

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 выделения процессов  
рассеяния векторных бозонов**

Научный руководитель

к.ф.-м.н., доцент

\_\_\_\_\_ Е. Ю. Солдатов

Консультант

\_\_\_\_\_ А. М. Петухов

Студент

\_\_\_\_\_ К. М. Савельев

Москва 2021

# Содержание

|                                                                                         |           |
|-----------------------------------------------------------------------------------------|-----------|
| <b>Введение</b>                                                                         | <b>2</b>  |
| <b>1 Методы</b>                                                                         | <b>5</b>  |
| 1.1 Алгоритм BDT . . . . .                                                              | 5         |
| 1.1.1 Принцип построения дерева решений . . . . .                                       | 5         |
| 1.1.2 Бустинг . . . . .                                                                 | 6         |
| 1.2 Алгоритм MLP . . . . .                                                              | 7         |
| 1.2.1 Принцип работы алгоритма . . . . .                                                | 7         |
| 1.2.2 Методы оптимизации . . . . .                                                      | 8         |
| 1.3 Оценка результатов работы алгоритмов . . . . .                                      | 9         |
| <b>2 Используемые данные</b>                                                            | <b>10</b> |
| 2.1 Устройство детектора . . . . .                                                      | 10        |
| 2.2 Исходные данные . . . . .                                                           | 11        |
| <b>3 Процесс работы и результаты</b>                                                    | <b>14</b> |
| 3.1 Распределение по кинематическим переменным . . . . .                                | 14        |
| 3.2 Проверка моделирования переменных . . . . .                                         | 17        |
| 3.2.1 Первый $Z\gamma$ QCD контрольный регион . . . . .                                 | 17        |
| 3.2.2 Второй $Z\gamma$ QCD контрольный регион . . . . .                                 | 20        |
| 3.2.3 $W\gamma$ контрольный регион . . . . .                                            | 22        |
| 3.3 Фиксированные отборы по переменным . . . . .                                        | 24        |
| 3.4 Тренировка и применение моделей . . . . .                                           | 25        |
| 3.5 Влияние использования Монте-Карло весов на результаты<br>обучения моделей . . . . . | 27        |
| <b>Заключение</b>                                                                       | <b>29</b> |
| <b>Список используемых источников</b>                                                   | <b>30</b> |

# 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 и на данный момент экспериментально не обнаруженный процесс электрослабого рождения  $Z$ -бозона, фотона совместно с двумя адронными струями 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 1a. 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 такого процесса представлен на рисунке 1б, а диаграммы электрослабого процесса, не являющегося рассеянием, представлен на рисунке 1в.

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 проводилась проверка оправданности применения методов машинного обучения, в ходе которой изучалось использование алгоритмов MLP [2] и BDT [2], и были установлены их оптимальные настройки. Также была проведена проверка соответствия Монте-Карло (МК) сгенерированных данных 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 TeV и интегральной светимостью  $20.3 \text{ fb}^{-1}$ , из-за недостаточной чувствительности было выполнено только для поиска аномальных вершин [3]. Это исследование нацелено на увеличение значимости измерения сечения процесса. Достижение порога значимости в  $5\sigma$  позволит подтвердить наблюдение исследуемого процесса. На данный момент опубликовано сечение процесса  $Z(\ell\ell)\gamma jj$  со значимостью  $4.7\sigma$  [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.

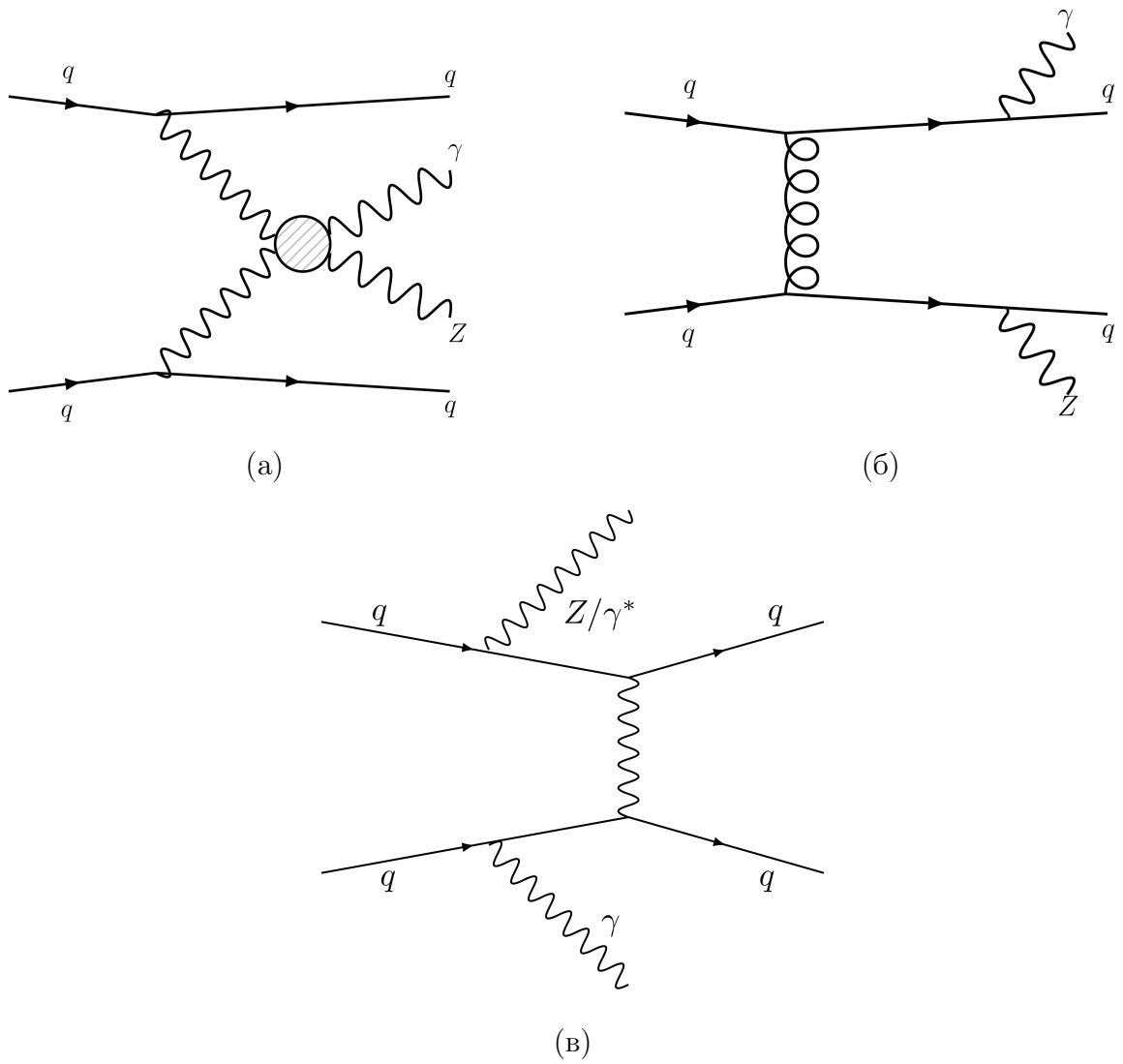


Рисунок 1 — Диаграммы процессов (а) образования посредством рассеяния векторных бозонов, (б) КХД и (в) электрослабое образование состояния  $Z\gamma jj$ .

# 1 Методы

## 1.1 Алгоритм BDT

Алгоритм BDT (Boosted Decision Trees) – это классификатор с бинарной древовидной структурой. Принцип его работы заключается в поочерёдном применении ограничений по различным переменным. Это 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 сигналоподобные или фоновоподобные в зависимости от коэффициента чистоты или от преобладания в них сигнальных, либо фоновых событий. Схематичный вид дерева решений изображен на рисунке [2](#).

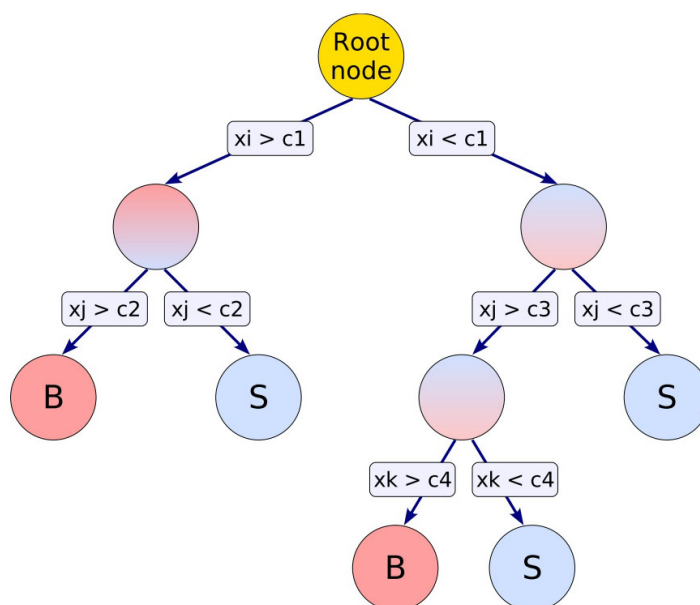


Рисунок 2 — Схематичный вид дерева решений.

## 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 последовательном 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.

При применении классификатора к набору данных, события поступают на вход каждому дереву решений, его отклик равен 1, если событие сигнальное и -1, – если фоновое. Откликом классификатора является взвешенная сумма откликов всех деревьев в лесу.

## 1.2 Алгоритм MLP

Алгоритм MLP (Multilayer Perceptron) – это одна из простейших моделей нейронных сетей. Принцип его работы заключается в последовательном применении линейных преобразований к переменным между слоями нейронов и передаточной функции для получения отклика. Передаточная функция необходима для обеспечения стабильности сходимости алгоритмов обучения и чаще всего является одной из разновидностей сигмоиды (1.1).

$$\sigma(x) = \frac{1}{1 + e^{-x}}, \quad \sigma(x) = \tanh x \quad (1.1)$$

Главными недостатками этого метода по сравнению с деревьями решений являются время обучения и склонность к переобучению. Также для корректной работы требуются преобразования входных данных.

### 1.2.1 Принцип работы алгоритма

Базовой структурной единицей нейронной сети является нейрон, работа которого заключается в суммировании входных значений, прибавлении константы, называемой смещением, и применении к полученной сумме передаточной функции.

Схема простейшего многослойного перцептрона с одним скрытым слоем представлена на рисунке 2. Нейроны в нём образуют слои, которые соединяются друг с другом связями, каждая из которых имеет вес. На вход сети подаётся вектор значений, длина которого равна количеству входных нейронов. Выходные значения нейронов из нижележащих слоёв, умножаясь на веса соответствующих связей, передаются в нейроны вышележащих слоёв. Таким образом, на выходе нейронной сети образуется вектор выходных значений, длина которого равна размерности выходного слоя.

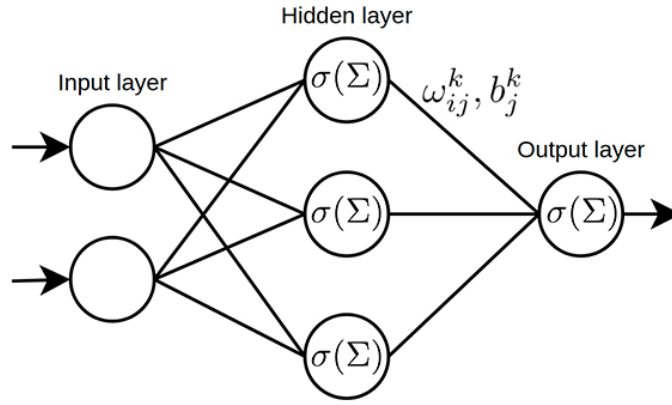


Рисунок 3 — Схема многослойного перцептрона, где  $\omega_{ij}^k$  — вес связи, соединяющий  $i$ -й нейрон  $k$ -1-го слоя с  $j$ -м нейроном  $k$ -го слоя,  $b_j^k$  — смещение  $j$ -го нейрона  $k$ -го слоя,  $\sigma$  — передаточная функция.

### 1.2.2 Методы оптимизации

Отклик нейронной сети в случае задачи классификации представляет из себя значение вероятности принадлежности события к одному из классов. Для обучения вычисляется функция потерь, которая характеризует то, насколько отклик нейросети соответствует истинному значению принадлежности события к классу. Часто в качестве функции потерь используется сумма квадратов отклонений отклика от истинных значений, вычисляемая по формуле

$$f_{loss}(\vec{y}) = \frac{1}{2} \sum (y_i - \hat{y}_i)^2 \quad (1.2)$$

где  $\vec{y}$  — вектор выходных значений,  $y_i$  — компоненты вектора входных значений,  $\hat{y}_i$  — истинные (ожидаемые) значения.

Процесс обучения заключается в оптимизации этой функции, то есть в поиске её локального минимума в пространстве, которое образовано весами связей сети и смещениями нейронов.

Обучение может осуществляться большим количеством способов, простейшим из которых является алгоритм обратного распространения ошибки [6]. Он представляет из себя одну из разновидностей метода градиентного спуска, при котором ошибка распространяется от выходов нейронной сети к её входам. Значения весовых коэффициентов и смещений изменяются на величину, пропорциональную соответствующей составляющей градиен-

та функции потерь. Величины изменения весов и смещений определяются по формулам

$$\Delta w_{ij}^k = -\alpha \frac{\partial f_{loss}}{\partial w_{ij}^k}, \quad \Delta b_j^k = -\alpha \frac{\partial f_{loss}}{\partial b_j^k} \quad (1.3)$$

где  $\alpha$  – константа обучения.

Другие более продвинутые алгоритмы – BFGS (Broyden–Fletcher–Goldfarb–Shanno algorithm) [7], ADAM (adaptive moment estimation) [8] и прочие являются итерационными алгоритмами численной оптимизации. Они обеспечивают быструю сходимость и высокую стабильность обучения, однако гораздо более сложны в подборе гиперпараметров.

## 1.3 Оценка результатов работы алгоритмов

Для оценки эффективности работы алгоритмов используются параметр значимости при применении ограничений к отклику классификатора (1.4), а также площадь под ROC-кривой, которая является функцией зависимости эффективности отбора сигнала (1.5) и фонового отклонения (1.6) как функций от значения ограничения по отклику. Эффективность сигнала определяется как доля сигнальных событий, которая остаётся после применения классификатора. Отклонение фона – это доля фоновых событий, исключаемых из исходного набора.

$$\sigma = \frac{S}{\sqrt{S+B}} \quad (1.4)$$

$$\varepsilon = \frac{S}{S_{\text{init}}} \quad (1.5)$$

$$\kappa = 1 - \frac{B}{B_{\text{init}}} \quad (1.6)$$

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

Также для всех моделей осуществлялась оценка переобучения. С помощью критерия согласия Пирсона оценивалось совпадение распределений отклика при применении модели на тестовой и тренировочной выборках.

## 2 Используемые данные

### 2.1 Устройство детектора

ATLAS [9] – это многоцелевой  $4\pi$ -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 энергосодержания частиц, мюонная система используется для определения импульса и направления пролёта мюонов. Магнитная система необходима для искривления траекторий заряженных частиц для определения их импульса.

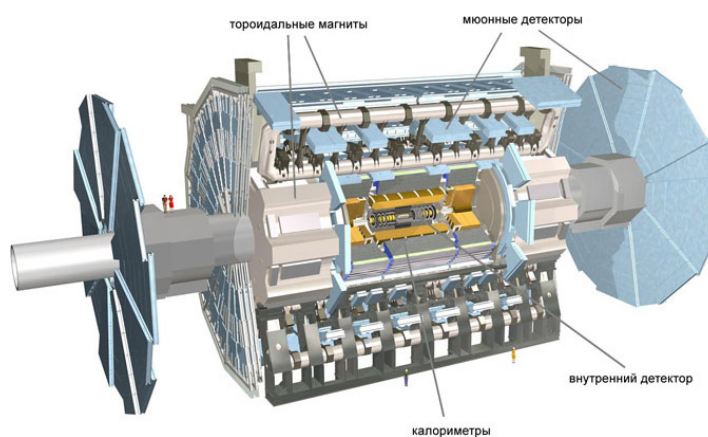


Рисунок 4 — Устройство детектора ATLAS.

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  $\phi$  отсчитывается в плоскости, перпендикулярной оси детектора, полярный  $\theta$  – от положительного направления оси  $z$ , направленной вдоль оси детектора. Обычно вместо угла  $\theta$  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  $139 \text{ fb}^{-1}$ . 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  $E_T^\gamma$  является триггерным. 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  $Z\gamma$  инклюзивный регион.

Фоновыми процессами являются МК моделированные 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.

- $Z(l\bar{l})\gamma$  – фон, связанный с потерей электрона или мюона.

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

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

Для анализа данные дополнительно разделяются на регионы. Это необходимо для верной оценки некоторых фоновых процессов с помощью реальных данных. Ограничения для регионов выбираются таким образом, чтобы в них была как можно большая доля событий того процесса, который необходимо оценить.  $W\gamma$  контрольный регион используется для уточнения числа событий и формы распределений фона  $t\bar{t}\gamma$ .  $Z\gamma$  контрольные регионы используются для уточнения фоновых процессов  $Z(\nu\nu)\gamma$ ,  $W\gamma$  QCD и  $W\gamma$  EWK. Критерии отбора для регионов приведены в таблице 2.

Таблица 1 — Критерии отбора событий.

| Переменная                               | Ограничение          |
|------------------------------------------|----------------------|
| $E_T^{miss}$                             | $> 120 \text{ GeV}$  |
| $E_T^\gamma$                             | $> 150 \text{ GeV}$  |
| 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 регионов.

| Переменная                         | Ограничение         |
|------------------------------------|---------------------|
| $W\gamma$ контрольный регион       |                     |
| $N_{\text{leptons}}$               | $\geq 1$            |
| $Z\gamma$ QCD контрольный регион 1 |                     |
| $N_{\text{leptons}}$               | $= 0$               |
| $m_{jj}$                           | $< 300 \text{ GeV}$ |
| $Z\gamma$ QCD контрольный регион 2 |                     |
| $N_{\text{leptons}}$               | $= 0$               |
| $m_{jj}$                           | $> 300 \text{ GeV}$ |
| $\zeta(\gamma)$                    | $> 0.6$             |
| $Z\gamma$ EWK сигнальный регион    |                     |
| $N_{\text{leptons}}$               | $= 0$               |
| $m_{jj}$                           | $> 300 \text{ GeV}$ |
| $\zeta(\gamma)$                    | $< 0.6$             |

$\zeta(\gamma)$  – 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  $\eta_\gamma$  – псевдобыстрота фотона,  $\eta_{j_1}$ ,  $\eta_{j_2}$  – псевдобыстроты струй.

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

## 3 Процесс работы и результаты

### 3.1 Распределение по кинематическим переменным

Были построены распределения по переменным, показавшим наибольшую эффективность в разделении сигнала и фона. К ним относятся

- $m_{jj}$  – инвариантная масса двух струй
- $\Delta Y(j_1, j_2)$  – разность быстрот двух струй
- $E_T^{\text{miss}}$  – недостающий поперечный импульс
- $p_T - \text{balance} = \frac{|\vec{p}_T^{\text{miss}} + \vec{p}_T^\gamma + \vec{p}_T^{j_1} + \vec{p}_T^{j_2}|}{E_T^{\text{miss}} + E_T^\gamma + p_T^{j_1} + p_T^{j_2}}$  – баланс поперечных импульсов
- $\eta(j_2)$  – псевдобыстрота второй по значению поперечного импульса струи
- $p_T(j_1)$  – поперечный импульс лидирующей струи
- $\eta(\gamma)$  – псевдобыстрота фотона
- $p_T - \text{balance}(\text{reduced}) = \frac{|\vec{p}_T^\gamma + \vec{p}_T^{j_1} + \vec{p}_T^{j_2}|}{E_T^\gamma + p_T^{j_1} + p_T^{j_2}}$  – сокращённый баланс поперечных импульсов
- $N_{\text{jets}}$  – число адронных струй
- $\sin \left| \frac{\Delta\varphi(j_1, j_2)}{2} \right|$
- $\Delta Y(j_1, \gamma)$  – разность псевдобыстрот между струёй и фотоном

В качестве фоновых событий использовалась сумма всех фонов, рассматриваемых в этом анализе.

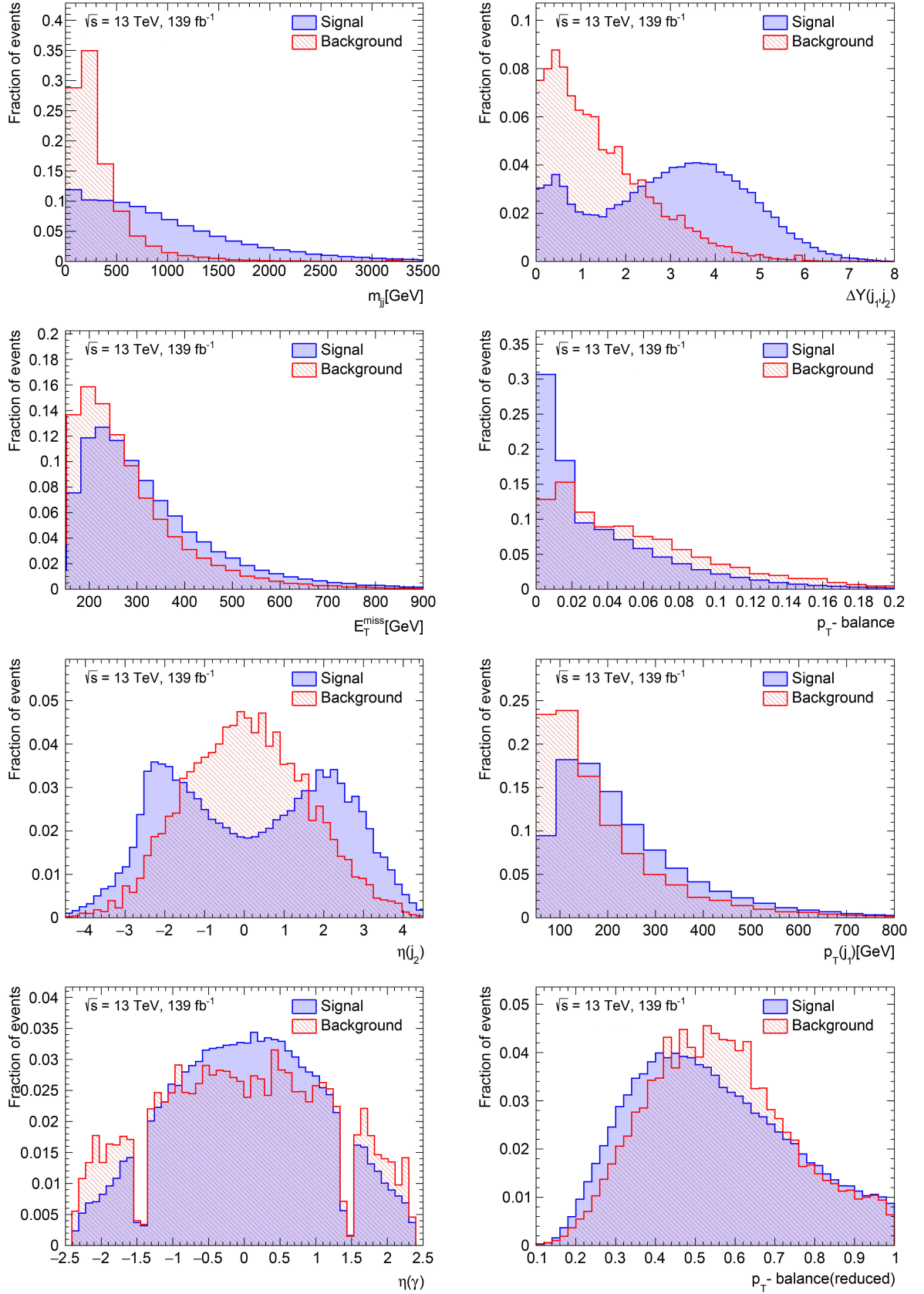


Рисунок 5 — Распределения по рассматриваемым в работе переменным, нормированные на полное число событий, для сигнала и фона.

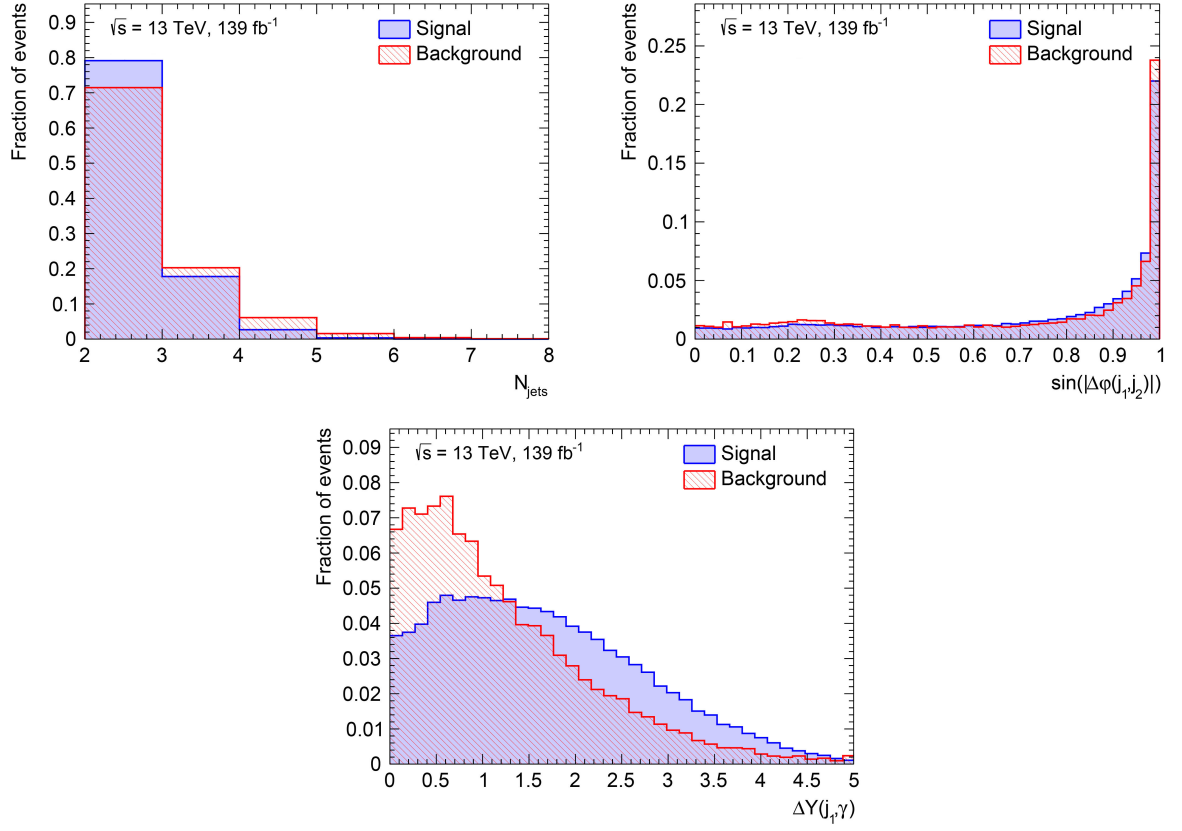


Рисунок 6 — Распределения по рассматриваемым в работе переменным, нормированным на полное число событий, для сигнала и фона.

## 3.2 Проверка моделирования переменных

Распределения переменных для МК-моделирования и реальных данных в трёх регионах, исследуемых в анализе, представлены на рисунках 7-13. 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.

### 3.2.1 Первый $Z\gamma$ QCD контрольный регион

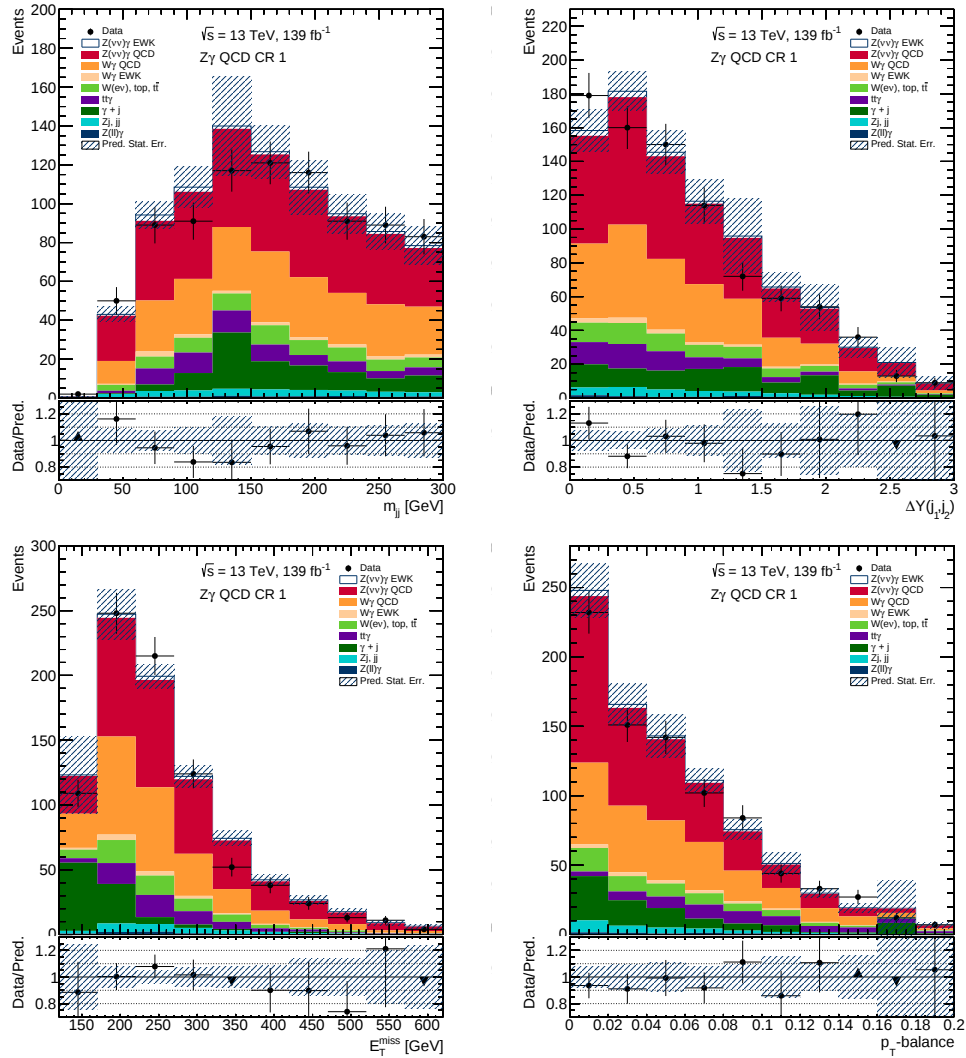


Рисунок 7 — Сравнение распределений переменных реальных и смоделированных данных для первого  $Z\gamma$  QCD контрольного региона. Штриховкой обозначена статистическая погрешность смоделированных данных.

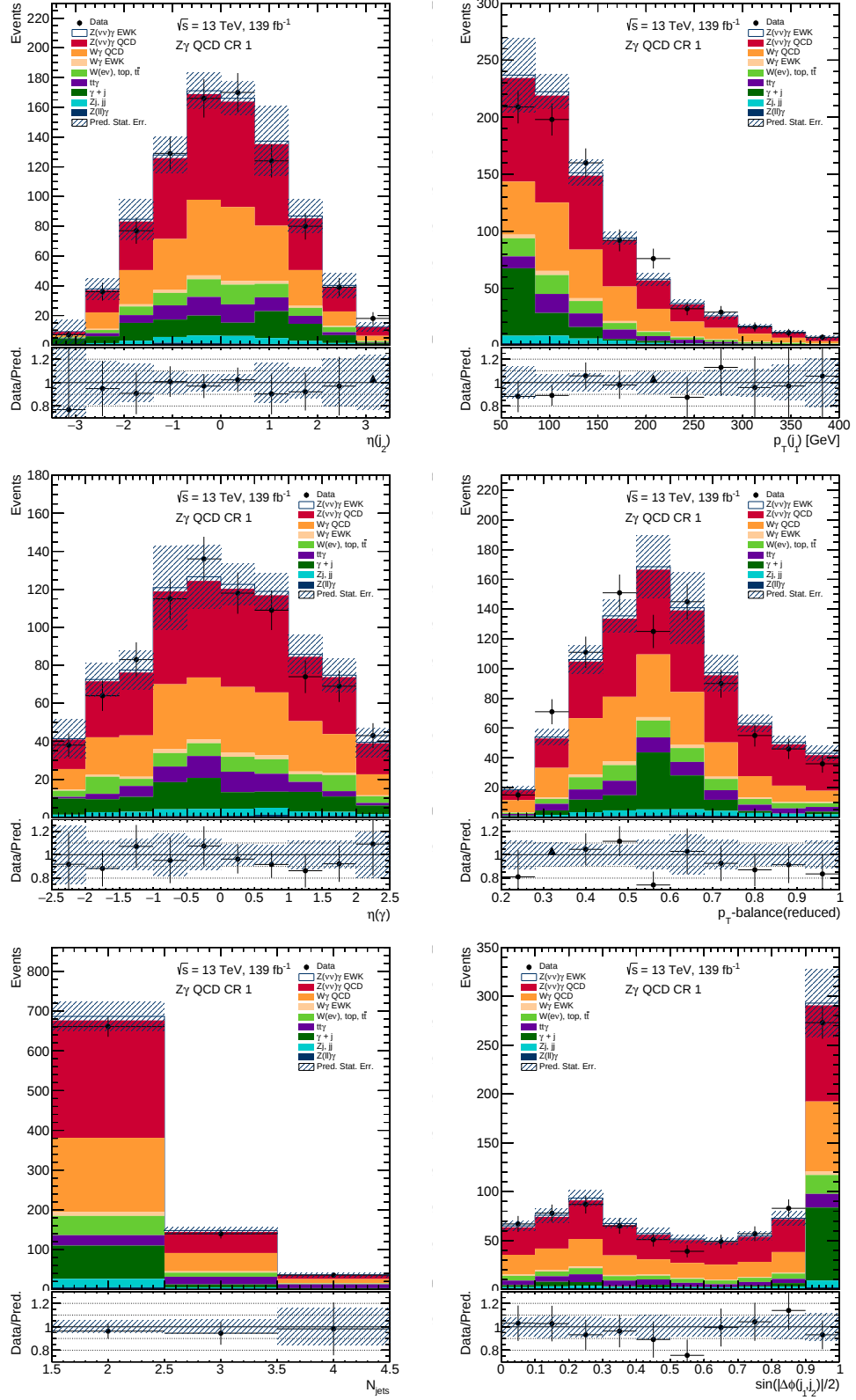


Рисунок 8 — Сравнение распределений переменных реальных и смоделированных данных для первого  $Z\gamma \text{ QCD}$  контрольного региона. Штриховкой обозначена статистическая погрешность смоделированных данных.

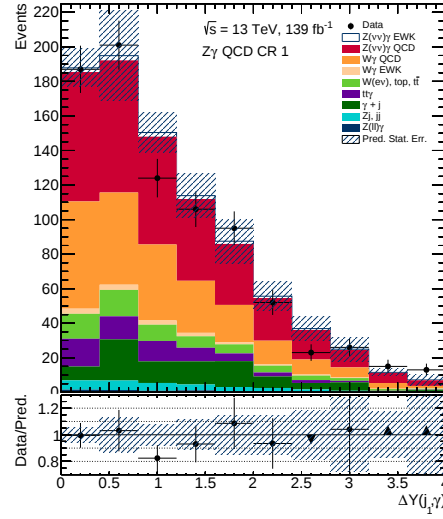


Рисунок 9 — Сравнение распределений переменных реальных и смоделированных данных для первого  $Z\gamma$  QCD контрольного региона. Штриховкой обозначена статистическая погрешность смоделированных данных.

### 3.2.2 Второй $Z\gamma$ QCD контрольный регион

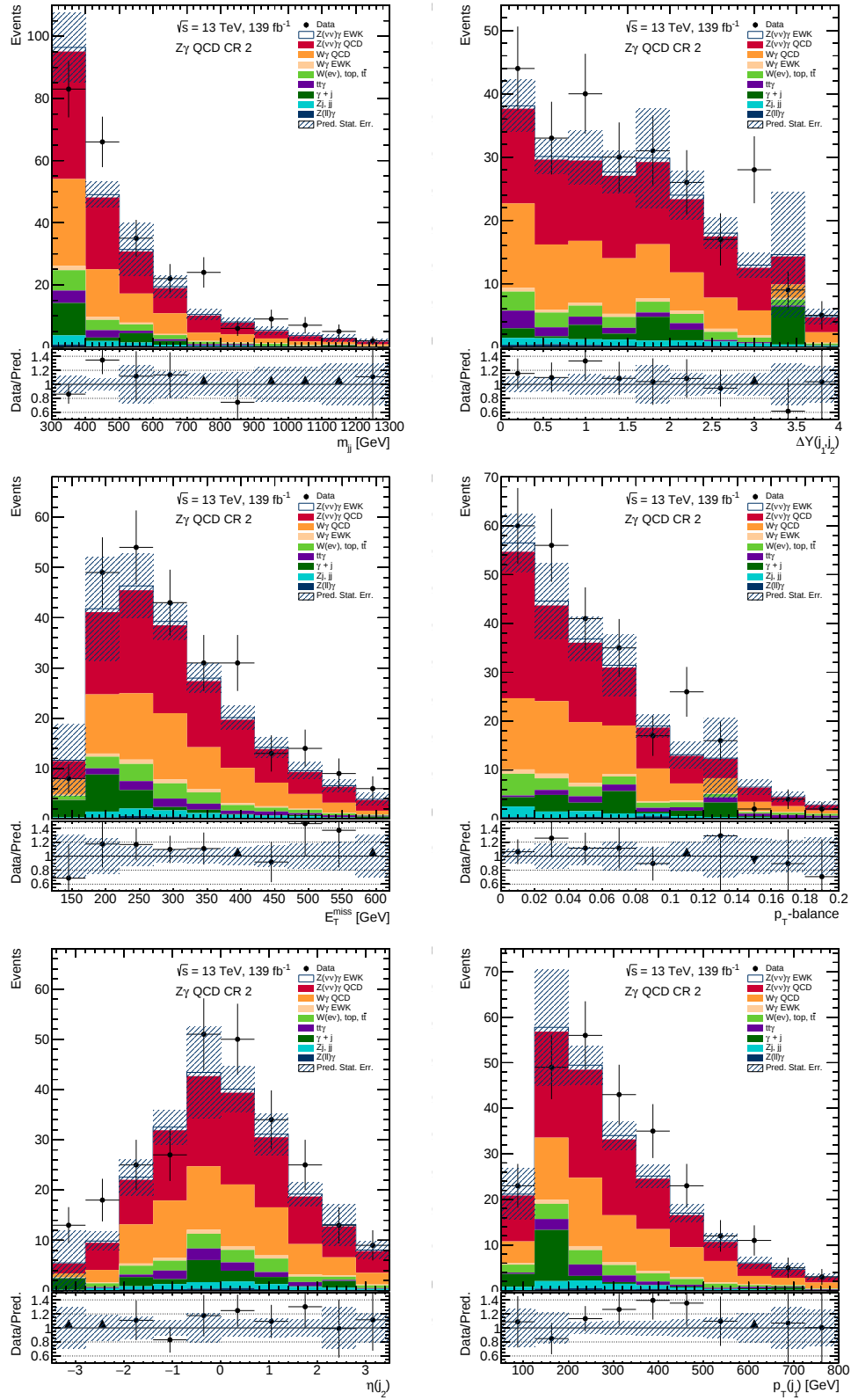


Рисунок 10 — Сравнение распределений переменных реальных и смоделированных данных для второго  $Z\gamma$  QCD контрольного региона. Штриховкой обозначена статистическая погрешность смоделированных данных.

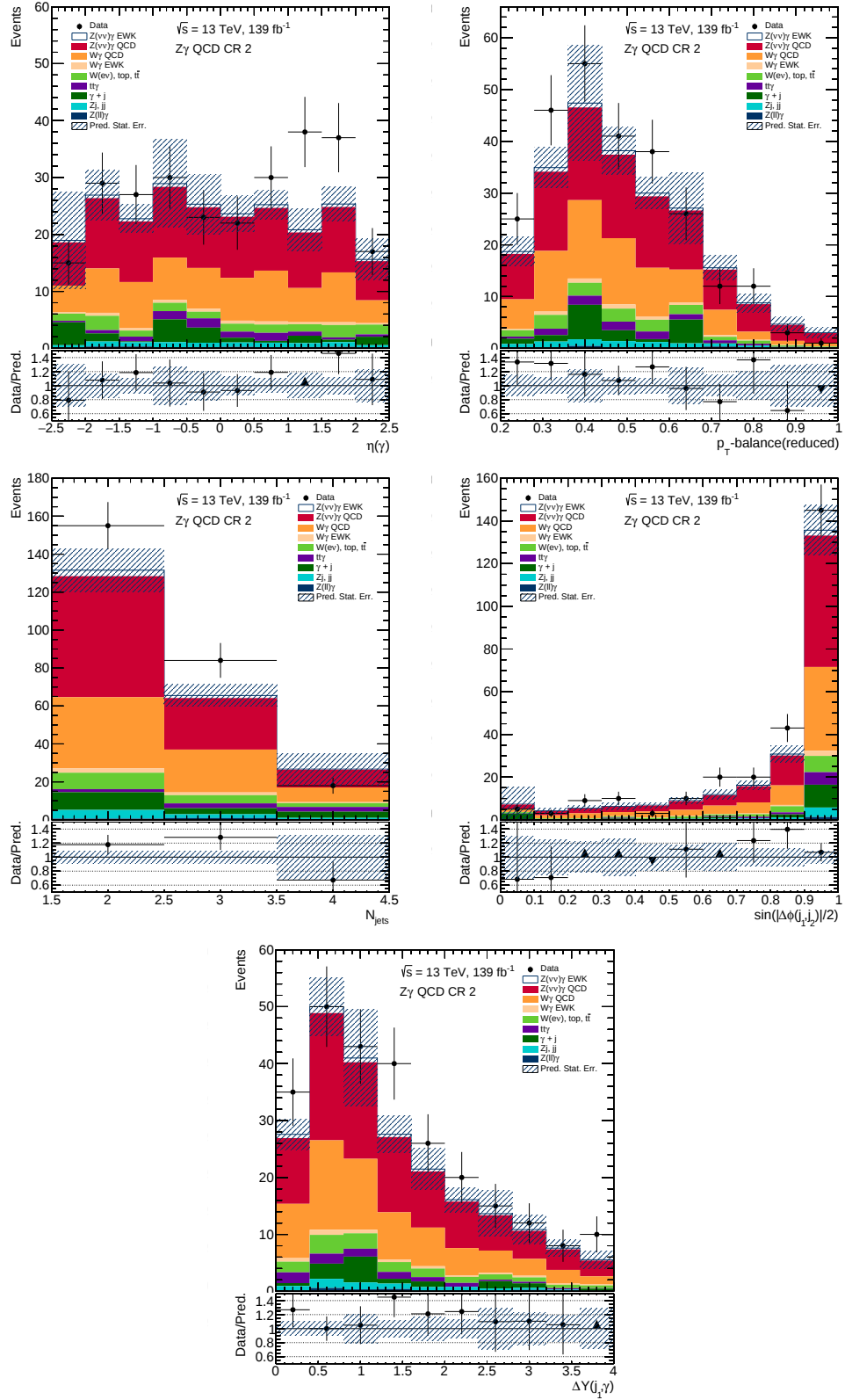


Рисунок 11 — Сравнение распределений переменных реальных и смоделированных данных для второго  $Z\gamma$  QCD контрольного региона. Штриховкой обозначена статистическая погрешность смоделированных данных.

### 3.2.3 $W\gamma$ контрольный регион

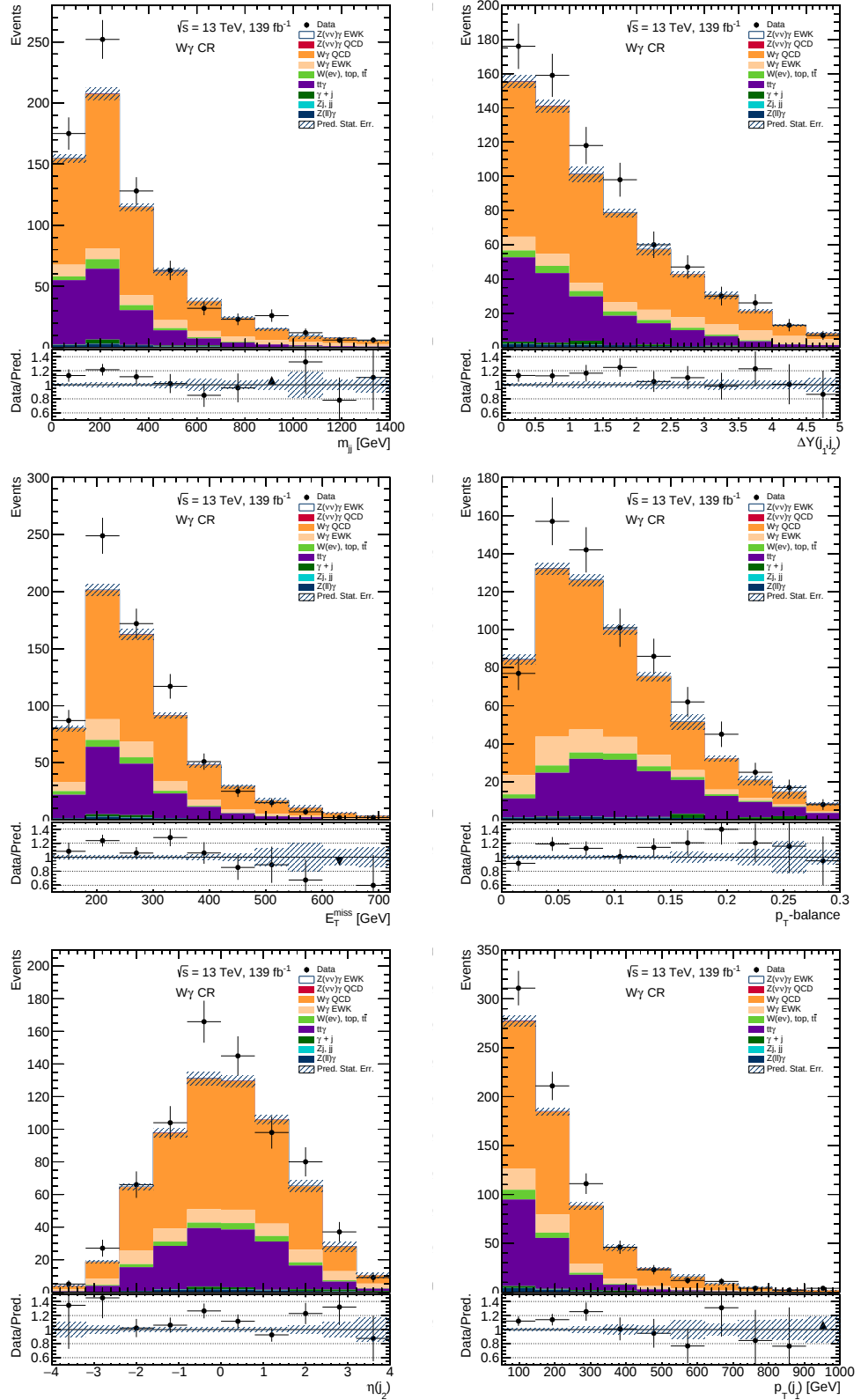


Рисунок 12 — Сравнение распределений переменных реальных и смоделированных данных для  $W\gamma$  контрольного региона. Штриховкой обозначена статистическая погрешность смоделированных данных.

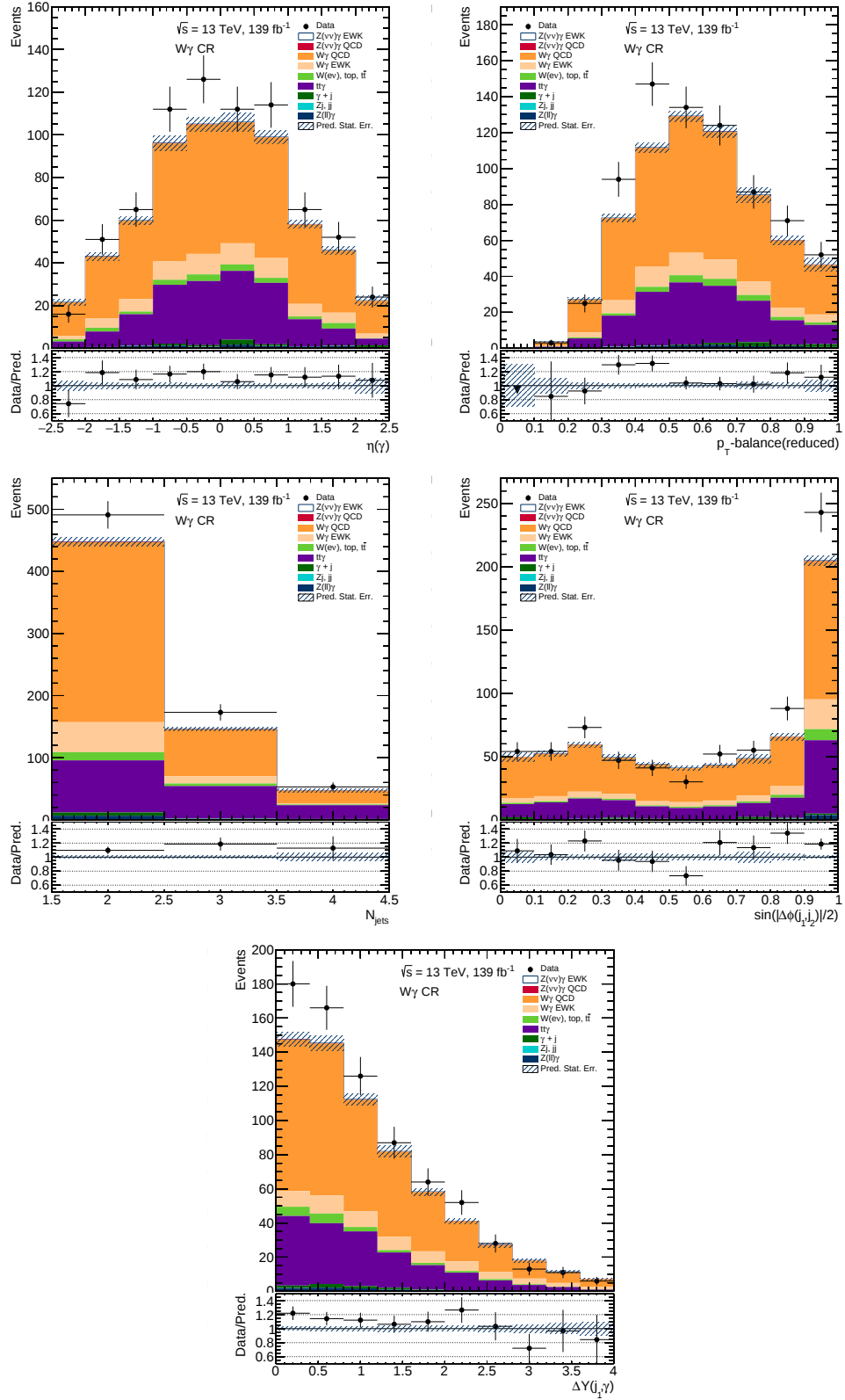


Рисунок 13 — Сравнение распределений переменных реальных и смоделированных данных для  $W\gamma$  контрольного региона. Штриховкой обозначена статистическая погрешность смоделированных данных.

### 3.3 Фиксированные отборы по переменным

Для сравнения эффективности применения методов машинного обучения были проведены фиксированные отборы по переменным центрально-сти фотона  $\zeta(\gamma)$ , инвариантной массы двух струй  $m_{jj}$ , баланса поперечных импульсов  $p_T - \text{balance}$  и разницы псевдобыстрот струй  $\Delta Y(j_1, i_2)$ , которые хорошо показали себя в предыдущем исследовании [3]. Была произведена оптимизация значимости для каждой комбинации ограничений. Самая большая значимость была получена для следующего порядка применения ограничений

$$\zeta(\gamma) \rightarrow m_{jj} \rightarrow p_T - \text{balance} \rightarrow \Delta Y(j_1, i_2) \quad (3.1)$$

Значения ограничений приведены в таблице 3. Максимум значимости и количества событий до и после применения ограничений представлены в таблице 6.

Таблица 3 — Порядок ограничений, обеспечивающий максимальную значимость.

| Переменная             | Ограничение         |
|------------------------|---------------------|
| $\zeta(\gamma)$        | $< 0.455$           |
| $m_{jj}$               | $> 697 \text{ GeV}$ |
| $p_T - \text{balance}$ | $< 0.064$           |
| $\Delta Y(j_1, i_2)$   | $> 2.227$           |

### 3.4 Тренировка и применение моделей

Для оптимизации работы алгоритмов и сравнения их эффективности был натренирован большой набор моделей BDT и MLP, которые создавались при помощи библиотеки машинного обучения TMVA [2], входящей в программный пакет ROOT [10]. Для подбора оптимальных параметров обучения модели тренировались с широким спектром комбинаций настроек, которые приведены в таблице 4.

Таблица 4 — Параметры обучения моделей.

| Модель | Параметр     | Значения                                                                                                                                          |
|--------|--------------|---------------------------------------------------------------------------------------------------------------------------------------------------|
| BDT    | NTrees       | 600, 500, 400, 300, 200, 100, 50                                                                                                                  |
|        | NCuts        | 20, 30, 40, 50                                                                                                                                    |
|        | MaxDepth     | 2, 3                                                                                                                                              |
|        | shrinkage    | 0.05, 0.1, 0.2, 0.5                                                                                                                               |
| MLP    | LearningRate | 0.001, 0.005, 0.01                                                                                                                                |
|        | BatchSize    | 3, 5, 10, 50                                                                                                                                      |
|        | HiddenLayers | N; N+10; N+20; N+50                                                                                                                               |
|        |              | N,N; N+10,N; N+20,N; N+50,N; N+5,N+5;<br>N+10,N+10; N+20,N+20; N+50,N+50                                                                          |
|        |              | N,N,N; N+10,N,N; N+20,N,N; N+50,N,N;<br>N+10,N+10,N; N+20,N+20,N; N+50,N+50,N;<br>N+5,N+5,N+5; N+10,N+10,N+10;<br>N+20,N+20,N+20; N+50,N+50,N+50; |

Для каждой модели были построены распределения отклика сигнала и фона при её применении к тренировочной и тестовой выборке для оценки переобучения. Также были построены графики значимости, для которых был определён максимум. Примеры распределений отклика и графиков значимости представлены на рисунках 14 и 15 соответственно. Сводные результаты применения моделей приведены в таблице 5, при этом перетренированные модели на уровне значимости 0.05 отбрасывались.

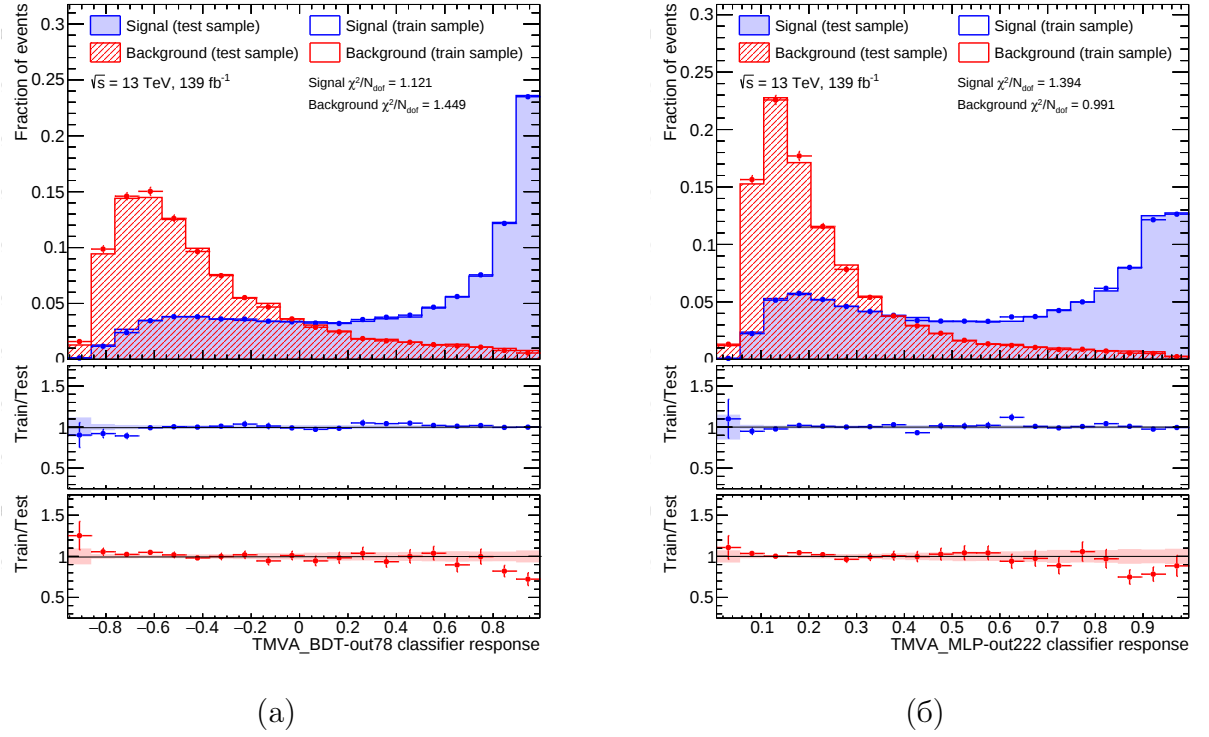


Рисунок 14 — Распределения отклика сигнала и фона для обучающей и тестовой выборки для моделей (а) BDT и (б) MLP.

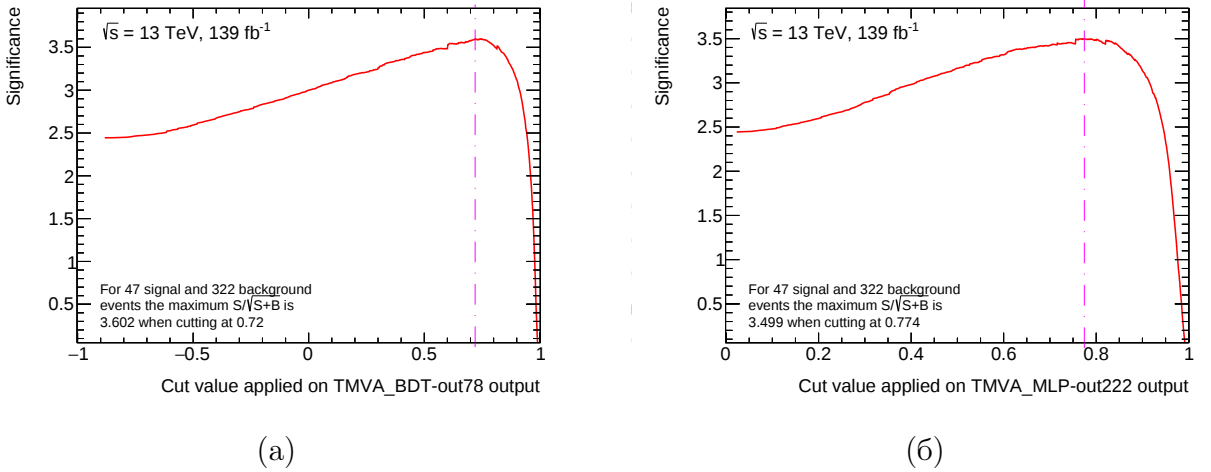


Рисунок 15 — Пример графиков зависимости значимости от ограничения по отклику для моделей (а) BDT и (б) MLP.

Максимальная величина значимости при применении фиксированных отборов более чем на одну величину стандартного отклонения меньше средних значимостей при применении моделей BDT и MLP, что говорит о превосходстве методов машинного обучения.

Таблица 5 — Значения максимальных и средних значений значимости для тренировок моделей с учётом и без учёта весов, где  $N_{\text{mod}}$  — число непереобученных моделей.

|     | С весами         |                   |                 | Без весов        |                   |                 |
|-----|------------------|-------------------|-----------------|------------------|-------------------|-----------------|
|     | $N_{\text{mod}}$ | Среднее           | Макс.           | $N_{\text{mod}}$ | Среднее           | Макс.           |
| BDT | 116              | $3.511 \pm 0.004$ | $3.60 \pm 0.04$ | 182              | $3.456 \pm 0.003$ | $3.55 \pm 0.03$ |
| MLP | 210              | $3.355 \pm 0.004$ | $3.57 \pm 0.05$ | 255              | $3.383 \pm 0.004$ | $3.58 \pm 0.06$ |

### 3.5 Влияние использования Монте-Карло весов на результаты обучения моделей

Для увеличения эффективности отделения сигнала от фона необходимо использование инструментов, обладающих большим функционалом и гибкостью настройки, по сравнению с TMVA, однако подавляющее большинство библиотек машинного обучения не поддерживает использование МК весов при обучении. Возникает необходимость проверки влияния использования весов событий на обучение моделей TMVA. Для этого был создан идентичный набор моделей без использования весов. Сводные результаты применения моделей, дающих наибольшую значимость, приведены в таблице 6. Настройки этих моделей приведены в таблице 7. Графики зависимости максимальной значимости от номера модели алгоритмов BDT и MLP представлены на рисунке 16.

Из результатов можно сделать вывод о том, что МК веса для набора, используемого в этом анализе не сильно влияют на результаты тренировки. Поэтому для создания моделей можно использовать библиотеки машинного обучения, не поддерживающие учёт МК весов.

Таблица 6 — Сравнение результатов применения фиксированных отборов и моделей с учётом и без учёта весов.  $\sigma_{\max}$  — максимальная значимость.

|                      | Вхожд.<br>сигнал | Вхожд.<br>фон | Кол-во<br>сигнала | Кол-во<br>фона  | $\sigma_{\max}$ |
|----------------------|------------------|---------------|-------------------|-----------------|-----------------|
| До отборов           | 90035            | 86888         | $46.9 \pm 0.2$    | $321.0 \pm 7.4$ | $2.44 \pm 0.03$ |
| Фикс. отборы         | 50362            | 11509         | $26.3 \pm 0.1$    | $39.7 \pm 2.5$  | $3.23 \pm 0.06$ |
| BDT78                | 52107            | 10267         | $27.2 \pm 0.1$    | $30.0 \pm 1.2$  | $3.60 \pm 0.04$ |
| BDT160 <sub>nw</sub> | 53480            | 10847         | $28.0 \pm 0.1$    | $34.2 \pm 1.0$  | $3.55 \pm 0.03$ |
| MLP222               | 52108            | 10529         | $27.2 \pm 0.1$    | $33.3 \pm 1.3$  | $3.57 \pm 0.05$ |
| MLP198 <sub>nw</sub> | 5183             | 9847          | $27.1 \pm 0.1$    | $30.2 \pm 0.9$  | $3.58 \pm 0.06$ |

Таблица 7 — Настройки моделей BDT и MLP с учётом и без учёта весов, обеспечивающие максимальную значимость.

| BDT       |       |                      | MLP            |        |                      |
|-----------|-------|----------------------|----------------|--------|----------------------|
|           | BDT78 | BDT160 <sub>nw</sub> |                | MLP222 | MLP198 <sub>nw</sub> |
| NTrees    | 400   | 200                  | LearningRate   | 0.01   |                      |
| NCuts     | 30    | 50                   | BatchSize      | 5      | 3                    |
| MaxDepth  | 2     | 3                    | HiddenLayers   | N+10,N |                      |
| shrinkage | 0.5   |                      | TrainingMethod | BP     |                      |
| BoostType | Grad  |                      | —              |        |                      |

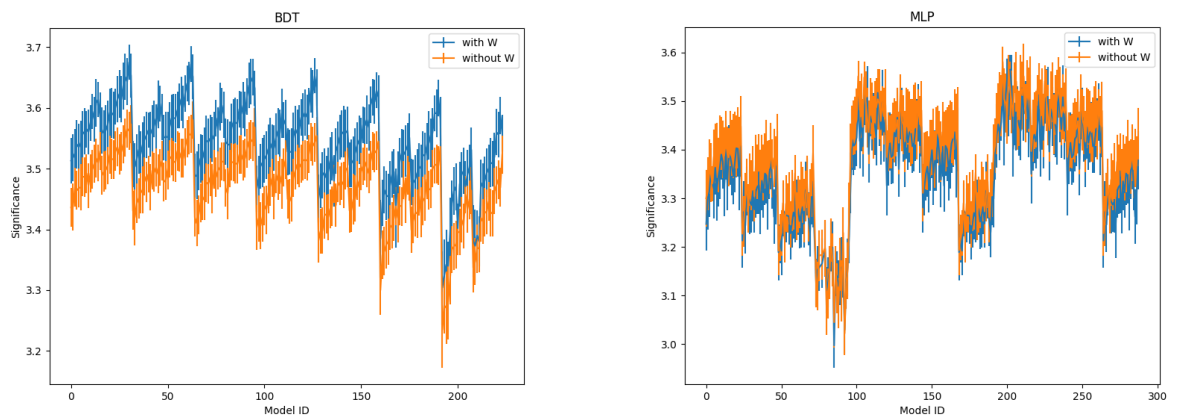


Рисунок 16 — Зависимости максимального значения значимости с использованием и без использования МК-весов событий при обучении для алгоритмов BDT и MLP от номера модели.

# Заключение

В процессе работы проведена проверка согласованности МК моделирования с реальными данными от столкновений, по результатам которой можно сказать, что модели, обученные на используемом в исследовании наборе МК данных могут применяться для классификации событий в реальных данных.

Было произведено сравнение результатов работы двух наиболее популярных алгоритмов машинного обучения MLP, BDT и стандартного метода фиксированных отборов по переменным при применении их к задаче разделения сигнальных и фоновых событий. Подобраны оптимальные настройки этих методов. Результат показал сильное превосходство алгоритмов MLP и BDT. С помощью применения алгоритма BDT удалось достичь значимости  $(3.60 \pm 0.04)\sigma$ . Максимальная значимость, полученная применением фиксированных отборов составляет  $(3.23 \pm 0.06)\sigma$ .

Проведено исследование влияния учёта МК весов событий при тренировке моделей. Определено, что учёт весов событий используемого набора МК данных незначительно влияет на качество работы моделей, что открывает возможность использования продвинутых библиотек машинного обучения, не поддерживающих использование МК весов событий при обучении.

В дальнейшем планируется перейти к более продвинутым моделям деревьев решений, таким как xgboost или lightgbm, и новым библиотекам нейронных сетей, таким как TensorFlow или PyTorch, для улучшения отделения сигнала от фона и получения большей значимости измерения ненулевого числа сигнальных событий.

## 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](#)].
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  $\sqrt{s}=8$  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. *Wu H.* A Stochastic Gradient Descent Theorem and the Back-Propagation Algorithm. — 2021. — arXiv: [2104.00539](https://arxiv.org/abs/2104.00539) [[math.OA](#)].
7. *Mokhtari A., Ribeiro A.* RES: Regularized Stochastic BFGS Algorithm // IEEE Transactions on Signal Processing. — 2014. — Dec. — Vol. 62, no. 23. — P. 6089–6104.
8. *Chen C.* [et al.]. Towards Practical Adam: Non-Convexity, Convergence Theory, and Mini-Batch Acceleration. — 2021. — arXiv: [2101.05471](https://arxiv.org/abs/2101.05471) [[cs.LG](#)].
9. *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.

10. *Brun R., Rademakers F.* ROOT — An object oriented data analysis framework // Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment. — 1997. — Vol. 389, no. 1. — P. 81–86. — New Computing Techniques in Physics Research V.