



---

# Measurements of observables sensitive to colour reconnection in $t\bar{t}$ events with the ATLAS detector at $\sqrt{s} = 13$ TeV

The ATLAS Collaboration

A measurement of observables sensitive to effects of colour reconnection in top-quark pair-production events is presented using  $139 \text{ fb}^{-1}$  of 13 TeV proton–proton collision data collected by the ATLAS detector at the LHC. Events are selected by requiring exactly one isolated electron and one isolated muon with opposite charge and two or three jets, where exactly two jets are required to be  $b$ -tagged. For the selected events, measurements are presented for the charged-particle multiplicity, the scalar sum of the transverse momenta of the charged particles, and the same scalar sum in bins of charged-particle multiplicity. These observables are unfolded to the stable-particle level, thereby correcting for migration effects due to finite detector resolution, acceptance and efficiency effects. The particle-level measurements are compared with different colour reconnection models in Monte Carlo generators. These measurements disfavour some of the colour reconnection models and provide inputs to future optimisation of the parameters in Monte Carlo generators.

# Contents

<b>1</b>	<b>Introduction</b>	<b>3</b>
<b>2</b>	<b>Colour reconnection in <math>t\bar{t}</math> events</b>	<b>4</b>
2.1	Colour reconnection models in PYTHIA 8	5
2.2	Colour reconnection models in HERWIG 7	6
<b>3</b>	<b>ATLAS detector</b>	<b>7</b>
<b>4</b>	<b>Samples of data and simulated events</b>	<b>8</b>
4.1	Simulation of $t\bar{t}$ , $tW$ and $t\bar{t}+V$ production	9
4.2	Simulation of $Z$ +jets and diboson production	9
<b>5</b>	<b>Object reconstruction and definitions of observables</b>	<b>10</b>
5.1	Detector-level object definitions	10
5.2	Particle-level object definitions	11
5.3	Event selection	12
5.4	Definition of the observables at particle level and detector level	12
<b>6</b>	<b>Background estimation and corrections to observables</b>	<b>13</b>
6.1	Event-based backgrounds	13
6.2	Track-based backgrounds	17
6.3	Corrections to observables	19
<b>7</b>	<b>Unfolding</b>	<b>21</b>
<b>8</b>	<b>Systematic uncertainties</b>	<b>22</b>
8.1	Detector systematic uncertainties	22
8.2	Track-based-background uncertainties	25
8.3	Event-based-background uncertainties	25
8.4	Unfolding-technique uncertainties	26
8.5	$t\bar{t}$ modelling uncertainties	26
8.6	Statistical uncertainties	28
8.7	Luminosity	28
8.8	Impact of scale variations in PYTHIA 8	31
<b>9</b>	<b>Results</b>	<b>33</b>
9.1	Comparisons with predictions from different generators	34
9.2	Comparison with different CR models in PYTHIA 8	36
9.3	PYTHIA 8 CR and UE parameters	36
9.4	HERWIG 7.2 CR models	39
9.5	Top-quark-specific CR models of PYTHIA 8	39
<b>10</b>	<b>Summary and conclusion</b>	<b>42</b>

<b>Appendices</b>	<b>42</b>
<b>A Migration matrices</b>	<b>42</b>
<b>B Absolute differential cross-sections</b>	<b>44</b>
<b>C Additional material for the measured distributions</b>	<b>44</b>

## 1 Introduction

Monte Carlo event generators are of the utmost importance for experimental particle physics as a tool providing a source of simulated events for analysis design, signal and background modelling, unfolding detector effects, and training of multivariate discriminants. A rich set of measurements provided by the Large Hadron Collider (LHC) experiments can be used to probe both soft and hard quantum chromodynamics (QCD), and thereby test many features of the Monte Carlo (MC) event generators. Even though the overall performance of the event generators is good, the understanding of some phenomena is insufficiently understood or can be significantly improved. One of these is the modelling of colour reconnection (CR). None of the models describing CR in hadron collisions is derived from first principles of QCD. Measurements are required to exclude certain models or to constrain their parameter space.

The simulation of the hard-scattering process is accompanied by the so-called underlying event (UE) consisting of multiple parton interactions (MPIs), initial- and final-state radiation (ISR and FSR), and remnants of the interacting protons. According to QCD the strong interactions between quarks and gluons are related to their colour charges. In order to trace the colour information in an event, MC event generators make use of the leading-colour (LC) approximation [1, 2]. Here, each quark is colour-connected to only one other parton (quark or gluon), while gluons are each connected to two other partons, since they carry both a colour and an anticolour charge. Additional colour lines are added to the final state by each MPI system, leading to a non-negligible possibility of phase-space overlaps between final states from different MPI systems and the hard-scattering process [3, 4]. Colour reconnection [5] models introduce a mechanism for reassigning colour connections between partons to resolve ambiguities between overlapping colour lines in space and time, and thus allow colour topologies different from those in the simple LC approach. The interactions and interference among the different processes leading to a given generated event are poorly understood and several models are implemented in the three parton-shower MC programs, PYTHIA 8 [6], HERWIG 7 [7], and SHERPA 2 [8].

In general, these models define some momentum-based measure of the colour configuration of partons in the event. In their workflow, the programs execute changes to the colour connections to reduce the defined measure and thereby correct for deficiencies of the LC approximation and allow colour connections between different MPIs. The modified colour configuration effectively leads to a reduction of the energy available to produce new particle–antiparticle pairs in the space between partons. Thus, the particle multiplicity is reduced and the produced particles carry, on average, more momentum than they would have in events without CR. Details can be found in Refs. [4, 9] and references therein.

An unresolved question is the involvement of the top quark and its decay products in CR. The typical hadronisation scale is about 1 fm, while the average transverse decay length of top quarks is 0.2 fm [10]. Thus, it is possible that top quarks interact with the colour fields stretching between partons in the final state [11]. However, in the current implementation of CR models, the reconnection algorithm is always

performed before the top quark decays, i.e. only the top quark itself takes part in the CR machinery, not its decay products.

In recent top-quark mass ( $m_{\text{top}}$ ) measurements, which reach a precision well below 1 GeV [12, 13], the uncertainty in the CR modelling is estimated to be as large as 400 MeV [13]. Thus, this uncertainty is becoming increasingly important as direct measurements of  $m_{\text{top}}$  become more precise.

Several measurements sensitive to the modelling of the UE and (partially) to CR in different scattering processes, e.g. minimum-bias events [14],  $Z$  boson production [15–17] and top-quark–antiquark ( $t\bar{t}$ ) production [18] events, have been performed at the LHC. However, dedicated measurements of CR in top-quark events have not been performed yet. This is particularly relevant since a well-defined prescription of the uncertainty due to CR is needed for top-quark mass measurements.

This article presents measurements of normalised differential cross-sections of  $t\bar{t}$  production as a function of observables sensitive to CR effects, namely the charged-particle multiplicity outside of jets ( $n_{\text{ch}}$ ), excluding leptons from the top-quark decay; the scalar sum of the transverse momenta of these charged particles ( $\sum_{n_{\text{ch}}} p_{\text{T}}$ ); and the double-differential measurement of the two quantities ( $\sum_{n_{\text{ch}}} p_{\text{T}}$  in bins of  $n_{\text{ch}}$ ). The last of these measurements contains partial information about the ‘standard’ observable which is often used in CR and UE measurements: the average transverse momentum of charged particles as a function of the charged-particle multiplicity. The observables are corrected for tracks from additional inelastic  $pp$  collisions in the same or neighbouring bunch crossings (pile-up), for tracks from secondary vertices, and for the effect of tracking inefficiencies. The distributions of the corrected observables are unfolded to the particle level in a fiducial phase space chosen to maximise the sensitivity to CR effects. The fiducial cross-section is obtained by integrating the differential cross-section.

In proton–proton collisions, the dominant top-quark production process is  $t\bar{t}$  production via the strong interaction. As a result of top-quark and top-antiquark decays, two on-shell  $W$  bosons emerge in  $t\bar{t}$  events. In order to select a nearly background-free sample of  $t\bar{t}$  events, this analysis targets events in which one  $W$  boson decays in the  $W \rightarrow e\nu$  channel and the second  $W$  boson decays via  $W \rightarrow \mu\nu$ .<sup>1</sup> Excluding dilepton final states with two charged leptons of the same flavour ( $e^+e^-$  or  $\mu^+\mu^-$ ) essentially eliminates the background arising from  $Z$ +jets production. Thus, the only relevant backgrounds are single top-quark production in the  $tW$  channel and events where jets are misidentified as leptons. The experimental signature of the selected events is given by exactly one prompt isolated electron, exactly one prompt isolated muon, and two or three hadronic jets – two of which each originate from a  $b$ -quark and are called  $b$ -jets. The selected events were recorded with the ATLAS detector [19] during Run 2 of the LHC in proton–proton collisions at a centre-of-mass energy of 13 TeV. The resulting data set corresponds to an integrated luminosity of 139 fb<sup>-1</sup>.

## 2 Colour reconnection in $t\bar{t}$ events

In the following, the CR models implemented in the two multipurpose MC generators, PYTHIA 8 and HERWIG 7, are described. The SHERPA 2 event generator has an MPI model based on the PYTHIA one, but without any CR model. CR is currently being implemented in the upcoming SHERPA 3.

---

<sup>1</sup> Events involving  $W \rightarrow \tau\nu$  decays with a subsequent decay of the  $\tau$ -lepton to either  $e\nu e\nu\tau$  or  $\mu\nu\mu\nu\tau$  are included in the signal.

## 2.1 Colour reconnection models in PYTHIA 8

PYTHIA 8 [20] uses the so-called Lund string model for the hadronisation, which is based on the observation that at large distances the potential energy of colour sources increases linearly with their separation. As soon as the potential energy reaches values similar to hadron masses, the string breaks and creates a new quark–antiquark pair. The two string segments then begin to stretch and break again until all the energy has been converted into quark–antiquark pairs connected by short string segments, identified as hadrons.

Three general CR models are currently implemented in PYTHIA 8:

### 1. MPI-based model (CR0)

Starting from the lowest- $p_T$  interaction in a set of MPIs the reconnection probability for an MPI-system with transverse momentum  $p_T$  is given by:

$$P_{\text{rec}}(p_T) = \frac{(R_{\text{rec}} p_{T0})^2}{(R_{\text{rec}} p_{T0})^2 + p_T^2},$$

where  $R_{\text{rec}}$  is a phenomenological parameter with a value in the ‘colour reconnection range’, i.e.  $0 \leq R_{\text{rec}} \leq 10$ , and  $p_{T0}$  is an energy-dependent dampening parameter used for MPIs. The latter parameter regularises the partonic cross-section to avoid divergence at low  $p_T$  and it depends on the centre-of-mass energy,  $\sqrt{s} = E_{\text{CM}}$ :

$$p_{T0} = p_{T0}^{\text{ref}} \left( \frac{E_{\text{CM}}}{E_{\text{CM}}^{\text{ref}}} \right)^{E_{\text{CM}}^{\text{pow}}},$$

where  $p_{T0}^{\text{ref}}$  is the value of  $p_{T0}$  at a reference energy  $E_{\text{CM}}^{\text{ref}}$ , and  $E_{\text{CM}}^{\text{pow}}$  is a tunable parameter. For each MPI system that undergoes a reconnection, partons from lower- $p_T$  MPI systems are added to the colour dipoles defined by the higher- $p_T$  MPI system, in a way that minimises the total colour-string length, defined as the sum of the logarithm of normalised squared invariant masses of all two parton combinations in the event. This model is used in all ATLAS simulated samples generated by PYTHIA 8 with the A14 set of tuned parameters [21].

### 2. QCD-based model (CR1)

This model is based on the CR0 model but uses a more complete treatment of the QCD multiplet structure which allows reconnections of dipoles to produce structures with three (anti)colour indices, referred to as ‘junctions’, thereby enhancing the production of (anti)baryons. Only reconnections which lower the string length are performed.

### 3. Gluon-move model (CR2)

In the gluon-move model, reconnections are performed in the same way as in the CR0 model, but only gluons are considered for reconnection. For each gluon, all the reconnections to all MPI systems are considered (not only the ones for softer MPIs), so in principle the colour connections from the hard interaction can be affected more significantly than in the default model. Again, only reconnections which lower the string length are performed.

In all three models the top-quark lifetime is long enough to avoid colour reconnection of the top-quark decay products with the rest of the event. Only the top quark itself is involved in the CR process and the top-quark decays are performed after the colour reconnections have taken place. Thus, there is also no CR among the top-quark decay products. The first tuning of these models to ATLAS data is documented in Ref. [22] and a recent tuning to CMS data is presented in Ref. [23].

Additionally, a class of CR models that is designed to affect the top-quark decay products separately is considered. These models are described in detail in Ref. [11]. Two different sets of gluons are considered: gluons radiated from the top-quark decay products ( $\{g_t\}$ ) and gluons from the rest of the event ( $\{g_r\}$ ). Performing colour exchanges in different ways leads to five models:

1. **Forced random (TCR1)**

A gluon from  $\{g_t\}$  is forced to exchange colours with a random gluon from the other set,  $\{g_r\}$ .

2. **Forced nearest (TCR2)**

A gluon from  $\{g_t\}$  is forced to exchange colours with the gluon from  $\{g_r\}$  that minimises  $m^2(g_t, g_r)$ .

3. **Forced farthest (TCR3)**

A gluon from  $\{g_t\}$  is forced to exchange colours with the gluon from  $\{g_r\}$  that maximises  $m^2(g_t, g_r)$ .

4. **Forced smallest  $\Delta\lambda$  (TCR4)**

A gluon from  $\{g_t\}$  is forced to exchange colours with the gluon from  $\{g_r\}$  for which the change in  $\lambda$  (available rapidity range of particle production) is smallest.

5. **Smallest  $\Delta\lambda$  (TCR5)**

This is the same as the previous model, except that gluons exchange colours only if  $\Delta\lambda < 0$ .

For all of these models a phenomenological strength parameter  $s$  is used, where each gluon from the set  $\{g_t\}$  is tested for reconnection with a probability  $s$ .

## 2.2 Colour reconnection models in HERWIG 7

HERWIG 7 uses the cluster model [24] to model the hadronisation of quarks into hadrons. After the parton-shower calculation, gluons are split into quark–antiquark pairs, and a cluster is formed from each colour-connected pair of quarks. Before hadrons are produced from clusters, CR can modify the configuration of the clusters.

Three general CR models are currently implemented in HERWIG 7:

- **Plain CR**

In the plain CR model [25], quarks from two clusters can be rearranged into two alternative clusters. A cluster is randomly chosen from the list of clusters and compared with all other clusters on the list. The rearrangement is performed for the combination of clusters with the lowest sum of the masses. The reconnection is accepted with a probability  $p_R$ , which is the only parameter of this model. This model typically leads to clusters with smaller invariant mass and thus the overall activity in the UE is reduced. This model is used in all ATLAS samples generated by HERWIG 7, with different versions using different sets of tuned parameters (‘tunes’) for the UE and the CR. The H7-UE-MMHT tune is used in HERWIG 7.0.4, the H7.1-Default tune is used in HERWIG 7.1.3, and the H7.2-Default tune is used in HERWIG 7.2.1. Also, in the latest version the MPI model was improved [26].

- **Statistical CR**

The statistical CR model uses the Simulated Annealing algorithm [27] to find the configuration of clusters that results in the absolute lowest value of the colour length, defined as the sum of the squared invariant masses of all clusters in the event. In this model the only possible reconnections which are not allowed are those connecting quarks and antiquarks produced in the non-perturbative splitting of gluons, which would lead to the production of a colour-singlet object [28].

- **Baryonic CR**

In contrast to the other two models, the baryonic CR model [29] uses a simple geometric picture of nearest neighbours, and the algorithm tries to find quarks that populate approximately the same phase-space region, based on their rapidity  $y$ . Suitable combinations of clusters are then categorised as mesonic or baryonic, and reconnected with probabilities  $p_R$  or  $p_B$  respectively. This model is only available since HERWIG 7.1.5.

### 3 ATLAS detector

The ATLAS detector [19] at the LHC covers nearly the entire solid angle around the collision point.<sup>2</sup> It consists of an inner tracking detector surrounded by a thin superconducting solenoid, electromagnetic and hadron calorimeters, and a muon spectrometer incorporating three large superconducting air-core toroidal magnets.

The inner-detector system (ID) is immersed in a 2 T axial magnetic field and provides charged-particle tracking in the range  $|\eta| < 2.5$ . The high-granularity silicon pixel detector covers the vertex region and typically provides four measurements per track, the first hit normally being in the insertable B-layer (IBL) installed before Run 2 [30, 31]. It is followed by the silicon microstrip tracker (SCT), which usually provides eight measurements per track. These silicon detectors are complemented by the transition radiation tracker (TRT), which enables radially extended track reconstruction up to  $|\eta| = 2.0$ . The TRT also provides electron identification information based on the fraction of hits (typically 30 in total) above a higher energy-deposit threshold corresponding to transition radiation.

The calorimeter system covers the pseudorapidity range  $|\eta| < 4.9$ . Within the region  $|\eta| < 3.2$ , electromagnetic calorimetry is provided by barrel and endcap high-granularity lead/liquid-argon (LAr) calorimeters, with an additional thin LAr presampler covering  $|\eta| < 1.8$  to correct for energy loss in material upstream of the calorimeters. Hadron calorimetry is provided by the steel/scintillator-tile calorimeter, segmented into three barrel structures within  $|\eta| < 1.7$ , and two copper/LAr hadron endcap calorimeters. The solid angle coverage is completed with forward copper/LAr and tungsten/LAr calorimeter modules optimised for electromagnetic and hadronic energy measurements respectively.

The muon spectrometer (MS) comprises separate trigger and high-precision tracking chambers measuring the deflection of muons in a magnetic field generated by the superconducting air-core toroidal magnets. The field integral of the toroids ranges between 2.0 and 6.0 Tm across most of the detector. A set of precision chambers covers the region  $|\eta| < 2.7$  with three layers of monitored drift tubes, complemented by cathode-strip chambers in the forward region, where the background is highest. The muon trigger system covers the range  $|\eta| < 2.4$  with resistive-plate chambers in the barrel, and thin-gap chambers in the endcap regions.

Interesting events are selected by the first-level trigger system implemented in custom hardware, followed by selections made by algorithms implemented in software in the high-level trigger [32]. The first-level

---

<sup>2</sup> ATLAS uses a right-handed coordinate system with its origin at the nominal interaction point (IP) in the centre of the detector and the  $z$ -axis along the beam pipe. The  $x$ -axis points from the IP to the centre of the LHC ring, and the  $y$ -axis points upwards. Cylindrical coordinates  $(r, \phi)$  are used in the transverse plane,  $\phi$  being the azimuthal angle around the  $z$ -axis. The pseudorapidity is defined in terms of the polar angle  $\theta$  as  $\eta = -\ln \tan(\theta/2)$ . Angular distance is measured in units of  $\Delta R \equiv \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$ .



trigger accepts events from the 40 MHz bunch crossings at a rate below 100 kHz, which the high-level trigger further reduces in order to record events to disk at about 1 kHz.

An extensive software suite [33] is used in data simulation, in the reconstruction and analysis of real and simulated data, in detector operations, and in the trigger and data acquisition systems of the experiment.

## 4 Samples of data and simulated events

The analysis uses proton–proton ( $pp$ ) collision data which were recorded with the ATLAS detector during Run 2 of the LHC in the years from 2015 to 2018 at a centre-of-mass energy of 13 TeV. After applying data-quality requirements [34], the combined data set corresponds to an integrated luminosity of  $139 \text{ fb}^{-1}$  with an uncertainty of 1.7% [35]. Split by years, the integrated luminosities are  $3.22 \text{ fb}^{-1}$  for 2015,  $33.0 \text{ fb}^{-1}$  for 2016,  $44.3 \text{ fb}^{-1}$  for 2017 and  $58.5 \text{ fb}^{-1}$  for 2018. The integrated luminosity of the data set is derived following a methodology similar to that detailed in Ref. [36]. The LUCID-2 detector [37] is used for the primary luminosity measurements. At the high instantaneous luminosity reached at the LHC, there are additional pile-up collisions. The average number of interactions per bunch crossing increased from 13 to 36 from 2015 to 2018.

Events were selected online during data taking by single-electron or single-muon triggers [38, 39]. Multiple triggers were used to increase the selection efficiency. The lowest-threshold triggers utilised isolation requirements to reduce the trigger rate. The isolated-lepton triggers had  $p_T$  thresholds of 20 GeV for muons and 24 GeV for electrons in 2015 data, and 26 GeV for both lepton types in 2016, 2017 and 2018 data. They were complemented by other triggers with higher  $p_T$  thresholds, but no isolation requirements, to increase the trigger efficiency.

Large sets of events simulating  $t\bar{t}$  and background processes were produced with event generator programs based on the Monte Carlo (MC) method to model the recorded and selected data. After event generation, the response of the ATLAS detector was simulated using either the GEANT4 toolkit [40] with a full detector model or a fast simulation [41] which employs a parameterisation of the calorimeter response [42]. Samples using the fast simulation were employed to estimate systematic uncertainties associated with the event generators.

To account for pile-up effects, simulated minimum-bias interactions were overlaid on the hard-scattering events and the resulting events were weighted to reproduce the observed pile-up distribution. The minimum-bias events are simulated using PYTHIA 8.186 [6] with a set of tuned parameters called the A3 tune [43] and the NNPDF2.3<sub>LO</sub> PDF set [44]. To account for differences in the pile-up profile between individual years in a suitable way, simulated events representing the detector conditions in the years 2015 and 2016 were weighted to reproduce the distribution of the number of interactions averaged over each bunch crossing observed in that data-taking period, while simulated events modelling the detector conditions in the years 2017 and 2018 were weighted to reproduce the actual distribution of the number of interactions per bunch crossing observed in that time. Finally, the simulated events were reconstructed using the same software as applied to the collision data, the same event selection requirements were applied and the selected events were passed through the same analysis chain. More details about the samples of simulated events are provided in the following subsections.



## 4.1 Simulation of $t\bar{t}$ , $tW$ and $t\bar{t}+V$ production

Samples of events simulating  $t\bar{t}$  and  $tW$  production were generated with the POWHEG BOX v2 [45–51] next-to-leading-order (NLO) matrix-element generator, setting  $m_{\text{top}} = 172.5$  GeV. The NNPDF3.0<sub>NLO</sub> PDF set [52] implementing the five-flavour scheme was used. Parton showers, hadronisation and the underlying event were modelled using PYTHIA 8.230 with the A14 tune [21] and the NNPDF2.3<sub>LO</sub> PDF set. The POWHEG BOX + PYTHIA generator set-up implements a matching scheme for their modelling of hard emissions that avoids double counting them. The matrix-element-to-parton-shower matching is steered by the  $h_{\text{damp}}$  parameter, which controls the  $p_{\text{T}}$  of the first additional gluon emission beyond the LO Feynman diagram and therefore regulates the high- $p_{\text{T}}$  emission against which the  $t\bar{t}$  system recoils. Harder emissions in the parton shower are then vetoed. Event generation was run with  $h_{\text{damp}} = 1.5 \times m_{\text{top}}$  [53]. The renormalisation and factorisation scales were set dynamically on an event-by-event basis, namely to  $\mu_r = \mu_f = \sqrt{m_{\text{top}}^2 + p_{\text{T}}^2(t)}$  for  $t\bar{t}$  production and to  $\mu_r = \mu_f = m_{\text{top}}$  for  $tW$  production. Decays of bottom and charm hadrons were simulated using the EVTGEN 1.6.0 program [54]. When generating  $tW$  events, the diagram-removal scheme [55] was employed to handle the interference with  $t\bar{t}$  production [53].

The cross-section of  $t\bar{t}$  production was scaled to the next-to-next-to-leading-order (NNLO) predictions by the TOP++ 2.0 program (see Ref. [56] and references therein), which include the resummation of next-to-next-to-leading logarithmic (NNLL) soft-gluon terms. A value of  $\sigma(t\bar{t}) = 832_{-51}^{+47}$  pb was used, assuming  $m_{\text{top}} = 172.5$  GeV. The total cross-section for  $tW$  production was computed at NLO in QCD with the HATHOR 2.1 program [57, 58] and the simulated event sample was scaled to  $\sigma(tW + \bar{t}W) = 71.7 \pm 3.8$  pb.

The production of a  $t\bar{t}$  pair in association with a vector boson  $V$  was modelled at NLO using the MADGRAPH5\_AMC@NLO 2.3.3 generator [59] with the NNPDF3.0<sub>NLO</sub> set of PDFs. The events were interfaced to PYTHIA 8.210 for modelling of the parton showers, using the A14 tune and the NNPDF2.3<sub>LO</sub> PDF set. The decays of bottom and charm hadrons were simulated using EVTGEN 1.2.0. Cross-sections were calculated at NLO QCD and NLO EW accuracy using MADGRAPH5\_AMC@NLO as reported in Ref. [60]. The predicted values are  $0.88_{-0.11}^{+0.09}$  pb and  $0.60_{-0.07}^{+0.08}$  pb for  $t\bar{t}+Z$  and  $t\bar{t}+W$ , respectively, where the uncertainties were estimated from variations of  $\alpha_s$  and the renormalisation and factorisation scales.

## 4.2 Simulation of $Z$ +jets and diboson production

Backgrounds from  $Z$  boson production in association with jets, especially heavy-flavour jets, and on-shell diboson production ( $WW$ ,  $WZ$  and  $ZZ$ ) were simulated with the SHERPA 2.2.1 generator [8]. To simulate  $Z$ +jets (diboson) production, NLO-accurate matrix elements for up to two additional partons (one additional parton) and LO-accurate matrix elements for up to four (three) additional parton emissions were calculated with the COMIX [61] and OPENLOOPS 1 [62–64] libraries. The matrix elements used for diboson production contain all diagrams with four electroweak vertices. The default SHERPA parton shower [65], based on Catani–Seymour dipole factorisation, and the cluster hadronisation model [66] were used. The samples employ the dedicated set of tuned parameters developed by the SHERPA authors and the NNPDF3.0<sub>NLO</sub> PDF set. The NLO matrix elements of a given jet multiplicity were matched to the parton shower using a colour-exact variant of the MC@NLO algorithm [67]. Different jet multiplicities were then merged into an inclusive sample using an improved CKKW matching procedure [68, 69] which was extended to NLO accuracy using the MEPS@NLO prescription [70]. The merging threshold was set to 20 GeV.

The  $Z$ +jets sample was normalised to an NNLO prediction [71] of the total cross-section, obtained with the FEWZ package [72]. Considering the dilepton event selection used in the analysis, the simulation

efficiency of the  $Z$ +jets samples was increased by forcing the  $Z$  bosons to decay into a pair of charged leptons ( $e^+e^-$ ,  $\mu^+\mu^-$  or  $\tau^+\tau^-$ ). In diboson events, at least one of the two bosons was required to decay leptonically. The simulated diboson event samples were normalised to the total cross-sections provided by SHERPA at NLO in QCD.

## 5 Object reconstruction and definitions of observables

Event candidates are identified by means of isolated electrons and muons, and jets, some of which are  $b$ -tagged. In the following, the definitions of these reconstructed objects at detector level, together with the corresponding objects reconstructed at particle level, as well as the event selection are described. In addition the analysis observables are introduced.

### 5.1 Detector-level object definitions

Vertices are reconstructed from at least two ID tracks with  $p_T > 500$  MeV. The primary vertex of an event is defined as the vertex with the highest sum of  $p_T^2$  of all associated ID tracks [73].

Tracks are reconstructed from the measured positions of ‘hits’ in the ID caused by the passage of charged particles [74]. The track reconstruction consists of an iterative track-finding algorithm seeded by combinations of at least three silicon-detector hits followed by a combinatorial Kalman filter [75] to build track candidates based on hits compatible with the extrapolated trajectory. Ambiguities between the track candidates are then resolved by applying quality criteria to suppress combinations of hits unlikely to originate from a single charged particle. Finally, a high-resolution track fit is performed, resulting in more precise track reconstruction parameters. Track candidates are required to pass the *Tight Primary* selection [76]. In particular, they must have  $p_T > 500$  MeV and be within  $|\eta| < 2.5$ . At least one hit in the two innermost layers is required if the extrapolated track crosses the sensitive region of an active sensor module. The number of hits in the pixel and SCT detectors must be larger than 9 for  $|\eta| \leq 1.65$  or larger than 11 for  $|\eta| > 1.65$ , with no more than 2 missing SCT hits on a track if the respective SCT modules are operational. Additionally, track-to-vertex-matching criteria are applied to mitigate the effect of pile-up tracks and improve the rejection of tracks from secondary interactions: the transverse impact parameter calculated relative to the beam line must satisfy  $|d_0| < 0.5$  mm, and the longitudinal impact parameter calculated relative to the event primary vertex must satisfy  $|z_0 \sin(\theta)| < 0.5$  mm, where  $\theta$  is the polar angle of the track.

Electron candidates are reconstructed from clusters of energy deposited in the electromagnetic calorimeter with a matched track reconstructed in the ID [77]. The pseudorapidity of clusters,  $\eta_{\text{cluster}}$ , is required to be in the range  $|\eta_{\text{cluster}}| < 2.47$ . However, clusters are excluded if they are in the transition region  $1.37 < |\eta_{\text{cluster}}| < 1.52$  between the barrel and endcap electromagnetic calorimeters. A likelihood-based method is used to simultaneously evaluate several properties of electron candidates, including shower shapes in the electromagnetic calorimeter, track quality, and transition radiation produced in the TRT. Electron candidates must have  $p_T > 25$  GeV and satisfy the *Tight* identification criteria as defined in Ref. [77]. Electrons are required to be isolated from other activity by applying the *Tight* requirements on the sum of nearby energy in the calorimeter and the sum of the momenta of nearby tracks [77].

Muon candidates are reconstructed by combining tracks in the MS with tracks in the ID [78, 79] and must have  $|\eta| < 2.5$ ,  $p_T > 25$  GeV, satisfy the *Medium* identification criteria, and pass the *Tight* isolation requirements as defined in Ref. [79].

The tracks associated with electron and muon candidates must satisfy slightly different requirements on  $d_0$ , i.e.  $d_0$  significance requirements of  $|d_0/\sigma_{d_0}| < 5.0$  for electrons and  $|d_0/\sigma_{d_0}| < 3.0$  for muons are imposed. The lepton energy and momentum scales and resolutions are calibrated in data by using  $Z \rightarrow ee$  [77] or  $Z \rightarrow \mu\mu$  events and inspecting the dilepton mass spectrum near the  $Z$ -boson peak.

Jets are reconstructed from particle-flow objects [80] with the anti- $k_t$  clustering algorithm [81, 82] using a radius parameter of 0.4. Their energy is calibrated [83], and they must fulfil  $p_T > 25$  GeV and  $|\eta| < 2.5$ . Jets with  $p_T < 60$  GeV and  $|\eta| < 2.5$  are required to pass a requirement on the jet-vertex-tagger (JVT) discriminant [84] to suppress jets originating from pile-up collisions. The JVT-discriminant is required to be above 0.5, which corresponds to an efficiency of 92% for non-pile-up jets. Jets containing  $b$ -hadrons are identified ( $b$ -tagged) with the DL1r algorithm [85], which uses a deep feed-forward neural network with several  $b$ -tagging algorithms as input [86]. The algorithms exploit the impact parameters of charged-particle tracks, parameters of reconstructed secondary vertices and the topology of  $b$ - and  $c$ -hadron decays inside the jets. In order to strongly reduce the misidentification rate of  $c$ -jets and light-flavour ( $u$ ,  $d$  or  $s$ ) jets, a specific working point of the DL1r algorithm was defined and calibrated, using Run 2 data and the standard calibration technique, i.e. a likelihood-based method in a sample highly enriched in  $t\bar{t}$  events [85]. The  $b$ -tagging efficiency for jets that originate from the hadronisation of  $b$ -quarks is 70% in simulated  $t\bar{t}$  events.

To avoid double-counting objects satisfying more than one set of selection criteria, a procedure called *overlap removal* is applied. Reconstructed objects are removed in the following order: electrons sharing an ID track with a muon; jets within  $\Delta R = 0.2$  of an electron, thereby avoiding double-counting electron energy deposits as jets; electrons within  $\Delta R = 0.4$  of a remaining jet, to reduce the impact of non-prompt electrons; jets within  $\Delta R = 0.2$  of a muon if they have less than three associated tracks; muons within  $\Delta R = 0.4$  of a remaining jet, reducing the rate of non-prompt muons.

In order to build observables in this analysis, tracks within  $\Delta R = 0.4$  of a jet or within  $\Delta R = 0.01$  of the selected electron or muon are discarded.

## 5.2 Particle-level object definitions

Particle-level objects are defined in simulated events by using only stable particles with a mean lifetime  $\tau > 30$  ps, using the following criteria.

Charged particles are required to have  $p_T > 500$  MeV and  $|\eta| < 2.5$ . Charged particles that are generated through interactions with matter, referred to as secondary particles, are excluded.

Electrons and muons are required not to originate from a generated hadron in the MC event, either directly or through a  $\tau$ -lepton decay. This ensures that they originate from a  $W$ -boson decay without requiring a direct match. The four-momenta of the bare leptons are modified (dressed) by adding the four-momenta of prompt photons within  $\Delta R = 0.1$ . The dressed leptons must then have  $p_T > 25$  GeV and  $|\eta| < 2.5$ .

Jets are clustered using all stable particles, excluding those used in the definition of dressed leptons and prompt neutrinos, using the anti- $k_t$  algorithm with a radius parameter of 0.4. Jets are required to

have  $p_T > 25$  GeV and  $|\eta| < 2.5$  and are identified as  $b$ -jets via ghost-matching [87] to a  $b$ -hadron with  $p_T > 5$  GeV.

To avoid double-counting, particle-level objects are subject to overlap removal criteria. Dressed muons and electrons with a separation  $\Delta R < 0.4$  from a jet are excluded.

In order to obtain a track collection outside of jets, charged particles within  $\Delta R = 0.4$  of a jet or within  $\Delta R = 0.01$  of the electron or muon from the top-quark decay are excluded.

### 5.3 Event selection

Events are required to have at least one primary vertex, exactly one electron and one muon with  $p_T > 25$  GeV for the 2015 data-taking period and  $p_T > 27$  GeV for data taken from 2016 to 2018 to match the trigger thresholds. Furthermore, the events are required to contain two or three jets with  $p_T > 25$  GeV. Exactly two of the jets must be  $b$ -tagged jets. The third jet is mainly accepted in order to increase the statistics of the selected sample while keeping the modelling uncertainties manageable. Hadronic resonances are suppressed by requiring the dilepton invariant mass to satisfy  $m_{\ell\ell} > 15$  GeV. Finally, events with track multiplicities  $n_{\text{trk,out}}$  (defined below) above 100 are excluded because the dominant contribution is due to pile-up tracks.<sup>3</sup>

Events with an opposite-sign (OS)  $e\mu$  pair define the nominal sample, whilst events with a same-sign (SS)  $e\mu$  pair are used to estimate the background from fake leptons. The particle-level events are required to pass the same selection requirements as the data selection but applied to the particle-level objects. In the case of the lepton  $p_T$  the lower requirement of 25 GeV is applied.

### 5.4 Definition of the observables at particle level and detector level

#### Particle-level observables

Particle-level observables are built from charged particles, as defined in Section 5.2. The three observables are defined as:

- $n_{\text{ch}}$ : Charged-particle multiplicity with the binning defined as [0, 10, 20, 30, 40, 50, 60, 80, 100].
- $\sum_{n_{\text{ch}}} p_T$ : Scalar sum of the transverse momenta of the selected charged particles with the binning defined as [0, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100,  $\infty$ ) in units of GeV. This observable is highly correlated with the charged-particle multiplicity, but contains additional information about the kinematic configuration of the event.
- $\sum_{n_{\text{ch}}} p_T$  in bins of  $n_{\text{ch}}$ : This two-dimensional observable is built from the above two variables, as the  $\sum_{n_{\text{ch}}} p_T$  in bins of  $n_{\text{ch}}$  with the binning defined in Table 1.

The binning of the observables was chosen according to a predefined procedure described in Section 7.

---

<sup>3</sup> The fraction of selected data events with track multiplicities above 100 is below 0.2%.

Table 1: Binning used for the measured  $\sum_{n_{\text{ch}}} p_{\text{T}}$  in bins of  $n_{\text{ch}}$  observable.

$n_{\text{ch}}$ bin	$\sum_{n_{\text{ch}}} p_{\text{T}}$ binning [GeV]
[0, 20)	[0, 10, $\infty$ )
[20, 40)	[0, 35, $\infty$ )
[40, 60)	[0, 60, $\infty$ )
[60, 80)	[0, 80, $\infty$ )
[80, 100]	[0, $\infty$ )

### Detector-level observables

At detector level, the observables are calculated using reconstructed tracks with the same binning as at particle level. The corresponding variables are then defined as:

- $n_{\text{trk,out}}$ : Number of selected tracks outside of jets.
- $\sum_{n_{\text{trk,out}}} p_{\text{T}}$ : Scalar sum of the transverse momenta of these tracks.
- $\sum_{n_{\text{trk,out}}} p_{\text{T}}$  in bins of  $n_{\text{trk,out}}$ : The  $\sum_{n_{\text{trk,out}}} p_{\text{T}}$  in bins of  $n_{\text{trk,out}}$ .

Before these variables are used in the unfolding, track-based backgrounds are subtracted and further corrections are applied, as explained in Section 6.2.

## 6 Background estimation and corrections to observables

Two different sources can lead to background contamination in the measured observables: event-based contributions from background processes with a final state similar to  $t\bar{t}$  dileptonic events, discussed in Section 6.1, and track-based contributions originating from sources other than the hard-scattering process, discussed in Section 6.2. Since the track-based background has to be subtracted from the measured variables  $n_{\text{trk,out}}$  and  $\sum_{n_{\text{trk,out}}} p_{\text{T}}$  the following procedure is applied:

- The pile-up and secondary-particle contribution per event is estimated in a stochastic way. This procedure is detailed in Section 6.2.
- A tracking-efficiency correction (TEC) is applied to the background-subtracted observables to ease the unfolding. This procedure is detailed in Section 6.3.

### 6.1 Event-based backgrounds

After event selection, several background processes with a final state similar to  $t\bar{t}$  dileptonic events pass the selection. Sources that contain top quarks contribute to the background, with  $tW$  events containing a single top quark being the dominant contribution. The production of  $t\bar{t}+V$  events, with  $V$  being either a  $W$  or a  $Z$  boson, contributes an irreducible background. Events that contain either two electroweak bosons or a  $Z$  boson produced in association with jets,  $Z(\rightarrow \tau\tau)+\text{jets}$ , can be misidentified as a  $t\bar{t}$  dilepton event. Finally, there is a contribution from multijet processes when at least one of the reconstructed leptons is wrongly reconstructed as an isolated lepton, arising from either a heavy-flavour hadron decay, an electron from a photon conversion, a jet misidentified as an electron, or a muon produced from decay-in-flight of a pion or kaon. This last category is collectively called the fake-lepton background.

The fake-lepton background contribution is estimated using a partially data-driven approach. Events with an enriched fake-lepton contribution are obtained from the observed same-sign  $e\mu$  sample. The total number of fake-lepton background events,  $N^{\text{fake}}$ , is given by:

$$N^{\text{fake}} = R \cdot \left( N^{\text{data,SS}} - N^{\text{prompt,SS}} \right),$$

$$R = \frac{N^{\text{fake,OS}}}{N^{\text{fake,SS}}},$$

where  $N^{\text{data,SS}}$  is the number of same-sign events in data,  $N^{\text{prompt,SS}}$  is the estimated number of prompt same-sign events from simulation, and  $R$  is the ratio of opposite-sign ( $N^{\text{fake,OS}}$ ) to same-sign ( $N^{\text{fake,SS}}$ ) events with fake leptons in simulation, with  $R$  being  $2.4 \pm 0.1$ . The kinematic distributions of the fake-lepton background are determined from the simulated  $t\bar{t}$  events with single-lepton and dilepton final states.

The composition of the same-sign samples is illustrated in Figure 1. Electron and muon  $p_T$  distributions obtained from same-sign selected data events are shown together with the corresponding predictions from simulated events. The contributions are broken down into prompt leptons and various fake-lepton categories. The prompt contributions are about 35% of the same-sign sample. They include wrong-sign (WS) contributions, dominated by  $t\bar{t}$  dilepton events where the electron charge sign has been wrongly reconstructed, and right-sign (RS) contributions, with same-sign prompt leptons, from  $t\bar{t}+V$  events,  $Z$ +jets and diboson production. The fake-lepton contributions are dominated by electrons from photon conversions, which represent 62% of the same-sign sample. Sub-leading contributions come from leptons produced in semileptonic decays of heavy-flavour hadrons, and other sources, such as misidentified hadrons or decays-in-flight of pions and kaons. A conservative 50% uncertainty is assigned to the fake-lepton background [88].

All other background processes are modelled by simulated events and the expected number of events is calculated using the theoretical cross-sections and the acceptance from the simulated samples scaled to the integrated luminosity. The expected and the observed event yields are shown in Table 2. The estimated purity of  $t\bar{t}$  dilepton events is approximately 96%, with the backgrounds from single top quarks being the largest impurities. Observed distributions of the leading jet  $p_T$  and the jet multiplicity compared with the model obtained from simulated events are shown in Figure 2. In the leading jet  $p_T$  distribution, a slope can be seen in the ratio of data to prediction. This is in agreement with previous differential cross-section measurements by ATLAS [88–91], and CMS [92–95], which have consistently observed a softer top-quark  $p_T$  spectrum (and related distributions) than what is predicted by NLO+PS MC generators. This discrepancy is at least partially due to missing NNLO corrections [96–100]. The difference between data and prediction is covered by the systematic uncertainties.

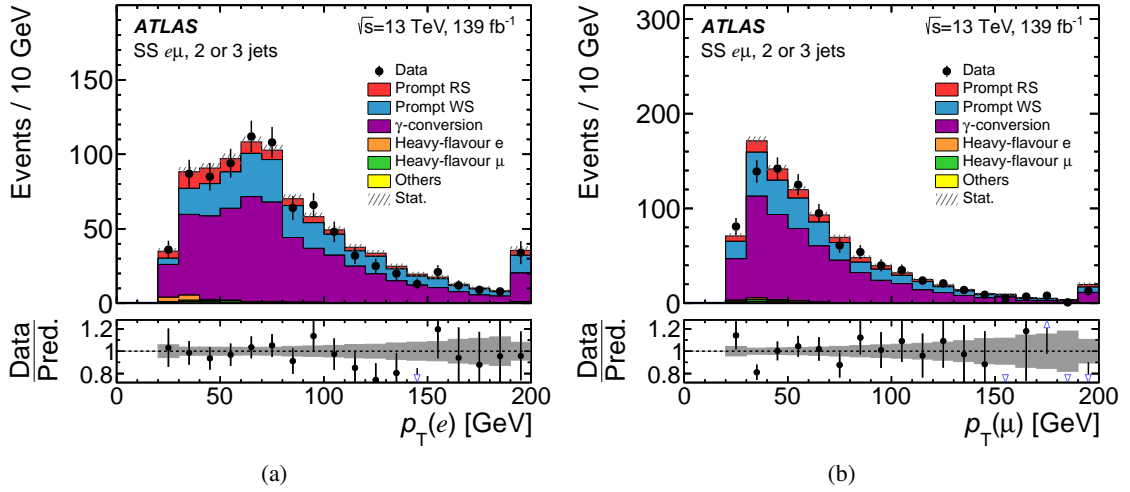


Figure 1: Distributions of (a) the electron  $p_T$  and (b) the muon  $p_T$ , in events with a same-sign  $e\mu$  pair and two or three jets, of which two are  $b$ -tagged jets. The simulation's prediction is normalised to the number of expected events and broken down into contributions where both leptons are prompt RS or WS, or one is a fake lepton from a photon conversion originating from a top-quark decay or from background, from heavy-flavour decay or from other sources. The lower panel shows the ratio of data to the prediction in each bin. The hatched (grey) uncertainty band represents the statistical uncertainty due to limited size of the simulated samples. Events beyond the  $x$ -axis range are included in the last bin. The blue triangular markers in the lower panels point to data-to-prediction ratio values which lie beyond the shown  $y$ -axis range.

Table 2: Event yields obtained after the event selection. The expected event yields from  $t\bar{t}$  production and the various background processes are compared with the observed event yield. The fractional contributions from  $t\bar{t}$  production and the background processes to the expected event yield is given in %. The processes labelled by 'Others' include production of  $Z$ +jets and diboson background events. The uncertainties include the MC statistical uncertainty and the normalisation uncertainty.

Process	Events	Fraction [%]
$t\bar{t}$	$190\,600 \pm 9\,500$	96
Single top	$6\,310 \pm 640$	3.2
Fake leptons	$1\,320 \pm 660$	0.66
$t\bar{t}+V$	$213 \pm 26$	0.11
Others	$91 \pm 21$	0.046
Total	$199\,000 \pm 11\,000$	
Data	195 507	



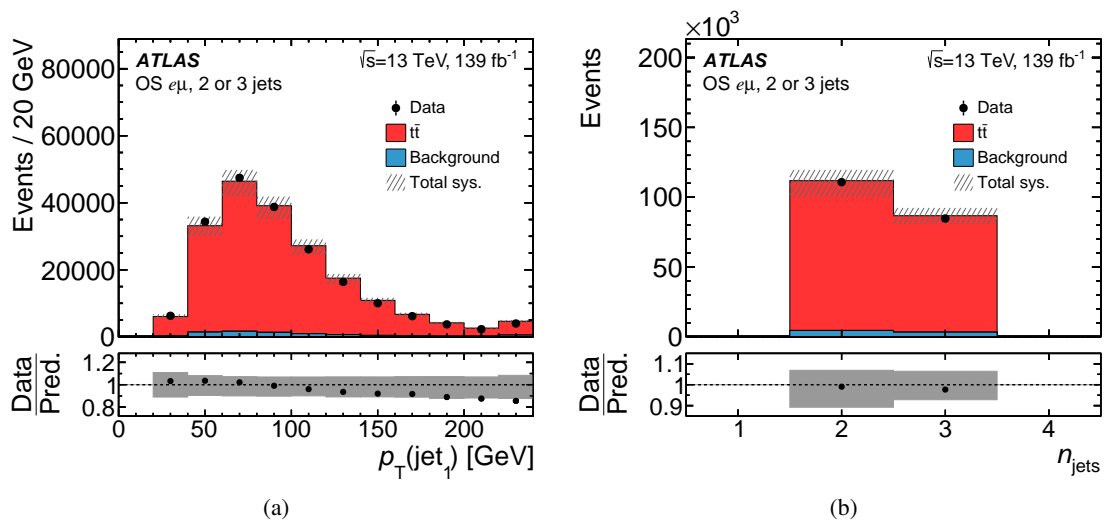


Figure 2: Distributions of (a) the leading jet  $p_T$  and (b) the jet multiplicity in events with an opposite-sign  $e\mu$  pair and two or three jets, of which two are  $b$ -tagged jets. The observed data is compared with the expectation from simulated  $t\bar{t}$  and background events, where the background includes contributions from  $tW$ ,  $t\bar{t}+V$ ,  $Z$ +jets and diboson processes and the estimated fake-lepton background. The lower panel shows the ratio of data to the prediction in each bin. The hatched (grey) uncertainty band includes the MC statistical uncertainty, background normalisation uncertainties, detector systematic uncertainties and  $t\bar{t}$  modelling systematic uncertainties. Events beyond the  $x$ -axis range are included in the last bin.

## 6.2 Track-based backgrounds

Selected tracks other than the hard-scatter primary tracks can arise from additional sources. Tracks originating from additional  $pp$  collisions, referred to as pile-up, are the dominant background source. In addition, secondary particles produced through interactions with matter rather than as a result of direct fragmentation processes form a small contribution, while secondary particles produced in heavy-flavour decays are negligible, since tracks within jets are excluded. Finally, tracks that are formed by a random combination of hits or from a combination of hits from several particles, referred to as fake tracks, are considered. Corrected quantities  $n_{\text{trk,prim}}$  and  $\sum_{n_{\text{trk,prim}}} p_T$  are obtained from the measured observables  $n_{\text{trk,out}}$  and  $\sum_{n_{\text{trk,out}}} p_T$  at detector level by subtracting these contributions on an event-by-event basis in the following way:

$$n_{\text{trk,prim}} = n_{\text{trk,out}} - c_{\text{PU}}(\mu, n_{\text{trk,out}}) \cdot n_{\text{PU}} - c_{\text{sec}} \cdot n_{\text{sec}}, \quad (1)$$

$$\sum_{n_{\text{trk,prim}}} p_T = \sum_{n_{\text{trk,out}}} p_T - \sum_{i=1}^{c_{\text{PU}} \cdot n_{\text{PU}}} p_T^{\text{corr}} - \sum_{i=1}^{c_{\text{sec}} \cdot n_{\text{sec}}} p_T, \quad (2)$$

where  $\mu$  is the number of interactions per bunch crossing,  $n_{\text{PU}}$  is the number of pile-up tracks,  $n_{\text{sec}}$  is the number of secondary-particle tracks including the contribution from fake tracks, and  $c_{\text{PU}}$ , binned in  $\mu$  and  $n_{\text{trk,out}}$ , and  $c_{\text{sec}}$  are scale factors to correct for differences between simulation and observed data. The corrected  $\sum_{n_{\text{trk,prim}}} p_T$  in bins of the  $n_{\text{trk,prim}}$  observable is built from Eqs. (1) and (2), and thus no additional background estimation is needed.

The pile-up and secondary-particle contamination ( $n_{\text{PU}}$  and  $n_{\text{sec}}$ ) per event is subtracted in the measured observables in a stochastic way based on a method using MC templates. The determination of the two contributions is based on the assumption that they are uncorrelated, which was also checked to be true in simulation. The procedure is performed as follows:

- $n_{\text{trk,prim}}$ : Templates are created from simulated  $t\bar{t}$  events for both the pile-up and secondary-particle track multiplicities, in bins of  $n_{\text{trk,out}}$ . The background contamination in  $n_{\text{trk,out}}$  is expected to be higher for higher track multiplicities, since more tracks are likely to be selected from additional pile-up vertices overlapping with the primary vertex. For each event a random number is drawn from each template, corresponding to the  $n_{\text{trk,out}}$  value of the event. In this way the contamination is subtracted in the measured observables in a stochastic way.
- $\sum_{n_{\text{trk,prim}}} p_T$ : A procedure similar to that for the  $n_{\text{trk,prim}}$  observable is used, where templates are created from simulated  $t\bar{t}$  events for pile-up and secondary-particle track  $p_T$ . In order to compensate for deficiencies in the modelling of pile-up in simulation a reweighting method is used. To obtain a corrected  $p_T$  distribution ( $p_T^{\text{corr}}$ ), tracks matched to pile-up vertices, in contrast to nominal tracks matched to the primary vertex, are compared between simulation and observed data and their ratio is used as a reweighting function. This function is used to reweight the  $p_T$  distribution obtained from pile-up tracks matched to the primary vertex. Since the contribution of tracks from secondary particles is small, no correction is applied in this case. These templates, together with the previously estimated  $n_{\text{PU}}$  ( $n_{\text{sec}}$ ), are then used to estimate the pile-up (secondary-particle) track contamination in the measured  $\sum_{n_{\text{trk,out}}} p_T$  observable, by drawing random numbers  $n_{\text{PU}}$  ( $n_{\text{sec}}$ ) times from the corresponding track  $p_T$  template for each event.

In order to validate the procedure a closure test is performed, comparing the obtained distributions with distributions built from true pile-up tracks in MC events. Differences between the expected and observed

rates of pile-up tracks and secondary-particle tracks are accounted for by the scale factors  $c_{\text{PU}}$  and  $c_{\text{sec}}$ . Templates of the  $z_0$  and  $d_0/\sigma_{d_0}$  distributions are created from simulated  $t\bar{t}$  events for hard-scatter primary tracks, pile-up tracks and secondary-particle tracks. The distributions for pile-up tracks are corrected for differences between collision data and simulation using high-purity pile-up tracks originating from pile-up vertices. Two separate binned maximum-likelihood fits are performed to estimate  $c_{\text{PU}}$  and  $c_{\text{sec}}$ . Studies have shown that these two contribution can be determined independently in the following way: to obtain  $c_{\text{PU}}$ , hard-scatter primary-track and secondary-particle-track templates of the  $z_0$  distribution are used, while to obtain  $c_{\text{sec}}$ , hard-scatter primary-track and pile-up-track templates of the  $d_0/\sigma_{d_0}$  distribution are used. The scale factors of each contribution in the two fits are kept free-floating. Studies have shown that  $c_{\text{PU}}$  depends on  $\mu$  and the track multiplicity. Therefore,  $c_{\text{PU}}$  is determined for different intervals of  $\mu$  and  $n_{\text{trk,out}}$ . Examples of fitted  $z_0$  and  $d_0/\sigma_{d_0}$  distributions are shown in Figure 3 and a summary of the estimated  $c_{\text{PU}}(\mu, n_{\text{trk,out}})$  is shown in Table 3. The obtained scale factors range from 0.9 to 1.4, where higher values are obtained for higher  $\mu$  and  $n_{\text{trk,out}}$  values. A value of  $2.34 \pm 0.02$  is estimated for  $c_{\text{sec}}$ . The estimation of systematic uncertainties of  $c_{\text{PU}}$  is described in Section 8.2.

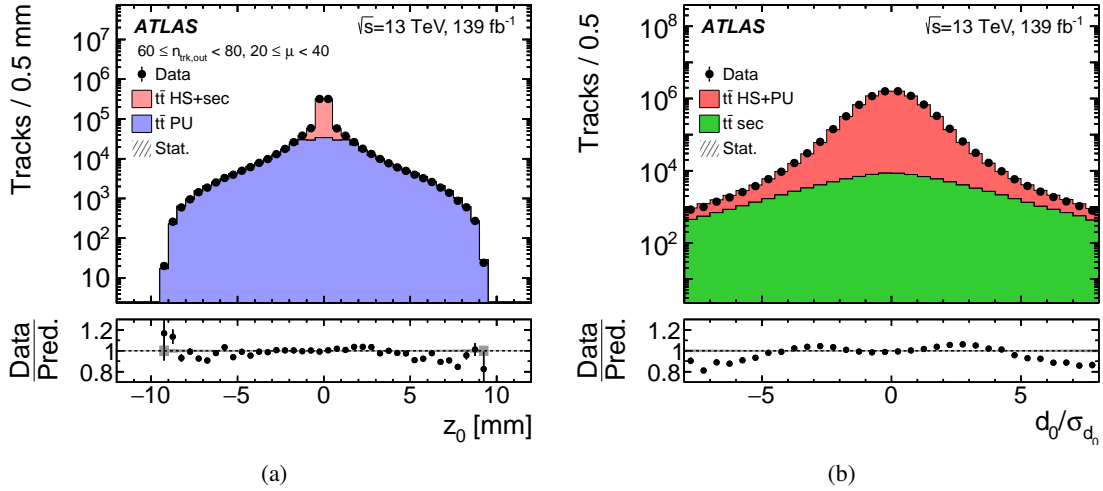


Figure 3: Illustration of the estimation of the pile-up (secondary-particle) track background by fitting the  $z_0$  ( $d_0/\sigma_{d_0}$ ) distribution. As representative examples, the  $z_0$  distribution is shown in the region with  $20 \leq \mu < 40$  and  $60 \leq n_{\text{trk,out}} < 80$  in (a) and the fitted  $d_0/\sigma_{d_0}$  distribution is shown in (b). The stacked histograms are normalised to the fit result. The lower panel shows the ratio of data to the prediction in each bin. The hatched (grey) uncertainty band represents the statistical uncertainty due to the limited size of the simulated samples.

Table 3: Summary of the estimated pile-up scale factors  $c_{\text{PU}}$ , parameterised in  $\mu$  and  $n_{\text{trk,out}}$ . All values have a statistical precision of 0.01.

Region	$n_{\text{trk,out}} < 20$	$20 \leq n_{\text{trk,out}} < 40$	$40 \leq n_{\text{trk,out}} < 60$	$60 \leq n_{\text{trk,out}} < 80$	$80 \leq n_{\text{trk,out}} \leq 100$
$\mu < 20$	0.91	1.04	0.97	1.05	1.08
$20 \leq \mu < 40$	0.91	1.08	1.08	1.07	1.11
$\mu \geq 40$	0.95	1.15	1.23	1.27	1.36

### 6.3 Corrections to observables

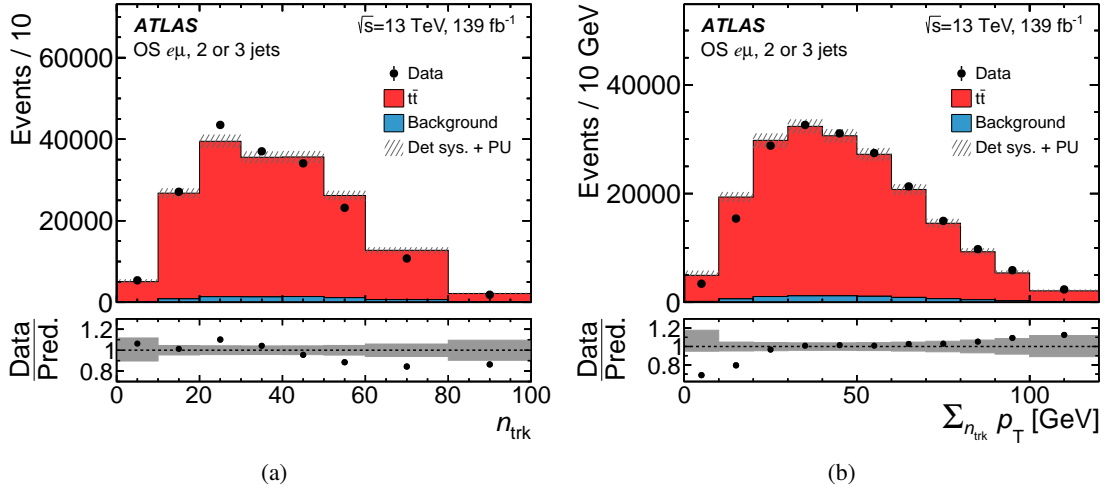
To place the leading elements of the migration matrix on the principal diagonal, the background-subtracted quantities,  $n_{\text{trk,prim}}$  and  $\sum_{n_{\text{trk,prim}}} p_{\text{T}}$ , are scaled with correction factors  $C_{\text{ntrk}}$  and  $C_{\text{sum-pt}}$ , respectively, accounting for the limited track-reconstruction efficiency:

$$n_{\text{trk}} = \frac{n_{\text{trk,prim}}}{C_{\text{ntrk}}(n_{\text{trk,prim}})} \quad \sum_{n_{\text{trk}}} p_{\text{T}} = \frac{\sum_{n_{\text{trk,prim}}} p_{\text{T}}}{C_{\text{sum-pt}}(\sum_{n_{\text{trk,prim}}} p_{\text{T}})} \quad (3)$$

The TEC factors are calculated in bins of the background-subtracted quantity by averaging the ratio of the quantity at the background-subtracted level to its value at particle level across all selected events of the nominal simulated  $t\bar{t}$  sample. The function  $C_{\text{ntrk}}(n_{\text{trk,prim}})$  has values between 0.68 and 0.89 in its range from 0 to 100, while the function  $C_{\text{sum-pt}}(\sum_{n_{\text{trk,prim}}} p_{\text{T}})$  rises stepwise from 0.66 to 0.89 across the bins of its full range. The corrections are applied to the observables in measured data events, in simulated events from all scattering processes (i.e. the  $t\bar{t}$  process and all background processes), and in simulated events of the samples with systematically varied parameters. The corrected quantities,  $n_{\text{trk}}$  and  $\sum_{n_{\text{trk}}} p_{\text{T}}$ , are the final corrected-level observables forming the distributions to be unfolded, and are shown in Figure 4. The corrected  $\sum_{n_{\text{trk}}} p_{\text{T}}$  in bins of the  $n_{\text{trk}}$  observable, shown in Figure 4(c), is built from Eq. (3) without any additional correction. The uncertainty band includes detector systematic uncertainties and the pile-up background-track-rate uncertainty. Within these experimental uncertainties there are considerable differences between the observed data and the nominal prediction, reflecting non-optimal tuning of the UE and MPI. Table 4 summarises the naming convention for the observables at different levels of the analysis.

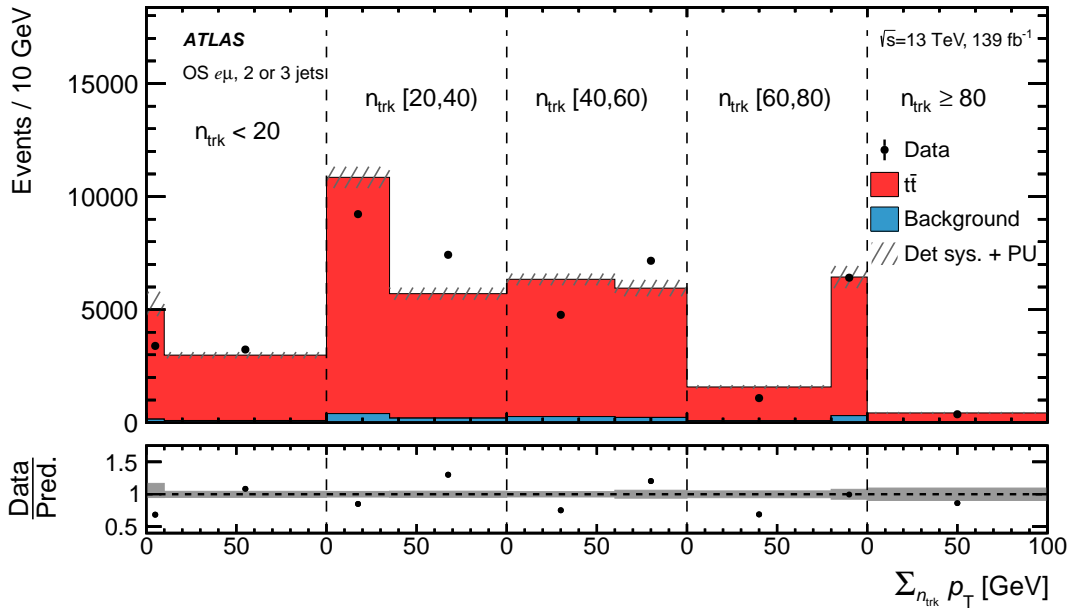
Table 4: Naming convention for the observables at different levels of the analysis. At the background-subtracted level the contributions of tracks from pile-up collisions and tracks from secondary vertices are subtracted. At the corrected level the TEC is applied. The observables at particle level are the analysis results.

Analysis level	Number of charged particles outside of jets	Scalar sum of track $p_{\text{T}}$
Detector level	$n_{\text{trk,out}}$	$\sum_{n_{\text{trk,out}}} p_{\text{T}}$
Background-subtracted level	$n_{\text{trk,prim}}$	$\sum_{n_{\text{trk,prim}}} p_{\text{T}}$
Corrected level	$n_{\text{trk}}$	$\sum_{n_{\text{trk}}} p_{\text{T}}$
Particle level	$n_{\text{ch}}$	$\sum_{n_{\text{ch}}} p_{\text{T}}$



(a)

(b)



(c)

Figure 4: Distributions of (a)  $n_{\text{trk}}$ , (b)  $\sum n_{\text{trk}} p_T$  and (c)  $\sum n_{\text{trk}} p_T$  in bins of  $n_{\text{trk}}$  for events with an opposite-sign  $e\mu$  pair and two or three jets, of which two are  $b$ -tagged jets. The corrected-level observed distributions are compared with the expectation from simulated events, broken down into contributions from  $t\bar{t}$  and the background processes  $tW$ ,  $t\bar{t}+V$ ,  $Z$ +jets and diboson, together with the estimated fake-lepton background. The estimated contributions from pile-up tracks and secondary particles is subtracted from measured data events and simulated  $t\bar{t}$  events. The TEC is applied to all processes. The lower panel shows the ratio of observed data to the prediction in each bin. The hatched (grey) uncertainty band includes detector systematic uncertainties and the pile-up background-track-rate uncertainty. Events beyond the  $x$ -axis range are included in the last bin.

## 7 Unfolding

The differential cross-section distributions in the fiducial phase space defined in Section 5.3 are obtained from the corrected-level events using an unfolding technique that corrects for acceptance and certain detector effects. The iterative Bayesian method [101] as implemented in RooUnfold [102] is used for this unfolding.

The unfolding procedure starts from the observables described in the previous section, i.e. after subtracting the track-based backgrounds and applying the TEC, and consists of two stages. First, the event-based background contributions  $N_{\text{bkg}}$ , summarised in Table 2, are subtracted bin-by-bin from the track-based background-subtracted and tracking-efficiency-corrected distribution  $N_{\text{corr}}$ . The contributing processes are normalised either to the corresponding theory predictions or, in the case of the fake-lepton background, to the partially data-driven estimate. Second, the distributions are unfolded using a detector response model, the migration matrix  $\mathcal{M}$ , obtained from simulated  $t\bar{t}$  events. As part of this second step, two correction factors are applied to correct for non-overlap of the fiducial phase space at corrected level and particle level. The corrections account for events that fall within the fiducial phase space of one level but not the other. The extraction of the absolute differential cross-section for an observable  $X$ ,  $X$  being  $n_{\text{trk}}$ ,  $\sum_{n_{\text{trk}}} p_{\text{T}}$  or  $\sum_{n_{\text{trk}}} p_{\text{T}}$  in bins of  $n_{\text{trk}}$ , is then summarised by the equation:

$$\frac{d\sigma^{\text{fid}}}{dX^i} \equiv \frac{1}{\mathcal{L} \cdot \Delta X^i} \cdot \frac{1}{\epsilon^i} \cdot \sum_j \mathcal{M}_{ij}^{-1} \cdot f_{\text{acc}}^j \cdot (N_{\text{corr}}^j - N_{\text{bkg}}^j)$$

where the index  $i$  ( $j$ ) iterates over the bins of observable  $X$  at particle (corrected) level,  $\Delta X^i$  is the bin width,  $\mathcal{L}$  is the integrated luminosity, the inversion of  $\mathcal{M}$  symbolises unfolding using the iterative Bayesian method, and  $f_{\text{acc}}$  and  $\epsilon$  are the phase-space correction factors. These last two parameters are defined as:

$$f_{\text{acc}} = \left( \frac{N_{\text{corr}\wedge\text{part}}}{N_{\text{corr}}} \right) \quad \epsilon = \left( \frac{N_{\text{corr}\wedge\text{part}}}{N_{\text{part}}} \right)$$

The number  $N_{\text{corr}\wedge\text{part}}$  indicates the number of corrected-level events that satisfy the particle-level selection and  $N_{\text{part}}$  is the number of events passing the particle-level selection. The response model and phase-space correction factors are obtained from simulated  $t\bar{t}$  events.

The inclusive fiducial cross-section is obtained by integrating the unfolded differential cross-section over the kinematic bins, and its value is used to compute the normalised differential cross-section  $1/\sigma \cdot d\sigma^{\text{fid}}/dX^i$ .

The binning is chosen according to the resolution of each observable such that the fraction of events in the diagonal bins is greater than 50% and the data statistical uncertainty in each bin is below 10%. The unfolding is performed using three iterations. This value is chosen because using more iterations produces only small event-count changes in each of the bins and gives lower correlation factors between them.

The unfolding procedure was validated by performing four tests. To ensure that the results are not biased by the MC generator used in the unfolding procedure, a study was performed in which the particle-level spectra from the default POWHEG BOX + PYTHIA simulation are altered by changing the shape of distributions using either a linear or sinusoidal function. Here, maximal variations of 40% are considered. Alternatively, the shapes are also changed by a smooth function that describes the ratio of the background-subtracted data distribution to the distribution obtained from simulated  $t\bar{t}$  events. Here, variations of up to 30% are applied. The reweighted corrected-level  $t\bar{t}$  distributions are then unfolded using the nominal migration matrices and

unfolding corrections to see if the reweighted particle-level distributions can be recovered. These tests are referred to as ‘stress tests’. The altered shapes are recovered within statistical uncertainties by the unfolding procedure in the case of the linear reweighting, while deviations of 5% in the tails are observed for the sinusoidal reweighting and the MC-to-data reweighting. In order to explore the impact of discrepancies between data and MC simulations in variables other than the unfolded observables, a cross-reweighting scenario is also considered. This uses the same procedure as in the stress tests above, but the reweighting function is obtained from one of the one-dimensional observables, either  $n_{\text{ch}}$  or  $\sum n_{\text{ch}} p_{\text{T}}$ , and applied to the other two observables. The reweighting function is defined as a smooth function that describes the ratio of the measured distribution (unfolded data) to the distribution obtained from simulated  $t\bar{t}$  events at the particle level. Similarly to the nominal stress tests, the altered shapes are recovered with a maximum deviation of 3.6%. Non-closure deviations after applying the background-subtracted data to  $t\bar{t}$  MC ratio reweighting and cross-reweighting scenario are added as systematic uncertainties (see Section 8.4).

## 8 Systematic uncertainties

Systematic uncertainties from several sources affect the measured observables. Experimental uncertainties are associated with the reconstruction of the four-momenta of final-state objects: electrons, muons, jets, and tracks. Systematic uncertainties in the estimation of the background event rates and uncertainties related to the unfolding procedure are taken into account. Modelling uncertainties are related to parameters and methods used in MC event generation, and the statistical uncertainties are also considered. The total uncertainty is then calculated, assuming all the systematic uncertainties are uncorrelated. Finally, the sensitivity of the measured observables to variations of initial- and final-state radiation and variation of  $h_{\text{damp}}$  is tested. In the following, the different uncertainty components are explained in more detail.

### 8.1 Detector systematic uncertainties

Detector systematic uncertainties, due to residual differences between data and MC simulations after calibration, and uncertainties in corrective scale factors are propagated through the analysis and evaluated for  $t\bar{t}$  production and the  $tW$  background. Non-significant uncertainties with negligible differences between the systematically varied distribution and the nominal distribution at the corrected level are not considered further. The uncertainties are evaluated by building distributions from events contributed by the systematically varied  $t\bar{t}$  production process and the systematically varied background processes such as  $tW$ . The resulting distribution is then unfolded after subtracting the nominal background, using the nominal migration matrix and unfolding corrections. The relative deviation from the nominal unfolded distribution is then assigned as a systematic uncertainty. The following sources are considered:

- **Jet uncertainties**

The jet energy scale (JES) uncertainties are derived from test-beam data, in situ calibration measurements and simulation, and are parameterised in bins of jet  $p_{\text{T}}$  and  $\eta$  [83]. Its uncertainty is decomposed into a set of 30 uncorrelated components. Sources of uncertainty contributing to the JES uncertainties include the  $\eta$  intercalibration, jet flavour composition and response, differences between jets induced by  $b$ -quarks and those from gluons or light-quarks, detector modelling, non-closure, punch-through losses and several pile-up properties. The effect of single high- $p_{\text{T}}$  particles is also included. Furthermore, uncertainties related to the jet energy resolution (JER) are considered. The uncertainty of the JVT efficiency [84] is found to be negligible.



Since this analysis makes use of  $b$ -tagging, the associated uncertainties in tagging efficiency for  $b$ -jets,  $c$ -jets, and light-flavour jets are considered by using an uncertainty model containing nine/four/four independent variations for the  $b$ -/ $c$ -/light-flavour jet calibrations and two components for the MC-based extrapolation to high- $p_T$  jets [85].

The combination of these uncertainties is labelled ‘Jet’ in Figure 5, with the main contribution coming from the jet flavour response.

- **Track uncertainties**

Systematic uncertainties associated with  $d_0$  and  $z_0$  impact parameter resolutions are applied, based on measurements of minimum-bias data and  $Z \rightarrow \mu\mu$  events and parameterised in  $p_T$  and  $\eta$ . In addition, the impact parameter resolution is particularly sensitive to the accuracy of the simulation’s modelling of the number of inactive modules in the pixel detector [76]. An uncertainty is estimated by randomly disabling 5% of the pixel modules.

Weak modes in the alignment uncertainties are accounted for by applying bias corrections to tracks as a function of their angular position to simulate the impact of such modes. Variations biasing  $d_0$  and  $z_0$  are taken into account, as well as sagitta biases [103].

Finally, uncertainties in the track reconstruction efficiency as well as track fake-rates are evaluated. Two main sources of uncertainty are considered for the track reconstruction efficiency: amount-of-material uncertainties and the physics model used in simulation. The effects of these uncertainties are assessed by comparing the efficiency in samples with the different configurations: the passive material of the ID is scaled by 5%, the passive material of the IBL is scaled by 10%, the passive material in the Patch Panel 0 (PP0) region is scaled by 25%, and finally the GEANT4 physics model is varied [104]. The estimation of the track fake-rate is based on observations of differences in the non-linear component of the evolution of track multiplicity as a function of  $\langle\mu\rangle$  as described in Ref. [76] and the uncertainty is assumed to be 100%.

The combination of these uncertainties is labelled ‘Track’ in Figure 5, with the main contribution coming from the global tracking efficiency.

- **Lepton uncertainties**

Uncertainties due to the lepton trigger, reconstruction and selection efficiencies in simulation are estimated from a tag-and-probe method applied to electrons and muons from leptonic decays of  $Z$  bosons and  $J/\psi$  mesons using methods similar to those in Refs. [78, 105]. The same processes, using methods described in Refs. [77, 78], are used to estimate uncertainties in the lepton momentum scale and resolution. The combination of these uncertainties is labelled ‘Lepton’ in Figure 5.

- **Average number of additional pile-up events**

Scale factors are applied to reweight simulated events in order to obtain the distribution of the average number of additional pile-up events corresponding to the distribution obtained in data. An uncertainty in these reweighting scale factors is considered, based on the difference between the instantaneous luminosities in data and simulation, and is labelled as ‘ $\mu$ -reweight’.

Figure 5 shows the fractional uncertainty from each of the above-mentioned categories for the three observables. The dominant uncertainties are track and jet uncertainties, and they are observed to be largest at low or high values of the measured distributions.

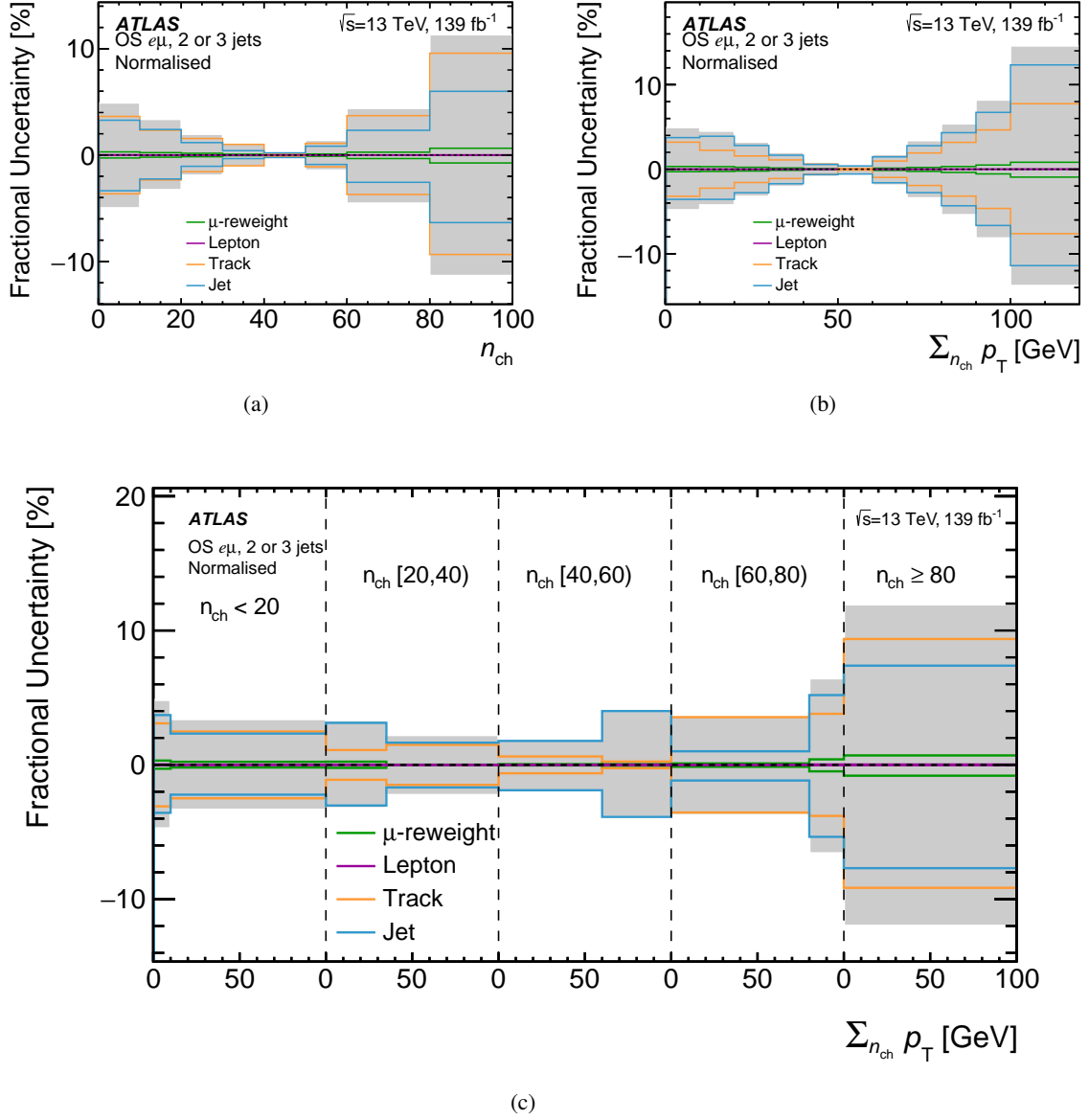


Figure 5: Fractional uncertainties in the measured normalised differential cross-section as a function of (a)  $n_{\text{ch}}$ , (b)  $\sum_{n_{\text{ch}}} p_{\text{T}}$ , and (c)  $\sum_{n_{\text{ch}}} p_{\text{T}}$  in bins of  $n_{\text{ch}}$  for all experimental systematic uncertainties. The  $x$ -axis in (c) is split into five bins of  $n_{\text{ch}}$  by the dashed vertical lines and  $\sum_{n_{\text{ch}}} p_{\text{T}}$  is presented in each bin. The grey bands represent the presented components added in quadrature. Events beyond the  $x$ -axis range are included in the last bin in (a) and (b), and in the corresponding last bin of each slice in (c).

## 8.2 Track-based-background uncertainties

The uncertainty in the rate of background pile-up tracks is evaluated by subtracting a varied number of pile-up tracks from the number of reconstructed tracks in  $t\bar{t}$  simulated events. The variation is given by a re-evaluated pile-up scale factor  $c_{\text{PU}}$ , described in Section 6.2, from either systematically varied samples, alternative samples, or sideband fits. The resulting distribution is then unfolded using the nominal migration matrix and efficiency. The relative deviation from the nominal unfolded distribution is assigned as a systematic uncertainty. The following uncertainties in the scale factor are considered:

- The uncertainty associated with the  $z_0$  impact parameter resolution is evaluated using systematically varied  $t\bar{t}$  samples and it is found to be smaller than 2%.
- The uncertainty in the reweighting procedure used to correct the modelling of pile-up tracks in simulation is evaluated by varying the selection of pile-up vertices and leads to an uncertainty of 1.4% in the pile-up scale factor.
- Uncertainties due to the influence of hard-scatter primary tracks are estimated by fitting a sideband region with  $|z_0| > 1$  mm. The resulting uncertainty ranges from less than 1% for most regions to as much as 14% in the high  $n_{\text{trk,out}}$  and low  $\mu$  region.
- The uncertainty due to different beam conditions and alignment settings is evaluated by re-fitting the rate of background pile-up tracks for different collision-data periods and leads to an uncertainty of 0.4% in the pile-up scale factor.
- Uncertainties in the modelling of parton showers, hadronisation and the underlying event, are evaluated by comparing the nominal samples to alternative samples for which POWHEG BOX is interfaced to HERWIG 7.1.3 instead of PYTHIA 8.230. The main difference is caused by the different underlying-event models and thus by a modified  $z_0$  distribution of the hard-scatter primary tracks used in the fit. The largest and dominant uncertainty is found for low  $n_{\text{trk,out}}$  values, being at a level of 17%, while otherwise the uncertainty is a few percent.

All pile-up scale-factor uncertainties are added in quadrature and the resulting relative uncertainties are presented in Table 5. The largest uncertainty is observed for events with  $n_{\text{trk,out}} < 20$ , mainly due to differences in the modelling of the UE between PYTHIA and HERWIG. Finally, the uncertainty due to the non-closure of the subtraction method in a few bins is added in quadrature to the combined pile-up scale-factor uncertainty and labelled as ‘PU subtraction’ in Figure 8.

Table 5: The total pile-up scale-factor relative uncertainty parameterised in  $\mu$  and  $n_{\text{trk}}$  and expressed in percent.

Region	$n_{\text{trk,out}} < 20$	$20 \leq n_{\text{trk,out}} < 40$	$40 \leq n_{\text{trk,out}} < 60$	$60 \leq n_{\text{trk,out}} < 80$	$80 \leq n_{\text{trk,out}} \leq 100$
$\mu < 20$	16	3	3	5	15
$20 \leq \mu < 40$	16	1	1	3	9
$\mu \geq 40$	18	2	2	4	5

## 8.3 Event-based-background uncertainties

Uncertainties due to the normalisation of the background processes are assessed by varying the normalisation of each background process according to its uncertainty. The following uncertainties, based on inclusive

cross-section uncertainties, are considered: single-top-quark and  $t\bar{t}+V$  processes 10%, fake leptons 50%, and other processes 20%. The combination of these uncertainties is labelled ‘Background rates’.

Uncertainties related to  $tW$  background modelling are assessed using samples where the renormalisation and factorisation scales in the matrix-element calculations and the intensity of initial-state and final-state radiation are varied. These uncertainties are labelled ‘ $tW$  Scale  $\mu_r$ ’, ‘ $tW$  Scale  $\mu_f$ ’, ‘ $tW$  ISR  $\alpha_s$ ’ and ‘ $tW$  FSR’, respectively. The parameter variations in the generation of these samples are identical to those used for the nominal  $t\bar{t}$  process, described in Section 8.5. The uncertainty from the scheme for removing the overlap of the  $tW$  process with  $t\bar{t}$  production is evaluated by comparing the nominal  $tW$  sample, using the diagram-removal scheme, with a sample produced with an alternative scheme (diagram subtraction) [55] and is labelled as ‘ $tW$  DS vs. DR’.

The above uncertainties are evaluated by building distributions where the  $t\bar{t}$  and systematically varied background contributions are added together. Each resulting distribution is then unfolded after subtracting the nominal background, using the nominal migration matrix and efficiency. Again the relative deviation from the measured unfolded distribution is assigned as a systematic uncertainty.

Figure 6 shows the fractional uncertainty from each of the above-mentioned categories for the three observables and the combination of these uncertainties is labelled ‘Event background’ in Figure 8.

#### 8.4 Unfolding-technique uncertainties

Based on the non-closure of the stress tests using the MC distribution reweighted to background-subtracted data and the cross-reweighting scenario, an unfolding-technique uncertainty is defined. It was checked that the cross-reweighting scenario together with generator uncertainties cover potential differences in the responses of the measured observables due to track- $p_T$  differences between simulation and data. The reweighted corrected-level  $t\bar{t}$  distribution is unfolded using the nominal migration matrix and unfolding corrections. The envelope of relative differences between these unfolded distributions is assigned as a systematic uncertainty, referred to as the ‘Unfolding technique’ in Figure 8. For the distributions of  $n_{\text{ch}}$  and  $\sum_{n_{\text{ch}}} p_T$ , the unfolding-technique uncertainty ranges between 0.1% and 2.2%, and is less than the statistical uncertainty of the data in all bins except the first one in each of these distributions. In the two-dimensional measurement of  $\sum_{n_{\text{ch}}} p_T$  in bins of  $n_{\text{ch}}$  the unfolding-technique uncertainty ranges between 0.3% and 3.6%, and exceeds the statistical uncertainty of the data in five of the nine bins.

#### 8.5 $t\bar{t}$ modelling uncertainties

The choice of MC generator and its settings determines the kinematic properties of simulated  $t\bar{t}$  events. In order to assess uncertainties in different detector responses or different modellings of the event kinematics, an alternative MC sample is unfolded using the nominal PYTHIA 8 migration matrix and corrections and then compared with the generated ‘truth’ spectrum of that sample. The systematic uncertainty is then given by the size of the non-closure, which is propagated as a relative uncertainty. This procedure is applied to all following uncertainties if not stated otherwise.

Uncertainties in modelling parton showers, hadronisation, the underlying event, and colour reconnections are assigned by comparing the nominal samples with alternative samples for which POWHEG BOX was interfaced to HERWIG 7.1.3 instead of PYTHIA 8.230. When generating parton showers with HERWIG,

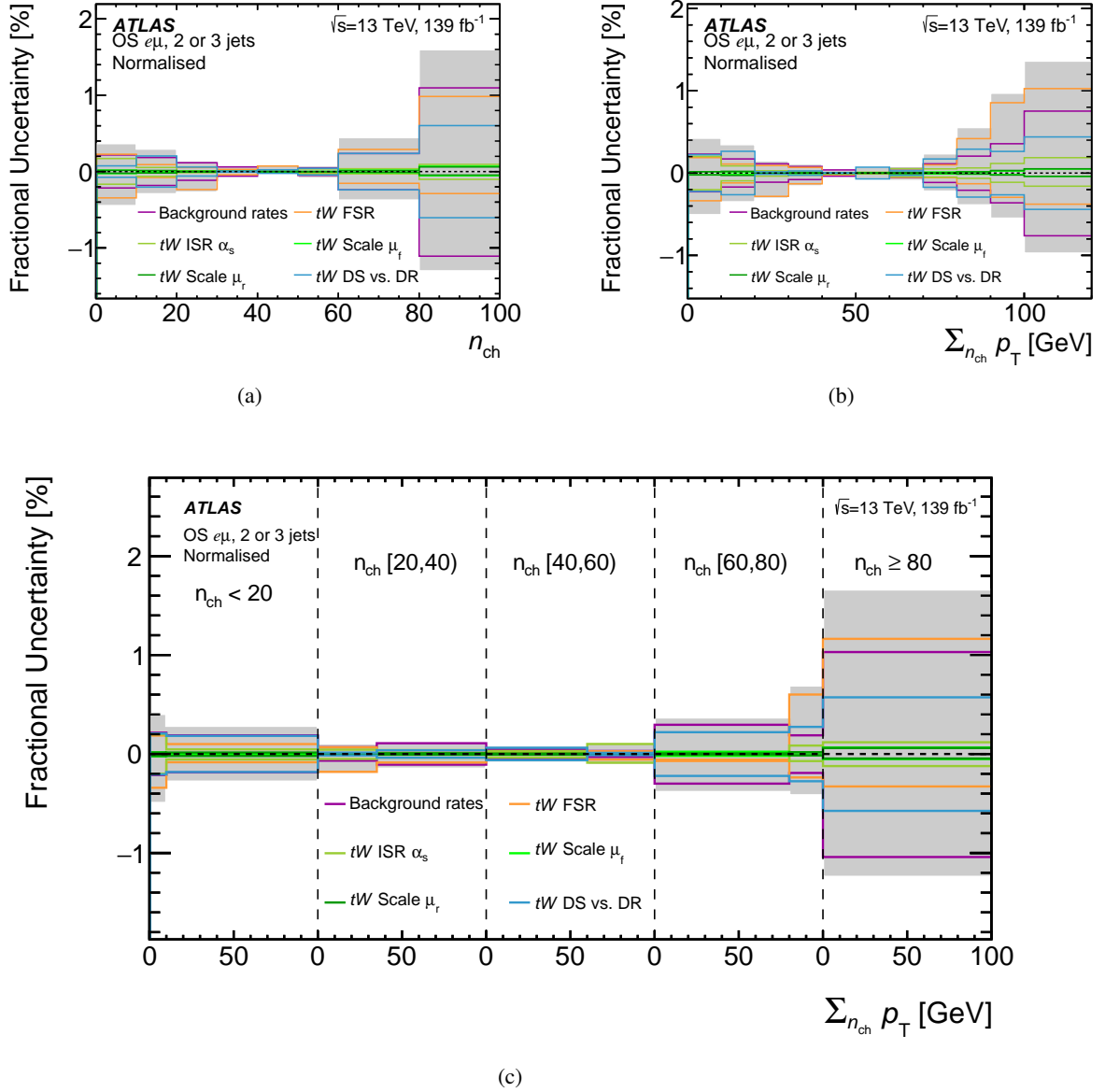


Figure 6: Fractional uncertainties in the measured normalised differential cross-section as a function of (a)  $n_{\text{ch}}$ , (b)  $\sum_{n_{\text{ch}}} p_{\text{T}}$ , (c)  $\sum_{n_{\text{ch}}} p_{\text{T}}$  in bins of  $n_{\text{ch}}$  for event-based background rate uncertainties and modelling uncertainties for the  $tW$  process. The  $x$ -axis in (c) is split into five bins of  $n_{\text{ch}}$  by the dashed vertical lines and  $\sum_{n_{\text{ch}}} p_{\text{T}}$  is presented in each bin. The grey bands represent the sum in quadrature of the presented components. Events beyond the  $x$ -axis range are included in the last bin in (a) and (b), and in the corresponding last bin of each slice in (c).

the MMHT2014<sub>LO</sub> [106] PDF set was used with the H7-UE-MMHT [7] set of tuned parameters. This uncertainty is labelled as ‘H7’ in the following.

Uncertainties related to the choice of renormalisation and factorisation scales in the matrix-element calculations, labelled as ‘Scale  $\mu_r$ ’ and ‘Scale  $\mu_f$ ’, are evaluated by varying the scales in a correlated way by factors of 2 and 0.5. The scale variations are implemented as generator weights in the nominal sample.

The uncertainty due to the scale choice for matching the matrix-element calculation of the  $t\bar{t}$  process to the parton shower is estimated using an additional  $t\bar{t}$  sample produced with the  $h_{\text{damp}}$  parameter set to  $3 \times m_{\text{top}}$  instead of  $1.5 \times m_{\text{top}}$ , while keeping all other generator settings the same as for the nominal sample of  $t\bar{t}$  events. This uncertainty is labelled as ‘ $h_{\text{damp}}$ ’ in the following.

Uncertainties in the intensity of initial-state and final-state radiation, labelled as ‘ISR  $\alpha_s$ ’ and ‘FSR’, are assessed by varying the parameter Var3c of the A14 parton-shower tune within the uncertainties of the tune and by varying the renormalisation scale  $\mu_r$  of final-state radiation, at which the strong coupling constant  $\alpha_s$  is evaluated, by factors of 0.5 and 2.0. The two variations, of Var3c and  $\mu_r$ , are handled independently.

Uncertainties in the modelling of colour reconnection effects are studied using samples with modified CR settings or using different CR models. A detailed description of the tuned parameter settings is given in Ref. [22], and the CR models are described in Section 2.1. The final CR uncertainty is then given by the envelope of the non-closure differences of CR0–CR2 comparisons and labelled as ‘CR’.

The uncertainty due to the choice of PDF is evaluated using the PDF4LHC15 combined PDF set [107] with 30 symmetric eigenvectors. The  $t\bar{t}$  production sample is reweighted with the central value and eigenvectors of the combined PDF set to construct a set of systematically varied templates. This uncertainty is labelled as ‘PDF4LHC’.

In all uncertainty evaluations mentioned above the alternative samples or reweighted samples are normalised to the total cross-section of the nominal samples.

Figure 7 shows the fractional uncertainty from each of the above-mentioned categories for the three observables. The dominant uncertainties are uncertainties in the modelling of the parton shower and hadronisation and in the choice of colour reconnection model, especially in the tails of the distributions. The combination of all  $t\bar{t}$ -modelling-related uncertainties is labelled ‘Signal modelling’ in Figure 8.

## 8.6 Statistical uncertainties

Statistical uncertainties are evaluated using pseudo-experiments, where the input distributions are varied and then unfolded. The RMS of the resulting distribution for each bin is used as an estimator of the uncertainty. Two sets of pseudo-experiments are used, one to account for the statistical uncertainty of the data and one to account for the limited size of the MC samples. The combination of these uncertainties is labelled ‘Statistical’ in Figure 8.

## 8.7 Luminosity

The uncertainty in the combined 2015–2018 integrated luminosity is 1.7% [35], obtained using the LUCID-2 detector [37] for the primary luminosity measurements. The combination of the uncertainty in the luminosity and all detector uncertainties is labelled ‘Detector’ in Figure 8.

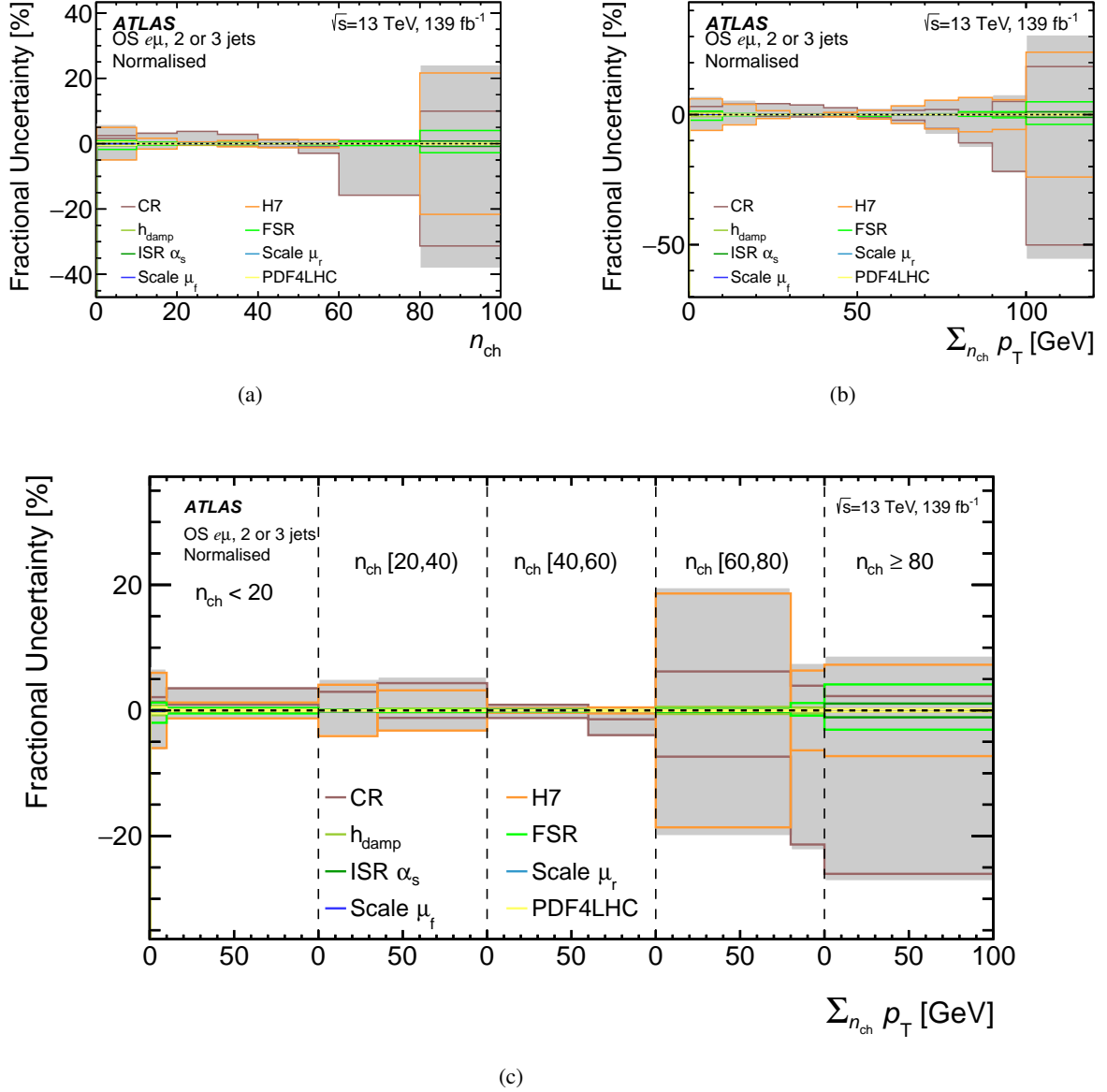


Figure 7: Fractional uncertainties in the measured normalised differential cross-section as a function of (a)  $n_{\text{ch}}$ , (b)  $\sum_{n_{\text{ch}}} p_{\text{T}}$ , (c)  $\sum_{n_{\text{ch}}} p_{\text{T}}$  in bins of  $n_{\text{ch}}$  for all  $t\bar{t}$  modelling uncertainties. The  $x$ -axis in (c) is split into five bins of  $n_{\text{ch}}$  by the dashed vertical lines and  $\sum_{n_{\text{ch}}} p_{\text{T}}$  is presented in each bin. The grey bands represent the sum in quadrature of the presented components. Events beyond the  $x$ -axis range are included in the last bin in (a) and (b), and in the corresponding last bin of each slice in (c).



A summary of the fractional uncertainty from each of the previously mentioned categories is shown for the three observables in Figure 8. The pile-up tracks background and the modelling uncertainties generally have the largest impact on the measurement.

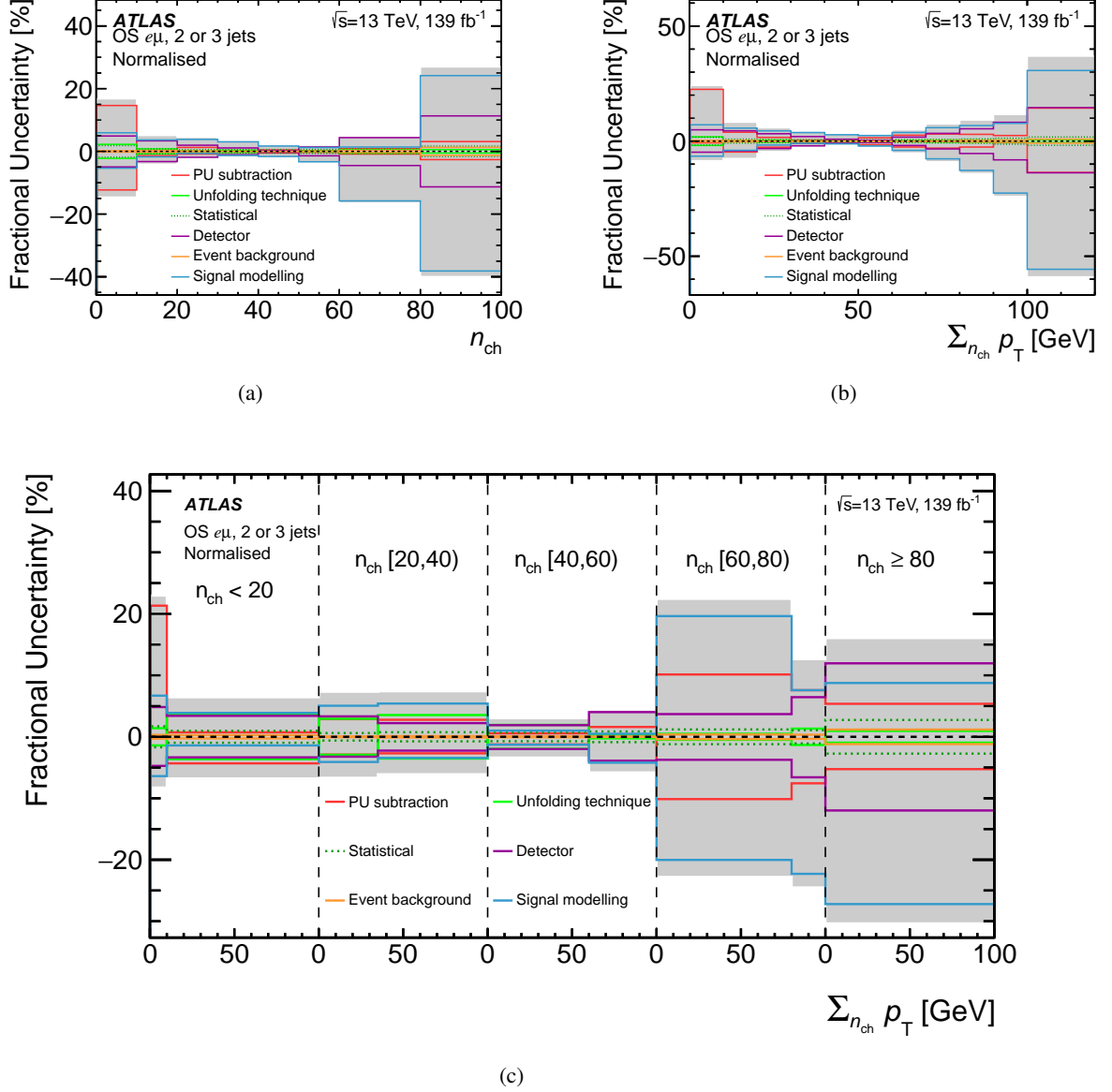
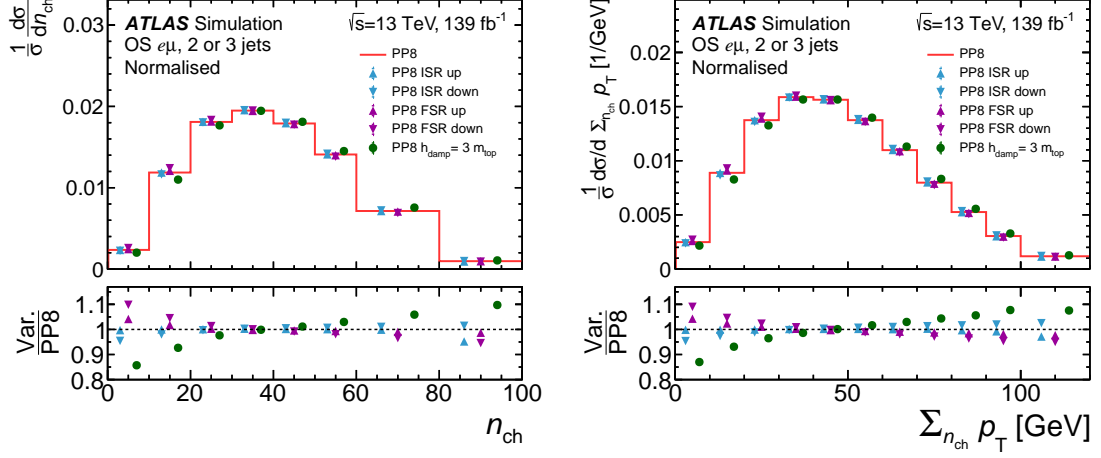


Figure 8: Fractional uncertainties in the measured normalised differential cross-section as a function of (a)  $n_{\text{ch}}$ , (b)  $\sum_{n_{\text{ch}}} p_T$ , (c)  $\sum_{n_{\text{ch}}} p_T$  in bins of  $n_{\text{ch}}$  for all systematic and statistical uncertainties. The x-axis in (c) is split into five bins of  $n_{\text{ch}}$  by the dashed vertical lines and  $\sum_{n_{\text{ch}}} p_T$  is presented in each bin. The grey bands represent the sum in quadrature of the presented components. Events beyond the x-axis range are included in the last bin in (a) and (b), and in the corresponding last bin of each slice in (c).

## 8.8 Impact of scale variations in PYTHIA 8

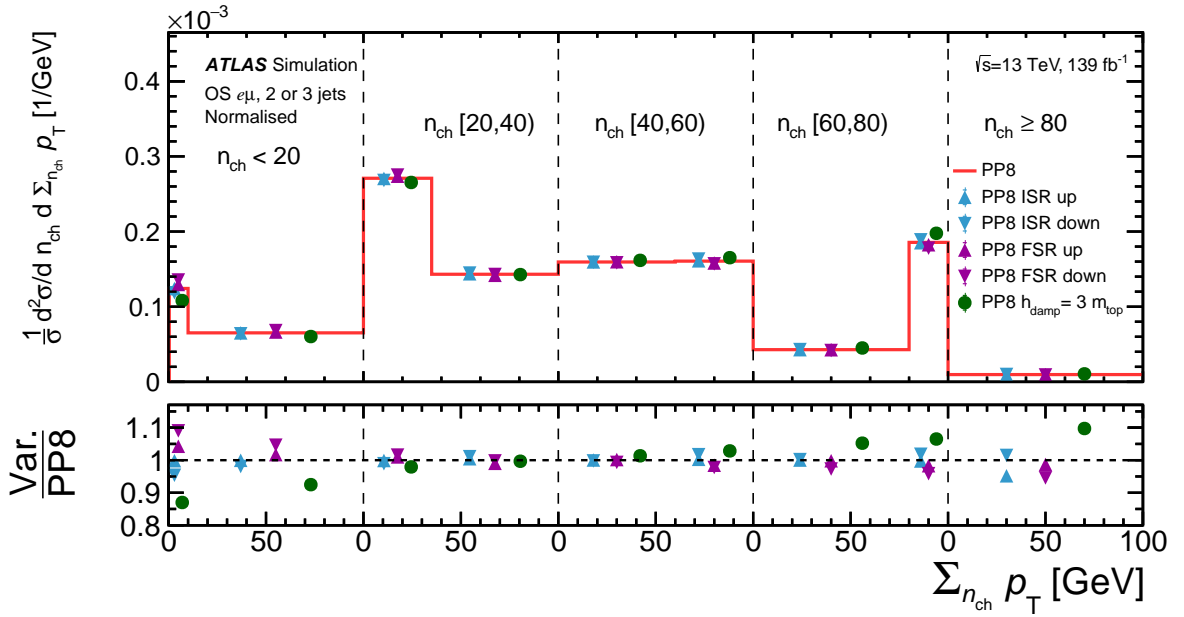
In order to understand the sensitivity of the measured differential cross-section distributions to variations of the POWHEG BOX + PYTHIA 8 set-up which are not directly connected with CR or UE modelling, the differential cross-sections are compared with those with varied ISR/FSR and  $h_{\text{damp}}$  parameter values.

Additional radiation can produce additional charged particles either by leakage from high- $p_T$  jets or by producing more soft jets. Thus, the normalised differential cross-sections are compared in Figure 9 for variations of the  $h_{\text{damp}}$ , ISR and FSR parameters in the nominal POWHEG BOX + PYTHIA 8 set-up as described in Section 8.5. The largest differences are seen for variations of  $h_{\text{damp}}$ . In the bulk of the  $n_{\text{ch}}$  and  $\sum_{n_{\text{ch}}} p_T$  distributions the differences are only a few percent. However, for low and high values of  $n_{\text{ch}}$  and  $\sum_{n_{\text{ch}}} p_T$ , the differences are as large as 15%. The influence of FSR and ISR is much weaker and is only significant for low  $\sum_{n_{\text{ch}}} p_T$  values. It can also be seen that the effect of ISR is the exact opposite of that of FSR for these bins. Based on these observations, it can be concluded that the measurement is only weakly sensitive to modelling of additional radiation in  $t\bar{t}$  events and that differences seen among the predictions from various generator set-ups can be attributed mainly to differences in the CR or UE modelling.



(a)

(b)



(c)

Figure 9: Normalised differential cross-section as a function of (a)  $n_{\text{ch}}$ , (b)  $\sum n_{\text{ch}} p_{\text{T}}$ , and (c)  $\sum n_{\text{ch}} p_{\text{T}}$  in bins of  $n_{\text{ch}}$ . The  $x$ -axis in (c) is split into five bins of  $n_{\text{ch}}$  by the dashed vertical lines and  $\sum n_{\text{ch}} p_{\text{T}}$  is presented in each bin. The nominal PYTHIA 8 prediction is compared with the predictions from different parameter variations in PYTHIA 8. Events beyond the  $x$ -axis range are included in the last bin. The lower panels show the ratio of each parameter-varied prediction to the nominal prediction in each bin.

## 9 Results

In this section, the analysis results are presented by comparing the normalised differential cross-sections with predictions from various MC models with different colour reconnection models and settings. The absolute differential cross-sections can be used to calculate the inclusive cross-section in the fiducial phase-space, and then the normalised differential cross-sections. At particle level the following requirements, as defined in Section 5.3, are imposed: events are required to have exactly one electron and one opposite-sign muon with  $p_T > 25$  GeV and  $|\eta| < 2.5$  and contain two or three jets with  $p_T > 25$  GeV and  $|\eta| < 2.5$ , where exactly two of the jets must be  $b$ -tagged. Requirements of  $m_{\ell\ell} > 15$  GeV and  $n_{\text{ch}} < 100$  are also applied. The resulting cross-section is:

$$\sigma_{t\bar{t}}^{\text{fid}} = 4.94_{-0.41}^{+0.43}(\text{syst}) \pm 0.03(\text{stat}) \pm 0.08(\text{lumi}) \text{ pb.}$$

In order to quantify how well the measured differential cross-sections agree with the various model predictions, their respective  $\chi^2$  values are calculated. The formula for calculating the  $\chi^2$  value, given the unfolded data, the model, and the covariance matrix, is defined as follows:

$$\chi^2 = V_{N_b}^T \cdot \text{Cov}_{N_b}^{-1} \cdot V_{N_b},$$

where  $N_b$  is the number of bins in the spectrum,  $V_{N_b}$  is the vector of differences between the unfolded data and the prediction, and  $\text{Cov}_{N_b}$  represents the covariance matrix.

The covariance matrix incorporates both the statistical and the systematic uncertainties. The systematic uncertainties include experimental uncertainties, signal modelling and background-related uncertainties as described in the previous chapter. In addition, uncertainties in the theoretical predictions, which are independent of the CR modelling, are considered and referred to as ‘theory uncertainties’. These theory uncertainties include scale variations in the matrix element and parton shower as well as the  $h_{\text{damp}}$  variation. The relative difference between the unfolded variation and the nominal unfolded distribution is taken as uncertainty.

The covariance matrix is obtained by performing pseudo-experiments in the following way: each bin of the measured distribution at the corrected level is varied according to a Poisson distribution with the number of observed events in this bin as the mean value. Gaussian-distributed shifts are coherently added for each detector-modelling systematic uncertainty and the MC statistical uncertainty by scaling each Poisson-fluctuated bin with the expected relative variation from the associated systematic uncertainty effect. In the case of a one-sided uncertainty, the corresponding contribution is symmetrised. The resulting distribution is then corrected with the nominal event-based background subtraction and passed through the standard unfolding procedure using the nominal corrections. Modelling uncertainties and an unfolding-procedure uncertainty are then coherently added with additional Gaussian-distributed shifts. For the normalised differential cross-sections,  $V_{N_b}$  is replaced by  $V_{N_b-1}$ , which is the vector of differences between the unfolded data and the prediction, discarding one of the  $N_b$  elements. Consequently,  $\text{Cov}_{N_b-1}$  is the sub-matrix derived from the full matrix of the normalised differential cross-sections by discarding the corresponding row and column. Hence, the obtained sub-matrix is invertible and the  $\chi^2$  does not depend on which element is discarded. In practice the last row and column is removed.

In addition, a covariance matrix including correlations of systematic uncertainties between the two observables  $n_{\text{ch}}$  and  $\sum_{n_{\text{ch}}} p_T$  is calculated and used to derive a so-called global  $\chi^2$  value, denoted by ‘Global( $n_{\text{ch}}, \sum_{n_{\text{ch}}} p_T$ )’. All of the covariance matrices have large off-diagonal elements, mainly due to the

modelling uncertainties. In the following comparisons of the differential cross-sections with predictions from different event generators using the nominal settings, predictions from CR models implemented in PYTHIA 8 and CR models implemented in HERWIG 7 are presented and discussed.

## 9.1 Comparisons with predictions from different generators

In Figure 10 the normalised differential cross-sections are compared with the predictions obtained from POWHEG+PYTHIA 8.230, SHERPA 2.2.10, and POWHEG+HERWIG 7 using three versions of HERWIG 7, namely 7.0.4, 7.1.3 and 7.2.1. These three versions have different UE and CR tunes and partially modified MPI models. Table 6 shows the  $\chi^2$ /NDF values resulting from the comparisons of the measured differential cross-sections with the corresponding prediction from these generators for each observable. Figure 10 shows that the unfolded data disagree with the predictions from SHERPA 2.2.10, which does not include CR effects, thus illustrating their importance. The  $n_{\text{ch}}$  measured cross-section is approximately equally well described by PYTHIA 8.230 and HERWIG 7, while the  $\sum_{n_{\text{ch}}} p_{\text{T}}$  measured normalised cross-section has a much better  $\chi^2$  value for HERWIG 7. With increasing HERWIG 7 version number, the predicted and observed  $n_{\text{ch}}$  distributions show worse agreement, while for  $\sum_{n_{\text{ch}}} p_{\text{T}}$  the level of agreement remains about the same. The distribution of  $\sum_{n_{\text{ch}}} p_{\text{T}}$  in bins of  $n_{\text{ch}}$  is best described by the three versions of HERWIG 7 in comparison with PYTHIA 8.230 and SHERPA 2. Relative to PYTHIA 8.230, the lower  $\chi^2$  value for SHERPA 2 is due to the inclusion of the theory uncertainties. Overall, the best agreement is achieved using HERWIG 7.0.4.

Table 6: The  $\chi^2$  and NDF for measured normalised differential cross-sections obtained by comparing the different predictions with the unfolded data. Global( $n_{\text{ch}}, \sum_{n_{\text{ch}}} p_{\text{T}}$ ) denotes the scenario in which the covariance matrix is built including correlations of systematic uncertainties between the two observables  $n_{\text{ch}}$  and  $\sum_{n_{\text{ch}}} p_{\text{T}}$ .

Observable	$n_{\text{ch}}$	$\sum_{n_{\text{ch}}} p_{\text{T}}$	Global( $n_{\text{ch}}, \sum_{n_{\text{ch}}} p_{\text{T}}$ )	$\sum_{n_{\text{ch}}} p_{\text{T}}$ in bins of $n_{\text{ch}}$
NDF	7	10	17	8
Generator set-up	$\chi^2$			
POWHEG+PYTHIA 8.230	62	106	434	224
CR0	55	113	629	129
CR1	98	60	581	158
CR2	58	179	402	238
POWHEG+HERWIG 7.0.4	39	16	145	29
POWHEG+HERWIG 7.1.3	53	42	188	41
POWHEG+HERWIG 7.2.1	78	25	313	87
POWHEG+HERWIG Baryonic CR	75	20	241	29
POWHEG+HERWIG Stat CR	23	40	121	39
SHERPA 2.2.10	77	211	263	124

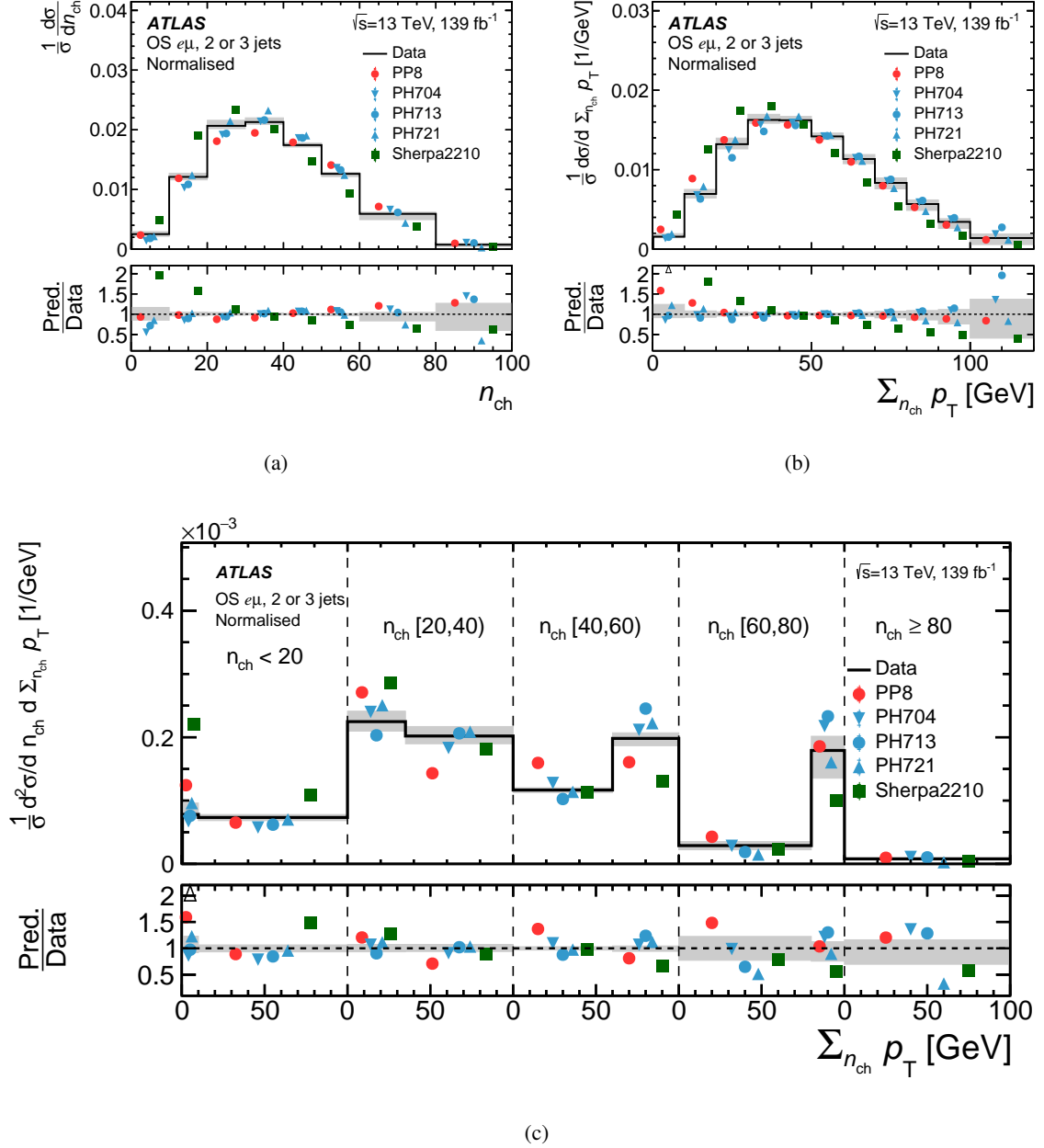


Figure 10: Normalised differential cross-section as a function of (a)  $n_{ch}$ , (b)  $\Sigma n_{ch} p_T$ , and (c)  $\Sigma n_{ch} p_T$  in bins of  $n_{ch}$ . The  $x$ -axis in (c) is split into five bins of  $n_{ch}$  by the dashed vertical lines and  $\Sigma n_{ch} p_T$  is presented in each bin. Unfolded data are shown as the black line with the grey band corresponding to the total uncertainty in each bin. The results are compared with the predictions of different MC event generators. Events beyond the  $x$ -axis range are included in the last bin. The lower panels show the ratio of each prediction to data in each bin. The black triangular markers in the lower panels point to prediction-to-data ratio values which lie beyond the shown  $y$ -axis range.

## 9.2 Comparison with different CR models in PYTHIA 8

In Figure 11, the normalised differential cross-sections are compared with predictions from different CR models implemented in PYTHIA 8. Table 6 shows  $\chi^2/\text{NDF}$  values resulting from the comparisons between the measured differential cross-sections and the predictions. The CR0 model is based on the same MPI-based CR model as the nominal POWHEG+PYTHIA 8.230 sample. One can see that the unfolded data is in better agreement with the sample using the dedicated CR0 set of tuned parameters than with the nominal sample without a dedicated CR tune, but that the global  $\chi^2$  is somewhat larger. For the CR1 model,  $\sum_{n_{\text{ch}}} p_{\text{T}}$  is described best, while  $n_{\text{ch}}$  is worst among the three CR models. In contrast, the  $\sum_{n_{\text{ch}}} p_{\text{T}}$  distribution, obtained with the CR2 model, is in worst agreement with data, while the other distributions have  $\chi^2/\text{NDF}$  values comparable to those obtained from POWHEG+PYTHIA 8.230 with the A14 set of tuned parameters. It can also be seen that neither these CR models nor the A14 tune can describe the lower part of the  $\sum_{n_{\text{ch}}} p_{\text{T}}$  distribution.

## 9.3 PYTHIA 8 CR and UE parameters

In this section, the normalised differential cross-sections are compared with the predictions obtained from PYTHIA 8 with variations of specific parameters connected with the CR and UE modelling. These variations are applied to the nominal PYTHIA 8 set-up using CR0 and the A14 tune. The varied parameters are:

- CR range parameter  $R_{\text{rec}}$  is set to its maximal value of  $R_{\text{rec}} = 10$ , such that the reconnection probability reaches saturation (maxCR). The default is 1.71.
- CR is switched off, i.e.  $R_{\text{rec}} = 0$  (noCR).
- MPI parameter  $p_{\text{T}0}^{\text{ref}}$ , by default set to 2.09, is lowered to 2.0 and raised to 2.2.
- UE activity is varied by using the Var1 eigentune of the A14 tune [21]. This eigentune includes variations of  $\alpha_s$  and variations of  $R_{\text{rec}}$ .

In Figure 12 the normalised differential cross-sections are compared with both the predictions with these varied parameter values and the predictions of the nominal PYTHIA 8 sample. The predictions without CR are not compatible with the data, while predictions using a maximal probability for colour reconnection in the default PYTHIA 8 CR0 model are still compatible with the measurement. Another observation is that the overall  $\sum_{n_{\text{ch}}} p_{\text{T}}$  value is not sensitive to the CR strength, but it is sensitive within the  $n_{\text{ch}}$  ranges. Finally, a clear sensitivity to the variation of  $p_{\text{T}0}^{\text{ref}}$  is observed and the presented measurement would suggest a value of  $p_{\text{T}0}^{\text{ref}}$  lower than the one used in the A14 tune. Thus, it can be concluded that for further tuning of CR models both the parameters of the CR model itself and the parameters of the UE should be included.



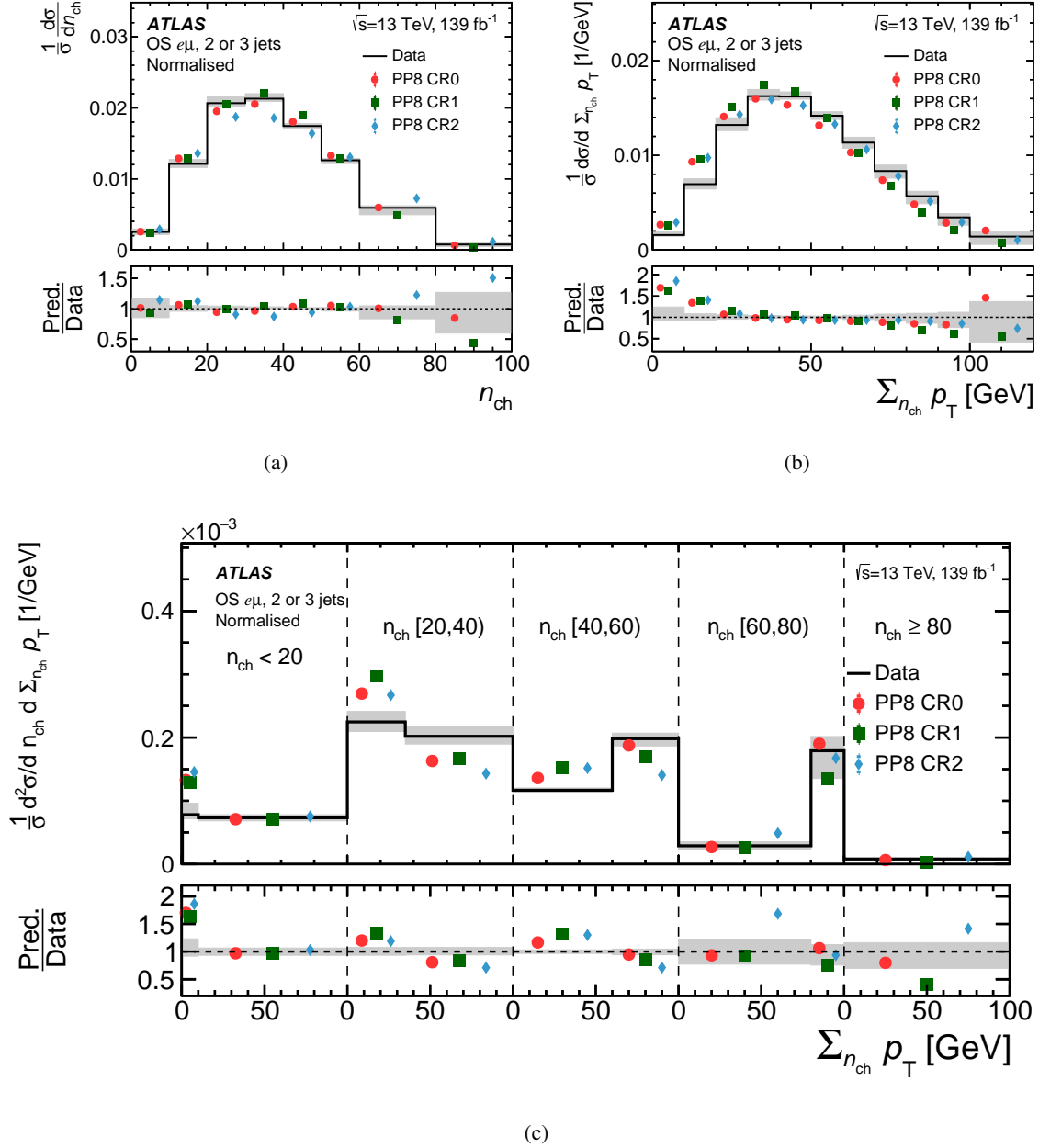
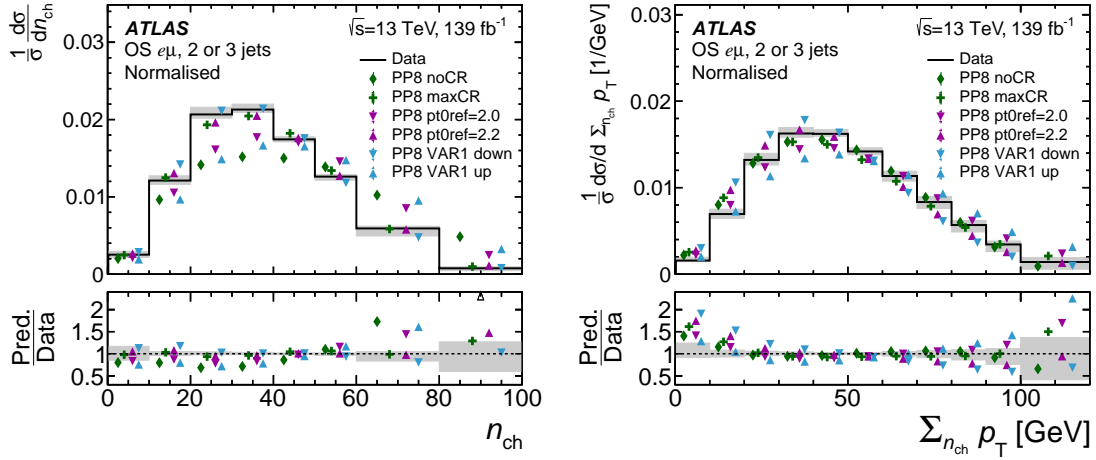
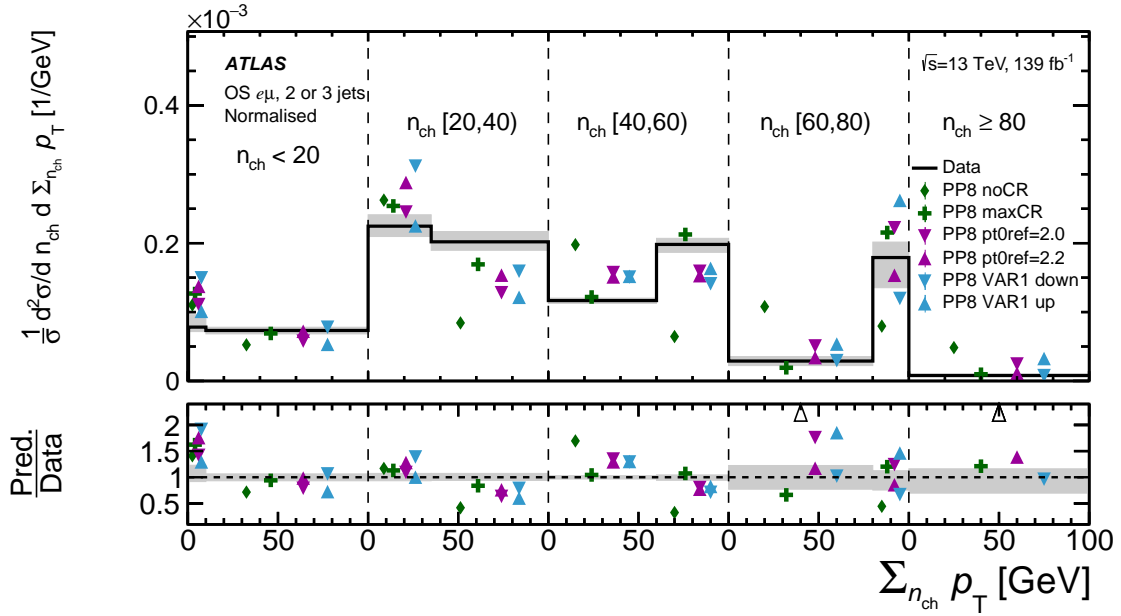


Figure 11: Normalised differential cross-section as a function of (a)  $n_{\text{ch}}$ , (b)  $\sum n_{\text{ch}} p_T$ , and (c)  $\sum n_{\text{ch}} p_T$  in bins of  $n_{\text{ch}}$ . The  $x$ -axis in (c) is split into five bins of  $n_{\text{ch}}$  by the dashed vertical lines and  $\sum n_{\text{ch}} p_T$  is presented in each bin. Unfolded data are shown as the black line with the grey band corresponding to the total uncertainty in each bin. The results are compared with the predictions of different CR models in PYTHIA 8. Events beyond the  $x$ -axis range are included in the last bin. The lower panels show the ratio of each prediction to data in each bin.



(a)

(b)



(c)

Figure 12: Normalised differential cross-section as a function of (a)  $n_{\text{ch}}$ , (b)  $\sum n_{\text{ch}} p_T$ , and (c)  $\sum n_{\text{ch}} p_T$  in bins of  $n_{\text{ch}}$ . The  $x$ -axis in (c) is split into five bins of  $n_{\text{ch}}$  by the dashed vertical lines and  $\sum n_{\text{ch}} p_T$  is presented in each bin. Unfolded data are shown as the black line with the grey band corresponding to the total uncertainty in each bin. The results are compared with the predictions of different CR and UE parameter variations in PYTHIA 8. Events beyond the  $x$ -axis range are included in the last bin. The lower panels show the ratio of each prediction to data in each bin. The black triangular markers in the lower panels point to prediction-to-data ratio values which lie beyond the shown  $y$ -axis range.

## 9.4 HERWIG 7.2 CR models

In this section, the normalised differential cross-sections are compared with the predictions obtained from different CR models implemented in HERWIG 7.2, as described in Section 2.2. In Figure 13 the normalised differential cross-sections are compared with distributions from both the predictions of the different CR models and predictions without any CR. Similarly to PYTHIA 8, predictions without CR do not describe data well, but in the case of HERWIG 7.2 some deficiencies are compensated for by the UE modelling. Table 6 shows  $\chi^2$  values resulting from the comparisons between the measured differential cross-sections and the predictions. For the  $n_{\text{ch}}$  distribution the newly introduced statistical CR model performs the best, while for  $\sum_{n_{\text{ch}}} p_{\text{T}}$  the baryonic CR model agrees better with the observed distribution. Thus, even without dedicated retuning of the alternative CR models, they agree better with the measured data than the default plain CR model. In general, agreement with the  $\sum_{n_{\text{ch}}} p_{\text{T}}$  distribution is better than for any other generator considered in this analysis.

## 9.5 Top-quark-specific CR models of PYTHIA 8

In this section, the normalised differential cross-sections are compared with the predictions obtained from different top-quark-specific CR models implemented in PYTHIA 8, as described in Section 2.1. In Figure 14 the normalised differential cross-sections are compared with the distributions from the different predictions, where the strength parameter ( $s$ ) is set to either  $s = 0.1$  or  $s = 1.0$ . Distributions obtained using  $s = 1.0$  represent the maximal CR effect of the respective model. The spread between the distributions obtained with  $s = 0.1$  and  $s = 1.0$  is therefore a measure of the sensitivity of the observables to the corresponding model. Of the five models studied, only TCR1 and TCR3 show sizeable differences between the two settings.

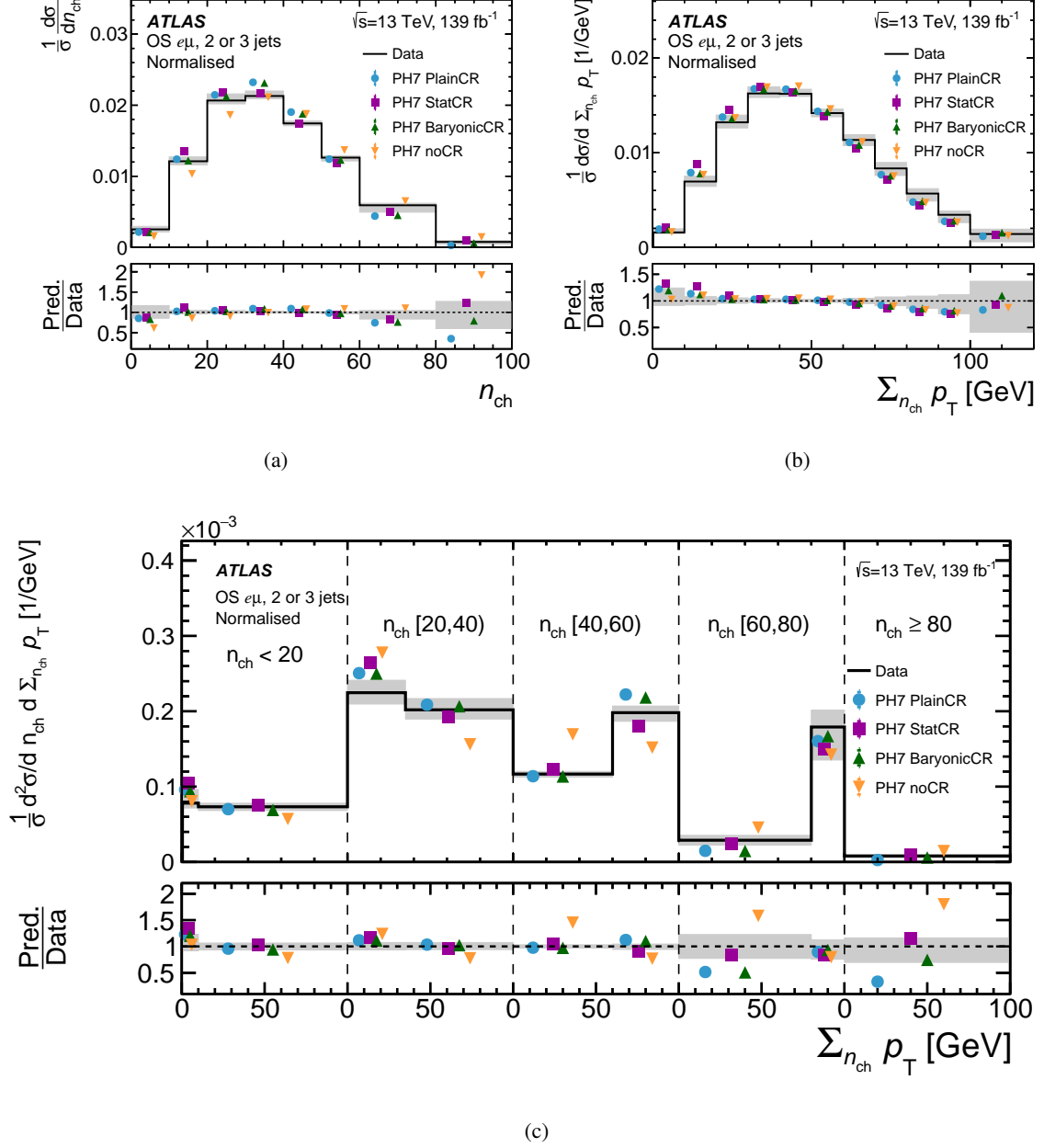


Figure 13: Normalised differential cross-section as a function of (a)  $n_{\text{ch}}$ , (b)  $\sum_{n_{\text{ch}}} p_T$ , and (c)  $\sum_{n_{\text{ch}}} p_T$  in bins of  $n_{\text{ch}}$ . The  $x$ -axis in (c) is split into five bins of  $n_{\text{ch}}$  by the dashed vertical lines and  $\sum_{n_{\text{ch}}} p_T$  is presented in each bin. Unfolded data are shown as the black line with the grey band corresponding to the total uncertainty in each bin. The results are compared with the predictions of different CR models in HERWIG 7.2. Events beyond the  $x$ -axis range are included in the last bin. The lower panels show the ratio of each prediction to data in each bin.

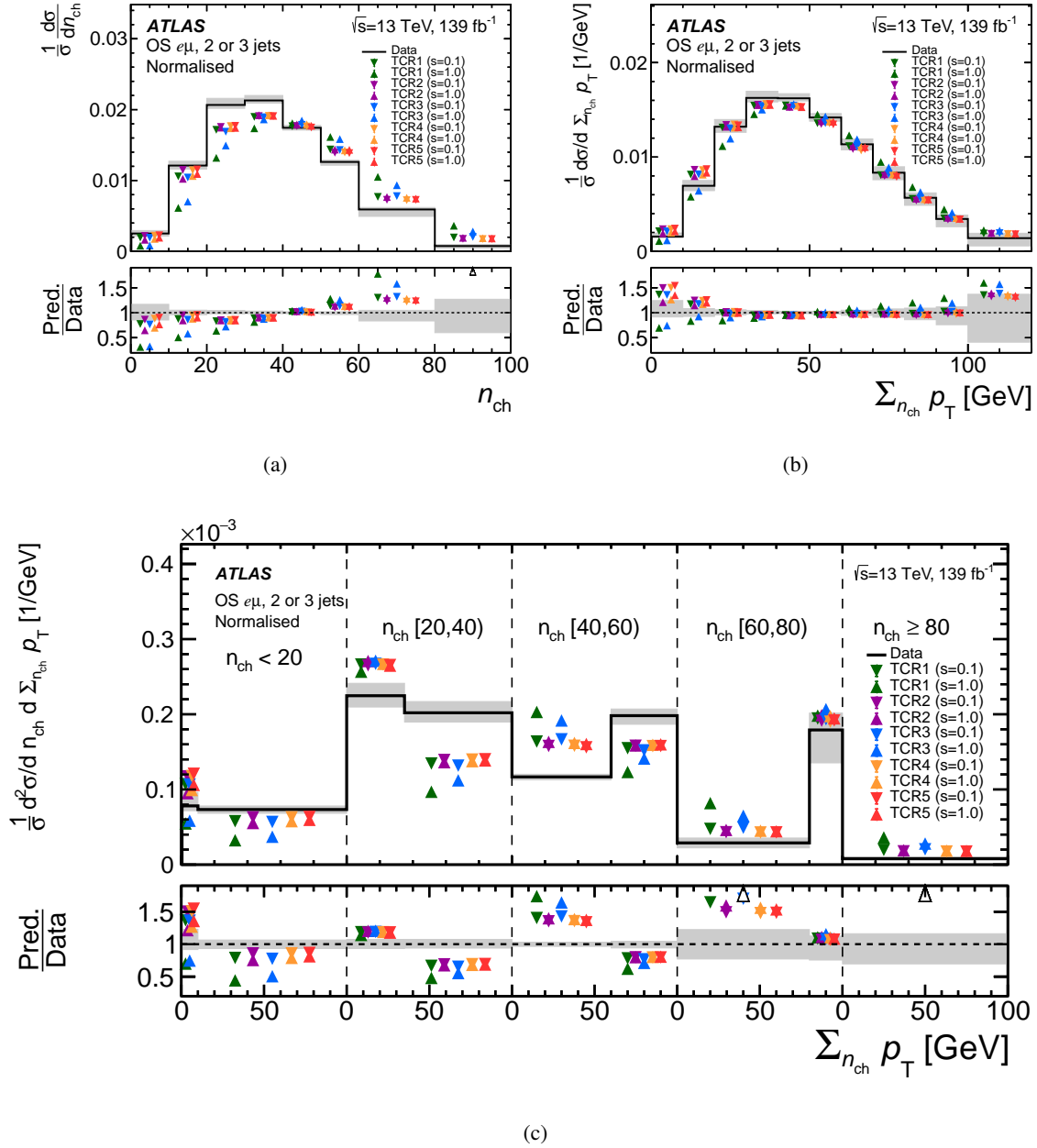


Figure 14: Normalised differential cross-section as a function of (a)  $n_{\text{ch}}$ , (b)  $\Sigma n_{\text{ch}} p_T$ , and (c)  $\Sigma n_{\text{ch}} p_T$  in bins of  $n_{\text{ch}}$ . The  $x$ -axis in (c) is split into five bins of  $n_{\text{ch}}$  by the dashed vertical lines and  $\Sigma n_{\text{ch}} p_T$  is presented in each bin. Unfolded data are shown as the black line with the grey band corresponding to the total uncertainty in each bin. The results are compared with the predictions of different top-quark-specific CR models in PYTHIA 8. Events beyond the  $x$ -axis range are included in the last bin. The lower panels show the ratio of each prediction to data in each bin. The black triangular markers in the lower panels point to prediction-to-data ratio values which lie beyond the shown  $y$ -axis range.

## 10 Summary and conclusion

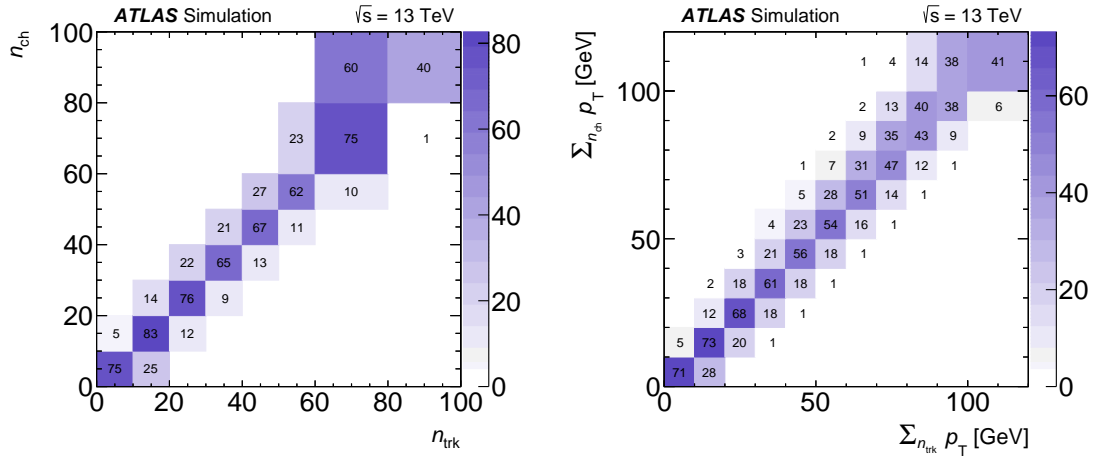
A measurement of three observables sensitive to colour reconnection effects in  $t\bar{t}$  events is presented using a data sample of  $139\text{ fb}^{-1}$  recorded by the ATLAS experiment at the LHC in proton–proton collisions at a centre-of-mass energy of  $\sqrt{s} = 13\text{ TeV}$  from 2015 to 2018. The charged-particle multiplicity, the scalar sum of their transverse momenta, and the scalar sum in bins of charged-particle multiplicity are measured in a fiducial phase space in events with exactly one isolated electron and one isolated muon with opposite charge, and two or three jets, where exactly two jets are required to be  $b$ -tagged. The observables are unfolded to stable-particle level and they are compared with predictions from parton-shower programs. The measurement has a precision of 2%–10% in the central bins and up to 50% in the outer bins. The combination of the first two observables, and a double-differential measurement of the two quantities, provide stringent constraints on the colour reconnection models in mainstream Monte Carlo generators. Using these results as a benchmark, more precise models can be developed in a dedicated tuning effort. Distributions of  $\sum_{n_{\text{ch}}} p_{\text{T}}$  obtained from HERWIG 7.2 are able to describe data well, while predictions from PYTHIA 8.2 for  $n_{\text{ch}}$  are better than those for  $\sum_{n_{\text{ch}}} p_{\text{T}}$ . The event generator SHERPA 2.2.10 does not include a colour reconnection model, which should lead to a higher particle multiplicity because the suppression mechanism from CR is missing. Instead, it predicts too few particles, and also its  $\sum_{n_{\text{ch}}} p_{\text{T}}$  distribution is softer than measured in data. Among the studied PYTHIA 8 colour reconnection models the best agreement with data is achieved with the default MPI-based set of tuned parameters and the largest disagreement is found for the set of parameters based on the gluon-move model. In the case of the MPI-based model, variations of the colour-reconnection range parameter to its minimal and maximal values are investigated and, similarly to SHERPA, a model without colour reconnection is excluded, while a maximal strength is still compatible with the measurement. Finally, top-quark-specific colour reconnection models are compared with the measured distributions, and sensitivity to strength-parameter variations is mainly seen in the distribution of  $\sum_{n_{\text{ch}}} p_{\text{T}}$  in bins of  $n_{\text{ch}}$  for two of the five models. Variations of the PYTHIA 8 MPI model parameter  $p_{\text{T}0}^{\text{ref}}$  showed that a value lower than the one used in the A14 set of tuned parameters is suggested by the measurement. In the case of HERWIG 7.2 the newly introduced statistical colour-reconnection model agrees best for the  $n_{\text{ch}}$  distribution, while for the  $\sum_{n_{\text{ch}}} p_{\text{T}}$  distribution the baryonic CR model agrees best.

The effects of varying the colour reconnection models can be used to develop estimates of a colour reconnection uncertainty. The chosen observables are not only sensitive to parameters related to colour reconnection models, but also to the MPI parameters. Thus, these results could be used as input to future tuning of Monte Carlo generators for both colour reconnection and MPI parameters, which should be handled simultaneously.

## Appendices

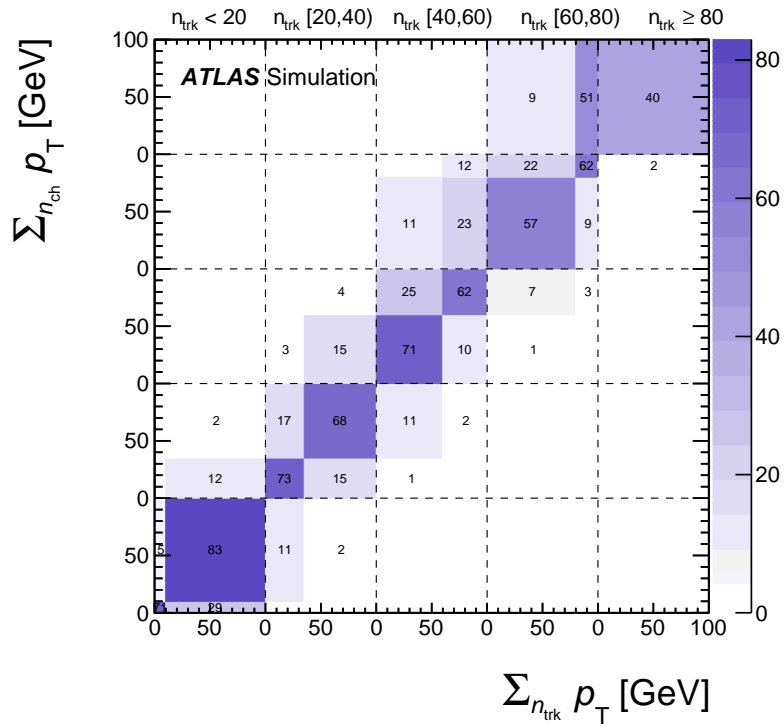
### A Migration matrices

Figure 15 shows the migration matrices used as input to the unfolding method for the observables  $n_{\text{ch}}$ ,  $\sum_{n_{\text{ch}}} p_{\text{T}}$ , and  $\sum_{n_{\text{ch}}} p_{\text{T}}$  in bins of  $n_{\text{ch}}$ . The corrected-level observable is shown on the  $x$ -axis, while the particle-level observable is shown on the  $y$ -axis. The values are normalised so that each row sums to unity.



(a)

(b)



(c)

Figure 15: Migration matrix of the observables (a)  $n_{\text{trk}}$ , (b)  $\sum n_{\text{trk}} p_T$  and (c)  $\sum n_{\text{trk}} p_T$  in bins of  $n_{\text{trk}}$ . The values are normalised per row and shown as percentages. The numerical value of the bin is displayed if it is  $> 1\%$ . Events beyond the  $x$ - or  $y$ -axis range are included in the last bin.

## B Absolute differential cross-sections

Absolute differential cross-sections are compared with predictions from various MC event generators and presented in Figures 16–19.

## C Additional material for the measured distributions

In addition to the quantitative comparisons of normalised differential cross-sections with predictions from different MC models, this appendix presents two alternative  $\chi^2$  scenarios:

- Using only uncertainties described in Section 8, i.e. modelling uncertainties are only included as non-closure uncertainties, while the extra theory uncertainties defined in Section 9 are omitted. This scenario is referred to as ‘Total’ in the following tables.
- Using the detector covariance matrix, and adding modelling uncertainties, scale variations in the matrix element and parton shower as well as the  $h_{\text{damp}}$  variation only to the diagonal elements of the covariance matrix. This scenario is referred to as ‘De-correlate modelling’ in the following tables.

The  $\chi^2$  values corresponding to these two scenarios for the normalised differential cross-sections are shown in Table 7, while the  $\chi^2$  values for the absolute differential cross-sections are shown in Table 8. In both tables, a clear reduction of the  $\chi^2$  values can be seen in the ‘De-correlate modelling’ scenario compared to the ‘Total’ scenario. This clearly shows that the modelling uncertainty correlations between the bins are the reason for the large  $\chi^2$  values in the case of the ‘Total’ scenario. The results for the two scenarios follow similar trends. The  $n_{\text{ch}}$  measured cross-section is approximately equally well described by PYTHIA 8.230 and HERWIG 7, while the  $\sum_{n_{\text{ch}}} p_{\text{T}}$  measured normalised and absolute differential cross-sections are in much better agreement with HERWIG 7. With increasing HERWIG 7 version number, the  $n_{\text{ch}}$  distribution agreement gets worse, while the  $\sum_{n_{\text{ch}}} p_{\text{T}}$  agreement with remains at a similar level. The distribution of  $\sum_{n_{\text{ch}}} p_{\text{T}}$  in bins of  $n_{\text{ch}}$  is described best by the three versions of HERWIG 7, not so well by PYTHIA 8.230, and worst by SHERPA 2. Overall, the best agreement is achieved using HERWIG 7.0.4.



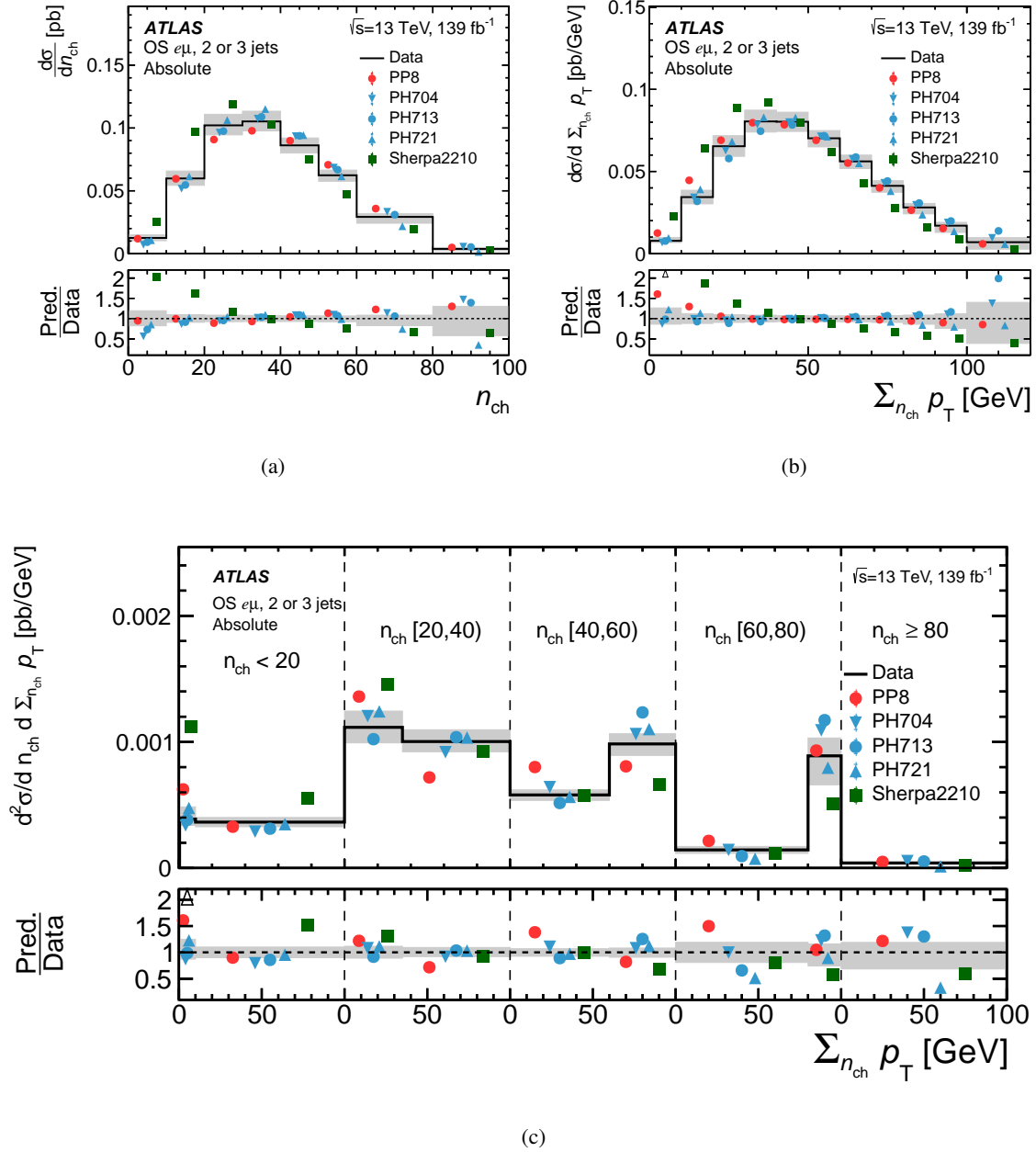


Figure 16: Absolute differential cross-section as a function of (a)  $n_{ch}$ , (b)  $\Sigma n_{ch} p_T$ , and (c)  $\Sigma n_{ch} p_T$  in bins of  $n_{ch}$ . Unfolded data are shown as the black line with the grey band corresponding to the total uncertainty in each bin. The results are compared with the predictions of different MC event generators. Events beyond the  $x$ -axis range are included in the last bin. The lower panels show the ratio of each prediction to data in each bin. The black triangular markers in the lower panels point to prediction-to-data ratio values which lie beyond the shown  $y$ -axis range.

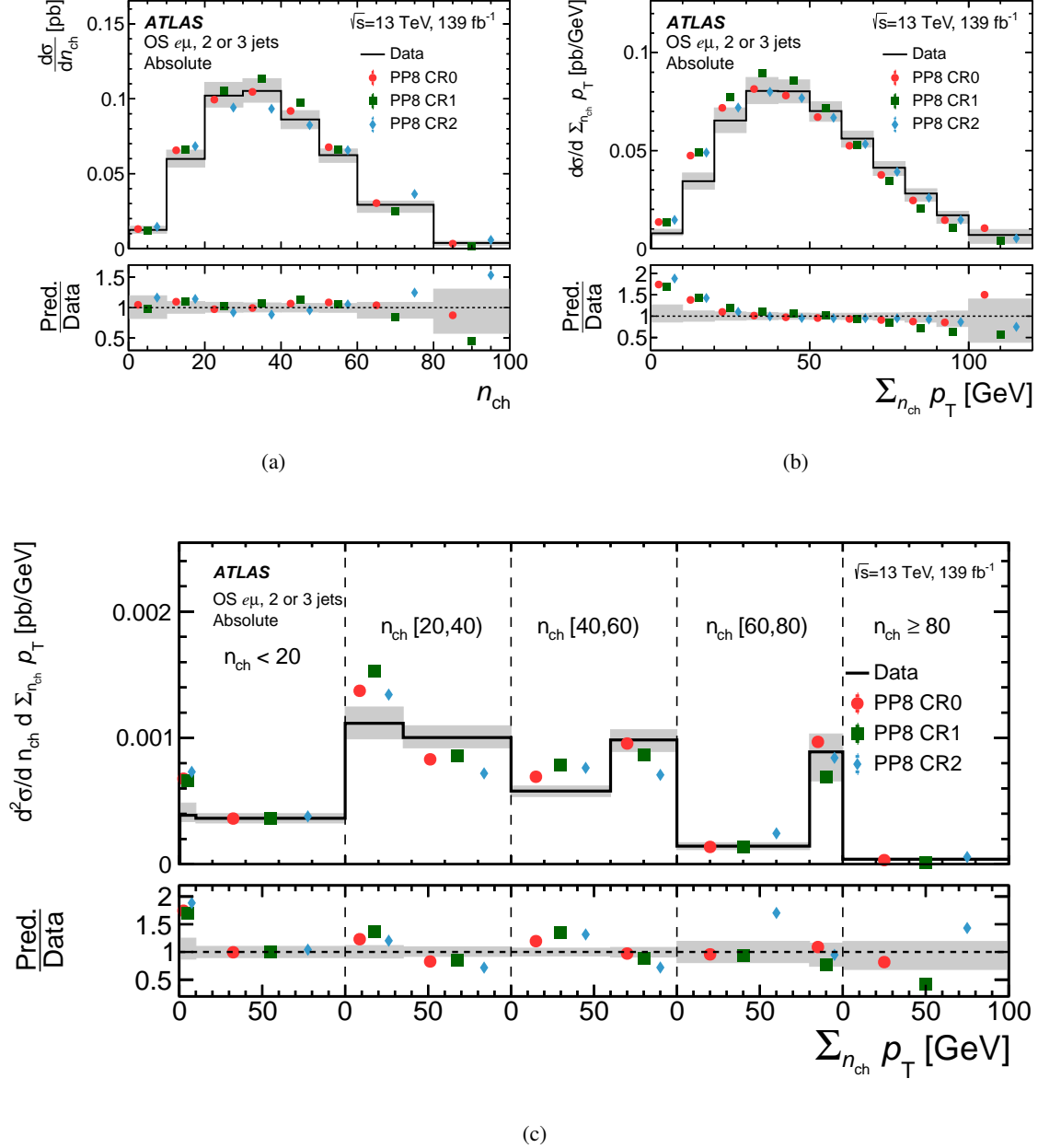
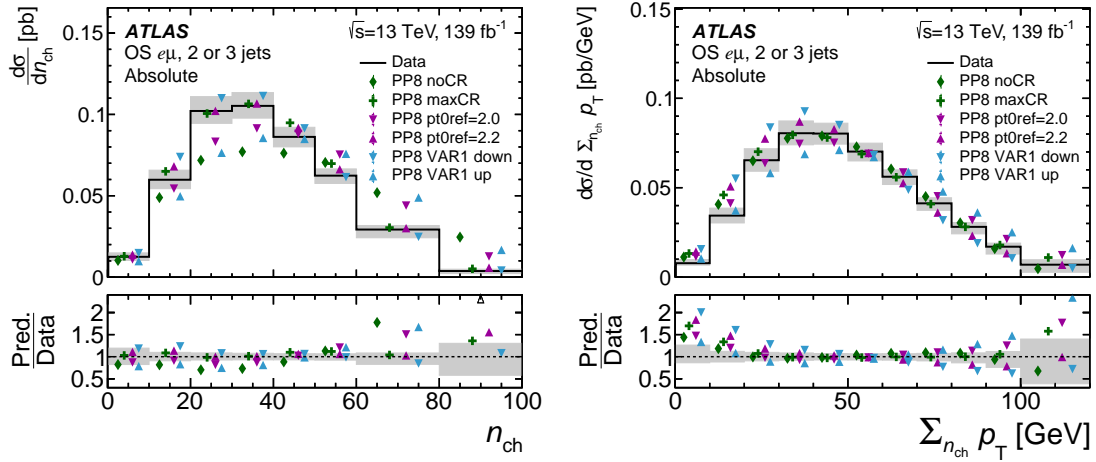
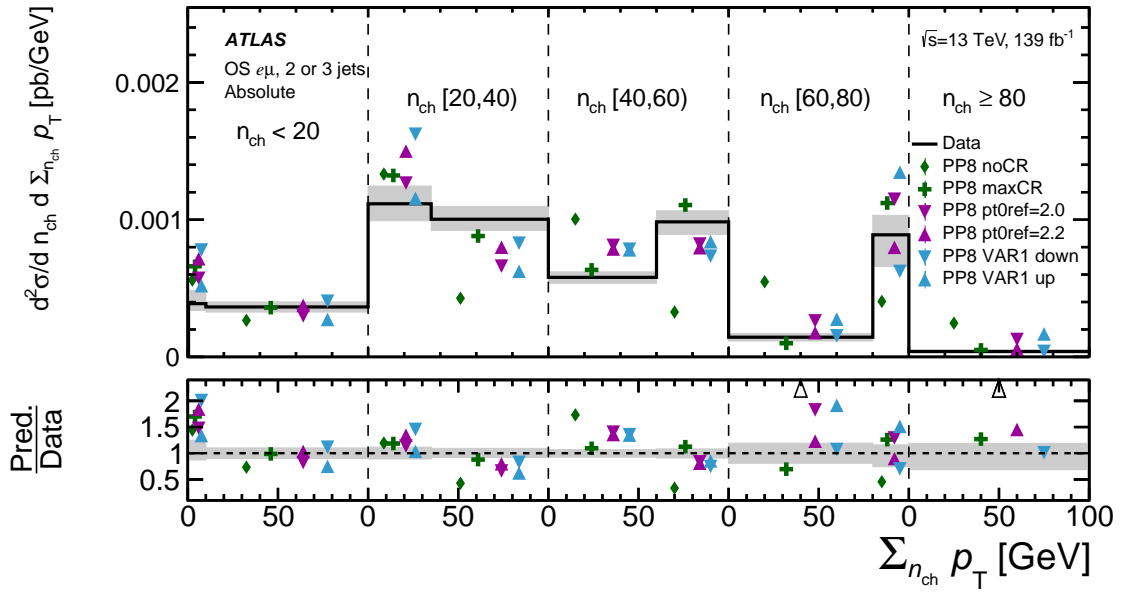


Figure 17: Absolute differential cross-section as a function of (a)  $n_{\text{ch}}$ , (b)  $\Sigma n_{\text{ch}} p_T$ , and (c)  $\Sigma n_{\text{ch}} p_T$  in bins of  $n_{\text{ch}}$ . Unfolded data are shown as the black line with the grey band corresponding to the total uncertainty in each bin. The results are compared with the predictions of different CR models in PYTHIA 8. Events beyond the  $x$ -axis range are included in the last bin. The lower panels show the ratio of each prediction to data in each bin.



(a)

(b)



(c)

Figure 18: Absolute differential cross-section as a function of (a)  $n_{\text{ch}}$ , (b)  $\sum n_{\text{ch}} p_T$ , and (c)  $\sum n_{\text{ch}} p_T$  in bins of  $n_{\text{ch}}$ . Unfolded data are shown as the black line with the grey band corresponding to the total uncertainty in each bin. The results are compared with the predictions of different CR and UE parameter variations in PYTHIA 8. Events beyond the  $x$ -axis range are included in the last bin. The lower panels show the ratio of each prediction to data in each bin. The black triangular markers in the lower panels point to prediction-to-data ratio values which lie beyond the shown  $y$ -axis range.

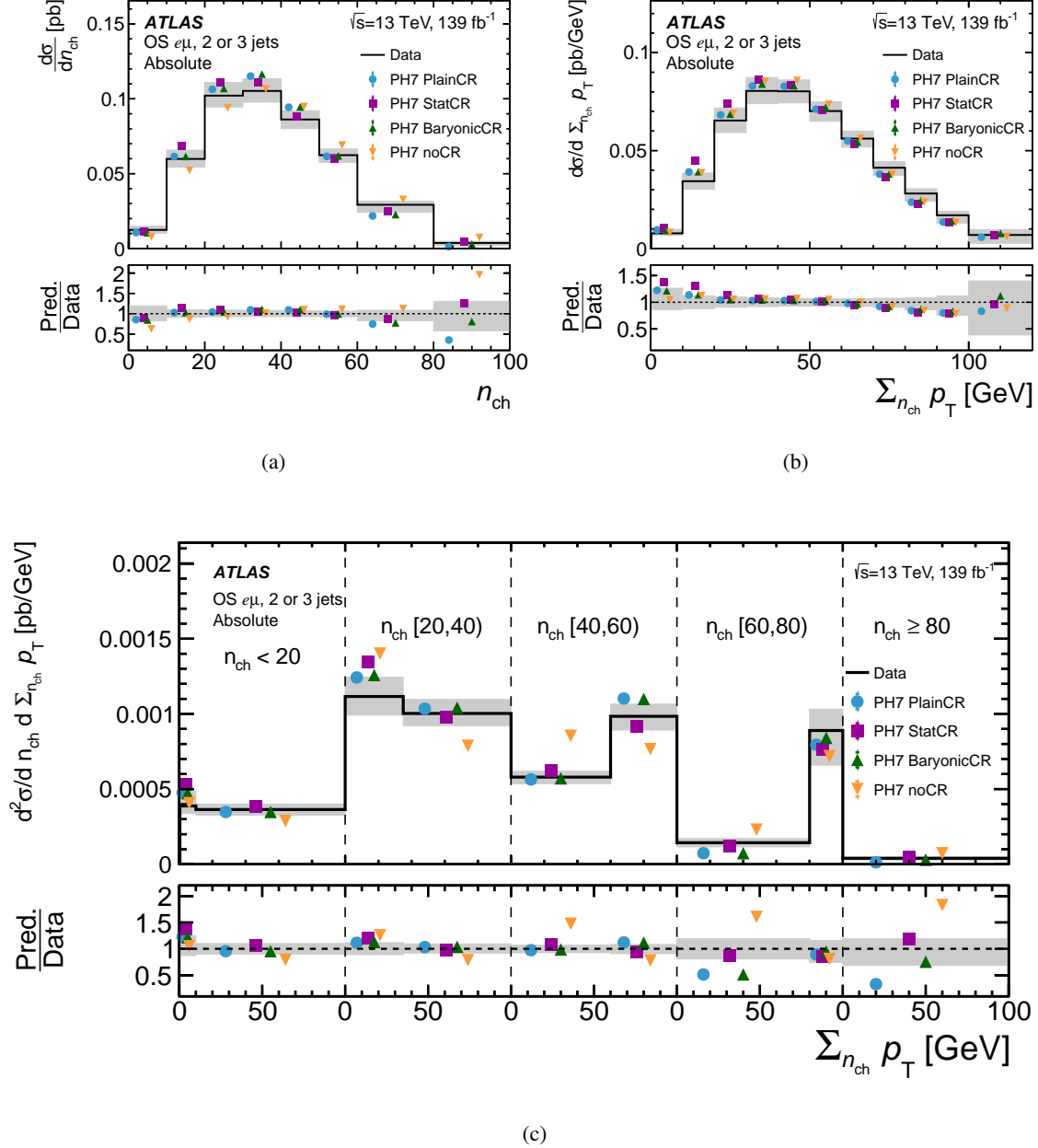


Figure 19: Absolute differential cross-section as a function of (a)  $n_{\text{ch}}$ , (b)  $\Sigma n_{\text{ch}} p_T$ , and (c)  $\Sigma n_{\text{ch}} p_T$  in bins of  $n_{\text{ch}}$ . Unfolded data are shown as the black line with the grey band corresponding to the total uncertainty in each bin. The results are compared with the predictions of different CR models in HERWIG 7.2. Events beyond the  $x$ -axis range are included in the last bin. The lower panels show the ratio of each prediction to data in each bin.

Table 7: The  $\chi^2$  and NDF for the measured normalised differential cross-sections obtained by comparing the different predictions with the unfolded data. The values corresponding to the ‘Total’ scenario are given first, followed by the ‘De-correlate modelling’ scenario.

Observables	$n_{\text{ch}}$	$\sum_{n_{\text{ch}}} p_{\text{T}}$	Global( $n_{\text{ch}}, \sum_{n_{\text{ch}}} p_{\text{T}}$ )	$\sum_{n_{\text{ch}}} p_{\text{T}}$ in bins of $n_{\text{ch}}$
NDF	7	10	17	8
Generator set-up	$\chi^2$			
<b>Total</b>				
POWHEG+PYTHIA 8.230	73	293	573	307
CR0	69	296	791	254
CR1	96	131	635	610
CR2	151	570	880	424
POWHEG+HERWIG 7.0.4	80	17	206	184
POWHEG+HERWIG 7.1.3	84	44	241	61
POWHEG+HERWIG 7.2.1	91	27	380	405
POWHEG+HERWIG Baryonic CR	107	22	288	118
POWHEG+HERWIG Stat CR	23	76	145	117
SHERPA 2.2.10	459	950	1409	1453
<b>De-correlate modelling</b>				
POWHEG+PYTHIA 8.230	24	18	63	155
CR0	8	37	53	53
CR1	18	23	49	107
CR2	47	40	106	125
POWHEG+HERWIG 7.0.4	19	2	21	28
POWHEG+HERWIG 7.1.3	14	11	20	27
POWHEG+HERWIG 7.2.1	32	5	37	19
POWHEG+HERWIG Baryonic CR	24	5	30	16
POWHEG+HERWIG Stat CR	7	13	18	10
SHERPA 2.2.10	148	155	250	137

Table 8: The  $\chi^2$  and NDF for the measured absolute differential cross-sections obtained by comparing the different predictions with the unfolded data. The nominal values are first shown followed by the ‘Total’ scenario and the ‘De-correlate modelling’ scenario.

Observables	$n_{\text{ch}}$	$\sum n_{\text{ch}} p_{\text{T}}$	Global( $n_{\text{ch}}, \sum n_{\text{ch}} p_{\text{T}}$ )	$\sum n_{\text{ch}} p_{\text{T}}$ in bins of $n_{\text{ch}}$
NDF	8	11	19	9
Generator set-up	$\chi^2$			
PowHEG+PYTHIA 8.230	88	121	486	270
CR0	88	143	738	153
CR1	176	77	702	216
CR2	96	209	468	279
PowHEG+HERWIG 7.0.4	56	22	162	45
PowHEG+HERWIG 7.1.3	80	44	198	54
PowHEG+HERWIG 7.2.1	128	33	342	99
PowHEG+HERWIG Baryonic CR	115	24	262	36
PowHEG +HERWIG Stat CR	30	55	142	54
SHERPA 2.2.10	144	275	414	144
Total				
PowHEG+PYTHIA 8.230	97	327	606	325
CR0	101	337	888	278
CR1	169	153	762	642
CR2	176	643	899	525
PowHEG+HERWIG 7.0.4	98	21	227	191
PowHEG+HERWIG 7.1.3	106	52	262	72
PowHEG+HERWIG 7.2.1	128	32	402	420
PowHEG+HERWIG Baryonic CR	139	25	307	125
PowHEG+HERWIG Stat CR	30	90	163	156
SHERPA 2.2.10	536	1093	1615	1624
De-correlate modelling				
PowHEG+PYTHIA 8.230	8	9	22	72
CR0	2	17	23	28
CR1	8	22	31	53
CR2	11	19	35	83
PowHEG+HERWIG 7.0.4	8	1	9	11
PowHEG+HERWIG 7.1.3	5	6	10	19
PowHEG+HERWIG 7.2.1	10	4	13	20
PowHEG+HERWIG Baryonic CR	7	10	11	15
PowHEG+HERWIG Stat CR	4	10	15	9
SHERPA 2.2.10	54	100	127	87

## References

- [1] Z. Nagy and D. E. Soper, *Parton showers with quantum interference: Leading color, spin averaged*, *JHEP* **03** (2008) 030, arXiv: [0801.1917 \[hep-ph\]](#).
- [2] A. Buckley et al., *General-purpose event generators for LHC physics*, *Phys. Rept.* **504** (2011) 145, arXiv: [1101.2599 \[hep-ph\]](#).
- [3] A. Ortiz Velasquez, P. Christiansen, E. Cuautle Flores, I. A. Maldonado Cervantes and G. Paic, *Color Reconnection and Flowlike Patterns in pp Collisions*, *Phys. Rev. Lett.* **111** (2013) 042001, arXiv: [1303.6326 \[hep-ph\]](#).
- [4] J. R. Christiansen and P. Z. Skands, *String formation beyond leading colour*, *JHEP* **08** (2015) 003, arXiv: [1505.01681 \[hep-ph\]](#).
- [5] T. Sjöstrand, ‘Colour reconnection and its effects on precise measurements at the LHC’, 2013, arXiv: [1310.8073 \[hep-ph\]](#).
- [6] T. Sjöstrand, S. Mrenna and P. Skands, *A brief introduction to PYTHIA 8.1*, *Comput. Phys. Commun.* **178** (2008) 852, arXiv: [0710.3820 \[hep-ph\]](#).
- [7] J. Bellm et al., *Herwig 7.0/Herwig++ 3.0 release note*, *Eur. Phys. J. C* **76** (2016) 196, arXiv: [1512.01178 \[hep-ph\]](#).
- [8] E. Bothmann et al., *Event generation with Sherpa 2.2*, *SciPost Phys.* **7** (2019) 034, arXiv: [1905.09127 \[hep-ph\]](#).
- [9] J. Bellm, C. B. Duncan, S. Gieseke, M. Myska and A. Siódmok, *Spacetime colour reconnection in Herwig 7*, *Eur. Phys. J. C* **79** (2019) 1003, arXiv: [1909.08850 \[hep-ph\]](#).
- [10] Particle Data Group, P. Zyla et al., *Review of Particle Physics*, *PTEP* **2020** (2020) 083C01.
- [11] Argyropoulos, Spyros and Sjöstrand, Torbjorn, *Effects of color reconnection on  $t\bar{t}$  final states at the LHC*, *JHEP* **11** (2014) 043, arXiv: [1407.6653 \[hep-ph\]](#).
- [12] ATLAS Collaboration, *Measurement of the top quark mass in the  $t\bar{t} \rightarrow \text{lepton}+\text{jets}$  channel from  $\sqrt{s} = 8\text{ TeV}$  ATLAS data and combination with previous results*, *Eur. Phys. J. C* **79** (2019) 290, arXiv: [1810.01772 \[hep-ex\]](#).
- [13] CMS Collaboration, *Measurement of the top quark mass in the all-jets final state at  $\sqrt{s} = 13\text{ TeV}$  and combination with the lepton+jets channel*, *Eur. Phys. J. C* **79** (2019) 313, arXiv: [1812.10534 \[hep-ex\]](#).
- [14] ATLAS Collaboration, *Measurement of charged-particle distributions sensitive to the underlying event in  $\sqrt{s} = 13\text{ TeV}$  proton–proton collisions with the ATLAS detector at the LHC*, *JHEP* **03** (2017) 157, arXiv: [1701.05390 \[hep-ex\]](#).
- [15] CMS Collaboration, *Measurement of the underlying event activity in inclusive Z boson production in proton–proton collisions at  $\sqrt{s} = 13\text{ TeV}$* , *JHEP* **07** (2018) 032, arXiv: [1711.04299 \[hep-ex\]](#).
- [16] ATLAS Collaboration, *Measurement of distributions sensitive to the underlying event in inclusive Z-boson production in pp collisions at  $\sqrt{s} = 7\text{ TeV}$  with the ATLAS detector*, *Eur. Phys. J. C* **74** (2014) 3195, arXiv: [1409.3433 \[hep-ex\]](#).
- [17] ATLAS Collaboration, *Measurement of distributions sensitive to the underlying event in inclusive Z boson production in pp collisions at  $\sqrt{s} = 13\text{ TeV}$  with the ATLAS detector*, *Eur. Phys. J. C* **79** (2019) 666, arXiv: [1905.09752 \[hep-ex\]](#).

- [18] CMS Collaboration, *Study of the underlying event in top quark pair production in pp collisions at 13 TeV*, *Eur. Phys. J. C* **79** (2019) 123, arXiv: [1807.02810 \[hep-ex\]](#).
- [19] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, *JINST* **3** (2008) S08003.
- [20] C. Bierlich et al., *A comprehensive guide to the physics and usage of PYTHIA 8.3*, (2022), arXiv: [2203.11601 \[hep-ph\]](#).
- [21] ATLAS Collaboration, *ATLAS Pythia 8 tunes to 7 TeV data*, ATL-PHYS-PUB-2014-021, 2014, URL: <https://cds.cern.ch/record/1966419>.
- [22] ATLAS Collaboration, *A study of different colour reconnection settings for Pythia8 generator using underlying event observables*, ATL-PHYS-PUB-2017-008, 2017, URL: <https://cds.cern.ch/record/2262253>.
- [23] CMS Collaboration, *CMS PYTHIA 8 colour reconnection tunes based on underlying-event data*, (2022), arXiv: [2205.02905 \[hep-ex\]](#).
- [24] B. R. Webber, *A QCD model for jet fragmentation including soft gluon interference*, *Nucl. Phys. B* **238** (1984) 492.
- [25] Gieseke, Stefan and Röhr, Christian and Siódmok, Andrzej, *Colour reconnections in Herwig++*, *Eur. Phys. J. C* **72** (2012) 2225, arXiv: [1206.0041 \[hep-ph\]](#).
- [26] J. Bellm, S. Gieseke and P. Kirchgaesser, *Improving the description of multiple interactions in Herwig*, *Eur. Phys. J. C* **80** (2020) 469, arXiv: [1911.13149 \[hep-ph\]](#).
- [27] S. Kirkpatrick, C. D. Gelatt Jr. and M. P. Vecchi, *Optimization by Simulated Annealing*, *Science* **220** (1983) 671.
- [28] D. Reichelt, P. Richardson and A. Siodmok, *Improving the simulation of quark and gluon jets with Herwig 7*, *Eur. Phys. J. C* **77** (2017) 876, arXiv: [1708.01491 \[hep-ph\]](#).
- [29] S. Gieseke, P. Kirchgaesser and S. Plätzer, *Baryon production from cluster hadronisation*, *Eur. Phys. J. C* **78** (2018) 99, arXiv: [1710.10906 \[hep-ph\]](#).
- [30] ATLAS Collaboration, *ATLAS Insertable B-Layer: Technical Design Report*, ATLAS-TDR-19; CERN-LHCC-2010-013, 2010, URL: <https://cds.cern.ch/record/1291633>, Addendum: ATLAS-TDR-19-ADD-1; CERN-LHCC-2012-009, 2012, URL: <https://cds.cern.ch/record/1451888>.
- [31] B. Abbott et al., *Production and integration of the ATLAS Insertable B-Layer*, *JINST* **13** (2018) T05008, arXiv: [1803.00844 \[physics.ins-det\]](#).
- [32] ATLAS Collaboration, *Performance of the ATLAS trigger system in 2015*, *Eur. Phys. J. C* **77** (2017) 317, arXiv: [1611.09661 \[hep-ex\]](#).
- [33] ATLAS Collaboration, *The ATLAS Collaboration Software and Firmware*, ATL-SOFT-PUB-2021-001, 2021, URL: <https://cds.cern.ch/record/2767187>.
- [34] ATLAS Collaboration, *ATLAS data quality operations and performance for 2015–2018 data-taking*, *JINST* **15** (2020) P04003, arXiv: [1911.04632 \[physics.ins-det\]](#).



- [35] ATLAS Collaboration, *Luminosity determination in pp collisions at  $\sqrt{s} = 13$  TeV using the ATLAS detector at the LHC*, ATLAS-CONF-2019-021, 2019, URL: <https://cds.cern.ch/record/2677054>.
- [36] ATLAS Collaboration, *Luminosity determination in pp collisions at  $\sqrt{s} = 8$  TeV using the ATLAS detector at the LHC*, *Eur. Phys. J. C* **76** (2016) 653, arXiv: [1608.03953](https://arxiv.org/abs/1608.03953) [[hep-ex](#)].
- [37] G. Avoni et al., *The new LUCID-2 detector for luminosity measurement and monitoring in ATLAS*, *JINST* **13** (2018) P07017.
- [38] ATLAS Collaboration, *Performance of electron and photon triggers in ATLAS during LHC Run 2*, *Eur. Phys. J. C* **80** (2020) 47, arXiv: [1909.00761](https://arxiv.org/abs/1909.00761) [[hep-ex](#)].
- [39] ATLAS Collaboration, *Performance of the ATLAS muon triggers in Run 2*, *JINST* **15** (2020) P09015, arXiv: [2004.13447](https://arxiv.org/abs/2004.13447) [[hep-ex](#)].
- [40] GEANT4 Collaboration, S. Agostinelli et al., *GEANT4 – a simulation toolkit*, *Nucl. Instrum. Meth. A* **506** (2003) 250.
- [41] ATLAS Collaboration, *The ATLAS Simulation Infrastructure*, *Eur. Phys. J. C* **70** (2010) 823, arXiv: [1005.4568](https://arxiv.org/abs/1005.4568) [[physics.ins-det](#)].
- [42] ATLAS Collaboration, *The simulation principle and performance of the ATLAS fast calorimeter simulation FastCaloSim*, ATL-PHYS-PUB-2010-013, 2010, URL: <https://cds.cern.ch/record/1300517>.
- [43] ATLAS Collaboration, *The Pythia 8 A3 tune description of ATLAS minimum bias and inelastic measurements incorporating the Donnachie–Landshoff diffractive model*, ATL-PHYS-PUB-2016-017, 2016, URL: <https://cds.cern.ch/record/2206965>.
- [44] R. D. Ball et al., *Parton distributions with LHC data*, *Nucl. Phys. B* **867** (2013) 244, arXiv: [1207.1303](https://arxiv.org/abs/1207.1303) [[hep-ph](#)].
- [45] P. Nason, *A new method for combining NLO QCD with shower Monte Carlo algorithms*, *JHEP* **11** (2004) 040, arXiv: [hep-ph/0409146](https://arxiv.org/abs/hep-ph/0409146).
- [46] S. Frixione, G. Ridolfi and P. Nason, *A positive-weight next-to-leading-order Monte Carlo for heavy flavour hadroproduction*, *JHEP* **09** (2007) 126, arXiv: [0707.3088](https://arxiv.org/abs/0707.3088) [[hep-ph](#)].
- [47] S. Frixione, P. Nason and C. Oleari, *Matching NLO QCD computations with parton shower simulations: the POWHEG method*, *JHEP* **11** (2007) 070, arXiv: [0709.2092](https://arxiv.org/abs/0709.2092) [[hep-ph](#)].
- [48] S. Alioli, P. Nason, C. Oleari and E. Re, *NLO single-top production matched with shower in POWHEG: s- and t-channel contributions*, *JHEP* **09** (2009) 111, arXiv: [0907.4076](https://arxiv.org/abs/0907.4076) [[hep-ph](#)], Erratum: *JHEP* **02** (2010) 011.
- [49] S. Alioli, P. Nason, C. Oleari and E. Re, *A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX*, *JHEP* **06** (2010) 043, arXiv: [1002.2581](https://arxiv.org/abs/1002.2581) [[hep-ph](#)].
- [50] E. Re, *Single-top Wt-channel production matched with parton showers using the POWHEG method*, *Eur. Phys. J. C* **71** (2011) 1547, arXiv: [1009.2450](https://arxiv.org/abs/1009.2450) [[hep-ph](#)].

- [51] R. Frederix, E. Re and P. Torrielli, *Single-top  $t$ -channel hadroproduction in the four-flavour scheme with POWHEG and aMC@NLO*, *JHEP* **09** (2012) 130, arXiv: [1207.5391 \[hep-ph\]](#).
- [52] R. D. Ball et al., *Parton distributions for the LHC run II*, *JHEP* **04** (2015) 040, arXiv: [1410.8849 \[hep-ph\]](#).
- [53] ATLAS Collaboration, *Studies on top-quark Monte Carlo modelling for Top2016*, ATL-PHYS-PUB-2016-020, 2016, URL: <https://cds.cern.ch/record/2216168>.
- [54] D. J. Lange, *The EvtGen particle decay simulation package*, *Nucl. Instrum. Meth. A* **462** (2001) 152.
- [55] S. Frixione, E. Laenen, P. Motylinski, C. White and B. R. Webber, *Single-top hadroproduction in association with a  $W$  boson*, *JHEP* **07** (2008) 029, arXiv: [0805.3067 \[hep-ph\]](#).
- [56] M. Czakon and A. Mitov, *Top++: A program for the calculation of the top-pair cross-section at hadron colliders*, *Comput. Phys. Commun.* **185** (2014) 2930, arXiv: [1112.5675 \[hep-ph\]](#).
- [57] M. Aliev et al., *HATHOR – HAdronic Top and Heavy quarks crOss section calculatoR*, *Comput. Phys. Commun.* **182** (2011) 1034, arXiv: [1007.1327 \[hep-ph\]](#).
- [58] P. Kant et al., *HatHor for single top-quark production: Updated predictions and uncertainty estimates for single top-quark production in hadronic collisions*, *Comput. Phys. Commun.* **191** (2015) 74, arXiv: [1406.4403 \[hep-ph\]](#).
- [59] 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 \[hep-ph\]](#).
- [60] D. de Florian et al., *Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector*, (2016), arXiv: [1610.07922 \[hep-ph\]](#).
- [61] T. Gleisberg and S. Höche, *Comix, a new matrix element generator*, *JHEP* **12** (2008) 039, arXiv: [0808.3674 \[hep-ph\]](#).
- [62] F. Buccioni et al., *OpenLoops 2*, *Eur. Phys. J. C* **79** (2019) 866, arXiv: [1907.13071 \[hep-ph\]](#).
- [63] F. Cascioli, P. Maierhöfer and S. Pozzorini, *Scattering Amplitudes with Open Loops*, *Phys. Rev. Lett.* **108** (2012) 111601, arXiv: [1111.5206 \[hep-ph\]](#).
- [64] A. Denner, S. Dittmaier and L. Hofer, *COLLIER: A fortran-based complex one-loop library in extended regularizations*, *Comput. Phys. Commun.* **212** (2017) 220, arXiv: [1604.06792 \[hep-ph\]](#).
- [65] S. Schumann and F. Krauss, *A parton shower algorithm based on Catani–Seymour dipole factorisation*, *JHEP* **03** (2008) 038, arXiv: [0709.1027 \[hep-ph\]](#).
- [66] J.-C. Winter, F. Krauss and G. Soff, *A modified cluster-hadronization model*, *Eur. Phys. J. C* **36** (2004) 381, arXiv: [hep-ph/0311085](#).
- [67] S. Höche, F. Krauss, M. Schönherr and F. Siegert, *A critical appraisal of NLO+PS matching methods*, *JHEP* **09** (2012) 049, arXiv: [1111.1220 \[hep-ph\]](#).

- [68] S. Catani, F. Krauss, B. R. Webber and R. Kuhn, *QCD Matrix Elements + Parton Showers*, [JHEP \*\*11\*\* \(2001\) 063](#), arXiv: [hep-ph/0109231](#).
- [69] S. Höche, F. Krauss, S. Schumann and F. Siegert, *QCD matrix elements and truncated showers*, [JHEP \*\*05\*\* \(2009\) 053](#), arXiv: [0903.1219 \[hep-ph\]](#).
- [70] S. Höche, F. Krauss, M. Schönherr and F. Siegert, *QCD matrix elements + parton showers. The NLO case*, [JHEP \*\*04\*\* \(2013\) 027](#), arXiv: [1207.5030 \[hep-ph\]](#).
- [71] C. Anastasiou, L. Dixon, K. Melnikov and F. Petriello, *High-precision QCD at hadron colliders: Electroweak gauge boson rapidity distributions at next-to-next-to leading order*, [Phys. Rev. D \*\*69\*\* \(2004\) 094008](#), arXiv: [hep-ph/0312266](#).
- [72] K. Melnikov and F. Petriello, *Electroweak gauge boson production at hadron colliders through  $\mathcal{O}(\alpha_s^2)$* , [Phys. Rev. D \*\*74\*\* \(2006\) 114017](#), arXiv: [hep-ph/0609070](#).
- [73] ATLAS Collaboration, *Vertex Reconstruction Performance of the ATLAS Detector at  $\sqrt{s} = 13$  TeV*, ATL-PHYS-PUB-2015-026, 2015, URL: <https://cds.cern.ch/record/2037717>.
- [74] ATLAS Collaboration, *Performance of the ATLAS track reconstruction algorithms in dense environments in LHC Run 2*, [Eur. Phys. J. C \*\*77\*\* \(2017\) 673](#), arXiv: [1704.07983 \[hep-ex\]](#).
- [75] R. Frühwirth, *Application of Kalman filtering to track and vertex fitting*, [Nucl. Instrum. Meth. A \*\*262\*\* \(1987\) 444](#).
- [76] ATLAS Collaboration, *Early Inner Detector Tracking Performance in the 2015 Data at  $\sqrt{s} = 13$  TeV*, ATL-PHYS-PUB-2015-051, 2015, URL: <https://cds.cern.ch/record/2110140>.
- [77] ATLAS Collaboration, *Electron and photon performance measurements with the ATLAS detector using the 2015–2017 LHC proton–proton collision data*, [JINST \*\*14\*\* \(2019\) P12006](#), arXiv: [1908.00005 \[hep-ex\]](#).
- [78] ATLAS Collaboration, *Muon reconstruction performance of the ATLAS detector in proton–proton collision data at  $\sqrt{s} = 13$  TeV*, [Eur. Phys. J. C \*\*76\*\* \(2016\) 292](#), arXiv: [1603.05598 \[hep-ex\]](#).
- [79] ATLAS Collaboration, *Muon reconstruction and identification efficiency in ATLAS using the full Run 2 pp collision data set at  $\sqrt{s} = 13$  TeV*, [Eur. Phys. J. C \*\*81\*\* \(2021\) 578](#), arXiv: [2012.00578 \[hep-ex\]](#).
- [80] ATLAS Collaboration, *Jet reconstruction and performance using particle flow with the ATLAS Detector*, [Eur. Phys. J. C \*\*77\*\* \(2017\) 466](#), arXiv: [1703.10485 \[hep-ex\]](#).
- [81] M. Cacciari, G. P. Salam and G. Soyez, *The anti- $k_t$  jet clustering algorithm*, [JHEP \*\*04\*\* \(2008\) 063](#), arXiv: [0802.1189 \[hep-ph\]](#).
- [82] M. Cacciari, G. P. Salam and G. Soyez, *FastJet user manual*, [Eur. Phys. J. C \*\*72\*\* \(2012\) 1896](#), arXiv: [1111.6097 \[hep-ph\]](#).
- [83] ATLAS Collaboration, *Jet energy scale and resolution measured in proton–proton collisions at  $\sqrt{s} = 13$  TeV with the ATLAS detector*, [Eur. Phys. J. C \*\*81\*\* \(2020\) 689](#), arXiv: [2007.02645 \[hep-ex\]](#).

- [84] ATLAS Collaboration, *Performance of pile-up mitigation techniques for jets in pp collisions at  $\sqrt{s} = 8$  TeV using the ATLAS detector*, *Eur. Phys. J. C* **76** (2016) 581, arXiv: [1510.03823 \[hep-ex\]](#).
- [85] ATLAS Collaboration, *ATLAS b-jet identification performance and efficiency measurement with  $t\bar{t}$  events in pp collisions at  $\sqrt{s} = 13$  TeV*, *Eur. Phys. J. C* **79** (2019) 970, arXiv: [1907.05120 \[hep-ex\]](#).
- [86] ATLAS Collaboration, *Performance of b-jet identification in the ATLAS experiment*, *JINST* **11** (2016) P04008, arXiv: [1512.01094 \[hep-ex\]](#).
- [87] M. Cacciari, G. P. Salam and G. Soyez, *The Catchment Area of Jets*, *JHEP* **04** (2008) 005, arXiv: [0802.1188 \[hep-ph\]](#).
- [88] ATLAS Collaboration, *Measurement of the  $t\bar{t}$  production cross-section and lepton differential distributions in  $e\mu$  dilepton events from pp collisions at  $\sqrt{s} = 13$  TeV with the ATLAS detector*, *Eur. Phys. J. C* **80** (2020) 528, arXiv: [1910.08819 \[hep-ex\]](#).
- [89] ATLAS Collaboration, *Measurements of top-quark pair differential and double-differential cross-sections in the  $\ell$ +jets channel with pp collisions at  $\sqrt{s} = 13$  TeV using the ATLAS detector*, *Eur. Phys. J. C* **79** (2019) 1028, arXiv: [1908.07305 \[hep-ex\]](#),  
Erratum: *Eur. Phys. J. C* **80** (2020) 1092.
- [90] ATLAS Collaboration, *Differential  $t\bar{t}$  cross-section measurements using boosted top quarks in the all-hadronic final state with  $139\text{fb}^{-1}$  of ATLAS data*, (2022), arXiv: [2205.02817 \[hep-ex\]](#).
- [91] ATLAS Collaboration, *Measurements of top-quark pair differential cross-sections in the lepton+jets channel in pp collisions at  $\sqrt{s} = 8$  TeV using the ATLAS detector*, *Eur. Phys. J. C* **76** (2016) 538, arXiv: [1511.04716 \[hep-ex\]](#).
- [92] CMS Collaboration, *Inclusive and differential cross section measurements of single top quark production in association with a Z boson in proton–proton collisions at  $\sqrt{s} = 13$  TeV*, *JHEP* **02** (2021) 107, arXiv: [2111.02860 \[hep-ex\]](#).
- [93] CMS Collaboration, *Measurement of differential  $t\bar{t}$  production cross sections using top quarks at large transverse momenta in pp collisions at  $\sqrt{s} = 13$  TeV*, *Phys. Rev. D* **103** (2021) 052008, arXiv: [2008.07860 \[hep-ex\]](#).
- [94] CMS Collaboration, *Measurements of  $t\bar{t}$  differential cross sections in proton–proton collisions at  $\sqrt{s} = 13$  TeV using events containing two leptons*, *JHEP* **02** (2019) 149, arXiv: [1811.06625 \[hep-ex\]](#).
- [95] CMS Collaboration, *Measurement of the differential cross section for top quark pair production in pp collisions at  $\sqrt{s} = 8$  TeV*, *Eur. Phys. J. C* **75** (2015) 542, arXiv: [1505.04480 \[hep-ex\]](#).
- [96] S. Catani, S. Devoto, M. Grazzini, S. Kallweit and J. Mazzitelli, *Top-quark pair production at the LHC: Fully differential QCD predictions at NNLO*, *JHEP* **07** (2019) 100, arXiv: [1906.06535 \[hep-ph\]](#).
- [97] M. Czakon et al., *Top-pair production at the LHC through NNLO QCD and NLO EW*, *JHEP* **10** (2017) 186, arXiv: [1705.04105 \[hep-ph\]](#).
- [98] M. Czakon, D. Heymes and A. Mitov, *Dynamical scales for multi-TeV top-pair production at the LHC*, *JHEP* **04** (2017) 071, arXiv: [1606.03350 \[hep-ph\]](#).

- [99] M. Czakon, D. Heymes and A. Mitov,  
*High-Precision Differential Predictions for Top-Quark Pairs at the LHC*,  
*Phys. Rev. Lett.* **116** (2016) 082003, arXiv: [1511.00549 \[hep-ph\]](#).
- [100] M. Czakon, P. Fiedler and A. Mitov,  
*Total Top-Quark Pair-Production Cross Section at Hadron Colliders Through  $O(\alpha_S^4)$* ,  
*Phys. Rev. Lett.* **110** (2013) 252004, arXiv: [1303.6254 \[hep-ph\]](#).
- [101] G. D’Agostini, *A multidimensional unfolding method based on Bayes’ theorem*,  
*Nucl. Instrum. Meth. A* **362** (1995) 487, ISSN: 0168-9002.
- [102] T. Auye, ‘Unfolding algorithms and tests using RooUnfold’,  
*Proceedings, 2011 Workshop on Statistical Issues Related to Discovery Claims in Search Experiments and Unfolding (PHYSTAT 2011)* (CERN, Geneva, Switzerland, 17th–20th Jan. 2011) 313, arXiv: [1105.1160 \[physics.data-an\]](#).
- [103] ATLAS Collaboration, *Alignment of the ATLAS Inner Detector in Run-2*,  
*Eur. Phys. J. C* **80** (2020) 1194, arXiv: [2007.07624 \[hep-ex\]](#).
- [104] ATLAS Collaboration, *Study of the material of the ATLAS inner detector for Run 2 of the LHC*,  
*JINST* **12** (2017) P12009, arXiv: [1707.02826 \[hep-ex\]](#).
- [105] ATLAS Collaboration, *Electron reconstruction and identification in the ATLAS experiment using the 2015 and 2016 LHC proton–proton collision data at  $\sqrt{s} = 13$  TeV*,  
*Eur. Phys. J. C* **79** (2019) 639, arXiv: [1902.04655 \[hep-ex\]](#).
- [106] L. Harland-Lang, A. Martin, P. Motylinski and R. Thorne,  
*Parton distributions in the LHC era: MMHT 2014 PDFs*, *Eur. Phys. J. C* **75** (2015) 204,  
arXiv: [1412.3989 \[hep-ph\]](#).
- [107] J. Butterworth et al., *PDF4LHC recommendations for LHC Run II*, *J. Phys. G* **43** (2016) 023001,  
arXiv: [1510.03865 \[hep-ph\]](#).