

(Re)interpretation of the LHC
results for new physics:

Publication and reuse of ML models

7th workshop of the
LHC Reinterpretation Forum

12 Dec 2022

CERN and online

Simulation-based reinterpretation (“recasting”)

- ▶ Aims at reproducing experimental analyses in Monte Carlo simulation
- ▶ Nowadays well established for traditional cut-based analyses. Information needed: cf. arXiv:2003.07868

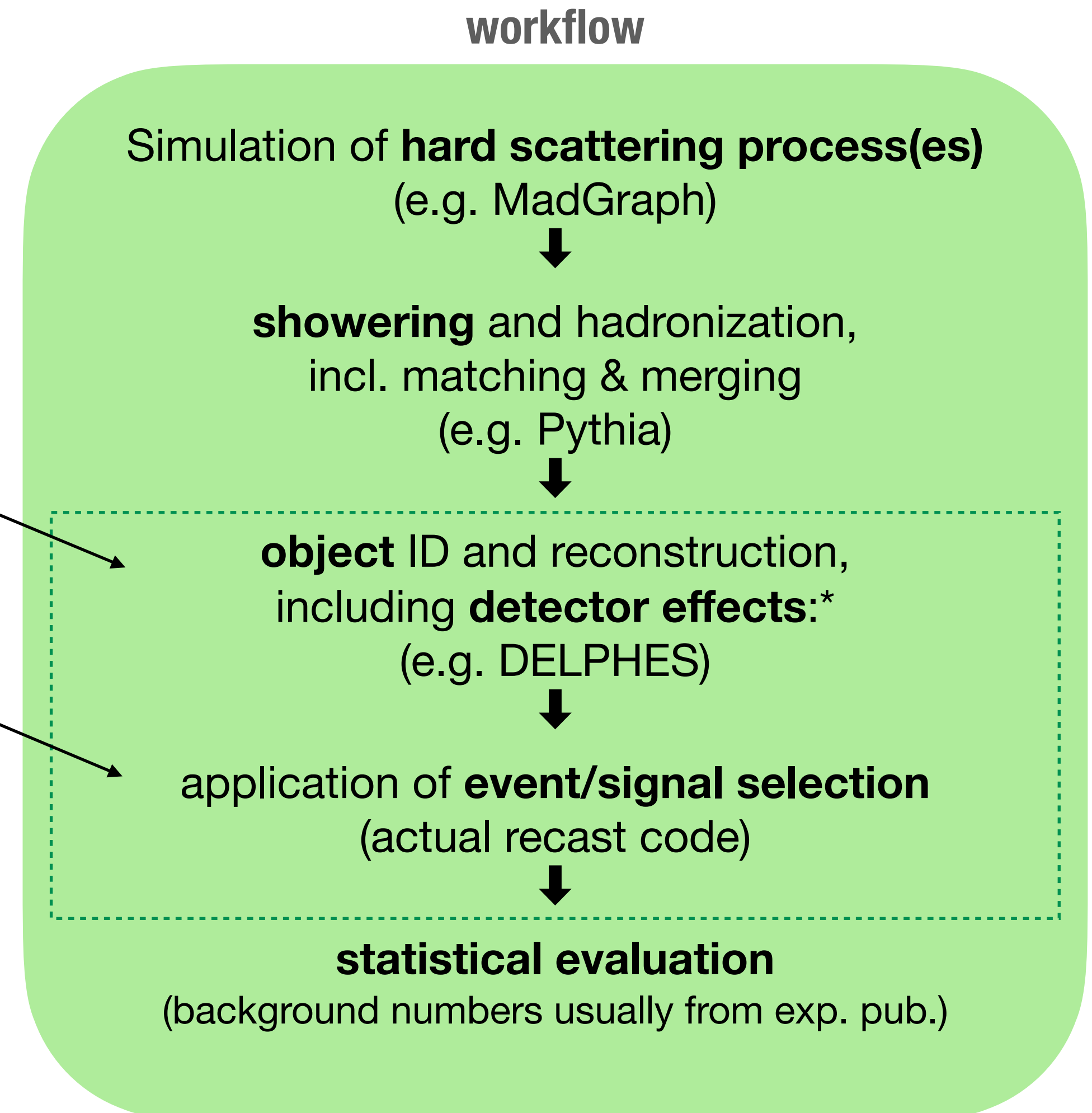
object definitions;
identification, tagging,
reconstruction efficiencies

detailed preselection and
signal (+control) region cuts

- ▶ However, more and more analyses exploit ML techniques to gain in sensitivity

e.g. ML-based taggers,
signal/bkg discrimination with ML classifiers

Pb: how can we reuse those?



* except for detector-unfolded results (Rivet/Contur)

ML as a bottleneck for reinterpretation?

- ▶ More and more analyses exploit ML techniques to gain in sensitivity.
- ▶ Serious difficulty for analysis preservation and reuse unless
 - resulting id/reco **efficiencies can be** (and are!) **parametrised** in terms of quantities accessible in a simulation, e.g., p_t , η , ...
 - the actual **ML model is published** in appropriate form.

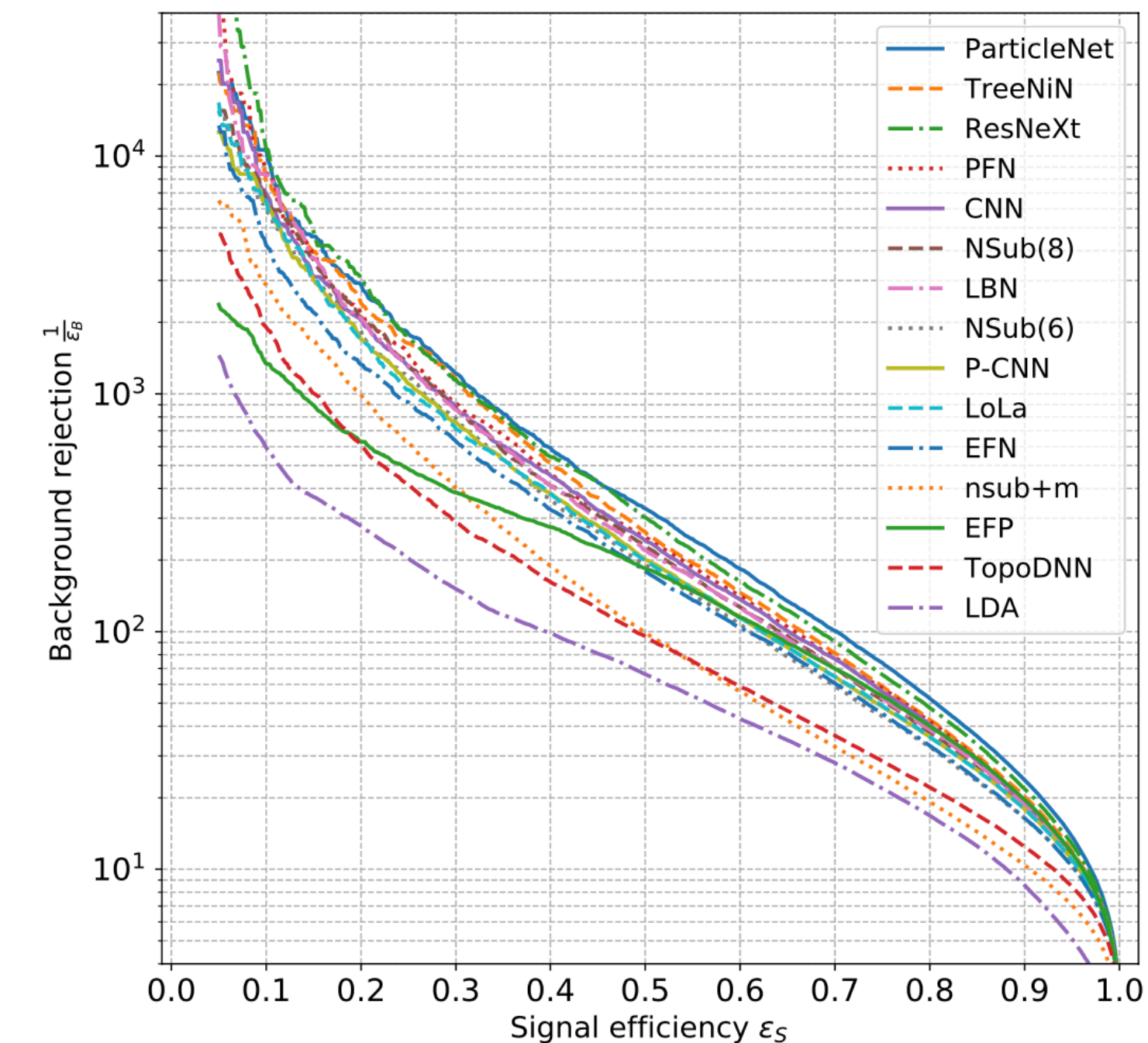
Caveat: input variables need to be physics quantities that can be matched in a simulation

Two analyses where the latter has been attempted:

- ▶ **ATLAS-SUSY-2018-22** (0-lepton gluino/squark search)
published **BDT weights** as XML file
→ RAMP seminar by Kenta Uno
- ▶ **ATLAS-SUSY-2019-04** (1-2 leptons + jets RPV search)
published neural network as **ONNX** file
→ RAMP seminar by Javier Montejo Berlingen

RAMP: Reinterpretation Auxiliary Material Presentation

Example: top taggers



13 different algorithms: image-based (2),
4-vector-based (5), theory-inspired (6) taggers

“The Machine Learning Landscape of Top Taggers”
G. Kasieczka, T. Plehn et al., arXiv:1902.09914

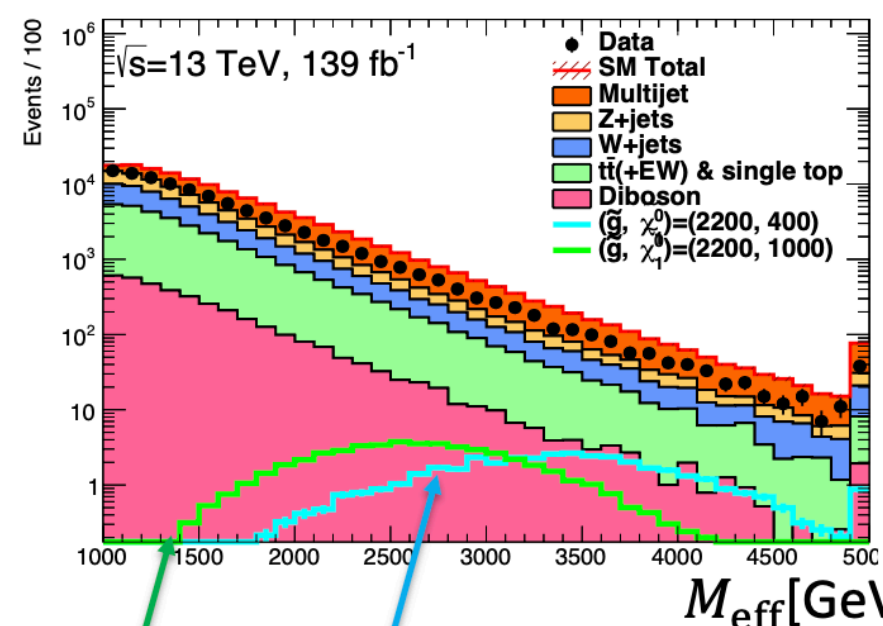
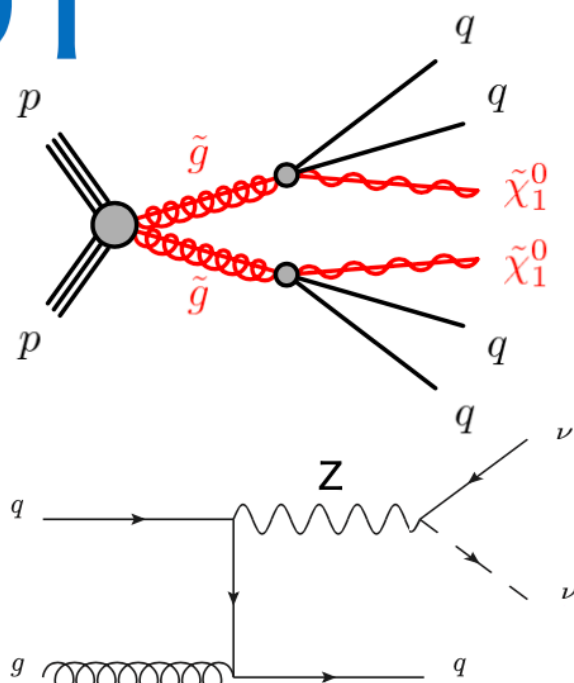
Kenta Uno on ATLAS-SUSY-2018-22

Analysis approach: BDT

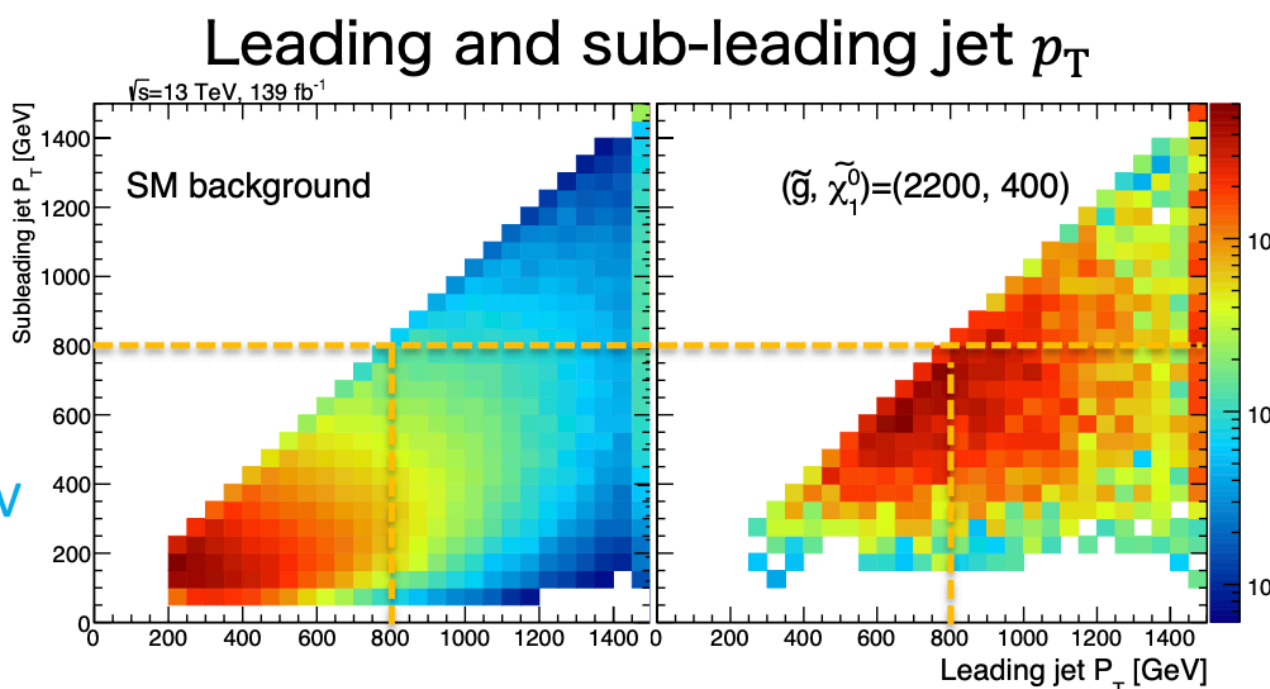
Key variable in the previous analysis: M_{eff}

$$M_{\text{eff}} = \sum_i |p_T(i)| + E_T^{\text{miss}}$$

- However, it is not efficient as $\tilde{\chi}_1^0$ mass is large



$m(\tilde{g}) = 2200$ GeV, $m(\tilde{\chi}_1^0) = 400$ GeV
 $m(\tilde{g}) = 2200$ GeV, $m(\tilde{\chi}_1^0) = 1000$ GeV



Using correlation between variables is still efficient.

→ BDT analysis is useful for SUSY search

2021/4/9

5

Discussion point 1

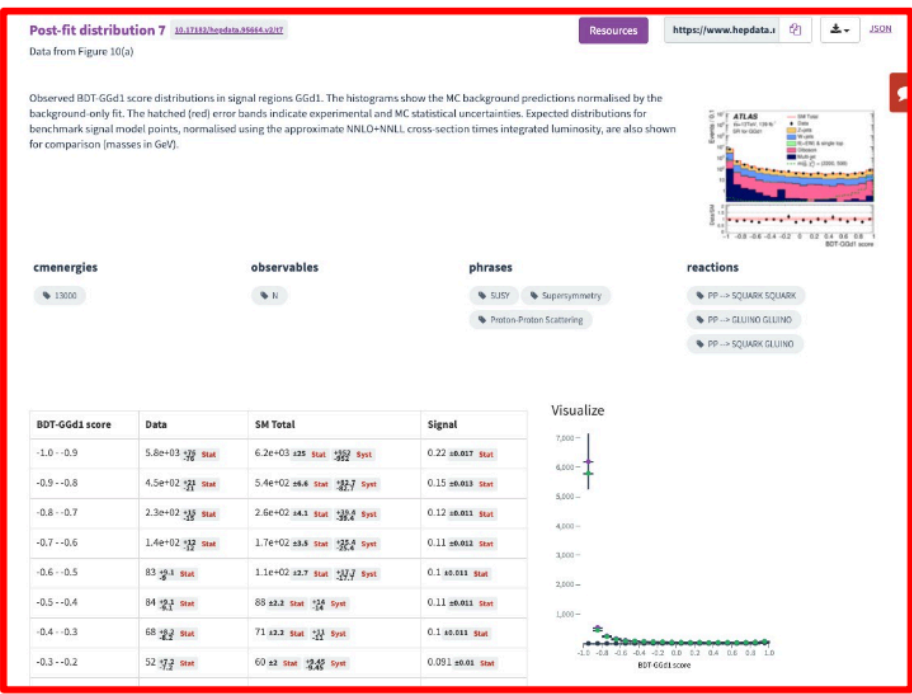
- The auxiliary material provided for the BDT reinterpretation
 - We published HepData materials [[Link](#)]
 - Provided xml files of BDT classifiers
 - ZeroLepton2018-SRBDT-weight.tar.gz
 - Also provided analysis snipped code
 - ZeroLeptonBDT2018.cxx
 - By using this code, you can use same BDT classifiers.
 - Full likelihood is provided in the hepdata material

- ✂ Signal events: fast sim (reco)
- ✂ Background events: full reco

We provide data, SM total and signal yields in each BDT score bin and the acceptance and efficiencies of BDT SRs for simplified GG models.

→ I think it might help everyone..

2021/4/9



14

10 -12 input variables: jet $p_T, \eta, E_T^{\text{miss}}, m_{\text{eff}}, \text{Aplanarity}$

Javier Montejo Berlingen on ATLAS-SUSY-2019-04

Reinterpretation material

Additional resources

C++ code snippet with the implementation of the analysis selection at truth-level

- Can be used with SimpleAnalysis framework

SLHA files for benchmark signals

ONNX files for the neural networks for the EWK analysis

- Not possible to use lwttn because the architecture of one of the layers is not supported

Upcoming: **ROOT workspaces**, containing data and the fitted background model in all SRs

- not using pyhf as it doesn't support our parameterised background model

HEPData

Search for R-parity violating supersymmetry in a final state containing leptons and many jets

Aad, Georges et al.

Search for R-parity violating supersymmetry in a final state containing leptons and many jets

the ATLAS experiment using $\sqrt{s} = 13$ TeV proton-proton collision data

The ATLAS collaboration

Aad, Georges , Abbott, Braden Keim , Abbott, D , Abud, Adam , Abeling, Kira , Abhayasinghe, Des , Abidi, Haider , Abramowicz, Halina , Abreu, He , Abulaiti, Yiming

CERN-EP-2021-066, 2021.

[https://doi.org/10.1007/JHEP04\(2021\)104](https://doi.org/10.1007/JHEP04(2021)104)

INSPIRE Resources

Abstract (data abstract)
CERN-LHC.

A search for R-parity violating supersymmetry in a final state characterised by high jet multiplicity, at least one light lepton and either zero or at least three b -tagged jets. The search uses 139 fb^{-1} of $\sqrt{s} = 13$ TeV proton collision data collected by the ATLAS experiment during Run 2 of the Large Hadron Collider. The results are interpreted in the context of R-parity-violating supersymmetry.

Additional Publication Resources

filter

Common Resources 4

Distribution: 1 2

Distribution: 2 2

Distribution: 3 2

Distribution: 4 2

Distribution: 5 2

Distribution: 6 2

Distribution: 7 2

Distribution: 8 2

Distribution: 9 2

Distribution: 10 2

Distribution: 11 2

Distribution: 12 2

Distribution: 13 2

Distribution: 14 2

External Link

web page with auxiliary material

View Resource

C++ File

Code snippet with the implementation of the analysis selection at truth-level

Download

tgz File

SLHA files for benchmark signals

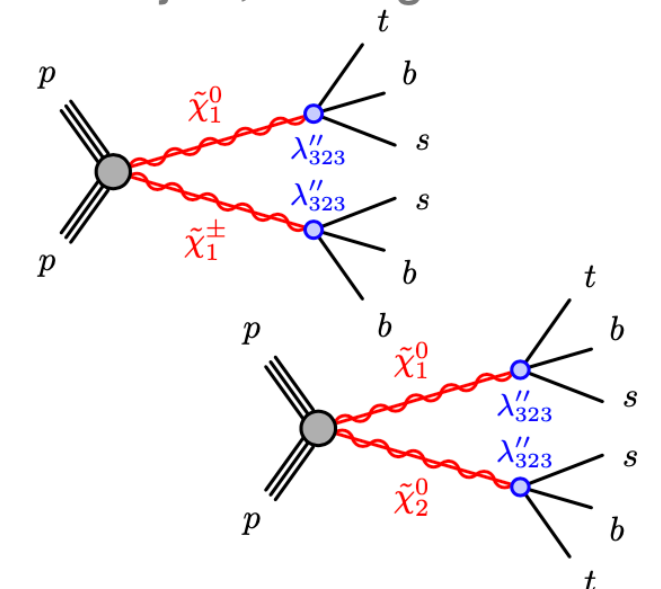
Download

tgz File

ONNX files for the neural networks for the EWK analysis

Download

Electroweak production
6 or 8 jets, 4 b-tag



Snowmass white paper on data and analysis preservation and reinterpretation

S. Bailey et al., arXiv:2203.10057

Reinterpretation and Recasting Recommendations

- 5.1:** Encourage that **reinterpretability and reuse be kept in mind early on in the analysis design.** This concerns, for instance, the **choice of input parameters in ML models**, the full specification of the fiducial phase space of a measurement in terms of the final state, including any vetos applied, and generally the choice of non-overlapping regions and standard naming of shared nuisances to facilitate the combination of analyses.
- 5.2:** Develop a common set of standards for the naming and documentation of shared nuisances.
- 5.3:** Improve the coordination among the different public reinterpretation frameworks with the goal of a centralised database of recast codes, common input/output formats, and a unified statistical treatment.
- 5.4:** Encourage the **FAIR**-ification of codes and **data products** from (theory) reinterpretation studies outside the experimental collaborations at the same level of sophistication as asked for experimental analyses and results. Suitable repositories are, e.g., GitHub and Zenodo; appropriate versioning is essential.