Available on CMS information server

CMS IN 2007/031

CMS Internal Note

The content of this note is intended for CMS internal use and distribution only

June 23, 2007

Study of V+jets in bins of boson P_T and jet multiplicity for 2007 $1\ fb^{-1}$ Monte Carlo production at CMS

M. Pierini, M. Spiropulu

European Laboratory for Nuclear Research CERN, Geneva, Switzerland

Abstract

We study the cross section of W/Z+jets events at the LHC, binned in boson P_T and jet multiplicity. We use the LO ALPGEN Monte Carlo generator to calculate the matrix element of the considered processes. The parton shower, the hadronization and the simulation of the underlying event is provided by PYTHIA. We apply parton shower matching using the MLM scheme, and calculate the expected number of events in $1\ fb^{-1}$ of proton-proton collisions at $\sqrt{s} = 14$ TeV per jet multiplicity bin in the range [1,5]. We thus deduce the number of events to be asked before unweighing in order to obtain the Monte Carlo statistics (after parton shower matching) corresponding to $1\ fb^{-1}$. We provide the necessary phase space grid (`.grid2` files) for each P_T bin and jet multiplicity bin, to be used for centralized production at CMS.

Contents

1 Introduction

The study of missing energy on transverse plane (MET) at LHC represents one of the available strategies to search for new particles produced by proton-proton collisions.

In fact, almost all New Physics (NP) models providing a Dark Matter candidates are characterized by a discrete symmetry (e.g. R-parity for supersymmetric (SUSY) models, T-parity in Little Higgs models, etc.) which prevents the lightest particle of the new physics spectrum to decay. This feature implies that the decay of any NP particle produced at LHC will generate a cascade, which will include one or more massive dark matter candidates escaping from the detector. The cascade also implies the production of high P_T leptons and/or jets. The relevant Standard Model (SM) background to these processes is the production of W+jets and Z+jets events in the tails of the boson P_T distribution, expecially when the Z decays to invisible.

In this study, we use LO Monte Carlo calculations of QCD matrix elements to evaluate the expected background for different jet multiplicities and in different bins of $P_T(W/Z)$. These values, together with the generation efficiency, allow to calculate the number of events to produce in order to obtain a statistics which is comparable to what we expect to collect with the first $1fb^{-1}$ of data. Using these values, we provide a proposal for the 2007 generation of W/Z+jet samples in CMS.

2 Setup of Monte Carlo Generator

In this work, we consider the case of event generation performed by the ALPGEN Monte Carlo generator [1], a matrix element (ME) calculator based on the ALPHA algorithm [2]. The calculation of the cross section for a given hard process is performed at the leading order (LO), through a numerical procedure which is organized in several steps. In this section, we give a short description of how the events are generated. Then, we describe the general and the process-specific settings that we used in this study.

2.1 Parton-level generation and cross section evaluation

The first step in the event generation (`imode` = `0` or `1`) is the production of a *weighted* dataset, performed according to the following procedure:

- The parameters required to define the hard process are passed to the code. These include the selection of jet multiplicity, the mass of possible heavy quarks, rapidity and transverse momentum cuts, etc.
- A first phase-space integration cycle is performed, with the goal of exploring how the cross-section is distributed in phase-space and among the possible contributing subprocesses. Event by event, the following steps take place:
  - one subprocess is randomly selected;
  - a point in phase-space is randomly selected, consistent with the required kinematic acceptance cuts;
  - the initial-state parton luminosity is evaluated for the chosen subprocess, and one among the possible flavor configurations is selected (see later);
  - spin and color for each parton are randomly assigned;
  - the matrix element is evaluated, and the weight of the event is obtained after inclusion of the phase-space and parton-luminosity factors. A bookkeeping of the weights is kept for each individual subprocess and phase-space sub-volume.
- At the end of the first integration cycle, a map of the cross-section distribution among the different subprocesses and in phase-space is available. It will be used for subsequent integration cycles, where the phase-space and subprocess random sampling will be weighted by the respective probability distributions.
- After the completion of a series of warm-up integration cycles (whose number is specified at the beginning of the run by the user), the optimized integration grids are stored in a (`.grid1`) file. A final large-statistics run is then performed and (for `imode` = `1`) the result is stored into a `.wgt` file. After the generation of each event, its kinematics is analyzed and histograms are filled. In addition, a second grid is saved (`.grid2` file), obtained from the combination of the warm-up cycles and the last run. Because of the larger number of generated events, this grid is more precise that the warm-up grid and should be taken as a reference for the centralized production of large statistics samples.

Since the generation of weighted parton-level events is typically by far the most CPU intensive component of the calculation, expecially for large jet multiplicities, the storage of weighted events allows to build up event data sets which can then be used efficiently for studies of hadronization systematics or realistic detector simulations. In addition, weighted events can be used to produce a dataset with different decays of the boson (selected during the unweighing stage).

In the second mode of operation (`imode = 2`) the code generates unweighted parton level events and stores them into a `.unw` file, for subsequent evolution via a code for parton-shower (PYTHIA [3] in the case of this study).

In this mode of operation the matrix-element calculation generates all the flavor and color information necessary for the complete shower evolution. The kinematic, flavor and color data for a given event are stored in a file, and are read in by the shower MC, which will process the event.

When performing the parton showering and hadronization process, a matching algorithm is applied, in order to assure a correspondence between the partons of the ME calculation and the hadron clusters of the final state. This additional step implies a matching efficiency which is $\sim 10\%$ ($\sim 50\%$) for high (low) jet multiplicity. In this paper, we use the so-called MLM algorithm [4], which is implemented by default in ALPGEN.

2.2 ALPGEN Version and Parameters

In this study, we use ALPGEN `v2.12`, which provides some bug fixing with respect to the previous version. In this version, some of the default parameters (such as beam energy, detector acceptance, and colliding particles) are fixed to typical Tevatron values. In addition, the specific features of this study demands some customization of the generated parton spectrum (this study is done in bins of $P_T(W/Z)$). In ALPGEN, the default setting can be modified in two ways:

- by changing the values of the parameters taken as input. This can be done interactively, or passing to the executable a data card.
- by coding in the `user.f` file (which is specific of the considered process) any additional cuts at the generation level.

In this study we used both the provided ALPGEN parameters and the user defined routine for selection of the boson P_T. All the changes we applied (with respect to the default release) are documented in the rest of this section.

2.3 Electroweak couplings

In the current version of the ALPHA code the input couplings are derived from the standard $SU(3) \times SU(2)_L \times U(1)_Y$ tree level Lagrangian. As a default for all processes, ALPGEN employs `iewopt = 3`. The ALPGEN authors have studied alternative options and determined that the changes induced on cross sections are at most a few percent. Gauge and Higgs boson widths are calculated at tree level after the couplings have been selected. With the exception of the class of processes involving several gauge bosons, the boson widths are set to 0; the ALPHA algorithm, on which ALPGEN is based, uses fixed widths in the propagators.

When `iewopt=3` is specified, the values of m_Z, m_W, and G_F are taken as input, from which the other electroweak parameters are extracted:

$$\begin{aligned} &\sin^2\theta_W = 1 - (m_W/m_Z)^2 = 0.2222, \\ &g = (4\sqrt{2}\,G_F)^{1/2} m_W = 0.6532, \\ &\alpha_{em}(m_Z) = (g\ \sin\theta_W)^2/4\pi = 1/132.5\ . \end{aligned} \tag{1}$$

In our study, this default setting is adopted.

2.4 Setting of Weighted Events Generation

The customization needed for the generation of weighted events is common to W+jets and Z+jets processes. In this section, we consider the case of W+jets, but all the conclusions apply to the other case as well.

When running in `imode` = `1` [1], a set of mandatory parameters has to be specified. These parameters can be passed to the executable as a card file

```
> ./wjetgen < input
```

or they are directly asked by the program when running. A minimal input file contains five parameters, the meaning of which is clear from the input file provided as an example by the authors:

```
1 !  imode
myprocess !  label for files
0 !  start with:  0=new grid, 1=previous warm-up grid, 2=previous generation
grid
10000 2 !  Nevents/iteration, N(warm-up iterations)
100000 !  Nevents generated after warm-up .
```

The label is used to name all the output files (`<label>.wgt`, `<label>.unw`, etc.). The other parameters define the size of the warm-up cycles and of the final sample. As a rule of thumb, we required a number of warm-up cycles equal to the jet multiplicity plus two. For each cycle, we required one million events in the case of low multiplicity and low P_T of the vector boson, increasing the value when going to larger values of both the quantities. The detailed list of values is given in sec. 4. The number of events after warm-up was taken big enough to assure an error on the computed cross section below the 1% level. When generating the events, we checked that the calculated cross sections for the same process in different warm-up cycles was stable within a $\sim 10\%$ tolerance. In the case that one of these two condition was not satisfied, the generation was repeated requiring larger statistics to be generated.

Optionally, the user can specify other settings, which overwrite the default values of the corresponding parameter. First of all, we need to set the flags so that LHC collisions are computed. This is particularly important since (up to the version `v2.12` considered for this study) the default values are still related to the Tevatron beam configuration. In order to change it, the user has to specify [2]:

```
ih2 1    !  Select pp (1) or ppbar (-1) collisions
ebeam 7000 !  beam energy in CM frame
```

Then, we need to modify the acceptance and minimum-momentum cuts for the partons, in order to adapt that to the geometry and performances of the CMS detector:

```
etajmax 5.  !  max|eta| for light jets
ptjmin 20.  !  minimum pt for light jets
```

while for charged leptons the default upper cut on η (`etalmax` = `10`) is large enough that in practise no cut is applied. As a last step, we pass two parameter:

```
ickkw 0 !  CKKW scale option:  set to 1 to enable jet-parton matching
```

to properly prepare the generated events for matching. This setting should be considered as a default for CMS, unless for those processes (such as $4b$+jets) for which a reasonably efficient matching is difficult with the available algorithms.

The expression for the renomarlization/factorization scale Q is process dependent and the user should take particular care when deciding what to use. Among the five possibilities, namely

```
iqopt=0=> Q=qfac
```

[1] The difference between running in `imode` = `1` and `imode` = `0` is only the fact that for `imode` = `1` the weighted events are dumped in an output file, while no output event is written on disk in the case of `imode` = `0`.

[2] In general, the user is advised to run at least once for each process asking `imode` = `4`, for which a `par.list` file containing the default parameters is written on disk. In this way, it will be possible to have a better control on what is actually generated

```
iqopt=1=> Q=qfac*sqrtm_W^2+ sum(pt_jet^2)
iqopt=2=> Q=qfac*mW
iqopt=3=> Q=qfac*sqrtm_W^2+ pt_W^2
iqopt=4=> Q=qfac*sqrtsum(pt_jet^2),
```

we decided to use `iqopt=1`, which corresponds to a better description of the Tevatron data and is also advised by the ALPGEN authors. A similar choice is done for when considering Z+jets processes:

```
iqopt=0=> Q=qfac
iqopt=1=> Q=qfac*sqrtm0^2+ sum(pt_jet^2)
iqopt=2=> Q=qfac*m0
iqopt=3=> Q=qfac*sqrtm_0^2+ pt_Z^2
iqopt=4=> Q=qfac*sqrtsum(pt_jet^2),
```

where `m0` is the invariant dilepton mass. It is important to stress the fact that ALPGEN also allows to apply a rescale factor to the scale Q.

```
qfac 1 !  Options for rescaling the factorization/renormalization scale Q
```

This factor is fixed to one by default (`qfac = 1`). As we know from the Tevatron experience, changing these two parameters (`iqopt` and `qfac`) is a powerful handle to tune the Monte Carlo predictions (which, we remind, are calculated at LO) to the data. This means that the values used here are going to be studied using data, when a sufficient statistics will be available to perform a data/Monte Carlo comparison study, similar to what was done at the Tevatron.

In addition, we would like to mention a relevant difference with respect to the ALPGEN v2.05 production (see appendix). In that case we used ALPGEN v2.05, in which the setting `iqopt=1` and `iqopt=3` were swapped. Because of the (above mentioned) dependence of the cross section on this parameter, this difference alone introduces a 20% discrepancy among the cross sections for low jet multiplicity.

The setting of Z+jets simulation requires additional care, because of the presence of additional requirements on the dilepton mass, namely:

```
mllmin= 40.  !  min dilepton inv mass
mllmax= 200.  !  max dilepton inv mass
```

These default values are consistent for our need, but they might not be optimal for other exclusive cases, such as EW studies where the datasets may be binned in mass bins. This is another example of the fact that, when using with ALPGEN, it is of primary importance to produce a `par.list` file for each process. Note that ALPGENactually generates $\ell\ell$+jets (γ/Z production) and that the mass range is important when we form the SM ratio of $W\ell\nu$+jets/$Z\ell\ell$+jets. In fact we expect the ratio to be slighly lower than 10 for the inclusive ratio ($W\ell\nu$+jets/$Z/\gamma\ell\ell$+jets) and between 10 and 11 when the mass window of the $\ell\ell$ system is around the Z-pole. Note also that when we bin in boson P_T in the case of the $Z/\gamma\ell\ell$+jets it in fact the P_T of the Z that we use.

2.5 Setting of Unweighted Events Generation

Running in `imode = 2`, ALPGEN generates a set of unweighted events and provides a calculation of the cross section for the considered process. In the case of a W (Z) +jets, the calculated matrix element refers to a charged lepton+neutrino (two charged leptons) final state, the quoted cross sections referring to a single lepton family. All spin correlations and finite width effects are accounted for. In the flavor assignment, the code selects by default electrons. The EW parameters are fixed by default using the option `iewopt=3` (see eq. (2)). The run produces a `.unw` file, which can then be passed to a parton shower Monte Carlo code to produce the actual dataset.

At this stage, it is possible to use the same sample of weighted events to generated unweighted samples corresponding to the same process, but with different flavor content. This is done initializing a set of flags which are specific of each process, and (if possible) keeping it in a database for provenance.

In the case of W+jets, one has to set the value of the `iwdecmode` flag to one of the following options:

- $1 \ = \ e\bar{\nu}_e$

- $2 = \mu\bar{\nu}_\mu$
- $3 = \tau\bar{\nu}_\tau$
- $4 = \ell\bar{\nu}_\ell$
- $5 = q\bar{q}'$
- $6 =$ fully inclusive

where q indicates a light quark. In the case of Z+jets, the Z is forced to decay to two leptons. To select the flavor of the final state, the user has to set the `izdecmode` flag to one of the following values:

- $1 = e^+e^-$
- $2 = \mu^+\mu^-$
- $3 = \tau^+\tau^-$
- $4 = \ell\ell$.

When choosing `izdecmode = 4`, the user can select $Z \to \ell^+\ell^-$ (`ilep = 1`) or $Z \to \nu\bar{\nu}$ (`ilep = 0`). In both cases, the sum over the three lepton families is implicit.

To conclude, let us remark the fact that different decays can correspond to the same value of `izdecmode` for different processes. For example, a larger set of options for Z decays are available when Z are produced using `vbjet` rather than `zjet`. In this case, in fact, the `izdecmode` flag can be set to

- $1 = \nu\bar{\nu}$
- $2 = \ell^+\ell^-$
- $3 = q\bar{q}$
- $4 = b\bar{b}$
- $5 =$ fully inclusive

where q indicates a light quark. This means that, if the user is interested to $Z \to b\bar{b}$+ Njets (in order to measure the mass resolution of $b\bar{b}$ final states using Z as a candle), `vbjet` should be used instead of `zjets`. This and other aspects of ALPGEN are fully documented in the manual [1], which the user is advised to read before deciding the process to use.

2.6 User Filter at Generation Level

Using ALPGEN, the user has the possibility of adding a set of cuts on the partons, which allow to calculate the matrix element of the considered process with additional kinematic or angular cuts. This is possible by editing the `<process>usr.f` file, which is present in each `process/` directory. In particular, this file contains a subroutine `usrcut`, where the user can implement customized selection code.

We used this feature to bin the cross section as a function of the vector-boson P_T. For instance, in the case of `wjetwork/wjetusr.f`, we modified the code as follows (for the case of the first bin, i.e. $P_T < 100$ GeV/c^2):

```
c-------------------------------------------------------------------
subroutine usrcut(lnot,wusr)
c-------------------------------------------------------------------
c PRIMARY CUTS ALREADY APPLIED TO PHASE-SPACE GENERATION:
c ptjmin < pt(jet) < ptjmax for all light jets
c -etajmax < eta(jet) < etajmax for all light jets
c delta R(jj) > drjmin for all (light jet, light jet) pairs
c pt(lept)>ptlmin etmiss > minetmiss
c abs(eta(lept)) < etalmax
```

```
c lepton/jet isolation
c
c USE THIS ROUTINE TO ENFORCE OTHER CUTS
implicit none
include 'alpgen.inc'
include 'wjet.inc'
integer lnot
double precision wusr

c definition of ptw variable for tails cut
real ptw
c
lnot=0
wusr=1d0

c ptw tails cu
ptw=sqrt(pw(1)**2+pw(2)**2)
if(ptw.gt.100) goto 10

return
10 lnot= 1
end
```

Similar code is implemented in `zjetwork/zjetusr.f` for the case of Z+jets process and for different P_T bins.

3 Parton Shower And Matching

In this section, we discuss the aspects of the simulation which are related to the showering of the partons and the hadronization. We describe the matching of the parton shower to the matrix elements and the approaches proposed in literature. We conclude giving a detailed explanation on how the parton shower simulation of ALPGEN unweighted samples is done in the CMSSW framework and the changes we applied for this study.

3.1 Matching of Parton Showers and Matrix Elements

The evolution from the parton-level process to the final state actually observed in the detector is the critical step for any jet simulation. This evolution, performed at $\sqrt{s} = 14$ TeV using one of the parton-shower Monte Carlo generators available, introduces a problem of double counting of hard parton emissions, which generates an undesired dependence of the calculated cross section on the (angular and kinematic) parameters applied on the partons at the generation level. This dependence can be reduced applying an algorithm which avoids double counting of hard processes.

To be more specific, consider the case of a final state with $N + 1$ jets. This configuration can be obtained starting from the production of $N + 1$ partons in a hard proton-proton interaction, followed by the production of one jet from each parton, or from a final state with N partons, when a gluon or quark is emitted in the shower simulation, within enough momentum and at a sufficiently large angle to start a new jet. In a jet production at high energy, the presence of several hard scales (such as the jet transverse energies or the dijet invariant masses) breaks the factorization among the parton generation and the shower evolution. This means that there is an intermediate region of the phase space for which the two different ways of producing a final state of $N + 1$ jets actually correspond to the same process. At present, a rigorous solution to this problem does not exist. On the other hand, several approaches have been proposed in the literature [5] to reduce this dependence. Regardless of the approach used, a matching procedure should go through the following steps:

- Definition of a jet measure and calculation of all the relevant cross sections for a given process (e.g. for $pp \to X$, calculate $\sigma(pp \to X + Njets)$ with $N = 0, 1, ..., N_{max}$).
- Production of hard-partons samples, with probability proportional to the cross section.

- Use of a dynamical and kinematic criterion that accepts or rejects events, taking into account both the running of the coupling constants and the probability for the parton to propagate in a certain pattern of the phase space without branching (Sudakov effect).
- Generation of parton shower for each leg, with the constraint of keeping the number of final jets equal to the number of initial partons.

The differences among the approaches are represented by the choice of the jet measure, of the angular and kinematic requirements that define the rejection of the events, and the requirements that constrain the parton shower algorithm to generate one jet from one parton in the final step. We now give a brief description of some of these approaches.

In the study presented here, we used the MLM matching algorithm, which is provided as a default in ALPGEN and has proved a good description of the Tevatron data. On the other hand, we will not necessarily follow this approach when dealing with real data. In fact, only a full data/Monte Carlo comparison study will allow to understand which is the best matching strategy, in terms of data description and associated systematic error. In the meanwhile, the reader should keep in mind the fact that several possibilities are available and other approaches might appear in the future.

The matching algorithm applied in this study is different than what was done in the past (see appendix), since the lower cut on jet p_T was set moved from 15 GeV/c to 20 GeV/c.

3.1.1 CKKW Matching

The CKKW matching [6], implemented in the SHERPA [7] parton shower generator, is based on a full separation among the acceptance/rejection procedure and the shower evolution. According to this approach, after the hadronization two particles are considered to belong to different jets if their $k_\perp$ distance

$$k_\perp^{(ij)} = \sqrt{2 \min p_\perp^i, p_\perp^j [\cosh(\eta^i - \eta^j) - \cos(\phi^i - \phi^j)]} \tag{2}$$

is larger than a critical value $k_{\perp,0}$. At the same time, each jet is required to have a transverse momentum larger than $k_{\perp,0}$. The matrix elements are then reweighted according to Sudakov factors, which take into account those terms that would appear in the shower evolution. After the reweighting is applied and a suitable starting condition for the shower generation is fixed, the shower is developed vetoing *any* hard emission, the probability of which is already accounted for by the Sudakov reweighting.

3.1.2 Dipole Cascade and CKKW Matching

The dipole cascade approach, implemented in ARIADNE [8], replaces the usual $1 \rightarrow 2$ branching processes with $2 \rightarrow 3$ splittings. Form instance, gluon radiation is modeled as coherently produced by color-anticolor charged parton pairs. The emission in the dipole cascade is ordered according to the transverse momentum, defined as

$$p_\perp^2 = \frac{s_{12} s_{23}}{s_{123}} \tag{3}$$

where s_{ij} is the invariant mass of the partons i and j and the index 2 corresponds to the emitted parton. When applied to hadronic collisions, this method treats all the emissions as originating from a dipole final states. This implies the need of a modeling for several possible branching processes [5].

When implementing the CKKW matching in the context of dipole shower, some changes are required with respect to the description of the CKKW approach given above. First, unlike the *standard* CKKW approach, the emission scale is not reconstructed using the $k_\perp$ algorithm. Instead, the output of the matrix element calculation is evolved through a full parton shower simulation, i.e. for each parton configuration the full simulation by ARIADNE is taken as the starting point of the matching procedure. The shower simulation produces a set of intermediate partons S_i and the corresponding emission scale $p_{\perp i}$.

Then, for each parton, a trial emission is simulated, using the intermediate scale $p_{\perp i}$ for each S_i state. If the $p_\perp$ of the emitted parton is larger than $p_{\perp i}$, the state is rejected. This is equivalent of keeping the events according to the no-emission probability of each line of the shower, as calculated by ARIADNE, i.e. to introduce the Sudakov suppression after the shower simulation.

3.1.3 MLM Algorithm

The MLM algorithm [4] is defined by the following rules:

- Generate parton-level configurations for all final states up to a multiplicity N. Partons are constrained by requiring
$$p_T^{part} > p_T^{min} \quad , \quad |\eta_{part}| < \eta_{max} \quad , \quad \Delta R > R_{min}, \tag{4}$$
where p_T^{part} and η_{part} are the momentum and pseudorapidity of the light partons, and ΔR is the distance among partons in the (η, ϕ) plane.

- Perform shower evolution using one of the available shower Monte Carlo codes, e.g (for ALPGEN) PYTHIA or HERWIG.

- Apply a jet algorithm to the partons of the shower evolution, before hadronization, generating a list of *clusters*. Each cluster is defined by the minimum E_T, E_T^{clus}, and by a jet cone size R_{clus} [3)].

- Associate the initial partons to one of the clusters:
  - starting from the highest p_T parton, find the cluster with minimum distance ΔR from it. If $\Delta R < R_{match}$ (where R_{match} is the *matching radius*) the parton is matched.
  - remove the matched cluster from the list of clusters and iterate the procedure starting from the next to highest energetic parton.

- Unless each parton is matched to a cluster, the event is rejected.

- If the event we are considering has a number of jets $n < N$ and a number of clusters $N_{clust} > n$, the event is rejected. Otherwise, it contributes to the exclusive sample with $N_{clust} = n$.

- If $n = N$ (i.e to the largest parton multiplicity considered when generating parton level events), the events with $N_{clus} > N$ are not rejected, provided that the additional clusters are softer than each of the matched clusters.

- After matching, the exclusive samples with $n < N$ are combined to the sample with $n = N$ to define the inclusive sample.

The MLM matching is implemented in ALPGEN using the GETJET cone algorithm for cluster definition, but any alternative choice is allowed.

Note that each of the exclusive samples should be weighted according to its individual LO cross section. The final inclusive sample can be reweighted using a NLO cross-section calculation such as MCFM [9], provided that the same requirements are used at the NLO calculation.

3.2 Parton Shower and Hadronization in CMSSW

The unweighted events produced by ALPGEN can be evolved to a hadron shower in the CMSSW framework, by using a `.cfg` interface among ALPGEN and one of the available Monte Carlo simulators of shower evolution. At present, the choice of the shower simulator is limited to PYTHIA and HERWIG, since other solutions are not supported by ALPGEN. In addition, an interface for HERWIG is not yet available in CMSSW, so that at the end there was not even a decision to take.

For this study, we used PYTHIA `v6.409` in `CMSSW 1_4_2` through the interface provided with the CMSSW release, namely `CMSSW_1_4_2 GeneratorInterface/AlpgenInterface/test/Alpgen.cfg`. Starting from the parton level configurations generated by ALPGEN, PYTHIA takes care of simulating the parton shower evolution and the hadronization of the final state, as well as the particles associated to the underlying event. When used through the `CMSSW` interface, it requires as inputs the `.unw` and `_unw.par` files generated by ALPGEN in the `imode` = `2` run.

The interface provides a very good starting example, but some changes were required. In particular, we changed the following two parameters as follows:

[3)] These parameters are related to the p_T^{min} and R_{min} parameters applied on the partons before the shower; but they are not necessarily the same.

```
"RXpar(1) = 20.  !  ETCLUS : minET(CLUS)",
"RXpar(2) = 0.7 !  RCLUS : deltaR(CLUS)"
```

in order to apply the MLM matching algorithm as it was done with the ALPGEN `v2.05` CMS production in 2005/2006. The user should pay attention to the fact that, in the case of the largest jet multiplicity for the considered process the MLM matching has to be applied inclusively. This means that also the following line of the interface has to be modified:

```
"IXpar(2) = 1 !  inclus./exclus.  sample:  0/1",
```

when the decided largest number of jet is considered. When running through this interface, PYTHIA produces a standard EDM `.root` file, where the collection of generated showers are saved, including the underlying event. The collection can then be processed through the full CMS detector simulation and trigger and analyzed with standard tools.

4 Results

In this section, we present the results of the generation, which allow to estimate the needed requirements to simulated a sample of $\sim 1\text{fb}^{-1}$.

The simulations are performed for W+Njets and Z+Njets events, as a function of jet multiplicity ($1 \leq$Njets≤ 5) and with increasing values of P_T. Using as a metric the DCO4 production binning, where a cut on $\hat{p}_T$ determined the boson P_T bins (see Appendix), we implemented a different boson P_T binning (inclding the low boson P_T bin) as follows:

- $0 < P_T(W) < 100$ GeV/c
- $100 < P_T(W) < 300$ GeV/c
- $300 < P_T(W) < 800$ GeV/c
- $800 < P_T(W) < 1600$ GeV/c
- $1600 < P_T(W) < 3200$ GeV/c
- $3200 < P_T(W) < 5000$ GeV/c

In addition, W/Z+0jets samples were also generated, to provide a reference bin. As we will discuss in the final section, these two samples require very large statistics. It may be wise to consider $2 \rightarrow 1$ generation with PYTHIA or other dedicated generators for W+0jets and Z+0jets samples.

In the rest of this section, results are given in terms of:

- For W+Njets and Z+Njets with Njets< 6, a set of six tables (one for each P_T bin) detailing the size of the samples in input, after weighted generation and after unweighting. The unweighting procedure by default forces the vector bosons to decay as $W \rightarrow e\nu_e$ and $Z \rightarrow e^+e^-$.
- For W+Njets and Z+Njets, a set of six tables (one for each P_T bin) detailing the efficiency of the matching procedure, the number of events expected in a sample of $\sim 1\text{fb}^{-1}$ (from the ALPGEN calculation of the cross section), and the number of events needed as input to obtain a comparable Monte Carlo statistics, after matching. For these results, inclusive $W \rightarrow \ell\nu_\ell$ $Z \rightarrow \ell^+\ell^-$, and $Z \rightarrow \nu_\ell^+\nu_\ell^-$ decays are considered.
- A set of six plots (one for each P_T bin) showing the ratio of calculated cross section for W+Njets over Z+Njets.
- Distribution of inclusive and exclusive cross sections for W+Njets and Z+Njets (as a function of jet multiplicity) for the first P_T bin.

sample	warmup PS-Grid	N_{asked}	N_{wgt}	σ_{wgt} (pb)	N_{unwgt}	σ_{unwgt} (pb)
w0j	1M,3	5M	3979189	18063	3558154	18063
w1j[$0 < P_T(W) < 100$ GeV/c]	1M,4	5M	2802796	5207	645408	5207
w2j[$0 < P_T(W) < 100$ GeV/c]	1M,5	5M	140895	2407	34681	2164
w3j[$0 < P_T(W) < 100$ GeV/c]	1M,6	25M	177108	1080	14816	784
w4j[$0 < P_T(W) < 100$ GeV/c]	2M,7	50M	92994	477	20591	268
w5j[$0 < P_T(W) < 100$ GeV/c]	4M,8	100M	124586	202	14400	85

Table 1: $\sigma(e\nu + N \text{ jets})$ at the LHC, for $P_T(W) < 100.$ GeV/c

sample	warmup PS-Grid	N_{asked}	N_{wgt}	σ_{wgt} (pb)	N_{unwgt}	σ_{unwgt} (pb)
w1j[$100 < P_T(W) < 300$ GeV/c]	1M,4	5M	909132	183	327890	183
w2j[$100 < P_T(W) < 300$ GeV/c]	1M,5	5M	157985	281	3205	230
w3j[$100 < P_T(W) < 300$ GeV/c]	1M,6	25M	42645	250	15684	164
w4j[$100 < P_T(W) < 300$ GeV/c]	2M,7	50M	220186	168	19897	86
w5j[$100 < P_T(W) < 300$ GeV/c]	4M,8	100M	134313	95	12736	37

Table 2: $\sigma(e\nu + N \text{ jets})$ at the LHC, for $100 < P_T(W) < 300$ GeV/c.

4.1 W+jets Generation

From Tab. 1 to Tab. 6 we give the details for the generation of weighted and unweighted samples of W+jets events. For the unweighting procedure, we assumed $W \to \ell\nu_\ell$. The number of events asked in generation and the number of warm-up cycles has been chosen to provide an error on the cross section 1%.

4.2 Z+jets Generation

From Tab. 7 to Tab. 12 we give the details for the generation of weighted and unweighted samples of Z+jets events. For the unweighting procedure, we assumed $Z \to \ell^+\ell^-$. The number of events asked in generation and the number of warm-up cycles has been chosen to provide an error on the cross section 1%. To give a sense of the time scale to produce a stable grid, the last bin in multiplicity and $P_T(W)$ of the Z+jets took 11 dayes in a 5.6 GHz machine.

4.3 Size of Monte Carlo Samples for 1 fb^{-1}

From Tab. 13 to Tab. 30 we summarize the expected and needed events corresponding to a statistics of 1 fb^{-1}.

sample	warmup PS-Grid	N_{asked}	N_{wgt}	σ_{wgt} (pb)	N_{unwgt}	σ_{unwgt} (pb)
w1j[$300 < P_T(W) < 800$ GeV/c]	1M,4	5M	231166	3.2	150406	3.2
w2j[$300 < P_T(W) < 800$ GeV/c]	1M,5	5M	67506	8.3	17563	6.3
w3j[$300 < P_T(W) < 800$ GeV/c]	1M,6	25M	209791	11.3	17583	6.5
w4j[$300 < P_T(W) < 800$ GeV/c]	2M,7	50M	246415	10.8	13366	4.8
w5j[$300 < P_T(W) < 800$ GeV/c]	4M,8	100M	25162	7.0	1970	2.4

Table 3: $\sigma(e\nu + N$ jets) at the LHC, for $300 < P_T(W) < 800$ GeV/c.

sample	warmup PS-Grid	N_{asked}	N_{wgt}	σ_{wgt} (pb)	N_{unwgt}	σ_{unwgt} (pb)
w1j[$800 < P_T(W) < 1600$ GeV/c]	1M,4	5M	190277	0.03	138150	0.03
w2j[$800 < P_T(W) < 1600$ GeV/c]	1.5M,5	7.5	53553	0.10	11623	0.07
w3j[$800 < P_T(W) < 1600$ GeV/c]	1.5M,6	40M	142943	0.17	17707	0.09
w4j[$800 < P_T(W) < 1600$ GeV/c]	3M,7	80M	93995	0.19	12961	0.07
w5j[$800 < P_T(W) < 1600$ GeV/c]	6M,8	160M	160903	0.16	5431	0.05

Table 4: $\sigma(e\nu + N$ jets) at the LHC, for $800 < P_T(W) < 1600$ GeV/c.

sample	warmup PS-Grid	N_{asked}	N_{wgt}	σ_{wgt} (pb)	N_{unwgt}	σ_{unwgt} (pb)
w1j[$1600 < P_T(W) < 3200$ GeV/c]	1M,4	5M	70049	0.0003	27604	0.0003
w2j[$1600 < P_T(W) < 3200$ GeV/c]	2M,5	10M	86091	0.0012	8294	0.0008
w3j[$1600 < P_T(W) < 3200$ GeV/c]	2M,6	50M	43910	0.0021	4624	0.0010
w4j[$1600 < P_T(W) < 3200$ GeV/c]	4M,7	100M	91533	0.0025	6768	0.0009
w5j[$1600 < P_T(W) < 3200$ GeV/c]	8M,8	200M	127767	0.0021	4375	0.0006

Table 5: $\sigma(e\nu + N$ jets) at the LHC, for $1600 < P_T(W) < 3200$ GeV/c.

sample	warmup PS-Grid	N_{asked}	N_{wgt}	σ_{wgt} (pb)	N_{unwgt}	σ_{unwgt} (pb)
w1j[$3200 < P_T(W) < 5000$ GeV/c]	1M,4	5M	112526	1.3E-7	44321	1.3E-7
w2j[$3200 < P_T(W) < 5000$ GeV/c]	2.8M,5	14M	38446	6.5E-7	3496	4.1E-7
w3j[$3200 < P_T(W) < 5000$ GeV/c]	2.8M,6	70M	86487	1.3E-6	8684	5.7E-7
w4j[$3200 < P_T(W) < 5000$ GeV/c]	6M,7	140M	64358	1.3E-6	3622	4.5E-7
w5j[$3200 < P_T(W) < 5000$ GeV/c]	12M,8	280M	69408	1.0E-6	4324	2.5E-7

Table 6: $\sigma(e\nu + N \text{ jets})$ at the LHC, for $3200 < P_T(W) < 5000$ GeV/c.

sample	warmup PS-Grid	N_{asked}	N_{wgt}	σ_{wgt} (pb)	N_{unwgt}	σ_{unwgt} (pb)
z1j[$0 < P_T(Z) < 100$ GeV/c]	1M,4	5M	1138753	551	157128	551
z2j[$0 < P_T(Z) < 100$ GeV/c]	1M,5	5M	81565	254	27648	234
z3j[$0 < P_T(Z) < 100$ GeV/c]	1M,6	25M	62000	115	30643	87
z4j[$0 < P_T(Z) < 100$ GeV/c]	2M,7	50M	163392	51	20540	29
z5j[$0 < P_T(Z) < 100$ GeV/c]	4M,8	100M	81495	21	12843	9

Table 7: $\sigma(ee + N \text{ jets})$ at the LHC, for $0 < P_T(Z) < 100$ GeV/c.

sample	warmup PS-Grid	N_{asked}	N_{wgt}	σ_{wgt} (pb)	N_{unwgt}	σ_{unwgt} (pb)
z1j[$100 < P_T(Z) < 300$ GeV/c]	1M,4	5M	1012432	22	121903	22
z2j[$100 < P_T(Z) < 300$ GeV/c]	1M,5	5M	75954	35	18756	29
z3j[$100 < P_T(Z) < 300$ GeV/c]	1M,6	25M	135682	31	26277	21
z4j[$100 < P_T(Z) < 300$ GeV/c]	2M,7	50M	165372	20	18692	11
z5j[$100 < P_T(Z) < 300$ GeV/c]	4M,8	100M	132143	11	10791	4

Table 8: $\sigma(ee + N \text{ jets})$ at the LHC, for $100 < P_T(Z) < 300$ GeV/c.

sample	warmup PS-Grid	N_{asked}	N_{wgt}	σ_{wgt} (pb)	N_{unwgt}	σ_{unwgt} (pb)
z1j[$300 < P_T(Z) < 800$ GeV/c]	1M,4	5M	737654	0.42	109125	0.42
z2j[$300 < P_T(Z) < 800$ GeV/c]	1M,5	5M	101709	1.09	12295	0.83
z3j[$300 < P_T(Z) < 800$ GeV/c]	1M,6	25M	255644	1.47	19046	0.86
z4j[$300 < P_T(Z) < 800$ GeV/c]	2M,7	50M	72354	1.39	12106	0.62
z5j[$300 < P_T(Z) < 800$ GeV/c]	4M,8	100M	125175	1.02	8198	0.34

Table 9: $\sigma(ee + N$ jets$)$ at the LHC, for $300 < P_T(Z) < 800$ GeV/c.

sample	warmup PS-Grid	N_{asked}	N_{wgt}	σ_{wgt} (pb)	N_{unwgt}	σ_{unwgt} (pb)
z1j[$800 < P_T(Z) < 1600$ GeV/c]	1M,4	5M	430491	0.004	150017	0.004
z2j[$800 < P_T(Z) < 1600$ GeV/c]	1.5M,5	7.5M	129059	0.013	12296	0.009
z3j[$800 < P_T(Z) < 1600$ GeV/c]	1.5M,6	40M	246401	0.021	18861	0.011
z4j[$800 < P_T(Z) < 1600$ GeV/c]	3M,7	80M	118652	0.021	11193	0.009
z5j[$800 < P_T(Z) < 1600$ GeV/c]	6M,8	160M	71939	0.020	5854	0.010

Table 10: $\sigma(ee + N$ jets$)$ at the LHC, for $800 < P_T(Z) < 1600$ GeV/c.

sample	warmup PS-Grid	N_{asked}	N_{wgt}	σ_{wgt} (pb)	N_{unwgt}	σ_{unwgt} (pb)
z1j[$1600 < P_T(Z) < 3200$ GeV/c]	1M,4	5M	587902	0.00003	67129	0.00003
z2j[$1600 < P_T(Z) < 3200$ GeV/c]	2M,5	10M	92582	0.00015	9468	0.00010
z3j[$1600 < P_T(Z) < 3200$ GeV/c]	2M,6	50M	142331	0.00027	10946	0.00013
z4j[$1600 < P_T(Z) < 3200$ GeV/c]	4M,7	100M	59083	0.00032	5645	0.00011
z5j[$1600 < P_T(Z) < 3200$ GeV/c]	8M,8	200M	33478	0.00030	422	0.00007

Table 11: $\sigma(ee + N$ jets$)$ at the LHC, for $1600 < P_T(Z) < 3200$ GeV/c.

sample	warmup PS-Grid	N_{asked}	N_{wgt}	σ_{wgt} (pb)	N_{unwgt}	σ_{unwgt} (pb)
z1j[$3200 < P_T(Z) < 5000$ GeV/c]	1M,4	5M	617649	1.7E-8	123469	1.7E-8
z2j[$3200 < P_T(Z) < 5000$ GeV/c]	2.8M,5	14M	86959	8.6E-8	11543	5.4E-8
z3j[$3200 < P_T(Z) < 5000$ GeV/c]	2.8M,6	70M	107959	1.6E-7	9913	7.0E-8
z4j[$3200 < P_T(Z) < 5000$ GeV/c]	6M,7	140M	113130	1.7E-7	1657	5.3E-8
z5j[$3200 < P_T(Z) < 5000$ GeV/c]	12M,8	280M	120276	1.2E-7	3361	2.9E-8

Table 12: $\sigma(ee + N \text{ jets})$ at the LHC, for $3200 < P_T(Z) < 5000$ GeV/c.

sample	ϵ_{match}	$N_{ev/fb}$	$N_{generate}$
w0j	0.84	4.53E+07	7.61E+07
w1j[$0 < P_T(W) < 100$ GeV/c]	0.59	9.24E+06	1.21E+08
w2j[$0 < P_T(W) < 100$ GeV/c]	0.39	2.54E+06	9.36E+08
w3j[$0 < P_T(W) < 100$ GeV/c]	0.25	5.88E+05	3.97E+09
w4j[$0 < P_T(W) < 100$ GeV/c]	0.15	1.24E+05	1.95E+09
w5j[$0 < P_T(W) < 100$ GeV/c]	0.33	8.43E+04	1.78E+09

Table 13: Summary of expected and required Monte Carlo statistics for $W \to \ell\nu_\ell$+ Njets ($\ell = e, \mu, \tau$), corresponding to 1 fb^{-1} of data collected at LHC, for $0 < P_T(W) < 100$ GeV/c.

sample	ϵ_{match}	$N_{ev/fb}$	$N_{generate}$
w1j[$100 < P_T(W) < 300$ GeV/c]	0.47	2.56E+05	8.37E+06
w2j[$100 < P_T(W) < 300$ GeV/c]	0.33	2.25E+05	1.08E+09
w3j[$100 < P_T(W) < 300$ GeV/c]	0.22	1.07E+05	7.87E+08
w4j[$100 < P_T(W) < 300$ GeV/c]	0.15	3.77E+04	6.49E+08
w5j[$100 < P_T(W) < 300$ GeV/c]	0.36	3.96E+04	8.61E+08

Table 14: Summary of expected and required Monte Carlo statistics for $W \to \ell\nu_\ell$+ Njets ($\ell = e, \mu, \tau$), corresponding to 1 fb^{-1} of data collected at LHC, for $100 < P_T(W) < 300$ GeV/c.

sample	ϵ_{match}	$N_{ev/fb}$	$N_{generate}$
w1j[$300 < P_T(W) < 800$ GeV/c]	0.30	2.91E+03	3.21E+05
w2j[$300 < P_T(W) < 800$ GeV/c]	0.22	4.04E+03	5.34E+06
w3j[$300 < P_T(W) < 800$ GeV/c]	0.16	3.07E+03	2.79E+07
w4j[$300 < P_T(W) < 800$ GeV/c]	0.11	1.56E+03	5.37E+07
w5j[$300 < P_T(W) < 800$ GeV/c]	0.43	3.03E+03	3.60E+08

Table 15: Summary of expected and required Monte Carlo statistics for $W \to \ell\nu_\ell$+ Njets ($\ell = e, \mu, \tau$), corresponding to 1 fb^{-1} of data collected at LHC, for $300 < P_T(W) < 800$ GeV/c.

sample	ϵ_{match}	$N_{ev/fb}$	$N_{generate}$
w1j[$800 < P_T(W) < 1600$ GeV/c]	0.19	1.59E+01	2.99E+03
w2j[$800 < P_T(W) < 1600$ GeV/c]	0.15	3.15E+01	1.33E+05
w3j[$800 < P_T(W) < 1600$ GeV/c]	0.12	3.01E+01	5.89E+05
w4j[$800 < P_T(W) < 1600$ GeV/c]	0.09	1.88E+01	1.35E+06
w5j[$800 < P_T(W) < 1600$ GeV/c]	0.43	5.92E+01	4.04E+06

Table 16: Summary of expected and required Monte Carlo statistics for $W \to \ell\nu_\ell$+ Njets ($\ell = e, \mu, \tau$), corresponding to 1 fb^{-1} of data collected at LHC, for $800 < P_T(W) < 1600$ GeV/c.

sample	ϵ_{match}	$N_{ev/fb}$	$N_{generate}$
w1j[$1600 < P_T(W) < 3200$ GeV/c]	0.16	1.34E-01	1.48E+02
w2j[$1600 < P_T(W) < 3200$ GeV/c]	0.13	3.04E-01	2.77E+03
w3j[$1600 < P_T(W) < 3200$ GeV/c]	0.11	3.47E-01	3.38E+04
w4j[$1600 < P_T(W) < 3200$ GeV/c]	0.08	2.24E-01	4.05E+04
w5j[$1600 < P_T(W) < 3200$ GeV/c]	0.43	7.43E-01	7.81E+04

Table 17: Summary of expected and required Monte Carlo statistics for $W \to \ell\nu_\ell$+ Njets ($\ell = e, \mu, \tau$), corresponding to 1 fb^{-1} of data collected at LHC, for $1600 < P_T(W) < 3200$ GeV/c.

sample	ϵ_{match}	$N_{ev/fb}$	$N_{generate}$
w1j[$3200 < P_T(W) < 5000$ GeV/c]	0.15	5.73E-05	4.40E-02
w2j[$3200 < P_T(W) < 5000$ GeV/c]	0.14	1.68E-04	4.93E+00
w3j[$3200 < P_T(W) < 5000$ GeV/c]	0.12	1.99E-04	1.38E+01
w4j[$3200 < P_T(W) < 5000$ GeV/c]	0.09	1.24E-04	5.22E+01
w5j[$3200 < P_T(W) < 5000$ GeV/c]	0.44	3.31E-04	4.86E+01

Table 18: Summary of expected and required Monte Carlo statistics for $W \to \ell\nu_\ell$+ Njets ($\ell = e, \mu, \tau$), corresponding to 1 fb^{-1} of data collected at LHC, for $3200 < P_T(W) < 5000$ GeV/c.

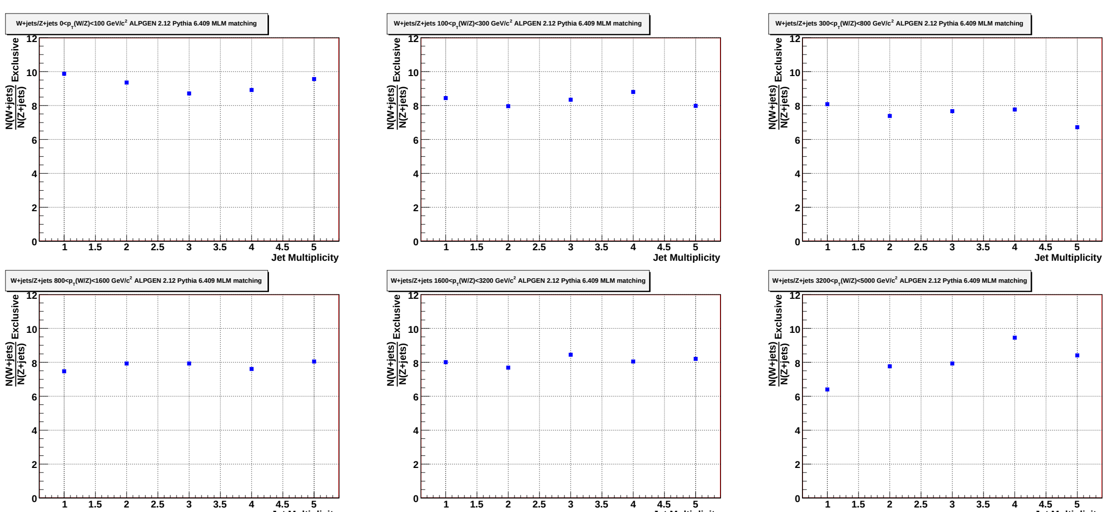

Figure 0: Ratio of exclusive $\sigma(W + Njets)/\sigma(Z + Njets)$ as a function of jet multiplicity, after the matching procedure has been applied.

5 1 fb^{-1} generation needs

For production using the PA and providing the phase space .grid2 we sum the number of events in the different jet multiplicity bins for 1 fb^{-1} (up to 5) and the different boson P_T bins. We sum the samples for which the number of events to be simulated and reconstructed is greater than 0.1 M events per sample (the samples for the high boson P_T bins are smaller samples and we propose to process them at the Tier-2s of the CSA07 exercise).

sample	ϵ_{match}	$N_{ev/fb}$	$N_{generate}$
z0j	0.82	4.41E+06	4.34E+07
z1j[$0 < P_T(Z) < 100$ GeV/c]	0.57	9.35E+05	5.26E+07
z2j[$0 < P_T(Z) < 100$ GeV/c]	0.39	2.71E+05	1.27E+08
z3j[$0 < P_T(Z) < 100$ GeV/c]	0.26	6.75E+04	2.12E+08
z4j[$0 < P_T(Z) < 100$ GeV/c]	0.16	1.39E+04	2.14E+08
z5j[$0 < P_T(Z) < 100$ GeV/c]	0.32	8.81E+03	2.14E+08

Table 19: Summary of expected and required Monte Carlo statistics for $Z \to \ell^+\ell^-$+ Njets ($\ell = e, \mu, \tau$), corresponding to 1 fb^{-1} of data collected at LHC, for $0 < P_T(Z) < 100$ GeV/c.

sample	ϵ_{match}	$N_{ev/fb}$	$N_{generate}$
z1j[$100 < P_T(Z) < 300$ GeV/c]	0.45	3.04E+04	2.76E+06
z2j[$100 < P_T(Z) < 300$ GeV/c]	0.33	2.83E+04	2.31E+07
z3j[$100 < P_T(Z) < 300$ GeV/c]	0.21	1.29E+04	5.90E+07
z4j[$100 < P_T(Z) < 300$ GeV/c]	0.13	4.28E+03	8.51E+07
z5j[$100 < P_T(Z) < 300$ GeV/c]	0.37	4.96E+03	1.24E+08

Table 20: Summary of expected and required Monte Carlo statistics for $Z \to \ell^+\ell^-$+ Njets ($\ell = e, \mu, \tau$), corresponding to 1 fb^{-1} of data collected at LHC, for $100 < P_T(Z) < 300$ GeV/c.

sample	ϵ_{match}	$N_{ev/fb}$	$N_{generate}$
z1j[$300 < P_T(Z) < 800$ GeV/c]	0.29	3.61E+02	5.79E+04
z2j[$300 < P_T(Z) < 800$ GeV/c]	0.22	5.47E+02	1.01E+06
z3j[$300 < P_T(Z) < 800$ GeV/c]	0.15	4.00E+02	3.39E+06
z4j[$300 < P_T(Z) < 800$ GeV/c]	0.11	2.01E+02	7.65E+06
z5j[$300 < P_T(Z) < 800$ GeV/c]	0.44	4.50E+02	1.25E+07

Table 21: Summary of expected and required Monte Carlo statistics for $Z \to \ell^+\ell^-$+ Njets ($\ell = e, \mu, \tau$), corresponding to 1 fb^{-1} of data collected at LHC, for $300 < P_T(Z) < 800$ GeV/c.

sample	ϵ_{match}	$N_{ev/fb}$	$N_{generate}$
z1j[$800 < P_T(Z) < 1600$ GeV/c]	0.20	2.13E+00	3.53E+02
z2j[$800 < P_T(Z) < 1600$ GeV/c]	0.15	3.97E+00	1.63E+04
z3j[$800 < P_T(Z) < 1600$ GeV/c]	0.11	3.79E+00	7.08E+04
z4j[$800 < P_T(Z) < 1600$ GeV/c]	0.09	2.47E+00	1.97E+05
z5j[$800 < P_T(Z) < 1600$ GeV/c]	0.43	7.36E+00	4.67E+05

Table 22: Summary of expected and required Monte Carlo statistics for $Z \to \ell^+\ell^-$+ Njets ($\ell = e, \mu, \tau$), corresponding to 1 fb^{-1} of data collected at LHC, for $800 < P_T(Z) < 1600$ GeV/c.

sample	ϵ_{match}	$N_{ev/fb}$	$N_{generate}$
z1j[$1600 < P_T(Z) < 3200$ GeV/c]	0.16	1.67E-02	7.62E+00
z2j[$1600 < P_T(Z) < 3200$ GeV/c]	0.13	3.95E-02	3.11E+02
z3j[$1600 < P_T(Z) < 3200$ GeV/c]	0.1	4.11E-02	1.79E+03
z4j[$1600 < P_T(Z) < 3200$ GeV/c]	0.08	2.79E-02	5.90E+03
z5j[$1600 < P_T(Z) < 3200$ GeV/c]	0.46	9.06E-02	9.29E+04

Table 23: Summary of expected and required Monte Carlo statistics for $Z \to \ell^+\ell^-$+ Njets ($\ell = e, \mu, \tau$), corresponding to 1 fb^{-1} of data collected at LHC, for $1600 < P_T(Z) < 3200$ GeV/c.

sample	ϵ_{match}	$N_{ev/fb}$	$N_{generate}$
z1j[$3200 < P_T(Z) < 5000$ GeV/c]	0.15	8.96E-06	2.43E-03
z2j[$3200 < P_T(Z) < 5000$ GeV/c]	0.14	2.16E-05	1.82E-01
z3j[$3200 < P_T(Z) < 5000$ GeV/c]	0.12	2.51E-05	1.48E+00
z4j[$3200 < P_T(Z) < 5000$ GeV/c]	0.09	1.31E-05	1.28E+01
z5j[$3200 < P_T(Z) < 5000$ GeV/c]	0.44	3.93E-05	7.5

Table 24: Summary of expected and required Monte Carlo statistics for $Z \to \ell^+\ell^-$+ Njets ($\ell = e, \mu, \tau$), corresponding to 1 fb^{-1} of data collected at LHC, for $3200 < P_T(Z) < 5000$ GeV/c.

sample	ϵ_{match}	$N_{ev/fb}$	$N_{generate}$
z0j	0.82	8.82E+06	8.68E+07
z1j[$0 < P_T(Z) < 100$ GeV/c]	0.57	1.87E+06	1.05E+08
z2j[$0 < P_T(Z) < 100$ GeV/c]	0.39	5.42E+05	2.54E+08
z3j[$0 < P_T(Z) < 100$ GeV/c]	0.26	1.35E+05	4.24E+08
z4j[$0 < P_T(Z) < 100$ GeV/c]	0.16	2.77E+04	4.29E+08
z5j[$0 < P_T(Z) < 100$ GeV/c]	0.32	1.76E+04	4.29E+08

Table 25: Summary of expected and required Monte Carlo statistics for $Z \rightarrow \nu_\ell \bar{\nu}_\ell$+ Njets ($\ell = e, \mu, \tau$), corresponding to 1 fb^{-1} of data collected at LHC, for $0 < P_T(Z) < 100$ GeV/c.

sample	ϵ_{match}	$N_{ev/fb}$	$N_{generate}$
z1j[$100 < P_T(Z) < 300$ GeV/c]	0.45	6.08E+04	5.52E+06
z2j[$100 < P_T(Z) < 300$ GeV/c]	0.33	5.65E+04	4.62E+07
z3j[$100 < P_T(Z) < 300$ GeV/c]	0.21	2.57E+04	1.18E+08
z4j[$100 < P_T(Z) < 300$ GeV/c]	0.13	8.56E+03	1.70E+08
z5j[$100 < P_T(Z) < 300$ GeV/c]	0.37	9.91E+03	2.49E+08

Table 26: Summary of expected and required Monte Carlo statistics for $Z \rightarrow \nu_\ell \bar{\nu}_\ell$+ Njets ($\ell = e, \mu, \tau$), corresponding to 1 fb^{-1} of data collected at LHC, for $100 < P_T(Z) < 300$ GeV/c.

sample	ϵ_{match}	$N_{ev/fb}$	$N_{generate}$
z1j[$300 < P_T(Z) < 800$ GeV/c]	0.29	7.22E+02	1.16E+05
z2j[$300 < P_T(Z) < 800$ GeV/c]	0.22	1.09E+03	2.02E+06
z3j[$300 < P_T(Z) < 800$ GeV/c]	0.15	8.00E+02	6.78E+06
z4j[$300 < P_T(Z) < 800$ GeV/c]	0.11	4.03E+02	1.53E+07
z5j[$300 < P_T(Z) < 800$ GeV/c]	0.44	9.00E+02	2.49E+07

Table 27: Summary of expected and required Monte Carlo statistics for $Z \rightarrow \nu_\ell \bar{\nu}_\ell$+ Njets ($\ell = e, \mu, \tau$), corresponding to 1 fb^{-1} of data collected at LHC, for $300 < P_T(Z) < 800$ GeV/c.

sample	ϵ_{match}	$N_{ev/fb}$	$N_{generate}$
z1j[$800 < P_T(Z) < 1600$ GeV/c]	0.20	4.25E+00	7.05E+02
z2j[$800 < P_T(Z) < 1600$ GeV/c]	0.15	7.95E+00	3.25E+04
z3j[$800 < P_T(Z) < 1600$ GeV/c]	0.11	7.57E+00	1.42E+05
z4j[$800 < P_T(Z) < 1600$ GeV/c]	0.09	4.95E+00	3.93E+05
z5j[$800 < P_T(Z) < 1600$ GeV/c]	0.43	1.47E+01	9.35E+05

Table 28: Summary of expected and required Monte Carlo statistics for $Z \to \nu_\ell \bar{\nu}_\ell$+ Njets ($\ell = e, \mu, \tau$), corresponding to 1 fb^{-1} of data collected at LHC, for $800 < P_T(Z) < 1600$ GeV/c.

sample	ϵ_{match}	$N_{ev/fb}$	$N_{generate}$
z1j[$1600 < P_T(Z) < 3200$ GeV/c]	0.16	3.34E-02	1.52E+01
z2j[$1600 < P_T(Z) < 3200$ GeV/c]	0.13	7.91E-02	6.22E+02
z3j[$1600 < P_T(Z) < 3200$ GeV/c]	0.1	8.21E-02	3.59E+03
z4j[$1600 < P_T(Z) < 3200$ GeV/c]	0.08	5.57E-02	1.18E+04
z5j[$1600 < P_T(Z) < 3200$ GeV/c]	0.46	1.81E-01	1.86E+05

Table 29: Summary of expected and required Monte Carlo statistics for $Z \to \nu_\ell \bar{\nu}_\ell$+ Njets ($\ell = e, \mu, \tau$), corresponding to 1 fb^{-1} of data collected at LHC, for $1600 < P_T(Z) < 3200$ GeV/c.

sample	ϵ_{match}	$N_{ev/fb}$	$N_{generate}$
z1j[$3200 < P_T(Z) < 5000$ GeV/c]	0.15	1.79E-05	4.86E-03
z2j[$3200 < P_T(Z) < 5000$ GeV/c]	0.14	4.32E-05	3.64E-01
z3j[$3200 < P_T(Z) < 5000$ GeV/c]	0.12	5.01E-05	2.97E+00
z4j[$3200 < P_T(Z) < 5000$ GeV/c]	0.09	2.62E-05	2.56E+01
z5j[$3200 < P_T(Z) < 5000$ GeV/c]	0.44	7.87E-05	1.50E+01

Table 30: Summary of expected and required Monte Carlo statistics for $Z \to \nu_\ell \bar{\nu}_\ell$+ Njets ($\ell = e, \mu, \tau$), corresponding to 1 fb^{-1} of data collected at LHC, for $3200 < P_T(Z) < 5000$ GeV/c.

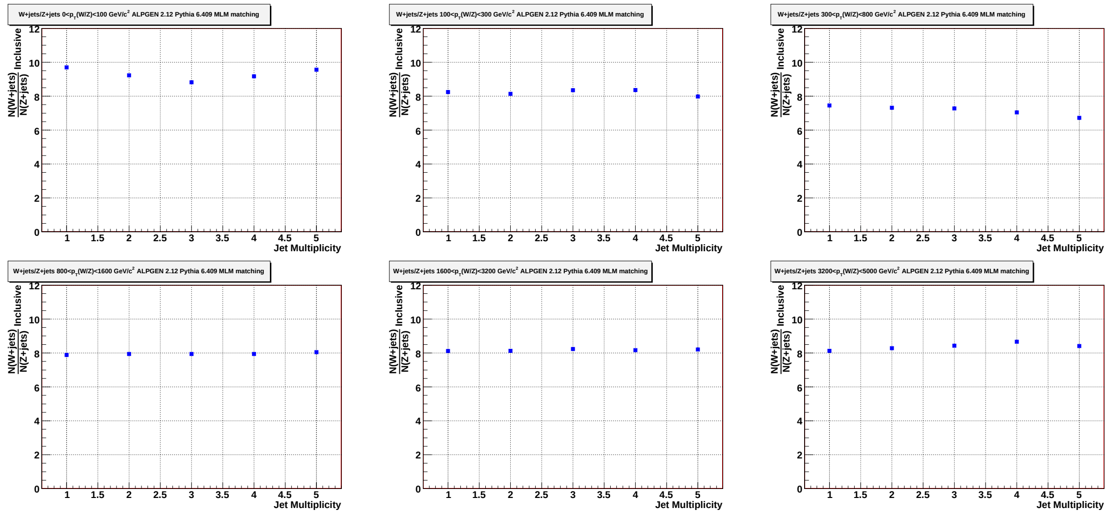

Figure 1: Ratio of inclusive $\sigma(W+Njets)/\sigma(Z+Njets)$ as a function of jet multiplicity, for different $P_T(W/Z)$ bins, after the matching procedure has been applied.

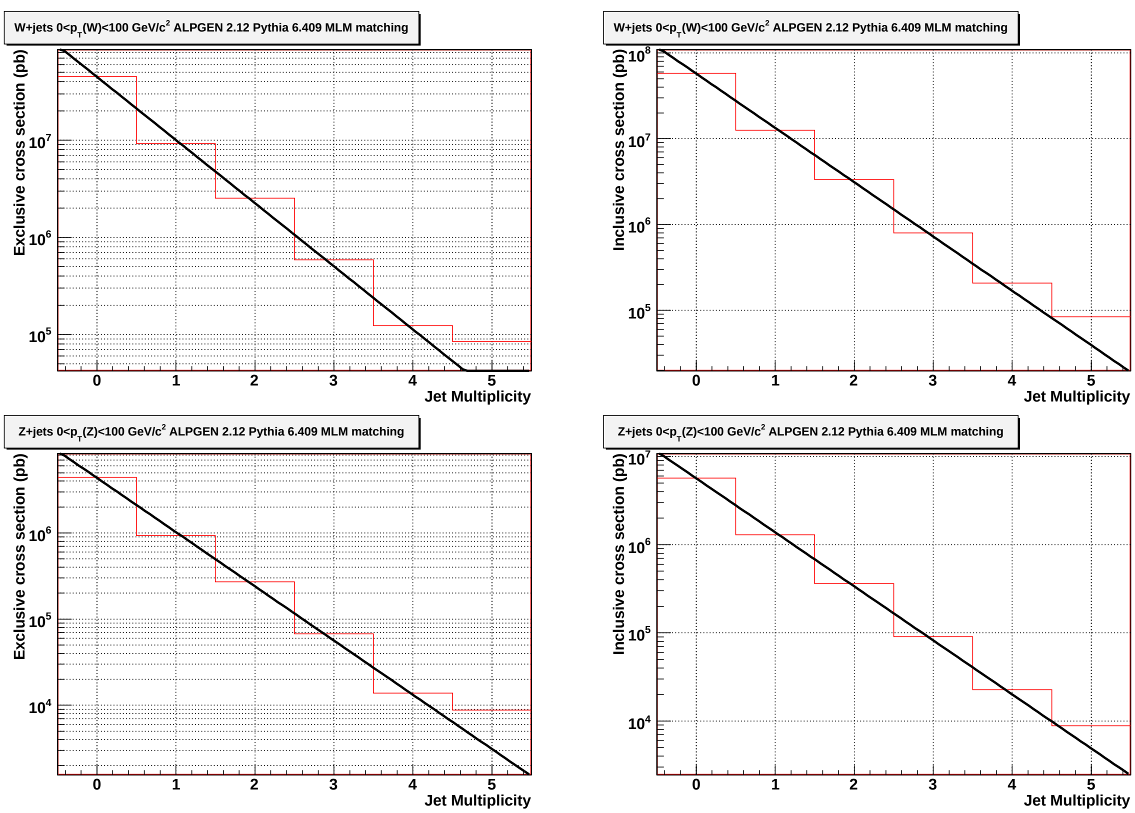

Figure 2: Exclusive (left) and inclusive (right) $\sigma(W+Njets)$ (top) and $\sigma(Z+Njets)$ (bottom) with $P_T(W/Z) < 100$ GeV/c, as a function of jet multiplicity, after the matching procedure has been applied. The slope of the fit gives $\sigma(N\ jets)/\sigma(N+1\ jets) \sim 4.3$.

To generate 1 fb^{-1} in the $W+0$ jets bin (we don't need to bin this production in boson P_T) with ALPGEN we need $\sim$ 45 M events. Correspondingly, the Z visible sample is 4.4 M and the invisible 8.8 M in the 0 jet bin. **We propose to include in the CSA07 production 100 pb^{-1} of W+0 and Z+0 jets ammounting to 4.5 M W's, 880 k invisible Z's and 440 k visible Z's**.

For 1 fb^{-1} the sum of W+jets datasets for jet multiplicities $\geq$1 is then 13 M events (for the 3 flavors table 13, 14). The corresponding Z+jets datasets is 1.2 M events to the charged leptons (table 19, jet multiplicities 1,2) and 2.6

W/Z+0jets events in 1 fb^{-1}		
Sample	Number of Events	Comment
W+0jets	45 M	all flavours
Z+0jets	4.4 M	all flavours visible
Z+0jets	8.8 M	all flavours invisible

Table 31: Number of expected events in 1 fb^{-1} for W/Z+0jets.

M for the invisible ones. This values are obtained summing all the datasets (for a given channel) containing more than 0.1 M events.

The smaller datasets (high boson P_T and all multiplicities) amount to

- $\sim$ 176 k (we can afford to scale them to 10 fb^{-1}, $\sim$ 1 M) for the W+jets
- $\sim$ 173 k Z+jets visible and $\sim$ 211 k invisible.

The final CSA07 W/Z+jets cocktail may consist of $\sim$22 M events as follows:

- 4.5 M W+0 jets (all flavors, 100 pb^{-1})
- of 900 k invisible Z's corresponding to 100 pb^{-1} we propose 100 k (for checks, matching, ISR etc)
- 450 k visible Z's (all flavors, 100 pb^{-1})
- 13 M W+$\geq$ 1jets (all flavors, binned in boson P_T and jet multiplicity, $\sim$ 1 fb^{-1})
- 1.2 M Z+$\geq$ 1jets (all flavors visible, binned in boson P_T and jet multiplicity, $\sim$ 1 fb^{-1})
- 2.6 M Z+$\geq$ 1jets (invisible, binned in boson P_T and jet multiplicity $\sim$ 1 fb^{-1})
- $\sim$ 0.5 M residual W/Z+jets (all flavors, visible/invisible, the high P_T bins and jet multiplicities that require small statistics $\sim$ 1 fb^{-1}).

Note that the figures above correspond to the LO cross section for these processes. Refer to the appendix for the $t\bar{t}$ production needs which could possibly be done in the Tier2s of the CSA07. However note that the .grid2s need to be recomputed with the latest version of ALPGEN.

6 Discussion

We studied the cross section of W/Z+jets events at the LHC, binned in boson p_T and jet multiplicity. We used the LO ALPGEN Monte Carlo generator to calculate the matrix element of the considered processes. The parton shower and the hadronization were provided by PYTHIA. We applied parton shower matching using the MLM scheme, and calculated the expected number of events in 1 fb^{-1} of proton-proton collisions at $\sqrt{s} = 14$ TeV per jet multiplicity bin in the range [1,5].

We found a 15% difference in cross section for low jet multiplicity and low P_T values with respect to the previous ALPGEN v2.05 results (see appendix), explained by the different scale value corresponding to the `iqopt=1` setting.

In addition, we observe an *increase* of the cross section with *increasing* jet multiplicity for p_T bins higher that the first. We understand this to be aan effect of the the threshold on the $P_T(W/Z)$ that corresponds to a similar cut on the total of the recoiling system. This in turns corresponds to a similar cut on the jet p_T if there is only one jet, while in general it produces a complicated cut on the phase space of the multiple jets event, depending on the momenta and the opening angles of the final state partons. The different portion of the phase space can compensate the decrease of the cross section with increasing multiplicity, generating the enhancement we observe

W/Z+$\geq$ 1jets events in 1 fb^{-1}		
Sample	Number of Events	Comment
W+$\geq$ 1jets	13 M	all flavours sum of samples with $\geq$ 0.1M each binned in jet multiplicity and boson P_T
Z+$\geq$ 1jets	1.2 M	all flavours visible sum of samples with $\geq$ 0.1M each binned in jet multiplicity and boson P_T
Z+$\geq$ 1jets	2.6 M	all flavours invisible sum of samples with $\geq$ 0.1M each binned in jet multiplicity and boson P_T

Table 32: Number of expected events in 1 fb^{-1} for W/Z+$\geq$1jets.

W/Z+jets events in 1 fb^{-1}		
Sample	Number of Events	Comment
W+jets	176 K	all flavours sum of samples with small statistics binned in jet multiplicity and boson P_T
Z+jets	173K	all flavours visible sum of samples with small statistics binned in jet multiplicity and boson P_T
Z+jets	211 K	all flavours invisible sum of samples with small statistics binned in jet multiplicity and boson P_T

Table 33: Number of expected events in 1 fb^{-1} for W/Z+jets (low statistics bins).

for Njets=2,3. This creates a bias for the W/Z ratio at the high P_T bins. The ratio $W\ell\nu$+jets/$Z/\gamma\ell\ell$+jets is slighly lower than 10 even in the low P_T bin, as expected (see section 2.4). We have checked that it is indeed between 10 and 11 when the mass window of the $\ell\ell$ system is around the Z-pole.

From these results, we deduce the number of events to be asked before unweighing in order to obtain the Monte Carlo statistics (after parton shower matching) corresponding to $1\ fb^{-1}$. We present a proposal for the $W/Z+$ jets [0,5] cocktail composition that amounts to a total of ~ 23 M events.

sample	warmup PS-Grid	N_{events}	N_{wgt}	σ_{wgt} (pb)	N_{unwgt}	σ_{unwgt} (pb)	$N_{matched}$	matched (pb)	ϵ_{match}
w0j	2M,3	6M	5.9M	1.3E4	2.1M	1.3E4	1.6M	1.03E4	0.8
w1j	1M,3	1M	0.5M	0.4E4	0.1M	0.4E4	0.06M	0.25E4	0.6
w2j	1M,6	5M	76k	0.2E4	26k	0.2E4	10k	0.07E4	0.4
w3j	1M,7	25M	0.1M	0.1E4	27k	785	7k	195	0.25
w4j	1M,8	25M	70k	557	8.4k	340	1.2k	51	0.15
w5j	2M,9	35M	97k	260	5.2k	150	0.5k	14	0.09
w6j	1M,9	50M	36k	119	1.5k	70	0.5k	23	0.3

Table 34: $\sigma(e\nu_e + N \text{ jets})$ at the LHC.

sample	warmup PS-Grid	N_{events}	N_{wgt}	σ_{wgt} (pb)	N_{unwgt}	σ_{unwgt} (pb)	$N_{matched}$	matched (pb)	ϵ_{match}
z0j	1M,4	2M	0.56M	980	90k	970	70k	747	0.76
z1j	1M,4	2M	0.37M	360	60k	363	34k	209	0.6
z2j	1M,6	10M	194k	193	47k	174	18k	66	0.37
z3j	2M,7	50M	0.13M	102	27k	73	7k	17	0.24
z4j	1M,8	40M	0.13M	52	15k	31	2k	4.5	0.14
z5j	3M,8	50	74k	25	5.2k	14.6	0.5k	1.3	0.09
z6j *	3M,9	50M	50k	1.1	4k	6.7	1.3K	2.2	0.3

Table 35: $\sigma(e^+e^- + N \text{ jets})$ at the LHC.

A APPENDIX: ALPGEN v2.05 production

In Tab. 34 (Tab. 35) we report the result of a similar study of W/Z+jets samples, based on DCO4 (ALPGEN v2.05) generation.

The main differences, with respect to the study presented in this document, are:

- A lower cut on $p_T > 15$ GeV/c for jets was applied, instead of $p_T > 20$ GeV/c of the v2.12 version.
- Even if `iqopt=1` was used also in this case, the value of the renormalization scale was `qfac*sqrtm_W/Z^2+ pt_W/Z^2` rather than `qfac*sqrtm_W/Z^2+ sum(pt_jet^2)`, because of a swap in the definition of `iqopt=1` and `iqopt=3` options in that ALPGEN version.

In Fig. 3, we show the exclusive and inclusive cross section, as a function of jet multiplicity, for W+jets and Z+jets samples. We obtained $\sigma(Njets)/\sigma(N+1jets) \sim 3.8$ (~ 3.5) for $\sigma(W+Njets)$ ($\sigma(Z+Njets)$) events. .

A binning in p_T was performed on the DC04 sample, produced using PYTHIA. The binning was performed according to the following requirements on $\hat{p}_T$:

- $0 < \hat{p}_T < 40$ (400K events)
- $10 < \hat{p}_T < 100$ (600K events)
- $25 < \hat{p}_T < 170$ (800K events)
- $42.5 < \hat{p}_T < 300$ (600K events)

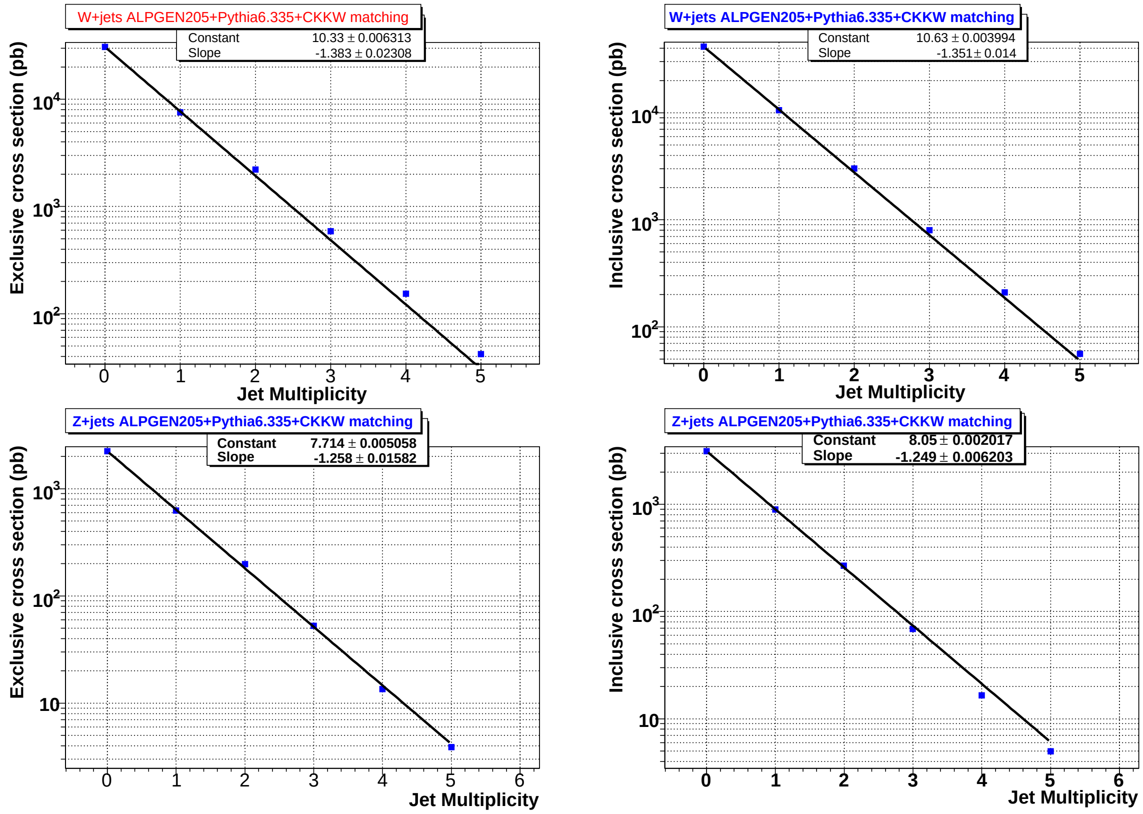

Figure 3: Exclusive (left) and inclusive (right) $\sigma(W+N jets)$ (top) and $\sigma(Z+N jets)$ (bottom) with $p_T(W/Z) < 100$ GeV/c, as a function of jet multiplicity, from ALPGEN v2.05 production. The slope of the fit gives $\sigma(N\ jets)/\sigma(N+1\ jets) \sim 3.8$ ($\sigma(N\ jets)/\sigma(N+1\ jets) \sim 3.$) for $\sigma(W+N jets)$ ($\sigma(Z+N jets)$) events.

- $75 < \hat{p}_T < 500$ (300K events)
- $125 < \hat{p}_T < 800$ (200K events)
- $200 < \hat{p}_T < 1400$ (100K events)
- $350 < \hat{p}_T < 2200$ (80K events)
- $550 < \hat{p}_T < 3200$ (50K events)
- $800 < \hat{p}_T < 4400$ (20K events)

corresponding to a luminosity of $1\ fb^{-1}$ ($15\ fb^{-1}$) for low $\hat{p}_T$ (high $\hat{p}_T$) bins.

ALPGEN v2.05 setting has also been used to generate $t\bar{t}$ events. The values of cross sections for the various steps of ALPGEN simulation and of the matching efficiency are given in Tab. 36.

Using these numbers, one can calculate the required yields to ask in generation to obtain a statistics comparable to $1\ fb^{-1}$ recorded at $\sqrt{s} = 14TeV$. In particular, knowing the luminosity $\mathcal{L}$, the number of matched events N_{match}, the number of events asked N_{events}, and the cross section after matching σ_{match}, one can write

$$N^{\mathcal{L}}_{events} = \frac{N_{events}}{N_{match}} \times \mathcal{L} \times \sigma_{match} \tag{A-1}$$

Applying this expression to the numbers of Tab. 36, we obtain the following number of events to generate for the required statistics at 1 fb^{-1} :

- ask for 18.6 M of tt+0jets for a yield of 146 K
- ask for 148 M for tt+1jets for a yield of 148 K

sample	warmup PS-Grid	N_{events}	N_{wgt}	σ_{wgt} (pb)	N_{unwgt}	σ_{unwgt} (pb)	$N_{matched}$	matched (pb)	ϵ_{match}
tt0j	1M,3	7M	0.6M	487	114k	214	55k	73	0.4
tt1j	1M,4	7M	0.32M	1112	26k	297	7k	74	0.25
tt2j	1M,4	20M	65k	1212	3k	237	519	43	0.18
tt3j	1M,6	30M	0.2M	996	7k	143	800	16	0.11
tt4j	1M,8	70M	0.13M	685	7k	70	500	5.2	0.07
tt5j	1M,7	50M	85k	400	1.5k	29	76	1.5	0.05

Table 36: $\sigma(t\bar{t} + N \text{ jets}) foritdecmod = 6(allhadronic)$ at the LHC.

- 332 M for tt+2jets for a yield of 86 K
- 1200 M for tt+3jets for a yield of 32 K
- 1456 M for tt+4jets for a yield of 10.4 K
- 1974 M for tt+5jets for a yield of 3 K

The total projected number of $t\bar{t}$ events to be simulated and recontsructed for 1 fb^{-1} and up to 5 jets is then $\sim$ 450,000 events (in datasets that could in priciple run in the Tier-2s for CSA07). The study above will be redone with the latest version of ALPGEN and CMSSW, but the yields are not expected to vary significantly. Note that this calculation is based on the LO cross section. If we want to be ready for the data (all-orders included), a factor of $\sim$2 needs to be applied giving approximately 1 M events for 1 fb^{-1} [10].

Some thought needs to put in the range between 1 and 10 fb^{-1} production as then the statistics available will be enough to use the top as a standard model candle and the interest will turn into binned production in the P_T of the top.

References

[1] M. L. Mangano, M. Moretti, F. Piccinini, R. Pittau and A. D. Polosa, JHEP **0307** (2003) 001 [arXiv:hep-ph/0206293].

[2] F. Caravaglios and M. Moretti, Phys. Lett. B **358** (1995) 332 [arXiv:hep-ph/9507237].

[3] T. Sjostrand, S. Mrenna and P. Skands, JHEP **0605** (2006) 026 [arXiv:hep-ph/0603175].

[4] M. L. Mangano, in preparation. See also M. L. Mangano, M. Moretti, F. Piccinini and M. Treccani, JHEP **0701** (2007) 013 [arXiv:hep-ph/0611129].

[5] S. Hoche, F. Krauss, N. Lavesson, L. Lonnblad, M. Mangano, A. Schalicke and S. Schumann, arXiv:hep-ph/0602031.

[6] S. Catani, F. Krauss, R. Kuhn and B. R. Webber, JHEP **0111** (2001) 063 [arXiv:hep-ph/0109231]. F. Krauss, JHEP **0208** (2002) 015 [arXiv:hep-ph/0205283].

[7] A. Schalicke and F. Krauss, JHEP **0507** (2005) 018 [arXiv:hep-ph/0503281]. T. Gleisberg, S. Hoche, F. Krauss, A. Schalicke, S. Schumann and J. C. Winter, JHEP **0402** (2004) 056 [arXiv:hep-ph/0311263].

[8] L. Lonnblad, Comput. Phys. Commun. **71** (1992) 15.

[9] J. M. Campbell and R. K. Ellis, Phys. Rev. D **62** (2000) 114012 [arXiv:hep-ph/0006304].

[10] N. Kidonakis and R. Vogt, Phys. Rev. D **68** (2003) 114014 [arXiv:hep-ph/0308222].