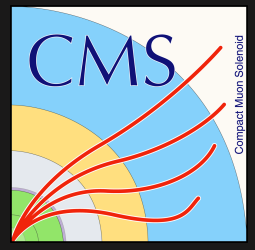
MC template-based search for Higgs boson decays into a pair of muons targeting the VBF production channel

Dmitry Kondratyev, on behalf of the CMS collaboration

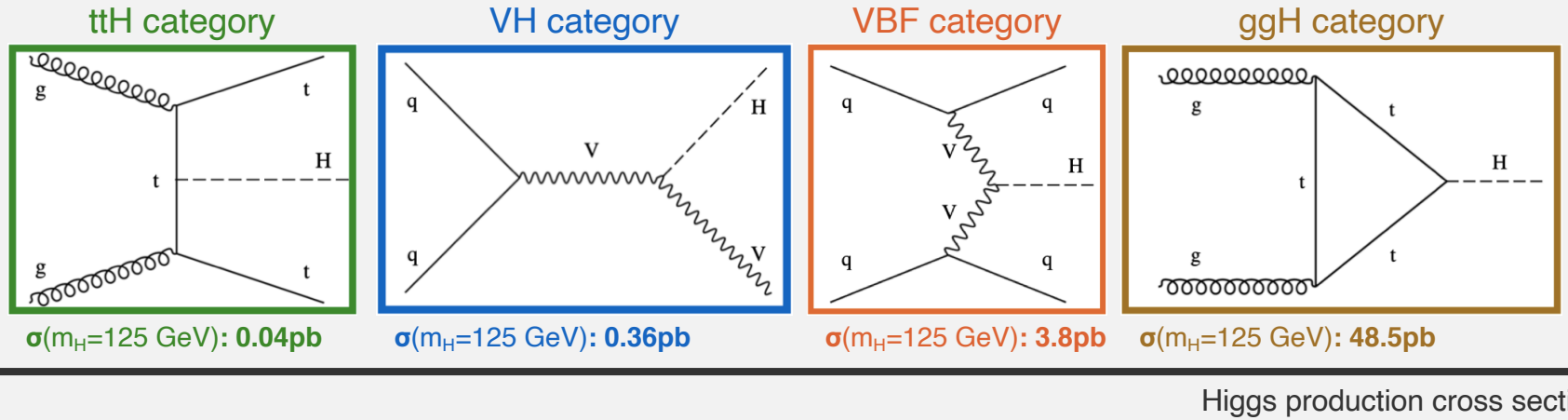
Higgs 2020 Conference: Young Scientists Forum
October 26–30, 2020



PURDUE UNIVERSITY

CMS H→$\mu\mu$ analysis strategy

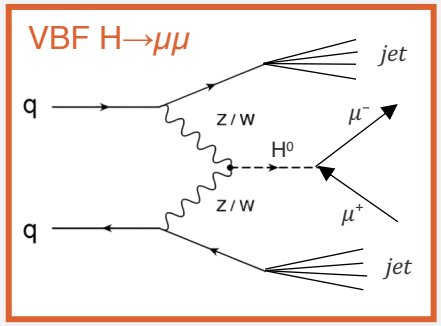
- H→$\mu\mu$ analysis provides **the first measurement of Higgs coupling to 2nd generation fermions**
- CMS recently reported **the first evidence** of H→$\mu\mu$ decay (full Run 2 data): arXiv:2009.04363
- Search performed in **four mutually exclusive categories**, corresponding to different Higgs production modes:



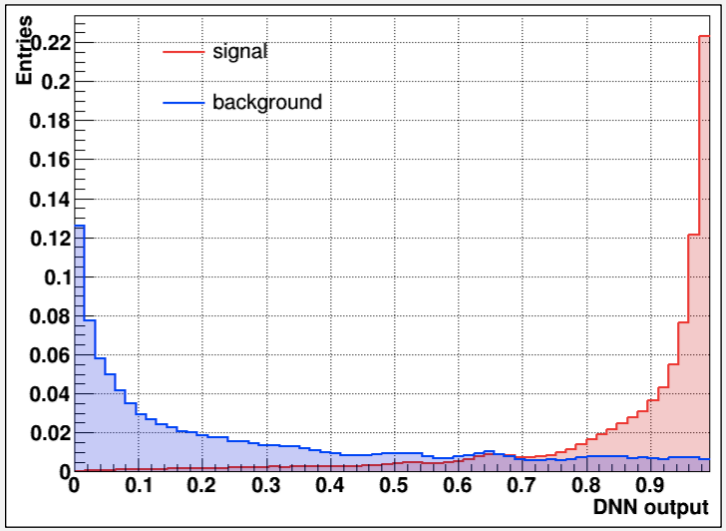
- **ttH**, **VH** and **ggH** categories use **data-driven strategy:**
  - background estimated from parametric fit of data in mass sidebands
  - BDT trained for data sub-categorization to enhance sensitivity
- For **VBF** category, **MC-template based strategy** was developed
  - ~20% improvement w.r.t. data-driven approach in VBF channel

MC-based strategy in VBF channel

- Signal and background predictions obtained from **MC simulation** of all SM processes
- Data is fit to "background" and "signal-plus-background" hypotheses
- Observable for the fit should be sensitive to $S/\sqrt{S+B}$
- Use **distinctive signature of VBF H→*μμ* events**:
  - 125 GeV peak in dimuon invariant mass
  - Two forward high-p_T jets



- Train a **Deep Neural Network (DNN)** to construct observable.
  Use 21 input features:
  - dimuon mass and experimental mass resolution
  - kinematic properties of jets and dijet pair
  - kinematic properties of dimuon-jet system
- Higher DNN output scores correspond to more signal-like events

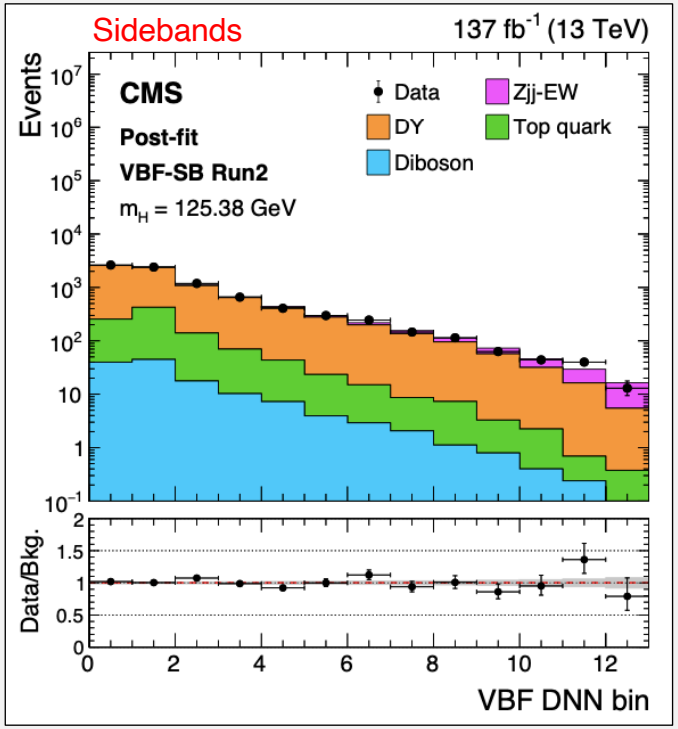


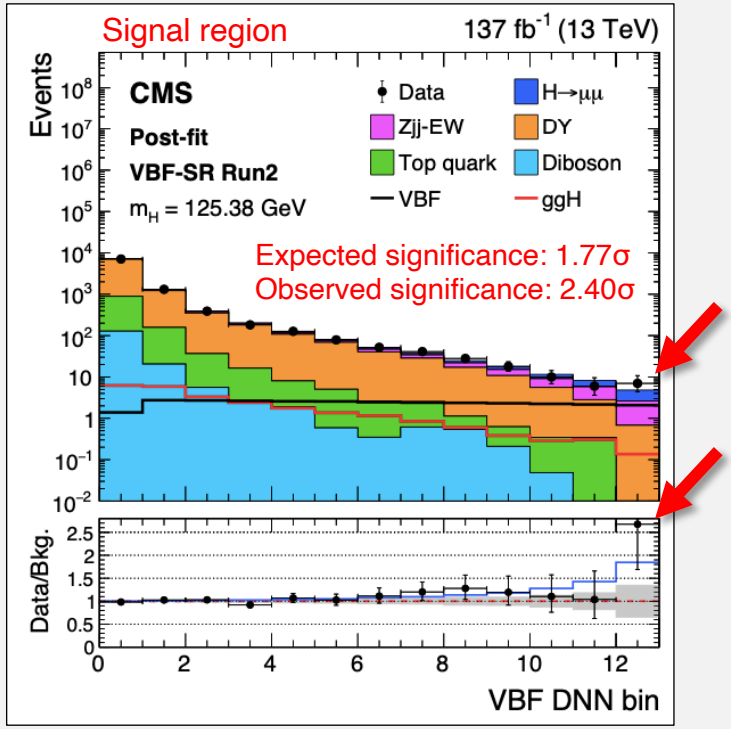
Systematic uncertainties

- Background estimated from MC simulation ⇒ uncertainties on MC prediction are crucial
- Considered systematics affect both **shape and normalization** of DNN output templates
- Systematics with largest impacts on signal significance:
  - **Theory uncertainties** (possible mismodelling in MC generators):
    - parton shower modeling (Pythia8 vs. Herwig)
    - QCD factorization and renormalization scale for background processes
  - **Experimental uncertainties**:
    - jet energy scale end resolution
    - uncertainty on background acceptance due to misreconstruction of forward jets (fake jets from pile-up and detector noise)
- Uncertainties are included in the fit as **log-normal nuisance parameters**
- Systematics from the same source are correlated across channels and mass regions

MC template fit

- **Simultaneous fit** to data performed in two $m_{\mu\mu}$ regions:
  - **Signal region**: **$115 < m_{\mu\mu} < 135$ GeV**
  - **Sidebands**: **$110 < m_{\mu\mu} < 115$ GeV** or **$135 < m_{\mu\mu} < 150$ GeV**
- Fit in sidebands used to constrain systematics
- Fit in signal region used to extract VBF H→$\mu\mu$ signal
- **Excess of data observed in the highest-score (highest signal purity) DNN bin**

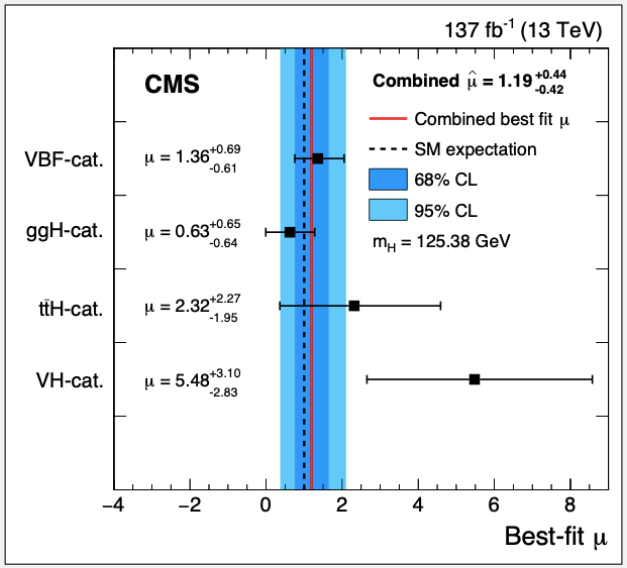




Results

- Combination of four exclusive channels (ggH, VBF, VH, ttH) lead to observation of **the first evidence of H→*μμ* decay**
- Observed signal strength in agreement with SM prediction
- Dedicated MC-based strategy for VBF channel has become crucial for achieving this result

Production category	Observed (expected) signif.	Observed (expected) UL on μ
VBF	2.40 (1.77)	2.57 (1.22)
ggH	0.99 (1.56)	1.77 (1.28)
$t\bar{t}H$	1.20 (0.54)	6.48 (4.20)
VH	2.02 (0.42)	10.8 (5.13)
Combined $\sqrt{s} = 13$ TeV	2.95 (2.46)	1.94 (0.82)
Combined $\sqrt{s} = 7,\ 8,\ 13$ TeV	2.98 (2.48)	1.93 (0.81)



Q: Why was it necessary to have a different strategy for VBF category?

A: data-driven approach performs worse in the VBF channel, due to lack of data in mass sidebands (needed for background estimation from parametric fit)

Q: Could other channels benefit from MC-template based approach?

A: ggH category suffers from large irreducible Drell-Yan background ⇒ large uncertainties on MC background. In ttH and VH, data-driven approach was chosen as simpler and more robust