A COMBINED FIT TO THE HIGGS BRANCHING RATIOS AT ILD

ILD Analysis/Software Meeting

Vincent Boudry, Jean-Claude Brient, Fabricio Jimenez, Jonas Kunath (LLR)

3 mars 2021

OUTLINE

A combined fit to the Higgs Branching Ratios at ILD

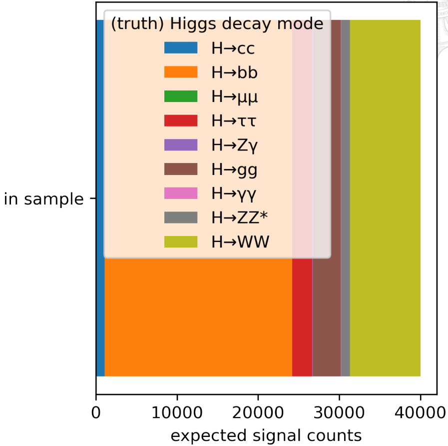
Previously :
Following up on JCB's talk at LCWS2019.

Currently :
Still a work in progress.

HIGGS-BRS ALL-IN-ONE

Trying to adapt a τ branching ratio paper from ALEPH to Higgs at ILD :

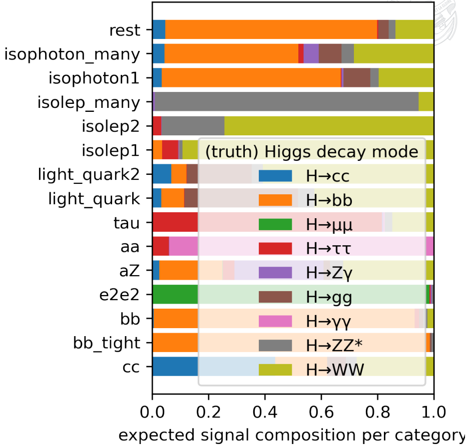
1. Build a (Higgsstrahlung) sample with all Higgs decay modes present.
2. Construct categories to separate the decay modes (& background) as well as possible.
3. Fit the Higgs branching ratios to the observed category counts.



HIGGS-BRS ALL-IN-ONE

Trying to adapt a τ branching ratio paper from ALEPH to Higgs at ILD :

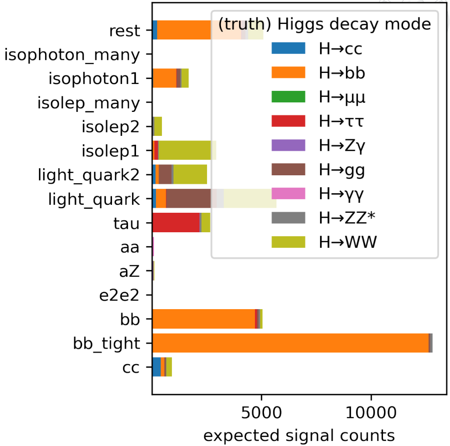
1. Build a (Higgsstrahlung) sample with all Higgs decay modes present.
2. Construct categories to separate the decay modes (& background) as well as possible.
3. Fit the Higgs branching ratios to the observed category counts.



HIGGS-BRS ALL-IN-ONE

Trying to adapt a τ branching ratio paper from ALEPH to Higgs at ILD :

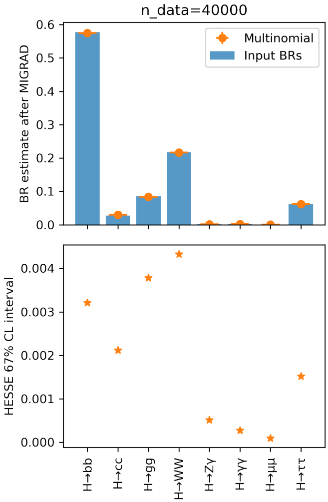
1. Build a (Higgsstrahlung) sample with all Higgs decay modes present.
2. Construct categories to separate the decay modes (& background) as well as possible.
3. Fit the Higgs branching ratios to the observed category counts.



HIGGS-BRS ALL-IN-ONE

Trying to adapt a τ branching ratio paper from ALEPH to Higgs at ILD :

1. Build a (Higgsstrahlung) sample with all Higgs decay modes present.
2. Construct categories to separate the decay modes (& background) as well as possible.
3. Fit the Higgs branching ratios to the observed category counts.



HIGGS-BRS ALL-IN-ONE

Trying to adapt a τ branching ratio paper from ALEPH to Higgs at ILD :

1. Build a (Higgsstrahlung) sample with all Higgs decay modes present.
2. Construct categories to separate the decay modes (& background) as well as possible.
3. Fit the Higgs branching ratios to the observed category counts.

Advantages

- A model independent extraction of all branching ratios (at once).
- Independent of any Higgs production cross section measurement.
- Gaussian errors $\rightarrow$ multinomial errors, as everything is in the same sample.

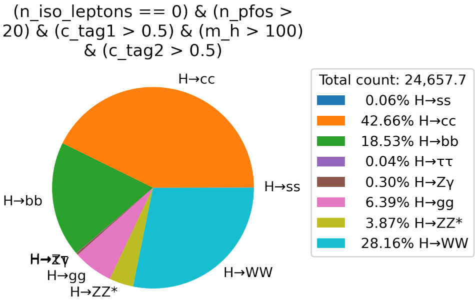
CURRENT SIMPLIFICATIONS

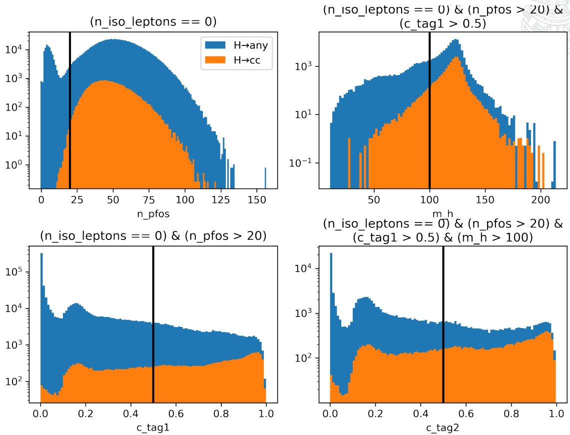
Reconstructed events from the new $\sqrt{s} = 250$ GeV SetA mass production (`ILD_l5_o1_v02`, `v02-02`).

- So far, only using the signal (Higgsstrahlung).
  - $\geq$ 100k simulated events/decay mode.
- $ZH \rightarrow \nu\bar{\nu}H$, so no confusion between Z and H at this stage.
- MC-Veto on overlay particles (and `LCFIPlus` re-vertexing on remaining PFOs).
  - Not necessary, but should redo some category boundaries (see backup).

Assumption for the fit (later) : $\sigma_{ZH}^{\text{unpol.}} \approx 200$ fb, 2000 fb^{-1} H-20 scenario, catch 10 % of the signal $\rightarrow O(40,000)$ Higgs events.

CATEGORY BUILDING EXAMPLE : CC



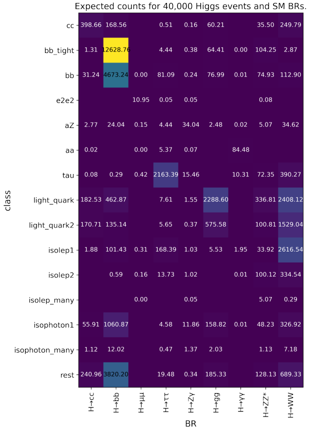


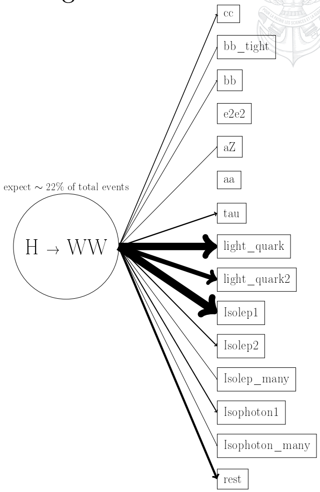
CATEGORIES

- Reuse common tools :
  - `LCFIPlus` for jet construction and flavor tagging.
  - `IsolatedLeptonTaggingProcessor`, `IsolatedPhotonTaggingProcessor`.
- Custom variables motivated by specific channels :
  - $M_{\mu^+\mu^-}$, $M_{\gamma\gamma}$.
  - Recoil of the leading (Isolated) Photon against $(E = M_H, \vec{p} = 0)$ at rest ($\approx 60 - 100$ GeV for $H \to Z\gamma$).
  - ...
- Only use so-far not categorized events in classes further down in the list.

CATEGORY MATRIX

This shows how events from a given BR distribute among categories. E.g. $H \to W^+W^-$.

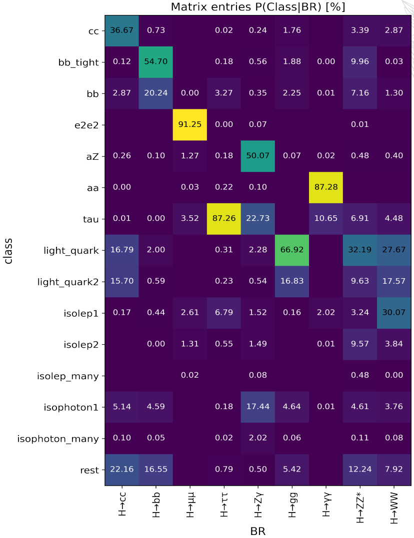




OPTIMIZATION - SETUP

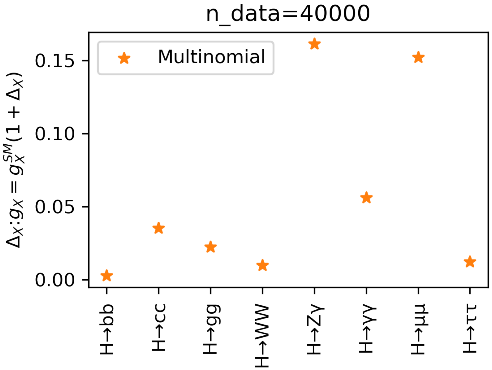
BRs from minimization through iminuit.

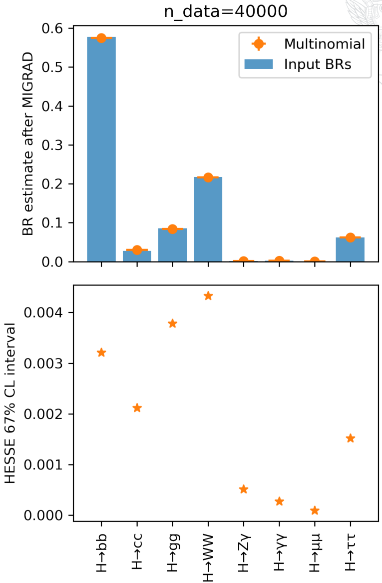
- $\vec{S} = M \cdot \vec{B} = \vec{f}(\vec{B})$, with
  - $\vec{S}$: The signal counts per category ($S = data - bkg$).
  - M : The matrix build from simulated events, as outlined above.
  - $\vec{B}$: The target. Use the BRs in the simulation as starting values.
- The cost function : Multinomial log-likelihood.
  - $-\ln\mathcal{L} = -N_{\text{data}} \sum_i S_i \ln\left(\sum_j M_{ij} B_j\right)$.
  - $B_{H\to ZZ^*} = 1 - \sum_{i\neq H\to ZZ^*} B_i$.
  - Each BR constrained to $[0, 1]$.



OPTIMIZATION - RESULTS

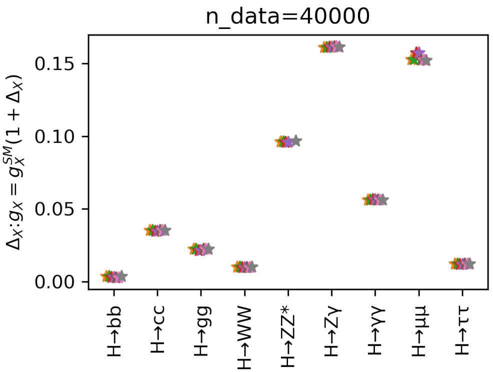
The fitted BR^{min} reproduces BR^{true} within its uncertainties.

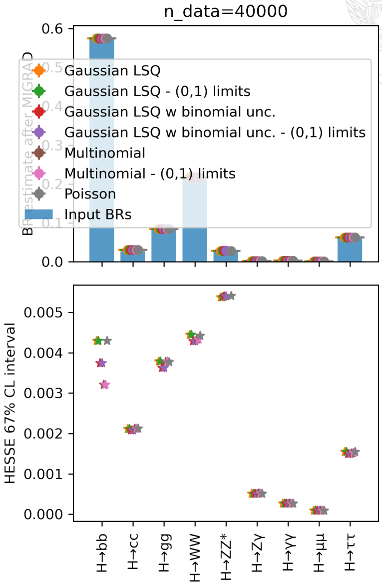




OPTIMIZATION - RESULTS

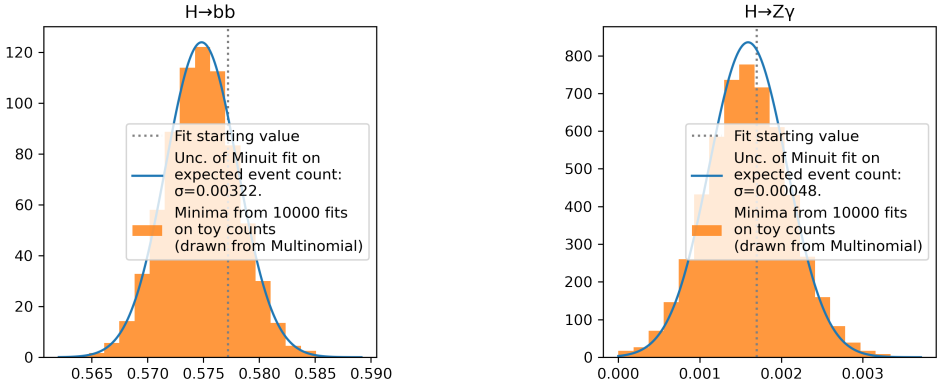
Effect of the likelihood definition on the uncertainties.





OPTIMIZATION - VALIDITY CHECK

Toy study : Draw from multinomial (N_{data} fixed).
Shown : 2 of the toy fit distributions for multinomial $\ln\mathcal{L}$ with $[0,1]$ boundaries.

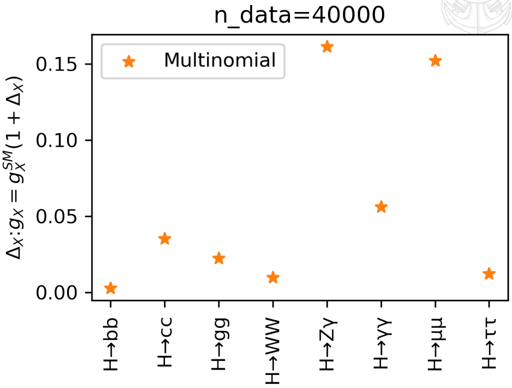


CONCLUSION

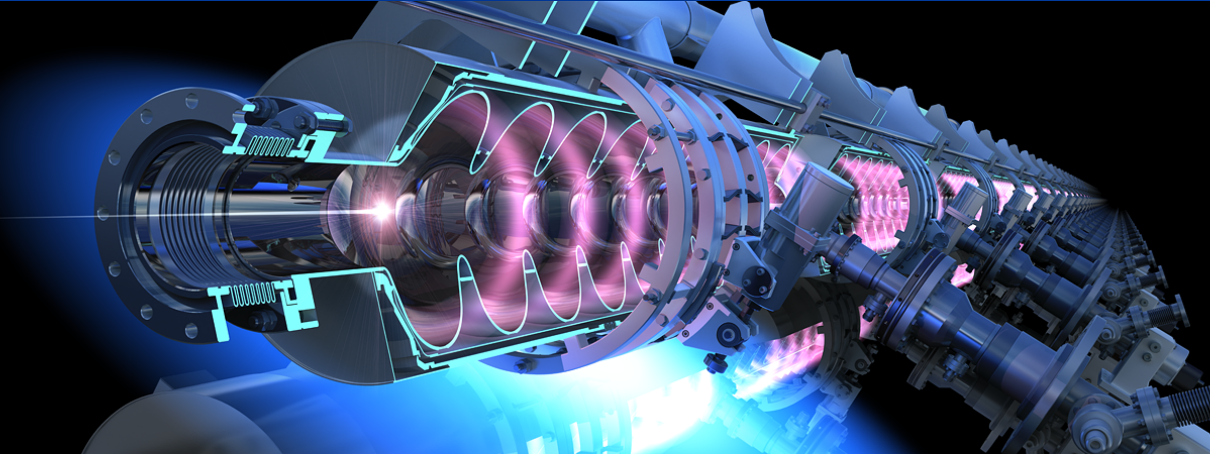
- An extraction of all branching ratios without direct dependence on $\sigma_{ZH}(, \sigma_{\text{WW-fusion}})$.
- Can deal with large changes in the Higgs BRs (see backup).

More work needed :

- Signal selection/background rejection omitted so far.
- Overlap with dedicated analyses on $ZH, Z \rightarrow (e^+e^-, \mu^+\mu^-, \nu\bar{\nu})$.

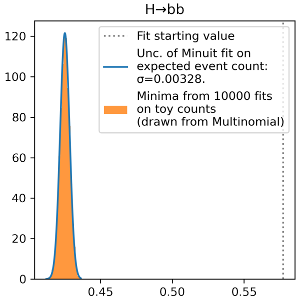


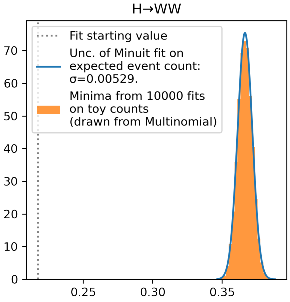
5
BACK-UP

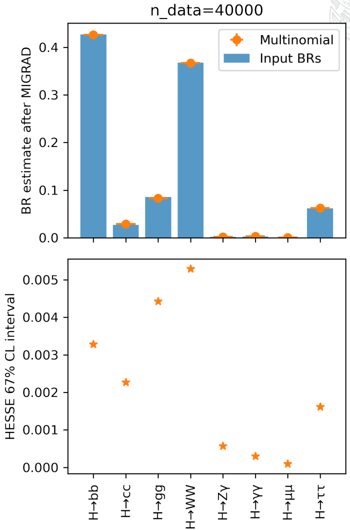


FIT IN A NON-SM SCENARIO

Assume 57.7% $\rightarrow$ 42.7% $BR(H \rightarrow b\bar{b})$, 21.8% $\rightarrow$ 36.8% $BR(H \rightarrow W^+W^-)$.



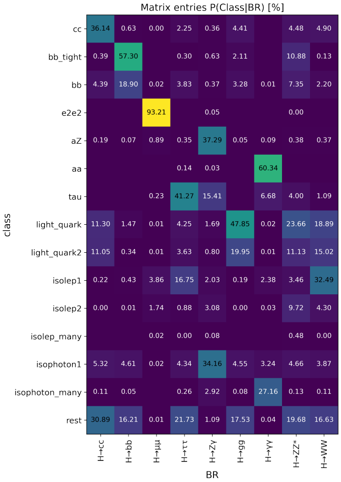




EXPECTED COUNTS PER (CATEGORY, BR) PAIR

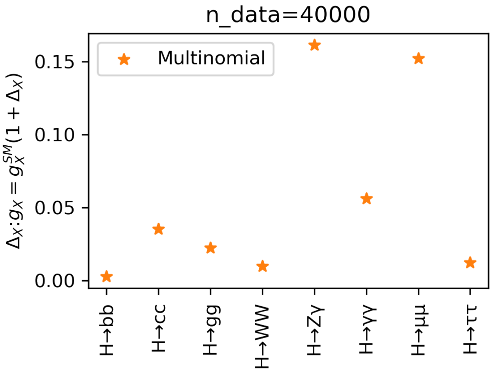
Removed ILC overlay particles.

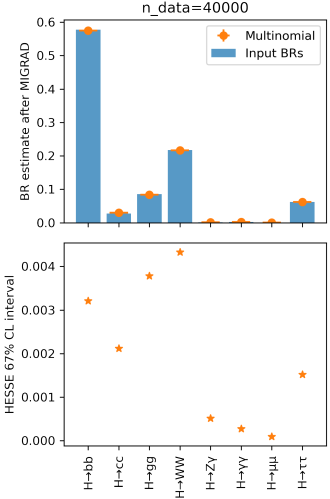
With overlay, categories not adapted yet.



EFFECT OF OVERLAY PARTICLES ON THE FIT

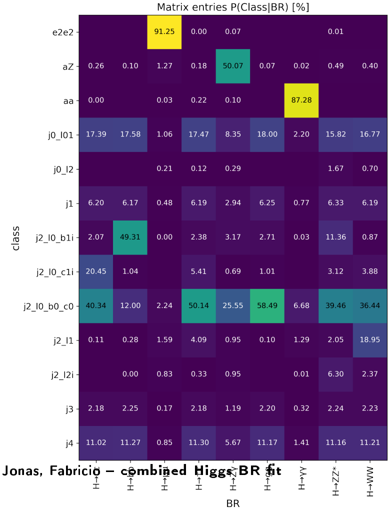
All fits on Higgsstrahlung events without additional background events.
Category definitions are **not** adapted to the scenario with overlay events.

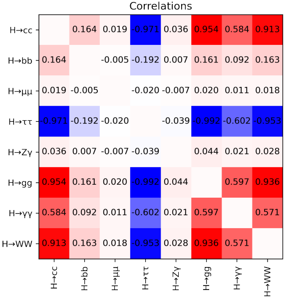


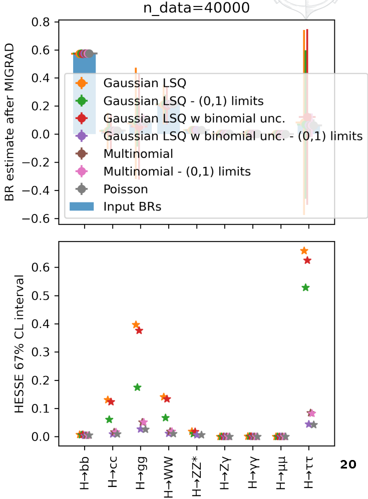


A FIT WITH CATEGORIES THAT ARE TOO CORRELATED

With a highly correlated category definition/fit, the uncertainties of the `MINUIT` fit are huge, and details in the likelihood function definition matter.







CATEGORY DEFINITIONS


```
no_iso = '(n_iso_leptons == 0) & (n_iso_photons == 0) & '
categories = dict(
    cc=no_iso+'(n_pfos > 20) & (b_tag2 < 0.7) & (m_h > 100) & (c_tag1 > 0.5) & (c_tag2 > 0.5)',
    bb_tight=no_iso+'(b_tag1 > .8) & (b_tag2 > .8)',
    bb=no_iso+'(b_tag1 > .8)',
    e2e2='(e2e2_mass > 100) & (e2e2_mass < 130)',
    aZ='(aZ_a_energy > 20) & (aZ_a_energy < 50) & (aZ_other_mass > 75) & (aZ_other_mass < 100) &
    ↪  (abs(aZ_a_cos_theta) < 0.9)',
    aa='(n_iso_photons > 0) & (e_h > 125) & (n_pfos < 15) & (n_iso_leptons == 0) & (abs(aZ_a_cos_theta) < 0.9) &
    ↪  (aZ_a_energy > 35)',
    tau='(n_pfos < 15) & (n_iso_leptons == 0)',
    # tau='(y_45 < 1e-5) & (n_iso_leptons == 0)',
    light_quark=no_iso+'(b_tag1 + c_tag1 < .5)',
    light_quark2=no_iso+'(b_tag2 + c_tag2 < .5)',

    # Isolepton collector.
    isolep1='(n_iso_leptons == 1)',
    isolep2='(n_iso_leptons == 2)',
    isolep_many='(n_iso_leptons > 2)',
    isophoton1='(n_iso_photons == 1)',
    isophoton_many='(n_iso_photons > 1)',

    rest='(hDecay > 0)',  # Cannot have empty expression -> Always true.
)
```

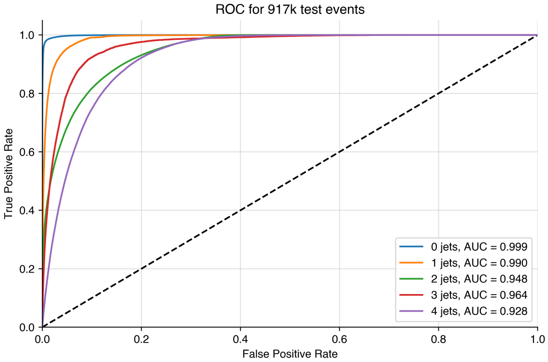
BUILDING NUMBER OF JETS LABEL

Not currently in use, but we investigate if it could help.

Trained a BDT for labeling events by the number of jets using the y and m variables.

- We use 20 input variables :
  - The masses of individual and combined (up to 4) jets, m.
  - The y_{ij} values from the jet clustering algorithm.
- The target used for training is the true number of hadronic jets.
- Trained the BDT using 100k events from the signal sample.
- Performed a randomized search on the following hyper parameters (**best**) :
  - `eta` : [0.05, 0.10, **0.20**, 0.30, 0.35] ,
  - `max_depth` : [2, 3, 4, **5**, 6],
  - `gamma` : [3, 1, **0.5**, 1e-01, 1e-02],
  - `colsample_bytree` : [0.5, **0.6**, 0.7].

BUILDING NUMBER OF JETS LABEL



- Multiclass ROC using one-versus-rest strategy.
- Studies in progress to understand how well this classifier performs, particularly when tested in subsets of data (i.e. categories).
- Having a # jets predicted can help us build optimal categories for Higgs BR measurements.