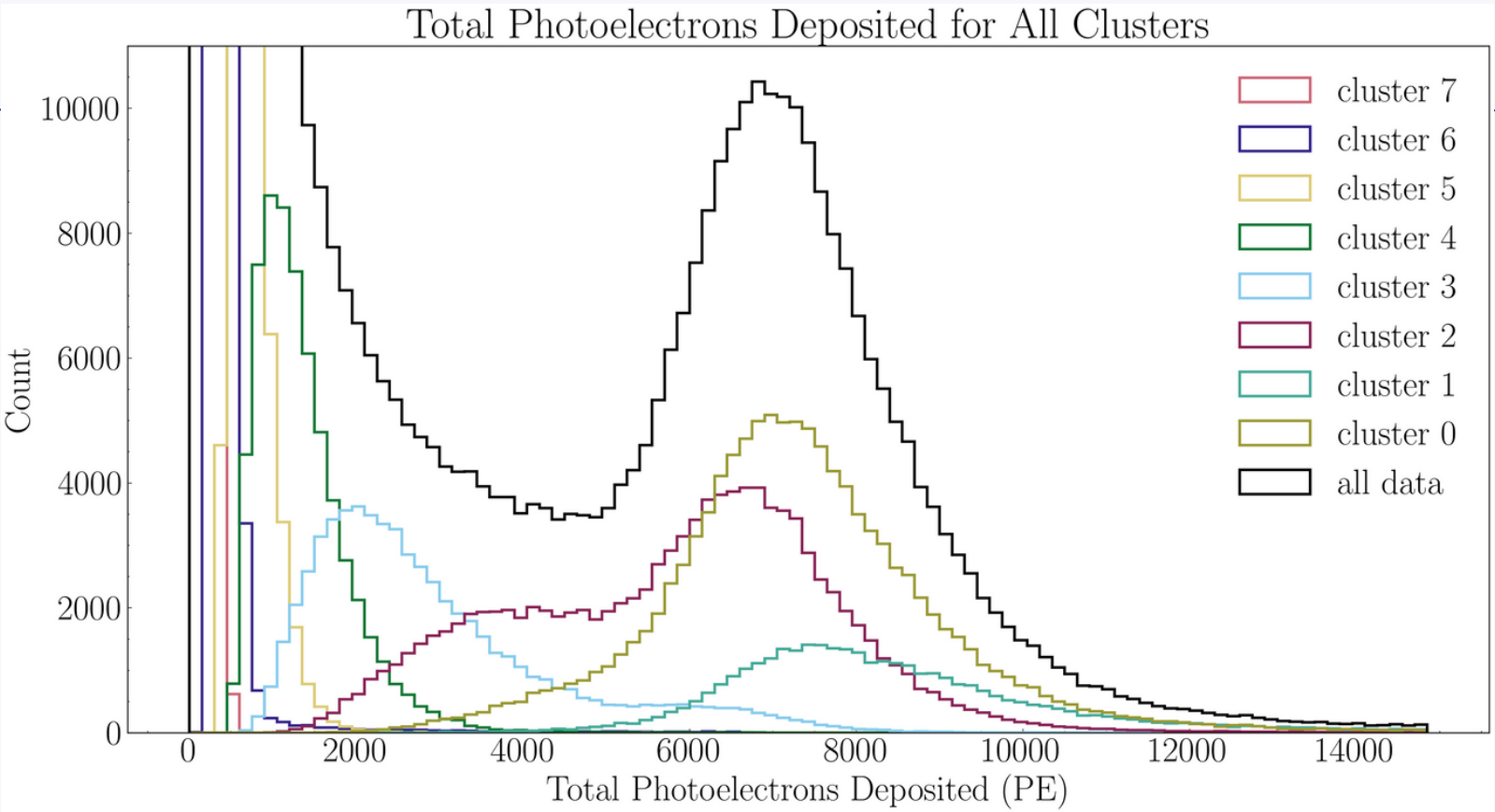








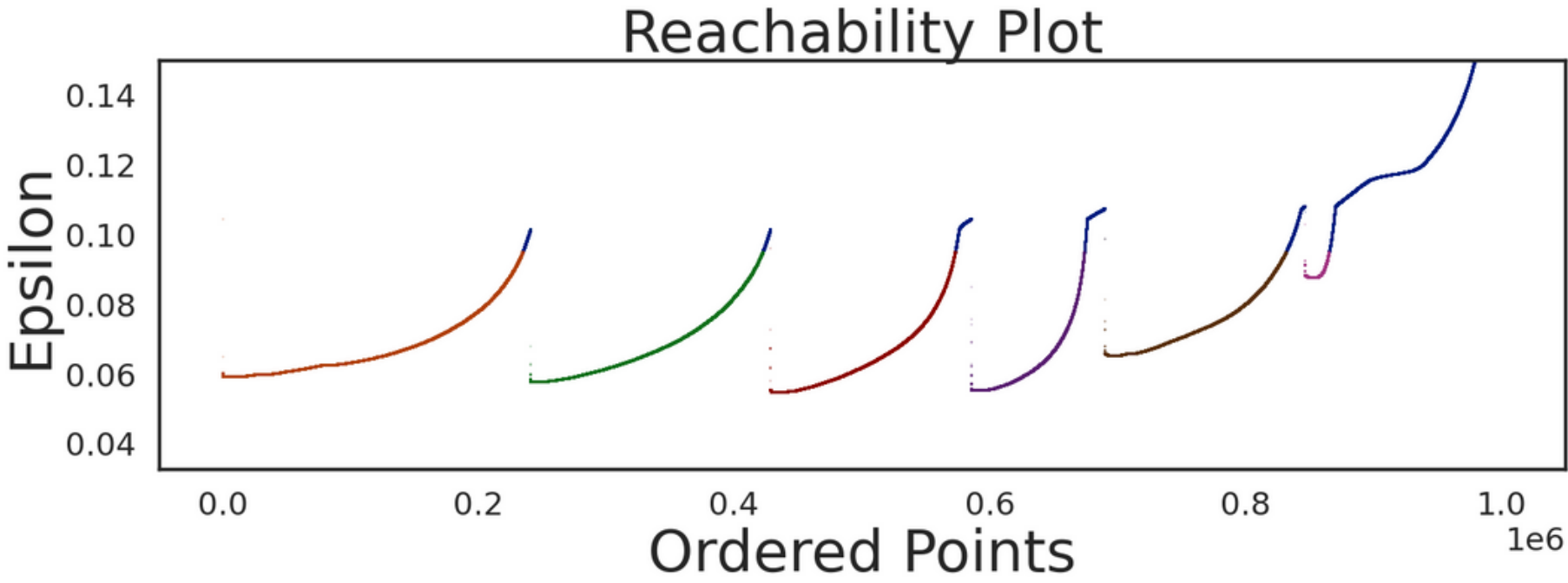
- To test robustness, 24 hours dataset was divided into 24 independent subsets where the algorithms were run independently.
- To test precision, each subset of one hour was run 10 times.
- Average results and deviations where measured.



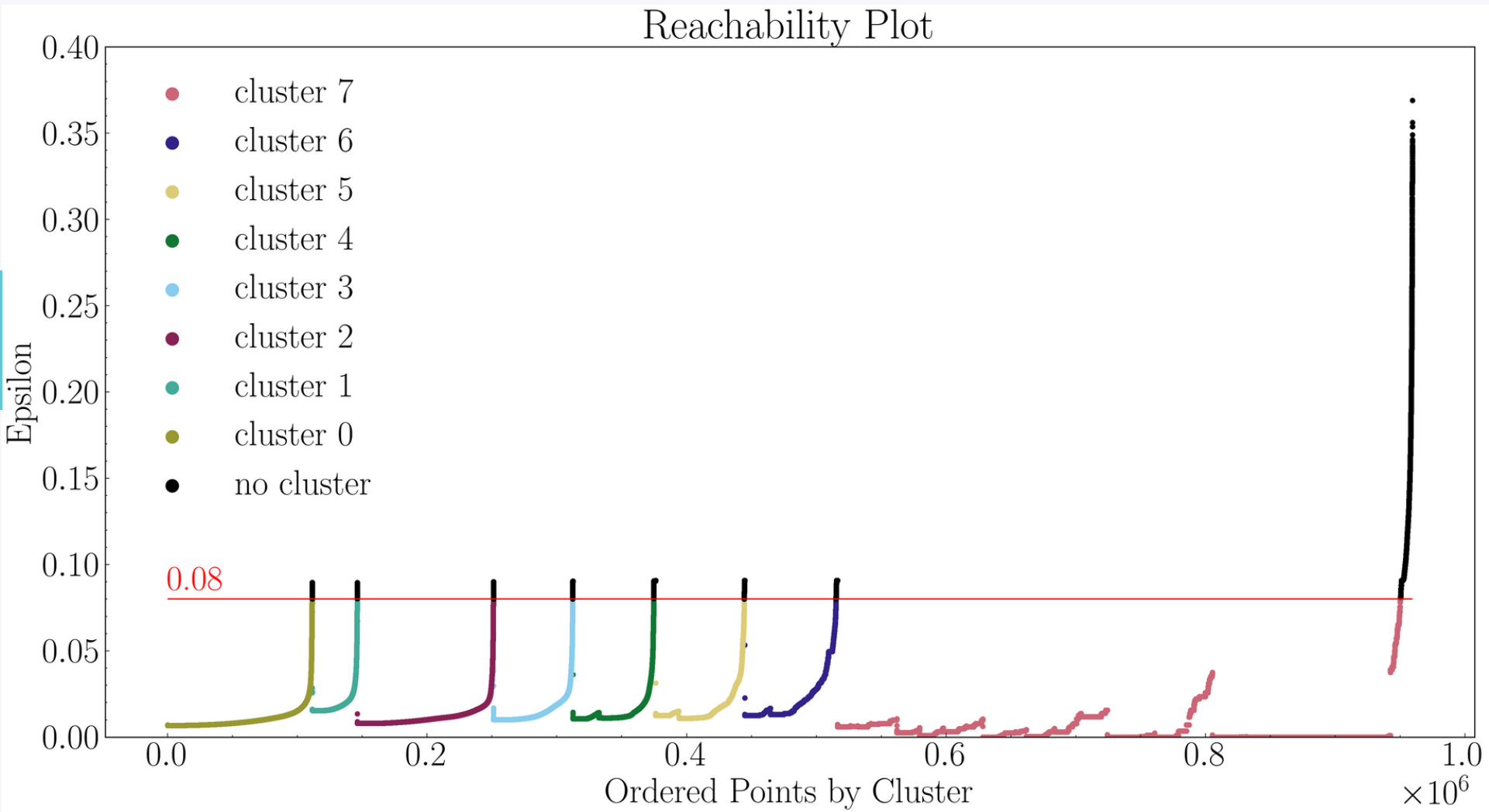
|      | Cluster 0 |        | Cluster 2 |        |
|------|-----------|--------|-----------|--------|
|      | Avg       | Std    | Avg       | Std    |
| Muon | 0.9658    | 0.0024 | 0.9727    | 0.0030 |

- cluster 0 -> ~ 97% muons, ~ 1% electrons/positrons
- cluster 1 -> ~ 89% muons, ~ 5% electrons/positrons
- cluster 2 -> ~ 97% muons, ~ 1% electrons/positrons
- cluster 3 -> ~ 58% muons, ~ 23% photons
- cluster 4 -> ~40% photons, ~ 27% muons, ~22% electrons/positrons
- cluster 5 -> ~62% photons, 22% electrons/positrons
- cluster 6 -> ~70% photons, 20% electrons/muons
- cluster 7 -> ~80% photons, ~1% electrons/mouns

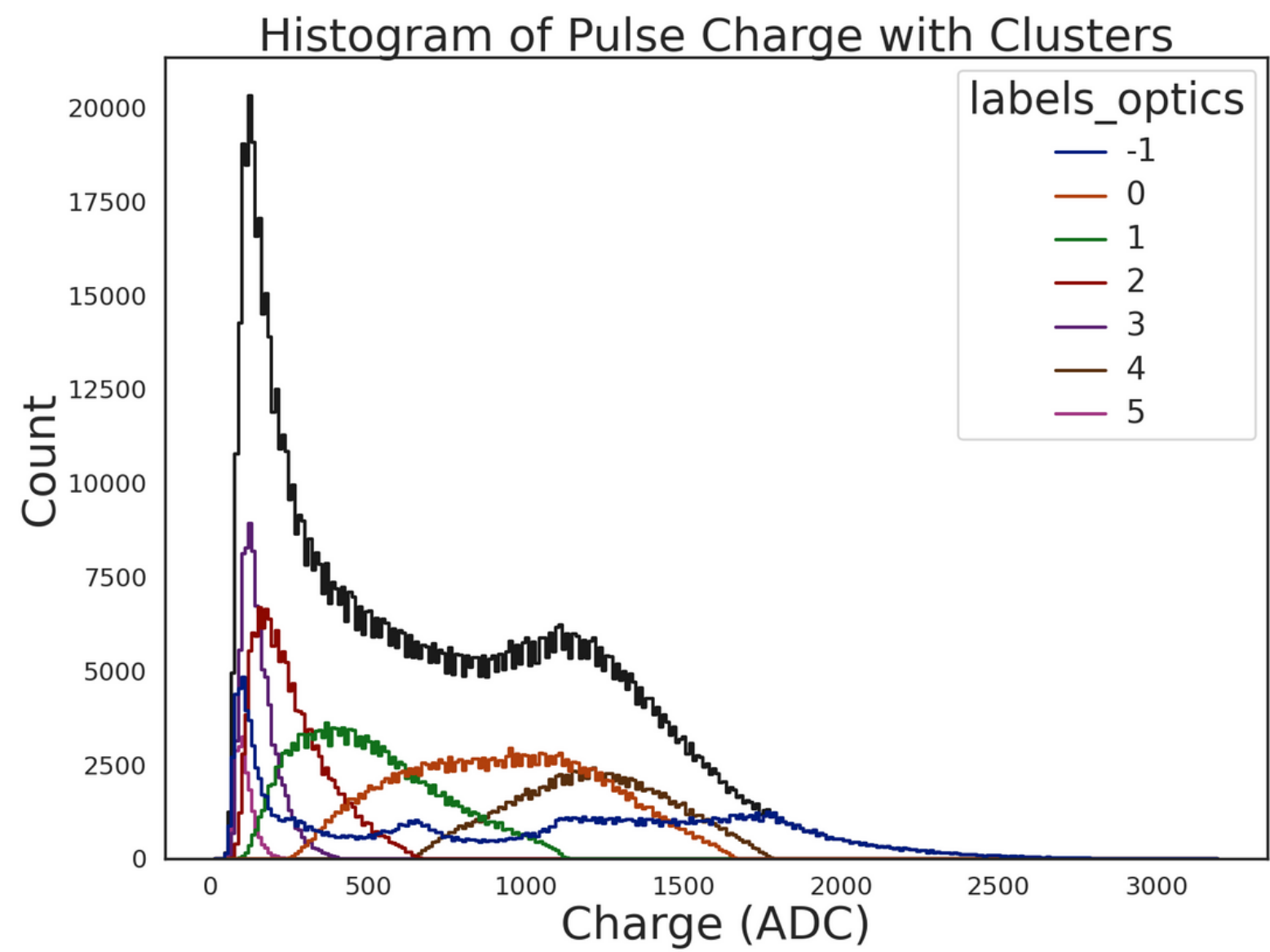
Real  
Data



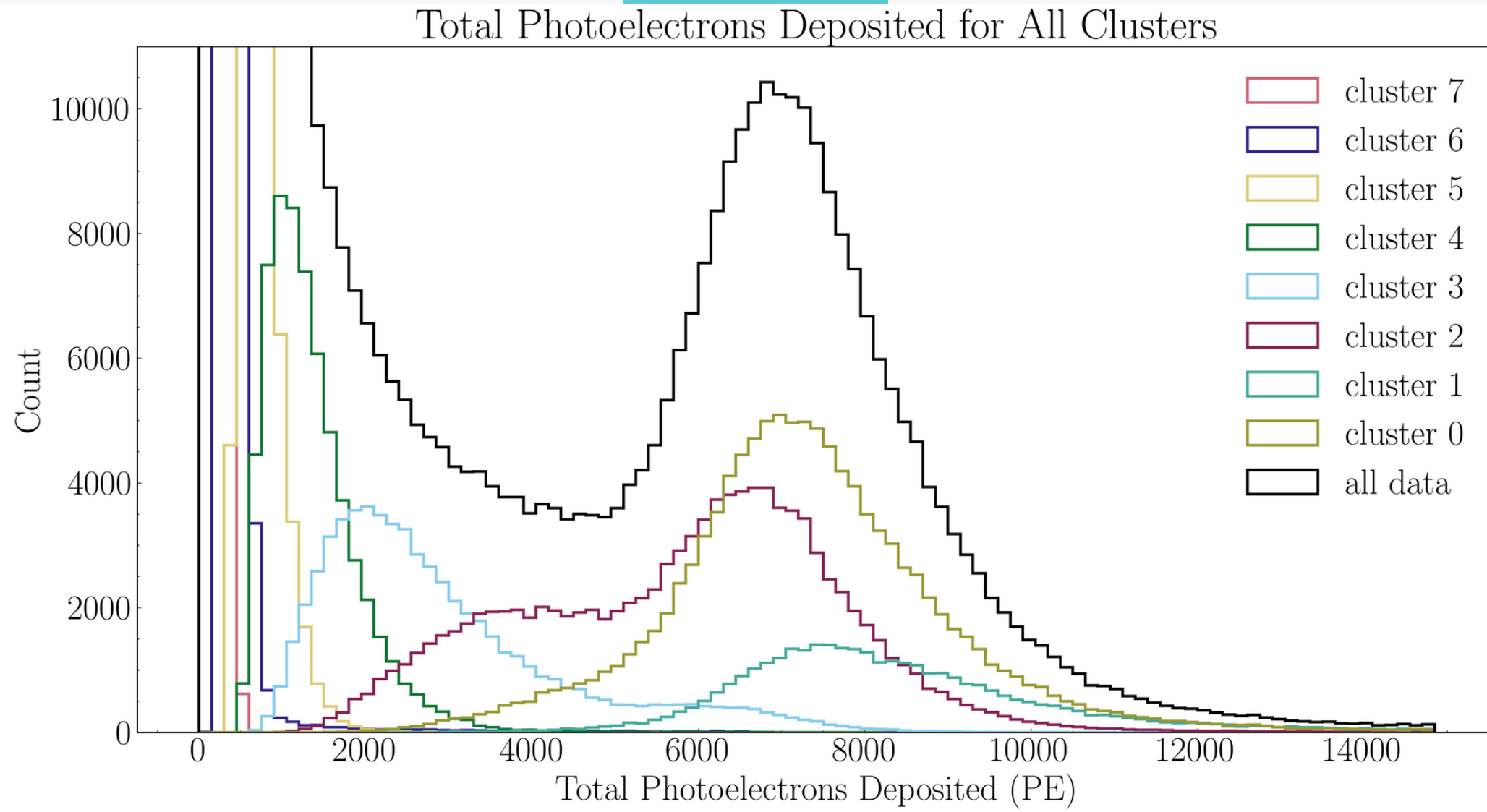
Sim Data



Real Data



Sim Data



## Conclusions

- For the actual data: even with a highly noisy feature space, the reachability plot shows clear cluster structures.
- Cluster groups where located where secondary particle contributions are primarily expected to appear.
- The algorithm is able to separate VEM from other arrival trajectories into the WCD (geometry of the detector is important).

## Future steps:

- Better data quality (high sampling rate) and fine tuning to attend the electromagnetic components and for neutron detection
- Further cleaning and study of the feature space.
- Test the algorithm in other WCDs (e.g. in Antarctica) using pre-trained models
- Additional validation using simulated data and using actual labeled data (for specific cases).

# Thank you!

Contact: Ticiano Torres Peralta  
[ttorres@herrera.unt.edu.ar](mailto:ttorres@herrera.unt.edu.ar)

AVE. AND STD. OF PARTICLE CONTRIBUTIONS FOR 24 ONE HOUR  
RUNS PER CLUSTER

|                   | Cluster 0 |        | Cluster 1 |        | Cluster 2 |        |
|-------------------|-----------|--------|-----------|--------|-----------|--------|
|                   | Avg       | Std    | Avg       | Std    | Avg       | Std    |
| Photon            | 0.0141    | 0.0012 | 0.0513    | 0.0034 | 0.0091    | 0.0015 |
| Electron/Positron | 0.0161    | 0.0011 | 0.0453    | 0.0025 | 0.0135    | 0.0015 |
| Muon              | 0.9658    | 0.0024 | 0.8925    | 0.0059 | 0.9727    | 0.0030 |
| Hadron            | 0.0001    | 0.0000 | 0.0000    | 0.0000 | 0.0001    | 0.0000 |
| Neutron           | 0.0020    | 0.0002 | 0.0049    | 0.0003 | 0.0022    | 0.0002 |
| Other             | 0.0020    | 0.0002 | 0.0060    | 0.0002 | 0.0033    | 0.0023 |

AVE. AND STD. OF PARTICLE CONTRIBUTIONS FOR 24 ONE HOUR  
RUNS PER CLUSTER (CONT.)

|                   | Cluster 3 |        | Cluster 4 |        | Cluster 5 |        | Cluster 6 |        | Cluster 7 |        |
|-------------------|-----------|--------|-----------|--------|-----------|--------|-----------|--------|-----------|--------|
|                   | Avg       | Std    | Avg       | Std    | Avg       | Std    | Avg       | Std    | Avg       | Std    |
| Photon            | 0.2308    | 0.0096 | 0.4645    | 0.0087 | 0.6245    | 0.0051 | 0.7045    | 0.0043 | 0.8060    | 0.0061 |
| Electron/Positron | 0.1520    | 0.0055 | 0.2278    | 0.0044 | 0.2292    | 0.0025 | 0.2001    | 0.0025 | 0.0916    | 0.0007 |
| Muon              | 0.5846    | 0.0141 | 0.2786    | 0.0115 | 0.1276    | 0.0067 | 0.0671    | 0.0037 | 0.0130    | 0.0006 |
| Hadron            | 0.0000    | 0.0000 | 0.0000    | 0.0000 | 0.0000    | 0.0000 | 0.0000    | 0.0000 | 0.0000    | 0.0000 |
| Neutron           | 0.0138    | 0.0007 | 0.0134    | 0.0005 | 0.0090    | 0.0003 | 0.0212    | 0.0015 | 0.0823    | 0.0061 |
| Other             | 0.0187    | 0.0008 | 0.0158    | 0.0008 | 0.0097    | 0.0007 | 0.0071    | 0.0003 | 0.0071    | 0.0003 |

AVE. AND STD. OF PARTICLE CONTRIBUTIONS FOR ONE HOUR  
RUNS REPEATED 10 TIME PER CLUSTER

|                   | Cluster 0 |        | Cluster 1 |        | Cluster 2 |        |
|-------------------|-----------|--------|-----------|--------|-----------|--------|
|                   | Avg       | Std    | Avg       | Std    | Avg       | Std    |
| Photon            | 0.0151    | 0.0007 | 0.0522    | 0.0011 | 0.0105    | 0.0009 |
| Electron/Positron | 0.0173    | 0.0008 | 0.0461    | 0.0010 | 0.0148    | 0.0007 |
| Muon              | 0.9632    | 0.0017 | 0.8908    | 0.0022 | 0.9702    | 0.0015 |
| Hadron            | 0.0001    | 0.0000 | 0.0001    | 0.0000 | 0.0001    | 0.0000 |
| Neutron           | 0.0023    | 0.0002 | 0.0049    | 0.0002 | 0.0022    | 0.0001 |
| Other             | 0.0021    | 0.0002 | 0.0059    | 0.0002 | 0.0030    | 0.0024 |

AVE. AND STD. OF PARTICLE CONTRIBUTIONS FOR ONE HOUR  
RUNS REPEATED 10 TIME PER CLUSTER (CONT.)

|                   | Cluster 3 |        | Cluster 4 |        | Cluster 5 |        | Cluster 6 |        | Cluster 7 |        |
|-------------------|-----------|--------|-----------|--------|-----------|--------|-----------|--------|-----------|--------|
|                   | Avg       | Std    | Avg       | Std    | Avg       | Std    | Avg       | Std    | Avg       | Std    |
| Photon            | 0.2349    | 0.0040 | 0.4682    | 0.0041 | 0.6232    | 0.0031 | 0.7063    | 0.0032 | 0.8078    | 0.0038 |
| Electron/Positron | 0.1536    | 0.0025 | 0.2262    | 0.0019 | 0.2306    | 0.0023 | 0.1988    | 0.0013 | 0.0921    | 0.0004 |
| Muon              | 0.5796    | 0.0056 | 0.2772    | 0.0034 | 0.1278    | 0.0034 | 0.0674    | 0.0014 | 0.0131    | 0.0001 |
| Hadron            | 0.0000    | 0.0000 | 0.0000    | 0.0000 | 0.0000    | 0.0000 | 0.0000    | 0.0000 | 0.0000    | 0.0000 |
| Neutron           | 0.0135    | 0.0005 | 0.0131    | 0.0003 | 0.0090    | 0.0003 | 0.0205    | 0.0011 | 0.0800    | 0.0041 |
| Other             | 0.0183    | 0.0003 | 0.0152    | 0.0004 | 0.0094    | 0.0001 | 0.0069    | 0.0002 | 0.0069    | 0.0001 |