

Modelling and tuning in top quark physics

Marek Schönherr

CERN

TOP2018 – 17 Sep 2018



Monte-Carlo usage

Monte-Carlo Event Generators generally used in two different ways

Theory prediction

- parameters in perturbative regime dictated by first principles or theory bias
- non-perturbative parameters determined universally in well-defined limited set of observables (same as PDFs)

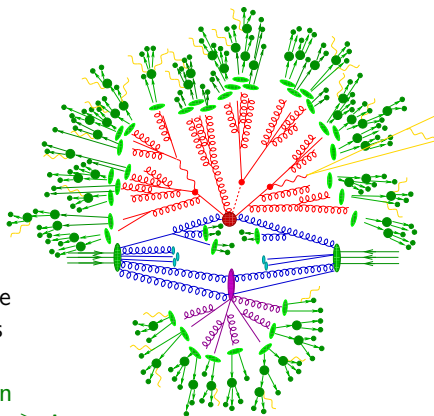
Data modelling

- all available parameters tuned to best reproduce measured data of specific process
- resulting distribution not predictive, but very useful to determine acceptances, efficiencies, etc.

Both valid use cases, but need to be aware which one one is using

Tuning strategies

- MCs build by factorising collisions into different stages with characteristic energy regimes
 - ⇒ hard scattering, parton evolution, proton frag., multiple interactions, hadronisation, hadron decays
- this factorisation also means, each stage independent of the details of the others
 - eg. hadronisation only dependent on colours, flavours & momenta of parton ensemble at a relative separation $\sim t_{IR} \gtrsim \Lambda_{QCD}$
- each stage is tuned as much individually as possible
 - hadron decay parameters fitted to decay data from b -/ c -factories, etc.
 - hadronisation tuned to e^+e^- data at various energies (b -fac., SLD, LEP)
 - multiple interactions, beam remnants, etc. tuned to hadron collider data



General purpose Monte-Carlo Event Generators

	HERWIG 7	PYTHIA 8	SHERPA
PS	\tilde{q} -Shower Dipole Shower	Default Parton Shower DIRE VINCIA	CSSHOWER DIRE
MI	soft gluon model & hard scat. model (JIMMY-based)	sophisticated interleaved model	old PYTHIA-style non-interleaved
Had	Cluster Interface to Lund String	Lund String	Cluster* Interface to Lund String
MB	MinBias	MinBias	–

Tuning strategies

well documented Skands, Carrazza, Rojo in [arXiv:1404.5630](https://arxiv.org/abs/1404.5630)

Generally use χ^2 minimisation with

$$\chi^2(\vec{x}) = \frac{1}{N} \sum_{i \in \mathcal{O}} w_i \frac{(\text{MC}_i(\vec{x}) - \text{Data}_i)^2}{\sigma_{i,\text{Data}}^2}$$

with parameter set \vec{x} and weights w_i to determine the relative importance of observable/bin on tune.

Very much automated using PROFESSOR.

[Buckley et.al. arXiv:0907.2973](https://arxiv.org/abs/0907.2973)

Constrained by run-time of the generator.

Thus, most observables tuned at LOPS (incl. ME corrections).

In SHERPA, some observables at MEPS (LO multijet merging with low jet multiplicity).

Tune must be restricted to observables that are reliably described through LOPS.

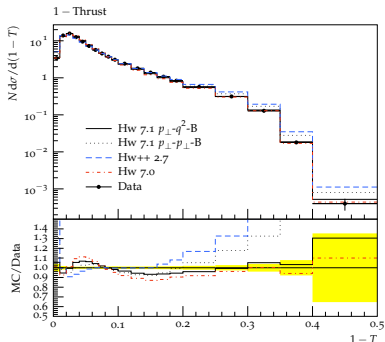
Tuning FSR and hadronisation parameters

General strategy

- 1 tune light quark parameters
- 2 tune b -quark parameters
- 3 tune c -quark parameters
(c -meson abundances receive large contribs from b rates)
- 4 retune light quark parameters in light of new b/c params

models generally inadequate

→ balance observables



Data from SLD, LEP, JADE, TASSO, b -fac., etc. used
Colour-reconnection generally not needed in e^+e^-

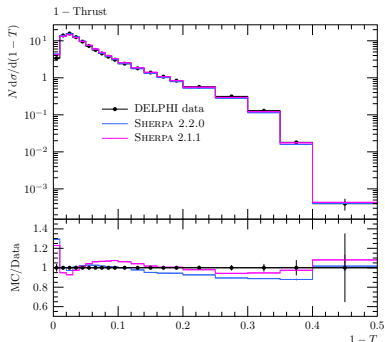
Tuning FSR and hadronisation parameters

General strategy

- ① tune light quark parameters
- ② tune b -quark parameters
- ③ tune c -quark parameters
(c -meson abundances receive large contribs from b rates)
- ④ retune light quark parameters in light of new b/c params

models generally inadequate

→ balance observables



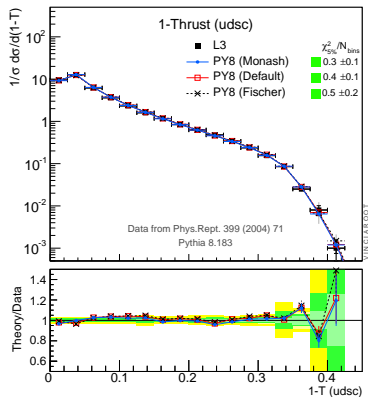
Data from SLD, LEP, JADE, TASSO, b -fac., etc. used
 Colour-reconnection generally not needed in e^+e^-

Tuning FSR and hadronisation parameters

General strategy

- ① tune light quark parameters
- ② tune b -quark parameters
- ③ tune c -quark parameters
(c -meson abundances receive large contribs from b rates)
- ④ retune light quark parameters
in light of new b/c params

models generally inadequate
→ balance observables



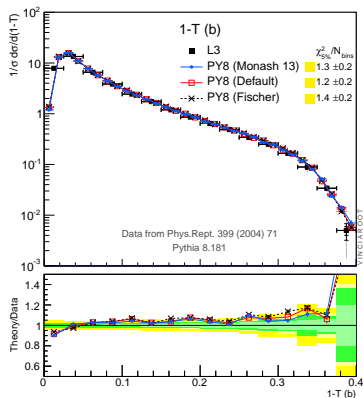
Data from SLD, LEP, JADE, TASSO, b -fac., etc. used
Colour-reconnection generally not needed in e^+e^-

Tuning FSR and hadronisation parameters

General strategy

- ① tune light quark parameters
- ② tune b -quark parameters
- ③ tune c -quark parameters
(c -meson abundances receive large contribs from b rates)
- ④ retune light quark parameters
in light of new b/c params

models generally inadequate
→ balance observables



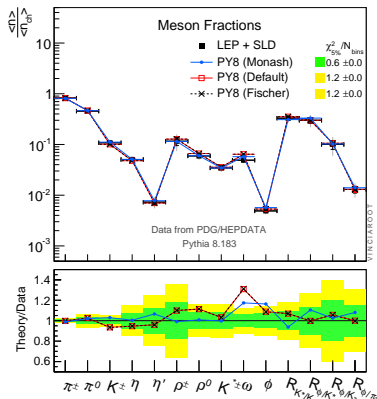
Data from SLD, LEP, JADE, TASSO, b -fac., etc. used
Colour-reconnection generally not needed in e^+e^-

Tuning FSR and hadronisation parameters

General strategy

- ① tune light quark parameters
- ② tune b -quark parameters
- ③ tune c -quark parameters
(c -meson abundances receive large contribs from b rates)
- ④ retune light quark parameters
in light of new b/c params

models generally inadequate
→ balance observables



Data from SLD, LEP, JADE, TASSO, b -fac., etc. used
Colour-reconnection generally not needed in e^+e^-

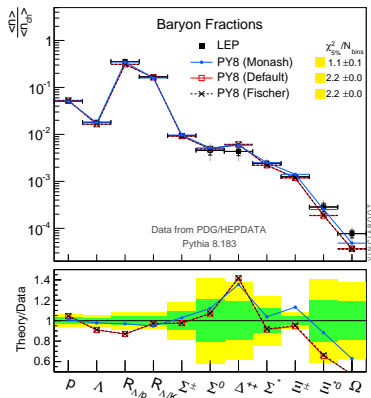
Tuning FSR and hadronisation parameters

General strategy

- 1 tune light quark parameters
- 2 tune b -quark parameters
- 3 tune c -quark parameters
(c -meson abundances receive large contribs from b rates)
- 4 retune light quark parameters
in light of new b/c params

models generally inadequate
→ balance observables

Data from SLD, LEP, JADE, TASSO, b -fac., etc. used
Colour-reconnection generally not needed in e^+e^-



Tuning underlying event, beam remnants, etc.

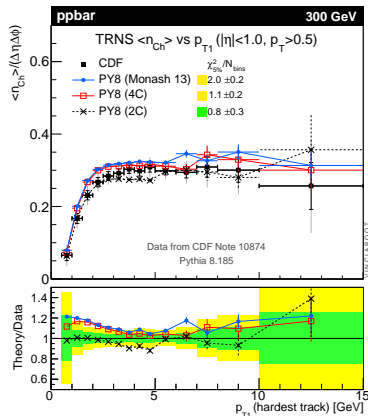
General strategy

- 1 tune intrinsic transverse momentum to first few bins in $p_T^{e^+e^-}$ distribution in DY
- 2 tune UE, beam remnants and colour reconnection using underlying event observables

Use data from UA5, Tevatron & LHC at various collider energies

UE in leading jet and DY.

No tune by generator authors currently uses top data.



Some observables receive sizable contriubs from multijet final states
 \Rightarrow **only include in tune if modelled properly**



Tuning underlying event, beam remnants, etc.

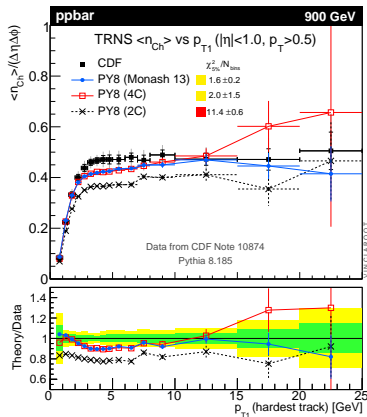
General strategy

- 1 tune intrinsic transverse momentum to first few bins in $p_T^{e^+e^-}$ distribution in DY
- 2 tune UE, beam remnants and colour reconnection using underlying event observables

Use data from UA5, Tevatron & LHC at various collider energies

UE in leading jet and DY.

No tune by generator authors currently uses top data.



Some observables receive sizable contribs from multijet final states
 \Rightarrow **only include in tune if modelled properly**

Tuning underlying event, beam remnants, etc.

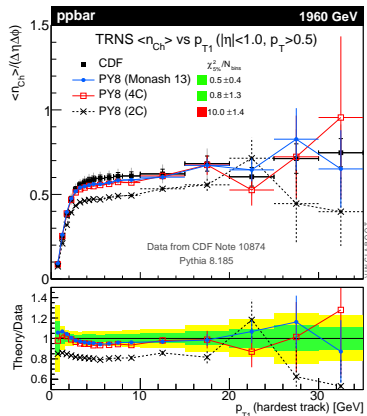
General strategy

- 1 tune intrinsic transverse momentum to first few bins in $p_T^{e^+e^-}$ distribution in DY
- 2 tune UE, beam remnants and colour reconnection using underlying event observables

Use data from UA5, Tevatron & LHC at various collider energies

UE in leading jet and DY.

No tune by generator authors currently uses top data.



Some observables receive sizable contribs from multijet final states
 \Rightarrow **only include in tune if modelled properly**

Tuning underlying event, beam remnants, etc.

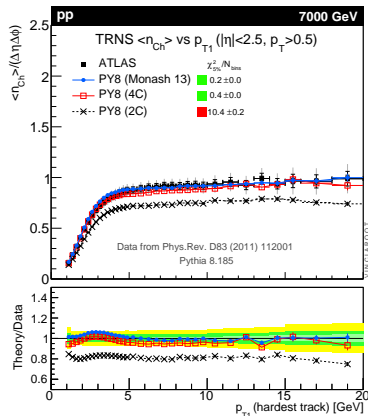
General strategy

- 1 tune intrinsic transverse momentum to first few bins in $p_T^{e^+e^-}$ distribution in DY
- 2 tune UE, beam remnants and colour reconnection using underlying event observables

Use data from UA5, Tevatron & LHC at various collider energies

UE in leading jet and DY.

No tune by generator authors currently uses top data.



**Some observables receive sizable contriubs from multijet final states
 ⇒ only include in tune if modelled properly**



Tuning underlying event, beam remnants, etc.

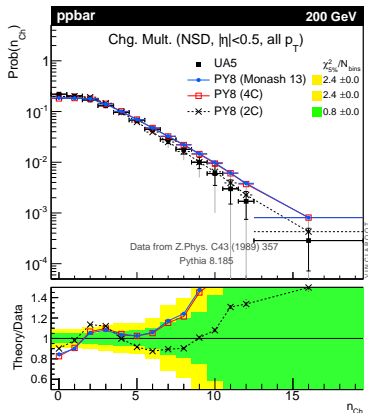
General strategy

- 1 tune intrinsic transverse momentum to first few bins in $p_T^{e^+e^-}$ distribution in DY
- 2 tune UE, beam remnants and colour reconnection using underlying event observables

Use data from UA5, Tevatron & LHC at various collider energies

UE in leading jet and DY.

No tune by generator authors currently uses top data.



Some observables receive sizable contriubs from multijet final states
 \Rightarrow **only include in tune if modelled properly**

Tuning underlying event, beam remnants, etc.

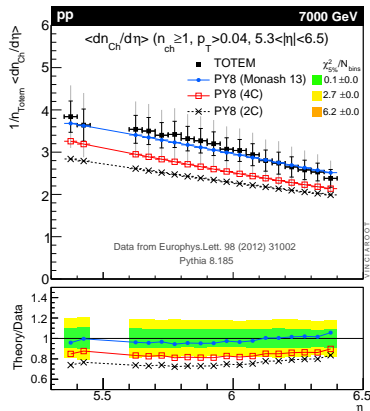
General strategy

- 1 tune intrinsic transverse momentum to first few bins in $p_T^{e^+e^-}$ distribution in DY
- 2 tune UE, beam remnants and colour reconnection using underlying event observables

Use data from UA5, Tevatron & LHC at various collider energies

UE in leading jet and DY.

No tune by generator authors currently uses top data.



**Some observables receive sizable contribs from multijet final states
 \Rightarrow only include in tune if modelled properly**



Tuning underlying event, beam remnants, etc.

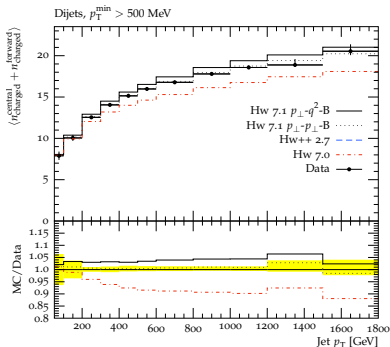
General strategy

- 1 tune intrinsic transverse momentum to first few bins in $p_T^{\ell^+\ell^-}$ distribution in DY
- 2 tune UE, beam remnants and colour reconnection using underlying event observables

Use data from UA5, Tevatron & LHC at various collider energies

UE in leading jet and DY.

No tune by generator authors currently uses top data.



Some observables receive sizable contribs from multijet final states
⇒ only include in tune if modelled properly

Tuning underlying event, beam remnants, etc.

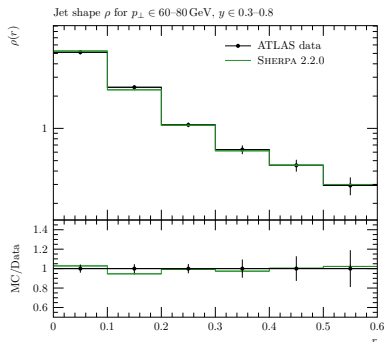
General strategy

- 1 tune intrinsic transverse momentum to first few bins in $p_T^{\ell^+\ell^-}$ distribution in DY
- 2 tune UE, beam remnants and colour reconnection using underlying event observables

Use data from UA5, Tevatron & LHC at various collider energies

UE in leading jet and DY.

No tune by generator authors currently uses top data.



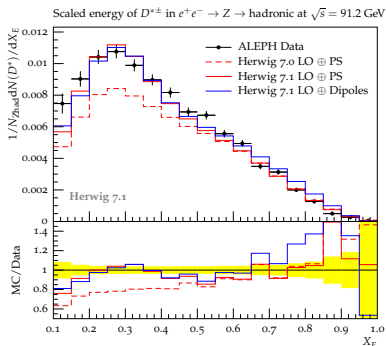
**Some observables receive sizable contrihs from multijet final states
 \Rightarrow only include in tune if modelled properly**



Better modelling trumps tuning

Dichotomy of MC gens

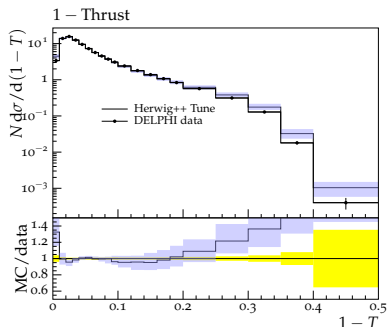
- all models are incomplete, most implementations incomplete
 - yet address everything as universally as possible
 - example: hadron decays
 - > 100 hadrons
 - > 2000 decay channels
- internal hadron decay packages have to address all of them to a reasonable accuracy



replace internal implementation by interface to EVTGEN

Not everything that can be fixed by tuning should be fixed by tuning.

Uncertainties



Tune uncertainties

Tune performed by minimising

$$\chi^2(\vec{x}) = \frac{1}{N} \sum_{i \in \mathcal{O}} w_b \frac{(\text{MC}_i(\vec{x}) - \text{Data}_i)^2}{\sigma_{i,\text{Data}}^2}$$

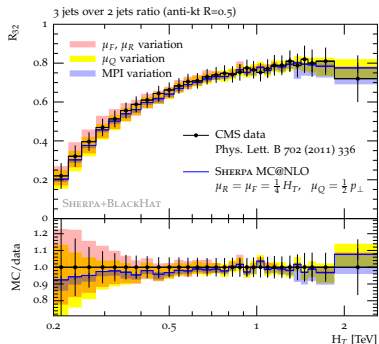
Define **Eigentunes** as set of eigen-vectors in χ^2 potential, normalised to some predefined $\Delta\chi^2$ (same as PDF)

Large parameter space, possible reduction for specific observables

Model uncertainties:

Systematic model variation, similar to what is done in the Perugia tunes

Uncertainties



Tune uncertainties

Tune performed by minimising

$$\chi^2(\vec{x}) = \frac{1}{N} \sum_{i \in \mathcal{O}} w_b \frac{(\text{MC}_i(\vec{x}) - \text{Data}_i)^2}{\sigma_{i,\text{Data}}^2}$$

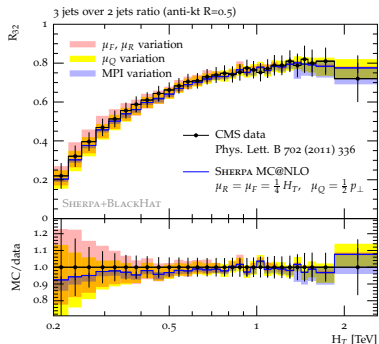
Define **Eigentunes** as set of eigen-vectors in χ^2 potential, normalised to some predefined $\Delta\chi^2$ (same as PDF)

Large parameter space, possible reduction for specific observables

Model uncertainties:

Systematic model variation, similar to what is done in the Perugia tunes

Uncertainties



Model uncertainties:

Systematic model variation, similar to what is done in the Perugia tunes

Tune uncertainties

Tune performed by minimising

$$\chi^2(\vec{x}) = \frac{1}{N} \sum_{i \in \mathcal{O}} w_b \frac{(\text{MC}_i(\vec{x}) - \text{Data}_i)^2}{\sigma_{i,\text{Data}}^2}$$

Define **Eigentunes** as set of eigen-vectors in χ^2 potential, normalised to some predefined $\Delta\chi^2$ (same as PDF)

Large parameter space, possible reduction for specific observables

Combination of LO tunes with state-of-the-art calculations

Care must be taken when using combining (N)NLO partonic calculations with LO tuned MC.

Example:

- (N)NLO calculation require (N)NLO PDF.
- Consistency of logarithmic resummation requires continuous evolution of α_s and PDF across scales, ie. tools.
In practice: same PDF in ME and PS.
- UE models require LO PDF for probabilistic interpretation
- Proton sum rules require consistency of PDF across primary and multiple interaction. In practice: same PDF in UE and PS.

⇒ **compromise, depending on relevance.**

All perturbative parameters (μ_R , μ_F , $t_{\text{shower}}^{\text{start}}$, h_{damp} , etc.) should be put to the theoretically preferred value and then varied by a conventional amount to arrive at an uncertainty.

Influence on top quark observables

Top quark processes are unique, in that they

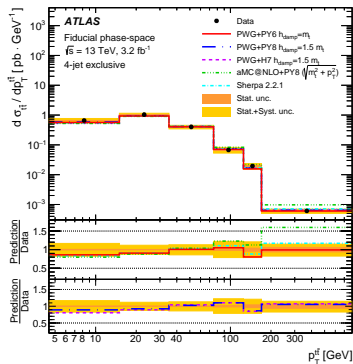
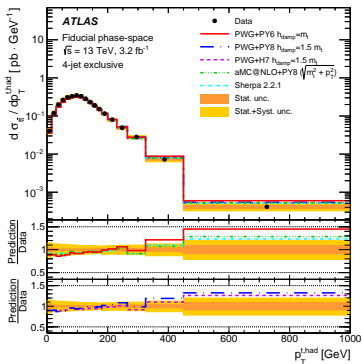
- always b -quarks involved
 - almost always gluons involved
- ⇒ both not as constrained as light quarks

What to do?

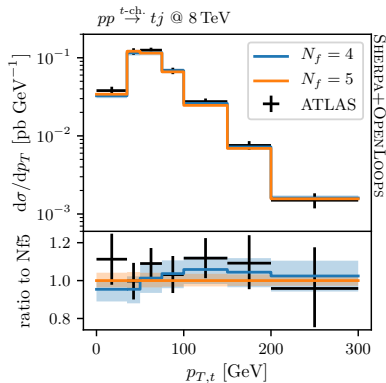
- include top quark processes in tuning?
 - lose predictivity for all observables tuned to
 - + optimal reproduction of data
- find proxies where to constrain models for g and b related parameters better!
 - under way: $\Delta R(B, B)$ measurement for $t\bar{t}b\bar{b}$ observables
 - gluon jet data at LEP and LHC to constrain gluon jet fragmentation
 - etc.

General top quark observables

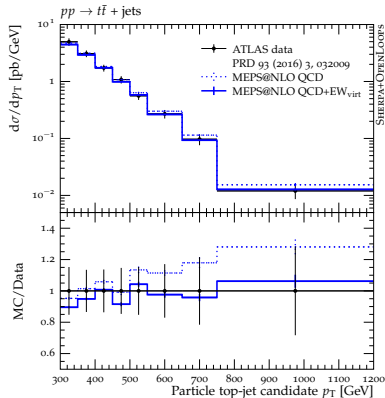
General top quark observables not driven by secondary b -quark production or precise details of non-perturbative effects. These are well described by the current models with the current tunes.



Example: Single-top and $t\bar{t}$



MC@NLO $pp \rightarrow t\bar{t}$



multijet merged

$pp \rightarrow t\bar{t} + 0, 1j @ \text{NLO} + 2, 3, 4 @ \text{LO}$

incl. approx. EW corrections

Conclusions

- It is important to clearly distinguish the (at least) two types of Monte-Carlo Event Generator usage:
Theory prediction vs. data reproduction
- Tunes by generator authors currently do not involve top quark data. MC calculations with these tunes can be treated as predictions for all top quark observables.
- Mismodelling not necessarily related to need for more tuning. Always investigate impact of better theory input.
- All perturbative parameters (μ_R , μ_F , $t_{\text{shower}}^{\text{start}}$, h_{damp} , etc.) should be put to the theoretically preferred value and then varied by a conventional amount to arrive at an uncertainty.

Thank you for your attention!