EFT fitting exercise

A. Gilbert for CMS

LHC EFT WG Area 4 Meeting | 14 February 2022

Introduction

- First steps towards fitting exercise: https://gitlab.cern.ch/agilbert/eft-exercise
  - Many corners cut, simplifications made
  - Focus on difficulties and areas that should be improved
- Reminder: twiki page to document conventions for future cross-experiment combination
  - Currently not fully in sync with recommendations - but no major showstoppers anticipated

From the November LHC EFT WG Meeting

Sketch of the exercise 1/2

1. Agree on models and conventions (see twiki)
   - SMEFT basis, input parameters, flavour symmetry, operators
   - Parametrization: linear/quadratic?, LO/NLO?, parameter settings...
   - Uncertainties: what to include, how to correlate
   - Some details later
2. Identify input analyses
   - From EW, Higgs, and top sector
   - With reasonable physics case for combination
   - For re-interpretation need: cross-sections, covariance, and Rivet routine (EFT interpretation in paper *not* a requirement)
   - Collected list of candidate analyses – possible choices discussed later
3. Derive particle level parametrizations with agreed models and conventions
   - Typically MadGraph+SMEFT(sim|atNLO) $\rightarrow$ Pythia8 $\rightarrow$ Rivet
   - ATLAS and CMS could could run each others Rivet routines for some processes, as cross check
   - Also cross-check with theory publications

3 / 12

From the November LHC EFT WG Meeting

Sketch of the exercise 2/2

4. Create models in RooFit
   - Multivariate Gaussian, potentially full model in some cases
   - Nuisance parameters for systematics that should be correlated
5. Perform first combined fits
   - Linear and quadratic, 1D and multidimensional
   - Start with analyses for which we first finish steps 1-4 (possibly start with few EW measurements)
6. Study correlation of uncertainties
   - For theory uncertainties, some test cases for experimental uncertainties
   - Publically available information not enough to correlate all relevant parameters
7. More in-depth studies
   - More involved schemes to evaluate theory uncertainties and validity
   - Acceptance corrections (e.g. for decays in STXS analysis)
   - Study of overlap in event selections

4 / 12

Steps

- **Assumption**: even if analysis paper makes EFT interpretation, we will likely need to re-derive parameterisation with consistent EFT choices and full set of operators
- EFT choices:
  - Warsaw basis as implemented in **SMEFTsim 3.0.2**
  - **topU3l** flavour symmetry
  - Include both linear (c_i) and quadratic ($c_i c_j$) terms in the parameterisation
  - Consider, a priori, all 129 CP-even operators
- Choose input analyses, one from each of Higgs, top and EW
  - Not considering EFT sensitivity at this point
  - Technical requirements:
    - Diff/fiducial measurement with covariance matrix & RIVET routine available
  - Extract fitted values, uncertainties and covariance matrix directly from HEPData entries
- Construct EFT parameterisation using EFT2Obs
- Build RooFit model with multi-variate Gaussian pdf from cov. matrix, making cross section parameters functions of the EFT coefficients
  - Fit & NLL scans as we would in a normal analysis

HIG-19-015: H→γγ

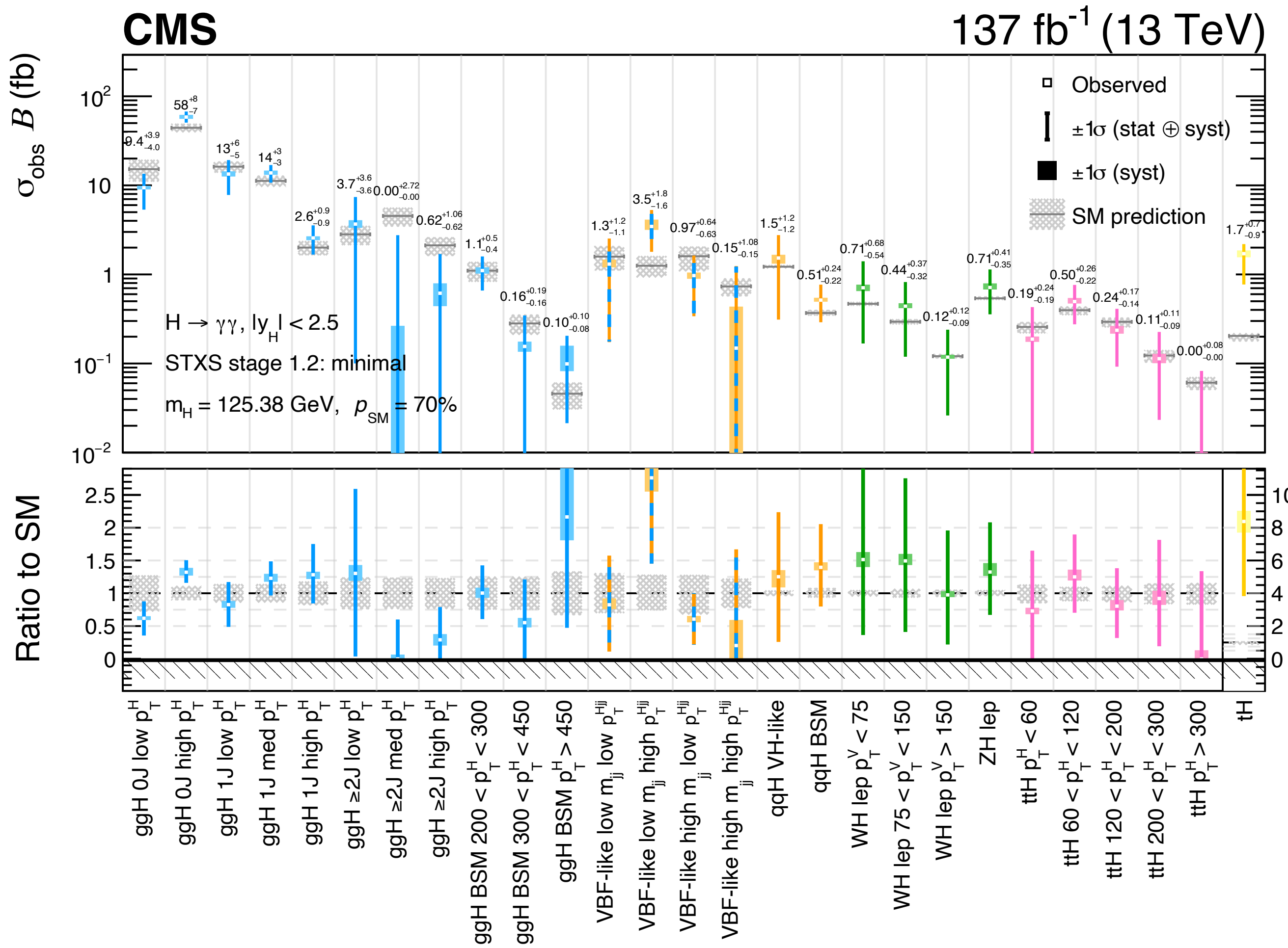
- EFT reinterpretations of the STXS measurements already done
  - Common RIVET routine for bot
- So far only considering qqH bins (VBF + V(had)H) here, and no decay info
- NB: work here not yet in sync with CMS Higgs group (who are more advanced in several aspects)

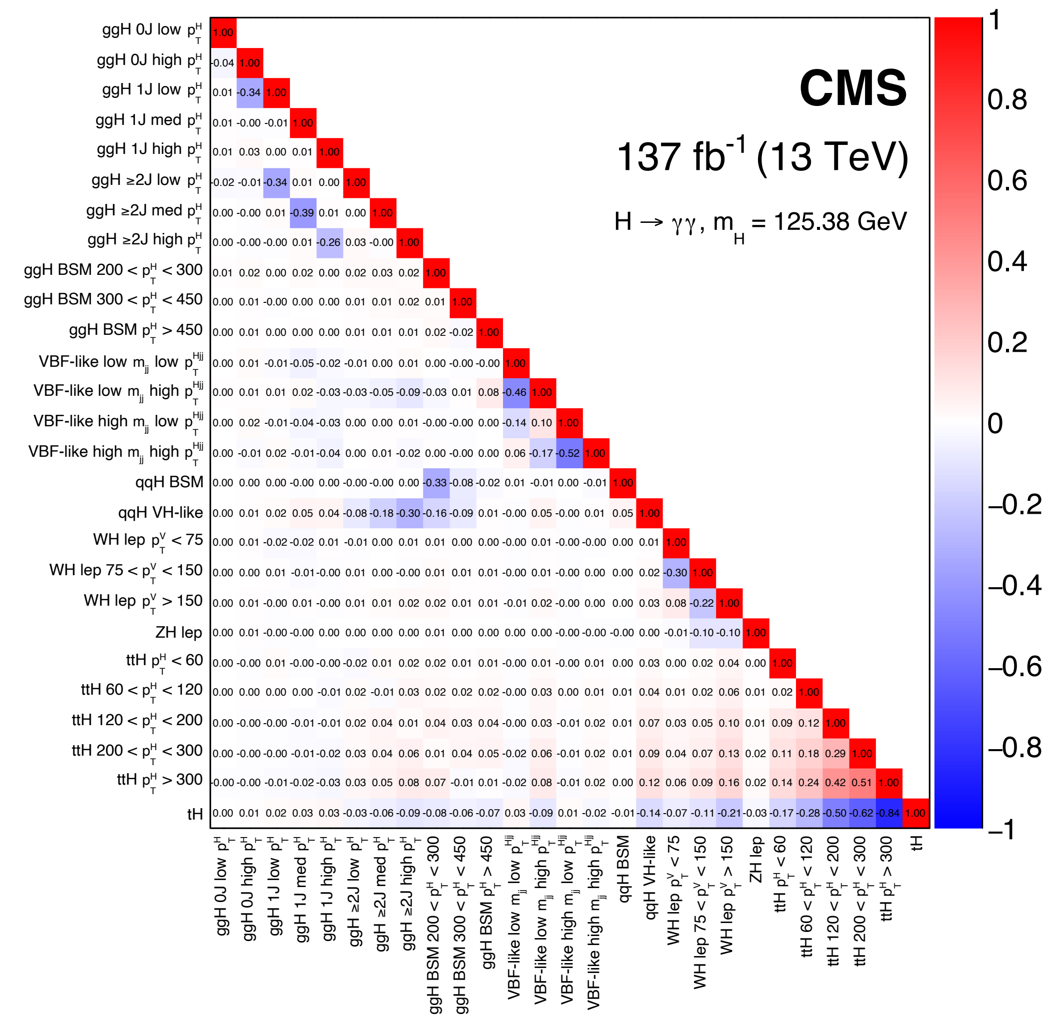
Tasks:

☐ Add all production modes

☐ Add decay partial width parameterisations

☐ Consider linearisation in prod x decay parameteriation

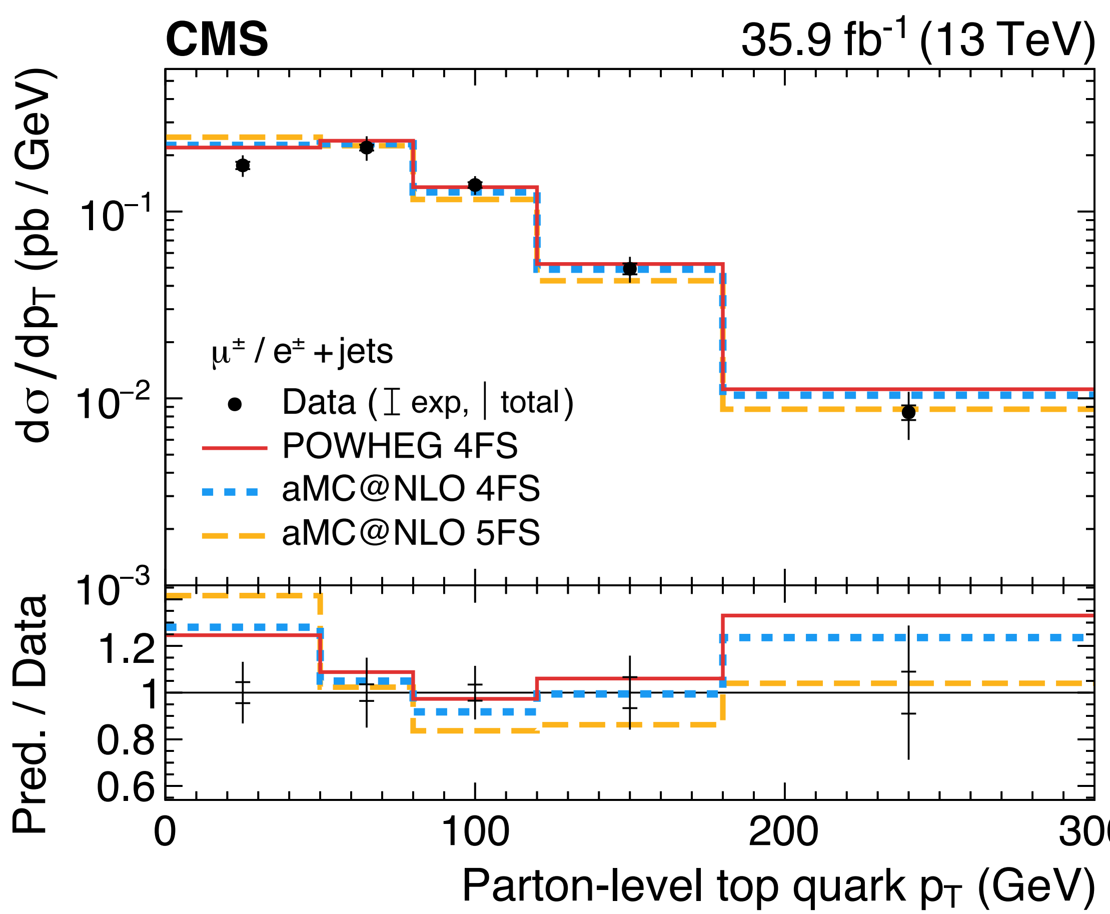


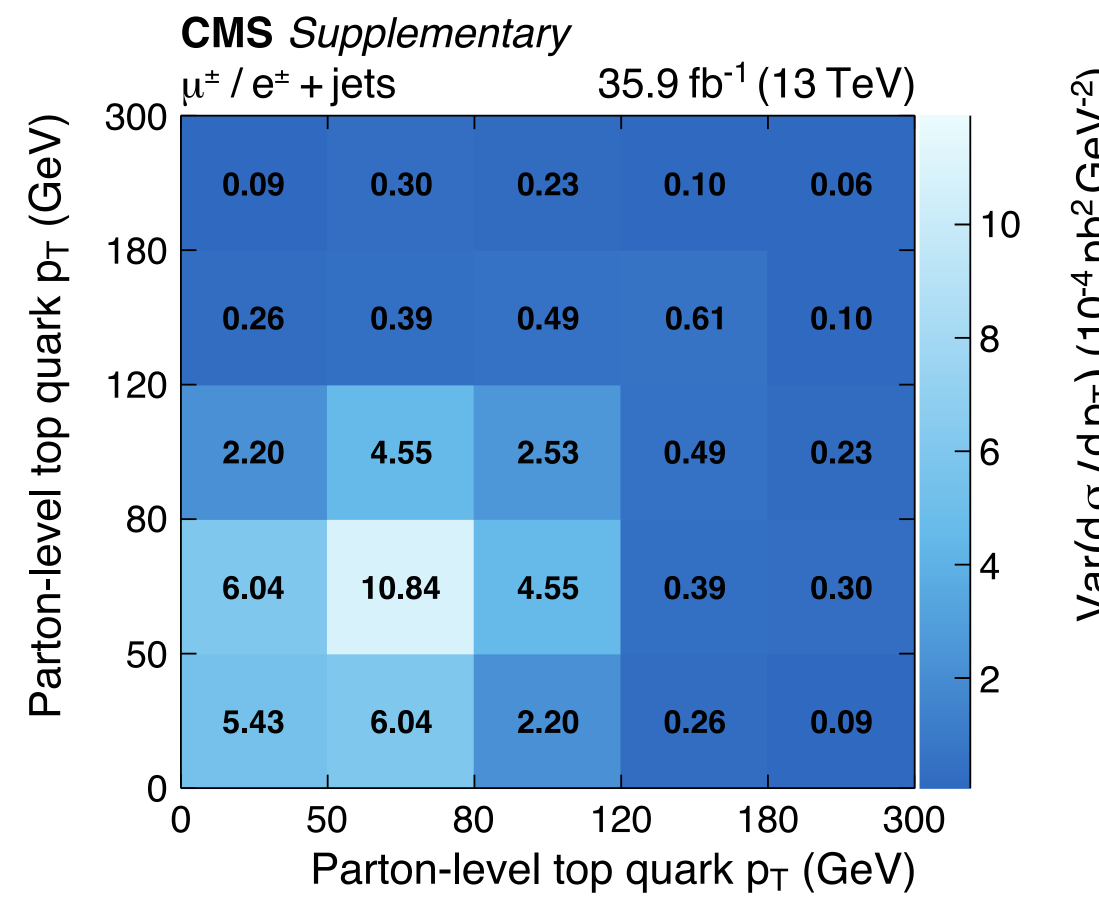


TOP-17-023: single top (t-channel)

- Paper contains several differential observables, just use top quark p_T here
- Fitted signal yields unfolded to parton and particle level:

$$\chi^2 = (\vec{y} - \boldsymbol{R}\epsilon\vec{x})^{\mathrm{T}} \boldsymbol{V}_y^{-1} (\vec{y} - \boldsymbol{R}\epsilon\vec{x}) + \underbrace{\tau^2 \left\| \boldsymbol{L}(\vec{x} - \vec{x}_0) \right\|^2}_{\text{regularisation}} + \lambda \sum_i (\vec{y} - \boldsymbol{R}\epsilon\vec{x})_i ,$$





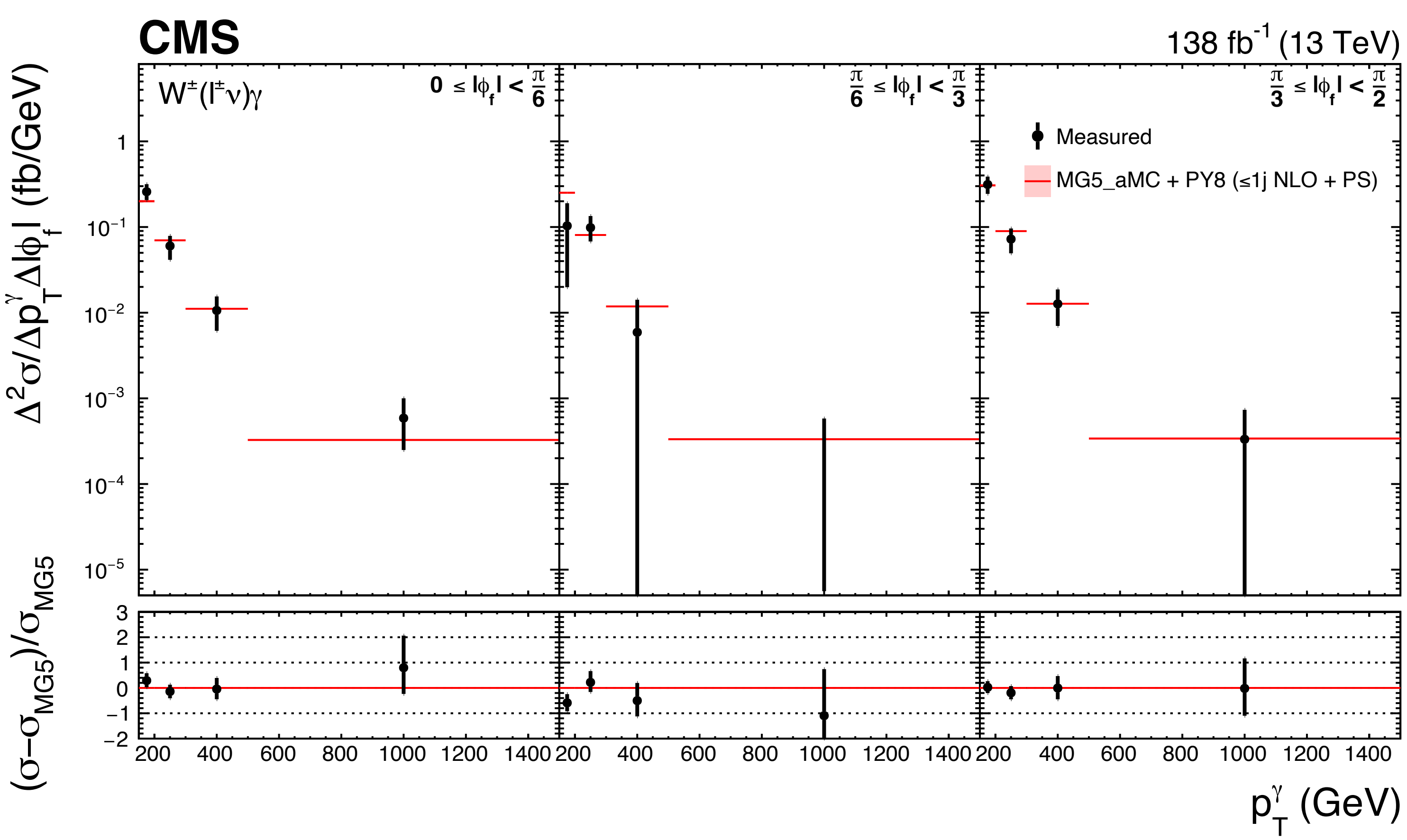
Tasks:

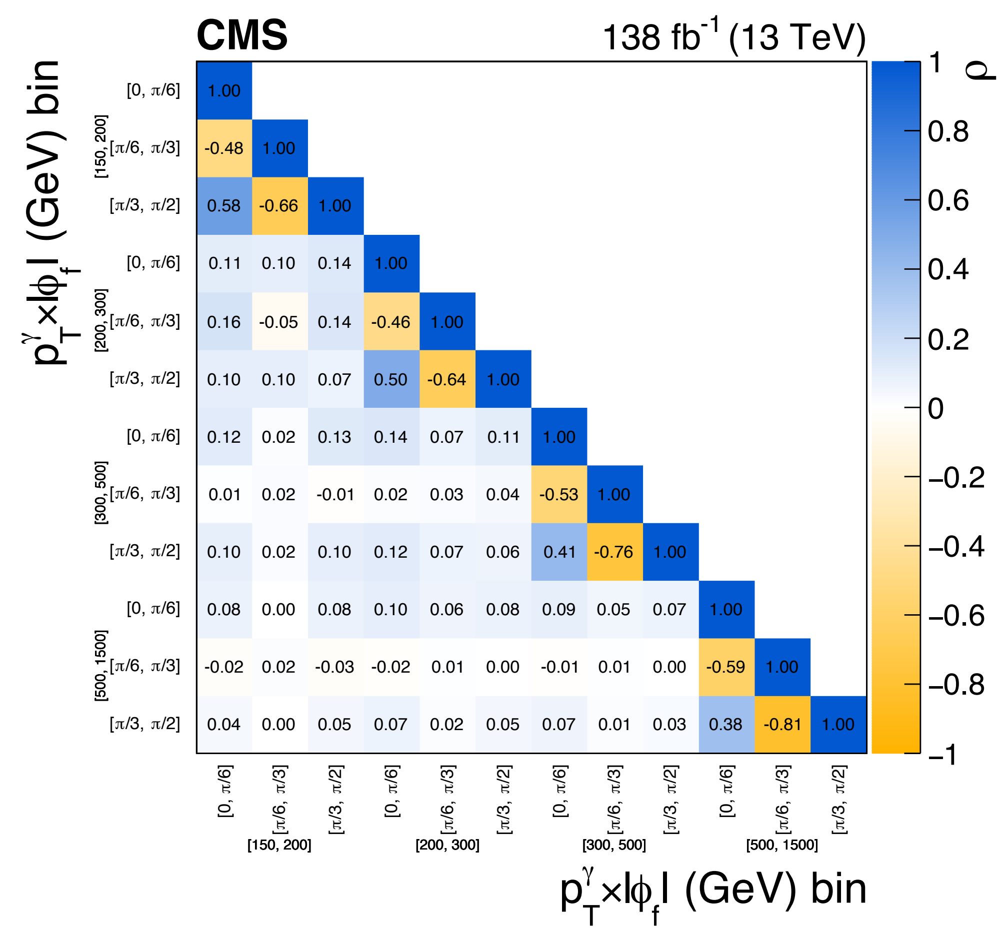
☐ How can/should we include in a future full likelihood comb?

☐ What to take as the “SM prediction”… need to add theory uncertainties

SMP-20-005 (Wγ differential)

- Double-differential cross section in p_T^γ x $|\varphi_f|$
  - Latter enhances sensitivity to interference term with some operators
  - Paper includes limits on single operator (c_W), using earlier version of SMEFTsim
  - Similar approach to Higgs STXS: likelihood fit with parameterisation of fiducial bins





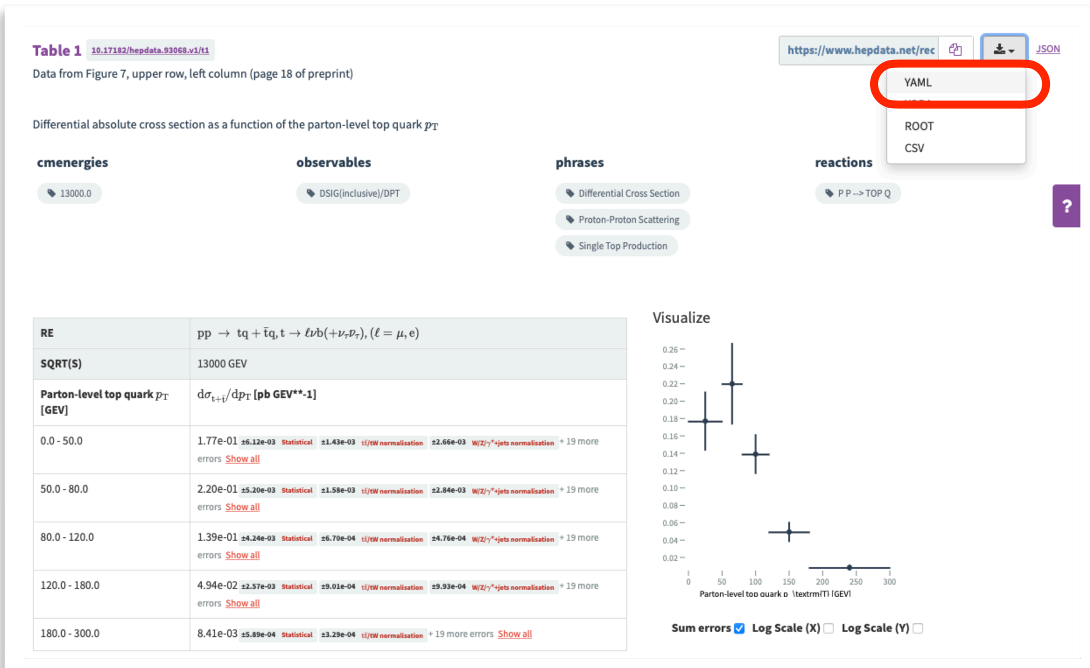
Tasks:

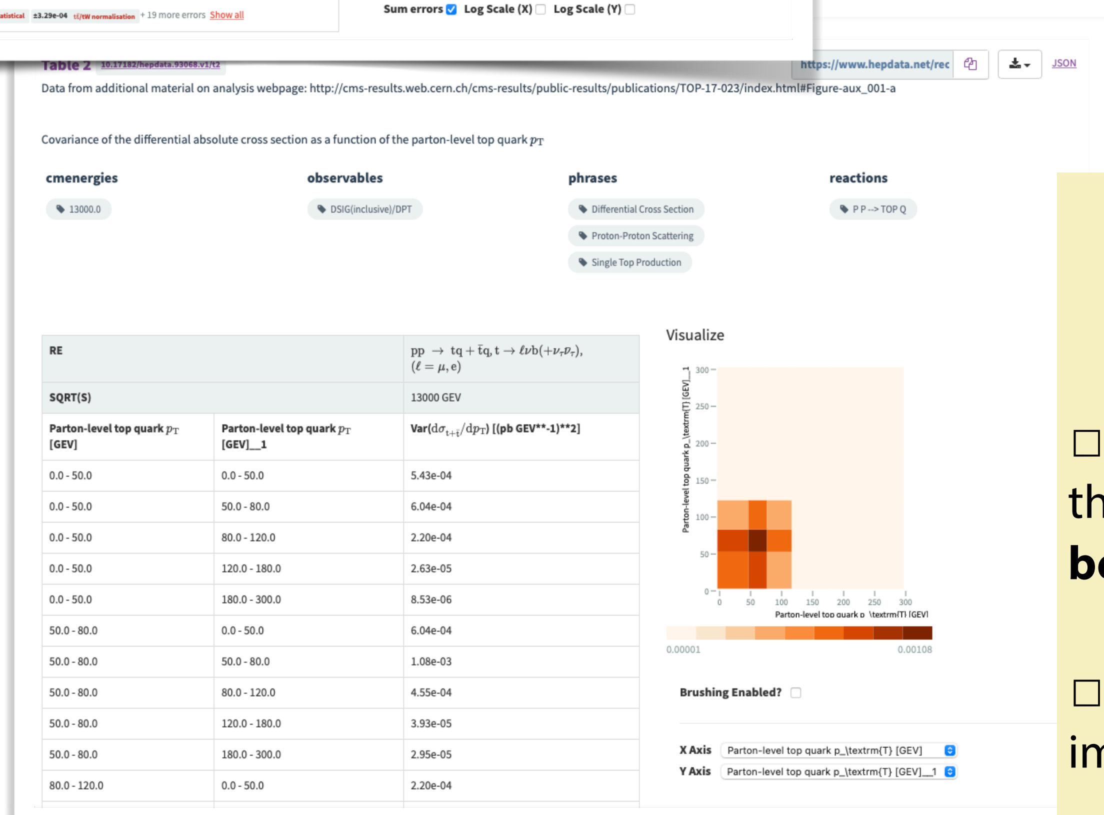
☐ SM theory uncertainties on prediction not taken into account

☐ Good analysis to investigate limits of Gaussian approx.

Step 1: extract HEPData inputs

- All steps currently performed in one script: testFit.py





- For each analysis extract:
  - Vectors of measurements and total uncertainties, relative to the SM predictions: $\vec{\mu}$, $\vec{\sigma}$
  - Construct covariance matrix: $\mathbf{C}$
- Each analysis records info in slightly different format

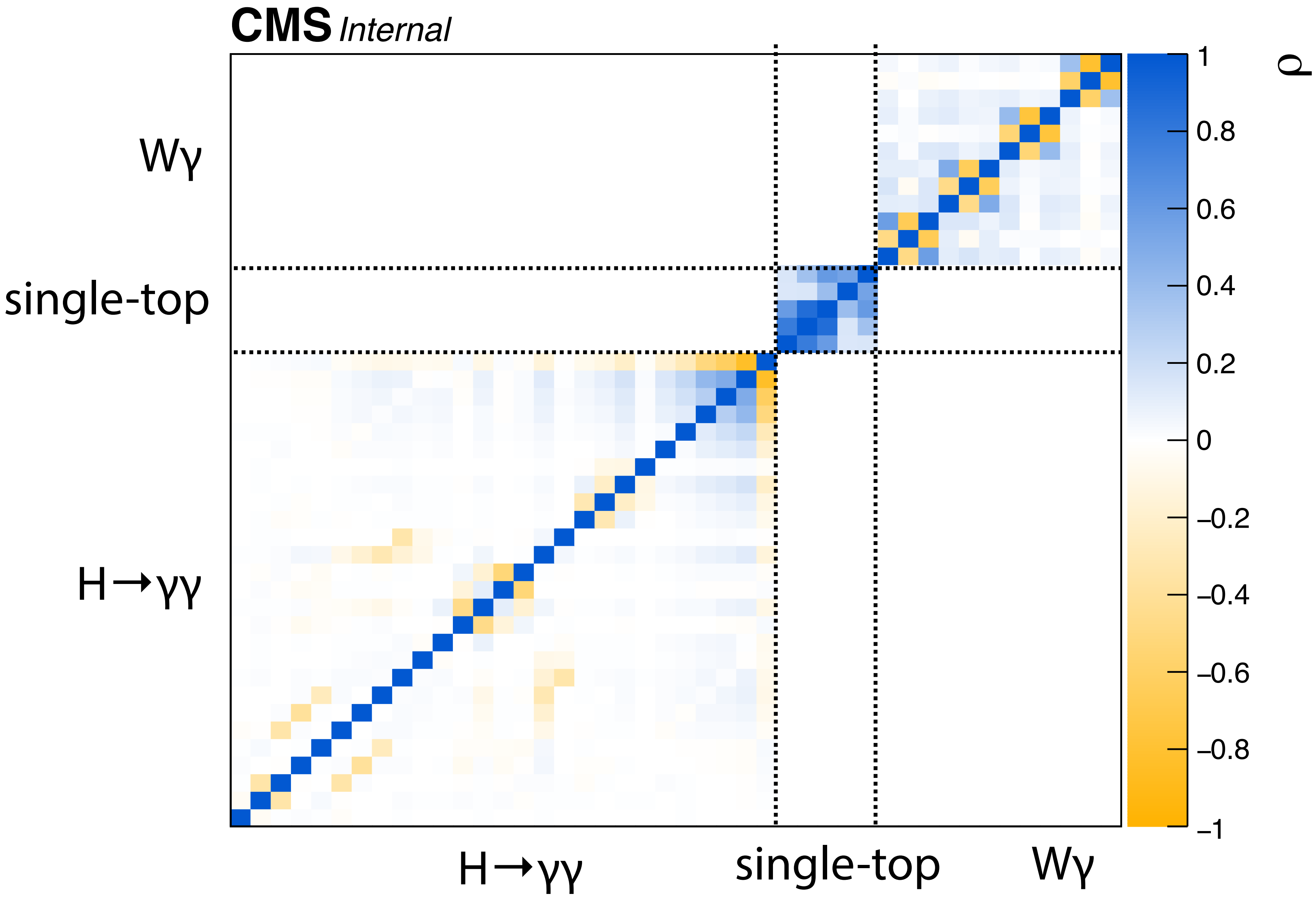
Tasks:

☐ Make a separate parsing step (customised for each analysis) that dumps info in full standardised format - **this part could be CMS/ATLAS common?**

☐ Eventually will want to check SM theory uncertainties are implemented - chance to add inter-channel correlations

Step 2: combined covariance matrix

- Combine $\vec{\mu}$ for all analyses, and create trivial block-diagonal covariance matrix:



- Defines multi-variate Gaussian PDF for set of cross section parameters $\vec{x}$

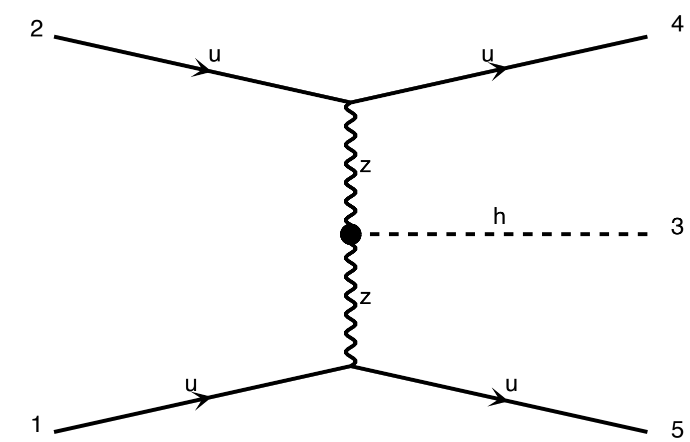
$$P(\vec{x}\,|\,\vec{\mu}) = \exp\left[-\frac{(\vec{x}-\vec{\mu})\mathbf{C}^{-1}(\vec{\mathbf{x}}-\vec{\mu})}{\sqrt{\det(2\pi\mathbf{C})}}\right]$$

Step 3: generate EFT parameterisations

- Using EFT2Obs with analysis RIVET routines
  - For now, using simplest LO process definitions:

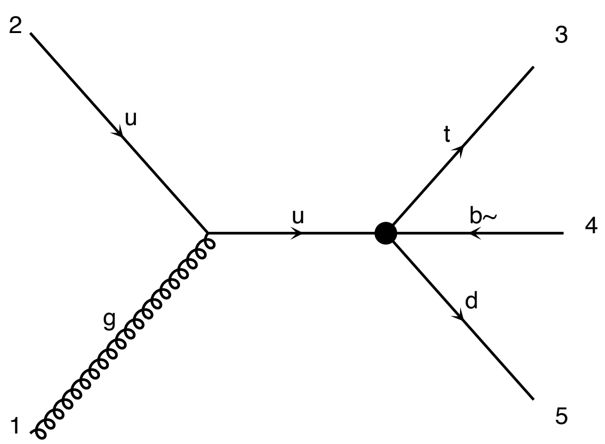
qqH:

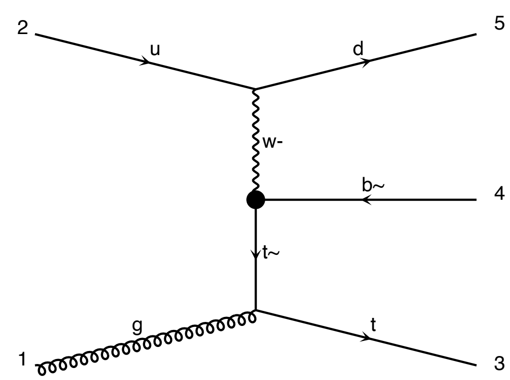
```
generate p p > h j j QCD=0 NP<=1
```



single-top:

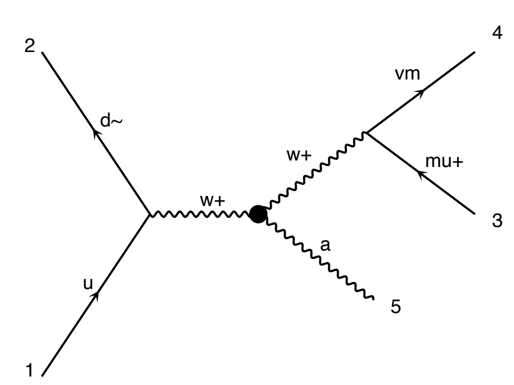
```
generate p p > t b~ j $$ w+ w- NP<=1 @0
add process p p > t~ b j $$ w+ w- NP<=1 @1
```

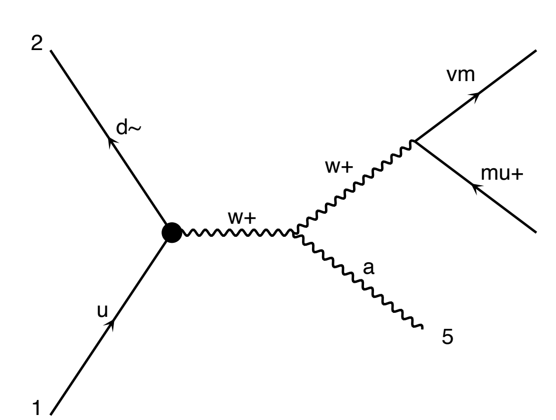




Wγ:

```
generate p p > lep nu a QED=4 NP<=1 SMHLOOP=0
```





Tasks:

- ☐ Consider adding additional jet emission processes and using merging
- ☐ For some processes consider SMEFT@NLO (can test how well QCD/EFT factorise)
- ☐ Propagator corrections?

Step 3: generate EFT parameterisations

- EFT2Obs will generate events and run reweighting
- Challenges:
  - Including all 129 CP-even operators in the reweighting step prohibitive: (8514 points!)
  - Can restrict to only operators that actually affect process
    - Not trivial to determine this (can be done by hand, but time consuming) - see next slide
- Three processes sensitive to **21/129** operators

Tasks:

✅ Develop automatic detection of all operators that affect the generated process
☐ Decays with madspin are supported, but no EFT effects will be considered - inconsistent?

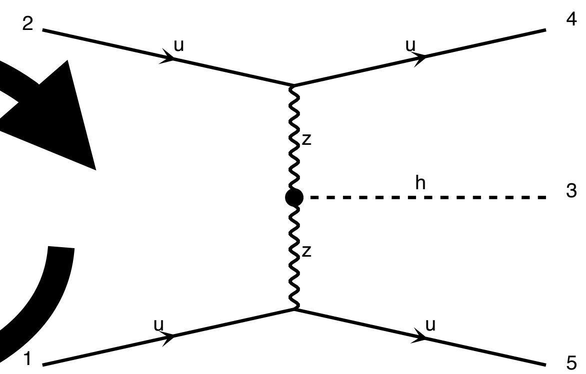
Operator		Top	qqH	Wγ
cHj3	c_{Hj}^{3}	✅	✅	✅
cHl3	c_{Hl}^{3}	✅	✅	✅
cll1	$c_{\mathrm{ll}}^{\mathrm{Prime}}$	✅	✅	✅
cHDD	c_{HD}		✅	✅
cHWB	c_{HWB}		✅	✅
cbGRe	c_{bG}	✅		
cbWRe	c_{bW}	✅		
cHQ3	c_{HQ}^{3}	✅		
cHtbRe	c_{Htb}	✅		
cQj31	c_{Qj}^{31}	✅		
cqj38	c_{Qj}^{38}	✅		
ctGRe	c_{tG}	✅		
ctWRe	c_{tW}	✅		
cHG	c_{HG}			
cHB	c_{HB}		✅	
cHbox	$c_{H\Box}$		✅	
cHd	c_{Hd}		✅	
cHj1	c_{Hj}^{1}		✅	
cHu	c_{Hu}		✅	
cHW	c_{HW}		✅	
clj3	c_{lj}^{3}			✅
cW	c_{W}			✅

Finding process-specific parameters

- Script developed by Matthew Knight, now part of EFT2Obs
- Analyses a Madgraph process directory to find subset of model parameters that affect process

```
generate p p > h j j QCD=0 NP<=1
```

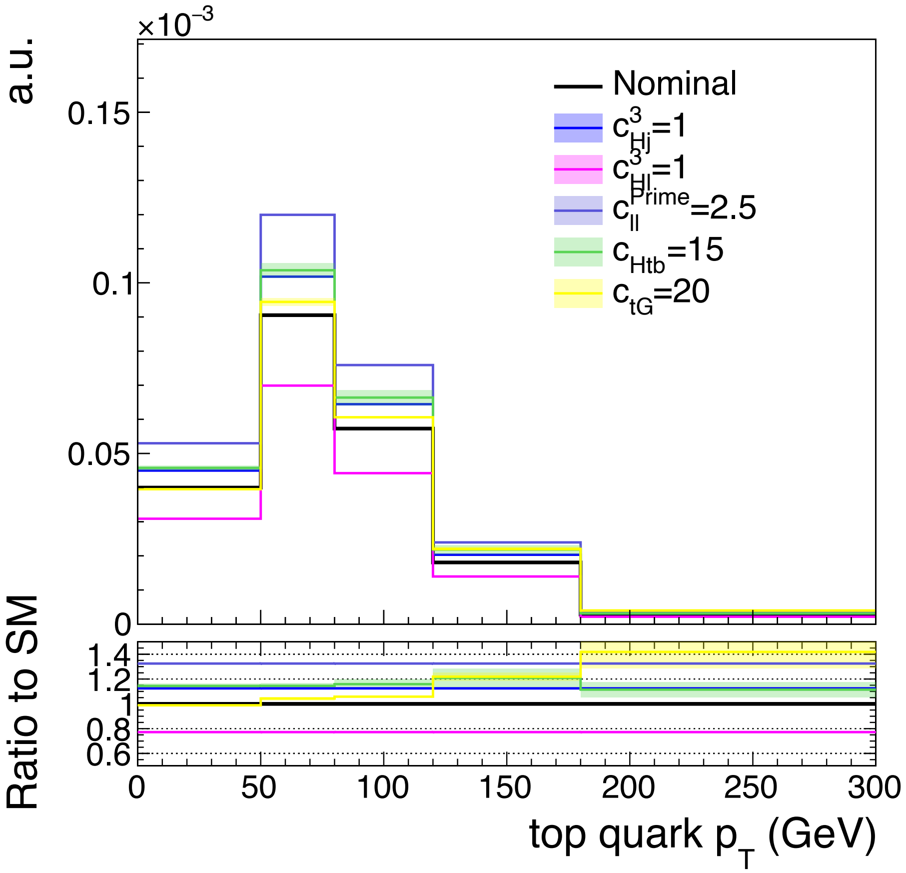
```
python scripts/auto_detect_operators.py -p qqH-SMEFTsim3
```

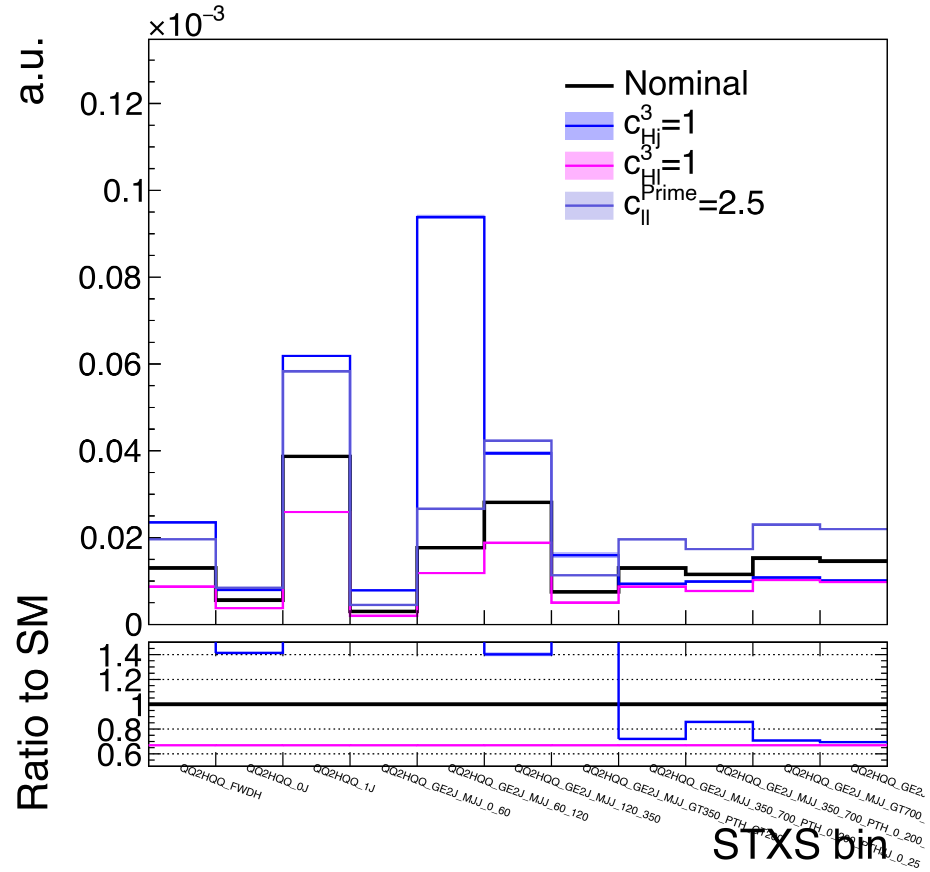


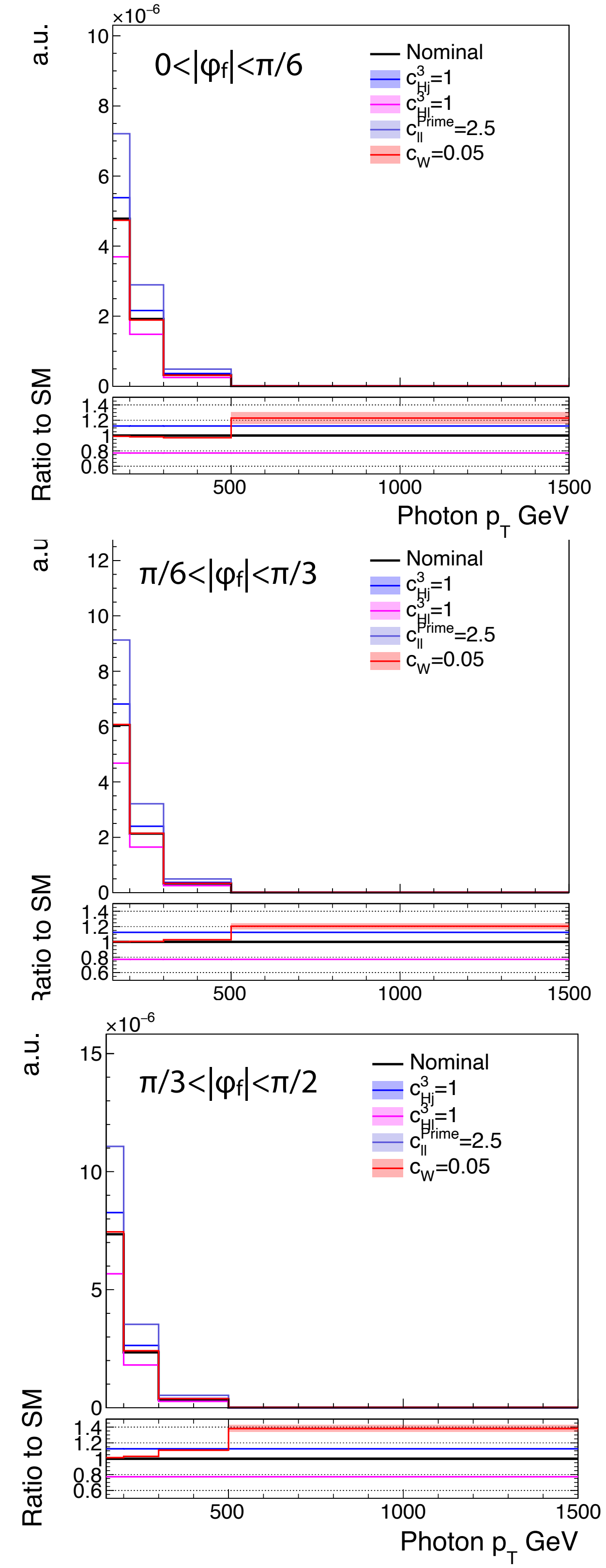

```
>> Loading model: SMEFTsim_topU3l_MwScheme_UFO
>> Possible parameters: ['cG', 'cW', 'cH', 'cHbox', 'cHDD', 'cHG', 'cHW', 'cHB', 'cHWB', 'cuHRe', 'ctHRe', 'cdHRe',
'cbHRe', 'cuGRe', 'ctGRe', 'cuWRe', 'ctWRe', 'cuBRe', 'ctBRe', 'cdGRe', 'cbGRe', 'cdWRe', 'cbWRe', 'cdBRe', 'cbBRe',
'cHj1', 'cHQ1', 'cHj3', 'cHQ3', 'cHu', 'cHt', 'cHd', 'cHbq', 'cHudRe', 'cHtbRe', 'cjj11', 'cjj18', 'cjj31', 'cjj38',
'cQj11', 'cQj18', 'cQj31', 'cQj38', 'cQQ1', 'cQQ8', 'cuu1', 'cuu8', 'ctt', 'ctu1', 'ctu8', 'cdd1', 'cdd8', 'cbb',
'cbd1', 'cbd8', 'cud1', 'ctb1', 'ctd1', 'cbu1', 'cud8', 'ctb8', 'ctd8', 'cbu8', 'cutbd1Re', 'cutbd8Re', 'cju1', 'cQu1',
'cju8', 'cQu8', 'ctj1', 'ctj8', 'cQt1', 'cQt8', 'cjd1', 'cjd8', 'cQd1', 'cQd8', 'cbj1', 'cbj8', 'cQb1', 'cQb8',
'cjQtu1Re', 'cjQtu8Re', 'cjQbd1Re', 'cjQbd8Re', 'cjujd1Re', 'cjujd8Re', 'cjujd11Re', 'cjujd81Re', 'cQtjd1Re',
'cQtjd8Re', 'cjuQb1Re', 'cjuQb8Re', 'cQujb1Re', 'cQujb8Re', 'cjtQd1Re', 'cjtQd8Re', 'cQtQb1Re', 'cQtQb8Re', 'ceHRe',
'ceWRe', 'ceBRe', 'cHl1', 'cHl3', 'cHe', 'cll', 'cll1', 'clj1', 'clj3', 'cQl1', 'cQl3', 'cee', 'ceu', 'cte', 'ced',
'cbe', 'cje', 'cQe', 'clu', 'ctl', 'cld', 'cbl', 'cle', 'cledjRe', 'clebQRe', 'cleju1Re', 'cleQt1Re', 'cleju3Re',
'cleQt3Re']
>> Relevant parameters: ['cHB', 'cHDD', 'cHW', 'cHWB', 'cHbox', 'cHd', 'cHj1', 'cHj3', 'cHl3', 'cHu', 'cll1']
>> Relevant parameters from propagator corrections: []
>> Final relevant parameters: ['cHB', 'cHDD', 'cHW', 'cHWB', 'cHbox', 'cHd', 'cHj1', 'cHj3', 'cHl3', 'cHu', 'cll1']
```

Step 3: generate EFT parameterisations

- Result: parameterisation for each bin as $1 + \sum_i A_i c_i + \sum_{i,j} B_{ij} c_i c_j$







- Subset of operators shown for each process
- Mixture of normalisation and shape altering effects
- Anticipate redundant degrees of freedom among all operators considered

Step 4: Inject parameterisation

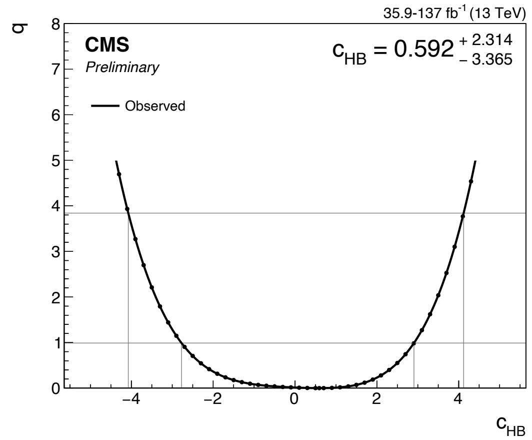
- EFT2Obs outputs JSON-format parameterisation for each bin
- The fitting script reads these, and replaces free parameters $x_i \rightarrow x_i(c_j)$
- Construct a likelihood function from "data" ($\vec{\mu}$) and the reparameterised multivariate Gaussian PDF
- In principle can then fit for all c_i
  - Some manual tuning of parameter ranges required

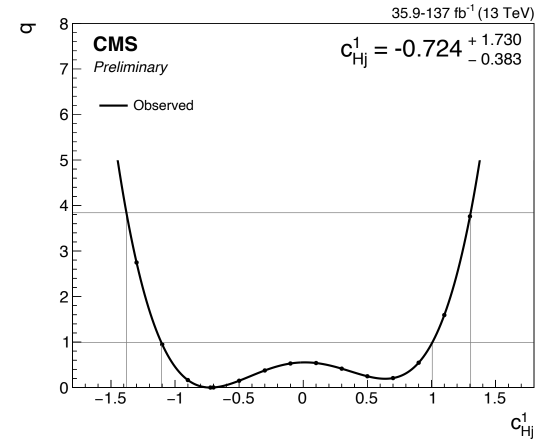
Tasks:

- ☐ Currently requires manual alignment of bin labels in hepData and from EFT2Obs
- ☐ Automatic tuning (rescaling) of the parameter ranges would be useful
- ☐ Anticipate need to optionally inject acceptance corrections in this stage


```
{
  "bin_labels": [
    "pt_t_bin_0",
    "pt_t_bin_1",
    "pt_t_bin_2",
    "pt_t_bin_3",
    "pt_t_bin_4"
  ],
  "edges": [
    [
      0.0,
      50.0
    ],
    [
      50.0,
      80.0
    ],
    [
      80.0,
      120.0
    ],
    [
      120.0,
      180.0
    ],
    [
      180.0,
      300.0
    ]
  ],
  "bins": [
    [
      [
        -0.003070670529258986,
        5.0886748780813544e-05,
        "cbgre"
      ],
      [
        -0.005930811202120159,
        0.0023186485688322465,
        "cbwre"
      ],
      [
        0.12124548772256405,
        1.6352508475238297e-06,
        "chj3"
      ],
      [
        -0.24250224594130354,
        0.0,
        "chl3"
      ],
      [
        0.12124548772256405,
        1.6352508475238297e-06,
        "chq3"
      ],
      [
        0.001983512945884298,
        0.0007162703096425856,
        "chtbre"
      ],
      [
        0.12124548772256405,
        1.6352508475238297e-06,
        "cll1"
      ],
```

Result

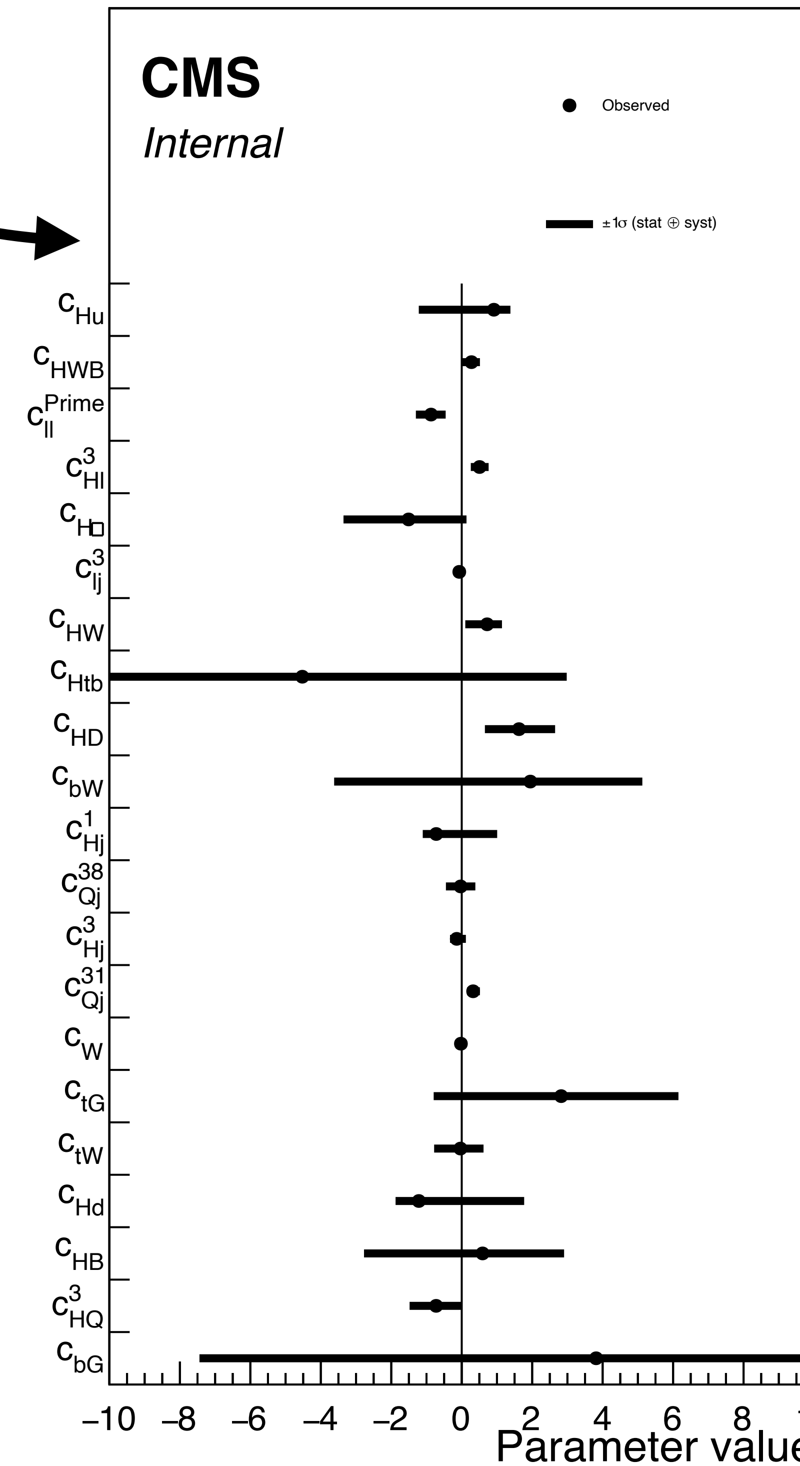




. . .

- Degeneracy between parameters means we cannot fit all 21 simultaneously

- Instead, fit and scan each one at a time, keeping all other $c_i = 0$



Tasks:

☐ Implement **principle component analysis** to identify DoFs we can fit simultaneously

☐ Add flexibility to fit with or without quadratic $c_i c_j$ terms

☐ Improve workflow for plotting NLL scans, interval extraction, making summary plot, etc (currently borrowed from HComb tools)

☐ Add a visualisation of constraint coming from each input analysis (so we can quickly assess which is the most sensitive for each operator)

Summary & discussion

- Basic EFT combination tool set up
- For analyses already fitting cross sections directly (likelihood unfolding) conceptually straightforward to move to full likelihood combination
- Next steps:
  - Implement similar set of ATLAS analyses
  - Cross-validation of parameterisations and fit results
  - Further studies of uncertainties, validity, truncation, handling flat directions, etc
  - Include more analyses to start exploring physics sensitivity