

Generators and Datasets for Simulation Production Campaign

Sakib Rahman
On behalf of the ePIC Production WG
University of Manitoba
5 February 2024

Latest Completed Production Campaign ([23.12.0](#))

For live campaign updates, follow the [firehose](#) mattermost channel.

To see what was run in past campaigns, review our campaign history pages for [reconstructed output files](#) and [full geant4 simulation output files](#).

Update is provided at the end of the campaign via email/web. To learn how to access files on xrootd, review our [FAQ](#) page.

To learn about our simulation and analysis framework, review the tutorials on the [ePIC Collaboration Landing Page](#).

Example:

Listing and viewing Reconstructed Outputs for a particular dataset

xrdfs [root://dtn-eic.jlab.org](#) ls
[/work/eic2/EPIC/RECO/23.12.0/epic_craterlake/DIS/NC/18x275/minQ2=10](#)

where the different segments indicate [server address](#), [base address](#), [detector config](#), [physics process](#) and [beam properties](#) respectively.

The corresponding event generator files will be at
[/work/eic2/EPIC/EVGEN/DIS/NC/18x275/minQ2=10](#)
and geant4 simulation output (for small subset) will be at
[/work/eic2/EPIC/FULL/23.12.0/epic_craterlake/DIS/NC/18x275/minQ2=10](#)

Copy files locally using xrdcp or open them in root directly

Current Campaign (24.02.0)

We have organized a group of datasets (Physics Processes and Backgrounds) that will be run every campaign (moving gradually towards version control for everything). Besides regular production campaign datasets, we will accommodate charter requests from Physics Working Groups depending on availability of resources after approval by Analysis and Software coordinators (AC/SCs).

Datasets in Simulation Campaign Workflow

Generator versioning and input cards reproducibility

Generation of the hepmc input events

[Afterburner](#) if appropriate and [conversion to hepmc3.tree.root](#)

Get approval from AC/SCs

[Responsibility of Physics Working Groups](#)

Upload to xrootd at `root://dtn-eic.jlab.org//work/eic2/EPIC/EVGEN`

[Campaigns](#) repository stores location of specific datasets in csv files and calculates timing benchmarks

Total number of events in each hepmc file is divided into 2 hr runtime segments and submitted as jobs to open science grid

[Responsibility of Production WG](#)

EIC / campaigns / datasets



Feature add proton beamgas dataset pythia8.306 1.0
Sakib Rahman authored 1 week ago

e8603d31



main

datasets / BACKGROUNDS / pythia8.306-1.0_ProtonBeamGas_100GeV.csv

Find file

Blame

History

Permalink



pythia8.306-1.0_ProtonBeamGas_100GeV.csv 113 B



Blame

Edit



BACKGROUNDS/BEAMGAS/proton/pythia8.306-1.0/100GeV/pythia8.306-1.0_ProtonBeamGas_100GeV_run*

hepmc3.tree.root

0

0

Index

Segment

```
jug_xl> rahmans@ifarm1802:/w/eic-scshef2104/users/rahmans$ xrdls root://dtn-eic.jlab.org ls /work/eic2/EPIC/EVGEN/BACKGROUNDS/BEAMGAS/proton/pythia8.306-1.0/100GeV | head -n 2
/work/eic2/EPIC/EVGEN/BACKGROUNDS/BEAMGAS/proton/pythia8.306-1.0/100GeV/pythia8.306-1.0_ProtonBeamGas_100GeV_run001.hepmc3.tree.root
/work/eic2/EPIC/EVGEN/BACKGROUNDS/BEAMGAS/proton/pythia8.306-1.0/100GeV/pythia8.306-1.0_ProtonBeamGas_100GeV_run002.hepmc3.tree.root
jug_xl> rahmans@ifarm1802:/w/eic-scshef2104/users/rahmans$ xrdls root://dtn-eic.jlab.org ls
/work/eic2/EPIC/RECO/23.12.0/epic_craterlake/BACKGROUNDS/BEAMGAS/proton/pythia8.306-1.0/100GeV | head -n 2
/work/eic2/EPIC/RECO/23.12.0/epic_craterlake/BACKGROUNDS/BEAMGAS/proton/pythia8.306-1.0/100GeV/pythia8.306-1.0_ProtonBeamGas_100GeV_run001.0000.eicrecon.tree.edm4eic.root
/work/eic2/EPIC/RECO/23.12.0/epic_craterlake/BACKGROUNDS/BEAMGAS/proton/pythia8.306-1.0/100GeV/pythia8.306-1.0_ProtonBeamGas_100GeV_run001.0001.eicrecon.tree.edm4eic.root
```

Physics in Production (Unversioned)

Dataset	Expected Corehours	Generator
DIS NC 10x100 [minQ2=1, minQ2=10, minQ2=100, minQ2=1000] 18x275 [minQ2=1, minQ2=10, minQ2=100, minQ2=1000] 5x41 [minQ2=1, minQ2=10, minQ2=100]	49k 98k 24k	Pythia8
DIS CC 10x100 [minQ2=100, minQ2=1000] 18x275 [minQ2=100, minQ2=1000] 5x41 [minQ2=100]	21k 47k 7k	Pythia8
EXCLUSIVE TCS ABCONV [10x100 (hel minus), 18x275(hel minus/plus), 5x41(hel minus/plus)]	45k	?
EXCLUSIVE UCHANNEL PI0 UCHANNEL RHO	1k 0.3k	?
EXCLUSIVE DVCS ABCONV	14k	?

Physics in Production (Unversioned)

Dataset	Expected Corehours	Generator
SINGLES 3to50degrees(e-, e+, pi-, pi+, pi0, kaon-, kaon+, gamma, proton) 45to135degrees(e-, e+, pi-, pi+, pi0, kaon-, kaon+, gamma, proton) 130to177degrees(e-, e+, pi-, pi+, pi0, kaon-, kaon+, gamma, proton) etaScan(e-, mu-, gamma)	29k 24k 23k 9k	

Physics in Production (Versioned)

Dataset	Expected Corehours	Generator
SIDIS 10x100 [q2_0to1] 18x275 [q2_0to1] 5x41 [q2_0to1]	75k 144k 46k	pythia6-eic 1.0.0
LAMBDA	?	LambdaGen pythia8.306-1.0
DVMP	0.3k	DVMPdataset EplC1.0.0-1.0
DEMP	22k	DEMPGen 1.0.0
EXCLUSIVE DIFFRACTIVE PHI ABCONV PHOTOPRODUCTION JPSI PHOTOPRODUCTION JPSI ABCONV	51k ? ?	SARTREdataset sartre-1.39-1.0

Backgrounds In Production (Unversioned)

Dataset	Expected Corehours	Generator
MERGED (10x100)	0.06k	?
ELECTRON BEAMGAS	?	?

Backgrounds In Production (Versioned)

Dataset	Expected Corehours	Generator
PROTON BEAMGAS [275 GeV, 100 GeV]	33k	pythia8.306-1.0

Criteria for MCEGs to be Included in Production

Must support the hepmc3 output format

Then the [hepmc3ascii2root](#) tool can be used to convert from hepmc3 ascii text to hepmc3.tree.root, the uniform standard adopted for all datasets in production.

Must not duplicate effort

We do not have the resources to run multiple generators for resource-intensive physics processes. For example: high Q2 DIS with Pythia8, Pythia6, Djangoh, etc. Physics WGs need to choose a reference generator for each process and inform Production WG.

Must satisfy data and analysis preservation requirements

Generator source code and scripts, steering files, meta data relevant to reproducing each dataset should be tracked in a publicly accessible repository. See next slide for more information on version control.

The repository should also contain a readme file describing all the steps necessary to completely reproduce a particular set of files.

Later, the Production WG will group these repositories together as submodules in an umbrella repository under [EIC](#) github namespace.

Version Control for Production Datasets

Follow the [input preprocessing guidelines](#)

Case 1: You control both generator source code and scripts/steering files. Consider the following repository versioning scheme:

Major.Minor.Patch

- Major: Changes that invalidate previously generated hepmp3.tree.root files.
- Minor: Changes that make previously generated hepmp3.tree.root files less ideal to use.
- Patch: Changes that don't modify the generated hepmp3.tree.root files in any statistically relevant way. For example, the first release for such a repository can be tagged 1.0.0 and incremented accordingly.

Case 2: You don't control the generator source code but only the scripts/steering files.Consider a two part scheme:

Part1-Part2

- Part1: Semantic version of the generator that was used to create the dataset
- Part2: Major.Minor for your steering files. For example, if you used Pythia6.428 and some steering files to create the dataset, the first release should be tagged Pythia6.428-1.0

File Organization for Upload to Xrootd

Organization of files	Example
<physics processes>/<generator repository release tag>/<electron momentum>x<proton momentum>/q2_<minimum q2>to<maximum q2> /<generator repository release tag>_<physics processes>_<electron momentum>x<proton momentum>_q2_<minimum q2>to<maximum q2>_run<index>.hepmp3.tree.root	DIS/NC/pythia6.428-1.0/10x100/q2_10to100/pythia6.428-1.0_DIS-NC_10x100_q2_10to100_run001.hepmp3.tree.root

Summary and Conclusion

- 1) Process of bringing new datasets into production (same for regular campaigns and special charters)
Physics WGs (Reproducible Generation) => Software and Analysis Coordinators (Approval) => Production WG (Transfer and Running)
- 2) Event generators must meet the eligibility criteria to be included in production with regards to output format (hepmc3.tree.root), version tracking and effort consolidation.
- 3) Please attend the upcoming ePIC Software, Computing, and Physics Discussion on Feb. 14 (<https://indico.bnl.gov/event/21772/>). We would appreciate feedback from Physics WGs in reviewing the list of physics processes and their corresponding physics generators for inclusion in regular or charter productions.