

# Методы повышения чувствительности анализа и оценка фона $Z$ +струи в исследовании ассоциированного рождения $Z$ -бозона с фотоном в эксперименте ATLAS

Научный руководитель: Солдатов Е.Ю.

Консультант: Пятиизбянцева Д.Н.

Студент: Казакова К.К



Кафедра физики  
элементарных  
частиц

Москва 2022



# Описание процессов и мотивация

## Исследуемый процесс: ассоциированное рождение Z-бозона с фотоном

**Сигнальный процесс:**  $H \rightarrow Z(\ell\ell)\gamma$

- **Фоновые процессы:**  $Z+\text{jets}$  ( $\text{jet} \rightarrow \gamma$ ),  $Z\gamma$ ,  $Z\gamma jj$

**Мотивация:** 1) поиск бозона Хиггса

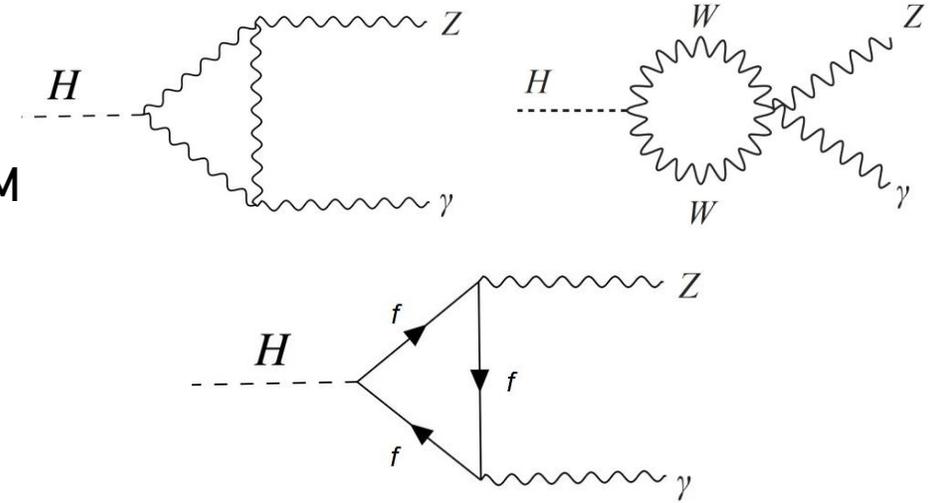
2) процесс высокочувствителен к отклонениям от СМ

**Цель анализа:** повышение чувствительности исследования

**Задачи анализа:** 1) Создание категоризации

2) Сравнение формы фона и данных в категориях

3) Фитирование смоделированного сигнала в категориях



**Сигнальный процесс:** ассоциированное рождение  $Z\gamma$

- **Фоновые процессы:**  $W(\rightarrow l\nu)\gamma$ ,  $t\bar{t}\gamma$ ,  $e\rightarrow\gamma$ ,  $\gamma+\text{jet}$ ,  $Z(\rightarrow\ell\ell)\gamma$ , **jet $\rightarrow\gamma$  (14%)**

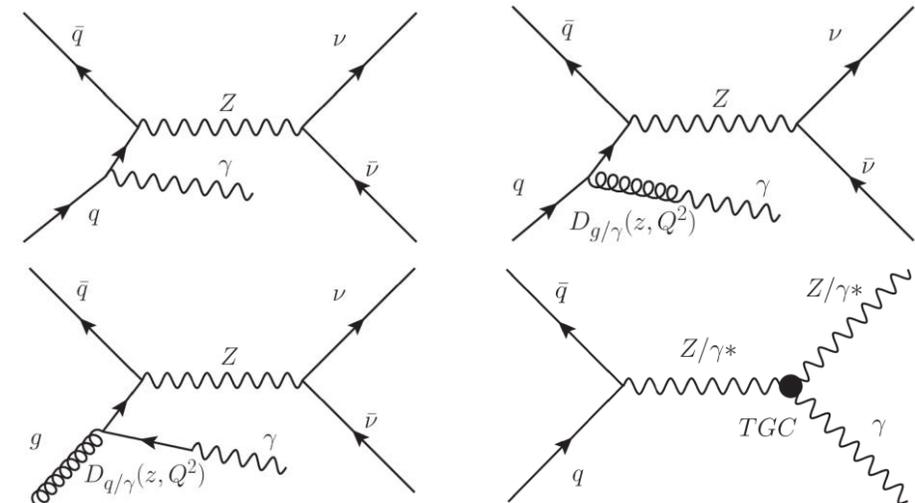
**Мотивация:** процесс высокочувствителен к отклонениям от СМ

**Цель анализа:** оценка числа фоновых событий  $\text{jet}\rightarrow\gamma$

**Задачи анализа:** 1) Оценка числа фоновых событий ABCD-методом

2) Создание альтернативного метода оценки

3) Оценка статистических и систематических погрешностей



# Используемые отборы

## Отбор фотонов

- В работе используются три фотонные изоляции: FixedCutTight, FixedCutTightCaloOnly и FixedCutLoose
  - Фотон идентифицируется как <<жесткий>> (tight), если он удовлетворяет всем критериям формы ЭМ ливня
- Если по крайней мере один из критериев формы ЭМ ливня нарушается, то фотон идентифицируется как <<мягкий>> (loose').

Изоляция	Калориметрическая изоляция	Трековая изоляция
FixedCutTightCaloOnly	$E_T^{\text{cone40}} - 0.022 \cdot p_T^\gamma < 2.45 \text{ ГэВ}$	-
FixedCutTight	$E_T^{\text{cone40}} - 0.022 \cdot p_T^\gamma < 2.45 \text{ ГэВ}$	$p_T^{\text{cone20}} / p_T^\gamma < 0.05$
FixedCutLoose	$E_T^{\text{cone20}} - 0.065 \cdot p_T^\gamma < 0 \text{ ГэВ}$	$p_T^{\text{cone20}} / p_T^\gamma < 0.05$

## Процесс: $H \rightarrow Z(\ell\ell)\gamma$

- Выбран канал распада Z-бозона в пару электронов или в пару мюонов
- $81.2 \text{ ГэВ} < m_{\ell\ell} < 101.2 \text{ ГэВ}$ ,  $p_T / m_{\ell\ell} > 0.12$
- Кандидаты в бозон Хиггса восстанавливаются по массе Z-бозона и фотону с наибольшей поперечной энергией:  $123.59 \text{ ГэВ} < m_{\ell\ell\gamma} < 126.59 \text{ ГэВ}$

Отбор	Электроны	Мюоны	Фотоны
$p_T^\gamma$	$> 10 \text{ ГэВ}$	$> 10 \text{ ГэВ}$	$> 10 \text{ ГэВ}$
$ \eta $	$ \eta  < 2.47$ кроме $1.37 <  \eta  < 1.52$	$ \eta  < 2.7$	$ \eta  < 2.37$ кроме $1.37 <  \eta  < 1.52$
Идентификация	Мягкая	Средняя	Жёсткая
Изоляция	FCLoose	FCLoose	FCLoose

## Процесс: ассоциированное рождение $Z\gamma$

- Выбран нейтринный канал распада Z-бозона ( $Z \rightarrow \nu\nu$ )

Переменная	Ограничение
$E_T^{\text{miss}}$	$> 130 \text{ ГэВ}$
$E_T^\gamma$	$> 150 \text{ ГэВ}$
Число жёстких фотонов	$N_\gamma = 1$
Число лептонов	$N_e = 0, N_\mu = 0$
Значимость $E_T^{\text{miss}}$	$> 11$
$ \Delta\phi(E_T^{\text{miss}}, \gamma) $	$> 0.6$
$ \Delta\phi(E_T^{\text{miss}}, j_1) $	$> 0.4$



отборы + изоляция – сигнальная область

# Категоризация для $H \rightarrow Z(\ell)\gamma$



Чувствительность:  $\xi = \frac{N_S}{\sqrt{N_S + N_B}}$

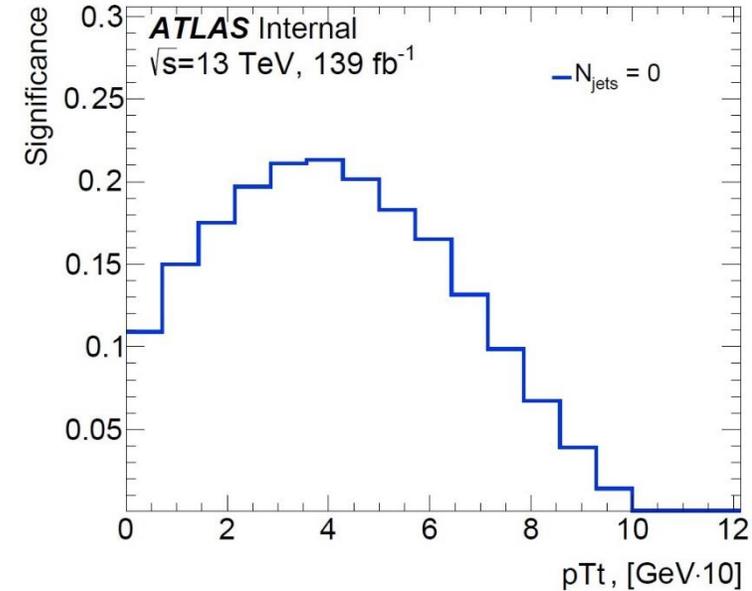
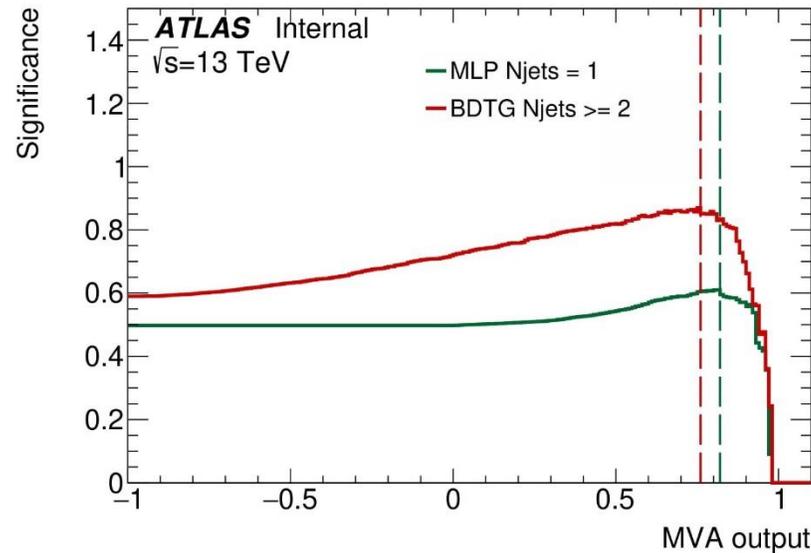
$N_S$  – число сигнальных событий (VBF, ggH, VH, ttH)

$N_B$  – число фоновых событий (Z+jets, Z $\gamma$ , Z $\gamma$ jj)

- Для категории  $N_{jets} = 0$  производилось разделение по переменной  $p_{Tt}$ . Максимум чувствительности достигается при **40 ГэВ**

- Для категории  $N_{jets} = 1$  применялся алгоритм машинного обучения **MLP**. Максимум чувствительности достигается при отклике **MLP = 0.82**

- Для категории  $N_{jets} \geq 2$  применялся классификатор **BDTG**. Максимум чувствительности достигается при отклике **BDTG = 0.76**



➡ Получено 6 категорий

Суммарная чувствительность для всех категорий:  $\xi = 1.29 \pm 0.02$

Чувствительность без применения категоризации:  $\xi = 0.840 \pm 0.003$

$$p_{Tt} = 2 \cdot \frac{|p_x^Z p_y^\gamma - p_x^\gamma p_y^Z|}{p_T^{Z-\gamma}}$$

# Сравнение и фитирование для $H \rightarrow Z(\ell)\gamma$

Моделирование фонового распределения  $\Rightarrow$   
 процесс его параметризации  $\Rightarrow$  в каждой  
 категории форма фона должна описываться  
 аналитическими функциями

Распределения в 4 других категориях в back-up

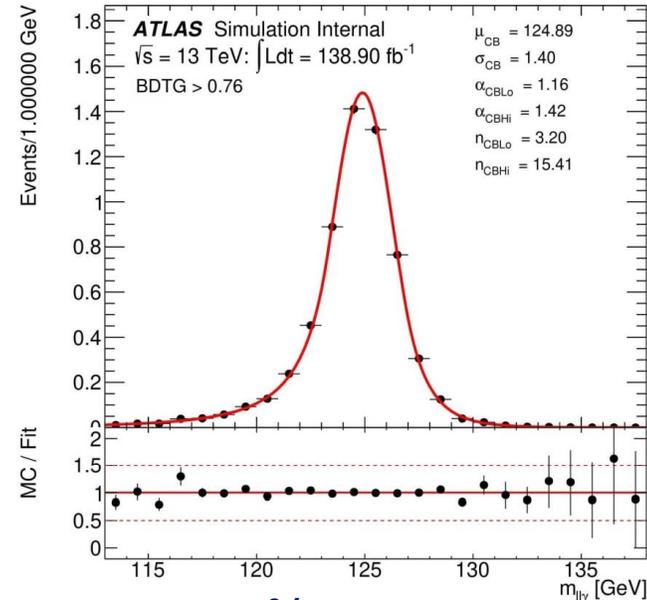
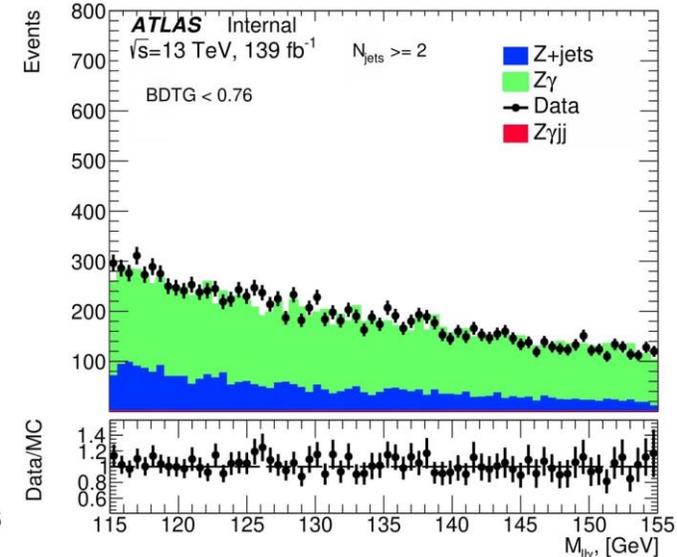
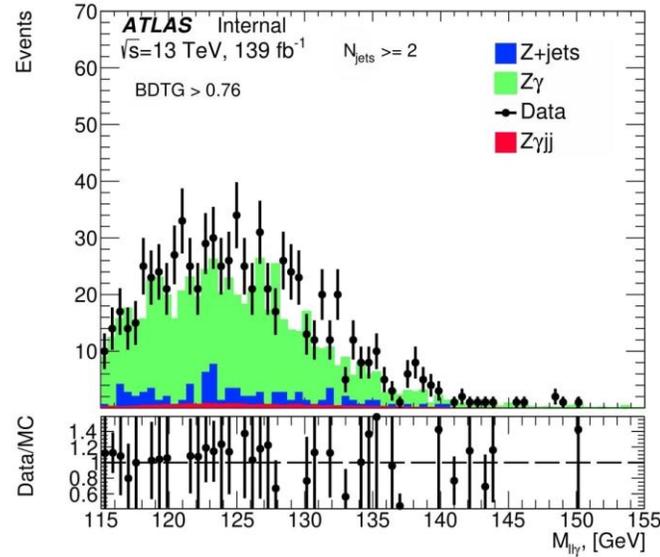
Моделирование сигнального распределения  $\Rightarrow$   
 процесс его параметризации

Функция фитирования для сигнала:

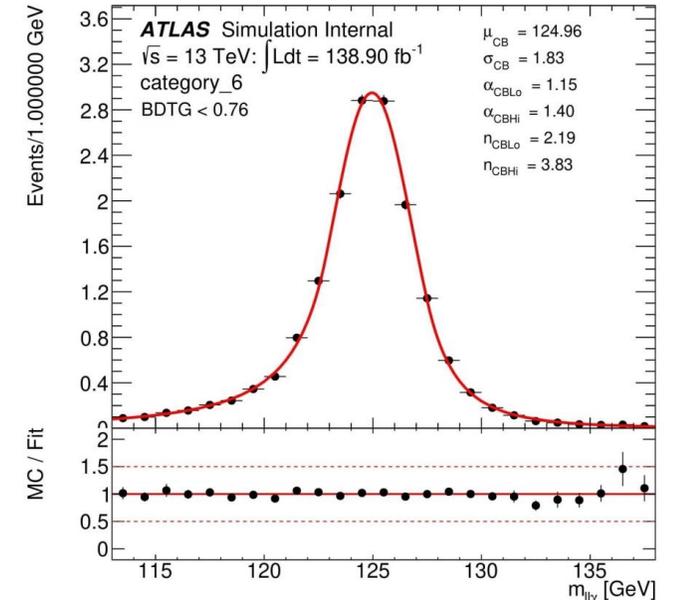
**Double-Sided Crystal Ball**

$$N \cdot \begin{cases} e^{t^2/2}, & -\alpha_{Lo} \leq t \leq \alpha_{Lo}, \\ \frac{e^{-0.5\alpha_{Lo}^2}}{\left[ \frac{\alpha_{Lo}}{n_{Lo}} \left( \frac{n_{Lo}}{\alpha_{Lo}} - \alpha_{Lo} - t \right) \right]^{n_{Lo}}}, & t < -\alpha_{Lo}, \\ \frac{e^{-0.5\alpha_{Hi}^2}}{\left[ \frac{\alpha_{Hi}}{n_{Hi}} \left( \frac{n_{Hi}}{\alpha_{Hi}} - \alpha_{Hi} + t \right) \right]^{n_{Hi}}}, & t > \alpha_{Hi}, \end{cases}$$

где  $t = \Delta m_H / \sigma_{CB}$ ,  $\Delta m_H = M_{Z\gamma} - \mu_{CB}$



$\chi^2/N_{\text{dof}} = 0.89$



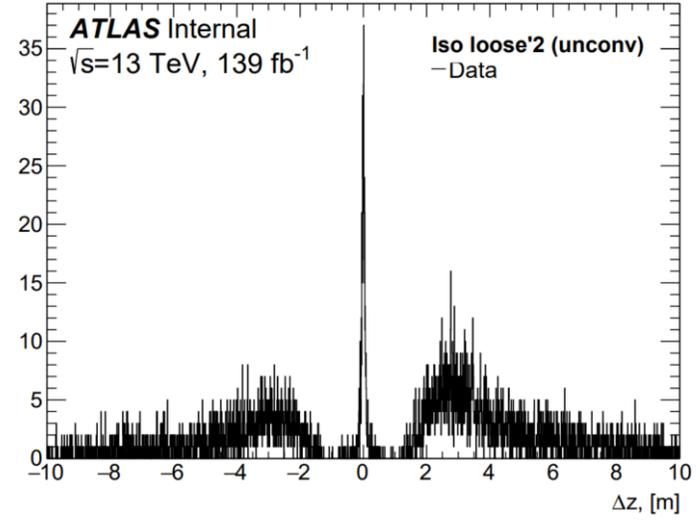
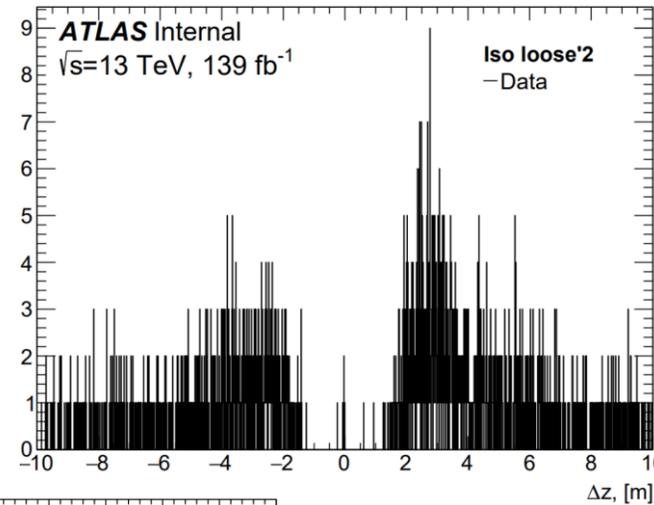
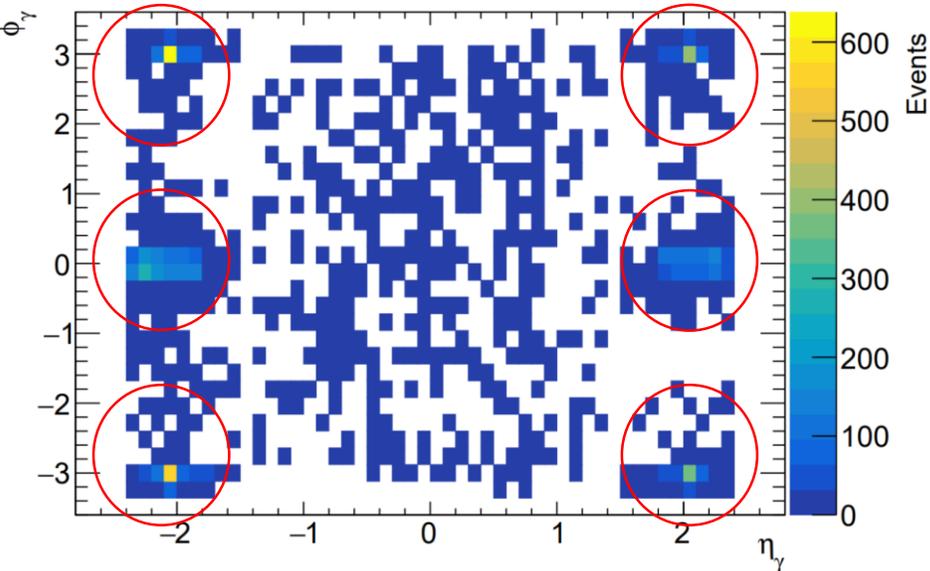
$\chi^2/N_{\text{dof}} = 1.28$

# Фон, обусловленный конфигурацией пучка

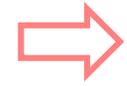
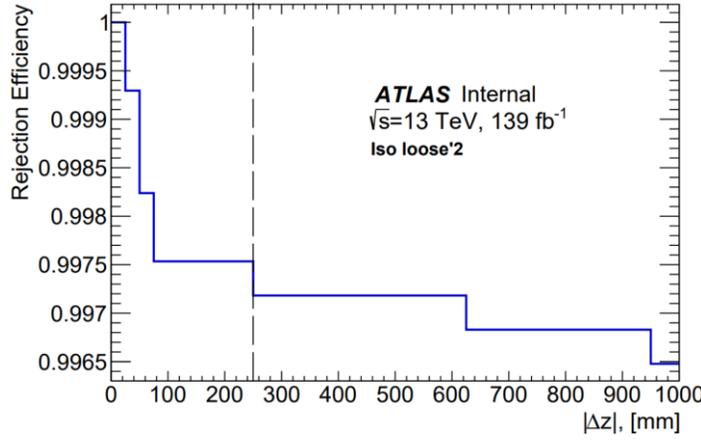
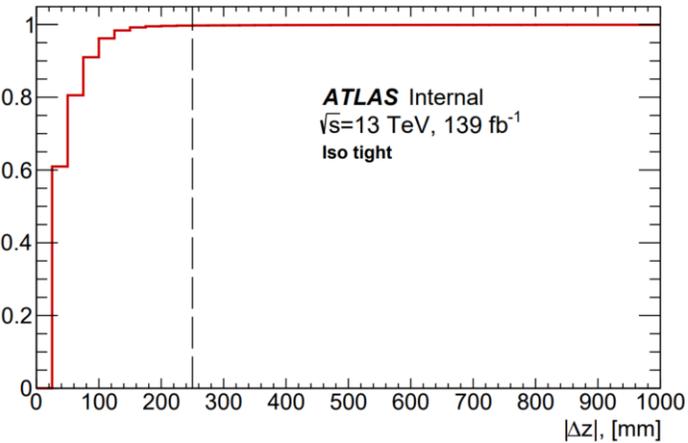
Протоны взаимодействуют с веществом в трубе  $\Rightarrow$  высокоэнергетичные мюоны создают большое энерговыделение посредством излучения  $\Rightarrow$  **ложные струи**

Преимущественно фон состоит из неконверсионных фотонов

С учётом отборов  $|\phi| < 0.2$ ,  
 $|\phi| \in [2.9, 3.2]$  и  $|\eta| > 1.7$



$$\Delta z = z_\gamma - z_{vtx}$$



- Наиболее оптимальному отбору соответствует  $|\Delta z| < 250 \text{ mm}$
- Значение для эффективности отбора в сигнальной области составляет  $\epsilon_{\text{accept}} = 99.7 \pm 0.9 \%$
- Значение для эффективности режекции  $\epsilon_{\text{rej}} = 100 \pm 3 \%$

# ABCD-метод оценки фона jet $\rightarrow \gamma$ (I)

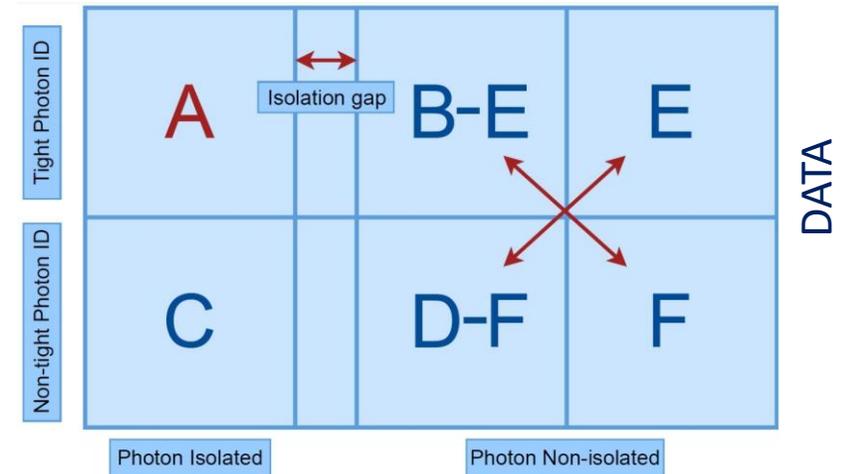
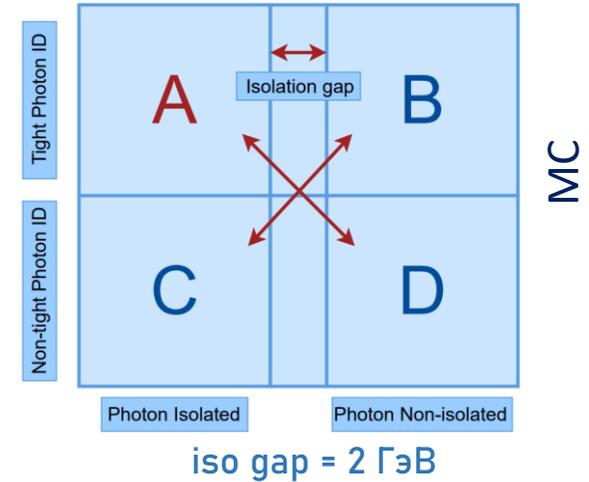
В качестве переменных используются идентификационные и изоляционные критерии для фотонов, которые не должны коррелировать

Определяется несколько типов мягких фотонов (*loose*'), для которых нарушаются различные критерии формы ЭМ ливня

$$R = \frac{N_A N_D}{N_C N_B}$$

## Результат оптимизации R фактора

R-фактор	loose'2	loose'3	loose'4	loose'5
FixedCutTight, без ограничения				
МК	$1.05 \pm 0.15$	$1.14 \pm 0.15$	$1.19 \pm 0.14$	$1.39 \pm 0.17$
Данные	$1.4 \pm 0.3$	$1.3 \pm 0.3$	$1.3 \pm 0.3$	$1.3 \pm 0.3$
FixedCutTight, ограничение = 25.45 ГэВ				
МК	$1.06 \pm 0.15$	$1.15 \pm 0.16$	$1.21 \pm 0.15$	$1.40 \pm 0.17$
Данные	$1.01 \pm 0.18$	$1.02 \pm 0.18$	$1.01 \pm 0.18$	$1.01 \pm 0.17$
FixedCutTight, трек. инверсия + ограничение = 25.45 ГэВ				
МК	$1.01 \pm 0.12$	$1.15 \pm 0.12$	$1.29 \pm 0.13$	$1.58 \pm 0.16$
Данные	$1.07 \pm 0.10$	$1.13 \pm 0.10$	$1.15 \pm 0.10$	$1.15 \pm 0.10$
FixedCutTightCaloOnly				
МК	$1.06 \pm 0.10$	$1.14 \pm 0.11$	$1.22 \pm 0.10$	$1.40 \pm 0.12$
Данные	$1.07 \pm 0.10$	$1.13 \pm 0.10$	$1.15 \pm 0.10$	$1.15 \pm 0.10$



$$R_{data} = \frac{N_{B-E}^{data} N_F^{data}}{N_{D-F}^{data} N_E^{data}}$$

➔ Выбрана наиболее оптимальная изоляция **FixedCutTightCaloOnly**

# ABCD-метод оценки фона jet $\rightarrow \gamma$ (II)

- Число событий в областях определяется как:

$$\begin{cases} N_A = N_A^{\text{sig}} + N_A^{\text{bkg}} + N_A^{\text{jet} \rightarrow \gamma}; \\ N_B = c_B N_A^{\text{sig}} + N_B^{\text{bkg}} + N_B^{\text{jet} \rightarrow \gamma}; \\ N_C = c_C N_A^{\text{sig}} + N_C^{\text{bkg}} + N_C^{\text{jet} \rightarrow \gamma}; \\ N_D = c_D N_A^{\text{sig}} + N_D^{\text{bkg}} + N_D^{\text{jet} \rightarrow \gamma}; \end{cases}$$

$$c_i = \frac{N_i^{\text{sig}}}{N_A^{\text{sig}}}$$

$$\tilde{N}_i = N_i - N_i^{\text{bkg}}$$

## Параметры утечки сигнала в КО

	$c_B$	$c_C$	$c_D$
Значение	$0.0713 \pm 0.0002$	$0.00879 \pm 0.00007$	$0.00070 \pm 0.00002$

С учётом R фактора на данных, получим уравнение:

$$N_A^{\text{sig}} = \tilde{N}_A - R(\tilde{N}_B - c_B N_A^{\text{sig}}) \frac{\tilde{N}_C - c_C N_A^{\text{sig}}}{\tilde{N}_D - c_D N_A^{\text{sig}}}$$

- Количества данных и фоновых событий в каждой области:

	Data	$W\gamma$ QCD	$W\gamma$ EWK	$W(e\nu), top, tt$	$tt\gamma$	$\gamma + \text{jet}$	$Z(ll)\gamma$
A	$24946 \pm 158$	$3655 \pm 22$	$145.9 \pm 0.7$	$3070 \pm 12$	$213 \pm 3$	$5016 \pm 52$	$270 \pm 4$
B	$5163 \pm 72$	$337 \pm 8$	$14.1 \pm 0.2$	$140.9 \pm 0.5$	$21.9 \pm 1.0$	$161 \pm 9$	$15.1 \pm 1.3$
C	$1586 \pm 40$	$32 \pm 2$	$1.42 \pm 0.07$	$41.92 \pm 0.14$	$2.2 \pm 0.3$	$36 \pm 4$	$2.4 \pm 0.4$
D	$2805 \pm 53$	$3.0 \pm 0.6$	$0.21 \pm 0.03$	$0 \pm 0$	$0.82 \pm 0.19$	$0.8 \pm 0.4$	$0.19 \pm 0.11$

Решение уравнения имеет вид:

$$N_A^{\text{sig}} = \frac{b - \sqrt{b^2 - 4ac}}{2a}, \text{ где}$$

$$\begin{cases} a = c_D - R c_B c_C; \\ b = \tilde{N}_D + c_D \tilde{N}_A - R(c_B \tilde{N}_C + c_C \tilde{N}_B); \\ c = \tilde{N}_D \tilde{N}_A - R \tilde{N}_C \tilde{N}_B. \end{cases}$$

Подставляя найденное решение в систему уравнений, получим **центральное значение событий в сигнальной области с учетом R фактора на данных:  $N_A^{\text{jet} \rightarrow \gamma} = 1960$**

# Оценка погрешностей

**Статистическая погрешность:** числа событий в каждой контрольной области были независимо проварьированы на  $\pm 1\sigma$  для данных и всех фонов. Полученные значения просуммированы в квадратурах.

**Итоговая статистическая погрешность:**  $\delta = 4\%$

## **Систематическая погрешность:**

- варьирование изоляционного промежутка и использование альтернативных loose' (24%)
- погрешность от различных МК генераторов и моделей партонных ливней (9%)
- погрешности на эффективность реконструкции фотона (1.4%)

$$\bullet \sigma_{\text{iso}}^{\text{CB}} = \delta_{\text{iso}}^{\text{eff}} \cdot (c_B + 1) / c_B$$

$$\bullet \sigma_{\text{ID}}^{\text{CC}} = \delta_{\text{ID}}^{\text{eff}} \cdot (c_C + 1) / c_C$$

$$\bullet \sigma_{\text{iso}}^{\text{CD}} = \delta_{\text{iso}}^{\text{eff}} \cdot (c_B + 1) / c_B$$

$$\bullet \sigma_{\text{ID}}^{\text{CD}} = \delta_{\text{ID}}^{\text{eff}} \cdot (c_C + 1) / c_C$$

$$\left\{ \begin{array}{l} \delta_{\text{iso}}^{\text{eff}} = 0.013 \\ \delta_{\text{ID}}^{\text{eff}} = 0.013 \end{array} \right\}$$



$$N_A^{\text{jet} \rightarrow \gamma} = 1960 \pm 80(\text{стат.}) \pm 510(\text{сист.})$$

МК Zj и Multijet **предсказывают  $1560 \pm 1240$  событий**

Центральное значение	1960 $\pm$ 80
loose'3	-334
loose'4	-397
loose'5	-472
Изоляционный зазор +0.15 ГэВ	+33
Изоляционный зазор -0.15 ГэВ	-22

**Итоговая систематическая погрешность:**  $\delta = 26\%$

# Метод максимального правдоподобия

- Альтернативный способ оценки фона

Функция правдоподобия: 
$$L(N_{ji} | f_{F_{ji}}, f_{N_j}) = \prod_{j=A}^{B,C,D} \prod_{i=1}^{N_{bins}} Pois(N_{ji} | \nu_{b_{ji}} + \nu_{\gamma_{ji}} f_{F_{ji}} + \nu_{s_{ji}} f_{N_j})$$

$f_{F_{ji}}$  - параметр, на который умножается оцениваемый фон в бинах в областях A, B, C, D

$f_{N_j}$  - параметр, на который умножается МК сигнал в областях A, B, C, D

$\nu_{b_{ji}}$ ,  $\nu_{s_{ji}}$  и  $\nu_{\gamma_{ji}}$  - количество событий в МК фонах, МК сигнале и оцениваемом фоне в бинах в областях A, B, C, D соответственно

$$\ln L = \sum_{j,i} Pois(N_{ji} | \nu_{b_{ji}} + \nu_{\gamma_{ji}} f_{F_{ji}} + \nu_{s_{ji}} f_{N_j})$$

$$\frac{\partial \ln L}{\partial f_{F_{ji}}} = 0, \quad \frac{\partial \ln L}{\partial f_{N_j}} = 0 \quad \Rightarrow \quad \text{поиск минимума с помощью пакета RooFit}$$

Преимущества модели:

- учёт биннига внутри регионов
- не требует оптимизации R фактора на МК и на данных

Условия для фита:

$$1 = \frac{\nu_{\gamma_{Ai}} f_{F_{Ai}} \cdot \nu_{\gamma_{Di}} f_{F_{Di}}}{\nu_{\gamma_{Bi}} f_{F_{Bi}} \cdot \nu_{\gamma_{Ci}} f_{F_{Ci}}}$$

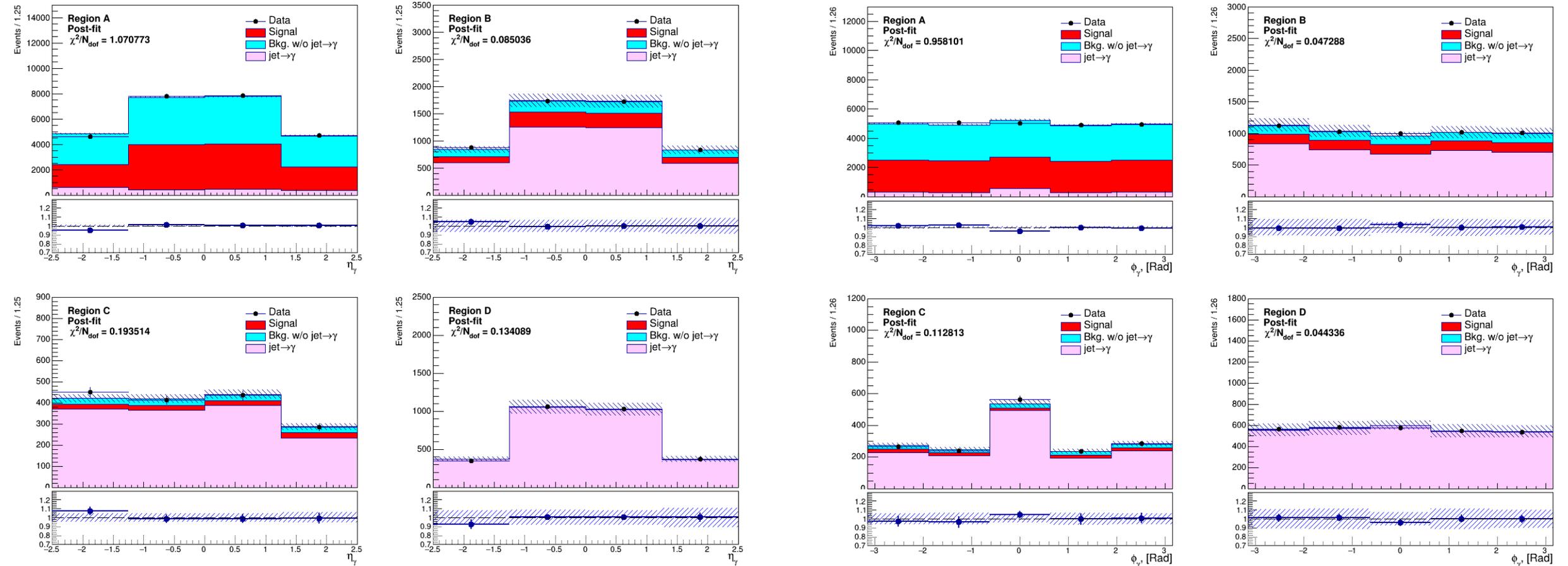
$$f_{F_{Bi}} = f_{F_{Di}}$$

Оценка:

$$N_{jet \rightarrow \gamma} = \nu_{\gamma_{Ai}} f_{F_{Ai}}$$

# Результаты фитирования

- Фитирование производилось для двух различных переменных с разным биннингом



- Биннинг выбирался на основе значения  $\chi^2/N_{\text{dof}}$
- Оценка центрального числа событий методом максимального правдоподобия в сигнальной области A для  $\eta_\gamma$  составила **1882**, для  $\phi_\gamma$  оценка составила **1743**

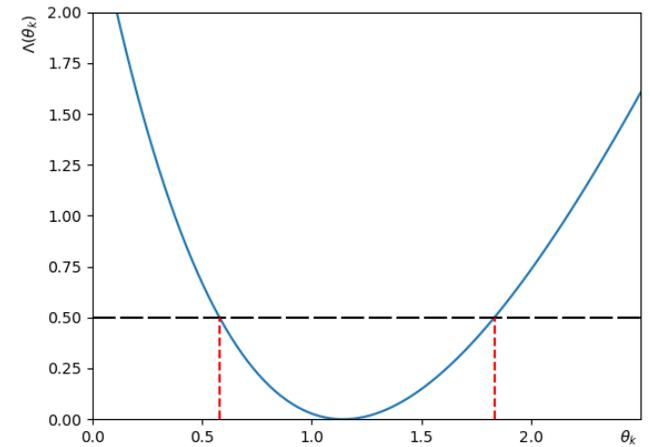
# Оценка погрешностей для ММП

## Статистическая погрешность:

- Отношение функций правдоподобий

$$\lambda(\theta_k) = \frac{L(\theta_k, \hat{\theta}_{l \neq k})}{L(\hat{\theta}_k, \hat{\theta}_{l \neq k})} \Rightarrow \Lambda = -\ln \lambda(\theta_k), \text{ где } \theta_k = f_{F_{ji}}, f_{N_j}$$

Значения, при которых  $\Lambda = 0.5$  соответствуют  $\pm\sigma$



визуализация функции  $\Lambda$

Статистические погрешности оценивались с помощью RooFit и составили:

для  $\eta_\gamma$   $1882^{+74}_{-70}$ , для  $\phi_\gamma$   $1743^{+69}_{-65}$

$\delta_{\text{stat}} = 4\%$

## Систематическая погрешность:

- Оценивалась аналогичным для стандартного ABCD-метода образом

$\eta_\gamma$

Центральное значение ( $\eta_\gamma$ )	$1882^{+74}_{-70}$
loose'3	-401
loose'4	-447
loose'5	-512
Изоляционный зазор +0.15 GeV	-5
Изоляционный зазор -0.15 GeV	+16
Различные генераторы	-150

$\delta_{\text{sist}} = 28\%$

Центральное значение ( $\phi_\gamma$ )	$1743^{+69}_{-65}$
loose'3	-353
loose'4	-406
loose'5	-467
Изоляционный зазор +0.15 GeV	-1
Изоляционный зазор -0.15 GeV	-4
Различные генераторы	-155

$\phi_\gamma$

$\delta_{\text{sist}} = 28\%$

Итоговая оценка методом ММП:

$$N_A^{\text{jet} \rightarrow \gamma} = 1882^{+74}_{-70} \text{ (стат.)} \pm 527 \text{ (сист.)}$$

$$N_A^{\text{jet} \rightarrow \gamma} = 1743^{+69}_{-65} \text{ (стат.)} \pm 488 \text{ (сист.)}$$

➤ Оценки для разных переменных **совпадают** в пределах погрешностей со значениями, полученными стандартным ABCD-методом, а также согласуются друг с другом

# Заключение

В соответствии с поставленными задачами в результате данной работы:

- **создана категоризация** для процесса  $H \rightarrow Z(\ell)\gamma$  и **повышена чувствительность анализа** со значения  $\xi = 0.840 \pm 0.003$  до значения  $\xi = 1.29 \pm 0.02$ . Произведено **сравнение формы фона** и данных, а также **фитирование смоделированного сигнала** в категориях
- **получен оптимальный отбор** по координатной переменной  $|\Delta z|$  для ассоциированного рождения  $Z\gamma$ , в результате фон, обусловленный конфигурацией пучка, **подавлен на  $99.7 \pm 0.9\%$**
- **оценено центральное значение фоновых событий  $jet \rightarrow \gamma$**  для  $Z\gamma$  в сигнальной области A методом ABCD, а также оценены статистические и систематические погрешности. В результате получено  $N_A^{jet \rightarrow \gamma} = 1960 \pm 80(\text{стат.}) \pm 510(\text{сист.})$ . Погрешность оценки фона, полученной на основе данных, значительно меньше погрешности оценки из МК
- **разработан способ** оценки  $jet \rightarrow \gamma$  **на основе метода максимального правдоподобия** и получена оценка числа фоновых событий в сигнальной области A для различных переменных, которая составляет  $N_A^{jet \rightarrow \gamma} = 1882_{-70}^{+74}(\text{стат.}) \pm 527(\text{сист.})$  для  $\eta_\gamma$  и  $N_A^{jet \rightarrow \gamma} = 1743_{-65}^{+69}(\text{стат.}) \pm 488(\text{сист.})$  для  $\phi_\gamma$ . Результаты совпадают в пределах погрешностей со значением, полученным ABCD-методом, а также согласуются друг с другом

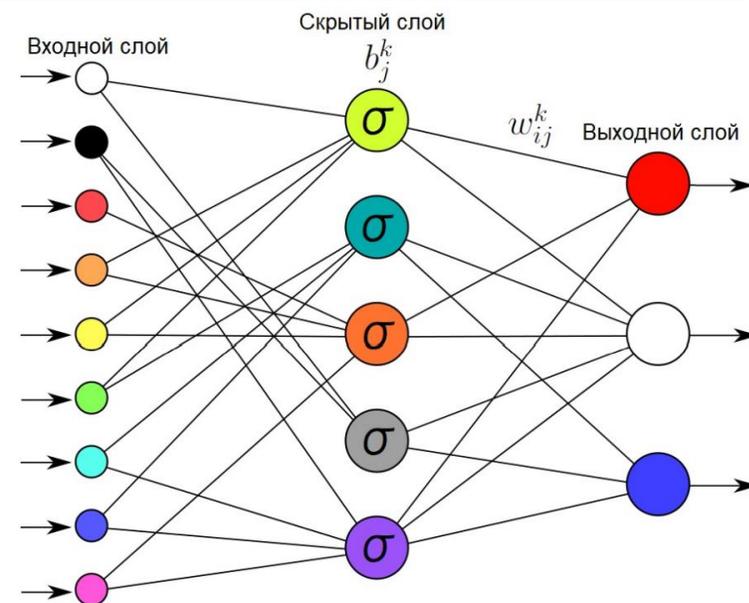
**BACK-UP**

# Машинное обучение и ММП

## Алгоритм MLP:

Функция активации:  $\sigma(x) = \frac{1}{1 + e^{-x}}$ ,  $\sigma(x) = \tanh(x)$ .

Нейроны в разных слоях соединены между собой взвешенными связями. При передаче значения от одного нейрона к другому оно домножается на соответствующий вес связи.



## Классификатор BDTG:

В каждом узле определяется такая переменная и отбор по ней, которые наилучшим образом разделяют события.

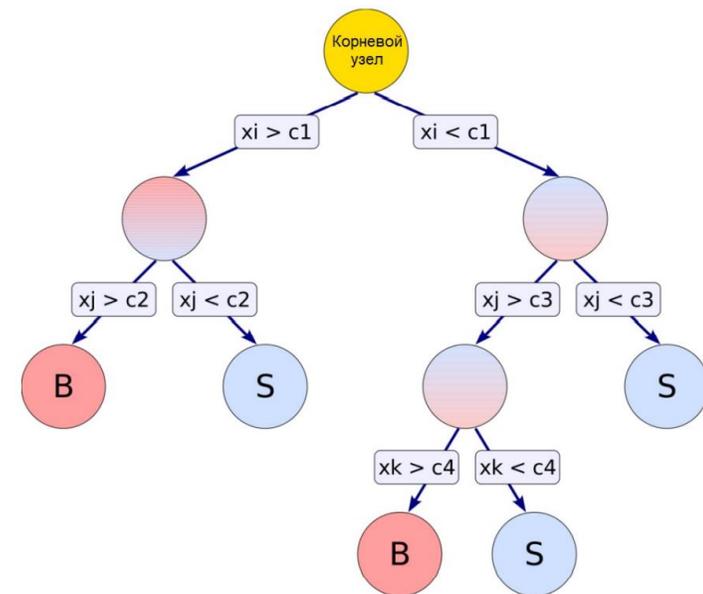
## Метод МП:

$$\mathcal{L}_n(X_1, X_2, \dots, X_n | \vec{\theta}) = \prod_{i=1}^n \mathcal{L}_1(X_i | \vec{\theta})$$

$$\text{Оценка: } \hat{\vec{\theta}} = \operatorname{argmax} \mathcal{L}(X_1, X_2, \dots, X_n | \vec{\theta})$$

Решение системы дает искомую оценку

$$\left\{ \begin{array}{l} \frac{\partial}{\partial \theta_1} \ln \mathcal{L}(X_1, X_2, \dots, X_n | \vec{\theta}) = 0, \\ \frac{\partial}{\partial \theta_2} \ln \mathcal{L}(X_1, X_2, \dots, X_n | \vec{\theta}) = 0, \\ \dots \dots \\ \frac{\partial}{\partial \theta_k} \ln \mathcal{L}(X_1, X_2, \dots, X_n | \vec{\theta}) = 0, \end{array} \right.$$



# Критерии рабочих точек и К0

$loose'2 : w_{s3}, F_{side}$

$loose'3 : w_{s3}, F_{side}, \Delta E$

$loose'4 : w_{s3}, F_{side}, \Delta E, E_{ratio}$

$loose'5 : w_{s3}, F_{side}, \Delta E, E_{ratio}, w_{tot}$

$$R_{data} = \frac{N_{B-E}^{data} N_F^{data}}{N_{D-F}^{data} N_E^{data}}$$

$$\Delta R = \sqrt{\Delta \eta^2 + \Delta \phi^2}$$

**A (CP):**  $E_T^{cone40} - 0.022 p_T^Y < 2.45$ , tight

**B (K0):**  $2.45 + isogap < E_T^{cone40} - 0.022 p_T^Y$ , tight

**C (K0):**  $E_T^{cone40} - 0.022 p_T^Y < 2.45$ , non-tight

**D (K0):**  $2.45 + isogap < E_T^{cone40} - 0.022 p_T^Y$ , non-tight

- $w_{s3}$  – ширина электромагнитного ливня с использованием трёх стриповых (первых слоёв ЭМ калориметра) слоёв вокруг стрипового слоя с максимальной энергией
- $F_{side}$  – доля энергии вне трёх стриповых слоёв, но внутри семи слоёв
- $\Delta E$  – разница энергий стриповых слоёв, где в одном слое выделилась вторая по величине энергия, и слоя, где выделилась наименьшая энергия
- $E_{ratio}$  – отношение разности энергий, ассоциированных с наиболее высоким и вторым по величине выделением энергии к сумме этих энергий
- $w_{tot}$  – полная поперечная ширина ливня

Параметры утечки	Различные МК генераторы и модели партонных ливней		
	MadGraph+Pythia8, Sherpa 2.2	MadGraph+Herwig7, MadGraph+Pythia8	$\delta$
$c_B$	$0.0713 \pm 0.0002$	$0.1000 \pm 0.0011$	29%
$c_C$	$0.00879 \pm 0.00007$	$0.0092 \pm 0.0003$	4%
$c_D$	$0.00070 \pm 0.00002$	$0.00099 \pm 0.00010$	29%
$jet \rightarrow \gamma$	1960	1785	9%

# Ассоциированное рождение $Z\gamma$

## Фоновые процессы:

- конечные состояния  $\tau\nu\gamma$  и  $l\nu\gamma$  от КХД и электрослабого рождения  $W\gamma$ , где  $\tau$  распадается на адроны, или где электрон или мюон от распада  $\tau$  или  $W$  не регистрируются детектором;
- события  $\gamma+$  струя, в которых большой  $E_T^{\text{miss}}$  возникает из комбинации реального  $E_T^{\text{miss}}$  от нейтрино в распадах тяжелых кварков и от неверно измеренной энергии струй;
- события  $W(e\nu)$ ,  $t$ -кварк и  $t\bar{t}$ , где электрон в конечном состоянии неверно идентифицируется как фотон ( $e \rightarrow \gamma$ );
- события от рождения  $t\bar{t}\gamma$ , когда один или оба  $W$  бозона от распада  $t$ -кварка распадаются на лептоны. Эти лептоны либо распадаются на  $\tau$ -лептоны, которые либо распадаются на адроны, либо не реконструируются;
- $Z(\nu\bar{\nu}) +$  струя и многоструйные события, где одна из струй неверно идентифицируется как фотон ( $jet \rightarrow \gamma$ );
- события  $Z(ll) + \gamma$  (преимущественно  $\tau$ -лептоны), где  $\tau$  распадается на адроны или когда электрон или мюон от распада  $\tau$  или  $Z$  не регистрируется.

**Потерянный поперечный импульс**  $\vec{p}_T^{\text{miss}}$

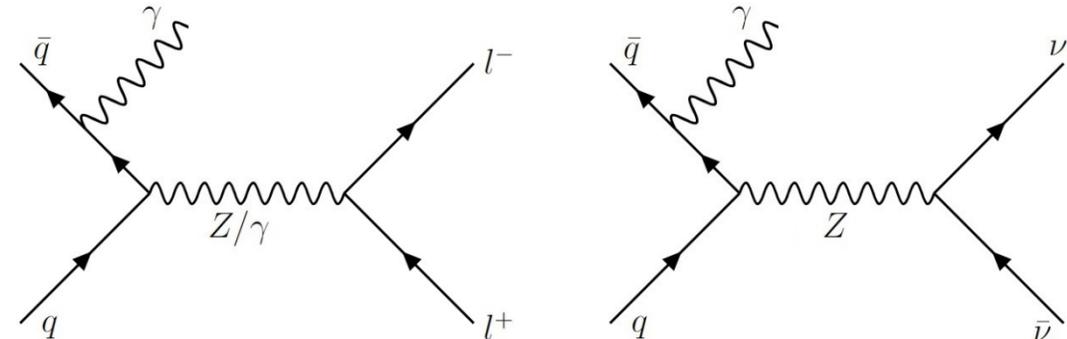
Определяется как сумма поперечных импульсов частиц в конечном

состоянии со знаком минус  $\vec{p}_T^{\text{miss}} = -\sum \vec{p}_T^f$

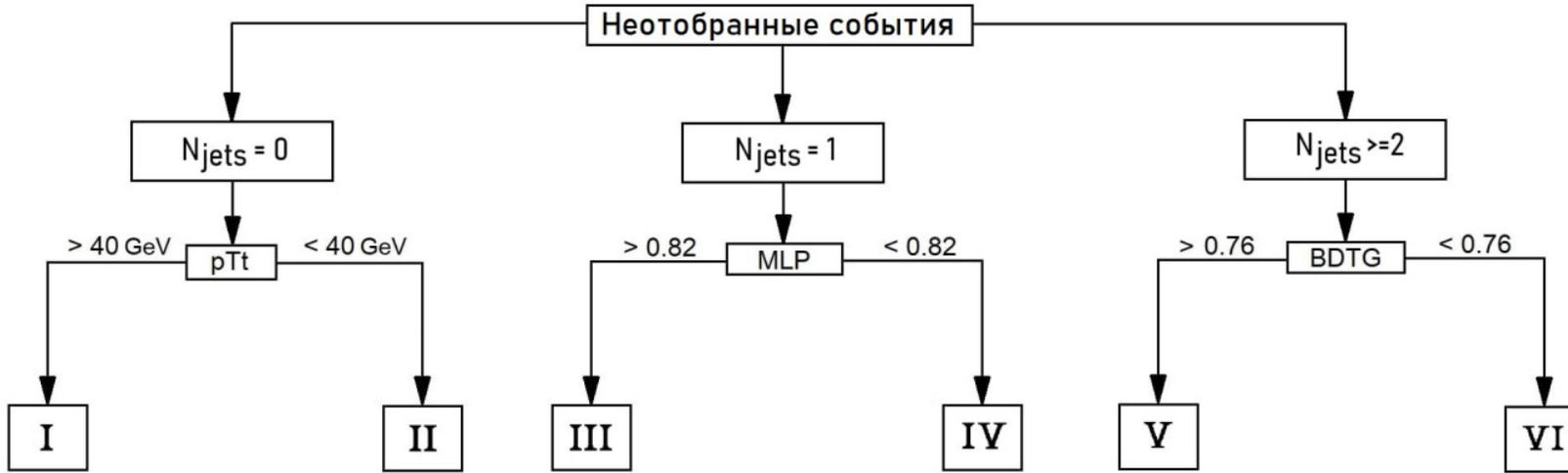
**Значимость**  $E_T^{\text{miss}} = E_T^{\text{miss}^2} / (\sigma_L^2 (1 - \rho_{LT}^2))$

где  $\sigma_L$  – дисперсия измерения потерянного поперечного импульса в продольном направлении

$\rho_{LT}$  – корреляционный фактор измерения продольной и поперечной компонент потерянного поперечного импульса



# Категоризация

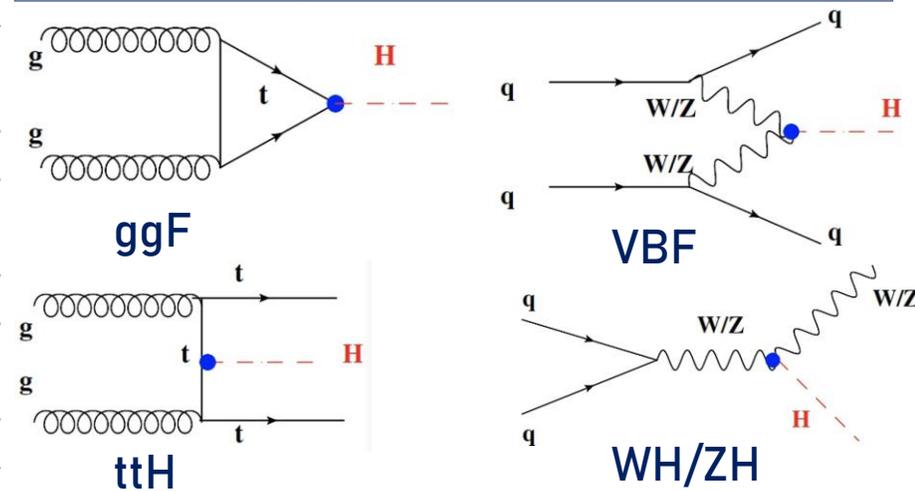


Формула Пуассона:

$$\xi_{\text{Poisson}} = \sqrt{2 \cdot \left[ (S + B) \times \ln \left( 1 + \frac{S}{B} \right) - S \right]}$$

№	$N_{jets}$	Категория	$N_S$	$N_B$	$S/\sqrt{(S+B)}$	$\xi_{\text{Poisson}}$
1	0	$p_{Tt}^{ll\gamma} > 40 \text{ ГэВ}$	$1.54 \pm 0.05$	$50 \pm 3$	$0.215 \pm 0.010$	$0.217 \pm 0.010$
2	0	$p_{Tt}^{ll\gamma} < 40 \text{ ГэВ}$	$34.25 \pm 0.18$	$6516 \pm 20$	$0.423 \pm 0.002$	$0.424 \pm 0.002$
$\Sigma$			$35.79 \pm 0.19$	$6566 \pm 20$	$0.475 \pm 0.010$	$0.476 \pm 0.010$
3	1	MLP $> 0.82$	$14.32 \pm 0.11$	$504 \pm 11$	$0.629 \pm 0.008$	$0.635 \pm 0.008$
4	1	MLP $< 0.82$	$14.46 \pm 0.16$	$2711 \pm 29$	$0.277 \pm 0.003$	$0.277 \pm 0.003$
$\Sigma$			$28.8 \pm 0.2$	$3215 \pm 31$	$0.687 \pm 0.009$	$0.693 \pm 0.009$
5	$\geq 2$	BDTG $> 0.76$	$9.30 \pm 0.07$	$105 \pm 3$	$0.870 \pm 0.012$	$0.895 \pm 0.014$
6	$\geq 2$	BDTG $< 0.76$	$17.51 \pm 0.07$	$1933 \pm 12$	$0.396 \pm 0.002$	$0.398 \pm 0.002$
$\Sigma$			$26.81 \pm 0.10$	$2038 \pm 13$	$0.956 \pm 0.012$	$0.980 \pm 0.014$
$\Sigma$			$91.4 \pm 0.3$	$11819 \pm 39$	$1.269 \pm 0.019$	$1.29 \pm 0.02$
$\Sigma$	-	-			$0.837 \pm 0.003$	$0.840 \pm 0.003$

## Способы образования бозона Хиггса



# Алгоритм MLP

## Настройки алгоритма:

функция активации: **tanh**

количество эпох обучения: **600**

структура скрытых слоёв: **N + 10, N, N**

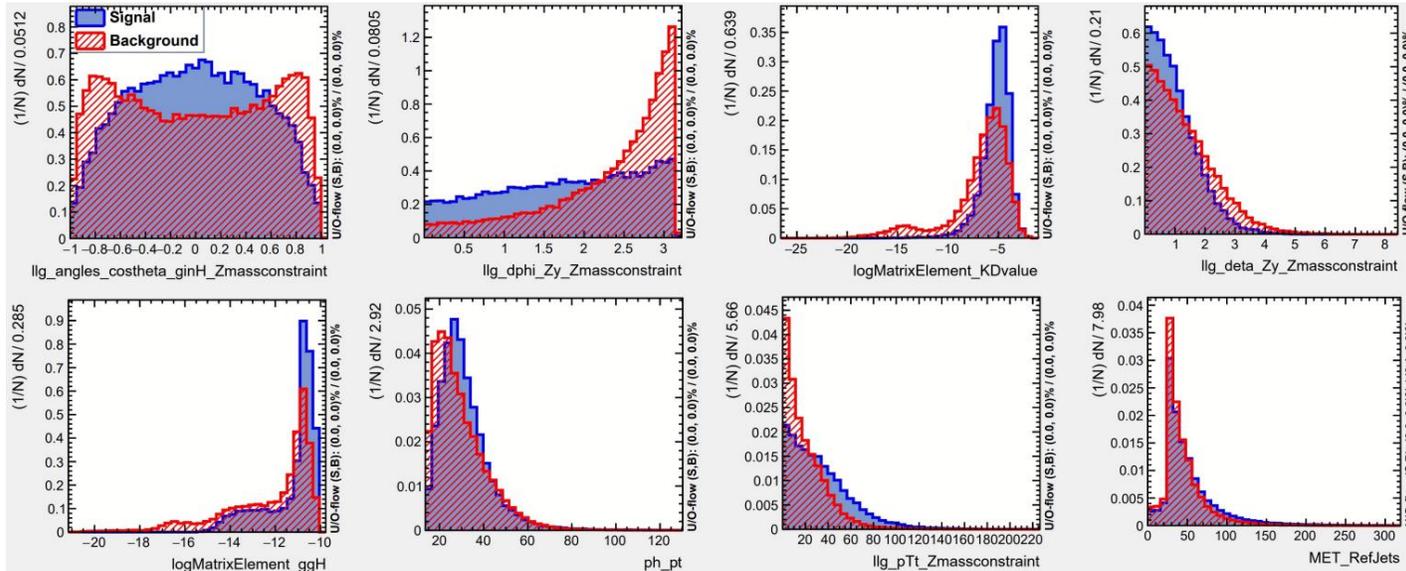
преобразования входных данных: **N**

количество итераций: **5**

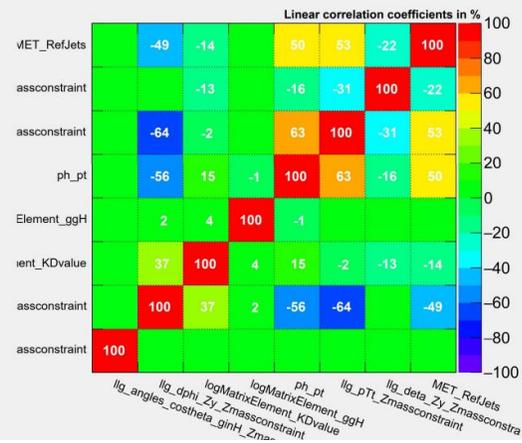
$$\langle S^2 \rangle = 1/2 \cdot \int \frac{(yS - yB)^2 * dy}{yS + yB}$$

$$I_i = \bar{x}_i^2 \sum_{j=1}^n (\omega_{ij}^{(1)})^2, \quad i = 1, \dots, n_{var}$$

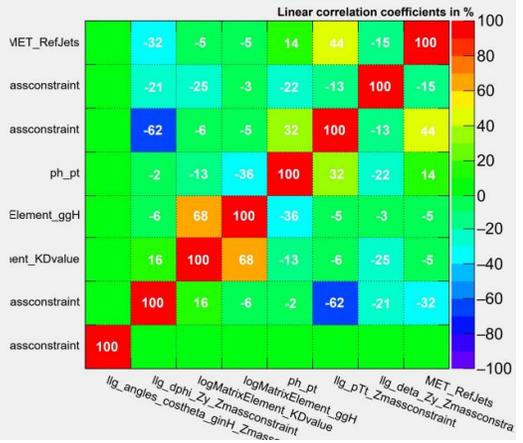
Переменная	Определение	$\langle S^2 \rangle$	$I_m$	Ранг
$p_{Tt}^{ll\gamma}$	$p_T$ частицы, ортогональный вектору разности между импульсами $Z$ -бозона и $\gamma$	0.084	1370	1
$p_T^\gamma$	Поперечный импульс фотона	0.030	1184	2
$\log ME_{kDVal}$	$2 \cdot [\log(ME_{ggH}) - \log(ME_{bkg})]$	0.119	475	3
$MET_{Jets}$	Потерянный поперечный импульс системы jj	0.017	147	4
$\Delta\eta^{Z,\gamma}$	Разница псевдобыстрот между $Z$ -бозоном и $\gamma$	0.033	44	5
$\cos^{H,Z\gamma} \theta$	Косинус угла между $H \rightarrow Z\gamma$ и $Z\gamma$	0.031	42	6
$\Delta\phi^{Z,\gamma}$	Азимутальный угол между $Z$ -бозоном и $\gamma$	0.011	36	7
$\log ME_{ggH}$	Логарифм матричного элемента ggH	0.118	17	8



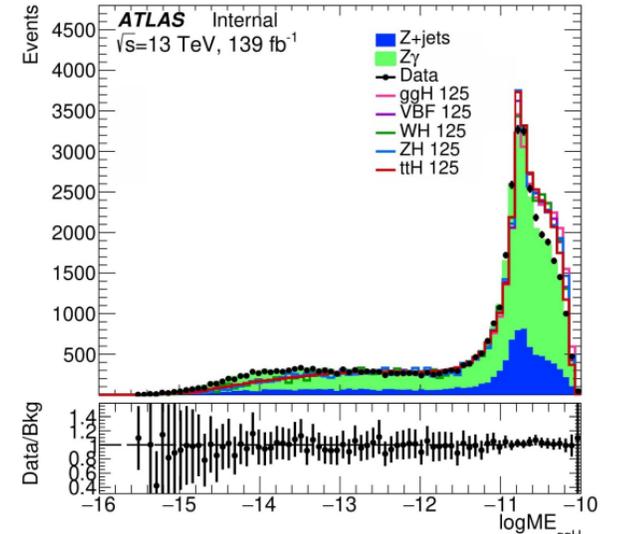
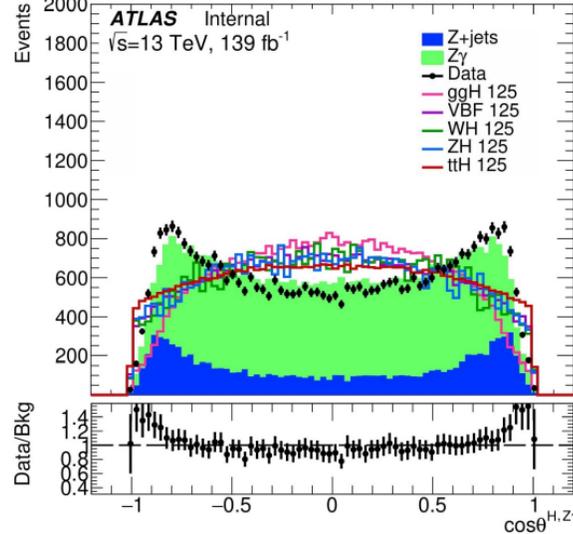
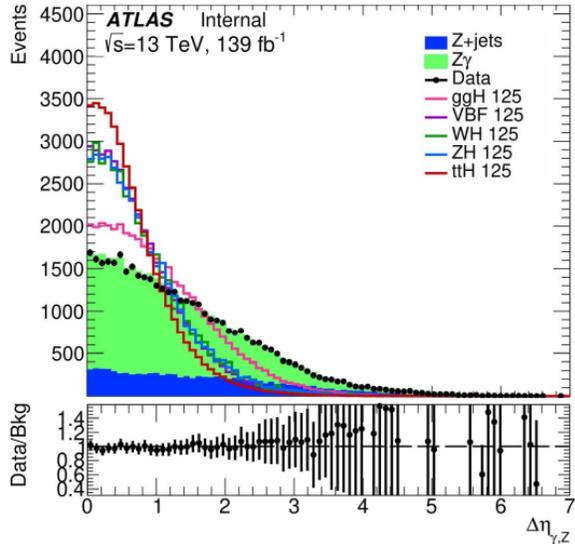
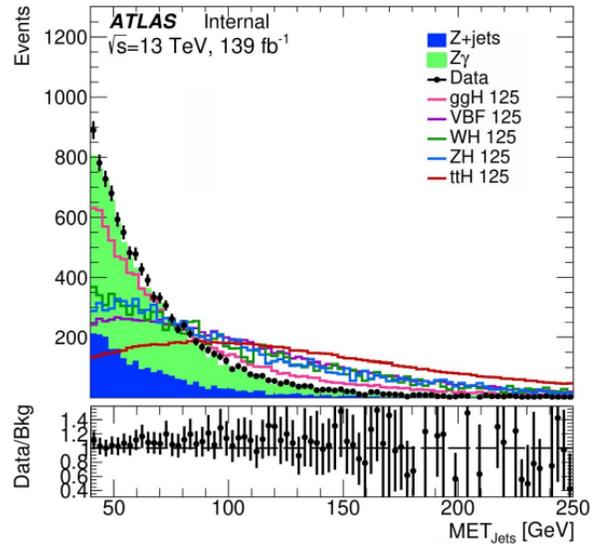
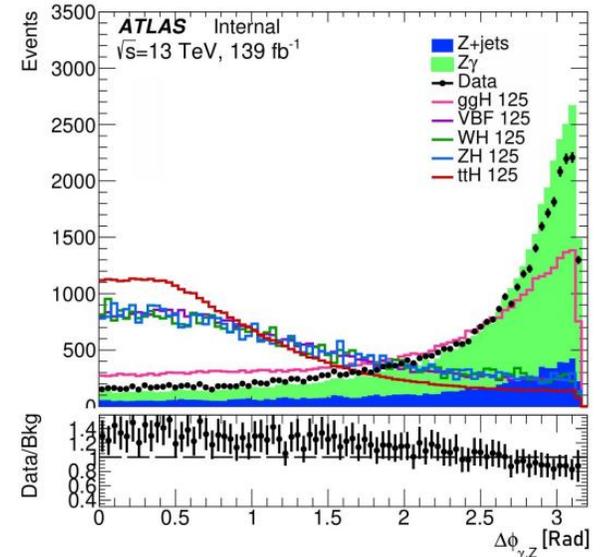
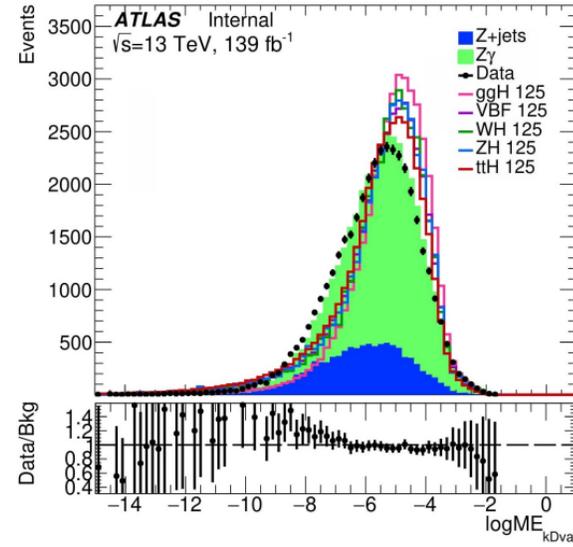
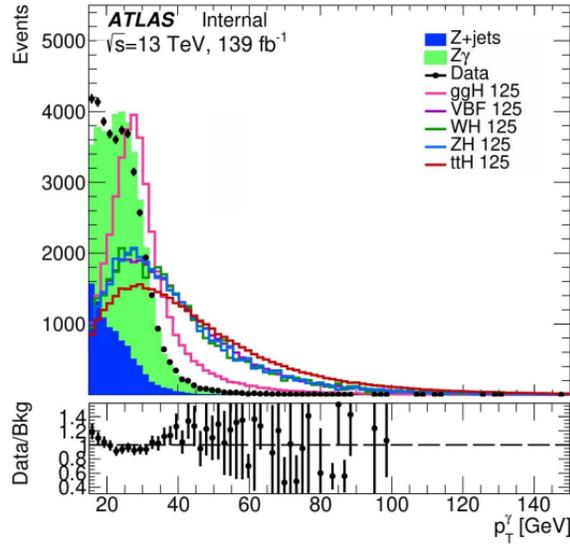
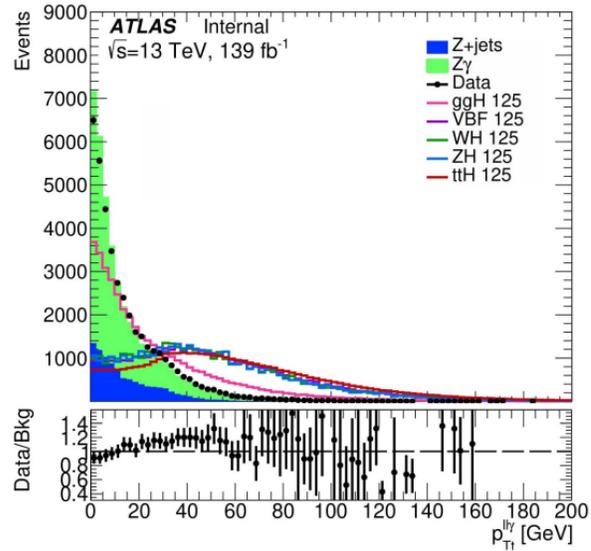
Correlation Matrix (signal)



Correlation Matrix (background)



# Алгоритм MLP



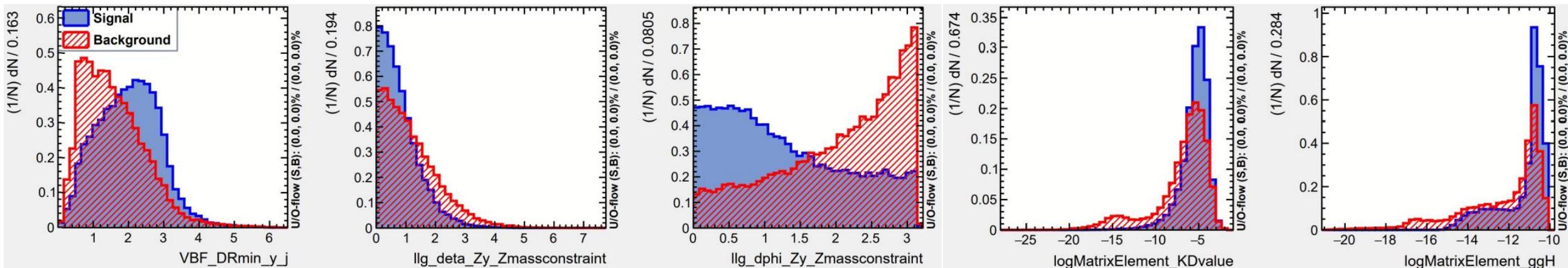
# Классификатор BDTG

## Настройки классификатора:

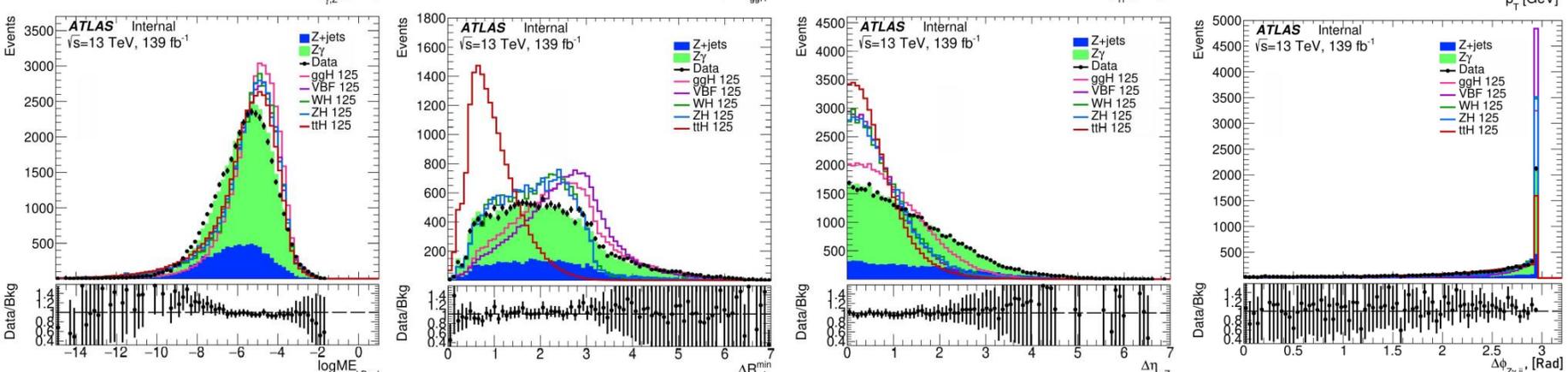
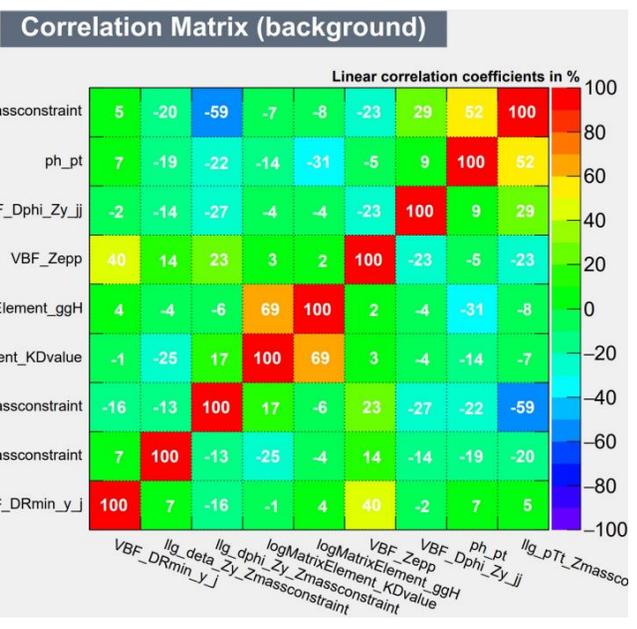
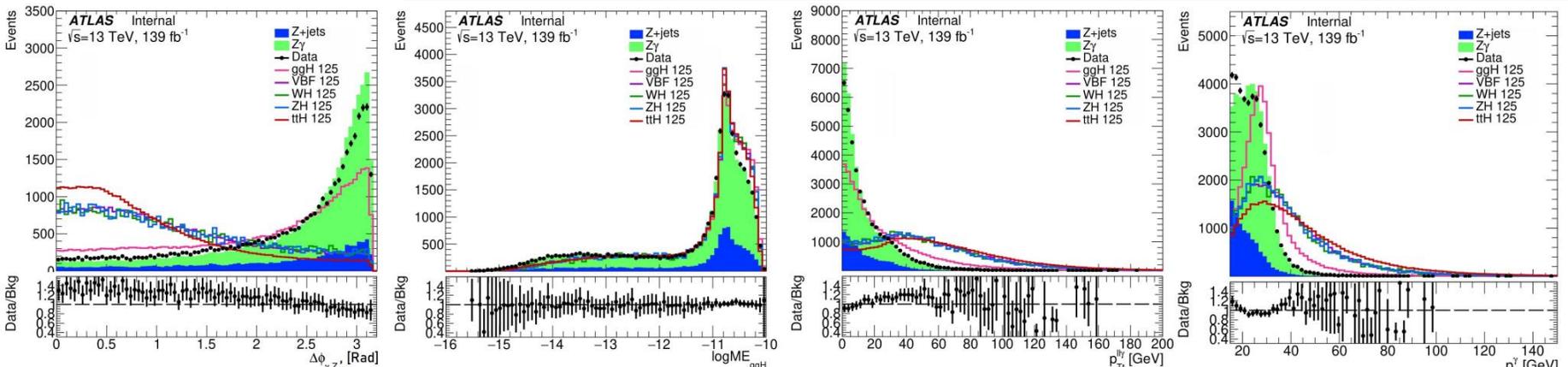
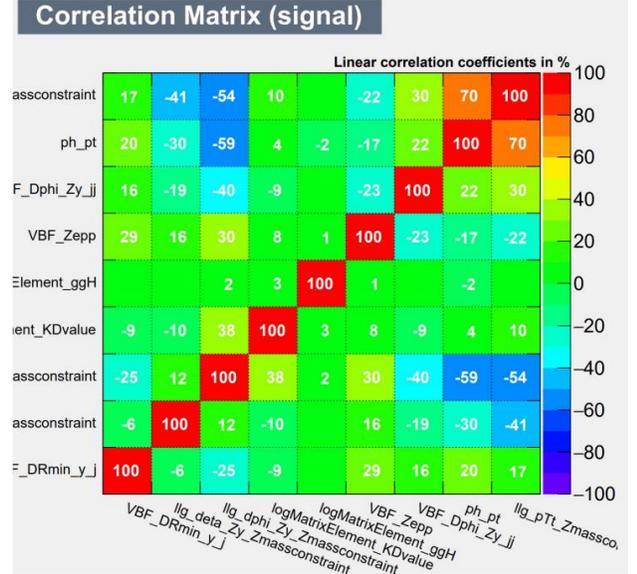
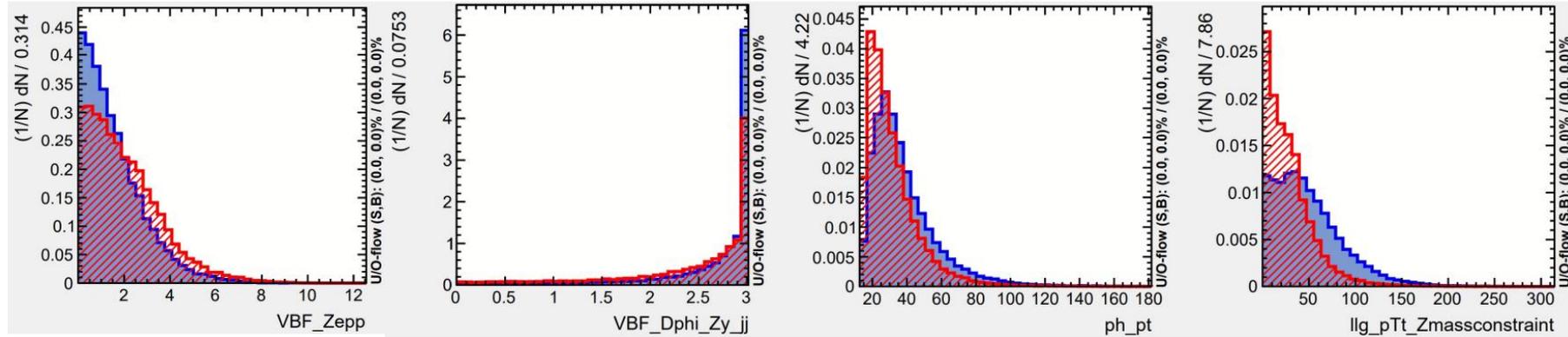
количество деревьев решений: **2000**  
 минимальный процент событий в узле: **2.5%**  
 тип бустинга BoostType: **Grad**  
 константа обучения для алгоритма  
 градиентного бустинга: **0.1**  
 количество отрезков, на которые  
 разбивается диапазон значений  
 переменной для определения оптимального  
 отбора nCuts: **20**  
 максимальная глубина дерева MaxDepth: **3**

Переменная	Определение	$\langle S^2 \rangle$	$I_m$	Ранг
$\Delta\phi^{Z,\gamma}$	Азимутальный угол между $Z$ -бозоном и $\gamma$	0.157	0.140	1
$\Delta\phi_{Z\gamma,jj}$	Азимутальный угол между $Z\gamma$ и системой $jj$	0.030	0.036	2
$\Delta\eta^{Z,\gamma}$	Разница псевдобыстрот между $Z$ -бозоном и $\gamma$	0.052	0.111	3
$\log ME_{kDVal}$	$2 \cdot [\log(ME_{ggH}) - \log(ME_{bkg})]$	0.092	0.110	4
$\Delta R_{\gamma,j}^{\min}$	Минимальное расстояние между $Z\gamma$ и системой $jj$	0.072	0.110	5
$\log ME_{ggH}$	Логарифм матричного элемента $ggH$	0.124	0.110	6
$\eta^{Z\text{pp}}$	$ \eta_{Z\gamma} - 0.5 \cdot (\eta_{j1} + \eta_{j2}) $	0.026	0.100	7
$p_{Tt}^{ll\gamma}$	$p_T$ частицы, ортогональный вектору разности между импульсами $Z$ -бозона и фотона	0.110	0.096	8
$p_T^\gamma$	Поперечный импульс фотона	0.048	0.095	9

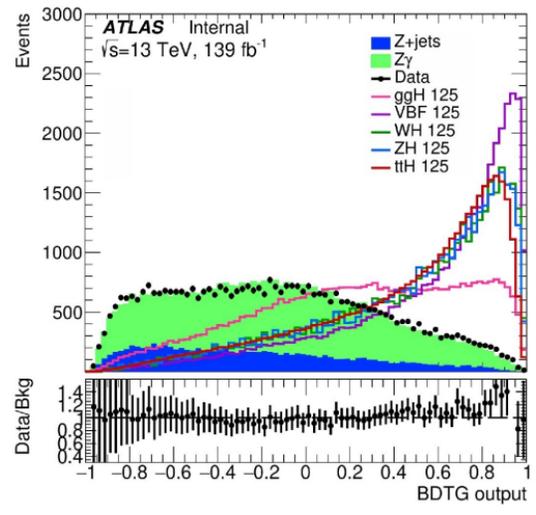
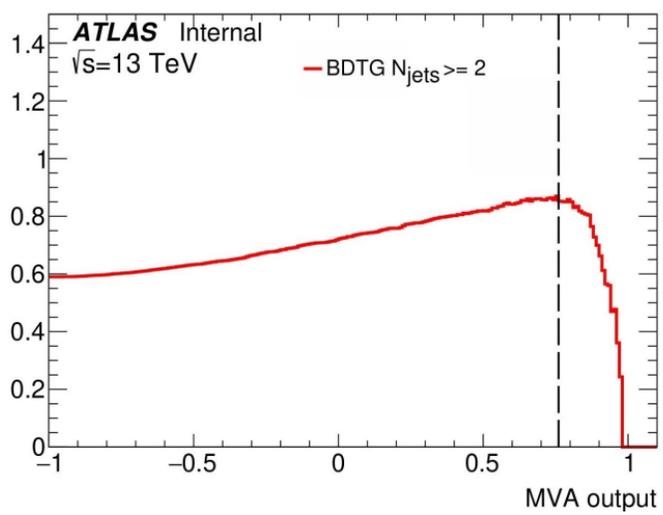
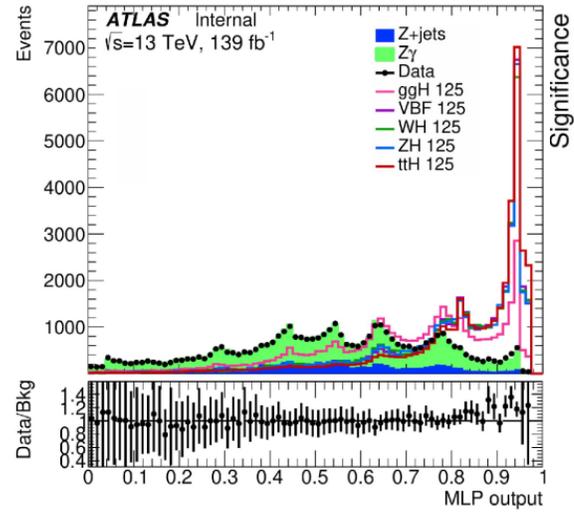
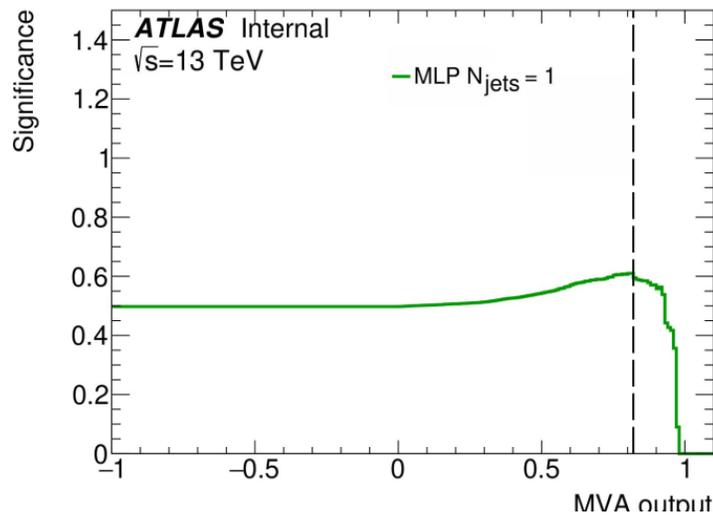
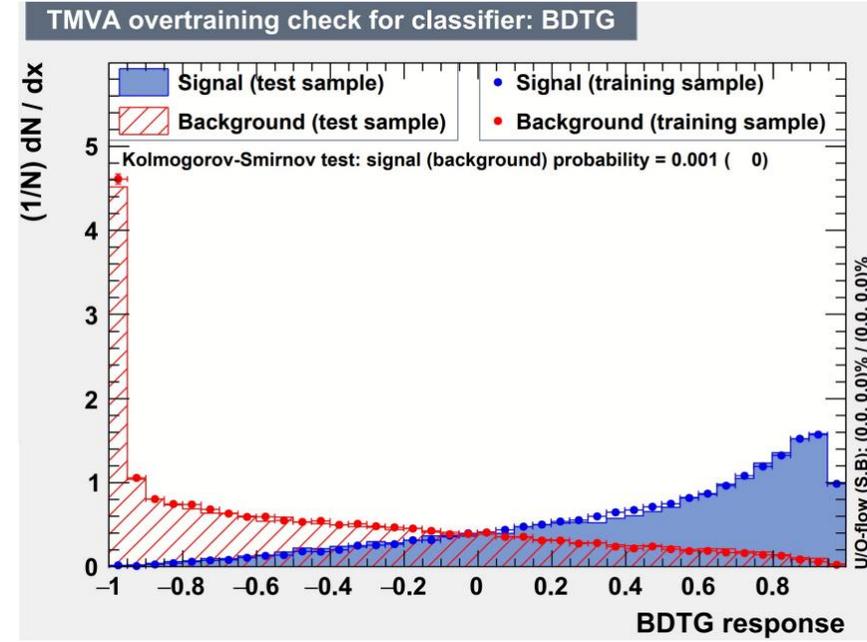
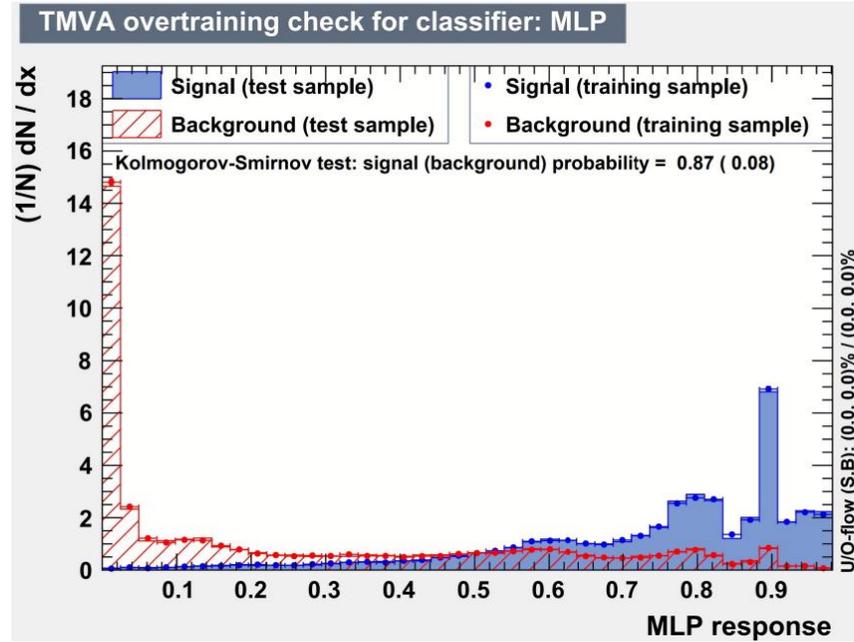
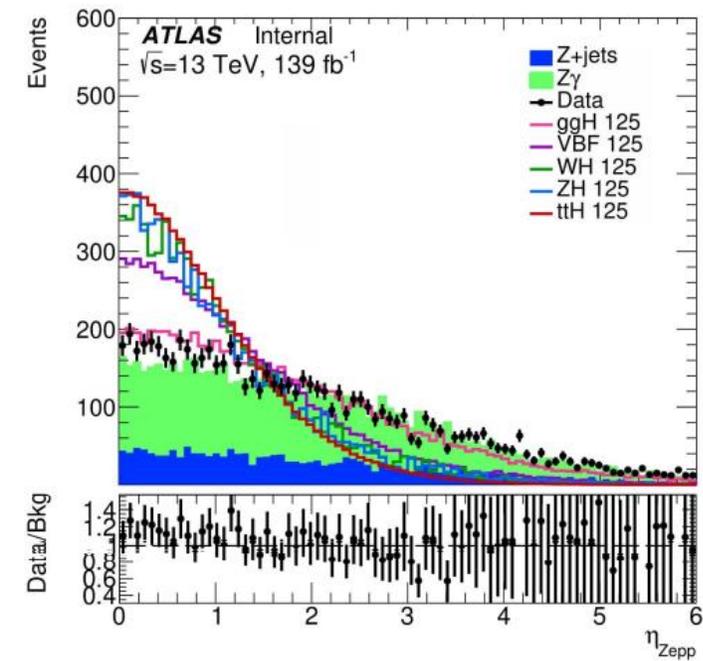
$$I_m = |a_m| \sqrt{(1.0 - s_m)}$$



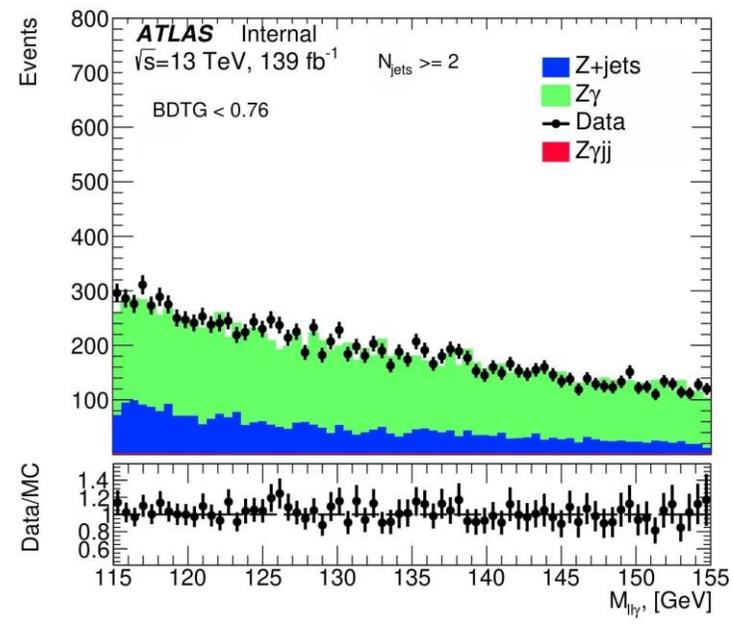
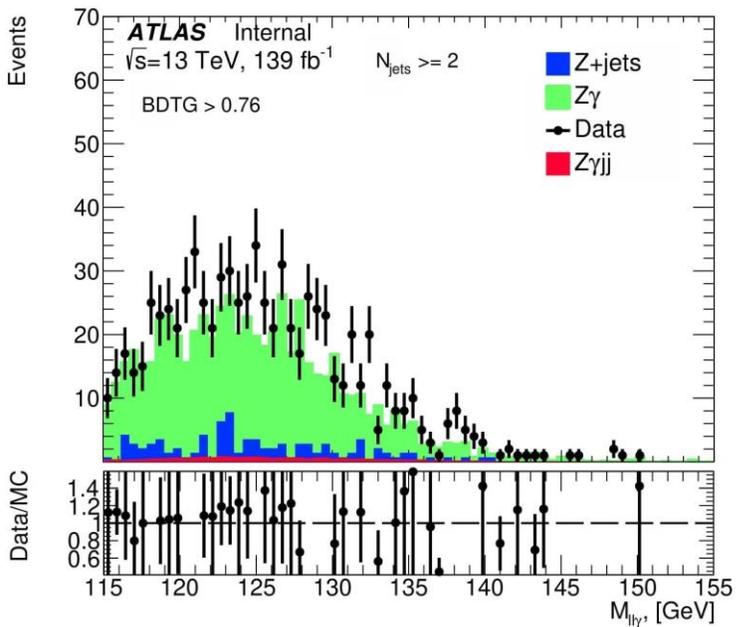
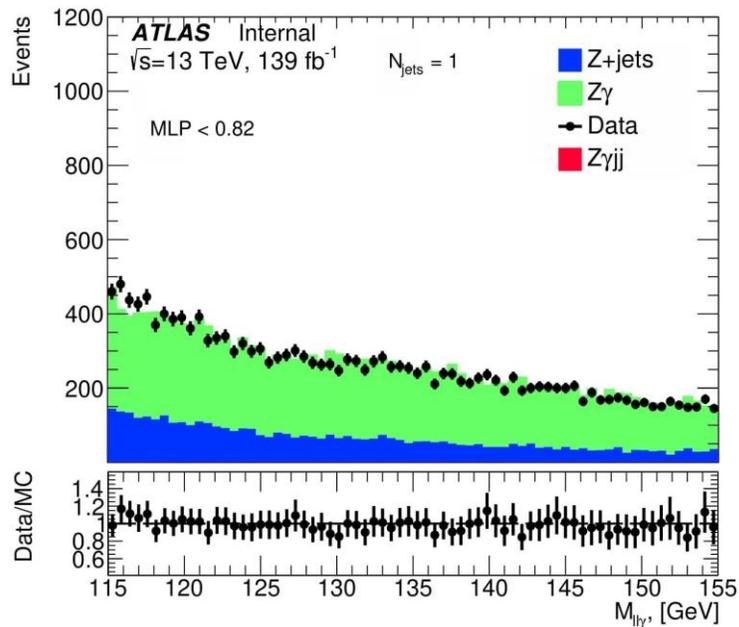
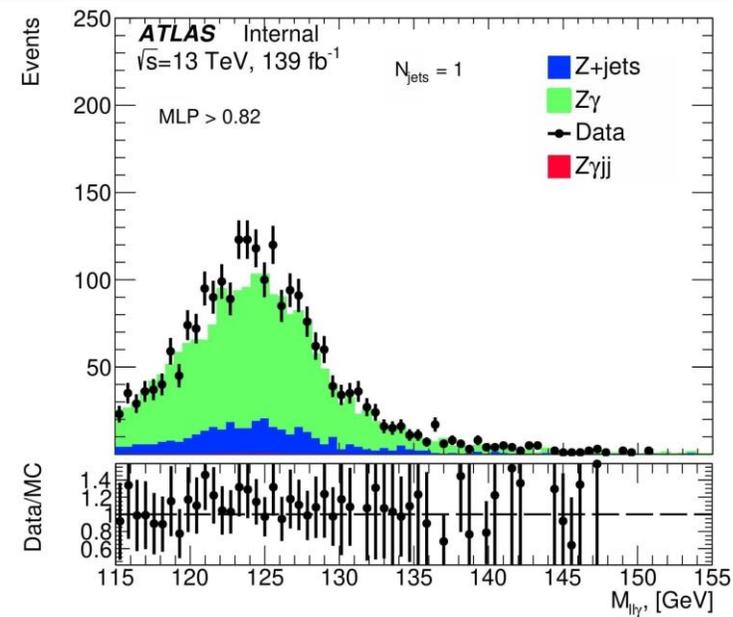
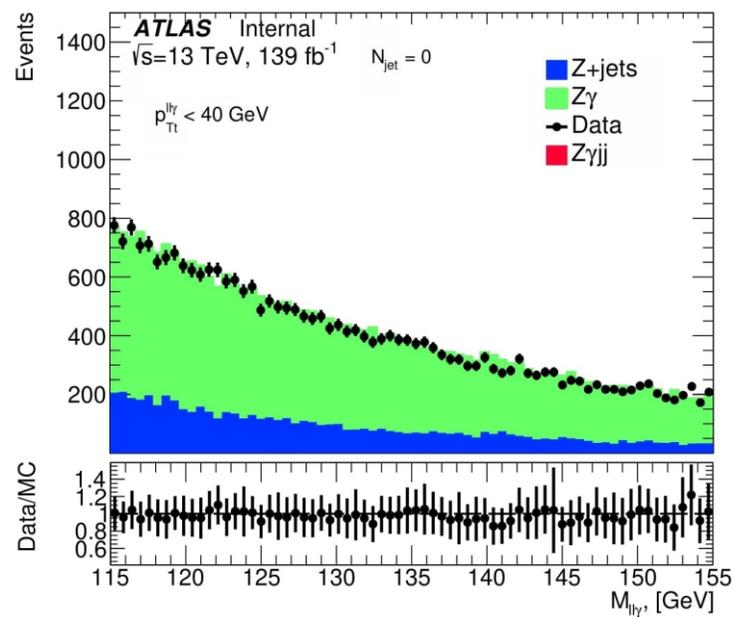
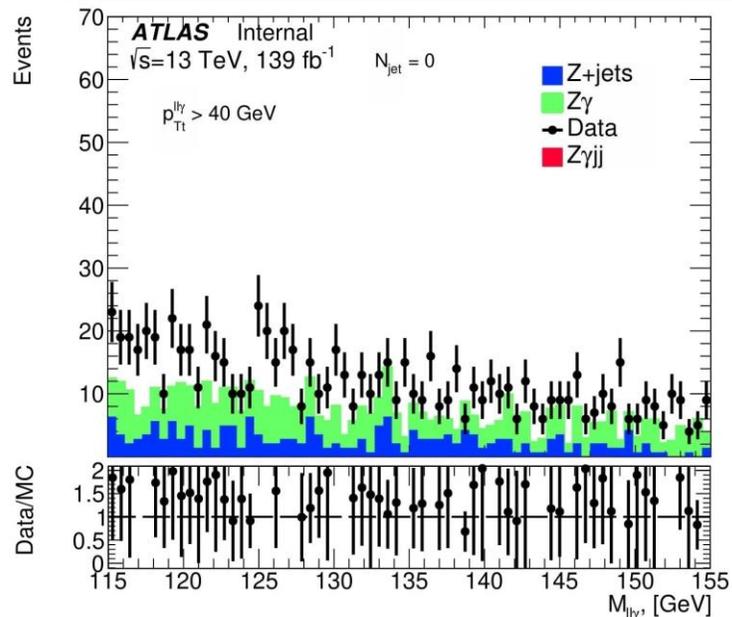
# Классификатор BDTG



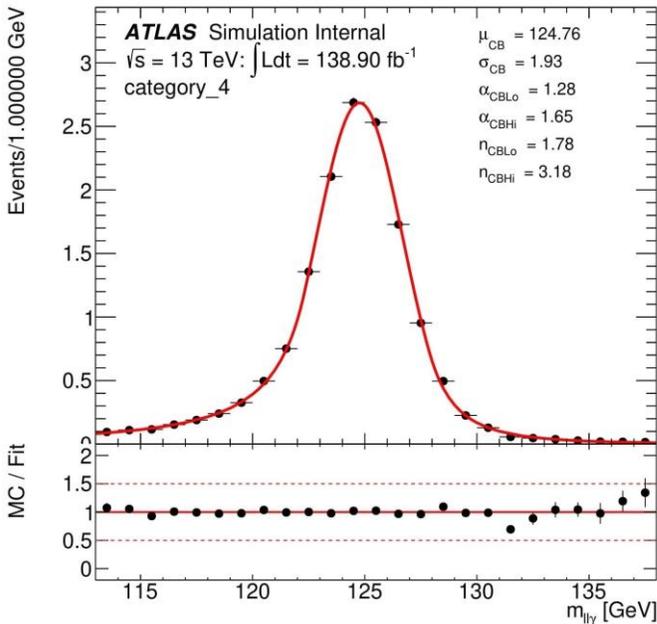
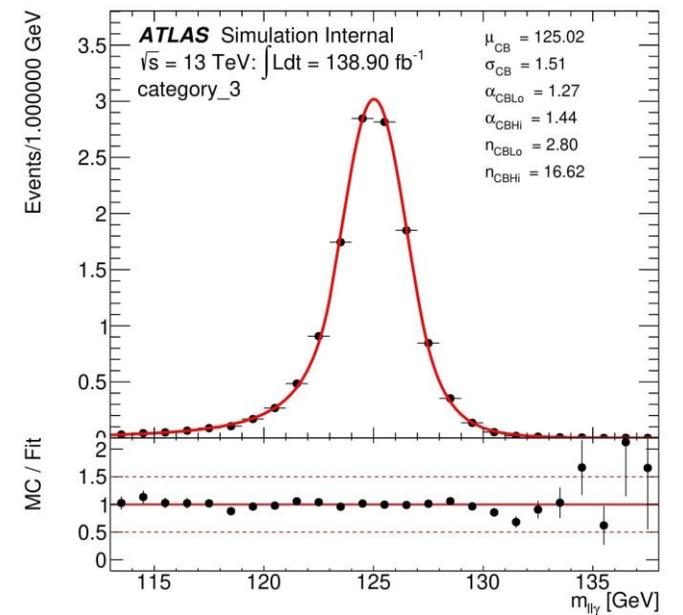
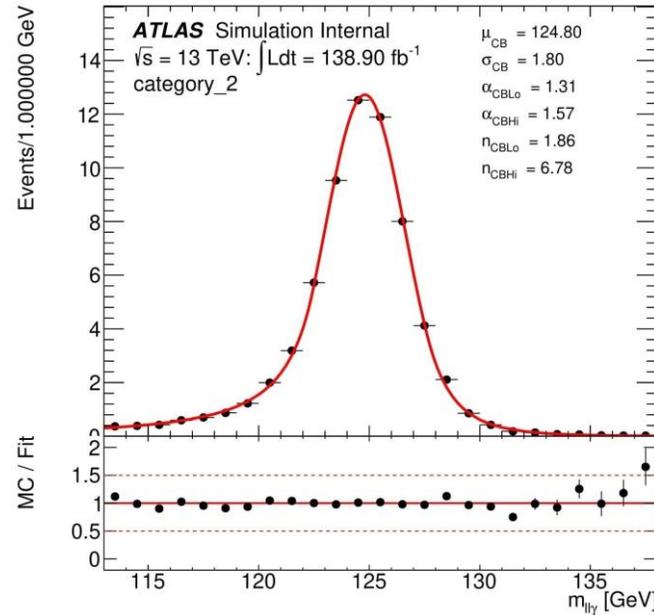
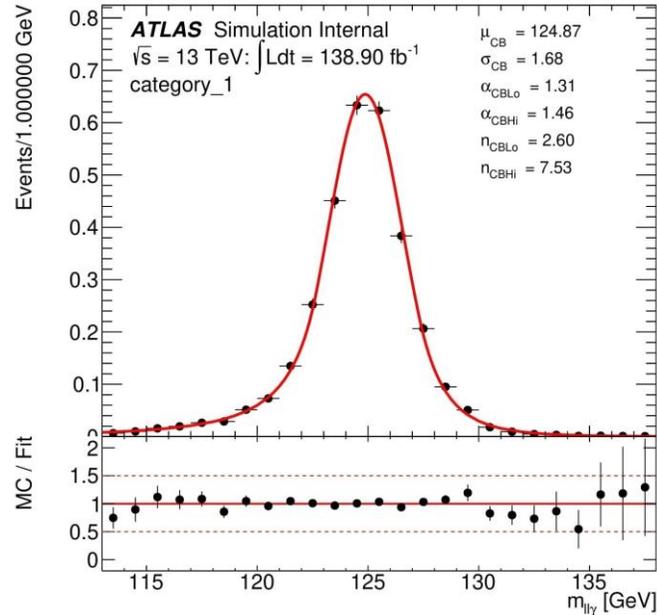
# MLP и BDTG



# Сравнение фона и данных

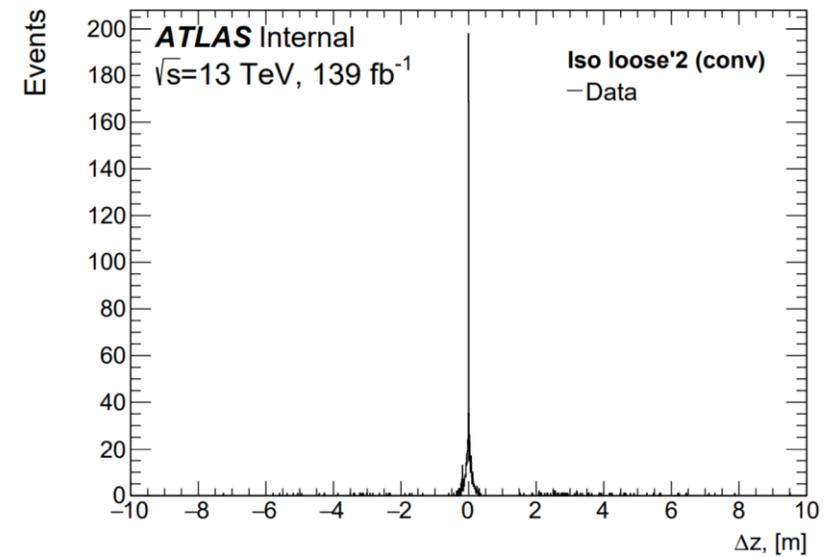
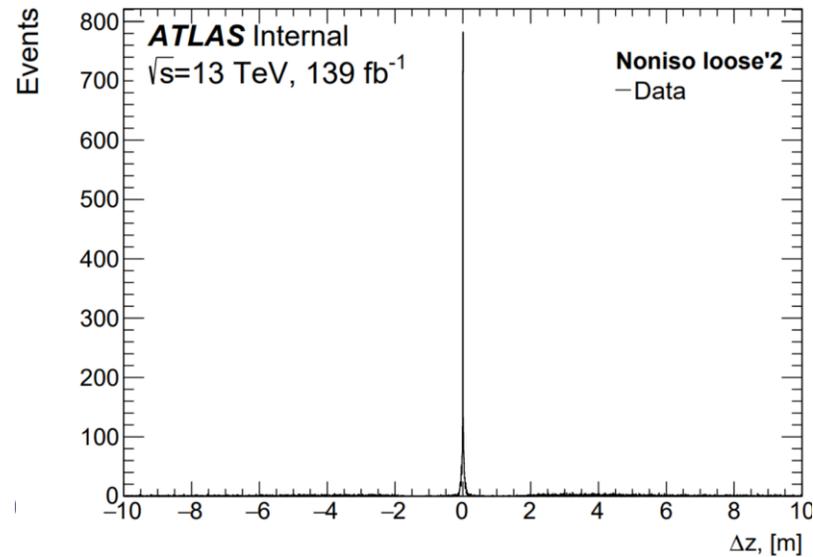
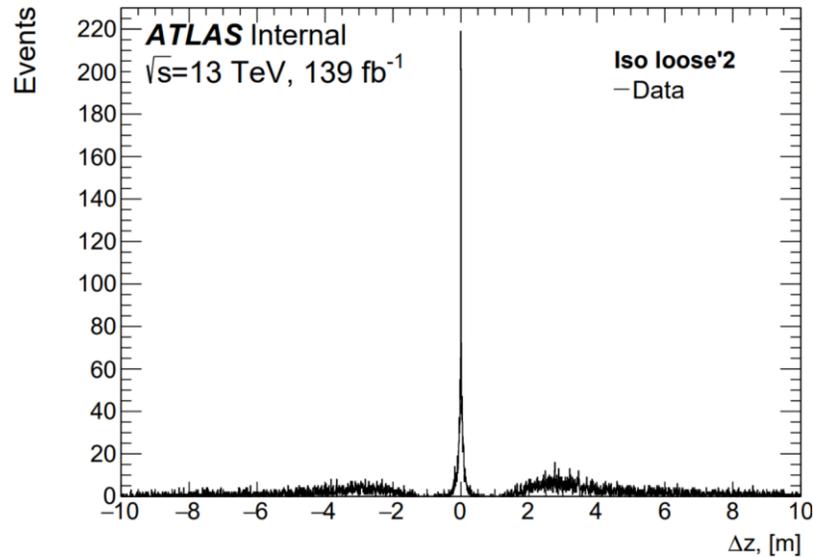
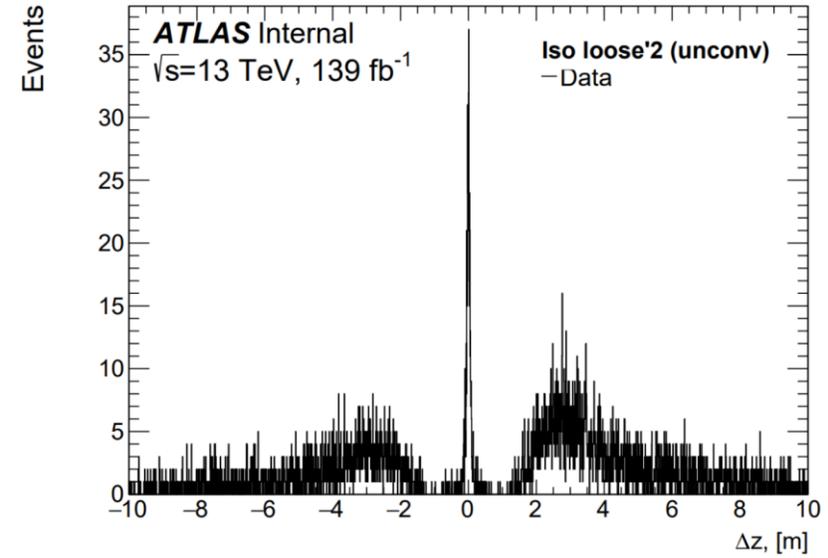
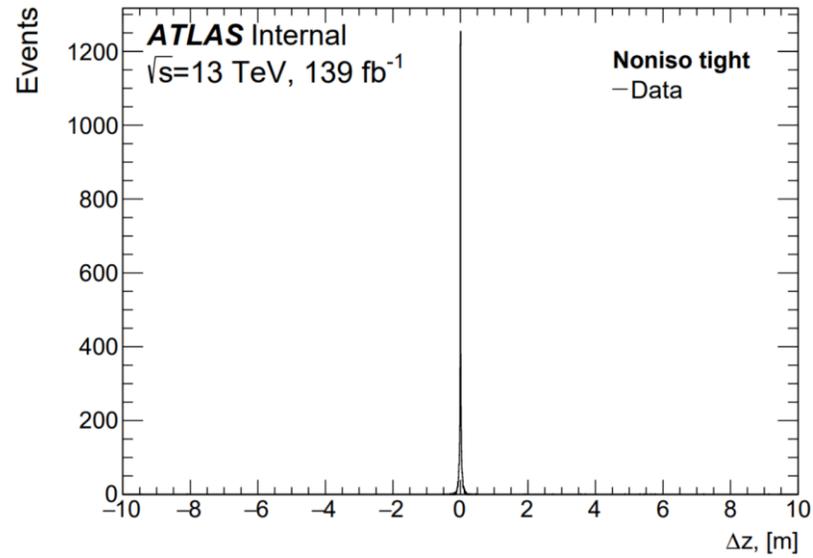
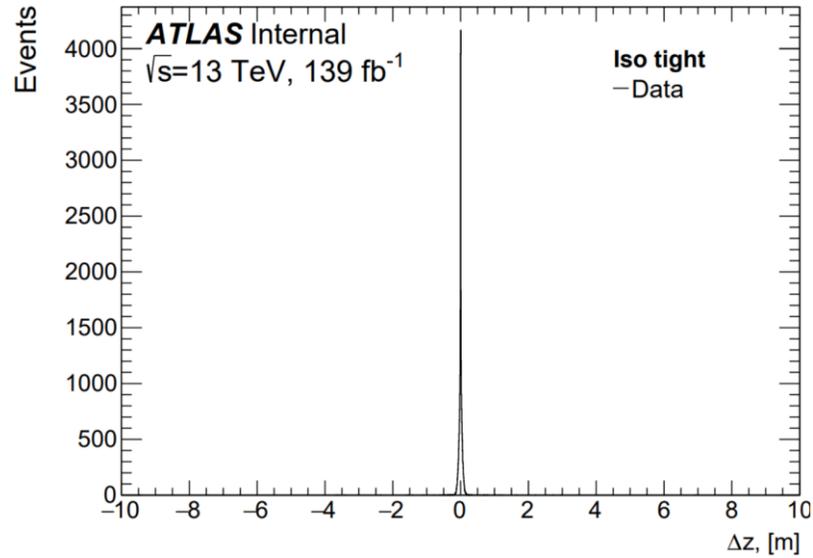


# Фитирование сигнала

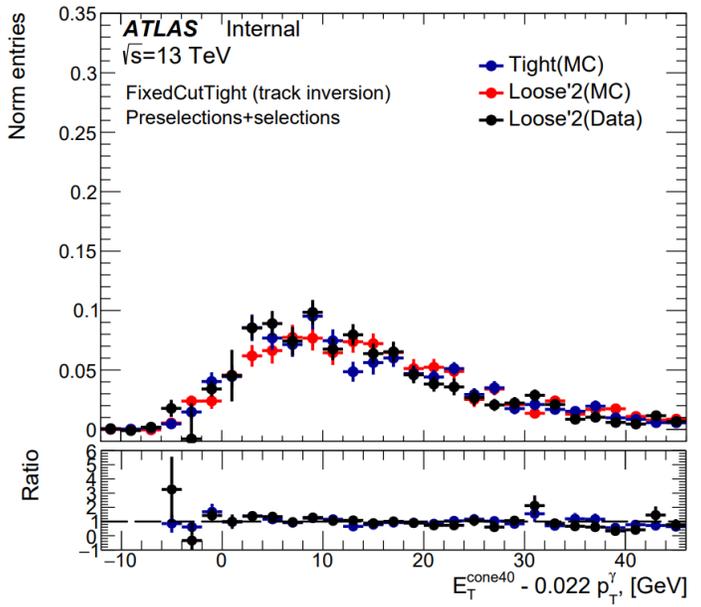
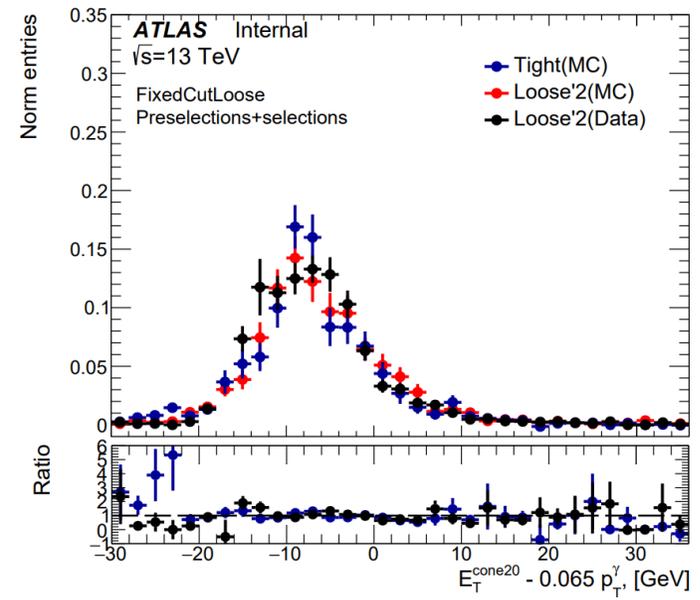
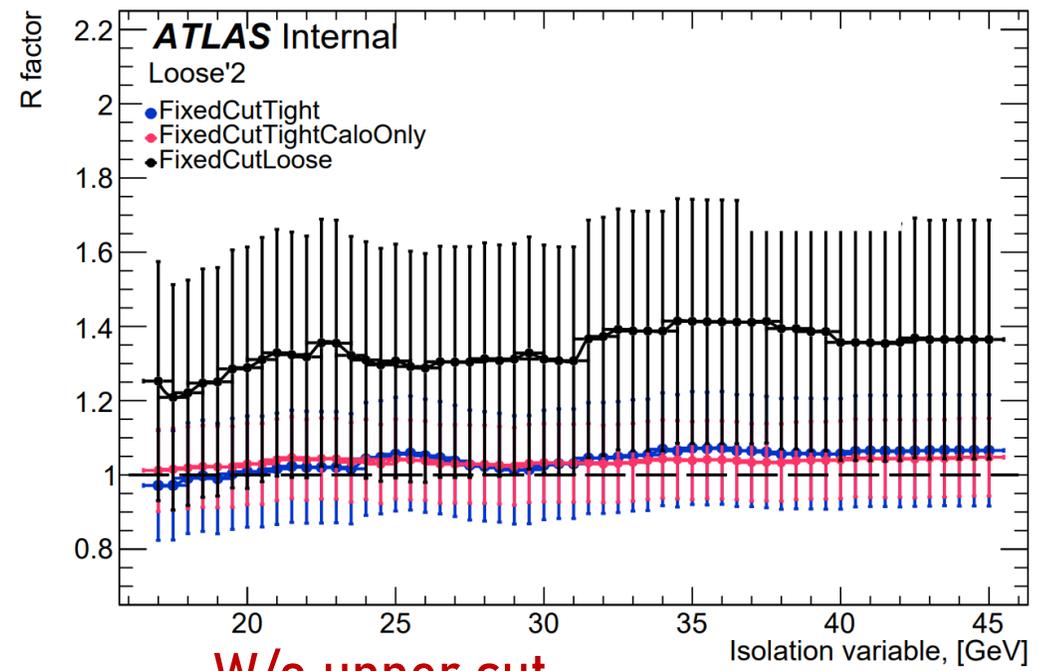
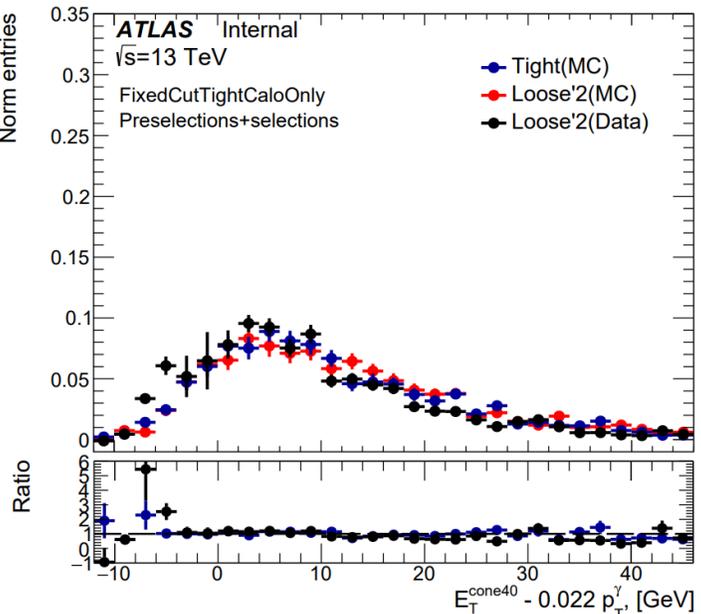
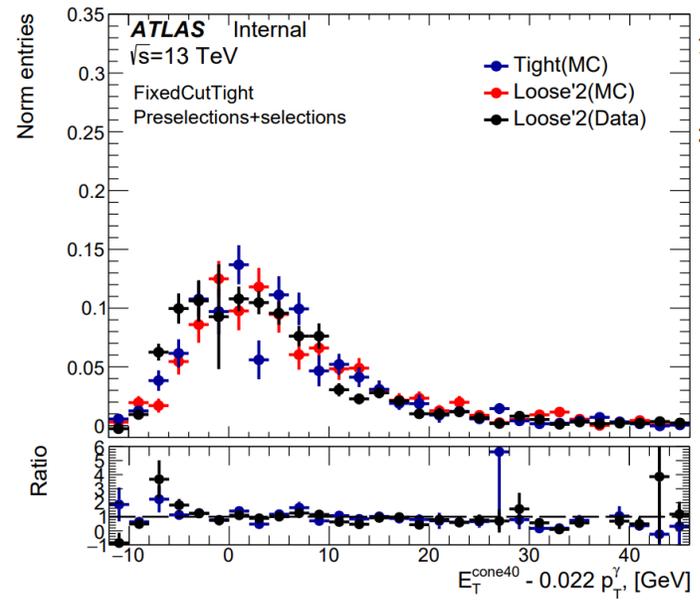


Категория	Значение массы, ГэВ	$\sigma_{CB}$ , ГэВ	$\chi^2/Ndf$
$N_{jet} = 0, p_{Tt}^{ll\gamma} > 40 \text{ ГэВ}$	$124.87 \pm 1.49$	$1.68 \pm 0.96$	0.86
$N_{jet} = 0, p_{Tt}^{ll\gamma} < 40 \text{ ГэВ}$	$124.80 \pm 0.48$	$1.80 \pm 0.64$	3.41
$N_{jet} = 1, \text{MLP} > 0.82$	$125.02 \pm 0.65$	$1.51 \pm 0.90$	1.84
$N_{jet} = 1, \text{MLP} < 0.82$	$124.76 \pm 0.86$	$1.93 \pm 1.02$	2.10
$N_{jet} \geq 2, \text{BDTG} > 0.76$	$124.89 \pm 0.92$	$1.40 \pm 0.95$	0.89
$N_{jet} \geq 2, \text{BDTG} < 0.76$	$124.96 \pm 1.11$	$1.83 \pm 1.13$	1.28

# Фон, обусловленный конфигурацией пучка



# Распределение по изоляции и R (loose'2)



W/o upper cut

Isolation	$R_{MC}$ factor
FCTight	$1.05 \pm 0.15$
FCCaloOnly	$1.06 \pm 0.10$
FCLoose	$1.29 \pm 0.28$
FCT (inversion)	$1.01 \pm 0.12$

- Корреляция для изоляции FixedCutLoose  
➡ изоляция не рассматривается
- Другие рабочие точки в **backup**

# Оптимизация R фактора на данных (I/IV)

- Изоляция: **FixedCutTight**, без ограничения сверху

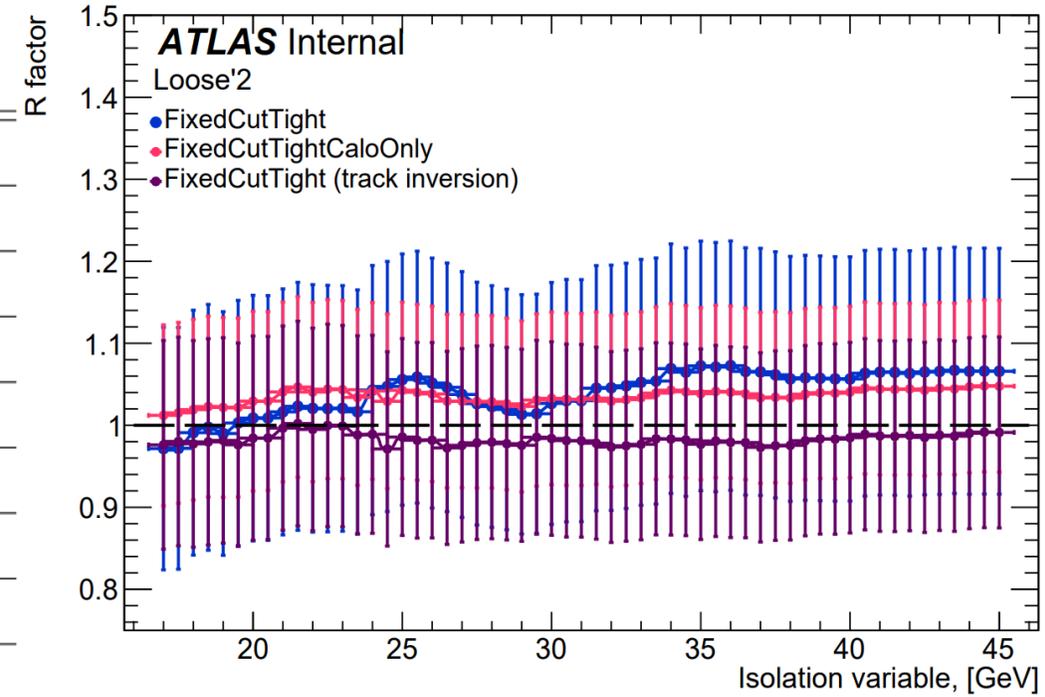
**FixedCutTight**, (w/o upper cut)

MC

	loose'2	loose'3	loose'4	loose'5
R-factor	$1.05 \pm 0.15$	$1.14 \pm 0.15$	$1.19 \pm 0.14$	$1.39 \pm 0.17$

Data-driven

Cut	loose'2	loose'3	loose'4	loose'5
<b>7.95</b>	$1.6 \pm 0.3$	$1.5 \pm 0.3$	$1.4 \pm 0.3$	$1.4 \pm 0.3$
<b>8.45</b>	$1.5 \pm 0.3$	$1.5 \pm 0.3$	$1.4 \pm 0.3$	$1.4 \pm 0.3$
<b>8.95</b>	$1.4 \pm 0.3$	$1.3 \pm 0.3$	$1.3 \pm 0.3$	$1.3 \pm 0.3$
<b>9.45</b>	$1.6 \pm 0.4$	$1.5 \pm 0.4$	$1.5 \pm 0.4$	$1.5 \pm 0.3$
<b>9.95</b>	$1.6 \pm 0.4$	$1.5 \pm 0.4$	$1.7 \pm 0.4$	$1.6 \pm 0.4$



- Оптимизация **не применима** из-за корреляций и ограниченной статистики

# Оптимизация R фактора на данных (II/IV)

- Изоляция: **FixedCutTight**, ограничение сверху 25.45 GeV

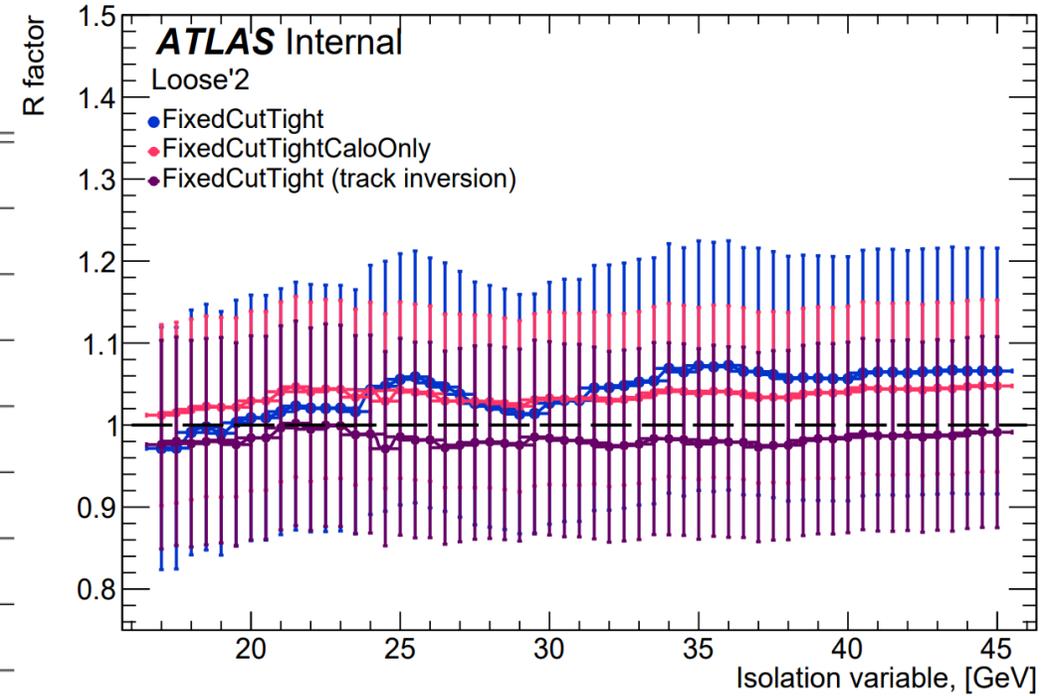
**FixedCutTight**, (upper cut = 25.45 GeV)

MC

	loose'2	loose'3	loose'4	loose'5
R-factor	$1.06 \pm 0.15$	$1.15 \pm 0.16$	$1.21 \pm 0.15$	$1.40 \pm 0.17$

Data-driven

Cut	loose'2	loose'3	loose'4	loose'5
8.45	$1.1 \pm 0.2$	$1.1 \pm 0.2$	$1.03 \pm 0.18$	$1.06 \pm 0.18$
8.95	$0.96 \pm 0.18$	$0.97 \pm 0.17$	$0.96 \pm 0.17$	$0.97 \pm 0.16$
9.05	$1.01 \pm 0.18$	$1.02 \pm 0.18$	$1.01 \pm 0.18$	$1.01 \pm 0.17$
9.45	$1.08 \pm 0.19$	$1.10 \pm 0.19$	$1.10 \pm 0.19$	$1.12 \pm 0.18$
9.95	$1.03 \pm 0.18$	$1.03 \pm 0.18$	$1.16 \pm 0.19$	$1.16 \pm 0.19$
10.45	$1.1 \pm 0.2$	$1.1 \pm 0.2$	$1.2 \pm 0.2$	$1.2 \pm 0.2$
10.95	$1.2 \pm 0.2$	$1.2 \pm 0.2$	$1.3 \pm 0.2$	$1.3 \pm 0.2$



Для loose'2 для данных  $\delta = 18\%$

- Достаточно большая погрешность

# Оптимизация R фактора на данных (III/IV)

- Изоляция: **FixedCutTight** с инверсией трековой изоляции

**FixedCutTight (inversion)**, (w/o upper cut)

MC

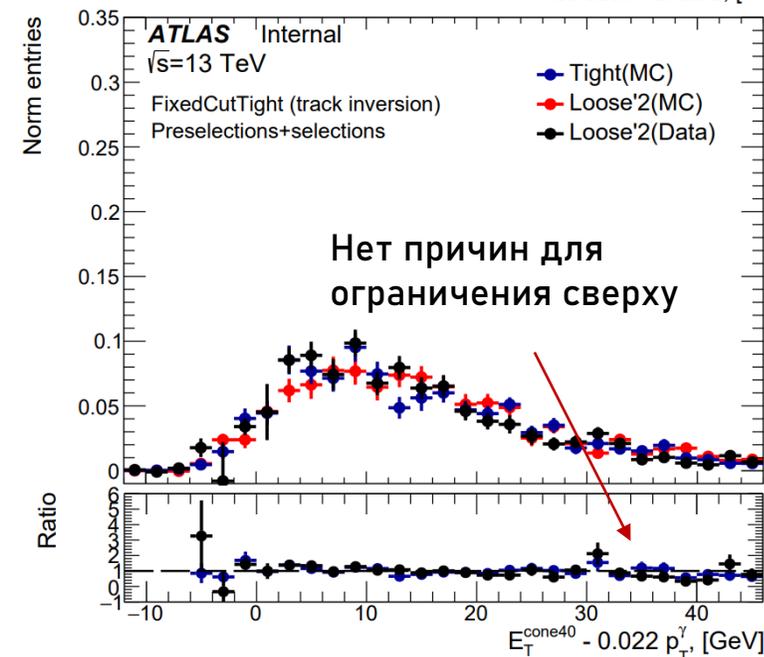
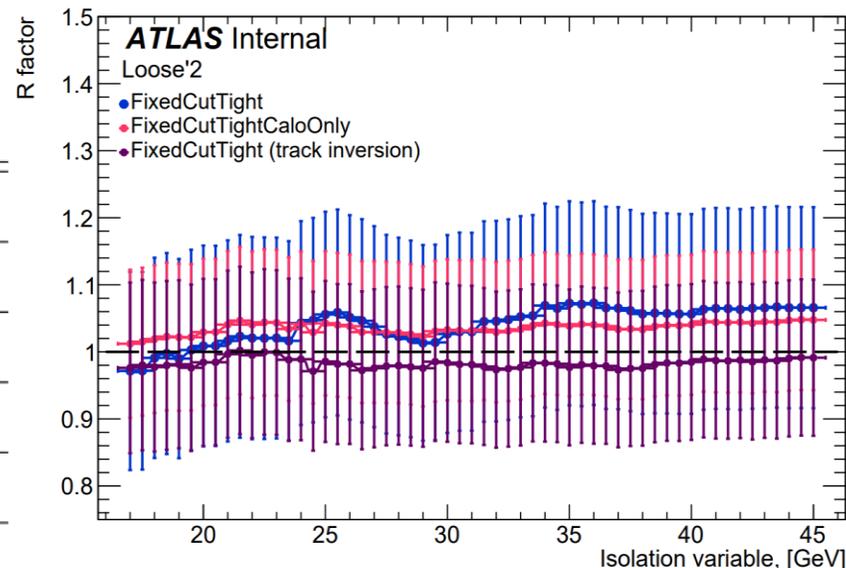
	loose'2	loose'3	loose'4	loose'5
R-factor	$1.01 \pm 0.12$	$1.15 \pm 0.12$	$1.29 \pm 0.13$	$1.58 \pm 0.16$

Data-driven

Cut	loose'2	loose'3	loose'4	loose'5
9.45	$1.09 \pm 0.13$	$1.15 \pm 0.13$	$1.09 \pm 0.11$	$1.13 \pm 0.11$
9.95	$1.08 \pm 0.12$	$1.16 \pm 0.12$	$1.11 \pm 0.11$	$1.13 \pm 0.10$
10.20	$1.07 \pm 0.12$	$1.13 \pm 0.12$	$1.09 \pm 0.10$	$1.12 \pm 0.10$
10.45	$1.09 \pm 0.12$	$1.14 \pm 0.12$	$1.10 \pm 0.10$	$1.14 \pm 0.10$
10.95	$1.18 \pm 0.13$	$1.23 \pm 0.12$	$1.17 \pm 0.10$	$1.20 \pm 0.10$

Для loose'2 для данных  $\delta = 11\%$

- Оптимизация применима



# Оптимизация R фактора на данных (IV/IV)

- Изоляция: **FixedCutTightCaloOnly**, без верхнего ограничения

**FixedCutTightCaloOnly, (w/o upper cut)**

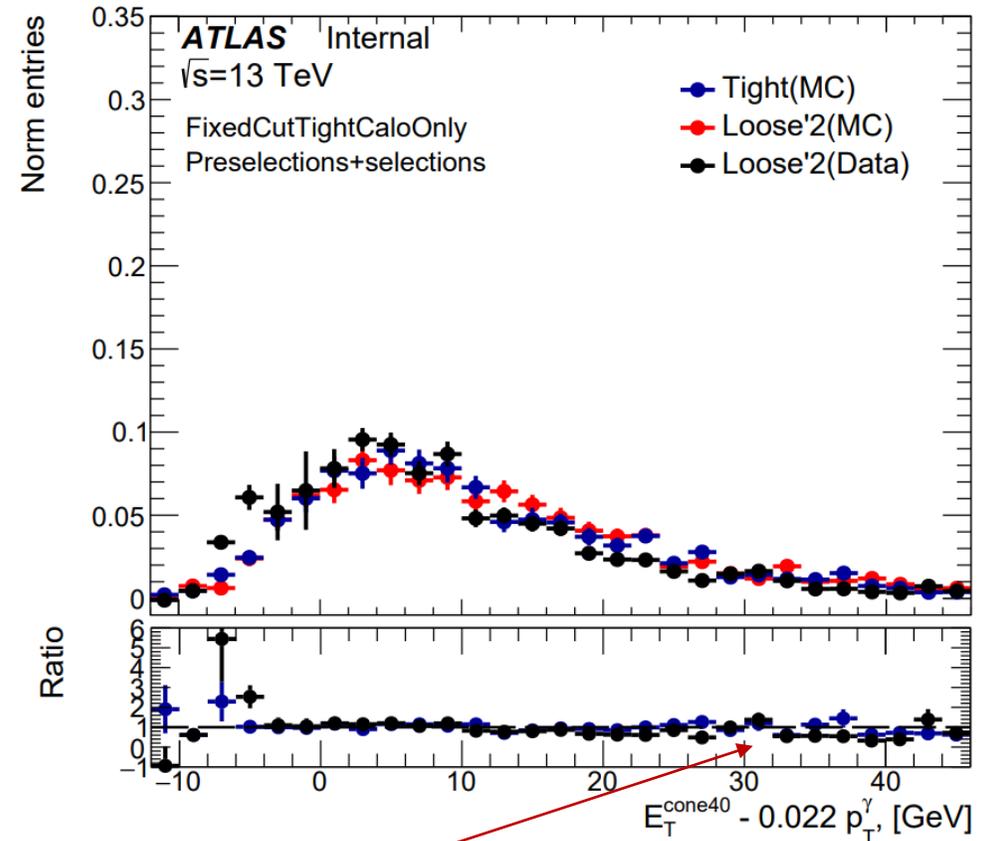
MC

	loose'2	loose'3	loose'4	loose'5
R-factor	$1.06 \pm 0.10$	$1.14 \pm 0.11$	$1.22 \pm 0.10$	$1.40 \pm 0.12$

Data-driven

Cut	loose'2	loose'3	loose'4	loose'5
9.45	$1.08 \pm 0.11$	$1.14 \pm 0.11$	$1.12 \pm 0.10$	$1.13 \pm 0.10$
9.95	$1.07 \pm 0.10$	$1.13 \pm 0.10$	$1.15 \pm 0.10$	$1.15 \pm 0.10$
10.45	$1.09 \pm 0.10$	$1.14 \pm 0.10$	$1.14 \pm 0.10$	$1.15 \pm 0.10$
10.95	$1.18 \pm 0.11$	$1.23 \pm 0.11$	$1.21 \pm 0.10$	$1.22 \pm 0.10$
11.45	$1.23 \pm 0.11$	$1.27 \pm 0.11$	$1.22 \pm 0.10$	$1.22 \pm 0.10$

- Применимые значения R фактора на данных



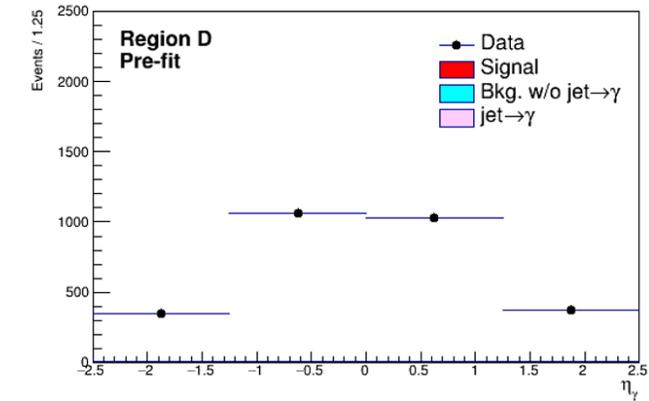
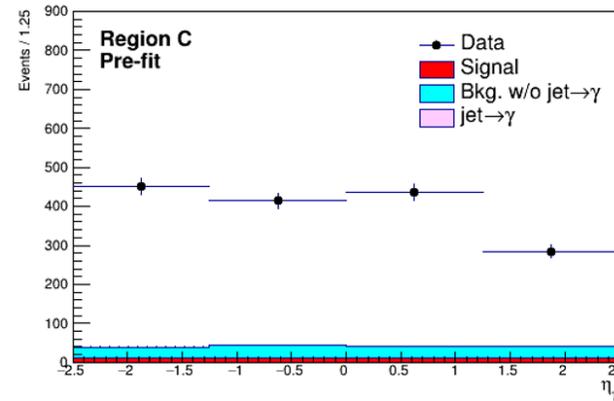
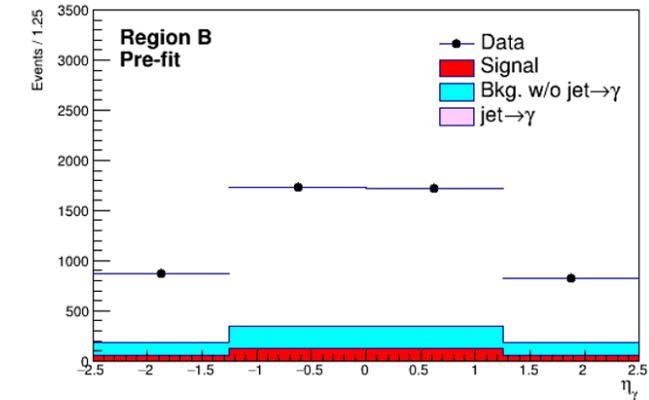
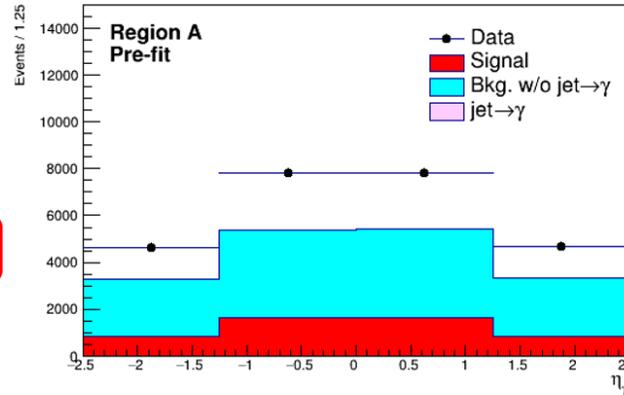
Нет причин для ограничения сверху

Для loose'2 для данных  $\delta = 9\%$

# Метод максимального правдоподобия

Variable: $\eta_\gamma$			
$N_{\text{bins}}$	Estimate	R factor	A: $\chi^2/N_{\text{dof}}$
3	$1965^{+78}_{-74}$	$1.09 \pm 0.05$	0.70
4	$1882^{+74}_{-70}$	$1.06 \pm 0.04$	1.07
5	$1834^{+74}_{-69}$	$1.03 \pm 0.03$	0.90
6	$1794^{+78}_{-74}$	$1.01 \pm 0.03$	0.83
7	$1744^{+80}_{-77}$	$0.98 \pm 0.03$	0.72
8	$1694^{+75}_{-68}$	$0.96 \pm 0.02$	0.45

Variable: $\phi_\gamma$			
$N_{\text{bins}}$	Estimate	R factor	A: $\chi^2/N_{\text{dof}}$
3	$1780^{+74}_{-69}$	$0.99 \pm 0.05$	1.53
4	$1801^{+72}_{-68}$	$0.99 \pm 0.04$	0.34
5	$1743^{+69}_{-65}$	$0.97 \pm 0.04$	0.96
6	$1784^{+72}_{-65}$	$0.99 \pm 0.03$	0.51
7	$1723^{+70}_{-65}$	$0.97 \pm 0.03$	0.82
8	$1763^{+73}_{-68}$	$0.98 \pm 0.03$	0.43

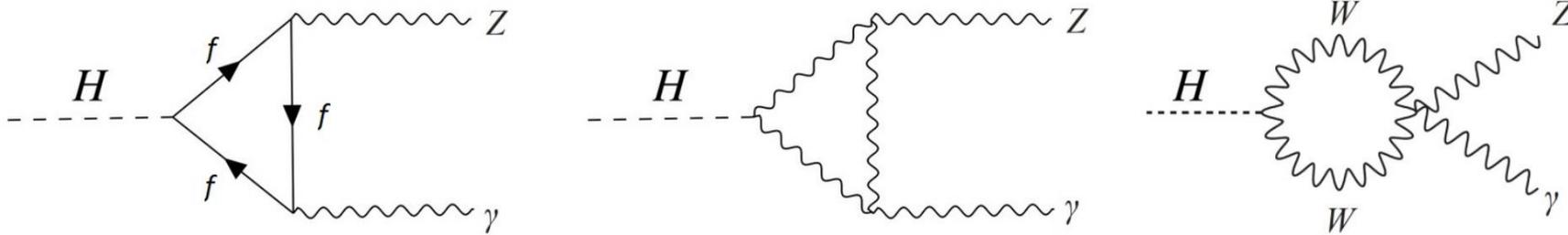


$$E_T^{\text{cone40}} \text{ И } E_T^{\text{cone20}}$$

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

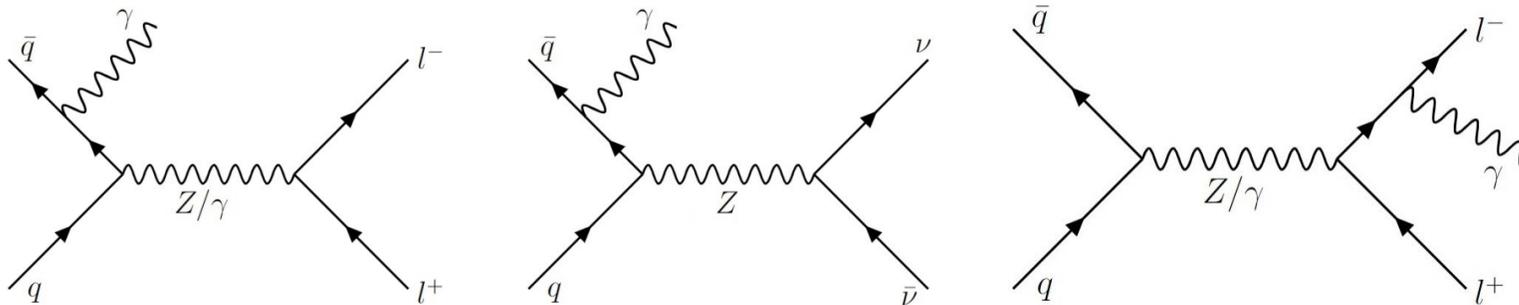
поперечная энергия, выделившаяся в калориметре в изолированном регионе (в конусе вокруг кластера, исключая сердцевину кластера) внутри конуса раствором  $\Delta R = 0.4$  и  $0.2$

# Процессы $H \rightarrow Z(\ell)\gamma$ и $Z\gamma$

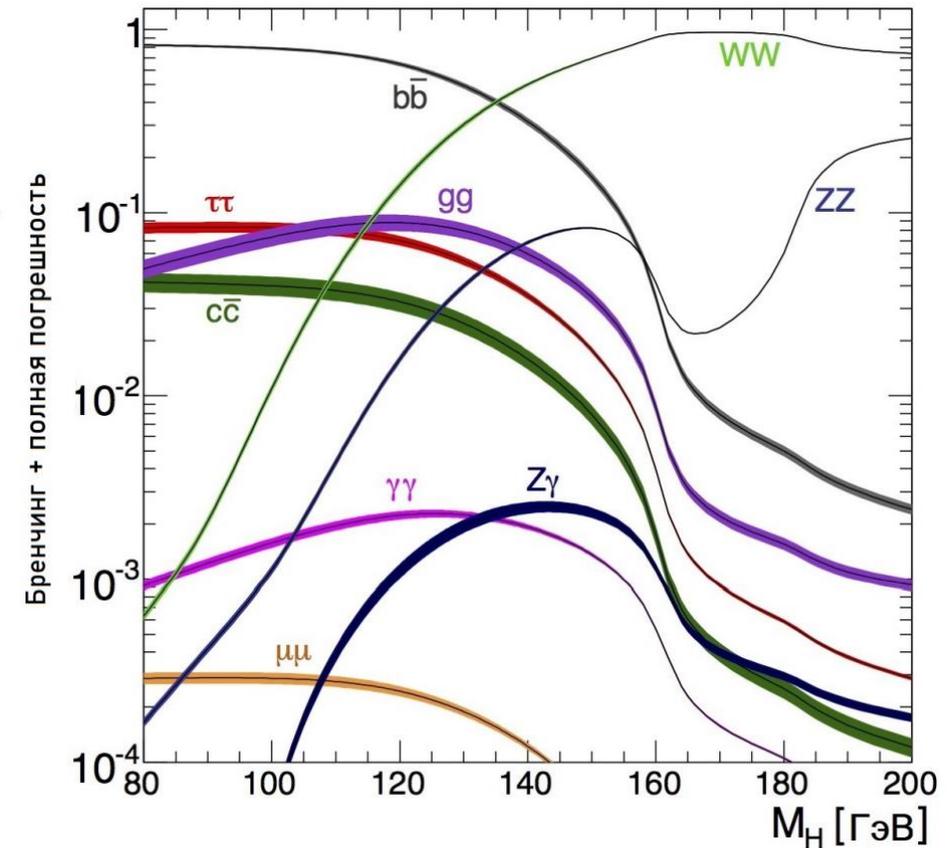


Сечение  $H \rightarrow Z(\ell)\gamma$ :  
 $\sigma = 0.0843 \pm 0.005$  пбн

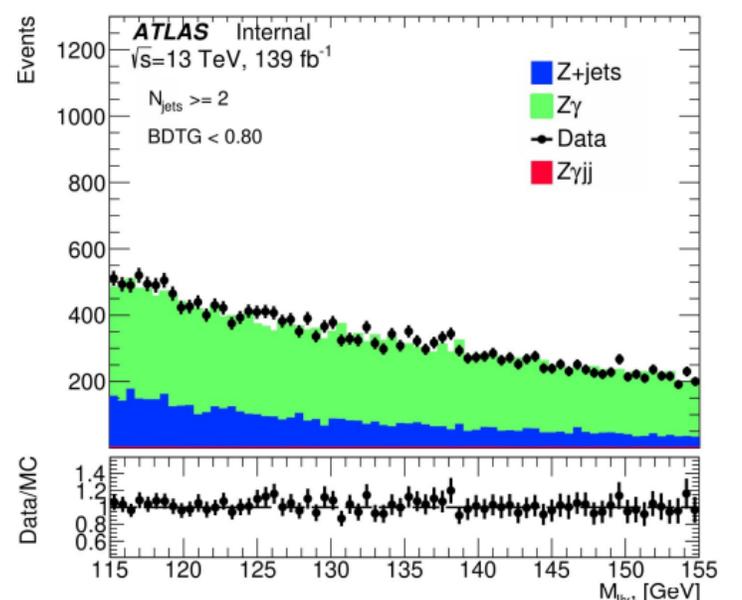
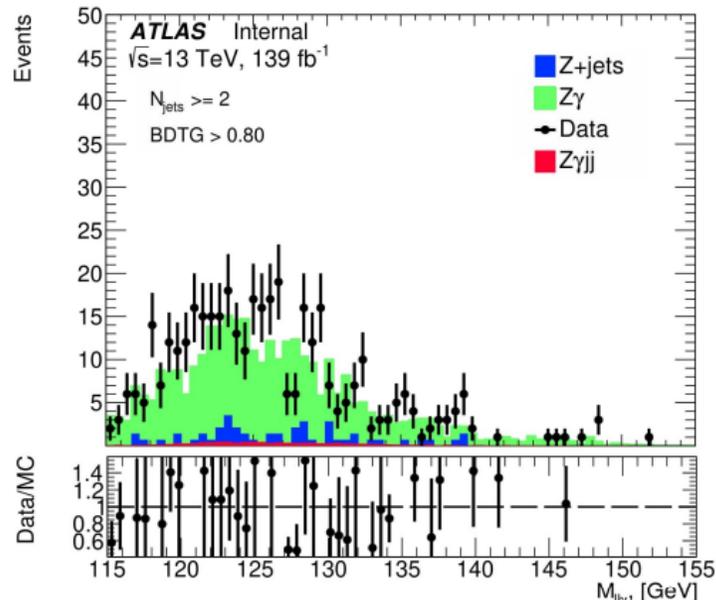
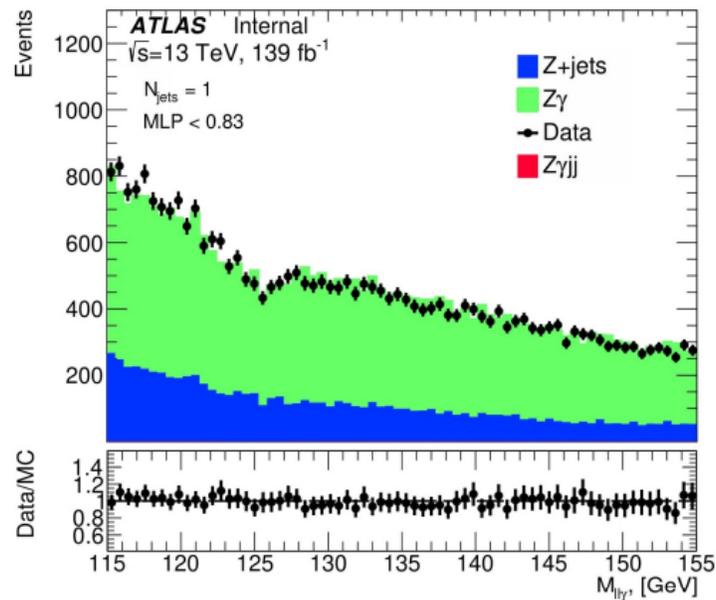
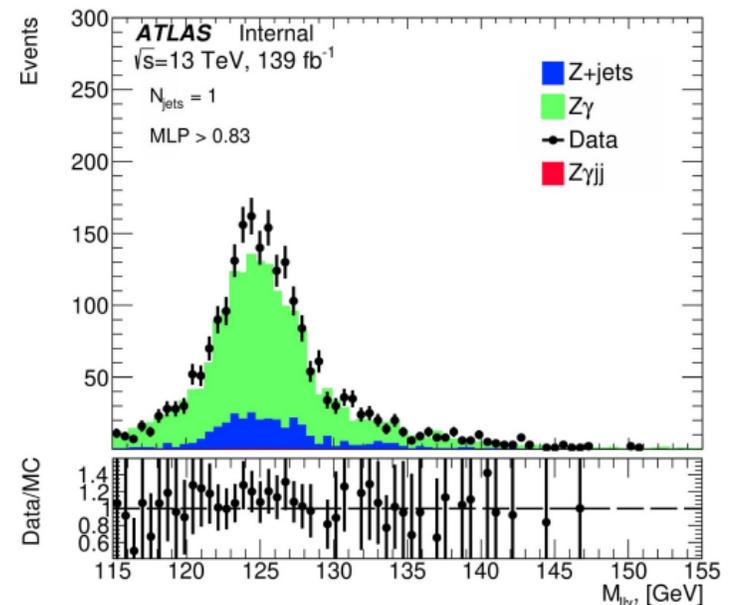
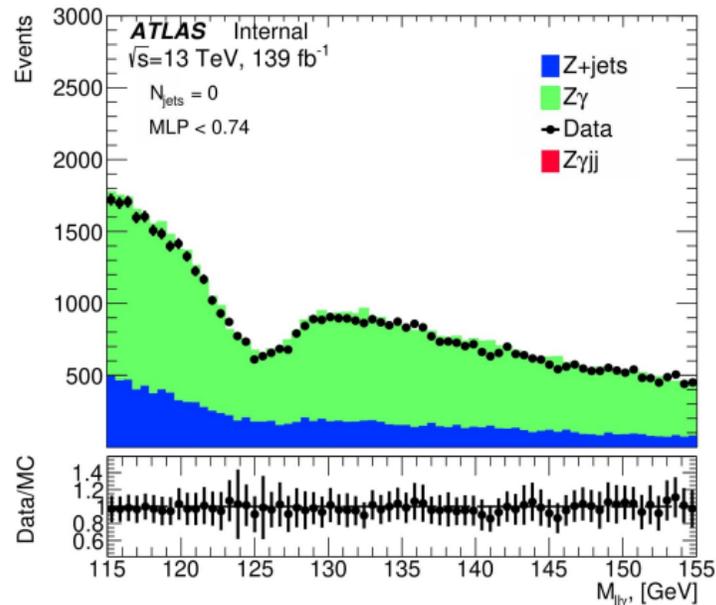
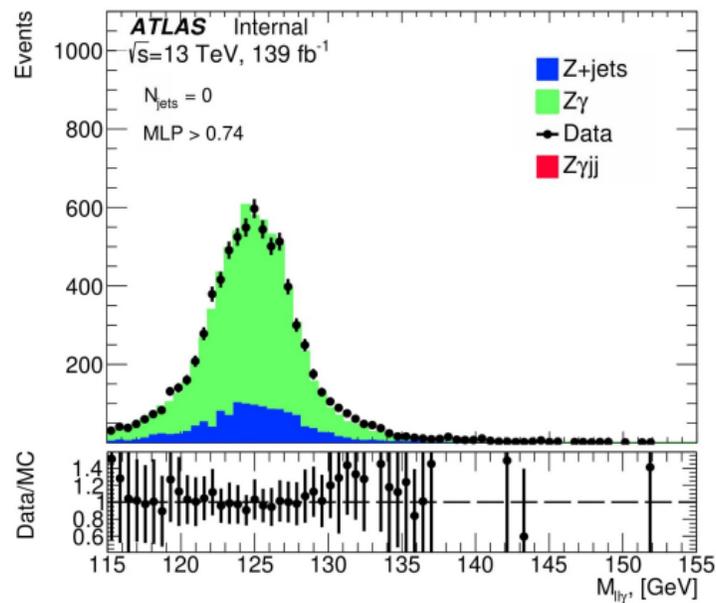
Текущий результат:  $m_H = 125.09 \pm 0.21$ (стат.)  $\pm 0.11$ (сист.) ГэВ



Канал образования	$m_H$ , ГэВ	Кол-во событий	p-tag	DSID	Сечение, пбн
ggH, Z(l)gam	125	1.436M	p4062(m16a/e)	345316	48.52
VBF, Z(l)gam	125	1.155M	p4062(m16a/e)	345833	3.779
WpH, Z(l)gam	125	60k	p4062(m16a/e)	345320	-
WmH, Z(l)gam	125	60k	p4062(m16a/e)	345321	1.369
ZH, Z(l)gam	125	120k	p4062(m16a/e)	345322	0.7612
ttH, Z(l)gam	125	5.398M	p4062(m16a/e)	346198	0.5065



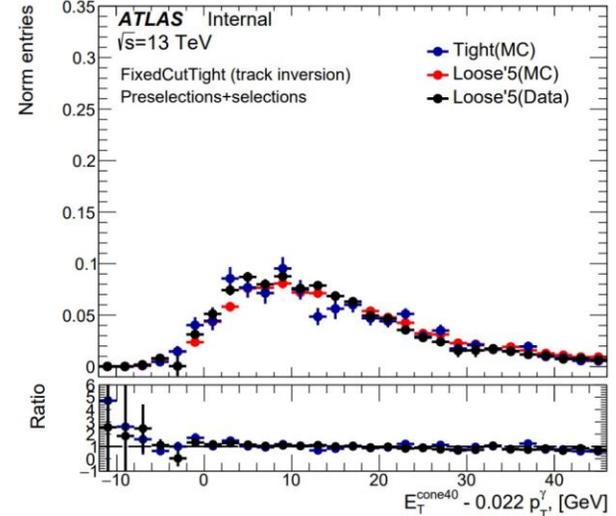
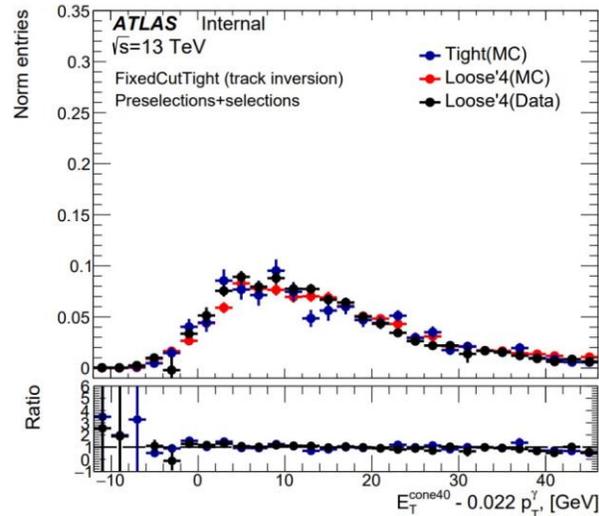
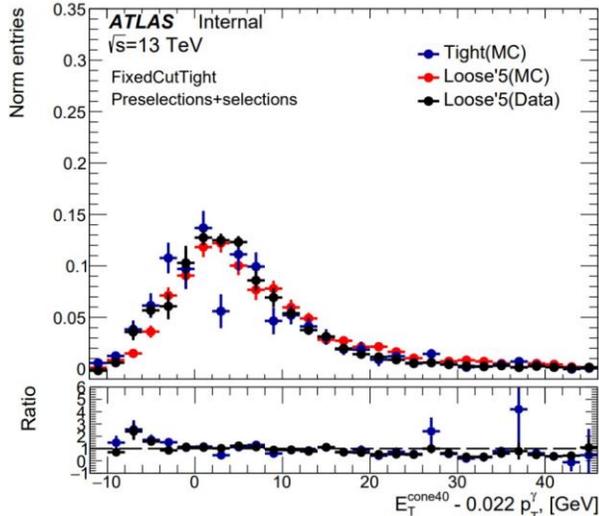
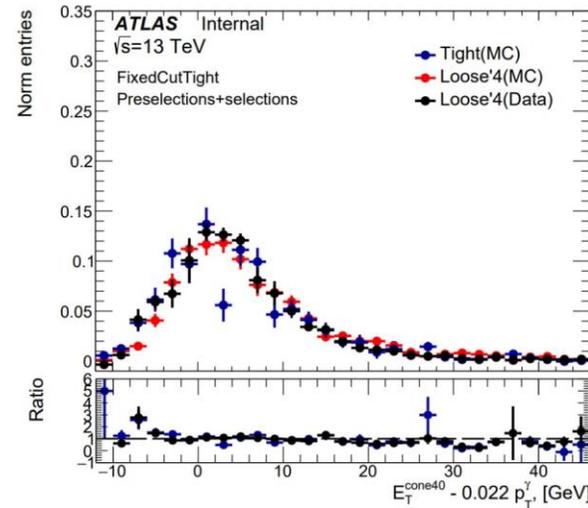
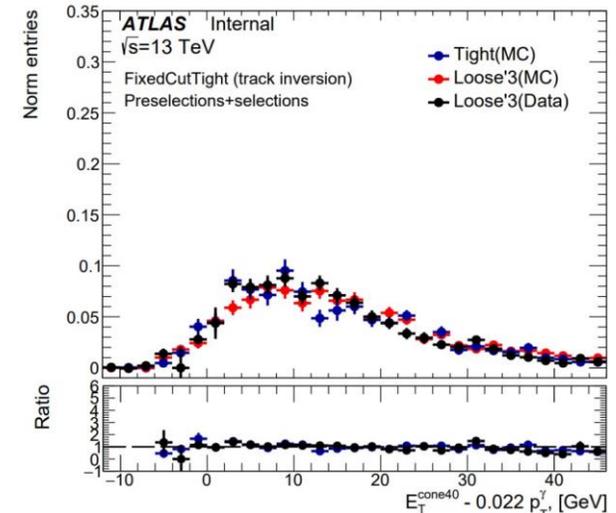
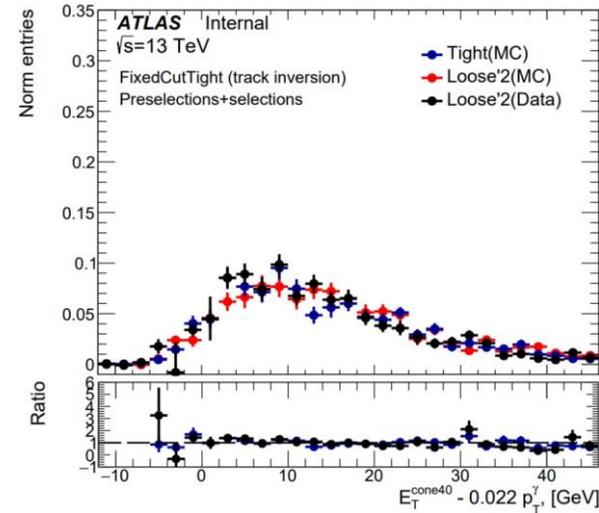
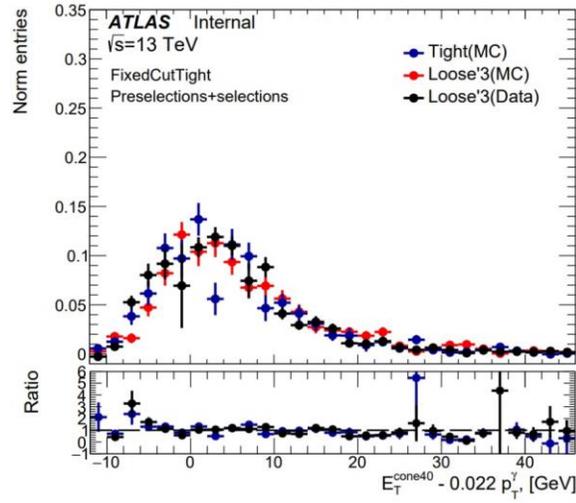
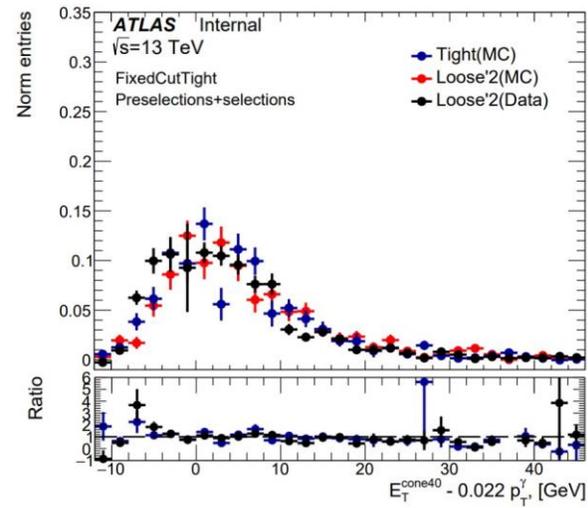
# Неприменимая категоризация



# Распределения по изоляции

## FixedCutTight

## FixedCutTight (трековая инверсия)



# Распределения по изоляции

## FixedCutTightCaloOnly

## FixedCutLoose

