ATL-PHYS-PUB-2011-11

CMS NOTE-2011/005

Procedure for the LHC Higgs boson search combination in Summer 2011

The ATLAS Collaboration
The CMS Collaboration
The LHC Higgs Combination Group

August 18, 2011

Abstract

In this note, we report the results of the technical combination exercises conducted by the group during Winter-Spring 2011 and summarize the decisions taken in preparation for the statistical combination of the Standard Model Higgs boson searches at the LHC. The procedure to be used for the combination in Summer 2011 is explicitly detailed to avoid potential biases from decisions taken after the data have been collected.

Contents

1 Introduction

The discovery of the mechanism for electroweak symmetry breaking is one of the keystones of the Large Hadron Collider (LHC) physics program. By summer of 2011, ATLAS [1] and CMS [2] will have results with over 1 fb^{-1} of data that should allow LHC to make very strong statements on the Standard Model (SM) Higgs boson in a wide mass range [3, 4].

In December of 2010, the LHC Higgs Combination Group (LHC-HCG) was formed with the aim of preparing for a combination of ATLAS and CMS SM Higgs search results. This report summarises the efforts of the LHC-HCG over the last few months towards this goal. The outline of the report is as follows:

- Sections 2 and 3 define the procedures for characterising exclusion of a signal or an observation of excesses to be used for the combination in summer 2011.
- Section 4 defines Higgs mass points for which the ATLAS+CMS combination is expected to be performed.
- Then, in Section 5, we summarise which systematic errors will be correlated between ATLAS and CMS and how the errors will be modelled in general.
- In Section 6, we outline the expected format of presenting the final results.
- In Section 7, we document the results of the technical exercises with toy analysis models (synchronisation and validation).
- After giving a summary, we make a few closing remarks on the overall experience of the last six months and an outlook for the future.

2 Limit setting procedure for the summer 2011

In this section, we summarise the arrived-at procedure for computing exclusion limits, which is based on the modified frequentist method, often referred to as CL_s [5–10]. To fully define the method, we specify the choice of the test statistic and how we treat nuisance parameters in the construction of the test statistic and in generating pseudo-data. To put the method in a broader context, a brief overview of statistical methods used in high energy physics is given in Appendix A.

In this section, the expected SM Higgs boson event yields will be generically denoted as s, backgrounds—as b. These will stand for event counts in one or multiple bins or for unbinned probability density functions, whichever approach is used in an analysis. It has become customary to express null results of the SM-like Higgs searches as a limit on a *signal strength modifier* μ (also referred to as R) that is taken to change the SM Higgs boson cross sections of all production mechanisms by exactly the same scale μ. Note that the decay branching ratios are assumed to be unchanged.

Predictions for both signal and background yields, prior to the scrutiny of the *observed data entering the statistical analysis*, are subject to multiple uncertainties that are handled by introducing nuisance parameters θ, so that signal and background expectations become functions of the nuisance parameters: $s(\theta)$ and $b(\theta)$.

All sources of uncertainties are taken to be either 100%-correlated (positively or negatively) or uncorrelated (independent). Partially correlated errors are either broken down to sub-components that can be said to be either 100% correlated or uncorrelated, or declared to be 100% / 0% correlated, whichever is believed to be appropriate or more conservative. This allows us to include all constraints in the likelihood functions in a clean factorised form.

The systematic error *pdf*s $\rho(\theta|\tilde{\theta})$, where $\tilde{\theta}$ is the default value of the nuisance parameter, reflect our degree of belief on what the true value of θ might be. Both the form of these *pdf*s to be used in the combination and the question of which errors are to be taken as correlated between ATLAS and CMS are discussed in detail in Section 5.

Next, we take a conceptual step to re-interpret systematic error *pdf*s $\rho(\theta|\tilde{\theta})$ as posteriors arising from some real or imaginary measurements $\tilde{\theta}$, as given by the Bayes' theorem:

$$\rho(\theta|\tilde{\theta}) \;\sim\; p(\tilde{\theta}\,|\,\theta)\cdot\pi_\theta(\theta), \tag{1}$$

where $\pi_\theta(\theta)$ functions are hyper-priors for those "measurements". As will be shown later, the *pdf*s we chose to work with (normal, log-normal, gamma distribution) can be easily re-formulated in such a context, while keeping $\pi_\theta(\theta)$ flat.

Such a shift in the point of view allows one to represent all systematic errors in a frequentist context. By writing a systematic error *pdf* as the posterior $\rho(\theta|\tilde{\theta})$ constructed from a fictional auxiliary "measurement", the *pdf* $p(\tilde{\theta}\,|\,\theta)$ for that auxiliary measurement can be used to constrain the likelihood of the main measurement in a frequentist calculation. Furthermore, the auxiliary "measurement" *pdf* $p(\tilde{\theta}\,|\,\theta)$ can be used to construct sampling distributions of the test statistic following the pure frequentist language (in contrast to the Bayesian-frequentist hybrid used at LEP and Tevatron—see Appendix A for details).

The following enumerated list specifies explicitly the entire procedure.

2.1 Observed limits

1. Construct the likelihood function $\mathcal{L}(\text{data}|\mu,\theta)$

$$\mathcal{L}(\text{data}\,|\,\mu,\theta) = \text{Poisson}\,(\,\text{data}\,|\,\mu\cdot s(\theta)+b(\theta)\,)\cdot p(\tilde{\theta}|\theta)\,. \tag{2}$$

Here "data" represents either the actual experimental *observation* or *pseudo-data* used to construct sampling distributions to be discussed further below. The parameter μ is the signal strength modifier and θ represents the full suite of nuisance parameters.

Poisson $(\,\text{data}\,|\,\mu s+b\,)$ stands either for a product of Poisson probabilities to observe n_i events in bins i:

$$\prod_i \frac{(\mu s_i+b_i)^{n_i}}{n_i!}\; e^{-\mu s_i-b_i}\,, \tag{3}$$

or for an unbinned likelihood over k events in the data sample:

$$k^{-1}\prod_i(\mu S f_s(x_i)+B f_b(x_i))\;\cdot\; e^{-(\mu S+B)}\,. \tag{4}$$

In the latter equation, $f_s(x)$ and $f_b(x)$ are *pdf*s of signal and background of some observable(s) x, while S and B are total event rates expected for signal and backgrounds.

2. To compare the compatibility of the *data* with the *background-only* and *signal+background* hypotheses, where the signal is allowed to be scaled by some factor μ, we construct the test statistic $\tilde{q}_\mu$ [11] based on the profile likelihood ratio:

$$\tilde{q}_\mu \;=\; -2\ln\frac{\mathcal{L}(\text{data}|\mu,\hat{\theta}_\mu)}{\mathcal{L}(\text{data}|\hat{\mu},\hat{\theta})}, \quad \text{with a constraint } 0\le\hat{\mu}\le\mu \tag{5}$$

where $\hat{\theta}_\mu$ refers to the conditional maximum likelihood estimators of θ, given the signal strength parameter μ and "data" that, as before, may refer to the actual experimental observation or pseudo-data (toys). The pair of parameter estimators $\hat{\mu}$ and $\hat{\theta}$ correspond to the global maximum of the likelihood.

The lower constraint $0 \le \hat{\mu}$ is dictated by physics (signal rate is positive), while the upper constraint $\hat{\mu} \le \mu$ is imposed by hand in order to guarantee a one-sided (not detached from zero) confidence interval. Physics-wise, this means that upward fluctuations of the data such that $\hat{\mu} > \mu$ are not considered as evidence against the signal hypothesis, namely a signal with strength μ.

Note that this definition of the test statistic differs from what has been used at LEP (where "profiling" of systematic errors was not used) and at Tevatron (where systematic errors were profiled, but μ in the denominator was fixed at zero). See Appendix A for details.

3. Find the *observed* value of the test statistic $\tilde{q}_\mu^{obs}$ for the given signal strength modifier μ under test.

4. Find values of the nuisance parameters $\hat{\theta}_0^{obs}$ and $\hat{\theta}_\mu^{obs}$ best describing the experimentally *observed* data (i.e. maximising the likelihood as given in Eq. 2), for the *background-only* and *signal+background* hypotheses, respectively.

5. Generate toy Monte Carlo pseudo-data to construct *pdf*s $f(\tilde{q}_\mu|\mu,\hat{\theta}_\mu^{\text{obs}})$ and $f(\tilde{q}_\mu|0,\hat{\theta}_0^{\text{obs}})$ assuming a signal with strength μ in the *signal+background* hypothesis and for the *background-only* hypothesis ($\mu = 0$). These distributions are shown in Fig. 1. Note, that for the purposes of generating a *pseudo-dataset*, the nuisance parameters are fixed to the values $\hat{\theta}_\mu^{\text{obs}}$ or $\hat{\theta}_0^{\text{obs}}$ obtained by fitting the observed data, but are allowed to float in fits needed to evaluate the test statistic. This way, in which the nuisance parameters are fixed to their maximum likelihood estimates, has good coverage properties [12].

6. Having constructed $f(\tilde{q}_\mu|\mu,\hat{\theta}_\mu^{\text{obs}})$ and $f(\tilde{q}_\mu|0,\hat{\theta}_0^{\text{obs}})$ distributions, we define two *p*-values to be associated with the actual observation for the *signal+background* and *background-only* hypotheses, p_μ and p_b:

$$p_\mu \;=\; P(\,\tilde{q}_\mu \ge \tilde{q}_\mu^{obs}\,|\,\text{signal+background}) \;=\; \int_{\tilde{q}_\mu^{\text{obs}}}^{\infty} f(\tilde{q}_\mu|\mu,\hat{\theta}_\mu^{\text{obs}})\,d\tilde{q}_\mu\,, \tag{6}$$

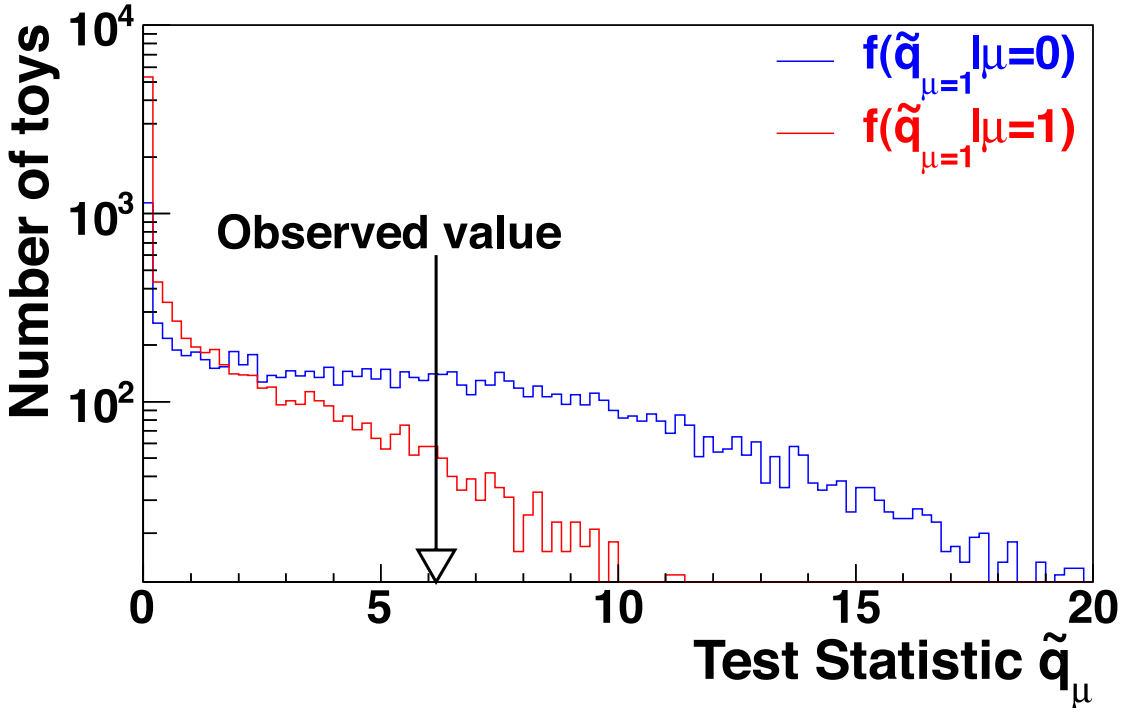

Figure 1: Test statistic distributions for ensembles of pseudo-data generated for *signal+background* and *background-only* hypotheses. See the text for definitions of the test statistic and methodology of generating pseudo-data.

$$1 - p_b \;=\; P(\,\tilde{q}_\mu \geq \tilde{q}_\mu^{obs}\,|\,\text{background-only}) \;=\; \int_{q_0^{\rm obs}}^{\infty} f(\tilde{q}_\mu|0,\hat{\theta}_0^{\rm obs})\,d\tilde{q}_\mu\,, \qquad (7)$$

and calculate $\mathrm{CL_s}(\mu)$ as a ratio of these two probabilities [1]

$$CL_s(\mu) = \frac{p_\mu}{1-p_b} \qquad (8)$$

7. If, for $\mu = 1$, $\mathrm{CL_s} \leq \alpha$, we would state that the SM Higgs boson is excluded with $(1-\alpha)$ $\mathrm{CL_s}$ confidence level (C.L.). It is known that the $\mathrm{CL_s}$ method gives conservative limits, i.e. the actual confidence level is higher than $(1-\alpha)$. See Appendix A for more details.

8. To quote the 95% Confidence Level upper limit on μ, to be further denoted as $\mu^{95\%\mathrm{CL}}$, we adjust μ until we reach $\mathrm{CL_s} = 0.05$.

2.2 Expected limits

The most straightforward way for defining the expected median upper-limit and $\pm 1\sigma$ and $\pm 2\sigma$ bands for the *background-only* hypothesis is to generate a large set of background-

[1] Note that we define p_b as $p_b = P(\,\tilde{q}_\mu < \tilde{q}_\mu^{obs}\,|\,\text{background-only})$, excluding the point $\tilde{q}_\mu = \tilde{q}_\mu^{obs}$. With these definitions one can identify p_μ with CL_{s+b} and p_b with $1 - CL_b$.

only pseudo-data and calculate CL_s and $\mu^{95\%CL}$ for each of them, as if they were real data (Fig. 2 (left)). Then, one can build a cumulative probability distribution of results by starting integration from the side corresponding to low event yields (Fig. 2 (right)). The point at which the cumulative probability distribution crosses the quantile of 50% is the median expected value. The $\pm 1\sigma$ (68%) band is defined by the crossings of the 16% and 84% quantiles. Crossings at 2.5% and 97.5% define the $\pm 2\sigma$ (95%) band.

Despite being logically very straightforward, this prescription is not too practical from the computational point of view due to the high CPU demand. If N is the number of "toys" being generated in the internal loop of calculations of the desired quantity and M is a number of pseudo-data sets for which such computation is performed, then the number of times the likelihoods would have to be evaluated in such a linear procedure is $N \cdot M$.

To save on the CPU consumption, we use the fact that the distributions of the test statistic for a given μ do not depend on the pseudo-data, so they can be computed only once. The computation of the p-values for each pseudo-data then requires the test statistic to be evaluated only once for each trial value of μ, and the total number of evaluations is proportional to $N + M$ instead of $N \cdot M$.

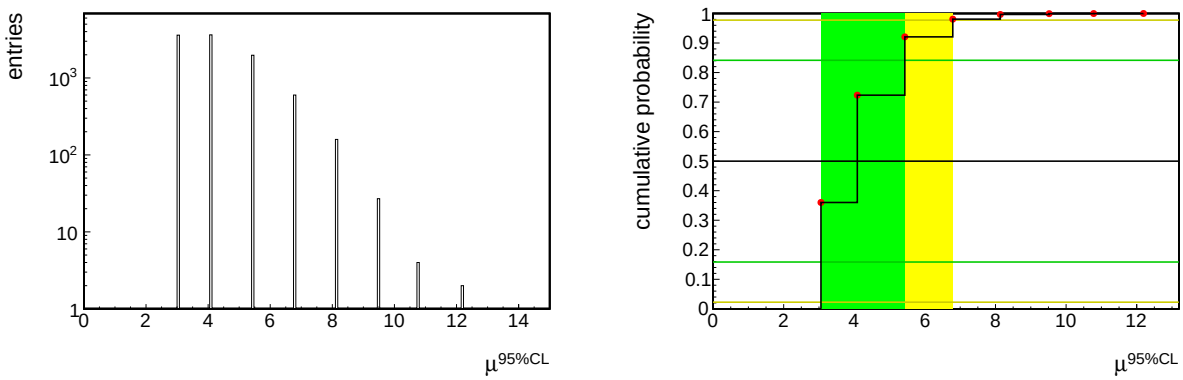

Figure 2: (Left) An example of differential distribution of possible limits on μ for the *background-only* hypothesis ($s = 1$, $b = 1$, no systematic errors). (Right) Cumulative probability distribution of the plot on the left with 2.5%, 16%, 50%, 84%, and 97.5% quantiles (horizontal lines) defining the median expected limit as well as the $\pm 1\sigma$ (68%) and $\pm 2\sigma$ (95%) bands for the expected value of μ for the *background-only* hypothesis.

3 Quantifying an excess of events for summer 2011

3.1 Fixed Higgs boson mass m_H

The presence of the signal is quantified by the *background-only* p-value, i.e. the probability for the background to fluctuate and give an excess of events as large or larger than the observed one. As before, this requires defining a test statistic and the construction of its sampling distribution. For a given Higgs boson mass hypothesis m_H, the test statistic

used is q_0:

$$q_0 \;=\; -2\ln\frac{\mathcal{L}(\text{data}|0,\hat{\theta}_0)}{\mathcal{L}(\text{data}|\hat{\mu},\hat{\theta})} \quad \text{and } \hat{\mu}\geq 0. \tag{9}$$

The constraint $\hat{\mu}\geq 0$ gives an accumulation of the test statistic at zero for events with downward fluctuations, since we are not interested in interpreting a deficit of events with respect to the expected background on an equal footing with an excess. Following the frequentist convention for treatment of nuisance parameters as discussed in Section 2, we build the distribution $f(q_0|0,\hat{\theta}_0^{\rm obs})$ by generating pseudo-data for nuisance parameters around $\hat{\theta}_0^{\rm obs}$ and event counts following Poisson probabilities under the assumption of the *background-only* hypotheses. An example of such a q_0 distribution is shown in Fig. 3. From such a distribution, one can evaluate the p-value corresponding to a given experimental observation $q_0^{\rm obs}$ as follows:

$$p_0 \;=\; P(q_0\geq q_0^{obs}) \;=\; \int_{q_0^{obs}}^{\infty} f(q_0|0,\hat{\theta}_0^{\rm obs})\,dq_0. \tag{10}$$

To convert the p-value into a significance Z, we adopt the convention of a "one-sided Gaussian tail":

$$p=\int_Z^{\infty}\frac{1}{\sqrt{2\pi}}\exp(-x^2/2)\,dx \;=\; \frac{1}{2}\,P_{\chi^2_1}(Z^2), \tag{11}$$

where, $P_{\chi^2_1}$ stands for survival function of χ^2 for one degree of freedom.

The 5σ significance ($Z=5$) would correspond in this case to $p_b \;=\; 2.8\times 10^{-7}$. Evaluation of such low probabilities may become impractical in terms of the CPU demand. The solid line in Fig. 3 is the χ^2 distribution for one degree of freedom. One can see that, by simply relying on the asymptotic behaviour of the likelihood ratio test statistic q_0, a fair *estimate* of p-values (and corresponding significances) can be obtained from the observed value $q_0^{\rm obs}$ itself, without having to generate pseudo-data [2]:

$$p^{estimate}=\frac{1}{2}\left[1-\text{erf}\left(\sqrt{q_0^{\rm obs}/2}\right)\right]. \tag{12}$$

The p-value discussed above is evaluated at a fixed m_H and can be referred to as a *local* p-value. Since we test the *background-only* hypothesis many times as we scan m_H, we must take into account this dilution effect associated with the multiple testing, also known as a trial factor or look-elsewhere effect.

3.2 Estimating the look-elsewhere effect

In the Higgs boson search, the Higgs boson mass parameter m_H is undefined for the *background-only* hypothesis, and therefore the standard regularity conditions of Wilks'

[2] In practice, it is known that such an asymptotic behaviour works very well even for cases with very few expected events.

theorem [13] do not apply. That is one cannot construct a unique test statistic encompassing all possible signals and having asymptotic χ^2-behaviour. Hence, specialised methods are required for quantifying the compatibility of a given observation with the *background-only* hypothesis.

The *global* test statistic to be associated with the search in some broad mass range can be written as follows:

$$q_0(\hat{m}_H) = \max_{m_H} q_0(m_H). \tag{13}$$

In the asymptotic regime and for very small p-values, a procedure exists and is well described in reference [14] that is largely based on Davies' result [15]. Following these references, the p-value of the global test statistic can be written as follows:

$$p_b^{global} \;=\; P(q_0(\hat{m}_H) > u) \;\leq\; \langle N_u \rangle + \frac{1}{2} P_{\chi^2_1}(u) \tag{14}$$

where $\langle N_u \rangle$ is the average number of up-crossings of the likelihood ratio scan $q_0(m_H)$ at a level u. The definition of up-crossings is illustrated in Fig. 4. The ratio of *global* and *local* p-values is often referred to as the *trial factor*.

The average number of up-crossings at two levels u and u_0 are related via the following formula

$$\langle N_u \rangle \;=\; \langle N_{u_o} \rangle \, e^{-(u-u_o)/2}, \tag{15}$$

which allows one evaluate the term $\langle N_u \rangle$ at the high level u from measuring the average number of up-crossings $\langle N_{u_o} \rangle$ at some lower reference level u_0.

When one has a well defined background model, then the number of low-threshold up-crossings $\langle N_{u_o} \rangle$ can be measured by generating a relatively small set of pseudo-data. In many analyses, such a background model indeed can be constructed. However, the use of cuts or multivariate analysis (MVA) selections optimised independently for different Higgs boson masses does not allow one to construct a background model that would be guaranteed to account for all correlations between nearby test mass points.

The foreseen way around this is to count the number of up-crossings with the data themselves. Indeed, when the look-elsewhere effect is large (and this is the only case when we really care to evaluate it), the number of up-crossings at low thresholds will be large and reasonably well measured[3]. This procedure should give us a fair estimate of the trial factor by which we need to "de-rate" the *local* p-value derived from the maximal value $q_0(\hat{m}_H)$ observed in the scan. It should be noted that there is no direct relation between the number of mass points and the trial factor since the latter is determined by the mass resolutions of the search channels.

For example, let us assume that by performing a scan over Higgs boson masses m_H, we find that the maximum value $q_0(\hat{m}_H)$ is 9, which, according to Eq. 12, gives an estimated *local* p-value of 0.13% and *local* significance of 3σ (Eq. 11). Next, let us assume that the measured number of up-crossings at level $u_o = 1$ (*local* 1σ-significance) is measured to be 8. Then, the *global* p-value corresponding to the observed excess (with the *local* p-value of 0.13% or 3σ-significance) can be derived from the Eq. 14 and is about 15%. Therefore, the trial factor for a *local* 3σ excess in this example is about 100.

[3]In the presence of a signal, this number might be biased by one unit.

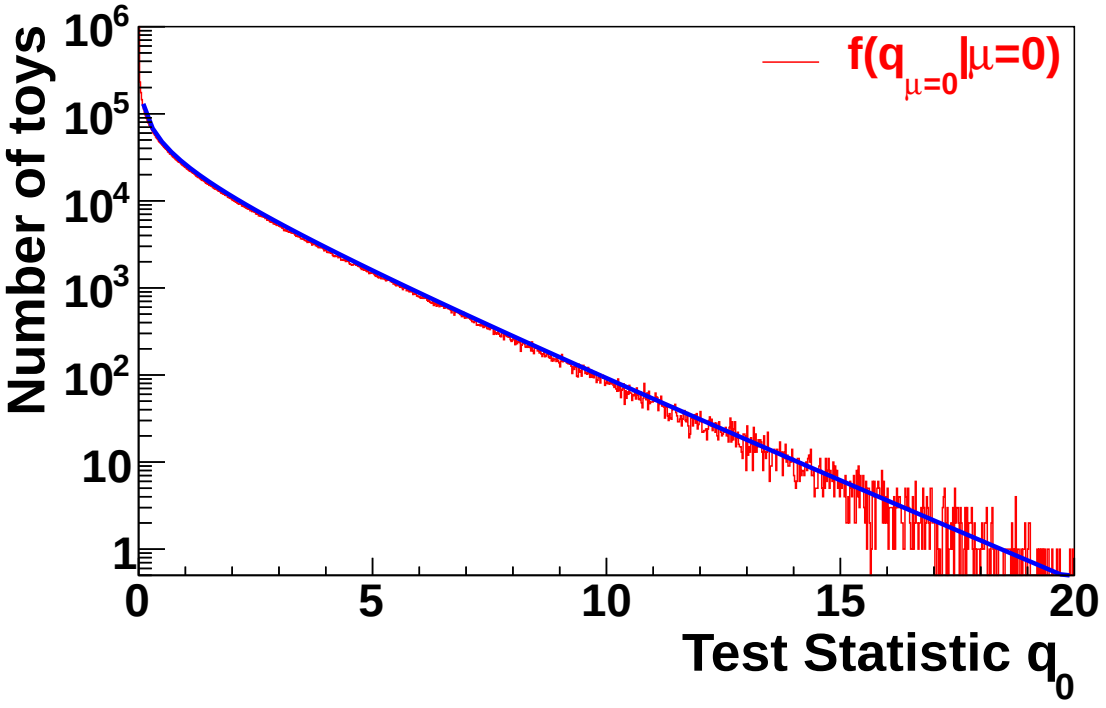

Figure 3: Distribution $f(q_0|0,\hat{\theta}_0^{\rm obs})$ of the test statistic q_0 as obtained by generating pseudo-data (toys) for the *background-only* hypotheses.

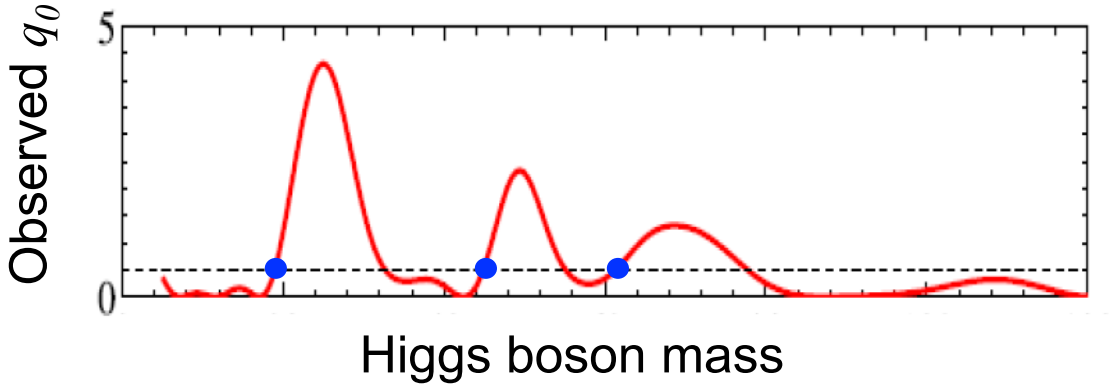

Figure 4: An illustration of a hypothetical scan of the test statistic q_0 vs m_H for some data. Up-crossings for a given threshold value u are shown with blue points.

4 Higgs mass points

The choice of mass points for the combination is driven by the $H \to 2\gamma$ and $H \to ZZ \to 4\ell$ analyses that look for a narrow peak over the continuum background. Figure 5 shows the expected $\delta m_{\gamma\gamma}$ and $\delta m_{4\mu}$ resolutions as well as the Higgs half-width $\Gamma_H/2$. The test masses in the SM Higgs search should not be much farther apart than the observable width of the Higgs peak. A simple model with a Gaussian-shaped signal and flat background shows that if we choose to step in $1\sigma_m$ increments, the loss of sensitivity for a Higgs boson with a mass right in the middle between the chosen test masses is less than 5%. With $2\sigma_m$ increments, the loss of sensitivity can be as high as 20%. The increments in the mass steps are therefore chosen to be close to $1\sigma_m$, as shown in Fig. 5. Table 1 summarizes the chosen mass points. Initially, we will not use less than 1 GeV binning until we have tuned the $H \to \gamma\gamma$ response.

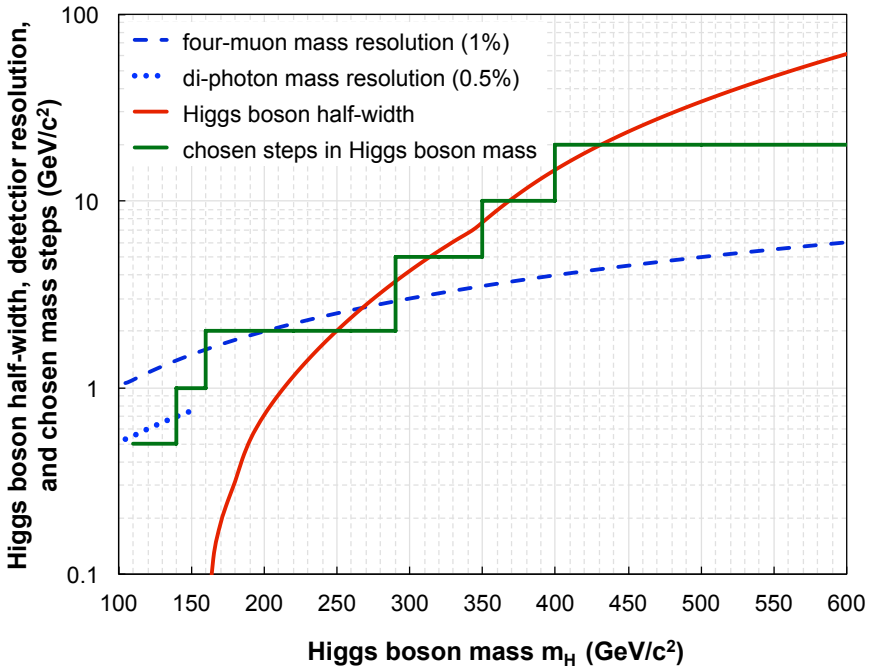

Figure 5: Expected detector resolutions for reconstructing two photons $\delta m_{\gamma\gamma}$ (blue dotted) and four muons $\delta m_{4\mu}$ (blue dashed) as well as the intrinsic Higgs half-width $\Gamma_H/2$ (red) as a function of the Higgs mass m_H. The chosen size of mass steps for the Higgs search analyses is shown in green.

Table 1: The chosen Higgs mass points for which all analyses going into the overall Higgs search combination should provide their results (within the range of their sensitivity).

Mass range (GeV/c^2)	Step size (GeV/c^2)	Number of points	Step size is driven by
110-140	0.5	61	$\delta m_{\gamma\gamma}$ for the best category of photons
140-160	1	20	$\delta m_{4\mu}$
160-290	2	65	$\delta m_{4\mu}$ and $\Gamma/2$
290-350	5	12	$\Gamma/2$
350-400	10	5	$\Gamma/2$
400-600	20	10	$\Gamma/2$ at the beginning of the range
TOTAL		173	

5 Systematic Uncertainties

5.1 Systematic uncertainty probability density functions

Systematic uncertainties on observables are handled by introducing nuisance parameters θ with a probability density function, *pdf*, $\rho(\theta)$ with some $\tilde{\theta}$ associated with the best estimate of the nuisance (e.g., mean, median, peak) and some other parameter characterising the overall shape of the *pdf*, and in particular its width. Different choices of pdf are described as follow:

- Nuisance parameters, unconstrained by any a priori considerations and/or measurements not involving the data going into the statistical analysis, are assigned flat *priors*.

- The Gaussian *pdf* is a frequent choice for systematic uncertainties. It is well-suited for describing uncertainties on parameters that can be both positive and negative:

$$\rho(\theta) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(\theta-\tilde{\theta})^2}{2\sigma^2}\right) \tag{16}$$

  Technically, an observable A with best estimate $\tilde{A}$ and the ascribed Gaussian relative uncertainties σ_A can be simulated by generating random values of θ from the *normal* distribution with $\tilde{\theta} = 0$ and $\sigma = 1$ and by writing $A = \tilde{A}\cdot(1+\sigma_A\cdot\theta)$. Two observables A and B with 100% positively correlated uncertainties—of not necessarily the same scale—can be generated by using $A = \tilde{A}\cdot(1+\sigma_A\cdot\theta)$ and $B = \tilde{B}\cdot(1+\sigma_B\cdot\theta)$. The 100% negative correlations are constructed by using $\sigma_A > 0$ and $\sigma_B < 0$.

  However, the Gaussian *pdf* is not suitable for positively defined observables like cross sections, cut efficiencies, luminosity, etc. The common (and arguably not particularly elegant) solution is to truncate the Gaussian at or just above zero.

- An alternative option is to use the log-normal *pdf* that allows one to avoid all pathologies/difficulties of the truncated Gaussian;

$$\rho(\theta) = \frac{1}{\sqrt{2\pi}\ln(\kappa)} \exp\left(-\frac{(\ln(\theta/\tilde{\theta}))^2}{2(\ln\kappa)^2}\right)\frac{1}{\theta}. \tag{17}$$

  The width of the log-normal *pdf* is characterised by κ (e.g. $\kappa = 1.10$ implies that the observable can be larger or smaller by a factor 1.10, both deviation having a chance of 16%). For small uncertainties, the Gaussian with a relative uncertainties ϵ and the log-normal with $\kappa = 1+\epsilon$ (or $\kappa = e^{\epsilon}$) are asymptotically identical, while the log-normal *pdf* is certainly a more appropriate choice for very large uncertainties (e.g. "a factor of two uncertainty" maps nicely onto log-normal with $\kappa = 2$). Figure 6 (left) shows log-normal distributions with different κ values. The log-normal distribution has a longer tail with respect to the Gaussian and goes to zero at $\theta = 0$. It is the log-normal *pdf* that is chosen for all uncertainties that are deemed to be correlated between ATLAS and CMS (see next section).

Technically, an observable A with best estimate $\tilde{A}$ and the ascribed log-normal uncertainties κ_A can be simulated by generating random values of θ from the *normal* distribution (Eq. 16) with $\tilde{\theta} = 0$ and $\sigma = 1$ and by writing $A = \tilde{A} \cdot \kappa_A^\theta$. Two observables A and B with 100% positively correlated uncertainties—of not necessarily the same scale—can be generated by using $A = \tilde{A} \cdot \kappa_A^\theta$ and $B = \tilde{B} \cdot \kappa_B^\theta$. The 100% negative correlations are constructed by using $\kappa_A > 1$ and $\kappa_B < 1$.

- The gamma distribution is adopted for describing statistical uncertainties associated with a number of Monte Carlo events in simulation (after applying all cuts) or a number of observed events in a data control sample. In both cases, the event rate n in the signal search region can be related to the number of events N in MC or data via a simple relationship $n = \alpha \cdot N$. Ignoring uncertainties on α that are to be dealt with separately, the uncertainties on the predicted rate n associated with the observation of N events is described by the gamma distribution as given by Eq. 18:

$$\rho(n) = \frac{1}{\alpha} \, \frac{(n/\alpha)^N}{N!} \, \exp(-n/\alpha). \tag{18}$$

  This form can be easily derived using the Bayesian methodology and assuming that the prior $\pi(n)$ is flat. The most probable value for n is αN, the mean value is $\alpha(N+1)$, and the dispersion is $\alpha\sqrt{N^2+1}$. Note that $N = 0$ is a perfectly allowable situation, resulting in the exponential *pdf* for n, with the maximum at $n = 0$, mean $= \alpha$, and dispersion $= \alpha$. Gamma distributions with different numbers of events observed in control samples are shown in Fig. 6 (right).

  Uncertainties modelled by gamma distributions can be found in both ATLAS and CMS analyses, but they are never correlated between ATLAS and CMS, nor would they be unless both experiments would decided to rely on the very same observations.

  The mapping between Bayesian posterior *pdf*s $\rho(\theta|\tilde{\theta})$ and corresponding frequentist auxiliary measurement *pdf*'s $p(\tilde{\theta}\,|\,\theta)$ as discussed in Section 2 and represented by Eq. 1 for the uncertainties discussed in this section is shown in Table 2.

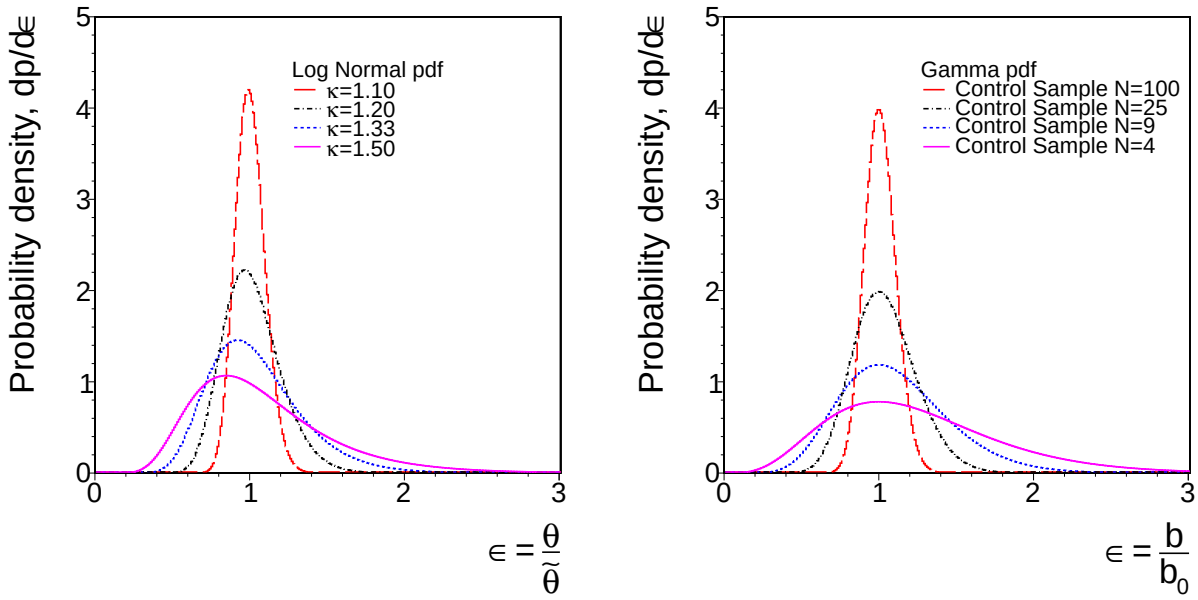

Figure 6: (Left) Log-normal distributions with $\kappa = 1.10$, 1.20, 1.33 and 1.50. (Right) Gamma distribution with the number of events in a control sample $B = 100$, 25, 9 and 4.

Table 2: Mapping between Bayesian posterior *pdf*s $\rho(\theta|\tilde{\theta})$ and corresponding frequentist auxiliary measurement *pdf*'s $p(\tilde{\theta}\,|\,\theta)$ and "primordial" prior $\pi_\theta(\theta)$ as discussed in Section 2 and represented by Eq. 1 for the uncertainties discussed in this section.

Type of uncertainties	Bayesian posterior $\rho(\theta\|\tilde{\theta})$	Frequentist $p(\tilde{\theta}\,\|\,\theta)$	Prior $\pi_\theta(\theta)$
Unconstrained	flat	flat	flat
Gaussian/Log-normal	$\rho(\theta\,\|\,\tilde{\theta}) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{(\theta-\tilde{\theta})^2}{2}\right)$	$p(\tilde{\theta}\,\|\,\theta) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{(\tilde{\theta}-\theta)^2}{2}\right)$	flat
Statistical uncertainties	$\rho(\theta\,\|\,N) = \frac{\theta^N}{N!}\ \exp(-\theta)$	$p(N\,\|\,\theta) = \frac{\theta^N}{N!}\ \exp(-\theta)$	flat

5.2 Uncertainties correlated between experiments

Currently, we identify four main groups of such correlated uncertainties that we associate with:

- PDF+α_s uncertainties
- theoretical renormalisation/factorisation scale uncertainties
- uncertainties in modelling underlying event and parton showering
- experimental uncertainties on luminosities

Theoretical uncertainties can be looked at from three different points of view:

- Uncertainties on the total cross sections σ_{tot}. These are an important starting point. However, they are not necessarily applicable to actual physics analyses where various experimental cuts restrict the final phase space.
- Uncertainties on the acceptance $\mathcal{A}$. These are very important for analyses aiming at setting limits on overall cross sections from measurements performed in a restricted phase space.
- Uncertainties on the cross section within the limited acceptance, i.e. $\mathcal{A} \cdot \sigma_{tot}$. These uncertainties are needed when one attempts to set limits by combining analyses of varying sensitivity for different Higgs production mechanisms. A priori, the level of correlations between uncertainties on $\mathcal{A}$ and σ_{tot} is not known.

5.2.1 Naming convention

Nuisance parameters with the same name appearing in different analyses (within one or both experiments) are taken to be 100% correlated. Different names imply no correlations. Any two sources of uncertainties that are believed to be only partially correlated are either broken further down to the independent sub-contributions or declared to be correlated/uncorrelated, whichever is believed to be more appropriate or more conservative.

To avoid accidental correlations in the combination of two experiments, uncertainties specific to each experiment will have a prefix ATLAS or CMS. Uncertainties without such prefixes are assumed to be 100% correlated between the two experiments.

5.2.2 Total cross sections

Breaking up systematic uncertainties associated with PDF+α_s uncertainties into truly independent sources would imply painstaking work with nearly no impact on the final results. Also, this option does not really work in the context of taking envelopes of multiple PDF sets as prescribed by the LHC Higgs Cross Section group. The other possible extreme is to have all processes bluntly 100% correlated. This appears to be too simplistic. As a compromise, we adopt the following approximation.

First, we group all processes in three categories based on the prevailing production source. Then, we assume that PDF+α_s systematic uncertainties between all processes in one group are 100% positively correlated and not correlated between processes from different groups. This results in three nuisance parameters as shown in Table 3. The detailed matrix of PDF uncertainty correlations, as calculated by the CTEQ collaboration [16], can be found in Appendix B. It shows that the chosen scheme for correlating PDF uncertainties between different processes is fair. In those cases where we see sizable deviations, the adopted scheme generally implies more conservative results.

We assume that all physics processes have uncorrelated QCD scale uncertainties, except for a few very closely related processes (W/Z, WW/WZ/ZZ) that we treat as 100% correlated. The list of independent nuisance parameters characterising theoretical uncertainties in cross section calculations is given in Table 3.

The cross section uncertainties for the Higgs boson production are taken from the LHC Higgs Cross Section Group report [17]. The PDF+α_s and the renormalisation/factorisation scale uncertainties are treated separately. The prescription recommended by the LHC Higgs Cross Section Group [17] will be considered in the future.

5.2.3 Acceptance uncertainties

For setting limits on a total cross section times branching ratio of a particular production mechanism and decay mode of a signal, one is interested in the uncertainties on the acceptance $\mathcal{A}$, which is the ratio of (cross section with cuts) / (full cross section). Depending on the cuts, some uncertainties may cancel out in this ratio, while others may remain independent.

Uncertainties of a similar type arise when one uses a data-driven technique for evaluating some particular background event rate n in a signal region by extrapolating from an observation of N events in a control region. The two can be related via a so-called extrapolation factor α: $n = \alpha \cdot N$. When the extrapolation factor is derived from theory/MC, α = (cross section with cut set A) / (cross section with cut set B).

Given that the cuts are ever evolving entities, calculations of the acceptance and extrapolation factor uncertainties are to be performed within the ATLAS and CMS Higgs physics groups.

We currently assume that the acceptance and extrapolation factor uncertainties are independent from the total cross section uncertainties, except for the acceptance associated with jet counting in the $gg \to H \to WW + 0/1/2$-jets analyses. This exception is discussed in the next section.

The naming convention for such uncertainties is AAA_BBB_ACCEPT or AAA_BBB_EXTRAP, where AAA identifies the original source of uncertainty (pdf, QCDscale, UEPS), while BBB gives an indication of what process or method the uncertainty is associated with with.

Should ATLAS and CMS use similar cuts and techniques, the uncertainties will be assumed to be 100% correlated between the two collaborations. This will have to be decided on a case-by-case basis. At this stage, in the context of extrapolation factors, we identify two very similar data-driven techniques used by ATLAS and CMS for predicting WW and $t\bar{t}$ background contributions in the $H \to WW \to 2\ell 2\nu + 0$-jets signal regions. The uncertainties, listed in Table 3, are dominated by QCD scale uncertainties.

5.2.4 Cross section times acceptance uncertainties for $gg \to H + 0/1/2$-jets

As discussed in the previous section, uncertainties on acceptance of all cuts except for jet counting are treated as independent from the total cross section. Most of the time, being so much smaller than the total cross section uncertainties, such sub-leading acceptance uncertainties can actually be neglected.

However, the uncertainties associated with jet counting in the $gg \to H + 0/1/2$-jets sub-processes, i.e., the fractions of events falling into the 0-, 1-, and 2-jet bins, are very sensitive to the choice of QCD scales. In fact, the *exclusive* $0/1/2$ jet bin cross sections uncertainties are larger than the total cross section uncertainty and have both negative and positive correlations. The LHC Higgs Cross Section Group recommends that it is the inclusive cross sections for $gg \to H+ \geq 0$-jets, $gg \to H+ \geq 1$-jets, $gg \to H+ \geq 2$-jets that have independent theoretical uncertainties. Hence, one can find the three corresponding nuisance parameters in Table 3. The procedure of propagating *inclusive* cross section uncertainties into *exclusive* 0, 1, and 2-jet bins is described in Appendix C.

5.2.5 Uncertainties in modelling underlying event and parton showering

Besides already discussed PDFs and QCD scales, uncertainties in modeling the underlying event (UE) activity and parton showering (PS) are yet another potential source of uncertainties in evaluation of acceptance and extrapolation factors. The current prescription for their evaluation is to compare results obtained with UE/PS modeling available in different generators (e.g. Pythia, Herwig, Sherpa). Note that the primary interaction ME generator does not have to be the same as a UE/PS generator (e.g., it could be Powheg). The log-normal parameter κ is defined as follows:

$$\kappa = \frac{\text{Yield[ME + UE/PS(generator B)]}}{\text{Yield[ME + UE/PS(generator A)]}}. \tag{19}$$

5.2.6 Instrumental uncertainties

For now, luminosity uncertainties are the only instrumental uncertainties that we take as 100%-correlated between ATLAS and CMS. In time, the luminosity uncertainties may be split into correlated and uncorrelated components.

Table 3: List of nuisance parameters for systematic uncertainties assumed to be 100% correlated between ATLAS and CMS.

PDF+α_s uncertainties

nuisance	groups of physics processes
pdf_gg	$gg \to H$, $t\bar{t}H$, VQQ, $t\bar{t}$, tW, tb (s-channel), $gg \to VV$
pdf_qqbar	VBF H, VH, V, VV, $\gamma\gamma$
pdf_qg	tbq (t-channel), γ+jets

QCD scale uncertainties

nuisance	groups of physics processes
QCDscale_ggH	total inclusive $gg \to H$
QCDscale_ggH1in	inclusive $gg/qg \to H+ \geq 1$ jets
QCDscale_ggH2in	inclusive $gg/qg \to H+ \geq 2$ jets
QCDscale_qqH	VBF H
QCDscale_VH	associate VH
QCDscale_ttH	$t\bar{t}H$
QCDscale_V	W and Z
QCDscale_VV	WW, WZ, and ZZ up to NLO
QCDscale_ggVV	$gg \to WW$ and $gg \to ZZ$
QCDscale_ZQQ	Z with heavy flavor $q\bar{q}$-pair
QCDscale_WQQ	W with heavy flavor $q\bar{q}$-pair
QCDscale_ttbar	$t\bar{t}$, single top productions are lumped here for simplicity

Phenomenological uncertainties

nuisance	groups of physics processes
UEPS	all processes sensitive to modeling of UE and PS

Acceptance uncertainties

nuisance	comments
QCDscale_WW_EXTRAP	extrap. factor α for deriving WW bkgd in HWW analysis
QCDscale_ttbar_EXTRAP	extrap. factor α for deriving $t\bar{t}$ bkgd in HWW analysis

Instrumental uncertainties

nuisance	comments
lumi	uncertainties in luminosities

6 Format of presenting results

The results of the ATLAS + CMS Higgs search combination will be presented in the following forms

- A scan of *local* p-values, i.e. probabilities $P(q_0 \geq q_0^{obs}|m_H)$, vs test Higgs boson mass m_H will characterise how significant upward departures in the observed values of q_0^{obs} approximately are. We refer to these as *local* (and use "approximately" in the above sentence), since these p-values do not include the overall trial factor associated with the look-elsewhere effect. Figure 7 gives an example of such a scan. We will show approximate p-values as derived from the asymptotic χ^2-like distribution expected for q_0 as given by Eq. 12. When practical, the *local* p-values will be calculated by using toys according to Eq. 10.

- The look-elsewhere effect will be quantified following the procedure described in Sec. 3.2.

- The CL_s scan vs test Higgs boson mass, similar to the one shown in Figure 8 [18] (this plot is borrowed from the Spring 2011 Tevatron Higgs search combination), will quantify the confidence levels at which the Standard Model Higgs boson is excluded for different m_H hypotheses. The median expected CL_s values together with $\pm1\sigma$ and $\pm2\sigma$ bands will be also presented. Higgs boson masses for which $\mathrm{CL}_s < \alpha$ will be said to be excluded at the $(1-\alpha)$ confidence level.

- 95% C.L. limits $\mu^{95\%\mathrm{CL}}$ on the Higgs boson production cross section strength modifier μ vs test mass m_H, similar to the one shown in Figure 9, will be also presented, together with the median expected and $\pm1\sigma$ and $\pm2\sigma$ bands. This plot shows by what factor the SM Higgs boson cross section must be modified to be excluded at 95% C.L.

The numerical summary of the obtained results will be presented in the following form:

Table 4: Numerical results of the ATLAS+CMS Higgs search combination. Observed values are shown in bold font, expected—in plain font.

m_H	*local* p-value		$CL_s(\mu=1)$	$\mu^{95\%\mathrm{CL}}$					
(GeV/c^2)	**from toys**	**approx.**	**obs** (exp)	**obs**	-2σ	-1σ	median	$+1\sigma$	$+2\sigma$
110	**xxx**	**xxx**	**xxx** (xxx)	**xxx**	xxx	xxx	xxx	xxx	xxx
600	**xxx**	**xxx**	**xxx** (xxx)	**xxx**	xxx	xxx	xxx	xxx	xxx

What is presented here is the minimum of information. The experiments may agree to show additional information to illustrate and support the interpretation of the results.

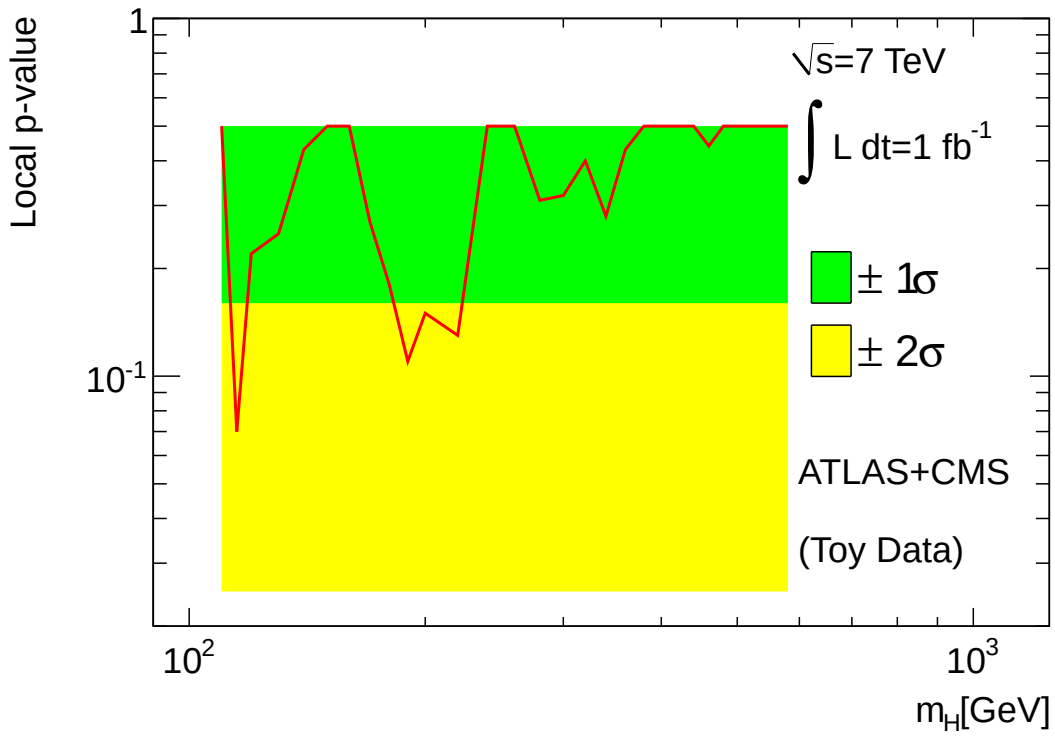

Figure 7: *local* p-value scan vs m_H. This plot does not correspond to any MC or data analysis. To help guide the eye, the n-sigma significance levels are highlighted with colour bands.

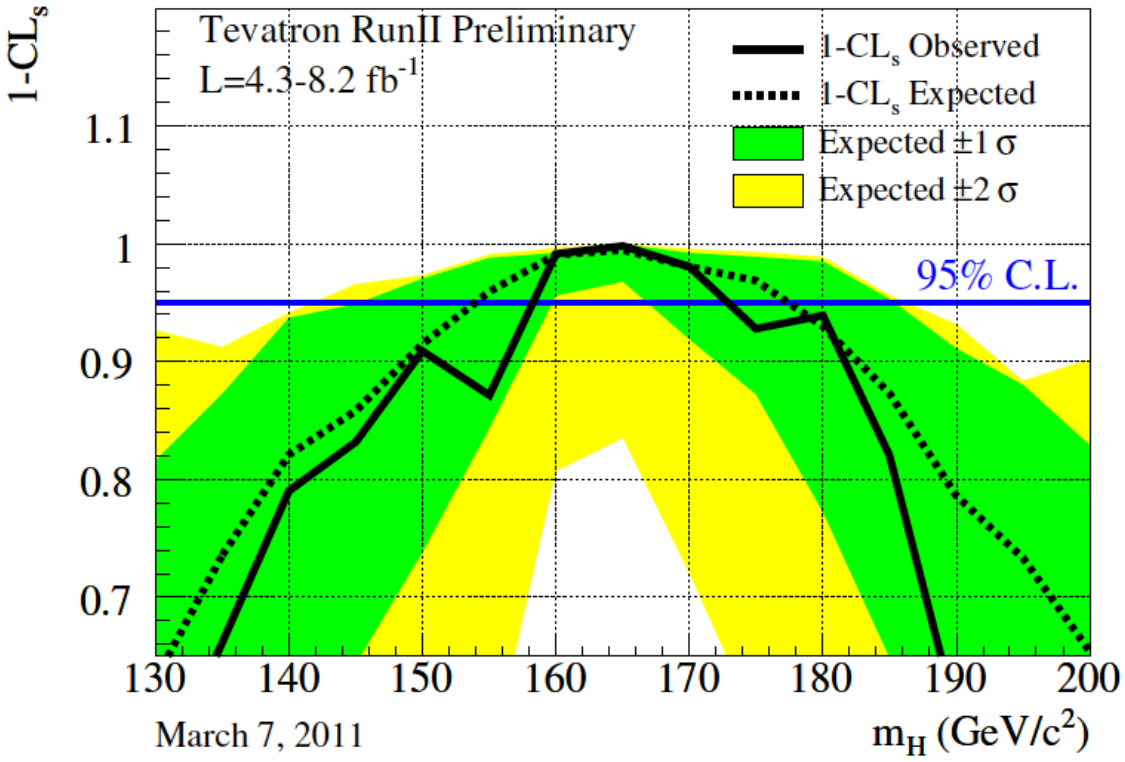

Figure 8: CL_s scan vs m_H. The solid line shows the observed values of $(1 - \mathrm{CL}_s)$. The green/yellow bands indicate $\pm 1\sigma$ and $\pm 2\sigma$ intervals for the expected values under the *background-only* hypothesis. The median expectation is shown with the dashed line.

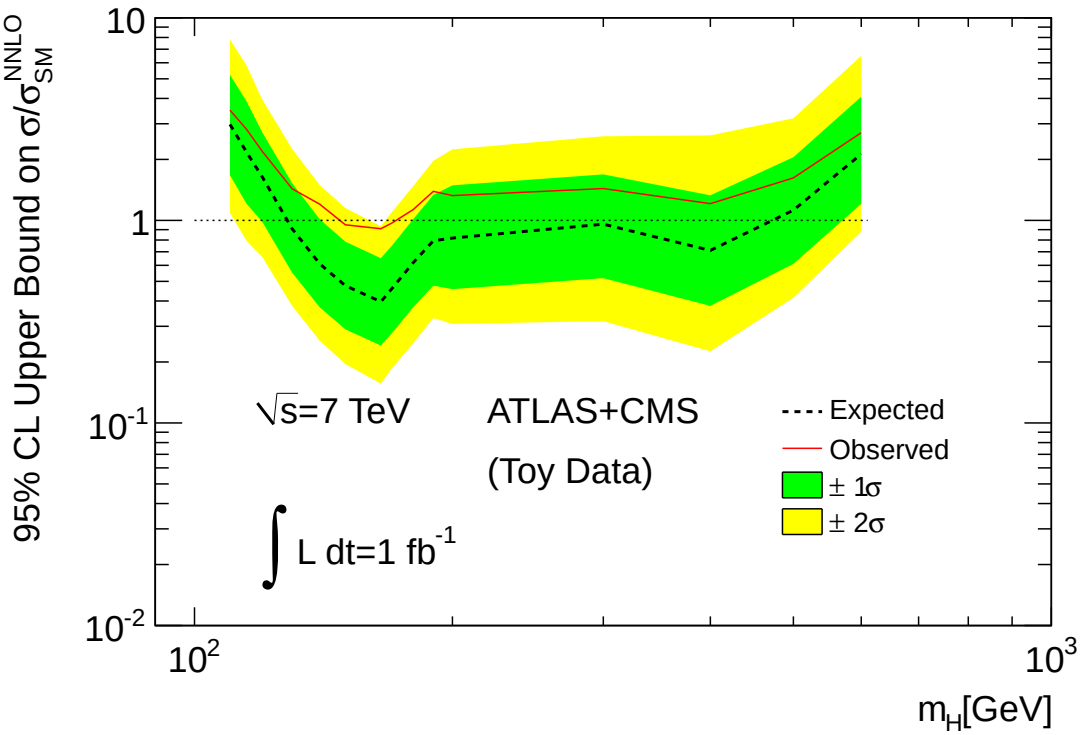

Figure 9: 95% C.L. limits $\mu^{95\%\mathrm{CL}}$ on the Higgs boson production cross section strength modifier μ ($\sigma = \mu\sigma_{SM}^{NNLO}$) vs Higgs boson mass m_H. This plot does not correspond to any MC or data analysis. The solid line shows the observed limit. The green/yellow bands indicate $\pm 1\sigma$ and $\pm 2\sigma$ intervals for the expected limits under the *background-only* hypothesis. The median expectation is shown with the dashed line.

7 Technical combination exercises (validation and synchronisation)

This section describes the Higgs combinations of ATLAS and CMS toy data that were performed to exercise the combination tools and framework. Both ATLAS and CMS have chosen to work in the common framework of RooStats [19]. It provides a common platform for exchanging so-called *Workspaces* that contain all the information needed for the statistical analyses and simplifies the logistic of data exchange between collaborations. Moreover, RooStats offers a diverse set of statistical methods that one can exercise starting from the very same workspace. Having all these benefits, the package is still under development, to which we have contributed by providing quick feedback based on the results of our exercises. More technical details on RooStats can be found in Appendix D

In order to validate and synchronise calculations of the desired quantities, the combination exercise proceeded as follows. ATLAS and CMS prepared their own Workspaces for a given analysis or combination of analyses. All analysis models were based on toy pseudo-data. No real data were involved in these exercises. Then, each collaboration would perform statistical analysis on its own workspace, on the workspace of the other collaboration, and then would build its own ATLAS+CMS combined workspace and perform statistical analysis on it. The three results (ATLAS-only, CMS-only, ATLAS+CMS) obtained by each collaboration were required to match within the quoted statistical precision of the calculations.

The statistical methods used were as follows:

- Exclusion limits obtained by using the *Profile Likelihood approximation* (see Appendix A.1.3) are the very first step of synchronisation. Although this method does not give accurate exclusion limits, it is very fast computationally, which allowed us to validate that joint likelihoods independently built by ATLAS and CMS from the single-experiment inputs are indeed identical. It is these joint likelihoods that are at the heart of the final statistical methods adopted for the Summer 2011 combination. For synchronisation purposes, we use "limits" on μ as given by Eq. 33 in Appendix A.

- Exclusion limits obtained with the *LEP-type CL_s prescription* (see Appendix A) are the next step toward the final version of the CL_s construction. Since the LEP approach does not involve profiling of nuisance parameters, these calculations are relatively fast as well.

- Exclusion limits obtained with the *LHC-type CL_s prescription* (see Sec. 2) that have been agreed on for the Summer 2011 combination were the final step of synchronisation. This approach now involves profiling of systematic errors and requires substantial CPU power. In calculations of limits on the signal strength modifier μ, one goes via steps of assessing values of the test statistic q, p-values for *signal+background* and *background-only* and their ratio CL_s, which makes the full suite of quantities that would be needed for presenting the statistical interpretation of the Higgs boson search combination.

Since both ATLAS and CMS used the same underlying RooFit and RooStats code, the scope of crosschecks across the two collaborations may be thought to be somewhat limited. However, this procedure has proved to be very useful and allowed us to validate and debug the way the combined models are constructed starting from the ATLAS and CMS models and how the basic RooStats and RooFit libraries are used.

As a separate crosscheck, all CMS-only results have been validated using the independent code L&S [20] that does not rely on RooStats and uses RooFit in a very limited capacity for functional *pdf*s.

Whenever disagreements of results were observed, we were able to track them down to either plain bugs or more subtle misinterpretations of the input information provided by the collaborations. In other words, the technical synchronisation exercise proved to be extremely valuable and prepared us for the forthcoming combinations with the 2011 data.

7.1 $H \to WW \to \ell\ell\nu\nu + 0jets$

The first combination exercise undertaken used toy analyses for the simplest $H \to WW$ channel in the di-leptonic final state with no hadronic jets. The goal of this exercise was to perform a first exchange of inputs and produce a combined exclusion limit in which some systematic uncertainties were treated as correlated across the experiments.

Model details

For this exercise, the measurements in both experiments were treated as multichannel counting experiments. The likelihood function is therefore written as the product of Poisson terms for each channel times the product of all the constraint terms for the nuisance parameters θ associated to the systematic uncertainties.

$$\mathcal{L} = \prod_{i \in \text{obs.}} \text{Poisson}\left(n_i \,|\, \nu_i(\mu, \theta)\right) \cdot \prod_{j \in \text{nui.s}} \text{Constraint}(\theta_j, \tilde{\theta}_j) \tag{20}$$

For convenience, the θ_i are normalised so that the constraint is always a normal distribution with zero mean and unit variance, and all non-universal terms enter only in the relationship between parameters and expected yield in the signal regions $\nu_i(\mu,\theta)$. For uncertainties related to the statistical uncertainty in the control regions or in the simulation, the associated nuisance parameter is the expected yield in that region, and the constraint term is a Poisson likelihood for $\tilde{\theta}_j$ observed events and θ_j expected ones; this is mathematically equivalent to a Gamma distribution over θ_j with most probable value $\tilde{\theta}_j$.

The correlation of the uncertainties across the experiments is implemented by using the same nuisance parameter θ_i to describe the same uncertainty in the two models[4]. The combined likelihood is constructed by multiplying together the two likelihoods removing the duplicated constraint terms from correlated uncertainties.

In this first exercise, only two sources of systematic uncertainties were treated as correlated: the normalisation of the luminosity, driven by machine-dependent uncertainties, and the inclusive Higgs production cross section through the gluon fusion process, driven by theoretical uncertainties (the contribution from other production modes to the $H \to WW + 0j$ final state is negligible).

The ATLAS model has 3 signal channels and 3 main control regions that enter the likelihood directly as observables, plus other sidebands that are modelled as constraints. It contains 17 ATLAS-specific nuisance parameters, plus the two associated with the luminosity and Higgs production cross section. The CMS model has 4 signal channels corresponding to the leptonic final states; measurements from sidebands enter the likelihood only through constraint terms for the nuisance parameters. In total it contains 35 CMS-specific nuisance parameters plus the two correlated ones.

[4]Only multiplicative corrections are considered to be eligible for correlations: we assume that sidebands or simulated samples are private to each collaboration and therefore the associated uncertainties are uncorrelated.

Obtained Results

At the time of the exercise, no decision had yet been taken on the preferred statistical method for computing the exclusion limit at LHC. To make the exercise possible, we therefore decided to use two simple and well established methods, for which statistical code was available in the two collaborations: the profile likelihood asymptotic approximation, and the LEP-like hybrid method. The two methods are described in detail in Appendix A.

Three combination "handshakes" have been performed:

- Observed limits for each experiment separately and for the combination for a range of mass values, using the profile likelihood asymptotic approximation. The results computed by the two collaborations are in perfect agreement (Table 5).

- CL_s values for SM Higgs ($\mu = 1$) hypotheses, computed with the LEP-type CL_s method. The results were found to be in agreement within the computational accuracy given by the number of toy experiments used, 10^4 (Table 6).

- Observed limit for the combined model at $m_H = 140$ GeV/c^2 computed with LEP-type CL_s method to better than 1% computational accuracy. The result computed by the two collaborations are in a good agreement: 0.766 ± 0.006 from CMS, 0.7673 ± 0.0014 from ATLAS.

Table 5: $H \to WW + 0j$ combination exercise: computed exclusion limits on $\mu = \sigma/\sigma_{SM}$ with the profile likelihood asymptotic approximation. The agreement is better than one per mil.

m(H)	**ATLAS computation**			**CMS computation**		
GeV/c^2	Comb.	ATLAS	CMS	Comb.	ATLAS	CMS
120	3.968	3.734	6.709	3.968	3.734	6.709
130	1.601	1.652	2.493	1.601	1.652	2.493
140	0.828	1.041	1.186	0.828	1.041	1.186
150	0.451	0.784	0.551	0.451	0.784	0.551
160	0.314	0.555	0.369	0.314	0.555	0.369
170	0.290	0.653	0.314	0.290	0.653	0.314
180	0.327	0.811	0.357	0.327	0.811	0.357
190	0.623	1.211	0.742	0.623	1.211	0.742
200	0.861	1.661	1.017	0.861	1.661	1.017

Table 6: $H \to WW + 0j$ combination exercise: computed CL_s values for the SM Higgs ($\mu = 1$) hypotheses with LEP-type CL_s method. The agreement is within the quoted computational precision. The "-" indicates that the information is not available. The 0 corresponds to $< 10^{-4}$.

m(H)	**ATLAS computation**			**CMS computation**		
GeV/c^2	Comb.	ATLAS	CMS	Comb.	ATLAS	CMS
120	0.597 ± 0.008	0.578 ± 0.010	0.812 ± 0.006	0.586	-	0.806
130	0.154 ± 0.004	0.240 ± 0.007	0.389 ± 0.006	0.166	0.237	0.392
140	0.014 ± 0.002	0.087 ± 0.004	0.052 ± 0.003	0.015	0.088	0.056
150	0.0004 ± 0.0003	0.033 ± 0.003	0.0013 ± 0.0005	0.000	0.031	0.001
160	0	0.005 ± 0.001	0	0.000	0.005	0.000
170	0	0.012 ± 0.002	0	0.000	-	0.000
180	0	0.037 ± 0.003	0	0.000	0.038	0.000
190	0.005 ± 0.001	0.148 ± 0.005	0.011 ± 0.002	0.005	0.135	0.011
200	0.027 ± 0.002	0.242 ± 0.007	0.048 ± 0.003	0.025	0.234	0.050

7.2 $H \to WW \to \ell\ell\nu\nu + 0/1/2 - jets$

The second technical combination exercise was again used the $H \to WW$ analysis in the di-leptonic final state, but now also considered those categories of events with one and two jets. The goal of this second exercise was to have a better treatment of all the systematic uncertainties of theoretical origin, and to increase the complexity of the model.

Model Details

The two analyses were still modelled as multi-channel counting experiments, so the likelihood function had the same structure as in the previous exercise.

The systematic uncertainties considered for correlations across the experiments were:

- the scale of the luminosity measurement;
- the effect of PDF uncertainties on the production cross sections, handled separately for the processes dominated by the three partonic initial states gg, qq, $q\bar{q}$, and gq,
- the uncertainties on the cross sections coming from higher orders, estimated varying the renormalisation and factorisation scales. These uncertainties were accounted for separately for $gg \to H$, VBF H, associated $H + W/Z$ production and for the backgrounds $q\bar{q} \to V$ ($V = W/Z$), $qq \to VV$, $gg \to VV$ and $t\bar{t}$.

For simplicity, the backgrounds from single top and the associated $t + W$ production were treated as part of the larger $t\bar{t}$ background. For the ATLAS model, the scale uncertainties for WW and $t\bar{t}$ were further separated into the uncertainty on the inclusive cross section and the uncertainty on the extrapolation between signal region and sideband, and the two terms were treated as uncorrelated. When combining the two likelihoods in this exercise, the uncertainties on the inclusive WW, $t\bar{t}$ cross sections from the ATLAS model have been taken as correlated with the uncertainty on the accepted cross section for the same processes in the CMS model.

The ATLAS model included 9 signal channels and 12 control channels treated as observables. There are 24 ATLAS-specific nuisance parameters plus 13 theoretical uncertainties eligible for correlation with CMS.

The CMS model included 9 signal channels, and control regions were included only through constraints terms. There are 32 CMS-specific parameters plus 11 theoretical uncertainties eligible for correlation with ATLAS.

Eventually the combined model contains 70 nuisance parameters of which 10 are correlated across the two experiments. Four parameters are eligible for correlation but were not correlated for lack of a counterpart in the other model because it was considered negligible (PDF uncertainty for gq processes, scale uncertainty on the $H + W/Z$ process) or because the uncertainties were factorised differently (WW and $t\bar{t}$ as described earlier).

Obtained Results

For this exercise, one Higgs mass point was considered, namely 140 GeV/c^2. The same three handshakes as for the previous exercise were done:

- Exclusion limits on $\mu = \sigma/\sigma_{SM}$ from the profile likelihood approximation (Table 7). The agreement is better than one per mil.
- CL_s values for SM Higgs hypothesis in the hybrid LEP-like approach (Table 8). The agreement is within the quoted computational precision.
- Exclusion limit for the combined models at $m(H) = 140$ GeV/c^2 computed with LEP-type CL_s method to better than 1% computational accuracy. The agreement between the result computed by ATLAS, 0.519 ± 0.003, and by CMS, 0.508 ± 0.003, was considered satisfactory[5].

Table 7: $H \to WW + 0/1/2j$ combination exercise: computed exclusion limits on $\mu = \sigma/\sigma_{SM}$ at $m(H) = 140$ GeV/c^2 with the profile likelihood asymptotic approximation.

Model	**ATLAS computation**	**CMS computation**
ATLAS	0.802547	0.802548
CMS	0.426186	0.426186
Combined	0.355680	0.355681

Table 8: $H \to WW + 0j$ combination exercise: computed CL_S values for the SM Higgs ($\mu = 1$) hypotheses with LEP-type CL_s method.

Model	**CMS computation**	**ATLAS computation**
ATLAS	0.1036 ± 0.0018	0.1075 ± 0.0050
CMS	0.0009 ± 0.0003	0.0016 ± 0.0011
Combined	0.0014 ± 0.0003	0.0032 ± 0.0011

[5]The discrepancy would be 2.5 standard deviations. However, the values of μ are determined from an interpolation from a grid of tested μ values, and the reported uncertainties include only the statistical uncertainties on the CL_S values for each grid point and not a systematic uncertainty from the choice of interpolation model.

7.3 $(H \to WW) + (H \to \gamma\gamma) + (H \to ZZ \to 4\ell)$

The third combination exercise used a significantly more complex model, in which also the $H \to \gamma\gamma$ and $H \to ZZ \to 4\ell$ channels have been considered. The goals of these exercises were to test models in which the distribution of a continuous variable like the di-photon mass is used in the computation of the limit.

Model Details

For the two latter channels, the analyses are modelled as a search for an excess in the $\gamma\gamma$ and 4ℓ invariant mass distributions. In each channel i, the data are modelled as a sum of signal and background components j with the expected normalisations $\nu_{i,j}(\mu,\theta)$ and shapes $f_{i,j}(m|\theta)$:

$$f_i(m|\mu,\theta) = \sum_j \frac{\nu_{i,j}(\mu,\theta)}{\nu_i(\mu,\theta)} \cdot f_{i,j}(m|\nu,\theta) \qquad \nu_i^{\text{tot}} = \sum_j \nu_{i,j}(\mu,\theta). \tag{21}$$

The negative logarithm of the likelihood function for a single channel can be summed over the observed events as

$$-\log \mathcal{L}_i = \sum_{e=1}^{n_i} [-\log f_i(m_e|\mu,\theta)] + n_i \log(\nu_i^{\text{tot}}) - \nu_i^{\text{tot}}, \tag{22}$$

up to terms depending only on n_i which would cancel out when taking the ratio of two likelihood functions for the same data but different values of μ and θ.

The overall likelihood is then built as the product of the individual likelihoods and of the constraint terms just like in the counting experiment case.

It is technically convenient to treat all channels entering the combination in an uniform way. Therefore the $H \to WW$ counting experiment has been re-written introducing a dummy variable x with range $[0, 1]$ and taking all $f_{i,j}(x)$ to be equal to the uniform distribution; this new expression is completely equivalent to the one using Poisson likelihoods.

The models included in this combination were: the ATLAS and CMS $H(\to WW \to \ell\ell\nu\nu) + 0/1/2j$ models of the previous exercises, the ATLAS and CMS $H \to \gamma\gamma$ models, and a CMS $H \to ZZ \to 4\ell$ model [6]. The combined model contains about 5800 unbinned events separated in 37 exclusive categories. There are in total 98 nuisance parameters, 10 of which are correlated across the experiments like in the previous combination exercise).

Obtained Results

Just like in the previous exercise, only a single Higgs mass point was considered, $m_H = 140$ GeV/c^2. Similar handshakes to those of the previous exercise were done: exclusion limits on $\mu = \sigma/\sigma_{SM}$ from the profile likelihood approximation for all the channels separately and for the combination, and the exclusion limit for the combined model using the

[6] There was an initial technical issue with the implementation of the ATLAS $H \to ZZ \to 4\ell$ model at the time, so it was left out at the beginning to allow the exercise to proceed.

LEP-type CL_s Bayesian-frequentist method. The results for the profile likelihood approximation are in excellent agreement (Table 9), and the hybrid Bayesian-frequentist ones agree within their computational accuracies (0.636 ± 0.005 from ATLAS, 0.626 ± 0.004 from CMS).

After the ATLAS $H \to ZZ \to 4\ell$ toy model became available, we exercised limit calculations of the ultimate *LHC-type* CL_s method as defined in Section 2. Results of calculations agree within the computational precision and are shown in Table 10.

Table 9: $(H \to WW) + (H \to \gamma\gamma) + (H \to ZZ \to 4\ell)$ combination exercise: exclusion limits on $\mu = \sigma/\sigma_{SM}$ at $m(H) = 140$ GeV/c^2 with the profile likelihood asymptotic approximation.

Model	**CMS computation**	**ATLAS computation**	**difference (%)**
ATLAS WW	0.7073	0.7073	-
ATLAS $\gamma\gamma$	5.7725	5.7721	-
CMS WW	0.4248	0.4248	-
CMS $\gamma\gamma$	4.2997	4.3000	-
CMS ZZ	1.1679	1.1679	-
ATLAS combined	0.7100	0.7100	-
CMS combined	0.3444	0.3444	-
All combined	**0.2724**	**0.2724**	-

Table 10: $(H \to WW) + (H \to \gamma\gamma) + (H \to ZZ \to 4\ell)$ combination exercise: exclusion limits on $\mu = \sigma/\sigma_{SM}$ at $m(H) = 140$ GeV/c^2 with the LHC-type CL_s method.

Model	**CMS computation**	**ATLAS computation**	**difference (%)**
ATLAS WW	0.76 ± 0.01	0.76 ± 0.02	0%
ATLAS $\gamma\gamma$	5.76 ± 0.02	5.80 ± 0.03	+1%
ATLAS ZZ	4.32 ± 0.05	4.25 ± 0.02	-2%
CMS WW	0.517 ± 0.003	0.526 ± 0.006	+2%
CMS $\gamma\gamma$	3.96 ± 0.01	4.00 ± 0.04	+1%
CMS ZZ	1.691 ± 0.004	1.660 ± 0.040	-2%
ATLAS combined	0.667 ± 0.009	0.674 ± 0.022	+1%
CMS combined	0.426 ± 0.005	0.439 ± 0.005	+3%
All combined	0.410 ± 0.005	0.408 ± 0.014	**-0.5%**

8 Summary

The LHC Higgs Combination Group was formed in December 2010 to prepare ATLAS and CMS Collaborations for the forthcoming Higgs search combinations with the 2011 data. Over the time period of six months, the group achieved the following goals as documented in this report:

- established the common methods for reporting exclusion limits and quantifying excesses,
- agreed on the initial set of common systematic errors between ATLAS and CMS, on their modelling and correlations,
- formulated the format of presenting Higgs search results,
- exercised statistical methods and software tools with toy models of Higgs searches in order to validate and synchronise the overall combination procedure.

The group is ready to combine Higgs search results from ATLAS and CMS.

Outlook

At the time of writing, no major issues remain unresolved. Many hurdles have been overcome to pave the way toward combined ATLAS and CMS Higgs results in 2011. It is our belief that, should any new issues arise, they will be addressed in the same spirit in which the current work has been conducted: discussions and agreement. The report presented here is by no means the final word on combining ATLAS and CMS Higgs search results. We fully expect that the techniques presented here will evolve and be refined further.

Acknowledgements

We would like to thank the ATLAS statistics forum and CMS statistics committee for their extremely valuable and continuous feedback and for the guiding suggestions and corrections. We would like to acknowledge the role of the LHC Higgs Cross Section group that helped settle a number of non-trivial questions on correlations of theoretical errors for exclusive final states of Higgs boson production in association with jets. The prompt response of the group on the request to produce SM Higgs boson production cross sections and branching ratios for the fine grid of Higgs boson mass points needed for the combination was simply spectacular. We would also like to thank the ATLAS and CMS Higgs working groups for their close involvement in the overall effort and for preparing analysis Workspaces for performing technical exercises as reported in this document.

A Brief overview of statistical methods

This Appendix briefly accounts for the different statistical approaches aiming to characterise a non-observation of a signal or establish a significant excess of events. We refrain from judgemental statements on the pros and cons of different methods and simply account for what has been used in the past. For a more comprehensive overview one can refer, for example, to Refs. [21, 22]. The methods chosen for the combination in Summer 2011 are discussed in more detail in Sections 2 and 3.

In the following subsections, the expected Standard Model Higgs event yields will be generically denoted as s, backgrounds as b. These will stand for event counts in one or multiple bins or for unbinned probability density functions, whichever approach is used in an analysis. Predictions for both signal and background yields, prior to the scrutiny of the data entering the statistical analysis, are subject to multiple uncertainties that are handled by introducing nuisance parameters θ, so that signal and background expectations become functions of the nuisances: $s(\theta)$ and $b(\theta)$. The actual observed events will be denoted as *data* or *observation*.

A.1 Limits

The Bayesian and the classical frequentist, with a number of modifications, are two statistical approaches commonly used in high energy physics for characterising the absence of a signal.

Both methods allow one to quantify the level of incompatibility of data with a signal hypothesis, which is expressed as a confidence level (C.L.). It is common to require a 95% C.L. for "excluding" a signal, this is however a convention. The probabilistic interpretation of C.L. as the chance of being right or wrong when stating the non-existence of a signal is not straightforward and the subject of a vast body of literature.

In addition, in an analysis targeting a specific signal production mechanism and a particular decay mode, one can also set *approximately* model-independent limits on signal cross section times branching ratio ($\sigma \times \mathrm{BR}$) or *somewhat better defined* limits on cross section times branching ratio times experimental acceptance ($\sigma \times \mathrm{BR} \times \mathcal{A}$). The latter are less useful for testing various theories unless a model of the experimental acceptance $\mathcal{A}$ is also provided.

In a combination of multiple analyses sensitive to different signal production mechanisms and different decay modes, presenting results in a form of limits on $\sigma \times \mathrm{BR}$ or $\sigma \times \mathrm{BR} \times \mathcal{A}$ is impossible. The customary alternative for SM Higgs searches is to set limits on a common *signal strength modifier* μ that is taken to change the cross sections of all production mechanisms by exactly the same scale. Decay branching ratios are assumed to be those given by the Standard Model. The Standard Model Higgs is said to be excluded at, say, 95%C.L., when the 95% C.L. limit on μ drops to one, i.e. $\mu_{95\%CL} = 1$. In the next sub-sections, we will follow this convention and discuss limits on the common signal strength modifier μ.

A.1.1 Bayesian approach

In the Bayesian approach, the Bayes theorem is invoked to assign a degree of belief to the Higgs hypothesis by calculating the posterior "probability density function" $L(\mu)$ on the signal strength μ:

$$L(\mu) = \frac{1}{C} \int_{\theta} p(\text{data}|\mu s + b)\ \rho_{\theta}(\theta)\ \pi_{\mu}(\mu)\ d\theta. \tag{23}$$

The functions $\rho_{\theta}(\theta)$ are *pdf*s describing our prior belief in the scale and description of the uncertainties on signal and background event yields. The choice of these *pdf*s is discussed in Section 5. The function $\pi_{\mu}(\mu)$ is the prior on the signal strength, which is commonly taken to be flat for $\mu \geq 0$ and zero otherwise. Other priors are possible, but have hardly ever been used in high energy physics. The constant C is set to make the overall posterior function $L(\mu)$ normalised to unity. Integration over nuisance parameters in the above equation is known as marginalisation.

The Bayesian one-sided 95% C.L. limits on μ are extracted from the following equation:

$$\int_{0}^{\mu_{95\%CL}} L(\mu)\ d\mu\ =\ 0.95. \tag{24}$$

By definition, the Bayesian methodology obeys the likelihood principle since the inference is based on the data alone. The Bayesian approach is among the three methods described in the PDG.

A.1.2 Frequentist approach and its modifications

Classical frequentist

The classical frequentist approach is formulated for the case of no systematic uncertainties and begins from defining a test statistic q_{μ} designed to discriminate signal-like from background-like events. The test statistic compresses all signal-vs-background discriminating information into one number. By the Neyman-Pearson lemma, the ratio of likelihoods Q is the most powerful discriminator. For a number of reasons, the actual quantity used is a logarithm of the ratio, or more accurately, $-2lnQ$:

$$q_{\mu}\ =\ -2\ \ln \frac{\mathcal{L}(data|\mu s + b)}{\mathcal{L}(data|b)}, \tag{25}$$

where $\mathcal{L}(data|rate)$ is simply a product of Poisson probabilities for number of either *observed* or *simulated* events in each sub-channel, given the expected signal and background rates. One can see that events with $q_{\mu} > 0$ are more likely to appear under the *background-only* hypothesis than the *background+signal* assumption.

It is to be noted that this test statistic was used by LEP and the Tevatron, but not at the LHC, where the profile-likelihood test statistic $\tilde{q}_{\mu}$ is used (see table 11) due to its known asymptotic properties (see A.1.3).

Having defined the test statistic, next one constructs *pdf*s of the chosen test statistic q_{μ} under the *signal+background* hypothesis by means of "tossing" toy pseudo-observations according to the very same Poisson probabilities. Using these *pdf*s, one can then evaluate

the probability $P(q_\mu \geq q_\mu^{data} | \mu s + b)$ for the *observed* value q_μ^{data} to be as or less compatible with the *background+signal* hypothesis. Such a probability is denoted as CL_{s+b}. In the classical frequentist approach, one says that the signal is excluded at, say, 95% C.L., if $\mathrm{CL}_{s+b} = 0.05$.

However, such a definition has a pitfall: by taking the signal strength equal to zero, one expects, by construction, that $\mathrm{CL}_{s+b} \leq 0.05$ with a 5% chance—hence, 5% of all searches will end up excluding a signal of zero strength. In this case, one must appreciate the actual statistical meaning of what has been observed in such cases: that is, a downward fluctuation of the background. To prevent, at least partially, our inference of a signal from such downward fluctuations, a number of solutions have been suggested.

Modifications of the classical frequentist method

- Feldman and Cousins [23] introduced a method of constructing unified (i.e. one/two-sided) confidence intervals based on the likelihood-ratio test statistic:

$$q_\mu \;=\; -2 \ln \frac{\mathcal{L}(data|\mu s + b)}{\mathcal{L}(data|\hat{\mu} s + b)}, \quad \text{with a constraint: } 0 \leq \hat{\mu} \tag{26}$$

  where $\hat{\mu}$ maximises the likelihood $\mathcal{L}(data|\mu s + b)$. Such construction automatically protects the limits on signal strength from the undesired effects of downward fluctuations of background, preserves the proper frequentist coverage, and does not suffer from under-coverage due to having to make flip-flop decisions between reporting one-sided upper limits (no excess) and two-sided intervals when a significant excess of events is observed. One can force the FC method to report one-sided limits no matter what—the price is over-coverage for the cases when one observes an excess of events. The Feldman-Cousins approach is among the three methods described in the PDG.

- At the time of LEP, the so-called modified frequentist approach was introduced with the same goal to "protect" our judgement on a very weak signal strength when downward fluctuations occur [5–7]. In this method, in addition to $\mathrm{CL}_{s+b} = P(q_\mu \geq q_\mu^{data} \,|\, \mu s + b)$, one also calculates $\mathrm{CL}_b = P(q_\mu \geq q_\mu^{data} \,|\, b)$, by "tossing" pseudo-data for *background-only* event rate, and, then calculates the quantity CL_s as the ratio of these two probabilities:

$$\mathrm{CL}_s \;=\; \frac{\mathrm{CL}_{s+b}}{\mathrm{CL}_b}. \tag{27}$$

  In the modified frequentist approach, it is this value, CL_s, that is required to be less than or equal to 0.05 in order to declare the 95% C.L. exclusion. By construction, the CL_s-based limits are one-sided. The price of the protection from background downward fluctuations is a gradual increase in the over-coverage as one observes fewer and fewer events. For an observation right on the top of the *background-only* expectation ($\mathrm{CL}_b \sim 0.5$), CL_s is about twice as large as CL_{s+b}. The modified frequentist approach is among the three methods described in the PDG.

- Recently, another approach of Power-Constrained Limits (PCL) was proposed [24]. It prescribes using results from the classical frequentist method ($\mathrm{CL}_{s+b} = 0.05$),

unless the observed limit is below the 50%-quantile of the expected *background-only* results (the experimental sensitivity) . This means that the power of the test with respect to the alternative background only hypothesis is not allowed to go below 50%. In this case when a large downward fluctuation is observed, the reported limit is the one corresponding to the experimental sensitivity. By construction, the limit is one-sided. The price of protection from downward fluctuations by imposing the "power constraint" is an over-coverage when one observes downward fluctuations below the experimental sensitivity.

Introducing systematic uncertainties

Systematic uncertainties on signal and background rates, $s(\theta)$ and $b(\theta)$, are introduced via modifications to the test statistic itself and/or the way pseudo-data are generated. In the following, the prior *pdf* for the nuisance θ will be written as $\rho(\theta|\tilde{\theta})$, where $\tilde{\theta}$ is the "nominal" value of the nuisance parameter.

- One can choose to keep the test statistic given by Eq. 25 or Eq. 26 unchanged and evaluate them using the *nominal* values of the signal and background rates, i.e. at $s(\tilde{\theta})$ and $b(\tilde{\theta})$. The effect of systematic uncertainties is then introduced via modifying $s(\theta)$ and $b(\theta)$ before each pseudo-data set is generated by drawing random numbers from the $\rho(\theta|\tilde{\theta})$ distributions. This method was first introduced to the field by Cousins and Highland [25] and is now known as hybrid Bayesian-frequentist, since the treatment of nuisance parameters in this case is explicitly Bayesian. This is how nuisance parameters were handled at LEP.

- At Tevatron, the hybrid Bayesian-frequentist approach to "tossing" pseudo-data remained the same, but the test statistic was redefined. The Poisson-like likelihoods can be extended to include the nuisance parameter *pdf*s $\rho(\theta|\tilde{\theta})$

$$\mathcal{L}(data|\mu,\theta) = \text{Poisson}\,(\,data\,|\,\mu\cdot s(\theta)+b(\theta)\,)\cdot\rho(\theta|\tilde{\theta}) \tag{28}$$

  Before taking the ratio, both the numerator and denominator likelihoods can be maximised with respect to nuisance parameters. The test statistic then would take the following form:

$$q_\mu \;=\; -2\,\ln\frac{\mathcal{L}(data|\mu,\hat{\theta}_\mu)}{\mathcal{L}(data|0,\hat{\theta}_0)} \tag{29}$$

  where $\hat{\theta}_\mu$ and $\hat{\theta}_0$ are maximum likelihood estimators for the *signal+background* hypothesis (with the signal strength factor μ) and for the *background-only* hypothesis ($\mu = 0$). This is the test statistic used at Tevatron.

- A one-sided test statistics which does not allow the signal to become negative is the profile likelihood test statistic [11]

$$\tilde{q}_\mu \;=\; -2\,\ln\frac{\mathcal{L}(data|\mu,\hat{\theta}_\mu)}{\mathcal{L}(data|\hat{\mu},\hat{\theta})},\quad 0\le\hat{\mu}\le\mu \tag{30}$$

The pair of parameters $\hat{\mu}$ and $\hat{\theta}$ gives the global maximum of the likelihood. The additional constraint $\hat{\mu} \leq \mu$ ensures that the obtained limits are one-sided. The advantage of this test statistic is that its pdf distribution can be approximated by asymptotic formulae based on Wilks and Wald theorems, as derived in Ref. [11] (see Appendix A.1.3).

- Yet another way to treat nuisance parameters is to re-interpret the systematic uncertainty *pdf*s $\rho(\theta|\tilde{\theta})$ as posteriors of some *real* or *imaginary* measurements. Such re-interpretation allows one to build sampling distributions without explicit Bayesian marginalisation. It is this approach to constructing sampling distributions of the test statistic that is chosen for the ATLAS+CMS Higgs search combination in Summer 2011. It is described in detail in Section 2.

From the overview presented in this section, the CL_s procedure chosen for the summer 2011 combination actually differs in details from the ones used at LEP and Tevatron (which were also different). For comparison purposes, all the differences are summarised in Table 11 below. The LEP prescription does not allow one to take full advantage of the constraints imposed on the nuisance parameters by the data used in the statistical analysis. The Tevatron and LHC versions of CL_s, though constructed differently, in practice—as we find—give nearly identical results. The benefit of the LHC-type CL_s is that it uses a test statistic with the desired asymptotic properties. Also, the sampling distributions of the test statistic can be built following the pure frequentist language.

Table 11: Comparison of CL_s definitions as used at LEP, Tevatron, and adopted for the summer 2011 Higgs combination at LHC.

	Test statistic	Profiled?	Test statistic sampling
LEP	$q_\mu = -2 \ln \frac{\mathcal{L}(data\|\mu,\tilde{\theta})}{\mathcal{L}(data\|0,\tilde{\theta})}$	no	Bayesian-frequentist hybrid
Tevatron	$q_\mu = -2 \ln \frac{\mathcal{L}(data\|\mu,\hat{\theta}_\mu)}{\mathcal{L}(data\|0,\hat{\theta}_0)}$	yes	Bayesian-frequentist hybrid
LHC	$\tilde{q}_\mu = -2 \ln \frac{\mathcal{L}(data\|\mu,\hat{\theta}_\mu)}{\mathcal{L}(data\|\hat{\mu},\hat{\theta})}$	yes $(0 \leq \hat{\mu} \leq \mu)$	frequentist

A.1.3 Profile Likelihood Asymptotic Approximation

If we remove the physical requirement $\hat{\mu} > 0$ from the test statistic $\tilde{q}_\mu$ based on the profile likelihood ratio (Equation 30) then we find

$$q_\mu = -2 \ln \frac{\mathcal{L}(data|\mu,\hat{\theta}_\mu)}{\mathcal{L}(data|\hat{\mu},\hat{\theta})}, \quad \hat{\mu} \leq \mu \tag{31}$$

Following Wilks theorem, in the asymptotic regime, q_μ is expected to have half a χ^2 distribution for one degree of freedom (under signal+background experiments). The value of μ that makes

$$\frac{1}{2}q_\mu = 1.35 \tag{32}$$

would correspond to a one-sided $\mathrm{CL}_{s+b} = 0.05$ probability. Another popular choice is

$$\frac{1}{2}q_\mu = 1.92, \tag{33}$$

which is an ad hoc adjustment: it corresponds to $\mathrm{CL}_{s+b} = 0.025$ and, hence, would match $\mathrm{CL}_s = 0.05$, when an observation is right on top of the background-only expectations and, hence, $\mathrm{CL}_b = 0.5$.

However, with the physical requirement $\hat{\mu} > 0$, the asymptotic behaviour of $f(\tilde{q}_\mu|signal+background)$ (where $\tilde{q}_\mu$ is the test statistic used in this combination) does not follow half a χ^2 anymore, yet, it follows a well defined formula [11]

$$f(\tilde{q}_\mu|\mu) = \frac{1}{2}\delta(\tilde{q}_\mu) + \begin{cases} \frac{1}{2}\frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{\tilde{q}_\mu}}e^{-\tilde{q}_\mu/2} & 0 < \tilde{q}_\mu \leq \mu^2/\sigma^2 \ , \\ \frac{1}{\sqrt{2\pi}(2\mu/\sigma)}\exp\left[-\frac{1}{2}\frac{(\tilde{q}_\mu+\mu^2/\sigma^2)^2}{(2\mu/\sigma)^2}\right] & \tilde{q}_\mu > \mu^2/\sigma^2 \ . \end{cases} \tag{34}$$

where

$$\sigma^2 = \frac{\mu^2}{q_{\mu,A}} \tag{35}$$

$q_{\mu,A}$ is the test statistics evaluated with the Asimov data set, i.e. the expected background and the nominal nuisance parameters (setting all fluctuations to be zero).

In the same reference one can also find asymptotic formulae for $f(\tilde{q}_\mu|background)$ from which one can easily derive the median expected limits and their bands, using the Asimov representative data set, without performing any toy Monte Carlo experiment. It is also shown there that in the asymptotic limit, the two test statistics, $\tilde{q}_\mu$ and q_μ (Equations 30 and 31) are equivalent, leading to the same p-values. Which means that in the asymptotic limit, it is sometimes more convenient to use the simpler asymptotic formulae of q_μ. Using these formulae one can easily derive asymptotic relations which easily solve for the upper limit with the CL_s method.

$$CL_s = 0.05 = \frac{1 - \Phi(\sqrt{q_\mu})}{\Phi(\sqrt{q_{\mu,A}} - \sqrt{q_\mu})} \tag{36}$$

Φ^{-1} is the quantile (inverse of the cumulative distribution) of the standard Gaussian. The median and expected error bands are given by

$$\mu_{up+N} = \sigma(\Phi^{-1}(1 - \alpha\Phi(N)) + N) \tag{37}$$

with $\alpha = 0.05$ (μ can be taken as μ_{up}^{med} in the calculation of σ). Note that for $N = 0$ we find the median expected CL_S limit

$$\mu_{up}^{med} = \sigma\Phi^{-1}(1 - 0.5\alpha) = \sigma\Phi^{-1}(0.975) \tag{38}$$

For situations with small numbers of events, the asymptotic result is not guaranteed and is in fact known to give very biased (over-optimistic) results.

A.2 Quantifying an excess of events

In the case of observing an excess of events, characterisation of it begins with evaluating the p-value of the upward fluctuation of the background-only hypothesis. This can be done by "tossing" background-only pseudo-data and building up the corresponding *pdf* for the test statistic of choice.

The four test statistics as given in Equations 25, 29, 26, 30 can be used. The first two would probably use $\mu = 1$, while the profile likelihood ratio is constructed for $\mu = 0$ and $\hat{\mu}$ either unconstrained or constrained to be positive, which makes no difference on the tail of the distribution. For the first two test statistics, observations with a large excess of events would form a left-hand tail, while the profile likelihood test statistic would stretch to the right.

The p-value, i.e. the probability of getting an observation as or less compatible as seen in data for the background-only hypothesis, is then defined as $P(q_1 \leq q_1^{data})$ for the test statistics given by Equations 25, 29 and $P(q_0 \geq q_0^{data})$ for the profile likelihood test statistic given by Equations 26 and 30.

The p-value can be converted into significance Z via either of the two conventions (one-sided or two-sided normal distribution tail probability):

$$p = \int_Z^\infty \frac{1}{\sqrt{2\pi}} \exp(-x^2/2)\, dx \tag{39}$$

$$p = 2\int_Z^\infty \frac{1}{\sqrt{2\pi}} \exp(-x^2/2)\, dx \tag{40}$$

In the asymptotic regime the profile likelihood test statistic (Eq. 9) has the very attractive property of being distributed as a half χ^2 for one degree of freedom, which allows one to approximately estimate the significance, Z, as defined by Equation 39 from the following simple formula:

$$Z = \sqrt{q_0^{data}}. \tag{41}$$

The asymptotic approximation gives very satisfactory results for significance estimations even when one is far from the asymptotic regime.

B Correlations of PDF-associated uncertainties

The following tables show the level of correlations between different backgrounds and Standard Model Higgs production modes. Fig. 10 gives correlations between different backgrounds. Fig. 11 show correlations between different Standard Model Higgs production mechanisms as well as between Higgs production modes and different backgrounds.

In the current mode of combination, cells of the same colour are taken to be 100% correlated, while cells with no fill color are assumed to have no correlations. We follow an intuitive rule of thumb that assuming positively 100% correlated errors is more conservative than weak or negative correlations and that assuming no correlations is more conservative than negatively correlated errors. In general, this is true for signal-signal and signal-background correlations. For background-background correlations, this is also true, except for special cases of deriving (constraining) one background from measuring event rates associated with another one.

There is not a simple solution that would cover all possible situations. The choice of congregating all signal and background processes in three major groups based on the prevailing LO initial states is simply a compromise. As one can see from the tables, the choice we made on grouping different processes is sensible and the differences usually imply that we stay on the conservative side.

Backgrounds

	Z	W	ZZ	WW	WZ	Wγ	WQQ	ZQQ	ggWW	ggZZ	ttbar	tW	tb	tbq
Z	1	0.95	0.67	0.70	0.95	0.9	0.43/0.53	0.08	-0.67	-0.75	-0.74	-0.81	0.59	-0.29
W	0.95	1	0.52/0.69	0.60/0.71	0.88/1.0	0.90/0.80	0.39/0.50	0.08	-0.67	-0.74	-0.73	-0.8	0.57	-0.29
ZZ	0.67	0.52/0.69	1	0.97	0.54/0.73	0.62	0.78/0.87	-0.09	-0.36	-0.34	-0.17	-0.81	0.9	-0.23
WW	0.70	0.60/0.71	0.97	1	0.63/0.75	0.69	0.80/0.86	-0.02	-0.34	-0.33	-0.20	-0.33	0.94	-0.08
WZ	0.95	0.88/1.0	0.54/0.73	0.63/0.75	1	0.9	0.55	0.1	-0.64	-0.71	-0.71	-0.73	0.61	-0.34
Wγ	0.9	0.90/0.80	0.62	0.69	0.9	1	0.63/0.53	0.32	-0.44	-0.54	-0.68	0.61	0.61	0
WQQ	0.43/0.53	0.39/0.50	0.78/0.87	0.80/0.86	0.55	0.63/0.53	1	0.08	-0.12	-0.12	-0.05	-0.15	0.64	-0.32
ZQQ	0.08	0.08	-0.09	-0.02	0.1	0.32	0.08	1	0.54	0.36	-0.26	-0.05	-0.03	0.59
ggWW	-0.67	-0.67	-0.36	-0.34	-0.64	-0.44	-0.12	0.54	1	0.98	0.65	0.81	-0.28	0.63
ggZZ	-0.75	-0.74	-0.34	-0.33	-0.71	-0.54	-0.12	0.36	0.98	1	0.79	0.91	-0.27	0.55
ttbar	-0.74	-0.73	-0.17	-0.20	-0.71	-0.68	-0.05	-0.26	0.65	0.79	1	0.97	-0.12	0.17
tW	-0.81	-0.8	-0.81	-0.33	-0.73	0.61	-0.15	-0.05	0.65	0.91	0.97	1	-0.25	0.31
tb	0.59	0.57	0.9	0.94	0.61	0.61	0.64	-0.03	-0.28	-0.27	-0.12	-0.25	1	0.04
tbq	-0.29	-0.29	-0.23	-0.08	-0.34	0	-0.32	0.59	0.63	0.55	0.17	0.31	0.04	1

Figure 10: Correlations of PDF-associated errors between different backgrounds.

m_H=120

	ggH	VBF	WH	ZH	ttH	Z	W+/W-	ZZ	WW	WZ	Wγ	WQQ	ZQQ	ggWW	ggZZ	ttbar	tW	tb	tbq
ggH	1	-0.57	-0.23	-0.14	-0.6	0.01	0.03	0.02	-0.20	0.04	0.23	-0.14	0.95	0.47	0.28	-0.35	-0.12	-0.24	0.52
VBF	-0.57	1	0.63/0.73	0.76	0.09	0.43	0.26/0.41	0.79	0.72	0.28/0.43	0.28/0.37	0.52/0.71	-0.41	-0.47	-0.4	-0.10	-0.28	0.65	-0.25
WH	-0.23	0.63/0.73	1	0.93	0	0.62	0.52/0.64	0.92	0.93	0.65/058	0.65/0.56	0.79/0.95	-0.02	-0.29	-0.28	-0.15	-0.28	0.99/0.77	0.05/-0.30
ZH	-0.14	0.76	0.93	1	0.03	0.64	0.53/0.66	0.99	0.99	0.55/0.71	0.63	0.83	-0.07	-0.31	-0.3	-0.14	-0.28	0.93	-0.14
ttH	-0.6	0.09	0	0.03	1	-0.61	-0.6	0	-0.05	-0.58	-0.64	0.04	-0.5	0.03	0.56	0.94	0.84	0.02	-0.07

m_H=160

	ggH	VBF	WH	ZH	ttH	Z	W+/W-	ZZ	WW	WZ	Wγ	WQQ	ZQQ	ggWW	ggZZ	ttbar	tW	tb	tbq
ggH	1	-0.61	-0.29	-0.35	-0.24	-0.32	-0.32	-0.35	-0.29	-0.29	-0.06	-0.12	0.9	0.82	0.68	0.1	0.33	-0.27	0.67
VBF	-0.61	1	0.62	0.74	0.2	0.35	0.19/0.34	0.75	0.66	0.20/0.36	0.19/0.28	0.46/0.70	-0.47	-0.46	-0.37	-0.03	-0.22	0.6	-0.29
WH	-0.29	0.62	1	0.93	0.1	0.55	0.52	0.9	0.93	0.56	0.56	0.93	-0.07	-0.26	-0.23	-0.07	-0.21	1	0.03
ZH	-0.35	0.74	0.93	1	0.16	0.54	0.43/0.58	0.98	0.97	0.45/0.63	0.52	0.93	-0.14	-0.29	-0.25	-0.04	-0.2	0.91	-0.16
ttH	-0.24	0.2	0.1	0.16	1	-0.59	-0.58	0.03	-0.03	-0.56	-0.62	-0.05	-0.54	0.33	0.51	0.92	0.8	0.04	-0.12

m_H=200

	ggH	VBF	WH	ZH	ttH	Z	W+/W-	ZZ	WW	WZ	Wγ	WQQ	ZQQ	ggWW	ggZZ	ttbar	tW	tb	tbq
ggH	1	-0.5	-0.26	-0.3	0.13	-0.59	-0.59	-0.36	-0.32	-0.55	-0.33	-0.11	0.68	0.98	0.93	0.5	0.69	-0.27	0.67
VBF	-0.5	1	0.60/0.73	0.72	0.26	0.28	0.13/0.28	0.7	0.62	0.15/0.30	0.12/0.20	0.40/0.69	-0.52	-0.44	-0.34	0.02	-0.17	0.55	-0.32
WH	-0.26	0.60/0.73	1	0.92	0.2	0.44	0.44/0.38	0.89	0.86	0.48/0.41	0.47/0.36	0.78/0.74	-0.15	-0.24	-0.2	0	-0.15	0.98/0.69	0
ZH	-0.3	0.72	0.92	1	0.24	0.46	0.34/0.51	0.95	0.93	0.37/0.56	0.43	0.74/0.85	-0.19	-0.3	-0.22	0.02	-0.14	0.88	-0.2
ttH	0.13	0.26	0.2	0.24	1	-0.57	-0.57	0.03	-0.03	-0.55	-0.63	0.03	-0.56	0.29	0.48	0.9	0.78	0.03	-0.15

m_H=300

	ggH	VBF	WH	ZH	ttH	Z	W+/W-	ZZ	WW	WZ	Wγ	WQQ	ZQQ	ggWW	ggZZ	ttbar	tW	tb	tbq
ggH	1	-0.16	-0.08	-0.09	0.66	-0.8	-0.79	-0.31	-0.31	-0.76	-0.64	-0.11	0.12	0.9	0.97	0.92	0.98	-0.23	0.43
VBF	-0.16	1	0.53/0.72	0.68	0.29	0.16	0.04/0.19	0.6	0.51	0.05/0.20	0.03	0.27/0.65	-0.57	-0.42	-0.31	0.09	-0.11	0.44	-0.39
WH	-0.08	0.53/0.72	1	0.92	0.23	0.32	0.20/0.36	0.82	0.80/0.71	0.34/0.37	0.30/0.20	0.68/0.64	-0.24	-0.22	-0.16	0.1	-0.06	0.89	-0.06
ZH	-0.09	0.68	0.92	1	0.27	0.32	0.20/0.38	0.87	0.82	0.21/0.44	0.26	0.61/0.81	-0.29	-0.25	-0.18	0.11	-0.07	0.79	-0.28
ttH	0.66	0.29	0.23	0.27	1	-0.6	-0.59	-0.05	-0.12	-0.58	-0.65	-0.04	-0.58	0.28	0.47	0.9	0.78	-0.04	-0.17

m_H=500

	ggH	VBF	WH	ZH	ttH	Z	W+/W-	ZZ	WW	WZ	Wγ	WQQ	ZQQ	ggWW	ggZZ	ttbar	tW	tb	tbq
ggH	1	0.09	0.05	0.05	0.91	-0.78	-0.76	-0.25	-0.28	-0.75	-0.73	-0.13	-0.3	0.63	0.78	0.99	0.97	-0.2	0.15
VBF	0.09	1	0.38/0.70	0.6	0.24	0.073	0.0/0.12	0.47	0.37	0/0.12	-0.08	0.11/0.59	-0.58	-0.4	-0.29	0.1	-0.08	0.29	-0.48
WH	0.05	0.38/0.70	1	0.9	0.16	0.19	0.09/0.26	0.69	0.64	0.20/0.20	0.14/0.09	0.55/0.53	-0.3	-0.21	-0.14	0.14	-0.02	0.73	-0.12
ZH	0.05	0.6	0.9	1	0.16	0.22	0.09/0.29	0.77	0.68	0.10/0.34	0.12	0.44/0.74	-0.35	-0.27	-0.19	0.13	-0.05	0.65	-0.37
ttH	0.91	0.24	0.16	0.16	1	-0.63	-0.61	-0.18	-0.23	-0.61	-0.69	-0.14	-0.57	0.3	0.48	0.89	0.79	-0.15	-0.14

Figure 11: Correlations of PDF-associated errors between different SM Higgs production mechanisms as well as between Higgs production modes and different backgrounds.

C Systematic errors in exclusive 0/1/2-jet bins for $gg \to H$ process

The consensus of theorists working in the context of the LHC Higgs Cross Section Group is that it is the *inclusive* cross sections $\sigma_{\geq 0}$, $\sigma_{\geq 1}$, $\sigma_{\geq 2}$ that should be assumed to have independent theoretical errors. Hence, the three independent nuisance parameters are to be associated with uncertainties on these *inclusive* cross sections. These nuisance parameters are labelled as *QCDscale_ggH*, *QCDscale_ggH1in*, *QCDscale_ggH2in*.

However, the actual Higgs search analyses are often split into *exclusive* final states with 0, 1, and 2 jets. Such a choice is dictated by background considerations and—for purposes of the combination of analyses—the necessity to keep all observations mutually exclusive. This section defines the agreed-on procedure for assigning systematic errors on the *exclusive* final states and their cross-channel correlations.

Note that the overall errors on the exclusive final states are larger than the error on the total cross section. Also, it is important to note that some κ's are greater than one, while the others are smaller. This is a manifestation of negative correlations of errors between exclusive final states.

Prescription summary

Take the total $gg \to H$ cross section from the Higgs cross section group Yellow Report (YR). Convert the relative QCD scale uncertainties ϵ_+ and ϵ_- (both are positive numbers) from YR to log-normal κ.	σ_{gg}^{YR} $\kappa^{YR} = \sqrt{\exp(\epsilon_+) \cdot \exp(\epsilon_-)}$
Acceptance of events into 0, 1, 2 jet bins is evaluated at the level of the full detector simulation. The associated per-bin effective cross sections to be used in the analysis are:	$\sigma_{gg}^{YR} \cdot \mathcal{A}_0^{det}$ $\sigma_{gg}^{YR} \cdot \mathcal{A}_1^{det}$ $\sigma_{gg}^{YR} \cdot \mathcal{A}_2^{det}$
Using the parton level fixed-order program HNNLO and parton-level cuts closely resembling lepton/jet/MET cuts in the analysis, calculate *exclusive* cross sections for the default QCD scale (TBD) and their variation by changing the scale by a factor of 2 up/down. From these numbers, construct inclusive cross sections and derive their uncertainties. Replace the total CS error with that from YR.	σ_0, σ_1, σ_2 $\sigma_{\geq 0} = \sigma_0 + \sigma_1 + \sigma_2$, $\sigma_{\geq 1} = \sigma_1 + \sigma_2$, $\sigma_{\geq 2} = \sigma_2$ $\kappa_{\geq 0}$, $\kappa_{\geq 1}$, $\kappa_{\geq 2}$ ~~$\kappa_{\geq 0}$~~ $\to \kappa_{\geq 0}^{YR}$
Calculate exclusive theoretical 0, 1, 2 jet bin fractions:	$f_0 = \sigma_0/\sigma_{\geq 0}$ $f_1 = \sigma_1/\sigma_{\geq 0}$ $f_2 = \sigma_2/\sigma_{\geq 0}$

Nuisance parameter name	0-jet bin	1-jet bin	2-jet bin
QCDscale_ggH	$\kappa = \left(\kappa^{YR}\right)^{\frac{1}{f_0}}$	-	-
QCDscale_ggH1in	$\kappa = \left(\kappa_{\geq 1}\right)^{-\frac{f_1+f_2}{f_0}}$	$\kappa = \left(\kappa_{\geq 1}\right)^{-\frac{f_1+f_2}{f_1}}$	-
QCDscale_ggH2in	-	$\kappa = \left(\kappa_{\geq 2}\right)^{-\frac{f_2}{f_1}}$	$\kappa = \kappa_{\geq 2}$

Numerical example

The following tables give a numerical example for $m_H = 160$ GeV/c^2. HNNLO cuts: two leptons with $p_T > 20$ GeV and $|\eta| < 2.5$; MET > 30 GeV (p_T of the two-neutrino system); consider only those jets that have $p_T > 30$ GeV and $|\eta| < 3.0$.

Convert the relative QCD scale uncertainties ϵ_+ and ϵ_- (both are positive numbers) from YR to log-normal κ.	$\epsilon_+ = 0.109$, $\epsilon_- = 0.072$ $\kappa^{YR} = \sqrt{\exp(0.109) \cdot \exp(0.072)} = 1.095$
Using the parton level fixed-order program HNNLO and parton-level cuts closely resembling lepton/jet/MET cuts in the analysis, calculate *exclusive* cross sections for the default QCD scale (TBD) and their variation by changing the scale by a factor of 2 up/down. From these numbers, construct inclusive cross sections and derive their uncertainties. Replace the total CS error with that from YR.	$\sigma_{\geq 0} = [\text{default Q}]^{[\text{Q}/2]}_{[2\text{Q}]} = 41.19^{45.55}_{37.11}$ $\sigma_{\geq 1} = [\text{default Q}]^{[\text{Q}/2]}_{[2\text{Q}]} = 12.59^{15.45}_{10.11}$ $\sigma_{\geq 2} = [\text{default Q}]^{[\text{Q}/2]}_{[2\text{Q}]} = 2.39^{3.95}_{1.51}$ $\kappa_{\geq 0} = \sqrt{\frac{45.55}{41.19} \cdot \frac{41.19}{37.11}} = \sqrt{1.11 \cdot 1.11} = 1.11$ $\kappa_{\geq 1} = \sqrt{\frac{15.45}{12.59} \cdot \frac{12.59}{10.11}} = \sqrt{1.25 \cdot 1.23} = 1.24$ $\kappa_{\geq 2} = \sqrt{\frac{3.95}{2.39} \cdot \frac{2.39}{1.511}} = \sqrt{1.58 \cdot 1.65} = 1.62$ Replace $\kappa_{\geq 0} = 1.11$ with 1.095 from YR
Calculate exclusive theoretical 0, 1, 2 jet bin fractions:	$f_0 = \sigma_0/\sigma_{\geq 0} = 0.69$ $f_1 = \sigma_1/\sigma_{\geq 0} = 0.25$ $f_2 = \sigma_2/\sigma_{\geq 0} = 0.06$

Nuisance name	0-jet bin	1-jet bin	2-jet bin
QCDscale_ggH	$\kappa = \left(\kappa^{YR}\right)^{\frac{1}{f_0}} = 1.14$	-	-
QCDscale_ggH1in	$\kappa = (\kappa_{\geq 1})^{-\frac{f_1+f_2}{f_0}} = 0.91$	$\kappa = (\kappa_{\geq 1})^{-\frac{f_1+f_2}{f_1}} = 1.30$	-
QCDscale_ggH2in	-	$\kappa = (\kappa_{\geq 2})^{-\frac{f_2}{f_1}} = 0.89$	$\kappa = \kappa_{\geq 2} = 1.62$

Derivation

We start out from assuming that errors are not too large and we can relate the log-normal and relative errors as follows: $\kappa_{\geq n} = \exp(\epsilon_{\geq n})$. Then, variations in cross sections $\sigma_{\geq 0}$, $\sigma_{\geq 1}$, $\sigma_{\geq 2}$ are independent and can be written as

$$\tilde{\sigma}_{\geq 0} = \sigma_{\geq 0} \cdot (\kappa_{\geq 0})^x = \sigma_{\geq 0} \cdot \exp(\epsilon_{\geq 0} \cdot x) = \sigma_{\geq 0} \, (1 + \epsilon_{\geq 0} \cdot x),$$

$$\tilde{\sigma}_{\geq 1} = \sigma_{\geq 1} \cdot (\kappa_{\geq 1})^y = \sigma_{\geq 1} \cdot \exp(\epsilon_{\geq 1} \cdot y) = \sigma_{\geq 1} \, (1 + \epsilon_{\geq 1} \cdot y),$$

$$\tilde{\sigma}_{\geq 2} = \sigma_{\geq 2} \cdot (\kappa_{\geq 2})^z = \sigma_{\geq 2} \cdot \exp(\epsilon_{\geq 2} \cdot z) = \sigma_{\geq 2} \, (1 + \epsilon_{\geq 2} \cdot z),$$

where $\epsilon_{\geq n}$ are relative errors and x, y, z are independent nuisance parameters with normal distributions).

$$\begin{aligned}
\tilde{\sigma}_0 &= \tilde{\sigma}_{\geq 0} - \tilde{\sigma}_{\geq 1} \\
&= \sigma_{\geq 0} \, (1 + \epsilon_{\geq 0} \cdot x) - \sigma_{\geq 1} \, (1 + \epsilon_{\geq 1} \cdot y) \\
&= (\sigma_{\geq 0} - \sigma_{\geq 1}) + \sigma_{\geq 0} \epsilon_{\geq 0} \cdot x - \sigma_{\geq 1} \epsilon_{\geq 1} \cdot y \\
&= \sigma_0 + \sigma_0 \tfrac{1}{f_0} \epsilon_{\geq 0} \cdot x - \sigma_0 \tfrac{f_1 + f_2}{f_0} \epsilon_{\geq 1} \cdot y \\
&= \sigma_0 \cdot \left(1 + \tfrac{1}{f_0} \epsilon_{\geq 0} \cdot x - \tfrac{f_1 + f_2}{f_0} \epsilon_{\geq 1} \cdot y \right) \\
&= \sigma_0 \cdot \left(1 + \tfrac{1}{f_0} \epsilon_{\geq 0} \cdot x \right) \cdot \left(1 - \tfrac{f_1 + f_2}{f_0} \epsilon_{\geq 1} \cdot y \right) \\
&= \sigma_0 \cdot e^{\frac{1}{f_0} \epsilon_{\geq 0} \cdot x} \cdot e^{-\frac{f_1 + f_2}{f_0} \epsilon_{\geq 1} \cdot y} \\
&= \sigma_0 \left[\left(e^{\epsilon_{\geq 0}} \right)^{\frac{1}{f_0}} \right]^x \cdot \left[\left(e^{\epsilon_{\geq 1}} \right)^{-\frac{f_1 + f_2}{f_0}} \right]^y \\
&= \sigma_0 \left[\left(\kappa_{\geq 0} \right)^{\frac{1}{f_0}} \right]^x \cdot \left[\left(\kappa_{\geq 1} \right)^{-\frac{f_1 + f_2}{f_0}} \right]^y,
\end{aligned}$$

from where one can see that the exclusive 0-jet bin cross section is subject to uncertainties driven by two independent nuisance parameters x and y and their effect can be written as log-normal with κ's recalculated from the original errors $\kappa_{\geq n}$ on *inclusive* cross sections and *exclusive* fractions f_n.

The effect of nuisance parameters on the *exclusive* cross section σ_1 can be calculated in the exact same manner:

$$\tilde{\sigma}_1 = \tilde{\sigma}_{\geq 1} - \tilde{\sigma}_{\geq 2} = \quad ... \quad = \sigma_1 \left[\left(\kappa_{\geq 1} \right)^{\frac{f_1 + f_2}{f_1}} \right]^y \cdot \left[\left(\kappa_{\geq 2} \right)^{-\frac{f_2}{f_1}} \right]^z.$$

D Technical tools

Implementation of the statistical procedures described above requires a few ingredients: the data themselves, the ability to evaluate the likelihood function at arbitrary parameter points (μ, θ) given an arbitrary dataset, the ability to generate pseudo-data for an arbitrary parameter point, and a prior $\pi(\mu, \theta)$ for Bayesian and hybrid methods. This implies that we must have the probability model $\mathcal{L}(\text{data}_c|\mu, \theta)$ and not just the observed likelihood function. Providing the full probability model for a broad class of models that may describe binned or unbinned data parametrised in $\mathcal{O}(50)$ parameters is challenging and requires dedicated technology. The RooFit and RooStats projects have been developed to meet this challenge. RooFit, which originated in the BaBar experiment, provides the modelling language and the software interfaces and implementation for representing the data and the probability model, as well as the ability to generate pseudo data from the model and find the maximum likelihood estimates $\hat{\mu}$, $\hat{\theta}$, and $\hat{\hat{\theta}}(\mu)$ via MINUIT [26]. RooStats provides higher-level statistical tools for various statistical methods, including the ones outlined above [19].

The probability models for the individual channels (indexed by c) $\mathcal{L}_c(\text{data}_c|\mu, \theta)$ have been implemented in software using the RooFit modelling language, often with the aid of dedicated scripting or factories that construct models of a specific form. A class called `ModelConfig` stores the meta-data necessary for the RooStats statistical tools to use the model in a generic way. The full structure is managed by a class called `RooWorkspace`, which can be saved into a ROOT file using the ROOT persistency and I/O technology.

The individual probability models $\mathcal{L}_c(\text{data}_c|\mu, \theta)$ are formed by individual analysis groups and stored in these workspace files. The combined model is formed using a `RooSimultaneous` object that associates the individual datasets and model terms and identifies the common parameter of interest μ, the nuisance parameters for the experimental systematics common within an experiment, and the nuisance parameters for theoretical uncertainties that are common to ATLAS and CMS

$$\mathcal{L}(\text{data}|\mu, \theta) = \prod_c \mathcal{L}_c(\text{data}_c|\mu, \theta) \ . \tag{42}$$

The correct description of the correlated effect of a common source of uncertainty requires coordination of the parametrisation between the different channels. Some level of customisation is possible post-facto, though we prefer the original workspace to be parametrised appropriately.

References

[1] The ATLAS Experiment at the CERN Large Hadron Collider. *JINST*, 3:S08003, 2008.

[2] S. Chatrchyan et al. The cms experiment at the cern lhc. *JINST*, 3:S08004, 2008.

[3] Further investigations of ATLAS Sensitivity to Higgs Boson Production in different assumed LHC scenarios. (ATL-PHYS-PUB-2011-001), 2011.

[4] The CMS physics reach for searches at 7 TeV. (CMS NOTE 2010/008), 2010.

[5] A. L. Read. Presentation of search results: the CLs technique. *J. Phys. G: Nucl. Part. Phys.*, 28, 2002.

[6] A. L. Read. Modified frequentist analysis of search results (the CLs method). *in Proceedings of the First Workshop on Confidence Limits, CERN, Geneva, Switzerland*, 2000.

[7] Thomas Junk. Confidence level computation for combining searches with small statistics. *Nucl.Instrum.Meth.*, A434:435–443, 1999.

[8] W. Fisher. Collie: A confidence level limit evaluator. D0 note 5595, June 2009.

[9] W. Fisher. Systematics and limit calculations. Report No. FERMILAB-TM-2386-E, 2006.

[10] Tom Junk. Sensitivity, Exclusion and Discovery with Small Signals, Large Backgrounds, and Large Systematic Uncertainties. CDF/DOC/STATISTICS/PUBLIC/8128, October 2007.

[11] Glen Cowan, Kyle Cranmer, Eilam Gross, and Ofer Vitells. Asymptotic formulae for likelihood-based tests of new physics. *Eur.Phys.J.*, C71:1554, 2011.

[12] K. Cranmer. Statistical challenges for searches for new physics at the lhc. *Proceedings of Phystat05, Oxford University Press, Editors Louis Lyons,Muge Karagoz Unel*, pages pp. 112–124, 2005.

[13] S.S. Wilks. The large-sample distribution of the likelihood ratio for testing composite hypotheses. *Ann. Math. Statist.*, 9:pp. 60–62, 1938.

[14] Eilam Gross and Ofer Vitells. Trial factors for the look elsewhere effect in high energy physics. *The European Physical Journal C - Particles and Fields*, 70:525–530, 2010. 10.1140/epjc/s10052-010-1470-8.

[15] R.B. Davies. Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika*, 74:pp. 33–43, 1987.

[16] CTEQ Collaboration. http://cteq.org.

[17] LHC Higgs Cross Section Working Group, S. Dittmaier, C. Mariotti, G. Passarino, and R. Tanaka (Eds.). Handbook of LHC Higgs Cross Sections: 1. Inclusive Observables. *CERN-2011-002*, CERN, Geneva, 2011.

[18] The TEVNPH Working Group (for CDF and D0 collaborations). Combined cdf and d0 upper limits on standard model higgs boson production with up to 8.2 fb^{-1} of data. *FERMILAB-CONF-11-044-E, CDF Note 10441, D0 Note 6184*, March 15, 2011.

[19] Lorenzo Moneta, Kevin Belasco, Kyle Cranmer, Alfio Lazzaro, Danilo Piparo, et al. The RooStats Project. *PoS*, ACAT2010:057, 2010. [arXiv:1009.1003].

[20] Chen, M. and Korytov, A. Limits and signifcance. https://mschen.web.cern.ch/mschen/LandS/.

[21] G. Cowan. Statistical Data Analysis. *Clarendon Press, Oxford*, 1998.

[22] K. Nakamura et al. Particle Data Group. *J. Phys. G*, 37:075021, 2010.

[23] Robert D. Cousins Gary J. Feldman. Unified approach to the classical statistical analysis of small signals. *Phys. Rev. D*, 57(7):3873–3889, 1998.

[24] G. Cowan, K. Cranmer, E. Gross, and O. Vitells. Power-Constrained Limits. 2011. [arXiv:1105.3166].

[25] Robert D. Cousins and Virgil L. Highland. Incorporating systematic uncertainties into an upper limit. *Nucl.Instrum.Meth.*, A320:331–335, 1992. Revised version.

[26] Wouter Verkerke. Statistical Software for the LHC. *PHYSTAT-LHC Workshop on Statistical Issues for LHC Physics*, 2008. oai:cds.cern.ch:1021125. `http://cdsweb.cern.ch/record/1099988`.