

# 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.

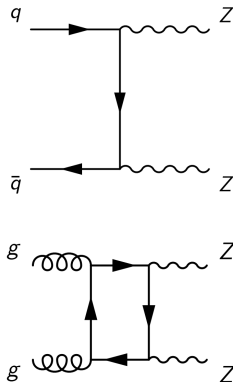
Москва, 2022



# 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;
- ▶  $76 \text{ ГэВ} < M_{ll} < 106 \text{ ГэВ}$ ;
- ▶  $E_T^{miss} > 70 \text{ ГэВ}$ .

Сигнал	
QCD ZZ (~ 18%)	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 (~ 48%)	рождение Z-бозона и струи, с распадом Z-бозона в пару заряженных лептонов и большим ложным потерянным поперечным импульсом
tt (~ 11.5%)	рождение пары топ-кварков и последующим распадом включающим конечное состояние $ll\nu\nu$ (не резонансное рождение $ll\nu\nu$ )
WZ (~ 12%)	рождение пары бозонов Z и W, с распадом Z-бозона в пару заряженных лептонов и лептонным распадом W
WW (~ 5.2%)	рождение пары W с распадом в $ll\nu\nu$ (не резонансное рождение $ll\nu\nu$ )
Wt (~ 2.8%)	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



# Оптимизация и настройка классификатора

6 различных критериев разделения:

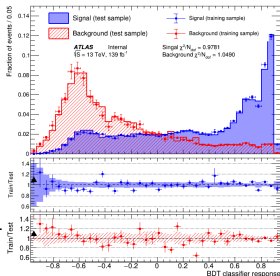
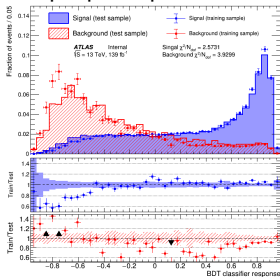
- ▶ CrossEntropy
- ▶ GinilIndex (Default)
- ▶ GinilIndexWithLaplace
- ▶ MisClassificationError
- ▶ SDivSqrtSPlusB ( $S/\sqrt{S+B}$ )
- ▶ RegressionVariance

Оптимизация гиперпараметров:

Option	Default	Best
NTrees	400	200
Shrinkage	0.1	0.3
MaxDepth	3	2
MinNodeSize	5%	0.2%

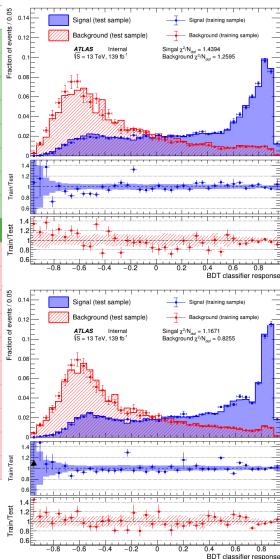
Увеличение значимости с  $44.3 \pm 0.4$  до  $46.1 \pm 0.4$

Результаты оптимизации гиперпараметров:



# 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  $46.1 \pm 0.4$  до  $46.8 \pm 0.4$

# Фит. Определение контрольных и сигнального регионов.

## Расслабленный вариант ФП:

SR	WZ(3l)	Non-resonant	Zj
MET > 70 GeV			MET > 70 GeV
m(l) in [76; 106] GeV			
dR(l) < 2.2			dR(l) < 2.2
dΦ(Z, MET) > 1.3			dΦ(Z, MET) > 1.3
B-veto			
MET/HT > 0			
METsig > 7			METsig < 7
	mT(W) > 60 GeV	3l events removed in preselection	

Фит в сигнальном регионе происходил по переменной  $BDTscore$ , в контрольных по переменной  $p_T^Z$

## Строгий вариант ФП:

SR	WZ(3l)	Non-resonant	Zj
MET > 70 GeV			MET > 70 GeV
m(l) in [76; 106] GeV			
dR(l) < 1.8			dR(l) < 1.8
dΦ(Z, MET) > 2.3			dΦ(Z, MET) > 2.3
B-veto			
MET/HT > 0.5			
METsig > 10			METsig < 9
	mT(W) > 60 GeV		

Фит в сигнальном и контрольных регионах происходил по переменной  $p_T^Z$

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

Фит в контрольных регионах происходил на экспериментальных данных, в сигнальном регионе на данных Азимова



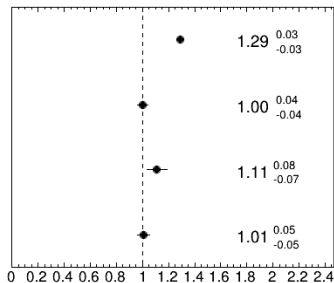
# Фит. Результаты.

## Значения нормфакторов:

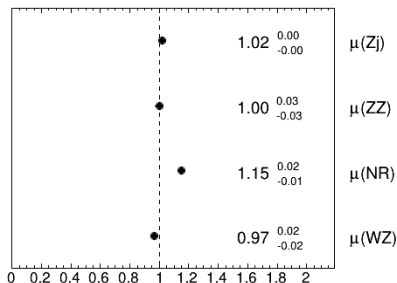
Строгий вариант ФП

Расслабленный вариант ФП

*ATLAS* Internal



*ATLAS* Internal



Значимость: 23.0

Значимость: 36.5



## Заключение:

- ▶ Оптимизировались различные параметры классификатора BDTG
- ▶ В ходе оптимизации получен стабильный алгоритм
- ▶ По форме распределения отклика классификатора производился фит в сигнальном регионе
- ▶ С фитом по отклику классификатора удалось достичь бОльшей значимости ( $23.0 \rightarrow 36.5$ ) и уменьшить стат. ошибку для  $\mu(ZZ)$  ( $0.04 \rightarrow 0.03$ ) по сравнению с фитом по переменной  $p_T^Z$ .

## Дальнейшие планы:

- ▶ Добавить систематические погрешности в фит
- ▶ Использовать в фите оценки фонов, оцениваемых data driven методами

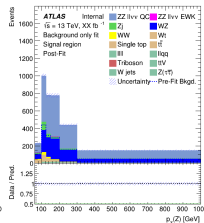
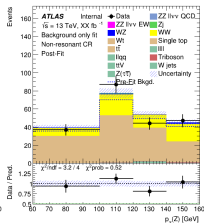
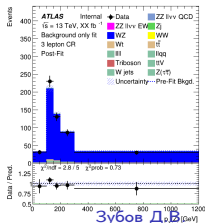
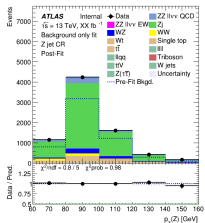
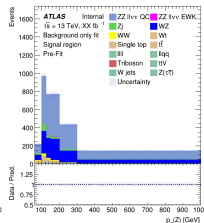
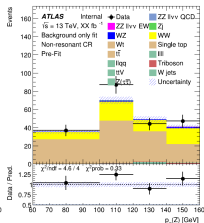
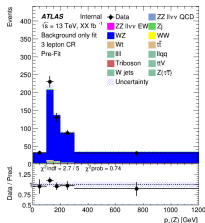
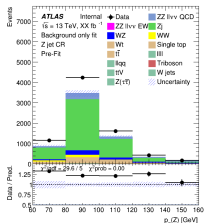
backup

# 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.

# Фит. Распределения до и после для строгого варианта ФП.

- ▶ Сначала фит производился в контрольных регионах на данных для определения нормфакторов основных фоновых процессов.
- ▶ Затем с полученными нормфакторами производился фит во всех четырех регионах для определения нормфактора для сигнала



## Initial selection:

- ▶ All\_Good GRL events
- ▶ 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
- ▶  $76 < M_{ll} < 106$  GeV
- ▶ MET > 70 GeV.

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

Signal	
QCD ZZ	345666, 345723
EWK ZZ	363724
Background	
Z+jets, $Z_{\tau\tau}$	364100-364141
WZ	363358, 364253, 364284
TOP	410472, 410644, 410645, 410658, 410659
WW	361600, 361606, 345718
Wt	410648, 410649
llll	345071, 345706, 364250
VVV	364242-364249, 363507-363509
llqq	363356
ttV	410081, 410155-410157

	Strong	Relax
Signal		
QCD ZZ	1914±16	4409±23
EWK ZZ	(4.49±0.04) × 3	(16.38±0.07) × 3
Total signal	1927±16	4458±23
Background		
Zj	155±19	12184±290
WZ	933±8	3116±15
TOP	125±2	2829±11
WW	62.6±1.4	1356±7
Wt	37±2	729±10
llll	47.2±1.0	170.5±1.5
VVV	12.93±0.16	38.2±0.3
llqq	0.9±0.3	35.9±1.3
ttV	0.77±0.11	4.46±0.26
Total bkg.	1378±21	20540±292
Significance	43.9±0.4	30.1±0.3