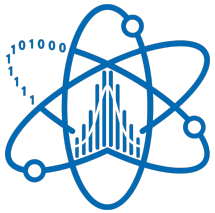
National Research
Tomsk
State
University



Лаборатория
анализа данных
физики высоких энергий
Томского
государственного
университета

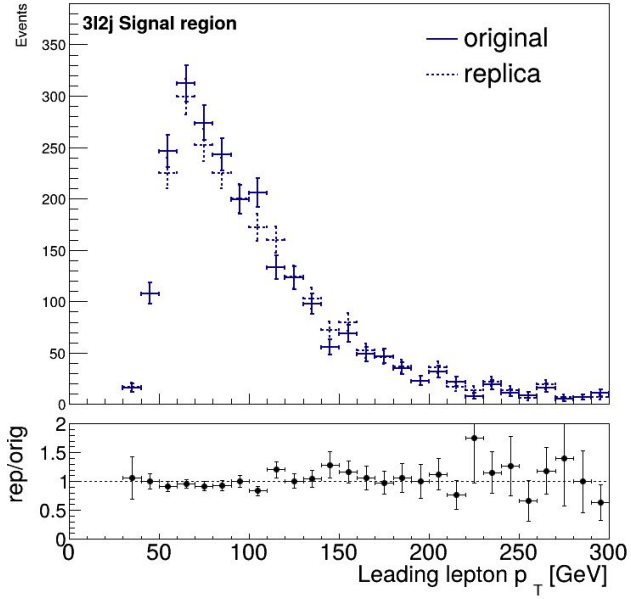
Физический анализ данных

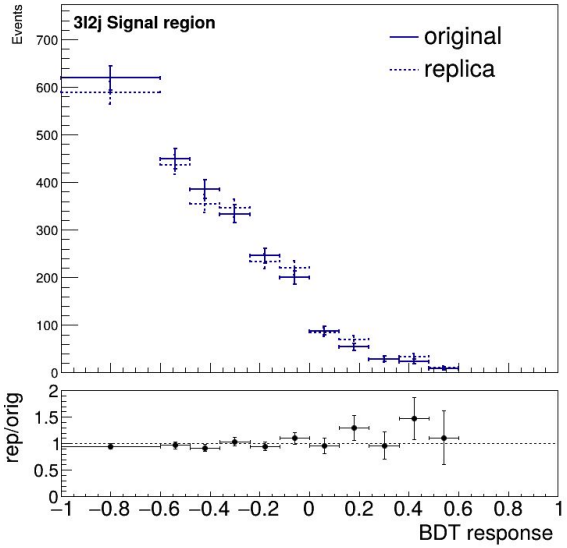
Томский Государственный Университет

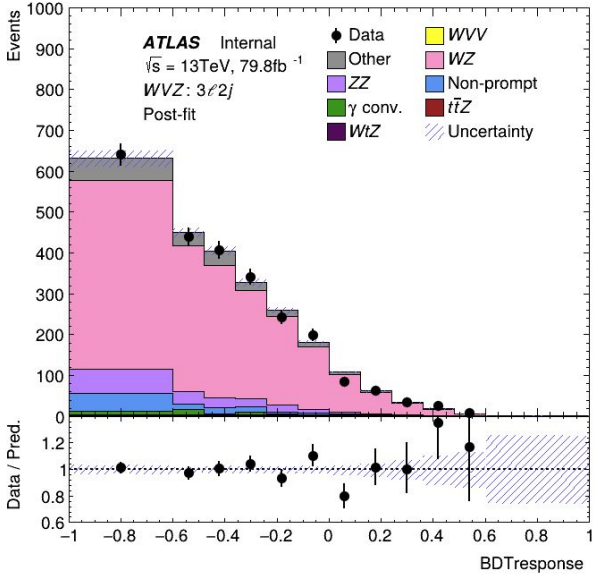
Мария Диденко

Distribution of variables: signal region

- 3l2j SR selection is applied
- Events after SR selection: orig=2438, repl=2407
- The shapes agree within statistical fluctuations (ratio plot: replica/original ≈ 1 ± stat)



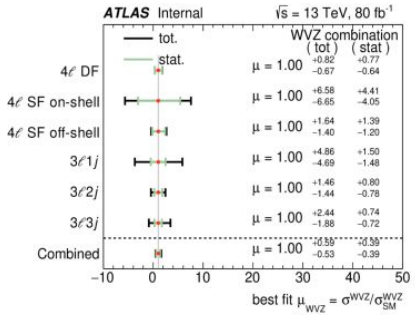


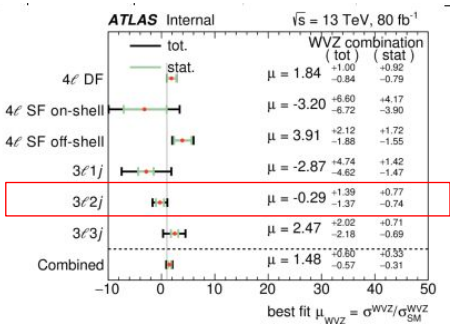


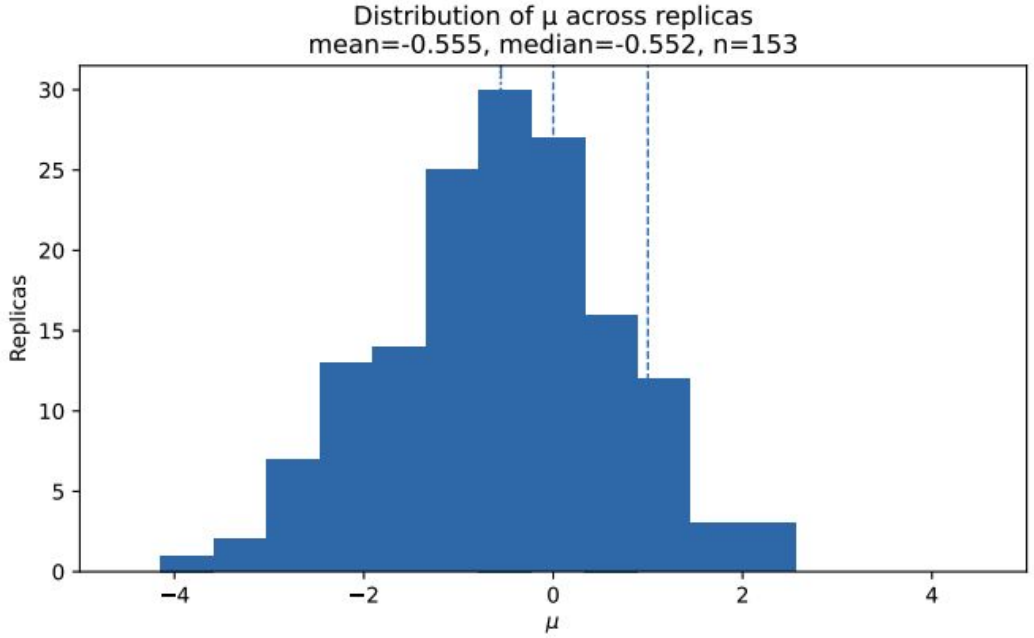
μ distribution

Most of tasks are still running

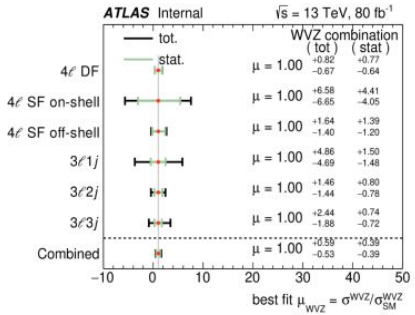
Checked 153 tasks

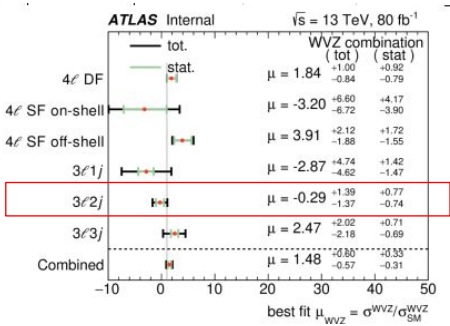


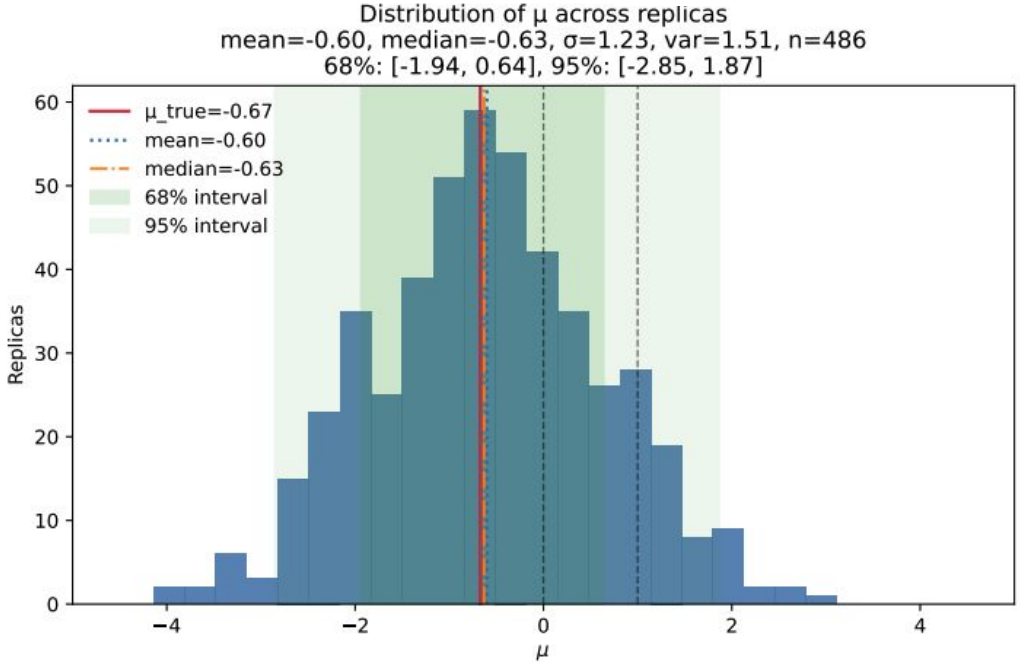




μ distribution







Leptons + at least 3 jets (real data)

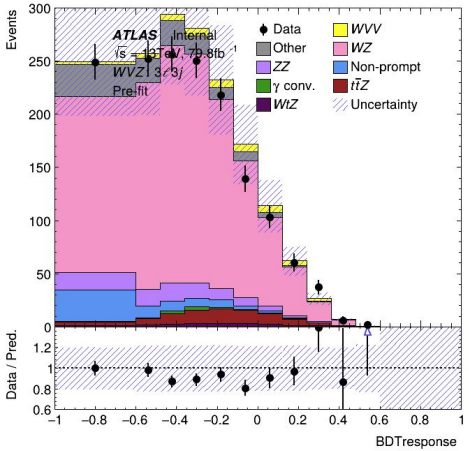
- Checked the remaining **3ℓ + ≥3j** region using **real data**.
- In the configuration file, the variable **newBDTG_15_3l3j** was specified, but it was **missing in the ntuples**.
- Used **newBDTG_32_3l3j_1** from the ntuples instead.
- The obtained signal strength **(μ = 2.57)** is close to the reference value **(μ = 2.47)**

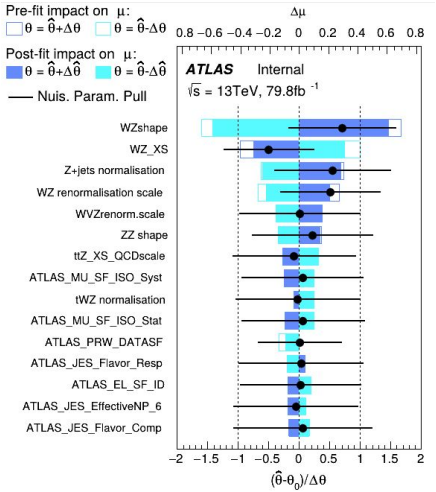
```
###--- 3L preselection && at least 3 jets ---###

Region: three_lep_presel_atLeast_3jets
  Type: SIGNAL
  DataType: DATA
  Variable: newBDTG_15_3l3j,13,-1.,1.
  VariableTitle: BDT response
```

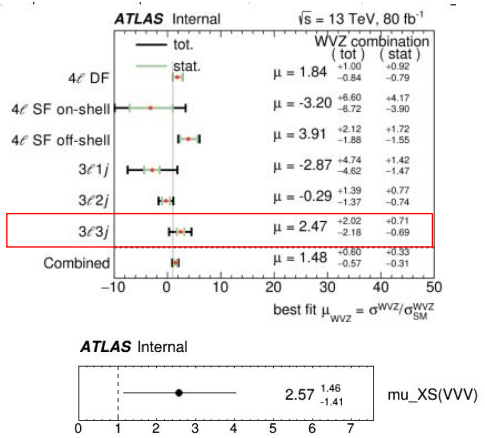
in the config

newBDTG_28_3l3j
newBDTG_32_3l3j
newBDTG_32_3l3j_1

in the ntuples

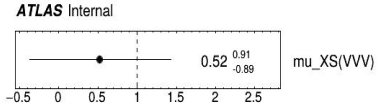


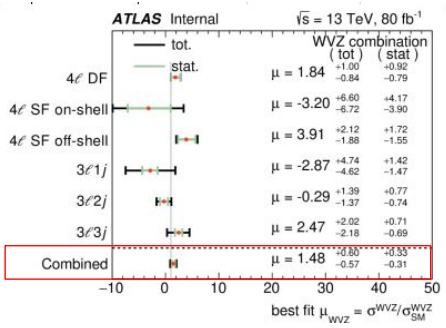


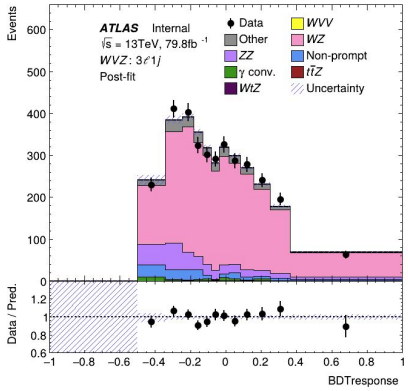


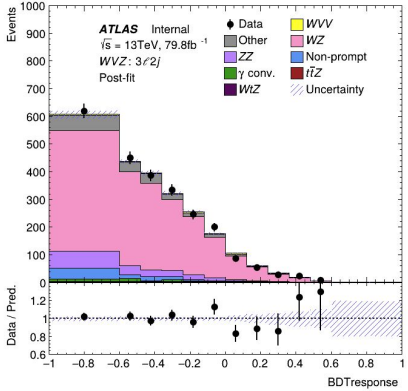
Combined regions (real data)

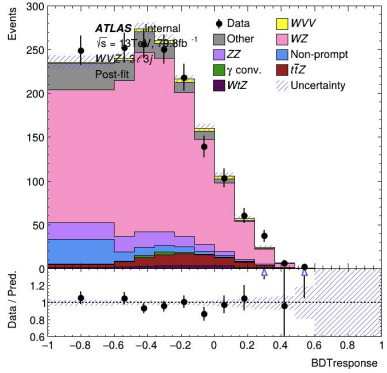
- Combined results from **three regions**: **3ℓ + 1j**, **3ℓ + 2j**, and **3ℓ + ≥3j**.
- Used **real data** to verify consistency of the combined fit.
- The **post-fit BDT distributions** (bottom plots) show good agreement between data and prediction across all regions.
- The obtained **combined signal strength** is **μ = 0.52**, while the **reference value** reported in the publication is **μ = 1.48**:
  - The **difference mainly comes from the 3ℓ + 2j region**:
    - -0.67 vs -0.29

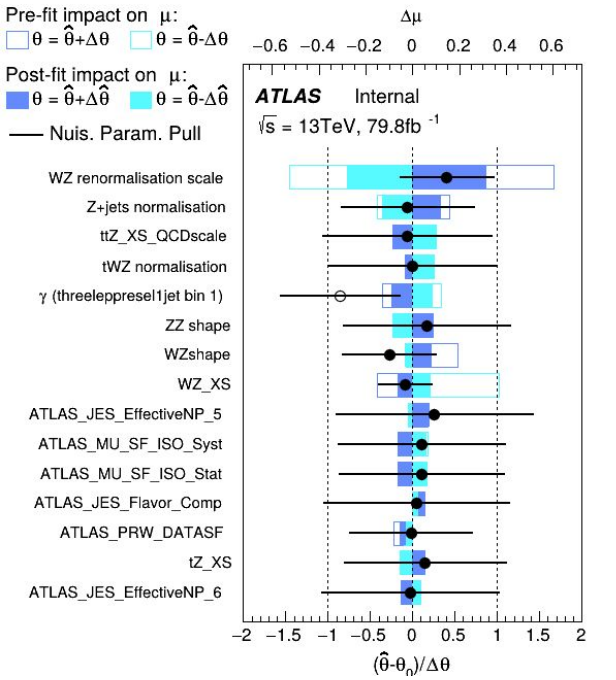






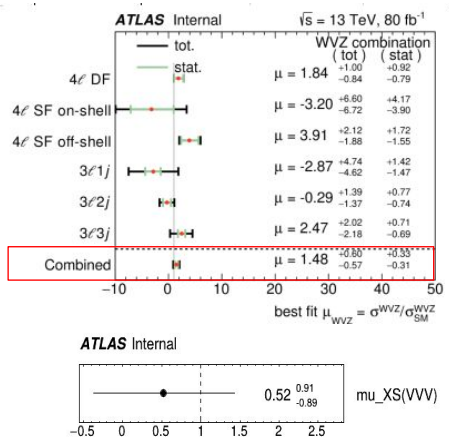


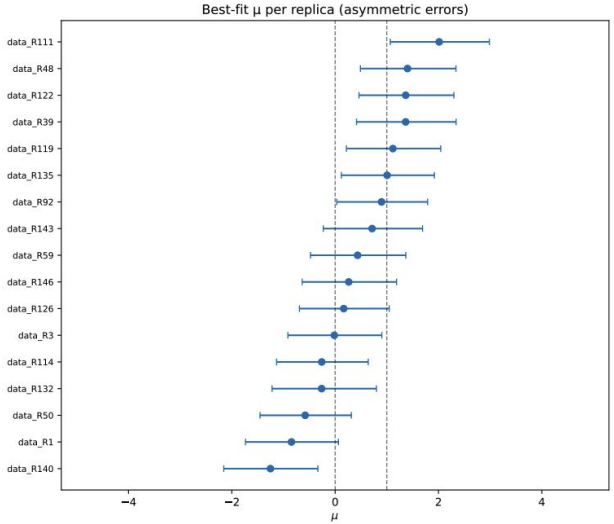


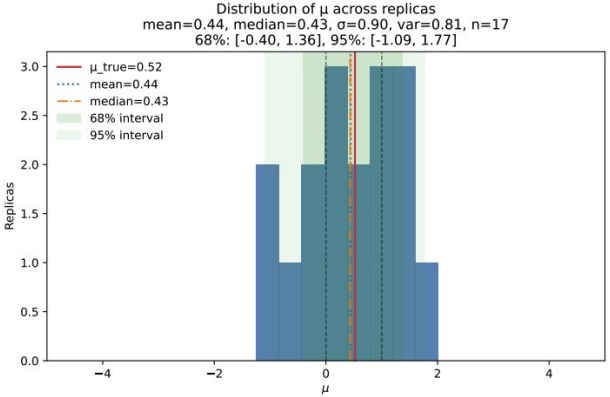


Combined regions (replicas)

- Many replicas **fail during generation**, when TRExFitter tries to run all options simultaneously (**`nwfsdpr`**).
- Currently, I'm running **one option at a time** (**n**), and then plan to process the remaining ones.
- So far, **17 replicas** have been successfully generated.
- The **distribution of μ** is close to the original result (**μ = 0.52**),
  but **more statistics are needed** to confirm the stability of the result.

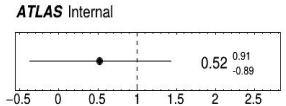


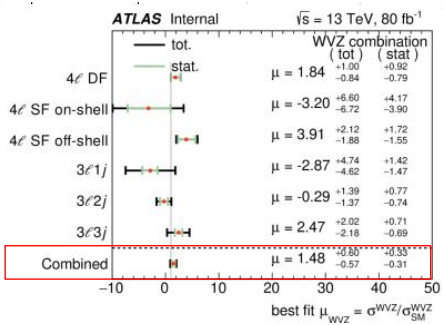




Replicas: 17
μ mean: **0.4429**, median: 0.4341, std: 0.9023
Avg errUp: 0.9335, Avg |errDown|: 0.9045
-- vs μ_true=0.52 --
Bias : -0.07706
RMSE : 0.8787
Std/Var : 0.9023 / 0.8141
Replicas coverage 68% : 76.5%
Replicas coverage 95% : 100.0%
Global 68% CI [-0.402, 1.36] => hit? True
Global 95% CI [-1.09, 1.77] => hit? True

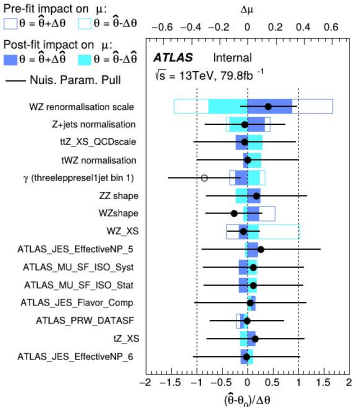
Combined regions (replicas)



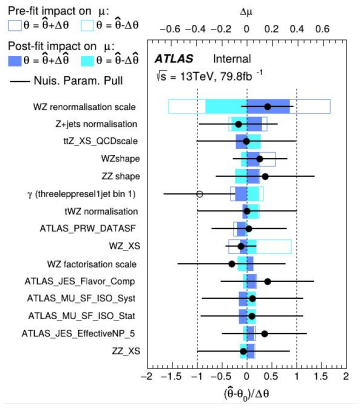


Each replica includes random statistical variations in data, which slightly change the fitted nuisance parameters and their impact on μ.

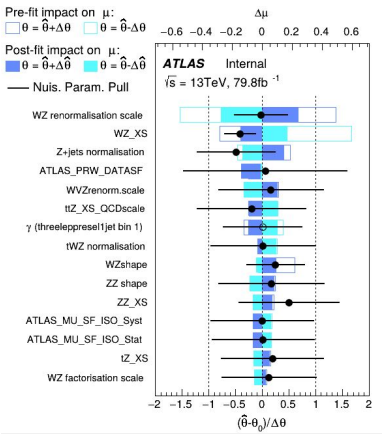
Real Data



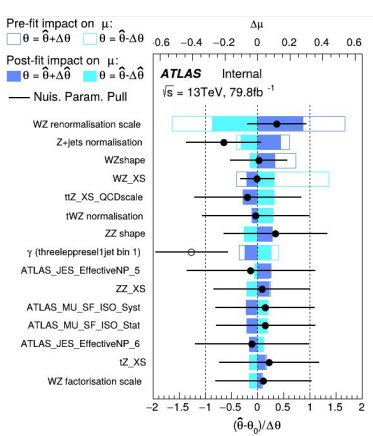
R3



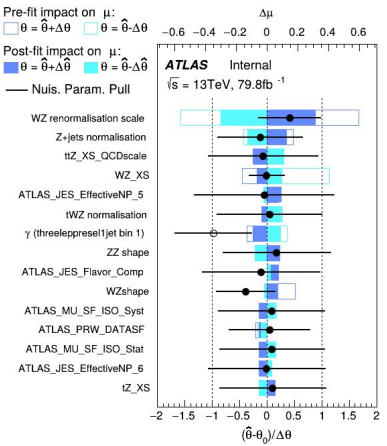
R39



R48



R50



HEP data

Measurement of the total and differential cross-sections of ttW production: a good starting point

- Ready bootstrap replicas already included in the workspace ✅
- Working fitting scripts available ✅
- Regularization disabled, fitting launched (XRooFit) ✅
  - evaluate the statistical variation of the data
- Running 1000 replicas: in progress ⏩

Fit results (without regularization):

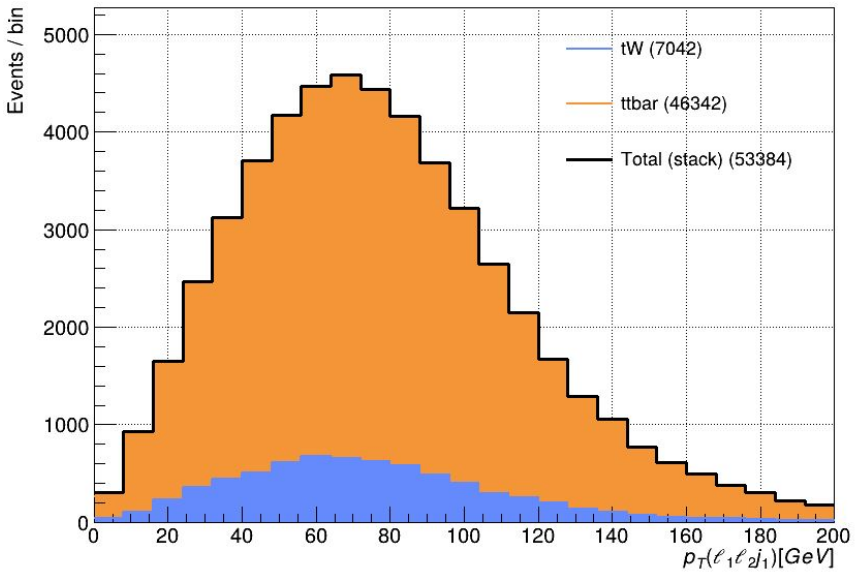
- NLL: 454.056
- Norm_ttW = 1.1393 ± 0.0932

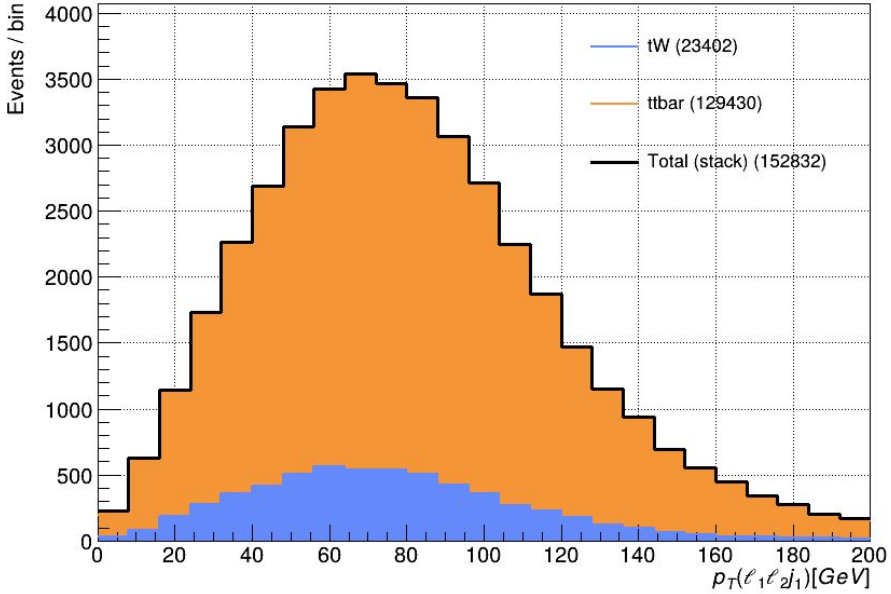
Measurement of t-channel production of single top quarks and antiquarks

- JSON file with data: understanding of its structure is required (or conversion JSON → ROOT workspace)
- Several fitting options available: PyHF or TRExFitter
- repeating the ttW structure: conversion JSON → YAML workspace

BDT ntuples

- **Additionally:** the ntuples for the BDT are ready, and the statistics have been increased by a factor of three.





Thank you!