

Machine Learning Anomaly Detection techniques for the search of new physics at LHC with ATLAS experiment

Annalisa Berti, Nuno Castro, Rute Pedro

2/10/26

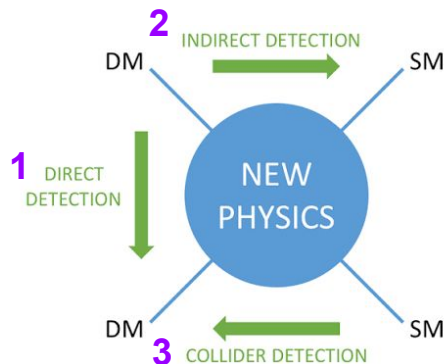
Ref. UI/BD/154673/2023

The search for DM

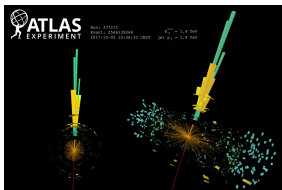
The hypothesis of Dark Matter comes from a long history of experimental evidence entailing that something in the current cosmological model is wrong.

- Galaxies rotational speed [1]
- Gravitational lensing effects
- The anisotropies in the CMB [2]

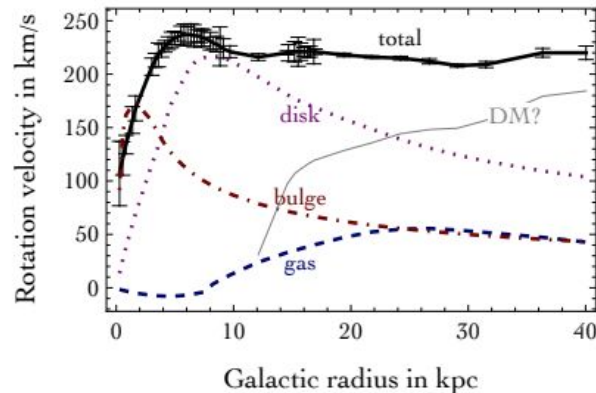
Something only-gravitationally interactive, non-relativistic and long-lived is out there and, nowadays, there are three ways we look for it:



1. Nuclear recoil from DM-nucleus scattering
2. High energy photon, neutrino, antimatter coming from high-density DM regions
3. **Missing transverse momentum**



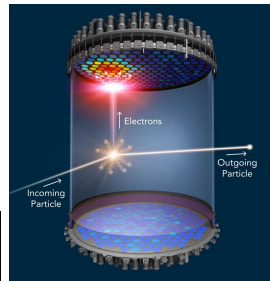
UGC03205 (denser, baryonic, more stars)



Cirelli, Marco, Alessandro Strumia, and Jure Zupan. "Dark matter." *arXiv preprint [arXiv:2406.01705](https://arxiv.org/abs/2406.01705)* (2024).

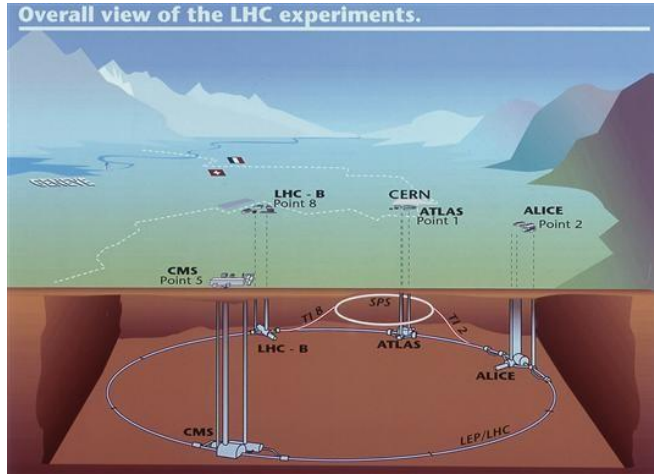


[AMS-02 experiment](#)



[LUX-ZEPLIN experiment](#)

ATLAS experiment on the Large Hadron Collider



The LHC is the biggest particle accelerator in the world. It has **4 major detectors** with different functionings and purposes.

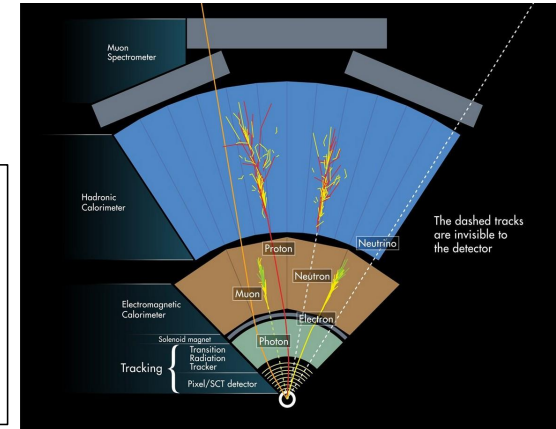
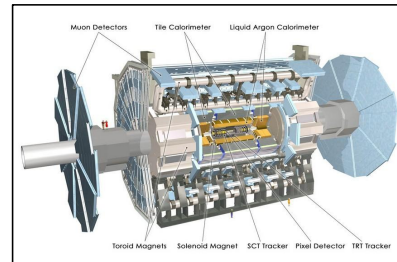
The state of the art technology allows to have **proton-proton collisions** with a center of mass energy of 13.6TeV.

After each collision, a cascade of processes is generated, which final products are detected by the detector.

The ATLAS detector is composed by a **sub-system of detectors** each dedicated to a different kind of particle.

The whole system allows for:

- ❑ Particle Identification
- ❑ Collision vertex reconstruction
- ❑ Momentum measurement
- ❑ Energy measurement



Search for new physics in the mono-jet channel

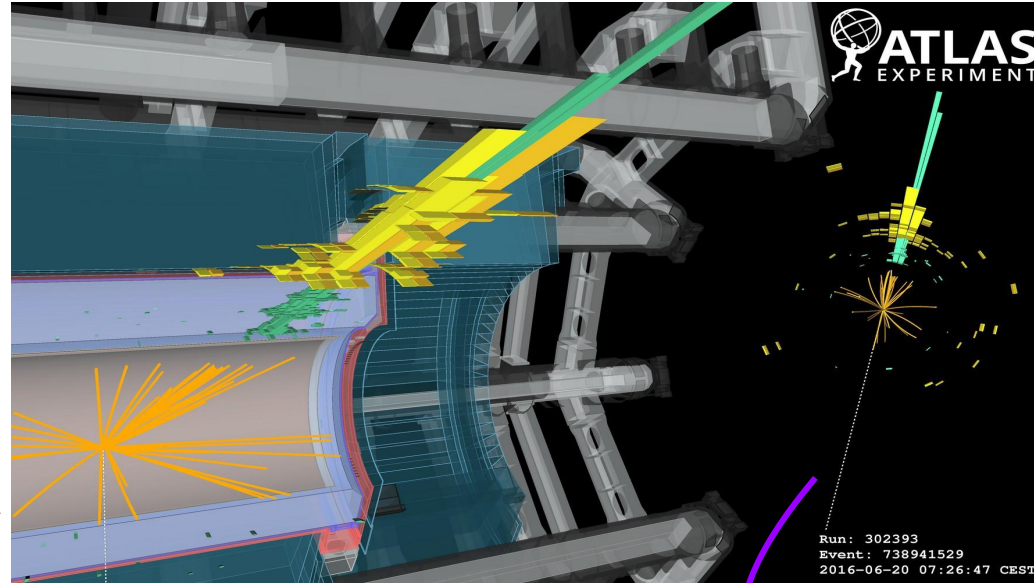
Targets events with:

- High missing transverse momentum ($ET_{\text{miss}} > 250 \text{ GeV}$).

Dark matter, like neutrinos, would leave the detector unseen and show up as missing transverse momentum (stable and non-interacting).

- At least one large-radius jet (the experimental signature of quarks, gluons and boosted W, Z or H bosons).

A highly energetic jet recoiling against the missing momentum acts as a tag. It marks events where the invisible particles were produced together with visible radiation, so the momentum imbalance can be observed and the events stand out from typical collisions.



Missing Transverse Energy

$$p_{\text{T}}^{\text{miss}} = - \left(\underbrace{\sum_{\text{selected electrons}} p_{\text{T}}^e + \sum_{\text{accepted photons}} p_{\text{T}}^{\gamma} + \sum_{\text{accepted } \tau\text{-leptons}} p_{\text{T}}^{\tau} + \sum_{\text{selected } \mu} p_{\text{T}}^{\mu}}_{\text{hard term}} + \underbrace{\sum_{\text{accepted jets}} p_{\text{T}}^{\text{jet}} + \sum_{\text{unused tracks}} p_{\text{T}}^{\text{track}}}_{\text{soft term}} \right)$$

Search for new physics at collider experiments

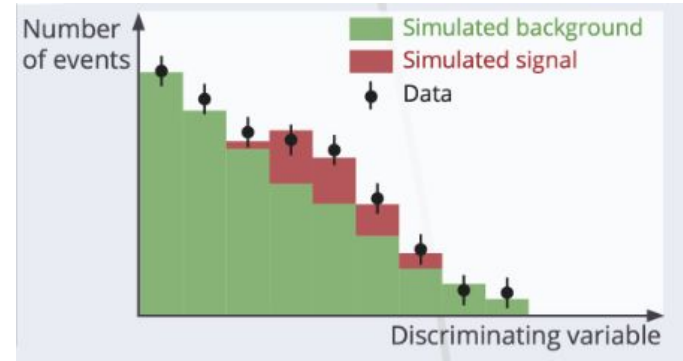
In canonical searches, you target a specific beyond standard model theory or effective theory and, looking at the data,

- First, you compare data and the simulated background-only distribution.

If the probability of observing that data, given the background-only distribution is lower than 3×10^{-7} $\rightarrow$ You claim a discovery!

- If the agreement with the background is acceptable, you repeat the procedure scanning the signal-plus-background hypothesis.

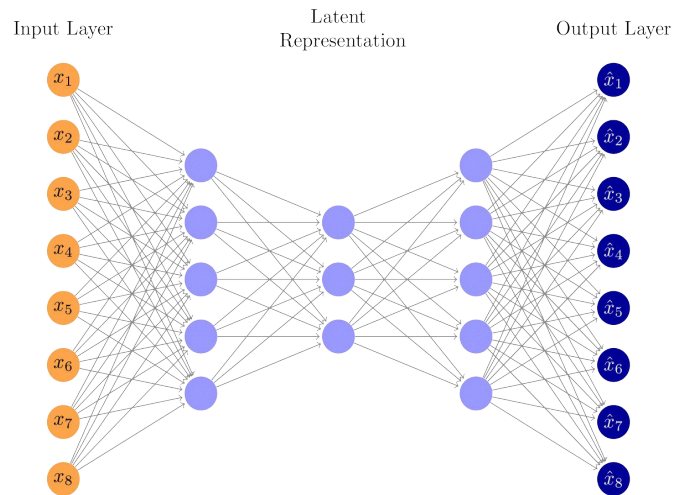
Every signal excess you exclude sets an upper limit on the signal cross-section, that can then be translated into limits in the parameter space of your theory.



A novel approach with Anomaly Detection

Anomaly detection is a machine learning technique designed to spot anomalies in data.

It finds an application in many different areas of the LHC experiments ([search for new physics](#), jet tagging, trigger selection, [data quality](#)).



A widely used architecture is the Autoencoder.

The objective is to fit a deep neural network (which is basically a very complex, non-linear function) designed to reconstruct the input data, using part of the simulated background.

Then give simulated and real data as input to the trained model.

Eventually, the reconstruction error distributions can be used to lead the statistical analysis described in the previous slide:

1. Investigate any significant deviation from the background-only model.
2. Choose a set of benchmark signal models to set upper limits.

PhD progress

To get to the point in which an ATLAS analysis can make meaningful physical claims, it must first do the following:

1. Object definition: the process of turning the detector-level information into well-defined physics objects that can be used in the analysis. 
2. Individuate the background processes populating your channel and
 - a. if simulable, use the centrally produced MC simulated samples, 
 - b. if not, refine the event selection criterium to cut them out.
3. Propagate the instrumental and theoretical uncertainties on the simulated data. 
4. (in our case) train a robust and efficient AD model. 
5. Define control regions where data can be fitted to obtain good data/MC agreement. 

Then one can perform the statistical data analysis on the AD score.

PhD progress - background composition

Analysis cuts to get the mono-jet all-hadronic channel:

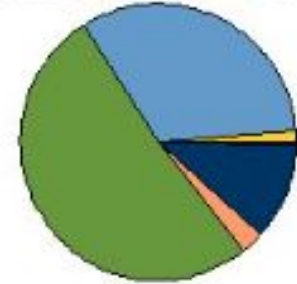
- Lowest un-pre-scaled MET trigger
- $\text{MET} > 200\text{GeV}$
- **Lepton Veto** (we select a specific kind of final states for the jet)
- $n \text{ Large R Jets} \geq 1$



Background processes simulated:

- **Z + jets** (Z can decay into a couple of neutrinos appearing in the detector as pure missing energy)
- **W + jets** (W can decay in a lepton-neutrino couple and the lepton doesn't get reconstructed)
- Diboson (same as above)
- **Ttbar** (each top quark can decay hadronically or semi-leptonically with the lepton mis-reconstructed via Wb)
- TtbarX (same as above)
- Single top (same as above)
- Other rarer events

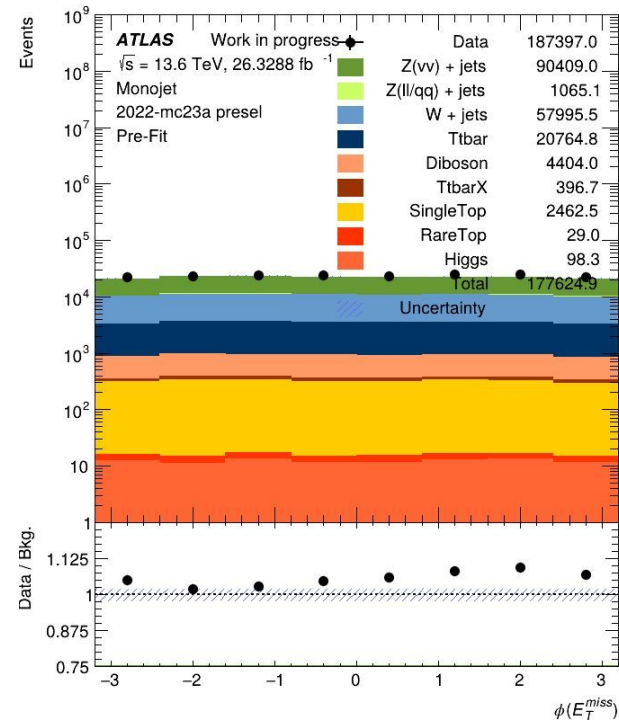
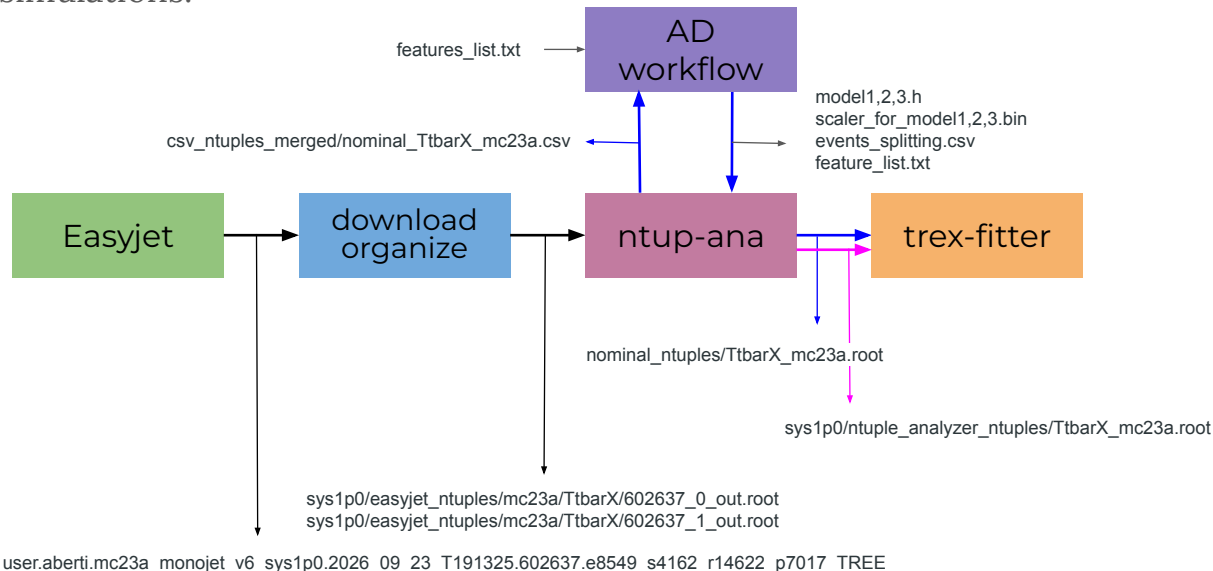
2022-mc23a PRESEL



PhD progress - setup workflow

The workflow to produce the data and MC samples with the object definition and selection cuts needed in the analysis is set and solid.

It can process Run2 ('15-'18) and Run3 ('22-'24) samples and the respective simulations.

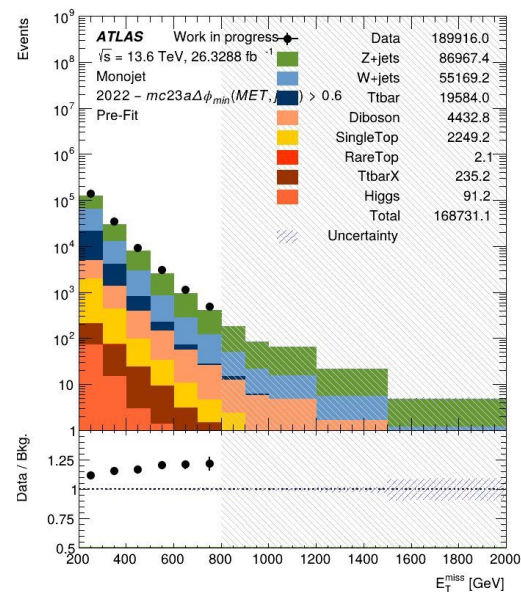
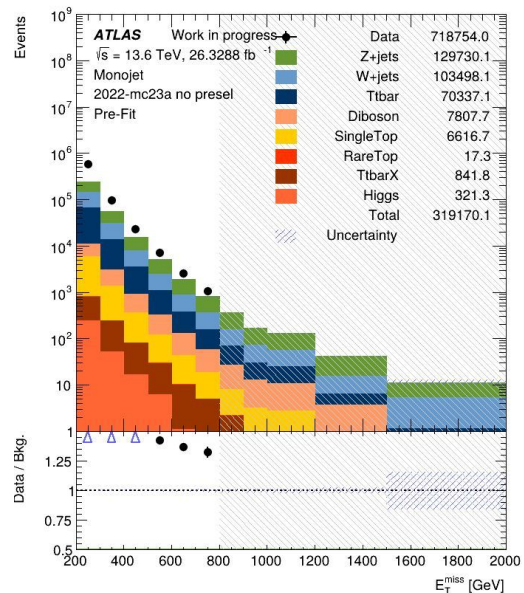
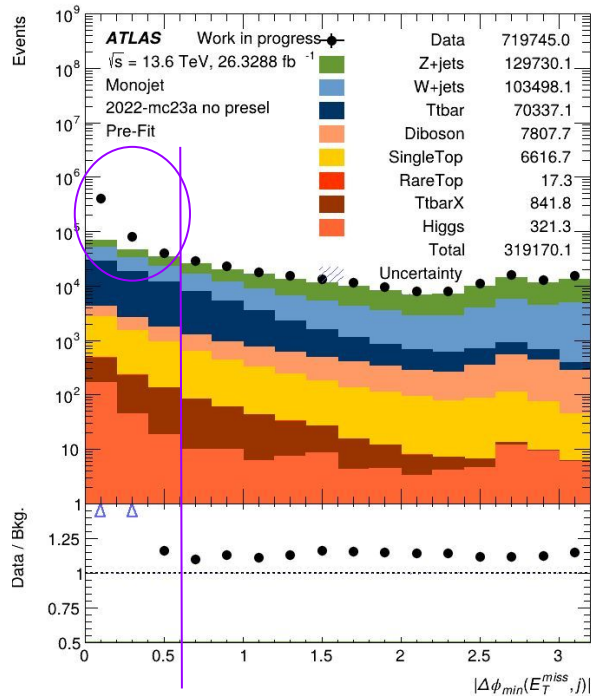


Missing energy distribution angle on the plane transversal to the beam.

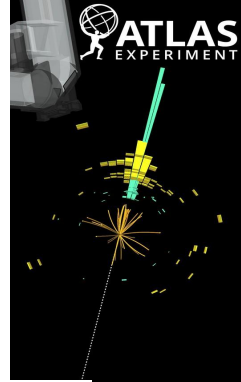
PhD progress - preselection

After the analysis selection, another selection has appeared necessary to remove from the data the contributions from instrumental induced background and events containing fake jets.

- $\Delta\phi_{\min}(\text{MET}, \text{jets}) > 0.6$ (angular distance between missing pT and its closest jet).



Data/MC agreement in preselection region before and after the $\Delta\phi_{\min}(\text{MET}, \text{jets})$ cut.

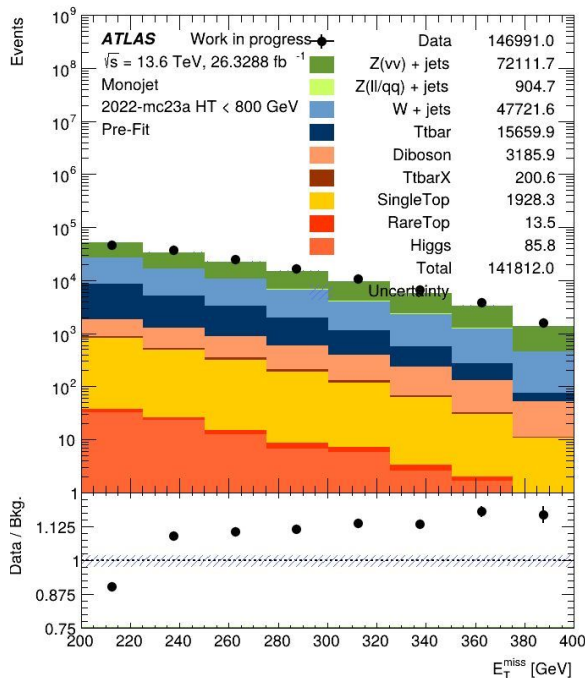


PhD progress - uncertainty propagation

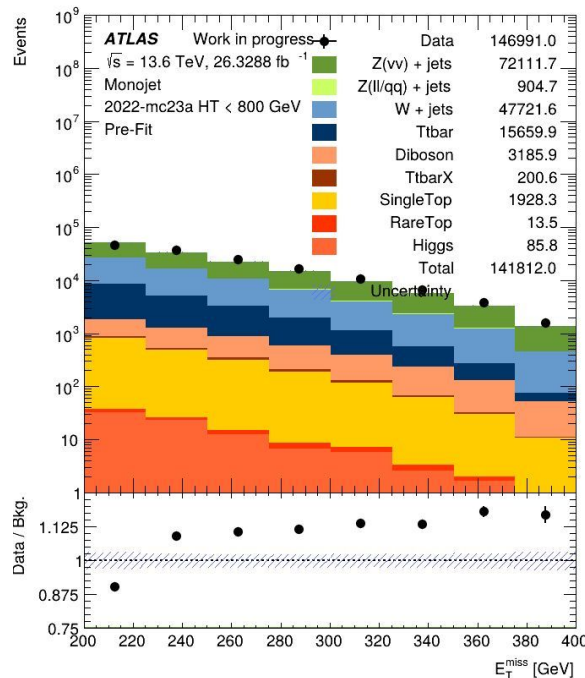
The uncertainties can be grouped into instrumental uncertainties (from the detector and its description) and modeling uncertainties (from the theoretical assumptions used in MC simulation).

All relevant sources of systematic uncertainty, instrumental and modeling, are propagated as nuisance parameters in the profile likelihood fit

Missing energy distribution in the control region ($HT < 800$ GeV) without uncertainties.

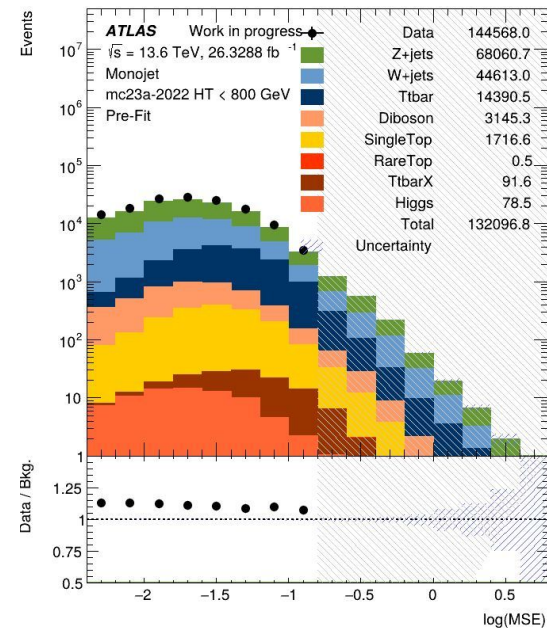
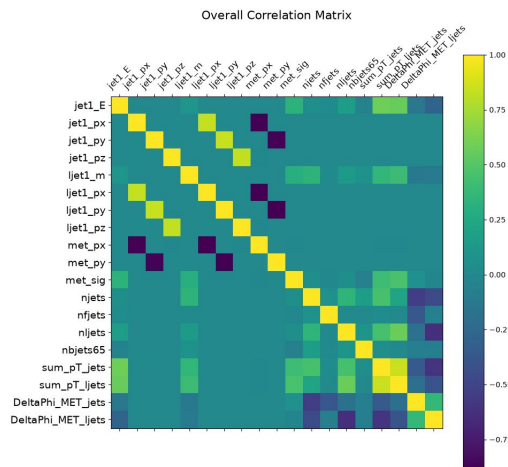


Missing energy distribution in the control region ($HT < 800$) with luminosity and jet-related uncertainties.



PhD progress - Machine Learning model

- The first tests were done with a very basic architecture (30,10,7,10,30).
- Currently we are playing around with the input features in order to find the ones giving the best performance of the model.
- Features used:
 - leading small R jet energy and momentum
 - leading large R jet energy, momentum, and mass
 - missing momentum and missing energy significance
 - number of small/small and b-tagged/large/forward jets
 - scalar sum of the total jets momenta
 - $\Delta\phi_{\min}(\text{MET}, \text{small R jets})$, $\Delta\phi_{\min}(\text{MET}, \text{large R jets})$



Data/MC agreement for anomaly score distributions ($\log(\text{MSE})$) in control region ($HT < 800 \text{ GeV}$)

The features are mostly uncorrelated, and in case of correlation it's expected.

Conclusions

Machine learning is an extremely useful tool across the LHC experiments, from data quality monitoring to the search for new physics.

Dark Matter is one of the biggest mysteries of modern science and a light, weakly interacting particle (WIMP) in the GeV-TeV range is experimentally accessible at the LHC.

The mono-jet channel is ideal for a signal-agnostic search, because WIMPs are stable, non-interacting particles that escape the detector as missing transverse momentum.

Anomaly Detection changes the approach. No tested BSM theory has yet delivered a Dark Matter candidate, so AD opens the door to inverting the theory-experiment relationship: the data guide the search, not the model.

Even without a discovery, AD adds a new possibility: setting upper limits on many different signal models at once. Sensitivity will likely be lower than a dedicated search, but the number of possible signal models is unlimited while time and resources are not, so testing more models simultaneously is a real advantage.

Thank you for the
attention!