**THE EUROPEAN
PHYSICAL JOURNAL C**

Tagging the initial-state gluon

Simone Caletti[1,a], **Oleh Fedkevych**[1,b], **Simone Marzani**[1,c], **Daniel Reichelt**[2,d]

[1] Dipartimento di Fisica, Università di Genova and INFN, Sezione di Genova, Via Dodecaneso 33, 16146 Genoa, Italy
[2] Institut für Theoretische Physik, Georg-August-Universität Göttingen, Friedrich-Hund-Platz 1, 37077 Göttingen, Germany

Abstract We study the production of an electroweak boson in association with jets, in processes where the jet with the highest transverse momentum is identified as quark-initiated. The quark/gluon tagging procedure is realised by a cut on a jet angularity and it is therefore theoretically well-defined and exhibits infrared and collinear safety. In this context, exploiting resummed perturbation theory, we are able to provide theoretical predictions for transverse momentum distributions at a well-defined and, in principle, systematically improvable accuracy. In particular, tagging the leading jet as quark-initiated allows us to enhance the initial-state gluon contribution. Thus these novel transverse momentum distributions are potentially interesting observables to probe the gluonic degrees of freedom of the colliding protons.

1 Introduction

Over the past decade our understanding of the internal structure of jets has increased tremendously. Thanks to the applications of the methods of perturbative QCD, the field has become mature and jet substructure algorithms that are, at the same time, performant and robust have been developed. Observables originally designed for searching new physics are now the target of unfolded measurements that can be compared to theoretical predictions, at the precision level, see e.g. [1–8]. Furthermore, ideas developed by the jet substructure community have found applications in other contexts of particle physics. For instance, after it was realised that the momentum fraction that characterises the splitting identified by the SoftDrop algorithm [9] follows a distribution dictated by the QCD splitting functions [10–13], this observable has become a standard way of probing interactions with the quark/gluon plasma, see e.g. [14] and references therein.

This work is part of an incipient effort to find innovative ways of applying jet substructure techniques to the broader LHC physics program and to provide the necessary tools so that this cross-pollination can bear its fruits. In this context, measurements of the internal structure of highly energetic jets produced in association with a boosted electroweak boson offer an ideal playground for such studies, because the leptonic decay of the Z boson offers a valuable trigger. These events are not only relevant for background studies[1] but, as we shall detail in the following, they open up novel possibilities to employ jet substructure for Standard Model measurements.

Substructure observables, such as jet angularities [15], measure the pattern of the hadronic activity within a jet. For this reason, they are often employed as tagging variables that aim to distinguish jets that have been originated by elementary particles carrying different colour degrees of freedom, e.g. colour singlets versus QCD partons or, even, quarks versus gluons. Although there exist more powerful quark/gluon (q/g) discriminants than a cut on a jet angularity, this procedure is theoretically well-defined, infra-red and collinear (IRC) safe and, hence, its behaviour can be understood with perturbative methods.

In the following, we study Z+jet production, requiring that the jet with the highest transverse momentum has been identified as quark-initiated. As preliminarily explored in [16], by tagging the final-state, we indirectly bias particular partonic sub-processes. If the leading jet is tagged as quark-initiated, then, at leading order, the subprocess that features a quark and gluon in the initial state is enhanced. Thus, this procedure could potentially provide us with a new handle on the determination of the gluon parton distribution function (PDF). The main result of this study is the calculation

a e-mail: simone.caletti@ge.infn.it

b e-mail: oleh.fedkevych@ge.infn.it

c e-mail: simone.marzani@ge.infn.it (corresponding author)

d e-mail: daniel.reichelt@uni-goettingen.de

---

[1] This process represents the main background for the associated production of a Higgs and an electroweak boson in the boosted regime, where the decay products of the Higgs are reconstructed in one jet.

of the transverse momentum distribution of the Z boson in events where the leading jet is identified as quark-initiated. Because of IRC safety of the tagging procedure, we are able to compute this distribution at a well-defined and, in principle improvable, accuracy. Therefore, this observable could be directly included in standard fits of PDFs. The results presented here account for the resummation of the tagging parameter at the *next-to-leading logarithmic* (NLL) accuracy, matched to fixed order predictions at $O\left(\alpha_s^2\right)$ with respect to the Born contribution, henceforth denoted as next-to-leading order (NLO).[2]

2 Enhancing the gluon contribution

We start by considering the channel fractions f_{ij} that measure, to a given order in perturbation theory, the contribution to the Z (or jet) transverse momentum distributions from the subprocess initiated by partons i and j, which, for brevity, we indicate as σ_{ij}^a, where a could be the Z boson or the leading jet J:

$$f_{ij}^a = \frac{\sigma_{ij}^a}{\sigma_{qq}^a + \sigma_{qg}^a + \sigma_{gg}^a}, \quad a = Z, J, \tag{1}$$

where qq and qg include any combination of quarks and anti-quarks. Our aim is to study how tagging a quark-jet in the final-state changes the relative contributions of the various partonic subprocesses. To this purpose, we define

$$\widetilde{f}_{ij}^a = \frac{\widetilde{\sigma}_{ij}^a}{\widetilde{\sigma}_{qq}^a + \widetilde{\sigma}_{qg}^a + \widetilde{\sigma}_{gg}^a}, \quad a = Z, J, \tag{2}$$

where the tilde indicates "after tagging". Furthermore, in what follows we will be interested in the fractions of the event (before and after tagging) which feature at least one gluon in the initial state. To this purpose, we define

$$f_g^a = f_{qg}^a + f_{gg}^a, \quad \text{and} \quad \widetilde{f}_g^a = \widetilde{f}_{qg}^a + \widetilde{f}_{gg}^a. \tag{3}$$

We will refer to this fraction as *gluon channel purity*.

We begin our discussion in a simplified setting, which is nevertheless enough to capture the essential physics points. We will then validate our conclusions using Monte Carlo (MC) parton shower simulations. If we consider the Born approximation, then we have no gg contribution and, because

the Z boson and the jet are back-to-back, we obtain $f_{ij}^Z = f_{ij}^J$. Fragmentation of the hard partons can lead to transverse momentum imbalance, however, we can limit ourselves to the leading logarithmic (LL) regime, where all parton splittings happen in the soft and collinear limit, with no recoil.[3] Therefore, dropping the superscript a, we have

$$f_g = \frac{\sigma_{qg}}{\sigma_{qq} + \sigma_{qg}}. \tag{4}$$

We note that, within our approximation, the fraction of events with a (properly defined) final-state quark can be considered as a proxy for the gluon channel purity. We find that, for $p_{tJ} \geq 100$ GeV, $f_g \simeq 0.85$ and it exhibits a rather mild dependence on the transverse momentum. Next, we note that in our approximation we simply have

$$\widetilde{f}_g = \frac{\varepsilon_q \sigma_{qg}}{\varepsilon_g \sigma_{qq} + \varepsilon_q \sigma_{qg}} = \left(1 + \frac{1 - f_g}{f_g} \frac{\varepsilon_g}{\varepsilon_q}\right)^{-1}, \tag{5}$$

where ε_q is the efficiency of the tagger to correctly label quark jets and ε_g the false-positive rate, which measures how often gluon jets are wrongly labelled as quarks. We note that a perfect quark-tagger with $\varepsilon_q = 1$ and $\varepsilon_g = 0$ returns tagged events that have been originated by one gluon in the initial state, i.e. $\widetilde{f}_g = 1$. On the other hand, with an efficiency of 50%, which corresponds to tossing a coin, we recover Eq. (4).

A rather common class of q/g taggers exhibits at LL a property known as Casimir scaling (CS) [15], namely the tagging efficiencies are related by $\varepsilon_g = \left(\varepsilon_q\right)^{C_A/C_F}$, where the exponent is given by the ratio of the Casimir operators in the appropriate colour representation, C_F for quarks and C_A for gluons. This property emerges because the LL distributions of both quarks and gluons exhibit the same Sudakov-like behaviour, with a coefficient determined by the appropriate colour factor. Thus, for a CS tagger, we find

$$\widetilde{f}_g^{\mathrm{CS}} = \left(1 + \frac{1 - f_g}{f_g} \varepsilon_q^{\frac{C_A}{C_F} - 1}\right)^{-1}. \tag{6}$$

For instance, a CS tagger with $\varepsilon_q = 0.65$, yields $\widetilde{f}_g^a \simeq 0.9$. This can be pushed to 0.95 if the tighter working point $\varepsilon_q = 0.35$ is considered. Taggers that obey CS are not the most performant, but they are interesting for us because they are under good theoretical control. The efficiencies ε_i can be computed in QCD using resummed perturbation theory to a well-defined and, in principle improvable, theoretical accuracy. Furthermore, the inclusion of higher logarithmic corrections generally leads to an improvement with respect

---

[2] The fixed-order counting is slightly different than what is usually employed in standard transverse-momentum distribution. This has to do with the fact that, in order to obtain a non-vanishing value of any angularity, the jet must have at least two constituents. Thus, in what follows we will refer to the lowest-order ($2 \rightarrow 2$) scattering as Born approximation, while (N)LO will be reserved for the contributions with one (two) additional emission(s).

---

[3] We will lift this approximation when dealing with actual simulations. However, we can always choose a set of kinematical cuts that favours the back-to-back configuration.

to strict CS, so that, depending on the specifics of the tagging procedure, the target gluon channel purity $\widetilde{f}_g^a \simeq 0.95$ can be achieved at a reasonable working point. In the following we construct a CS tagger that is based on a particular class of jet substructure observables known as jet angularities.

3 Jet angularities as a quark/gluon tagger

Jet angularities [15] are defined as

$$\lambda_\alpha = \sum_{i \in \text{jet}} \frac{p_{t,i}}{p_{t\,J}} \left(\frac{\Delta_i}{R_0} \right)^\alpha, \qquad (7)$$

where the sum runs over the constituents of the hardest jet in the event and $\Delta_i = \sqrt{(y_i - y_J)^2 + (\phi_i - \phi_J)^2}$ is the distance in the azimuth-rapidity plane of particle i from the jet axis. We define jets with the anti-k_t clustering algorithm [17] with radius R_0 and standard E-scheme for recombination. IRC safety requires $\alpha > 0$, while angularities with $\alpha \leq 1$ are sensitive to recoil against soft emissions [18]. In order to circumvent this issue, when $\alpha \leq 1$, the jet axis is obtained using the Winner–Take–All (WTA) recombination scheme [19]. We also consider groomed jets. In this case, we recluster the jet with the Cambridge–Aachen algorithm [20,21] and apply the SoftDrop grooming algorithm with parameters z_{cut} and β [9]. The angularity is then computed on the constituents of the groomed jet, with the WTA prescription adopted for angularities with $\alpha \leq 1$.

Because of the different colour factor, angularity distributions for quark- and gluon-initiated jets peak at different values. We can exploit this separation and define our q/g tagger through a cut on the jet angularity. In particular, a jet with $\lambda_\alpha < \lambda_{\text{cut}}$ will be labelled as a quark jet:

$$\widetilde{\sigma}_{ij}^a = \int_0^{\lambda_{\text{cut}}} d\lambda_\alpha \frac{d\sigma_{ij}^a}{d\lambda_\alpha}, \qquad (8)$$

where we have introduced the differential angularity distribution. The implicit dependence of the tagged distribution on the transverse momentum, on the angularity exponent α and, optionally, on the SoftDrop parameters z_{cut} and β, is understood.

From the simple CS analysis above, we have concluded that we should work with quark efficiencies $\varepsilon_q \simeq 0.35$, in order to reach a purity of initial-state gluons around 0.95. The value of the angularity cut that is necessary to achieve this working point for the tagger clearly depends on the angular exponent α in Eq. (7) as well as on the parameters z_{cut} and β of the SoftDrop algorithm, should we wish to employ groomed jets. Different theoretical considerations can guide us with this choice. First of all, we would like to preserve calculability, i.e. we want to cement our findings in perturbative

field theory. Thus, we would like our tagger to be as insensitive as possible to non-perturbative (NP) contributions such as hadronisation corrections and the Underlying Event (UE). Secondly, although we can calculate transverse momentum spectra in resummed perturbation theory [22–24], we aim for perturbative stability. Thus, we favour working points for the tagger for which λ_{cut} is not too small.

In order to turn the above considerations into a quantitative study we use simulated data obtained with the MC event generator PYTHIA 8.303 [25]. The UE is simulated according to the model presented in [26–28] and hadronisation effects according to the Lund string model [29,30]. Throughout the paper, we use the NNPDF 3.0 NLO set of PDFs [31]. We consider the inclusive production of a pair of oppositely charged muons in proton–proton collisions at 13 TeV centre-of-mass energy, requiring that the invariant mass of the muon pair to be within 70 and 110 GeV. Jets are clustered with the anti-k_t algorithm with $R_0 = 0.4$ and ordered in transverse momentum. Henceforth, the jet will be implicitly considered to be the hardest one and we will refer to the muon-antimuon pair as the Z boson. With this in mind, the fiducial volume is defined following Ref. [8]: $p_{t\,\mu} > 26$ GeV, $p_{t\,Z} > 30$ GeV, $p_{t\,J} > 15$ GeV, $|\eta_\mu| < 2.4$, and $|y_{\text{jet}}| < 1.7$. Furthermore, in order to enforce back-to-back configurations, we impose $\left| \frac{p_{t\,J} - p_{t\,Z}}{p_{t\,J} + p_{t\,Z}} \right| < 0.3$, $\left| \phi_{\text{jet}} - \phi_Z \right| > 2$.

For event selection and analysis we employ RIVET [32, 33]. Jet reconstruction is done with FAST JET [34], and the SoftDrop implementation in the FAST JET `contrib` is used. We distinguish two different stages of the simulation: "parton-level", i.e. with parton shower effects only, and "hadron-level", i.e. with hadronisation and UE included.

By keeping the two partonic processes of interest separate, we compute Receiver Operating Characteristic (ROC) curves that show the mis-tag rate (gluon efficiency) ε_g as a function of the signal (quark) efficiency. They are computed for different values of the angularity exponent $\alpha = 0.5, 1, 2$ in the ungroomed case and for SoftDrop jets with $z_{\text{cut}} = 0.1$ and $\beta = 0, 1$. We show hadron-level results in Fig. 1, as well as the ratios to their parton-level counterparts, which we take as a measure of NP contributions. The dotted portion of the latter indicates that the efficiency ε_q is dominated by splittings in the non-perturbative region, as determined, for instance, in Ref. [24]. We also show the target line, which fixes the gluon efficiency ε_g as a function of ε_q, for given f_g and $\widetilde{f}_g$,

$$\varepsilon_g = \frac{f_g (1 - \widetilde{f}_g)}{\widetilde{f}_g (1 - f_g)} \varepsilon_q, \qquad (9)$$

which is easily derived from Eq. (5). The slope shown in Fig. 1 is determined by the original $f_g = 0.85$ and target gluon purity $\widetilde{f}_g = 0.95$. The intersections of each ROC
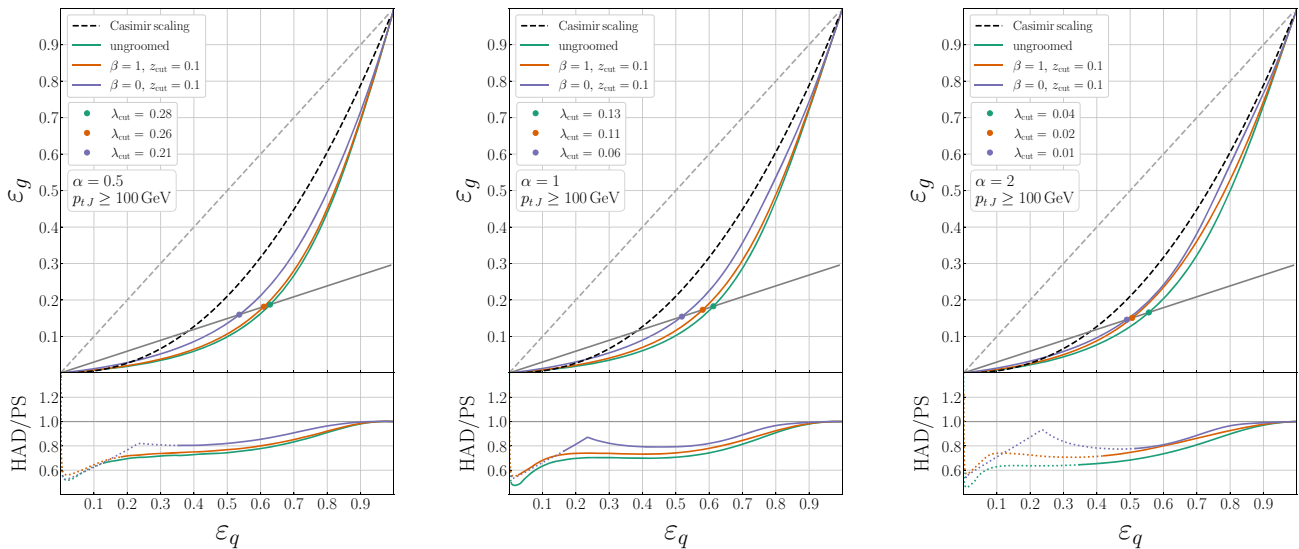
Fig. 1 Receiver Operating Characteristic (ROC) curves for different values of the angularity exponent $\alpha = 0.5, 1, 2$, from left to right obtained with MC simulation using PYTHIA. In each plot, different curves refer to ungroomed jets and SoftDrop jets with $z_{cut} = 0.1$ and $\beta = 0, 1$. For comparison we also show the ROC curves corresponding to a random classifier (dashed grey) and the CS ones (dashed black). Straight lines represent the target performance, as detailed in the text.

At the bottom of each plot we show the ratio between each ROC curve obtained at hadron-level and the corresponding parton-level result. The dashed portion of these lines indicate the region sensitive to splittings with relative transverse momentum below 1 GeV, where NP effects are expected to be sizeable. This region of phase-space is sensitive to shower cut-off effects, which cause the observed kinks

curve with the target line set our tagger working points. Corresponding values of λ_{cut} are reported in the figure. The choice of the tagger working points is clearly not unique. For instance, we could have optimised the signal-to-background ratio by choosing on each ROC curve the point that is closest to the $(1, 0)$ corner.

The analysis of Ref. [24] tells us that larger values of α are under better theoretical control. However, as it is clear from Fig. 1, lower values of α have better performance, essentially because of their increased sensitivity to the collinear region. Thus, the choice $\alpha = 1$ appears to be a good compromise between performance and robustness. We note that the use of SoftDrop does not always provide us with improvements on the size of the NP contributions. This might be related to the fact that in order to obtain the same efficiency ε_q we need to cut groomed jets at lower values of λ_{cut}, where NP physics may be more prominent. Furthermore, we note that we are working with a rather small jet radius, which prevents large contributions from the UE. We expect (light) grooming to be beneficial, should one consider larger jet radii. Finally, configurations with aggressive SoftDrop, i.e. $\beta = 0$, typically result in worse performance, because important information is groomed away. Thus, in what follows, we shall focus on the $\alpha = 1$ case either with no grooming or with $\beta = 1$, $z_{cut} = 0.1$.

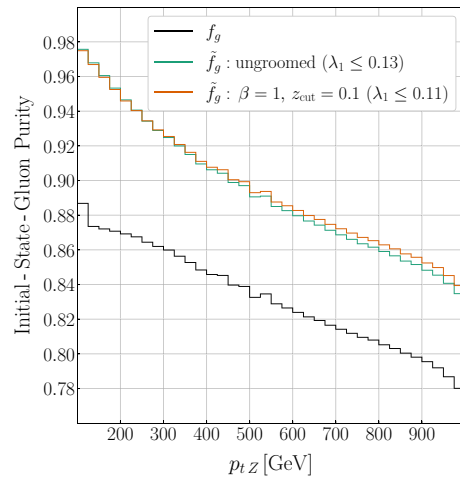We now quantify the gain in the initial-state gluon purity that we obtain after tagging. Figure 2 shows the fractions

Fig. 2 The initial-state gluon purity before (f_g) and after tagging ($\tilde{f}_g$) obtained with PYTHIA simulations, as a function of the transverse momentum of the Z boson

$\tilde{f}_g^Z$ as a function of the Z boson transverse momentum for the selected taggers, with f_g^Z also shown for comparison. We first note that the performance of both taggers is very good, leading to gluon channel purities that exceed our 95% target. Analogous conclusions can also be drawn if we plot our results as a function of $p_{t\,J}$. However, as we will argue shortly, in this context, the transverse momentum distribution of the Z boson is a more robust observable. We also notice

that the gluon channel purities decrease with $p_{t\,Z}$. This is due to the fact that our taggers are defined looking at their efficiencies with $p_{t\,J} > 100\,$GeV and, therefore, we expect them to work better at the lower end of the transverse momentum spectrum. This loss in performance could be cured by adjusting the cut on the angularity as a function of $p_{t\,J}$. However, in this first study, we prefer to keep our framework simple.

4 Transverse momentum distributions

We now provide theoretical predictions for our observables of interest, namely transverse-momentum distributions, in the presence of tagging. Our calculation includes the resummation of logarithms of λ_{cut} at NLL accuracy and is matched to NLO. Thanks to a flavour-dependent matching procedure, *cf.* also [35,36], we are able to achieve NLO+NLL' accuracy. We also include a bin-by-bin NP correction factor obtained with MC simulations.[4]

Our results are reported in Fig. 3, where we show the $p_{t\,Z}$ distribution for events where the highest-p_t jet is quark initiated, for the taggers selected for this study. We show the $p_{t\,Z}$ distribution with no-tagging, with tagging on standard jets and with tagging on SoftDrop jets. We note that NP corrections are rather sizeable at low $p_{t\,Z}$, making this observable most reliable in the high transverse momentum region. The latter is actually per se interesting because it allows us to probe the proton dynamics described by the PDFs at large values of the momentum fraction x, i.e. in a kinematic region where they are less constrained. To qualitatively assess the values of x which would be accessible, we show, on the upper horizontal axis, the Born-level momentum fraction computed at central (zero) rapidities:

$$\bar{x} = x_{1,2} = \left.\frac{p_t e^{\pm y_J} + \sqrt{p_t^2 + m_Z^2} e^{\pm y_Z}}{\sqrt{S}}\right|_{y_J = y_Z = 0}.$$

We have also studied $p_{t\,J}$ distributions. However, as anticipated, these distributions turn out to be less robust, essentially because jet dynamics can be significantly altered by the cut on the angularity, as well as by the grooming procedure.[5] In contrast, the $p_{t\,Z}$ spectrum is inclusive with respect to the jet activity and thus, for a given working point of the tagger, less dependent on the details of the tagging procedure itself. We can therefore take modifications in such distribution as

---

[4] See Ref. [24] for details about the calculation and its numerical implementation in the resummation plugin [37,38] to the SHERPA [39,40] framework, including perturbative uncertainties, obtained by varying the perturbative (renormalisation, factorisation and resummation) scales and NP corrections. We use COMIX [41] in conjunction with Open-Loops [42] and Recola [43,44] for the fixed order calculation. The NP corrections are based on SHERPA parton shower simulations at MC@NLO accuracy [45,46] hadronised with SHERPA's cluster fragmentation model [47].

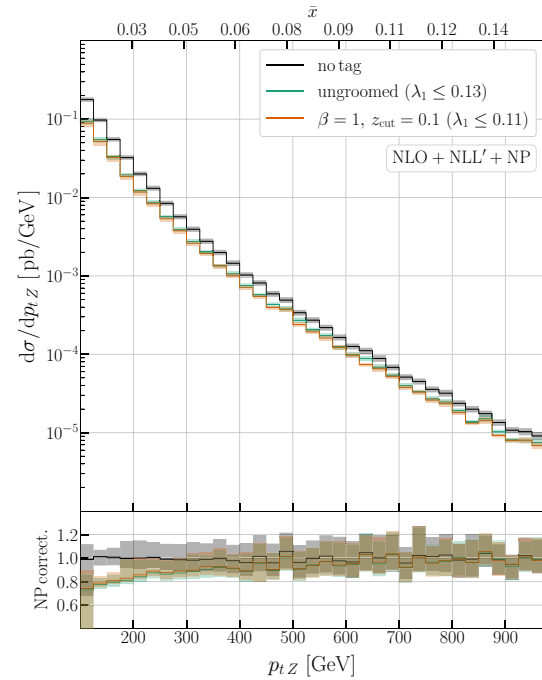[5] Note that if $\beta = 0$, the groomed $p_{t\,J}$ distribution is not even IRC safe.

Fig. 3 Transverse momentum distribution of the Z boson in Z+jet events, with the leading jet tagged as quark-initiated, according to our operational definition, detailed in the text. The untagged distribution is also shown for comparison. The NLO+NLL' calculation is supplemented with a NP correction factor, shown at the bottom

more directly related to the bias that the tagger induces on the composition of the partonic initial state, which is what we want to achieve.

5 Conclusions and future developments

We have shown how q/g tagging can be successfully applied to Z+jets events in order to significantly enhance the gluon-initiated contributions. In particular, our tagger is realised through a simple cut on a jet angularity, which is an IRC safe observable and therefore can be studied in perturbation theory. Exploiting MC simulations, we have performed a study of their efficiencies and their dependence on NP effects, exploring different angularities and different levels of grooming. We have explicitly studied the transverse momentum of the Z boson, providing theoretical predictions that included both all-order resummation and matching to NLO. We have shown that we can achieve initial-state gluon purities close to 95%.

We see several possible directions for future work in this context. First, we would like to assess the impact of this type of observables on PDF fits. In particular, while our study does show an increase in the gluon purity, this comes at the cost of a noticeable reduction of the available dataset. Thus, despite the relative large cross-section of the process

we are considering, we may need to optimise the tagger's working point, taking this further aspect into consideration. For instance, as mentioned above, we could choose λ_{cut} so that the signal-to-background ratio is maximised. Second, the results presented in this study are based on a calculation of the angularity spectra, and hence of closely related efficiencies, at NLO+NLL$'$. The resummed calculation can be promoted to higher accuracy [48–50]. The fixed-order can also be improved by including the two-loop correction to Z+1 jet (see [51] and references therein). An improvement in the description of the angularity distribution away from $\lambda_\alpha = 0$ is much more challenging because it requires Z+2 partons at NNLO, which may become available in the near future. Finally, with the aim of enhancing performance while maintaining calculability, it would be interesting to consider more powerful, albeit more sophisticated, taggers. In this context, the Les Houches multiplicity [16] is very promising and, although its theoretical understanding is only in its infancy, we believe that achieving NLL accuracy is within reach.

Data Availability Statement This manuscript has no associated data or the data will not be deposited. [Authors' comment: For this publication only computer-generated pseudo data were used. These have been obtained with either publicly available Monte Carlo event generators or with private code, which is made available by the Authors upon request.]

References

1. ATLAS Collaboration, G. Aad et al., Observation of electroweak production of two jets and a Z-boson pair with the ATLAS detector at the LHC, arXiv:2004.10612
2. ATLAS Collaboration, G. Aad et al., Measurement of soft-drop jet observables in pp collisions with the ATLAS detector at $\sqrt{s}$ =13 TeV. Phys. Rev. D **101**(5), 052007 (2020). arXiv:1912.09837
3. ATLAS Collaboration, G. Aad et al., Properties of jet fragmentation using charged particles measured with the ATLAS detector in pp collisions at $\sqrt{s}$ = 13 TeV. Phys. Rev. D **100**(5), 052011 (2019). arXiv:1906.09254
4. ATLAS Collaboration, M. Aaboud et al., Measurement of the soft-drop jet mass in pp collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector. Phys. Rev. Lett. **121**(9), 092001 (2018). arXiv:1711.08341
5. ATLAS Collaboration, M. Aaboud et al., Measurement of jet-substructure observables in top quark, W boson and light jet production in proton–proton collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector. JHEP **08**, 033 (2019). arXiv:1903.02942
6. CMS Collaboration, A.M. Sirunyan et al., 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
7. CMS Collaboration, A.M. Sirunyan et al., Measurement of the groomed jet mass in PbPb and pp collisions at $\sqrt{s_{\mathrm{NN}}}$ = 5.02 TeV. *JHEP* **10**, 161 (2018). arXiv:1805.05145
8. CMS Collaboration, Study of quark and gluon jet substructure in dijet and Z+jet events from pp collisions. CMS-PAS-SMP-20-010 (2021)
9. A.J. Larkoski, S. Marzani, G. Soyez, J. Thaler, Soft drop. JHEP **05**, 146 (2014). arXiv:1402.2657
10. A.J. Larkoski, S. Marzani, J. Thaler, Sudakov safety in perturbative QCD. Phys. Rev. D **91**(11), 111501 (2015). arXiv:1502.01719
11. A. Larkoski, S. Marzani, J. Thaler, A. Tripathee, W. Xue, Exposing the QCD splitting function with CMS open data. Phys. Rev. Lett. **119**(13), 132003 (2017). arXiv:1704.05066
12. A. Tripathee, W. Xue, A. Larkoski, S. Marzani, J. Thaler, Jet substructure studies with CMS open data. Phys. Rev. D **96**(7), 074003 (2017). arXiv:1704.05842
13. P. Cal, K. Lee, F. Ringer, W.J. Waalewijn, The soft drop momentum sharing fraction z_g beyond leading-logarithmic accuracy. arXiv:2106.04589
14. Y. Chen, Jet substructure and parton splitting: an experimental overview. Nucl. Phys. A **1005**, 122010 (2021)
15. A.J. Larkoski, J. Thaler, W.J. Waalewijn, Gaining (mutual) information about quark/gluon discrimination. JHEP **11**, 129 (2014). arXiv:1408.3122
16. S. Amoroso et al., Les Houches 2019: Physics at TeV Colliders: Standard Model Working Group Report, in *11th Les Houches Workshop on Physics at TeV Colliders: PhysTeV Les Houches (PhysTeV 2019) Les Houches, France, June 10-28, 2019*, 2020. arXiv:2003.01700
17. M. Cacciari, G.P. Salam, G. Soyez, The anti-k_t jet clustering algorithm. JHEP **04**, 063 (2008). arXiv:0802.1189
18. A.J. Larkoski, G.P. Salam, J. Thaler, Energy correlation functions for jet substructure. JHEP **06**, 108 (2013). arXiv:1305.0007
19. A.J. Larkoski, D. Neill, J. Thaler, Jet shapes with the broadening axis. JHEP **04**, 017 (2014). arXiv:1401.2158
20. Y.L. Dokshitzer, G.D. Leder, S. Moretti, B.R. Webber, Better jet clustering algorithms. JHEP **08**, 001 (1997). arXiv:hep-ph/9707323
21. M. Wobisch, T. Wengler, Hadronization corrections to jet cross-sections in deep inelastic scattering, in Monte Carlo generators for

HERA physics. Proc. Workshop Hamburg Germany **1998–1999**, 270–279 (1998). arXiv:hep-ph/9907280

22. Z.-B. Kang, K. Lee, F. Ringer, Jet angularity measurements for single inclusive jet production. JHEP **04**, 110 (2018). arXiv:1801.00790

23. Z.-B. Kang, K. Lee, X. Liu, F. Ringer, Soft drop groomed jet angularities at the LHC. Phys. Lett. B **793**, 41–47 (2019). arXiv:1811.06983

24. S. Caletti, O. Fedkevych, S. Marzani, D. Reichelt, S. Schumann, G. Soyez, V. Theeuwes, Jet Angularities in Z+jet production at the LHC. JHEP **07**, 076 (2021). arXiv:2104.06920

25. T. Sjöstrand, S. Ask, J.R. Christiansen, R. Corke, N. Desai, P. Ilten, S. Mrenna, S. Prestel, C.O. Rasmussen, P.Z. Skands, An introduction to PYTHIA 8.2. Comput. Phys. Commun. **191**, 159–177 (2015). arXiv:1410.3012

26. T. Sjöstrand, Multiple parton–parton interactions in hadronic events, in *Proceedings, 23RD International Conference on High Energy Physics, July 16–23, 1986* (Berkeley CA, 1985)

27. T. Sjöstrand, M. van Zijl, A multiple interaction model for the event structure in hadron collisions. Phys. Rev. D **36**, 2019 (1987)

28. T. Sjöstrand, P.Z. Skands, Multiple interactions and the structure of beam remnants. JHEP **03**, 053 (2004). arXiv:hep-ph/0402078

29. B. Andersson, G. Gustafson, G. Ingelman, T. Sjöstrand, Parton fragmentation and string dynamics. Phys. Rep. **97**, 31–145 (1983)

30. T. Sjöstrand, Jet fragmentation of nearby partons. Nucl. Phys. B **248**, 469–502 (1984)

31. NNPDF Collaboration, R.D. Ball et al., Parton distributions for the LHC Run II. JHEP **04**, 040 (2015). arXiv:1410.8849

32. A. Buckley, J. Butterworth, D. Grellscheid, H. Hoeth, L. Lonnblad, J. Monk, H. Schulz, F. Siegert, Rivet user manual. Comput. Phys. Commun. **184**, 2803–2819 (2013). arXiv:1003.0694

33. C. Bierlich et al., Robust independent validation of experiment and theory: Rivet version 3. SciPost Phys. **8**, 026 (2020). arXiv:1912.05451

34. M. Cacciari, G.P. Salam, G. Soyez, FastJet user manual. Eur. Phys. J. C **72**, 1896 (2012). arXiv:1111.6097

35. A. Banfi, G.P. Salam, G. Zanderighi, Infrared safe definition of jet flavor. Eur. Phys. J. C **47**, 113–124 (2006). arXiv:hep-ph/0601139

36. J. Baron, D. Reichelt, S. Schumann, N. Schwanemann, V. Theeuwes, Soft-drop grooming for hadronic event shapes. arXiv:2012.09574

37. E. Gerwick, S. Hoeche, S. Marzani, S. Schumann, Soft evolution of multi-jet final states. JHEP **02**, 106 (2015). arXiv:1411.7325

38. N. Baberuxki, C.T. Preuss, D. Reichelt, S. Schumann, Resummed predictions for jet-resolution scales in multijet production in e^+e^- annihilation. JHEP **04**, 112 (2020). arXiv:1912.09396

39. T. Gleisberg, S. Hoeche, F. Krauss, M. Schonherr, S. Schumann, F. Siegert, J. Winter, Event generation with SHERPA 1.1. JHEP **02**, 007 (2009). arXiv:0811.4622

40. Sherpa Collaboration, E. Bothmann et al., Event Generation with Sherpa 2.2. SciPost Phys. **7**(3), 034 (2019). arXiv:1905.09127

41. T. Gleisberg, S. Hoeche, Comix, a new matrix element generator. JHEP **12**, 039 (2008). arXiv:0808.3674

42. F. Buccioni, J.-N. Lang, J.M. Lindert, P. Maierhöfer, S. Pozzorini, H. Zhang, M.F. Zoller, OpenLoops 2. Eur. Phys. J. C **79**(10), 866 (2019). arXiv:1907.13071

43. S. Actis, A. Denner, L. Hofer, J.-N. Lang, A. Scharf, S. Uccirati, RECOLA: REcursive computation of one-loop amplitudes. Comput. Phys. Commun. **214**, 140–173 (2017). arXiv:1605.01090

44. B. Biedermann, S. Bräuer, A. Denner, M. Pellen, S. Schumann, J.M. Thompson, Automation of NLO QCD and EW corrections with Sherpa and Recola. Eur. Phys. J. C **77**, 492 (2017). arXiv:1704.05783

45. S. Frixione, B.R. Webber, Matching NLO QCD computations and parton shower simulations. JHEP **06**, 029 (2002). arXiv:hep-ph/0204244

46. S. Hoeche, F. Krauss, M. Schonherr, F. Siegert, A critical appraisal of NLO+PS matching methods. JHEP **09**, 049 (2012). arXiv:1111.1220

47. J.-C. Winter, F. Krauss, G. Soff, A Modified cluster hadronization model. Eur. Phys. J. C **36**, 381–395 (2004). arXiv:hep-ph/0311085

48. C. Frye, A.J. Larkoski, M.D. Schwartz, K. Yan, Precision physics with pile-up insensitive observables, arXiv:1603.06375

49. 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

50. A. Kardos, A.J. Larkoski, Z. Trócsányi, Groomed jet mass at high precision. Phys. Lett. B **809**, 135704 (2020). arXiv:2002.00942

51. A. GehrmannDeRidder, T. Gehrmann, E.W.N. Glover, A. Huss, T.A. Morgan, Precise QCD predictions for the production of a Z boson in association with a hadronic jet. Phys. Rev. Lett. **117**(2), 022001 (2016). arXiv:1507.02850

52. J.D. Hunter, Matplotlib: a 2D graphics environment. Comput. Sci. Eng. **9**(3), 90–95 (2007)

53. T. Oliphant, *Guide to NumPy* (Trelgol Publishing, New York, 2006)