Improving Signal Significance of SUSY Compressed Scenarios by ML Algorithms

Jorge Fraga, R. Rodriguez, J. Solano, J. Molano, C. Avila

[arXiv:2106.06813]

Accepted for publication in IJMPA

Universidad de los Andes
Departamento de Física

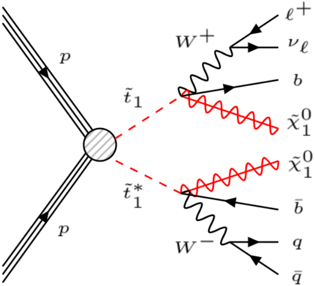
VII UNIANDES PARTICLE PHYSICS SCHOOL

Bogota - Ecuador

December 6th, 2022

Search for SUSY stop particles

- A study of four different machine learning (ML) algorithms is performed to determine the best technique to disentangle a SUSY signal from its SM backgrounds
- We measure the quality of the ML algorithms through their impact on signal significance.



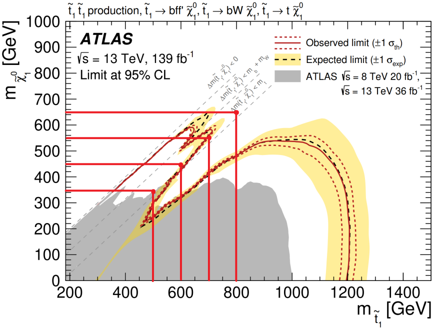
Feynman diagram for stop pair production with three body decay and semileptonic final states (with $\ell = e, \mu$)

Signal samples

We simulate proton-proton collisions with a $\sqrt{s} = 13$ TeV and an integrated luminosity of 140 fb^{-1}

Table: MSSM signal parameters for each benchmark. Signal sample parameters are produced using SoftSusy [3]

Parameters	*s500*	*s600*	*s700*	*s800*
$m(\tilde{t})$ [GeV]	500	600	700	800
$m(\tilde{\chi}_1^0)$ [GeV]	350	450	550	650
σ (NNLO-NLL) [fb]	0.6090	0.2050	0.0783	0.0326
Ωh^2	2.94	5.34	5.97	8.74
m_0 [GeV]	2450	2600	2560	2400
$m_{1/2}$ [GeV]	800	950	1240	1400
A_0 [GeV]	-5950	-6500	-7200	-7300
$\tan\beta$	25	25	35	32



We focus on the production of SUSY stops, along the compressed spectra:

$$m_W < \Delta_m = m_{\tilde{t}_1} - m_{\tilde{\chi}_1^0} < m_t$$

in the mass range of $m_{\tilde{t}_1} \in [500, 800]$ GeV

Backgrounds

Table: LO cross-sections for SM backgrounds at $\sqrt{s} = 13$ TeV.

Background	Sub-processes	$\sigma_{LO} \cdot BR[pb]$
$t\bar{t}$ 1L		220.7
$t\bar{t}$ 2L		40.17
W+jets	$H_T \in [100, 200]$ GeV	51.89
	$H_T \in [200, 400]$ GeV	17.68
	$H_T \in [400, \infty]$ GeV	2.703
Single Top	Wt	13.16
	t-channel (tbj)	15.79
	t-channel (tj)	5.166
	s-channel (tb)	0.8814
$t\bar{t} + V$	$t\bar{t} + W$	0.3455
	$t\bar{t} + Z$	0.5856
Diboson	WW	9.849
	WZ	4.253
	ZZ	3.752

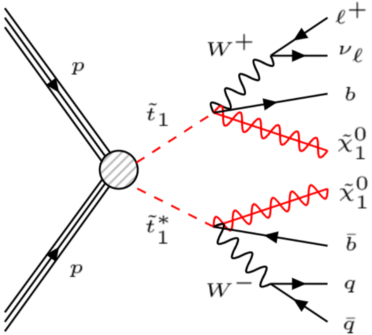
Pre-selection criteria

Table: General pre-selection requirements. These generate an starting point for the comparison.

$N(\ell)$ ($\ell = e, \mu$)	=1 ($p_T > 25$ GeV, $\|\eta\| < 2.5$)
$N(j)$	≥ 4 ($p_T > 25$ GeV, $\|\eta\| < 2.5$)
N(b-tag)	≥ 1
$\min \Delta\phi(j_{1,2}, \vec{p}_T^{\,miss})$	> 0.4
m_T^{lep}	> 90 GeV
E_T^{miss}	> 90 GeV

where the transverse mass of the lepton is defined as:

$$m_T^{lep} = \sqrt{2\, p_T^{\ell} E_T^{miss} [1 - \cos\Delta\phi(p_T^{miss}, p_T^{\ell})]}$$



Statistical Significance

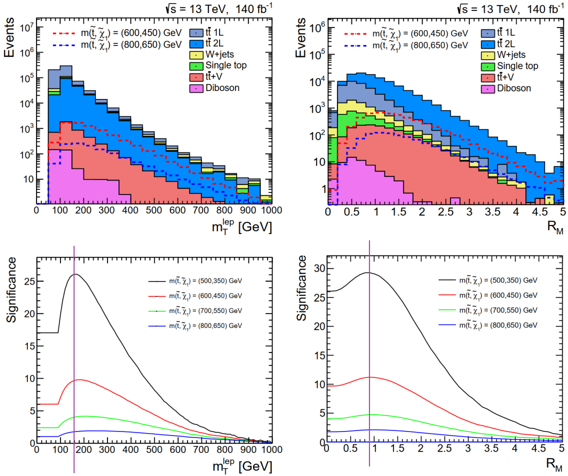
- The significance of signal events S standing on top of a background B, which has an uncertainty σ_B is

$$\text{Sig} = \sqrt{2}\left\{(S+B)\ln\left[\frac{(S+B)(B+\sigma_B^2)}{B^2+(S+B)\sigma_B^2}\right]-\frac{B^2}{\sigma_B^2}\ln\left[1+\frac{\sigma_B^2 S}{B(B+\sigma_B^2)}\right]\right\}^{1/2}$$

  for the case of S and B having Gaussian distributions [4].
- This offers a more conservative approach w.r.t. the naive definition

$$Sig = \frac{S}{\sqrt{S+B}}$$

Optimized selections for the cut-and-count method



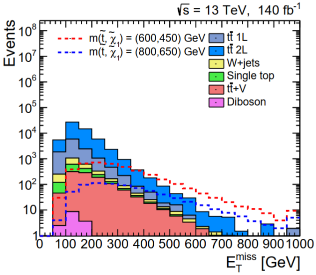
$$m_T^{lep} > 160 \text{ GeV} \qquad R_M = E_T^{miss}/p_T^{j_0} > 0.8$$

$$\Delta\phi(p_T^{\text{b-jet}}, \vec{p}_T^{\,miss}) > 0.4$$

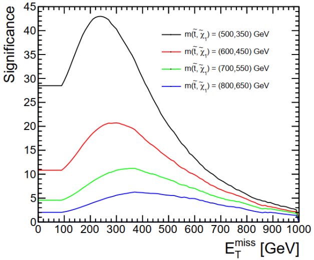
Significance for the cut-and-count method

Preselection +

m_T^{lep}	> 160 GeV
E_T^{miss}/p_T^{j0}	> 0.8
$\Delta\phi(p_T^{\text{b-jet}}, \vec{p}_T^{miss})$	> 0.4



	E_T^{miss} cut	Sig
s500	240	43.0
s600	300	20.7
s700	360	11.2
s800	370	6.27



ML algorithms

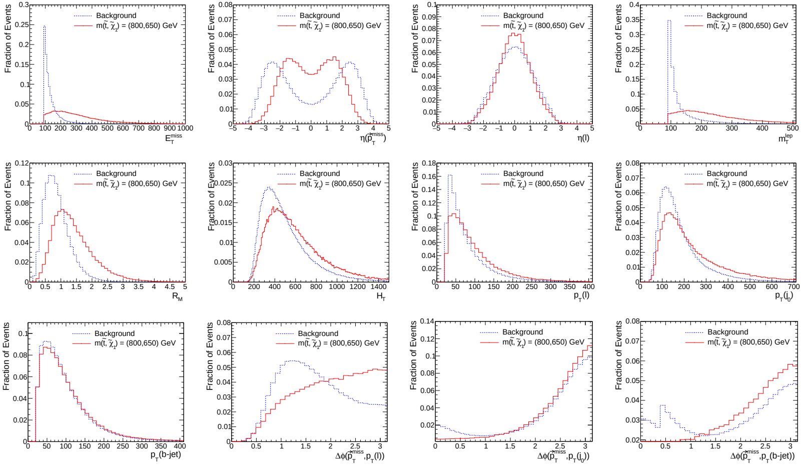
For the classification task, we train and evaluate a set of four ML algorithms:

- LR A **logistic regression** [5] that is a classifier based on regression analysis using a logistic model
- RF The **Random Forest** [6] is a bagging ensemble of decision trees
- XG The Optimized Gradient Boosting Classifier **XGBoost** [7] is a meta-model that iterates over a set of weak classifiers.
- NN A feed-forward **Neural Network** [8] to build a decision boundary after stacking several hidden layers of non-linear relationships.

Table: Physical variables used for training the ML classifiers.

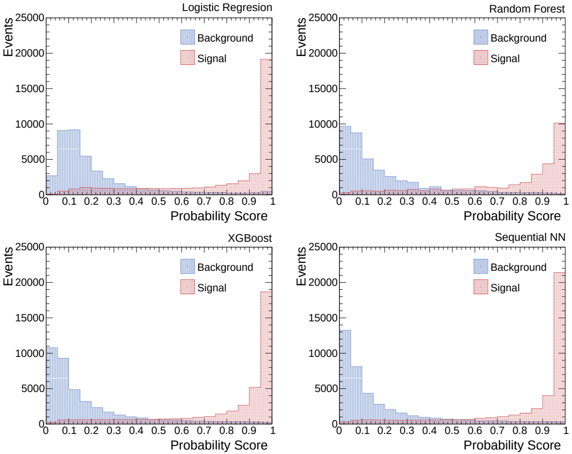
Input Variable
E_T^{miss}
$\eta(\ell)$
$\eta(\vec{p}_T^{\,miss})$
H_T
$p_T(\ell)$
$p_T(j_0)$
p_T(b-jet)
$\Delta\phi(\vec{p}_T^{\,miss}, p_T(\ell))$
$\Delta\phi(\vec{p}_T^{\,miss}, p_T(j_0))$
$\Delta\phi(\vec{p}_T^{\,miss}, p_T(b\text{-jet}))$
m_T^{lep}
R_M

Variables for ML training for the $s800$ benchmark



Testing for the ML algorithms

- We use a binary classifier that categorizes between signal and background classes.
- Probability score distributions for the $m(\tilde{t}, \tilde{\chi}) = (800, 650)$ GeV



Evaluation for ML algorithms

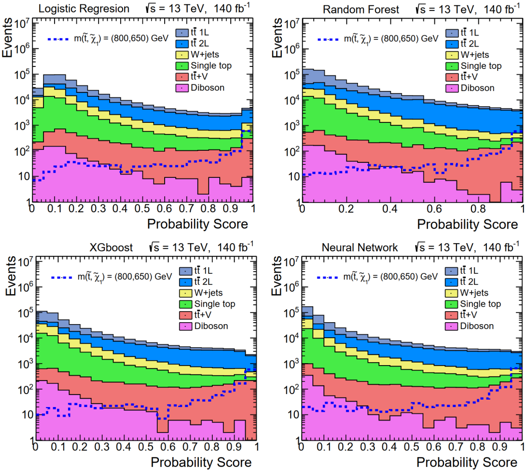
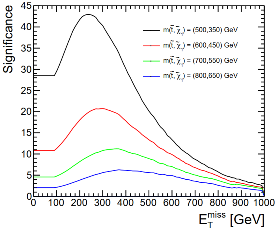
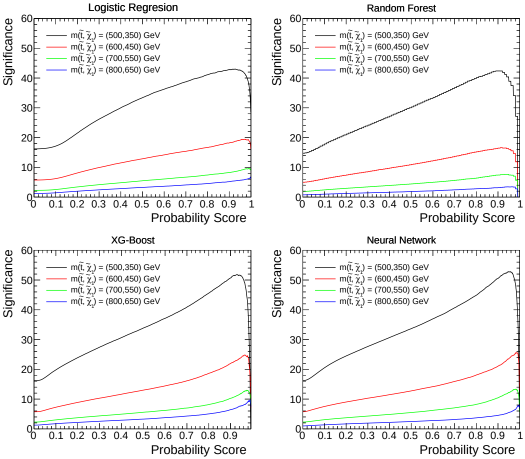


Figure: Probability score distributions for the $m(\tilde{t}, \tilde{\chi}) = (800, 650)$ GeV signal.

Significance comparison



Significance for cut-and-count method



Significance for ML algorithms

Gains in significance for ML algorithms

- Gain in significance of the ML algorithms with respect to the cut-and-count methodology.

Signal benchmark	Cut-and-Count	LR		RF		XG		NN	
	Sig	Sig	Gain	Sig	Gain	Sig	Gain	Sig	Gain
s500	43.0	42.7	-1%	50.0	16%	51.0	19%	51.9	21%
s600	20.7	19.3	-7%	23.1	12%	24.2	17%	25.2	22%
s700	11.2	9.48	-15%	12.1	8%	13.2	18%	13.4	20%
s800	6.27	5.24	-16%	6.66	6%	7.34	17%	8.18	31%

- RF, XG and NN algorithms show a better performance in separating the SUSY signals from backgrounds, when compared to the cut-and-count methodology.
- LR algorithm shows in average a poorer performance getting worse for higher masses.

Conclusions

- XG and NN algorithms show the best performance in separating the SUSY signals from backgrounds, when compared to the cut-and-count method.
- Note that XGBoost is much simpler to implement than the NN, as less hyper-parameters need to be optimized.
- High potential of the ML algorithms as efficient alternatives and consistent tools for searching physics BSM.
- This study has been accepted for publication in Int. Journal. Mod Phys A. [arXiv:2106.06813]

Thanks for
your attention

References

The ATLAS collaboration, *Search for new phenomena with top quark pairs in final states with one lepton, jets, and missing transverse momentum in pp collisions at* $\sqrt{s}$= *13 TeV with the ATLAS detector,* JHEP 04 (2021) 174 [arXiv:2012.03799 [hep-ex]]

The CMS collaboration, *Search for direct top squark pair production in events with one lepton, jets, and missing transverse momentum at 13 TeV with the CMS experiment,* JHEP, 05, (2020) [arXiv:1912.08887 [hep-ex]]

B. C. Allanach, *SOFTSUSY: a program for calculating supersymmetric spectra,* Comput. Phys. Commun., 143, (2002) [arXiv:hep-ph/0104145]

Cowan, G., Cranmer, K., Gross, E. et al. *Asymptotic formulae for likelihood-based tests of new physics*. Eur. Phys. J. C 71, 1554 (2011).

F. Pedregosa et al, *Scikit-learn: Machine Learning in Python,* [arXiv:1201.0490 [cs.LG]]

T. Oshiro, P. Perez, J. Baranauskas, *How Many Trees in a Random Forest?,* International Workshop on Machine Learning and Data Mining in Pattern Recognition, (2012)

L. Mason , J. Baxter , P. Bartlett , M. Frean, *Boosting Algorithms as Gradient Descent,* International Conference on Neural Information Processing, (1999)

P. Baldi, K. Bauer, C. Eng, P. Sadowski, D. Whiteson, *Jet Substructure Classification in High-Energy Physics with Deep Neural Networks,* Phys. Rev. D, **9**, (2016) [arXiv:1603.09349 [hep-ex]]

Backup Slides

Search for $\tilde{t}\tilde{t}^*$ in ATLAS and CMS @ 13TeV

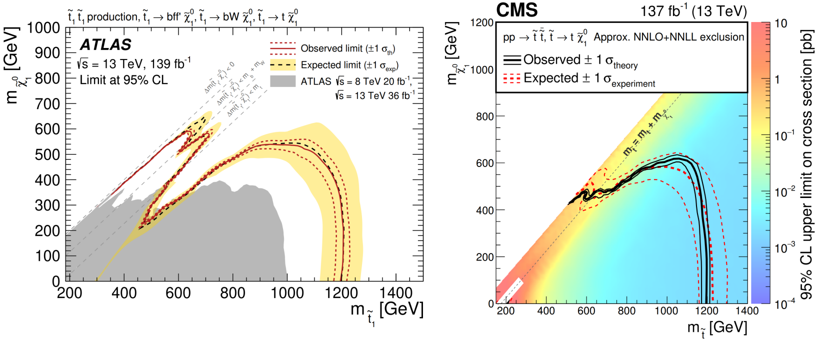


Figure: Exclusion limits for neutralinos ($\tilde{\chi}_1^0$) and stop ($\tilde{t}$) for ATLAS [1] and CMS [2] experiments.

Optimization of m_T^{lep} and E_T^{miss}

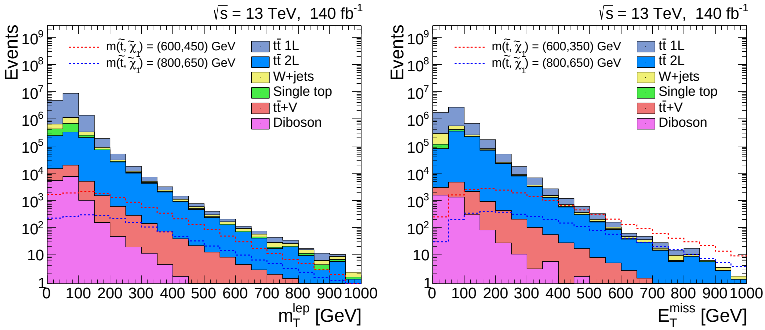


Figure: Distributions for transverse mass m_T^{lep} (left panel) before the 90 GeV selection and E_T^{miss} (right panel) after the $m_T^{lep} > 90$ GeV selection. Two benchmark signals are displayed for comparison.

SIgnal and Background yields after baseline selections

	Signal			
	s500	*s600*	*s700*	*s800*
Initial Events	85260	28700	10962	4564
$N(\ell = e, \mu) = 1,$	53691	17961	6780	2790
$N(j) \geq 4$	43076	14548	5534	2281
$N(\text{b-tag}) \geq 1$	37431	12641	4812	1982
$\min \Delta\phi(j_{1,2}, \vec{p}_T^{\,miss}) > 0.4$	34319	11637	4441	1835
$m_T^{lep} > 90$ GeV	22845	7904	3104	1315
$E_T^{miss} > 90$ GeV	20483	7203	2867	1225

	Backgrounds						
	$t\bar{t}$ 1L	$t\bar{t}$ 2L	$W+$ jets	Single Top	$t\bar{t}+V$	Diboson	Total
Initial Events	45437200	8269769	36395000	5649060	190500	3566700	99508229
$N(\ell = e, \mu) = 1,$	29025400	3737550	23455900	3924690	54862	1609600	61808002
$N(j) \geq 4$	18117600	1270150	2765200	911986	44499	77820	23187255
$N(\text{b-tag}) \geq 1$	15729000	1112230	942017	743815	38818	18444	18584324
$\min \Delta\phi(j_{1,2}, \vec{p}_T^{\,miss}) > 0.4$	14000300	989968	808791	657935	34318	16262	16507574
$m_T^{lep} > 90$ GeV	2064180	370894	153778	98123	8281	2072	2697328
$E_T^{miss} > 90$ GeV	389030	193888	20444	18842	4023	349	626576

Optimization of number of events

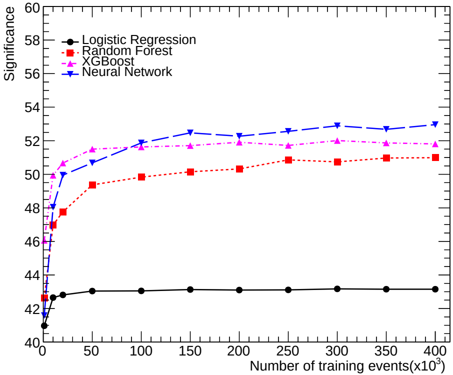


Figure: Significance as a function of the number of events in the training set for the *s500* signal mass point and the four ML algorithms.

ML algorithms: LR

- The Logistic Regression we apply a logit regression model with an optimized solver that uses a coordinate descent algorithm with regularization,
- This adds a squared penalization to model coefficients.
- We use regularization in order to reduce the over-fitting and improve numerical stability during the training phase.
- Other parameters like the tolerance and the inverse of regularization strength were optimized with no improvement in significance.

ML algorithms: RF

- For the Random Forest classifier, we propose a forest with 50 trees, and a Gini Impurity metric for splits and bootstrap.
- The number of trees was optimized using a grid-search over the number of estimators in the model in the range of [50,100] trees.
- Furthermore, we use the Gini impurity metric because it is less intensive than other split criteria when splitting nodes in decision trees.

ML algorithms: XG

- For the Optimized Gradient Boosting, we use an ensemble of 100 trees boosted using the tree booster (*gbtree*).
- We chose a tree booster because we are iterating over weak decision trees.
- The number of the trees in this model was selected after a grid-search in the range of [50,500] trees, where the significance was optimized. In that sense, each decision tree has 100 estimators.
- Additionally, the η parameter has been optimized in the range [0.01,0.30], getting a value of 0.30. Note that a XGBoost is also an ensemble classifier like the Random Forest.
- However, the XGBoost is based on boosting weak learners (shallow decision trees) that use gradient descent algorithms over the error metric. On the other hand, the random forest is a bagging of fully grown decision trees which are ensembled.

ML algorithms: NN

- For the NN we propose a fully-connected neural network with *sigmoid* activation in the last layer and rectified linear units (*ReLu*) activation in previous layers. We optimized the activation for the intermediate layers between *tanh* and *sigmoid*.
- We performed a hyper-parameter optimization on the learning rate, number of layers and nodes per layer. The learning rate was optimized in the range $[10^{-7}, 10^{-2}]$, obtaining an optimal value of 10^{-4}. Layers were optimized in the range [3,10] obtaining an optimal value of 4 layers. Nodes for the three intermediate *ReLu* layers were optimized in the range [4,128] obtaining optimal values of 32, 16 and 8 nodes for the first, second and third layers, respectively.
- We normalized the feature distributions fed to the neural network through a Quantile transformer algorithm with gaussian output.

Event yields for the cut-and-count method

Table: Maximal significance expected for each signal benchmark point after applying the indicated E_T^{miss} requirement. Signal and background yields are shown for statistical reference.

	E_T^{miss} cut	Signal Events	Backgrounds $t\bar{t}$ 1L	$t\bar{t}$ 2L	W+jets	ST	$t\bar{t}$+V	VV	Total	Sig	S/B
s500	240	5591	627	3769	89	46	269	2	4802	43.0	1.2
s600	300	1530	210	1270	40	19	133	1	1673	20.7	0.91
s700	360	496	67	462	18	8	65	1	621	11.2	0.79
s800	370	239	53	391	16	7	58	1	526	6.27	0.45

Event yields for NN algorithm

Table: Maximal significance expected for each signal benchmark point after applying the respective probability score threshold for the NN algorithm. Signal and detailed background yields are shown for statistical reference.

	Threshold	Signal	Backgrounds								
	cut	Events	$t\bar{t}$ 1L	$t\bar{t}$ 2L	W+jets	ST	$t\bar{t}$+V	VV	Total	Sig	S/B
s500	0.96	7745	740	4563	310	162	351	1	6127	51.9	1.3
s600	0.97	2308	291	1894	128	74	206	0	2593	25.2	0.89
s700	0.98	800	121	841	57	30	126	0	1175	13.4	0.68
s800	0.99	351	75	431	37	19	80	0	642	8.18	0.55

Event yields for XGboost algorithm

Table: Maximal significance expected for each signal benchmark point after applying the respective probability score threshold for the XG algorithm. Signal and detailed background yields are shown for statistical reference.

	Threshold cut	Signal Events	Backgrounds							Sig	S/B
			$t\bar{t}$ 1L	$t\bar{t}$ 2L	W+jets	ST	$t\bar{t}$+V	VV	Total		
s500	0.93	7724	669	4988	262	147	372	1	6439	51.0	1.2
s600	0.97	1900	183	1338	88	41	161	0	1811	24.2	1.0
s700	0.98	687	91	621	39	24	87	0	862	13.2	0.80
s800	0.98	380	109	708	51	27	103	0	998	7.34	0.38

Gains in significance for ML algorithms

Table: Gain in significance of the ML algorithms with respect to the cut-and-count methodology. Accuracy in the estimate of signal significance is also provided.

Signal	Cut and Count	LR		RF		XG		NN	
benchmark	Sig	Sig	Gain	Sig	Gain	Sig	Gain	Sig	Gain
s500	43.0±0.52	42.7±0.45	-1%	50.0±0.50	16%	51.0±0.52	19%	51.9±0.53	21%
s600	20.7±0.48	19.3±0.39	-7%	23.1±0.42	12%	24.2±0.50	17%	25.2±0.51	22%
s700	11.2±0.47	9.48±0.32	-15%	12.1±0.40	8%	13.2±0.47	18%	13.4±0.43	20%
s800	6.27±0.39	5.24±0.25	-16%	6.66±0.31	6%	7.34±0.36	17%	8.18±0.45	31%