Optimisation of large-radius jet reconstruction for the ATLAS detector in 13 TeV proton-proton collisions

ATLAS Collaboration

DOI
10.1140/epjc/s10052-021-09054-3

Publication date
2021

Document Version
Final published version

Published in
European Physical Journal C

License
CC BY

Citation for published version (APA):
ATLAS Collaboration (2021). Optimisation of large-radius jet reconstruction for the ATLAS detector in 13 TeV proton-proton collisions. European Physical Journal C, 81, [334].
https://doi.org/10.1140/epjc/s10052-021-09054-3
Optimisation of large-radius jet reconstruction for the ATLAS detector in 13 TeV proton–proton collisions

ATLAS Collaboration*  
CERN, 1211 Geneva 23, Switzerland

Received: 12 September 2020 / Accepted: 16 March 2021 / Published online: 19 April 2021  
© CERN for the benefit of the ATLAS collaboration 2021

Abstract  Jet substructure has provided new opportunities for searches and measurements at the LHC, and has seen continuous development since the optimization of the large-radius jet definition used by ATLAS was performed during Run 1. A range of new inputs to jet reconstruction, pile-up mitigation techniques and jet grooming algorithms motivate an optimisation of large-radius jet reconstruction for ATLAS. In this paper, this optimisation procedure is presented, and the performance of a wide range of large-radius jet definitions is compared. The relative performance of these jet definitions is assessed using metrics such as their pileup stability, ability to identify hadronically decaying W bosons and top quarks with large transverse momenta. A new type of jet input object, called a ‘unified flow object’ is introduced which combines calorimeter- and inner-detector-based signals in order to achieve optimal performance across a wide kinematic range. Large-radius jet definitions are identified which significantly improve on the current ATLAS baseline definition, and their modelling is studied using $pp$ collisions recorded by the ATLAS detector at $\sqrt{s} = 13$ TeV during 2017.

Contents

1 Introduction .................................. 1  
2 The ATLAS detector, data and simulated events .... 3 
3 Objects and algorithms ....................... 4  
   3.1 Jet input objects ............................ 4  
   3.1.1 Stable generator-level particles .......... 4  
   3.1.2 Inner detector tracks .................... 4  
   3.1.3 Topological clusters .................... 4 
   3.1.4 Particle-flow objects (PFOs) ............ 5  
   3.1.5 Track-CaloClusters (TCCs) ............. 6  
3.2 Jet-input-level pile-up mitigation algorithms .... 6  
   3.2.1 Constituent subtraction (CS) .......... 6  
   3.2.2 SoftKiller (SK) ....................... 6  
   3.2.3 Pile-up per particle identification (PUPPI) 7  
3.3 Grooming algorithms ........................ 7  
   3.3.1 Trimming .............................. 7  
   3.3.2 Pruning ............................... 7  
   3.3.3 Soft-drop (SD) ....................... 8  
   3.3.4 Recursive soft-drop (RSD) and bottom-up soft-drop (BUSD) 8  
4 Performance metrics ......................... 8  
   4.1 Tagging performance ...................... 9  
   4.2 Pile-up stability .......................... 12  
   4.2.1 Pile-up stability of the $W$ boson jet mass peak position ....... 12  
   4.2.2 Pile-up stability of a simple tagger .......... 12  
4.3 Topological sensitivity ..................... 12  
5 Unified flow objects (UFOs) .................. 13  
6 Performance survey ........................... 16  
   6.1 Tagging performance ...................... 16 
   6.2 Pile-up stability .......................... 17 
6.3 Topological sensitivity ..................... 18  
7 Comparison of calibrated jet definitions .......... 19  
   7.1 Simulation-based jet energy and mass scale calibrations .......... 20  
   7.2 Comparison of calibrated jet definition performance .... 23  
   7.2.1 Jet mass and $p_T$ resolution .......... 23  
   7.2.2 Jet mass + JSS tagging performance .... 26  
7.3 Data-to-simulation comparisons ............... 26  
8 Concluding remarks .......................... 30  
References .................................... 31

1 Introduction

High-energy particle collisions such as those produced in the Large Hadron Collider (LHC) at CERN can result in the production of massive particles (e.g. $W/Z/H$ bosons and top quarks) with large Lorentz boosts. When such particles decay, their decay products become collimated, or ‘boosted’, in the direction of the progenitor particle. For massive particles that are sufficiently boosted, it is advantageous to reconstruct their hadronic decay products as a single large-
radius (large-$R$) jet. Such large-$R$ jets capture a characteristic, multi-pronged jet substructure from the two-body or three-body decays of hadronically decaying $W$, $Z$ and $H$ bosons and top quarks, which is distinct from the radiation pattern of a light-quark- or gluon-initiated jet.

The substructure of boosted hadronic systems is complicated by the presence of soft radiation from several sources, which degrades performance when reconstructing jet substructure observables. In particular, soft radiation from the underlying event and uncorrelated radiation from additional $pp$ interactions concurrent with the hard-scattering event of interest (pile-up interactions) can degrade the jet mass resolution and other jet substructure quantities, which are critical to boosted object identification. These effects are amplified by the use of a large radius for jet reconstruction [25–28], which incorporates more uncorrelated energy. During Run 1, the average number of pile-up interactions per LHC bunch crossing was roughly 20. This number increased to $\sim 34$ in the Run 2 dataset, although some events during this period were recorded with up to 70 pile-up interactions. The average number of pile-up collisions is expected to increase further during Run 3 and will reach $\sim 200$ pile-up interactions during high-luminosity LHC operations [29]. As experimental conditions become more challenging, the choices made when reconstructing large-$R$ jets will need to evolve to maintain optimal performance.

There is no single way to reconstruct a jet, and several choices must be made at the level of a physics analysis to define the jets which will be used. Jets at the LHC are typically reconstructed from some set of input objects (‘jet inputs’, or simply ‘inputs’ throughout) using a sequential recombination algorithm with a user-specified radius parameter ($R$). Once a jet input type is chosen, it may be preprocessed before jet reconstruction, for example, to mitigate the effects of pile-up. After jet reconstruction, a grooming algorithm may be applied to the jets which preferentially removes soft and/or wide-angled radiation from the reconstructed jet, to further suppress contributions from pile-up and the underlying event and to enhance the resolution of the jet mass and other substructure observables.

Large-$R$ jets are typically reconstructed by ATLAS using the anti-$k_t$ algorithm [30] and a radius parameter $R = 1.0$. The choice of recombination scheme and radius parameter has been studied previously [31], and is not revisited in these studies. ATLAS large-$R$ jet reconstruction has so-far been based on topological cluster inputs reconstructed only using calorimeter-based energy measurements. These clusters provide excellent energy resolution, but do not accurately represent the positions of individual particles within jets with large transverse momentum ($p_T$), particularly in areas where the energy density is large or the calorimeter granularity is coarse. This can result in degraded performance when the resolution of individual particles becomes relevant, for instance, when reconstructing the mass of showers which are so collimated that they are not spatially resolved by the ATLAS calorimeter’s granularity. In order to better reconstruct the angular distributions of charged particles within jets, several particle-flow (PFlow) algorithms which were developed and commissioned by ATLAS during Run 2 are considered. These include a PFlow implementation designed to improve $R = 0.4$ jet performance at low $p_T$ [32], and a variant designed to reconstruct jet substructure at the highest transverse momenta, called Track-CaloClusters (TCCs) [7,33]. In this work, a union of PFlow and TCCs called ‘Unified Flow Objects’ (UFOs) is established to provide optimal performance across a wider kinematic range than is possible with either particle-flow objects (PFOs) or TCCs alone, which are each found to perform well in distinct kinematic regions. Jet inputs may also be preprocessed using one or several of the many input-object-level pile-up mitigation techniques which have been developed, such as constituent subtraction [34,35], Voronoi subtraction [36], SoftKiller [37], and pile-up per particle identification (PUPPI) [38]. Various input types and pile-up mitigation algorithms can be combined to create pile-up-robust inputs to jet reconstruction, adding additional complexity to the search for optimal performance.

Grooming algorithms are another tool which may be used to remove undesirable radiation from jets after they have been reconstructed. The performance of several grooming algorithms was studied by ATLAS in detail using Run 1 data [39] and during preparations for Run 2 [40], including the jet trimming [41], pruning [42], and mass drop filtering [43] algorithms. Based on these studies, large-$R$ jets groomed with the trimming algorithm using parameter choices of $R_{\text{sub}} = 0.2$ and $f_{\text{cut}} = 5\%$ were found to be optimal for ATLAS with Run 2 conditions. Since the completion of these studies, several additional jet grooming algorithms have been proposed, including the modified mass drop (mMDT) [44] and soft-drop (SD) [45] algorithms, and their recent extensions: bottom-up soft-drop (BUSD) and recursive soft-drop (RSD) [46].

The development of new input objects, pile-up mitigation techniques and jet grooming algorithms by the experimental and phenomenological communities motivates a thorough reoptimisation of the large-$R$ jet definition used by ATLAS. In this paper, the jet tagging and substructure performance of 171 distinct combinations of the different jet inputs, pile-up mitigation techniques and grooming algorithms is evaluated using Run 2 conditions. The performance of different jet definitions is compared in the context of several metrics, which quantify their tagging performance, their pile-up stability,
and the sensitivity of their mass response to different jet substructure topologies. The performance in data is also studied to ensure the validity of the conclusions from the Monte Carlo studies.

The remaining sections of this document are structured as follows. The ATLAS detector is described in Sect. 2, along with aspects of the 2017 pp dataset and details of the simulated events used to perform these studies. An overview of the jet reconstruction techniques surveyed by these studies is provided in Sect. 3. Several metrics are used to determine the optimal jet definition, as well as to understand the behaviour of individual algorithms. Due to the large number of possible large-$$R$$ jet definitions, a two-stage optimisation is performed to determine which of these exhibit the best performance. In the first stage, presented in Sect. 4, the metrics which will be used to evaluate the relative performance of all jet definitions are established by studying the performance of a limited set of jet definitions. The observations made from these comparisons motivate an union of the existing particle-flow and TCC input objects; this new input object type is presented in Sect. 5. The results of the complete survey of jet definitions are presented in Sect. 6. UFO-based definitions which perform consistently well are selected for further study. This smaller list of jet definitions, each of which improves on the current ATLAS baseline large-$$R$$ jet definition, is calibrated using simulated events, and a more detailed comparison of their performance in terms of their tagging performance and jet $$p_T$$ and mass resolutions as well as their performance in data is made in Sect. 7. In an appendix, more details of the interaction between pile-up interactions and topological cluster formation are provided.

2 The ATLAS detector, data and simulated events

The ATLAS detector [47–49] consists of three principal subsystems. The inner detector (ID) provides tracking of charged particles within $$|\eta| < 2.5$$ using silicon pixel and microstrip detectors, as well as a transition radiation tracker which provides a large number of hits in the ID’s outermost layers in addition to particle identification information. This subsystem is immersed in an axial magnetic field generated by a 2 T solenoid. A sampling calorimeter surrounds the ID and barrel solenoid, providing energy measurements of electromagnetically and hadronically interacting particles within $$|\eta| < 4.9$$, and is followed by a muon spectrometer.

The electromagnetic showers of electrons and photons are measured with a high-granularity liquid argon (LAr) calorimeter, consisting of a barrel module within $$|\eta| < 1.475$$ and two endcaps from 1.365 < $$|\eta| < 3.2$$. Hadronic showers are measured using a steel/scintillator tile calorimeter within $$|\eta| < 1.7$$ and with a pair of LAr/copper endcaps within 1.5 < $$|\eta| < 3.2$$. In the forward region, a LAr/copper and LAr/tungsten forward calorimeter measures showers of both kinds within 3.2 < $$|\eta| < 4.9$$.

The muon spectrometer is based one barrel and two endcap superconducting toroidal magnets. Precision chambers provide measurements for all muons within $$|\eta| < 2.7$$, and separate trigger chambers allow the online selection of events with muons within $$|\eta| < 2.4$$.

As writing events to disk at the nominal LHC collision rate of 40 MHz is currently unfeasible, a two-level trigger system is used to select events for analysis. The hardware-based Level-1 trigger accepts events at a rate of ~100 kHz using a subset of available detector information. The software-based High-Level Trigger then reduces the event rate to ~1 kHz, which is retained for further analysis.

Studies presented in this paper utilise a dataset of proton–proton collisions delivered by the LHC in 2017 with centre-of-mass-energy $$\sqrt{s} = 13$$ TeV and collected with the ATLAS detector. Data containing high-$$p_T$$ dijet events were selected using a single-jet trigger, and the leading anti-$$k_t$$, $$R = 1.0$$ jet is required to have $$p_T$$ above 600 GeV. All data are required to meet standard ATLAS quality criteria [50]; data taken during periods when detector subsystems were not functional, which contain significant contamination from detector noise, or where there were detector read-out problems are discarded. The resulting dataset has an integrated luminosity of 44.3 fb$$^{-1}$$ and an associated luminosity uncertainty of 2.4% [51], obtained using the LUCID-2 detector [52] for the primary luminosity measurements.

The simulated event samples used to perform these studies were generated using PYTHIA 8.186 [53,54] with the NNPDF2.3 LO [55] set of parton distribution functions (PDF), a $$p_T$$-ordered parton shower, Lund string hadronisation [56,57], and the A14 set of tuned parameters (tune) [58]. These samples provide ‘background’ jets which originate from high-energy quark and gluon scattering (using a 2 → 2 matrix element), and ‘signal’ jets originating from high-$$p_T$$ W boson and top quark decays across a wide kinematic range. The signal W jets were produced using a BSM spin-1 W’ → WZ → qqqq model including only hadronic W and Z decays. The signal top quark jets are taken from a BSM Z’ → tt model, where the top quarks may decay either hadronically or semileptonically. In order to remove dependence on the specific BSM physics models used to generate these jets, the $$p_T$$ spectrum of signal jets is always reweighted to match that of background jets [59]. Strcightforward particle-level containment definitions are used to ensure
that the signal jets provide samples of two- and three-pronged jet topologies: the decay partons of the W boson or top quark are required to be within $\Delta R = 0.75$ of the particle-level jet axis. Top jets containing leptonic W boson decays are rejected using particle-level information.

All simulated events were passed through the complete ATLAS detector simulation \cite{60} based on GEANT4 \cite{61} using the FTFP_BERT_ATL model \cite{60}. The effect of pile-up was modelled by overlaying the hard-scatter event with minimum-bias $pp$ collisions generated by PYTHIA 8.210 with the A3 tune \cite{62} and the NNPDF2.3 LO PDF set. The number of pile-up vertices was reweighted to match the data events, which have an average of 38 simultaneous interactions per bunch crossing in the 2017 dataset. Pile-up events are overlaid such that each subdetector reconstructs the effect of signals from adjacent bunch crossings (‘out-of-time’ pile-up) as well as those from the same bunch crossing as the hard-scatter event (‘in-time’ pile-up) \cite{63}.

### 3 Objects and algorithms

This section provides a brief overview of different jet input object, pile-up mitigation and grooming options. All jets discussed in these studies are reconstructed using the anti-$k_t$ algorithm as implemented in \textsc{FastJet} \cite{64} with radius parameter $R = 1.0$. All jets used in these results are required to have a minimum $p_T$ of 300 GeV, and to be within $\eta < 1.2$.

The complete set of jet input object types, pile-up mitigation and grooming algorithms surveyed is summarised in Table 1. In some cases, additional algorithms or settings were studied but were not found to produce results which differed significantly from those presented here. Notes have been made in Sect. 4 when appropriate regarding these omitted jet definitions, and they are indicated in Table 1 by an asterisk (*).

#### 3.1 Jet input objects

##### 3.1.1 Stable generator-level particles

Particle-level jets, or ‘truth jets’, are reconstructed in simulated events at generator level. All detector-stable particles from the hard-scattering process with a lifetime $\tau$ in the laboratory frame such that $c\tau > 10$ mm are used. Particles that are expected to leave only negligible energy depositions in the calorimeter, i.e. muons and neutrinos, are excluded.

Ungroomed particle-level jets are used as the reference objects for selections throughout these studies in order to ensure that the same set of reconstructed jets are selected for comparison, regardless of the jet input objects used in reconstruction or grooming algorithm applied. In studies of simulated jets, unless otherwise specified, ungroomed particle-level jets are geometrically matched ($\Delta R < 0.75$) to ungroomed reconstructed jets, and kinematic selections are applied to the ungroomed particle-level jet four-vector.

Particle-level jets are also taken as the reference for simulation-based ATLAS jet calibrations, and for studies of the jet energy and mass resolution. In this circumstance, they are groomed using the same algorithm and parameters as the reconstructed jets to which they are being compared (Sect. 7).

#### 3.1.2 Inner detector tracks

Tracks are reconstructed from charged-particle hits in the inner detector. In order to ensure that only well-reconstructed tracks from the hard scattering are used, track quality criteria are applied. The ‘loose’ quality working point is used, which places requirements on the number of silicon hits in each subdetector \cite{65}. Tracks are associated to the primary vertex (PV) of the hard interaction by placing a requirement on the track distance of closest approach to the PV along the z axis, $|z_0 \sin \theta| < 2.0$ mm. The PV is selected as the vertex with the highest scalar $p_T^2$ sum of tracks associated with it using transverse and longitudinal impact parameter requirements. In addition, tracks are required to have $p_T > 500$ MeV and to be within the tracking volume ($|\eta| < 2.5$).

#### 3.1.3 Topological clusters

Jets reconstructed from ATLAS calorimeter information are built from ‘topoclusters’ \cite{66}, which are three-dimensional groupings of topologically connected calorimeter cells. Topoclusters are formed using iterated ‘seed’ and ‘collect’ steps based on the absolute value of the signal significance in a cell relative to the expected noise, $\sigma_{\text{noise}}$, which considers both electronic noise and stochastic noise from pile-up interactions. Cells with signal significance over $4\sigma_{\text{noise}}$ in an event are allowed to seed topocluster formation, and their neighbouring cells with significance over $2\sigma_{\text{noise}}$ are subsequently included. This step is repeated until all adjacent cells have a significance below $2\sigma_{\text{noise}}$, at which point all neighbouring cells are added to the cluster ($2\sigma_{\text{noise}}$). If this process results in a cluster with two or more local energy maxima, a splitting algorithm is used to separate the showers. The energies of the resulting set of clusters are calibrated at the electromagnetic (EM) scale, and all clusters are taken to be massless.

An additional calibration using the local cell weighting (LCW) scheme is applied to form clusters whose energy is calibrated at the correct particle-level scale \cite{66}. This weighting scheme classifies energy depositions as either electromagnetic- or hadronic-like using a variety of cluster moments, and accounts for the non-compensating response of the calorimeter, out-of-cluster energy, and for energy deposited in the dead material within the detector.
Table 1 Summary of pile-up mitigation algorithms, jet inputs, and grooming algorithms, the abbreviated names used throughout this work, and the relevant parameters tested for each algorithm. UFOs are introduced in Sect. 5. Algorithms marked with an asterisk (*) were studied, but were not found to produce results significantly different from other configurations. Such results are not presented in these studies.

| Algorithm                                      | Abbreviation | Settings                        |
|------------------------------------------------|--------------|---------------------------------|
| Jet input objects                              | Topological Clusters | Topoclusters | N/A |
| Particle-Flow                                  | PFlow        | N/A                             |
| Track-Caloclusters                             | TCCs         | N/A                             |
| Unified Flow Objects                           | UFOs         | N/A                             |
| Pile-up mitigation algorithms                  | Constituent Subtraction | CS | $A_F = 0.01$ |
|                                                  |              | $\Delta R_{\text{max}} = 0.25$ | $\alpha = 0$ |
| Voronoi Subtraction (*)                        | VS           | N/A                             |
| SoftKiller                                     | SK           | $\ell = 0.6$                    |
| Pile-up Per Particle Identification            | PUPPI        | $R_{\text{min}} = 0.001$        |
|                                                  |              | $R_0 = 0.3$                     |
|                                                  |              | $a = 200 \text{ MeV}$           |
|                                                  |              | $b = 14 \text{ MeV}$            |
| Jet grooming algorithms                        | Soft-Drop    | $z_{\text{cut}} = 0.1$          |
|                                                  |              | $\beta = 0, 1, 2(*)$            |
| Bottom-up Soft-Drop                            | BUSD         | $z_{\text{cut}} = 0.05, 0.1$    |
| Recursive Soft-Drop                            | RSD          | $z_{\text{cut}} = 0.05, 0.1$    |
|                                                  |              | $\beta = 0, 1, 2(*)$            |
|                                                  |              | $N = 3, 5(*)$, $\infty$         |
| Pruning                                        | N/A          | $z_{\text{cut}} = 0.15$         |
| Trimming                                       | N/A          | $R_{\text{cut}} = 0.25$         |
|                                                  |              | $f_{\text{cut}} = 5\%, 9\%$    |
|                                                  |              | $R_{\text{sub}} = 0.1, 0.2$     |

Finally, the angular coordinates ($\eta$ and $\phi$) of topoclusters are recalculated relative to the primary vertex of the event, instead of the geometric centre of the ATLAS detector. A detailed description of topocluster reconstruction and calibration is provided in Ref. [66].

3.1.4 Particle-flow objects (PFOs)

Particle-flow (PFlow) reconstruction combines track- and calorimeter-based measurements and results in improved jet energy and mass resolution, and improved pile-up stability relative to jets reconstructed from topoclusters alone [32, 67]. Double-counting of contributions from the momentum measurement of charged particles in the inner detector and their energy measurement from the calorimeters is removed using a cell-based energy subtraction.

The PFlow algorithm first attempts to match each selected track to a single topocluster in the calorimeter, using topoclusters calibrated to the EM scale, and tracks selected using the “tight” quality working point [65]. The track momentum and the topocluster position are used to compute the expected energy deposition in the calorimeter by the particle that created the track. It is not uncommon for a single particle to deposit energy in multiple topoclusters. For each track/topocluster system, the PFlow algorithm evaluates the probability that the particle’s energy was deposited in more than one topocluster, and may include additional topoclusters in the track/topocluster system if they are necessary to reconstruct the full shower energy. The expected energy deposited in the calorimeter by the particle that produced the track is subtracted, cell-by-cell, from the associated topoclusters. If the associated calorimeter energy following this subtraction is consistent with the expected shower fluctuations of a single particle, the remaining calorimeter energy is removed.

Topoclusters which are not matched to any tracks are assumed to contain energy deposited by neutral particles and are left unmodified. In the cores of jets, particles are often produced at higher energies and in dense environments, decreasing the advantages of using inner-detector-based measurements of charged particles. To account for this degradation of inner tracker performance, the shower subtraction is gradually disabled for tracks with momenta below 100 GeV if the
energy $E_{\text{clus}}$ deposited in the calorimeter in a cone of size $\Delta R = 0.15$ around the extrapolated track trajectory satisfies

$$\frac{E_{\text{clus}} - \langle E_{\text{dep}} \rangle}{\sigma (E_{\text{dep}})} > 33.2 \times \log_{10}(40 \text{ GeV/} p_T^{\text{trk}}) ,$$

where $E_{\text{dep}}$ is the expected energy deposition from a charged pion. The subtraction is completely disabled for tracks with $p_T > 100$ GeV when this condition is satisfied.

After the PFlow algorithm has run to completion, the collection of particle-flow objects (PFOs) consists of tracks, and both modified and unmodified topoclusters. Charged PFOs which are not matched to the PV are removed in order to reduce the contribution from pile-up; this procedure is referred to as ‘Charged Hadron Subtraction’ (CHS) [68,69].

### 3.1.5 Track-CaloClusters (TCCs)

Track-CaloClusters (TCCs) [33] were developed in the context of searches for massive BSM diboson resonances [7]. These constituents combine calorimeter- and inner-detector-based measurements in a manner which is optimised for jet substructure reconstruction performance in the highest-$p_T$ based measurements in a manner which is optimised for jet substructure reconstruction performance in the highest-$p_T$ jets. Unlike PFlow, which uses the expected energy deposits of single particles to determine the contributions of individual tracks to clusters, the TCCs use the energy information from topoclusters and angular information from tracks.

The TCC algorithm starts by attempting to match each ‘loose’ track in the event (from both the hard-scatter and pile-up vertices) to topoclusters calibrated to the local hadronic scale in the calorimeter. In the case where one track matches one topocluster, the $p_T$ of the TCC object is taken from the topocluster, while its $\eta$ and $\phi$ coordinates are taken from the track. In more complex situations where multiple tracks are matched to multiple topoclusters, several TCC objects are created (where the TCC multiplicity is equal to the track multiplicity): each TCC object is given some fraction of the momentum of the topocluster, where that fraction is determined from the ratios of momenta of the matched tracks. TCC angular properties ($\eta, \phi$) are taken directly from the unmodified inner detector tracks, and their mass is set to zero.

As in PFlow reconstruction, unmatched topoclusters are included in the TCC objects as unmodified neutral objects.

### 3.2 Jet-input-level pile-up mitigation algorithms

Prior to jet reconstruction, the set of input objects may be preprocessed by one or by a combination of several input-level pile-up mitigation algorithms. When reconstructing jets from topoclusters, these algorithms are applied to the entire set of inputs. When incorporating tracking information, the PV provides an additional, powerful method to reject charged particles from pile-up interactions. In this case, these additional pile-up mitigation algorithms are applied only to the neutral PFOs or TCCs in an event before jet finding.

#### 3.2.1 Constituent subtraction (CS)

Constituent Subtraction [34] is a per-particle method of performing area subtraction [70] on jet input objects. The catchment area [26] of each input object is defined using ghost association: massless particles called ‘ghosts’ are overlaid on the event uniformly, with $p_T$ satisfying

$$p_T^g = A_g \times \rho ,$$

where $A_g$, the area of the ghosts, is set to 0.01 and $p_T^g$ corresponds to the expected contribution from pile-up radiation in a small $\Delta \eta \times \Delta \phi$ area of 0.1 $\times$ 0.1. For each event, the pile-up energy density $\rho$ is estimated as the median of the $p_T/A$ distribution of the $R = 0.4$ $k_t$ [71] jets in the event. These jets are reconstructed without a $p_T$ requirement, but are required to be within $|\eta| < 2.0$. The total $p_T$ of all of the ghosts is equal to the expected average pile-up contribution, based on the estimated value of $\rho$.

After the ghosts have been added, the distance $\Delta R_{i,k}$ between each cluster $i$ and ghost $k$ is given by

$$\Delta R_{i,k} = \sqrt{(\eta_i - \eta_k)^2 + (\phi_i - \phi_k)^2} .$$

The cluster-ghost pairs are then sorted in order of ascending $\Delta R_{i,k}$, and the algorithm proceeds iteratively through each $(i,k)$ pair, modifying the $p_T$ of each cluster and ghost by

- If $p_{T,i} \geq p_{T,k}$: $p_{T,i} \rightarrow p_{T,i} - p_{T,k} , \quad p_{T,k} \rightarrow 0$;
- otherwise: $p_{T,k} \rightarrow p_{T,k} - p_{T,i} , \quad p_{T,i} \rightarrow 0$.

until $\Delta R_{i,k} > \Delta R_{\text{max}}$, where $\Delta R_{\text{max}}$ is a free parameter of the algorithm taken to be 0.25 in this study, based on studies of $R = 0.4$ jet performance [72]. Any ghosts remaining after the subtraction are eliminated.

In the authors’ description of this algorithm, a correction is also applied for the mass of the input object. Since all neutral ATLAS jet inputs are defined to be massless, this correction is unnecessary in the ATLAS implementation.

#### 3.2.2 SoftKiller (SK)

The SoftKiller (SK) [37] algorithm applies a $p_T$ cut to input objects. This cut is chosen on an event-by-event basis such that the value of $\rho$ after the selection is approximately zero. To achieve this, the event is divided into an $\eta-\phi$ grid of user-specified length scale, chosen to be $\ell = 0.6$, based on studies

\footnote{In the original formulation, there is also the option to make a $p_T^2$-dependent distance metric. Only values of $\alpha = 0$ were considered in Ref. [34], and so only this configuration is considered in these studies.}
of $R = 0.4$ jet performance \[72\]. The $p_T$ cut is determined in order to make half of the grid spaces empty after it is applied (input objects are removed from all grid cells, not just the half which are empty following SK).

To account for detector-level effects, where input objects may not consist purely of hard-scatter or pile-up contributions (see appendix), the best performance is achieved by applying some form of area subtraction to input objects before applying SK. In these studies, SK is always applied to inputs after the CS algorithm; this combination is indicated as ‘CS + SK’.

An alternative approach to assigning areas to jet input objects is based on Voronoi tessellation \[36\] and was studied both in isolation and in conjunction with the SoftKiller algorithm. Both variants of this alternative were found to perform similarly to the CS + SK results presented here.

### 3.2.3 Pile-up per particle identification (PUPPI)

‘Pile-up per particle identification’, or PUPPI \[38\], is a pile-up-mitigation algorithm which assigns each input object $i$ a likelihood to have originated from a pile-up interaction based on its kinematic properties and proximity to charged hard-scatter particles matched to the event’s PV. This likelihood is given by

$$\alpha_i = \log \left( \sum_j \frac{p_T^j}{\Delta R_{ij}} \times \Theta (R_{\text{min}} \leq \Delta R_{ij} \leq R_0) \right),$$

where the index $j$ tracks the charged inputs matched to the PV, $R_0$ is the maximum radial distance at which inputs may be matched to each other, $R_{\text{min}}$ is the minimum radial distance of matching, $\Delta R_{ij}$ is the angular distance between an input object and a charged hard-scatter particle, and $\Theta$ is the Heaviside step function. The value of $R_{\text{min}}$ is generally taken to be very small, and is chosen to be 0.001 in these studies. The value of $R_0$ is chosen to be 0.3.

Once $\alpha$ has been calculated for all input objects, then the following quantity is determined:

$$\chi^2_i = \Theta (\alpha_i - \bar{\alpha}_{\text{PU}}) \times \frac{(\alpha_i - \bar{\alpha}_{\text{PU}})^2}{\sigma_{\text{PU}}}.$$

where $\bar{\alpha}_{\text{PU}}$ is the mean value of $\alpha$ for all charged pile-up input objects in the event, and $\sigma_{\text{PU}}$ is the RMS of that same distribution. The four-momentum of each neutral input $i$ is then weighted by

$$w_i = F_{\chi^2} \times \frac{1}{\text{NDF}=1} \left( \chi^2_i \right),$$

where $F_{\chi^2}$ is the cumulative distribution of the $\chi^2$ distribution, eliminating all neutral inputs $i$ whose calculated value of $\alpha_i$ is less than $\bar{\alpha}_{\text{PU}}$.

In order to suppress additional noise, a $p_T$ cut is applied to the remaining input objects after they have been reweighted. This cut is dependent on the number of reconstructed primary vertices ($N_{\text{PV}}$), and is determined by

$$p_T,\text{cut} = a + b \times N_{\text{PV}}$$

where the parameters $a$ and $b$ are user-specified. For these studies, the parameters are chosen to be $a = 200$ MeV and $b = 14$ MeV, based on studies of the $R = 0.4$ PFlow jet energy resolution.

While PUPPI could technically be applied to topoclusters, the principles of the algorithm depend strongly on the matching of neutral input objects to nearby charged particles from the hard-scatter event. It is therefore more effective for particle-flow-type algorithms. Due to the large number of free parameters, and since it has only been optimised for ATLAS PFlow jets with $R = 0.4$, PUPPI is only applied to PFlow jets.

### 3.3 Grooming algorithms

#### 3.3.1 Trimming

Trimming \[41\] was designed to remove contamination from soft radiation in the jet by excluding regions of the jet where the energy flow originates mainly from the underlying event, pile-up, or initial-state radiation (ISR), in order to improve the resolution of the jet energy and mass measurements. In Run I \[31\], it was also found to be effective in mitigating the effects of pile-up on large-$R$ jets. To trim a large-$R$ jet, the jet constituents are reclustered into subjets of a user-specified radius $R_{\text{sub}}$ using the $k_t$ algorithm. Subjets with $p_T$ less than some user-specified fraction $f_{\text{cut}}$ of the $p_T$ of the original ungroomed jet are discarded; their constituents are removed from the final groomed jet.

#### 3.3.2 Pruning

Pruning \[42\] proposes a modification of the jet clustering sequence, which removes splittings that are assessed as likely to pull in soft radiation from pile-up interactions and the underlying event. This is achieved by determining a ‘pruning radius’ such that hard prongs fall into separate subjets, while discarding softer radiation outside of these prongs. The constituents of the large-$R$ jet are reclustered using the Cambridge–Aachen (C/A) algorithm \[73,74\] to form an angle-ordered cluster sequence. At each step of the clustering sequence, the softer subjet is discarded if it is either too soft or wide-angled, enforced by requiring

$$\Delta R_{12} \geq R_{\text{cut}} \times \frac{2 M_{12}}{p_T,12} \frac{1}{p_T,12}$$

where $z \leq z_{\text{cut}}$. 

$\odot$ Springer
where $\Delta R_{12}$, $M_{12}$, and $p_{T,12}$ are respectively the angular distance, the mass, and the transverse momentum of the subjet pair at a given step in the clustering sequence, and $z = \min \left( \frac{p_{T,1}, p_{T,2}}{p_{T,1} + p_{T,2}} \right)$. The parameters $R_{\text{cut}}$ and $z_{\text{cut}}$ are user-defined, and respectively control the amount of wide-angled and soft radiation which is removed by the pruning algorithm.

3.3.3 Soft-drop (SD)

Soft-drop [45] is a technique for removing soft and wide-angle radiation from a jet. In this algorithm, the constituents of the large-$R$ jet are reclustered using the C/A algorithm, creating an angle-ordered jet clustering history. Then, the clustering sequence is traversed in reverse (starting from the widest-angled radiation and iterating towards the jet core). At each step in the clustering sequence, the kinematics of the splitting are tested with the condition

$$\min(p_{T,1}, p_{T,2}) \over p_{T,1} + p_{T,2} < z_{\text{cut}} \left( \frac{\Delta R_{12}}{R} \right)^{\beta},$$

where the subscripts 1 and 2 respectively denote the harder and softer branches of the splitting, and the parameters $z_{\text{cut}}$ and $\beta$ dictate the amount of soft and wide-angled radiation which is removed. If the splitting fails this condition, the lower-$p_T$ branch of the clustering history is removed, and the declustering process is repeated on the higher-$p_T$ branch. If the condition is satisfied, the process terminates and the remaining constituents form the groomed jet.

If $\beta = 0$, SD suppresses radiation purely based on the $p_T$, while larger values of $\beta$ allow more soft radiation to remain within the groomed jet when it is sufficiently collinear. SD with $\beta = 0$ is equivalent to the modified Mass Drop Tagger (MDT) algorithm [31,75]. SD grooming has an intrinsic quality which is not shared by the trimming or pruning algorithms: certain jet substructure observables are calculable beyond leading-logarithm accuracy following the application of SD [75–81].

3.3.4 Recursive soft-drop (RSD) and bottom-up soft-drop (BUSD)

The standard soft-drop algorithm aims to find the first hard splitting in the jet clustering history in order to define a groomed jet. In the case of a multi-pronged decay, this treatment may not be sufficient to remove enough soft radiation from the jet, since the SD condition may be satisfied before removing all of this energy. A recursive extension of the SD algorithm (‘recursive soft-drop,’ or RSD) has been proposed [46], in which the algorithm continues recursively along the harder branch of the C/A clustering sequence until $N$ hard splittings have been found. The case of $N=1$ is equivalent to the standard SD algorithm, while for larger values of $N$, a larger fraction of the jet may be traversed by the grooming algorithm. When $N = \infty$, the entire C/A sequence is traversed by the grooming algorithm regardless of the number of hard splittings found.

Bottom-up soft-drop (BUSD) [46] instead incorporates the SD criteria within the jet clustering algorithm, similar to pruning. In these studies, the ‘local’ version of BUSD is implemented, which is applied after initial jet reconstruction. Using this approach, jets are reconstructed with the anti-$k_T$ algorithm, and then reclustered using a modified version of the C/A algorithm, where particles $i$ and $j$ with the smallest distance $d_{ij} = \Delta R_{ij}/R_0$ are combined to create a new pseudojet given by

$$p_{ij} = \begin{cases} \max(p_i, p_j), & \text{if the soft-drop condition fails}, \\ p_i + p_j, & \text{otherwise}. \end{cases}$$

The results of applying local BUSD are expected to be similar to those of RSD with $N = \infty$, since both algorithms begin with the same set of constituents per jet and groom the entire C/A clustering sequence.

Other configurations for the SD family of algorithms were studied, including $\beta = 2$ grooming, but were not found to give results significantly different from those reported in detail.

4 Performance metrics

In order to survey the relative performance of all considered large-$R$ jet definitions, several metrics must be established which probe relevant aspects of their behaviour in the context of large-$R$ jet reconstruction and calibration by ATLAS. It is not feasible to calibrate each of the definitions studied (even with a simulation-based approach, as in Sect. 7), and so these metrics have been chosen in order to be robust against differences caused by calibration. The metrics selected include the tagging performance of high-$p_T$ $W$ bosons and top quarks, the stability of the jets in the presence of pile-up interactions, and the degree to which a jet definition’s mass scale depends on the signal- or background-like substructure of the jet.

In this section, the behaviour of each metric is illustrated using a reduced list of jet definitions that have been selected to highlight the interplay between different aspects of jet reconstruction. For each metric, jets reconstructed from topological clusters, particle-flow and track-calocluster input objects are compared, with and without pile-up mitigation. Two grooming algorithms are also compared for each jet input: trimming with $R_{\text{cut}} = 0.2$ and $f_{\text{cut}} = 0.05$, and soft-drop with $\beta = 1.0$ and $z_{\text{cut}} = 0.1$. The trimming algorithm is chosen because it is the current baseline definition used by ATLAS. The soft-drop algorithm is chosen as an alternative which has demonstrated good performance, as is shown in Sect. 6.
Results of the complete survey of all jet definitions summarised in Table 1 are provided in Sect. 6.

4.1 Tagging performance

Many analyses using large-\(R\) jets rely on a tagger to distinguish between different types of jets, such as distinguishing between the decay of a high-\(p_T\), hadronically decaying top quark and a jet originating from a high-energy quark or gluon. Such boosted-particle taggers range in complexity from simple mass cuts to complex machine-learning algorithms [82–84]. While the complete optimisation of a jet tagger is outside the scope of this work, it is important to compare the tagging performance of different jet definitions in terms of their background rejection (defined as the reciprocal of the background-jet tagging efficiency) at fixed signal-jet tagging efficiency. This may be done using a simple tagger based on the jet mass and a jet substructure (JSS) observable. In order to study the tagging performance for different jet topologies, taggers are created for high-\(p_T\) \(W\) bosons and top quarks by combining the jet mass with another jet substructure observable which is sensitive to either two- or three-pronged signal jet topologies.

The jet mass, as defined by

\[
M_{\text{jet}} = \sqrt{\left(\frac{1}{N_{\text{jet}}} \sum_{i \in \text{jet}} E_i \right)^2 - \left(\frac{1}{N_{\text{jet}}} \sum_{i \in \text{jet}} \vec{p}_i \right)^2},
\]

where \(i\) are the constituents of the jet, is typically one of the most powerful variables that can be used to discriminate between different types of jets.

To tag boosted \(W\) decays, which have a two-pronged structure, the \(D_2\) observable [85–87] is used with a choice of angular exponent \(\beta = 1.0\). This observable is a ratio of three-point to two-point energy–energy correlation functions which has been used by ATLAS in \(W\) taggers since Run 1 [39,82].

For boosted top quark decays, which have a three-pronged structure, \(\tau_{32}\) with the winner-take-all axis configuration [88,89] is used. This observable is a ratio of two \(N\)-subjettiness variables, which tests the compatibility of a jet’s substructure with a particular \(N\)-pronged hypothesis. ATLAS has incorporated \(\tau_{32}\) into its top taggers, whether simple or complex, since Run 1 [59,82].

Unlike a mass-only tagger, where more aggressive grooming can improve the jet mass resolution at the cost of grooming away additional information contained within a jet’s soft radiation, a mass + JSS tagger relies on such soft radiation to achieve better background rejection. Such taggers are a more realistic approximation to the expected future tagging performance of any given jet definition (which will use more sophisticated techniques), and are amenable to this survey of many jet definitions.

For both the \(W\) and top taggers, the tagging algorithm proceeds similarly: first, a fixed signal-efficiency (\(\epsilon_{\text{sig}}\)) mass window is selected, where the window is defined to be the minimum mass range which contains 68% of the signal mass distribution. This window should select the signal jet mass peak. A one-sided cut is then applied to \(D_2\) or \(\tau_{32}\), and background rejection (\(1/\epsilon_{\text{bkg}}\)) is compared at a fixed signal efficiency taken to be \(\epsilon_{\text{sig}} = 50\%\). This signal efficiency working point is representative of taggers used by ATLAS in physics analysis, and the results were not found to depend strongly on the working point which was selected. The relative performance of various jet definitions in terms of their background rejection at a fixed signal efficiency point was noted to typically provide a consistent ordering of jet definitions before and after applying a simulation-based calibration, and so this metric was selected instead of possible alternatives such as the Receiver Operating Characteristic (ROC) curve integral.

The background rejection for the boosted \(W\) boson tagger is shown as a function of signal tagging efficiency in Fig. 1 for two \(p_T\) bins: a low-\(p_T\) bin (300 GeV < \(p_T\) < 1500 GeV), and a high-\(p_T\) bin (1000 GeV < \(p_T\) < 1500 GeV), where kinematic requirements are placed on the \(p_T\) of the ungroomed particle-level jet which is associated with the detector-level jet under study (Sect. 3.1.1). The low-\(p_T\) bin represents the regime where the \(W\) decay products are boosted just enough to be contained within a single large-\(R\) jet, while the high-\(p_T\) bin represents the regime where the decay products are more collimated and may begin to merge. The performance in these two regions is expected to be different due to detector effects and algorithmic differences. Similarly, the background rejection of the top tagger is shown in Fig. 2, except the lower \(p_T\) bin is chosen to be 500 GeV < \(p_T\) < 1000 GeV, since the larger mass of the top quark results in less collimation of its decay products.

Better alternatives to the baseline topocluster jet definition are clearly visible. At low \(p_T\), PFlow reconstruction results in the best performance for \(W\) boson and top tagging, while TCCs have a lower background rejection than topocluster jets. At high \(p_T\), TCCs provide a significantly better background rejection than the other options, although PFlow still provides an improvement over topocluster reconstruction.

The application of CS + SK pile-up mitigation has very little effect for the high-\(p_T\) jet, but for the low-\(p_T\) \(W\) tagger, it significantly improves the background rejection for soft-drop jets, which are more susceptible to pile-up than trimmed jets. This effect is seen for all three jet input types, but it is pronounced for topocluster inputs, which do not use tracking information to remove pile-up. Top tagging performance benefits more from adopting soft-drop grooming than \(W\) tagging: background rejection increases when tagging top
Fig. 1 Background rejection as a function of signal efficiency for a tagger using the jet mass and $D_2$ for W boson jets at (a), (c), (e) low $p_T$, and (b), (d), (f) high $p_T$. Several different jet input object types are shown: (a), (b) topoclusters, (c), (d) particle-flow objects and (e), (f) track-caloclusters. Jet $p_T$ and $\eta$ cuts before tagging are made using the ungroomed particle-level large-$R$ jet matched to each of the groomed reconstructed large-$R$ jets. Jets groomed with the trimming ($R_{\text{sub}} = 0.2$, $f_{\text{cut}} = 0.05$) and soft-drop ($\beta = 1.0$, $z_{\text{cut}} = 0.1$) algorithms are shown. The background rejection factor of the baseline topocluster-based trimmed collection at a fixed signal tagging efficiency of 50% is indicated with a $\ast$. 
Fig. 2 Background rejection as a function of signal efficiency for a tagger using the jet mass and $\tau_{32}$ for top quark jets at a, c, d low $p_T$, and b, d, f high $p_T$. Several different jet input object types are shown: a, b topoclusters, c, d particle-flow objects and e, f track-caloclusters. Jet $p_T$ and $\eta$ cuts before tagging are made using the ungroomed particle-level large-$R$ jet matched to each of the groomed reconstructed large-$R$ jets. Jets groomed with the trimming ($R_{sub} = 0.2$, $f_{cut} = 0.05$) and soft-drop ($\beta = 1.0$, $z_{cut} = 0.1$) algorithms are shown. The background rejection factor of the baseline topocluster-based trimmed collection at a fixed signal tagging efficiency of 50% is indicated with a ⋆.
quarks regardless of the input object type or $p_T$ bin when soft-drop is chosen.

### 4.2 Pile-up stability

Two metrics are used to study the pile-up stability of jet definitions in order to determine which definitions are sufficiently insensitive to pile-up. The first quantifies the effect on the jet mass scale by studying how the $W$ boson mass peak position changes as a function of pile-up, and provides a handle with which to assess the impact of pile-up on a jet's hard structure. The second quantifies the impact on substructure observables by studying the pile-up dependence of $W$ boson tagging efficiency, in order to quantify how pile-up contributions alter the soft radiation patterns within jets.

A related study of the effects of pile-up on topocluster reconstruction is presented in an appendix of this publication, utilising a new technique which propagates particle-level information about hard-scatter and pile-up energy depositions through the ATLAS reconstruction procedure.

#### 4.2.1 Pile-up stability of the $W$ boson jet mass peak position

Jet substructure observables such as the jet mass are particularly sensitive to pile-up; the contribution of pile-up to the jet mass scales approximately with the jet radius cubed [90].

Figure 3 shows a subset of the trimmed mass distribution of $W$ jets in bins of $N_{PV}$ for various jet input object types, demonstrating that pile-up can visibly alter the average value and width of the jet mass distribution. This effect is quantified using a simple metric. In bins of $N_{PV}$, the core of the $W$ mass peak is iteratively fit with a Gaussian distribution. The trend of the fitted peak position versus $N_{PV}$ is then fit with a line. The slope of this line is a measure of the sensitivity of the jet mass to PU: a larger magnitude indicates larger pile-up. The second quantifies the impact on substructure observables by studying the pile-up dependence of $W$ boson tagging efficiency, in order to quantify how pile-up contributions alter the soft radiation patterns within jets.

As pile-up levels increase, the signal efficiency of the $W$ tagger tends to decrease, although the opposite behaviour is often observed for TCC jets. Similarly to what was found when studying the $W$ mass peak position metric (Sect. 4.2.1), topocluster inputs are the least stable. After pile-up mitigation, the pile-up stability of all inputs, including TCCs, improves. The trends in stability as a function of grooming algorithm are the same as for the $W$ mass peak position.

#### 4.3 Topological sensitivity

ATLAS calibrates large-$R$ jets using a procedure which involves simulation-based and in situ methods [91]. For the simulation-based calibration, the average jet energy and mass scale in reconstructed jets are calibrated to the average scale of jets at particle level, using a sample of jets originating from light quarks and gluons (Sect. 7.1). These light-quark- and gluon-derived calibrations are also currently applied to all jets, including to signal jets (e.g. $W/Z/H/t$ jets). Dependence of the jet energy and mass scale on the progenitor of the jet is undesirable: if the jet mass scale for signal and background jets with similar kinematics is different, then the signal jets will receive an incorrect calibration factor.

In order to examine the topology dependence of the jet mass scale for different jet definitions, the ratio of the mean value of the uncalibrated jet mass response, $R_m = m_{\text{true}}/m_{\text{true}}$, for signal $W$ jets to that of background jets is constructed within a bin of large-$R$ jet $p_T$, $\eta$ and mass. Deviations from unity will result in non-closure in the mass response for signal jets following calibration (Sect. 7.1). This effect is relevant at low $p_T$, where $W$ jets may be contained.
Fig. 3 Pile-up dependence of the $W$ boson jet mass reconstructed using (a) topoclusters, (b) particle-flow objects and (c) track-caloclusters. Distributions are shown for the trimming grooming algorithm ($R_{\text{sub}} = 0.2$, $f_{\text{cut}} = 0.05$), with unmodified jet input objects. Jet $p_T$ and $\eta$ cuts before tagging are made using the ungroomed particle-level large-$R$ jet matched to each of the groomed reconstructed large-$R$ jets.

within an $R = 1.0$ jet, but top quarks are not; therefore, only $W$ jets and background jets are considered in this context. The baseline topocluster-based trimmed large-$R$ jet definition used by the ATLAS experiment exhibits a difference for signal jets of 4% by this metric; therefore, deviations from unity of 4% or less have not been found to be problematic at later stages of the calibration workflow [91], given the current level of calibration precision.

Figure 6 shows the jet mass response for signal and background jets built from topological clusters and groomed with either the trimming or soft-drop grooming algorithms. The low-$p_T$ bin, where this topological effect is most pronounced, is shown. A larger sensitivity to the signal- or background-like nature of the jet is observed for soft-drop grooming, which retains more soft radiation. The application of pile-up mitigation can exacerbate topological differences in the jet mass scale by altering the distribution of soft jet constituents differently depending on the jet’s signal- or background-like topology.

5 Unified flow objects (UFOs)

After observing the behaviour of the jet input objects currently used by ATLAS in physics analyses (topoclusters, PFOs and TCCs), it is clear even from the reduced set of jet definitions (Sect. 4) that no single jet definition is optimal according to all metrics. While TCCs significantly improve tagging performance at high $p_T$, their performance is typically worse than the baseline topocluster-based trimmed jet definition at low $p_T$, and they are more sensitive to pile-up than other definitions. Jets reconstructed from PFOs can...
Fig. 4 The value of the fitted $W$ boson mass peak as a function of the number of primary vertices, $N_{PV}$. Several different jet input object types are shown: a topoclusters, b particle-flow objects and c track-caloclusters. Jet $p_T$ and $\eta$ cuts before tagging are made using the ungroomed particle-level large-$R$ jet matched to each of the groomed reconstructed large-$R$ jets. Jets groomed with the trimming ($R_{\text{sub}} = 0.2$, $f_{\text{cut}} = 0.05$) and soft-drop ($\beta = 1.0$, $z_{\text{cut}} = 0.1$) algorithms are shown improve on the baseline definition for the entire $p_T$ range, but their tagging performance is significantly worse than that of TCC jets at high $p_T$ when given the same grooming algorithm.

The relative performance of these jet definitions can be understood by reflecting on how different inputs are reconstructed. For low-$p_T$ particles, PFOs are designed to improve the correspondence between particles and reconstructed objects. However, as the particle $p_T$ increases or the environment close-by to the particle becomes dense, the inner detector’s momentum resolution deteriorates, and so the PFlow subtraction algorithm is gradually disabled in order to avoid degradation of the jet energy resolution.

The cluster splitting scheme used for TCCs does not utilise a detailed understanding of the correlation between tracks and clusters, and instead is designed to resolve many (charged) particles without double counting their energy. When splitting low-energy topoclusters, this can result in an incorrect redistribution of the cluster’s energy, while for high-energy clusters, the ability to resolve many particles increases the relative tagging performance of TCCs over other definitions. TCCs exhibit pile-up instabilities at low $p_T$, where the mass scale decreases as the number of pile-up interactions increases. This trend is the opposite of what is observed for jets reconstructed from topoclusters and PFOs, and occurs because the TCC algorithm splits clusters into more components when additional tracks from pile-up interactions are present in the reconstruction procedure.

These observations motivate the development of a new jet input object, which combines desirable aspects of PFO and
The signal efficiency of a W boson tagger as a function of the number of primary vertices, \( N_{\text{PV}} \). Several different jet input object types are shown: (a) topoclusters, (b) particle-flow objects and (c) track-caloclusters. Jet \( p_T \) and \( \eta \) cuts before tagging are made using the ungroomed particle-level large-\( R \) jet matched to each of the groomed reconstructed large-\( R \) jets. Jets groomed with the trimming (\( R_{\text{sub}} = 0.2, f_{\text{cut}} = 0.05 \)) and soft-drop (\( \beta = 1.0, z_{\text{cut}} = 0.1 \)) algorithms are shown.

The signal efficiency of a W boson tagger as a function of the number of primary vertices, \( N_{\text{PV}} \). Several different jet input object types are shown: (a) topoclusters, (b) particle-flow objects and (c) track-caloclusters. Jet \( p_T \) and \( \eta \) cuts before tagging are made using the ungroomed particle-level large-\( R \) jet matched to each of the groomed reconstructed large-\( R \) jets. Jets groomed with the trimming (\( R_{\text{sub}} = 0.2, f_{\text{cut}} = 0.05 \)) and soft-drop (\( \beta = 1.0, z_{\text{cut}} = 0.1 \)) algorithms are shown.

TCC reconstruction in order to achieve optimal overall performance across the full kinematic range. These new inputs are called Unified Flow Objects (UFOs).

The UFO reconstruction algorithm is illustrated in Fig. 7. The process begins by applying the standard ATLAS PFlow algorithm (Sect. 3.1.4). Charged PFOs which are matched to pile-up vertices are removed. The remaining PFOs are classified into different categories: neutral PFOs, charged PFOs which were used to subtract energy from a topocluster, and charged PFOs for which no subtraction was performed due to their high momentum or being located in a dense environment. Jet-input-level pile-up mitigation algorithms may now be applied to the neutral PFOs if desired. A modified version of the TCC splitting algorithm is then applied to the remaining PFOs: only tracks from the hard-scatter vertex are used as input to the splitting algorithm, in order to avoid pile-up instabilities. Any tracks which have been used for PFlow subtraction are not considered, as they have already been well-matched and their expected contributions have been subtracted from the energy in the calorimeter. The TCC algorithm then proceeds as described in Sect. 3.1.5, using the modified collection of tracks to split neutral and unsubtracted charged PFOs instead of topoclusters. This approach provides the maximum benefit of PFlow subtraction at lower particle \( p_T \), and cluster splitting where the benefit is maximal at high particle \( p_T \).

The performance of UFOs is illustrated in Figs. 8 and 9 according to the same metrics as for other jet input objects in Sect. 4. The increased tagging performance of UFOs is demonstrated across both the low and high \( p_T \) ranges in.
Fig. 6 Distribution of the jet mass response in W jets and q/g jets reconstructed from topoclusters. The mass response is constructed following application of the a trimming ($R_{\text{sub}} = 0.2$, $f_{\text{cut}} = 0.05$) or b soft-drop ($\beta = 1.0$, $z_{\text{cut}} = 0.1$) grooming algorithms at both truth and detector level. Jet $p_T$ and $\eta$ selections are made using the ungroomed particle-level large-$R$ jet matched to each of the groomed detector-level large-$R$ jets. The uncertainties from the fits are typically less than 0.005. A particle-level mass-window cut with 68% signal efficiency is applied to both the groomed signal and background jets.

Fig. 8, where their performance is superior to that of TCC jets at high $p_T$, and becomes similar to that of PFlow jets as $p_T$ decreases.

UFOs are naturally pile-up-stable due to the inclusion of only charged-particle tracks matched to the primary vertex, similar to the ATLAS PFlow algorithm. Figure 9 demonstrates the additional stability that an input-level pile-up mitigation algorithm such as CS + SK can offer when it is applied to neutral particles (calorimeter deposits), especially at low $p_T$.

The topological dependence of UFOs is not enhanced relative to the other jet definitions previously studied, and options exist with sensitivity equal to or below that of the baseline topocluster-based trimmed definition which improve on other aspects of jet performance.

6 Performance survey

The metrics described in Sect. 4 are used to study the performance of all jet definitions listed in Table 1, with the addition of UFOs. This provides a more complete understanding of the interplay between the different aspects of jet reconstruction. The results are summarised in Figs. 10, 11, 12, 13 and 14.

6.1 Tagging performance

A comparison of the background rejection of the W tagger at the 50% signal tagging efficiency working point is shown in Fig. 10 for two $p_T$ bins: a low-$p_T$ bin ($300 \text{ GeV} < p_T^{\text{true, ungroomed}} < 500 \text{ GeV}$), and a high-$p_T$ bin ($1000 \text{ GeV} < p_T^{\text{true, ungroomed}} < 1500 \text{ GeV}$).

Several trends are apparent from the performance of the taggers. As seen in Sect. 4, for a fixed grooming algorithm, PFO reconstruction improves on topocluster reconstruction for both $p_T$ bins, while TCCs improve background rejection even further at high $p_T$. In both cases, UFO reconstruction is able to match or improve on the performance of other jet inputs for both $p_T$ bins. In general, pile-up mitigation improves $W$ tagging performance for all input types. The effects of pile-up mitigation are more apparent at low $p_T$, where soft pile-up radiation has a larger impact on the reconstruction of $D_2$. At high $p_T$, pile-up mitigation significantly improves the performance of TCC jets. This is related to the greater impact of pile-up mitigation for TCCs on the background mass distribution than the signal distribution, which increases the background rejection.

The tagging performance varies significantly among the different grooming algorithms and parameter choices. For trimming algorithms, smaller values of $R_{\text{sub}}$ or larger values of $f_{\text{cut}}$ result in reduced tagging performance, regardless of the jet input type. These parameter choices correspond to more aggressive grooming, indicating that some of the softer radiation is important for effectively tagging different types of jets. An analogous observation is made for SD jets, where small values of $\beta$, or large values of $z_{\text{cut}}$ generally result in degraded tagging performance.

A similar set of results is seen for the top tagger in Fig. 11. In the low-$p_T$ bin, PFlow jets typically outperform both topocluster and TCC jets, while TCC jets outperform the other input object types at high $p_T$. Again, UFO jets are able
Fig. 7 An illustration of the unified flow object reconstruction algorithm

to match or improve the performance compared to the other jet input types in both $p_T$ bins. Pile-up mitigation tends to improve results, particularly at low $p_T$, as observed for $W$ taggers, although in a few cases the background rejection deteriorates. The baseline trimming algorithm works well for all input object types, but at low $p_T$, the background rejection may be improved by 50% by instead using a SD algorithm with lighter grooming. The standard SD algorithm with $\beta = 1$ and $z_{\text{cut}} = 0.1$ works particularly well, although recursive and bottom-up variants can also provide comparable performance.

In general, the tagging performance of jets constructed out of UFOs matches or exceeds that of jets reconstructed out of any other input type.

6.2 Pile-up stability

The slopes of the fitted average $W$ boson jet mass as a function of $N_{\text{PV}}$ are shown in Fig. 12 for each of the surveyed jet definitions. The uncertainties in the fitted slope values tend to be negligible compared to the differences between reported values. Among jet input types, PFOs and UFOs are the most pile-up-stable. PFOs, TCCs, and UFOs are all more pile-up-stable than topoclusters, due to the ability to easily remove charged particles from pile-up vertices. As discussed in Sect. 4, the fitted value of the TCC $W$ mass peak position decreases as a function of $N_{\text{PV}}$ for most grooming algorithms, although for lighter grooming algorithms which are more affected by pile-up, the slope is sometimes positive. This effect is exacerbated by the use of CS + SK, and for CS + SK TCCs, all of the studied trends are negative.

There are significant differences in the pile-up stability of different jet grooming algorithms. In general, all studied configurations of trimming are stable. For SD, RSD and BUSD, stability depends on the parameter choice. Larger values of $\beta$, where more soft and wide-angled radiation is retained, have a larger pile-up dependence. As expected, for the same value of $z_{\text{cut}}$, RSD and BUSD are more stable than the standard SD definition.

For all input types, with the exception of TCCs, jet-input-level pile-up mitigation techniques improve the pile-up stability of the jet definitions. Since too much energy is already subtracted for TCCs because of the inclusion of pile-up tracks in their reconstruction, any additional subtraction further degrades performance. For other jet inputs, the use of pile-up mitigation reduces the pile-up sensitivity so that it is better than or equivalent to the pile-up sensitivity from the baseline trimmed topocluster jet definition. This is true even for lightly groomed algorithms (e.g. RSD with $z_{\text{cut}} = 0.05$, $\beta = 1$, $N = 3$), where CS + SK improves stability by a factor of 20. While PUPPI improves the pile-up stability of PFOs, the performance of CS + SK PFOs is better overall, sometimes by more than a factor of two. This improvement is seen for nearly all grooming algorithms. The pile-up stability of UFOs is similar to that of PFOs, which is expected since the modified TCC splitting step does not remove pile-up particles.

The change in signal efficiency of the $D_2$ tagger as a function of $N_{\text{PV}}$ is shown in Fig. 13. Uncertainties in the reported values from the fitting procedure tend to be negligible (sub-percent level). As pile-up levels increase, the signal efficiency of the $W$ tagger tends to decrease. As observed when studying the $W$ mass peak position metric, topocluster inputs are...
Background rejection as a function of signal efficiency for a tagger using (top row) the jet mass and $D_2$ for $W$ boson jets, or (bottom row) the jet mass and $\tau_3$ for top quark jets. These results are shown in (left) low-$p_T$ and (right) high-$p_T$ bins, and include a comparison of different jet input object types, including topoclusters, particle-flow objects, track-caloclusters and unified flow objects. The large-$R$ jets are groomed using the trimming algorithm ($R_{\text{sub}} = 0.2$, $f_{\text{cut}} = 0.05$). The background rejection factor of the baseline topocluster-based trimmed collection at a fixed signal tagging efficiency of 50% is indicated with a $\star$.

The least stable. After pile-up mitigation, the pile-up stability of all inputs, including TCCs, improves by this metric. The trends in stability as a function of grooming algorithm are the same as for the $W$ mass position. While CS + SK is typically still more performant than PUPPI, the degree of improvement is not as large as that observed when studying the pile-up stability of the $W$ jet mass peak-position.

### 6.3 Topological sensitivity

In order to examine the topology dependence of the jet energy and mass scale for different jet definitions, the ratio of the mean value of the uncalibrated jet mass response for $W$ jets to that of background jets is constructed. These values can be significantly different, as seen in Sect. 4. Deviations from unity will result in non-closure in the mass response following calibration. This effect is largest at low $p_T$, where the reconstruction of $W$ jets is relevant. As seen in Fig. 14, the baseline topocluster-based trimmed large-$R$ jet definition used by the ATLAS experiment shows a score of around 4% in this metric, and so small deviations from unity are not problematic.

The topology dependence is increased by the application of jet-input-level pile-up mitigation algorithms. In general, TCCs show the most sensitivity, which can reach 20% after pile-up mitigation algorithms are applied. The topological sensitivity is increased for all inputs after the application of CS + SK, regardless of the grooming algorithm applied. This effect is generally lower for UFOs than for other jet...
inputs, even after pile-up mitigation algorithms are applied; the behaviour of PFlow jets is similar.

7 Comparison of calibrated jet definitions

The tagging performance of a jet definition will have the largest impact on the sensitivity of searches for new physics performed by ATLAS, and so it is the primary metric used to determine which definitions are important for further study. The pile-up stability and topological sensitivity of the jet mass scale are also important, but since the performance of the baseline topocluster-based trimmed jet definition is still adequate, they are primarily used to distinguish between otherwise similar jet definitions. The primary motivation for choosing UFO-based definitions for further study is their $W$ boson and top quark tagging performance.

Based on their optimal tagging performance over the entire kinematic range of interest, in addition to the increased pile-up stability achieved by utilising tracking information in the jet definition, only jets reconstructed from UFOs are considered further. Several grooming algorithms are promising: soft-drop ($\beta = 1.0, z_{\text{cut}} = 0.1$) jets perform well when tagging high-$p_T$ top quarks, while the RSD ($\beta = 1.0, z_{\text{cut}} = 0.05, N = \infty$) and BUSD ($\beta = 1.0, z_{\text{cut}} = 0.05$) extensions provide further improvements for high-$p_T$ $W$ bosons. Trimmed UFO jets ($f_{\text{cut}} = 0.05, R_{\text{sub}} = 0.2$) also provide competitive performance in certain regions. These four UFO jet definitions were selected for calibration and further study, as summarised in Table 2 in the category ‘studied definitions.’
Fig. 10 Background rejection at 50% signal efficiency for a tagger using the jet mass and $D_2$ for W boson jets at (a) low $p_T$, and (b) high $p_T$. Jet $p_T$ and $\eta$ cuts before tagging are made using the ungroomed particle-level large-$R$ jet matched to each of the groomed reconstructed large-$R$ jets. The current baseline topocluster-based trimmed collection is indicated with a ★.

7.1 Simulation-based jet energy and mass scale calibrations

A simulation-based calibration is derived using PYTHIA dijet events for each of the UFO collections which were selected for further study, as well as for additional large-$R$ jet definitions which will permit comparisons of each aspect of the jet definition which is studied. These jet definitions are listed in Table 2. This calibration follows the methodology in Ref. [91], and restores the average reconstructed jet $p_T$ and mass scales (JES, JMS) to those of the particle-level references. For each jet definition, a reference set of particle-level jets are reconstructed as described in Sect. 3.1.1, and the same grooming algorithm is applied as that used for the detector-level jet definition.

Detector-level jets are matched to particle-level jets using a procedure which minimises the distance $\Delta R = \sqrt{(\Delta \phi)^2 + (\Delta \eta)^2}$. The $p_T$ and mass responses are defined respectively as $R_{p_T} = \langle p_{T\text{reco}} / p_{T\text{true}} \rangle$ and $R_m = \langle m_{\text{reco}} / m_{\text{true}} \rangle$, where the ‘reco’ quantities correspond to the value of the jet energy or mass before any calibration has been applied. The truth quantities are defined using particle-level jets, reconstructed following the procedure described in Sect. 3.1.1. The average response is determined using a Gaussian fit to the core of each response distribution.

For the JES calibration, these fits are performed in bins of jet energy and detector pseudorapidity $\eta^{jet}$, defined as the jet pseudorapidity calculated relative to the geometrical centre of the ATLAS detector. This parameterisation yields...
Fig. 11 Background rejection at 50% signal efficiency for a tagger using the jet mass and $\tau_3$ for top quark jets at a low $p_T$, and b high $p_T$. Jet $p_T$ and $\eta$ cuts before tagging are made using the ungroomed particle-level large-$R$ jet matched to each of the groomed reconstructed large-$R$ jets. The current baseline topocluster-based trimmed collection is indicated with a *.

a more accurate representation of the active calorimeter cells than that obtained when using the pseudorapidity calculated relative to the PV, and results in an improved evaluation of the calorimeter response. The JES correction factor, $c_{\text{JES}} = 1/R_{\text{Pt}}$, is smoothed in energy and $\eta_{\text{det}}$, and is applied to the four-momentum of the reconstructed jet as a multiplicative scale factor. A correction to the jet $\eta$ (‘$\Delta \eta$’ below) is also applied to correct for biases with respect to the particle-level reference in certain detector regions [92]. The JES correction is similar for each of the four CS + SK UFO jet definitions which are calibrated, regardless of the grooming algorithm which is applied.

After the JES correction has been applied, the jet mass scale calibration is derived using the same procedure in bins of $E_{\text{reco}}, \eta_{\text{det}}$, and $\log(m_{\text{reco}}/E_{\text{reco}})$. The jet mass calibration factor $c_{\text{JMS}} = 1/R_m$ is applied only to the mass of the jet, keeping the jet energy fixed and thus allowing the $p_T$ to vary. This factor is also a smooth function of the large-$R$ jet kinematics. The reconstructed large-$R$ jet kinematics are thus given by:

\[
\text{Jet Constituent Type} \quad \text{Unmodified} \quad \text{CS+SK} \quad \text{PFlow} \quad \text{TCC} \quad \text{UFO}
\]

\[
\text{Jet Mass} \quad \text{Soft Drop} \quad z_{\text{ud}} = 0.1, \beta = 0.0 \quad z_{\text{ud}} = 0.1, \beta = 0.1 \quad z_{\text{ud}} = 0.05, \beta = 0.0, N = 3 \quad z_{\text{ud}} = 0.05, \beta = 0.0, N = 4 \quad z_{\text{ud}} = 0.05, \beta = 0.0, N = 5 \quad z_{\text{ud}} = 0.05, \beta = 1.0, N = 3
\]

\[
\text{Recursive SD} \quad z_{\text{ud}} = 0.1, \beta = 0.0, N = 3 \quad z_{\text{ud}} = 0.05, \beta = 0.0, N = 3 \quad z_{\text{ud}} = 0.05, \beta = 0.0, N = 4 \quad z_{\text{ud}} = 0.05, \beta = 0.0, N = 5 \quad z_{\text{ud}} = 0.05, \beta = 0.0, N = 6 \quad z_{\text{ud}} = 0.05, \beta = 0.0, N = 7
\]

\[
\text{Bottom-up SD} \quad z_{\text{ud}} = 0.1, \beta = 0.0, N = 4 \quad z_{\text{ud}} = 0.05, \beta = 0.0, N = 4 \quad z_{\text{ud}} = 0.05, \beta = 0.0, N = 5 \quad z_{\text{ud}} = 0.05, \beta = 0.0, N = 6 \quad z_{\text{ud}} = 0.05, \beta = 0.0, N = 7 \quad z_{\text{ud}} = 0.05, \beta = 0.0, N = 8
\]

\[
\text{Purging} \quad z_{\text{ud}} = 0.1, \beta = 0.0, N = 4 \quad z_{\text{ud}} = 0.05, \beta = 0.0, N = 4 \quad z_{\text{ud}} = 0.05, \beta = 0.0, N = 5 \quad z_{\text{ud}} = 0.05, \beta = 0.0, N = 6 \quad z_{\text{ud}} = 0.05, \beta = 0.0, N = 7 \quad z_{\text{ud}} = 0.05, \beta = 0.0, N = 8
\]

\[
\text{Trimming} \quad z_{\text{ud}} = 0.1, \beta = 0.0, N = 4 \quad z_{\text{ud}} = 0.05, \beta = 0.0, N = 4 \quad z_{\text{ud}} = 0.05, \beta = 0.0, N = 5 \quad z_{\text{ud}} = 0.05, \beta = 0.0, N = 6 \quad z_{\text{ud}} = 0.05, \beta = 0.0, N = 7 \quad z_{\text{ud}} = 0.05, \beta = 0.0, N = 8
\]
Fig. 12 Pile-up dependence of the value of the fitted $W$ boson mass peak at low $p_T$. Jet $p_T$ and $\eta$ cuts before tagging are made using the ungroomed particle-level large-$R$ jet matched to each of the groomed reconstructed large-$R$ jets. The current baseline topocluster-based trimmed collection is indicated with a $\star$. The $z$-axis colour range is based on the difference of the baseline collection from a slope of 0. This makes differences between definitions more discernible than those between very unstable collections, which may have values beyond the axis range.

Fig. 13 Pile-up dependence of a $D_2$ cut on the $W$ boson jet selection efficiency at low $p_T$. Jet $p_T$ and $\eta$ cuts before tagging are made using the ungroomed particle-level large-$R$ jet matched to each of the groomed reconstructed large-$R$ jets. The current baseline topocluster-based trimmed collection is indicated with a $\star$. The $z$-axis colour range is based on the difference of the baseline collection from a slope of 0. This makes differences between definitions more discernible than those between very unstable collections, which may have values beyond the axis range.

\[
E_{\text{reco}} = c_{\text{JES}} E_0, \quad m_{\text{reco}} = c_{\text{JES}} c_{\text{JMS}} m_0, \quad \eta_{\text{reco}} = \eta_0 + \Delta \eta,
\]

\[
p_{T\text{reco}} = c_{\text{JES}} \sqrt{E_0^2 - c_{\text{JMS}} m_0^2 \cosh(\eta_0 + \Delta\eta)},
\]

where the quantities $E_0$, $m_0$ and $\eta_0$ refer to the jet properties prior to any calibration, but following the jet grooming procedure. The JMS correction is mostly similar for each of the four CS + SK UFO jet definitions which are studied, but differences in the size of the correction become largest for massive jets at high $p_T$. Figure 15 presents the average jet mass response $R_m$ for jets with a particle-level jet mass equal to that of the $W$ boson, for the four CS + SK UFO jet definitions which are calibrated. The response for large-$R$ jets with this mass is obtained by directly taking a profile through the smoothed response maps. High-$p_T$ trimmed jets require a...
Fig. 14  Ratio of the mean value of mass response in \( W \) jets to that in \( q/g \) jets at low \( p_T \). Kinematic selections before tagging are made using the ungroomed particle-level large-\( R \) jet matched to each of the groomed reconstructed large-\( R \) jets. The current baseline topocluster-based trimmed collection is indicated with a star.

Table 2  Summary of the jet reconstruction algorithms, jet-input-level pile-up mitigation algorithms, and grooming algorithms which were determined to merit calibration and further study. Several promising UFO-based definitions are calibrated, as well as other definitions which enable comparisons of the impact of varying different aspects of jet definitions.

| Category          | Input objects       | Grooming algorithm | Configuration |
|-------------------|---------------------|--------------------|---------------|
| Baseline          | LCW Topoclusters    | Trimmed            | \( R_{sub} = 0.2, f_{cut} = 0.05 \) |
| definitions       | TCCs                | Trimmed            | \( R_{sub} = 0.2, f_{cut} = 0.05 \) |
| Studied           | CS + SK UFOs        | SD                 | \( z_{cut} = 0.1, \beta = 1.0 \) |
| definitions       | CS + SK UFOs        | RSD                | \( z_{cut} = 0.05, \beta = 1.0, N = \infty \) |
| Additional        | UFOs                | Trimmed            | \( R_{sub} = 0.2, f_{cut} = 0.05 \) |
| definitions       | PFOs                | Trimmed            | \( R_{sub} = 0.2, f_{cut} = 0.05 \) |
|                   | UFOs                | SD                 | \( z_{cut} = 0.1, \beta = 1.0 \) |

smaller calibration factor than jets which are groomed using the SD, RSD or BUSD algorithms. This indicates that there are differences in the high-\( p_T \) behaviour of grooming algorithms: trimming removes more pile-up from jets at high \( p_T \), bringing the average JMS of these jets closer to particle level before calibration.

All figures where JES + JMS calibrations have been applied to the large-\( R \) jet four-vector are labelled ‘JES + JMS’.

7.2 Comparison of calibrated jet definition performance

7.2.1 Jet mass and \( p_T \) resolution

The expected large-\( R \) jet mass resolution, defined to be the 68% interquartile range divided by twice the median of the distribution, is shown in Fig. 16 for samples of signal jets. For these studies (as for all studies in this document), the baseline trimmed topocluster mass is used directly, rather than the combined mass [91] (which incorporates additional measurements from the inner tracking detector), allowing a direct comparison of the unmodified performance of the different jet definitions. In Fig. 16a and b, the resolution for all UFO jet definitions is shown to be better than for the baseline trimmed topocluster definition, particularly at high \( p_T \). The expected mass resolution of UFO jets is stable across the entire \( p_T \) spectrum. In the low-\( p_T \) region the mass resolution of UFO jets is typically similar to that of topocluster jets, while in the high-\( p_T \) region, it more closely follows the behaviour of TCC jets. For hadronically decaying high-\( p_T \) top quarks, UFOs improve the jet mass resolution relative to topocluster-based jets by 26%, and by 40% for high-\( p_T \) hadronically decaying \( W \) bosons.

In order to help factorise the performance gains from various sources, comparisons of the jet mass resolution are also provided for several other calibrated jet definitions. Figure 16c and d show a comparison of the four unmodified input object types using the trimming algorithm. In general,
Fig. 15 The jet mass response for UFO CS + SK large-\( R \) jets which have been groomed with a trimming, b soft-drop, c recursive soft-drop and d bottom-up soft-drop. The jet mass response is presented as a function of jet pseudorapidity for several values of the jet transverse momentum from 200 GeV to 2 TeV, for jets with a particle-level mass equal to the \( W \) boson mass. The mass responses for large-\( R \) jets with this mass are obtained by directly taking a profile through the smoothed response maps.

at high-\( p_T \) the mass resolution of top quarks is better than that of \( W \) bosons due to the fact that \( W \) bosons are lighter, and their decay products are typically more collimated, making the calorimeter granularity relevant at lower values of \( p_T \). UFO jets outperform topocluster and TCC jets for both \( W \) boson and top quark jets. PFlow jets are also found to be more performant than topocluster and TCC jets for top quark jets, although their performance deteriorates for highly boosted \( W \) bosons. The trimming and soft-drop algorithms are compared for UFO jets with and without CS + SK pile-up mitigation in Fig. 16e and f. The application of CS + SK does not significantly alter the mass resolution of trimmed UFO jets; however, it is found to improve the mass resolution for soft-drop jets at low \( p_T \) by nearly 40%.

The large-\( R \) jet \( p_T \) resolution for background jets is shown in Fig. 17, determined as the one-standard-deviation width of Gaussian fits to the \( R_{p_T} \) distributions divided by their fitted mean. The \( p_T \) resolution of trimmed topocluster jets is superior to that of either TCC trimmed jets or any of the UFO jet definitions studied. UFO jets do not use the LC correction because PFOs are reconstructed using topoclusters at the EM scale, which results in a degraded correlation between the particle-level and detector-level large-\( R \) jet \( p_T \). While TCC jets take topoclusters calibrated to the LC scale as input, the energy resolution of TCC trimmed jets is worse than for topocluster trimmed jets, while the UFO trimmed jet resolution is almost identical to the resolution of PFlow trimmed jets. This indicates that the energy resolution degradation of TCC is due to the inclusion of pile-up tracks in the energy sharing, since these are not included in the UFO implementation.
Fig. 16 The jet mass resolution for (a), (c), (e) W boson jets, and (b), (d), (f) top quark jets as a function of $p_T$. In (a), (b) the relative performance of the studied UFO definitions is compared with the current ATLAS baseline topocluster and TCC jets, while in (c), (d) only jet input object types are compared, and in (e), (f) the impact of pile-up mitigation is highlighted.
The jet $p_T$ resolution in dijet events. In (a) the relative performance of the studied UFO definitions is compared with the current ATLAS baseline topocluster and TCC jets, while in (b) only jet input object types are compared, and in (c) the impact of pile-up mitigation is highlighted.

### 7.2.2 Jet mass + JSS tagging performance

In this section, a comparison of the tagging performance of the calibrated jet definitions is reported. Instead of considering a single efficiency working point (Sect. 4.1), the tagging performance is studied using ROC curves. Figures 18 and 19 show the tagger background rejection as a function of the tagger signal efficiency, using the same jet mass + jet substructure taggers discussed in Sect. 4.1: a fixed mass-window cut with 68% signal efficiency is applied, and then a one-sided $D_2$ or $\tau_2$ cut is made to obtain the desired signal efficiency.

When tagging high-$p_T$, hadronically decaying $W$ bosons (Fig. 18), the considered UFO definitions bring significant improvement over the LCTopo and TCC definitions. At high $p_T$, UFOs outperform the baseline topocluster-based jet definition in terms of their background rejection by about 120% at a fixed signal-tagging efficiency of 50%. Use of the recursive or bottom-up soft-drop grooming algorithm is noted to further improve performance over the trimmed UFO definition by an additional 10% for a signal efficiency of 50%, and the application of CS + SK pile-up mitigation is also found to increase performance by roughly 10% when it is applied in conjunction with the soft-drop grooming algorithm.

When tagging high-$p_T$, hadronically decaying $W$ bosons (Fig. 18), the considered UFO definitions bring significant improvement over the LCTopo and TCC definitions. At high $p_T$, UFOs outperform the baseline topocluster-based jet definition in terms of their background rejection by about 120% at a fixed signal-tagging efficiency of 50%. Use of the recursive or bottom-up soft-drop grooming algorithm is noted to further improve performance over the trimmed UFO definition by an additional 10% for a signal efficiency of 50%, and the application of CS + SK pile-up mitigation is also found to increase performance by roughly 10% when it is applied in conjunction with the soft-drop grooming algorithm.

### 7.3 Data-to-simulation comparisons

Robust modelling of jet substructure is crucial to reduce uncertainties related to Monte Carlo modelling of parton showers in physics analyses that rely on jet-substructure-based techniques. To verify the accuracy of the simulation, predictions were generated at the detector level for several
Fig. 18 Background rejection as a function of signal efficiency for a tagger using the jet mass and $D_{2}$ for $W$ boson jets at (left) low $p_{T}$, and (right) high $p_{T}$. In (a), (b) the relative performance of the studied UFO definitions is compared with the current ATLAS baseline topocluster and TCC jets, while in (c), (d) only jet input object types are compared, and in (e), (f) the impact of pile-up mitigation is highlighted. The background rejection factor of the baseline topocluster-based trimmed collection at a fixed signal tagging efficiency of 50% is indicated with a $\star$. 
Fig. 19 Background rejection as a function of signal efficiency for a tagger using the jet mass and $\tau_3$ for top quark jets at (left) low $p_T$, and (right) high $p_T$. In a, b the relative performance of the studied UFO definitions is compared with the current ATLAS baseline topocluster and TCC jets, while in c, d only jet input object types are compared, and in e, f the impact of pile-up mitigation is highlighted.
Fig. 20 Data-to-simulation comparisons of 

(a) the groomed jet mass, 

(b) the number of constituents, 

(c) the groomed jet $D_2$ and 

(d) the groomed jet $\tau_{32}$. Only statistical uncertainties are displayed, and the statistical uncertainty of the simulation is negligible compared to that of the data. The ratio of simulation to data is provided in the lower panel of each figure.

Jet substructure observables in high-$p_T$ dijet events using PYTHIA and reconstructed using the full ATLAS detector simulation [60] based on GEANT4 [61]. The results are compared with the distributions observed in data collected during 2017. Events are selected using the lowest unprescaled single large-$R$ jet trigger. This trigger is fully efficient for ungroomed large-$R$ jets with $p_T > 600$ GeV. Data are required to pass a series of quality requirements and cleaning cuts. In addition, overlap removal and pile-up reweighting are applied. Events are required to have at least one jet with a groomed jet $p_T$ above 600 GeV, and all jets are required to have $p_T > 600$ GeV and $|\eta| < 1.2$. When studying the behaviour of $\tau_{32}$ and $D_2$, the jet mass is required to be greater than 40 GeV. Data and simulated events are required to pass the same event selection.

The observed data are compared with simulated dijet events in Fig. 20. The jet mass, number of jet constituents, $D_2$, and $\tau_{32}$ are studied. Only statistical uncertainties are displayed, and the statistical uncertainty of the simulation is negligible compared to that of the data. In general, the level of agreement between data and simulation for the UFO jets is similar to that of topocluster trimmed jets, indicating that this level of agreement is tolerable for general use on ATLAS. The exception to this is the number of constituents, which is known to be modelled poorly [66]. The modelling is improved for UFO jets relative to topocluster-based trimmed jets, particularly at large constituent multiplicities.

The background rejection for the mass + JSS taggers described in Sect. 6 is shown in Fig. 21 as a function of the large-$R$ jet $p_T$, where taggers are created for each $p_T$ bin, using the 50% signal efficiency working point. For the
Fig. 21 Data-to-simulation comparisons of the background rejection for groomed jets for (a) the mass + $D_2$ W tagger, and (b) the mass + $\tau_{32}$ top tagger.

$W$ tagger, agreement between data and simulation is similar for all jet definitions, while for the top taggers, agreement is slightly worse for UFO jets than for the topocluster trimmed definition.

8 Concluding remarks

The development of jet substructure techniques has enabled new searches and measurements, boosting the sensitivity of the Large Hadron Collider experiments to the physics of and beyond the Standard Model. This paper has presented a set of performance comparisons in order to determine the most promising large-$R$ jet definitions for use in future analyses, with a focus on comparing different jet input objects, pile-up mitigation algorithms and jet grooming algorithms.

A new type of jet input, called a Unified Flow Object, has been proposed which incorporates tracking information into jet substructure reconstruction by combining particle-flow reconstruction for low-$p_T$ particles and cluster splitting for particles at high $p_T$ and in dense environments. These UFO inputs can increase the background rejection of jet taggers across a wide kinematic range by up to 120% for a simple $W$ tagger at 50% signal efficiency, and up to 135% for a simple top tagger at 50% signal efficiency when compared with the current baseline trimmed topocluster large-$R$ jet definition.

The application of CS + SK pile-up mitigation has been shown to stabilise and augment performance as a function of the number of pile-up interactions, which will be crucial in the face of the difficult experimental conditions to come during future LHC data-taking periods. Pile-up mitigation increases the number of experimentally viable grooming configurations to include options which do not groom soft radiation aggressively enough to be considered with unmodified jet inputs.

Several promising grooming algorithms were compared using large-$R$ CS + SK UFO jets. Definitions incorporating soft-drop grooming and its extensions, recursive soft-drop and bottom-up soft-drop, all outperform the baseline ATLAS trimming configuration in terms of high-$p_T$ $W$ and top quark tagging using simple taggers. These collections are viable for general-purpose use in the challenging experimental conditions of the LHC only due to the improvements in jet inputs and pile-up mitigation algorithms. The soft-drop definition using $z_{\text{cut}} = 0.1$ and angular exponent $\beta = 1.0$ outperforms all other candidates when identifying high-$p_T$ top quarks, and is competitive to within 5–10% of the considered RSD and BUSD options when tagging boosted $W$ bosons. These jets also exhibit good pile-up stability and a tolerable sensitivity to topological effects, according to the metrics studied. This definition provides superior jet mass resolution for low-$p_T$ $W$ jets when compared with RSD and BUSD options. Due to its wide range of applicability, it is concluded that the CS + SK UFO soft-drop ($\beta = 1.0$, $z_{\text{cut}} = 0.1$) large-$R$ jet definition provides the best performance for use as a general-purpose jet definition in ATLAS physics analyses.

Acknowledgements We thank CERN for the very successful operation of the LHC, as well as the support staff from our institutions without whom ATLAS could not be operated efficiently. We acknowl-
edge the support of ANPCyT, Argentina; YerPhI, Armenia; ARC, Australia; BMWF and FWF, Austria; ANAS, Azerbaijan; SSTC, Belarus; CNPq and FAPESP, Brazil; NSERC, NRC, and CFIL, Canada; CERN: ANID, Chile; CAS, MOST and NSFC, China; COLCIENCIAS, Colombia; MSMT CR, MPO CR and VSC CR, Czech Republic; DNF and DNRSNCP, Denmark; IN2P3-CNRS and CEA-DRF/IRFU, France; SRNSFG, Georgia; BMBF, HGF and MPG, Germany; GSRT, Greece; RGC and Hong Kong SAR, China; ISF and Benoziyo Center, Israel; INFN, Italy; MEXT and JSPS, Japan; CNRST, Morocco; NWO, Netherlands; RCN, Norway; MNISW and NCN, Poland; FCT, Portugal; MNE/IFA, Romania; JINR; MES of Russia and NRC KI, Russian Federation; MESTD, Serbia; MSSR, Slovakia; ARRS and MIZS, Slovenia; DST/NRF, South Africa; MICINN, Spain; SRC and Wallace Foundation, Sweden; SERI, SNSF and Cantons of Bern and Geneva, Switzerland; MOST, Taiwan; TAEK, Turkey; STFC, United Kingdom; DOE and NSF, USA. In addition, individual groups and members have received support from BCKDF, CANARIE, Compute Canada, CRC and IVADO, Canada; Beijing Municipal Science & Technology Commission, China; COST, ERC, ERDF, Horizon 2020 and Marie Skłodowska-Curie Actions, European Union; Investissements d’Avenir Labex, Investissements d’Avenir IDEX and ANR, France; DFG and AvH Foundation, Germany; Herakleitos, Thales and Aristeia programmes co-financed by EU-ESF and the Greek NSRF, Greece; BSF-NSF and GIF, Israel; La Caixa Banking Foundation, CERCA Programme Generalitat de Catalunya and PROMETEO and GenT Programmes Generalitat Valenciana, Spain; Göran Gustafssons Stiftelse, Sweden; The Royal Society and Leverhulme Trust, United Kingdom. The crucial computing support from all WLCG partners is acknowledged gratefully, in particular from CERN, the ATLAS Tier-1 facilities at TRIUMF (Canada), NDGF (Denmark, Norway, Sweden), CC-IN2P3 (France), KIT/GridKA (Germany), INFN-CNAF (Italy), NL-T1 (The Netherlands), PIC (Spain), ASGC (Taiwan), RAL (UK) and BNL (USA), the Tier-2 facilities worldwide and large non-WLCG resource providers. Major contributors of computing resources are listed in Ref. [93].

Data Availability Statement This manuscript has no associated data or the data will not be deposited. [Authors’ comment: All ATLAS scientific output is published in journals, and preliminary results are made available in Conference Notes. All are openly available, without restriction on use by external parties beyond copyright law and the standard conditions agreed by CERN. Data associated with journal publications are also made available: tables and data from plots (e.g. cross section values, likelihood profiles, selection efficiencies, cross section limits, …) are stored in appropriate repositories such as HEPDATA (http://hepdata.cedar.ac.uk/). ATLAS also strives to make additional material related to the paper available that allows a reinterpretation of the data in the context of new theoretical models. For example, an extended encapsulation of the analysis is often provided for measurements in the framework of RIVET (http://rivet.hepforge.org/).]

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, and indicate if changes were made. You must not distribute this article, or any part of it, for any commercial purpose. The full text of this article is available at http://creativecommons.org/licenses/by/4.0/.

Funded by SCOAP3.

References

1. R. Kogler et al., Jet substructure at the large hadron collider: experimental review. Rev. Mod. Phys. 91, 045003 (2019). arXiv:1803.06991 [hep-ex]
2. A.J. Larkoski, I. Moult, B. Nachman, Jet substructure at the Large Hadron Collider: a review of recent advances in theory and machine learning. Phys. Rep. 841, 1 (2020). arXiv:1709.04464 [hep-ph]
3. ATLAS Collaboration, Search for heavy particles decaying into top-quark pairs using lepton-plus-jets events in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. Eur. Phys. J. C 78, 565 (2018). arXiv:1804.10823 [hep-ex]
4. ATLAS Collaboration, Search for supersymmetry in final states with missing transverse momentum and multiple b-jets in proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. JHEP 06, 107 (2018). arXiv:1711.01901 [hep-ex]
5. ATLAS Collaboration, Search for Dark Matter Produced in Association with a Higgs Boson Decaying to $bb$ using 36 $fb^{-1}$ of pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS Detector. Phys. Rev. Lett. 119, 181804 (2017). arXiv:1707.01302 [hep-ex]
6. ATLAS Collaboration, Search for chargino and neutralino production in final states with a Higgs boson and missing transverse momentum at $\sqrt{s} = 13$ TeV with the ATLAS detector. Phys. Rev. D 100, 012006 (2019). arXiv:1812.09432 [hep-ex]
7. ATLAS Collaboration, Search for diboson resonances in hadronic final states in 139 $fb^{-1}$ of pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. JHEP 09, 091 (2019). arXiv:1906.08589 [hep-ex]
8. CMS Collaboration, A multi-dimensional search for new heavy resonances decaying to boosted WW, WZ, or ZZ bosons pairs in the dijet final state at 13 TeV. Eur. Phys. J. C 80, 237 (2020). arXiv:1906.05977 [hep-ex]
9. CMS Collaboration, Search for production of Higgs boson pairs in the four b quark final state using large-area jets in proton–proton collisions at $\sqrt{s} = 13$ TeV. JHEP 01, 040 (2019). arXiv:1808.01473 [hep-ex]
10. CMS Collaboration, Search for single production of vector-like quarks decaying to a top quark and a W boson in proton–proton collisions at $\sqrt{s} = 13$ TeV. Eur. Phys. J. C 79, 90 (2019). arXiv:1809.08597 [hep-ex]
11. CMS Collaboration, Search for low mass vector resonances decaying into quark–antiquark pairs in proton–proton collisions at $\sqrt{s} = 13$ TeV. Phys. Rev. D 100 (2019). arXiv:1909.04114 [hep-ex]
12. CMS Collaboration, Inclusive search for highly boosted Higgs bosons decaying to bottom quark-antiquark pairs in proton–proton collisions at $\sqrt{s} = 13$ TeV. JHEP 12, 085 (2020). arXiv:2006.13251 [hep-ex]
13. ATLAS Collaboration, Measurement of the Soft-Drop Jet Mass in pp Collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. Phys. Rev. Lett. 121, 092001 (2018). arXiv:1711.08341 [hep-ex]
14. ATLAS Collaboration, Measurement of soft-drop jet observables in pp collisions with the ATLAS detector at $\sqrt{s} = 13$ TeV. Phys. Rev. D 101, 052007 (2020). arXiv:1912.09837 [hep-ex]
15. ATLAS Collaboration, Measurement of the jet mass in high transverse momentum $Z(\rightarrow bb)\gamma$ production at $\sqrt{s} = 13$ TeV using the ATLAS detector. Phys. Lett. B 812, 135991 (2021). arXiv:1907.07093 [hep-ex]
16. ATLAS Collaboration, Measurement of the Lund Jet Plane Using Charged Particles in 13 TeV Proton-Proton Collisions with the ATLAS Detector. Phys. Rev. Lett. 124, 222002 (2020). arXiv:2004.03540 [hep-ex]
17. CMS Collaboration, Measurement of the jet mass distribution and top quark mass in hadronic decays of boosted top quarks in pp collisions at $\sqrt{s} = 13$ TeV. Phys. Rev. Lett. 124 (2020). arXiv:1911.03800 [hep-ex]
18. CMS Collaboration, Measurements of the differential jet cross section as a function of the jet mass in dijet events from proton-proton collisions at $\sqrt{s} = 13$ TeV. JHEP 11, 113 (2018). arXiv:1807.05974 [hep-ex]

19. CMS Collaboration, Measurement of jet substructure observables in $t\bar{t}$ events from proton–proton collisions at $\sqrt{s} = 13$ TeV. Phys. Rev. D 98 (2018). arXiv:1808.07340 [hep-ex]

20. STAR Collaboration, Measurement of Groomed Jet Substructure Observables in pp Collisions at $\sqrt{s} = 200$ GeV with STAR (2020). arXiv:2003.02114 [hep-ex]

21. LHCb Collaboration, Study of $J/\psi$ production in jets. Phys. Rev. Lett. 118, 192001 (2017). arXiv:1701.05116 [hep-ex]

22. LHCb Collaboration, Measurement of charged hadron production in Z-tagged jets in proton-proton collisions at $\sqrt{s} = 8$ TeV. Phys. Rev. Lett. 123, 232001 (2019). arXiv:1904.08878 [hep-ex]

23. ALICE Collaboration, Exploration of jet substructure using iterative declustering in pp and Pb–Pb collisions at LHC energies. Phys. Lett. B 802, 135227 (2020). arXiv:1905.02512 [nucl-ex]

24. ALICE Collaboration, First direct observation of the dead-cone effect. Nucl. Phys. A 1005, 121905 (2021). arXiv:2004.05968 [hep-ex]

25. M. Dasgupta, L. Magnnea, G.P. Salam, Non-perturbative QCD effects in jets at hadron colliders. JHEP 02, 055 (2008). arXiv:0712.3014 [hep-ph]

26. M. Cacciari, G.P. Salam, G. Soyez, The anti-kt jet clustering algorithm. JHEP 04, 005 (2008). arXiv:0802.1188 [hep-ph]

27. M. Cacciari, J. Rojo, G.P. Salam, G. Soyez, Jet reconstruction in heavy ion collisions. Eur. Phys. J. C 71, 1539 (2011). arXiv:1010.1759 [hep-ph]

28. G. Soyez, G.P. Salam, J. Kim, S. Dutta, M. Cacciari, Pileup subtraction for jets. Phys. Rev. Lett. 110, 162001 (2013). arXiv:1211.2811 [hep-ph]

29. High-luminosity large hadron collider (HL-LHC): technical design report V. 0.1, 4/2017, ed. by G. Apollinari et al (2017).

30. CMS Collaboration, Measurements of jet substructure observables in $t\bar{t}$ events from proton–proton collisions at $\sqrt{s} = 8$ TeV. Phys. Rev. Lett. 123, 232001 (2019). arXiv:1904.08878 [hep-ex]

31. ATLAS Collaboration, Performance of jet substructure techniques. JHEP 11, 113 (2018). arXiv:1807.05974 [hep-ex]

32. ATLAS Collaboration, Jet reconstruction and performance using FastJet. JHEP 02, 055 (2008). arXiv:0712.3014 [hep-ph]

33. ATLAS Collaboration, JHEP 02, 055 (2008). arXiv:0712.3014 [hep-ph]

34. ATLAS Collaboration, ATLAS Pythia 8 tunes to 7 TeV data, ATL-CONF-2014-021 (2014). https://cds.cern.ch/record/1966419

35. ATLAS Collaboration, Performace of Boosted W Boson Identification with the ATLAS Detector, ATL-PHYS-PUB-2014-004 (2014). https://cds.cern.ch/record/1690048

36. ATLAS Collaboration, Performance of Boosted W Boson Identification with the ATLAS Detector, ATL-PHYS-PUB-2014-004 (2014).

37. ATLAS Collaboration, The anti-kt jet clustering algorithm. JHEP 02, 084 (2010). arXiv:0912.1342 [hep-ph]

38. ATLAS Collaboration, The anti-kt jet clustering algorithm. JHEP 02, 084 (2010). arXiv:0912.1342 [hep-ph]

39. ATLAS Collaboration, Identification of boosted, hadronically decaying $W$ bosons and comparisons with ATLAS data taken at $\sqrt{s} = 8$ TeV. Eur. Phys. J. C 76, 154 (2016). arXiv:1510.05821 [hep-ex]
65. ATLAS Collaboration, Early Inner Detector Tracking Performance in the 2015 Data at $\sqrt{s} = 13$ TeV, ATL-PHYS-PUB-2015-051 (2015) https://cds.cern.ch/record/2110140
66. ATLAS Collaboration, Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1, Eur. Phys. J. C 77, 490 (2017) arXiv:1603.02934 [hep-ex]
67. ATLAS Collaboration, Jet energy scale and resolution measured in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector (2020). arXiv:2007.02645 [hep-ex]
68. CMS Collaboration, Pileup mitigation at CMS in 13 TeV data. JINST 15, P09018 (2020). arXiv:2003.00503 [hep-ex]
69. CMS Collaboration, Particle-flow reconstruction and global event description with the CMS detector. JINST 12, P10003 (2017). arXiv:1706.04965 [hep-ex]
70. G. Soyez, Jet areas as a tool for background subtraction, in 25th Winter Workshop on Nuclear Dynamics (2009) arXiv:0905.2851 [hep-ph]
71. S.D. Ellis, D.E. Soper, Successive combination jet algorithm for hadron collisions. Phys. Rev. D 48, 3160 (1993). arXiv:hep-ph/9305266
72. ATLAS Collaboration, Constituent-level pile-up mitigation techniques in ATLAS, ATLAS-CONF-2017-065 (2017) https://cds.cern.ch/record/2281055
73. Y.L. Dokshitzer, G. Leder, S. Moretti, B. Webber, Better jet clustering algorithms. JHEP 08, 001 (1997). arXiv:hep-ph/9707323
74. M. Wobisch, T. Wengler, 'Hadronization Corrections to Jet Cross Sections in Deep-Inelastic Scattering', Workshop on Monte Carlo Generators for HERA Physics (Plenary Starting Meeting) 270 (1998). arXiv:hep-ph/9907280
75. S. Marzani, L. Schunk, G. Soyez, A study of jet mass distributions with grooming. JHEP 07, 132 (2017). arXiv:1704.02210 [hep-ph]
76. S. Marzani, L. Schunk, G. Soyez, The jet mass distribution after soft drop. Eur. Phys. J. C 78, 96 (2018). arXiv:1712.05105 [hep-ph]
77. C. Frye, A.J. Larkoski, M.D. Schwartz, K. Yan, Factorization for groomed jet substructure beyond the next-to-leading logarithm. JHEP 07, 064 (2016). arXiv:1603.09338 [hep-ph]
78. C. Frye, A.J. Larkoski, M.D. Schwartz, K. Yan, Precision physics with pile-up insensitive observables (2016). arXiv:1603.06375 [hep-ph]
79. Z.-B. Kang, K. Lee, X. Liu, F. Ringer, Soft drop groomed jet angularities at the LHC. Phys. Lett. B 793, 41 (2019). arXiv:1811.06983 [hep-ph]
80. Z.-B. Kang, K. Lee, X. Liu, F. Ringer, The groomed and ungroomed jet mass distribution for inclusive jet production at the LHC. JHEP 10, 137 (2018). arXiv:1803.03645 [hep-ph]
81. A.H. Hoang, S. Mantry, A. Pathak, I.W. Stewart, Extracting a short distance top mass with light grooming. Phys. Rev. D 100, 074021 (2019). arXiv:1708.02586 [hep-ph]
82. ATLAS Collaboration, Performance of top-quark and W-boson tagging with ATLAS in Run 2 of the LHC. Eur. Phys. J. C 79, 375 (2019). arXiv:1808.07858 [hep-ex]
83. CMS Collaboration, Identification techniques for highly boosted W bosons that decay into hadrons. JHEP 12, 017 (2014). arXiv:1410.4227 [hep-ex]
84. CMS Collaboration, Identification of heavy, energetic, hadronically decaying particles using machine-learning techniques. JINST 15, P09018 (2020). arXiv:2004.08262 [hep-ex]
85. A.J. Larkoski, G.P. Salam, J. Thaler, Energy correlation functions for jet substructure. JHEP 06, 108 (2013). arXiv:1305.0007 [hep-ph]
86. A.J. Larkoski, I. Moulit, D. Neill, Power counting to better jet observables. JHEP 12, 009 (2014). arXiv:1409.6298 [hep-ph]
87. A.J. Larkoski, I. Moulit, D. Neill, Analytic boosted boson discrimination. JHEP 05, 117 (2016). arXiv:1507.03018 [hep-ph]
88. J. Thaler, K. Van Tilburg, Identifying boosted objects with N-subjettiness. JHEP 03 (2011). ISSN: 1029-8479
89. J. Thaler, K. Van Tilburg, Maximizing boosted top identification by minimizing N-subjettiness. JHEP 02 (2012). ISSN: 1029-8479
90. ATLAS Collaboration, Jet mass and substructure of inclusive jets in $\sqrt{s} = 7$ TeV pp collisions with the ATLAS experiment. JHEP 05, 128 (2012). arXiv:1203.4606 [hep-ex]
91. ATLAS Collaboration, In situ calibration of large-radius jet energy and mass in 13 TeV proton-proton collisions with the ATLAS detector. Eur. Phys. J. C 79, 135 (2019). arXiv:1807.09477 [hep-ex]
92. ATLAS Collaboration, Monte Carlo Calibration and Combination of In-situ Measurements of Jet Energy Scale, Jet Energy Resolution and Jet Mass in ATLAS, ATLAS-CONF-2015-037 (2015). https://cds.cern.ch/record/2044941
93. ATLAS Collaboration, ATLAS Computing Acknowledgements, ATL-SOFT-PUB-2020-001. https://cds.cern.ch/record/2717821
Graduate School of Science, Osaka University, Osaka, Japan
Department of Physics, University of Oslo, Oslo, Norway
Department of Physics, Oxford University, Oxford, UK
LPNHE, Sorbonne Université, Université de Paris, CNRS/IN2P3, Paris, France
Department of Physics, University of Pennsylvania, Philadelphia, PA, USA
Konstantinov Nuclear Physics Institute of National Research Centre “Kurchatov Institute”, PNPI, St. Petersburg, Russia
Department of Physics and Astronomy, University of Pittsburgh, Pittsburgh, PA, USA

Laboratório de Instrumentação e Física Experimental de Partículas - LIP, Lisbon, Portugal; Departamento de Física, Faculdade de Ciências, Universidade de Lisboa, Lisbon, Portugal; Departamento de Física, Universidad de Coimbra, Coimbra, Portugal; Centro de Física Nuclear da Universidade de Lisboa, Lisbon, Portugal; Departamento de Física, Universidad del Minho, Braga, Portugal; Departamento de Física Teórica y del Cosmos, Universidad de Granada, Granada, Spain; Dep Física and CEFITEC of Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, Caparica, Portugal; Instituto Superior Técnico, Universidade de Lisboa, Lisbon, Portugal

Institute of Physics of the Czech Academy of Sciences, Prague, Czech Republic
Czech Technical University in Prague, Prague, Czech Republic
Faculty of Mathematics and Physics, Charles University, Prague, Czech Republic
Particle Physics Department, Rutherford Appleton Laboratory, Didcot, UK
IRFU, CEA, Université Paris-Saclay, Gif-sur-Yvette, France
Santa Cruz Institute for Particle Physics, University of California Santa Cruz, Santa Cruz, CA, USA
Department of Física, Pontificia Universidad Católica de Chile, Santiago, Chile; Universidad Andres Bello, Departament de Física, Santiago, Chile; Instituto de Alta Investigación, Universidad de Tarapacá, Santiago, Chile; Departamento de Física, Universidad Técnica Federico Santa María, Valparaiso, Chile

Universidade Federal de São João del Rei (UFSJ), São João del Rei, Brazil
Department of Physics, University of Washington, Seattle, WA, USA
Department of Physics and Astronomy, University of Sheffield, Sheffield, UK
Department of Physics, Shinshu University, Nagano, Japan
Department Physik, Universität Siegen, Siegen, Germany
Department of Physics, Simon Fraser University, Burnaby, BC, Canada
SLAC National Accelerator Laboratory, Stanford, CA, USA
Physics Department, Royal Institute of Technology, Stockholm, Sweden
Departments of Physics and Astronomy, Stony Brook University, Stony Brook, NY, USA
Department of Physics and Astronomy, University of Sussex, Brighton, UK
School of Physics, University of Sydney, Sydney, Australia
Institute of Physics, Academia Sinica, Taipei, Taiwan

Faculty of Science, Tbilisi University, Tbilisi, Georgia; High Energy Physics Institute, Tbilisi State University, Tbilisi, Georgia
Department of Physics, Technion, Israel Institute of Technology, Haifa, Israel
Raymond and Beverly Sackler School of Physics and Astronomy, Tel Aviv University, Tel Aviv, Israel
Department of Physics, Aristotle University of Thessaloniki, Thessaloniki, Greece
International Center for Elementary Particle Physics and Department of Physics, University of Tokyo, Tokyo, Japan
Graduate School of Science and Technology, Tokyo Metropolitan University, Tokyo, Japan
Department of Physics, Tokyo Institute of Technology, Tokyo, Japan
Tomsk State University, Tomsk, Russia
Department of Physics, University of Toronto, Toronto, ON, Canada
Department of Physics and Astronomy, University of California Irvine, Irvine, CA, USA
Department of Physics and Astronomy, University of Uppsala, Uppsala, Sweden
Department of Physics, University of Illinois, Urbana, IL, USA
Instituto de Física Corpuscular (IFIC), Centro Mixto Universidad de Valencia-CSIC, Valencia, Spain
Department of Physics, University of British Columbia, Vancouver, BC, Canada
