Dataset description

Melissa van Beekveld

July 24, 2020

We generate LHC events corresponding to a center-of-mass energy of 13 TeV. Events for the background and signal processes are generated at leading order (LO) with up to two additional partons in the matrix element using the event generator MG5_aMC@NLO v6.3.2 (Madgraph) and versions above with the NNPDF PDF set using 5 flavors in the definition of the proton. Madgraph is interfaced to Pythia 8.2, that handles the showering of the matrix element level generated events. The matching with the parton shower, needed in the case when one or more additional jets are generated in Madgraph, is done using the MLM merging prescription. Then, a quick detector simulation is performed with Delphes 3, using a modified version of the ATLAS detector card. Pileup is not included in this dataset. A repository of the data scripts that are used to generate the events is on GitHub [1].
The final state objects, as described in Table 1, are stored in a one-line-per-event text (csv) file, where each line has variable length and contains 3 event-specifiers, followed by the kinematic features for each object in the event. The format of CSV files is:

`event ID; process ID; event weight; MET; METphi; obj1, E1, pt1, eta1, phi1; obj2, E2, pt2, eta2, phi2; ...`

The `event ID` is an event specifier. It is an integer to identify the generation of that particular event, included for debugging purposes only. The `process ID` is a string referring to the process that generated the event (see the Les Houches BSM proceedings). The event weight w is defined as 1 for all *background* SM events, thereby creating a weighted SM data set. To recover the cross section (after channel cuts), one takes

$$\sigma_{\text{after cuts}} = \frac{N_{\text{lines}}}{\mathcal{L}}, \tag{1}$$

where the luminosity $\mathcal{L}$ (in fb^{-1}) is given in the name of the `background.csv` file, i.e.

`background_chan1_7.79.csv`

is the background file belonging to channel 1 with $\mathcal{L} = 7.79$ fb^{-1}.
For the *signal* events, the weight is defined as

$$w = \frac{\sigma_{\text{after cuts}}}{N_{\text{lines}}}. \tag{2}$$

Probably, the weights do not have to be used. If you decide to include the weights and are unsure how to do it, pop in a question to the Slack channel.

[1] `https://github.com/melli1992/unsupervised_darkmachines`

An event consists of a variable number of objects. An event is stored when at least one of the following requirements are fulfilled:

- At least one (b)-jet with transverse momentum $p_T > 60$ GeV and pseudorapidity $|\eta| < 2.8$, or
- at least one electron with $p_T > 25$ GeV and $|\eta| < 2.47$, except for $1.37 < |\eta| < 1.52$, or
- at least one muon with $p_T > 25$ GeV and $|\eta| < 2.7$, or
- at least one photon with $p_T > 25$ GeV and $|\eta| < 2.37$.

Of course, these are unrealistic trigger requirements, but we aim to create a flexible data set that allows for different types of studies that might need different selection criteria. The η-restriction on the electrons models a veto in the crack regions as often applied in ATLAS analyses. Such a veto can also be applied to photons by the user. Note that for the processes with the largest cross sections ($W^{\pm}/\gamma/Z$ + jets and QCD jet production) we have applied cuts on $H_T > 100$ GeV and 600 GeV respectively to make the data generation manageable. The observable H_T is defined as the scalar sum of the transverse momenta of all jets (with $p_{T,j_i} > 20$ GeV and $|\eta_{j_i}| < 2.8$):

$$H_T = \sum_i |p_{T,j_i}|. \tag{3}$$

Therefore, if one includes any of these processes in their analysis, one must make sure that the same cuts are also applied to the the other processes, which is the reason that we apply this cut in some of the channels. The requirements on the final states objects that are stored are

- (b-)jet: $p_T > 20$ GeV and $|\eta| < 2.8$,
- electron/muon: $p_T > 15$ GeV and $|\eta| < 2.7$,
- photon: $p_T > 20$ GeV and $|\eta| < 2.37$.

This means that, for example, a jet with $p_T = 10$ GeV is not included in the dataset. The detector simulation as performed by Delphes removes any electrons with $|\eta| > 2.5$, as the reconstruction efficiency is set to 0 beyond that point.
The scale choice is set dynamically by Madgraph during the event generation. The resulting cross

Symbol ID	**Object**
`j`	jet
`b`	b-jet
`e-`	electron (e^-)
`e+`	positron (e^+)
`m-`	muon (μ^-)
`m+`	antimuon (μ^+)
`g`	photon (γ)

Table 1: Definition of symbols used for final-state objects. Only b-quark jets are tagged, no τ- or c-jets have been defined.

sections are not reweighted with any of the available higher-order and/or resummed cross sections. All relevant SM (background) processes that have been generated are summarized in the Les Houches BSM proceedings.

The channels are defined as follows:

- Channel 1
  - $\not{E}_T \geq 200$ GeV
  - $\not{E}_T / H_T \geq 0.2$ (H_T as in Eq. (3))
  - $N(j_{p_T>50\text{ GeV}}) \geq 4$ (j can either be a jet or a b-jet)
  - $N(j_{p_T>200\text{ GeV}}) \geq 1$
  - $H_T \geq 600$ GeV
- Channel 2a
  - $\not{E}_T \geq 50$ GeV
  - $N(l_{p_T>15\text{ GeV}}) \geq 3$
- Channel 2b
  - $\not{E}_T \geq 50$ GeV
  - $N(l_{p_T>15\text{ GeV}}) \geq 2$ ($l = \mu^{\pm}, e^{\pm}$)
  - $H_T \geq 50$ GeV
- Channel 3
  - $\not{E}_T \geq 100$ GeV
  - $H_T \geq 600$ GeV

We have created the following signals:

- Channel 1
  - Z' (2 TeV) + monojet (`monojet_Zp2000.0_DM_50.0_chan1.csv`), Z' decays fully invisible
  - Z' (200 GeV) + single (anti-)top (`monotop_200_A_chan1.csv`), Z' decays fully invisible
  - SUSY R-parity violating stop stop (1 TeV) production (`stlp_st1000_chan1.csv`), stop decay into leptons and b-quarks
  - SUSY R-parity violating squark-squark production (`sqsq1_sq1400_neut800_chan1.csv` 1.4 TeV squark and 800 GeV neutralino), squarks decaying to jets
  - SUSY gluino-gluino production (`glgl1400_neutralino1100_chan1.csv` 1.4 TeV gluino and 1.1 TeV neutralino, `glgl1600_neutralino800_chan1.csv` 1.6 TeV gluino and 800 GeV neutralino), gluinos decay to jets and MET
  - SUSY stop-stop (1 TeV) production (`stop2b1000_neutralino300_chan1.csv`), stops decaying to a top quark and a neutralino (300 GeV)

  - SUSY squark-squark production (`sqsq_sq1800_neut800_chan1.csv` 1.8 TeV squark and 800 GeV neutralino), squarks decaying to jets and MET

- Channel 2a
  - Z' (50 GeV) in lepton violating $U(1)_{L_\mu - L_\tau}$ (`pp23mt_50_chan2a.csv` with three leptons in final state, `pp24mt_50_chan2a.csv` with 4 leptons in final state), Z' decays to leptons and neutrinos (MET)
  - SUSY gluino-gluino production (`gluino_1000.0_neutralino_1.0_chan2a.csv`), gluino decaying to a top–anti-top-quark pair and MET
  - SUSY chargino-neutralino(2) production (`chaneut_cha400_neut200_chan2a.csv`, `chaneut_cha200_neut50_chan2a.csv`, `chaneut_cha250_neut150_chan2a.csv`, `chaneut_cha300_neut100_chan2a.csv`, chargino and neutralino2 masses as in file names)

- Channel 2b
  - Z' (50 GeV) in lepton violating $U(1)_{L_\mu - L_\tau}$ (`pp23mt_50_chan2b.csv` with three leptons in final state, `pp24mt_50_chan2b.csv` with 4 leptons in final state), Z' decays to leptons and neutrinos (MET)
  - SUSY R-parity violating stop stop (1 TeV) production (`stlp_st1000_chan2b.csv`), stop decay into leptons and b-quarks
  - SUSY chargino-chargino production (`chacha_cha300_neut140_chan2b.csv`, `chacha_cha400_neut60_chan2b.csv`, `chacha_cha600_neut200_chan2b.csv`, chargino and neutralino1 masses as in file name), chargino decays into leptons and MET
  - SUSY gluino-gluino (1 TeV) production (`gluino_1000.0_neutralino_1.0_chan2b.csv`), gluino decaying to a top–anti-top-quark pair and MET
  - SUSY chargino-neutralino(2) production (`chaneut_cha200_neut50_chan2b.csv`, `chaneut_cha250_neut150_chan2b.csv`, chargino and neutralino2 masses as in file names)

- Channel 3
  - Z' (2 TeV) + monojet (`monojet_Zp2000.0_DM_50.0_chan3.csv`), Z' decays fully invisible
  - Z' (2 TeV) + monoV (`monoV_Zp2000.0_DM_50.0_chan3.csv`), Z' decays fully invisible
  - Z' (200 GeV) + single (anti-)top (`monotop_200_A_chan3.csv`), Z' decays fully invisible
  - SUSY R-parity violating stop stop (1 TeV) production (`stlp_st1000_chan3.csv`), stop decay into leptons and b-quarks
  - SUSY R-parity violating squark-squark production (`sqsq1_sq1400_neut800_chan3.csv` 1.4 TeV squark and 800 GeV neutralino), squarks decaying to jets
  - SUSY gluino-gluino production (`glgl1400_neutralino1100_chan3.csv` 1.4 TeV gluino and 1.1 TeV neutralino, `glgl1600_neutralino800_chan3.csv` 1.6 TeV gluino and 800 GeV neutralino), gluinos decay to jets and MET
  - SUSY gluino-gluino (1 TeV) production (`gluino_1000.0_neutralino_1.0_chan3.csv`), gluino decaying to a top–anti-top-quark pair and MET
  - SUSY stop-stop (1 TeV) production (`stop2b1000_neutralino300_chan3.csv`), stops decaying to a top quark and a neutralino (300 GeV)

– SUSY squark-squark production (`sqsq_sq1800_neut800_chan3.csv` 1.8 TeV squark and 800 GeV neutralino), squarks decaying to jets and MET