Оценка фона, обусловленного неверной идентификацией адронной струи как фотона 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 2022

Описание процесса и мотивация

Исследуемый процесс: $Z(vv)\gamma$ Фоновые процессы: $W(\rightarrow lv)\gamma$, tt γ , $e\rightarrow\gamma$, γ +jet, jet $\rightarrow\gamma$, $Z(\rightarrow ll)\gamma$ (14%) Мотивация: исследуемый процесс высокочувствителен к отклонениям от Стандартной Модели <u>Цель анализа:</u> оценить число фоновых событий, обусловленных неверной идентификацией адронной струи как фотона (jet $\rightarrow\gamma$)

(c)

(d)

Задачи анализа:

- оценить число фоновых событий ABCD-методом
- разработать метод оценки, основанный на методе максимального правдоподобия, далее произвести оценку и сравнить результат с ABCD-методом
- оценить статистические и систематические погрешности числа фоновых событий для обоих методов

Используемые отборы

Сигнал: электрослабое рождение Z(vv)ү и КХД Z(vv)ү

Предотборы

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

Переменная	Ограничение
Значимость $E_{\rm T}^{\rm miss}$	> 11
$ \Delta \phi(E_{ m T}^{ m miss},\gamma) $	> 0.6
$ \Delta \phi(E_{ m T}^{ m miss},j_1) $	> 0.4

Отборы

Регион с такими предотборами и отборами + изоляция сигнальный регион

Отбор фотонов

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

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

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

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

Результат оптимизации R фактора



R- фактор	loose'2	loose'3	loose'4	loose'5			
	FixedCutTight, без ограничения						
MK	1.05 ± 0.15	1.14 ± 0.15	1.19 ± 0.14	1.39 ± 0.17			
Данные	1.4 ± 0.3	1.3 ± 0.3	1.3 ± 0.3	1.3 ± 0.3			
]	FixedCutTigh	t, ограничени	$1\mathrm{e}=25.45~\mathrm{Gm}^2$	B			
MK	1.06 ± 0.15	1.15 ± 0.16	1.21 ± 0.15	1.40 ± 0.17			
Данные	1.01 ± 0.18	1.02 ± 0.18	1.01 ± 0.18	1.01 ± 0.17			
FixedCut	Гight, трек. и	нверсия + ог	раничение =	25.45 ГэВ			
MK	1.01 ± 0.12	1.15 ± 0.12	1.29 ± 0.13	1.58 ± 0.16			
Данные	1.07 ± 0.10	1.13 ± 0.10	1.15 ± 0.10	1.15 ± 0.10			
FixedCutTightCaloOny							
MK	1.06 ± 0.10	1.14 ± 0.11	1.22 ± 0.10	1.40 ± 0.12			
Данные	1.07 ± 0.10	1.13 ± 0.10	1.15 ± 0.10	1.15 ± 0.10			





Выбрана наиболее оптимальная изоляция FixedCutTightCaloOnly

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

Число событий в регионах определяется как:

 $N_A = N_A^{\text{sig}} + N_A^{\text{bkg}} + N_A^{\text{jet} \to \gamma};$

 $N_B = c_{\rm B} N_A^{\rm sig} + N_B^{\rm bkg} + N_B^{\rm jet \to \gamma};$ $N_C = c_{\rm C} N_A^{\rm sig} + N_C^{\rm bkg} + N_C^{\rm jet \to \gamma};$

 $N_D = c_D N_A^{\text{sig}} + N_D^{\text{bkg}} + N_D^{\text{jet} \to \gamma};$

Параметры утечки сигнала в КО

		c_B	c_C	c_D
>	Значение	0.0713 ± 0.0002	0.00879 ± 0.00007	0.00070 ± 0.00002

С учётом R фактора на данных, получим уравнение:

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

Количества данных и фоновых событий в каждом регионе

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

Решение уравнения имеет вид:
$$N_{\rm A}^{\rm sig} = \frac{b - \sqrt{b^2 - 4ac}}{2a}$$
, где
$$\begin{cases} a = c_D - Rc_B c_C; \\ b = \widetilde{N}_{\rm D} + c_D \widetilde{N}_{\rm A} - R(c_B \widetilde{N}_{\rm C} + c_C \widetilde{N}_{\rm B}) \\ c = \widetilde{N}_{\rm D} \widetilde{N}_{\rm A} - R \widetilde{N}_{\rm C} \widetilde{N}_{\rm B}. \end{cases}$$

 $c_i = \frac{N_i^{\rm sig}}{N_A^{\rm sig}} \qquad \square$

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

Подставляя найденное решение в систему уравнений, получим <mark>центральное значение событий в</mark> сигнальном регионе с учетом R фактора на данных: N_A^{jet → γ} = 1960

Оценка погрешностей

Статистическая погрешность: числа событий в каждой контрольной области были независимо

проварьированы на ±1σ для данных и всех фонов. Полученные значения просуммированы в квадратурах. Итоговая статистическая погрешность: δ = 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%)

•
$$\sigma_{
m iso}^{
m c_B}=\delta_{
m iso}^{
m eff}\cdot(c_B+1)/c_B$$

•
$$\sigma_{\text{ID}}^{c_{\text{C}}} = \delta_{\text{ID}}^{\text{eff}} \cdot (c_{C} + 1)/c_{C}$$

• $\sigma_{\text{iso}}^{c_{\text{D}}} = \delta_{\text{iso}}^{\text{eff}} \cdot (c_{B} + 1)/c_{B}$
 $\begin{cases} \delta^{\text{eff}}_{\text{iso}} = 0.013 \\ \delta^{\text{eff}}_{\text{ID}} = 0.013 \end{cases}$

•
$$\sigma_{\text{iso}}^{\text{c}_{\text{D}}} = \delta_{\text{iso}}^{\text{eff}} \cdot (c_C + 1)/c_B$$
 (

Central value	$1960{\pm}83$
loose'3	-334
$\mathbf{loose'4}$	-397
$\mathbf{loose'5}$	-472
Isolation gap $+0.15$ GeV	+33
Isolation gap -0.15 GeV	-22

$$N_A^{jet \to \gamma}$$
 = 1960 ± 83(стат.) ± 510(сист.)

МК Zj и Multijet предсказывают 1560 ± 1243 событий

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_{Fji} - параметр, на который умножается оцениваемый фон в каждом бине для регионов А, В, С, D
 f_{Nj} - параметр, на который умножается МК сигнал для регионов А, В, С, D
 \u03c6/b_{ji}, \u03c6/s_{ji} и \u03c6/\u03c6_{ji}, - количество событий в МК фонах, МК сигнале и оцениваемом фоне в каждом бине для регионов А, В, С, D соответственно

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

 $\frac{\partial \ln L}{\partial f_{F_{ji}}} = 0, \frac{\partial \ln L}{f_{N_j}} = 0$ ССС помощью пакета RooFit

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

Преимущества модели:

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

<u>Оценка:</u> N $_{jet \to \gamma} = \nu_{\gamma_{Ai}} f_{F_{Ai}}$

Результаты фитирования

Фитирование производилось для двух различных переменных с разным биннигом



- Биннинг выбирался на основе значения χ2/N_{dof}
- Оценка центрального числа событий методом максимального правдоподобия в сигнальной области А для п_у составила 1882, для ф_у оценка составила 1743

Оценка погрешностей для ММП

Значения, при которых Л

Статистическая погрешность:

• Отношение функций правдоподобий

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

Статистические погрешности оценивались с помощью RooFit и составили: для _¶ 1882⁺⁷⁴₋₇₀, для _{\$\phi_{\gamma}\$} 1743⁺⁶⁹₋₆₅ δ_{stat} = 4%



Систематическая погрешность:



		1000+74
	Центральное значение (η_{γ})	1882^{+74}_{-70}
	loose'3	-401
n.	loose'4	-447
·γ	loose'5	-512
	Изоляционный зазор +0.15 GeV	-5
δ _{sist} = 28%	Изоляционный зазор -0.15 GeV	+16
	Различные генераторы	-150

Центральное значение (ϕ_{γ})	1743_{-65}^{+69}	
loose'3	-353	
loose'4	-406	Φ
loose'5	-467	Ψγ
Изоляционный зазор +0.15 GeV	-1	<u> </u>
Изоляционный зазор -0.15 GeV	-4	δ _{sist} = 28%
Различные генераторы	-155	

2.00 V(θ^k)

1.75

1.50

1.25

0.75

Итоговая оценка методом ММП:

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

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

Полученные оценки для разных переменных совпадают в пределах погрешностей со значениями, полученными стандартным ABCD-методом, а также согласуются друг с другом 9/10

Заключение

В соответствии с поставленной задачей в результате данной работы:

оптимизированы регионы, используемые в двумерном методе боковых интервалов для исследуемого процесса

оценено центральное значение фоновых событий jet→γ в сигнальном регионе ABCD методом, а также оценены статистические и систематические погрешности, в результате получено N_A^{jet → γ} = 1960 ± 83(стат.) ± 510(сист.)

работа над методом максимального правдоподобия <mark>близка к завершению. Получена оценка</mark> числа фоновых событий для различных переменных, которая составляет №_д^{jet} → ^ү

= 1882⁺⁷⁴₋₇₀ (стат.) ± 527(сист.) для п_ү и N_A^{jet → γ} = 1743⁺⁶⁹₋₆₅(стат.) ± 488(сист.) для ф_γ. Результаты совпадают в пределах погрешностей со значением, полученным ABCD-методом, а также согласуются друг с другом

В дальнейшем планируется:

- 1) завершить работу над методом максимального правдоподобия
- оптимизировать значения R фактора и получить оценку фона jet→γ с учетом адронного канала распада W(τν)

BACK-UP

Фон от столкновения пучков частиц



Фон, обусловленный конфигурацией пучка



Критерии рабочих точек и КО

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

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

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

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



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

∆Е – 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

	Различные МК генераторы и модели партонных ливней			
Параметры утечки	MadGraph+Pythia8, Sherpa 2.2 MadGraph+Herwig7, MadGraph+Pythia8			
$c_{ m B}$	0.0713 ± 0.0002	0.1000 ± 0.0011	29%	
$c_{ m C}$	0.00879 ± 0.00007	0.0092 ± 0.0003	4%	
c_{D}	0.00070 ± 0.00002	0.00099 ± 0.00010	29%	
$jet ightarrow \gamma$	1960	1785	9%	

A (CP): E_T^{cone40} - 0.022 p_T^Y < 2.45, tight B (KO): 2.45 + isogap < E_T^{cone40} - 0.022 p_T^Y, tight C (KO): E_T^{cone40} - 0.022 p_T^Y < 2.45, non-tight D (KO): 2.45 + isogap < E_T^{cone40} - 0.022 p_T^Y, non-tight

Распределение по изоляции и R (loose'2)



Оптимизация R фактора на данных (I/IV)



• Оптимизация не применима из-за корреляций и ограниченной статистики

Оптимизация R фактора на данных (II/IV)

 Изоляция: FixedCutTight, ограничение сверху 25.45 GeV

${f FixedCutTight, (upper \ cut = 25.45 \ GeV)}$							
MC							
	loose'2 loose'3 loose'4 loose'5						
R-factor	1.06 ± 0.15	1.15 ± 0.16	1.21 ± 0.15	1.40 ± 0.17			

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



Для loose'2 для данных δ = 18%

Достаточно большая погрешность

Оптимизация R фактора на данных (III/IV)



Оптимизация применима

Для loose'2 для данных δ = 11%

E^{cone40} - 0.022 p^γ_τ, [GeV]

Оптимизация R фактора на данных (IV/IV)



Применимые значения R фактора на данных

Для loose'2 для данных δ = 9%

Метод максимального правдоподобия

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

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





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