Pythia News & Plans (+ Tuning Recommendations) Peter Skands (Monash University) & Leif Gellersen (Lund University)

New Updated Comprehensive Reference (2022):

A COMPREHENSIVE GUIDE TO THE PHYSICS AND USAGE OF PYTHIA 8.3

Authors: Christian Bierlich, Smita Chakraborty, Nishita Desai, Leif Gellersen, Ilkka Helenius, Philip Ilten, Leif Lönnblad, Stephen Mrenna, Stefan Prestel, Christian T. Preuss, Torbjörn Sjöstrand, Peter Skands, Marius Utheim, and Rob Verheyen. SciPost Physics Codebases 8 (2022).



ATLAS Generator Infrastructure and Tuning Meeting

https://arxiv.org/abs/2203.11601

~ 315 pages

July 2023







Are shower variations appropriate to use in conjunction with heavyflavour fragmentation?

(Since the appropriate value of r_b or r_c varies with α_s , these variations might break agreement with LEP tuning)

Quick answer is yes

But the second point is more subtle.

Parameter hierarchies

Tuning: the higher up the chain you change something, the more it will affect the large-scale event structure -> Start at the top, and work your way down. α_s is a perturbative parameter \rightarrow affects jet rates, IR safe observables, ... r_{h} is a non-perturbative parameter (just like aLund, bLund, sigmaPT, ...) Ideally, should not attempt to determine a non-perturbative parameter if the perturbative ones are not appropriate.





Changing r_b to "compensate" for change in $\alpha_s(M_7)$?



b quark loses too little energy to radiation Increasing r_h forces it to lose more to hadronisation.







What happens to other distributions?



Illustrations



What does it

3-jet events have a smaller $\langle x_B \rangle$ than 2-jet events

So if you don't get the relative mixture of 2- to 3-jet events right, then you would be in unsafe territory trying to fit lower-scale non-perturbative parameters to an inclusive measurement of $\langle x_{\rm B} \rangle$.

What can you do?

Use the value for the shower α_s that gets the "right" 3/2 jet ratio at LEP Or: use multi-leg NLO merging (~ NNLO matching) Or: use reweighting to measured 3-jet distributions? ratio.

Or: say you want to keep the B hadron energy fraction constant. (My feeling is this overcompensates.)

Similar comments for many other tuning parameters, eg aLund, bLund, sigmaPT, ...

Parameter Hierarchies Example: Summary

- Or: use $\langle x_{\rm B} \rangle$ in an exclusive 2-jet sample that does not depend on the relative 2-to-3-jet





Are there plans for producing a Monash-level tune for VINCIA showers?

Quick answer is yes: (partly done already; more work needed for LHC)

VINCIA default hadronisation parameters Substantial changes across recent versions; full-fledged pp retune not done yet. → UE modelling not as good as Monash & not cross checked at 13 TeV (yet).

Recent effort started in Pythia to produce new default tunes (Lead: S. Mrenna)

LO and NLO level tunes (with new generation of PDFs) + uncertainties Including both the SimpleShower and VINCIA showers. Timescale ~ 1 year (?)

Feedback on Questions from ATLAS

- Is this an area where collaboration between ATLAS and PYTHIA developers would be useful?
- **PartonShowers:model = 2** (VINCIA) already automatically switches to a dedicated set of
 - (eg different values for aLund, bLund, etc) → similar level of agreement at LEP as Monash.
- Only ~3 people work on VINCIA [PZS, C T Preuss & L Gellersen, all with other big projects]; Very interested in feedback, validations, suggestions, from ATLAS!



6

VINCIA Default Tune

Name	VINCIA Default	Monash
BeamRemnants:primordialKThard	0.400	1.800
BeamRemnants:primordialKTsoft	0.250	0.900
ColourReconnection:range	1.750	1.800
MultipartonInteractions:alphaSorder	2	1
MultipartonInteractions:alphaSvalue	0.119	0.130
MultipartonInteractions:ecmPow	0.210	0.215
MultipartonInteractions:expPow	1.750	1.850
MultipartonInteractions:pT0Ref	2.240	2.280
SigmaProcess:alphaSorder	2	1
SigmaProcess:alphaSvalue	0.119	0.130
StringFlav:etaPrimeSup	0.100	0.120
StringFlav:etaSup	0.500	0.600
StringFlav:mesonCvector	1.300	$0.880 \rightarrow 1.5$
StringFlav:mesonSvector	0.530	0.550
StringFlav:mesonUDvector	0.420	0.500
StringFlav:popcornSmeson	0.750	0.500
StringFlav:popcornSpair	0.750	0.900
StringFlav:probQQ1toQQ0	0.025	0.028
StringFlav:probQQtoQ	0.077	0.081
StringFlav:probSQtoQQ	1.000	0.915
StringFlav:probStoUD	0.205	0.217
StringPT:sigma	0.305	0.335
StringZ:aExtraDiquark	0.900	0.970
StringZ:aLund	0.450	0.680
StringZ:bLund	0.800	0.980
StringZ:rFactB	0.850	0.855
StringZ:rFactC	1.150	$1.320 \rightarrow 1.5$
TimeShower:interleaveResDec	on	off

Main Reference:

Sector Showers for Hadron Collisions Helen Brooks, Christian T. Preuss, PZS JHEP 07 (2020) 032

Dedicated study of VBF with Powheg matching and sector-CKKWL merging:

A Study of QCD Radiation in VBF Higgs Production with Vincia and Pythia Stefan Höche Stephen Mrenna, Shay Payne, Christian T. Preuss, PZS SciPost Phys. 12 (2022) 1, 010

Also:

Multipole QED shower (+ plans to apply to hadron decays) Interleaved Resonance Decays (+ plans to apply to hadron decays) Full-fledged EW shower Several research projects underway LO matrix-element corrections NLO matrix-element corrections NNLO matching









ISR: Drell-Yan pT spectrum



BeamRemnants:primordialKThard	0.400	1.800
BeamRemnants:primordialKTsoft	0.250	0.900

Note:

Vincia Default Tune



Note: expect LO matrix-element corrections for many processes ~ soon



Vincia Default Tune

Multi-parton interactions: UE at 7 TeV

(Sorry, I did not manage to generate much statistics ahead of talk but wanted to show you with current version, 8.309)



Note: will probably be updated (and faster) with new QCD CR in upcoming versions



9

Great if we could have a way to dump out "effective standalone" Pythia configs from the ATLAS interface

- As is the case for Herwig and Sherpa.
- Requires more standardising of the Pythia executable? It differs from the other gens in not having one main steering binary

The quick answer is yes (but only partly)

You can dump all changed settings

bool Settings::writeFile(string toFile, bool writeAll = false) bool Settings::writeFile(ostream& os = cout, bool writeAll = false) write current settings to a file or to an ostream. argument toFile, os : file or stream on which settings are written. argument writeAll (default = off): normally only settings that have been changed are written, but if true then all settings are output. **Note:** the method returns false if it fails.

bool Settings::writeFileXML(ostream& os = cout) write out the information stored in xmldoc to be used later to initialize Settings through an input stream.

So far no equivalent for ParticleData (but not hard to implement?)

And of course you still need main program + any UserHooks (+ ext libraries) etc you link.

Feedback on Questions from ATLAS



Could we get a summary of what has happened in the API?

That might allow us to simplify/standardise our interface code, especially around all the special cases that are coded in there. Some interest in having an "ATLAS example" executable that does "the ATLAS things", but is that the wrong framing? The executable idea is fine, but we really want to have all our current treatments callable from the Pythia library, so both the standalone executable and the interface just consist of very minimal calling of API routines. Then it is easy to synchronise. Otherwise the "Athena ATLAS" and "Pythia ATLAS" modes will not be synced and will drift apart.

Sounds (to me) mostly like an ATLAS internal discussion / internal consistency?

No plans from our side to make significant changes to the Pythia "API" Question: is there something you cannot achieve with current structure?



RecoilToTop: Coherence in Top Decay: 2nd emission

Second (and subsequent) emissions in top decay

Not controlled by PowHeg, nor by Pythia's MECs.



tg RF antenna: Phase space & recoils set by: t - g = b + WCollective recoil

- Not as important as 1st em. Still highly significant if goal is per-mille precision on m_{r}



g - t dipole treated as g - b: Phase space & recoils set by **b** Affects *b* fragmentation





RecoilToTop

PYTHIA allows different coherence/recoil options in top decays

Recently made a dedicated UserHook "recoilToTop" (for use with recToCol = off) \rightarrow 8.310 in code! Theoretically the "least bad" option (in absence of Vincia-style RF antennae). Needs validations & feedback.



- g t dipole treated as g b:





What do the MECorrections actually do

- Are they also applicable for processes with offshell contributions like bb4l? There are three different options given in the manual: TimeShower:MEafterFirst, TimeShower:MEcorrections,
 - TimeShower:MEextended.
- Could you explain a bit what they do?

Feedback on Questions from ATLAS







Matrix-Element Corrections

Modify parton shower

Use splitting kernels \propto full matrix element for 1st emission:

Parton Shower
$$\frac{P(z)}{Q^2} \rightarrow \frac{P'(z)}{Q^2} = \frac{P(z)}{Q^2} \underbrace{\frac{|M_{n+1}|^2}{\sum_i P_i(z)/Q_i^2 |M_n|^2}}_{\text{MEC}}$$
 (suppressing \mathfrak{a}_s and Jacobian factors)

Proces

Done in PYTHIA for all SM decays and many BSM ones

Based on systematic classification of spin/colour structures

Difficult to generalise beyond one emission

Parton-shower expansions complicated & can have "dead zones" Achieved in VINCIA (by devising showers that have simple expansions)

Bengtsson, Sjöstrand, PLB 185 (1987) 435

Norrbin, Sjöstrand, NPB 603 (2001) 297

- (Also used to account for mass effects, and for a few simple hard processes like Drell-Yan.)

Giele, Kosower, Skands, PRD 84 (2011) 054003 Fischer et al, arXiv:1605.06142







(MECs with Loops: POWHEG)





Acronym stands for: **Po**sitive Weight Hardest Emission Generator.



PowHeg is widely applied/available, can be used with PYTHIA, HERWIG, SHERPA

Subtlety 1: Connecting with parton shower

Truncated Showers & Vetoed Showers

Subtlety 2: Avoiding (over)exponentiation of hard radiation

Controlled by "hFact" parameter (POWHEG)



TimeShower:MEcorrections (default = on) flaq Use of matrix element corrections where available; on/off = true/false.

TimeShower:MEextended (default = on) flaq processes that are implemented.

This should at least provide relevant mass dampening for massive radiators and recoilers.

Only has a meaning if MEcorrections above is switched on.

TimeShower:MEafterFirst (default = on) flag Use of matrix element corrections also after the first emission, for dipole ends of the same system that did not yet radiate. Only has a meaning if MEcorrections above is switched on.

Use matrix element corrections also for $1 \rightarrow n$ and $2 \rightarrow n$ processes where no matrix elements are encoded, by an attempt to match on to one of the $1 \rightarrow 2$



Pythia — Other Recent Activity and Plans

Perturbative Accuracy

- **PanScales** showers \leftrightarrow Pythia 8 (PZS to spend 1-year sabbatical at Oxford from August onwards)

Hadronisation and Tuning

- Colour Reconnections & String Junctions (QCD CR model being revisited, Dipole Swing) Strangeness (Close-Packing & Rope hadronization) QED corrections in hadron decays (esp B hadrons, viz PHOTOS) B decays (with EVTGEN, Monash-Warwick Alliance) Forward proton spectra NA61/Shine
- Automated Shower and Hadronisation Uncertainties

Heavy-Ion Physics, Collectivity, Cosmic Rays, Low-Energy Scattering

- Angantyr / Gleipnir: Pythia-based modelling of Heavy-Ion collisions without medium Hadronic Rescattering and Low-Energy Cross Sections EIC and Neutrino-Ion?
- Cosmic-Ray Air Showers (variable beams & energies)

VINCIA: NNLO matching, iterated ME corrections, $2 \rightarrow 4$ branchings, 2^{nd} -order kernels, recoil strategies, ...



Tuning: Some general comments. What do you want it to be?



Sensible

A set of physically sensible central parameter values, with good universality.

What does "physically sensible" and "good universality" really mean? Understanding MC models: hierarchies, universalities, and sensitivities.



Sophisticated

High-precision & specialised parameter sets, with reliable uncertainties

Tuning in the context of NⁿLO matching & precision applications. Theory uncertainties. Rigorous scientific analyses of parameter spaces.



A pure optimisation problem. The **best fit** you can get. Ask questions later.





Risky? Overfitting, oversimplification, GIGO, black-box syndrome, tunnel vision, loss of insight & scientific rigour, Tyranny of Carlo,...







Reliable Uncertainties and Preventing Overfitting

Monash Tune: 5% flat sanity-limit Theory Uncertainty to prevent overfitting

distributions with unknown correlations.

(Monash Tune was done by eye, so this was simply a matter of judgement.)

Use Pythia to map correlations between observables and incorporate in tuning?

Professor's eigentunes may be prone to artifacts of overtuning

See eg arXiv:1812.07424 for examples (and slightly more elaborate way to address issue but still fundamentally based on the flat 5% sanity limit)

There is still a need to develop reliable well-motivated uncertainty variations

Beyond "eigentunes" (Perugia had simple ones, Monash had none)

- Can this be improved on? Using better theory uncertainty estimates? & sensitivities? Would like TH uncertainties to get to ~ χ^2_{red} ~ 1. Not well-defined across multiple
- E.g., well-measured peak will dominate, with arbitrarily tiny uncertainties, at price of not spanning range in tails/asymptotics. Unclear interplay with genuine theory uncertainties.

- Ideally also propose *method* for how to obtain them, and justify or improve on the 5% approach.



Data Preservation: <u>HEPDATA</u>

Online database of experimental measurement results Please make sure all published results make it there

Analysis Preservation: <u>RIVET</u>

Large library of encoded analyses + data comparisons Main analysis & constraint package for event generators All your analysis are belong to RIVET

Updated validation plots: <u>MCPLOTS.CERN.CH</u>

Online plots made from Rivet analyses

Want to help? Connect to LHC@home project Test4Theory

Reproducible tuning: <u>PROFESSOR</u>, AUTOTUNES, APPRENTICE (& more?)

Automated tuning (& more)

Resources





Menu

→ Front Page

- → LHC@home 2.0
- → Generator Versions
- → Generator Validation
- → Update History
- → User Manual and Reference

Analysis filter:

→ ALL pp/ppbar
→ ALL ee
Specific analysis:
→[
→ Latest analyses

Z (Drell-Yan)

- → Jet Multiplicities
- $\rightarrow 1/\sigma d\sigma(Z)/d\phi_n^*$
- \rightarrow d σ (Z)/dpTZ
- $\rightarrow 1/\sigma d\sigma (\dot{Z})/dpTZ$

W

- → Charge asymmetry vs η
- → Charge asymmetry vs N_{iet}
- → dσ(jet)/dpT
- → Jet Multiplicities

Top (MC only)

- → Δφ (ttbar)
- $\rightarrow \Delta y$ (ttbar)
- \rightarrow | Δ y| (ttbar)
- → M (ttbar)
- → pT (ttbar)
- → Cross sections
- → y (ttbar)
- → Asymmetry
- → Individual tops

Bottom

- → η Distributions
- → pT Distributions
- → Cross sections

Underlying Event : TRNS : Σ(pT) vs pT1

 Generator Group:
 General-Purpose MCs
 Soft-Inclusive MCs
 Alpgen
 Herwig++
 Pythia 6
 Pythia 8
 Sherpa

 Subgroup:
 Defaults
 LHC Tunes
 C++ Generators
 Tevatron vs
 LHC tunes

pp @ 7000 GeV



meplots.cern.cl







22

Join us at LHC@home Test4Theory







Future Plans





Extra Slides

Parameters (in PYTHIA): FSR pQCD Parameters

Matching $a_s(m_Z)$ a_s Running

Additional Matrix Elements included?

At tree level / one-loop level? Using what matching scheme?

The value of the strong coupling

In PYTHIA, you set an effective value for $\alpha_s(m_Z^2) \Leftrightarrow$ choice of k in $\alpha_s(kp_\perp^2)$

Renormalization Scheme and Scale for α_s

1- vs 2-loop running, MSbar / CMW scheme, choice of k in $\alpha_s(kp_\perp^2)$, cf /

Subleading Logs

• • •

Ordering variable, coherence treatment, effective $1 \rightarrow 3$ (or $2 \rightarrow 4$), recoil strategy, ...

Branching Kinematics (z definitions, local vs global momentum conservation), hard parton starting scales / phase-space cutoffs, masses, non-singular terms,



Parameters (in PYTHIA): String Tuning

Hadron energy fractions



Fragmentation Function

p_⊤ in string breaks



Scale of string-breaking process Shower cutoff and $\langle p_{\perp} \rangle$ in string breaks

Meson Multiplets



Mesons

Baryon Multiplets

Baryons

colour reconnections (junctions), ... ?







Strangeness suppression, **Vector/Pseudoscalar**, η , η' , ...

Baryon-to-meson ratios, Spin-3/2 vs Spin-1/2, "popcorn",





Example: Effective Value of Strong Coupling

PYTHIA 8 (hadronization on) vs LEP: Thrust

$$T = \max_{\vec{n}} \left(\frac{\sum_{i} |\vec{p_i} \cdot \vec{n}|}{\sum_{i} |\vec{p_i}|} \right) \qquad 1 - 1$$



Using effective $\alpha_{s}(M_{Z}) = 0.12$



Example: Effective Value of Strong Coupling

PYTHIA 8 (hadronization on) vs LEP: Thrust

$$T = \max_{\vec{n}} \left(\frac{\sum_{i} |\vec{p_i} \cdot \vec{n}|}{\sum_{i} |\vec{p_i}|} \right) \qquad 1 - 1$$



Using effective $\alpha_s(M_Z) = 0.14$



Wait ... is this Crazy?

Best result

Obtained with $\alpha_s(M_Z) \approx 0.14$

≠ World Average ~ 0.118

Effective value of α_{s} depends on the order and scheme Baseline MC \approx Leading Order + LL resummation Other leading-Order extractions of $\alpha_{s} \approx 0.13 - 0.14$ Effective scheme interpreted as "CMW" \rightarrow 0.13 2-loop running $\rightarrow 0.127$; NNLO Matching $\rightarrow 0.12$ Hartgring, Laenen, PZS, JHEP 10 (2013) 127; see also backup slides

Not so crazy (but does rely on "magic" mathematical accident in Z decay) Let parameters vary to a level consistent with the (limited) formal accuracy. Sanity check = consistency with other determinations at a similar formal order, within the uncertainty at that order (including a CMW-like scheme redefinition to go to 'MC scheme')



Catani, Marchesini, Webber, Nucl.Phys.B 349 (1991) 635-654

To improve systematically \rightarrow Merging at NLO



Example 2: Sensitivity to Hadronization Parameters

PYTHIA 8 (hadronization on) Vs (hadronization off)

Important point: These observables are IR safe -> minimal hadronisation corrections Big differences in how sensitive each of these are to hadronisation & over what range



Large sensitivities to "lower" phenomena break the divide-and-conquer simplification.

Another **important** point: **peaks** of distributions are all where **HAD** sensitivity is highest!

University



31

... and sensitivity to fixed-order corrections



These points are relatively insensitive to **both** hadronization **and** matching/merging

(Adding nuisance terms $\Delta P(z) \propto Q^2$ to the splitting kernels beyond shower accuracy)



Hadronization Corrections: Fragmentation Tuning

Now use infrared sensitive observables - sensitive to hadronization + first few bins of previous (IR safe) ones



Tutorial

Monash 🧑 University



Fragmentation Tuning



If treated like a black box, we could tune the shape of the momentum spectrum solely by modifying eg the



34

Identified Particles

V/P, B/M, B_{3/2}/B_{1/2}, strange/unstrange, Heavy, ...



Could be completely mistured if looking **only** at inclusive charged $\ln(x)$ spectrum

Point: include observables with **direct** sensitivity to each parameter you include.

Plenty of observables have **direct** sensitivity to strangeness (& other PID) fractions

Monash 🔬 University



GIGO



Large changes in strangeness or vector/pseudoscalar ratios do modify the momentum spectrum



At the cost of totally destroying agreement with observables that are directly sensitive to those parameters



Parameters (in PYTHIA): Initial-State Radiaton



At tree level / one-loop level? What matching scheme?



Relation between Q_{PS} and Q_{F} (Vetoed showers? Suppressed? cf matching)

I-F colour-flow interference effects (eg VBF & Tevatron $t\bar{t}$ asym) & interleaving

A small additional amount of "unresolved" kT

Fermi motion + unresolved ISR emissions + low-x effects?







Drell-Yan pT distribution



ISR + Primordial kT



Note: Q.M. requires physical observable!

University



Beware Process Dependence!





→ we should ensure we do MECs / matching / merging if we want to use them (or something equivalent to that.)



Number of MPI



interactions \rightarrow size of overall activity





Strings per Interaction





- Infrared Regularization scale p_{+0} for the QCD 2→2 (Rutherford) scatterings used for multiple parton
- Note: strongly correlated with choice of PDF set! (low-x gluon)
- Proton transverse mass distribution \rightarrow difference between central (more active) vs peripheral (less active) collisions
- **Color correlations between multiple-parton-interaction** systems (aka colour reconnections — relative to LC)
 - \rightarrow shorter or longer strings \rightarrow less or more hadrons per interaction
- **Evolution of UE**, $\langle dN/d\eta \rangle$, ... with collider CM energy Cast as energy evolution of p_{T0} parameter.





Bad Example: Why dN/dŋ is useless (by itself)

$\langle dN_{ch}/d\eta \rangle$ often used as main constraint on models of minimum-bias physics



Can get right $\langle N_{ch} \rangle$ with completely wrong

Another interesting observable is the forward-backward elation, defined in the following way. Consider two sin pseudorapidity: one between $\Delta \eta/2$ and $\Delta \eta/2 + 1$ rd) the other between $-\Delta n/2$ and $-(\Delta n/2 \pm 1)$

of what is included up to this point, even rather drastic variations are insufficient to come anywhere near an explanation of the data.

41



Relative size of this plateau / min-bias depends on pT0, PDF, and b-profile

As you trigger on progressively higher p_T , the entire event increases ...

... until you reach a plateau ("max-bias") also called the "jet pedestal" effect **Interpreted as impact-parameter effect** Qualitatively reproduced by MPI models







Branching fractions and decay modelling



Collective Effects (in pp)

Colour Reconnections (& effects on precision measurements like m_{top}) Strangeness Enhancements (eg close-packing, ropes, ...) Flow-like effects (eg close-packing, string shoving, ...)



Forward Physics

 \bullet \bullet \bullet

Beam-Remnant Handling Diffractive Modelling (incl hard diffraction, Pomeron substructure) Total and Elastic Cross-section parametrisatons



