NanoAODRun1

Open Data Workshop, 02.08.2022

<u>Achim Geiser</u>, DESY Hamburg, for the CMS/DPOA NanoAODRun1 subgroup

Direct contributors to ntuple code, validation, and production (within CMS):
Afiq Anuar, Hamed Bakshiansohi, Armando Bermudez, Edgar Carrera, A.G., Nur Zulaiha Jomhari, Hannes Jung, Clemens Lange, Dietrich Liko, Josry Metwally, Qun Wang, Stefan Wunsch, Heng Yang, Yewon Yang

Direct contributors to usage examples and their documentation (including external contributors):
Matthew Bellis, A.G., Florian Lorkowski, Ihor Melnyk, Leonardo Olivi, Jinal Shah†, Fabian Staeger, Lennox Wood, Stefan Wunsch, Yewon Yang

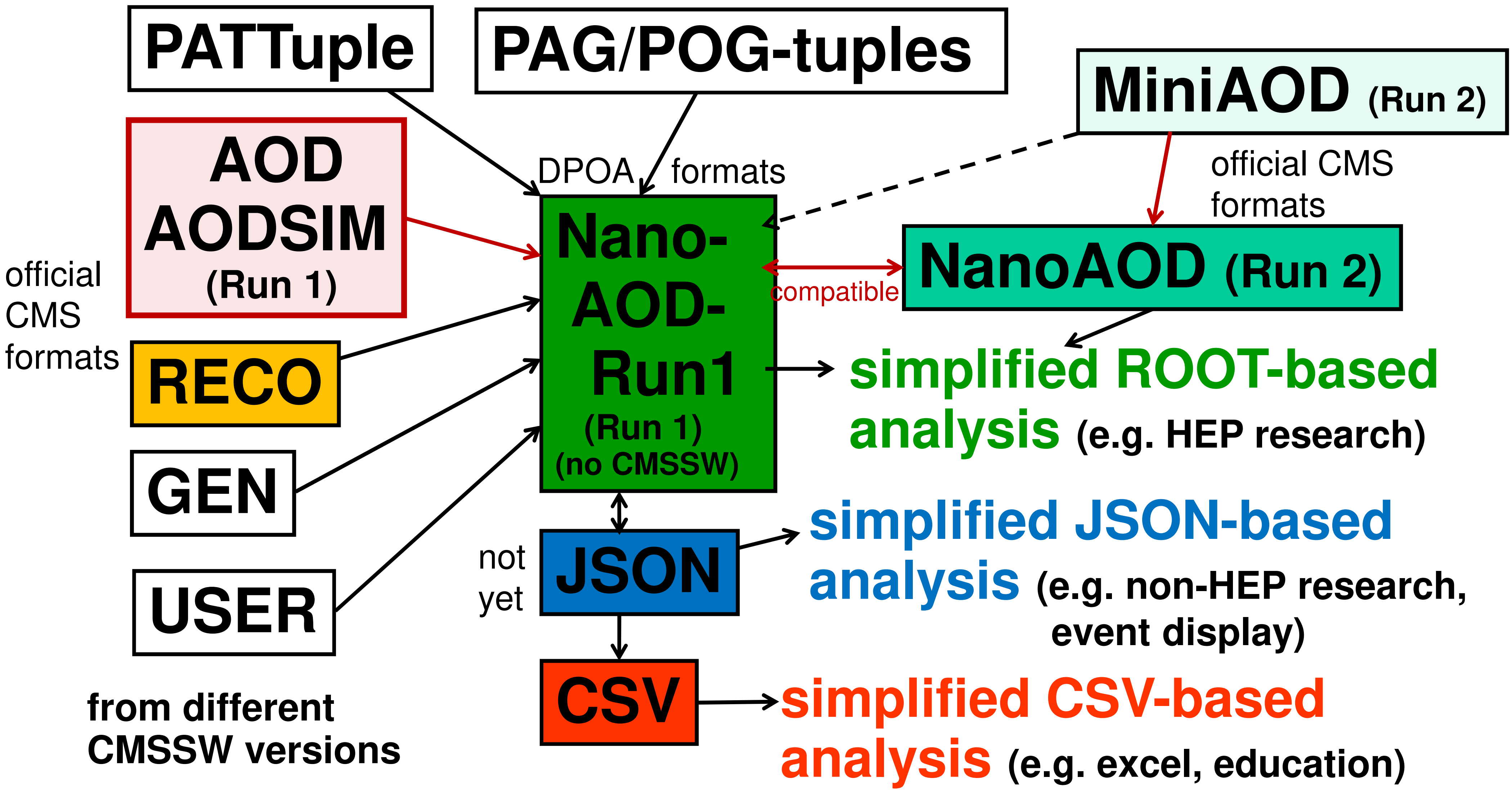
- What is NanoAODRun1?
- Documentation and explanations
- Usage examples for the CMS Open Data workshop, partially `hands on'
  (focusing on handling of (di)muons, triggers, and datasets; examples also using other objects in preparation)

Listeners might want to open the slides also from the workshop indico in order to click or copy/paste some of the links given in this contribution.

See also instructions on the workshop tutorial on github.

Vision for simplified DPOA data format (CMS data)
partially implemented

Design common flat ntuple format for all datasets (remove CMSSW dependence)



What is NanoAODRun1?

- Derived data format provided by CMS DPOA for Run 1 Open Data, retaining significant research analysis capability. Single format/content for all years/datasets/analyses.
- Predefined flat Root ntuple, compatible with recent 'default' Root versions (or other programs with interface to Root ntuples), bypassing the use of AOD, VMs, containers
- Suitable for ~50% of research level physics analyses. Others should continue to use AOD.
- Data volume and processing time reduced by orders of magnitude w.r.t. AOD.
- Conceptually backwards compatible to most already existing AOD-based Open Data examples (i.e., the results can be reproduced, often exactly) -> example provided.
- Physically backwards compatible to many already existing educational Open Data examples based on "reduced" educational nanoAOD versions -> example provided.
- At least partially forward compatible to future official CMS Run2 nanoAOD releases.
- Many of the already publicly available official CMS NanoAODtools can be applied.
- Publishable results can be produced (or reproduced), as with CMS Run2 NanoAOD -> example provided for reproduction of published CMS distribution.

The physics objects and their use

- The physics objects in NanoAODRun1 are conceptually the same as those encountered in the previous POET AOD (miniAOD) exercises
  https://github.com/cms-opendata-analyses/PhysObjectExtractorTool
  -> no explicit (re)explanations here

- The physics is the same for Run 1 and Run2.
  Only the center-of mass and some detector and software aspects differ.

- Everything you learned previously about physics objects remains valid.
  Only the variable naming scheme and the computing environments differ.

- Why would you want to analyze Run 1 data when Run 2 data are available?
  Run 2 (Open) data have higher energy.
  Run 1 (Open) data have higher luminosity/more statistics
  (until more Run 2 data will have been publicly released)

- Why would you want to use NanoAODRun1 rather than AOD?
  easier handling, no CMSSW, no need for containers (you can still use them, of course),
  but less control over low level features of the objects.
  E.g. vertices and leptons, photons, jets, … but no longer all tracks, PF objects, …

The code/implementation and documentation

- NanoAODRun1 ntuple producing code available on (still internal) CERN gitlab
  https://gitlab.cern.ch/cms-opendata/cms-opendata-nanoaodplus/nanoaodplus_v1
- runs on all public AOD and AODSIM data sets, in particular Run 1
- not intended to be run by users, but available for inspection
  -> will be exported to public git soon

- users should directly use public NanoAODRun1 ntuples produced by CMS/DPOA
- ntuple content validated within DPOA, feedback on `overlooked features' welcome.
- initially being made available on selected subset of 2010, 2011 and 2012 pp samples (mostly data, some MC), as needed for the Open Data workshop.
- More samples will be added on best effort basis as needed/requested by users

- **NanoAODRun1 ntuple format documentation** (to be further improved):
- **WorkBook**
  https://twiki.cern.ch/twiki/bin/view/CMSPublic/WorkBookNanoAODRun1

NanoAODRun1 WorkBook

TWiki > CMS Web > DataPreservationOpenAccess > DPOANanoAODlike > DPOAWorkBookNanoAODplus > DPOAWorkBookNanoAODplustemp (2022-04-14, AchimGeiser)

This is the workbook documentation for NanoAODRun1, modeled after WorkBookNanoAOD

The DPOA NanoAODRun1 data tier

The NanoAODRun1 format is a NanoAOD-like research-level Ntuple format for CMS Run 1 data, readable with bare ROOT and containing the per-event information that is needed in most generic analyses. Ir nanoAOD sprit, of order 50% of all Run 1 physics analyses should be able to use this format (see e.g. here), with a gain of 1-2 orders of magnitude in terms of CPU and disk usage. In contrast to CMS Nano is derived from MiniAOD and supported by XPOG, it is generated directly from AOD with completely independent code, and is supported by the CMS DPOA legacy data sets subgroup. This is a much smaller (to NanoAODv8/9, the content of NanoAODRun1 is not endorsed by CMS as a whole. Nevertheless, as documented below, there is large overlap in functionality and content between NanoAODRun1 and Nano common analyses are possible. Since no intermediate data tier exists, NanoAODRun1 also contains extra variables taken directly from AOD. These variables are also called 'plus' extensions. User feedback fc NanoAODRun1 improvements is highly welcome.

Note that there is also a NanoAOD-like educational reduced NanoAOD format, which is sometimes also plainly referred to as "NanoAOD" in the Open Data context, and has a partially compatible but much m aimed at specific educational/pedagogical exercises rather than capability for general full physics analysis.

NanoAODRun1 ntuples for relevant Open Data and MC sets are being/will be produced and made available by the DPOA group, so normal users should not need to deal with the ntuple production code exce purposes. Nevertheless, the single set of code for the generation of NanoAODRun1, suited for all AOD and RECO Open Data sets (with auto-recognition of different compilation options for different CMSSW v in the so far CMS-internal repository https://gitlab.cern.ch/cms-opendata/cms-opendata-nanoaodplus/nanoaodplus_v1 and can be used by experts (the general CMS software guide may help). Nonexpert: much more pedagogical POET code (for CMSSW_5_3_32 only).

The almost final NanoAODRun1 ntuple version documented here is NanoAODplus_v0.9b. It will be updated to the essentially unchanged NanoAODRun1_v1 for the public release.

https://twiki.cern.ch/twiki/bin/view/CMSPublic/WorkBookNanoAODRun1

modeled after "official" WorkBookNanoAOD
https://twiki.cern.ch/twiki/bin/view/CMSPublic/WorkBookNanoAOD
(for CMS data from 2016 onwards, not yet publicly released)

Automatic documentation example: Muon

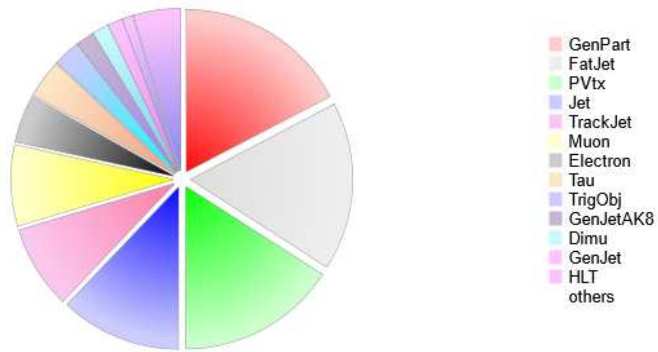
https://twiki.cern.ch/twiki/pub/CMSPublic/WorkBookNanoAODRun1/doc_DYJetsToLL_M-50_7TeV.html#Muon

Muon [back to top]

Object property	Type	Description
Muon_Chi2	Float_t	tracker muon chi2/ndof
Muon_Id	Int_t	Unique identifier of muon, for list-independent cross-referencing
Muon_charge	Int_t	electric charge
Muon_dxy	Float_t	dxy (with sign) wrt first PV, in cm
Muon_dxyBest	Float_t	dxy (with sign) wrt best PV, in cm
Muon_dxyErr	Float_t	dxy uncertainty, in cm
Muon_dxybs	Float_t	dxy (with sign) wrt beam spot, in cm
Muon_dz	Float_t	dz (with sign) wrt first PV, in cm
Muon_dzBest	Float_t	dz (with sign) wrt best PV, in cm
Muon_dzErr	Float_t	dz uncertainty, in cm
Muon_eta	Float_t	eta
Muon_gChi2	Float_t	global muon chi2/ndof
Muon_genPartFlav	UChar_t	Flavour of genParticle for MC matching to status==1 muons: 1 = prompt muon (including gamma*->mu mu), 15 = muon from prompt tau, 5 = muon from b, 4 = muon from c, 3 = muon from light or unknown, 0 = unmatched
Muon_genPartIdx	Int_t(index to Genpart)	GenPart index of the associated gen object (-1 if none)
Muon_geta	Float_t	global muon eta
Muon_gnPix	Int_t	number of valid pixel hits, global muon
Muon_gnValid	Int_t	number of valid tracker hits, global muon
Muon_gnValidMu	Int_t	Muon_gnValidMu
Muon_gphi	Float_t	global muon phi
Muon_gpt	Float_t	global muon pt
Muon_highPtId	UChar_t	high-pT cut-based ID (1 = tracker high pT, 2 = global high pT, which includes tracker high pT)
Muon_highPurity	Bool_t	satisfies high purity criteria
Muon_ip3d	Float_t	3D impact parameter wrt first PV, in cm
Muon_ip3dBest	Float_t	3D impact parameter wrt best PV, in cm
Muon_isArbitrated	Bool_t	isGood(TrackerMuonArbitrated)
Muon_isGlobal	Bool_t	muon is global muon
Muon_isGood	Bool_t	isGood(TMAnyStationTight)
Muon_isGoodAng	Bool_t	isGood(TMLastStationAngTight)
Muon_isGoodLast	Bool_t	isGood(TMLastStationTight)
Muon_isMini	Bool_t	can be found on MiniAOD
Muon_isNano	Bool_t	can be found on NanoAOD
Muon_isPFcand	Bool_t	muon is PF candidate
Muon_isRPCcand	Bool_t	RPC Muon
Muon_isStandAlone	Bool_t	StandAlone Muon
Muon_isTracker	Bool_t	muon is tracker muon

Automatic documentation (NanoAODTools)

testnanoall94_DY.root (141.829 Mb, 50000 events, 2.90 kb/event)



example:
2012 DYtoLL MC

instructions how to generate will be added after the workshop

Event data

collection	kind	vars	items/evt	kb/evt	b/item	plot	%	cumulative %	
GenPart	collection	24	20.28	0.503	25.4		17.3%	17.3%	100.0%
FatJet	collection	13	17.30	0.487	28.8		16.8%	34.1%	82.7%
PVtx	collection	19	15.26	0.462	31.0		15.9%	50.0%	65.9%
Jet	collection	16	10.23	0.350	35.0		12.0%	62.1%	50.0%
TrackJet	collection	9	15.00	0.240	16.4		8.3%	70.4%	37.9%
Muon	collection	70	1.35	0.234	177.3		8.1%	78.4%	29.6%
Electron	collection	49	0.76	0.137	184.4		4.7%	83.1%	21.6%
Tau	collection	17	3.96	0.110	28.4		3.8%	86.9%	16.9%
TrigObj	collection	6	7.83	0.077	10.0		2.6%	89.5%	13.1%
GenJetAK8	collection	5	3.03	0.051	17.1		1.7%	91.3%	10.5%
Dimu	collection	29	0.34	0.047	143.3		1.6%	92.9%	8.7%
GenJet	collection	5	2.39	0.041	17.7		1.4%	94.3%	7.1%
HLT	singleton	421	1.00	0.029	30.0		1.0%	95.3%	5.7%
Photon	collection	10	0.80	0.025	32.4		0.9%	96.2%	4.7%
MET	singleton	7	1.00	0.025	25.3		0.9%	97.1%	3.8%
PV	singleton	8	1.00	0.022	22.3		0.8%	97.8%	2.9%
L1	singleton	127	1.00	0.018	18.3		0.6%	98.4%	2.2%
OtherPV	vector	2	3.00	0.012	4.2		0.4%	98.9%	1.6%
GenPV	singleton	5	1.00	0.011	11.7		0.4%	99.2%	1.1%
CaloMET	singleton	3	1.00	0.011	10.9		0.4%	99.6%	0.8%
Trig	singleton	25	1.00	0.005	5.2		0.2%	99.8%	0.4%
event	variable	1	1.00	0.002	2.3		0.1%	99.9%	0.2%

First 22 Open Data papers from CMS data as proxy for what you might want to do with our data

should already be reproducible with current NanoAODRun1

should be feasible with future version ("full" NanoAOD compatibility)

continues to require AOD and/or its metadata (personal judgement)

e.g. tracks or full PF object set

CMS Open Data related research papers

publication status early 2021

Compare also the generic lists of documents citing Open Data and CMS on new inSpire and old inSpire.

- Jet Substructure Studies with CMS Open Data, A. Tripathee et al., Apr 19, 2017, MIT-CTP-4890, arXiv:1704.05842, Phys. Rev. D96 (2017) 074003; 2010 data, 33 citations.
- Exposing the QCD Splitting Function with CMS Open Data, A. Larkoski et al., Apr 17, 2017, MIT-CTP-4891, arXiv:1704.05066, Phys. Rev. Lett. 119 (2017) 132003; 2010 data, 34 citations. Also see corresponding news and blog.
- related preliminary results were presented at the FCC week 2015, Boost2015, Moriond 2016, and Boost2017.
- Application of a Convolutional Neural Network for image classification to the analysis of collisions in High Energy Physics, C.F. Madrazo et al., Inspire, arXiv:1708.07034, EPJ Web Conf. 214 (2019) 06017; 2011 data, 7 citations.
- Towards automation of data quality system for CERN CMS experiment, Yandex collaboration, M Borisyak et al, arXiv:1709.08607, J. Phys. Conf. Ser. 898 (2017) 092041; 2010 data, 7 citations.
- Expressing Parallelism with ROOT, D. Piparo et al, J.Phys.Conf.Ser. 898 (2017) no.7, 072022, 2011 MC, 2 citations.
- Fast and Accurate Simulation of Particle Detectors Using Generative Adversarial Networks, P. Musella and F. Pandolfi, arXiv:1805.00850, Comput. Softw. Big Sci. 2 (2018) 8; 2011 MC, 31 citations.
- End-to-End Physics Event Classification with the CMS Open Data: Applying Image-based Deep Learning on Detector Data to Directly Classify Collision Events at the LHC, M. Andrews et al., arXiv:1807.11916, EPJ Web Conf. 214 (2019) 06031, published in Comput. Softw. Big Sci. 4 (2020) 1, 6; 2012 MC, 17 citations.
- Searching in CMS Open Data for Dimuon Resonances with Substantial Transverse Momentum, C. Cesarotti et al., arXiv:1902.04222, Phys. Rev. D100 (2019) 015021; 2012 data+MC, 4 citations.
- Intermittency in pseudorapidity space of pp collisions at = 7 TeV, Z. Zong et al., EPJ Web Conf. 206 (2019) 09004, 2011 data, no citations yet.
- End-to-End Jet Classification of Quarks and Gluons with the CMS Open Data, M. Andrews et al., arXiv:1902.08276, Nucl.Instrum.Meth. A977 (2020) 164304; 2012 MC, 7 citations; also arXiv:1910.07029.
- Testing Non-Standard Sources of Parity Violation in Jets at √s=8 TeV with CMS Open Data, C.G. Lester and M. Schott, arXiv:1904.11195, JHEP 12 (2019) 120; 2012 data+MC, 5 citations.
- Processing Columnar Collider Data with GPU-Accelerated Kernels, J. Pata and M. Spiropulu, arXiv:1906.06242, 1 citation.
- On the model dependence of fiducial cross-section measurements, G. Facini et al., arXiv:1906.01278, Mod. Phys. Lett. A34 (2019) no.38, 2050065; CMS MC, 1 citation.
- Opportunities and Challenges of Standard Model Production Cross Section Measurements at 8 TeV using CMS Open Data, A. Apyan et al., arXiv:1907.08197, JINST 15 (2020) 01, P01009; 2012 data+MC, 3 citations.
- Explicit Jet Veto as a Tool to Purify the Underlying Event in the Drell-Yan Process Using CMS Open Data, S.P. Mehdiabadi and A. Fahim, arXiv:1907.08842, J. Phys. G 46 (2019) 9, 095003; 2012, data+MC, 1 citation.
- Exploring the Space of Jets with CMS Open Data, P.T. Komiske et al, arXiv:1908.08542, Phys. Rev. D 101 (2020) 3, 034009; 2012 data+MC, 9 citations.
- Interaction Networks for the Identification of Boosted $H\to b\bar{b}$ Decays, E.A. Moreno et al., Sep 26, 2019, arXiv:1909.12285, Phys.Rev. D102 (2020) 012010; 2016 MC,16 citations.
- Study of Di-muon Production Process in pp Collision in CMS Data from Symmetry Scaling Perspective, S. Bhaduri, A. Bhaduri and D. Ghosh, arXiv:1911.09928, Advances in High Energy Physics (2020), 2011 and 2012 data, 1 citation.
- The Hidden Geometry of Particle Collisions, Patrick T. Komiske, Eric M. Metodiev, Jesse Thaler (MIT). Apr 21, 2020. 56 pp.arXiv:2004.04159, JHEP 2007 (2020) 006; 2012 Jet data, 9 citations.
- Quantum-inspired Machine Learning on high-energy physics data, Marco Trenti et al., arXiv:2004.13747, 1 citation.
- Adversarially Learned Anomaly Detection on CMS Open Data: re-discovering the top quark, Oliver Knapp, Guenther Dissertori, Olmo Cerri, Thong Q. Nguyen, Jean-Roch Vlimant et al. May 4, 2020. arXiv:2005.01598, 2012 single muon data, 7 citations.
- Lorentz and permutation invariants of particles III: constraining non-standard sources of parity violation, Christopher G. Lester, Ward Haddadin, Ben Gripaios (Cambridge U.). Aug 12, 2020. 66 pp, arXiv:2008.05206, 2012 JETHT data, no citations yet.

First usage examples provided at this workshop

https://twiki.cern.ch/twiki/bin/view/CMSPublic/NanoAODRun1Examples

also/alternatively see instructions on workshop github

Plain root ntuples offer a multitude off different ways to work with the data;
examples provided here only provide a first glimpse from a subjective selection.
Will be expanded to more possibilities
Please do not hesitate to provide your own, maybe even still within this workshop!

Interactively exploring the ntuple content with plain ROOT (2011 DY MC dataset)

Use of precompiled C++ code with plain ROOT (2010B Mu dataset)
Reproduction of existing originally AOD-based Open Data validation example

Using ROOT+RDataFrame with C++, python or Jupyter notebook (2012B/C DoubleMuParked datasets)
Modification of existing simplified nanoAOD-based Open Data outreach example

Using ROOT+RDataFrame with C++ (2011A DoubleMuon and MuOnia datasets)
to reproduce published CMS plot using trigger, quality and (re)vertexing information
Example designed from scratch for this workshop (documentation to be improved).

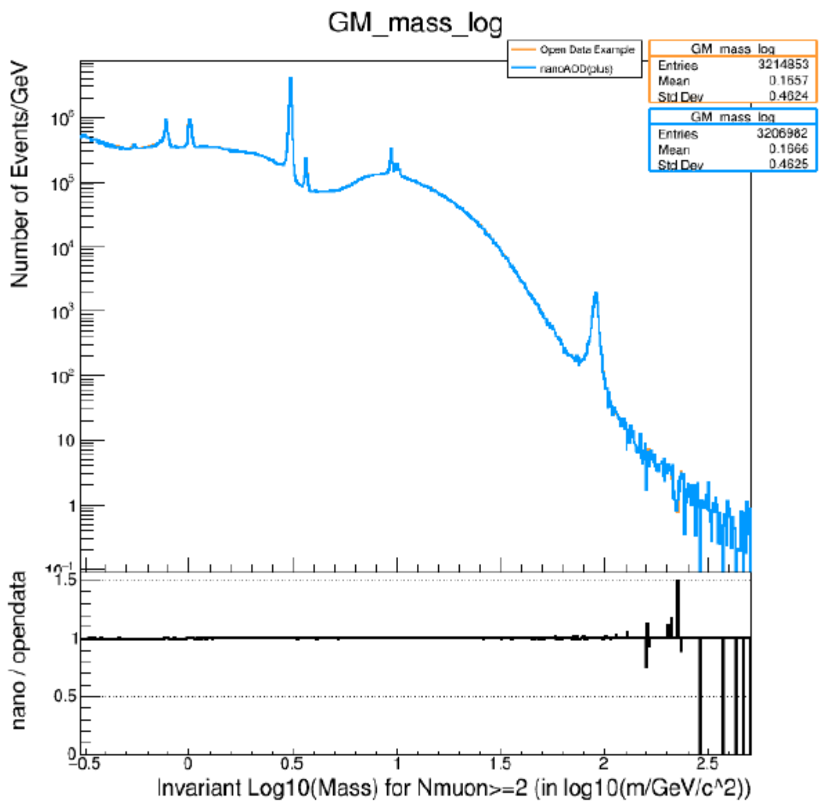
Using plain ROOT with interactive CINT to reproduce published CNS plot
= semi-interactive variant of above

Exercise reproducing existing AOD example from NanoAODRun1

reference: 2010 Mu/Monitor dataset validation example, http://opendata.cern.ch/record/460

https://twiki.cern.ch/twiki/pub/CMSPublic/NanoAODRun1Examples/MuHistos_eospublic.cxx

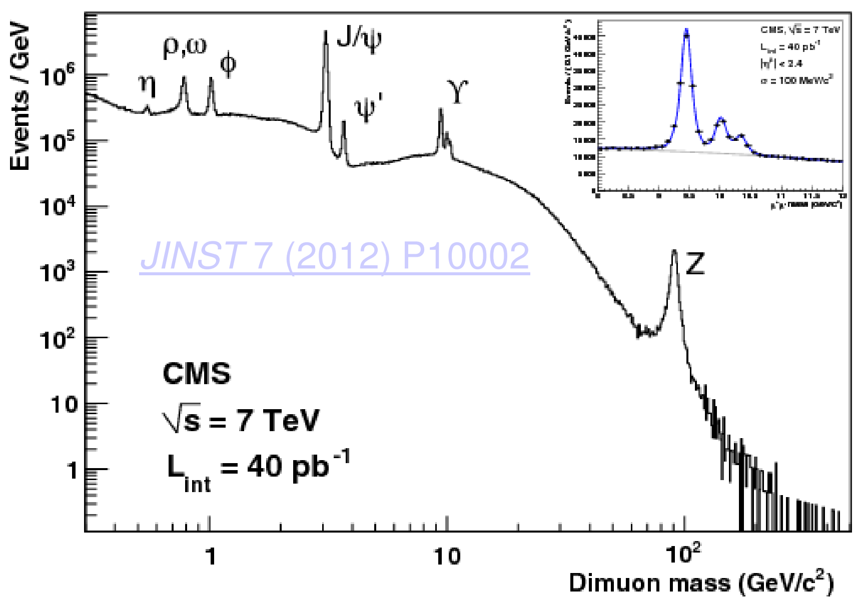
From summer student projects 2019/21



Open Data AOD example

NanoAODRun1 recoded example

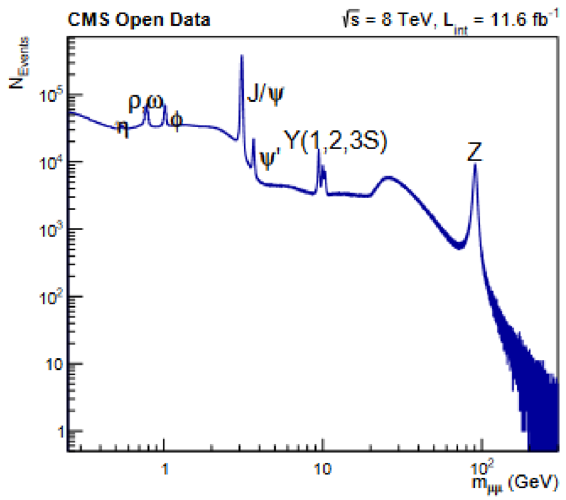
similar (not identical) published distribution:

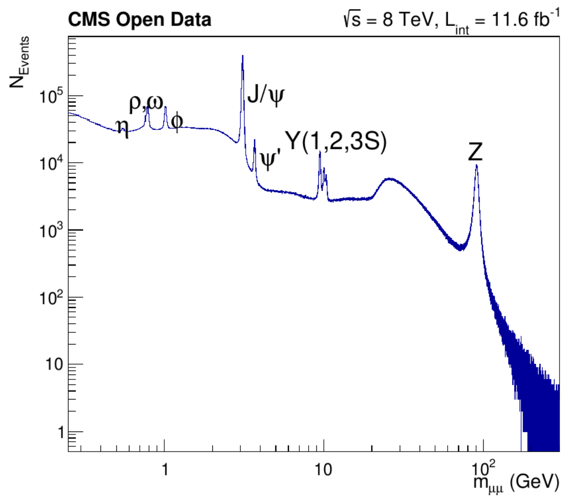


Exercise using existing educational "reduced NanoAOD" example (1 line modification)

8 TeV Dimuon spectrum example, http://opendata.cern.ch/record/12342

Invariant mass of unlike sign muon pairs,
Existing educational reduced nanoAOD (left) vs. NanoAODRun1 (right)





only one small change to existing Open data example code. E.g.
https://twiki.cern.ch/twiki/pub/CMSPublic/NanoAODRun1Examples/dimuonSpectrum2012_eospublic.C

runs on both reduced educational NanoAOD and NanoAODRun1.
several variants provided:
RdataFrame, C++, Python, (Jupyter Notebook)

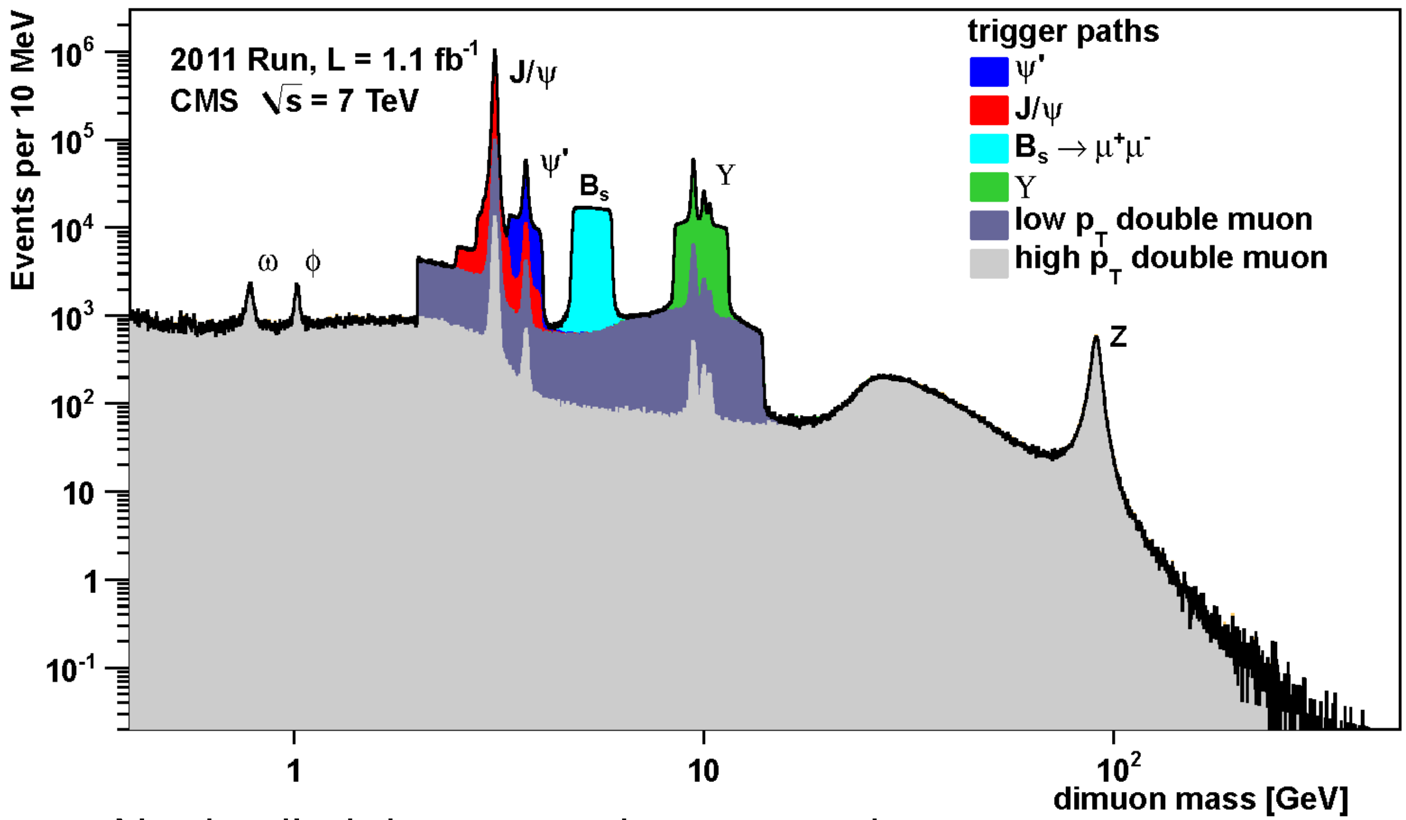
Public plot suited for “fast” reproduction from NanoAODRun1

First 1.1 pb^{-1} of 2011 pp data, including trigger, quality and data set selections

(not possible starting from educational example, suitable for ‘online’ exercise)

CMS detector performance
EPJ Web of Conferences 70 (2014) 00028 (page 6)

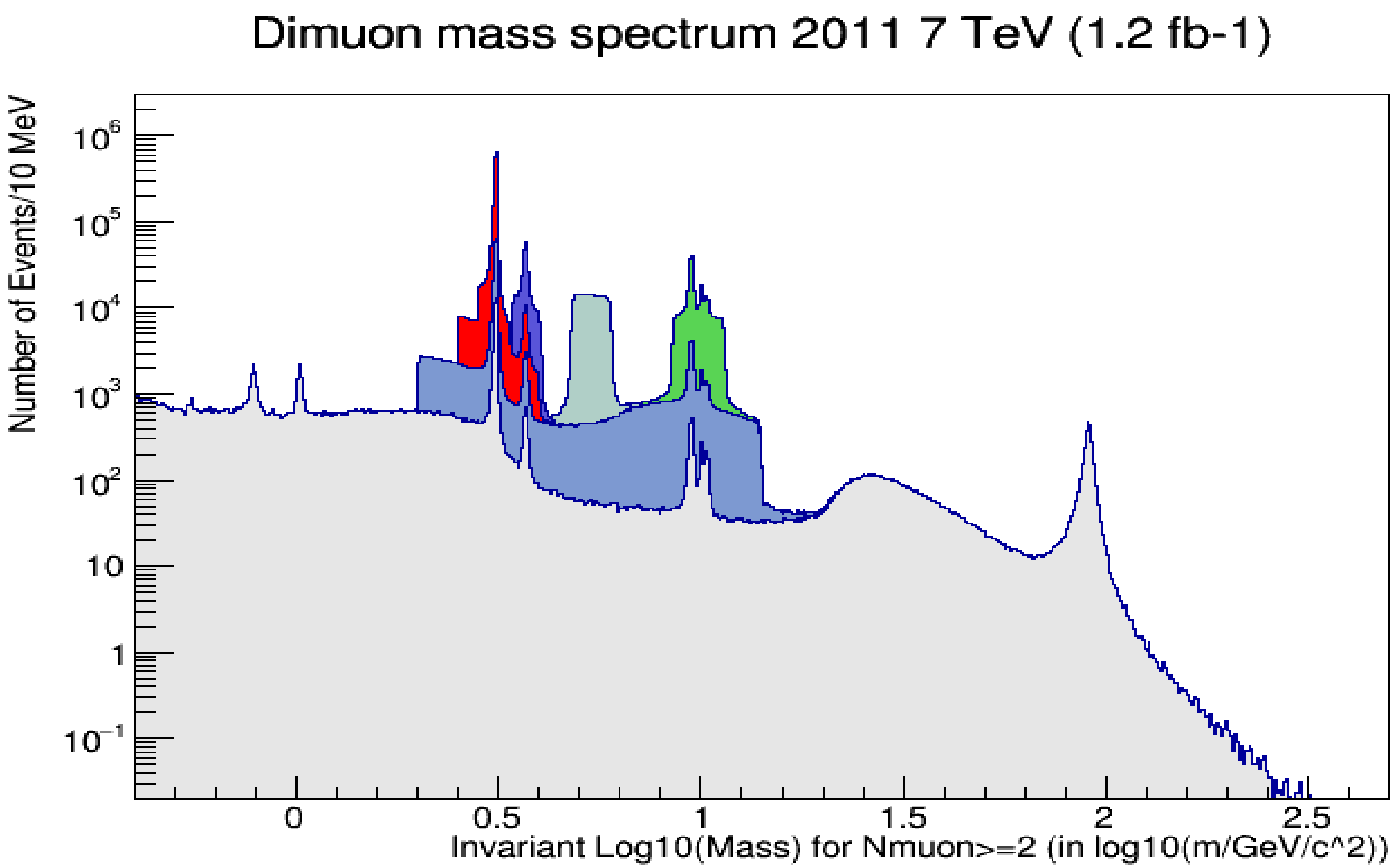
The CMS trigger system, TRG-12-001
JINST 12 (2017) 01, P01020 (page 83)



No detailed documentation, so need to guess cuts

Public plot reproduced from NanoAODRun1

First 1.2 pb^{-1} of 2011 pp data, including trigger, quality, and data set selections
(not possible starting from educational example, suitable for ‘online’ exercise)



Implemented in CINT (interactive C)

Involves cuts on
- datasets: DoubleMuon & MuOnia
- triggers various
- pt cuts various
- some auxiliary mass cuts
- medium quality
- dimuon vertex refit

Example code will be released

Backup

Reproduce public/published 2011 invariant mass spectrum

Some explanations of cuts to reproduce paper content
(very close, but not exact, since original plot was produced from prompt reconstruction/quality selection):

2011 1.1 fb^{-1}, not full dataset: apply Golden JSON quality selection (already done at level of ntuple production) **and restrict run range to early runs** (from ntuple).
recalculate offline luminosity with brilcalc: 1.2 fb^{-1} (was 1.1 fb^{-1} online)

Superposition of data from two different datasets: DoubleMu and MuOnia
-> read both and remove overlap

Superposition of data from selected groups of triggers
-> read events satisfying specific triggers only, depending on dataset

Data cleaned by selection cuts
-> apply quality cuts and kinematic cuts to both muon and dimuon systems, dependent upon trigger selection

Remove pileup background (muons from different vertices)
-> use precomputed/revertexed dimuon candidates from same vertex

Trigger matching possible from ntuple but not yet applied (necessary for validity of trigger efficiencies)
-> will be added in future version of usage example

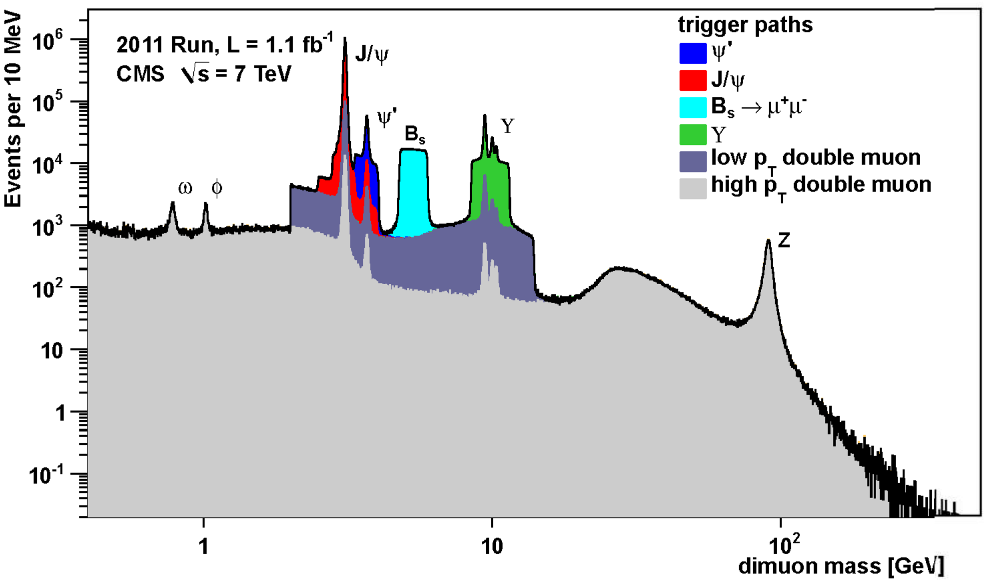
All of this possible with NanoAODRun1 (or AOD), mostly not with educational version(s)

Reproduce public/published 2011 invariant mass spectrum

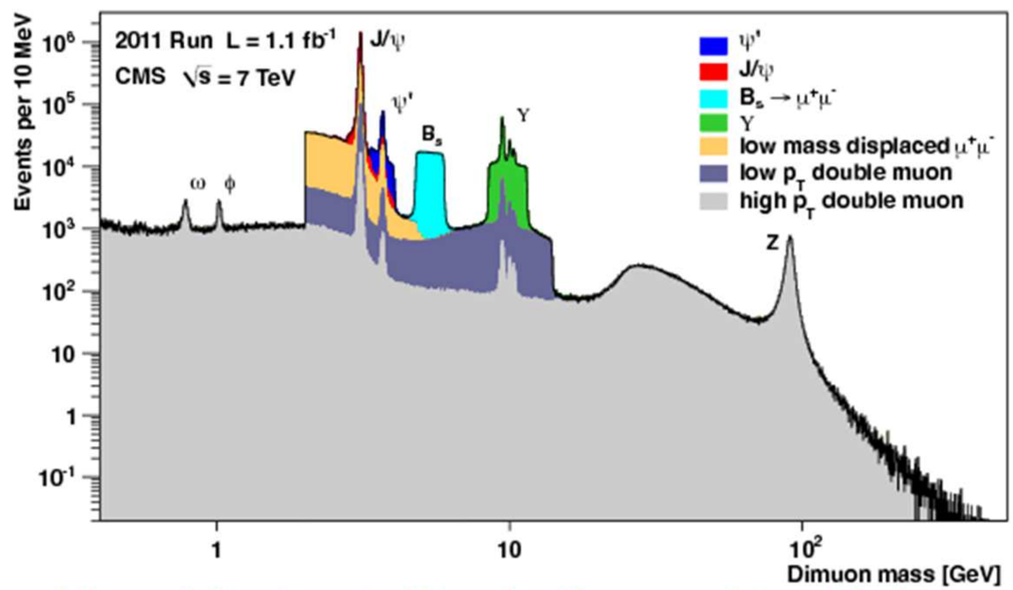
First part of 2011 pp data, including trigger, data set and quality selections
(not possible starting from educational example, fast, suitable for interactive online exercise)

Two variants:

CMS detector performance
EPJ Web of Conferences 70 (2014) 00028 (page 6)



The CMS trigger system, TRG-12-001
JINST 12 (2017) 01, P01020 (page 83)



Unfortunately: no documentation of details of cuts for either of the two (editors/authors could not help)
-> need to make educated guesses

-> two challenges: technicalities
details of event selection (NanoAODRun1 has sufficient information to reflect any cuts)

Reminder: POET vs. nanoAODplus

POET	**nanoAODplus**
both are for **research level applications**	
suited for **all analyses**	suited for ~**50% of analyses** (nanoAOD concept)
both have ~**50% direct content “dumps”** and ~**50% “derived” variables/objects**	
examples for access to AOD physics detector objects, and derived objects. **freely tunable** by users. -> users need to **make reasonable choices** themselves	fully **validated choices** partially already made for users (nanoAOD concept) -> users need to understand fewer technicalities -> **provide documentation** for which analyses the choices made are useful
users need **legacy CMSSW** -> VMs or containers	users need **plain ROOT only**, do not need VMS or container encapsulation
typically CPU days/weeks/months -> need **computing farm**	typically CPU minutes/hours/days -> need **laptop**
input from **CMSSW 5_3_32** only	input from **CMSSW 4_2_8** (2010), **5_3_32** (2011/12), **7_6_4** (2015), **10_6_26** (UL, validation)
not forward compatible with Run 2	partially **forward compatible** with Run 2 nanoAOD (2023 release of 2016 data)
both have imperfections, and **profit from user feedback**	

The currently available NanoAODRun1 data and MC sets

Data:

2010A: Mu, MuOnia, EG
2010B: Mu, MuOnia, Electron, Jet
2011A: SingleMu, DoubleMu, MuOnia, SingleElectron, DoubleElectron, Jet
2011B: SingleMu, DoubleMu, MuOnia, SingleElectron, DoubleElectron, Jet
2012B: SingleMu, DoubleMuParked, MuOnia, SingleElectron, DoubleElectron
2012C: SingleMu, DoubleMuParked, MuOnia, SingleElectron, DoubleElectron

MC:

more in pipeline

2011 DY MC
2011 ttbar->LL MC
2012 ttbar MC

Using NanoAODRun1 with TChain

Issue: Different individual NanoAODRun1 ntuples cover different data subranges
-> they will have different HLT_xxx and L1_xxx trigger bits stored.
A TChain will only make the triggers of the first file available.

Possible solutions:

- Use the files merged using the haddnano.py NanoAODTool (already done for you).
  -> trigger bits will all be merged, but files get very large (10-100 Gb).
  Preferred solution if your system has enough resources to handle these files (usually the case for institutional resources, not always on private resources).
- Use the trigger summary variables Trig_xxx instead (or no trigger variables at all for simple applications). Then, chaining the small ntuples is not an issue.
- Make an external preselection (skimming) to select only events having the triggers set that you want, and store a smaller merged ntuple (documentation will be provided later)
- Produce your histograms separately from each small file, and add the histograms in a later step. This will always work with any system, but might be a bit tedious.