Implementation of BDT and NN searches in CheckMATE

Krzysztof Rolbiecki
(U. Warsaw)

Reinterpretation & OpenMAPP mini-workshop
Grenoble, 17 – 20 June 2024

NATIONAL SCIENCE CENTRE
POLAND

Contents

- NN example 1, arxiv:2106.09609, SUSY-2019-04
- NN example 2, arXiv:2211.08028, SUSY-2018-30
- Boosted decision tree, arxiv:2010.14293, SUSY-2018-22
- BDT example 2, arXiv:2209.13935, SUSY-2019-02

Examples available at https://github.com/CheckMATE2/checkmate2 but not yet merged into master branch

CheckMATE for recasting

- CheckMATE is a general tool for recasting arbitrary model
- Accepts events as .hepmc, .lhe; integration with Pythia and MadGraph
- based on Delphes for detector simulation
- using existing LHC searches calculates a limit on a given parameter point
- From SLHA file to the limit in one click
- one can easily constrain models that were not covered in the original ATLAS/CMS search
- Currently around 50 searches at 13 TeV coded, including 24 with full luminosity
- long-lived particles branch
- https://checkmate.hepforge.org/ and https://github.com/CheckMATE2/checkmate2

NN example
arXiv: 2106.09609

Arxiv:2106.09609

- Search for RPV-SUSY in final states with leptons and many jets (0 or 3 b-tagged)
- Signal regions count the number of jets with different pT thresholds; in general, 6-15 jets, at least 1 lepton
- Target: stops, gluino and EW higgsinos/winos
- EW signal: neutralino -> tbs; chargino -> bbs
- EW SR: 1 lepton, =6 jets, >=4 b-jets, NN discriminant
- NN released as ONNX files (in total 5, each for different jet multiplicity 4,5,6,7,8); unfortunately, very little information is provided

NN and CheckMATE implementation

- Using ONNX Runtime, https://onnxruntime.ai
- C++ library, analysis is performed on the event-by-event basis

- NN has 65 inputs: jets energy, rapidity, azimuth; MET, b-jet multiplicity; distance between jets and leading lepton, etc; some high-level combinations of jet momenta (as invariant masses);
- Problematic: b-tagging score for each jet based on DLR1 b-tagging algorithm (the pseudocode takes fixed values: 5 for b-jet; 1 for non-b-jet)

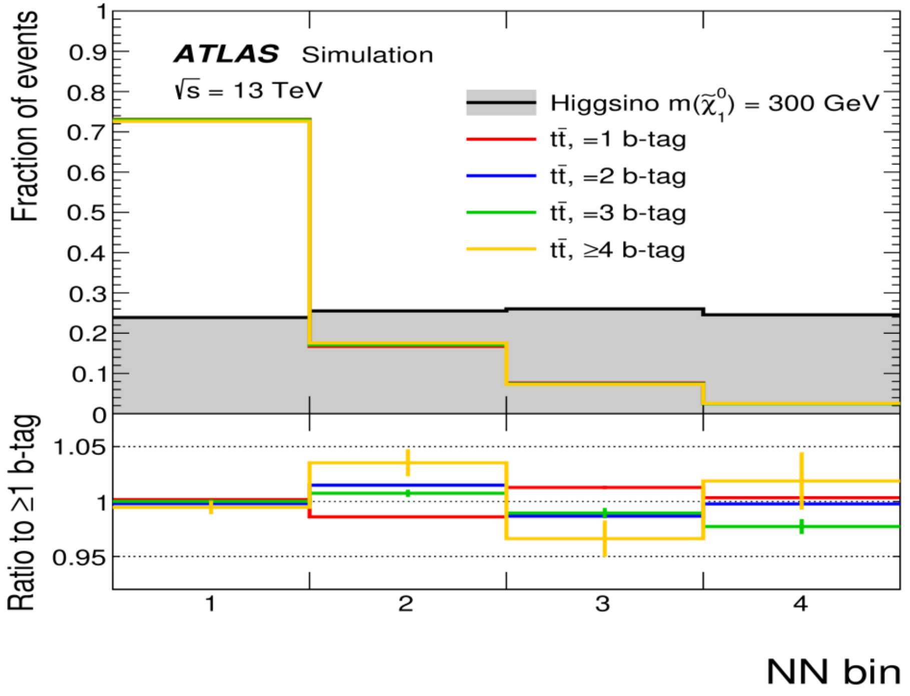
Validation - cutflow

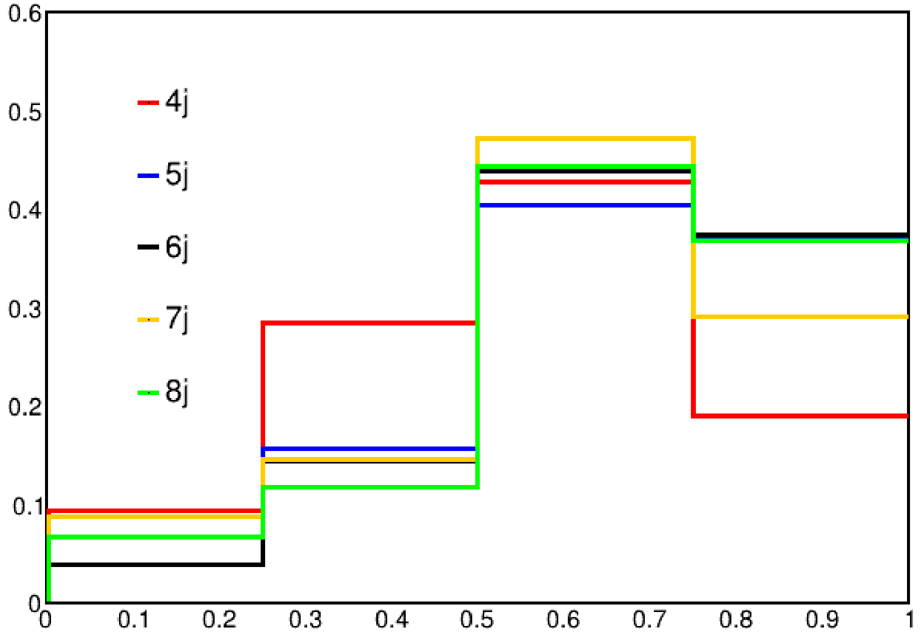
- Pretty much everything went wrong
- Clearly a problem with lepton id
- Too few events with high jet multiplicity
- After b-tagging things look better
- After NN inference the results are somewhat random (not surprising after looking at the histogram)
- Perhaps a general problem with signal modelling

$\tilde{\chi}_{1,2} \to tbs$	ATLAS	CheckMATE
All	14491	14491
Lead lep $p_T > 27$ GeV	5413	3304
$==$ 4 jets	681	549
$==$ 5 jets	1101	766
$==$ 6 jets	1188	734
$==$ 7 jets	840	466
$==$ 8 jets	420	186
$==$ 4 jets, $\geq$ 4b	7	5.4
$==$ 5 jets, $\geq$ 4b	29	33
$==$ 6 jets, $\geq$ 4b	57	43
$==$ 7 jets, $\geq$ 4b	61	41
$==$ 8 jets, $\geq$ 4b	39	21
NN_{4j} bin 4	2.9	0.7
NN_{5j} bin 4	8.7	6
NN_{6j} bin 4	17.4	11
NN_{7j} bin 4	18.9	19
NN_{8j} bin 4	14.5	10

Validation

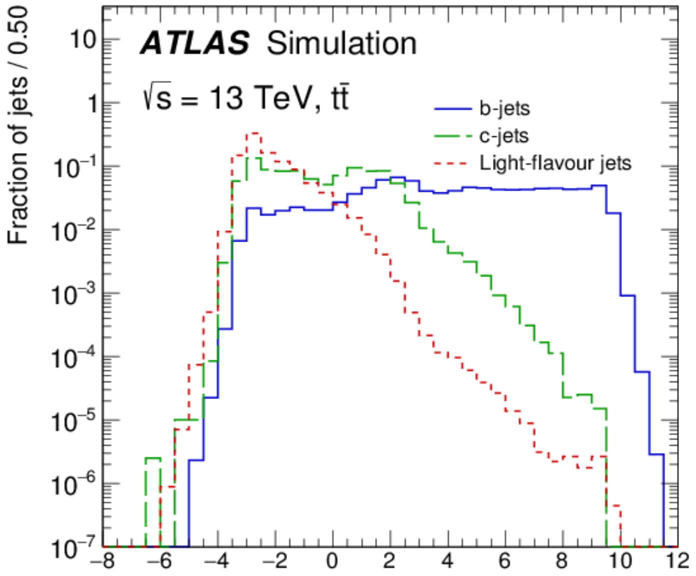
- Cutflows do not seem very useful; compare figures of NN output (unclear definitions of "bins" though)
- ATLAS distribution flat for signal
- Should be insensitive to b-tag, but it's not apparent in recast (e.g. large variation when using different ad-hoc b-tag scores)



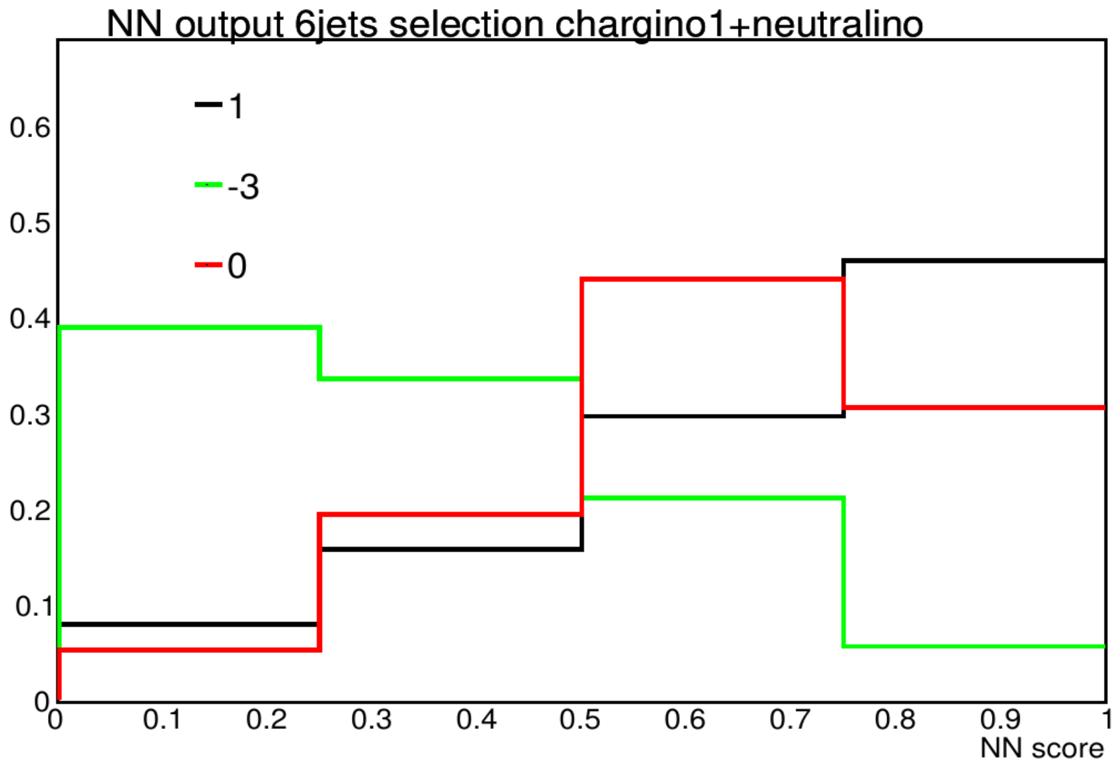


Comparison for light jets

- Jets are assigned a score using a BDT tagger
- In recast we take ad-hoc values following SimpleAnalysis snippet
- The result seems to be very sensitive for reasonable choices of tagging score
- -1 seems to reproduce ATLAS better than 0 from SimpleAnalysis



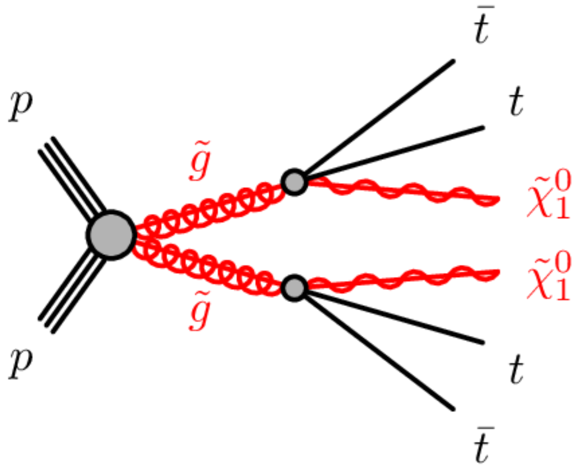
FTAG-2018-01

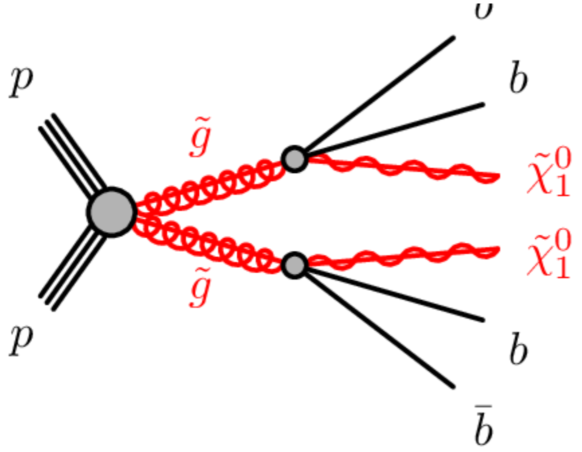


Another NN example
arXiv: 2211.08028

arXiv: 2211.08028

- Search for gluinos decaying to 3rd generation quarks
- Final state: at least 4 jets, at least 3 b-tagged jets, MET, 0-1 leptons (more allowed in NN analysis)
- 8 NN signal regions: 4 for gluino decaying to top pair and 4 for gluino decaying to bottom pair (still it is one net)
- The choice of the desired SR is via the last three inputs (I.e. decay type, and target masses)
- The NN has 87 input parameters: jet (small and large R) momenta, lepton momenta, MET and b-tag category (binary)
- The output gives separate background and signal probabilities





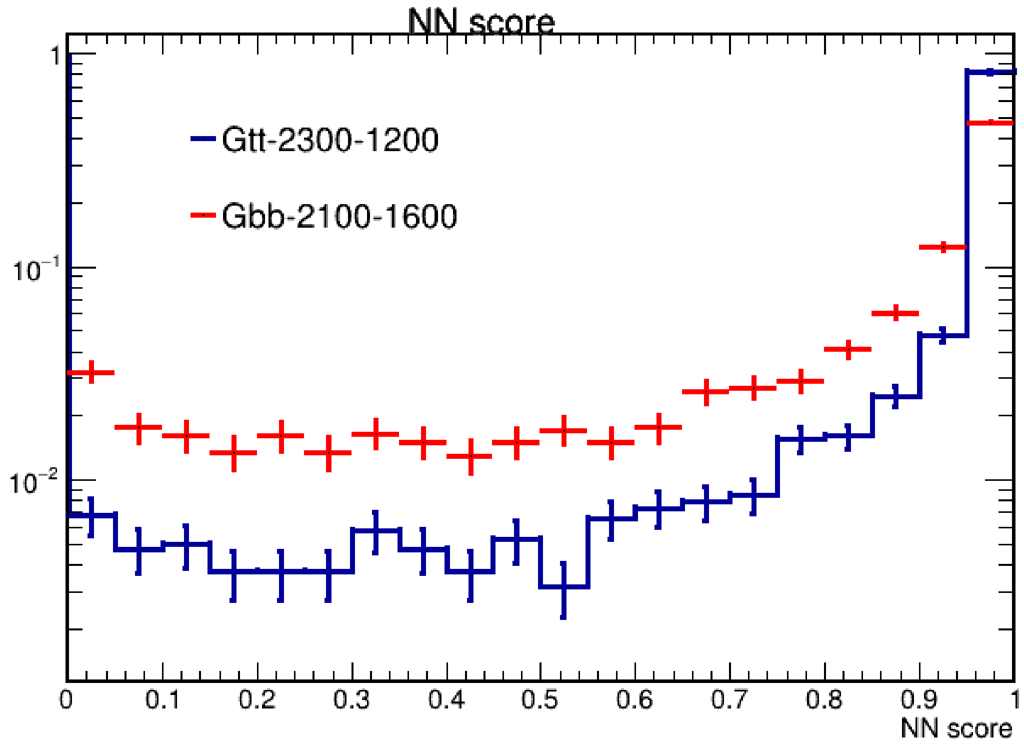
Validation

- Reasonable agreement across all channels
- The efficiency somewhat lower but similar and consistent effect also seen for cut-and-count analysis
- This is still a preprint so not too much material for comparison and testing was rather quick (nothing changed actually)

	ATLAS	CheckMATE
Gtt selection		
Common rquirem.	7.66	7.30
SR-Gtt-2100-1	2.63	1.94
SR-Gtt-1800-1	2.80	2.11
SR-Gtt-2300-1200	2.95	2.62
SR-Gtt-1900-1400	0.19	0.27
Gbb selection		
Common rquirem.	80	65
SR-Gbb-2800-1400	22	14
SR-Gbb-2300-1000	21	14
SR-Gbb-2100-1600	6.20	6.80
SR-Gbb-2000-1800	0.19	0.58

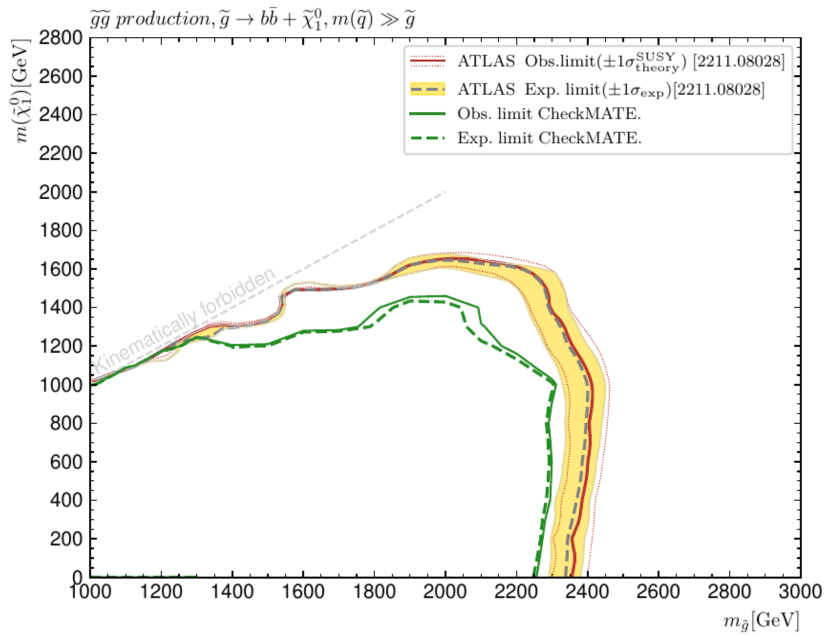
NN output comparison

- True signal is Gtt m_{gluino} = 2400, m_{neut} = 1000
- Target signal for Gtt-2300-1200
- Less compatible with Gbb as it should be
- The last bin pronounced for Gbb but the cutoff is at 0.9993 so the acceptance is actually very low
- No figures from ATLAS for comparison

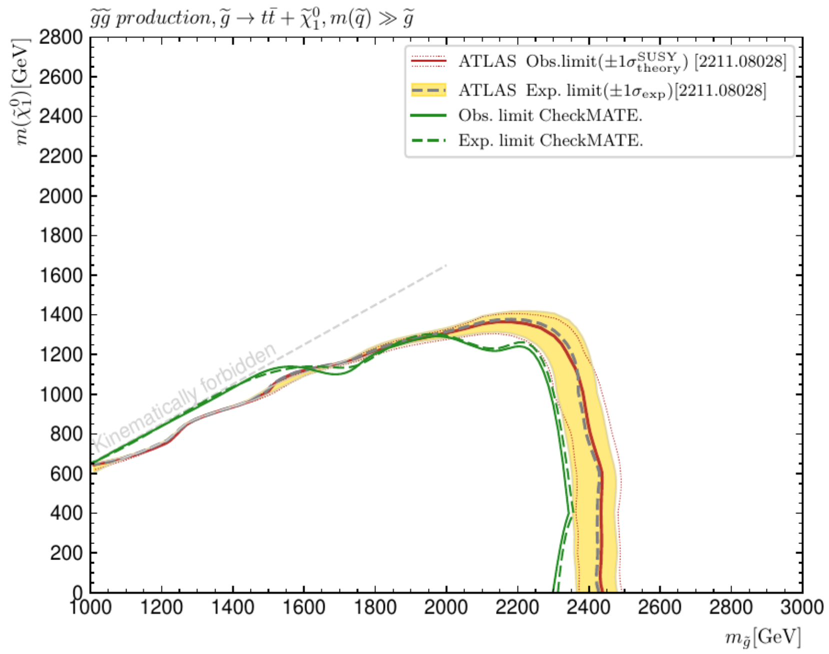


Validation

- Reasonable agreement (within 30%) for b-squarks
- Disagreement mainly due to b-tagging efficiency (not specifically tuned)



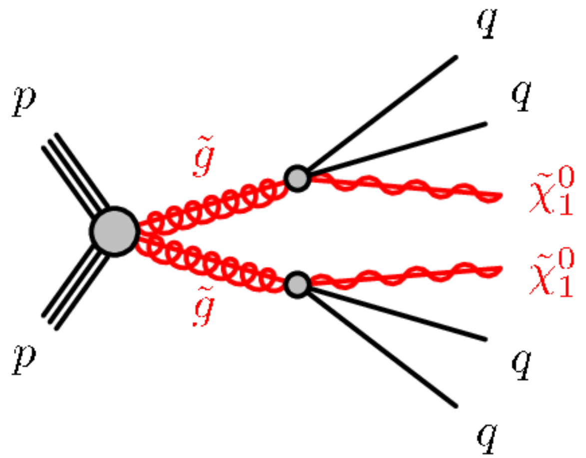
- Better yet for t-squarks (note a complicated final state)

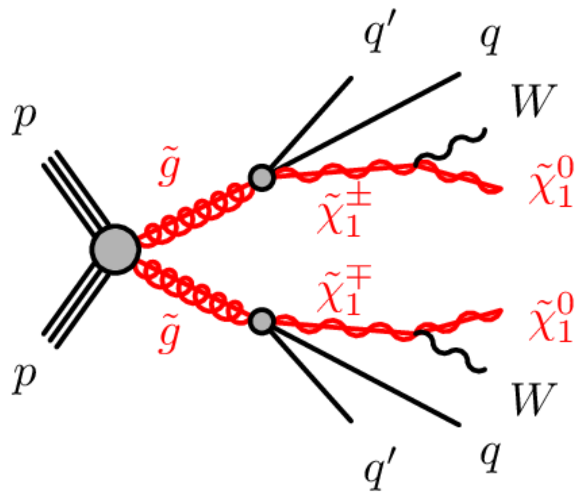


Boosted Decision Tree example

arXiv: 2010.14293

- Search for squarks and gluinos
- Final state: 2-6 jets + MET
- Principal variables: m_{eff} , leading jet pT and MET
- Multi-bin signal regions
- 8 boosted decision tree SRs, targeting gluino production with
  direct and indirect decays
- BDT weights released as .xml files for using with ROOT Toolkit for Multivariate Data Analysis -- TMVA





Validation

- Generally excellent agreement
- Here example for model-independent discovery channels
- See Iñaki's talk for details of multi-bin fits and validation
- Full note:

checkmate.hepforge.org

	Selection	$m_{\tilde{q}} = 1200$ GeV, $m_{\tilde{\chi}_1^0} = 600$ GeV		$m_{\tilde{q}} = 1400$ GeV, $m_{\tilde{\chi}_1^0} = 600$ GeV		$m_{\tilde{q}} = 1600$ GeV, $m_{\tilde{\chi}_1^0} = 400$ GeV	
		ATLAS	CM	ATLAS	CM	ATLAS	CM
Generated MC events		10000	10000	6000	10000	6000	10000
Common Requirements	Preselection, $E_T^{\text{miss}} > 300$ GeV, $p_T(j_1) > 200$ GeV, $m_{\text{eff}} > 800$ GeV	1763	1780	541	546	174	176
	jet multiplicity ≥ 2	1763	1780	541	546	174	176
	Cleaning cuts	1746	–	535	–	173	–
SR-2j-1600	$\Delta\phi(j_{1,2,(3)}, E_T^{\text{miss}}) > 0.8$	1433	1434	431	433	136	139
	$\Delta\phi(j_{i>3}, E_T^{\text{miss}}) > 0.4$	1377	1353	411	410	129	130
	$p_T(j_2) > 250$ GeV	853	850	311	310	111	112
	$\lvert\eta(j_{1,2})\rvert < 2.0$	836	832	306	305	109	110
	$E_T^{\text{miss}}/\sqrt{H_T} > 16$ GeV$^{1/2}$	568	554	228	227	86.4	87.3
	$m_{\text{eff}}(\text{incl.}) > 1600$ GeV	366	362	202	195	83.5	84.2
SR-2j-2200	$\Delta\phi(j_{1,2,(3)}, E_T^{\text{miss}}) > 0.4$	1603	1619	483	492	156	158
	$\Delta\phi(j_{i>3}, E_T^{\text{miss}}) > 0.2$	1567	1566	470	476	151	153
	$p_T(j_1) > 600$ GeV	509	514	269	259	120	121
	$E_T^{\text{miss}}/\sqrt{H_T} > 16$ GeV$^{1/2}$	337	339	201	188	94.6	95.7
	$m_{\text{eff}}(\text{incl.}) > 2200$ GeV	101	96	108	101	76.1	76.4
SR-2j-2800	$\Delta\phi(j_{1,2,(3)}, E_T^{\text{miss}}) > 0.8$	1433	1434	431	433	136	138
	$\Delta\phi(j_{i>3}, E_T^{\text{miss}}) > 0.4$	1377	1352	411	410	129	130
	$p_T(j_2) > 250$ GeV	853	850	311	311	111	112
	$\lvert\eta(j_{1,2})\rvert < 1.2$	655	653	235	239	82.3	84.3
	$E_T^{\text{miss}}/\sqrt{H_T} > 16$ GeV$^{1/2}$	439	433	173	178	64.6	66.4
	$m_{\text{eff}}(\text{incl.}) > 2800$ GeV	15.6	10.5	18.8	15.1	29.1	27.0

BDT input

Input variables	GGd1	GGd2	GGd3	GGd4
$E_{\mathrm{T}}^{\mathrm{miss}}$ [GeV]	•	–	•	
$p_{\mathrm{T}}(j)$ [GeV]	$p_{\mathrm{T}}(j_1), p_{\mathrm{T}}(j_2), p_{\mathrm{T}}(j_3), p_{\mathrm{T}}(j_4)$			
$\eta(j)$	$\eta(j_1), \eta(j_2), \eta(j_3), \eta(j_4)$			
Aplanarity	•			–
m_{eff} [GeV]	•			
Total number of input variables	11	10	11	10

Input variables	GGo1	GGo2	GGo3	GGo4
$E_{\mathrm{T}}^{\mathrm{miss}}$ [GeV]	–			•
$p_{\mathrm{T}}(j)$ [GeV]	$p_{\mathrm{T}}(j_1), p_{\mathrm{T}}(j_2), p_{\mathrm{T}}(j_3), p_{\mathrm{T}}(j_4), p_{\mathrm{T}}(j_5)$	$p_{\mathrm{T}}(j_1), p_{\mathrm{T}}(j_2), p_{\mathrm{T}}(j_3), p_{\mathrm{T}}(j_4)$	$p_{\mathrm{T}}(j_1), p_{\mathrm{T}}(j_2), p_{\mathrm{T}}(j_3), p_{\mathrm{T}}(j_5), p_{\mathrm{T}}(j_6)$	$p_{\mathrm{T}}(j_1), p_{\mathrm{T}}(j_2), p_{\mathrm{T}}(j_3), p_{\mathrm{T}}(j_4)$
$\eta(j)$	$\eta(j_1), \eta(j_2), \eta(j_3), \eta(j_4), \eta(j_5)$	$\eta(j_1), \eta(j_2), \eta(j_3), \eta(j_4)$	$\eta(j_1), \eta(j_2), \eta(j_3), \eta(j_5), \eta(j_6)$	$\eta(j_1), \eta(j_2), \eta(j_3), \eta(j_4)$
Aplanarity	•			–
m_{eff} [GeV]	•			
Total number of input variables	12	10	12	10

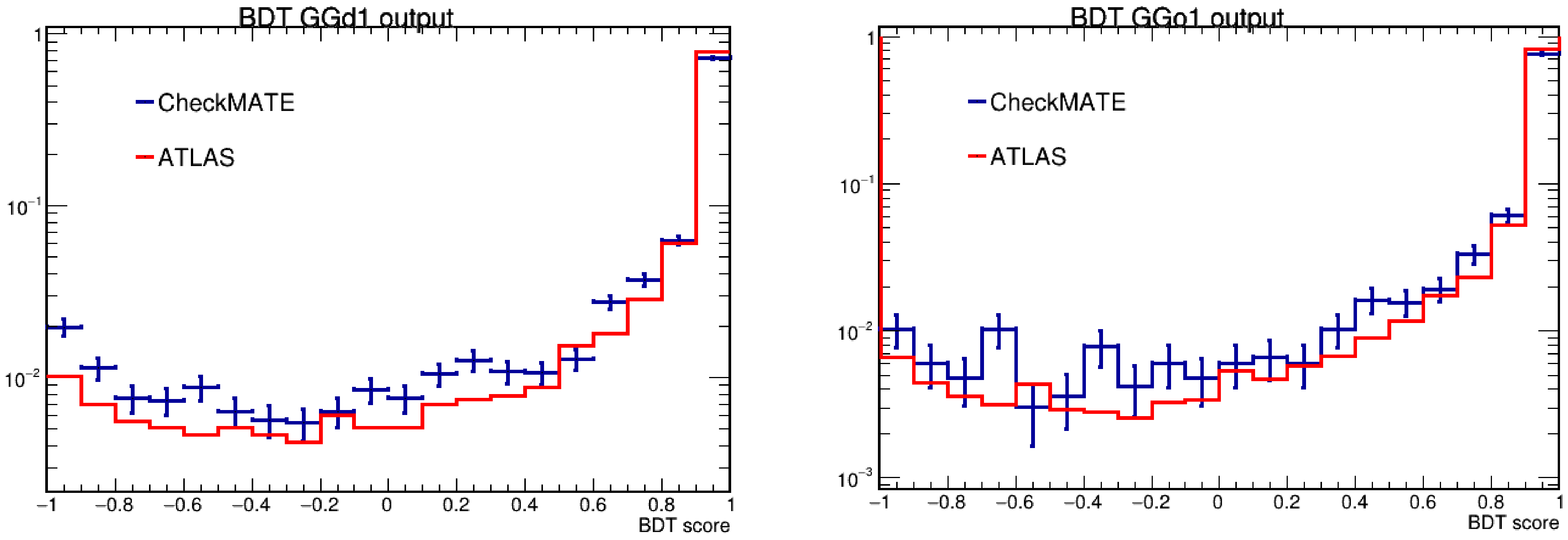
BDT validation

- Each SR targets direct gluino decays for specific range in $\Delta m = m_g - m_{neut}$
- GGd1: Δm = 1600-1900 GeV
- GGd2: Δm = 1000-1400 GeV
- GGd3: Δm = 600-1000 GeV
- GGd4: Δm = 200-600 GeV
- Overall, very good agreement
- No specific exclusion contour to compare with

$m_{\tilde{g}}$	2200		2200		1800		1400	
$m_{\tilde{\chi}_1^0}$	500		1000		1000		1000	
	A	C	A	C	A	C	A	C
GGd1	14.1	12.5	7.04	5.5	5.5	4.2	3.0	3.0
GGd2	14.3	13.4	11.4	10.1	19.4	14.3	8.8	9.9
GGd3	14.4	14.1	14.4	13.8	71.7	62.0	49.1	43.7
GGd3	2.9	3.4	6.0	6.1	60.5	54.0	89.6	85.8

A = ATLAS; C = CheckMATE

BDT output comparison – preferred for validation

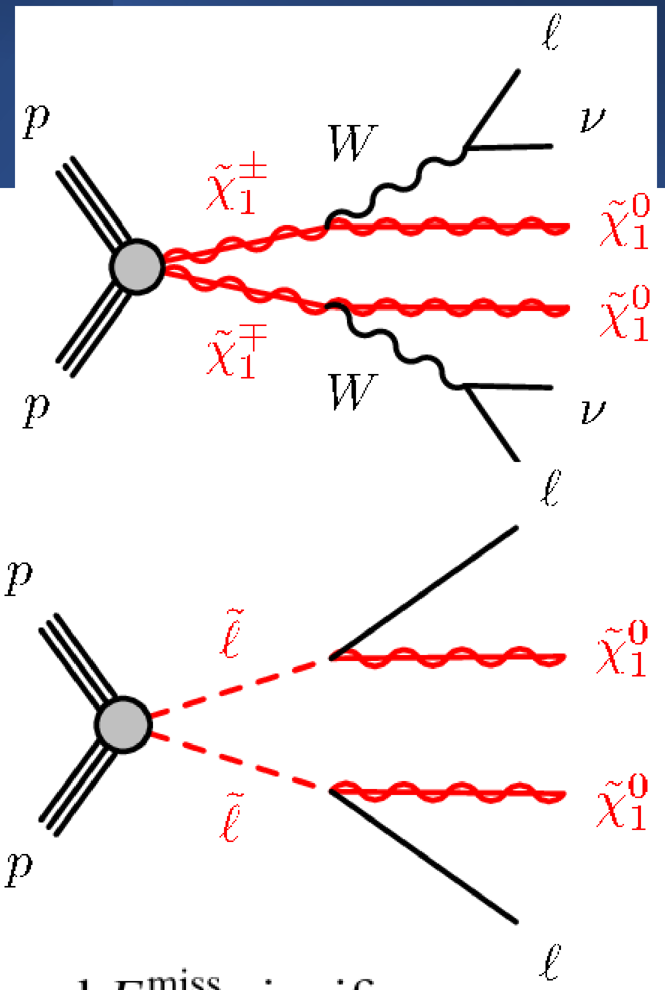


- Both signal regions show good agreement
- GGd1 = direct decay; GGo1 = one step decay

BDT example 2

arXiv:2209.13935

arXiv: 2209.13935



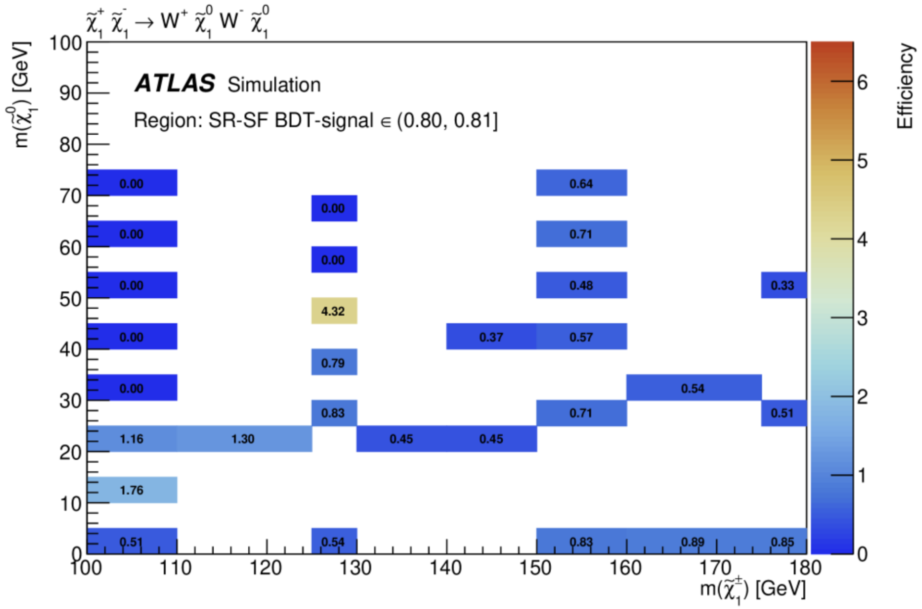
- Search for sleptons and charginos
- Final state: 2 leptons+MET, up to 1 jet
- BDTs for chargino signal (0J SF/DF)
- Also multibin with BDT output but no likelihood :(
- And yet a new way of implementation and providing trees: this time using MVAUtils
- Variables for BDT: $m_{\ell\ell}$, $\Delta\phi_{\mathrm{boost}}$, $\Delta\phi_{p_{\mathrm{T}}^{\mathrm{miss}},\ell_1}$, $\Delta\phi_{p_{\mathrm{T}}^{\mathrm{miss}},\ell_2}$, $\cos\theta^*_{\ell\ell}$, and $E_{\mathrm{T}}^{\mathrm{miss}}$ significance.
  $p_{\mathrm{T}}^{\ell_1}$, $p_{\mathrm{T}}^{\ell_2}$, $E_{\mathrm{T}}^{\mathrm{miss}}$, m_{T2}^0,

Validation pending

$m\left(\tilde{\chi}_1^{\pm},\right) = (125, 25)$ GeV	Number of events
Total events	75711.3
Events with only 2 leptons, with $p_T > 9$ GeV	16790.0
Events passing cleaning cuts	16788.8
Trigger & $p_T^{\ell_1} > 27$ GeV	15053.2
$n_{jet} < 2$	12072.1
$m_{\ell\ell} > 11$ GeV	12020.2
Events with Opposite Sign leptons	11959.7
E_T^{miss} significance > 8	2101.0
$m_{T2}^{0} > 50$ GeV	1564.2
$n_{jet} = 0$	1026.5
Events with Same Flavour leptons	521.1
$m_{\ell\ell} < 76$ GeV or $m_{\ell\ell} > 106$ GeV	427.8
BDT-signal ¿ 0.77	58.4
BDT-other ¡ 0.01	35.5
Events with Different Flavour leptons	505.4
BDT-signal ¿ 0.81	100.1

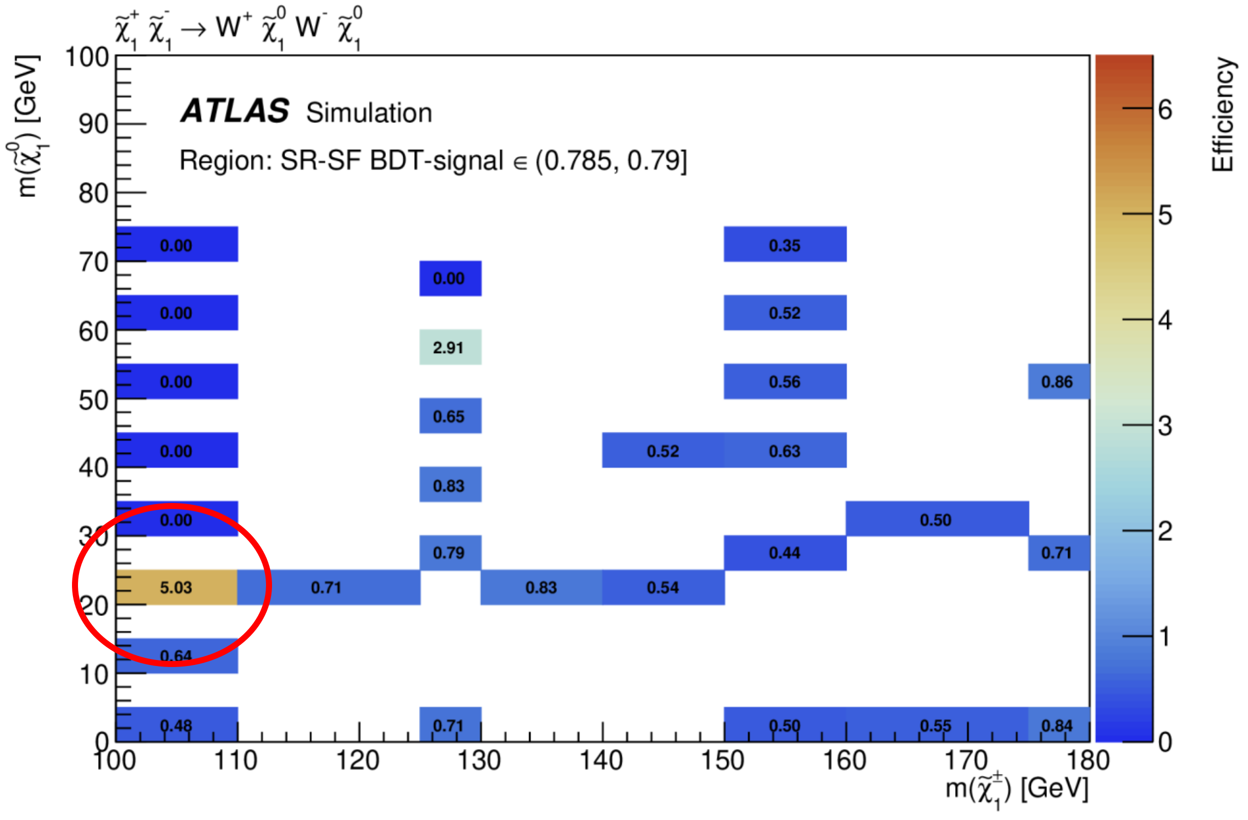
- Just one cutflow provided
  CM efficiency for BDT seems to be similar, but I suspect large sensitivity to MC details

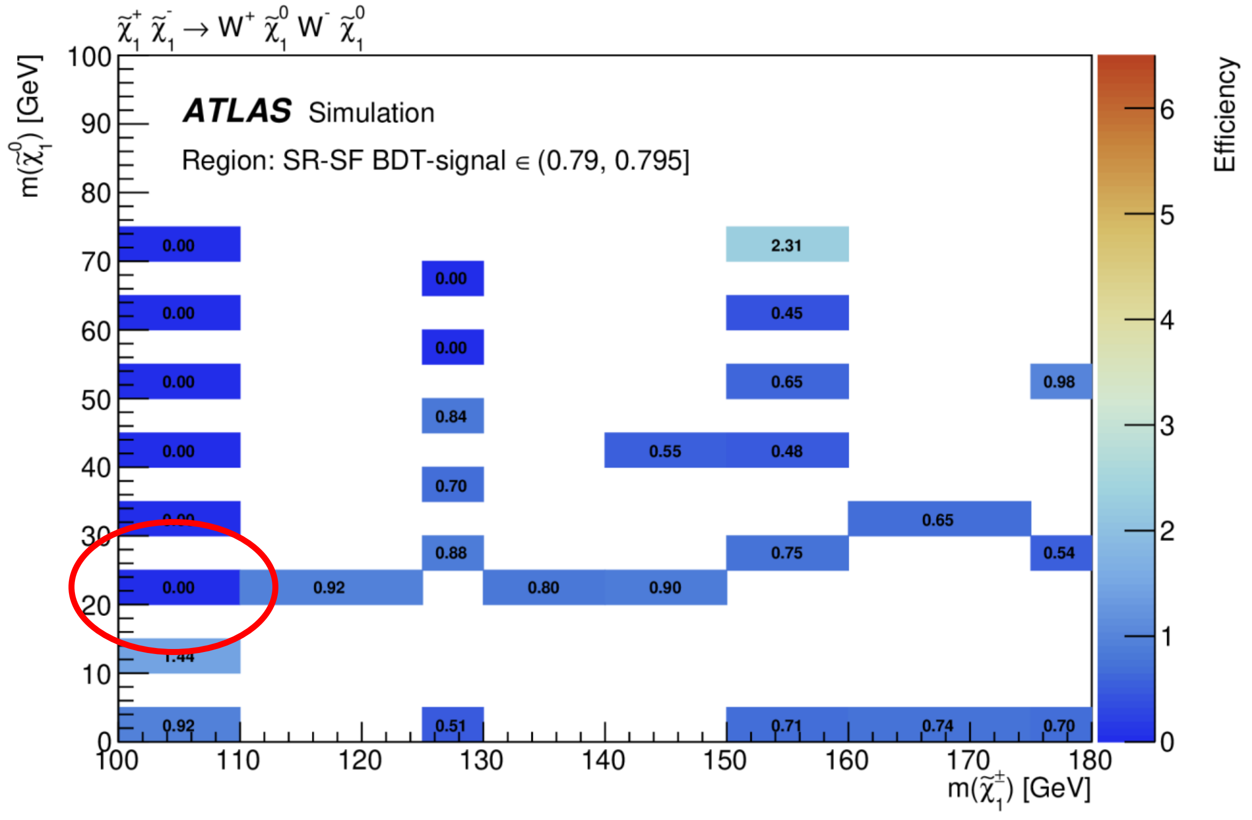
- Efficiency maps for chargino signal bin-by-bin in BDT score distribution (moderately useful)
- Again, BDT distribution for benchmark models is not provided



To be continued...

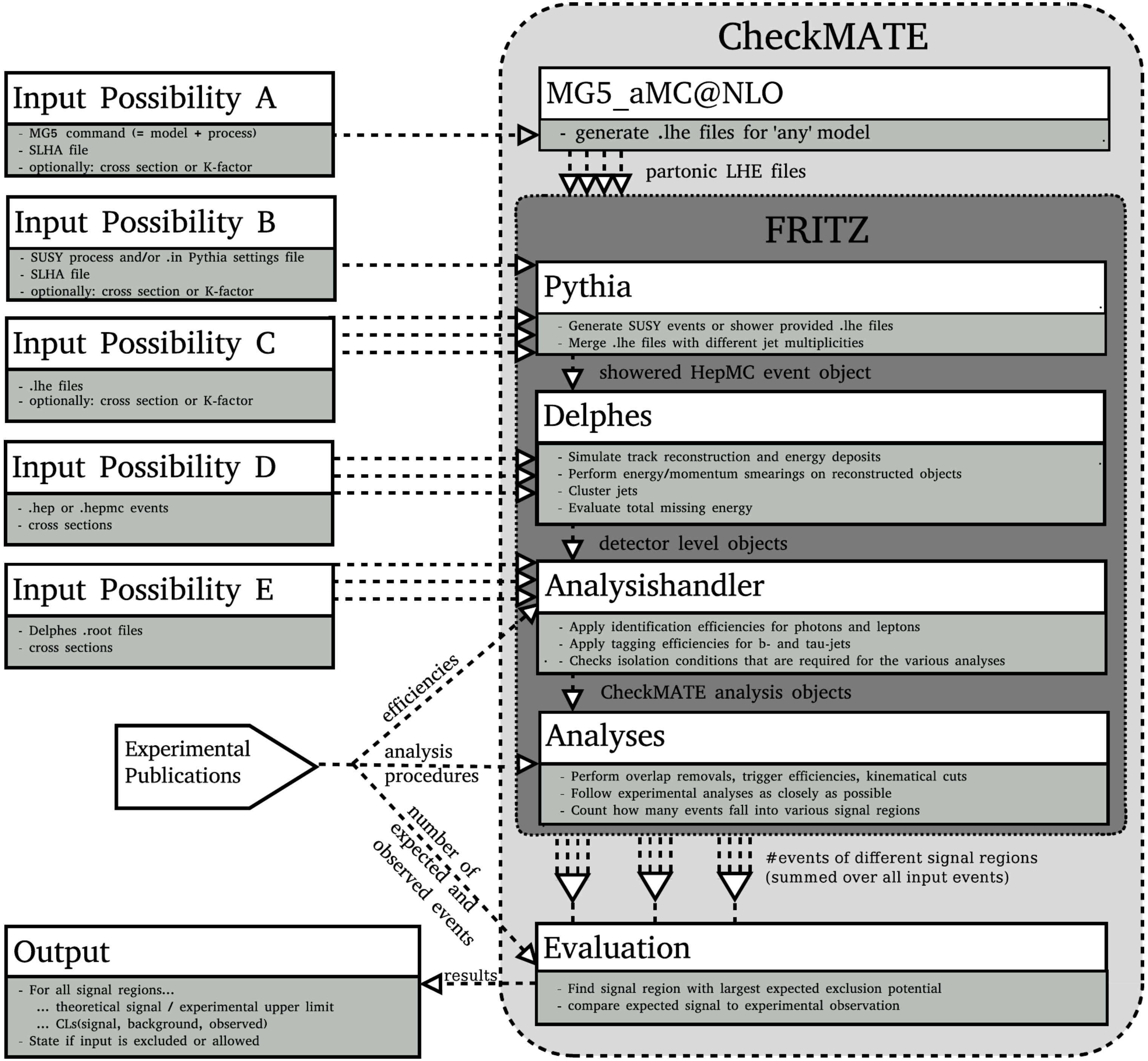
Some features in validations plots are difficult to understand





Conclusions

- Mixed experience with NN: one search is generally problematic the other one was pretty straightforward
- Analysis pseudocodes are invaluable for understanding details of inference
- Comparison with SimpleAnalysis would be super useful
- BDT implementation without problems, good agreement even at the detailed comparison for multijets
- Generally, histograms of BDT/NN scores for benchmark models would be strongly preferred for reasonable validation
- At the technical level we now have three different implementations...
- Will future searches be MVAUtils compatible?
- Outlook soon: reinterpretation studies of non-SUSY models – MUED for start


CheckMATE
MG5_aMC@NLO
- generate .lhe files for 'any' model
partonic LHE files
FRITZ
Pythia
- Generate SUSY events or shower provided .lhe files
- Merge .lhe files with different jet multiplicities
showered HepMC event object
Delphes
- Simulate track reconstruction and energy deposits
- Perform energy/momentum smearings on reconstructed objects
- Cluster jets
- Evaluate total missing energy
detector level objects
Analysishandler
- Apply identification efficiencies for photons and leptons
- Apply tagging efficiencies for b- and tau-jets
- Checks isolation conditions that are required for the various analyses
CheckMATE analysis objects
Analyses
- Perform overlap removals, trigger efficiencies, kinematical cuts
- Follow experimental analyses as closely as possible
- Count how many events fall into various signal regions
#events of different signal regions
(summed over all input events)
Evaluation
- Find signal region with largest expected exclusion potential
- compare expected signal to experimental observation
Input Possibility A
- MG5 command (= model + process)
- SLHA file
- optionally: cross section or K-factor
Input Possibility B
- SUSY process and/or .in Pythia settings file
- SLHA file
- optionally: cross section or K-factor
Input Possibility C
- .lhe files
- optionally: cross section or K-factor
Input Possibility D
- .hep or .hepmc events
- cross sections
Input Possibility E
- Delphes .root files
- cross sections
Experimental Publications
efficiencies
analysis procedures
number of expected and observed events
results
Output
- For all signal regions...
... theoretical signal / experimental upper limit
... CLs(signal, background, observed)
- State if input is excluded or allowed