

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ
ВЫСШЕГО ОБРАЗОВАНИЯ
«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ ЯДЕРНЫЙ УНИВЕРСИТЕТ МИФИ»
(НИЯУ МИФИ)

УДК 539.120.71

ОТЧЁТ
О НАУЧНО-ИССЛЕДОВАТЕЛЬСКОЙ РАБОТЕ

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

Научный руководитель

к.ф.-м.н., доцент

_____ Е. Ю. Солдатов

Научный руководитель

_____ А. М. Петухов

Студент

_____ К. М. Савельев

Москва 2021

Содержание

Введение	2
1 Методы	5
1.1 Композиция деревьев решений	5
1.1.1 Принцип построения дерева решений	5
1.1.2 Бустинг	5
1.1.3 Оценка результатов работы алгоритма	6
1.2 Библиотека LGBM	7
1.3 Отбор переменных	7
1.4 Подбор настроек классификаторов	8
1.5 Использование переменных третьей струи	8
2 Используемые данные	10
2.1 Устройство детектора	10
2.2 Исходные данные	11
3 Процесс работы и результаты	15
3.1 Отбор переменных	15
3.2 Тренировка и применение моделей	16
Заключение	18
Список используемых источников	19
A Распределения по переменным	20
B Проверка моделирования переменных	23
B.1 Первая $Z\gamma$ QCD контрольная область	23
B.2 Вторая $Z\gamma$ QCD контрольная область	26
B.3 $W\gamma$ контрольная область	29

Введение

Стандартная модель (СМ) даёт достаточно точные качественные и количественные предсказания свойств элементарных частиц и их взаимодействий. Однако, существует ряд явлений, которые не могут быть описаны в рамках Стандартной модели. Одни из самых известных: явление осцилляций нейтрино, из которого следует наличие у него массы, что противоречит СМ; тёмная материя, косвенные признаки наличия которой наблюдаются в большом масштабе в астрономических наблюдениях; факт барионной асимметрии. Всё это свидетельствует о том, что современная СМ не является всеобъемлющей. Это даёт стимул к поискам отклонений, которые могут привести к открытию более совершенной модели для описания взаимодействий элементарных частиц.

Целью исследования является поиск отклонений от предсказаний СМ при рассеянии векторных бозонов, $VV \rightarrow VV$, где $V = W/Z/\gamma$. Для исследования был выбран высокочувствительный к отклонениям СМ процесс электрослабого рождения Z -бозона, фотона совместно с двумя адронными струями ($Z\gamma jj$) с последующим распадом Z -бозона на нейтрино и антинейтрино.

Выбор нейтрального канала распада связан с его достаточно большой вероятностью (20%) [1] и возможностью отделения сигнала в отличии от распада по адронному каналу, вероятность которого составляет около 70%. Лептонный канал распада не рассматривался из-за его сравнительно низкой вероятности ($\sim 6.7\%$).

Этот процесс невозможно отделить от других электрослабых процессов с тем же конечным состоянием. Поэтому его изучение возможно только посредством рассмотрения всех процессов электрослабого образования конечного состояния $Z\gamma jj$. Они включают в себя как процессы рассеяния векторных бозонов, чувствительных к изменениям параметров Стандартной модели, так и прочие электрослабые процессы. Пример диаграммы процесса рассеяния представлен на рисунке 1а. Кинематические параметры частиц в конечном состоянии позволяют отделить их от КХД процессов

с тем же конечным состоянием, которые являются основным фоном наряду с экспериментальными фонами и обладают сечением почти в сотню раз превышающим фон изучаемого процесса. Пример диаграммы электрослабого процесса, не являющегося рассеянием, представлен на рисунке 1в, а пример диаграммы КХД процесса представлен на рисунке 1б.

Применение одномерных фиксированных отборов при отделении сигнальных событий от фона при изучении процессов электрослабого рождения Z -бозона, фотона и двух адронных струй не позволяет достаточно хорошо отделить сигнал от фона, из-за чего точность измерения сечения получается низкой, а измерение – неточным. Поэтому проводилось исследование использования алгоритма машинного обучения «Композиция деревьев решений» (Boosted Decision Trees, BDT) [2]. Были изучены подходы к использованию параметров третьей струи по величине поперечного импульса в событии и были установлены оптимальные настройки моделей машинного обучения. Также была проведена проверка соответствия данных, сгенерированных с помощью метода Монте-Карло (МК), реальным данным.

Эффективное разделение сигнальных и фоновых событий даст возможность определить величину сечения исследуемого процесса с большей точностью.

В предыдущем исследовании, опубликованном коллаборацией ATLAS, использовались данные столкновений с энергией в системе центра масс 8 ТэВ и интегральной светимостью 20.3 фб^{-1} , из-за недостаточной чувствительности оно было выполнено только для поиска аномальных вершин [3]. Это исследование нацелено на увеличение значимости измерения сечения процесса. Достижение порога значимости в 5σ позволит подтвердить наблюдение исследуемого процесса. На данный момент опубликовано сечение процесса $Z(\ell\ell)\gamma jj$ со значимостью 4.7σ [4]. Результаты измерения параметров этого процесса могут использоваться для поиска аномалий в вершине $WWZ\gamma$, а также вершин $ZZZ\gamma$, $ZZ\gamma\gamma$ и $Z\gamma\gamma\gamma$, запрещённых в СМ.

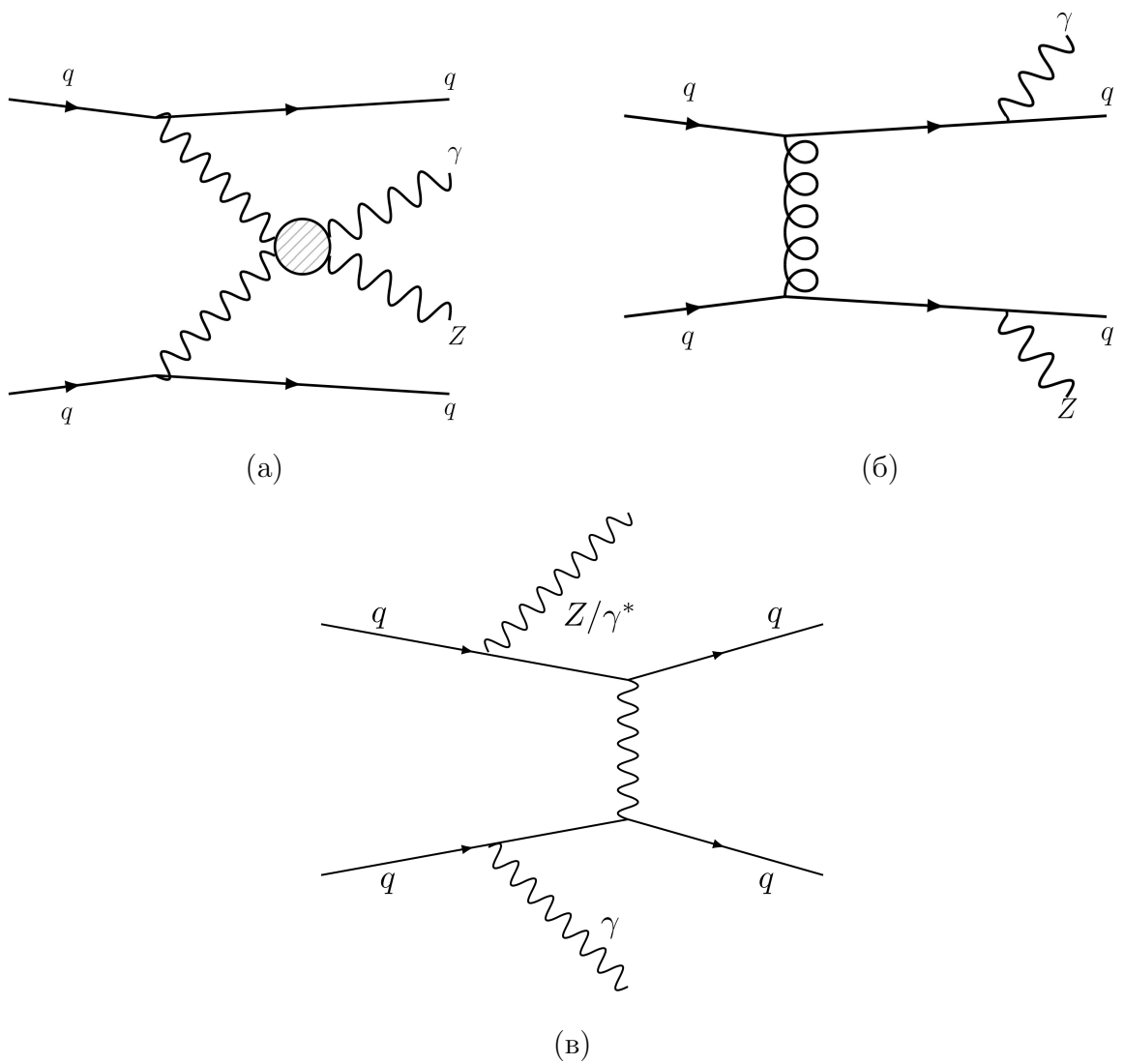


Рисунок 1 — Диаграммы процессов образования посредством рассеяния векторных бозонов (а), КХД (б) и электрослабое образование (в) состояния $Z\gamma jj$.

1 Методы

1.1 Композиция деревьев решений

Композиция деревьев решений, созданная с помощью градиентного бустинга (Boosted Decision Trees, BDT) – это классификатор с бинарной древовидной структурой. Принцип его работы заключается в поочерёдном применении ограничений по различным переменным с построением дерева решений. Это, в отличие от применения фиксированных отборов, позволяет разбить фазовое пространство на множество областей, которые классифицируются как сигнальные или фоновые.

1.1.1 Принцип построения дерева решений

Входные данные попадают в корневой узел дерева, далее производятся отборы по переменным так, чтобы максимизировать коэффициент разделение сигнала и фона. Затем из этих отборов выбирается тот, который обеспечивает максимальное разделение событий. Процесс повторяется для каждого дочернего узла до тех пор, пока количество событий в каком-либо из них не станет меньше установленного. Далее все узлы классифицируются как сигналopodobные или фонopodobные в зависимости от коэффициента чистоты или от преобладания в них сигнальных, либо фоновых событий. Схематичный вид дерева решений изображен на рисунке 2.

1.1.2 Бустинг

Недостатком деревьев решений является их чувствительность к флуктуациям в исходных данных. Например, из-за флуктуации одной переменной в тренировочном наборе данных может сильно повлиять на структуру итогового дерева решений.

Этой проблемы можно избежать, прибегнув к бустингу [5]. Суть этого алгоритма заключается в создании леса деревьев решений. При последо-

вательном создании каждого дерева веса событий тренировочного образца изменяются таким образом, чтобы максимизировать влияние на построение дерева тех переменных, которые были неправильно классифицированы на предыдущих шагах. При этом каждому дереву присваивается вес, который отражает его эффективность в разделении событий.

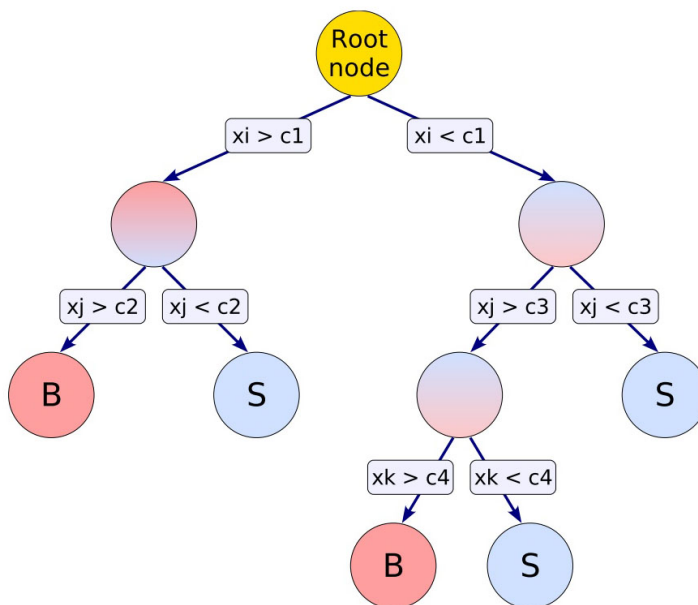


Рисунок 2 — Схематичный вид дерева решений.

При применении классификатора к набору данных, события поступают на вход каждому дереву решений, его отклик равен 1, если событие сигнальное и -1, – если фоновое. Отклик классификатора – непрерывная величина, лежащая в пределах $[-1;1]$ и являющаяся взвешенной суммой откликов всех деревьев в лесу. Распределение по отклику можно использовать для разделения сигнальных и фоновых событий.

1.1.3 Оценка результатов работы алгоритма

Для оценки эффективности работы алгоритма используются параметр значимости при применении ограничений к отклику классификатора (1.1), а также площадь под ROC-кривой, которая является функцией зависимости эффективности отбора сигнала (1.2) и фонового отклонения (1.3) как функций от значения ограничения по отклику. Эффективность сигнала определяется как доля сигнальных событий, которая остаётся по-

сле применения классификатора. Отклонение фона – это доля фоновых событий, исключаемых из исходного набора.

$$\sigma = \frac{S}{\sqrt{S+B}} \quad (1.1)$$

$$\varepsilon = \frac{S}{S_{\text{init}}} \quad (1.2)$$

$$\kappa = 1 - \frac{B}{B_{\text{init}}} \quad (1.3)$$

где S – число сигнальных событий, B – число фоновых событий, S_{init} и B_{init} – число сигнальных и фоновых событий в исходном наборе соответственно.

Также для всех моделей осуществлялась оценка переобучения. Переобучение – это явление при котором модель хорошо проявляет себя при применении к тренировочной выборке, но плохо работает при применении к событиям из тестовой выборки.

С помощью критерия согласия Пирсона оценивалось совпадение распределений отклика при применении модели на тестовой и тренировочной выборках.

1.2 Библиотека LGBM

Библиотека LightGBM [6] предоставляет реализации большого спектра алгоритмов деревьев решений и использует подход, основанный на гистограммах, который позволяеткратно уменьшить время обучения моделей. Также эта библиотека поддерживает работу с недостающими значениями. Во время разделения по какой либо из переменных при построении дерева решений события с недостающими значениями игнорируются. Далее узел, в который попадают события с недостающим значением переменной выбирается так, чтобы коэффициент разделения был наибольшим.

1.3 Отбор переменных

Для отбора переменных для обучения классификатора используется «N+1». Его принцип заключается в следующем. Изначально список пере-

менных для обучения пуст. В него добавляется переменная-кандидат. На основе текущего списка создаётся классификатор и определяется метрика эффективности разделения событий (в работе в качестве метрики используется значимость). Затем переменная убирается из списка и процесс повторяется для всех остальных переменных-кандидатов. После этого выбирается переменная, использование которой обеспечивает наибольший прирост метрики разделения. Она окончательно добавляется в список переменных и исключается из переменных-кандидатов. Затем весь алгоритм повторяется до тех пор, пока список переменных-кандидатов не будет пуст.

Далее строится зависимость значения метрики разделения от выбранных переменных. Список переменных выбирается как первые k переменных до момента, когда метрика перестаёт расти.

1.4 Подбор настроек классификаторов

Для подбора оптимальных настроек классификаторов для обучения было создано большое число моделей с случайными значениями гиперпараметров. Для каждой модели строились распределения отклика сигнала и фона при её применении к тестовой и тренировочной выборке для оценки переобучения. Перетренированные модели отбрасывались с помощью критерия Пирсона на уровне значимости 0.05. После этого выбирались модели с наибольшей значимостью разделения событий.

Случайный выбор настроек обеспечивает намного более быстрый подбор оптимальных значений по сравнению с перебором настроек «по сетке».

1.5 Использование переменных третьей струи

Для лучшего разделения сигнальных и фоновых событий можно дополнительно использовать для обучения классификатора переменные, учитывающие параметры третьей по значению поперечного импульса струи в событии. Однако, как видно из рисунка 3, примерно в 75% событий третья струя отсутствует.

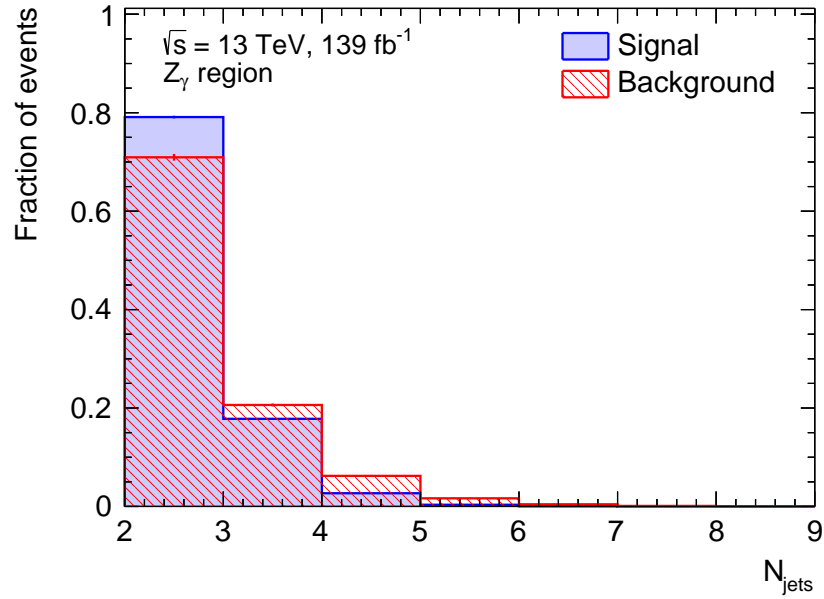


Рисунок 3 — Распределение по числу струй в событии.

Для решения проблемы отсутствия значений переменных у части событий предлагается использовать три стратегии по работе с отсутствующим данными, которые представлены ниже

- B_1 - разделение событий на две категории: с $N_{\text{jets}} = 2$ и $N_{\text{jets}} > 2$. При этом для каждой категории обучается отдельный классификатор ($B_1^{N_{\text{jets}}=2}$ и $B_1^{N_{\text{jets}}>2}$). В регионе с количеством струй $N_{\text{jets}} > 2$ могут использоваться переменные, учитывающие параметры третьей струи. Затем результаты для этих категорий комбинируются;
- B_2 - добавление фиксированных значений параметрам третьей струи для событий с $N_{\text{jets}} = 2$;
- B_3 - использование встроенных в библиотеку LightGBM алгоритмов работы с отсутствующими данными.

Также для сравнения используется стратегия B_0 , никак не учитывающая параметры третьей струи.

2 Используемые данные

2.1 Устройство детектора

ATLAS [7] – это многоцелевой 4π -детектор, один из четырёх крупнейших детекторов на Большом адронном коллайдере, расположенном в Европейской организации по ядерным исследованиям CERN, Женева, Швейцария. Он состоит из внутреннего трекового детектора, электромагнитного и адронного калориметров, магнитной системы, а также мюонной системы. Устройство детектора ATLAS изображено на рисунке 4. Трековый детектор предназначен для определения треков заряженных частиц для измерения их импульса. Калориметры необходимы для измерения энерговыведения частиц, мюонная система используется для определения импульса и направления пролёта мюонов. Магнитная система необходима для искривления траекторий заряженных частиц для определения их импульса.

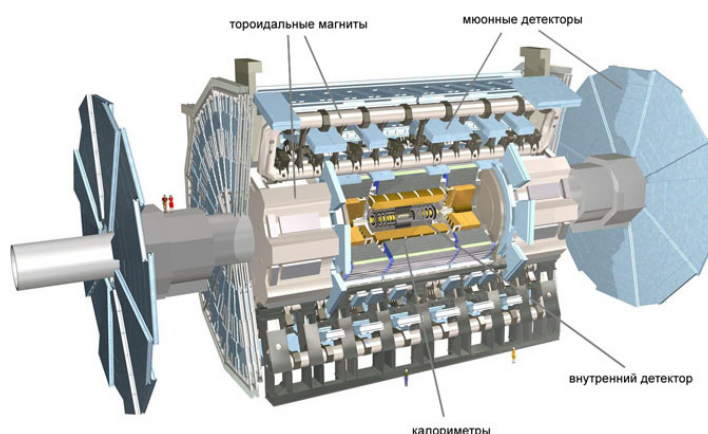


Рисунок 4 — Устройство детектора ATLAS.

Триггерная система детектора состоит из нескольких уровней. Она снижает частоту событий с десятков мегагерц до сотен герц, отбирая события, представляющие интерес для анализа.

Для описания направления вылета частиц, используется цилиндрическая система координат. Азимутальный угол ϕ отсчитывается в плоскости, перпендикулярной оси детектора, полярный θ – от положительного направления оси z , направленной вдоль оси детектора. Обычно вместо угла θ используется величина псевдобыстроты (2.1), дающая более равномерное распределение частиц, рождённых при столкновении.

$$\eta = -\ln \operatorname{tg} \frac{\theta}{2} \quad (2.1)$$

2.2 Исходные данные

Работа проводится с данными, полученными методом МК моделирования протон-протонного столкновения в детекторе ATLAS на Большом адронном коллайдере с энергией в системе центра масс 13 TeV и интегральной светимости 139 fb^{-1} . К исходным данным применены ограничения для отбора кандидатов на процессы с конечным состоянием, содержащим Z -бозон, фотон и две адронные струи (Z -бозон распадается на нейтрино и антинейтрино, которые дают недостающий поперечный импульс). Ограничения перечислены в таблице 1. Ограничение накладываемое на E_T^γ обеспечивает эффективность отбора событий 98.5%. Условия на число фотонов, струй соответствует конечному состоянию процесса. Лептонное вето отсеивает процессы с лептонами в конечном состоянии. Угловые ограничения оптимизированы таким образом, чтобы максимально подавлять прочие фоны. Применённые отборы образуют $Z\gamma$ инклюзивную контрольную область.

Фоновыми процессами являются смоделированные с помощью МК процессы:

- $Z(\nu\nu)\gamma$ QCD – КХД образование Z -бозона с фотоном с последующим распадом Z -бозона на нейтрино и антинейтрино.
- $W\gamma$ QCD и $W\gamma$ EWK – КХД и электрослабые процессы рождения W -бозона с фотоном.
- $t\bar{t}\gamma$ – образование пары топ кварков с фотоном.

- $Z(ll)\gamma$ – фон, связанный с потерей электрона или мюона.

А также фоны, оцениваемые из данных, связанные с неправильной идентификацией частиц или неверным измерением параметров:

- $W(e\nu)$, top , $t\bar{t}$ – фон, связанный с идентификацией электрона как фотона.
- $\gamma + j$ – фон, связанный с неверным измерением недостающего поперечного импульса.
- Zj , jj – фон, связанный с идентификацией струи как фотона.

Для анализа данные дополнительно разделяются на области фазового пространства. Это необходимо для верной оценки некоторых фоновых процессов с помощью реальных данных. Ограничения для областей выбираются таким образом, чтобы в них была как можно большая доля событий того процесса, который необходимо оценить. $W\gamma$ контрольная область используется для уточнения числа событий и формы распределений фона $tt\gamma$. $Z\gamma$ контрольные области используются для уточнения фоновых процессов $Z(\nu\nu)\gamma$, $W\gamma$ QCD и $W\gamma$ EWK. Критерии отбора для областей приведены в таблице 2. Графическое представление критериев отбора для областей представлено на рисунке 5.

Таблица 1 — Критерии отбора событий.

Переменная	Ограничение
E_T^{miss}	>120 GeV
E_T^γ	>150 GeV
Число фотонов	$N_\gamma = 1$
Число струй	$N_{jets} \geq 2$
Число лептонов	$N_e = 0, N_\mu = 0$
$ \Delta\phi(\gamma, \vec{p}_T^{miss}) $	> 0.4
$ \Delta\phi(j_1, \vec{p}_T^{miss}) $	> 0.3
$ \Delta\phi(j_2, \vec{p}_T^{miss}) $	> 0.3

Таблица 2 — Определение контрольных и сигнальных областей.

Переменная	Ограничение
$W\gamma$ контрольная область	
N_{leptons}	≥ 1
$Z\gamma$ QCD контрольная область 1	
N_{leptons}	$= 0$
m_{jj}	$< 300 \text{ GeV}$
$Z\gamma$ QCD контрольная область 2	
N_{leptons}	$= 0$
m_{jj}	$> 300 \text{ GeV}$
$\zeta(\gamma)$	> 0.6
$Z\gamma$ EWK сигнальная область	
N_{leptons}	$= 0$
m_{jj}	$> 300 \text{ GeV}$
$\zeta(\gamma)$	< 0.6

где $\zeta(\gamma)$ – центральность фотона, определяемая формулой

$$\zeta(\gamma) = \left| \frac{\eta_\gamma - \frac{\eta_{j_1} + \eta_{j_2}}{2}}{\eta_{j_1} - \eta_{j_2}} \right| \quad (2.2)$$

где η_γ – псевдобыстрота фотона, η_{j_1}, η_{j_2} – псевдобыстроты струй.

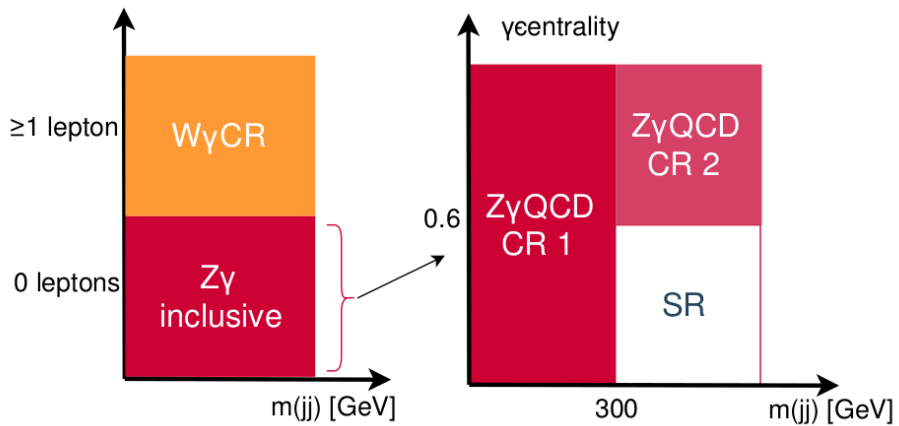


Рисунок 5 — Определение контрольных и сигнальных областей.

Обучение моделей производилось в $Z\gamma$ инклюзивной области. Определение значимости проводилось в $Z\gamma jj$ EWK сигнальной области. Для избежания переобучения моделей из обучающей выборки были исключены фоны, содержащие малое число событий: $\gamma + j$; Zj , jj ; $Z(l)\gamma$.

3 Процесс работы и результаты

3.1 Отбор переменных

Отбор переменных для обучения производился с помощью метода «N+1». Графики зависимости значимости при отборах по отклику классификаторов от используемого для обучения набора переменных представлены на рисунке 6. Список отобранных переменных и их распределения представлены в аппендиксе А. Проверка моделирования в контрольных областях представлена в аппендиксе В.

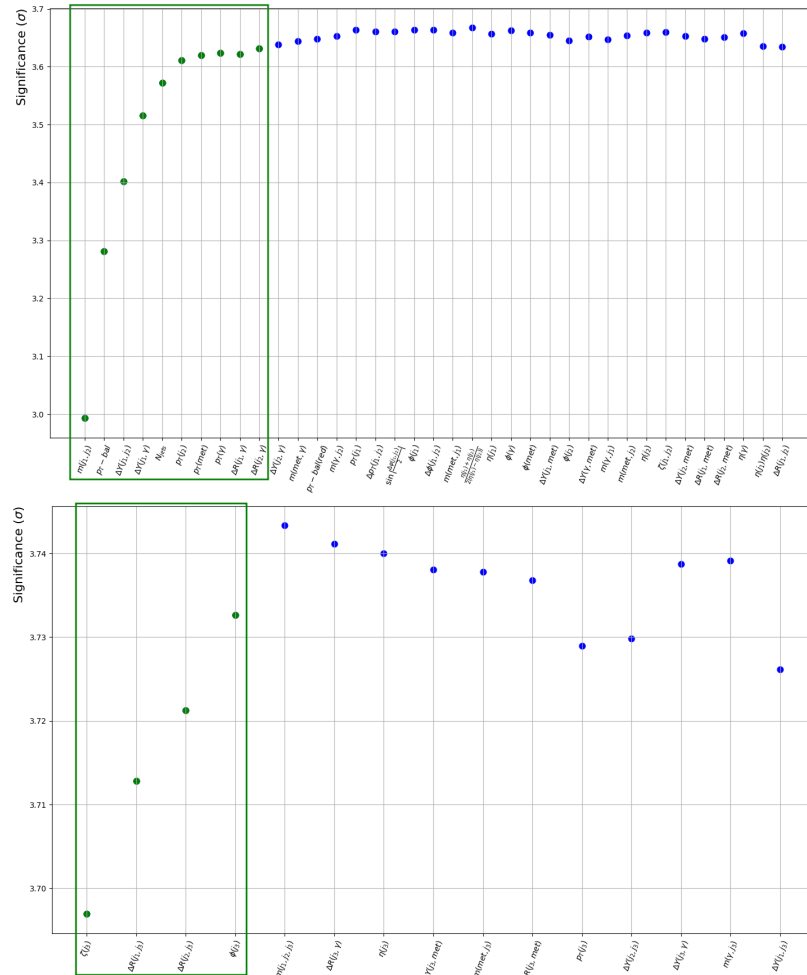
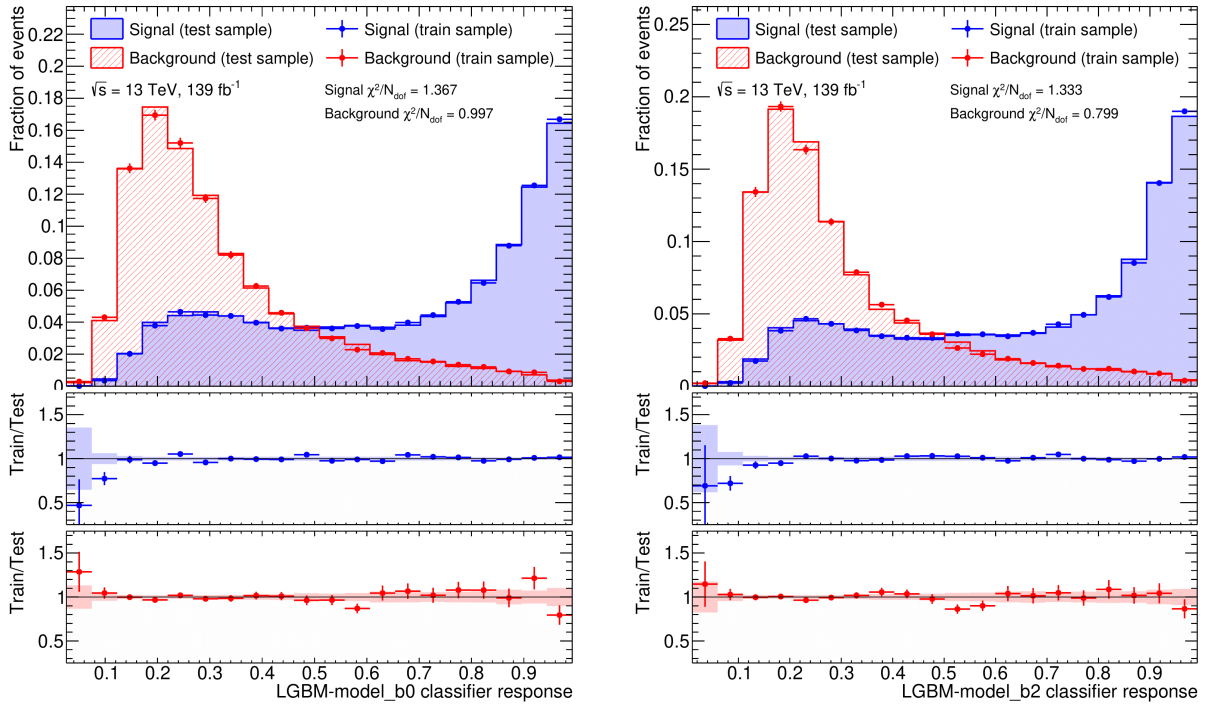


Рисунок 6 — Распределения отклика сигнала и фона для обучающей и тестовой выборки для стратегий B_0 (а) и B_2 (б)

3.2 Тренировка и применение моделей

Для оптимизации работы каждой из стратегий и сравнения их эффективности был создан большой набор классификаторов с случайными значениями настроек и выбраны лучшие. Для каждого классификатора были построены распределения по отклику для сигнала и фона при применении к тренировочной и к тестовой выборкам с целью проверки переобучения. Примеры такого распределения для стратегий B_0 и B_2 представлены на рисунке 7. Сводные результаты результатов работы классификаторов с использованием различных стратегий представлены в таблице 3.



(a)

(б)

Рисунок 7 — Графики зависимости значимости при отборах по отклику классификаторов от используемого для обучения набора переменных.

Таблица 3 — Сводные результаты по созданию классификаторов.

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

Как можно видеть, использование параметров третьей струи при обучении даёт прирост в значимости, но значительных отличий между стратегиями нет. Однако, стратегия B_1 требует создания и оптимизации настроек двух классификаторов, а при использовании стратегии B_2 необходимо учитывать области значений переменных третьей струи для корректного задания фиксированных значений параметров третьей струи для событий с $N_{\text{jets}} = 2$. Поэтому предпочтение отдаётся использованию автоматизированной стратегии B_3 .

Заключение

В процессе работы был произведён отбор оптимальных переменных для обучения классификаторов. Для этих переменных было проверено качество их согласованности с данными. По её результатам можно сказать, что модели, обученные на используемом в исследовании наборе МК данных, могут применяться для классификации событий в реальных данных.

Было произведено сравнение стратегий использования переменных третьей струи. Для этого были оптимизированы настройки классификаторов. Выявлено, что использование параметров третьей струи приводит к росту максимальной значимости, максимальное значение которой составило $(3.71 \pm 0.04)\sigma$. Существенной разницы между стратегиями не наблюдается, однако предпочтение отдаётся стратегии V_3 из-за простоты использования и отсутствия необходимости обучения двух классификаторов.

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

СПИСОК ИСПОЛЬЗУЕМЫХ ИСТОЧНИКОВ

1. *Group P. D.* [et al.]. Review of Particle Physics // Progress of Theoretical and Experimental Physics. — 2020. — Aug. — Vol. 2020, no. 8. — eprint: <https://academic.oup.com/ptep/article-pdf/2020/8/083C01/34673722/ptaa104.pdf> ; — 083C01.
2. *Hoecker A.* [et al.]. TMVA - Toolkit for Multivariate Data Analysis. — 2007. — arXiv: [physics/0703039](https://arxiv.org/abs/physics/0703039) [[physics.data-an](https://arxiv.org/abs/physics/0703039)].
3. *Aaboud M.* [et al.]. Studies of $Z\gamma$ production in association with a high-mass dijet system in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector // Journal of High Energy Physics. — 2017. — July. — Vol. 2017, no. 7.
4. *Khachatryan V.* [et al.]. Measurement of the cross section for electroweak production of $Z\gamma$ in association with two jets and constraints on anomalous quartic gauge couplings in proton–proton collisions at $s=8$ TeV // Physics Letters B. — 2017. — Vol. 770. — P. 380–402.
5. *Friedman J. H.* Greedy function approximation: A gradient boosting machine // Ann. Stat. — 2001. — Vol. 29, no. 5. — P. 1189–1232.
6. *Ke G.* [et al.]. LightGBM: A Highly Efficient Gradient Boosting Decision Tree // Advances in Neural Information Processing Systems. Vol. 30 / ed. by I. Guyon [et al.]. — Curran Associates, Inc., 2017.
7. *Collaboration T. A.* [et al.]. The ATLAS Experiment at the CERN Large Hadron Collider // Journal of Instrumentation. — 2008. — Aug. — Vol. 3, no. 08. — S08003–S08003.

А Распределения по переменным

Были построены распределения по переменным, отобранным для обучения классификатора методом «N+1».

- m_{jj} – инвариантная масса двух струй
- $p_T - \text{balance} = \frac{|\vec{p}_T^{\text{miss}} + \vec{p}_T^\gamma + \vec{p}_T^{j_1} + \vec{p}_T^{j_2}|}{E_T^{\text{miss}} + E_T^\gamma + p_T^{j_1} + p_T^{j_2}}$ – 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 струей и фотоном
- $\zeta(j_3)$ – центральность третьей струи
- $\Delta R(j_1, j_3)$ – угловое расстояние между лидирующей струей и третьей струей
- $\Delta R(j_2, j_3)$ – угловое расстояние между второй струей и третьей струей
- $\varphi(j_3)$ – азимутальный угол вылета третьей струи

Лидирующая, вторая и третья струи – это струи, расположенные в порядке возрастания поперечного импульса. В качестве фоновых событий использовалась сумма всех фонов, рассматриваемых в этом анализе. Угловое расстояние между двумя объектами определяется как $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\varphi)^2}$. На рисунках 8-9 представлены распределения по отобранным переменным.

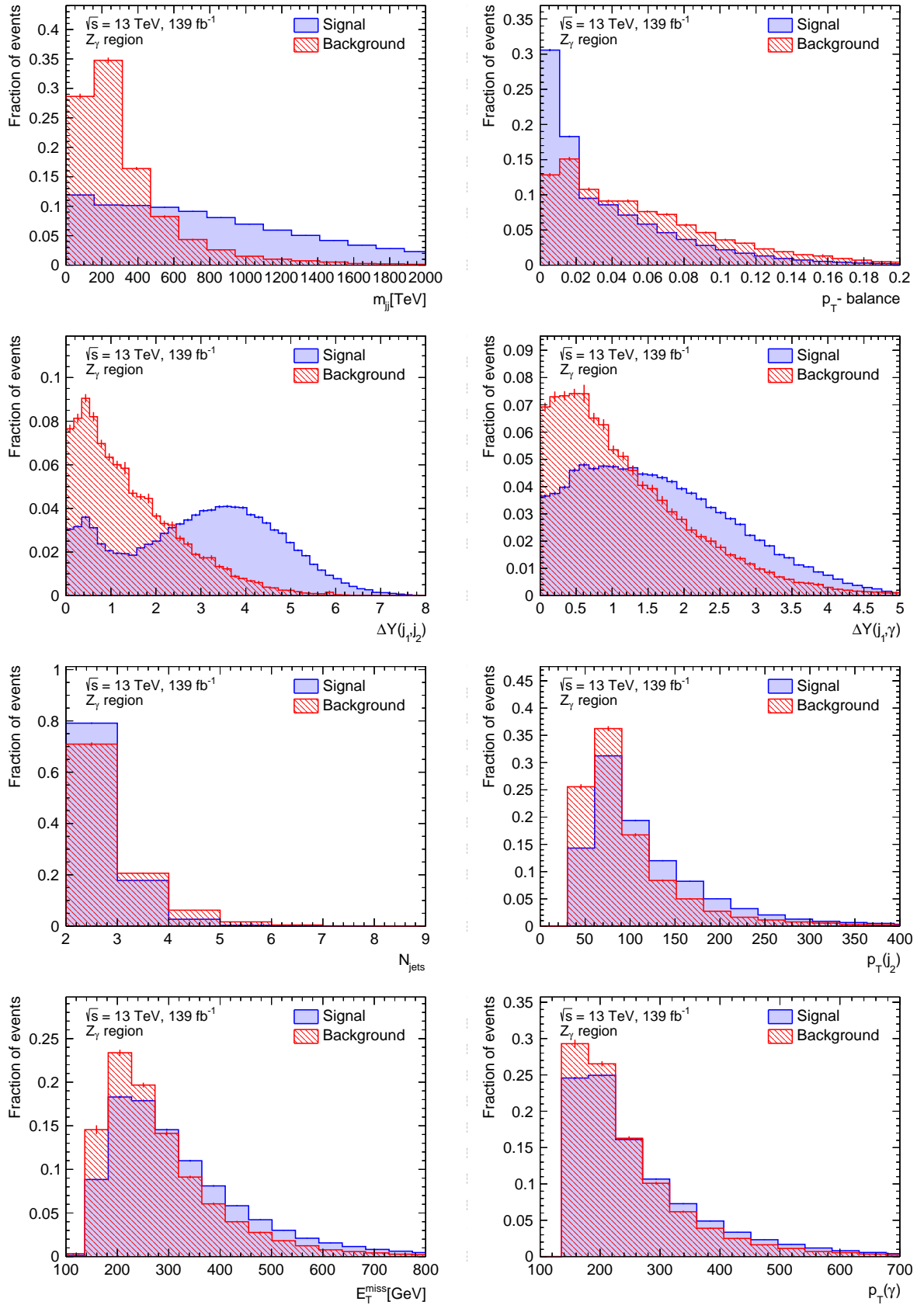


Рисунок 8 — Распределения по отобранным переменным, нормированные на полное число событий, для сигнала и фона.

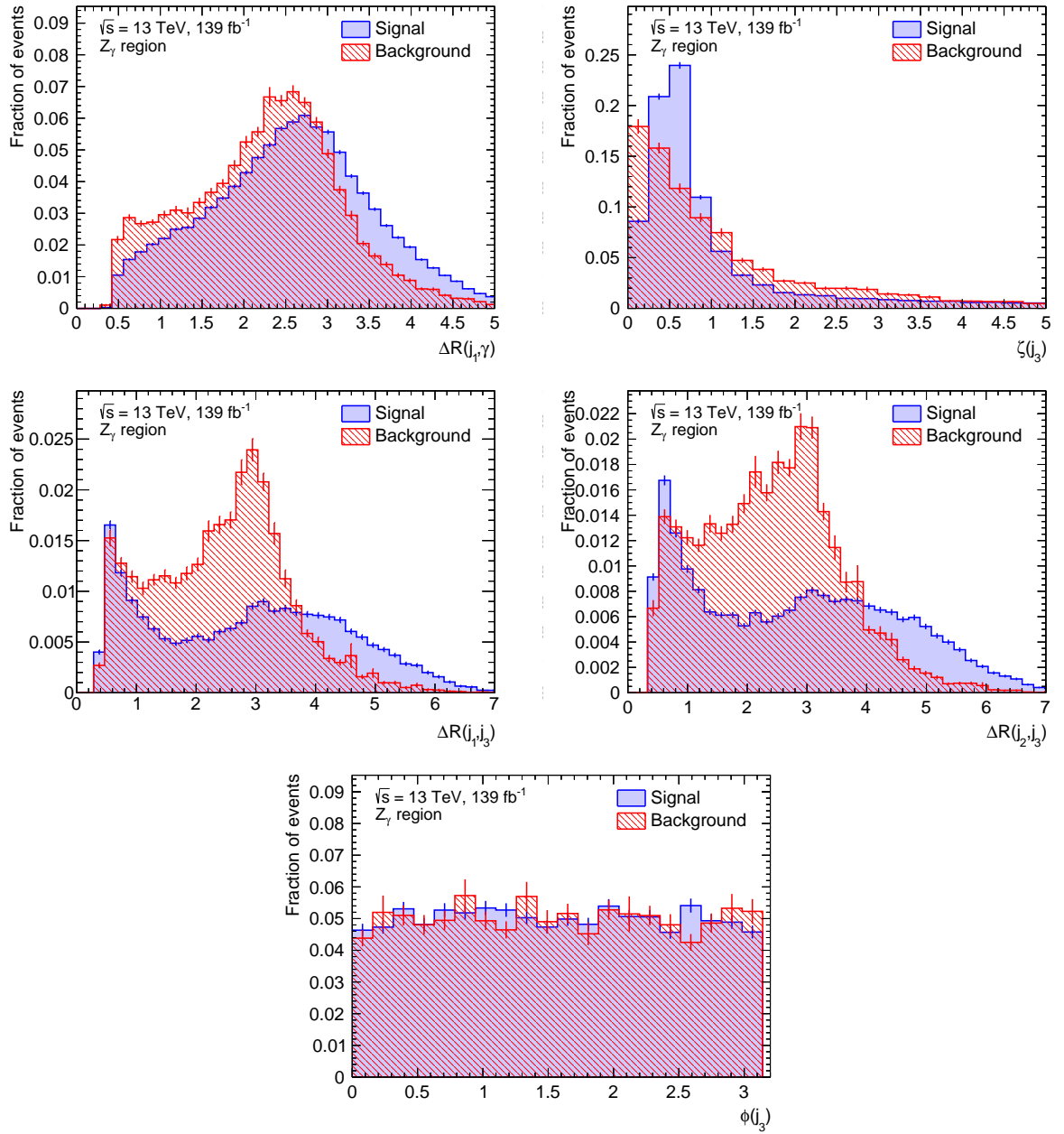


Рисунок 9 — Распределения по отобранным в работе переменным, нормированным на полное число событий, для сигнала и фона.

В Проверка моделирования переменных

Распределения переменных для МК-моделирования и реальных данных в трёх контрольных областях, исследуемых в анализе, представлены на рисунках 10-18. По ним можно видеть, что МК смоделированные данные довольно хорошо согласуются с реальными данными. Поэтому можно сказать, что модели, обученные на смоделированных данных, могут применяться к реальным данным для классификации событий.

В.1 Первая $Z\gamma$ QCD контрольная область

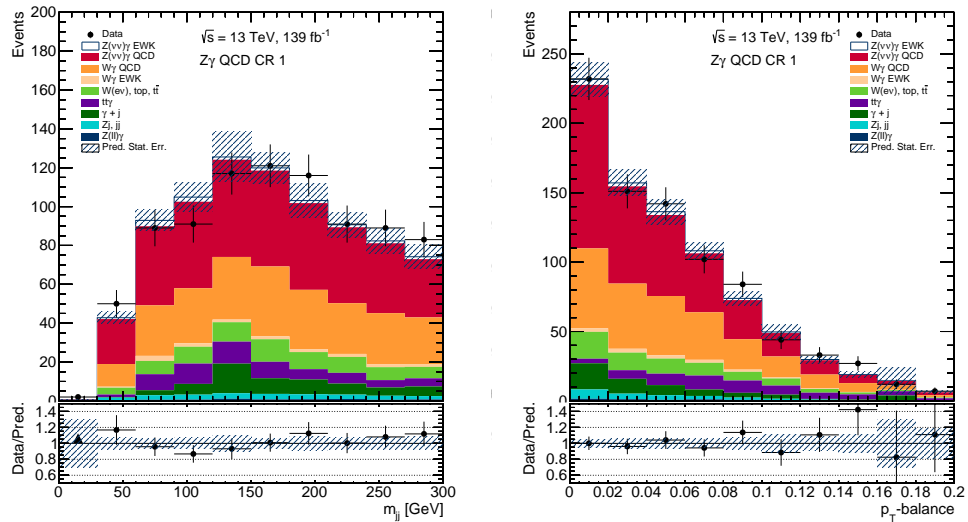


Рисунок 10 — Сравнение распределений переменных реальных и смоделированных данных для первой $Z\gamma$ QCD контрольной области. Штриховкой обозначена статистическая погрешность смоделированных данных.

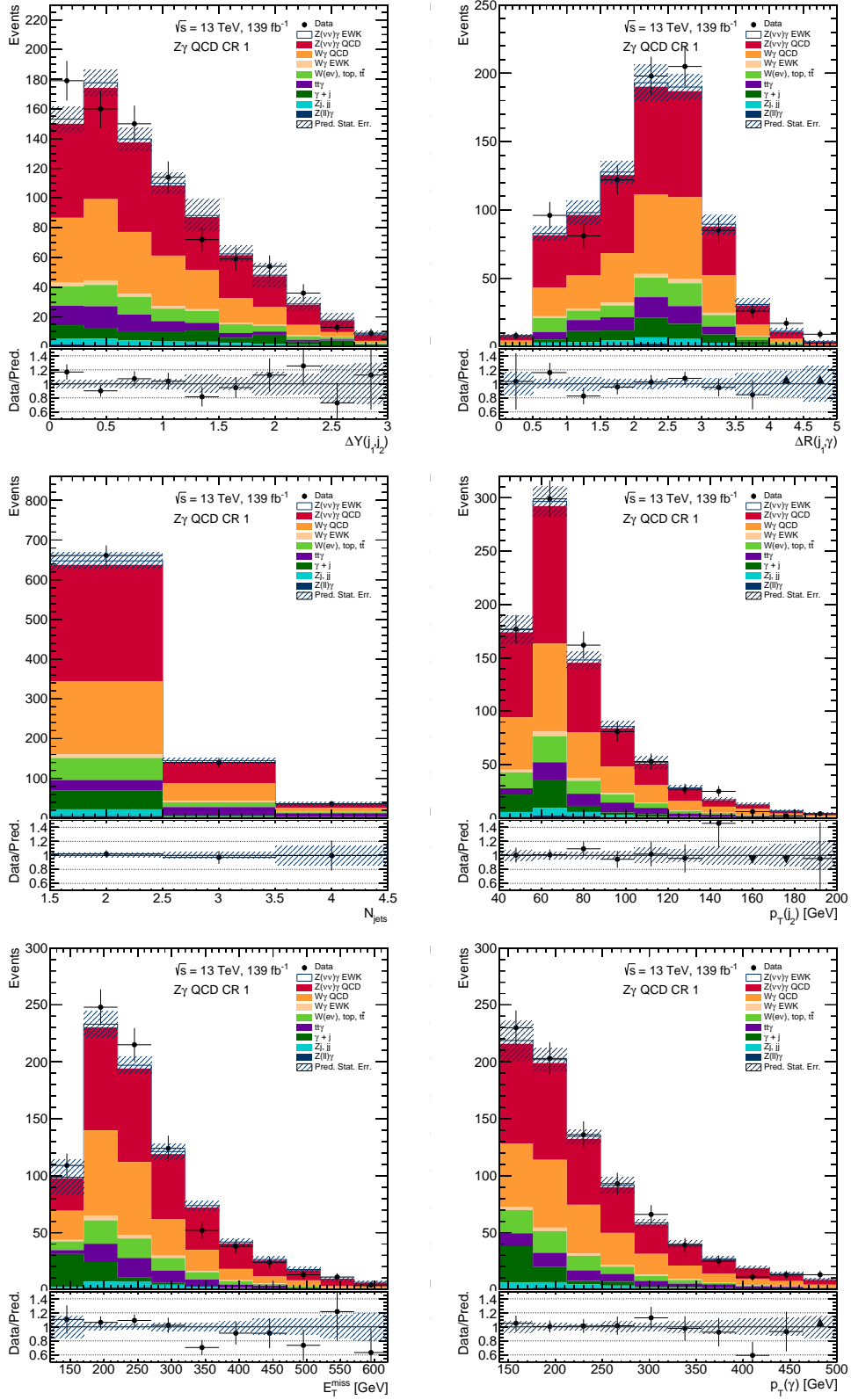


Рисунок 11 — Сравнение распределений переменных реальных и смоделированных данных для первой $Z\gamma$ QCD контрольной области. Штриховкой обозначена статистическая погрешность смоделированных данных.

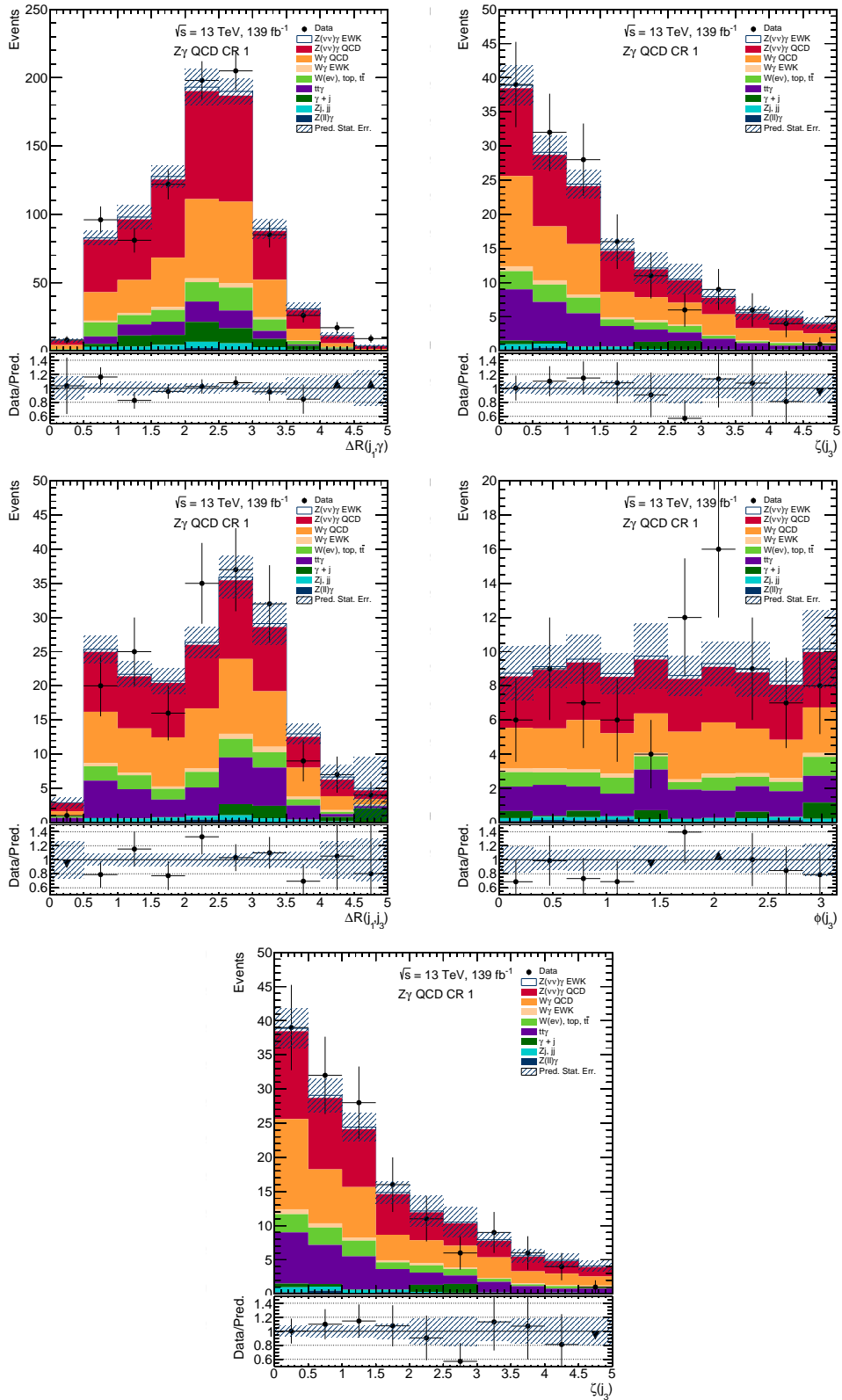


Рисунок 12 — Сравнение распределений переменных реальных и смоделированных данных для первой $Z\gamma$ QCD контрольной области. Штриховкой обозначена статистическая погрешность смоделированных данных.

В.2 Вторая $Z\gamma$ QCD контрольная область

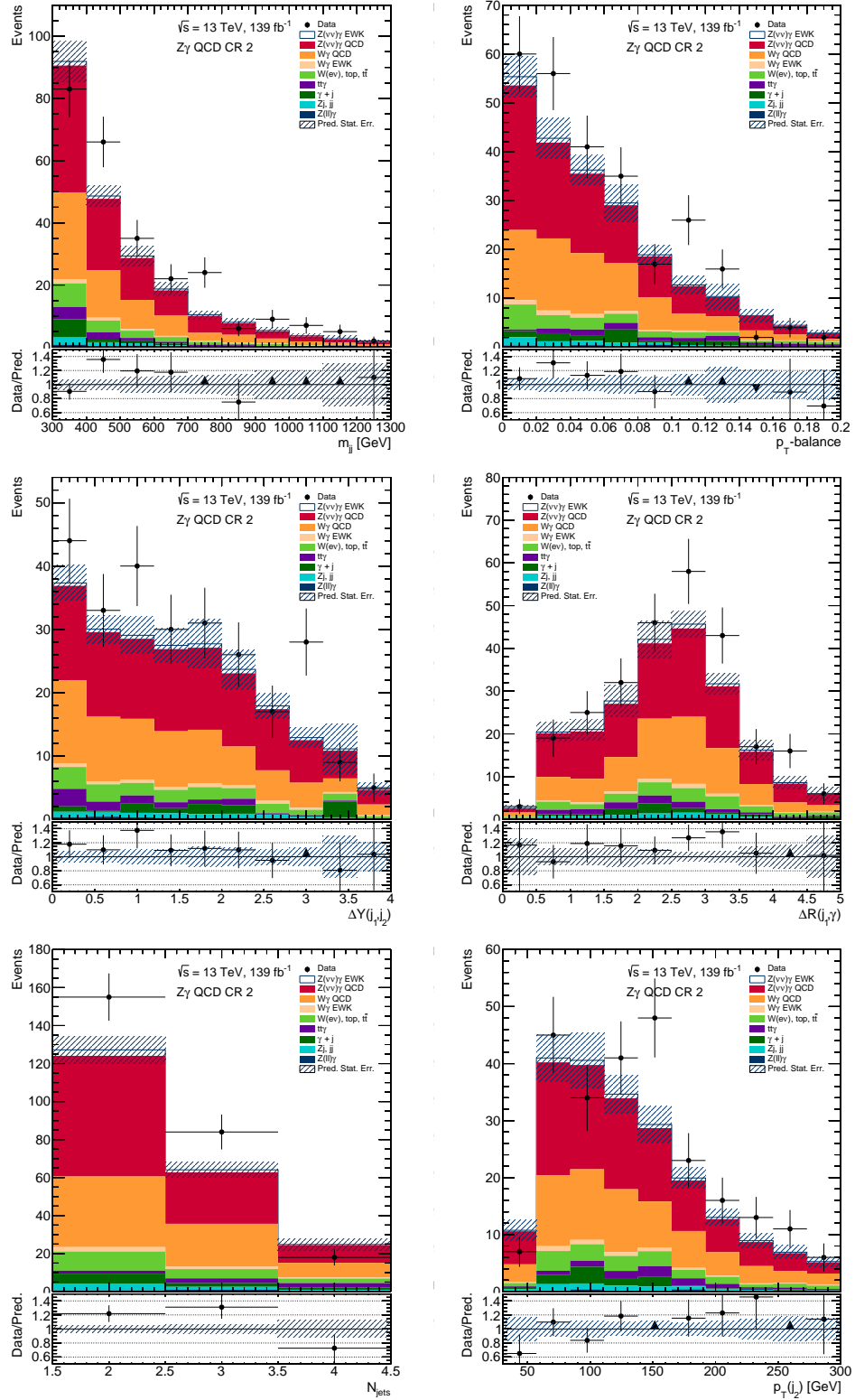


Рисунок 13 — Сравнение распределений переменных реальных и смоделированных данных для второй $Z\gamma$ QCD контрольной области. Штриховкой обозначена статистическая погрешность смоделированных данных.

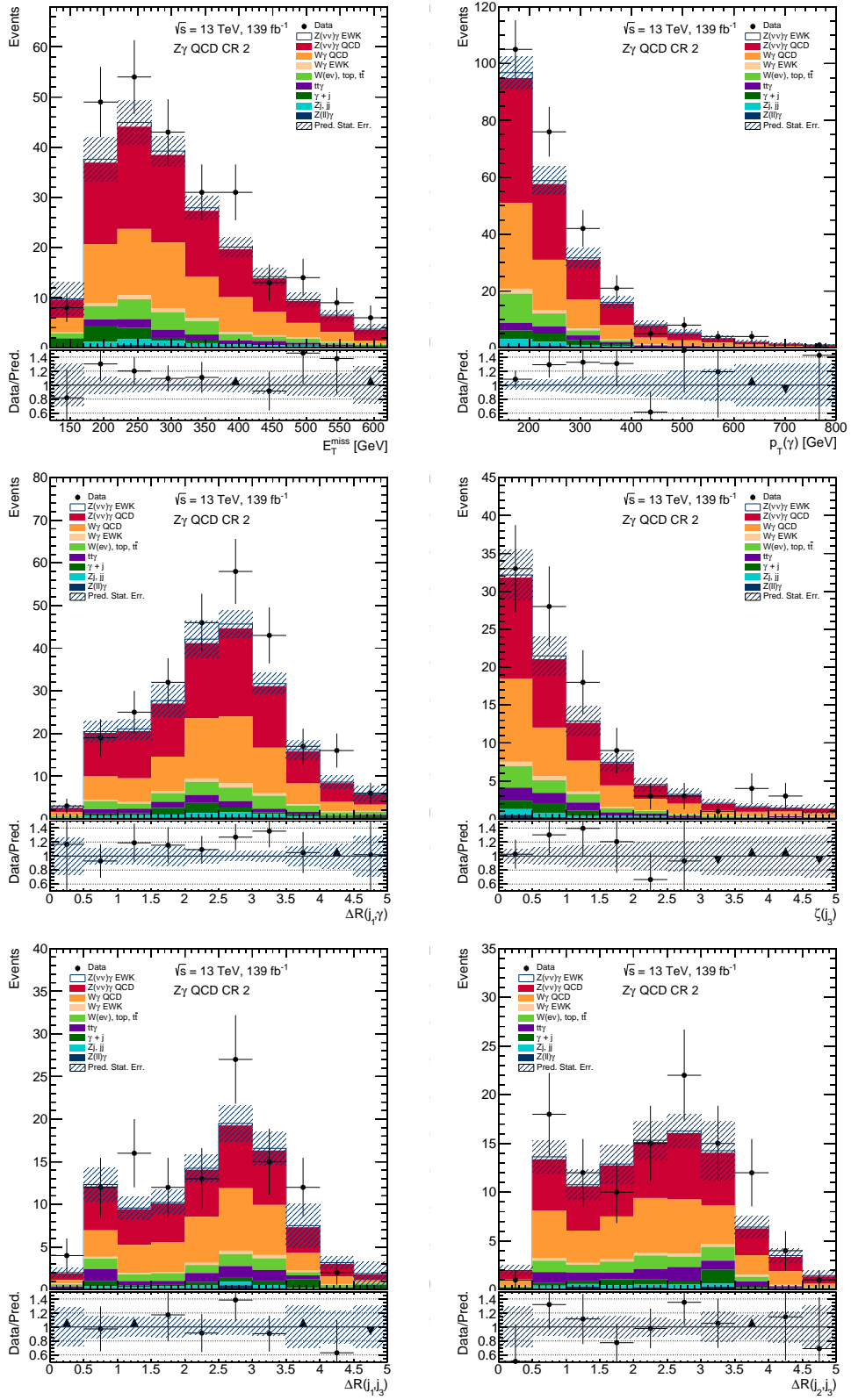


Рисунок 14 — Сравнение распределений переменных реальных и смоделированных данных для второй $Z\gamma$ QCD контрольной области. Штриховкой обозначена статистическая погрешность смоделированных данных.

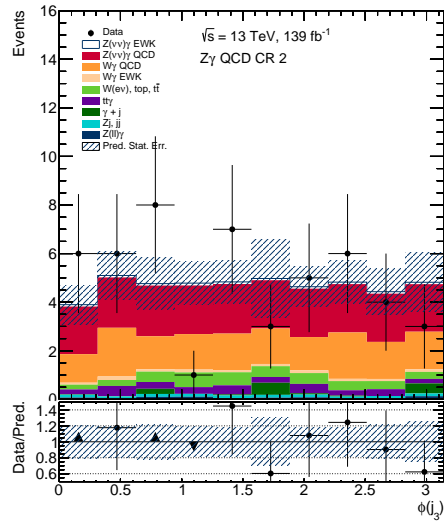


Рисунок 15 — Сравнение распределений переменных реальных и смоделированных данных для второй $Z\gamma$ QCD контрольной области. Штриховкой обозначена статистическая погрешность смоделированных данных.

В.3 $W\gamma$ контрольная область

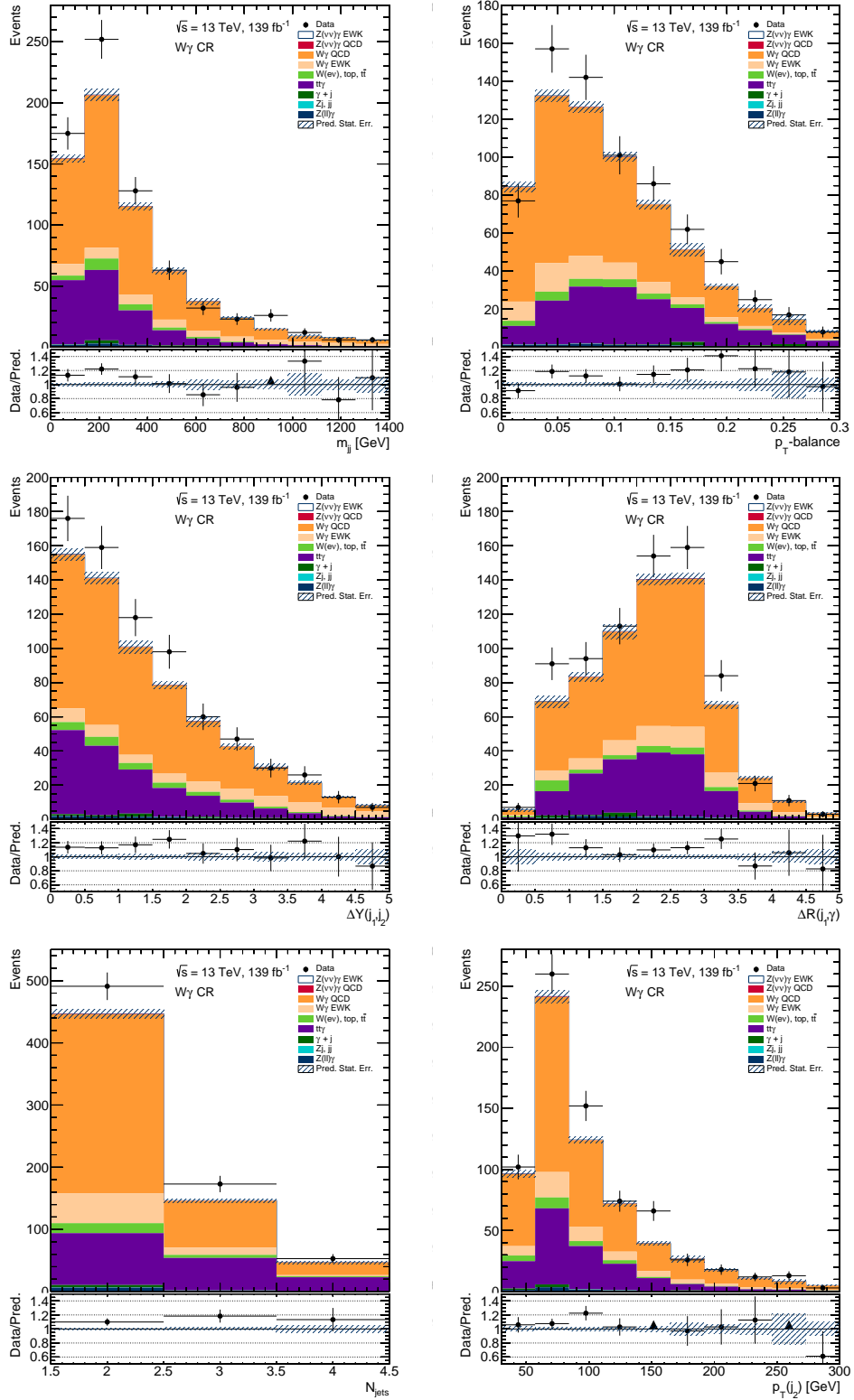


Рисунок 16 — Сравнение распределений переменных реальных и смоделированных данных для $W\gamma$ QCD контрольной области. Штриховкой обозначена статистическая погрешность смоделированных данных.

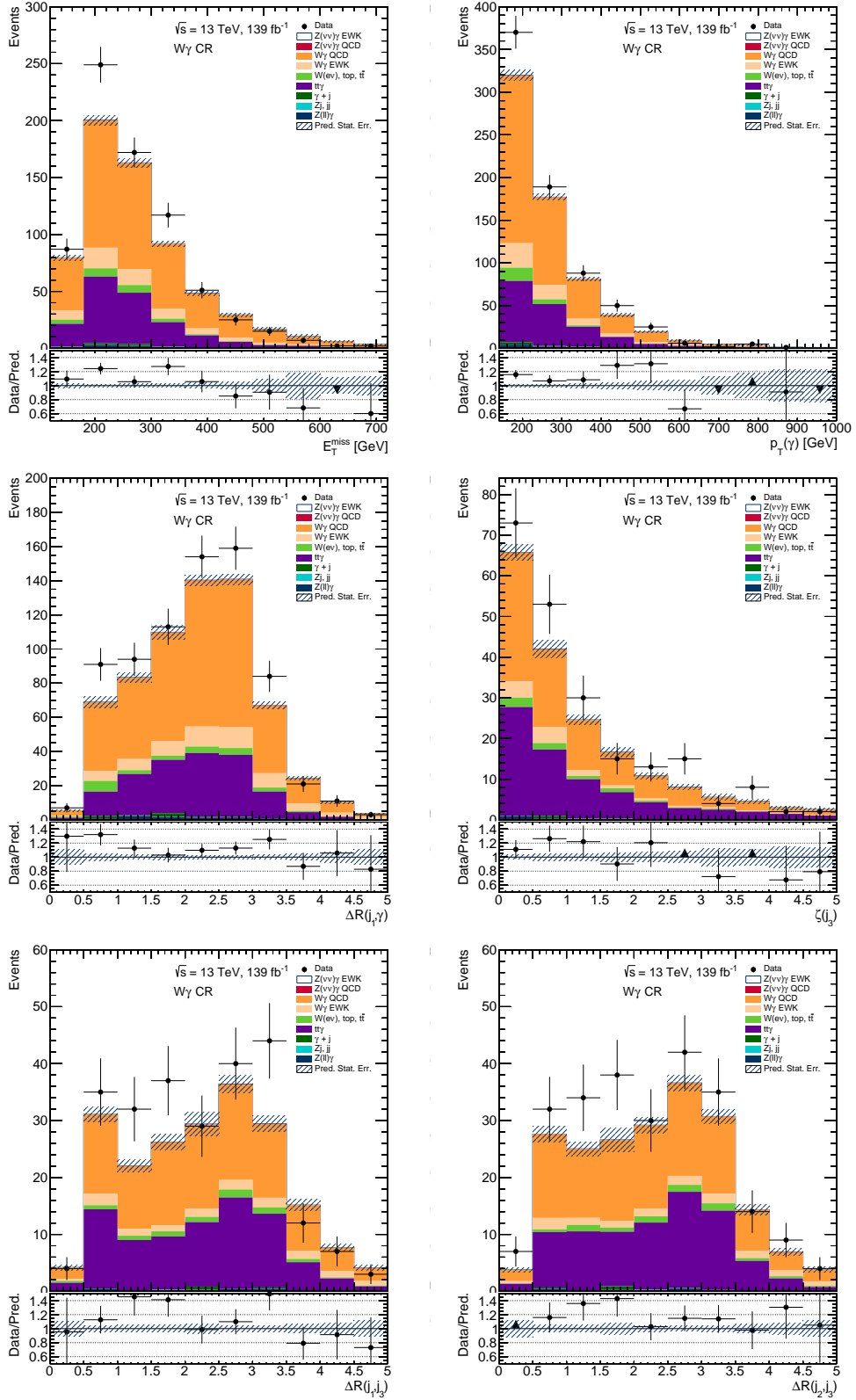


Рисунок 17 — Сравнение распределений переменных реальных и смоделированных данных для $W\gamma$ QCD контрольной области. Штриховкой обозначена статистическая погрешность смоделированных данных.

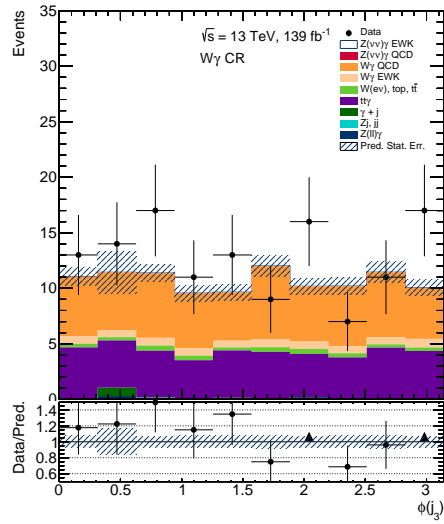


Рисунок 18 — Сравнение распределений переменных реальных и смоделированных данных для $W\gamma$ QCD контрольной области. Штриховкой обозначена статистическая погрешность смоделированных данных.