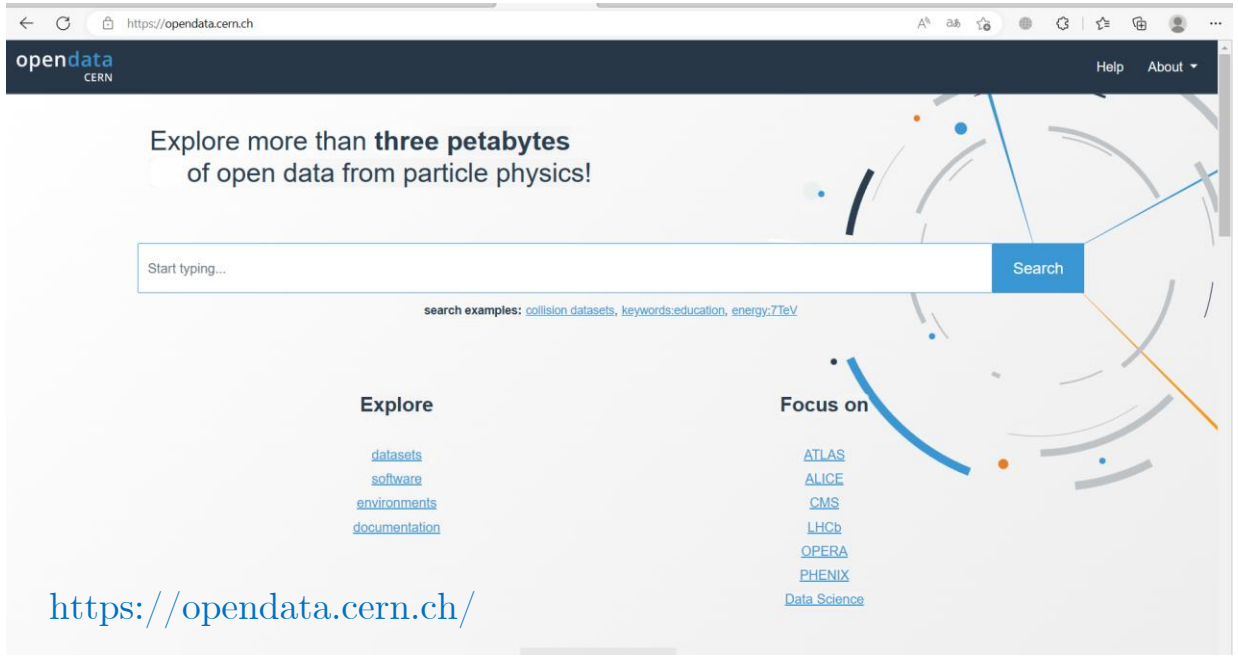
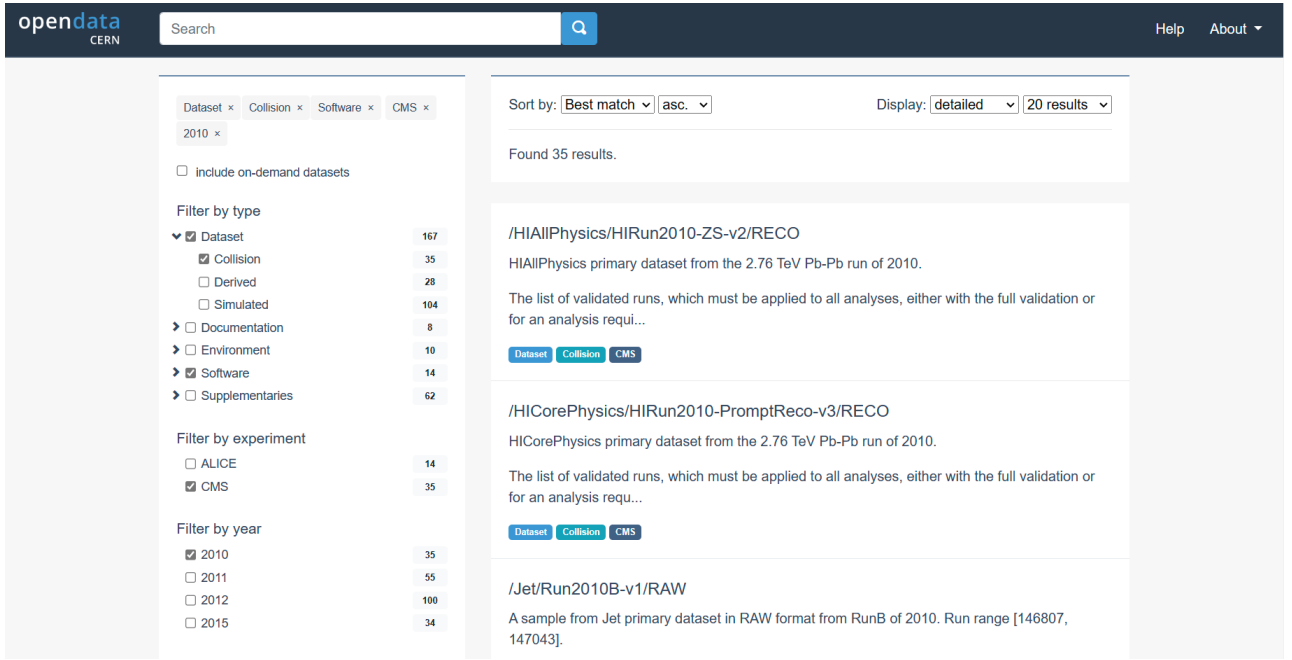
Data analysis of Run3 and RUN2015 Open Data of the CMS experiment at the LHC

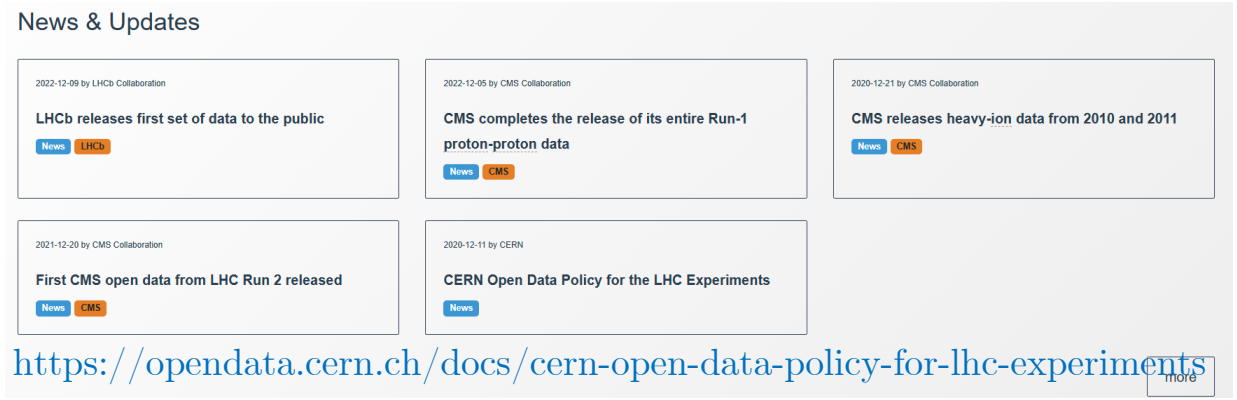
Tumasayn A.[1], Hayrapetyan A.[1], Korsakov I.[2], Lanyov A.[3], Shmatov S.[2]
[1]Yerevan Physics Institute, 2 Alikhanyan Brothers Street, Yerfi
[2]Meshcheryakov Laboratory of Information Technologies, JINR
[3]Veksler and Baldin Laboratory of High Energy Physics, JINR

Open Data portal(1)



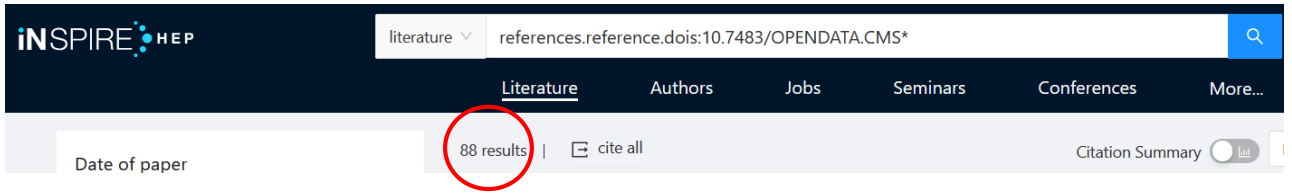
https://opendata.cern.ch/





https://opendata.cern.ch/docs/cern-open-data-policy-for-lhc-experiments

~ 90 articles



https://inspirehep.net/literature?sort=mostrecent&size=25&page=1&q=references.reference.dois%3A10.7483%2FOPENDATA.CMS%2A

Open Data portal(2)

2014, Nov — 2010 pp, 50%

First release, virtual machine environment

2016, Apr — 2011 pp, 50%

Simulated samples, validation examples, basic tools

2017, Dec — 2012 pp, 50%

More usage examples (Higgs), Jupyter notebooks

2019, Jul — 2010 pp, 100%

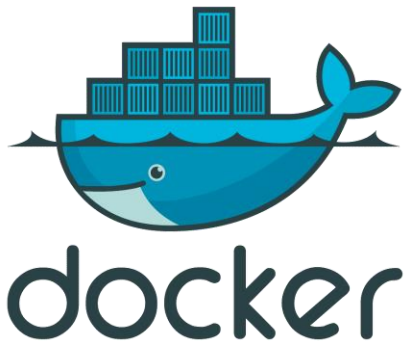
ML samples, special datasets, docker containers, simulated data generation tools

2020, Aug — 2011 pp, 100%

First examples of automated workflows, improved tools

2020, Dec — 2010-11 HI, 100%

First heavy-ion data release



2021, Dec — 2015 pp, 99%

First Run-2 data release, slimmer data format

2022, Dec — 2012 pp, 100%

Full Run-1 pp data release, improved usage examples

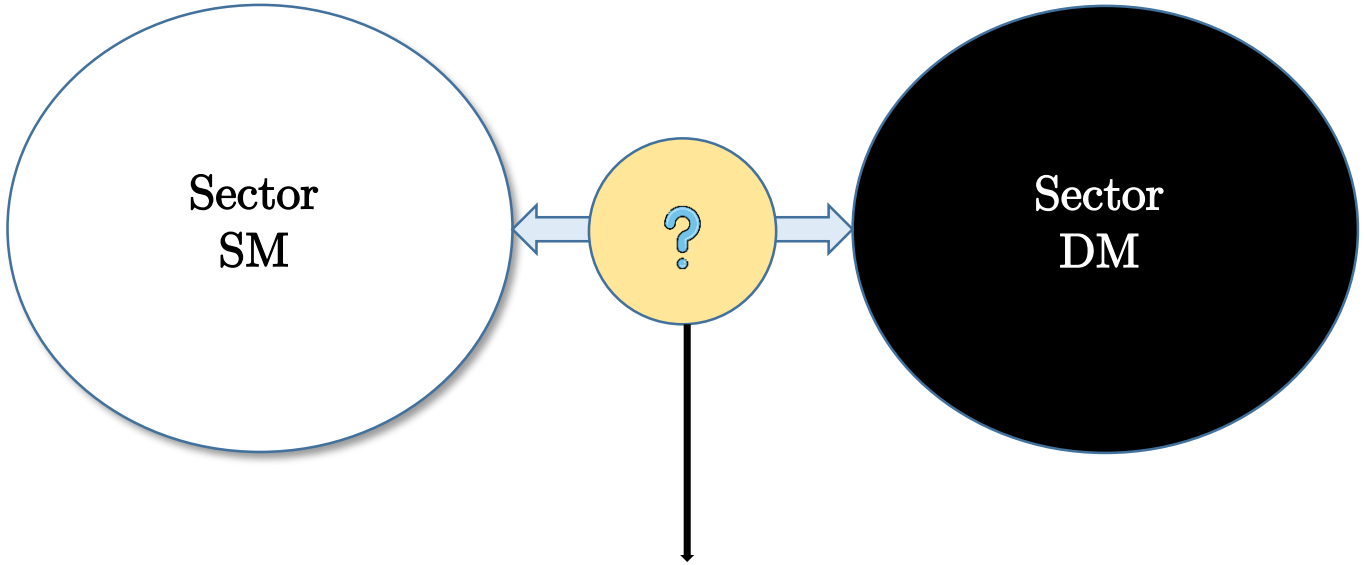
2023, Sept — 2013 HI, 100%

Full Run-1 heavy-ion data, extended usage examples

2024, Apr — 2016 pp, 50%

First NanoAOD, updated usage instructions

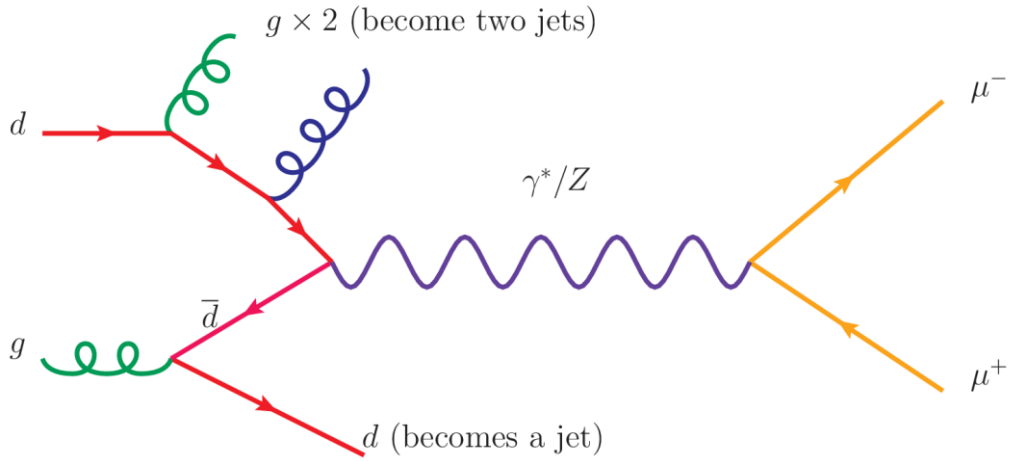
A simplified dark matter model – as a validation of using Open Data for research in BSM physics



Axial-vector mediator
$g_q = 0.1$,
$g_l = 0.01$.
$g_{DM} = 1.0$

Vector mediator
$g_q = 0.01$,
$g_l = 0.01$.
$g_{DM} = 1.0$

The Drell-Yan process is the main background in the search for signals of new physics beyond the Standard Model (SM), particularly in the search for candidate particles for the role of the Dark Matter



Open Data and generated event samples. Muons selections

$L_{\text{int}} \approx 2649\ pb^{-1}$

Reconstructed datasets :
SingleMuon/Run2015D-16Dec2015v1/MINIAOD

Muon selections

- high accuracy of the p_T calculation: $\delta p_T/p_T < 0.3$
- Isolation: (IsoPt < 0.1)
- Identification: “Global” and “Traker” muons, Tracker layers>5, PixelHits>0, MuonHits>0
- Kinematic cuts: $|\eta| < 2.4$, $p_{T1} > 22$ GeV, $p_{T2} > 10$ GeV
- HLT trigger: HLT IsoMu20, HLT IsoTkMu20

Drell-Yan
PYTHIA 8
MADGRAPH5 aMC@NLO

ttbar, tW, tbarW
PYTHIA 8
POWHEG v2

WW, ZZ, WZ
PYTHIA 8

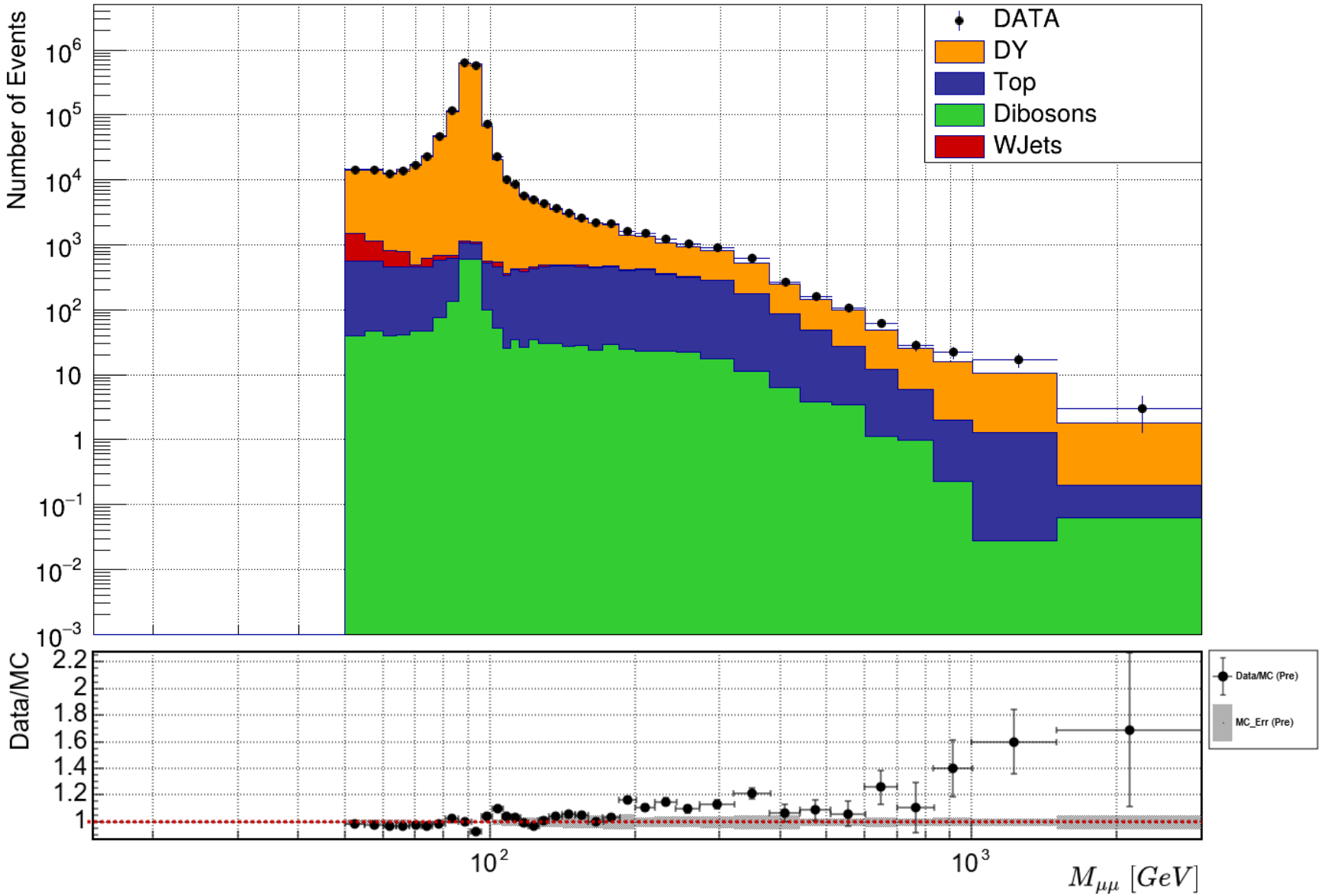
W+jets
PYTHIA 8
MADGRAPH5 aMC@NLO

Analysis of 1 data set with 1 TB ~ 8 hours

Dimuons invariant mass (Open Data)

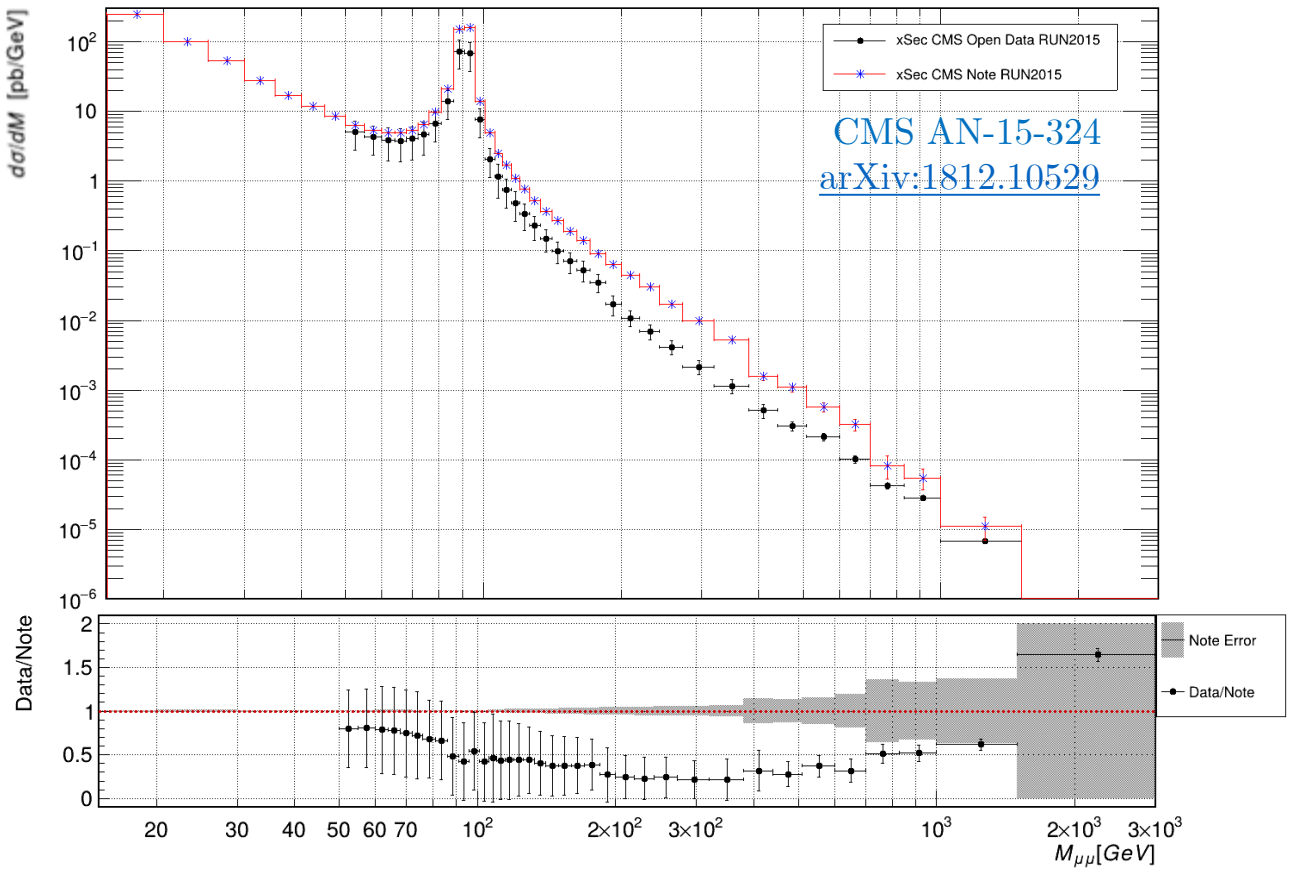
$$k = \frac{\sigma_{MC} L_{ing}}{N_{MC}}$$

MC normalization coefficient

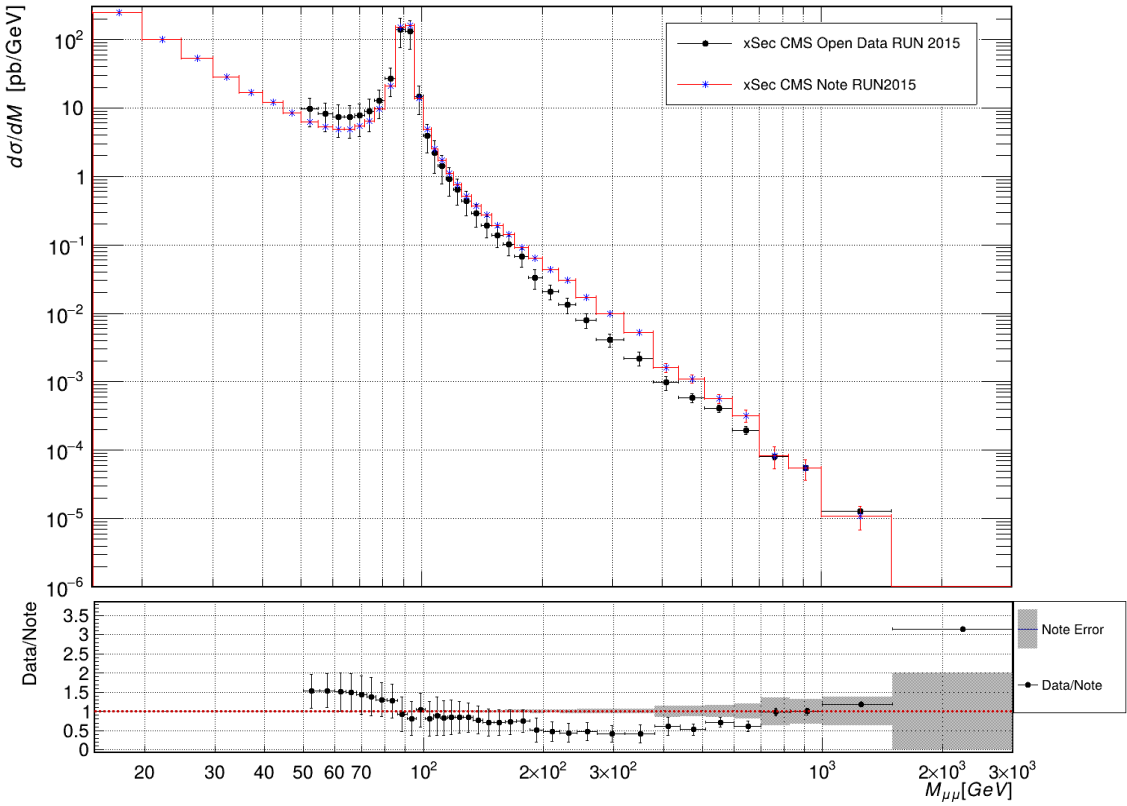


Cross section

Before normalization on σ_Z^{Note}
CMS Open Data RUN2015 ~2.6 fb^{-1}



After normalization on σ_Z^{Note}
CMS Open Data RUN2015 ~2.6 fb^{-1}



Data sets RUN2022 and selections

Data

/Muon/Run2022C-18Sep2023-v3/MINIAOD
/Muon/Run2022D-22Sep2023-v1/MINIAOD ~8 fb^{-1}
/Muon/Run2022E-22Sep2023-v1/MINIAOD
/Muon/Run2022F-22Sep2023-v2/MINIAOD ~27 fb^{-1}
/Muon/Run2022G-22Sep2023-v1/MINIAOD

MC Signal Background

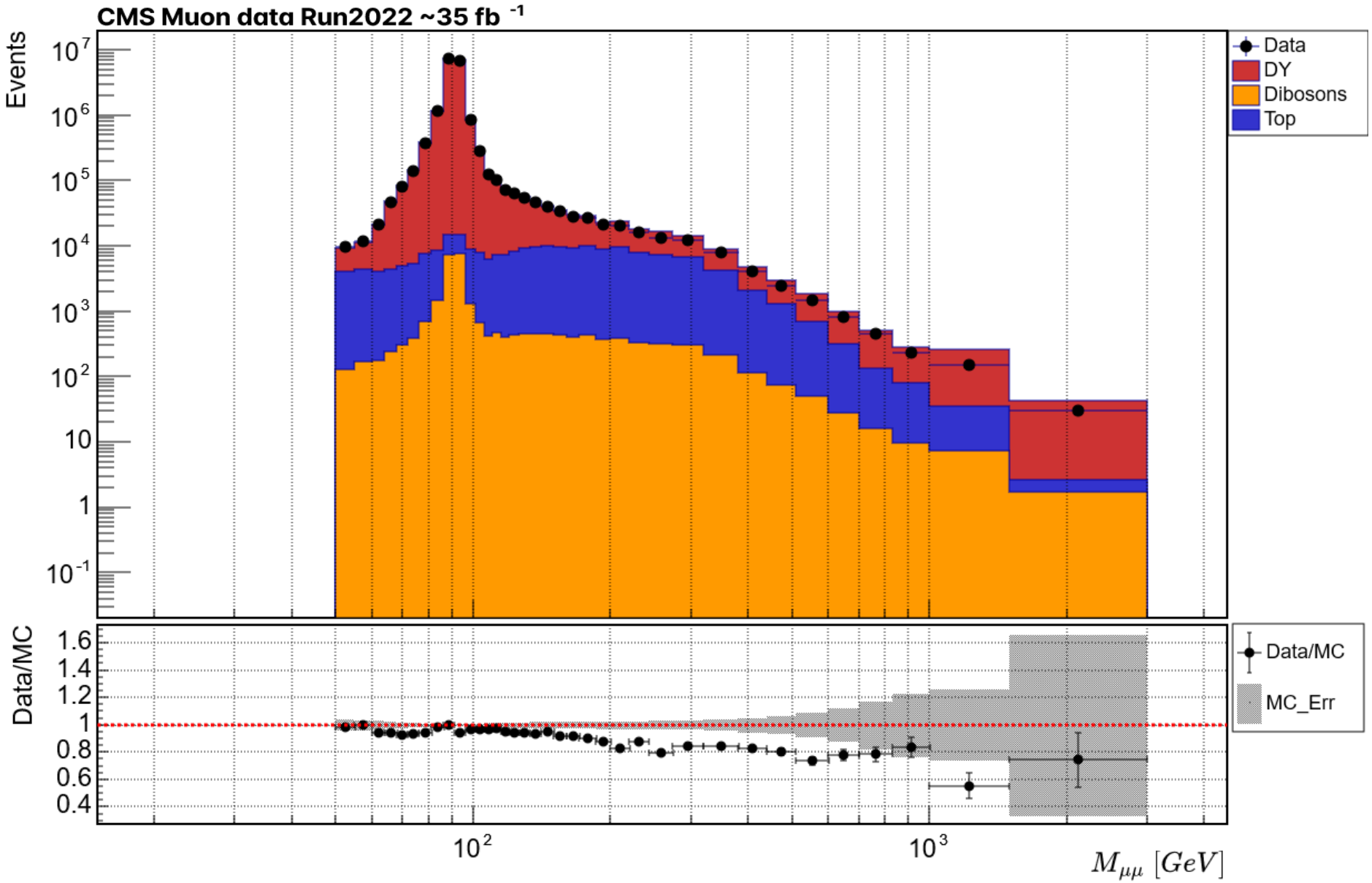
/DYToLL-M50_pre/postEE pythia8
/TT_TuneCP5_pre/postEE pythia8
/TWminusto2L2Nu_pre/postEE pythia8
/TbarWplusto2L2Nu_pre/postEE pythia8
/WW_TuneCP5_pre/postEE pythia8
/WZ_TuneCP5_pre/postEE pythia8
/ZZ_TuneCP5_pre/postEE pythia8

Name	Selection cut
Kinematic	$p_{T1} > 30\ GeV/c, \quad \lvert\eta\rvert < 2.4$ $p_{T2} > 30\ GeV/c, \quad \lvert\eta\rvert < 2.4$
Muon identification	isTightMuon()
Impact parameter	$\lvert d_{xy}\rvert < 2\ mm$
Isolation	$\sum_i \left(Iso_{(\text{R04ChargeHadronPt})} + Iso_{(\text{R04NeutralHadronPt})}\right) / p_i^T(\mu) < 0.2$
Trigger	HLT_IsoMu27_v

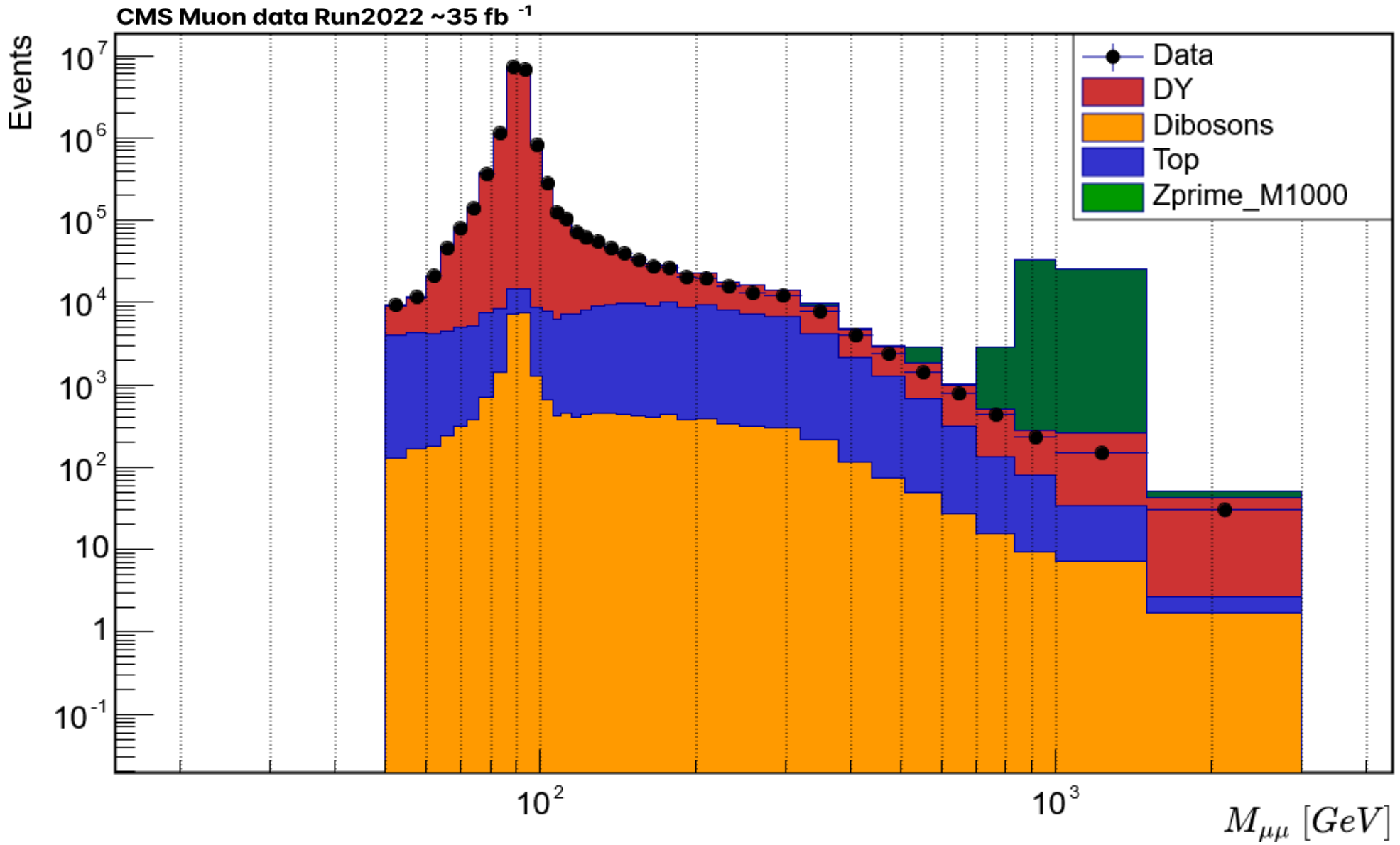
HTCondor High Throughput Computing

Analysis of 1 data set with 24 TB (~500 jobs): ~ 4 hours

Dimuons invariant mass (RUN2022)



Example of first simulation



PYTHIA 8

Summary

Open Data RUN3

•The results of the analysis of CMS Open Data from RUN2015 with an integrated luminosity of ~2.6 fb^{-1} are presented and a center-of-mass energy of 13 TeV.

•The differential cross-section of the Drell-Yan process as a function of the invariant mass of the muon pair has been calculated.

•The results are shown to be in agreement with the predictions of the Standard Model.

•A comparison of the differential cross-section distribution of the Drell-Yan process with the results published by the CMS collaboration is demonstrated.

•A Drell-Yan analysis was performed for the full RUN2022 dataset containing muons, with an integrated luminosity of ~35 pb^{-1}.

•A pair of muons was selected using isolation criteria and muon identification.

•The invariant mass of the muon pair and their kinematic parameters were calculated.

•The obtained results were compared with the predictions of the Standard Model.

These results are planned to be used in the future analysis of simplified dark matter model.

The work plan for 2025. Participation in events 2020-2024

•XV International School-Conference "Current Problems in Micro-World Physics," August 27 – September 3, 2023, Minsk, Belarus, “Studying muon pair production based on open data from CMS,” session talk.

•XXIII International Baikal Summer School on Elementary Particle Physics and Astrophysics, July 11–18, 2023, Bolshie Koty, Russia, “Studying muon pair production based on open data from CMS,” session talk.

•High-Energy Physics Conference, September 11-14, 2023, Yerevan, Armenia, participant.

•59th meeting of the PAC for Nuclear Physics, January 22, 2024, poster presentation.

•Scientific Session of the Nuclear Physics Section of the JINR, April 1–5, 2024, speaker.
•2024 ASIA EUROPE PACIFIC SCHOOL OF HIGH-ENERGY PHYSICS, June 12–25, 2024, attendee.

•Mathematical Modeling and Computational Physics, October 20–25, 2024, speaker.

•Multiple presentations at CMS collaboration working meetings (CERN, Geneva) at JINR.

•Study of Generator Performance. Modeling event sets for signal processes.

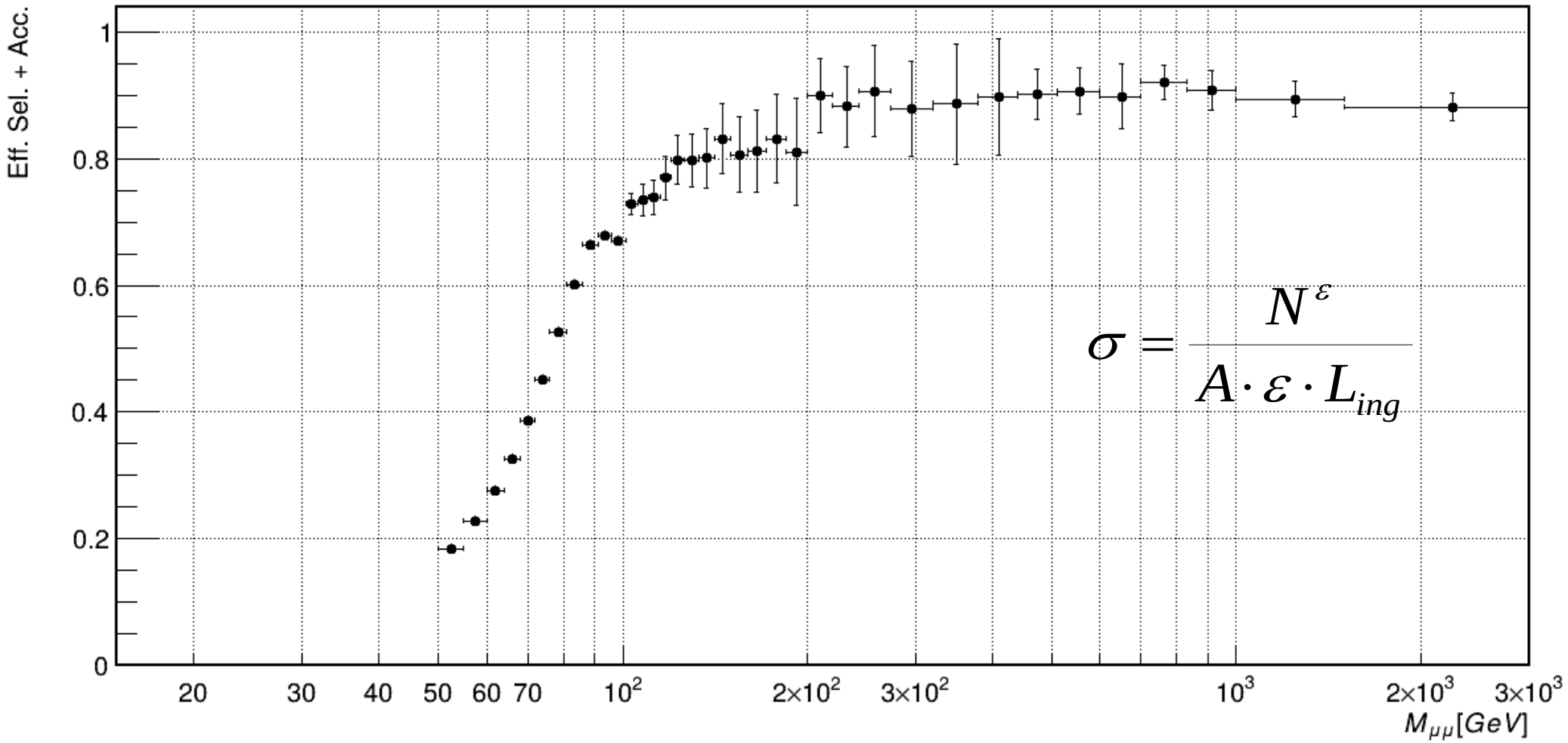
•Include Drell-Yan process corrections in the analysis. Do this for both open data and RUN3.

•Perform the calculation of systematic uncertainties.

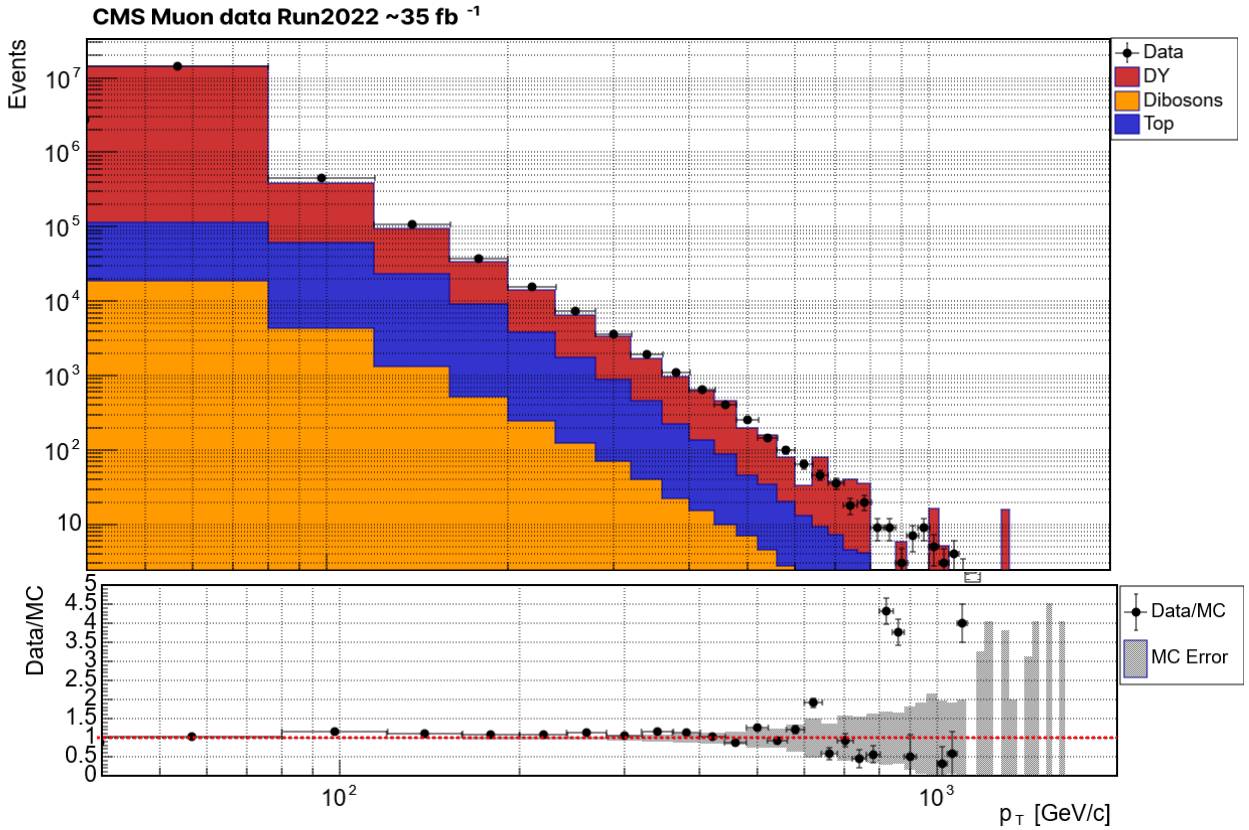
Thank you for your attention!

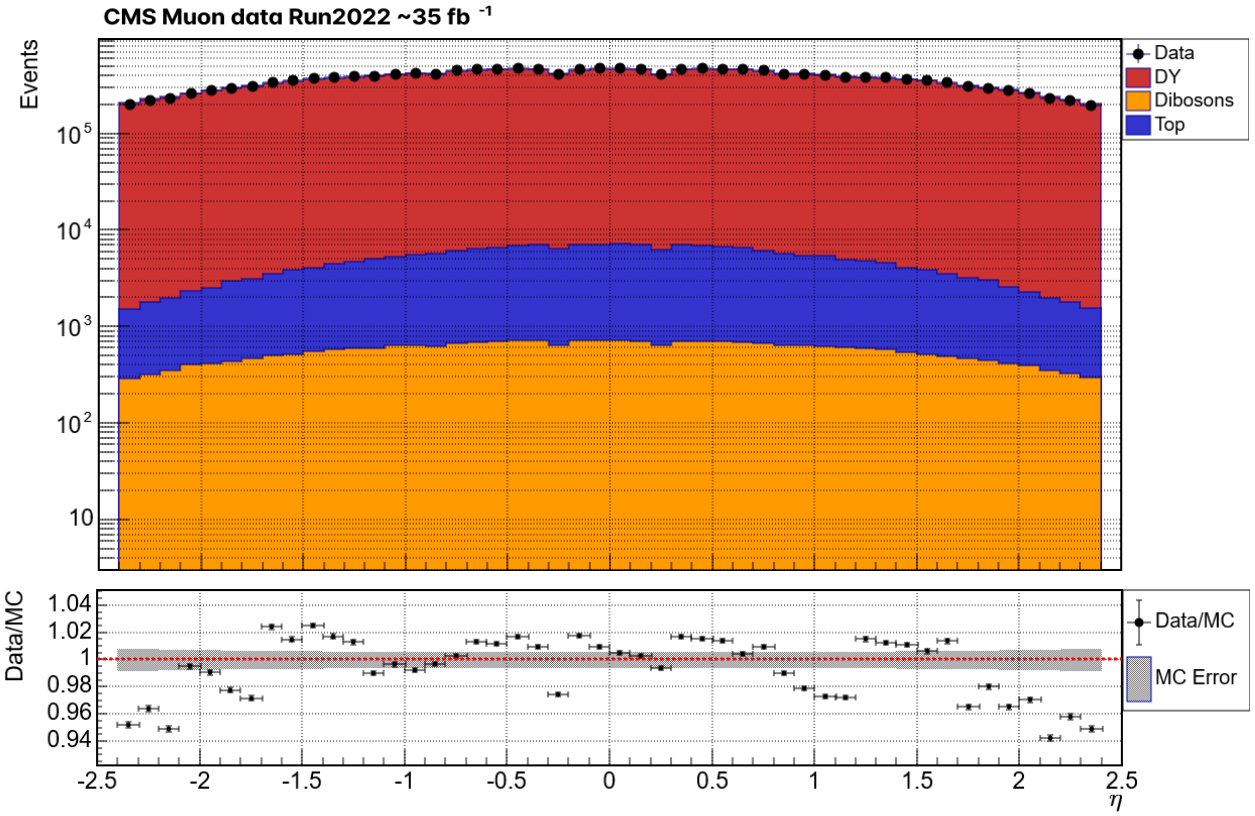
Efficiency

$$E(Full\ Eff.) = A(Eff.Acc.)\cdot\varepsilon, \quad \varepsilon(Eff.selections) = \varepsilon_{trigger}\cdot\varepsilon_{identfication}\cdot\varepsilon_{iso}\ldots$$

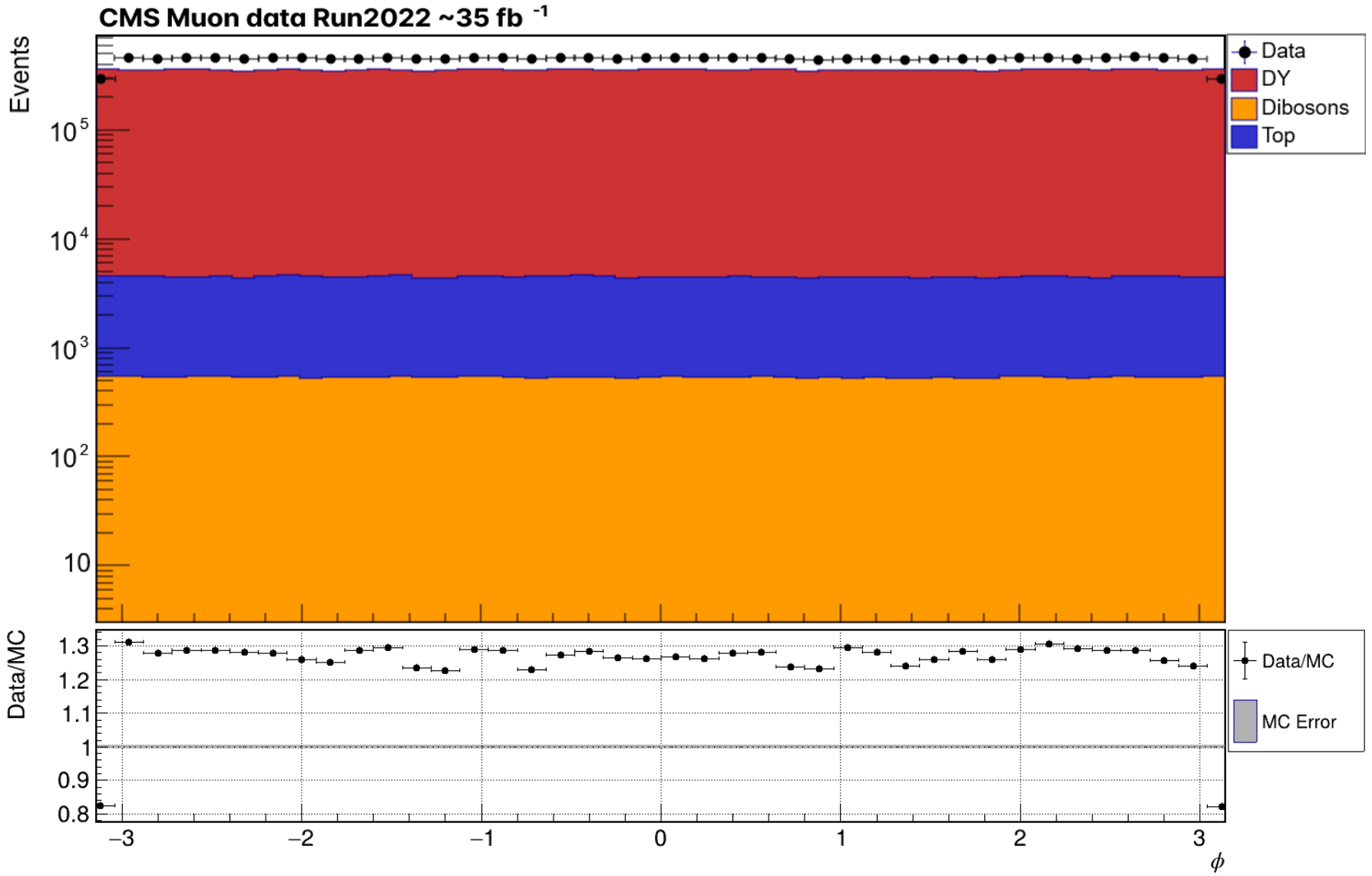


Kinematic parameters(RUN2022)





Kinematic parameters(2)



Run2022

Run2022B

nfill	nrun	nls	ncms	totdelivered(/fb)	totrecorded(/fb)
8	19	4639	4638	0.101540816	0.096555539

hltpath	nfill	nrun	ncms	totdelivered(/fb)	totrecorded(/fb)
HLT_IsoMu27_v16	4	9	3381	0.093277632	0.089078990

Run2022C

nfill	nrun	nls	ncms	totdelivered(/fb)	totrecorded(/fb)
39	106	25897	25894	5.248152376	5.010409016

hltpath	nfill	nrun	ncms	totdelivered(/fb)	totrecorded(/fb)
HLT_IsoMu27_v17	39	106	25894	5.247673905	5.010409016

Run2022D

nfill	nrun	nls	ncms	totdelivered(/fb)	totrecorded(/fb)
10	45	10920	10919	3.140588880	2.970045129

hltpath	nfill	nrun	ncms	totdelivered(/fb)	totrecorded(/fb)
HLT_IsoMu27_v17	10	45	10919	3.140353182	2.970045129

Run2022E

nfill	nrun	nls	ncms	totdelivered(/fb)	totrecorded(/fb)
23	51	21698	21698	6.046849293	5.806955207

hltpath	nfill	nrun	ncms	totdelivered(/fb)	totrecorded(/fb)
HLT_IsoMu27_v18	23	51	21698	6.046849293	5.806955207

Run2022F

nfill	nrun	nls	ncms	totdelivered(/fb)	totrecorded(/fb)
47	118	51589	51589	18.696392501	17.781901464

hltpath	nfill	nrun	ncms	totdelivered(/fb)	totrecorded(/fb)
HLT_IsoMu27_v18	47	118	51589	18.696392501	17.781901464

Run2022G

nfill	nrun	nls	ncms	totdelivered(/fb)	totrecorded(/fb)
7	21	8242	8242	3.241980255	3.082753036

hltpath	nfill	nrun	ncms	totdelivered(/fb)	totrecorded(/fb)
HLT_IsoMu27_v18	7	21	8242	3.241980255	3.082753036