Suppressing pile-up contributions in the formation of topological clusters in ATLAS

Giulia Fazzino

Supervised by: Dr. Chris Malena Delitzsch

Alma Mater Studiorum Università di Bologna, Université Clermont Auvergne,
Technische Universität Dortmund

30 September, 2024

International Master
Advanced Methods
in Particle Physics

Overview

1. Introduction: key concepts
2. Motivation
3. Cluster Classification with DNNs
4. Results: Pile-Up Identification
5. Summary and Outlook

Pile-up in proton-proton collisions

- Protons circulate in the LHC arranged in **bunches**, each containing $\sim 10^{11}$ particles, to reach high instantaneous luminosities
- At each bunch crossing, many interactions happen simultaneously

  $\Longrightarrow$ The interaction with the highest total energy is called **hard scatter** while all other collisions are referred to as **pile-up**
- Pile-up is a background component $\Longrightarrow$ has to be mitigated

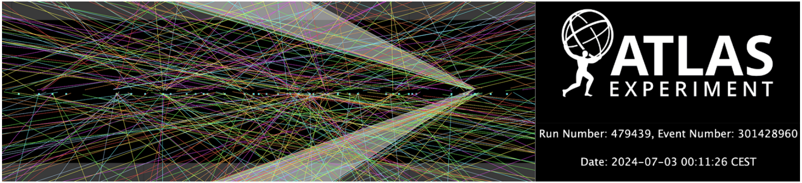
Pile-up in proton-proton collisions

- Protons circulate in the LHC arranged in **bunches**, each containing $\sim 10^{11}$ particles, to reach high instantaneous luminosities

- At each bunch crossing, many interactions happen simultaneously

  $\Longrightarrow$ The interaction with the highest total energy is called **hard scatter** while all other collisions are referred to as **pile-up**

- Pile-up is a background component $\Longrightarrow$ has to be mitigated



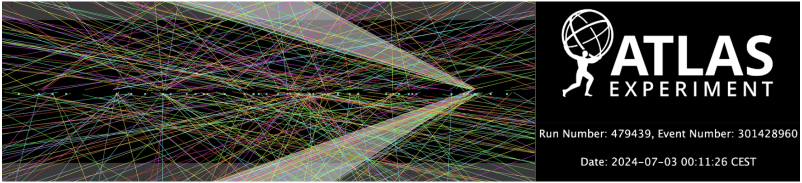
https://twiki.cern.ch/twiki/bin/view/AtlasPublic/EventDisplayRun3Collisions

Pile-up in proton-proton collisions

- Protons circulate in the LHC arranged in **bunches**, each containing $\sim 10^{11}$ particles, to reach high instantaneous luminosities
- At each bunch crossing, many interactions happen simultaneously

  $\Longrightarrow$ The interaction with the highest total energy is called **hard scatter** while all other collisions are referred to as **pile-up**

- Pile-up is a background component $\Longrightarrow$ has to be mitigated



https://twiki.cern.ch/twiki/bin/view/AtlasPublic/EventDisplayRun3Collisions

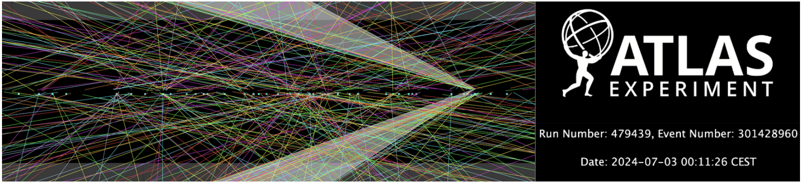
Pile-up in proton-proton collisions

- Protons circulate in the LHC arranged in **bunches**, each containing $\sim 10^{11}$ particles, to reach high instantaneous luminosities

- At each bunch crossing, many interactions happen simultaneously

  $\Longrightarrow$ The interaction with the highest total energy is called **hard scatter** while all other collisions are referred to as **pile-up**

- Pile-up is a background component $\Longrightarrow$ has to be mitigated



https://twiki.cern.ch/twiki/bin/view/AtlasPublic/EventDisplayRun3Collisions

Pile-up in ATLAS

- ATLAS is one of the **general purpose** detectors at the LHC

$\Longrightarrow$ It operates at very high luminosities

$\Longrightarrow$ Measurements include a lot of pile-up

In the simulations used for these studies:

Pile-up in ATLAS

- ATLAS is one of the **general purpose** detectors at the LHC

$\Longrightarrow$ It operates at very high luminosities

$\Longrightarrow$ Measurements include a lot of pile-up

In the simulations used for these studies:

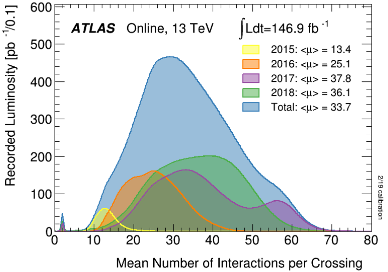
Pile-up in ATLAS

- ATLAS is one of the **general purpose** detectors at the LHC

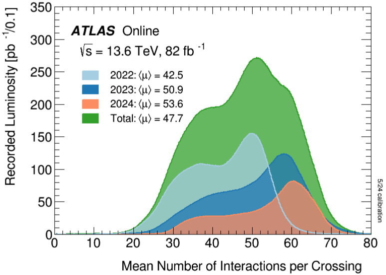
$\Longrightarrow$ It operates at very high luminosities

$\Longrightarrow$ Measurements include a lot of pile-up

In the simulations used for these studies:



https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2



https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun3

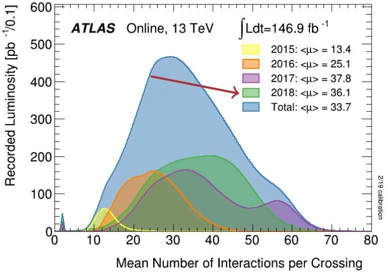
Pile-up in ATLAS

- ATLAS is one of the **general purpose** detectors at the LHC

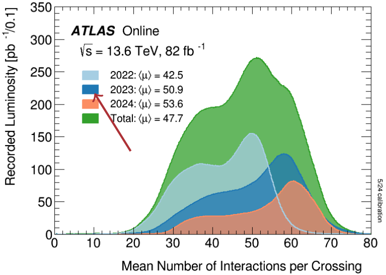
$\Longrightarrow$ It operates at very high luminosities

$\Longrightarrow$ Measurements include a lot of pile-up

In the simulations used for these studies:



https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2



https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun3

Topological clusters

- Group of **geometrically close** calorimeter cells showing a signal with high enough **significance**

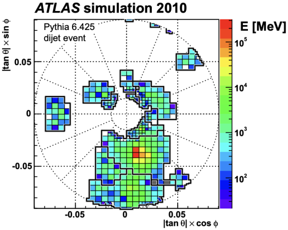
$$\varsigma_{\text{cell}}^{\text{EM}} = E_{\text{cell}}^{\text{EM}} \,/\, \sigma_{\text{noise, cell}}^{\text{EM}}$$

Measured (uncalibrated) cell energy
Average expected noise

- They are used as inputs for **jet reconstruction**

! Each topocluster can contain contributions from both PU and HS

- Current PU-suppression methods are based on tracks (PFlow), momentum (CS + SK) or cell timing (since Run 3)



https://doi.org/10.48550/arXiv.1603.02934

Topological clusters

- Group of **geometrically close** calorimeter cells showing a signal with high enough **significance**

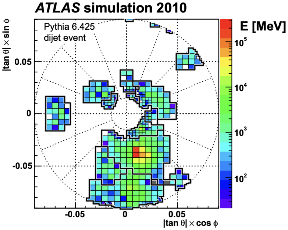
$$\varsigma^{EM}_{\text{cell}} = E^{EM}_{\text{cell}} \,/\, \sigma^{EM}_{\text{noise, cell}}$$

Measured (uncalibrated) cell energy
Average expected noise

- They are used as inputs for **jet reconstruction**

! Each topocluster can contain contributions from both PU and HS

- Current PU-suppression methods are based on tracks (PFlow), momentum (CS + SK) or cell timing (since Run 3)



https://doi.org/10.48550/arXiv.1603.02934

Topological clusters

- Group of **geometrically close** calorimeter cells showing a signal with high enough **significance**

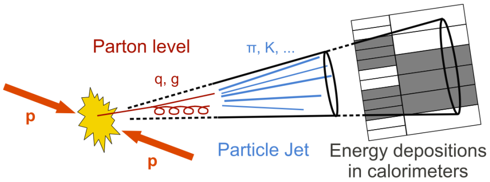
$$\varsigma_{\text{cell}}^{\text{EM}} = E_{\text{cell}}^{\text{EM}} \,/\, \sigma_{\text{noise, cell}}^{\text{EM}}$$

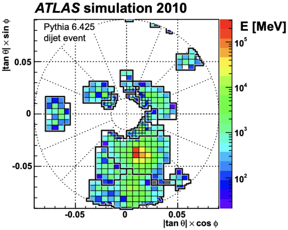
Measured (uncalibrated) cell energy
Average expected noise

- They are used as inputs for **jet reconstruction**

! Each topocluster can contain contributions from both PU and HS

- Current PU-suppression methods are based on tracks (PFlow), momentum (CS + SK) or cell timing (since Run 3)





https://cms.cern/news/jets-cms-and-determination-their-energy-scale

https://doi.org/10.48550/arXiv.1603.02934

Topological clusters

- Group of **geometrically close** calorimeter cells showing a signal with high enough **significance**

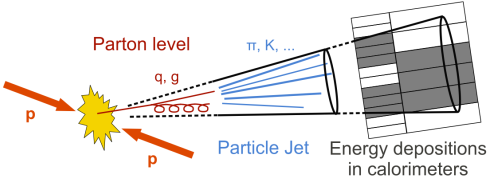
$$\varsigma^{\mathrm{EM}}_{\mathrm{cell}} = E^{\mathrm{EM}}_{\mathrm{cell}} \;/\; \sigma^{\mathrm{EM}}_{\text{noise, cell}}$$

Measured (uncalibrated) cell energy
Average expected noise

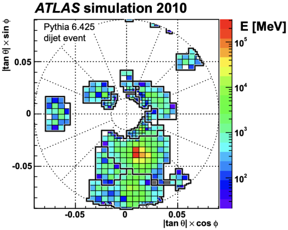
- They are used as inputs for **jet reconstruction**

! Each topocluster can contain contributions from both PU and HS

- Current PU-suppression methods are based on tracks (PFlow), momentum (CS + SK) or cell timing (since Run 3)



https://cms.cern/news/jets-cms-and-determination-their-energy-scale



https://doi.org/10.48550/arXiv.1603.02934

Topological clusters

- Group of **geometrically close** calorimeter cells showing a signal with high enough **significance**

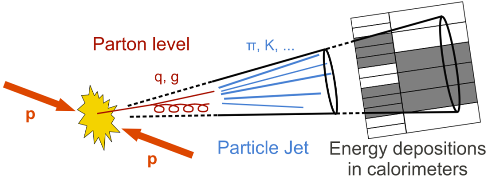
$$\varsigma^{\mathrm{EM}}_{\mathrm{cell}} = E^{\mathrm{EM}}_{\mathrm{cell}} \,/\, \sigma^{\mathrm{EM}}_{\mathrm{noise,\,cell}}$$

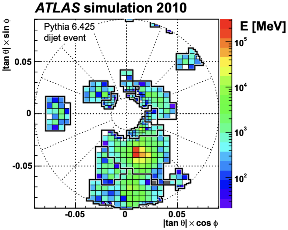
Measured (uncalibrated) cell energy
Average expected noise

- They are used as inputs for **jet reconstruction**

! Each topocluster can contain contributions from both PU and HS

- Current PU-suppression methods are based on tracks (PFlow), momentum (CS + SK) or cell timing (since Run 3)





https://cms.cern/news/jets-cms-and-determination-their-energy-scale
https://doi.org/10.48550/arXiv.1603.02934

Digging deeper: why PU suppression?

- The presence of PU can either add energy to the HS topoclusters or create new clusters
- The **topocluster response** distribution is wider in high PU conditions

$R^{EM}_{clus} = \frac{E^{EM}_{clus}}{E^{dep}}$

E^{EM}_{clus} = cluster energy measured at EM scale (uncalibrated)

E^{dep} = deposited energy in the cluster (truth energy)

$\Longrightarrow$ PU suppression can improve energy calibration

Digging deeper: why PU suppression?

- The presence of PU can either add energy to the HS topoclusters or create new clusters
- The **topocluster response** distribution is wider in high PU conditions

$R^{EM}_{clus} = \frac{E^{EM}_{clus}}{E^{dep}}$

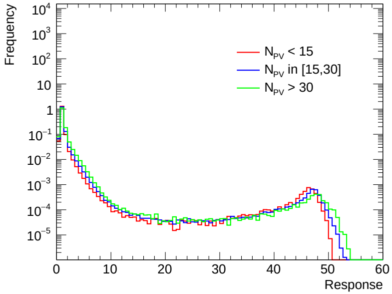
E^{EM}_{clus} = cluster energy measured at EM scale (uncalibrated)

E^{dep} = deposited energy in the cluster (truth energy)

$\Longrightarrow$ PU suppression can improve energy calibration

Digging deeper: why PU suppression?

- The presence of PU can either add energy to the HS topoclusters or create new clusters
- The **topocluster response** distribution is wider in high PU conditions



$$R^{EM}_{clus} = \frac{E^{EM}_{clus}}{E^{dep}}$$

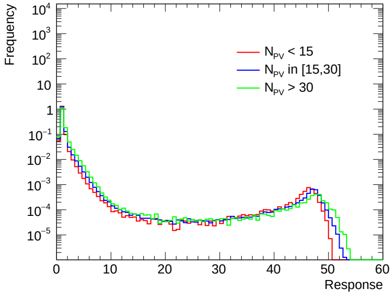
E^{EM}_{clus} = cluster energy measured at EM scale (uncalibrated)

E^{dep} = deposited energy in the cluster (truth energy)

$\Longrightarrow$ PU suppression can improve energy calibration

Digging deeper: why PU suppression?

- The presence of PU can either add energy to the HS topoclusters or create new clusters
- The **topocluster response** distribution is wider in high PU conditions



$$R_{\text{clus}}^{\text{EM}} = \frac{E_{\text{clus}}^{\text{EM}}}{E^{\text{dep}}}$$

$E_{\text{clus}}^{\text{EM}}$ = cluster energy measured at EM scale (uncalibrated)

E^{dep} = deposited energy in the cluster (truth energy)

$\Longrightarrow$ PU suppression can improve energy calibration

Effects of PU on cluster response

- $R^{\mathrm{EM}}_{\mathrm{clus}}$ has a non-trivial distribution
- PU contributions result in a pronounced **right tail**
- Higher distance between the mean and the median $\Longrightarrow$ more skewness

$\Longrightarrow$ PU suppression is especially important for **low-energy** clusters

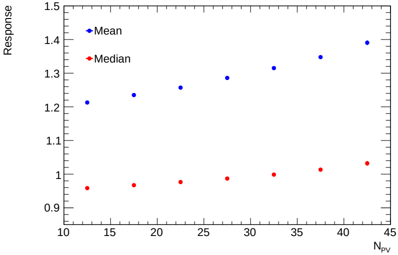
Effects of PU on cluster response

- $R^{\mathrm{EM}}_{\mathrm{clus}}$ has a non-trivial distribution
- PU contributions result in a pronounced **right tail**
- Higher distance between the mean and the median $\implies$ more skewness

$\implies$ PU suppression is especially important for **low-energy** clusters

Effects of PU on cluster response

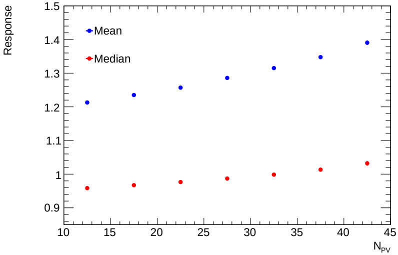
- $R^{\mathrm{EM}}_{\mathrm{clus}}$ has a non-trivial distribution
- PU contributions result in a pronounced **right tail**
- Higher distance between the mean and the median $\implies$ more skewness

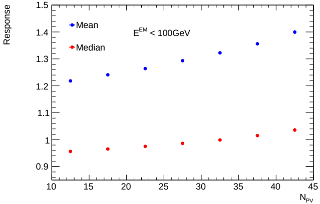


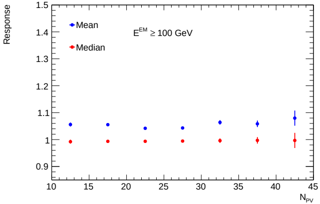
$\implies$ PU suppression is especially important for **low-energy** clusters

Effects of PU on cluster response

- $R^{\mathrm{EM}}_{\mathrm{clus}}$ has a non-trivial distribution
- PU contributions result in a pronounced **right tail**
- Higher distance between the mean and the median $\implies$ more skewness







$\implies$ PU suppression is especially important for **low-energy** clusters

Key idea

- **Goal**: Suppress pile-up contributions to calorimeter topoclusters
- **Strategy**:
  - Distinguish, using **Deep Neural Networks**:
    1. Clusters with only pile-up contributions (**PU-only**)
    2. Clusters with contributions from both hard scatter and pile-up (**Mixed**)
    3. Clusters with only hard scatter contributions (**HS-only**)
  - Remove (1) from jet reconstruction inputs
  - Correct (2) to remove pile-up contributions
- Two DNNs have been implemented for this purpose:
  - One to identify HS-only clusters
  - One to identify PU-only clusters

Key idea

- **Goal**: Suppress pile-up contributions to calorimeter topoclusters
- **Strategy**:
  - Distinguish, using **Deep Neural Networks**:
    1. Clusters with only pile-up contributions (**PU-only**)
    2. Clusters with contributions from both hard scatter and pile-up (**Mixed**)
    3. Clusters with only hard scatter contributions (**HS-only**)
  - Remove (1) from jet reconstruction inputs
  - Correct (2) to remove pile-up contributions
- Two DNNs have been implemented for this purpose:
  - One to identify HS-only clusters
  - One to identify PU-only clusters

Key idea

- **Goal**: Suppress pile-up contributions to calorimeter topoclusters
- **Strategy**:
  - Distinguish, using **Deep Neural Networks**:
    1. Clusters with only pile-up contributions (**PU-only**)
    2. Clusters with contributions from both hard scatter and pile-up (**Mixed**)
    3. Clusters with only hard scatter contributions (**HS-only**)
  - Remove (1) from jet reconstruction inputs
  - Correct (2) to remove pile-up contributions
- Two DNNs have been implemented for this purpose:
  - One to identify HS-only clusters
  - One to identify PU-only clusters

Key idea

- **Goal**: Suppress pile-up contributions to calorimeter topoclusters
- **Strategy**:
  - Distinguish, using **Deep Neural Networks**:
    1. Clusters with only pile-up contributions (**PU-only**)
    2. Clusters with contributions from both hard scatter and pile-up (**Mixed**)
    3. Clusters with only hard scatter contributions (**HS-only**)
  - Remove (1) from jet reconstruction inputs
  - Correct (2) to remove pile-up contributions
- Two DNNs have been implemented for this purpose:
  - One to identify HS-only clusters
  - One to identify PU-only clusters

Network(s) Overview

- The networks have the **same structure**, optimized for the HS-only identification task, with the Run 2 MC Simulation
- **Architecture**:
  - 512 nodes (`ReLu`) → batch normalization
  - 256 nodes (`ReLu`) → batch normalization
  - 128 nodes (`ReLu`) → batch normalization
  - 64 nodes (`ReLu`) → batch normalization → 30% dropout
  - 32 nodes (`ReLu`) → batch normalization → 30% dropout
  - 16 nodes (`ReLu`) → batch normalization
  - 1 node (`sigmoid`)
- **Performance score**: ROC-AUC
- **Loss**: binary crossentropy

Network(s) Overview

- The networks have the **same structure**, optimized for the HS-only identification task, with the Run 2 MC Simulation
- **Architecture**:
  - 512 nodes (`ReLu`) $\rightarrow$ batch normalization
  - 256 nodes (`ReLu`) $\rightarrow$ batch normalization
  - 128 nodes (`ReLu`) $\rightarrow$ batch normalization
  - 64 nodes (`ReLu`) $\rightarrow$ batch normalization $\rightarrow$ 30% dropout
  - 32 nodes (`ReLu`) $\rightarrow$ batch normalization $\rightarrow$ 30% dropout
  - 16 nodes (`ReLu`) $\rightarrow$ batch normalization
  - 1 node (`sigmoid`)
- **Performance score**: ROC-AUC
- **Loss**: binary crossentropy

Network(s) Overview

- The networks have the **same structure**, optimized for the HS-only identification task, with the Run 2 MC Simulation
- **Architecture**:
  - 512 nodes (`ReLu`) $\rightarrow$ batch normalization
  - 256 nodes (`ReLu`) $\rightarrow$ batch normalization
  - 128 nodes (`ReLu`) $\rightarrow$ batch normalization
  - 64 nodes (`ReLu`) $\rightarrow$ batch normalization $\rightarrow$ 30% dropout
  - 32 nodes (`ReLu`) $\rightarrow$ batch normalization $\rightarrow$ 30% dropout
  - 16 nodes (`ReLu`) $\rightarrow$ batch normalization
  - 1 node (`sigmoid`)
- **Performance score**: ROC-AUC
- **Loss**: binary crossentropy

Input Features Choice

- Not all the topocluster features are important for these tasks, thus the feature number can be reduced

1. Keep only features with correlation below 75%
2. Evaluate feature **permutation importance** for the remaining set of (24) features
3. Remove features with low importance
4. **Final set**: 14 (13) input features

$\Longrightarrow$ The most important features will also be less pile-up robust

Permutation importance

Shuffle randomly the values of one feature at a time and look at how much the performance (*here the ROC-AUC*) changes

Input Features Choice

- Not all the topocluster features are important for these tasks, thus the feature number can be reduced

1. Keep only features with correlation below 75%
2. Evaluate feature **permutation importance** for the remaining set of (24) features
3. Remove features with low importance
4. **Final set:** 14 (13) input features

$\Longrightarrow$ The most important features will also be less pile-up robust

Permutation importance

Shuffle randomly the values of one feature at a time and look at how much the performance (*here the ROC-AUC*) changes

Input Features Choice

- Not all the topocluster features are important for these tasks, thus the feature number can be reduced

1. Keep only features with correlation below 75%
2. Evaluate feature **permutation importance** for the remaining set of (24) features
3. Remove features with low importance
4. **Final set:** 14 (13) input features

$\Longrightarrow$ The most important features will also be less pile-up robust

Permutation importance

Shuffle randomly the values of one feature at a time and look at how much the performance (*here the ROC-AUC*) changes

Input Features Choice

- Not all the topocluster features are important for these tasks, thus the feature number can be reduced

1. Keep only features with correlation below 75%
2. Evaluate feature **permutation importance** for the remaining set of (24) features
3. Remove features with low importance
4. **Final set:** 14 (13) input features

$\Longrightarrow$ The most important features will also be less pile-up robust

Permutation importance

Shuffle randomly the values of one feature at a time and look at how much the performance (*here the ROC-AUC*) changes

Input Features Choice

- Not all the topocluster features are important for these tasks, thus the feature number can be reduced

1. Keep only features with correlation below 75%
2. Evaluate feature **permutation importance** for the remaining set of (24) features
3. Remove features with low importance
4. **Final set**: 14 (13) input features

$\Longrightarrow$ The most important features will also be less pile-up robust

Permutation importance

Shuffle randomly the values of one feature at a time and look at how much the performance (*here the ROC-AUC*) changes

Input Features

1. Cluster probability to be **generated** by an **EM** shower
2. **Distance** of the cluster from nominal vertex
3. Weighted first moment of cell **signal density** distribution
4. **Distance** of the cluster from calorimeter frontface
5. Fraction of **energy** in **EM** calorimeter
6. Cluster **timing**
7. Total **number of cells** in the cluster
8. Cluster **isolation**
9. Signal **quality** in Tile calorimeter
10. **Variance** of cell **timing** distribution
11. Signal **quality** in LAr calorimeter
12. Second moment of radial **distances** of cells to the principal cluster axis
13. **Energy dispersion** perpendicular to main cluster axis
14. Cluster **energy** (solely for HS-only identification)

Input Features

1. Cluster probability to be **generated** by an **EM** shower
2. **Distance** of the cluster from nominal vertex
3. Weighted first moment of cell **signal density** distribution
4. **Distance** of the cluster from calorimeter frontface
5. Fraction of **energy** in **EM** calorimeter
6. Cluster **timing**
7. Total **number of cells** in the cluster
8. Cluster **isolation**
9. Signal **quality** in Tile calorimeter
10. **Variance** of cell **timing** distribution
11. Signal **quality** in LAr calorimeter
12. Second moment of radial **distances** of cells to the principal cluster axis
13. **Energy dispersion** perpendicular to main cluster axis
14. Cluster **energy** (solely for HS-only identification)

Data

- 2018 (**Run 2**) & 2023 (**Run 3**) simulated di-jet samples
- Jet $p_T < 1800\,\text{GeV}$
- 13 input cluster features
- 2 classes of clusters:
  1. **PU-only**: $E^{\text{dep}} < 1\,\text{MeV}$ and $R^{\text{EM}}_{\text{clus}} > 4$
  2. **Mixed**: $E^{\text{dep}} > 1\,\text{MeV}$ or $R^{\text{EM}}_{\text{clus}} < 4$
- Is the definition fine?
  - ✓ In a HS-only sample, taken from a simulation without pile-up, only a fraction of $2 \cdot 10^{-5}$ clusters satisfies (1)

Data

- 2018 (**Run 2**) & 2023 (**Run 3**) simulated di-jet samples
- Jet $p_T < 1800\,\text{GeV}$
- 13 input cluster features
- 2 classes of clusters:
  1. **PU-only**: $E^{\text{dep}} < 1\,\text{MeV}$ and $R^{\text{EM}}_{\text{clus}} > 4$
  2. **Mixed**: $E^{\text{dep}} > 1\,\text{MeV}$ or $R^{\text{EM}}_{\text{clus}} < 4$
- Is the definition fine?
  - ✓ In a HS-only sample, taken from a simulation without pile-up, only a fraction of $2 \cdot 10^{-5}$ clusters satisfies (1)

Data

- 2018 (**Run 2**) & 2023 (**Run 3**) simulated di-jet samples
- Jet $p_T < 1800\,\text{GeV}$
- 13 input cluster features
- 2 classes of clusters:
  1. **PU-only**: $E^{\text{dep}} < 1\,\text{MeV}$ and $R^{\text{EM}}_{\text{clus}} > 4$
  2. **Mixed**: $E^{\text{dep}} > 1\,\text{MeV}$ or $R^{\text{EM}}_{\text{clus}} < 4$
- Is the definition fine?
  - ✓ In a HS-only sample, taken from a simulation without pile-up, only a fraction of $2 \cdot 10^{-5}$ clusters satisfies (1)

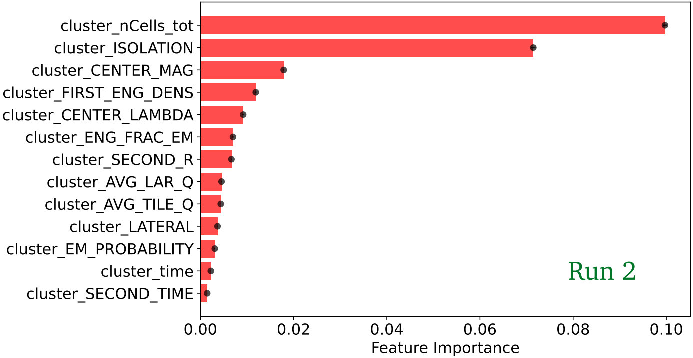
Data

- 2018 (**Run 2**) & 2023 (**Run 3**) simulated di-jet samples
- Jet $p_T < 1800\,\text{GeV}$
- 13 input cluster features
- 2 classes of clusters:
  1. **PU-only**: $E^{\text{dep}} < 1\,\text{MeV}$ and $R^{\text{EM}}_{\text{clus}} > 4$
  2. **Mixed**: $E^{\text{dep}} > 1\,\text{MeV}$ or $R^{\text{EM}}_{\text{clus}} < 4$
- Is the definition fine?
  - ✓ In a HS-only sample, taken from a simulation without pile-up, only a fraction of $2 \cdot 10^{-5}$ clusters satisfies (1)

Data

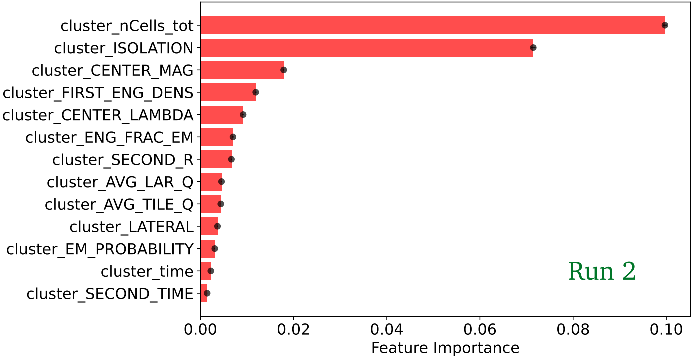
- 2018 (**Run 2**) & 2023 (**Run 3**) simulated di-jet samples
- Jet $p_T < 1800\,\text{GeV}$
- 13 input cluster features
- 2 classes of clusters:
  1. **PU-only**: $E^{\text{dep}} < 1\,\text{MeV}$ and $R^{\text{EM}}_{\text{clus}} > 4$
  2. **Mixed**: $E^{\text{dep}} > 1\,\text{MeV}$ or $R^{\text{EM}}_{\text{clus}} < 4$
- Is the definition fine?
  - ✓ In a HS-only sample, taken from a simulation without pile-up, only a fraction of $2 \cdot 10^{-5}$ clusters satisfies (1)

Feature Importance



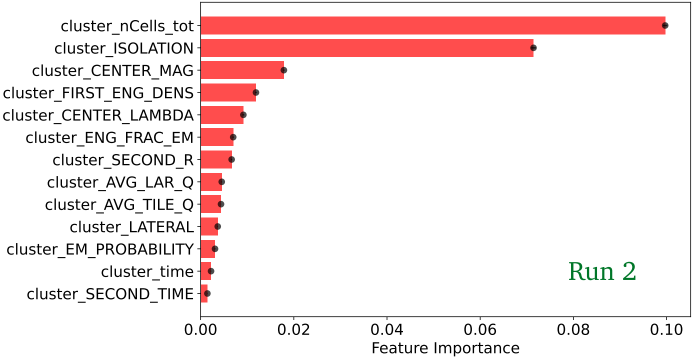
- PU-only clusters are smaller (lower number of cells) and more isolated
- Very similar results for Run 3

Feature Importance



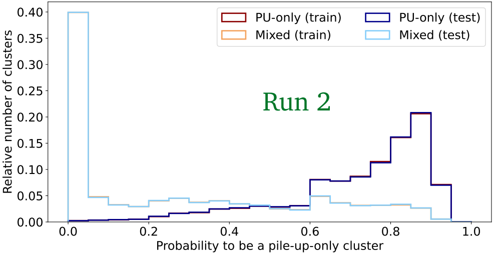
- PU-only clusters are smaller (lower number of cells) and more isolated
- Very similar results for Run 3

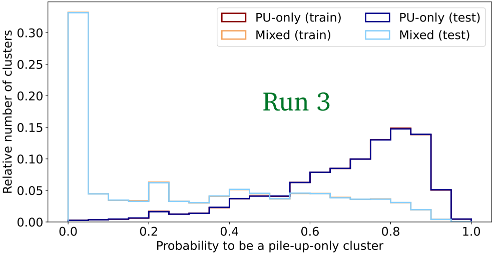
Feature Importance



- PU-only clusters are smaller (lower number of cells) and more isolated
- Very similar results for Run 3

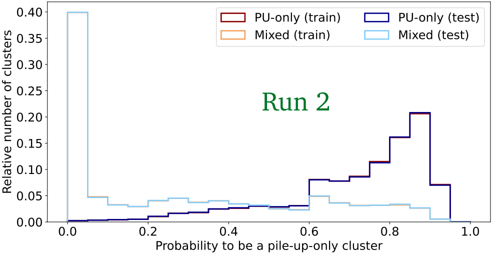
Classification Output

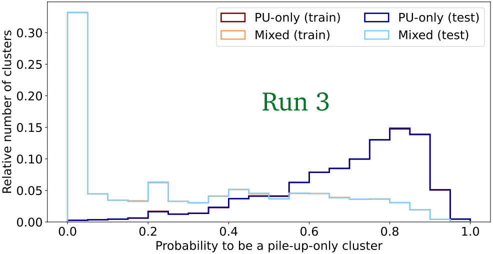




- The highest possible output is never reached
- Good separation in both cases
- Could be improved with a more strict mixed clusters definition

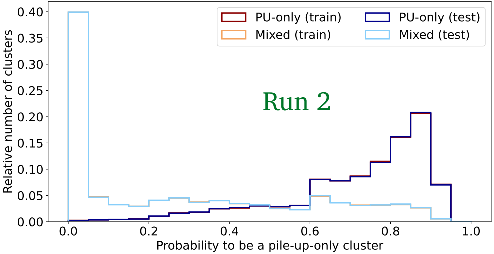
Classification Output

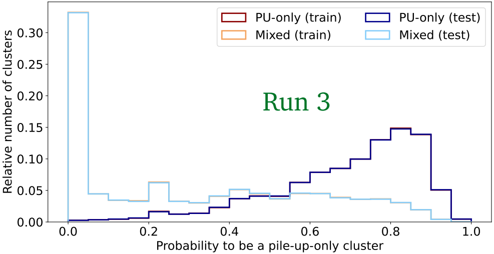




- The highest possible output is never reached
- Good separation in both cases
- Could be improved with a more strict mixed clusters definition

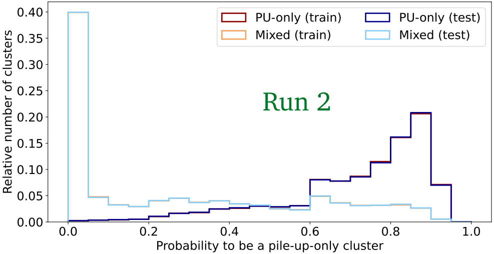
Classification Output

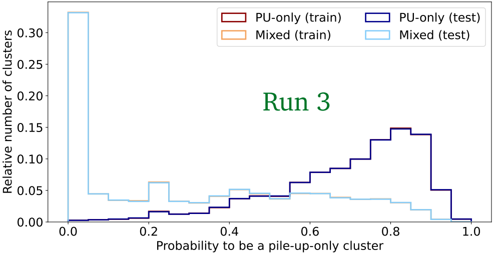




- The highest possible output is never reached
- Good separation in both cases
- Could be improved with a more strict mixed clusters definition

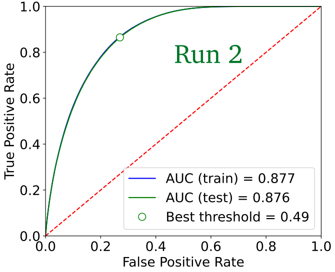
Classification Output

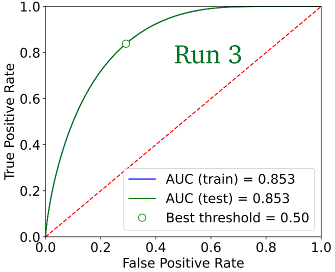




- The highest possible output is never reached
- Good separation in both cases
- Could be improved with a more strict mixed clusters definition

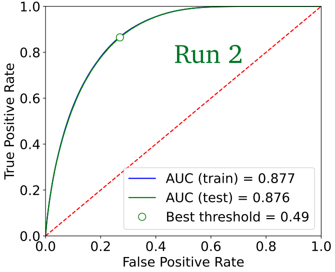
Classification Performance

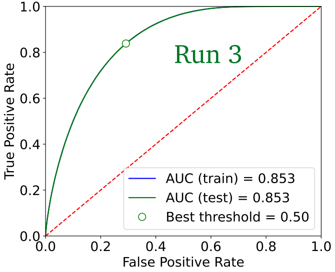




- **High AUC** in both cases
- Pile-up **suppression strategy**:
  1. Choose a threshold on the network output (*here: upper corner of the ROC curve*)
  2. Keep only clusters with output below the threshold

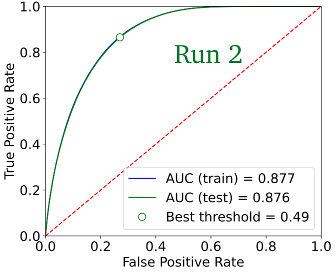
Classification Performance

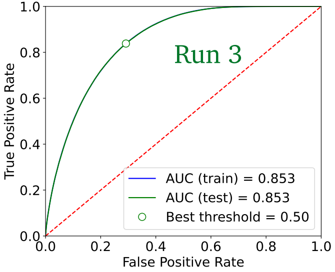




- **High AUC** in both cases
- Pile-up **suppression strategy**:
  1. Choose a threshold on the network output (*here: upper corner of the ROC curve*)
  2. Keep only clusters with output below the threshold

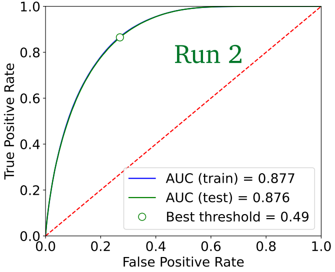
Classification Performance

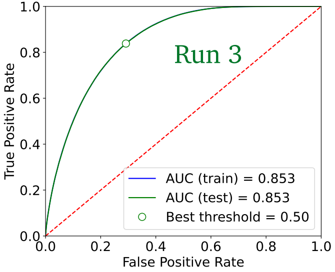




- **High AUC** in both cases
- Pile-up **suppression strategy**:
  1. Choose a threshold on the network output (*here: upper corner of the ROC curve*)
  2. Keep only clusters with output below the threshold

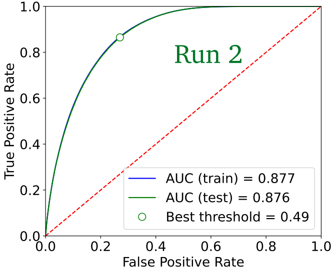
Classification Performance

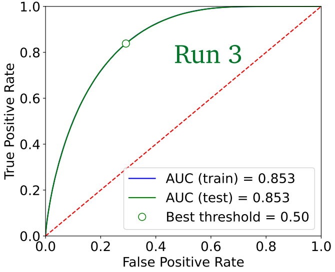




- **High AUC** in both cases
- Pile-up **suppression strategy**:
  1. Choose a threshold on the network output (*here: upper corner of the ROC curve*)
  2. Keep only clusters with output below the threshold

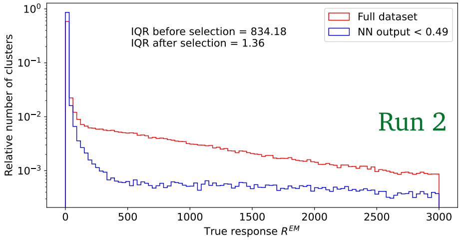
Classification Performance

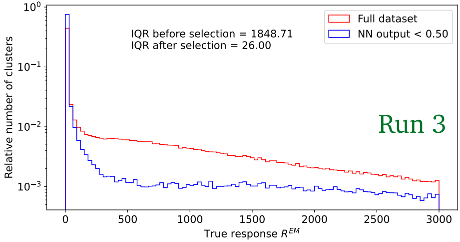




- **High AUC** in both cases
- Pile-up **suppression strategy**:
  1. Choose a threshold on the network output (*here: upper corner of the ROC curve*)
  2. Keep only clusters with output below the threshold

Topocluster Response After Pile-Up Suppression

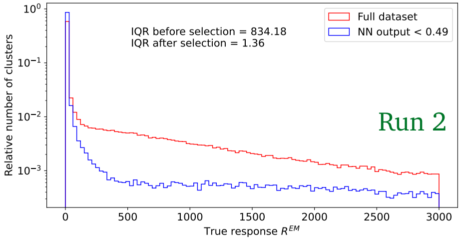


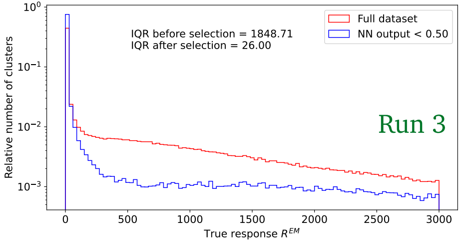


- High IQR before suppression because PU-only clusters can have very high responses (small denominator in $R^{\mathrm{EM}}_{\mathrm{clus}}$)
- The **IQR decreases remarkably** after PU-suppression

IQR	Before	After
Run 2	834	1
Run 3	1849	26

Topocluster Response After Pile-Up Suppression



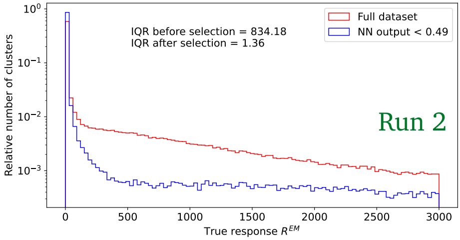


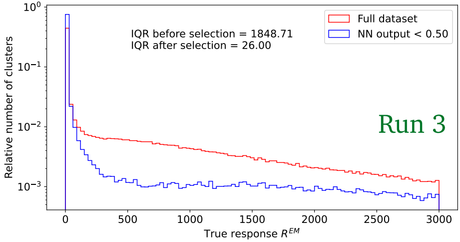
- High IQR before suppression because PU-only clusters can have very high responses (small denominator in $R^{\mathrm{EM}}_{\mathrm{clus}}$)

- The **IQR decreases remarkably** after PU-suppression

IQR	Before	After
Run 2	834	1
Run 3	1849	26

Topocluster Response After Pile-Up Suppression





- High IQR before suppression because PU-only clusters can have very high responses (small denominator in $R^{\text{EM}}_{\text{clus}}$)
- The **IQR decreases remarkably** after PU-suppression

IQR	Before	After
Run 2	834	1
Run 3	1849	26

Summary

- The presence of pile-up degrades collider measurements, especially at high luminosity
- The **topocluster response** worsens with high pile-up. It could be improved by:
  - Removing pile-up only clusters
  - Suppressing pile-up contributions in mixed clusters
- A **definition of pile-up-only** clusters in MC simulations can be obtained based on their energy and response
- Suppressing pile-up-only clusters results in a **more narrow** distribution for $R^{\mathrm{EM}}_{\mathrm{clus}}$, with its IQR decreasing by:
  - 3 orders of magnitude in Run 2
  - 2 orders of magnitude in Run 3

Summary

- The presence of pile-up degrades collider measurements, especially at high luminosity
- The **topocluster response** worsens with high pile-up. It could be improved by:
  - Removing pile-up only clusters
  - Suppressing pile-up contributions in mixed clusters
- A **definition of pile-up-only** clusters in MC simulations can be obtained based on their energy and response
- Suppressing pile-up-only clusters results in a **more narrow** distribution for $R^{\mathrm{EM}}_{\mathrm{clus}}$, with its IQR decreasing by:
  - 3 orders of magnitude in Run 2
  - 2 orders of magnitude in Run 3

Summary

- The presence of pile-up degrades collider measurements, especially at high luminosity
- The **topocluster response** worsens with high pile-up. It could be improved by:
  - Removing pile-up only clusters
  - Suppressing pile-up contributions in mixed clusters
- A **definition of pile-up-only** clusters in MC simulations can be obtained based on their energy and response
- Suppressing pile-up-only clusters results in a **more narrow** distribution for $R^{\text{EM}}_{\text{clus}}$, with its IQR decreasing by:
  - 3 orders of magnitude in Run 2
  - 2 orders of magnitude in Run 3

Summary

- The presence of pile-up degrades collider measurements, especially at high luminosity
- The **topocluster response** worsens with high pile-up. It could be improved by:
  - Removing pile-up only clusters
  - Suppressing pile-up contributions in mixed clusters
- A **definition of pile-up-only** clusters in MC simulations can be obtained based on their energy and response
- Suppressing pile-up-only clusters results in a **more narrow** distribution for $R^{\mathrm{EM}}_{\mathrm{clus}}$, with its IQR decreasing by:
  - 3 orders of magnitude in Run 2
  - 2 orders of magnitude in Run 3

Outlook: Where Could This Go Next?

- The pile-up contributions in mixed clusters could be quantified by using the **DigiTruth** method https://cds.cern.ch/record/2677419
- One multi-class DNN could be implemented to classify all **three classes** at the same time
- The topocluster classification network could be combined with the one used for **energy calibration**
- The possibility to **define** mixed clusters **more strictly** could be explored

Thank you for listening!

Outlook: Where Could This Go Next?

- The pile-up contributions in mixed clusters could be quantified by using the **DigiTruth** method https://cds.cern.ch/record/2677419
- One multi-class DNN could be implemented to classify all **three classes** at the same time
- The topocluster classification network could be combined with the one used for **energy calibration**
- The possibility to **define** mixed clusters **more strictly** could be explored

Thank you for listening!

Outlook: Where Could This Go Next?

- The pile-up contributions in mixed clusters could be quantified by using the **DigiTruth** method https://cds.cern.ch/record/2677419
- One multi-class DNN could be implemented to classify all **three classes** at the same time
- The topocluster classification network could be combined with the one used for **energy calibration**
- The possibility to **define** mixed clusters **more strictly** could be explored

Thank you for listening!

Outlook: Where Could This Go Next?

- The pile-up contributions in mixed clusters could be quantified by using the **DigiTruth** method https://cds.cern.ch/record/2677419
- One multi-class DNN could be implemented to classify all **three classes** at the same time
- The topocluster classification network could be combined with the one used for **energy calibration**
- The possibility to **define** mixed clusters **more strictly** could be explored

Thank you for listening!

Outlook: Where Could This Go Next?

- The pile-up contributions in mixed clusters could be quantified by using the **DigiTruth** method https://cds.cern.ch/record/2677419
- One multi-class DNN could be implemented to classify all **three classes** at the same time
- The topocluster classification network could be combined with the one used for **energy calibration**
- The possibility to **define** mixed clusters **more strictly** could be explored

Thank you for listening!

Backup – PU suppression at topocluster level

- Constituent Subtraction + Soft Killer:
  - Divide the event in patches and evaluate pile-up density $\rho = \underset{i \in \text{pathces}}{\text{median}} \left\{ \frac{p_{T,i}}{A_i} \right\}$
  - **CS**: overlay the event with ghosts, whose momentum is determined by ρ, and modify the topoclusters momentum by subtracting those of the close-by ghosts
  - **SK**: impose a momentum threshold, chosen as the minimal cut for which $\rho = 0$

- Since Run 3: if a **cell time** is outside the collision time window ($|t_{\text{cell}}| > 25\,\text{ns}$) and its significance is below 10, it isn't included in the topocluster

Backup – PU suppression with PFlow objects

- Particle Flow objects are built by trying to match each topocluster to a **track**

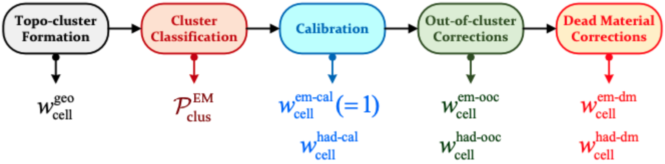
$\Longrightarrow$ PU contributions coming from charged vertices are naturally suppressed

- Neutral contributions are suppressed by **Pile-Up Per Particle Identification** (PUPPI):
  - Assign to each neutral object a probability α_i to originate from the primary vertex (based on its proximity to charged HS products)
  - Use the distribution of α for charged PU products as a reference (let's call its mean value $\bar{\alpha}$ and RMS σ)
  - Remove objects with $\alpha_i < \bar{\alpha}$, re-weight the others as: $w = F_{\chi^2}(\Theta(\alpha_i - \bar{\alpha})\frac{\alpha_i - \bar{\alpha}}{\sigma^2})$

Backup – Topocluster Energy Calibration

- Two possible methods:

  1. **Local Hadronic Cell Weighting** (LCW) calibration (currently used):



https://doi.org/10.1140/epjc/s10052-017-5004-5

  2. **Machine Learning** based calibration:
     - A DNN reconstructs the topocluster response R^{DNN}
     - The true energy is $E^{\text{dep}} = E_{\text{clus}}^{\text{EM}} / R^{\text{DNN}}$

https://cds.cern.ch/record/2866591

Backup – Formulae for Network Description

- `ReLu` function: $f(x) = \max\{0, x\}$
- `Sigmoid` function: $f(x) = \frac{1}{1+e^{-x}}$
- **Binary cross-entropy**:

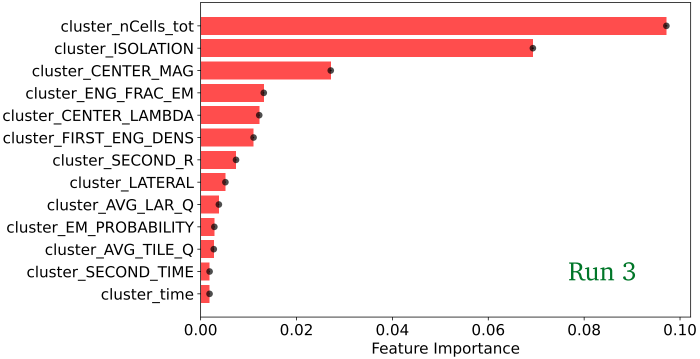
  $$\mathcal{L}(y, p) = -\frac{1}{N}\sum_{i=1}^{N}\left[y_i \log(p_i) + (1-y_i)\log(1-p_i)\right]$$

  - Weight of each term: $w_i = \frac{N}{2 \cdot N_i}$
- **Permutation importance** of feature j,evaluated with K repetitions:

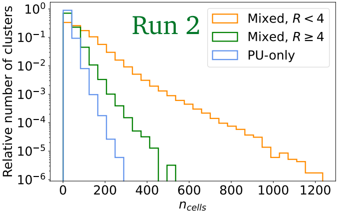
  $$i_j = s - \frac{1}{K}\sum_{k=1}^{K} s_{k,j}$$

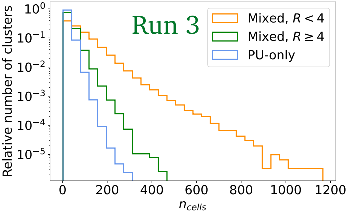
  - s is the network score
  - $s_{k,j}$ is the score after shuffling feature j, in repetion k

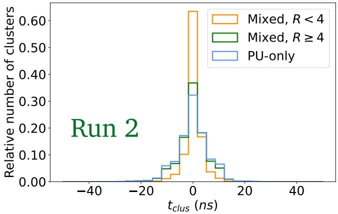
Backup – Feature importance in Run 3

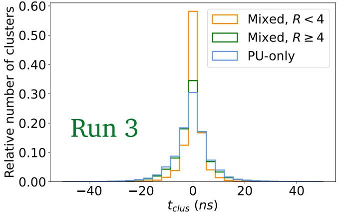


Backup – Feature Distributions, PU Identification – I

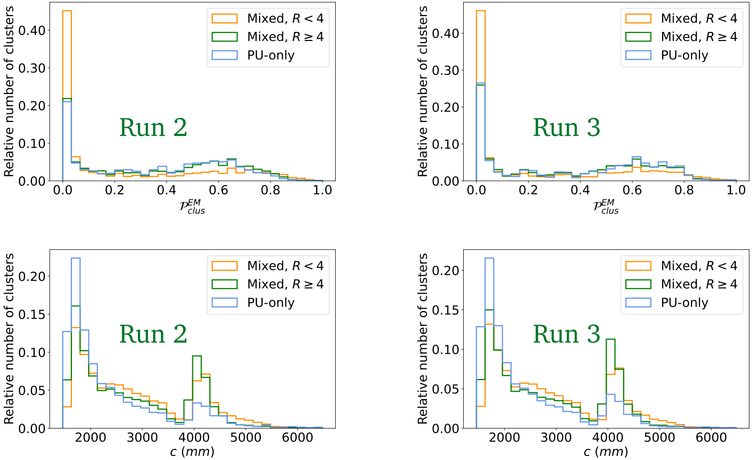




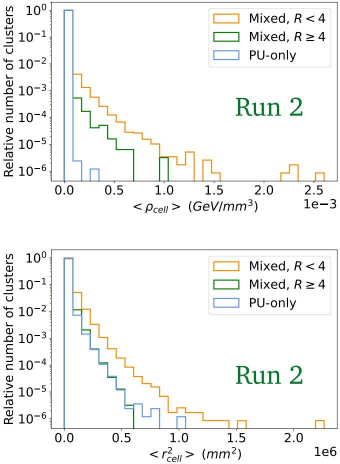




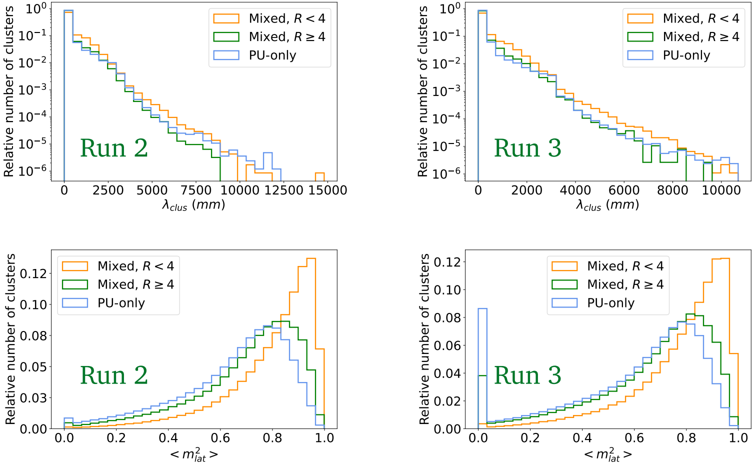
Backup – Feature Distributions, PU Identification – II



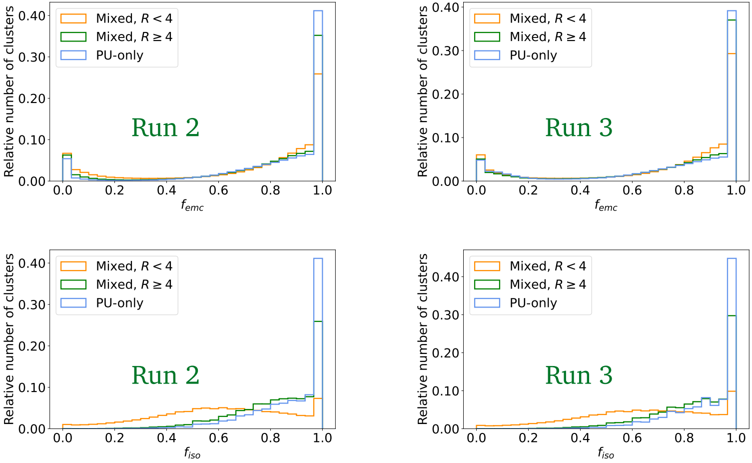
Backup – Feature Distributions, PU Identification – III



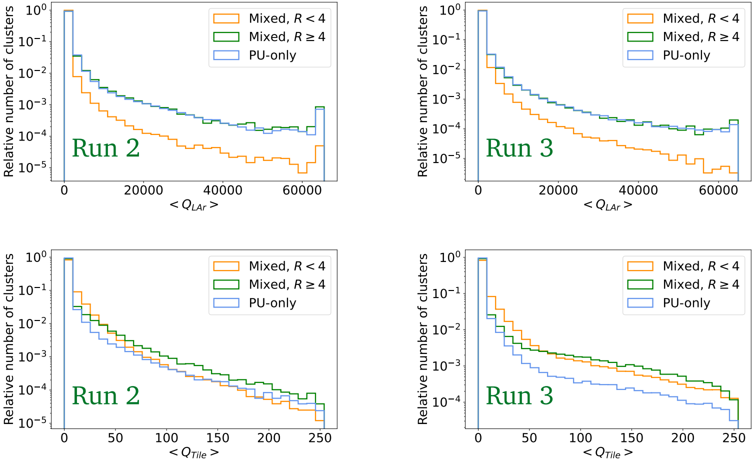
Backup – Feature Distributions, PU Identification – IV



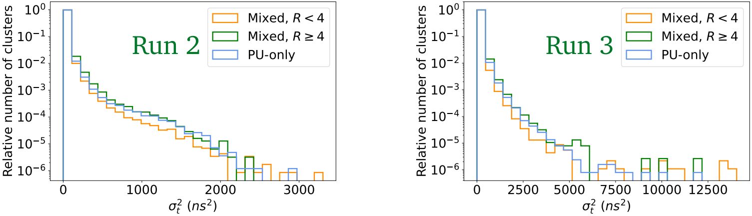
Backup – Feature Distributions, PU Identification – V



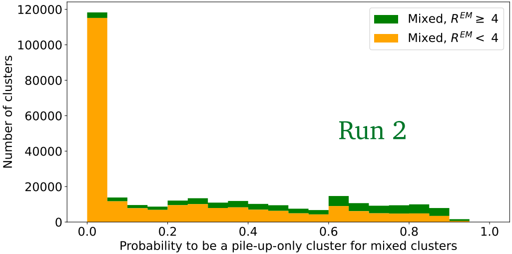
Backup – Feature Distributions, PU Identification – VI

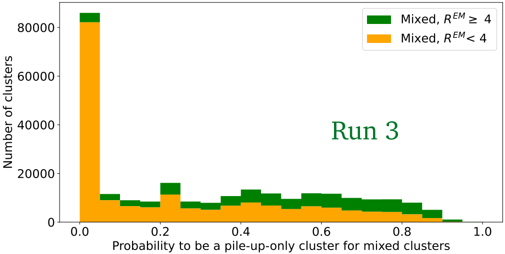


Backup – Feature Distributions, PU Identification – VII

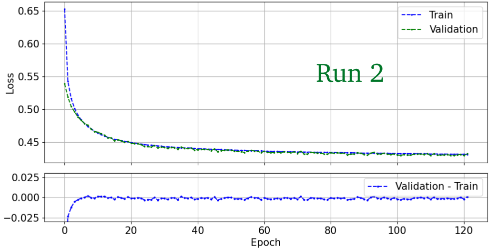


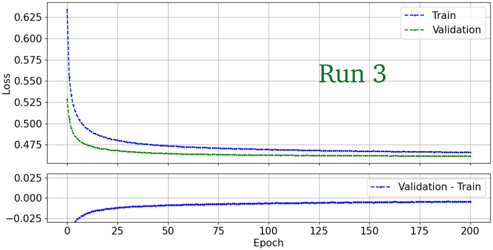
Backup – Probability to Be PU-Only of Mixed Clusters





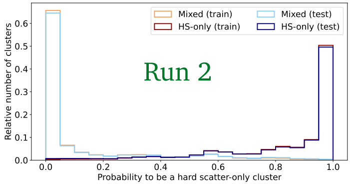
Backup – PU Identification, Loss Curves

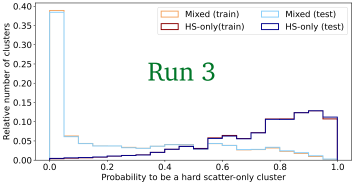




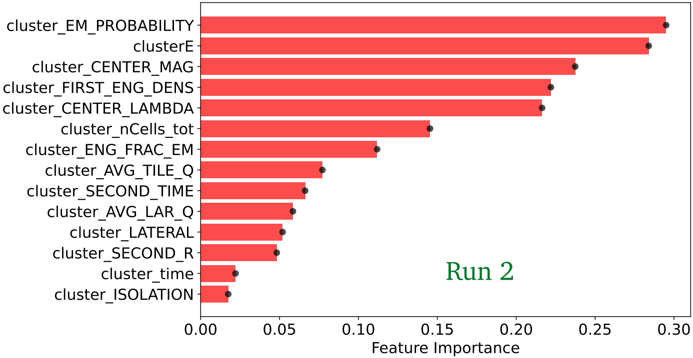
Backup – HS Identification, Output

- Same jet p_T and simulated years as PU Identification
- **HS-only**: simulation without pile-up
- **Mixed**: full simulation, clusters from events with $\mu > 20$ to avoid including HS-only
- **Classification output**:
  - Worse for Run 3, possibly due to the additional PU suppression

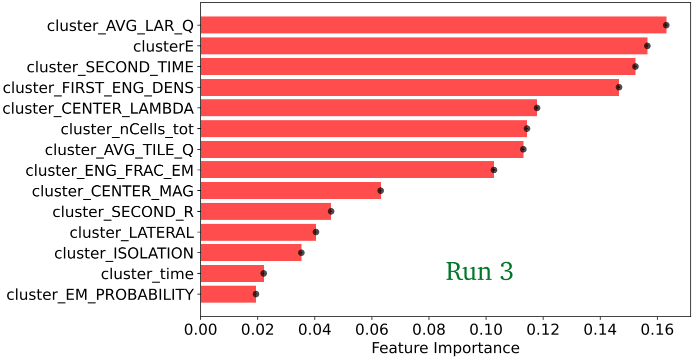




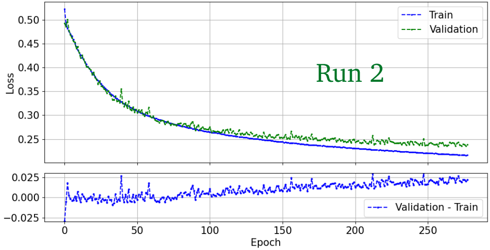
Backup – HS Identification, Feature Importance

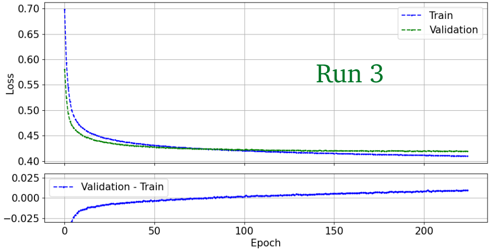


Backup – HS Identification, Feature Importance

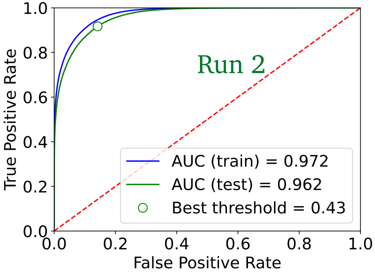


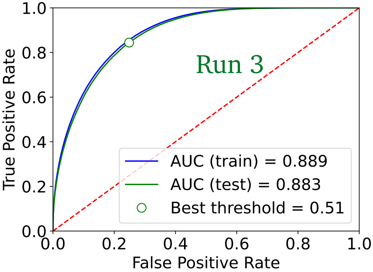
Backup – HS Identification, Loss Curves





Backup – HS Identification, Performance





- Worse performance for Run 3
  - Possibly due to additional PU suppression in Run 3 (timing cut)

Backup – HS Identification, Responses