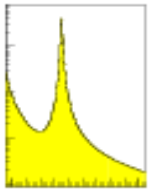
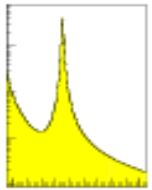




Event 4

Machine Learning Techniques in the CMS Higgs to Di-muons Search

Darin Acosta, Pierluigi Bortignon, Dimitri Bourilkov, Andrew Brinkerhoff, Andrew Carnes, Sergei Gleyzer, Brendan Regnery

University of Florida
for the CMS Collaboration

CHEP 2018, Sofia, Bulgaria, July 10, 2018

Physics Motivation

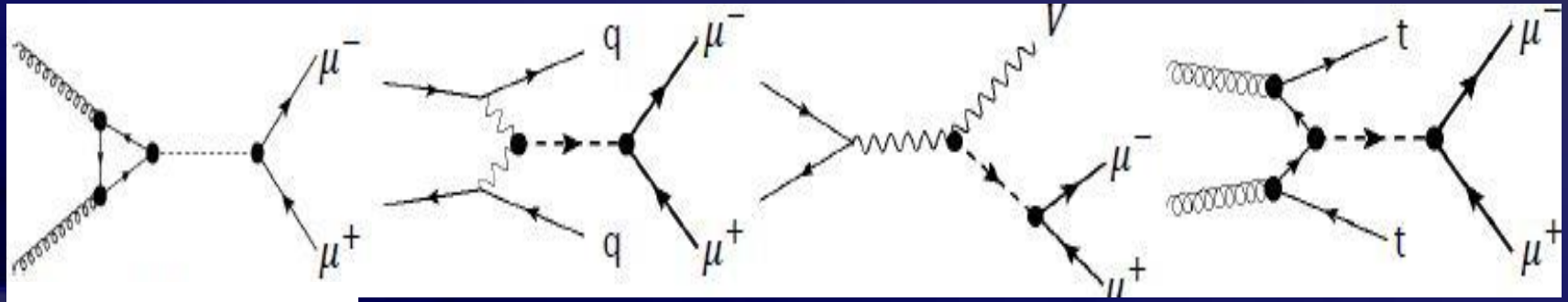
- Probe the Higgs couplings to the 2nd generation (couplings to the 3rd generation and to gauge bosons show SM-like features)
- A very difficult analysis:
  - small branching ratio of the muon decay ~ $2 \cdot 10^{-4}$
  - large irreducible backgrounds: dominated by Drell-Yan to $\mu\mu$ and di-top to $\mu\mu$ (+ di-bosons and single top)
- On the plus side: excellent mass resolution for muon pairs, narrow Higgs peak
- HIG-17-009 (PAS & PAPER, 35.9 fb^{-1}, 2016 data)

Production Channels

Several channels with different cross sections, topologies and backgrounds:

gg fusion, VV fusion, associate VH, ttH

ratios 1 : 0.078 : 0.046 : 0.010



V = W± or Z

Event Classification

- Boosted Decision Tree (BDT) in TMVA (ROOT): binary signal-background separation; BDT score [-1, 1]
- Signal trained on: gg→H (ggF), VV→H (VBF), associate VH
- Variables with some discriminating power for S/B used

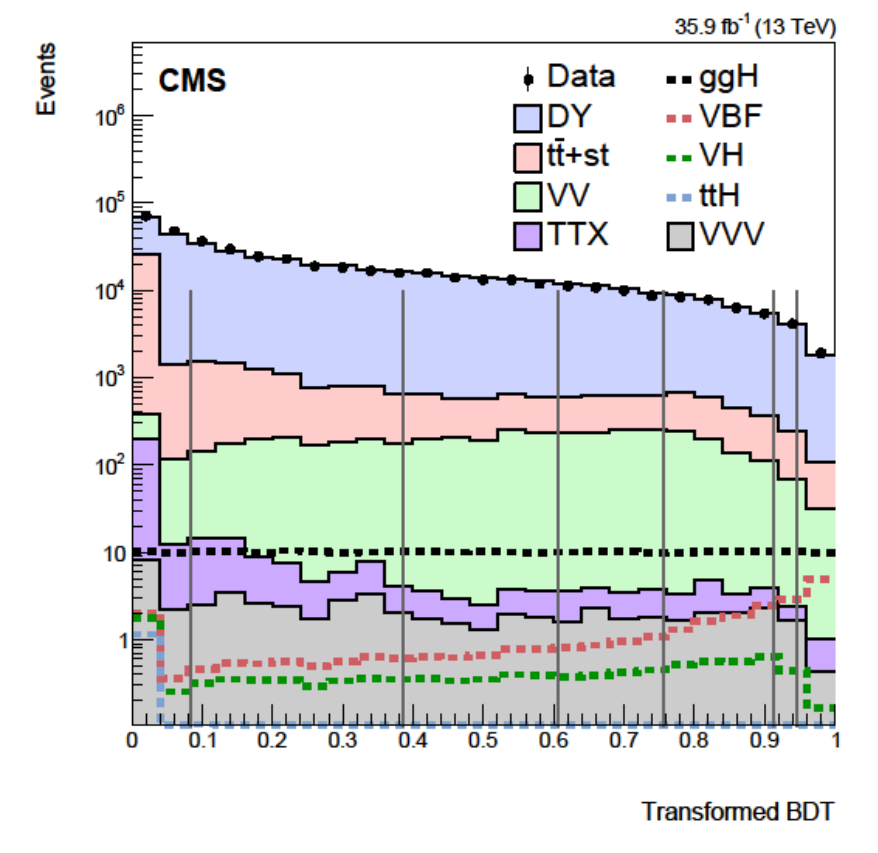
- The p_T and η of the dimuon system
- The $|\delta\eta|$ and $|\delta\phi|$ between the muons
- The η values of the two highest–p_T jets
- The masses of the two highest–mass dijet pairs
- The $|\delta\eta|$ bewteen the jets in the two highest–mass pairs
- The number of jets with $|\eta| < 2.4$ and $|\eta| > 2.4$
- The number of jets passing the CSVv2 medium b-tag working point
- The E_T^{miss}

Variable Ranking by BDT

- BDT ranks the used variables by importance
- The di-muon system p_T and η come on top
- Jet variables important for VV→H
- Signal sample split in 3 independent chunks: training, test, limit setting
- Background: train/test

Feature	Importance
Dimuon p_T	5.342e-02
Dimuon η	3.896e-02
$\lvert\delta\phi(\mu\mu)\rvert$	3.500e-02
Number of medium b-tags	2.842e-02
$\eta(jet1)$	1.724e-02
E_T^{miss}	1.706e-02
Number of forward jets	1.218e-02
$\lvert\delta\eta(jj_1)\rvert$	8.310e-03
Number of central jets	8.036e-03
$\eta(jet2)$	7.460e-03
$\lvert\delta\eta(\mu\mu)\rvert$	6.793e-03
$M(jj_1)$	6.546e-03
$\lvert\delta\eta(jj_2)\rvert$	3.304e-03
$M(jj_2)$	2.199e-03

BDT Output



Transformed BDT response in quantile in data and MC:

sum of all signal events in all production processes has a uniform distribution. The VBF channel has the highest BDT scores. B-tagging helps to suppress the top backgrounds.

Decision Tree Autocategorizer I

- "Greedily" optimize the sensitivity (and the expected upper limit)
- Events categorized on:
  - BDT score: from -1 ⇒ background-like to +1 ⇒ signal-like
  - the maximum muon $|\eta|$: proxy for the mass resolution (2.8 to 7.6 GeV FWHM for barrel/endcap categories: 0.0-0.9-1.9-2.4 in $|\eta|$)
- Events (S/B) in 0.5 GeV bins for [120, 130] GeV
- Significance of each bin i given by S/√B: add in quad

$$\text{Net Significance} = \sum_{c,i} S_{c,i}^2 / B_{c,i}$$

- We start with one (inclusive) category c: 1st iteration

Decision Tree Autocategorizer II

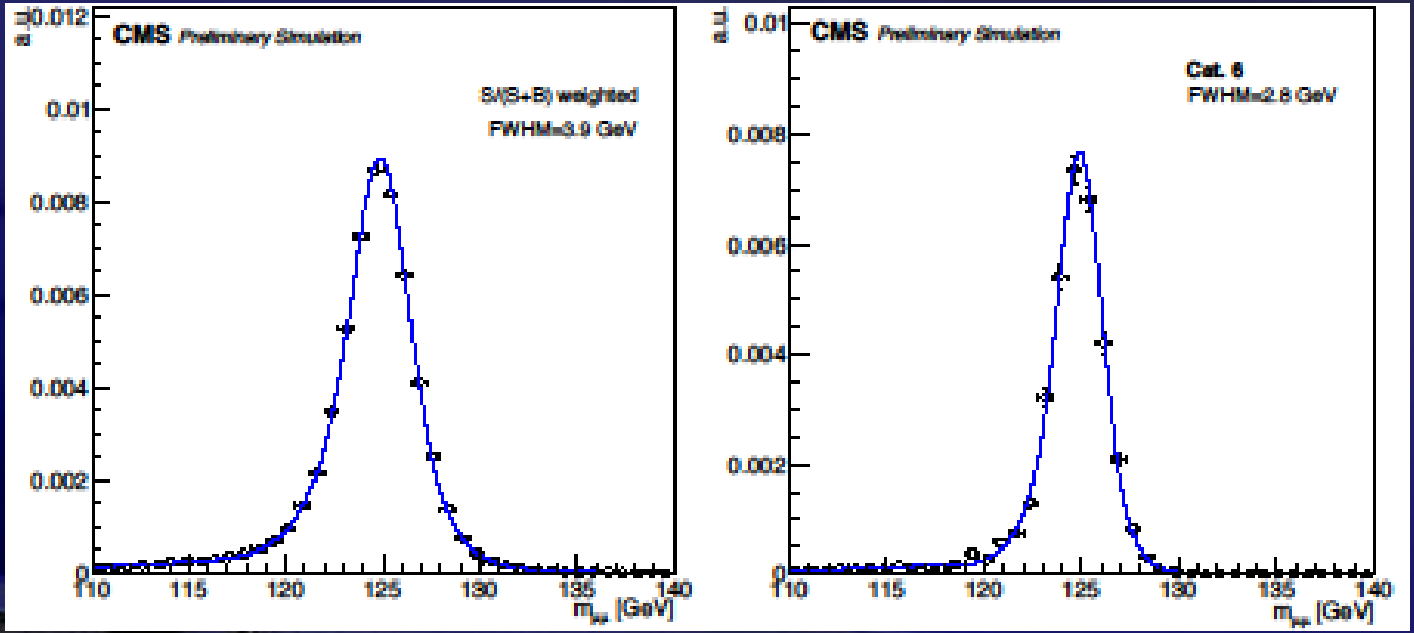
- Fine binning ⇒ central bins contribute most ⇒ account for resolution
- Check all split values for $|\eta|$ and BDT score and find the values giving maximum gain for $c \rightarrow c1 + c2$

$$\text{Gain} = \sum_i S_{c1,i}^2 / B_{c1,i} + \sum_i S_{c2,i}^2 / B_{c2,i} - \sum_i S_{c,i}^2 / B_{c,i}$$

- Split in 2 categories by max gain - either on BDT score or $|\eta|$: 2nd iteration
- Next iterations: repeat the procedure on the newly created categories, "greedily" going for the max gains
- Stop when desired # of categories is reached

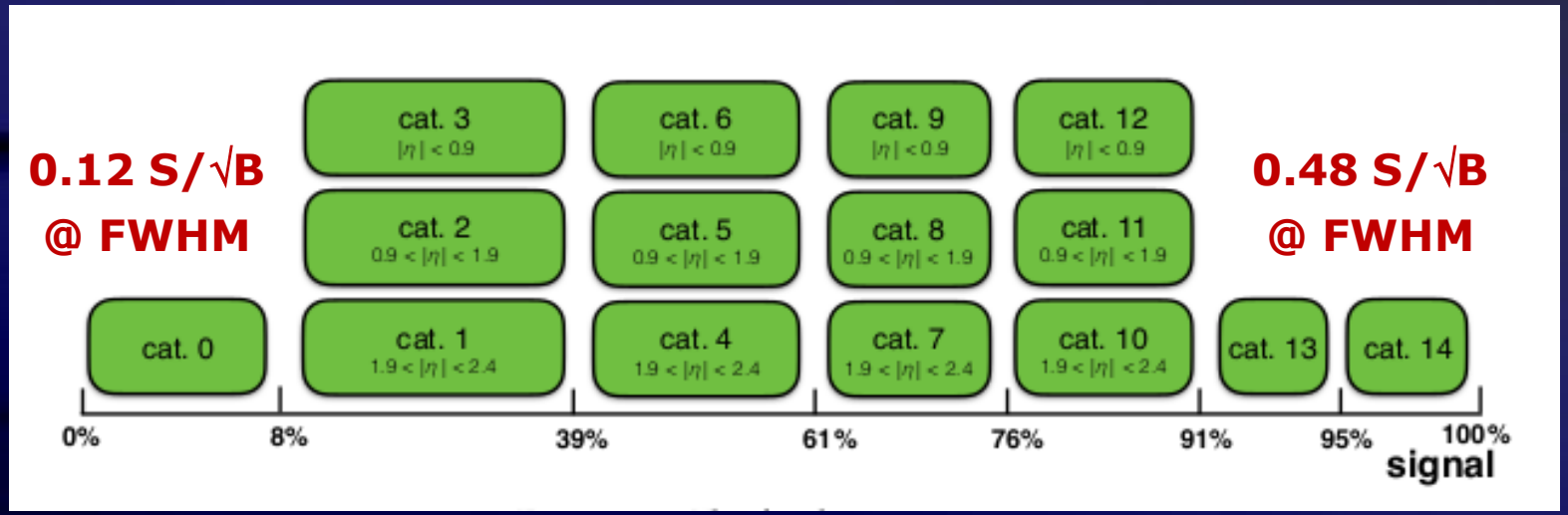
Mass Resolution - Categories

Signal model compared to MC: weighted sum of contributions from all processes and all categories, and for one of the best - category 6 (normalized to unity)



Final Categories I

- Simplification: rounding some of the cuts (~ no loss)
- 15 event categories
- Gain of 23% relative to the 7/8 TeV BDT categories



BDT response quantile [%]

Final Categories II

BDT response quantile [%]	Maximum muon $\lvert\eta\rvert$	ggH [%]	VBF [%]	WH [%]	ZH [%]	ttH [%]	Signal	Bkg./GeV @125 GeV	FWHM [GeV]	Bkg. fit function	$S/\sqrt{B}$ @FWHM
0 – 8	$\lvert\eta\rvert < 2.4$	4.9	1.3	3.3	6.3	32	21.2	3.13×10^3	4.2	mBW $\cdot B_{deg4}$	0.12
8 – 39	$1.9 < \lvert\eta\rvert < 2.4$	5.6	1.7	3.9	3.5	1.3	22.3	1.34×10^3	7.2	mBW $\cdot B_{deg4}$	0.16
8 – 39	$0.9 < \lvert\eta\rvert < 1.9$	10	2.8	6.5	6.4	5.2	41.1	2.24×10^3	4.1	mBW $\cdot B_{deg4}$	0.29
8 – 39	$\lvert\eta\rvert < 0.9$	3.2	0.8	1.9	2.1	3.5	12.7	7.83×10^2	2.9	mBW $\cdot B_{deg4}$	0.18
39 – 61	$1.9 < \lvert\eta\rvert < 2.4$	2.9	1.7	2.7	2.7	0.3	11.8	4.37×10^2	7.0	mBW $\cdot B_{deg4}$	0.14
39 – 61	$0.9 < \lvert\eta\rvert < 1.9$	7.2	3.3	6.1	5.2	1.3	29.2	9.70×10^2	4.0	mBW $\cdot B_{deg4}$	0.31
39 – 61	$\lvert\eta\rvert < 0.9$	3.6	1.1	2.6	2.2	0.9	14.5	4.81×10^2	2.8	mBW	0.26
61 – 76	$1.9 < \lvert\eta\rvert < 2.4$	1.2	1.5	1.8	1.7	0.2	5.2	1.48×10^2	7.6	mBW $\cdot B_{deg4}$	0.11
61 – 76	$0.9 < \lvert\eta\rvert < 1.9$	4.8	3.6	4.5	4.4	0.7	20.3	5.12×10^2	4.2	mBW $\cdot B_{deg4}$	0.29
61 – 76	$\lvert\eta\rvert < 0.9$	3.2	1.6	2.3	2.1	0.6	13.1	3.22×10^2	3.0	mBW	0.28
76 – 91	$1.9 < \lvert\eta\rvert < 2.4$	1.2	3.1	2.2	2.1	0.2	5.8	1.04×10^2	7.1	mBW $\cdot B_{deg4}$	0.14
76 – 91	$0.9 < \lvert\eta\rvert < 1.9$	4.4	8.7	6.2	6.0	1.1	20.3	3.60×10^2	4.2	mBW $\cdot B_{deg4}$	0.35
76 – 91	$\lvert\eta\rvert < 0.9$	3.1	4.0	3.8	3.6	0.9	13.7	2.36×10^2	3.2	mBW	0.34
91 – 95	$\lvert\eta\rvert < 2.4$	1.7	6.4	2.5	2.6	0.5	8.6	96.0	4.0	mBW	0.28
95 – 100	$\lvert\eta\rvert < 2.4$	2.0	19	1.5	1.4	0.7	13.7	83.4	4.1	mBW	0.48
0 – 100	$\lvert\eta\rvert < 2.4$	59	61	51	52	49	253	1.30×10^4	3.9		

Outlook

- Boosted Decision Trees help to improve the separation of signal and background for the different channels of the very challenging $H \to \mu\mu$ search
- By creating automatic event categories, the highest sensitivity of the analysis can be achieved
- Future work: try additional variables and deep neural networks