

Классификация лептонных распадов W бозона методами машинного обучения в р-р столкновениях при $\sqrt{s} = 13$ ТэВ в эксперименте ATLAS.

Научный руководитель: Солдатов Е.Ю.

Научный консультант: Пономаренко Д.Е.

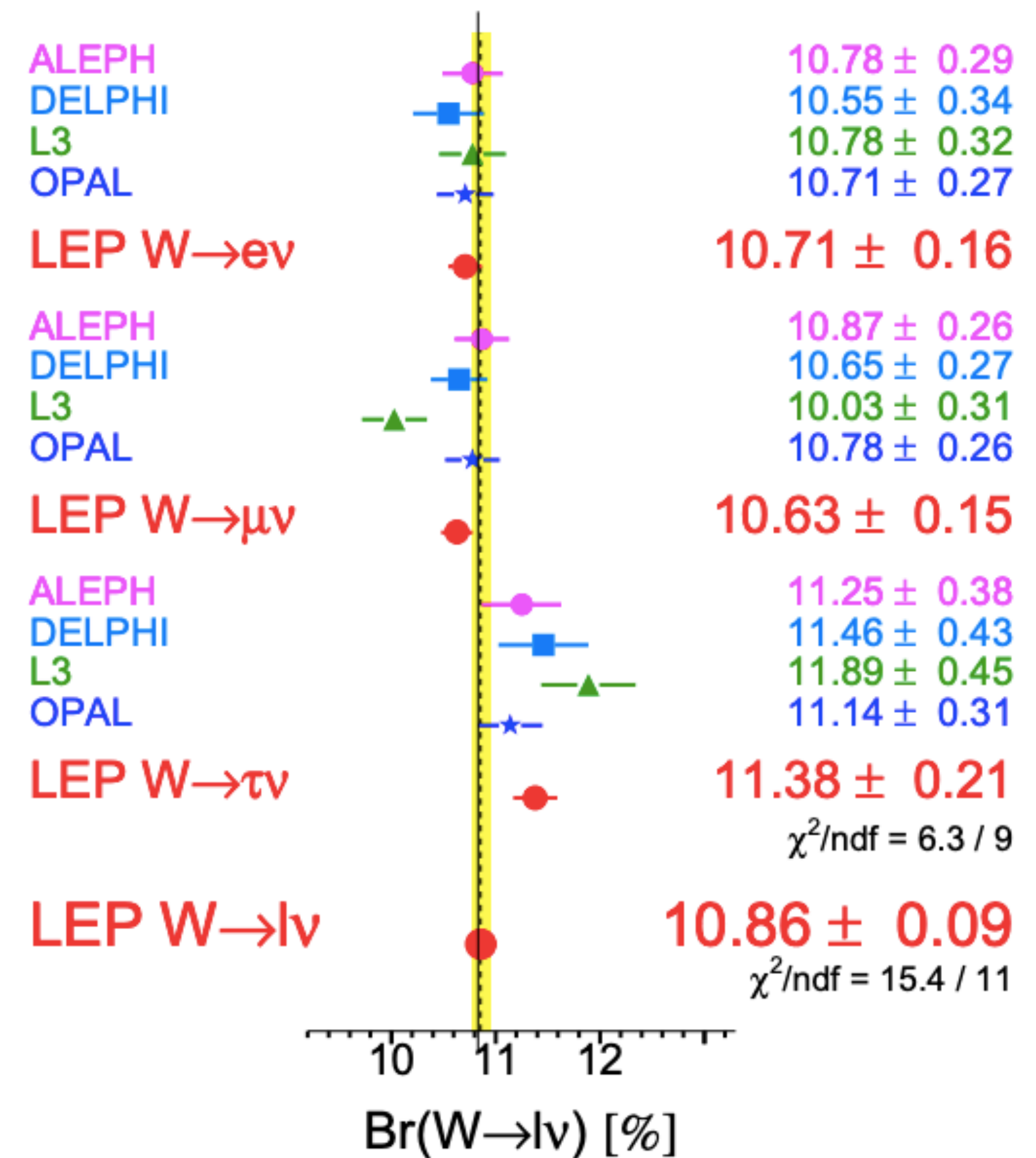
Выполнил: Толкачев Г.А.

29.06.2021

Мотивация

- ▶ В анализе данных с LEP имеется указание на возможное отклонение в отношении отношений сечения двух процессов лепленного распада W бозона $B(W \rightarrow \tau\nu)/B(W \rightarrow \mu\nu)$.
- ▶ На эксперименте ATLAS рождается большое количество W . Можно гарантировать хорошую статистику для распадов $W \rightarrow \tau\nu \rightarrow \mu\nu\nu$. Имеется больше содержание фонового процесса $W \rightarrow \mu\nu$, который имеет схожую сигнатуру.
- ▶ Цель улучшить результат измерения отношения отношений сечений с помощью переменной отклика классификатора BDT.

W Leptonic Branching Ratios



$$B(W \rightarrow \tau\nu)/B(W \rightarrow \mu\nu) = 1.070 \pm 0.026$$

Этапы работы

1

Подбор оптимального классификатора и его параметров.

2

Обучение и валидация классификатора.
Классификация данных.

3

Оптимизация разбиения отклика классификатора на интервалы.
Измерение лептонной универсальности.

Используемые данные

1

Монте-Карло данные, соответствующие условиям реальных протон-протонных столкновений эксперимента ATLAS во втором сеансе набора данных в 2015 году. С полной интегральной светимостью 3.9 fb^{-1} .

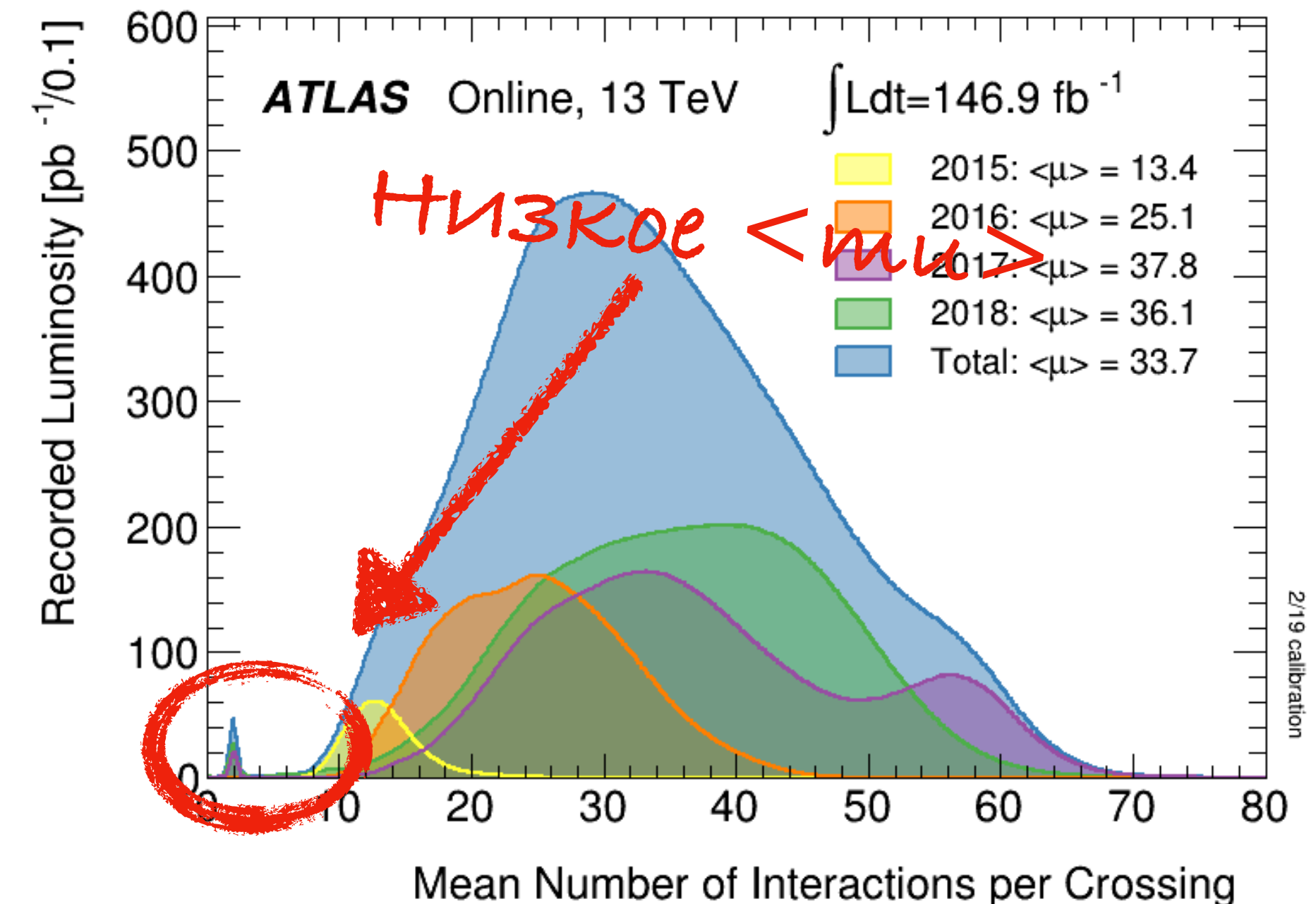
Экспериментальные данные, набранные на детектора ATLAS в 2017 и 2018 году с полной интегральной светимостью 340 pb^{-1} и низким $\langle \mu \rangle$. При столкновении протон-протонных пучков с энергией 13 ТэВ.

Монте-Карло данные, соответствующие условиям реальных протон-протонных столкновений эксперимента ATLAS во втором сеансе набора данных в 2017 и 2018 году.

2

3

$\langle \mu \rangle$ - среднее число взаимодействий при столкновении



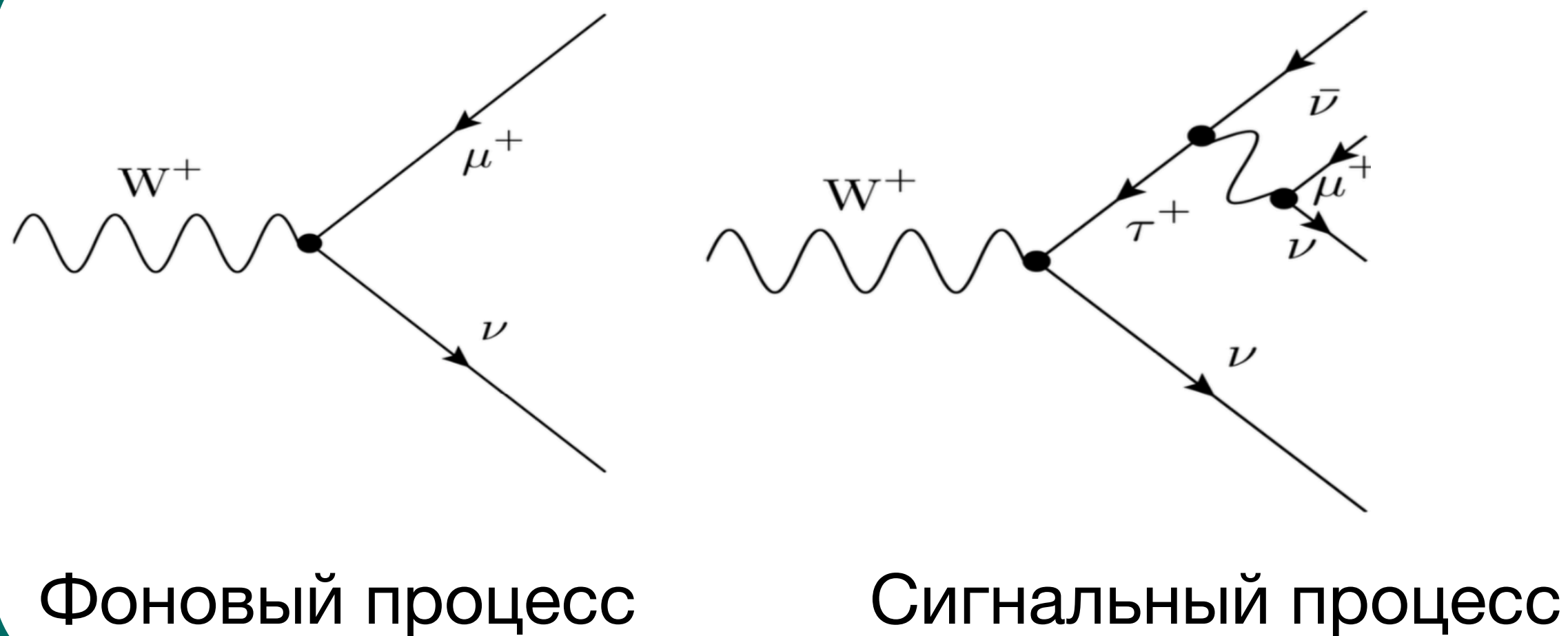
Подбор оптимального классификатора и его параметров

Критерии отбора событий для региона, который характеризуется большим содержанием сигнальных событий распадов $W \rightarrow \tau \nu$

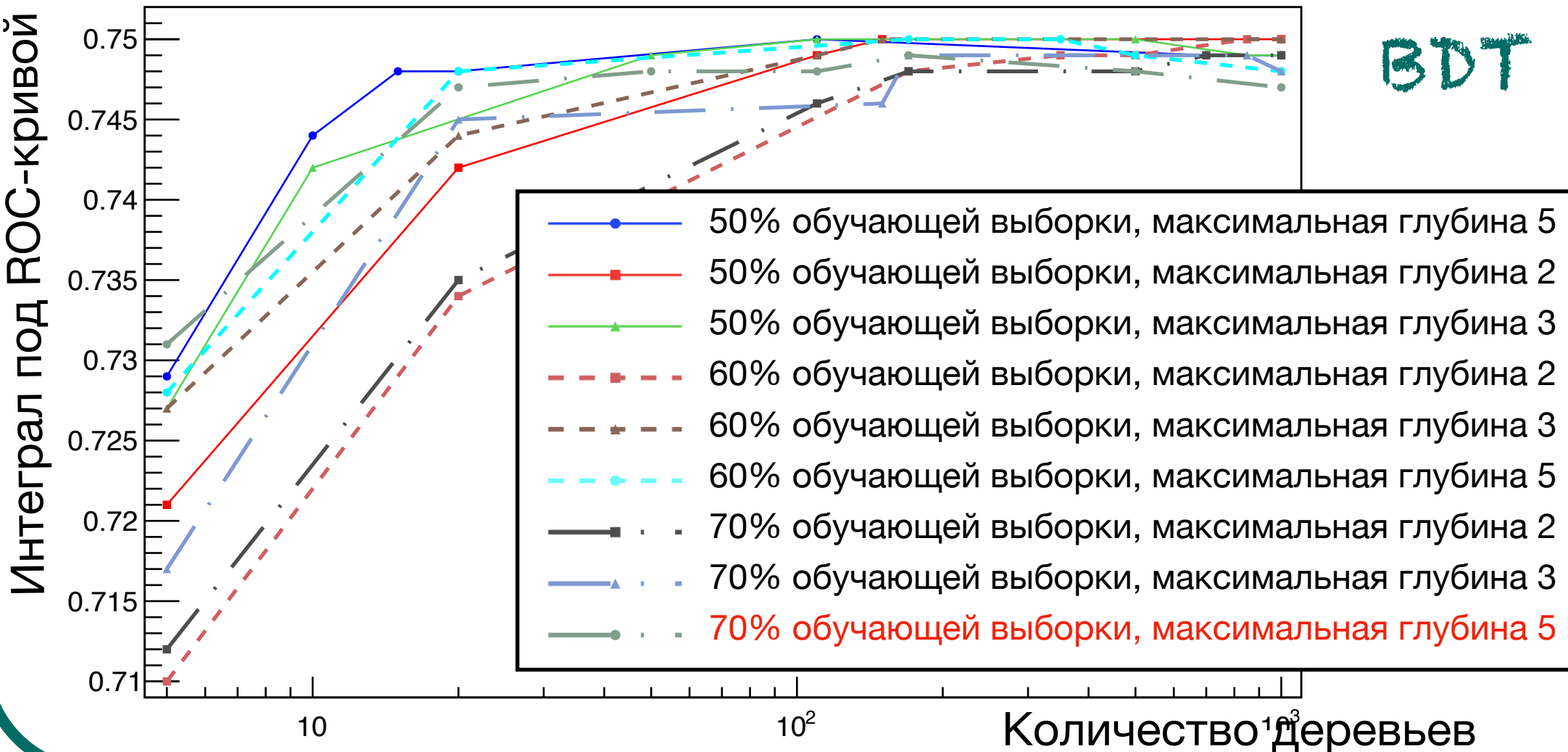
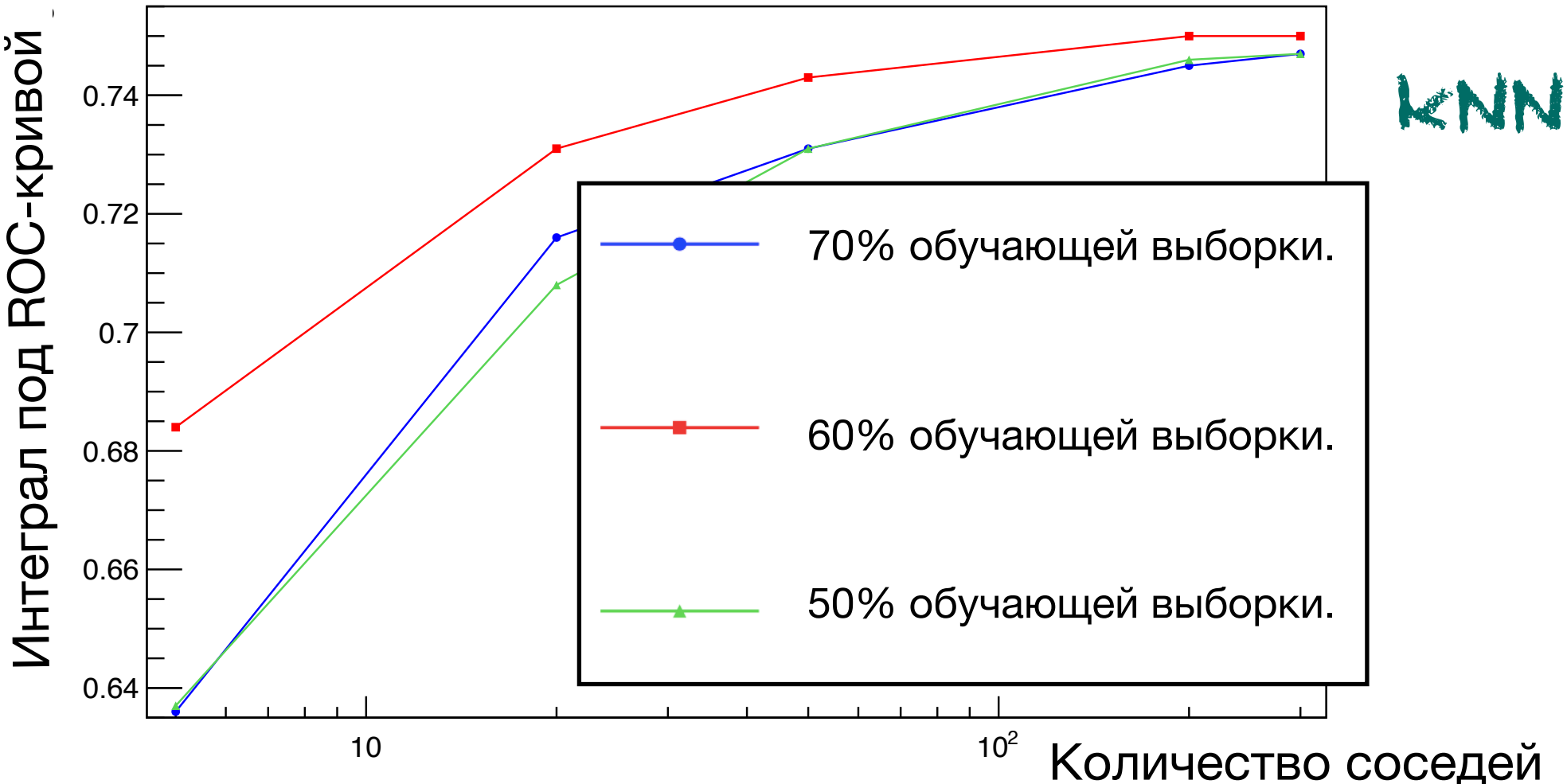
p_T	E_T^{miss}	m_T	N_{lep}	$ \eta $	b-струй	ID	ISO
$> 25 \text{ ГэВ}$	$> 25 \text{ ГэВ}$	$> 40 \text{ ГэВ}$	1	< 2.47	0	medium	tight

Для обучения классификатора были использованы переменные:
 $p_T, m_T, \eta, \phi_{miss}, \phi_{lep}, E_T^{miss}, \Delta\phi(l - E_T^{miss})$.

В обучении использовалось 66779 сигнальных событий и 623885 фоновых событий

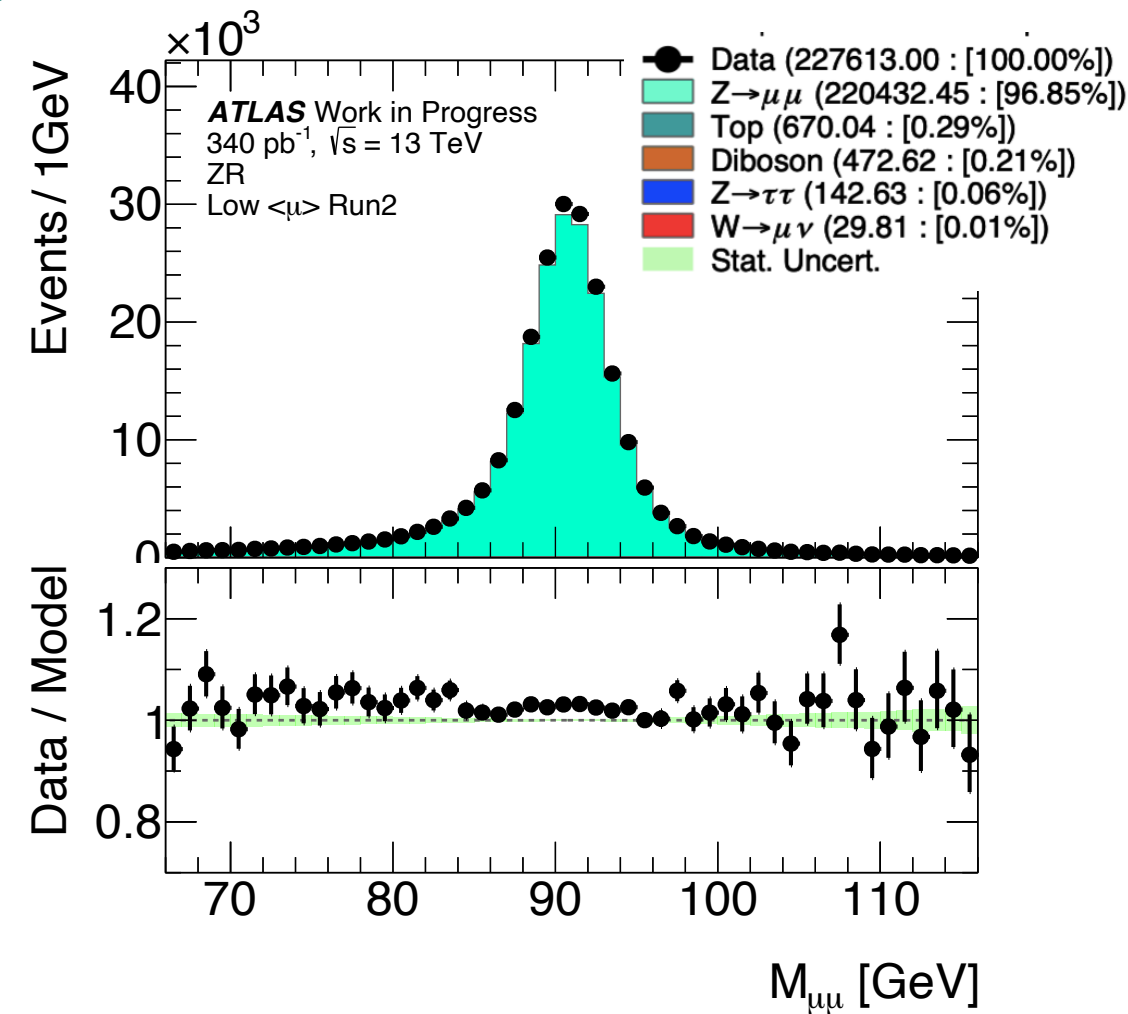


Зависимость значения интеграла под ROC-кривой от параметров классификатора



Контрольные регионы

Z регион

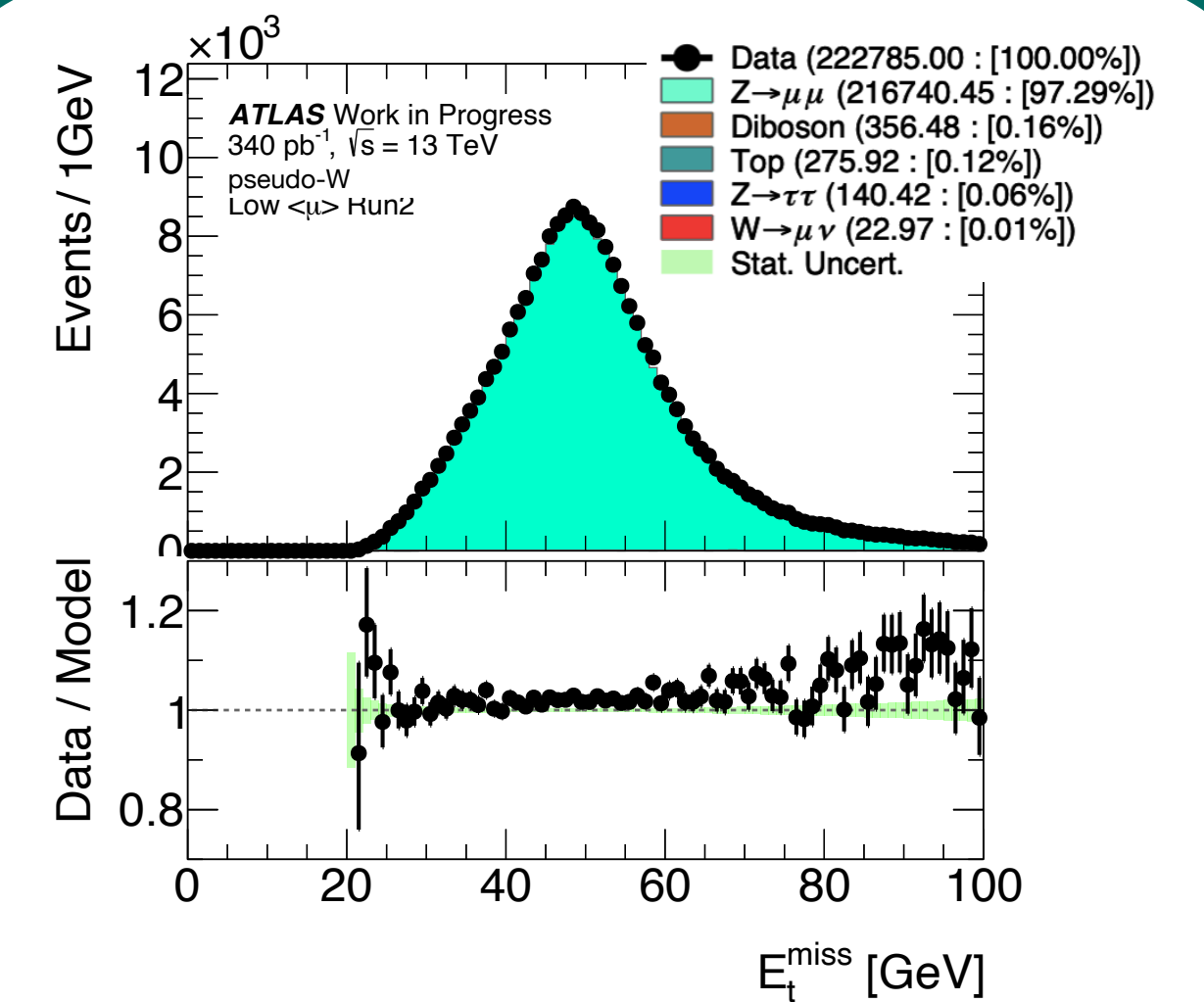


Критерии отбора для Z региона

p_T	M_{ll}	ID	topoetcone20/ p_T
> 20 ГэВ	66 ГэВ $< m < 116$ ГэВ	tight	< 0.1

- ▶ Экспериментальные данные хорошо согласуются с Монте-Карло.
- ▶ Необходим для валидации сигнального региона и модели машинного обучения.

Псевдо-W регион

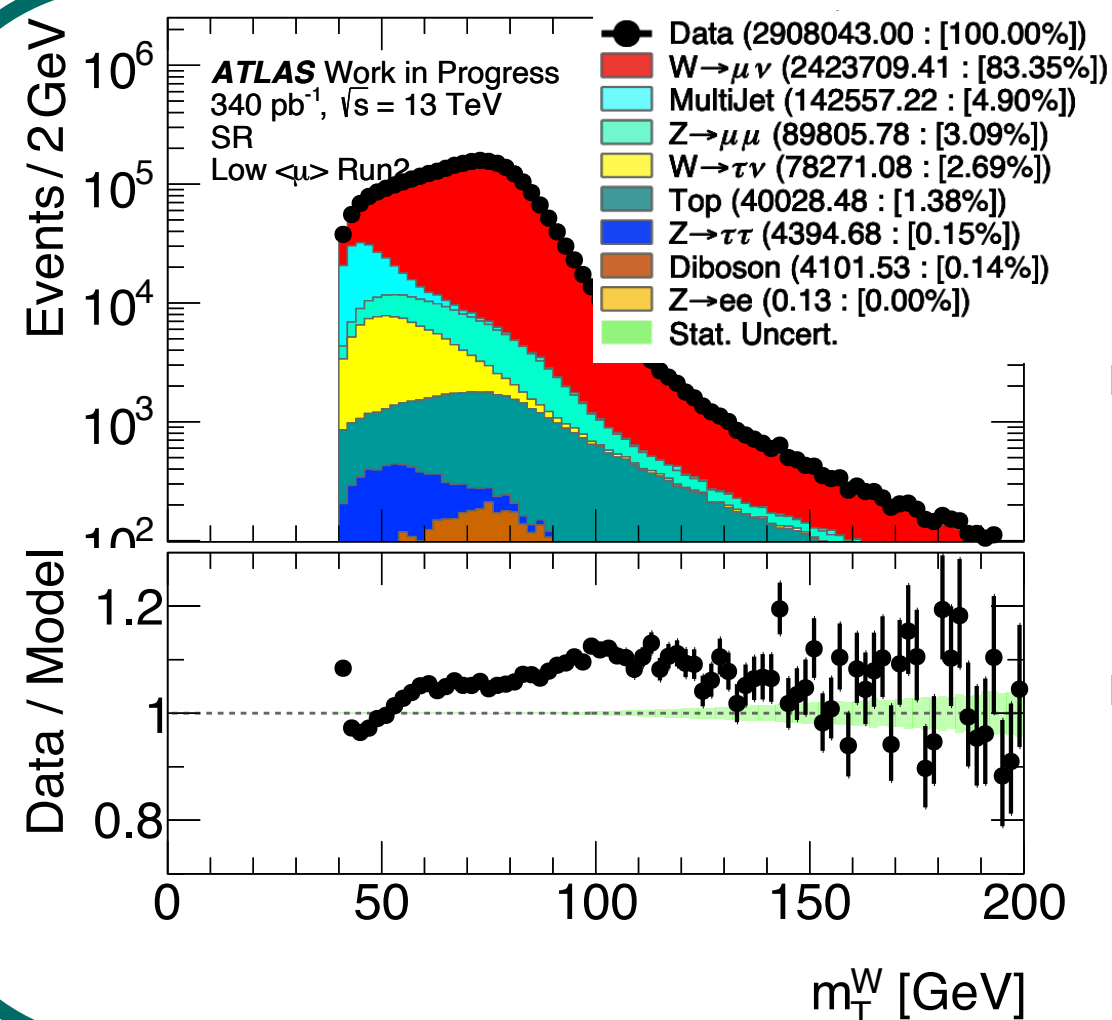


- ▶ Созданный регион на основе Z региона

$$E_{T,pseudoW}^{miss} = E_T^{miss} + rand(lep).$$

- ▶ Критерии отбора как для Z региона.
- ▶ Имитирует сигнатуру сигнального региона, сохраняя хорошее соответствие Монте-Карло и экспериментальных данных.

Сигнальный регион



Критерии отбора для сигнального региона

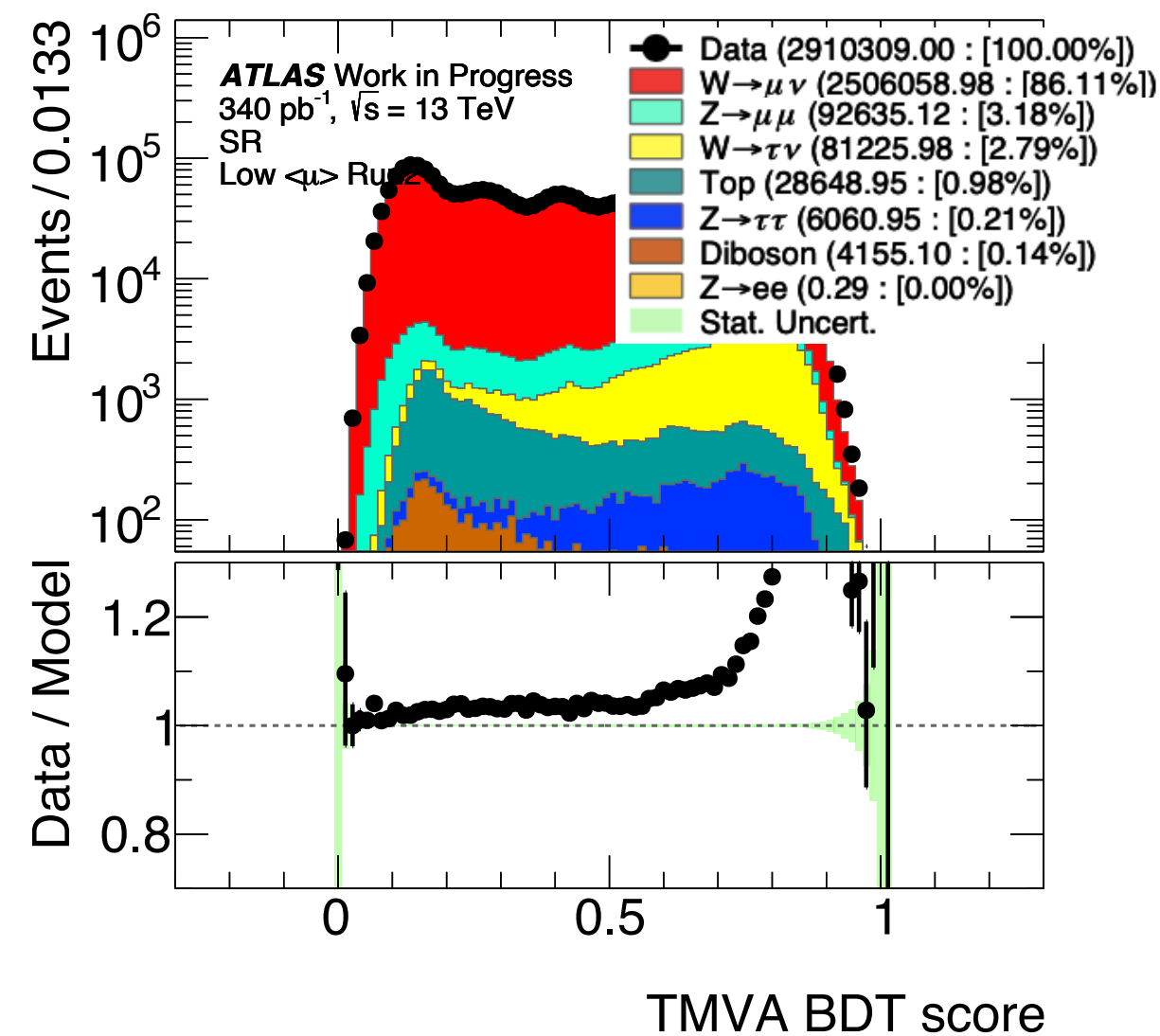
p_T	E_T^{miss}	m_T	N_{lep}	ptvarcone20/ p_T	topoetcone20/ P_T
> 20 ГэВ	> 20 ГэВ	> 40 ГэВ	1	< 0.1	< 0.1

- ▶ Характеризуется большим содержанием событий распадов $W \rightarrow \tau\nu$.
- ▶ Используется для получения отклика классификатора BDT.

$$m_T^W = \sqrt{2p_T E_T^{miss} (1 - \cos(\Delta\phi(l - E_T^{miss})))}$$

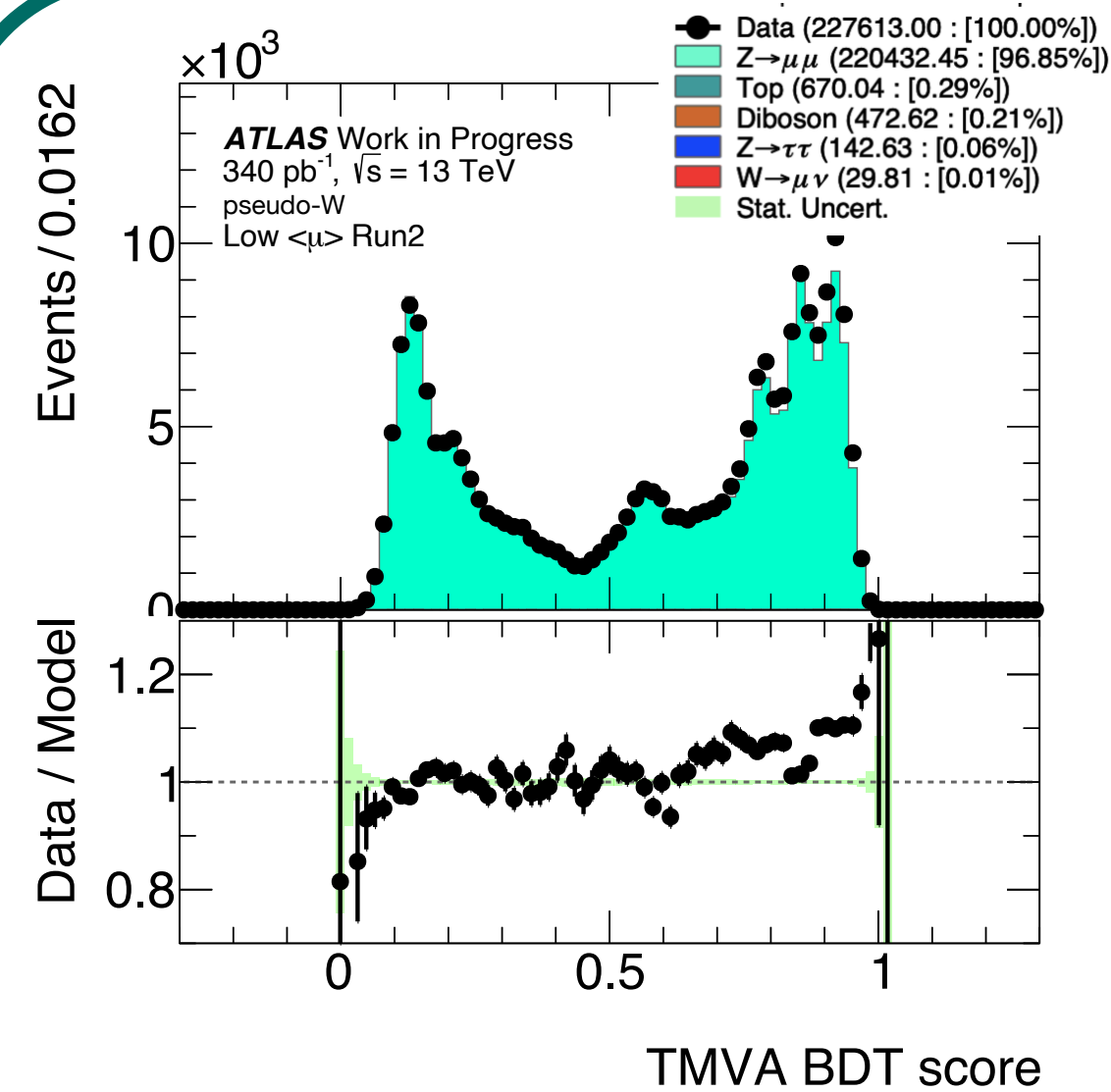
Результаты классификации событий

Классификация данных сигнального региона без учета оценки КХД фона



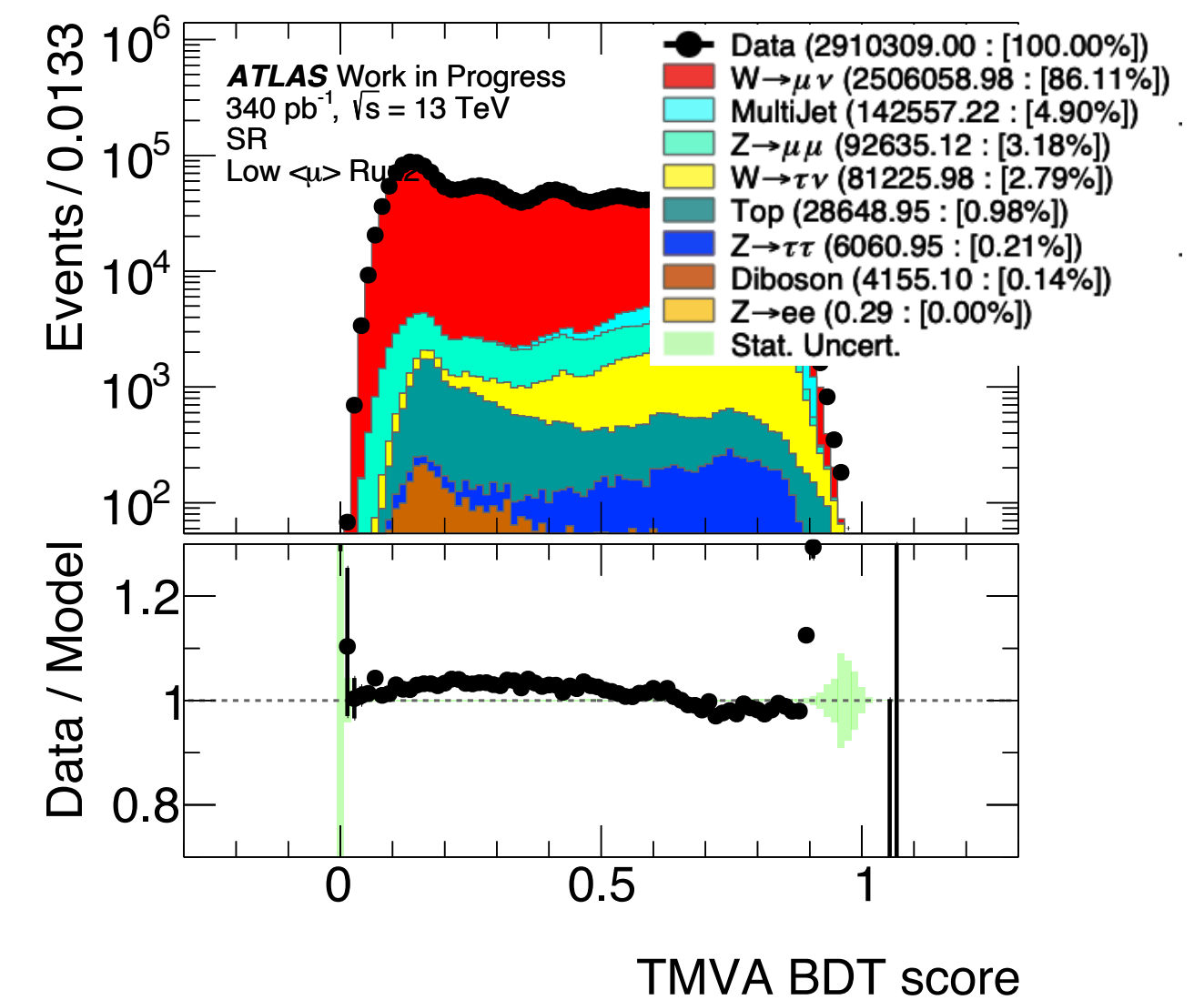
- ▶ Отклонения в области (0, 0.6) связано с плохим моделированием Монте-Карло.
- ▶ Возможной причиной отклонений в области (0.6, 1.0) является неучтенный КХД фон.
- ▶ Для подтверждения данной гипотезы необходимо провести классификацию данных псевдо-W региона.

Классификация данных псевдо-W региона



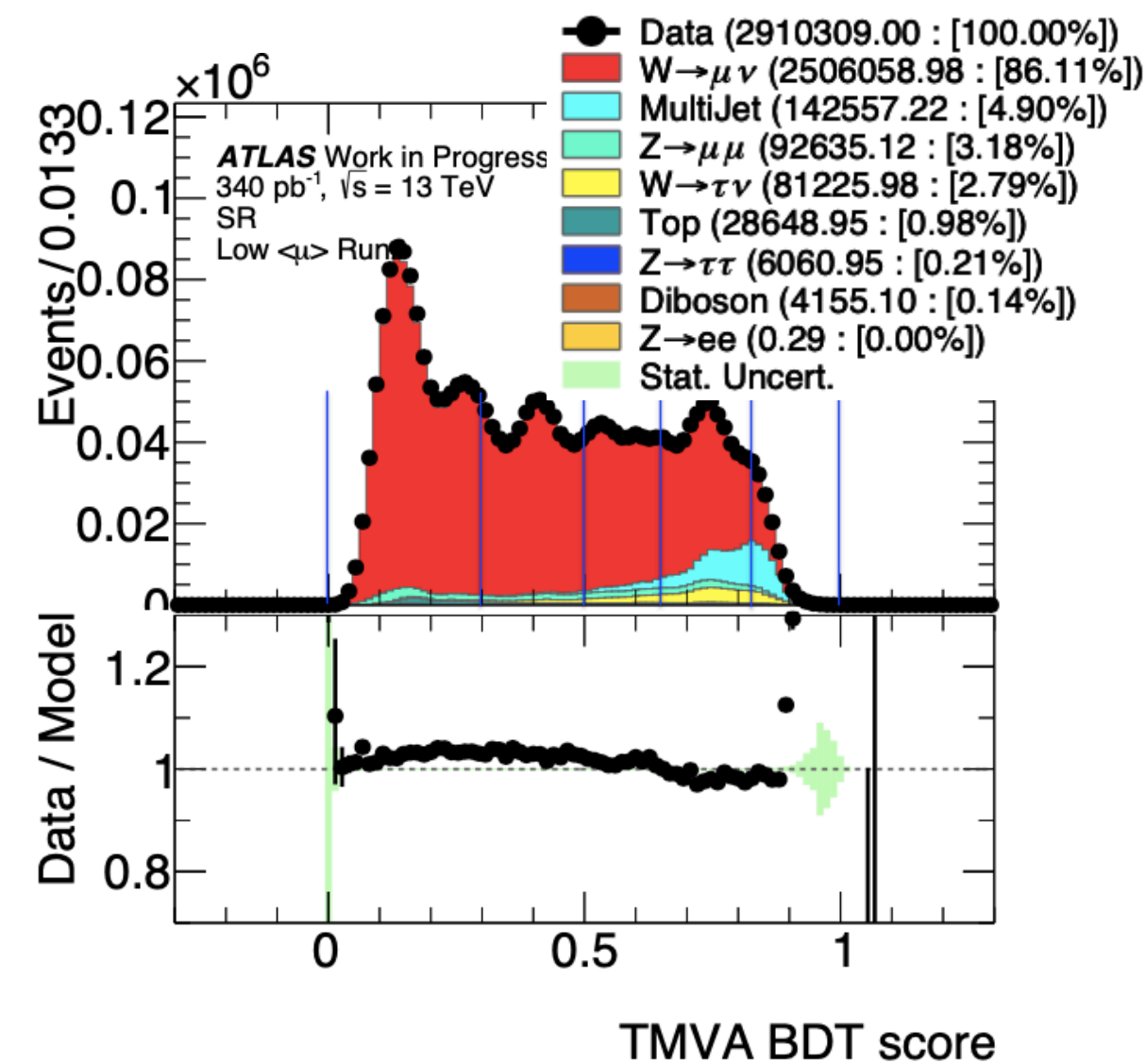
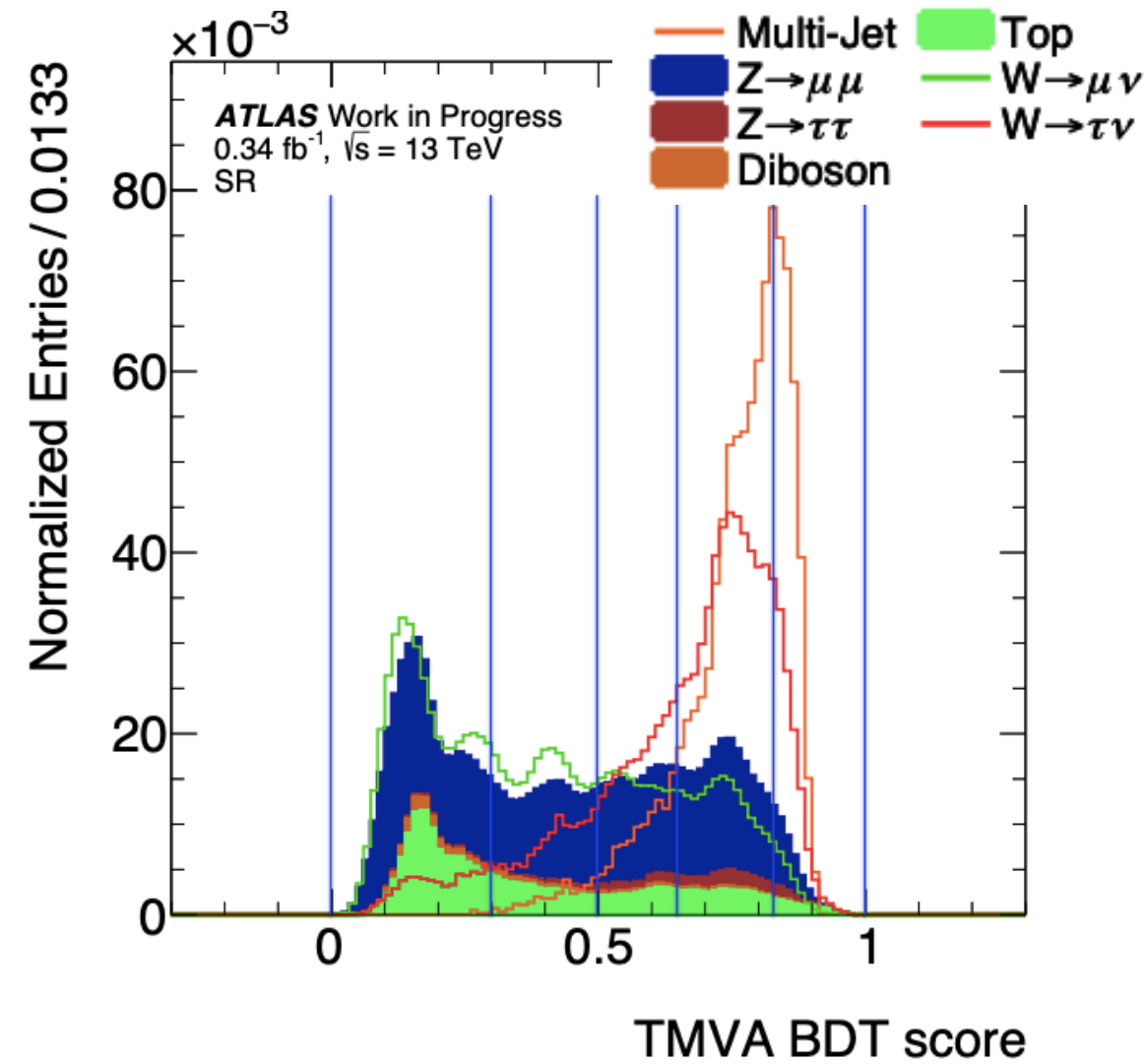
- ▶ Отклик на Монте—Карло данных хорошо согласуется с откликом на экспериментальных данных.
- ▶ Отклонение в классификации данных сигнального региона в области (0.6, 1) с большой вероятностью является КХД фон.

Классификация данных сигнального региона с учетом оценки КХД фона



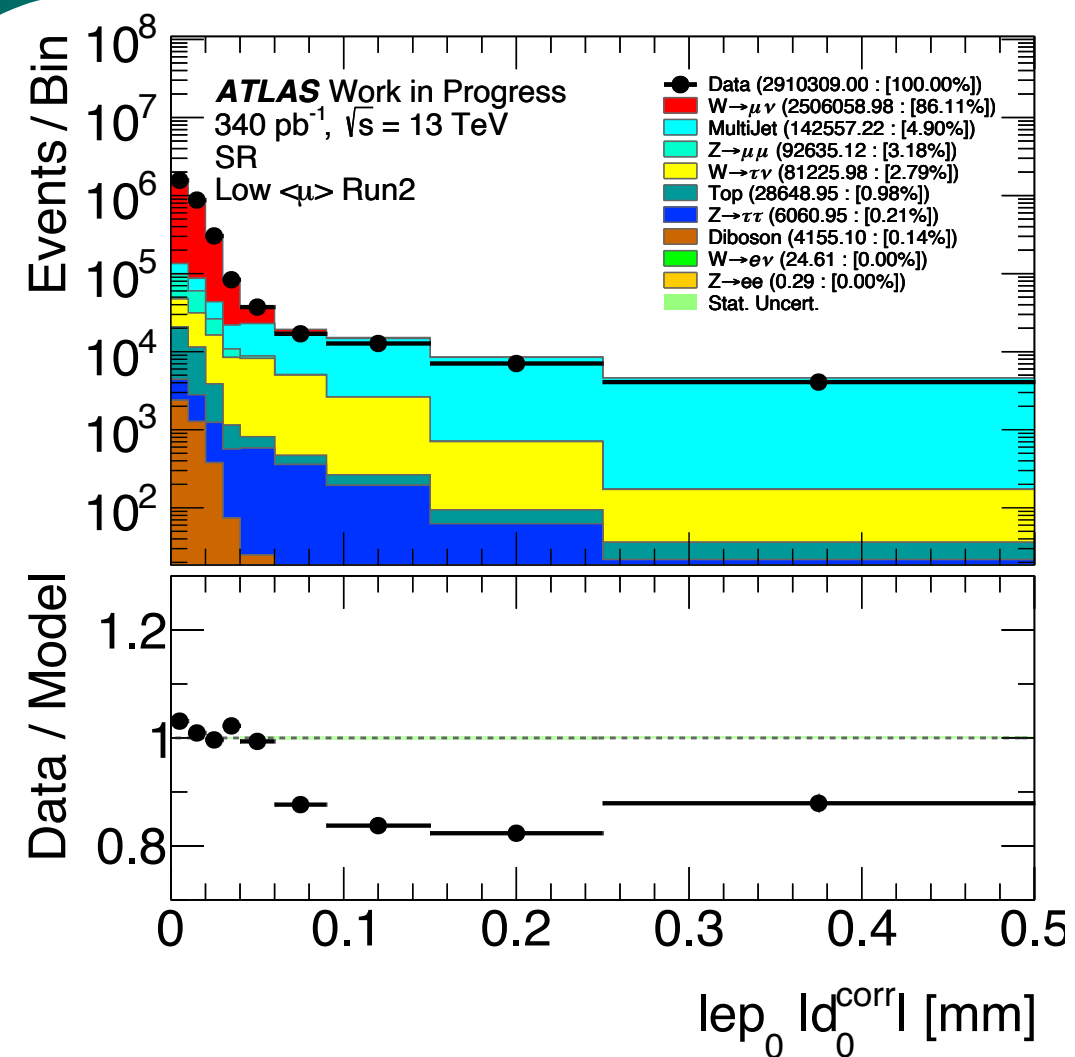
- ▶ Отклонения в области (0.9, 1) может быть связано как с плохим моделированием Монте-Карло, так и с плохой оценкой КХД фона.

Оптимизация разбиения распределения BDT на интервалы



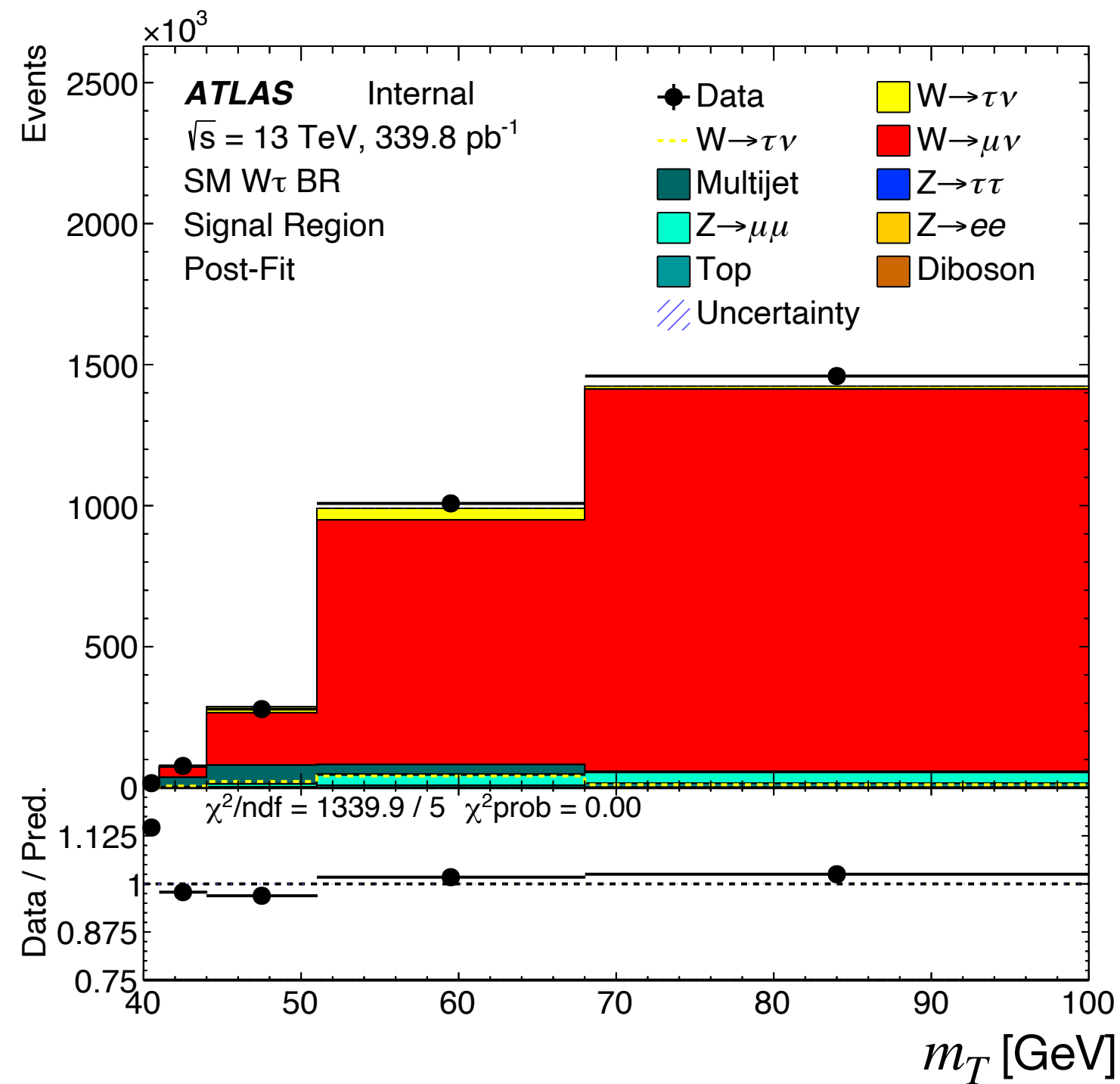
- Для получения оптимального результата необходимо провести оптимизацию разбиения распределения новой переменной отклика модели BDT.
- Разбиение проводится таким образом чтобы быть наиболее чувствительным к изменению формы сигнального и фонового распределения.

$$\text{BDT} = [0.0, 0.3, 0.5, 0.65, 0.83, 1.0]$$



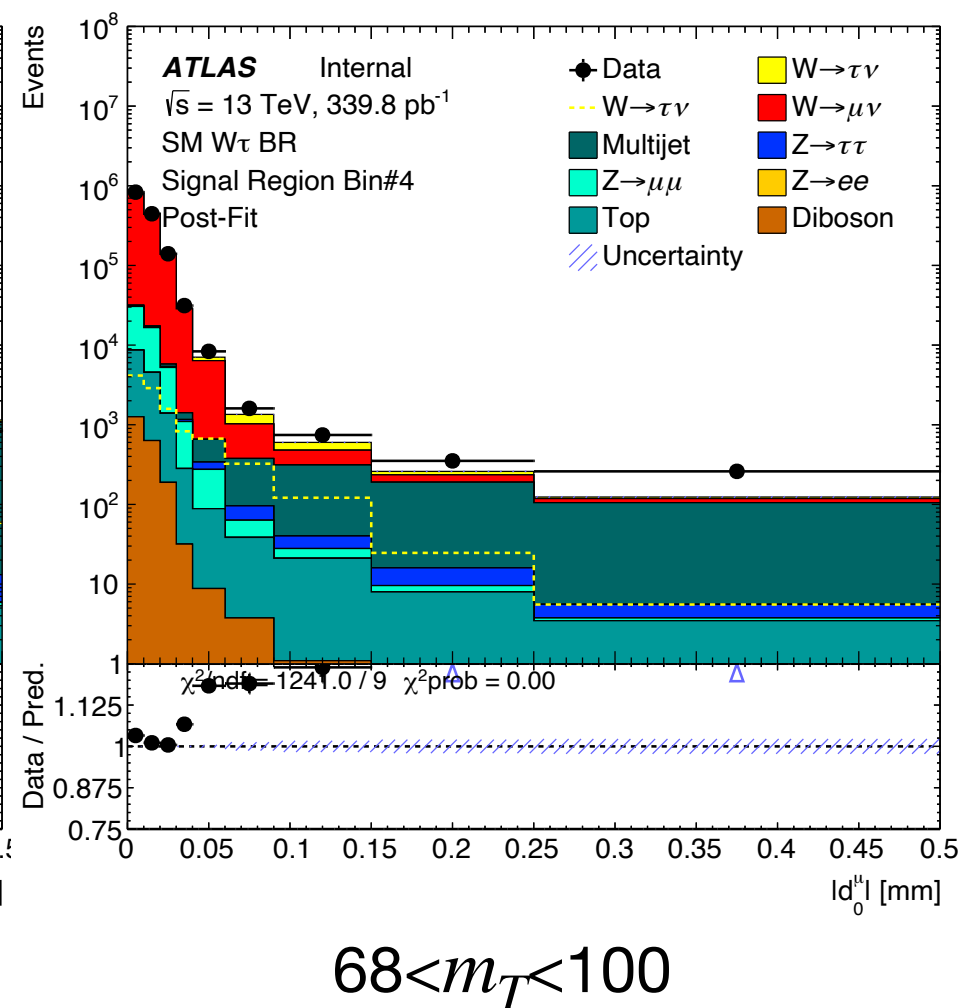
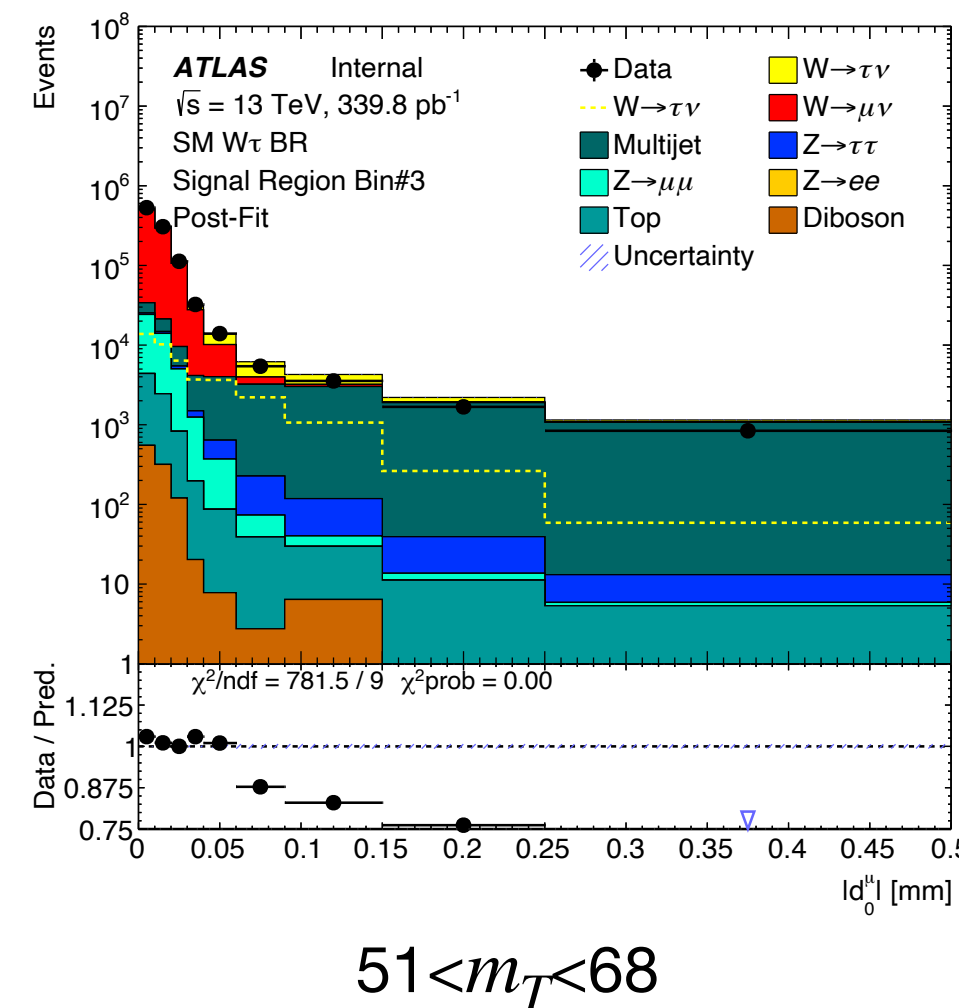
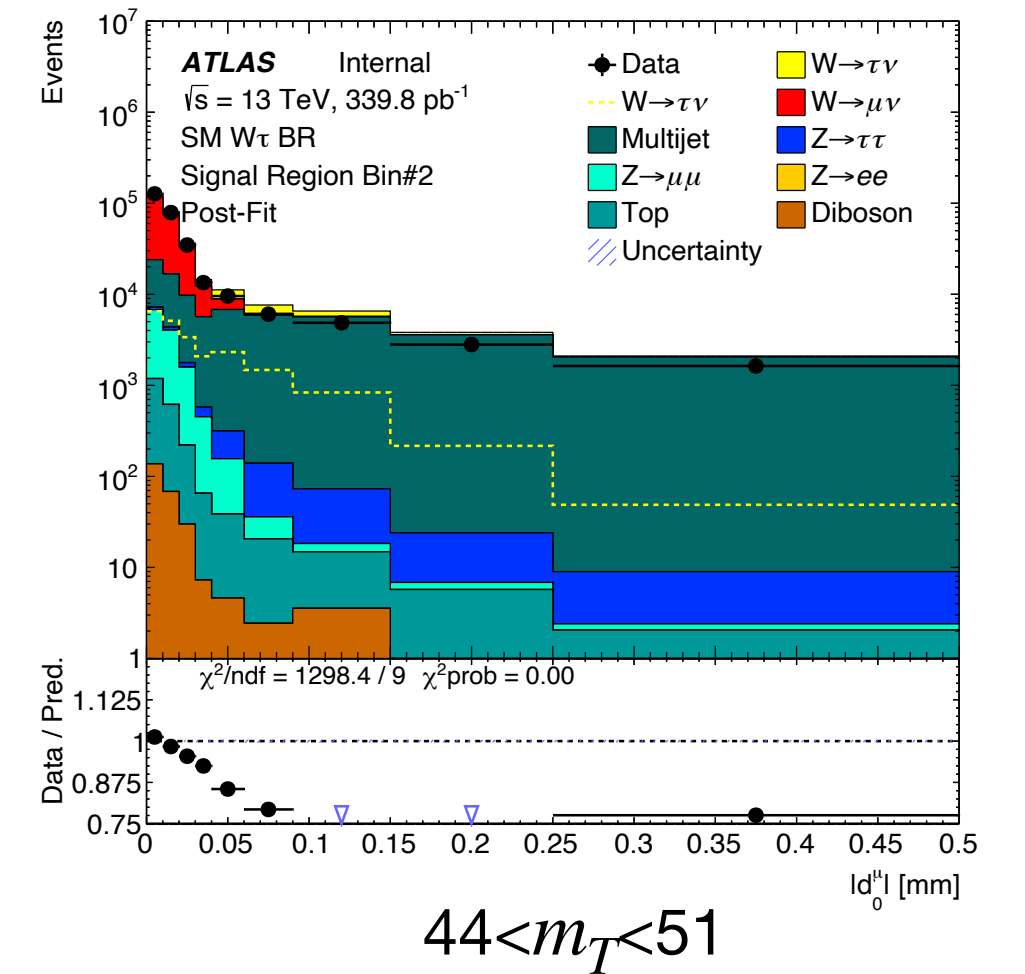
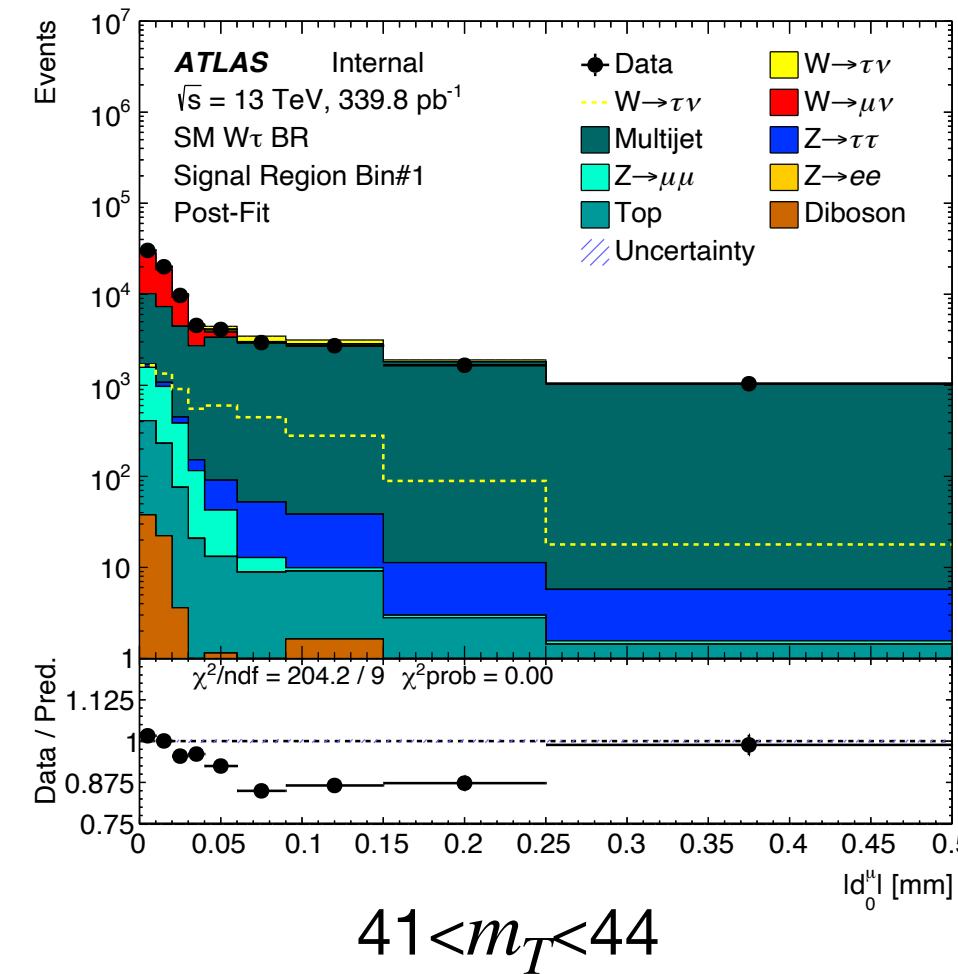
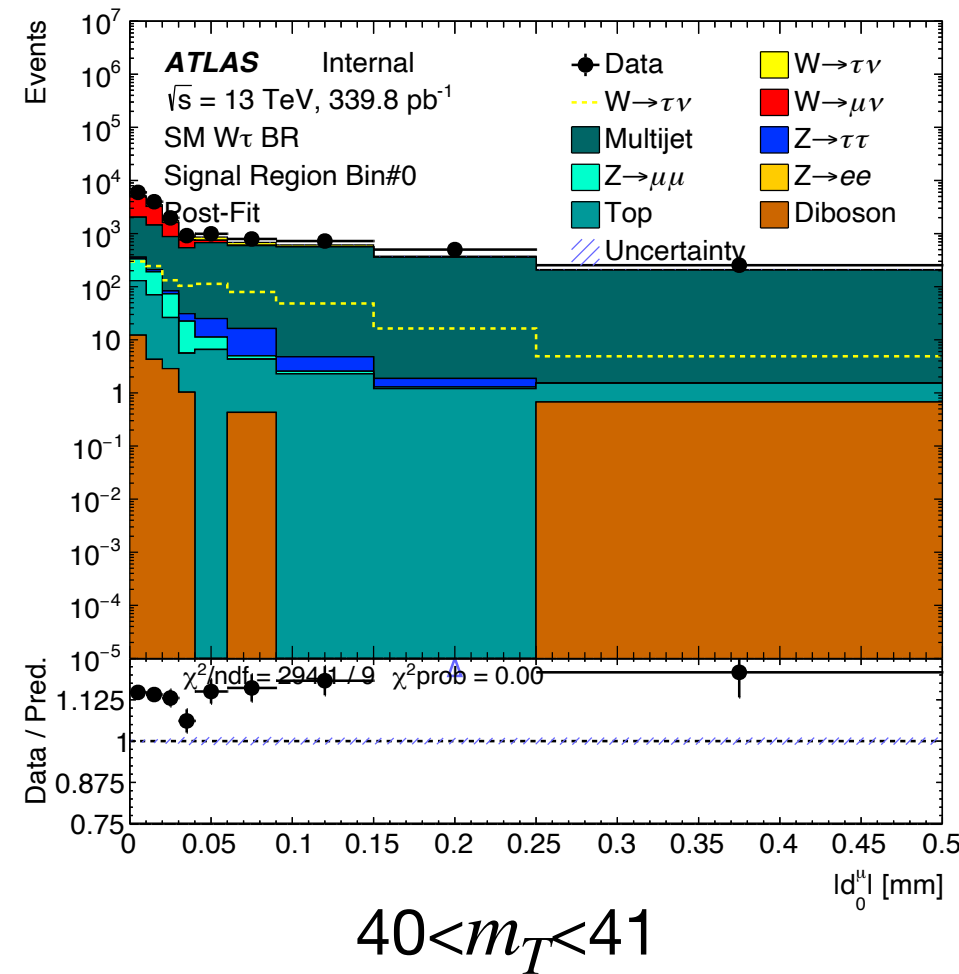
- При проведении измерения разбиение на бины сигнального региона производится в двумерном распределении прицельного параметра d_0 и отклика классификатора BDT или поперечной массы m_T .
- d_0 является чувствительной переменной при идентификации распадов $W \rightarrow \tau\nu$. При распаде $W \rightarrow \tau\nu$ распределение d_0 более шире, чем для распадов при прямых распадах $W \rightarrow \mu\nu$.
- Отклонения в отношении экспериментальных и МК данных обусловлены различающимися значениями сечения пучка протонных сгустков.

Результаты измерения с использованием переменной поперечной массы m_T



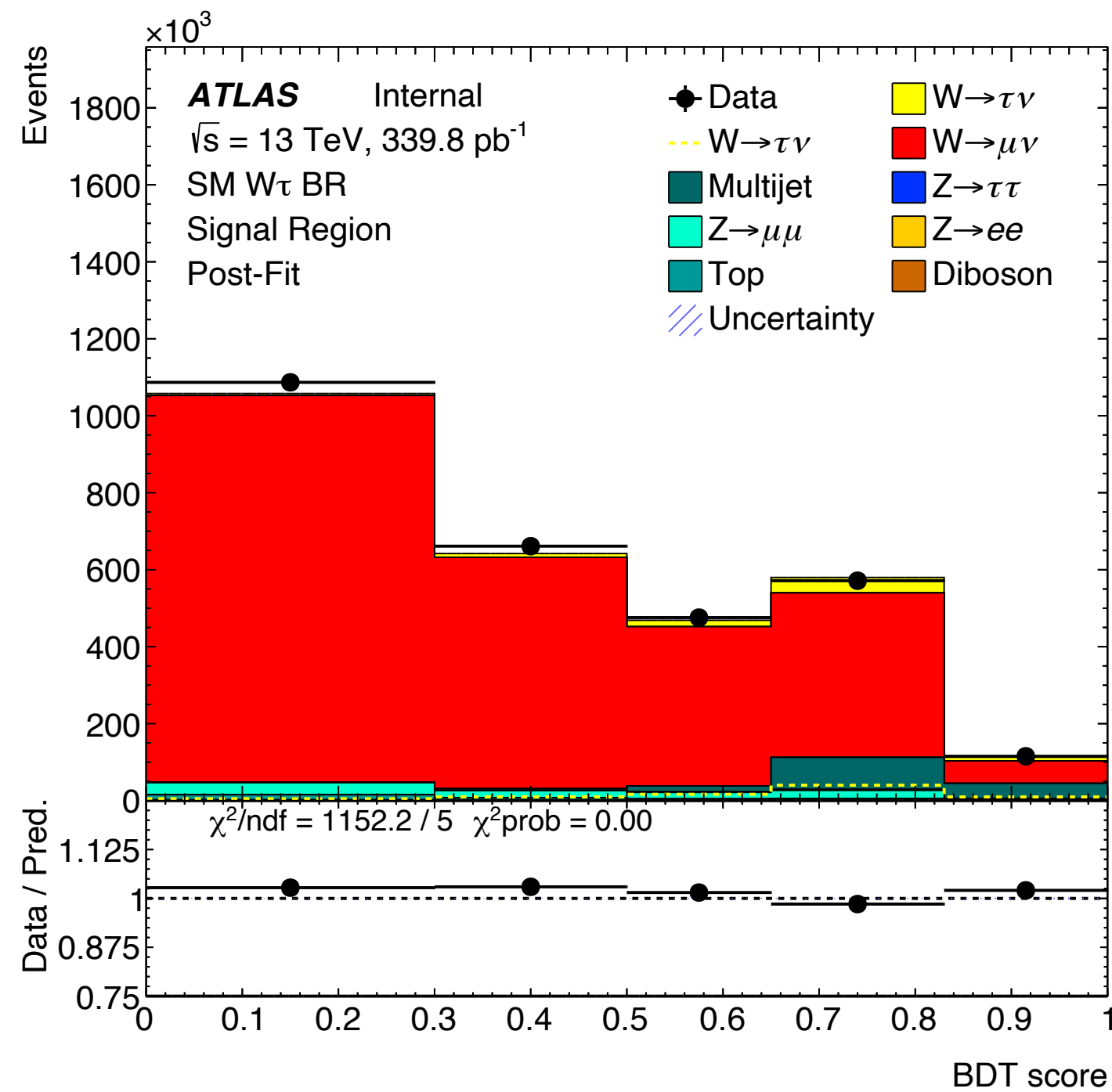
Результат измерения на данных Азимова

$$R_{\tau/\mu} = 1.000^{+0.015}_{-0.015}(\text{stat})^{+0.022}_{-0.022}(\text{syst})$$



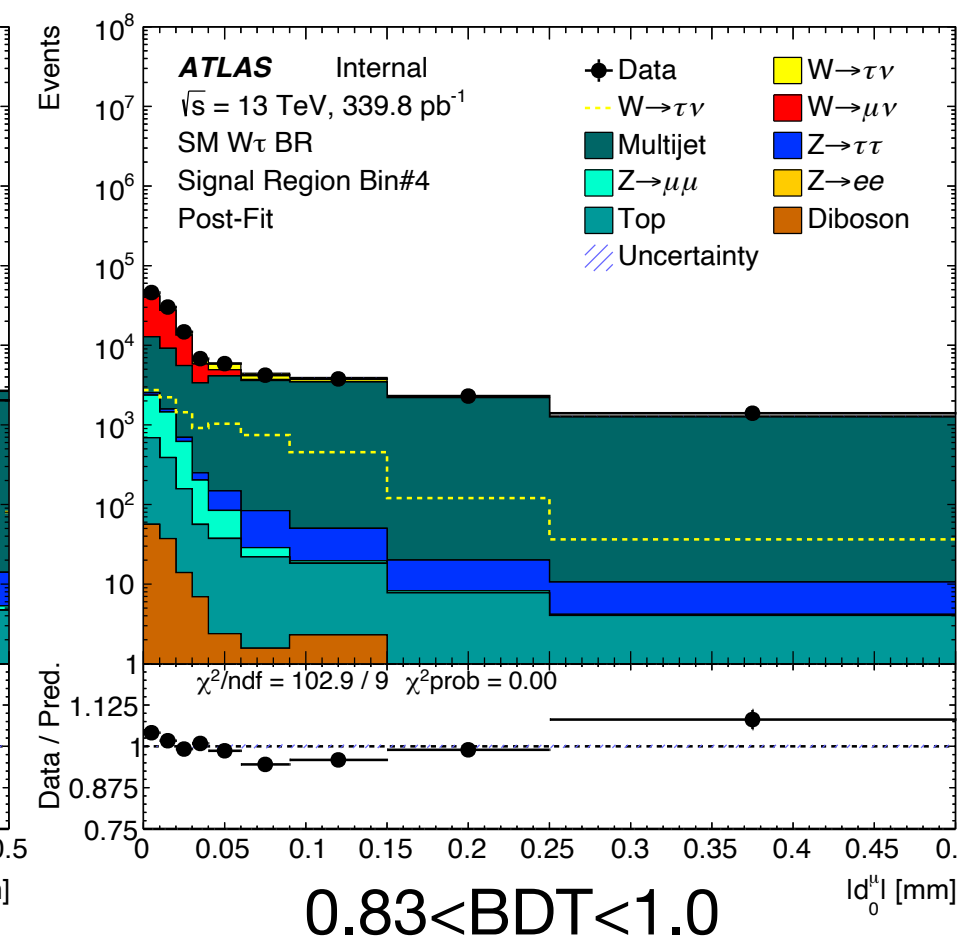
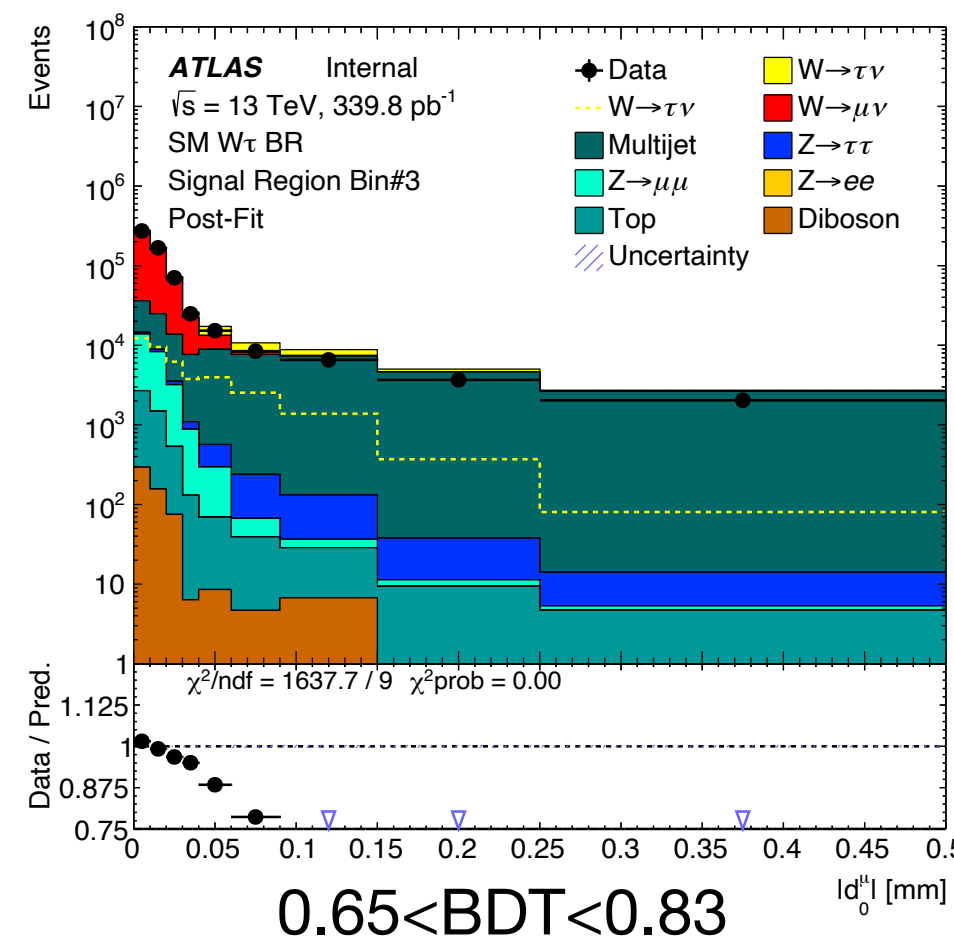
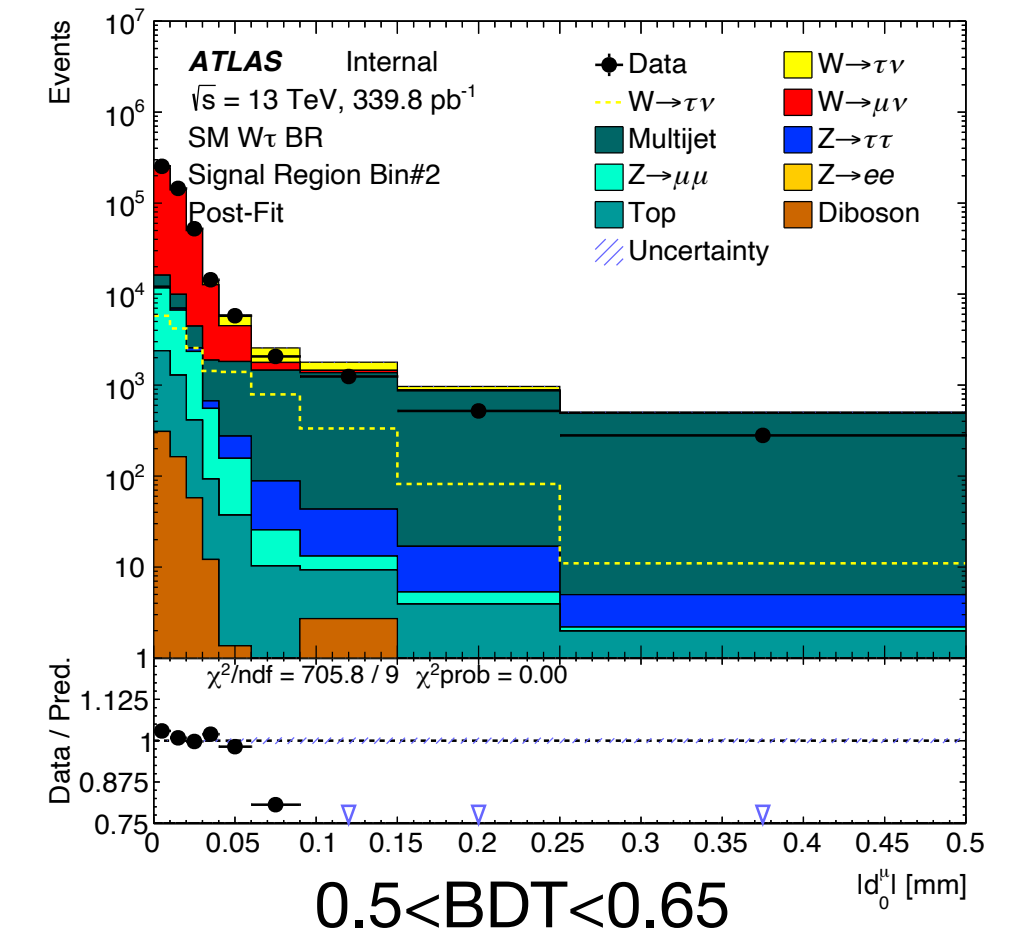
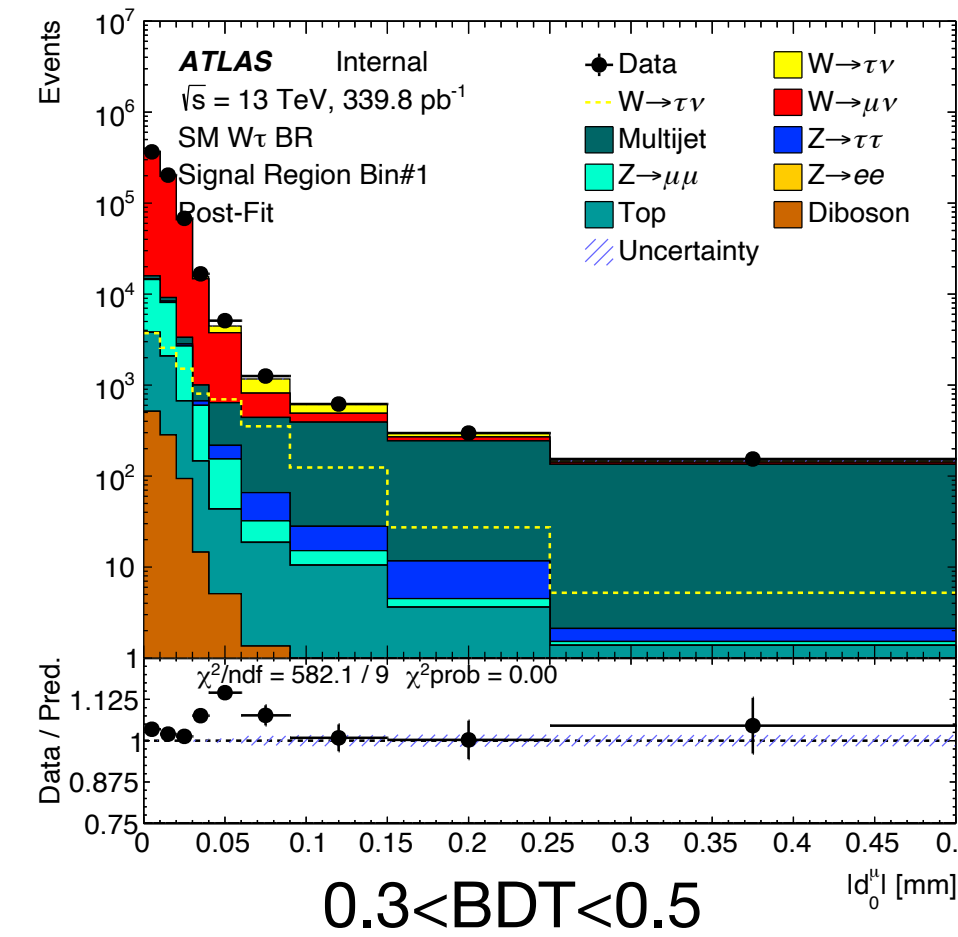
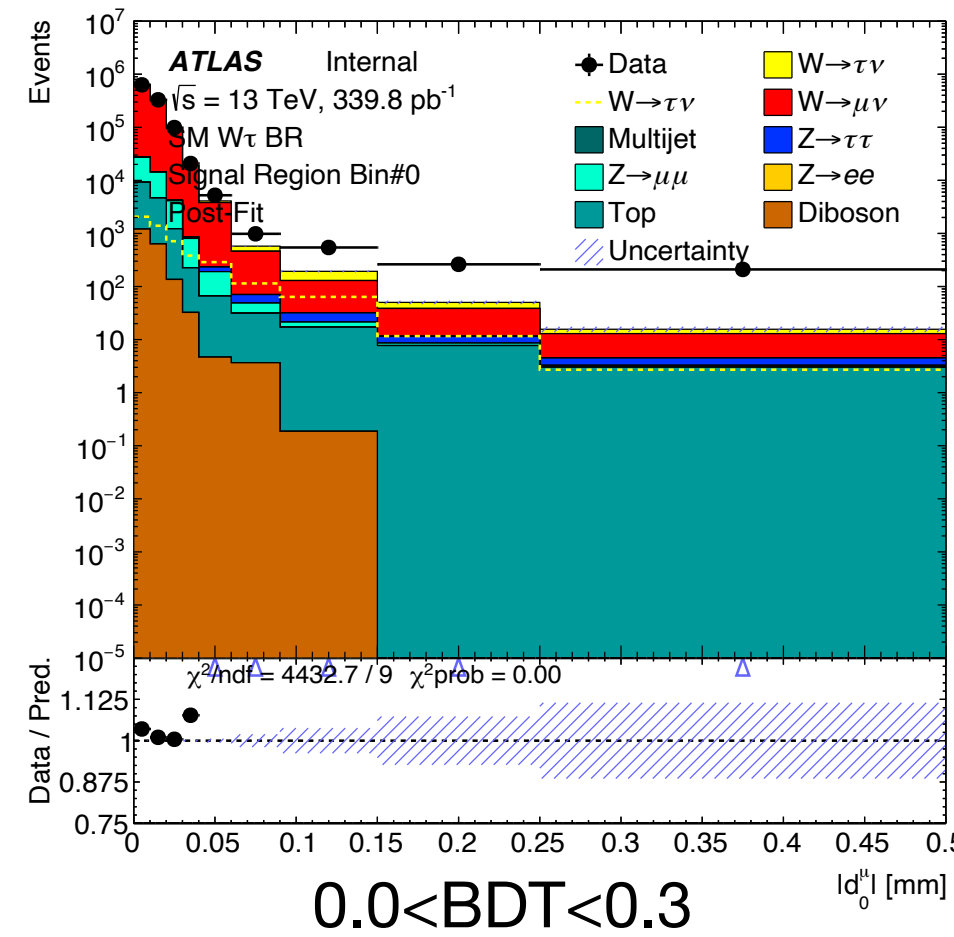
На рисунках показано сравнение экспериментальных и МК данных. Это сделано для того, чтобы показать имеющиеся отклонения в распределении d_0 .

Результаты измерения с использованием новой переменной отклика классификатора BDT



Результат измерения на данных Азимова

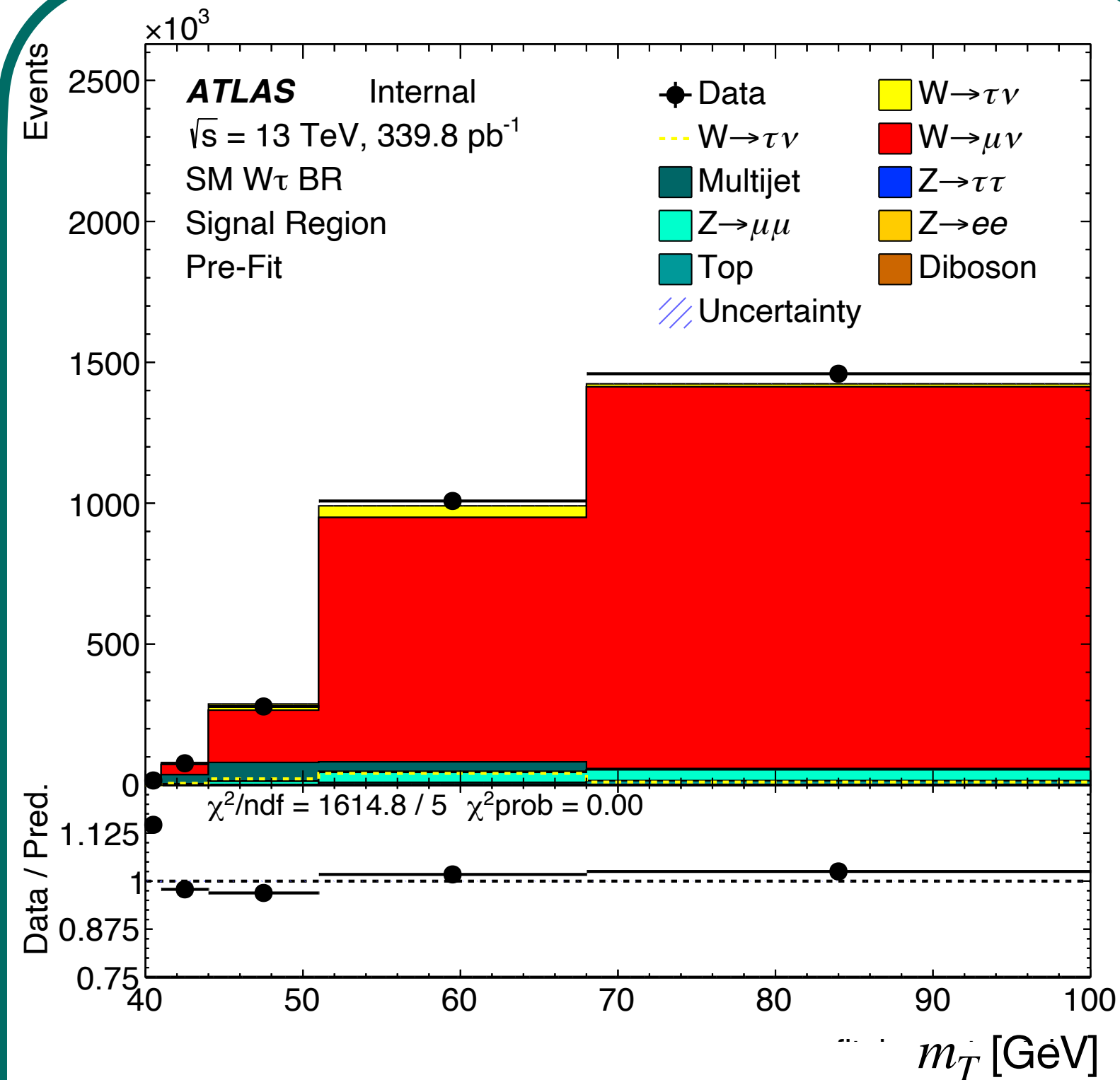
$$R_{\tau/\mu} = 1.000^{+0.015}_{-0.015}(\text{stat})^{+0.027}_{-0.027}(\text{syst})$$



На рисунках показано сравнение экспериментальных и МК данных. Это сделано для того, чтобы показать имеющиеся отклонения в распределении d_0 .

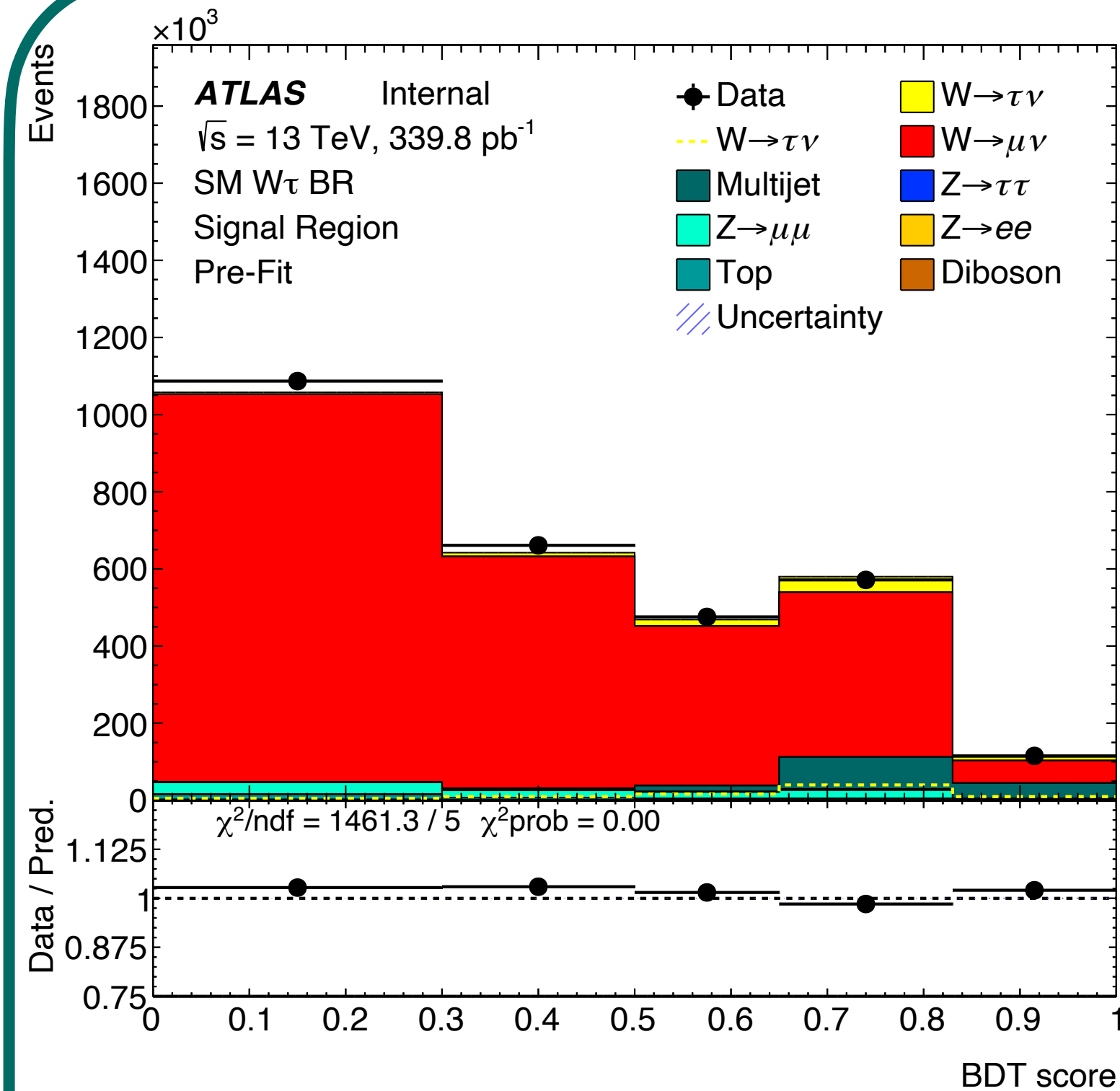
Результаты

Измерение с использованием переменной поперечной массы



$$R_{\tau/\mu} = 1.000^{+0.015}_{-0.015}(\text{stat})^{+0.022}_{-0.022}(\text{syst})$$

Измерение с использованием новой переменной отклика классификатора BDT



$$R_{\tau/\mu} = 1.000^{+0.015}_{-0.015}(\text{stat})^{+0.027}_{-0.027}(\text{syst})$$

- Сравнение двух результатов говорит о том, что использование переменной отклика классификатора BDT при измерении не влияет на статистическую погрешность данных, при этом увеличивает систематическую погрешность измерения на 0.005.
- Классификатор не получает новой информации для классификации событий из кинематических переменных в сравнении с информацией о кинематике событий, заложенной в m_T .

Заключение

- ▶ Выполнен подбор подходящего классификатора и его параметров. Критерием выбора оптимального классификатора выступал интеграл под ROC-кривой. Сделан выбор в пользу BDT классификатора использующего для классификации 70% обучающей выборки и архитектуру из 200 деревьев с максимальной глубиной в 5 ветвей.
- ▶ Получен отклик классификатора событий для экспериментальных данных и данных Монте-Карло сигнального региона. Исследованы наблюдаемые расхождения между Монте-Карло и экспериментальными данными с помощью псевдо- W региона. Показано, что расхождения между Монте-Карло и экспериментальными данными в распределении отклика классификатора BDT могут быть связаны как с плохим моделированием Монте-Карло так и с плохой оценкой КХД фона.
- ▶ Сделана оптимизация разбиения на интервалы новой переменной отклика классификатора BDT.
- ▶ Проведено измерение лептонной универсальности с использованием новой переменной отклика классификатора BDT. Для сравнения результата измерения проведено такое же измерение с использованием m_T . Выполнено сравнение результатов. Сделан вывод о том, что классификатор не получает новой информации из кинематических переменных по сравнению с информацией, заложенной в m_T .
- ▶ Выполнено ознакомление с пакетом ROOT, библиотекой машинного обучения TMVA. Получены начальные сведения о Стандартной модели, лептонной универсальности, экспериментах по проверке лептонной универсальности, эксперименте ATLAS и машинном обучении. Также были изучены следующие пакеты для проведения анализа данных: xTauReader, HAPPy, TRExFitter. Выполнен ряд работ по модернизации xTauReader фреймворка. За счет данных улучшений удалось значительно повысить вычислительную производительность на этапе классификации данных.

Спасибо за внимание!

Дополнительные слайды

Использованные данные

Экспериментальные данные

В работе использовались экспериментальные данные, набранные на детектора ATLAS в 2017 и 2018 году во время режима с низкой светимостью 340 пБ^{-1} . При столкновении протон-протонных пучков с энергией 13 ТэВ.

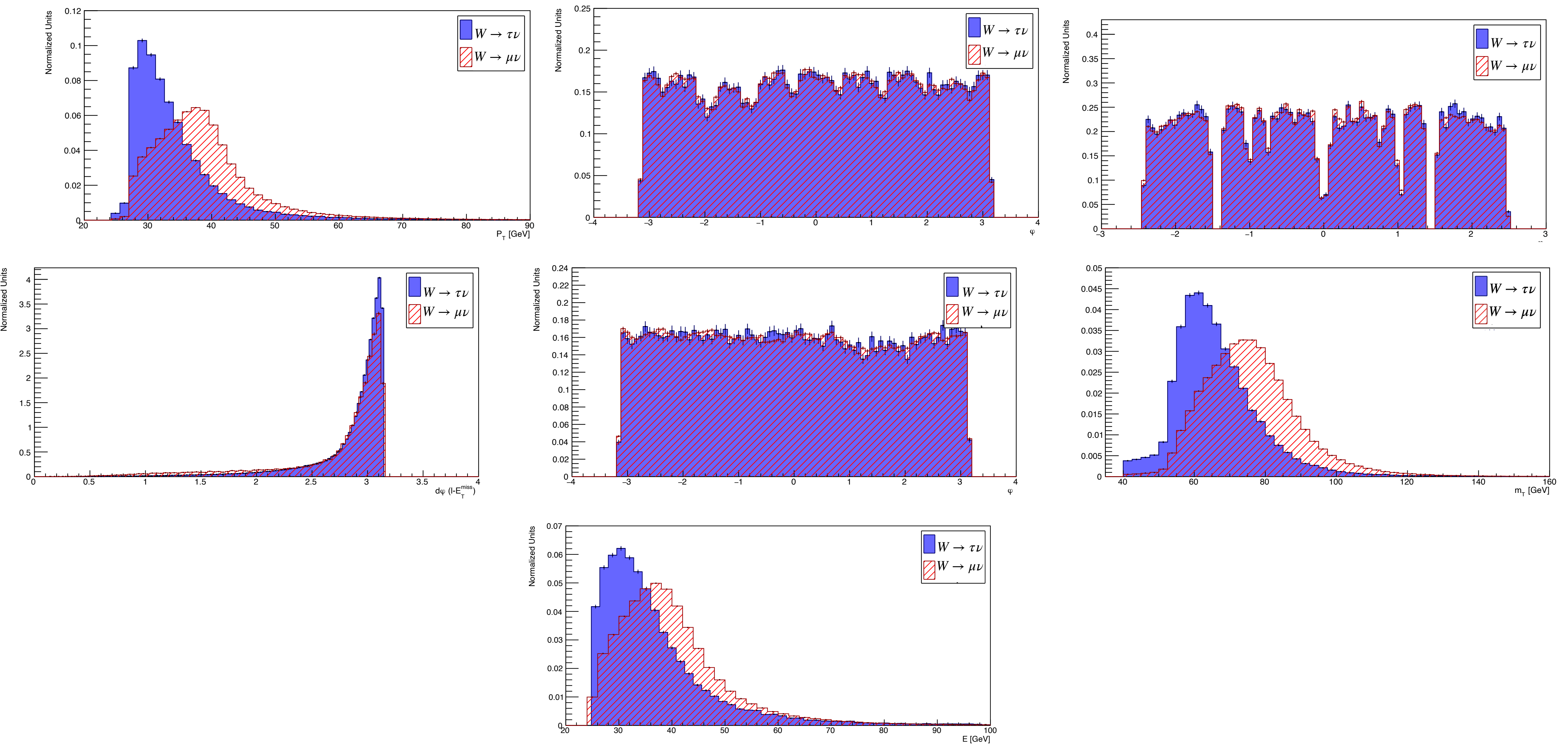
Монте-Карло моделирование

Смоделированные данные были получены методом Монте-Карло с помощью генераторов Pythia и Sherpa и прошли всю цепочку реконструкций на условии реальных протон - протонных столкновений эксперимента ATLAS

Список Монте-Карло используемых в работе.

Sample	DSID	Generator	xs [pb]
$W^+ \rightarrow \mu\nu$	361101	PowhegPythia8EvtGen	11500.9
$W^- \rightarrow \mu\nu$	361104	PowhegPythia8EvtGen	8579.31
$W^+ \rightarrow \tau\nu$	361102	PowhegPythia8EvtGen	11500.9
$W^- \rightarrow \tau\nu$	361105	PowhegPythia8EvtGen	8579.31
$Z \rightarrow \tau\tau$	361108	PowhegPythia8EvtGen	1950.63
$Z \rightarrow \mu\mu$	361107	PowhegPythia8EvtGen	1950.63
Top	410013	PhPy8EG_P2012	35.8455
Top	410014	PhPy8EG_P2012	35.8244
Top	410470	PhPy8EG	729.77
Top	410642	PhPy8EG	36.993
Top	410643	PhPy8EG	22.174
Top	410644	PowhegPythia8EvtGen	2.06146
Top	410645	PowhegPythia8EvtGen	1.28867
Top	410646	PowhegPythia8EvtGen	35.8486
Diboson	363356	Sherpa_221_PDF30	2.20355
Diboson	363358	Sherpa_221_PDF30	3.4328
Diboson	363359	Sherpa_221_PDF30	24.708
Diboson	363360	Sherpa_221_PDF30	24.724
Diboson	363489	Sherpa_221_PDF30	11.42
Diboson	364250	Sherpa_221_PDF30	1.2523
Diboson	364253	Sherpa_221_PDF30	4.579
Diboson	364254	Sherpa_221_PDF30	12.501
Diboson	364255	Sherpa_221_PDF30	3.2344

Сравнение сигнальных и фоновых событий

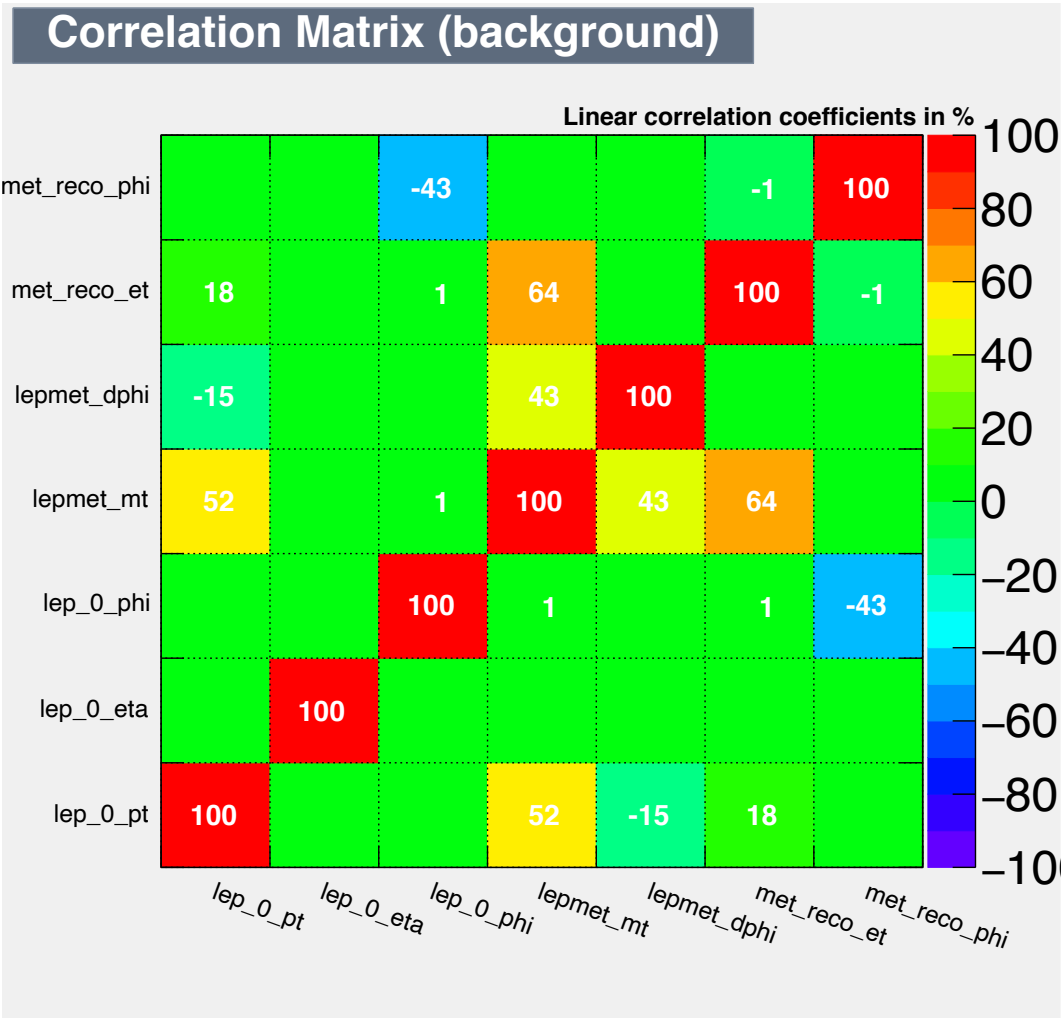
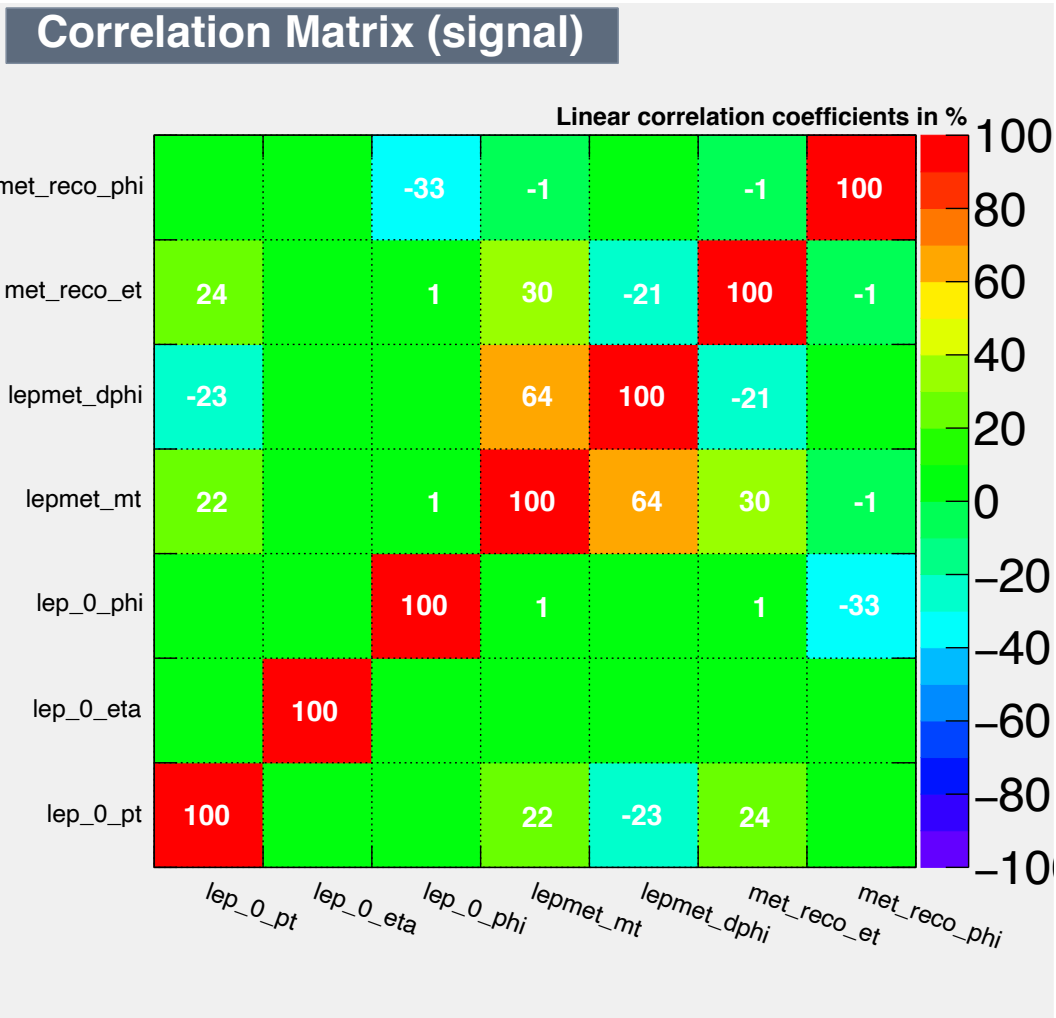


Обучение классификатора BDT

Значимость кинематических переменных используемых в обучении

Переменная	P_T	M_T	$\Delta\phi$	E_T^{miss}	η	MET ϕ	ϕ_{lep}
Значимость	1.8e-01	2.6e-01	4.4e-02	1.2e-01	2.3e-04	1.2e-04	1.3e-04

Матрицы корреляции



Параметры классификатора:

Количество деревьев - 200

Глубина деревьев - 5

Отношение тренировочных и тестовых событий 70/30.

Количество событий:

Сигнальных - 182806

Фоновых - 732000

Сигнальные события:

$$W \rightarrow \tau \nu \rightarrow \mu \nu \nu$$

Фоновые события:

$$W \rightarrow \mu \nu$$

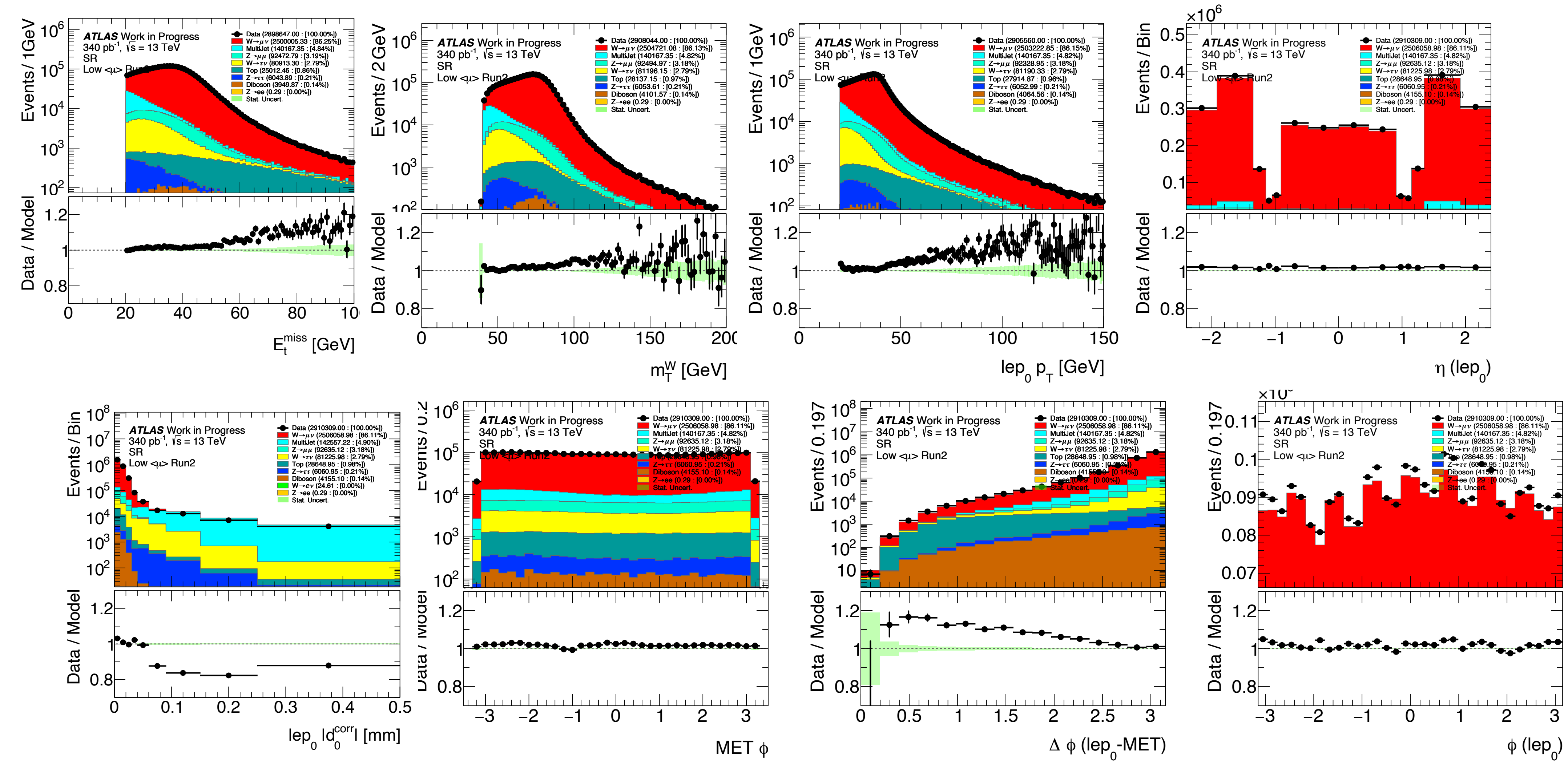
$$Z \rightarrow \mu \mu$$

Распады t кварка

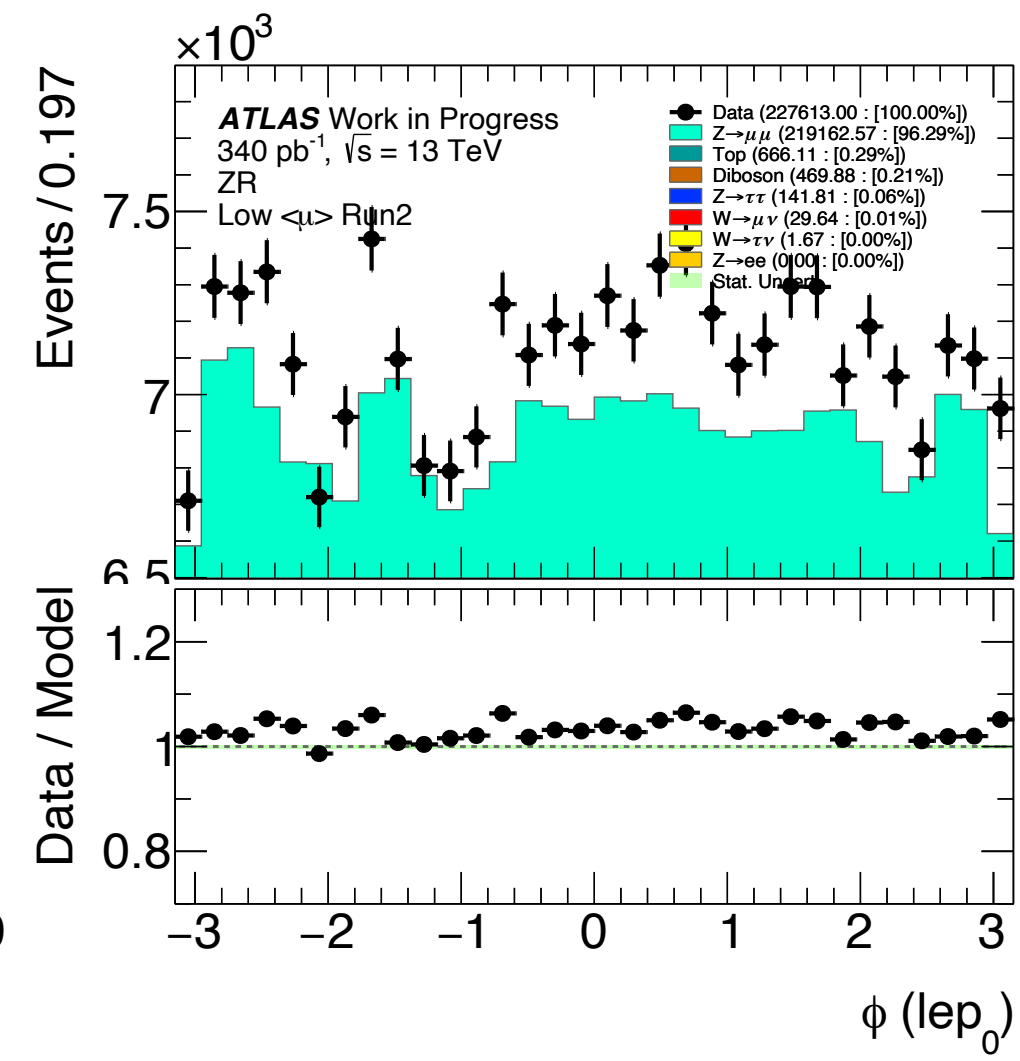
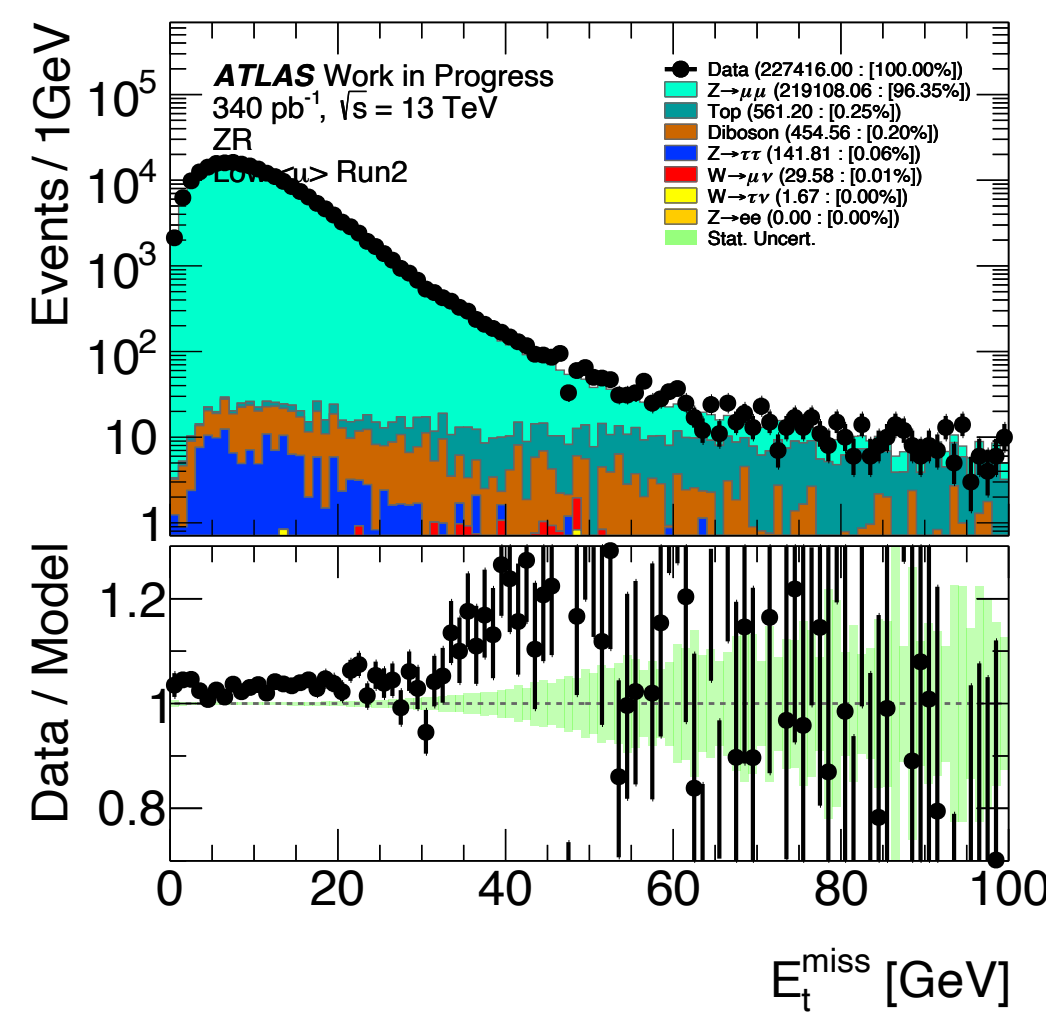
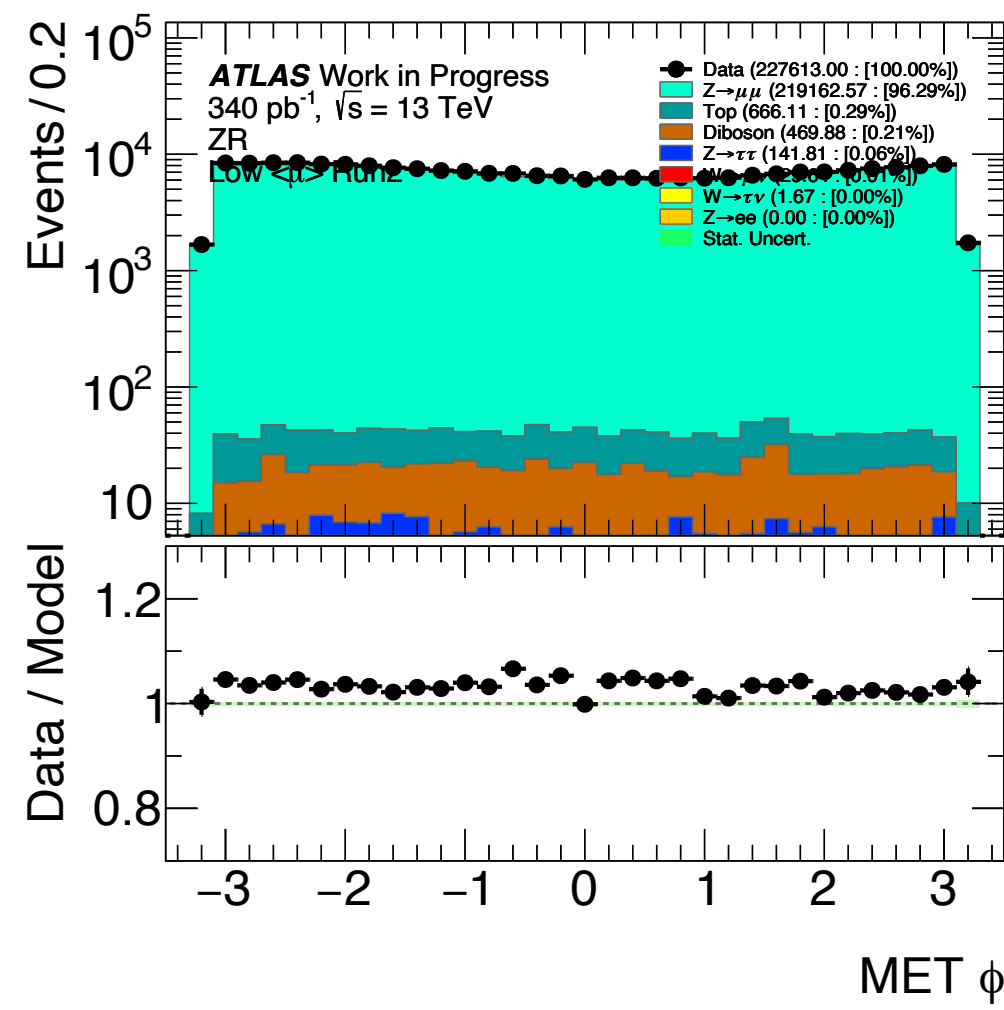
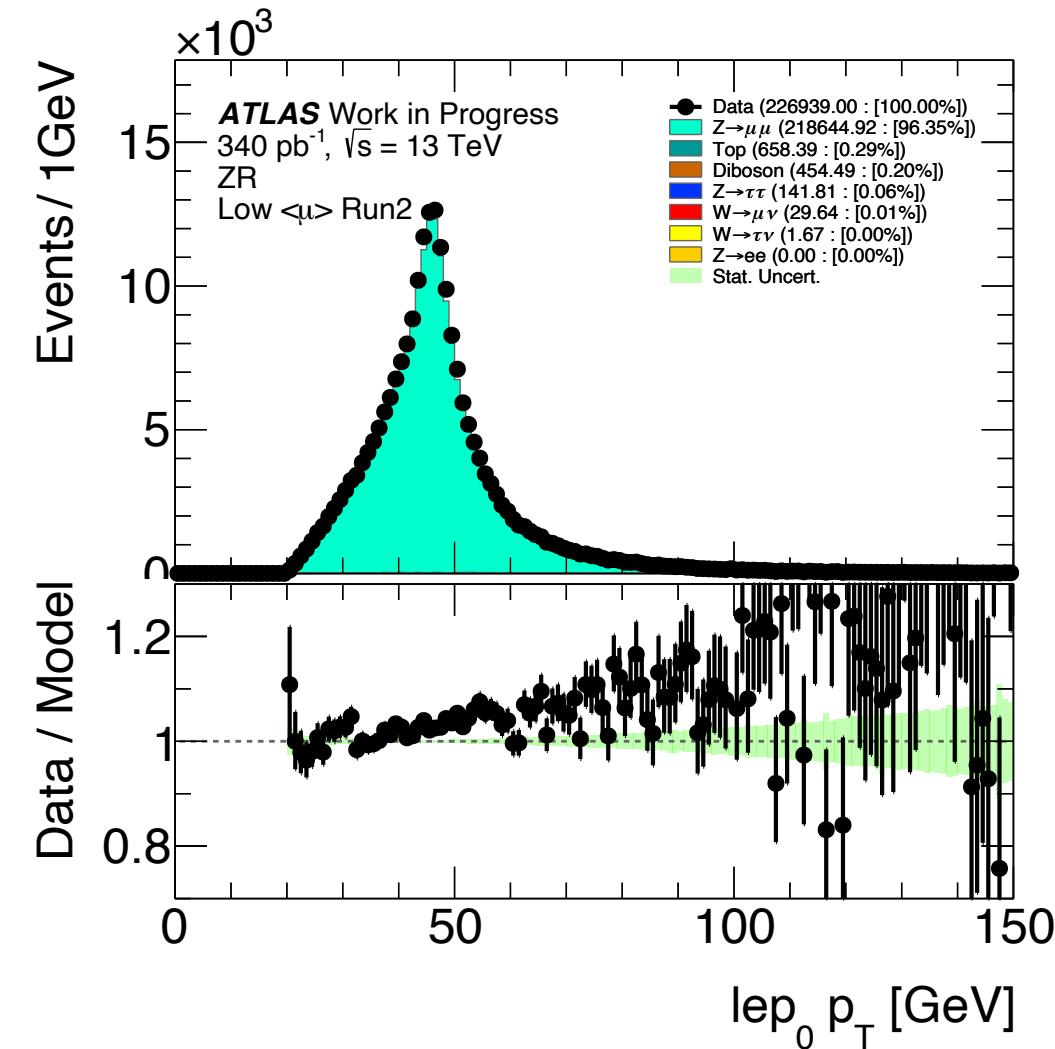
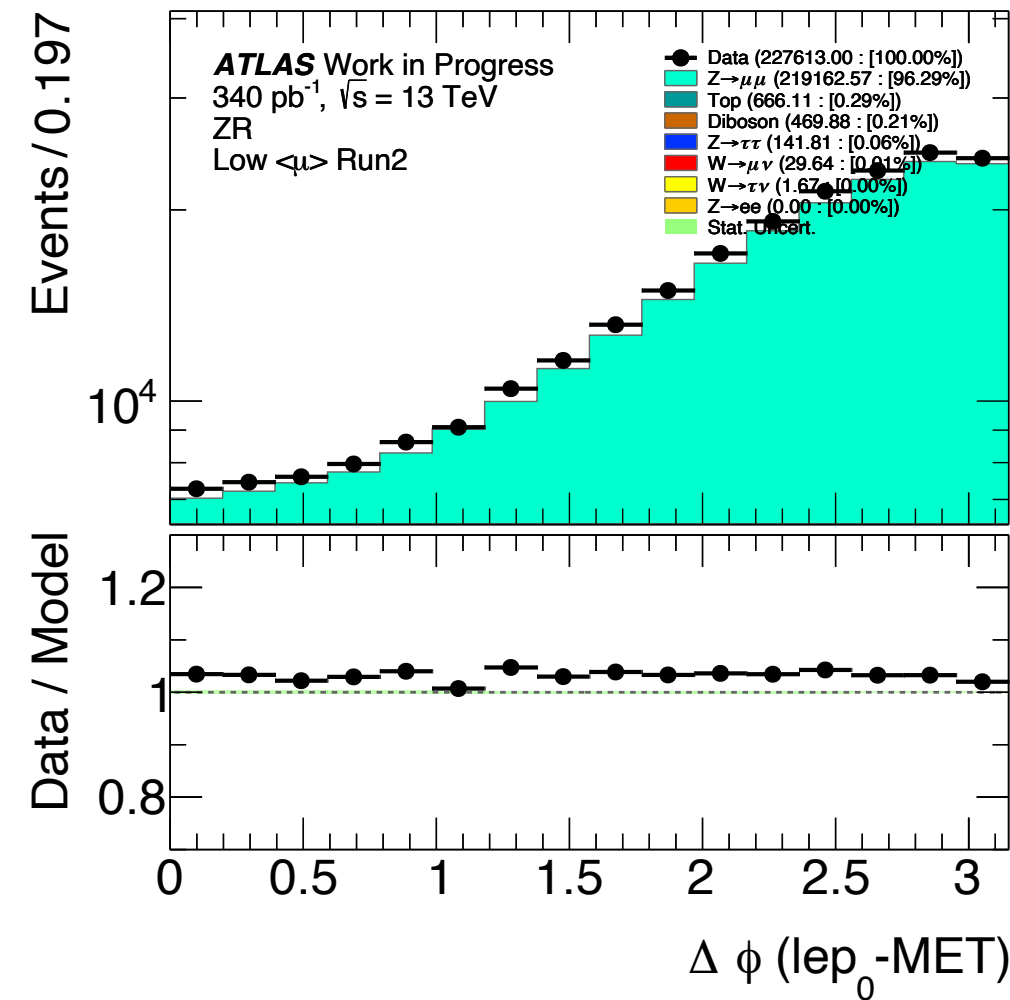
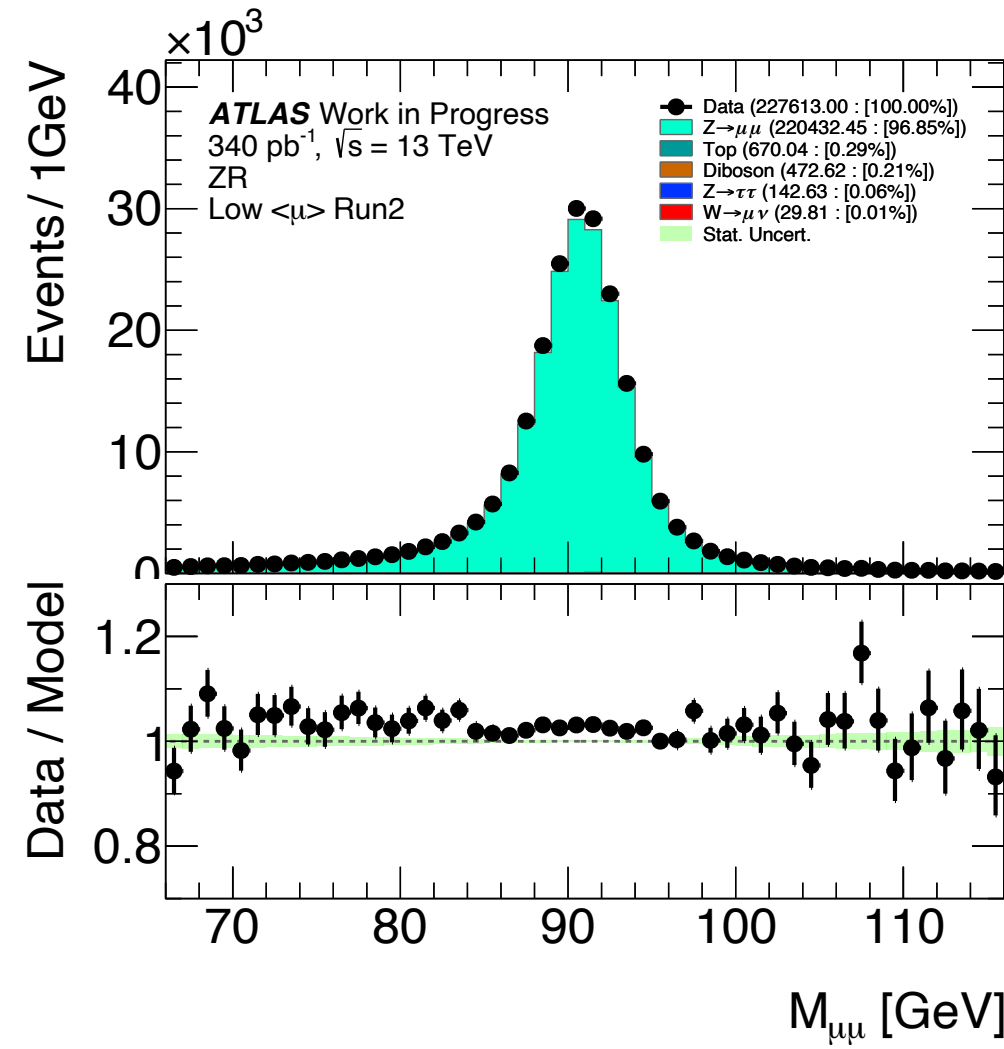
Дибозонные процессы

Значение интеграла под ROC кривой - 0.795

Сигнальный регион



Z регион



Функция правдоподобия

$$L = \prod_b^{bins \in ZR} Pois(N_b^{data}, \mu_{ZR} N_b^{MC}) \prod_b^{bins \in SR} Pois(N_b^{data}, \mu_W [\mu_{sig} N_b^\tau + N_b^l] + \mu_{ZR} N_b^{MC} + N_b^{QCD}) \prod_j^{NPs} G(\theta_j^0, \theta_j) \quad (1)$$

Задание интересующего параметра

- μ_{sig} - Отношение отношений сечени.
- μ_l, μ_{ZR} - Нормировочные коэффициенты.
- N_b^τ - Вклад от распадов $W \rightarrow \tau\nu$.
- N_b^l - Вклад от распадов $W \rightarrow \mu\nu$.
- N_b^{QCD} - Вклад КХД фона.
- N_b^{MC} - Вклад остальных процессов.

$$\mu_{sig} = \left(\frac{BR(W \rightarrow \tau\nu) BR(\tau \rightarrow \mu\nu\nu)}{BR(W \rightarrow \tau\nu)} \right)_{data} / \left(\frac{BR(W \rightarrow \tau\nu) BR(\tau \rightarrow \mu\nu\nu)}{BR(W \rightarrow \tau\nu)} \right)_{MC} \quad (2)$$

$$BR(\tau \rightarrow \mu\nu\nu) = 17.39 \% \Rightarrow \mu_{sig} = \left(\frac{BR(W \rightarrow \tau\nu)}{BR(W \rightarrow \tau\nu)} \right)_{data} / \left(\frac{BR(W \rightarrow \tau\nu)}{BR(W \rightarrow \tau\nu)} \right)_{MC} \quad (3)$$

$$\left(\frac{BR(W \rightarrow \tau\nu)}{BR(W \rightarrow \mu\nu)} \right)_{MC} = 1 \Rightarrow \mu_{sig} = R_{\tau/\mu} = \frac{BR(W \rightarrow \tau\nu)}{BR(W \rightarrow \mu\nu)} \quad (4)$$

Учет систематических погрешностей

Используемые систематические погрешности

muon_eff_reco

MuRecoSys_corr_BKGFRAC

MuRecoSys_corr_HIGHTOLOWEXTRAPOLATION

MuRecoSys_corr_LUMIUNCERT

MuRecoSys_corr_MATCHING_SYM

MuRecoSys_corr_MCXSEC

MuRecoSys_corr_PtDEPENDENCY

MuRecoSys_corr_QCDTEMPLATE

MuRecoSys_corr_SUPRESSIONSCALE

MuRecoSys_corr_TRUTH_SYM

MuTTVASys_corr_BKGFRAC

MuTTVASys_corr_LUMIUNCERT

MuTTVASys_corr_MCXSEC

MuTTVASys_corr_QCDTEMPLATE

MuTTVASys_corr_SUPRESSIONSCALE

muon_eff_iso

MuIsoSys_corr_dRmuj

MuIsoSys_corr_eta

MuIsoSys_corr_isoBkg

MuIsoSys_corr_mass

MuIsoSys_corr_probeQ

MuIsoSys_corr_tagIsol

MuIsoSys_corr_tpdR

muon_eff_trig

MuTrigSys_PtEta_corr_dPhi_avg

MuTrigSys_PtEta_corr_iso_tight_avg

MuTrigSys_PtEta_corr_noIP_avg

MuTrigSys_PtEta_corr_iso_loose_avg

MuTrigSys_PtEta_corr_mlltight_avg

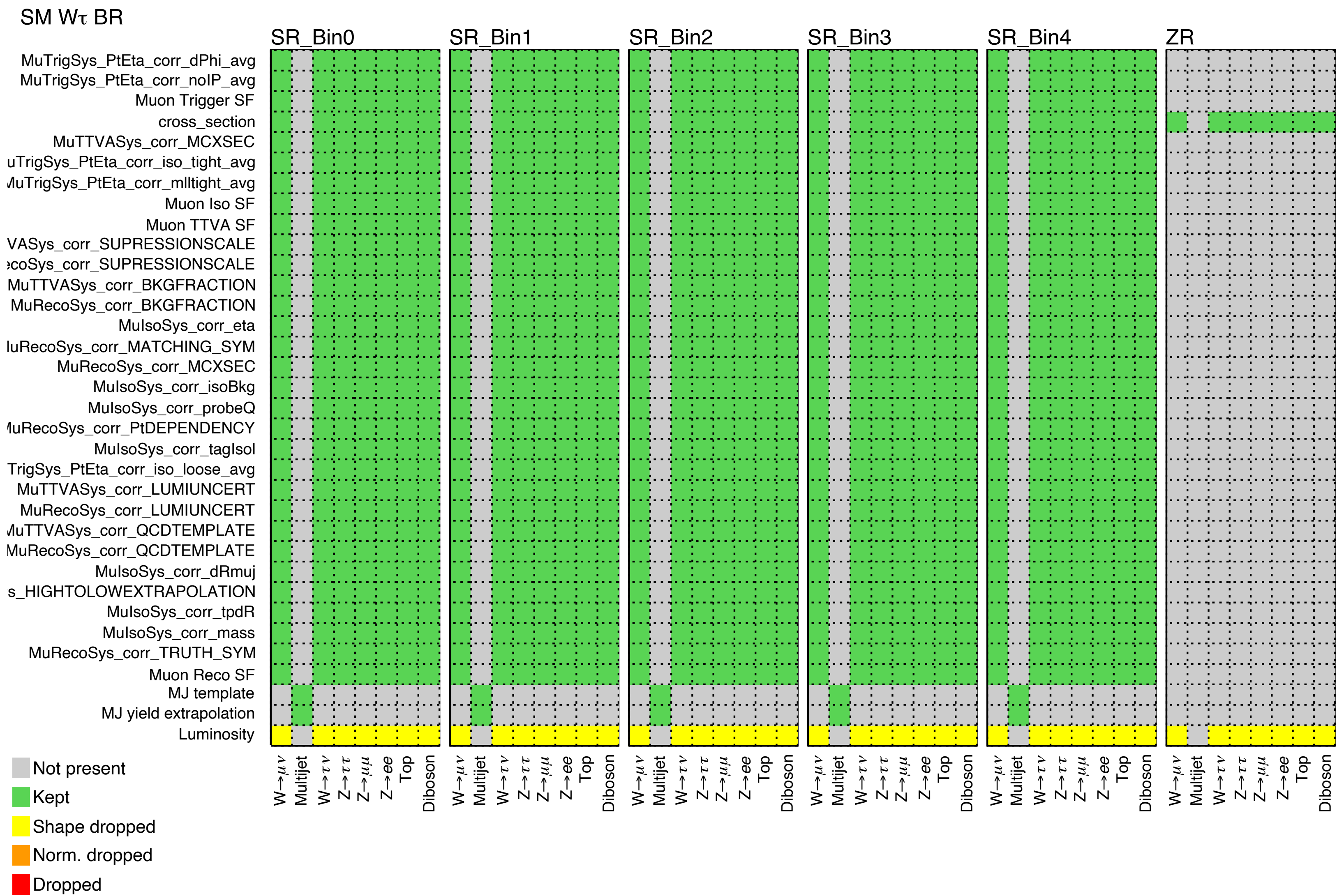
Вклад групп систематических погрешностей, входящих в полную систематическую погрешность.

	Контрольная переменная	
	m_T	BDT
Группа погрешностей	Погрешность	
Полная систематическая погрешность	0.022	0.027
Статистическая погрешность МК	0.014	0.017
Погрешность на оценку КХД	0.018	0.023
NormFactors	$< 10^{-3}$	0.003
cross_section	0.005	$< 10^{-3}$
lumi	0.008	$< 10^{-3}$
muIso_STAT	0.008	$< 10^{-3}$
muIso_SYS	0.008	0.011
muReco_STAT	0.007	0.001
muReco_SYS	0.008	$< 10^{-3}$
muTTVA_STAT	0.008	$< 10^{-3}$
muTTVA_SYS	0.008	$< 10^{-3}$
muTrig_STAT	0.008	0.009
muTrig_SYS	0.009	0.005

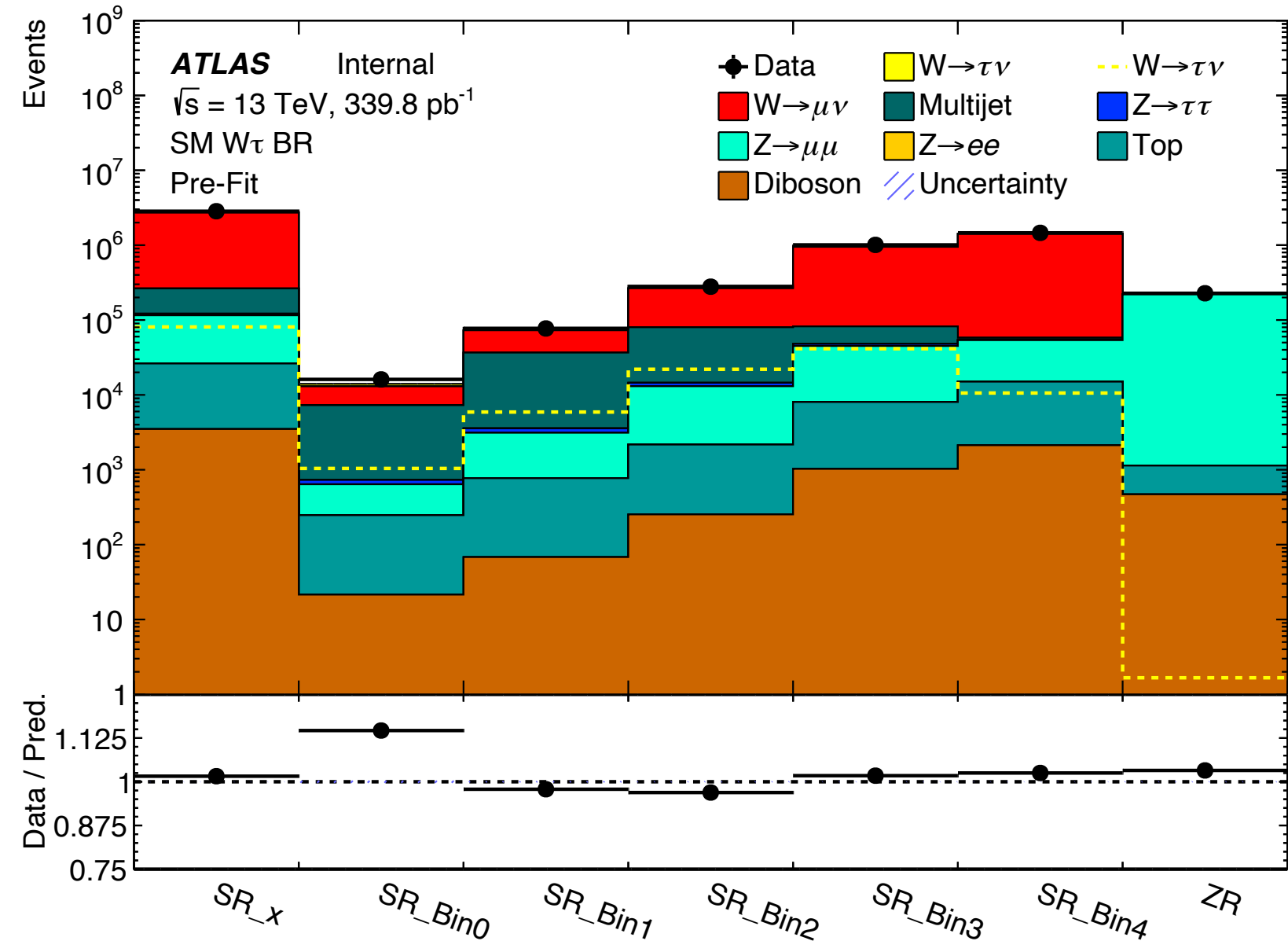
При разделении всех систематических погрешностей на группы, квадратичная сумма вкладов погрешностей от всех групп отличается от полной систематической погрешности. Это связано с корреляцией групп систематических погрешностей

Учет систематических погрешностей

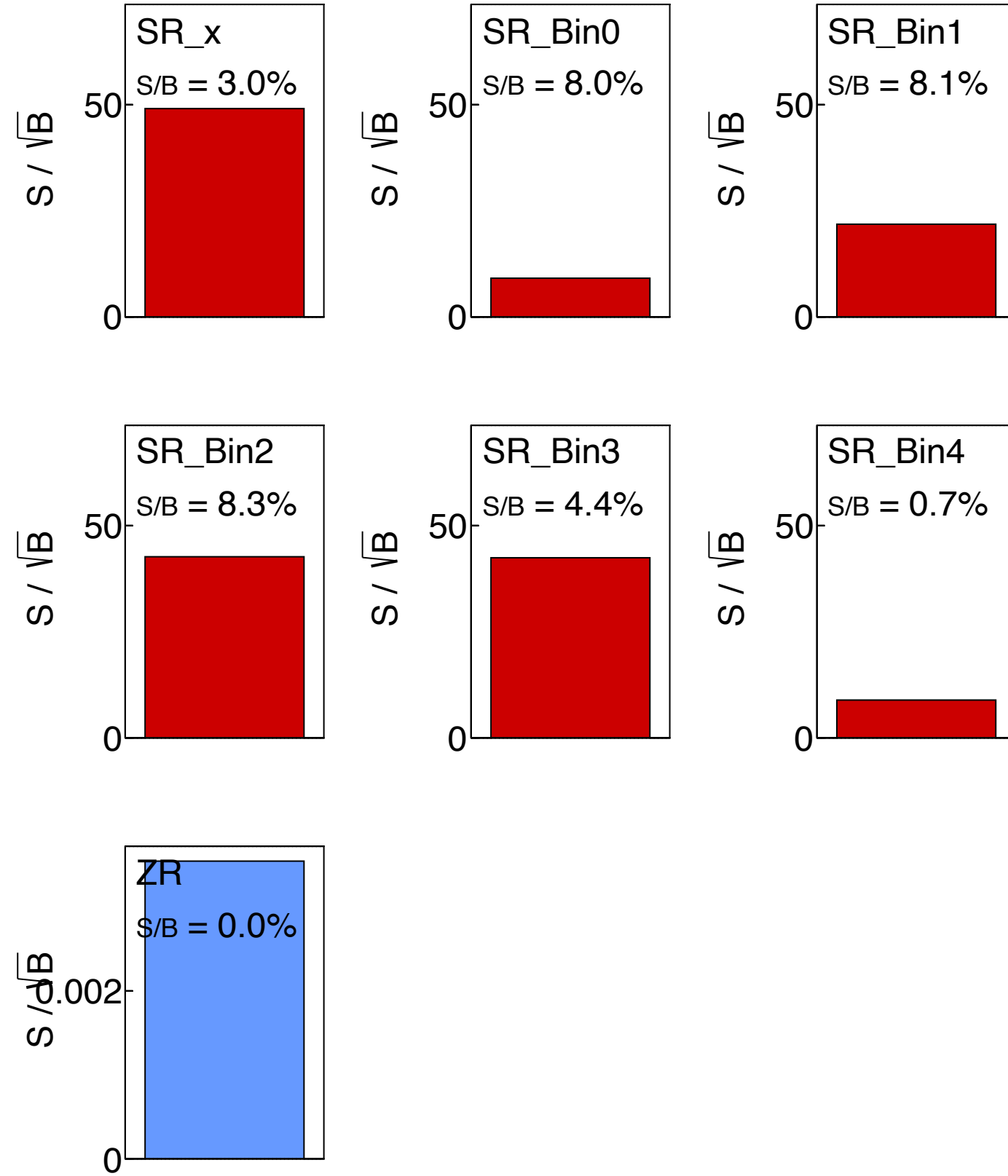
Использованные систематические погрешности для каждого процесса и региона.



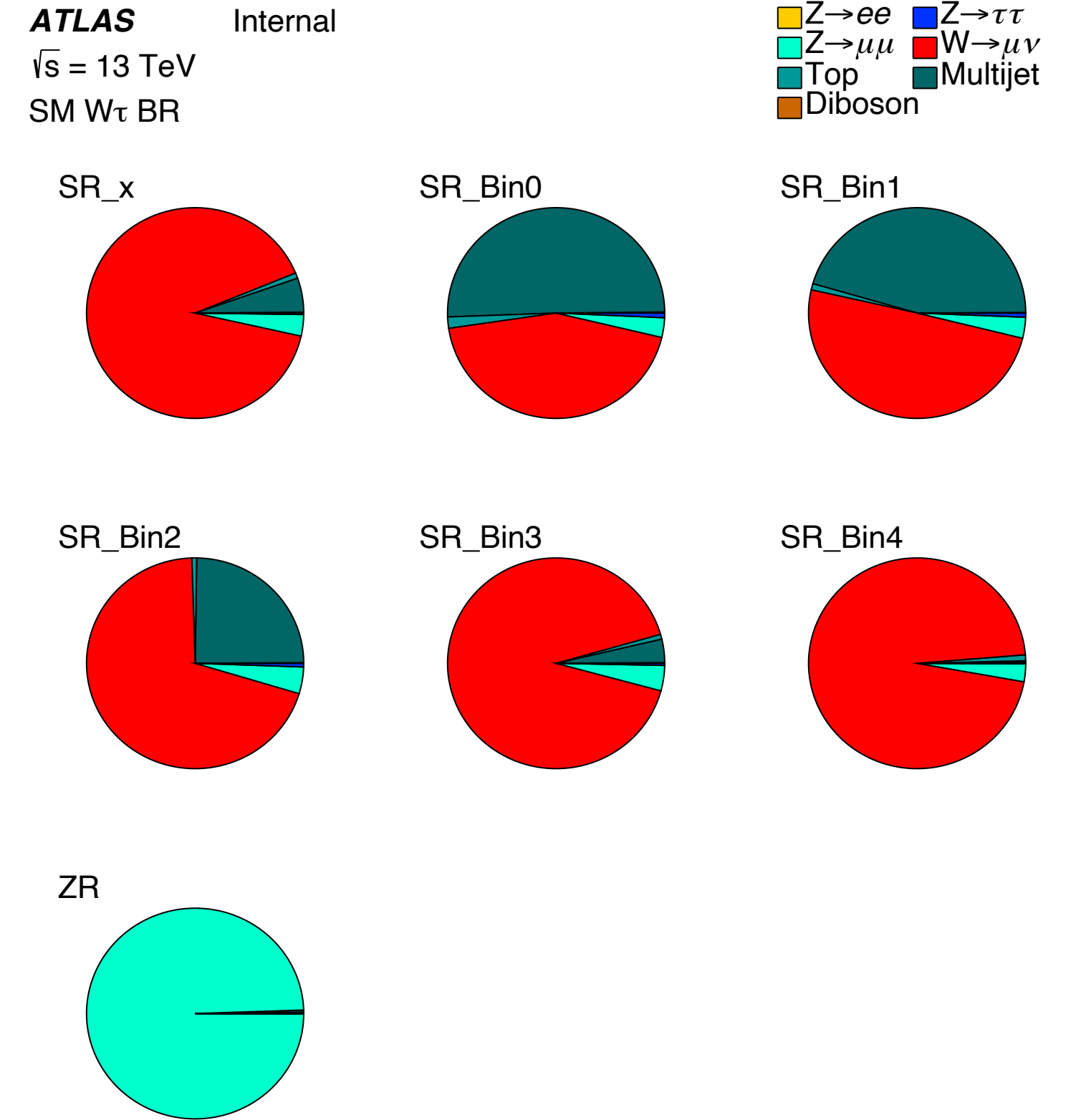
m_T



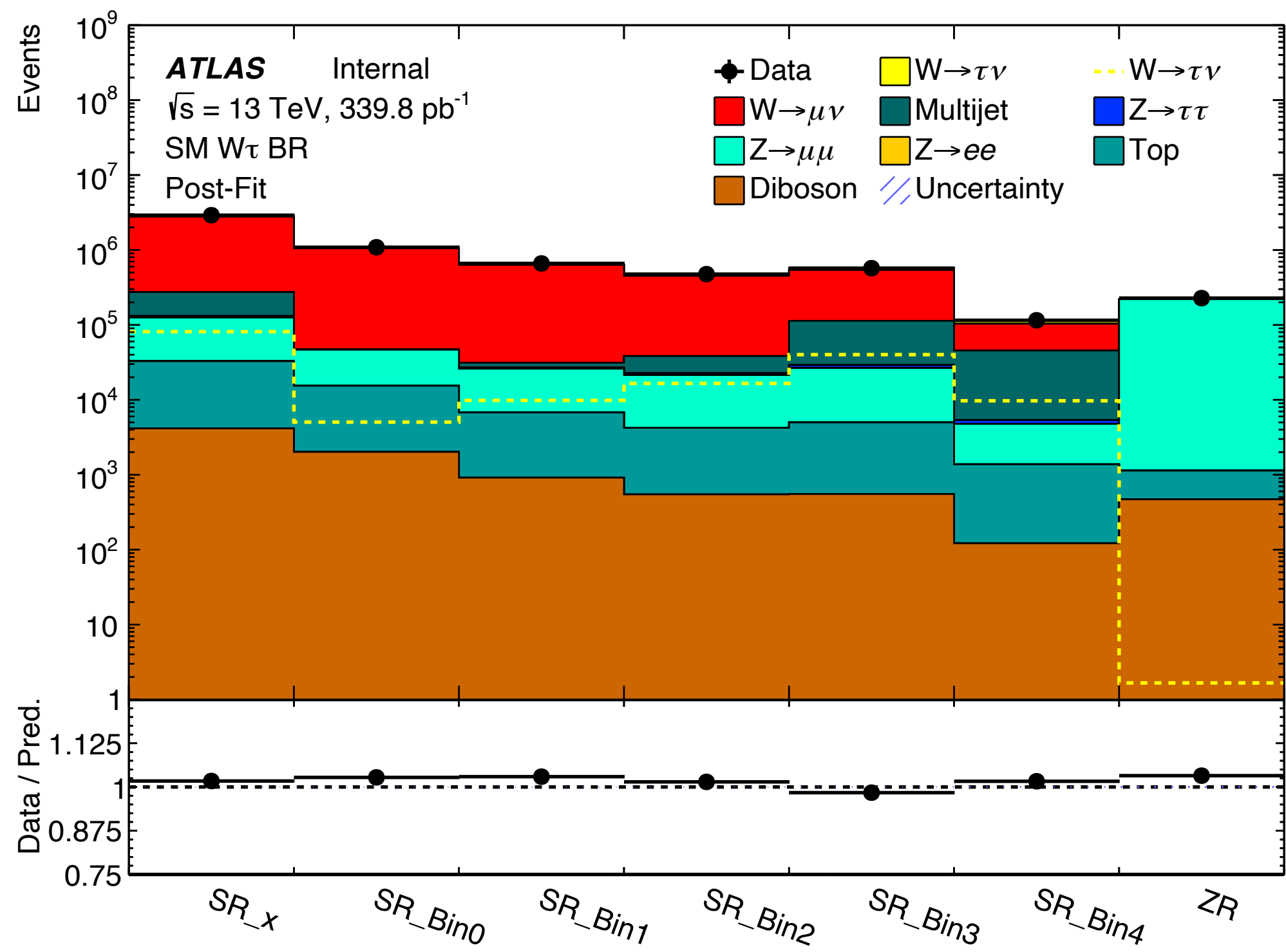
ATLAS Internal
 $\sqrt{s} = 13 \text{ TeV}, 339.8 \text{ pb}^{-1}$
 SM $W\tau$ BR



ATLAS Internal
 $\sqrt{s} = 13 \text{ TeV}$
 SM $W\tau$ BR

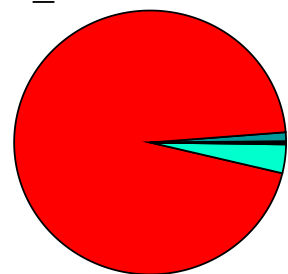


BDT

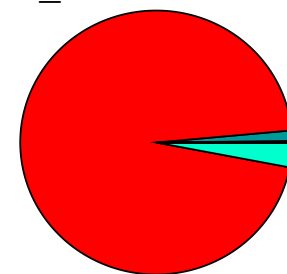


ATLAS Internal
 $\sqrt{s} = 13 \text{ TeV}$
SM $W\tau$ BR

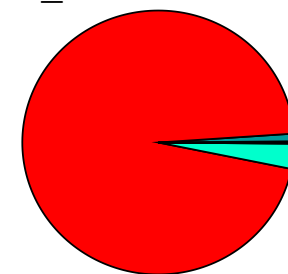
SR_x



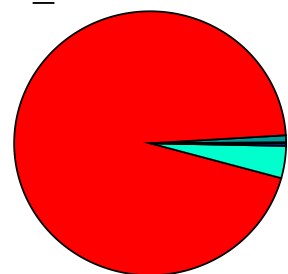
SR_Bin0



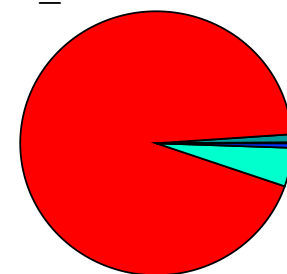
SR_Bin1



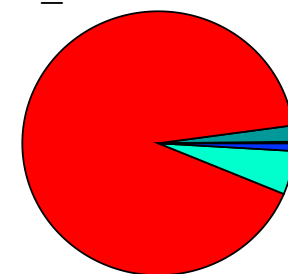
SR_Bin2



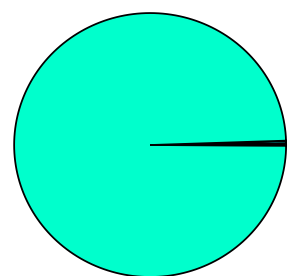
SR_Bin3



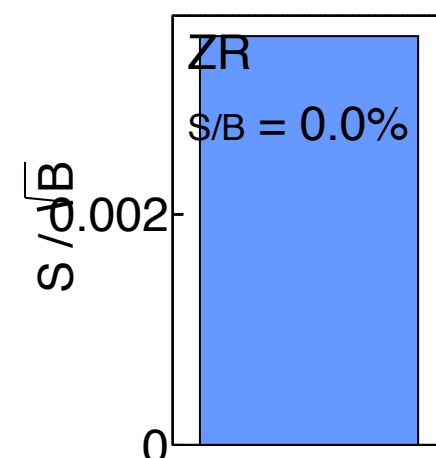
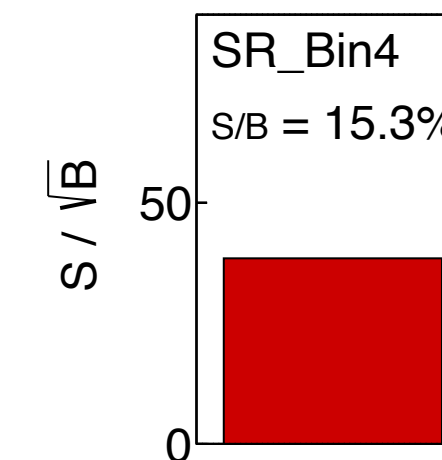
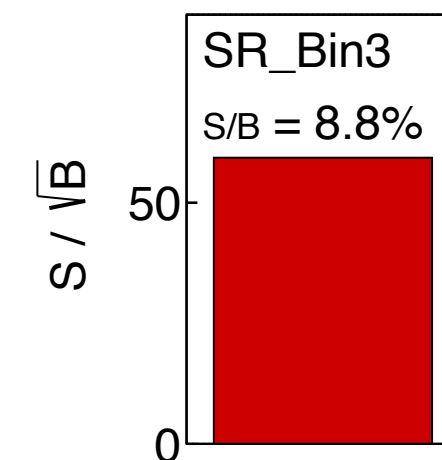
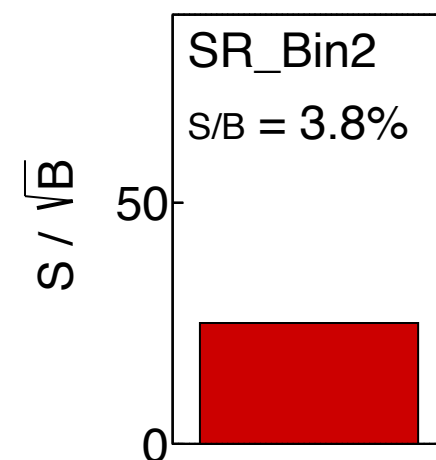
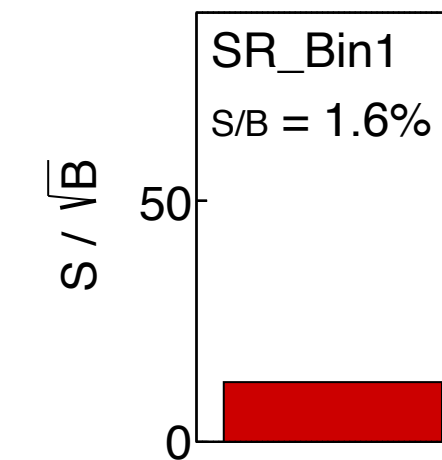
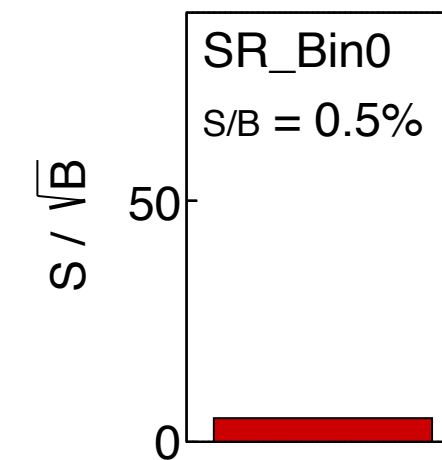
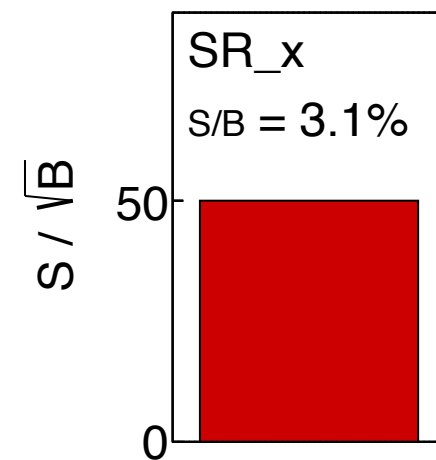
SR_Bin4



ZR



ATLAS Internal
 $\sqrt{s} = 13 \text{ TeV}$, 339.8 pb^{-1}
SM $W\tau$ BR



Прицельный параметр d0

