Eur. Phys. J. C manuscript No. (will be inserted by the editor)

The jet mass distribution after Soft Drop

Simone Marzani^{a,1}, Lais Schunk^{b,2}, Gregory Soyez^{c,3}

December 2017

Abstract We present a first-principle computation of the mass distribution of jets which have undergone the grooming procedure known as Soft Drop. This calculation includes the resummation of the large logarithms of the jet mass over its transverse momentum, up to next-to-logarithmic accuracy, matched to exact fixed-order results at next-to-leading order. We also include non-perturbative corrections obtained from Monte-Carlo simulations and discuss analytic expressions for hadronisation and Underlying Event effects.

Introduction. The study of jets at the Large Hadron Collider (LHC) has recently taken a new turn with new substructure observables [1, 2] amenable to precise theory calculations [3–5], including genuine theory uncertainty bands, and corresponding experimental measurements from both the CMS [6] and ATLAS [7] collaborations. The substructure techniques we concentrate on are usually referred to as grooming and they aim to reduce sensitivity to non-perturbative corrections and pileup.

A first series of studies has focused on the jet mass after applying the (modified) MassDrop Tagger (mMDT) [1, 8] in dijet events, as measured by the CMS collaboration [6]. On the theory side, the description of this observable requires to match a resummed calculation, important in the small-mass region, to fixed-order results, which are important for large masses. The former are obtained analytically, including to all orders terms enhanced by the large logarithms of p_t^2/m^2 with p_t the jet transverse momentum and m the (groomed) jet mass. The latter is obtained from fixed-order Monte-Carlo simulations. To date, two theory calculations are available: a SCET-based next-to-leading logarithmic (NLL) resummation in the small z_{cut} limit, matched to leading order (LO) results [4], and our previous study

matching a leading logarithmic resummation, including finite (but small) z_{cut} effects, to next-to-leading order results [5]. Comparing both predictions, we see a small NLL effect at small mass and non-negligible NLO corrections at large mass.

The goal of the present letter is to extend our mMDT study from Ref. [5] to the case of Soft Drop [2], i.e. allowing for a non-zero value of the angular exponent β . When $\beta \neq 0$, the logarithmic counting differs from the mMDT case, essentially because Soft Drop retains soft-collinear radiation, which is always groomed away by mMDT. In this case, the SCET-based calculation from Ref. [4] reaches NNLL accuracy and it is matched, in the dijet case, to LO fixed-order results. Here, we present the results of a NLL resummation matched to NLO fixed-order accuracy.

After a brief review of the Soft Drop procedure, we will present our results first in the resummation region, then matched to fixed-order. We then provide an analytic estimate of non-perturbative corrections, extending to the Soft Drop case the analytic results obtained in Ref. [1] for the mMDT. We conclude by providing and discussing our final predictions, including the theory uncertainty bands. These have already been compared to experimental data in [7], where a good agreement was found, especially in the perturbative region.

Soft Drop. For a given jet, the Soft Drop procedure first reclusters the constituents of the jet with the Cambridge/Aachen algorithm [9] into a single jet j. Starting from j, it then applies the following iterative procedure:

- 1. undo the last clustering step $j \rightarrow j_1, j_2$, with $p_{t1} > p_{t2}$.
- 2. stop the procedure if the Soft Drop condition is met:

$$\frac{\min(p_{t1}, p_{t2})}{p_{t1} + p_{t2}} > z_{\text{cut}} \left(\frac{\theta_{12}}{R}\right)^{\beta},\tag{1}$$

where $z_{\rm cut}$ and β are free parameters, $\theta_{12}^2 = \Delta y_{12}^2 + \Delta \phi_{12}^2$ and R the original jet radius.

¹Dipartimento di Fisica, Università di Genova and INFN, Sezione di Genova, Via Dodecaneso 33, 16146, Italy

²Deutsches Elektronen-Synchroton DESY, Notkestraße 85, 22607 Hamburg, Germany

³IPhT, CEA Saclay, CNRS UMR 3681, F-91191 Gif-Sur-Yvette, France

^ae-mail: simone.marzani@ge.infn.it

^be-mail: lais.sarem.schunk@desy.de

ce-mail: gregory.soyez@ipht.fr

3. otherwise, set $j = j_1$ and go back to 1, or stop if j_1 has no further substructure.

The limit $\beta \to 0$ corresponds to the mMDT.

NLL resummation. We consider the cumulative cross-section for the ratio $m^2/(p_t R)^2$ to be smaller than some value ρ , integrated over the $\mathcal{O}(\alpha_s^2)$ matrix element for the Born-level production of 2 jets, in a given p_t bin:

$$\Sigma_{\text{NLL}}(\rho; p_{t1}, p_{t2}) = \int_{p_{t1}}^{p_{t2}} dp_t \sum_{i} \frac{d\sigma_{\text{jet,LO}}^{(i)}}{dp_t} \frac{e^{-R_i(\rho) - \gamma_E R_i'(\rho)}}{\Gamma(1 + R_i'(\rho))}, (2)$$

where we have separated contributions from different flavour channels, R'_i is the derivative of R_i wrt $\log(1/\rho)$ and the radiator R_i is given by

$$R_{i}(\rho) = \frac{C_{i}}{2\pi\alpha_{s}\beta_{0}^{2}} \left\{ \left[W(1-\lambda_{B}) - \frac{W(1-\lambda_{c})}{1+\beta} - 2W(1-\lambda_{1}) + \frac{2+\beta}{1+\beta} W(1-\lambda_{2}) \right] - \frac{\alpha_{s}K}{2\pi} \left[\log(1-\lambda_{B}) - \frac{\log(1-\lambda_{c})}{1+\beta} + \frac{2+\beta}{1+\beta} \log(1-\lambda_{2}) - 2\log(1-\lambda_{1}) \right] + \frac{\alpha_{s}\beta_{1}}{\beta_{0}} \left[V(1-\lambda_{B}) - \frac{V(1-\lambda_{c})}{1+\beta} - 2V(1-\lambda_{1}) + \frac{2+\beta}{1+\beta} V(1-\lambda_{2}) \right] \right\},$$
(3)

where

$$\lambda_{c} = 2\alpha_{s}\beta_{0}\log(1/z_{\text{cut}}), \qquad \lambda_{\rho} = 2\alpha_{s}\beta_{0}\log(1/\rho), \qquad (4)$$

$$\lambda_{1} = \frac{\lambda_{\rho} + \lambda_{B}}{2}, \qquad \lambda_{2} = \frac{\lambda_{c} + (1+\beta)\lambda_{\rho}}{2+\beta}, \qquad (5)$$

and $\lambda_B = 2\alpha_S \beta_0 B_i$ appears due to hard-collinear splittings, and $W(x) = x \log(x)$, $V(x) = \frac{1}{2} \log^2(x) + \log(x)$.

Note that α_s is calculated using the exact two-loop running coupling, at the scale $p_t R$, and, in order to reach NLL accuracy, it is evaluated in the CMW scheme [10]. Furthermore, compared to the original results [2], the hard-collinear contributions have been expressed as corrections to double-logarithm arguments. In practice, this is equivalent to replacing $P_i(z) \rightarrow (2C_i/z)\Theta(z < e^{B_i})$. This introduces unwanted NNLL terms but has the advantage to give well-defined and positive resummed distributions which, in turn, makes the matching to fix order easier.

To avoid any potential issue related to the Landau pole, appearing in a region anyway dominated by hadronisation, we have frozen the coupling at a scale $\mu_{fr} = 1$ GeV. Corresponding expressions can be found e.g. in Ref. [11].

Matching to NLO. The Soft Drop mass distributions for the dijet processes can be calculated at fixed order at $\mathcal{O}(\alpha_s^4)$, i.e. up to NLO accuracy. This is available for example using the NLOJet++ [12] generator to simulate $2 \to 3$ events

at LO and NLO. Jets are then clustered with the anti- k_t algorithm [13] as implemented in FastJet-3.2.2 [14]. In what follows, we have used the CT14 PDF set [15].

NLO mass distributions need to be matched to our NLL resummed results. For this, the LO jet mass distribution needs to be separated in flavour channels, while the flavour separation of the NLO jet mass distribution is instead subleading. At $\mathcal{O}(\alpha_s^3)$ a jet has at most two constituent and the only case where the flavour is ambiguous is when a jet is made of two quarks (or a quark and an anti-quark of different flavours). We (arbitrarily) treat this as a quark jet, an approximation which is valid at our accuracy. To keep the required flavour information in NLOJet++, we have used the patch introduced in Ref. [16].

To avoid artefacts at large mass, the endpoint of the resummed calculation is matched to the endpoint of the perturbative distribution by replacing

$$\log\left(\frac{1}{\rho}\right) \to \log\left(\frac{1}{\rho} - \frac{1}{\rho_{\text{max},i}} + e^{-B_q}\right) \tag{6}$$

in the resummed results [17]. The endpoints of the LO and NLO distributions are found to be (see Appendix B of Ref. [5]) $\rho_{\text{max,LO}} \approx 0.279303$ and $\rho_{\text{max,NLO}} \approx 0.44974$, for R = 0.8.

Finally, the matching between NLL and NLO results in each p_t bin can be done using log-R matching given by [16]

$$\Sigma_{\text{NLL+NLO}}(\rho) = \left[\sum_{i} \Sigma_{\text{NLL}}^{(i)} \exp\left(\frac{\Sigma_{\text{LO}}^{(i)} - \Sigma_{\text{NLL,LO}}^{(i)}}{\sigma_{\text{jet,LO}}^{(i)}}\right)\right]$$
(7)

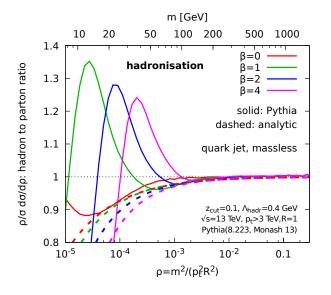
$$\times \exp \left(\frac{\bar{\Sigma}_{\text{NLO}} - \Sigma_{\text{NLL,NLO}}}{\sigma_{\text{jet,LO}}} - \sum_{i} \frac{(\Sigma_{\text{LO}}^{(i)})^2 - (\Sigma_{\text{NLL,LO}}^{(i)})^2}{\sigma_{\text{jet,LO}}^{(i)}\sigma_{\text{jet,LO}}} \right).$$

In this expression, $\Sigma_{\rm NLL}^{(i)}$ is given by Eq. (2), trivially split in flavour channels. $\Sigma_{\rm NLL,LO}^{(i)}$ and $\Sigma_{\rm NLL,NLO}$ (summed over flavour channels) are the expansion of $\Sigma_{\rm NLL}^{(i)}$ to LO $(\mathscr{O}(\alpha_s^3))$ and NLO $(\mathscr{O}(\alpha_s^4))$, respectively. For the fixed-order part

$$\Sigma_{\text{LO}}^{(i)} = -\int_{0}^{1} d\rho' \frac{d\sigma_{\text{mass,LO}}^{(i)}}{d\rho'} + \sigma_{\text{jet,NLO}}^{(i)}, \tag{8}$$

$$\bar{\Sigma}_{\rm NLO} = -\int_{\rho}^{1} d\rho' \frac{d\sigma_{\rm mass,NLO}}{d\rho'},\tag{9}$$

where $d\sigma_{\text{mass,(N)LO}}/d\rho$ denotes the mass distribution at (N)LO as obtained from NLOJet++ and $\sigma_{\text{jet,(N)LO}}$ the (N)LO correction to the inclusive jet cross-section in the p_t bin under consideration. These expressions also require the inclusive jet cross-section, both at LO and NLO, to be split in flavour channels. This is done as for the 3-jet LO distribution above using the flavour-aware NLOJet++ version used in [16]. Alternatively, we have also used the (R-)matching scheme given by Eq. (3.28) of [16].



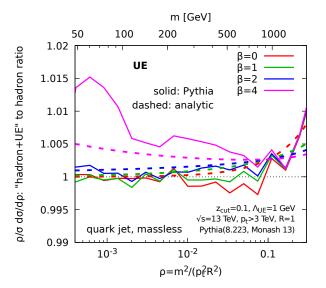


Fig. 1 Comparison of our theoretical estimate of hadronisation corrections to what is implemented in a standard Monte-Carlo parton shower, for different values of the angular exponent β , $z_{\text{cut}} = 0.1$ and R = 0.8. Quark jets are considered and hadron masses are neglected. Left: hadronisation corrections (i.e. ratio of hadron level to parton level) for $\Lambda_{\text{hadr}} = 0.4$ GeV; right: Underlying Event corrections (i.e. ratio of distributions with and without UE) for $\Lambda_{\text{UE}} = 1$ GeV.

From Eq. (7) it is trivial to obtain differential distributions in bins of ρ . Normalised distributions can then be obtained by dividing the result by the NLO inclusive jet cross-section $\sigma_{\rm jet,LO} + \sigma_{\rm jet,NLO}$.

The uncertainties on the distributions come from four sources: renormalisation and factorisation scales, resummation uncertainty and matching uncertainty. The first two are estimated using the 7-point rule [18]. The resummation uncertainties are obtained by varying ρ in Eqs. (2) and (3) between $\rho/2$ and ρ , introducing the appropriate correction — $\pm \log(2)R'$ in the exponent in (2) — to maintain NLL accuracy. The matching uncertainty is estimated by considering both the log-R and R matching schemes. We take the central value from the central scale choice and the uncertainty from the envelope of the 11 scale variations.²

Non-perturbative corrections. Power corrections induced by non-perturbative (NP) effects can be estimated for Soft Drop using the same approach as the equivalent calculation for mMDT presented in Section 8.3.3 of Ref. [1]. We have to take into account two effects: (i) the mass of the SD jet will be affected by NP corrections, (ii) NP effects can shift the momentum of the subjets and alter the SD condition.

First, the mass shift can be written as (see [19]) $\delta m^2 = C_R \Lambda_{\text{hadr}} p_t R_{\text{eff}}$, where R_{eff} is the effective jet radius after grooming, i.e. for a mass m and subjets passing the Soft Drop condition with a momentum fraction z, $R_{\text{eff}} = m/(p_t \sqrt{z(1-z)})$.

Following the same steps as in Ref. [1] we obtain³

$$\frac{d\sigma}{dm}\Big|_{\text{hadr}}^{(m \text{ shift})} = \frac{d\sigma}{dm}\Big|_{\text{pert}} \left(1 + \frac{C_R \Lambda_{\text{hadr}}}{m} \frac{z_{\text{SD}}^{-1/2} - \Delta_i}{L_{\text{SD}} + B_i}\right), \tag{10}$$

with
$$z_{\rm SD}=z_{\rm cut}^{\frac{2}{2+\beta}}\left(\frac{m}{p_tR}\right)^{\frac{2\beta}{2+\beta}},$$
 $L_{\rm SD}=\log(1/z_{\rm SD})$ and

$$\Delta_q = \frac{3\pi}{8} \quad \text{and} \quad \Delta_q = \frac{(15C_A - 6n_f T_R)\pi}{32C_A}.$$
(11)

Then, hadronisation will shift the momentum of the softer subjet by an average $\delta p_t = -C_A \Lambda_{\rm hadr}/R_{\rm eff}$, where we have taken into account that the softer subjet typically corresponds to a gluon emission. This means that emissions which were perturbatively passing the Soft Drop condition, with $z_{\rm SD} < z < z_{\rm SD} - \delta p_t/p_t$, will fail the Soft Drop condition after hadronisation, leading to a reduction of the cross-section

$$\frac{d\sigma}{dm}\Big|_{\text{hadr}}^{(p_t \text{ shift})} = \frac{d\sigma}{dm}\Big|_{\text{pert}} \left(1 - \frac{C_A \Lambda_{\text{hadr}}}{m} \frac{z_{\text{SD}}^{-1/2}}{L_{\text{SD}} + B_i}\right).$$
(12)

The final hadronisation correction includes both (10) and (12). Both terms are proportional to $\frac{\Lambda_{\text{hadr}}}{p_t} \left(\frac{p_t}{m}\right)^{\frac{2+2\beta}{2+\beta}}$, which increases with β and has the appropriate limits for $\beta \to \infty$ and $\beta \to 0$.

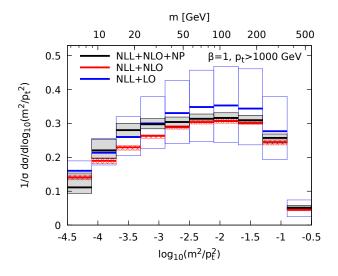
A similar calculation can be carried out for the Underlying Event (UE) contamination. In this case we have $\delta p_t = \Lambda_{\rm UE} \pi R_{\rm eff}^2$ and $\delta m^2 = \frac{1}{2} \Lambda_{\rm UE} p_t R_{\rm eff}^4$. Following the same steps as above, we find

$$\frac{d\sigma}{dm}\Big|_{\text{IJF}} = \frac{d\sigma}{dm}\Big|_{\text{pert}} \left(1 + \frac{\Lambda_{\text{UE}}m^2}{p_t^3 R^3} \frac{z_{\text{SD}}^{-2}(1 - f_{m,i})}{L_{\text{SD}} + B_i}\right), \tag{13}$$

¹Note that this normalisation procedure gives consistent results when computing the uncertainties on the matched distributions.

²Seven factorisation and renormalisation scales, two resummation scales and two matching schemes.

³Although, instead of averaging $R_{\rm eff}$ over z, we have kept explicit the z dependence of $R_{\rm eff}$ and averaged the final correction over z.



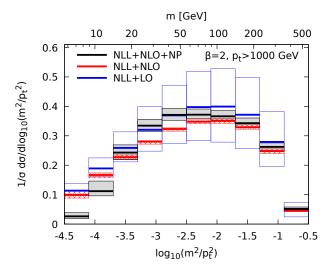


Fig. 2 Resummed and match theoretical predictions for the Soft Drop jet mass distribution, for two different values of the angular exponent $\beta = 1$ (left) and $\beta = 2$ (right), $z_{\text{cut}} = 0.1$ and R = 0.8. The colours correspond to different accuracy of the calculation, as detailed in the legend.

where the 1 in the numerator corresponds to the p_t shift and the $f_{m,i}$ term corresponds to mass-shift effects, with

$$f_{m,q} = \frac{1 + 3z_{\text{SD}} + 2z_{\text{SD}}^2(3L_{\text{SD}} - 2)}{4},$$

$$f_{m,g} = \frac{1 + 2z_{\text{SD}} + 3z_{\text{SD}}^2(2L_{\text{SD}} - 1)}{4} + \frac{n_f T_R}{C_A} z_{\text{SD}} (1 - z_{\text{SD}}).$$
(14)

This time, both sources of corrections give an effect proportional to $\frac{\Lambda_{\rm UE}}{p_t} \left(\frac{p_t}{m}\right)^{\frac{2\beta-4}{2+\beta}}$, which increase with β and has the expected $\Lambda_{\rm UE} p_t/m^2$ behaviour in the limit $\beta \to \infty$.

In Fig. 1, we compare our analytic findings (dashed lines) with the Monte-Carlo simulations, obtained with Pythia 8.223 [20] (Monash 13 [21] tune, solid lines). We consider both hadronisation corrections (left) and UE effects (right), for a range of β values. UE effects are seen to be much smaller than hadronisation corrections. In the region where $\Lambda_{\rm hadr,UE} \ll m \ll p_t$, our analytic calculation captures the main features observed in the simulations, including the increase with β and the global trend in ρ . At smaller mass, Pythia simulations exhibit a peak in the hadronisation corrections which is beyond the scope of our power-correction calculation.

Even if the above analytic approach to estimating NP effects is helpful for a qualitative understanding, it is unclear how reliable it would be for phenomenology. For example, hadron masses, which are neglected here, would have an additional effect, even at large mass. Thus, the analytic estimates can, at best, be trusted up to a fudge factor and analytic results can not be trusted at small mass (see also [22]).

As for our mMDT calculation [5], for our final predictions, we have therefore decided to estimate NP corrections using different Monte-Carlo simulations: Herwig 6.521 [23] with the tune AUET2 [24], Pythia 6.428 [25] with the Z2 [26] and Perugia 2011 [27] tunes, and Pythia 8.223 [20] with

the 4C [28] and Monash 13 [21] tunes. For each Monte-Carlo, we compute the ratio between the full simulation and the parton level. The average result is taken as the average NP correction, and the envelope as the uncertainty which is added in quadrature to the perturbative uncertainty.

Final predictions and conclusions. Our final predictions, are presented for $\beta=1$ (left) and $\beta=2$ (right) in Fig 2. To highlight our key observations, we present our final results at NLL matched to NLO and including NP corrections (labelled NLL+NLO+NP), as well as pure perturbative results (NLL+NLO) and results obtained when matching to LO only (NLL+LO). The most striking feature that we observe is that matching to NLO not only affects quite significantly the central value of our prediction, but also significantly reduces the uncertainty across the entire spectrum.

Then, we see that NP corrections remain small over a large part of the spectrum, although they start being sizeable at larger mass when the angular exponent β increases. The fact that Soft Drop observables can be computed precisely in perturbative QCD, with small NP corrections, makes them interesting for further phenomenological studies, including other observables like angularities or attempts to extract the strong coupling constant from fits to the data.

Finally, we note that these predictions have recently been compared to experimental results obtained by the ATLAS collaboration in Ref. [7]. A good overall agreement between data and theory is observed, especially in the region where NP corrections are small.

Acknowledgements We thank Ben Nachman and Gavin Salam for many useful discussions. SM and LS would like to thank the IPhT Saclay for hospitality during the course of this project. GS's work is supported in part by the French Agence Nationale de la Recherche, under grant ANR-15-CE31-0016 and by the ERC Advanced Grant Higgs@LHC (No. 321133).

References

- 1. M. Dasgupta, A. Fregoso, S. Marzani and G. P. Salam, JHEP **1309** (2013) 029 [arXiv:1307.0007].
- 2. A. J. Larkoski, S. Marzani, G. Soyez and J. Thaler, JHEP **1405** (2014) 146 [arXiv:1402.2657].
- 3. C. Frye, A. J. Larkoski, M. D. Schwartz and K. Yan, arXiv:1603.06375.
- 4. C. Frye, A. J. Larkoski, M. D. Schwartz and K. Yan, JHEP **1607** (2016) 064 [arXiv:1603.09338].
- 5. S. Marzani, L. Schunk and G. Soyez, JHEP **1707** (2017) 132 [arXiv:1704.02210].
- CMS Collaboration [CMS Collaboration], CMS-PAS-SMP-16-010.
- 7. M. Aaboud *et al.* [ATLAS Collaboration], arXiv:1711.08341.
- 8. J. M. Butterworth, A. R. Davison, M. Rubin and G. P. Salam, Phys. Rev. Lett. **100** (2008) 242001 [arXiv:0802.2470].
- Y. L. Dokshitzer, G. D. Leder, S. Moretti and B. R. Webber, JHEP 9708 (1997) 001 [hep-ph/9707323];
 M. Wobisch and T. Wengler, In *Hamburg 1998/1999, Monte Carlo generators for HERA physics* 270-279 [hep-ph/9907280].
- S. Catani, B. R. Webber and G. Marchesini, Nucl. Phys. B 349 (1991) 635.
- 11. M. Dasgupta, L. Schunk and G. Soyez, JHEP **1604** (2016) 166 [arXiv:1512.00516].
- S. Catani and M. H. Seymour, Nucl. Phys. B 485 (1997) 291 Erratum: [Nucl. Phys. B 510 (1998) 503] [hep-ph/9605323]; Z. Nagy, Phys. Rev. D 68 (2003) 094002 [hep-ph/0307268].
- M. Cacciari, G. P. Salam and G. Soyez, JHEP 0804 (2008) 063 [arXiv:0802.1189].

- M. Cacciari and G. P. Salam, Phys. Lett. B 641 (2006)
 [hep-ph/0512210]; M. Cacciari, G. P. Salam and G. Soyez, Eur. Phys. J. C 72 (2012) 1896 [arXiv:1111.6097].
- S. Dulat *et al.*, Phys. Rev. D **93** (2016) no.3, 033006 [arXiv:1506.07443].
- A. Banfi, G. P. Salam and G. Zanderighi, JHEP 1006 (2010) 038 [arXiv:1001.4082].
- 17. S. Catani, L. Trentadue, G. Turnock and B. R. Webber, Nucl. Phys. B **407** (1993) 3.
- M. Cacciari, S. Frixione, M. L. Mangano, P. Nason and G. Ridolfi, JHEP 0404 (2004) 068 [hep-ph/0303085].
- 19. M. Dasgupta, L. Magnea and G. P. Salam, JHEP **0802** (2008) 055 [arXiv:0712.3014].
- 20. T. Sjostrand, S. Mrenna and P. Z. Skands, Comput. Phys. Commun. **178** (2008) 852 [arXiv:0710.3820].
- 21. P. Skands, S. Carrazza and J. Rojo, Eur. Phys. J. C **74** (2014) no.8, 3024 [arXiv:1404.5630].
- 22. M. Dasgupta, F. A. Dreyer, G. P. Salam and G. Soyez, JHEP **1606** (2016) 057 [arXiv:1602.01110].
- G. Corcella, I. G. Knowles, G. Marchesini, S. Moretti,
 K. Odagiri, P. Richardson, M. H. Seymour and
 B. R. Webber, hep-ph/0210213.
- 24. [ATLAS Collaboration], ATL-PHYS-PUB-2011-008, ATL-COM-PHYS-2011-329.
- 25. T. Sjostrand, S. Mrenna and P. Z. Skands, JHEP **0605** (2006) 026 [hep-ph/0603175].
- 26. R. Field, arXiv:1010.3558.
- P. Z. Skands, Phys. Rev. D 82 (2010) 074018
 [arXiv:1005.3457]; B. Cooper, J. Katzy,
 M. L. Mangano, A. Messina, L. Mijovic and P. Skands,
 Eur. Phys. J. C 72 (2012) 2078 [arXiv:1109.5295].
- 28. R. Corke and T. Sjostrand, JHEP **1103** (2011) 032 [arXiv:1011.1759].