

Выделение сигнала в инклюзивном процессе
 $pp \rightarrow ZZ \rightarrow \ell\nu\nu$ в эксперименте АТЛАС на
данных второго сеанса работы БАК.

Зубов Дмитрий

Национальный исследовательский ядерный университет «МИФИ»

Научный руководитель: Солдатов Е.Ю.

Консультант: Петухов А.М.

13.06.2023

Мотивация и цели

Мотивация:

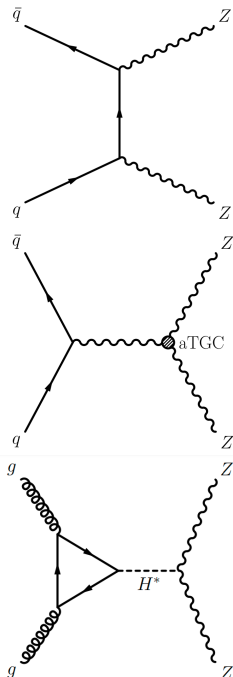
- ▶ Отсутствие ZZZ или $ZZ\gamma$ вершин взаимодействия \rightarrow косвенный поиск эффектов, предсказанных теориями за пределами СМ.
- ▶ $pp \rightarrow ZZ$ важный фон для процесса рождения бозона Хиггса.

Цели анализа:

- ▶ Получить интегральное и дифференциальные сечения для переменных $p_T^{\ell\ell} = p_T^Z$, $\Delta\phi(\ell\ell)$, m_{T}^{ZZ} , N_{jets} , $m(j_1, j_2)$.
- ▶ Поставить пределы на аномальные тройные вершины взаимодействия.

Задача:

- ▶ Измерение сечения рождения пар Z-бозонов с использованием классификаторов машинного обучения для повышения точности оценки.



Подходы к оценке интегрального сечения в анализе.

1. Подход со «строгим» предотбором событий:

- ▶ Выделение сигнальной области с наилучшим сохранением сигнала и наилучшим подавлением фона.
- ▶ Разработка методики оценки фонов.
- ▶ Оценка интегрального сечения.

2. Подход с «расслабленным» предотбором событий и применение классификатора машинного обучения для лучшего разделения сигнала и фона:

- ▶ Выбор расслабленной сигнальной области с относительно хорошим подавлением фона и с большей степенью сохранения сигнала.
- ▶ Создание классификатора машинного обучения.
- ▶ Разработка методики оценки фонов.
- ▶ Оценка интегрального сечения.

Inclusive $ZZ \rightarrow ll\nu\nu$

- ▶ 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;
- ▶ $76 \text{ ГэВ} < M_{ll} < 106 \text{ ГэВ}$;
- ▶ $E_T^{miss} > 70 \text{ ГэВ}$.
- ▶ Отбор объектов описан в бэкапе

Signal	
ZZ ($\sim 0.7\%$)	Рождение двух Z-бозонов и последующий распад в $ll\nu\nu$
Background	
Zj ($\sim 85.6\%$)	рождение Z-бозона и струи, с распадом Z-бозона в пару заряженных лептонов и большим ложным потерянными поперечным импульсом
tt ($\sim 11.0\%$)	рождение пары топ-кварков и последующим распадом включающим конечное состояние $ll\nu\nu$ (не резонансное рождение $ll\nu\nu$)
WZ ($\sim 1.0\%$)	рождение пары бозонов Z и W, с распадом Z-бозона в пару заряженных лептонов и лептонным распадом W
WW ($\sim 0.5\%$)	рождение пары W с распадом в $ll\nu\nu$ (не резонансное рождение $ll\nu\nu$)
Wt ($\sim 0.9\%$)	рождение W и топ-кварка и распадом в конечное состояние, содержащее $ll\nu\nu$ (не резонансное рождение $ll\nu\nu$)
Other ($4l, llqq, VVV, Z(\tau\tau), W + \text{jets}$)	Фоновые процессы, которые вносят незначительный вклад в общее число событий и оцениваются с помощью МК

Оптимизация отбора событий. Строгий предотбор событий

- В процессе оптимизации искались пороги на переменные, при которых достигается максимум сигнальной значимости:

$$Z = \sqrt{2 \times [(S + B) \times \ln(1 + (S/B)) - S]}$$

- Сигнальная значимость рассматривалась как функция нескольких переменных и поиск оптимального вектора оптимизируемых отборов осуществлялся перебором всевозможных вариантов ограничения фазового пространства.

Переменная	До	После
E_T^{miss} , ГэВ	—	>70
ΔR_{ll}	—	<1.8
$\Delta\phi(E_T^{miss}, p_T^l)$	—	>2.3
N_{b-jets}	—	<1
E_T^{miss} значимость	—	>10

Сигнал

QCD ZZ	7600±30	1946±15
EWK ZZ	262±2	13.0±0.04
Всего сигнала	7860±30	1959±15

Фон

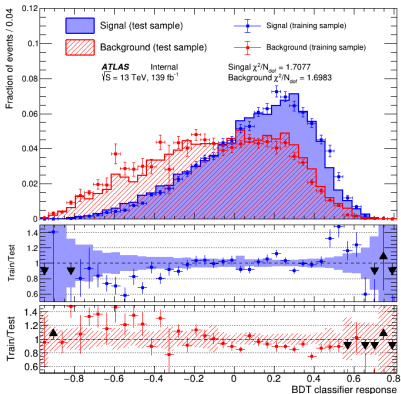
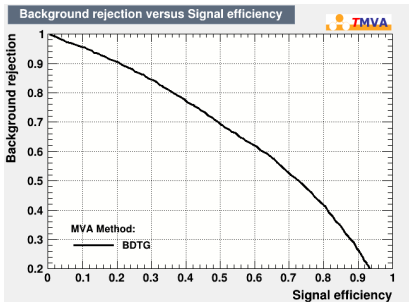
Zj	963000±4000	180±20
WZ	11340±30	945±8
tt	123340±80	131±2
WW	5093±13	64.0±1.5
Wt	10250±40	41±3
VVV	41.8±0.3	7.88±0.10
Other	282±2	0.79±0.11
Всего фона	1123000±4000	1368±20

Сигнальная значимость	5.43±0.02	44.7±0.4
------------------------------	------------------	-----------------

Результаты тренировки классификатора с «жестким» предотбором событий

Переменная	Жесткий отбор
E_T^{miss} , ГэВ	>70
E_T^{miss} значимость	>10
ΔR_{ll}	<1.8
$\Delta\phi(\vec{E}_T^{miss}, \vec{p}_T^{ll})$	>2.3
N_{b-jets}	<1

- ▶ Не удалось достичь хорошего разделения сигнала и фона
- ▶ Классификатор переобучен



Выделение «расслабленной» сигнальной области

- Для выделения сигнальной области с расслабленными предотборами производилась оптимизация порогов на переменные.

- В ходе этой оптимизации искался вектор порогов на переменные, соответствующий максимуму сигнальной значимости при условии, что число сигнальных событий >4500 .

Переменная	Расслабленный отбор	Жесткий отбор
E_T^{miss} , ГэВ	>70	>70
E_T^{miss} значимость	>7	>10
ΔR_{ll}	<2.2	<1.8
$\Delta\phi(\vec{E}_T^{miss}, \vec{p}_T^l)$	>1.3	>2.3
Число b-струй	<1	<1

Сигнал

QCD ZZ	4410 ± 20	1946 ± 15
EWK ZZ	57.8 ± 0.9	13.0 ± 0.4
Всего сигнала	4470 ± 20	1959 ± 15

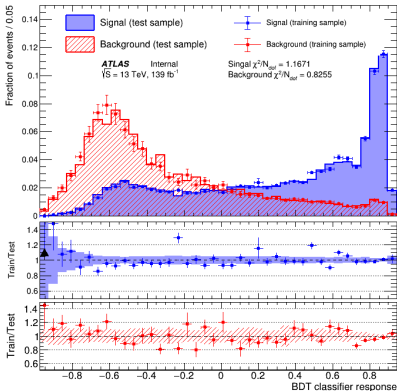
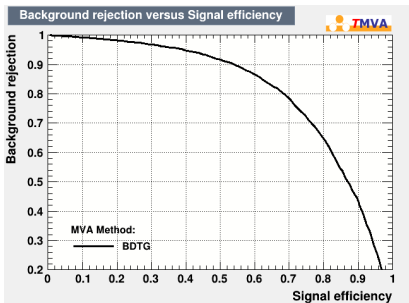
Фон

Zj	12200 ± 300	180 ± 20
WZ	3116 ± 15	945 ± 8
tt	2829 ± 11	131 ± 2
WW	1352 ± 7	64.0 ± 1.5
Wt	729 ± 10	41 ± 3
VVV	1771 ± 0.17	7.88 ± 0.10
Other	4.46 ± 0.26	0.79 ± 0.11
Всего фона	20400 ± 300	1370 ± 20

Результаты тренировки классификатора с расслабленным предотбором событий

Переменная	Расслабленный отбор
E_T^{miss} , ГэВ	>70
E_T^{miss} значимость	>7
ΔR_{ll}	<2.2
$\Delta\phi(\vec{E}_T^{miss}, \vec{p}_T^l)$	>1.3
N_{b-jets}	<1

- ▶ Сигнал и фон хорошо разделяются
- ▶ Максимальная сигнальная значимость 46.8 ± 0.4
- ▶ Гиперпараметры и переменные оптимизированы



Описание фита

Интегральное сечение и фон оцениваются в фите, путем максимизации функции правдоподобия:

$$\mathcal{L}(\mu, \theta) = \prod_r \left[\prod_{i \in \text{bins}} \text{Pois}(N_i^{\text{data}} | \mu \nu_i^s \eta^s(\theta) + \nu_i^b \eta^b(\theta)) \right] \cdot \prod_i^{\text{nuis. par.}} \mathcal{L}(\theta_i),$$

$N(\nu)$ — наблюдаемое (прогнозируемое) количество событий

μ — коэффициент нормировки сигнала (сила сигнала), $\mu = \nu^s / N^s$.

θ — коэффициенты нормировки фона и систематические неопределенности.

η — отражает влияние систематических неопределенностей на число событий в бине.

В фите 4 области (включая сигнальную) и 4 коэффициента нормировки (включая силу сигнала).

Сейчас фит к наблюдаемым данным выполняется **только в контрольных областях** и набор данных Азимова используется вместо наблюдаемых данных в сигнальной области.

Следующая статистика используется для вычисления ожидаемой значимости и неопределенности $\hat{\mu}$:

$$q(\mu, \hat{\mu}, \hat{\theta}) = -2 \ln \lambda(\mu, \hat{\mu}, \hat{\theta}) = -2 \ln \frac{\mathcal{L}(\mu, \hat{\theta}(\mu))}{\mathcal{L}(\hat{\mu}, \hat{\theta})}, \quad Z_{\text{disc}}^{\text{exp.}} = \sqrt{q(\mu = 1)_{\text{A}}}. \quad (1)$$

Определение контрольных и сигнальной областей.

Расслабленный вариант ФП:

Переменная	SR	WZ (3 ℓ)	NR ($e\mu$)	Zj
E_T^{miss} , ГэВ	>70		>70	
ΔR_{ll}	<2.2		<2.2	
$\Delta\phi(E_T^{miss}, \vec{p}_T^l)$, рад	>1.3		>1.3	
E_T^{miss} значимость		>7		[4;7]
m_T^W , ГэВ		>60		

Фит в сигнальной области происходил по переменной BDT_{score} , в контрольных по переменной ρ_T^Z

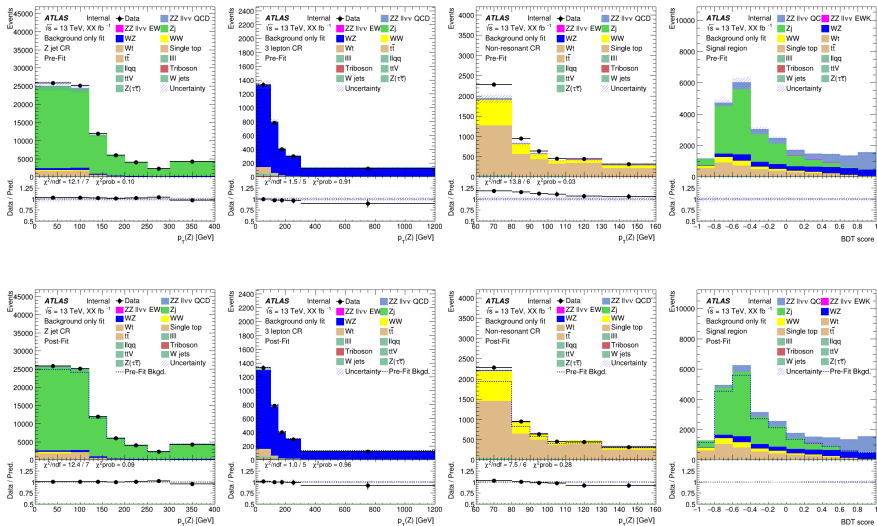
Жесткий вариант ФП:

Переменная	SR	WZ (3 ℓ)	NR ($e\mu$)	Zj
E_T^{miss} , ГэВ	>70		>70	
ΔR_{ll}	<1.8		<1.8	
$\Delta\phi(E_T^{miss}, \vec{p}_T^l)$, рад	>2.3		>2.3	
E_T^{miss} значимость		>10		[4;9]
m_T^W , ГэВ		>60		

Фит в сигнальной и контрольных областях происходил по переменной ρ_T^Z

- ▶ **SR** - область фазового пространства, в которой доля сигнальных событий максимальна.
- ▶ **WZ(3 ℓ)** - область фазового пространства, в которой доля событий процесса WZ максимальна.
- ▶ **Non-resonant** - область фазового пространства, в которой доля событий процессов нерезонансного рождения $\ell^+\ell^-$ максимальна.
- ▶ **Zj** - область фазового пространства, в которой доля событий процесса Zj максимальна.

Фит. Распределения до и после для расслабленного варианта ФП.



Фит. Результаты.

Полученное значение силы сигнала μ_{ZZ} применяется при вычисления наблюдаемого сечения: $\sigma_{meas.} = \mu_{ZZ} \cdot \sigma_{SM}$

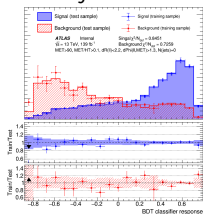
	«Жесткий» фит по $\rho^T(Z)$	«Расслабленный» фит по $\rho^T(Z)$	«Расслабленный» фит по отклику BDT
μ_{ZZ}	$1.00^{+0.04}_{-0.04}(\text{stat})^{+0.06}_{-0.05}(\text{syst})$	$1.00^{+0.04}_{-0.04}(\text{stat})^{+0.06}_{-0.05}(\text{syst})$	$1.00^{+0.03}_{-0.03}(\text{stat})^{+0.06}_{-0.05}(\text{syst})$
μ_{Zj}	$1.31^{+0.03}_{-0.03}(\text{stat})^{+0.07}_{-0.07}(\text{syst})$	$1.13^{+0.01}_{-0.01}(\text{stat})^{+0.06}_{-0.06}(\text{syst})$	$1.13^{+0.01}_{-0.01}(\text{stat})^{+0.06}_{-0.06}(\text{syst})$
μ_{NR}	$1.11^{+0.08}_{-0.07}(\text{stat})^{+0.05}_{-0.05}(\text{syst})$	$1.15^{+0.02}_{-0.02}(\text{stat})^{+0.05}_{-0.05}(\text{syst})$	$1.15^{+0.02}_{-0.02}(\text{stat})^{+0.05}_{-0.05}(\text{syst})$
μ_{WZ}	$1.01^{+0.05}_{-0.05}(\text{stat})^{+0.06}_{-0.05}(\text{syst})$	$0.97^{+0.02}_{-0.02}(\text{stat})^{+0.06}_{-0.05}(\text{syst})$	$0.97^{+0.02}_{-0.02}(\text{stat})^{+0.06}_{-0.05}(\text{syst})$
Ожидаемая значимость	16.8	16.2	26.1

Фит по отклику классификатора BDT в расслабленном фазовом пространстве показывает значительно большую ожидаемую значимость.

Изменения классификатора

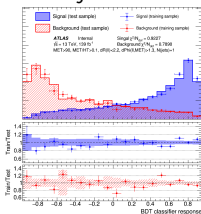
- ▶ Обучение классификаторов для 4-ех категорий с количеством струй 0, 1, 2 и более 2
- ▶ Добавление переменных, описывающих струи
- ▶ Оптимизация гиперпараметров

$N_{jets}=0$



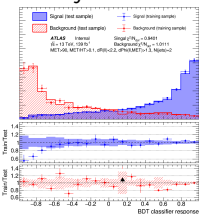
p-Value(Signal) = 0.65
p-Value(Bckg) = 0.80

$N_{jets}=1$



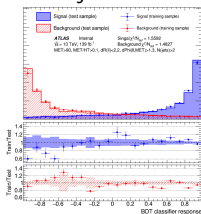
p-Value(Signal) = 0.55
p-Value(Bckg) = 0.72

$N_{jets}=2$



p-Value(Signal) = 0.53
p-Value(Bckg) = 0.44

$N_{jets}>2$



p-Value(Signal) = 0.057
p-Value(Bckg) = 0.080

Оптимальные гиперпараметры:

- ▶ NTrees: 400
- ▶ Shrinkage: 0.2
- ▶ MinNodeSize: 10%

	Signal	Bkg	Z
$N_{jets}=0$	1311.96	690.149	40.51
$N_{jets}=1$	639.818	611.829	22.63
$N_{jets}=2$	234.449	266.918	12.78
$N_{jets}>2$	139.62	278.742	7.78
Total	2325.85	1847.91	46.36

Заключение

- ▶ Велась работа по созданию классификатора.
- ▶ Использование классификатора при извлечении сигнала показало снижение статистических ошибок и увеличение ожидаемой.
- ▶ Получены стабильные классификаторы с хорошей разделяющей способностью для каждой категории.

Следующие шаги:

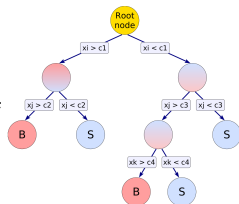
- ▶ Выполнить фит с учетом систематических погрешностей в ослабленной сигнальной области сравнить его с фитом в строгой сигнальной области.

backup

Decision trees with gradient boosting (BDTG)

Decision tree

- ▶ A decision tree is a binary tree : a sequence of cuts paving the phase-space of the input variables
- ▶ Repeated yes/no decisions on each variables are taken for an event until a stop criterion is fulfilled
- ▶ Trained to maximize the purity of signal nodes (or the impurity of background nodes)



Advantages:

- ▶ Decision trees are independent of monotonous variable transformations
- ▶ Weak variables are ignored and do not deteriorate performance

Disadvantages:

- ▶ Decision trees are extremely sensitive to the training samples, therefore to overtraining
- ▶ Slightly different training samples can lead to radically different DT

Boosting

- ▶ Sequentially apply the DT algorithm to reweighted (boosted) versions of the training data
- ▶ Each model in the series trains upon its predecessor's mistakes, trying to correct them
- ▶ Works very well on non-optimal decision tree (small number of nodes)
- ▶ There are different boosting algorithms and in our work we use the gradient descent

Classifier training parameters

Hyperparameters:

- ▶ Number of trees;
- ▶ Max depth of the decision tree allowed
- ▶ Minimum percentage of training events required in a leaf node
- ▶ Number of grid points in variable range used in finding optimal cut in node splitting
- ▶ Shrinkage (Learning rate)

Separation into training and test sample:

- ▶ Random division in equal proportion

1. E_T^{miss} ;
2. Object-based E_T^{miss} -significance;
3. $m_T(ZZ)$;
4. $\Delta R(\ell\ell)$;
5. $p_T^{\ell 1}$;
6. $p_T^{\ell 2}$;
7. $p_T^Z = p_T^{\ell\ell}$
8. $\frac{p_T^Z}{m_T(ZZ)}$;
9. $H_T = p_T^{\ell 1} + p_T^{\ell 2} + \sum_i p_{T}(j_i)$;
10. E_T^{miss} / H_T ;
11. $\rho_Z = \frac{p_T^Z}{p_T^{\ell 1} + p_T^{\ell 2}}$;
12. $\Delta\varphi(\vec{p}_T^{\ell 1}, \vec{p}_T^{\ell 2})$;
13. $\Delta\varphi(\vec{p}_T^{\text{miss}}, \vec{p}_T^{\ell\ell})$;
14. $\frac{p_T^{\ell 1}}{p_T^{\ell 2}}$;
15. $\Delta\eta(\ell\ell)$;
16. $m(\ell\ell)$;
17. y_Z ;
18. N_{jets} ;
19. $\frac{E_T^{\text{miss}} + \sum_i p_{T}(j_i) - p_T^Z}{p_T^Z}$;
20. V_T

Object selection

Electrons

- ▶ Likelihood medium
- ▶ $\text{lead} > 30 \text{ ГэВ}$
- ▶ $\text{sublead} > 20 \text{ ГэВ}$
- ▶ $|\eta| \text{ calo cluster} < 2.47$
- ▶ $|\Delta(z_0) \cdot \sin(\theta)| < 0.5 \text{ мм}$
- ▶ $|d_0\text{-significance}| < 5$
- ▶ Isolation WP
FixedCutLoose
- ▶ Crack region veto
- ▶ Исключение пересечений с мюонами и струями

Muons

- ▶ Medium
- ▶ $|\eta| < 2.5$
- ▶ $\text{lead} > 30 \text{ ГэВ}$
- ▶ $\text{sublead} > 20 \text{ ГэВ}$
- ▶ Combined muons
- ▶ $|\Delta(z_0) \cdot \sin(\theta)| < 0.5 \text{ мм}$
- ▶ $|d_0\text{-significance}| < 3$
- ▶ Isolation WP
PflowLoose_FixedRad
- ▶ Исключение пересечений со струями

Jets

- ▶ AntiKt4EMPFLOW
- ▶ $> 30 \text{ ГэВ}$
- ▶ $|\eta| < 4.5$
- ▶ $\text{JVT} > 0.5$
- ▶ Event-level cleaning for LooseBad jets
- ▶ Tight WP, rebuilt with METMaker using selected leptons and all calibrated jets

Изменения в сигнальной области

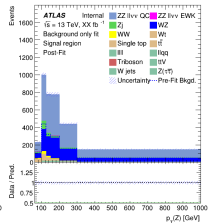
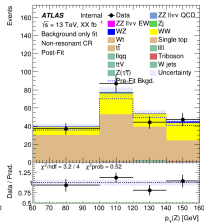
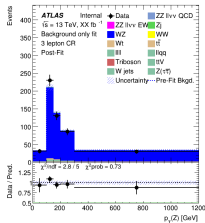
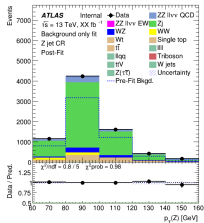
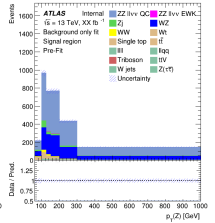
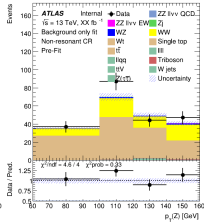
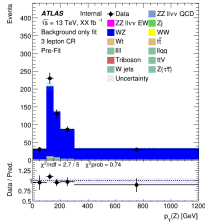
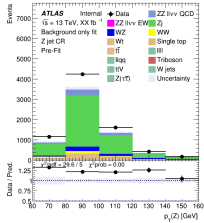
- ▶ Отказ от определения CP с переменной E_T^{miss} -значимости из-за сложности в использовании на уровне генератора.
- ▶ Добавлены переменные E_T^{miss}/H_T и $m_{\ell\ell}$
- ▶ Оптимизация отбора событий также проводилась путем максимизации значимости сигнала Z с использованием многомерного подхода.

	Строгий	Расслабленный		Строгий	Расслабленный
$m_{\ell\ell}$, ГэВ	$\in [80; 100]$	$\in [76; 106]$			
$E_T^{\text{miss.}}$, ГэВ	> 110	> 90	Сигнал	1562 ± 15	3810 ± 20
$\Delta R(\ell\ell)$	< 1.8	< 2.2	Фон	1007 ± 17	25000 ± 300
$\Delta\varphi(E_T^{\text{miss.}}, \vec{\ell\ell})$	> 2.7	> 1.3	Z	41.1 ± 0.4	23.5 ± 0.2
$E_T^{\text{miss.}}/H_T$	> 0.65	> 0.1			
$N_{\text{b-jets}}$	$= 0$	$= 0$			

Selection optimization details

- ▶ $E_T^{miss.}$, [50; 1500] GeV, a step of 10 GeV;
- ▶ $\Delta R(\ell\ell)$, [0; 4], a step of 0.1;
- ▶ $\Delta\varphi(\text{miss}, \ell\ell)$, [0; 3.15], a step of 0.1;
- ▶ $E_T^{miss.}/H_T$, [0; 2], a step of 0.05;
- ▶ $N_{\text{b-jets}}$, events with $\{0, 1, 2, 3, \geq 4\}$ b-jets.

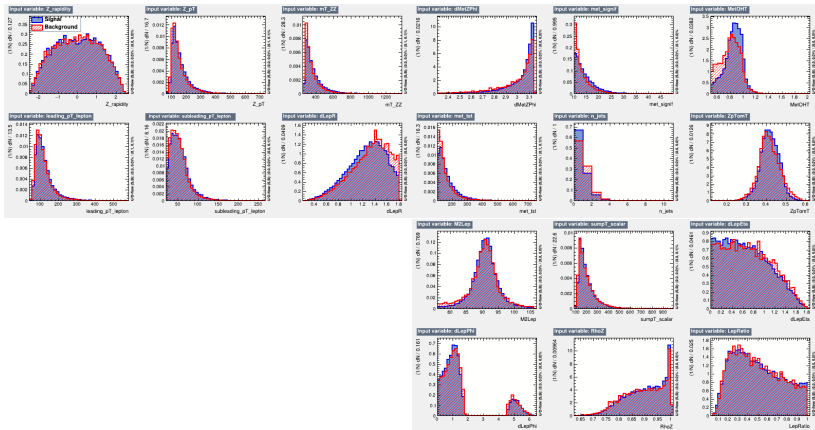
Fit. Before and after distributions for the strict version of the phase space



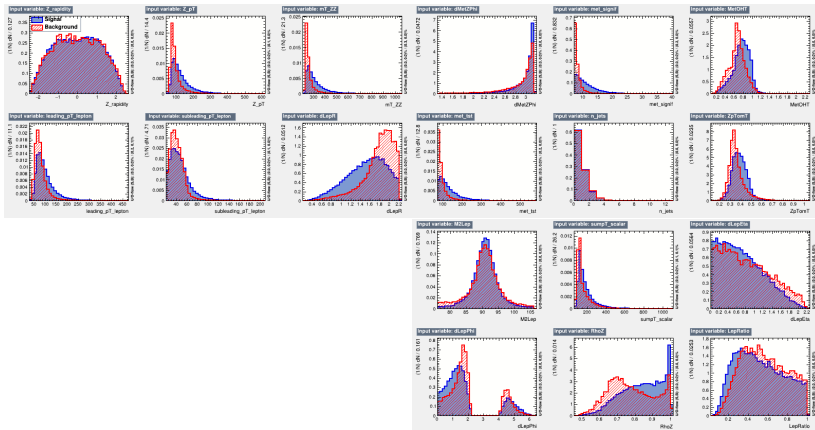
Theoretical		Experimental	
PDF	3.5%	Lepton.	2.0%
Scale	2.0%	Jet.	2.0%
UEPS	2.0%	$E_T^{miss.}$	1.1%

Таблица: Main sources of Theoretical и Experimental errors

Variables. Strict preselection.

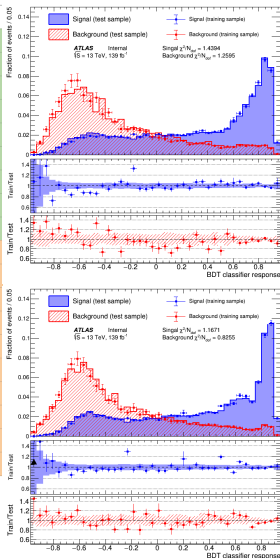


Variables. Loose preselection



Отбор переменных

Variable	auROC
M2Lep	0.79267
dMetZPhi	0.80029
dLepR	0.80434
MetOHT	0.80567
n_jets	0.80690
leading_pT_lepton	0.80715
mT_ZZ	0.80740
subleading_pT_lepton	0.80748
frac_pT	0.80818
sumpT_scalar	0.80809
met_tst	0.80813
LepRatio	0.80799
dLepPhi	0.80787
dLepEta	0.80693
Z_rapidity	0.80623
Z_pT	0.80393
sumpT_vector	0.80371
ZpTomT	0.79994
RhoZ	0.79746



- ▶ Идея в том, чтобы измерить важность переменной, глядя на сколько увеличивается $auROC$, когда переменная добавляется.
- ▶ Отбор начинается с одной переменной с наибольшим $auROC$ и последовательно добавляет переменную из оставшихся $N - n$ с самым высоким $auROC$.
- ▶ Это предполагает обучение BDT для каждого из $N - n$ комбинации для определения $auROC$ и нахождения лучшей комбинации.

Увеличение значимости с 46.1 ± 0.4 до 46.8 ± 0.4

Training setup

Variable	$N_{jets}=0$	$N_{jets}=1$	$N_{jets}=2$	$N_{jets}>2$
mT_{ZZ}	✓	✓	✓	✓
leading_pT_lepton	✓	✓	✓	✓
subleading_pT_lepton	✓	✓	✓	✓
dLepR	✓	✓	✓	✓
dMetZPhi	✓	✓	✓	✓
E_T^{miss} signif.	✓	✓	✓	✓
frac_pT	✓	✓	✓	✓
MetOHT	✓	✓	✓	✓
M2Lep	✓	✓	✓	✓
leading_jet_pt	—	✓	✓	✓
leading_jet_rapidity	—	✓	✓	✓
second_jet_pt	—	—	✓	✓
second_jet_rapidity	—	—	✓	✓
m_{jj}	—	—	✓	✓
dY _{jj}	—	—	✓	✓
jet_vsum_pt	—	—	✓	✓
jet_vsum_eta	—	—	✓	✓
jet_vsum_phi	—	—	✓	✓

Preselection	
E_T^{miss} , GeV	>90
ΔR_{ll}	<2.2
$\Delta\phi(E_T^{miss}, p_T^l)$	>1.3
N_{b-jets}	<1
E_T^{miss}/H_T	>0.1

Hyperparameter optimization

- ▶ Hyperparameter optimisation (HPO) has been used to improve separation power while maintaining stability.
- ▶ Optimised metrics — Z , with condition $p\text{-value} > 0.05\%$
- ▶ Training and test samples were randomly allocated each time, avoiding bias (SplitMode=random:SplitSeed=0)
- ▶ Hyperparameters (HP) under consideration:
 - ▶ NTrees: 100, 200, 300, 400, 600, 800, 1000
 - ▶ Shrinkage: 0.02, 0.04, 0.06, 0.08, 0.1, 0.2, 0.3, 0.4, 0.5, 1.0
 - ▶ MinNodeSize: 0.2%, 0.5%, 1%, 2%, 5%, 10%, 20%

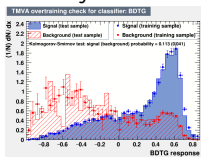
$$Z = \sqrt{2 \times [(S + B) \times \ln(1 + (S/B)) - S]}$$

A detailed comparison of all classifiers is shown in the [table](#).

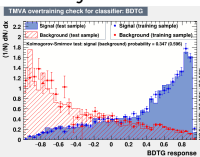
Изменения классификатора

- ▶ Обучение классификаторов для 4-ех категорий с количеством струй 0, 1, 2 и более 2
- ▶ Добавление переменных, описывающих струи

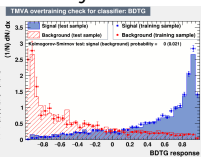
$N_{jets}=0$



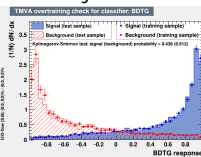
$N_{jets}=1$



$N_{jets}=2$



$N_{jets}>2$



- ▶ Уменьшилась статистика
- ▶ Для некоторых категорий классификатор получается переобученным при прежних настройках
- ▶ Необходимость обновления оптимизации гиперпараметров

Оптимизация гиперпараметров

- ▶ Hyperparameter optimisation (HPO) has been used to improve separation power while maintaining stability.
- ▶ Optimised metrics — Z, with condition $p\text{-value} > 0.05\%$
- ▶ Training and test samples were randomly allocated each time, avoiding bias (SplitMode=random:SplitSeed=0)
- ▶ Hyperparameters (HP) under consideration:
 - ▶ NTrees: 100, 200, 300, 400, 600, 800, 1000
 - ▶ Shrinkage: 0.02, 0.04, 0.06, 0.08, 0.1, 0.2, 0.3, 0.4, 0.5, 1.0
 - ▶ MinNodeSize: 0.2%, 0.5%, 1%, 2%, 5%, 10%, 20%

$$Z = \sqrt{2 \times [(S + B) \times \ln(1 + (S/B)) - S]}$$

A detailed comparison of all classifiers is shown in the [table](#).

	Signal	Bkg	Signif.
--	--------	-----	---------

Z-optimized

$N_{jets}=0$	1311.96	690.149	40.51
$N_{jets}=1$	639.818	611.829	22.63
$N_{jets}=2$	234.449	266.918	12.78
$N_{jets}>2$	139.62	278.742	7.78
Total	2325.85	1847.91	46.36

AUC-optimized (previous results)

$N_{jets}=0$	1415.59	837.336	40.35
$N_{jets}=1$	653.12	621.09	22.91
$N_{jets}=2$	244.35	301.53	12.62
$N_{jets}>2$	167.0	439.5	7.67
Total	2480.06	2041.89	47.23

No jet separation (previous results)	2237.84	1673.63	46.53
--------------------------------------	---------	---------	-------

$$Z = \frac{1}{\sqrt{2 \times [(S + B) \times \ln(1 + (S/B)) - S]}}$$

Selection of classifier hyperparameters

- ▶ The performance of the classifier varies from training to training. In other words, a classifier that previously showed excellent performance may perform worse the next time with the same settings.
- ▶ During HPO, specific HPs were not selected, but patterns in setting, stability, and separation power were observed:
 0. The larger Z is, the more unstable the classifier is. Therefore, it is essential to find a compromise set of HP.
 1. Classifiers with the highest Z usually have a relatively large number of trees.
 2. Classifiers with the highest Z usually have a relatively large number of trees (400-1000) and shrinkage (0.1-0.5).
 3. At large values of minnodesize (20%), the classifiers had the lowest separating power, but at the value of this parameter of 10%, stable classifiers with large Z were observed.
- ▶ Thus, one set of hyperparameters was defined for all categories, at which the classifier has good stability and separability:
 - ▶ NTrees: 400
 - ▶ Shrinkage: 0.2
 - ▶ MinNodeSize: 10%

Inclusive $ZZ \rightarrow ll\nu\nu$

- ▶ Vertex with 2 tracks with $p_T > 1$ GeV
- ▶ Two same flavour opposite-sign leptons (e+e- OR mu+mu-), leading $p_T > 30$ GeV, subleading $p_T > 20$ GeV
- ▶ Veto on any additional lepton with Loose ID and $p_T > 7$ GeV
- ▶ $76 < M_{\ell\ell} < 106$ GeV
- ▶ $E_T^{\text{miss}} > 70$ GeV.

Process	% of background	
	Strict	Loose
$WZ \rightarrow l\nu ll$ — one missing l mimics the signal topology	68%	12%
$Z(\rightarrow ll) + \text{jets}$ — lepton pair with mismeasured E_T^{miss}	15%	69%
$WW \rightarrow l\nu l'\nu'$ — non resonant production of a lepton pair	3%	3%
$Wt, t, t\bar{t}, ttV$ — non resonant production of a lepton pair via t -quark	9%	15%
Other backgrounds: $4l, llqq, VVV, Z(\tau\tau), W + \text{jets}$	5%	1%