ATLAS Note

ATL-COM-OREACH-2019-002

19th June 2019

For internal circulation only

ATL-OREACH-INT-2019-001
19 June 2019

Proposal for an ATLAS endorsed 13 TeV dataset for outreach purposes

Meirin Oan Evans[a], Even Simonsen Haaland[b], Arturo Sanchez Pineda[c], Kate Shaw[a], Farid Ould-Saada[b], Leonid Serkin[c], Jean-Francois Arguin[d], Susana Cabrera Urban[e], Dimitris Fassouliotis[f], Christine Kourkoumelis[f], Stelios Vourakis[f], Terrence Wyatt[g], Magnar Kopangen Bugge[b], Eirik Gramstad[b], Sascha Mehlhase[h], Agni Bethani[g], Thomas James Stevenson[a]

[a] *University of Sussex (GB)*
[b] *University of Oslo (NO)*
[c] *INFN Gruppo Collegato di Udine and ICTP, Trieste (IT)*
[d] *Université de Montréal (CA)*
[e] *University of Valencia and CSIC (ES)*
[f] *National and Kapodistrian University of Athens (GR)*
[g] *University of Manchester (GB)*
[h] *LMU Munich (DE)*

This document describes the motivations and educational aims of a new ATLAS Open Data release of datasets. The plan is to release a set of ROOT ntuple files for real data recorded by the ATLAS detector at a centre-of-mass energy of 13 TeV, plus an extensive collection of Monte Carlo datasets. The note outlines the design, content and internal structure of the ntuple files. Also outlined are the different educational use cases that are considered to decide which datasets to produce and which variables to include. This note is intended as a proposal to ATLAS members of the dataset to be released, and also as documentation for ATLAS members that then go on to use the dataset. The note follows a deep review of the different feedback given by ATLAS and non-ATLAS members that have been using the previous release for several years.

Contents

List of Figures

List of Tables

1 Introduction

The ATLAS Collaboration, as well as most other high-energy physics collaborations and experiments, rely on public funding to support their scientific programmes. Even more importantly, their continuity depends on a constant integration and replacement of human power – students and professionals that need to be trained, usually by senior members of those collaborations. We can call it “Knowledge Transfer", and it goes in both ways: to reattribute to society and to keep running the – even larger – scientific endeavours worldwide. Moreover, for that knowledge transfer to happen between members and non-members of the collaboration it is necessary to have real public datasets, tools and documentation.

An important part of ATLAS' current Open Access model regarding recorded and simulated data is to release datasets with a focus on education, training and outreach [1]. This mandate supports the creation of multiple platforms, projects, software and educational products that are used all over the planet. Examples of these platforms are the ATLAS and CERN Open Data websites [2, 3].

With that approach in mind, ATLAS has already released trillions of simulated and recorded collision events at a centre-of-mass energy of 8 TeV in 2016 [4–6], together with analysis tools and web-based documentation on the websites mentioned above.

Following the 2016 release (that included 1 fb^{-1} of real data), ATLAS members as well as external collaborators and users reported a range of activities [7]. This note documents the overall design, physics object content and selection, example analysis proposals and collections of samples for the next ATLAS Open Data release.

The new set of samples is based on ATLAS data and Monte Carlo (MC) datasets recorded and produced at a centre-of-mass energy of 13 TeV. This release would be the first 13 TeV public dataset released by an LHC experiment. The format of the new samples follows the 8 TeV samples. Amongst other reasons, this keeps backward compatibility of the documentation, software and tools developed by ATLAS, external partners and users. The samples are ROOT [8] files containing TTree objects with almost a hundred variables.

The design and content of the TTree object is based on input from multiple collaboration members. These include researchers and educators that are using the 8 TeV samples, and others who want to explore new academic programmes or educational exercises for their students using real LHC data. Other formats are under consideration, but they will be simple translations of the aforementioned ROOT TTree into tabular formats, when possible and relevant.

There is a significant number of use cases for this soon-to-come release of 10 fb^{-1} of real data, hundreds of fb^{-1} more in MC and resources. These data, MC and resources will be invaluable for the ATLAS Collaboration, as well as any other people that would like to explore the world of experimental particle physics and the computer science behind its data analysis.

The following three figures give an idea of the most relevant changes in the number of samples and the content of 13 TeV datasets for this 2019 release proposal with respect to the 2016 release at 8 TeV:

- Figure 1 shows the evolution of the Open Data datasets from the 8 TeV release (2016) to the 13 TeV release (2019).
- Figure 2 shows the evolution of the Open Data Higgs and BSM MC signals from the 8 TeV release (2016) to the 13 TeV release (2019).
- Figure 3 shows the evolution of the Open Data ntuple TTree structure from the 8 TeV release (2016) to the 13 TeV release (2019).

2 Proposed real and simulated datasets

2.1 Dataset layout

The dataset to be published needs to meet several demands. It has to enable institutes with diverse needs to implement their desired laboratory courses and present an interesting dataset for students outside ATLAS.

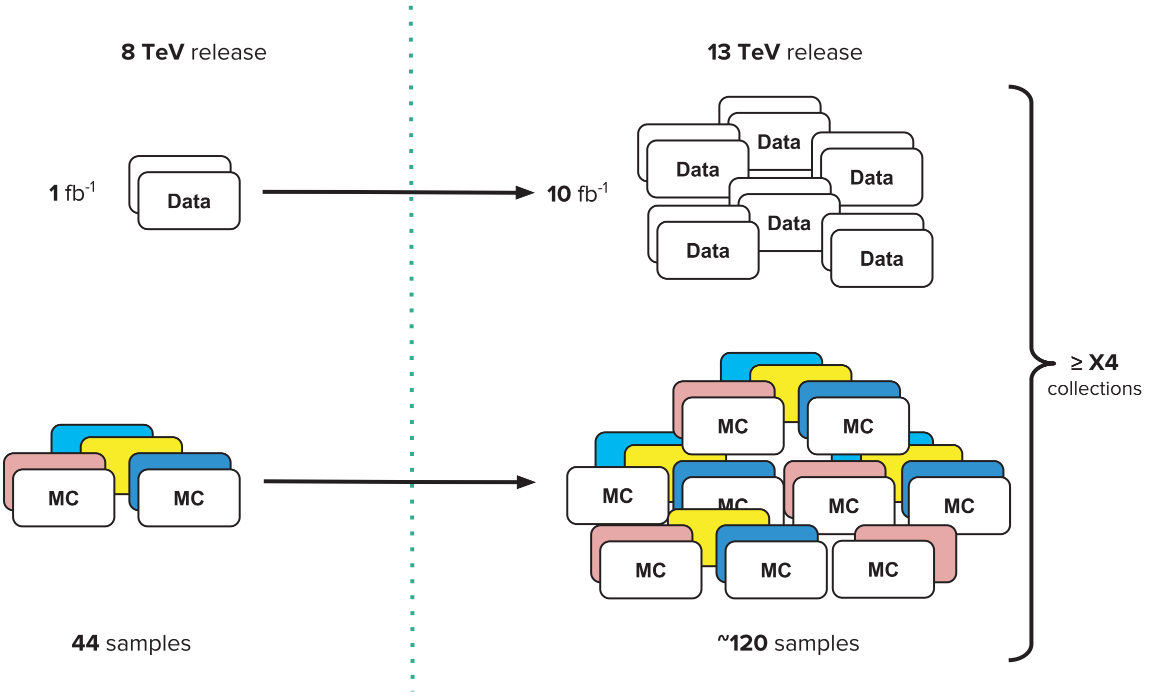

Figure 1: Evolution of the Open Data datasets from the 8 TeV release (2016) to the 13 TeV release (2019)

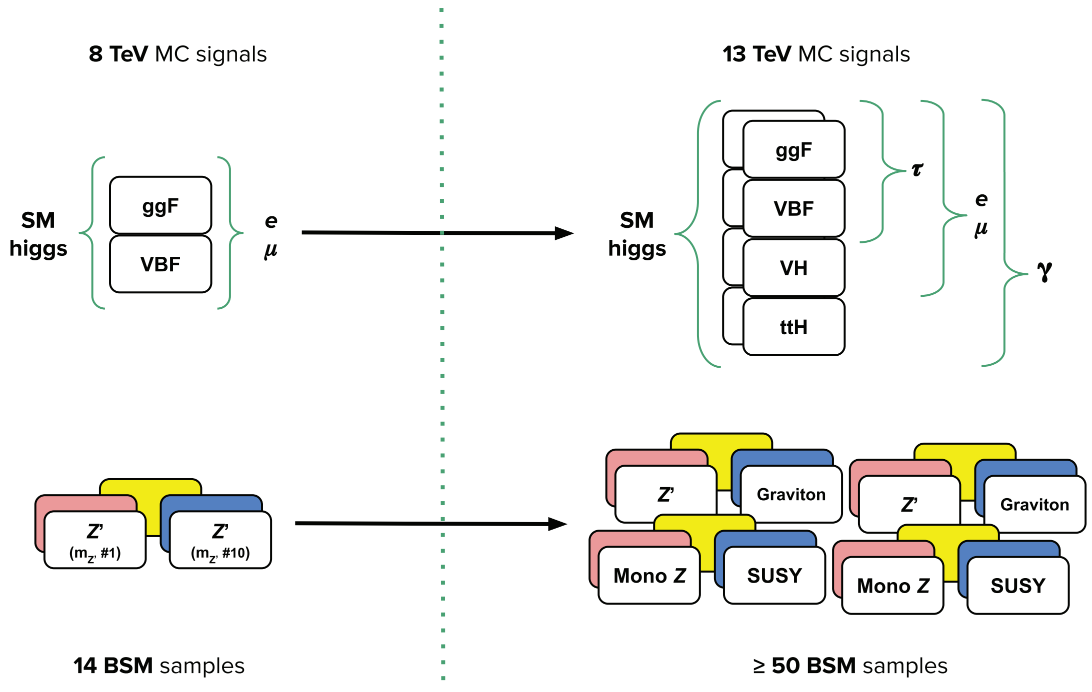

Figure 2: Evolution of the Open Data Higgs and Beyond Standard Model (BSM) MC signals from the 8 TeV release (2016) to the 13 TeV release (2019)

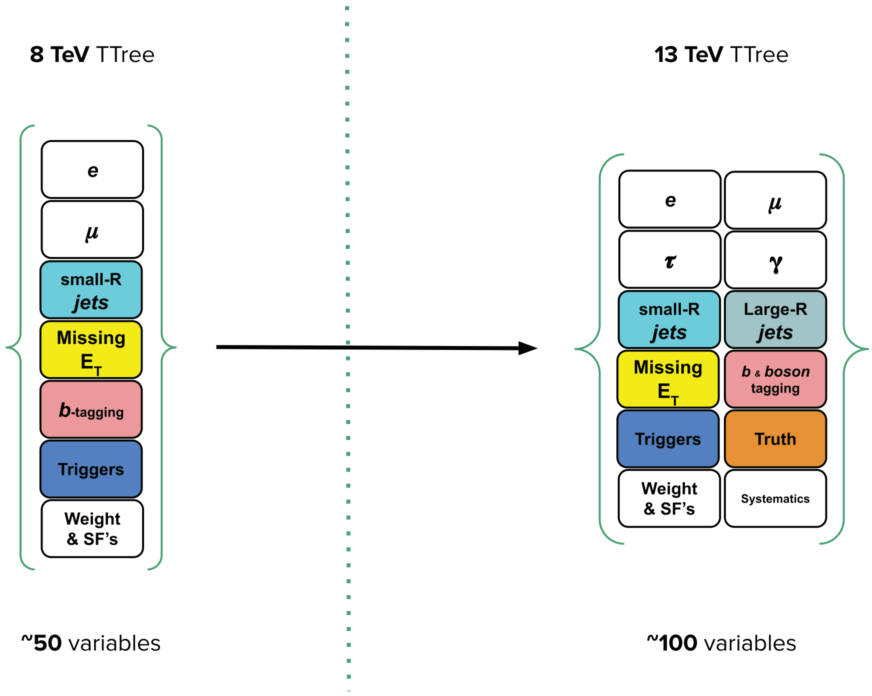

Figure 3: Evolution of the Open Data ntuple TTree structure from the 8 TeV release (2016) to the 13 TeV release (2019)

At the same time it should still be analysable on older commodity hardware as one wants to have a low technical entry barrier. Furthermore, the rules laid down in the ATLAS Data Access Policy [1] have to be honoured. Compromising between these demands affects the size and content of the dataset.

It is proposed to set the size of the dataset to about 10 fb^{-1}. Laboratory courses doing searches may have enough statistics to exclude models with strong signals in a meaningful phase space. Furthermore, Standard Model analyses such as a measurement of $H \to ZZ \to \ell\ell\ell\ell$ will have suitable statistics for relatively precise results.

The information content includes only basic information: four vectors, quality flags, an estimate for systematics, some basic event information and truth information for some datasets. Furthermore, the published dataset will not contain information for precise systematics treatment or serious data-driven methods like the Matrix Method [9]. Such a lean information layout will help the students retain a clear view of the data, and will help them to focus on the educational targets at hand, rather than being swamped by technical details. It is also expected that such a dataset will be processed faster on commodity hardware.

Currently, the ideal layout for the dataset is a modified version of the ntuple produced by the $HH \to bb\tau\tau$ version of the CxAODFramework [10, 11], using the “Goldcrest" $HH \to bb\tau\tau$ CxAOD production tag [12]. This framework sets up AnalysisBase release 2.4.28. The ntuple layout is defined in Tables 1 and 2. Electrons and muons are combined into a lepton collection and distinguished using lep_type to ensure as much continuity as possible for users that have experienced 8 TeV ATLAS Open Data. This also emphasises how electrons and muons are different to hadronic taus (which are new since 8 TeV ATLAS Open Data), by having separate tau branches.

By employing a stable release of the $HH \to bb\tau\tau$ version of the CxAODFramework, one would have the advantage of having a documented and well tested foundation for the definition of the dataset.

2.2 Intended use

This note is intended as a proposal to ATLAS members of the dataset to be released, and also as documentation for ATLAS members that then go on to use the dataset. These public data will be for educational use only. They do not qualify to reproduce actual ATLAS analysis results or produce publishable results. The fact that full systematics will not be provided, among other simplifications, restricts people from publishing actual results. The limited information provided sets the desired bounds of usage.

2.3 Capabilities and limitations of the proposed dataset

The provided content of the proposed datasets translates directly to the type of physics analyses that are feasible. It may support studies such as selection optimisation studies, searches, and measurements such as $H \to ZZ \to \ell\ell\ell\ell$. In addition it may be used as a testbed for new data-analysis techniques external to ATLAS, e.g. kinematic fits. To highlight in a public way the new analysis techniques people have created, we intend to invite users to add content to the atlas-outreach-data-tools GitHub repository [14].

The proposed datasets can be used for educational and physics purposes with different levels of task difficulty. At a beginner level, one could visualise the content of the datasets and produce simple distributions. An intermediate-level task would consist of making histograms with collision data after some basic selection, while advanced-level tasks would allow for a deeper look into the ATLAS data, with possibilities of measuring real event properties and physical quantities.

A non-exhaustive list of possible tasks with the proposed datasets include: comparisons of several distributions of event variables for simulated signal and background events; finding variables that are able to separate signal from background (jet multiplicity, transverse momenta of jets and leptons, lepton isolation, b-tagging, missing transverse energy, angular distributions); development and modification of cuts on these variables in order to enrich the signal over background separation; optimisation of the signal-over-background ratio and estimation of the purity based on simulation only; comparisons of the selection efficiency between data and simulation.

Advanced-level tasks might include: reconstruction of the truth objects (quarks or bosons) by assigning the detector objects (jets, leptons, missing energy) to the hypothetical decay trees; estimation of the impact of several sources of systematic uncertainties (luminosity uncertainty, b-tagging efficiency, trigger efficiency, uncertainty on background events) by adding approximate and conservative values; derivation of production cross sections and masses of objects.

The content does not support the creation of unfolded distributions, measuring efficiency corrected cross sections, and searches for signals that are not supported by signal datasets made available by ATLAS.

Only limited truth object information is provided, in order to facilitate the derivation of detector responses. Some experimental uncertainties are added in the datasets to provide ground for systematic-uncertainty estimation studies. With the aim of facilitating particle-level studies that do not require detailed knowledge of the detector, high-level object information and limited low-level object information is provided.

2.4 Size of the dataset

The integrated luminosity of the dataset needed for the individual exercises has been estimated to be 10 fb^{-1}. The total needed space will be determined by the size of the preselected real data and the size of the simulated data. To lessen the need for large storage resources the simulated datasets will be preselected. The final storage needs can be seen in Tables **??** to 31. The derivations used can be found in the Open Data CxAOD production framework [15]. It is estimated that 85 GB will be needed for the lepton (STDM4) collections, 3 GB for the diphoton (HIGG1D1) collection, 2 GB for the 1τ+1lep (HIGG4D2) collection, 100 MB for the 2τ (HIGG4D3) collection, 310 MB for the 1τ+0lep (HIGG4D5) collection, 2 GB for the fatjet (i.e. large-R jet) (HIGG5D2) collection and 1 GB for the truth (STDM4) collection. These collections will be provided separately as ntuples with different content. A total of approximately 100 GB of storage will be needed for all real and simulated data.

2.5 Proposed Monte Carlo datasets

Utilising simulated data in laboratory courses can teach students about background composition, statistical aspects of MC generation and data/MC comparison. A standard set of simulated SM processes includes W+jets, Z+jets, $t\bar{t}$, single-top, QCD, diboson, and diphoton datasets. Here the QCD datasets may be neglected due to the application of a one-lepton, one-tau or diphoton skim. For searches, the basic set of SM processes is complemented by simulations of Z', graviton and Higgs processes. The final datasets are described in Tables 10 to 31 in the Appendix. 85 GB of disk space is required for the lepton (STDM4) collection, 3 GB for the diphoton (HIGG1D1) collection, 2 GB for the 1τ+1lep (HIGG4D2) collection, 100 MB for the 2τ (HIGG4D3) collection, 310 MB for the 1τ+0lep (HIGG4D5) collection, 2 GB for the fatjet (i.e. large-R jet) (HIGG5D2) collection and 1 GB for the truth (STDM4) collection. These collections will be provided separately as ntuples with different content. Bundled together the simulated and real data require about 100 GB in storage.

2.6 Hosting of the datasets

It is proposed to host the dataset on the ATLAS and CERN Open Data portals [2, 3] in ROOT ntuple format, produced using ROOT 6.04/16 [16]. Currently, ATLAS stores 8 TeV Open Data in ROOT ntuple format on these portals, which in total has a size of about 8 GB. Storing an additional dataset of an order of magnitude larger in size should not pose a problem.

3 Motivation for research-driven education

The importance of the continuous integration of qualified personnel to the ATLAS Experiment is paramount. This endeavour started 25 years ago and is planned to last at least another 15 years from now. With such a lifetime it is easy to imagine that ATLAS members can spend their entire professional career associated with the experiment. Of course, this is not always the case, therefore constant knowledge transfer to students and young researchers is needed to keep running the experiment and the infrastructure around it.

In the meantime, ATLAS and the other LHC experiments are writing (and re-writing) the high-energy and particle physics textbooks in many subjects: from detector development and technology, to Data Acquisition Systems; from novel data-analysis techniques, to new (or more precise) physics results. This requires that professors and senior researchers around the world design and perform exercises with their trainees as close as possible to the reality of that new physics knowledge.

This provides strong motivation to make a collaborative effort to release meaningful and reliable samples for training and education. The 2016 release of ATLAS Open Data at 8 TeV succeeded in providing such samples. However, the 2016 release does not contain enough statistics to actually study the Higgs. Students are left with few events around the Higgs mass, which is somewhat disappointing since they expect to make a discovery with the 1 fb^{-1} of Open Data currently available. The next sections explore the physics analysis examples that influenced the design of the 13 TeV ATLAS Open Data samples.

3.1 Searching for the Higgs boson in the $H \to \gamma\gamma$ channel

One possible analysis to carry out with students during a laboratory course might be the study of $H \to \gamma\gamma$. The aim is to show students how this channel was used in the Higgs discovery. To reproduce selection requirements to reconstruct a simple invariant-mass peak, the following photon variables would be needed: four-momenta, isolation and quality criteria, and information whether photons were trigger and truth matched. The necessary MC would include the main Higgs production mechanisms; gluon–gluon fusion (ggH), vector-boson fusion (VBFH) and vector-boson associated production (VH). The production of a top-quark pair in association with a Higgs boson (ttH) would also be nice to include since this was only recently observed. To bring the current research frontier into the classroom, the most recent observations are important to cover. Processes that have been observed very recently are also often considered more attractive to students. For a simple analysis without using multivariate techniques, up to 10 fb^{-1} and a centre-of-mass energy of 13 TeV might be required to observe a statistically significant (3σ) Higgs peak in this channel alone.

3.2 Extending the 8 TeV Open Data example analyses

The ATLAS Open Data project has been used to teach undergraduate introductory courses in particle physics. These courses could include a general (qualitative) introduction followed by a discussion of experimental methods and finish with an “appetiser" to theoretical physics, i.e. quantum electrodynamics. The part discussing experimental methods could include e.g. three weeks using Open Data.

Example analyses could be released alongside the proposed datasets outlined in this document. Students could choose projects based on some of the example analyses provided in the Open Data project, and each would have a suggested analysis to start from. The analysis examples are briefly described below:

- Z boson analysis (suggested starting point: ZAnalysis.py): code containing a study regarding the Z mass peak, Z+jet balancing for jet calibration, Z+jet invariant mass.
- W boson analysis (suggested starting point: WAnalysis.py): code containing a study regarding the W charge asymmetry and its evolution with p_{T}, W+1 jet selection, fake-lepton study.
- $t\bar{t}$ 1-lepton analysis (suggested starting point: TTbarAnalysis.py): code containing a study regarding cross-section measurement, top mass measurement, boosted W boson inside $t\bar{t}$ events, $t\bar{t}$ invariant mass.

- WZ and ZZ analyses (suggested starting points: WZAnalysis.py and ZZAnalysis.py): code containing a study regarding the WZ charge asymmetry, p_{T} of the Z boson in WZ events, ZZ invariant mass, WZ invariant mass.
- $t\bar{t}$ 2-lepton analysis (suggested starting point: TTbarAnalysis.py): code containing selections for $t\bar{t}$ 2-lepton, cross-section measurement, b-jet content of the dataset.
- WW and $Z \to \tau\tau$ analyses (suggested starting point: ZAnalysis.py): code containing an extraction of a WW signal (starting from a Z boson selection), extraction of $Z \to \tau\tau$ signal, cross-section measurement of $Z \to \tau\tau$, WW+ 2 jets
- Search for new physics using same-sign lepton pairs (suggested starting point: ZAnalysis.py): code containing a study and suppression of fake leptons, suppression of charge-flip electrons, searching for a bump in the mass distribution.

If the university physics curriculum includes a mandatory programming course prior to taking this course, students would not need an introduction to programming for this course. Four hours of course time during the week in a computing laboratory, would provide a chance for a teacher and teaching assistant to answer questions. Students could either form teams of about seven, each picking one of the above projects, or teams of two working on the same project. Another possibility is to have only one simpler project which consists of re-discovering the Z and W bosons, the top quark and diboson (WZ) production, and demonstrate that these particles' behaviours are consistent with the SM.

The higher centre-of-mass energy and increased luminosity of a new release of Open Data using 13 TeV data will allow to study $H \to ZZ$, which will be very attractive to students. An addition of simulated Supersymmetry (SUSY) processes will allow to search for new physics with only slight modifications of the $t\bar{t}$ and WZ analyses, and will allow to bring the subject of dark matter into courses (since the lightest SUSY particle is considered to be a good dark matter candidate). From the educational and outreach points-of-view, a highlight for the new release could be the capability to inject these signals in the real data. Also, the addition of fat jets, which can be used to reconstruct boosted objects (W/Z/top/H) in the new release could allow to design interesting projects that use modern particle-physics tools.

3.3 Reconstructing the invariant and transverse masses of Standard Model bosons

A university laboratory course already studying $Z \to \ell\ell$ ($\ell = e, \mu$) could be extended to include $Z \to \tau\tau$, $H \to \tau\tau$ and $H \to \gamma\gamma$. In the same way, extending $W \to \ell\nu_\ell$ to $W \to \tau\nu_\tau$ would be possible. Having been introduced to the concept of invariant and transverse mass, students would be encouraged to discover the entire actual analyses themselves. Additionally to the first release of Open Data, these analyses would require information on hadronic tau and photon candidates. Some very simple systematic-uncertainty estimations will also be useful to teach students the importance of systematics. With 10 fb^{-1} at 13 TeV, studies of diboson events, including $H \to ZZ \to \ell\ell\ell\ell$, will be possible.

3.4 Searching for the Higgs boson and new physics

The previous release of ATLAS Open Data at 8 TeV has been used in several courses at the University of Oslo for the past years, and three different projects have been developed [7, 17]. Some parts of these projects can unfortunately not be completely carried out with the current Open Data release due to a lack of information/variables and/or low statistics.

A release of a 13 TeV dataset with higher integrated luminosity, and additional variables (e.g. photon information), would allow for nice extensions and realisations of the already existing projects.

The first project focuses on the 4ℓ final state, what it can tell us about the SM, and in particular the Higgs boson. As already mentioned in Section 3.3, a release of a dataset with higher integrated luminosity (and energy) would allow students to study the $H \to ZZ$ process, which would be very attractive.

The second project focuses on new forces and extra space dimensions, and the students do a search for the Z' and/or the graviton by studying the invariant mass distribution of two leptons. These two different signals could be further characterised by studying their angular distributions, since the Z' and the graviton carry different spins (spin-1 and spin-2, respectively). A part of this project that currently can not be carried out is to also study the diphoton channel, and do a search for Higgs (spin-0) and graviton production by making use of the diphoton invariant mass and angular distributions. This would however be possible with the proposed 13 TeV dataset.

The third project focuses on searches for SUSY and dark matter by looking at final states with two leptons and missing transverse energy. The simplest signal models to study in these cases are direct production of sleptons ($\tilde{\ell}^+\tilde{\ell}^-$) and production of mono-Z with associated dark matter candidate particles. The necessary variables are already contained in the existing 8 TeV dataset. However, additional variables (e.g. truth information and systematics) would allow for more sophisticated analyses for advanced students.

In general all three projects are closely related to the research done by the ATLAS group in Oslo, and they serve as excellent introductions to the field, in particular for first-year master students, or bachelor students that are curious about experimental particle physics. A new release of Open Data would bring these projects even closer to the current research frontier. It would also, to a larger extent, allow students to select the analyses they find most interesting, without the limitations of lacking variables or low statistics.

3.5 Study of t-channel single-top production

Another analysis that would be possible to carry out with undergraduate students, e.g. during a two weeks summer programme dedicated to introduce them into the field of top-quark physics, is the study of the t-channel single-top production. This study will consist of three steps. First, they will perform a simple generator-level study where they will reconstruct the top-quark mass using the true four-momenta of the top-quark decay products: $t \to W(\to \ell\nu)b$.

Later, they will reconstruct the top-quark mass using the four-momenta of detector-level physics objects: a lepton, a b-jet and the missing transverse momentum of the event. From the detector-level physics objects, they can determine neutrino longitudinal momentum, and then compare the rescaled missing transverse momentum with the true transverse momentum of the neutrino. Finally, they will perform a simple cut-based selection of t-channel single-top-quark events using the previously reconstructed top-quark mass and three additional discriminating variables and they will evaluate signal and background event yields.

3.6 Optimisation of event selection criteria for discovering new physics with ATLAS

Analysis paths exploring thousands of events in a batch mode could be used for third year undergraduate laboratories [18]. They would aim to enable students to simulate the researcher's analysis by optimising a set of relevant criteria in order to maximise the signal to background ratio. Students could study both decays of a Z boson to two leptons, and decays of Higgs bosons to four leptons.

For each particle they could be provided with three sets of events: a) Monte Carlo signal events b) Monte Carlo background events and c) real data.

In the case of the Z boson analysis, the background would be constructed by W+jets events with one same flavour opposite sign additional lepton. In the case of the Higgs boson search, the irreducible background $ZZ \rightarrow \ell\ell\ell\ell$ would be used together with two sources of reducible backgrounds Z+ jets and $t\bar{t}$. For the latter, the four-lepton-filtered datasets would be used.

The variables which could be used from the Open Data ntuples would be the kinematic information of the leptons (p_{T}, η, ϕ), the charge and the identification type (electron or muon). Furthermore, the variables which are implemented in the standard analysis to suppress the reducible backgrounds could also be used. These are the impact parameter significances (lep_trackd0pvunbiased, lep_tracksigd0pvunbiased, and lep_z0) and the track and calorimeter isolation variables (lep_ptcone30 and lep_etcone20) normalised to the lepton p_{T}.

Students could then use a user friendly GUI (see Figure 4) which lists the group of criteria (such as p_{T} of leptons, impact parameter significance and isolation) and optimise each one separately by inspecting the histograms provided for simulated signal and background events and maximising the corresponding significance, which is displayed as a function of the discriminating variable value (see Figure 5).

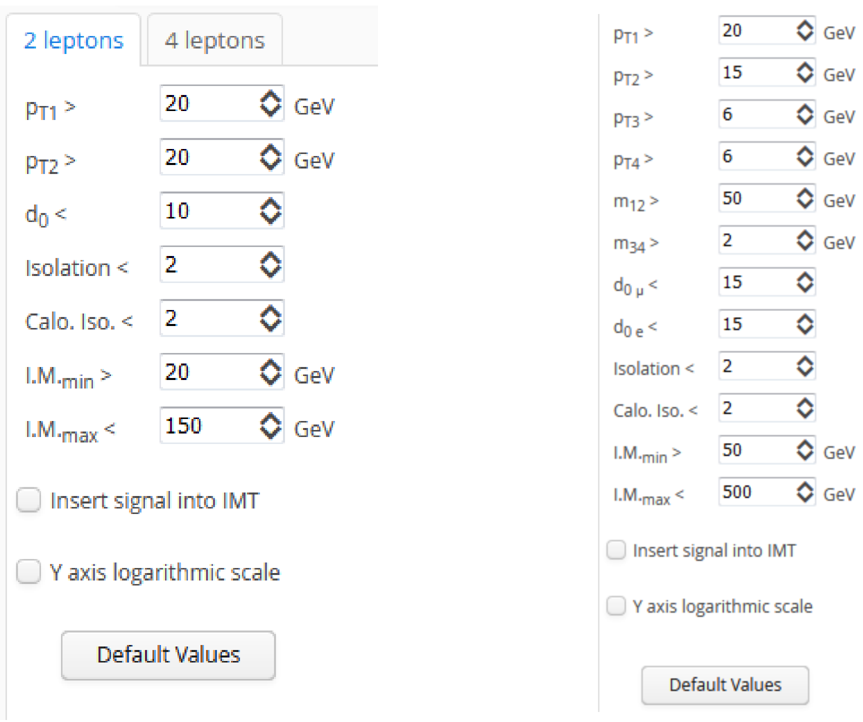

Figure 4: HYPATIA batch processing GUI to adjust the criteria to be optimised for a two lepton analysis (left) and a four lepton analysis (right).

Figure 5: Example of adjusting event selection criteria definition. The upper plot represents signal MC distribution, the middle plot the background MC distribution and the lower plot the corresponding significance. The x-axes correspond to isolation value and the y-axes to bin entries.

Many students could follow the above described analysis over a number of academic years. Nevertheless at the end of the day, while there are enough real data events for the data and simulation comparison in the $Z \to \ell\ell$ study, there are too few real data events for the data and simulation comparison in the $H \to ZZ \to \ell\ell\ell\ell$ study with 1 fb^{-1} of 8 TeV Open Data (see Figure 6). With the proposed release of 10 fb^{-1} of real data at 13 TeV this would finally be possible.

4 Conclusion

This note details the design, requirements, technical details and benefits of publishing an ATLAS endorsed dataset at a centre-of-mass energy of 13 TeV. We describe a series of educational projects and physics analyses that justify and will profit from such a dataset, as well as a comparison with respect to the previous ATLAS public release dataset at a centre-of-mass energy of 8 TeV in 2016. In terms of total luminosity, the size of the recorded data is proposed to be 10 fb^{-1}. Proper sizes for Monte Carlo samples describing necessary SM and BSM processes have been calculated taking into account the feedback of users and ATLAS analysis groups to be able to produce meaningful analysis examples. The content of the data samples will be rather basic: replicating a final-analysis kind of dataset and using only high-level variables such as the kinematics of reconstructed physics objects, quality flags, necessary event information and a minimal amount of truth and systematic information. Once published, both data and MC samples are proposed to be hosted on the ATLAS [2] and CERN [3] Open Data portals, having a total size of around 150 GB. The published datasets will foster educational efforts at the high-school and university levels by providing the material for hands-on physics exercises at different levels of complexity with the aim to reach as many people as possible in ATLAS and non-ATLAS members' institutions around the world.

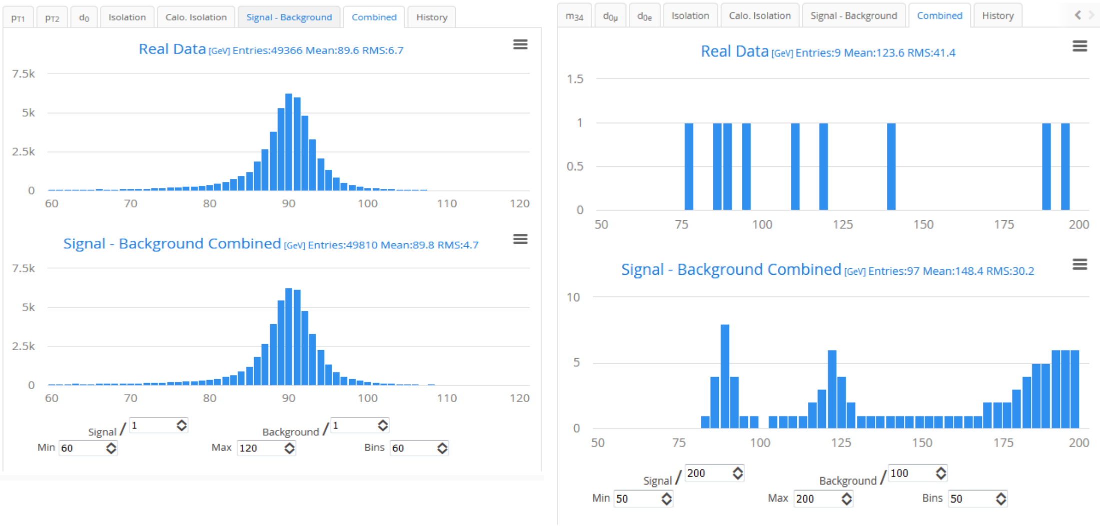

Figure 6: Data and simulation comparisons for the (a) $Z \to \ell\ell$ analysis and (b) $H \to ZZ(^*) \to \ell\ell\ell\ell$ analysis. In each case the upper plot corresponds to real data, while the lower plot corresponds to the combined signal and background MC data with user imposed normalisation.

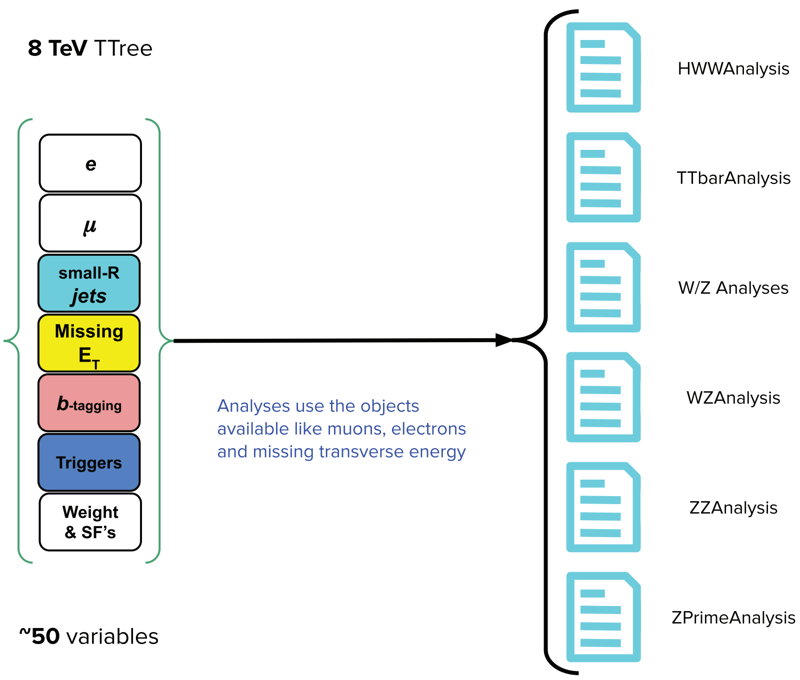

Figure 7: Reviewed Open Data physics analysis examples released together with the 8 TeV (2016) public samples.

branchname	type	description
runNumber	int	run identifier
eventNumber	int	event identifier
channelNumber	int	simulated data dataset ID e.g. $WW \to \ell\nu\ell\nu$ 361600
mcWeight	float	weight of a simulated event
SF_PILEUP	float	scalefactor for pileup reweighting
SF_ELE	float	scalefactor for electron efficiency, only $\neq$1 if electron selected & tagged
SF_MUON	float	scalefactor for muon efficiency, only $\neq$1 if muon selected & tagged
SF_PHOTON	float	scalefactor for photon efficiency, only $\neq$1 if photon selected & tagged
SF_TAU	float	scalefactor for tau efficiency, only $\neq$1 if tau selected & tagged
SF_BTAG	float	scalefactor for b-tagging algorithm, only $\neq$1 if b-jet selected & tagged
SF_LepTRIGGER	float	scalefactor for different operating efficiencies of used lepton triggers
SF_PhotonTRIGGER	float	scalefactor for different operating efficiencies of used photon triggers
SF_TauTRIGGER	float	scalefactor for different operating efficiencies of used tau triggers
SF_DiTauTRIGGER	float	scalefactor for different operating efficiencies of used ditau triggers
trigE	bool	boolean whether a standard trigger has fired in the egamma stream
trigM	bool	boolean whether a standard trigger has fired in the muon stream
trigP	bool	boolean whether a standard trigger has fired in the photon stream
trigT	bool	boolean whether a standard trigger has fired in the tau stream
trigDT	bool	boolean whether a standard trigger has fired in the ditau stream
lep_n	int	number of preselected leptons
lep_truthMatched	vector<bool>	boolean indicating whether the lepton is matched to a simulated lepton
lep_trigMatched	vector<bool>	boolean signifying whether the lepton is triggering the event
lep_pt	vector<float>	transverse momentum of the lepton
lep_eta	vector<float>	pseudo-rapidity of the lepton
lep_phi	vector<float>	azimuthal angle of the lepton
lep_E	vector<float>	energy of the lepton
lep_z0	vector<float>	z-coordinate of the track associated to the lepton wrt. primary vertex
lep_charge	vector<int>	charge of the lepton
lep_type	vector<int>	number signifying the lepton type (e or μ)
lep_isTightID	vector<bool>	boolean indicating whether lepton satisfies tight ID reconstruction criteria
lep_ptcone30	vector<float>	scalar sum of track p_{T} in a cone of R=0.3 around lepton
lep_etcone20	vector<float>	scalar sum of track E_{T} in a cone of R=0.2 around lepton
lep_d0	vector<float>	d_0 of track associated to lepton at point of closest approach (p.o.a.)
lep_d0sig	vector<float>	d_0 significance of track associated to lepton at p.o.a.
met_et	float	transverse energy of the missing momentum vector
met_phi	float	azimuthal angle of the missing momentum vector
jet_n	int	number of preselected jets
jet_pt	vector<float>	transverse momentum of the jet
jet_eta	vector<float>	pseudo-rapidity of the jet
jet_phi	vector<float>	azimuthal angle of the jet
jet_E	vector<float>	energy of the jet
jet_jvt	vector<float>	jet vertex tagging of the jet
jet_trueflav	vector<int>	flavour of the simulated jet
jet_truthMatched	vector<bool>	boolean indicating whether the jet is matched to a simulated jet
jet_MV2c10	vector<float>	weight from algorithm based on Multi-Variate technique

Table 1: Proposed event, lepton, missing transverse momentum and jet variable content of a data dataset for educational purposes. The content presented is a further slimmed down version of the ntuple produced by the $HH \to bb\tau\tau$ version of the CxAODFramework.

branchname	type	description
photon_n	int	number of preselected photons
photon_truthMatched	vector<bool>	boolean indicating whether the photon is matched to a simulated photon
photon_trigMatched	vector<bool>	boolean signifying whether the photon is triggering the event
photon_pt	vector<float>	transverse momentum of the photon
photon_eta	vector<float>	pseudo-rapidity of the photon
photon_phi	vector<float>	azimuthal angle of the photon
photon_E	vector<float>	energy of the photon
photon_isTightID	vector<bool>	boolean indicating whether photon satisfies tight ID reconstruction criteria
photon_ptcone30	vector<float>	scalar sum of track p_{T} in a cone of R=0.3 around photon
photon_etcone20	vector<float>	scalar sum of track E_{T} in a cone of R=0.2 around photon
photon_convType	vector<int>	information whether and where the photon was converted
fatjet_n	int	number of preselected fatjets
fatjet_pt	vector<float>	transverse momentum of the fatjet
fatjet_eta	vector<float>	pseudo-rapidity of the fatjet
fatjet_phi	vector<float>	azimuthal angle of the fatjet
fatjet_E	vector<float>	energy of the fatjet
fatjet_m	vector<float>	invariant mass of the fatjet
fatjet_truthMatched	vector<int>	information whether the fatjet is matched to a simulated fatjet
fatjet_D2	vector<float>	weight from algorithm for W/Z boson tagging
fatjet_tau32	vector<float>	weight from algorithm for top quark tagging
tau_n	int	number of preselected taus
tau_pt	vector<float>	transverse momentum of the tau
tau_eta	vector<float>	pseudo-rapidity of the tau
tau_phi	vector<float>	azimuthal angle of the tau
tau_E	vector<float>	energy of the tau
tau_isTightID	vector<bool>	boolean indicating whether tau satisfies tight ID reconstruction criteria
tau_truthMatched	vector<bool>	boolean indicating whether the tau is matched to a simulated tau
tau_trigMatched	vector<bool>	boolean signifying whether the tau is triggering the event
tau_nTracks	vector<int>	number of tracks in the tau decay
tau_BDTid	vector<float>	Boosted Decision Tree identification number of the tau
ditau_m	float	mass of ditau system, from the Missing Mass Calculator [13]
truth_pt	vector<float>	transverse momentum of the simulated particle in t decay chain (t, W, b, lep, ν)
truth_eta	vector<float>	pseudo-rapidity of the simulated particle in t decay chain (t, W, b, lep, ν)
truth_phi	vector<float>	azimuthal angle of the simulated particle in t decay chain (t, W, b, lep, ν)
truth_E	vector<float>	energy of the simulated particle in t decay chain (t, W, b, lep, ν)
truth_pdgid	vector<int>	PDG ID number of the simulated particle in t decay chain (t, W, b, lep, ν)
lep_pt_syst	vector<float>	quadrature sum of systematic shifts that affect lep_pt, to provide an estimate
met_et_syst	float	quadrature sum of systematic shifts that affect met_pt, to provide an estimate
jet_pt_syst	vector<float>	quadrature sum of systematic shifts that affect jet_pt, to provide an estimate
photon_pt_syst	vector<float>	quadrature sum of systematic shifts that affect photon_pt, to provide an estimate
fatjet_pt_syst	vector<float>	quadrature sum of systematic shifts that affect fatjet_pt, to provide an estimate
tau_pt_syst	vector<float>	quadrature sum of systematic shifts that affect tau_pt, to provide an estimate

Table 2: Proposed photon, fatjet, hadronic tau truth information and systematics estimation variable content of a data dataset for educational purposes. The content presented is a further slimmed down version of the ntuple produced by the $HH \to bb\tau\tau$ version of the CxAODFramework.

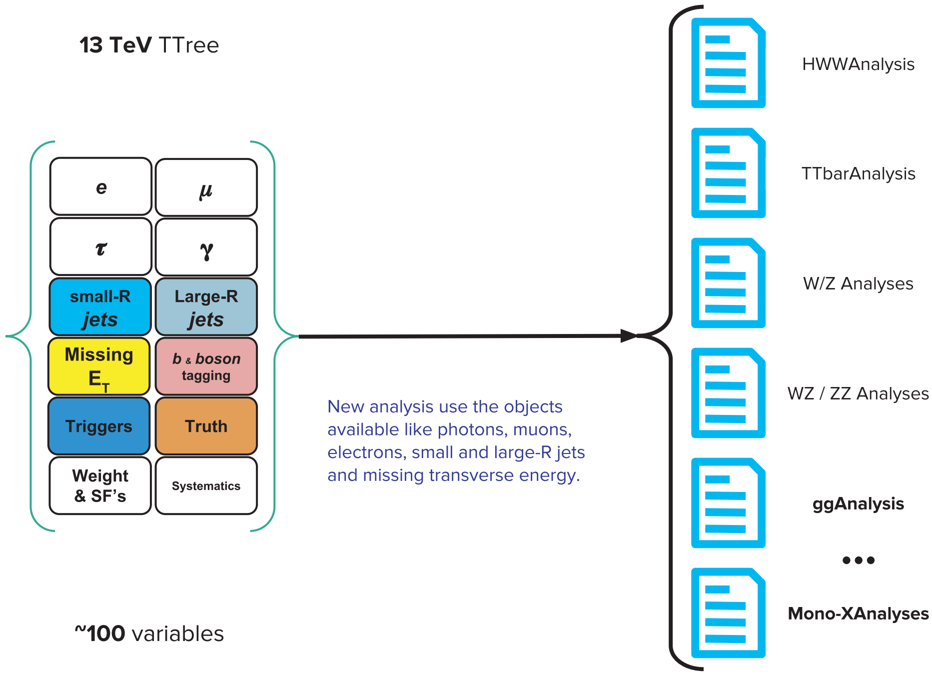

Figure 8: Proposed Open Data physics analysis examples to be released using the 13 TeV (2019) public samples.

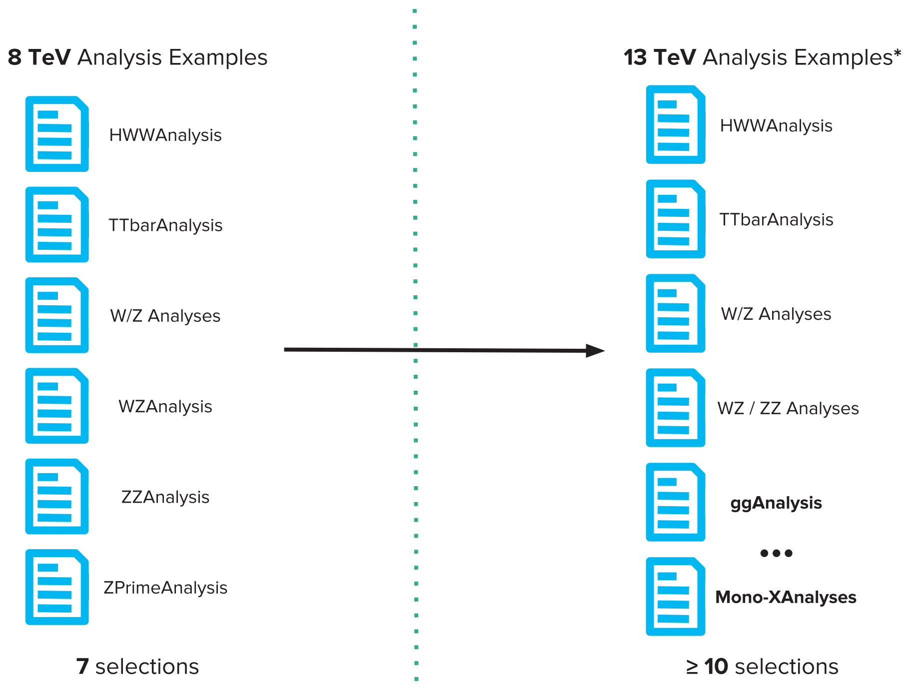

Figure 9: Evolution of the Open Data physics analysis examples from those using the 8 TeV samples (2016) to those using the 13 TeV samples (2019).
*The idea is that once the 13 TeV samples are public, ATLAS and non-ATLAS members can contribute to the development of those analysis examples.

List of contributions

Meirin Oan Evans	Editing, table compilation, appendix writing, “Searching for the Higgs boson in the $H \to \gamma\gamma$ channel" analysis
Even Simonsen Haaland	Editing, “Searching for the Higgs boson and new physics" analysis
Arturo Sanchez Pineda	ATLAS Outreach Data & Tools group coordination
Kate Shaw	ATLAS Open Data concept
Farid Ould-Saada	BSM processes and “Searching for the Higgs boson and new physics" analysis
Leonid Serkin	“Capabilities and Limitations of the Proposed Dataset" section
Jean-Francois Arguin	“Extending the provided example analyses" analysis
Susana Cabrera Urban	“Study of t-channel single top production" analysis
Dimitris Fassouliotis	“Optimisation of event selection criteria for discovering new physics with ATLAS” analysis
Christine Kourkoumelis	“Optimisation of event selection criteria for discovering new physics with ATLAS” analysis
Stelios Vourakis	‘Optimisation of event selection criteria for discovering new physics with ATLAS” analysis
Terrence Wyatt	“Reconstructing the invariant and transverse masses of Standard Model bosons" analysis
Magnar Kopangen Bugge	BSM processes and cross sections
Eirik Gramstad	BSM processes and cross sections
Sascha Mehlhase	editorial work, concept, comments
Agni Bethani	Tuple making code design
Thomas James Stevenson	Tuple making code design

Appendix

A Detailed List of MC Datasets

~85 GB will be needed for the lepton collections, ~3 GB for the diphoton collection, ~2 GB for the 1τ+1lep collection, 100 MB for the 2τ collection, 310 MB for the 1τ+0lep collection, ~2 GB for the fatjet collection and 1 GB for the truth collection. Bundled together the simulated and measured data require about ~100 GB in storage.

The cross sections and DSIDs for the SM and $Z' \rightarrow t\bar{t}$ processes quoted were taken from the Central MC15 Production List page [19]. Cross sections for leptonic Z', graviton, mono-Z and SUSY processes were taken from AMI [20]. Higgs to diphoton cross sections were taken from the paper "Measurements of Higgs boson properties in the diphoton decay channel with 36 fb^{-1} of pp collision data at $\sqrt{s}$ = 13 TeV with the ATLAS detector". Higgs branching ratios were taken from the Handbook of LHC Higgs Cross Sections [21].

period	$N_{\text{events}}^{\text{measured}}$	$N_{\text{events}}^{\text{pre}}$	$N_{\text{events}}^{\text{tuple}}$	$\mathcal{L}$/fb^{-1}	size/MB
A	361236098	47317312	10430587	0.547	1678
B	369173081	101669759	34876174	1.949	5910
C	623945258	152132517	51062670	2.884	8667
D	872671484	230557180	72797928	4.684	12362

Table 3: Breakdown of the dataset of measured data with exactly 1 lepton per event, with a total luminosity of 10.064 fb^{-1}. $N_{\text{events}}^{\text{measured}}$ denotes the number of events prior to preselection and $N_{\text{events}}^{\text{pre}}$ the number of events after preselection. $N_{\text{events}}^{\text{tuple}}$ is the number of events after further preselection. $\mathcal{L}$ denotes the luminosity of the dataset. Period A was made by combining the runs 297730, 298595, 298609, 298633, 298687, 298690, 298771, 298773, 298862, 298967, 299055, 299144, 299147, 299184, 299243, 299584 and 300279. Period B was made by combining the runs 300345, 300415, 300418, 300487, 300540, 300571, 300600, 300655, 300687, 300784, 300800, 300863 and 300908. Period C was made by combining the runs 301912, 301918, 301932, 301973, 302053, 302137, 302265, 302269, 302300, 302347, 302380, 302391 and 302393. Period D was made by combining the runs 302737, 302831, 302872, 302919, 302925, 302956, 303007, 303079, 303201, 303208, 303264, 303266, 303291, 303304, 303338, 303421, 303499 and 303560. The measured data in the 1 lepton collection were selected using the same preselection as applied on the simulated data in the 1 lepton collection. The total needed storage for 1 lepton measured data with exactly 1 lepton per event is around 28.617 GB.

period	$N_{\text{events}}^{\text{measured}}$	$N_{\text{events}}^{\text{pre}}$	$N_{\text{events}}^{\text{tuple}}$	$\mathcal{L}/\text{fb}^{-1}$	size/MB
A	361236098	47317312	535499	0.547	143
B	369173081	101669759	2459378	1.949	523
C	623945258	152132517	3587885	2.884	761
D	872671484	230557180	5490432	4.684	1162

Table 4: Breakdown of the dataset of measured data with at least 2 leptons per event, with a total luminosity of 10.064 fb^{-1}. $N_{\text{events}}^{\text{measured}}$ denotes the number of events prior to preselection and $N_{\text{events}}^{\text{pre}}$ the number of events after preselection. $N_{\text{events}}^{\text{tuple}}$ is the number of events after further preselection. $\mathcal{L}$ denotes the luminosity of the dataset. Period A was made by combining the runs 297730, 298595, 298609, 298633, 298687, 298690, 298771, 298773, 298862, 298967, 299055, 299144, 299147, 299184, 299243, 299584 and 300279. Period B was made by combining the runs 300345, 300415, 300418, 300487, 300540, 300571, 300600, 300655, 300687, 300784, 300800, 300863 and 300908. Period C was made by combining the runs 301912, 301918, 301932, 301973, 302053, 302137, 302265, 302269, 302300, 302347, 302380, 302391 and 302393. Period D was made by combining the runs 302737, 302831, 302872, 302919, 302925, 302956, 303007, 303079, 303201, 303208, 303264, 303266, 303291, 303304, 303338, 303421, 303499 and 303560. The measured data in the 2 lepton collection were selected using the same preselection as applied on the simulated data in the 2 lepton collection. The total needed storage for 2 lepton measured data with at least 2 leptons per event is around 2.589 GB.

period/run	$N_{\text{events}}^{\text{measured}}$	$N_{\text{events}}^{\text{pre}}$	$N_{\text{events}}^{\text{tuple}}$	$\mathcal{L}/\text{fb}^{-1}$	size/MB
A	361236098	2069198	430348	0.547	89.6
B	369173081	7010247	1528729	1.949	317.8
C	623945258	10321765	2237193	2.884	462.0
D	872671484			4.684	

Table 5: Breakdown of the dataset of measured data in the diphoton collection with a total luminosity of 10.064 fb^{-1}. $N_{\text{events}}^{\text{measured}}$ denotes the number of events prior to preselection and $N_{\text{events}}^{\text{pre}}$ the number of events after preselection. $N_{\text{events}}^{\text{tuple}}$ is the number of events after further preselection. $\mathcal{L}$ denotes the luminosity of the dataset. Period A was made by combining the runs 297730, 298595, 298609, 298633, 298687, 298690, 298771, 298773, 298862, 298967, 299055, 299144, 299147, 299184, 299243, 299584 and 300279. Period B was made by combining the runs 300345, 300415, 300418, 300487, 300540, 300571, 300600, 300655, 300687, 300784, 300800, 300863 and 300908. Period C was made by combining the runs 301912, 301918, 301932, 301973, 302053, 302137, 302265, 302269, 302300, 302347, 302380, 302391 and 302393. Period D was made by combining the runs 302737, 302831, 302872, 302919, 302925, 302956, 303007, 303079, 303201, 303208, 303264, 303266, 303291, 303304, 303338, 303421, 303499 and 303560. The measured data in the diphoton collection were selected using the same preselection as applied on the simulated data in the diphoton collection. The total needed storage for diphoton measured data is around ~1 GB.

period	$N_{\text{events}}^{\text{measured}}$	$N_{\text{events}}^{\text{pre}}$	$N_{\text{events}}^{\text{tuple}}$	$\mathcal{L}/\text{fb}^{-1}$	size/MB
A	361236098	19767030	49620	0.547	12.4
B	369173081	46411468	170700	1.949	42.9
C	623945258	69363987	259259	2.884	65.0
D	872671484	113197747	374120	4.684	94.0

Table 6: Breakdown of the dataset of measured data in the 1 hadronic τ + 1 leptonic τ collection with a total luminosity of 10.064 fb^{-1}. $N_{\text{events}}^{\text{measured}}$ denotes the number of events prior to preselection and $N_{\text{events}}^{\text{pre}}$ the number of events after preselection. $N_{\text{events}}^{\text{tuple}}$ is the number of events after further preselection. $\mathcal{L}$ denotes the luminosity of the dataset. Period A was made by combining the runs 297730, 298595, 298609, 298633, 298687, 298690, 298771, 298773, 298862, 298967, 299055, 299144, 299147, 299184, 299243, 299584 and 300279. Period B was made by combining the runs 300345, 300415, 300418, 300487, 300540, 300571, 300600, 300655, 300687, 300784, 300800, 300863 and 300908. Period C was made by combining the runs 301912, 301918, 301932, 301973, 302053, 302137, 302265, 302269, 302300, 302347, 302380, 302391 and 302393. Period D was made by combining the runs 302737, 302831, 302872, 302919, 302925, 302956, 303007, 303079, 303201, 303208, 303264, 303266, 303291, 303304, 303338, 303421, 303499 and 303560. The measured data in the 1 hadronic τ + 1 leptonic τ collection were selected using the same preselection as applied on the simulated data in 1 hadronic τ + 1 leptonic τ collection. The total needed storage for 1 hadronic τ + 1 leptonic τ measured data is around 214 MB.

period	$N_{\text{events}}^{\text{measured}}$	$N_{\text{events}}^{\text{pre}}$	$N_{\text{events}}^{\text{tuple}}$	$\mathcal{L}/\text{fb}^{-1}$	size/MB
A	361236098	6565575	10488	0.547	2.9
B	369173081	12373001	41499	1.949	10.8
C	623945258	19101749	62763	2.884	16.2
D	872671484	28072391	104718	4.684	27.1

Table 7: Breakdown of the dataset of measured data in the 2 hadronic τ + 0 lepton collection with a total luminosity of 10.064 fb^{-1}. $N_{\text{events}}^{\text{measured}}$ denotes the number of events prior to preselection and $N_{\text{events}}^{\text{pre}}$ the number of events after preselection. $N_{\text{events}}^{\text{tuple}}$ is the number of events after further preselection. $\mathcal{L}$ denotes the luminosity of the dataset. Period A was made by combining the runs 297730, 298595, 298609, 298633, 298687, 298690, 298771, 298773, 298862, 298967, 299055, 299144, 299147, 299184, 299243, 299584 and 300279. Period B was made by combining the runs 300345, 300415, 300418, 300487, 300540, 300571, 300600, 300655, 300687, 300784, 300800, 300863 and 300908. Period C was made by combining the runs 301912, 301918, 301932, 301973, 302053, 302137, 302265, 302269, 302300, 302347, 302380, 302391 and 302393. Period D was made by combining the runs 302737, 302831, 302872, 302919, 302925, 302956, 303007, 303079, 303201, 303208, 303264, 303266, 303291, 303304, 303338, 303421, 303499 and 303560. The measured data in the 2 hadronic τ + 0 lepton collection were selected using the same preselection as applied on the simulated data in the 2 hadronic τ + 0 lepton collection. The total needed storage for 2 hadronic τ + 0 lepton measured data is around 57 MB.

run	$N_{\text{events}}^{\text{measured}}$	$N_{\text{events}}^{\text{pre}}$	$N_{\text{events}}^{\text{tuple}}$	$\mathcal{L}/\text{pb}^{-1}$	size/MB
A	361236098	9385153	226308	0.547	50.1
B	369173081	13885436	131628	1.949	30.1
C	623945258	21913530	198942	2.884	45.5
D	872671484	32604484	324365	4.684	74.3

Table 8: Breakdown of the dataset of measured data in the 1 hadronic τ + 0 lepton collection with a total luminosity of 10.064 fb^{-1}. $N_{\text{events}}^{\text{measured}}$ denotes the number of events prior to preselection and $N_{\text{events}}^{\text{pre}}$ the number of events after preselection. $N_{\text{events}}^{\text{tuple}}$ is the number of events after further preselection. $\mathcal{L}$ denotes the luminosity of the dataset. Period A was made by combining the runs 297730, 298595, 298609, 298633, 298687, 298690, 298771, 298773, 298862, 298967, 299055, 299144, 299147, 299184, 299243, 299584 and 300279. Period B was made by combining the runs 300345, 300415, 300418, 300487, 300540, 300571, 300600, 300655, 300687, 300784, 300800, 300863 and 300908. Period C was made by combining the runs 301912, 301918, 301932, 301973, 302053, 302137, 302265, 302269, 302300, 302347, 302380, 302391 and 302393. Period D was made by combining the runs 302737, 302831, 302872, 302919, 302925, 302956, 303007, 303079, 303201, 303208, 303264, 303266, 303291, 303304, 303338, 303421, 303499 and 303560. The measured data in the 1 hadronic τ + 0 lepton collection were selected using the same preselection as applied on the simulated data in the 1 hadronic τ + 0 lepton collection. The total needed storage for 1 hadronic τ + 0 lepton measured data is around 200 MB.

run	$N_{\text{events}}^{\text{measured}}$	$N_{\text{events}}^{\text{pre}}$	$N_{\text{events}}^{\text{tuple}}$	$\mathcal{L}/\text{pb}^{-1}$	size/MB
A	361236098			0.547	
B	369173081			1.949	
C	623945258			2.884	
D	872671484			4.684	

Table 9: Breakdown of the dataset of measured data in the lepton + fatjet collection with a total luminosity of 10.064 fb^{-1}. $N_{\text{events}}^{\text{measured}}$ denotes the number of events prior to preselection and $N_{\text{events}}^{\text{pre}}$ the number of events after preselection. $N_{\text{events}}^{\text{tuple}}$ is the number of events after further preselection. $\mathcal{L}$ denotes the luminosity of the dataset. Period A was made by combining the runs 297730, 298595, 298609, 298633, 298687, 298690, 298771, 298773, 298862, 298967, 299055, 299144, 299147, 299184, 299243, 299584 and 300279. Period B was made by combining the runs 300345, 300415, 300418, 300487, 300540, 300571, 300600, 300655, 300687, 300784, 300800, 300863 and 300908. Period C was made by combining the runs 301912, 301918, 301932, 301973, 302053, 302137, 302265, 302269, 302300, 302347, 302380, 302391 and 302393. Period D was made by combining the runs 302737, 302831, 302872, 302919, 302925, 302956, 303007, 303079, 303201, 303208, 303264, 303266, 303291, 303304, 303338, 303421, 303499 and 303560. The measured data in the lepton + fatjet collection were selected using the same preselection as applied on the simulated data in the lepton + fatjet collection. The total needed storage for lepton + fatjet measured data is around ~40 MB.

process	DSID	Generator	σ/pb	FE	f_k	$N_\mathrm{events}^\mathrm{generated}$	$\mathcal{L}/\mathrm{fb}^{-1}$	$N_\mathrm{events}^\mathrm{tuple}$	size/MB
$t\bar{t} \to \ell + X$	410000	PowHeg+Pythia	696.11	0.5442	1.195	49386600	109.09	15747860	4690
single top wt-chan	410013	PowHeg+Pythia	34.009	1	1.054	4985800	135.74	898422	250
single antitop wt-chan	410014	PowHeg+Pythia	33.989	1	1.054	4985600	138.05	910949	250
single top s-chan	410025	PowHeg+Pythia	2.0514	1	1.0048	997800	484.08	274207	70
single antitop s-chan	410026	PowHeg+Pythia	1.2615	1	1.0215	995400	772.45	282568	70
Z+Jets ee	361106	PowHeg+Pythia	1901.1	1	1.026	79045597	40.53	16442978	3220
Z+Jets $\mu\mu$	361107	PowHeg+Pythia	1901.1	1	1.0261	77497800	39.73	12817065	2430
Z+Jets $\tau\tau$	361108	PowHeg+Pythia	1901.1	1	1.0261	29546000	15.15	986942	210
W^++Jets $e\nu$	361100	PowHeg+Pythia	11306.0	1	1.0172	41870000	3.64	~30000000	~2000
W^++Jets $\mu\nu$	361101	PowHeg+Pythia	11306.0	1	1.0172	39493600	3.43	13396822	2360
W^++Jets $\tau\nu$	361102	PowHeg+Pythia	11306.0	1	1.0172	59343600	5.16	550901	110
W^-+Jets $e\nu$	361103	PowHeg+Pythia	8283.1	1	1.0358	29886000	3.48	9395683	1690
W^-+Jets $\mu\nu$	361104	PowHeg+Pythia	8283.1	1	1.0358	31915400	3.72	11080210	1960
W^-+Jets $\tau\nu$	361105	PowHeg+Pythia	8283.1	1	1.0358	19945400	2.32	410455	80
$ZZ \to qq\ell^+\ell^-$	363356	Sherpa	15.564	1	0.14158	5317000	2413		
$WZ \to qq\ell^+\ell^-$	363358	Sherpa	3.4328	1	1	5124000	1493		
$W_\mathrm{p}W_\mathrm{m} \to qq\ell\nu_\ell$	363359	Sherpa	24.708	1	1	6673000	270.07		
$W_\mathrm{p}W_\mathrm{m} \to \ell\nu_\ell qq$	363360	Sherpa	24.724	1	1	7115000	287.78		
$WZ \to \ell\nu_\ell qq$	363489	Sherpa	11.42	1	1	7100000	621.72		
$ZZ \to \ell^+\ell^-\ell^+\ell^-$	363490	Sherpa	1.2578	1	1	17825300	14172		
$WZ \to \ell\nu_\ell\ell^+\ell^-$	363491	Sherpa	4.6049	1	1	15772084	3425		
$ZZ \to \ell^+\ell^-\nu\nu$	363492	Sherpa	12.466	1	1	14803000	1187		
$WZ \to \ell\nu_\ell\nu\nu$	363493	Sherpa	3.2286	1	1	5922600	1834		

Table 10: Full top, Z, inclusive W and diboson datasets with exactly 1 lepton. The factor FE denotes the filter efficiency for a given dataset and f_k is used for rescaling the leading order estimate to next to leading order in perturbative QCD. The total size of all simulated top, Z, inclusive W and diboson datasets with exactly 1 lepton is about ~60 GB. $N_\mathrm{events}^\mathrm{generated}$ denotes the number of events prior to preselection and $N_\mathrm{events}^\mathrm{tuple}$ is the number of events after preselection and requiring exactly 1 lepton per event.

process	DSID	Generator	σ/pb	FE	f_k	$N_{\text{events}}^{\text{generated}}$	$\mathcal{L}$/fb^{-1}	$N_{\text{events}}^{\text{tuple}}$	size/MB
W+Jets $\mu\nu$ no jets $b$veto 0-70	364156	Sherpa	19149	0.8245	0.9702	24723000	1.61	8395656	1520
W+Jets $\mu\nu$ with jets $b$veto	364157	Sherpa	19142	0.13091	0.9702	9847000	4.05	3666436	710
W+Jets $\mu\nu$ with b 0-70	364158	Sherpa	19138	0.044641	0.9702	17226200	20.78	5825715	1130
W+Jets $\mu\nu$ no jets $b$veto 70-140	364159	Sherpa	945.52	0.6743	0.9702	14788000	23.91	5938336	1280
W+Jets $\mu\nu$ with jets $b$veto 70-140	364160	Sherpa	945.44	0.24335	0.9702	9853800	44.14	4324093	970
W+Jets $\mu\nu$ with b 70-140	364161	Sherpa	944.14	0.084708	0.9702	19639000	253.1	7627342	1720
W+Jets $\mu\nu$ no jets $b$veto 140-280	364162	Sherpa	339.7	0.60292	0.9702	9882000	49.73	4307540	1010
W+Jets $\mu\nu$ with jets $b$veto 140-280	364163	Sherpa	339.88	0.29253	0.9702	7408000	76.8	3414233	830
W+Jets $\mu\nu$ with b 140-280	364164	Sherpa	339.64	0.110298	0.9702	24585000	676.43	10156719	2500
W+Jets $\mu\nu$ no jets $b$veto 280-500	364165	Sherpa	72.079	0.54768	0.9702	4940000	128.98	2212383	570
W+Jets $\mu\nu$ with jets $b$veto 280-500	364166	Sherpa	72.1	0.32016	0.9702	2958000	132.08	1359828	370
W+Jets $\mu\nu$ with b 280-500	364167	Sherpa	72.058	0.12542	0.9702	2959500	337.53	1250347	350
W+Jets $\mu\nu$ 500-1000	364168	Sherpa	15.006	1	0.9702	5910500	405.97	2663218	770
W+Jets $\mu\nu$ >1000	364169	Sherpa	1.2348	1	0.9702	3779000	3154	x	x
W+Jets $e\nu$ no jets $b$veto 0-70	364170	Sherpa	19153	0.82467	0.9702	24740000	1.61	7298004	1330
W+Jets $e\nu$ with jets $b$veto	364171	Sherpa	19145	0.13086	0.9702	9853500	4.05	3282722	640
W+Jets $e\nu$ with b 0-70	364172	Sherpa	19138	0.04482	0.9702	17242400	20.72	5152119	990
W+Jets $e\nu$ no jets $b$veto 70-140	364173	Sherpa	944.98	0.67483	0.9702	14660500	23.7	5668677	1220
W+Jets $e\nu$ with jets $b$veto 70-140	364174	Sherpa	945.74	0.24413	0.9702	9818400	43.83	4252664	950
W+Jets $e\nu$ with b 70-140	364175	Sherpa	945.79	0.10341	0.9702	9801900	103.3	3682272	820
W+Jets $e\nu$ no jets $b$veto 140-280	364176	Sherpa	339.67	0.59875	0.9702	9879000	50.07	4347004	1000
W+Jets $e\nu$ with jets $b$veto 140-280	364177	Sherpa	339.87	0.28876	0.9702	7410000	77.82	2759997	670
W+Jets $e\nu$ with b 140-280	364178	Sherpa	339.64	0.10898	0.9702	24677800	687.19	10298381	2520
W+Jets $e\nu$ no jets $b$veto 280-500	364179	Sherpa	72.074	0.5483	0.9702	4923800	128.42	2282543	590
W+Jets $e\nu$ with jets $b$veto 280-500	364180	Sherpa	72.105	0.31969	0.9702	2963400	132.51	1431891	390
W+Jets $e\nu$ with b 280-500	364181	Sherpa	72.091	0.13706	0.9702	2958000	308.56	1290393	360
W+Jets $e\nu$ 500-1000	364182	Sherpa	15.047	1	0.9702	5916800	405.3	2821018	820
W+Jets $e\nu$ >1000	364183	Sherpa	1.2344	1	0.9702	3947000	3296	1955311	600
W+Jets $\tau\nu$ no jets $b$veto 0-70	364184	Sherpa	19155	0.82462	0.9702	24784000	1.62	445592	90
W+Jets $\tau\nu$ with jets $b$veto	364185	Sherpa	19149	0.13152	0.9702	9865600	4.04	226449	50
W+Jets $\tau\nu$ with b 0-70	364186	Sherpa	19148	0.045111	0.9702	17273200	20.61	333628	70
W+Jets $\tau\nu$ no jets $b$veto 70-140	364187	Sherpa	945.02	0.67562	0.9702	14808500	23.91	650945	140
W+Jets $\tau\nu$ with jets $b$veto 70-140	364188	Sherpa	946.23	0.24247	0.9702	9860000	44.3	494988	110
W+Jets $\tau\nu$ with b 70-140	364189	Sherpa	945.71	0.10391	0.9702	9857000	103.39	411187	90
W+Jets $\tau\nu$ no jets $b$veto 140-280	364190	Sherpa	339.78	0.59872	0.9702	9899000	50.15	630080	150
W+Jets $\tau\nu$ with jets $b$veto 140-280	364191	Sherpa	339.66	0.28477	0.9702	7315000	77.95	494208	120
W+Jets $\tau\nu$ with b 140-280	364192	Sherpa	339.35	0.10599	0.9702	9834000	281.81	570262	140
W+Jets $\tau\nu$ no jets $b$veto 280-500	364193	Sherpa	72.091	0.54837	0.9702	4931200	128.57	358142	90
W+Jets $\tau\nu$ with jets $b$veto 280-500	364194	Sherpa	71.994	0.31881	0.9702	2956400	132.76	218067	60
W+Jets $\tau\nu$ with b 280-500	364195	Sherpa	71.945	0.13597	0.9702	2954100	311.26	193784	50
W+Jets $\tau\nu$ 500-1000	364196	Sherpa	15.052	1	0.9702	5355000	366.69	404545	120
W+Jets $\tau\nu$ >1000	364197	Sherpa	1.2341	1	0.9702	3946000	3296	317213	100

Table 11: Full sliced W datasets with exactly 1 lepton. The factor FE denotes the filter efficiency for a given dataset and f_k is used for rescaling the leading order estimate to next to leading order in perturbative QCD. The total size of all simulated sliced W datasets with exactly 1 lepton is about ~28 GB. $N_{\text{events}}^{\text{generated}}$ denotes the number of events prior to preselection and $N_{\text{events}}^{\text{tuple}}$ is the number of events after preselection and requiring exactly 1 lepton per event.

process	DSID	Generator	σ/pb	FE	f_k	$N_{\text{events}}^{\text{generated}}$	$\mathcal{L}$/fb^{-1}	$N_{\text{events}}^{\text{tuple}}$	size/MB
$WmH \to \ell\nu_\ell bb$ M_{H} = 125 GeV	345053	PowHeg+Pythia	0.17555	1	1	975900	5559		
$WpH \to \ell\nu_\ell bb$ M_{H} = 125 GeV	345054	PowHeg+Pythia	0.16458	1	1	1979950	12030		

Table 12: Full Higgs signal datasets with exactly 1 lepton. The factor FE denotes the filter efficiency for a given dataset and f_k is used for rescaling the leading order estimate to next to leading order in perturbative QCD. The total size of all simulated Higgs datasets with exactly 1 lepton is about ~100 MB. $N_{\text{events}}^{\text{generated}}$ denotes the number of events prior to preselection and $N_{\text{events}}^{\text{tuple}}$ is the number of events after preselection and requiring exactly 1 lepton per event.

process	DSID	Generator	σ/pb	FE	f_k	$N_{\text{events}}^{\text{generated}}$	$\mathcal{L}$/fb^{-1}	$N_{\text{events}}^{\text{tuple}}$	size/MB
$Z' \to t\bar{t}$ $M_{Z'}$ = 400 GeV	301322	Pythia	8.9857	1	1	199200	22.17	31697	9.3
$Z' \to t\bar{t}$ $M_{Z'}$ = 500 GeV	301323	Pythia	8.7385	1	1	199600	22.84	35678	10.7
$Z' \to t\bar{t}$ $M_{Z'}$ = 750 GeV	301324	Pythia	3.1201	1	1	199000	63.78	39762	12.2
$Z' \to t\bar{t}$ $M_{Z'}$ = 1000 GeV	301325	Pythia	1.1261	1	1	199800	177.43	41026	12.9
$Z' \to t\bar{t}$ $M_{Z'}$ = 1250 GeV	301326	Pythia	0.45981	1	1	199200	433.22	41243	13.1
$Z' \to t\bar{t}$ $M_{Z'}$ = 1500 GeV	301327	Pythia	0.20685	1	1	198800	961.08	39920	12.7
$Z' \to t\bar{t}$ $M_{Z'}$ = 1750 GeV	301328	Pythia	0.10016	1	1	199800	1995	38413	12.2
$Z' \to t\bar{t}$ $M_{Z'}$ = 2000 GeV	301329	Pythia	0.051346	1	1	199800	3891	36249	11.6
$Z' \to t\bar{t}$ $M_{Z'}$ = 2250 GeV	301330	Pythia	0.027481	1	1	199200	7249	34299	11.0
$Z' \to t\bar{t}$ $M_{Z'}$ = 2500 GeV	301331	Pythia	0.015226	1	1	198200	13017	32235	10.4
$Z' \to t\bar{t}$ $M_{Z'}$ = 2750 GeV	301332	Pythia	0.0086884	1	1	199800	22996	30322	9.8
$Z' \to t\bar{t}$ $M_{Z'}$ = 3000 GeV	301333	Pythia	0.0050843	1	1	195800	38511	28003	9.0

Table 13: Full Z' signal datasets with exactly 1 lepton. The factor FE denotes the filter efficiency for a given dataset and f_k is used for rescaling the leading order estimate to next to leading order in perturbative QCD. The total size of all simulated Z' datasets with exactly 1 lepton is about 130 MB. $N_{\text{events}}^{\text{generated}}$ denotes the number of events prior to preselection and $N_{\text{events}}^{\text{tuple}}$ is the number of events after preselection and requiring exactly 1 lepton per event.

process	DSID	Generator	σ/pb	FE	f_k	$N_\mathrm{events}^\mathrm{generated}$	$\mathcal{L}$/fb^{-1}	$N_\mathrm{events}^\mathrm{tuple}$	size/MB
$\tilde{g}\tilde{g} \rightarrow t\tilde{t}t\tilde{t} \rightarrow tttt$ + DM $M_{\tilde{\mathrm{g}}}$ = 1.2 TeV M_DM = 1 GeV	370114	MadGraph+Pythia	0.057037	1	1	100000	1753	25064	11.2
$\tilde{g}\tilde{g} \rightarrow t\tilde{t}t\tilde{t} \rightarrow tttt$ + DM $M_{\tilde{\mathrm{g}}}$ = 1.2 TeV M_DM = 600 GeV	370118	MadGraph+Pythia	0.057002	1	1	100000	1754	22957	10.2
$\tilde{g}\tilde{g} \rightarrow t\tilde{t}t\tilde{t} \rightarrow tttt$ + DM $M_{\tilde{\mathrm{g}}}$ = 1.4 TeV M_DM = 1 GeV	370129	MadGraph+Pythia	0.015756	1	1	100000	6347	25526	11.3
$\tilde{g}\tilde{g} \rightarrow t\tilde{t}t\tilde{t} \rightarrow tttt$ + DM $M_{\tilde{\mathrm{g}}}$ = 1.6 TeV M_DM = 1 GeV	370144	MadGraph+Pythia	0.004747	1	1	99000	20855	25102	11.1
$\tilde{t}\tilde{t} \rightarrow tt$ + DM $M_{\tilde{\mathrm{t}}}$ = 450 GeV M_DM = 1 GeV	388240	MadGraph+Pythia	0.88424	1	1	50000	56.55	9109	3.0
$\tilde{t}\tilde{t} \rightarrow tt$ + DM $M_{\tilde{\mathrm{t}}}$ = 500 GeV M_DM = 1 GeV	387154	MadGraph+Pythia	0.46603	1	1	20000	42.92	3656	1.2
$\tilde{t}\tilde{t} \rightarrow tt$ + DM $M_{\tilde{\mathrm{t}}}$ = 500 GeV M_DM = 200 GeV	387157	MadGraph+Pythia	0.46702	1	1	50000	107.06	8956	2.9
$\tilde{t}\tilde{t} \rightarrow tt$ + DM $M_{\tilde{\mathrm{t}}}$ = 600 GeV M_DM = 1 GeV	387163	MadGraph+Pythia	0.15518	1	1	49000	315.76	9444	3.1

Table 14: SUSY signal datasets with exactly 1 lepton. The factor FE denotes the filter efficiency for a given dataset and f_k is used for rescaling the leading order estimate to next to leading order in perturbative QCD. The total size of the simulated SUSY datasets with exactly 1 lepton is about 50 MB. $N_\mathrm{events}^\mathrm{generated}$ denotes the number of events prior to preselection and $N_\mathrm{events}^\mathrm{tuple}$ is the number of events after preselection and requiring exactly 1 lepton per event.

process	DSID	Generator	σ/pb	FE	f_k	$N_\mathrm{events}^\mathrm{generated}$	$\mathcal{L}/\mathrm{fb}^{-1}$	$N_\mathrm{events}^\mathrm{tuple}$	size/MB
single top t-chan	410011	PowHeg+Pythia	43.739	1	1.0094	4986200	112.94	1480868	490
single antitop t-chan	410012	PowHeg+Pythia	25.778	1	1.0193	4989800	189.9	1551986	520

Table 15: Full truth information datasets with exactly 1 lepton. The factor FE denotes the filter efficiency for a given dataset and f_k is used for rescaling the leading order estimate to next to leading order in perturbative QCD. The total size of all truth information 1 datasets with exactly 1 lepton is about 1 GB. The datasets have been subjected to the same skimming procedure as the 1 lepton SM datasets. $N_\mathrm{events}^\mathrm{generated}$ denotes the number of events prior to preselection and $N_\mathrm{events}^\mathrm{tuple}$ is the number of events after preselection.

process	DSID	Generator	σ/pb	FE	f_k	$N_\mathrm{events}^\mathrm{generated}$	$\mathcal{L}$/fb^{-1}	$N_\mathrm{events}^\mathrm{tuple}$	size/MB
$t\bar{t} \to \ell + X$	410000	PowHeg+Pythia	696.11	0.5442	1.195	49386600	109.09	2910544	921.9
single top wt-chan	410013	PowHeg+Pythia	34.009	1	1.054	4985800	135.74	165555	49.6
single antitop wt-chan	410014	PowHeg+Pythia	33.989	1	1.054	4985600	138.05	167835	50.4
single top s-chan	410025	PowHeg+Pythia	2.0514	1	1.0048	997800	484.08	7086	2.1
single antitop s-chan	410026	PowHeg+Pythia	1.2615	1	1.0215	995400	772.45	7513	2.2
Z+Jets ee	361106	PowHeg+Pythia	1901.1	1	1.026	79045597	40.53	21848557	4877.5
Z+Jets $\mu\mu$	361107	PowHeg+Pythia	1901.1	1	1.0261	77497800	39.73	22122823	4885.0
Z+Jets $\tau\tau$	361108	PowHeg+Pythia	1901.1	1	1.0261	29546000	15.15	182181	44.5
W^++Jets $e\nu$	361100	PowHeg+Pythia	11306.0	1	1.0172	41870000	3.64	41503	10.2
W^++Jets $\mu\nu$	361101	PowHeg+Pythia	11306.0	1	1.0172	39493600	3.43	35185	8.6
W^++Jets $\tau\nu$	361102	PowHeg+Pythia	11306.0	1	1.0172	59343600	5.16	1774	0.6
W^-+Jets $e\nu$	361103	PowHeg+Pythia	8283.1	1	1.0358	29886000	3.48	31894	7.8
W^-+Jets $\mu\nu$	361104	PowHeg+Pythia	8283.1	1	1.0358	31915400	3.72	30308	7.3
W^-+Jets $\tau\nu$	361105	PowHeg+Pythia	8283.1	1	1.0358	19945400	2.32	1409	0.4
$ZZ \to qq\ell^+\ell^-$	363356	Sherpa	15.564	1	0.14158	5317000	2413	1403154	396.4
$WZ \to qq\ell^+\ell^-$	363358	Sherpa	3.4328	1	1	5124000	1493	1316638	372.9
$W_\mathrm{p}W_\mathrm{m} \to qq\ell\nu_\ell$	363359	Sherpa	24.708	1	1	6673000	270.07	13375	4.0
$W_\mathrm{p}W_\mathrm{m} \to \ell\nu_\ell qq$	363360	Sherpa	24.724	1	1	7115000	287.78	14246	4.2
$WZ \to \ell\nu_\ell qq$	363489	Sherpa	11.42	1	1	7100000	621.72	28200	8.4
$ZZ \to \ell^+\ell^-\ell^+\ell^-$	363490	Sherpa	1.2578	1	1	17825300	14172		
$WZ \to \ell\nu_\ell\ell^+\ell^-$	363491	Sherpa	4.6049	1	1	15772084	3425		
$ZZ \to \ell^+\ell^-\nu\nu$	363492	Sherpa	12.466	1	1	14803000	1187		
$WZ \to \ell\nu_\ell\nu\nu$	363493	Sherpa	3.2286	1	1	5922600	1834		
ZZ+Jets 4ℓ filter	364250	Sherpa	1.252	1	1	17635900	14086	4190712	1281.1
$t\bar{t}W$	410155	Pythia	0.54822	1	1.096	7432900	12371		
$t\bar{t}Z \to e^+e^- + X$	410218	Pythia	0.036865	1	1.12	1408800	34121	789536	319.2
$t\bar{t}Z \to \mu^+\mu^- + X$	410219	Pythia	0.036868	1	1.12	1409600	34137	576598	233.6

Table 16: Full top, Z, W and diboson datasets with at least 2 leptons. The factor FE denotes the filter efficiency for a given dataset and f_k is used for rescaling the leading order estimate to next to leading order in perturbative QCD. The total size of all simulated top, Z, W and diboson datasets with at least 2 leptons is about x GB. $N_\mathrm{events}^\mathrm{generated}$ denotes the number of events prior to preselection and $N_\mathrm{events}^\mathrm{tuple}$ is the number of events after preselection and requiring at least 2 leptons per event.

process	DSID	Generator	σ/pb	FE	f_k	$N_{\mathrm{events}}^{\mathrm{generated}}$	$\mathcal{L}/\mathrm{fb}^{-1}$	$N_{\mathrm{events}}^{\mathrm{tuple}}$	size/MB
$Z \to \mu^+\mu^-$ $M_{\mu\mu} = 10 - 40$ GeVb-veto 0-70	364198	Sherpa	2414.3	0.96536	0.9751				
$Z \to \mu^+\mu^-$ $M_{\mu\mu} = 10 - 40$ GeVwith b 0-70	364199	Sherpa	2414.2	0.034445	0.9751				
$Z \to \mu^+\mu^-$ $M_{\mu\mu} = 10 - 40$ GeVb-veto 70-280	364200	Sherpa	50.33	0.89306	0.9751				
$Z \to \mu^+\mu^-$ $M_{\mu\mu} = 10 - 40$ GeVwith b 70-280	364201	Sherpa	50.29	0.11212	0.9751				
$Z \to \mu^+\mu^-$ $M_{\mu\mu} = 10 - 40$ GeVb-veto 280+	364202	Sherpa	3.2398	0.85373	0.9751				
$Z \to \mu^+\mu^-$ $M_{\mu\mu} = 10 - 40$ GeVwith b 280+	364203	Sherpa	3.2813	0.16027	0.9751				
$Z \to e^+e^-$ $M_{\mathrm{ee}} = 10 - 40$ GeVb-veto 0-70	364204	Sherpa	2415.3	0.96520	0.9751				
$Z \to e^+e^-$ $M_{\mathrm{ee}} = 10 - 40$ GeVwith b 0-70	364205	Sherpa	2415.5	0.034696	0.9751				
$Z \to e^+e^-$ $M_{\mathrm{ee}} = 10 - 40$ GeVb-veto 70-280	364206	Sherpa	50.354	0.89320	0.9751				
$Z \to e^+e^-$ $M_{\mathrm{ee}} = 10 - 40$ GeVwith b 70-280	364207	Sherpa	50.487	0.1087	0.9751				
$Z \to e^+e^-$ $M_{\mathrm{ee}} = 10 - 40$ GeVb-veto 280+	364208	Sherpa	3.2539	0.85485	0.9751				
$Z \to e^+e^-$ $M_{\mathrm{ee}} = 10 - 40$ GeVwith b 280+	364209	Sherpa	3.2526	0.15351	0.9751				
$Z \to \tau^+\tau^-$ $M_{\tau\tau} = 10 - 40$ GeVb-veto 0-70	364210	Sherpa	2416.0	0.96534	0.9751				
$Z \to \tau^+\tau^-$ $M_{\tau\tau} = 10 - 40$ GeVwith b 0-70	364211	Sherpa	2414.4	0.034676	0.9751				
$Z \to \tau^+\tau^-$ $M_{\tau\tau} = 10 - 40$ GeVb-veto 70-280	364212	Sherpa	50.376	0.89310	0.9751				
$Z \to \tau^+\tau^-$ $M_{\tau\tau} = 10 - 40$ GeVwith b 70-280	364213	Sherpa	50.468	0.10985	0.9751				
$Z \to \tau^+\tau^-$ $M_{\tau\tau} = 10 - 40$ GeVb-veto 280+	364214	Sherpa	3.2852	0.85552	0.9751				
$Z \to \tau^+\tau^-$ $M_{\tau\tau} = 10 - 40$ GeVwith b 280+	364215	Sherpa	3.2813	0.15581	0.9751				

Table 17: Full low-mass Drell-Yan datasets with at least 2 leptons. The factor FE denotes the filter efficiency for a given dataset and f_k is used for rescaling the leading order estimate to next to leading order in perturbative QCD. The total size of all simulated low-mass Drell-Yan datasets with at least 2 leptons is about ~28 GB. $N_{\mathrm{events}}^{\mathrm{generated}}$ denotes the number of events prior to preselection and $N_{\mathrm{events}}^{\mathrm{tuple}}$ is the number of events after preselection and requiring at least 2 leptons per event.

process	DSID	Generator	σ/pb	FE	f_k	$N_{\text{events}}^{\text{generated}}$	$\mathcal{L}$/fb^{-1}	$N_{\text{events}}^{\text{tuple}}$	size/MB
$gg \to H \to \tau\tau$ M_{H} = 125 GeV	341122	PowHeg+Pythia	1.90820026	0.12277	1.4547	1522300	4467	104490	26.7
VBF$H \to \tau\tau$ M_{H} = 125 GeV	341155	PowHeg+Pythia	0.2420117	0.12271	0.9788	2078800	71516	403327	105.3
$ZH \to$ any+$ZZ(\to \ell\ell\ell\ell)$ M_{H} = 125 GeV	341947	Pythia	0.0002089	0.010256	1	150000	70012374	59325	20.4
$WH \to ZZ \to \ell\ell\ell\ell$ M_{H} = 125 GeV	341964	Pythia	0.0003769	1	1	149400	396392	61009	20.9
VBF$H \to ZZ \to \ell\ell\ell\ell$ M_{H} = 125 GeV	344235	PowHeg+Pythia	0.0004633012	1	1	985000	2126047	583980	185.6
$ZH \to \ell^+\ell^- bb$ M_{H} = 125 GeV	345055	PowHeg+Pythia	0.044837	1	1	2968500	66206		
$gg \to ZH \to \ell^+\ell^- bb$ M_{H} = 125 GeV	345057	PowHeg+Pythia	0.005803	1	1	678000	116836		
$gg \to H \to ZZ \to \ell\ell\ell\ell$ M_{H} = 125 GeV	345060	PowHeg+Pythia	0.0060239	1	1	985000	163515	525407	162.3
VBF$H \to WW \to \ell\nu\ell\nu$ M_{H} = 125 GeV	345323	PowHeg+Pythia	3.7365	0.51007	1	1175000	616.51	425132	108.2
$gg \to H \to WW \to \ell\nu\ell\nu$ M_{H} = 125 GeV	345324	PowHeg+Pythia	28.301	0.49374	1	1972000	141.13	628689	155.4
$WH \to qqWW(\to \ell\nu\ell\nu)$ M_{H} = 125 GeV	345325	PowHeg+Pythia	0.86202	1	1	246000	285.38	55035	16.3
$WH \to \ell\nu WW(\to \ell\nu\ell\nu)$ M_{H} = 125 GeV	345327	PowHeg+Pythia	0.27864	1	1	99000	355.30	41887	12.3
$ZH \to qqWW(\to \ell\nu\ell\nu)$ M_{H} = 125 GeV	345336	PowHeg+Pythia	0.76093	1	1	245000	321.97	48665	14.8
$ZH \to \ell\ell WW(\to \ell\nu\ell\nu)$ M_{H} = 125 GeV	345337	PowHeg+Pythia	0.076214	1	1	297000	3897	168829	53.7
$ZH \to \nu\nu WW(\to \ell\nu\ell\nu)$ M_{H} = 125 GeV	345445	PowHeg+Pythia	0.1501	1	1	198000	1319	40364	10.4

Table 18: Full Higgs signal datasets with at least 2 leptons. The factor FE denotes the filter efficiency for a given dataset and f_k is used for rescaling the leading order estimate to next to leading order in perturbative QCD. The total size of all simulated Higgs datasets with at least 2 leptons is about ~1000 MB. Though not very big the datasets have been subjected to the same skimming procedure as the 2 lepton SM datasets. $N_{\text{events}}^{\text{generated}}$ denotes the number of events prior to preselection and $N_{\text{events}}^{\text{tuple}}$ is the number of events after preselection and requiring at least 2 leptons per event.

process	DSID	Generator	σ/pb	FE	f_k	$N_{\text{events}}^{\text{generated}}$	$\mathcal{L}$/fb^{-1}	$N_{\text{events}}^{\text{tuple}}$	size/MB
$Z' \to t\bar{t}$ $M_{Z'}$ = 400 GeV	301322	Pythia	8.9857	1	1	199200	22.17	5776	1.8
$Z' \to t\bar{t}$ $M_{Z'}$ = 500 GeV	301323	Pythia	8.7385	1	1	199600	22.84	6333	2.0
$Z' \to t\bar{t}$ $M_{Z'}$ = 750 GeV	301324	Pythia	3.1201	1	1	199000	63.78	6646	2.1
$Z' \to t\bar{t}$ $M_{Z'}$ = 1000 GeV	301325	Pythia	1.1261	1	1	199800	177.43	6498	2.1
$Z' \to t\bar{t}$ $M_{Z'}$ = 1250 GeV	301326	Pythia	0.45981	1	1	199200	433.22	5937	2.0
$Z' \to t\bar{t}$ $M_{Z'}$ = 1500 GeV	301327	Pythia	0.20685	1	1	198800	961.08	5337	1.8
$Z' \to t\bar{t}$ $M_{Z'}$ = 1750 GeV	301328	Pythia	0.10016	1	1	199800	1995	4816	1.7
$Z' \to t\bar{t}$ $M_{Z'}$ = 2000 GeV	301329	Pythia	0.051346	1	1	199800	3891	4278	1.5
$Z' \to t\bar{t}$ $M_{Z'}$ = 2250 GeV	301330	Pythia	0.027481	1	1	199200	7249	3741	1.3
$Z' \to t\bar{t}$ $M_{Z'}$ = 2500 GeV	301331	Pythia	0.015226	1	1	198200	13017	3333	1.2
$Z' \to t\bar{t}$ $M_{Z'}$ = 2750 GeV	301332	Pythia	0.0086884	1	1	199800	22996	3093	1.1
$Z' \to t\bar{t}$ $M_{Z'}$ = 3000 GeV	301333	Pythia	0.0050843	1	1	195800	38511	2674	1.0
$Z' \to ee$ $M_{Z'}$ = 2000 GeV	301215	Pythia	0.0088432	1	1	19800	2239	14938	3.8
$Z' \to ee$ $M_{Z'}$ = 3000 GeV	301216	Pythia	0.00080617	1	1	19600	24312	15054	3.8
$Z' \to ee$ $M_{Z'}$ = 4000 GeV	301217	Pythia	0.00010351	1	1	19800	191286	14912	3.8
$Z' \to ee$ $M_{Z'}$ = 5000 GeV	301218	Pythia	0.000018319	1	1	18000	982586	12554	3.2
$Z' \to \mu\mu$ $M_{Z'}$ = 2000 GeV	301220	Pythia	0.0088801	1	1	983000	110697	526455	129.5
$Z' \to \mu\mu$ $M_{Z'}$ = 3000 GeV	301221	Pythia	0.00080295	1	1	988000	1230463	516562	126.1
$Z' \to \mu\mu$ $M_{Z'}$ = 4000 GeV	301222	Pythia	0.00010332	1	1	986000	9543167	501025	113.7
$Z' \to \mu\mu$ $M_{Z'}$ = 5000 GeV	301223	Pythia	0.000018334	1	1	999000	54488928	492452	119.2
Mono-Z M_{DM} = 10 GeV	303511	MadGraph+Pythia	11.55	0.117	1	10000	7.40	3311	0.8
Mono-Z M_{DM} = 100 GeV	303512	MadGraph+Pythia	0.4682	0.65182	1	10000	32.77	4058	1.0
Mono-Z M_{DM} = 200 GeV	306085	MadGraph+Pythia	0.1424	0.78542	1	10000	89.41	4484	1.1
Mono-Z M_{DM} = 300 GeV	303513	MadGraph+Pythia	0.063965	0.82651	1	10000	189.15	4796	1.2
Mono-Z M_{DM} = 400 GeV	306093	MadGraph+Pythia	0.031865	0.85903	1	10000	365.32	4875	1.2
Mono-Z M_{DM} = 500 GeV	305710	MadGraph+Pythia	0.018275	0.87831	1	10000	623.01	4955	1.3
Mono-Z M_{DM} = 600 GeV	306103	MadGraph+Pythia	0.01136	0.88255	1	10000	997.43	5146	1.3
Mono-Z M_{DM} = 700 GeV	305711	MadGraph+Pythia	0.007416	0.89535	1	10000	1506	5249	1.3
Mono-Z M_{DM} = 800 GeV	306109	MadGraph+Pythia	0.005016	0.90148	1	10000	2211	5227	1.3
Mono-Z M_{DM} = 2000 GeV	303514	MadGraph+Pythia	0.0001636	0.93449	1	10000	65410	5427	1.4
	302681	MadGraph+Pythia	0.031435	1	1	20000	636.23	9238	3.3
	302687	MadGraph+Pythia	0.018827	1	1	15000	796.73	6994	2.5
	302701	MadGraph+Pythia	0.005172	1	1	20000	3867		
	302708	MadGraph+Pythia	0.001489	1	1	19000	12760	8759	3.1
	309070	MadGraph+Pythia	0.00028772	1	1	19000	66036	8607	3.0

Table 19: Full Z' and Mono-Z signal datasets with at least 2 leptons. The factor FE denotes the filter efficiency for a given dataset and f_k is used for rescaling the leading order estimate to next to leading order in perturbative QCD. The total size of all simulated Z' and Mono-Z datasets with at least 2 leptons is about 530 MB. Though not very big the datasets have been subjected to the same skimming procedure as the 2 lepton SM datasets. $N_{\text{events}}^{\text{generated}}$ denotes the number of events prior to preselection and $N_{\text{events}}^{\text{tuple}}$ is the number of events after preselection and requiring at least 2 leptons per event.

process	DSID	Generator	σ/pb	FE	f_k	$N_{events}^{generated}$	$\mathcal{L}/fb^{-1}$	N_{events}^{tuple}	size/MB
$G \to ee\ M_G = 750$ GeV	305550	Pythia	2.4052	1	1	49000	20.37	27411	7.4
$G \to ee\ M_G = 1000$ GeV	305553	Pythia	0.57205	1	1	47000	82.16	25922	7.1
$G \to ee\ M_G = 2000$ GeV	305556	Pythia	0.011431	1	1	98000	4199	24801	6.8
$G \to ee\ M_G = 3000$ GeV	305559	Pythia	0.00069865	1	1	49000	70135	24627	6.8
$G \to ee\ M_G = 4000$ GeV	305562	Pythia	0.000064079	1	1	49000	764681	24237	6.6
$G \to \mu\mu\ M_G = 750$ GeV	305568	Pythia	2.3952	1	1	50000	20.88	27927	7.5
$G \to \mu\mu\ M_G = 1000$ GeV	305571	Pythia	0.56835	1	1	50000	87.97	27469	7.5
$G \to \mu\mu\ M_G = 2000$ GeV	305574	Pythia	0.011387	1	1	50000	4391	25885	7.1
$G \to \mu\mu\ M_G = 3000$ GeV	305577	Pythia	0.00070262	1	1	50000	71162	25006	6.9
$G \to \mu\mu\ M_G = 4000$ GeV	305580	Pythia	0.000064403	1	1	48000	776361	x	x
$G \to ZZ \to \ell\ell\ell\ell\ M_G = 200$ GeV	307431	MadGraph+Pythia	1.86	1	1	29000	15.59	15772	5.2
$G \to ZZ \to \ell\ell\ell\ell\ M_G = 500$ GeV	307434	MadGraph+Pythia	0.02373	1	1	27000	1138	19619	6.9
$G \to ZZ \to \ell\ell\ell\ell\ M_G = 1000$ GeV	303329	MadGraph+Pythia	0.0004122	1	1	3000	7278	2210	0.8
$G \to ZZ \to \ell\ell\ell\ell\ M_G = 1500$ GeV	307439	MadGraph+Pythia	0.00003702	1	1	30000	810373	21670	7.7
$G \to ZZ \to \ell\ell\ell\ell\ M_G = 2000$ GeV	303334	MadGraph+Pythia	0.0000057	1	1	5000	877193	2908	1.0

Table 20: Full graviton signal datasets with at least 2 leptons. The factor FE denotes the filter efficiency for a given dataset and f_k is used for rescaling the leading order estimate to next to leading order in perturbative QCD. The total size of all simulated graviton datasets with at least 2 leptons is about ~10 MB. Though not very big the datasets have been subjected to the same skimming procedure as the 2 lepton SM datasets. $N_{events}^{generated}$ denotes the number of events prior to preselection and N_{events}^{tuple} is the number of events after preselection and requiring at least 2 leptons per event.

process	DSID	Generator	σ/pb	FE	f_{k}	$N_{\mathrm{events}}^{\mathrm{generated}}$	$\mathcal{L}$/fb^{-1}	$N_{\mathrm{events}}^{\mathrm{tuple}}$	size/MB
$\tilde{g}\tilde{g} \rightarrow \tilde{t}\tilde{t}\tilde{t}\tilde{t} \rightarrow tttt + \mathrm{DM}\ M_{\tilde{g}} = 1.2\ \mathrm{TeV} M_{\mathrm{DM}} = 1\ \mathrm{GeV}$	370114	MadGraph+Pythia	0.057037	1	1	100000	1753	12096	5.7
$\tilde{g}\tilde{g} \rightarrow \tilde{t}\tilde{t}\tilde{t}\tilde{t} \rightarrow tttt + \mathrm{DM}\ M_{\tilde{g}} = 1.2\ \mathrm{TeV} M_{\mathrm{DM}} = 600\ \mathrm{GeV}$	370118	MadGraph+Pythia	0.057002	1	1	100000	1754	12564	5.9
$\tilde{g}\tilde{g} \rightarrow \tilde{t}\tilde{t}\tilde{t}\tilde{t} \rightarrow tttt + \mathrm{DM}\ M_{\tilde{g}} = 1.4\ \mathrm{TeV} M_{\mathrm{DM}} = 1\ \mathrm{GeV}$	370129	MadGraph+Pythia	0.015756	1	1	100000	6347	11509	5.4
$\tilde{g}\tilde{g} \rightarrow \tilde{t}\tilde{t}\tilde{t}\tilde{t} \rightarrow tttt + \mathrm{DM}\ M_{\tilde{g}} = 1.6\ \mathrm{TeV} M_{\mathrm{DM}} = 1\ \mathrm{GeV}$	370144	MadGraph+Pythia	0.004747	1	1	99000	20855	11067	5.2
$\tilde{t}\tilde{t} \rightarrow tt + \mathrm{DM}\ M_{\tilde{t}} = 450\ \mathrm{GeV} M_{\mathrm{DM}} = 1\ \mathrm{GeV}$	388240	MadGraph+Pythia	0.88424	1	1	50000	56.55	1601	0.6
$\tilde{t}\tilde{t} \rightarrow tt + \mathrm{DM}\ M_{\tilde{t}} = 500\ \mathrm{GeV} M_{\mathrm{DM}} = 1\ \mathrm{GeV}$	387154	MadGraph+Pythia	0.46603	1	1	20000	42.92	656	0.2
$\tilde{t}\tilde{t} \rightarrow tt + \mathrm{DM}\ M_{\tilde{t}} = 500\ \mathrm{GeV} M_{\mathrm{DM}} = 200\ \mathrm{GeV}$	387157	MadGraph+Pythia	0.46702	1	1	50000	107.06	1690	0.6
$\tilde{t}\tilde{t} \rightarrow tt + \mathrm{DM}\ M_{\tilde{t}} = 600\ \mathrm{GeV} M_{\mathrm{DM}} = 1\ \mathrm{GeV}$	387163	MadGraph+Pythia	0.15518	1	1	49000	315.76	1652	0.6
$\tilde{\chi}_1^{\pm}\tilde{\chi}_2^0 \rightarrow WZ(\rightarrow \ell\nu\ell\ell)+\not{E}_{\mathrm{T}}\ M_{\tilde{\chi}_1^{\pm}} = 100\ \mathrm{GeV} M_{\tilde{\chi}_1^0} = 0\ \mathrm{GeV}$	392226	MadGraph+Pythia	23.525	0.67285	1	20000	1.26	11686	3.3
$\tilde{\chi}_1^{\pm}\tilde{\chi}_2^0 \rightarrow WZ(\rightarrow \ell\nu\ell\ell)+\not{E}_{\mathrm{T}}\ M_{\tilde{\chi}_1^{\pm}} = 350\ \mathrm{GeV} M_{\tilde{\chi}_1^0} = 0\ \mathrm{GeV}$	392220	MadGraph+Pythia	0.18475	0.76781	1	10000	70.5	7009	2.1
$\tilde{\chi}_1^{\pm}\tilde{\chi}_2^0 \rightarrow WZ(\rightarrow \ell\nu\ell\ell)+\not{E}_{\mathrm{T}}\ M_{\tilde{\chi}_1^{\pm}} = 400\ \mathrm{GeV} M_{\tilde{\chi}_1^0} = 0\ \mathrm{GeV}$	392217	MadGraph+Pythia	0.10480	0.76994	1	10000	123.93	7099	2.1
$\tilde{\chi}_1^{\pm}\tilde{\chi}_2^0 \rightarrow WZ(\rightarrow \ell\nu\ell\ell)+\not{E}_{\mathrm{T}}\ M_{\tilde{\chi}_1^{\pm}} = 500\ \mathrm{GeV} M_{\tilde{\chi}_1^0} = 0\ \mathrm{GeV}$	392223	MadGraph+Pythia	0.03899	0.77285	1	5000	165.93	3644	1.1
$\tilde{\chi}_1^{\pm}\tilde{\chi}_2^0 \rightarrow WZ(\rightarrow qq\ell\ell)+\not{E}_{\mathrm{T}}\ M_{\tilde{\chi}_1^{\pm}} = 0.5\ \mathrm{TeV} M_{\tilde{\chi}_1^0} = 0.1\ \mathrm{TeV}$	392302	MadGraph+Pythia	0.03896	0.65405	1	5000	196.22	2947	0.9
$\tilde{\chi}_1^{\pm}\tilde{\chi}_2^0 \rightarrow WZ(\rightarrow qq\ell\ell)+\not{E}_{\mathrm{T}}\ M_{\tilde{\chi}_1^{\pm}} = 0.3\ \mathrm{TeV} M_{\tilde{\chi}_1^0} = 0.1\ \mathrm{TeV}$	392304	MadGraph+Pythia	0.34915	0.62499	1	10000	45.83	5412	1.6
$\tilde{\chi}_1^{\pm}\tilde{\chi}_2^0 \rightarrow WZ(\rightarrow qq\ell\ell)+\not{E}_{\mathrm{T}}\ M_{\tilde{\chi}_1^{\pm}} = 0.3\ \mathrm{TeV} M_{\tilde{\chi}_1^0} = 0.2\ \mathrm{TeV}$	392308	MadGraph+Pythia	0.35045	0.62466	1	10000	45.68	4947	1.5
$\tilde{\chi}_1^{\pm}\tilde{\chi}_2^0 \rightarrow WZ(\rightarrow qq\ell\ell)+\not{E}_{\mathrm{T}}\ M_{\tilde{\chi}_1^{\pm}} = 0.4\ \mathrm{TeV} M_{\tilde{\chi}_1^0} = 0\ \mathrm{TeV}$	392317	MadGraph+Pythia	0.1049	0.65427	1	10000	145.7	5708	1.7
$\tilde{\chi}_1^{\pm}\tilde{\chi}_2^0 \rightarrow WZ(\rightarrow qq\ell\ell)+\not{E}_{\mathrm{T}}\ M_{\tilde{\chi}_1^{\pm}} = 0.5\ \mathrm{TeV} M_{\tilde{\chi}_1^0} = 0\ \mathrm{TeV}$	392323	MadGraph+Pythia	0.03904	0.65878	1	5000	194.41	2924	0.9
$\tilde{\chi}_1^{\pm}\tilde{\chi}_2^0 \rightarrow WZ(\rightarrow qq\ell\ell)+\not{E}_{\mathrm{T}}\ M_{\tilde{\chi}_1^{\pm}} = 0.4\ \mathrm{TeV} M_{\tilde{\chi}_1^0} = 0.3\ \mathrm{TeV}$	392324	MadGraph+Pythia	0.1048	0.63997	1	10000	149.10	5105	1.5
$\tilde{\chi}_1^{\pm}\tilde{\chi}_2^0 \rightarrow WZ(\rightarrow qq\ell\ell)+\not{E}_{\mathrm{T}}\ M_{\tilde{\chi}_1^{\pm}} = 0.1\ \mathrm{TeV} M_{\tilde{\chi}_1^0} = 0\ \mathrm{TeV}$	392326	MadGraph+Pythia	23.595	0.52366	1	20000	1.62	9697	2.8
$\tilde{\chi}_1^{\pm}\tilde{\chi}_2^0 \rightarrow WZ(\rightarrow qq\ell\ell)+\not{E}_{\mathrm{T}}\ M_{\tilde{\chi}_1^{\pm}} = 0.2\ \mathrm{TeV} M_{\tilde{\chi}_1^0} = 0.1\ \mathrm{TeV}$	392330	MadGraph+Pythia	1.7228	0.59586	1	20000	19.48	9921	2.9
$\tilde{\chi}_1^{\pm}\tilde{\chi}_2^0 \rightarrow WZ(\rightarrow qq\ell\ell)+\not{E}_{\mathrm{T}}\ M_{\tilde{\chi}_1^{\pm}} = 0.5\ \mathrm{TeV} M_{\tilde{\chi}_1^0} = 0.3\ \mathrm{TeV}$	392332	MadGraph+Pythia	0.03899	0.65563	1	5000	195.60	2785	0.9
$\tilde{\chi}_1^{\pm}\tilde{\chi}_2^0 \rightarrow WZ(\rightarrow qq\ell\ell)+\not{E}_{\mathrm{T}}\ M_{\tilde{\chi}_1^{\pm}} = 0.6\ \mathrm{TeV} M_{\tilde{\chi}_1^0} = 0.1\ \mathrm{TeV}$	392354	MadGraph+Pythia	0.01654	0.67099	1	5000	450.52	3023	0.9
$\tilde{\chi}_1^{\pm}\tilde{\chi}_2^0 \rightarrow WZ(\rightarrow qq\ell\ell)+\not{E}_{\mathrm{T}}\ M_{\tilde{\chi}_1^{\pm}} = 0.6\ \mathrm{TeV} M_{\tilde{\chi}_1^0} = 0\ \mathrm{TeV}$	392356	MadGraph+Pythia	0.01656	0.668	1	5000	451.99	3050	0.9
$\tilde{\chi}_1^{\pm}\tilde{\chi}_2^0 \rightarrow WZ(\rightarrow qq\ell\ell)+\not{E}_{\mathrm{T}}\ M_{\tilde{\chi}_1^{\pm}} = 0.7\ \mathrm{TeV} M_{\tilde{\chi}_1^0} = 0.4\ \mathrm{TeV}$	392361	MadGraph+Pythia	0.007672	0.67526	1	4000	772.11	2407	0.8
$\tilde{\chi}_1^{\pm}\tilde{\chi}_2^0 \rightarrow WZ(\rightarrow qq\ell\ell)+\not{E}_{\mathrm{T}}\ M_{\tilde{\chi}_1^{\pm}} = 0.7\ \mathrm{TeV} M_{\tilde{\chi}_1^0} = 0.1\ \mathrm{TeV}$	392364	MadGraph+Pythia	0.007674	0.66687	1	5000	977.03	3006	0.9
$\tilde{\chi}_1^{\pm}\tilde{\chi}_2^0 \rightarrow WZ(\rightarrow qq\ell\ell)+\not{E}_{\mathrm{T}}\ M_{\tilde{\chi}_1^{\pm}} = 0.7\ \mathrm{TeV} M_{\tilde{\chi}_1^0} = 0\ \mathrm{TeV}$	392365	MadGraph+Pythia	0.007676	0.6786	1	5000	959.89	3072	0.9

Table 21: Part (a) of SUSY signal datasets with at least 2 leptons. The factor FE denotes the filter efficiency for a given dataset and f_{k} is used for rescaling the leading order estimate to next to leading order in perturbative QCD. The total size of part (a) of the simulated SUSY datasets with at least 2 leptons is about 50 MB. Though not very big the datasets have been subjected to the same skimming procedure as the 2 lepton SM datasets. $N_{\mathrm{events}}^{\mathrm{generated}}$ denotes the number of events prior to preselection and $N_{\mathrm{events}}^{\mathrm{tuple}}$ is the number of events after preselection and requiring at least 2 leptons per event.

process	DSID	Generator	σ/pb	FE	f_k	$N_{\text{events}}^{\text{generated}}$	$\mathcal{L}/\text{fb}^{-1}$	$N_{\text{events}}^{\text{tuple}}$	size/MB
$\tilde{\chi}_1^+\tilde{\chi}_1^- \to \tilde{\ell}\nu\ell\tilde{\nu} \to \ell\nu\ell\nu + \not{E}_T$ $M_{\tilde{\chi}_1^\pm}$=0.2 TeV$M_{\tilde{\chi}_1^0}$=100 GeV	392501	MadGraph+Pythia	0.93842	0.46715	1	25000	57.03	13632	3.4
$\tilde{\chi}_1^+\tilde{\chi}_1^- \to \tilde{\ell}\nu\ell\tilde{\nu} \to \ell\nu\ell\nu + \not{E}_T$ $M_{\tilde{\chi}_1^\pm}$=0.2 TeV$M_{\tilde{\chi}_1^0}$=150 GeV	392502	MadGraph+Pythia	0.93826	0.40376	1	14000	36.96	4590	1.2
$\tilde{\chi}_1^+\tilde{\chi}_1^- \to \tilde{\ell}\nu\ell\tilde{\nu} \to \ell\nu\ell\nu + \not{E}_T$ $M_{\tilde{\chi}_1^\pm}$=0.3 TeV$M_{\tilde{\chi}_1^0}$=100 GeV	392504	MadGraph+Pythia	0.18834	0.53084	1	24000	240.05	15595	3.9
$\tilde{\chi}_1^+\tilde{\chi}_1^- \to \tilde{\ell}\nu\ell\tilde{\nu} \to \ell\nu\ell\nu + \not{E}_T$ $M_{\tilde{\chi}_1^\pm}$=0.3 TeV$M_{\tilde{\chi}_1^0}$=250 GeV	392506	MadGraph+Pythia	0.18832	0.4276	1	14000	173.86	4617	1.2
$\tilde{\chi}_1^+\tilde{\chi}_1^- \to \tilde{\ell}\nu\ell\tilde{\nu} \to \ell\nu\ell\nu + \not{E}_T$ $M_{\tilde{\chi}_1^\pm}$=0.4 TeV$M_{\tilde{\chi}_1^0}$=100 GeV	392507	MadGraph+Pythia	0.05594	0.55667	1	25000	802.82	16869	4.3
$\tilde{\chi}_1^+\tilde{\chi}_1^- \to \tilde{\ell}\nu\ell\tilde{\nu} \to \ell\nu\ell\nu + \not{E}_T$ $M_{\tilde{\chi}_1^\pm}$=0.4 TeV$M_{\tilde{\chi}_1^0}$=300 GeV	392509	MadGraph+Pythia	0.05598	0.51064	1	25000	874.57	14203	3.6
$\tilde{\chi}_1^+\tilde{\chi}_1^- \to \tilde{\ell}\nu\ell\tilde{\nu} \to \ell\nu\ell\nu + \not{E}_T$ $M_{\tilde{\chi}_1^\pm}$=0.5 TeV$M_{\tilde{\chi}_1^0}$=300 GeV	392513	MadGraph+Pythia	0.02055	0.55669	1	25000	2185	16686	4.3
$\tilde{\chi}_1^+\tilde{\chi}_1^- \to \tilde{\ell}\nu\ell\tilde{\nu} \to \ell\nu\ell\nu + \not{E}_T$ $M_{\tilde{\chi}_1^\pm}$=0.6 TeV$M_{\tilde{\chi}_1^0}$=300 GeV	392517	MadGraph+Pythia	0.0086272	0.56632	1	25000	5117	17380	4.4
$\tilde{\chi}_1^+\tilde{\chi}_1^- \to \tilde{\ell}\nu\ell\tilde{\nu} \to \ell\nu\ell\nu + \not{E}_T$ $M_{\tilde{\chi}_1^\pm}$=0.7 TeV$M_{\tilde{\chi}_1^0}$=1 GeV	392518	MadGraph+Pythia	0.003966	0.58451	1	25000	10784	17666	4.5
$\tilde{\chi}_1^+\tilde{\chi}_1^- \to \tilde{\ell}\nu\ell\tilde{\nu} \to \ell\nu\ell\nu + \not{E}_T$ $M_{\tilde{\chi}_1^\pm}$=0.7 TeV$M_{\tilde{\chi}_1^0}$=300 GeV	392521	MadGraph+Pythia	0.0039656	0.57637	1	25000	10938	17602	4.5
$\tilde{\ell}\tilde{\ell} \to \ell\ell$ + DM $M_{\tilde{\ell}}$ = 100 GeVM_{DM} = 1 GeV	392916	MadGraph+Pythia	1.46	0.55255	1	10000	12.4	5538	1.4
$\tilde{\ell}\tilde{\ell} \to \ell\ell$ + DM $M_{\tilde{\ell}}$ = 200 GeVM_{DM} = 1 GeV	392918	MadGraph+Pythia	0.10125	0.63868	1	8000	123.71	5019	1.3
$\tilde{\ell}\tilde{\ell} \to \ell\ell$ + DM $M_{\tilde{\ell}}$ = 300 GeVM_{DM} = 1 GeV	392920	MadGraph+Pythia	0.018705	0.66521	1	10000	803.68	6594	1.7
$\tilde{\ell}\tilde{\ell} \to \ell\ell$ + DM $M_{\tilde{\ell}}$ = 500 GeVM_{DM} = 1 GeV	392924	MadGraph+Pythia	0.0017865	0.68467	1	9000	7358	6005	1.6
$\tilde{\ell}\tilde{\ell} \to \ell\ell$ + DM $M_{\tilde{\ell}}$ = 100 GeVM_{DM} = 50 GeV	392925	MadGraph+Pythia	1.4885	0.54858	1	10000	12.25	5024	1.3
$\tilde{\ell}\tilde{\ell} \to \ell\ell$ + DM $M_{\tilde{\ell}}$ = 200 GeVM_{DM} = 100 GeV	392936	MadGraph+Pythia	0.1023	0.63191	1	10000	154.69	6164	1.6
$\tilde{\ell}\tilde{\ell} \to \ell\ell$ + DM $M_{\tilde{\ell}}$ = 500 GeVM_{DM} = 100 GeV	392942	MadGraph+Pythia	0.0017945	0.68505	1	10000	8135	6660	1.7
$\tilde{\ell}\tilde{\ell} \to \ell\ell$ + DM $M_{\tilde{\ell}}$ = 300 GeVM_{DM} = 200 GeV	392951	MadGraph+Pythia	0.018885	0.65847	1	10000	804.17	6284	1.6
$\tilde{\ell}\tilde{\ell} \to \ell\ell$ + DM $M_{\tilde{\ell}}$ = 400 GeVM_{DM} = 300 GeV	392962	MadGraph+Pythia	0.00521	0.65989	1	10000	2909	6414	1.6
$\tilde{\ell}\tilde{\ell} \to \ell\ell$ + DM $M_{\tilde{\ell}}$ = 500 GeVM_{DM} = 300 GeV	392964	MadGraph+Pythia	0.001796	0.67467	1	10000	8253	6823	1.8
$\tilde{\ell}\tilde{\ell} \to \ell\ell$ + DM $M_{\tilde{\ell}}$ = 600 GeVM_{DM} = 1 GeV	392982	MadGraph+Pythia	0.00067895	0.68457	1	10000	21515	6787	1.7
$\tilde{\ell}\tilde{\ell} \to \ell\ell$ + DM $M_{\tilde{\ell}}$ = 600 GeVM_{DM} = 300 GeV	392985	MadGraph+Pythia	0.00067915	0.68434	1	10000	21516	6820	1.7
$\tilde{\ell}\tilde{\ell} \to \ell\ell$ + DM $M_{\tilde{\ell}}$ = 700 GeVM_{DM} = 1 GeV	392996	MadGraph+Pythia	0.0002983	0.68634	1	10000	48844	6783	1.7
$\tilde{\ell}\tilde{\ell} \to \ell\ell$ + DM $M_{\tilde{\ell}}$ = 700 GeVM_{DM} = 300 GeV	392999	MadGraph+Pythia	0.0002983	0.68684	1	10000	48808	6898	1.8

Table 22: Part (b) of SUSY signal datasets with at least 2 leptons. The factor FE denotes the filter efficiency for a given dataset and f_k is used for rescaling the leading order estimate to next to leading order in perturbative QCD. The total size of part (b) of the simulated SUSY datasets with at least 2 leptons is about 60 MB. Though not very big the datasets have been subjected to the same skimming procedure as the 2 lepton SM datasets. $N_{\text{events}}^{\text{generated}}$ denotes the number of events prior to preselection and $N_{\text{events}}^{\text{tuple}}$ is the number of events after preselection and requiring at least 2 leptons per event.

process	DSID	Generator	σ/pb	FE	f_k	$N_\mathrm{events}^\mathrm{generated}$	$\mathcal{L}/\mathrm{fb}^{-1}$	$N_\mathrm{events}^\mathrm{tuple}$	size/MB
single top t-chan	410011	PowHeg+Pythia	43.739	1	1.0094	4986200	112.94	36847	14.2
single antitop t-chan	410012	PowHeg+Pythia	25.778	1	1.0193	4989800	189.90	38934	15.0

Table 23: Full truth information datasets with at least lepton. The factor FE denotes the filter efficiency for a given dataset and f_k is used for rescaling the leading order estimate to next to leading order in perturbative QCD. The total size of all truth information 1 datasets with at least 2 leptons is about 30 MB. The datasets have been subjected to the same skimming procedure as the 2 lepton SM datasets. $N_\mathrm{events}^\mathrm{generated}$ denotes the number of events prior to preselection and $N_\mathrm{events}^\mathrm{tuple}$ is the number of events after preselection.

process	DSID	Generator	σ/pb	FE	f_k	$N_{\text{events}}^{\text{generated}}$	$\mathcal{L}$/fb^{-1}	$N_{\text{events}}^{\text{tuple}}$	size/MB
ttH	341081	Herwig	0.00115938	0.00228	1	927400	350837850	576492	225
$gg \to H$	343981	PowHeg+Pythia	28.3	0.00228	1.580807142	1976000	19373	1054714	231
VBFH	345041	PowHeg+Pythia	3.7363	0.00228	1	921000	108114	497470	114
WH	345318	PowHeg+Pythia	0.86204	0.00228	1	248000	126180	113765	32
ZH	345319	PowHeg+Pythia	0.76087	0.00228	1	471000	271504	230901	63
$G \to \gamma\gamma$ M_{G} = 500 GeV	302460	Pythia	34.771	1	1	20000	0.58	x	x
$G \to \gamma\gamma$ M_{G} = 1000 GeV	302461	Pythia	1.2256	1	1	20000	16.32	x	x
$G \to \gamma\gamma$ M_{G} = 2000 GeV	302462	Pythia	0.024346	1	1	20000	82.15	x	x
$G \to \gamma\gamma$ M_{G} = 3000 GeV	302463	Pythia	0.001505	1	1	20000	13289	x	x
$G \to \gamma\gamma$ M_{G} = 4000 GeV	302464	Pythia	0.00013698	1	1	20000	146007	x	x

Table 24: Full $H \to \gamma\gamma$ signal datasets in the diphoton collection. The factor FE denotes the filter efficiency for a given dataset and f_k is used for rescaling the leading order estimate to next to leading order in perturbative QCD. The total size of all simulated datasets in the diphoton collection is about ~1 GB. $N_{\text{events}}^{\text{generated}}$ denotes the number of events prior to preselection and $N_{\text{events}}^{\text{tuple}}$ is the number of events after preselection.

process	DSID	Generator	σ/pb	FE	f_k	$N_{\text{events}}^{\text{generated}}$	$\mathcal{L}/\text{fb}^{-1}$	$N_{\text{events}}^{\text{tuple}}$	size/MB
$t\bar{t} \to \ell + X$	410000	PowHeg+Pythia	696.11	0.5442	1.195	49296600	108.90	529068	183.53
single top t-chan	410011	PowHeg+Pythia	43.739	1	1.0094	1996600	45.22	4202	1.28
single antitop t-chan	410012	PowHeg+Pythia	25.778	1	1.0193	1994200	75.9	4911	1.49
single top wt-chan	410013	PowHeg+Pythia	34.009	1	1.054	1994200	55.63	12785	4.23
single antitop wt-chan	410014	PowHeg+Pythia	33.989	1	1.054	1994000	55.66	12564	4.18
single top s-chan	410025	PowHeg+Pythia	2.0514	1	1.0048	997800	484.08	1785	0.57
single antitop s-chan	410026	PowHeg+Pythia	1.2615	1	1.0215	995400	772.45	1897	0.60
Z+Jets ee	361106	PowHeg+Pythia	1901.1	1	1.026	61106597	31.33	280913	72.71
Z+Jets $\mu\mu$	361107	PowHeg+Pythia	1901.1	1	1.0261	1998400	1.02	2327	0.69
Z+Jets $\tau\tau$	361108	PowHeg+Pythia	1901.1	1	1.0261	29546000	15.15	27362	7.34
W^++Jets $e\nu$	361100	PowHeg+Pythia	11306.0	1	1.0172	25544800	2.22	13674	3.55
W^++Jets $\mu\nu$	361101	PowHeg+Pythia	11306.0	1	1.0172	1996000	0.17	1447	0.38
W^++Jets $\tau\nu$	361102	PowHeg+Pythia	11306.0	1	1.0172	1979400	0.17	107	0.05
W^-+Jets $e\nu$	361103	PowHeg+Pythia	8283.1	1	1.0358	17905400	2.09	20474	5.23
W^-+Jets $\mu\nu$	361104	PowHeg+Pythia	8283.1	1	1.0358	1997000	0.23	1627	0.42
W^-+Jets $\tau\nu$	361105	PowHeg+Pythia	8283.1	1	1.0358	1999800	0.23	174	0.07
$ZZ \to qq\ell^+\ell^-$	363356	Sherpa	15.564	1	0.14158	5317000	2413		
$WZ \to qq\ell^+\ell^-$	363358	Sherpa	3.4328	1	1	5124000	1493		
$W_p W_m \to qq\ell\nu_\ell$	363359	Sherpa	24.708	1	1	6673000	270.07		
$W_p W_m \to \ell\nu_\ell qq$	363360	Sherpa	24.724	1	1	7115000	287.78		
$WZ \to \ell\nu_\ell qq$	363489	Sherpa	11.42	1	1	7100000	621.72		
$ZZ \to \ell^+\ell^-\ell^+\ell^-$	363490	Sherpa	1.2578	1	1	17825300	14172		
$WZ \to \ell\nu_\ell\ell^+\ell^-$	363491	Sherpa	4.6049	1	1	15772084	3425		
$ZZ \to \ell^+\ell^-\nu\nu$	363492	Sherpa	12.466	1	1	14803000	1187		
$WZ \to \ell\nu_\ell\nu\nu$	363493	Sherpa	3.2286	1	1	5922600	1834		
$gg \to H \to \tau\tau$	341123	PowHeg+Pythia	1.90820026	0.4548	1.4546	1446900	1146	84632	23.60
VBF$H \to \tau\tau$	341156	PowHeg+Pythia	0.24201136	0.45539	0.9788	2087900	19355	81876	23.79

Table 25: Full top, Z, W, diboson and $H \to \tau\tau$ simulated datasets in the 1 hadronic τ + 1 leptonic τ collection. The factor FE denotes the filter efficiency for a given dataset and f_k is used for rescaling the leading order estimate to next to leading order in perturbative QCD. The total size of all top, Z, W, diboson and $H \to \tau\tau$ simulated datasets in the 1 hadronic τ + 1 leptonic τ collection is about 330 MB. $N_{\text{events}}^{\text{generated}}$ denotes the number of events prior to preselection and $N_{\text{events}}^{\text{tuple}}$ is the number of events after preselection.

process	DSID	Generator	σ/pb	FE	f_k	$N_{\text{events}}^{\text{generated}}$	$\mathcal{L}$/fb^{-1}	$N_{\text{events}}^{\text{tuple}}$	size/MB
$Z \to \mu^+\mu^-$ $M_{\mu\mu}$ = 10 – 40 GeVb-veto 0-70	364198	Sherpa	2414.3	0.96536	0.9751				
$Z \to \mu^+\mu^-$ $M_{\mu\mu}$ = 10 – 40 GeVwith b 0-70	364199	Sherpa	2414.2	0.034445	0.9751				
$Z \to \mu^+\mu^-$ $M_{\mu\mu}$ = 10 – 40 GeVb-veto 70-280	364200	Sherpa	50.33	0.89306	0.9751				
$Z \to \mu^+\mu^-$ $M_{\mu\mu}$ = 10 – 40 GeVwith b 70-280	364201	Sherpa	50.29	0.11212	0.9751				
$Z \to \mu^+\mu^-$ $M_{\mu\mu}$ = 10 – 40 GeVb-veto 280+	364202	Sherpa	3.2398	0.85373	0.9751				
$Z \to \mu^+\mu^-$ $M_{\mu\mu}$ = 10 – 40 GeVwith b 280+	364203	Sherpa	3.2813	0.16027	0.9751				
$Z \to e^+e^-$ M_{ee} = 10 – 40 GeVb-veto 0-70	364204	Sherpa	2415.3	0.96520	0.9751				
$Z \to e^+e^-$ M_{ee} = 10 – 40 GeVwith b 0-70	364205	Sherpa	2415.5	0.034696	0.9751				
$Z \to e^+e^-$ M_{ee} = 10 – 40 GeVb-veto 70-280	364206	Sherpa	50.354	0.89320	0.9751				
$Z \to e^+e^-$ M_{ee} = 10 – 40 GeVwith b 70-280	364207	Sherpa	50.487	0.1087	0.9751				
$Z \to e^+e^-$ M_{ee} = 10 – 40 GeVb-veto 280+	364208	Sherpa	3.2539	0.85485	0.9751				
$Z \to e^+e^-$ M_{ee} = 10 – 40 GeVwith b 280+	364209	Sherpa	3.2526	0.15351	0.9751				
$Z \to \tau^+\tau^-$ $M_{\tau\tau}$ = 10 – 40 GeVb-veto 0-70	364210	Sherpa	2416.0	0.96534	0.9751				
$Z \to \tau^+\tau^-$ $M_{\tau\tau}$ = 10 – 40 GeVwith b 0-70	364211	Sherpa	2414.4	0.034676	0.9751				
$Z \to \tau^+\tau^-$ $M_{\tau\tau}$ = 10 – 40 GeVb-veto 70-280	364212	Sherpa	50.376	0.89310	0.9751				
$Z \to \tau^+\tau^-$ $M_{\tau\tau}$ = 10 – 40 GeVwith b 70-280	364213	Sherpa	50.468	0.10985	0.9751				
$Z \to \tau^+\tau^-$ $M_{\tau\tau}$ = 10 – 40 GeVb-veto 280+	364214	Sherpa	3.2852	0.85552	0.9751				
$Z \to \tau^+\tau^-$ $M_{\tau\tau}$ = 10 – 40 GeVwith b 280+	364215	Sherpa	3.2813	0.15581	0.9751				

Table 26: Full low-mass Drell-Yan simulated datasets in the 1 hadronic τ + 1 leptonic τ collection. The factor FE denotes the filter efficiency for a given dataset and f_k is used for rescaling the leading order estimate to next to leading order in perturbative QCD. The total size of all low-mass Drell-Yan simulated datasets in the 1 hadronic τ + 1 leptonic τ collection is about x MB. $N_{\text{events}}^{\text{generated}}$ denotes the number of events prior to preselection and $N_{\text{events}}^{\text{tuple}}$ is the number of events after preselection.

process	DSID	Generator	σ/pb	FE	f_k	$N_{\text{events}}^{\text{generated}}$	$\mathcal{L}$/fb^{-1}	$N_{\text{events}}^{\text{tuple}}$	size/MB
Z+Jets $\tau\tau$	361108	PowHeg+Pythia	1901.1	1	1.0261	29546000	15.15	20999	6
$gg \to H \to \tau\tau$	341124	PowHeg+Pythia	1.90819703	0.42194	1.4546	1543100	1318	26556	7
VBF$H \to \tau\tau$	341157	PowHeg+Pythia	0.24201165	0.42191	0.9788	2089900	20911	81307	24

Table 27: Full simulated datasets in the 2 hadronic τ + 0 lepton collection. The factor FE denotes the filter efficiency for a given dataset and f_k is used for rescaling the leading order estimate to next to leading order in perturbative QCD. The total size of all simulated datasets in the 2 hadronic τ + 0 lepton collection is about 40 MB. $N_{\text{events}}^{\text{generated}}$ denotes the number of events prior to preselection and $N_{\text{events}}^{\text{tuple}}$ is the number of events after preselection.

process	DSID	Generator	σ/pb	FE	f_k	$N_{\text{events}}^{\text{generated}}$	$\mathcal{L}$/fb^{-1}	$N_{\text{events}}^{\text{tuple}}$	size/MB
W^++Jets $\tau\nu$	361102	PowHeg+Pythia	11306.0	1	1.0172	59343600	5.16	2462	0.91
W^-+Jets $\tau\nu$	361105	PowHeg+Pythia	8283.1	1	1.0358	19945400	2.32	2509	0.90
W+Jets $\tau\nu$ no jets $b$veto 0-70	364184	Sherpa	19155	0.82462	0.9702	24784000	11668	94	0.06
W+Jets $\tau\nu$ with jets $b$veto	364185	Sherpa	19149	0.13152	0.9702	9865600	6719	41	0.03
W+Jets $\tau\nu$ with b 0-70	364186	Sherpa	19148	0.045111	0.9702	17273200	11670	83	0.07
W+Jets $\tau\nu$ no jets $b$veto 70-140	364187	Sherpa	945.02	0.67562	0.9702	14808500	320146	1023	0.27
W+Jets $\tau\nu$ with jets $b$veto 70-140	364188	Sherpa	946.23	0.24247	0.9702	9860000	250727	982	0.26
W+Jets $\tau\nu$ with b 70-140	364189	Sherpa	945.71	0.10391	0.9702	9857000	230455	847	0.23
W+Jets $\tau\nu$ no jets $b$veto 140-280	364190	Sherpa	339.78	0.59872	0.9702	9899000	1121004	32582	8.05
W+Jets $\tau\nu$ with jets $b$veto 140-280	364191	Sherpa	339.66	0.28477	0.9702	7415000	750555	23622	6.17
W+Jets $\tau\nu$ with b 140-280	364192	Sherpa	339.35	0.10599	0.9702	24595900	2201658	60101	15.79
W+Jets $\tau\nu$ no jets $b$veto 280-500	364193	Sherpa	72.091	0.54837	0.9702	4931200	892504	61268	16.04
W+Jets $\tau\nu$ with jets $b$veto 280-500	364194	Sherpa	71.994	0.31881	0.9702	2956400	512309	31643	8.71
W+Jets $\tau\nu$ with b 280-500	364195	Sherpa	71.945	0.13597	0.9702	2954100	478683	26023	7.37
W+Jets $\tau\nu$ 500-1000	364196	Sherpa	15.052	1	0.9702	5945000	1277937	85971	25.42
W+Jets $\tau\nu$ >1000	364197	Sherpa	1.2341	1	0.9702	3946000	1148849	62567	20.44

Table 28: Full simulated datasets in the 1 hadronic τ + 0 lepton collection. The factor FE denotes the filter efficiency for a given dataset and f_k is used for rescaling the leading order estimate to next to leading order in perturbative QCD. The total size of all simulated datasets in the 1 hadronic τ + 0 lepton collection is about 110 MB. $N_{\text{events}}^{\text{generated}}$ denotes the number of events prior to preselection and $N_{\text{events}}^{\text{tuple}}$ is the number of events after preselection.

process	DSID	Generator	σ/pb	FE	f_k	$N_{\text{events}}^{\text{generated}}$	$\mathcal{L}/\text{fb}^{-1}$	$N_{\text{events}}^{\text{tuple}}$	size/MB
$t\bar{t} \to \ell + X$	410000	PowHeg+Pythia	696.11	0.5442	1.195	50566600	111.70	~19206795	2
single top t-chan	410011	PowHeg+Pythia	43.739	1	1.0094	2276000	51.55	34019	12.1
single antitop t-chan	410012	PowHeg+Pythia	25.778	1	1.0193	4989800	189.90	28713	10.2
single top wt-chan	410013	PowHeg+Pythia	34.009	1	1.054	4985800	135.74	72198	28.3
single antitop wt-chan	410014	PowHeg+Pythia	33.989	1	1.054	4985600	138.05	71531	28.0
single top s-chan	410025	PowHeg+Pythia	2.0514	1	1.0048	997800	484.08	11461	4.1
single antitop s-chan	410026	PowHeg+Pythia	1.2615	1	1.0215	319400	247.86	2068	0.8
Z+Jets ee	361106	PowHeg+Pythia	1901.1	1	1.026	79045597	40.53	~16442978	~3220
Z+Jets $\mu\mu$	361107	PowHeg+Pythia	1901.1	1	1.0261	77497800	39.73	22027	7.9
Z+Jets $\tau\tau$	361108	PowHeg+Pythia	1901.1	1	1.0261	29546000	15.15	3415	1.2
W^++Jets $e\nu$	361100	PowHeg+Pythia	11306.0	1	1.0172	41870000	3.64	9943	3.2
W^++Jets $\mu\nu$	361101	PowHeg+Pythia	11306.0	1	1.0172	39493600	3.43	7965	2.7
W^++Jets $\tau\nu$	361102	PowHeg+Pythia	11306.0	1	1.0172	59343600	5.16	1392	0.7
W^-+Jets $e\nu$	361103	PowHeg+Pythia	8283.1	1	1.0358	29886000	3.48	7522	2.5
W^-+Jets $\mu\nu$	361104	PowHeg+Pythia	8283.1	1	1.0358	31915400	3.72	5785	1.9
W^-+Jets $\tau\nu$	361105	PowHeg+Pythia	8283.1	1	1.0358	19945400	2.32	962	0.5
$WW \to \ell\nu\ell\nu$	361600	PowHeg+Pythia	10.636	1	1	5926000	557.16	51368	17.9
$WZ \to \ell\nu\ell\ell$	361601	PowHeg+Pythia	4.5106	1	1	4985000	1105	39659	15.4
$WZ \to \ell\nu\nu\nu$	361602	PowHeg+Pythia	2.7781	1	1	982600	353.69	10402	3.2
$ZZ \to \ell\ell\ell\ell$	361603	PowHeg+Pythia	1.2663	1	1	3920000	3096	7178	2.9
$ZZ \to \nu\nu\ell\ell$	361604	PowHeg+Pythia	0.92296	1	1	981000	1063	4714	1.6
$WW \to \ell\nu qq$	361606	PowHeg+Pythia	44.177	1	1	4343000	98.31	33886	12.0
$WZ \to qq\ell\ell$	361607	PowHeg+Pythia	3.2882	1	1	1469000	446.75	27751	11.1
$WZ \to \ell\nu qq$	361609	PowHeg+Pythia	10.095	1	1	9693000	960.18	127465	45.7
$ZZ \to qq\ell\ell$	361610	PowHeg+Pythia	2.2761	1	1	3933000	1728	38860	15.3

Table 29: Full top, Z, W and diboson datasets in the lepton + fatjet collection. The factor FE denotes the filter efficiency for a given dataset and f_k is used for rescaling the leading order estimate to next to leading order in perturbative QCD. The total size of all simulated top, Z, W and diboson datasets in the lepton + fatjet collection is about ~XXX GB. $N_{\text{events}}^{\text{generated}}$ denotes the number of events prior to preselection and $N_{\text{events}}^{\text{tuple}}$ is the number of events after preselection.

process	DSID	Generator	σ/pb	FE	f_k	$N_{\text{events}}^{\text{generated}}$	$\mathcal{L}$/fb^{-1}	$N_{\text{events}}^{\text{tuple}}$	size/MB
W+Jets $\mu\nu$ no jets $b$veto 0-70	364156	Sherpa	19149	0.8245	0.9702	24723000	1.61	102	0.12
W+Jets $\mu\nu$ with jets $b$veto	364157	Sherpa	19142	0.13091	0.9702	9847000	4.05	53	0.06
W+Jets $\mu\nu$ with b 0-70	364158	Sherpa	19138	0.044641	0.9702	17226200	20.78	107	0.10
W+Jets $\mu\nu$ no jets $b$veto 70-140	364159	Sherpa	945.52	0.6743	0.9702	14788000	23.91	144	0.13
W+Jets $\mu\nu$ with jets $b$veto 70-140	364160	Sherpa	945.44	0.24335	0.9702	9853800	44.14	125	0.12
W+Jets $\mu\nu$ with b 70-140	364161	Sherpa	944.14	0.084708	0.9702	19639000	253.1	265	0.26
W+Jets $\mu\nu$ no jets $b$veto 140-280	364162	Sherpa	339.7	0.60292	0.9702	9882000	49.73	32435	9.09
W+Jets $\mu\nu$ with jets $b$veto 140-280	364163	Sherpa	339.88	0.29253	0.9702	7408000	76.8	18215	5.40
W+Jets $\mu\nu$ with b 140-280	364164	Sherpa	339.64	0.110298	0.9702	24585000	676.43	55473	16.89
W+Jets $\mu\nu$ no jets $b$veto 280-500	364165	Sherpa	72.079	0.54768	0.9702	4940000	128.98	443369	130.00
W+Jets $\mu\nu$ with jets $b$veto 280-500	364166	Sherpa	72.1	0.32016	0.9702	2958000	132.08	187891	57.96
W+Jets $\mu\nu$ with b 280-500	364167	Sherpa	72.058	0.12542	0.9702	2919500	332.96	163974	50.70
W+Jets $\mu\nu$ 500-1000	364168	Sherpa	15.006	1	0.9702	5910500	405.97	1889106	651.62
W+Jets $\mu\nu$ >1000	364169	Sherpa	1.2348	1	0.9702	3959000	3305	1822749	710.65
W+Jets $e\nu$ no jets $b$veto 0-70	364170	Sherpa	19153	0.82467	0.9702	24740000	1.61	220	0.17
W+Jets $e\nu$ with jets $b$veto	364171	Sherpa	19145	0.13086	0.9702	9853500	4.05	112	0.10
W+Jets $e\nu$ with b 0-70	364172	Sherpa	19138	0.04482	0.9702	17242400	20.72	194	0.13
W+Jets $e\nu$ no jets $b$veto 70-140	364173	Sherpa	944.98	0.67483	0.9702	14660500	23.7	511	0.25
W+Jets $e\nu$ with jets $b$veto 70-140	364174	Sherpa	945.74	0.24413	0.9702	9818400	43.83	306	0.12
W+Jets $e\nu$ with b 70-140	364175	Sherpa	945.79	0.10341	0.9702	5401900	103.3	233	0.12
W+Jets $e\nu$ no jets $b$veto 140-280	364176	Sherpa	339.67	0.59875	0.9702	9879000	50.07	46844	13.44
W+Jets $e\nu$ with jets $b$veto 140-280	364177	Sherpa	339.87	0.28876	0.9702	7360000	77.82	28933	8.71
W+Jets $e\nu$ with b 140-280	364178	Sherpa	339.64	0.10898	0.9702	24677800	687.19	87012	26.80
W+Jets $e\nu$ no jets $b$veto 280-500	364179	Sherpa	72.074	0.5483	0.9702	4923800	128.42	550845	164.90
W+Jets $e\nu$ with jets $b$veto 280-500	364180	Sherpa	72.105	0.31969	0.9702	2963400	132.51	250094	77.70
W+Jets $e\nu$ with b 280-500	364181	Sherpa	72.091	0.13706	0.9702	2958000	308.56	219286	71.32
W+Jets $e\nu$ 500-1000	364182	Sherpa	15.047	1	0.9702	5911800	405.3	964639	337.66
W+Jets $e\nu$ >1000	364183	Sherpa	1.2344	1	0.9702	3947000	3296	1977331	801.62
W+Jets $\tau\nu$ no jets $b$veto 0-70	364184	Sherpa	19155	0.82462	0.9702	24784000	1.62	7	0.02
W+Jets $\tau\nu$ with jets $b$veto	364185	Sherpa	19149	0.13152	0.9702	9865600	4.04	5	0.02
W+Jets $\tau\nu$ with b 0-70	364186	Sherpa	19148	0.045111	0.9702	17273200	20.61	3	0.02
W+Jets $\tau\nu$ no jets $b$veto 70-140	364187	Sherpa	945.02	0.67562	0.9702	14808500	23.91	22	0.05
W+Jets $\tau\nu$ with jets $b$veto 70-140	364188	Sherpa	946.23	0.24247	0.9702	9270000	41.65	x	x
W+Jets $\tau\nu$ with b 70-140	364189	Sherpa	945.71	0.10391	0.9702	9857000	103.39	18	0.03
W+Jets $\tau\nu$ no jets $b$veto 140-280	364190	Sherpa	339.78	0.59872	0.9702	9899000	50.15	7469	2.11
W+Jets $\tau\nu$ with jets $b$veto 140-280	364191	Sherpa	339.66	0.28477	0.9702	7405000	78.91	4329	1.30
W+Jets $\tau\nu$ with b 140-280	364192	Sherpa	339.35	0.10599	0.9702	24819900	711.26	3040	0.95
W+Jets $\tau\nu$ no jets $b$veto 280-500	364193	Sherpa	72.091	0.54837	0.9702	4931200	128.57	103499	30.77
W+Jets $\tau\nu$ with jets $b$veto 280-500	364194	Sherpa	71.994	0.31881	0.9702	2956400	132.76	44839	13.95
W+Jets $\tau\nu$ with b 280-500	364195	Sherpa	71.945	0.13597	0.9702	2954100	311.26	37962	12.09
W+Jets $\tau\nu$ 500-1000	364196	Sherpa	15.052	1	0.9702	5945000	407.1	338738	117.00
W+Jets $\tau\nu$ >1000	364197	Sherpa	1.2341	1	0.9702	3946000	3296	319788	125.76

Table 30: Full sliced W datasets in the lepton + fatjet collection. The factor FE denotes the filter efficiency for a given dataset and f_k is used for rescaling the leading order estimate to next to leading order in perturbative QCD. The total size of all simulated sliced W datasets in the lepton + fatjet collection is about ~3 GB. $N_{\text{events}}^{\text{generated}}$ denotes the number of events prior to preselection and $N_{\text{events}}^{\text{tuple}}$ is the number of events after preselection.

process	DSID	Generator	σ/pb	FE	f_k	$N_{\text{events}}^{\text{generated}}$	$\mathcal{L}/\text{fb}^{-1}$	$N_{\text{events}}^{\text{tuple}}$	size/MB
$Z' \to t\bar{t}$ $M_{Z'}$ = 400 GeV	301322	Pythia	8.9857	1	1	199200	22.17	790	0.3
$Z' \to t\bar{t}$ $M_{Z'}$ = 500 GeV	301323	Pythia	8.7385	1	1	199600	22.84	1446	0.6
$Z' \to t\bar{t}$ $M_{Z'}$ = 750 GeV	301324	Pythia	3.1201	1	1	199000	63.78	15300	5.9
$Z' \to t\bar{t}$ $M_{Z'}$ = 1000 GeV	301325	Pythia	1.1261	1	1	199800	177.43	29427	11.5
$Z' \to t\bar{t}$ $M_{Z'}$ = 1250 GeV	301326	Pythia	0.45981	1	1	199200	433.22	35494	14.1
$Z' \to t\bar{t}$ $M_{Z'}$ = 1500 GeV	301327	Pythia	0.20685	1	1	198800	961.08	36985	14.8
$Z' \to t\bar{t}$ $M_{Z'}$ = 1750 GeV	301328	Pythia	0.10016	1	1	199800	1995	36920	15.0
$Z' \to t\bar{t}$ $M_{Z'}$ = 2000 GeV	301329	Pythia	0.051346	1	1	199800	3891	35353	14.4
$Z' \to t\bar{t}$ $M_{Z'}$ = 2250 GeV	301330	Pythia	0.027481	1	1	199200	7249	33813	13.9
$Z' \to t\bar{t}$ $M_{Z'}$ = 2500 GeV	301331	Pythia	0.015226	1	1	198200	13017	31832	13.1
$Z' \to t\bar{t}$ $M_{Z'}$ = 2750 GeV	301332	Pythia	0.0086884	1	1	199800	22996	29764	12.4
$Z' \to t\bar{t}$ $M_{Z'}$ = 3000 GeV	301333	Pythia	0.0050843	1	1	195800	38511	27368	11.3
$\tilde{g}\tilde{g} \to t\bar{t}t\bar{t} \to tttt$ +DM $M_{\tilde{g}}$ = 1.2 TeV M_{DM} = 1 GeV	370114	MadGraph+Pythia	0.057037	1	1	100000	1753	35809	19.8
$\tilde{g}\tilde{g} \to t\bar{t}t\bar{t} \to tttt$ +DM $M_{\tilde{g}}$ = 1.2 TeV M_{DM} = 600 GeV	370118	MadGraph+Pythia	0.057002	1	1	100000	1754	26245	13.9
$\tilde{g}\tilde{g} \to t\bar{t}t\bar{t} \to tttt$ +DM $M_{\tilde{g}}$ = 1.4 TeV M_{DM} = 1 GeV	370129	MadGraph+Pythia	0.015756	1	1	100000	6347	36476	20.3
$\tilde{g}\tilde{g} \to t\bar{t}t\bar{t} \to tttt$ +DM $M_{\tilde{g}}$ = 1.6 TeV M_{DM} = 1 GeV	370144	MadGraph+Pythia	0.004747	1	1	99000	20855	35951	20.0
$\tilde{t}\tilde{t} \to tt$ + DM $M_{\tilde{t}}$ = 450 GeV M_{DM} = 1 GeV	388240	MadGraph+Pythia	0.88424	1	1	50000	56.55	3896	1.6
$\tilde{t}\tilde{t} \to tt$ + DM $M_{\tilde{t}}$ = 500 GeV M_{DM} = 1 GeV	387154	MadGraph+Pythia	0.46603	1	1	20000	42.92	1921	0.8
$\tilde{t}\tilde{t} \to tt$ + DM $M_{\tilde{t}}$ = 500 GeV M_{DM} = 200 GeV	387157	MadGraph+Pythia	0.46702	1	1	50000	107.06	3422	1.4
$\tilde{t}\tilde{t} \to tt$ + DM $M_{\tilde{t}}$ = 600 GeV M_{DM} = 1 GeV	387163	MadGraph+Pythia	0.15518	1	1	49000	315.76	6483	2.6
$\tilde{\chi}_1^{\pm}\tilde{\chi}_2^0 \to WZ(\to \ell\nu\ell\ell)+\not{E}_T$ $M_{\tilde{\chi}_1^{\pm}}$ =100 GeV $M_{\tilde{\chi}_1^0}$=0 GeV	392226	MadGraph+Pythia	23.525	0.67285	1	20000	1.26	317	0.1
$\tilde{\chi}_1^{\pm}\tilde{\chi}_2^0 \to WZ(\to \ell\nu\ell\ell)+\not{E}_T$ $M_{\tilde{\chi}_1^{\pm}}$ =350 GeV $M_{\tilde{\chi}_1^0}$=0 GeV	392220	MadGraph+Pythia	0.18475	0.76781	1	10000	70.50	1613	0.6
$\tilde{\chi}_1^{\pm}\tilde{\chi}_2^0 \to WZ(\to \ell\nu\ell\ell)+\not{E}_T$ $M_{\tilde{\chi}_1^{\pm}}$ =400 GeV $M_{\tilde{\chi}_1^0}$=0 GeV	392217	MadGraph+Pythia	0.10480	0.76994	1	10000	123.93	1995	0.7
$\tilde{\chi}_1^{\pm}\tilde{\chi}_2^0 \to WZ(\to \ell\nu\ell\ell)+\not{E}_T$ $M_{\tilde{\chi}_1^{\pm}}$ =500 GeV $M_{\tilde{\chi}_1^0}$=0 GeV	392223	MadGraph+Pythia	0.03899	0.77285	1	5000	165.93	1402	0.5

Table 31: Full BSM signal datasets in the lepton + fatjet collection. The factor FE denotes the filter efficiency for a given dataset and f_k is used for rescaling the leading order estimate to next to leading order in perturbative QCD. The total size of all simulated BSM datasets in the lepton + fatjet collection is about 210 MB. $N_{\text{events}}^{\text{generated}}$ denotes the number of events prior to preselection and $N_{\text{events}}^{\text{tuple}}$ is the number of events after preselection.

References

[1] ATLAS Collaboration, *ATLAS Data Access Policy*, http://opendata.cern.ch/record/413, 2014.

[2] URL: http://opendata.atlas.cern.

[3] URL: http://opendata.cern.ch.

[4] F. Socher et al., *Proposal for an ATLAS endorsed data set for Outreach Purposes*, tech. rep. ATL-COM-OREACH-2015-001, CERN, 2015, URL: https://cds.cern.ch/record/2034188.

[5] *Review Studies for the ATLAS Open Data Dataset*, tech. rep. ATL-OREACH-PUB-2016-001, CERN, 2016, URL: https://cds.cern.ch/record/2203649.

[6] URL: http://atlas.cern/updates/atlas-news/explore-lhc-data-new-atlas-educational-platform.

[7] *Review of ATLAS Open Data 8 TeV datasets, tools and activities*, tech. rep. ATL-OREACH-PUB-2018-001, CERN, 2018, URL: https://cds.cern.ch/record/2624572.

[8] R. Brun and F. Rademakers, *ROOT — An object oriented data analysis framework*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **389** (1997) 81, New Computing Techniques in Physics Research V, ISSN: 0168-9002, URL: http://www.sciencedirect.com/science/article/pii/S016890029700048X.

[9] M. Aaboud et al., *Measurement of the photon identification efficiencies with the ATLAS detector using LHC Run-1 data*, Eur. Phys. J. **C76** (2016) 666, arXiv: 1606.01813 [hep-ex].

[10] *The CxAODFramework physics analysis framework*, URL: https://twiki.cern.ch/twiki/bin/view/AtlasProtected/CxAODFramework#Introduction.

[11] *Framework to produce CxAODs from DAODs for 13 TeV ATLAS Open Data.* URL: https://gitlab.cern.ch/meevans/OpenCxAOD8.

[12] *Goldcrest setup tag for the $HH \rightarrow bb\tau\tau$ version of the CxAOD Framework*, URL: https://gitlab.cern.ch/CxAODFramework/FrameworkSub_HH_bbtautau/tree/505d5a37580df1661b5c987556eb3dd36333e230.

[13] A. Elagin, P. Murat, A. Pranko, and A. Safonov, *A new mass reconstruction technique for resonances decaying to $\tau\tau$*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **654** (2011) 481, ISSN: 0168-9002, URL: http://www.sciencedirect.com/science/article/pii/S0168900211014112.

[14] *ATLAS Outreach data and tools GitHub account*, URL: https://github.com/atlas-outreach-data-tools.

[15] *Derivation grid sample lists*, URL: https://gitlab.cern.ch/meevans/OpenCxAOD8/tree/master/FrameworkSub_HH_bbtautau/In.

[16] *ROOT Release 6.04/16 - 2016-03-17*, URL: https://root.cern.ch/content/release-60416.

[17] *Advanced research projects*,
URL: https://www.mn.uio.no/fysikk/english/research/projects/zpath/research-based-education/advanced-research-projects.html.

[18] D. Fassouliotis, C. Kourkoumelis, and S. Vourakis,
Introducing HEP to high-school and university students through ATLAS event analysis tools,
EPJ Web of Conferences **164** (07018 2017).

[19] URL: https://twiki.cern.ch/twiki/bin/view/AtlasProtected/CentralMC15ProductionList#Overview_of_all_DSIDs_with_corre.

[20] *AMI service for ATLAS*, URL: https://ami.in2p3.fr/.

[21] J. R. Andersen et al., *Handbook of LHC Higgs Cross Sections: 3. Higgs Properties*,
(2013), ed. by S. Heinemeyer, C. Mariotti, G. Passarino, and R. Tanaka,
arXiv: 1307.1347 [hep-ph].