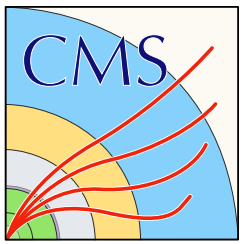
H(125) $\rightarrow a_1 a_1 \rightarrow 4\tau$: MVA Approach (Update)

S. Choudhury [3], S. Consuegra Rodríguez [1], E. Gallo [1], A. Kalogeropoulos [2], T. Lenz [1], *Danyer Pérez Adán* [1], and A. Raspereza [1]

[1]DESY-Hamburg, [2]Princeton University, [3]Indian Institute of Science

June 2019

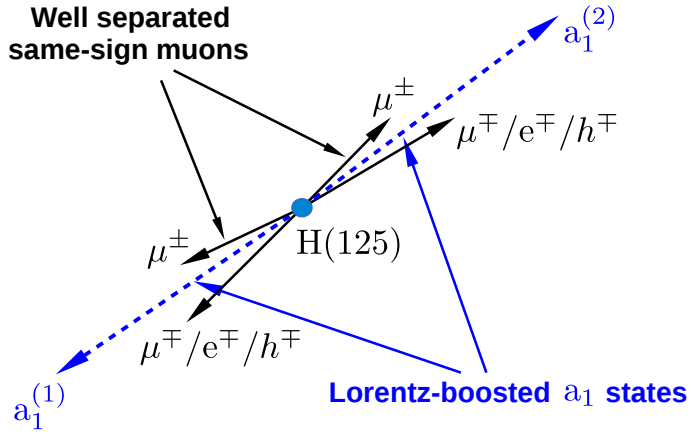


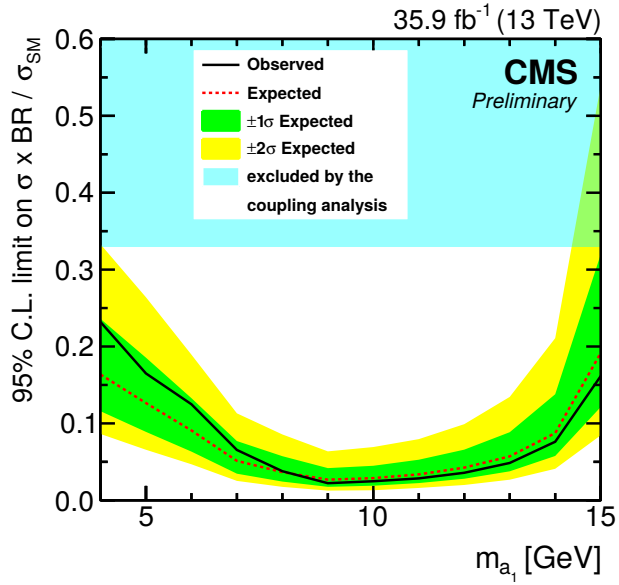
HELMHOLTZ RESEARCH FOR GRAND CHALLENGES

Reminder of 2016 Analysis (CMS-PAS-HIG-18-006)

...

- Highly boosted a_1 bosons
  - Collimated decay products
  - Non-isolated leptons in final state
- Exploit $a_1 \rightarrow \tau_\mu \tau_{1-prong}$ decays
- Same-sign-dimuon Trigger used
- Two isolated muon-track pairs within a $\Delta R_{Iso} = 0.5$
- Signal extraction by means of 2D (m_1, m_2) distribution
- Background shape estimation from Data
- Loss of sensitivity near the boundaries of the mass interval





Selection and Analysis Strategy

New Selection of the 1-prong candidates. See ↷

https://indico.desy.de/indico/event/22965/contribution/3/material/slides/0.pdf

Control Regions (Update)

- "Loose-Iso"
  - At least one of the muon-track pairs is loosely isolated: 1 or more surrounding "soft-tracks"
  - Used for determination of background shape
- "Loose-Iso-N23"
  - At least one of the muon-track pairs is loosely isolated: 2 or 3 surrounding "soft-tracks"
  - Used for background validation
- "Loose-Iso-N45"
  - At least one of the muon-track pairs is loosely isolated: 4 or 5 surrounding "soft-tracks"
  - Used for background validation
- "Semi-Iso"
  - One muon-track pair is not isolated
  - Used for background validation and background shape uncertainty

Signal Extraction

- BDT classification output with Signal and Background normalizations floating

Pros and Cons encountered with the new approach

Previous results ...

- *"The new selection and MVA approach improve and stabilize the sensitivity in the mass range (4-15 GeV). We could extend our mass range up 19 or 21 GeV"*

  $\Rightarrow$ In order to convince ourselves that the mass range could actually be extended up to 21 GeV new Signal MC samples were generated (see next slides)

  $\Rightarrow$ Since interpolation with high dimensional PDFs is almost impossible, the missing mass points (11,12,13 and 14 GeV) were also generated (see next slides)

- *"The BDT is missing signal statistics ..."*

  $\Rightarrow$ The poor signal acceptance limits the statistics for training, therefore new Signal MC samples with 250-300 times higher acceptance were generated (see next slides)

Signal Samples

New generated nominal mass points: ma = 11, 12, 13, 14, 17, 19 and 21 GeV

- ggH:
  - /SUSYGluGluToHToAA_AToTauTau_M-ma_TuneCUETP8M1_13TeV_pythia8/dperezad-RunIISummer16_MINIAOD-28028af67189b3de7224b79195bd0e1d/USER
- VBF:
  - /SUSYVBFToHToAA_AToTauTau_M-ma_TuneCUETP8M1_13TeV_pythia8/dperezad-RunIISummer16_MINIAOD-28028af67189b3de7224b79195bd0e1d/USER
- VH:
  - /SUSYVH_HToAA_AToTauTau_M-ma_TuneCUETP8M1_13TeV_pythia8/dperezad-RunIISummer16_MINIAOD-28028af67189b3de7224b79195bd0e1d/USER
- ttH:
  - /SUSYttH_HToAA_AToTauTau_M-ma_TuneCUETP8M1_13TeV_pythia8/dperezad-RunIISummer16_MINIAOD-28028af67189b3de7224b79195bd0e1d/USER

Signal Samples (Same Charge Double Muon Filter)

New mass points (ma = 4 .. 21 GeV) with Gen-Level Filter

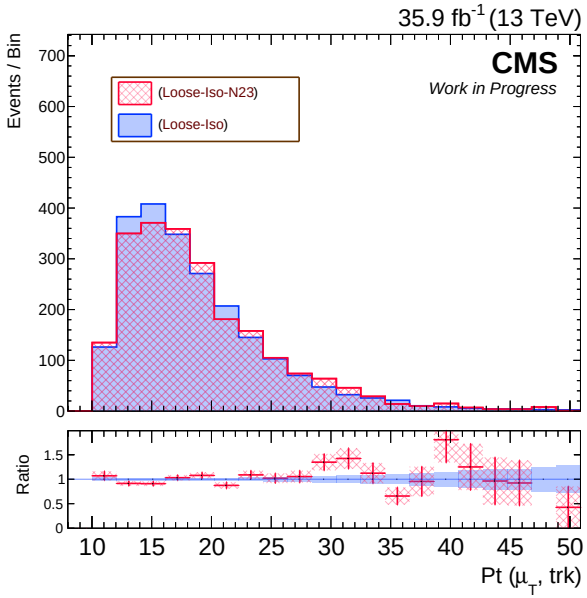
- Gen-Level Filter applied to $gg \rightarrow H(125) \rightarrow 4\tau$:
  - two same-charge μ
  - $Pt > 10(20)$ GeV
  - $|\eta| < 2.4$
  - $\Delta R(\mu, \mu) > 1.5$
- SS DM Trigger Efficiency $> 85\%$: $\rightarrow$ The acceptance for these samples is increased by a factor of ~ 250
- $25K - 70K$ events available for trainning with the current selection
- The GEN-SIM and MINIAOD formats of this samples can be found in:

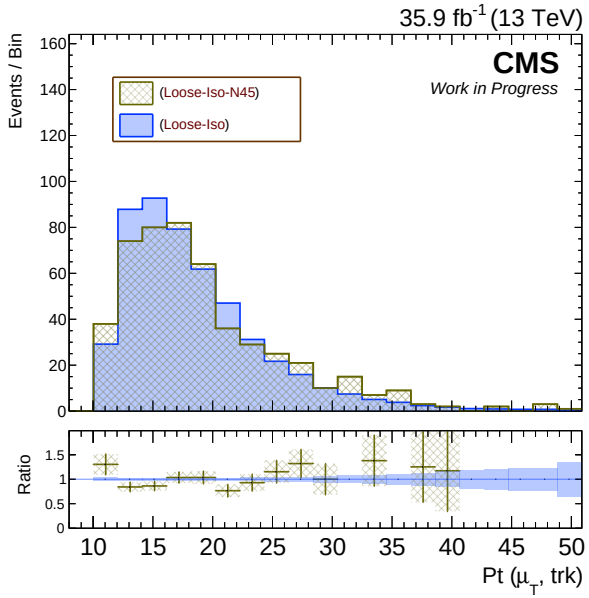
  /pnfs/desy.de/cms/tier2/store/user/dperezad/MC_Production/2016/H2aa_4tau_SCDoMuFilter/ggH/
- Even when these samples where generated with 2016 conditions (MC tuning), they could still be usefull for MVA training in 2017 and 2018, one just need to reprocess the GEN-SIM format with the detector conditions of those years

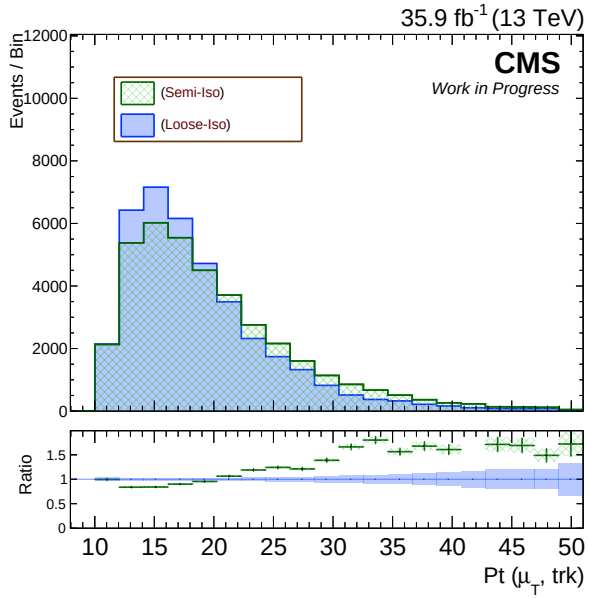
Background Validation

- The distributions of each of the 10 variables used in the BDT was compared in the different validation regions
- The largest difference was found to be between the "Loose-Iso" (Bkgd Model) and the "Semi-Iso" validation region

Reconstructed Pt of the trailing muon and corresponding 1-prong track



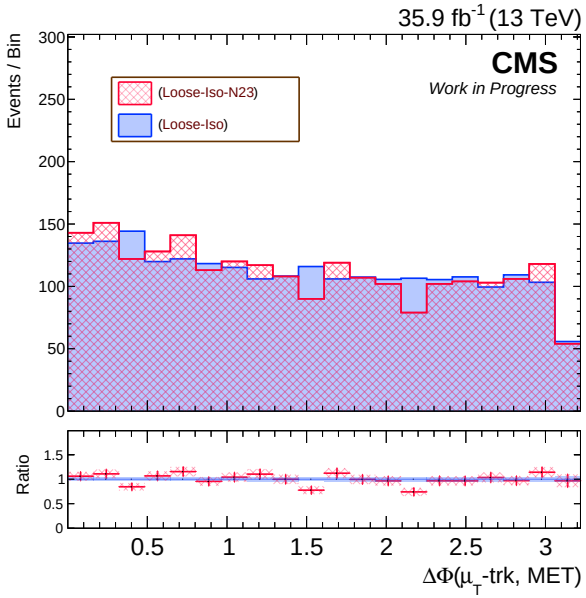


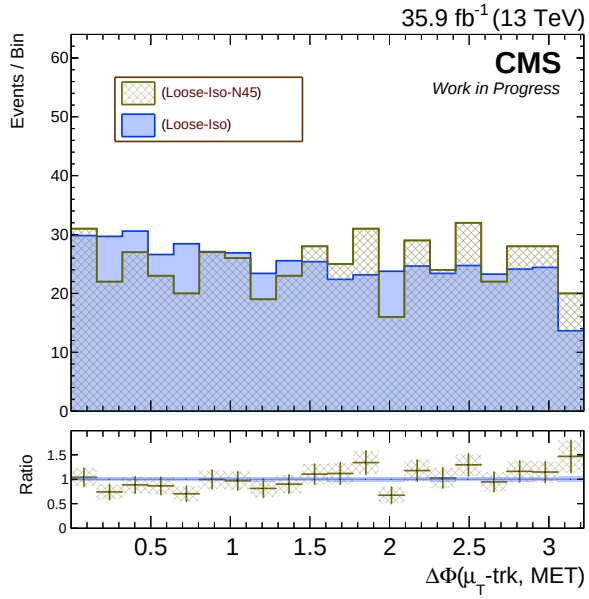


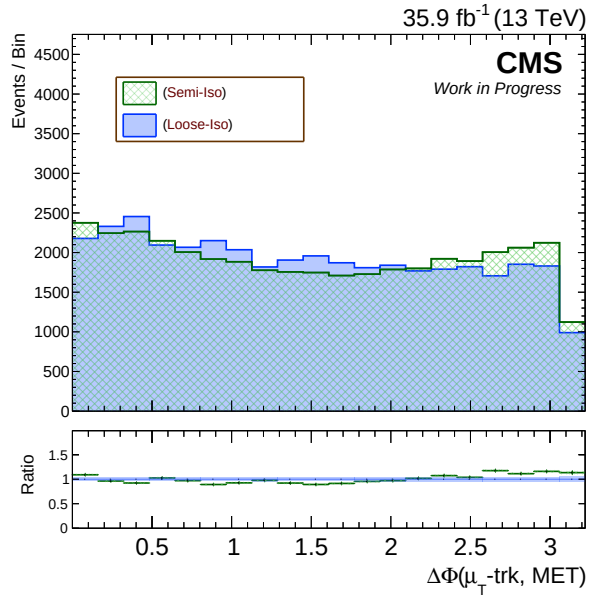
Background Validation

- The distribution of each of the 10 variables used in the BDT was compared in the different validation regions
- The largest difference was found to be between the "Loose-Iso" (Bkgd Model) and the "Semi-Iso" validation region

$\Delta\Phi$ between the MET and the 4 vector of trailing muon and its surrounding track



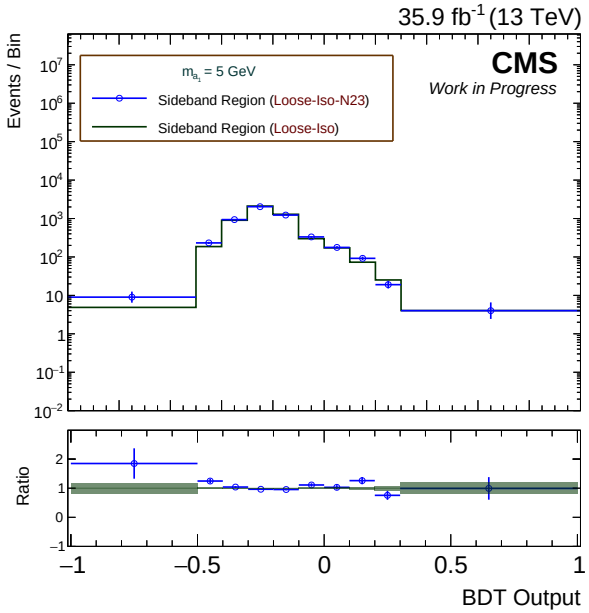


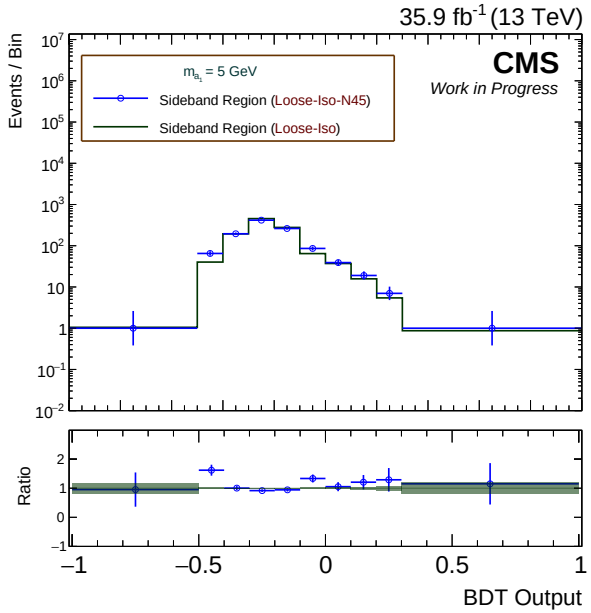


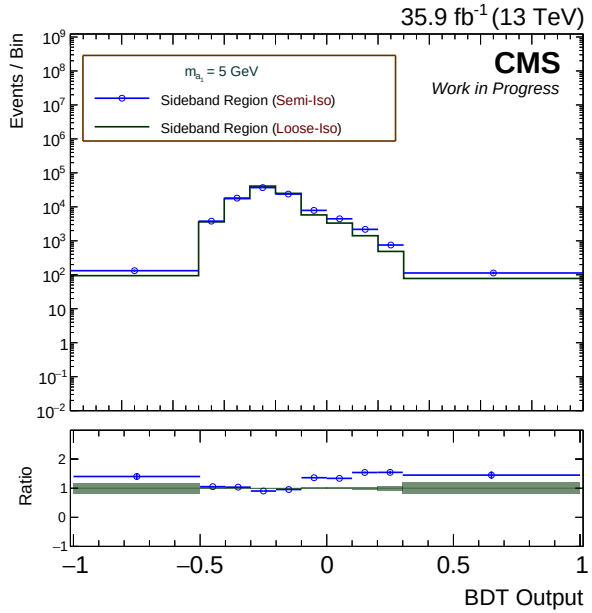
Background Validation

...

- The BDT is used to classify the events, and the distribution obtained in the "Loose-Iso" (Background Model) is compared to the ones obtained in "Loose-Iso-N23", "Loose-Iso-N45" and "Semi-Iso"
- The largest difference (as expected from the variable comparison) is between "Loose-Iso" and "Semi-Iso"
- The difference between "Loose-Iso" and "Semi-Iso" is taken as a background shape uncertainty in the limit extraction





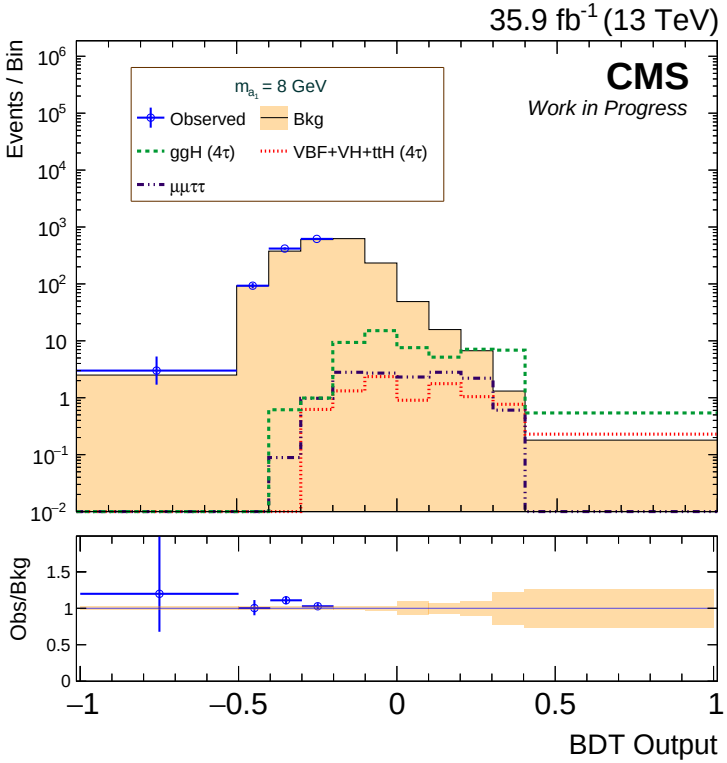


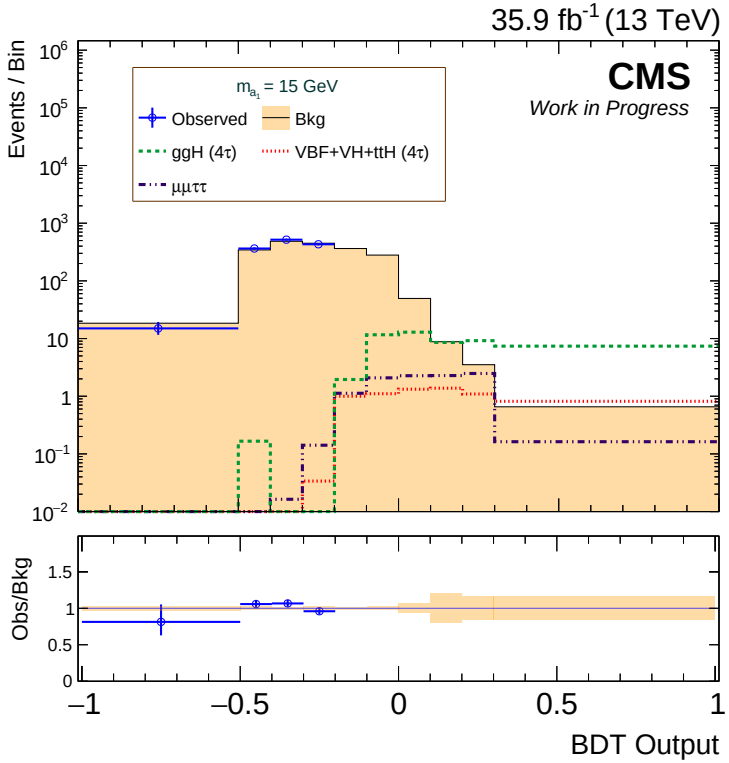
Updated results with 2016 Dataset

...

- Background distribution is obtained after performing fit to data under the background-only hypothesis

$$\text{Branching ratio : } B(H(125) \to a_1 a_1) \cdot B^2(a_1 \to \tau\tau) = 20\%$$





Updated Results with 2016 dataset. Comparison with 2D discriminant.

...

- Expected limits are set in terms of 95% CL on $\frac{\sigma}{\sigma_{SM}} \times B(H(125) \to a_1 a_1) \cdot B^2(a_1 \to \tau\tau)$

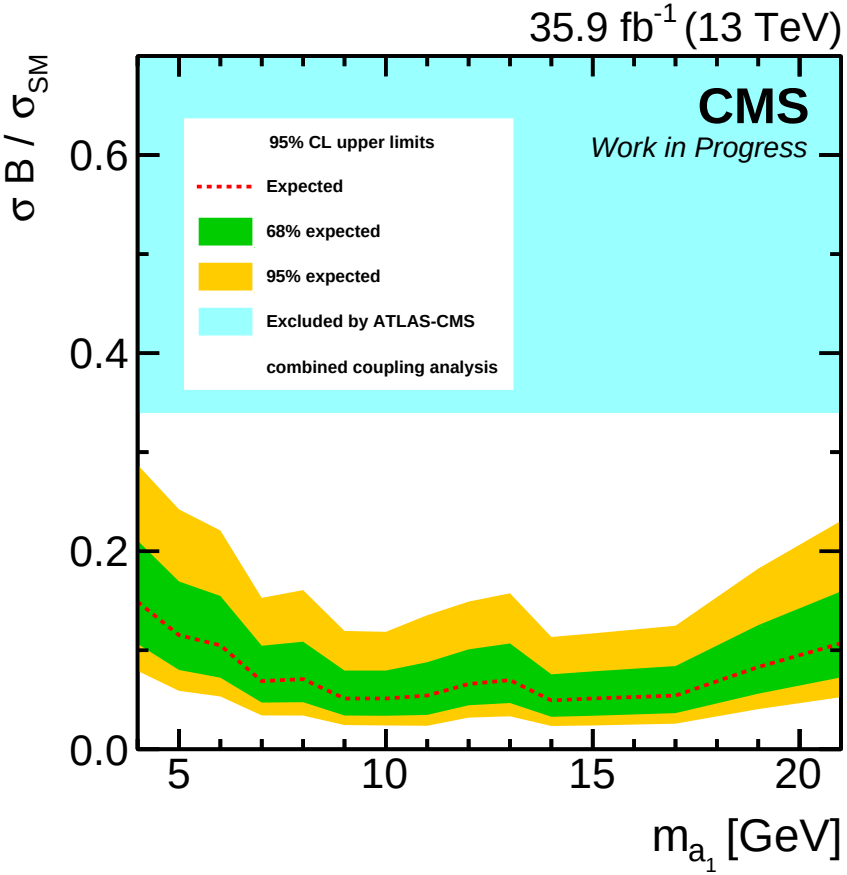


Figure: MVA (BDT) discriminator based

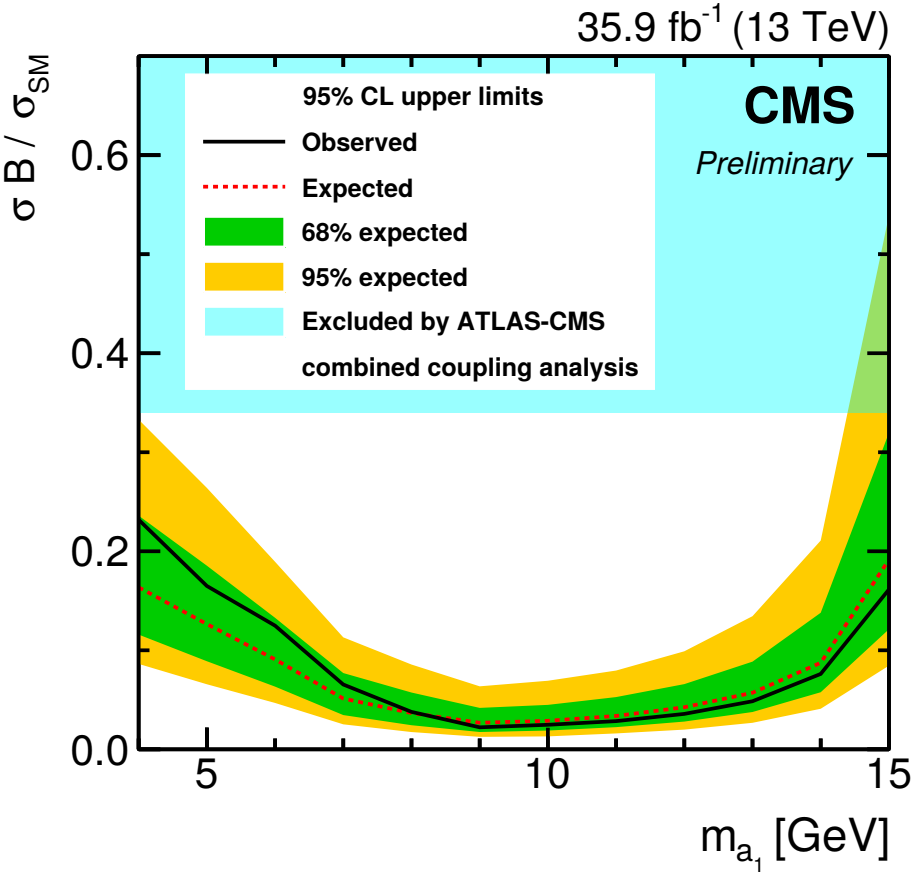


Figure: 2D discriminator based

Summary

...

- The analysis is performing very well in the mass interval from 4 to 21 GeV
- The MVA based approach improves the analysis for low and high masses
- For full Run2 analysis we would need to officially request the missing mass points for 2016 and all samples for 2017 and 2018 (in case we decide to proceed with the MVA based approach)

Summary

...

- The analysis is performing very well in the mass interval from 4 to 21 GeV
- The MVA based approach improves for low and high masses by a factor ~ 4
- For full Run2 analysis we would need to officially request the missing mass points for 2016 and all samples for 2017 and 2018 (in case we decide to proceed with the MVA based approach)

Thanks for your attention!

Backup