

ОПТИМИЗАЦИЯ ОТБОРА СОБЫТИЙ ДЛЯ АНАЛИЗА $ZZ \rightarrow \ell\nu\nu$ НА ДАННЫХ ЭКСПЕРИМЕНТА ATLAS С $\sqrt{S} = 13$ ТэВ

Зубов Д.В.

НИЯУ МИФИ

Научный руководитель: Солдатов Е.Ю.

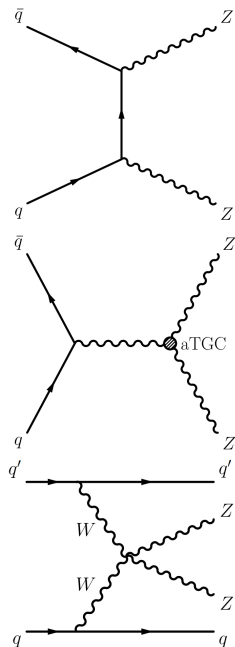
Консультанты: Петухов А.М., Пятиизбянцева Д.Н.

Москва, 2022

Введение

Актуальность и мотивация:

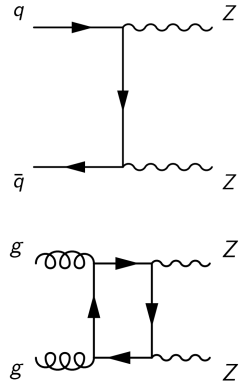
- ▶ Прецизионное измерение двухбозонных процессов один из способов проверки Стандартной модели и пертурбативной КХД на масштабе энергий несколько ТэВ
- ▶ Измерение $aTGC$ и $aQGC$ является косвенным поиском новой физики
- ▶ Многие расширения СМ предсказывают новые скалярные, векторные или тензорные резонансы, которые могут распадаться на пары электрослабых бозонов.
- ▶ Векторное рассеяние бозонов (VBS) - ключевой процесс для исследования механизма нарушения электрослабой симметрии.



Инклюзивный процесс $ZZ \rightarrow ll\nu\nu$

- ▶ В событии два разноименно-заряженных лептона одного аромата (e^+e^- или $\mu^+\mu^-$), при этом, поперечный импульс первого больше 30 ГэВ, второго больше 20 ГэВ;
- ▶ Вето на третий заряженный лептон;
- ▶ $76 \text{ ГэВ} < M_{ll} < 106 \text{ ГэВ}$;
- ▶ $E_T^{\text{miss}} > 70 \text{ ГэВ}$.

Сигнал	
QCD ZZ	КХД рождение двух Z-бозонов и последующий распад в $ll\nu\nu$
EWK ZZ	Электрослабое рождение двух Z-бозонов и последующий распад в $ll\nu\nu$
Фон	
Zj	рождение Z-бозона и струи, с распадом Z-бозона в пару заряженных лептонов и большим ложным потеряннным поперечным импульсом
tt	рождение пары топ-кварков и последующим распадом включающим конечное состояние $ll\nu\nu$ (не резонансное рождение $ll\nu\nu$)
WZ	рождение пары бозонов Z и W, с распадом Z-бозона в пару заряженных лептонов и лептонным распадом W
WW	рождение пары W с распадом в $ll\nu\nu$ (не резонансное рождение $ll\nu\nu$)
Wt	рождение W и топ-кварка и распадом в конечное состояние, содержащее $ll\nu\nu$ (не резонансное рождение $ll\nu\nu$)
VVV	рождение трех векторных бозонов ($V = W$ или Z)
Other (ttV, ttVV)	рождение пары топ-кварков и одного или двух векторных бозонов



Разделение сигнальных и фоновых событий с помощью алгоритмов машинного обучения

- ▶ Машинное обучение применялось в целях максимального разделения сигнальных и фоновых событий и более точной оценки числа сигнальных событий
- ▶ Более точная оценка сигнальных событий позволит измерить сечение с меньшей погрешностью
- ▶ В качестве алгоритма был выбран BDTG - алгоритм на основе леса деревьев решений использующий градиентный спуск.

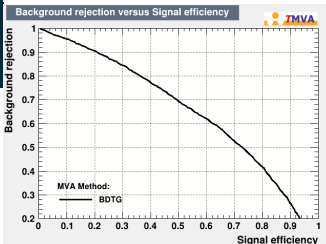
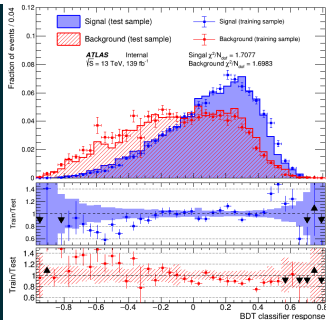
«Жесткий» отбор событий

- ▶ Ограничения на переменные были получены в ходе максимизации сигнальной значимости

Переменная	До	После		До	После
E_T^{miss} значимость	—	>10	Сигнал		
E_T^{miss} , ГэВ	—	—	QCD ZZ	7596 ±28	1914 ±16
ΔR_{ll}	—	<1.8	EWK ZZ	262 ±2	12.7 ±0.4
$\Delta\phi(E_T^{miss}, p_T^{ll})$	—	>2.3	Total signal	7858 ±28	1926 ±15
Число b-струй	—	<1	Фон		
E_T^{miss}/H_T	—	>0.5	Zj	962833 ±4057	155 ±19
			WZ	11338 ±29	933 ±8
Сигнальная значимость	7.43±0.03	44.1±0.4	tt	123340 ±73	125 ±2
Число сигнальных событий	7858 ±28	1926 ±16	WW	5093 ±13	62.4 ±1.5
Число фоновых событий	(1123 ±4)·10 ³	1368 ±21	Wt	10251 ±41	37 ±3
			VVV	41.8 ±0.3	7.83 ±0.10
			Other	282 ±2	0.77 ±0.11
			Total bkg.	(1123 ±4)·10 ³	1368 ±22

Результаты тренировки классификатора с «жестким» предотбором событий

Rank	Variable	Separation
1	MetOHT	3.772e-02
2	met_signif	3.701e-02
3	dMetZPhi	1.752e-02
4	met_tst	1.715e-02
5	mT_ZZ	1.450e-02
6	ZpTomT	1.277e-02
7	n_jets	1.129e-02
8	dLepR	1.022e-02
9	M2Lep	1.006e-02
10	Z_pT	9.471e-03
11	leading_pT_lepton	5.227e-03
12	subleading_pT_lepton	4.602e-03
13	dLepPhi	4.249e-03
14	RhoZ	3.643e-03
15	dLepEta	3.622e-03
16	sumpT_scalar	2.712e-03
17	LepRatio	1.607e-03
18	Z_rapidity	1.392e-03



- ▶ Не удалось достичь хорошего разделения сигнала и фона
- ▶ Классификатор переобучен
- ▶ ROC-кривая почти прямая

«Расслабленный» предотбор событий

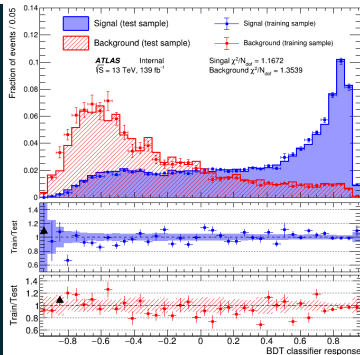
- ▶ В событии два разноименно-заряженных лептона одного аромата (e^+e^- или $\mu^+\mu^-$), при этом, поперечный импульс первого больше 30 ГэВ, второго больше 20 ГэВ;
- ▶ Вето на третий заряженный лептон;
- ▶ $76 \text{ ГэВ} < M_{ll} < 106 \text{ ГэВ}$;
- ▶ $E_T^{\text{miss}} > 70 \text{ ГэВ}$.

Variable	Strong	Relax
$E_T^{\text{miss}} \text{ signif.}$	>10	>7
$E_T^{\text{miss}}, \text{ GeV}$	>70	>70
ΔR_{ll}	<1.8	<2.2
$\Delta\phi(\vec{E}_T^{\text{miss}}, \vec{p}_T^{\text{ll}})$	>2.3	>1.3
$N_{b\text{-jets}}$	<1	<1
E_T^{miss}/H_T	>0.5	>0

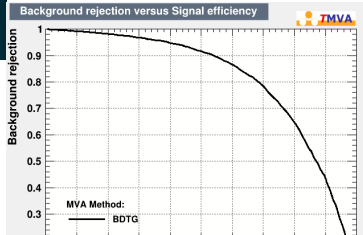
	Strong	Relax
Signal		
QCD ZZ	1914 ± 16	4409 ± 23
EWK ZZ	12.7 ± 0.4	57.8 ± 0.9
Total signal	1926 ± 16	4467 ± 23
Background		
Zj	155 ± 19	12184 ± 290
WZ	933 ± 8	3116 ± 15
tt	125 ± 2	2829 ± 11
WW	62.4 ± 1.4	1352 ± 7
Wt	37 ± 2	729 ± 10
llll	47.2 ± 1.0	170.5 ± 1.5
VVV	7.83 ± 0.10	17.71 ± 0.17
llqq	0.9 ± 0.3	35.9 ± 1.3
Other	0.77 ± 0.11	4.46 ± 0.26
Total bkg.	1368 ± 21	20439 ± 291
Total		
Total	3295 ± 26	24906 ± 292
data ¹	3584 ± 60	29968 ± 173

Результаты тренировки классификатора с расслабленным предотбором событий

```
-----  
Rank : Variable : Separation  
-----  
1 : met_signif : 2.819e-01  
2 : met_tst : 2.089e-01  
3 : mT_ZZ : 2.048e-01  
4 : Z_pT : 1.352e-01  
5 : dLepR : 1.175e-01  
6 : leading_pT_lepton : 9.075e-02  
7 : ZpTomT : 8.831e-02  
8 : MetOHT : 8.319e-02  
9 : RhoZ : 6.550e-02  
10 : sumpT_scalar : 5.821e-02  
11 : dLepPhi : 4.992e-02  
12 : subleading_pT_lepton : 4.072e-02  
13 : dMetZPhi : 1.799e-02  
14 : LepRatio : 1.723e-02  
15 : dLepEta : 1.560e-02  
16 : M2Lep : 1.416e-02  
17 : Z_rapidity : 1.292e-03  
18 : n_jets : 3.590e-04  
-----
```

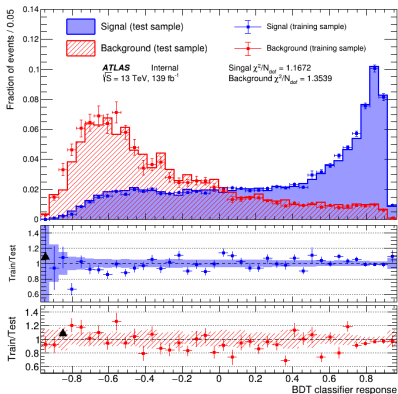


- ▶ Сигнал и фон хорошо разделяются
- ▶ Значение χ^2/NDF близко к 1 для сигнала и фона

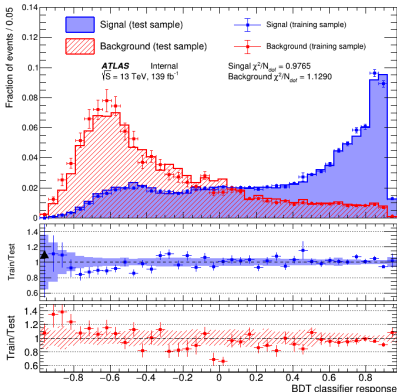


Исключение скоррелированных переменных

Все переменные



Удалены скоррелированные переменные



При удалении скоррелированных переменных улучшилось согласие отклика классификатора тренировочной и тестовой выборок.

Оптимизация гиперпараметров классификатора.

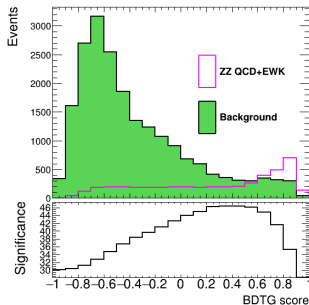
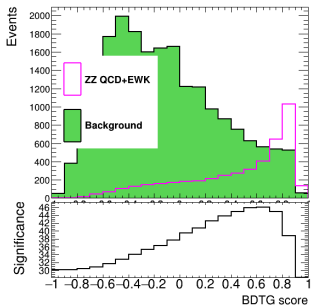
Оптимизация происходила перебором всех вариантов параметров по сетке с максимизацией площади под ROC-кривой.

Начальный набор гиперпараметров:

Option	Value
NTrees	200
Boost Type	Grad
MaxDepth	2
MinNodeSize	2%
nCuts	30

Оптимизированный набор гиперпараметров:

Option	Value
NTrees	800
Boost Type	Grad
MaxDepth	3
MinNodeSize	5%
nCuts	20



Оптимизация гиперпараметров не привела к значительным улучшениям.

Заключение:

- ▶ Были рассмотрены различные варианты обучения классификатора.
- ▶ В ходе оптимизации получен стабильный алгоритм
- ▶ Максимально достижимая сигнальная значимость с помощью классификатора BDTG увеличилась с 44.1 ± 0.4 до 46 ± 0.3 .

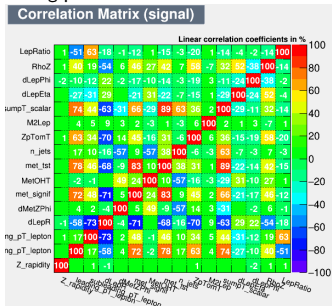
Дальнейшие планы:

- ▶ Использовать полученный классификатор для оценки числа сигнальных событий.

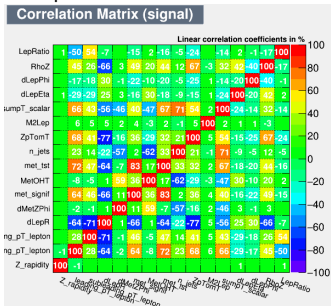
backup

Correlation matrices

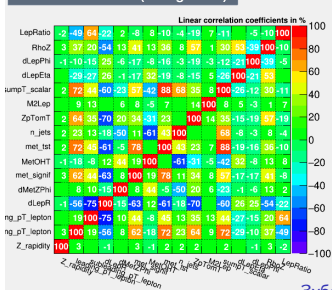
Strong preselection



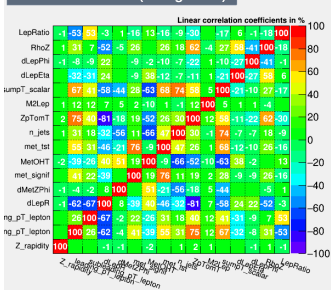
Relax preselection



Correlation Matrix (background)



Correlation Matrix (background)



Hao's slides from 21 Dec 2020: (ANA-HIGG-2018-26-INT2)

EW correction

- Apply EW correction based on truth met for QCD ZZ sample (345723 + 345666)
- More detail in section 'bkgEstimation_ZZ' in : [ANA-HIGG-2018-26-INT2](#)

50<met_Truth<60	correction_factor=0.95049841
60<met_Truth<70	correction_factor=0.94603957
70<met_Truth<80	correction_factor=0.94116453
80<met_Truth<90	correction_factor=0.93749643
90<met_Truth<100	correction_factor=0.93427814
100<met_Truth<120	correction_factor=0.92784113
120<met_Truth<140	correction_factor=0.91831967
140<met_Truth<160	correction_factor=0.90832711
160<met_Truth<180	correction_factor=0.89836000
180<met_Truth<200	correction_factor=0.88769693
200<met_Truth<250	correction_factor=0.87195970
250<met_Truth<300	correction_factor=0.84713314
300<met_Truth<350	correction_factor=0.82531651
350<met_Truth<400	correction_factor=0.80451086
400<met_Truth<450	correction_factor=0.78480511
450<met_Truth<500	correction_factor=0.76750962
500<met_Truth<600	correction_factor=0.74312196
600<met_Truth<700	correction_factor=0.71224842
700<met_Truth<800	correction_factor=0.68098102
800<met_Truth<1000	correction_factor=0.66162730

global_weight* = correction_factor

2020/12/21

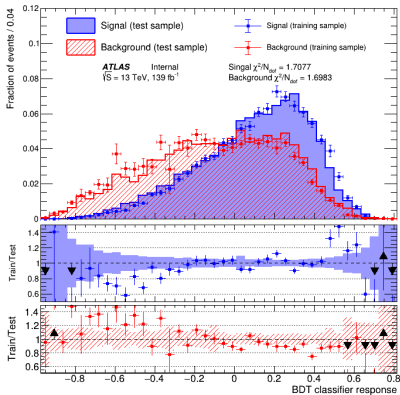
Hao Xu (USTC)

2

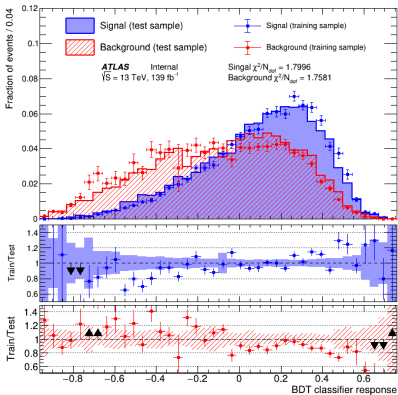
345666.Sherpa__222__NNPDF30NNLO_llvZZ.deriv.DAOD_STDM3.e6240_s3126_r9364_p4252.root
345666.Sherpa__222__NNPDF30NNLO_llvZZ.deriv.DAOD_STDM3.e6240_s3126_r10201_p4252.root
345666.Sherpa__222__NNPDF30NNLO_llvZZ.deriv.DAOD_STDM3.e6240_s3126_r10764_p4252.root

Remove correlated variables

All variables



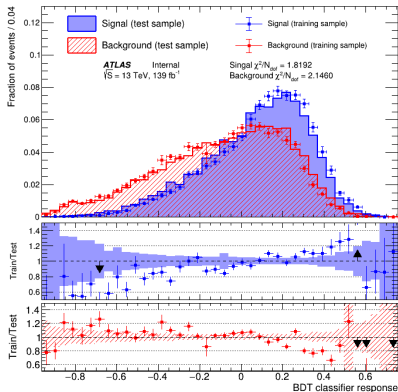
Correlated variables are removed



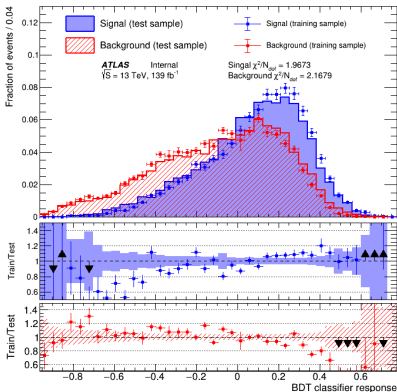
No improvement with removing correlated variables.

Remove Zj , $llqq$, ttV and $ttVV$ backgrounds

All variables



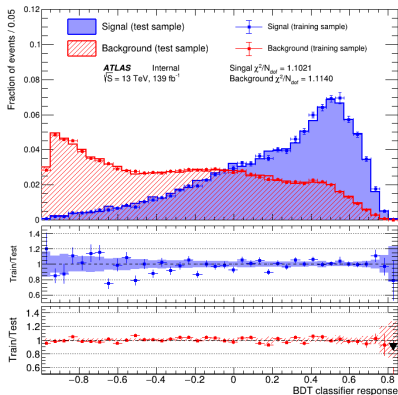
Correlated variables are removed



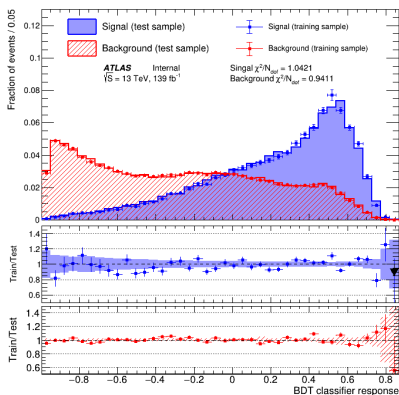
Removing Zj , $llqq$, ttV and $ttVV$ backgrounds leads to lower stat error for classifier response for Background sample, but doesn't lead to improving of stability of classifier.

Remove Zj, llqq, ttV and ttVV backgrounds. Relax.

All variables



Correlated variables are removed



Removing correlated variables increases $\frac{\chi^2}{NDF}$.

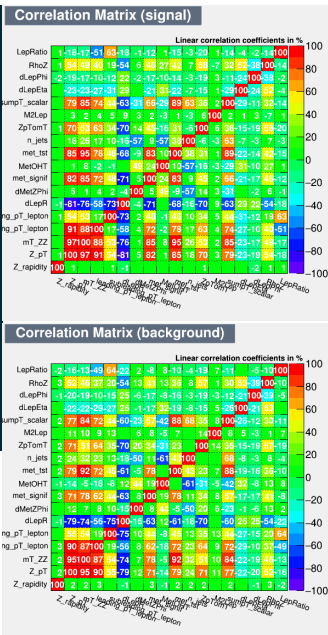
Removing Zj, llqq, ttV and ttVV backgrounds increases $\frac{\chi^2}{NDF}$ and decreases stat error for background distribution.

Используемые переменные

Rank	Variable	Separation
1	MetOHT	3.772e-02
2	met_signif	3.701e-02
3	dMetZPhi	1.752e-02
4	met_tst	1.715e-02
5	mT_ZZ	1.450e-02
6	ZpTomT	1.277e-02
7	n_jets	1.129e-02
8	dLepR	1.022e-02
9	M2Lep	1.006e-02
10	Z_pT	9.471e-03
11	leading_pT_lepton	5.227e-03
12	subleading_pT_lepton	4.602e-03
13	dLepPhi	4.249e-03
14	RhoZ	3.643e-03
15	dLepEta	3.622e-03
16	sumpT_scalar	2.712e-03
17	LepRatio	1.607e-03
18	Z_rapidity	1.392e-03

$$\langle S^2 \rangle = \frac{1}{2} \int \frac{(\hat{y}_S(y) - \hat{y}_B(y))^2}{\hat{y}_S(y) + \hat{y}_B(y)} dy$$

mT_ZZ и Z_pT — кандидаты на исключение

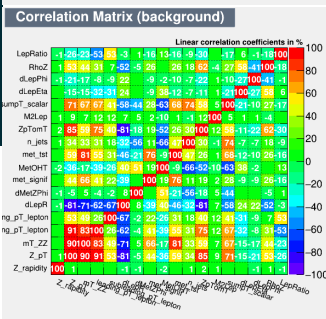
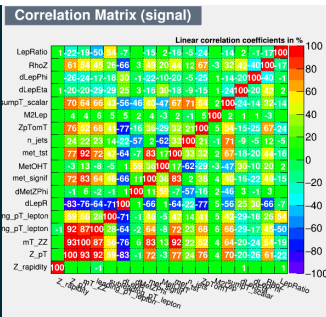


Используемые переменные.

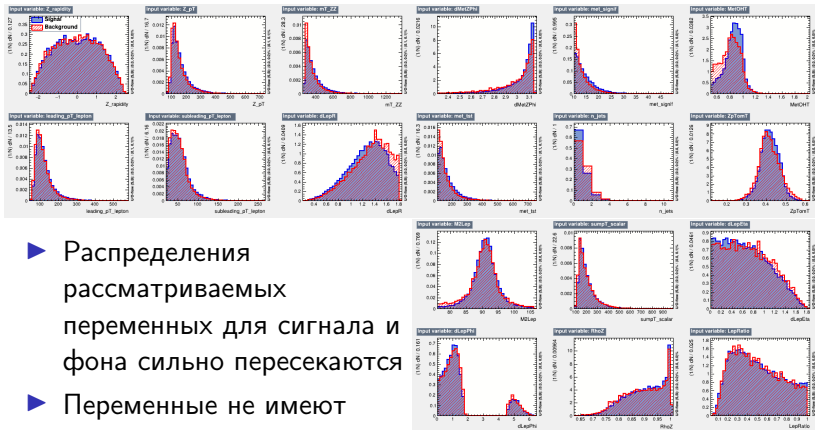
Rank	Variable	Separation
1	met_signif	2.819e-01
2	met_tst	2.089e-01
3	mT_ZZ	2.048e-01
4	Z_pT	1.352e-01
5	dLepR	1.175e-01
6	leading_pT_lepton	9.075e-02
7	ZpTomT	8.831e-02
8	MetOHT	8.319e-02
9	RhoZ	6.550e-02
10	sumpT_scalar	5.821e-02
11	dLepPhi	4.992e-02
12	subleading_pT_lepton	4.072e-02
13	dMetZPhi	1.799e-02
14	LepRatio	1.723e-02
15	dLepEta	1.560e-02
16	M2Lep	1.416e-02
17	Z_rapidity	1.292e-03
18	n_jets	3.590e-04

$$\langle S^2 \rangle = \frac{1}{2} \int \frac{(\hat{y}_S(y) - \hat{y}_B(y))^2}{\hat{y}_S(y) + \hat{y}_B(y)} dy$$

mT_ZZ и Z_pT — кандидаты на исключение



Используемые переменные



Используемые переменные

