Национальный исследовательский ядерный университет
«МИФИ»

Кафедра физики элементарных частиц №40

Научная исследовательская работа студента на тему:

Применение методов машинного обучения для выделения процессов рассеяния векторных бозонов

Научный руководитель
к.ф.-м.н., доцент
Солдатов Евгений Юрьевич

Консультант
Петухов Александр Максимович

г. Москва 2021

Работа
студента 3-ого курса
Савельева Константина
Михайловича
ИЯФиТ

Введение

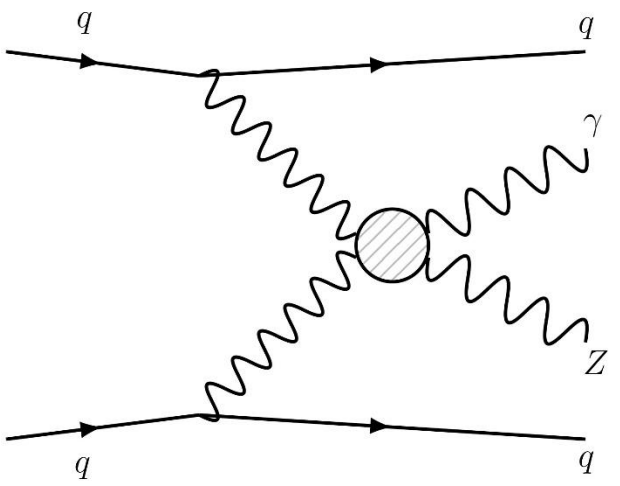


Диаграмма процесса рассеяния векторных бозонов

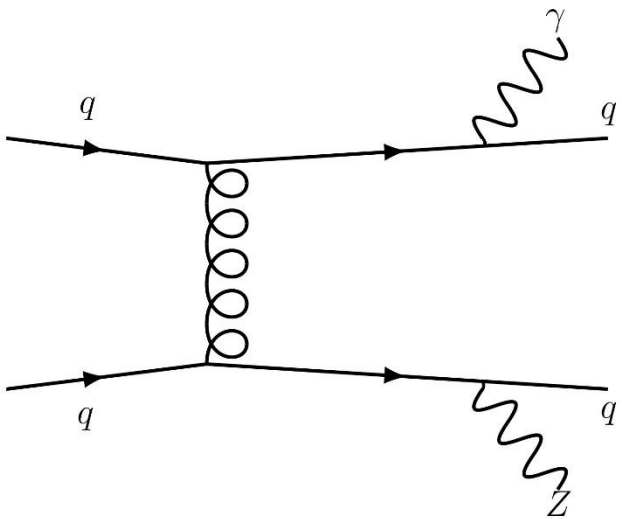
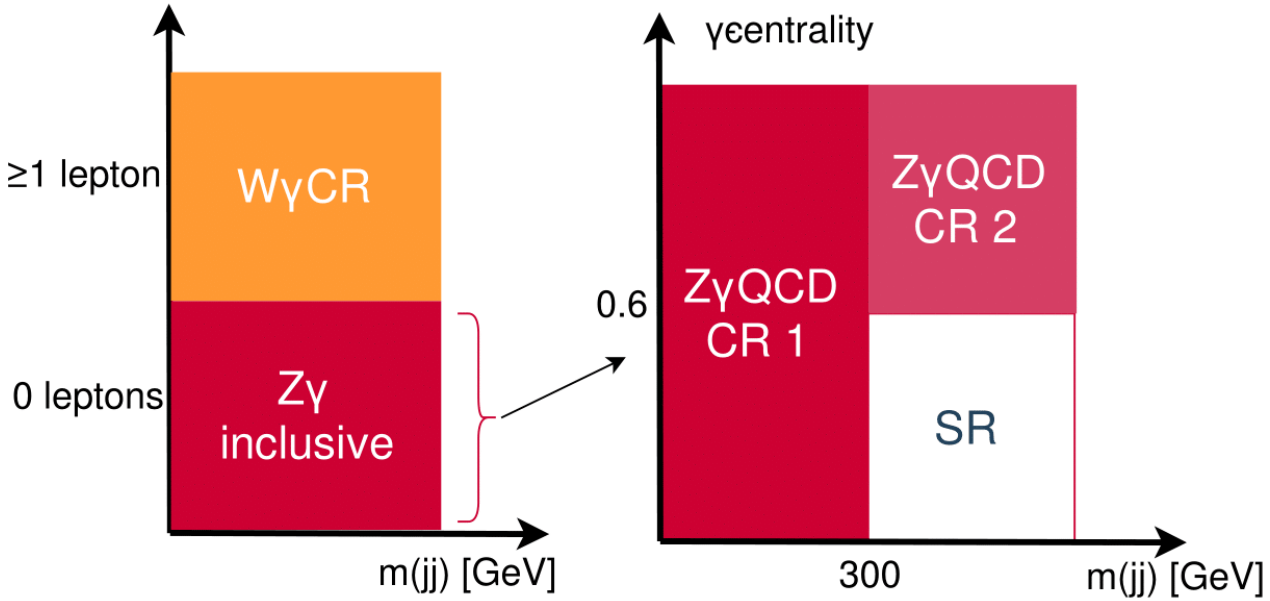


Диаграмма процесса КХД образования состояния Zγjj

- Целью работы является поиск редкого, ранее не наблюдавшегося 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. Одномерные фиксированные отборы не дают достаточной значимости, поэтому в работе исследовалось применение алгоритмов машинного обучения к разделению событий.

Используемые данные

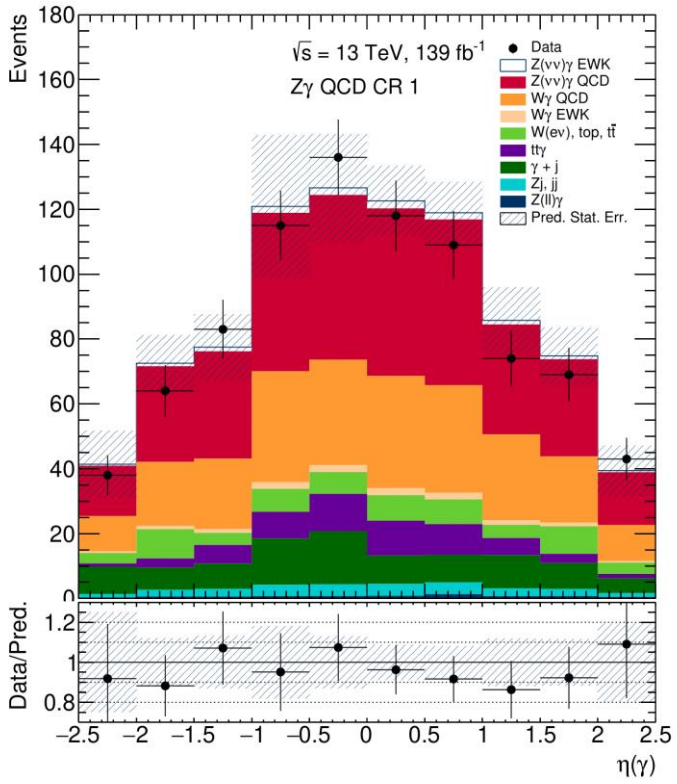
Работа производилась с Монте-Карло моделированными данными протон-протонных столкновений в детекторе ATLAS на БАКе с энергией 13TeV и интегральной светимостью 139 fb^{-1} и реальными данные, набранными в течении 2015-2018 гг.



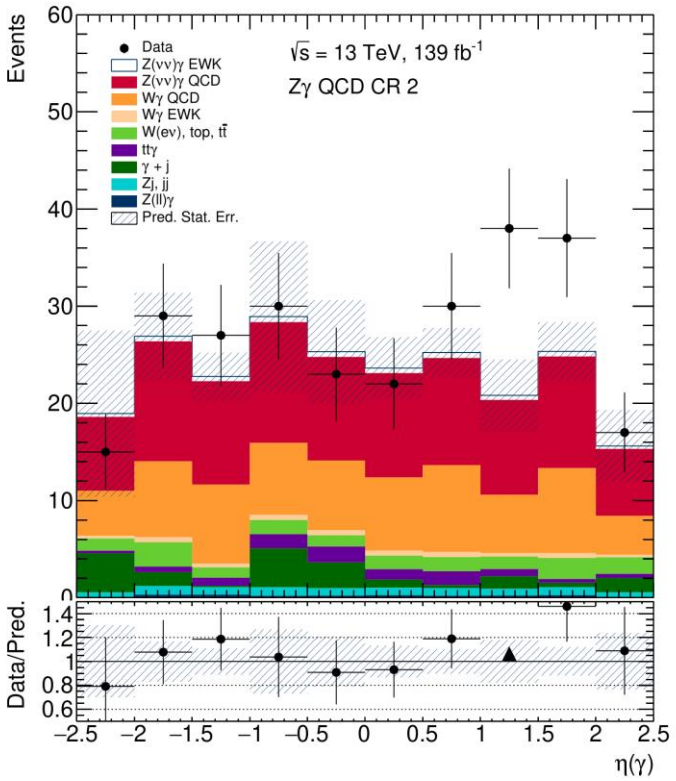
Проверка Монте-Карло моделирования данных

Для проверки согласованности Монте-Карло сгенерированных данных реальным данным в трёх регионах были построены распределения переменных, используемых для обучения моделей.

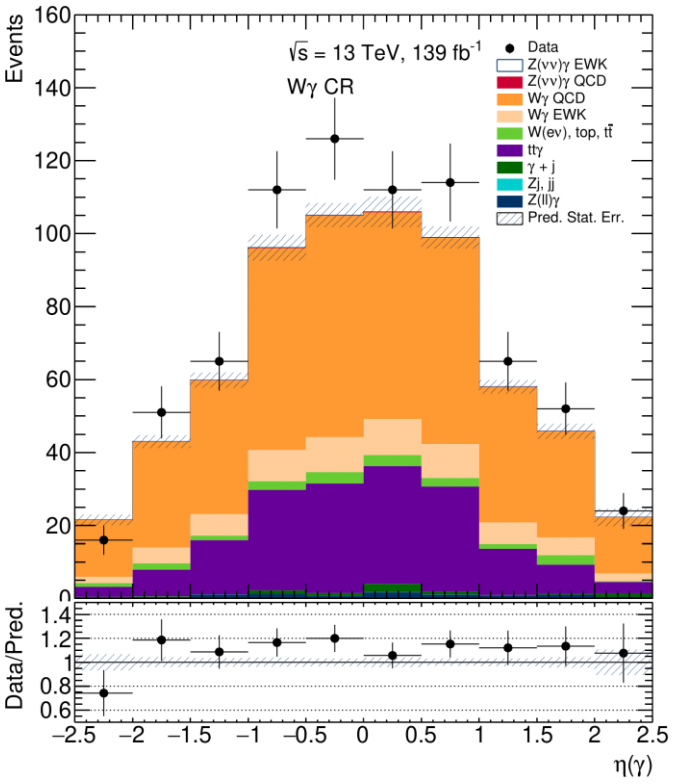
Первый Zγ QCD контрольный регион



Второй Zγ QCD контрольный регион



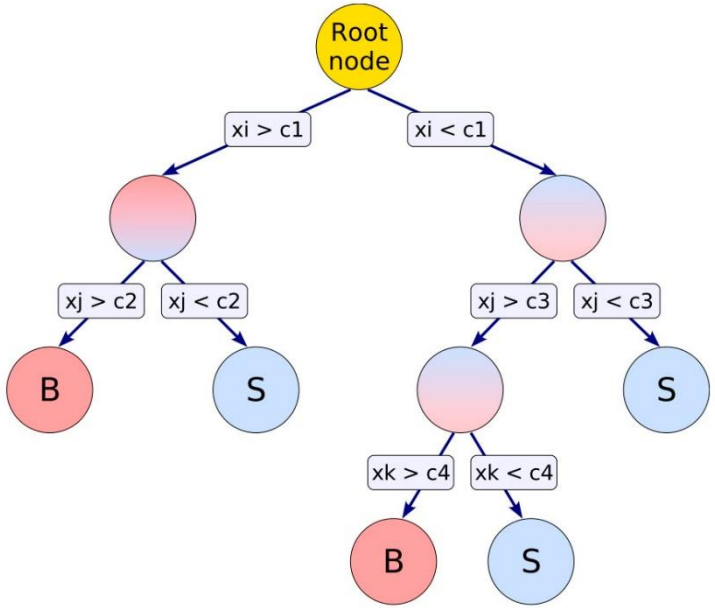
Wγ контрольный регион

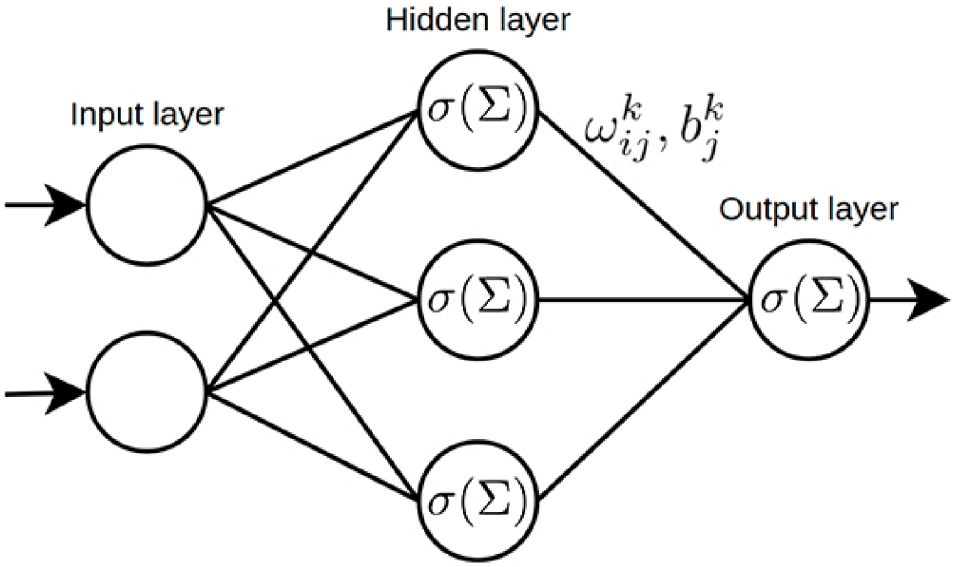


Алгоритмы BDT и MLP

Boosted Decision Trees (BDT) – это классификатор с бинарной древовидной структурой, позволяющий разбивать фазовое пространство на множество областей.

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



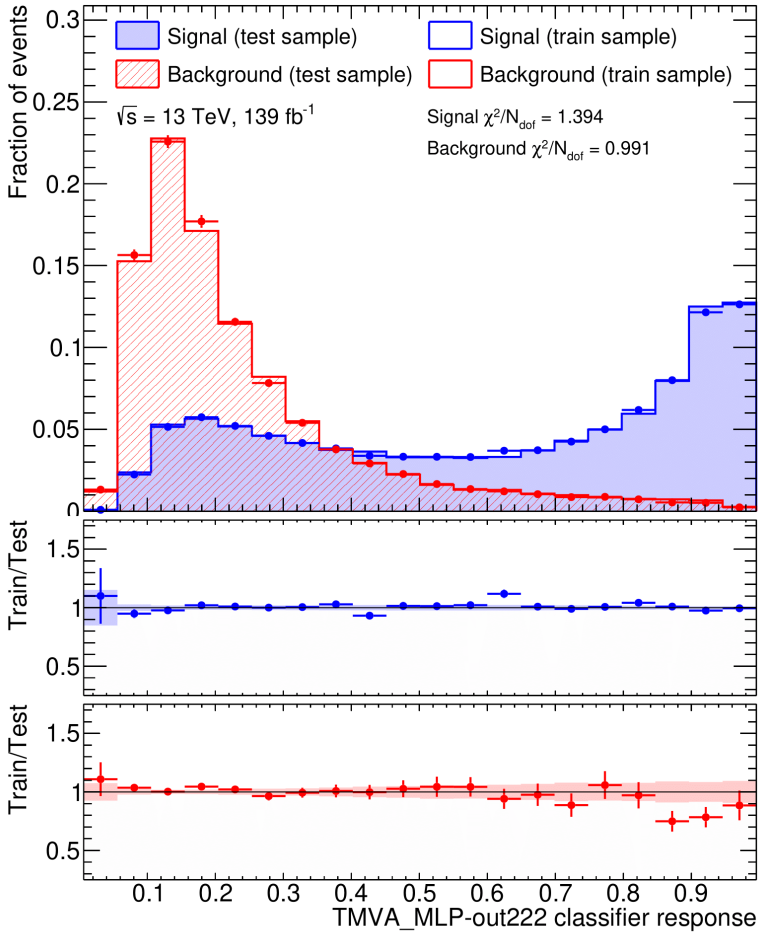


Алгоритм MLP (Multilayer Perceptron) – это одна из простейших моделей нейронных сетей.

- большей склонностью к переобучению,
- большое число параметров для настройки
- требует минимальной подготовки входных переменных
- может, в зависимости от структуры сети, иметь довольно большое время обучения.

Тренировка моделей. Подбор параметров.

В процессе работы был создан большой набор моделей с различными настройками. Для проверки перетренировки моделей сравнивались распределения отклика для тренировочной и тестовой выборок.

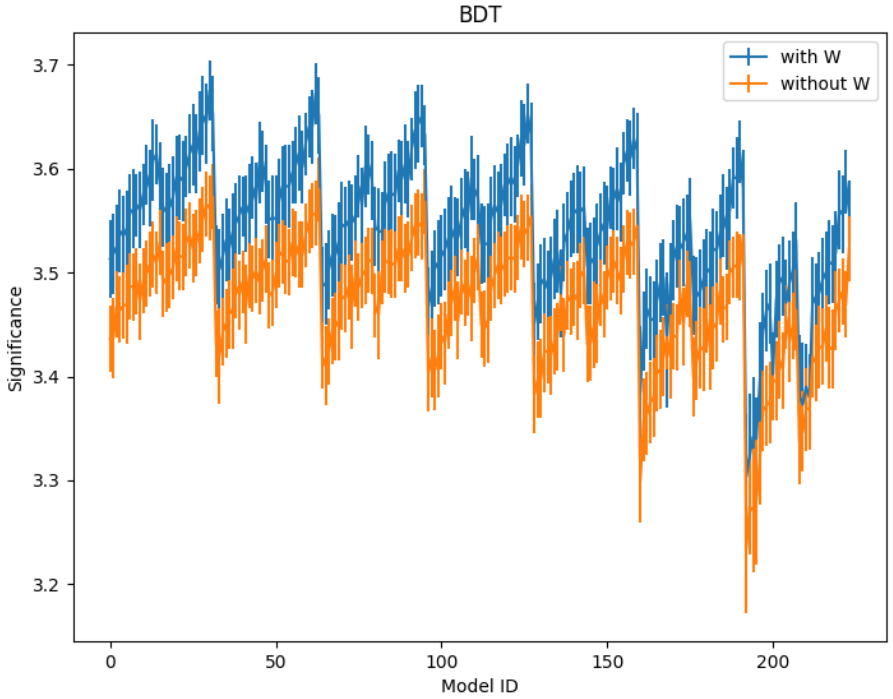


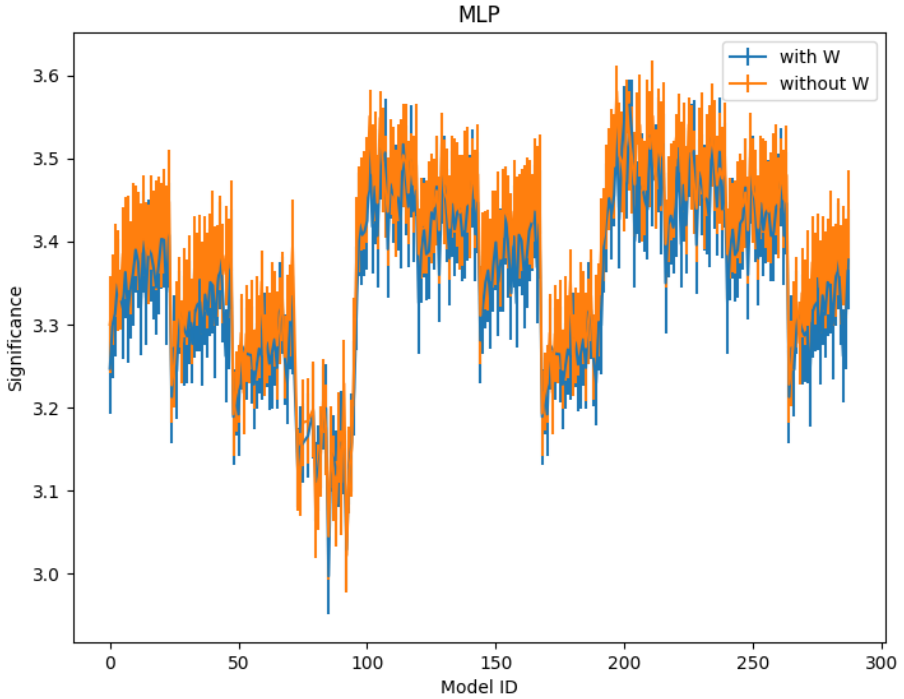
Были отобраны лучшие модели, обеспечивающие максимальную значимость. Методы BDT и MLP показали примерно одинаковый результат, больший по сравнению с фиксированными отборами.

	Вхожд. сигнал	Вхожд. фон	Кол-во сигнала	Кол-во фона	$\sigma_{\max}$
До отборов	90035	86888	46.9 ± 0.2	321.0 ± 7.4	2.44 ± 0.03
Фикс. отборы	50362	11509	26.3 ± 0.1	39.7 ± 2.5	3.23 ± 0.06
BDT78	52107	10267	27.2 ± 0.1	30.0 ± 1.2	3.60 ± 0.04
$\mathrm{BDT160_{nw}}$	53480	10847	28.0 ± 0.1	34.2 ± 1.0	3.55 ± 0.03
MLP222	52108	10529	27.2 ± 0.1	33.3 ± 1.3	3.57 ± 0.05
$\mathrm{MLP198_{nw}}$	5183	9847	27.1 ± 0.1	30.2 ± 0.9	3.58 ± 0.06

Влияние Монте-Карло весов событий на обучение

Подавляющее большинство библиотек машинного обучения, необходимых для перехода к более продвинутым моделям, не поддерживают использование Монте-Карло весов при обучении. Поэтому была проведена проверка того, как влияют веса на обучение моделей. Был натренирован идентичный набор моделей, но без учёта весов. Влияние весов на алгоритм BDT заметно, а MLP с весами и без показывает примерно одинаковые результаты.





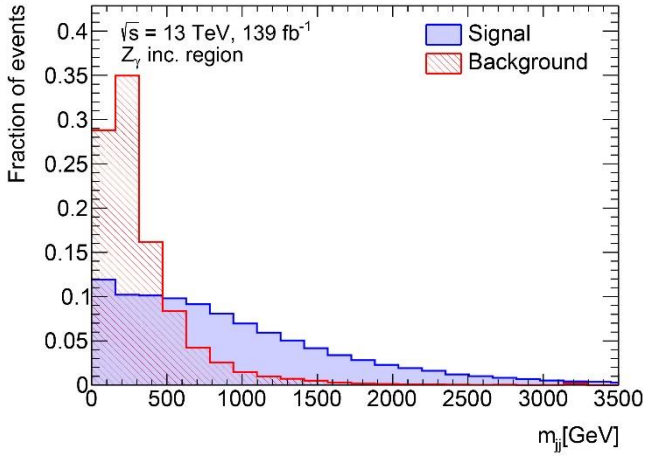
	С весами			Без весов		
	N_{mod}	Среднее	Макс.	N_{mod}	Среднее	Макс.
BDT	116	3.511 ± 0.004	3.60 ± 0.04	182	3.456 ± 0.003	3.55 ± 0.03
MLP	210	3.355 ± 0.004	3.57 ± 0.05	255	3.383 ± 0.004	3.58 ± 0.06

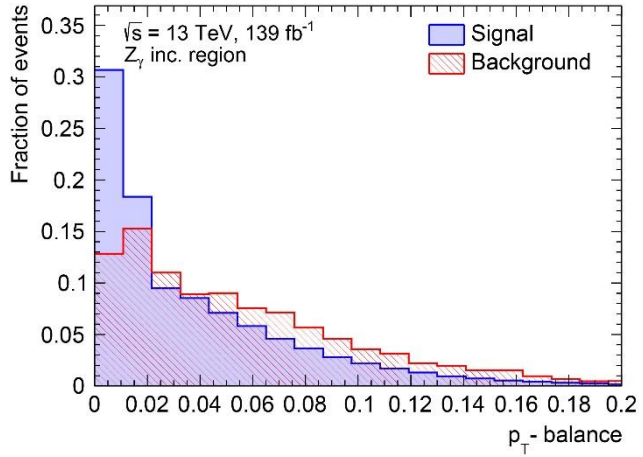
Заключение

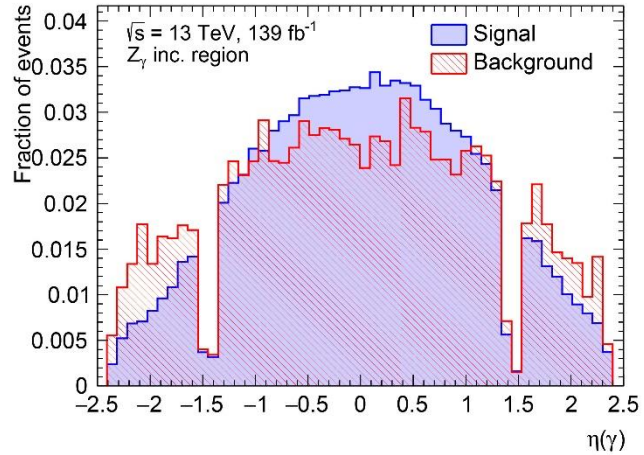
- Проведена проверка согласованности МК моделирования с реальными данными от столкновений.
- Было произведено сравнение результатов работы двух наиболее популярных алгоритмов машинного обучения MLP, BDT и стандартного метода фиксированных отборов.
- Проведено исследование влияния учёта МК весов событий при тренировке моделей.

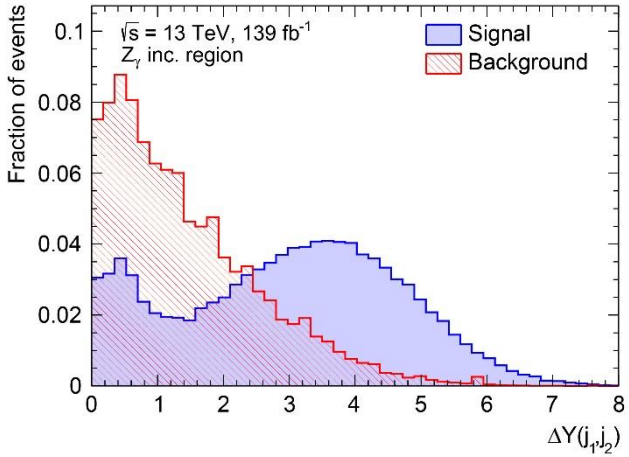
В дальнейшем планируется перейти к более продвинутым моделям деревьев решений, таким как xgboost или lightgbm, и новым библиотекам нейронных сетей, таким как TensorFlow или PyTorch, для улучшения отделения сигнала от фона и получения большей значимости и их тщательная настройка.

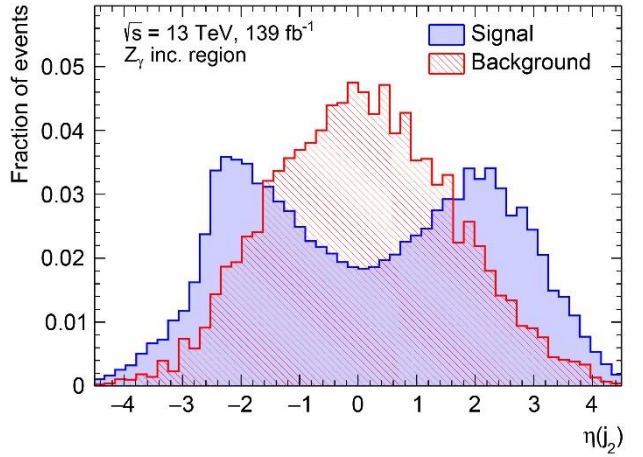
Спасибо за внимание!

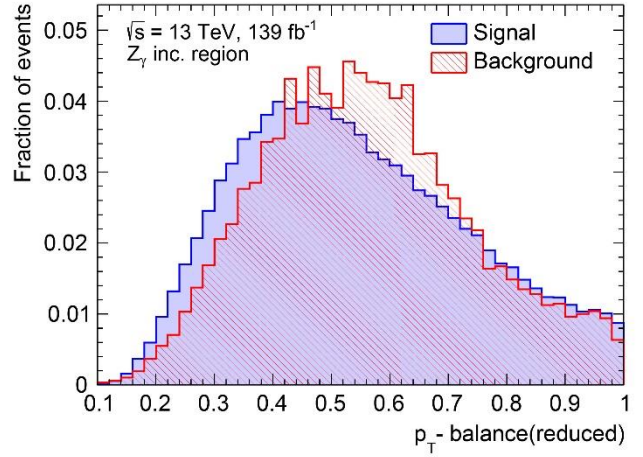

√s = 13 TeV, 139 fb-1
Zγ inc. region
Signal
Background
Fraction of events
mjj[GeV]

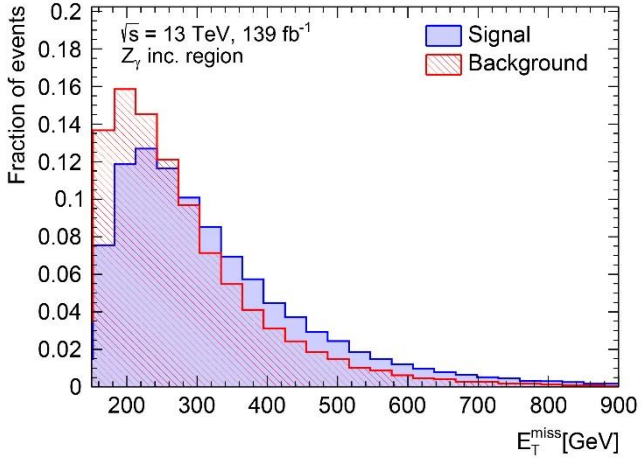

√s = 13 TeV, 139 fb-1
Zγ inc. region
Signal
Background
Fraction of events
pT- balance

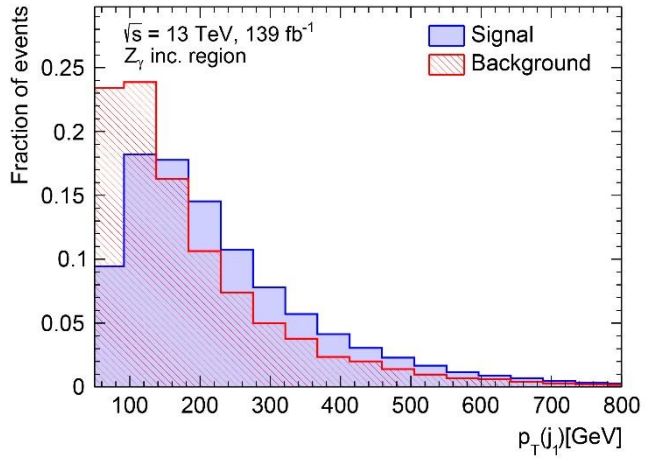

√s = 13 TeV, 139 fb-1
Zγ inc. region
Signal
Background
Fraction of events
η(γ)

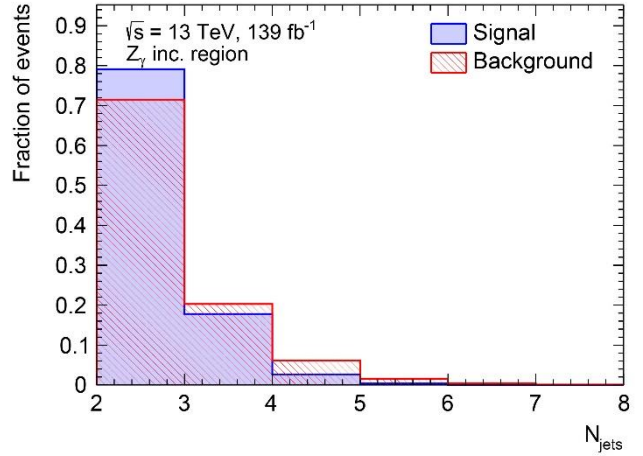

√s = 13 TeV, 139 fb-1
Zγ inc. region
Signal
Background
Fraction of events
ΔY(j1,j2)

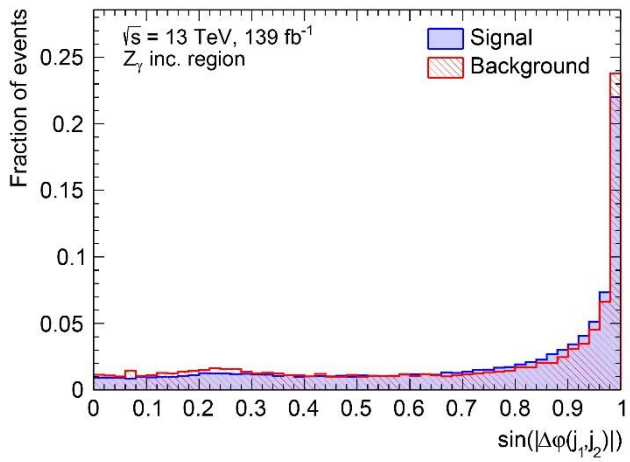

√s = 13 TeV, 139 fb-1
Zγ inc. region
Signal
Background
Fraction of events
η(j2)

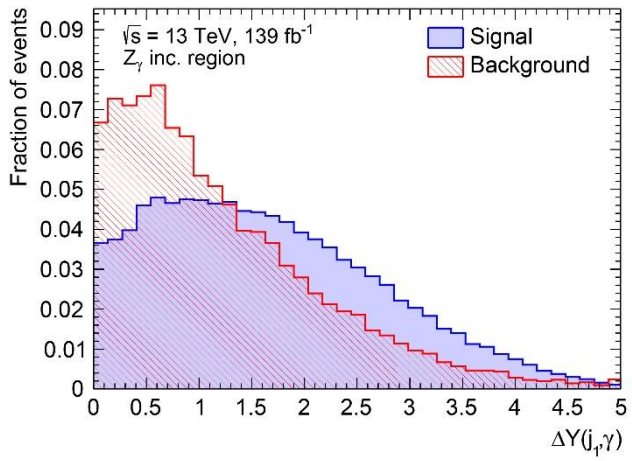

√s = 13 TeV, 139 fb-1
Zγ inc. region
Signal
Background
Fraction of events
pT- balance(reduced)


√s = 13 TeV, 139 fb-1
Zγ inc. region
Signal
Background
Fraction of events
ETmiss[GeV]


√s = 13 TeV, 139 fb-1
Zγ inc. region
Signal
Background
Fraction of events
pT(j1)[GeV]


√s = 13 TeV, 139 fb-1
Zγ inc. region
Signal
Background
Fraction of events
Njets


√s = 13 TeV, 139 fb-1
Zγ inc. region
Signal
Background
Fraction of events
0.25
0.2
0.15
0.1
0.05
0
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1
sin(|Δφ(j1,j2)|)


√s = 13 TeV, 139 fb-1
Zγ inc. region
Signal
Background
Fraction of events
0.09
0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0
0 0.5 1 1.5 2 2.5 3 3.5 4 4.5 5
ΔY(j1,γ)

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	$N_e = 0, N_\mu = 0$
$\lvert\Delta\phi(\gamma, \vec{p}_T^{\,miss})\rvert$	> 0.4
$\lvert\Delta\phi(j_1, \vec{p}_T^{\,miss})\rvert$	> 0.3
$\lvert\Delta\phi(j_2, \vec{p}_T^{\,miss})\rvert$	> 0.3

Переменная	Ограничение
$W\gamma$ контрольный регион	
N_{leptons}	≥ 1
$Z\gamma$ QCD контрольный регион 1	
N_{leptons}	$= 0$
m_{jj}	< 300 GeV
$Z\gamma$ QCD контрольный регион 2	
N_{leptons}	$= 0$
m_{jj}	> 300 GeV
γ-centrality	> 0.6
$Z\gamma$ EWK сигнальный регион	
N_{leptons}	$= 0$
m_{jj}	> 300 GeV
γ-centrality	< 0.6

- m_{jj} – инвариантаная масса двух струй
- $\Delta Y(j_1, j_2)$ – разность быстрот двух струй
- E_T^{miss} – недостающий поперечный импульс
- $p_T - \text{balance} = \frac{|\vec{p}_T{}^{miss} + \vec{p}_T{}^{\gamma} + \vec{p}_T{}^{j1} + \vec{p}_T{}^{j2}|}{E_T^{miss} + E_T{}^{\gamma} + p_T{}^{j1} + p_T{}^{j2}}$ – 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
- $p_T - \text{balance(reduced)} = \frac{|\vec{p}_T{}^{\gamma} + \vec{p}_T{}^{j1} + \vec{p}_T{}^{j2}|}{E_T^{\gamma} + p_T{}^{j1} + p_T{}^{j2}}$ – сокращённый баланс поперечных импульсов
- N_{jets} – число адронных струй
- $\sin|\frac{\Delta\varphi(j_1,j_2)}{2}|$
- $\Delta Y(j_1, \gamma)$ – разность псевдобыстрот между струёй и фотоном

$$p_T - \text{balance} = \frac{|\vec{p}_T^{\,miss} + \vec{p}_T^{\,\gamma} + \vec{p}_T^{\,j_1} + \vec{p}_T^{\,j_2}|}{E_T^{miss} + E_T^{\gamma} + p_T^{j_1} + p_T^{j_2}}$$

$$p_T - \text{balance(reduced)} = \frac{|\vec{p}_T^{\,\gamma} + \vec{p}_T^{\,j_1} + \vec{p}_T^{\,j_2}|}{E_T^{\gamma} + p_T^{j_1} + p_T^{j_2}}$$

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

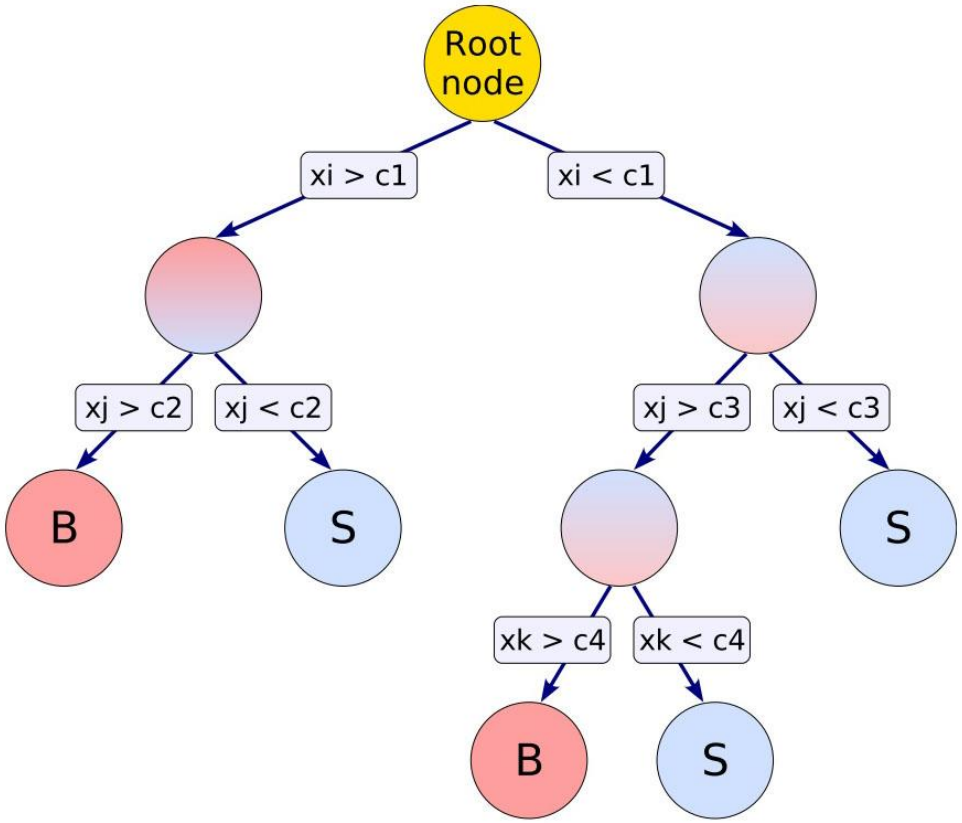
где η_γ – псевдобыстрота фотона, η_{j_1}, η_{j_2} – псевдобыстроты струй.

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

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; 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; N+10,N+10,N; N+20,N+20,N; N+50,N+50,N; N+5,N+5,N+5; N+10,N+10,N+10; N+20,N+20,N+20; N+50,N+50,N+50;

BDT			MLP		
	BDT78	$BDT160_{nw}$		MLP222	$MLP198_{nw}$
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		–		

Алгоритм Boosted Decision Trees (BDT)

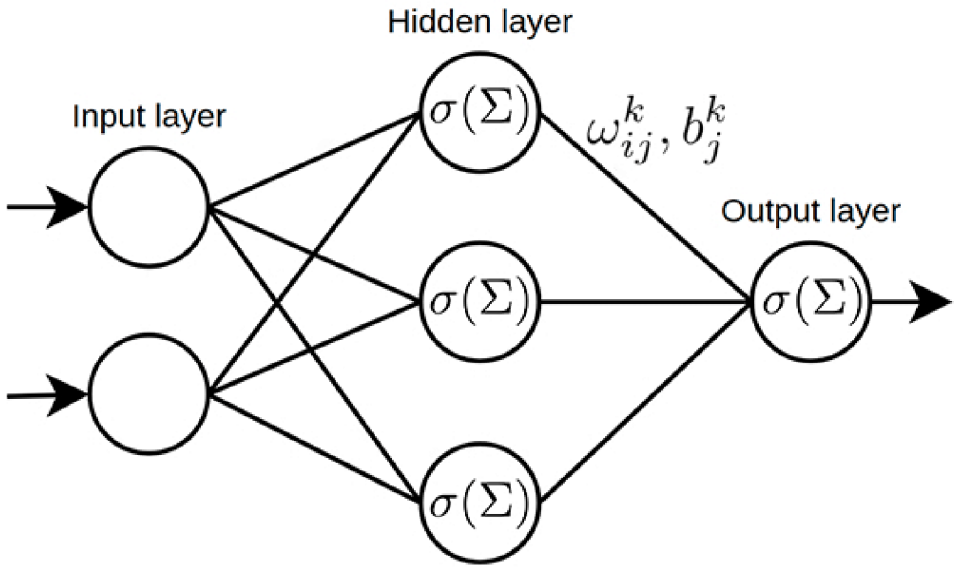


Работа метода заключается в создании леса бинарных деревьев решений.

Принцип их построения заключается в том, что для каждого узла дерева определяется переменная и ограничение по ней, которые обеспечивают максимальное разделение сигнала и фона в дочернем узле.

Для того, чтобы исключить влияние флуктуаций в данных, применяемых для обучения классификатора, используется бустинг. Он заключается в создании леса деревьев решений, в котором при создании каждого последующего дерева алгоритм переоцениевает веса событий таким образом, чтобы уделить больше внимания неверно классифицированным событиям.

Алгоритм Multilayer Perceptron



Принцип его работы заключается в последовательном применении линейных преобразований к переменным между слоями нейронов и передаточной функции для получения отклика.

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

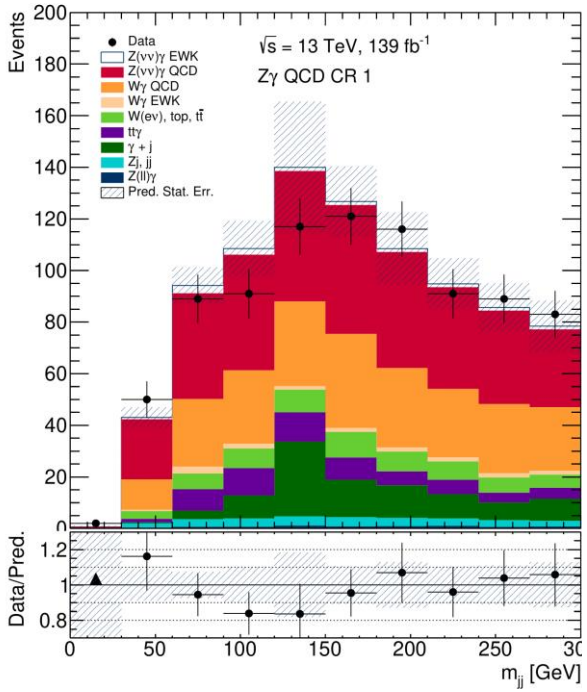
Одним из методов численной оптимизации определяется минимум функции ошибок, которая определяет качество классификации событий, например

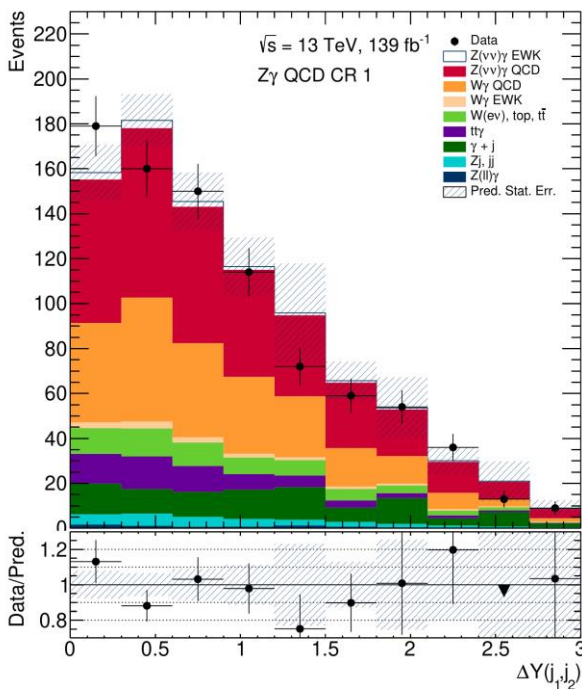
$$f_{loss}(\vec{y}) = \frac{1}{2}\sum(y_i - \hat{y}_i)^2$$

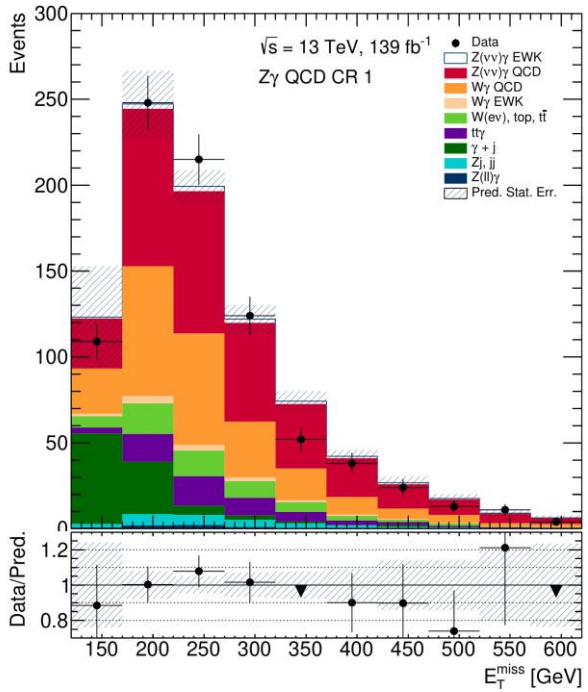
Простейшим методом оптимизации является алгоритм обратного распространения ошибки, при котором ошибка распространяется от выходов сети к её входам. Веса и смещения на каждой итерации изменяются на величины:

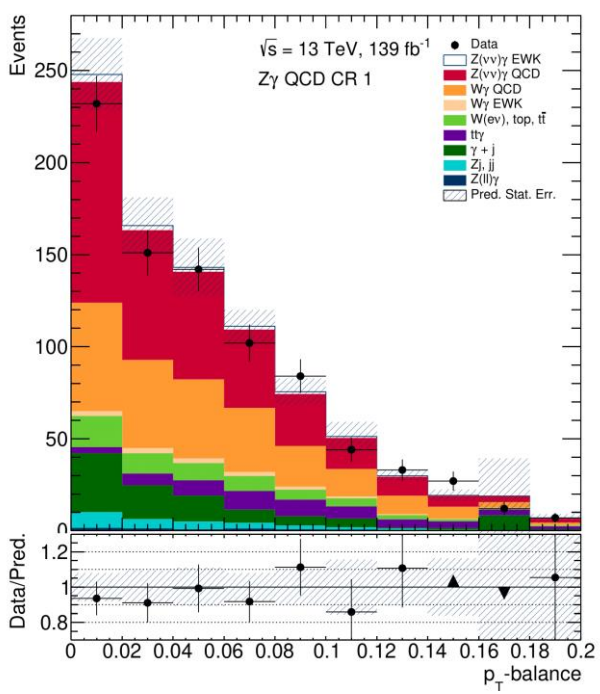
$$\Delta w_{ij}^k = -\alpha\frac{\partial f_{loss}}{\partial w_{ij}^k}, \qquad \Delta b_j^k = -\alpha\frac{\partial f_{loss}}{\partial b_j^k}$$

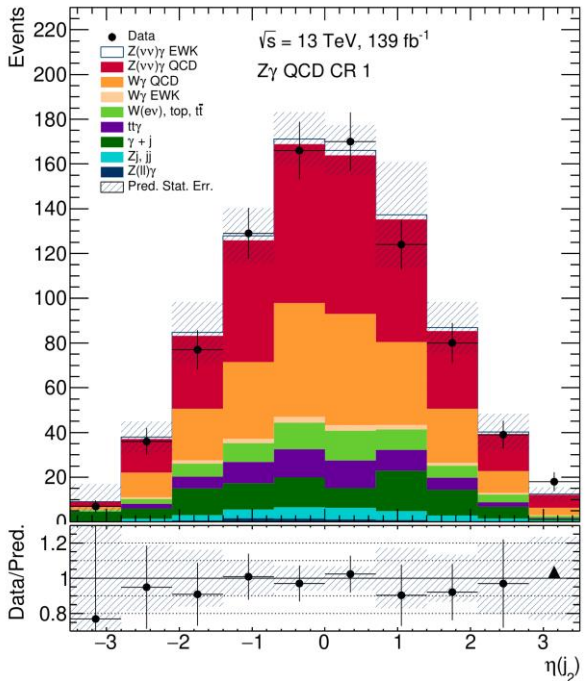
Первый Zγ QCD контрольный регион

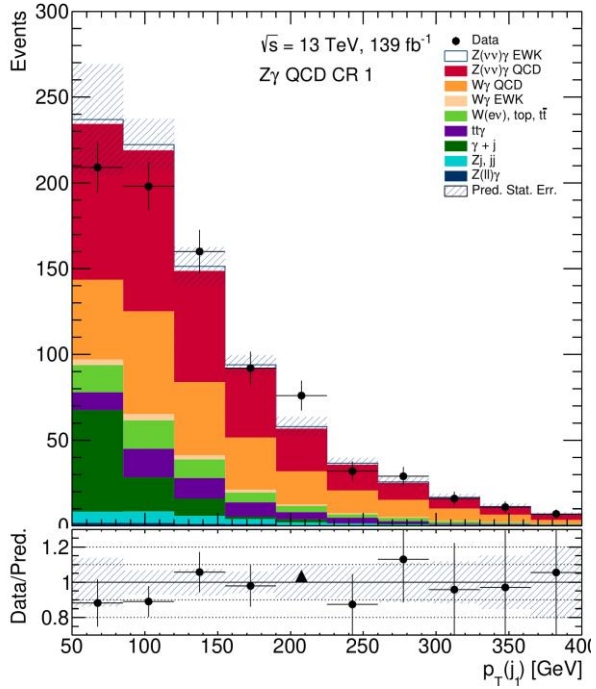


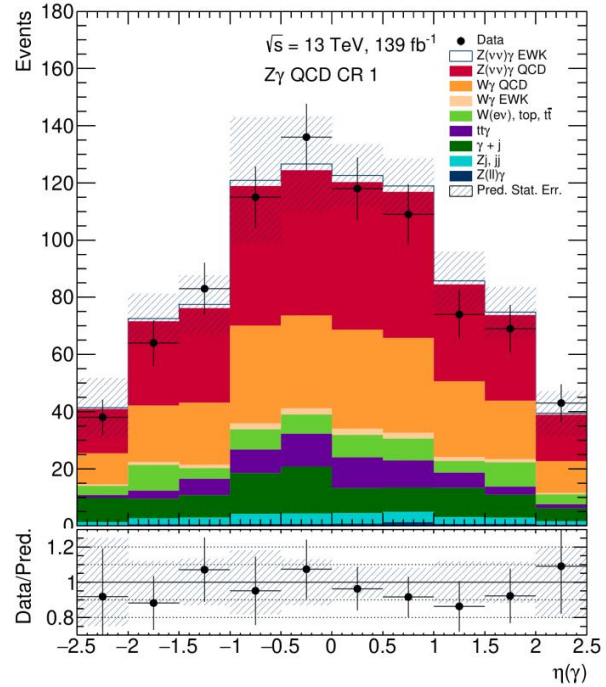


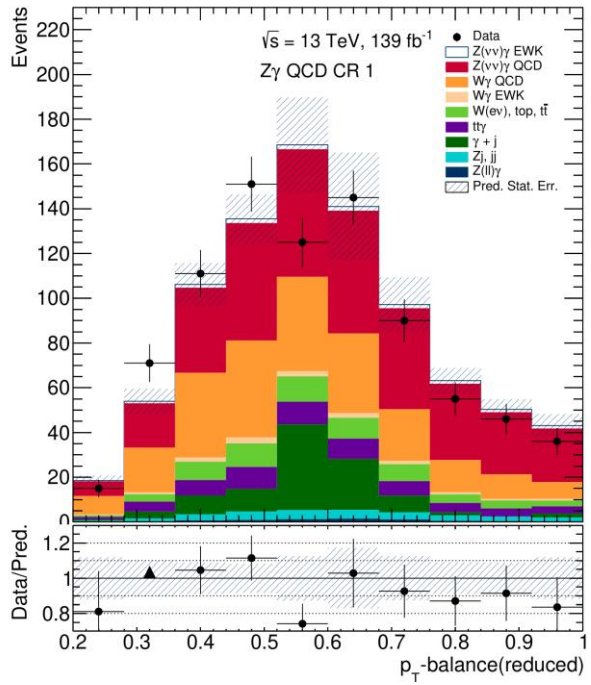


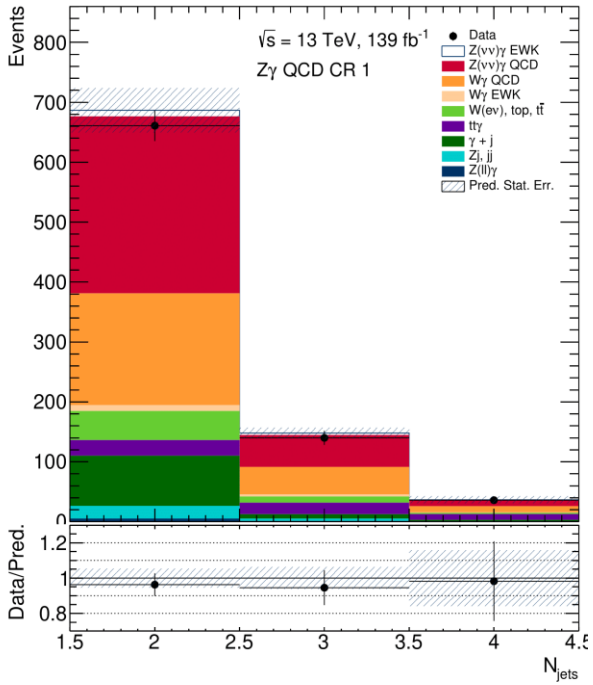


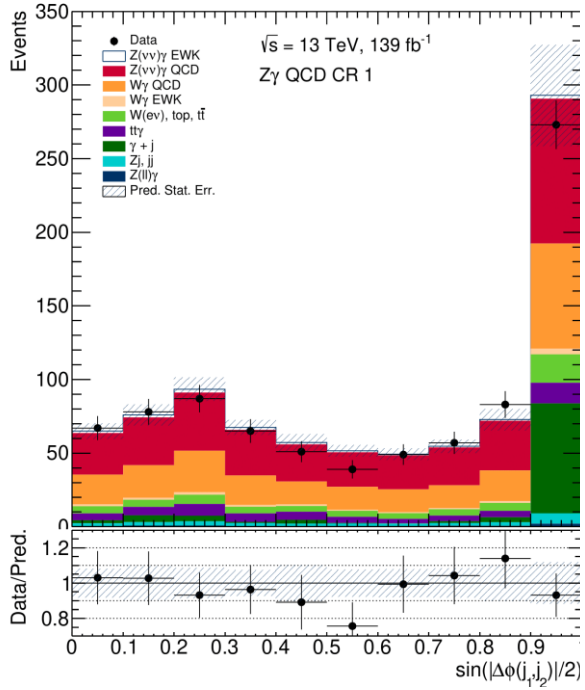

Events
√s = 13 TeV, 139 fb-1
Zγ QCD CR 1
Data
Z(νν)γ EWK
Z(νν)γ QCD
Wγ QCD
Wγ EWK
W(eν), top, tt̄
ttγ
γ + j
Zj, jj
Z(ll)γ
Pred. Stat. Err.
Data/Pred.
η(j2)

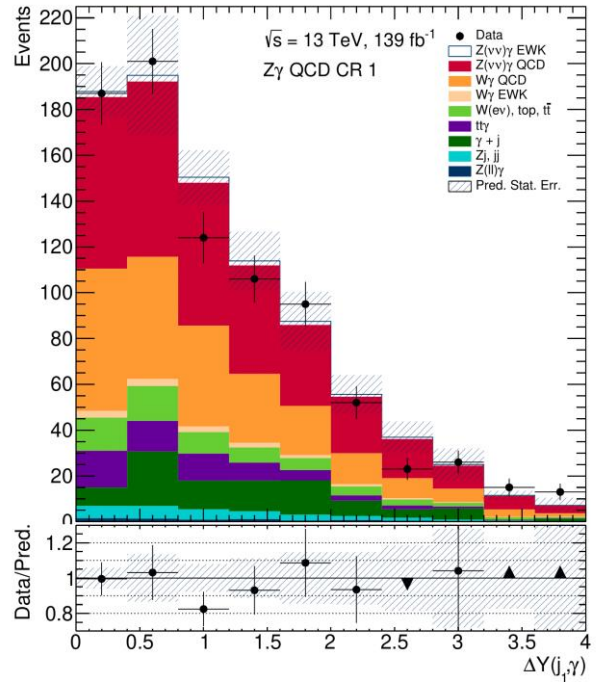

Events
√s = 13 TeV, 139 fb-1
Zγ QCD CR 1
Data
Z(νν)γ EWK
Z(νν)γ QCD
Wγ QCD
Wγ EWK
W(eν), top, tt̄
ttγ
γ + j
Zj, jj
Z(ll)γ
Pred. Stat. Err.
Data/Pred.
pT(j1) [GeV]


Events
√s = 13 TeV, 139 fb-1
Zγ QCD CR 1
Data
Z(νν)γ EWK
Z(νν)γ QCD
Wγ QCD
Wγ EWK
W(eν), top, tt̄
ttγ
γ + j
Zj, jj
Z(ll)γ
Pred. Stat. Err.
Data/Pred.
η(γ)

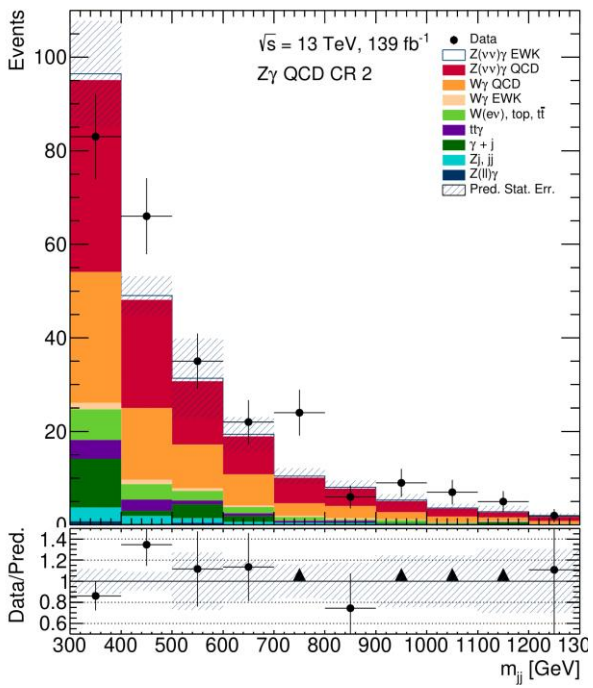

Events
√s = 13 TeV, 139 fb-1
Zγ QCD CR 1
Data
Z(νν)γ EWK
Z(νν)γ QCD
Wγ QCD
Wγ EWK
W(eν), top, tt̄
ttγ
γ + j
Zj, jj
Z(ll)γ
Pred. Stat. Err.
Data/Pred.
pT-balance(reduced)

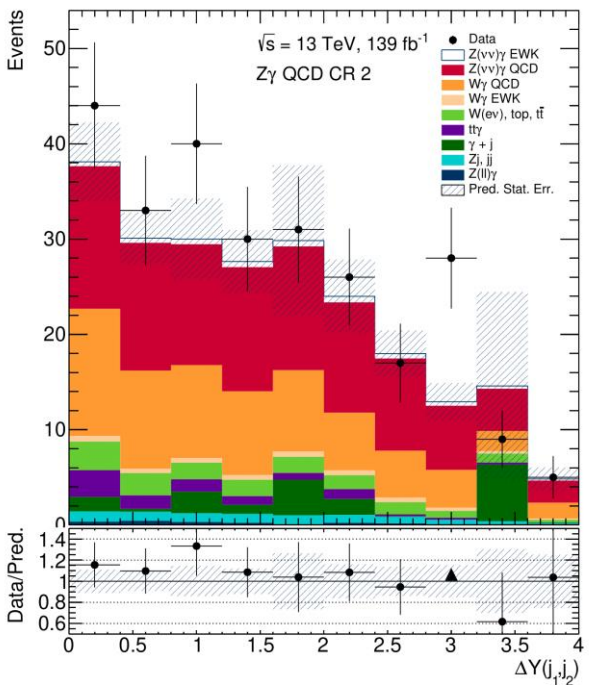

Events
√s = 13 TeV, 139 fb-1
Zγ QCD CR 1
Data
Z(νν)γ EWK
Z(νν)γ QCD
Wγ QCD
Wγ EWK
W(eν), top, tt̄
ttγ
γ + j
Zj, jj
Z(ll)γ
Pred. Stat. Err.
Data/Pred.
Njets

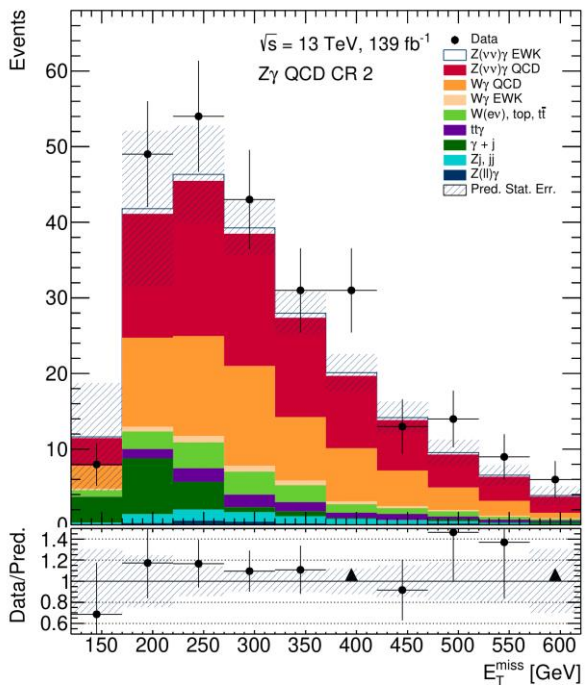

Events
√s = 13 TeV, 139 fb-1
Zγ QCD CR 1
Data
Z(νν)γ EWK
Z(νν)γ QCD
Wγ QCD
Wγ EWK
W(eν), top, tt̄
ttγ
γ + j
Zj, jj
Z(ll)γ
Pred. Stat. Err.
Data/Pred.
sin(|Δφ(j1,j2)|/2)

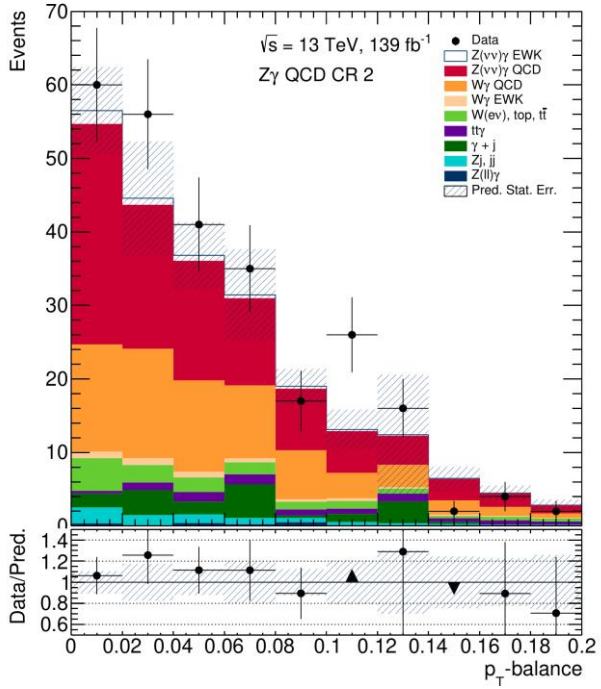

Events
√s = 13 TeV, 139 fb-1
Zγ QCD CR 1
Data
Z(νν)γ EWK
Z(νν)γ QCD
Wγ QCD
Wγ EWK
W(eν), top, tt̄
ttγ
γ + j
Zj, jj
Z(ll)γ
Pred. Stat. Err.
Data/Pred.
ΔY(j1,γ)

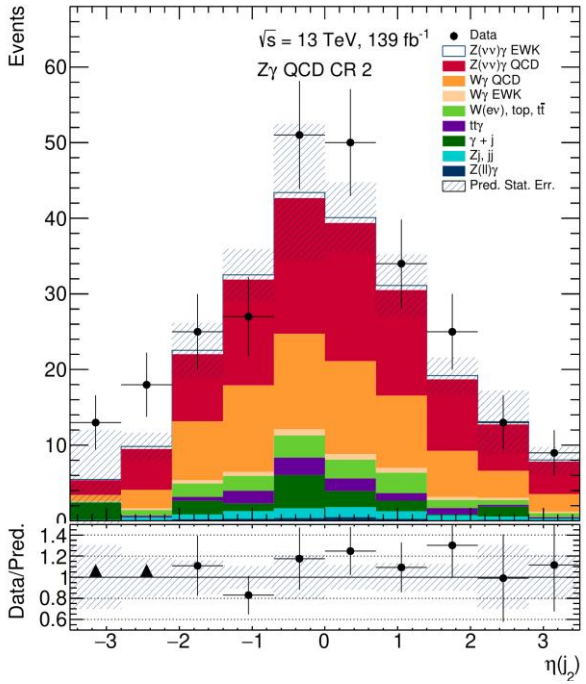
Второй Zγ QCD контрольный регион

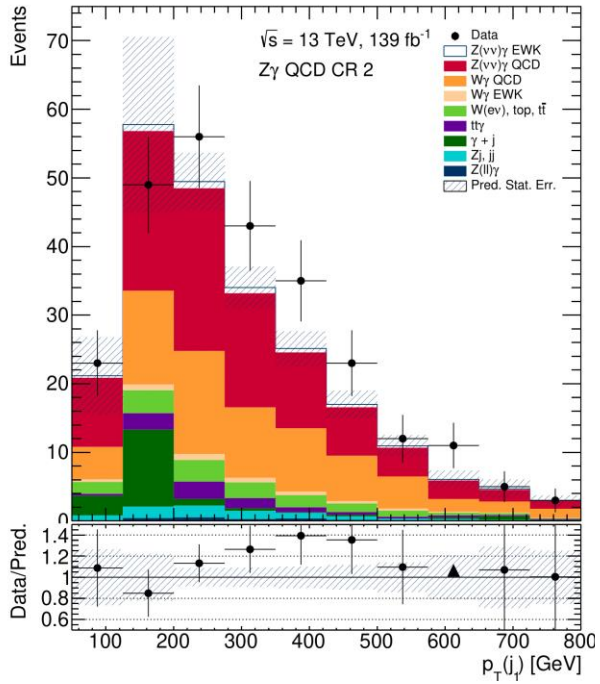


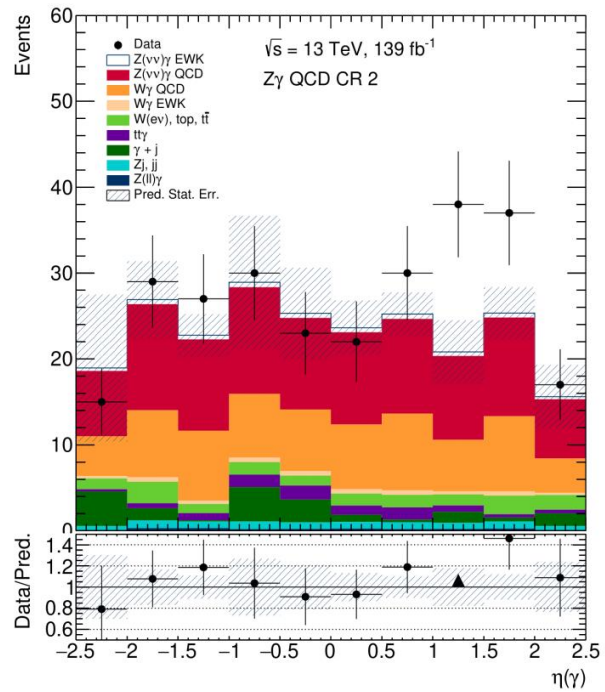


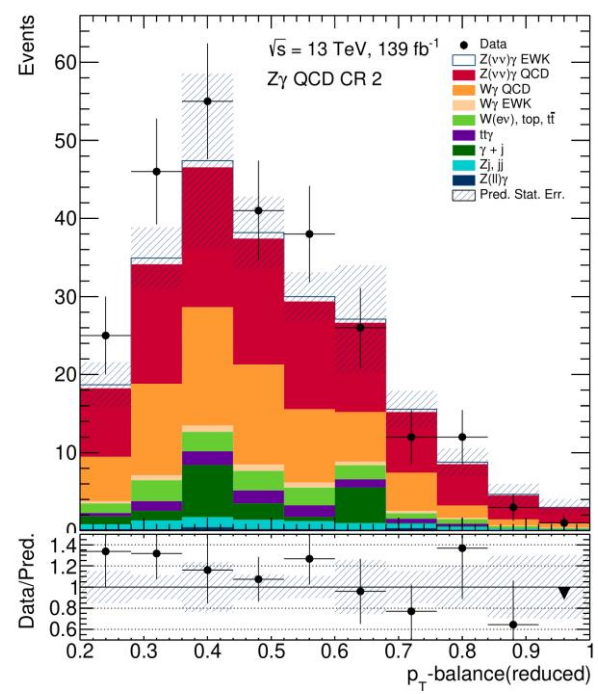


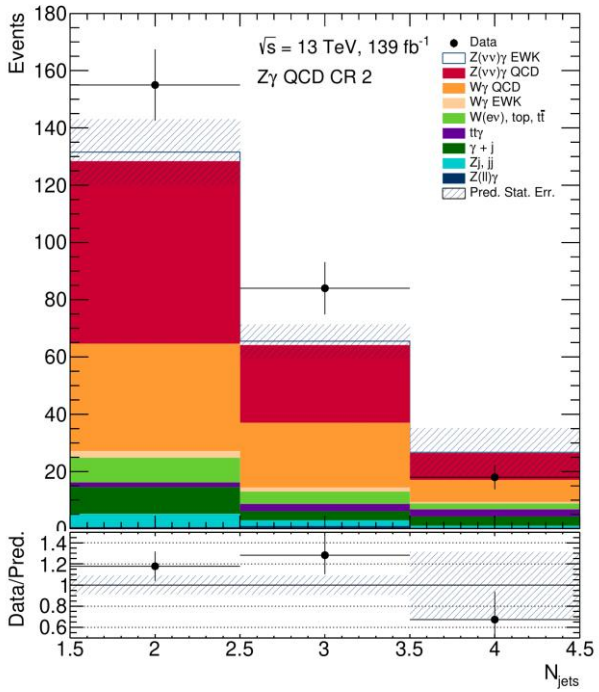


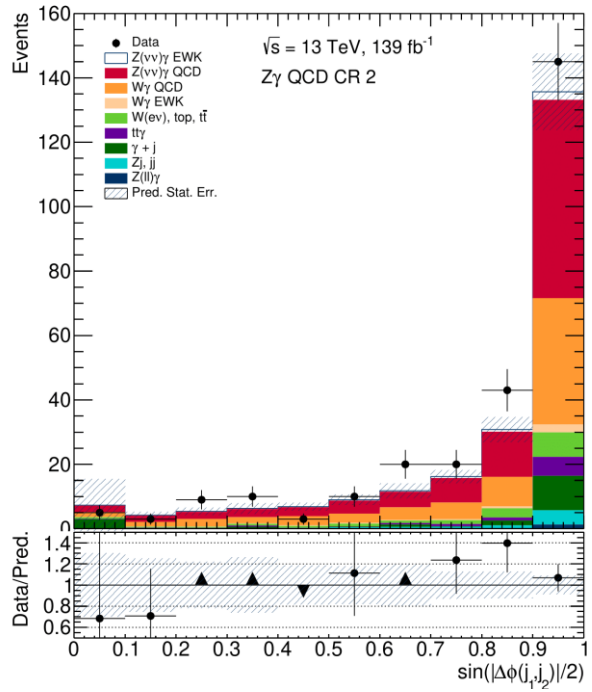

Events
√s = 13 TeV, 139 fb-1
Zγ QCD CR 2
Data
Z(νν)γ EWK
Z(νν)γ QCD
Wγ QCD
Wγ EWK
W(eν), top, tt̄
ttγ
γ + j
Zj, jj
Z(ll)γ
Pred. Stat. Err.
Data/Pred.
η(j2)

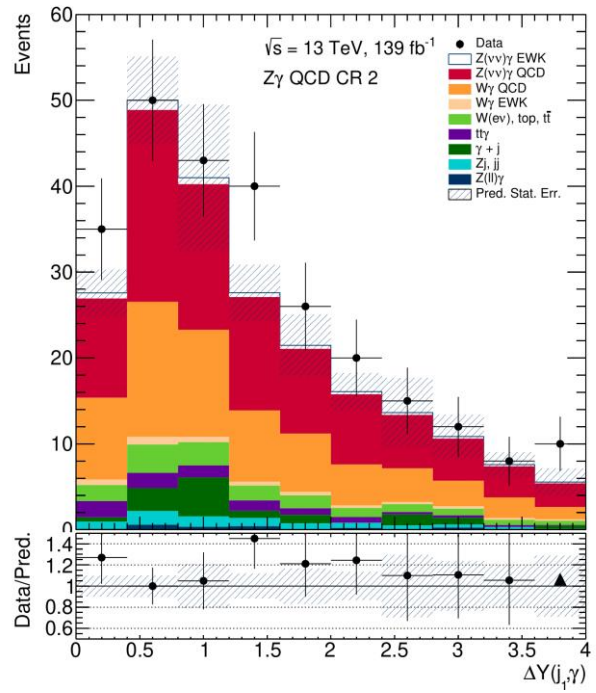

Events
√s = 13 TeV, 139 fb-1
Zγ QCD CR 2
Data
Z(νν)γ EWK
Z(νν)γ QCD
Wγ QCD
Wγ EWK
W(eν), top, tt̄
ttγ
γ + j
Zj, jj
Z(ll)γ
Pred. Stat. Err.
Data/Pred.
pT(j1) [GeV]


Events
√s = 13 TeV, 139 fb-1
Zγ QCD CR 2
Data
Z(νν)γ EWK
Z(νν)γ QCD
Wγ QCD
Wγ EWK
W(eν), top, tt̄
ttγ
γ + j
Zj, jj
Z(ll)γ
Pred. Stat. Err.
Data/Pred.
η(γ)

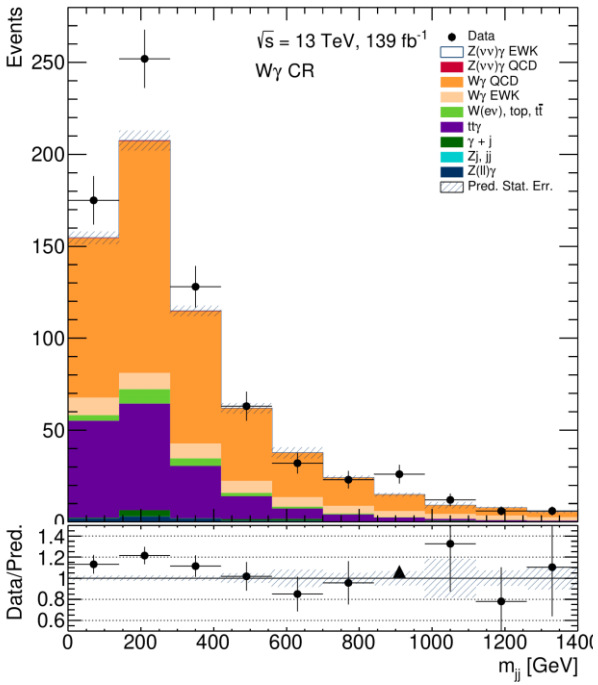

Events
√s = 13 TeV, 139 fb-1
Zγ QCD CR 2
Data
Z(νν)γ EWK
Z(νν)γ QCD
Wγ QCD
Wγ EWK
W(eν), top, tt̄
ttγ
γ + j
Zj, jj
Z(ll)γ
Pred. Stat. Err.
Data/Pred.
pT-balance(reduced)

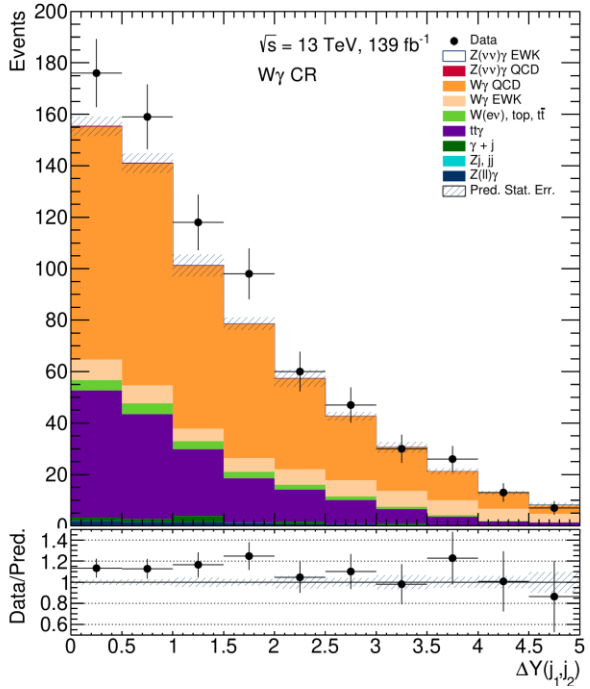

Events
√s = 13 TeV, 139 fb-1
Zγ QCD CR 2
Data
Z(νν)γ EWK
Z(νν)γ QCD
Wγ QCD
Wγ EWK
W(eν), top, tt̄
ttγ
γ + j
Zj, jj
Z(ll)γ
Pred. Stat. Err.
Data/Pred.
Njets

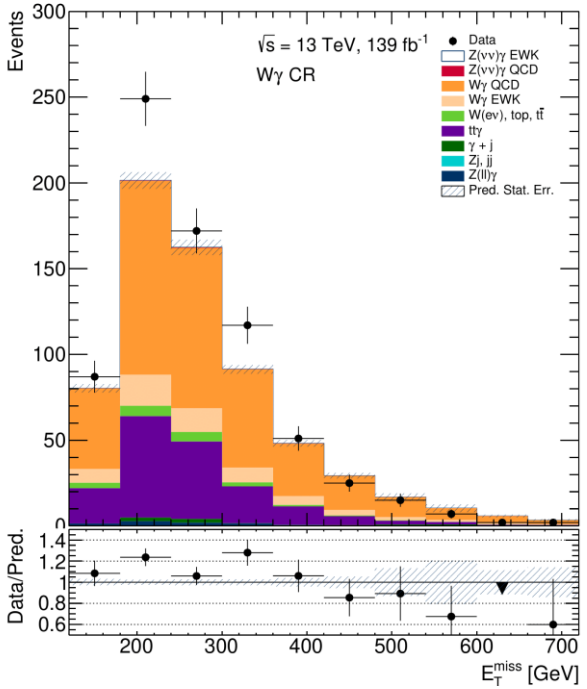

Events
√s = 13 TeV, 139 fb-1
Zγ QCD CR 2
Data
Z(νν)γ EWK
Z(νν)γ QCD
Wγ QCD
Wγ EWK
W(eν), top, tt̄
ttγ
γ + j
Zj, jj
Z(ll)γ
Pred. Stat. Err.
Data/Pred.
sin(|Δφ(j1,j2)|/2)

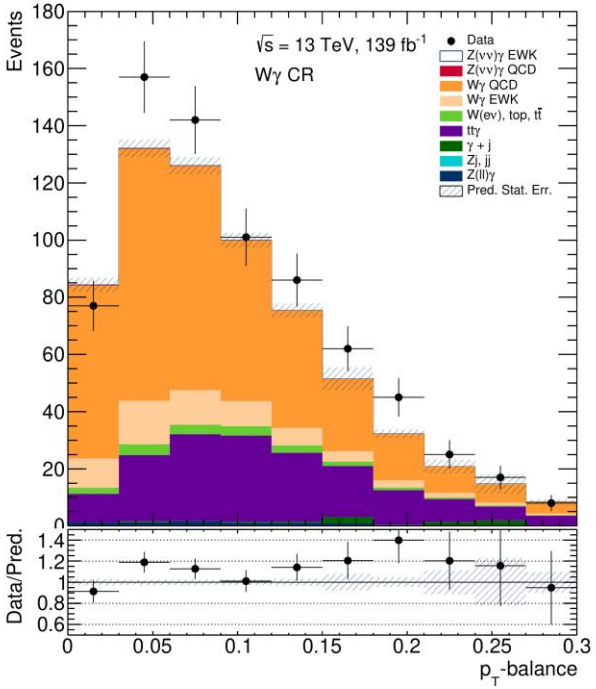

Events
√s = 13 TeV, 139 fb-1
Zγ QCD CR 2
Data
Z(νν)γ EWK
Z(νν)γ QCD
Wγ QCD
Wγ EWK
W(eν), top, tt̄
ttγ
γ + j
Zj, jj
Z(ll)γ
Pred. Stat. Err.
Data/Pred.
ΔY(j1,γ)

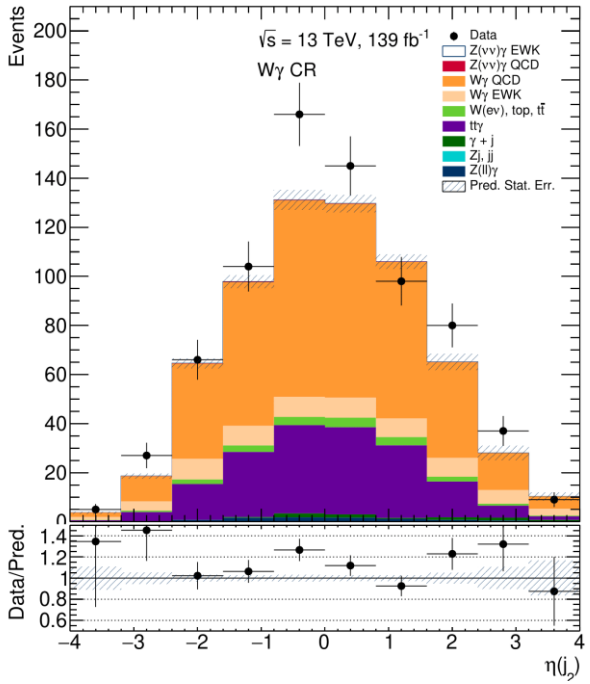
Wγ контрольный регион

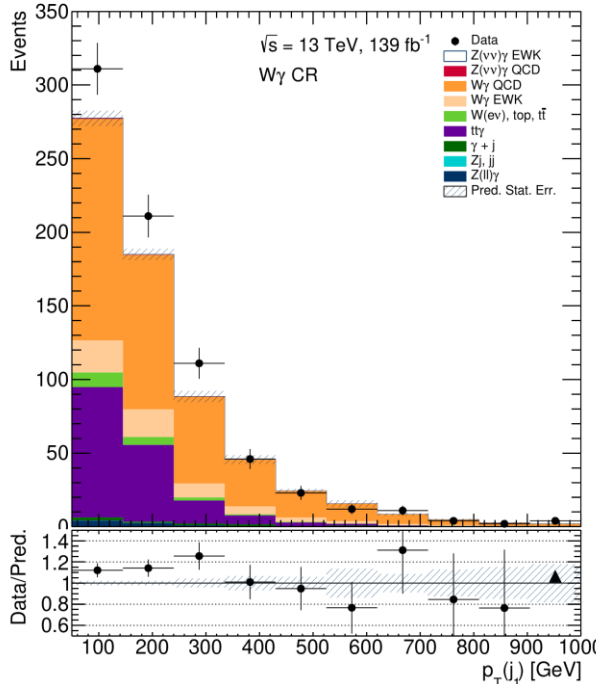


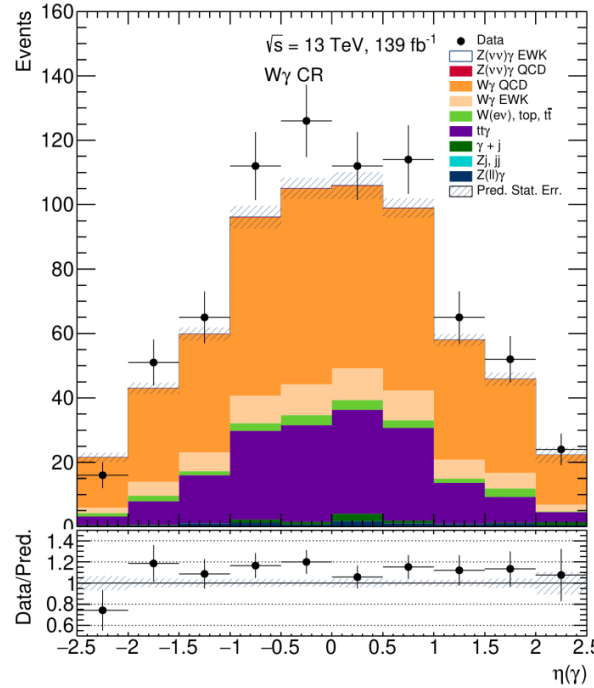


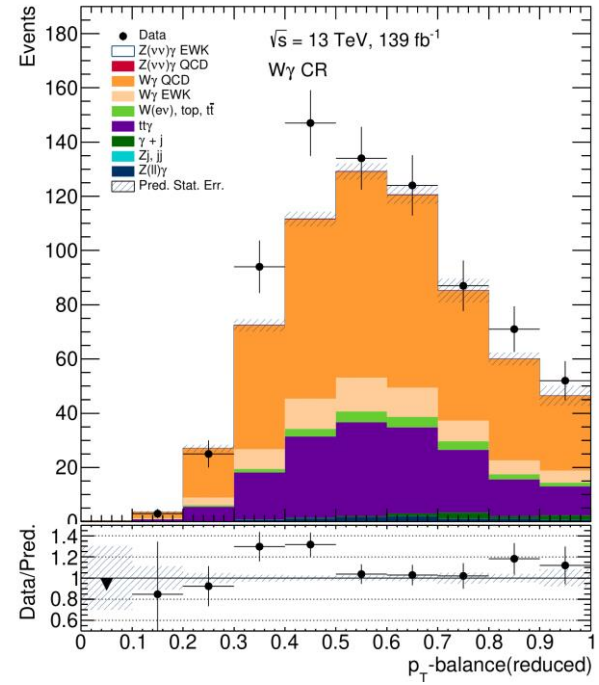


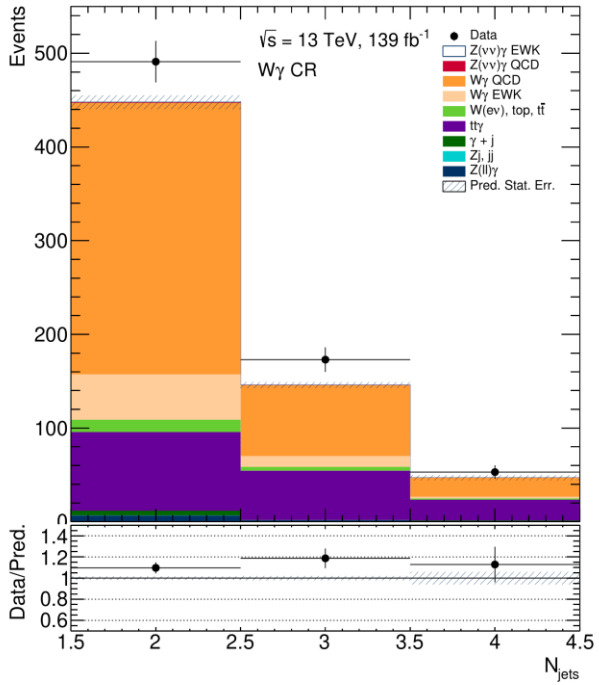


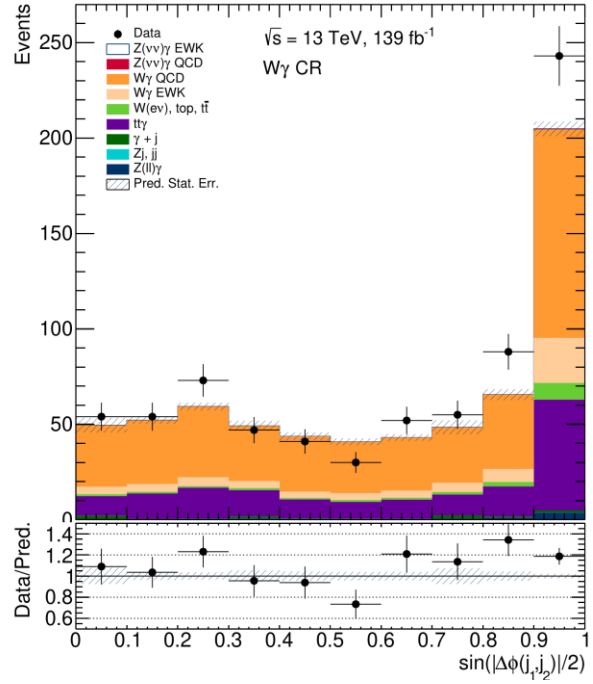

√s = 13 TeV, 139 fb-1
Wγ CR
Data
Z(νν)γ EWK
Z(νν)γ QCD
Wγ QCD
Wγ EWK
W(eν), top, tt̄
ttγ
γ + j
Zj, jj
Z(ll)γ
Pred. Stat. Err.
Events
Data/Pred.
η(j2)

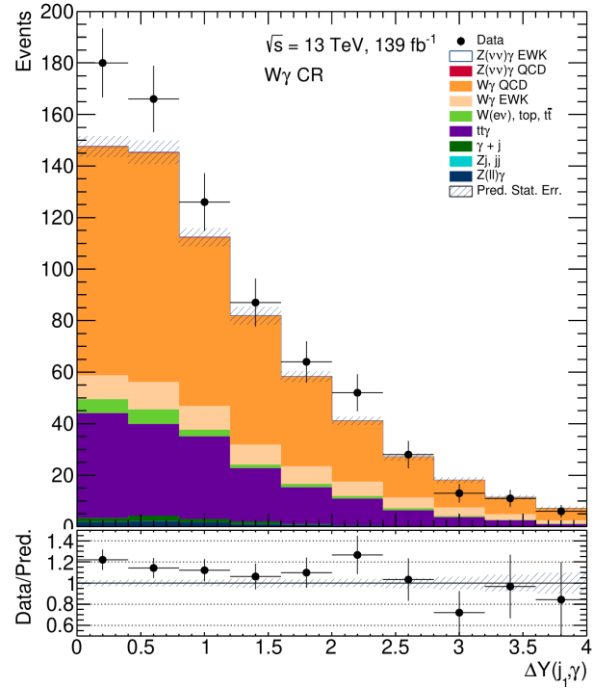

√s = 13 TeV, 139 fb-1
Wγ CR
Data
Z(νν)γ EWK
Z(νν)γ QCD
Wγ QCD
Wγ EWK
W(eν), top, tt̄
ttγ
γ + j
Zj, jj
Z(ll)γ
Pred. Stat. Err.
Events
Data/Pred.
pT(j1) [GeV]


√s = 13 TeV, 139 fb-1
Wγ CR
Data
Z(νν)γ EWK
Z(νν)γ QCD
Wγ QCD
Wγ EWK
W(eν), top, tt̄
ttγ
γ + j
Zj, jj
Z(ll)γ
Pred. Stat. Err.
Events
Data/Pred.
η(γ)


√s = 13 TeV, 139 fb-1
Wγ CR
Data
Z(νν)γ EWK
Z(νν)γ QCD
Wγ QCD
Wγ EWK
W(eν), top, tt̄
ttγ
γ + j
Zj, jj
Z(ll)γ
Pred. Stat. Err.
Events
Data/Pred.
pT-balance(reduced)


√s = 13 TeV, 139 fb-1
Wγ CR
Data
Z(νν)γ EWK
Z(νν)γ QCD
Wγ QCD
Wγ EWK
W(eν), top, tt̄
ttγ
γ + j
Zj, jj
Z(ll)γ
Pred. Stat. Err.
Events
Data/Pred.
Njets


√s = 13 TeV, 139 fb-1
Wγ CR
Data
Z(νν)γ EWK
Z(νν)γ QCD
Wγ QCD
Wγ EWK
W(eν), top, tt̄
ttγ
γ + j
Zj, jj
Z(ll)γ
Pred. Stat. Err.
Events
Data/Pred.
sin(|Δφ(j1,j2)|/2)


√s = 13 TeV, 139 fb-1
Wγ CR
Data
Z(νν)γ EWK
Z(νν)γ QCD
Wγ QCD
Wγ EWK
W(eν), top, tt̄
ttγ
γ + j
Zj, jj
Z(ll)γ
Pred. Stat. Err.
Events
Data/Pred.
ΔY(j1,γ)