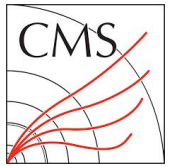
Parametric Models in Combine

Hands-on exercise

Kyle Cormier (UZH), Aliya Nigamova (UHH) and Nicholas Wardle (IC)

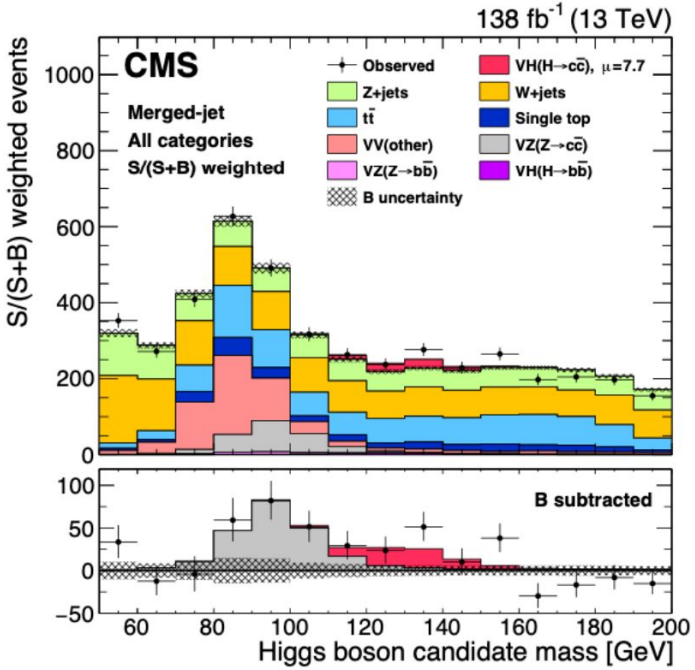


Introduction

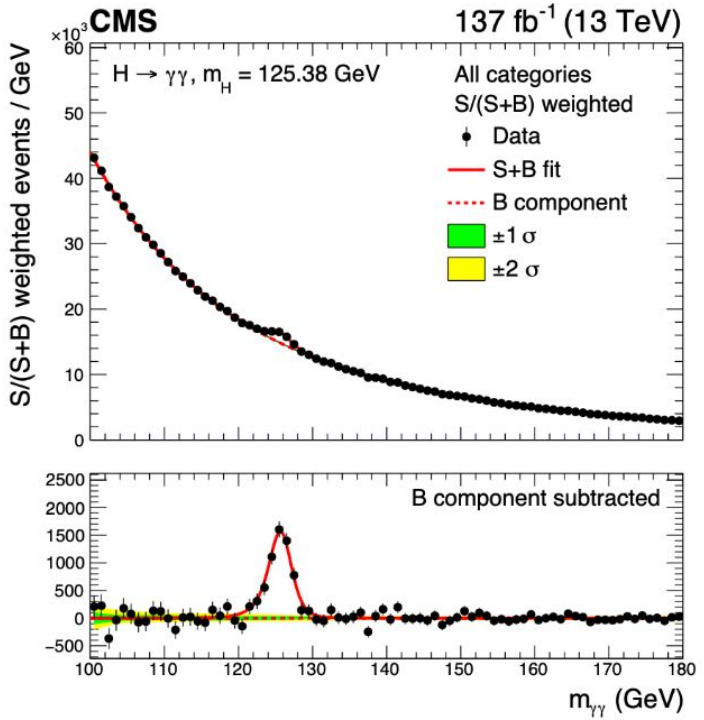
- Use this Link for the “Parametric Models” exercises (under the Tutorials pages)

- Broadly, there are two types of fit we perform in combine
  - **Templated-based**: no simple analytic form to describe signal/background
    - Use templates (histograms) for signal/background and shape systematic uncertainties
    - Example covered earlier
  - **Parametric-based**: signal and background models can be described by analytic functions
    - E.g. Gaussian signal on top of falling exponential background
    - This exercise will be based on a parametric fitting analysis

Introduction



Template



Parametric

Getting started

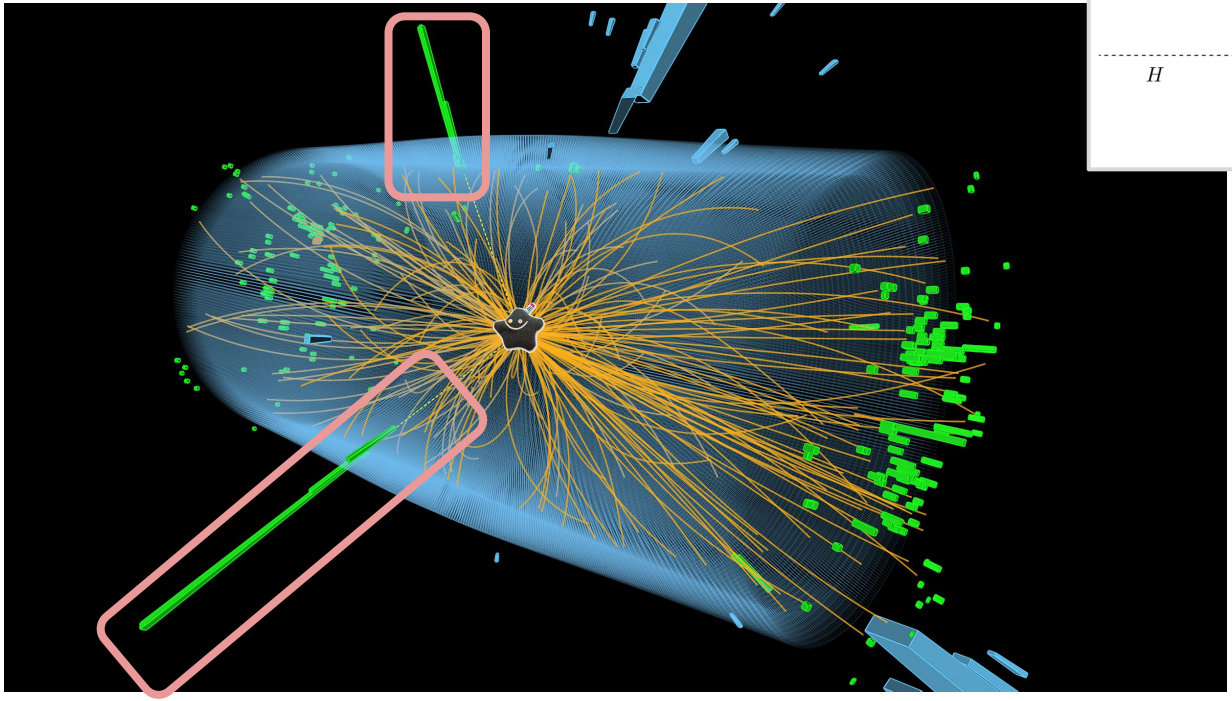
- By now you should have a working setup of combine v9.2.0
  - If not then follow the Combine Setup instructions in the parametric exercise docs
- The inputs and scripts required to run the tutorial are contained in the area
  - `data/tutorials/parametric_exercise`
- The exercise is split into six parts, which cover: *Parametric model building, Simple fits (MLE), Systematic uncertainties, Toy generation and bias studies (extra: Discrete profiling & multi signal models)*
- Throughout the tutorial there are a number of questions/tasks to complete.
  - These are marked by the **bullet points**. If you are unsure of the answer then we are happy to discuss!
- All of the code is available in python (.py) scripts
  - Commented out! When you understand the code you can uncomment (block by block) and run with e.g.:

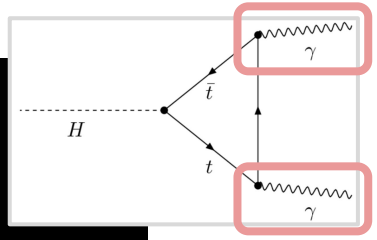
```
python3 construct_models_part1.py
```

Analysis overview

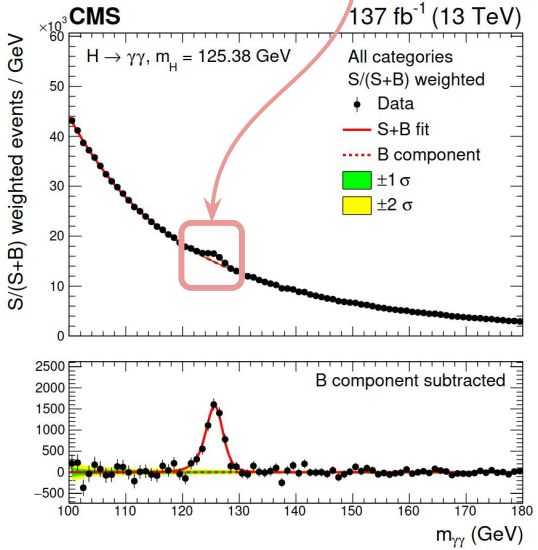
$$m_{\gamma\gamma} = \sqrt{2E_1E_2(1-\cos\alpha)}$$

- Based on one of the most famous parametric fitting analyses: H -> $\gamma\gamma$





Higgs boson appears as bump above background!



Part 1: datacard_part1.txt

Can use * to tell combine to calculate these automatically

```
# Datacard example for combine tutorial 2023 (part 1)
------------------------------------------
imax 1
jmax 1
kmax *
------------------------------------------

shapes      ggH          Tag0      workspace_sig.root      workspace_sig:model_ggH_Tag0
shapes      bkg_mass     Tag0      workspace_bkg.root      workspace_bkg:model_bkg_Tag0
shapes      data_obs     Tag0      workspace_bkg.root      workspace_bkg:data_Tag0

------------------------------------------
bin             Tag0
observation     -1
------------------------------------------
bin             Tag0          Tag0
process         ggH           bkg_mass
process         0             1
rate            181.01        1.0
------------------------------------------
```

Number of channels/categories

Number of backgrounds

Number of nuisance parameters

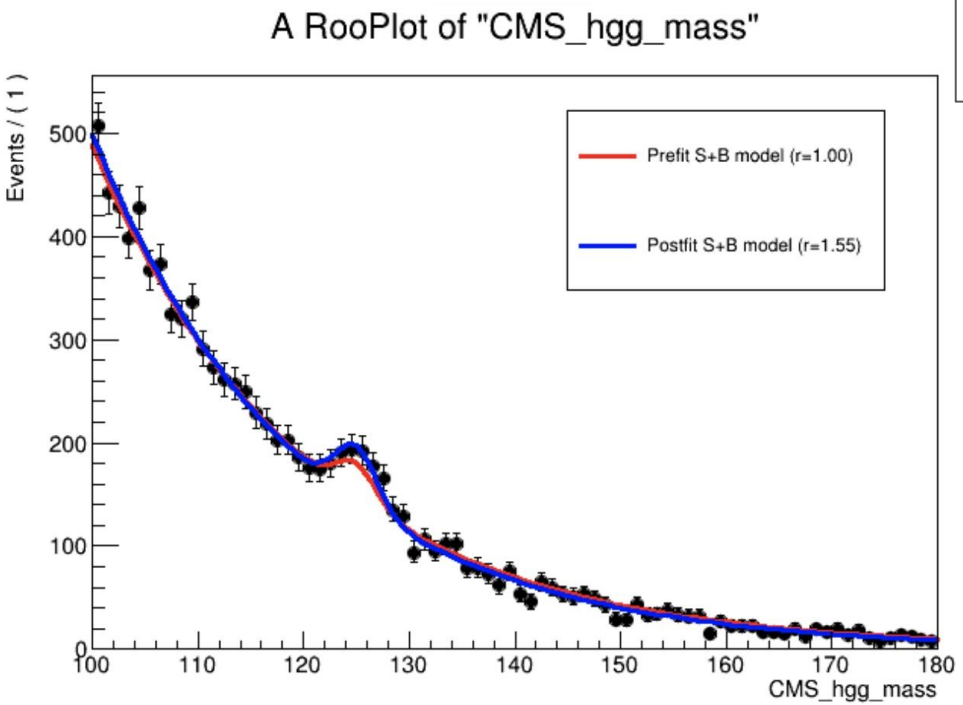
Maps processes to parametric models

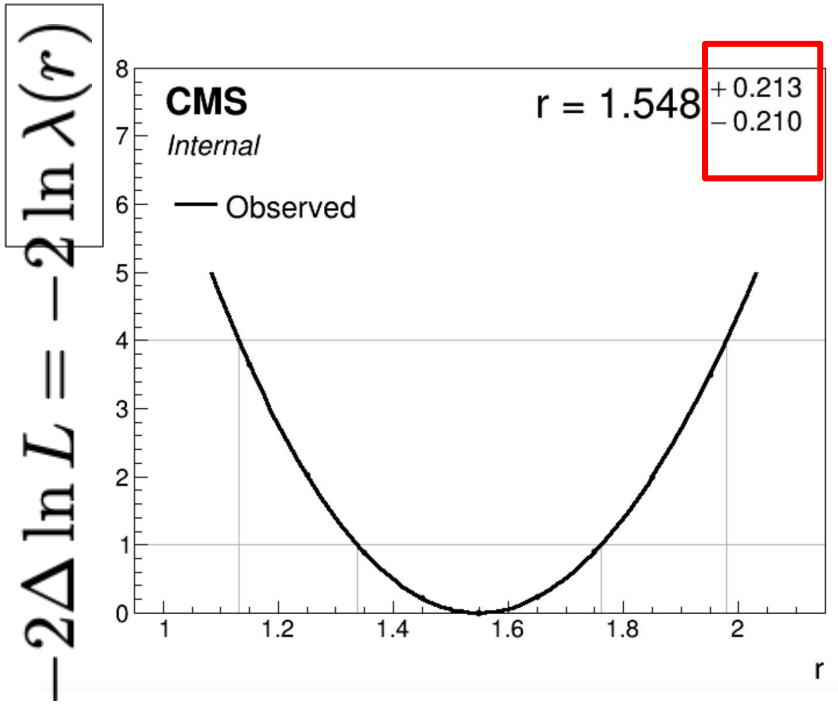
List of categories/channels
Observation is number of observed events: -1 means take from data

Processes in each channel with expected yield (rate)

We use 0 (or negative integer) for signal process, and positive integer for background

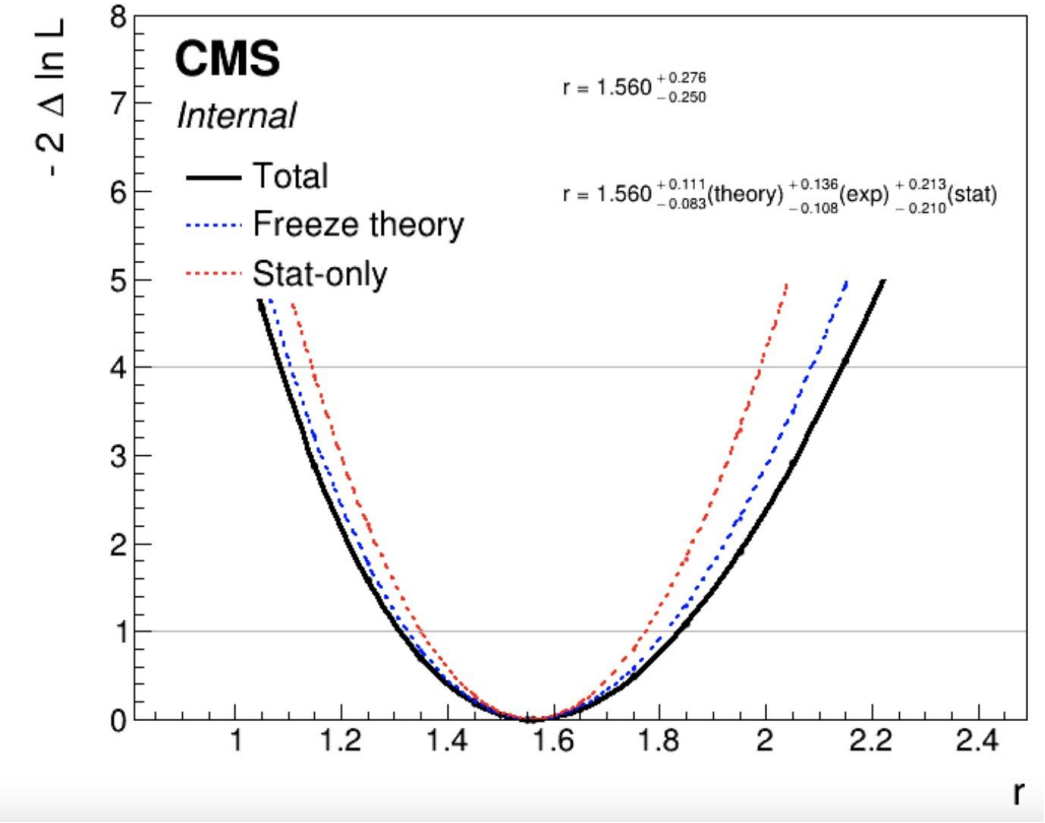
Part 2: Simple Fits





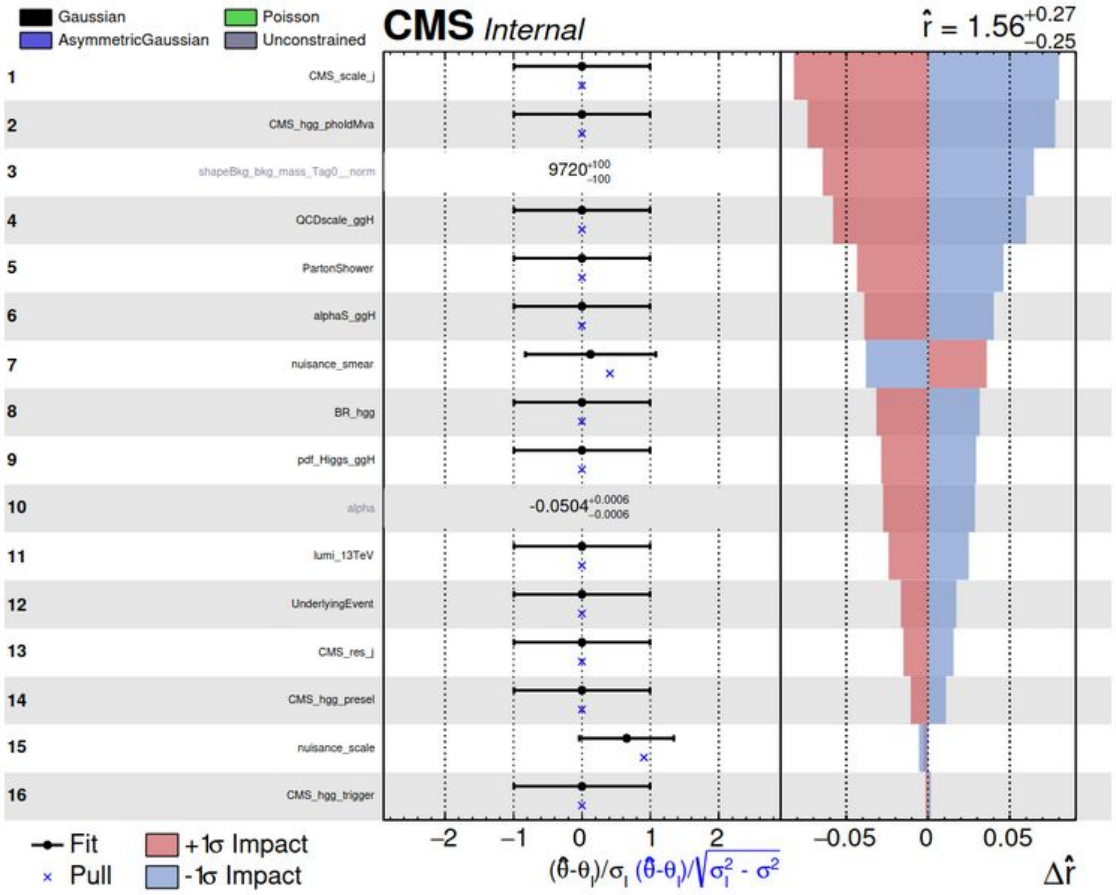
Remember Glen's lectures:
"**Graphical method" for uncertainty**

Part 3: Uncertainties



Grouping nuisance parameters to freeze gives us estimate of contributions to total uncertainty.

Part 3: Uncertainties



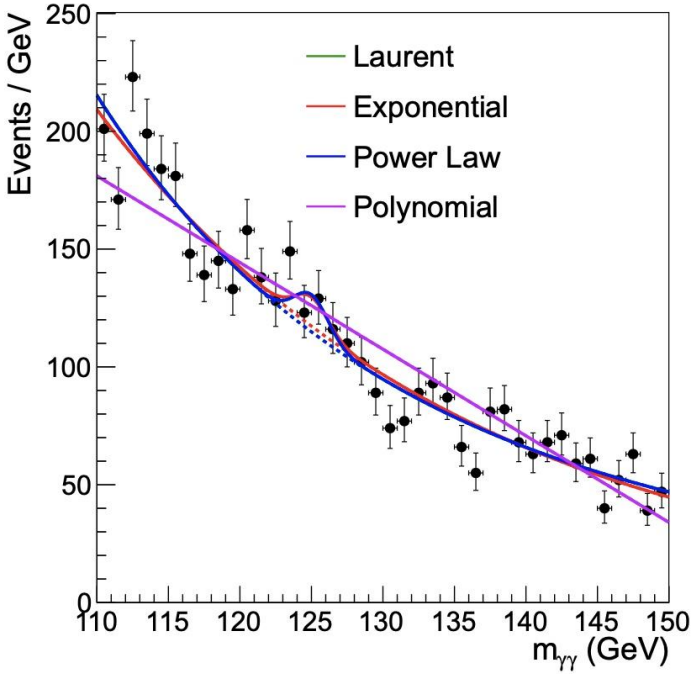
Finer detail breakdown via impacts

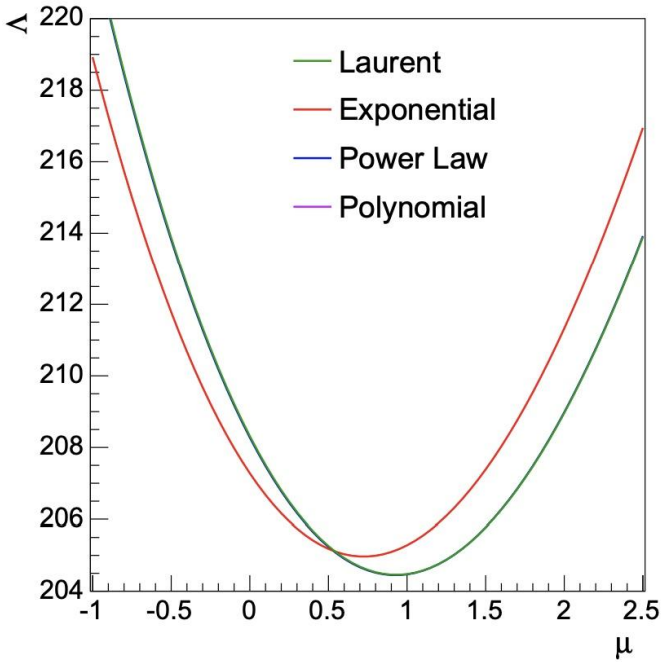
$$\Delta\hat{r}^{\pm} = \hat{r}\left(\hat{\theta} \pm \Delta\theta\right) - \hat{r}$$

Post-fit nuisance parameters (and uncertainties) and **pulls**

Part 4: Bias studies

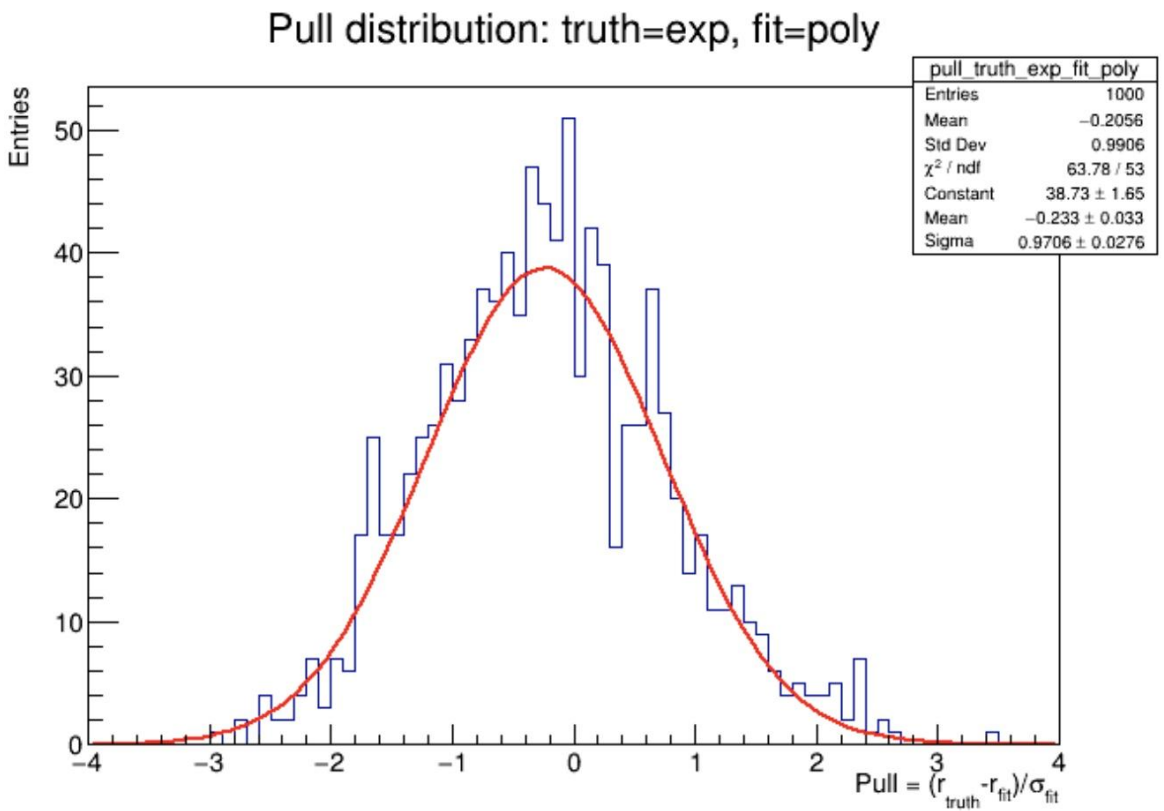
What if we don't *know* the right background function?





JINST 10 PO4015

Part 4: Bias studies



Remember Bias definition of estimator

$$b = E[\hat{\theta}] - \theta$$

Here we can define the bias of the estimator based on the chosen model assuming a different model for the background!

Summary

Combine is (one of) our tools for performing statistics calculations BUT

→ Statistics is more than just running some code!

→ Understanding what you are calculating is crucial to getting good results (see Glen's lectures, PDG statistics review, statistics for physicists textbooks…)

→ Sometimes things don't work out well

→ Fit instabilities, incorrect model assumptions, Bugs!

→ Always make diagnostic checks for your results

Combine let's you make these checks/tests too, read more about what you can do with combine in the docs
https://cms-analysis.github.io/HiggsAnalysis-CombinedLimit

Summary

Combine is (one of) our tools for performing statistics calculations BUT

→ Statistics is more than just running some code!

→ Understanding what you are calculating is crucial to getting good results (see Glen's lectures, PDG statistics review, statistics for physicists textbooks…)

→ Sometimes things don't work out well

→ Fit instabilities, incorrect model assumptions, Bugs!

→ Always make diagnostic checks for your results

Combine let's you make these checks/tests too, read more about what you can do with combine in the docs
https://cms-analysis.github.io/HiggsAnalysis-CombinedLimit

Questions?