

Data-driven approaches to pile-up treatment at the LHC

Hans Van Haevermaet*

University of Antwerp (BE)

E-mail: hans.van.haevermaet@cern.ch

Experiments in the upcoming high-luminosity runs at the LHC face the challenges of very large pile-up. Primary techniques to deal with this are based on precise vertex reconstruction by tracking detectors. Outside tracker acceptances, however, lie regions of much interest for many aspects of the LHC physics program, and treatments of pile-up in these regions rely more strongly on Monte Carlo simulations. But here, one is also approaching parts of the phase space where the tuning of Monte Carlo event generators becomes subject to increasingly large uncertainties. In this study we explore a new data-driven approach that treats pile-up and does not spoil the physics of the signal process. In particular we are interested in restoring final state jet correlations, and present this with numerical illustrations for the Drell-Yan + jet production process.

*The European Physical Society Conference on High Energy Physics
22-29 July 2015
Vienna, Austria*

*Speaker.

1. Introduction

Experiments at hadron colliders operating at a very high luminosity face the challenge of pile-up, i.e. a large number of additional hadron-hadron collisions per bunch crossing. This results in a high probability that the final state of a particular process drowns in pile-up activity. At the Large Hadron Collider (LHC), pile-up consists of about 20 pp collisions on average in Run I data, it increases to a level of over 50 at the current Run II data, and even more for future higher-luminosity runs [1–11]. Advanced vertexing techniques have been developed to deal with high pile-up environments in regions covered by tracking detectors, however, more generally, experiments rely on Monte Carlo simulations that include pile-up modelling to compare with data. But this introduces a significant model dependence, especially in more forward regions where currently less precise measurements are available to constrain Monte Carlo models.

In this study we propose a different, purely data-driven approach to treat high pile-up, rather than to rely on Monte Carlo methods. Our main focus is to deal with the presence of high transverse momentum (p_T) jets produced from pile-up events independent of the primary interaction vertex, in a region where tracking detectors are not available to identify them. A typical application would be Higgs production through vector boson fusion, where associated jets may be produced outside the tracking detector acceptances, e.g. $|\eta| > 3$. The issue we try to address here is thus quite different from those for which most of the existing pile-up correction methods are designed. These are the jet transverse momentum pedestal, due to the bias in the jet p_T from added pile-up particles in the jet cone, and the clustering of overlapping soft particles from multiple collisions into jets.

In the following we will therefore use an existing method to correct for the contribution of soft pile-up particles in the event, and introduce a new approach to solve the issue of misidentified high p_T jets in regions where no track and vertex information is available. The primary aim is to look for a method that can restore correlations between final state particles or jets, and can be used outside tracking detector acceptances without relying on a Monte Carlo model. To accomplish this we suggest to use a minimum bias data sample recorded during high pile-up runs, and apply an event-mixing technique with this sample, to retrieve the true signal from the measured signal distribution in high pile-up. The proposed method applies to the regime of high pile-up, which is relevant for the LHC and future high-luminosity colliders. One of the features of the method is that it does not require any dedicated data-taking in low pile-up runs, since the data required for event mixing can be 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 our new approach using Drell-Yan lepton pair production associated with jets as a case study, and discuss the main physical consequences of additional pile-up collisions.

2. Drell-Yan plus associated jet production at high pile-up

Consider the associated production of a Drell-Yan lepton pair via Z -boson exchange and a jet, at a centre-of-mass energy of $\sqrt{s} = 13$ TeV. We require 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 Pythia8 [12] with the

4C tune [13], and jets are reconstructed with the anti- k_T algorithm [14] with distance parameter $R = 0.5$.

Fig. 1 then shows the Z-boson p_T spectrum for Z+jet events for $N_{PU} = 0, 20$ and 50 ; and for comparison the inclusive Z-boson spectrum for $N_{PU} = 0$. One sees that the inclusion of pile-up has a large effect on the Z+jet p_T spectrum: the statistics increases drastically, the shape of the distribution changes, and its peak is shifted towards lower values. This is because the Z+jet sample becomes completely dominated by pile-up, i.e. even with a $p_T^{(jet)} > 30$ GeV requirement basically all events are tagged as being Z+jet, so that the measured Z+jet spectrum approaches the inclusive Z-boson spectrum.

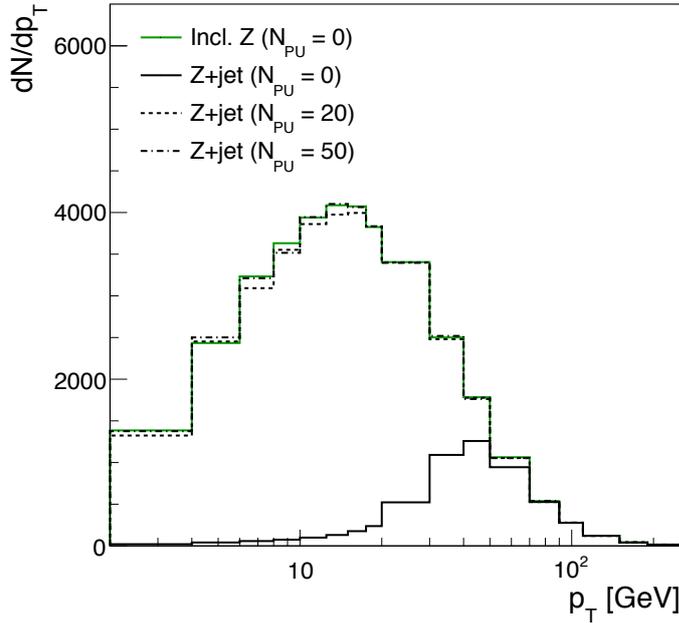


Figure 1: Effect of pile-up on the Z-boson p_T spectrum for Z+jet production at the LHC.

More precisely, one can identify two main pile-up contributions: (a) the presence of a large amount of soft particles in the event which leads to a bias (jet pedestal) in jet p_T due to added pile-up particles in the jet cone, and can make fake jets when these soft particles are clustered together; and (b) a large probability that genuine high p_T jets come from independent pile-up events. Several methods exist to correct for the first type (a) of contributions: these include techniques based on jet vertex fraction [3] or charged hadron subtraction [5, 15], the PUPPI method [16], and the SoftKiller method [17]. They can correct the p_T of individual final state particles very well, but are not able to treat correlated observables. The same applies to approaches that use jet substructure information, such as jet cleansing [18].

In this study we will apply SoftKiller [17], a new event-wide particle-level pile-up removal method that can also be used with calorimeter information only, to remove the contributions from soft particles in the event. Fig. 2 presents the results of the SoftKiller (SK) subtraction at $N_{PU} = 50$, and illustrates the effects of pile-up on the leading jet and Z-boson p_T spectra.

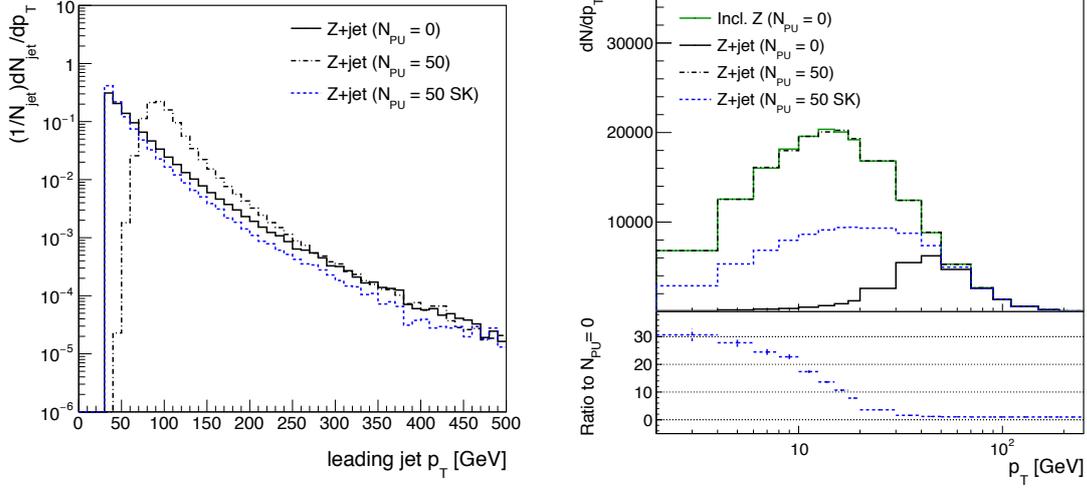


Figure 2: Application of SoftKiller [17] to Z+jet production. Left: (a) the leading jet p_T spectrum; right: (b) the Z-boson p_T spectrum.

Fig. 2a confirms that the leading jet p_T spectrum can be corrected with the SoftKiller method rather well, thus efficiently removing the jet p_T pedestal, while Fig. 2b, on the other hand, still shows a large discrepancy in the Z-boson p_T spectrum after the SoftKiller subtraction. In the higher p_T part of the spectrum there is no need for any correction, but at low p_T values significant contributions are still present from misidentified pile-up jets. These need to be properly treated, particularly in regions outside tracker acceptances where vertex information is not available [6].

3. Uncorrelated event samples and the jet-mixing method

To treat the remaining contributions, beyond soft particles and the jet p_T pedestal, we propose to apply jet mixing techniques [19–22] using uncorrelated samples. The main idea is that the signal in a high pile-up scenario is obtained by mixing the signal without pile-up with a minimum bias sample of data at high pile-up. To identify the contribution of high p_T jets coming from independent pile-up events, we thus construct a signal plus pile-up scenario in a data-driven way, where physics objects from pile-up background are added to the event before selection criteria are applied. The technique is designed to treat the region of high N_{PU} , where $(N_{\text{PU}} + 1)/N_{\text{PU}} \approx 1$.

We use the Drell-Yan + jets case study to illustrate the method by taking a sample (Monte Carlo in this exercise) containing N_{PU} minimum bias events after SoftKiller subtraction, mixing this with the signal at zero pile-up, and then perform the same event selection requiring a jet with $p_T^{(\text{jet})} > 30$ GeV and $|\eta^{(\text{jet})}| < 4.5$.

The results are presented in Fig. 3 for $N_{\text{PU}} = 50$ and $N_{\text{PU}} = 100$. The solid black curve is the “true” Z+jet signal, the dashed blue curve is the high pile-up SoftKiller-corrected result, and the solid red curve is the high pile-up SoftKiller + jet-mixing corrected result. We regard the dashed blue curve as pseudodata in high pile-up, and the long-dashed red curve is the jet-mixed curve, obtained as described above. The result of the mixing method (solid red curve) is then given by a

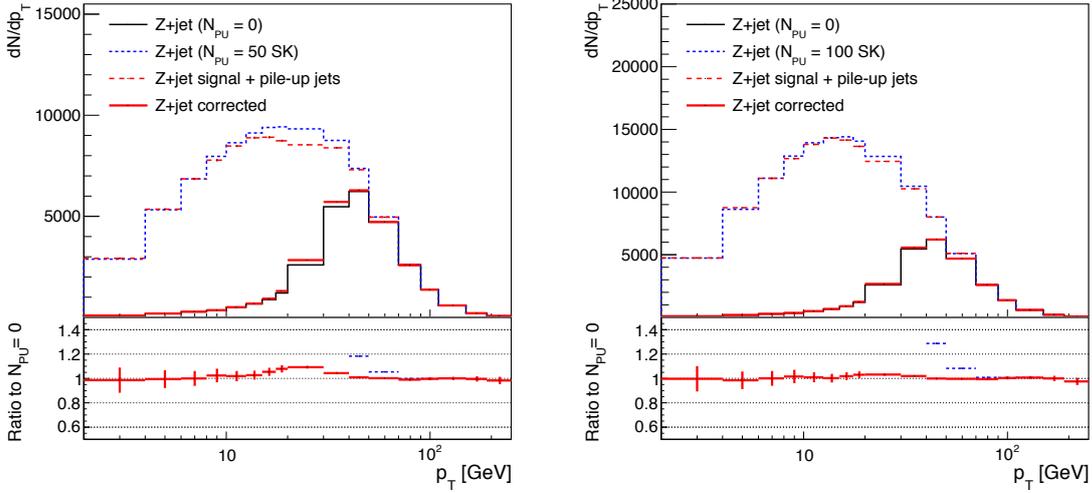


Figure 3: The Z-boson p_T spectrum in Z+jet production corrected with the jet-mixing method. Left: (a) $N_{PU} = 50$; right: (b) $N_{PU} = 100$.

simple “unfolding”, defined by multiplying the signal at zero pile-up by the ratio of the pseudodata (dashed blue) curve to the jet-mixed (long-dashed red) curve. Without using any Monte Carlo method, the true signal is extracted nearly perfectly from the jet-mixed sample.

In addition to the closure test shown above, we have checked the model dependence by applying the jet-mixing technique to different starting distributions, and verified that in this case as well the unfolding returns the true signal with good approximation. Furthermore, we checked the performance of the jet-mixing method as function of N_{PU} , and found that its accuracy even improves with increasing the number of pile-up interactions (which can be seen already comparing figures 3a and 3b). We also performed checks by looking at the behaviour of the jet resolution after the jet-mixing method is applied, and see that here too the features of the signal distributions at zero pile-up are well reproduced. The aforementioned results can be found in [23]. The main advantages of this jet-mixing method are that it can be used with data recorded in high pile-up, and it does not depend on Monte Carlo models for the pile-up correction.

4. Conclusions

Current methods developed to treat pile-up at the LHC use precise vertex and track information to remove all pile-up particles and correct the jet transverse momentum. They are however limited to the phase space where tracking detectors are available, outside this region one relies in general on Monte Carlo simulations to model pile-up effects. This leads to significant model dependences, especially in regions where these models are not well constrained by measurements.

In this study we proposed a new data-driven approach to treat pile-up, which is model independent, and is able to restore final state correlations even in regions where no track or vertex information is available. In particular it is designed to correct for the misidentification of high- p_T jets from independent pile-up events, an effect that is not treated in existing methods outside tracking detector regions.

The main ingredient is a minimum bias data sample recorded during high pile-up runs, which is used to construct a jet-mixing method. From this we have shown, using Z +jet processes, that one can extract successfully true signal distributions from the mixed sample, thus circumventing any model dependence implied by the use of Monte Carlo event generators.

This is a general technique, and can be used in any measurement to restore correlations between final-state particles. The method can be applied to the high pile-up regime and does not require special runs at low pile-up: the needed samples are recorded at the same time as the signal events, leading to no loss in luminosity. The use of this method thus implies good prospects both for precision Standard Model studies, e.g. in Drell-Yan and Higgs production [24–26], and for searches for beyond Standard Model physics in high pile-up regimes.

5. Acknowledgements

The results in these proceedings have been obtained in collaboration with F. Hautmann and H. Jung. Many thanks to the organizers and conveners of the EPS HEP 2015 conference for giving us the opportunity to present our studies.

References

- [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] F. Hautmann, H. Jung and H. Van Haevermaet, arXiv:1508.07811 [hep-ph].
- [24] P. Cipriano *et al.*, Phys. Rev. D **88** (2013) 097501.
- [25] H. Van Haevermaet *et al.*, PoS DIS **2014** (2014) 163.
- [26] U. Langenegger, M. Spira and I. Strebel, arXiv:1507.01373 [hep-ph].