https://sites.google.com/view/mark-d-goodsell

Latest developemnts in HackAnalysis and MadAnalysis

Mark Goodsell

@PitifulRed

https://realselfenergy.blogspot.com

https://goodsell.pages.in2p3.fr/hackanalysis/

https://goodsell.pages.in2p3.fr/bsmart/

LPTHE
LABORATOIRE DE PHYSIQUE
THEORIQUE ET HAUTES ENERGIES

SORBONNE UNIVERSITÉ

PROJET FINANCÉ PAR L'ANR - PROJECT FUNDED BY THE ANR

cnrs

Overview

- Update on the activities from the LPTHE affiliates (Benj Fuks & myself)
- Justification for existence of HackAnalysis
- Latest developments
- Latest inclusions in MadAnalysis and future ideas

Why HackAnalysis?

Exist several major frameworks for ‘full’ recasting:

- RIVET: amazing for SM processes, great support by ATLAS members, no root, YODA for histograms. For BSM all detector stuff is abstracted into “projections.” Relies on being handed hepmc files (nb weight treatment).
- MadAnalysis: gold standard for transparency and reproducability. Can use either Delphes or SFS (no root required, but hepmc instead).
- CheckMATE: good ideas for running points quickly, loads of analyses being added, mainly intended to be used as a black box, requires root and Delphes.
- GAMBIT: intended for global scans as black box, great ideas for fast detector simulation, some compromises in favour of speed vs accuracy.
- ADL: no need to introduce here, relies on root & hepmc.

In 2020, I wanted to recast the CMS disappearing track search, and none of them were usable: if you want some feature (finite size of detector, disappearing tracks in this case) you better contact the authors.

CMS-EXO-20-004: cuts

- Triggers followed by a cut on MET (120 GeV) – *without muons*
- At least one high-pT jet (110 GeV), no jets within $|\Delta\Phi| < 0.5$ of the MET vector
- Remaining cuts are all on the tracks: pT > 55 GeV
- Sufficiently isolated
- No missing hits in the pixel detector, no missing inner hits
- Sufficiently separated ($\Delta R < 0.5$) from jets, ($\Delta R < 0.15$) from electrons, *muons*
- Must actually disappear! That means, >2 missing outer hits, < 10 GeV calorimeter energy around the track.
- Extra complication: data split into 6 different data taking periods! 2015, 2016A/B, 2017, 2018A/B (due to malfunctioning parts of detector)
- Signal regions depend on number of tracker layers that have been hit!

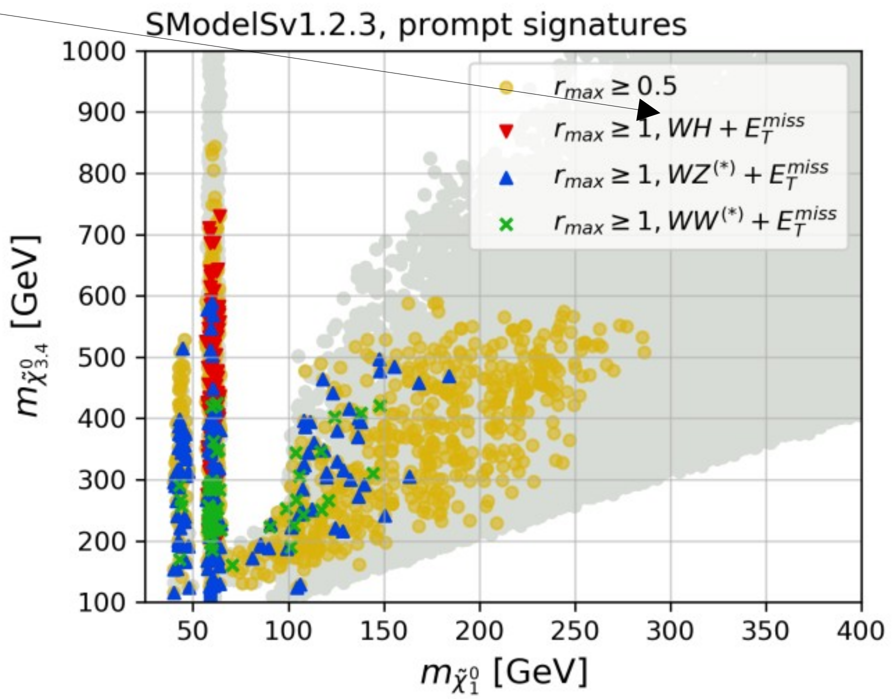
Pileup!

Long-lived charginos get mistaken for muons and get included in MET calculation!

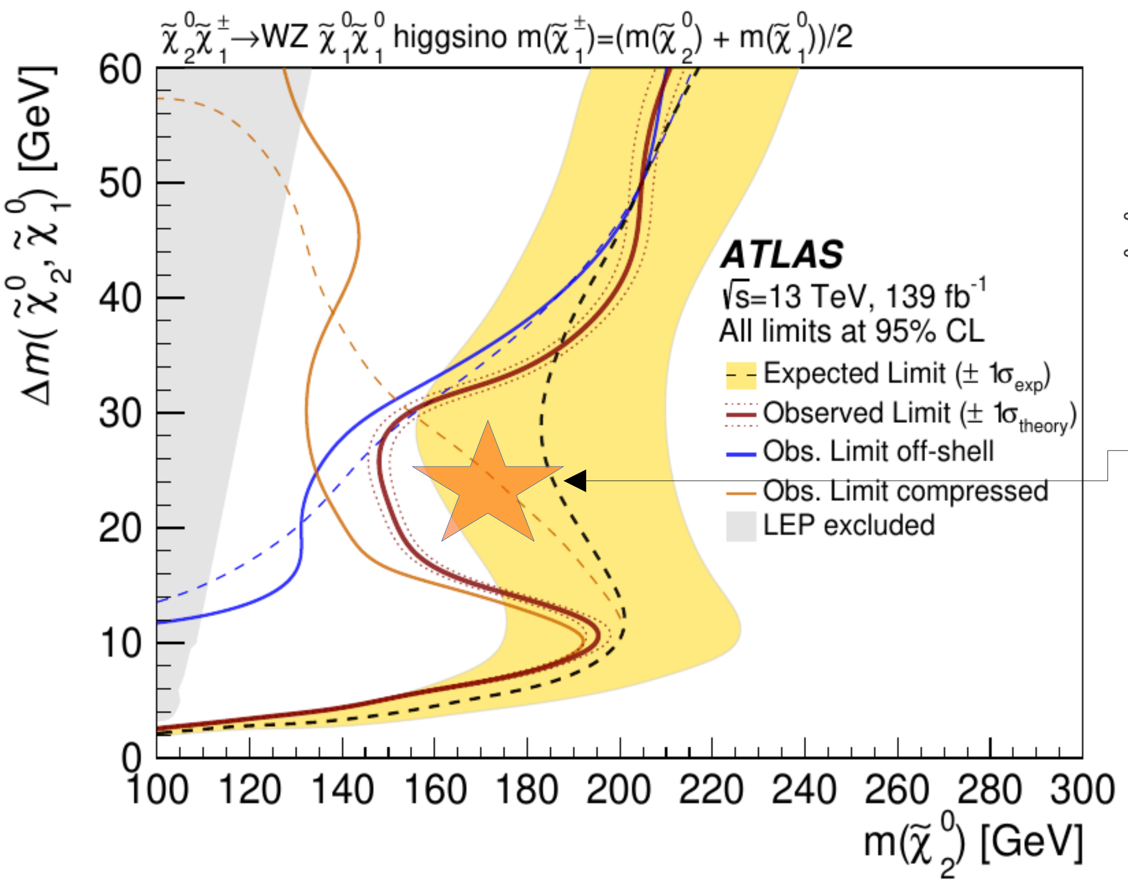
Pixels, hits, track isolation, forget about using any standard detector simulation!

- I was also interested in electroweakino searches for Dirac Gaugino models (SciPostPhys.9.4.047 with Kraml, Reyes and Williamson)
- I tried to use GAMBIT, was proposed “ColliderBit standalone solo” but could not use it. (Issues with the pythia code produced by MadGraph, etc etc).
- Some of the most powerful relevant analyses weren’t extant anyway.

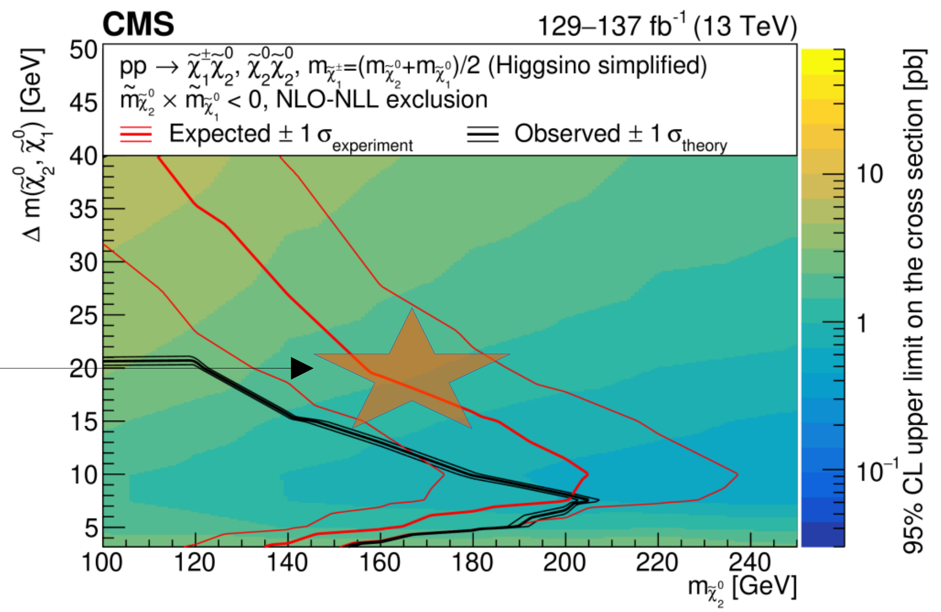
- For the recast of ATLAS-SUSY-2018-09 in MadAnalysis:
  - EWino searches have fairly small efficiencies → need to simulate large numbers of events.
  - At the time needed to use Delphes → generate large root files, tinkering with the efficiencies during development of the recast was **painful.**
  - Implementing dynamic isolation requirements (as ATLAS required) impossible in Delphes (in-built routines were removing too many leptons), so have to do it in the analysis anyway.
  - At the time there was no pyhf interface in MadAnalysis



Excesses in soft lepton searches



Combination from ATLAS-SUSY-2019-09

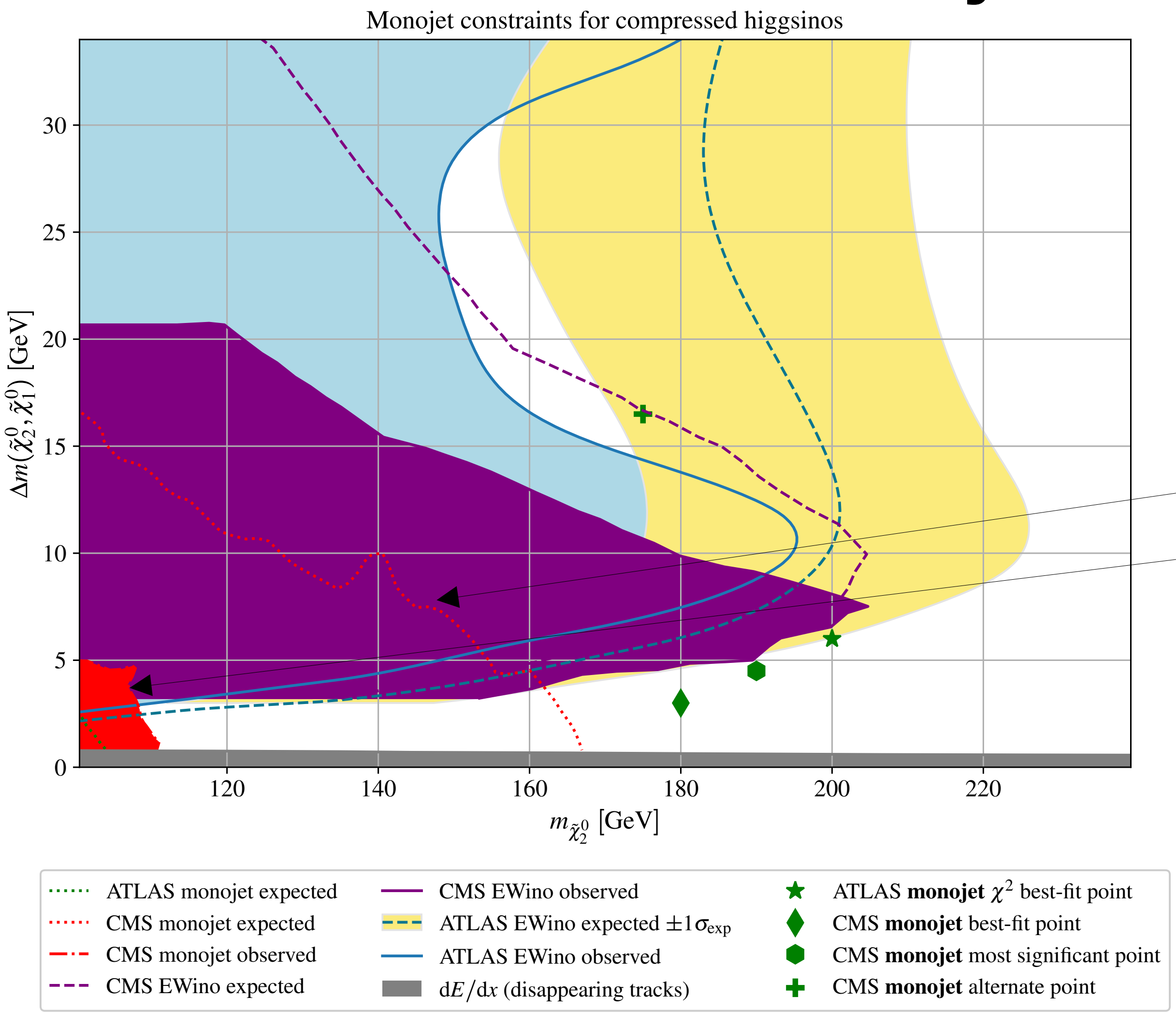


Equivalent limits on Higgsinos from CMS-SUS-18-004

We pointed out in arXiv:2311.17149 (with D. Agin, T. Murphy, B. Fuks) that there are also excesses in monojets that overlap with these

- Performed scan in the plane of the Higgsino scenario using originally 1M events per point, then **8M per point** in fine region, grid with increments of 10 GeV in Δm and 0.5 GeV in $m_{\tilde{\chi}_2^0}$
- Computed NLO+NLL cross-sections with `resummino` (and `BSMArt`)
- Recast using `MadAnalysis`
- Statistical analysis using `spey`
- Need to use 'best' signal region only for ATLAS (no combinations)

Excesses in Monojet searches

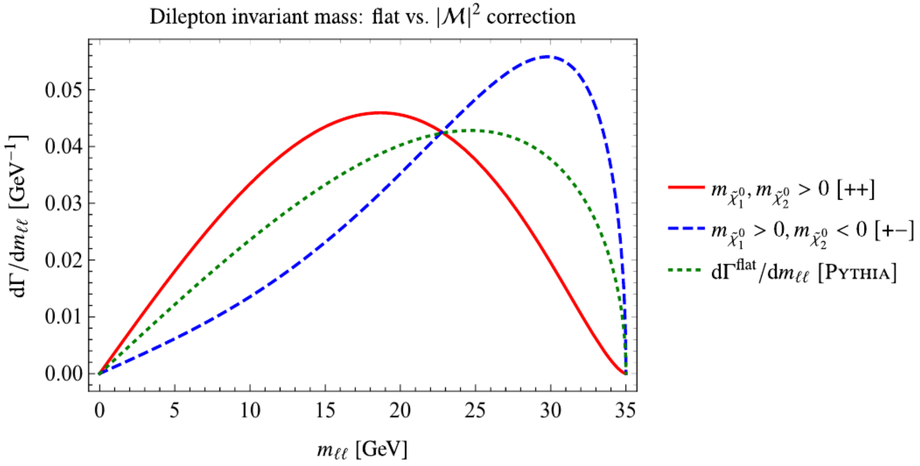


- Close the Higgsino hole!
- ATLAS search much less powerful
- But: found a massive difference between expected and observed limits …

…. in ***both ATLAS and CMS analyses***

Recasting – and using – the ATLAS soft lepton searches is challenging:

- 2-lepton search uses RestFrames – contains > 50 c++ files, needs root + minuit
- Need detailed invariant-mass reconstruction of decay products → not possible with generation of events in pythia
- Tiny efficiencies → simulate large numbers of events
- 3-lepton search uses MET significance



Cut	ATLAS	HackAnalysis
Initial number of events ($\mathcal{L} \times \sigma$)	1.0	1.0
Initial number of events ($\mathcal{L} \times \sigma_{\geq 1\text{jet}}$)	2.3×10^{-1}	5.0×10^{-1}
E_T^{miss} trigger	2.8×10^{-2}	1.2×10^{-1}
2 leptons	4.2×10^{-3}	6.1×10^{-3}
veto 3GeV $< m_{\ell\ell} <$ 3.2GeV	3.9×10^{-3}	5.7×10^{-3}
$\min(\Delta\phi(\text{any jet}, \mathbf{p}_T^{\text{miss}})) > 0.4$	3.8×10^{-3}	5.3×10^{-3}
$\Delta\phi(j_1, \mathbf{p}_T^{\text{miss}}) > 2.0$	3.7×10^{-3}	5.2×10^{-3}
$1 < m_{\ell\ell} < 60$ GeV	3.3×10^{-3}	4.0×10^{-3}
$\Delta R_{ee} > 0.30$, $\Delta R_{\mu\mu} > 0.05$, $\Delta R_{e\mu} > 0.20$	2.9×10^{-3}	4.0×10^{-3}
Leading lepton $p_T > 5$ GeV	2.4×10^{-3}	3.3×10^{-3}
Number of jets ≥ 1	2.3×10^{-3}	3.3×10^{-3}
Leading jet $p_T >$100 GeV	2.1×10^{-3}	2.5×10^{-3}
Number of b-tagged jets $= 0$	1.8×10^{-3}	2.2×10^{-3}
$m_{\tau\tau} < 0$ or > 160 GeV	1.5×10^{-3}	1.9×10^{-3}
ee or $\mu\mu$	1.5×10^{-3}	1.9×10^{-3}
$m_T^{\ell_1} < 60$ GeV	1.3×10^{-3}	1.6×10^{-3}
$E_T^{\text{miss}} > 200$	6.5×10^{-4}	8.9×10^{-4}
$\max(0.85, 0.98 - 0.02 \times m_{\ell\ell}) < R_{\text{ISR}} < 1.0$	4.9×10^{-4}	5.6×10^{-4}
sub-leading lepton $p_T > \min(10, 2 + m_{\ell\ell}/3)$	4.7×10^{-4}	5.5×10^{-4}
$m_{\ell\ell} < 60$ GeV	4.7×10^{-4}	5.5×10^{-4}
$m_{\ell\ell} < 40$ GeV	4.7×10^{-4}	5.5×10^{-4}
$m_{\ell\ell} < 30$ GeV	4.7×10^{-4}	5.5×10^{-4}
$m_{\ell\ell} < 20$ GeV	4.7×10^{-4}	5.5×10^{-4}
$m_{\ell\ell} < 10$ GeV	4.7×10^{-4}	5.5×10^{-4}
$m_{\ell\ell} < 5$ GeV	4.7×10^{-4}	5.5×10^{-4}
$m_{\ell\ell} < 3$ GeV	3.2×10^{-4}	3.6×10^{-4}
$m_{\ell\ell} < 2$ GeV	1.3×10^{-4}	1.5×10^{-4}

Table 5: Signal region comparison for ATLAS-SUSY-2018-16 signal region SR_E_hi

Goals for HackAnalysis:

- Major recasting packages have become monolithic: they do what they are intended to do incredibly well, but it’s hard to get them to do something else. I want to be able to add new features easily and without breaking something. E.g.:
  - RestFrames
  - Pileup in fast sim
  - FastJet features such as pileup subtraction, etc etc
  - Finite detector size
- Would be ideal to take advantage of the best ideas of each.
- Want to speed up development of new analyses (mainly) for MadAnalysis – this means no compromises in precision.
- Ideally should be as simple as possible to port to other frameworks.
- Therefore also want a minimum of external dependencies (e.g. root can be difficult to install/unavailable on clusters).

Intro to HackAnalysis

- Implementation of MadAnalysis-style analysis structure (init(), Execute(), Finalise(); AddRegionSelection(..), AddCut(..)) so you can almost convert to MA5 syntax with a perl script … but based on `heputils` – can take advantage of GAMBIT binning functions/efficiency functions/syntax.
- YODA for plotting/histogramming (and can also read efficiency information in YODA files provided on HEPData).
- Basic Makefile rather than configure scripts, cmake etc – easier to add your own code.
- External dependencies: YODA, hepmc2, fastjet, pythia, openmp. ONNX and zlib as options.
- Four running modes:
  - analysePYTHIA.exe for pythia event generation (super fast + dirty)
  - analysePYTHIA_LHE.exe for reading lhe files + showering internally
  - analyseHEPMC.exe mainly for compatibility/checking against MA5
  - analyseHAEVENT.exe for reading pre-processed events
- Piloted by a `yaml` file
- Can include pileup (code for generating min bias events included)

```
---
analyses:
 - DT_CMS
 - HSCP_ATLAS

settings:
  nevents: 1000
  cores: 1
  Include Pileup: false
  Efficiency Filename: LO.eff
  Cutflow Filename: LO_cf.eff
  Histogram Filename: LO.yoda
  Config file: LO.cfg
```

This should be called via

```
./analysePYTHIA.exe LOpythia.yaml
```

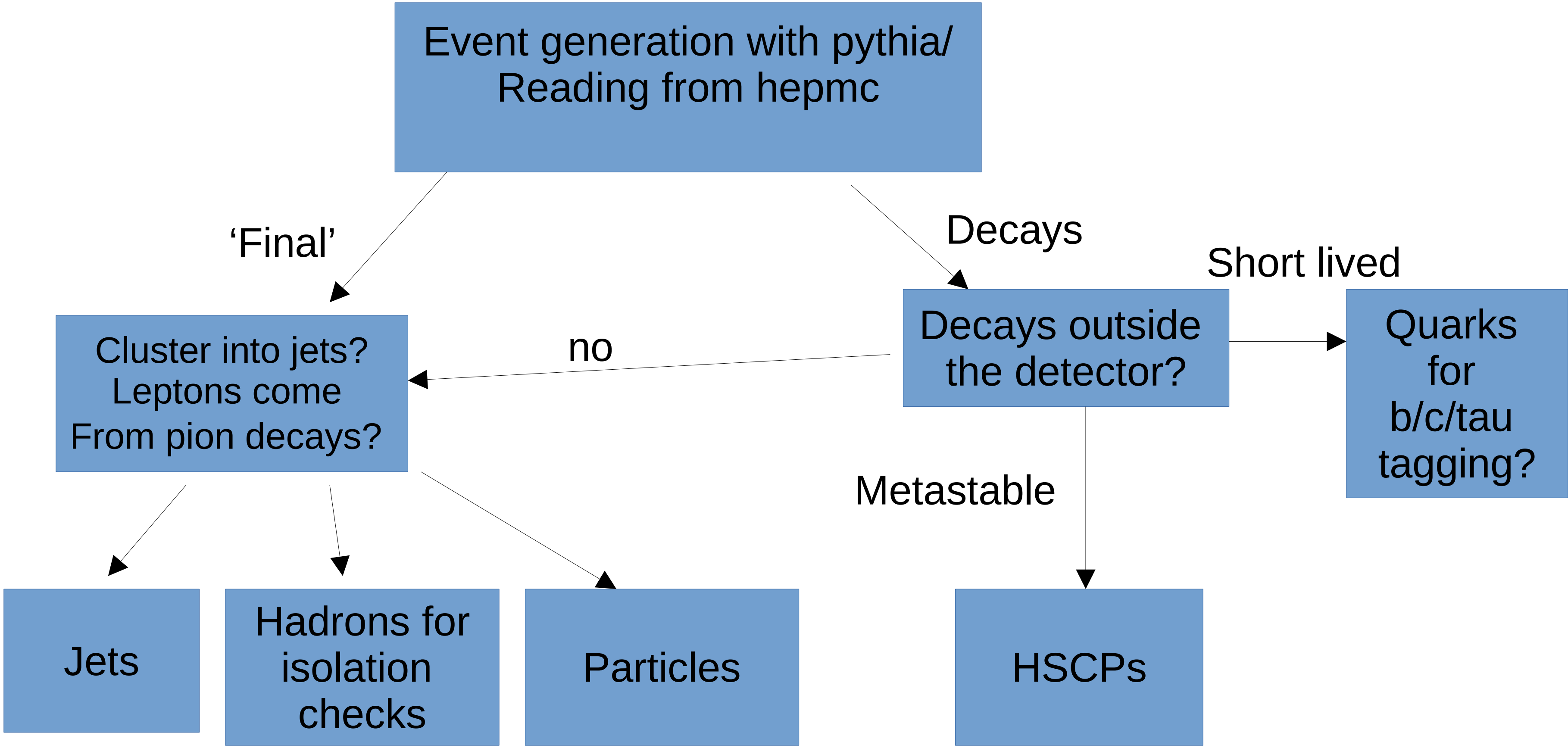
HackAnalysis 2 new features

Described in the manual arXiv:2406.10042

- Simple inclusion of new particles via a QNUMBERS file (or directly in yaml input)
- Multiple ‘detector’ simulations
- Compressed event format
- Automatic systematic uncertainties
- RestFrames, Eigen, Nelder-Mead minimiser, MT2
- ONNX interface
- Json output for cutflows, weight info, etc: can be used for merging runs
- Scripts for merging runs, printing cutflows in LaTeX
- Python scripts for running stats (exclusion/signal strength limits/p-values/likelihoods) through pyhf, spey and toy-based single bin
- Interface with BSMArt for scanning – handling the generation of events in MadGraph, gridpack generation, etc – and convergence checking
- New and old analyses

Write your own filling function! Maximum flexibility to use e.g. avanced fastjet features without breaking something, etc etc

Detector ‘filling’ routines



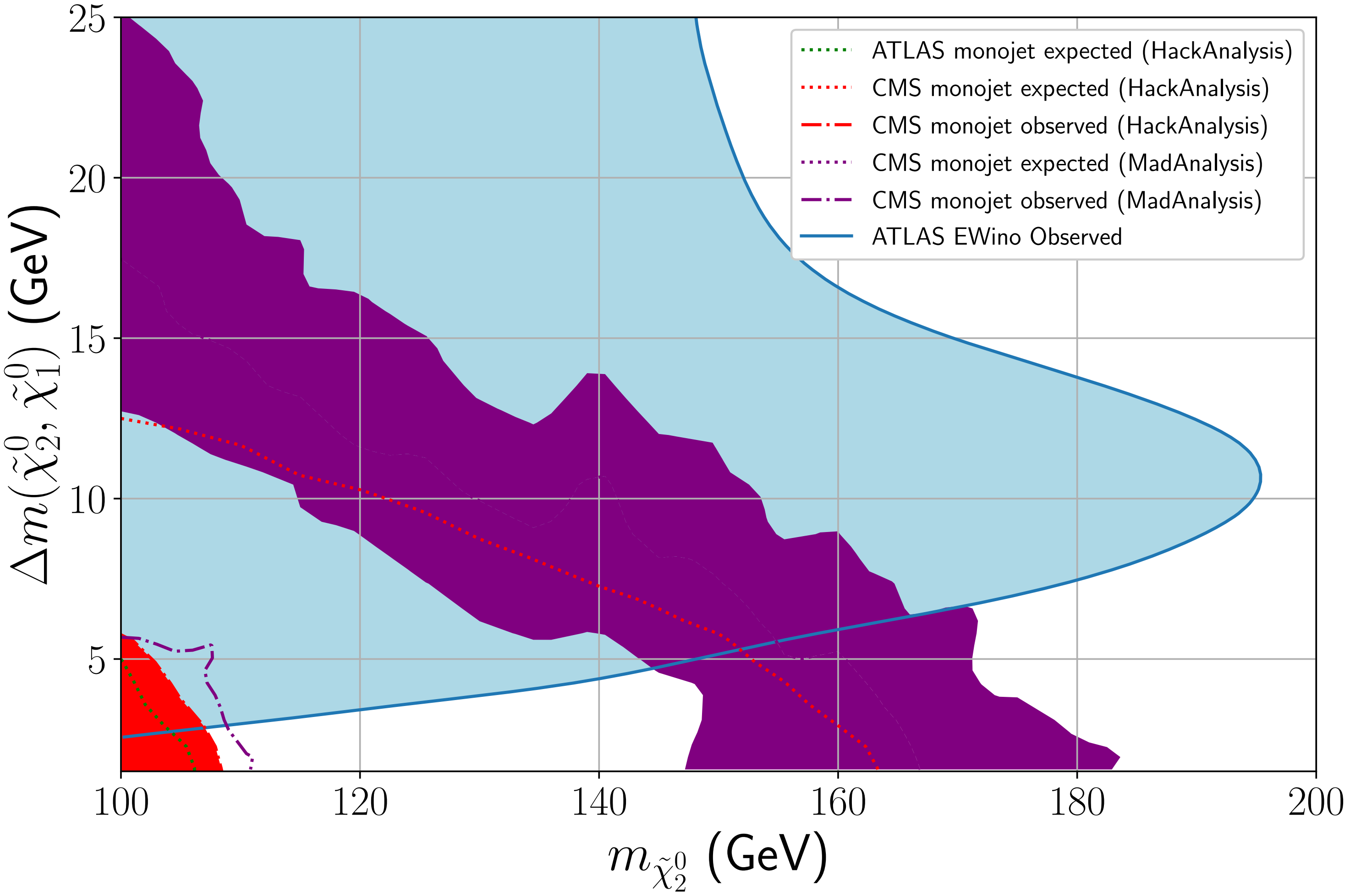
Gridpacks, batches

Many features to make running/prototyping as fast as possible:

- Generate gridpack in MadGraph → run in ‘read-only’ mode, one gridpack run per core to generate lhe events. One .lhe file per core.
  - Can then shower directly running pythia.
  - With gridpacks can easily run batches of points of any size.
  - If not too large: put MG5 output + gridpacks on ramdisk (/dev/shm) → no writing to disk at all during run!
  - Extra bonus: can then do convergence checks after each batch
- Store events in a compressed reco format. E.g. 100k event sample:
  - 7.2 GB `.hepmc` (!!!)
  - 19 MB `.lhe.gz`
  - 10 MB `.ha.gz`
- Store one reco file/core → can rerun sample in multicore mode. Incredibly fast.
- Can choose to keep hadrons for isolation or discard.

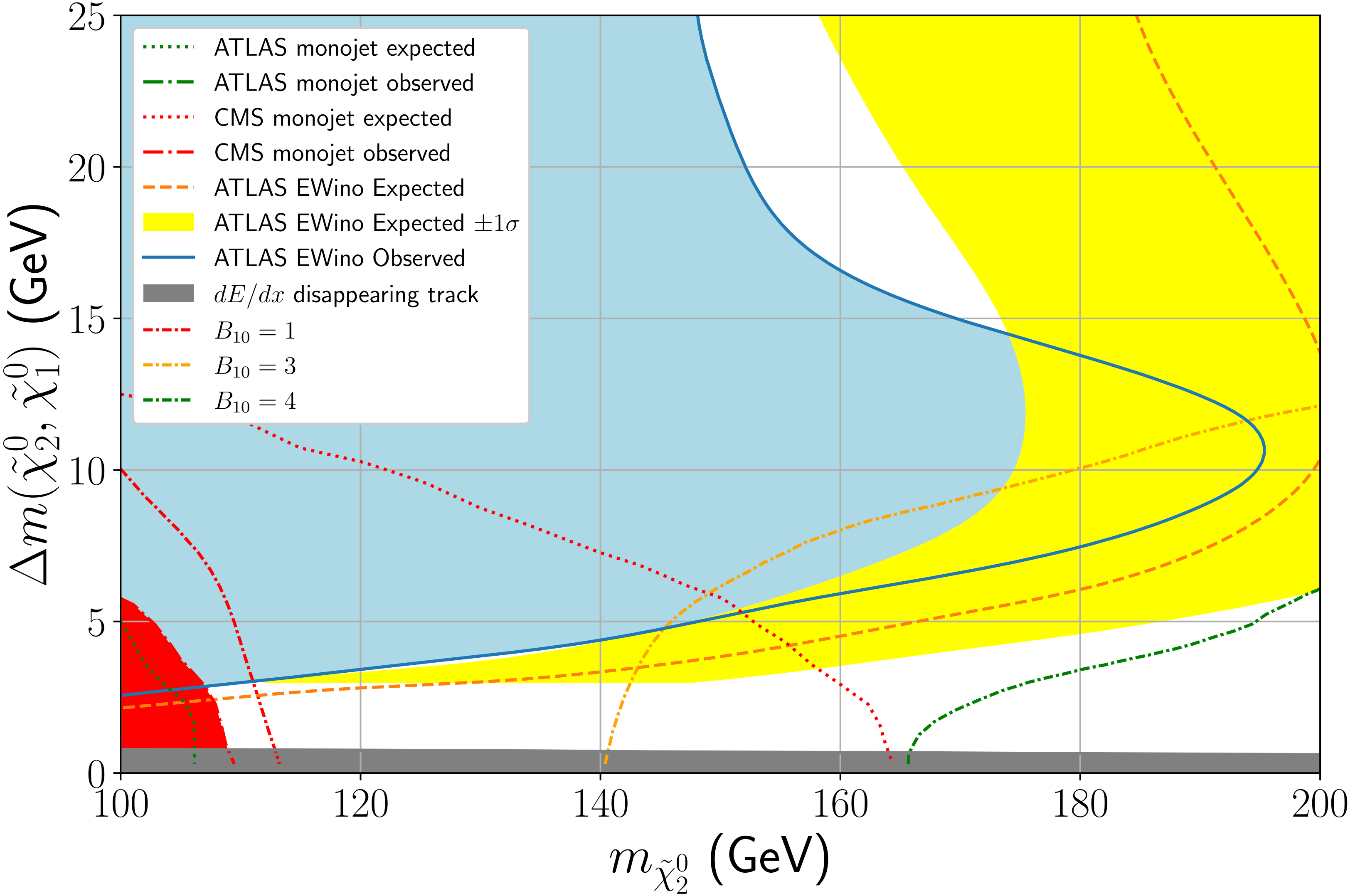
E.g. running 19.2M events/point using gridpack mode via BSMArt takes about 4 hours/point/batch of 3.2M events on 8 cores on lxplus …

vs 16 hours to run 2M events via MadAnalysis



And this is without using ramdisk/batches

Can also compute likelihood ratios (~ Bayes factor) via spey:



Automatic systematic uncertainties, TeX outputs:

First cut: overall systematics

Subsequent cuts: uncertainty on cut efficiency

Cut	ATLAS	HackAnalysis
All weighted events	1.0	$1.0^{+0.00}_{-0.00}$ (stat) $^{+16.1\%}_{-10.4\%}$ (syst)
$N_{\mathrm{jets},25} \geq 2$	8.8×10^{-1}	$9.0^{+0.01}_{-0.01} \times 10^{-1}$ (stat) $^{+0.4\%}_{-0.5\%}$ (syst)
1 signal lepton	7.9×10^{-1}	$7.9^{+0.02}_{-0.02} \times 10^{-1}$ (stat) $^{+0.4\%}_{-0.6\%}$ (syst)
Second baseline lepton veto	7.6×10^{-1}	$7.9^{+0.02}_{-0.02} \times 10^{-1}$ (stat) $^{+0.4\%}_{-0.6\%}$ (syst)
$m_{\mathrm{T}} > 50$ GeV	7.0×10^{-1}	$7.3^{+0.02}_{-0.02} \times 10^{-1}$ (stat) $^{+0.4\%}_{-0.6\%}$ (syst)
$E_{\mathrm{T}}^{\mathrm{miss}} > 180$ GeV	6.0×10^{-1}	$6.2^{+0.02}_{-0.02} \times 10^{-1}$ (stat) $^{+0.5\%}_{-0.7\%}$ (syst)
$N_{\mathrm{jets}} \leq 3$	5.0×10^{-1}	$4.6^{+0.02}_{-0.02} \times 10^{-1}$ (stat) $^{+1.8\%}_{-1.8\%}$ (syst)
$N_{\mathrm{b-jets}} = 2$	2.2×10^{-1}	$2.2^{+0.02}_{-0.02} \times 10^{-1}$ (stat) $^{+2.9\%}_{-2.6\%}$ (syst)
$m_{\mathrm{bb}} > 50$ GeV	2.2×10^{-1}	$2.2^{+0.02}_{-0.02} \times 10^{-1}$ (stat) $^{+2.9\%}_{-2.6\%}$ (syst)
$E_{\mathrm{T}}^{\mathrm{miss}} > 240$ GeV	1.9×10^{-1}	$1.9^{+0.02}_{-0.02} \times 10^{-1}$ (stat) $^{+2.9\%}_{-2.6\%}$ (syst)
$m_{\mathrm{bb}} \in$ [100,140] GeV	1.4×10^{-1}	$1.6^{+0.01}_{-0.01} \times 10^{-1}$ (stat) $^{+3.0\%}_{-2.7\%}$ (syst)
$m_{\ell,\mathrm{b}_1} > 120$ GeV	1.3×10^{-1}	$1.5^{+0.01}_{-0.01} \times 10^{-1}$ (stat) $^{+3.0\%}_{-2.7\%}$ (syst)
$m_{\mathrm{T}} > 240$ GeV	9.6×10^{-2}	$1.1^{+0.01}_{-0.01} \times 10^{-1}$ (stat) $^{+3.2\%}_{-2.8\%}$ (syst)
$m_{\mathrm{CT}} > 180$ GeV	8.3×10^{-2}	$9.4^{+0.12}_{-0.12} \times 10^{-2}$ (stat) $^{+3.4\%}_{-2.8\%}$ (syst)
$m_{\mathrm{CT}} \in$ [180,230] GeV	1.6×10^{-2}	$1.5^{+0.05}_{-0.05} \times 10^{-2}$ (stat) $^{+4.4\%}_{-3.3\%}$ (syst)
$m_{\mathrm{CT}} \in$ [230,280] GeV	1.8×10^{-2}	$1.7^{+0.05}_{-0.05} \times 10^{-2}$ (stat) $^{+4.5\%}_{-2.9\%}$ (syst)
$m_{\mathrm{CT}} > 280$ GeV	5.0×10^{-2}	$6.2^{+0.10}_{-0.10} \times 10^{-2}$ (stat) $^{+2.9\%}_{-2.8\%}$ (syst)

Table 3: ATLAS-SUSY-2019-08, Signal regions HM for parameter point $m_{\tilde{\chi}_2^0}/m_{\tilde{\chi}_1^\pm}, m_{\tilde{\chi}_1^0} = (750, 100)$ GeV. The final lines correspond to regions HMdisc,HMlow,HMmed and HMhigh respectively.

Analyses

List of implemented analyses

HackAnalysis Name	$\mathcal{L}(\text{fb}^{-1})$	Reference	Collab. Page	HEPData
ATLAS_SUSY_2018_16	139	arXiv:1911.12606	ATLAS-SUSY-2018-16	1767649
ATLAS_SUSY_2019_09_offshell	139	arXiv:2106.01676	ATLAS-SUSY-2019-09	1866951
ATLAS_SUSY_2019_09_onshell	139	arXiv:2106.01676	ATLAS-SUSY-2019-09	1866951
CMS_EXO_20_004	139	arXiv:2107.13021	CMS-EXO-20-004	1893308
ATLAS_EXOT_2018_06	139	arXiv:2102.10874	ATLAS-EXOT-2018-06	1847779
ATLAS_SUSY_2019_08	139	arXiv:1909.09226	ATLAS-SUSY-2019-08	1755298
ATLAS_SUSY_2017_04_2body	139	arXiv:1907.05163	ATLAS-SUSY-2017-04	
ATLAS_SUSY_2017_04_3body	139	arXiv:1907.05163	ATLAS-SUSY-2017-04	
DT_CMS	38.4	arXiv:2004.05153	CMS-EXO-19-010	1790827
DT_CMS	101	arXiv:1804.07321	CMS-EXO-16-044	1669245
HSCP_ATLAS	36.1	arXiv:1902.01636	ATLAS-SUSY-2016-32	1718558
ATLAS_SUSY_2017_04_2body	32.8	arXiv:2112.05163	ATLAS-SUSY-2017-04	1745920
ATLAS_SUSY_2017_04_3body	32.8	arXiv:2112.05163	ATLAS-SUSY-2017-04	1745920

Intention is not to create a competitor database

Idea: prototype + scan in HA2, then export

Alternative workflow is I develop the HA2 version at the same time as a MA5 one (e.g. by student …) for cross-checks

MadAnalysis Developments

- HEPData integration: MadAnalysis implementations now linked directly (extensive discussion yesterday)
- ONNX interface (nearly ready): Artur Oudot (ATLAS @ LPNHE) has been working on it with one analysis.
- Uncertainties in plots (ask Jack Araz).
- Interface to spey (nearly ready; prior Simplified Likelihood code should not be used).
- Restframes (in progress, debugging).
- Piloting via BSMArt (already possible for both PAD and Expert mode).

Aside on systematics

- RIVET and MadAnalysis read weight info from the hepmc files, but ***there is no convention for weight naming.***
- MadGraph has a method of naming the weights, but information is lost:
- In contrast, in .lhe files the weights are defined in blocks and the method for combining the weights is retained.
  - ➔ In HackAnalysis systematics are computed using the weight info from the .lhe files.

Future directions for MA5/HA2 and OpenMAPP

Done(?)

No	Task name	Lead	Partners	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
1.1	Extension of HEPData functionalities	7	4,5,7	x	x	x	x	x	x																		
1.2	Interfacing to HEPData	4	1,2,4,5,6,7	x	x	x	x	x	x	x	x	x	x	x	x												
1.3	Interfacing to LHC Open Event Data	8	2,3,8							x	x	x	x	x	x												
2.1	(Meta)Database of implemented analyses	2	1,2,4,5,6,8			x	x	x	x	x	x	x	x	x	x	x	x	x									
2.2	Common interface for analysis steering and output	4	2,4,5,6,7			x	x	x	x	x	x	x	x	x	x												
2.3	MC interface and common validation infrastructure for recasting frameworks	3	2,3,4,5,6,8													x	x	x	x	x	x	x	x				
3.1	Statistical models	1	1,2,4-8								x	x	x	x	x	x	x	x	x								
3.2	Enable combinations of analyses	2	1,2,4,5,6,8												x	x	x	x	x	x	x	x	x	x			
3.3	Physics case study	1	1-6,8																					x	x	x	x

Through spey

To discuss!

Future directions for MA5 and OpenMAPP

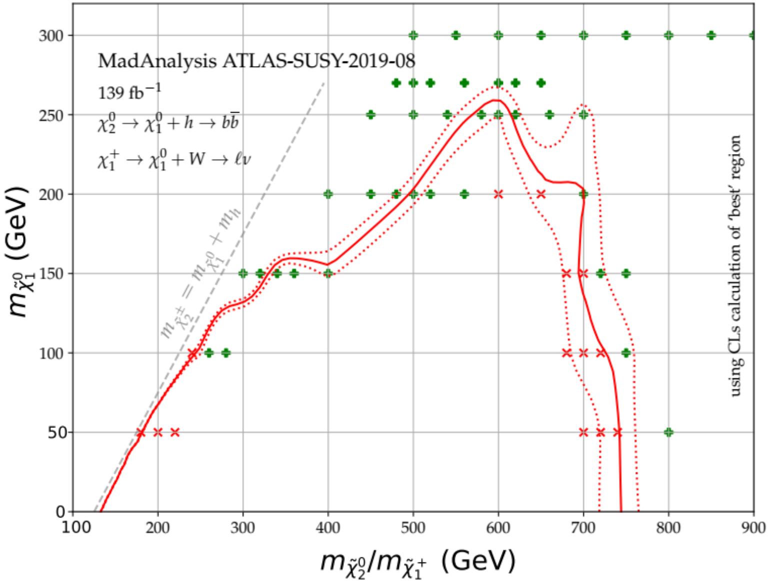
- Convergence checking + batch mode (anticipated, should yield substantial speedups)
- In the OpenMAPP document we promised HDF5 interface for event storage. This is to be discussed.
- Interoperability with other frameworks? Via HackAnalysis
- Improved connection of MA5/HA2 with SModelS via BSMArt???

BACKUP

Recast of ATLAS-SUSY-2019-08

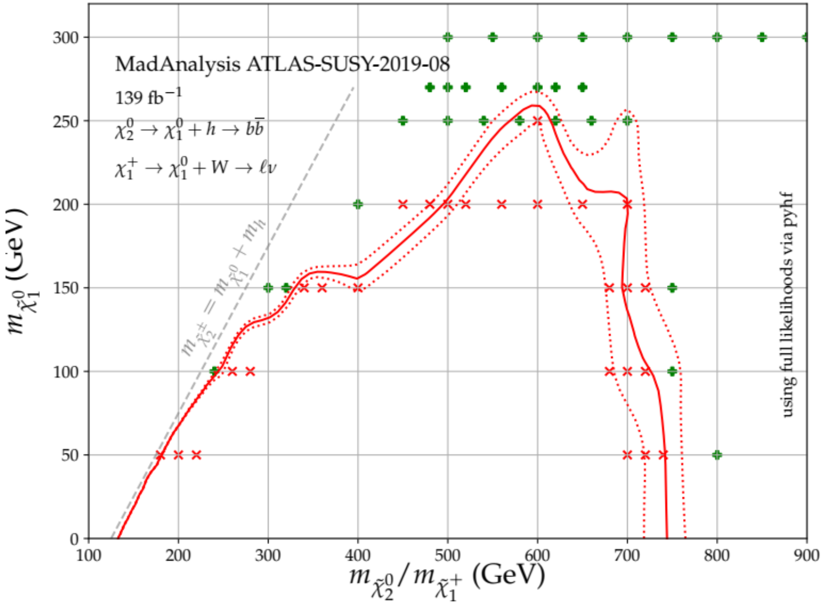
WH signature through Higgs to bb and W to leptons @ 139fb^{-1}

Looked for a Wino NLSP and Bino LSP in **MSSM**



Exclusions using 'best' region

Full likelihoods in pyhf



Exclusions using private implementation + pyhf

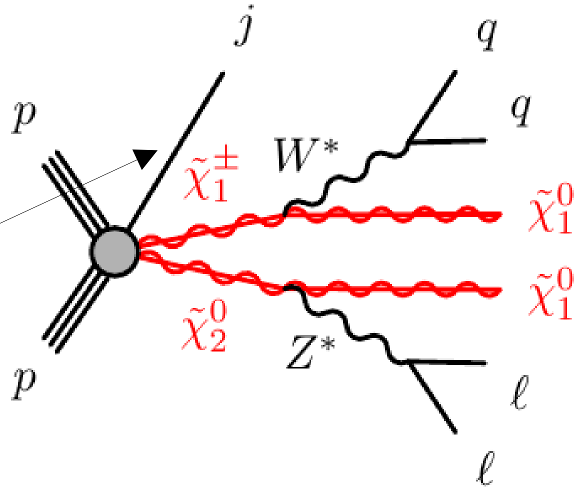
Examining these excesses would require recasts which were not available (see later) – recasts of electroweak searches are sporadic:

- GAMBIT conducted a study with massless gravitinos (arXiv:2303.09082)
- The only soft lepton search they had is the 36 inv. fb by CMS …
- … that is also MadAnalysis

Search label	Luminosity	Source
ATLAS_2BoostedBosons	139 fb^{-1}	ATLAS hadronic chargino/neutralino search [100]
ATLAS_0lep	139 fb^{-1}	ATLAS 0-lepton search [101]
ATLAS_0lep_stop	36 fb^{-1}	ATLAS 0-lepton stop search [102]
ATLAS_1lep_stop	36 fb^{-1}	ATLAS 1-lepton stop search [103]
ATLAS_2lep_stop	139 fb^{-1}	ATLAS 2-lepton stop search [104]
ATLAS_2OSlep_Z	139 fb^{-1}	ATLAS stop search with Z/H final states [105]
ATLAS_2OSlep_chargino	139 fb^{-1}	ATLAS 2-lepton chargino search [106]
ATLAS_2b	36 fb^{-1}	ATLAS 2-b-jet stop/sbottom search [107]
ATLAS_3b	24 fb^{-1}	ATLAS 3-b-jet Higgsino search [108]
ATLAS_3lep	139 fb^{-1}	ATLAS 3-lepton chargino/neutralino search [109]
ATLAS_4lep	139 fb^{-1}	ATLAS 4-lepton search [110]
ATLAS_MultiLep_strong	139 fb^{-1}	ATLAS leptons + jets search [111]
ATLAS_PhotonGGM_1photon	139 fb^{-1}	ATLAS 1-photon GGM search [112]
ATLAS_PhotonGGM_2photon	36 fb^{-1}	ATLAS 2-photon GGM search [113]
ATLAS_Z_photon	80 fb^{-1}	ATLAS Z + photon search [114]
CMS_0lep	137 fb^{-1}	CMS 0-lepton search [115]
CMS_1lep_bb	36 fb^{-1}	CMS 1-lepton + b-jets chargino/neutralino search [116]
CMS_1lep_stop	36 fb^{-1}	CMS 1-lepton stop search [117]
CMS_2lep_stop	36 fb^{-1}	CMS 2-lepton stop search [118]
CMS_2lep_soft	36 fb^{-1}	CMS 2 soft lepton search [119]
CMS_2OSlep	137 fb^{-1}	CMS 2-lepton search [120]
CMS_2OSlep_chargino_stop	36 fb^{-1}	CMS 2-lepton chargino/stop search [121]
CMS_2SSlep_stop	137 fb^{-1}	CMS 2 same-sign lepton stop search [122]
CMS_MultiLep	137 fb^{-1}	CMS multilepton chargino/neutralino search [123]
CMS_photon	36 fb^{-1}	CMS 1-photon GMSB search [124]
CMS_2photon	36 fb^{-1}	CMS 2-photon GMSB search [125]
CMS_1photon_1lepton	36 fb^{-1}	CMS 1-photon + 1-lepton GMSB search [126]

Analysis	Short Description	Implemented by
CMS-SUS-16-033	Supersymmetry in the multijet plus missing energy channel (35.9 fb-1)	F. Ambrogi and J. Sonneveld
CMS-SUS-16-039	Electroweakinos in the SS2L, 3L and 4L channels (35.9 fb-1)	B. Fuks and S. Mondal
CMS-SUS-16-048	Compressed electroweakinos with soft leptons (35.9 fb-1)	B. Fuks J.Y. Araz
CMS-SUS-16-052	SUSY in the 1l + jets channel (36 fb-1)	D. Sengupta
CMS-SUS-17-001	Stops in the OS dilepton mode (35.9 fb-1)	S.-M. Choi, S. Jeong, D.-W. Kang *et al.*
CMS-SUS-19-006	SUSY in the HT/missing HT channel (137 fb-1)	M. Mrowietz, S. Bein, J. Sonneveld
CMS-B2G-17-014	Vector-like quarks with charge 5/3 with same-sign dileptons (35.9/fb)	J. Salko, L. Panizzi
CMS-EXO-16-010	Mono-Z-boson (2.3 fb-1)	B. Fuks
CMS-EXO-16-012	Mono-Higgs (2.3 fb-1)	S. Ahn, J. Park, W. Zhang
CMS-EXO-16-022	Long-lived leptons (2.6 fb-1)	J. Chang M. Ustch, M. Goodsell
CMS-EXO-17-009	Leptoquark pair production in the electron(s)+jets channel (35.9 fb-1)	T. Murphy
CMS-EXO-17-011	WR and heavy neutrino in the 2l2j mode (35.9 fb-1)	A. Jueid, B. Fuks
CMS-EXO-17-015	Leptoquarks + dark matter in the 1mu+1jet+met channel (77.4 fb-1)	A. Jueid and B. Fuks
CMS-EXO-17-030	Pairs of trijet resonances (35.9 fb-1)	Y. Kang, J. Kim, J. Choi, S. Yun
CMS-EXO-19-002	Type-III seesaw and top-philic scalars with multileptons (137/fb)	E. Conte, R. Ducrocq
CMS-EXO-19-010	CMS disappearing tracks (139/fb)	M. Goodsell
CMS-EXO-20-002	WR and heavy neutrino in the 2l2j mode (138 fb-1)	A. Jueid, B. Fuks
CMS-EXO-20-004	Dark matter in the multi-jet+met channel (137 fb-1)	A. Albert
CMS-HIG-18-011	Exotic Higgs decay in the 2 muons + 2 b-jet channel via 2 pseudoscalars (35.9 fb-1)	J.B. Lee and J. Lee
CMS-TOP-17-009	SM four-top analysis (35.9 fb-1)	L. Darmé and B. Fuks
CMS-TOP-18-003	SM four-top analysis (137 fb-1)	L. Darmé and B. Fuks

The simplified scenarios used (we focussed on ‘higgsino’) involve W/Z decays of charginos/neutralinos, e.g.:



Can’t we look at the monojet + MET?
Classic claim that ‘higgsinos aren’t constrained by monojets’ comes because for *pure* higgsinos only one process is relevant:

$$pp \to \tilde{\chi}_1^0 \tilde{\chi}_0^2$$

All the others leave charged tracks

But when we have a mass splitting should include:

$$pp \to \tilde{\chi}_1^0 \tilde{\chi}_0^2, \tilde{\chi}_1^0 \tilde{\chi}_1^{\pm}, \tilde{\chi}_2^0 \tilde{\chi}_1^{\pm}, \tilde{\chi}_1^{\pm} \tilde{\chi}_1^{\mp}$$

Above the disappearing track limit have prompt decays + soft leptons/jets

To test this we used recasts of ATLAS and CMS monojet searches in MadAnalysis:

CMS-EXO-20-004 →

- MET > 250 GeV
- DeepAK algorithm to categorise leading jet as mono-W/Z/j
- Veto on leptons pT > 10 GeV
- Veto on bjets
- **Recast provided by CMS!!!**
- **Simplified likelihood also provided!!**

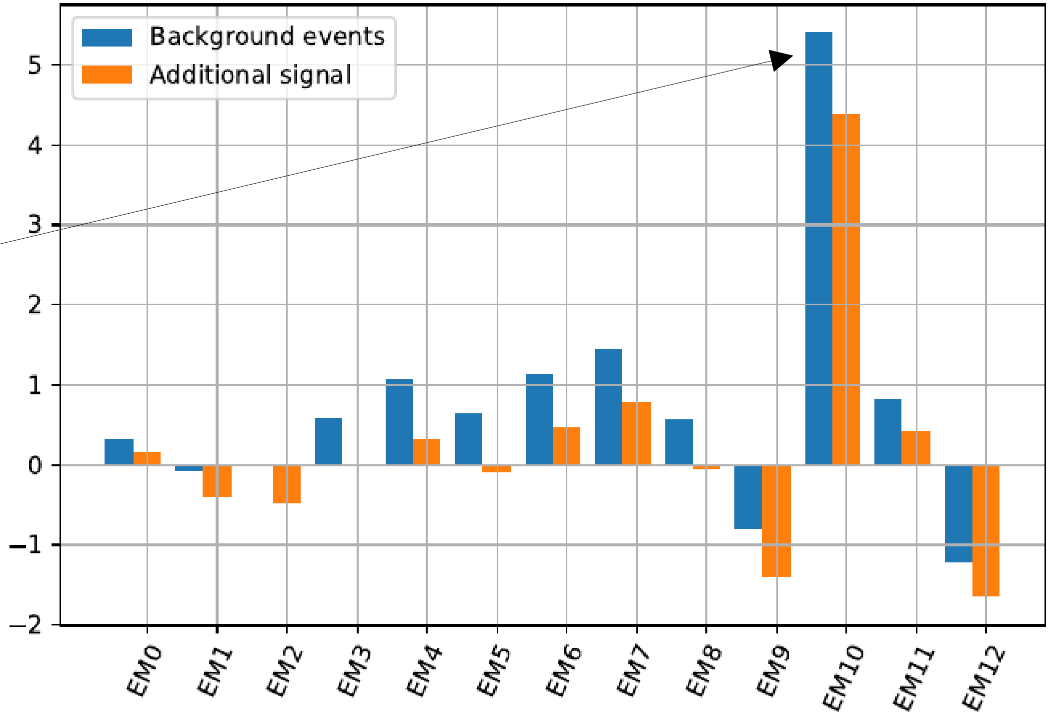
ATLAS-EXOT-2018-06 →

- MET > 200 GeV
- 13 exclusive bins in MET, largest > 1200 GeV
- Veto on leptons/photons pT > 7 GeV
- *Up to 3 additional jets allowed*
- **Recast performed by us (Diyar Agin)**
- **No likelihood information provided**

So we wanted to examine the excesses more closely

Exclusive Signal Region		
Region	Predicted	Observed
EM0	1 783 000 ± 26 000	1 791 624
EM1	753 000 ± 9000	752 328
EM2	314 000 ± 3500	313 912
EM3	140 100 ± 1600	141 036
EM4	101 600 ± 1200	102 888
EM5	29 200 ± 400	29 458
EM6	10 000 ± 180	10 203
EM7	3870 ± 80	3986
EM8	1640 ± 40	1663
EM9	754 ± 20	738
EM10	359 ± 10	413
EM11	182 ± 6	187
EM12	218 ± 9	207

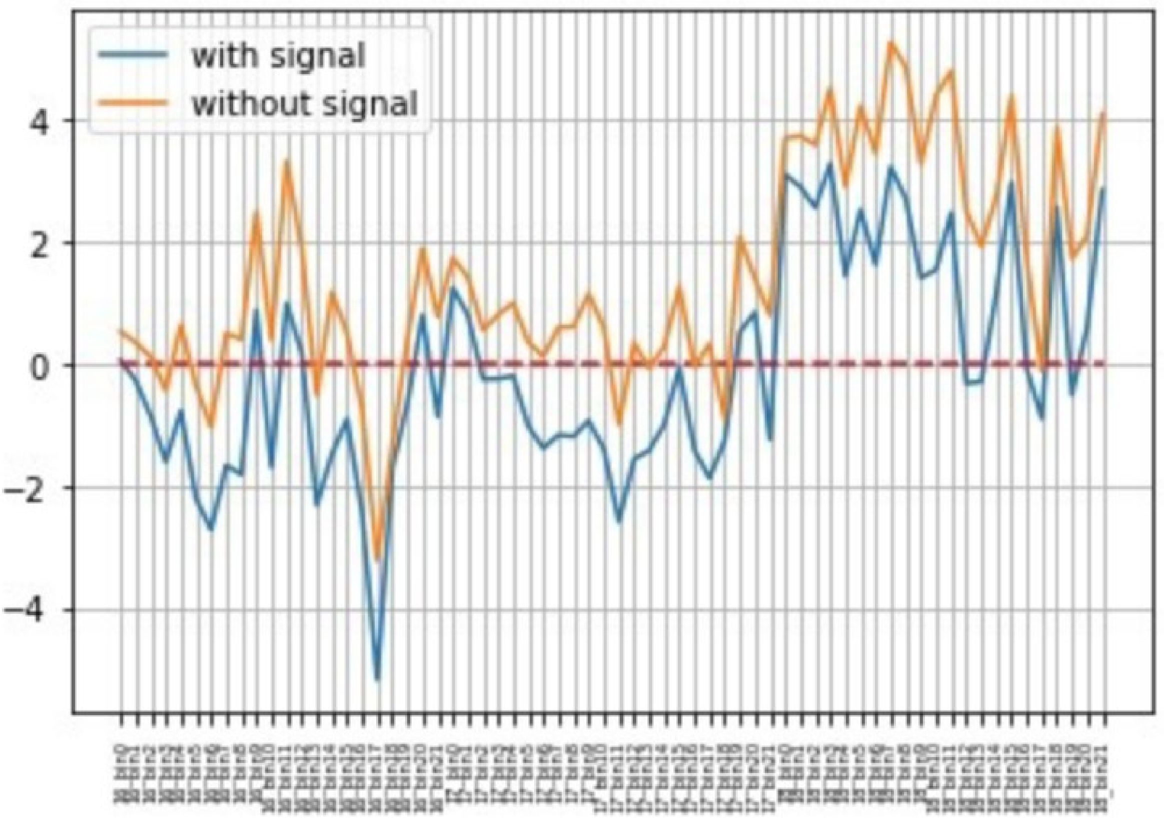
ATLAS have several small excesses and one giant one



But it is still just about visible when binned as inclusive regions:

Inclusive Signal Region		
Region	Predicted	Observed
IM0	3 120 000 ± 40 000	3 148 643
IM1	1 346 000 ± 16 000	1 357 019
IM2	597 000 ± 8000	604 691
IM3	286 000 ± 4000	290 779
IM4	146 400 ± 2300	149 743
IM5	45 550 ± 1000	46 855
IM6	16 800 ± 500	17 397
IM7	7070 ± 240	7194
IM8	3180 ± 130	3208
IM9	1560 ± 80	1545
IM10	720 ± 60	807
IM11	407 ± 34	394
IM12	223 ± 19	207

No such tables in the CMS paper … but we can inspect the accompanying HEPData and find many excesses in both low and high MET regions.



… and since we have statistical info, can look for ‘best fit’ points

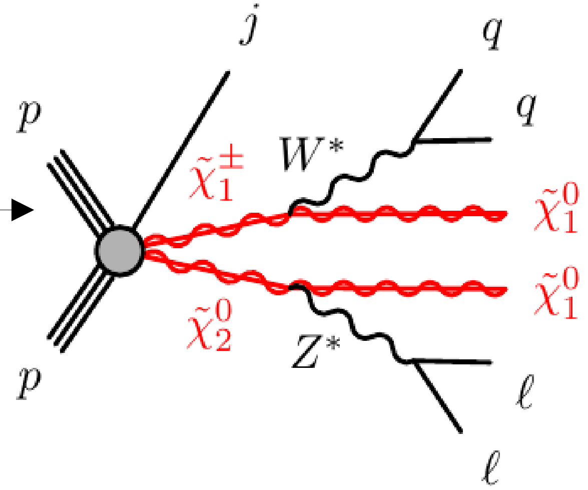
- For ATLAS we construct a naive χ^2 as a function of signal strength
- For CMS, can compute p-values for every point, varying signal strength

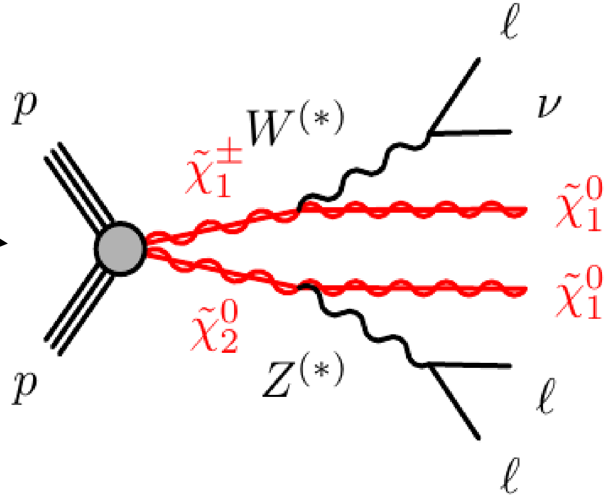
… now for soft leptons

ATLAS Analyses

2 ATLAS searches for:

- 2 soft leptons + ISR jet + (some) MET
  - ATLAS-SUSY-2018-16 where the excess is seen
- 3 soft leptons + either MET *or lepton trigger*
  - ATLAS-SUSY-2019-09 with no clear excess





ATLAS-SUSY-2018-16

Variable	Preselection requirements	
	2ℓ	$1\ell 1T$
Number of leptons (tracks)	= 2 leptons	= 1 lepton and ≥ 1 track
Lepton p_{T} [GeV]	$p_{\mathrm{T}}^{\ell_1} > 5$	$p_{\mathrm{T}}^{\ell} < 10$
$\Delta R_{\ell\ell}$	$\Delta R_{ee} > 0.30$, $\Delta R_{\mu\mu} > 0.05$, $\Delta R_{e\mu} > 0.2$	$0.05 < \Delta R_{\ell\mathrm{track}} < 1.5$
Lepton (track) charge and flavor	$e^{\pm}e^{\mp}$ or $\mu^{\pm}\mu^{\mp}$	$e^{\pm}e^{\mp}$ or $\mu^{\pm}\mu^{\mp}$
Lepton (track) invariant mass [GeV]	$3 < m_{ee} < 60$, $1 < m_{\mu\mu} < 60$	$0.5 < m_{\ell\mathrm{track}} < 5$
J/ψ invariant mass [GeV]	veto $3 < m_{\ell\ell} < 3.2$	veto $3 < m_{\ell\mathrm{track}} < 3.2$
$m_{\tau\tau}$ [GeV]	< 0 or > 160	no requirement
$E_{\mathrm{T}}^{\mathrm{miss}}$ [GeV]	> 120	> 120
Number of jets	≥ 1	≥ 1
Number of b-tagged jets	$= 0$	no requirement
Leading jet p_{T} [GeV]	≥ 100	≥ 100
$\min(\Delta\phi(\text{any jet}, \mathbf{p}_{\mathrm{T}}^{\mathrm{miss}}))$	> 0.4	> 0.4
$\Delta\phi(j_1, \mathbf{p}_{\mathrm{T}}^{\mathrm{miss}})$ †	≥ 2.0	≥ 2.0

Variable	Electroweakino SR Requirements			
	SR–E–low	SR–E–med	SR–E–high	SR–E–$1\ell 1T$
$E_{\mathrm{T}}^{\mathrm{miss}}$ [GeV]	[120, 200]	[120, 200]	> 200	> 200
$E_{\mathrm{T}}^{\mathrm{miss}}/H_{\mathrm{T}}^{\mathrm{lep}}$	< 10	> 10	–	> 30
$\Delta\phi(\mathrm{lep}, \mathbf{p}_{\mathrm{T}}^{\mathrm{miss}})$	–	–	–	< 1.0
Lepton or track p_{T} [GeV]	$p_{\mathrm{T}}^{\ell_2} > 5 + m_{\ell\ell}/4$	–	$p_{\mathrm{T}}^{\ell_2} > \min(10, 2 + m_{\ell\ell}/3)$	$p_{\mathrm{T}}^{\mathrm{track}} < 5$
$M_{\mathrm{T}}^{\mathrm{S}}$ [GeV]	–	< 50	–	–
$m_{\mathrm{T}}^{\ell_1}$ [GeV]	[10, 60]	–	< 60	–
R_{ISR}	[0.8, 1.0]	–	$[\max(0.85, 0.98 - 0.02 \times m_{\ell\ell}),\ 1.0]$	–

Restframes quantities to identify ISR jet

From ATLAS-SUSY-2018-16

	Signal Region	N_{obs}	N_{exp}	$\langle\epsilon\sigma\rangle^{95}_{obs}$ [fb]	S^{95}_{obs}	S^{95}_{exp}	$p(s=0)$
SR–E	$m_{\ell\ell} < 1$	0	1.0 ± 1.0	0.022	3.0	$3.0^{+1.3}_{-0.0}$	0.50
	$m_{\ell\ell} < 2$	46	44 ± 6.8	0.15	21	19^{+7}_{-5}	0.38
	$m_{\ell\ell} < 3$	90	77 ± 12	0.29	41	31^{+11}_{-9}	0.18
	$m_{\ell\ell} < 5$	151	138 ± 18	0.38	52	43^{+16}_{-11}	0.24
	$m_{\ell\ell} < 10$	244	200 ± 19	0.62	86	49^{+26}_{-13}	0.034
	$m_{\ell\ell} < 20$	383	301 ± 23	0.95	132	61^{+22}_{-16}	0.0034
	$m_{\ell\ell} < 30$	453	366 ± 27	1.04	144	70^{+26}_{-20}	0.0065
	$m_{\ell\ell} < 40$	492	420 ± 30	0.96	134	74^{+29}_{-20}	0.027
	$m_{\ell\ell} < 60$	583	520 ± 35	0.97	135	84^{+32}_{-23}	0.063

2.7σ local excess for ‘signal model with unconstrained normalisation’

Maybe a different model would give a stronger significance?
(what we’re looking into now ...)

CMS Analysis details

Targets the same W-Z channel, but now ‘all-in-one’ analysis with whole of Run 2 data subsuming preliminary one:

CMS-SUS-18-004

Search region	Low-MET		Med-MET	High-MET	Ultra-MET
	Raw $p_{\mathrm{T}}^{\mathrm{miss}}$	$p_{\mathrm{T}}^{\mathrm{miss}}$	$p_{\mathrm{T}}^{\mathrm{miss}}$	$p_{\mathrm{T}}^{\mathrm{miss}}$	$p_{\mathrm{T}}^{\mathrm{miss}}$
2ℓ-Ewk	> 125	$(125, 200]$	$(200, 240]$	$(240, 290]$	> 290
2ℓ-Stop	> 125	$(125, 200]$	$(200, 290]$	$(290, 340]$	> 340
3ℓ-Ewk	> 125	$(125, 200]$		> 200	

Very similar to ATLAS

Except: maximum pT for leptons

…. and no RestFrames

Naively should be more permissive regarding models

Variable	2ℓ-Ewk Low-MET	2ℓ-Ewk Higher-MET	2ℓ-Stop Low-MET	2ℓ-Stop Higher-MET	3ℓ-Ewk Low-MET	3ℓ-Ewk Higher-MET
N_{lep}	2	2	2	2	3	3
$p_{\text{T}}(\ell_1)$ [GeV] for e(μ)	$(5, 30)$	$(5(3.5), 30)$	$(5, 30)$	$(5(3.5), 30)$	$(5, 30)$	$(5(3.5), 30)$
$p_{\text{T}}(\ell_2)$ [GeV] for e(μ)	$(5, 30)$	$(5(3.5), 30)$	$(5, 30)$	$(5(3.5), 30)$	$(5, 30)$	$(5(3.5), 30)$
$p_{\text{T}}(\ell_3)$ [GeV] for e(μ)	—	—	—	—	$(5, 30)$	$(5(3.5), 30)$
1 OS pair	✓	✓	✓	✓	✓	✓
1 OSSF pair	✓	✓	✓	—	✓	✓
$\Delta R(\ell_i \ell_j)$ $(i, j = 1, 2, 3,\ i \neq j)$	—	> 0.3	—	> 0.3	—	> 0.3
$M_{\text{SFOS}}(\ell\ell)$ ($M^{\text{min}}_{\text{SFOS}}(\ell\ell)$ in 3ℓ) [GeV]	$(4, 50)$	$(1, 50)$	$(4, 50)$	$(1, 50)$	$(4, 50)$	$(1, 50)$
$M^{\text{max}}_{\text{SFAS}}(\ell\ell)$ (AS=any sign) [GeV]	—	—	—	—	< 60	—
$M_{\text{SFOS}}(\ell\ell)$ ($M^{\text{min}}_{\text{SFOS}}(\ell\ell)$ in 3ℓ) [GeV]	veto $(3, 3.2)$ and $(9, 10.5)$					
$p_{\text{T}}(\ell\ell)$ [GeV]	> 3		> 3		—	
Leading jet "Tight lepton veto"	✓		✓		—	
$m_{\text{T}}(\ell_i, p^{\text{miss}}_{\text{T}})$ [GeV] $(i = 1, 2)$	< 70		—		—	
H_{T} [GeV]	> 100					
$p^{\text{miss}}_{\text{T}} / H_{\text{T}}$	$(2/3, 1.4)$		$(2/3, 1.4)$		—	
$\text{N}_b(p_{\text{T}} > 25\,\text{GeV})$	$= 0$					
$M_{\tau\tau}$ [GeV]	veto $(0, 160)$		veto $(0, 160)$		—	

Recasting the CMS soft lepton search

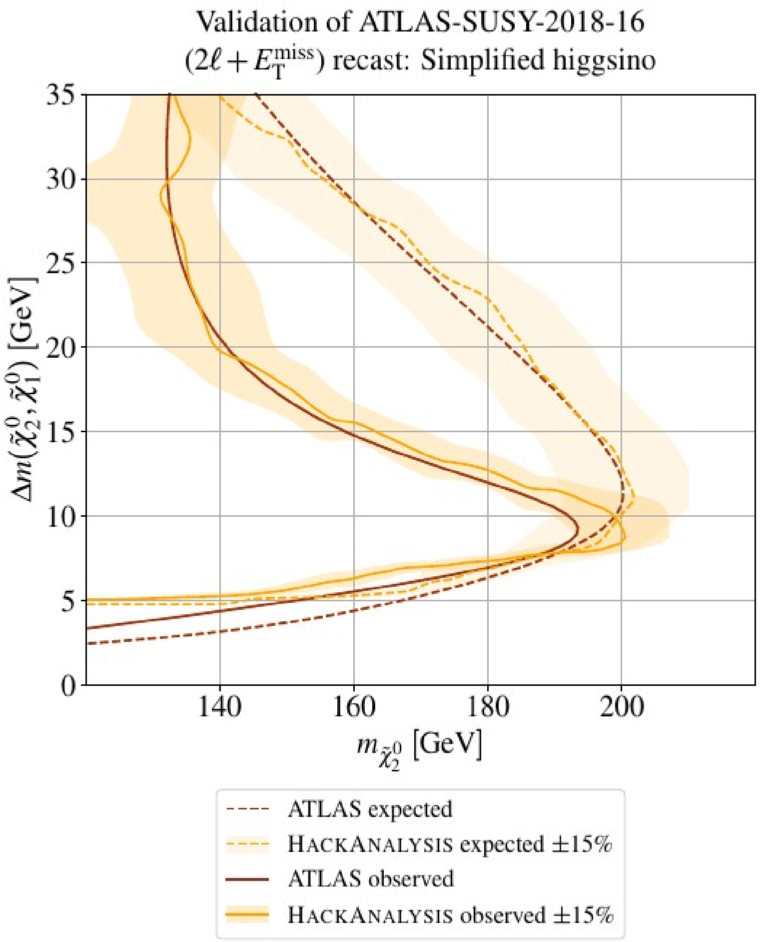
- Efficiency info from previous soft lepton analysis (lepton reconstruction, b-tagging)
- What is lacking at the moment is MET reconstruction/trigger efficiency … but we hope we are nearly there.
- Not much information available on the differences in year, but I dealt with that before
- Cutflows and simplified likelihood were provided.
- Then will be able to compare excesses in CMS & ATLAS, and compare models

Recasting the ATLAS soft lepton analyses

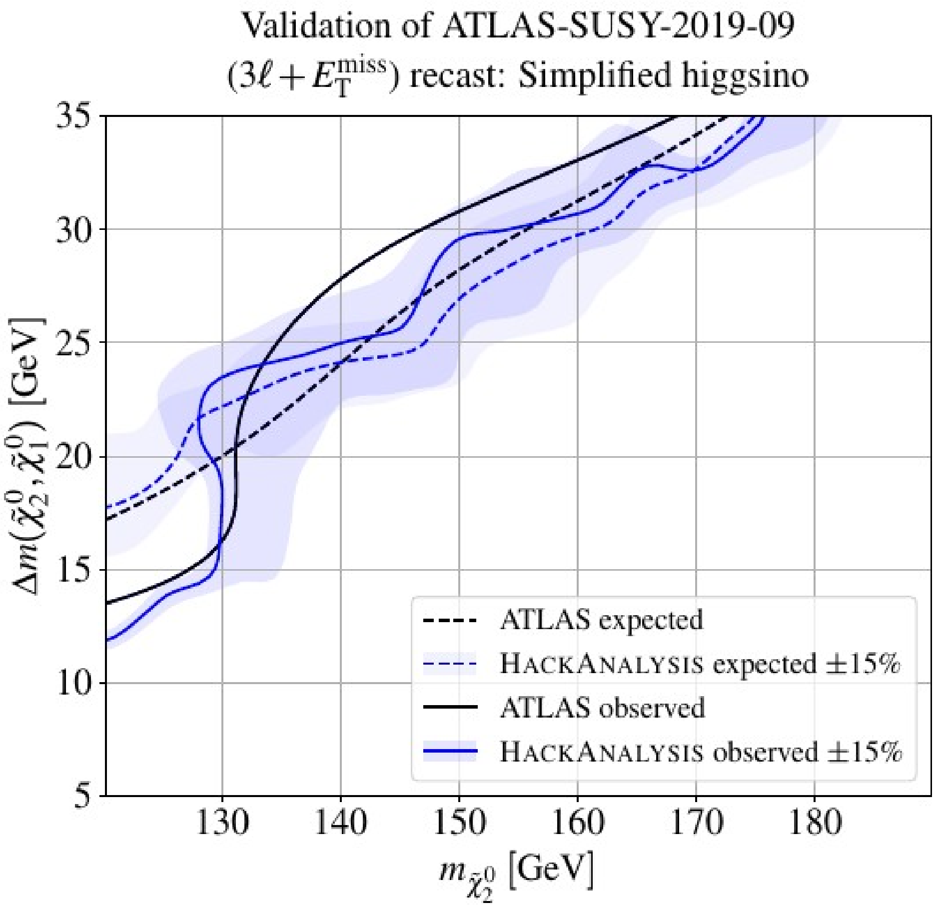
- Needed to recast two separate analyses to cover the ‘excess’ region
- Both analyses provided pseudocode, cutflows and pyhf statistical models
- Efficiency information was provided for both

However:

- Restframes requires the dreaded root
- Many SimpleAnalysis routines required by both are not public
- The 3-lepton analysis requires lepton triggers and efficiency information was very hard to dig out/emulate



2 soft leptons



3 soft leptons

Application: realistic MSSM models

- Typical cross-sections for EWinos around 200 GeV are about 1pb
- Searches are therefore sensitive to efficiencies around $10^{-4} - 10^{-5}$
- In toy model, decay of Z → leptons with ~ 10% branching ratio
- When generating samples for the toy model, can bias event generation – only need to simulate O(1M) events to get sufficient statistics:

$$\text{rel. uncert.} = \frac{1}{\sqrt{\epsilon N}} \longrightarrow N \sim 0.01/\epsilon \text{ for } 10\%$$

BUT:

- Realistic MSSM points have complicated decay chains involving intermediate EWinos.
- End up having to simulate O(10M) events per point (lose half from MLM matching too)
- We postponed full scans to the second paper