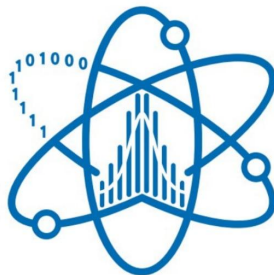


НЕР Анализ данных

Томский Государственный Университет

16 июня 2025



Направления

- **Повторение анализа на ATLAS Open Data:**
 - *Измерение сечений рождения пар топ-кварков ($t\bar{t}$) и Z-бозонов на данных 2015 года ($\sqrt{s} = 13$ ТэВ)*
[\[https://arxiv.org/abs/1612.03636\]](https://arxiv.org/abs/1612.03636)
- **Рождение и дифференциальные распределения J/Ψ, ω и φ мезонов** в протон-протонных столкновениях на Большом адронном коллайдере (выпускная квалификационная работа Ильи Полищука под руководством Ирины и Вадима).
- **Оценка эффективности алгоритмов машинного обучения** для задач классификации в анализах данных физики высоких энергий.

Измерение сечений рождения пар топ-кварков ($t\bar{t}$) и Z-бозонов на данных 2015 года

.....

Завершён продакшн **ntuples** с ослабленной преселекцией, чтобы сохранить всю информацию, релевантную **для всех анализируемых регионов** (SR, VR, CR), а также для оценки вклада QCD:

- включены все данные и MC-сэмплы, систематики активированы
- суммарное время подготовки ntuples — около одной недели

Базовая преселекция:

- Триггеры
- Pileup reweighting
- **runEventCleaning: True** — удаление событий без primary vertex, плохих джетов и др.
- Электроны:
 - идентификация **LooseLH**, изоляция **Loose_VarRad**
 - Отбор: $p_T > 10$ ГэВ, $|\eta| < 2.47$
- Мюоны:
 - Качество: **Loose**, изоляция **Loose_VarRad**
 - Отбор: $p_T > 10$ ГэВ
- Overlap Removal

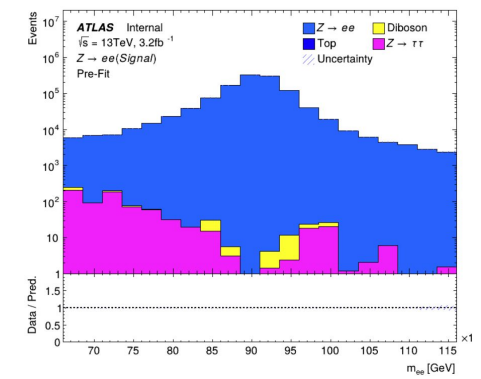
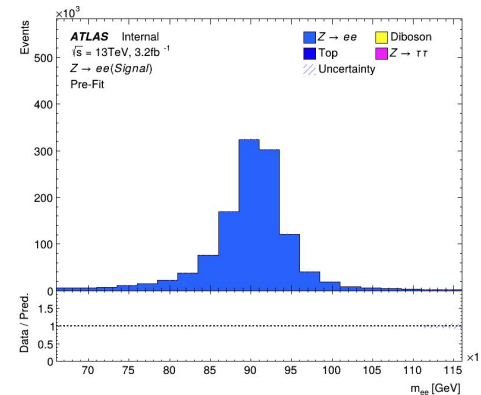
Измерение сечений рождения пар топ-кварков ($t\bar{t}$) и Z-бозонов на данных 2015 года

Настроена **конфигурация** TRexFitter:

- Определены SR/VR/CR-регионы, настроены входные ntuple-файлы
- Использованы выборки:
 - $Z \rightarrow ee$ (сигнал), многобозонный фон, $t\bar{t}$, $Z \rightarrow \tau\tau$
- Включена систематика по люминесценции (0.83%).

Проверена полная цепочка анализа на небольшой статистике — процедура отработывает корректно:

- **создание гистограмм**, которые используются TRexFitter для построения модели правдоподобия (likelihood fit)
- формирование объекта **RooWorkspace**, используемого для проведения фитинга
- генерация **pre-fit гистограмм** (распределения данных и MC до фитинга) и проверка графиков
- запуск процедуры **фитинга**
- получение **post-fit** распределений (данные и MC после фитинга)



Построение гистограмм с **Asimov** данными

Измерение сечений рождения пар топ-кварков ($t\bar{t}$) и Z-бозонов на данных 2015 года

В процессе:

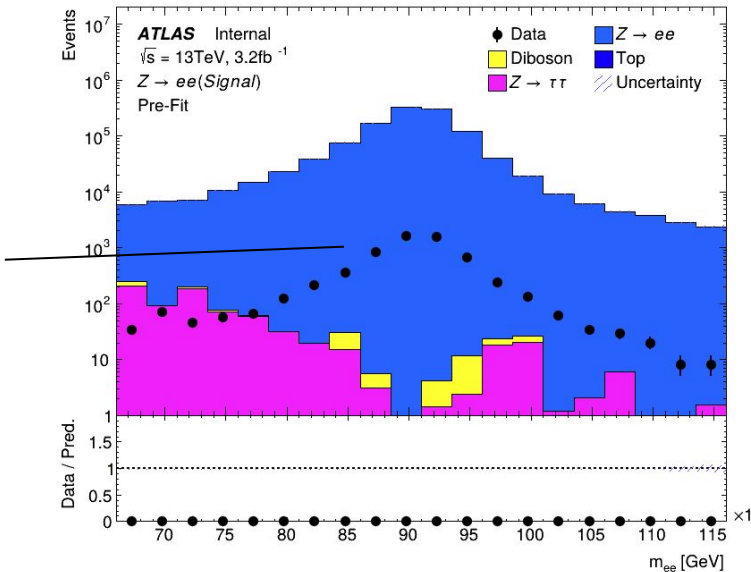
- подготовка гистограмм на **полной статистике** с корректным включением весов: **нормировка по сечению, pile-up, светимость**

Следующий этап:

- Включение систематик в анализ и их пошаговая отладка.
- настройка фита

Форма МС соответствует данным, но:

- **МС сильно выше данных по норме** во всём диапазоне



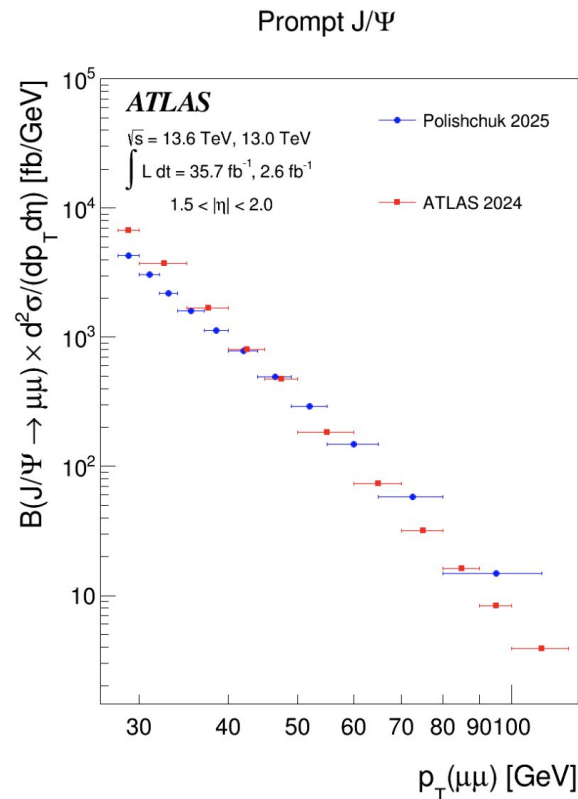
Рождение и дифференциальные распределения J/Ψ , ω и ϕ мезонов

- Использованы данные ATLAS за 2022 год.
- Анализ проводится по мюонным каналам: $J/\psi \rightarrow \mu^+\mu^-$, $\omega \rightarrow \mu^+\mu^-$, $\phi \rightarrow \mu^+\mu^-$
- Рассчитаны дифференциальные и двойные дифференциальные сечения

Сравнение с публикациями:

- Результаты сравниваются с данными ATLAS и ALICE:
 - хорошее соответствие для **вторичных** J/ψ и ω ;
 - расхождение в диапазоне $p_T = 28\text{--}34$ ГэВ для **первичных** J/ψ .

Защита Ильи: 18 июня



Алгоритмы машинного обучения для задач классификации

- Классификация событий на сигнальные и фоновые — ключевая задача большинства анализов на Большом адронном коллайдере (LHC).
- Традиционно используются методы машинного обучения, такие как Boosted Decision Trees (BDT) и нейросети (Neural Networks).
- **Цель:** Оценить эффективность современных методов ML — Transformers, GAN, RNN, GNN, Diffusion Models и др. — для задач классификации в физике частиц.
- Анализ будет проводиться на основе открытых данных ATLAS OpenData и репозитория HEPData
 - Пример первого анализа: *Measurement of the production cross-section for a Higgs boson in association with a vector boson in the $H \rightarrow WW^* \rightarrow \ell\nu\ell\nu$ channel in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*
 - Проверяем MC статистику после применения селекции, чтобы эффективно тренировать ML
- **Ожидаемый результат:** подготовка научной публикации с предварительным названием «Оценка эффективности алгоритмов машинного обучения для задач классификации в анализах данных физики высоких энергий»
- Новый постдок Неда Фироз с 1 июля

Thank you!

Данные Open Data Atlas 2015-2016

Related datasets

Run 2 2015 [proton-proton collision data](#)
[ATLAS DAOD_PHYSLITE format Run 2 2015 proton-proton collision data](#)

Run 2 2016 [proton-proton collision data](#)
[ATLAS DAOD_PHYSLITE format Run 2 2016 proton-proton collision data](#)

MC simulation electroweak boson nominal samples
[ATLAS DAOD_PHYSLITE format MC simulation electroweak boson nominal samples](#)

MC simulation exotic signal samples
[ATLAS DAOD_PHYSLITE format MC simulation exotic signal samples](#)

MC simulation Higgs nominal samples
[ATLAS DAOD_PHYSLITE format MC simulation Higgs nominal samples](#)

MC simulation Higgs systematic variation samples
[ATLAS DAOD_PHYSLITE format MC simulation Higgs systematic variation samples](#)

MC simulation QCD jet nominal samples
[ATLAS DAOD_PHYSLITE format MC simulation QCD jet nominal samples](#)

MC simulation QCD jet systematic variation samples
[ATLAS DAOD_PHYSLITE format MC simulation QCD jet systematic variation samples](#)

MC simulation SUSY signal samples
[ATLAS DAOD_PHYSLITE format MC simulation SUSY signal samples](#)

MC simulation top nominal samples
[ATLAS DAOD_PHYSLITE format MC simulation top nominal samples](#)

MC simulation top systematic variation samples
[ATLAS DAOD_PHYSLITE format MC simulation top systematic variation samples](#)

~9 миллиардов событий (реальные данные и симуляции) при
общем размере ~66.87 ТБ.

Open Data Atlas (<https://opendata.cern.ch/record/80020>)

Формат данных: 33 fb-1 DAOD PhysLight format

- Данные, собранные детектором ATLAS в рамках экспериментов на Большом адронном коллайдере за 2015-2016 год
- 7 групп симулированных физических процессов, включающих номинальные процессы (электрослабые бозоны, Хиггс, QCD-джеты, топ-кварки) и вариации с различными параметрами этих процессов необходимые для оценки систематических ошибок измерений.

Production ntuples

Доступ к системам ATLAS:

- Успешно настроен доступ к инфраструктуре ATLAS с inter-ноды кластера ТГУ
- Отлажена система аутентификации и взаимодействия с основными сервисами ATLAS (Rucio, CVMFS и др.)
- Завершено расширение доступа на все вычислительные ноды — полная интеграция анализа в распределённую среду

Распределённые вычисления:

- Настроена система распределённых заданий HTCondor на кластере ТГУ
- Проведено тестирование отправки задач и мониторинга, адаптирован интерфейс фреймворка под многонодовую архитектуру

Предпродакшн и тестирование:

- Запущен предварительный продакшн ntuples на моделированных данных (MC)
 - Цель: протестировать стабильность цепочки обработки и поведение фреймворка в реальных условиях
-

Измерение сечений рождения пар топ-кварков ($t\bar{t}$) и Z-бозонов

Реализован фреймворк для преобразования данных из формата DAOD PhysLight в удобный и читаемый формат ntuples, пригодный для последующего физического анализа

- Репозиторий: <https://git.hep.tsu.ru/mdidenko/framework/>
- Построен на базе **Athena AnalysisBase v25.2.45**

Выходной ntuple: дерево **analysis**, с сохранением информации по событиям, электронам, мюонам и джетам

- Базовая преселекция
- Подключение триггеров 2015 года (электронные и мюонные цепочки, полное соответствие данным)
- Реализация отбора объектов (электроны, мюоны, джеты)
- Удаление перекрытий (OR) между электронами, мюонами и джетами

Все параметры управления заданы в **конфигурационном YAML-файле**

- Возможности адаптации под конкретные задачи:
 - включение/отключение систематик (**runSystematics: False**)
 - переопределение контейнеров, ID, изоляций и триггеров
 - гибкая настройка выходных переменных и структуры ntuple

Фреймворк может использоваться для различных задач ATLAS без изменения кода — достаточно скорректировать конфигурацию

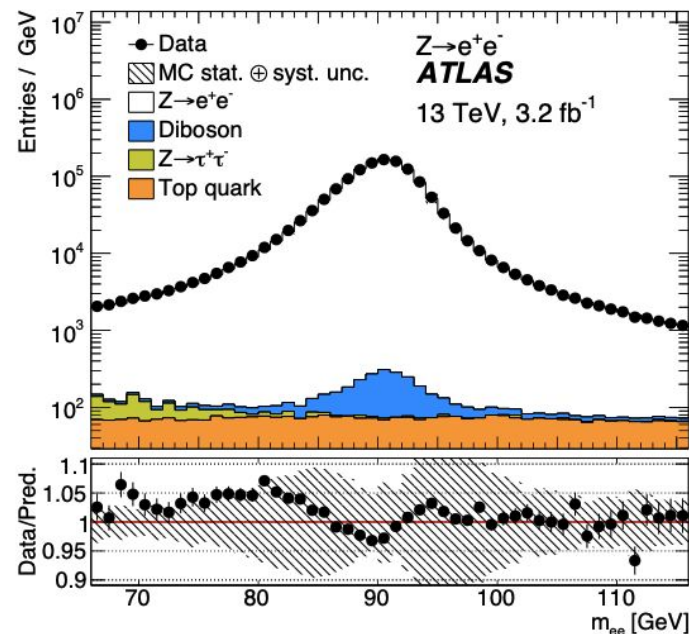
Продакшн на данных Run 2

Готовится первый продакшн на реальных данных 2015 года

- Данные: 3.2 fb^{-1} ($\sim 9 \text{ TB}$)
- MC: $Z \rightarrow ll$ (электроны и мюоны)
- Ожидаемая длительность: 2 недели

Планируемые первые результаты анализа

- Предварительные распределения отношения данных к MC (data/MC)
- Переменная: инвариантная масса Z-бозона
- Включить систематические неопределенности
- Ожидаемое завершение подготовки распределений — через 2 недели



<https://arxiv.org/abs/1612.03636>