Оценка фона Z+струи в исследовании ассоциированного рождения Z-бозона с фотоном

Научные руководители: Солдатов Е.Ю.
Пятиизбянцева Д.Н.
Студент: Казакова К.К

Кафедра физики элементарных частиц

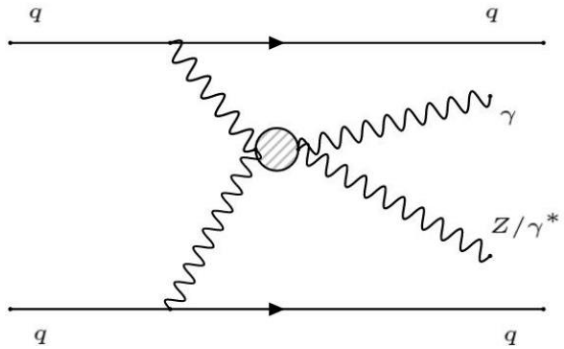
Москва 2021

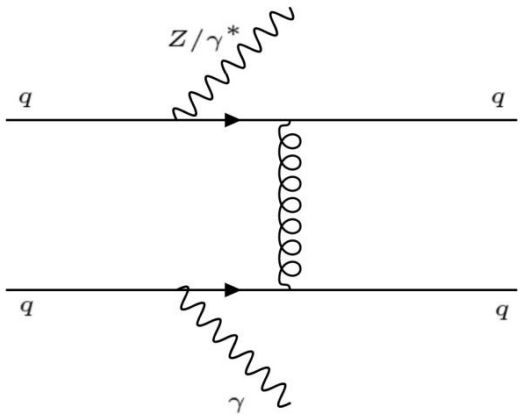
Описание процесса и мотивация

Исследуемые процессы: электрослабый Z(νν)γ и КХД Z(νν)γ

Фоновые процессы: W(→lν)γ, ttγ, e→γ, γ+jet, jet→γ, Z(→ll)γ

Мотивация: исследуемые процессы высокочувствительны к отклонениям от Стандартной Модели





Цели анализа:

- оптимально подавить фоновые события, обусловленные конфигурацией пучка
- оценить число фоновых событий, обусловленных неверной идентификацией адронной струи как фотона (jet→γ) с помощью ABCD-метода
- оценить статистические и систематические погрешности числа фоновых событий
- оценить число фоновых событий с помощью метода максимального правдоподобия и сравнить результат со стандартным ABCD-методом

Используемые отборы

- Сигнал: электрослабый Z(νν)γ и КХД Z(νν)γ

Предотборы

Переменная	Ограничение
$E_{\mathrm{T}}^{\mathrm{miss}}$	> 130 ГэВ
E_{T}^{γ}	> 150 ГэВ
Число фотонов	$N_\gamma = 1$
Число лептонов	$N_e = 0,\ N_\mu = 0$

Отборы

Переменная	Ограничение
Значимость $E_{\mathrm{T}}^{\mathrm{miss}}$	> 11
$\lvert\Delta\phi(E_{\mathrm{T}}^{\mathrm{miss}}, \gamma)\rvert$	> 0.7
$\lvert\Delta\phi(E_{\mathrm{T}}^{\mathrm{miss}}, j_1)\rvert$	> 0.4
$\lvert\Delta\phi(E_{\mathrm{T}}^{\mathrm{miss}}, j_2)\rvert$	> 0.3

⇨ Регион с такими предотборами и отборами – инклюзивный регион

Отбор фотонов

- В анализе используется три различные фотонные изоляции: FixedCutTight, FixedCutTightCaloOnly и FixedCutLoose

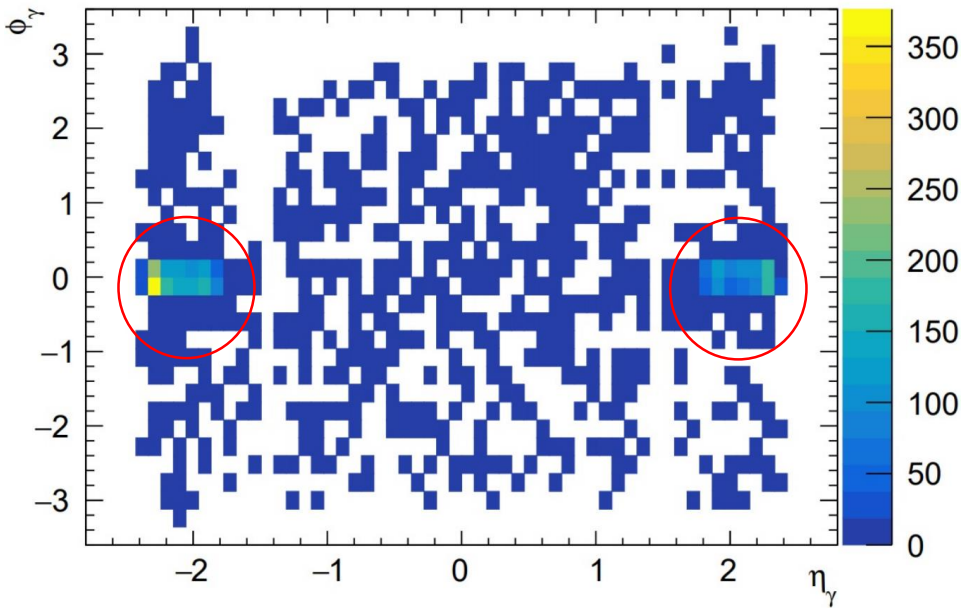
- Фотону присваивается значение <<жесткий>> (tight), если он удовлетворяет всем критериям формы ЭМ ливня

Изоляция	Калориметрическая изоляция	Трековая изоляция
FixedCutTightCaloOnly	$E_{\mathrm{T}}^{\mathrm{cone40}} - 0.022{\cdot}p_{\mathrm{T}}^{\gamma} < 2.45$ ГэВ	-
FixedCutTight	$E_{\mathrm{T}}^{\mathrm{cone40}} - 0.022{\cdot}p_{\mathrm{T}}^{\gamma} < 2.45$ ГэВ	$p_{\mathrm{T}}^{\mathrm{cone20}}/p_{\mathrm{T}}^{\gamma} < 0.05$
FixedCutLoose	$E_{\mathrm{T}}^{\mathrm{cone20}} - 0.065{\cdot}p_{\mathrm{T}}^{\gamma} < 0$ ГэВ	$p_{\mathrm{T}}^{\mathrm{cone20}}/p_{\mathrm{T}}^{\gamma} < 0.05$

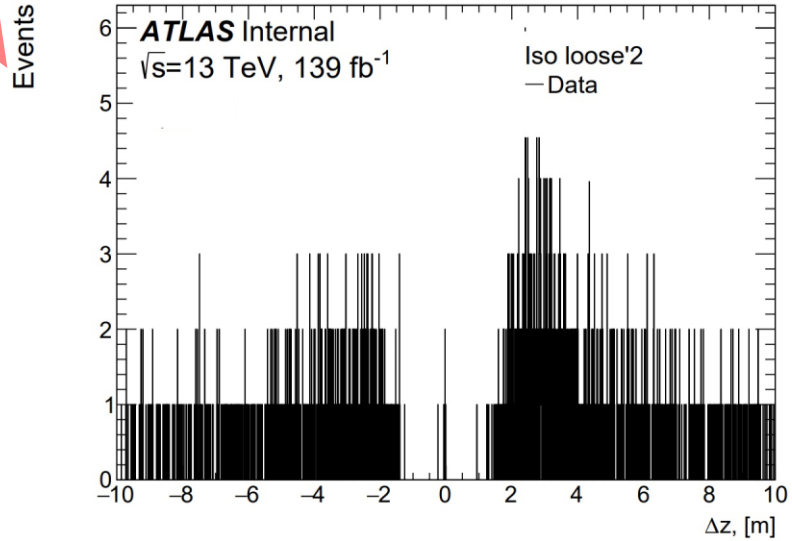
- Если по крайнем мере один из критериев формы ЭМ ливня нарушается, то фотону присваивается значение <<мягкий>> (loose').

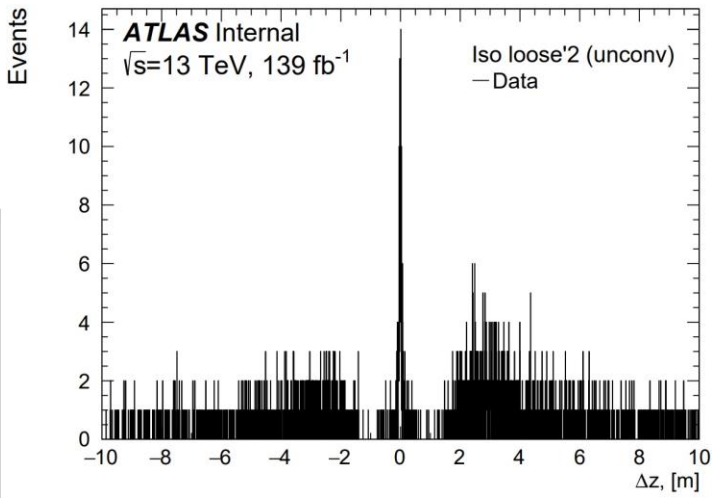
Фон, обусловленный конфигурацией пучка

- Изолированная область с <<мягкими>> фотонами содержит фоновые события, которые обусловлены конфигурацией пучка
- Преимущественно фон состоит из неконверсионных фотонов

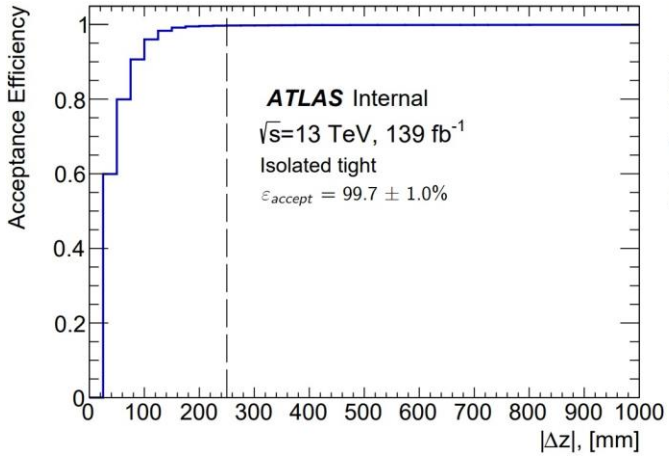


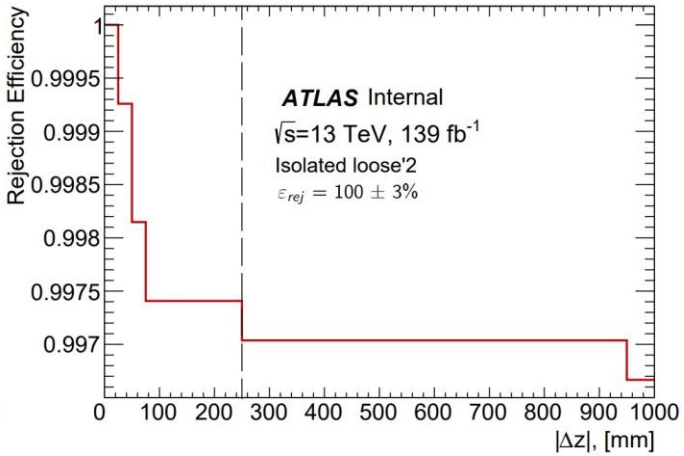
Ограничение на |φ| < 0.2 и |η| > 1.7





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





- Наиболее эффективному отбору соответствует |Δz| < 250 mm
- Значение для эффективности сигнала составляет ϵ_{accept}= 99.7 ± 1.0 %
- Значение для эффективности режекции ϵ_{rej} = 100 ± 3 %

ABCD-метод оценки фона jet → γ (I)

- В качестве переменных используются идентификационные и изоляционные критерии для фотонов, они не должны коррелировать
- Применена изоляция FixedCutTight

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

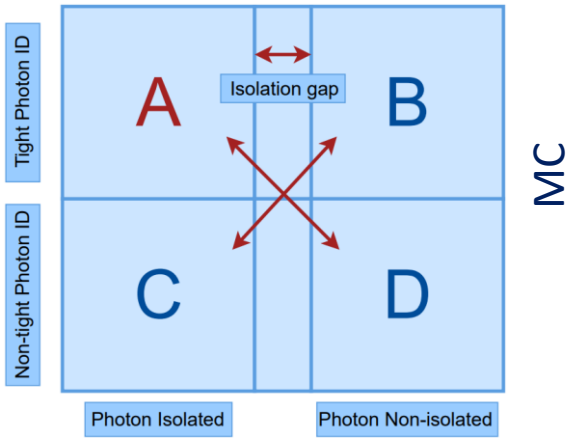
A (CP): $E_T^{cone40} - 0.022\, p_T^{\gamma} < 2.45$, tight

B (KO): $2.45 + \text{isogap} < E_T^{cone40} - 0.022\, p_T^{\gamma}$, tight

C (KO): $E_T^{cone40} - 0.022\, p_T^{\gamma} < 2.45$, non-tight

D (KO): $2.45 + \text{isogap} < E_T^{cone40} - 0.022\, p_T^{\gamma}$, non-tight

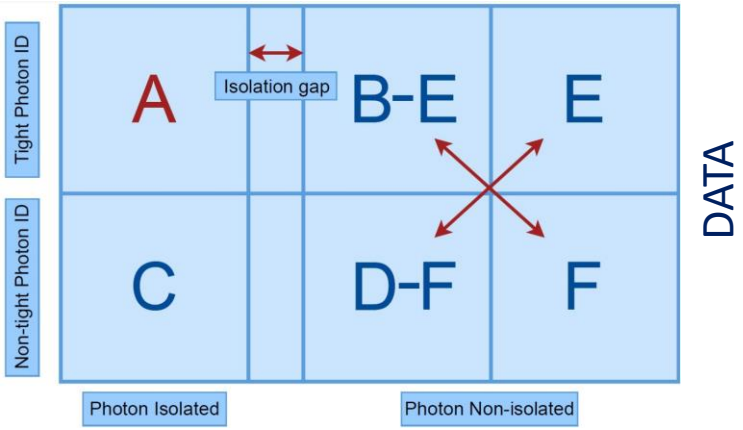
$+\ p_T^{cone20}/p_T^{\gamma} < 0.05$



iso gap = 2 ГэВ

- <<Мягкие>> фотоны (loose') удовлетворяют не всем критериям формы ЭМ ливня

$$R_{\text{data}} = \frac{N^{\text{data}}_{\text{B-E}} N^{\text{data}}_{\text{F}}}{N^{\text{data}}_{\text{D-F}} N^{\text{data}}_{\text{E}}}$$



Оценка R фактора до оптимизации

FixedCutTight	loose'2	loose'3	loose'4	loose'5
MC	1.10 ± 0.16	1.18 ± 0.15	1.24 ± 0.15	1.45 ± 0.17
Data-driven	2.0 ± 0.5	1.8 ± 0.5	1.7 ± 0.4	1.7 ± 0.4

⇨ Необходимо переоптимизировать регионы из-за сильной корреляции

ABCD-метод оценки фона jet → γ (II)

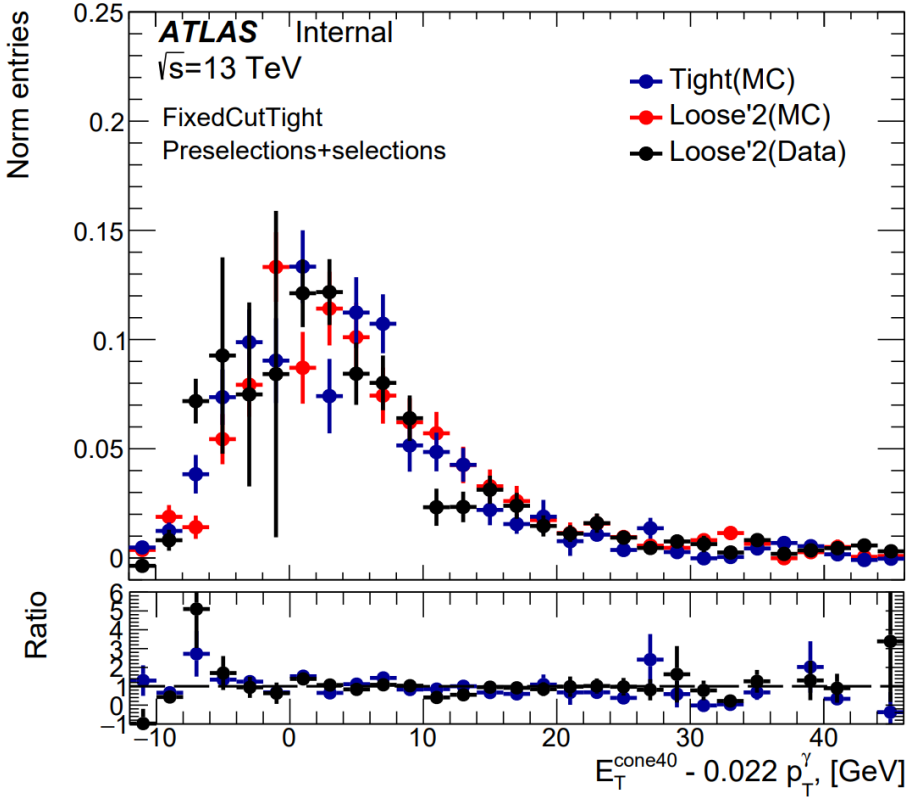
- Наиболее оптимальное ограничение сверху = 25.45 ГэВ

FixedCutTight, ограничение = 25.45 GeV				
MC				
	loose'2	loose'3	loose'4	loose'5
R-factor	1.12 ± 0.16	1.21 ± 0.17	1.28 ± 0.16	1.44 ± 0.18
Data-driven				
Cut	loose'2	loose'3	loose'4	loose'5
8.75	1.0 ± 0.3	1.0 ± 0.3	1.0 ± 0.3	1.1 ± 0.3

- R фактор на данных имеет крайне большую погрешность

⇨ применим инверсию трековой изоляции в неизолированных КО

FixedCutTight (инверсия), (ограничение = 25.45 GeV)				
MC				
	loose'2	loose'3	loose'4	loose'5
R-factor	1.07 ± 0.13	1.23 ± 0.14	1.34 ± 0.14	1.62 ± 0.17
Data-driven				
Cut	loose'2	loose'3	loose'4	loose'5
9.95	1.18 ± 0.12	1.23 ± 0.11	1.22 ± 0.10	1.26 ± 0.10
10.45	1.12 ± 0.11	1.16 ± 0.11	1.16 ± 0.10	1.22 ± 0.10
10.95	1.20 ± 0.12	1.25 ± 0.11	1.24 ± 0.11	1.30 ± 0.11



- Точка разделения КО B-E и E равна 10.45 ГэВ
- Наилучшей точкой выбрана loose'2

ABCD-метод оценки фона jet → γ (III)

- Каждый регион можно разбить следующим образом:

$$\begin{cases} N_A = N_A^{\text{sig}} + N_A^{\text{bkg}} + N_A^{\text{jet}\to\gamma}; \\ N_B = c_{\text{B}} N_A^{\text{sig}} + N_B^{\text{bkg}} + N_B^{\text{jet}\to\gamma}; \\ N_C = c_{\text{C}} N_A^{\text{sig}} + N_C^{\text{bkg}} + N_C^{\text{jet}\to\gamma}; \\ N_D = c_{\text{D}} N_A^{\text{sig}} + N_D^{\text{bkg}} + N_D^{\text{jet}\to\gamma}; \end{cases}$$

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

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

Параметры утечки сигнала в КО

	c_B	c_C	c_D
Значение	0.00645±0.00018	0.0091±0.0002	0.00011±0.00002

С учётом R фактора на данных, получим уравнение:

$$N_{\text{A}}^{\text{sig}} = \widetilde{N}_{\text{A}} - R(\widetilde{N}_{\text{B}} - c_{\text{B}} N_{\text{A}}^{\text{sig}}) \frac{\widetilde{N}_{\text{C}} - c_{\text{C}} N_{\text{A}}^{\text{sig}}}{\widetilde{N}_{\text{D}} - c_{\text{D}} N_{\text{A}}^{\text{sig}}}$$

- Количества фоновых событий в каждом регионе

	Data	Wγ QCD	Wγ EWK	W($e\nu$), top, tt	$tt\gamma$	γ + jet	$Z(ll)\gamma$	$W(\tau\nu)$
A	blinded ± 49	681 ± 8	65.6 ± 0.5	180 ± 3	184 ± 3	258 ± 16	20.6 ± 1.1	13 ± 4
B	619 ± 25	3.8 ± 0.5	0.23 ± 0.03	1.052 ± 0.014	1.3 ± 0.2	13 ± 3	0.12 ± 0.06	51 ± 4
C	67 ± 8	6.0 ± 0.7	0.68 ± 0.05	2.24 ± 0.03	2.0 ± 0.3	2.3 ± 1.2	0.14 ± 0.06	11 ± 3
D	431 ± 21	0.03 ± 0.03	0.011 ± 0.006	0 ± 0	0.11 ± 0.06	0 ± 0	0 ± 0	48 ± 4

Решение уравнения имеет вид: $N_{\text{A}}^{\text{sig}} = \dfrac{b - \sqrt{b^2 - 4ac}}{2a}$, где

$$\begin{cases} a = c_D - R c_B c_C; \\ b = \widetilde{N}_{\text{D}} + c_D \widetilde{N}_{\text{A}} - R(c_B \widetilde{N}_{\text{C}} + c_C \widetilde{N}_{\text{B}}); \\ c = \widetilde{N}_{\text{D}} \widetilde{N}_{\text{A}} - R \widetilde{N}_{\text{C}} \widetilde{N}_{\text{B}}. \end{cases}$$

- Подставляя найденное решение в систему уравнений, получим оценку фоновых событий в сигнальном регионе $N_A^{\text{jet}\to\gamma} = 54$

Оценка погрешностей

Статистическая погрешность: числа событий в каждой КО были независимо проварьированы на ±1σ для данных и всех фонов. Полученные значения были просуммированы в квадратурах.

Итоговая статистическая погрешность: δ = 28%

Центральное значение	54^{+15}_{-15}
loose'3	-4
loose'4	+10
loose'5	+8
Изоляционный зазор +0.5 ГэВ	±0
Изоляционный зазор -0.9 ГэВ	+1

Систематическая погрешность:

- варьирование изоляционного промежутка и использование альтернативных loose' (19%)
- погрешность от различных МК генераторов и моделей партонных ливней (0%)

Параметры утечки	Различные МК генераторы и модели партонных ливней		
	MadGraph+Pythia8, Sherpa 2.2	MadGraph+Herwig7, MadGraph+Pythia8	δ
c_B	0.00645 ± 0.00018	0.0039 ± 0.0006	40%
c_C	0.0091 ± 0.0002	0.0093 ± 0.0008	2%
c_D	0.00011 ± 0.00002	0.00015 ± 0.00008	27%
$jet \to \gamma$	54^{+15}_{-15}	54^{+15}_{-15}	0%

- погрешности на эффективность реконструкции фотона (1.3%)
  - $\sigma^{c_B}_{iso} = \delta^{eff}_{iso} \cdot (c_B + 1)/c_B$
  - $\sigma^{c_C}_{ID} = \delta^{eff}_{ID} \cdot (c_C + 1)/c_C$
  - $\sigma^{c_D}_{iso} = \delta^{eff}_{iso} \cdot (c_B + 1)/c_B$
  - $\sigma^{c_D}_{ID} = \delta^{eff}_{ID} \cdot (c_C + 1)/c_C$

$\delta^{eff}_{iso} = 0.013$
$\delta^{eff}_{ID} = 0.013$

$N_A^{jet \to \gamma} = 54 \pm 15$(стат.) ± 10(сист.)

МК наборы Zj и Multijet предсказывают 17 ± 12 событий

Итоговая систематическая погрешность: δ = 19%

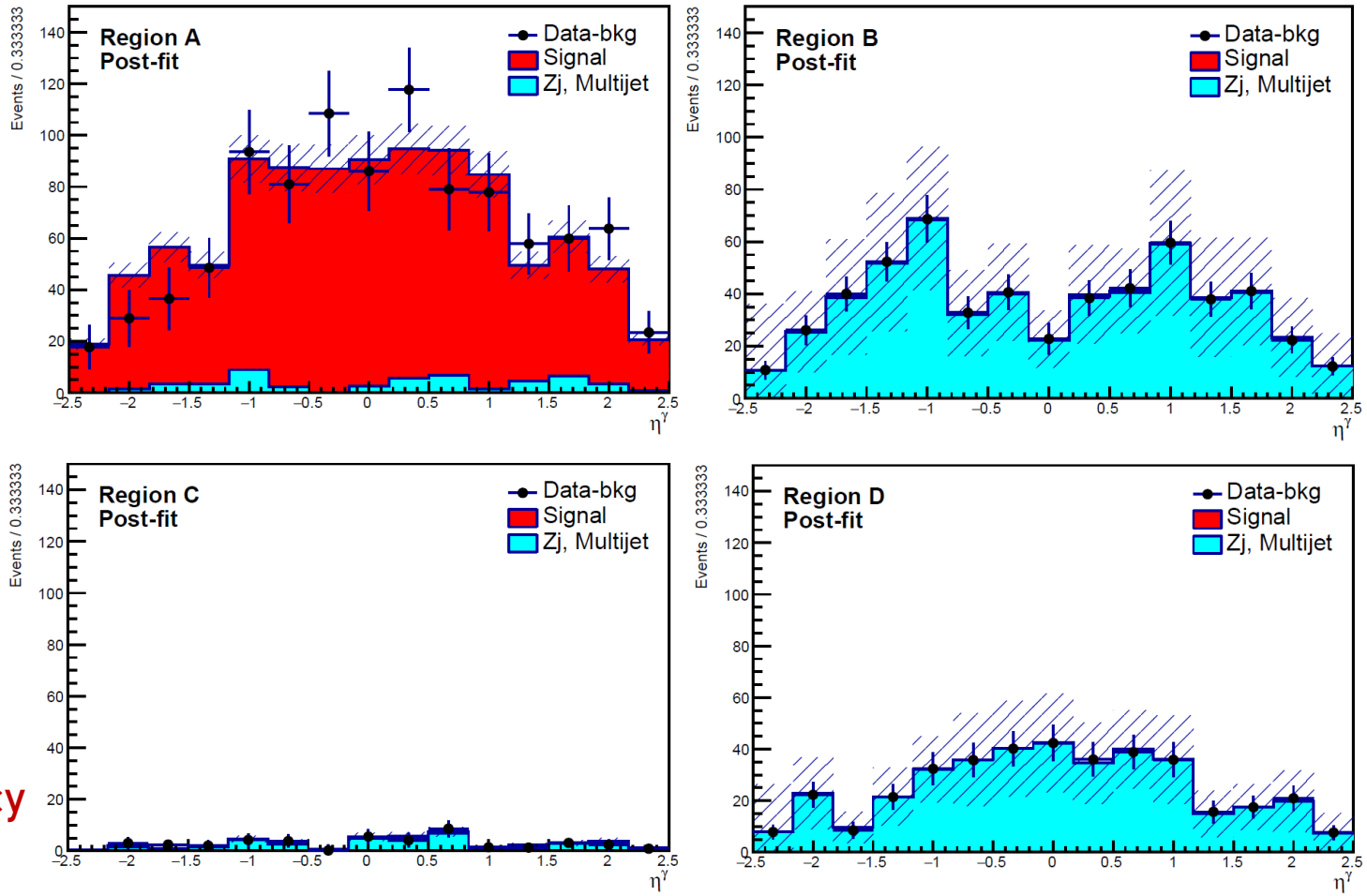
Метод максимального правдоподобия

- Альтернативный способ оценки фона
- Модель создана на основе пакета TRooFit
- Требование R фактор = 1

Преимущества модели:

- учёт биннига внутри регионов
- возможность использования значения корреляционного фактора, полученного на основе данных

Идея: фитирование формы сигнала и исследуемого фона данными за вычетом фона, который не относится к исследуемому процессу



- Оценка событий методом максимального правдоподобия в сигнальной области A составила $N_A^{jet \to \gamma} = 51 \pm 6$, что совпадает с оценкой, полученной стандартным ABCD-методом

Заключение

В соответствии с поставленной задачей в результате данного анализа:

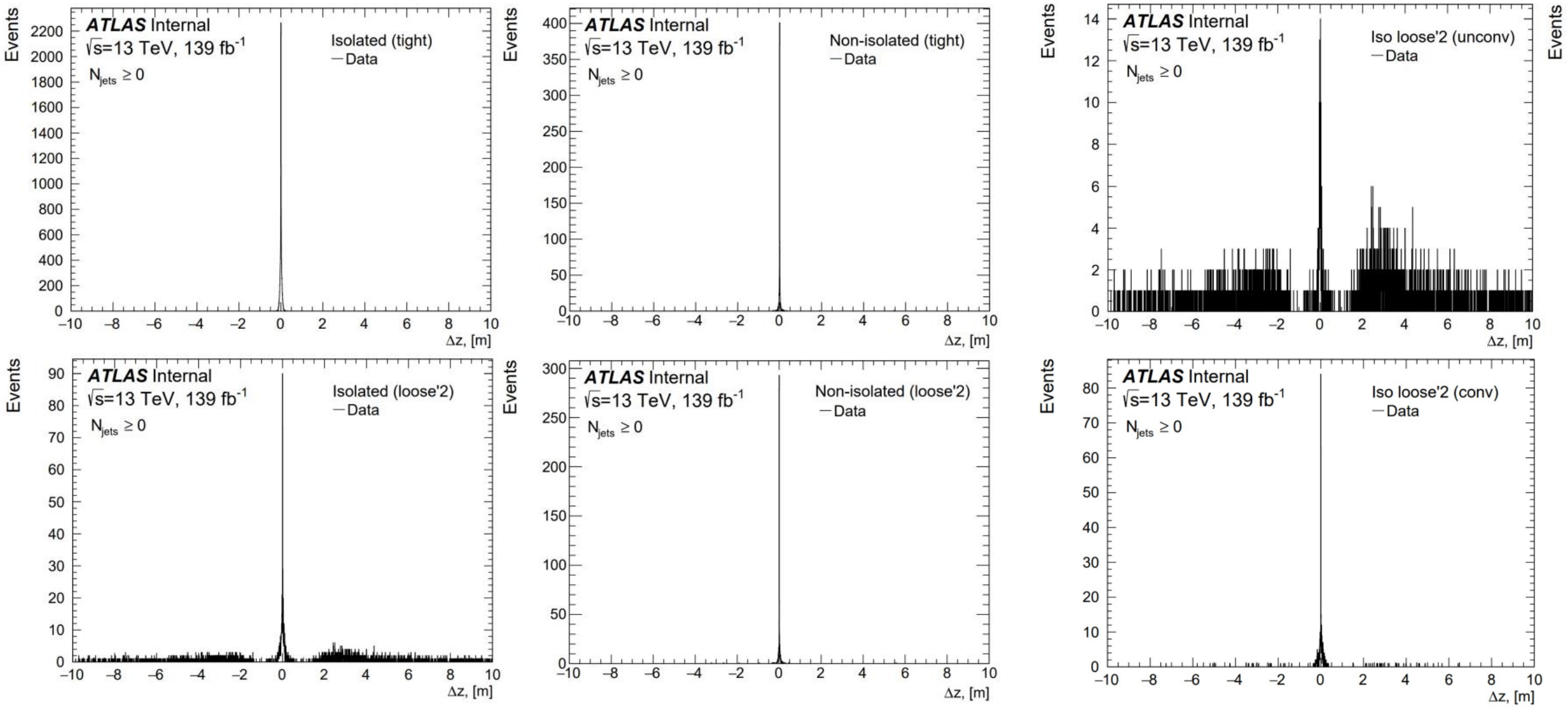
- выбран наиболее оптимальный отбор на координатную переменную Δz для подавления фона, обусловленного конфигурацией пучка
- оптимизированы регионы двумерного метода боковых интервалов для исследуемого процесса
- получена оценка центрального значения фоновых событий jet→γ в сигнальной области, а также оценены статистические и систематические погрешности, в результате получено $N_A^{jet \to \gamma} = 54 \pm 15(\text{стат.}) \pm 10(\text{сист.})$
- начата работа над методом максимального правдоподобия и получена первичная оценка числа фоновых событий, которая составляет $N_A^{jet \to \gamma} = 51 \pm 6$, что совпадает в пределах погрешностей со значением, полученным ABCD-методом

В дальнейшем планируется:

1) переоптимизировать регионы ABCD-метода с учетом отборов по новым переменным
2) продолжить работу над методом максимального правдоподобия
3) получить оценку фона с помощью Матричного Метода

BACK-UP

Фон от столкновения пучков частиц



Оптимизация ABCD-метода

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

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

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

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

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

- ws3 - ширина электромагнитного ливня с использованием трёх стриповых (первых слоёв ЭМ калориметра) слоёв вокруг стрипового слоя с максимальной энергией
- Fside - энергия вне трёх стриповых слоёв, но внутри семи слоёв
- ΔE - разница энергий стриповых слоёв, где в одном слое выделилась вторая по величине энергия, и слоя, где выделилась наименьшая энергия
- Eratio - отношение разности энергий, ассоциированных с наиболее высоким и вторым по величине выделением энергии к сумме этих энергий;
- wtot - полная поперечная ширина ливня

Uc = 29.45 ГэВ, Mc = 11.45 ГэВ

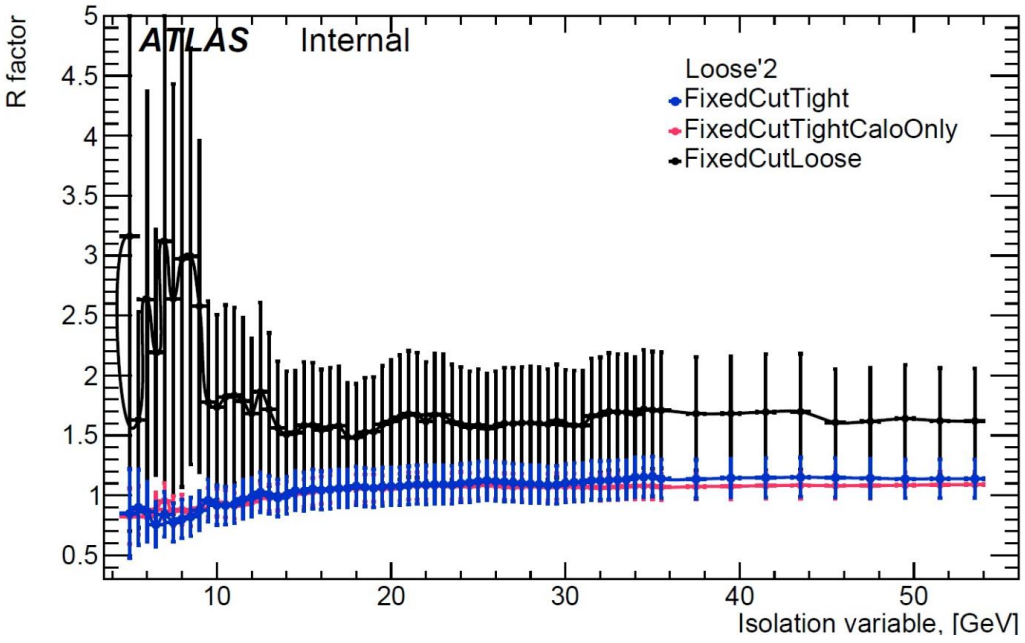
FixedCutTight	loose'2	loose'3	loose'4	loose'5
MC	1.09 ± 0.16	1.17 ± 0.16	1.24 ± 0.15	1.41 ± 0.17
Data-driven	2.0 ± 0.5	1.8 ± 0.5	1.7 ± 0.4	1.8 ± 0.4
FixedCutTightCaloOnly	loose'2	loose'3	loose'4	loose'5
MC	1.08 ± 0.11	1.16 ± 0.11	1.24 ± 0.11	1.40 ± 0.12
Data-driven	1.30 ± 0.12	1.33 ± 0.11	1.27 ± 0.10	1.29 ± 0.10
FixedCutLoose	loose'2	loose'3	loose'4	loose'5
MC	1.6 ± 0.3	1.6 ± 0.3	1.6 ± 0.3	2.0 ± 0.4
Data-driven	2.6 ± 1.0	2.1 ± 0.8	2.2 ± 0.8	2.05 ± 0.7

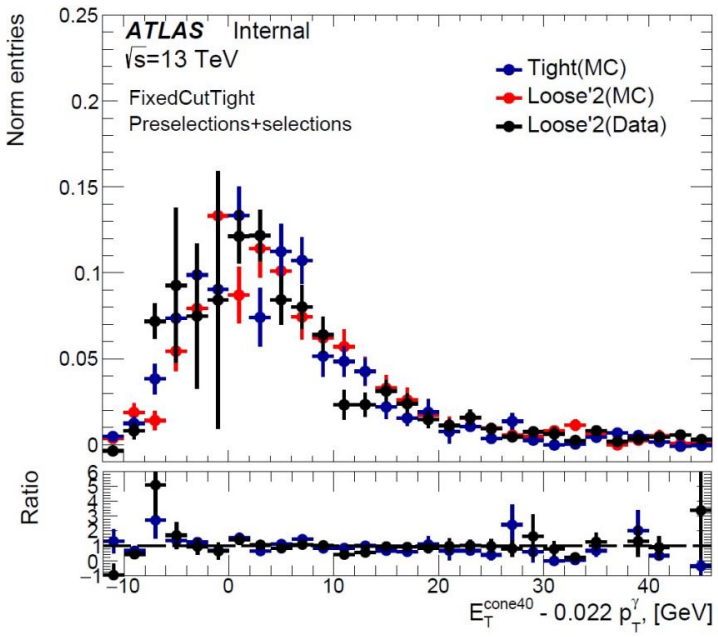
Оптимизация ABCD-метода

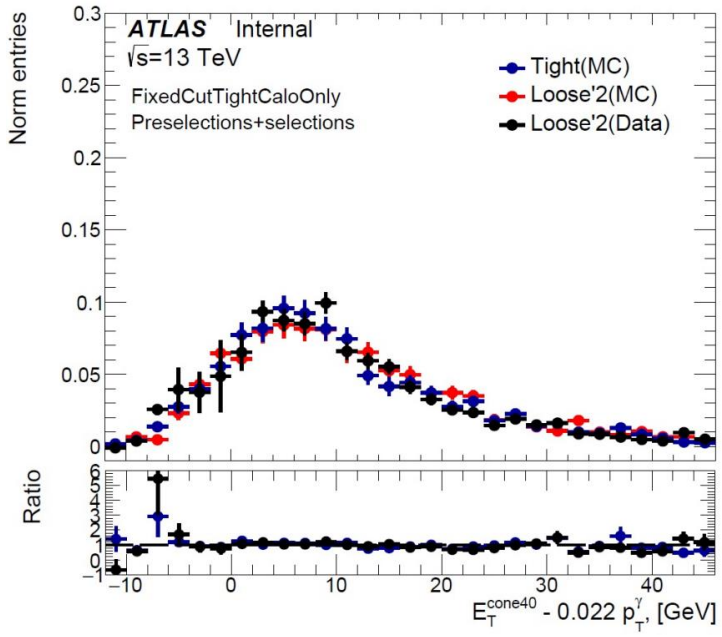
FixedCutTight, ограничение = 25.45 GeV				
MC				
	loose'2	**loose'3**	**loose'4**	**loose'5**
R-factor	1.12 ± 0.16	1.21 ± 0.17	1.28 ± 0.16	1.44 ± 0.18
Data-driven				
Cut	**loose'2**	**loose'3**	**loose'4**	**loose'5**
6.95	1.1 ± 0.4	1.2 ± 0.4	1.2 ± 0.4	1.2 ± 0.4
7.45	1.2 ± 0.3	1.1 ± 0.3	1.1 ± 0.3	1.1 ± 0.3
7.95	1.2 ± 0.3	1.2 ± 0.3	1.2 ± 0.3	1.2 ± 0.3
8.45	1.1 ± 0.3	1.1 ± 0.3	1.1 ± 0.3	1.1 ± 0.3
8.75	1.0 ± 0.3	1.0 ± 0.3	1.0 ± 0.3	1.1 ± 0.3
8.95	1.1 ± 0.3	1.1 ± 0.3	1.1 ± 0.3	1.1 ± 0.3
9.45	1.3 ± 0.3	1.3 ± 0.3	1.3 ± 0.3	1.3 ± 0.3

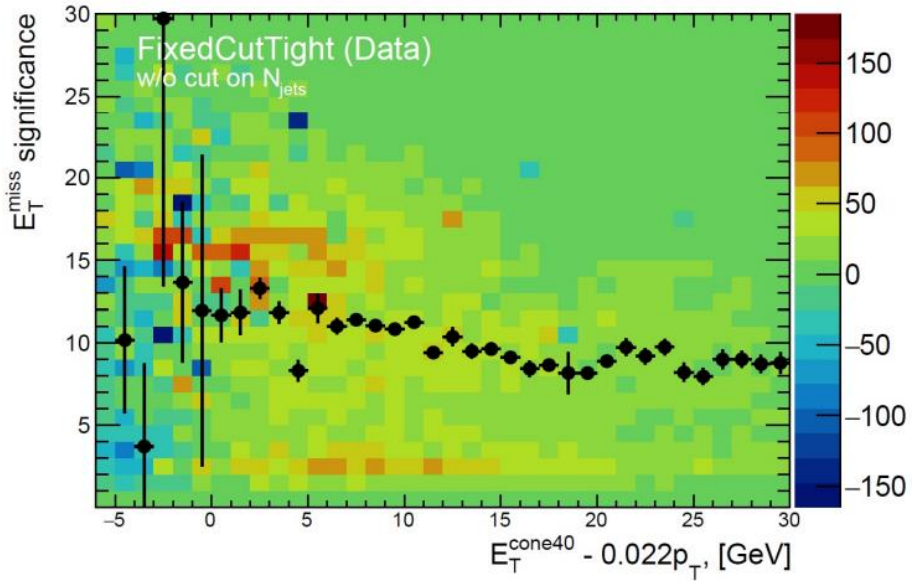
FixedCutTight (инверсия), (ограничение = 25.45 GeV)				
MC				
	loose'2	**loose'3**	**loose'4**	**loose'5**
R-factor	1.07 ± 0.13	1.23 ± 0.14	1.34 ± 0.14	1.62 ± 0.17
Data-driven				
Cut	**loose'2**	**loose'3**	**loose'4**	**loose'5**
9.95	1.18 ± 0.12	1.23 ± 0.11	1.22 ± 0.10	1.26 ± 0.10
10.45	1.12 ± 0.11	1.16 ± 0.11	1.16 ± 0.10	1.22 ± 0.10
10.95	1.20 ± 0.12	1.25 ± 0.11	1.24 ± 0.11	1.30 ± 0.11
11.45	1.26 ± 0.13	1.32 ± 0.12	1.28 ± 0.11	1.33 ± 0.11

Оптимизация ABCD-метода









FixedCutTightCaloOnly, без ограничения на U_{cut}				
МК				
	loose'2	**loose'3**	**loose'4**	**loose'5**
R-factor	1.10 ± 0.11	1.19 ± 0.11	1.27 ± 0.11	1.45 ± 0.12

Данные				
M_{cut}	**loose'2**	**loose'3**	**loose'4**	**loose'5**
10.45	1.21 ± 0.10	1.25 ± 0.10	1.21 ± 0.09	1.23 ± 0.09

Метод максимального правдоподобия

Основное предположение ABCD-метода: $\boxed{\widetilde{N}_A = \widetilde{m}\widetilde{N}_B, \widetilde{N}_C = \widetilde{m}\widetilde{N}_D}$

Количество собыйти в данных по регионам: $\mathbf{data} = \{N_A,\ N_B,\ N_C,\ N_D\}$

Функция правдоподобия:

$$L(\mathbf{data}|\widetilde{N_B}, \widetilde{N_D}, \widetilde{m}) = (N_A|\widetilde{m}\widetilde{N}_B)Pois(N_B|\widetilde{N}_B)Pois(N_C|\widetilde{m}\widetilde{N}_D)Pois(N_D|\widetilde{N}_D)$$

$$L(\mathbf{data}|\widetilde{N_B}, \widetilde{N_D}, \widetilde{m}) = \frac{e^{-\widetilde{N}_{tot}} \cdot (\widetilde{m}\widetilde{N}_B)^{N_A} \cdot (\widetilde{N}_B)^{N_B} \cdot (\widetilde{m}\widetilde{N}_D)^{N_C} \cdot (\widetilde{N}_D)^{N_D}}{N_A! \cdot N_B! \cdot N_C! \cdot N_D!}$$

$$\widetilde{N}_{tot} = \widetilde{m}\widetilde{N}_B + \widetilde{N}_B + \widetilde{m}\widetilde{N}_D + \widetilde{N}_D$$

Дифференцирование по свободным параметрам:

$$\begin{cases} \dfrac{\partial L}{\partial \widetilde{m}} = -\widetilde{N}_B - \widetilde{N}_D + \dfrac{N_A}{\widetilde{m}} + \dfrac{N_C}{\widetilde{m}} = 0, \\ \dfrac{\partial L}{\partial \widetilde{N}_B} = -\widetilde{m} - 1 + \dfrac{N_A}{\widetilde{N}_B} + \dfrac{N_B}{\widetilde{N}_B} = 0, \\ \dfrac{\partial L}{\partial \widetilde{N}_D} = -\widetilde{m} - 1 + \dfrac{N_C}{\widetilde{N}_D} + \dfrac{N_D}{\widetilde{N}_D} = 0. \end{cases}$$

$$\Rightarrow$$

$$\widetilde{m} = \frac{N_A + N_C}{N_B + N_D}$$

$$\widetilde{N}_B = \frac{(N_A + N_B)(N_B + N_D)}{N_A + N_B + N_C + N_D}$$

$$\widetilde{N}_D = \frac{(N_C + N_D)(N_B + N_D)}{N_A + N_B + N_C + N_D}$$