Data-driven analyses in the search for the Higgs decay to two muons

Xunwu Zuo, on behalf of the CMS collaboration

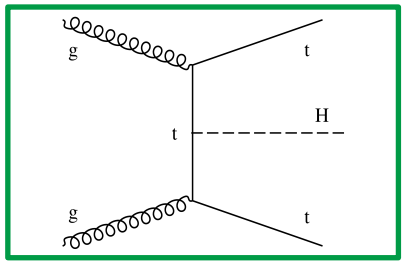
October 29, 2020

Context

- The $H \to \mu\mu$ analysis by CMS on Run 2 dataset [arXiv 2009.04363] is performed in **4 event categories**, with **2 different analysis strategies**.

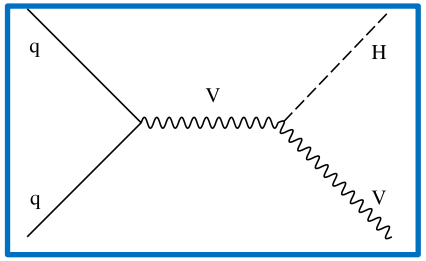
Data-driven

ttH category



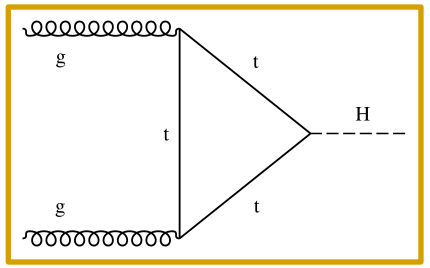
Tagged by b-jets from the top decay

VH category



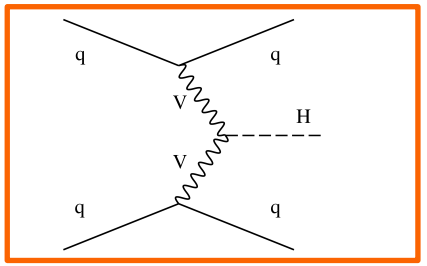
Tagged by leptons from the decay of W or Z

ggH category



Dimuon final state, no extra b-jets or leptons

VBF category



Tagged by two energetic jets accompanying the Higgs

MC-based

- The **data-driven (DD) strategy**: sub-categorization based on MVA discriminators uncorrelated with $m_{\mu\mu}$, and signal extraction by parametric fit to the $m_{\mu\mu}$ spectrum in each sub-category.
- The **MC-based strategy:** MVA discriminator with $m_{\mu\mu}$ information, no sub-categorization, and signal extraction by template fit to the MVA discriminator.

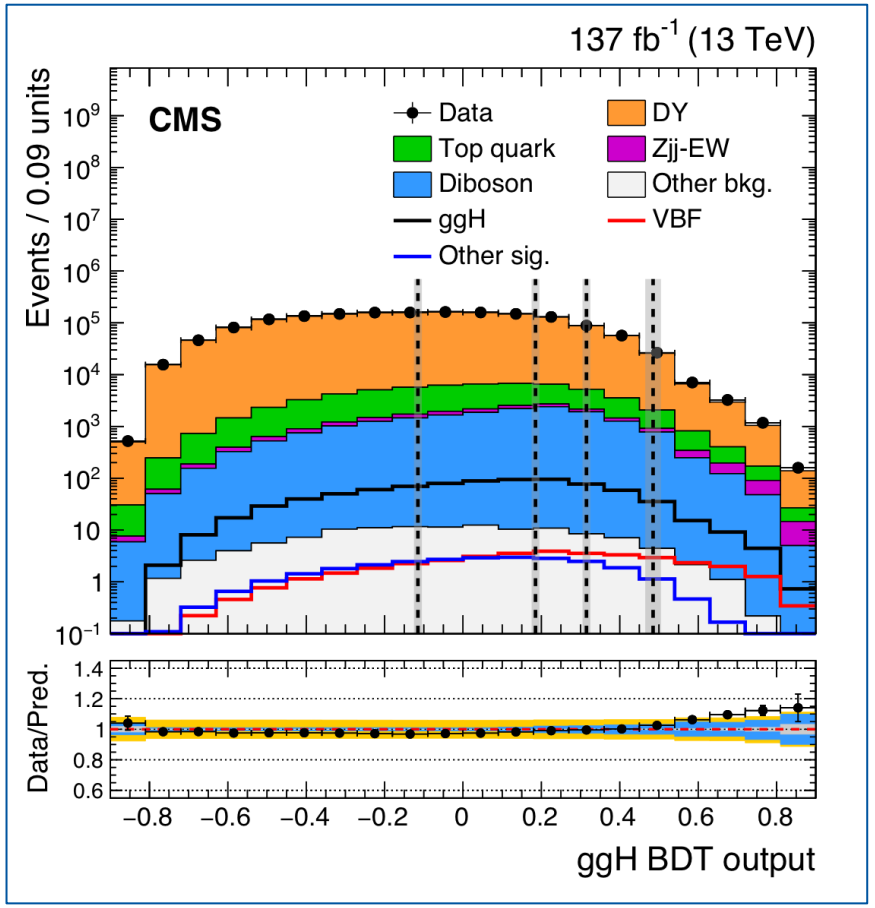
DD strategy – BDT categorization

- The analyses are performed independently in the **ggH**, **ttH**, and **VH** categories, following very similar procedures.

- A BDT is trained in each event category.
  - Use input variables uncorrelated with $m_{\mu\mu}$.
  - Involve mass resolution in categorization by using $1/\sigma(m_{\mu\mu})$ as (per-event) weight for signal events in training.

- BDT categorization boundaries optimized for overall expected significance

BDT in ggH category

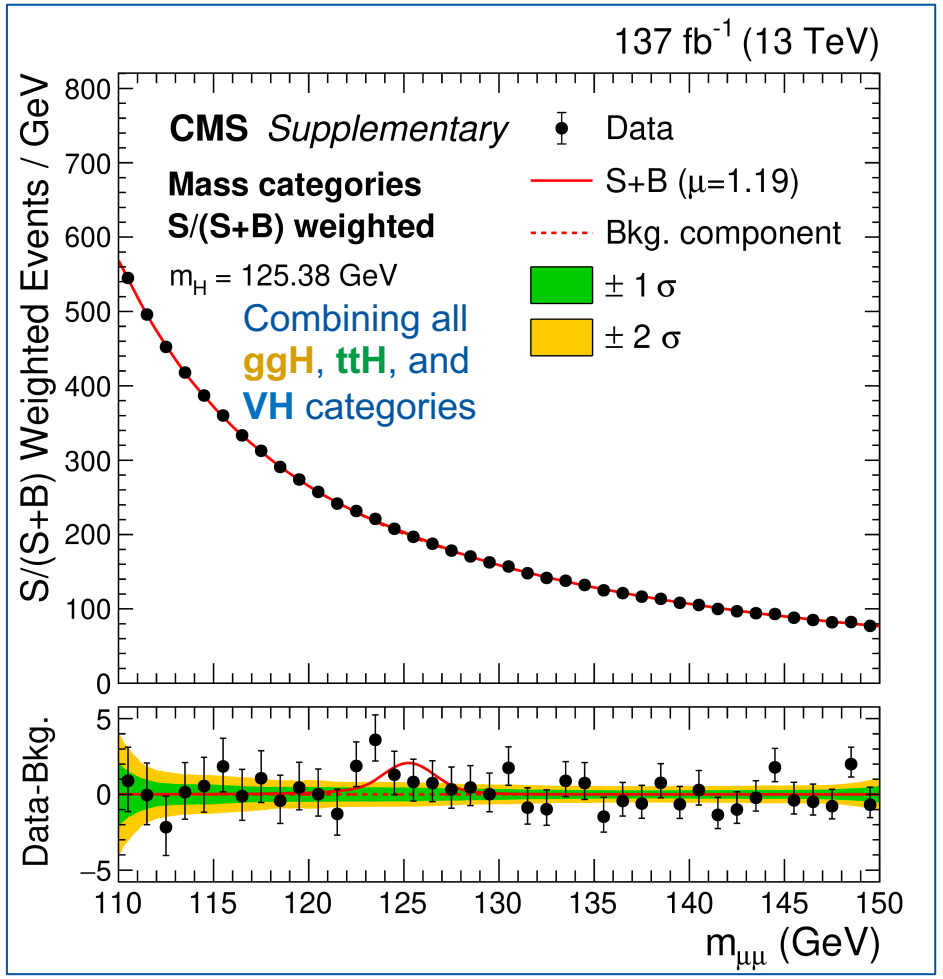


DD strategy – fit procedures

- The **signals** are modeled independently in each sub-category by **double-sided crystal ball functions.**

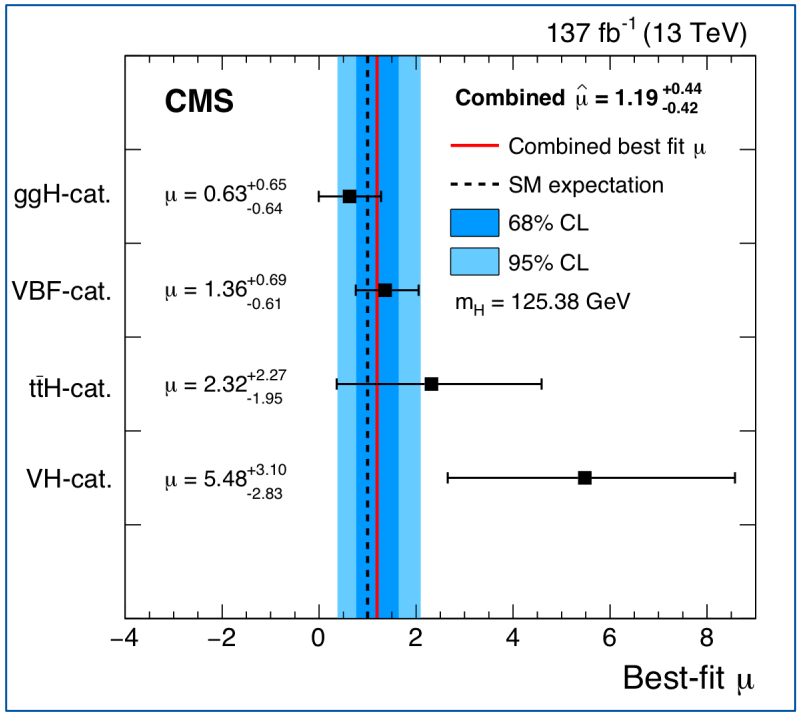
- The **background** is modeled by **parametric functions**, chosen from a **group of functional forms.**
  - In the ggH categories in particular, the background is **discretely profiled**, taking the envelope of the likelihood scans of multiple functions.

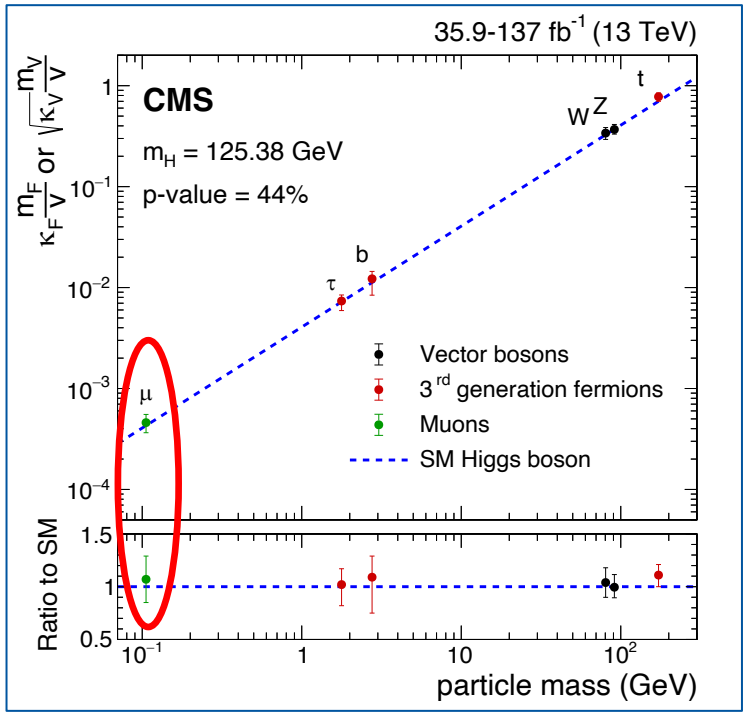
- Studies on toy datasets to evaluate the potential measurement bias induced by the choice of background function.



Combined results

- Data-driven approach used in **ggH**, **ttH**, and **VH** categories, covering most of the phase-space of the Higgs production.
- The combination of the DD and MC-based analyses amounts to the **first evidence of the Higgs to muon decay** and the **most precise measurement of the Higgs to muon coupling**!





Back-up

Discrete profile likelihood

- Different function candidates are fitted to the data.
- Likelihood of different fits overlaid, and the maximum of them, or the minimum in -log(L), is taken for every different signal strength hypothesis.
- This profile of likelihood is used in the evaluation of signal strength.