

Statistical combination of experimental results in ATLAS

Stefan Gadatsch^a *on behalf of the ATLAS collaboration*

2nd September, 2016


^aCERN




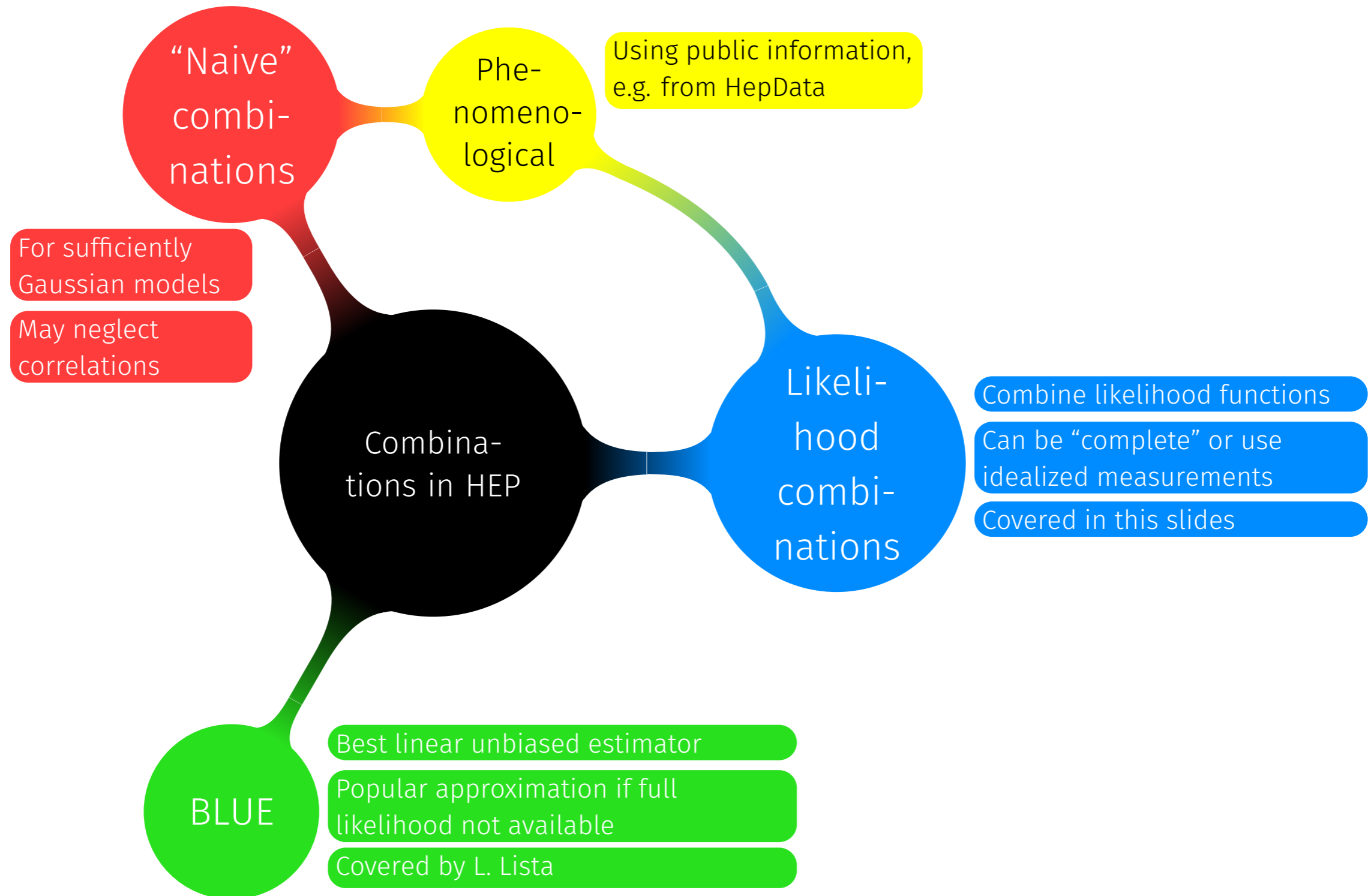
- Combinations are **increasingly important** for HEP to perform the single most powerful test of a hypothesis, e.g. for a **measurement** or assessing the **compatibility** of measurements
- **Dedicated analyses** for each production and decay mode target specific properties
- **Example:** Higgs boson mass and couplings
- Searches are often performed more **inclusively**, complicating statistical combinations
- Many analyses perform combinations already **implicitly**
- **Example:** data driven background estimates, subsidiary measurements, etc.

Channel	References for individual publications		Signal strength [μ] from results in this paper (Section 5.2)		Signal significance [σ]	
	ATLAS	CMS	ATLAS	CMS	ATLAS	CMS
$H \rightarrow \gamma\gamma$	[91]	[92]	1.14 ^{+0.27} _{-0.25} (^{+0.26} _{-0.24})	1.11 ^{+0.25} _{-0.23} (^{+0.23} _{-0.21})	5.0 (4.6)	5.6 (5.1)
$H \rightarrow ZZ$	[93]	[94]	1.52 ^{+0.40} _{-0.34} (^{+0.32} _{-0.27})	1.04 ^{+0.32} _{-0.26} (^{+0.30} _{-0.25})	7.6 (5.6)	7.0 (6.8)
$H \rightarrow WW$	[95, 96]	[97]	1.22 ^{+0.23} _{-0.21} (^{+0.21} _{-0.20})	0.90 ^{+0.23} _{-0.21} (^{+0.23} _{-0.20})	6.8 (5.8)	4.8 (5.6)
$H \rightarrow \tau\tau$	[98]	[99]	1.41 ^{+0.40} _{-0.36} (^{+0.37} _{-0.33})	0.88 ^{+0.30} _{-0.28} (^{+0.31} _{-0.29})	4.4 (3.3)	3.4 (3.7)
$H \rightarrow bb$	[100]	[101]	0.62 ^{+0.37} _{-0.37} (^{+0.39} _{-0.37})	0.81 ^{+0.45} _{-0.43} (^{+0.45} _{-0.43})	1.7 (2.7)	2.0 (2.5)
$H \rightarrow \mu\mu$	[102]	[103]	-0.6 ^{+3.6} _{-3.6} (^{+3.6} _{-3.6})	0.9 ^{+3.6} _{-3.5} (^{+3.3} _{-3.2})		
$t\bar{t}H$ production	[77, 104, 105]	[107]	1.9 ^{+0.8} _{-0.7} (^{+0.7} _{-0.7})	2.9 ^{+1.0} _{-0.9} (^{+0.9} _{-0.8})	2.7 (1.6)	3.6 (1.3)

The work of many people

- This talk will be focussed on **likelihood combinations**
 - Other approaches, e.g. BLUE, will be covered in different talks  L. Lista
- Key concepts will be explained with the Higgs coupling combination as an example
- Statistical inference also covered by other speakers

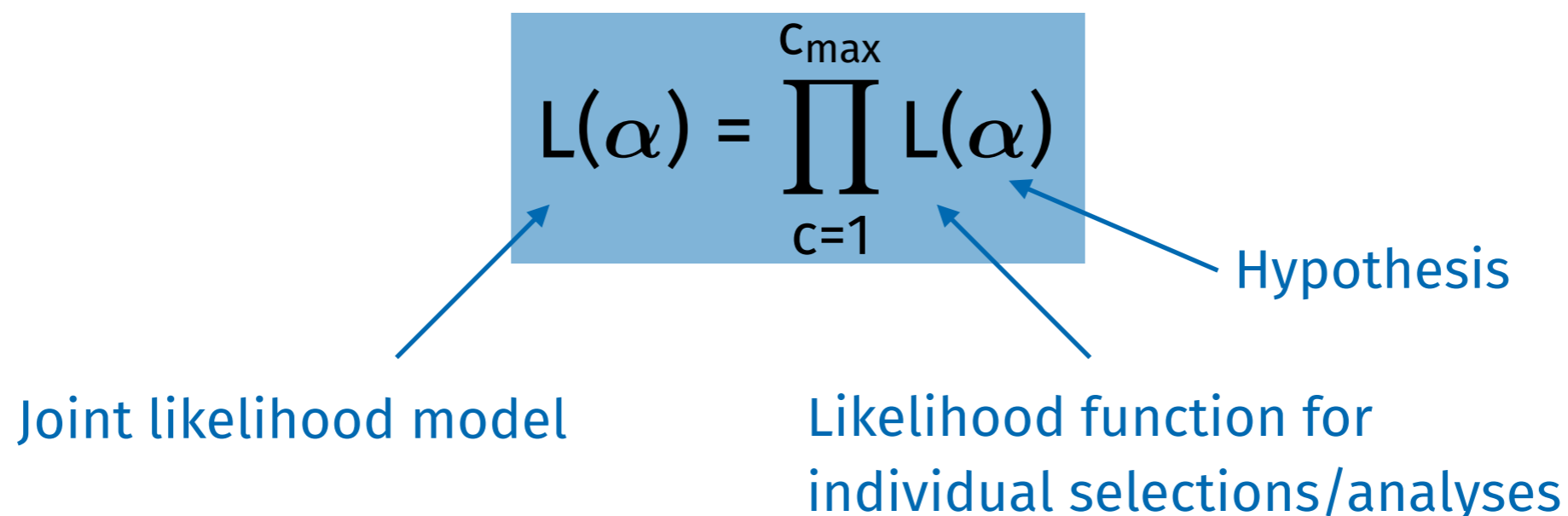
 E. Gross, F. Matorras, S. Schmitt, S. Biondi, ...



- Given a set of **measurements** and a **hypothesis**, a **likelihood function** is defined as the probability of the data under this hypothesis, formulated as a **probability density function**
 - Simplest example: idealised measurement

$$\theta = \tilde{\theta} \pm \sigma \rightarrow \text{Gaussian}(\tilde{\theta} | \theta, \sigma)$$

- Disjoint selections of the data are simultaneously described by a **product over the respective likelihood functions**



The diagram illustrates the joint likelihood model equation: $L(\alpha) = \prod_{c=1}^{C_{\max}} L(\alpha)$. A blue box highlights the equation. Three blue arrows point from labels below to parts of the equation: one from 'Joint likelihood model' to the left-hand side $L(\alpha)$, one from 'Likelihood function for individual selections/analyses' to the individual $L(\alpha)$ term in the product, and one from 'Hypothesis' to the C_{\max} index in the product.

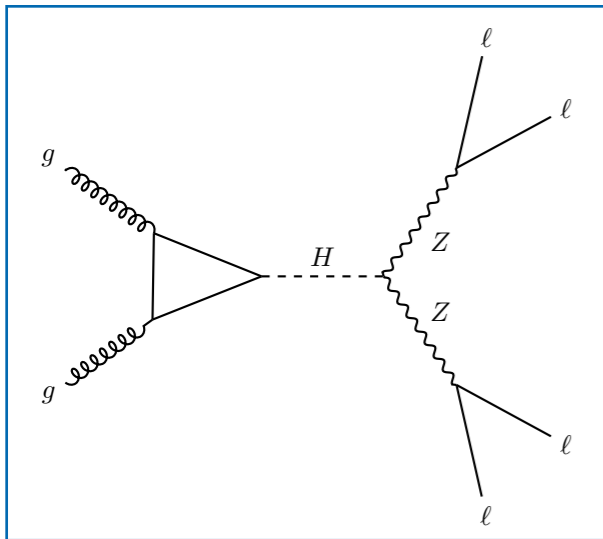
$$L(\alpha) = \prod_{c=1}^{C_{\max}} L(\alpha)$$

Joint likelihood model

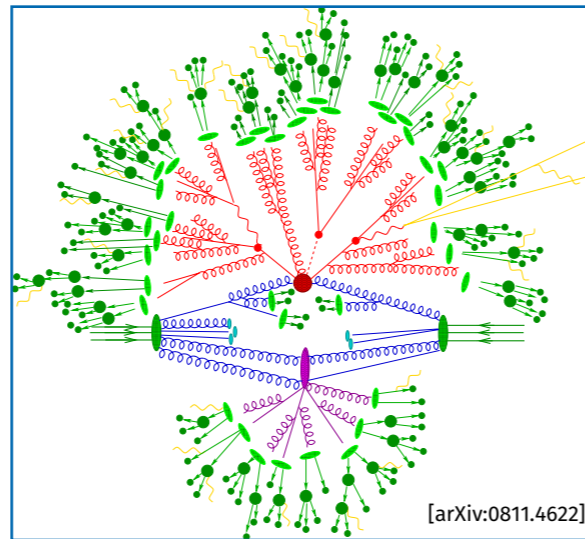
Likelihood function for individual selections/analyses

Hypothesis

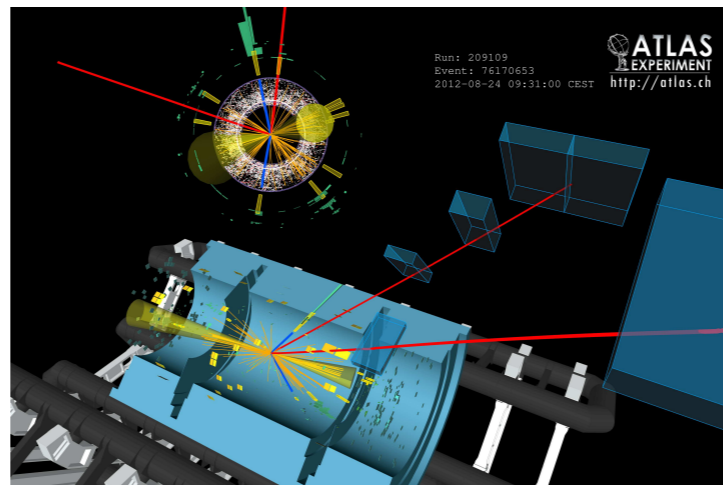
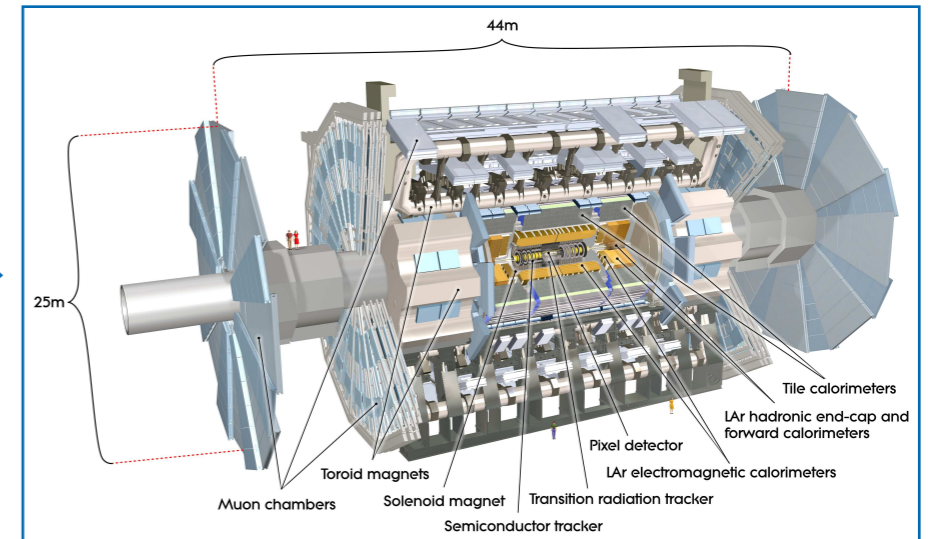
Hard scatter



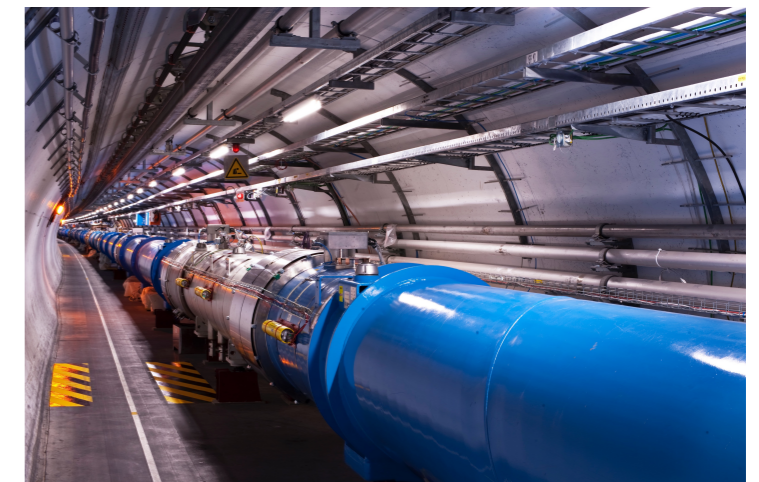
Soft physics



Detector simulation

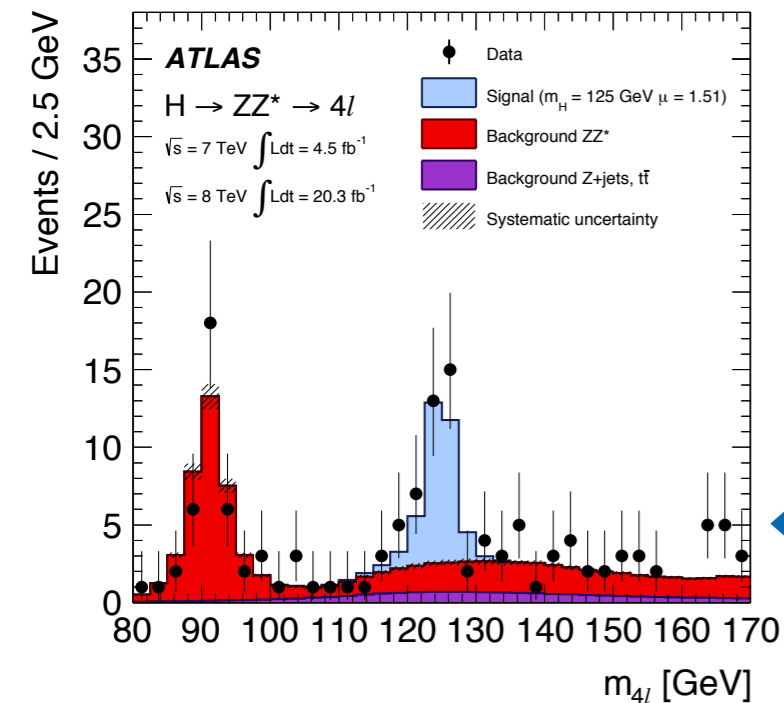


LHC data



Reconstruction

Selection

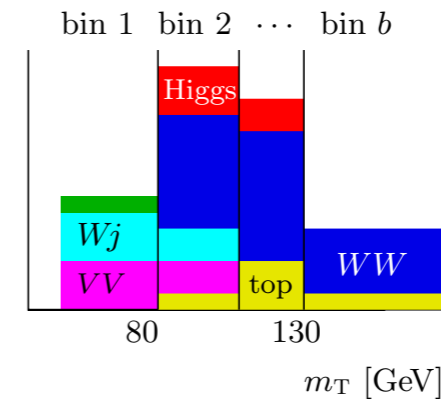


Typical HEP measurements



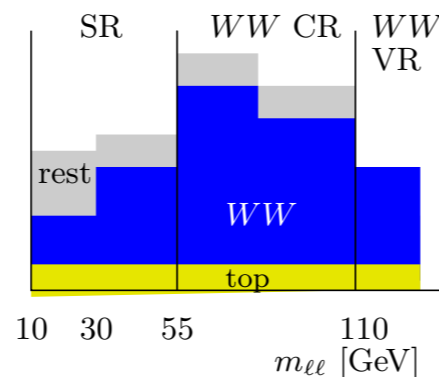
- Typically analyses consist of one or more **disjoint selections** enriched in signal or background like events
- Categories are hardly ever pure, i.e. various processes enter each selection
- The measurement can be a **single number** or the data can be augmented with one or more discriminating observables

(a) Signal region for $n_j = 0, e\mu$ category

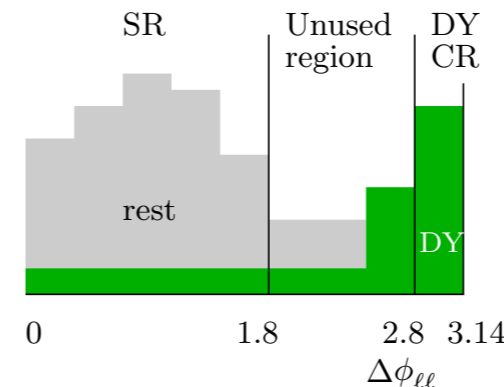


SR shown in (a) has Poisson terms in \mathcal{L}

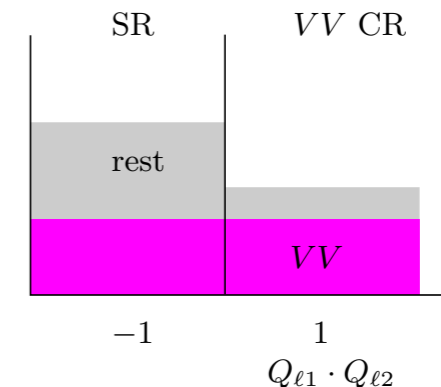
(b) WW
Apply β_{WW} to N_{WW}



(c) Drell-Yan
Apply β_{DY} to N_{DY}

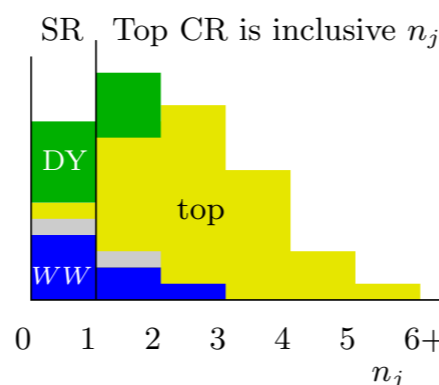


(d) VV
Apply β_{VV} to N_{VV}

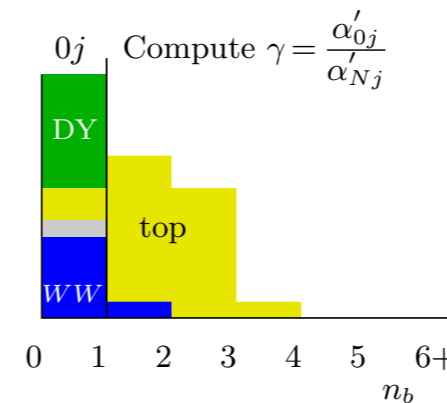


Profiled CRs in (b, c, d) have Poisson terms in \mathcal{L}

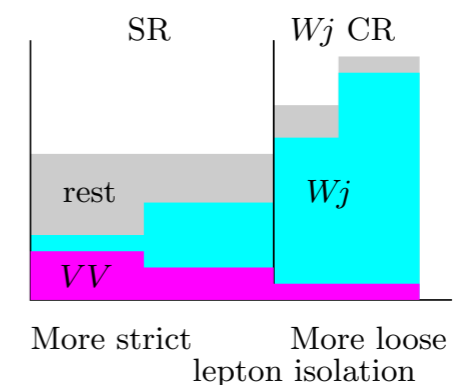
(e) Top quark
Apply β_{top} to N_{top}



(f) $n_b \geq 1$ data
Apply γ^2 to β_{top}



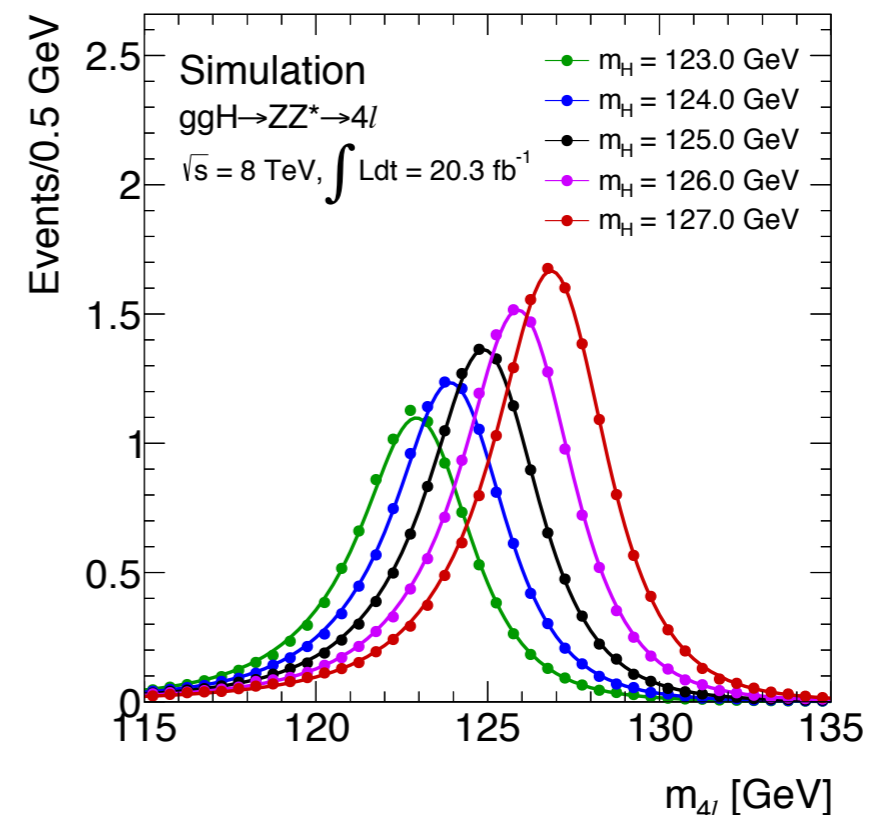
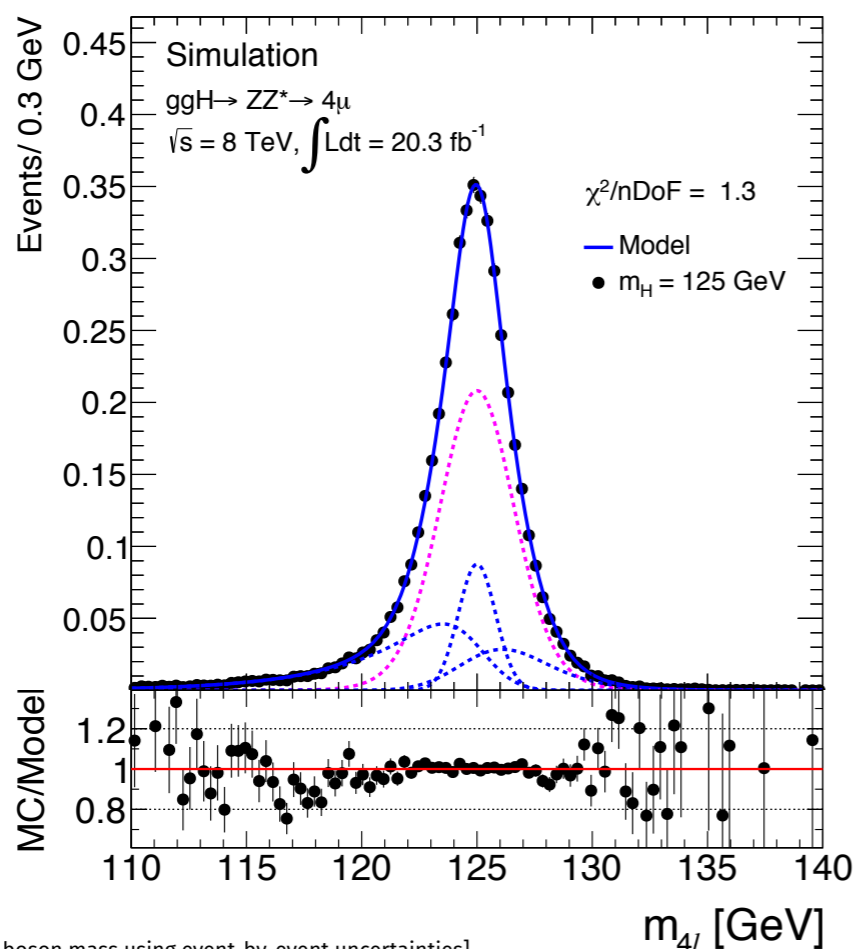
(g) Wj
 N_{Wj} in bins b



Regions (a-d) in fit (e-g) **not** in fit

Nonprofiled CRs in (e, f, g) have **no** Poisson term in \mathcal{L}

- In rare cases (e.g. measuring m_H in Higgs decays to four leptons) it is possible to construct an **analytical model**
- Requires **convolving** the predicted **detector response** with the truth **lineshape** per event (typically a Breit-Wigner for resonances)
- Response varies e.g. for different lepton flavours
- Can be extended to include per-event uncertainties



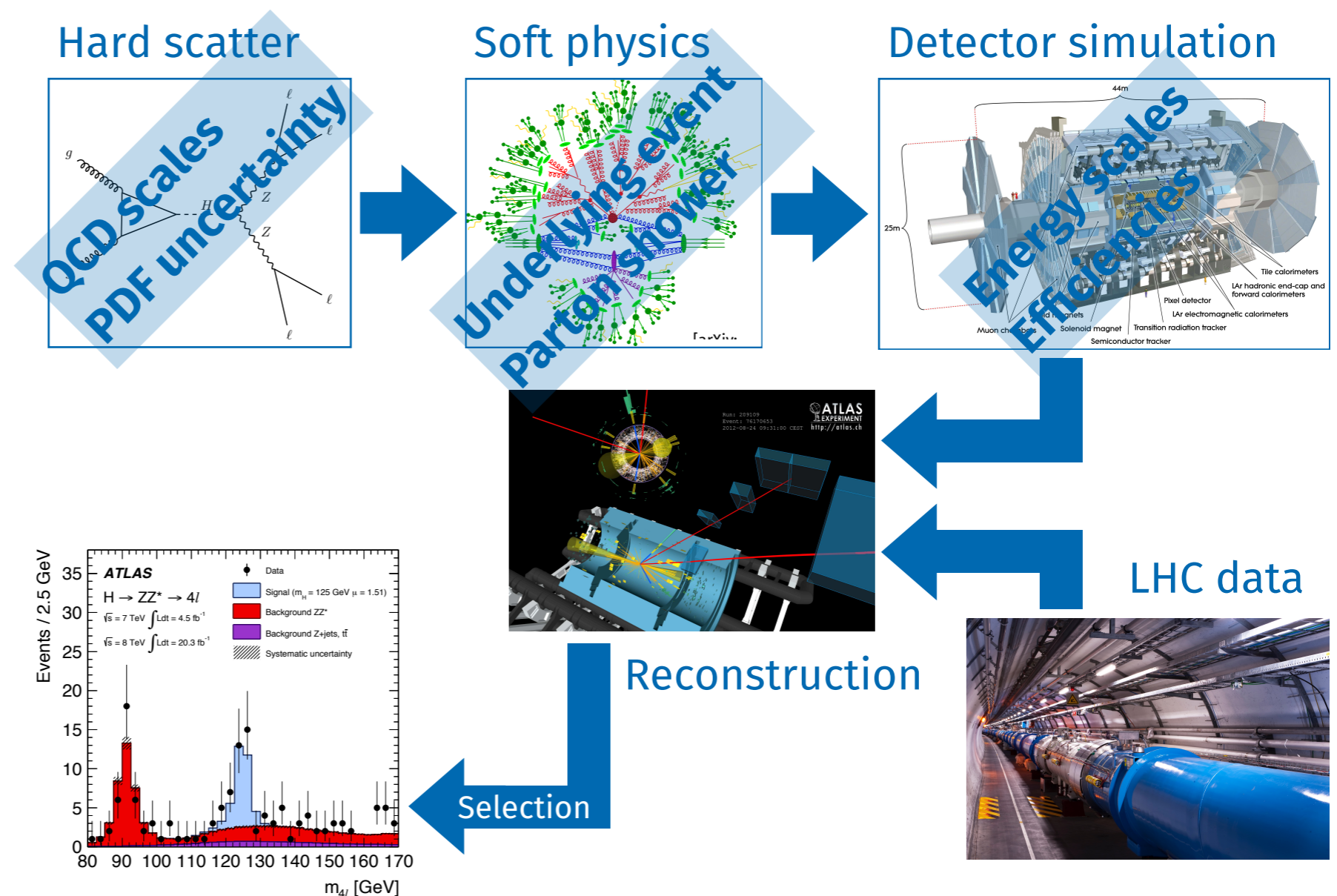
Incorporating systematic uncertainties

- Counts and distributions are subject to many uncertainties
- Capture the effect of each by introducing **nuisance parameters**
- Often only the maximum likelihood estimate from **subsidiary measurements** known

🗨️ E. Gross

- Information not sufficient to construct full likelihood model

- **Empirically** parametrise effect of each nuisance from available simulation for discrete model parameters



Assumption: likelihood of subsidiary measurement and effect on physics measurement factorise

“Combination” of selections

Data

$$L(\mathcal{D}, \mathcal{G} | \alpha) = \prod_{c=1}^{C_{\max}} \left[\text{Poisson}(n_c | \nu_c(\alpha)) \prod_{e=1}^{n_c} f_c(x_{ce} | \alpha) \right]$$

Global observables, i.e. MLE of nuisances from subsidiary measurements

expected counts

Empirically derive dependence on nuisances from **morphing** between “nominal”, “up” and “down” variations, i.e. simulation for discrete model parameters

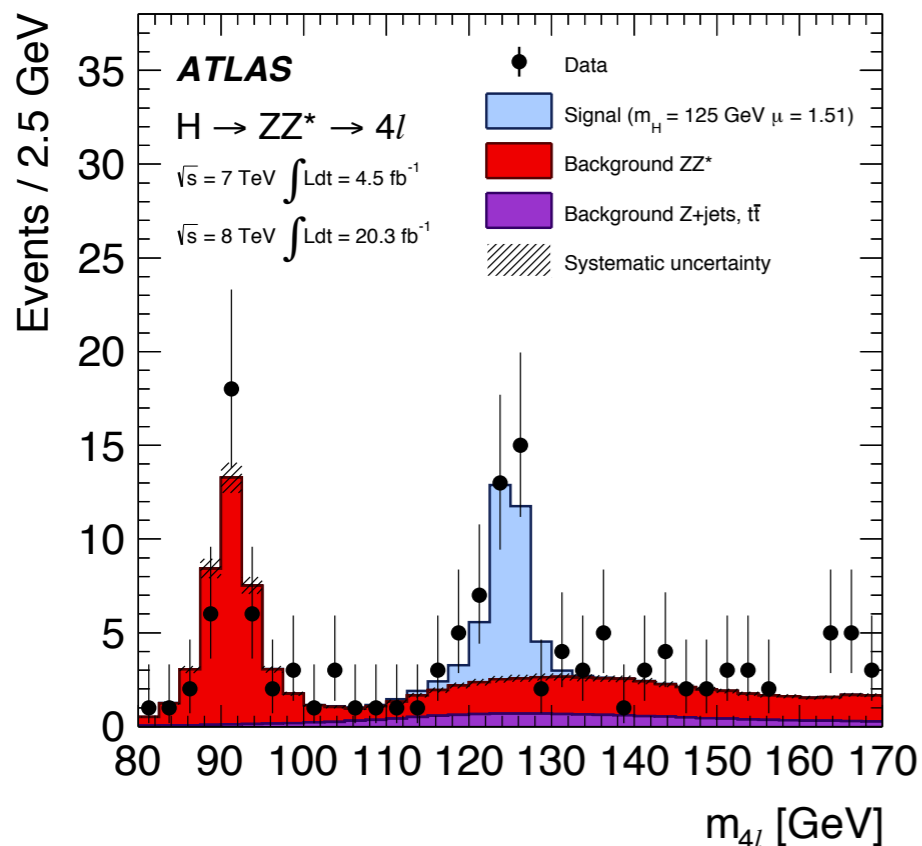
Augment data from count using observable

“Combination” of physics and subsidiary measurements

$$\times \prod_{p \in \mathcal{S}} f_p(\tilde{\theta}_p | \theta_p, \sigma_p)$$

Subsidiary measurement, typically **assume** Gauss or Poisson distribution

- Model building based on the **RooFit** package
 - **Key concept:** represent mathematical model by C++ objects from which the likelihood function is build
 - Support for normalisation relies on numeric techniques when analytical integrals not available
 - Support for conditional probability models, convolutions, etc.
- Commonly adopted by the **LHC experiments**



RooFit

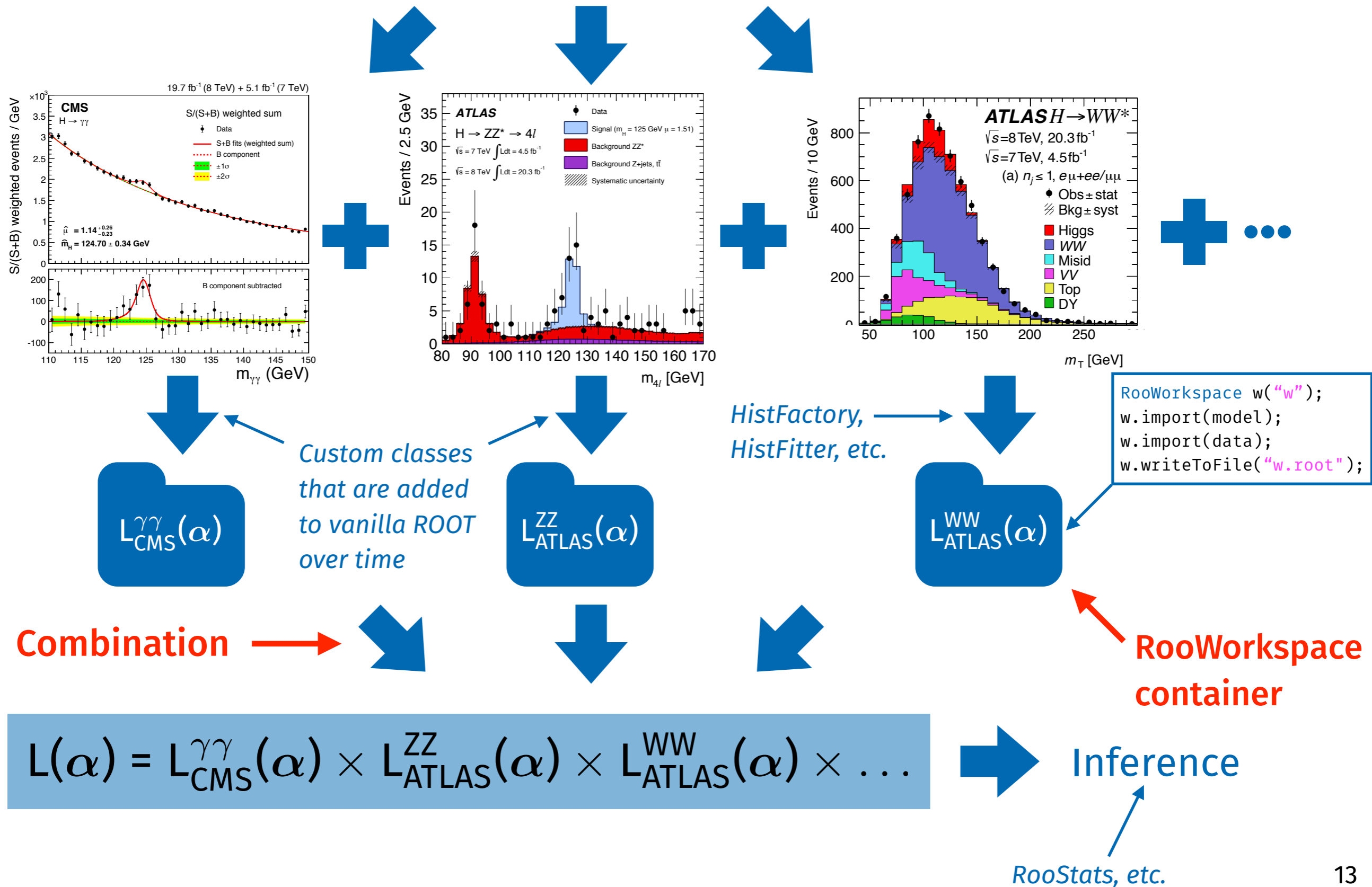
```
RooRealVar s("s", "s", 5);
RooRealVar b("b", "b", 50);
RooRealVar n("n", "n", 0, 100);
RooRealVar mu("mu", "mu", 0, 10);
RooFormulaVar nexp("nexp", "nexp", "mu*s+b",
  RooArgList(mu, s, b));
RooPoisson model("model", "model", n, nexp);
```

```
RooDataSet data("data", "data", n);
n=55;
data.add(n);
```



The analysis workflow

Analysers specialised in simulation, specific measurements, etc.



- Combining involves more than building a **simultaneous PDF** and implementing **parameterisations**

Detector uncertainties follow strategy of individual experiments
Theoretical uncertainties on inclusive cross sections are generally correlated between experiments, but correlation between processes is often not clear and depends on phase space
Acceptance and **efficiency** related uncertainties not correlated
Monte Carlo statistics related nuisances affect a single bin only

Eigenbasis, multivariate subsidiary measurements, template morphing, ...

Correlations

Fully or partially correlated parameters, correlated and uncorrelated parts, ...

E.g. luminosity, jet energy scale

Likelihood combinations

Harmonisation

Cross sections, pdf sets, uncertainties, ...

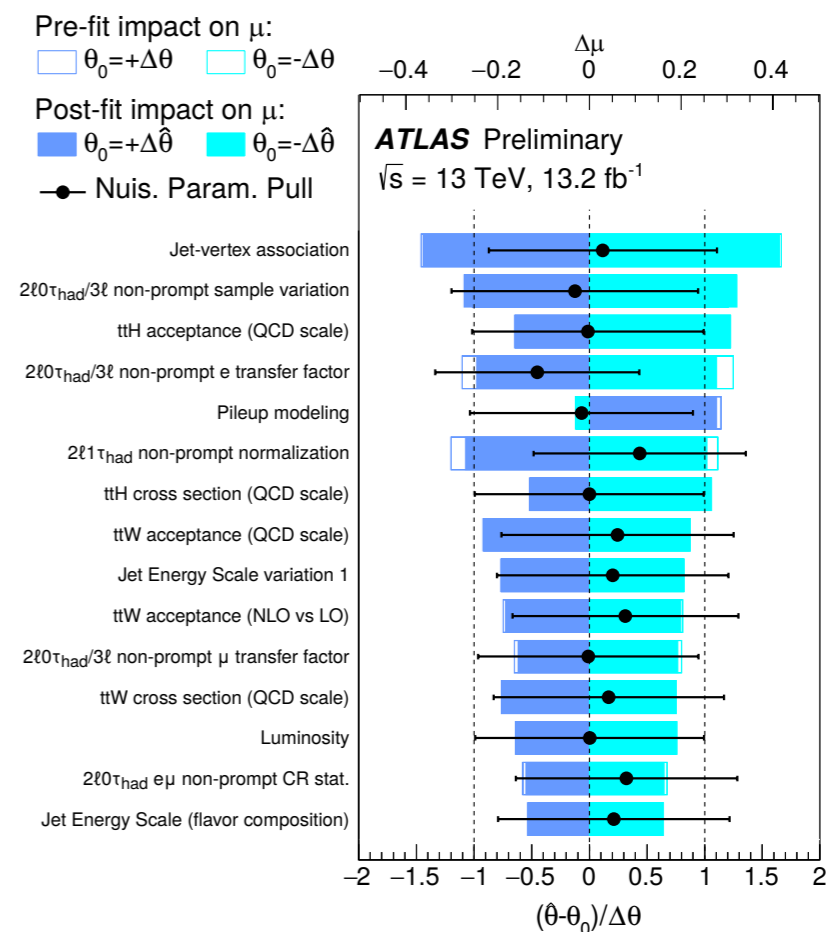
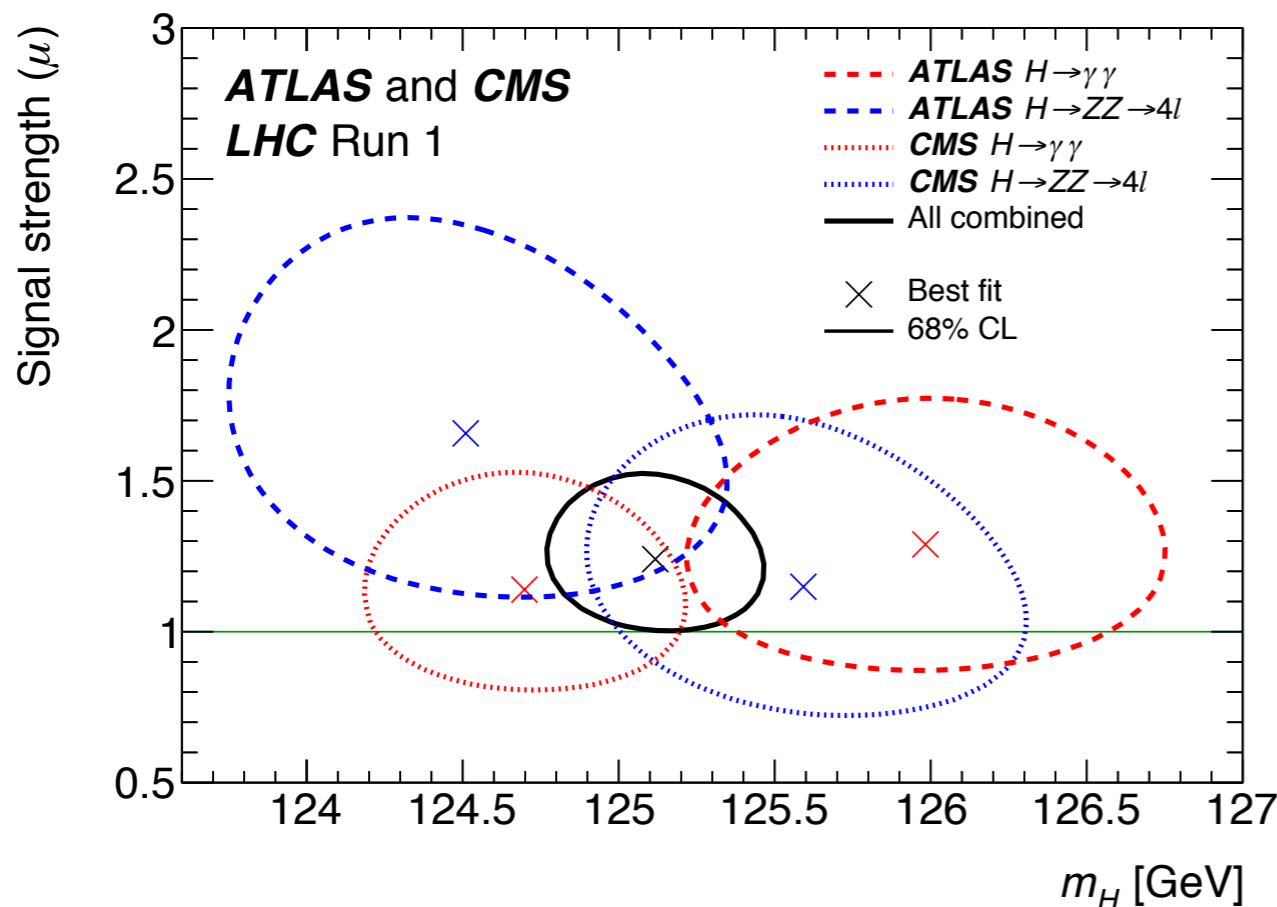
Cross checks

Independent tools and analysis teams cross checking pdf, nuisance parameters and inference

Book-keeping

Almost 600 disjoint selections with one, two, or more observables and more than 4200 nuisance parameters

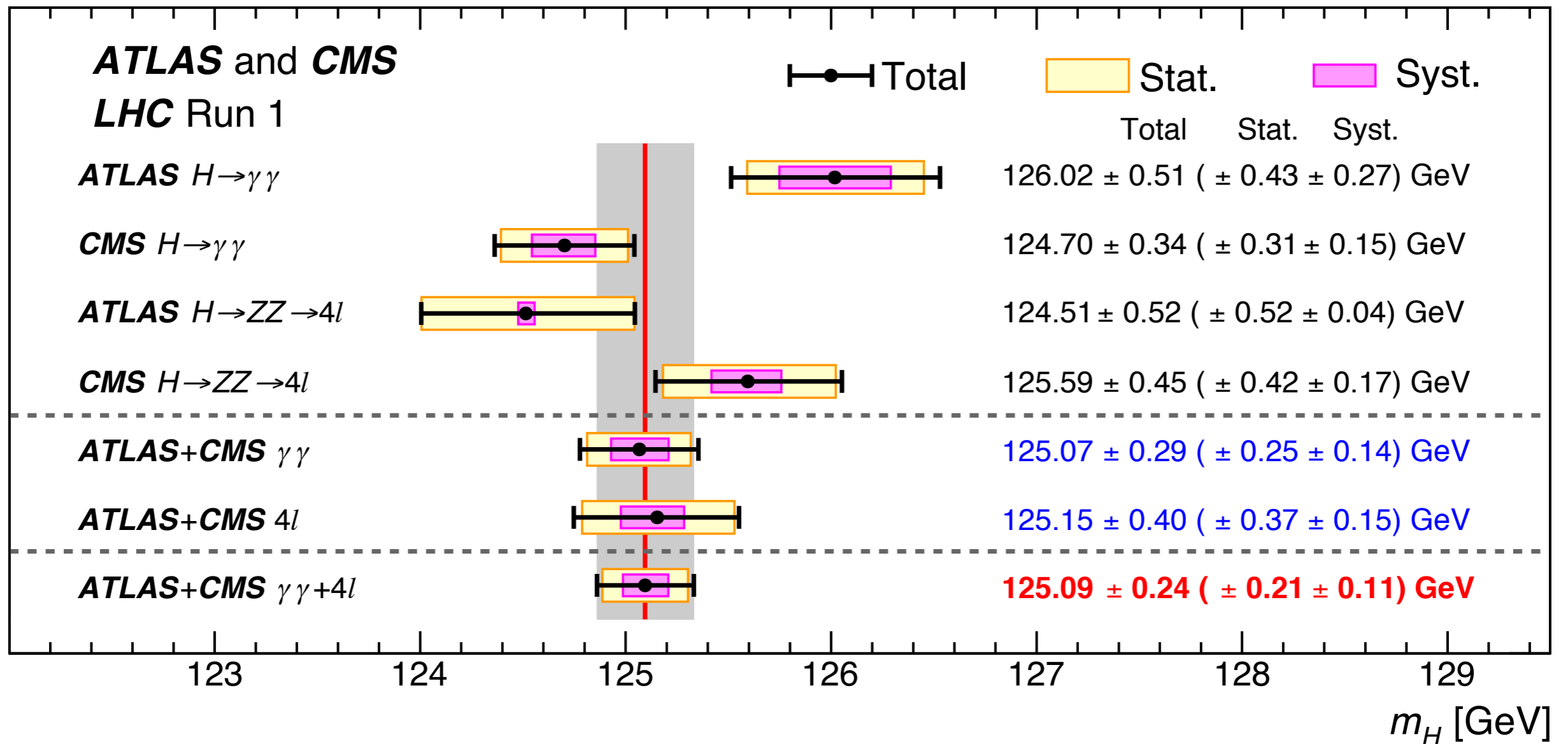
- Follow standard **frequentist methods** for constructing confidence intervals, hypothesis testing and evaluating the tensions
 - Often the **asymptotic approximation** is used 💬 E. Gross
- **RooStats** provides a series of tools for this tasks
 - General concepts are completely **decoupled** from model building and work on any probability model and dataset provided
- Internally developed tools for **additional cross checks**
 - E.g. pulls, constraints, correlations and impact of nuisance parameters



- Complex likelihood function may lead to **instabilities** and **performance issues**, even though many aspects are simplified
 - Minimisation of the likelihood function takes several hours (with Hesse days) and requires large (> 7G) memory nodes
 - Largest fraction of time spend in **interpolation**
- Several strategies to **reduce number of interpolations**
 - Disentangle “**real**” effect of systematic variations from **statistical noise**, e.g. using **bootstrap replicas**
 - Performance groups provide **reduced set of variations**
 - Depending on the measurement performed, non-significant uncertainties can be **pruned** on analysis level
 - **Examples:** evaluate the variation of the counts and templates or the correlation with measurements
- Complementary **technical developments**
 - Constant term optimisation, caching, vectorisation, binned evaluation

 D. Börner

Naive combinations vs. full model (I)

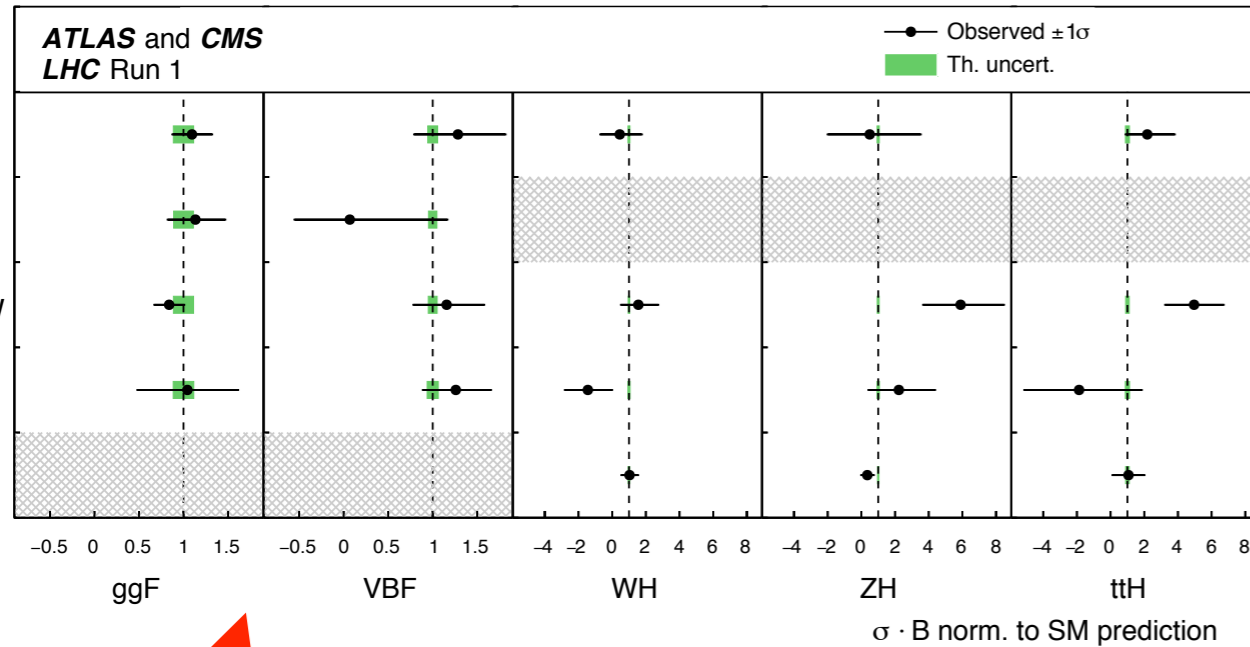


In a gaussian regime, e.g. statistically limited analyses with no complex correlations, weighted average is a good approximation:

$$m_H^{\gamma\gamma+4l} = 125.11 \pm 0.22 \text{ GeV}$$

Naive combinations vs. full model (II)

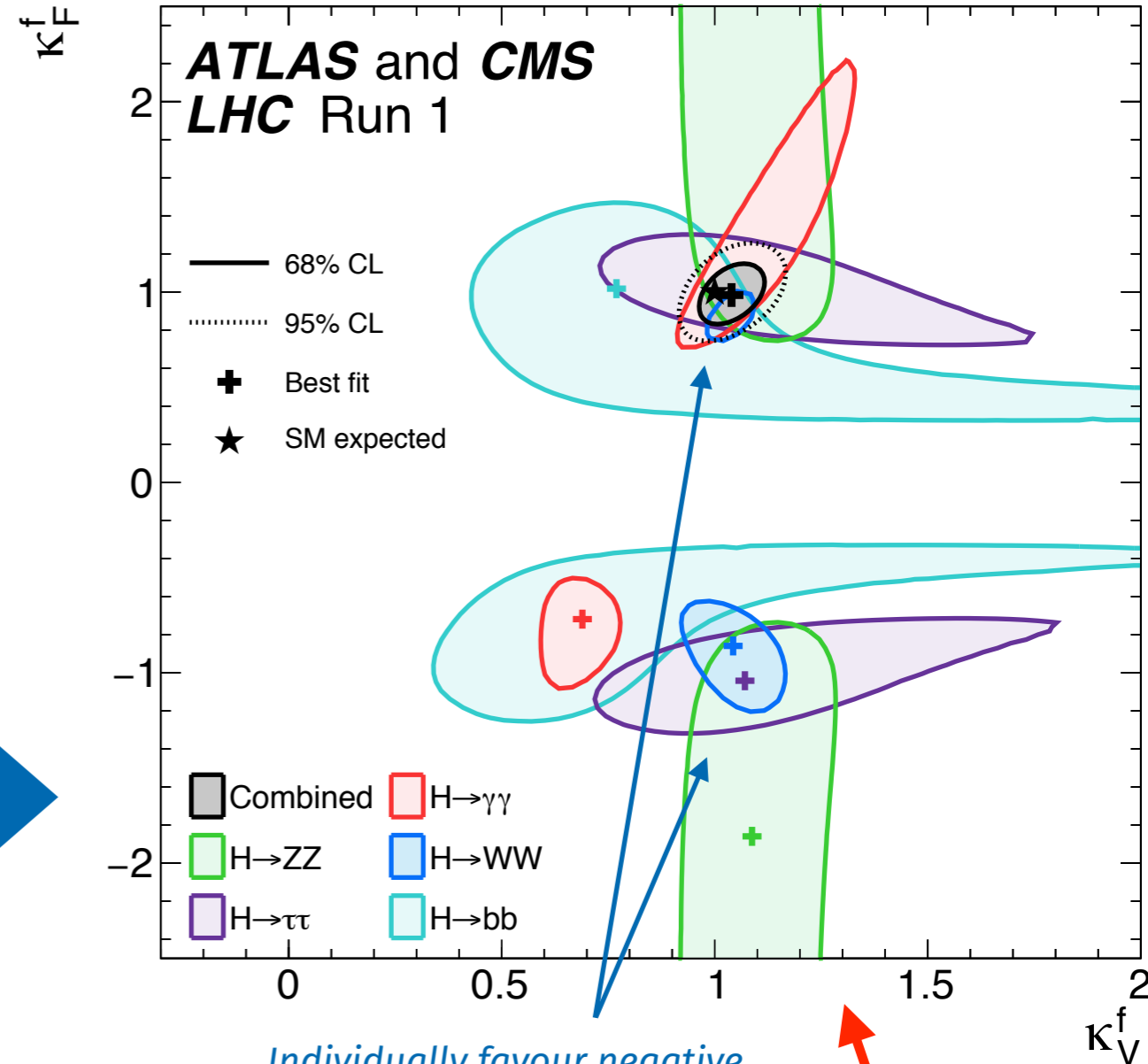
For more complex models, naive combinations are not feasible and may not reproduce results accurately



Measurements

Hides all intrinsic measurements, e.g. background rates, phase space regions, etc.

Requires knowledge of parametrisation, correlations, etc.



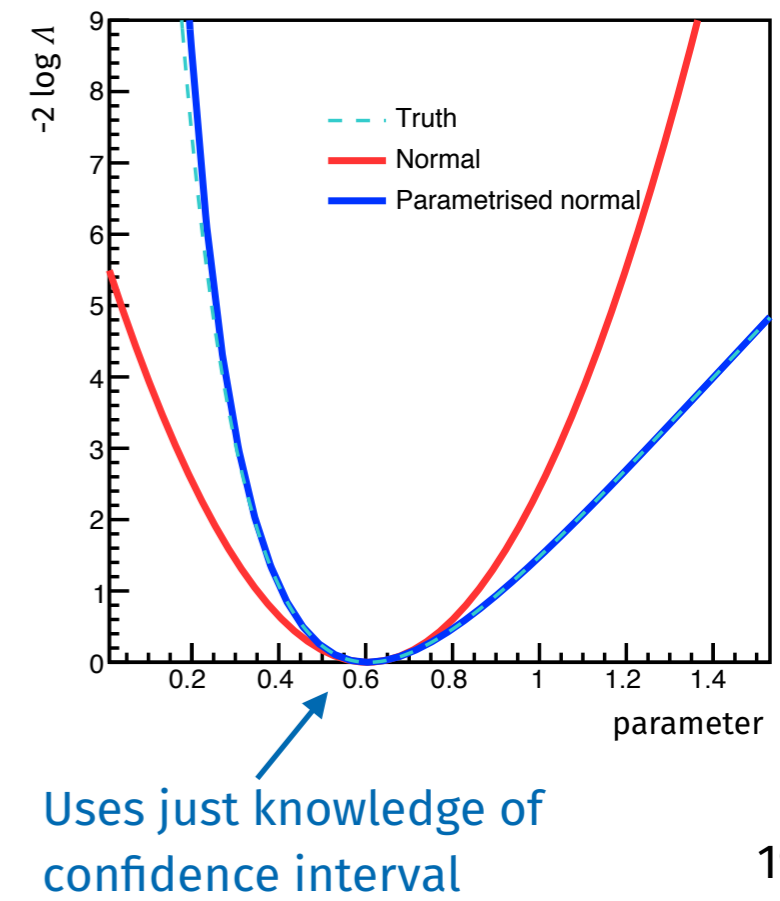
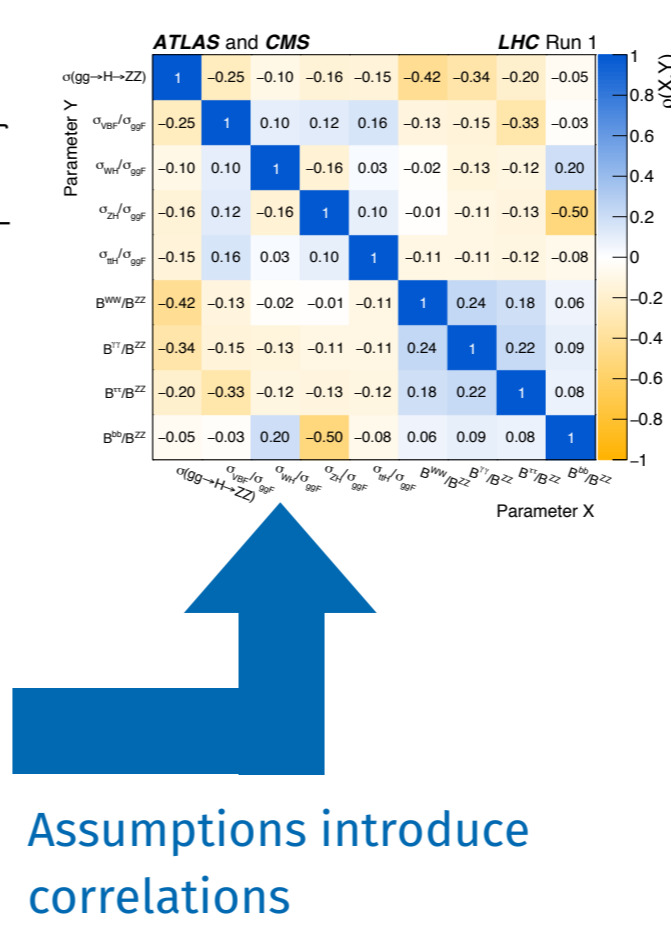
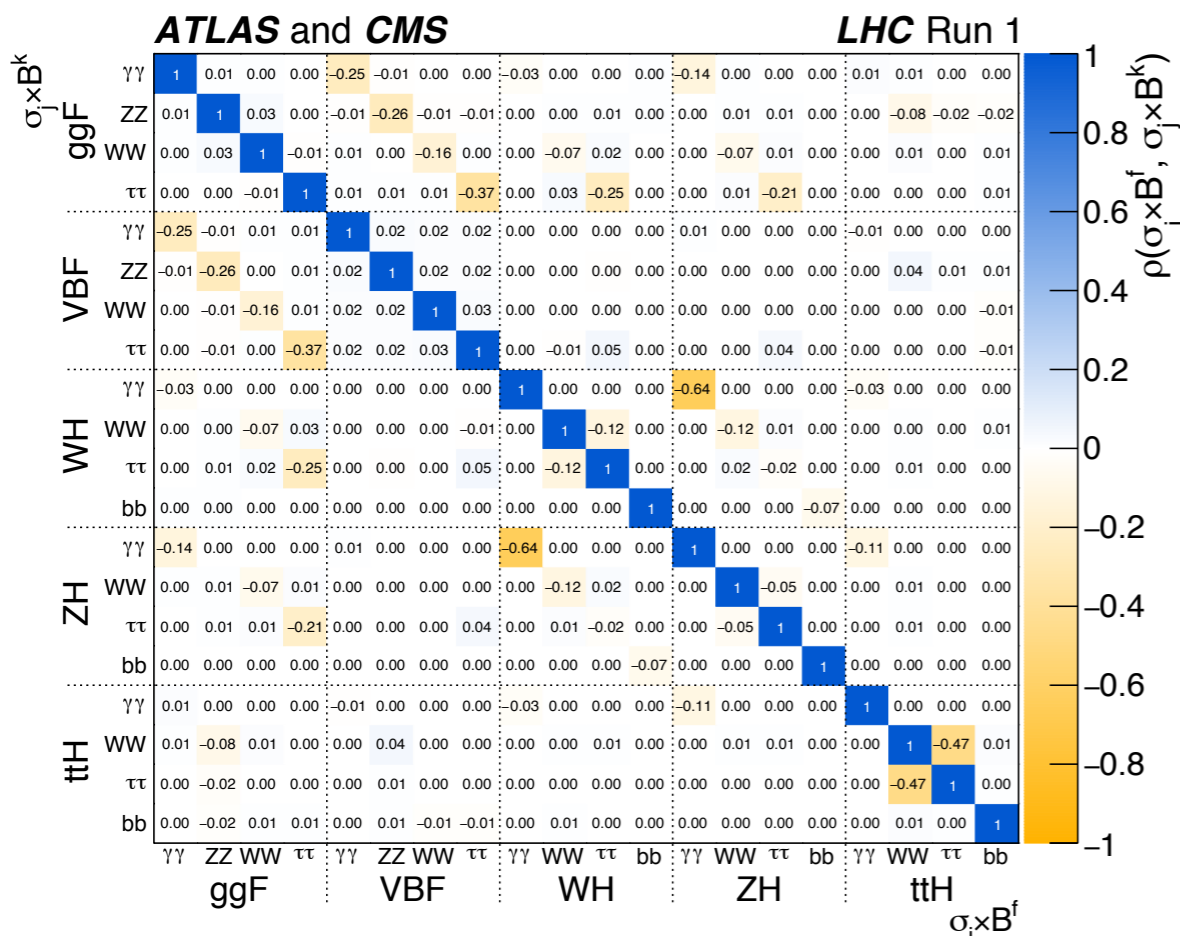
Individually favour negative relative signs, but combination favours positive

Interpretation

Making the results available (I)

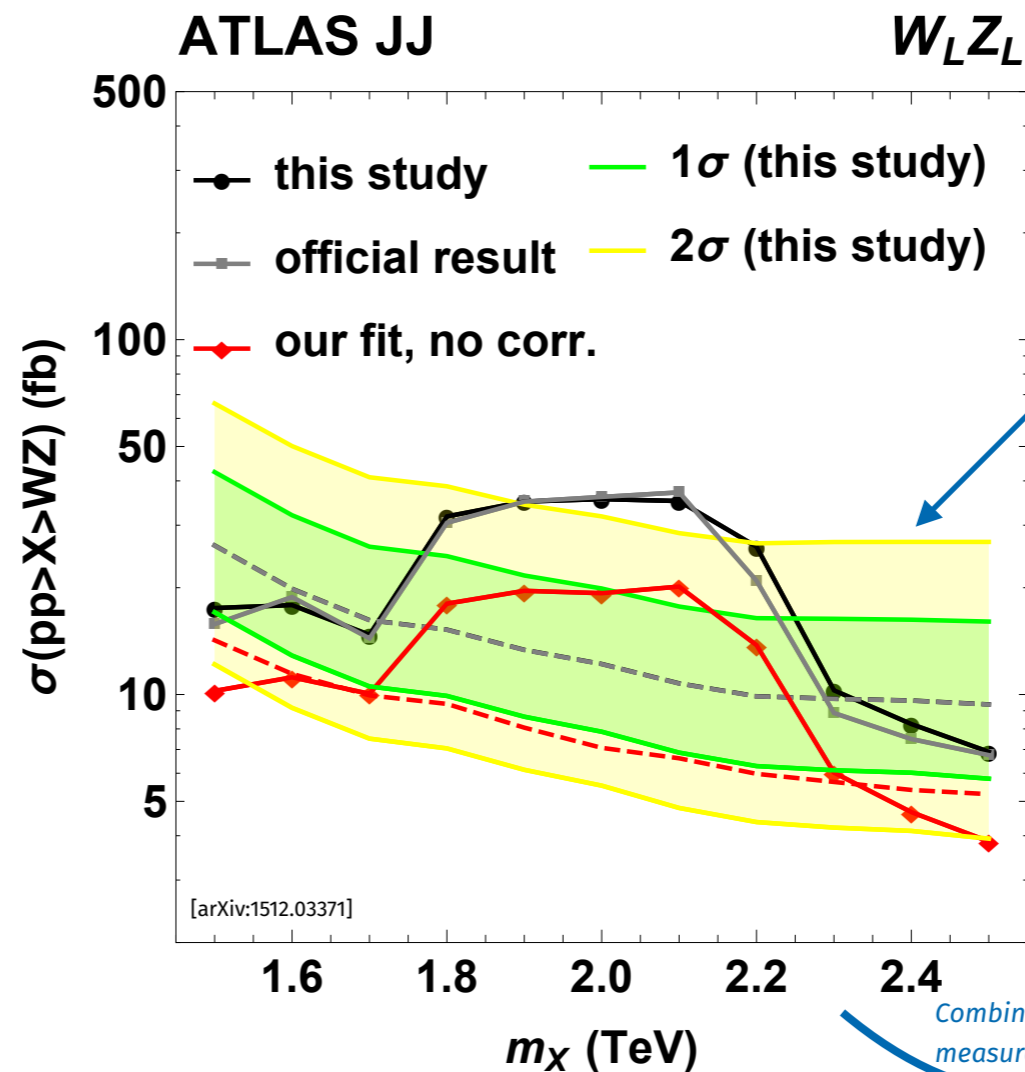
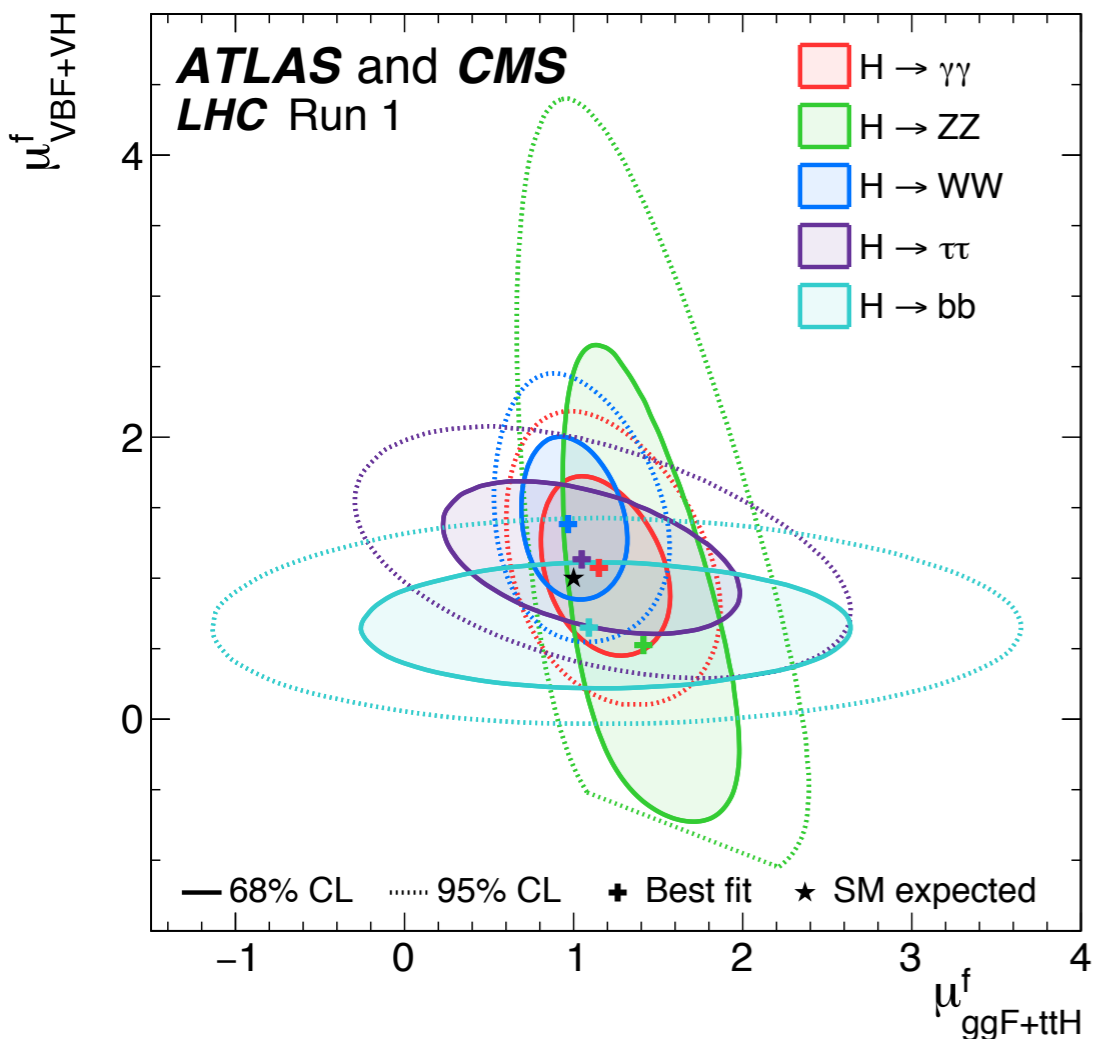
Ongoing efforts to make results from generic models available for re-interpretations

- Typically quote 1σ and 2σ **confidence intervals**
 - Sufficient to perform **naive combination** of results, but may differ significantly from full model when neglecting correlations, etc.
 - **Correlation matrix** at the minimum provides additional information
 - Can build (**parametrised**) **normal distributions** that reproduce likelihood scans if model sufficiently Gaussian

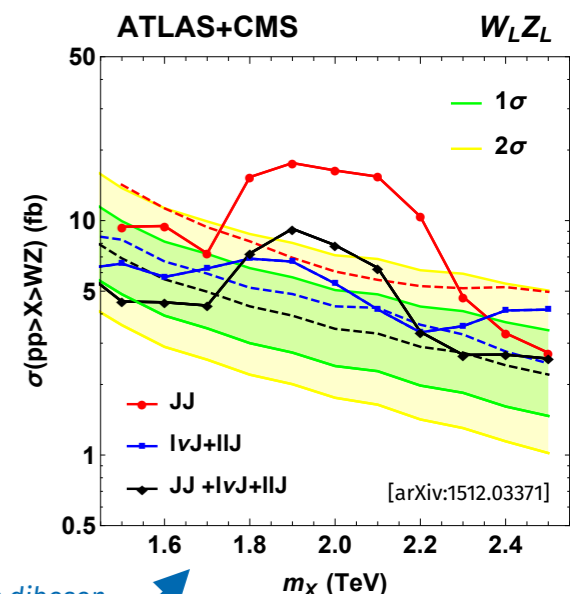


Making the results available (II)

- Often provide (multi-dimensional) profile likelihood scans to compare predictions, e.g. from simulating custom model
- Improving habit to make templates available, e.g. on **HepData**
 - Build simplified likelihood that can be combined
- Within the collaboration, share the persisted pdf, e.g. for BSM interpretations



Phenomenological parametrisation using HepData material



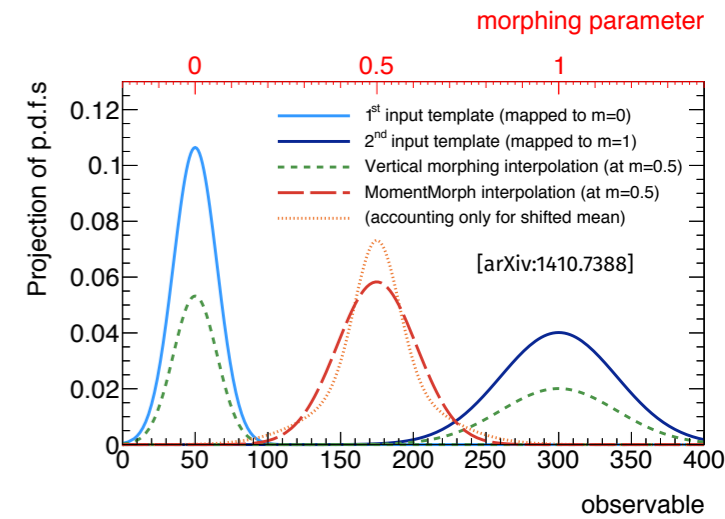
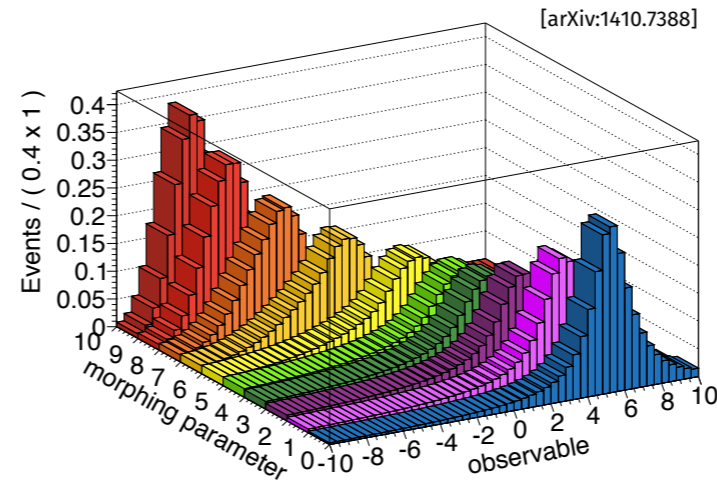
Combine diboson measurements

- With increasing data statistics, more sophisticated **theoretical models** can be tested, but also **systematic uncertainties** and **statistical modelling** gain in importance
 - Experimental systematics will shrink with more data being available for calibration measurements
 - Theoretical uncertainties will shrink when calculations improve
- This poses **many questions** and requires **technical developments**:
 - *What distributions of **subsidiary measurements** are correct?*
 - *Does the **parametrisation of an systematic uncertainty** capture the “real” effect, or introduces it too much detail?*
 - *Is the interpolation between sampling points for **response mapping** accurate, e.g. hadronisation/fragmentation model and non-factorisable uncertainties?*

	Best fit μ	Uncertainty				
		Total	Stat	Expt	Thbgd	Thsig
ATLAS + CMS (measured)	1.09	+0.11 -0.10	+0.07 -0.07	+0.04 -0.04	+0.03 -0.03	+0.07 -0.06
ATLAS + CMS (expected)		+0.11 -0.10	+0.07 -0.07	+0.04 -0.04	+0.03 -0.03	+0.07 -0.06
ATLAS (measured)	1.20	+0.15 -0.14	+0.10 -0.10	+0.06 -0.06	+0.04 -0.04	+0.08 -0.07
ATLAS (expected)		+0.14 -0.13	+0.10 -0.10	+0.06 -0.05	+0.04 -0.04	+0.07 -0.06
CMS (measured)	0.97	+0.14 -0.13	+0.09 -0.09	+0.05 -0.05	+0.04 -0.03	+0.07 -0.06
CMS (expected)		+0.14 -0.13	+0.09 -0.09	+0.05 -0.05	+0.04 -0.03	+0.08 -0.06

- Combinations are gaining in importance in HEP
- Combining **full likelihood preserves all available information**, i.e. correlations, composition of selections, subsidiary measurements, etc.
- Build the likelihood function using **Roofit**, persist it in a sharable **RooWorkspace** container, and perform inference using **Roostats** tools
- Combination requires **harmonisation**, poses **technical challenges** and drives **developments**
- Naive combinations can be done quickly, but do often not reproduce full results
- Ongoing efforts to make results available to community

Backup



Handles multidimensional morphing for systematics, but no simple interface

Typically used for modelling effect of systematic uncertainties on templates



Mainly used for physics measurements, e.g. m_H , or CP mixing

Uncertainties on the Higgs boson mass

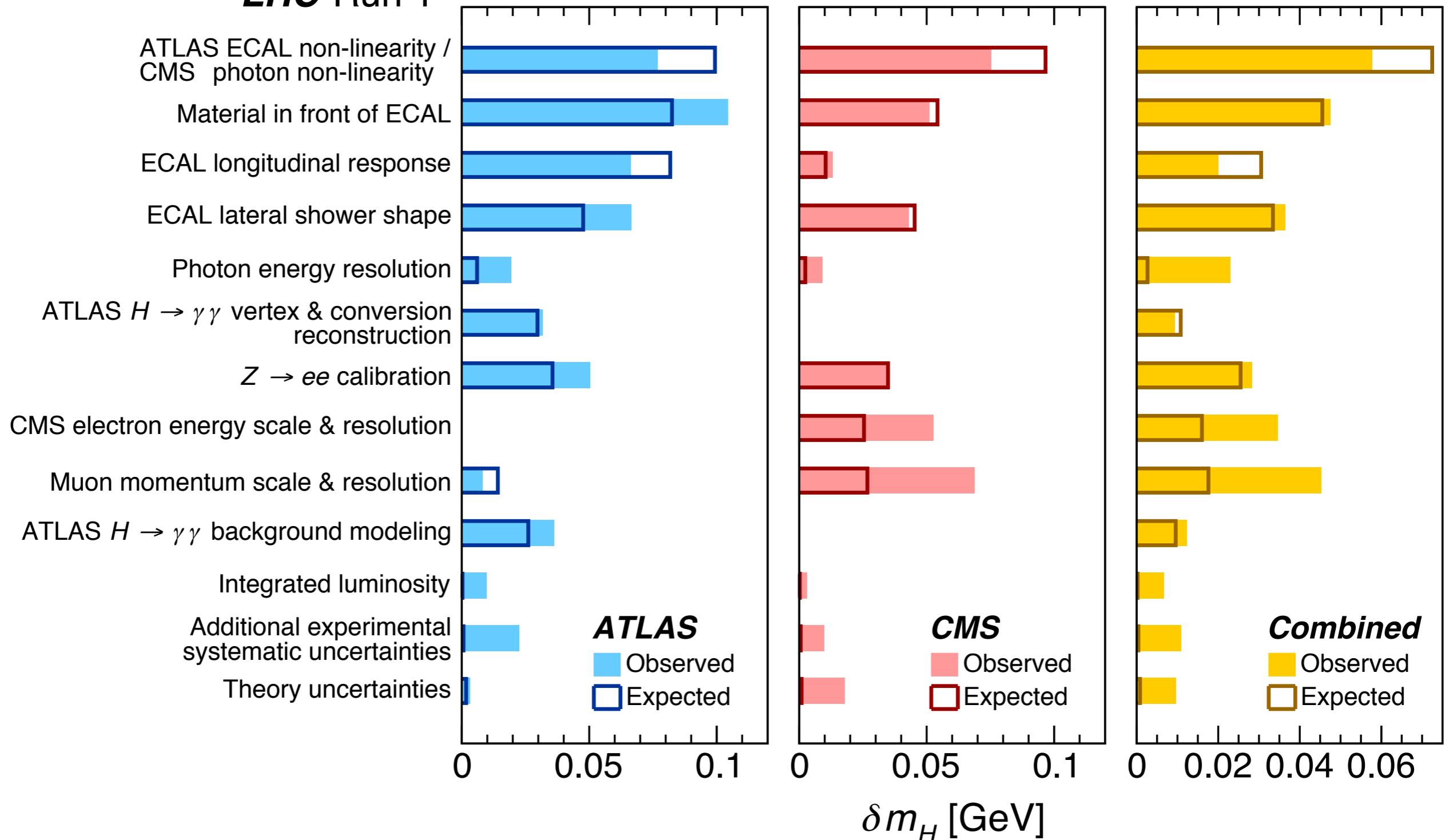


ATLAS and CMS LHC Run 1

Uncertainty in ATLAS
combined result

Uncertainty in CMS
combined result

Uncertainty in LHC
combined result



Uncertainties in the Higgs combination



Parameter	SM prediction	Best fit value	Uncertainty				Best fit value	Uncertainty				Best fit value	Uncertainty			
			Stat	Expt	Thbgd	Thsig		Stat	Expt	Thbgd	Thsig		Stat	Expt	Thbgd	Thsig
			ATLAS+CMS					ATLAS					CMS			
$\sigma(gg \rightarrow H \rightarrow ZZ)$ [pb]	0.51 ± 0.06	$0.59^{+0.11}_{-0.10}$	$+0.11$ -0.10	$+0.02$ -0.01	$+0.01$ -0.01	$+0.01$ -0.01	$0.77^{+0.19}_{-0.17}$	$+0.19$ -0.16	$+0.04$ -0.03	$+0.02$ -0.02	$+0.01$ -0.01	$0.44^{+0.14}_{-0.12}$	$+0.13$ -0.11	$+0.04$ -0.03	$+0.01$ -0.01	$+0.02$ -0.01
$\sigma_{\text{VBF}}/\sigma_{ggF}$	0.082 ± 0.009	$0.109^{+0.034}_{-0.027}$	$+0.029$ -0.024	$+0.013$ -0.009	$+0.006$ -0.004	$+0.010$ -0.008	$0.079^{+0.035}_{-0.026}$	$+0.030$ -0.023	$+0.014$ -0.009	$+0.008$ -0.005	$+0.009$ -0.006	$0.138^{+0.073}_{-0.051}$	$+0.061$ -0.046	$+0.033$ -0.019	$+0.014$ -0.006	$+0.015$ -0.010
σ_{WH}/σ_{ggF}	0.037 ± 0.004	$0.031^{+0.028}_{-0.026}$	$+0.024$ -0.022	$+0.012$ -0.012	$+0.008$ -0.008	$+0.003$ -0.002	$0.054^{+0.036}_{-0.026}$	$+0.031$ -0.023	$+0.012$ -0.008	$+0.014$ -0.009	$+0.007$ -0.004	$0.005^{+0.044}_{-0.037}$	$+0.037$ -0.028	$+0.021$ -0.023	$+0.010$ -0.008	$+0.003$ -0.001
σ_{ZH}/σ_{ggF}	0.0216 ± 0.0024	$0.066^{+0.039}_{-0.031}$	$+0.032$ -0.025	$+0.018$ -0.012	$+0.014$ -0.012	$+0.005$ -0.003	$0.013^{+0.028}_{-0.014}$	$+0.021$ -0.012	$+0.013$ -0.005	$+0.013$ -0.005	$+0.003$ -0.002	$0.123^{+0.076}_{-0.053}$	$+0.063$ -0.046	$+0.038$ -0.022	$+0.019$ -0.013	$+0.009$ -0.005
σ_{tH}/σ_{ggF}	0.0067 ± 0.0010	$0.0220^{+0.0068}_{-0.0057}$	$+0.0055$ -0.0048	$+0.0031$ -0.0023	$+0.0023$ -0.0020	$+0.0014$ -0.0010	$0.0126^{+0.0066}_{-0.0053}$	$+0.0052$ -0.0042	$+0.0031$ -0.0023	$+0.0024$ -0.0020	$+0.0013$ -0.0007	$0.0340^{+0.0158}_{-0.0116}$	$+0.0121$ -0.0097	$+0.0085$ -0.0051	$+0.0048$ -0.0036	$+0.0026$ -0.0015
B^{WW}/B^{ZZ}	$8.09 \pm < 0.01$	$6.7^{+1.6}_{-1.3}$	$+1.5$ -1.2	$+0.4$ -0.3	$+0.4$ -0.3	$+0.3$ -0.2	$6.5^{+2.1}_{-1.6}$	$+2.0$ -1.4	$+0.6$ -0.4	$+0.5$ -0.4	$+0.3$ -0.2	$7.1^{+2.9}_{-2.1}$	$+2.6$ -1.8	$+1.0$ -0.7	$+0.7$ -0.5	$+0.4$ -0.3
$B^{\gamma\gamma}/B^{ZZ}$	0.0854 ± 0.0010	$0.069^{+0.018}_{-0.014}$	$+0.018$ -0.014	$+0.003$ -0.002	$+0.002$ -0.001	$+0.002$ -0.002	$0.062^{+0.024}_{-0.018}$	$+0.023$ -0.017	$+0.007$ -0.004	$+0.002$ -0.001	$+0.003$ -0.002	$0.079^{+0.034}_{-0.023}$	$+0.032$ -0.023	$+0.009$ -0.005	$+0.003$ -0.002	$+0.004$ -0.003
$B^{\tau\tau}/B^{ZZ}$	2.36 ± 0.05	$1.77^{+0.59}_{-0.46}$	$+0.52$ -0.41	$+0.27$ -0.20	$+0.05$ -0.04	$+0.06$ -0.04	$2.17^{+1.07}_{-0.74}$	$+0.89$ -0.64	$+0.53$ -0.35	$+0.16$ -0.10	$+0.17$ -0.09	$1.56^{+0.90}_{-0.61}$	$+0.78$ -0.54	$+0.45$ -0.26	$+0.07$ -0.05	$+0.07$ -0.04
B^{bb}/B^{ZZ}	21.5 ± 1.0	$4.2^{+4.4}_{-2.6}$	$+2.8$ -2.0	$+2.3$ -1.1	$+2.5$ -1.2	$+0.4$ -0.2	$9.6^{+10.1}_{-5.7}$	$+7.4$ -4.4	$+4.5$ -2.4	$+5.1$ -2.7	$+1.3$ -0.5	$3.7^{+4.1}_{-2.4}$	$+3.1$ -2.0	$+1.8$ -0.9	$+1.9$ -1.1	$+0.4$ -0.2
			$+16.8$ -9.0	$+13.9$ -7.9	$+6.3$ -2.8	$+6.7$ -3.3	$+2.1$ -0.9	$+29.3$ -11.8	$+24.2$ -10.5	$+10.9$ -3.3	$+11.8$ -4.0	$+4.0$ -1.2	$+23.4$ -10.4	$+12.7$ -3.8	$+12.2$ -4.4	$+2.5$ -0.9

Parameter	Best fit value	Uncertainty				Best fit value	Uncertainty				Best fit value	Uncertainty			
		Stat	Expt	Thbgd	Thsig		Stat	Expt	Thbgd	Thsig		Stat	Expt	Thbgd	Thsig
		ATLAS+CMS					ATLAS					CMS			
κ_{gZ}	$1.09^{+0.11}_{-0.11}$	$+0.09$ -0.09	$+0.02$ -0.02	$+0.00$ -0.01	$+0.06$ -0.05	$1.20^{+0.16}_{-0.15}$	$+0.14$ -0.14	$+0.03$ -0.03	$+0.02$ -0.02	$+0.07$ -0.06	$0.99^{+0.14}_{-0.13}$	$+0.12$ -0.12	$+0.03$ -0.04	$+0.01$ -0.01	$+0.06$ -0.04
λ_{Zg}	$1.27^{+0.23}_{-0.20}$	$+0.18$ -0.16	$+0.10$ -0.07	$+0.06$ -0.05	$+0.10$ -0.08	$1.07^{+0.26}_{-0.22}$	$+0.21$ -0.18	$+0.10$ -0.06	$+0.07$ -0.06	$+0.09$ -0.07	$1.47^{+0.45}_{-0.34}$	$+0.35$ -0.28	$+0.22$ -0.14	$+0.11$ -0.10	$+0.13$ -0.09
λ_{tg}	$1.78^{+0.30}_{-0.27}$	$+0.21$ -0.20	$+0.13$ -0.11	$+0.09$ -0.09	$+0.14$ -0.11	$1.40^{+0.34}_{-0.33}$	$+0.25$ -0.24	$+0.14$ -0.15	$+0.12$ -0.14	$+0.14$ -0.09	$-2.26^{+0.50}_{-0.53}$	$+0.43$ -0.39	$+0.22$ -0.23	$+0.04$ -0.18	$+0.14$ -0.21
λ_{WZ}	$0.88^{+0.10}_{-0.09}$	$+0.09$ -0.08	$+0.03$ -0.03	$+0.03$ -0.02	$+0.02$ -0.01	$0.92^{+0.14}_{-0.12}$	$+0.13$ -0.11	$+0.04$ -0.03	$+0.03$ -0.03	$+0.02$ -0.02	$-0.85^{+0.13}_{-0.15}$	$+0.11$ -0.13	$+0.05$ -0.06	$+0.04$ -0.04	$+0.01$ -0.03
$ \lambda_{\gamma Z} $	$0.89^{+0.11}_{-0.10}$	$+0.10$ -0.09	$+0.03$ -0.02	$+0.01$ -0.02	$+0.02$ -0.01	$0.87^{+0.15}_{-0.13}$	$+0.15$ -0.13	$+0.04$ -0.04	$+0.02$ -0.01	$+0.02$ -0.02	$0.91^{+0.17}_{-0.14}$	$+0.16$ -0.14	$+0.04$ -0.03	$+0.02$ -0.02	$+0.02$ -0.02
$ \lambda_{\tau Z} $	$0.85^{+0.13}_{-0.12}$	$+0.12$ -0.10	$+0.07$ -0.06	$+0.01$ -0.02	$+0.02$ -0.01	$0.96^{+0.21}_{-0.18}$	$+0.18$ -0.15	$+0.10$ -0.09	$+0.04$ -0.03	$+0.03$ -0.02	$0.78^{+0.20}_{-0.17}$	$+0.17$ -0.15	$+0.10$ -0.08	$+0.01$ -0.02	$+0.02$ -0.01
$ \lambda_{bZ} $	$0.58^{+0.16}_{-0.20}$	$+0.12$ -0.17	$+0.07$ -0.06	$+0.07$ -0.07	$+0.03$ -0.04	$0.61^{+0.24}_{-0.24}$	$+0.20$ -0.19	$+0.09$ -0.12	$+0.10$ -0.10	$+0.04$ -0.03	$0.47^{+0.26}_{-0.17}$	$+0.23$ -0.13	$+0.06$ -0.07	$+0.11$ -0.08	$+0.00$ -0.04
		$+0.25$ -0.22	$+0.09$ -0.07	$+0.08$ -0.05	$+0.06$ -0.05	$+0.36$ -0.29	$+0.31$ -0.26	$+0.12$ -0.09	$+0.11$ -0.08	$+0.08$ -0.05	$+0.38$ -0.37	$+0.32$ -0.34	$+0.15$ -0.11	$+0.11$ -0.09	$+0.07$ -0.07