

Energy scaling of minimum-bias tunes

H. Schulz¹, P.Z. Skands^{2,a}¹Institut für Physik, Humboldt-Universität zu Berlin, Berlin, Germany²Theoretical Physics, CERN, 1211 Geneva 23, Switzerland

Received: 12 April 2011 / Published online: 10 May 2011

© The Author(s) 2011. This article is published with open access at Springerlink.com

Abstract We propose that the flexibility offered by modern event-generator tuning tools allows for more than just obtaining “best fits” to a collection of data. In particular, we argue that the universality of the underlying physics model can be tested by performing several, mutually independent, optimizations of the generator parameters in different physical regions. For regions in which these optimizations return similar and self-consistent parameter values, the model can be considered universal. Deviations from this behavior can be associated with a breakdown of the modeling, with the nature of the deviations giving clues as to the nature of the breakdown. We apply this procedure to study the energy scaling of a class of minimum-bias models based on multiple parton interactions (MPI) and p_{\perp} -ordered showers, implemented in the PYTHIA 6.4 generator. We find that a parameter controlling the strength of color reconnections in the final state is the most important source of non-universality in this model.

1 Introduction

The main virtue of general-purpose Monte Carlo event generators (sometimes called “shower” Monte Carlos, although they are normally relied on for many other physics aspects as well) is their ability to provide a complete and fully differential picture of collider final states, down to the level of individual particles. This allows them to be used as detailed—albeit approximate—theoretical references for measurements performed at accelerators like the LHC, against which models of both known and ‘new’ physics can be tested. The achievable accuracy depends both on the inclusiveness of the chosen observable (with more inclusive observables generally being more precisely predicted) and on the sophistication of the simulation itself. An important driver for the

latter is obviously the development of improved theoretical models, e.g., by including matching to higher-order matrix elements, more accurate resummations, or better non-perturbative models; but it also depends crucially on the available constraints on the remaining free parameters of the model. Using existing data to constrain these is referred to as generator tuning.

The recent minimum-bias measurements from the LHC experiments, in particular those at center-of-mass energy $\sqrt{s} = 7$ TeV, have highlighted the question to what extent the energy scaling of total cross sections and differential distributions are consistent with model-based extrapolations from lower energies, or whether they exhibit any non-trivial departures from such predictions. Most of the LHC collaborations have already gone some way towards studying this question, by including comparisons of specific models and tunes to the data in their publications. In the short term, such comparisons are useful both as immediate tests of commonly used models, and to illustrate the current amount of theoretical uncertainty surrounding a particular distribution. They also provide a set of well-defined theoretical reference curves for future studies. However, the conclusions that can be drawn from comparisons of individual tunes of specific models on single distributions are necessarily limited. In order to obtain more general conclusions, a more coherent and over-arching look at both the data and the models is needed.

In this study, we shall make use of the PROFESSOR tuning tool [1] to provide such a look. Specifically, rather than performing one global fit to all the data, as is usually done, we instead use PROFESSOR to perform several independent optimizations of the model parameters for a range of different collider energies. At each energy, we use the same set of minimum-bias observables, modulo the limitations imposed by different detector acceptances, trigger conditions, and correction procedures. We thereby seek to obtain a data-driven map of the preferred energy dependence of each of the tuned parameters. This can then be compared to the functional dependence assumed by the underlying model,

^a e-mail: peter.skands@cern.ch

thereby furnishing not just a “best fit” of the model parameters, but also a consistency check on the universality of the underlying physics model itself.

We emphasize that this sort of consistency check is not limited to energy scaling alone. In all generality, a consistency check on the underlying physics model can be obtained by performing independent optimizations in any two (or more) “physics windows” that the modeling provides or assumes relations between. Moreover, with the recent advent of automated tuning tools, it is now becoming possible to explore many such independent optimizations with only modest investments of computing and manpower. In regions in which consistent parameter sets are obtained, with predictions that are acceptably close to the data, the model can be considered as interpolating well, i.e., it is universal. If not, a breakdown in the ability of the model ability to span different physical regimes has been identified, and can be addressed.

For simplicity, we concentrate on one particular model here, the ‘new’ interleaved multiple-interactions model [2, 3] implemented in the PYTHIA 6 event generator [4]. All parameters not explicitly subjected to optimization in this study are those of the ‘Perugia 0’ tune [5, 6]. We note that this model has significant similarities with, but is not identical to, the one implemented in PYTHIA 8 [7].

In Sect. 2, we give brief overviews of the theory model, the PROFESSOR/RIVET tuning framework, and the data sets we have used to do the tuning. Section 3 contains our main study of the energy scaling of three main model parameters at currently existing colliders. We round off with conclusions and outlook in Sect. 4.

2 Setup

In this section, we briefly describe the theoretical models we take as starting points, emphasizing in particular the parameters relevant to this study (Sect. 2.1), the tuning framework we will be using (Sect. 2.2), and the data sets that have been included (Sect. 2.3).

2.1 Theoretical model

As mentioned above, we consider the interleaved model of p_\perp -ordered showers and multiple parton–parton interactions of [2, 3], as implemented in PYTHIA 6.4.23 [4], specifically the ‘Perugia 0’ tune of that model [6] unless otherwise specified.

In this class of models, pioneered by [8], a unified approach is taken to the modeling of all inelastic non-diffractive events, in which dijet production and its associated underlying event (UE) is viewed merely as the hard high- p_\perp tail of minimum-bias (MB), without any sharp

modeling distinction between the two. The fundamental building block for the model is the dijet cross section, computed at leading order in perturbative QCD, whose dominant component is simple low- p_\perp Rutherford scattering (t -channel gluon exchange).

For $p_\perp \rightarrow 0$, i.e., when two partons scatter via the exchange of a very soft gluon, this cross section exhibits a divergence whose ultimate origin is similar to that of initial-state bremsstrahlung. And similarly to what is done for bremsstrahlung, the model recasts this divergence in terms of a unitarized (Sudakov-suppressed) finite cross section for the hardest scattering accompanied by a divergent number of successively softer multiple parton interactions (MPI), a number which is ultimately regulated by the introduction of an infrared regularization scale. We do not intend to give a full account of unitarization here, but instead refer the reader to [2, 4, 8] for details on it in the context of the MPI models here discussed. We also note that an interesting exploration of the relation between this effective scale and perturbative low- x physics was recently carried out in [9]. Finally, a pedagogical and more general discussion of underlying-event models in general-purpose Monte Carlo generators can be found in the recent review [10].

In this study, we focus on three main parameters of the resulting type of model: the infrared regularization scale, the proton transverse mass distribution, and the color-reconnection strength, as follows.

2.1.1 Infrared regularization scale

The fact that long-wavelength gluons only see a coherent sum of the color charges in the hadronic substructure—color screening—is assumed to ultimately regulate the divergent number of parton–parton scatterings, similarly to how the non-perturbative cutoff in parton showers regulates the number of parton shower emissions. In the model we consider here, a smooth regulator is introduced, by modifying the divergent parts of the cross section (including the strong coupling since we use the standard MC scale choice $\alpha_s(p_\perp^2)$) as follows,

$$\alpha_s(p_\perp^2) \frac{dp_\perp^2}{p_\perp^4} \rightarrow \alpha_s(p_{\perp 0}^2 + p_\perp^2) \frac{dp_\perp^2}{(p_{\perp 0}^2 + p_\perp^2)^2}, \quad (1)$$

where $p_{\perp 0}$ physically expresses the scale at which the color screening effect is supposed to become active. This parameter, which we call the *infrared regularization scale*, constitutes the main free parameter for all models of this type, with low values yielding more soft MPI activity (in the limit that it is taken to zero, the original unregulated behavior would be reobtained). In the PYTHIA model, it is assumed to have a power-law scaling with the CM energy,

$$p_{\perp 0}(\sqrt{s}) = \text{PARP}(82) \cdot \left(\frac{\sqrt{s}}{\text{PARP}(89)} \right)^{\text{PARP}(90)}, \quad (2)$$

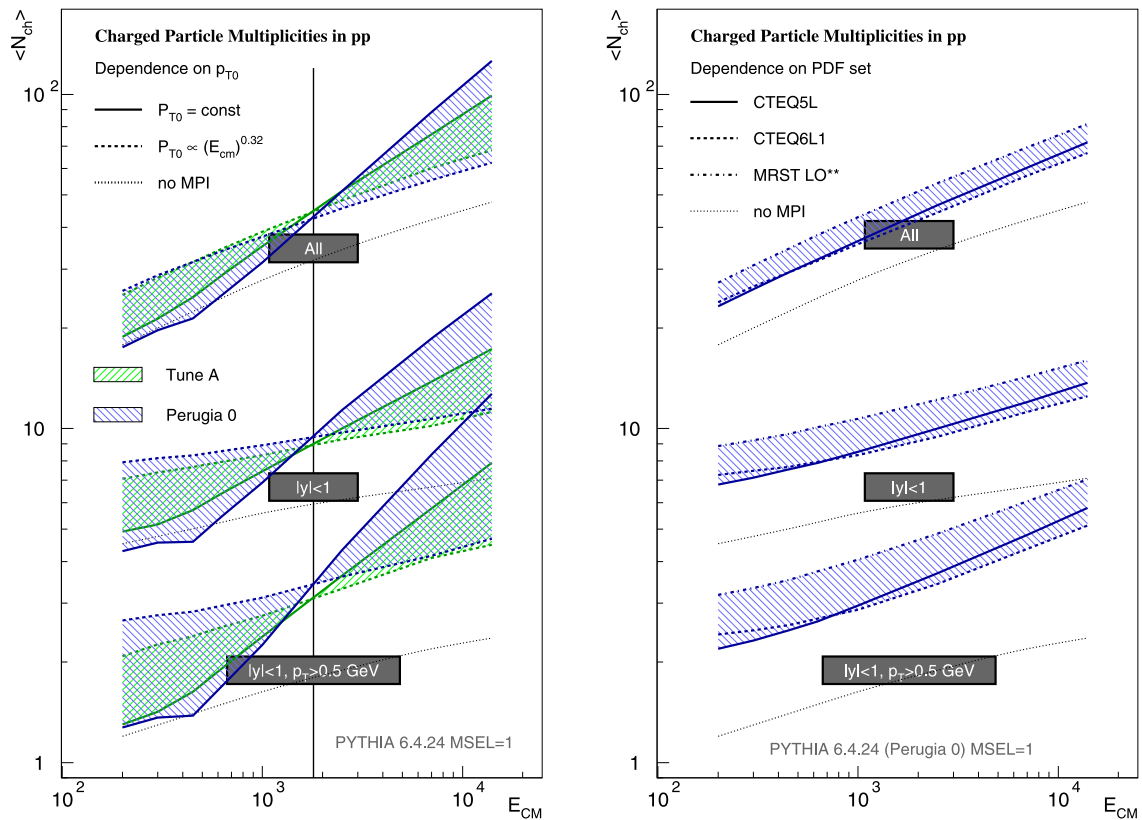


Fig. 1 Energy scaling of charged-particle multiplicities in pp in three different phase space regions (*top*: inclusive, *middle*: central, *bottom*: central hard). *Left*: Dependence on the scaling of the $p_{\perp 0}$ parameter for two different PYTHIA models, represented by Tune A and Perugia 0,

respectively. The *solid vertical line* represents the reference energy, 1800 GeV, at which PARP(82) is defined for both models. *Right*: Dependence on the PDF set, for the Perugia 0 model. For reference, Tune A without MPI is also shown (*dotted lines*)

where PARP(82), PARP(89), and PARP(90) are tunable parameters. Roughly speaking, PARP(82) gives the value of $p_{\perp 0}$ (in GeV) at a fixed reference CM energy = PARP(89) (also in GeV), and PARP(90) determines the scaling behavior of $p_{\perp 0}$ away from that energy. Below, instead of assuming the form, (2), we shall fit for $p_{\perp 0}$ independently at several different values of \sqrt{s} . (Technically, we do this by fixing PARP(89) to the energy of the relevant collider and fitting for PARP(82) which can then be interpreted directly as $p_{\perp 0}$ at that energy.) We can then check whether the resulting points lie on a curve that is consistent with the functional form of (2) or not.

To give the reader a more concrete idea of the dependence of the overall event activity on the assumed scaling form, the left-hand pane of Fig. 1 illustrates the scaling behavior of charged-particle multiplicities in non-diffractive minimum-bias events,¹ for two different assumptions of the energy

scaling of the $p_{\perp 0}$ parameter which we consider comparatively extreme:

1. Solid lines: constant $p_{\perp 0}$, i.e., PARP(90) = 0.0, resulting in a very fast growth of multiplicity with energy.
2. Dashed lines: PARP(90) = 0.32, i.e., $p_{\perp 0}$ varying as $(\sqrt{s})^{0.32}$, resulting in a multiplicity growth with energy which is comparable to the case without MPI, shown with dotted lines.

For completeness, both of the two PYTHIA min-bias models are included here, represented by Tune A and Perugia 0, respectively. (For reference, the default for Tune A is a scaling power of 0.25. For Perugia 0, it is 0.26.) In both cases, three phase space regions are shown: inclusive (top), central (middle), and central hard (bottom), with phase space cuts as indicated in the gray shaded boxes. Comparing model curves with equal scaling assumptions (solid with solid and dashed with dashed), it is evident that the two models have somewhat different intrinsic scaling properties, even with the same assumption for the scaling of $p_{\perp 0}$. Thus, the value and scaling of $p_{\perp 0}$ alone is not sufficient to fix the energy scaling completely.

¹Specifically, the generated events correspond to running PYTHIA in its “minimum-bias” mode with diffraction switched off. We permit ourselves this somewhat unphysical definition here, since the illustration is intended for qualitative purposes only. (For the numerical studies later in this report, we use a full inelastic sample that includes diffraction.)

It is also worth noting that both models require *at least one* partonic scattering per hadron–hadron collision, hence it is not possible for the average multiplicity to drop below the “no-MPI” case. This causes a rather abrupt change in the scaling behavior at low energies for those curves that intersect the no-MPI (dotted) one.

On the right-hand pane of Fig. 1, we illustrate another important dependence, on the PDF set used. Both Tune A and Perugia 0 were made using the CTEQ5L set [11] (solid lines). Changing to either CTEQ6L1 [12] (dashed lines) or MRST LO** [13] (dot-dashed lines) affects both the value of the average multiplicities as well as their scaling behavior and distribution in phase space. In particular, the LO** set generates a significantly larger activity than its CTEQ cousins. It is therefore not generally possible to separate the choice of PDF set from the $p_{\perp 0}$ choice.

2.1.2 Transverse mass distribution

A further important aspect of the model is the shape of the assumed proton matter distribution. In MPI models, the probability for additional parton–parton interactions to occur in a given collision is proportional to the amount of matter overlap between the colliding beam particles in that collision, which in turn depends on their impact parameter, b . If the proton structure is very uniform (e.g., a featureless pion/gluon cloud), the differences between peripheral and central collisions will be quite small, while a strongly peaked distribution (e.g., valence lumps/hot spots) can make the activity in central collisions much higher than in peripheral ones. Thus, while we may think of the infrared regularization scale above as determining the *average* number of multiple parton interactions, the b profile affects how much this number can *deviate* from the mean in peripheral vs. central events. In the overlap model used for the Perugia tunes, the overlap function (the time-integrated convolution of two proton mass distributions, see [2, 8]) is cast as

$$O(b) \propto \exp(-b^d) \quad (3)$$

with the power d a free parameter whose range is normally taken to be from $d = 1$ (exponential, representing a very peaked structure) to $d = 2$ (Gaussian, representing a smooth structure). Note that the normalization of this distribution is fixed to unity. Note also that b is given in an arbitrary unit; since the only dimensionful quantity is the total cross section, which is fixed by a Donnachie–Landshoff formula [14], the b shape does not affect the total cross section at all in this type of model, and only the dimensionless ratio $b/\langle b \rangle$ appears in the explicit calculations.²

²For completeness, we note that, while there is thus formally a dependence on the overall proton-proton impact parameter b in the model,

The power, d , appears as the parameter PARP(83) in PYTHIA.³ It is not assumed to change with energy, i.e.,

$$d(\sqrt{s}) = \text{PARP}(83). \quad (4)$$

By making separate tunes at each energy individually, we will obtain a data-driven test of the validity of this assumption.

Since all expressions are cast in terms of the dimensionless ratio of the impact parameter relative to its average, the assumed shape also does not greatly affect the *average* event activity. The main consequence of different b profiles thus lies in the *shape* of distributions, with a smooth matter profile generating narrower ones than more lumpy profiles, a consequence of the latter allowing for larger event-to-event fluctuations. The fact that the average multiplicity is not greatly affected by the choice of impact parameter profile is illustrated on the left-hand pane of Fig. 2, where the b dependence of the Perugia 0 tune has been varied between a Gaussian and an Exponential overlap distribution without substantially altering neither the average values nor their scaling with energy.

2.1.3 Color reconnection strength

The last main aspect of the modeling we shall be concerned with here is the strength of the color reconnections (CR) that are used to model the collapse of the color wave function in the final state. The so-called ‘color annealing’ models employed by the Perugia tunes were described in detail in [6, 15–17] and are qualitatively similar to the Generalized-Area-Law (GAL) models developed earlier by the Uppsala group [18]. Briefly summarized, each MPI corresponds to one or two color exchanges between the beams (depending on whether a quark or a gluon is exchanged, respectively). In the $N_C \rightarrow \infty$ limit used to represent color topologies in MC generators, every such exchange must be neutralized at the hadronization stage. In PYTHIA’s case this is modeled by the formation of strings, which subsequently hadronize to produce observable particles.

In the most naive $N_C \rightarrow \infty$ treatment, each such string would be completely independent of the others. However, since the number of real-world colors is finite, $N_C = 3$, and since the strings generated by MPI all traverse the same rapidity region (between the remnants) there is some reason

there is no actual space-time representation of the collision, and hence no dependence on the direction of b nor on the individual parton–parton impact parameters.

³Strictly speaking, this form of the matter profile is only selected for MSTP(82) = 5. See the PYTHIA documentation on MSTP(82) for how to select other matter profiles, such as the double-Gaussian one [4].

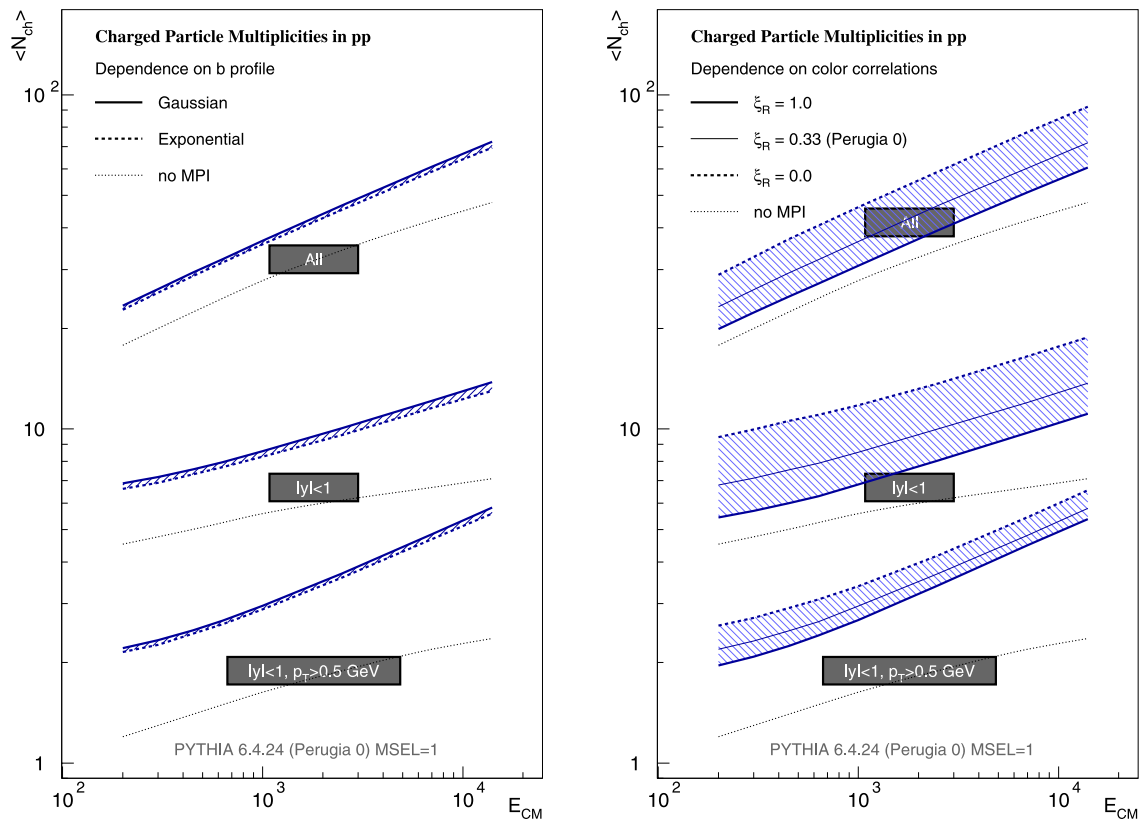


Fig. 2 Energy scaling of charged-particle multiplicities in pp in three different phase space regions (*top*: inclusive, *middle*: central, *bottom*: central hard). *Left*: two different impact parameter profiles. *Right*: three

different color-reconnection strengths. For reference, Tune A without MPI is also shown (*dotted lines*). For all other curves, the parameters of Perugia 0 were used, except for the modifications indicated on the plots

to suppose that the collapse of the color wavefunction is instead more complicated and/or that the strings after formation interact to fuse or cut each other up. In the string picture, such effects should be driven by a minimization of the total space-time area spanned by the strings (the so-called area law for classical strings, as measured, e.g., by the λ measure [19, 20]). Even without understanding the dynamics in detail, we may therefore reasonably suspect that the end result will be shorter string pieces, which in turn will produce fewer, but more energetic, particles, i.e., a harder fragmentation spectrum.

At least within the p_{\perp} -ordered PYTHIA 6 modeling, some such mechanism does appear to be empirically necessary in order to properly describe the observed increase of the mean p_{\perp} of charged tracks with track multiplicity in min-bias events [21, 22].

The annealing models developed in [6, 15–17] are all formulated in terms of one main parameter: the basic color-reconnection/string-interaction strength, ξ_R , given by PARP(78) in the code. The larger this parameter is, the stronger the reconnection effect, and the faster the rise of $\langle p_{\perp} \rangle(N_{ch})$. However, since these models were only intended as crude toy models, nothing has so far been said as to their

possible dependence on the energies of the colliding beams. The only scaling built into the models is thus a rough scaling with the number of MPI in an event, or in the most detailed variant (so far used only for the Perugia 2010 and Perugia K tunes [6]) the number of overlapping string pieces in each rapidity region. The fundamental reconnection probability is assumed constant, i.e.,

$$\xi_R(\sqrt{s}) = \text{PARP}(78). \quad (5)$$

Again, by making separate tunes at each energy individually, we will obtain a data-driven test of the validity of this assumption.

The consequence of varying PARP(78) from zero to one is illustrated in the right-hand pane of Fig. 2. We observe that the average multiplicity at each energy can be modified by up to a factor of 2 by this particular color-reconnection model, but note also that the relative scaling *between* energies stays virtually independent of ξ_R . Nonetheless, since any model with, for instance, an energy-dependent ξ_R would interpolate between our curves—leading to a different effective energy dependence—we must still conclude that the energy scaling of other models could be qualitatively different from this.

Table 1 Parameter sampling-ranges used for the MC generator runs at each energy

\sqrt{s}	630 GeV	900 GeV	1.8 TeV	1.96 TeV	7 TeV	Global fit
PARP(82)	1.066 ... 1.979	1.169 ... 2.171	1.4 ... 2.6	1.431 ... 2.658	1.993 ... 3.702	1.0 ... 3.0
PARP(78)			0.0 ... 0.7			
PARP(83)			1.0 ... 2.0			
PARP(90)						0.16 ... 0.34

Finally, we note that the CR model employed by the Perugia 0 tune actually depends on one more parameter, PARP(77), which acts to suppress reconnections among high- p_{\perp} string pieces. Since this parameter is tightly correlated with PARP(78) and since it only affects details of the high- p_{\perp} tail of the p_{\perp} distribution, we have kept it fixed to its Perugia-0 value in this study. However, we did check that allowing its value to change or even fixing it to zero did not qualitatively alter any of our conclusions.

2.1.4 Remarks on diffraction

Finally, we should emphasize a point that is especially relevant for the modeling of low-multiplicity minimum-bias collisions. As mentioned above, the model we have outlined here attempts primarily to describe inelastic, non-diffractive events. It would therefore have to be complemented by a separate modeling of the diffractive parts of the cross section, to the extent that the measurements are sensitive to these, as, e.g., in low-multiplicity minimum-bias events. In this study, we use the diffractive modeling provided by the PYTHIA 6 generator itself. We note that this modeling does not include diffractive jet production and is presumably not reliable for high- p_{\perp} and/or high-mass diffractive particle production. We therefore investigate the energy scaling of the resulting combined min-bias model for two subsamples of events—one that includes the low-multiplicity region, and one that excludes it, as will be discussed in the description of the data sets below.

2.2 Tuning framework

The PROFESSOR tuning framework [1] relies on the construction of a fast analytic model of the generator by bin-wise parameterizations of the generator's response to shifts in parameter-space. These parameterizations are performed within a hypercube of user-specified volume inside the generator parameter space. (It is up to the user to make sure that the boundaries of this hypercube correspond to meaningful generator settings.) The comparison to experimental analyzes is performed via the RIVET [23, 24] analysis tool, which in turn relies on the HEPMC [25] event record format and on the HEPDATA repository [26].

An important point is to choose analyses in RIVET that contain observables that are sensitive to the tuning-parameters. The PROFESSOR system contains tools that help to identify those observables that are not sensitive to the parameters in question (prof-sensitivities) and to confirm that the sampling-hypercube is chosen in such a way that the Monte-Carlo-generator runs enclose the experimental data (prof-envelopes).

The actual tuning stage consists of a numerical minimization of a goodness-of-fit (GOF) measure constructed from the parameterizations $f^{(b)}(\mathbf{p})$, experimental data \mathcal{R}_b and most important a weight w_b for each bin of a set of observables to tune to:

$$\chi^2(\mathbf{p}) = \sum_{\mathcal{O}} \sum_{b \in \mathcal{O}} w_b \cdot \frac{(f^{(b)}(\mathbf{p}) - \mathcal{R}_b)^2}{\Delta_b^2}. \quad (6)$$

The bin-weights w_b can be seen as the main user-input to the tuning-stage; they help to emphasize or exclude certain regions of an observable in the numerical fitting procedure. The return value of the fit will be a point \mathbf{p} that minimizes the GOF given in (6). This parameter set may then be subjected to explicit validation in terms of comparison of the observable as predicted from the parameterizations at \mathbf{p} with the experimental data. Finding a set of weights that leads to a satisfying description of observables usually requires some iteration.

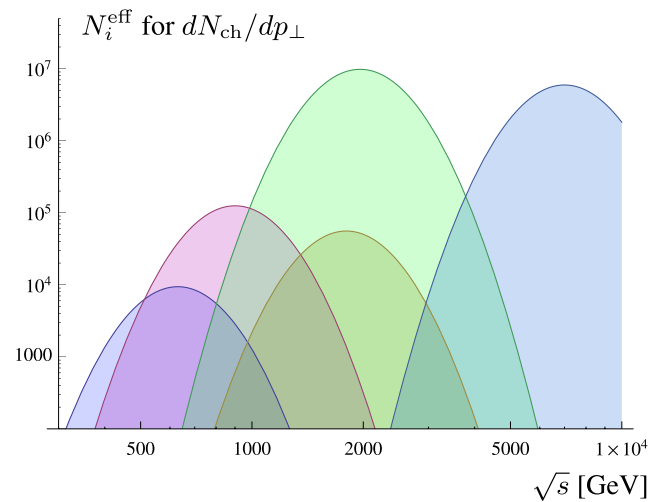
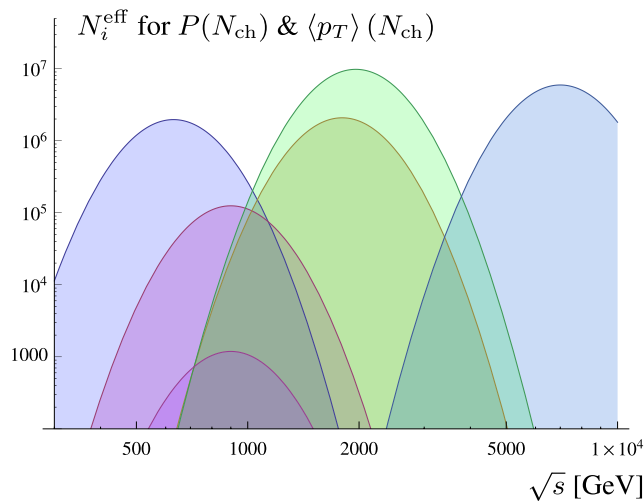
For each collider energy, we perform a “local” tune that includes only data from that particular energy. For these, we manually set PARP(89) equal to the given energy and let PROFESSOR optimize for PARP(82), which can then be interpreted directly as a $p_{\perp 0}$ value at that energy. We used the Perugia 0 parameter settings as center of our parameter sampling hypercube, and used the Perugia 0 energy scaling to define the sampling range for PARP(82) at each energy. An overview of the parameter sampling-ranges is given in Table 1.

We also perform a “global” tune, using data from all energies simultaneously, with an approximate relative weighting that attempts to take into account that the amount of data—and hence the statistical power—at each energy varies. The number of events contained in each data sample we used is listed in Table 2. (These data sets will be discussed in more

Table 2 *Top row:* Number of events contained in each data sample used, and the effective event numbers, N_i^{eff} , for each observable, for each sample used for that observable. *Bottom row:* illustration of the

relative sizes of each of the samples and their contributions to N_i^{eff} , according to (7), for *Bottom left:* $P(N_{\text{ch}})$ & $\langle p_T \rangle(N_{\text{ch}})$ and *Bottom right:* $dN_{\text{ch}}/dp_{\perp}$

Analysis	\sqrt{s} [GeV]	N_{evt}	N_i^{eff}	
			$P(N_{\text{ch}})$ & $\langle p_T \rangle(N_{\text{ch}})$	$dN_{\text{ch}}/dp_{\perp}$
CDF 1988 [30]	630	9400		47k
CDF 2002 [31]	630	1963157	2.0M	
UA5 1989 [32]	900	1189	0.8M	
ATLAS 2010 [33]	900	124782	0.8M	162k
CDF 1988 [30]	1800	55700		9.2M
CDF 2002 [31]	1800	2079558	11.2M	
CDF Run-II [21, 34]	1960	9788000	11.7M	9.8M
ATLAS 2010 [33]	7000	5395000	5.4M	5.4M
$\max(N_i^{\text{eff}})$			11.7M	



detail in Sect. 2.3.) In order to take into account the possibility that several measurements of the same observable may have been performed at closely spaced energies (e.g., Tevatron Run I and II), we define an effective total number of events for each observable, for each collider energy, as follows:

$$N_i^{\text{eff}} = \sum_j N_j e^{-r_{ij}^2/(2\sigma_E^2)}, \quad (7)$$

where j runs over all included measurements of the given observable at all energies, $r_{ij} = \log_2(E_i/E_j)$ provides a logarithmic measure of the distance between two energies, and we have chosen an “energy resolution parameter” of $\sigma_E = 1/3$ so that energies spaced a factor of 2 or more apart correspond to being 3σ away from each other and will therefore effectively contribute independently, while measurements closer than a factor $2^{1/3} \sim 1.25$ in energy will blend into each other, being resolved by less than 1σ . (We note that this parameter could be varied to help estimate the

uncertainty on the tuning, but the question of more rigorous uncertainties is not a simple one and reaches beyond the scope we aim to address here.) The effective event numbers computed in this way are given, for each observable, in the two rightmost columns of Table 2, with the figures below illustrating the effective contributions of each sample to the total at each energy.

The most naive weight normalization would be to simply let all the samples enter with unit weights. This would unavoidably bias the fit towards the energies at which most statistics has been collected. Using N_i^{eff} , we may instead attempt to normalize the statistical power in such a way as to force each energy to enter with approximately equal weight, regardless of the amount of statistics collected. This could be achieved by weighting each sample by

$$w_i^{\text{eff}} = \frac{\max(N_i^{\text{eff}})}{N_i^{\text{eff}}}, \quad (8)$$

which we refer to as “linear” reweighting. In this scheme, two event samples at widely spaced energies, containing for instance 10k and 1M events, respectively, would receive weights 100 and 1, respectively. I.e., the power of the measurement with the *lowest* statistics would be artificially enhanced, in order for the global fit not to be totally dominated by the higher-statistics one. This strategy should be considered extreme, however, since it grants no benefit at all to measurements performed with superior statistics, and one therefore risks being overly sensitive to noise in the poorly measured ones. As an intermediate compromise, we propose to let the samples enter with relative weights

$$w_i^{\text{eff}} = \sqrt{\frac{\max(N_i^{\text{eff}})}{N_i^{\text{eff}}}} \quad (9)$$

such that the samples in the example above would enter with weights 10 and 1, rather than 100 and 1. We refer to this as “square root” reweighting.

Obviously, we do not intend this to define a rigorous procedure, but view it as a first attempt at highlighting and addressing the disparate statistical powers available in the various sets.

As a final comment, we note that a certain freedom exists for the parameterization of the generator response which allows for systematic checks of the validity of obtained tuning results which can also be turned into typical spreads which we use as rough uncertainty estimates in this study. These estimates and their properties are described in greater detail in [27]. We include them as light shaded (cyan) bands on the plots in the following subsections, added in quadrature with the ordinary χ^2 -based fit uncertainties computed by Minuit, which are shown as darker shaded bands.

2.3 Data sets

To study the energy dependence of parameters properly, we need to make sure that the experimental data at each energy has either been corrected for detector effects in a comprehensible way or that the uncorrected data is presented with enough information on efficiencies so that it can be used with RIVET. Further, the distributions we choose must be both sensitive to the parameters we want to tune and, simultaneously, consistently available at all collider energies with phase-space regions and trigger conditions not too different from one to the next.

A minimal set of minimum-bias distributions that matches these requirements for our selection of tuning parameters and for energies ranging from 630 GeV to 7 TeV is:

- The charged-particle multiplicity-distribution (N_{ch})
- The charged-particle p_{\perp} -distribution
- The charged-particle $\langle p_{\perp} \rangle$ vs. N_{ch} -distribution

We would have liked to extend the reach of our study by including data from the RHIC experiments but were unable to find any published (corrected) data for the $\langle p_{\perp} \rangle$ vs. N_{ch} distribution at 200 GeV. The same is true for older experiments such as SFM, where multiplicities and p_{\perp} -distributions were measured in pp-reactions at 62 GeV and lower. As has recently been emphasized [28], it would also have been interesting to add forward–backward correlations to the list of variables, since these are particularly sensitive to the mix of short- vs. long-distance processes, but in the energy range we consider, only UA5 has so far reported measurements [29], with numbers at Tevatron and LHC energies not (yet) available.

We consider two separate event samples: one where all events with at least one charged track are included, corresponding to a conventional min-bias definition (though excluding the “zero bin” when comparing to experiments that include it in their measurements), and one where only events with $N_{\text{ch}} \geq 6$ are included, corresponding to a sample in which diffractive contributions are expected to be strongly suppressed. Due to the ambiguities discussed earlier concerning the treatment of diffraction, we shall base our conclusions mainly on the $N_{\text{ch}} \geq 6$ sample, using the other as a further counter-check into the diffractive region. Since measurements with an explicit $N_{\text{ch}} \geq 6$ definition have so far only been carried out by ATLAS, the closest we can get for other experiments is to suppress the five first bins in $P(N_{\text{ch}})$ and $\langle p_{\perp} \rangle$ vs. N_{ch} , and adjusting the normalization of the former such that the remaining bins sum to unity.

A further complication concerns the phase-space (“fiducial”) regions measured by the different experiments. Although no two experiments have exactly the same coverage, it is here of great help that all of the experiments we consider have taken data at least two energies, and in the best cases also in several different phase space regions. This effectively allows us to construct a kind of bootstrapped path among the different regions. In fact, without such counter-checks, the method we propose here would be badly compromised—one could then never be certain whether a deviation in the optimized parameters is caused by energy dependence or by the difference in phase space regions. We therefore encourage the RHIC, Tevatron, and LHC experiments in their efforts to make measurements using several different combinations of trigger conditions, phase space regions, and collider energies. Ultimately, it is by such comprehensive and systematic sets of measurements, and by the comprehensive tests that they enable, that we may establish a truly reliable modeling of collider final states.

The observables, ranges, and weights used for both the $N_{\text{ch}} \geq 1$ and $N_{\text{ch}} \geq 6$ samples are given in Table 3. Larger weights are given to the high-multiplicity tails of the $P(N_{\text{ch}})$ and $\langle p_{\perp} \rangle(N_{\text{ch}})$ distributions to emphasize their asymptotic slopes. As our definition for “high multiplicity”, we took

Table 3 Observables and ranges included in the study. The trigger (“Trig”) conditions are as follows: *TC1*: CDF Run I MB [30], *TC2*: CDF Run II MB[21], *TU1*: UA5 MB [30], *TA**: ATLAS trigger requir-ing ≥ 1 (≥ 6) charged particles within $|\eta| < 2.5$ and $p_{\perp} > 0.5$ GeV for the $N_{\text{ch}} \geq 1$ ($N_{\text{ch}} \geq 6$) sample

Exp/Trig	\sqrt{s} [GeV]	Ref.	Observable	$ \eta _{\text{max}}$	$p_{T\text{min}}$ [GeV]	Fitrangle		Weight	
						$N_{\text{ch}} \geq 1$	$N_{\text{ch}} \geq 6$	Local	Global
CDF/TC1	630	[30]	dN/dp_{\perp}	1.0	0.4	0.4–3	0.4–3	1.0	15.8
CDF/TC1	630	[31]	$P(N_{\text{ch}})$	1.0	0.4	1–16	6–16	1.0	2.4
			$P(N_{\text{ch}})$	1.0	0.4	16–30	16–30	5.0	12.1
			$\langle p_{\perp} \rangle$ vs. N_{ch}	1.0	0.4	1–16	6–16	1.0	2.4
CDF/TC1	630	[31]	$\langle p_{\perp} \rangle$ vs. N_{ch}	1.0	0.4	16–30	16–30	5.0	12.1
UA5/TU1	900	[32]	$P(N_{\text{ch}})$	0.5	0.0	1–19	6–19	1.0	3.8
UA5/TU1	900	[32]	$P(N_{\text{ch}})$	0.5	0.0	19–30	19–30	5.0	19.1
ATL/TA*	900	[33]	dN/dp_{\perp}	2.5	0.5	0.5–10.0	0.5–10.0	1.0	8.5
ATL/TA*	900	[33]	$P(N_{\text{ch}})$	2.5	0.5	1–25	6–25	1.0	3.8
ATL/TA*	900	[33]	$P(N_{\text{ch}})$	2.5	0.5	25–60	25–60	5.0	19.1
ATL/TA*	900	[33]	$\langle p_{\perp} \rangle$ vs. N_{ch}	2.5	0.5	1–25	6–25	1.0	3.8
			$\langle p_{\perp} \rangle$ vs. N_{ch}	2.5	0.5	25–60	25–60	5.0	19.1
CDF/TC1	1800	[30]	dN/dp_{\perp}	1.0	0.4	0.4–10.0	0.4–10.0	1.0	1.1
CDF/TC1	1800	[31]	$P(N_{\text{ch}})$	1.0	0.4	1–17	6–17	1.0	1.0
CDF/TC1	1800	[31]	$P(N_{\text{ch}})$	1.0	0.4	17–40	17–40	5.0	5.0
CDF/TC1	1800	[31]	$\langle p_{\perp} \rangle$ vs. N_{ch}	1.0	0.4	1–17	6–17	1.0	1.0
CDF/TC1	1800	[31]	$\langle p_{\perp} \rangle$ vs. N_{ch}	1.0	0.4	17–40	17–40	5.0	5.0
CDF/TC2	1960	[21]	dN/dp_{\perp}	1.0	0.4	0.4–15.0	0.4–15.0	1.0	1.0
CDF/TC2	1960	[34]	$P(N_{\text{ch}})$	1.0	0.4	1–18	6–18	1.0	1.0
CDF/TC2	1960	[34]	$P(N_{\text{ch}})$	1.0	0.4	18–30	18–30	5.0	5.0
CDF/TC2	1960	[21]	$\langle p_{\perp} \rangle$ vs. N_{ch}	1.0	0.4	1–18	6–18	1.0	1.0
CDF/TC2	1960	[21]	$\langle p_{\perp} \rangle$ vs. N_{ch}	1.0	0.4	18–30	18–30	5.0	5.0
ATL/TA*	7000	[33]	dN/dp_{\perp}	2.5	0.5	0.5–40	0.5–40.0	1.0	1.5
ATL/TA*	7000	[33]	$P(N_{\text{ch}})$	2.5	0.5	1–49	6–49	1.0	1.5
ATL/TA*	7000	[33]	$P(N_{\text{ch}})$	2.5	0.5	49–70	49–70	5.0	7.4
ATL/TA*	7000	[33]	$\langle p_{\perp} \rangle$ vs. N_{ch}	2.5	0.5	1–49	6–49	1.0	1.5
ATL/TA*	7000	[33]	$\langle p_{\perp} \rangle$ vs. N_{ch}	2.5	0.5	49–70	49–70	5.0	7.4

the N_{ch} value that came closest to separating out the 1% highest-multiplicity events, for each measurement.

3 Consistency of energy scaling

In this section, we study the degree to which the parameters obtained for a best-fit tune across all included data sets and collider energies are consistent with those obtained when we include only specific subsets of the data. In particular, we focus on the consistency on the assumed energy scaling by comparing the results of the global fit to results obtained at each energy separately.

We study three specific questions:

- Is the assumed scaling law governing the infrared regularization scale for multiple parton interactions consistent with what one finds when optimizing the tuning at each energy separately?
- Is the transverse mass distribution of the proton (assumed unchanging with energy) consistent with what one finds when optimizing the tuning at each energy separately?
- Is the assumed color reconnection strength (assumed unchanging with energy) consistent with what one finds when optimizing the tuning at each energy separately?

The evolution of the infrared regularization scale with energy is depicted in the top row of Fig. 3, with the left-hand panes showing the results for the $N_{\text{ch}} \geq 1$ sample and the right-hand panes those for $N_{\text{ch}} \geq 6$. The scaling of Perugia 0

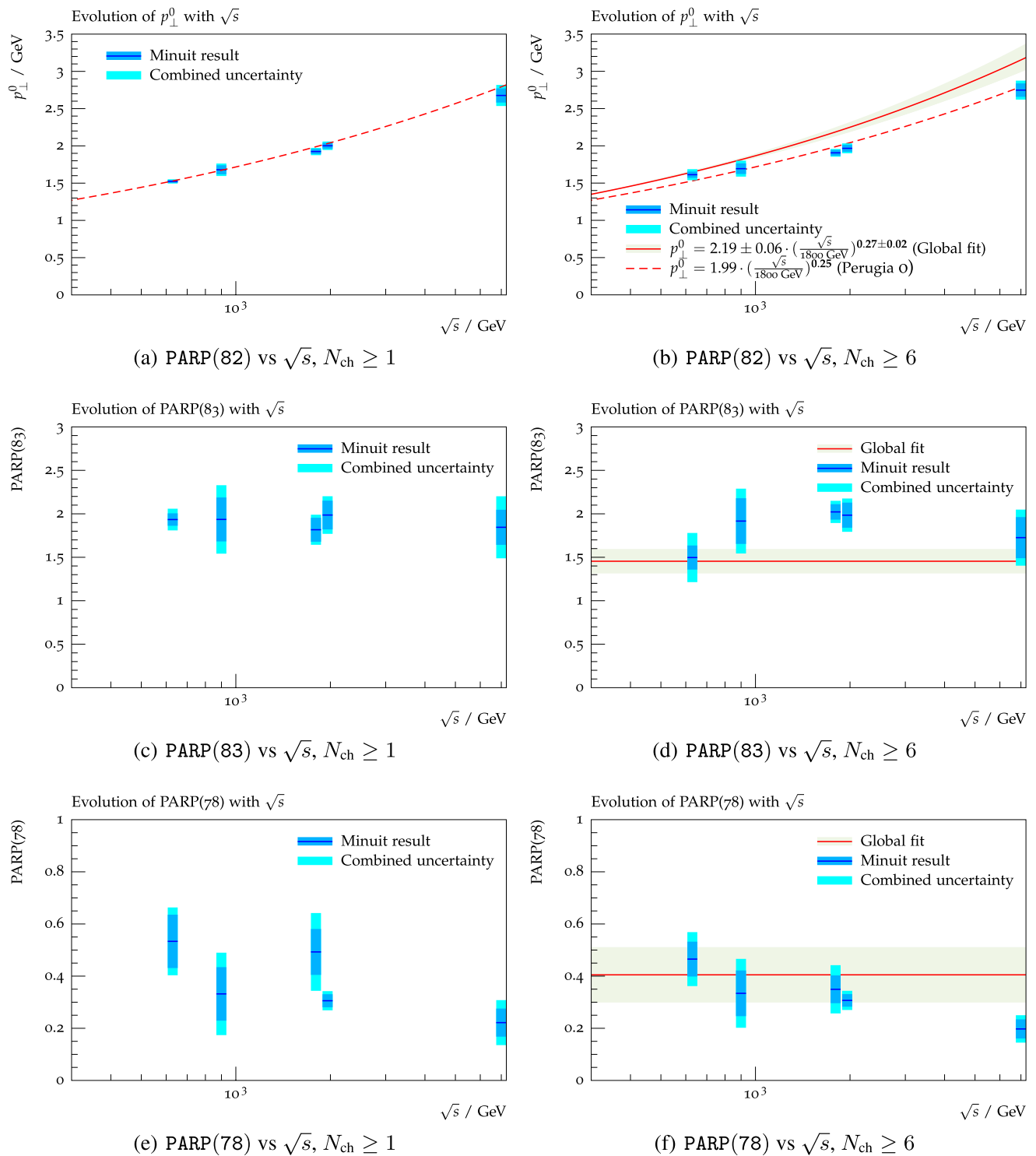


Fig. 3 (Color online) Energy dependence of the three tune parameters, from top to bottom: PARP(82), PARP(83), and PARP(78). Independent optimizations (blue/shaded lines) compared to global fit curve (red solid curves). Left: $N_{\text{ch}} \geq 1$ sample. Right: $N_{\text{ch}} \geq 6$ sample

(red dashed lines) is compared to the global fit (red solid line in light shaded band) and to the independent optimizations (blue horizontal lines inside cyan bands). As mentioned in Sect. 2.2, the inner (darker blue) bands correspond to the fit

parameter uncertainties calculated by Minuit and the outer (lighter cyan) bands include an estimate of PROFESSOR's interpolation uncertainty [27] as well, with the two added in quadrature.

We make four conclusions concerning PARP(82). One, that the results of the independent optimizations are consistent with the functional form represented by (2) and hence we find no evidence for a need for any significant departure from the model assumptions in this energy range. Two, that the different individual data sets appear to be consistent and reconcilable within this modeling context—the two different CDF measurements, the left- and right-hand sample definitions, and the ATLAS and UA5 measurements at 900 GeV, all appear to give consistent parameters. Three, that the light shaded (cyan) bands are very small, and that the tune result can therefore be considered technically stable. And finally, that the global fit *does not* coincide with the independent optimizations. Although not huge, this deviation hints that one or more of the other parameters must be exhibiting a non-universal behavior.

Turning now to the scaling of the transverse shape parameter, PARP(83), this is particularly interesting since it could reveal whether minimum-bias collisions at different energies effectively probe a different “average proton shape”. Such a variation could, e.g., be generated by correlations between b and x (see, e.g., [35–38]), folded with the different x ranges that are accessible at each energy. Roughly speaking, we might then expect to see a slightly more lumpy average proton at lower energies, consistent with a higher average x at those energies, and a smoother proton at higher energies/lower average x . Results of the local and global tunes for PARP(83) are shown in the middle panes of Fig. 3.

In our main $N_{\text{ch}} \geq 6$ sample, shown in the right-hand pane of Fig. 3, there may be some weak evidence for such a trend, with a close-to-Gaussian proton ($\text{PARP}(83) = 2$) favored by the high-energy data and a slightly more peaked distribution favored by the 630-GeV data. However, note that the shaded bands are here larger, indicating that the fit is less well constrained than it was for PARP(82). Also, when we include the lowest-multiplicity bins, in the left-hand pane, the trend disappears. Our tentative conclusion is therefore that the uncertainties are too large to make any firm conclusions, but that the model at least appears to be self-consistent within those uncertainties. Further studies at lower energies (e.g., including pp data from RHIC and/or further Tevatron studies at 630 GeV) and/or attempting to isolate different effective x ranges at higher energies, e.g., by using different rapidity and/or trigger regions, could contribute significantly to probing this question further. More theoretical work to improve the understanding of the relationship between the language used here and that of other phenomenological models, as was done, e.g., by [39], would also be valuable and could open the possibility for a consistent “importation” of constraints from related physical models into the Monte Carlo context.

Returning to the present study, note also that PROFESSOR furnishes us with one additional key piece of information. When optimizing several parameters simultaneously,

we not only get the optimized values for each parameter separately; we also get a correlation matrix between them. When interpreting our results, one should therefore be aware that there is a strong correlation between PARP(82) and PARP(83). Hence, there is still a possibility that PARP(83) could have a more significant energy dependence, to be traded off against that of PARP(82). However, since the fit result for PARP(82) was, itself, quite stable, we consider this possibility something of a minority report, not favored by the central fits; but also not completely excluded.

Finally, the Color Reconnection strength, PARP(78), is—perhaps not surprisingly—the least well constrained parameter, with the individual fit results showing a preferred scaling with energy that is not accounted for by the underlying model. Given the large uncertainties surrounding color correlations and final-state interactions in hadron collisions, this can be considered a reminder that, although the models do make attempts at incorporating this kind of phenomena, our understanding is still very far from complete or reliable. For the time being, any global tuning relying on the CR models considered in this study would be forced to make a compromise between the high- and low-energy data. Pragmatic alternatives for physics studies at specific colliders would range from giving that particular energy a larger weight in the global fit to simply abandoning a global fit altogether. Although clearly not theoretically satisfactory, the latter may be a useful strategy for applications in which the Monte Carlo modeling is only used as a sophisticated differential “parameterization” of the behavior of the data. In particular at 7 TeV, large data samples are now becoming available that probe many different and complementary phase space regions in detail, allowing a fairly complete set of constraints to be obtained for that particular collider energy.

4 Conclusions

We have argued that the capabilities of modern tuning tools can and should be used for more than just making “best fits” to a collection of data. For example, by making independent optimizations of the MC generator parameters for several different collider energies, we have here obtained a data-driven test of the universality of the generator modeling. Three of the most important generator parameters controlling the underlying-event and minimum-bias physics in the PYTHIA 6 generator were included in the study, corresponding to: the infrared regularization scale for multiple parton interactions (MPI), the proton transverse shape, and the strength of color reconnections (CR). A brief discussion of each of these parameters was given in Sect. 2.1. The PROFESSOR tool used for the tunings as well as a weighting strategy that attempts to take the size of disparate statistical samples and measurements at closely spaced energies

Table 4 Tuning results obtained with the maximum information parameterizations. The errors quoted are those calculated by Minuit

\sqrt{s}	630 GeV	900 GeV	1.8 TeV	1.96 TeV	7 TeV	Global fit
PARP(89)	630.0	900.0	1800.0	1960.0	7000.0	1800.0
Tuning to observables with $N_{\text{ch}} \geq 1$						
PARP(78)	0.53 ± 0.10	0.33 ± 0.10	0.49 ± 0.09	0.31 ± 0.03	0.22 ± 0.05	
PARP(82)	1.52 ± 0.02	1.68 ± 0.06	1.92 ± 0.03	2.00 ± 0.04	2.68 ± 0.10	
PARP(83)	1.93 ± 0.07	1.94 ± 0.25	1.82 ± 0.14	1.99 ± 0.17	1.85 ± 0.20	
Tuning to observables with $N_{\text{ch}} \geq 6$						
PARP(78)	0.47 ± 0.07	0.33 ± 0.09	0.35 ± 0.05	0.31 ± 0.02	0.20 ± 0.04	0.40 ± 0.11
PARP(82)	1.61 ± 0.04	1.69 ± 0.07	1.91 ± 0.04	1.97 ± 0.05	2.75 ± 0.01	2.19 ± 0.06
PARP(83)	1.50 ± 0.14	1.92 ± 0.03	2.02 ± 0.09	1.98 ± 0.01	1.73 ± 0.02	1.45 ± 0.14
PARP(90)						0.27 ± 0.02

formally into account were described in Sect. 2.2. The data sets consisted of minimum-bias measurements of charged particle multiplicities, p_{\perp} spectra, and the average of p_{\perp} vs. multiplicity, as described in Sect. 2.3.

Our numerical results were presented and discussed in Sect. 3. We find that the result of independent optimizations of the IR regularization scale, energy by energy, are consistent with the power-law behavior assumed by the model, at least within the energy range we were able to probe, from 630 to 7000 GeV. The transverse matter distribution may exhibit mild deviations from universality, a question which data in particular from minimum-bias measurements at RHIC could help shed further light on. Finally, the optimal value of the color-reconnection strength appears to vary significantly with energy, with lower values preferred at higher energies, in contrast to the intrinsic assumption of a constant strength in the model. This confirms the theoretical evaluation, that the CR modeling is currently the largest source of theoretical ambiguity, and emphasizes that at least the models investigated here cannot be considered truly universal over the studied energy range.

The parameter values corresponding to each of our “local” (i.e., energy-by-energy) tunes as well as those of a “global” one are collected in Table 4. All other parameter values were taken to be those of the Perugia 0 tune [6]. These new tunes have been included in PYTHIA starting from version 6.4.25, with tune numbers 360–365 (see the PYTHIA update notes for details).

Finally, we argue that procedures similar to the one followed here can be used to advantage also in other contexts, to give a clearer picture of which regions the modeling is able to describe with approximately universal parameters, which in turn helps isolate the genuinely problematic areas more easily. The trend of the optimized parameters to deviate in one or another direction in the problematic regions may also give clues as to the root of the problem, though such interpretations should be made in conjunction with a

good understanding of the correlations between the parameters and a careful evaluation of the possible missing physics components in the modeling. One clear possibility to apply this type of strategy to present measurements would be to perform independent optimizations using different complementary phase space regions at each energy.

Acknowledgements This work was supported in part by the Marie Curie research training network “MCnet” (contract number MRTN-CT-2006-035606). H.S. acknowledges support from the German Research Foundation (DFG) in the framework of Graduate School 1504.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

1. A. Buckley, H. Hoeth, H. Lacker, H. Schulz, J.E. von Seggern, Eur. Phys. J. C **65**, 331 (2010). [arXiv:0907.2973](#)
2. T. Sjöstrand, P.Z. Skands, J. High Energy Phys. **03**, 053 (2004). [hep-ph/0402078](#)
3. T. Sjöstrand, P.Z. Skands, Eur. Phys. J. C **39**, 129 (2005). [hep-ph/0408302](#)
4. T. Sjöstrand, S. Mrenna, P.Z. Skands, J. High Energy Phys. **05**, 026 (2006). [hep-ph/0603175](#)
5. P. Bartalini et al. (eds.) (2010). [arXiv:1003.4220](#)
6. P.Z. Skands, Phys. Rev. D **82**, 074018 (2010). [arXiv:1005.3457](#)
7. T. Sjöstrand, S. Mrenna, P.Z. Skands, Comput. Phys. Commun. **178**, 852 (2008). [arXiv:0710.3820](#)
8. T. Sjöstrand, M. van Zijl, Phys. Rev. D **36**, 2019 (1987)
9. M. Ryskin, A. Martin, V. Khoze (2011). [arXiv:1102.2844](#)
10. A. Buckley et al. (2011). [arXiv:1101.2599](#)
11. CTEQ Collaboration, H.L. Lai et al., Eur. Phys. J. C **12**, 375 (2000). [hep-ph/9903282](#)
12. CTEQ Collaboration, J. Pumplin et al., J. High Energy Phys. **07**, 012 (2002). [hep-ph/0201195](#)
13. A. Sherstnev, R.S. Thorne, Eur. Phys. J. C **55**, 553 (2008). [arXiv:0711.2473](#)
14. A. Donnachie, P.V. Landshoff, Phys. Lett. B **296**, 227 (1992). [hep-ph/9209205](#)

15. M. Sandhoff, P.Z. Skands (2006). [hep-ph/0604120](#)
16. C. Buttar et al. (2006). [hep-ph/0604120](#)
17. P.Z. Skands, D. Wicke, Eur. Phys. J. C **52**, 133 (2007). [hep-ph/0703081](#)
18. J. Rathsmann, Phys. Lett. B **452**, 364 (1999). [hep-ph/9812423](#)
19. B. Andersson, G. Gustafson, B. Söderberg, Z. Phys. C **20**, 317 (1983)
20. B. Andersson, G. Gustafson, B. Söderberg, Nucl. Phys. B **264**, 29 (1986)
21. CDF Collaboration, T. Aaltonen et al., Phys. Rev. D **79**, 112005 (2009). [arXiv:0904.1098](#)
22. ATLAS Collaboration, G. Aad et al., Phys. Lett. B **688**, 21 (2010). [arXiv:1003.3124](#)
23. B.M. Waugh et al. (2006). [hep-ph/0605034](#)
24. A. Buckley et al. (2010). [arXiv:1003.0694](#)
25. M. Dobbs, J.B. Hansen, Comput. Phys. Commun. **134**, 41 (2001)
26. A. Buckley, M. Whalley (2010). [arXiv:1006.0517](#)
27. A. Buckley et al. (2010). [arXiv:1005.5357](#)
28. K. Wraight, P. Skands, Eur. Phys. J. C (2011, submitted). [arXiv:1101.5215](#)
29. UA5 Collaboration, R. Ansorge et al., Z. Phys. C **37**, 191 (1988)
30. CDF Collaboration, F. Abe et al., Phys. Rev. Lett. **61**, 1819 (1988)
31. CDF Collaboration, D.E. Acosta et al., Phys. Rev. D **65**, 072005 (2002)
32. UA5 Collaboration, R.E. Ansorge et al., Z. Phys. C **43**, 357 (1989)
33. ATLAS Collaboration, CERN Report No. ATLAS-COM-CONF-2010-031 (2010, unpublished) (Was originally 'ATL-COM-PHYS-2010-268')
34. CDF Collaboration, N. Moggi et al., PUB-NOTE **9936** (2009)
35. G. Pancheri, A. Grau, R.M. Godbole, Y.N. Srivastava (2010). [arXiv:1007.0208](#)
36. B. Blok, Y. Dokshitzer, L. Frankfurt, M. Strikman (2010). [arXiv:1009.2714](#)
37. L. Frankfurt, M. Strikman, C. Weiss (2010). [arXiv:1009.2559](#)
38. R. Corke, T. Sjöstrand (2011). [arXiv:1101.5953](#)
39. T. Rogers, M. Strikman, Phys. Rev. D **81**, 016013 (2010). [arXiv:0908.0251](#)