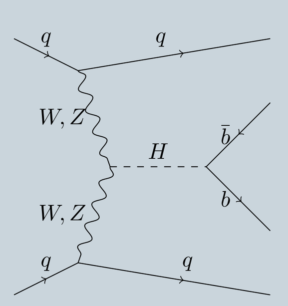
Search for the SM Higgs boson produced by vector boson fusion and decaying to bottom quarks

EPSHEP 2013 Stockholm, Sweden, 18-24 July, 2013

Sara Alderweireldt, S. Bansal, T. Cornelis, X. Janssen and N. Van Remortel – *Universiteit Antwerpen, Wilrijk, Belgium*
(SARA.ALDERWEIRELDT@CERN.CH)
P. Azzurri, J. Bernardini, D. Caiulo, P. Spagnolo and C. Vernieri – *INFN & Scuola Normale, Pisa, Italy*
K. Kousouris and S. De Visscher – *CERN, Geneva, Switzerland*
for the CMS collaboration

Introduction and strategy

Signal topology



- the signal is characterized by a 4-jet final state:
  - two **quark jets** with large pseudo-rapidity difference
  - two centrally produced **b-jets**
- no colour flow between the b-jets and the VBF jets

Background

- dominant background = **QCD multijet production** (after trigger ~ 4 orders of magnitude)
- **other backgrounds** ordered by importance:
  - Z+jets
  - $t\bar{t}$
  - single-top
  - W+jets

Search strategy

1. use a **topological trigger** on signal main properties:
   - jets with large $\Delta\eta$
   - two b-jets
   - …
2. use **multivariate methods** to exploit maximally the differences between signal and QCD
   - maintain orthogonality to the $m_{b\bar{b}}$ distribution
3. **fit the $m_{b\bar{b}}$ distribution** and look for a resonance structure on top of the background

Cross section and branching ratio

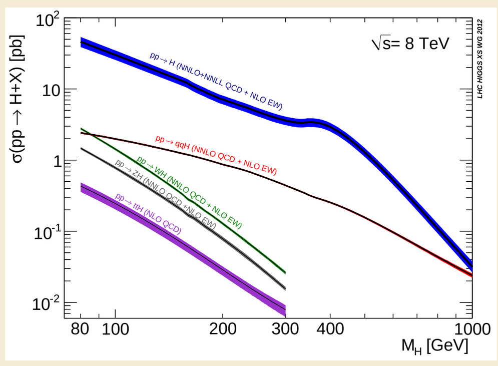


Fig.1: Cross section $\sigma(pp \to HX)$.

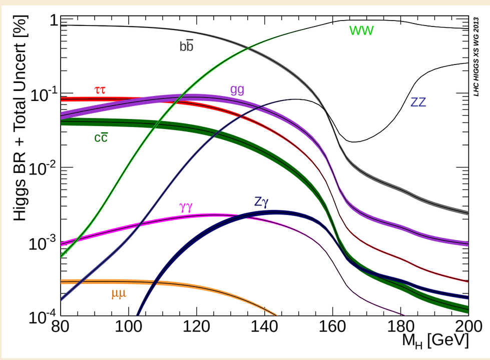


Fig.2: Branching ratio.

- VBF mode has significantly **larger cross section** than VH or ttH modes which have been studied before.
- Contribution to the $b\bar{b}$ combination needs to be studied.

Preselection

- $p_\perp > 85, 70, 60, 40$ GeV
- $m_{q\bar{q}} > 300$ GeV
- $\Delta\eta_{q\bar{q}} > 2.5$
- $\Delta\phi_{b\bar{b}} < 2.0$

Trigger (1)

L1 & HLT trigger paths

L1:

- **triple-jet**: $p_\perp$ cuts for three jets + η requirements (at least two jets central)

HLT:

- both CALO (2) & PF (4) **quad-jet** paths (overlap ~50%, rate ~10 Hz)
- cuts:
  - **$p_\perp$ cuts** on the four hardest jets
  - further kinematic cuts on **$\Delta\eta_{qq}$ and m_{qq}**
  - **b-tagging** (CALO paths: track-counting / PF paths: combined secondary vertex)
  - PF path extra: keep the cuts on CALO jets, with lowered thresholds
- finally: OR of all the paths

Efficiency w.r.t. reference trigger

- measured with respect to a single-jet trigger with online $p_\perp$ threshold of 80 GeV (checked to be unbiased in our phase space)
- small efficiency at the plateau, due to heavy-flavour content of the signal trigger (no b-tag requirement at preselection level)
- the trigger efficiency measurement is **statistics limited**

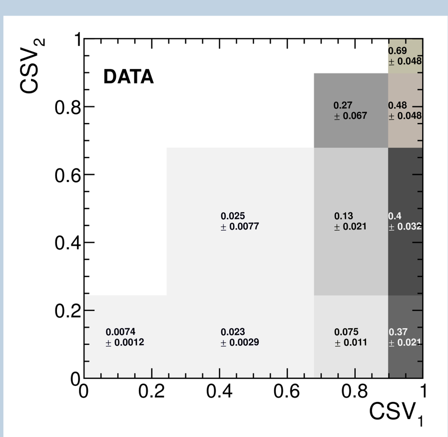


Fig.3: 2D efficiency map for data.

- display efficiency in **2D map**, binned in CSV b-tag of the two most b-tagged jets
- 2D maps used to compute a **correction factor** on simulated trigger efficiencies
- average correction factors: ~85%

Signal vs. background discrimination (2)

The use of **multivariate techniques** makes it possible to achieve maximum separation between signal and background – taking into account not only the **distributions of the discriminating variables** , but also their **correlations**.

Since the **search strategy relies on a background fit of the $m_{b\bar{b}}$ spectrum**, it is critical that the correlation between the multivariate discriminant and $m_{b\bar{b}}$ is minimal: a cut on the former should not affect significantly the shape of the latter. Careful examination of possible variables leads to the list below:

$\lvert\Delta\eta_{q\bar{q}}\rvert$	$\lvert\Delta\eta\rvert$ between the two *least* b-tagged jets	$\text{b-tag}_{0,1}$	b-tag value of the two *most* b-tagged jets
$\lvert\Delta\eta\rvert\text{max} - \lvert\Delta\eta_{q\bar{q}}\rvert$	difference between maximal and least b-tagged $\Delta\eta$ (ideally 0)	$\text{qgl}_{2,3}$	quark-gluon discriminator (JME) for the two *least* b-tagged jets
$m_{q\bar{q}}$	invariant mass of the $q\bar{q}$ pair	η_3	η of the third b-tagged jet
$\eta^{\text{boost}}_{q\bar{q}}$	average η of the $q\bar{q}$ pair	$\lvert\cos\theta\rvert, \lvert\cos\alpha\rvert$	polar angles of $p_{q1} \times p_{q2}$ and $p_{q1} + p_{q2}$ in the $b\bar{b}$ CM frame
$H_{T\text{soft}}$	scalar $p_\perp$ sum of the track-jets with $p_\perp > 1$ GeV formed by PV tracks that do not belong to jets		

The ANN in data & definition of categories

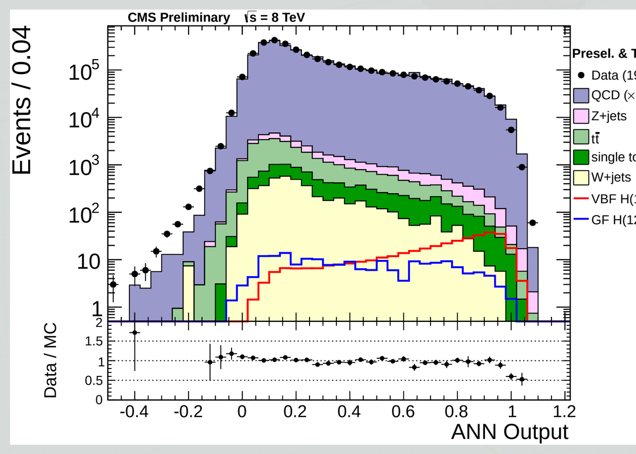


Fig.4: ANN output distribution, after offline event preselection.

The **ANN response for signal** peaks towards unity, but has a considerable tail towards lower values. In order to utilize all ANN response information efficiently, **the search is conducted simultaneously in four bins of the ANN response** (categories). The number of categories and the boundaries between them are chosen such that the sensitivity is maximized. The category with the lowest ANN response (< 0.52) is not used in the signal search.

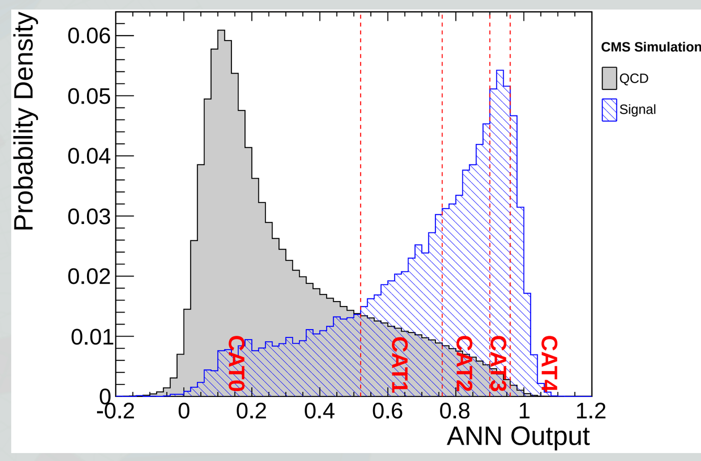


Fig.5: Definition of categories in the ANN output distribution.

Results (3)

Fit to the data

The background model consists of three components (null hypothesis): the contributions from the Z and the top -which have *normalization and template fixed*- and the contribution from QCD -for which a 5th order Bernstein polynomial is used and which has *normalization and template floating.*

$$f(x) = N_Z \cdot Z(x) + N_T \cdot T(x) + N_{QCD} \cdot B^{(5)}(x)$$

with: $B^{(5)}(x) = \sum_{\nu=0}^{5} \beta_\nu \, b_{5,\nu}(x)$ and $b_{5,\nu}(x) = \binom{5}{\nu} \cdot x^\nu \cdot (1-x)^{5-\nu}$

When considering the signal hypothesis, there is an additional signal template with *free normalization.*

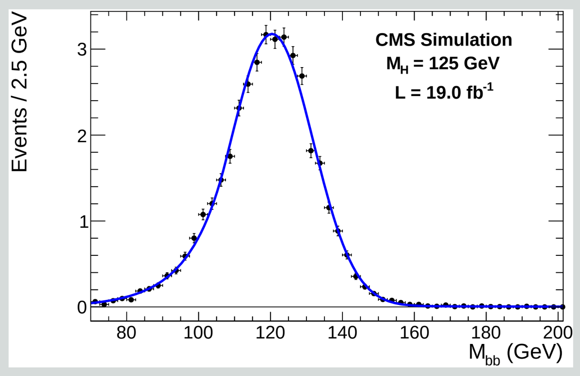


Fig.6: Fit to the $m_{b\bar{b}}$ distribution.

The fit is performed in the **70 GeV $< m_{b\bar{b}} <$ 250 GeV mass range** and for **each ANN category** separately:

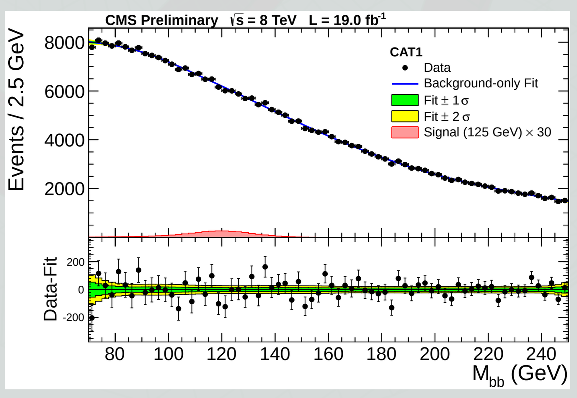


Fig.7: Fit to the $m_{b\bar{b}}$ spectrum, for category 1 ($0.52 <$ ANN < 0.76).

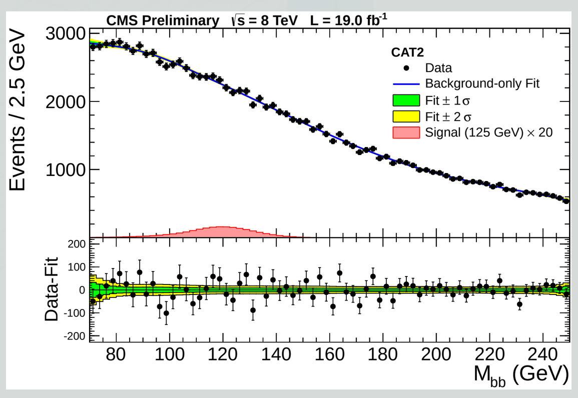


Fig.8: Fit to the $m_{b\bar{b}}$ spectrum, for category 2 ($0.76 <$ ANN < 0.90).

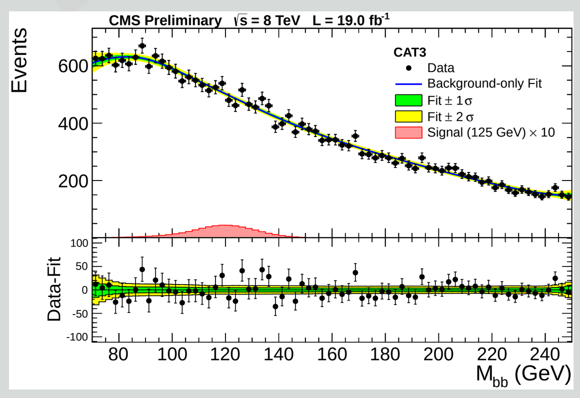


Fig.9: Fit to the $m_{b\bar{b}}$ spectrum, for category 3 ($0.90 <$ ANN < 0.96).

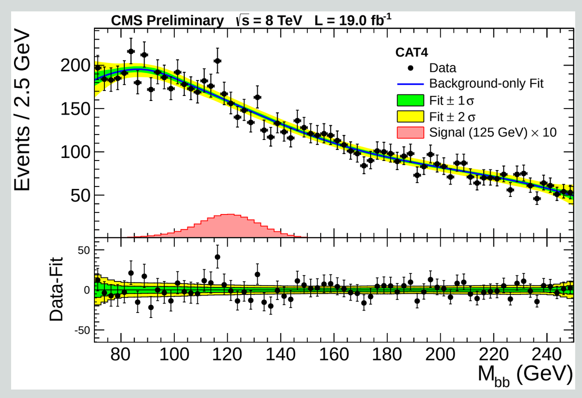


Fig.10: Fit to the $m_{b\bar{b}}$ spectrum, for category 4 (ANN > 0.96).

95% Asymptotic CL limits were derived for five mass points from 115 to 135 GeV:

- combined result including **categories 1-2-3-4** (category 0 has no improvement on the limit and is only used as a template for toys)
- a small signal excess is observed
- combined result at $M_H = 125$ GeV:
  - excluding $3.64 \times$ SM
  - expected $3.02 \times$ SM
  - fitted signal strength: $\mu = \sigma/\sigma_{SM} = 0.7 \pm 1.4$

M_H	exp. limit	obs. limit	exp. significance	obs. significance
115	2.43	2.40	0.85	0.00
120	2.74	3.15	0.75	0.36
125	3.02	3.64	0.68	0.49
130	3.36	4.03	0.62	0.48
135	4.14	5.18	0.50	0.47

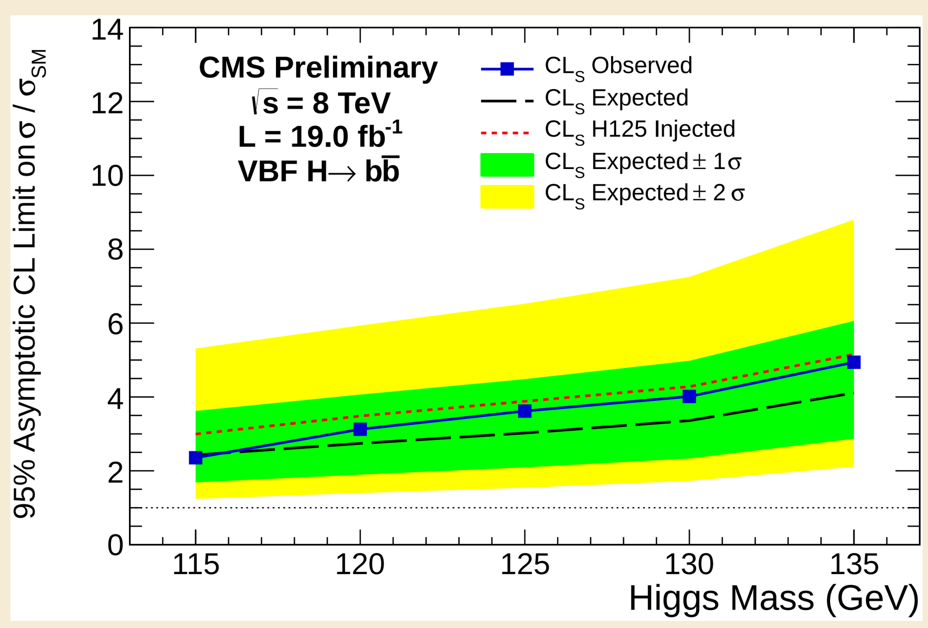


Fig.11: 95% Asymptotic CL limits on σ/σ_{SM}.

CMS PAS HIG-13-011: Search for the standard model Higgs boson produced in vector boson fusion, and decaying to bottom quarks. The CMS Collaboration.