Thoughts on documenting profile likelihood fits in publications

Georges Aad

CPPM, Aix-Marseille Université, CNRS/IN2P3, Marseille, France

Higgs Toppings Workshop
Probing Top-Higgs Interactions at the LHC
May 2018, Benasque, Spain

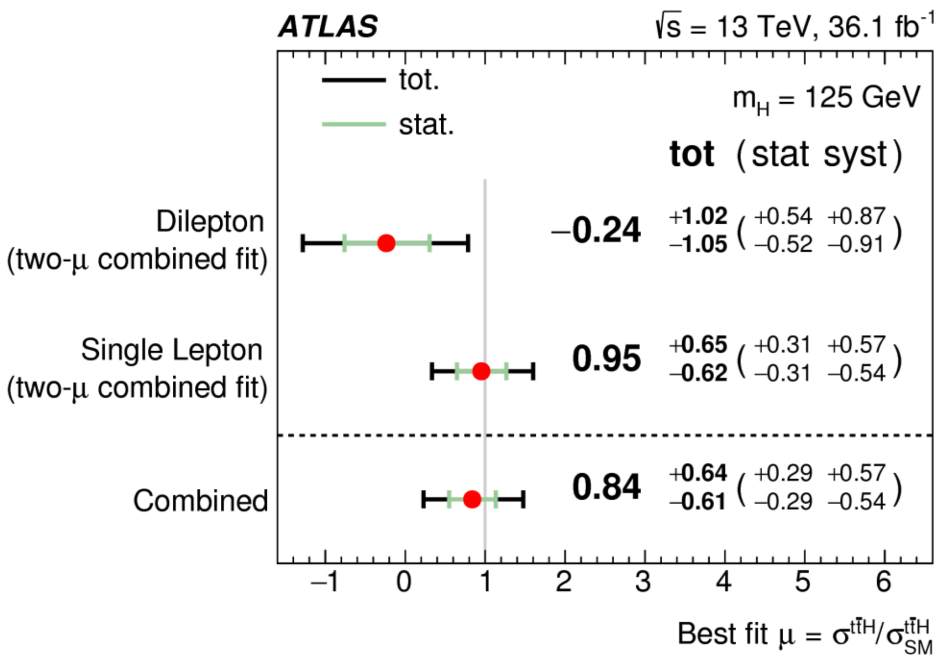
Subjects for Discussion

1. How/what to publish concerning the signal
2. Material needed for a minimal understanding of the fit model and the results
3. How to interpret results concerning the backgrounds

Lets Start by Being Consistent

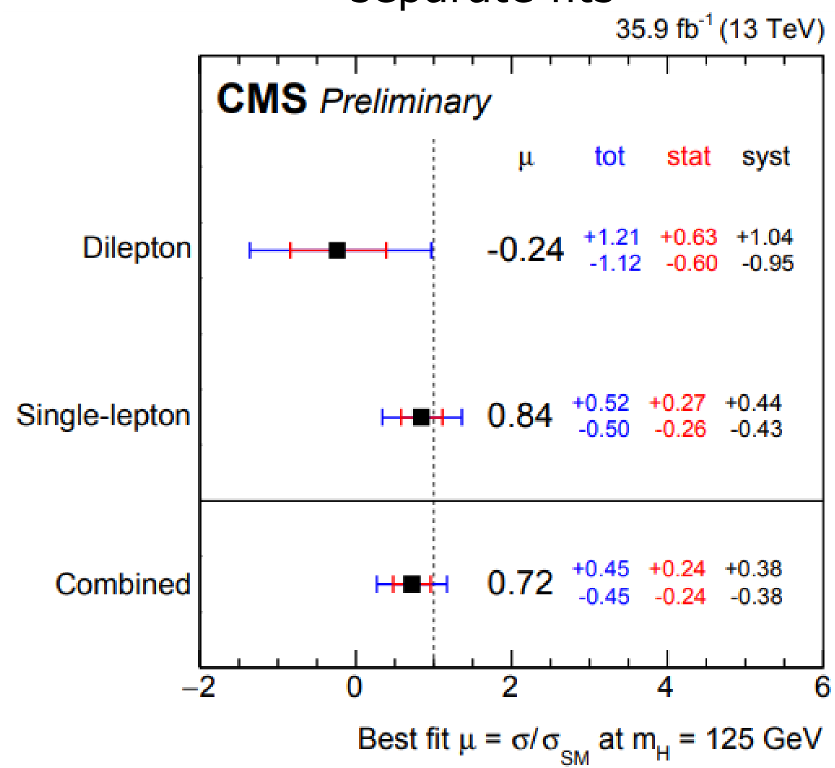
ATLAS

Individual channel mu from combined fit
Separate fit results also in the paper



CMS

Individual channel mu from separate fits



- Interesting to have both in the publications
  - Especially when combining many channels (helps assessing compatibility)
- We can argue what to use for the main results

mu or cross-section

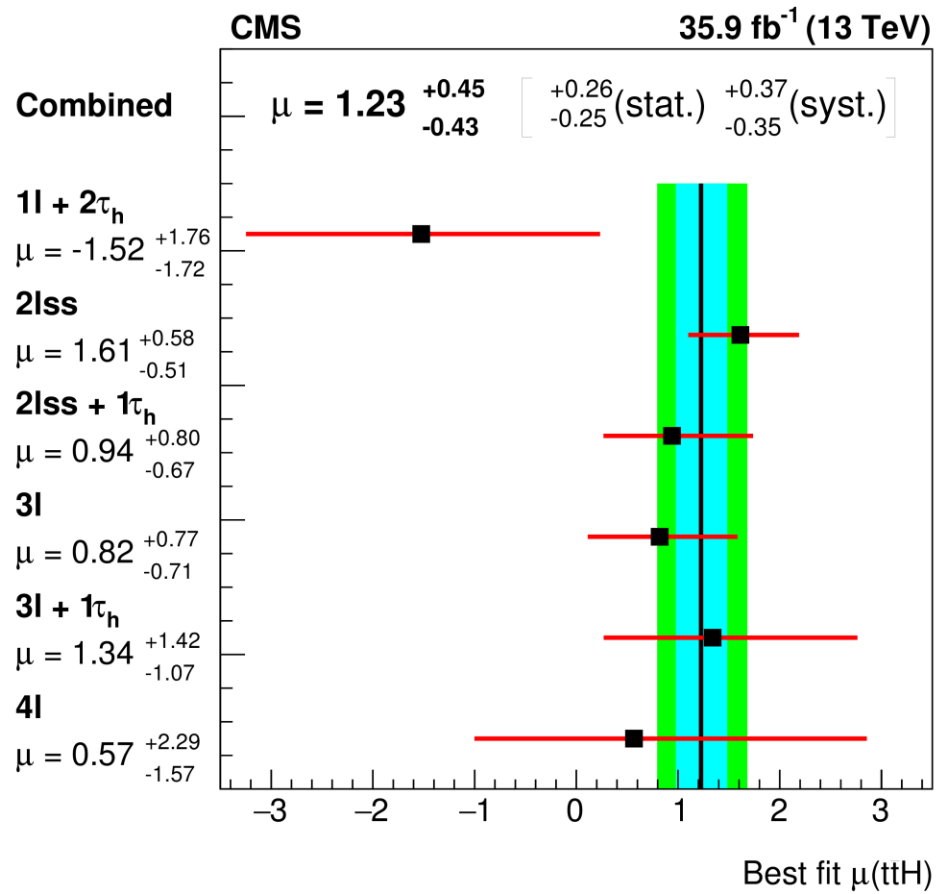
- Publishing mu or signal cross-section is (almost) equivalent
  - Assuming proper systematics are added on the signal acceptance
  - Assuming MC is good enough to extrapolate to the full phase space (like any other measurement anyway)
  - These assumptions become more important as the signal uncertainties become dominant
- Mu allows to directly see possible deviations from the SM
- Cross section does not depend on reference ttH MC
  - Theory cross-section uncertainties removed from measured cross-section
  - Separate values for different center-of-mass energy

ATLAS:

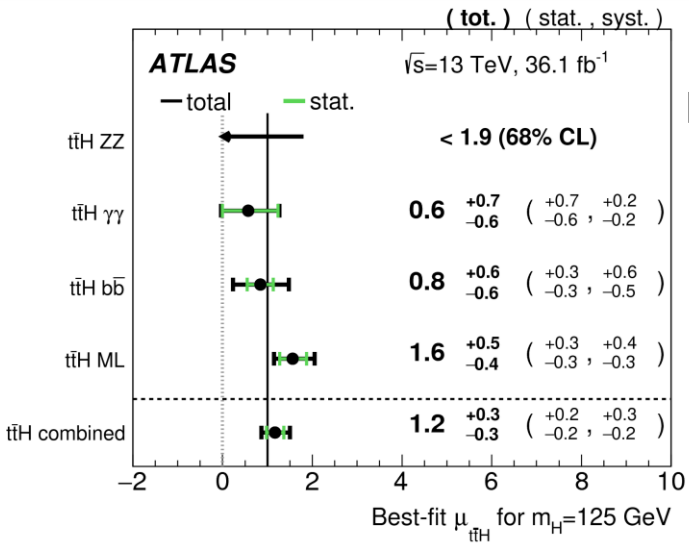
$\mu = 1.17 \pm 0.19\,(\text{stat.})\,^{+0.27}_{-0.23}\,(\text{syst.}).$ → $\sigma(ttH) = 590^{+160}_{-150}\,fb$

Separate mu for Different Final States

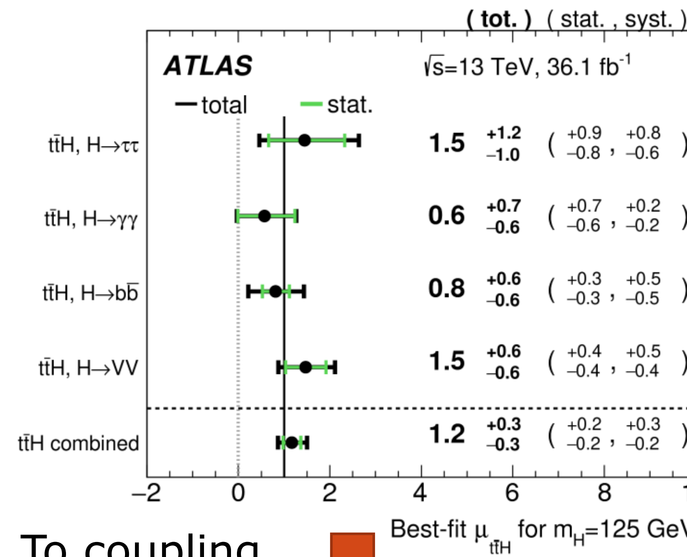
- Currently the standard way of presenting results
- Very important to continue to provide this information
- Important to quantify the compatibility between the channels
- Can hint to new physics that is totally independent of ttH
  - Especially for channels with no reconstructed Higgs



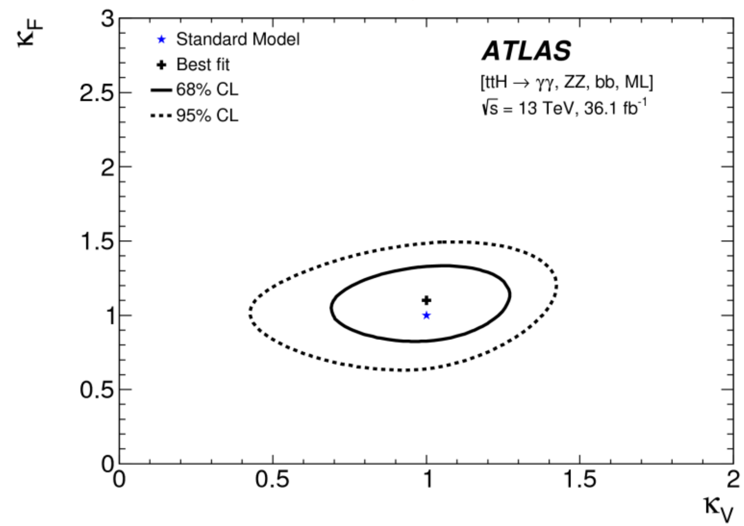
Additional Interpretations of the Results



From analysis channels to decay modes



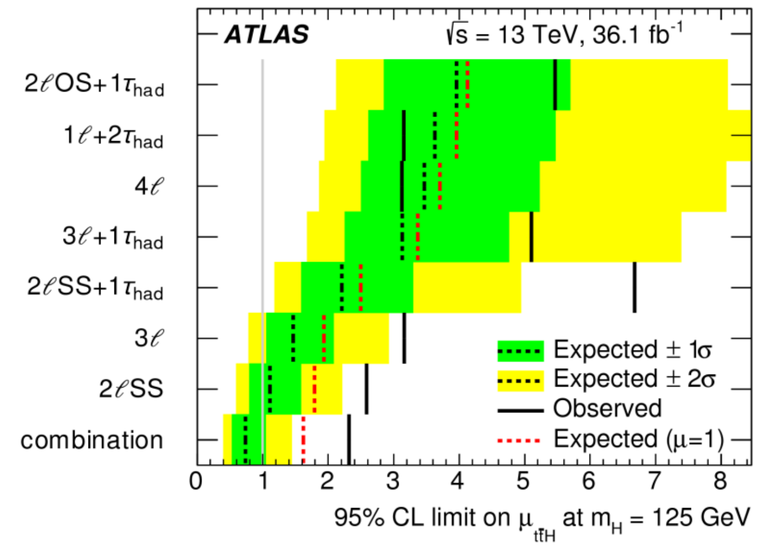
To coupling modifiers



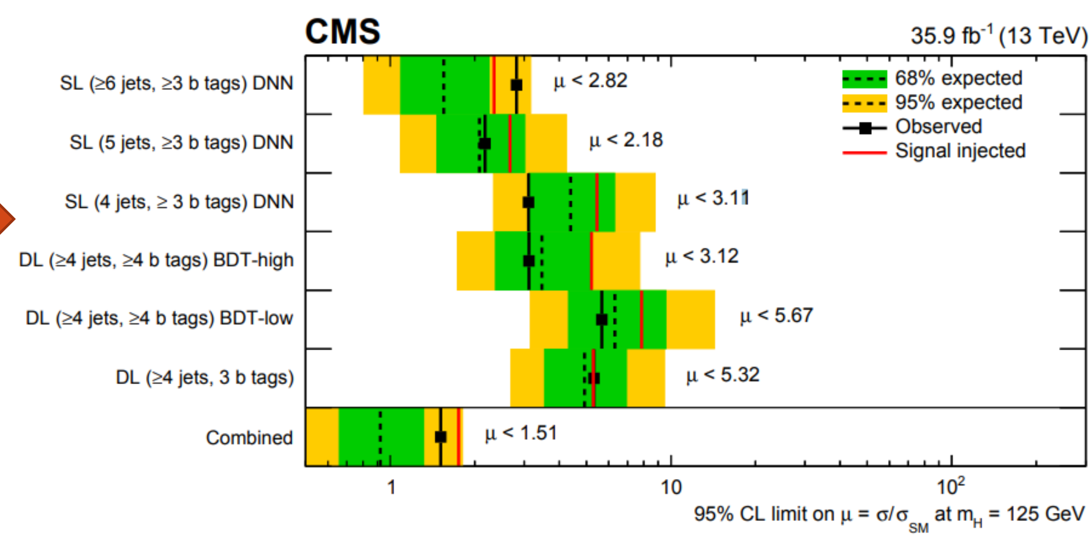
- Physics interpretations become more important with increased sensitivity
- How useful is it to assess different Higgs decay modes and coupling modifiers for the ttH channel alone?
- Is there other useful interpretations?

We are Still Publishing Limits

- Publishing a limit on a signal that is observed is weird
- Is this really useful for anything after the discovery?



This plot is actually nice to assess the power of each analysis category
But why not using mu?



Subjects for Discussion

1. How/what to publish concerning the signal

2. Material needed for a minimal understanding of the fit model and the results

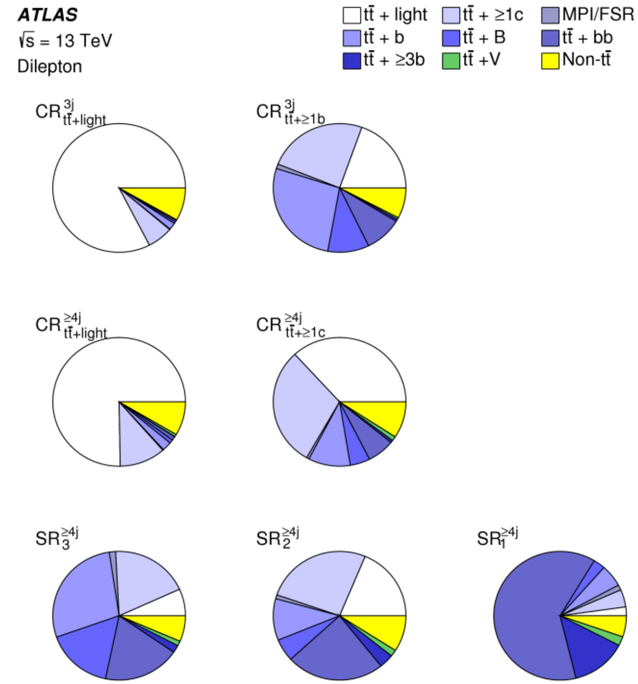
3. How to interpret results concerning the backgrounds

Critics/Feedback are Mandatory

- Very complex analysis for many ttH channels
  - Combination of multiple final states
  - Complex strategies with various MVA techniques
  - Large backgrounds with important systematics
  - Small and very sensitive signal
  - Huge profile likelihood fit: hundreds of bins and nuisance parameters
  - Highly non-trivial to understand and compare the results
- Mandatory information for a minimal understanding
  - Background composition (pre/post fit)
  - Size and impact of important systematics
  - Validation procedures for the modeling and robustness of the fit
  - ...?
- This information can be communicated in various ways
- Simplest and efficient way: sharing the workspace
  - Why not?
  - We are going to do it at some point for the combination

Background Composition pre/post fits

- Standard: most analysis provide the yields in various regions pre- and post-fit
- Mandatory to understand the main background components
- Mandatory to see the impact of the fit on individual backgrounds



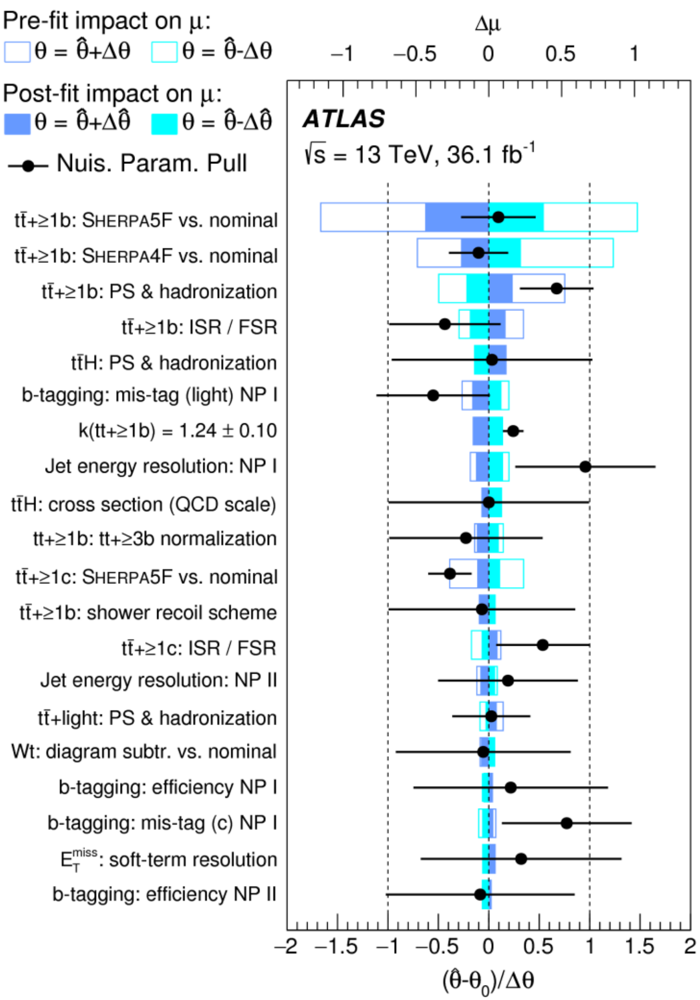
Process	$t\bar{t}H$ node		$t\bar{t}+b\bar{b}$ node		$t\bar{t}$+2b node		$t\bar{t}$+b node		$t\bar{t}+c\bar{c}$ node		$t\bar{t}$+lf node	
	pre-fit (post-fit) yields											
$t\bar{t}$+lf	1982	(1381)	1280	(897)	852	(595)	916	(661)	243	(172)	50	(36)
$t\bar{t}+c\bar{c}$	1150	(1415)	998	(1230)	636	(805)	444	(567)	115	(147)	16	(19)
$t\bar{t}$+b	549	(705)	575	(746)	314	(409)	253	(338)	28	(35)	4	(5)
$t\bar{t}$+2b	306	(233)	282	(215)	372	(293)	78	(62)	10	(8)	1	(0.8)
$t\bar{t}+b\bar{b}$	834	(769)	1156	(1082)	299	(266)	145	(129)	17	(15)	3	(2)
Single t	110	(116)	146	(145)	92	(82)	53	(53)	4	(4)	3	(3)
V+jets	38	(37)	78	(76)	34	(30)	10	(9)	7	(6)	0.6	(0.6)
$t\bar{t}$+V	80	(75)	58	(54)	31	(28)	11	(11)	4	(4)	0.4	(0.4)
Diboson	0.9	(0.9)	0.5	(0.5)	0.4	(0.4)	0.4	(0.4)	—	(—)	—	(—)
Total bkg.	5049	(4733)	4575	(4447)	2629	(2509)	1911	(1831)	429	(392)	77	(67)
± tot unc.	±1216	(±186)	±1156	(±142)	±603	(±80)	±422	(±65)	±107	(±14)	±18	(±3)
$t\bar{t}H$	142	(108)	53	(40)	24	(18)	10	(7)	2.1	(1.5)	0.30	(0.23)
± tot unc.	±19	(±15)	±8	(±6)	±3	(±2)	±1	(±1)	±0.2	(±0.2)	±0.03	(±0.03)
Data	4822		4400		2484		1852		422		76	

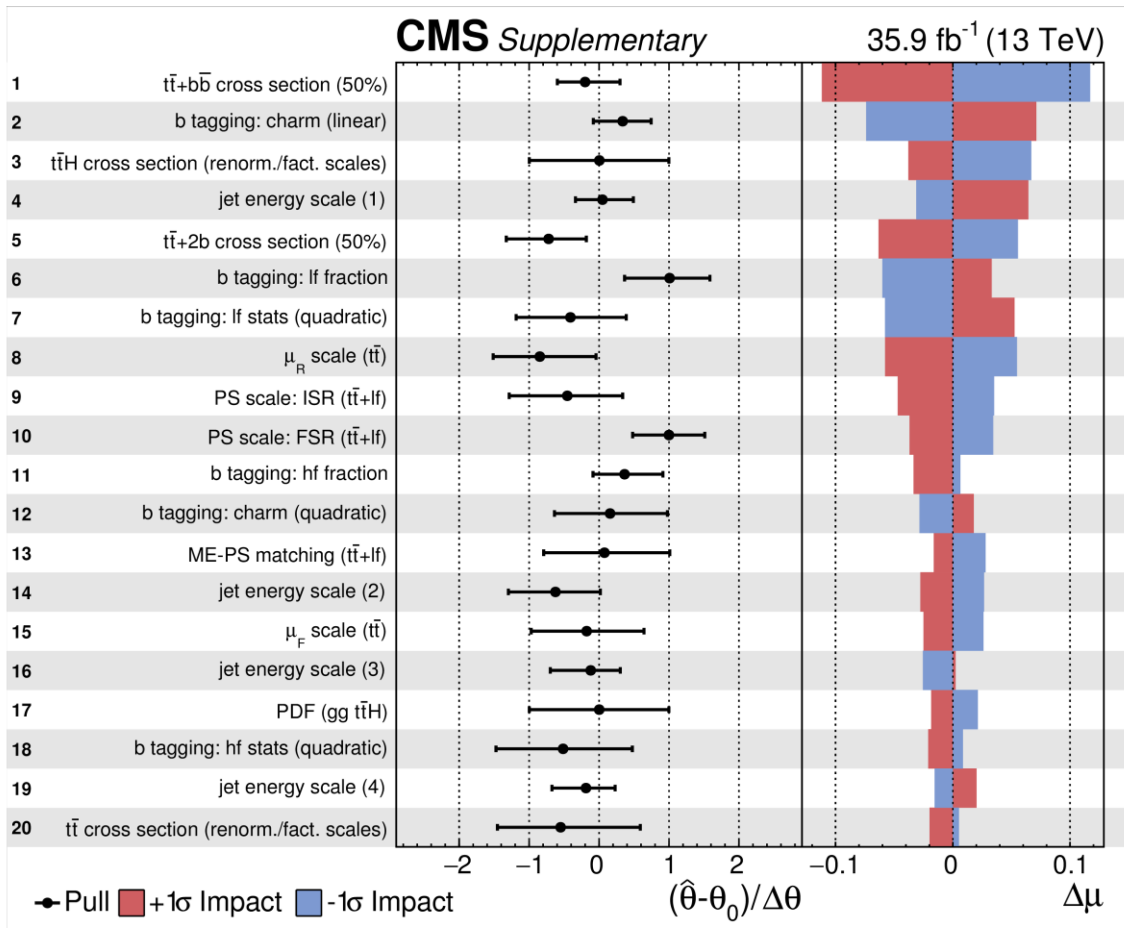
Pre/post fit ratio does not necessary represents cross sections changes

Highly affected by acceptance

Size and Impact of Important Systematics

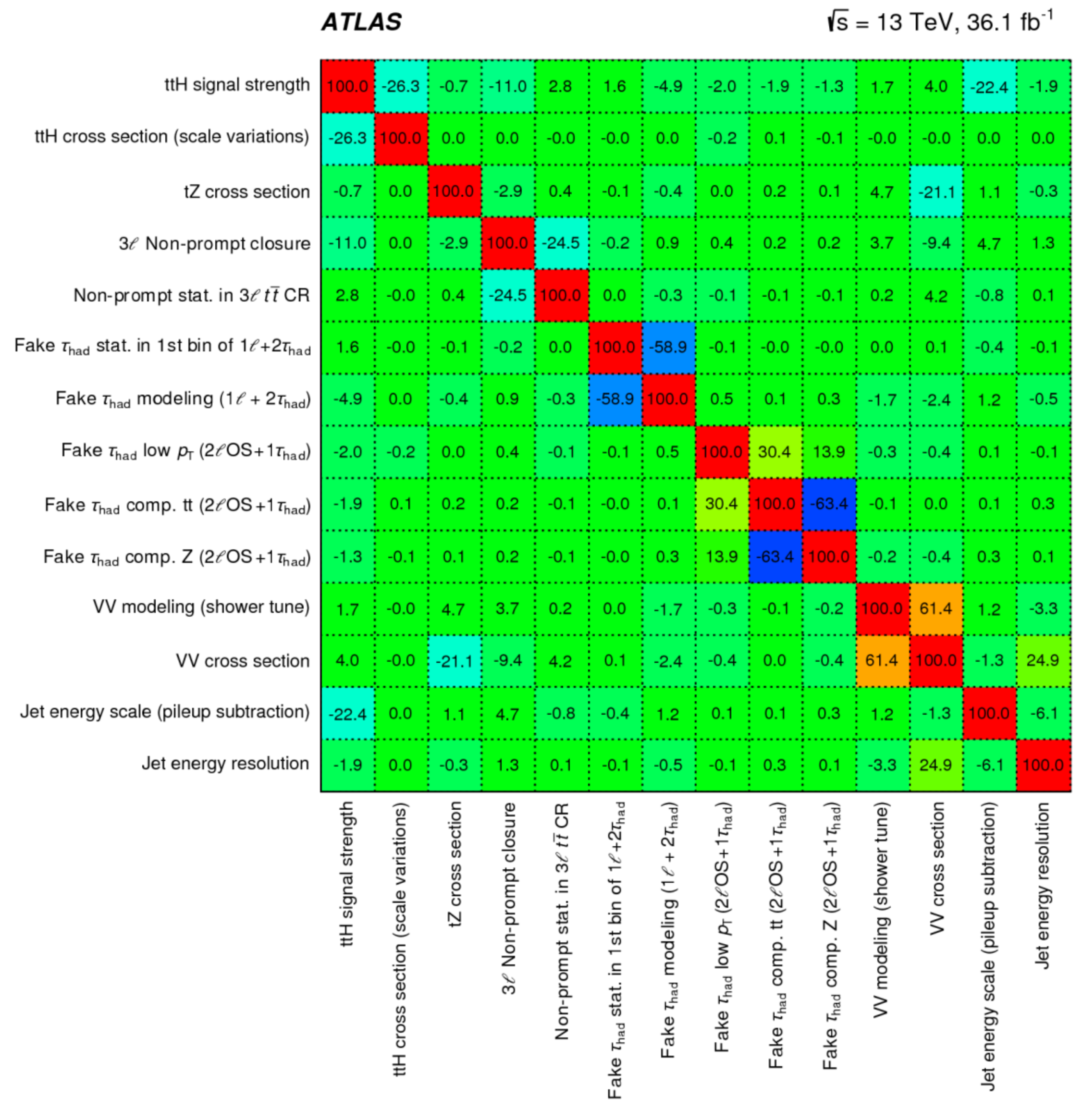
- Is there a need to motivate this one?
- Should we provide the norm parameters for important backgrounds in addition?





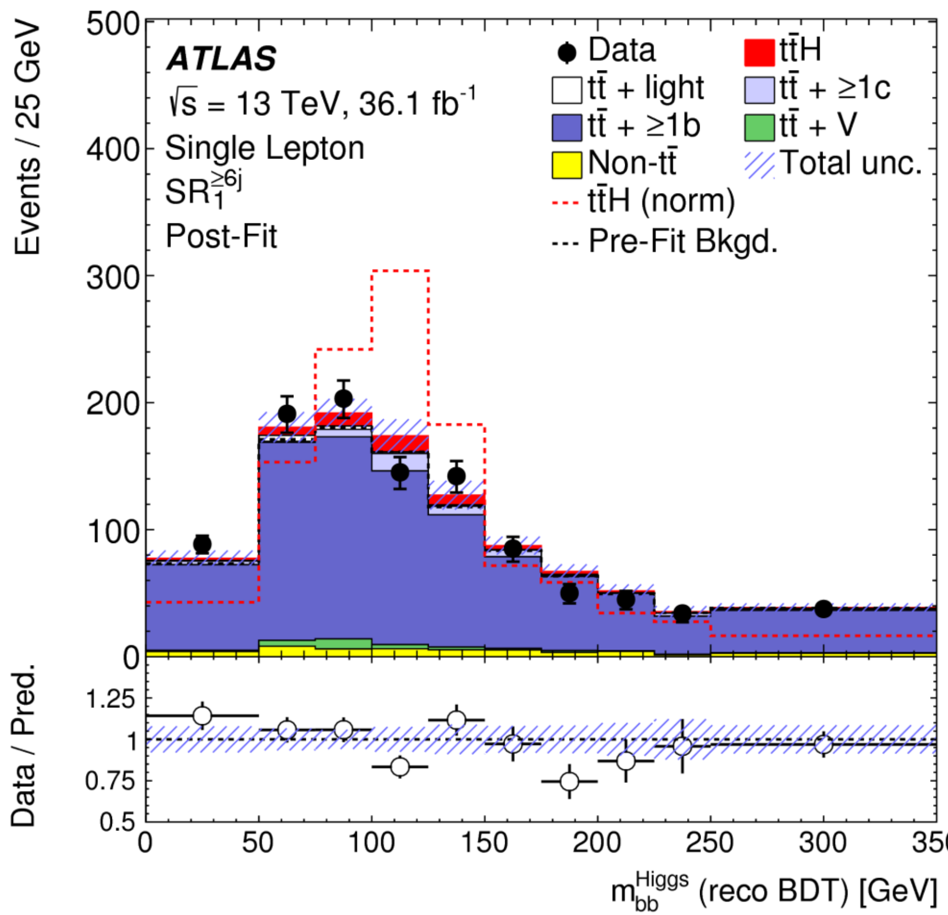
Correlations

Although hard to interpret, large correlations can carry very important information about the fit



Post-fit Plots of Key variables

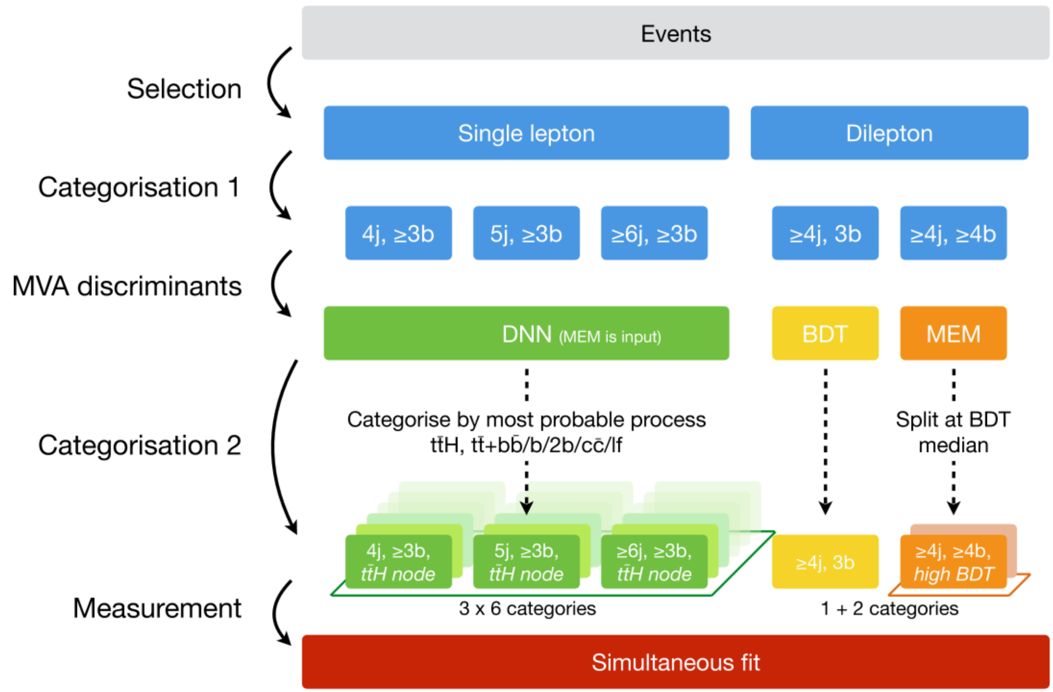
- Important to check input variables to the MVA post-fit
  - Using the final fit (with MVA)
- Also some key variables are always nice to see
  - E.g. m(b,b) in ttH(bb)
- Very long list of plots
  - Regions x Variables
- An unbiased subset as additional material would be very nice to have



Documenting Analysis Procedure

- Complex strategy with many layers
- Not always clear what drives the sensitivity
  - Even for the analyzers themselves
- Important to give hints on what actually matters
- Do we need to go further and quantify the impact of various components
  - Not always feasible
  - Not always useful

Nice plot CMS



- How to document this in less than 100 pages?
  - Good luck for the editors !!

Documenting Syst Model

- Complex systematics models
  - Hard to understand the exact procedure
  - Not always trivial to make the correspondence with ranking plots
- How to explain this clearer in less than 100 pages?
- Minimal requirement
  - Clarify and explain what matters

Source	Type	Remarks
Integrated luminosity	rate	Signal and all backgrounds
Lepton identification/isolation	shape	Signal and all backgrounds
Trigger efficiency	shape	Signal and all backgrounds
Pileup	shape	Signal and all backgrounds
Jet energy scale	shape	Signal and all backgrounds
Jet energy resolution	shape	Signal and all backgrounds
b tag hf fraction	shape	Signal and all backgrounds
b tag hf stats (linear)	shape	Signal and all backgrounds
b tag hf stats (quadratic)	shape	Signal and all backgrounds
b tag lf fraction	shape	Signal and all backgrounds
b tag lf stats (linear)	shape	Signal and all backgrounds
b tag lf stats (quadratic)	shape	Signal and all backgrounds
b tag charm (linear)	shape	Signal and all backgrounds
b tag charm (quadratic)	shape	Signal and all backgrounds
Renorm./fact. scales ($t\bar{t}$H)	rate	Scale uncertainty of NLO $t\bar{t}$H prediction
Renorm./fact. scales ($t\bar{t}$)	rate	Scale uncertainty of NLO $t\bar{t}$ prediction
Renorm./fact. scales ($t\bar{t}$+hf)	rate	Additional 50% rate uncertainty of $t\bar{t}$+hf predictions
Renorm./fact. scales (t)	rate	Scale uncertainty of NLO single t prediction
Renorm./fact. scales (V)	rate	Scale uncertainty of NNLO W and Z prediction
Renorm./fact. scales (VV)	rate	Scale uncertainty of NLO diboson prediction
PDF (gg)	rate	PDF uncertainty for gg initiated processes except $t\bar{t}$H
PDF (gg $t\bar{t}$H)	rate	PDF uncertainty for $t\bar{t}$H
PDF ($q\bar{q}$)	rate	PDF uncertainty of $q\bar{q}$ initiated processes ($t\bar{t}$+W,W,Z)
PDF (qg)	rate	PDF uncertainty of qg initiated processes (single t)
	pe	Renormalisation scale uncertainty of the $t\bar{t}$ ME generator, independent for additional jet flavours
	pe	Factorisation scale uncertainty of the $t\bar{t}$ ME generator, independent for additional jet flavours
	te	Initial state radiation uncertainty of the PS (for $t\bar{t}$ events), jet multiplicity dependent rate uncertainty, independent for additional jet flavours
	te	Final state radiation uncertainty (for $t\bar{t}$ events), jet multiplicity dependent rate uncertainty, independent for additional jet flavours
	te	NLO ME to PS matching, *hdamp* [?] (for $t\bar{t}$ events), jet multiplicity dependent rate uncertainty, independent for additional jet flavours
	te	Underlying event (for $t\bar{t}$ events), jet multiplicity dependent rate uncertainty, independent for additional jet flavours
	pe	Based on the NNPDF replicas, same for $t\bar{t}$H and additional jet flavours
	pe	Statistical uncertainty of the signal and background prediction due to the limited sample size

Systematic source	Description	$t\bar{t}$ categories
$t\bar{t}$ cross-section	Up or down by 6%	All, correlated
$k(t\bar{t}+\geq 1c)$	Free-floating $t\bar{t}+\geq 1c$ normalization	$t\bar{t}+\geq 1c$
$k(t\bar{t}+\geq 1b)$	Free-floating $t\bar{t}+\geq 1b$ normalization	$t\bar{t}+\geq 1b$
SHERPA5F vs. nominal	Related to the choice of NLO event generator	All, uncorrelated
PS & hadronization	POWHEG+HERWIG 7 vs. POWHEG+PYTHIA 8	All, uncorrelated
ISR / FSR	Variations of $\mu_{\rm R}$, $\mu_{\rm F}$, $h_{\rm damp}$ and A14 Var3c parameters	All, uncorrelated
$t\bar{t}+\geq 1c$ ME vs. inclusive	MG5_aMC@NLO+HERWIG++: ME prediction (3F) vs. incl. (5F)	$t\bar{t}+\geq 1c$
$t\bar{t}+\geq 1b$ SHERPA4F vs. nominal	Comparison of $t\bar{t}+b\bar{b}$ NLO (4F) vs. POWHEG+PYTHIA 8 (5F)	$t\bar{t}+\geq 1b$
$t\bar{t}+\geq 1b$ renorm. scale	Up or down by a factor of two	$t\bar{t}+\geq 1b$
$t\bar{t}+\geq 1b$ resumm. scale	Vary $\mu_{\rm Q}$ from $H_{\rm T}/2$ to $\mu_{\rm CMMPS}$	$t\bar{t}+\geq 1b$
$t\bar{t}+\geq 1b$ global scales	Set $\mu_{\rm Q}$, $\mu_{\rm R}$, and $\mu_{\rm F}$ to $\mu_{\rm CMMPS}$	$t\bar{t}+\geq 1b$
$t\bar{t}+\geq 1b$ shower recoil scheme	Alternative model scheme	$t\bar{t}+\geq 1b$
$t\bar{t}+\geq 1b$ PDF (MSTW)	MSTW vs. CT10	$t\bar{t}+\geq 1b$
$t\bar{t}+\geq 1b$ PDF (NNPDF)	NNPDF vs. CT10	$t\bar{t}+\geq 1b$
$t\bar{t}+\geq 1b$ UE	Alternative set of tuned parameters for the underlying event	$t\bar{t}+\geq 1b$
$t\bar{t}+\geq 1b$ MPI	Up or down by 50%	$t\bar{t}+\geq 1b$
$t\bar{t}+\geq 3b$ normalization	Up or down by 50%	$t\bar{t}+\geq 1b$

Fit Validation

- No standardized procedure available
- In general cross-checks are not published or even mentioned
- But cross-checks are mandatory for profile likehood fits with complex systematics model

- How to communicate these checks?
  - Probably impossible to publish the full cross check results
  - But important to clearly explain them
  - and provide a quantitative result when possible
    - Goodness of fit, impact on mu, …

Subjects for Discussion

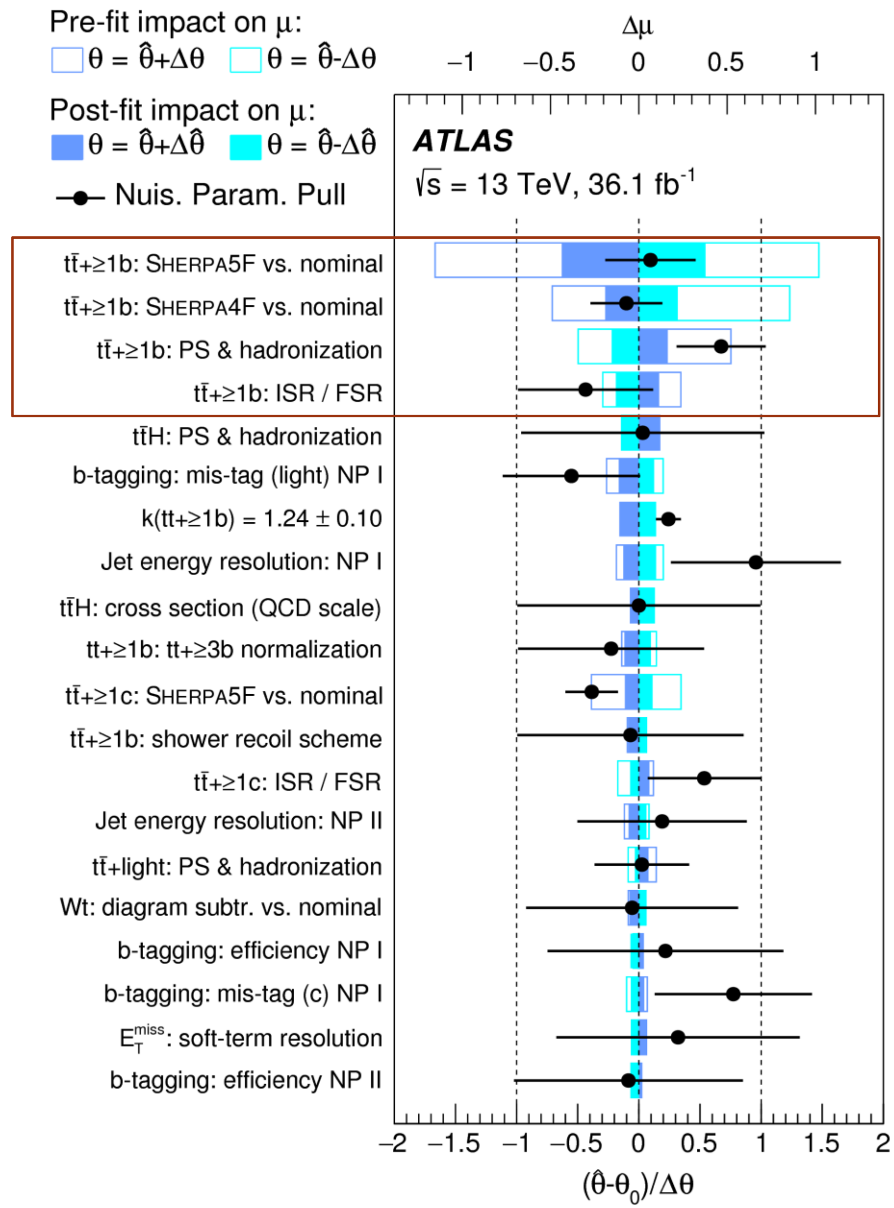
1. How/what to publish concerning the signal

2. Material needed for a minimal understanding of the fit model and the results

3. How to interpret results concerning the backgrounds

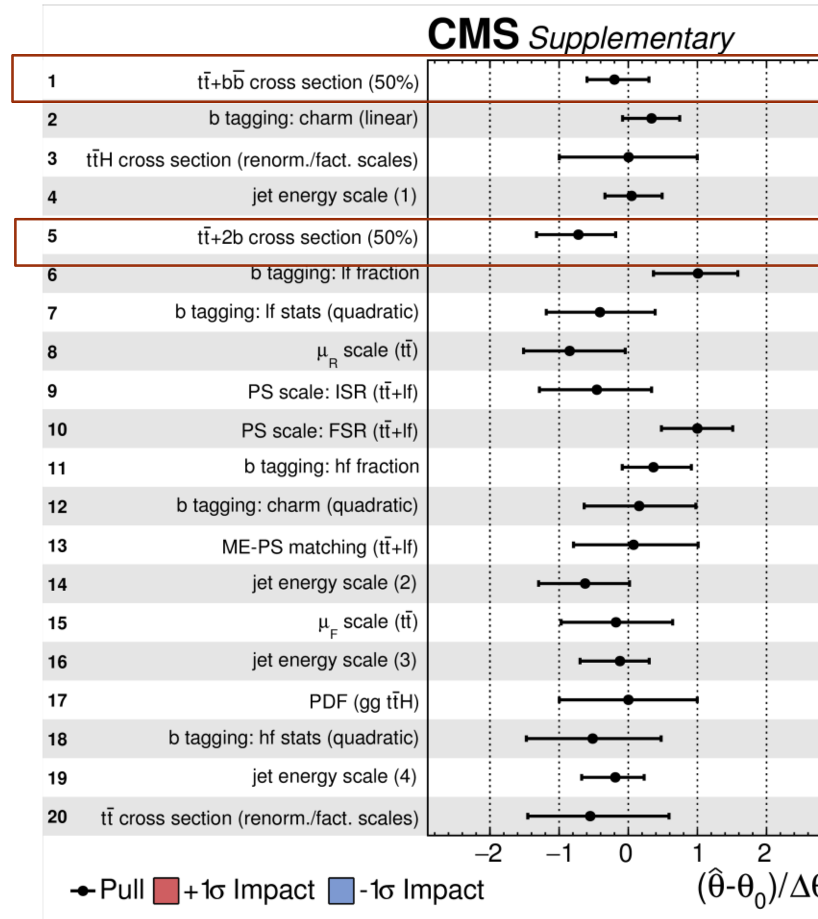
Interpretation of Pulls and Constraints

- Significant pulls and constraints in some nuisance parameters reflecting theory modeling on backgrounds
- Very dangerous to interpret those as measurements
  - Or make exclusion statements about 2-points systematics
  - Highly non trivial fit with complex correlations
  - Use proper unfolded measurements from dedicated analyses
- How clear is this in our papers?



Interpretation of Normalization Factors

- Free normalization factors on background are equivalent to mu
  - One can argue that this is also the case if a large prior is applied
- Should these be used as a measured cross sections?
  - Naively yes but …
  - Do we really want to use MC to extrapolate from the ttH corner of phase space?
  - Are we sure we covered the extrapolation uncertainties?
    - Enough acceptance/shape uncertainties?
- Be careful
  - Should be reflected in papers



ATLAS:

ttb: 1.24 ± 0.10

ttc: 1.63 ± 0.23

ttW: 0.92 ± 0.32

ttZ: 1.17 +0.25 −0.22

Extending the Analyses for Other Signals

- Measuring ttH, ttW and ttZ simultaneously was discussed
- This will need a modification of analysis strategy to increase sensitivity for those backgrounds
- In general this is not the priority but can be done in specific cases
- Why not providing a more general solution
  - We will never cover the wish lists from everyone
- Should the fit results (and whatever needed additional info) go to HEPData?

Conclusion

Don't argue much if you want this discussion to finish fast

Two more talks and you can go hiking …