

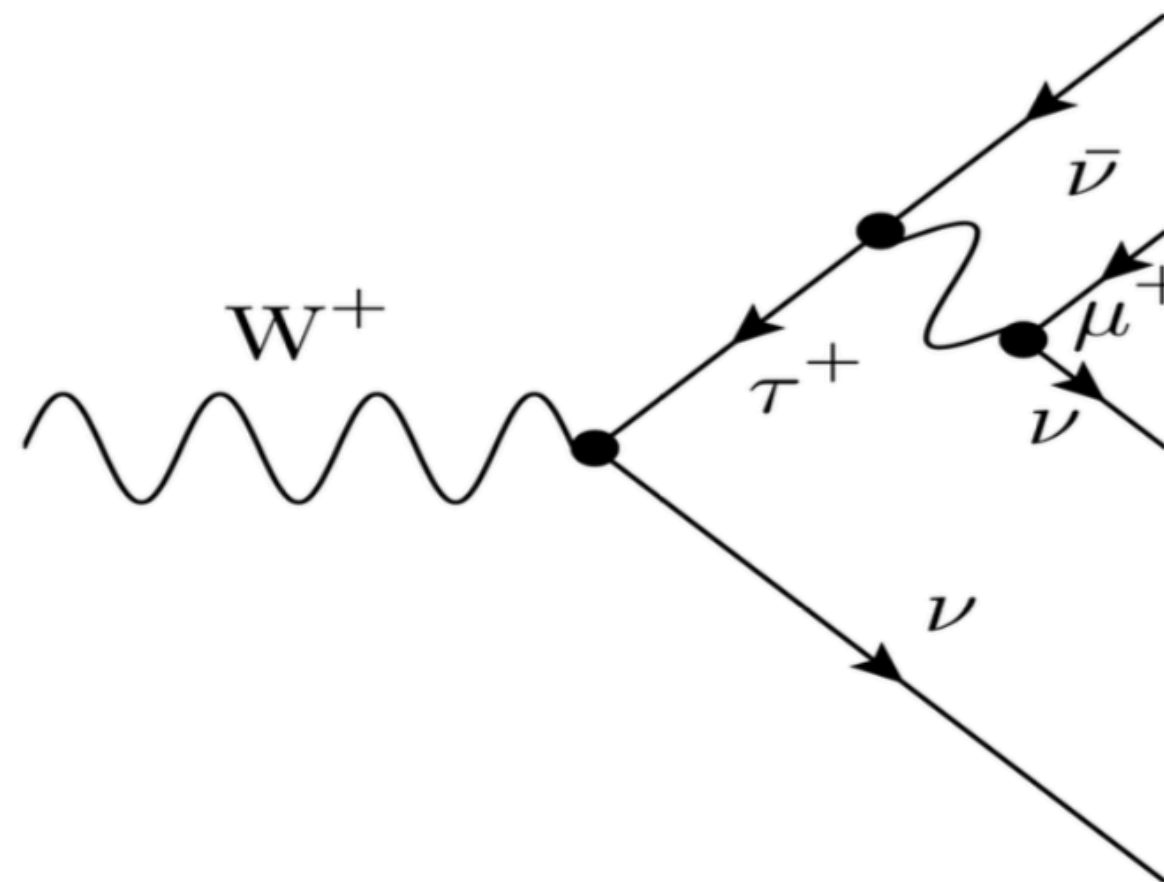
Классификация лептонных распадов W бозона методами машинного обучения в р-р столкновениях при $\sqrt{s} = 13$ ТэВ в эксперименте ATLAS.

Научный руководитель: Пономаренко Д.Е.
Выполнил: Толкачев Г.А.

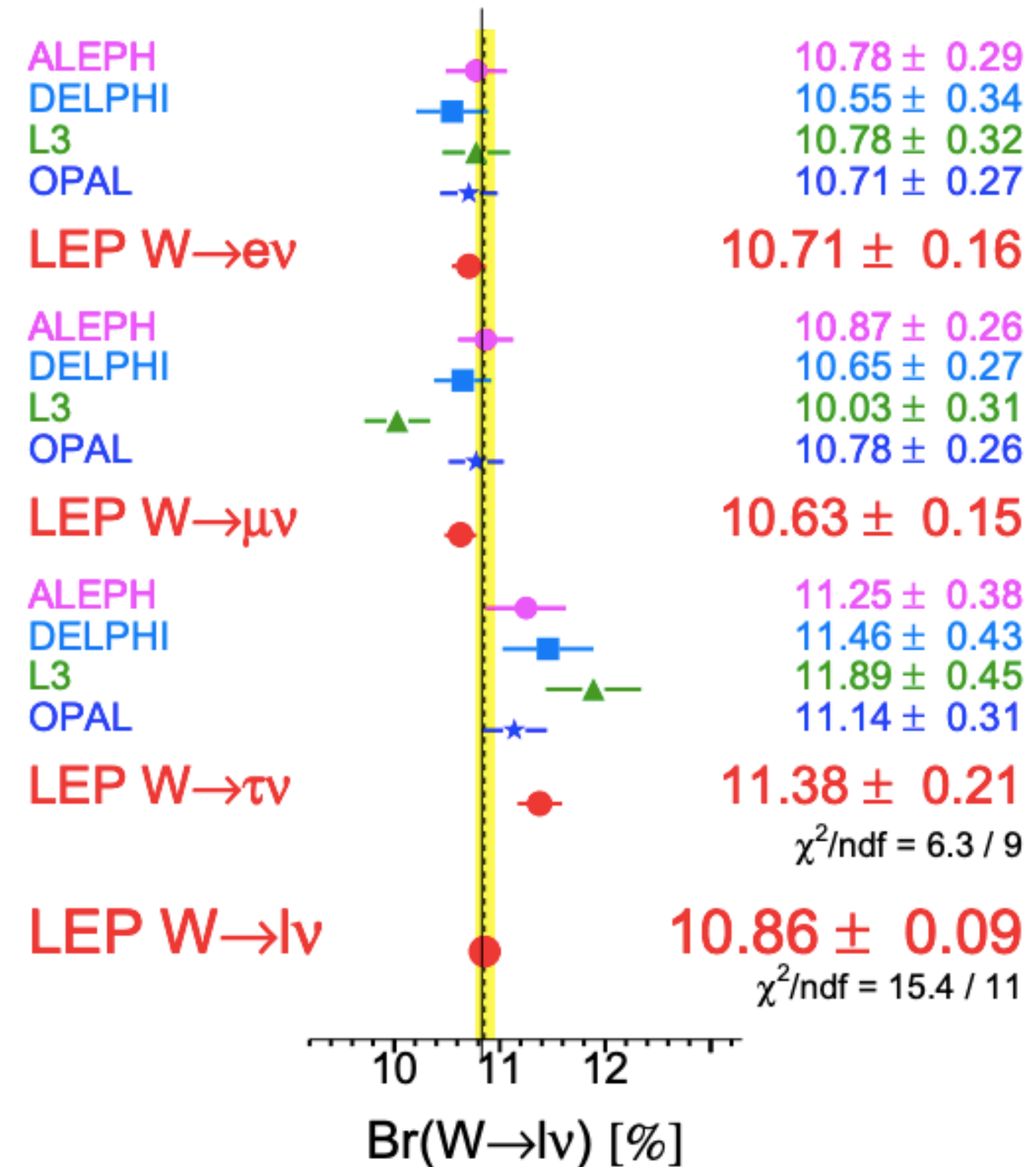
28.12.2020

Мотивация

- В анализе данных с LEP имеется указание на возможное отклонение в отношении отношений сечения двух процессов лепленного распада W бозона $Br(W \rightarrow \tau\nu \rightarrow \mu\nu\nu)/Br(W \rightarrow \mu\nu)$
- 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 найти оптимальную модель машинного обучения BDT для выделения сигнальных событий распада W бозона в τ .



W Leptonic Branching Ratios



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 условия реальных протон - протонных столкновений эксперимента ATLAS

Список Монте-Карло используемых в работе.

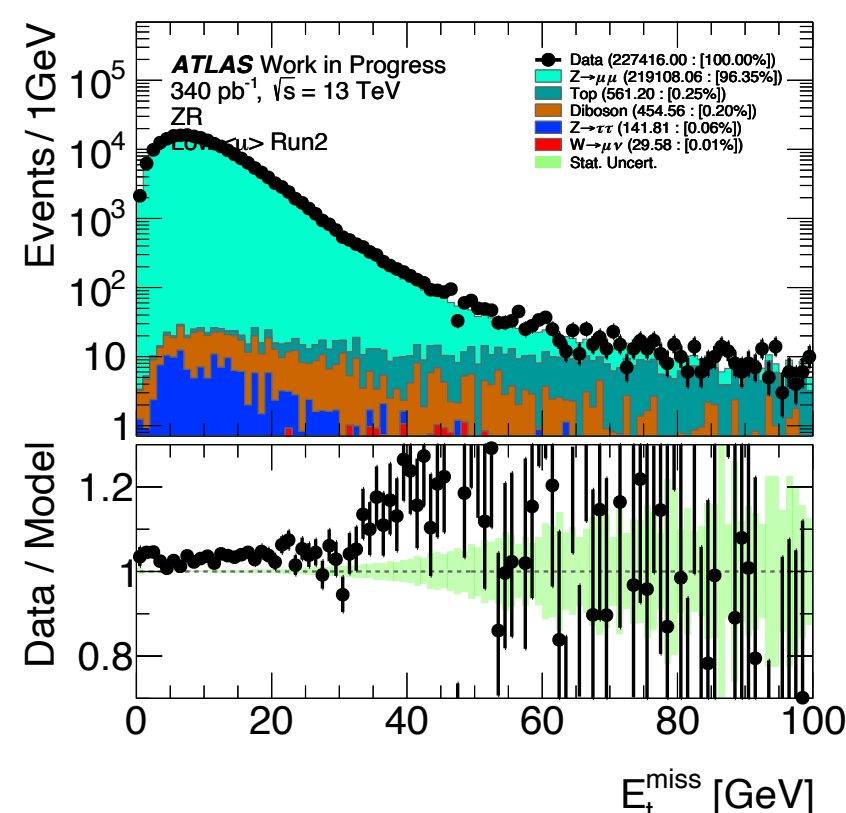
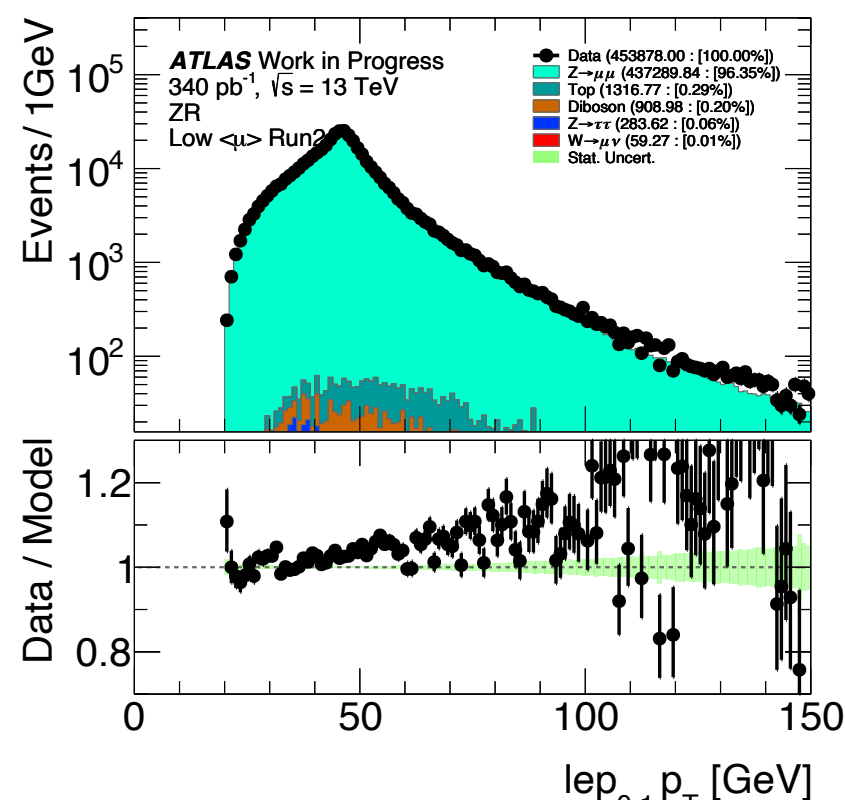
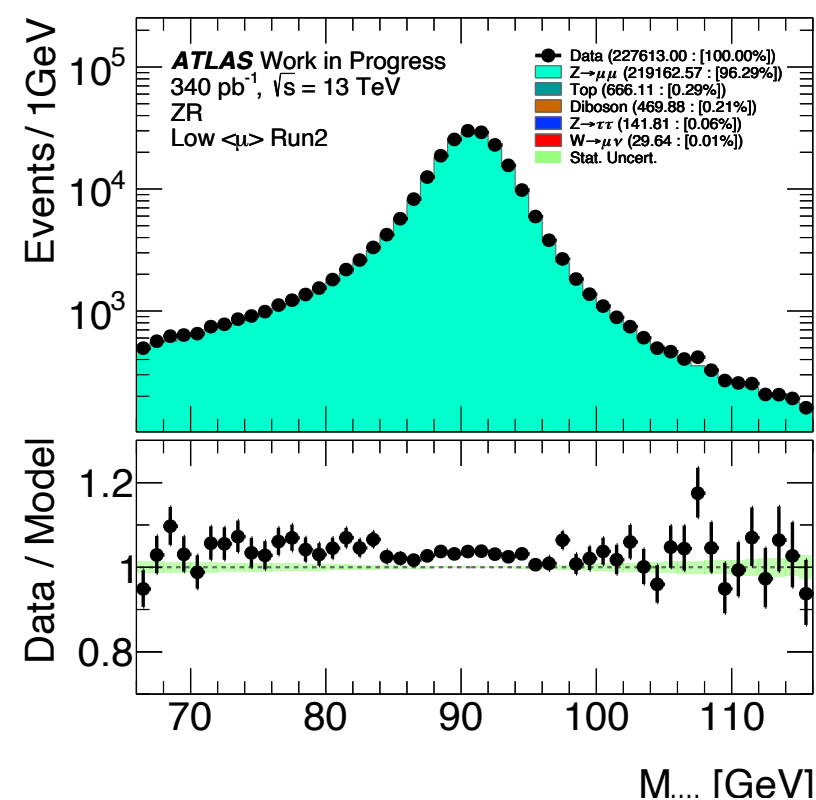
Sample	DSID	Generator	xs [pb]
$W^+ \rightarrow \mu \nu$	361101	PowhegPythia8EvtGen	11500.9
$W^- \rightarrow \mu \nu$	361104	PowhegPythia8EvtGen	8579.31
$W^+ \rightarrow \tau \nu$	361102	PowhegPythia8EvtGen	11500.9
$W^- \rightarrow \tau \nu$	361105	PowhegPythia8EvtGen	8579.31
$Z \rightarrow \tau \tau$	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

Контрольные регионы

Z регион

Кинематический отбор Z региона

P_T	M_{ll}	ID	Isolated
$> 20 \text{ ГэВ}$	$66 \text{ ГэВ} < m < 116 \text{ ГэВ}$	tight	medium

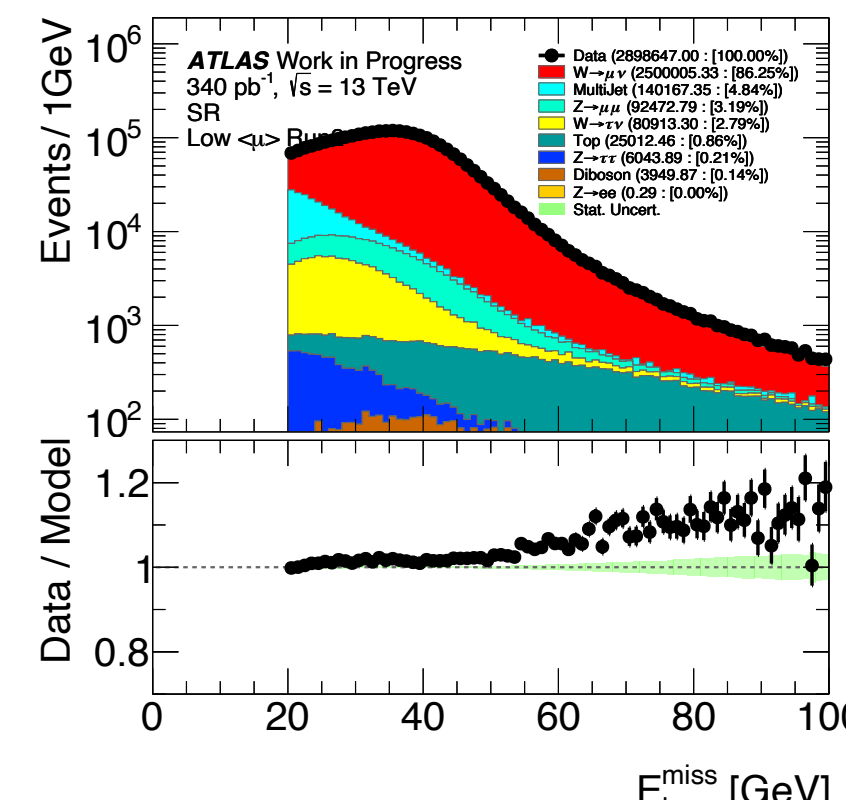
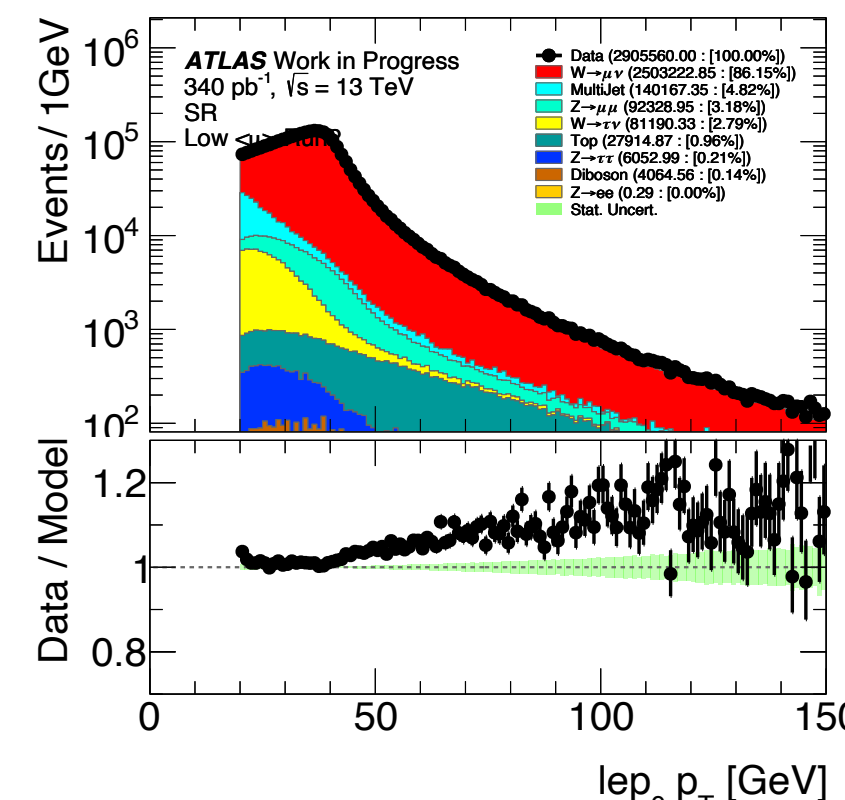
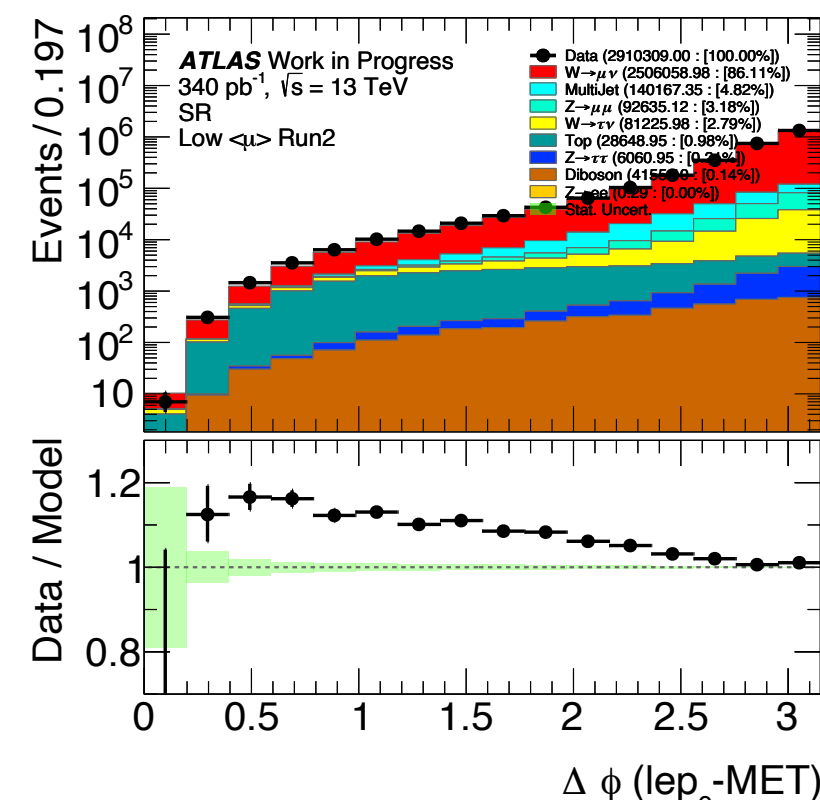


- Является контрольным регионом
- Экспериментальные данные хорошо согласуются с Монте-Карло данными.
- Необходим для проведения валидации сигнального региона и уменьшения статистической ошибки

Сигнальный регион

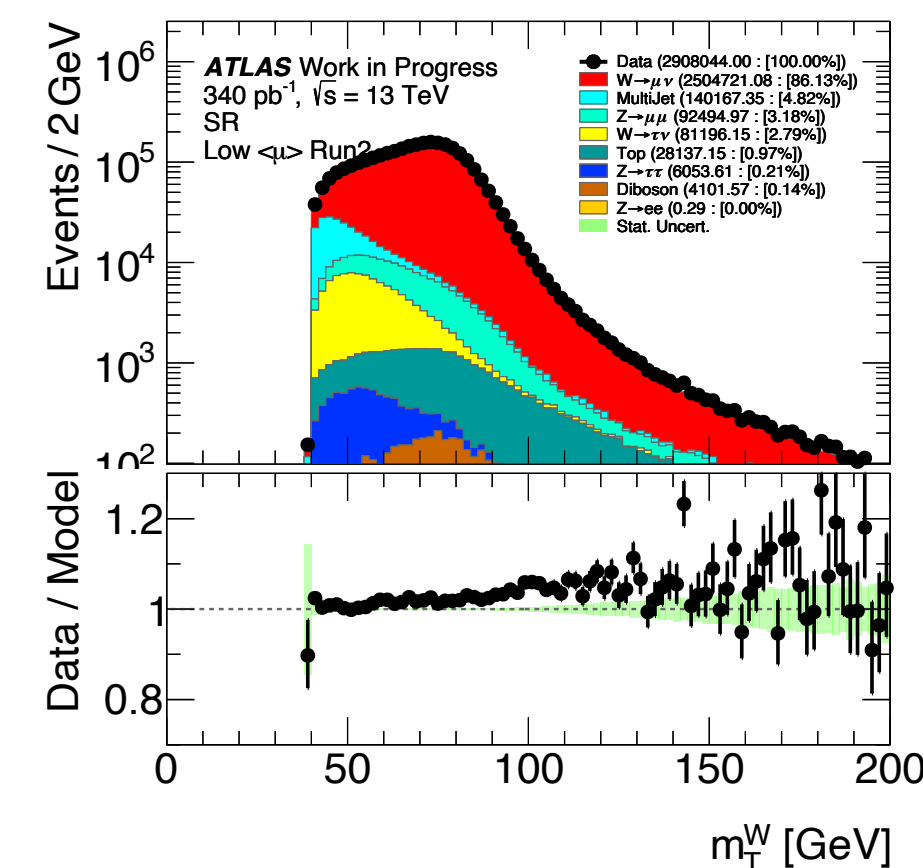
Кинематический отбор сигнального региона

P_T	E_T^{miss}	m_T	N lep	ptvarcone20/ P_T	ID	Isolated
$> 20 \text{ ГэВ}$	$> 20 \text{ ГэВ}$	$> 40 \text{ ГэВ}$	1	< 0.1	tight	medium



$$m_T^W = \sqrt{2P_T(l)P_T(\nu)(1 - \cos(\phi(l) - \phi(\nu)))}$$

- Содержит события с распадами W в τ
- С помощью данных сигнального региона получают отклик, необходимы для проведения измерений лептонной универсальности.
- Причиной отклонения Монте-Карло от экспериментальных данных в жесткой части P_T , при больших E_T^{miss} и m_T^W связано с плохим моделированием в генераторе PowhegPythia.

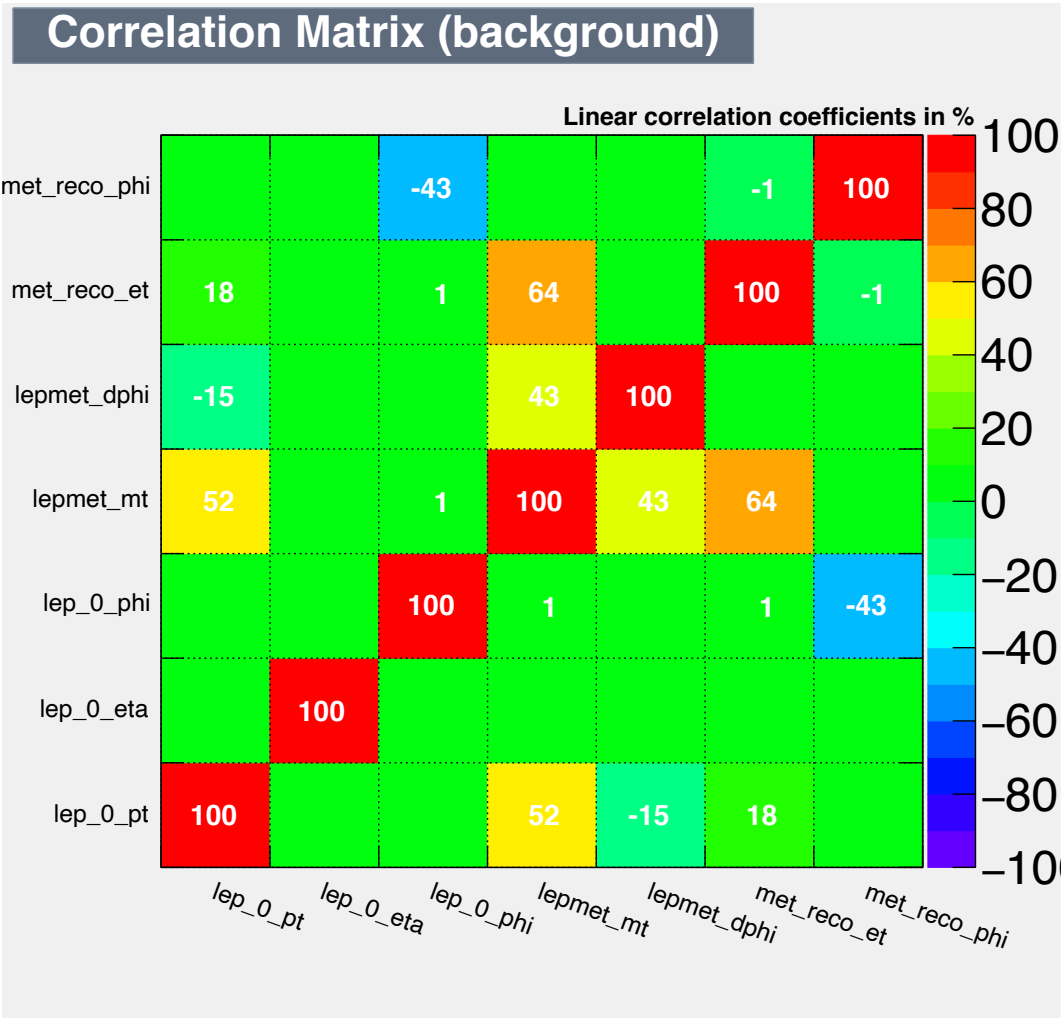
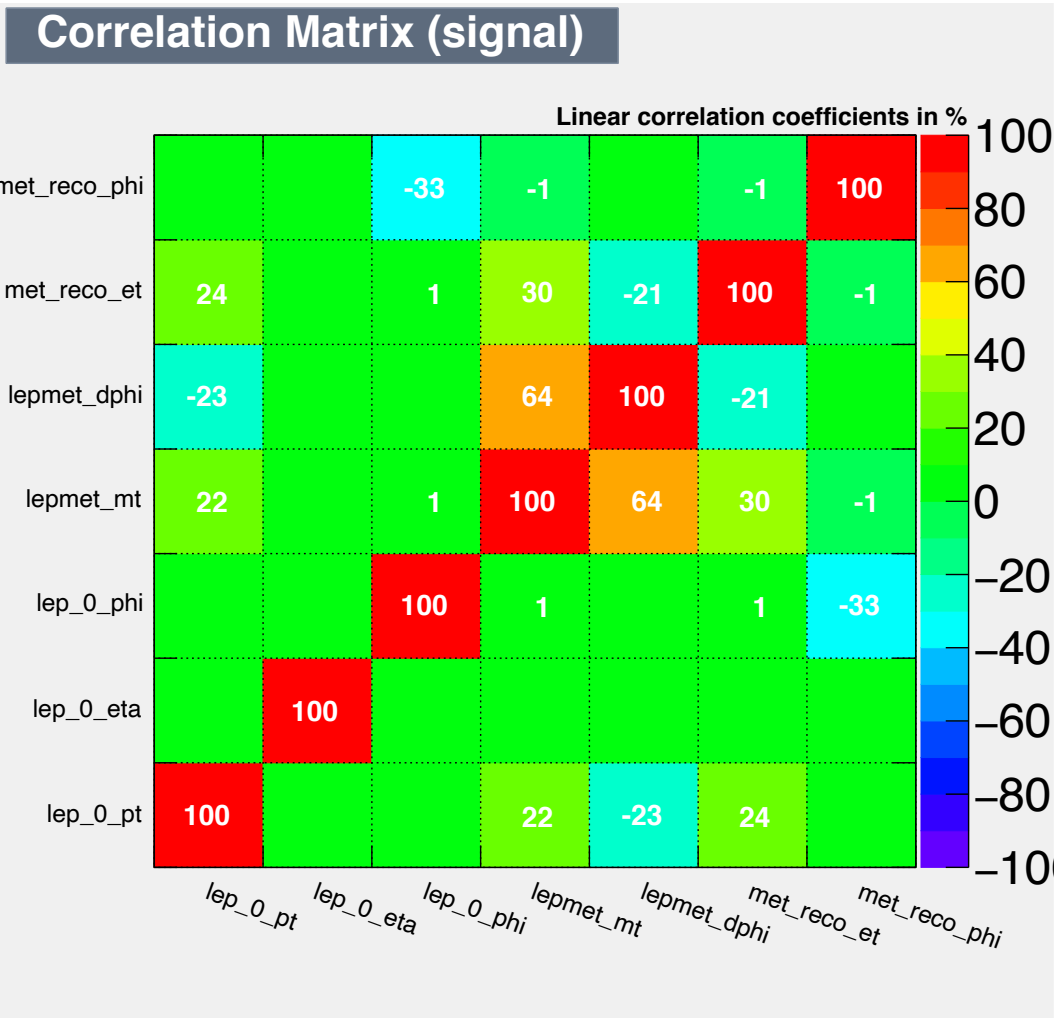


Обучение классификатора BDT

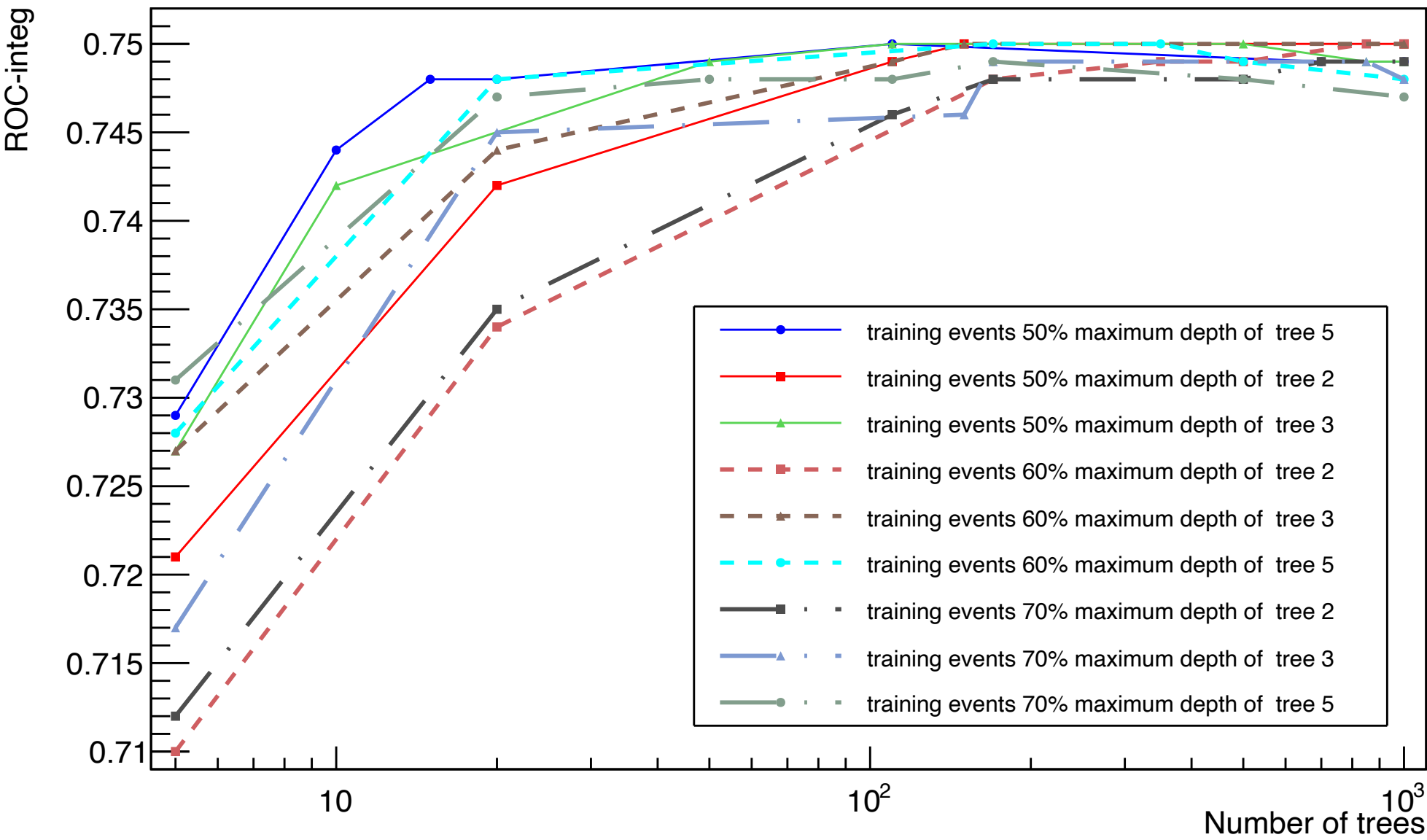
Значимость кинематических переменных используемых в обучении

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



Результат работы предыдущего семестра



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:

$$W \rightarrow \mu \nu$$

$$Z \rightarrow \mu \mu$$

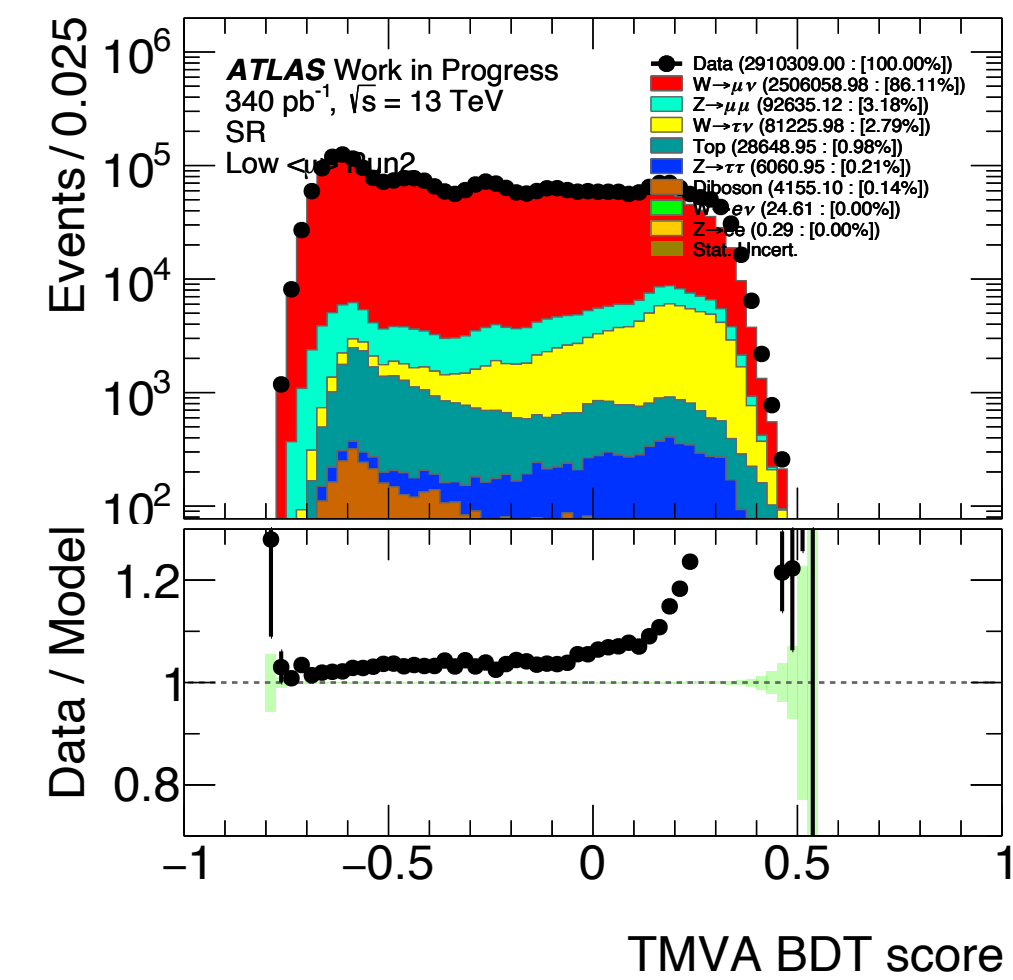
Распады t кварка

Дибозонные процессы

Значение интеграла под ROC кривой - 0.795

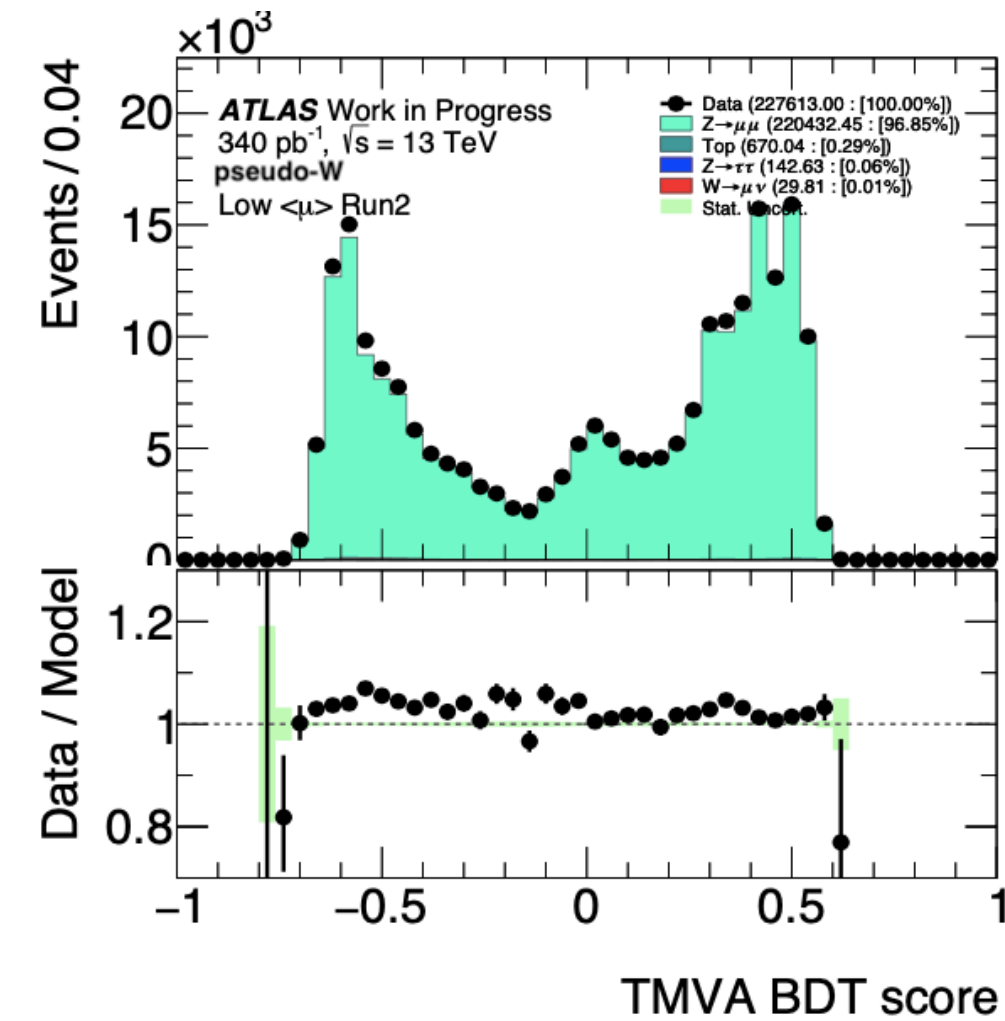
Результаты классификации событий

Классификация данных сигнального региона без учета оценки КХД фона



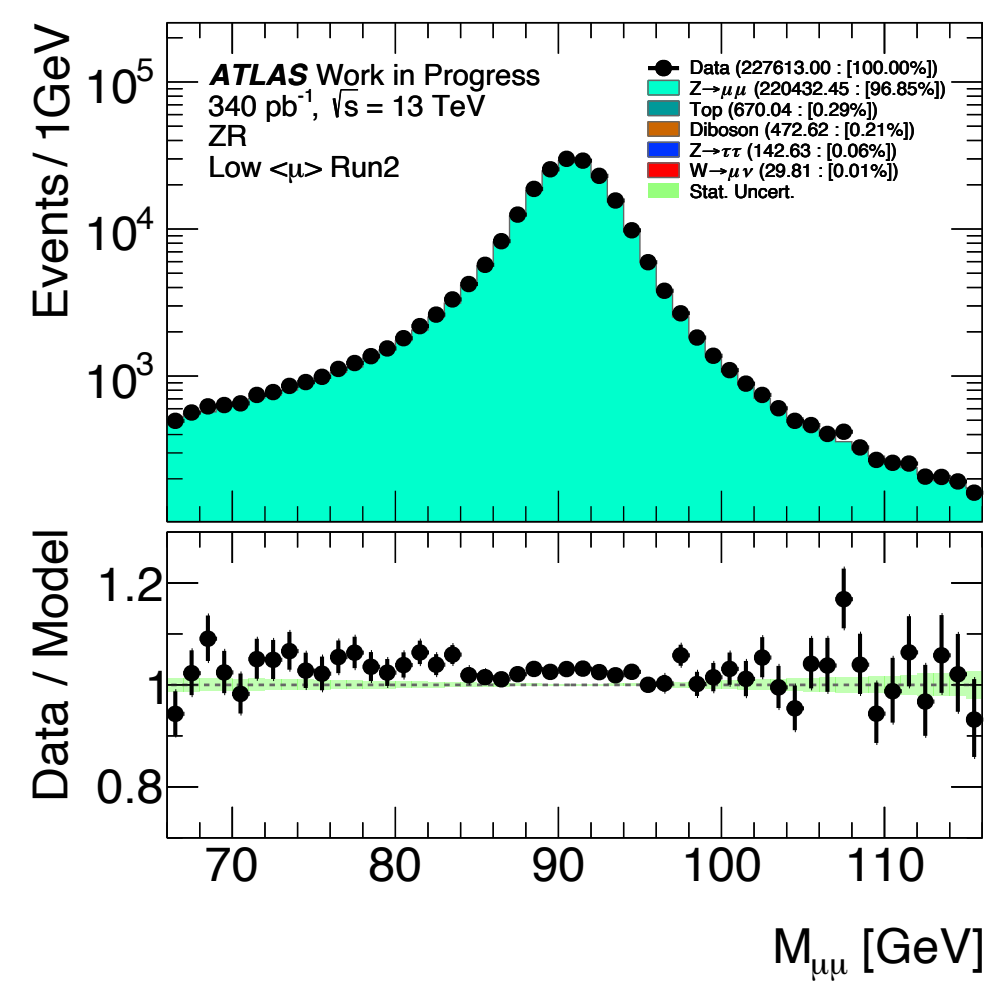
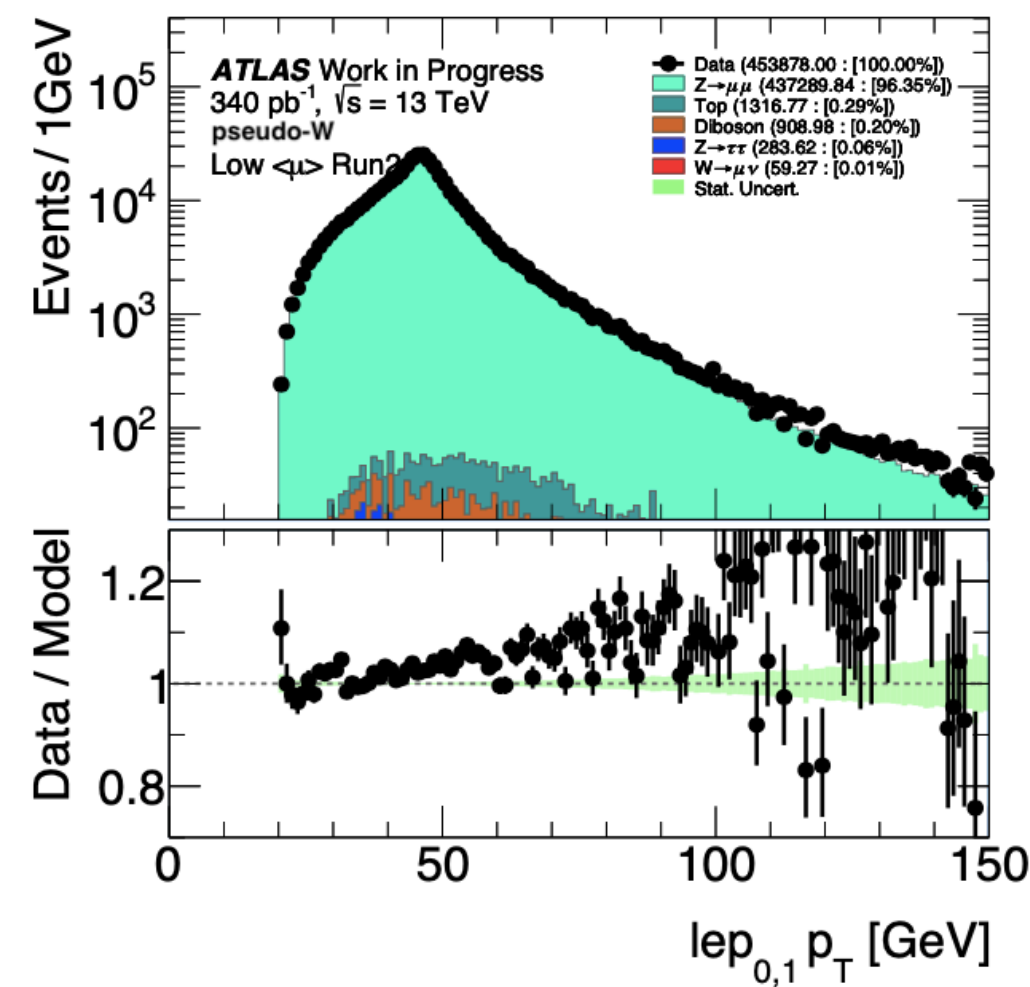
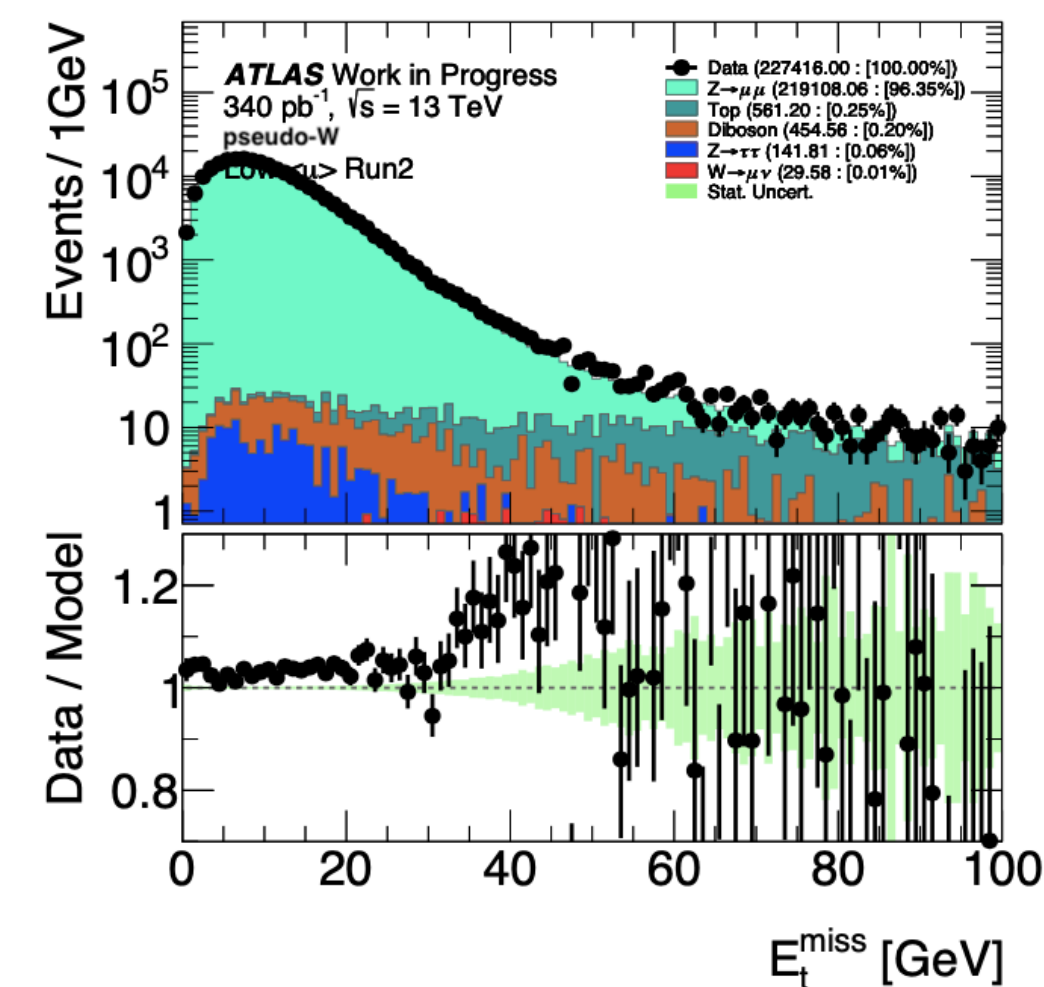
- Отклонения в области (-0.9, 0) связано с плохим моделированием Монте-Карло.
- Возможной причиной отклонений в области (0, 0.5) является неучтенный КХД фон.
- Для подтверждения данной гипотезы необходимо провести классификацию данных псевдо-W региона.

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, 0.5) с большой вероятностью является КХД фон

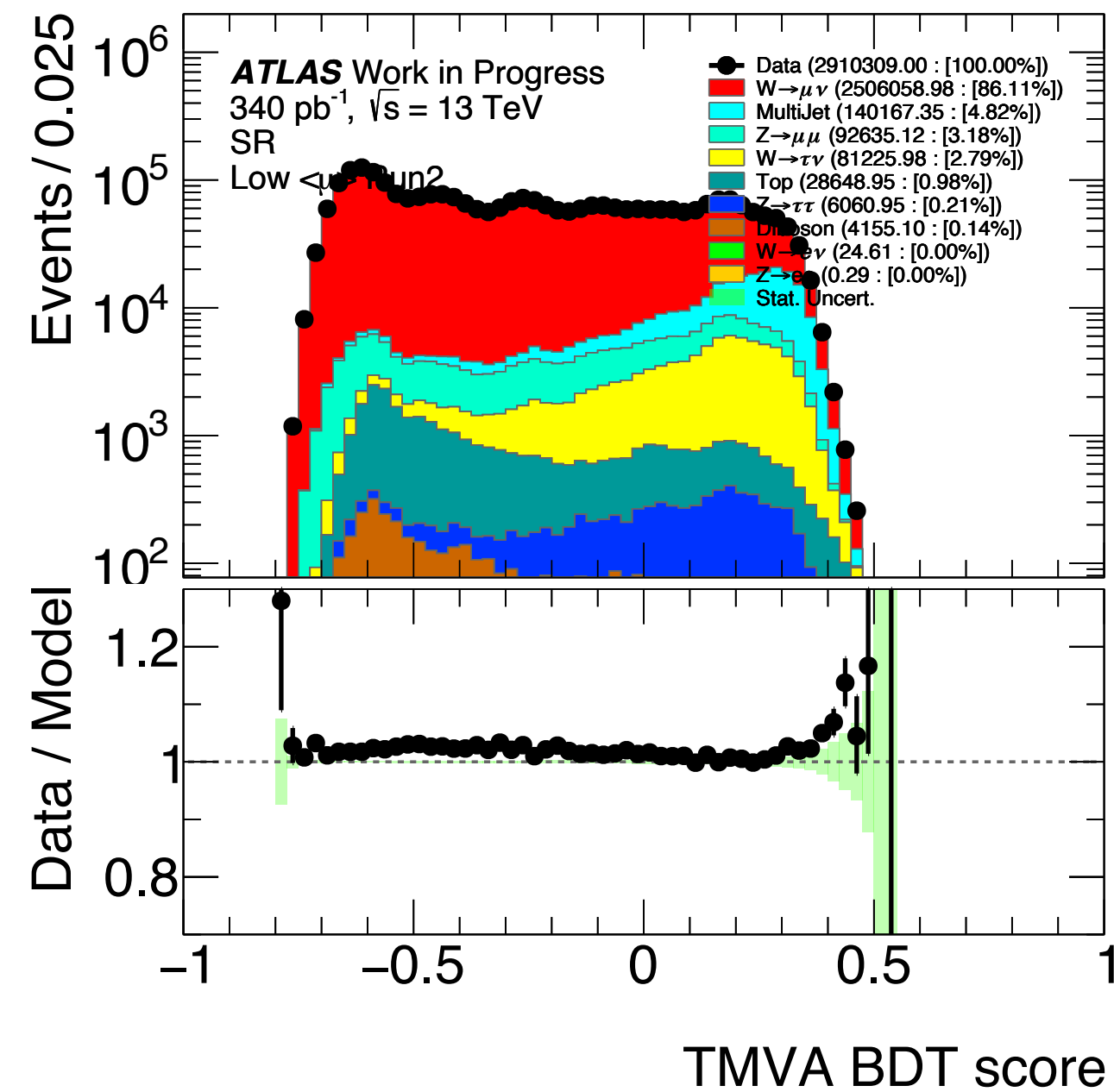
Псевдо-W регион



- Искусственно созданный регион на основе Z региона.
 $E_{T,pseudoW}^{miss} = E_T^{miss} + rand(lep),$
- Кинематический отбор Z региона.
- Имитируем сигнатуру сигнального региона, сохранения хорошее соответствие Монте-Карло и экспериментальных данных.

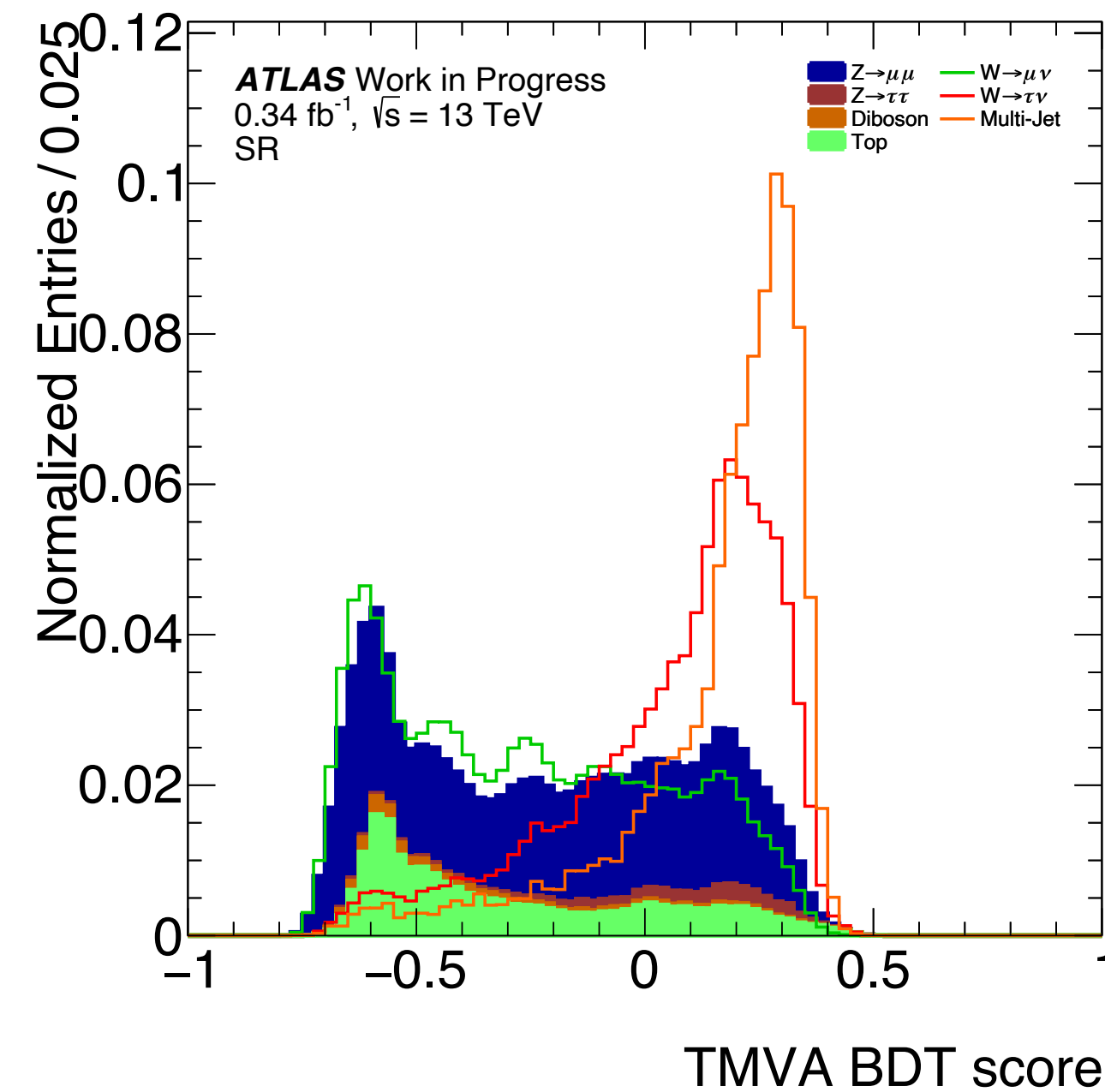
Результаты классификации событий

Классификация данных сигнального региона с учетом оценки КХД фона



- Отклонения в области (0.3, 0.5) может быть связано как с плохим моделированием Монте-Карло, так и с плохой оценкой КХД фона.
- Для выяснение причины отклонения необходимо сделать ограничение на переменную скоринга данных сигнального региона.
- Исследовать кинематические переменные с учетом данного ограничения

Сравнение откликов классификатора , после нормировки на интеграл.



- Видно явное отличие сигнала от фона.
- Форма распределения отклика для сигнала и КХД похожа.
- Форма распределений фоновых процессов тоже схожи.
- Модель BDT хорошо справляется с задачей классификации.
- Использование распределений отклика вместо поперечной поперечной массы m_T^W может улучшить точность измерения лептонной универсальности в базовом анализе.

Заключение

- Построены кинематические распределения для Монте-Карло данных и реальных данных 2017 и 2018 года с энергией 13 ТэВ и светимостью 340 пБ^{-1} эксперимента ATLAS. Показано их расхождение, в том числе объясняемое плохим моделированием Монте-Карло данных.
- Выполнена тренировка модели BDT, параметры которой были взяты из работы прошлого семестра. Получен отклик классификатора событий для экспериментальных данных и данных Монте-Карло сигнального региона без оценки КХД фона. Исследованы наблюдаемые расхождения между Монте-Карло и экспериментальными данными с помощью псевдо- W региона. После предоставления КХД фона в данный анализ была проведена классификация событий с учетом КХД фона. Сделан вывод о оставшемся расхождении в области от 0.3 до 0.5. Для выяснения причины расхождения планируется исследование кинематических переменных с добавлением ограничения на переменную отклика модели.
- Изучены пакеты для проведения анализа данных xTauReader и HAPPy.
- Выполнен ряд работ, которые связаны с модернизацией фреймворка xTauReader. За счет данных улучшений удалось значительно повысить производительность на этапе классификации данных. Была также устранена ошибка, связанная с переполнением оперативной памяти, при использовании библиотеки ROOT::TMVA.
- В качестве следующего шага работы планируется использование полученного отклика модели BDT в измерении лептонной универсальности и проверки его на влияние точности измерения.

Спасибо за внимание!

