IL NUOVO CIMENTO **45 C** (2022) 107
DOI 10.1393/ncc/i2022-22107-7

Communications: SIF Congress 2021

Simulating V+jets processes at ATLAS

G. Frattari
Department of Physics, Brandeis University - Waltham MA, USA

received 25 January 2022

Summary. — The production of a vector boson in association with hadronic jets is one of the most frequent background processes for precision measurements and new physics searches carried out in proton-proton collision at the LHC. One common issue for these analyses comes from the limited amount of events which is possible to produce with a fixed computing budget. Different methods have been investigated to optimise the statistical power of the simulated samples via phase-space biasing during the Monte Carlo event generation. Recent developments from the ATLAS Collaboration are presented.

1. – Introduction

The simulation of single electroweak vector-boson ($V = W, Z$) production in association with hadronic jets (V+jets) is an important ingredient of the ATLAS experiment [1] physics program. An accurate description of such processes is therefore crucial, especially in high jet-multiplicity final states. This can be achieved by multijet-merged Monte Carlo (MC) generators, where fixed order prediction for the production of a vector boson in association with multiple partons (quarks and gluons) up to NLO are combined with parton shower algorithms which take care of the generation of soft emissions needed to correctly describe the measured data. The reader is referred to ref. [2] for further details. The increased accuracy in the generation procedure leads to an increase in requirements of computational resources to produce the events. Furthermore, for an optimal description of the data, having at least the same number of MC events of the ones observed in data is mandatory.

In preparation of the future LHC [3] runs the computing and software framework of the ATLAS Collaboration has been reviewed, and a conceptual design report has been produced [4]. A significant amount of resources, both in terms of CPU consumption and disk usage is needed for the production of MC samples. A projection of the CPU requirements of the ATLAS Collaboration up to 2034 based on 2020 data shows how the current model is not sufficient to fit the future plans of the increased budget, and that substantial research and development are needed. Furthermore, about 20% of the CPU consumption is expected for the generation of MC events and about 41% of the CPU will be required for the simulation of particles interaction with the ATLAS detector. For these reasons, new methods for the generation of $V+$jets processes have been introduced in the Sherpa event generator [5]. The cross-section biasing method, of which an overview is presented in the next section, is set to replace the slicing approach as the default method to efficiently populate the phase-space of interest of the aforementioned measurements.

2. – Phase-space slicing and cross-section biasing

For the generation of $V+$jets samples with the Sherpa MC generator, a phase-space slicing procedure was adopted by the ATLAS Collaboration. This allows to generate in separate instances parts of the spectrum, and it is needed because the cross-section is steeply falling as a function of the transverse energy of the collision products. In fact, if events were generated according to their real probability the tail of the vector boson transverse momentum (p_{T}^{V}) distribution or of the sum of the transverse momenta of the hadronic jets produced (H_{T}) would be poorly populated. An alternative has been recently tested [6], and is referred to as *phase-space biasing.* The idea is to sample events according to a biased, or *enhanced*, value of the cross-section. This is achieved by multiplying the cross-section by the inverse of the biasing function $e(\mathcal{O},\Phi_m(\vec{x}))=\frac{\mathrm{d}\sigma}{\mathrm{d}\mathcal{O}_m}$, which depends on one observable, $\mathcal{O}$, and on the phase-space element $\Phi_m(\vec{x})$ of the $pp\to m+X$ process to be generated. To restore the correct cross-section distribution, each event weight is then multiplied by an additional weight $w_{i,m}^{\mathrm{bias}}=e(\mathcal{O},\Phi_m(\vec{x}))$. This method is called *differential enhancement* and is available via the *Enhance Observable* feature in the Sherpa event generator, and allows user defined observables to be adopted.

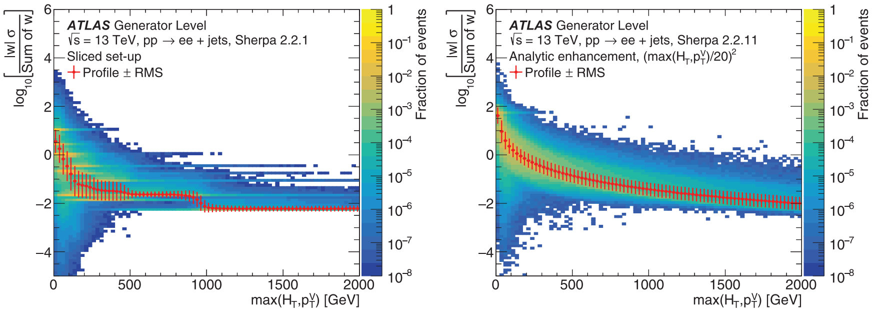
However, in the generation of V in association with multiple jets events the dependency of the biasing function on the phase-space element $\Phi_m(\vec{x})$ can lead to a significant spread of the distribution of the event weights, as each $V+\mathrm{x}$ partons process is biased with a different function. The relative MC statistical uncertainty (σ_{rel}) of the generated samples is given by

$$\sigma_{\mathrm{rel}}=\frac{\sqrt{\sum_i^{N_{\mathrm{raw}}}w_i^2}}{\sum_i^{N_{\mathrm{raw}}}w_i}=\frac{1}{\sqrt{N_{\mathrm{raw}}}}\sqrt{1+\frac{\sigma_{w_i}^2}{\langle w\rangle^2}}, \tag{1}$$

where N_{raw} is the number of generated events, $\langle w\rangle$ is the mean and σ_{w_i} is the standard deviation of the event weights. Hence, a large spread of the event weights distribution leads to a large statistical uncertainty for the sample. To recover this, a different method, namely the *analytic enhancement* was chosen for the simulation of $V+$ jets events, available via the *Enhance Function* feature of the Sherpa event generator. In this approach, the biasing function depends only on the user defined observable, and not on the phase-space element, like the parton multiplicity of the generated $V+\mathrm{x}$ final state.

For the generation of $V+$ jets processes it was chosen to enhance all the different $2\to X$ matrix-element-level processes (except for the $2\to 2$ one which has no bias applied) as a function of $\mathcal{O}=[\max(H_{\mathrm{T}},p_{\mathrm{T}}(V))/20\,\mathrm{GeV}]^2$.

The distribution of the logarithm of the event weights (scaled by the cross-section and normalised to the sum performed over all the events for a better visualisation) for the sample based on the phase-space slicing and for the one generated according to a biased cross-section are presented as a function of $\max(H_{\mathrm{T}},p_{\mathrm{T}})$ in fig. 1(a) and (b), respectively.



(a) (b)

Fig. 1. – Normalised distribution of the logarithm of the absolute value of the event weights, normalised to the sum performed over all the events and scaled by the cross-section, shown as a function of generator-level $\max(H_\mathrm{T}, p_\mathrm{T}(V))$ for (a) the sliced set-up and (b) for the one produced with differential enhancement of the cross-section. Figures from ref. [6].

The sliced set-up presents multiple populations which are organized in horizontal lines. Each of these represents a different slice which was generated independently. These different populations cause an increase in the RMS of the event weights (represented by the red bar in the plot), leading to a lower statistical power. When the cross-section biasing method is adopted instead, a smooth distribution of the event weights is observed, with a reduced RMS of the event weights up to ∼500 GeV.

As introduced in eq. (1), the statistical uncertainty of a set of weighted events depends on three terms: the number of raw events generated (N_raw), the mean of the event weights ($\langle w \rangle$) and their variance σ_w^2. A fourth component which impacts σ_rel is the fraction of events with a negative weight, f_nw. Negative weights are needed to implement in MC simulated samples predictions beyond the leading order, thus are unavoidable. The relative uncertainty of a MC sample with $f_\mathrm{nw} = 0.5$ diverges, therefore it is useful to monitor also such a quantity. Figure 2 summarises in one plot the effect of all the aforementioned components as a function of $\max(H_\mathrm{T}, p_\mathrm{T}(V))$. In the top pad the differential cross-section of an inclusive sample of $Z \to e^- e^+$ + jets events is presented as a function of $\max(H_\mathrm{T}, p_\mathrm{T}(V))$. The MC statistical uncertainty is reported on the right axis, and it is normalised such that the total number of raw events is the same for each set-up, $N_\mathrm{raw} = 10^7$. As anticipated, a large difference is observed between the set-up generated using the differential enhancement and the other two, for which the statistical uncertainty is comparable. From the middle pad it is possible to see that this difference is due to the distribution of the event weights. The lower pad shows a slightly larger fraction of negative weights in the tail of the distribution of the analytic enhancement set-up. Anyhow, this has a minor impact on the spread of the event weights. In the low p_T regime a reduction of f_nw was achieved in the latest developments of the SHERPA MC generator. Further details of the three set-ups can be found in ref. [6].

3. – Final remarks

The two approaches presented for the generation of MC events are characterised by both strengths and weaknesses. The slicing of the phase-space allows to control in an efficient manner the raw statistics in each of the different ranges, but it introduces a non-negligible spread of the MC event weights at the boundaries of each slice. Moreover, the granularity of the slicing is determined at the beginning of the generation stage.

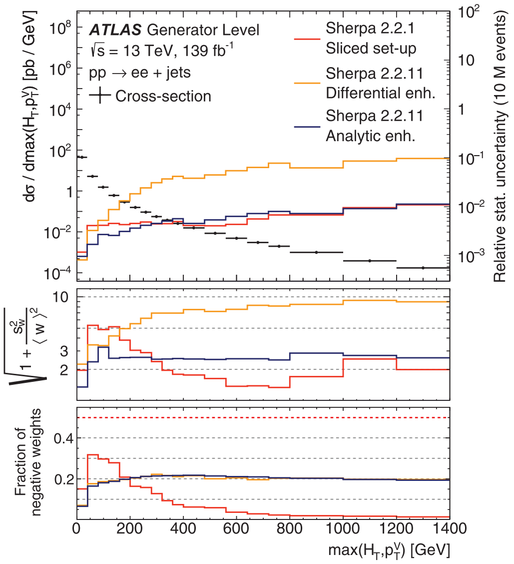


Fig. 2. – Summary plot comparing the statistical power of the sliced set-up (based on Sherpa v2.2.1), and the two samples generated with a biased cross-section (based on Sherpa v2.2.11), as a function of generator-level $\max(H_\mathrm{T}, p_\mathrm{T}(V))$. Figure from ref. [6].

The cross-section biasing method is a versatile way of shifting the bulk of the event generation in the desired phase-space thanks to the fact that the biasing function can be tuned for different needs. Furthermore, the statistical power of the event generation is fully predictable, as the cross-section biasing induces a continuous distribution of the MC event weights in the enhancement variable. The statistical performance of the two set-ups is comparable, but, as documented in ref. [6], the computing resources needed to generate a biased set-up are reduced with respect to a sliced one. The estimated reduction in terms of CPU consumption is of almost a factor 2 on the generation of a benchmark 200 M sample of $Z+\text{jets}$ events. For these reasons, the cross-section biasing method has been chosen for the next mass production of $V+\text{jets}$ MC events in the ATLAS Collaboration.

REFERENCES

[1] ATLAS Collaboration, *JINST*, **3** (2008) S08003.
[2] Buckley Andy *et al.*, *Phys. Rep.*, **504** (2011) 145, arXiv:1101.2599 [hep-ph].
[3] Evans Lyndon and Bryant Philip, *JINST*, **3** (2008) S08001.
[4] ATLAS Collaboration, CERN-LHCC-2020-015 (2020).
[5] Bothmann Enrico *et al.*, *SciPost Phys.*, **7** (2019) 034, arXiv:1905.09127 [hep-ph].
[6] ATLAS Collaboration, arXiv:2112.09588 [hep-ex] (Dec. 2021).