



Национальный исследовательский ядерный университет  
«МИФИ»

Кафедра физики элементарных частиц №40



Научная исследовательская работа студента на тему:

# Применение методов машинного обучения для разделения электрослабого и КХД процессов рождения $Z$ -бозона с фотоном

Научные руководители  
к.ф.-м.н., доцент  
Солдатов Евгений Юрьевич

Петухов Александр Максимович

г. Москва 2021

Работа  
студента 4-ого курса  
Савельева Константина  
Михайловича  
ИЯФит

# Введение

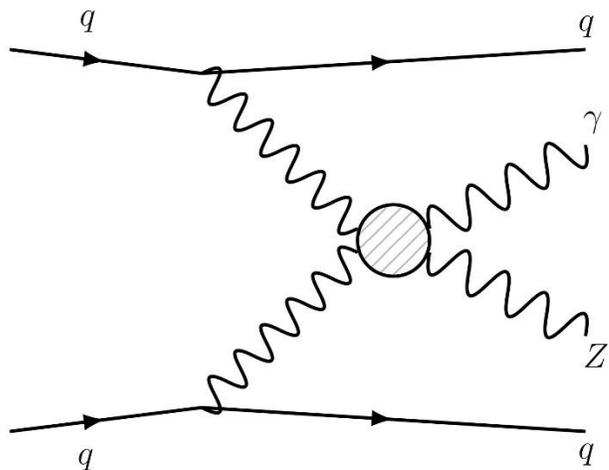


Диаграмма процесса рассеяния векторных бозонов

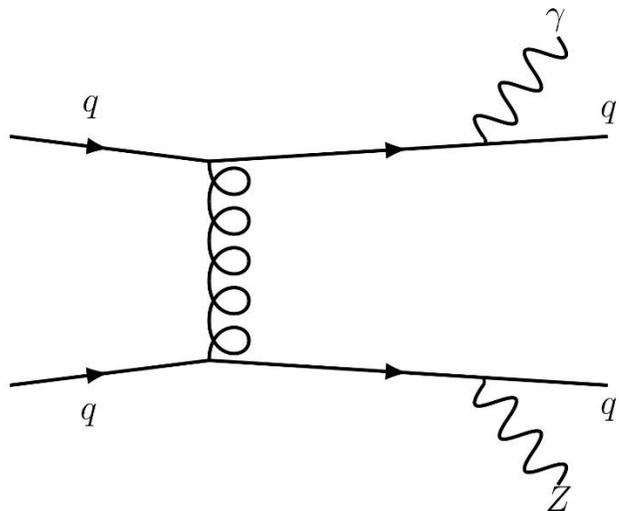


Диаграмма процесса КХД образования состояния  $Z\gamma jj$

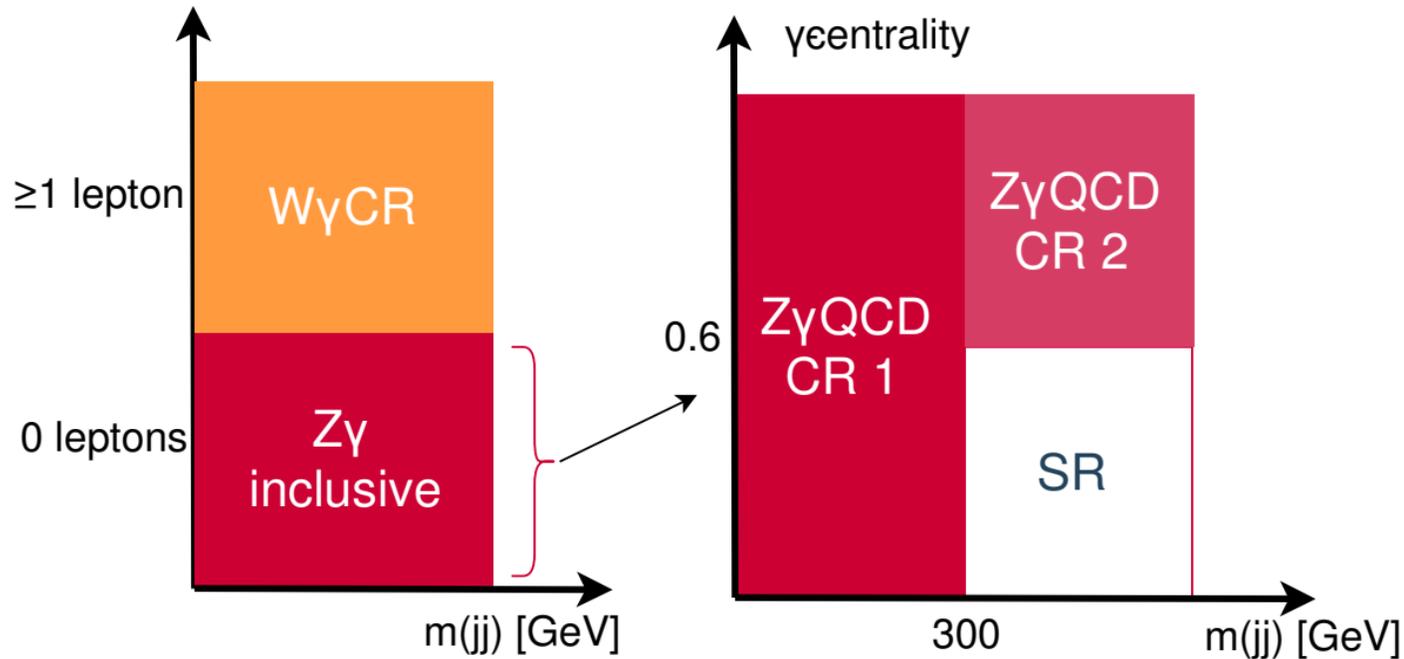
- Целью работы является изучение редкого процесса рассеяния векторных бозонов с рождением Z-бозона, фотона и двух адронных струй с последующим распадом Z-бозона на нейтрино и антинейтрино.
- Подобные процессы интересны с точки зрения поиска «новой физики» из-за их высокой чувствительности к отклонениям параметров от Стандартной модели.
- Выделение этого процесса является сложной задачей из-за высокого сечения основного фонового процесса – КХД образования идентичного конечного состояния.
- Для вычисления сечения исследуемого процесса с достаточной точностью необходимо эффективно отделять сигнальные события от фоновых. Классические одномерные фиксированные отборы не дают достаточной значимости, поэтому в работе исследовалось применение алгоритмов машинного обучения к разделению событий

Значимость 
$$\sigma = \frac{S}{\sqrt{S+B}}$$

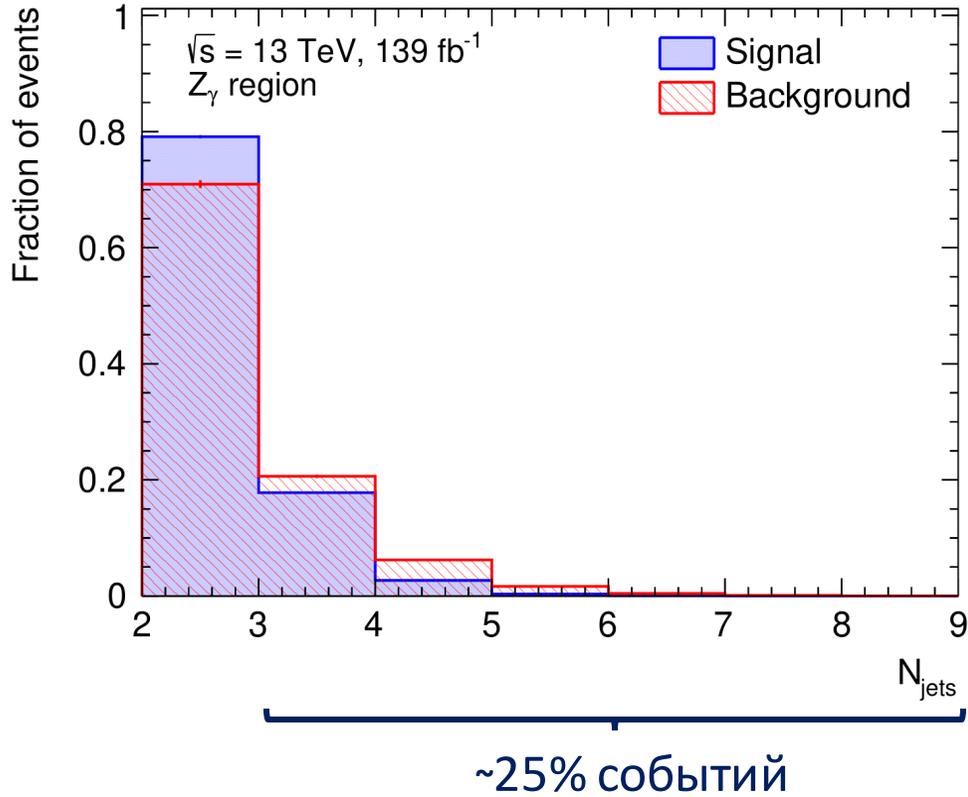
# Используемые данные

Работа производилась с Монте-Карло моделированными данными протон-протонных столкновений в детекторе ATLAS на БАК с энергией 13ТэВ и интегральной светимостью 139 фб-1 и реальными данные, набранными в течении 2015-2018 гг.

Используется композиция деревьев решения, созданная с помощью градиентного бустинга (Boosted Decision Trees, BDT) из библиотеки машинного обучения LightGBM. Она поддерживает работу с отсутствующими значениями.



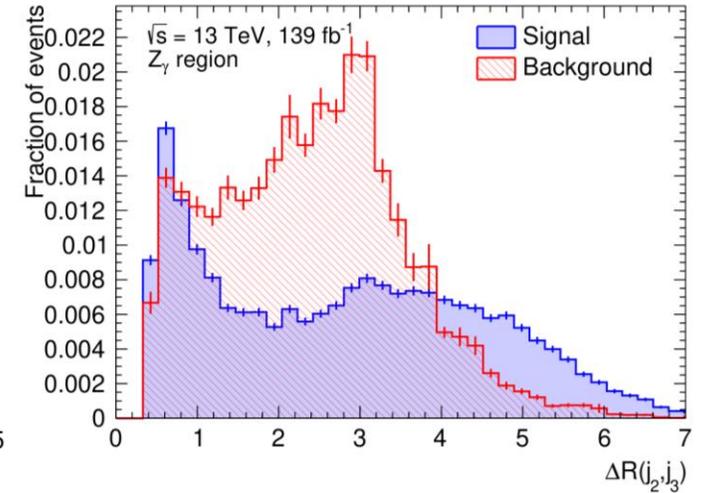
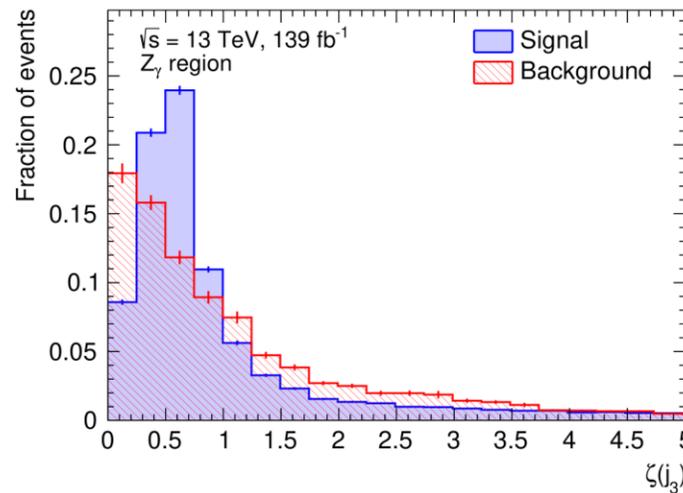
# Работа с переменными третьей струи



- $V_0$  - без использования параметров третьей струи

Стратегии по работе с отсутствующими значениями:

- $V_1$  - разделение событий на две категории: с  $N_{\text{jets}} = 2$  и  $N_{\text{jets}} > 2$
- $V_2$  - использование фиксированных значений параметров третьей струи в событиях с  $N_{\text{jets}} = 2$
- $V_3$  - использование встроенных алгоритмов работы с отсутствующими значениями



$$\zeta(x) = \left| \frac{\eta_x - \frac{\eta_{j_1} + \eta_{j_2}}{2}}{\eta_{j_1} - \eta_{j_2}} \right|$$

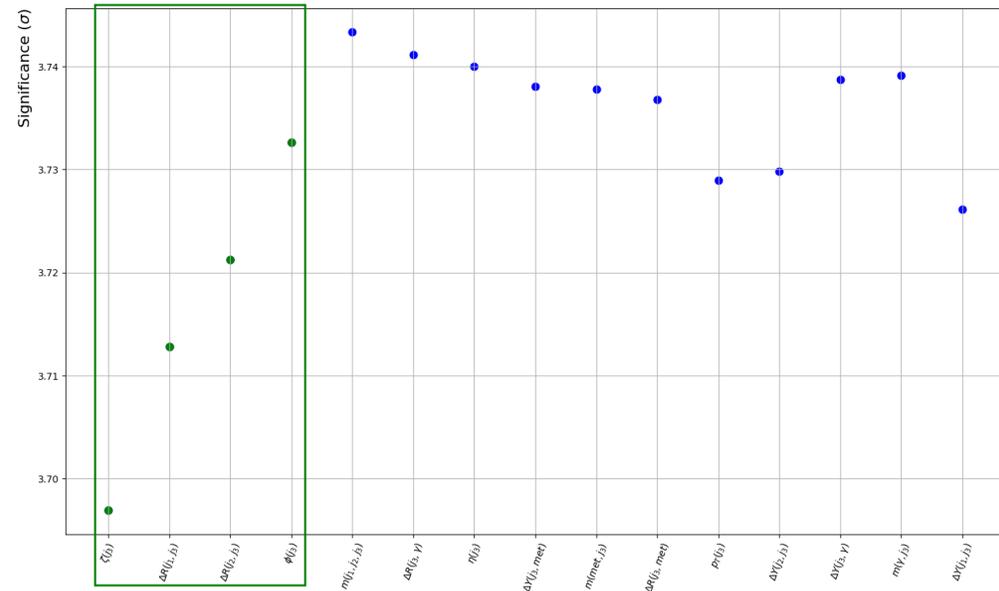
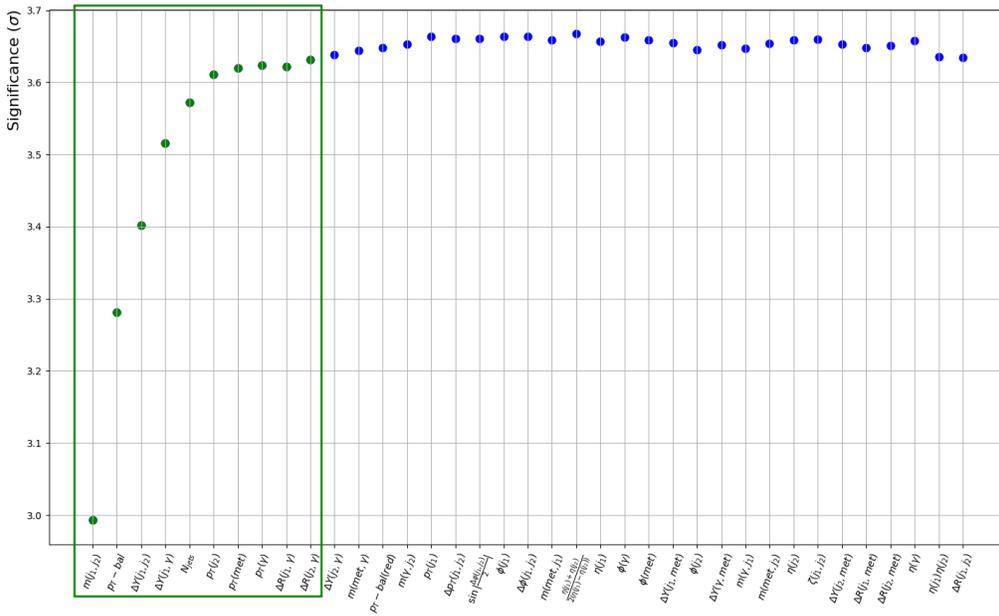
$$\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\varphi)^2}$$

# Отбор переменных для обучения

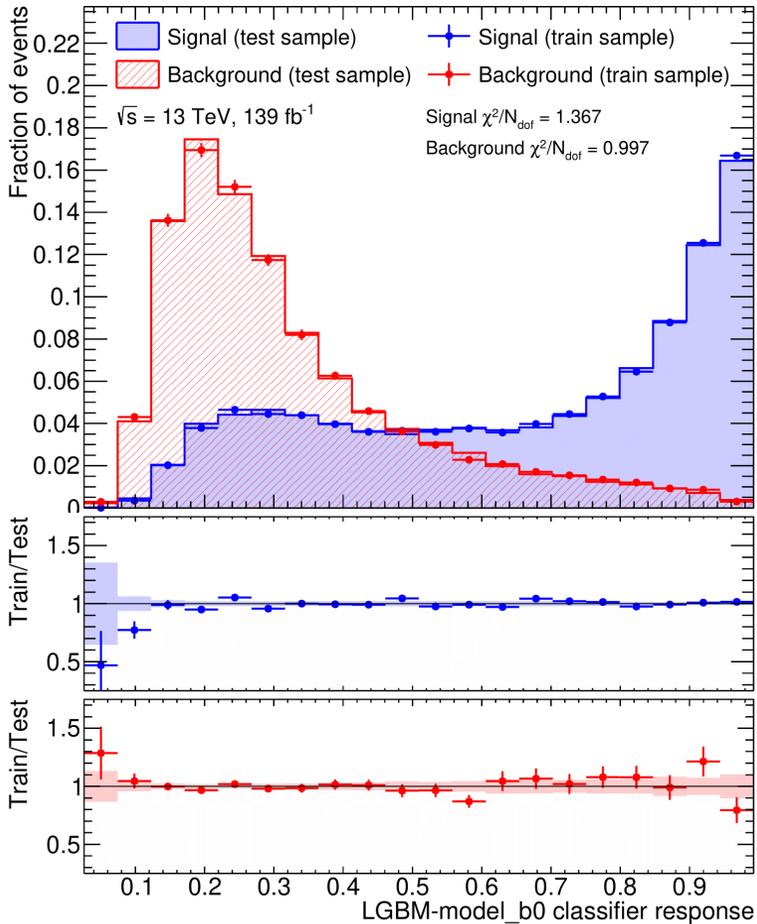
Переменные отбирались методом "N+1".

Сначала отбирались переменные, не использующие параметры третьей струи, затем были дополнительно отобраны переменные третьей струи. Использовались настройки классификаторов по умолчанию, а для подбора переменных третьей струи использовалась стратегия  $V_3$ .

- $m_{jj}$
- $p_T - \text{balance}$
- $\Delta Y(j_1, j_2)$
- $\Delta Y(j_1, \gamma)$
- $N_{\text{jets}}$
- $p_T(j_2)$
- $E_T^{\text{miss}}$
- $p_T(\gamma)$
- $\Delta R(j_1, \gamma)$
- $\zeta(j_3)$
- $\Delta R(j_1, j_3)$
- $\Delta R(j_2, j_3)$
- $\varphi(j_3)$



# Оптимизация настроек классификаторов



Стратегия	Кол-во сигнала	Кол-во фона	Значимость, $\sigma$
$B_0$	$26.57 \pm 0.13$	$28.1 \pm 0.9$	$3.60 \pm 0.03$
$B_1$	$28.69 \pm 0.13$	$31.2 \pm 1.0$	$3.71 \pm 0.04$
$B_2$	$26.92 \pm 0.13$	$26.2 \pm 0.8$	$3.70 \pm 0.03$
$B_3$	$28.50 \pm 0.13$	$31.5 \pm 0.9$	$3.68 \pm 0.03$

При использовании параметров третьей струи виден прирост значимости, однако значимости при использовании различных стратегий практически неразличимы.

Однако, использование стратегии  $B_1$  требует создания и оптимизации настроек для двух классификаторов. Использование стратегии  $B_2$  требует задания фиксированных значений для событий с  $N_{\text{jets}} = 2$ , причём значение зависит от переменной.

Поэтому предпочтение отдаётся использованию стратегии  $B_3$ .

# Заключение

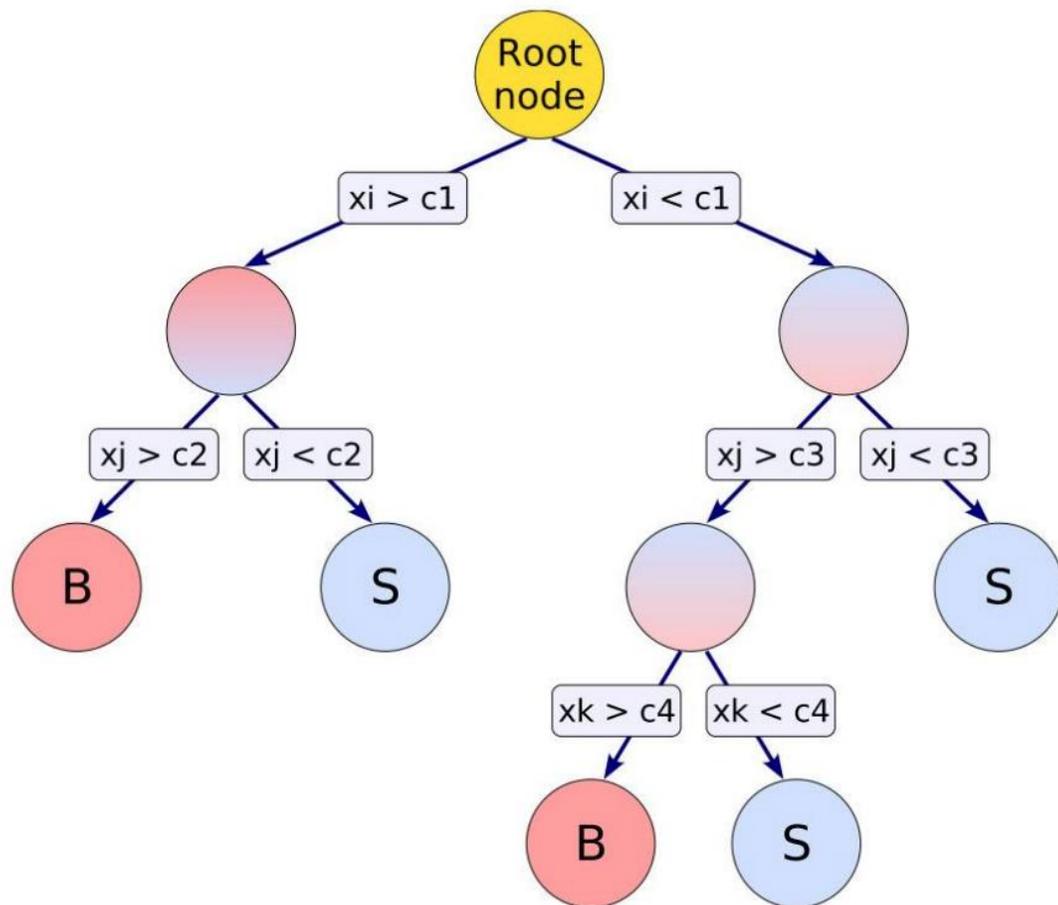
- Отобраны лучшие переменные для обучения классификаторов;
- Было произведено сравнение стратегий использования параметров третьей струи;
- В качестве наилучшей стратегии была выбрана стратегия  $V_3$ .

В дальнейшем планируется начать поиск новых переменных для обучения классификаторов исходя из теоретических соображений о физике процесса.

Спасибо за внимание!

# Back-up

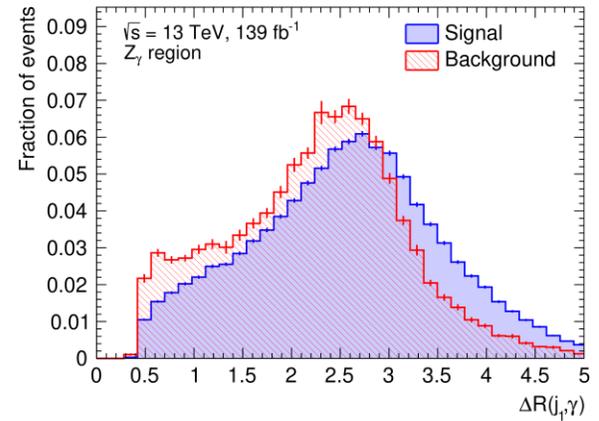
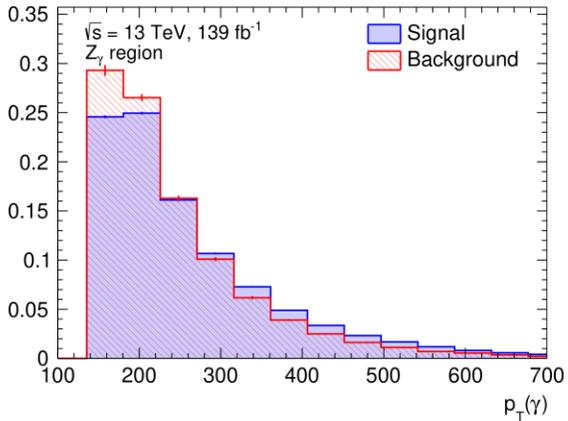
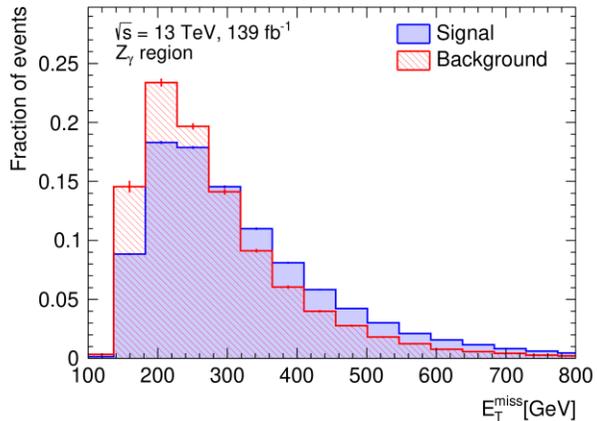
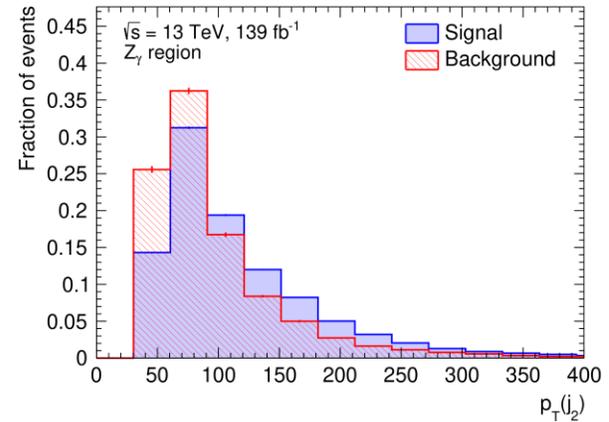
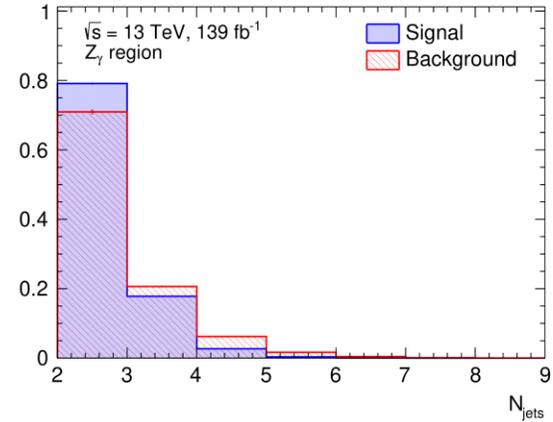
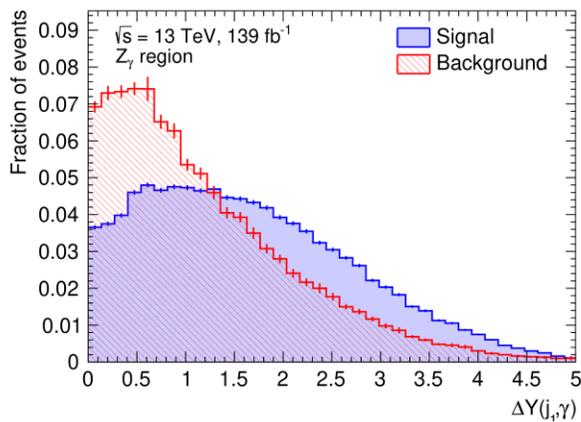
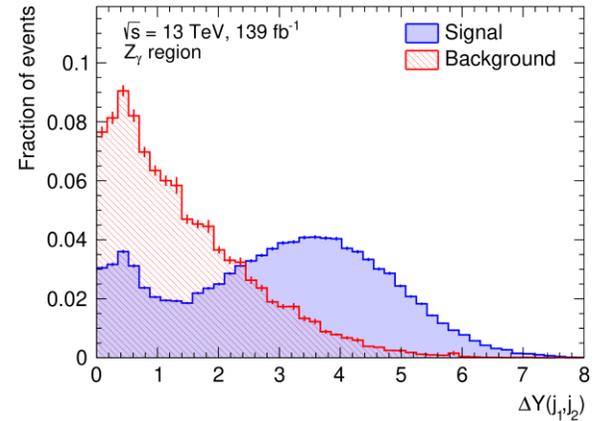
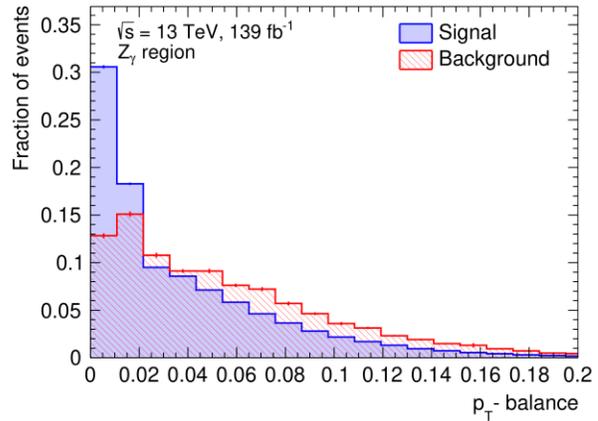
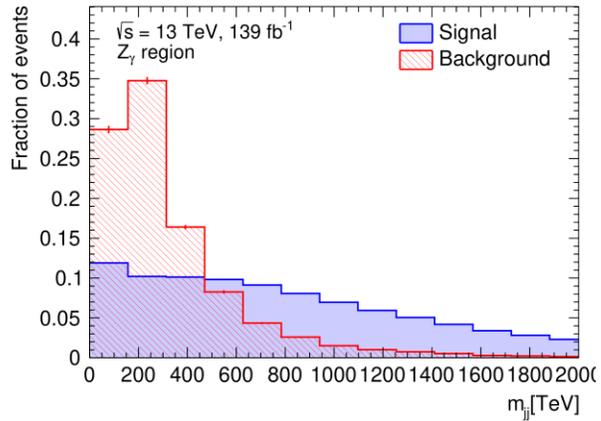
# Композиция деревьев решений

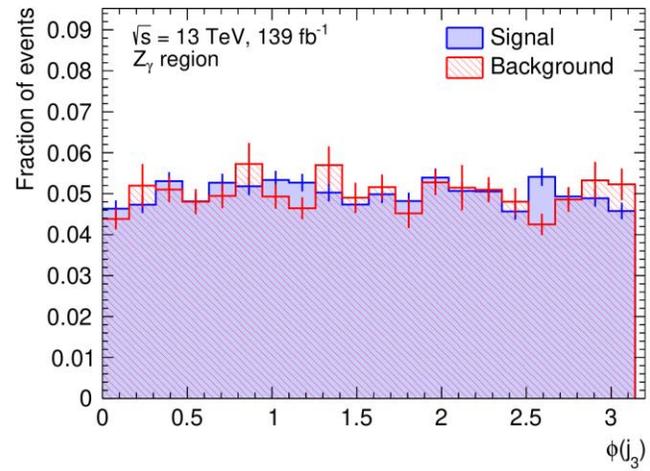
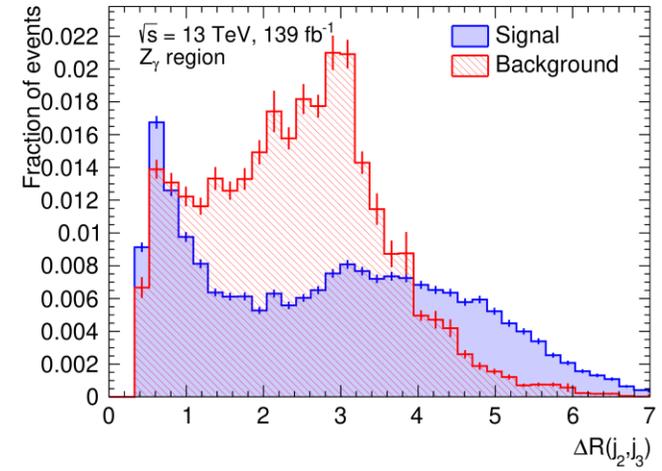
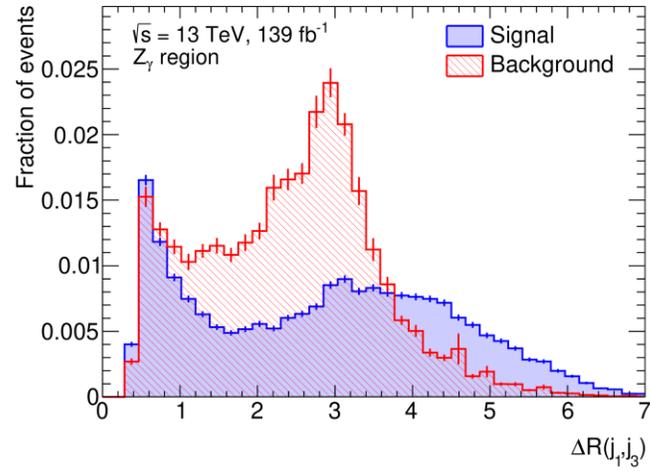
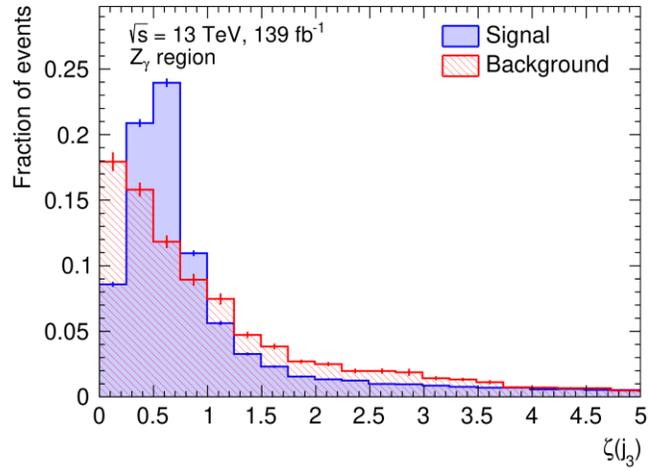


Композиция деревьев решения, созданная с помощью градиентного бустинга (Boosted Decision Trees, BDT) – это классификатор с бинарной древовидной структурой, позволяющий разбивать фазовое пространство на множество областей.

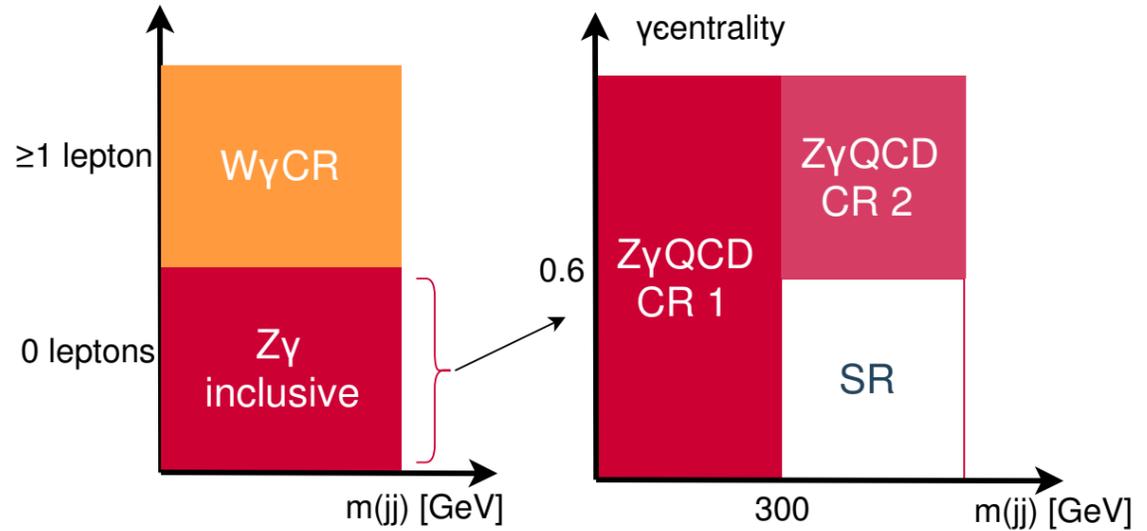
Основные характеристики метода:

- небольшое время обучения
- не склонен к переобучению из-за
- не требует подготовки переменных

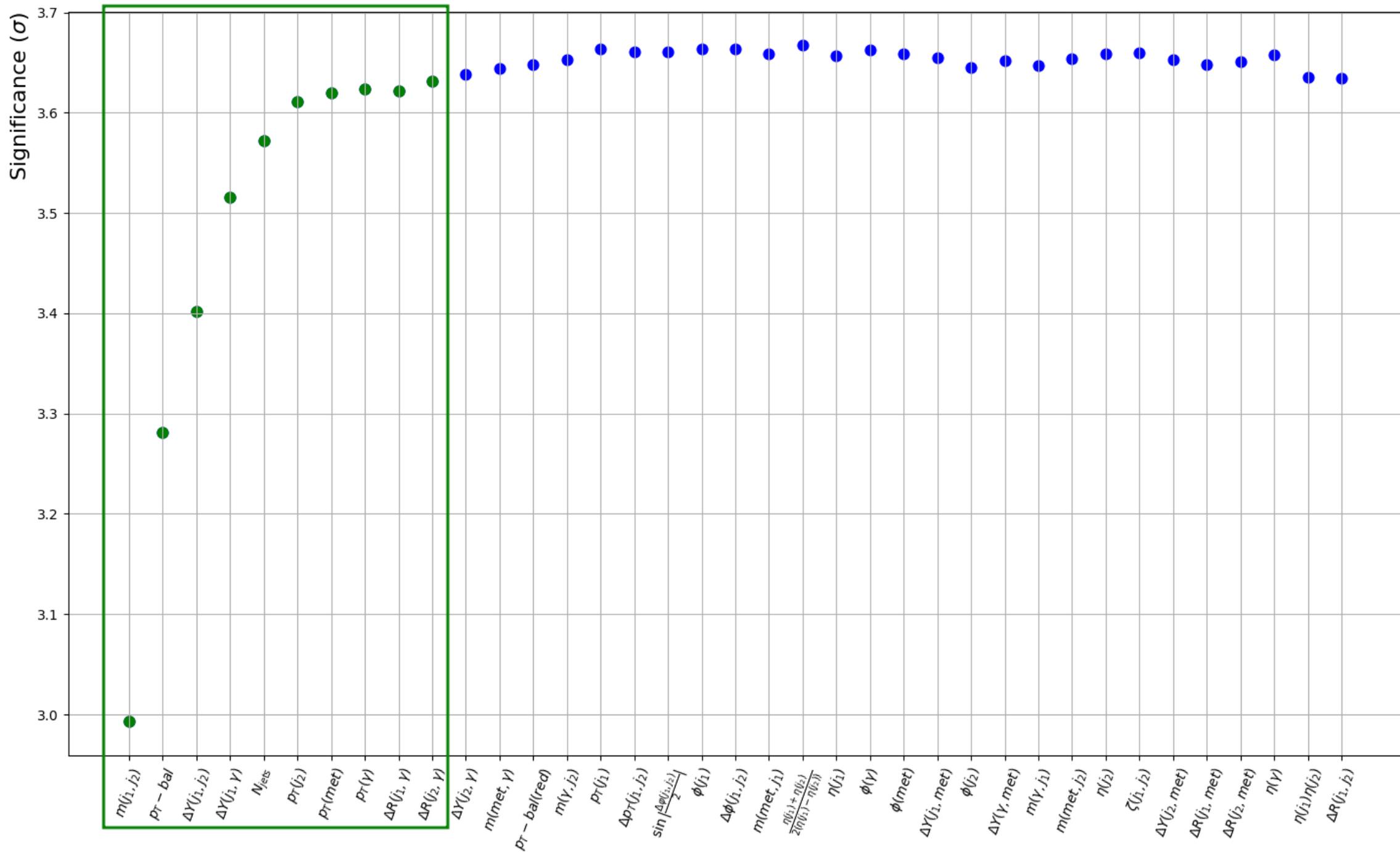


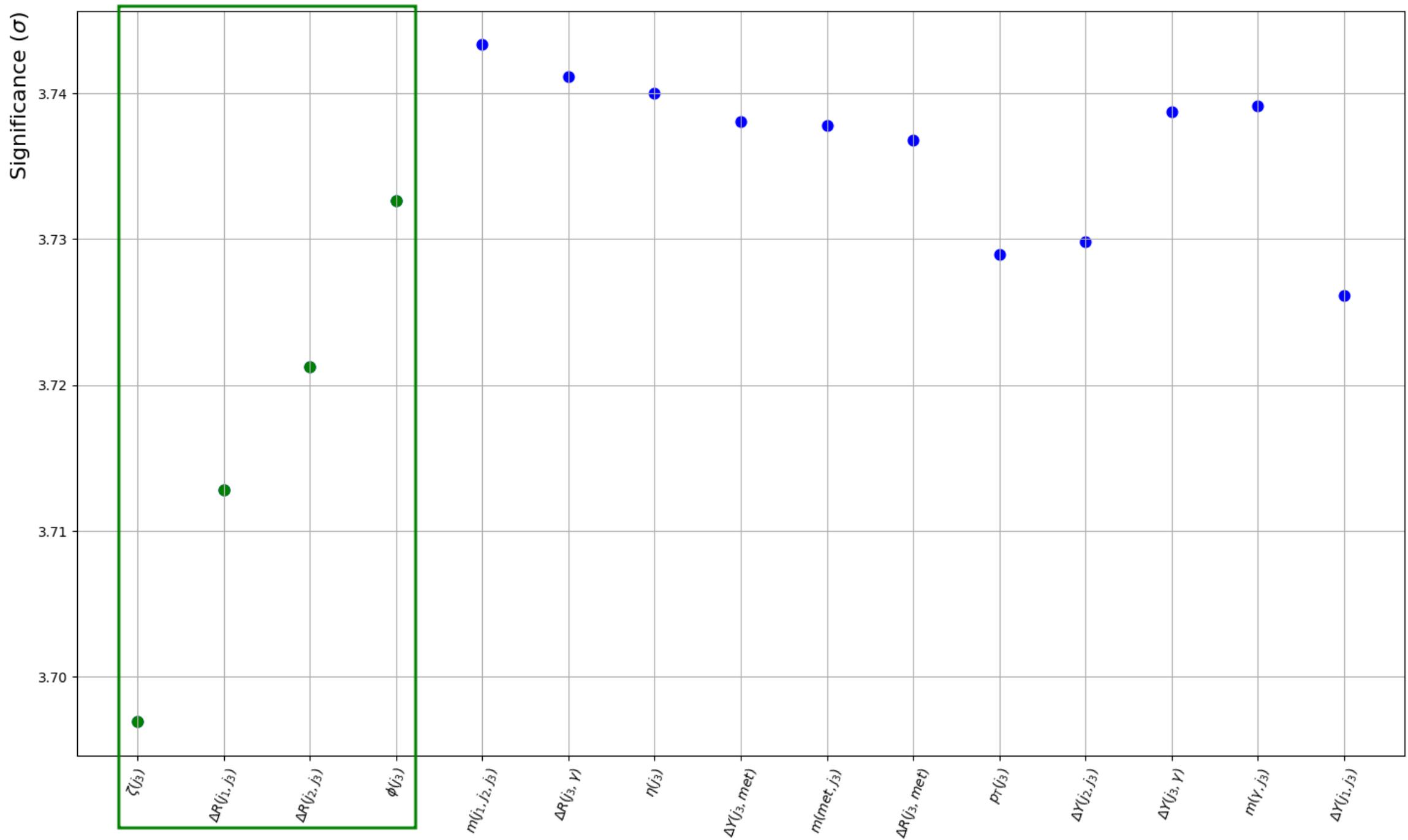


$W\gamma$ control region	
$N_{\text{leptons}}$	$\geq 1$
$Z\gamma jj$ QCD control region 1	
$N_{\text{leptons}}$	$= 0$
$m_{jj}$	$< 300 \text{ GeV}$
$Z\gamma jj$ QCD control region 2	
$N_{\text{leptons}}$	$= 0$
$m_{jj}$	$> 300 \text{ GeV}$
$\gamma$ -centrality	$> 0.6$
$Z\gamma jj$ EWK signal region	
$N_{\text{leptons}}$	$= 0$
$m_{jj}$	$> 300 \text{ GeV}$
$\gamma$ -centrality	$< 0.6$

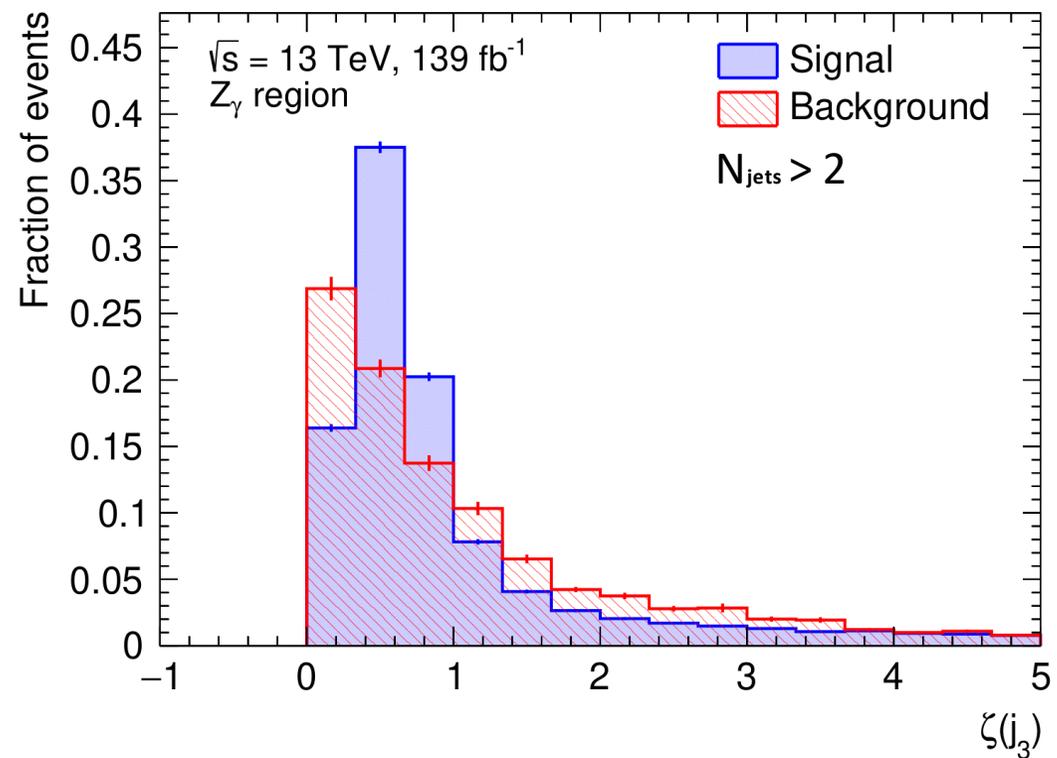
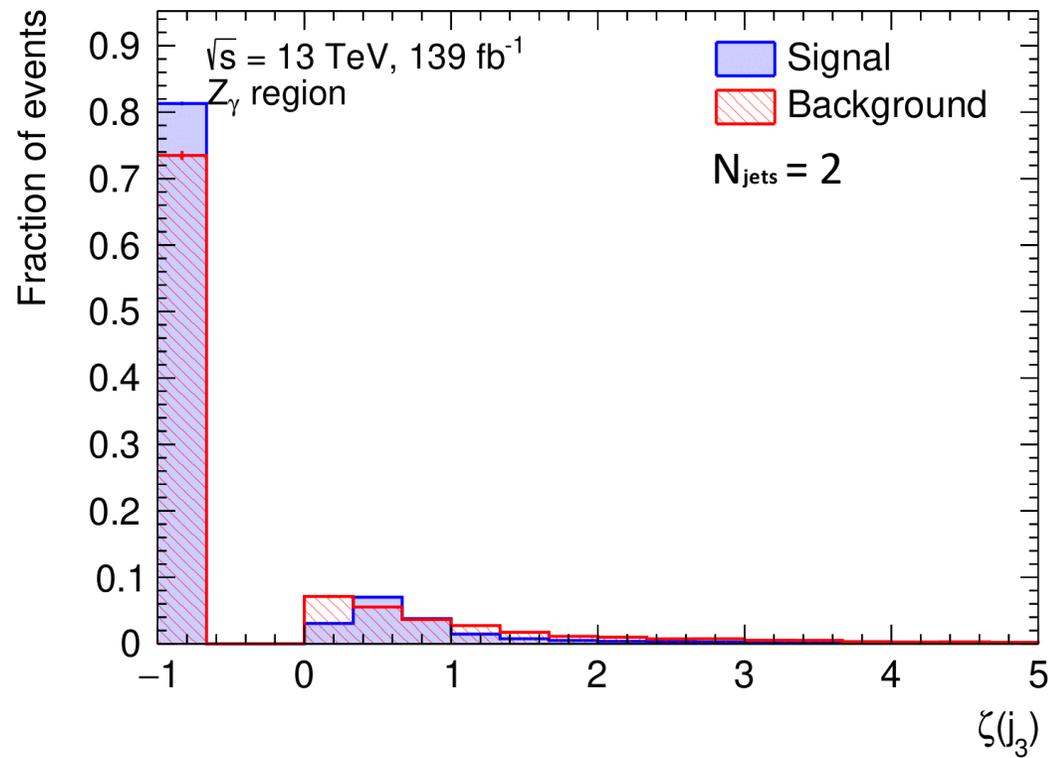


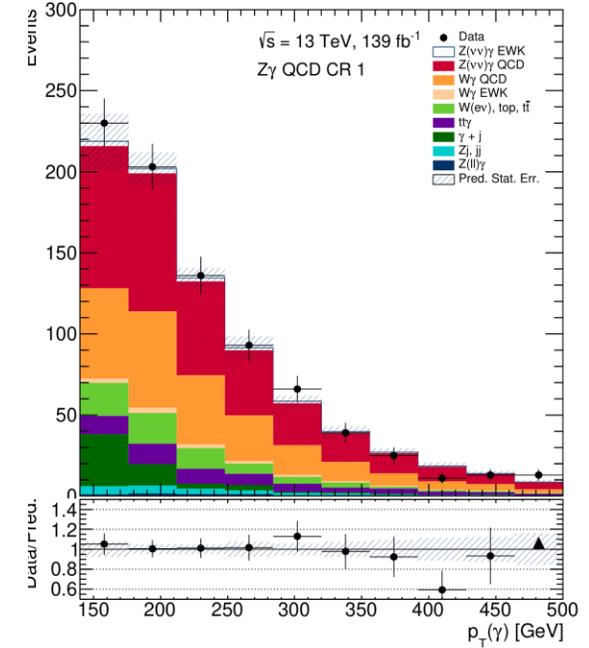
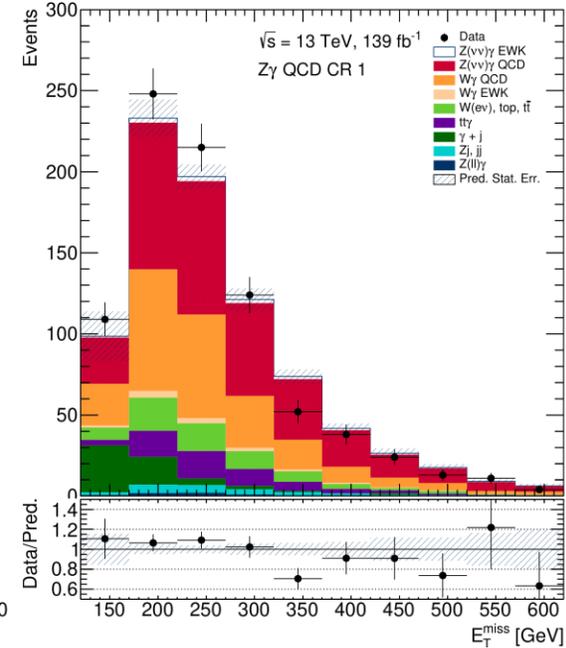
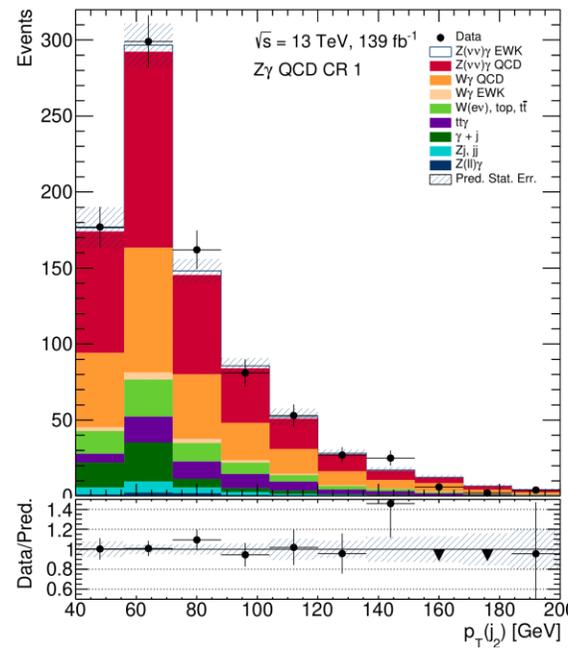
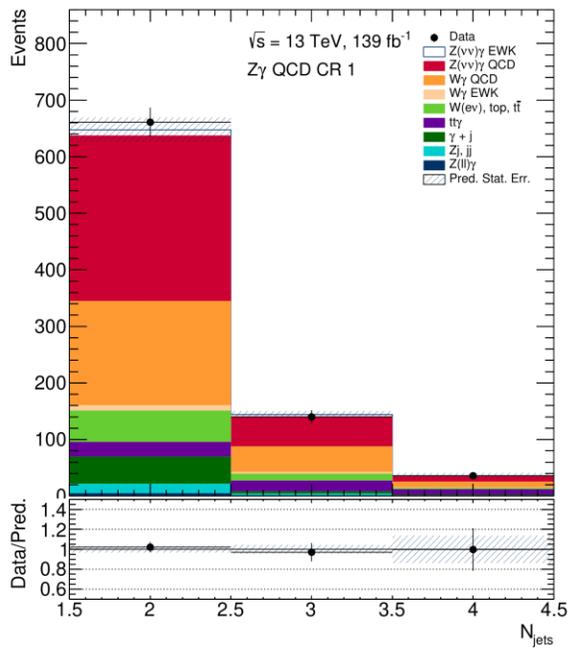
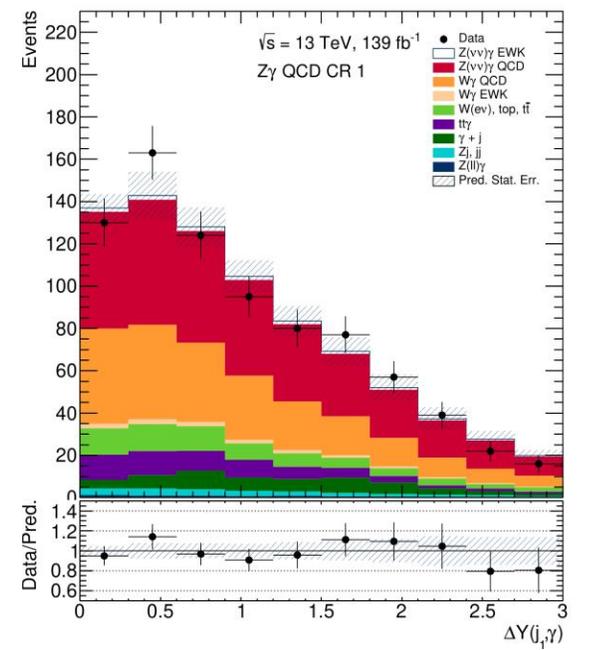
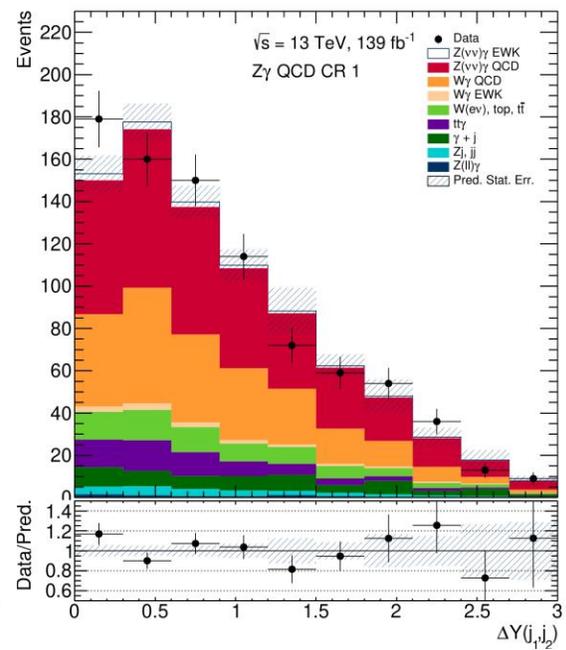
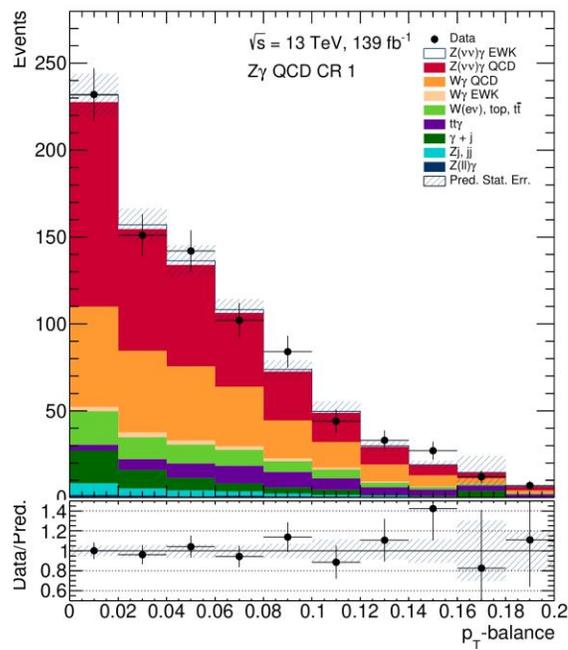
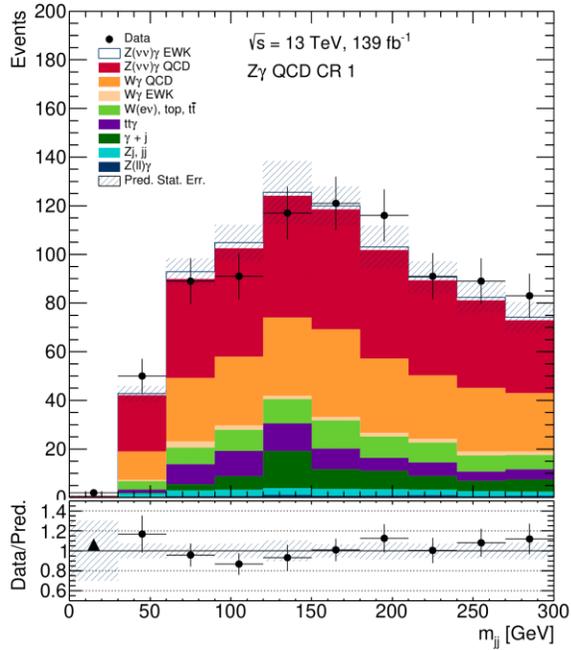
Selections	Cut Value
$E_{\text{T}}^{\text{miss}}$	$> 120 \text{ GeV}$
$E_{\text{T}}^{\gamma}$	$> 150 \text{ GeV}$
Number of tight isolated photons	$N_{\gamma} = 1$
Number of jets	$N_{\text{jets}} \geq 2$
Lepton veto	$N_{\text{e}} = 0, N_{\mu} = 0$
$E_{\text{T}}^{\text{miss}}$ significance	$> 12$
$ \Delta\phi(\gamma, \vec{p}_{\text{T}}^{\text{miss}}) $	$> 0.4$
$ \Delta\phi(j_1, \vec{p}_{\text{T}}^{\text{miss}}) $	$> 0.3$
$ \Delta\phi(j_2, \vec{p}_{\text{T}}^{\text{miss}}) $	$> 0.3$
$p_{\text{T}}^{\text{SoftTerm}}$	$< 16 \text{ GeV}$

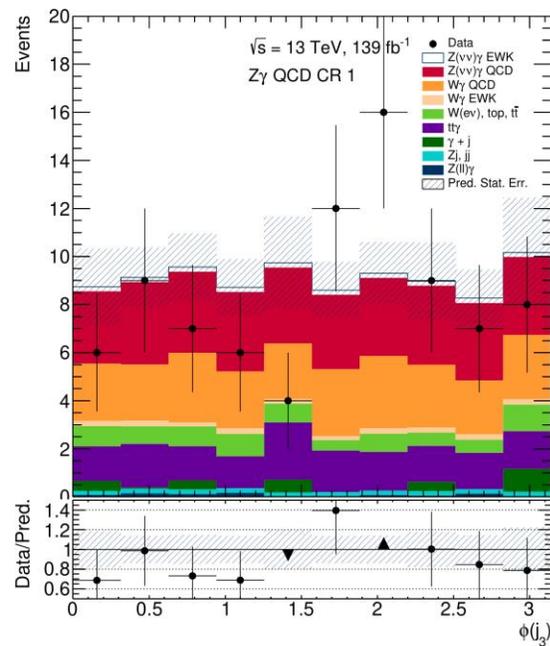
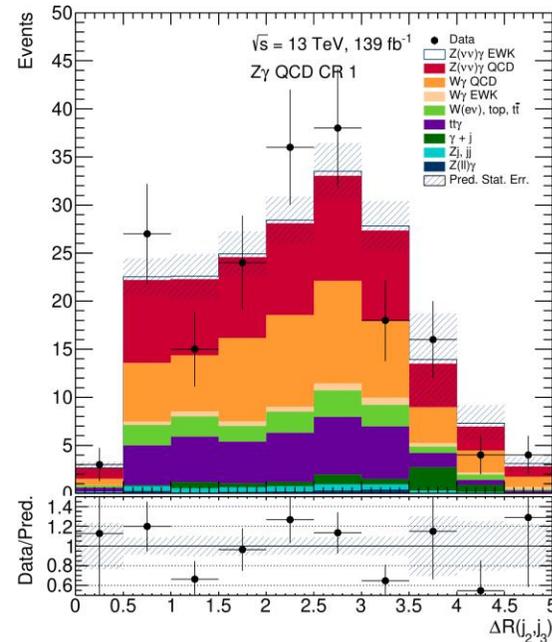
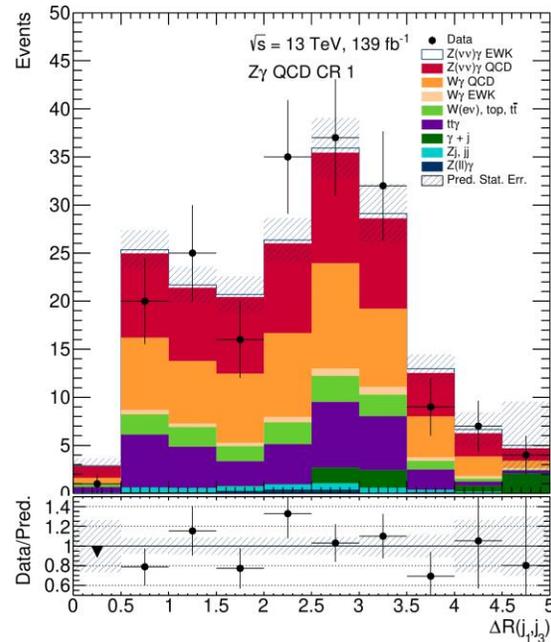
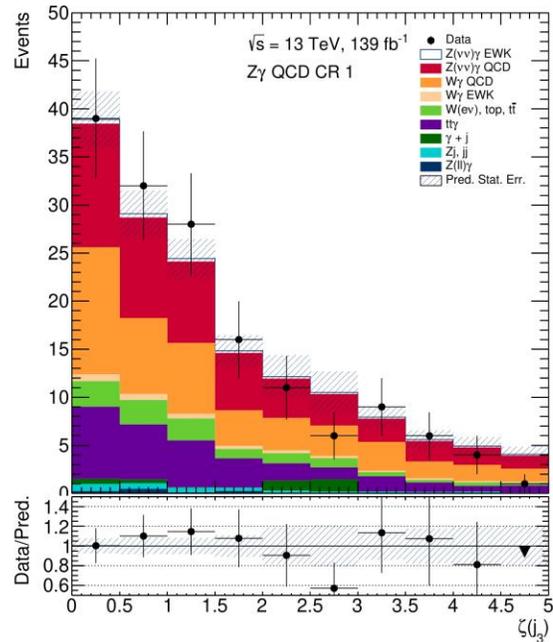
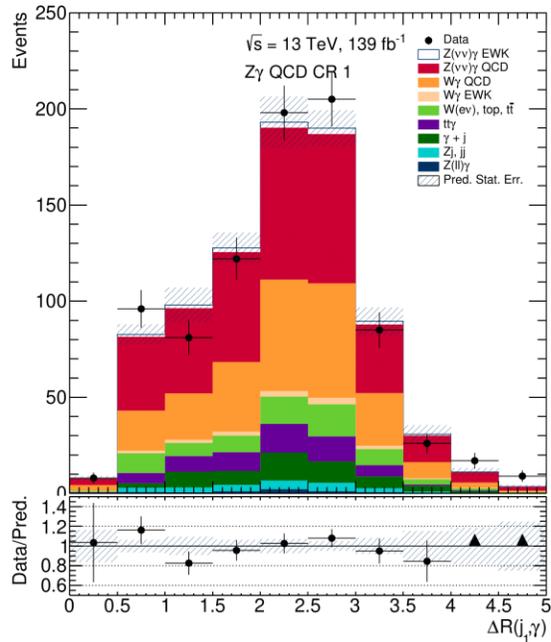




- $m_{jj}$  – инвариантная масса двух струй
- $p_T$  – balance =  $\frac{|\vec{p}_T^{miss} + \vec{p}_T^\gamma + \vec{p}_T^{j_1} + \vec{p}_T^{j_2}|}{E_T^{miss} + E_T^\gamma + p_T^{j_1} + p_T^{j_2}}$  – баланс поперечных импульсов
- $\Delta Y(j_1, j_2)$  – разность быстрот двух струй
- $\Delta Y(j_1, \gamma)$  – разность быстрот лидирующей струи и фотона
- $N_{jets}$  – число адронных струй
- $p_T(j_2)$  – поперечный импульс второй струи
- $E_T^{miss}$  – недостающий поперечный импульс
- $p_T(\gamma)$  – поперечный импульс фотона
- $\Delta R(j_1, \gamma)$  – угловое расстояние между лидирующей струёй и фотоном
- $\zeta(j_3)$  – центральность третьей струи
- $\Delta R(j_1, j_3)$  – угловое расстояние между лидирующей струёй и третьей струёй
- $\Delta R(j_2, j_3)$  – угловое расстояние между второй струёй и третьей струёй
- $\varphi(j_3)$  – азимутальный угол вылета третьей струи







	Вхожд. сигнал	Вхожд. фон	Кол-во сигнала	Кол-во фона	$\sigma_{\max}$
$N_{\text{jets}} > 1$	90035	86902	$46.7 \pm 0.2$	$304.9 \pm 4.60$	$2.49 \pm 0.02$
$N_{\text{jets}} = 2$	74237	60094	$38.6 \pm 0.2$	$195.2 \pm 3.7$	$2.52 \pm 0.02$
$N_{\text{jets}} > 2$	15798	26808	$8.12 \pm 0.07$	$109.7 \pm 2.8$	$0.75 \pm 0.01$
LightGBM					
$B_0$	51119	9212	$26.57 \pm 0.13$	$28.1 \pm 0.9$	$3.60 \pm 0.03$
$B_1^{N_{\text{jets}}=2}$	48505	8700	$25.21 \pm 0.12$	$26.0 \pm 1.0$	$3.49 \pm 0.04$
$B_1^{N_{\text{jets}}>2}$	6694	1374	$3.48 \pm 0.05$	$4.4 \pm 0.3$	$1.24 \pm 0.03$
$B_1$	55199	10074	$28.69 \pm 0.13$	$31.2 \pm 1.0$	$3.71 \pm 0.04$
$B_2$	51790	8494	$26.92 \pm 0.13$	$26.2 \pm 0.8$	$3.70 \pm 0.03$
$B_3$	54833	10105	$28.50 \pm 0.13$	$31.5 \pm 0.9$	$3.68 \pm 0.03$