Comparison and discussion on multi-parameter fits

Michele Pinamonti (ATLAS)

Markus Seidel, Pietro Vischia (CMS)

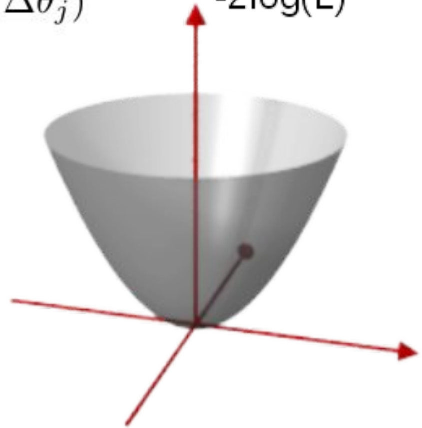
University and INFN Roma "Tor Vergata", University of Maryland, UC Louvain

LHCtopWG Meeting, CERN 20th - 21st Nov 2018

The profile likelihood fit technique

- **Profile likelihood fit (PLF)** = a *statistically meaningful* way of including **systematic uncertainties** in a maximum likelihood fit
    - systematics included as "***constrained***" **nuisance parameters**
    - the idea behind is that systematic uncertainties on the measurement of μ come from ***imperfect knowledge*** of parameters of the model (*S* and *B* prediction)

$$\mathcal{L}(\boldsymbol{n}, \boldsymbol{\theta}^0 | \mu, \boldsymbol{\theta}) = \prod_{i \in bins} \mathcal{P}(n_i | \mu \cdot S_i(\boldsymbol{\theta}) + B_i(\boldsymbol{\theta})) \times \prod_{j \in syst} \mathcal{G}(\theta_j^0 | \theta_j, \Delta\theta_j)$$



- The fit procedure becomes a **multi-dimensional Likelihood maximisation** problem
    - the fit **result** is not just the value (and uncertainty) on parameter(s) of interest (POI), but **a set of values** for all the parameters, including nuisance parameters:

$$(\hat{\mu}, \hat{\theta}_0, ...\hat{\theta}_{N-1}) : \mathcal{L}(\hat{\mu}, \hat{\boldsymbol{\theta}}) = max$$

    - usually **Wilks' theorem** and **asymptotic regime** used to estimate uncertainties and extracting exclusion limits and discovery significance **without integrations or toys**

Latest top results using profiling

ATLAS EXPERIMENT

- Searches for FCNC (tHq multi-lep, tHq yy, tHq bb, tZq)
- ttV, tZ
- Single top Wt and s-channel (8TeV)
- tt+gamma (only for fiducial cross-section, not for differential)

Outside TopWG, but top-related:

- ttH (bb, multi-lepton)
- Search for tt resonance
- 4 tops

CMS

"Standard measurements" moving to profiling:

- Cross-section (dilepton, lepton+jets, 5 TeV)
- Top mass (dilepton)

TOP searches and rare processes:

- 4 tops (same-sign, opposite-sign)
- FCNC, tZ, ttV

Related to TOP:

- ttH (leptonic, all-jet)

Pros and cons of profiling

Pros:

- **Systematics** are really **part of the fit** procedure ⇒ nice properties like:
  - *the precision always improves when adding more information, i.e. more bins*
- **Limit** setting, **significance** evaluation and **combination** of different analyses very natural
- Same procedure adopted by ATLAS and CMS
- ***In most cases*** → **reduction of total uncertainty**, thanks to constraints on nuisance parameters / in-situ calibration of systematic uncertainties

Cons:

- Minimisation procedure for complex fits can become **computationally intense**
- **Definition of systematic** uncertainties delicate:
  - complete systematic model with proper granularity needed
  - pulls and constraints on nuisance parameters need to be understood
- **Limited statistics in MC** (w.r.t. data) becoming an issue, especially for systematic uncertainties
- *Until recently*, only applicable / applied to signal strength / total cross-section measurements (but see last slides)

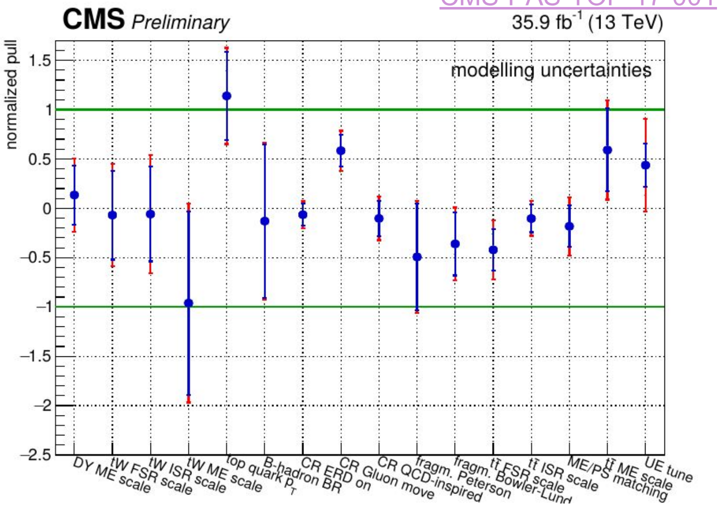
NP pulls, constraints and correlations

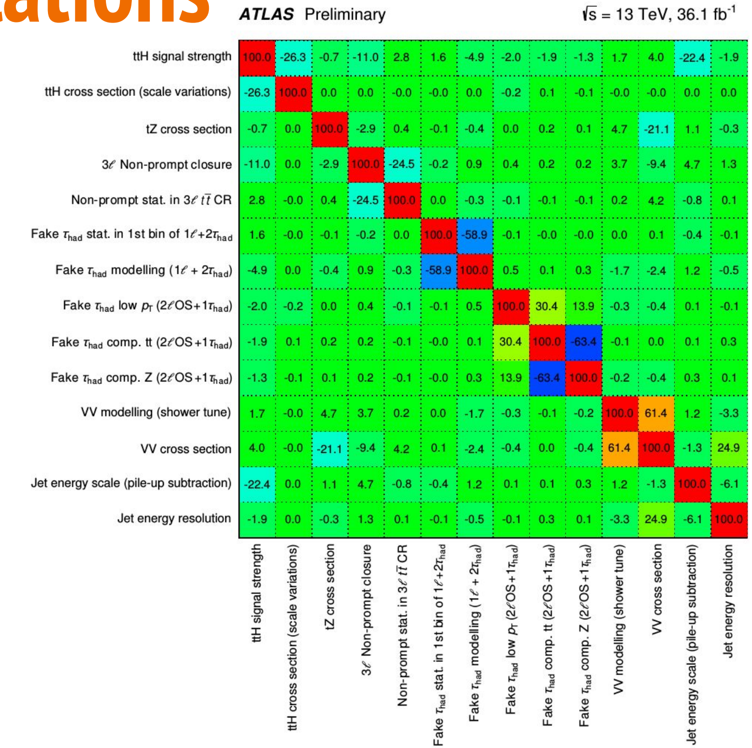
Phys. Rev. D 97, 072003

- Useful to **monitor** NP **pulls** and **constraints**:

CMS-PAS-TOP-17-001

$(\hat{\theta}-\theta_0)/\Delta\theta$



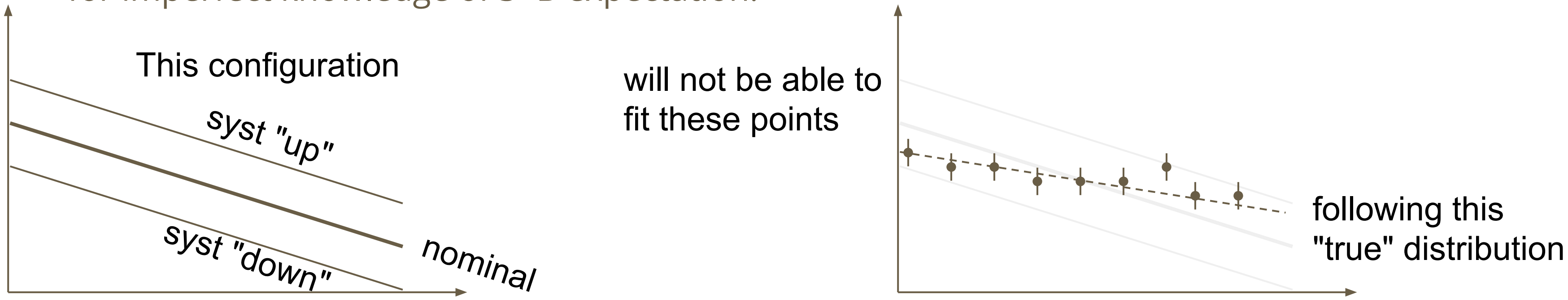


- Important to consider also NP **correlations**:
  - uncertainties on NPs (*and POI*) extracted from ***covariance matrix***, which includes ***correlation coefficients***
    - correlation **built by the fit**, even if completely independent / uncorrelated sources of uncertainty before the fit (*correlation in the improved knowledge of the parameters*)
    - (anti-)correlations can **reduce** total post-fit uncertainty!

Profiling issues

- The profile likelihood approach is **valid** with some **assumptions**
  - in particular, assumed that "*nature*" can be described by the model with ***a single combination of values*** for the parameters

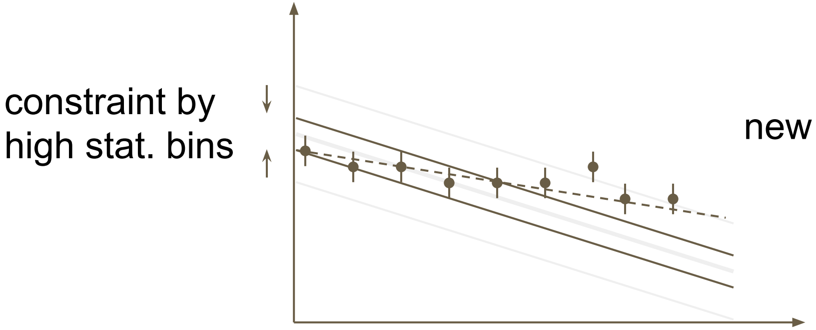
- Cannot just take *large uncertainties* hoping that they are enough to cover for imperfect knowledge of S+B expectation!



- "**Flexibility**" / "**granularity**" of the systematics model needs to be considered

The constraint issue

- Flexibility more and more **critical** when **statistical uncertainty** on data becomes less and less important w.r.t. systematics
  - e.g. taking the example before:



new physics!?

... no, data just following background *real* distribution...

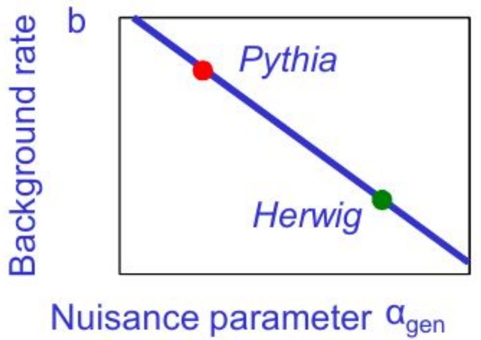
- More real examples:
  - **single JES** systematic NP across all jet energy spectrum allows high-stats low-energy control regions/bins to calibrate JES for high energy jets → intended?
  - **simple flat ± 50%** overall uncertainty on background, probably enough to cover uncertainties also in remote phase-spaces (*e.g. tails of distributions for W+HF-enriched selection*), but data in CRs will constrain it to <5%, propagated to SRs... → ok?

Theory modeling systematics

- ***Experimental systematics*** nowadays often well suited for profile likelihood application:
  - come from calibrations ⇒ gaussian constraint appropriate
  - broken-down into several independent/uncorrelated components (JES, *b*-tagging...)
- Different situation for **theory systematics**:
  - **difficulty 1:** what is the **distribution** of the subsidiary measurement?
  - **difficulty 2:** what are the **parameters** of the systematic?
    - can a combination of the included parameters describe **any possible** configuration?
    - is **any allowed value** of the parameter physically meaningful?

See: https://indico.cern.ch/event/287744/contributions/1641261/attachments/535763/738679/Verkerke_Statistics_3.pdf

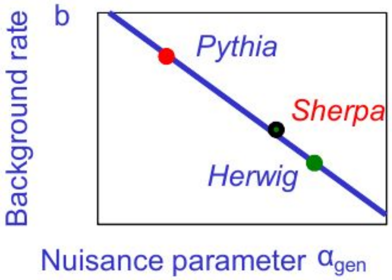
- ***The obviously tricky case:* "two point" systematics**
  - e.g. Herwig vs. Pythia as "parton shower and hadronization model uncertainty", as a single NP



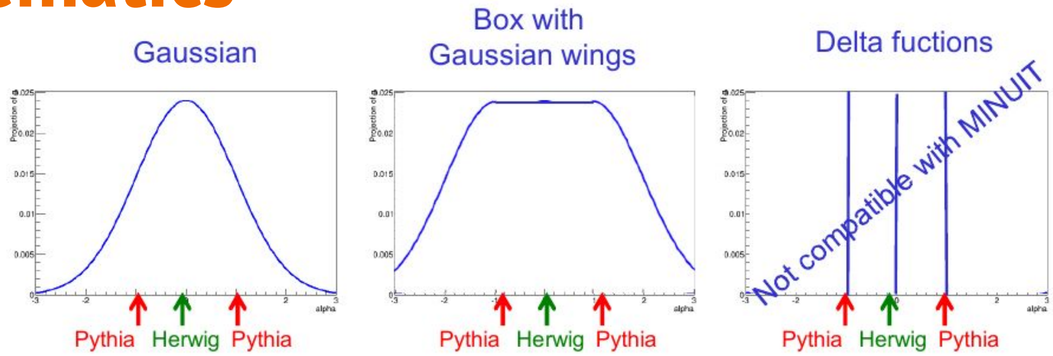
Theory modeling systematics

One-bin case:

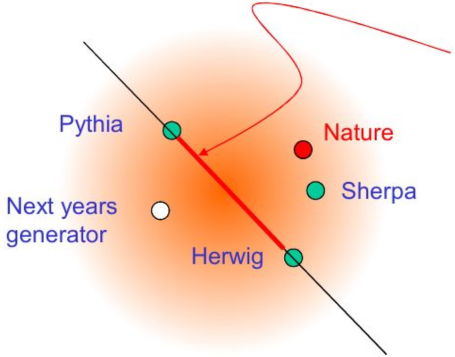
- reasonable to think that "Sherpa" can be between Herwig and Pythia



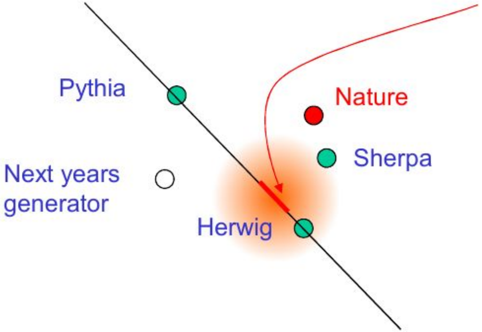
Which prior?



Pre-fit / non-constrained NP could be fine to cover for all possible models...

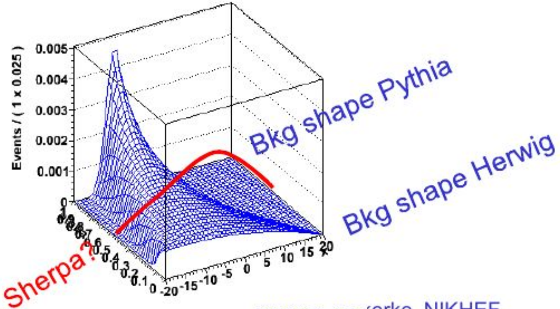


... but is this level of constraint ok?



Shape case:

- Sherpa can be different from linear combination of Py and Her...



Wouter Verkerke, NIKHEF

Theory modeling systematics

- *A not-so-obviously tricky case:*
  - scale uncertainties

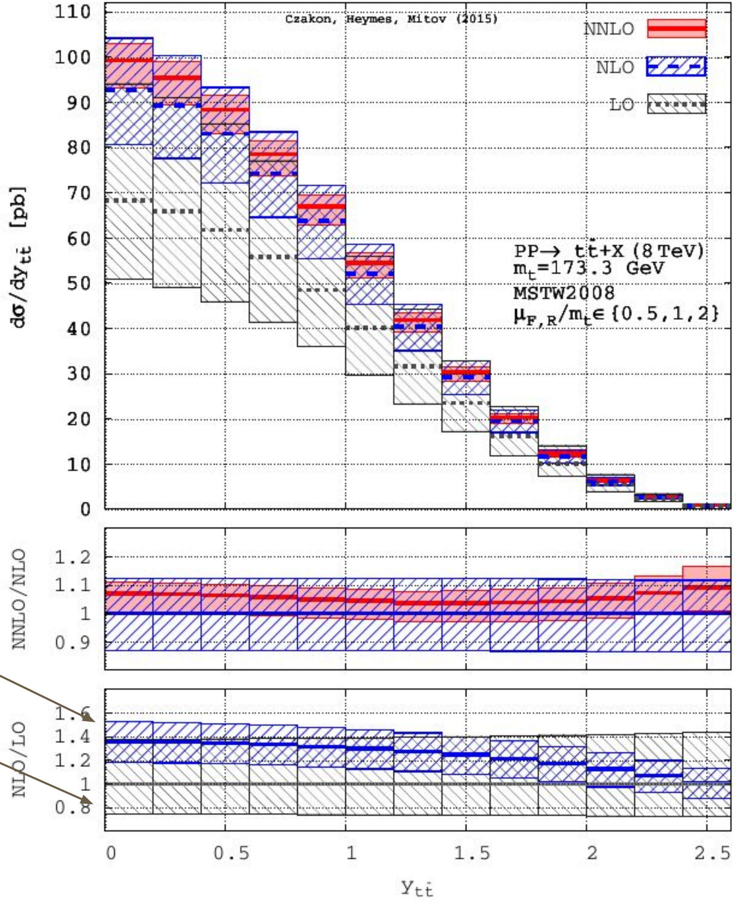
Take NLO scale variations
as uncertainty (missing NNLO MC)

⇒ flat uncertainty here,
and NNLO is within
uncertainty, but
NNLO/NLO is not flat!

+1

-1

Suppose data looks like NNLO, we measure y_{tt},
we constrain scale syst. in low y_{tt} bins
⇒ new physics at high y_{tt}?

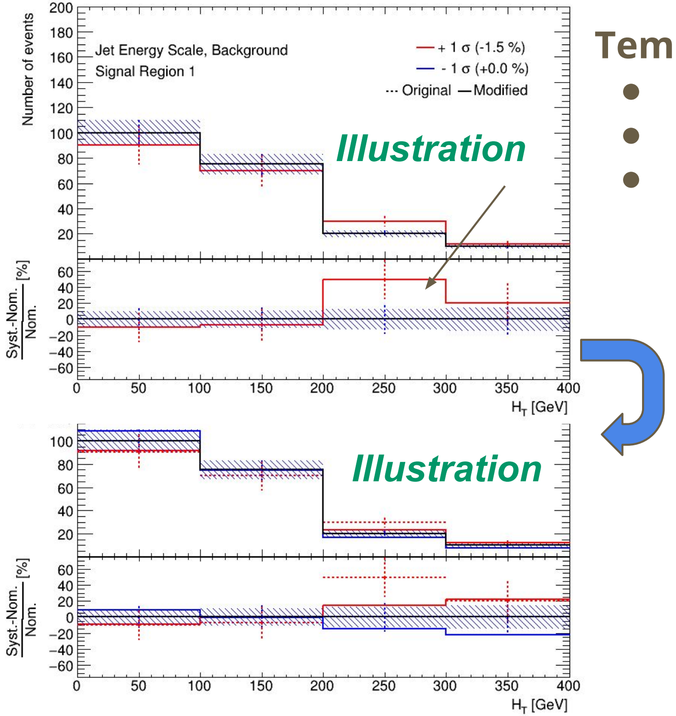


Phys. Rev. Lett. 116, 082003

- New idea by Frank Tackmann (talk at LHC EW precision group):
  replace scale uncertainties by taking the coefficients of higher-order corrections as floatable NP
  → can be constrained by data (if constraint ≲ 0.1: add structure of next order)

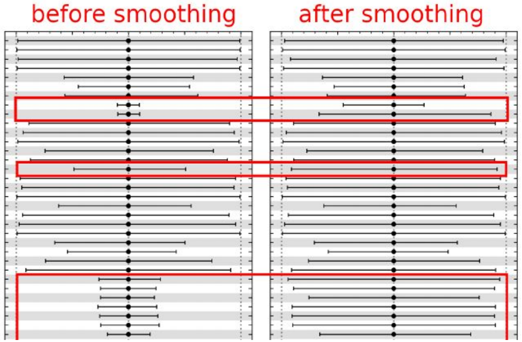
Statistical fluctuations in systematic templates

- See this talk at the latest ATLAS+CMS stat meeting
- Statistical fluctuations in templates used to define systematics can lead to artificial constraints
  ⇒ artificially small total uncertainty!



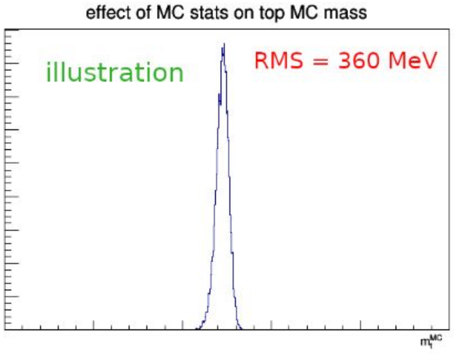
- Different ways to control / mitigate this effect:

Template smoothing:
- largely used by ATLAS
- different smoothing algorithms
- current recommendation:
  - choose case-by-case
  - check systs one-by-one
  - compare different choices



(taken from Defranchis' talk)

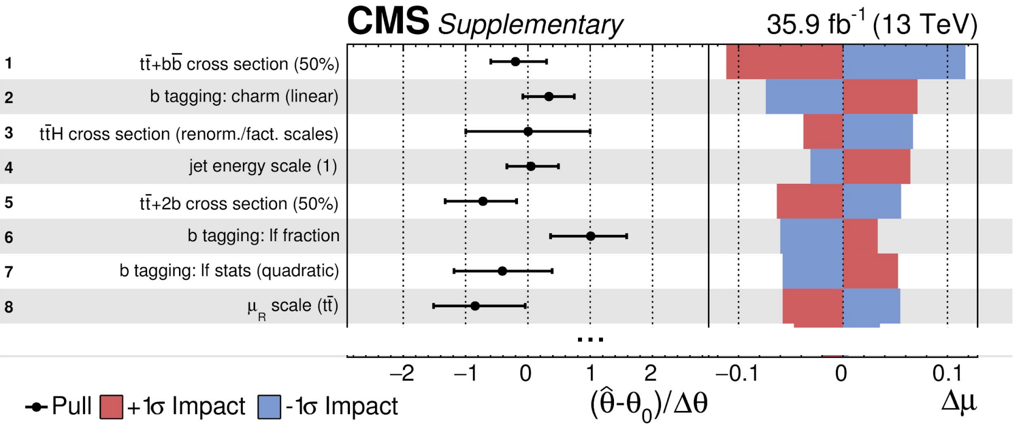
Pseudo-experiments:
- used in CMS (CMS-PAS-TOP-17-001)
- repeat fit N times with fluctuation
  of systematic templates
- spread of fitted results taken as additional uncertainty

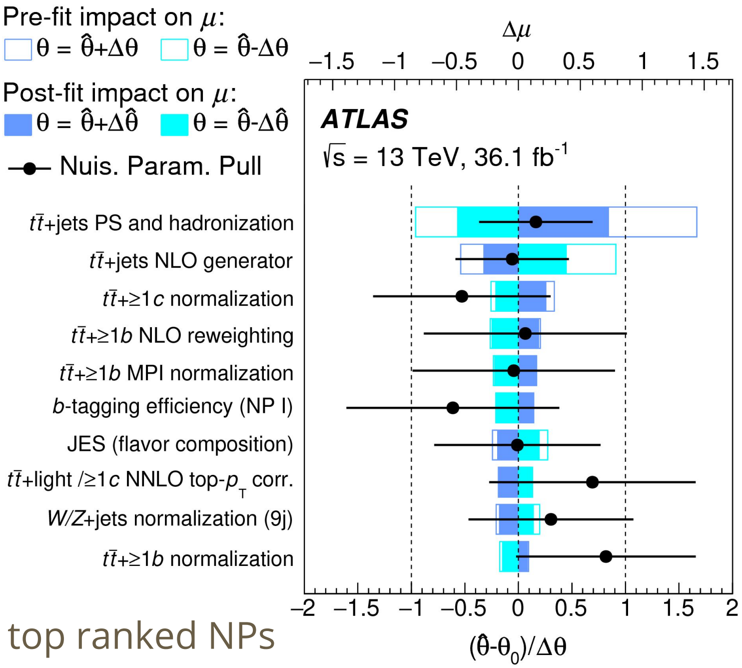


Impact of NP on the POI aka the "ranking plot"

- To answer the question "***which systematics are more important?***"
- The "**ranking plot**" shows *pre-fit* and *post-fit* **impact** of **individual NP** on the determination of μ:
  - **each NP fixed** to ± 1 pre-fit and post-fit sigmas ($\Delta\theta$ and $\Delta\hat{\theta}$ = uncertainty on $\hat{\theta}$)
  - fit re-done with *N-1* parameters
  - impact extracted as difference in **central value** of μ

arXiv:1804.03682





- Often combined with pulls and constraints on these top ranked NPs
- **Why not always publishing such a plot?**

arXiv:1811.02305

Breakdown of uncertainty in measurement

- To answer a similar but different question:
  - **how much of the total uncertainty** comes from a certain ***set of*** systematic uncertainties?
  - or similarly, how large is the pure "statistical uncertainty"?
    (keep in mind that)
- Recommended procedure
  (used by ATLAS and some CMS results):
  - **fix a group** of NPs to post-fit values
  - repeat the fit
  - look at **error on μ** this time
    and get $\Delta\mu$ as quadratic difference between full and reduced error
  - statistical uncertainty obtained by fixing all NPs

arXiv:1804.03682

Uncertainty source	$\pm\Delta\mu$ (observed)	$\pm\Delta\mu$ (expected)
Total experimental	$+0.15/-0.16$	$+0.19/-0.17$
b tagging	$+0.11/-0.14$	$+0.12/-0.11$
jet energy scale and resolution	$+0.06/-0.07$	$+0.13/-0.11$
Total theory	$+0.28/-0.29$	$+0.32/-0.29$
$t\bar{t}$+hf cross section and parton shower	$+0.24/-0.28$	$+0.28/-0.28$
Size of the simulated samples	$+0.14/-0.15$	$+0.16/-0.16$
Total systematic	$+0.38/-0.38$	$+0.45/-0.42$
Statistical	$+0.24/-0.24$	$+0.27/-0.27$
Total	$+0.45/-0.45$	$+0.53/-0.49$

- In some analysis **not fully clear** what is done
  (e.g. CMS-PAS-TOP-17-011, where individual sources larger that total uncertainty quoted...)

Profile likelihood fit for shape analyses

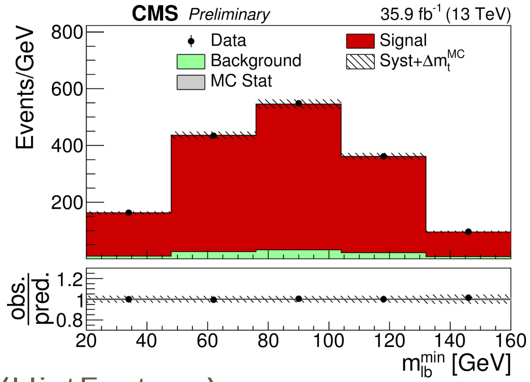
- **Usually** PLF applied just to **cross-section / signal strength** measurements
  - i.e. **POI** applied as ***multiplicative factor*** to signal process prediction:

$$\mathcal{L}(\boldsymbol{n}, \boldsymbol{\theta}^0 | \mu, \boldsymbol{\theta}) = \prod_{i \in bins} \mathcal{P}(n_i | \mu \cdot S_i(\boldsymbol{\theta}) + B_i(\boldsymbol{\theta})) \times \prod_{j \in syst} \mathcal{G}(\theta_j^0 | \theta_j, \Delta\theta_j)$$

- Nothing wrong in having the **POI(s) affecting** also the ***shape*** of the signal prediction (i.e. relative bin content):

$$\mu * S_i(\theta) \rightarrow S_i(\mu, \theta)$$

CMS-PAS-TOP-17-001



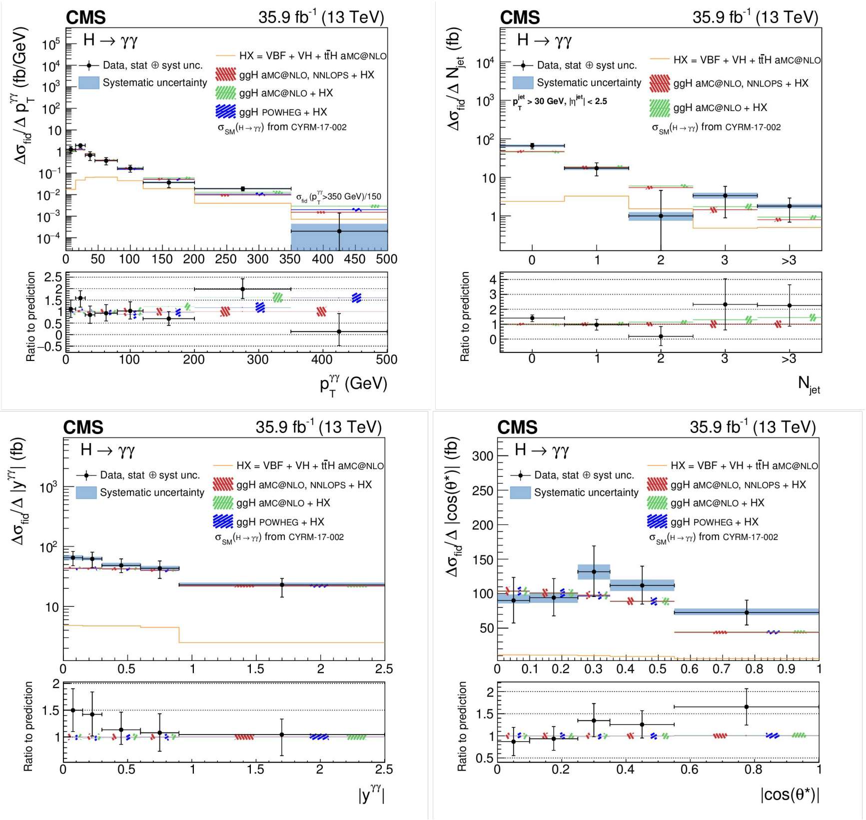
- This allows to perform **other kinds of measurements with PLF**:
  - *e.g. top mass, top width*

- Just ***technical problem*** of how to **interpolate** between different histogram templates for different values of the POI
  - not trivial if want to keep using existing tools to produce model (HistFactory)

Profile likelihood fit and unfolding

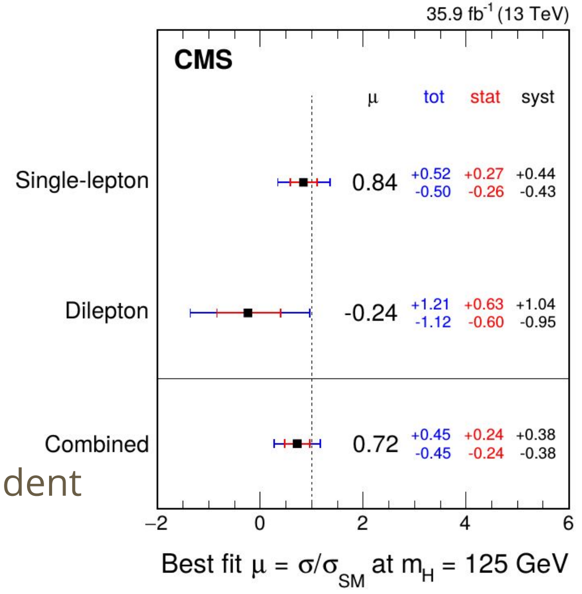
- ATLAS and CMS not used to combine **profiling** and **unfolding**:
  - unfolding procedures ***not compatible*** with PLF (e.g. *Bayesian unfolding*)
  - FBU does something similar to profiling (see ATLAS tt charge asymmetry)
- Conceptually no issue in unfolding with PLF:
  - fit = find values differential x-section that maximise the likelihood → "**Maximum Likelihood unfolding**"
- Already applied in CMS Higgs:
  - CMS H→γγ unfolds by fitting signal strength in different bins:

$$\mathcal{L}(\text{data}|\Delta\vec{\sigma}^{\text{fid}}, \vec{n}_{\text{bkg}}, \vec{\theta}_{\text{S}}, \vec{\theta}_{\text{B}}) = \prod_{i=1}^{n_{\text{cat}}}\prod_{j=1}^{n_{\text{b}}}\prod_{l=1}^{n_{m_{\gamma\gamma}}} \left(\frac{\sum_{k=1}^{n_b}\Delta\sigma_k^{\text{fid}} K_k^{ij}(\vec{\theta}_{\text{S}}) S_k^{ij}(m^l_{\gamma\gamma}|\vec{\theta}_{\text{S}}) L + n^{ij}_{\text{OOA}} S^{ij}_{\text{OOA}}(m^l_{\gamma\gamma}|\vec{\theta}_{\text{S}}) + n^{ij}_{\text{bkg}}B^{ij}(m^l_{\gamma\gamma}|\vec{\theta}_{\text{B}})}{n^{ij}_{\text{sig}}+n^{ij}_{\text{bkg}}}\right)^{n^{lij}_{\text{ev}}} \text{Pois}(n^{ij}_{\text{ev}}|n^{ij}_{\text{sig}}+n^{ij}_{\text{bkg}})\text{Pdf}(\vec{\theta}_{\text{S}})\text{Pdf}(\vec{\theta}_{\text{B}}),$$



Combination of measurements

- With the PLR approach, **combination** of different measurements is **natural**:
  - "just" ***add some more bins*** to the product
- However, important to consider **compatibility of models**:
  - **orthogonality** of channels:
    - bin contents in PLR supposed to be statistically independent
  - **same** definition of (set of) **POI**:
    - sometimes obvious, but not always
      (is μ applied to all the ttH, or just one decay channel? What about tH? ...)
  - **compatible** set of **systematics**:
    - most tricky part, especially for **ATLAS+CMS combinations**!
    - mainly dealing with the question "***which NPs are correlated between channels?***"
    - often cannot reach perfect solution, need to **test different correlation assumptions**
      *(notice that in PLR formalism systematics are either **fully correlated** or **fully uncorrelated**, even if this can be circumvented by splitting a nuisance into two components, and make one correlated, the other uncorrelated)*



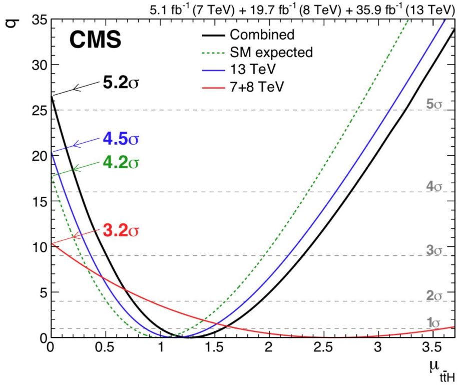
arXiv:1804.03682

Sanity check-list for profile likelihood fits

- Is your systematic model **complete and granular enough**?
    - are all relevant uncertainties, in relationship to the ***observables*** used, in place?
- Perform a fit on **Asimov data** and check all the nuisance parameter constraints:
    - are all the **constraints understood**?
        - e.g. is it **reasonable** to expect **improvements** on this systematic using top events?
        - pay even more attention to “***top ranked***” systematics
- Perform fit on **data** and check nuisance parameter constraints and pulls:
    - do you see the **same constraints** in Asimov and in data?
    - are the **pulls understood**?
        - is it reasonable for these NPs to **compensate for data/MC** disagreement?

- Special attention to **theory systematics**:
    - ask theorists what are the **limits** of their calculations
    - be careful if you seem to **constrain models** (CR, UE) **without a dedicated observable**
        - *and when adding dedicated observables: are you probing the same side of the coin?*
    - **2-point** systematics as well as **scale** uncertainties are delicate *(as shown in previous slides)*

Phys. Rev. Lett. 120 (2018) 231801



Conclusions

- Profile likelihood fit is a **powerful tool**, that could be used in ***any*** ATLAS or CMS top physics analysis
- Like many other nice toys in our field, it **cannot be used as a black box**:
  - **understanding** of the underlying concepts is needed
  - **pulls and constraints** of nuisance parameters have to be always scrutinised
- Important message:
  - in order to be used in a PLF analysis (*as for precision measurements performed with other tools*), the **set of systematic uncertainties** has to be ***complete*** and ***decomposed*** in all its independent sources
    - this is especially **challenging for theory / MC uncertainties**, where **close collaboration** between theory and experiment communities is needed

Backup

Splitting systematic uncertainties

- Usual answer from stats-gurus to the question "*what to do if worried by (over)constraints?*"
  - **redesign your analysis** in order to be less affected by systematics you don't want to constrain
  - or, **redesign you systematic model**,
    e.g. by ***splitting*** systematic uncertainties into more independent components

When possible, always consider splitting important and/or tightly constrained uncertainties, e.g.:

- Jet energy scale, b-tagging, … → use all the O(20-100) sources provided as output of the calibrations
- ME scale → vary μ_R and μ_F separately,
  consider $\mu_R+\mu_F$ in addition for shape analyses?
- PS scale → 7-point scale variation suggested by Peter Skands (presentation at CMS TOP workshop 2018)
  - \+ variations for non-singular terms
  - \+ flavor-dependent variations where relevant (m_t)?
- …