CERN-PH-TH/2011-228
DESY 11-124

Monte Carlo tuning in the presence of Matching

B. Cooper[1], J. Katzy[2], M. L. Mangano[3], A. Messina[3], L. Mijović[2], P. Skands[3]

[1] Department of Physics & Astronomy, University College London, WC1E 6BT, UK

[2] Deutsches Electron Synchroton, 22603 Hamburg, Germany

[3] European Organization for Nuclear Research, CERN CH-1211, Geneva 23, Switzerland

Abstract

We consider the impact of varying α_s choices (and scales) on each side of the so-called "matching scale" in MLM-matched matrix-element + parton-shower predictions of collider observables. We explain how inconsistent prescriptions can lead to counter-intuitive results and present a few explicit examples, focusing mostly on W/Z + jets processes. We give a specific prescription for how to improve the consistency of the matching and also address how to perform consistent tune variations (e.g., of the renormalization scale) around a central choice. Comparisons to several collider processes are included to illustrate the properties of the resulting improved matching, relying on AlpGen + Pythia 6, with the latter using the so-called Perugia 2011 tunes, developed as part of this effort.

1 Introduction

The theoretical description of multijet production in hadronic collisions is one of the key ingredients for the interpretation of the data from high-energy hadron colliders, the 1.96 TeV proton-antiproton Tevatron collider at Fermilab, and the proton-proton Large Hadron Collider (LHC) at CERN. Final states with multijets, possibly associated with electroweak gauge bosons, are in fact the dominant signature of the decay of heavy particles produced at high energy, whether in the Standard Model (top quarks and Higgs bosons), or in theories beyond the Standard Model (BSM), such as supersymmetry. The identification of these particles, and the study of their properties, requires an accurate modelling of the Standard Model (SM) sources of multijets. Great progress was achieved towards this goal in the past decade. On one side, the calculation of inclusive, parton-level, cross-sections to next-to-leading-order (NLO) in QCD has produced results for processes as complex as W+4 jets [1]. On the other, algorithms have been developed and implemented in numerical codes to provide a complete description of the hadronic final states emerging from processes with up to 6 jets, merging the exact leading-order (LO) calculation of the partonic matrix elements (ME) with the evolution, provided by so-called shower Monte Carlo (MC) codes, of the partonic shower (PS) and the subsequent hadronization of the partons in to physical hadrons.

The development of theoretical tools has been accompanied by experimental measurements, which provide the necessary validation test-bed for these calculations. Parton-level NLO calculations provide a first-principle description of inclusive final states: they have an intrinsic high degree of precision, due to the reduced dependence on the unphysical choice of a renormalization and factorization scale and, furthermore, are not subject to modelling uncertainties related to the details of the non-perturbative phase of the final state evolution. Calculations based on the merging of LO matrix elements, shower evolution and hadronization, on the other hand, while affected by the larger scale-setting uncertainty due to the LO approximation, provide a fully exclusive description of the final states, and are therefore more suitable for the experimental analyses. Their ultimate goal is not only to give reliable estimates of the inclusive jet rates and energy distributions, but also to reproduce properties of the final states such as the jet inner structure and the distribution of softer particles produced outside of the jets, including those resulting from the evolution of the fragments of the original colliding hadrons[1]. These properties, which depend on the details of the non-perturbative dynamics, can only be described through the phenomenological models embedded in the shower MC codes. The parameters of these phenomenological models need to be tuned using experimental data of some suitable observables. The factorization assumption built into any description of large-Q processes justifies the use of these same parameters in the prediction of different observables, and provides the basis for the predictive power of such tools. This assumption however must be validated with a direct comparison with data. Elements that need to be probed include the scaling with beam energy of the UE parameters, the universality of the parameters controlling the shower evolution and hadronization, and the overall independence of all parameters on the type of hard process. Deviations from the expected universality would highlight faults in the underlying modelling of effects beyond perturbative physics, or could be due to the insufficient precision of the perturbative description, in case NLO effects were to modify significantly the LO predictions. Differences compatible with the theoretical systematics of the LO approximation could however be reabsorbed by modifying the perturbative parameters that govern the LO systematics, for example the renormalization and factorization scales, or the matching variables used in the matrix-element/shower merging algorithm.

It is therefore important to understand the correlations between the effects of changing the soft and UE parameters on one side, and the perturbative parameters on the other. In this paper we present studies which

---

[1] We refer to the ensemble of these particles as the underlying event, or UE

demonstrate that, in the tuning of ME-PS matched predictions, it is vital that there is consistency in the treatment of α_S in both the ME and PS components. While this is a general issue for all shower MCs, we consider as an explicit example the merging of LO matrix elements with the Pythia 6 shower MC [2], as implemented in the framework of the AlpGen code [3], one of the reference tools for experimental multijet studies at the Tevatron and at the LHC. The most recent versions of Pythia (6.425) and AlpGen (2.14) codes were used for producing the results.

On the Pythia 6 side, we consider several different tune variations of the interleaved p_T-ordered parton-shower model [4], focusing on the so-called "Perugia" set of tunes of [5, 6], ranging from the Perugia 0 tune (from 2009) to the Perugia 2011 updates that have been developed as part of this work, including systematic up/down variations of the shower activity (see the Appendix and [5, 6] for details). We also compare to the "DW" tune [7] of the virtuality-ordered shower model [8, 9]. For Herwig [10], we include the "Jimmy" underlying-event model [11], with default parameters. We emphasize that the qualitative conclusions presented in this paper carry over to other shower models, including the ones implemented in Pythia 8 [12, 13] and Herwig ++ [14], but the quantitative aspects should still be considered limited to the particular tunes and shower models studied here. We rely on Fastjet [15] for jet clustering and have further used the Rivet-based [16] mcplots web site [17] for some of our comparisons.

The paper is organised as follows: in Section 2 we describe in detail the theoretical nature of the α_S consistency problem, and give a practical example of how it can be manifest in the prediction of high p_T observables. In Section 3 we show how a simple prescription can be applied to stabilise ME-PS tunings against this problem, propose a new tune for AlpGen + Pythia 6 matched predictions, and demonstrate the behaviour of this tune under tuning variations. In Section 4 we show that this new AlpGen + Pythia 6 tune is able to reproduce (within statistical errors) the Tevatron and LHC vector boson plus jets data. In addition we also the tune predictions to the jet shape measurements at the Tevatron and LHC. Finally we conclude in Section 5.

2 The Importance of Consistent α_S Treatment in ME-PS Matched Predictions

In this section we demonstrate that consistent treatment of α_S in ME-PS matched predictions is important in order to achieve the desired accuracy in the prediction of high p_T observables. We first present the theoretical arguments behind this, and then go on to show and explain that without adopting this approach one can observe undesirable and counter-intuitive effects on experimental observables.

2.1 Theoretical Background

The philosophy behind matching prescriptions such as the MLM one [18, 19] employed by AlpGen is to separate phase space cleanly into two distinct regions; a short-distance one, which is supposed to be described by matrix elements, and a long-distance one described by parton showers. In the long-distance region, real and virtual corrections, with the latter represented by Sudakov factors, are both generated by the shower and are intimately related by unitarity (for pedagogical reviews, see, e.g., [20, 21]). On the short-distance side, the real corrections are generated by the matrix elements while the virtual ones are still generated by the shower.

Much effort has gone into ensuring that the behaviour across the boundary between the two regions be as smooth as possible. CKKW showed [22] that it is possible to remove any dependence on this "matching scale" at NLL precision by careful choices of all ingredients in the matching; technical details of the implementation are important, and the dependence on the unphysical matching scale may be larger than NLL unless the implementation matches the theoretical algorithm precisely [23–25].

Especially when two different computer codes are used for matrix elements and showering, respectively (as when AlpGen or MadGraph [26] is combined with Pythia 6 or Herwig), inconsistent parameter sets between the two codes can jeopardise the consistency of the calculation and lead to unexpected results, as will be illustrated in the following sections.

To give a very simple theoretical example, suppose a matched matrix-element generator (MG) uses a different definition of α_s than the parton-shower generator (SG). Suppressing parton luminosity factors to avoid clutter, the real corrections, integrated over the hard part of phase space, for some arbitrary final state F, will then have the form

$$\sigma_{F+1}^{\mathrm{incl}} = \int_{Q_F^2}^{s} \mathrm{d}\Phi_{F+1}\ \alpha_s^{\mathrm{MG}}\ |M_{F+1}|^2\ , \tag{1}$$

where we have factored out the coupling corresponding to the "+1" parton and suppressed the dependence on any other couplings that may be present in $|M_{F+1}|^2$. The virtual corrections at the same order, generated by the shower off F, will have the form

$$\sigma_F^{\mathrm{excl}} = \sigma_F^{\mathrm{incl}} - \int \mathrm{d}\Phi_F \int_{Q_F^2}^{s} \frac{\mathrm{d}Q^2}{Q^2}\ \mathrm{d}z\ \sum_i \frac{\alpha_s^{\mathrm{SG}}}{2\pi} P_i(z)\ |M_F|^2\ +\ \mathcal{O}(\alpha_s^2)\ , \tag{2}$$

with $P_i(z)$ the DGLAP splitting kernels (or equivalent radiation functions in dipole or antenna shower approaches). If the two codes use the same definitions for the strong coupling, $\alpha_s^{\mathrm{SG}} = \alpha_s^{\mathrm{MG}}$, then the fact that $P(z)/Q^2$ captures the leading singularities of $|M_{F+1}|^2$ guarantees that the difference between the two expressions can at most be a non-singular term. Integrated over phase space, such a term merely leads to a finite $\mathcal{O}(\alpha_s)$ change to the total cross section, which is within the expected precision. Indeed, it is a central ingredient in both the MLM and (L)-CKKW matching prescriptions that a reweighting of the matched matrix elements be performed in order to ensure that the scales appearing in α_s match smoothly between the hard and soft regions. Thus, we may assume that the choice of renormalization scale after matching is $\mu \sim p_{\mathrm{T}}$ on both sides of the matching scale, where p_{T} is a scale characterising the momentum transfer at each emission vertex, as established by [27, 28] and encoded in the CKKW formalism [22].

In the case of the CKKW approach as implemented in the Sherpa MC framework [29], this prescription can be controlled exactly, since the matrix element and the shower evolution are part of the same computer code and hence naturally use the same α_s definition. This is also true in Lönnblad's variant [23] of the algorithm, used in Ariadne [30]. In the case of codes like AlpGen or Madgraph, on the other hand, an issue emerges. These codes are designed to generate parton-level event samples to be used with an arbitrary shower MC. Different shower MCs however use slightly different scales for the parton branchings, as a result of different approaches to the shower evolution, and may use different values of Λ_{QCD}, as a result of the tuning of the showers and/or underlying events. A possible mismatch therefore arises in the values of α_s used by the matrix-element calculation and those used by the shower.

If there is a mismatch in Λ_{QCD} or $\alpha_s(M_Z)$, then this will effectively generate a real-virtual difference whose leading singularities are proportional to

$$\alpha_s^2\ b_0 \ln\left(\frac{\Lambda_{\mathrm{MG}}^2}{\Lambda_{\mathrm{SG}}^2}\right) \frac{\mathrm{d}Q^2}{Q^2}\ \sum_i P_i(z)\ |M_F|^2\ . \tag{3}$$

	A	B	C
Λ_{MG}	Λ	$\frac{1}{2}\Lambda$	Λ
Λ_{SG}	Λ	Λ	$\frac{1}{2}\Lambda$

Table 1: The three cases, A, B, and C discussed in the text, for an arbitrary reference Λ value.

which is of next-to-leading logarithmic order (unless $\Lambda_{\mathrm{MG}} \sim \Lambda_{\mathrm{SG}}$, in which case it vanishes). Similarly, even if both matrix-element and shower codes are using the same Λ_{QCD}, but they use different running orders, then there will be an $\mathscr{O}(\alpha_s^3 \ln(p_\mathrm{T}^2/\Lambda^2))$ mismatch, which may also become large if $p_\mathrm{T} \gg \Lambda$.

To be more concrete, let us consider a specific example. Compare A) a matched MG+SG calculation which uses the same Λ_{QCD} value on both sides of the matching to B) a calculation in which the value used on the MG side is reduced to half its previous value but the SG one remains the same, as summarised by the two first columns of tab. 1. Going from case A to B, the following changes result:

1. The number of $(F+1)$ states added by the MG decreases, due to the lowering of the Λ_{QCD} value on the MG side, while the number of surviving F states remains constant, since the shower Sudakov is not modified. The total estimated cross section therefore decreases.
2. At the differential level, the smaller number of $(F+1)$ states combined with the unchanging number of F states implies smaller absolute jet cross sections and smaller fractions $\sigma_{\mathrm{jet}}/\sigma_{\mathrm{tot}}$.

Similarly we may consider what happens if C) we reduce the Λ_{QCD} value on the SG side instead, as summarised in the last column of tab. 1. Going from case A to C, the following changes result:

1. The number of $(F+1)$ states added by the MG remains constant, while the number of surviving F states increases, since the SG is generating fewer branchings. The total estimated cross section therefore *increases*.
2. Since the number of $(F+1)$ states is constant, while the shower is made less active, the final jets will actually be narrower, which *increases* the rate of reconstructed jets at any given fixed p_T value.
3. Since both the total cross section increases and the number of reconstructed additional jets also increases, jet *fractions* can either increase or decrease.

In particular, note the somewhat counter-intuitive effect that *decreasing* the shower α_s value actually *increases* the jet rates in a matched calculation, while it normally *decreases* them in a standalone shower calculation.

Since, as was discussed above, inconsistencies among the choices on the two sides can lead to differences at the NLL level, it is obviously important to ensure that they are consistent within a reasonable margin. This is particularly true in the context of event-generator tuning, in which specifically the NLL components of the shower description are sought to be optimized with respect to measured data, and hence changes at this level could effectively destroy the tuning.

Finally, we remind the reader that a change in Λ_{QCD} can be interpreted as a change in the opposite direction of the renormalization scale argument (for constant Λ_{QCD}), modulo small flavour threshold effects that we

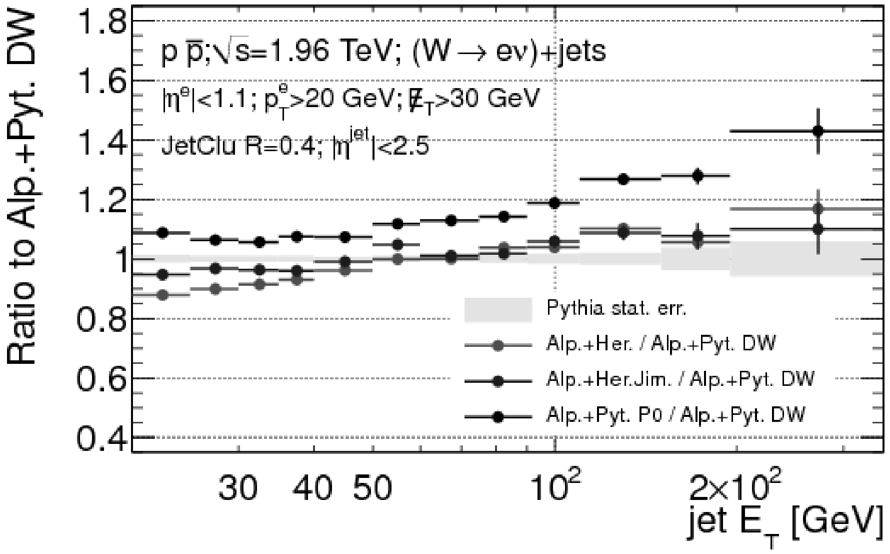

Figure 1: Ratio of predictions for the leading-jet E_T spectrum in W+jets final states at the Tevatron, obtained with AlpGen plus various MC codes and tunes. The leading jet observable is defined at the particle-level as in the CDF W+jets analysis [32].

shall ignore here. This is easy to realise from the definition of the coupling,

$$\alpha_s(k\mu^2) \overset{1-\text{loop}}{=} \frac{1}{b_0 \ln\left(k\mu^2/\Lambda^2\right)} = \frac{1}{b_0 \ln\left(\mu^2/(\frac{1}{k}\Lambda^2)\right)} . \tag{4}$$

Thus, we may write renormalization scale variations (e.g., by a factor of 2 in each direction) either by applying a prefactor directly on the renormalization scale argument of α_s or by applying the inverse of that factor to Λ_{QCD} while keeping the renormalization scale argument unchanged. Due to the technical structure of the codes, the former is more convenient in AlpGen (via the `ktfac` setting) whilst the latter is more convenient for Pythia 6.

2.2 Examples of the interplay between tunes and matching

In this section we give several examples of how the issues in ME-PS matching described in Section 2 can affect high-p_{T} observables using AlpGen interfaced to Pythia 6 with DW [7], Perugia 0 (P0) [5, 31] and Perugia 2010 (P2010) tunes [6].

In Fig. 1 we show the ratio of predictions for the transverse energy (E_T) spectrum of the leading jet (that jet with the highest E_T per event) in W+jet final states at the Tevatron, obtained by the merging of AlpGen with different shower codes; Herwig, Pythia 6 virtuality-ordered shower (DW), and p_{T}-ordered shower (Perugia 0). The differences between Herwig and Herwig plus Jimmy at small E_T can be explained by the different amounts of energy that, in the various cases, are deposited by the UE in the jet cones. In particular, as shown in Fig. 2, these differences can accommodate the slight shape discrepancy between data and AlpGen +Herwig that was noted, at small E_T, in the CDF study [32]. It is difficult, however, to attribute to the UE energy the significant differences seen in Fig. 1 at large E_T.

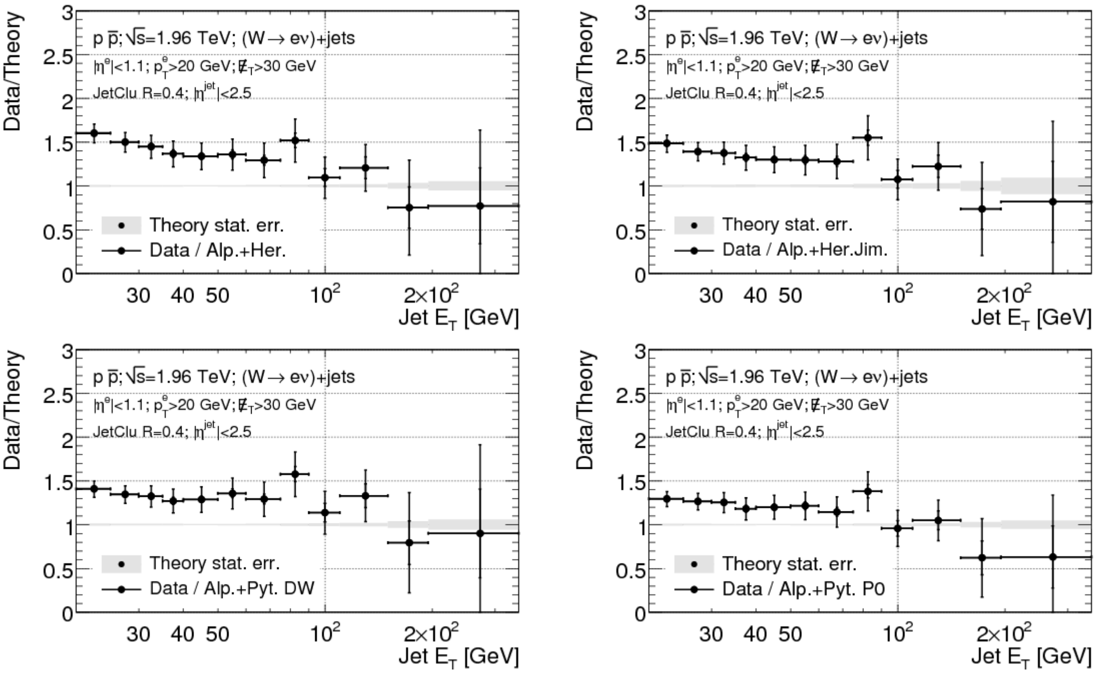

Figure 2: Comparison of CDF data [32] with the leading-jet E_T spectrum predicted by AlpGen plus various MC codes and tunes.

In order to investigate the source of the differences in the predictions, systematic parameter variations of the perturbative and non-perturbative model components of Pythia 6 have been studied using the Perugia family of Pythia 6 tunes with Perugia 0 as the central tune and Perugia Hard and Perugia Soft as the systematic variation tunes. The Perugia Soft and Perugia Hard tunes both use the same Parton Density Function (PDF) as Perugia 0, CTEQ5L [33], but differ in the values of Pythia 6 parameters controlling both perturbative and non-perturbative activity levels. In comparison to Perugia Soft, the Perugia Hard tune has more perturbative (initial and final state radiation) activity but less non-perturbative (multiple interactions, beam remnant and hadronization) activity. Perugia Soft on the other hand has less perturbative but more non-perturbative activity than the Perugia 0 tune. In order to investigate the interplay of the tuning variations with the MLM matching, the effect of the change of the tune on the physics observables in both the Pythia 6 standalone case and AlpGen + Pythia 6 case is presented.

In Fig. 3 the distribution of jet multiplicity (N_{jet}) in W+jets events is compared for events generated with the Perugia 0, Perugia Hard and Perugia Soft tunes. The N_{jet} observable is defined at the particle-level according to the definition used in the ATLAS measurement of the W+jets cross-section at $\sqrt{s}$=7 TeV [34]. Jets are clustered from stable particles using the anti-Kt jet algorithm [35] with the radius parameter $R = 0.4$, and considered in case they satisfy the following kinematic cuts: $p_{\text{T}} > 20$ GeV and $|\eta| < 2.8$. Comparisons are performed for both the Pythia 6 standalone (left) and AlpGen + Pythia 6 (right) cases. For the Pythia 6 standalone case we observe that Perugia Hard tune yields more high-p_{T} jets than the Perugia 0 tune while Perugia Soft yields less final state jets correspondingly. For the AlpGen + Pythia 6 case an opposite trend is observed: Perugia Hard tune yields less high-p_{T} jets and Perugia Soft tune yields more high-p_{T} jets. In order to determine which modelling components of the Perugia Soft and Perugia Hard tunes cause this behaviour, we considered the effect of varying individual sets of parameters of the Perugia Soft and Perugia Hard tunes in AlpGen + Pythia 6 predictions. Parameters were grouped according to the modelling aspect they control into Initial State Radiation (ISR), Final State Radiation including the FSR from the ISR partons (FISR), the Underlying Event (UE) and Colour Reconnections (CR) blocks[2]

Dedicated samples where only parameters of an individual block were varied in the ranges used in Perugia Soft and Perugia Hard tunes on top of the Perugia 0 tune were produced. The results of the study, in terms of the cross-section contribution of each AlpGen sub-sample (after MLM matching), are given in Table 2. As was already noted in Fig. 3, the cross-section for multijet production in the Perugia Hard case decreases with respect to Perugia 0, and vice versa for Perugia Soft. From Table 2, we see that the parameter blocks that produce this affect are the ISR and FISR blocks, while the impact of the CR and UE block variations on the cross-sections is negligible. In addition to the simultaneous variations of parameters in the blocks, we have also performed individual parameter variations for each of the parameters in order to check that potential correlations between the parameters do not affect the conclusions. Studies have also been performed for the Hadronization and Beam Remnant blocks of [6]. The variations of these parameters also had a negligible effect on the kinematic distributions and cross-section values.

In Fig. 4 we demonstrate that the increased parton shower activity can indeed lead to the reduced cross-section (and softer jet spectra) due to the increased rates at which the AlpGen + Pythia 6 events are vetoed during the MLM matching. In the figure the distributions of the events that pass or fail (ISVETO=0 or ISVETO$\neq$0) the MLM matching criterion are shown for the exclusive sub-sample of AlpGen + Pythia 6 Perugia 2010 W+jets events with exactly three additional partons from the matrix element in the final state[3]. Each of the distributions is normalised to unit area. The distributions are shown as a function of the largest p_{T} shower emission from the initial state radiation (left) and as a function of the largest p_{T} multiple proton-proton interaction. [4] In the left hand side figure we see that the events are rejected with higher probability, the larger the p_{T} of the hardest ISR branching in the event. Therefore, a Pythia 6 standalone tune which increases the ISR activity can, somewhat counter-intuitively, reduce the rate for multijet and hard emissions. In the right hand side we demonstrate that the events are accepted and rejected independently of the transverse momentum of the hardest multiple interaction in the event (which is the desired behaviour of the matching application used with the parton shower code).

To conclude, the origin of the differences observed in the predictions of tunes with different ISR/FSR activity matched to AlpGen is rather due to the mismatch between the jet-emission probability predicted by the matrix elements and by the shower. This comes from the mismatch in the value of α_S discussed earlier, arising from different values of Λ_{QCD} or from the use of a different evolution variable in the shower. If the value of α_S in the shower increases, the emission rate of additional jets during the shower evolution will increase. Since the matching algorithm rejects events with extra jets generated by the shower, to replace them with events where the jet is accounted for by a higher-order matrix element calculation, a larger value of α_S in the shower leads to a higher rejection rate. Unless this change in α_S is accompanied by a similar change in the matrix element calculation, the additional rejection is not compensated by the relative increase in rate for the higher-order parton-level contributions, leading to the effects reported in this section.

This important interplay between MC parameters, which are typically tuned to "soft" observables such as UE or the small-p_{T} DY spectrum, and the performance of the matching algorithms for "hard" observables, calls for particular attention when adopting new UE tunes in the framework of multijet studies with matrix-element matching. Along the same lines, it should be kept in mind that, tuning a stand-alone shower MC to

---

[2] The parameter blocks organisation is similar to the one introduced in [6] and are listed in A.2

[3] The observations in the text are largely independent on the final state parton multiplicity

[4] These p_{T} values are reported by Pythia 6 parameters `VINT(357)` (ISR) and `VINT(359)` (MPI) respectively.