ML methods for stop pair production search

19st June 2024

Daniele Dal Santo

UNIVERSITÄT BERN

AEC
ALBERT EINSTEIN CENTER
FOR FUNDAMENTAL PHYSICS

Motivation for a Run-3 stop search

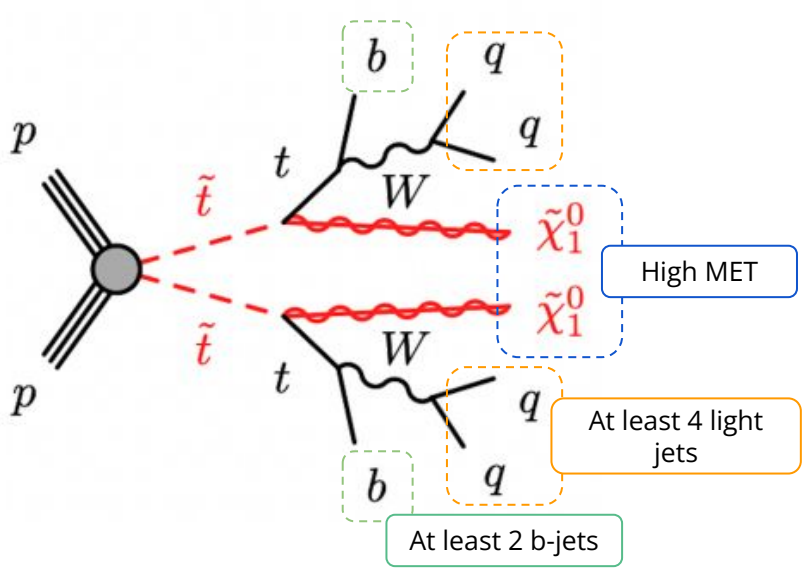
Search for stop pair production in all-hadronic tt+MET (0L) final state

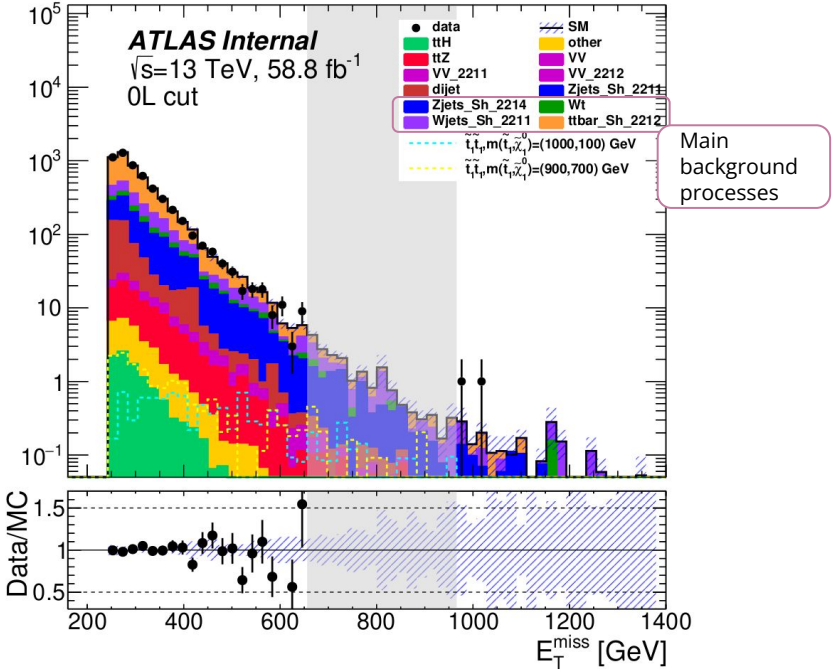
Aim: extend ATLAS Run 2 analysis thanks to

- increased statistics (> 90fb-1)
- increased center of mass energy during Run-3
- **improved signal background discrimination with ML**

The signal process

Search for stop pair production in all-hadronic tt+MET (0L) final state





stop OL: Run 2 vs early Run 3

In Run 2 analysis:

- only use final state information
- define regions employing physically motivated variables

In early Run 3 analysis:

- reconstruct the resolved top decays
- use classifier score to define regions
- (end-to-end classification is also being considered)

stop 0L: Run 2 vs early Run 3

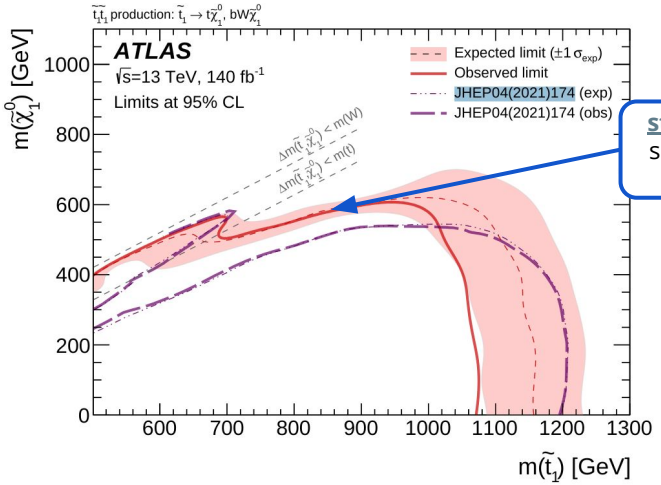
In Run 2 analysis:

- only use final state information
- define regions employing physically motivated variables

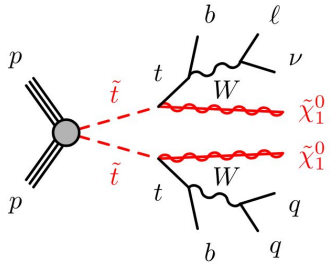
In early Run 3 analysis:

- reconstruct the resolved top decays
- use classifier score to define regions
- (end-to-end classification is also being considered)

Do we expect improvements?



stop 1L Run 2 (2nd wave): significant improvement in the compressed region



Truth information

Consider signal samples and the truth *W* and *top* 4-momenta:

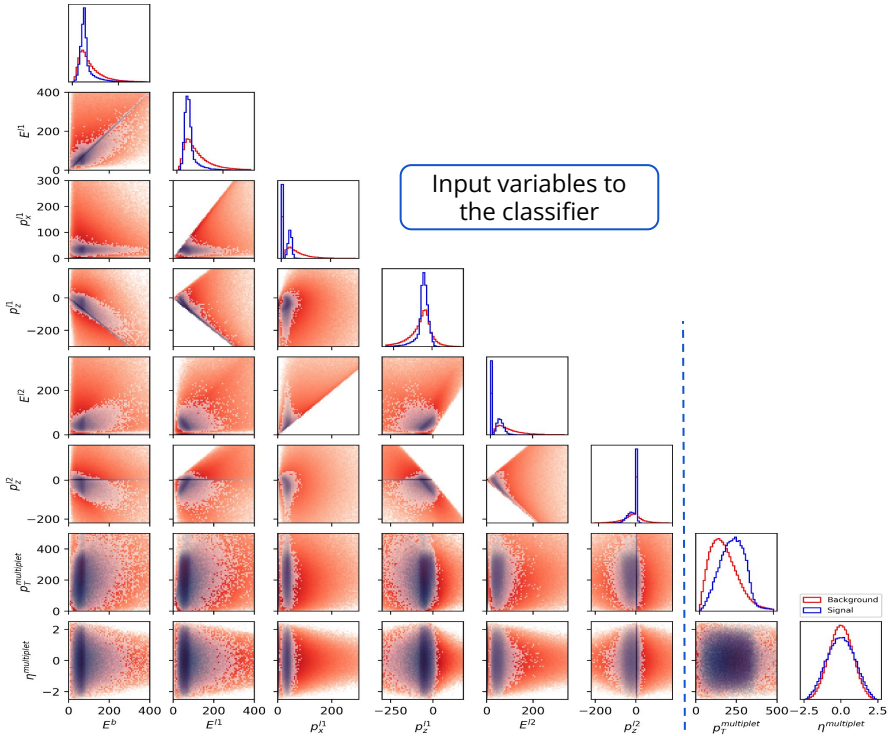
- apply a truth matching procedure
  - find the b-jet + true *W* that are in the decay cone of a true *top*
  - find the light jet(s) that are in the decay cone of a true *W* and ensure that the (di-)jet mass is close to W mass

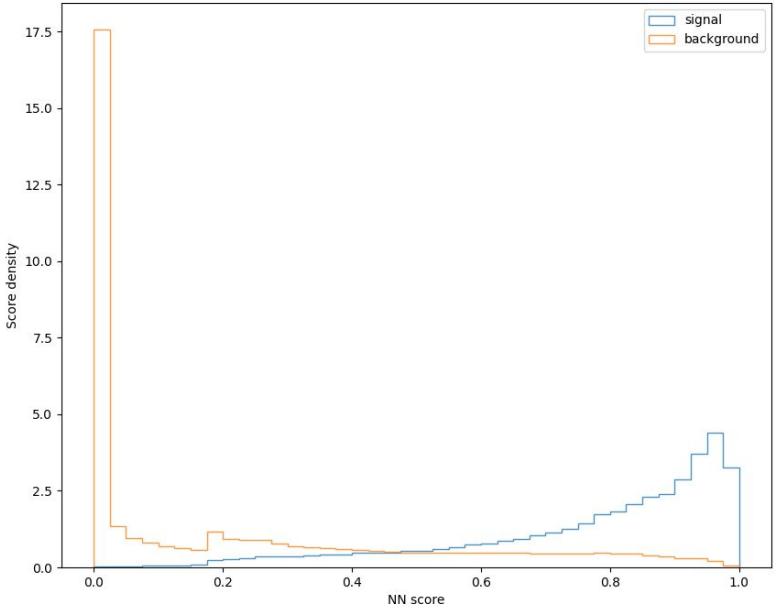
$$\Delta R_y < 2.1 \cdot m(t)/p_{\mathrm{T}}(t) \qquad \frac{|m_W - m(jj)|}{m_W} < 0.3$$

- this identifies up to 2 truth matched triplets (doublets)
- all the other multiplets are labeled as combinatorial background

Classifier for resolved top reconstruction

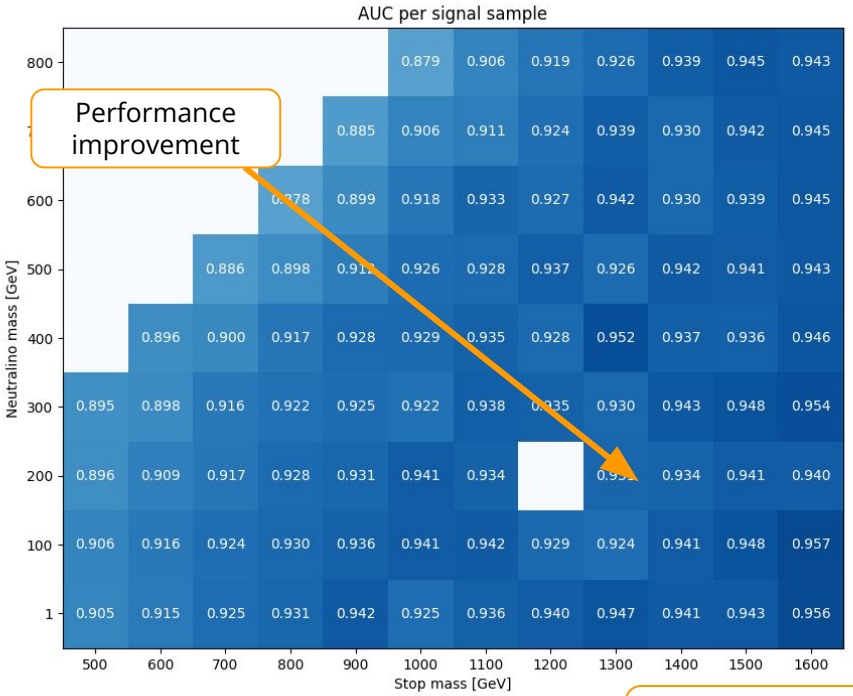
Idea: *train a classifier to distinguish between truth matched multiplets and combinatorial background*

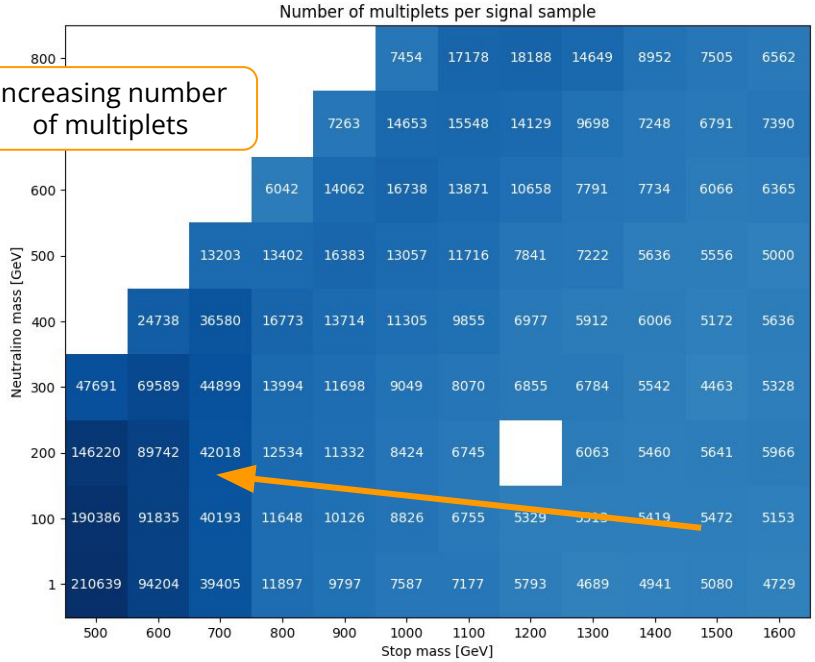




Classifier for resolved top reconstruction

Global AUC= 0.913





Performance gradient due to events being harder to reconstruct

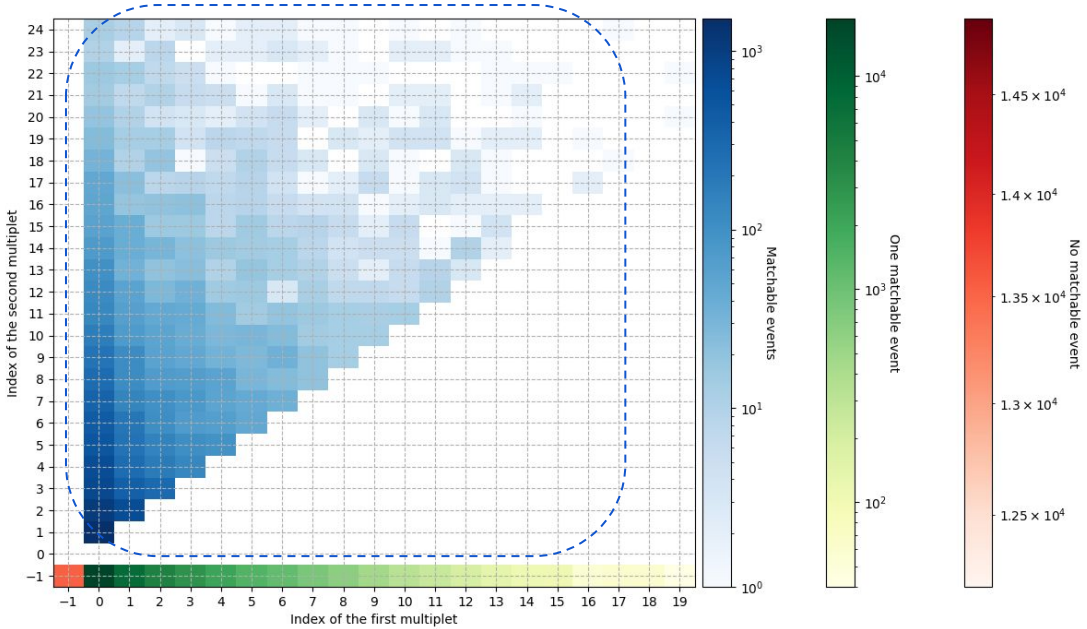
Classifier for resolved top reconstruction

Limit of the approach:

- not trivial to select the 2 multiplets to be matched with the 2 hadronic top decays
- the classifier doesn't have the full picture about the event

- Multiplets in the events are ordered according to NN score

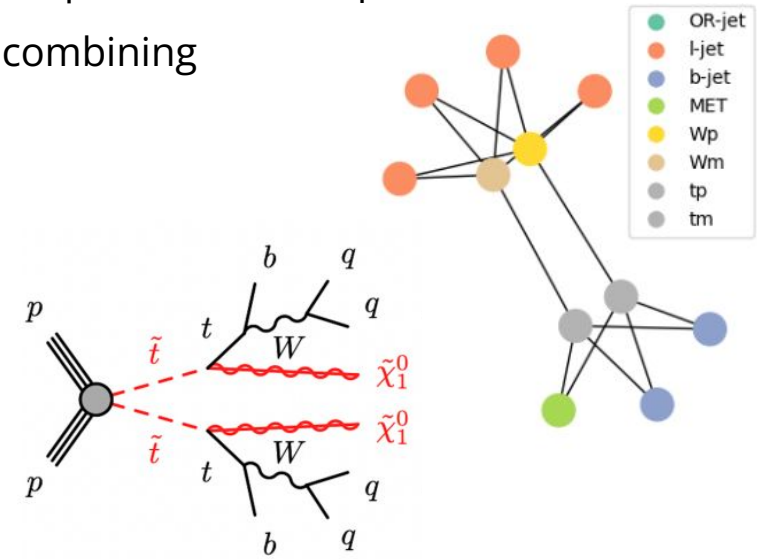
- The histogram shows the indices of the truth matched multiplet



Next steps

Moving to a transformer (*SPANet-like*, investigated by 1L group) or graph (*Topograph-like*, investigated by 0L group) NN approach:

- the full information about the event is exploited
- a score can be calculated to match every final-state particle with its parent
- it's possible to implement a multi-loss approach combining
  - truth-matching information
  - regression of kinematic variables of *W*s and *top*s
  - end-to-end signal vs background classification



ML methods for stop pair production search

19st June 2024

Daniele Dal Santo

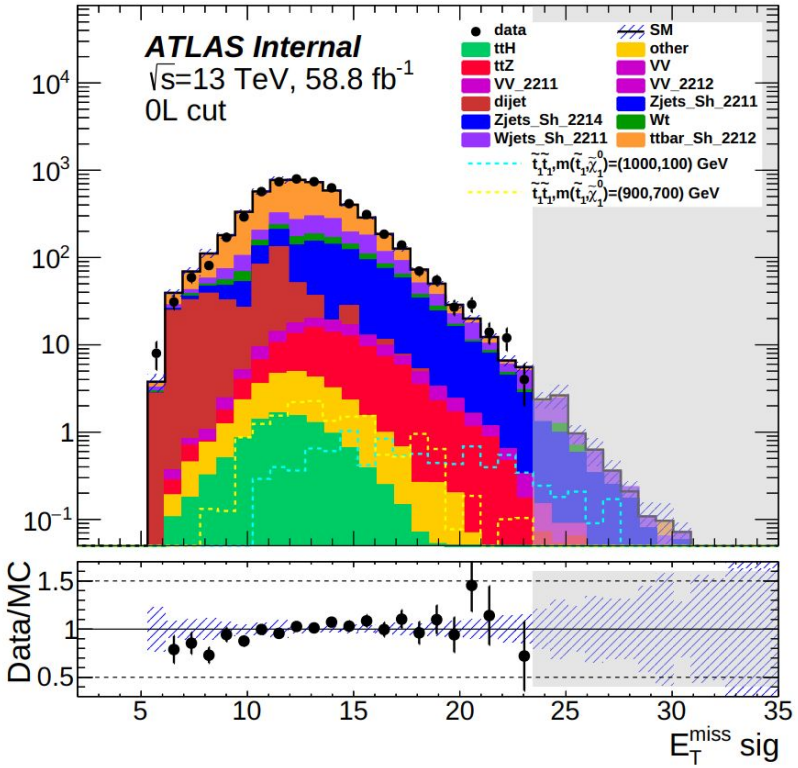
UNIVERSITÄT BERN

AEC
ALBERT EINSTEIN CENTER
FOR FUNDAMENTAL PHYSICS

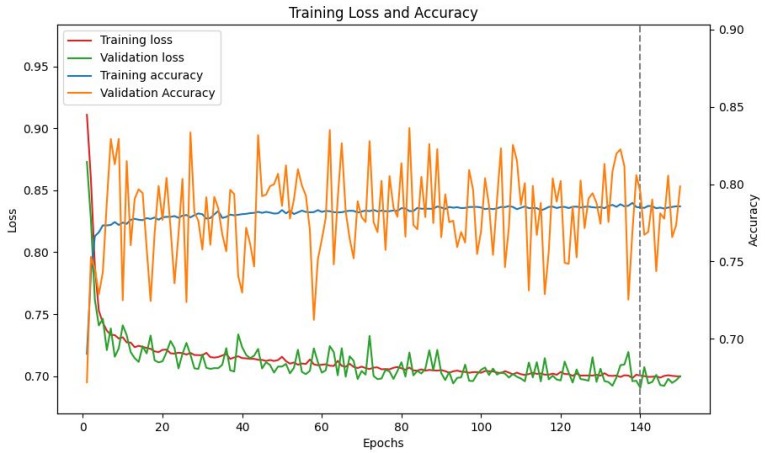
Preselections

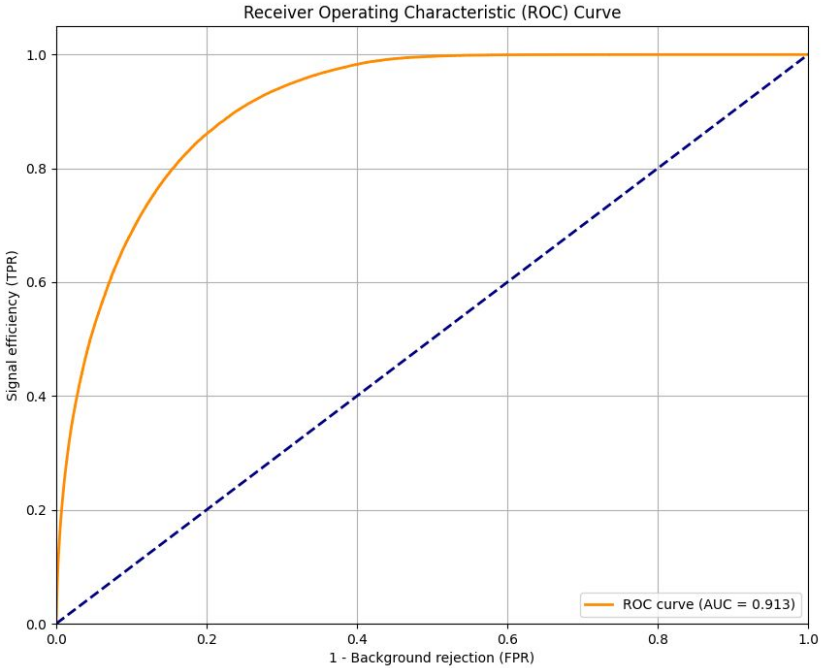
- MC20e vs data18

- $E_{\mathrm{T}}^{\mathrm{miss}} > 250$ GeV
- $n_{\mathrm{lep}} < 2$
- $n_{\mathrm{jets}} > 3$
- $n_{\mathrm{bjets}} > 0$
- $p_{\mathrm{T}}(\mathrm{2nd\ leading\ jet}) > 80$ GeV
- $p_{\mathrm{T}}(\mathrm{4th\ leading\ jet}) > 40$ GeV
- $\mathrm{min}\Delta\Phi(\mathrm{E}_{\mathrm{T}}^{\mathrm{miss}}, \mathrm{leading\ 4\ jets}) > 0.4$
- $E_{\mathrm{T}}^{\mathrm{miss}}\mathrm{significance} > 5$



Classifier for resolved top reconstruction





NN for resolved top reconstruction

Consider the two leading multiplets (in NN score) per event as identifying the top candidates