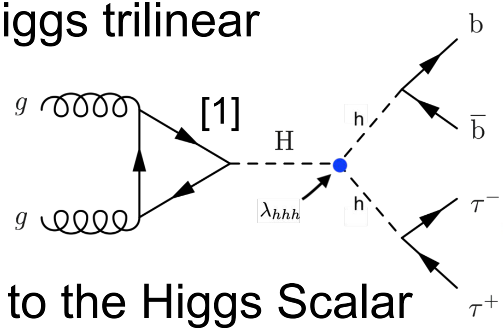
CMS Data Analysis School: $HH \rightarrow bb\tau\tau$ Exercise

Di-Higgs Production: The Physics Case

Non-Resonant Production

Allows measuring Higgs trilinear coupling λ_{hhh}



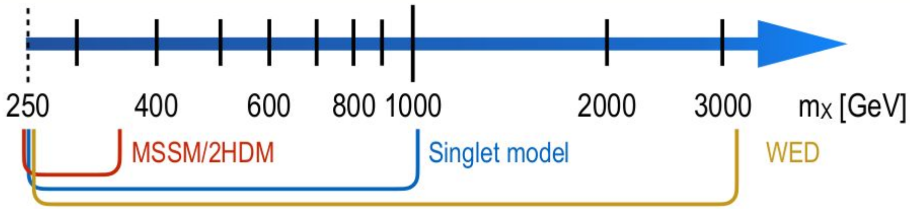
[1]

λ_{hhh} will give access to the Higgs Scalar potential

Data available in Run II are not sufficient to constrain SM hh production, but can constrain BSM models that predict an enhanced hh production

Resonant Production

Model-independent search of narrow width resonance not predicted by the SM

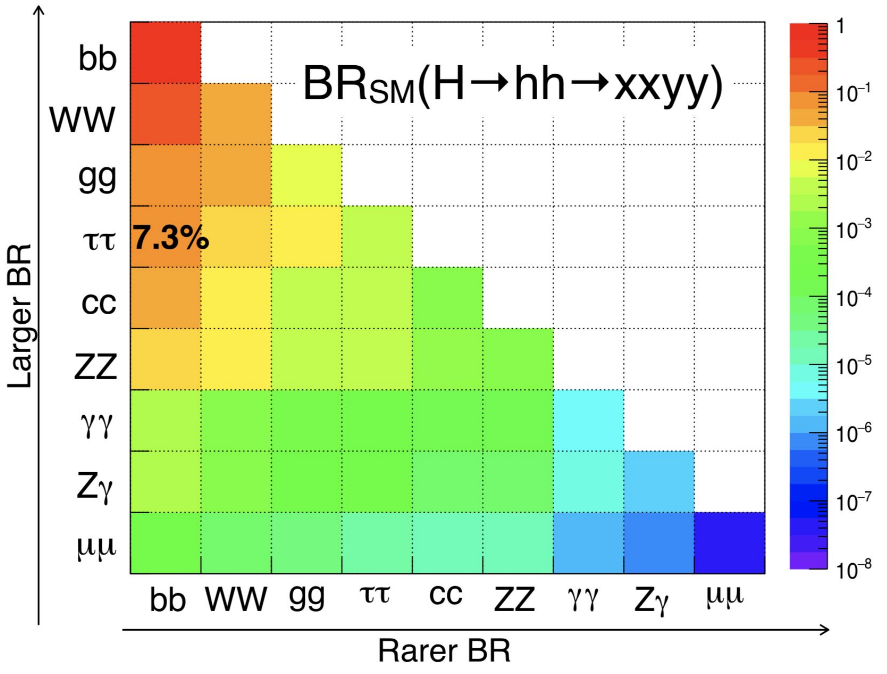


[2]

[1] https://wiki.physik.uzh.ch/cms/latex:feynman
[2]://indico.cern.ch/event/758324/contributions/3294150/attachments/1787235/2910327/CMSDAS_2019_HH_bbtautau.pdf

Why bbττ ?

1. Good trade-off between BR and signal purity
2. bb is the decay channel with highest BR and b-jets tagging; it is very efficient thanks to the precise reconstruction of primary and secondary vertices
3. Tau lepton reconstruction is quite challenging, but efficient algorithms have been developed across the years within the CMS collaboration



Analysis Workflow

Three different final states for H-> ττ are considered: $\tau_e\tau_h$, $\tau_u\tau_h$, $\tau_h\tau_h$

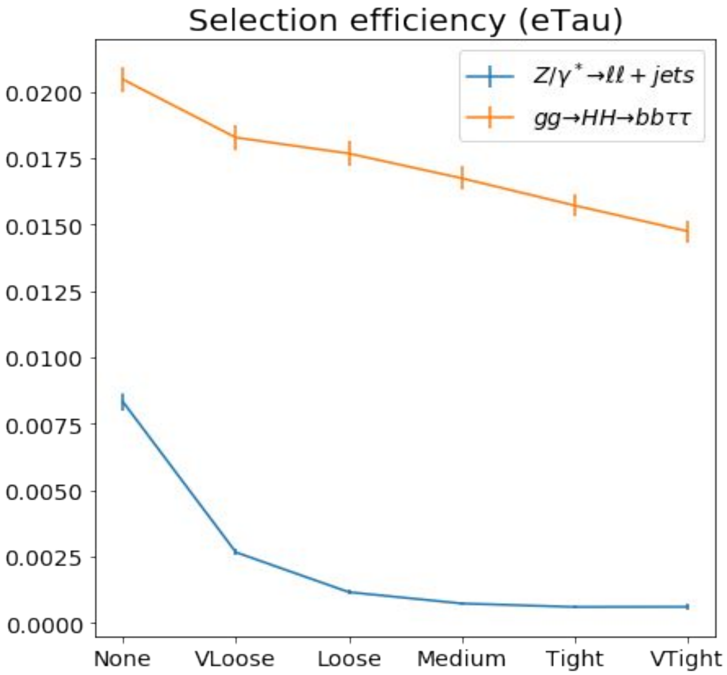
1. Determine the channel looking for isolated electrons, muons, or taus in the event (rejection of electrons faking taus)
2. SM H -> ττ analysis selections adding some specific cuts for this search are used
3. CSVv2 algorithm provided by the b-tag POG is used to select bb candidates
4. Invariant mass for H -> bb and H-> ττ candidates is required to be around the Higgs boson mass. Irreducible ttbar background is rejected applying an elliptical cut on these masses (cut optimization using standard techniques)
5. Limit extraction is performed in different ways:
   a. HH mass after a kinematic fit (resonant analysis)
   b. stranverse mass (non resonant analysis) (optimization using ML techniques)
6. study of the background composition and its impact on the signal sensitivity

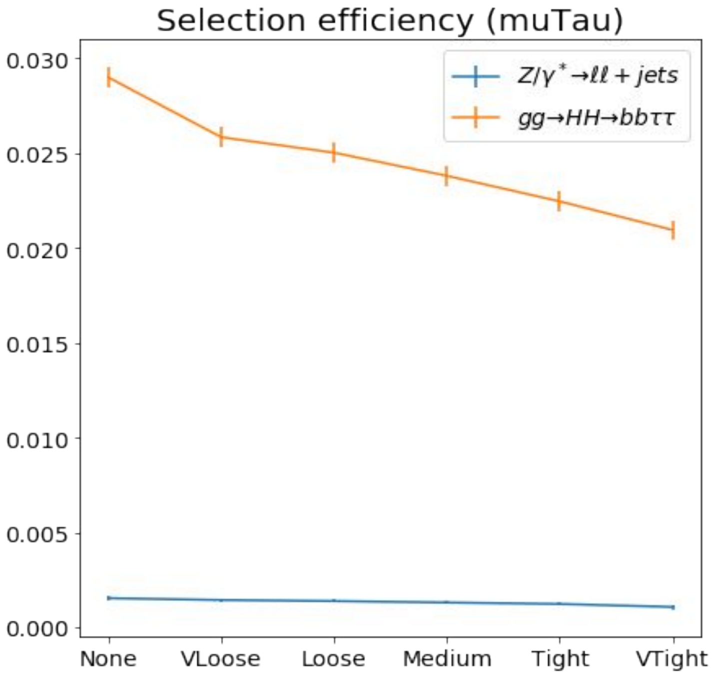
* exercises of this long exercise

Baseline selection studies

Selection efficiency

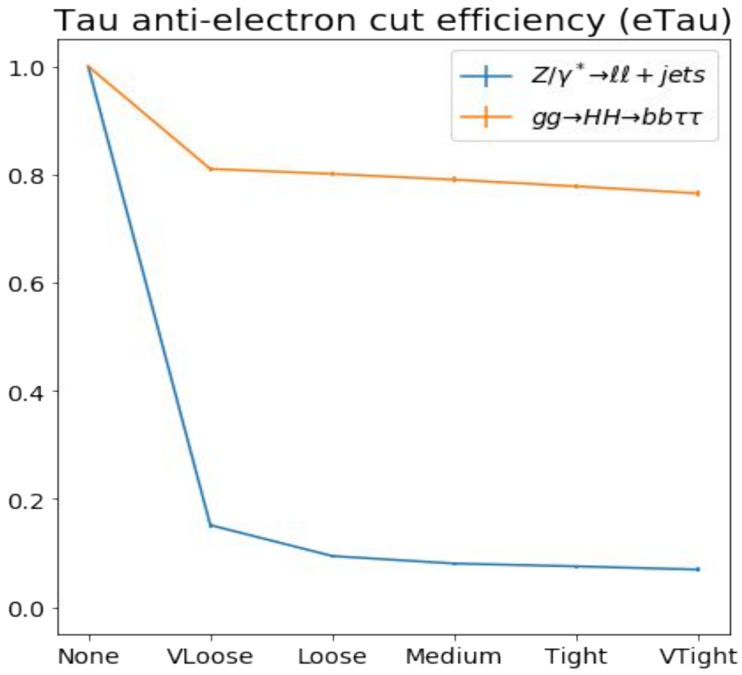
- Samples:
  - Signal: /gpfs/ddn/cms/user/cmsdas/2019/hh_bbtautau/miniAOD/GluGluToHHTo2B2Tau_node_SM/
  - DY: /gpfs/ddn/cms/user/cmsdas/2019/hh_bbtautau/miniAOD/DYJetsToLL_M-50

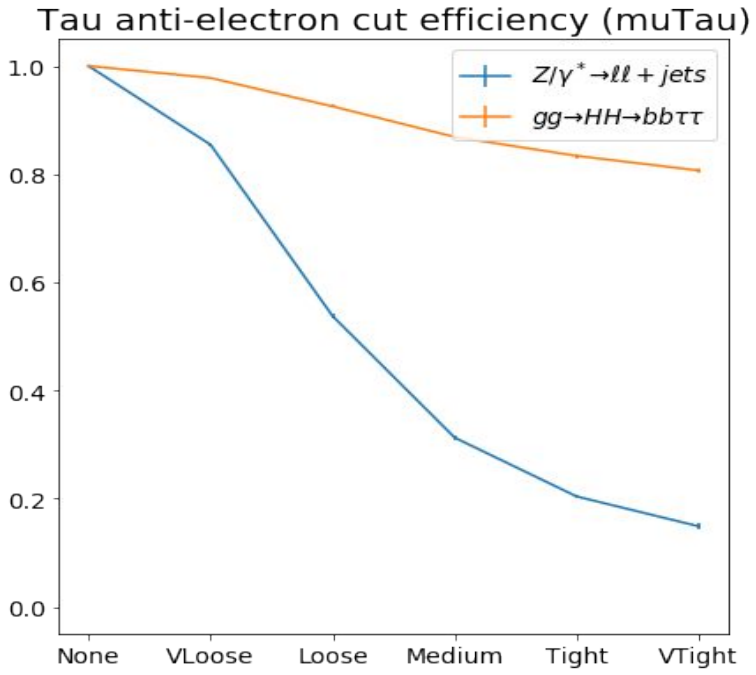




Efficiency of anti-electron cut

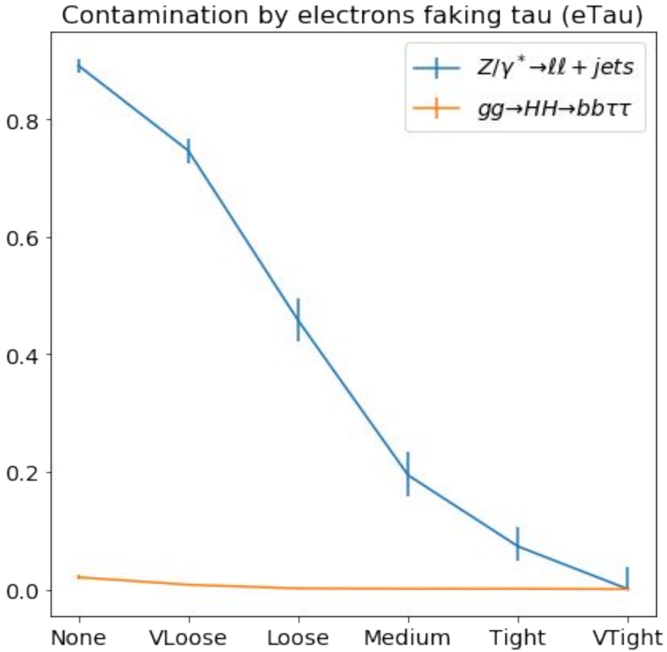
➢ Anti-electron cut applied in order to reject electron faking hadronic tau decay

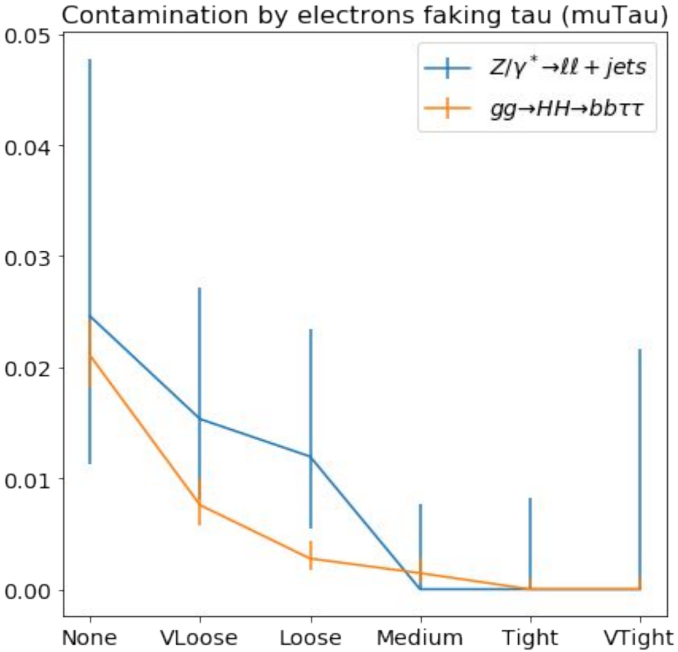




Contamination by electrons faking τ

- ➢ Contamination by electrons in final state
  - ○ reco tau matched to promt electron
  - ○ reco tau matched to electron from tau decay

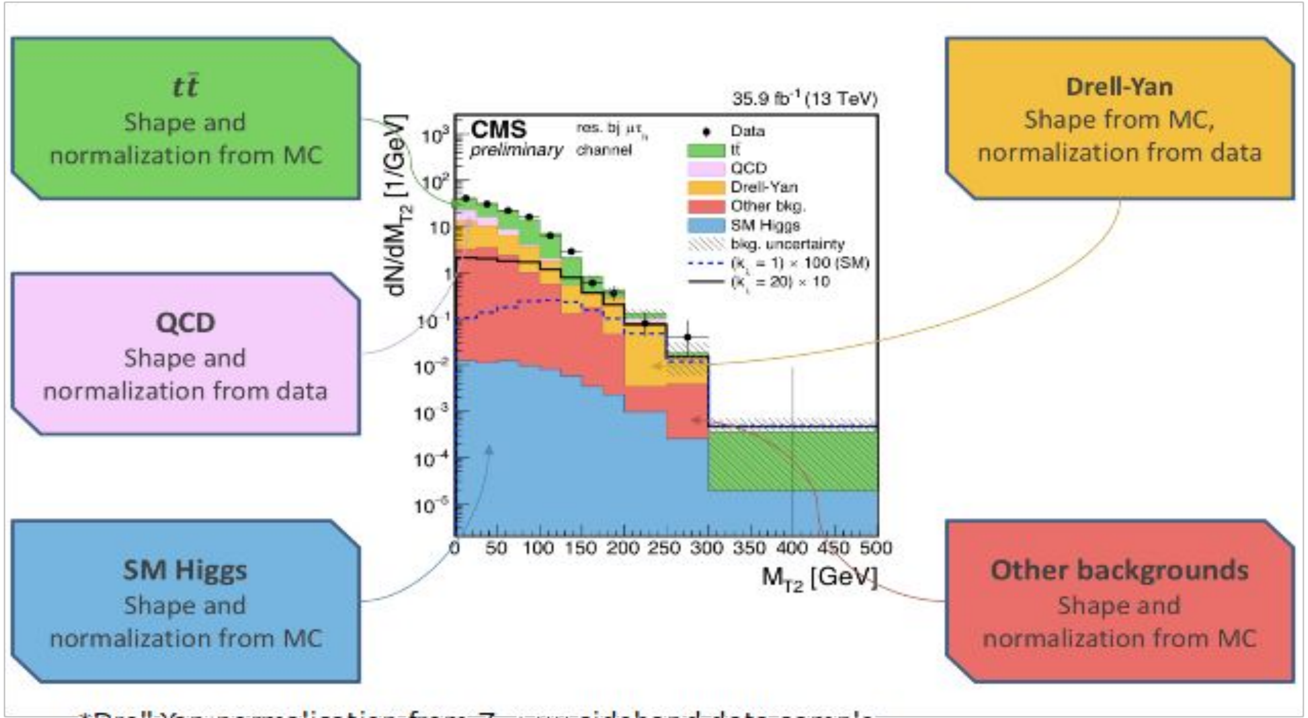




Significance Estimation

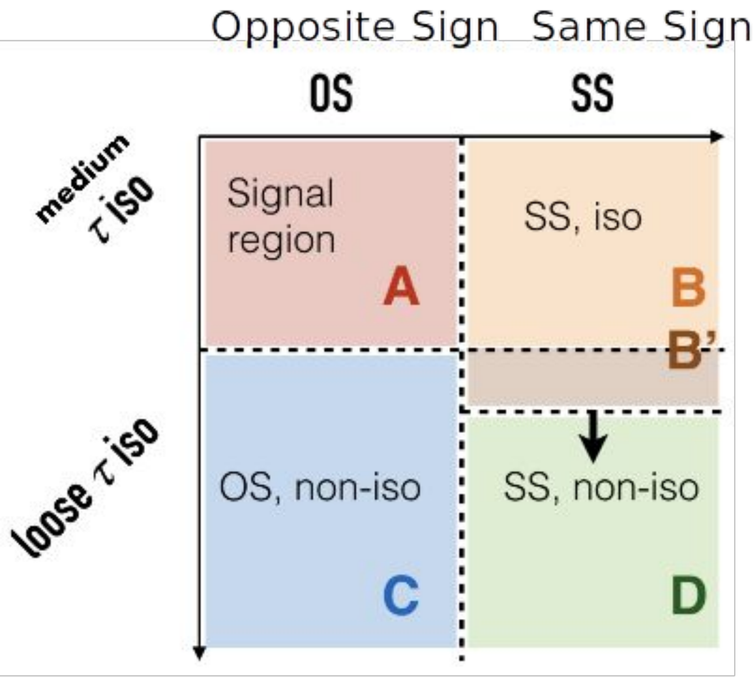
QCD Estimation

For each sideband region: QCD = Data - tt - Drell-Yan - SM Higgs - Other backgrounds



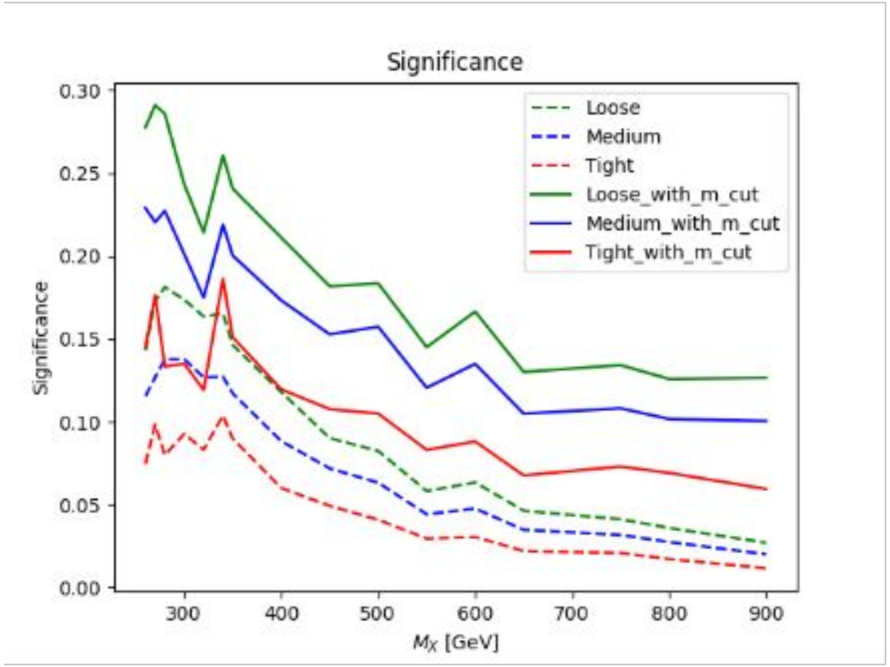
*Drell-Yan:normalisation from Z → μμ sideband data sample

QCD Estimation



- “ABCD” Method
- Signal region QCD: A = B * C / D
- Assumption: QCD Contribution from same/opp sign tau pairs are same
- Not completely true: opp/same sign extrapolation factor is calculated using data

Resonance production



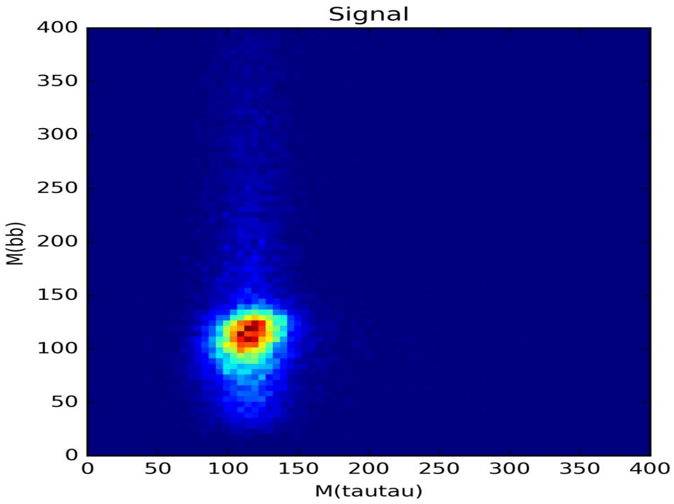
Non - Resonance production

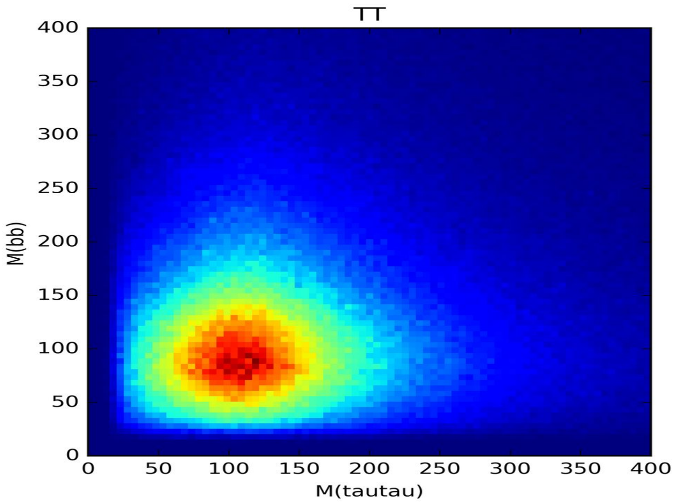
	No MT2 cuts	1σ MT2 cut around resonance peak
Loose	0.0050	0.0024
Medium	0.0039	0.0017
Tight	0.0025	0.0011

- With 1σ mass cuts around the resonance peak, significance improves.

- With 1σ mass cuts around the resonance peak, significance decreases.

H-H Candidate Selection Optimization





Mass window is optimized Accordingly to the resolution and mean value of m_tautau and m_bb distributions:

$$\left(\frac{m_{\tau\tau}\,[GeV] - 116}{35}\right)^2 + \left(\frac{m_{bb}\,[GeV] - 111}{45}\right)^2 < 1$$

Optimization workflow

- The distributions show:
    - irreducible backgrounds from ttbar process.
    - A long event tail, which is bad for MC modeling.
- Mass cut optimization on invariant mass of HH candidate was carried out on three channel (etau,mtau,tautau). The minimized parameters will give us optimal significance.
- Multitude of optimization was performed on ellipse's parameters (elliptical center and radius) by custom minimizer base on MINUIT, using most probable value and interquantile distance for center (118.5, 117.5) and radius (22.5, 14.2).

Optimization summary

Channel	Fitted Centers	Fitted Radius
eTau	x0 = 115 ± 2.7 y0 = 116 ± 16	a = 27 ± 0.4 b = 26 ± 0.5
muTau	x0 = 112 ± 97 y0 = 116 ± 97	a = 31 ± 60 b = 28 ± 28
tauTau	x0 = 112 ± 2 y0 = 113 ± 0.8	a = 26 ± 11 b = 13 ± 3

Further signal selection and machine learning

Machine Learning setup

selection:

- Baseline selections in each channel applied
- Elliptical H mass selection (not optimized) applied

$$\frac{(m_{bb} - 112)^2}{55^2} + \frac{(m_{sv} - 111)^2}{15^2} < 1$$

ML setup:

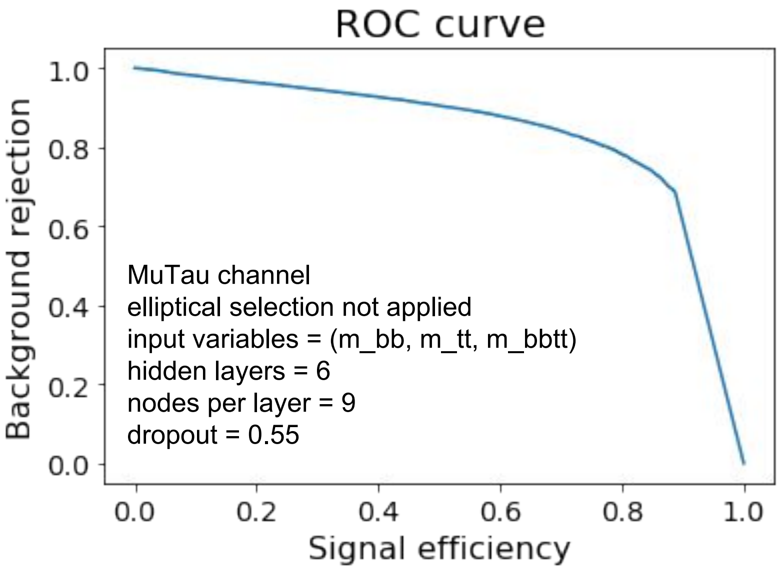
- Study on non-resonant analysis
- Separate training performed for each channel
- 20259 signal events and 29692 background in TauTau channel.
- 80% for training and 20% for testing
- input variables: (m_bb, m_tt, m_bbtt) + customized others
- O(5 layers, 10 nodes per layer)

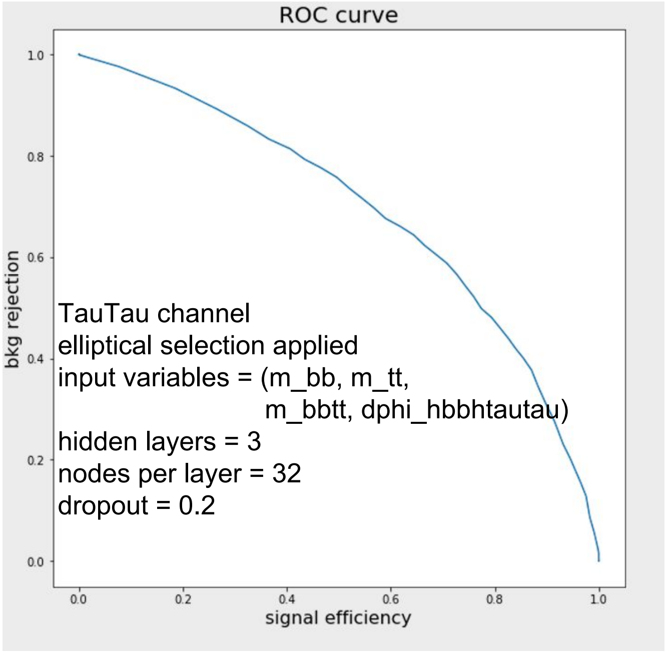
Note that in non-resonant analysis, signal is extracted from the “generalized transverse mass”. So m_bbtt can be used for training.

$$m_{T2} \equiv \min_{\boldsymbol{p}_{T1}+\boldsymbol{p}_{T2}=\boldsymbol{p}_T^{\tau\tau}} \{\max[m_T(m_{b1},\boldsymbol{p}_T^{b1},m_{vis}^{\tau 1},\boldsymbol{p}_{T1}),m_T(m_{b2},\boldsymbol{p}_T^{b2},m_{vis}^{\tau 2},\boldsymbol{p}_{T2})]\}$$

MVA performance

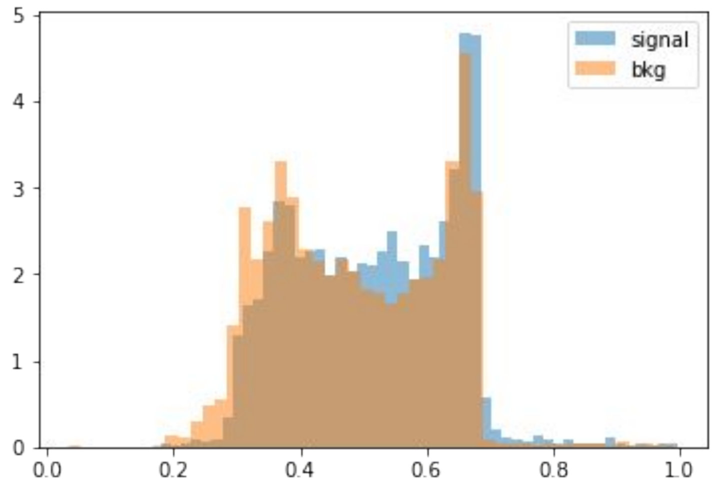
These are only very preliminarily optimized

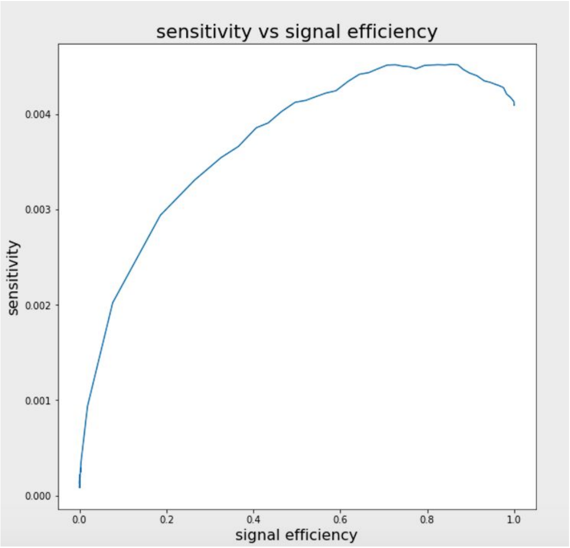


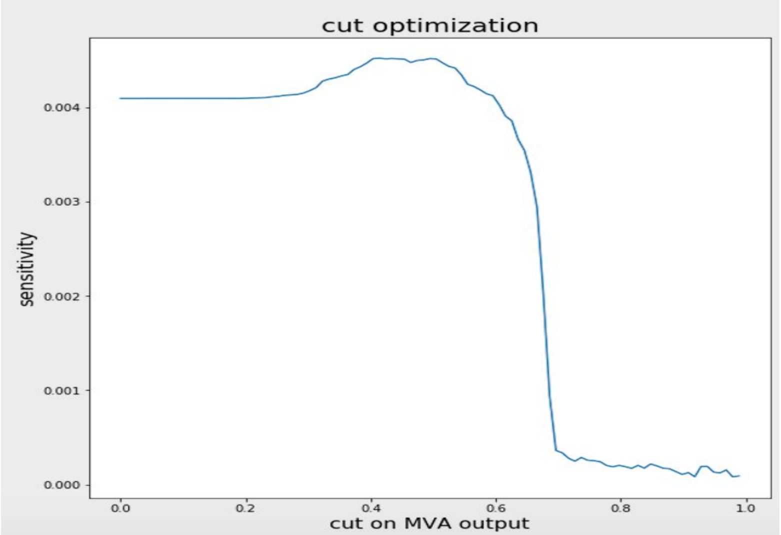


Cut efficiency and sensitivity

settings for this ML:
TauTau channel
elliptical selection applied
input variables = (m_bb, m_tt, m_bbtt, dphi_hbbhtautau)
hidden layers = 3
nodes per layer = 32
dropout = 0.2







Many thanks to Maria Teresa and Konstantin for helping us to understand something of this analysis!!!