Status Report

Hanyang University High Energy Physics Laboratory

Sungbeom Cho

(sucho@cern.ch)

Status

2025 Induction courses of my interest.
https://indico.cern.ch/event/1467816/timetable/

Create eos storage & CERNBox.

Ticket No: RQF3036547, **Opened:** 09-02-2025 12:07:21
Short description: EOS Home Folder Request
Message from Andreas Pfeiffer:

Dear Sunbeam Cho,

your CERNBox folder is visible on lxplus under /eos/user/s/sucho/ ...

HTH — stay safe and healthy !
cheers, Andreas

CMSSW

11:30 AM → 12:30 PM **Hand-on Offline activities**
Speaker: Phat Srimanobhas (Chulalongkorn University (TH))
CMSSW 101 | GMT20250205-101...

ML4DQM/DC

2:50 PM → 3:00 PM **Hands-on PPD/Data Quality Monitoring : Introduction**
Speaker: Antonio Vagnerini (University of Nebraska-Lincoln)
DQMoverview.pdf | GMT20250207-134...

3:05 PM → 3:25 PM **Hands-on DQM : Data Certification Automation (DC3) tutorial: golden-json creation**
Speaker: Cristiano Tarricone (Universita e INFN Torino (IT))
DChandsOn_CMSin...

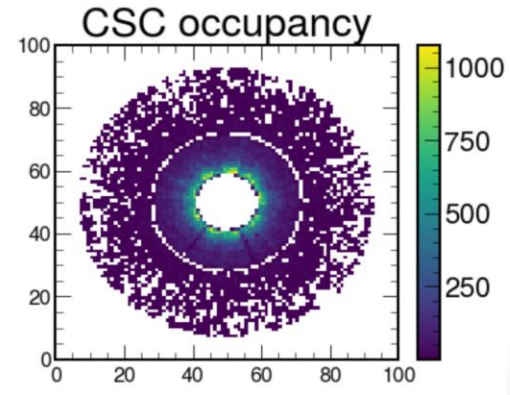
3:30 PM → 4:00 PM **Coffee Break**

4:00 PM → 4:30 PM **Hand-on PPD/Machine Learning prototyping using DIALS: CSC as benchmark**
Speaker: Marco Buonsante (Universita e INFN, Bari (IT))
CMS_Induction_DIA... | GMT20250207-150...

4:35 PM → 4:55 PM **Hand-on PPD/Model Deployment on DIALS: ML-json creation**
Speaker: Gabriel Moreira (Universidade do Estado do Rio de Janeiro (BR))
2025-02-07 Model d... | GMT20250207-153...

Hands on tutorials

```
Show2Dimg(summed_data["img"][0])
```




```
import numpy as np
import matplotlib.pyplot as plt
from cmsdials.auth.bearer import Credentials
from cmsdials import Dials
from cmsdials.filters import (
    RunFilters,
    LumisectionHistogram1DFilters,
    MLBadLumisectionFilters,
    MLModelsIndexFilters
)
```

MC Generation

4:00 PM → 4:25 PM **Hands-on PPD/Physics Data MonteCarlo Validation**
Speakers: Samadhan Kamble (Indian Institute of Technology Madras (IN)), Soumyadip Barman (Tata Institute of
CMS_Samadhan_In... | GMT20250205-150... | Induction-PdmV_So...

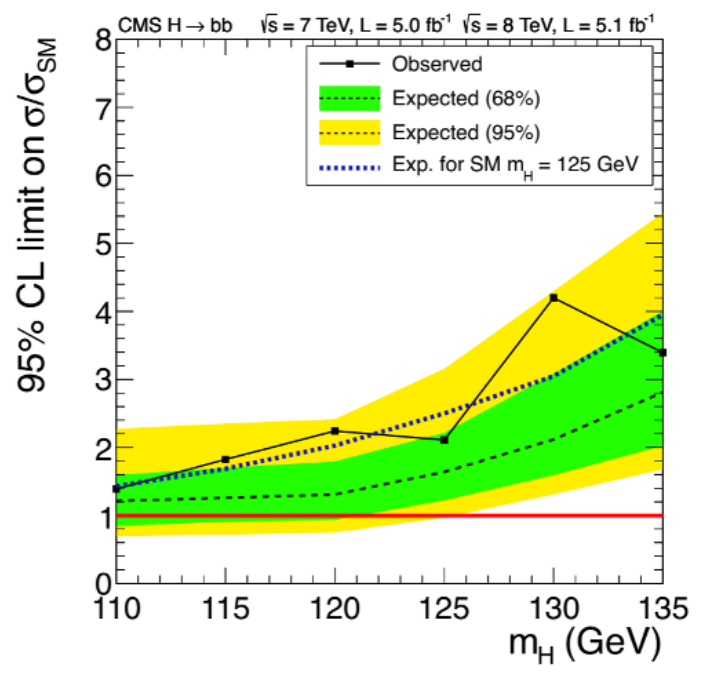
H(bb) reconstruction in SS2l

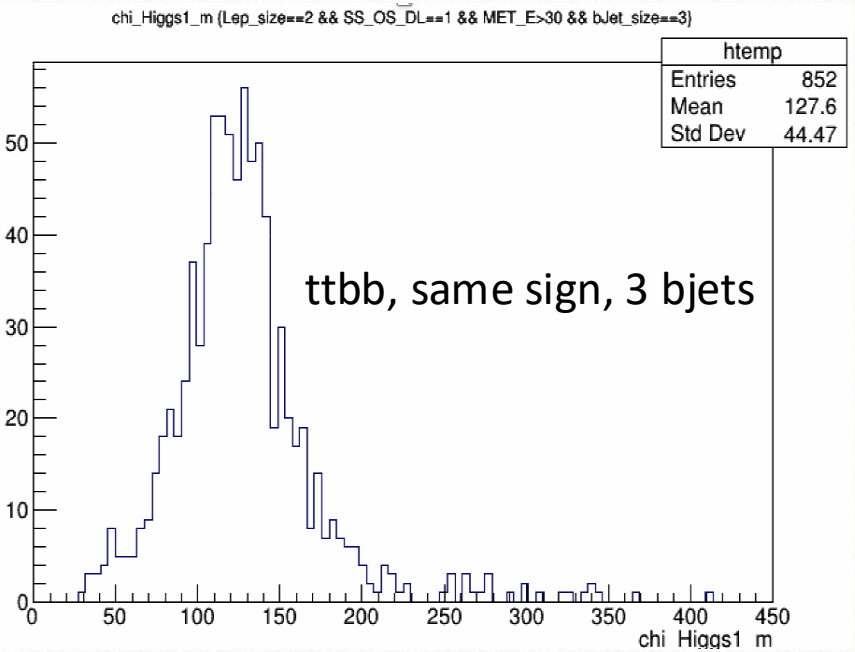
My personal opinion is, H(bb) itself only doesn't seem so nice..

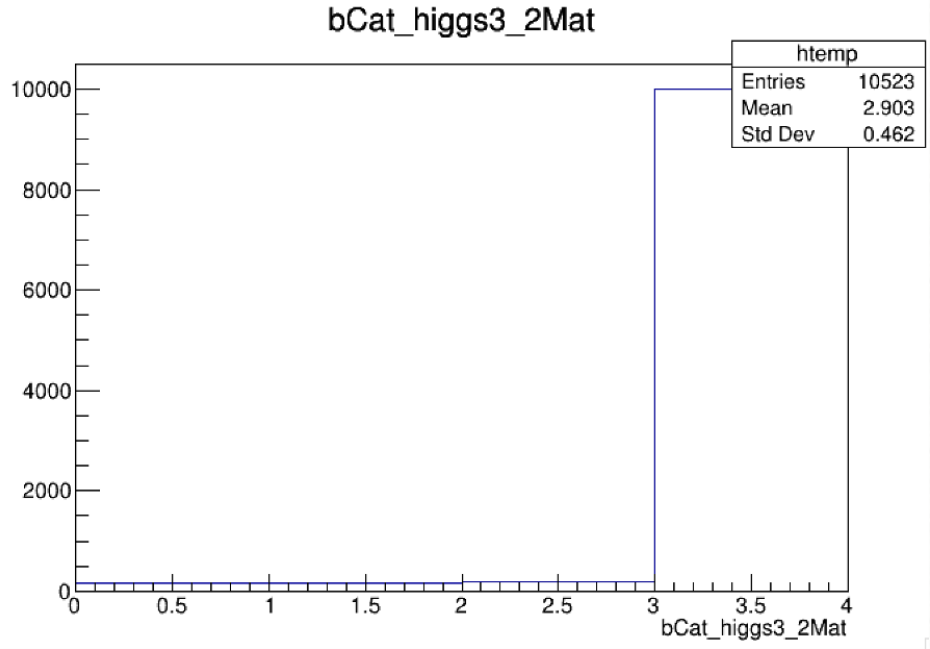
Reason1. Already reported, resolution of H>bb is unclear enough.

Reason2. I see combinatorial problem with **at least 3 jets.**

Reason3. Matchable event ratio = below 10%.



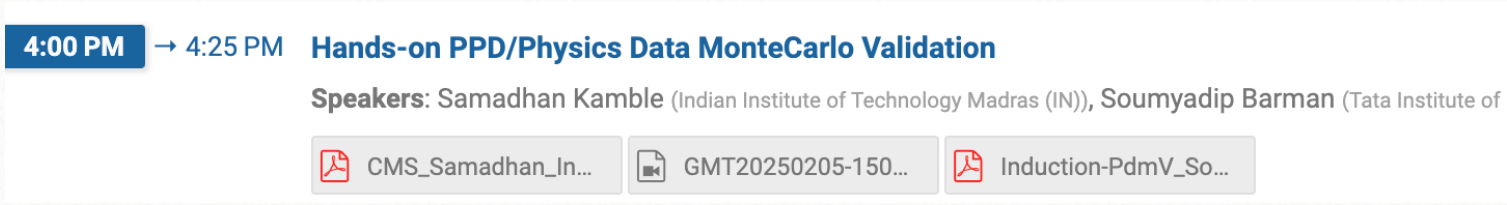




ttHH

My suggestion is, inspect leptons more.

And to be honest, try generating central generation of ttHH (bbWW).

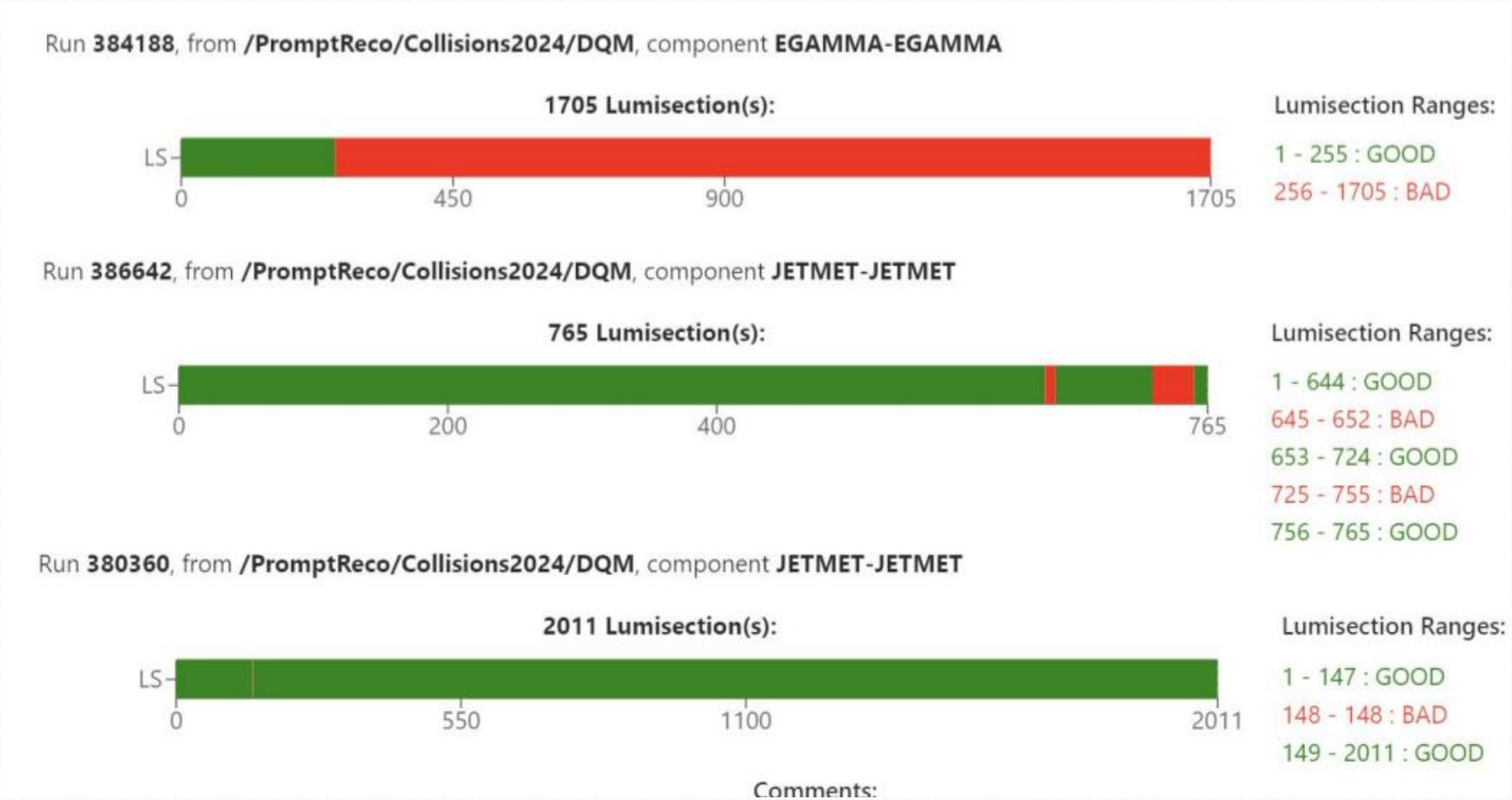


--> I can do this.

Back – up

Motivation of ML-DQM DC

In 2024, ~30fb of good for physics data comes from partially BAD runs certified by DC.



Taken from Cristiano's slide :
https://indico.cern.ch/event/1467816/timetable/

Motivations of ML-DQM DC

ML based DC is paradigm changing for the future.

1. Cope with larger data.
2. Reduce tedious human errors.
3. Closer inspection for data per lumi-section.
4. AE gives additional insights on detector performance.

Personal : Good starting point to engage in CMS workflow. Golden Json is what should understand.

Contribution? :

- No optimal official ML model yet. Lots of candidates, where we can also contribute.
- In CMS, CSC detector project is on going. --> I believe I can develop RPC model as well.

```
# Autentication
creds = Credentials.from_creds_file()
# workspace definition
dials = Dials(creds, workspace="csc")
```

*Starting point : Create "rpc workspace"

```
# Autentication
creds = Credentials.from_creds_file()
# workspace definition
dials = Dials(creds, workspace="rpc")
```

Workflow

Step 1. DIALS aggregate offline MEs into pandas format.

Step 2. ResNet AutoEncoder for anomaly detection. --> **There is no optimized model yet!**

Step 3. Create Golden Json.

https://github.com/Ma128-bit/17th-CMS-Induction-Course.git Hands on Tutorial

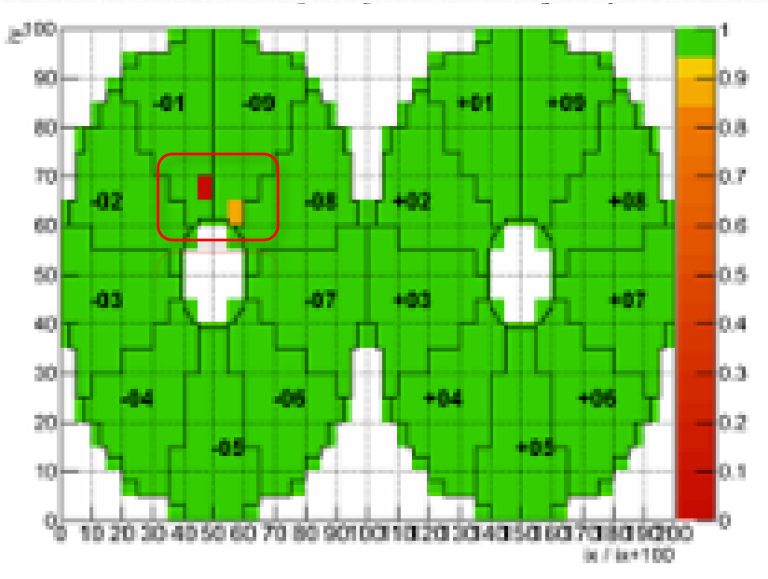
What is DIALS? : Data Inspector for Anomalous Lumi-Sections.

Python API, working on Jupiter.

Also it converts to pandas data frame easily, which is good for ML inputs.

```
# cmsdials utilities:
from cmsdials.auth.bearer import Credentials
from cmsdials import Dials
from cmsdials.filters import RunFilters, LumisectionFilters, MEFilters
from cmsdials.filters import LumisectionHistogram2DFilters
```

Anomaly Detection



We want to figure out..

Normal Data vs Anomalous Data, like we do in Online DQM.

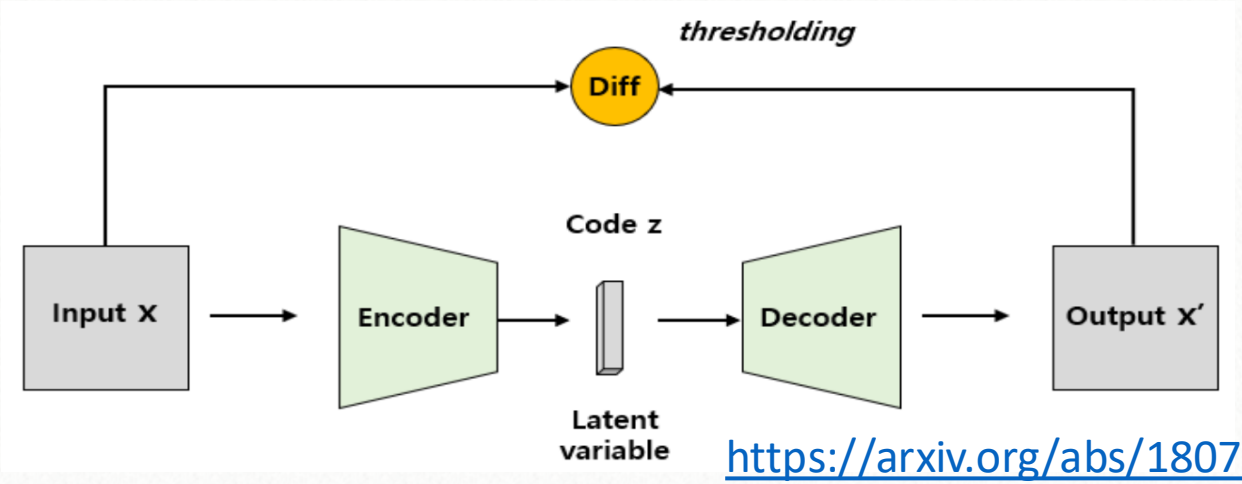
Anomaly Detection Concept.

Train with Normal Data == **Learn Normal Latent Space.**

Normal data will be reconstructed well from the **Normal Latent Spa**

Anomaly data will be poorly reconstructed from the **Normal Latent Space. --> Big Loss.**

***Loss is defined as "Reconstructed img – Input img difference"**



https://arxiv.org/abs/1807.02011