Approaches to H→γγ and H→ZZ differential measurement

LHC Higgs XS WG Fiducial meeting - 24/06/2015

Nicolas Chanon,
For the CMS Collaboration

Introduction and outline

Scope of the talk:

- Present CMS point of view on measuring fiducial and differential cross sections for the Higgs boson
- Will focus on $H \rightarrow \gamma\gamma$ and $H \rightarrow ZZ$ final states
- **Disclaimer**: we will not present final results, which are not yet public (coming soon)

We will discuss the following topics:

- Definition of fiducial phase-space
- Treatment of model dependence
- Unfolding procedure

Introduction

Scope of the analysis:

- Measure **Higgs boson fiducial and differential cross-sections in $H\rightarrow\gamma\gamma$ and $H\rightarrow ZZ$ channel**
- Compare with latest SM theory predictions
- Hints for new physics ?

$H\rightarrow\gamma\gamma$:

- **$H\rightarrow\gamma\gamma$** channel has a small branching ratio (~0.2%) but the analysis has **high selection efficiency (~40-45%)**
- About 300 reconstructed signal events are expected after selection (8 TeV Run I)
- But large background from the diphoton continuum and fakes

$H\rightarrow ZZ$:

- **$H\rightarrow ZZ\rightarrow 4l$** channel has a very small branching ratio (~0.01%) but the signature is very clean
- About 15 reconstructed signal events are expected after selection (8 TeV Run I)
- Small background from the ZZ continuum and fakes

Definition of fiducial phase-space: guidelines

Designing fiducial phase-space selection:

- **As close as possible to the reconstructed level selection**
  - Unfolding should only impact minimally the distributions: minimize extrapolation to larger phase-space
  - Translating reco-level isolation to generator-level is not obvious (resolution at low energy, pile-up, thresholds...)

- **Should minimize model dependence**
  - Try to have the acceptance similar for different “reasonable” signal models
  - This improves the robustness of the unfolding

- Definition should be easily implemented in different MC simulation

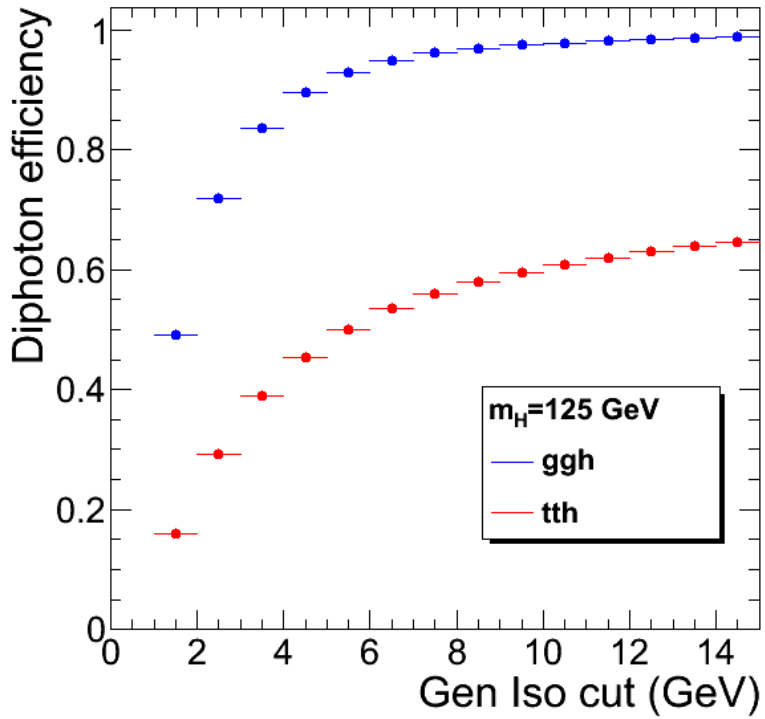
Definition of fiducial phase-space: H→γγ

Suggestions for the fiducial phase-space:

H→γγ

- Two photons with pT1/M>1/3 and pT2/M>1/4 within |η|<2.5 (close to reconstruction level selection)
- Isolation: for each photon, sum all stable generator level particles inside a 0.4 cone. Sum ET < 10 GeV

Loose cut of Sum ET < 10 GeV chosen to stand on the "plateau" of the isolation efficiency, for ggh and the busy tth events



Definition of fiducial phase-space: H→ZZ

Suggestions for the fiducial phase-space:

Lepton kinematics and isolation	
leading lepton p_T	$p_T > 20$ GeV
next-to-leading lepton p_T	$p_T > 10$ GeV
additional electrons (muons) p_T	$p_T > 7(5)$ GeV
pseudorapidity of electrons (muons)	$\|\eta\| < 2.5(2.4)$
p_T sum of all stable particles within $\Delta R < 0.4$ from lepton	less than $0.4 \cdot p_T$

Event topology	
existence of at least two SFOS lepton pairs, where leptons satisfy criteria above	
inv. mass of the Z_1 candidate	$40\,\text{GeV} < m(Z_1) < 120\,\text{GeV}$
inv. mass of the Z_2 candidate	$12\,\text{GeV} < m(Z_2) < 120\,\text{GeV}$
distance between selected four leptons	$\Delta R(\ell_i \ell_j) > 0.02$ for any $i \neq j$
inv. mass of any opposite sign lepton pair	$m(\ell^+\ell'^-) > 4\,\text{GeV}$
inv. mass of the selected four leptons	$105\,\text{GeV} < m_{4\ell} < 140\,\text{GeV}$
the selected four leptons must originate from the $H \to 4\ell$ decay	

close to reconstructed level selection

- **Unfold to born leptons** before final state radiation (can move to dressed leptons when more data will be available)

- **Isolation**:
  - do not include FSR photons and neutrinos in the isolation sum
  - Non-desirable effect of up to a factor 2 increase of the out of acceptance events if included

Some words about analysis methodology

Measuring differential distributions needs changes in the analysis strategy:

- For each observable, fit the mass to **extract signal strength μ in all bins simultaneously** (profile nuisance parameters across bins)

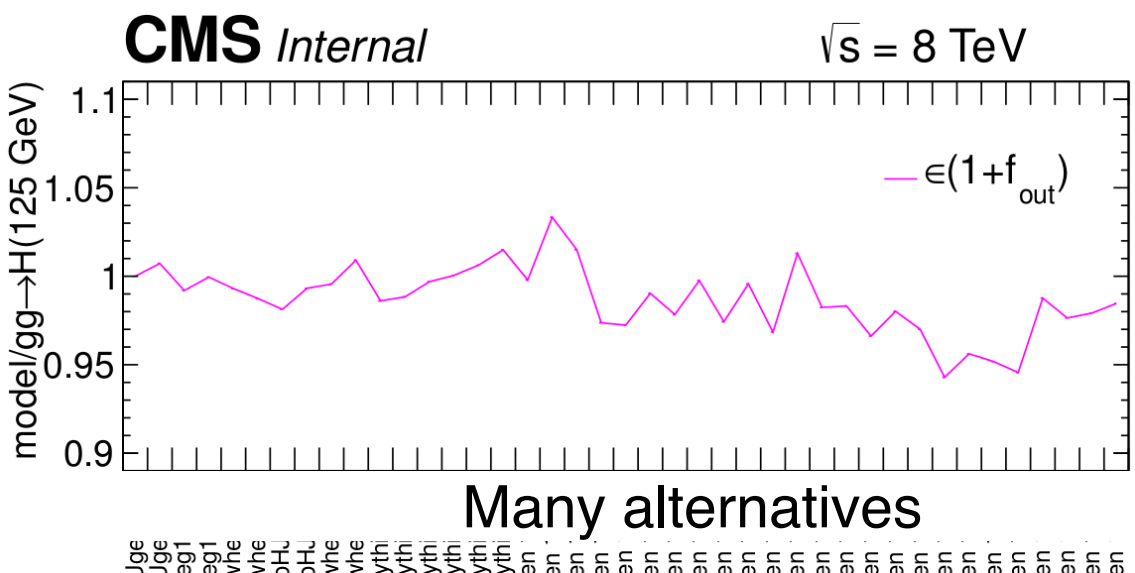
- **Analysis sensitivity should be evaluated using variables uncorrelated with the measured observable** (should not bias kinematics)
    - prefer mass-shape analysis and cut-based selection to kinematic BDT

- **Selection** needs to be as **model independent** as possible:
    - efficiencies should be independent of pT and eta
    - categories independent of pT
    - do not give privilege to higher pT region (we are also interested at low pT e.g. bottom/top effects in the ggh loop)

Treatment of the model dependence:

Suggested check (1): model dependence of the acceptance

- Check the impact of using a **different signal model** on the **efficiency** and the **fraction of reconstructed events that are outside of the fiducial volume**
- Many alternative samples are re-used from the Higgs spin measurement **(arxiv:1411.3441)**



example of H→ZZ:
7% effect (envelope)

Treatment of the model dependence:

Suggested check (2): model dependence of the unfolding

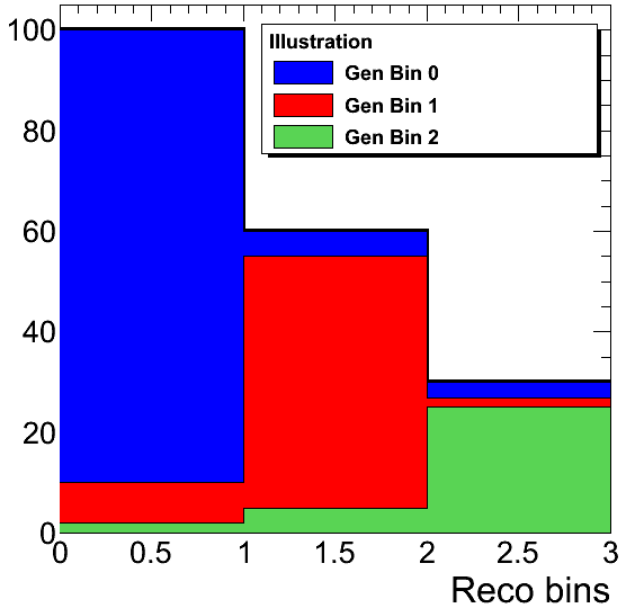
- Perform the **unfolding on a different signal model**
- Can for example **change the fraction of VBF+VH+tth** by x2 and x1/2
- Example of H→γγ: results in less than 5% variation of the measured cross section in, and up to 10% of the statistical uncertainty in each bin
- Can also vary the fraction of production modes or anomalous HVV couplings within the existing experimental constraints. Example of H→4l: model dependence is less than 1%

Suggested treatment of the model dependence:

- **Do not include the model dependence into the systematic uncertainties** (this is a theory uncertainty and after all we are primarily aiming at measuring the SM Higgs boson)
- **Quote** the impact of the model dependence on the measurement

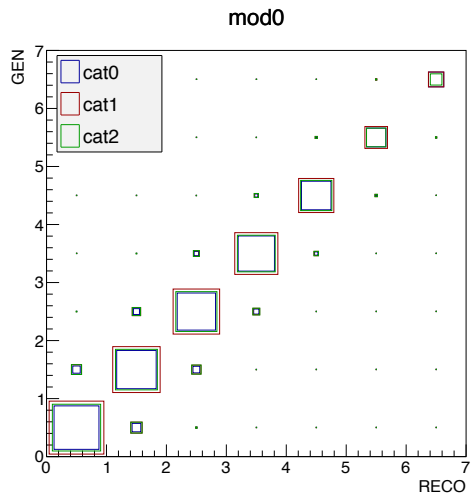
Unfolding procedure: response matrix

See also A. Marini's talk on Wed.



Why unfolding ?

- Unfolding is needed to **undo detector effects**
- There are bin by bin **migrations** between generator level and reconstructed level
- This needs to be corrected: use a **response matrix** to propagate the statistical uncertainty in each bin
- Selected events **falling out of acceptance** at generator level are collected in a special bin



The proposed method:

- **Fold in the response matrix** into the likelihood (perform measurement and unfolding at the same time)
- **Each cell** of the response matrix contains **normalization** (efficiency) and **signal model** for a given set of (gen, reco) **bin and analysis category**

Unfolding procedure (cont.)

Function to minimize

Response matrix

Gen level input distribution

$$\mathcal{F}(\mu_i) = -2\log\mathcal{L}\,K_{ij}\cdot\mu_i\,N_{gen,i}|N_{reco,j})$$

L is the usual **likelihood** function of B, S+B and the nuisance parameters

Signal strength to be measured for each bin of **gen level** input distribution

Reco level signal yields in each bins and categories

- **Same procedure is already used** e.g. for unfolding from μ in untagged and dijet-tagged categories to μ_{GGH}, μ_{VBF}
- This is **similar to response matrix inversion**, but:
- Usual method of estimating uncertainties (covariance matrix applied after μ measurement) is approximate (in particular for low statistics categories),
- Advantage: use the full likelihood including nuisance parameters

Unfolding procedure: comparison with other methods

No regularization applied:

- For bins with **sizeable off-diagonal cells**, bin migration implies an **increase of the statistical uncertainty**
- Regularization is usually used to **stabilize** the unfolding against **statistical fluctuations**
- Regularization **biases the best fit value** to reduce the total uncertainty: we believe it is not needed
- Regularization could be applied later if the covariance matrix is provided

Comparison with bin by bin correction:

- **The best fit value** is not always the same (migration can be important)
- Relative to bin-by-bin, the **statistical uncertainty is increased** in the bins where the off-diagonal elements are sizeable: up to 30% for Njets (due to jet energy resolution)
- Bin-by-bin method does not deal correctly with gen-reco bin migrations

Conclusions

- Proposed **fiducial phase-spaces** for **H→γγ and H→ZZ**

- Importance of **model independence** in designing the fiducial phase-space and the selection

- Proposed a method for **unfolding** the data simultaneously with the fit

Back-up slides