H(125) $\rightarrow a_1 a_1 \rightarrow 4\tau$: MVA Approach. Prospects for whole Run 2 analysis.

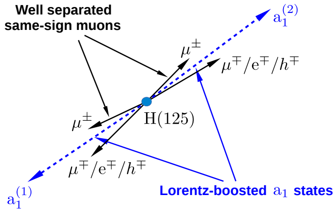
S. Choudhury [3], S. Consuegra Rodríguez [1], E. Gallo [1], A. Kalogeropoulos [2], T. Lenz [1],
Danyer Pérez Adán [1], and A. Raspereza [1]

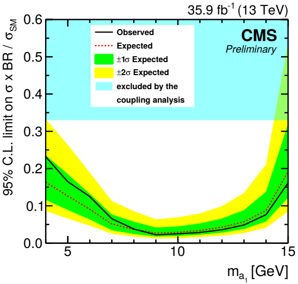
[1]DESY-Hamburg, [2]Princeton University, [3]Indian Institute of Science

April 2019

...

- 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





New Selection and Analysis Strategy

New Selection of the 1-prong candidates

- One hard ($p_T > 2.5$ GeV) track (1-prong) is selected around each muon:
  - Surrounding track must be in a cone of $\Delta R = 1.5$
  - If more than 1 track is found, the one with highest $pT_{sum} = (\overrightarrow{p_\mu} + \overrightarrow{p_{trk}})_T$ is selected
- The number of "all-tracks" and "soft-tracks" ($1.5 < p_T < 2.5$ GeV) within a $\Delta R = 0.5$ around each object is counted

Signal Region

- Each object must be completely Isolated

Control Regions

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

Signal Extraction: BDT classification

15 variables used in total (sorted by performance)

- $\Delta R(\mu_1, trk_1)$
- $\Delta R(\mu_2, trk_2)$
- $m(\mu_1, trk_1)$
- $m(\mu_1, trk_2)$
- $m(\mu_1, \mu_2)$
- $\Delta R(\mu_1, \mu_2)$
- $m(\mu_1, , trk_1, \mu_2, trk_2)$
- 8 others ...

Training

- Training performed with signal samples ("Signal-Region") vs data control region ("Semi-Iso")
- All signal channels (ggH, VBF, VH, ttH, $a_1 a_1 \rightarrow 2\mu 2\tau$) included and weighted by their cross section
- Fairly good statistics for background ($\sim 100K$) but very poor for signal ($\sim 2K$)

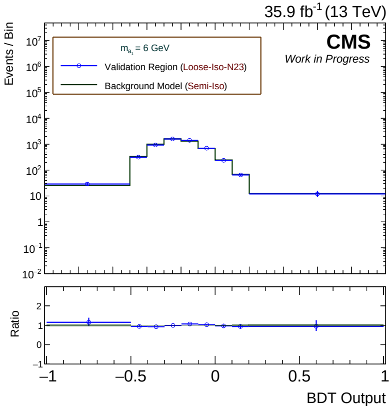
Procedure followed for signal extraction

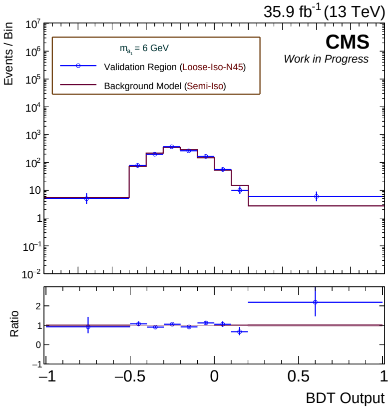
- Extract signal by means of a binned Max-likelihood fit applied to the BDT Output
- Performed with background and signal normalizations freely floating

Background Validation

...

- BDT distribution obtained in the "Semi-Iso" CR (Background Model) is compared to the ones obtained in "Loose-Iso-N23" and "Loose-Iso-N45"
- The difference between "Semi-Iso" and "Loose-Iso-N23" is taken as a shape uncertainty



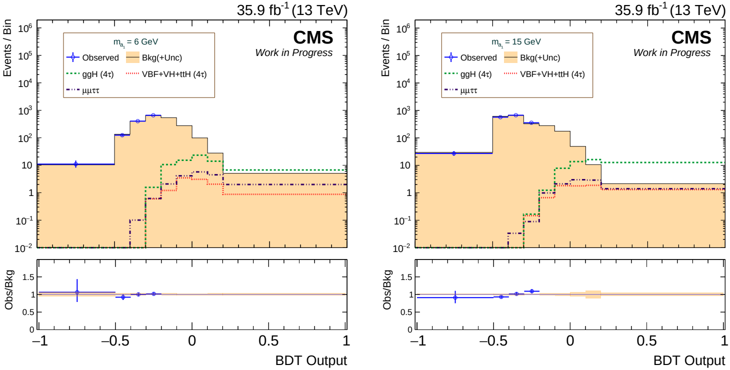


Final Discriminant

...

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

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



Results with 2016 dataset. Comparison with 2D discriminant.

...

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

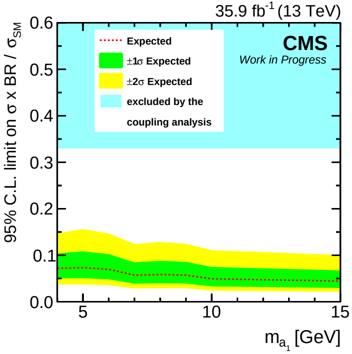


Figure: MVA (BDT) discriminator based

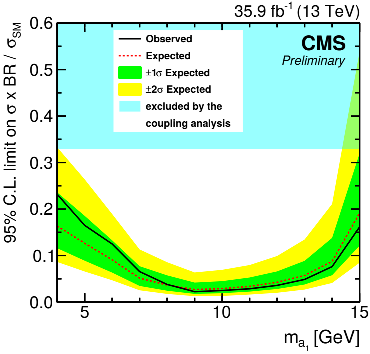


Figure: 2D discriminator based

...

- Mass points 8, 9 and 10 GeV get worse by a factor of ~ 1.8 because of the new selection
- Other mass points are improved by a factor that goes from 1.1 to 4.3

Extrapolation of results for whole Run 2 dataset.

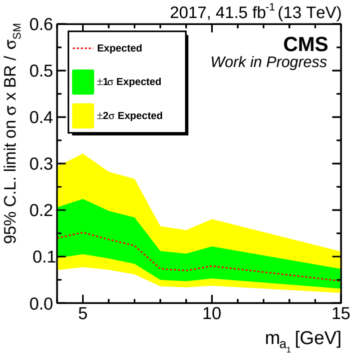
Issues ...

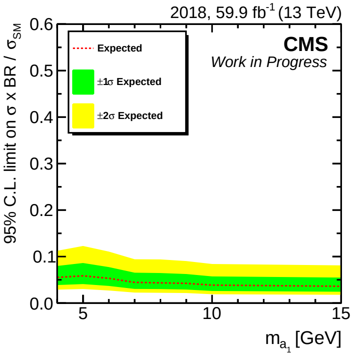
- Double Muon Trigger barely present in 2017 date taken (only $\sim 7\ fb^{-1}$)
- Either use those 7 fb^{-1} of DM data or use Single Muon Triggers
- Single Muon Triggers highly suppress signal acceptance for low mass points (anyway seems to be the best option)

Aproximations ...

- Since Signal samples for 2017 and 2018 not yet available, the 2016 samples were used in all years
- Expected limits obtained by generating Toys, not with the real data
- For 2017:
  - The Trigger IsoMu27 was used on 2017 SM dataset while for signal the Trigger IsoMu24 was used ($p_T^{Cut} > 28$ for both cases)
- For 2018:
  - While the 2018 dataset DM Ntuples become available ...
  - The 2016 DM dataset was used and the Bkg normalization was scaled up by the luminosity ratio
- Systematic uncertainties remain the same (pessimistic scenario ...?)
- Statistical uncertainties scaled down by luminosity ratio with respect to 2016

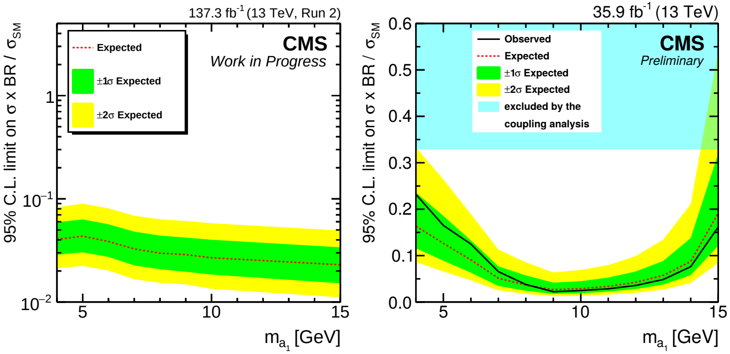
Extrapolation of results for whole Run 2 dataset.





...

- Misperformance of SM Trigger clearly seen for very low masses



...

- Improvement with respect to CMS-PAS-HIG-18-006 is of the order of **1 to 8**, depending on the mass point

Summary

...

- The new selection and MVA approach stabilize the sensitivity in the mass range
- We could extend our mass range up 19 or 21 GeV
- The BDT is missing signal statistics, and this clearly limits its potential (Maybe 2017 and 2018 signal samples would help ...)
- In 2017 we are not doing as well as in 2016 due to the absence of the Double Muon Trigger (... any idea different from using SM Triggers?)
- Despite of the "Pessimistic" scenario (in my opinion) considered here we have good improvements for low and high masses for the Run 2 analysis

To do list ...

- Request Signal MC samples (ggH, VBF, VH and ttH) for 2017 and 2018 (4 ... 15 ... 21 GeV) with 1 GeV step
- Request Signal MC samples (ggH, VBF, VH and ttH) for 2016 (11, 12, 13, 14, 16 ...21 GeV) with 1 GeV step
- It would be nice to have at least one set of those samples (2016 or 2017 or 2018) with 10 times more statistics. That would really increase the BDT performance.
- Ntuplize 2018 DM dataset

Thanks for your attention!

Backup