Measurement of cross-sections for production of a $Z$ boson in association with a flavor-inclusive or doubly $b$-tagged large-radius jet in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS experiment

The ATLAS Collaboration

We present measurements of cross-sections for production of a leptonically decaying $Z$ boson in association with a large-radius jet in 13 TeV proton-proton collisions at the LHC, using 36 fb$^{-1}$ of data from the ATLAS detector. Integrated and differential cross-sections are measured at particle-level in both a flavor-inclusive and a doubly $b$-tagged fiducial phase-space. The large-radius jet mass and transverse momentum, its kinematic relationship to the $Z$ boson, and the angular separation of $b$-tagged small-radius track-jets within the large-radius jet are measured. This measurement constitutes an important test of perturbative quantum chromodynamics in kinematic and flavor configurations relevant to several Higgs boson and beyond-Standard-Model physics analyses. The results highlight issues with modeling of additional hadronic activity in the flavor-inclusive selection, and a distinction between flavor-number schemes in the $b$-tagged phase-space.

© 2022 CERN for the benefit of the ATLAS Collaboration.
Reproduction of this article or parts of it is allowed as specified in the CC-BY-4.0 license.
1 Introduction

Since the proposal of a Higgs $\rightarrow b\bar{b}$ discovery channel based on the structure of a high-momentum jet [1], “boosted jet” methods have been a major feature of Higgs and other experimental analyses at the LHC [2–7]. In these, a large-radius jet (large-$R$ jet) is reconstructed, and then decomposed into smaller-radius (small-$R$) subjets whose structure enables identification of resonances whose decay products have been collimated by their parent’s large momentum. Such methods are also of high interest in searches for new physics, both because a high-mass new particle decaying into resonances naturally generates high-momentum merged jets, and because the high-momentum regime is particularly sensitive to modifications to Standard Model (SM) dynamics by new physics.

A good understanding of hard, collinear parton splittings in quantum chromodynamics (QCD) is key to measurements using large-$R$ jet methods, as these constitute the dominant background to boosted resonance signatures. Monte Carlo (MC) event-simulation methods include approximations such as factorization of partonic splittings in parton-shower algorithms, and slicing of emission phase-space between matrix-element and parton-shower sources. The extreme dynamics of boosted topologies include aspects of both collinear and high-$p_T$ physics, presenting a test for such factorizations.

The role of heavy-quark masses is also important: contradictory inclusions of mass effects in calculations for parton-shower MC generators and parton distribution functions (PDFs) mean that there is as yet no unambiguously more-correct simulation strategy for heavy-flavor production [8]. Additionally, the modeling of gluon-splitting into heavy quarks may benefit from renormalization-scale choices different from those developed for the more common cases of gluon emission from light quarks and gluons.

Due to these factors, empirical comparisons between MC models and collider measurements of boosted $b\bar{b}$ production are key to understanding and improving the validity of parton-shower MC simulations in this event topology, and closely related ones such as high-$p_T$ associated Higgs-boson production. In addition, this phase-space is also relatively novel for light-jet production, whose modeling can be similarly informed by equivalent measurements without the $b$-tagging requirement.

This analysis follows the $VH(b\bar{b})$ event-selection strategy of requiring boosted-jet production in association with a leptonically decaying vector boson, which is effective at reducing QCD backgrounds and provides an additional, experimentally clean proxy for each event’s characteristic momentum-transfer scale [1]. Previous measurements of the $Z+b(\bar{b})$ process, in a resolved rather than boosted event topology, were made using $\sqrt{s} = 7$ TeV proton–proton collisions by ATLAS, CMS and LHCb [9–13], at 8 TeV by CMS [14], and at 13 TeV by ATLAS and CMS [15, 16].

Additionally, studies of correlations between $B$ and $\bar{B}$ mesons have been performed by CMS at 7 TeV [17], and various properties of gluon splitting at 13 TeV have been measured by ATLAS [18]. In these papers the differential cross-section as a function of $\Delta R(b, \bar{b})$, the angular separation between the $b$-jets, was measured, and was found to be mismodeled by MC generators in the small-$\Delta R(b, \bar{b})$ region typical of gluon splitting into $b\bar{b}$. Given the importance of $\Delta R(b, \bar{b})$ modeling to techniques for reconstruction of boosted Higgs-bosons, this variable is also a target of this analysis. The higher 13 TeV center-of-mass energy produces a larger boosted-event population than in the 7 TeV studies, and in contrast to the recent ATLAS $Z+b\bar{b}$ cross-section measurement at 13 TeV, use is also made of small-radius charged-particle jets that allow precise measurement of the small angular separations.

In this paper, we present measurements of cross-sections for the production of a leptonically-decaying $Z$-boson in association with a large-$R$ jet, using data taken by the ATLAS detector at the LHC [19, 20] in
2015–2016. The measured cross-sections are differential in kinematic variables of the large-$R$ jet. An additional phase-space is defined by the requirement that the large-$R$ jet be doubly $b$-tagged: total and differential cross-sections, including as a function of the angular separation of the $b$-tagged subjets, are also measured in this phase-space. These measurements provide an important test of perturbative QCD in the boosted regime, including contributions both where a high-energy gluon splits to give a $b\bar{b}$ pair carrying a significant fraction of the jet momentum, and where secondary processes in which two $b$-tagged momentum flows exist within the jet but are not its dominant kinematic components.

The relevant perturbative-QCD issues are summarized in Section 2, the ATLAS detector is described in Section 3, and the MC event samples used in the analysis and for comparison with the resulting measurements are discussed in Section 4. The physics-object definitions at the reconstruction and particle levels are described in Section 5, followed by the event selection and observable definitions in Section 6. The correction for detector biases is treated in Section 7, and the sources and estimation of systematic uncertainties are described in Section 8. Finally, the detector-corrected observables are compared with current MC predictions in Section 9.

2 Theory context

Heavy-flavor partons in the initial state of a hard-scattering process are understood to arise mainly from perturbative gluon splittings into $b\bar{b}$ and $c\bar{c}$ quark–antiquark pairs, formalized as DGLAP QCD evolution [21–23]. But there is an ambiguity in this evolution in the usual factorization picture of perturbative QCD calculations, as to whether the emergence of heavy flavor is to be isolated into the partonic cross-section $\hat{\sigma}$, or is also permitted in the evolution of the PDFs which encode the initial-state proton structure. At present, this separation is strongly tied to the treatment of the heavy quark as having a finite or zero mass [8].

The picture without heavy-quark production in the PDF evolution – here, the absence of $b$-quarks – is termed the four-flavor number scheme (4FNS). In this, the $b$-quark density in the PDF is set to zero, and so the perturbative generation of initial-state $b$-quarks comes from explicit gluon splitting into a $b\bar{b}$ pair in the partonic matrix-element [24], usually including $b$-quark mass effects. A consequence is that in the four-flavor scheme there are always at least two participating $b$-quarks, although they may fall outside the experimental acceptance. By contrast, in the five-flavor number scheme (5FNS) the PDF evolution can generate initial-state $b$-quarks – again through gluon splitting, but now internalized in the functional form of the $b$-quark PDF. This allows matrix-element amplitudes where it is possible for only one $b$-quark to participate in the hard-scatter [24]. While the 5FNS initially seems the more complete treatment, its treatment of the initial-state $b$-quark is purely longitudinal (whereas gluon splitting in the matrix element generates transverse momentum), and to avoid non-cancellation of higher-order soft divergences the initial-state $b$-quark is currently treated as massless in standard PDF approaches [25]. In a hypothetical all-orders calculation the two schemes would give the same results, but for a truncated perturbation expansion they generally give different predictions. There are arguments in favour of both approaches: the 4FNS allows for transverse momentum exchange through the initial-state heavy quarks and hence might be expected to describe event kinematics better, while the 5FNS is able to make use of higher-order DGLAP resummation calculations in the PDF evolution, which are not present in matrix elements matched to parton showers. The importance of mass effects is expected to become less important for process scales $Q \gg m_b$, suggesting that higher-accuracy predictions may be obtained using the 5FNS in boosted event configurations [8, 26–28]. Recent computations such as the computation of $Z + b$ production...
at $O(\alpha_s^3)$ [29], and the NLO “fusing” scheme developed by Sherpa [30] combine desirable aspects of both schemes, and should also be dominated by the 5FNS in boosted phase-space. It is therefore important to compare experimental measurements of $b$-quark production with predictions using both of these schemes, to empirically test the accuracy and predictivity of the available theoretical approaches.

Theoretical uncertainties also arise in the production of $b$-quarks in the partonic final state. The usual parton shower formulation for parton or dipole splitting is derived in the collinear-emission limit, using the $p_T$ of the splitting as the characteristic (renormalization) scale, but this heuristic scale choice is only well-motivated for gluon-emission splitting functions [31]. The scale choice for gluon splitting, especially into heavy quarks, is hence an extrapolation requiring empirical testing. This is of great importance since uncertainties in heavy-flavor production by gluon splitting are a leading systematic limitation on the sensitivity of Higgs boson decays into $b\bar{b}$ in the $t\bar{t}H$, $VH$ (where $V$ is either a $W$ or $Z$ boson) and gluon-fusion channels [32, 33] — particularly in boosted-Higgs configurations where the two $b$-quarks are relatively collinear, similar to the gluon-splitting kinematics [34].

3 ATLAS detector

The ATLAS detector [19] is a multipurpose particle detector with a forward/backward-symmetric cylindrical geometry. The detector has a nearly $4\pi$ coverage in solid angle and consists of an inner tracking detector, electromagnetic and hadronic calorimeters, and a muon spectrometer.

The inner-detector system (ID) is immersed in a 2 T axial magnetic field and provides charged-particle tracking in the range $|\eta| < 2.5$. The high-granularity silicon pixel detector covers the vertex region and typically provides four measurements per track, the first hit normally being in the insertable B-layer (IBL) installed before Run 2 [35, 36]. It is followed by the silicon microstrip tracker (SCT), which usually provides eight measurements per track. These silicon detectors are complemented by the transition radiation tracker (TRT), which enables radially extended track reconstruction up to $|\eta| = 2.0$. The TRT also provides electron-identification information based on the fraction of hits (typically 30 in total) above a higher energy-deposit threshold corresponding to transition radiation.

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

The muon spectrometer (MS) comprises separate trigger and high-precision tracking chambers measuring the deflection of muons in a magnetic field generated by the superconducting air-core toroidal magnets. The field integral of the toroids ranges between 2.0 and 6.0 T m across most of the detector. Three layers of precision chambers, each consisting of layers of monitored drift tubes, covers the region $|\eta| < 2.7$.

1 ATLAS uses a right-handed coordinate system with its origin at the nominal interaction point (IP) in the center of the detector and the z-axis along the beam pipe. The x-axis points from the IP to the center of the LHC ring, and the y-axis points upward. Cylindrical coordinates $(r, \phi)$ are used in the transverse plane, $\phi$ being the azimuthal angle around the beam pipe. The pseudorapidity is defined in terms of the polar angle $\theta$ as $\eta = -\ln \tan(\theta/2)$. Angular distance is measured in units of $\Delta R = \sqrt{(\Delta \eta)^2 + (\Delta \phi)^2}$. 
complemented by cathode-strip chambers in the forward region, where the background is highest. The muon trigger system covers the range $|\eta| < 2.4$ with resistive-plate chambers in the barrel, and thin-gap chambers in the endcap regions.

A two-level trigger system is used to select events for further analysis [37, 38]. The first-level trigger is implemented in hardware and utilizes partial detector information to accept events at a rate of 100 kHz. The high-level trigger is based on software and reduces the rate of accepted events to 1 kHz. An extensive software suite [39] is used in the reconstruction and analysis of real and simulated data, in detector operations, and in the trigger and data-acquisition systems of the experiment.

4 Data and MC event samples

The data used in this measurement were collected during the LHC 2015 and 2016 $p p$-collision runs at $\sqrt{s} = 13$ TeV, corresponding to an integrated luminosity of 36.1 fb$^{-1}$. The uncertainty in the combined 2015–2016 integrated luminosity is 2.1% [40], obtained using the LUCID-2 detector [41] for the primary luminosity measurements. Analyzed events were required to have all ATLAS subdetectors fully operational, and stable beam conditions.

MC-simulated event samples were used in this analysis to estimate the contamination from background processes, correct the data from reconstruction level to particle level (unfolding), and for comparisons to the unfolded data. Four processes are considered in detail for this analysis: production of $Z$+jets, $W$+jets, $t\bar{t}$, and electroweak diboson events. Single top-quark production was shown to make a negligible contribution, as its acceptance is similar to the already small $W$+jets contribution, and its cross-section is orders of magnitude smaller. QCD multijet events were shown to be negligible using a data-driven method.

The $Z$+jets signal and $W$+jets background were simulated at next-to-leading-order (NLO) accuracy with the SHERPA 2.2.1 [42, 43] MC generator, matching additional hard parton emissions [44] to the parton shower algorithm based on the Catani–Seymour dipole formalism [45]. The MEPS@NLO prescription [46–49] was used with a merging threshold of 20 GeV to provide merged matrix-element and parton-shower calculations accurate at NLO in QCD for up to two additional partons and accurate at LO for up to four additional partons; the virtual QCD matrix-element components at NLO accuracy were provided by the OpenLoops library [50, 51]. In this configuration, $g \rightarrow b\bar{b}$ splittings can originate either from the matrix element or from the parton shower, depending on the transverse scale of the splitting relative to the CKKW merging scale [47, 52]. As the matrix-element portions of this calculation include $g \rightarrow b\bar{b}$ splittings, it is expected that the majority of boosted jets in this sample will have been initiated by matrix-element rather than parton-shower splittings, but that the $b\bar{b}$ splitting within them can arise from either matrix-element or parton-shower modeling. The 5FNS NNPDF3.0nnLO PDF set [53] with $\alpha_s(m_Z) = 0.118$ was used, in conjunction with the SHERPA authors’ standard set of tuned MC parameter values, referred to as the “tune”. The samples were normalized to the NNLO inclusive $Z/W$ cross-sections [54].

An alternative LO QCD $Z$+jets sample was simulated using MadGraph 2.2.2 [55] with up to four additional partons at matrix-element level, and using the NNPDF2.3LO set of PDFs [56]. This was interfaced with Pythia 8.186 [57] for modeling of the parton shower and underlying event, with use of the CKKW-L merging procedure [58, 59], and with bottom and charm hadron decays corrected by EvtGen 1.2.0 [60]. The A14 tune [61] and the 5FNS NNPDF2.3LO PDF set [56] with $\alpha_s = 0.13$ were used by Pythia 8. This sample was also normalized to the NNLO inclusive cross-section for use in MC-based background estimation.
The $t\bar{t}$ background was simulated using the *POWHEG-BOX v2* HVQ [62–65] generator at NLO with the CT10 PDF [66], and matched to the *PYTHIA* 8.186 [57] parton shower and hadronization with the A14 tune [61]. The top-quark mass was set to 172.5 GeV, and the $h_{\text{damp}}$ parameter – which controls the $p_T$ of the first additional emission beyond the Born configuration – was set to the mass of the top quark. This sample was normalized to the $t\bar{t}$ NNLO+NNLL cross-section [67].

The diboson processes ($WW$, $WZ$, and $ZZ$, with one of the bosons decaying hadronically and the other leptonically) were simulated using *SHERPA* 2.1.1 with a MEPS@NLO configuration similar to that used in the $Z+$jets and $W+$jets processes described above. The CT10NLO PDF set [68] with $\alpha_s = 0.118$ was used, with the corresponding *SHERPA* parton-shower tune.

All these MC event samples were processed through the ATLAS Geant4-based detector simulation [69, 70] and digitization system to produce inputs to object reconstruction equivalent to those from the detector data-stream in collision events. Pile-up — multiple $pp$ collisions in each hard-interaction bunch-crossing, as well as detector-response effects due to surrounding bunch-crossings — was emulated by pre-digitization overlay of simulated detector hits from multiple *PYTHIA* 8.186 inclusive QCD events using the A3 tune [71]. The composite events were reweighted in the analysis so the distribution of the number of overlays per simulated signal event matched the mean number of collisions per bunch-crossing, $\langle \mu \rangle$, in data.

In addition to these MC generator versions and configurations, signal samples were also produced only at particle level, using the newer *SHERPA* 2.2.10 generator in a configuration equivalent to that described above, as well as 4FNS and fusing variations, and the NLO *MADGRAPH5_AMC@NLO* 2.7.3 + *PYTHIA* 8.244 generator with FxFx merging in 5FNS and 4FNS modes. These variations are used in Section 9 for comparisons with unfolded observables in data.

## 5 Lepton and jet definitions

The objects used in this analysis to select events and define observables are charged leptons, large-$R$ jets, and (optionally $b$-tagged) small-$R$ track-jets. These are defined in this section, with detailed discussion of the systematic uncertainties associated with their reconstruction postponed to Section 8.

The final results of this analysis consist of observables measured in particle-level fiducial volumes, which closely match the reconstruction-level event- and object-selection, to minimize model-dependent extrapolations. In what follows, the physics objects are defined at both reconstruction level and particle level. Throughout this analysis, stable particles are defined to be those with a mean lifetime $\tau > 10 \text{ mm}/c$.

**Charged leptons:** The leptonically decaying $Z$ boson is identified by use of high-$p_T$ charged $e^\pm$ and $\mu^\pm$ pairs, including those from $\tau$-lepton decays.

At reconstruction level, identified electrons and muons were used, with contamination suppressed by use of “tight” and “medium” identification criteria for electrons and muons respectively [72, 73]. The lepton candidates were geometrically restricted to the active regions of the calorimeters and muon spectometer ($|\eta| < 2.47$ excluding the $1.37–1.52$ region for electrons, and $|\eta| < 2.47$ for muons), and have $p_T > 27$ GeV for both lepton flavors. Both the electrons and muons were required to be isolated from significant energy deposits in the calorimeter and from high-momentum tracks.

Corrections derived in $Z \rightarrow \ell\ell$ events were applied to account for differences in reconstruction and identification efficiencies between data and simulated events. Electron energies as measured by the
emagnetic calorimeter were calibrated to the electron energy in simulation, and discrepancies between data and simulation corrected [74]. The reconstructed muon momenta were similarly calibrated using a mix of simulation and data-driven methods [75]. Uncertainties associated with the lepton efficiencies, scales, and resolutions were propagated via systematic variations.

At particle level, the charged leptons are defined as final-state electrons and muons dressed with direct photons within a surrounding cone of size $\Delta R = 0.1$, with kinematic requirements of $|\eta| < 2.47$ and $p_T > 27$ GeV on the resulting objects. No explicit requirements of final-state isolation nor direct connection to the hard scattering are made in the fiducial lepton definition, hence events with two high-$p_T$ electrons or two muons in the final-state acceptance are treated as part of the signal even if one or both arise from $\tau$-lepton or heavy-flavor hadron decays.

**Large-$R$ jets:** A large-$R$ jet is required in this analysis as a proxy for a high-momentum, hadronically decaying or splitting object, e.g. a high-energy gluon.

In reconstructed events, large-$R$ jets were reconstructed from calibrated topological clusters of calorimeter cells [76], using the anti-$k_t$ algorithm with radius parameter $R = 1.0$ [77, 78]. The clustered jet’s energy and pseudorapidity were further calibrated using simulated data, and its mass was calibrated using a combination of calorimeter and tracking information [79, 80]. Pile-up and underlying-event contributions to the jets were suppressed by a dynamic trimming [81] procedure discarding clusters from $R = 0.2$ subjects with less than 5% of the original jet $p_T$. The trimmed jets were required to have $p_T > 200$ GeV and $|\eta| < 2$, to ensure that the majority of the jet lies within the tracker volume. Discrepancies between data and simulation in the jet calibration were treated as systematic uncertainties.

At particle level, all final-state particles are used as inputs to the anti-$k_t$, $R = 1.0$ jet algorithm, and trimming is applied with the same parameters as in reconstruction, effectively subtracting underlying-event contributions. Again the trimmed jets are required to have $p_T > 200$ GeV and $|\eta| < 2$.

**Subjets and $b$-tagging:** Small-$R$ subjets within the large-$R$ jet are used as proxies for the leading partons in the jet, e.g. the $b$-quarks in a high-energy $g \to b\bar{b}$ splitting. To achieve high angular resolution, inner-detector tracks are used in place of calorimeter information to construct these subjets.

At reconstruction level, the anti-$k_t$ algorithm with radius parameter $R = 0.2$ was used to construct track-jets from at least two inner-detector tracks matched to the primary vertex (see Section 6). The track-jets were required to have $p_T > 10$ GeV and $|\eta| < 2.5$. Identification of the track-jets as likely (or not) to have been initiated by $b$-quarks was provided by the ATLAS MV2c10 multivariate $b$-tagging algorithm [82] trained on leptonic $t\bar{t}$ events to achieve 70% $b$-tagging efficiency, with mis-tag rejection ratios of 7.1 and 120 for charmed and light jets respectively. Corrections were applied to compensate for differences in $b$-tagging efficiency and charm- and light-jet mis-tag rates observed between simulation and collision data. Systematic uncertainties from tracking, vertexing, and $b$-tagging calibration were evaluated.

At particle level, stable charged particles are used as the inputs to the $R = 0.2$ anti-$k_t$ jet algorithm, again with $p_T > 10$ GeV and $|\eta| < 2.5$. A charged-particle jet is considered to be $b$-tagged if a weakly decaying $b$-hadron, with $p_T > 5$ GeV, is associated with it by the ghost-association method [83].

Any small-$R$ track-jets or small-$R$ charged-particle jets matched to the large-$R$ jets by ghost-association are considered to be charged subjets of the large-$R$ jet. A large-$R$ jet is considered
\(b\)-tagged if any of its small-\(R\) subjets is \(b\)-tagged, and the number of \(b\)-tagged subjets is used to define subclasses of signal events.

A simple overlap-removal procedure was used at both reconstruction and particle levels to accommodate a single particle or object leaving multiple signatures in the detector. In particular, this procedure was motivated by the possibility that leptons from the \(Z\)-boson decay could also be recorded as an additional large-\(R\) jet. To correct for this, the angular separation between each lepton and large-\(R\) jet was computed, and if any were within \(\Delta R = 1.0\) of each other then the jet was removed. This cut has the additional effect of suppressing contributions from the dijet process with collinear \(Z\)-boson emission from quarks, while still admitting widely separated topologies in which the \(Z\) boson and the large-\(R\) jet are located in the same event hemisphere.

### 6 Event selection and observables

For both the data and reconstruction-level MC samples, single-electron and single-muon triggers [84–86] were used to select the subset of events of \textit{a priori} relevance to this analysis before application of offline event-selection cuts. The kinematic requirements on the leptons defined in the previous section ensured full trigger-efficiency for events in the analysis fiducial phase-space. Candidate events were required to have a primary vertex, defined as the vertex with the highest sum of track \(p_T^2\), and with at least two associated tracks of \(p_T > 400\) MeV.

Event selection cuts on the physics objects defined in Section 5 were applied equally to reconstruction-level events in data and MC simulation, and to particle-level MC events. To select signal events containing a leptonically decaying \(Z\)-boson candidate produced in association with a large-\(R\) jet, events were required to have exactly two charged leptons of the same flavor, \(\ell^2 \in \{e^\pm, \mu^\pm\}\), and at least one large-\(R\) jet. No opposite-charge requirement was placed on the charged-lepton pairs, but the invariant mass of the lepton pair, \(m_{\ell\ell}\), was required to be greater than 50 GeV to exclude the photon-dominated part of the Drell–Yan continuum.

This set of criteria defines the “inclusive” event-selection region for this analysis. A more exclusive “2-tag” region was defined as a subset of this by additionally requiring that the large-\(R\) jet contains exactly two \(b\)-tagged subjets. The numbers of events selected in these regions, for the dielectron and dimuon channels separately, and for both collision data and MC simulation, are shown in Table 1, omitting the single-top and multijet processes, which were shown to be negligible by MC studies and a data-driven background study respectively.

The large-\(R\) jet used to construct the observables below was chosen as the highest-\(p_T\) large-\(R\) jet in the inclusive selection, and the highest-\(p_T\) 2-tag large-\(R\) jet in the 2-tag selection. The 2-tag observables are hence not a subset of the inclusive observables, as the latter includes events in which the highest-\(p_T\) large-\(R\) jet does not contain two \(b\)-tagged subjets.

The differential observables measured in this analysis are:

**Properties of the large-\(R\) jet (\(J\)):** the large-\(R\) jet mass, \(m_J\), and transverse momentum, \(p_T^J\), for both the inclusive and 2-tag regions;

**Properties of the large-\(R\) jet and \(Z\)-boson system (\(Z+J\)):** the transverse momentum of their vector sum, \(p_T^{Z+J}\), and their separation in \(\phi\), \(\Delta \phi(Z, J)\), for the inclusive region only;
Table 1: Reconstruction-level event-selection yields (and statistical uncertainties of the expected yields) in the $ee$ and $\mu\mu$ channels from each process’s MC sample (with Sherpa 2.2.1 used for the $Z$+jets samples) with the normalizations discussed in Section 4, and from collision data. The single-top process was found to make a negligible contribution to all event selections and has been omitted. Multijet backgrounds were estimated to be negligible by a data-driven method.

| Process          | Inclusive $ee$ | Inclusive $\mu\mu$ | 2-tag $ee$ | 2-tag $\mu\mu$ |
|------------------|---------------|--------------------|-----------|----------------|
| $Z+b\bar{b}$    | $324 \pm 4$  | $305 \pm 4$       | $163.8 \pm 2.6$ | $157.2 \pm 2.5$ |
| $Z+c\bar{c}$    | $536 \pm 10$ | $530 \pm 9$       | $12.3 \pm 1.8$  | $19.3 \pm 2.0$  |
| $Z+bc$           | $89 \pm 2$   | $81 \pm 2$        | $14.6 \pm 1.2$  | $12.1 \pm 0.9$  |
| $Z+b$            | $2588 \pm 13$| $2423 \pm 12$     | $14.8 \pm 1.1$  | $12.4 \pm 1.3$  |
| $Z+c$            | $5073 \pm 32$| $4862 \pm 39$     | $5.5 \pm 1.3$   | $6.9 \pm 1.7$   |
| $Z+$light        | $53808 \pm 164$| $51206 \pm 145$  | $9.4 \pm 1.1$   | $11.1 \pm 1.5$  |
| $t\bar{t}$       | $5960 \pm 46$| $5204 \pm 43$     | $82.7 \pm 5.3$  | $75.4 \pm 5.6$  |
| $W$+jets         | $73 \pm 4$   | $7 \pm 1$         | $0.4 \pm 0.1$   | $< 0.1$         |
| Diboson           | $2042 \pm 17$| $1834 \pm 16$     | $21.5 \pm 1.4$  | $20.7 \pm 1.4$  |
| MC total         | $70493 \pm 175$ | $66452 \pm 158$  | $324.9 \pm 6.8$ | $315.1 \pm 7.2$ |
| Data             | $66481$       | $65304$            | $391$        | $384$           |

Subjet separation: the angular separation, $\Delta R(b, \bar{b})$, between the two $b$-tagged subjets in the 2-tag region.

These variables respectively measure the external and internal kinematics of the selected large-$R$ jet, the effect of additional QCD radiation on the $Z$+J event topology, and the kinematics of $g \rightarrow b\bar{b}$ splitting in the boosted regime. The total cross-sections in the fiducial volume for the inclusive and 2-tag event selections were also measured, via integration of the differential measurements.

A selection of these observables are shown at reconstruction level in Figure 1, illustrating the inclusive large-$R$ jet and $Z$+J $p_T$ distributions, and the 2-tag large-$R$ jet mass and $\Delta R(b, \bar{b})$ distributions in data and simulation. One $ee$ and one $\mu\mu$ dilepton observable is shown for each event-selection region: all observables were constructed separately for the electron and muon channels, to allow consistency checks and consideration of distinct detector effects before lepton-channel combination. Significant discrepancies between data and the sum of MC-modeled processes (including correction factors) are visible in several variables, particularly the inclusive large-$R$ jet and $Z$+J-system transverse momenta. This evidence of general mismodeling in the boosted phase-space at reconstruction-level motivated the publication of the detector-corrected forms of these observables.

These plots show the admixture of processes contributing to the two event-selection regions, with the inclusive event-selection dominated by the $Z$+light-parton (including $Z$+charm quark) and $t\bar{t}$ production processes, and the 2-tag region dominated by $Z+b\bar{b}$ with significant background contributions from diboson and $t\bar{t}$ production.

7 Correction of observables to particle level

The main result of this analysis is the set of differential cross-sections introduced in the previous section, corrected to the particle-level fiducial phase-space by unfolding: the deconvolution of biases introduced
Figure 1: Selected reconstruction-level observables, compared with pre-fit MC simulation with SHERPA 2.2.1 used for the \(Z+\)jets samples: the top row shows the inclusive-selection \(ee\) large-\(R\) jet \(p_T\) distribution (left) and \(\mu\mu\ Z+J\) \(p_T\) distribution (right), and the bottom row shows the 2-tag selection \(ee\) large-\(R\) jet mass and \(\mu\mu\ \Delta R(b, \bar{b})\) distributions. The MC statistical uncertainties are shown by the dark gray band and the total uncertainty, including in quadrature the systematic uncertainties detailed in Section 8, are shown by the light gray band. The statistical uncertainty of the data is given by the error bar on the data point.
by the detector and reconstruction algorithms. Presentation in the fiducial form assists comparison with results from different experiments and with theoretical predictions.

In this analysis, the unfolding was performed using the Fully Bayesian Unfolding (FBU) technique [87]. FBU directly performs a likelihood fit in the parameter space of signal cross-sections $\sigma$, plus a set of nuisance parameters $\nu$ that control background compositions and systematic uncertainties (to be described in Section 8). The FBU method hence gives access to a full posterior probability density in the space of signal and nuisance parameters, from which arbitrarily detailed correlation information may be extracted.

The FBU posterior probability for each observable is constructed as the product of Poisson probabilities over all reconstruction-level bins $i$ as a function of the model parameters $\theta = \{\sigma, \nu\}$,

$$P(\theta | d) \propto \mathcal{L}(d | \theta) \cdot \pi(\theta),$$

where $d$ is the set of observed bin counts in data, and $\pi(\theta)$ is a set of prior probability densities over the model parameters. The $\mathcal{L}$ term in this can be expressed as a product of Poisson likelihoods over the bins,

$$\mathcal{L}(d | \theta) = \prod_{i \in \text{bins}} \text{Poiss}(d_i, x_i(\theta)),
$$

where $x(\theta)$ is the set of expected total bin yields. This can be decomposed further into reconstruction-level background and signal cross-sections $b_i$ and $s_i$,

$$x_i(\theta) = \mathcal{L}(\nu) \cdot (b_i(\nu) + s_i(\theta))$$
$$= \mathcal{L}(\nu) \cdot \left( b_i(\nu) + \sum_j M_{ij}^{\text{p-R}}(\nu) \cdot \sigma_j \right),$$

where $\mathcal{L}$ is the integrated luminosity of the dataset and $M_{ij}^{\text{p-R}}$ is the response matrix $p(i|j)$ mapping particle-level bins $\{j\}$ to reconstruction-level bins $\{i\}$. Overflows outside the fiducial acceptance are included in the bin indexing, so migrations into and out of the acceptance are treated with the full machinery. This provides the full formalism necessary to relate the observed data $d_i$ to our parameters of interest, the particle-level signal cross-sections $\sigma_i$. In the inclusive region, the signal is defined as all $Z$+jets contributions, while in the 2-tag region it is only $Z+b\bar{b}$, with other $Z$+jets flavors now considered as part of the background.

The background cross-sections and response matrix were constructed from a set of MC-derived histogram templates, including predictions from the nominal MC samples ($\text{Sherpa } Z$+jets) with data/MC corrections, and a set of predictions from each systematic-uncertainty variation to be described in Section 8. Examples of nominal-model response-matrix templates for two observables are shown in Figure 2.

The unit nuisance parameters $\nu_k \in \nu$ were used to define linear interpolations of $b_i$ and $M_{ij}^{\text{p-R}}$ between templates corresponding to $\nu_k = 0$ and 1. The absolute deviations obtained from “up” and “down” variations of nuisance parameters were averaged into single positive deviations for use in this symmetrized form. Unit Gaussian priors were applied to all $\nu_k$ other than the luminosity uncertainty: this corresponds to a linear scaling of all rates and hence was given a positive log-normal prior with $\mu = 0$ and $\sigma = 0.021$ to avoid exploration of negative luminosities. The background normalizations were allowed to float with Gaussian prior widths discussed in Section 8, and flat, non-negative priors were imposed on the signal cross-section parameters $\sigma$.

For this analysis, the “hunfold” [88] implementation of FBU was used. This uses gradient ascent to maximize the posterior log-probability $\ln P(\theta)$, and then samples the posterior probability distribution.
using a proposal density in $\theta$ derived from the likelihood Hessian matrix at the maximum-likelihood point. In the unfolded observables of this analysis (see Section 9), each variable is constructed from a sum of electron and muon channels via a single FBU fit, with a “double width” concatenation of electron and muon response matrices used to simultaneously unfold the electron and muon distributions into combined particle-level distributions representing $Z$-boson decays into either lepton flavor.

8 Systematic uncertainties

The measurements in this analysis are affected by statistical uncertainties and by systematic uncertainties from detector-interaction and reconstruction processes, from MC modeling, and from the unfolding procedure. Estimates of these uncertainties were derived using standard methods described in this section, and were propagated through the unfolding procedure where they affected the final posterior distributions.

The main sources of experimental systematic uncertainty affecting these measurements were:

**Charged leptons:** energy/momentum scale and resolution, and reconstruction, identification, isolation, and trigger efficiencies. Systematic variations of the data/MC efficiency corrections and energy/momentum calibrations applied to the MC samples [74, 89] were used to define variations from their nominal templates in the FBU unfolding’s parameterized background cross-sections and signal response matrices.

**Large-$R$ jets:** energy scale, mass scale, energy resolution, and mass resolution. The jet energy scale (JES) and jet mass scale (JMS) uncertainties were based on the double-ratio of each variable (energy, mass) to its equivalent reconstructed from track-jets, between data and simulation; this construction permitted separation of the physics effects from the calorimeter reconstruction systematic uncertainties [90]. In this analysis, the JES and JMS uncertainties have been treated as fully uncorrelated; cross-checks assuming higher degrees of correlation had negligible effect. The uncertainty in the jet mass resolution (JMR) was determined by smearing the jet mass such that its resolution was degraded by 20%, and for the large-$R$ jet energy resolution (JER), an absolute uncertainty of 2% was used.
Flavor tagging: track-jet $b$-tagging efficiencies, and mis-tag rates for $c$-jets and light-flavor jets. The $b$-tagging efficiency and charm mis-tag rate in simulation were calibrated using the tag-and-probe method in $t\bar{t}$ events [91, 92], and the light-jet mis-tag rate was calibrated in dijet events [93]. A total of 25 diagonalized systematic uncertainties associated with these calibration factors were considered in this analysis.

Pile-up: pile-up reweighting uncertainty. The MC predictions were reweighted such that their distribution of the number of pile-up vertices matched the pile-up distribution measured in the data. The uncertainty from this procedure was propagated to the unfolding using variations of pile-up weights to account for the uncertainty in the pile-up estimation.

In addition to the above measures of imperfect understanding of the detector, uncertainties in the modeling of both the signal and background physics processes were propagated to the final measurement, via systematic variations in the unfolding components. As for the detector uncertainties, this propagation was implemented via linear interpolation of template distributions (including response matrices) in the unfolding machinery. For the signal process, standard 7-point variations of the renormalization and factorization scales by factors of two from the nominal values, the nominal PDF’s error-set, differences between the nominal PDF and the alternative CT14nnlo [94] and MMHT2014nnlo [95] central PDFs, and variations of $\alpha_s(m_Z)$ by $\pm 0.001$ were included. Approximate uncertainties in the matching procedure between matrix element and parton shower were evaluated by variations between the nominal Sherpa samples and the MadGraph5_AMC@NLO + Pythia 8 samples. The envelope of differences in the signal-process response matrix between Sherpa and MadGraph was also treated as a systematic uncertainty; a cross-check showed that the results were not sensitive to whether or not the Sherpa/MadGraph difference was split into several independently parametrized components.

Dedicated variation MC samples were used to evaluate the modeling of the $t\bar{t}$ background, considering the matrix element, parton shower model, and the dependence on initial- and final-state radiation settings and the $h_{\text{amp}}$ parameter. In addition, the quality of $t\bar{t}$ modeling was assessed in MC–data comparisons using an opposite-flavor $e\mu$ variation on the standard $ee\mu\mu$ selection; this resulted in additional mismodeling uncertainties of 30–50% being added in the first bins of the inclusive $p_T^{Z^{\pm}}$ observable. A conservative $\sigma = 0.2$ Gaussian prior, informed by the maximum reconstruction-level data/MC disagreement, was used for the normalization uncertainty of all background processes other than the non-signal $Z$+jets samples in the 2-tag region, for which a larger $\sigma = 0.5$ prior was used. This inflated uncertainty was assigned to reflect that a robust in situ measure of the 2-tag flavor fractions could not be obtained from $b$-tagging variable templates due to the low event-count in this analysis phase-space.

In addition, systematic errors arise from the finite sizes of the simulation samples. In principle, the statistical limitation of each bin corresponds to a nuisance parameter in the unfolding. But in practice many of these parameters have little effect, since most bins with low rates, e.g. very statistically limited off-diagonal elements of the response matrix, by construction do not contribute much to the result. This abundance of uncertain quantities creates an intractably large space of nuisance parameters in which the unfolding fit is unlikely to converge. A “pruning” procedure was hence implemented, both for statistical uncertainties in response-matrix estimates and for all the detector and modeling systematic uncertainties described. The pruning criterion was to remove nuisance parameters which produced a background variation of less than 5% in all bins (the background fractions being around 20% of the event yield in the inclusive region and 40% in the 2-tag region), and did not change any entries in the response matrix by at least 0.002. The effect was to prune jet-substructure and most lepton-calibration systematic uncertainties for all observables, jet-mass systematic uncertainties for the $p_T^{L}$ and $p_T^{Z^{\pm}}$ observables, and jet-mass resolution for all but the large-$R$ jet mass observable. The $b$-tagging uncertainties were partially pruned for the 2-tag region;
naturally, there are no such uncertainties in the inclusive region. The unpruned MC statistical uncertainties in total contributed a subleading $\sim 1\%$.

A comprehensive set of closure tests in the unfolding/nuisance-profiling procedure was performed, including closure tests with and without reweighting schemes to impose greater MC similarity to the reconstruction-level observables in data, stress-testing by reweighting Sherpa pseudodata with a MadGraph-derived response matrix and vice versa, and checks against bias from nonuniform signal priors. Non-closure effects from these tests, which were very small in nearly all bins, but rose to around 20% in single bins of the $p_T^J$ and $p_T^{Z+J}$ distributions, were added in quadrature to the mix of final uncertainties.

Summary systematic uncertainties are listed in Table 2 and illustrated for two observables in Figure 3. The summaries were obtained from the FBU posterior-distribution samples by computing the sample covariance matrix between all fit parameters, including nuisance parameters and signal bin-values, $\text{cov}_{ij} = \langle x_i x_j \rangle - \langle x_i \rangle \langle x_j \rangle$ for parameter indices $i$ and $j$. The absolute values of covariance-matrix rows for the elementary nuisance parameters were then summed in semantic groupings, e.g. the sets of nuisance parameters for electron reconstruction, jet reconstruction, $b$-tagging, etc. The resulting grouped-covariance entries were projected onto the signal-bin cross-section parameters via the relation $\sigma_b = \text{cov}_{bn} / \sqrt{\text{cov}_{nn}}$, where $b$ is the signal bin index, and $n$ the systematic nuisance index. By construction, these grouped uncertainties are symmetric. The total uncertainty, including statistical effects, is given by the larger standard deviation of each signal bin value. The largest systematic-group effects are from large-$R$ jet calibration and signal modeling in the inclusive region, and from background normalization and $b$-tagging calibration in the 2-tag region.
Table 2: Summary table of relative uncertainty magnitudes per observable. These uncorrelated estimates of systematic uncertainties’ contributions to the total uncertainty are based on projection of nuisance parameters onto the signal cross-section bin values via the likelihood-scan covariance matrix, and summing the elementary contributions in quadrature.

|                         | Inclusive                                                                 | 2-tag                                      |
|--------------------------|---------------------------------------------------------------------------|--------------------------------------------|
|                         | $p_{T}^{J}$                  | $m_{J}$                  | $p_{T}^{Z,J}$                  | $\Delta \phi (Z, J)$ | $p_{T}^{1}$                  | $m_{J}$                  | $\Delta R (b, \bar{b})$ |
| Scales & PDFs            | 0 – 10%                     | 3 – 7%                    | 2 – 9%                         | 1 – 4%                | 2 – 9%                     | 3 – 9%                    | 2 – 14%                |
| MC modeling              | 1 – 11%                     | 0 – 12%                   | 1 – 6%                         | 1 – 6%                | 0 – 15%                    | 4 – 9%                    | 2 – 9%                 |
| MC statistics            | 0 – 2%                      | 3 – 10%                   | 0 – 5%                         | 0 – 1%                | 4 – 12%                    | 3 – 11%                   | 2 – 4%                 |
| Bkg normalization        | 2 – 6%                      | 4 – 7%                    | 3 – 6%                         | 2 – 7%                | 9 – 37%                    | 11 – 23%                  | 11 – 26%               |
| $t\bar{t}$ modeling      | 1 – 22%                     | 3 – 13%                   | 2 – 13%                        | 2 – 11%               | 4 – 23%                    | 11 – 29%                  | 3 – 24%                |
| Jet reco.                | 5 – 50%                     | 6 – 11%                   | 5 – 21%                        | 7 – 22%               | 9 – 33%                    | 15 – 24%                  | 11 – 21%               |
| $b$-tagging              |                            |                           |                               |                      | 5 – 33%                    | 6 – 12%                   | 7 – 11%                |
| Electron reco.           | 0 – 2%                      | 1 – 7%                    | 2 – 7%                         | 1 – 3%                | 4 – 8%                     | 0 – 9%                    | 1 – 3%                 |
| Muon reco.               | 1 – 10%                     | 2 – 9%                    | 0 – 2%                         | 1 – 5%                | 1 – 3%                     | 1 – 7%                    | 0 – 2%                 |
| Pile-up                  | 0 – 9%                      | 0 – 3%                    | 0 – 2%                         | 0 – 1%                | 1 – 16%                    | 0 – 7%                    | 1 – 4%                 |
| Luminosity               | 3 – 4%                      | 0 – 7%                    | 3 – 7%                         | 4 – 5%                | 5 – 7%                     | 3 – 7%                    | 4 – 6%                 |
| Unfolding closure        | 0 – 21%                     | 0 – 0%                    | 0 – 18%                        | 0 – 0%                | 0 – 0%                     | 0 – 0%                    | 0 – 0%                 |
| Total syst.              | 7 – 62%                     | 11 – 27%                  | 10 – 33%                       | 9 – 27%               | 23 – 66%                   | 24 – 45%                  | 22 – 41%               |

9 Results

The posterior distributions of nuisance parameters common to the electron and muon channels were unfolded independently and found to be consistent before performing the simultaneous unfolding shown here. Other than in the inclusive-region $m_{J}$ distribution, the FBU procedure did not constrain most systematic-uncertainty nuisance parameters significantly. The nuisance parameter for the Sherpa vs. MadGraph modeling uncertainty was the exception: this was constrained in favor of Sherpa to 20–40% of the original prior width by all distributions. The background normalizations changed by at most a few percent, not significantly modifying the admixture of signal and background predicted by the MC programs, and least of all in the 2-tag fits.

The full multidimensional posterior-probability distribution is the most complete form of the measurement, but for histogram presentation each bin’s marginal probability distribution is used to define the central value and error bar; these correspond to the marginal median and marginal central 68% confidence range respectively.

The final unfolded differential cross-sections as functions of the event kinematics are presented in Figures 4 and 5, compared with NLO particle-level predictions from Sherpa 2.2.1 and 2.2.10, and MadGraph5_AMC@NLO 2.7.3 + Pythia 8.244. In this section, all predictions are normalized to their own calculated cross-section to allow an unbiased comparison of both the total rates and distribution shapes between the different generators.

The total fiducial cross-sections are measured by integration of the angular distributions (chosen because they do not have overflow bins) in both event-selection regions. The measured values are $\sigma^{\text{incl}} = 2.37 \pm 0.28$ pb for the inclusive selection, and $\sigma^{2\text{-tag}} = 14.6 \pm 4.6$ fb for the 2-tag selection.

The NLO Sherpa 2.2.10 and NLO MadGraph5_AMC@NLO + Pythia 8 generators predict slightly higher central values of the inclusive cross-section than the measured one, at 2.53 $\pm$ 1.25 pb and
Figure 4: Particle-level differential cross-sections in the inclusive event selection. The top row shows the large-$R$ jet $p_T$ (left) and mass (right), and the bottom row shows the $p_T$ of the $Z$+J system (left) and the azimuthal separation of the $Z$ and large-$R$ jet (right). The combined statistical and systematic uncertainty band from the FBU fit is shown. In the legend, “MGaMC” refers to NLO configurations of the MadGraph5_AMC@NLO generator, and “MG” to LO MadGraph, both run in conjunction with Pythia 8. All models are using the 5FNS.
Figure 5: Particle-level differential cross-sections in the 2-tag event selection. The top row shows the large-$R$ jet $p_T$ (left) and mass (right), and the bottom row shows the angular separation of $b$-tagged charged-particle subjets. The combined statistical and systematic uncertainty band from the FBU fit is shown. In the legend, “MGaMC” refers to NLO configurations of the MadGraph5_αMC@NLO generator, run in conjunction with PYTHIA 8, and “4/5F” refer to the flavor-number scheme used.
As noted in the review of reconstruction-level plots, the mismodeling of extra radiation by the older Sherpa 2.2.1 central prediction of 2.37 pb also agrees closely with the measurement. The large uncertainties in the NLO calculations are dominated by the effects of scale variations; systematic uncertainties are not well defined for the LO calculation, but the better nominal performance is evidently provided by the NLO generators. For the 2-tag cross-section, the NLO 5-flavor Sherpa 2.2.10 and MadGraph5_AMC@NLO central predictions describe the data well, with 14.9 ± 4.2 fb and 14.4 ± 1.9 fb respectively. The Sherpa fusing cross-section, mixing elements of the 4- and 5-flavor calculations, is close to the 5-flavor predictions, with 14.3 ± 4.8 fb. The 4-flavor Sherpa and MadGraph5_AMC@NLO calculations, and the previous 5-flavor Sherpa 2.2.1 prediction underestimate the 2-tag cross-section with 9.4 ± 3.1 fb, 4.4 ± 1.1 fb, and 9.1 fb respectively. This result underscores the expectation that the 5-flavor (or fusing) scheme is the more appropriate choice for heavy-quark production in this analysis phase-space, even for $b\bar{b}$-pair production.

The $\sigma^{2\text{-tag}} / \sigma^{\text{incl}}$ ratio of 2-tag to inclusive events seen in data is $(0.62 \pm 0.12)\%$, accounting for cancellations of shared systematic uncertainties between the inclusive and 2-tag cross-section estimates. This figure is reproduced well by Sherpa 2.2.10's $(0.59 \pm 0.39)\%$ and by MadGraph5_AMC@NLO + Pythia 8 with $(0.54 \pm 0.21)\%$. The older NLO Sherpa 2.2.1 and leading-order MadGraph + Pythia 8 estimates undershoot with 0.42% and 0.38% respectively. These cross-sections hence furnish new experimental discriminators between perturbative-QCD models of high-$p_T$ heavy-flavor production rates, despite the significant measurement uncertainties.

In the inclusive-selection differential distributions of Figure 4, the MadGraph5_AMC@NLO + Pythia 8 predictions can be seen to have the shapes in best agreement with data, not suffering from the excesses of activity common to the Sherpa models and leading-order MadGraph + Pythia 8 in the more extreme phase-space regions of high $p_T^J$ and $p_T^{Z+J}$, and small $\Delta\phi(Z, J)$. In this topology, where the $Z$+jets process becomes more like a dijet system with collinear $Z$-boson radiation, both generators display similar shape deviations with respect to the measurement, with the best agreement at low $p_T$, low mass, and low levels of additional event activity (as characterised by low $p_T^{Z+J}$ values and the back-to-back $\Delta\phi(Z, J) \sim \pi$ region). The excess in inclusive cross-section estimate for nominal MadGraph5_AMC@NLO +Pythia 8 can be seen to arise from a relatively small overpopulation with respect to Sherpa in the most populated bins of $p_T^J$ and $p_T^{Z+J}$, while its shapes typically match data to within 10%, whereas the other generators overestimate high-scale activity by 50–100%.

As noted in the review of reconstruction-level plots, the mismodeling of extra radiation by Sherpa and leading-order MadGraph + Pythia 8 is one of the most significant discrepancies between simulation and data observed in this analysis. Despite the large measurement uncertainties, this evidence of larger transverse recoil of the $Z$+J system in simulation than in data, as well as the higher-than-observed $p_T$ and mass of the large-$R$ jets, is an important result for inclusive QCD model development and tuning in this boosted phase-space.

In the 2-tag selection distributions in Figure 5, the larger uncertainties and much lower event counts mean that shape discrepancies are more difficult to discern: current shape modeling appears to be performing adequately, with relatively constant MC/data ratios for large-$R$ jet mass and $\Delta R(b, \bar{b})$. In particular, the large-$R$ jet mass in this region appears to be consistently well described by all MC models, with no sign of the excesses and model disagreements seen in the inclusive-region version of that observable. Further analysis with the complete Run 2 dataset will be required to discriminate between the models in this phase-space, beyond the evident favoring of the 5FNS for total $b\bar{b}$ production rate.
10 Conclusion

We have presented measurements of cross-sections for the production of a leptonically decaying $Z$ boson in association with a large-radius jet in LHC 13 TeV proton–proton collision events from the ATLAS 36 fb$^{-1}$ combined 2015–2016 data-taking run, corrected to a particle-level fiducial region. The observables presented are differential in kinematic variables of the $Z$ boson, the large-$R$ jet, and its associated small-$R$ $b$-tagged charged-particle jets. They are measured with a flavor-inclusive event selection and also within a “2-tag” event-selection region which adds a double $b$-hadron labeling requirement on the large-$R$ jet. The integrated cross-sections within the fiducial volumes of the event-selection regions have also been presented.

These cross-section estimates were extracted from data using the Fully Bayesian Unfolding formalism, effectively performing a posterior-probability fit over a combination of signal and background cross-section parameters, and various systematic uncertainties affecting the response of the detector. These measurements provide an important test of perturbative quantum chromodynamics, with particular emphasis on the production rates and kinematics of bottom quarks. These are a significant background to several important Higgs-boson searches, and are affected by significant theory and modeling uncertainties. The full data, correlations, and samples from the posterior-probability function are provided for use in event-generator tuning and model hypothesis-testing via public databases.

The differential cross-sections indicate significant mismodeling of QCD activity in the inclusive event selection by many MC models, with both the NLO SHERPA and LO MADGRAPH + PYTHIA 8 event generators predicting greater $p_T$ and azimuthal decorrelation in the $Z+J$ system than seen in the ATLAS data. The large-$R$ jet itself is consequently biased to higher $p_T$ and mass values than in data, although to a lesser extent than the deviations in the $Z+J$-system observables. The NLO MADGRAPH5_AMC@NLO + PYTHIA 8 model, by contrast, describes all distribution shapes well, with only a small overestimate of the inclusive fiducial cross-section. All models somewhat overestimate this cross-section, with recent SHERPA versions providing the best description.

The 2-tag selection, while its discrimination power is limited by the number of data events, does not appear to suffer from the same shape-modeling issues, and there is good shape agreement between the data and all MC models. The strongest feature observed in this event-selection region is in normalization, with models using the 4FNS approach significantly underestimating the rate of $b\bar{b}$ boosted-jet production. Five-flavor approaches with modern tools do much better, with both SHERPA 2.2.10 and MADGRAPH5_AMC@NLO providing accurate predictions for the 2-tag cross-section and the ratio of 2-tag to inclusive rates. This information is important for future use of MC-derived large-$R$ jet flavor composition in, for example, studies of the $VH(b\bar{b})$ process. As the result is statistically limited in the 2-tag region, the significant increase in integrated luminosity from the LHC full Run 2 dataset (and expected from the LHC Run 3 program) should provide a clearer view of how far the validity of MC modeling of heavy-flavor extends into this extreme event topology.

References

[1] J. M. Butterworth, A. R. Davison, M. Rubin, and G. P. Salam, 
"Jet Substructure as a New Higgs-Search Channel at the Large Hadron Collider," 
Phys. Rev. Lett. 100 (2008) 242001, arXiv: 0802.2470 [hep-ph].
[2] ATLAS Collaboration, Search for pair production of Higgs bosons in the $b\bar{b}b\bar{b}$ final state using proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector, JHEP 01 (2019) 030, arXiv: 1804.06174 [hep-ex].

[3] ATLAS Collaboration, Identification of boosted Higgs bosons decaying into $b$-quark pairs with the ATLAS detector at 13 TeV, Eur. Phys. J. C 79 (2019) 836, arXiv: 1906.11005 [hep-ex].

[4] 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 (2019) 040, arXiv: 1808.01473 [hep-ex].

[5] CMS Collaboration, Search for resonances decaying to a pair of Higgs bosons in the $b\bar{b}q\bar{q}\ell\nu$ final state in proton-proton collisions at $\sqrt{s} = 13$ TeV, JHEP 10 (2019) 125, arXiv: 1904.04193 [hep-ex].

[6] CMS Collaboration, Search for charged Higgs bosons decaying into a top and a bottom quark in the all-jet final state of $p\bar{p}$ collisions at $\sqrt{s} = 13$ TeV, JHEP 07 (2020) 126, arXiv: 2001.07763 [hep-ex].

[7] 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 (2020) 085, arXiv: 2006.13251 [hep-ex].

[8] D. Napoletano, “$V$ and heavy flavour production: status and recent theoretical developments,” 7th Large Hadron Collider Physics Conference (LHCP 2019) Puebla, Puebla, Mexico, May 20-25, 2019, arXiv: 1909.00143 [hep-ph].

[9] ATLAS Collaboration, Measurement of the cross-section for $b$-jets produced in association with a $Z$ boson at $\sqrt{s} = 7$ TeV with the ATLAS detector, Phys. Lett. B 706 (2012) 295, arXiv: 1109.1403 [hep-ex].

[10] ATLAS Collaboration, Measurement of differential production cross-sections for a $Z$ boson in association with $b$-jets in 7 TeV proton–proton collisions with the ATLAS detector, JHEP 10 (2014) 141, arXiv: 1407.3643 [hep-ex].

[11] CMS Collaboration, Measurement of the cross section and angular correlations for associated production of a $Z$ boson with $b$ hadrons in $pp$ collisions at $\sqrt{s} = 7$ TeV, JHEP 12 (2013) 039, arXiv: 1310.1349 [hep-ex].

[12] CMS Collaboration, Measurement of the production cross sections for a $Z$ boson and one or more $b$ jets in $pp$ collisions at $\sqrt{s} = 7$ TeV, JHEP 06 (2014) 120, arXiv: 1402.1521 [hep-ex].

[13] LHCb Collaboration, Measurement of the $Z+b$-jet cross-section in $pp$ collisions at $\sqrt{s} = 7$ TeV in the forward region, JHEP 01 (2015) 064, arXiv: 1411.1264 [hep-ex].

[14] CMS Collaboration, Measurements of the associated production of a $Z$ boson and $b$ jets in $pp$ collisions at $\sqrt{s} = 8$ TeV, Eur. Phys. J. C 77 (2017) 751, arXiv: 1611.06507 [hep-ex].

[15] ATLAS Collaboration, Measurements of the production cross-section for a $Z$ boson in association with $b$-jets in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector, JHEP 07 (2020) 044, arXiv: 2003.11968 [hep-ex].

[16] CMS Collaboration, Measurement of the associated production of a $Z$ boson with charm or bottom quark jets in proton–proton collisions at $\sqrt{s} = 13$ TeV, Phys. Rev. D 102 (2020) 032007, arXiv: 2001.06899 [hep-ex].
[17] CMS Collaboration, Measurement of \( \bar{B} \bar{B} \) Angular Correlations based on Secondary Vertex Reconstruction at \( \sqrt{s} = 7 \) TeV, JHEP 03 (2011) 136, arXiv: 1102.3194 [hep-ex].

[18] ATLAS Collaboration, Properties of \( g \to b \bar{b} \) at small opening angles in pp collisions with the ATLAS detector at \( \sqrt{s} = 13 \) TeV, Phys. Rev. D 99 (2019) 052004, arXiv: 1812.09283 [hep-ex].

[19] ATLAS Collaboration, The ATLAS Experiment at the CERN Large Hadron Collider, JINST 3 (2008) S08003.

[20] L. Evans and P. Bryant, LHC Machine, JINST 3 (2008) S08001.

[21] V. N. Gribov and L. N. Lipatov, Deep inelastic electron scattering in perturbation theory, Phys. Lett. B 37 (1971) 17.

[22] G. Altarelli and G. Parisi, Asymptotic freedom in parton language, Nuclear Physics B 126 (1977) 298.

[23] Y. L. Dokshitzer, Calculation of the structure functions for deep inelastic scattering and \( e^+ e^- \) annihilation by perturbation theory in quantum chromodynamics, Zh. Eksp. Teor. Fiz 73 (1977) 1216.

[24] D. Napoletano, Heavy quark mass effects in associated production, PoS DIS2018 (2018) 116, arXiv: 1808.04809 [hep-ph].

[25] F. Krauss and D. Napoletano, Towards a fully massive five-flavor scheme, Phys. Rev. D 98 (2018) 096002, arXiv: 1712.06832 [hep-ph].

[26] S. Forte, D. Napoletano, and M. Ubiali, Higgs production in bottom-quark fusion in a matched scheme, Phys. Lett. B 751 (2015) 331, arXiv: 1508.01529 [hep-ph].

[27] E. Bagnaschi, F. Maltoni, A. Vicini, and M. Zaro, Lepton-pair production in association with a \( b \bar{b} \) pair and the determination of the W boson mass, JHEP 07 (2018) 101, arXiv: 1803.04336 [hep-ph].

[28] G. Ridolfi, M. Ubiali, and M. Zaro, A fragmentation-based study of heavy quark production, JHEP 01 (2020) 196, arXiv: 1911.01975 [hep-ph].

[29] R. Gauld, A. Gehrmann-De Ridder, E. W. N. Glover, A. Huss, and I. Majer, Predictions for Z -Boson Production in Association with a b-Jet at \( O(\alpha_s^3) \), Phys. Rev. Lett. 125 (2020) 222002, arXiv: 2005.03016 [hep-ph].

[30] F. Krauss, D. Napoletano, and S. Schumann, Simulating \( b \)-associated production of Z and Higgs bosons with the SHERPA event generator, Phys. Rev. D 95 (2017) 036012, arXiv: 1612.04640 [hep-ph].

[31] R. K. Ellis, G. Marchesini, and B. R. Webber, Soft radiation in parton-parton scattering, Nucl. Phys. B 286 (1987) 643, [Erratum: Nucl. Phys. B 294 (1987) 1180].

[32] ATLAS Collaboration, Measurements of Higgs Bosons Decaying to Bottom Quarks from Vector Boson Fusion Production with the ATLAS Experiment at \( \sqrt{s} = 13 \) TeV, Eur. Phys. J. C 81 (2020) 537, arXiv: 2011.08280 [hep-ex].

[33] ATLAS Collaboration, Observation of \( H \to b \bar{b} \) decays and VH production with the ATLAS detector, Phys. Lett. B 786 (2018) 59, arXiv: 1808.08238 [hep-ex].
[34] ATLAS Collaboration, Measurement of the associated production of a Higgs boson decaying into b-quarks with a vector boson at high transverse momentum in pp collisions at √s = 13 TeV with the ATLAS detector, Phys. Lett. B 816 (2021) 136204, arXiv: 2008.02508 [hep-ex].

[35] ATLAS Collaboration, ATLAS Insertable B-Layer: Technical Design Report, ATLAS-TDR-19; CERN-LHCC-2010-013, 2010, url: https://cds.cern.ch/record/1291633, Addendum: ATLAS-TDR-19-ADD-1; CERN-LHCC-2012-009, 2012, url: https://cds.cern.ch/record/1451888.

[36] B. Abbott et al., Production and integration of the ATLAS Insertable B-Layer, JINST 13 (2018) T05008, arXiv: 1803.00844 [physics.ins-det].

[37] ATLAS Collaboration, Performance of the ATLAS trigger system in 2015, Eur. Phys. J. C 77 (2017) 317, arXiv: 1611.09661 [hep-ex].

[38] ATLAS Collaboration, Performance of the ATLAS Trigger System in 2015, Eur. Phys. J. C 77 (2017) 317, arXiv: 1611.09661 [hep-ex].

[39] ATLAS Collaboration, The ATLAS Collaboration Software and Firmware, ATL-SOFT-PUB-2021-001, 2021, url: https://cds.cern.ch/record/2767187.

[40] ATLAS Collaboration, Luminosity determination in pp collisions at √s = 13 TeV using the ATLAS detector at the LHC, ATLAS-CONF-2019-021, 2019, url: https://cds.cern.ch/record/2677054.

[41] G. Avoni et al., The new LUCID-2 detector for luminosity measurement and monitoring in ATLAS, JINST 13 (2018) P07017.

[42] E. Bothmann et al., Event Generation with Sherpa 2.2, SciPost Phys. 7 (2019) 034, arXiv: 1905.09127 [hep-ph].

[43] T. Gleisberg et al., Event generation with SHERPA 1.1, JHEP 02 (2009) 007, arXiv: 0811.4622 [hep-ph].

[44] T. Gleisberg and S. Höche, Comix, a new matrix element generator, JHEP 12 (2008) 039, arXiv: 0808.3674 [hep-ph].

[45] S. Schumann and F. Krauss, A Parton shower algorithm based on Catani-Seymour dipole factorisation, JHEP 03 (2008) 038, arXiv: 0709.1027 [hep-ph].

[46] S. Catani, F. Krauss, B. R. Webber, and R. Kuhn, QCD matrix elements + parton showers, JHEP 11 (2001) 063, arXiv: hep-ph/0109231.

[47] S. Höche, F. Krauss, S. Schumann, and F. Siegert, QCD matrix elements and truncated showers, JHEP 05 (2009) 053, arXiv: 0903.1219 [hep-ph].

[48] S. Höche, F. Krauss, M. Schönherr, and F. Siegert, QCD matrix elements + parton showers: The NLO case, JHEP 04 (2013) 027, arXiv: 1207.5030 [hep-ph].

[49] S. Höche, F. Krauss, M. Schönherr, and F. Siegert, A critical appraisal of NLO+PS matching methods, JHEP 09 (2012) 049, arXiv: 1111.1220 [hep-ph].

[50] F. Cascioli, P. Maierhöfer, and S. Pozzorini, Scattering Amplitudes with Open Loops, Phys. Rev. Lett. 108 (2012) 111601, arXiv: 1111.5206 [hep-ph].
[51] A. Denner, S. Dittmaier, and L. Hofer, 
*Collier: a fortran-based complex one-loop library in extended regularizations*, 
Comput. Phys. Commun. **212** (2017) 220, arXiv: 1604.06792 [hep-ph].

[52] S. Höche, F. Krauss, M. Schönherr, and F. Siegert, *NLO matrix elements and truncated showers*, 
JHEP **08** (2011) 123, arXiv: 1009.1127 [hep-ph].

[53] NNPDF Collaboration, R.D. Ball and others, *Parton distributions for the LHC Run II*, 
JHEP **04** (2015) 040, arXiv: 1410.8849 [hep-ph].

[54] C. Anastasiou, L. J. Dixon, K. Melnikov, and F. Petriello, *High-precision QCD at hadron colliders: Electroweak gauge boson rapidity distributions at next-to-next-to leading order*, 
Phys. Rev. D **69** (2004) 094008, arXiv: hep-ph/0312266.

[55] J. Alwall et al., *The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations*, 
JHEP **07** (2014) 079, arXiv: 1405.0301 [hep-ph].

[56] R. D. NNPDF Collaboration Ball et al., *Parton distributions with LHC data*, 
Nucl. Phys. B **867** (2013) 244, arXiv: 1207.1303 [hep-ph].

[57] T. Sjöstrand, S. Mrenna, and P. Skands, *A Brief Introduction to PYTHIA 8.1*, 
Comput. Phys. Commun. **178** (2008) 852, arXiv: 0710.3820 [hep-ph].

[58] L. Lönnblad, *Correcting the Colour-Dipole Cascade Model with Fixed Order Matrix Elements*, 
JHEP **05** (2002) 046, arXiv: hep-ph/0112284 [hep-ph].

[59] L. Lönnblad and S. Prestel, *Matching tree-level matrix elements with interleaved showers*, 
JHEP **03** (2012) 019, arXiv: 1109.4829 [hep-ph].

[60] D. J. Lange, *The EvtGen particle decay simulation package*, 
Nucl. Instrum. Meth. A **462** (2001) 152.

[61] *ATLAS Run 1 Pythia8 tunes*, tech. rep. ATL-PHYS-PUB-2014-021, CERN, 2014, 
url: https://cds.cern.ch/record/1966419.

[62] S. Frixione, P. Nason, and G. Ridolfi, 
*A positive-weight next-to-leading-order Monte Carlo for heavy flavour hadroproduction*, 
JHEP **09** (2007) 126, arXiv: 0707.3088 [hep-ph].

[63] P. Nason, *A new method for combining NLO QCD with shower Monte Carlo algorithms*, 
JHEP **11** (2004) 040, arXiv: hep-ph/0409146.

[64] S. Frixione, P. Nason, and C. Oleari, 
*Matching NLO QCD computations with parton shower simulations: the POWHEG method*, 
JHEP **07** (2007) 070, arXiv: 0709.2092 [hep-ph].

[65] S. Alioli, P. Nason, C. Oleari, and E. Re, *A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX*, 
JHEP **06** (2010) 043, arXiv: 1002.2581 [hep-ph].

[66] J. Pumplin et al., *New Generation of Parton Distributions with Uncertainties from Global QCD Analysis*, 
JHEP **07** (2002) 012, arXiv: hep-ph/0201195.

[67] M. Czakon, P. Fiedler, and A. Mitov, 
*Total Top-Quark Pair-Production Cross Section at Hadron Colliders Through O(αS^4)*, 
Phys. Rev. Lett. **110** (2013) 252004, arXiv: 1303.6254 [hep-ph].
[68] H.-L. Lai et al., New parton distributions for collider physics, Phys. Rev. D 82 (2010) 074024, url: https://link.aps.org/doi/10.1103/PhysRevD.82.074024.

[69] GEANT4 Collaboration, S. Agostinelli, et al., GEANT4 – a simulation toolkit, Nucl. Instrum. Meth. A 506 (2003) 250.

[70] ATLAS Collaboration, The ATLAS Simulation Infrastructure, Eur. Phys. J. C 70 (2010) 823, arXiv: 1005.4568 [physics.ins-det].

[71] ATLAS Collaboration, The ATLAS Simulation Infrastructure, Eur. Phys. J. C 70 (2010) 823, arXiv: 1005.4568 [physics.ins-det].

[72] ATLAS Collaboration, Measurement of the muon reconstruction performance of the ATLAS detector using 2011 and 2012 LHC proton–proton collision data at \( \sqrt{s} = 13 \) TeV, Eur. Phys. J. C 74 (2014) 3130, arXiv: 1407.3935 [hep-ex].

[73] ATLAS Collaboration, Measurement of the muon reconstruction performance of the ATLAS detector using 2011 and 2012 LHC proton–proton collision data, Eur. Phys. J. C 74 (2014) 3130, arXiv: 1407.3935 [hep-ex].

[74] ATLAS Collaboration, Measurement of the muon reconstruction performance of the ATLAS detector using 2011 and 2012 LHC proton–proton collision data, Eur. Phys. J. C 74 (2014) 3130, arXiv: 1407.3935 [hep-ex].

[75] ATLAS Collaboration, Optimisation of large-radius jet reconstruction for the ATLAS detector in 13 TeV proton–proton collisions, Eur. Phys. J. C 81 (2020) 334, arXiv: 2009.04986 [hep-ex].

[76] 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 (2019) 135, arXiv: 1807.09477 [hep-ex].

[77] ATLAS Collaboration, Optimisation of large-radius jet reconstruction for the ATLAS detector in 13 TeV proton–proton collisions, Eur. Phys. J. C 81 (2020) 334, arXiv: 2009.04986 [hep-ex].

[78] ATLAS Collaboration, FastJet user manual, Eur. Phys. J. C 72 (2012) 1896, arXiv: 1111.6097 [hep-ph].

[79] ATLAS Collaboration, FastJet user manual, Eur. Phys. J. C 72 (2012) 1896, arXiv: 1111.6097 [hep-ph].

[80] ATLAS Collaboration, FastJet user manual, Eur. Phys. J. C 72 (2012) 1896, arXiv: 1111.6097 [hep-ph].

[81] D. Krohn, J. Thaler, and L.-T. Wang, Jet trimming, JHEP 02 (2010) 084, arXiv: 0912.1342 [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 (2019) 375, arXiv: 1808.07858 [hep-ex].

[83] ATLAS Collaboration, Performance of top-quark and W-boson tagging with ATLAS in Run 2 of the LHC, Eur. Phys. J. C 79 (2019) 375, arXiv: 1808.07858 [hep-ex].
[85] ATLAS Collaboration, *Performance of the ATLAS muon triggers in Run 2*, JINST 15 (2020) P09015, arXiv: 2004.13447 [physics.ins-det].

[86] ATLAS Collaboration, *The ATLAS Inner Detector Trigger performance in pp collisions at 13 TeV during LHC Run 2* (2021), arXiv: 2107.02485 [hep-ex].

[87] G. Choudalakis, *Fully Bayesian Unfolding*, arXiv (2012), arXiv: 1201.4612.

[88] C. Pollard, [https://github.com/cspollard/hunfold](https://github.com/cspollard/hunfold), 2018.

[89] ATLAS Collaboration, *Muon reconstruction efficiency and momentum resolution of the ATLAS experiment in proton–proton collisions at √s = 7 TeV in 2010*, Eur. Phys. J. C 74 (2014) 3034, arXiv: 1404.4562 [hep-ex].

[90] ATLAS Collaboration, *Jet Calibration and Systematic Uncertainties for Jets Reconstructed in the ATLAS Detector at √s = 13 TeV*, tech. rep. ATL-PHYS-PUB-2015-015, CERN, 2015, url: [https://cds.cern.ch/record/2037613](https://cds.cern.ch/record/2037613).

[91] ATLAS Collaboration, *Measurements of b-jet tagging efficiency with the ATLAS detector using t¯t events at √s = 13 TeV*, JHEP 08 (2018) 089, arXiv: 1805.01845 [hep-ex].

[92] ATLAS Collaboration, *Measurement of b-tagging efficiency of c-jets in t¯t events using a likelihood approach with the ATLAS detector*, tech. rep. ATLAS-CONF-2018-001, CERN, 2018, url: [https://cds.cern.ch/record/2306649](https://cds.cern.ch/record/2306649).

[93] ATLAS Collaboration, *Calibration of light-flavour b-jet mistagging rates using ATLAS proton-proton collision data at √s = 13 TeV*, tech. rep. ATLAS-CONF-2018-006, CERN, 2018, url: [https://cds.cern.ch/record/2314418](https://cds.cern.ch/record/2314418).

[94] S. Dulat et al., *New parton distribution functions from a global analysis of quantum chromodynamics*, Phys. Rev. D 93 (2016) 033006, arXiv: 1506.07443 [hep-ph].

[95] L. A. Harland-Lang, A. D. Martin, P. Motylinski, and R. S. Thorne, *Parton distributions in the LHC era: MMHT 2014 PDFs*, Eur. Phys. J. C 75 (2015) 204, arXiv: 1412.3989 [hep-ph].