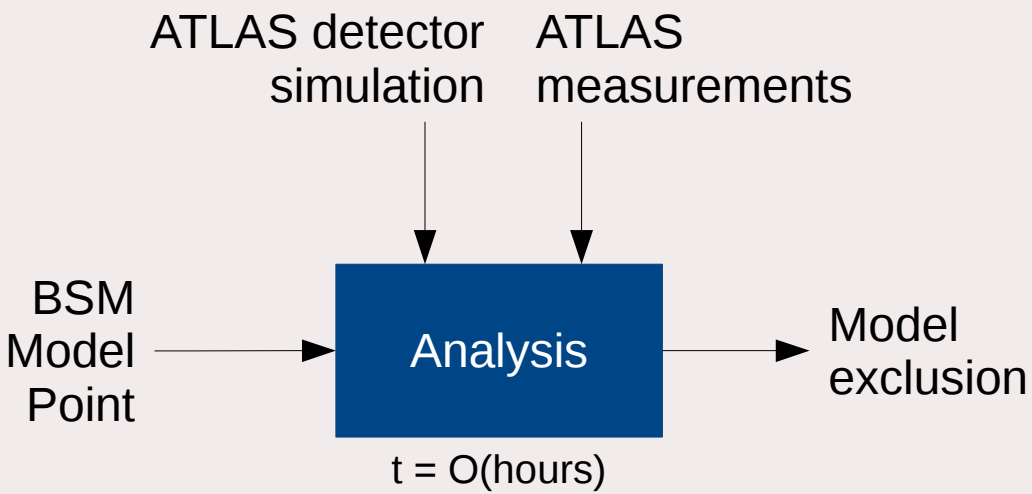
SUSY-AI: Generalizing LHC limits on Supersymmetry with Machine Learning

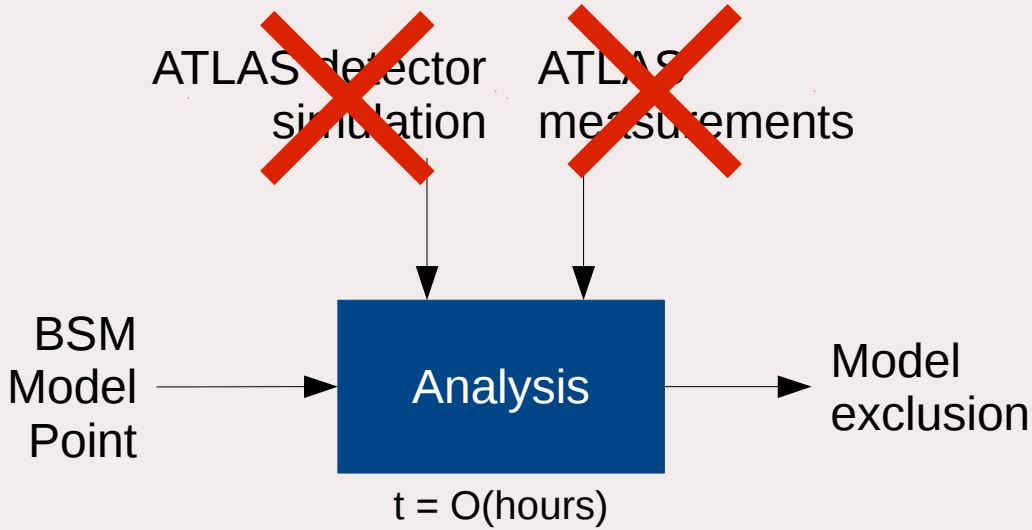
Sascha Caron, Jong Soo Kim, Krzysztof Rolbiecki, Roberto Ruiz de Austri, Bob Stienen

Radboud University

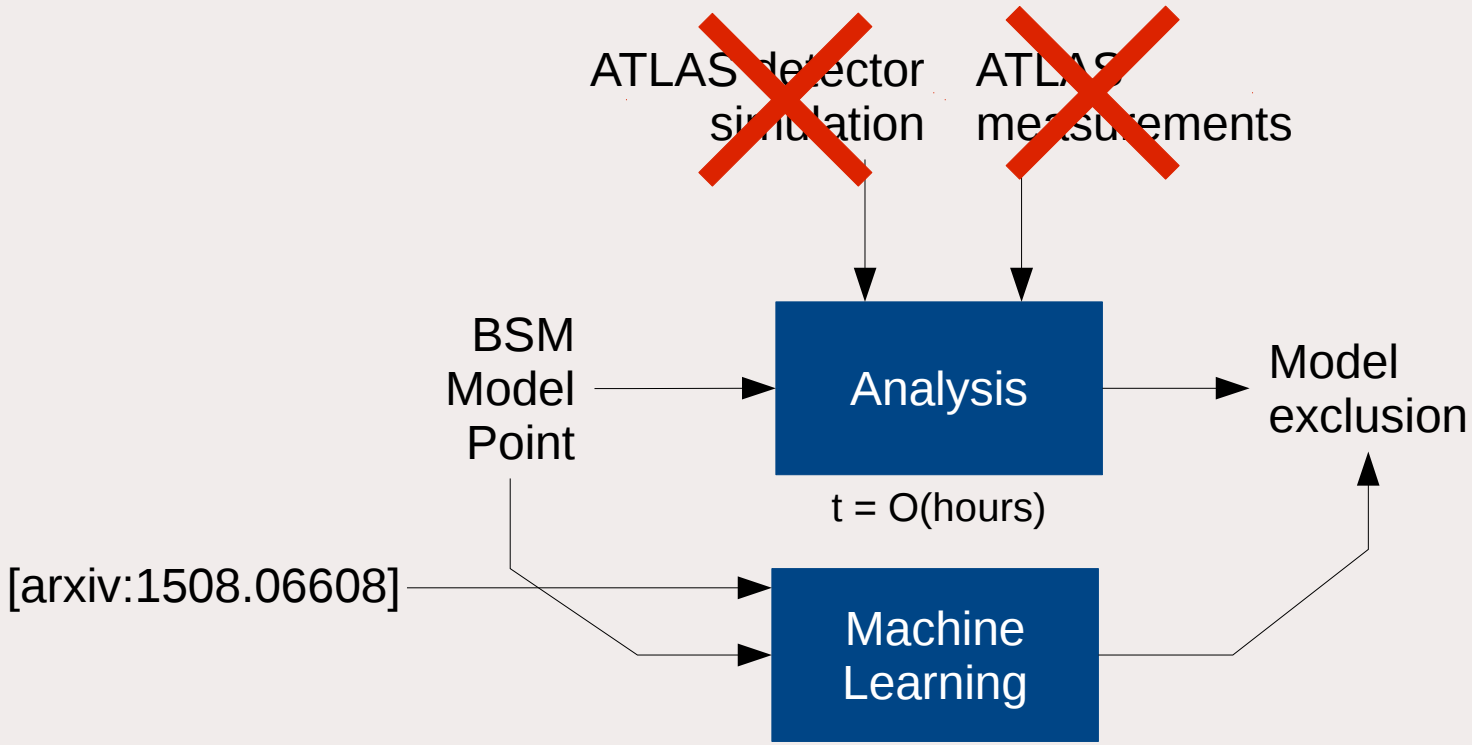
Exclusion analysis



Exclusion analysis



Exclusion analysis



Classification

Accuracy

$$\frac{TP + TN}{TP + FP + FN + TN}$$

Precision

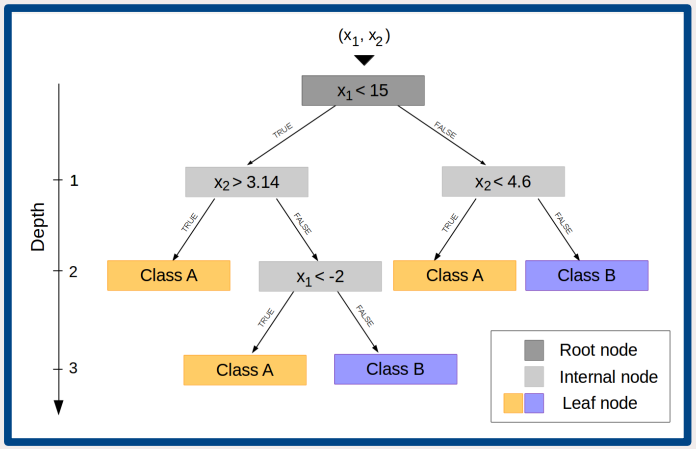
$$\frac{TP}{TP + FP}$$

Brier score

$$\frac{1}{N}\sum_{i=1}^{N}(p_i - t_i)^2$$

		True classification	
		Positive	Negative
Prediction	Positive	True positive (TP)	False Positive (FP) *Error of the first kind*
	Negative	False Negative (FN) *Error of the second kind*	True Negative (TN)

Random Forest



Boosting

- Combining multiple decision trees
- Classification based on majority vote

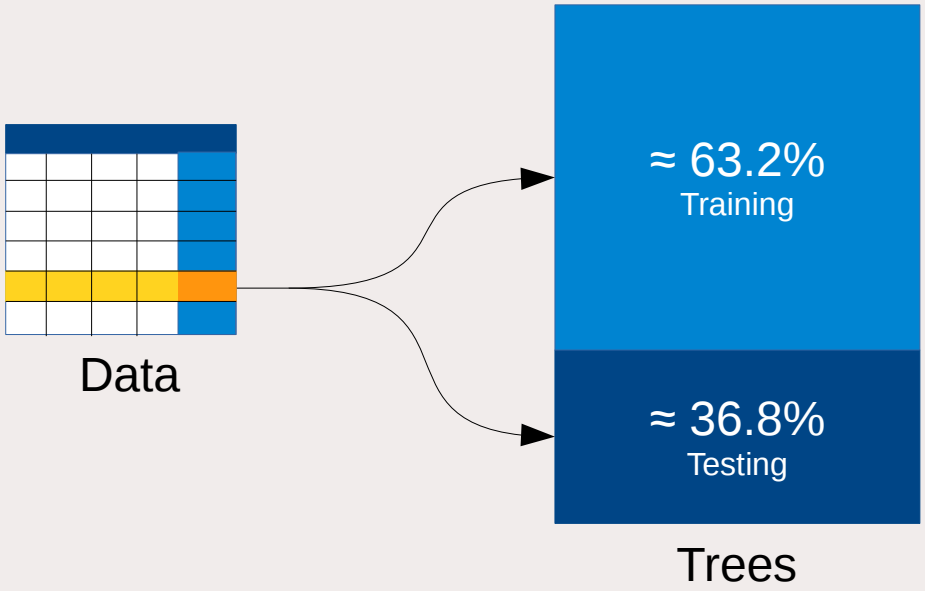
Random subspace method

- Only a random selection of features is considered for the split at each node.

Bootstrap aggregating (bagging)

- Each tree is trained on a subset of training data
- Subset generated by picking with replacement
- On average 63.2% unique attribute sets per tree

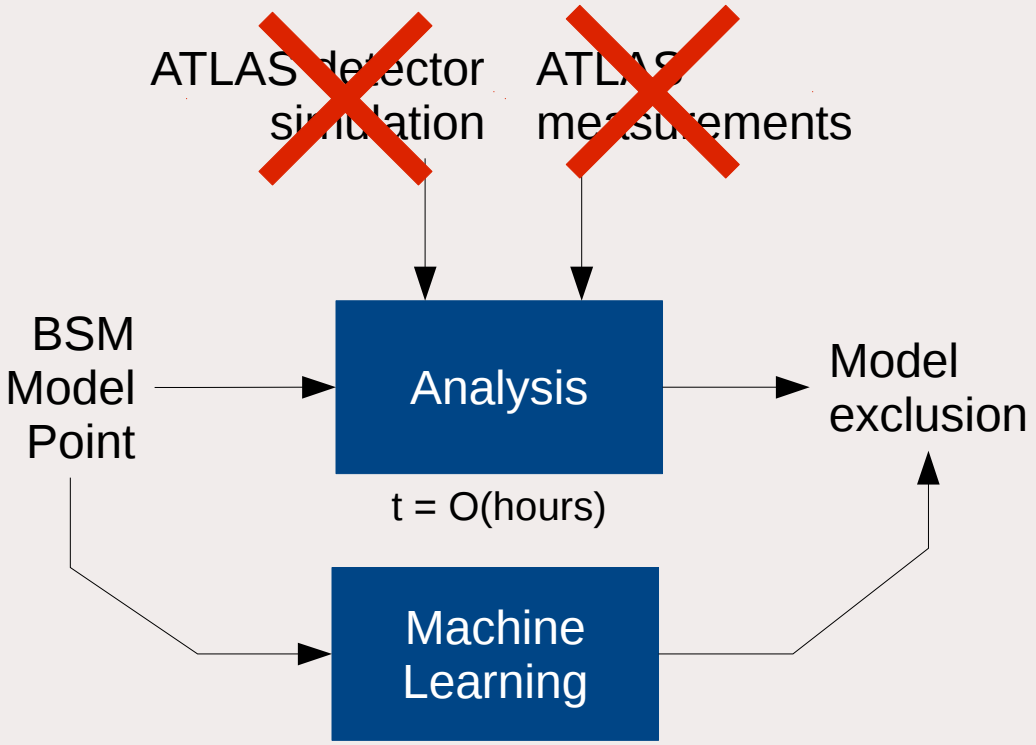
Out-of-bag estimation

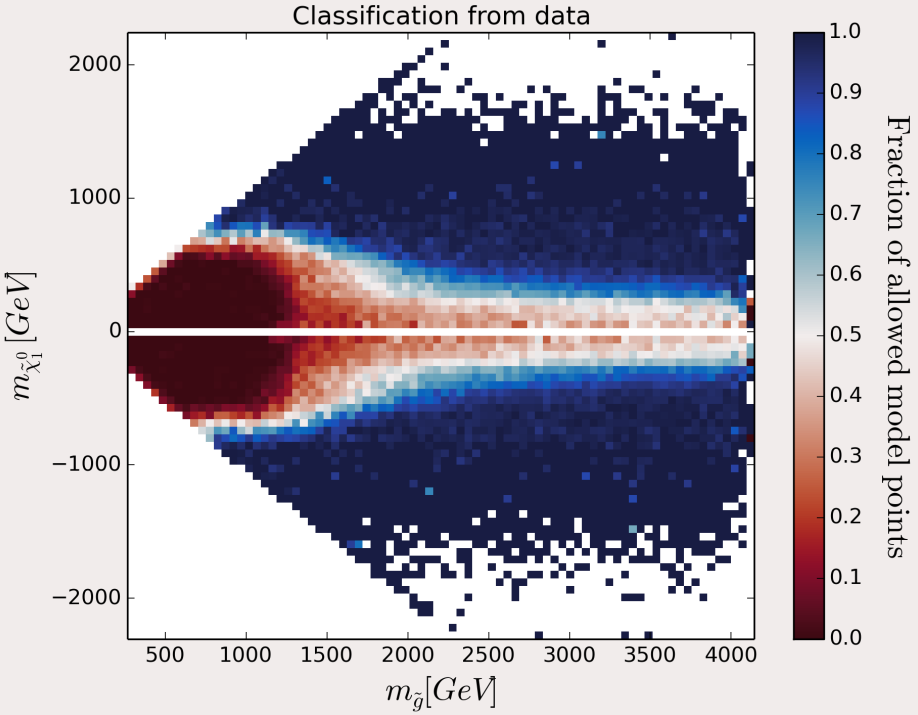


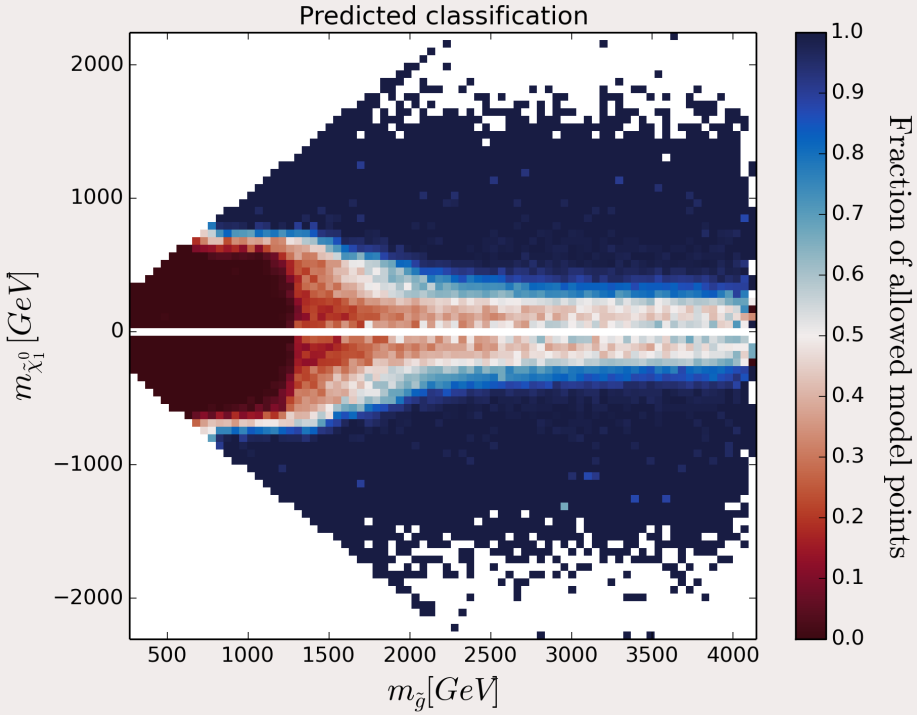
- ≈ 36.8% of the trees is not trained on attributeset i
- Needed: large amount of trees
- No seperate testing dataset needed!

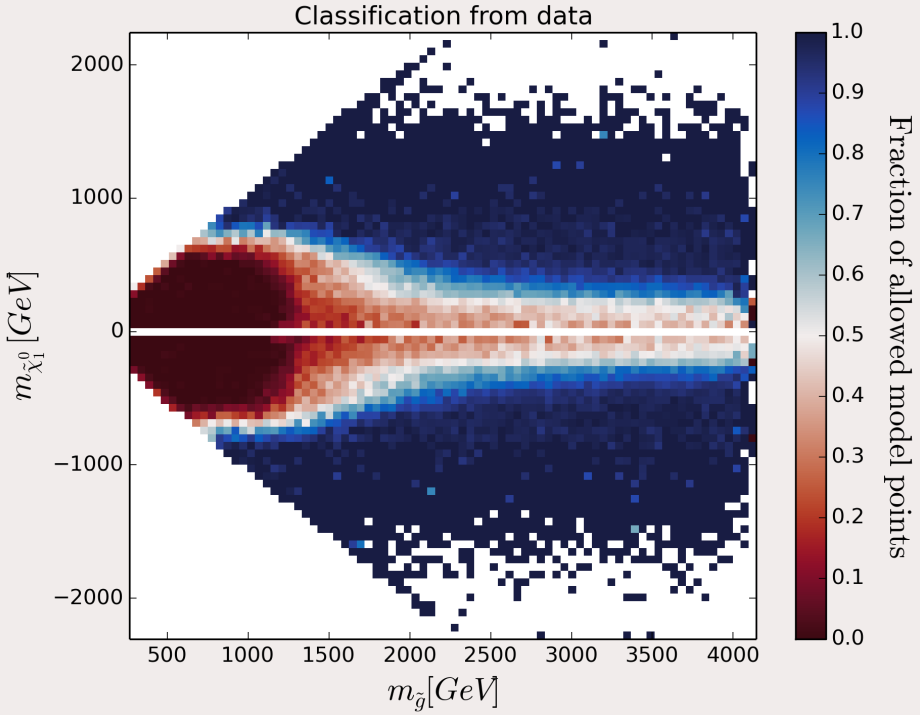
Machine Learning and ATLAS

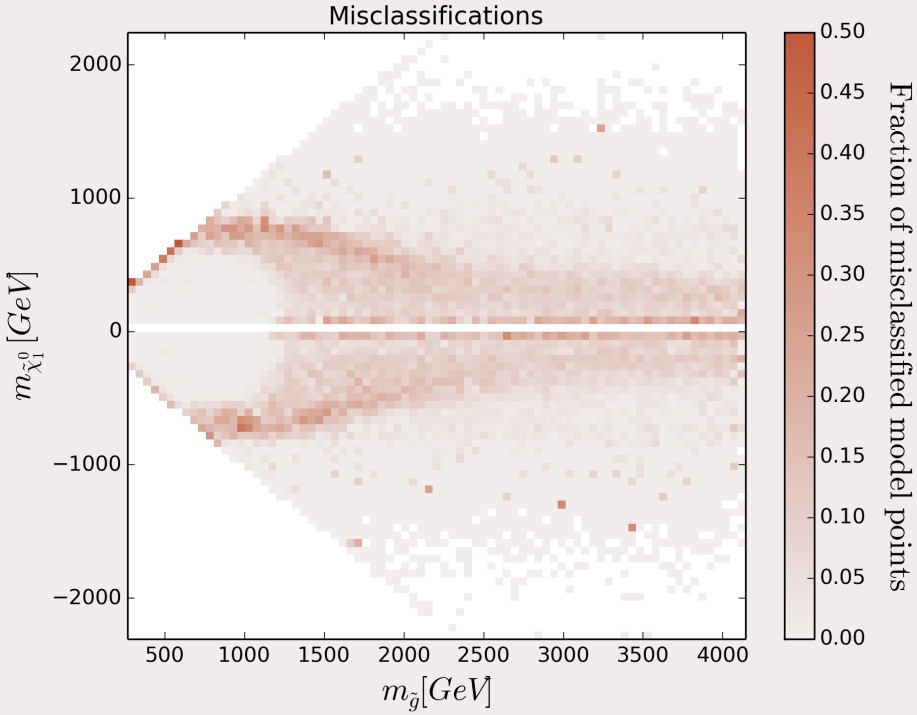
- Classification
- Random Forest
  - 870 trees
  - Maximum depth of 30
- MSSM parameter space
- 310,324 model points
  [arxiv:1508.06608]
- Out-of-bag estimation

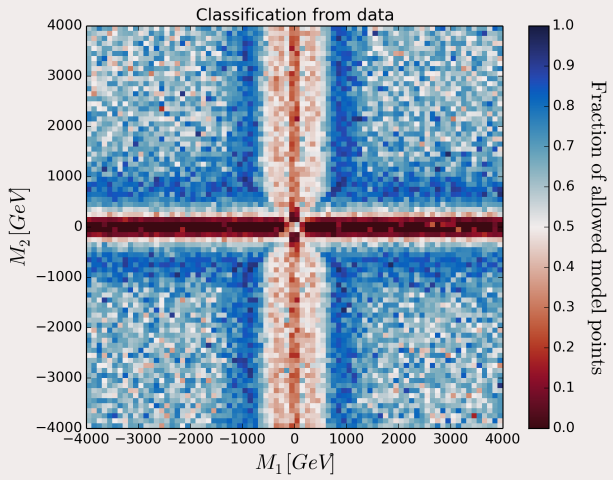


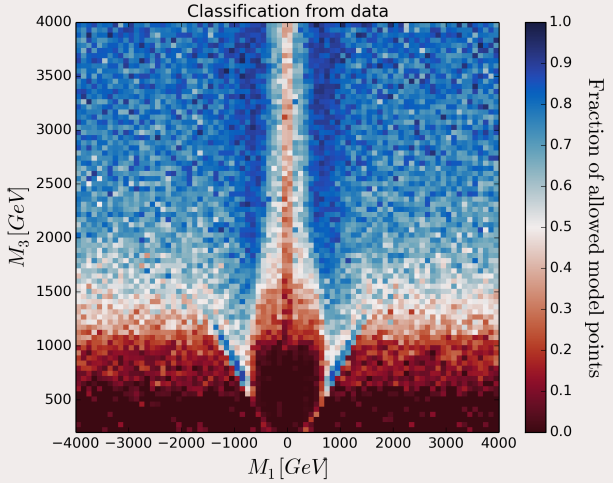

Classification from data
$m_{\tilde{\chi}_1^0}[GeV]$
$m_{\tilde{g}}[GeV]$
Fraction of allowed model points

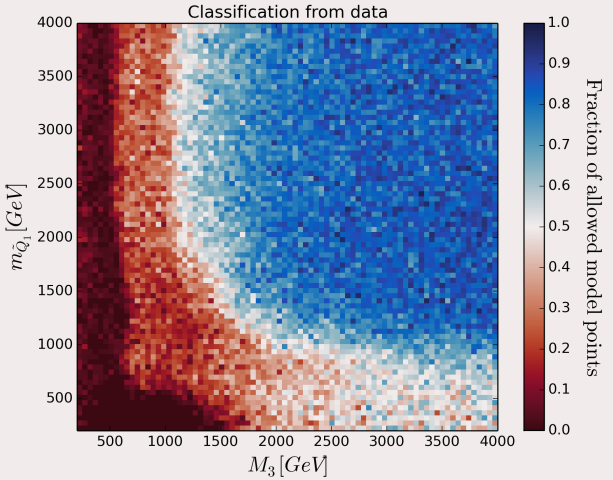

Predicted classification
$m_{\tilde{\chi}_1^0}[GeV]$
$m_{\tilde{g}}[GeV]$
Fraction of allowed model points

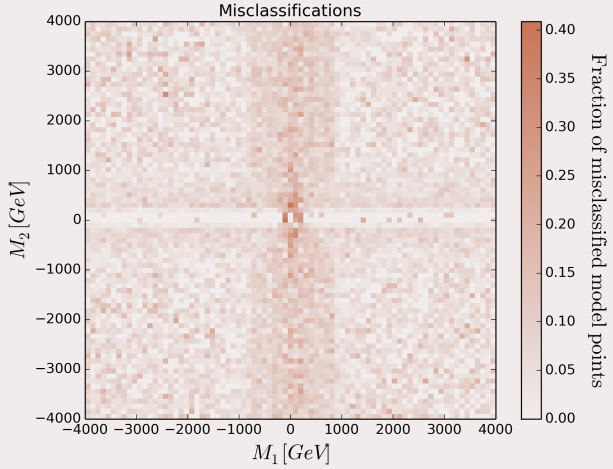

Classification from data
$m_{\tilde{\chi}_1^0}[GeV]$
$m_{\tilde{g}}[GeV]$
Fraction of allowed model points

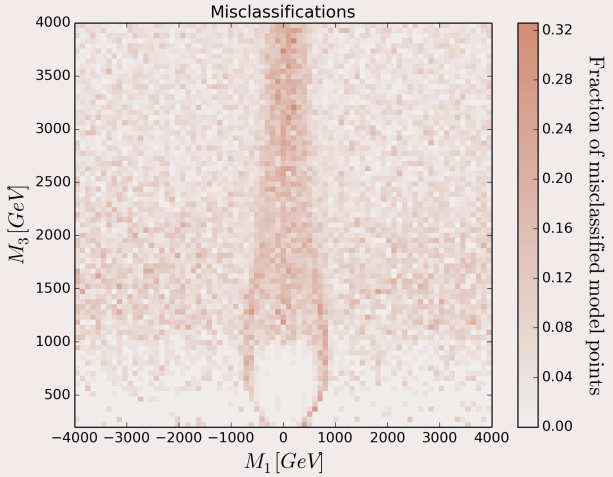

Misclassifications
$m_{\tilde{\chi}_1^0}[GeV]$
$m_{\tilde{g}}[GeV]$
Fraction of misclassified model points

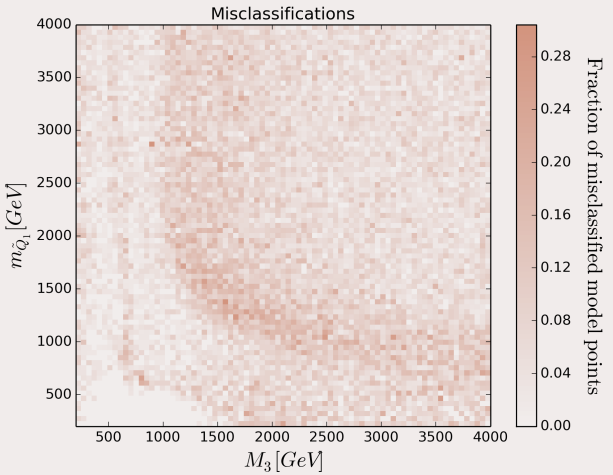

Classification from data
$M_2\,[GeV]$
$M_1\,[GeV]$
Fraction of allowed model points


Classification from data
$M_3\,[GeV]$
$M_1\,[GeV]$
Fraction of allowed model points


Classification from data
$m_{\tilde{Q}_i}\,[GeV]$
$M_3\,[GeV]$
Fraction of allowed model points

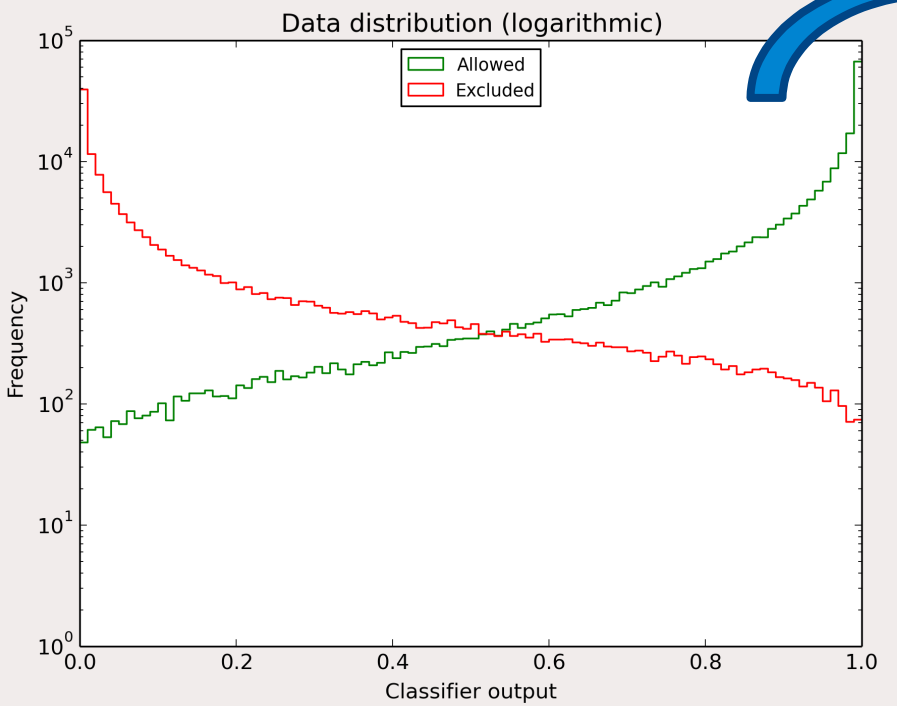

Misclassifications
$M_2\,[GeV]$
$M_1\,[GeV]$
Fraction of misclassified model points

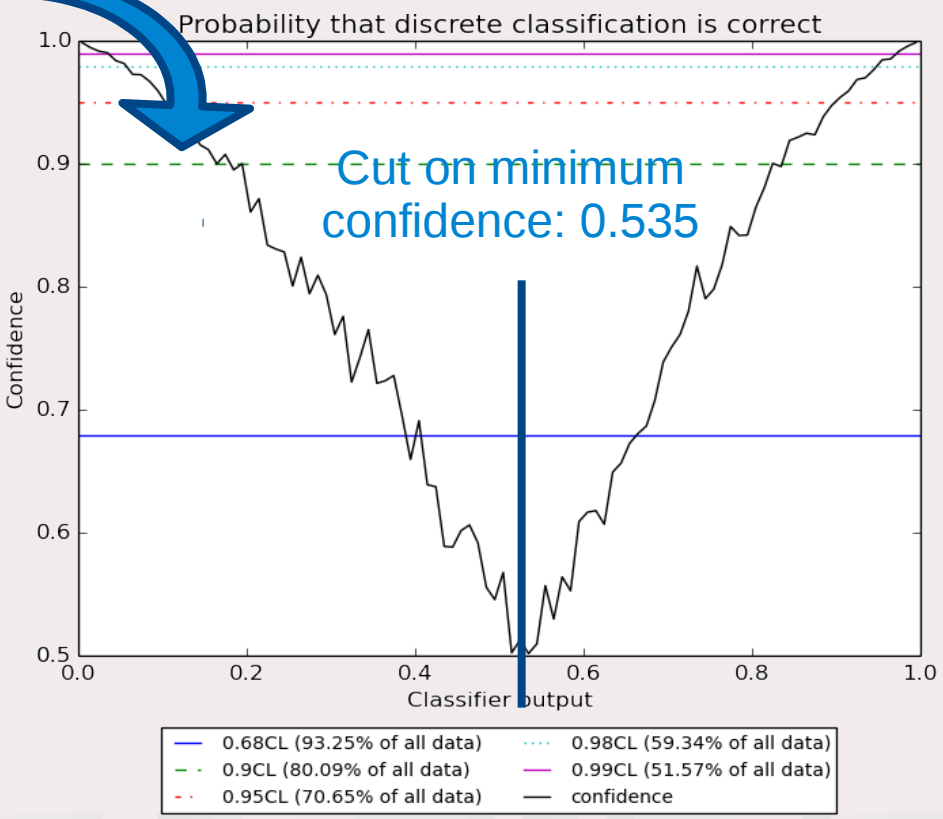

Misclassifications
$M_3\,[GeV]$
$M_1\,[GeV]$
Fraction of misclassified model points


Misclassifications
$m_{\tilde{Q}_i}\,[GeV]$
$M_3\,[GeV]$
Fraction of misclassified model points

Confidence

Ratio of majority class





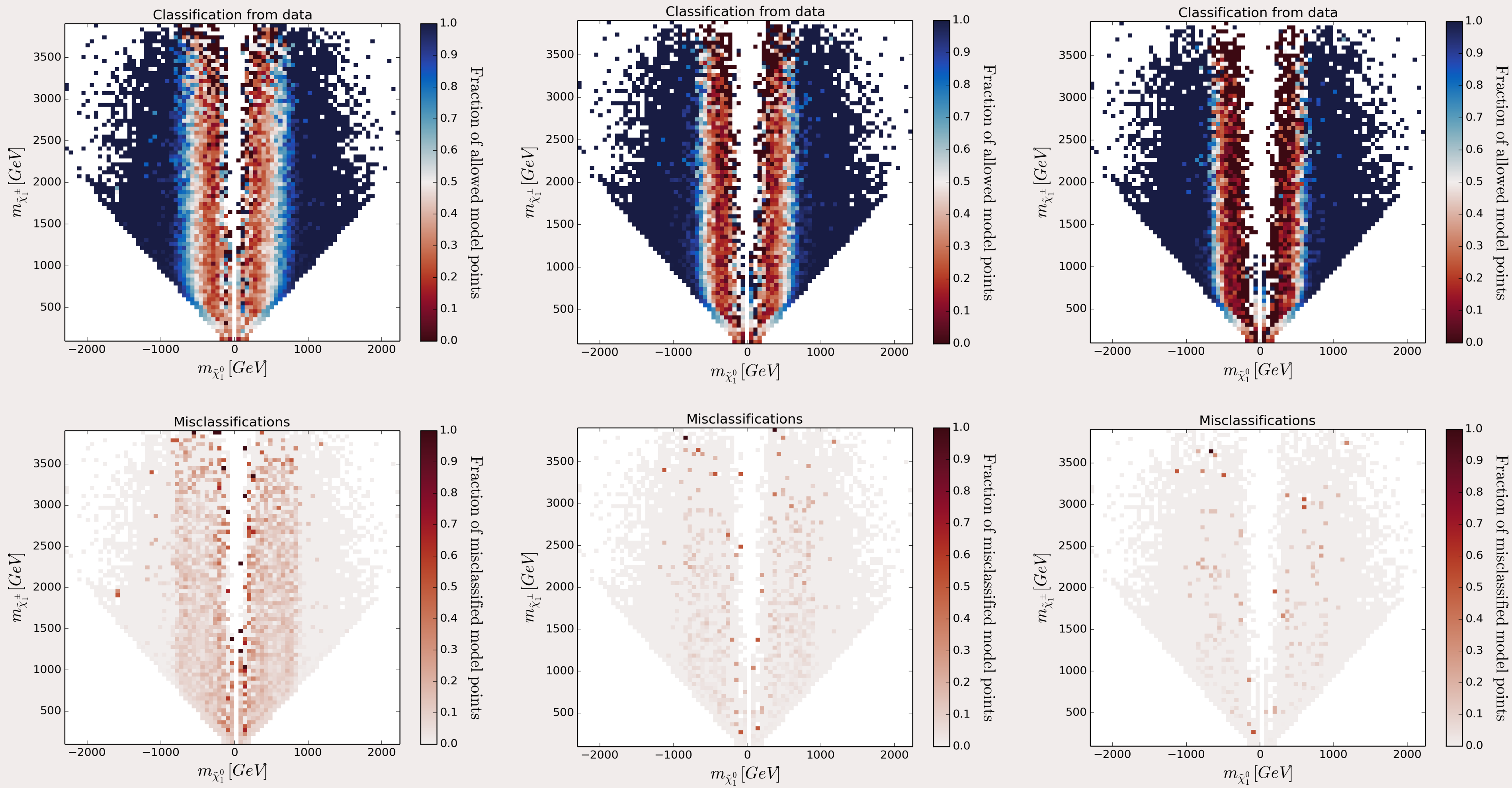
Performance measures

$$\frac{TP}{TP + FP}$$

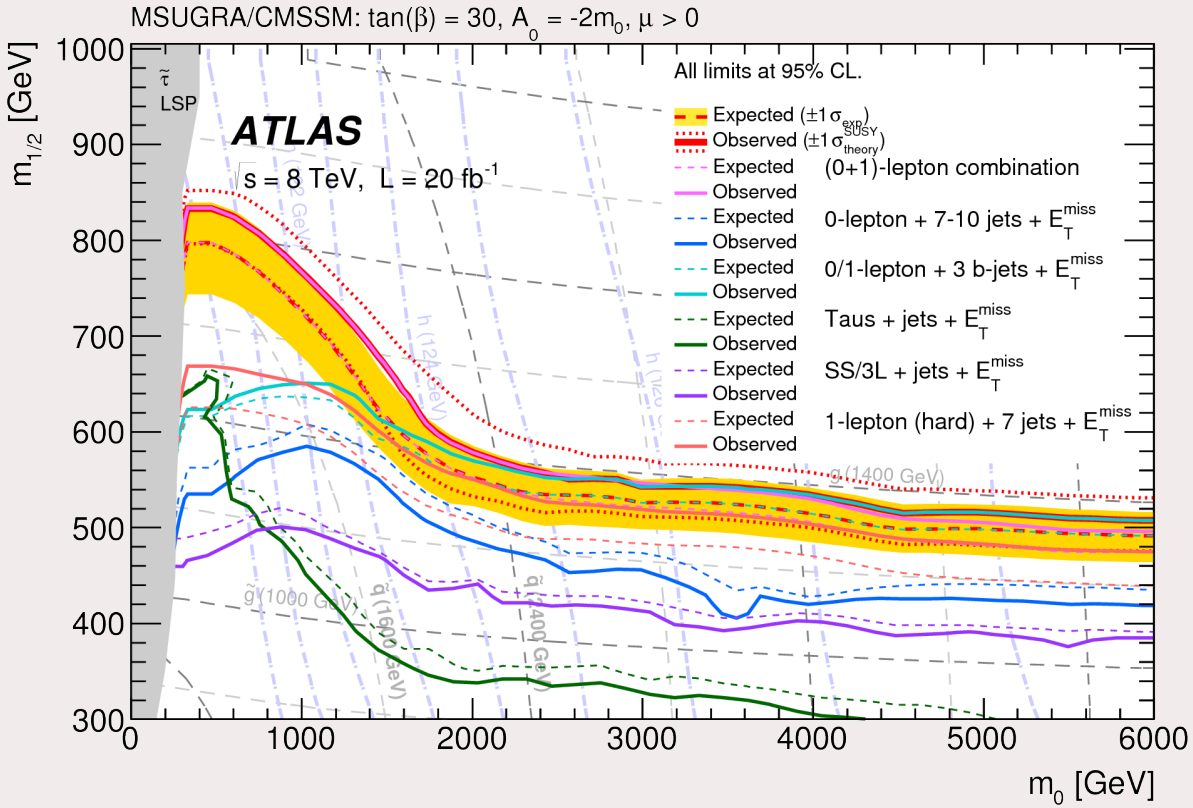
$$\frac{1}{N}\sum_{i=1}^{N}(p_i - t_i)^2$$

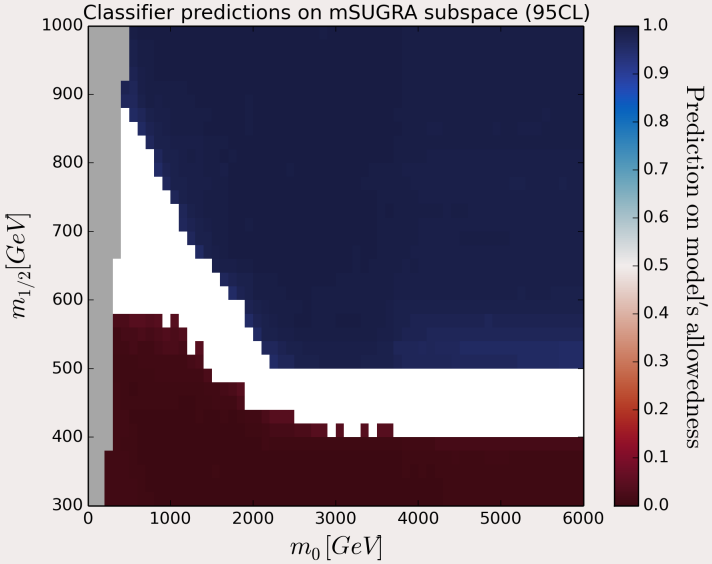
CL	#	Fraction	Accuracy	Precision	Brier
0.0	310324	1.0	0.93226	0.93951	0.04951
0.68	289371	0.93248	0.95735	0.96072	0.03573
0.95	219233	0.70646	0.99094	0.99092	0.00885
0.98	184230	0.59367	0.99543	0.99573	0.00452
0.99	160034	0.5157	0.99708	0.99747	0.00291

$$\frac{TP + TN}{TP + FP + FN + TN}$$

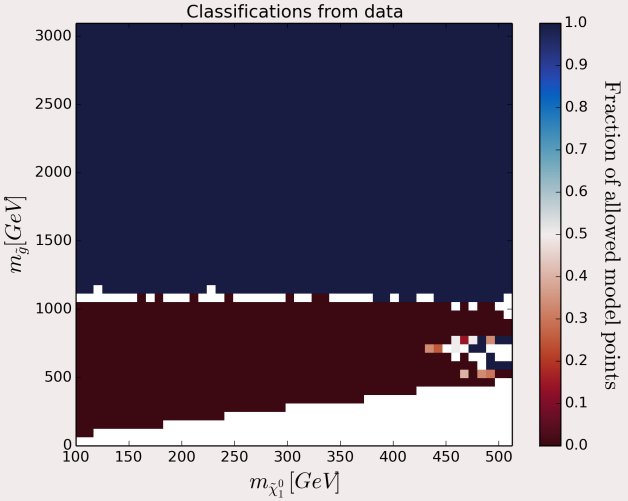

Classification from data
Classification from data
Classification from data
Fraction of allowed model points
Fraction of allowed model points
Fraction of allowed model points
Misclassifications
Misclassifications
Misclassifications
Fraction of misclassified model points
Fraction of misclassified model points
Fraction of misclassified model points
$m_{\tilde{\chi}_1^\pm}\,[GeV]$
$m_{\tilde{\chi}_1^0}\,[GeV]$

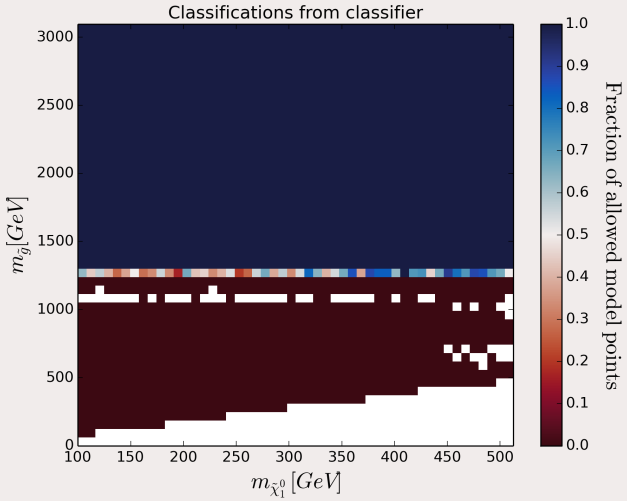
cMSSM

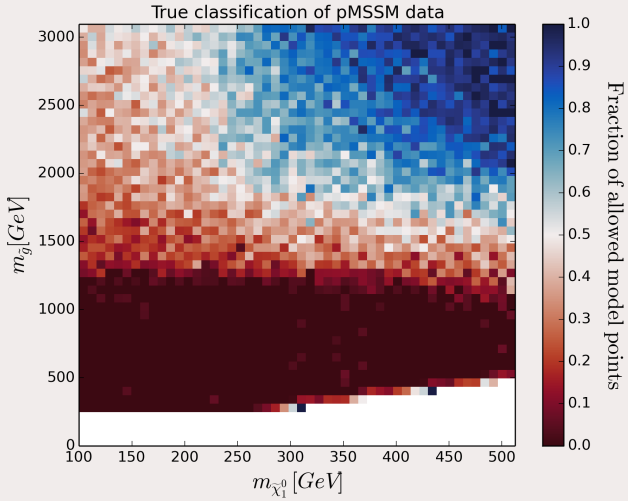

MSUGRA/CMSSM: tan(β) = 30, A0 = -2m0, μ > 0
All limits at 95% CL.
ATLAS
√s = 8 TeV, L = 20 fb-1
τ̃ LSP
Expected (±1σexp)
Observed (±1σtheory)
Expected
Observed
(0+1)-lepton combination
0-lepton + 7-10 jets + E_T^miss
0/1-lepton + 3 b-jets + E_T^miss
Taus + jets + E_T^miss
SS/3L + jets + E_T^miss
1-lepton (hard) + 7 jets + E_T^miss
g̃ (1400 GeV)
g̃ (1000 GeV)
q̃ (1600 GeV)
q̃ (1400 GeV)
m1/2 [GeV]
m0 [GeV]

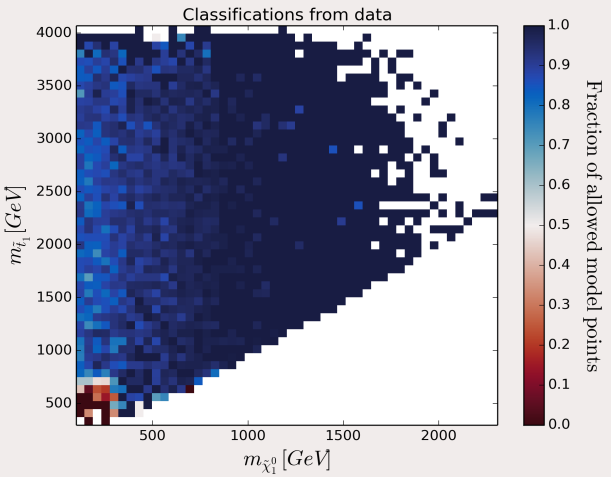

Classifier predictions on mSUGRA subspace (95CL)
m1/2 [GeV]
m0 [GeV]
Prediction on model's allowedness

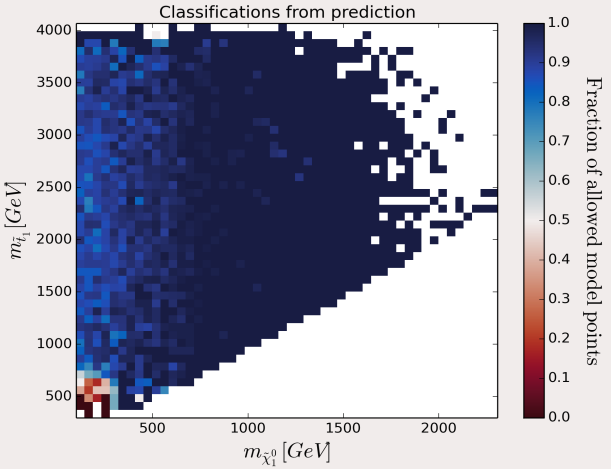
Natural SUSY

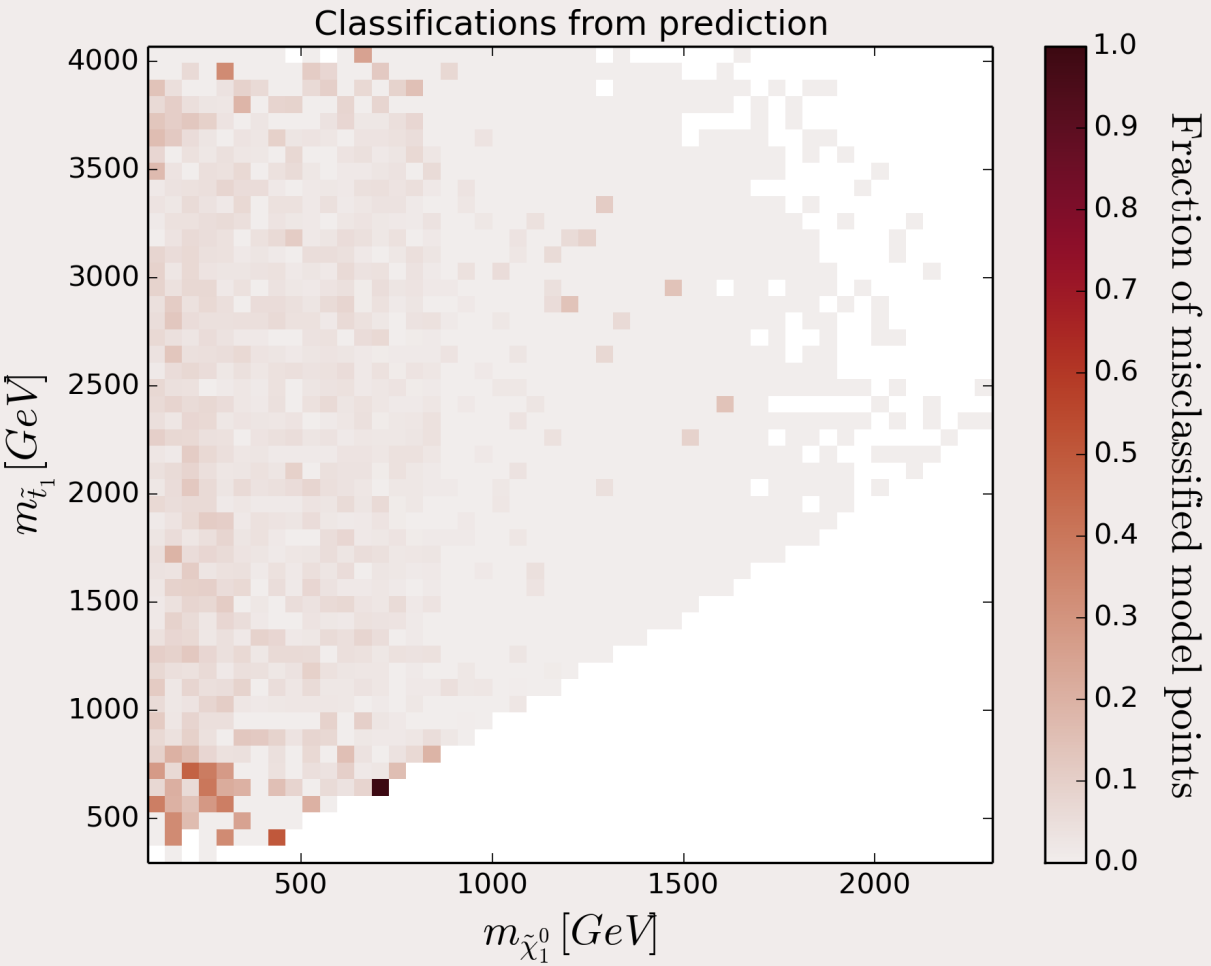


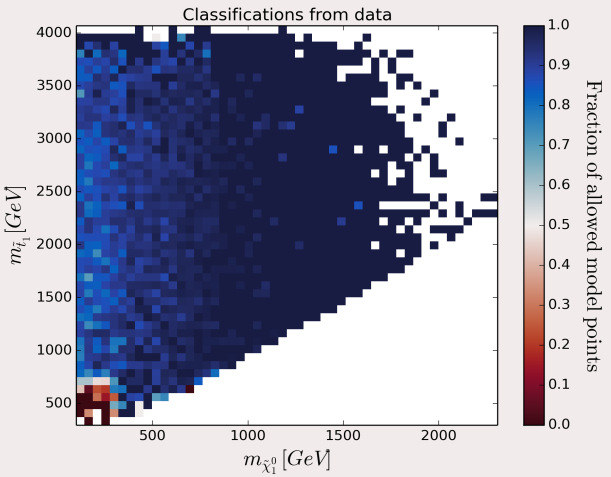


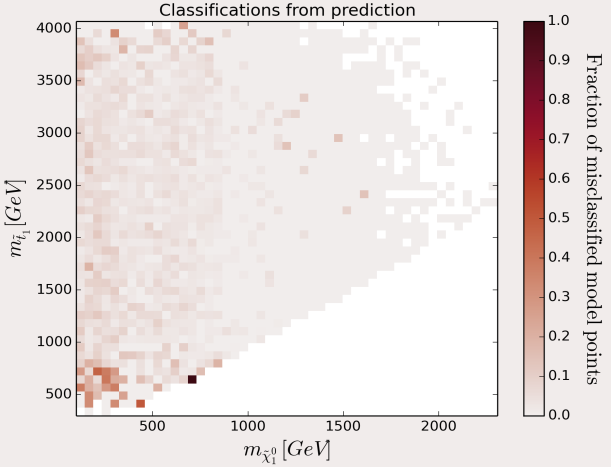


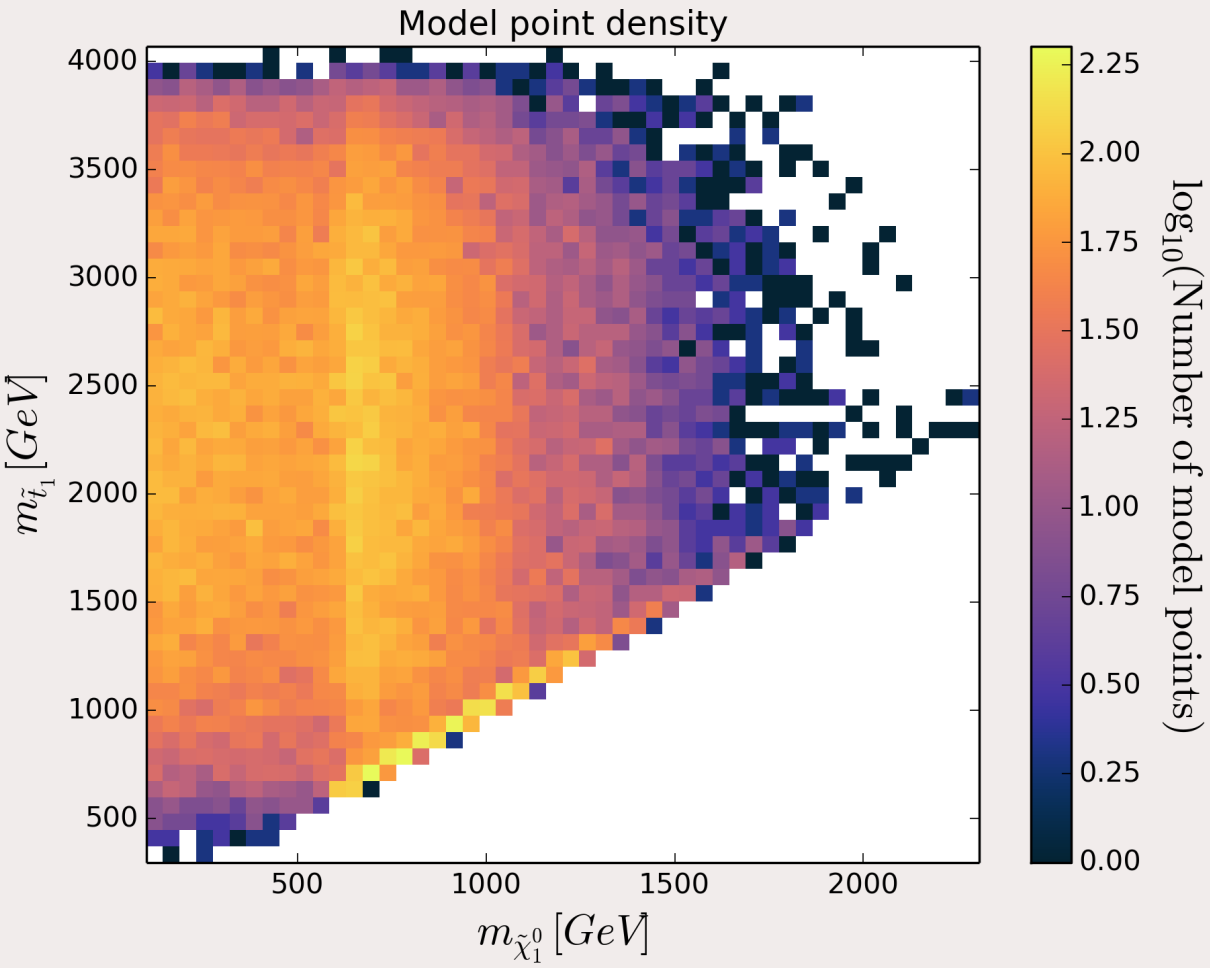

Classifications from data
$m_{\tilde{t}_1}\,[GeV]$
$m_{\tilde{\chi}_1^0}\,[GeV]$
Fraction of allowed model points

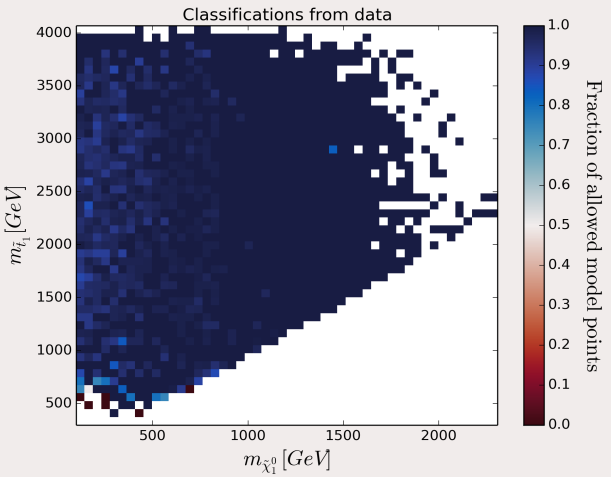

Classifications from prediction
$m_{\tilde{t}_1}\,[GeV]$
$m_{\tilde{\chi}_1^0}\,[GeV]$
Fraction of allowed model points

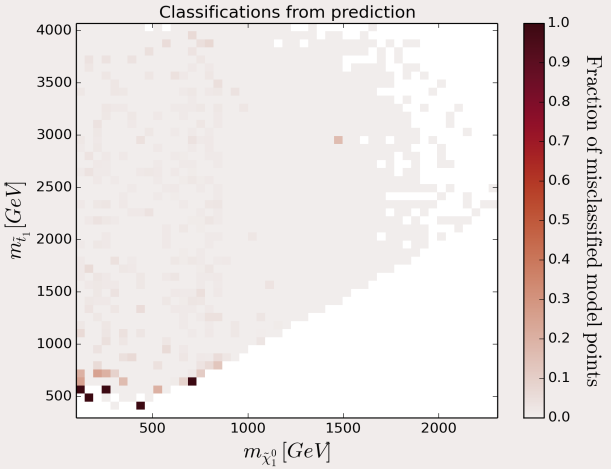

Classifications from prediction
$m_{\tilde{t}_1}\,[GeV]$
$m_{\tilde{\chi}_1^0}\,[GeV]$
Fraction of misclassified model points

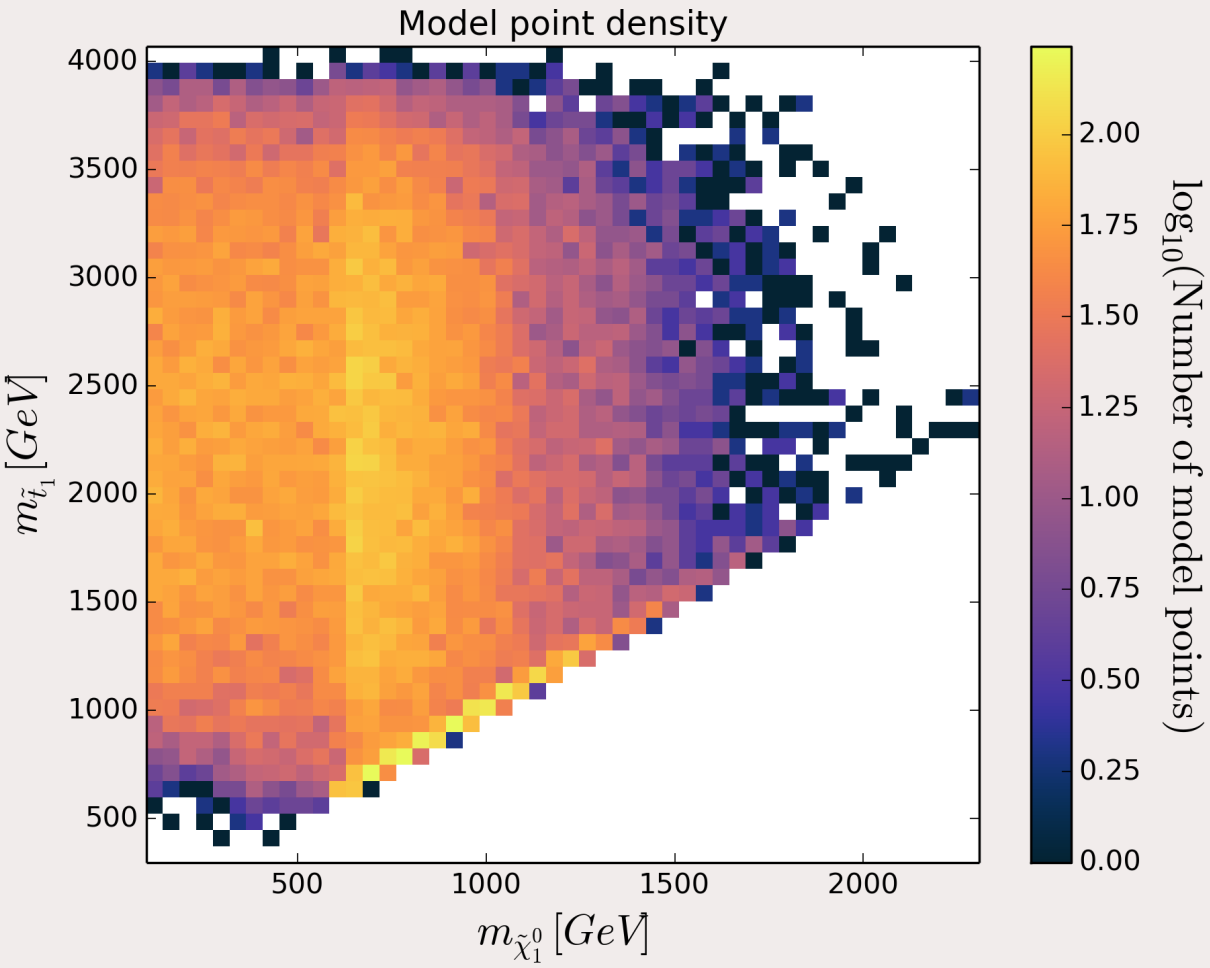

Classifications from data
$m_{\tilde{t}_1}[GeV]$
$m_{\tilde{\chi}_1^0}[GeV]$
Fraction of allowed model points


Classifications from prediction
$m_{\tilde{t}_1}[GeV]$
$m_{\tilde{\chi}_1^0}[GeV]$
Fraction of misclassified model points


Model point density
$m_{\tilde{t}_1}[GeV]$
$m_{\tilde{\chi}_1^0}[GeV]$
$\log_{10}$(Number of model points)


Classifications from data
$m_{\tilde{t}_1}\,[GeV]$
$m_{\tilde{\chi}_1^0}\,[GeV]$
Fraction of allowed model points


Classifications from prediction
$m_{\tilde{t}_1}\,[GeV]$
$m_{\tilde{\chi}_1^0}\,[GeV]$
Fraction of misclassified model points


Model point density
$m_{\tilde{t}_1}\,[GeV]$
$m_{\tilde{\chi}_1^0}\,[GeV]$
$\log_{10}$(Number of model points)

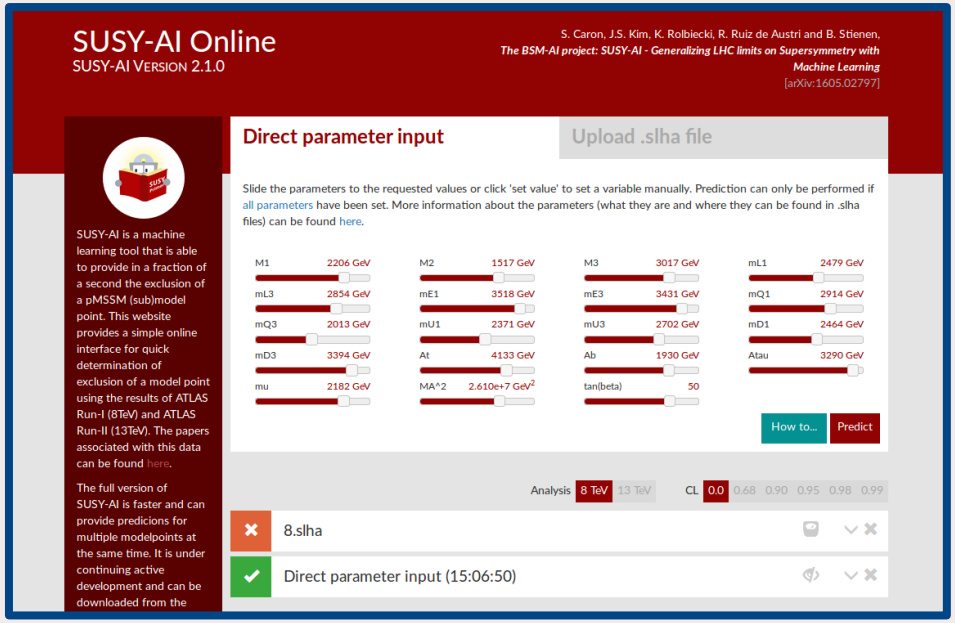
Benchmark tests (10.000 model points)

	Test 1		Test 2	
Coordinate selector	'1'		Function	
ID selector	None		'filename'	
Extra selector	None		3 variables	
	Time (s)	*Duration (s)*	*Time (s)*	*Duration (s)*
Reading files	0	97.177	0	137.044
Loading pickle	97.177	1.856	137.044	1.792
Mapping coordinates	99.033	0.006	138.836	0.007
Predicting	**99.039**	**1.048**	**138.843**	**1.340**
Calculating results	100.087	0.073	140.183	0.079
Apply min CL 0.95	100.160	0.019	140.262	0.048
Total run time	*100.179*		*140.310*	

SUSY-AI (Online)

- Tool has been published https://susyai.hepforge.org/
  - Python interface to classifier
  - Sklearn package for ML implementation

- Online interface http://susy-ai.org/
  - All functionalities except batch predictions
  - Predictions in < 2 seconds

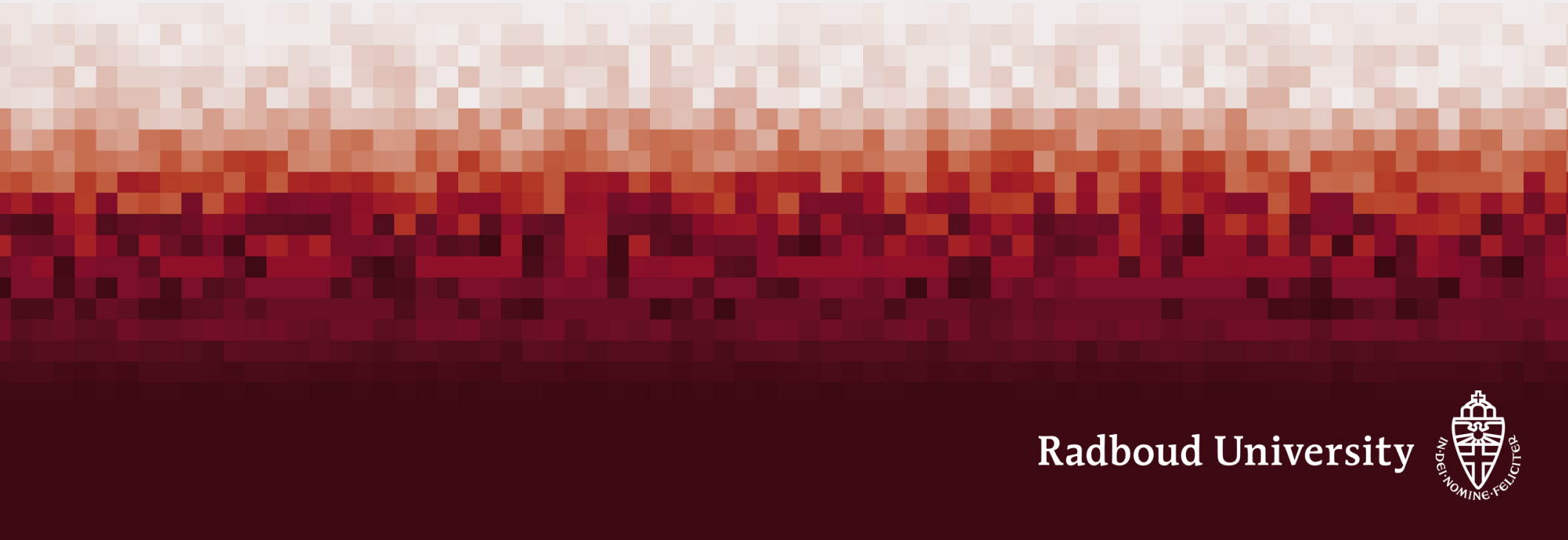


Summary / Conclusions

- ATLAS exclusion limits are predicted with 93.2% accuracy
- Programmatic interface (SUSY-AI: https://susyai.hepforge.org/)
- Online interface (http://susy-ai.org)

- More public data needed to broaden this use of Machine Learning

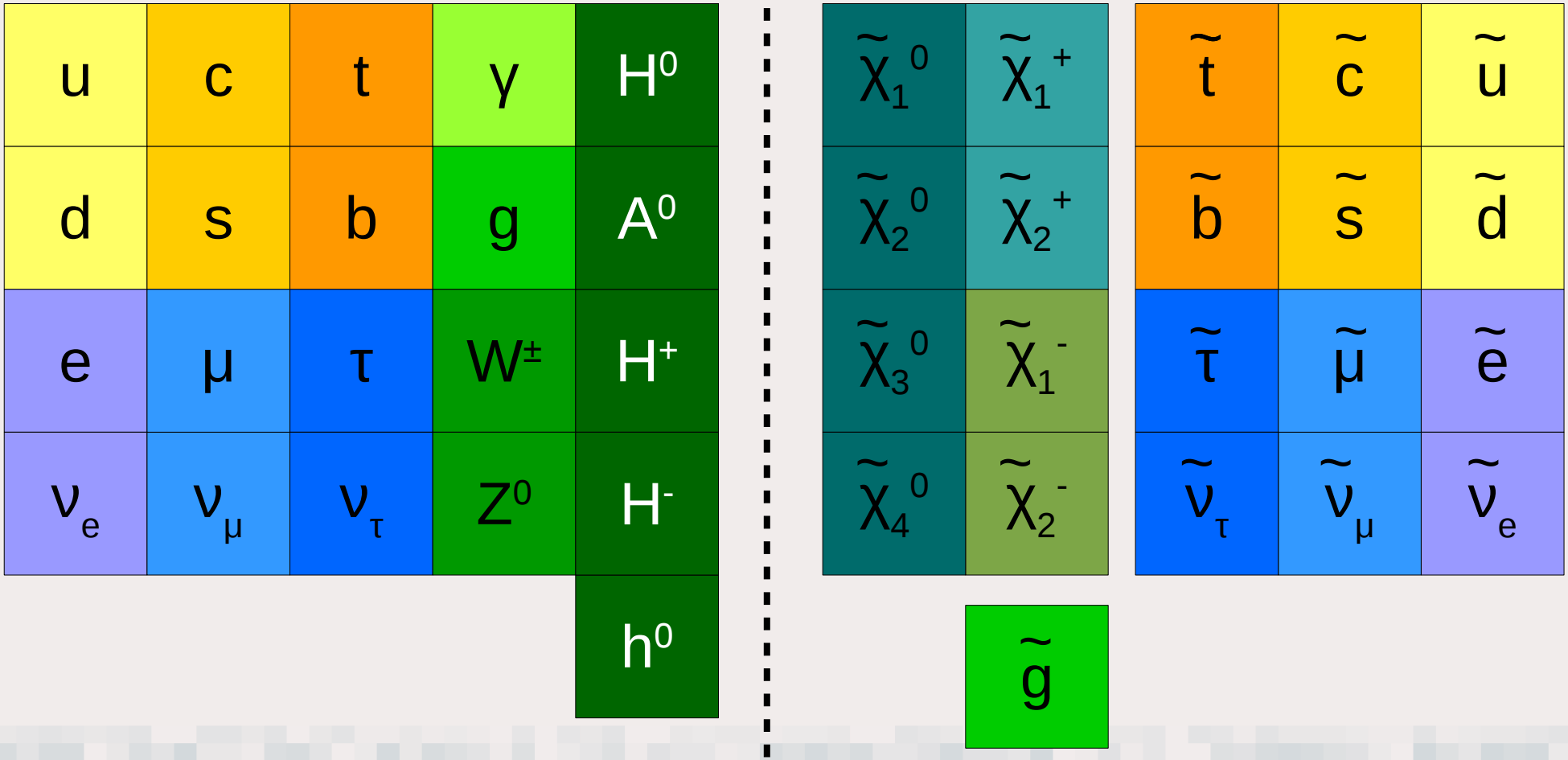
Thank you for your attention!



Radboud University

Backup slides

pMSSM particle content



pMSSM aannames

- Natuur wordt beschreven door R-pariteit bewarende MSSM
- Er worden geen aannames gemaakt op het brekingsmechanisme van supersymmetrie of over een Grand Unified Theory (GUT)
- Er is minimale schending van smaak
- Het lichtste neutralino is het lichtste supersymmetrische deeltje
- De eerste twee sfermion generaties zijn ontaard op hun massa
- De eerste twee generaties hebben verwaarloosbare Yukawa koppelingen

Performance measures

$$\frac{TP}{TP + FP}$$

$$\frac{1}{N}\sum_{i=1}^{N}(p_i - t_i)^2$$

CL	#	Fraction	Accuracy	Precision	Sensitivity	Brier	AUC
0.0	310324	1.0	0.93226	0.93951	0.94665	0.04951	0.98209
0.68	289371	0.93248	0.95735	0.96072	0.96835	0.03573	0.98783
0.95	219233	0.70646	0.99094	0.99092	0.99426	0.00885	0.99618
0.98	184230	0.59367	0.99543	0.99573	0.99672	0.00452	0.99767
0.99	160034	0.5157	0.99708	0.99747	0.99764	0.00291	0.99825

$$\frac{TP + TN}{TP + FP + FN + TN}$$

$$\frac{TP}{TP + FN}$$

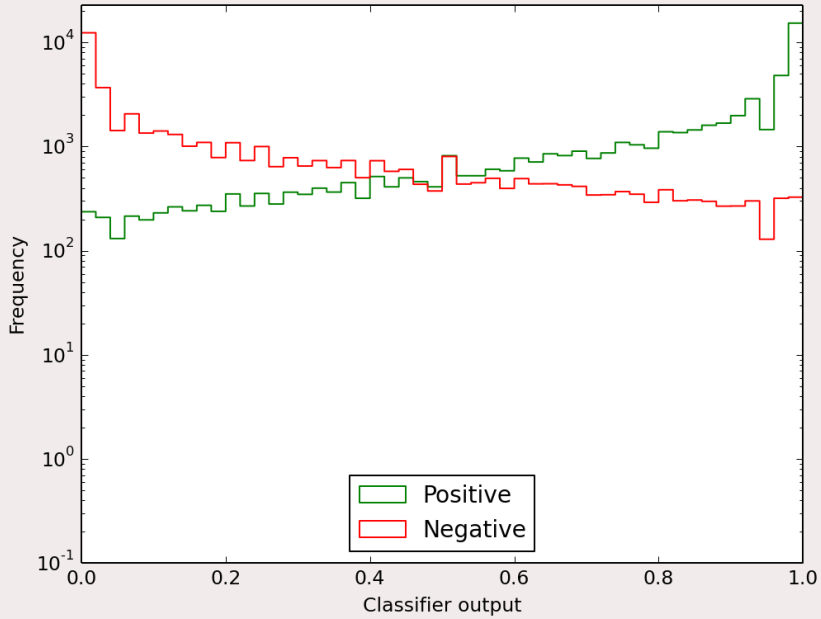
Feature importances

Parameter	Importance
mL1	0.020722329872
me1	0.0188857795865
mL3	0.0138133304874
me3	0.0143471537963
mQ1	0.0791210937176
mu1	0.0662304185267
md1	0.036920220724
mQ3	0.0261421684941
mu3	0.0181801062761
md3	0.0257523635519

Parameter	Importance
M1	0.0576284287751
M2	0.164160591853
mu	0.130095649773
M3	0.241522918816
At	0.0132556452076
Ab	0.0116453579659
Atau	0.0117438282989
mA2	0.0307053005307
tanbeta	0.0191273137473

Tuning of false classifications

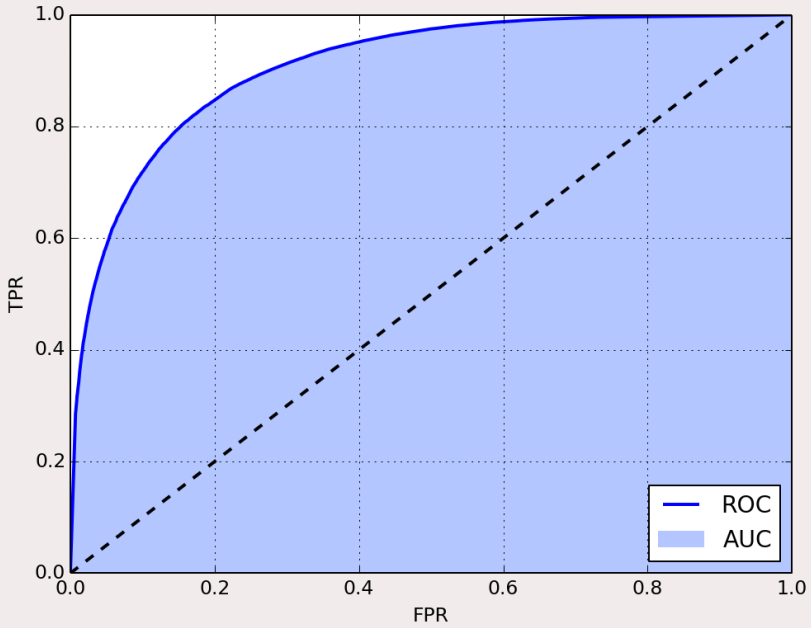
- Apply cut to go from continuous to binary prediction
  - Introduces errors of first and of second kind
- TPR en FPR depend on the location of this cut



$$\text{TPR} \frac{\text{TP}}{\text{TP} + \text{FN}} \qquad \text{FPR} \frac{\text{FP}}{\text{FP} + \text{TN}}$$

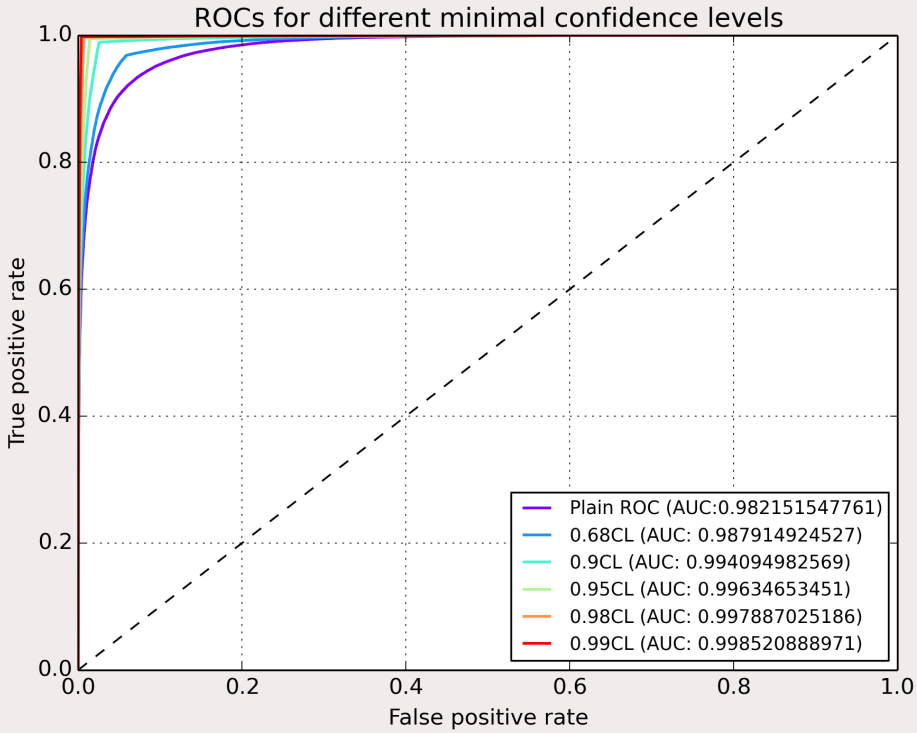
Receiver-Operator Characteristics (ROC)

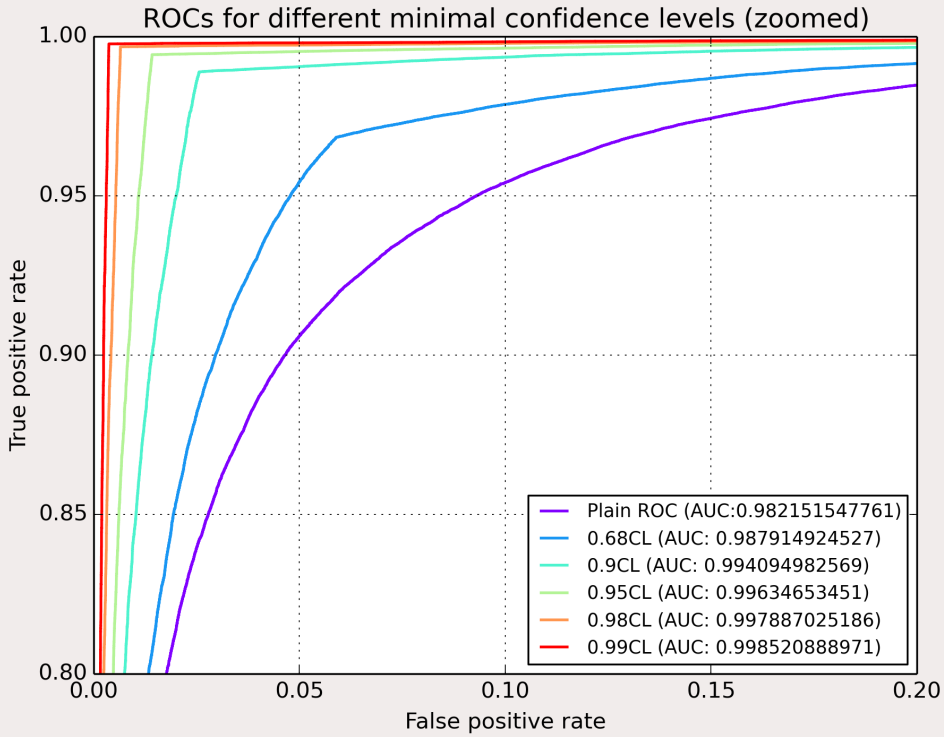
- Apply cut to go from continuous to binary prediction
  - Introduces errors of first and of second kind
- TPR en FPR depend on the location of this cut
- Block-shaped ROC curve means perfect classification
- Area under the curve (AUC) measure for classifier performance



$$\text{TPR} \quad \frac{\text{TP}}{\text{TP} + \text{FN}} \qquad \text{FPR} \quad \frac{\text{FP}}{\text{FP} + \text{TN}}$$

ROC curve


ROCs for different minimal confidence levels
True positive rate
False positive rate
0.0
0.2
0.4
0.6
0.8
1.0
Plain ROC (AUC:0.982151547761)
0.68CL (AUC: 0.987914924527)
0.9CL (AUC: 0.994094982569)
0.95CL (AUC: 0.99634653451)
0.98CL (AUC: 0.997887025186)
0.99CL (AUC: 0.998520888971)


ROCs for different minimal confidence levels (zoomed)
True positive rate
False positive rate
0.80
0.85
0.90
0.95
1.00
0.00
0.05
0.10
0.15
0.20
Plain ROC (AUC:0.982151547761)
0.68CL (AUC: 0.987914924527)
0.9CL (AUC: 0.994094982569)
0.95CL (AUC: 0.99634653451)
0.98CL (AUC: 0.997887025186)
0.99CL (AUC: 0.998520888971)

Pick chance in bagging

$$\mathrm{P(picked)} = 1 - \left(1 - \frac{1}{n}\right)^n$$

$$\lim_{n\to\infty} \mathrm{P(picked)} = \lim_{n\to\infty} 1 - \left(1 - \frac{1}{n}\right)^n = 1 - \frac{1}{e} \approx 0.632$$