Treating jet correlations in high pile-up at hadron colliders

F. Hautmann,[1,2] H. Jung,[3,4] and H. Van Haevermaet[4]

[1]*Rutherford Appleton Laboratory & University of Southampton*
[2]*Theoretical Physics, University of Oxford*
[3]*Deutsches Elektronen Synchrotron*
[4]*Elementaire Deeltjes Fysica, Universiteit Antwerpen*

Abstract

Experiments in the high-luminosity runs at the Large Hadron Collider face the challenges of very large pile-up. Primary techniques to deal with this are based on precise vertex and track reconstruction. Outside tracker acceptances, however, lie regions of interest for many aspects of the LHC physics program. We explore complementary approaches to pile-up treatment and propose a data-driven jet-mixing method which can be used outside tracker acceptances without depending on Monte Carlo generators. The method can be applied to treat correlation observables and take into account, besides the jet transverse momentum pedestal, effects of hard jets from pile-up.

1. Introduction. Experiments at hadron colliders operating with very high luminosity face the challenge of pile-up, namely, a very large number of overlaid hadron-hadron collisions per bunch crossing. At the Large Hadron Collider (LHC), in Run I data the pile-up is about 20 pp collisions on average, while it reaches the level of over 50 at Run II, and increases for higher-luminosity runs [1–11]. In regions covered by tracking detectors, advanced vertexing techniques have been developed to deal with environments characterized by high pile-up. More generally, experiments rely on Monte Carlo simulations which include pile-up for comparisons with data. However, this introduces a significant model dependence, especially in regions where no detailed and precise measurements are available to constrain Monte Carlo modeling.

In this paper we propose a different approach to treating high pile-up, with a view to employing data-driven methods rather than Monte Carlo methods. Our main focus is to deal with potentially large probabilities that jets with high transverse momenta are produced from pile-up events independent of the primary interaction vertex, in a region where tracking devices are not available to identify pile-up jets. A typical application would be Higgs production by vector boson fusion, where the associated jets may be produced outside the tracking detector acceptances. The issue we address is thus quite different from the issues that most of the existing methods for pile-up treatment are designed to deal with, which are the jet transverse momentum pedestal, due to the bias in the jet transverse momentum from added pile-up particles in the jet cone, and the clustering into jets of overlapping soft particles from pile-up.

In what follows we will therefore use standard existing methods to treat soft particles and the jet pedestal, and devise new approaches to tackle the issue of misidentification which arises, in addition, in cases where precise tracking and vertexing are not feasible. The aim is to look for methods which treat pile-up without spoiling the physics of the signal process and which can be used outside the tracking detector acceptances without depending on Monte Carlo modeling. To this end, we suggest using minimum bias (or jet) samples recorded from data in high pile-up runs and applying event-mixing techniques to relate, via these data samples, the "true" signal to the signal measured in high pile-up.

The approach does not address the question of a full detector simulation including pile-up. Rather, it focuses on how to extract physics signals with the least dependence on pile-up simulation, and how to use real data, rather than Monte Carlo events, at physics object level.

The proposed method applies to the regime of high pile-up which is relevant for the LHC as well as for future high-luminosity colliders. It is designed to treat not only inclusive variables but also correlations. One of the features of the method is that it does not require data-taking in dedicated runs at low pile-up. Rather, the data required for event mixing are recorded at the same time as the signal events in high pile-up runs, so that there is no loss in luminosity.

We will illustrate the approach using Drell-Yan lepton pair production associated with jets as a case study. We discuss two main physical consequences of pile-up collisions, the bias in the jet transverse momentum due to pile-up particles in the jet cone, and the misidentification of high transverse momentum jets from independent pile-up events. The method is general and can straightforwardly be extended to a large variety of processes affected by pile-up.

2. Drell-Yan plus jets at high pile-up as a case study. Let us consider the associated production of a Drell-Yan lepton pair via Z-boson exchange and a jet. We take the jet

transverse momentum and rapidity to be $p_T^{(\text{jet})} > 30$ GeV, $|\eta^{(\text{jet})}| < 4.5$, and the boson invariant mass and rapidity to be 60 GeV $< m^{(\text{boson})} < 120$ GeV, $|\eta^{(\text{boson})}| < 2$. Event samples are generated by PYTHIA 8 [12] with the 4C tune [13] for the different scenarios of zero pile-up and N_{PU} additional pp collisions at $\sqrt{s} = 13$ TeV. We reconstruct jets with the anti-k_T algorithm [14] with distance parameter $R = 0.5$. Results for the spectrum in the transverse momentum p_T of the Z-boson, for $Z + $ jet events, are shown in Fig. 1 for $N_{\text{PU}} = 0$, $N_{\text{PU}} = 20$ and $N_{\text{PU}} = 50$. For comparison we also show the inclusive $N_{\text{PU}} = 0$ Z-boson spectrum.

We see from Fig. 1 that the effects of pile-up on Z-boson $+$ jet scenarios are large. As a result of pile-up the shape of the p_T spectrum is changed and the peak is shifted to lower values. This can be interpreted by noting that, as the $Z + $ jet event sample becomes dominated by pile-up collisions, even with the jet transverse momentum selection cut $p_T^{(\text{jet})} > 30$ GeV the Z-boson p_T distribution in boson $+$ jet events will tend to approach the inclusive Drell-Yan spectrum, given by the solid green curve.
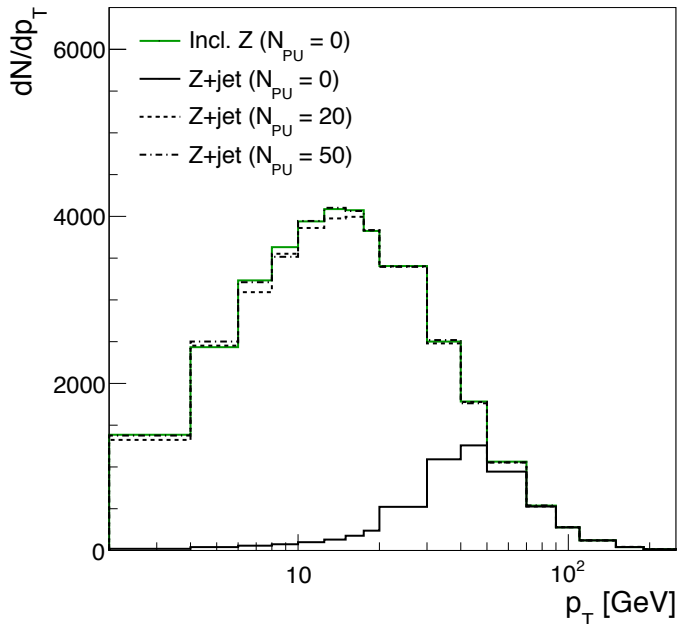

FIG. 1. *Effect of pile-up on the Z-boson transverse momentum p_T in Z-boson + jet production at the LHC.*

More precisely, we can identify two main implications of pile-up collisions: a large bias in the jet transverse momentum due to added pile-up particles in the jet cone leading to a jet pedestal, and a large probability that jets with high transverse momentum come from independent pile-up events.

Several methods exist to deal with the jet p_T pedestal. These include techniques based on the jet vertex fraction [3] and charged hadron subtraction [5, 15], the PUPPI method [16], the SoftKiller method [17]. These methods correct for transverse momenta of individual particles, but not for any mistagging. So do approaches inspired by jet substructure studies, such as jet cleansing [18]. In Fig. 2 we apply SoftKiller [17], a new event-wide particle-level
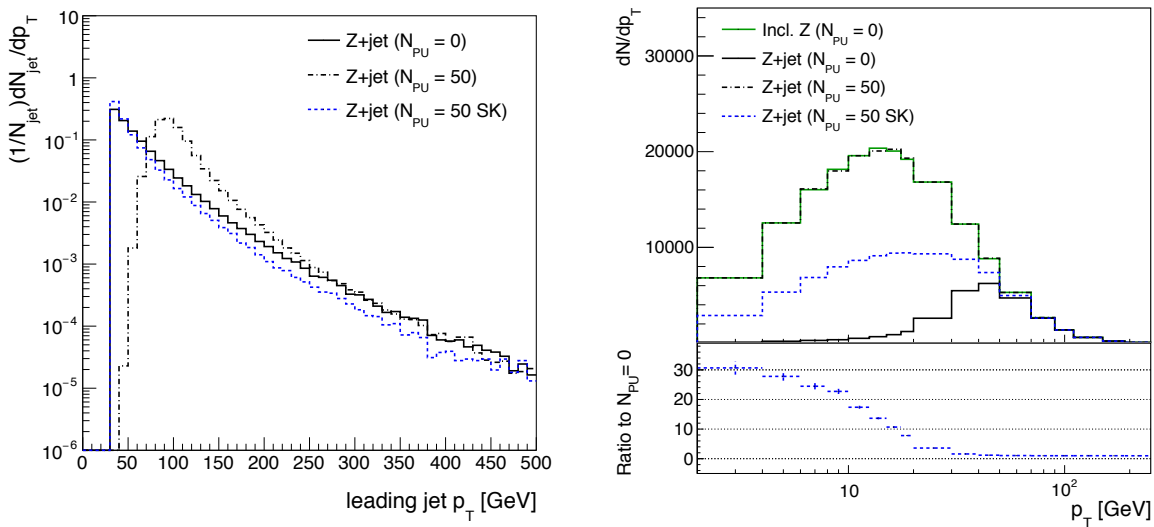
FIG. 2. *Application of SoftKiller [17] to Z-boson + jet production. Left: (a) the leading jet p_T spectrum; right: (b) the Z-boson p_T spectrum.*

pile-up removal method, which can also be used with calorimeter information only. We present results at zero pile-up ($N_{PU} = 0$), at pile-up $N_{PU} = 50$, and the result at pile-up $N_{PU} = 50$ with SoftKiller subtraction ($N_{PU} = 50$ SK).

Fig. 2 illustrates different physical effects of pile-up in the leading jet spectrum and in the Z-boson spectrum. In Fig. 2a we compute the leading jet p_T spectrum, and verify that SoftKiller efficiently removes the jet pedestal from pile-up: the zero pile-up jet spectrum (solid black curve) is shifted toward larger p_T by pile-up collisions (dot-dashed black curve for $N_{PU} = 50$) but the application of SoftKiller (dashed blue curve $N_{PU} = 50$ SK) corrects for this and restores the original signal with very good approximation. In Fig. 2b, on the other hand, we compute the Z-boson p_T spectrum. The solid black curve is the zero pile-up result, the dot-dashed black curve is the $N_{PU} = 50$ result, and the dashed blue curve is the result of applying SoftKiller. In the higher p_T part of the spectrum we observe that there is no need for any correction. In contrast, in the lower p_T part significant contributions are present from misidentified pile-up jets. These are not corrected for, and need to be properly treated, particularly in regions outside tracker acceptances where vertexing techniques cannot be relied on to identify pile-up jets [6]. We address this point next.

3. Uncorrelated event samples and jet mixing. To treat effects beyond soft particles and the jet p_T pedestal, we employ event mixing techniques [19–22] using uncorrelated samples. The main idea is that the signal in the pile-up scenario is obtained via mixing from the signal without pile-up and a minimum bias sample of data at high pile-up. Thus, to identify the contribution of the high p_T jets coming from independent pile-up events, we construct a signal plus pile-up scenario in a data-driven manner. We do this by adding physics objects from pileup background to event samples before selection criteria are applied. The approach is designed to treat the region of high number N_{PU} of pile-up events, where $(N_{PU} + 1)/N_{PU} \approx 1$.

We illustrate the method by taking a sample containing N_{PU} minimum bias events (which could be recorded data but we just take for illustration as Monte Carlo events), mixing this
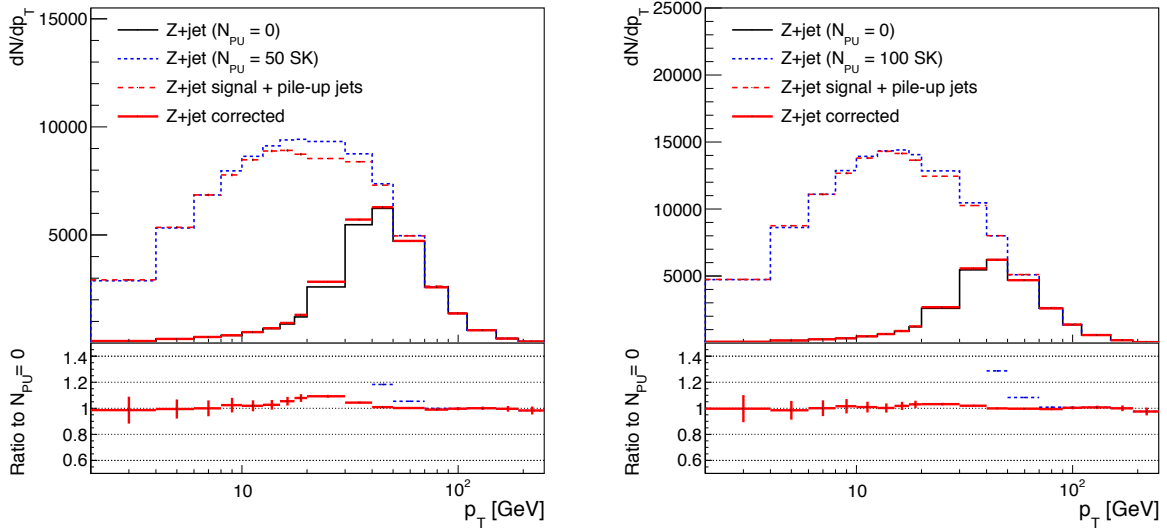
4

FIG. 3. *The Z-boson p_T spectrum in $Z + jet$ production from the jet mixing method. Left: (a) $N_{\mathrm{PU}} = 50$; right: (b) $N_{\mathrm{PU}} = 100$.*

with the signal at zero pile-up, and then requiring a jet with $p_T^{(\mathrm{jet})} > 30$ GeV, $|\eta^{(\mathrm{jet})}| < 4.5$. We extract the unbiased signal without relying on Monte Carlo algorithms.

Fig. 3 reports the result of carrying out this procedure, for $N_{\mathrm{PU}} = 50$ and $N_{\mathrm{PU}} = 100$. Here the solid black curve is the "true" Z-boson plus jet signal. The dashed blue curve is
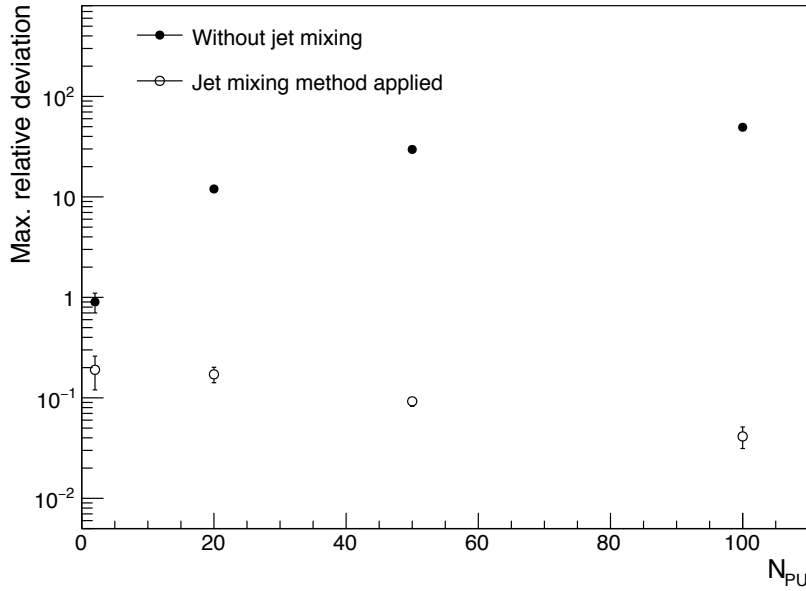

FIG. 4. *Maximal value of the relative deviation between the pile-up corrected signal and the true signal, with and without jet mixing, as a function of the number of pile-up collisions. Black dots: SoftKiller corrected result; open circles: jet mixing method applied.*

5

the high pile-up, SoftKiller-corrected result ($N_{\mathrm{PU}} = 50$ SK and $N_{\mathrm{PU}} = 100$ SK). As seen already in Fig. 2b, this is far from the solid black curve in the lower-p_T part of the spectrum. We regard the dashed blue curve as pseudodata in high pile-up. The long-dashed red curve is the jet mixing curve, obtained as described above by mixing the signal with the minimum bias sample. The result of the mixing method is then given as the solid red curve by a simple "unfolding", defined by multiplying the signal by the ratio of the pile-up (dashed blue) curve to the mixing (long-dashed red) curve. We see that without appealing to any Monte Carlo method the true signal is extracted nearly perfectly from the mixed sample.

In addition to the closure test carried out above, we have checked the model dependence by applying the mixing procedure to different starting distributions, and verified that in this case as well the unfolding returns the true signal.

In Fig. 4 we plot the maximal value of the relative deviation between the pile-up corrected signal and the true signal ((corrected - true)/true), for the SoftKiller case without jet mixing (black dots) and for the case with the jet mixing method applied (open circles), as a function of the number of pile-up collisions N_{PU}. We see that, while in the SoftKiller case, in which the jet pedestal is removed, the deviation from the true signal becomes larger as N_{PU} increases, the deviation does not increase with N_{PU} once the jet mixing method is applied to take account of the hard jets from pile-up.
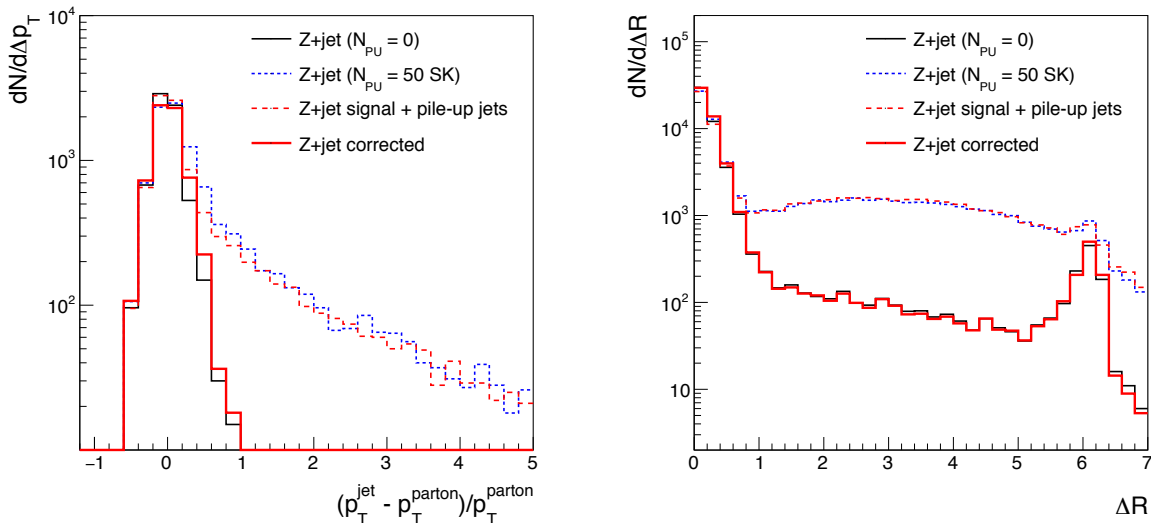

FIG. 5. *Effects on jet resolution. Left: (a) the parton-jet p_T correlation; right: (b) the ΔR distribution.*

The main advantages of this approach are that it can be used with data recorded in high pile-up, and it does not depend on Monte Carlo algorithms for pile-up correction. In addition, it is interesting to perform control checks by examining results for the jet resolution which we obtain from the jet mixing method. These are shown in Fig. 5. Fig. 5a reports the parton-jet p_T correlation, and Fig. 5b the distribution in $\Delta R = \sqrt{\Delta\phi^2 + \Delta\eta^2}$, where $\Delta\phi$ and $\Delta\eta$ are respectively the separation in azimuth and rapidity. We see that the features of the "true" signal are well reproduced.

4. Conclusions. Current methods to deal with pile-up at the LHC employ precise vertex and track reconstruction, in regions where these are available, and in general rely on Monte

Carlo simulations to model pile-up for data comparisons. The use of Monte Carlo event generators brings in a significant model dependence particularly in regions where these are not well constrained by measurements.

In this paper we have discussed a different, data-driven approach to the treatment of pile-up, which makes use of minimum bias or jet samples recorded from data taken in high pile-up runs, constructs mixing methods to extract the signal process, and thus circumvents the model dependence implied by the use of Monte Carlo generators.

The methodology is general, and can be applied in measurements to restore correlations between final-state particles. In such measurements two important kinds of pile-up effects are present, exemplified in the case of Z-boson plus jets which we have used for illustration, the jet p_T pedestal and the misidentification of high-p_T jets from independent pile-up events. While several methods exist to correct for the first effect (as well as for the related effect of the clustering into jets of overlapping soft particles), the second effect is not treated at present. We have proposed a jet-mixing method which treats this, and we have shown that it allows one to successfully extract the signal process from the mixed sample to within few percent.

The methods discussed in this paper can be applied to the high pile-up regime and do not require special runs at low pile-up. The data samples needed for jet mixing are recorded at the same time as the signal events. There is therefore no loss in luminosity. The advantages are that one can access the proper pile-up distribution and there is no need for pile-up reweighting.

The use of these methods thus implies good prospects both for precision Standard Model studies at moderate scales affected by pile-up, e.g. in Drell-Yan and Higgs production [23–25], and for searches for rare processes beyond Standard Model in high pile-up regimes.

[1] ATLAS Collaboration, ATLAS-CONF-2014-018.
[2] ATLAS Collaboration, ATLAS-CONF-2014-019.
[3] ATLAS Collaboration, ATLAS-CONF-2013-083.
[4] Z. Marshall [ATLAS Collaboration], J. Phys. Conf. Ser. **513** (2014) 022024.
[5] CMS Collaboration, CMS-PAS-JME-14-001.
[6] CMS Collaboration, CMS-PAS-JME-13-005.
[7] S. S. Ghosh [CMS Collaboration], arXiv:1502.05207 [hep-ex].
[8] J. Anderson *et al.*, Snowmass Energy Frontier Simulations White Paper (2013).
[9] A. Haas [ATLAS Coll.], talk at LHC Detector Simulations Workshop, CERN, March 2014.
[10] M. Hildreth [CMS Coll.], talk at LHC Detector Simulations Workshop, CERN, March 2014.
[11] S. Fartoukh, Phys. Rev. ST Accel. Beams **17** (2014) 111001.
[12] T. Sjöstrand, S. Mrenna and P. Z. Skands, Comput. Phys. Commun. **178** (2008) 852.
[13] R. Corke and T. Sjöstrand, JHEP **1103** (2011) 032.

[14] M. Cacciari, G. P. Salam and G. Soyez, JHEP **0804** (2008) 063.

[15] H. Kirschenmann [CMS Collaboration], PoS EPS-HEP2013 (2013) 433.

[16] D. Bertolini, P. Harris, M. Low and N. Tran, JHEP **1410** (2014) 59.

[17] M. Cacciari, G. P. Salam and G. Soyez, Eur. Phys. J. C **75** (2015) 59.

[18] D. Krohn, M. D. Schwartz, M. Low and L. T. Wang, Phys. Rev. D **90** (2014) 065020.

[19] D. Drijard, H. G. Fischer and T. Nakada, Nucl. Instrum. Meth. A **225** (1984) 367.

[20] S. Schael *et al.* [ALEPH Collaboration], Phys. Lett. B **606** (2005) 265.

[21] H. Schettler, DESY-THESIS-2013-036, CERN-THESIS-2013-265.

[22] B. Dutta, T. Kamon, N. Kolev and A. Krislock, Phys. Lett. B **703** (2011) 475.

[23] P. Cipriano *et al.*, Phys. Rev. D **88** (2013) 097501.

[24] H. Van Haevermaet *et al.*, PoS DIS **2014** (2014) 163.

[25] U. Langenegger, M. Spira and I. Strebel, arXiv:1507.01373 [hep-ph].