The MCPLOTS project

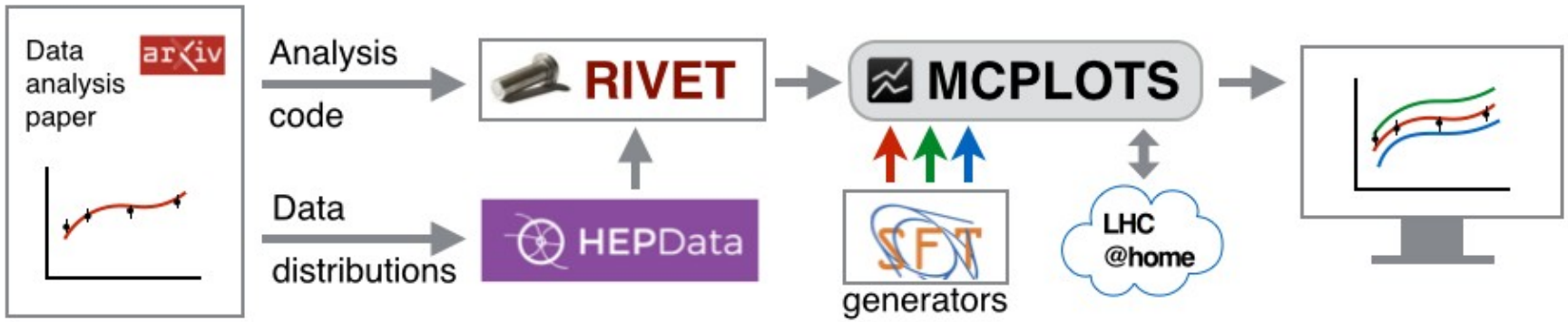
A. Karneyeu[1], N. Korneeva[2,3], P. Skands[2]

[1] University of Notre Dame
[2] Monash University
[3] National Research Tomsk Polytechnic University

DESY
27 May 2024

MCPLOTS : overview

http://mcplots.cern.ch/



https://rivet.hepforge.org/ https://ep-dep-sft.web.cern.ch/

https://www.hepdata.net/ https://lhcathome.web.cern.ch/

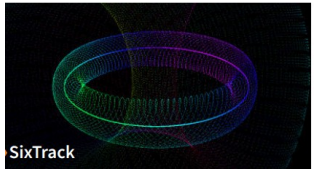
LHC@home

LHC @home | Volunteer computing for the LHC

Open Eng. 7 (2017) 1, 378-392

computing platform where voluteers donate idle time on their computers to their preferred project

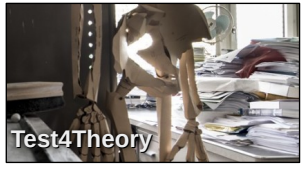
2004



Birth of the LHC@home : an outreach event for CERN's 50th anniversary

SixTrack simulation of beam dynamics runs on Windows/OSX/Linux

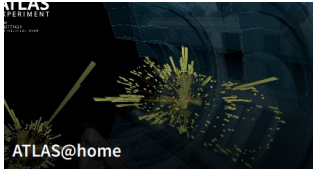
2010

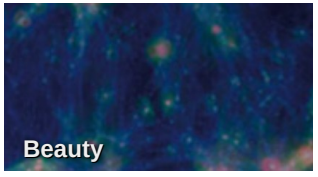


The first project to use virtualization (CernVM)

PoS ISGC2012 (2012) 036

2016

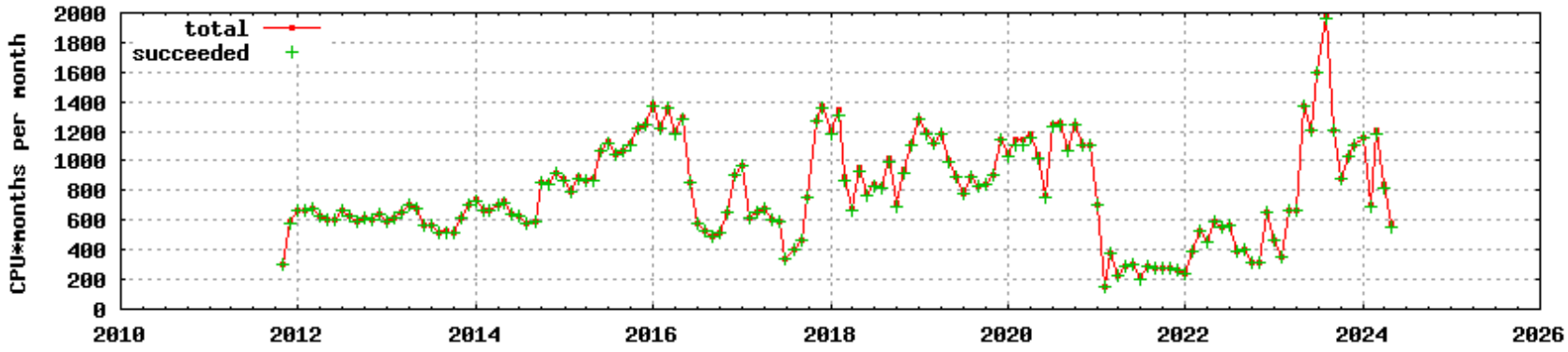




Main CERN experiments joined the platform

MCPLOTS and LHC@home

CPU resources accessed by the Test4Theory project (monthly averaged)

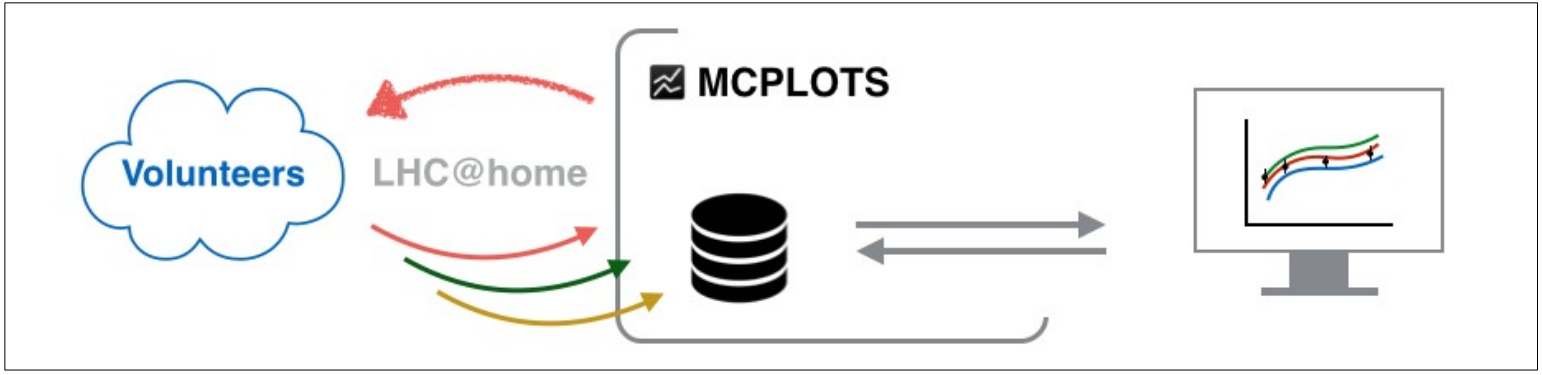


Key numbers :

~800 volunteers

~10^9 events per day

MCPLOTS and LHC@home



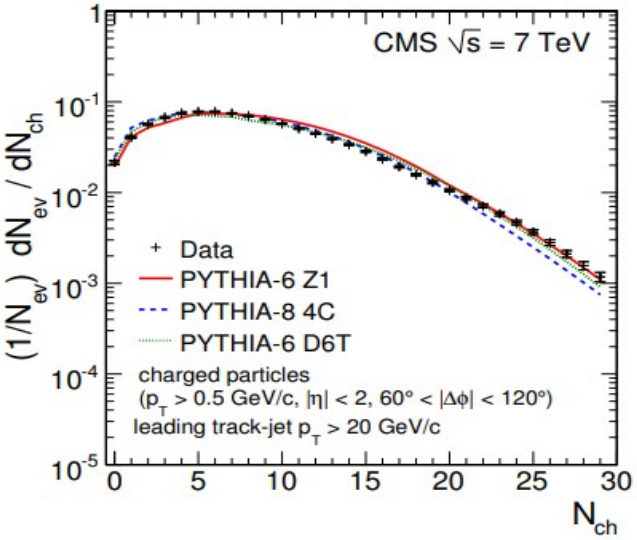
MCPLOTS workflow

Jobs (regular executable files) are distributed to volunteers
Completed ones are moved to MCPLOTS and stored on the **server**
Descriptions of obtained MC distributions are stored in the **database**
The website mcplots.cern.ch/ operates with **queries** to this DB

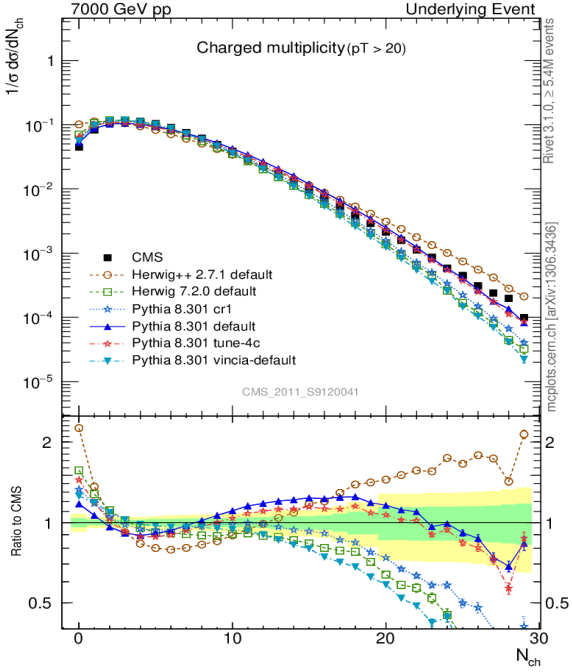
Plots

Generated distributions are collected on the MCPLOTS web server : dozens of generator-version-tune combinations for each data distribution ; they are plotted on the fly by a user's request.

Paper version



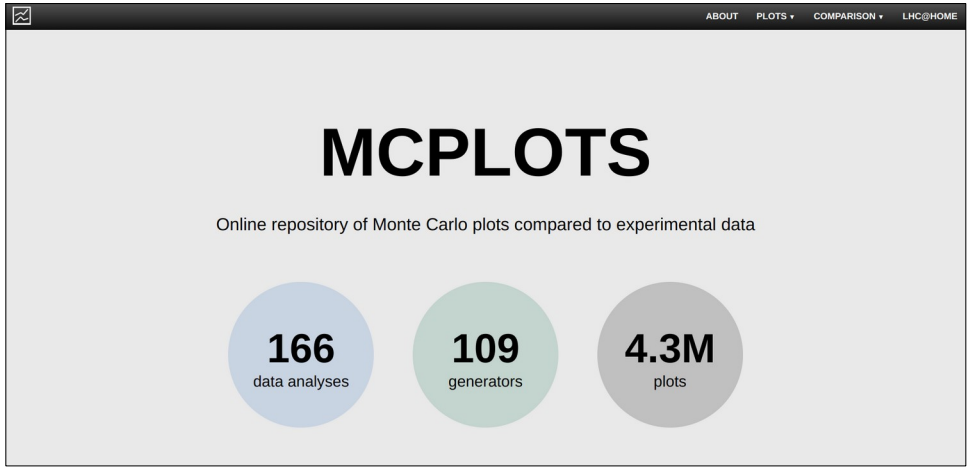
MCPLOTS version

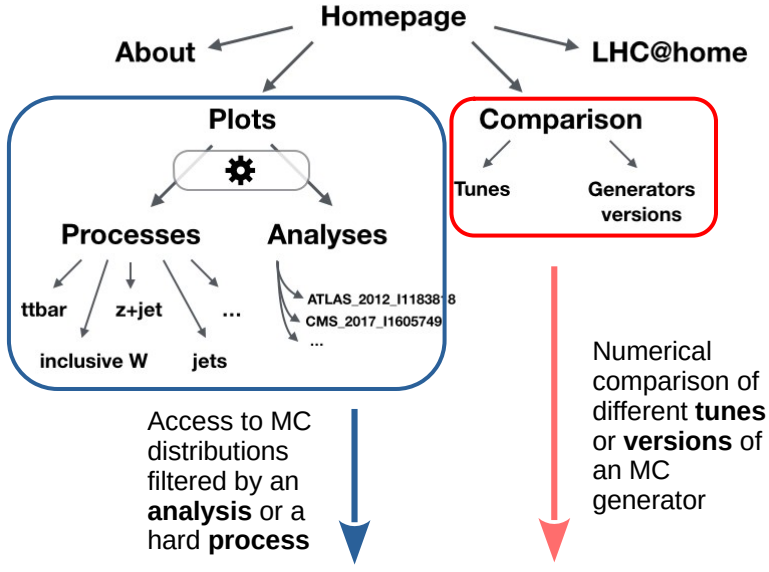


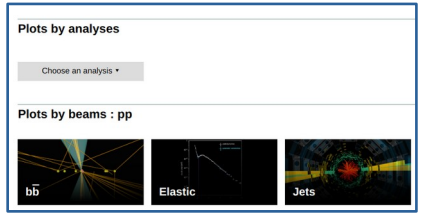
...or whichever generator-version-tune you want

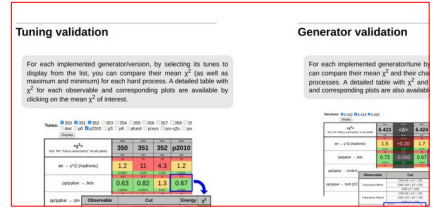
Website

http://mcplots.cern.ch/







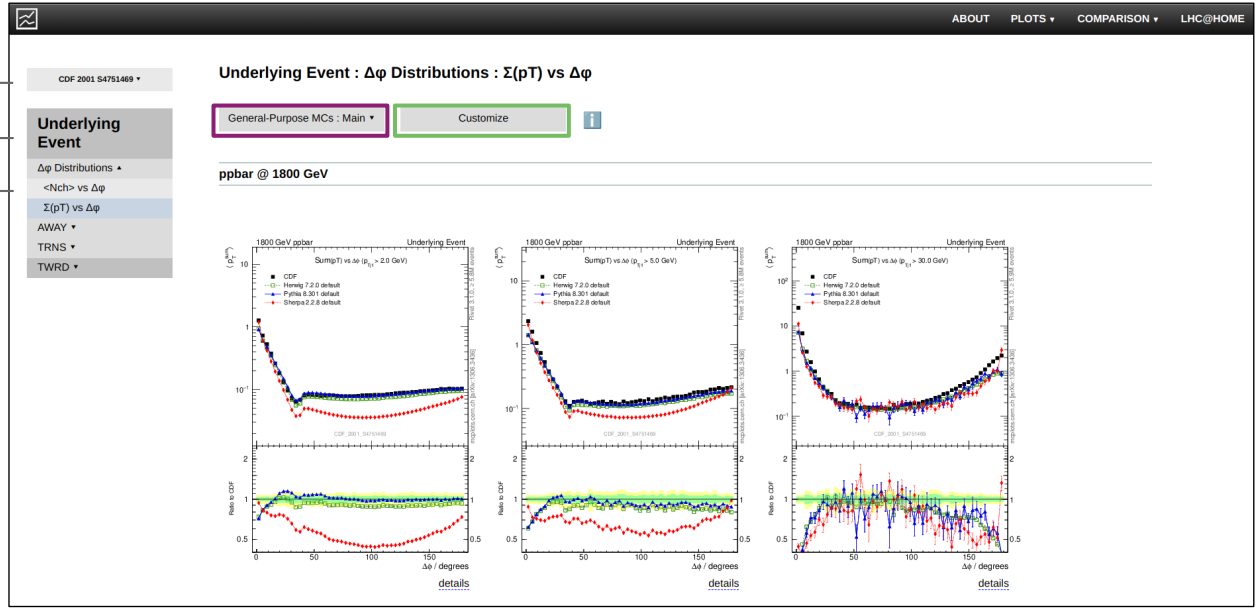


Website : plots

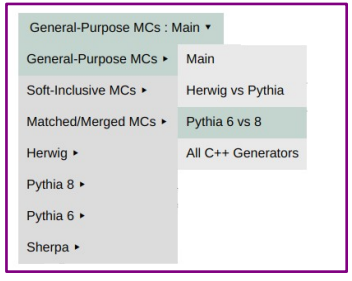
Analysis filter (if used)

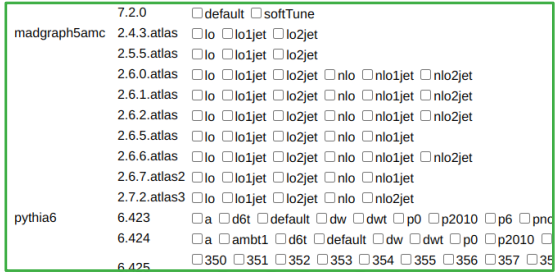
Hard process

Individual distributions



Possibility to choose what to plot either from **a pre-defined preset** or from **all MCs**

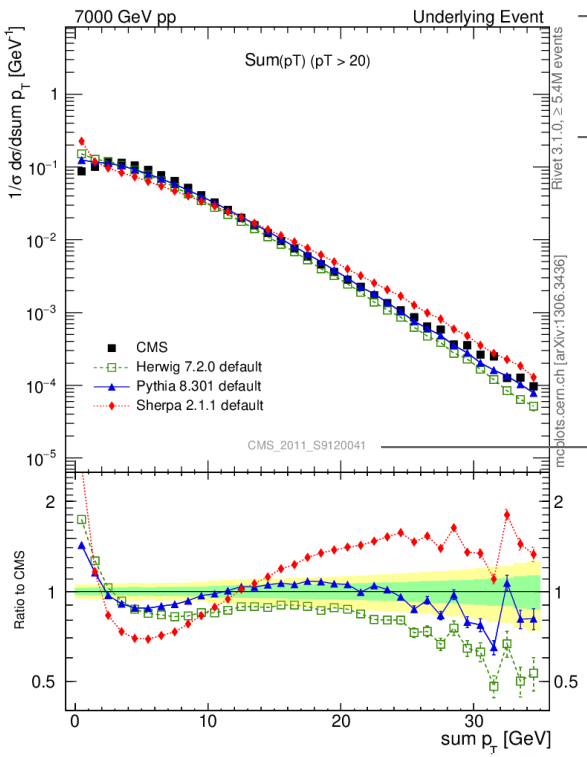




Website : plots

Beam parameters

Generator-version-tune
for each MC curve



Hard process

RIVET version and
# of MC events

RIVET reference

details

Download as: .pdf .eps .png .script.tgz #

CMS experiment: data | article paper

Herwig 7 (Def): data |generator card

Pythia 8 (Def): data |generator card

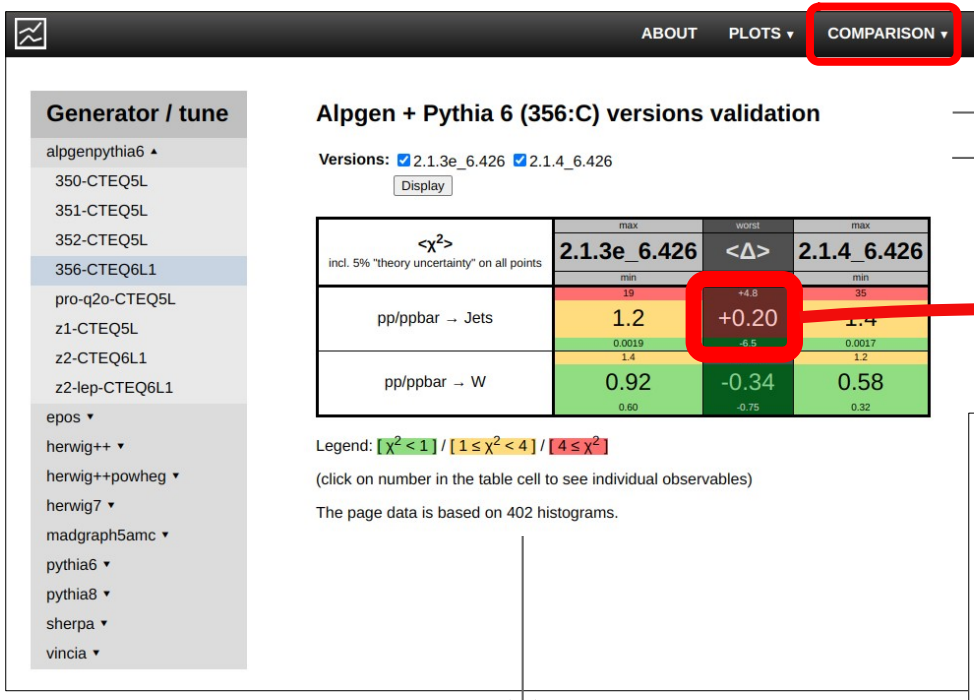
Sherpa (Def): data |generator card

Steering files and results
for each MC curve

Plot in higher resolution

Data distribution and article paper

Website : comparison



Generator (tune) name

Available versions to compare

χ² for individual observables

Number of distributions used to calculate <χ²>

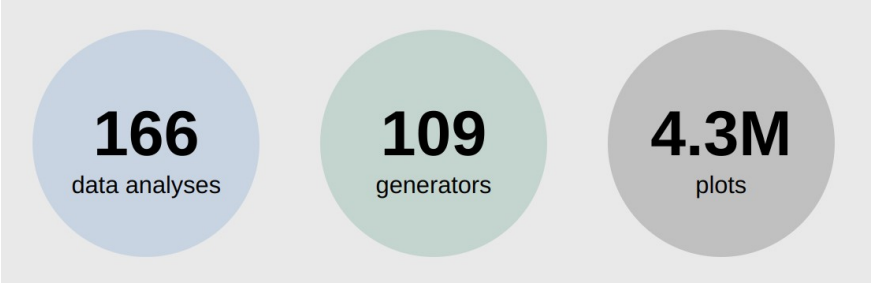
Details for Alpgen + Pythia 6 (356:C) v.2.1.3e_6.426 vs. v.2.1.4_6.426

pp/ppbar → Jets

Observable	Cut	Energy	$\chi^2_{+5\%}$ (2.1.3e_6.426)	Δ	$\chi^2_{+5\%}$ (2.1.4_6.426)
23-jet Correlation	CMS 2013 (Forward)	7000	5.2	-3.8	1.4
	CMS 2013 (Central)	7000	5.1	-3.8	1.3
ET(J1)	CDF 1994	1800	n/a*	-	n/a*
ET(J2)	CDF 1994	1800	n/a*	-	n/a*
Transverse Minor	CMS 90 < pT < 125	7000	0.41	+0.49	0.90
	CMS 125 < pT < 200	7000	0.99	-	n/a*
	CMS pT > 200	7000	n/a*	-	n/a*
Transverse Thrust	CMS 90 < pT < 125	7000	0.72	+0.88	1.6
	CMS 125 < pT < 200	7000	0.72	-	n/a*
	CMS pT > 200	7000	1.1	-	n/a*

Current status

Always shown on the main page :



Implemented generators : Alpgen, Epos, Herwig++ and Herwig7, MadGraph, Powheg-Box, Pythia6 and Pythia8, Sherpa, Vincia

1065 generator-version-tune combinations

166 data analyses with **5016 data distributions** implemented so far refer mostly to the *ee* and *pp* HEP collider experiments: ATLAS, CMS, D0 etc.

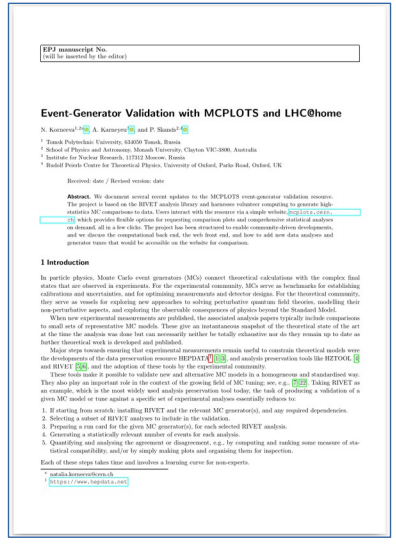
The repository is continously filled and the source code of the project is available :

https://gitlab.cern.ch/MCPLOTS/mcplots

Paper

The first paper was published ~10 year ago :
MCPLOTS: a particle physics resource based on volunteer computing
https://doi.org/10.1140/epjc/s10052-014-2714-9

The second one : **January 2024**
Event-Generator Validation with MCPLOTS and LHC@home
in arXiv : https://arxiv.org/abs/2401.10621
submitted to EPJ+


EPJ manuscript No.
(will be inserted by the editor)

Event-Generator Validation with MCPLOTS and LHC@home

N. Korneva[1,2,a], A. Karneyeu[3] and P. Skands[4,3]

[1] Tomsk Polytechnic University, 634050 Tomsk, Russia
[2] School of Physics and Astronomy, Monash University, Clayton VIC-3800, Australia
[3] Institute for Nuclear Research, 117312 Moscow, Russia
[4] Rudolf Peierls Centre for Theoretical Physics, University of Oxford, Parks Road, Oxford, UK

Received: date / Revised version: date

Abstract. We document several recent updates to the MCPLOTS event-generator validation resource. The project is based on the RIVET analysis library and harnesses volunteer computing to generate high-statistics MC comparisons to data. Users interact with the resource via a simple website, mcplots.cern.ch, which provides flexible options for requesting comparison plots and comprehensive statistical analyses on demand, all in a few clicks. The project has been structured to enable community-driven developments, and we discuss the computational back end, the web front end, and how to add new data analyses and generator tunes that would be accessible on the website for comparison.

1 Introduction

In particle physics, Monte Carlo event generators (MCs) connect theoretical calculations with the complex final states that are observed in experiments. For the experimental community, MCs serve as benchmarks for establishing calibrations and uncertainties, and for optimising measurements and detector designs. For the theoretical community, they serve as vessels for exploring new approaches to solving perturbative quantum field theories, modelling their non-perturbative aspects, and exploring the observable consequences of physics beyond the Standard Model.

When new experimental measurements are published, the associated analysis papers typically include comparisons to small sets of representative MC models. These give an instantaneous snapshot of the theoretical state of the art at the time the analysis was done but can necessarily neither be totally exhaustive nor do they remain up to date as further theoretical work is developed and published.

Major steps towards ensuring that experimental measurements remain useful to constrain theoretical models were the developments of the data preservation resource HEPDATA [1–3], and analysis preservation tools like HZTOOL [4] and RIVET [5,6], and the adoption of these tools by the experimental community.

These tools make it possible to validate new and alternative MC models in a homogeneous and standardised way. They also play an important role in the context of the growing field of MC tuning; see, e.g., [7–9]. Taking RIVET as an example, which is the most widely used analysis preservation tool today, the task of producing a validation of a given MC model or tune against a specific set of experimental analyses essentially reduces to:

1. If starting from scratch: installing RIVET and the relevant MC generator(s), and any required dependencies.
2. Selecting a subset of RIVET analyses to include in the validation.
3. Preparing a run card for the given MC generator(s), for each selected RIVET analysis.
4. Generating a statistically relevant number of events for each analysis.
5. Quantifying and analysing the agreement or disagreement, e.g., by computing and ranking some measure of statistical compatibility, and/or by simply making plots and organising them for inspection.

Each of these steps takes time and involves a learning curve for non-experts.

[a] natalia.korneva@cern.ch
[1] https://www.hepdata.net

- Full description of the updated repository and database structure
- Comprehensive user’s guide (the website functionality)
- Developer’s guide : how to implement
  - a new data analysis
  - a new generator (version)
  - a new generator tune
- Phase-space cuts discussion

Phase-space cuts

For many modern analyses using high-p_T jets we have to populate a corresponding phase-space region which is impossible without a generator-level hard-process cut

Example: `CMS_2013_I1265659` / Probing color coherence effects in pp collisions at √s=7 TeV

Event selection : 3-jets events with p_{T1} > **100** GeV

MCPLOTS implementation :

[beam]	[proc]	[Ecm]	[par]	[analysis histogram]	[obs]	[cuts]
pp	ttbar	13000	?	CMS_2013_I1265659_d01-x01-y01	jj.beta	cms-coh

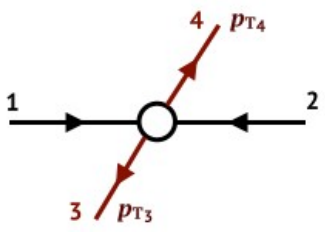
Data analysis cut : physical **particle-level** final state

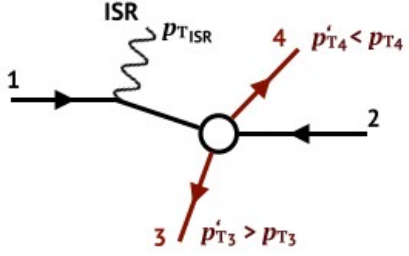
Generation cut : hard **partonic** process ;

it should be **broader** than the analysis cut

low enough *so as not to lose events that can pass the data-analysis cut after a PS*

high enough *to ensure the population of the desired phase-space region*





2 generic methods to determine an optimal generation cut

Phase-space cuts

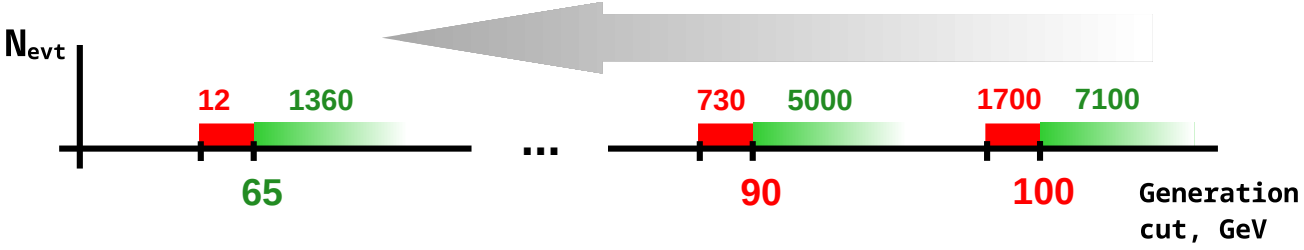
2 methods to determine an optimal generation cut

Illustration : `CMS_2013_I1265659` / Probing color coherence effects in pp collisions at √s=7 TeV

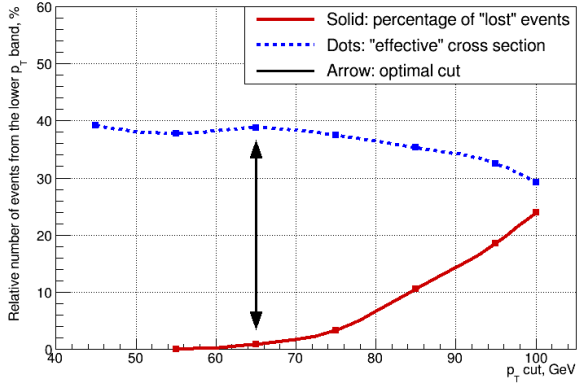
Event selection : 3-jets events with p_{T1} > **100** GeV

1. Estimation of $\frac{N_{lost}}{N_{test}}$

N_{lost} : *number of event that would not satisfy the given generation cut but which would pass the analysis cut*



Should not exceed a fraction of a percent



2. Estimation of the « effective » cross section $\sigma \cdot N_{test}$

Should be stabilised

Development

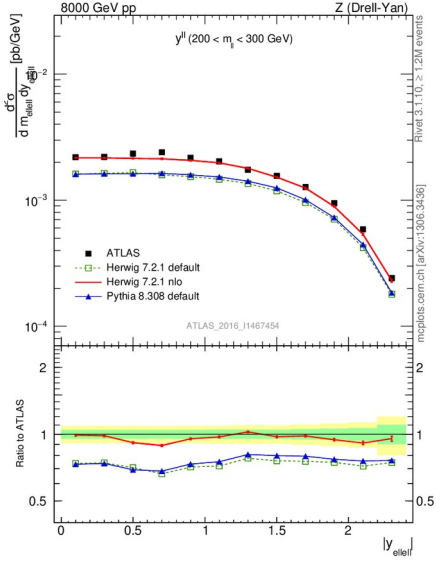
Matching/merging

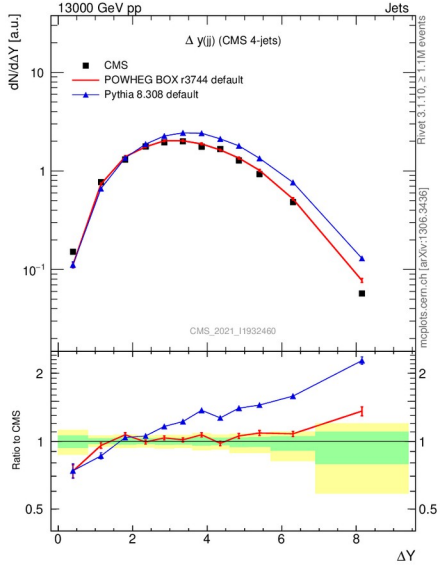
- NLO generator cards (in collaboration with the Herwig team)
- Various matching schemes to try


+

Variation of Powheg-Box/Pythia matching parameters to estimate matching systematics

Several simple processes are already implemented, with the default matching to the angular-ordered shower





Only default parameters are used now :

- `pTdef = 1`
- `pThard = 0`

Development

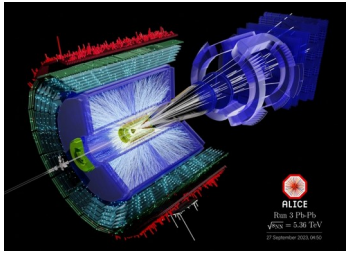
Expanding the coverage

166 data analyses with **5016 data distributions** implemented so far refer mostly to the *ee* and *pp* **HEP collider experiments**: ATLAS, CMS, D0 etc.

chromo

unified frontend to generators used for cosmic-ray studies

- MCPLOTS was presented to people from astroparticle community
- proposal from the CHROMO team to integrate it into MCPLOTS
- in process of discussing the possibility of installing CHROMO in CERN



- re-implementation of heavy ion runs
- new analyses and generators

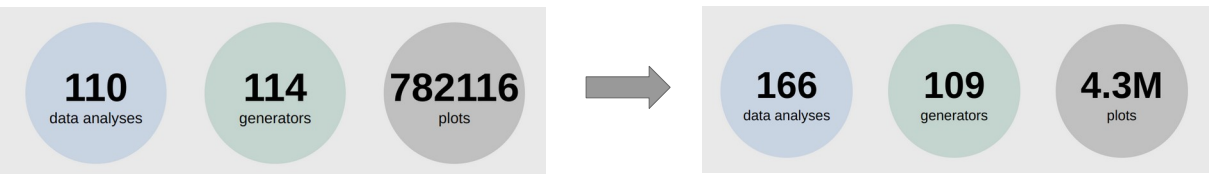
Development

Validation resource for the main CERN experiments

- Close work with the experiments on their validation tasks
- Automation of adding new generator versions for quicker validation
- Improving the website validation pages

Need more statistics !

Recent update :



- New data analyses have been implemented
- Number of **plots** are increased by **more than 4x**

Plan to **remove** some outdated versions/tunes to increase statistics in distributions of interest

Example :

Pythia 6
6 versions: 6.428, 6.427, 6.426, 6.425, 6.424, 6.423.
58 tunes: 345, 346, 347, 348, 349, 350, 351, 352, 353, 354, 355, 356...

Summary

- MCPLOTS : Online repository of Monte Carlo plots compared to experimental data
- CPU power: LHC@home
- Website: http://mcplots.cern.ch/
- Source code available to download

  https://gitlab.cern.ch/MCPLOTS/mcplots
- MCPLOTS paper

  https://arxiv.org/abs/2401.10621

BACKUP

Database

Information about the distributions

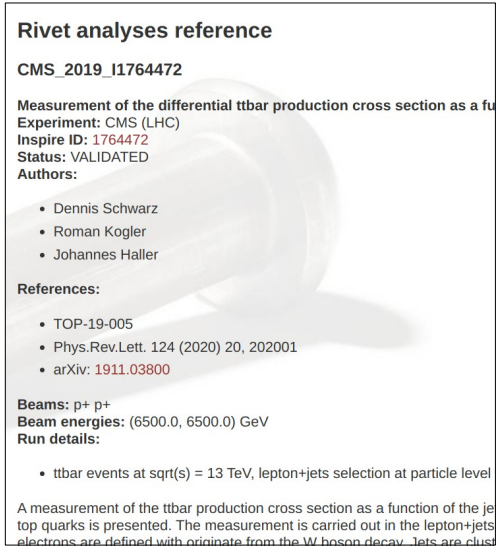
[id]	[fname]	[type]	[prc]	[obs]	[tune]	[exp]	[ref]	[hid]	[beam]	[Ecm]	[cuts]	[gen]	[ver]
90	dat/CDF_2005_S6217184-ppbar-1960/jets-js_int-cdf3-037-d07-x01-y01/pythia8-8.244-tune-2m.dat	mc	jets	js_int	tune-2m	CDF	CDF_2005_S6217184	d07-x01-y01	ppbar	1960	cdf3-037	pythia8	8.244
91	dat/CDF_2005_S6217184-ppbar-1960/jets-js_int-cdf3-037-d07-x01-y01/CDF_2005_S6217184.dat	data	jets	js_int		CDF	CDF_2005_S6217184	d07-x01-y01	ppbar	1960	cdf3-037		

The website operates with queries to this table

Data analysis implementation

Adding a new RIVET analysis to the MCPLOTS project = adding its description to the configuration files

[RIVET page of the analysis]



Rivet analyses reference

CMS_2019_I1764472

Measurement of the differential ttbar production cross section as a fu
Experiment: CMS (LHC)
Inspire ID: 1764472
Status: VALIDATED
Authors:

- Dennis Schwarz
- Roman Kogler
- Johannes Haller

References:

- TOP-19-005
- Phys.Rev.Lett. 124 (2020) 20, 202001
- arXiv: 1911.03800

Beams: p+ p+
Beam energies: (6500.0, 6500.0) GeV
Run details:

- ttbar events at sqrt(s) = 13 TeV, lepton+jets selection at particle level

A measurement of the ttbar production cross section as a function of the je
top quarks is presented. The measurement is carried out in the lepton+jets
electrons are defined with originate from the W boson decay. Jets are clust

[RIVET .plot file]

```
BEGIN PLOT /CMS_2019_I1764472/d01-x01-y01
Title=CMS, 13 TeV, jet mass in boosted top quark decays
XLabel=$m_\mathrm{jet}$
YLabel=$\frac{d\sigma}{dm_\mathrm{jet}} \frac{\mathrm{fb}}{\mathrm{GeV}}$
LogY=0
END PLOT

BEGIN PLOT /CMS_2019_I1764472/d02-x01-y01
Title=CMS, 13 TeV, normalized jet mass in boosted top quark decays
XLabel=$m_\mathrm{jet}$
YLabel=$\frac{1}{\sigma} \frac{d\sigma}{dm_\mathrm{jet}} \frac{1}{\mathrm{GeV}}$
LogY=0
END PLOT
```

$t\bar{t}$ differential x-section as a function of the jet mass...

...and its normalized version

→ 📝 `mcplots/scripts/mcprod/configuration/rivet-histograms.map :`

```
[beam] [proc]  [Ecm]   [par]      [analysis histogram]                 [obs]   [cuts]
 pp    ttbar   13000    -         CMS_2019_I1764472_d01-x01-y01        j.m     cms2019-ttboost
 pp    ttbar   13000    -         CMS_2019_I1764472_d02-x01-y01        j.m     cms2019-ttboost
```

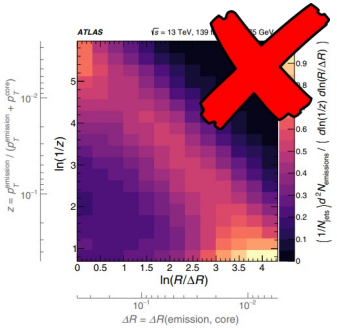
Data analysis implementation

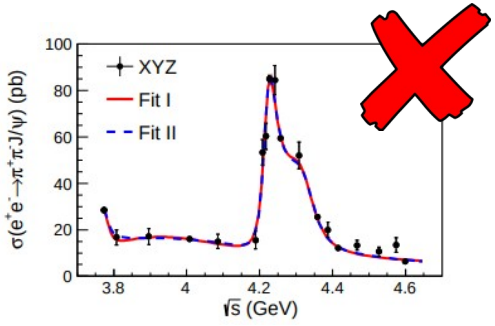
Another example : `ATLAS_2022_I2077570`

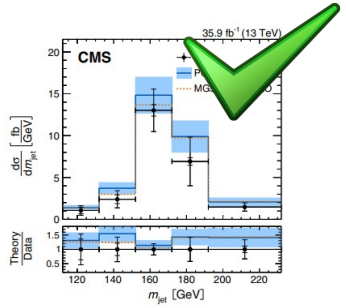
```
[beam] [proc]  [Ecm]   [par]  [analysis histogram]                                  [obs]    [cuts]
 pp     z1j     13000   55     ATLAS_2022_I2077570:LMODE=ELEL_d01-x01-y01   z.pt     atlas2022-zj
 pp     z1j     13000   55     ATLAS_2022_I2077570:LMODE=MUMU_d01-x01-y01   z.pt     atlas2022-zj
 pp     z1j     13000   55     ATLAS_2022_I2077570_d01-x01-y01              z.pt     atlas2022-zj
```

- ✔ Different running modes `LMODE`
- ✔ Phase-space cuts to be passed to MC generators `[par]`
  - minimum and maximum $\hat{p}_\perp$ of the final-state partons
  - their minimum and maximum invariant mass

Restrictions :
only one-dimensional distributions obtained with a fixed beam energy are allowed







Data analysis implementation

Displaying new plots : we have to define the correspondence between internal names and names displayed on the website.

`mcplots/scripts/mcprod/configuration/rivet-histograms.map :`

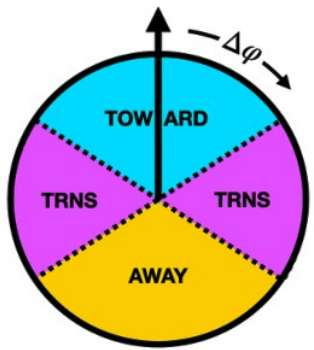
```
[beam] [proc]  [Ecm]   [par]      [analysis histogram]              [obs]              [cuts]
 pp     ue     7000     -         ATLAS_2010_S8894728_d12-x01-y03   avgpt-vs-nch-away  atlas4
```

Variable : mean p_T of the charged particles against the charged multiplicity in the AWAY region

Variable group : AWAY

Hard process : Underlying event

Additional information : atlas4



Data analysis implementation

📄 `mcplots/scripts/mcprod/configuration/rivet-histograms.map :`

```
[beam] [proc]  [Ecm]   [par]        [analysis histogram]              [obs]                [cuts]
 pp     ue     7000     -          ATLAS_2010_S8894728_d12-x01-y03   avgpt-vs-nch-away  atlas4
```

📝 `mcplots/www/mcplots.conf :`

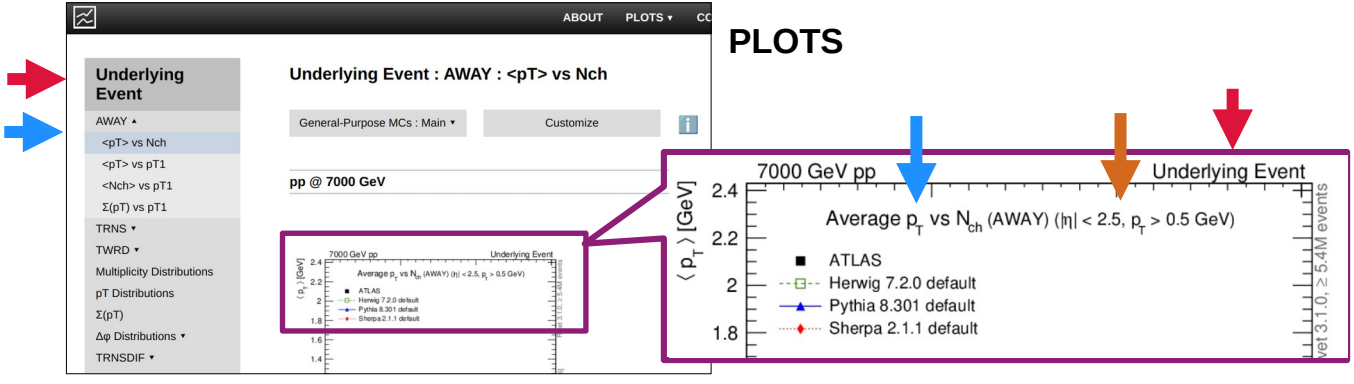
```
process_name = ! HTML name         ! name in LaTeX format
ue           = ! Underlying Event ! Underlying Event

observable name    = (HTML group name) ! HTML name          ! name in LaTeX format
avgpt-vs-nch-away = AWAY               ! &lt;pT&gt; vs Nch ! Average p_{T} vs N_{ch} (AWAY)

cut_name = ! HTML name       ! name in LaTeX format
atlas4   = ! ATLAS pT > 0.5 ! |#eta| < 2.5, p_{T} > 0.5 GeV/c
```

HTML name *for the website*
LaTeX format *for plots*

PLOTS



COMPARISON

pp/ppbar → Underlying Event

Observable	Cut	Energy	$\chi^2_{+5\%}$
<pT> vs Nch	ATLAS pT > 0.5	900	0.31
		7000	0.087
<pT> vs Nch	ATLAS pT > 0.5	900	0.72
		7000	0.16
<pT> vs Nch	ATLAS pT > 0.5	900	0.45
		7000	0.16
<pT> vs pT1	ATLAS pT > 0.5	900	0.23
		7000	0.12
<pT> vs pT1	ATLAS pT > 0.5	900	1.2
		7000	0.76
	CDF 2010	1960	n/a*