



Cutting with

AEACUS,

Plotting with

RHADAMANTHUS,

& Learning with

MINOS

Joel W. Walker
Sam Houston State University



Computational Tools for High Energy Physics & Cosmology

IP2I Lyon

November 22-26, 2021

Students: Kebur Fantahun, B. Ash Fernando, Nicolle Schachtner, Trenton Voth, Jesse Cantu, & William Ellsworth

Sample plots from work also with: Dutta, Gao, Kumar, Li, Maxin, Nanopoulos, Sandick, Sinha, Stengel

“ Then spake Zeus: . . . ‘The cases are now indeed judged ill and it is because . . . many . . . who have wicked souls are clad in fair bodies and ancestry and wealth, and . . . the judges are confounded . . . , having their own soul muffled in the veil of eyes and ears and the whole body. . . . They must be stripped bare of all those things . . . , beholding with very soul the very soul of each immediately. . . . [I] have appointed sons of my own to be judges; two from Asia, **Minos** and **Rhadamanthus**, and one from Europe, **Aeacus**. These . . . shall give judgement in the meadow at the dividing of the road, whence are the two ways leading, one to the Isles of the Blest . . . , and the other to Tartaros.’ ”

– Plato, *Gorgias* (trans. Lamb)

Unified Work Flow

- ❖ **MadGraph (+ Others):** Matrix Element Generation
- ❖ **MadEvent (+ Others):** Hard Scattering Simulation
- ❖ **Pythia (+ Others):** Showering and Hadronization
- ❖ **DELPHES:** Detector Simulation
(DEtector Level PHysics Emulation System)
- ❖ **AEACUS:** Statistics Computation & Cut Selection
- ❖ **RHADAMANTHUS:** Graphical Event Analysis
- ❖ **MINOS:** Machine Learning on Features



Package Notes

- ❖ AEACUS, RHADAMANTHUS, & MINOS are Perl & Python
- ❖ All Perl scripts are self contained - no libraries or installation
- ❖ RHADAMANTHUS calls MatPlotLib & Numpy
- ❖ MINOS calls XGBoost, MPL & Numpy
- ❖ Control is provided by simple reusable card files

AEACUS

(Algorithm Event Arbiter and Cut Selector)

Guiding Principles:

- ❖ It is important to separate **WHAT** from **HOW**
- ❖ It is important to document **UNAMBIGUOUSLY**
- ❖ It is important to streamline **REPRODUCTION**

AEACUS (Goals)

- ❖ Automate model recast comparison against LHC data
- ❖ Replicate most current search strategies for new physics
- ❖ Embody lightweight, consumer-level, standalone design
- ❖ Decouple specific usage from general functionality
- ❖ Render event cut strategies compactly & unambiguously
- ❖ Merge power & flexibility with uniformity & simplicity
- ❖ Decouple phenomenology from software maintenance

AEACuS (Function)

- ❖ Converts Delphes Root file to a lightweight extended LHCO format
- ❖ ANY information in the Delphes Root File (including substructure variables and fat-jet information) can be read and processed by AEACuS
- ❖ Event-by-event weights are read and handled consistently throughout
- ❖ Filters kinematics, geometry, isolation, charge & flavor
- ❖ Dilepton pair assembly (by like / unlike charge & flavor)
- ❖ Jet (Re)clustering (KT, C / A, Anti-KT) & Hemispheres (Lund, etc.)
- ❖ Missing E_T , scalar H_T , effective & invariant mass, ratios & products
- ❖ Transverse mass, 1- & 2-step asymmetric M_{T2} (with combinatorics), Tri-jet mass, α_T , Razor & α_R , Dilepton Z-balance, Lepton W-projection, $\Delta\phi$ (& biased $\Delta\phi^*$), Shape Variables (thrust & minor, spheri[o]city, F), Girth, + MORE
- ❖ Arbitrary user-described combinations of computable statistics

Language Vs. Framework

AEACuS is BOTH and it is FACTORIZABLE

- ❖ The AEACuS meta language is an ideal mechanism for large experiments (CMS/ATLAS) & small phenomenology groups to unambiguously propagate an approximate rendering of internal event selection strategies
- ❖ The AEACuS software tool is an ideal agent for the rapid and uniform projection of sophisticated event cut workflows onto new physics models

Cut Card Example

```
# 1412.0618 MT2 Han/Liu
# 1409.7058 Baer, Mustafayev, Tata

*** Object Reconstruction ***

# Bound pseudo-rapidity magnitude and transverse momentum
OBJ_ELE = PRM:[0,2.5], PTM:7
OBJ_MUO = PRM:[0,2.5], PTM:7
OBJ_TAU = PRM:[0,2.5], CUT:[0,0] # Tau veto
OBJ_JET = PTM:20, PRM:[0,4.5]

OBJ_JET_001 = SRC:+000, PTM:30, CUT:[1,1] # Monojet
OBJ_JET_002 = SRC:+001, PTM:100, PRM:[0,2.5], CUT:1 # Jet is hard
OBJ_JET_003 = SRC:+000, HFT:1, PRM:[0,2.5], CUT:[0,0] # B-veto

# Find OSSF Dilepton with smallest mass
OBJ_LEP_001 = SRC:+000, SET:[DIL,-1,+1,0,UNDEF], CUT:2
# Report mass of that dilepton
OBJ_LEP_002 = SRC:+001, EFF:SUM, OUT:MAS_001
# Report p_T of leading lepton
OBJ_LEP_003 = SRC:+001, CUT:[1,UNDEF,-1], OUT:PTM_001
# Report p_T of sub-leading lepton
OBJ_LEP_004 = SRC:[+001,-003], OUT:PTM_002

*** Global Event Selection / Statistics Computation ***

# Cut on MET
EVT_MET = CUT:100
# Compute DiTau mass statistic
EVT_TTM_001 = LEP:001, JET:001, OUT:1
# Compute generalized MT2
EVT_ATM_001 = MET:000, MOD:[GEN,LEP_003,LEP_004,150,150], OUT:1
# Compute delta-phi angle between MET and the leptons
EVT_MDP_001 = MET:000, LEP:003, OUT:1
EVT_MDP_002 = MET:000, LEP:004, OUT:1
# Compute delta-R and delta-phi between the leptons
EVT_ODR_001 = LEP:001, OUT:1
EVT_ODP_001 = LEP:001, OUT:1
```

- Define hierarchical groupings of Jets & Leptons to set event topology w/ inclusion “+” and exclusion “-”
- Filter on sign, flavor, b-tags, etc.
- [Min,Max] brackets set bounds
- The “SET” command calls a variety of subroutines (e.g. dilepton) to extract a subset of input objects
- The “EFF” command is similar, but returns a transformed object, e.g. a vector sum or reclustered jets

Cut Card Example

```
# 1412.0618 MT2 Han/Liu
# 1409.7058 Baer, Mustafayev, Tata

*** Object Reconstruction ***

# Bound pseudo-rapidity magnitude and transverse momentum
OBJ_ELE = PRM:[0,2.5], PTM:7
OBJ_MUO = PRM:[0,2.5], PTM:7
OBJ_TAU = PRM:[0,2.5], CUT:[0,0] # Tau veto
OBJ_JET = PTM:20, PRM:[0,4.5]

OBJ_JET_001 = SRC:+000, PTM:30, CUT:[1,1] # Monojet
OBJ_JET_002 = SRC:+001, PTM:100, PRM:[0,2.5], CUT:1 # Jet is hard
OBJ_JET_003 = SRC:+000, HFT:1, PRM:[0,2.5], CUT:[0,0] # B-veto

# Find OSSF Dilepton with smallest mass
OBJ_LEP_001 = SRC:+000, SET:[DIL,-1,+1,0,UNDEF], CUT:2
# Report mass of that dilepton
OBJ_LEP_002 = SRC:+001, EFF:SUM, OUT:MAS_001
# Report p_T of leading lepton
OBJ_LEP_003 = SRC:+001, CUT:[1,UNDEF,-1], OUT:PTM_001
# Report p_T of sub-leading lepton
OBJ_LEP_004 = SRC:[+001,-003], OUT:PTM_002

*** Global Event Selection / Statistics Computation ***

# Cut on MET
EVT_MET = CUT:100
# Compute DiTau mass statistic
EVT_TTM_001 = LEP:001, JET:001, OUT:1
# Compute generalized MT2
EVT_ATM_001 = MET:000, MOD:[GEN,LEP_003,LEP_004,150,150], OUT:1
# Compute delta-phi angle between MET and the leptons
EVT_MDP_001 = MET:000, LEP:003, OUT:1
EVT_MDP_002 = MET:000, LEP:004, OUT:1
# Compute delta-R and delta-phi between the leptons
EVT_ODR_001 = LEP:001, OUT:1
EVT_ODP_001 = LEP:001, OUT:1
```

- Compute statistics associated with referenced groups of kinematic objects, or with the event as a whole
- Computed statistics may be used downstream for channel sorting or plotting

Advanced Features

```
# CMS 1405.7570
# Electroweak SUSY with decays to l,W,Z,H
# With students Fantahun, Fernando, Schachtnner

*** Object Reconstruction ***
OBJ_ELE = PTM:10, PRM:[0.0,2.4]
OBJ_MUO = PTM:10, PRM:[0.0,2.4]
OBJ_TAU = PTM:20, PRM:[0.0,2.4]
OBJ_JET = PTM:30, PRM:[0.0,2.5]

OBJ_LEP_001 = SRC:+000, EMT:+3, CUT:[0,1] # zero or one tau
OBJ_LEP_002 = SRC:+000, CUT:[3,3] # exactly 3 of e, mu, tau
OBJ_LEP_003 = SRC:+002, PTM:20, CUT:1 # out of the 3 leptons, one >20 Gev

OBJ_JET_002 = SRC:+000, HFT:1, CUT:[0,0] # veto bjets

OBJ_LEP_004 = SRC:+002, EMT:-3, SET:[DIL,-1,0,50,UNDEF], CUT:0 # OSAF e/mu near 50 GeV
OBJ_LEP_005 = SRC:+004, EFF:SUM, OUT:MAS_001 # mass of the dilepton pair
OBJ_LEP_006 = SRC:[+002,-004], CUT:[1,UNDEF,-1] # remaining lepton

OBJ_LEP_007 = SRC:+000, EMT:-3, CUT:[1,UNDEF,-1] # harder of non-taus
OBJ_LEP_008 = SRC:[+001,+007], SET:[DIL,-1,0], CUT:0 # tau OSAF 1
OBJ_LEP_009 = SRC:+008, EFF:SUM, OUT:MAS_002 # mass of the dilepton pair

OBJ_LEP_010 = SRC:[+000,-007], EMT:-3, CUT:[1,UNDEF,-1] # softer of non-taus
OBJ_LEP_011 = SRC:[+001,+010], SET:[DIL,-1,0], CUT:0 # tau OSAF 2
OBJ_LEP_012 = SRC:+011, EFF:SUM, OUT:MAS_003 # mass of the dilepton pair

OBJ_LEP_013 = SRC:+002, SET:[DIL,-1,+1,91.2,UNDEF], CUT:0 # OSSF close to z

*** Global Event Selection ***
EVT_MET = CUT:50
# Transverse masses of unmerged lepton with MET
EVT_OTM_001 = LEP:006, MET:000, OUT:1
EVT_OTM_002 = LEP:010, MET:000, OUT:1
EVT_OTM_003 = LEP:007, MET:000, OUT:1

# Find the reconstructed M_LL
# closest to simulation of visible system for Z -> ditau
# 50 GeV for ditau -> e/mu or 60 GeV if one tau is hadronic
EVT_VAR_001 = KEY:{  
    IFE( LES( ABS( IFE(DEF($3),$3,$2) - 60 ), ABS($1-50)), IFE(DEF($3),$3,$2), $1 ),  
    MAS_001,MAS_002,MAS_003}, OUT:1
# Select corresponding transverse mass of MET + 3rd lepton system
EVT_VAR_002 = KEY:{  
    IFE( LES( ABS( IFE(DEF($3),$3,$2) - 60 ), ABS($1-50)), IFE(DEF($3),$6,$5), $4 ),  
    MAS_001,MAS_002,MAS_003,OTM_001,OTM_002,OTM_003}, OUT:1
```

- This example replicates a sophisticated CMS SUSY study for recasting
- LEP_004 holds the e/mu opposite sign / any flavor dilepton closest to 50 GeV
- LEP_007/010 combine a tau with either of the other e/mu
- In each case, the mass of the dilepton and the transverse mass (OTM) of the 3rd lepton with the MET is computed
- An OSSF dilepton closest to the Z is also reconstructed (13)

Advanced Features

```
# CMS 1405.7570
# Electroweak SUSY with decays to l,W,Z,H
# With students Fantahun, Fernando, Schachtner

*** Object Reconstruction ***
OBJ_ELE = PTM:10, PRM:[0.0,2.4]
OBJ_MUO = PTM:10, PRM:[0.0,2.4]
OBJ_TAU = PTM:20, PRM:[0.0,2.4]
OBJ_JET = PTM:30, PRM:[0.0,2.5]

OBJ_LEP_001 = SRC:+000, EMT:+3, CUT:[0,1] # zero or one tau
OBJ_LEP_002 = SRC:+000, CUT:[3,3] # exactly 3 of e, mu, tau
OBJ_LEP_003 = SRC:+002, PTM:20, CUT:1 # out of the 3 leptons, one >20 Gev

OBJ_JET_002 = SRC:+000, HFT:1, CUT:[0,0] # veto bjets

OBJ_LEP_004 = SRC:+002, EMT:-3, SET:[DIL,-1,0,50,UNDEF], CUT:0 # OSAF e/mu near 50 GeV
OBJ_LEP_005 = SRC:+004, EFF:SUM, OUT:MAS_001 # mass of the dilepton pair
OBJ_LEP_006 = SRC:[+002,-004], CUT:[1,UNDEF,-1] # remaining lepton

OBJ_LEP_007 = SRC:+000, EMT:-3, CUT:[1,UNDEF,-1] # harder of non-taus
OBJ_LEP_008 = SRC:[+001,+007], SET:[DIL,-1,0], CUT:0 # tau OSAF 1
OBJ_LEP_009 = SRC:+008, EFF:SUM, OUT:MAS_002 # mass of the dilepton pair

OBJ_LEP_010 = SRC:[+000,-007], EMT:-3, CUT:[1,UNDEF,-1] # softer of non-taus
OBJ_LEP_011 = SRC:[+001,+010], SET:[DIL,-1,0], CUT:0 # tau OSAF 2
OBJ_LEP_012 = SRC:+011, EFF:SUM, OUT:MAS_003 # mass of the dilepton pair

OBJ_LEP_013 = SRC:+002, SET:[DIL,-1,+1,91.2,UNDEF], CUT:0 # OSSF close to z

*** Global Event Selection ***
EVT_MET = CUT:50
# Transverse masses of unmerged lepton with MET
EVT_OTM_001 = LEP:006, MET:000, OUT:1
EVT_OTM_002 = LEP:010, MET:000, OUT:1
EVT_OTM_003 = LEP:007, MET:000, OUT:1

# Find the reconstructed M_LL
# closest to simulation of visible system for Z -> ditau
# 50 GeV for ditau -> e/mu or 60 GeV if one tau is hadronic
EVT_VAR_001 = KEY:{  
    IFE( LES( ABS( IFE(DEF($3),$3,$2) - 60 ), ABS($1-50)), IFE(DEF($3),$3,$2), $1 ),  
    MAS_001,MAS_002,MAS_003}, OUT:1
# Select corresponding transverse mass of MET + 3rd lepton system
EVT_VAR_002 = KEY:{  
    IFE( LES( ABS( IFE(DEF($3),$3,$2) - 60 ), ABS($1-50)), IFE(DEF($3),$6,$5), $4 ),  
    MAS_001,MAS_002,MAS_003,OTM_001,OTM_002,OTM_003}, OUT:1
```

- Search targets 3-lepton final states with mixed OS e/ μ and a hadronic tau
- Simulation: $Z \rightarrow \tau \tau$ visible mass ~ 50 GeV for e/ μ or ~ 60 GeV when a τ goes hadronic
- The τ is guaranteed to be OS with one of the e or μ
- A custom variable takes mass of the defined OS system closest to the sim. target
- The associated 3rd body transverse mass is stored too

Regions / Channels

```
*** Event Channel Filtering ***
CUT_ESC_001 = KEY:LEP_001, CUT:[0,0] # Tau Veto
CUT_ESC_002 = KEY:LEP_004, CUT:[2,2] # Force 2 OSAF elec/muon
CUT_ESC_004 = KEY:LEP_013, CUT:[2,2] # Force 2 OSSF

# missing energy
CUT_ESC_031 = KEY:MET_000, CUT:[50,100]
CUT_ESC_032 = KEY:MET_000, CUT:[100,150]
CUT_ESC_033 = KEY:MET_000, CUT:[150,200]
CUT_ESC_034 = KEY:MET_000, CUT:200

# invariant mass
CUT_ESC_511 = KEY:VAR_001, CUT:[0,100]
CUT_ESC_512 = KEY:VAR_001, CUT:100

# transverse mass
CUT_ESC_521 = KEY:VAR_002, CUT:[0,120]
CUT_ESC_522 = KEY:VAR_002, CUT:[120,160]
CUT_ESC_523 = KEY:VAR_002, CUT:160
```

```
# TABLE 13, 1405.7570
# Opposite sign mixed e/mu pair plus a hadronic tau
# from 0 to 100 GeV Invariant Mass
CUT_CHN_301 = ESC:[+511,+521,+031,-001,+002,-004]
CUT_CHN_302 = ESC:[+511,+521,+032,-001,+002,-004]
002,-004]
002,-004]
002,-004]
002,-004]
002,-004]
002,-004]
002,-004]
002,-004]
002,-004]
002,-004]
002,-004]
002,-004]
002,-004]
002,-004]
002,-004]
002,-004]
002,-004]
002,-004]
002,-004]
002,-004]
002,-004]
002,-004]
002,-004]
002,-004]
```

- Event selection cuts are “registered” for disjoint parameter regions
- For example, we define here a tau-veto, an OSAF and an OSSF sorting condition
- These cuts are NOT YET APPLIED, but only DEFINED

Regions / Channels

```
*** Event Channel Filtering ***
CUT_ESC_001 = KEY:LEP_001, CUT:[0,0] # Tau Veto
CUT_ESC_002 = KEY:LEP_004, CUT:[2,2] # Force 2 OSAF elec/muon
CUT_ESC_004 = KEY:LEP_013, CUT:[2,2] # Force 2 OSSF
```

- Many channels are defined very simply by subscribing to various cuts, without recomputation (fast)
- A minus sign inverts the cut
- here, we force a tau, and a MIXED (not SF) e/μ OS dilepton
- We then bin into channels on MET, invariant mass, and transverse mass

```
# TABLE 13, 1405.7570
# Opposite sign mixed e/mu pair plus a hadronic tau
# from 0 to 100 GeV Invariant Mass
CUT_CHN_301 = ESC:[+511,+521,+031,-001,+002,-004]
CUT_CHN_302 = ESC:[+511,+521,+032,-001,+002,-004]
CUT_CHN_303 = ESC:[+511,+521,+033,-001,+002,-004]
CUT_CHN_304 = ESC:[+511,+521,+034,-001,+002,-004]
CUT_CHN_311 = ESC:[+511,+522,+031,-001,+002,-004]
CUT_CHN_312 = ESC:[+511,+522,+032,-001,+002,-004]
CUT_CHN_313 = ESC:[+511,+522,+033,-001,+002,-004]
CUT_CHN_314 = ESC:[+511,+522,+034,-001,+002,-004]
CUT_CHN_321 = ESC:[+511,+523,+031,-001,+002,-004]
CUT_CHN_322 = ESC:[+511,+523,+032,-001,+002,-004]
CUT_CHN_323 = ESC:[+511,+523,+033,-001,+002,-004]
CUT_CHN_324 = ESC:[+511,+523,+034,-001,+002,-004]
# Greater than 100 GeV Invariant Mass
CUT_CHN_331 = ESC:[+512,+521,+031,-001,+002,-004]
CUT_CHN_332 = ESC:[+512,+521,+032,-001,+002,-004]
CUT_CHN_333 = ESC:[+512,+521,+033,-001,+002,-004]
CUT_CHN_334 = ESC:[+512,+521,+034,-001,+002,-004]
CUT_CHN_341 = ESC:[+512,+522,+031,-001,+002,-004]
CUT_CHN_342 = ESC:[+512,+522,+032,-001,+002,-004]
CUT_CHN_343 = ESC:[+512,+522,+033,-001,+002,-004]
CUT_CHN_344 = ESC:[+512,+522,+034,-001,+002,-004]
CUT_CHN_351 = ESC:[+512,+523,+031,-001,+002,-004]
CUT_CHN_352 = ESC:[+512,+523,+032,-001,+002,-004]
CUT_CHN_353 = ESC:[+512,+523,+033,-001,+002,-004]
CUT_CHN_354 = ESC:[+512,+523,+034,-001,+002,-004]
```

AEACUS Output

NNN	KEY_NNN	EPW_NNN	ENW_NNN	EZW_NNN	XPB_NNN	ABS_NNN	ERR_NNN	IPB_NNN	EFF_NNN	PRD_NNN								
000	CHN_000	124459	0	0	+6.40526E+02	+6.40526E+02	+1.60155E-01	+1.94307E+02	1.0000000000	1.0000000000								
001	TAU_000	106652	0	0	+5.48883E+02	+5.48883E+02	+1.48256E-01	+1.94307E+02	0.8569247704	0.8569247704								
002	OBJ_001	355	0	0	+1.82700E+00	+1.82700E+00	+8.55348E-03	+1.94307E+02	0.0033285827	0.0028523449								
003	JET_002	316	0	0	+1.62629E+00	+1.62629E+00	+8.06998E-03	+1.94307E+02	0.8901408451	0.0025389887								
004	JET_006	54	0	0	+2.77910E-01	+2.77910E-01	+3.33600E-03	+1.94307E+02	0.1708860759	0.0004338778								
005	MET_000	48	0	0	+2.47031E-01	+2.47031E-01	+3.14521E-03	+1.94307E+02	0.8888888889	0.0003856692								
006	ESC_001	24	0	0	+1.23504E-01	+1.23504E-01	+2.22389E-03	+1.94326E+02	0.4999534471	0.0001928165								
OBJ_001 LEP_001 JET_001																		
LEP_001 0.00000 0.00000																		
JET_001 0.00000 0.00000																		
EID_000	TAU_000	LEP_001	JET_001	JET_002	JET_006	ETA_001	PTM_001	MAS_001	ETA_002	PTM_002	ETA_003	PTM_003	MAS_003	ETA_004	PTM_004	CAL_000	MET_000	MHT_000
MEF_000	MT2_001	TTM_001	ODP_001	MDP_001	CTS_001	MT2_002	ODP_002	MDP_002	ODP_003	MDP_003	ODP_004	MDP_004	ODP_005	ODP_006	VAR_001	VAR_002	VAR_003	
VAR_004	VAR_005	VAR_006	VAR_011	VAR_012	VAR_013	VAR_014	VAR_021	WGT_000										
0000898	0	2	1	0	0	-0.029	28.3	46.5	-1.027	17.5	0.631	83.7	8.3	UNDEF	UNDEF	62.0	81.9	154.1
236.0	108.1	-75.4	2.261	0.112	0.461	5.1	2.971	2.373	UNDEF	3.083	0.710	UNDEF	UNDEF	UNDEF	+7.608E-01	+5.784E-01	+2.899E-02	
+9.299E-01	+7.727E-01	+5.587E-01	+3.459E-01	+2.140E-01	+1.023E+00	+0.000E+00	+1.594E+00	+5.146E-03										
0003430	0	2	1	0	0	-1.221	100.8	206.5	-0.308	90.4	-0.089	53.4	5.8	UNDEF	UNDEF	96.2	120.6	269.0
389.7	133.2	242.0	2.676	1.948	0.427	72.5	1.083	1.659	UNDEF	3.031	1.593	UNDEF	UNDEF	UNDEF	+7.226E-01	+8.117E-01	+8.399E-01	
+2.156E-01	+2.986E-01	+8.877E-02	+8.354E-01	+7.496E-01	+4.428E-01	+0.000E+00	+4.587E-01	+5.146E-03										
0004304	0	2	1	0	0	-0.521	108.2	190.9	0.475	73.8	1.873	41.5	6.1	UNDEF	UNDEF	18.7	31.1	223.5
254.7	131.7	109.3	2.409	2.400	0.461	41.9	2.393	1.475	UNDEF	0.007	1.481	UNDEF	UNDEF	UNDEF	+7.599E-01	+9.835E-01	+4.785E-01	
+8.849E-01	+4.422E-01	+9.539E-01	+3.475E+00	+2.368E+00	+1.334E+00	+0.000E+00	+7.566E-01	+5.146E-03										
0027442	0	2	1	0	0	-2.246	55.2	29.9	-1.594	28.1	-1.994	41.0	7.7	UNDEF	UNDEF	82.7	88.9	148.7
237.6	144.5	-83.1	0.370	2.237	0.315	76.2	1.940	2.607	UNDEF	2.106	1.570	UNDEF	UNDEF	UNDEF	+5.730E-01	+2.468E-01	+9.779E-01	
+3.799E-01	+9.208E-01	+9.636E-01	+6.214E-01	+3.157E-01	+4.615E-01	+0.000E+00	+5.839E-01	+5.146E-03										
0027631	0	2	1	0	0	0.596	97.8	105.7	-0.860	35.8	-0.858	46.3	8.2	UNDEF	UNDEF	47.3	61.8	312.7
374.5	153.3	-100.9	0.844	2.019	0.622	74.1	1.206	2.863	UNDEF	0.813	2.050	UNDEF	UNDEF	UNDEF	+8.969E-01	+8.965E-01	+5.342E-01	
+2.000E-03	+6.963E-01	+6.952E-01	+1.583E+00	+5.794E-01	+7.500E-01	+0.000E+00	+7.193E-01	+5.146E-03										
0030072	0	2	1	0	0	1.802	31.4	55.9	1.888	24.9	1.267	40.0	7.3	UNDEF	UNDEF	59.3	64.4	119.9
184.3	101.6	276.5	3.008	2.502	0.043	0.4	0.550	0.506	UNDEF	3.052	2.725	UNDEF	UNDEF	UNDEF	+8.579E-02	+4.892E-01	+9.470E-01	
+5.518E-01	+9.552E-01	+8.530E-01	+4.880E-01	+3.873E-01	+6.215E-01	+0.000E+00	+4.038E+00	+5.146E-03										

- ❖ Output is a set of tables reporting requested statistics & cut fractions
- ❖ It is often convenient to make no cuts at the lowest level, but only to compute
- ❖ Names such as “JET_001” have no invariant meaning - they are defined in a card_file

RHADAMANTHUS

(Recursively Heuristic Analysis, Display, And MANipulation:
The Histogram Utility Suite)

- ❖ Heuristic *adjective* \hyü-'ris-tik\ (www.merriam-webster.com)
: using experience to learn and improve :
involving or serving as an aid to learning, discovery, or problem-solving by experimental and especially trial-and-error methods <*heuristic techniques*> <*a heuristic assumption*>; also :
of or relating to exploratory problem-solving techniques that utilize self-educating techniques (as the evaluation of feedback) to improve performance <*a heuristic computer program*>

Plot Card Example

```
PLT_DAT_001 = DIR:"./M3/0b_4l", FIL:"BG:MEG:TTBAR*"
PLT_DAT_002 = DIR:"./M3/0b_4l", FIL:[ "BG:MEG:VVJJ*", "BG:MEG:ZJJJJ*", "BG:MEG:WJJJJ*" ]
PLT_DAT_003 = DIR:"./M3/0b_4l", FIL:"NMSSM:A:NMSSM*"

PLT_CHN_001 = DAT:[001,002,003], KEY:MET_000

PLT_HST_001 =
  IFB:300,
  CHN:001,
  LFT:0, RGT:1000, SPN:25,
  MIN:0.001, MAX:UNDEF,
  SUM:-1, NRM:0, AVG:3,
  LOG:1, LOC:0, CLR:0,
  TTL:"$4^+e/\mu$ with $0^+$ B-Jets, <RTS> = 14 TeV, <LUM> = 300 <IFB>",
  LBL:[ "<MET> Cut Threshold [GeV]", "Integrated Event Count" ],
  LGD:[
    "$t\overline{t} + 0-2\text{ Jets}",
    "$V,V + 0-2\text{ Jets} \& Z/W + 0-4\text{ Jets}",
    "NMSSM-A $\chi^0 \chi^0 + 0-2\text{ Jets}" ],
  OUT:"../Plots", NAM:"event_count_MET_0b_4l_300", FMT:"PDF"
```

Plot Card Example

```
PLT_DAT_001 = DIR:"./M3/0b_41", FIL:"BG:MEG:TTBAR*"
PLT_DAT_002 = DIR:"./M3/0b_41", FIL:[ "BG:MEG:VVJJ*", "BG:MEG:ZJJJJ*", "BG:MEG:WJJJJ*" ]
PLT_DAT_003 = DIR:"./M3/0b_41", FIL:"NMSSM:A:NMSSM*"
```

- Data Sets are built out of groups of “.cut” files from AEACuS
- Wildcards “*” are allowed to match multiple files
- Cross-sections are imported automatically
- Files with common trailing digits (name_NNN.cut) are averaged
- Files with unique names are summed

```
"$t\overline{t}+$ 0-2 Jets",
"$V\!, V+$ 0-2 Jets & $Z/W+$ 0-4 Jets",
"NMSSM-A $\chi^0 \chi^0+$ 0-2 Jets" ],
OUT:"./Plots", NAM:"event_count_MET_0b_41_300", FMT:"PDF"
```

Plot Card Example

```
PLT_DAT_001 = DIR:"./M3/0b_4l", FIL:"BG:MEG:TTBAR*"
PLT_DAT_002 = DIR:"./M3/0b_4l", FIL:[ "BG:MEG:VVJJ*", "BG:MEG:ZJJJJ*", "BG:MEG:WJJJJ*" ]
PLT_DAT_003 = DIR:"./M3/0b_4l", FIL:"NMSSM:A:NMSSM*"

PLT_CHN_001 = DAT:[001,002,003], KEY:MET_000
```

- Channels are built out of groups of datasets
- The plotting key refers to a statistic computed by AEACuS

```
SUM:-1, NRM:0, AVG:3,
LOG:1, LOC:0, CLR:0,
TTL:"$4^+e/\mu$ with $0^+$ B-Jets, <RTS> = 14 TeV, <LUM> = 300 <IFB>",
LBL:[ "<MET> Cut Threshold [GeV]", "Integrated Event Count" ],
LGD:[
    "$t\overline{t} + 0-2\text{ Jets}",
    "$V,V + 0-2\text{ Jets} \& Z/W + 0-4\text{ Jets}",
    "NMSSM-A $\chi^0 \chi^0 + 0-2\text{ Jets}" ],
OUT:". /Plots", NAM:"event_count_MET_0b_4l_300", FMT:"PDF"
```

Plot Card Example

- Histograms are built out of groups of channels
- Line continuation is indicated simply by indentation
- The luminosity may be specified in “IPB”, “IFB”, “IAB”, etc.

```
PLT_HST_001 =
  IFB:300,
  CHN:001,
  LFT:0, RGT:1000, SPN:25,
  MIN:0.001, MAX:UNDEF,
  SUM:-1, NRM:0, AVG:3.
```

- By default, events are oversampled and scaled down to the target luminosity
- There is a warning on scale factors < 1
- Optionally specify trim at exact luminosity “IFB:[300,-1]”
- Bins are specified by “LFT” = left, “RGT” = right, “SPN” = bin span
- Optionally “BNS” = number of bins may be used instead of one prior
- “MIN” and “MAX” provide optional manual limits on range

Plot Card Example

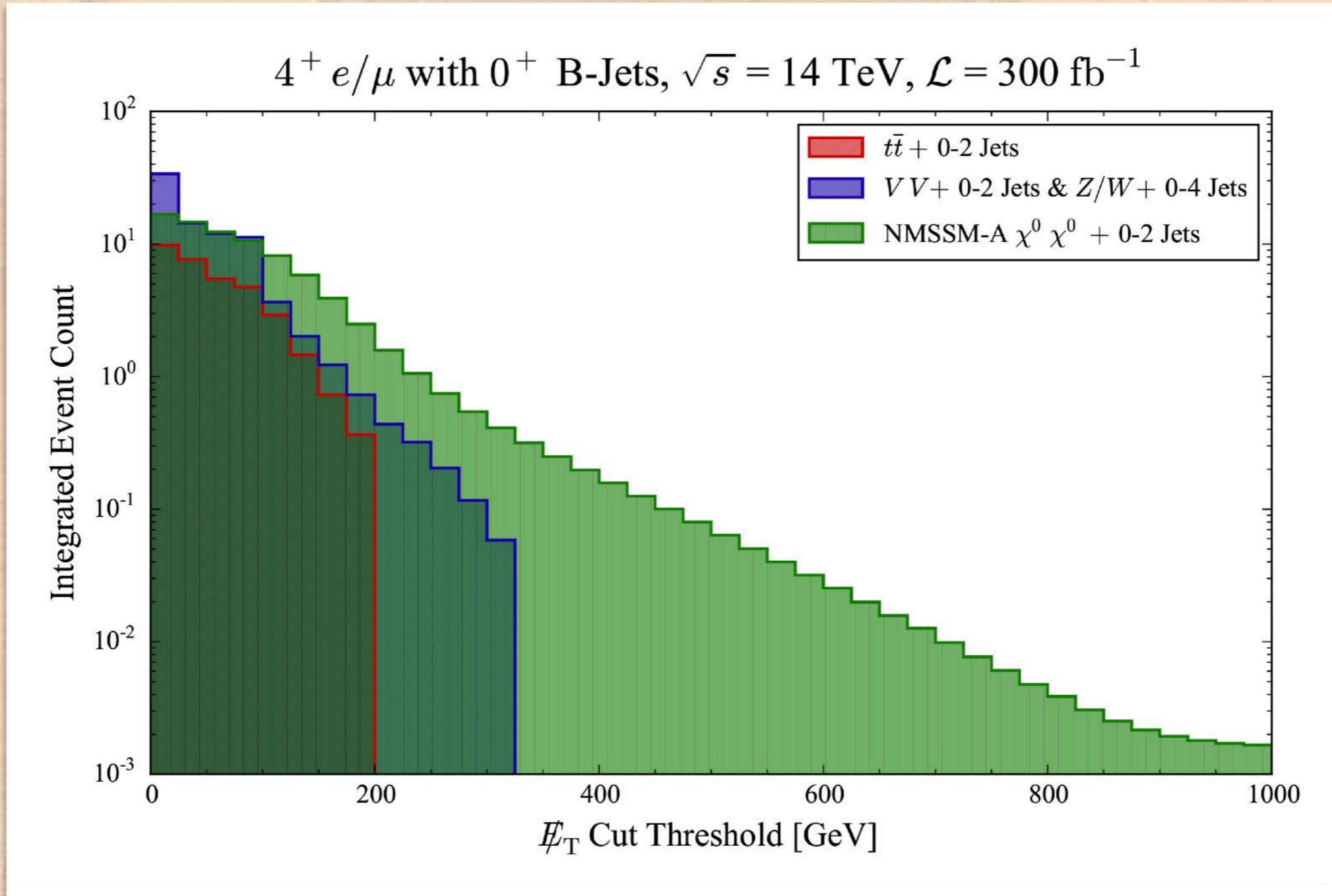
```
PLT_DAT_001 = DIR:"./M3/0b_41", FIL:"BG:MEG:TTBAR*"
```

- SUM +/- 1 compound bin counts to the right/left for threshold plots
- NRM facilitates normalization as for shape plots
- AVG engages bin smoothing with preservation of integrated counts
- LOG = 1/0 enables/disables logarithmic dependent axis

```
SUM:-1, NRM:0, AVG:3,  
LOG:1, LOC:0, CLR:0,  
TTL:"$4^+e/\mu$ with $0^+$ B-Jets, <RTS> = 14 TeV, <LUM> = 300 <IFB>",  
LBL:[ "<MET> Cut Threshold [GeV]", "Integrated Event Count" ],  
LGD:[  
    "$t\overline{t} + 0-2\text{ Jets}",  
    "$V,V + 0-2\text{ Jets} \& Z/W + 0-4\text{ Jets}",  
    "NMSSM-A $\chi^0 \chi^0 + 0-2\text{ Jets}" ],  
OUT:"./Plots", NAM:"event_count_MET_0b_41_300", FMT:"PDF"
```

- Inline LaTeX is used to input formulas for title, axis labels, and legends
- Several preconfigured notations are accessible via shorthand
- Available vector output formats include publication quality “EPS” & “PDF”
- Optionally specify intermediate Python source output “FMT:[PDF,1]”

Sample Plot Output



Optimize By Shape

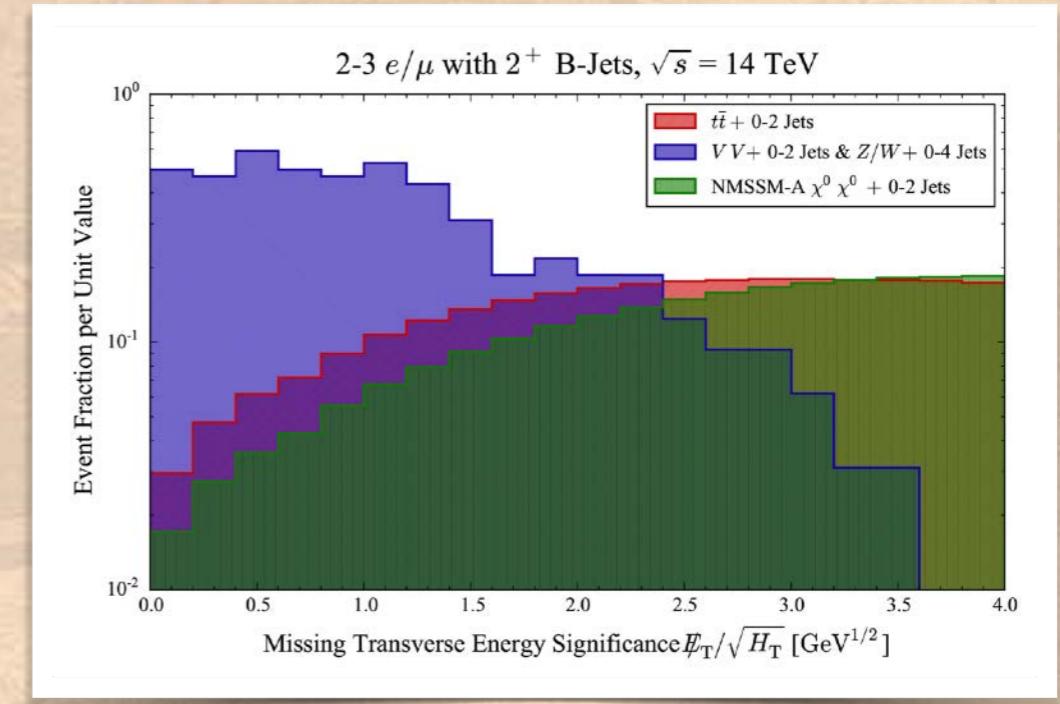
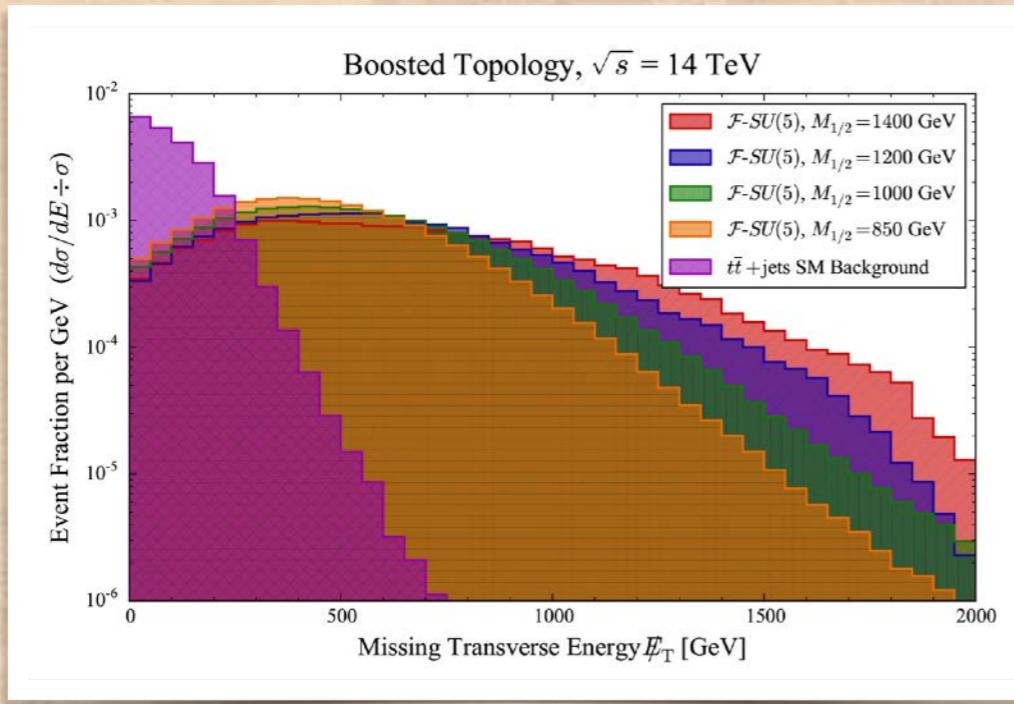
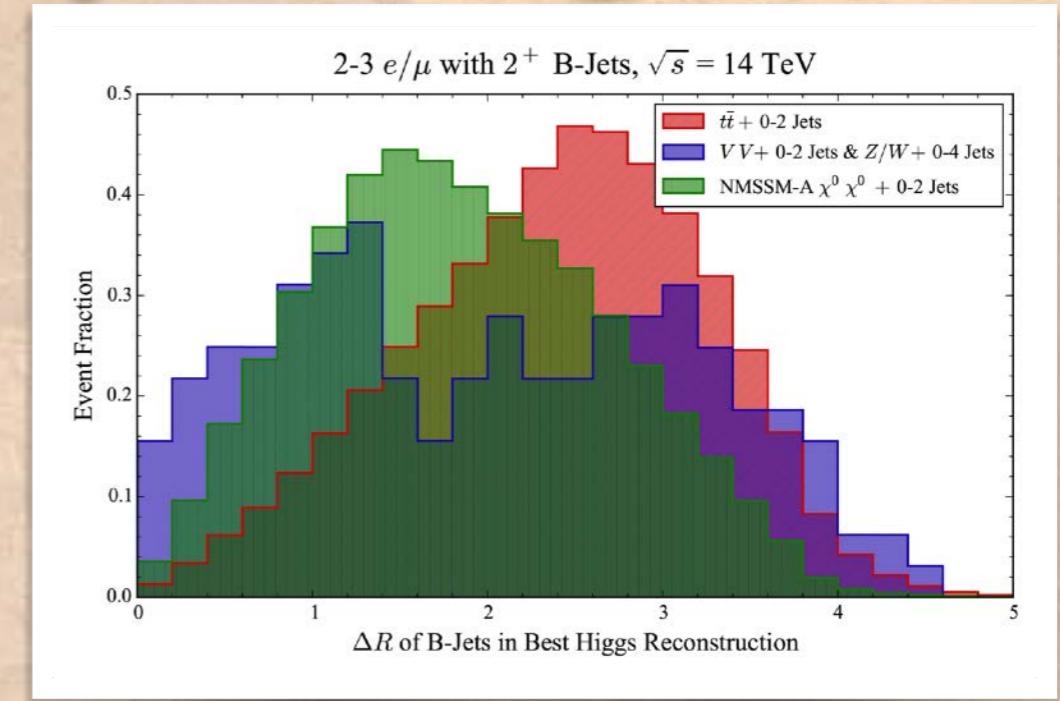
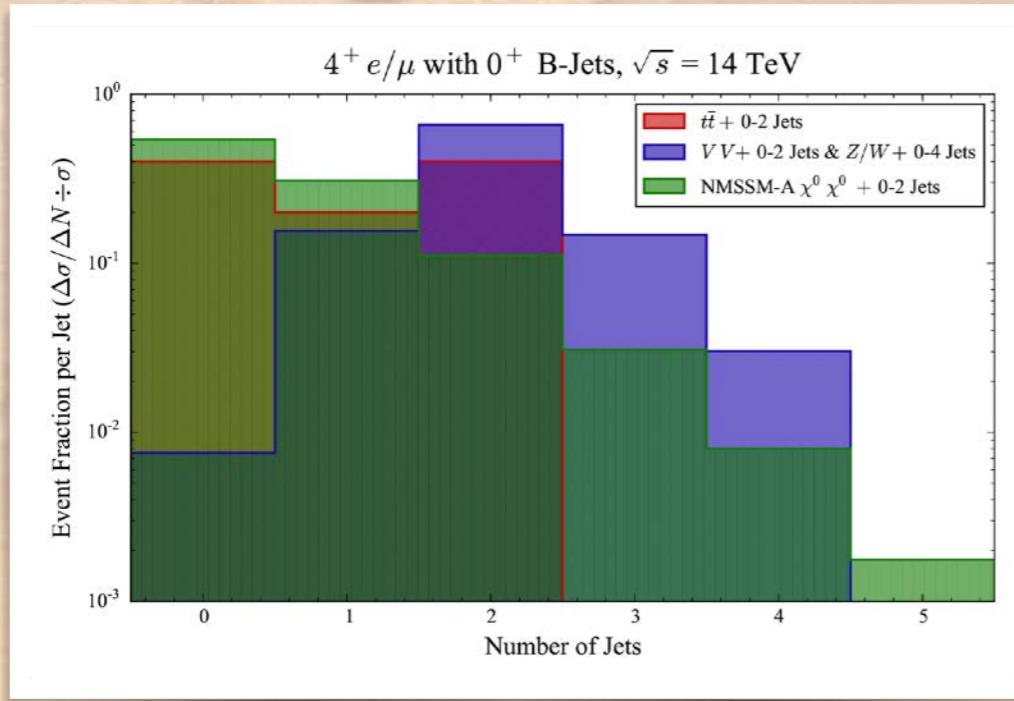
```
PLT_DAT_001 = DIR:"./Cuts", FIL:"Forward:BG:MEG:TTBAR_*"
PLT_DAT_002 = DIR:"./Cuts", FIL:"Forward:FSU5_VBF_25:850_*"
PLT_DAT_003 = DIR:"./Cuts", FIL:"Forward:FSU5_VBF_25:1000_*"
PLT_DAT_004 = DIR:"./Cuts", FIL:"Forward:FSU5_VBF_25:1200_*"
PLT_DAT_005 = DIR:"./Cuts", TTL:"Forward:FSU5_VBF_25:1400_*"

PLT_CHN
PLT_HST
```

- Shape plots are unit normalized
- They identify HOW to cut, e.g.
threshold min/max vs. window

```
LFT:0, RGT:2000, SPN:50,
MIN:0.000001, MAX:UNDEF,
SUM:0, NRM:1, AVG:3,
LOG:1, LOC:0, CLR:0,
TTL:"Boosted Topology, <RTS> = 14 TeV",
LBL:[ "Missing Transverse Energy <MET> [GeV]",
      "Event Fraction per Gev (<DEF>)" ],
LGD:[ "$\mathcal{F}$$\text{-SU(5)}$, $M_{1/2} = 1400$ GeV",
       "$\mathcal{F}$$\text{-SU(5)}$, $M_{1/2} = 1200$ GeV",
       "$\mathcal{F}$$\text{-SU(5)}$, $M_{1/2} = 1000$ GeV",
       "$\mathcal{F}$$\text{-SU(5)}$, $M_{1/2} = 850$ GeV",
       "$t\overline{t}+\text{jets}$ SM Background" ],
OUT:"./Plots", NAM:"met_shape_boosted_30", FMT:"PDF"
```

Optimize By Shape



Apply Selection Cuts

```
PLT_DAT_001 = DIR:"./M3/0b_41", FIL:"BG:MEG:TTBAR*"
PLT_DAT_002 = DIR:"./M3/0b_41", FIL:["BG:MEG:VVJJ*", "BG:MEG:ZJJJJ*", "BG:MEG:WJJJJ*"]
PLT_DAT_003 = DIR:"./M3/0b_41", FIL:"NMSSM:A:NMSSM*"

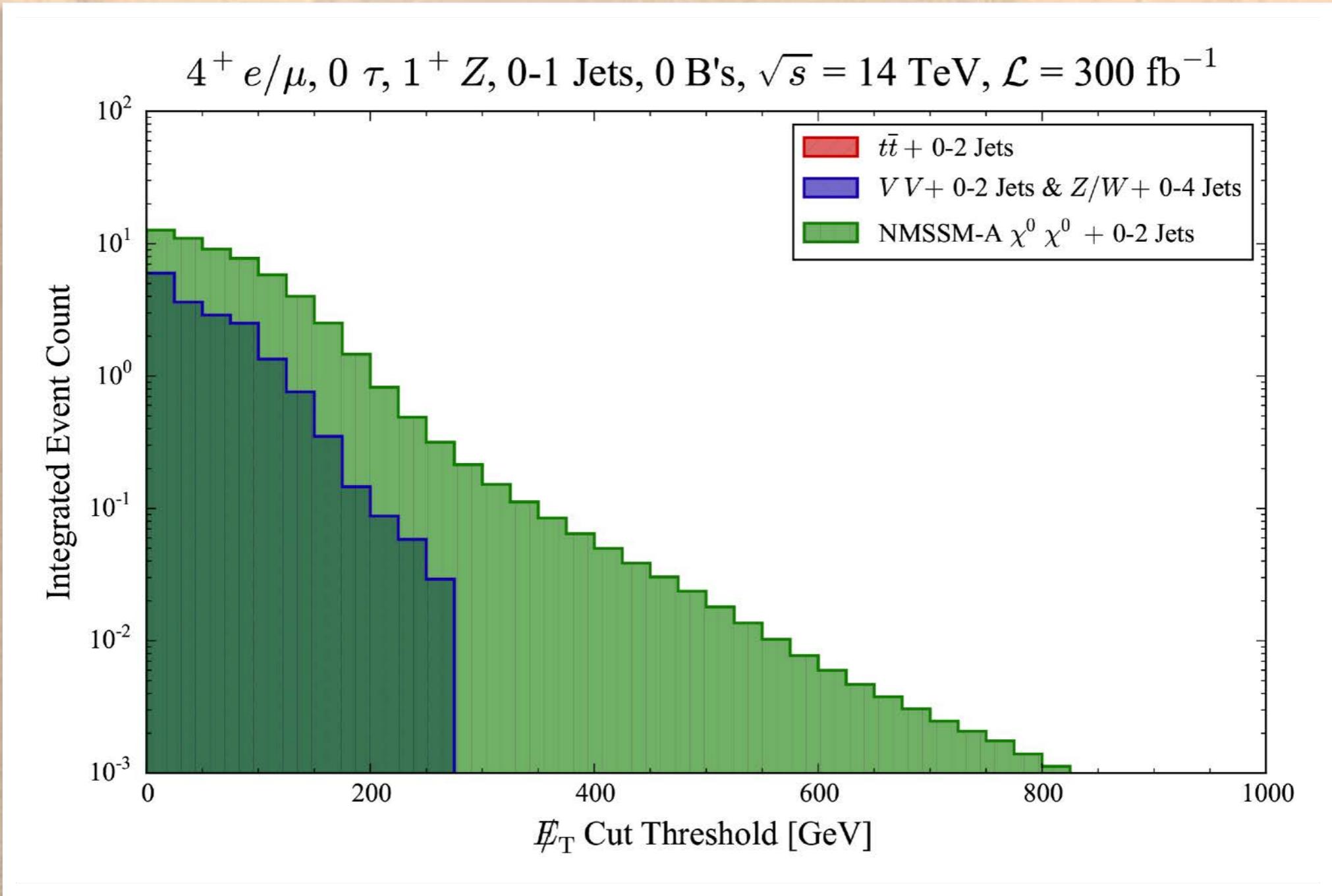
PLT_ESC_001 = KEY:LEP_002, CUT:[0,0] # Veto Taus
PLT_ESC_002 = KEY:LEP_005, CUT:1    # Force 1 Lepton pair in Z Window
PLT_ESC_003 = KEY:JET_000, CUT:[0,1] # Veto 2+ Jets
PLT_ESC_004 = KEY:JET_003, CUT:[0,0] # Veto B's

PLT_CHN_003 = DAT:[001,002,003], KEY:MET_000, ESC:[+001,+002,+003,+004]
```

- Event Selection Cuts (ESC) are registered by AEACus key and range
- Channels may subscribe to any number of registered cuts

```
SUM:-1, NRM:0, AVG:3,
LOG:1, LOC:0, CLR:0,
TTL:"$4^+e/\mu$, $0\tau$, $1^+Z$, 0-1 Jets, 0 B's, <RTS> = 14 TeV, <LUM> = 300 <IFB>",
LBL:[ "<MET> Cut Threshold [GeV]", "Integrated Event Count" ],
LGD:[ "$t\overline{t}$ 0-2 Jets",
       "$V,V+$ 0-2 Jets & $Z/W+$ 0-4 Jets",
       "NMSSM-A $\chi^0 \chi^0+$ 0-2 Jets" ],
OUT:".Plots", NAM:"event_count_MET_OPT_0b_41_300", FMT:"PDF"
```

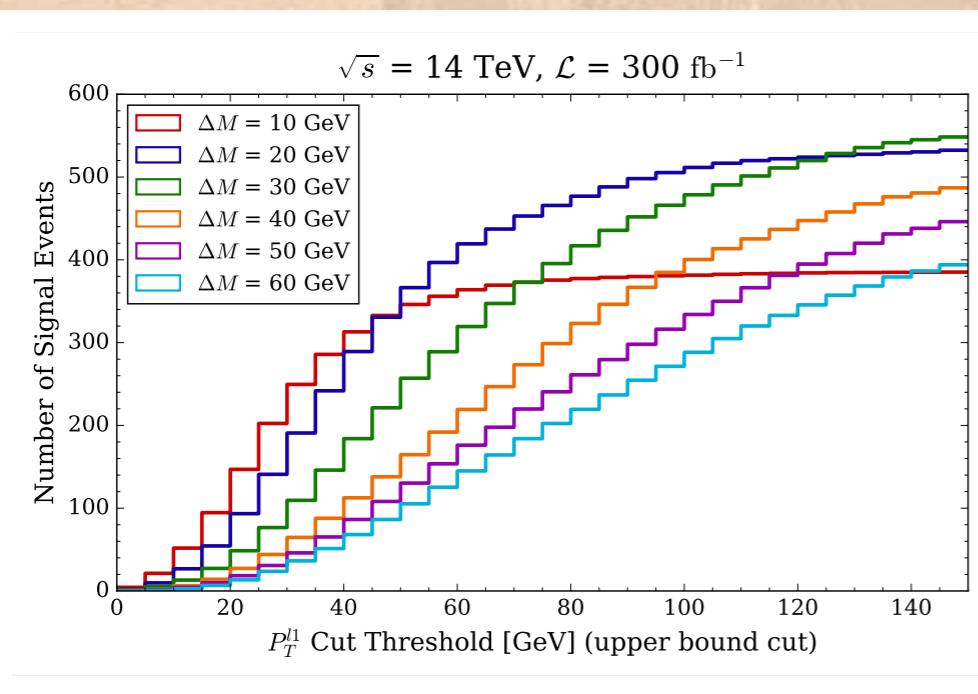
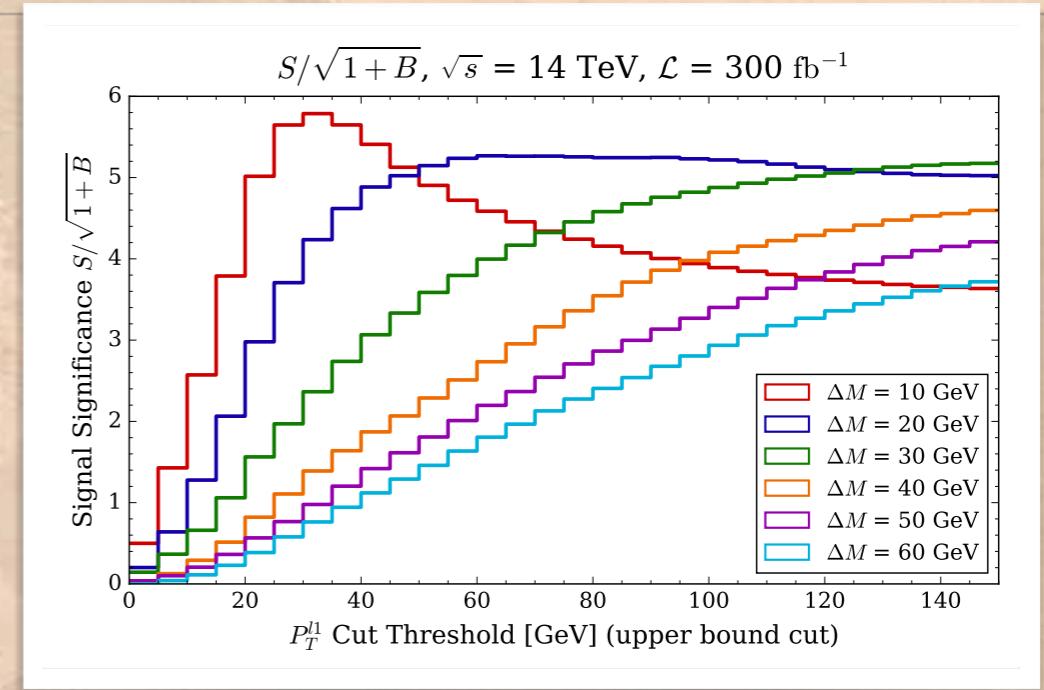
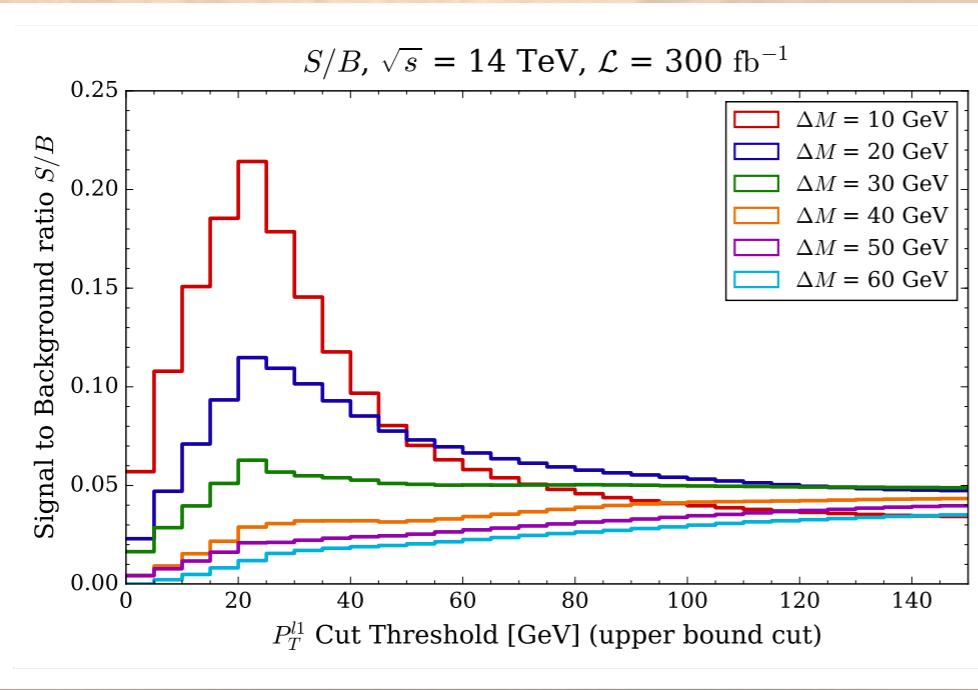
Optimized Plot Output



Transform Bin Channels

- ❖ User-defined functions of binned channels are allowed for specification of the dependent plotting variable
- ❖ Internal histogram object transparently applies the specified functional transformation bin-by-bin
- ❖ Channels with multiple data sets iterate automatically
- ❖ Single data sets expand to match large dimensionalities

Transform Bin Channels



- ❖ This is useful for taking arbitrary functions of merged channels, e.g. $S/1+B$, $S/\sqrt{1+B}$
- ❖ Useful for answering the question “WHERE to cut?”

Transform Bin Channels

```
PLT_DAT_001 = DIR:"./Cuts_LSD", FIL:"Jets:BG:MEG:TTBAR_"
PLT_DAT_002 = DIR:"./Cuts_LSD", FIL:"Jets:FSU5_VBF_25:850_"
PLT_DAT_003 = DIR:"./Cuts_LSD", FIL:"Jets:FSU5_VBF_25:1000_"
```

- Signal significance is computed here by combining Signal & BG
- Signal and BG use same key and subscribe to identical event selection cuts
- The single BG Channel is expanded to match four Signal Channels

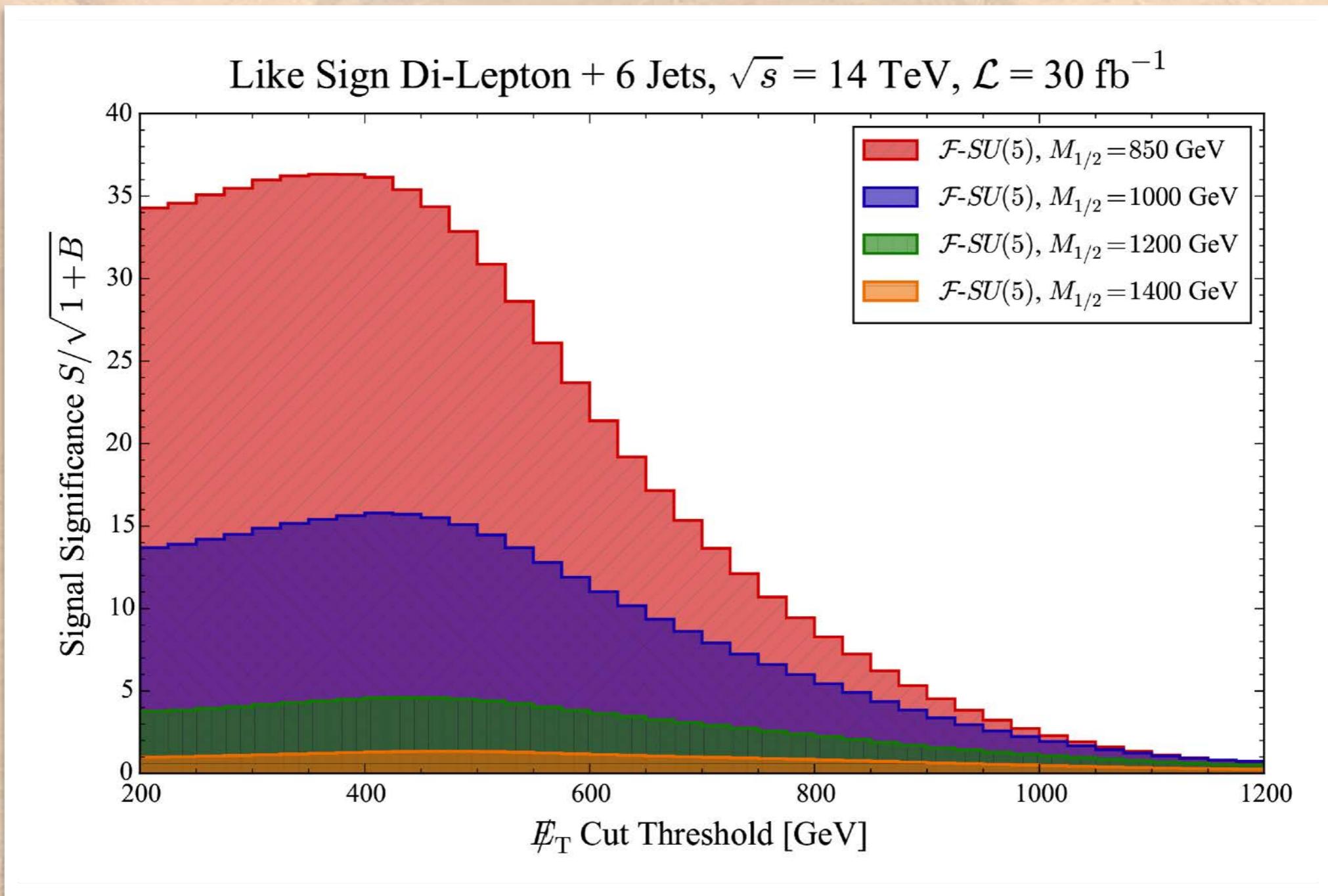
```
# One-dimensional background channel
PLT_CHN_001 = DAT:[001], KEY:MET_000, ESC:[+001,+002,+003,+004,+005]
# Four-dimensional signal channel
PLT_CHN_002 = DAT:[002,003,004,005], KEY:MET_000, ESC:[+001,+002,+003,+004,+005]

PLT_HST_002 =
    TFB:30
    CHN:{$2/SRT(1+$1),001,002},
    LFT:200, RGT:1200, SPN:25, BNS:UNDEF,
    MIN:0.0 MAX:UNDEF,
    SUM:-1, NRM:0, PER:UNDEF, AVG:3,
    LOG:0 CTX:0 LOG:0 CTB:0
```

- For a lower bound threshold plot, integrate “SUM” from the left “-1”

```
"signal significance <SIB>" ],
LGD:[
    "$\mathcal{F}$$-\mathrm{SU}(5)$, $M_{1/2} = 850$ GeV",
    "$\mathcal{F}$$-\mathrm{SU}(5)$, $M_{1/2} = 1000$ GeV",
    "$\mathcal{F}$$-\mathrm{SU}(5)$, $M_{1/2} = 1200$ GeV",
    "$\mathcal{F}$$-\mathrm{SU}(5)$, $M_{1/2} = 1400$ GeV"],
OUT:"./Plots", NAM:"met_sig_LSD_30", FMT:"PDF"
```

Transform Bin Channels



Transform Bin Channels

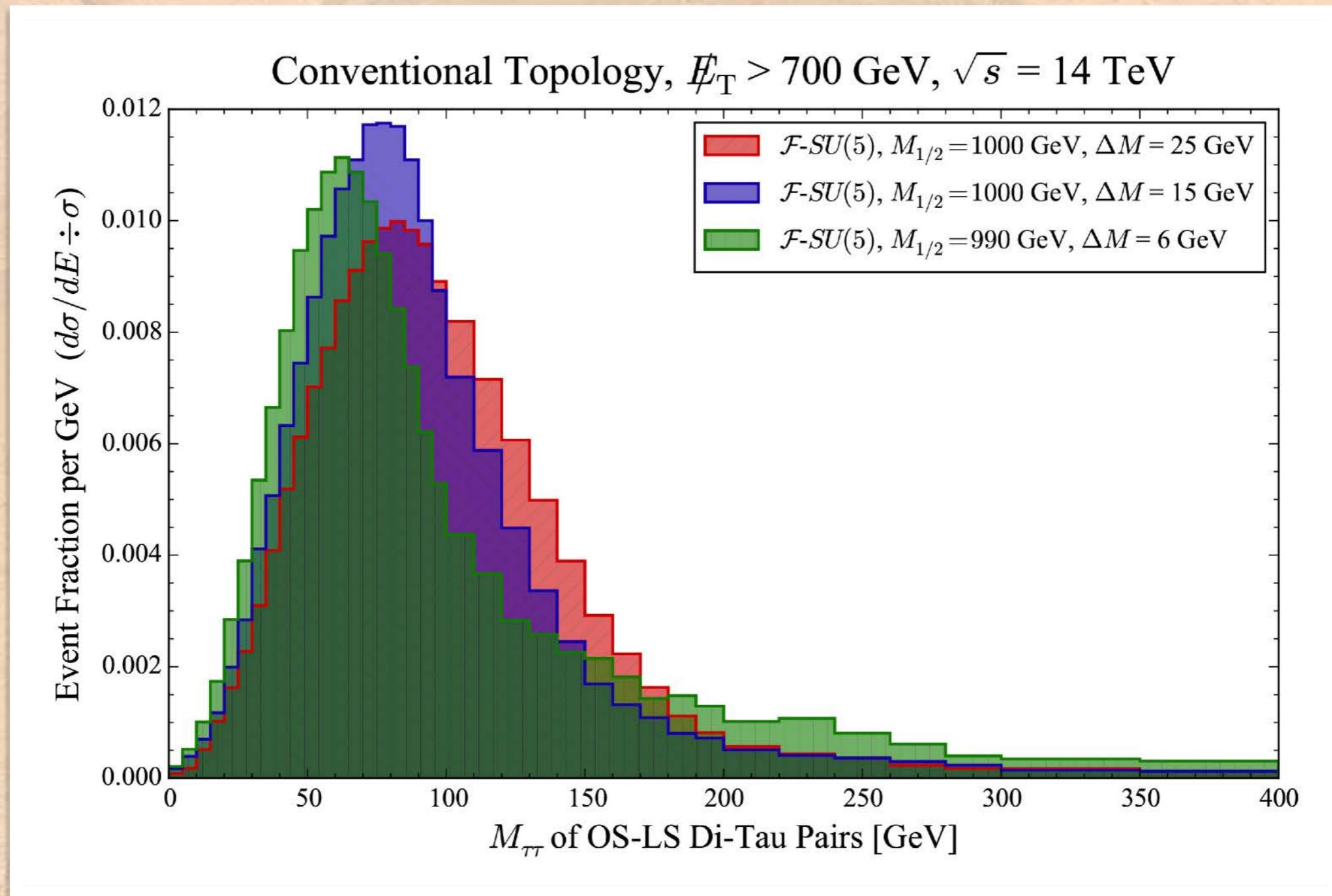
```
PLT_DAT_001 = DIR:"./Cuts_MT2", FIL:"Central:FSU5_VBF_25:1000_*"  
PLT_DAT_002 = DIR:"./Cuts_MT2", FIL:"Central:FSU5_VBF_15:1000_*"  
PLT_DAT_003 = DIR:"./Cuts_MT2", FIL:"Central:FSU5_VBF_6:990_*"  
  
PLT_ESC_001 = KEY:PTM_001, CUT:400      # Leading P_T Cut  
PLT_ESC_002 = KEY:PTM_002, CUT:200      # Sub-leading P_T Cut  
PLT_ESC_003 = KEY:MET_000, CUT:700      # MET Cut  
PLT_ESC_004 = KEY:DIL_001, CUT:1        # Same Sign Dilepton  
PLT_ESC_005 = KEY:DIL_002, CUT:1        # Opposite Sign Dilepton  
  
PLT_CHN_001 = DAT:[001,002,003], KEY:OIM_001, ESC:[+001,+002,+003,+004]  
PLT_CHN_002 = DAT:[001,002,003], KEY:OIM_001, ESC:[+001,+002,+003,+005]
```

```
PLT_HST_001 =  
    IFB:UNDEF,  
    CHN:{($2-$1),001,002},
```

- Opposite- minus Like-Sign dilepton counts are binned on invariant mass
- The signal is compared to itself, subscribing to different selection cuts
- The operation is repeated over each of three registered data sets
- There is an internal limiter ensuring positive semi-def bin values

```
OUT:"./Plots", NAM:"mtt_OS-LS_shape_DeltaM", FMT:"PDF"
```

Transform Bin Channels



Transform Bin Channels

```
PLT_DAT_001 = DIR:"./Cuts_MT2", FIL:"Central:FSU5_VBF_25:1000_*"  
PLT_DAT_002 = DIR:"./Cuts_MT2", FIL:"Central:FSU5_VBF_15:1000_*"  
PLT_DAT_003 = DIR:"./Cuts_MT2", FIL:"Central:FSU5_VBF_6:990_*"  
  
PLT_FSC_001 = KEV•PTM_001 CUT•400 # Leading P.T. Cut
```

- This example also demonstrates variable width binning
- Counts in wide bins are automatically scaled to preserve axis units
- The bin smoothing width “AVG” is independent for each data set

```
PLT_HST_001 =  
    IFB:UNDEF,  
    CHN:{($2-$1).001.002},  
    LFT:0, RGT:[100,200,300,400], SPN:[5,10,20,50]  
    MIN:0.0, MAX:UNDEF,  
    SUM:0, NRM:1, AVG:[3,3,4],  
    LOG:0, LOC:0, CLR:0,  
    TTL:"Conventional Topology, <MET> > 700 GeV, <RTS> = 14 TeV",  
    LBL:[ "$M_{\tau\tau}$ of OS-LS Di-Tau Pairs [GeV]",  
          "Event Fraction per GeV (<DEF>)" ],  
    LGD:[ "$\mathcal{F}$$-SSU(5)$, $M_{1/2} = 1000$ GeV, $\Delta M = 25$ GeV",  
          "$\mathcal{F}$$-SSU(5)$, $M_{1/2} = 1000$ GeV, $\Delta M = 15$ GeV",  
          "$\mathcal{F}$$-SSU(5)$, $M_{1/2} = 990$ GeV, $\Delta M = 6$ GeV" ],  
    OUT:". ./Plots", NAM:"mtt_OS-LS_shape_DeltaM", FMT:"PDF"
```

Transform Bin Channels

```
PLT_DAT_001 = DIR:"./M3/2b_21",
    FIL:[ "BG:MEG:TTBAR*", "BG:MEG:VVJJ*", "BG:MEG:ZJJJJ*", "BG:MEG:WJJJJ*" ]
PLT_DAT_002 = DIR:"./M3/2b_21", FIL:"NMSSM:A:NMSSM*"

PLT_ESC_001 = KEY:LEP_002, CUT:[0,0] # Veto Taus
PLT_ESC_002 = KEY:JET_007, CUT:1 # Force 1 B-Jet pair in Z/H Window
PLT_ESC_003 = KEY:LEP_005, CUT:1 # Force 1 Lepton pair in Z Window
PLT_ESC_004 = KEY:JET_010, CUT:[0,0] # Veto Single Track Jets
PLT_ESC_005 = KEY:ODR_001, CUT:[0,2.5] # Best Higgs Delta R < 2.5
PLT_ESC_006 = KEY:RHR_001, CUT:[2.0] # Met/root(HT) > 2

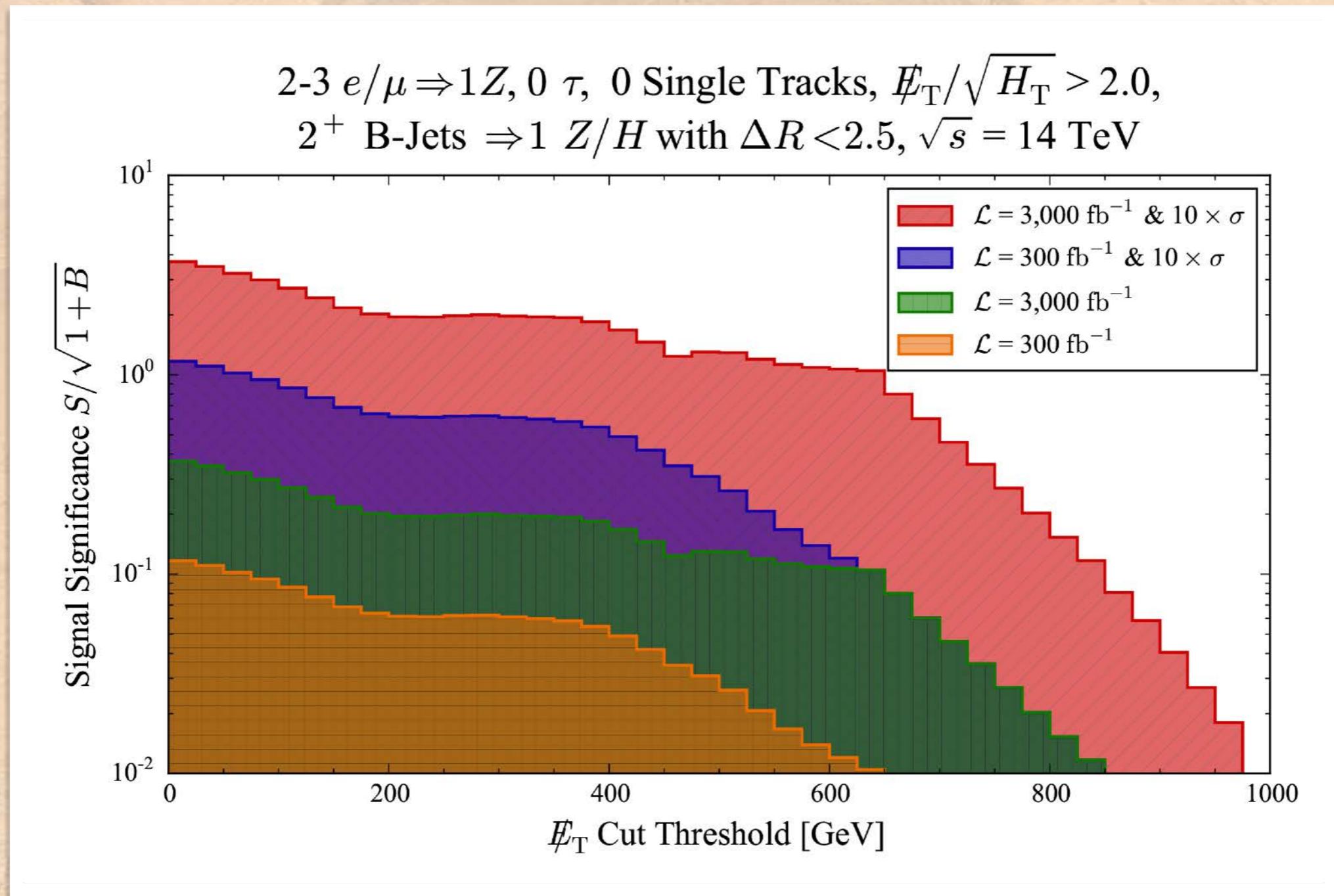
PLT_CHN_001 = DAT:001, KEY:MET_000, ESC:[+001,+002,+003,+004,+005,+006]
PLT_CHN_002 = DAT:002, KEY:MET_000, ESC:[+001,+002,+003,+004,+005,+006]

PLT_HST_001 =
    IFB:300,
    CHN:[
        {100*$2/SRT(1+10*$1),001,002},
        {10*$2/SRT(1+$1),001,002},
        {10*$2/SRT(1+10*$1),001,002},
        {$2/SRT(1+$1),001,002} ],
    ]
```

- Signal significance is again computed by combining Signal & BG Channels
- In this case the same channel is compared at two luminosity scale factors (1x,10x) and two cross section scale factors (1x,10x)

```
"<LUM> = 3,000 <IFB>",
"<LUM> = 300 <IFB>" ],
OUT:"./Plots", NAM:"event_count_MET_OPT_sig_2b_21_300", FMT:"PDF"
```

Transform Bin Channels

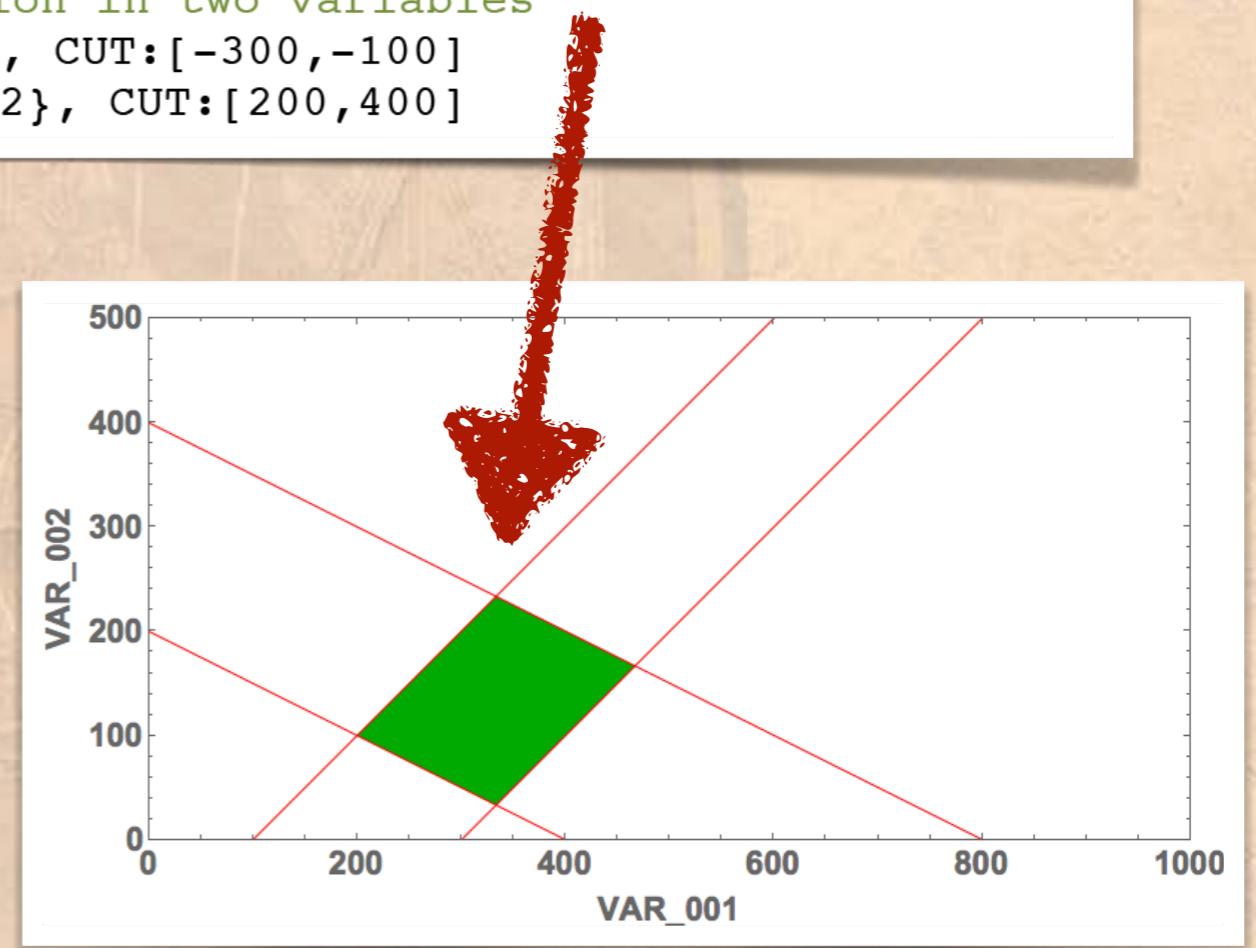


Transform Event Keys

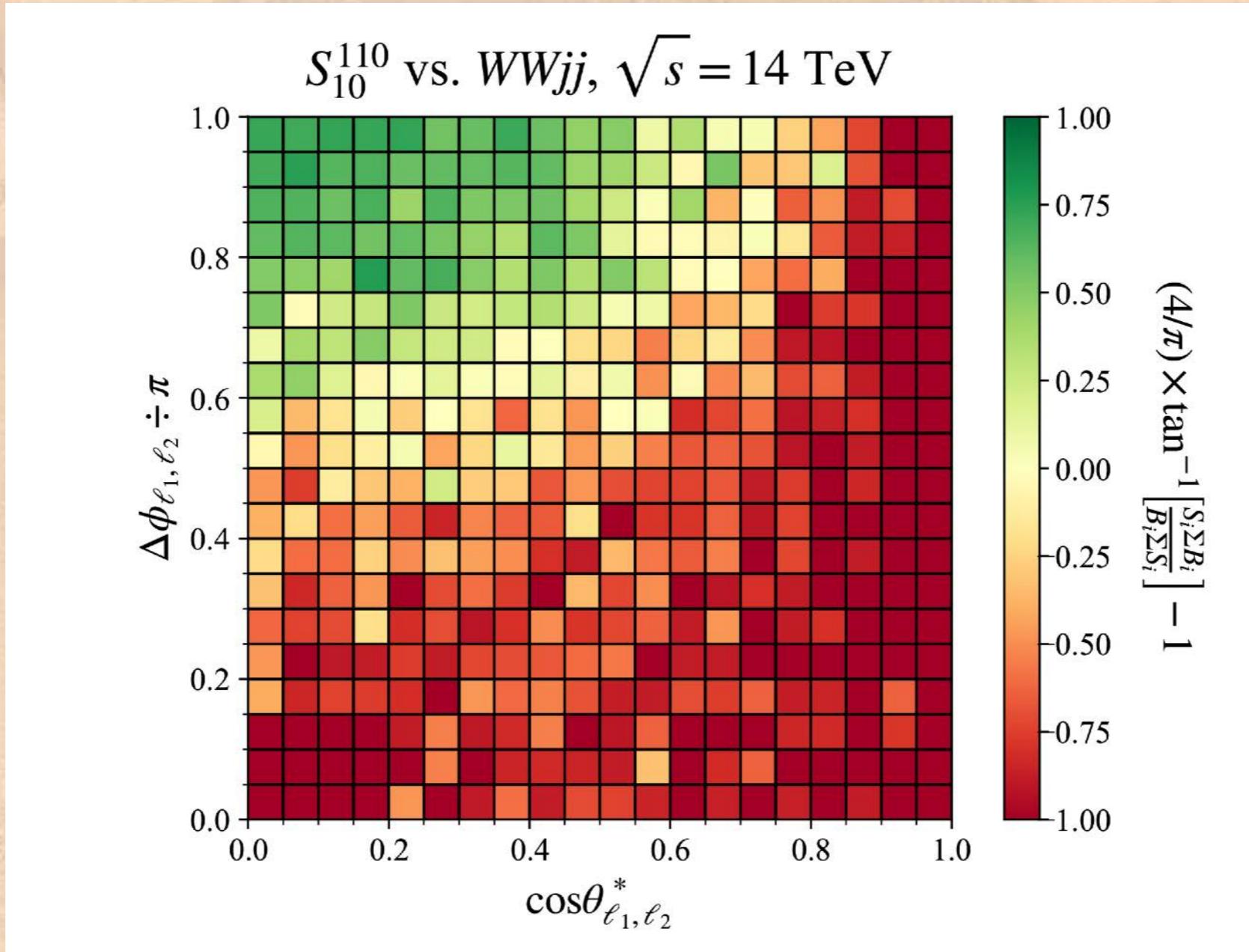
```
# Azimuthal Separation of two 4-vectors in range 0 to Pi
PLT_CHN_001 = DAT:[001,002,003], KEY:{PI()-ABS(PI()-ABS($2-$1)),PHI_001,PHI_002}

# Compound rhomboid selection region in two variables
PLT_ESC_001 = KEY:{$2-$1,VAR_001,VAR_002}, CUT: [-300,-100]
PLT_ESC_002 = KEY:{$2+$1/2,VAR_001,VAR_002}, CUT: [200,400]
```

- ❖ User-defined compound functions of event keys are allowed for event selection and for specification of the independent plotting variable
- ❖ Available functions include basic arithmetic, trigonometry, roots, powers, logarithms, exponentials, min, max, integer, modulus, and average



2-D Histograms



2-D Histograms

```
# Data sets are built from collections of files
DAT_001 = DIR:"./Cuts", FIL:"sl_110_n1_100_S14_001"
DAT_002 = DIR:"./Cuts", FIL:"S14_BG_WWJJ_*"

# Event channels are built from Data sets,
# referencing a pair (x then y) of keys,
# which may be functionally transformed
# using the curly-bracket notation
CHN_001 = DAT:001, KEY:[CTS_001,{\$1/PIE(),ODP_001}]
CHN_002 = DAT:002, KEY:[CTS_001,{\$1/PIE(),ODP_001}]
```

- Channels are built out of a 2-D ARRAY of KEYS
- You can define multiple channels in order to combine them in some way

```
# pairwise vectorized inputs. A second input
# to the title key references the colorbar.
H2D_001 =
CHN:{(4/PIE())*ATN(\$1,\$2)-1},001,002},
LFT:0, RGT:1, BNS:20, NRM:-1, MIN:-1, MAX:+1,
TTL:["\$S_{10}^{110}\$ vs. \$WWjj\$, \$\\sqrt{s} = 14\$ TeV",
      "\$(4/\\pi) \\times \\tan^{-1}\\left[\\frac{S_i \\Sigma B_i}{B_i \\Sigma S_i}\\right]-1\$"],
LBL:["\$\\cos \\theta^{*}_{\\ell_1,\\ell_2}\$", "\$\\Delta \\phi_{\\ell_1,\\ell_2} \\div \\pi\$"],
NAM:"H2D_CST_ODP"
```

2-D Histograms

```
# Data sets are built from collections of files
DAT_001 = DIR:"./Cuts", FIL:"sl_110_n1_100_S14_001"
DAT_002 = DIR:"./Cuts", FIL:"S14_BG_WWJJ_*
```

- Ultimately, use a SINGLE 2-Dimensional Channel for plotting
- LFT/RGT can be arrays if different x-y ranges are required
- MIN/MAX are now the range of “z” values

```
# Two-dimensional histograms are built from
# a single two-dimensional event channel,
# which may represent a cell-by-cell functional
# transformation of one or more basic channels.
# Binning specifications and axis labels accept
# pairwise vectorized inputs. A second input
# to the title key references the colorbar.

H2D_001 =
CHN:{(4/PIE())*ATN($1,$2)-1},001,002},
LFT:0, RGT:1, BNS:20, NRM:-1, MIN:-1, MAX:+1,
TTL:["$S_{10} \{110\}$ vs. $WWJJ$, $\sqrt{s} = 14$ TeV",
      "$(\frac{4}{\pi}) \times \tan^{-1}[\frac{S_i \Sigma B_i}{B_i \Sigma S_i}] - 1$"],
LBL:["$\cos \theta^{*}_{\ell 1, \ell 2}$", "$\Delta \phi_{\ell 1, \ell 2} / \pi$"],
NAM:"H2D_CST_ODP"
```

MINOS

(Machine Intelligent Optimization of Significance)

Intelligence *noun* \in-'te-lə-jən(t)s\ (www.merriam-webster.com)

1 a (1) : the ability to learn or understand or to deal with new or trying situations

: REASON *also*, : the skilled use of reason

4 : the ability to perform computer functions

What does MInOS do?

- ❖ MInOS automates BDT Machine Learning in a Collider Context
- ❖ It reads event features computed by AEACuS
- ❖ It correctly combines distinct / over-sampled MC by cross section
- ❖ It trains for optimal Signal/BG discrimination (XGBoost backend)
- ❖ It facilitates ensemble training against distinct BGs with merging
- ❖ It generates density, significance, feature importance, & ROC plots (MatPlotLib backend) from validation data (1/3 by default)
- ❖ It lets Pheno Projects skip overhead & get answers QUICKLY

What is a BDT?

- ❖ Boosted Decision Trees are a type of Supervised Machine Learning
- ❖ “Hypothesis **Boosting**” is a technique for combining a number of “weak learners” (here shallow **Decision Trees**) into a “strong learner”
- ❖ Each tree separates signal (class 1) from background (class 0) via successive forks at selected split points on one data feature at a time
- ❖ Each terminal leaf carries a score, totaled over trees for a result on (0,1)
- ❖ Later trees focus on misclassifications from earlier trees (boosting!)



Why BDTs for Physics?

- ❖ Binary classification problems (Signal vs. Background) are common
- ❖ We want to maximize discrimination power
- ❖ We want to eliminate bias and work efficiently
- ❖ We want to incorporate domain knowledge & expertise
- ❖ We want to understand what the machine learning learned

BDTs balance POWER with TRANSPARENCY

How is MInOS Set Up?

```
# Construct Data Sets From Files
MIN_DAT_001 = DIR:"./Cuts", FIL:["TTBarJJ_*"]
MIN_DAT_002 = DIR:"./Cuts", FIL:["ZZJJ_*", "AZJJ_*", "WZJJ_*", "WWJJ_*"]
MIN_DAT_003 = DIR:"./Cuts", FIL:"l1JJJJ_*"
MIN_DAT_101 = DIR:"./Cuts", FIL:"MulMulJJ_s1_110_n1_50_"

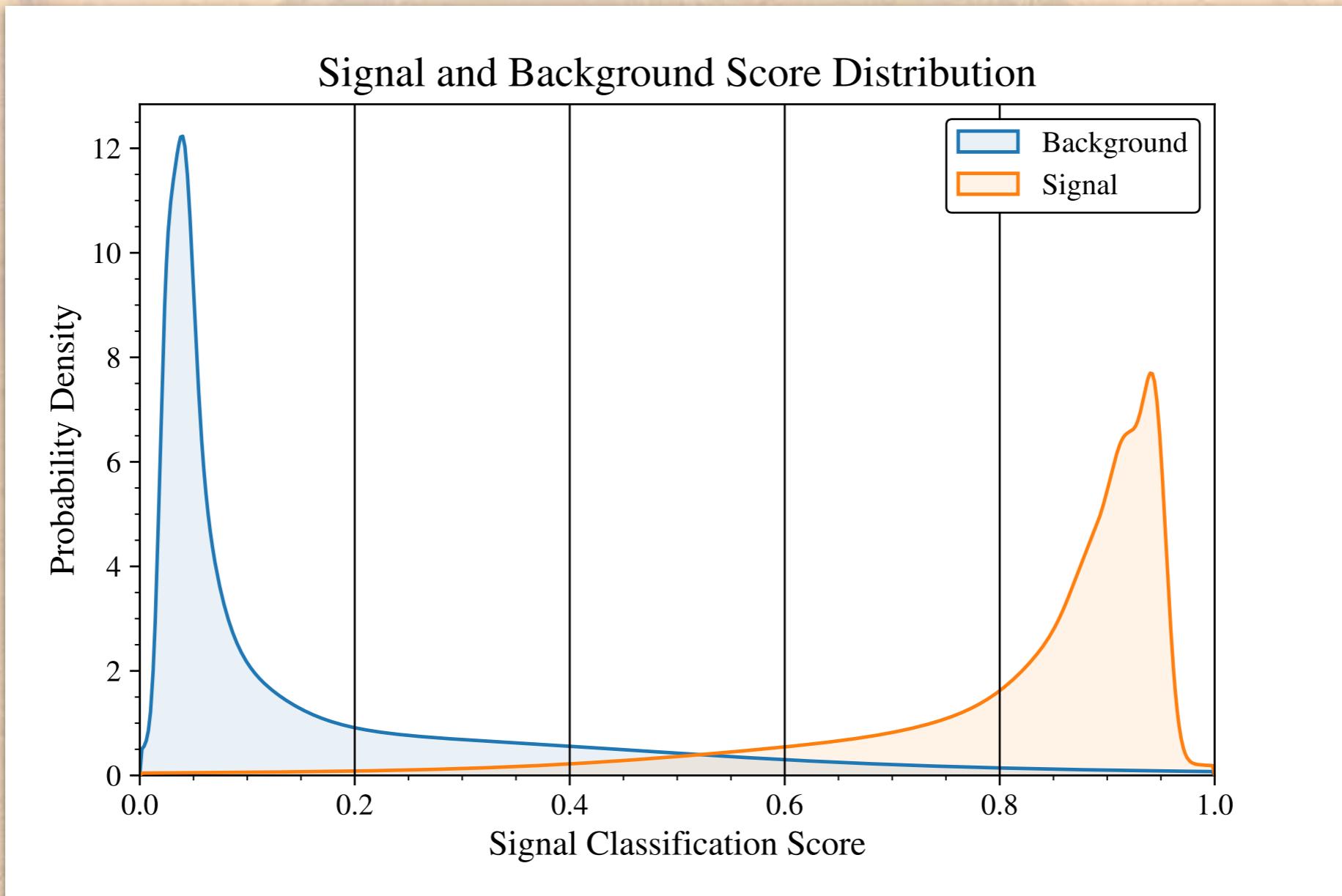
# Specify training features for inclusion (Channel 000 is for defaults)
MIN_CHN_000 = INC:[
    MAS_001,MAS_003,PTM_001,PTM_002,PTM_003,ETA_001,ETA_002,ETA_003,
    MET_000,MHT_000,MEF_000,TTM_001,CTS_001,ATM_001,ATM_002,
    MDP_001,MDP_002,MDP_003,ODP_001,ODP_002,ODP_004,
    VAR_001,VAR_002,VAR_004,VAR_011,VAR_012,VAR_013,VAR_021]

# Construct Channels from Data Sets
MIN_CHN_001 = DAT:[001,002,003], LBL:0
MIN_CHN_101 = DAT:101, LBL:1
# Construct Training from Channels
MIN_TRN_001 = CHN:[001,101]

# Apply Secondary Event Selections Before Training
MIN_ESC_001 = KEY:MAS_001, CUT:[101,81] # Z-Window Cut
MIN_ESC_002 = KEY:MET_000, CUT:75 # Require MET > 75 GeV
MIN_CHN_011 = DAT:[001,002,003], LBL:0, ESC:[+001,+002]
MIN_CHN_111 = DAT:101, LBL:1, ESC:[+001,+002]
MIN_TRN_011 = CHN:[011,111]
```

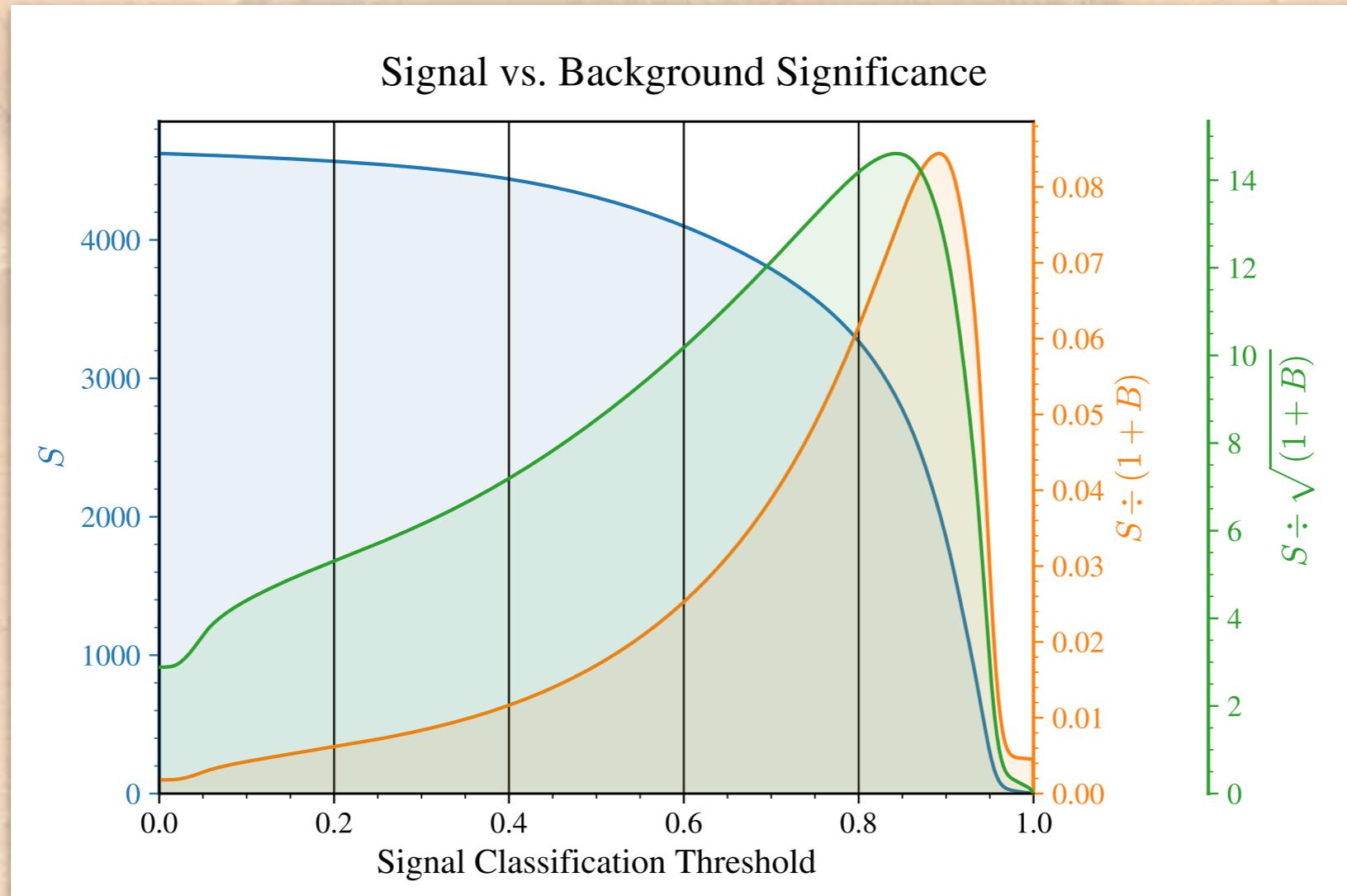
Card File from EXAMPLE_5 in GitHub Package

MInOS Output



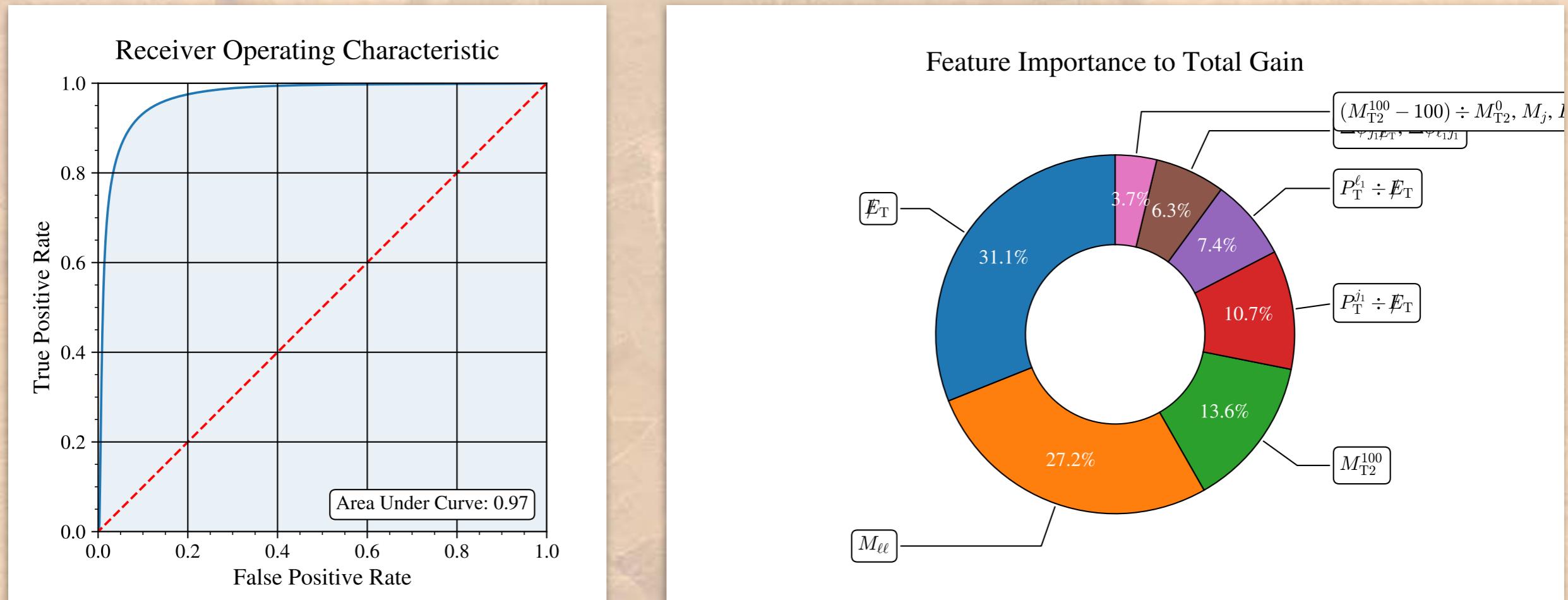
- ❖ Signal & Background Probability Density Visualizes Separation

MInOS Output



- ❖ Survival fraction of S , B as a function of the classification threshold are used to show achievable significance (at specified luminosity)

MInOS Output



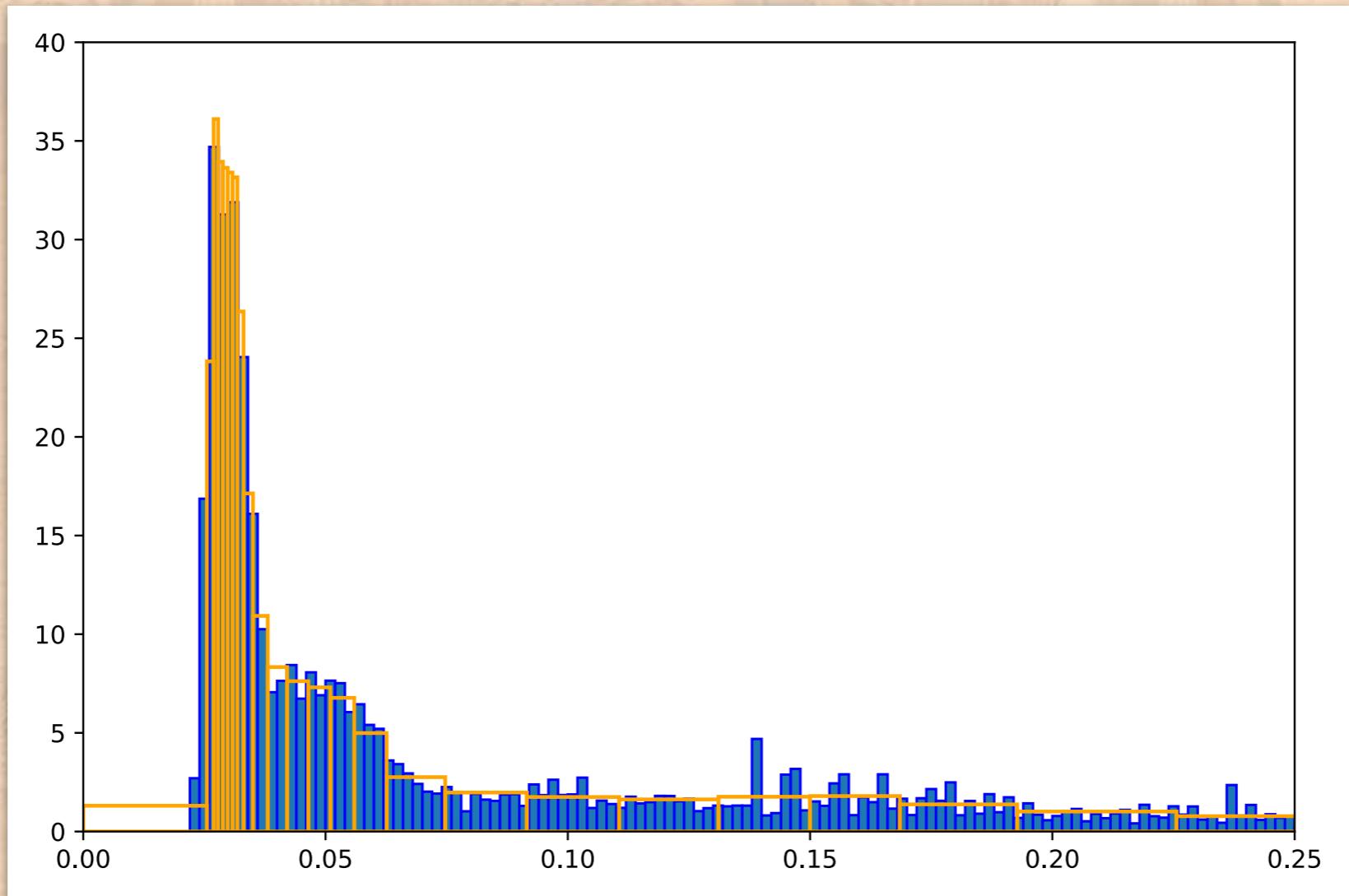
- ❖ The ROC curve is a standard metric of S/B separability
- ❖ A feature importance chart clarifies what is going on inside the BDT

Data Smoothing

- ❖ The statistical population of events grows sparse with harder cuts
- ❖ Smoothing may better approximate the reality of continuum data
- ❖ Naive interpolation (e.g. cubic spline) can induce unphysical artifacts
- ❖ We want to retain sharpness where clustering is real while washing out jitter where statistical event densities are low
- ❖ A proprietary multi-step solution is adopted to meet these goals

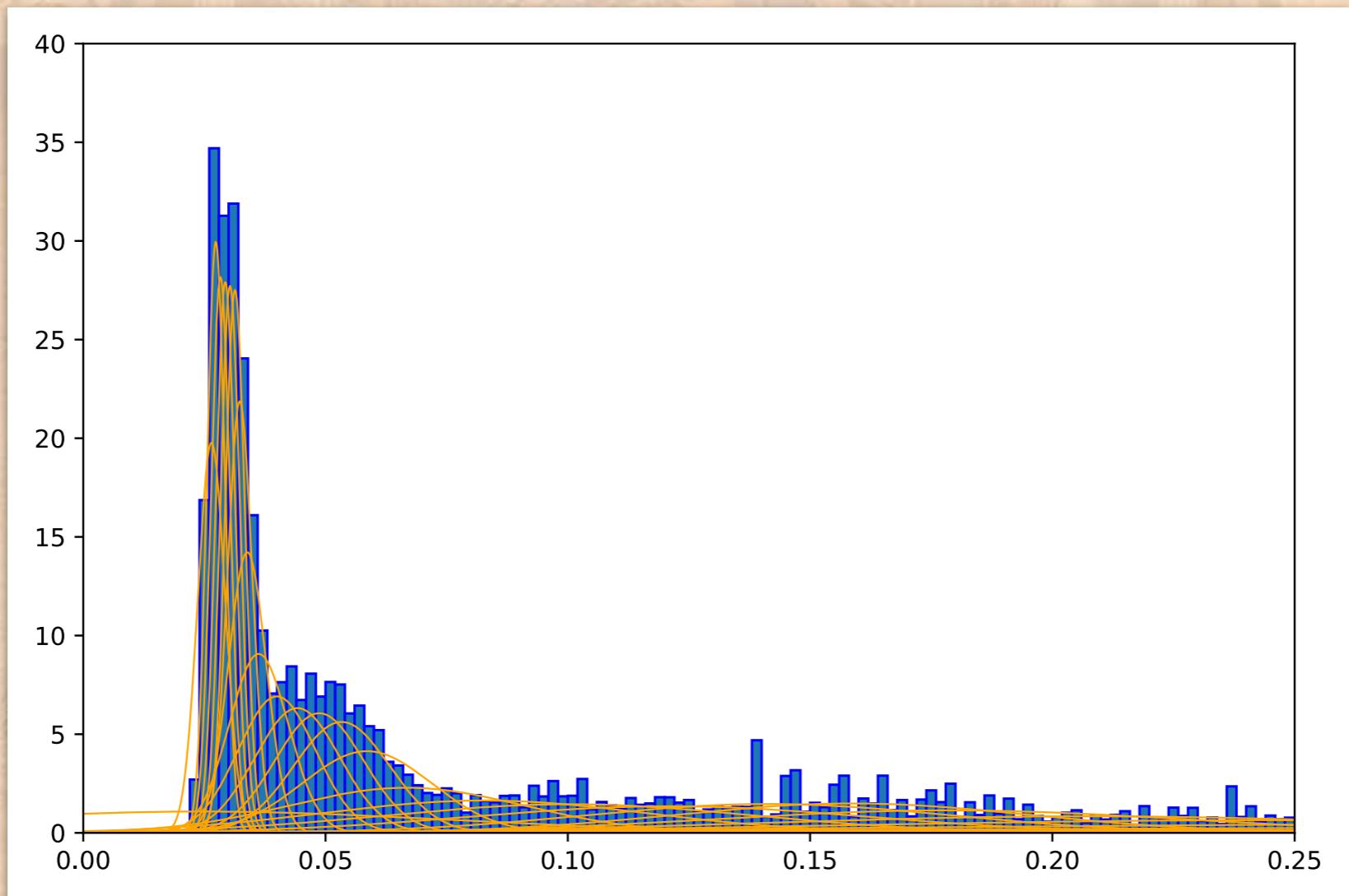
Data Smoothing

- ❖ First, we do variable-width binning with equal areas (cross sections)



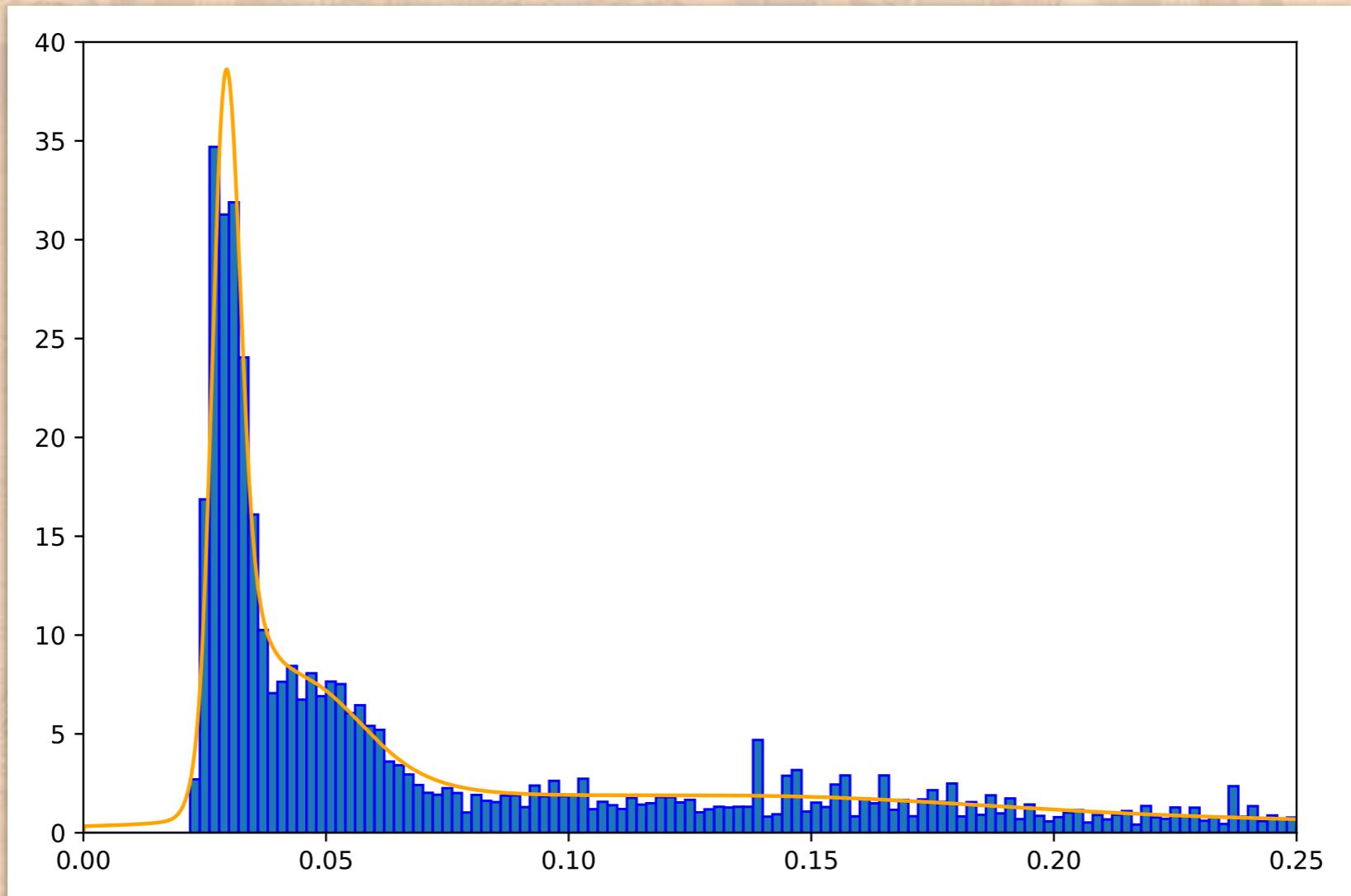
Data Smoothing

- ❖ Then, we populate narrow fixed bins via Gaussians \propto prior widths



Data Smoothing

- ❖ Finally, we sum and scale to generate a smooth density of area one



What is XGBoost?

- ❖ XGBoost (Extreme Gradient Boosting) by Tianqi Chen is a popular, innovative, widely available, and very fast BDT implementation
- ❖ Trees are built “greedily” (no backtracking), with the splitting feature, splitting value, and leaf score selected to optimize an objective \mathcal{L}
- ❖ This is guided by first (gradient) and second (Hessian) derivatives of the loss-function with respect to the class estimator of the nth object

$$g_n \equiv \frac{\partial \mathcal{L}_L}{\partial \hat{y}_n}$$
$$h_n \equiv \frac{\partial^2 \mathcal{L}_L}{\partial \hat{y}_n^2}$$

$$\delta \mathcal{L}_L \simeq \sum_{n=1}^N \left\{ g_n \delta \hat{y}_n + h_n \frac{\delta \hat{y}_n^2}{2} \right\}$$
$$\partial(\delta \mathcal{L}) / \partial(\delta \hat{y}_n) = 0,$$
$$\delta_0 \hat{y}_n = -g_n / h_n$$
$$(-\delta_0 \mathcal{L}_L) \simeq \sum_n g_n^2 / (2h_n)$$

XGBoost Details

- ❖ It is not possible to correct misclassifications per event.
- ❖ Rather, events with common features flow similarly through the decision tree, and “vote” for the score carried by their destination node
- ❖ The max $-(\delta_0 \mathcal{L})$ split is selected
- ❖ “Regulators” limit overtraining

$$G_\ell \equiv \sum_{n=1}^N g_n \times \delta_{\ell,\ell'(\vec{x}_n)}$$

$$H_\ell \equiv \sum_{n=1}^N h_n \times \delta_{\ell,\ell'(\vec{x}_n)}$$

$$\mathcal{L}_\Omega = \gamma L + \alpha \sum_{\ell=1}^L |s_\ell| + \lambda \sum_{\ell=1}^L \frac{s_\ell^2}{2}$$

$$\delta \mathcal{L} \simeq \sum_{\ell=1}^L \left\{ \gamma + \alpha |s_\ell| + G_\ell s_\ell + (H_\ell + \lambda) \frac{s_\ell^2}{2} \right\}$$

$$s_0^\ell = -\frac{G_\ell \pm \alpha}{H_\ell + \lambda}$$

$$(-\delta_0 \mathcal{L}) \simeq \sum_{\substack{\ell=1 \\ (|G_\ell| > \alpha)}}^L \left\{ \frac{(G_\ell \pm \alpha)^2}{2(H_\ell + \lambda)} - \gamma \right\}$$

Binary Logistic Regression

- ❖ The simplest objective is the Mean-Square Error
- ❖ MIInOS uses a binary logistic objective, yielding continuous classification scores on 0 to 1

$$\mathcal{L}_L = \sum_{n=1}^N \frac{(y_n - \hat{y}_n)^2}{2}$$

$$g_n \equiv \frac{\partial \mathcal{L}_L}{\partial \hat{y}_n} = (\hat{y}_n - y_n)$$

$$h_n \equiv \frac{\partial^2 \mathcal{L}_L}{\partial \hat{y}_n^2} = 1$$

“logistic” function $p = \frac{1}{1 + e^{-y}} \in \{0, 1/2, 1\}$

“logit” function $y = \ln \left(\frac{p}{1-p} \right) \in \{-\infty, 0, +\infty\}$

$$\mathcal{L}_L = - \sum_{n=1}^N \left\{ p_n \ln(\hat{p}_n) + (1 - p_n) \ln(1 - \hat{p}_n) \right\}$$

$$\begin{aligned} g_n &\equiv \frac{\partial \mathcal{L}_L}{\partial \hat{y}_n} = \frac{\partial \mathcal{L}_L}{\partial \hat{p}_n} \times \frac{\partial \hat{p}_n}{\partial \hat{y}_n} \\ &= -p_n \times (1 - \hat{p}_n) + \hat{p}_n \times (1 - p_n) \\ &= \hat{p}_n - p_n \end{aligned}$$

$$\begin{aligned} h_n &\equiv \frac{\partial^2 \mathcal{L}_L}{\partial \hat{y}_n^2} = \left(\frac{\partial^2 \mathcal{L}_L}{\partial \hat{y}_n \partial \hat{p}_n} = 1 \right) \times \frac{\partial \hat{p}_n}{\partial \hat{y}_n} \\ &= \hat{p}_n \times (1 - \hat{p}_n) \end{aligned}$$

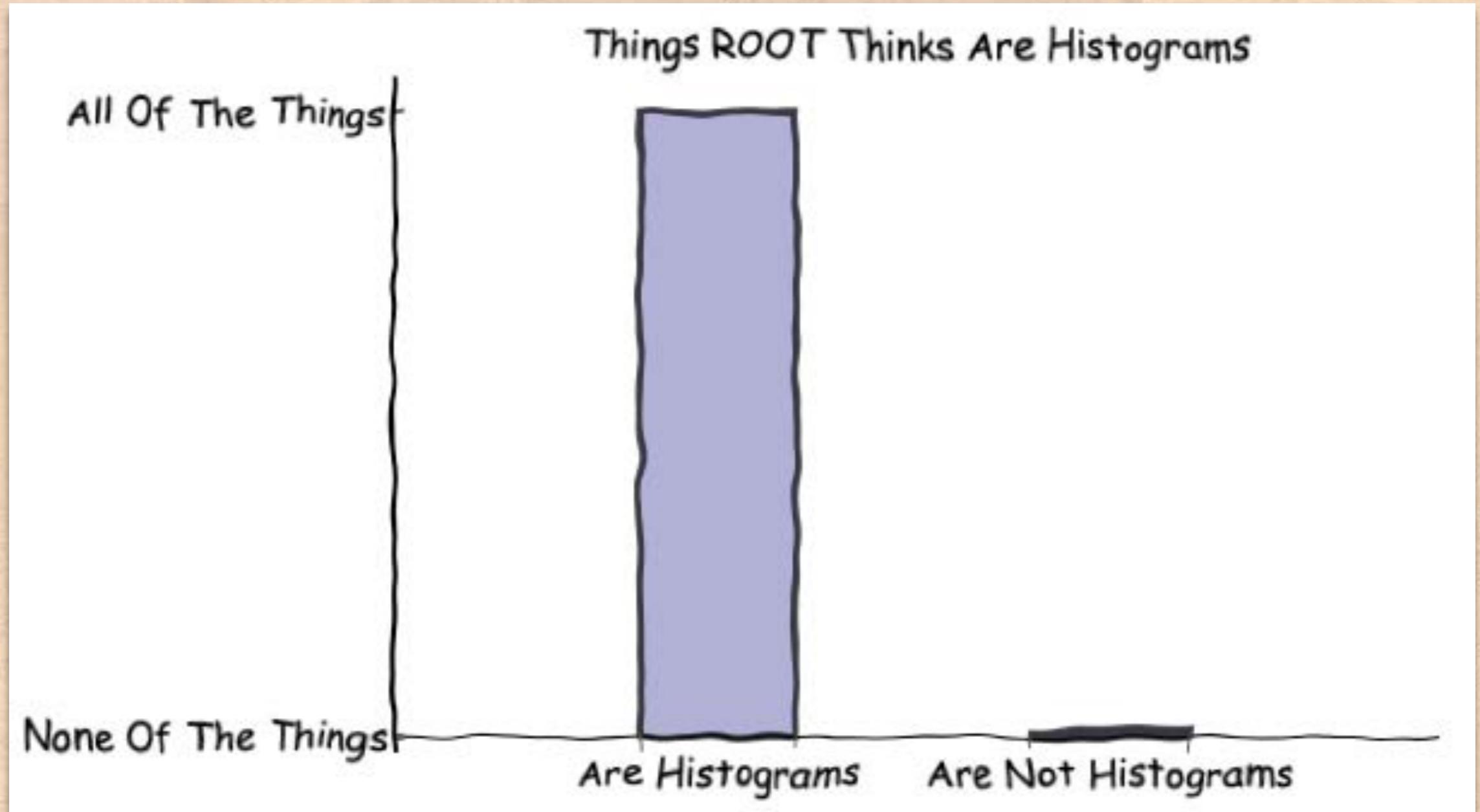
Note on Tuning / Weights

- ❖ A number of hyper-parameters, including the objective-level regulators γ , α , and λ are available to confront “Bias-Variance” issues
- ❖ In short, one must not add complexity without benefit (like χ^2/DOF)
- ❖ XGBoost also allows specification of maximal tree depth and count, with handles for “early stopping” or “pruning” when learning slows
- ❖ MIInOS balances data sets sent for training by separately normalizing the signal and background cross section to unity
- ❖ This stabilizes optimal numerical values of various hyper parameters, eliminates tension between intensive/extensive scaling, and induces a natural $\mathcal{O}(1)$ scale for the gradient, Hessian, and regulators

AEACUS, RHADAMANTHUS & MINOS

- ❖ The joint package is now ready to use, available at GitHub
- ❖ <https://github.com/joelwwalker/AEACuS>
- ❖ Please join today at 4:00 (local time) for a demo / tutorial
- ❖ Please contact author directly: **jwalker@shsu.edu**
- ❖ If you are interested in teaming up, borrowing features, building an analysis library, or doing validations, please Let Me Know!

Thank You!



“ Then spake Zeus: . . . ‘The cases are now indeed judged ill and it is because . . . many . . . who have wicked souls are clad in fair bodies and ancestry and wealth, and . . . the judges are confounded . . . , having their own soul muffled in the veil of eyes and ears and the whole body. . . . They must be stripped bare of all those things . . . , beholding with very soul the very soul of each immediately. . . . [I] have appointed sons of my own to be judges; two from Asia, **Minos** and **Rhadamanthus**, and one from Europe, **Aeacus**. These . . . shall give judgement in the meadow at the dividing of the road, whence are the two ways leading, one to the Isles of the Blest . . . , and the other to Tartaros.’ ”

– Plato, *Gorgias* (trans. Lamb)

