

RECEIVED: November 11, 2015

REVISED: January 18, 2016

ACCEPTED: February 1, 2016

PUBLISHED: February 19, 2016

# A study of multi-jet production in association with an electroweak vector boson

R. Frederix,<sup>a</sup> S. Frixione,<sup>b</sup> A. Papaefstathiou,<sup>c</sup> S. Prestel<sup>d</sup> and P. Torrielli<sup>e</sup>

<sup>a</sup>*Physik Department T31, Technische Universität München,  
James-Frank-Str. 1, Garching, D-85748 Germany*

<sup>b</sup>*INFN, Sezione di Genova,  
Via Dodecaneso 33, Genoa, I-16146 Italy*

<sup>c</sup>*PH Department, TH Unit, CERN,  
Geneva 23, CH-1211 Switzerland*

<sup>d</sup>*SLAC, National Accelerator Laboratory,  
2575 Sand Hill Road, Menlo Park, CA, 94025-7090 U.S.A.*

<sup>e</sup>*Dipartimento di Fisica, Università di Torino and  
INFN, Sezione di Torino,  
Via P. Giuria 1, Turin, I-10125 Italy*

*E-mail:* [rikkert.frederix@tum.de](mailto:rikkert.frederix@tum.de), [stefano.frixione@cern.ch](mailto:stefano.frixione@cern.ch),  
[apapaefs@cern.ch](mailto:apapaefs@cern.ch), [prestel@slac.stanford.edu](mailto:prestel@slac.stanford.edu), [torriell@to.infn.it](mailto:torriell@to.infn.it)

**ABSTRACT:** We consider the production of a single  $Z$  or  $W$  boson in association with jets at the LHC. We compute the corresponding cross sections by matching NLO QCD predictions with the HERWIG++ and PYTHIA8 parton showers, and by merging all of the underlying matrix elements with up to two light partons at the Born level. We compare our results with several 7-TeV measurements by the ATLAS and CMS collaborations, and overall we find a good agreement between theory and data.

**KEYWORDS:** NLO Computations, Hadronic Colliders

**ARXIV EPRINT:** [1511.00847](https://arxiv.org/abs/1511.00847)

---

## Contents

<b>1</b>	<b>Introduction and motivation</b>	<b>1</b>
<b>2</b>	<b>Technicalities and settings</b>	<b>3</b>
<b>3</b>	<b>Comparison to data</b>	<b>6</b>
3.1	$Z$ +jets	10
3.2	$W$ +jets	19
3.3	Inclusive observables	30
<b>4</b>	<b>Summary</b>	<b>35</b>

---

## 1 Introduction and motivation

The production of jets in association with a  $Z$  or a  $W$  boson plays a role of primary importance at high-energy hadron colliders. On the one hand, when regarded as backgrounds either or both of these processes are relevant to the determination of the properties of the Standard Model (SM) Higgs, of the top quark, and of the gauge sector of the SM, as well as playing a major role in many searches for new physics beyond the SM. On the other hand, they are also interesting in their own right, because they constitute an excellent testing ground for theoretical predictions. This stems from a variety of reasons: the high statistical accuracy of the measurements performed and foreseen by the LHC collaborations; the fact that these processes, especially if leptonic channels are considered, are simpler from a theoretical viewpoint than other SM processes with a similar number of final-state particles (e.g.  $t\bar{t}$ +jets); and because their complicated kinematics allow one to probe phase-space regions sensibly described either by Monte Carlo event generators (MCs henceforth), or by fixed-order calculations.

The features just mentioned are what render  $Z$ +jets and  $W$ +jets production the primary processes for the validation of multi-jet merging approaches. As a technical aside, we remind the reader that a matching formalism aims at including, in a consistent manner and in particular without double counting, the results of matrix elements computations into MCs, at a given order in perturbation theory and at a given number of final-state particles at the Born level. Conversely, a merging formalism combines several matched samples, which are identified by different parton multiplicities at the Born level. This implies that, by merging, one extends the scope of matched results towards kinematic regions defined by many, well-separated jets. These are indeed well studied by LHC collaborations: even in Run I, jet multiplicities exceeding five have been measured, with transverse momenta reaching the TeV range; in Run II, these impressive achievements will be easily surpassed. A good understanding of these characteristics, through the capability of merged predictions

to reproduce data, is vital for many new-physics searches, which are very often heavily reliant on theoretical predictions, and categorised in terms of jet multiplicities. Theoretically, a significant amount of information is involved in merged results, obviously including the merging formalisms themselves; detailed comparisons to data will help tell the various proposals apart, and ultimately suggest improvements. One also expects to become sensitive to large-logarithm effects of electroweak origin [1–4], whence the necessity of going beyond the present results, which are based on QCD computations.

Until recently, merging formalisms have relied on underlying computations of LO (i.e. tree-level only) accuracy. Techniques such as CKKW, CKKW-L, MLM, and their variants [5–14] have been systematically compared to each other (see e.g. ref. [10] for a study focusing on  $W$ +jets production), and to data. The extension of merging to NLO accuracy is a challenging theoretical problem [15–26]. Comprehensive comparisons among different approaches are so far lacking, and it is only lately that experimental collaborations have started to employ these results, with the idea of using them to replace gradually the LO-accurate ones. Of course, this assumes that mergings at the NLO will improve the description of the data given by their LO counterparts.

The aim of this paper is to use  $Z$ +jets and  $W$ +jets LHC data for a phenomenology validation of the FxFx NLO-merging formalism [21]. In order to rely on well-understood measurements, and to avoid making mistakes in the involved definitions of cuts and observables as defined by the experiments, we have limited ourselves to considering the 7-TeV analyses by ATLAS and CMS based on the full dataset (stemming from an integrated luminosity of about  $5 \text{ fb}^{-1}$ ), and associated with officially-supported Rivet [27] routines. Specifically, the focus of this paper will be the  $Z$ +jets results of ATLAS [28] and CMS [29, 30], and the  $W$ +jets results of ATLAS [31] and CMS [32]. We shall also briefly deal with the inclusive and the underlying-event analyses of refs. [33, 34], in order to raise some points of potential future interest that concern the interplay of tunable parameters in MCs with (NLO) matching and merging. In all cases, our predictions are obtained by means of the fully automated MADGRAPH5\_AMC@NLO framework [35]; parton showers are simulated with HERWIG++ [36–39] and PYTHIA8 [40, 41].

We conclude this section by giving the briefest possible introduction to the FxFx approach; the interested reader can find fuller details in the original paper [21] as well as in ref. [35]. FxFx is based on the MC@NLO matching procedure [42]. MC@NLO samples are constructed for the processes whose Born-level contributions are:

$$\mathcal{I}_1 + \mathcal{I}_2 \longrightarrow S + i \text{ partons}, \quad (1.1)$$

with  $i \geq 0$ , and where  $S$  is a set of  $p$  particles which does not contain any QCD massless partons. In the present case,  $S = \ell^+\ell^-$  or  $S = \ell\nu_\ell$  (with  $\ell = e$  or  $\mu$ ) for  $Z$ +jets and  $W$ +jets production respectively; by considering lepton pairs already at the matrix-element level, one takes into account exactly spin-correlation and off-shell effects. The samples corresponding to eq. (1.1) are defined in such a way that hard emissions are suppressed by means of a function that can be parametrised in terms of a hard mass scale (called the (FxFx) merging scale), except in the case of the largest multiplicity (i.e. the largest  $i$ ) considered. In other words, hard partons are mostly seen as entering a Born-level contribution (say, at  $i = i_0$ ),

rather than the real-emission correction to the previous multiplicity ( $i = i_0 - 1$ ). Such a suppression, enforced at the level of matrix elements, must be accompanied by suitable choices of shower scales. On top of this, matrix elements are also multiplied by appropriate Sudakov factors. Showered events are subject to an MLM-type rejection, quite analogous to the original one [10], except for the fact that a  $k_T$  jet-finding algorithm [43] is adopted, as is already the case in the LO MADGRAPH5 implementation [44]. The only difference w.r.t. an LO treatment is the fact that, at the NLO, the rejection procedure must be based on a jet-jet matching, rather than on a parton-jet one, in order to preserve IR safety. Finally, we point out that FxFx is non-unitary. In other words, the total rate resulting from an FxFx-merged sample is not necessarily equal to the total rate of the inclusive sample of lowest multiplicity, even before any final-state cuts are applied. While the rate thus obtained is, like any other observable, merging-scale dependent, it constitutes a genuine prediction of the method, in that it incorporates some of the contributions to the inclusive  $K$  factor at orders higher than NLO.

This paper is organised as follows: technical details are reported in section 2; we present our predictions and the comparisons to data in section 3 (section 3.1 for  $Z$ +jets, section 3.2 for  $W$ +jets, and section 3.3 for inclusive and small- $p_T$  observables); we conclude in section 4.

## 2 Technicalities and settings

The hard-event samples have been obtained with MADGRAPH5\_AMC@NLOv2.2.1 [35]. The whole procedure is fully automated: for either  $Z$ +jets or  $W$ +jets production, a single file of *unweighted* hard events is generated, and is subsequently showered by an MC. No post-processing (e.g. rescaling and combining samples relevant to different parton multiplicities) is necessary. The MLM rejection is performed on the fly, on a event-by-event basis, by the MCs. Event files are created with the following commands (here given for  $Z$ +jets production in order to be definite):

```
./bin/mg5_aMC
MG5_aMC> import model loop_sm-no_b.mass
MG5_aMC> define p = p b b~; define j = p
MG5_aMC> define l+ = e+ mu+; define l- = e- mu-
MG5_aMC> generate p p > l+ l- [QCD] @ 0
MG5_aMC> add process p p > l+ l- j [QCD] @ 1
MG5_aMC> add process p p > l+ l- j j [QCD] @ 2
MG5_aMC> output; launch
```

with `ickkw= 3` in `run_card.dat`.<sup>1</sup> The commands in the first two lines instruct MADGRAPH5\_AMC@NLO to perform a five-flavour computation, where the  $b$  quark is treated as massless and may appear in the initial states of partonic subprocesses. In the case of  $W^+W^-$  production (i.e. the two  $W$ 's are generated simultaneously), one simply replaces

<sup>1</sup>This setting tells the code that FxFx merging is to be used. For UNLOPS [24] one sets `ickkw= 4`.

the  $l^+ l^-$  pair above with  $l = e, \mu$ , with  $l = e^+, \mu^+, e^-, \mu^-$  and  $\nu l = \nu_e, \nu_\mu, \nu_{e^c}, \nu_{\mu^c}$ . As these command lines imply, all of our simulations are based on underlying matrix elements computed at the NLO accuracy, with up to two extra partons at the Born level (i.e. the highest final-state multiplicity considered at tree-level is equal to two leptons and three QCD partons): we do not consider any LO-accurate matrix elements for multiplicities larger than those that enter our NLO computations (i.e. four or higher). In the case of  $Z$ +jets production, both  $Z$  and  $\gamma^*$  contributions, as well as their interference, are included. For showering, PYTHIA8.210 [41] and HERWIG++2.7.1 [39] have been used. As far as PYTHIA8 is concerned, v8.210 embeds FxFx-related modules,<sup>2</sup> and includes all of the fixes to previous un-validated versions of those. In particular, the prescription on how to match the pseudo-jets of the short-distance cross section to jets after parton showering has now been aligned to that of HERWIG++. Furthermore, the numerical stability of the beam-remnant handling was improved to cope with some very small Bjorken- $x$  values that are induced by NNPDF-derived tunes. In HERWIG++, all FxFx modules are treated as an external plugin,<sup>3</sup> fully analogous to a generic user-defined analysis; thus, such modules must be compiled and loaded through the standard HERWIG++ input file.

The hard event files, as customary in MADGRAPH5\_AMC@NLO, contain additional weights that allow one to compute the hard-scale and PDF uncertainties according to the reweighting procedure introduced in ref. [45]; such weights are stored as dictated by the LHA v3.0 format [46]. The FxFx-specific modules of both PYTHIA8 and HERWIG++ are fully compatible with such a structure, and thus can handle all of the additional weights at the same time as the main event weight. Unfortunately, due to limitations in the present Rivet release, that package can only deal with one weight at a time; hence, at the level of analysis we have been forced to launch a separate Rivet run for each of the additional weights (therefore repeating exactly the same operations multiple times, since the kinematics of the event does not change). Although these operations are highly parallelisable, they could be avoided simply by giving Rivet the possibility of filling multiple histograms with the same kinematics and different weights.

In the computations of *the matrix elements* we have set the most relevant parameters as follows:

- $m_Z = 91.188$  GeV,  $m_W = 80.419$  GeV.
- $\Gamma_Z = 2.441$  GeV,  $\Gamma_W = 2.0476$  GeV.
- Lepton masses ( $e^\pm$  and  $\mu^\pm$ ) set equal to zero.<sup>4</sup>
- NNPDF2.3 [47] NLO PDFs (LHAGLUE number 230000), which also set the value of  $\alpha_s(m_Z) = 0.119$ . We work in a five-flavour scheme.

<sup>2</sup>FxFx merging in PYTHIA8 takes advantage of the available `UserHooks` facilities, meaning that the merging is (almost) completely decoupled from the event generator; note that all MLM-inspired merging methods are handled in such a way.

<sup>3</sup>These modules will be available in the forthcoming release of HERWIG++. In the meanwhile, they can be obtained upon request to the authors of this paper.

<sup>4</sup>The MCs will then put the charged leptons on their physical mass shells, but genuine lepton-mass effects are not included. We expect them to be utterly negligible for the observables we have considered.

- Central hard-scale choice:  $\mu_0 = H_T/2$ , with  $H_T$  the scalar sum of the transverse masses  $\sqrt{p_T^2 + m^2}$  of all final state particles. In merged samples, this is computed by using the  $2 \rightarrow V + 1$  process reconstructed by clustering back the given kinematic configuration;<sup>5</sup> on top of that, an  $\alpha_S$  reweighting is performed — see ref. [21] for more details.
- Hard-scale variations: independent,  $1/2\mu_0 < \mu_R, \mu_F < 2\mu_0$ .
- FxFx merging scales:<sup>6</sup>  $\mu_Q = 15, 25, \text{ and } 45 \text{ GeV}$ .
- $1/\alpha_{EM} = 132.507$ .
- Diagonal CKM matrix.

Most of the above settings may be checked by inspecting the headers of the hard event files, where they are included through copies of the relevant input cards. The PDFs have been obtained by means of the LHAPDF6 [48] package. Our simulations are based on 15M unweighted events for the sum of the two leptonic channels in  $Z$ +jets and in  $W$ +jets production. Given that lepton-mass effects are ignored here, the electron and muon channels basically account for half of those events each. In order to have a benchmark independent of merging, a further 5M events for the sum of the two decay channels have been generated for the inclusive MC@NLO sample of lowest multiplicity ( $i = 0$  in eq. (1.1)). We shall refer to these events as the “(fully) inclusive” samples. We finally recall that FxFx makes use of a  $k_T$ -clustering algorithm [43], with  $R = 1$ , both at the short-distance level and after parton showering. A jet at short distance is said to match a jet after parton showering if the  $k_T$  jet algorithm with minimal separation  $\mu_Q$  would combine two such objects into a single pseudo-jet.

In the MCs, the parameters have been set to their defaults<sup>7</sup> as defined by the MADGRAPH5\_AMC@NLO interface (see `amcatnlo.cern.ch`, “Special settings for the parton shower” under the “Help and FAQ” item), or by the native defaults. We point out that this implies, in particular, that the values of  $\alpha_S(M_Z)$  adopted by PYTHIA8 (equal to 0.1365 for initial- and final-state radiation, and to 0.130 for multiparton interactions) and by HERWIG++ (equal to 0.118, and internally evolved at two loops with a native scheme), as dictated by the respective tunes chosen here (see below), have not been modified. On the other hand, in the showers the same PDFs as in the hard-matrix element computations have been used, in order to guarantee that the formal NLO accuracy of the underlying MC@NLO simulation be respected. The underlying event tunes are Monash 2013 [49] in

<sup>5</sup>The scale  $\mu_0$  may not vanish, owing to the hard cuts imposed on the final-state leptons.

<sup>6</sup>We note that the merging scale need not be smaller than the minimal jet transverse momentum imposed in the analysis, and it is actually desirable that its range includes the latter. More details can be found at pages 56–57 of ref. [35], as well as at pages 18–19 of ref. [21].

<sup>7</sup>FxFx dictates shower-scale choices, as explained in ref. [21]. As far as fully-inclusive samples are concerned, the shower scales are assigned according to the MADGRAPH5\_AMC@NLO defaults (see pages 48–49 of ref. [35]); we point out that the latter are partly under the user’s control through the input parameter `shower_scale_factor`.

PYTHIA8, and UE-EE-3-CTEQ6L1 [37] in HERWIG++. Since neither of them has been obtained with NNPDF2.3 NLO PDFs, the quality of the comparisons to low- $p_T$  data might be degraded w.r.t. one performed by using the MCs standalone in conjunction with the same PDFs as in the tunes. While this is a common issue with NLO+PS simulations, and one that will be addressed possibly only in the context of a genuine NLO tune, the impact on  $Z$ +jets and  $W$ +jets observables is expected, and will be shown, to be rather minimal. We shall discuss this point in more details in section 3.3.

### 3 Comparison to data

This section contains the main results of this paper. We present  $Z$ +jets observables in section 3.1,  $W$ +jets observables in section 3.2; inclusive and underlying-event results are reported in section 3.3. In the remainder of this preamble we give some general information, and discuss features common to all analyses.

**Meaning of leptons.** In several of the Rivet routines we have used, results are obtained for both the bare and the dressed leptons. In the following, we shall restrict ourselves to considering only the latter, in view of their being more inclusive in QED radiation, and thus less sensitive to the modeling of QED showers in MCs; in general, however, the differences between bare and dressed predictions are fairly small. Consistently with such a choice, in our simulations both HERWIG++ and PYTHIA8 do feature QED showers; the input parameters that control them have been left equal to their default values as given by the MC authors.

**Event generation and efficiencies.** The hard jet cuts and widely different kinematic configurations relevant to  $Z$ +jets and  $W$ +jets production imply that it is generally time-consuming to accumulate sufficient statistics in MC simulations. We point out that, in the context of merging approaches that feature an MLM-type rejection procedure, the efficiency for finding a jet-jet match and hence not to reject an event is larger the larger are the generation cuts (with the obvious and usual condition that such cuts do not bias the physical results). When one makes several choices for the merging scale, there is therefore a trade-off between generating as many event samples as merging-scale choices with maximally-efficient generation cuts, and generating a single event sample, where the generation cuts are so that they do not bias any of the physics results obtained with the different merging scales. In order to reduce the hard-event generation time, and the number of associated files, we have adopted the latter option. However, we point out that the former option is a perfectly reasonable one too, which may actually be more convenient with large-scale computer clusters (where multiple hard-event generations can be launched in parallel). In the generation of our hard events, we have imposed  $p_T(j) \geq 8 \text{ GeV}$ , without any restrictions on jet rapidities;<sup>8</sup> the invariant mass of opposite-charge lepton pairs has

---

<sup>8</sup>Jets are obtained with an  $R = 1$   $k_T$ -clustering algorithm. We remark that in principle minimum- $p_T$  cuts are not mandatory in the Sudakov-reweighted version of the FxFx merging [21] (which is the default, used here and recommended for phenomenology applications), although they help increase significantly the generation efficiency.

	$\mu_Q = 15 \text{ GeV}$	$\mu_Q = 25 \text{ GeV}$	$\mu_Q = 45 \text{ GeV}$
HERWIG++	44%(2.7)	38%(3.2)	35%(3.5)
PYTHIA8	45%(4)	37%(4)	32%(4)

**Table 1.** Efficiencies of the MLM-type rejection in FxFx merging, rounded to the percent; in brackets, we report the corresponding oversampling factors (see the text for details).

been constrained to be larger than 40 GeV. The efficiencies (i.e. the probability that events not be rejected in the MLM procedure) resulting from these cuts that we have measured in our simulations are reported in table 1. On top of those, one needs to take into account the fact that a further loss of statistics is entailed by the presence of negatively-weighted events. The fraction of these is equal to about 25% in the FxFx samples for both  $Z$ +jets and  $W$ +jets production (for comparison, it is less than 10% in the fully-inclusive samples). In order to improve the statistics of the final physics plots, we have *oversampled* our hard events. In other words, the same events have been showered more than once, with care being taken to change the seeds that govern parton showers — in this way, a given hard event MLM-rejected with a certain shower configuration might be accepted by generating a different shower configuration. Of course, in order not to bias the final results the oversampling must not be too large; in this paper, we have used what is reported in round brackets in table 1 (a non-integer value implies that only a fraction of the hard-event file is considered; note that such a file is randomised at generation time).

**Choice of parameters in MC simulations.** Before turning to presenting our predictions, we briefly return to the fact that, as discussed in section 2, there are certain small inconsistencies in the parameter choices made at the level of hard matrix elements and in the parton showers; this is common in the context of NLO+PS simulations. In particular, we are concerned here with  $\alpha_s$  and the PDFs. For parton showering, we have set the  $\alpha_s(m_Z)$  input value equal to that relevant to the MC tunes, rather than as prescribed by the PDFs (note that either choice guarantees the NLO accuracy of the results). In the case of HERWIG++, the difference is negligible (0.118 vs 0.119), while it is larger<sup>9</sup> in the case of PYTHIA8 (0.1365 vs 0.119). We have verified that, by setting  $\alpha_s(m_Z) = 0.130$  in the HERWIG++ showers, the effects on some observables are sizable, and the agreement with data worsens. A similar degradation of the theory-data agreement is seen by choosing  $\alpha_s(m_Z) = 0.119$  in the PYTHIA8 showers, with lower  $\alpha_s$  values leading to lower jet rates and softer  $p_T$  spectra than in data. Thus, the conclusion of this heuristic study is that it appears to be best to set  $\alpha_s(m_Z)$  in the MCs equal to the value(s) that result(s) from tuning. As far as the PDFs are concerned, we have also showered events by choosing the LO version of the NNPDF2.3 sets (which formally spoils the NLO accuracy of the computations); some differences w.r.t. our standard simulations do appear also in this case, restricted to small- $p_T$  regions, but are generally much smaller than in the case of  $\alpha_s$  variations (with the

---

<sup>9</sup>However, note that HERWIG++ uses the Monte Carlo scheme [50] that, through a rescaling of  $\Lambda_{\text{QCD}}$  as obtained from the input  $\alpha_s(m_Z)$ , effectively gives larger values of  $\alpha_s(Q)$ , which are thereby closer to those of PYTHIA8.



	$\mu_Q = 15 \text{ GeV}$	$\mu_Q = 25 \text{ GeV}$	$\mu_Q = 45 \text{ GeV}$	inclusive	
$Z+\text{jets}$	2.055(−0.9%)	2.074	2.085(+0.5%)	2.012(−3.0%)	HW++
	2.168(+0.8%)	2.150	2.117(−1.5%)	2.011(−6.5%)	PY8
$W+\text{jets}$	20.60(−0.9%)	20.78	20.87(+0.4%)	19.96(−3.9%)	HW++
	21.71(+1.0%)	21.50	21.18(−1.5%)	19.97(−7.1%)	PY8

**Table 2.** Total rates (in nb) for the three different choices of the FxFx merging scale, as well as those for the inclusive (i.e. non-merged) samples, obtained with HERWIG++ (upper rows) and PYTHIA8 (lower rows). Relative differences w.r.t. the FxFx results obtained with the central merging scale are also reported in brackets.

exception of the underlying event analysis of ref. [34]). While this is reassuring, we point out that, given the correlations between  $\alpha_s$ , the PDFs, and the low-energy parameters that are set when tuning an MC, a more consistent theoretical treatment might necessitate beyond-LO tunes; we shall further comment on this point in section 3.3.

**Total inclusive rates.** In table 2 we report the predictions for the total rates (hence, independent of final-state cuts and jet definitions) that result from the FxFx-merged and the fully-inclusive samples; the latter are, by construction in the MC@NLO formalism, equal to those one obtains with fixed-order computations — indeed, the PYTHIA8 and HERWIG++ results in the last column agree to a 0.05% level, which is the statistical inaccuracy one expects from a 5M-event sample. There are two features in table 2 which are particularly worth remarking. Firstly, the merged results obtained with different merging scales are very close to each other. This gives one confidence on the fact that merging-scale systematics is under control, in spite of the large range chosen for  $\mu_Q$  variations. Secondly, the merged rates are a few percent larger than the fully-inclusive one, with the exact amount depending on the MC adopted for showering. This is a manifestation of the non-unitary behaviour of FxFx, and the MC-dependent amount of “unitarity violation” should be seen as an actual prediction associated with the given MC. On the other hand, the differences w.r.t. the fully-inclusive cross sections are not large, which is perfectly compatible with the moderate NNLO  $K$  factors for  $Z$  and  $W$  hadroproduction. We shall see that the small increase of the merged cross sections w.r.t. the inclusive ones is beneficial in terms of the comparisons to data.

**Normalisation of results.** The features just mentioned, and the predictivity they underpin, help us stress the following point. All of our predictions are reported with their native normalisation: in other words, *no rescaling* has been performed. While an overall re-normalisation by a constant (e.g. the NNLO/NLO fully-inclusive  $K$  factor) common to all observables is acceptable, we believe that the practice of rescaling theoretical results by factors that depend on the jet multiplicity leads to confusion, and especially when such a multiplicity is understood in the inclusive sense. Although by so doing one generally makes theory-data comparisons look better, one also tends to neglect the fact that merged results, especially at the NLO, are supposed to be predictive for both shapes and rates. At the very least, a rescaling dependent on the jet multiplicity renders it more difficult

to understand the strengths and weaknesses of a given merging approach, and to assess the overall predictivity of different merging techniques. The latter problem is clearly more acute in the case where the rescaling factors exhibit a non-negligible dependence on the jet multiplicity, and/or on the particular MC considered. As an example of both of these issues, we refer the reader to table 7, appendix A, of ref. [31], where the results of several state-of-the-art simulations are reported. From a purely theoretical viewpoint, the effects of a multiplicity-dependent rescaling have been considered e.g. in ref. [24]; figure 6 there, in particular, shows that such a rescaling may induce a large increase of the theoretical systematics, for cross sections characterised by large  $K$  factors.

**Generalities concerning differential observables.** All of the figures we shall present below have the same layout. In the main frame, the data are displayed as full black circles, with bars representing the associated errors as quoted by the experiments. The FxFx results are shown as green bands (labelled by “Var” in the plots), which give the full span of the theoretical uncertainties considered here: these are due to the variations of the hard scales, of the PDFs, and of the merging scale. Such bands have been obtained as follows. For a given merging scale, the envelope is constructed for the hard-scale dependence bin-by-bin, by taking the maximum and the minimum in each bin among the predictions associated with all possible hard-scale choices. Similarly, one constructs the PDF-dependence envelope according to the prescription of the PDF authors [47]. The two envelopes thus obtained are combined in quadrature, and the result is therefore the theoretical uncertainty band associated with the given merging scale. Finally, the envelope of these theoretical uncertainties relevant to the three merging scales is constructed. In the main frame we also report the central MC@NLO result for the fully-inclusive samples (i.e. non-merged) as a solid red histogram (labelled by “inc”). Below the main frame there are two insets. The upper one displays the fractional experimental errors as a yellow band. The ratios of the theoretical predictions over data are given as green bands (for the FxFx results) and as a solid red histogram (for the fully-inclusive MC@NLO results). The central FxFx predictions (i.e. obtained with reference hard scales and PDFs, and  $\mu_Q = 25$  GeV), divided by data, are displayed as full green triangles. Finally, dark-green error bars represent the fractional statistical uncertainties, in the case of  $\mu_Q = 25$  GeV. In the lower inset we show the ratios of the upper and lower end of the hard-scale-plus-PDF envelopes, over the central result; these are therefore theoretical fractional uncertainties, displayed as solid blue, dashed magenta, and dot-dashed light blue histograms for  $\mu_Q = 15, 25,$  and  $45$  GeV, respectively.

In order to facilitate the comparisons with published results, the observables in the plots are strictly labelled as in the original papers (i.e. the labels are inherited from the relevant Rivet routines). We warn the reader that this might imply some inconsistencies in the notations of different analyses.

Before turning to commenting the comparisons to  $V$ +jets data, we stress the following point: the observables we are interested in are characterised by a number of jets ( $N_{\text{jet}}$ ) that accompany the vector boson equal to or larger than one. For these, the MC@NLO fully-inclusive predictions are at most of LO accuracy, when  $N_{\text{jet}} = 1$ ; for  $N_{\text{jet}} > 1$ , they

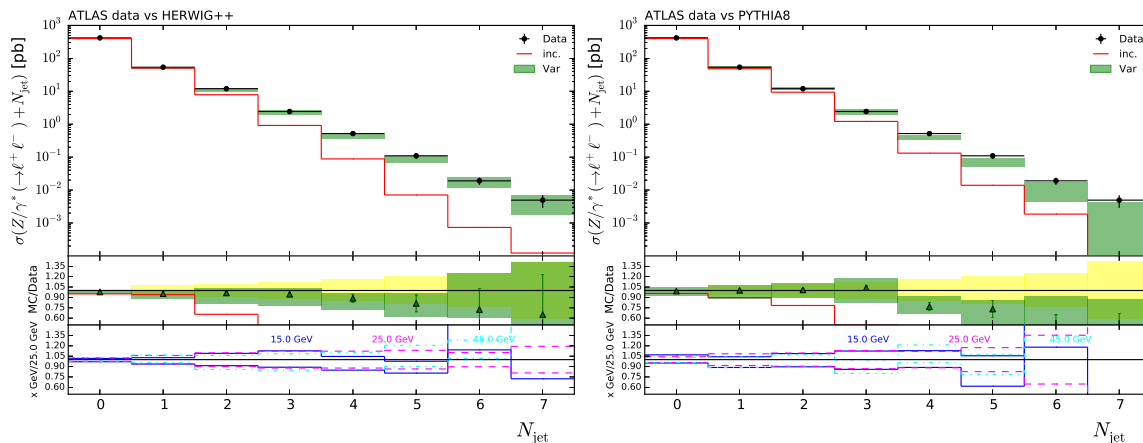
are MC-induced, hence of LL accuracy. This implies that for the present observables the comparison between data and fully-inclusive predictions is not particularly meaningful: hence, from a phenomenological viewpoint we believe it should not be considered. Note that when  $N_{\text{jet}} = 1$  the fully-inclusive MC@NLO results may actually be farther from data than those obtained from a  $V+1$ -parton LO sample, owing to the different choices of shower scales in the two simulations. Therefore, the *sole* reason for including the fully-inclusive MC@NLO predictions in the plots below is that of giving a benchmark against which the changes due to the inclusion of higher-multiplicity matrix elements in the FxFx procedure can be gauged in a purely theoretical way.

### 3.1 $Z$ +jets

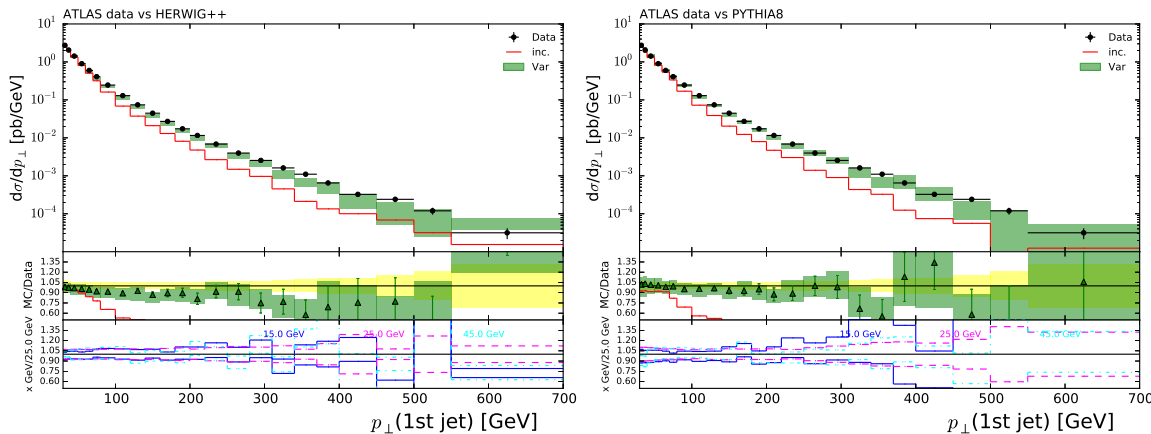
This section is devoted to the comparison between data and theoretical predictions relevant to  $Z$ +jets production. We shall make use of the ATLAS measurements of ref. [28], and of the CMS measurements of refs. [29, 30]. For each of these, a (non-exhaustive) summary of the characteristics of the analysis is given; however, we encourage the reader to check the original experimental papers for more details. In both the experimental results and in our simulations (see the beginning of section 2), a “ $Z$ ” is a shorthand for a lepton pair.

**ATLAS [28] (arXiv:1304.7098, Rivet analysis ATLAS\_2013\_I1230812).** Study of jet,  $Z$ , inclusive properties (the latter two defined by requiring the presence of at least one jet in the final state), and jet-jet correlations. Based on an integrated luminosity of  $4.6 \text{ fb}^{-1}$ , using both  $e^+e^-$  and  $\mu^+\mu^-$  pairs, with  $R = 0.4$  anti- $k_T$  jets [51] within  $p_T(j) > 30 \text{ GeV}$  and  $|y(j)| < 4.4$ . Further cuts ( $\ell$  denotes either an electron or a muon):  $p_T(\ell) \geq 20 \text{ GeV}$ ,  $66 \leq M(\ell\ell) \leq 116 \text{ GeV}$ ,  $\Delta R(j\ell) \geq 0.5$ ,  $\Delta R(\ell\ell) \geq 0.2$ ,  $|\eta(\mu)| \leq 2.4$ ,  $|\eta(e)| \leq 1.37$  and  $1.52 \leq |\eta(e)| \leq 2.47$ .

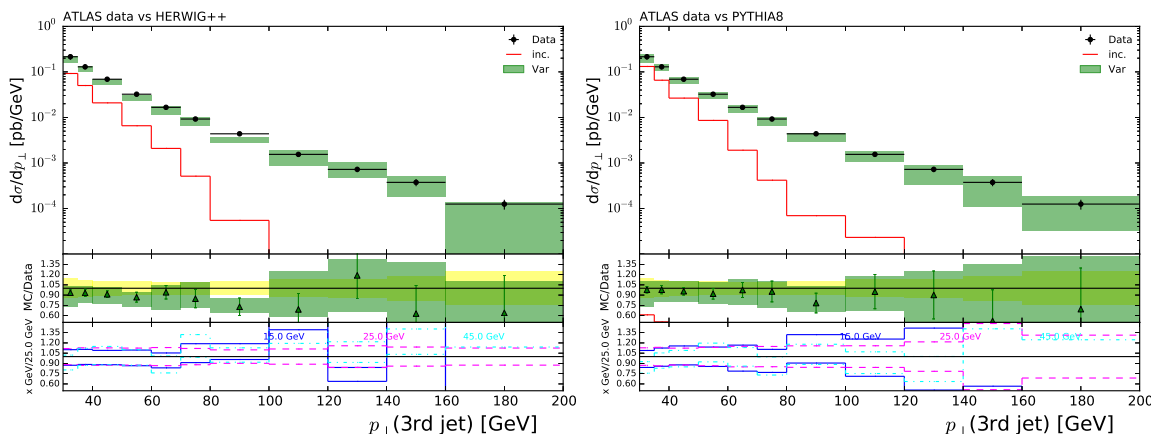
Each of the figures that follow is relevant to one observable and includes two panels, with identical data but different theoretical predictions (obtained by showering with HERWIG++ and with PYTHIA8). We present, in particular: figure 1: exclusive jet multiplicity; figure 2: transverse momentum of the 1<sup>st</sup> jet; figure 3: transverse momentum of the 3<sup>rd</sup> jet; figure 4: transverse momentum of the 4<sup>th</sup> jet; figure 5: rapidity of the 1<sup>st</sup> jet; figure 6: rapidity of the 3<sup>rd</sup> jet; figure 7: rapidity of the 4<sup>th</sup> jet; figure 8: rapidity distance between the two hardest jets; figure 9: invariant mass of the two hardest jets; figure 10: azimuthal distance between the two hardest jets; figure 11:  $\Delta R$  between the two hardest jets; figure 12:  $H_T$  of leptons and jets. Among the observables measured in ref. [28], those chosen here constitute a subset which is sufficiently representative, both in terms of kinematic characteristics and for the comparison with the two different MCs used in the present paper. The hardest, third-hardest, and fourth-hardest jets have been selected since formally they are predicted, given the matrix elements we have employed, at the NLO, LO, and LL accuracy respectively, and thus they cover all possible situations as far as the nominal predictivity of the simulations is concerned. The second-hardest jet, whose single-inclusive observables are not shown here, is expected to have a similar behaviour as the leading one, which is what we have indeed explicitly verified.



**Figure 1.** Exclusive jet multiplicity. Data from ref. [28], compared to HERWIG++ (left panel) and PYTHIA8 (right panel) predictions. The FxFx uncertainty envelope (“Var”) and the fully-inclusive central result (“inc”) are shown as green bands and red histograms respectively. See the end of section 2 for more details on the layout of the plots.



**Figure 2.** As in figure 1, for the transverse momentum of the 1<sup>st</sup> jet.



**Figure 3.** As in figure 1, for the transverse momentum of the 3<sup>rd</sup> jet.

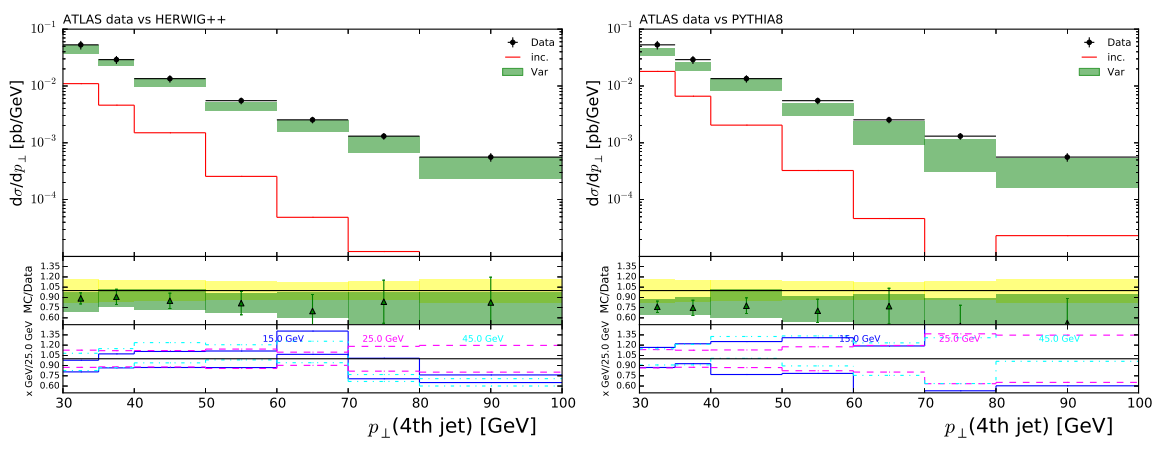


Figure 4. As in figure 1, for the transverse momentum of the 4<sup>th</sup> jet.

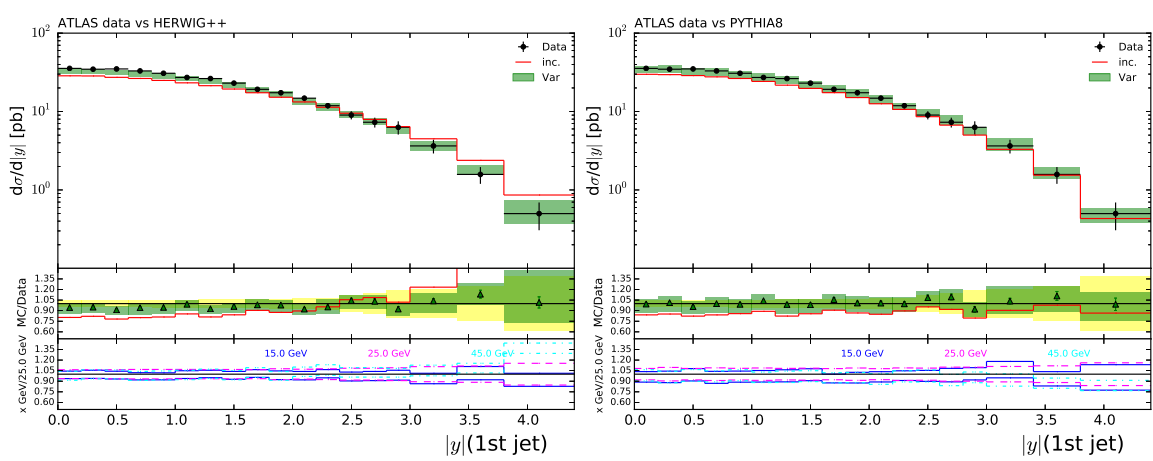


Figure 5. As in figure 1, for the rapidity of the 1<sup>st</sup> jet.

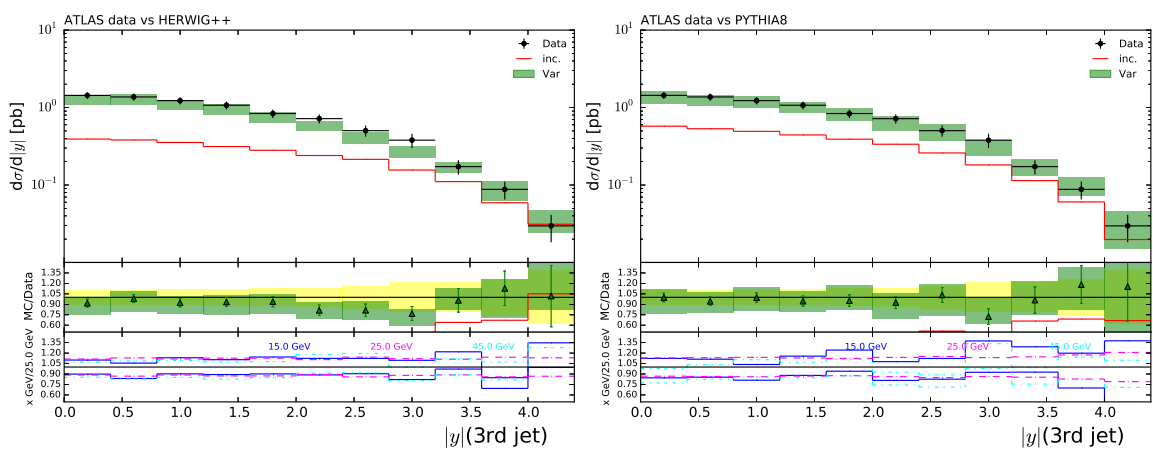


Figure 6. As in figure 1, for the rapidity of the 3<sup>rd</sup> jet.

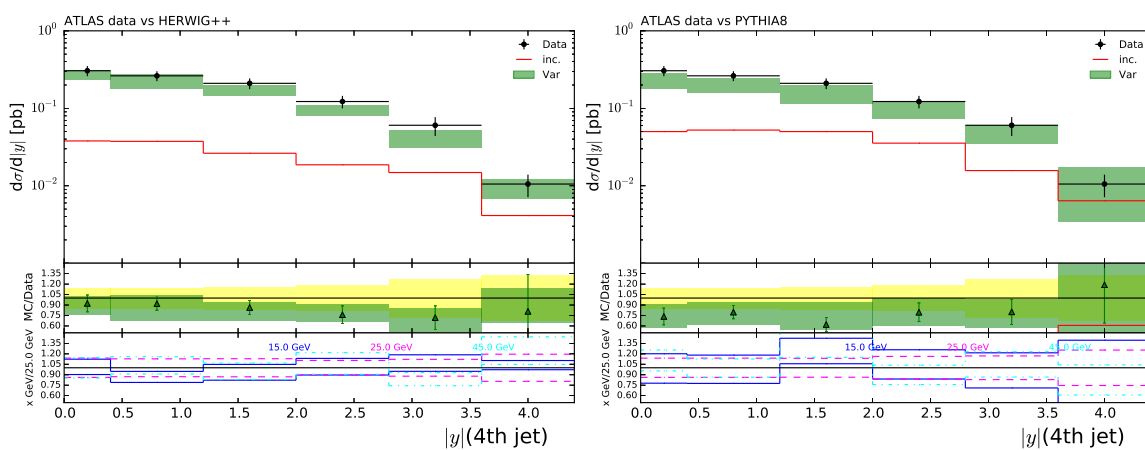


Figure 7. As in figure 1, for the rapidity of the 4<sup>th</sup> jet.

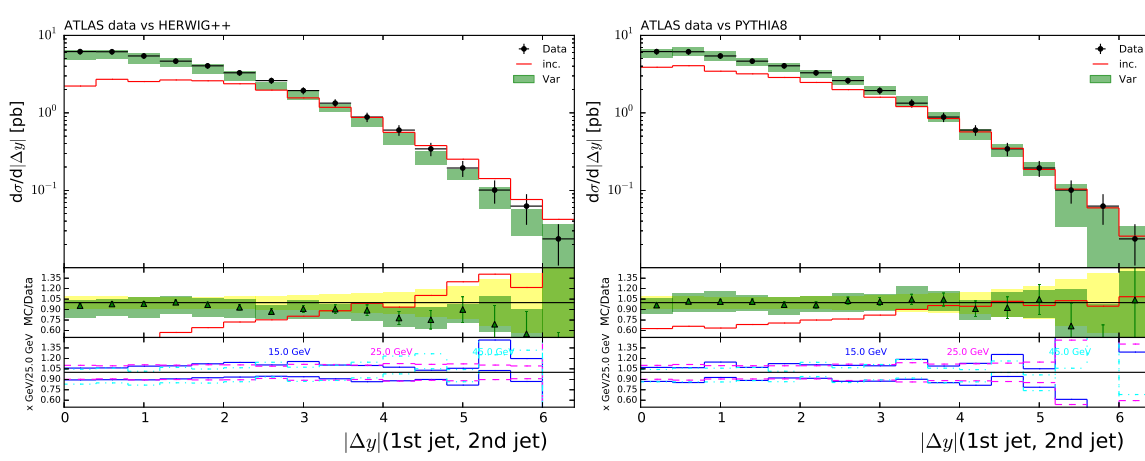


Figure 8. As in figure 1, for the rapidity distance between the two hardest jets.

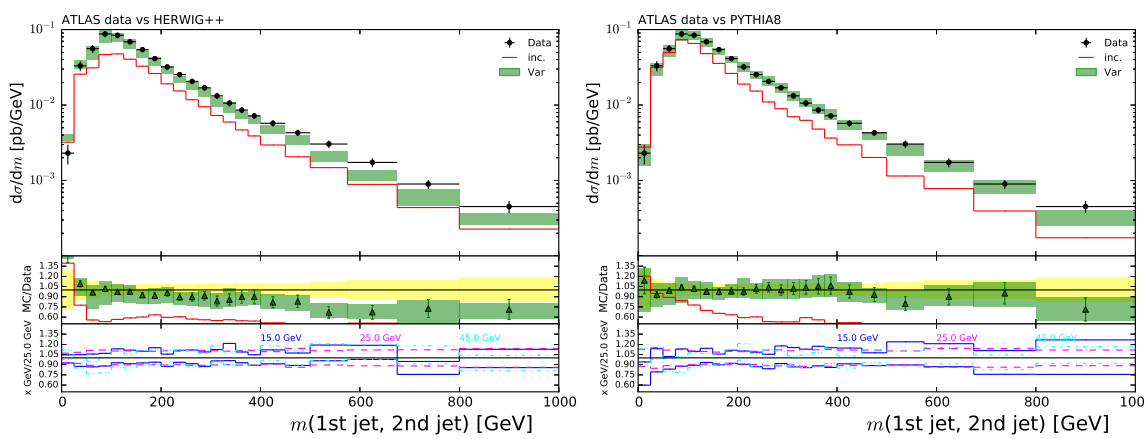


Figure 9. As in figure 1, for the invariant mass of the two hardest jets.

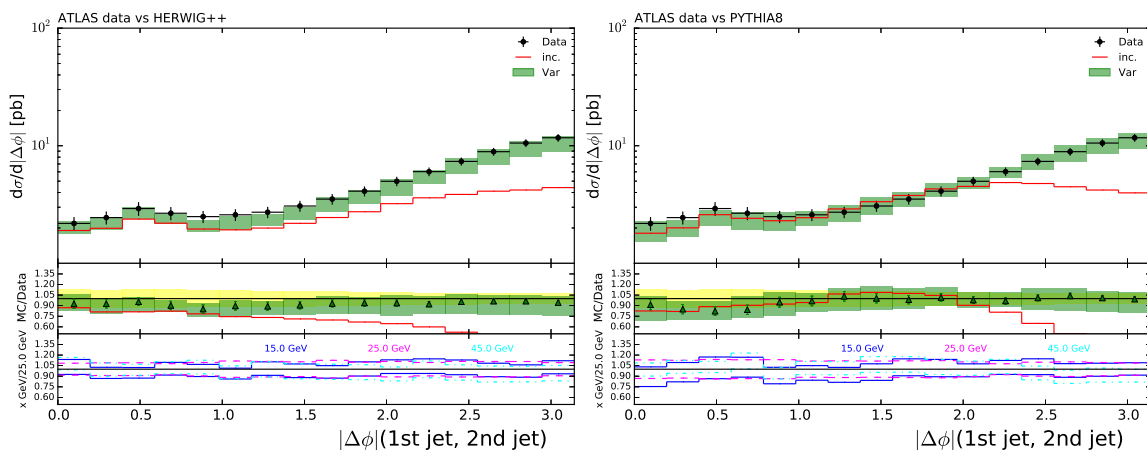


Figure 10. As in figure 1, for the azimuthal distance between the two hardest jets.

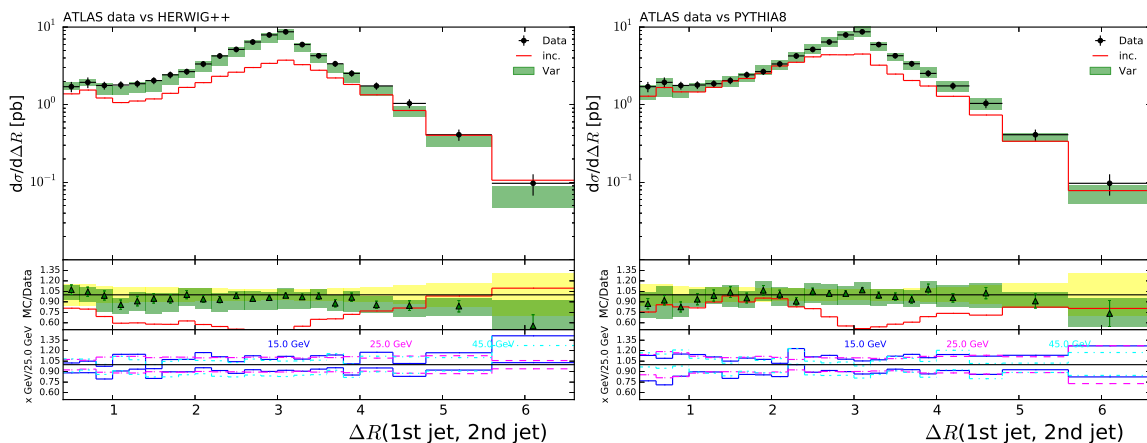


Figure 11. As in figure 1, for the  $\Delta R$  between the two hardest jets.

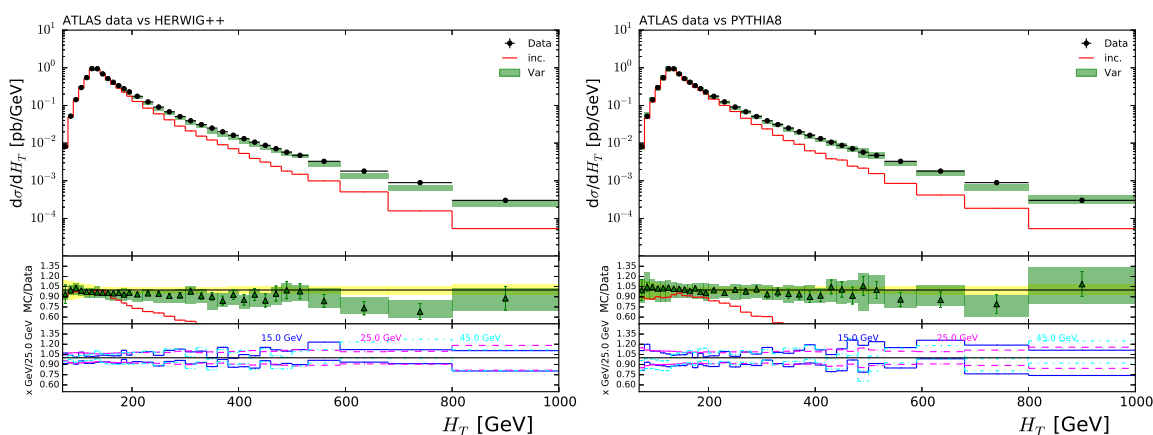


Figure 12. As in figure 1, for the  $H_T$  of leptons and jets.

The exclusive jet multiplicity (figure 1) is very well predicted by both MCs, up to  $N_{\text{jet}} = 3$ . Although in a statistically non-significant way, the central HERWIG++ prediction slightly undershoots the data, at variance with the PYTHIA8 one; this very minor difference between the two MCs is basically an overall effect, and can be accounted for by the total-rate results of table 2. The lack of high-multiplicity matrix elements starts to be visible for  $N_{\text{jet}} \geq 4$ , with PYTHIA8 dropping faster than HERWIG++ (whose central prediction is at the border of the data error band up to  $N_{\text{jet}} = 7$ ); it must be kept in mind that this multiplicity region is entirely dominated by MC effects, and formally of LL accuracy. The impact of multi-parton matrix elements, measured by the distance between the FxFx and the inclusive predictions, is dramatic.

The predictions for the single-jet transverse momenta of figures 2–4 tend to be marginally softer than data, although this trend is hardly statistically significant, except perhaps for the leading-jet distribution in the case of HERWIG++. It is worth remarking that, shape-wise, the agreement between theory and data is rather good even for the 4<sup>th</sup> jet, in spite of this being beyond matrix-element accuracy; for such a jet, the only difference between predictions and measurements is one of rates, whose values can be read off the  $N_{\text{jet}} = 4$  bin of figure 1. As expected, the differences between merged and inclusive predictions increase, in shape and rate, with the jet multiplicity. The situation for the single-jet rapidities (figures 5–7) is quite analogous to that of the  $p_T$ 's: they are described very well by both HERWIG++ and PYTHIA8 (bar the rate effect on the 4<sup>th</sup> jet). The almost identical FxFx predictions of the two MCs have no analogue at the inclusive level, where HERWIG++ and PYTHIA8 behave differently (with only the latter in reasonable agreement with data, and only for the leading jet); this is another clear indication of the benefits of the merging procedure. We finally point out that the theory-data comparison for the 2<sup>nd</sup> jet largely follows the same pattern as for the hardest jet, with reduced differences in hardness for the  $p_T$  distribution w.r.t. the latter case.

Figures 8–11 display four correlations, constructed with the two leading jets. Overall, the agreement between theory and data is good, and especially so in the case of PYTHIA8; the only exception is the invariant mass of the pair as predicted by HERWIG++, which is clearly softer than what is measured. Smaller discrepancies (i.e. within uncertainties) can be seen in the tail of the  $\Delta y$  distribution (slightly larger in the case of HERWIG++), at small  $\Delta\phi$  for PYTHIA8, and at large  $\Delta R$ ; note that the statistical errors start to be non-negligible in the tails of the  $\Delta y$  and  $\Delta R$  distributions. For all of these correlations, the differences between the merged and inclusive predictions are extremely significant.

Finally, the scalar transverse momentum sum of jets and leptons is shown in figure 12. Both MCs do well, with only a marginal tendency to be softer than data, more marked in the case of HERWIG++.

A general conclusion that can be drawn from the present theory-data comparison is that the benefits of merging extend beyond what one might naively expect on the basis of the multiplicities employed at the matrix element level. Even for observables that are entirely MC dominated, the presence of a few hard partons in the final state allows the MCs to stay within their natural range of validity, so that their underpinning approximations are not overstretched. While some of the  $p_T$ -related theoretical predictions are slightly softer



than data, the trend is generally not statistically significant; thus, there is no evidence for the necessity of including  $Z+3j$  matrix elements in the case of the observables studied here (bar the 4<sup>th</sup>-jet rate), since small differences could also be induced by e.g. different choices of input parameters in the matrix elements and/or the MCs.

**CMS [29] (arXiv:1310.3082, Rivet analysis CMS\_2013\_I1258128).** Study of rapidity distributions in  $Z+1$  jet events (i.e. exactly one jet). Based on an integrated luminosity of  $5\text{ fb}^{-1}$ , using both  $e^+e^-$  and  $\mu^+\mu^-$  pairs, with  $R = 0.5$  anti- $k_T$  jets, within  $p_T(j) > 30\text{ GeV}$  and  $|\eta(j)| < 2.4$ . Further cuts:  $p_T(\ell) \geq 20\text{ GeV}$ ,  $76 \leq M(\ell\ell) \leq 106\text{ GeV}$ ,  $|\eta(\ell)| \leq 2.1$ ,  $p_T(\ell\ell) \geq 40\text{ GeV}$ ,  $\Delta R(j\ell) \geq 0.5$ .

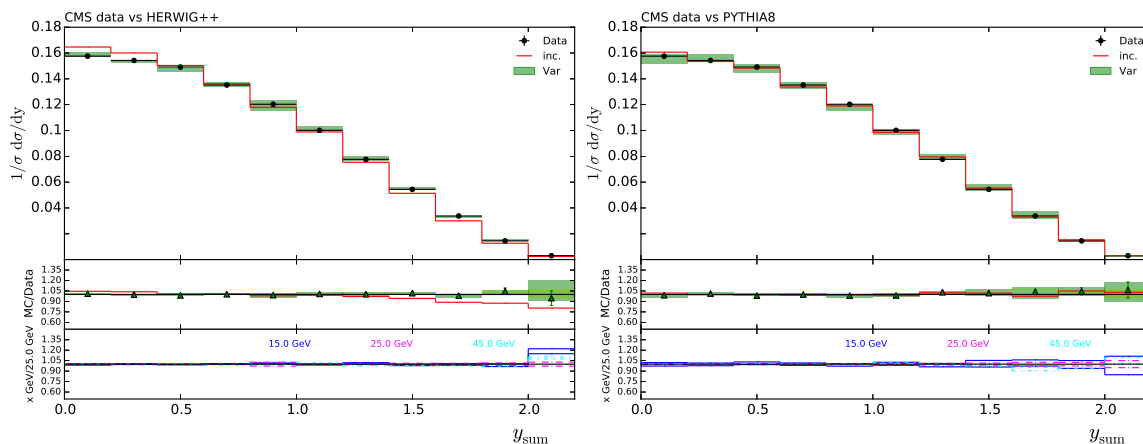
We present here two observables: in figure 13 and figure 14 the sum and the difference, respectively, of the rapidities of the  $Z$  and of the jet; these we have chosen for being the most involved cases among the measurements in ref. [29], and because their comparison with the theoretical LO+PS predictions considered by CMS was not entirely satisfactory.

As one can see from the figures, the agreement between merged predictions and data is excellent for both MCs. This result appears to be strongly driven by matrix-element effects, given the very significant differences between the HERWIG++ and PYTHIA8 predictions which result from the inclusive samples. This is especially true in the case of the rapidity difference, which in inclusive simulations is known to be affected by large MC systematics — for a detailed discussion on this point, see refs. [21, 52]. We point out that we have found a level of agreement identical to that of figures 13 and 14 also in the case of the single-inclusive rapidities (of the  $Z$  and the jet) measured in ref. [29].

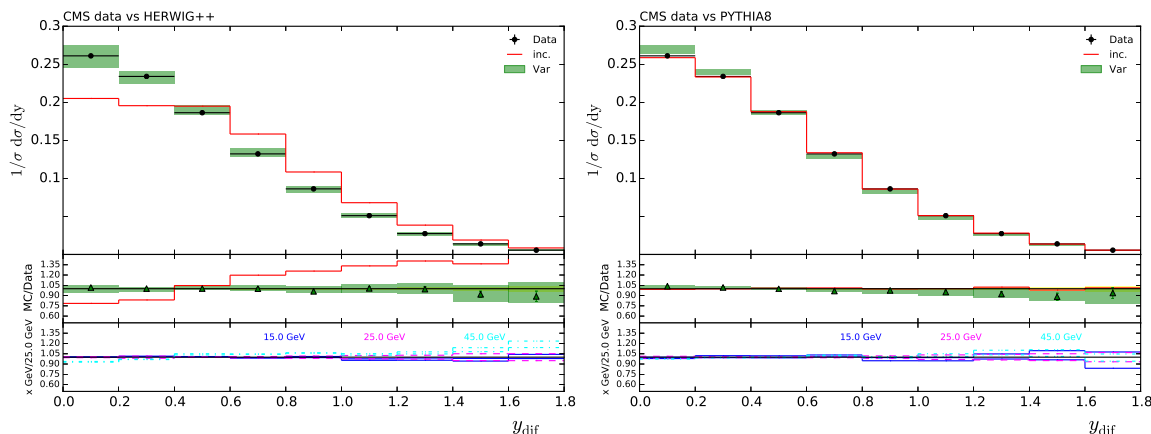
**CMS [30] (arXiv:1301.1646, Rivet analysis CMS\_2013\_I1209721).** Study of event shapes and azimuthal correlations. Based on an integrated luminosity of  $5\text{ fb}^{-1}$ , using both  $e^+e^-$  and  $\mu^+\mu^-$  pairs, with  $R = 0.5$  anti- $k_T$  jets within  $p_T(j) > 50\text{ GeV}$  and  $|\eta(j)| < 2.5$ . Further cuts:  $p_T(\ell) \geq 20\text{ GeV}$ ,  $71 \leq M(\ell\ell) \leq 111\text{ GeV}$ ,  $|\eta(\ell)| \leq 2.4$ ,  $\Delta R(j\ell) \geq 0.4$ . We point out that, owing to the hardness of the jets defined by CMS in this analysis (which is significantly larger than that relevant to refs. [28, 29]), and given the generation cuts we have chosen (see the beginning of section 3), our simulations have statistical accuracies relatively lower than those accumulated for the other measurements considered in this paper. For this reason, for certain observables we have limited ourselves to presenting only HERWIG++ results, after having checked the statistical compatibility of these with their PYTHIA8 counterparts, which have slightly larger fluctuations.

We have selected the following observables. Figure 15: azimuthal distance between the  $Z$  and the 1<sup>st</sup> jet; figure 16: transverse thrust; figure 17: azimuthal distance between the  $Z$  and the 1<sup>st</sup> jet, in events with at least two jets, and azimuthal distance between the 1<sup>st</sup> and the 2<sup>nd</sup> jet, in events with at least three jets; figure 18: azimuthal distance between the  $Z$  and the 1<sup>st</sup> jet, and transverse thrust, both in events with  $p_T(Z) \geq 150\text{ GeV}$ .

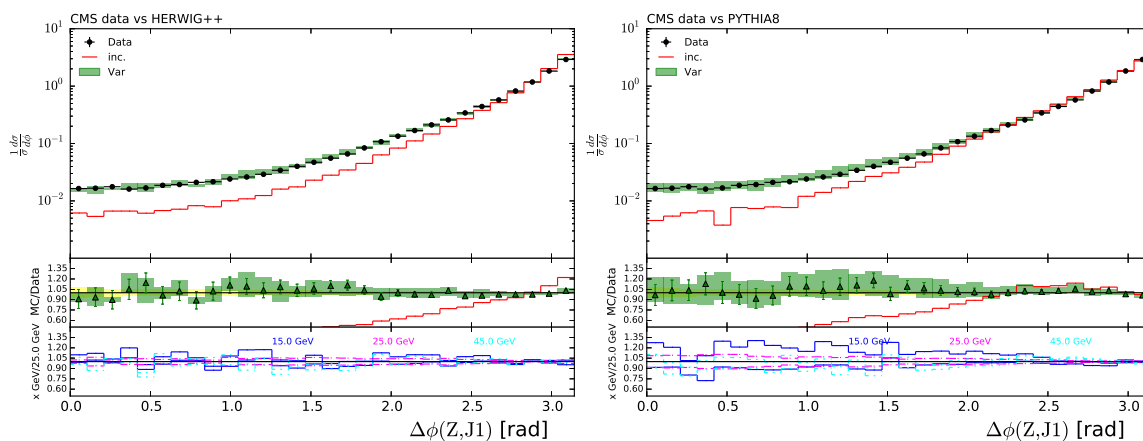
Two  $N_{\text{jet}} \geq 1$  observables are presented in figure 15 (distance between the  $Z$  and the leading jet in the azimuthal angle) and in figure 16 (transverse thrust), and compared to both HERWIG++ and PYTHIA8. When considered in their full ranges, these observables are very sensitive to the interplay between matrix elements, parton showers, and possibly



**Figure 13.** Sum of the rapidities of the  $Z$  and the 1<sup>st</sup> jet. Data from ref. [29], compared to HERWIG++ (left panel) and PYTHIA8 (right panel) predictions. The FxFx uncertainty envelope (“Var”) and the fully-inclusive central result (“inc”) are shown as green bands and red histograms respectively. See the end of section 2 for more details on the layout of the plots.



**Figure 14.** As in figure 13, for the difference of the rapidities of the  $Z$  and the 1<sup>st</sup> jet.



**Figure 15.** Azimuthal distance between the  $Z$  and the 1<sup>st</sup> jet. Data from ref. [30], compared to HERWIG++ (left panel) and PYTHIA8 (right panel) predictions. The FxFx uncertainty envelope (“Var”) and the fully-inclusive central result (“inc”) are shown as green bands and red histograms respectively. See the end of section 2 for more details on the layout of the plots.

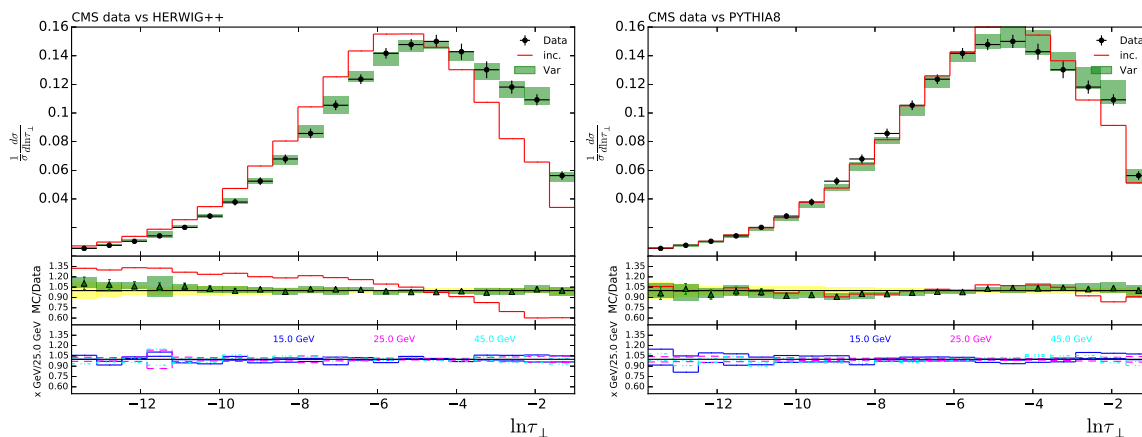


Figure 16. As in figure 15, for the transverse thrust.

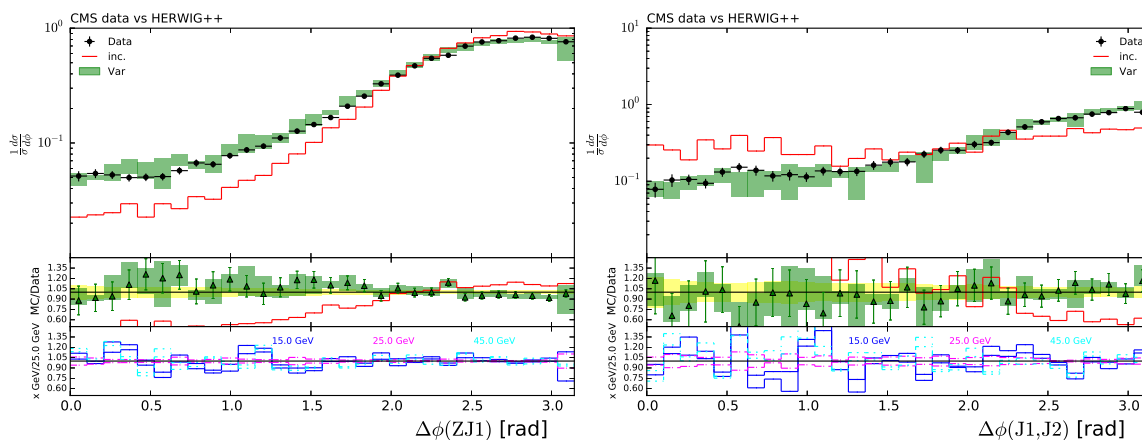


Figure 17. As in figure 15, for the azimuthal distance between the  $Z$  and the 1<sup>st</sup> jet, in events with at least two jets (left panel), and the azimuthal distance between the 1<sup>st</sup> and the 2<sup>nd</sup> jet, in events with at least three jets (right panel), both compared to HERWIG++ predictions.

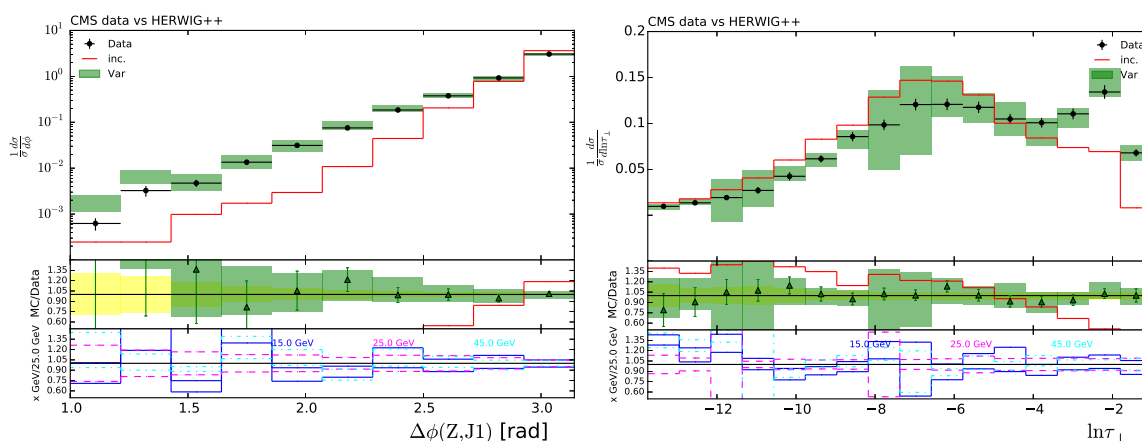


Figure 18. As in figure 17, for the azimuthal distance between the  $Z$  and the 1<sup>st</sup> jet (left panel), and the transverse thrust (right panel), both in events with  $p_T(Z) \geq 150$  GeV.

soft physics modelling,<sup>10</sup> and thus ultimately to the merging procedure. When increasing either the minimal jet multiplicity, or the  $Z$  transverse momentum, final-state objects are in fact less correlated (see ref. [30]). The agreement between the measurements and the FxFx theoretical predictions is excellent for both MCs, indicating that the cross-talk between merging and soft-physics modelling does not deteriorate the overall quality of event generation. On average, the merging systematics is larger with PYTHIA8 than with HERWIG++. Inspection of the lower insets shows that this is basically due to the behaviour of the  $\mu_Q = 15$  GeV PYTHIA8 sample, which has the largest fluctuations; however, from the upper insets one sees that the central predictions are quite consistent between the two MCs, and almost on top of the data in the whole ranges.

Two azimuthal distances ( $\Delta\phi(Z, j_1)$  with  $N_{\text{jet}} \geq 2$ , and  $\Delta\phi(j_1, j_2)$  with  $N_{\text{jet}} \geq 3$ ) are shown in figure 17, together with the corresponding HERWIG++ results. As expected, these are flatter than those of figure 15, i.e. more de-correlated. However, a certain amount of matrix-element-induced effect is still present, as one can infer by comparing the FxFx with the inclusive results, with the latter predicting less events in the  $\Delta\phi(Z, j_1) \rightarrow 0$  tail and a flatter  $\Delta\phi(j_1, j_2)$  distribution. The agreement of the theoretical predictions with data is again quite good, albeit in presence of larger systematics.

Finally, the same observables as in figures 15 and 16, but subject to a  $p_T(Z) \geq 150$  GeV cut, are compared to HERWIG++ predictions in figure 18. The only discrepancy with data is seen in the two leftmost bins of the azimuthal correlation, which are however affected by extremely large statistical errors (note, furthermore, that the bin with the smallest  $\Delta\phi(Z, j_1)$  value is not shown in figure 4 of ref. [30]). Observe that, in the case of the transverse thrust, the central FxFx result (green triangles) is in good agreement with data, and affected by a relatively small scale-and-PDF uncertainty (magenta dashed band); the large overall systematics is again dominated by the behaviour of the  $\mu_Q = 15$  GeV sample. We point out that in the context of the present analysis, which is characterised by very hard jets, such a choice for the merging scale is extreme, and might in fact be safely neglected.

As was the case for the two  $Z$ +jets analyses discussed previously, the differences between the FxFx and the fully-inclusive results are always very significant. The latter also exhibit a strong MC dependence, which is essentially absent in the former.

### 3.2 $W$ +jets

This section is devoted to the comparison between data and theoretical predictions relevant to  $W$ +jets production. We shall make use of the ATLAS measurements of ref. [31], and of the CMS measurements of ref. [32]. For each of these, a (non-exhaustive) summary of the characteristics of the analysis is given; however, we encourage the reader to check the original experimental papers for more details.

**ATLAS [31] (arXiv:1409.8639, Rivet analysis ATLAS\_2014\_I1319490).** Study of jet and inclusive properties (the latter defined by requiring the presence of at least one jet in the final state), and of jet-jet correlations. Based on an integrated luminosity of

<sup>10</sup>Multi-parton scatterings are conjectured to have a non-negligible impact on observables such as the transverse thrust [53].

4.6 fb<sup>-1</sup>, using both the electron and muon channels, with  $R = 0.4$  anti- $k_T$  jets within  $p_T(j) > 30$  GeV and  $|y(j)| < 4.4$ . Further cuts:  $p_T(\ell) \geq 25$  GeV,  $|\eta(\mu)| \leq 2.4$ ,  $|\eta(e)| \leq 1.37$  and  $1.52 \leq |\eta(e)| \leq 2.47$ ,  $E_T > 25$  GeV,  $m_T(\ell\nu) > 40$  GeV (see ref. [31] for the definition of the missing energy and the neutrino transverse momentum). Events are rejected if a second electron or muon that passes the cuts is present. Note that ATLAS treat  $W \rightarrow \tau\nu$  events as background, and consistently with that we have not generated this channel.

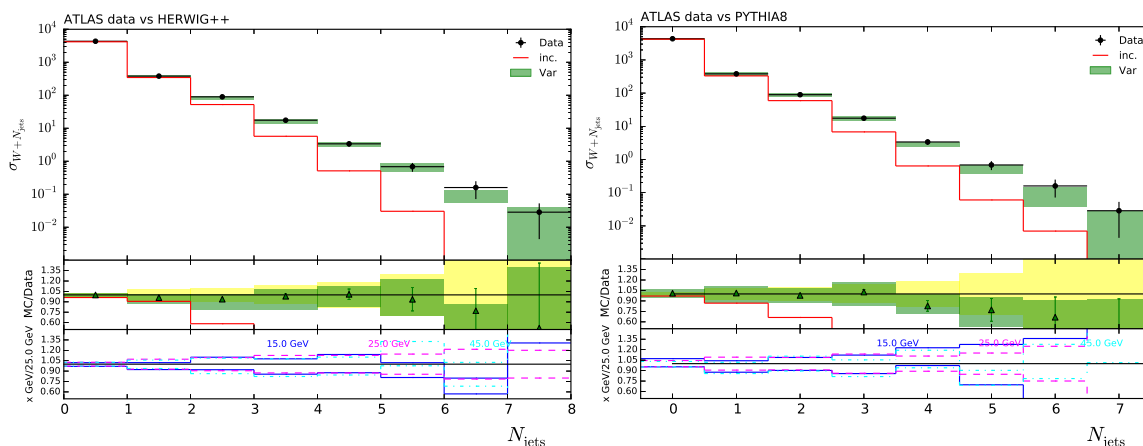
The observables that we have selected in this analysis are displayed in the following plots. Figure 19: exclusive jet multiplicity; figure 20: transverse momentum of the 1<sup>st</sup> jet; figure 21: transverse momentum of the 1<sup>st</sup> jet, in events with at least two jets; figure 22: transverse momentum of the 1<sup>st</sup> jet, in events with at least three jets; figure 23: rapidity of the 1<sup>st</sup> jet; figure 24: rapidity of the 2<sup>nd</sup> jet; figure 25: rapidity of the 3<sup>rd</sup> jet; figure 26: azimuthal distance between the two hardest jets; figure 27: rapidity distance between the two hardest jets; figure 28:  $\Delta R$  between the two hardest jets; figure 29: invariant mass of the two hardest jets; figure 30:  $H_T$ ; figure 31:  $H_T$  in events with at least three jets.

The comparison between theory and data for the exclusive jet multiplicity (figure 19) largely follows the same pattern as its analogue in  $Z$ +jets production, discussed above for the measurement of ref. [28] (see figure 1). If anything, in the present case the excellent agreement of the merged predictions with the data extends to larger  $N_{\text{jet}}$  values; however, we do not consider this fact to be particularly significant, since here one is in a regime completely dominated by MC effects, and thus subject to significant uncertainties. We point out that we obtain an identical level of agreement in the case of the inclusive jet multiplicity. This implies that, for our predictions, the analogues of the scale factors reported in table 7 of ref. [31] would all be quite consistent with each other.

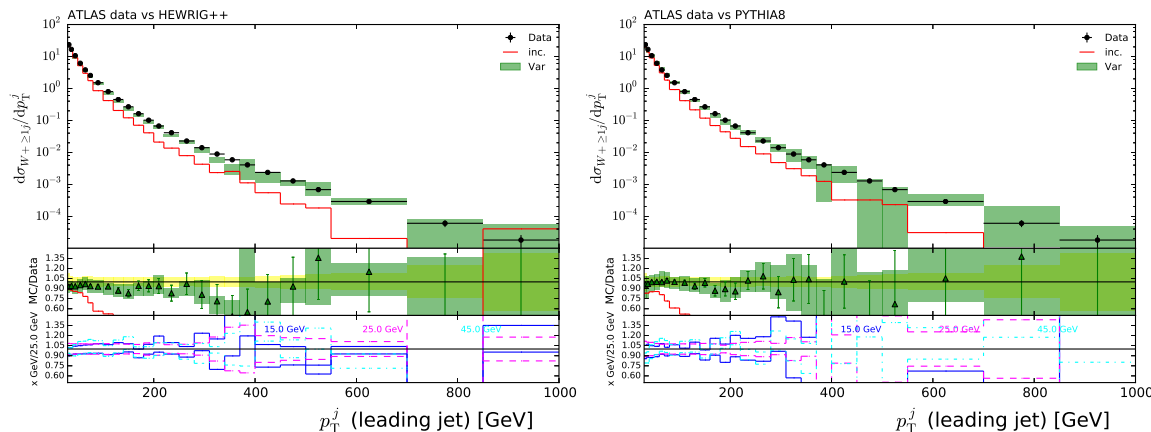
As far as the single-jet transverse momenta are concerned, we have considered that of the leading jet, in events characterised by different numbers of jets ( $N_{\text{jet}} \geq 1, 2, 3$  in figures 20, 21, and 22, respectively). The agreement between merged results and data is generally quite good. There is no indication of the predictions being softer than the measurements (as was marginally the case for the  $Z$ +jets analysis of ref. [28]); the clearest evidence of that, the  $N_{\text{jet}} \geq 1$  case as predicted by HERWIG++, is much weaker than its analogue in the  $Z$ +jets case (see figure 2). On the other hand, there is possibly an indication of the theory being lower than data at the smallest  $p_T$ 's, especially for  $N_{\text{jet}} \geq 2, 3$ , but this is not statistically very significant; we note that a similar trend has been observed in ref. [31] for several of the theoretical simulations considered there.

The rapidities of the three leading jets, displayed in figures 23–25, are in good agreement with the predictions of both HERWIG++ and PYTHIA8, in terms of both shapes and rates (the latter is consistent with what we have observed for the  $N_{\text{jet}}$  distribution). As in the case of  $Z$ +jets production, the similarity between the HERWIG++ and PYTHIA8 FxFx results has to be contrasted with their inclusive counterparts, whose shapes are significantly different from each other. Furthermore, such a similarity implies an almost complete independence upon the shower modelling of the merged predictions; this is in contrast to what is remarked in ref. [31] regarding the leading and subleading jet rapidities.

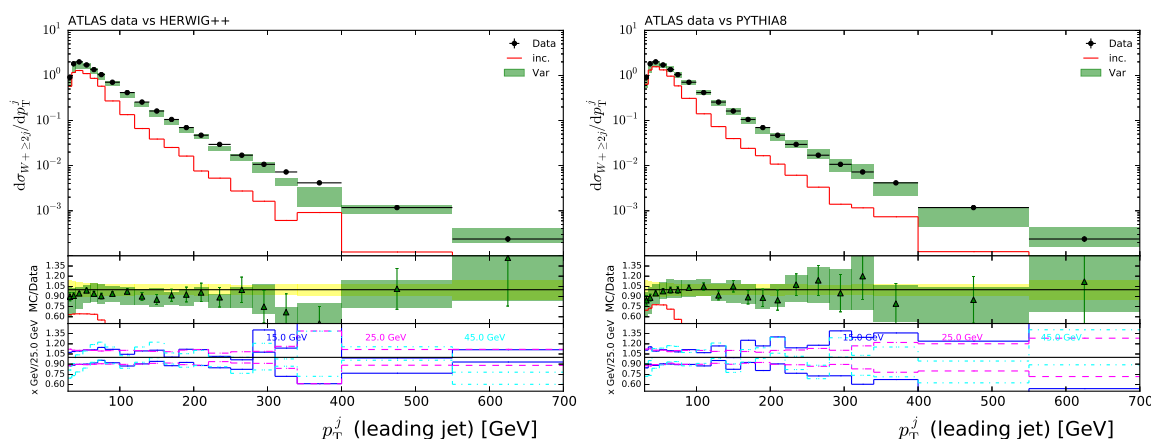
Correlations constructed with the two leading jets are reported in figures 26–29. The agreement between theory and data is generally quite good; we do not find any of the seem-



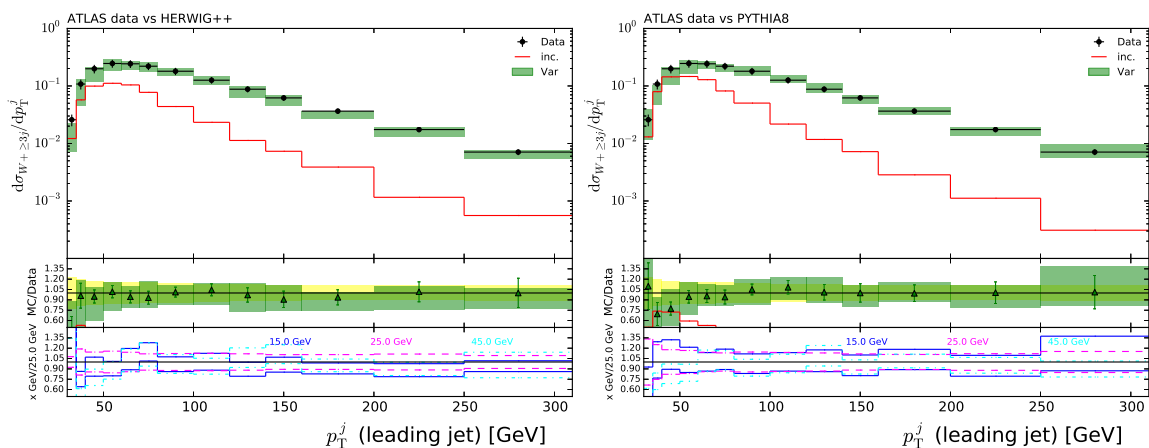
**Figure 19.** Exclusive jet multiplicity. Data from ref. [31], compared to HERWIG++ (left panel) and PYTHIA8 (right panel) predictions. The FxFx uncertainty envelope (“Var”) and the fully-inclusive central result (“inc”) are shown as green bands and red histograms respectively. See the end of section 2 for more details on the layout of the plots.



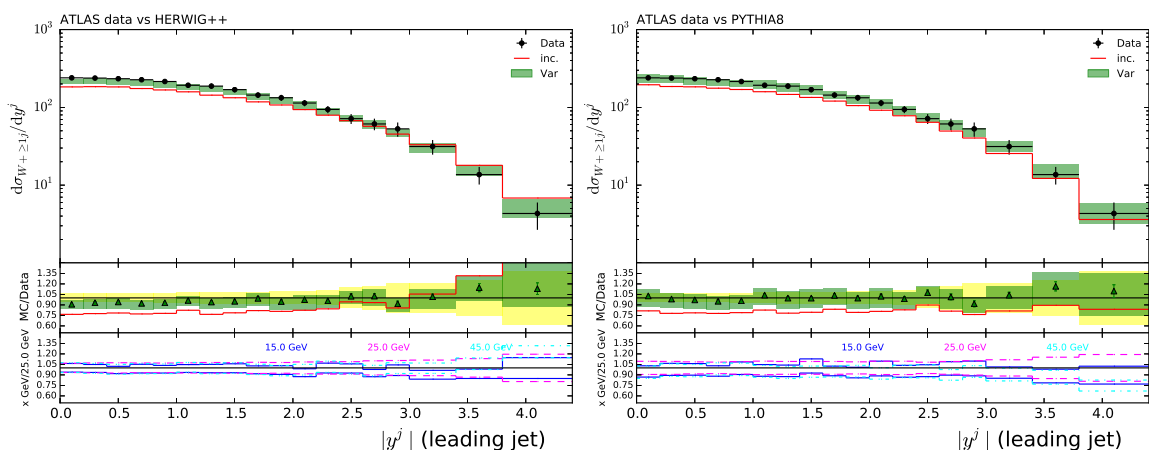
**Figure 20.** As in figure 19, for the transverse momentum of the 1<sup>st</sup> jet.



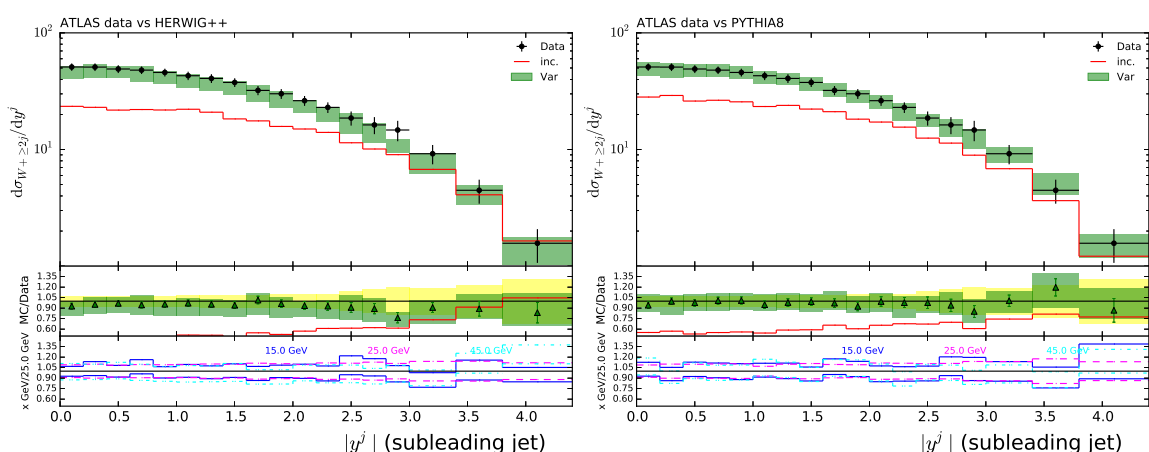
**Figure 21.** As in figure 19, for the transverse momentum of the 1<sup>st</sup> jet, in events with at least two jets.



**Figure 22.** As in figure 19, for the transverse momentum of the 1<sup>st</sup> jet, in events with at least three jets.



**Figure 23.** As in figure 19, for the rapidity of the 1<sup>st</sup> jet.



**Figure 24.** As in figure 19, for the rapidity of the 2<sup>nd</sup> jet.

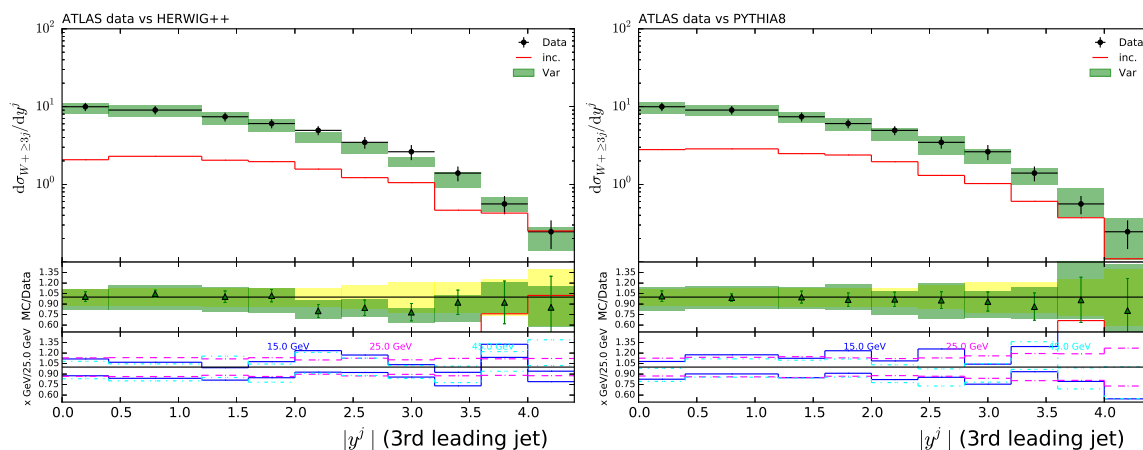


Figure 25. As in figure 19, for the rapidity of the 3<sup>rd</sup> jet.

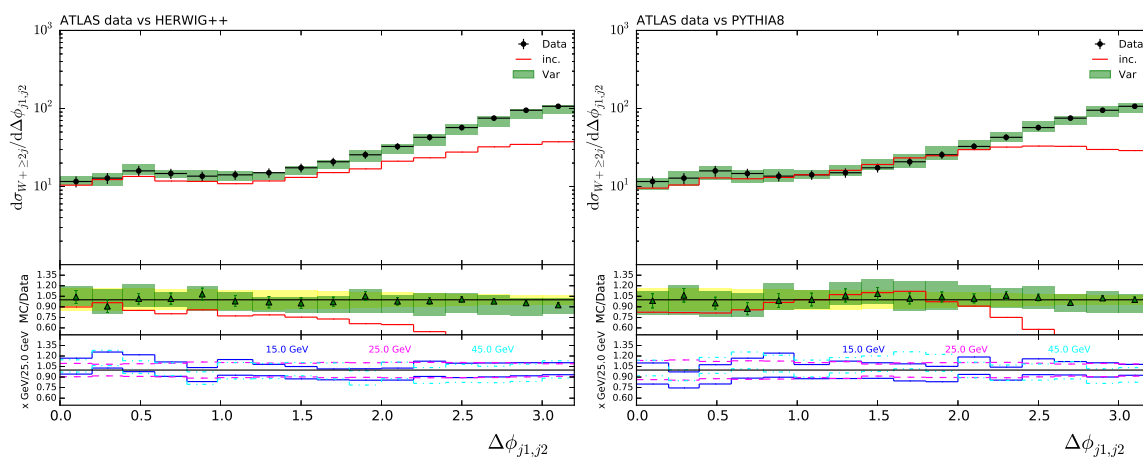


Figure 26. As in figure 19, for the azimuthal distance between the two hardest jets.

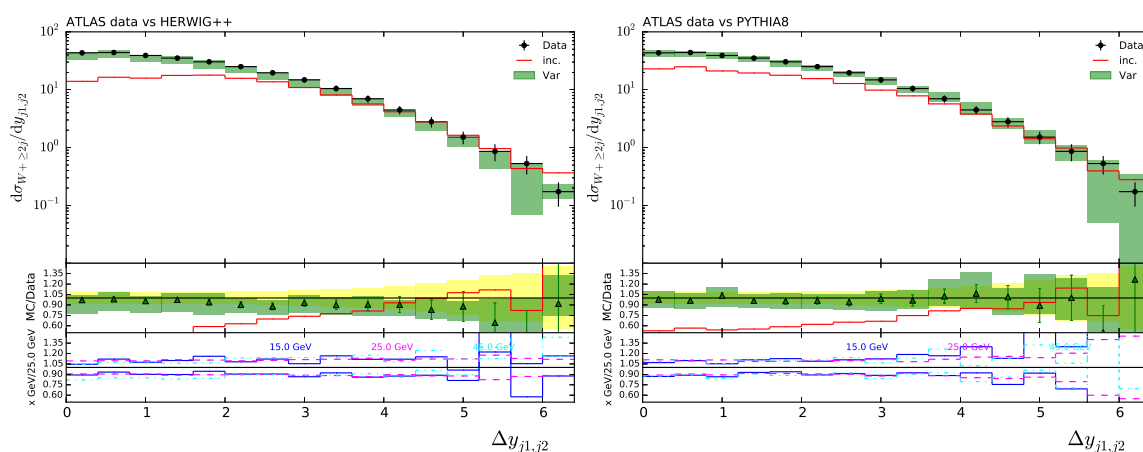


Figure 27. As in figure 19, for the rapidity distance between the two hardest jets.



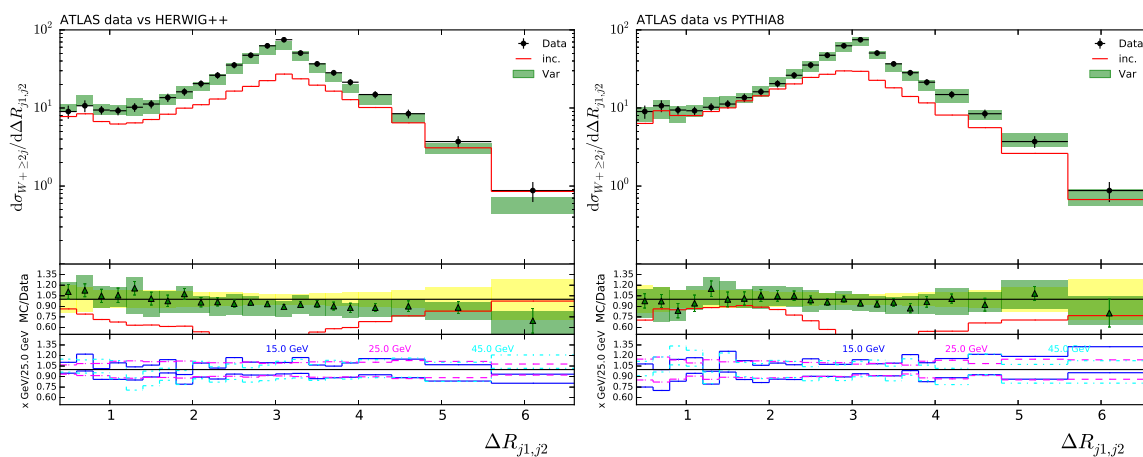


Figure 28. As in figure 19, for the  $\Delta R$  between the two hardest jets.

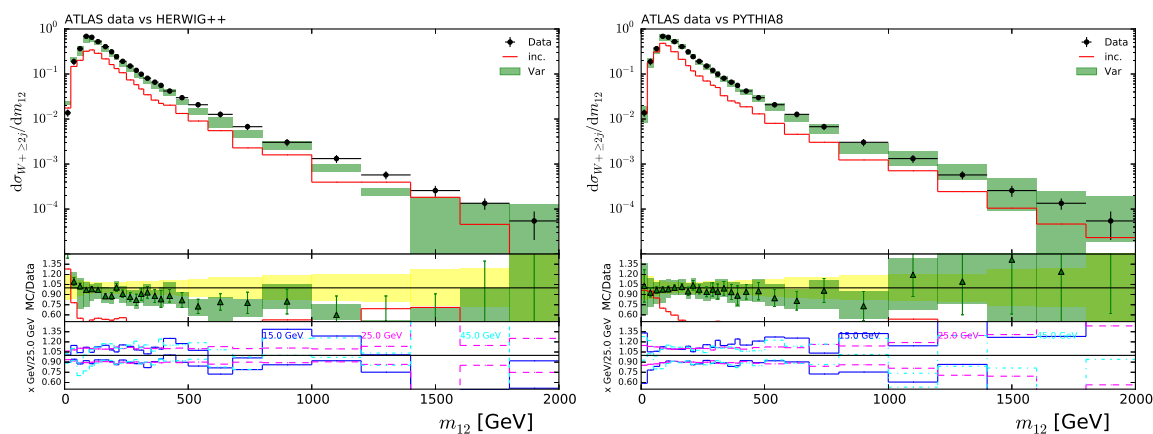


Figure 29. As in figure 19, for the invariant mass of the two hardest jets.

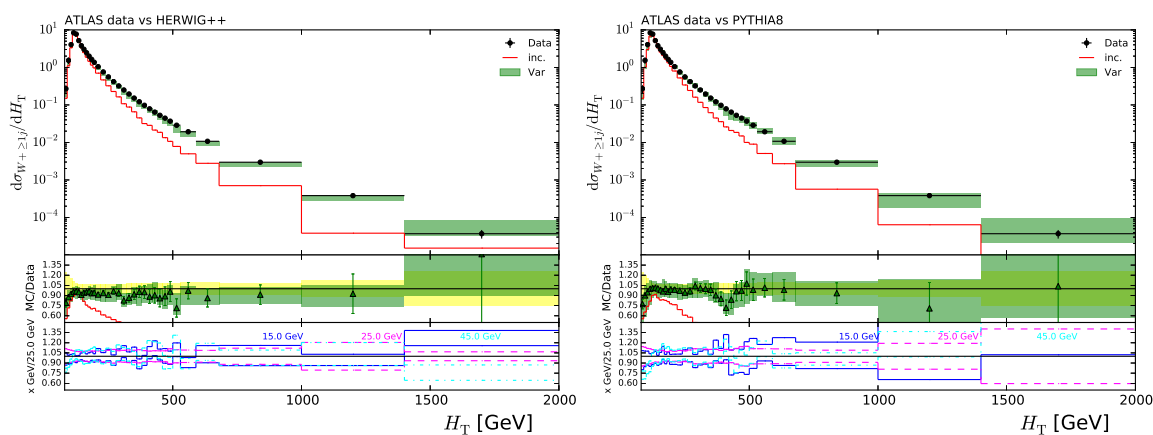
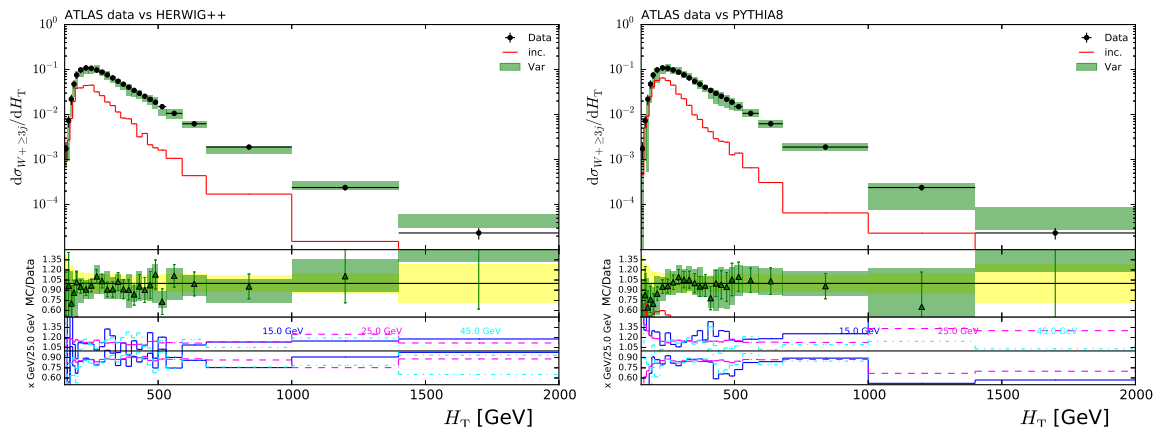


Figure 30. As in figure 19, for  $H_T$ .



**Figure 31.** As in figure 19, for  $H_T$  in events with at least three jets.

ingly large shape discrepancies<sup>11</sup> between the measurements and most of the theoretical predictions observed in ref. [31], particularly in the cases of  $\Delta y_{j_1 j_2}$  and  $\Delta R_{j_1 j_2}$ . Among our results, the only one whose shape is statistically only marginally compatible with that of the data is  $m_{j_1 j_2}$  as predicted by HERWIG++; we point out that a similar behaviour has been also found in the case of  $Z$ +jets production (see figure 9). Much smaller differences between the two MCs are present in the cases of rapidity and  $R$  distances, with PYTHIA8 slightly harder than HERWIG++.

We conclude the discussion of the results of ref. [31] with two  $H_T$  distributions, measured by imposing  $N_{\text{jet}} \geq 1$  (figure 30) and  $N_{\text{jet}} \geq 3$  (figure 31). The theory-data agreement is again good, within somewhat large systematics; our central predictions, taken at face value, have shapes and rates pretty compatible with those of the measurements. A possible exception is the very small  $H_T$  region, with theory decreasing faster than data — this is similar to the trend observed in ref. [31].

**CMS [32] (arXiv:1406.7533, Rivet analysis CMS\_2014\_I1303894).** Study of jet and inclusive properties (the latter defined by requiring the presence of at least one jet in the final state), and of correlations. Based on an integrated luminosity of  $5 \text{ fb}^{-1}$ , using only the muon channel, with  $R = 0.5$  anti- $k_T$  jets within  $p_T(j) > 30 \text{ GeV}$ ,  $|\eta(j)| < 2.4$ . Further cuts:  $p_T(\mu) > 24 \text{ GeV}$ ,  $|\eta(\mu)| < 2.1$ ,  $\Delta R(j\mu) \geq 0.5$ ,  $m_T(\mu\nu) > 50 \text{ GeV}$  (see ref. [32] for the definition of the missing energy and the neutrino transverse momentum); events must contain exactly one muon. We remark that a technical problem has occurred while running this Rivet analysis with PYTHIA8 for some merging scale, which we have failed to isolate and which has thus prevented us from reconstructing some of the observables in the simulation of such an MC. Since we believe that PYTHIA8 is already quite well tested in the comparison to the  $W$ +jets data of ref. [31] discussed previously, as well as for  $Z$ +jets production, for the observables in question we have limited ourselves to presenting the HERWIG++ results.

<sup>11</sup>We remark that, in the first version of this paper, data and theory were incompatible in the rightmost bin of the  $\Delta\phi_{j_1 j_2}$  distribution. This was due to an incorrect normalisation of that bin in the Rivet analysis, which has now been fixed. It is thus important that the revised version of ATLAS\_2014\_I1319490 be used in order to reproduce the results shown in figure 26.

We have chosen the observables that we consider in the following plots. Figure 32: exclusive jet multiplicity; figure 33: transverse momentum of the 1<sup>st</sup> jet; figure 34: pseudorapidity of the 1<sup>st</sup> jet; figure 35: azimuthal distance between the  $\mu$  and the 1<sup>st</sup> jet; figure 36:  $H_T$ ; figure 37: transverse momentum of the 2<sup>nd</sup> and of the 3<sup>rd</sup> jet; figure 38: pseudorapidity of the 2<sup>nd</sup> and of the 3<sup>rd</sup> jet; figure 39: azimuthal distance between the  $\mu$  and the 2<sup>nd</sup> jet, and between the  $\mu$  and the 3<sup>rd</sup> jet; figure 40:  $H_T$  in events with at least two and at least three jets.

The exclusive jet multiplicity (figure 32) is very well predicted by both MCs — one could repeat almost verbatim the same comments as for the analysis of ref. [31] (see figure 19).

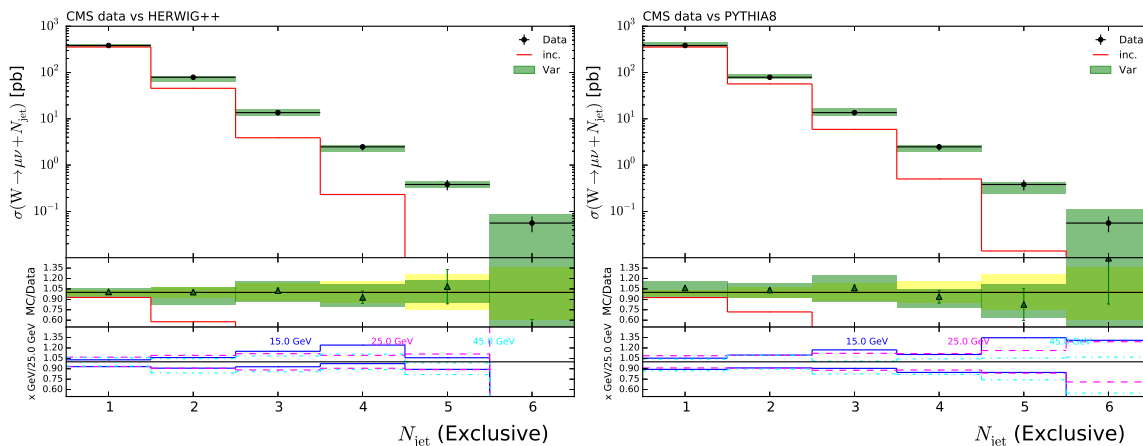
The inclusive leading-jet observables are reported in figure 33 ( $p_T$ ) and figure 34 (pseudorapidity). As far as the transverse momentum is concerned, both MCs tend to be slightly harder than data, an effect which is more visible in the case of PYTHIA8. This trend, which is statistically not very significant (especially in the case of HERWIG++), is similar to that observed in ref. [32]. If one had to regard our predictions as an NLO-upgraded version of those labelled “MadGraph” in ref. [32], one would clearly see a significant improvement w.r.t. the latter. However, we caution against taking this comparison too literally, if anything because the LO simulations reported in ref. [32] have been obtained with PYTHIA6. For what concerns the leading-jet pseudorapidity, both MCs give an excellent description of the data (and are thus basically identical to each other).

The azimuthal distance between the leading jet and the muon is shown in figure 35, with a fairly good theory-data agreement. For both MCs, predictions tend to be marginally steeper than data, although in a way which is not statistically significant. A similar trend is observed in ref. [32], and is slightly larger there than here (sometimes much larger, depending on the specific theory prediction).

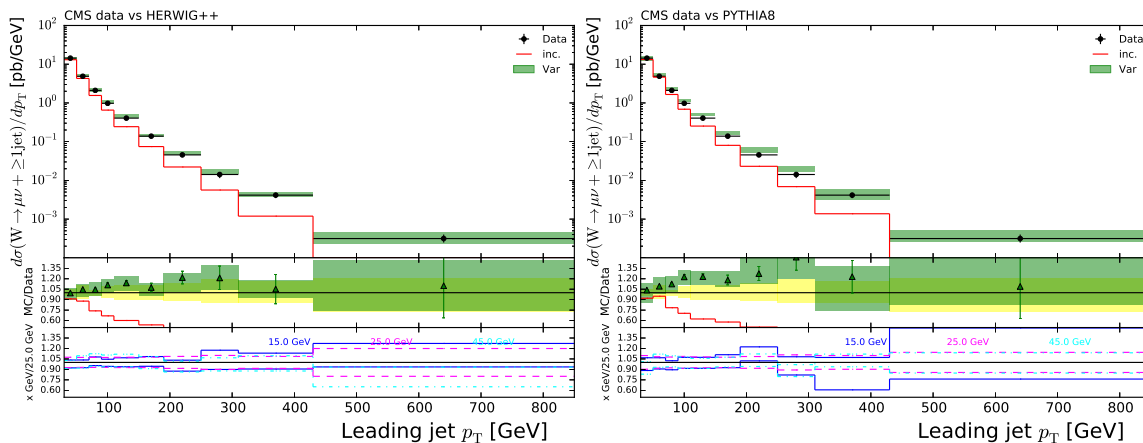
We present the  $N_{\text{jet}} \geq 1$   $H_T$  distribution in figure 36; we observe a pattern similar to that of the leading jet, with predictions slightly harder than data, and more markedly so in the case of PYTHIA8.

We now turn to observables which we compare only to HERWIG++ predictions. We start from single-jet ones, by considering the 2<sup>nd</sup> and 3<sup>rd</sup> jet transverse momenta (figure 37) and pseudorapidities (figure 38). There is no significant discrepancy between theory and data to report in any of these four plots. As far as the  $p_T$ 's are concerned, at variance with the case of the leading jet, our results have essentially the same hardness as the data; this has again to be compared to what is found in ref. [32] in the case of the MADGRAPH5 predictions.

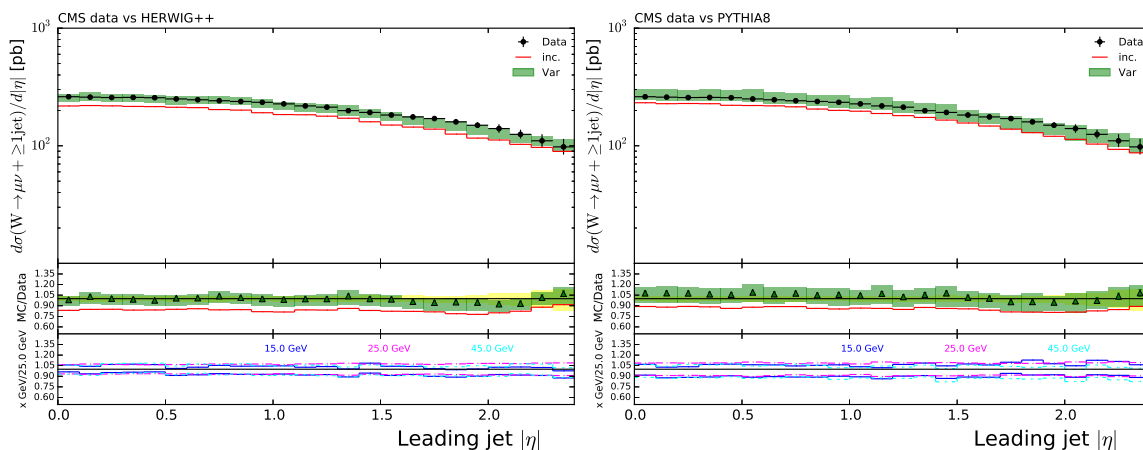
The  $\Delta\phi$  correlations between the muon and the 2<sup>nd</sup> and 3<sup>rd</sup> jets are given in figure 39, with our predictions again in fairly good agreement with data. Note that here the de-correlation is stronger than in the case of  $\Delta\phi(j_1, \mu)$ , but a small high-multiplicity matrix-element effect is still present, as can be seen by comparing the shapes of the merged results with those of the inclusive ones. Finally, in figure 40 we present the  $H_T$  distribution as measured in events with  $N_{\text{jet}} \geq 2$  and  $N_{\text{jet}} \geq 3$ . The agreement between theory and data is good, albeit with non-negligible systematics; in particular, our simulations are not harder than data, in contrast to what is found in ref. [32] in the case of LO-based predictions.



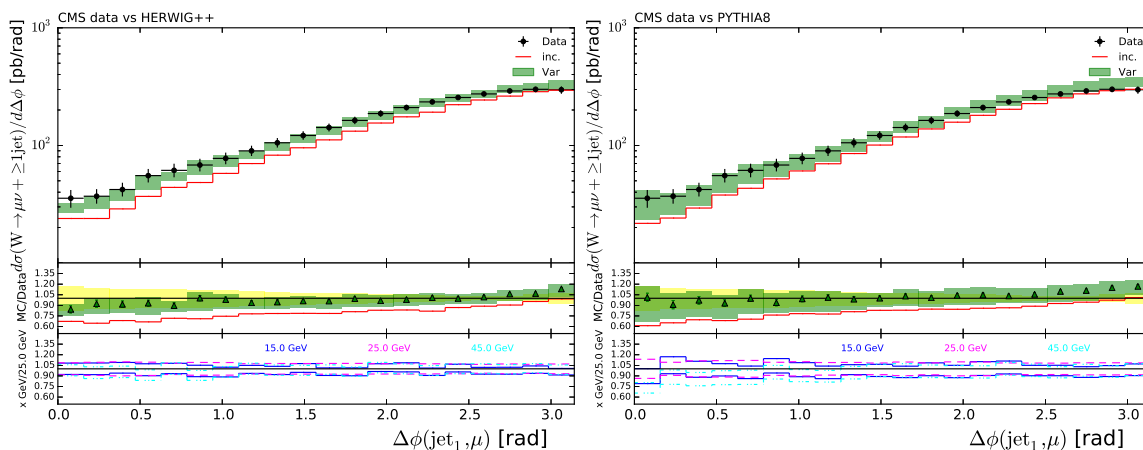
**Figure 32.** Exclusive jet multiplicity. Data from ref. [32], compared to HERWIG++ (left panel) and PYTHIA8 (right panel) predictions. The FxFx uncertainty envelope (“Var”) and the fully-inclusive central result (“inc”) are shown as green bands and red histograms respectively. See the end of section 2 for more details on the layout of the plots.



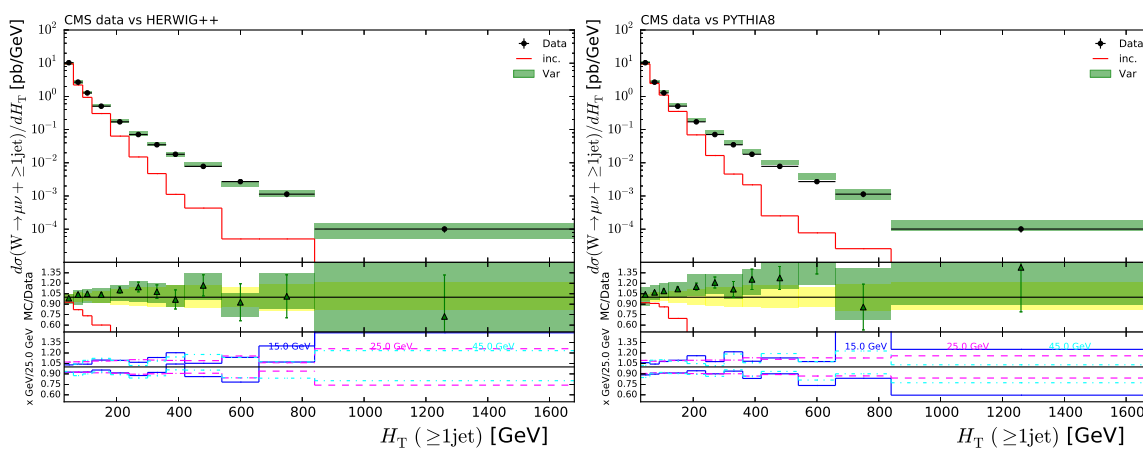
**Figure 33.** As in figure 32, for the transverse momentum of the 1<sup>st</sup> jet.



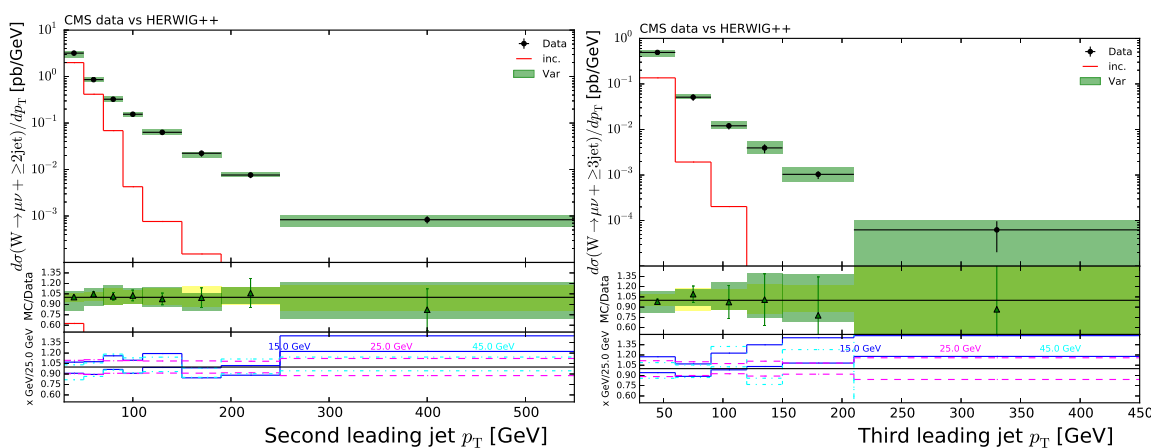
**Figure 34.** As in figure 32, for the pseudorapidity of the 1<sup>st</sup> jet.



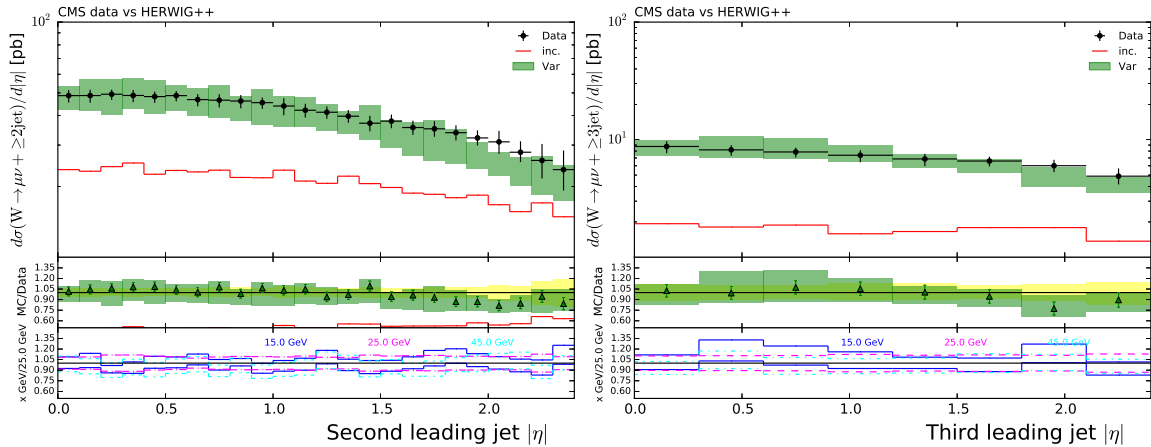
**Figure 35.** As in figure 32, for the azimuthal distance between the  $\mu$  and the 1<sup>st</sup> jet.



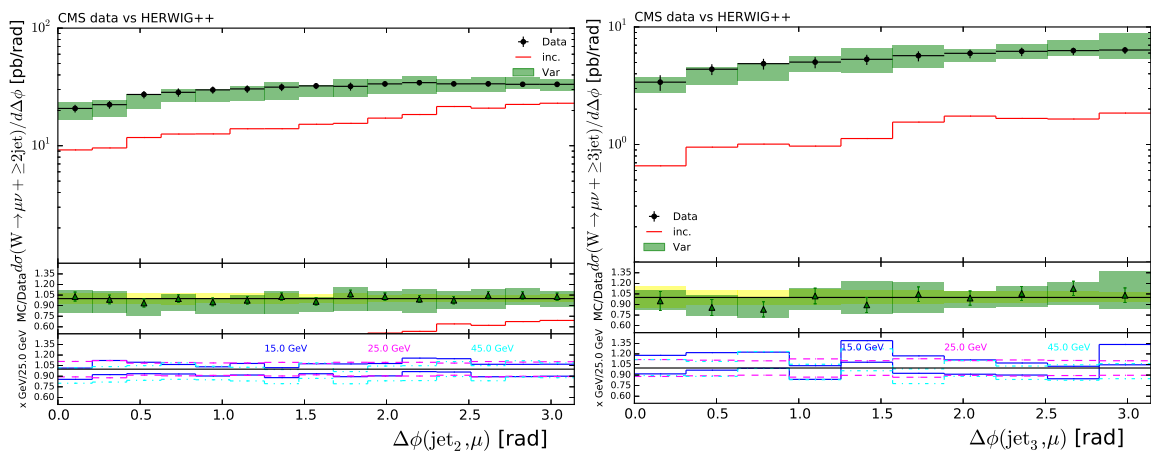
**Figure 36.** As in figure 32, for  $H_T$ .



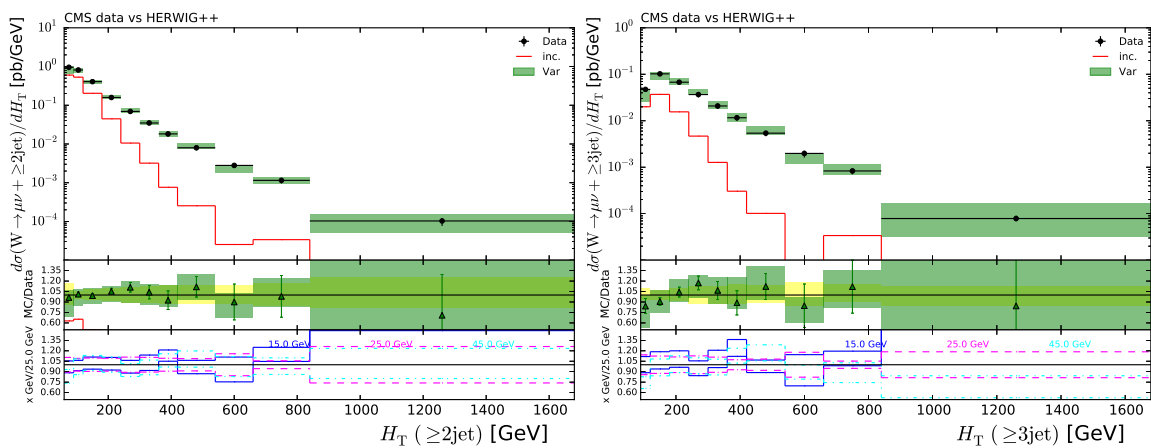
**Figure 37.** As in figure 32, for the transverse momentum of the 2<sup>nd</sup> (left panel) and of the 3<sup>rd</sup> (right panel) jet, both compared to HERWIG++ predictions.



**Figure 38.** As in figure 37, for the pseudorapidity of the 2<sup>nd</sup> (left panel) and of the 3<sup>rd</sup> (right panel) jet.



**Figure 39.** As in figure 37, for the azimuthal distance between the  $\mu$  and the 2<sup>nd</sup> jet (left panel), and between the  $\mu$  and the 3<sup>rd</sup> jet (right panel).



**Figure 40.** As in figure 37, for  $H_T$  in events with at least two (left panel) and at least three (right panel) jets.

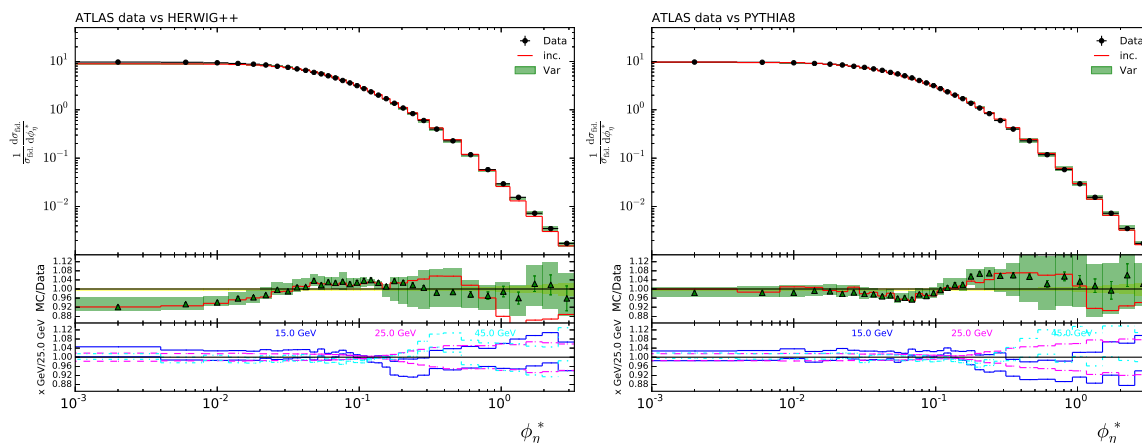
### 3.3 Inclusive observables

We have so far considered observables characterised by hard hadronic activity in association with a  $Z$  or a  $W$  boson, that emphasise the role that multi-parton matrix elements play in the context of merging techniques such as FxFx. On the other hand, one expects that physics-wise merged event samples improve on inclusive ones in every aspect, being equivalent to the latter for observables inclusive in QCD radiation, and able to treat seamlessly the transition between regimes with and without well-separated jets. These features are undoubtedly beneficial, since for all practical purposes the definition of what is meant by inclusive, or the occurrence of the transition between soft and hard kinematics, involve some arbitrariness, which becomes irrelevant when merged events are employed. Paradoxically, this might be problematic. In fact, MCs are tuned to data, and by far and large the tuning procedures are based on an LO inclusive picture (typically supplemented by matrix-element corrections); therefore, the tunes thus obtained, when used in a context of merged simulations (especially at the NLO), may degrade the quality of the theory-data agreement w.r.t. that which results from inclusive LO samples with LO-accurate parameters.

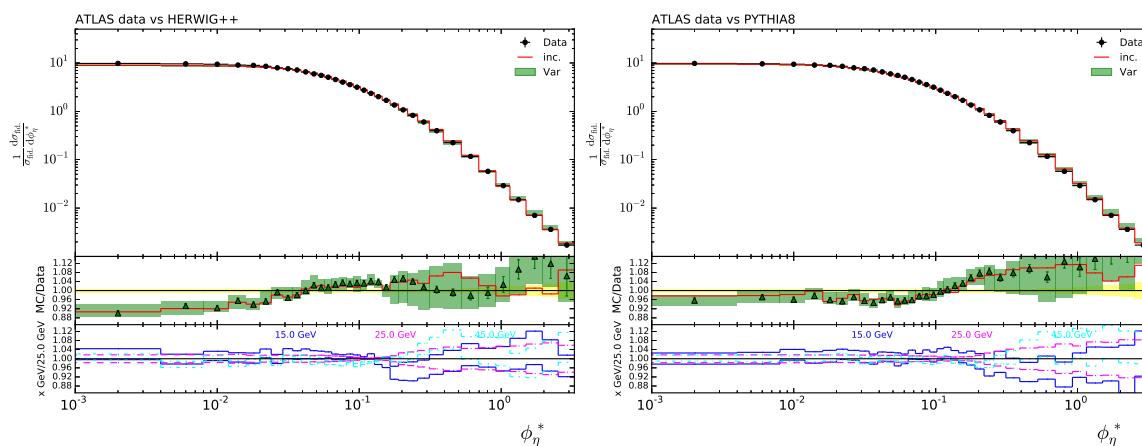
While in principle the solution to this problem is straightforward (i.e. to perform tunes with underlying (NLO-)merged samples), in practice there are several issues that require careful consideration (e.g. the possible interplay between multi-parton matrix elements and multi-particle underlying event (UE henceforth) models; the impossibility of imposing positivity constraints, simply because NLO PDFs are not positive definite; and so forth). A discussion of these issues is beyond the scope of the present work; we shall limit ourselves to considering the differences between merged and inclusive predictions for observables that, by construction, should be rather insensitive to hard radiation, using the measurements of refs. [33, 34] as benchmarks, analogously to what has been done in sects. 3.1 and 3.2. Such a comparison allows one to gauge the impact of the FxFx merging on the typical inclusive observables and thus, indirectly, to start addressing the question of whether NLO-merged tunes should be considered sooner than later.

**ATLAS [33] (arXiv:1211.6899, Rivet analysis ATLAS\_2012\_I1204784).** Measurement of the  $\phi_\eta^*$  angular correlation (see ref. [54] for its definition) in  $e^+e^-$  and  $\mu^+\mu^-$  production. Based on an integrated luminosity of  $4.6\text{ fb}^{-1}$ , within  $p_T(\ell) \geq 20\text{ GeV}$ ,  $66 \leq M(\ell\ell) \leq 116\text{ GeV}$ , and  $|\eta(\ell)| \leq 2.4$ . The Rivet analysis outputs the electron and muon results separately. To be definite, we shall discuss here only the comparison to  $e^+e^-$  data; the situation for  $\mu^+\mu^-$  data is very similar, with only minor differences which are not significant given the statistics of our theoretical simulations. The  $\phi_\eta^*$  observable is displayed as fully inclusive in the  $Z$  rapidity (figure 41), and within the  $|y_Z| < 0.8$ ,  $0.8 \leq |y_Z| < 1.6$ , and  $|y_Z| \geq 1.6$  regions (figure 42, figure 43, and figure 44, respectively). Note that  $\phi_\eta^*$ , on top of being inclusive, is also designed to emphasise the role of low- $p_T(Z)$  production.

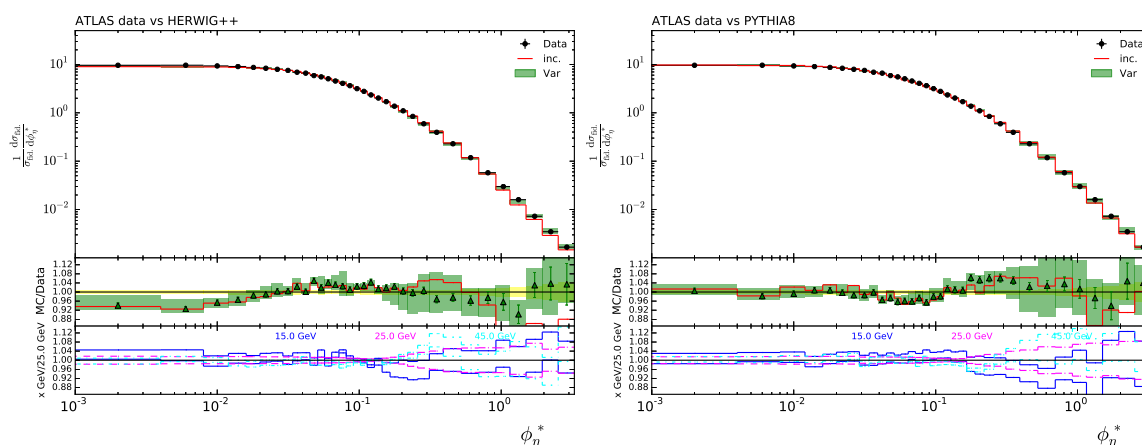
The results of figures 41–44 can be summarised as follows. Firstly, they confirm that at small  $\phi_\eta^*$  (i.e. at low  $p_T$ 's) the merged and inclusive predictions essentially coincide. However, the region where the two start to differ visibly depends on the specific MC employed:  $\phi_\eta^* \gtrsim 0.2$  for HERWIG++, and  $\phi_\eta^* \gtrsim 0.7$  for PYTHIA8, with some dependence on  $y_Z$ . Secondly, as implied by the previous item, at small  $\phi_\eta^*$  the level of the theory-data



**Figure 41.**  $\phi_\eta^*$  correlation in the  $e^+e^-$  channel, inclusive in the  $Z$  rapidity. Data from ref. [33], compared to HERWIG++ (left panel) and PYTHIA8 (right panel) predictions. The FxFx uncertainty envelope (“Var”) and the fully-inclusive central result (“inc”) are shown as green bands and red histograms respectively. See the end of section 2 for more details on the layout of the plots.

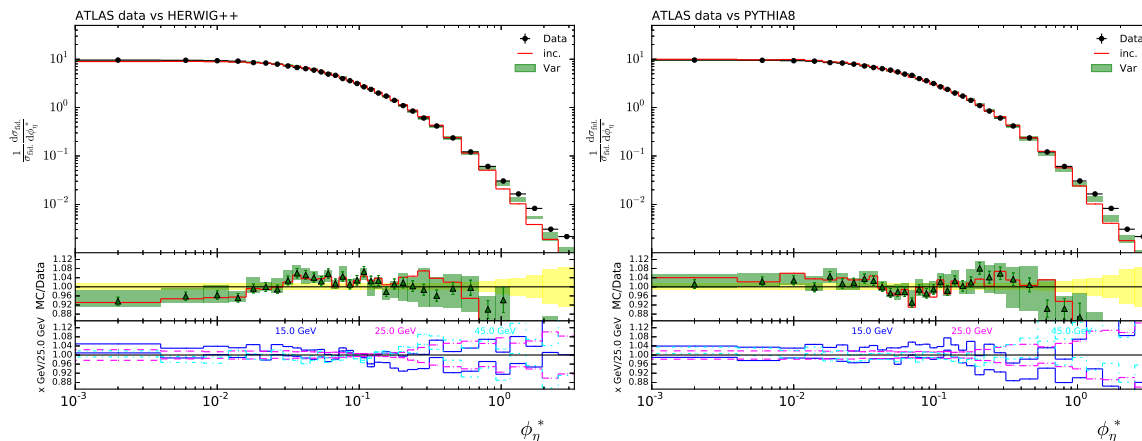


**Figure 42.** As in figure 41, for the  $|y_Z| < 0.8$  region.



**Figure 43.** As in figure 41, for the  $0.8 \leq |y_Z| < 1.6$  region.





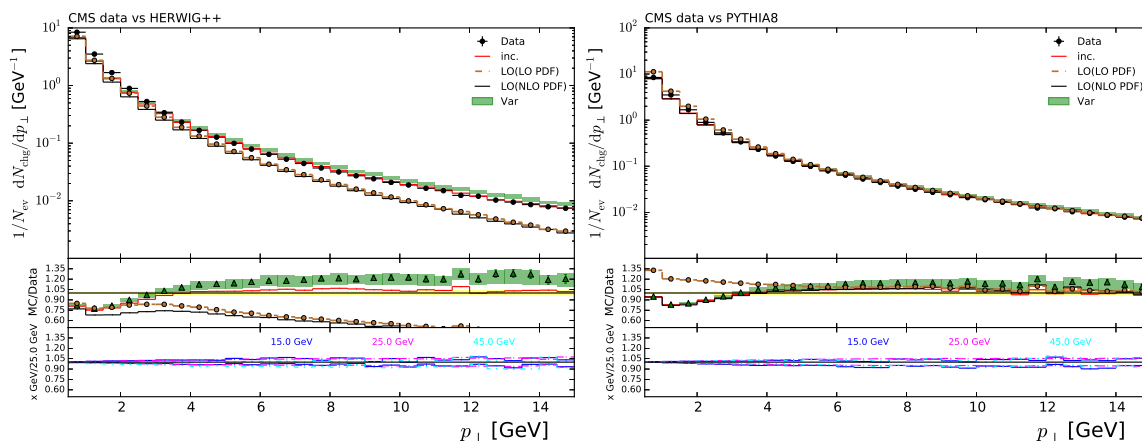
**Figure 44.** As in figure 41, for the  $|y_Z| \geq 1.6$  region.

agreement is driven by the MC, and therefore affected by the tune of the latter. Thus, the better description of the low- $\phi_\eta^*$  data provided by the PYTHIA8 simulations w.r.t. that of HERWIG++ must be interpreted in the context of the tunes used in this paper (see section 2). Thirdly, at large  $\phi_\eta^*$  the merged predictions from the two MCs tend to be rather consistent with each other. Since this region can largely be associated with hard-emission kinematic configurations,<sup>12</sup> this confirms what has been seen repeatedly in sects. 3.1 and 3.2. Namely, that in matrix-element-dominated regions features which are MC-specific matter much less; for the sake of the present discussion, this means essentially a significant *independence* of tunes (bar perhaps the choice of  $\alpha_s$  — see the beginning of section 3).

If one wants to use the  $\phi_\eta^*$  data as inputs for tuning, the implications of the previous discussion are the following. There exists an MC-specific small- $\phi_\eta^*$  region where the inclusion of high-multiplicity NLO matrix elements changes very little. In other words, if one tunes here, the results should be sufficiently “universal”, and allow one to employ them with simulations characterised by different perturbative accuracies. On the other hand, it may be desirable to have a larger lever arm when tuning, by including in the fits the intermediate- $\phi_\eta^*$  data (where again “intermediate” is an MC-dependent concept). This can be acceptable, but the resulting tunes will bear information on the underlying perturbative description, thus losing universality. It is this loss of universality which may be responsible for the degradation of the agreement with data alluded to at the beginning of this section. This phenomenon stems from what we may call *overtuning*: if the natural description of the intermediate region is one that requires information on higher perturbative orders, the lack of these in LO-based simulations will tend to be compensated (improperly) by the tuning, so that when such tunes are used with merged simulations a certain amount of double counting will be present.

**CMS [34] (arXiv:1204.1411, Rivet analysis CMS\_2012\_I1107658).** Measurement of the underlying event activity in  $\mu^+\mu^-$  production. Based on an integrated luminosity of

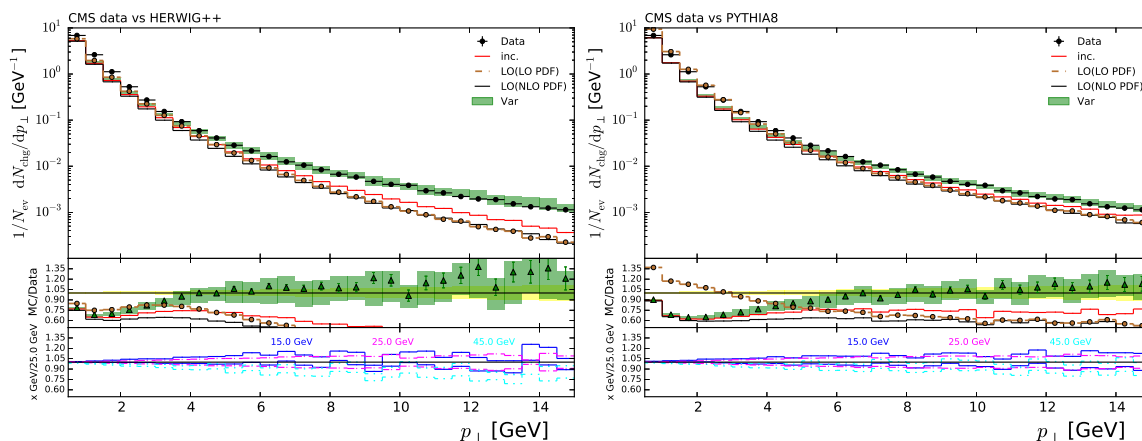
<sup>12</sup>This identification is hampered by the fact that the  $\phi_\eta^*$  observable squeezes the large- $p_T$  region in a few bins. From this viewpoint, an analysis in term of e.g. the transverse momentum of the lepton pair is superior w.r.t. the present one.



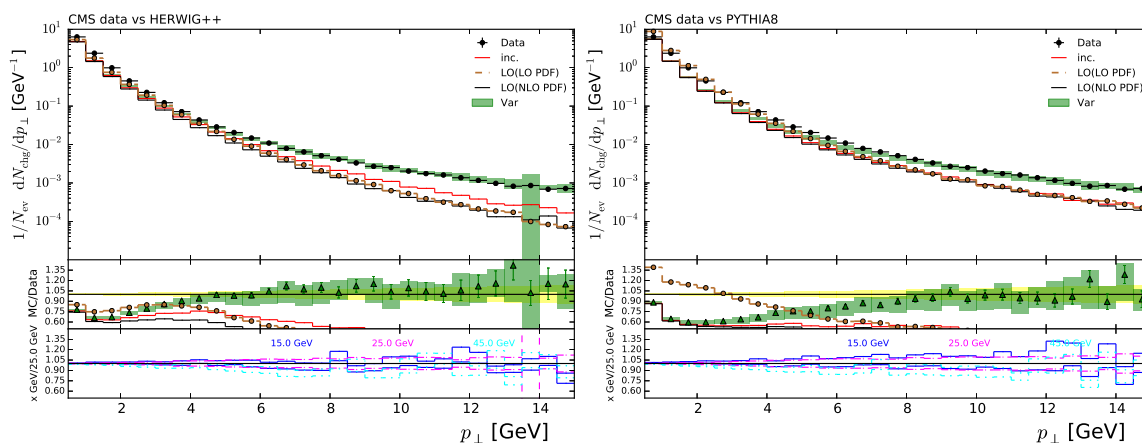
**Figure 45.** Inclusive charged particle multiplicity as a function of  $p_T$  in the away region. Data from ref. [34], compared to HERWIG++ (left panel) and PYTHIA8 (right panel) predictions. The FxFx uncertainty envelope (“Var”) and the fully-inclusive central result (“inc”) are shown as green bands and red histograms respectively. Two LO-accurate results are presented as solid black and dashed brown histograms. See the end of section 2 and the text of this section for more details on the layout of the plots.

$2.2 \text{ fb}^{-1}$ , within  $p_T(\mu) \geq 20 \text{ GeV}$ ,  $|\eta(\mu)| \leq 2.4$ , and either  $81 \leq M(\mu^+\mu^-) \leq 101 \text{ GeV}$  or  $p_T(\mu^+\mu^-) < 5 \text{ GeV}$ . The charged particles considered in the analysis must be within  $p_T > 0.5 \text{ GeV}$  and  $|\eta| < 2$ . We present them, as a function of the transverse momentum, for the away, transverse, and towards regions in figures 45, 46, and 47 respectively, in the case  $81 \leq M(\mu^+\mu^-) \leq 101 \text{ GeV}$ . These plots have the same layout as those shown so far, except for the fact that they contain two additional histograms in the main frame (their ratios to data are also reported in the upper insets). Such histograms are obtained with fully-inclusive simulations based on LO matrix elements, convoluted and showered either with NLO PDFs (solid black), or with LO PDFs (dashed brown, overlaid with full circles). We point out that the use of externally-generated LO events is associated with matrix-element corrections (MECs) in PYTHIA8; on the other hand, these corrections are not applied by HERWIG++.

By far and large, the towards, transverse, and away regions are defined as parallel, perpendicular, and anti-parallel to the direction of flight of the  $Z$ , respectively, through conditions on the azimuthal separation  $\Delta\phi$  between the  $Z$  and the charged tracks (for more details, we refer the reader to the original paper [34]). Thus, the away region is efficiently filled by the “first” hard parton that recoils against the  $Z$ . This is the reason why the merged and NLO-inclusive results are close to each other in figure 45 (with a better agreement between the two, and the data, in the case of PYTHIA8). Since PYTHIA8 applies MECs, the LO predictions are also in agreement with the NLO ones for such an MC starting from intermediate  $p_T$ ’s (and actually in the whole  $p_T$  range, significantly when NLO PDFs are employed — see later), while this is not the case for HERWIG++, given that hard matrix elements are completely lacking in this case (which allows one to see directly the dramatic impact of the hard recoil in the away region). Conversely, both the transverse and the towards region are designed to remove the contribution of the hard recoil; furthermore,



**Figure 46.** As in figure 45, for the transverse region.



**Figure 47.** As in figure 45, for the towards region.

higher-multiplicity hard configurations tend to contribute to the former but not to the latter. We see indeed in figures 46 and 47 that the differences between the merged and the non-merged (both NLO and LO) predictions are now significant. Note, in particular, that this is true also in the towards region (in spite of its being largely insensitive to hard radiation): this is a consequence of the fact that multi-parton matrix elements allow the MCs to emit more efficiently in all corners of the phase space, thanks to shorter colour strings or narrower radiation cones.

The agreement between data and NLO-merged predictions for  $p_T \gtrsim 6$  GeV (or lower, depending on the region and/or the MC) in all  $\Delta\phi$  regions and for both MCs<sup>13</sup> has to be compared with the fact that, with inclusive NLO or LO+MECs simulations, a satisfactory description of the data for such  $p_T$ 's can only be obtained in the away region. So while the observables considered here obviously provide one with excellent inputs for tuning, at the

<sup>13</sup>The only statistically-significant deviation is for HERWIG++ in the away region. In general, HERWIG++ is systematically slightly higher than PYTHIA8. Inspection of figures 45–47 suggests that, for these observables, HERWIG++ prefers a merging scale larger than PYTHIA8's.

same time the possibility of overtuning must be carefully assessed. The plots presented above show that with NLO-merged results the impact of tuning is mainly limited to small transverse momenta, regardless of the  $\Delta\phi$  region considered, while larger  $p_T$ 's are essentially unaffected. We note that this message comes out quite clearly *thanks to* the fact that the tunes we have adopted *do not* provide a good description of the small- $p_T$  data.<sup>14</sup>

Finally, as far as the small- $p_T$  cross sections are concerned, one sees immediately the expected good agreement between the two NLO-accurate results, in all the  $\Delta\phi$  regions. For  $p_T \lesssim 2\text{ GeV}$  (or larger), these two predictions also agree with the LO one obtained with NLO PDFs. This is interesting, because at low  $p_T$ 's we observe a very significant dependence (which disappears completely when moving towards larger  $p_T$ 's) of the LO simulations on whether the PDFs used are the NLO or the LO ones; furthermore, such a dependence is stronger for PYTHIA8 than for HERWIG++ (although the pattern NLO PDFs vs LO PDFs is similar in the two MCs). For what concerns the latter point, we have found that it is basically an accident: by changing the tunes and/or some of the MPI-related parameters in a given tune, the PDF dependence can be made weaker in PYTHIA8 and stronger in HERWIG++. This renders it clear that what drives the small- $p_T$  cross section is not the nominal perturbative accuracy of the PDFs in themselves, but rather their small- $x$  behaviours, through the smallest  $x$  values accessed within a given tune. In addition to this, we remark that at small  $p_T$ 's there is only a very minimal dependence on the perturbative accuracy of the short-distance results, when these are all convoluted and showered with the *same* PDFs. The conclusion is that a proper description of the small- $p_T$  data can only be achieved in the context of a fully self-consistent treatment of non-perturbative “parameters” (including the PDFs), and that this is largely independent of the nature of the hard events used in the simulations. We point out that such (approximate) universality is subject to the condition that one does not tune over the soft-hard transition region (let alone over the hard one): the very same considerations made above for the analysis of ref. [33] apply here. These findings heuristically confirm, in particular, the expectation that NLO+PS simulations based on NLO PDFs and tunes obtained with LO PDFs are liable to worsen the agreement with data at small transverse momenta *for fully-inclusive observables*. They also imply, however, that in the absence of a proper full-NLO tune a pragmatic solution is that of tuning using LO(+MECs) matrix elements with NLO PDFs, provided that one restricts oneself to genuine low- $p_T$  data.

## 4 Summary

In this paper we have presented a comprehensive comparison between NLO+PS theoretical predictions obtained by merging different partonic multiplicities according to the FxFx method, and several measurements performed by ATLAS and CMS for the leptonic channels of  $Z$ +jets and  $W$ +jets production. In order to deal only with well-established quantities and to minimise the number of operations on experimental results and definitions, we have

---

<sup>14</sup>We have however verified that the agreement theory-data significantly improves at small  $p_T$ 's e.g. by adopting the 4C tune in PYTHIA8 and the UE-EE-SCR-CTEQ6L1 tune in HERWIG++.

considered 7-TeV LHC datasets with high statistical accuracy (based on an integrated luminosity of about  $5\text{ fb}^{-1}$ ), and associated with a public Rivet routine. The primary aim of such a comparison is the validation of the FxFx procedure in an environment fully realistic from a phenomenological viewpoint.

Our computations have been carried out in the automated MADGRAPH5\_AMC@NLO framework, with parton showers performed by HERWIG++ and PYTHIA8. Thus, we have validated the relevant computer programs also in a technical sense: a single sample of unweighted hard events is produced simply through input commands (e.g. there is no *a posteriori* recombination of different multiplicities), and both HERWIG++ and PYTHIA8 automatically handle FxFx-merged samples.

The results we have obtained are based on underlying NLO matrix elements, with Born-level final states characterised by a lepton pair (thus taking fully into account the off-shellness of the vector boson, and spin correlations) and either zero, one, or two extra QCD partons. Therefore, the largest (tree-level) final-state multiplicity in our simulations features two leptons and three partons. We did not merge any LO-accurate samples for multiplicities larger than those included in our NLO matrix elements. We have used the same hard-event sample for all the  $Z$ +jets analyses (and analogously for the  $W$ +jets ones); in other words, we have not performed any analysis-specific event generation.

In the context of merged simulations, and especially for those that are NLO accurate, the scaling in the number of jets ( $N_{\text{jet}}$ ) must be regarded as a genuine theoretical prediction. For this reason, all of our differential distributions are presented as they result from the parton showers, in shapes *and* rates. We believe that, at least when the understanding of production mechanisms is paramount, the rescaling of theoretical results by an  $N_{\text{jet}}$ -dependent factor, extracted from data, is detrimental, because it entails a loss of predictivity, and tends to obscure the various strengths and weaknesses of the specific combination of matrix elements, merging method, and MC considered.

The overall agreement between our predictions and the  $V$ +jets data is good for both HERWIG++ and PYTHIA8. There is a very limited number of observables for which minor differences between theory and data are statistically significant, and these are also the only cases where the results of the two MCs may visibly differ. The general agreement between HERWIG++ and PYTHIA8 implies a significant reduction of the dependence on MC-specific assumptions, in favour of an increase in predictivity due to the inclusion of matrix-element information through merging. We also point out that the theory-data agreement extends to jet multiplicities which one would not expect to be predicted with the highest accuracy, given the matrix elements we have employed. It thus appears that NLO computations with up to two extra final-state partons are able to capture the most important dynamical effects in  $V$ +jets production, providing the MCs with suitable initial conditions for parton showers, which are never stretched outside their defining approximations. If taken at face value (i.e. ignoring both the theory and experimental uncertainties) certain  $p_T$ -related distributions in  $Z$ +jets production are softer in our predictions than in data; however, given the general picture it seems unlikely that this is a hint that  $Z$ +3 jets NLO matrix elements would be necessary (although of course it would be appealing to compute them). The only indication of this is in the theory being lower than data in the 4<sup>th</sup>-jet rate. We note that the typical

LO-merged samples used by the experiments for  $V$ +jets production feature much larger parton multiplicities (at least four extra, and typically more). It is therefore an interesting question that of whether such high-multiplicity LO matrix elements are truly *all* needed in order to compensate for the lack of genuine NLO effects. Conversely, one may wonder whether merging procedures at the NLO are in any case superior to their LO counterparts, irrespective of the matrix elements included in the computations (obviously, beyond a minimal threshold).

One aspect where NLO mergings improve by construction on LO ones is in the reduction of the theoretical systematics. In this paper, we have assessed three sources of uncertainties, due to the choices of hard scales, PDFs, and merging scale. We remind the reader that in MADGRAPH5\_AMC@NLO the variations of hard scales and PDFs do not entail any separate runs, since the results for each individual choice are stored in the hard-event file as companion weights, correlated with the main event weight (it is such a correlation that allows one to use a single shower history for all weights, and hence to significantly improve the numerical stability of the final results). It is crucial that compatibility with this feature be systematically incorporated in modern simulation and analysis software (such as Rivet). As far as merging-scale dependence is concerned, we have considered a rather large range, in order to be as conservative as possible. Still, this kind of systematics appears to be under good control; in particular, shape-wise the different choices lead to fairly similar results.

We have also presented NLO-merged results for inclusive and underlying-event observables in  $Z$  production. As expected, in the kinematic regions associated with low- $p_T$  emissions such results are in agreement with those obtained from a fully-inclusive NLO sample; both are in fact ultimately determined by the underlying MC. In these cases, and for the specific tunes we have adopted in this paper, the description of the data is worse, and more MC-dependent, than for the  $Z$ +jets observables. This is just as well, because it gives one some further evidence that multi-jet observables are fairly independent of the general assumptions made in individual MCs, and of their tunes in particular.

We have not made any attempt to improve the description of low- $p_T$  observables by considering different tunes, but have limited ourselves to verifying that this is indeed possible. Rather, we have employed such observables to give explicit examples of how the use of merged predictions could lead to an issue we have called overtuning, which is basically a double-counting problem. In essence, one might use the differences between merged results and those adopted in the context of tuning (typically, LO plus matrix-element corrections) as a way to assess the possible impact of high-multiplicity matrix elements on tunable observables. When this impact is non negligible, a new tuning specific for merged samples may be desirable. Conversely, the differences mentioned above could help determine the kinematic regions to be used to obtain “universal” tunes, that can be safely adopted regardless of the perturbative accuracy of the simulations.

We have also shown that at small transverse momenta the role of the PDFs chosen for tuning is crucial. This implies that, in such kinematic regions, the use of NLO PDFs within MCs whose tunes are based on LO PDFs is less than ideal. Fortunately, our  $V$ +jets results demonstrate that this possible mismatch is essentially irrelevant in most of the phase space.

While eventually the derivation of proper NLO(-merged) tunes will offer the best overall option, a lower-cost alternative for the near future is that of performing MC tunes as they are done presently, but adopting NLO PDF sets.

Finally, we remark that while the validation of the FxFx merging carried out in this paper touches all of the general aspects of the procedure, it applies directly to the simulations of  $WH$ +jets and  $ZH$ +jets production, with  $H$  the SM Higgs, owing to the similarity of these processes with the  $V$ +jets ones. It is thus compelling to consider such Higgs associated channels in the context of merged predictions, in particular given the relevance of jet-multiplicity categorisation in experimental searches.

## Acknowledgments

We thank Fabio Maltoni for his comments on this manuscript, Peter Richardson and Torbjörn Sjöstrand for clarifications on and help with HERWIG++ and PYTHIA8, respectively, and Francesco Giuli for having tested independently MADGRAPH5\_AMC@NLO with  $V$ +jets production. We are indebted to Fabio Cossutti, Christian Gütschow, Joey Huston, Jan Kretschmar, Sascha Savin, and Alessandro Tricoli for having provided us with information and material relevant to ATLAS and CMS analyses. The work of RF is supported by the Alexander von Humboldt Foundation, in the framework of the Sofja Kovaleskaja Award Project “Event Simulation for the Large Hadron Collider at High Precision”, endowed by the German Federal Ministry of Education and Research. SF thanks the CERN TH Unit for hospitality during the course of this work. AP acknowledges support by the MCnet ITN FP7 Marie Curie Initial Training Network PITN-GA-2012-315877, and a Marie Curie Intra European Fellowship within the 7th European Community Framework Programme (grant no. PIEF-GA-2013-622071). SP is supported by the US Department of Energy under contract DE-AC02-76SF00515. The work of PT has received funding from the European Union Seventh Framework programme for research and innovation under the Marie Curie grant agreement N. 609402–2020 researchers: Train to Move (T2M). This work has been supported in part by the ERC grant 291377 “LHCtheory: Theoretical predictions and analyses of LHC physics: advancing the precision frontier”.

**Open Access.** This article is distributed under the terms of the Creative Commons Attribution License ([CC-BY 4.0](https://creativecommons.org/licenses/by/4.0/)), which permits any use, distribution and reproduction in any medium, provided the original author(s) and source are credited.

## References

- [1] P. Ciafaloni and D. Comelli, *Sudakov enhancement of electroweak corrections*, *Phys. Lett. B* **446** (1999) 278 [[hep-ph/9809321](#)] [[INSPIRE](#)].
- [2] M. Ciafaloni, P. Ciafaloni and D. Comelli, *Bloch-Nordsieck violating electroweak corrections to inclusive TeV scale hard processes*, *Phys. Rev. Lett.* **84** (2000) 4810 [[hep-ph/0001142](#)] [[INSPIRE](#)].
- [3] A. Denner and S. Pozzorini, *One loop leading logarithms in electroweak radiative corrections. 1. Results*, *Eur. Phys. J. C* **18** (2001) 461 [[hep-ph/0010201](#)] [[INSPIRE](#)].

- [4] A. Denner and S. Pozzorini, *One loop leading logarithms in electroweak radiative corrections. 2. Factorization of collinear singularities*, *Eur. Phys. J. C* **21** (2001) 63 [[hep-ph/0104127](#)] [[INSPIRE](#)].
- [5] S. Catani, F. Krauss, R. Kuhn and B.R. Webber, *QCD matrix elements + parton showers*, *JHEP* **11** (2001) 063 [[hep-ph/0109231](#)] [[INSPIRE](#)].
- [6] L. Lönnblad, *Correcting the color dipole cascade model with fixed order matrix elements*, *JHEP* **05** (2002) 046 [[hep-ph/0112284](#)] [[INSPIRE](#)].
- [7] F. Krauss, *Matrix elements and parton showers in hadronic interactions*, *JHEP* **08** (2002) 015 [[hep-ph/0205283](#)] [[INSPIRE](#)].
- [8] S. Mrenna and P. Richardson, *Matching matrix elements and parton showers with HERWIG and PYTHIA*, *JHEP* **05** (2004) 040 [[hep-ph/0312274](#)] [[INSPIRE](#)].
- [9] N. Lavesson and L. Lönnblad, *W + jets matrix elements and the dipole cascade*, *JHEP* **07** (2005) 054 [[hep-ph/0503293](#)] [[INSPIRE](#)].
- [10] J. Alwall et al., *Comparative study of various algorithms for the merging of parton showers and matrix elements in hadronic collisions*, *Eur. Phys. J. C* **53** (2008) 473 [[arXiv:0706.2569](#)] [[INSPIRE](#)].
- [11] S. Hoeche, F. Krauss, S. Schumann and F. Siegert, *QCD matrix elements and truncated showers*, *JHEP* **05** (2009) 053 [[arXiv:0903.1219](#)] [[INSPIRE](#)].
- [12] K. Hamilton, P. Richardson and J. Tully, *A Modified CKKW matrix element merging approach to angular-ordered parton showers*, *JHEP* **11** (2009) 038 [[arXiv:0905.3072](#)] [[INSPIRE](#)].
- [13] L. Lönnblad and S. Prestel, *Matching Tree-Level Matrix Elements with Interleaved Showers*, *JHEP* **03** (2012) 019 [[arXiv:1109.4829](#)] [[INSPIRE](#)].
- [14] L. Lönnblad and S. Prestel, *Unitarising Matrix Element + Parton Shower merging*, *JHEP* **02** (2013) 094 [[arXiv:1211.4827](#)] [[INSPIRE](#)].
- [15] N. Lavesson and L. Lönnblad, *Extending CKKW-merging to One-Loop Matrix Elements*, *JHEP* **12** (2008) 070 [[arXiv:0811.2912](#)] [[INSPIRE](#)].
- [16] K. Hamilton and P. Nason, *Improving NLO-parton shower matched simulations with higher order matrix elements*, *JHEP* **06** (2010) 039 [[arXiv:1004.1764](#)] [[INSPIRE](#)].
- [17] S. Hoche, F. Krauss, M. Schonherr and F. Siegert, *NLO matrix elements and truncated showers*, *JHEP* **08** (2011) 123 [[arXiv:1009.1127](#)] [[INSPIRE](#)].
- [18] W.T. Giele, D.A. Kosower and P.Z. Skands, *Higher-Order Corrections to Timelike Jets*, *Phys. Rev. D* **84** (2011) 054003 [[arXiv:1102.2126](#)] [[INSPIRE](#)].
- [19] S. Alioli, K. Hamilton and E. Re, *Practical improvements and merging of POWHEG simulations for vector boson production*, *JHEP* **09** (2011) 104 [[arXiv:1108.0909](#)] [[INSPIRE](#)].
- [20] S. Hoeche, F. Krauss, M. Schonherr and F. Siegert, *QCD matrix elements + parton showers: The NLO case*, *JHEP* **04** (2013) 027 [[arXiv:1207.5030](#)] [[INSPIRE](#)].
- [21] R. Frederix and S. Frixione, *Merging meets matching in MC@NLO*, *JHEP* **12** (2012) 061 [[arXiv:1209.6215](#)] [[INSPIRE](#)].
- [22] S. Plätzer, *Controlling inclusive cross sections in parton shower + matrix element merging*, *JHEP* **08** (2013) 114 [[arXiv:1211.5467](#)] [[INSPIRE](#)].



- [23] S. Alioli et al., *Combining Higher-Order Resummation with Multiple NLO Calculations and Parton Showers in GENEVA*, *JHEP* **09** (2013) 120 [[arXiv:1211.7049](#)] [[INSPIRE](#)].
- [24] L. Lönnblad and S. Prestel, *Merging Multi-leg NLO Matrix Elements with Parton Showers*, *JHEP* **03** (2013) 166 [[arXiv:1211.7278](#)] [[INSPIRE](#)].
- [25] K. Hamilton, P. Nason, C. Oleari and G. Zanderighi, *Merging  $H/W/Z + 0$  and 1 jet at NLO with no merging scale: a path to parton shower + NNLO matching*, *JHEP* **05** (2013) 082 [[arXiv:1212.4504](#)] [[INSPIRE](#)].
- [26] S. Alioli, C.W. Bauer, C. Berggren, F.J. Tackmann, J.R. Walsh and S. Zuberi, *Matching Fully Differential NNLO Calculations and Parton Showers*, *JHEP* **06** (2014) 089 [[arXiv:1311.0286](#)] [[INSPIRE](#)].
- [27] A. Buckley et al., *Rivet user manual*, *Comput. Phys. Commun.* **184** (2013) 2803 [[arXiv:1003.0694](#)] [[INSPIRE](#)].
- [28] ATLAS collaboration, *Measurement of the production cross section of jets in association with a Z boson in pp collisions at  $\sqrt{s} = 7$  TeV with the ATLAS detector*, *JHEP* **07** (2013) 032 [[arXiv:1304.7098](#)] [[INSPIRE](#)].
- [29] CMS collaboration, *Rapidity distributions in exclusive Z + jet and  $\gamma$  + jet events in pp collisions at  $\sqrt{s} = 7$  TeV*, *Phys. Rev. D* **88** (2013) 112009 [[arXiv:1310.3082](#)] [[INSPIRE](#)].
- [30] CMS collaboration, *Event shapes and azimuthal correlations in Z + jets events in pp collisions at  $\sqrt{s} = 7$  TeV*, *Phys. Lett. B* **722** (2013) 238 [[arXiv:1301.1646](#)] [[INSPIRE](#)].
- [31] ATLAS collaboration, *Measurements of the W production cross sections in association with jets with the ATLAS detector*, *Eur. Phys. J. C* **75** (2015) 82 [[arXiv:1409.8639](#)] [[INSPIRE](#)].
- [32] CMS collaboration, *Differential cross section measurements for the production of a W boson in association with jets in proton-proton collisions at  $\sqrt{s} = 7$  TeV*, *Phys. Lett. B* **741** (2015) 12 [[arXiv:1406.7533](#)] [[INSPIRE](#)].
- [33] ATLAS collaboration, *Measurement of angular correlations in Drell-Yan lepton pairs to probe Z/ $\gamma^*$  boson transverse momentum at  $\sqrt{s} = 7$  TeV with the ATLAS detector*, *Phys. Lett. B* **720** (2013) 32 [[arXiv:1211.6899](#)] [[INSPIRE](#)].
- [34] CMS collaboration, *Measurement of the underlying event in the Drell-Yan process in proton-proton collisions at  $\sqrt{s} = 7$  TeV*, *Eur. Phys. J. C* **72** (2012) 2080 [[arXiv:1204.1411](#)] [[INSPIRE](#)].
- [35] J. Alwall et al., *The automated computation of tree-level and next-to-leading order differential cross sections and their matching to parton shower simulations*, *JHEP* **07** (2014) 079 [[arXiv:1405.0301](#)] [[INSPIRE](#)].
- [36] M. Bahr et al., *HERWIG++ Physics and Manual*, *Eur. Phys. J. C* **58** (2008) 639 [[arXiv:0803.0883](#)] [[INSPIRE](#)].
- [37] S. Gieseke et al., *HERWIG++ 2.5 Release Note*, [arXiv:1102.1672](#) [[INSPIRE](#)].
- [38] K. Arnold et al., *HERWIG++ 2.6 Release Note*, [arXiv:1205.4902](#) [[INSPIRE](#)].
- [39] J. Bellm et al., *HERWIG++ 2.7 Release Note*, [arXiv:1310.6877](#) [[INSPIRE](#)].
- [40] T. Sjöstrand, S. Mrenna and P.Z. Skands, *A Brief Introduction to PYTHIA 8.1*, *Comput. Phys. Commun.* **178** (2008) 852 [[arXiv:0710.3820](#)] [[INSPIRE](#)].
- [41] T. Sjöstrand et al., *An Introduction to PYTHIA 8.2*, *Comput. Phys. Commun.* **191** (2015) 159 [[arXiv:1410.3012](#)] [[INSPIRE](#)].

- [42] S. Frixione and B.R. Webber, *Matching NLO QCD computations and parton shower simulations*, *JHEP* **06** (2002) 029 [[hep-ph/0204244](#)] [[INSPIRE](#)].
- [43] S. Catani, Y.L. Dokshitzer, M.H. Seymour and B.R. Webber, *Longitudinally invariant  $K_t$  clustering algorithms for hadron hadron collisions*, *Nucl. Phys. B* **406** (1993) 187 [[INSPIRE](#)].
- [44] J. Alwall, S. de Visscher and F. Maltoni, *QCD radiation in the production of heavy colored particles at the LHC*, *JHEP* **02** (2009) 017 [[arXiv:0810.5350](#)] [[INSPIRE](#)].
- [45] R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, R. Pittau and P. Torrielli, *Four-lepton production at hadron colliders: aMC@NLO predictions with theoretical uncertainties*, *JHEP* **02** (2012) 099 [[arXiv:1110.4738](#)] [[INSPIRE](#)].
- [46] J. Butterworth et al., *Les Houches 2013: Physics at TeV Colliders: Standard Model Working Group Report*, [arXiv:1405.1067](#) [[INSPIRE](#)].
- [47] R.D. Ball et al., *Parton distributions with LHC data*, *Nucl. Phys. B* **867** (2013) 244 [[arXiv:1207.1303](#)] [[INSPIRE](#)].
- [48] A. Buckley et al., *LHAPDF6: parton density access in the LHC precision era*, *Eur. Phys. J. C* **75** (2015) 132 [[arXiv:1412.7420](#)] [[INSPIRE](#)].
- [49] P.Z. Skands, S. Carrazza and J. Rojo, *Tuning PYTHIA 8.1: the Monash 2013 Tune*, *Eur. Phys. J. C* **74** (2014) 3024 [[arXiv:1404.5630](#)] [[INSPIRE](#)].
- [50] S. Catani, B.R. Webber and G. Marchesini, *QCD coherent branching and semiinclusive processes at large  $x$* , *Nucl. Phys. B* **349** (1991) 635 [[INSPIRE](#)].
- [51] M. Cacciari, G.P. Salam and G. Soyez, *The anti- $k_t$  jet clustering algorithm*, *JHEP* **04** (2008) 063 [[arXiv:0802.1189](#)] [[INSPIRE](#)].
- [52] P. Torrielli and S. Frixione, *Matching NLO QCD computations with PYTHIA using MC@NLO*, *JHEP* **04** (2010) 110 [[arXiv:1002.4293](#)] [[INSPIRE](#)].
- [53] J.R. Gaunt, *Glauber Gluons and Multiple Parton Interactions*, *JHEP* **07** (2014) 110 [[arXiv:1405.2080](#)] [[INSPIRE](#)].
- [54] A. Banfi, S. Redford, M. Vesterinen, P. Waller and T.R. Wyatt, *Optimisation of variables for studying dilepton transverse momentum distributions at hadron colliders*, *Eur. Phys. J. C* **71** (2011) 1600 [[arXiv:1009.1580](#)] [[INSPIRE](#)].