



Anatomy of the Higgs fits: A first guide to statistical treatments of the theoretical uncertainties

Sylvain Fichet^{a,b}, Grégory Moreau^{c,*}

^a *ICTP South American Institute for Fundamental Research, Instituto de Física Teórica,
São Paulo State University, Brazil*

^b *International Institute of Physics, UFRN, Av. Odilon Gomes de Lima, 1722, Capim Macio, 59078-400 Natal-RN,
Brazil*

^c *Laboratoire de Physique Théorique, Bât. 210, CNRS, Université Paris-sud 11, F-91405 Orsay Cedex, France*

Received 15 September 2015; received in revised form 19 February 2016; accepted 19 February 2016

Available online 2 March 2016

Editor: Tommy Ohlsson

Abstract

The studies of the Higgs boson couplings based on the recent and upcoming LHC data open up a new window on physics beyond the Standard Model. In this paper, we propose a statistical guide to the consistent treatment of the theoretical uncertainties entering the Higgs rate fits. Both the Bayesian and frequentist approaches are systematically analysed in a unified formalism. We present analytical expressions for the marginal likelihoods, useful to implement simultaneously the experimental and theoretical uncertainties. We review the various origins of the theoretical errors (QCD, EFT, PDF, production mode contamination...). All these individual uncertainties are thoroughly combined with the help of moment-based considerations. The theoretical correlations among Higgs detection channels appear to affect the location and size of the best-fit regions in the space of Higgs couplings. We discuss the recurrent question of the shape of the prior distributions for the individual theoretical errors and find that a nearly Gaussian prior arises from the error combinations. We also develop the bias approach, which is an alternative to marginalisation providing more conservative results. The statistical framework to apply the bias principle is introduced and two realisations of the bias are proposed. Finally, depending on the statistical treatment, the Standard Model prediction for the Higgs signal strengths is found to lie within either the 68% or 95% confidence level region obtained from the latest analyses of the 7 and 8 TeV LHC datasets.

* Corresponding author.

E-mail addresses: sylvain@ift.unesp.br (S. Fichet), moreau@th.u-psud.fr (G. Moreau).

1. Introduction and summary

Besides the historical discovery of a resonance around 125 GeV [1,2] that is most probably the Brout–Englert–Higgs boson responsible for the ElectroWeak (EW) symmetry breaking [3], the ATLAS and CMS Collaborations have provided a set of 88 rate measurements – based on the full dataset collected so far with luminosities of $\sim 5 \text{ fb}^{-1}$ at the center of mass energy $\sqrt{s} = 7 \text{ TeV}$ and $\sim 20 \text{ fb}^{-1}$ at $\sqrt{s} = 8 \text{ TeV}$ [4,5] (see also Refs. [6,7]) – that constitutes a new and precious source of indirect information on physics beyond the Standard Model (SM). Indeed, observing deviations of the Higgs boson rates with respect to their SM predictions would reveal the presence of an underlying theory while the absence of such deviations allows one to strongly constrain new models (see for example Ref. [8] for higher-dimensional models, Ref. [9] for composite Higgs theories and Ref. [10] for supersymmetric scenarios). So far, no signs from an unknown world have come out from the data, but this is only the beginning of a long exploration, given the expected LHC upgrades [11].

The fits of the Higgs rates (*cf.* Ref. [12] for the first set of analyses, Refs. [13–16] for the results after the Moriond 2013 winter conference and Refs. [4,5] for the latest official ATLAS and CMS analyses) are thus obviously important. Now certain aspects of these analyses remain to be worked out in order to obtain the final fits for testing new physics. First, the precise likelihood functions associated to the experimental rates (in particular their specific shapes and the complete correlations between channels) are not provided in the present public papers, although they might be expected at some point. Second, a major part of the theoretical uncertainties is due to QCD calculations of the Higgs production rates [17–20] and their treatments in the fits raise questions in the Higgs physics community (see Refs. [21,22] for recent discussions). Taking carefully into account these theoretical uncertainties is crucial for the Higgs fits due to the following reasons.

First, theoretical uncertainties can be sizeable with respect to the experimental ones. The QCD uncertainty on the gluon–gluon fusion mechanism dominantly involved in most of the Higgs discovery channels induces typically an error of $\sim 10\%$ on signal strengths (see Section 6), that is already comparable to the experimental error bars in several Higgs channels which reach values down to $\sim 20\%$ [4–7]. Besides, considering for instance the CMS perspectives at $\sqrt{s} = 14 \text{ TeV}$ with a luminosity of 300 fb^{-1} , the experimental error bars are around $\sim 5\%$ (with same systematic errors as today) for the diphoton final state and less than $\sim 10\%$ for the τ -lepton, Z and W boson channels [11] so that the theoretical error might even become the dominant one in some channels.

Second, theoretical uncertainties might be of the same magnitude as the main potential deviations due to new physics. For instance the maximal corrections to Higgs couplings estimated in Ref. [23] for characteristic composite Higgs and supersymmetric models¹ lead typically to deviations of the signal strengths between $\sim 2\%$ and tens of percent compared to SM. This is of

¹ In the case of no new states, related to the EW symmetry breaking, directly observed at the LHC.

the same order as the theoretical error mentioned above, so that one is precisely in the situation where the theoretical error deserves a careful treatment to test new physics scenarios.²

Therefore, in this paper, our primarily goal is to answer precisely the question: *what is the correct treatment of the theoretical uncertainties in the fits of the Higgs boson rates?*³ This seemingly simple question has lead us to several new developments, summarised in the three lines of work described in the paragraphs below.

First, we present a systematic survey of the various statistical treatments of the theoretical error and their applications to the Higgs fits within a unified formalism. We confront the frequentist and Bayesian frameworks,^{4,5} that prove to exhibit a certain degree of convergence at the level of accuracy of the present LHC data.⁶ We also compare the marginalisation and bias treatments. In the former, we consider the representative cases of Gaussian and flat combined priors because of the lack of knowledge inherent to the distribution of theoretical uncertainties.⁷ We find the Gaussian prior to be well motivated by the full combination of each individual theoretical uncertainty. It turns out that the choice of one among all these statistical approaches may affect significantly the determination of the Higgs properties. It is thus important to understand precisely the conceptual differences between these approaches. Finally, this survey is the opportunity to provide useful analytical expressions for the marginalised likelihood functions, including the theoretical correlations among the Higgs channels.

Second, we explain precisely the principle of bias⁸ and its fundamental differences with the marginalisation principle. The bias principle is more conservative than the marginalisation principle by construction and does not depend on the shape of the priors of the nuisance parameters. This thorough examination of the bias principle leads naturally to introduce a statistical framework for biasing. We propose two realisations of the bias, referred to as the extremal bias and the envelope method, that apply in both frequentist and Bayesian contexts. Regarding the error combinations, important differences arise between the marginalisation and bias frameworks.⁹

Third, we discuss and implement several improvements in the treatment of the theoretical uncertainties. (i) For the cross sections, the combinations of all the individual uncertainties are discussed exhaustively, including in particular the several errors constituting the parton PDF uncertainty. The so-called leading moment approximation is developed to facilitate the combination

² This intermediate situation is to be contrasted with the two extreme cases of expected signal strength deviations much higher than the theoretical error (which can then be neglected) or deviations well smaller (no hope to detect them). In both of these cases, a detailed treatment of the theoretical error would not be really needed to test new physics scenarios.

³ Throughout this paper, we use generically the expression “theoretical error” to denote any error on the SM prediction for the Higgs rates. This is a slight wording abuse, because certain of these errors like the ones from the PDF determination have a partial experimental origin.

⁴ Sometimes in the literature, there are inconsistencies in the sense that errors are combined in a frequentist way (combination depending on the prior shape) while the priors are convoluted in a Bayesian way (convolution via integrations).

⁵ A pure Bayesian fit of the Higgs rates has been carried out in Ref. [16].

⁶ To be contrasted with the preliminary study of Ref. [24] based on simulated Higgs data.

⁷ To the best of our knowledge, a flat prior for the theoretical uncertainty is for the first time applied to the Higgs fits. Notice also that the combination in quadrature of the theoretical and experimental errors, sometimes made in the literature, is equivalent to a marginalisation assuming Gaussian distributions for both sources of errors and neglecting the correlations. This is true in both frequentist and Bayesian cases.

⁸ A bias has been applied once in Ref. [14]. The analysis developed here improves the bias performed in Ref. [14] by including more effects like the production contamination, the individual scale/EFT/PDF errors, the branching fraction uncertainties, the correlations between Higgs channels and the Bayesian/frequentist cases.

⁹ For example, the PDF and amplitude uncertainties for the ggF mechanism are summed in quadrature in the Bayesian marginalisation, whereas they are linearly summed in the bias approach.

of such a high number of errors. (ii) The error contamination by various production modes and the errors on the Higgs branching ratios are taken into account. (iii) The correlations between the theoretical errors on the various Higgs detection channels are included.¹⁰ We show that these theoretical correlations induce significant shifts of the best-fit regions in the Higgs coupling parameter space. (iv) A Higgs fit with more conservative theoretical errors is shown to illustrate the potential impact from the imperfect knowledge of the magnitude of these errors.

For each of the statistical approaches developed along these three lines of work, we provide the up-to-date Higgs fit results based on the latest available data from the 7 and 8 TeV LHC, that can be readily used for new physics tests. From the theory side, we have updated the major gluon–gluon Fusion mechanism by using its reduced perturbative QCD error, issued from the recent calculation up to N³LO [25]. We have also included the theoretical uncertainty on this production mode due to the use of an Effective Field Theory in the amplitude calculation [25–27], so that the whole error on the cross section remains at $\sim 10\%$.

2. Statistical preliminaries

This section condenses the basic elements of frequentist and Bayesian statistics that will be used along the paper. In addition to statistical basics, the principle of bias is also presented.

2.1. Need-to-know frequentist and Bayesian statistics

In order to extract some information about a new physics model from a set of data, the central quantity to study is the likelihood function [28]. The likelihood function is equal to the conditional probability density for obtaining the observed data, taken as a function of the hypothesis. In the case of predictions made in a given hypothesis H with n parameters $\{\theta_n\} \equiv \theta$, the likelihood function reads

$$L(\theta) \equiv p(d|\theta), \quad (2.1)$$

where d represents the set of data. The overall factor of the likelihood is irrelevant for statistical analyses. In the present work, the data we will consider are the set of signal strength measurements from LHC and Tevatron, described in Section 4.1.

In particle physics, the likelihood function encloses a *statistical* uncertainty associated with the data. This is the uncertainty coming from the fluctuations inherent to the observation of a quantum process. This statistical uncertainty tends to zero in the limit of a large amount of data. However, other sources of uncertainty can be present, both on the experimental or the theoretical side. For example, uncertainties arise from the finite resolution of a detector, or from the finite accuracy of a computation. These *systematic* uncertainties do not depend on the amount of data, and need to be taken into carefully. In this paper, we are going to have a close look at the *theoretical* systematic uncertainties.

The starting point for modelling a systematic uncertainty is to explicitly parametrise it. Namely, one introduces a set of new parameters, $\delta \equiv \{\delta_i\}$, which explicitly modifies the likelihood,

$$L(\theta, \delta). \quad (2.2)$$

¹⁰ We notice that such correlations were included e.g. in Ref. [15] for the specific assumption of errors with Gaussian priors and neglecting the correlations among different Higgs production modes.

These new parameters are named *nuisance* parameters, as opposite to the θ 's which are considered as the parameters of interest. This step of parametrisation is common to the frequentist and Bayesian frameworks, and is fairly universal. Discrepancies will appear in the way the δ 's are treated, and will be at the center of our attention in the rest of the paper. Two fundamentally different points of view on how to treat the nuisance parameters, denoted as *marginalisation* and *bias*, will be further identified (in both the frequentist and Bayesian contexts).

In Bayesian statistics, model parameters are genuine random variables. They are associated with a so-called prior distribution, noted $\pi(\theta)$. In order to carry out a process of inference (for example, setting exclusion bounds), the relevant object to study is the posterior distribution,

$$p(H, \theta|d) \propto L(\theta) \pi(\theta). \quad (2.3)$$

In this framework, a so-called $1 - \alpha$ Bayesian credible region is defined by the domain $\Omega_\alpha = \{\theta \mid p(H, \theta|d) > p_\alpha\}$, where p_α is determined by the fraction of integrated posterior

$$\frac{\int_{\Omega_\alpha} d\theta p(H, \theta|d)}{\int_{\Omega} d\theta p(H, \theta|d)} = 1 - \alpha, \quad (2.4)$$

Ω being the whole parameter space. The $1 - \alpha$ Bayesian Credible (BC) contour is the boundary of Ω_α and it corresponds to the contour level defined as $\{\theta \mid p(H, \theta|d) = p_\alpha\}$. In what follows we will use the BC contours at

$$1 - \alpha = \{68.27\%, 95.45\%, 99.73\%\}. \quad (2.5)$$

In frequentist statistics, the likelihood function is employed to build a statistical test, like the likelihood ratio¹¹

$$q(\theta) = -2 \log \left[\frac{L(\theta)\pi(\theta)}{\max_{\theta \in \Omega} L(\theta)\pi(\theta)} \right]. \quad (2.6)$$

The probability density function (*pdf*) of this test is then computed by simulation (typically, using Monte-Carlo pseudo-data). The *pdf* of $q(\theta)$, noted f_q , can then be used to evaluate a **p**-value, typically of the form

$$\mathbf{p}(\theta) = \int_{q_d}^{\infty} f_q(q'|\theta) dq', \quad (2.7)$$

where q_d is the value given by the actual data. The $1 - \alpha$ confidence regions are then obtained by solving $\mathbf{p}(\theta) = \alpha$, *i.e.* the confidence regions are given by $\Omega_\alpha = \{\theta \mid \mathbf{p}(\theta) > \alpha\}$.

In the limit of a large number of observations, the likelihood becomes approximately Gaussian [29].¹² In this limit, confidence contours can be simply obtained by plotting isolines $q(\theta) = q_\alpha$, where the values q_α depend on the number of parameters of interest (see *e.g.*

¹¹ In classical frequentist statistics, hypotheses and parameters are not associated with probabilities. In this paper, for the frequentist side, we adopt the more general framework of hybrid Bayesian-frequentist statistics, in which a distribution can be attributed to a nuisance parameter. Conceptually, such distribution cannot be seen as a prior *pdf*, but corresponds to the likelihood for a real or imaginary measurement constraining the nuisance parameter (see Ref. [58, p. 4]). However, by abuse of language, we will sometimes use the term “prior” in frequentist statistics as well. Classical frequentist statistics are recovered by giving a flat shape to these frequentist “prior” distributions.

¹² Moreover, the parameter estimates will be distributed according to a multivariate Gaussian.

Ref. [30, statistics section]). This procedure is commonly used in the literature, even when the likelihood is not exactly Gaussian. We shall adopt this procedure throughout the paper. In the case where the likelihoods are bivariate (which will be the case of our example of Higgs fit), the threshold values are

$$q_\alpha = \{2.30, 6.18, 11.83\}. \quad (2.8)$$

In the Gaussian limit, these values match exactly the confidence levels $1 - \alpha = \{68.27\%, 95.45\%, 99.73\%\}$.

2.2. Treatment of nuisance parameters

2.2.1. Marginalisation principle

Having introduced the nuisance parameters δ ¹³ in the likelihood $L(\theta, \delta)$, the next step is to eliminate them. This will effectively deform the likelihood, enlarging the preferred regions, and possibly shift their central values. In the Bayesian framework, this is naturally done by integrating over δ , so that

$$L_B(\theta) = \int_{\mathcal{D}} d\delta L(\theta, \delta) \pi(\delta), \quad (2.9)$$

where $\pi(\delta)$ is the prior distribution for the δ parameters.¹⁴ This operation is named *marginalisation*. In the frequentist framework, the likelihood is instead maximised,

$$L_F(\theta) = \max_{\delta \in \mathcal{D}} [L(\theta, \delta) \pi(\delta)]. \quad (2.10)$$

This operation is usually named *profiling*. Here however, in order to emphasise the parallel between Bayesian and frequentist cases, we also refer to it as “marginalisation”. The outcome of Bayesian and frequentist marginalisation gives respectively the marginal likelihoods L_B and L_F . The best-fit regions are then obtained by using L_B and L_F in Eqs. (2.4) and (2.6), respectively. Finally, let us notice that in the frequentist case, it is clear that the marginalisation operation has the effect of selecting the values of δ preferred by the data.

2.2.2. Bias principle

The common feature of Bayesian and frequentist marginalisations is that nuisance parameters contribute to goodness-of-fit. This implies that the nuisance parameters can relax a tension among various measurements, which in turn induces a shift of the best-fit regions. In the context of the search for new physics, such a shift could also be characteristic of the presence of a new physics signal. It is thus of highest importance to correctly understand the effects of nuisance parameters, in order not to confuse systematic uncertainties with the presence of new physics!

In order to explicitly expose the shifts induced by nuisance parameters, and ultimately obtain more conservative results, a useful approach is to define a new operation, alternative to marginalising, with the requirement that the nuisance parameters *do not* contribute to goodness-of-fit. We will refer to this principle as *bias*, as opposite to the marginalisation principle. We will see that

¹³ Recall that we have defined δ as a set of nuisance parameters, $\delta \equiv \{\delta_i\}$. The subsequent integrations and maximisations will thus be multidimensional.

¹⁴ Throughout the paper we assume that the nuisance parameters are independent from the parameters of interest.

the bias principle provides results that are independent of the shape of the prior of the nuisance parameters.

The bias principle can be intuitively grasped as follows. Consider the likelihood $L(\theta, \delta)$ with a single nuisance parameter on the interval $\delta \in [\delta_a, \delta_b]$. Instead of marginalising over δ , one can look at the contours of the likelihood for various *discrete* values of δ , say $\delta = \delta_a, \delta_b$. For each value of δ , the contours are given by Eq. (2.4) (Bayesian) or Eq. (2.6) (frequentist). To obtain the contours, we can see that the likelihood is separately normalised for δ_a and δ_b . This normalisation is in general not the same for δ_a and δ_b . Because of this normalisation factor, no particular value of δ is preferred by the fit. It is this normalisation factor that concretely realises the bias principle.

In Bayesian statistics, the bias principle finds a general realisation as follows. The requirement one wants to implement is that the nuisance parameters δ do not contribute to goodness-of-fit. This is equivalent to ask that the δ do not have a preferred region once data are taken into account. To translate formally this condition, the relevant quantity to involve is the marginal posterior of δ , $p(\delta|d)$. To implement the bias principle, one should thus require $p(\delta|d)$ to be constant, which translates into the condition

$$\frac{\partial}{\partial \delta} p(\delta|d) = 0, \quad (2.11)$$

with

$$p(\delta|d) \propto \int_{\Omega} d\theta L(\theta, \delta) \pi(\delta) \pi(\theta). \quad (2.12)$$

We see that the condition (2.11) fixes the $\pi(\delta)$ prior to be

$$\pi(\delta) \propto \frac{1}{\int_{\Omega} d\theta L(\theta, \delta) \pi(\theta)}. \quad (2.13)$$

This peculiar prior is not independent on data, and is thus not orthodox with respect to the usual Bayesian philosophy. This is an expected consequence of biasing and all quantities are nevertheless well defined. It follows that the posterior for θ and δ has the form $L(\theta, \delta) \pi(\theta) / \int d\theta [L(\theta, \delta) \pi(\theta)]$. The Bayesian bias likelihood is then given by marginalising this particular posterior with respect to the nuisance parameters,

$$\bar{L}_B(\theta) \propto \int_{\mathcal{D}} d\delta \left[\frac{L(\theta, \delta)}{\int_{\Omega} d\theta L(\theta, \delta) \pi(\theta)} \right]. \quad (2.14)$$

In frequentist statistics, the bias principle is realised in a very similar way to the Bayesian case. The quantity telling how δ is constrained by the data is the marginal likelihood for δ (with its associated “prior”), $\max_{\theta \in \Omega} [L(\theta, \delta) \pi(\theta) \pi(\delta)]$, which selects the preferred θ for a given δ . One requires this marginal likelihood to be constant,

$$\frac{\partial}{\partial \delta} \max_{\theta \in \Omega} [L(\theta, \delta) \pi(\theta) \pi(\delta)] = 0. \quad (2.15)$$

This implies that the $\pi(\delta)$ “prior” satisfies

$$\pi(\delta) = \frac{1}{\max_{\theta \in \Omega} L(\theta, \delta) \pi(\theta)}. \quad (2.16)$$

The marginal likelihood of θ is then given by

$$\bar{L}_F(\theta) = \max_{\delta \in \mathcal{D}} \left[\frac{L(\theta, \delta)}{\max_{\theta \in \Omega} [L(\theta, \delta) \pi(\theta)]} \right]. \quad (2.17)$$

This operation is sometimes referred to as the envelope method. This is because, for a continuous domain \mathcal{D} , it draws continuous regions which are wider than the ones obtained by marginalising.¹⁵

Comparing the Bayesian and frequentist realisations of the bias principle, Eq. (2.14) and Eq. (2.17), it appears that the resulting bias operations are fully similar: the expressions Eq. (2.14) and Eq. (2.17) are identical up to interchanging maximisation and integration.

Let us finally comment about the best-fit regions for the bias likelihoods. The Bayesian bias is a particular case of Bayesian marginalisation with a well-chosen prior. The contours are thus obtained by integration, using \bar{L}_B in Eq. (2.4). For the frequentist bias, the bias likelihood \bar{L}_F can be treated using the usual likelihood ratio test and computing the associated p-value, as described in Eq. (2.6). We conclude that the best-fit regions for both the Bayesian and frequentist bias are well-defined.

Let us make an important comment which will turn useful for the frequentist treatments in Section 8. For a single δ in the discrete domain $\mathcal{D} = \{\delta_a, \delta_b\}$, the best-fit regions obtained by inserting the likelihood (2.17) in Eq. (2.6) reproduce exactly the ones in the discrete version of the bias described earlier in this subsection. Indeed, the normalised likelihood (2.17) will lead to a denominator equal to one in Eq. (2.6) and the role of this denominator in the contour definition will be played instead by the denominator of Eq. (2.17).

In this paper, we will refer to the general realisations of the bias principle given by Eqs. (2.14), (2.17) as the *envelope method*, for both the Bayesian and frequentist versions. In contrast, the discrete version of the bias previously introduced can be seen as a minimal realisation of this principle. In this paper, we will refer to it as the *extremal bias*, for both the Bayesian and frequentist versions.

3. Combinations of theoretical uncertainties

This section applies to any systematic uncertainties. Nevertheless, since in this paper our main focus is on theoretical uncertainties, we will readily use this term. In the previous section, we have seen that the correct procedure to incorporate theoretical uncertainties into the likelihood is to model these uncertainties using nuisance parameters and treat them using either the marginalisation or the bias approach. From the practical point of view, this step of marginalisation can be computationally heavy to carry out, both in the Bayesian and frequentist cases. Indeed, for each point in the space of parameters of interest, for n nuisance parameters, either a n -dimensional integration or a n -dimensional maximisation has to be done, whose complexity typically grows exponentially with n .

Because of the cost of exact marginalisation, it is a common practice in the high-energy physics community to combine certain uncertainties in a preliminary step, before carrying out

¹⁵ Using $L = e^{-\chi^2/2}$, one has the equivalent formulation of the envelope method in terms of χ^2 ,

$$\bar{\chi}^2(\theta) = \min_{\delta} \left[\chi^2(\theta, \delta) - 2 \log \pi(\theta) - \min_{\theta} [\chi^2(\theta, \delta) - 2 \log \pi(\theta)] \right]. \quad (2.18)$$

In case of classical frequentist statistics, $\pi(\theta)$ is a constant, so that the two $\log \pi(\theta)$ terms cancel.

the operation of marginalising. This approach of “preliminary combinations” should be followed with some care, because it can be approximative and may contain implicit assumptions. In this section, we revisit and develop the various operations of preliminary combination on a firm statistical ground.

3.1. Error modelisation

Let Q be an arbitrary quantity entering into a likelihood $L[Q]$. The uncertainty about Q can be modelled via a dependence of the form

$$Q \mapsto Q \times (1 + \delta \Delta), \quad (3.1)$$

where δ is the nuisance parameter, associated with a distribution $\pi(\delta)$, defined over the domain \mathcal{D} . Here and throughout this paper, without loss of generality, we let all the δ follow a “standard distribution”, such that all the information about the magnitude of the uncertainty will be contained in the coefficient Δ . With this parametrisation, Δ represents the *relative* uncertainty associated with Q . This linear model (3.1) is valid irrespective of the $\pi(\delta)$ distribution. The actual definition of π depends on the statistical approach adopted. In the Bayesian case, δ is a random variable, so that one chooses $E[\delta] = 0$, $V[\delta] = 1$.¹⁶ Note that the domain of δ can be either finite or infinite. In the hybrid frequentist case, one can follow the same conventions as for the Bayesian case. The classical frequentist case is equivalent to have a flat π , and one sets the domain to be $\mathcal{D} \equiv [-1, 1]$ in that case. For the errors we will consider, π will always be centred on zero.

Let us close this subsection by commenting on possible reparametrisations different from Eq. (3.1) and written as $L(\delta) = L(\phi[\delta'])$. Under this reparametrisation the prior is expressed as $\pi(\phi(\delta')) = \pi'(\delta')[\partial_{\delta'}\phi(\delta')]^{-1}$ in the Bayesian case. This reparametrisation appears as a change of variable in Eq. (2.9) and Eq. (2.10) (with a non-vanishing Jacobian) and in turn leaves invariant the marginalised Bayesian and frequentist likelihoods, respectively. Similarly, the frequentist bias likelihood (2.17) is unchanged by the reparametrisation. The invariance of the Bayesian bias likelihood (2.14) is in general not true.¹⁷ However, in the limit of small relative uncertainties, as will be the case throughout this paper, the Jacobian tends to be constant so that an approximate invariance under reparametrisation generally holds. Besides, our choice of definition (3.1) makes the relative uncertainty Δ explicitly appear, which can be used to perform error propagations and combinations under Taylor series expansions for $\Delta \ll 1$. For example, such an expansion leads to the linear sum in Eq. (3.2), neglecting higher-order quadratic terms.

3.2. Bayesian combination of theoretical uncertainties

In the Bayesian framework, a nuisance parameter δ is rigorously taken as a random variable with prior distribution π . In presence of various nuisance parameters, one may wish to combine

¹⁶ E and V respectively denote the expected value and variance operators, $E[\delta] = \int_{\mathcal{D}} d\delta \delta \pi(\delta)$ and $V[\delta] = \int_{\mathcal{D}} d\delta \delta^2 \pi(\delta) - (E[\delta])^2$.

¹⁷ Starting from the condition of Eq. (2.11), the bias formula is given by

$$\bar{L}_B(\theta) \propto \int_{\mathcal{D}'} d\delta' \left[\frac{L(\theta, \phi(\delta'))}{\int_{\Omega} d\theta L(\theta, \phi(\delta')) \pi(\theta)} \right]$$

which does not include a Jacobian factor necessary for recovering Eq. (2.14).

various sources of error, say δ_A and δ_B . A combination of these sources can be done if they appear systematically into a single combination inside the likelihood, $L[\delta_A \Delta_A + \delta_B \Delta_B]$. One can then define the combined error $\delta_C \Delta_C = \delta_A \Delta_A + \delta_B \Delta_B$, so that

$$L[\delta_C \Delta_C] \pi_C(\delta_C) \propto \int d\delta_A d\delta_B \delta[\delta_A \Delta_A + \delta_B \Delta_B - \delta_C \Delta_C] L[\delta_A \Delta_A + \delta_B \Delta_B] \pi_{A,B}(\delta_A, \delta_B), \quad (3.2)$$

where $\delta[x]$ is the Dirac distribution. Here $\pi_{A,B}$ is the common prior of δ_A, δ_B . If these are independent, one has $\pi_{A,B}(\delta_A, \delta_B) = \pi_A(\delta_A) \pi_B(\delta_B)$. Note that the integration over δ_C of the left-hand side of this equation recovers Eq. (2.9).

When δ_A and δ_B are independent, Eq. (2.9) implies that the distribution of δ_C is exactly given by a convolution product,

$$\pi_C\left(\frac{x_C}{\Delta_C}\right) = \int dx \pi_A\left(\frac{x}{\Delta_A}\right) \pi_B\left(\frac{x_C - x}{\Delta_B}\right). \quad (3.3)$$

The variable x can be seen as $\delta \Delta$. It is convenient to define $\bar{\pi}_C(x) = \pi_C\left(\frac{x}{\Delta_C}\right)$, so that the width of $\bar{\pi}_C$ is given by Δ_C . In contrast, recall that the width of π_C is always normalised to one by convention. Using the $\bar{\pi}$ definition, the convolution (3.3) can simply be written as

$$\bar{\pi}_C(x_C) = \int dx \bar{\pi}_A(x) \bar{\pi}_B(x_C - x), \quad (3.4)$$

or more shortly

$$\bar{\pi}_C = \bar{\pi}_A \star \bar{\pi}_B. \quad (3.5)$$

The resulting distribution π_C has in general a non trivial shape, except for example when both π_A and π_B are Gaussian, in which case π_C is Gaussian as well. In contrast, Eq. (3.3) implies that the magnitudes of the errors Δ_A, Δ_B are combined following

$$\Delta_C^2 = \Delta_A^2 + \Delta_B^2, \quad (3.6)$$

irrespective of the shape of the distributions. That is, the errors are always combined in quadrature, *i.e.* the variances always add-up. Note the Δ^2 's correspond to the variance of the $\bar{\pi}$ distributions.

In case of two independent sets of several correlated variables $\delta_{A,i}, \delta_{B,i}$ with respective covariance matrices $\mathcal{C}_A, \mathcal{C}_B$, combined as $\delta_{C,i} = \delta_{A,i} + \delta_{B,i}$,¹⁸ the combination is naturally generalised to

$$\mathcal{C}_C = \mathcal{C}_A + \mathcal{C}_B. \quad (3.7)$$

Again, this is independent of the prior shapes. The distribution of $\delta_{C,i}$ is again obtained using Eq. (3.2).

Finally, one may wish to combine nuisance parameters that are themselves correlated. In the case of two nuisance parameters δ_A, δ_B with a correlation coefficient ρ , one gets

$$\Delta_C^2 = \Delta_A^2 + \Delta_B^2 + 2\rho \Delta_A \Delta_B, \quad (3.8)$$

¹⁸ Note that in this case, for simplicity, we used a different convention from the one-variable case: we do not factor out the magnitude of the uncertainties (Δ_i) in front of the δ_i .

giving rise to a linear combination in the fully (anti-)correlated case $\rho = \pm 1$, and to Eq. (3.6) in the de-correlated case $\rho = 0$. The combination (3.8) is still independent of the prior shapes. Note that in this case π_C is still obtained from Eq. (3.2), but is not given anymore by a convolution product because π_A and π_B are not factorised anymore.

Finally, in the case of two sets of nuisance parameters $\delta_{A,i}$, $\delta_{B,i}$ with a relative correlation matrix \mathcal{C}_{AB} , one gets

$$\mathcal{C}_C = \mathcal{C}_A + \mathcal{C}_B + 2\mathcal{C}_{AB}. \quad (3.9)$$

All the results of this subsection are straightforward to derive using characteristic functions (see Appendix A).

In the limit $\Delta_A \gg \Delta_B$, it appears that $\pi_C \sim \pi_A$, *i.e.* the combined prior has mainly the shape of the leading uncertainty. In Section 3.4, we demonstrate that it is well justified to use Eq. (3.6), which is exact, together with the approximation $\pi_C \approx \pi_A$. Beyond the $\Delta_A \gg \Delta_B$ limit, if one wishes to care about the shape of π_C , a conservative approach is to consider both extreme cases $\pi_C = \pi_A$ and $\pi_C = \pi_B$. This is because the actual shape of π_C is always an intermediate distribution between π_A and π_B , as dictated by the convolution product.

3.3. Frequentist combination of theoretical uncertainties

Let us start again with the nuisance parameters δ_A , δ_B and their associated “prior” distribution $\pi_{A,B}$. If the nuisance parameters enter as a single combination in the likelihood, $L[\delta_A \Delta_A + \delta_B \Delta_B]$, one can define the nuisance parameter δ_C as above, and write

$$L[\delta_C \Delta_C] \pi_C(\delta_C) \propto \max_{\delta_A, \delta_B} \left[\delta[\delta_A \Delta_A + \delta_B \Delta_B - \delta_C \Delta_C] L[\delta_A \Delta_A + \delta_B \Delta_B] \pi_{A,B}(\delta_A, \delta_B) \right], \quad (3.10)$$

where again $\delta[x]$ is the Dirac distribution.¹⁹ We emphasize that this formula is exactly similar to the Bayesian one, Eq. (3.2), with integration replaced by marginalisation. When $\pi_{A,B}(\delta_A, \delta_B) = \pi_A(\delta_A)\pi_B(\delta_B)$, it appears then that the distribution of δ_C is given by

$$\pi_C \left(\frac{x_C}{\Delta_C} \right) \propto \max_x \left[\pi_A \left(\frac{x}{\Delta_A} \right) \pi_B \left(\frac{x_C - x}{\Delta_B} \right) \right]. \quad (3.11)$$

This formula has a convolution product structure, where the integration has been replaced by a maximisation. From that point, it is then possible to compute the frequentist correlation matrix, $\mathcal{C}_{ij}^{-1} = -\partial^2 \log L / \partial \theta_i \partial \theta_j$. The general formula for the combination of \mathcal{C}_A , \mathcal{C}_B is straightforward but tedious to compute. In sharp contrast with the Bayesian case, it appears in the frequentist case that the combination of the correlation matrices \mathcal{C}_A , \mathcal{C}_B accordingly to Eq. (3.11) depends on the shape of the π_A , π_B distributions.

In the particular case where both π_A , π_B are Gaussian, the combination appears to be in quadrature, as in the Bayesian case. The combination formulas then match exactly the Bayesian ones, Eqs. (3.6) and (3.7). Moreover π_C is also Gaussian. Another important particular case is the one of flat priors. In that case, π_C appears to be flat, and the combination is *linear*,

$$\Delta_C = \Delta_A + \Delta_B. \quad (3.12)$$

¹⁹ Here $\delta[x]$ can be taken as the regularised Dirac peak.

Note that no correlation matrix can be defined in the flat case.²⁰

In the case where δ_A and δ_B are correlated, they should be treated with a common “prior” as in the Bayesian case.

3.4. The leading moment approximation

Consider again the Bayesian case of a combination of two nuisance parameters, $\delta_C \Delta_C \equiv \delta_A \Delta_A + \delta_B \Delta_B$. Recall that the δ parameters have zero mean and have a standard distribution so that $E[\delta] = 0$, $V[\delta] = 1$. Assume further that the magnitude of the uncertainty B is small with respect to the uncertainty A ,

$$\Delta_A \gg \Delta_B. \quad (3.13)$$

When this condition is satisfied, the source of uncertainty B can be treated as a perturbation to the source of uncertainty A . Starting from this observation, one can obtain π_C up to Δ_B/Δ_A corrections (see Eq. (A.9)). This is demonstrated in [Appendix A](#) using characteristic functions. In particular, for independent variables, at the first non-trivial order in the expansion, one obtains that

$$\pi_C \approx \pi_A \quad (3.14)$$

$$\Delta_C^2 = \Delta_A^2 + \Delta_B^2. \quad (3.15)$$

Recall that π_C is determined by the convolution product $\bar{\pi}_C = \bar{\pi}_A \star \bar{\pi}_B$. Hence for $\Delta_A \gg \Delta_B$, one can intuitively expect that the shape of $\bar{\pi}_A$ and $\bar{\pi}_C$ are similar (see Eq. (3.14)), even though their widths are different (according to Eq. (3.15)). In case δ_A and δ_B are correlated, Eq. (3.15) has to be replaced by Eq. (3.8).

This “leading moment” approximation is useful in presence of a hierarchy between the magnitude of the various uncertainties. It dictates how to consistently capture the main effects of the uncertainties into the likelihood. This in turn allows one to obtain an approximate form for the combined priors, which opens up the possibility of obtaining analytical expressions for the marginal likelihoods.

The leading moment approximation also applies when δ_A and δ_B appear in various linear combinations within the likelihood. This situation typically happens when various observables are affected by the same source of uncertainty. The case of two nuisance parameters and two combinations is discussed in [Appendix A](#). One considers two combinations $\delta_{C_1} \Delta_{C_1} = \delta_A \Delta_{A_1} + \delta_B \Delta_{B_1}$, $\delta_{C_2} \Delta_{C_2} = \delta_A \Delta_{A_2} + \delta_B \Delta_{B_2}$. It is found that the $\Delta_{C_{1,2}}$ are obtained as in the one-combination case discussed above. The correlation coefficient between δ_{C_1} and δ_{C_2} requires more attention. If $\Delta_{A_1} \gg \Delta_{B_1}$, $\Delta_{A_2} \gg \Delta_{B_2}$, it is found to be approximately equal to one. This implies that the shapes of the distributions of δ_{C_1} , δ_{C_2} and δ_A are the same up to $\Delta_{B_{1,2}}/\Delta_{A_{1,2}}$ corrections (see Eq. (A.15)), that is

$$\pi_{C_1 C_2}(\delta_{C_1}, \delta_{C_2}) \approx \pi_A(\delta_{C_1}) \delta[\delta_{C_1} - \delta_{C_2}]. \quad (3.16)$$

From Eq. (3.16), it appears that the leading moment approximation reduces the number of nuisance parameters in the likelihood. In the case where $\Delta_{A_1} \gg \Delta_{B_1}$, $\Delta_{A_2} \ll \Delta_{B_2}$, it appears that

²⁰ In the multivariate case, $\delta_{A,i}$ and $\delta_{B,i}$ have in general a non-trivial domain \mathcal{D}_A , \mathcal{D}_B . The combined domain \mathcal{D}_C is given by the distance $||\delta_{C,i}||$ for which the centers of \mathcal{D}_A and \mathcal{D}_B are aligned with $\delta_{C,i}$ and the domain \mathcal{D}_A and \mathcal{D}_B share a single point. For example if \mathcal{D}_A , \mathcal{D}_B are “hyper-rectangles” with size $\Delta_{A,i}$, $\Delta_{B,i}$, the sizes simply add up just like in the one-dimensional case, $\Delta_{C,i} = \Delta_{A,i} + \Delta_{B,i}$.

the correlation coefficient between δ_{C_1} and δ_{C_2} is approximately equal to the correlation coefficient between δ_A and δ_B (see Eq. (A.16)), so that

$$\pi_{C_1 C_2} \approx \pi_{AB}. \quad (3.17)$$

In the particular case where δ_A and δ_B are independent, one has

$$\pi_{C_1 C_2} \approx \pi_{C_1} \pi_{C_2}, \quad \pi_{C_1} \approx \pi_A, \quad \pi_{C_2} \approx \pi_B. \quad (3.18)$$

In the other particular case where δ_A and δ_B are 100% correlated or anti-correlated, one has

$$\pi_{C_1 C_2}(\delta_{C_1}, \delta_{C_2}) \approx \pi_A(\delta_{C_1}) \delta[\delta_{C_1} \pm \delta_{C_2}]. \quad (3.19)$$

All the cases with more variables or more combinations can be deduced recursively from the case with two parameters and two combinations studied here.²¹

3.5. Combining uncertainties in the bias approach

We now analyse how the combination of uncertainties arises in the case of the method of bias. We still consider a combination of nuisance parameters $\delta_{A,B}$ entering in the likelihood as $L[\delta_A \Delta_A + \delta_B \Delta_B]$. Recall that in our conventions, δ is a random variable with a fixed domain, while Δ is a number representing the magnitude of the uncertainty. In the bias approach, by definition, the shape of the distribution of δ is set so that δ does not participate to the fit. The information about the uncertainty is thus encoded only in the *domain* of the variable $\delta \Delta$. The choice of this domain has some degree of arbitrariness. This choice depends on how conservative one wants the results to be. In the following we choose to let δ vary in the interval $[-1, 1]$ and we identify Δ as a 1σ error, *i.e.* the same way it is defined for the marginalisation.

The operation of Bayesian bias can be seen as a special case of marginalisation, where the prior is set by Eq. (2.13). As the likelihood we consider in this section depends only on the combination $\delta_A \Delta_A + \delta_B \Delta_B$, this peculiar prior depends only on the combination $\delta_A \Delta_A + \delta_B \Delta_B$ by construction. Let us denote it as $\pi_{\text{bias}}^B(\delta_A \Delta_A + \delta_B \Delta_B)$. In order to get the combination $\delta_C \Delta_C = \delta_A \Delta_A + \delta_B \Delta_B$, one applies the definition of Eq. (3.2) using the π_{bias}^B prior. It turns out that $\pi_C(\delta_C) = \pi_{\text{bias}}^B(\delta_C \Delta_C)$. This means that the domain of $\delta_C \Delta_C$ is given by the domain of $\delta_A \Delta_A + \delta_B \Delta_B$,

$$\mathcal{D}_{\delta_C \Delta_C} = \mathcal{D}_{\delta_A \Delta_A + \delta_B \Delta_B}. \quad (3.20)$$

When δ_A and δ_B are independent, one has simply

$$\Delta_C = \Delta_A + \Delta_B. \quad (3.21)$$

When δ_A and δ_B are 100% correlated positively (*i.e.* $\delta_A = \delta_B$), it turns out that one has again the combination

$$\Delta_C = \Delta_A + \Delta_B. \quad (3.22)$$

When δ_A and δ_B are 100% correlated negatively (*i.e.* $\delta_A = -\delta_B$), the combination reads

$$\Delta_C = |\Delta_A - \Delta_B|. \quad (3.23)$$

²¹ This leading moment approximation will be applied to the theoretical uncertainties on the Higgs rates in Sections 6.4 and 6.5.

Let us stress that the correlation between δ_A and δ_B is determined by their common domain $\mathcal{D}_{\delta_A \Delta_A, \delta_B \Delta_B}$. The above extreme cases are easily determined. The case of an intermediate correlation is trickier as it requires a precise definition of the domain. The case of an arbitrary correlation will not be needed throughout this paper. We see that the uncertainties are automatically combined *linearly* in the Bayesian bias method.

These results above can be applied recursively to more complex combinations. For example if $\delta_D \Delta_D = \delta_A \Delta_A + \delta_B \Delta_B + \delta_C \Delta_C$, with δ_A and δ_B 100% anti-correlated and δ_C independent from the two others, the bias combination gives

$$\Delta_D = |\Delta_A - \Delta_B| + \Delta_C. \quad (3.24)$$

Also, the bias combination applies in presence of various linear combinations (labelled by i) of the same nuisance parameters. In that case, the result of the combination is a common nuisance parameter δ , coming with different magnitudes Δ_i for each combination.

The frequentist bias has the same structure as the Bayesian bias. The starting point to determine the error combination is to use the frequentist version of the bias prior of Eq.(2.16) in Eq. (3.10). It follows that the frequentist combinations are the same as in the Bayesian case. We can thus conclude that in the bias approach, the preliminary combinations of uncertainties are done linearly, in both the frequentist and Bayesian cases. One should remark that such a combination is systematically more conservative than the combinations from both the Bayesian and frequentist marginalisations, as can be seen comparing Eqs. (3.21), (3.22), (3.23) with for example Eq. (3.8). Note that the combination in the frequentist marginalisation with flat prior (see *e.g.* Eq. (3.12)) is the same as the bias combination. Therefore the bias method is also more conservative than the standard marginalisation at the level of error combinations.

4. The Higgs boson rates

The couplings of the Higgs boson h are all predicted in the Standard Model, so that any deviation from the SM predictions would constitute a sign of the existence of physics beyond the SM. The Higgs couplings can be probed by collider experiments, which can produce the Higgs on-shell and observe its decays. This process of Higgs production followed by its decay is parametrised as

$$pp(p\bar{p}) \xrightarrow{X} h \rightarrow Y. \quad (4.1)$$

The SM Higgs production mechanisms accessible at the LHC (and Tevatron) are *i*) gluon–gluon fusion (ggF), *ii*) vector boson fusion (VBF), *iii*) associated production with an electroweak gauge boson $V = W, Z$ (VH), and *iv*) associated production with a $t\bar{t}$ pair (ttH). The main SM Higgs decays observed at the colliders are decays into gauge bosons, $h \rightarrow \gamma\gamma, ZZ, W^+W^-$, and into heavy fermions, $h \rightarrow b\bar{b}, \tau\bar{\tau}$. The production modes X and final states Y will be therefore taken in the following list,

$$X = \{\text{ggF, VBF, VH, ttH}\}, \quad (4.2)$$

$$Y = \{\gamma\gamma, ZZ, WW, b\bar{b}, \tau\bar{\tau}\}. \quad (4.3)$$

4.1. The data

The Higgs searches at ATLAS, CMS and the Tevatron are focused on a specific final state Y . For each final state, various channels are defined using mutually exclusive cuts. Throughout this

paper, these experimental channels will be labelled by lower case Latin indices (i, j, \dots). We will consider all the 88 channels. A given i contains the information on the final state and the specific channel. In the following, it will be sometimes useful to refer to the final state Y corresponding to a given channel i . We will use the short notation Y_i , meaning that Y is taken as a function of the variable i , *i.e.* $Y_i \equiv Y(i)$.

The results from Higgs searches at the LHC and the Tevatron are reported in terms of signal strengths μ_i^{ex} . A signal strength is defined as the ratio of the observed event number with the expected SM event number,

$$\mu_i^{\text{ex}} = \frac{N_i^{\text{ex}}}{N_i^{\text{SM}}} . \quad (4.4)$$

The predicted SM event rate of a process $pp(p\bar{p}) \xrightarrow{X} h \rightarrow Y$ is given, in the narrow width approximation, by $\mathcal{L} \sigma_X^{\text{SM}} B_Y^{\text{SM}}$. Here σ_X^{SM} is the production rate, B_Y^{SM} is the branching ratio $B_Y^{\text{SM}} = \Gamma_Y^{\text{SM}} / \sum_{Y'} \Gamma_{Y'}^{\text{SM}}$ and \mathcal{L} is the integrated luminosity. However, from the experimental viewpoint, all the production processes contribute to a given final state. Hence the Higgs production cross sections have to be weighted by a selection efficiency $\epsilon_{X,i}^{\text{SM}}$ encoding the effects of kinematical cuts. The actual expected event rates are thus given by

$$N_i^{\text{SM}} = \mathcal{L} \sum_X \epsilon_{X,i}^{\text{SM}} \sigma_X^{\text{SM}} B_i^{\text{SM}} , \quad (4.5)$$

where the notation B_i^{SM} is a shortcut for $B_{Y(i)}^{\text{SM}}$, *i.e.* the index i selects the final state Y . The experimental Higgs signal strengths have thus the form

$$\mu_i^{\text{ex}} = \frac{N_i^{\text{ex}}}{\mathcal{L} \sum_X \epsilon_{X,i}^{\text{SM}} \sigma_X^{\text{SM}} B_i^{\text{SM}}} . \quad (4.6)$$

Note that the kinematical cuts have been to some extent designed to disentangle the production modes, so that often one of the efficiencies will dominate over the others.

The experimental central values of the μ_i^{ex} , the associated statistical errors, the experimental systematic errors, and the selection efficiencies $\epsilon_{X,i}^{\text{SM}}$ that we will exploit in our analysis are taken from the following references. The statistical and experimental systematic errors are often combined within these references and will be denoted here as $\Delta\mu_i^{\text{ex}}$.

Regarding the ATLAS data, the diphoton final state results are taken from Ref. [31], the ZZ channel is from Ref. [32], the WW channel from Ref. [33], the $b\bar{b}$ from Ref. [34] and the $\tau\bar{\tau}$ from Ref. [35]. Results are presented as well in Ref. [6] and the combined channels are studied in Ref. [4].

As for the CMS results, the diphoton final state has been presented in Ref. [36], the ZZ channel measurements are provided in Ref. [37], the WW ones in Ref. [38], the $b\bar{b}$ in Ref. [39] and the $\tau\bar{\tau}$ in Ref. [40] (see also Ref. [7] and the combined channel analyses [5]).

Finally, the latest results from the Tevatron (D0 and CDF Collaborations) can be found in Refs. [41,42].

Apart from statistical and experimental systematic errors, certain theoretical errors on μ_i^{ex} are included in the public results. To the best of our knowledge, the combination between these experimental and theoretical uncertainties is often made in quadrature. We thus subtract in quadrature these theoretical errors from the provided total uncertainties. How to properly (re)introduce the theoretical errors constitutes the main topic of this paper, and will be discussed at length in the upcoming sections.

Finally, we mention that we do not include in our fits more challenging observables related to the Higgs pair production [43], off-shell effects, loop-induced $Z\gamma$ final state, electron/muon pair final states, final states induced by flavour-changing Higgs couplings, nor exotic or invisible final states. Some of those would require to introduce new parameters in the Lagrangian that we will consider in Eq. (4.7). The motivation is to keep a simple physical framework in order to discuss easily the statistical aspects. In any case, the present experimental limits on such Higgs observables are still not stringent enough to affect drastically the Higgs fits. Moreover, all the statistical concepts discussed throughout the paper can be simply extended to new Higgs observables.

4.2. New physics parametrisation

The new physics possibly lying beyond the SM may induce a distortion of the SM Higgs couplings. The correct way of dealing with the low-energy manifestation of heavy new physics is through the use of an effective Lagrangian (see *e.g.* Ref. [16] for global fits of the Higgs effective Lagrangian). The leading effects on the Higgs sector appear through dimension-6 operators. The effective Lagrangian then induces anomalous couplings between the Higgs and the SM particles. The anomalous couplings to weak bosons and to heavy fermions can be parametrised as

$$\begin{aligned} \mathcal{L}_H = & c_W g_{hWW} h W_\mu^+ W^{-\mu} + c_Z g_{hZZ} h Z_\mu^0 Z^{0\mu} \\ & - c_t y_t h \bar{t}_L t_R - c_b y_b h \bar{b}_L b_R - c_c y_c h \bar{c}_L c_R - c_\tau y_\tau h \bar{\tau}_L \tau_R + \text{h.c.} \end{aligned} \quad (4.7)$$

where $y_{t,b,c,\tau}$ are the SM Yukawa coupling constants (in mass eigenbasis), the subscript L/R indicates the fermion chirality, v is the Higgs vacuum expectation value, $g_{hWW} = 2M_W^2/v$ and $g_{hZZ} = M_Z^2/v$ are the EW gauge boson couplings. The $c_{W,Z,t,b,c,\tau}$ parameters are defined such that the limiting case $c_{W,Z,t,b,c,\tau} \rightarrow 1$ corresponds to the SM. New tensor structures are also generated by the effective Lagrangian but are not taken into account here.

Our focus being on theoretical uncertainties, we adopt a fairly simple parametrisation of the new physics effects. We assume universal deviations for fermion couplings, $c_f \equiv c_t = c_b = c_c = c_\tau$, and for weak bosons, $c_V \equiv c_W = c_Z$. The c_f are assumed to be real. Clearly, this simplified description of the new physics effects represents only a piece (operators with no extra derivatives) of the full dimension-6 effective Lagrangian. Having $c_W \approx c_Z$ and c_f universality is however approximately compatible with certain new physics scenarios, like for a warped extra-dimension with bulk custodial symmetry vanishing IR brane kinetic terms for EW gauge bosons [44,45].²² Having only two parameters in this simplified framework, the results of our fits will systematically be presented in the c_V – c_f plane.

In the hypothesis of the existence of a physics Beyond the SM (BSM) parametrised by c_V – c_f , the expected signal strength is given by

$$\mu_i^{\text{th}}[c_V, c_f] = \frac{N_i^{\text{BSM}}[c_V, c_f]}{N_i^{\text{SM}}} = \frac{\sum_X \epsilon_{X,i}^{\text{BSM}} \sigma_X^{\text{BSM}} B_i^{\text{BSM}}}{\sum_X \epsilon_{X,i}^{\text{SM}} \sigma_X^{\text{SM}} B_i^{\text{SM}}}, \quad (4.8)$$

N_i^{SM} being defined in Eq. (4.5). This is the theoretical prediction of the experimental signal strength defined in Eq. (4.6). Both BSM cross sections and branching ratios σ_X^{BSM} , B_i^{BSM} can be expressed in terms of the SM amplitudes and of c_V, c_f . The expressions can for example

²² Note that contrary to a widespread belief, $c_W = c_Z$ is not entirely justified by custodial symmetry [44].

be found in Ref. [46], whose procedure is closely followed here. In all generality, the BSM efficiencies are not the same as the ones of the SM either. However, this happens when couplings with new tensors structures are generated by new physics. In our simplified framework, this does not happen, such that one can safely take $\epsilon_{X,i}^{\text{BSM}} = \epsilon_{X,i}^{\text{SM}} \equiv \epsilon_X^i$.

The SM production cross sections and partial decay widths for the Higgs boson are taken, respectively, from the LHC Higgs cross section Working Group (LHCHWG) Ref. [17] (see also Refs. [18–20] as well as the recent N³LO ggF computation [25]) and Refs. [17,20]. These numerical results correspond to the rates calculated at the highest orders of EW and QCD corrections known so far (mixed EW–QCD at NNLO for the ggF mechanism [27] and at NLO for other Higgs production modes).

5. The Higgs likelihood

5.1. The base likelihood

Having introduced the statistical framework and the Higgs data in Sections 2 to 4, we can proceed with building the Higgs likelihood function. We define the *base* likelihood L_0 as the likelihood containing the central values of Higgs signal strengths, and the experimental uncertainties. The theoretical uncertainties are kept apart from now. Their inclusion into the base likelihood will be discussed at length in the next sections and is the central topic of this paper.

In absence of any experimental systematic errors, a signal strength variable follows a Poisson statistics, and the associated likelihood is thus a Poisson distribution. Whenever the event number is large enough, about $O(10)$ in practice, the likelihood can be approximated by a Gaussian. In contrast, in presence of systematic uncertainties, this approximation generally does not hold. In practice however, the complete likelihood resulting from the combination of statistical and experimental systematic errors is not provided in the experimental public results. We will therefore model the base likelihood using Gaussian distributions, just as if the shape came out only from the statistical error. Such an approximation is expected to be good as long as the systematic error is small with respect to the statistical error, as shown in Section 3.4 and Appendix A.

The observed rates in the current 88 channels (labelled by i, j) are potentially correlated, for example because of the experimental error on the luminosity. The base likelihood follows therefore a multivariate normal distribution,

$$L_\mu(\mu_i^{\text{th}}, \mu_i^{\text{ex}}) = \exp \left[-\frac{1}{2} \sum_{i,j} (\mu_i^{\text{th}} - \mu_i^{\text{ex}}) C_{ij}^{\text{ex}^{-1}} (\mu_j^{\text{th}} - \mu_j^{\text{ex}}) \right], \quad (5.1)$$

where C_{ij}^{ex} is the correlation matrix among all channels.

Ideally, each individual observed channel i must be considered in order to take into account all the experimental information available on the signal strengths. In practice, few elements of this correlation matrix have been provided by the Collaborations up to now. Therefore in the following, we will include only the diagonal elements of C_{ij}^{ex} , given by $C_{ii}^{\text{ex}} = (\Delta\mu_i^{\text{ex}})^2$, where $\Delta\mu_i^{\text{ex}}$ is the experimental uncertainty extracted from the public experimental results. For future releases, we encourage the experimental Collaborations to provide as many elements as possible for the correlation matrix of the individual signal strengths.²³

²³ Also, we suggest that both the magnitudes of the uncertainties $\Delta\mu_i^{\text{ex}}$ and the correlations should be presented without ambiguities, so that the people exterior to the Collaborations be able to properly reconstruct the likelihood function.

Alternatively, to perform the Higgs fits one could think of using the correlations between the combined observed rates, that are currently provided by the LHC Collaborations. Although instructive, these combined rates do not keep track of all information since they are grouping together different Higgs production modes (which were originally measured independently), like $\mu_{\text{VBF,VH}}^{\text{ex}}$ and $\mu_{\text{ggF,tH}}^{\text{ex}}$ for each Higgs decay channel [6,7]. Notice that such combined signal strengths also hide some information in the sense that they can result from summations over various exclusive selection cut categories.

5.2. The uncertainty on the signal strengths

The Higgs theoretical uncertainties we will refer to are the theoretical uncertainties associated with the *expected* event rates N_i^{SM} defined in Eq. (4.5), that are obtained through analytical and numerical computations in quantum field theory. These uncertainties will propagate both into the experimental signal strengths μ_i^{ex} and into the theoretical strengths μ_i^{th} , defined in Eqs. (4.6), (4.8). Following our conventions (see Section 3, Eq. (3.1)), the theoretical uncertainty on the Standard Model expected rate in a channel i is written under the form

$$N_i^{\text{SM}}(1 + \delta_i^N \Delta_i^N), \quad (5.2)$$

where δ_i^N is the nuisance parameter with $E[\delta_i^N] = 0$, $V[\delta_i^N] = 1$, and Δ_i^N represents the relative magnitude of the uncertainty.

The theoretical uncertainty on N_i^{SM} propagates to the experimental signal strength as

$$\mu_i^{\text{ex}}(1 + \delta_i^\mu \Delta_i^\mu) = \mu_i^{\text{ex}}(1 - \delta_i^N \Delta_i^N). \quad (5.3)$$

The case of the theoretical signal strength $\mu_i^{\text{th}} = N_i^{\text{BSM}}/N_i^{\text{SM}}$ is slightly trickier. Here we focus on the most realistic case where the deviations induced by new physics are small, so that the anomalous couplings c_a (with $a = (W, Z, t, b, c, \tau)$) are close to one, *i.e.* $|c_a - 1| \ll 1$. The contributions from new physics can be linearised with respect to the small parameters $c_a - 1$, so that the BSM event rate in the channel i can be written as

$$N_i^{\text{BSM}} = N_i^{\text{SM}} + \sum_a (c_a - 1) N_{a,i}^{\text{BSM}} + O((c_a - 1)^2). \quad (5.4)$$

In this expression, it appears that the leading source of uncertainty comes from the SM event rate uncertainty Δ_i^N . In the expression of μ_i^{th} , it turns out that this uncertainty cancels out at first order between the numerator (N_i^{BSM}) and the denominator (N_i^{SM}). The subleading uncertainties would then come from a term quadratic in Δ_i^N and from the relative uncertainty $(c_a - 1) \frac{\Delta N_{a,i}^{\text{BSM}}}{N_{a,i}^{\text{BSM}}}$ on the components $N_{a,i}^{\text{BSM}}$. Notice that one can reasonably expect similar QCD errors in the SM and BSM predictions so that $\frac{\Delta N_{a,i}^{\text{BSM}}}{N_{a,i}^{\text{BSM}}} \sim \Delta_i^N$. These higher-order contributions are subleading compared to the error on the experimental signal strength, given in Eq. (5.3), which is of order Δ_i^N . In the following, we will thus focus only on the uncertainty of the experimental signal strength $\mu_i^{\text{ex}}(1 + \delta_i^\mu \Delta_i^\mu)$.

5.3. The structure of the Higgs theoretical uncertainties

The theoretical uncertainty on N_i^{SM} comes from the errors on the Higgs cross sections σ_X^{SM} and partial decay widths Γ_Y^{SM} . Still following our conventions, these relative uncertainties are written as

$$\sigma_X^{\text{SM}}(1 + \delta_X^\sigma \Delta_X^\sigma), \quad (5.5)$$

$$\Gamma_Y^{\text{SM}}(1 + \delta_Y^\Gamma \Delta_Y^\Gamma). \quad (5.6)$$

The exact content of these errors will be discussed in details in the next section.

The uncertainty on the partial decay width propagates to the branching ratios. Defining the relative error on the branching ratios as $B_Y^{\text{SM}}(1 + \delta_Y^B \Delta_Y^B)$, one has²⁴

$$\delta_Y^B \Delta_Y^B = \sum_{Y'} \delta_{Y'}^\Gamma \Delta_{Y'}^\Gamma \left(B_{Y'}^{\text{SM}} - \delta_{YY'} \right). \quad (5.7)$$

The uncertainty from the cross sections and branching ratios then propagates to the signal strength (4.6) and is thus encoded in a factor $\mu_i^{\text{ex}}(1 + \delta_i^\mu \Delta_i^\mu)$ where

$$\delta_i^\mu \Delta_i^\mu = -\delta_i^N \Delta_i^N = -\frac{\sum_X \epsilon_X^i \sigma_X^{\text{SM}} \delta_X^\sigma \Delta_X^\sigma}{\sum_{X'} \epsilon_{X'}^i \sigma_{X'}^{\text{SM}}} - \delta_{Y_i}^B \Delta_{Y_i}^B, \quad (5.8)$$

$Y_i = Y(i)$ being the Y decay mode of the Higgs channel detection i . Note that the sign after the first equal symbol is just a convention if the errors are symmetric.

Finally, the errors on cross sections and partial widths come from several sources. One can write those generically as

$$\delta_X^\sigma \Delta_X^\sigma = \sum_n \delta_X^n \Delta_X^n, \quad (5.9)$$

$$\delta_Y^\Gamma \Delta_Y^\Gamma = \sum_{n'} \delta_Y^{n'} \Delta_Y^{n'} \quad (5.10)$$

with the relative errors $\Delta_X^n, \Delta_Y^{n'}$ to be detailed in the following.²⁵

Knowing the base likelihood of Eq. (5.1), and knowing where exactly the theoretical uncertainties enter, we have the complete Higgs likelihood as a function of all the quantities that will have to be treated statistically, namely the nuisance parameters and the effective BSM parameters,²⁶

$$L_\mu \left(\mu_i^{\text{th}}[c_V, c_f]; \mu_i^{\text{ex}}(1 + \delta_i^\mu \Delta_i^\mu) \right) = L_0 \left(c_V, c_f; \delta_X^n, \delta_Y^{n'} \right). \quad (5.11)$$

Rigorously, the next step is to eliminate the nuisance parameters, $\delta_X^n, \delta_Y^{n'}$, applying either the marginalisation or the bias method. In general these steps should be performed numerically, and are computationally heavy. Here however, we will use the methods of preliminary combinations advocated in Section 3. Then it will appear that the subsequent Higgs likelihoods are much lighter to treat.

²⁴ $\delta_{YY'}$ represents the Kronecker symbol.

²⁵ Throughout the paper, we will systematically denote the values of $\Delta_X^n, \Delta_Y^{n'}$ taken from the literature by $\Delta|_0$ or Δ^0 . The possible ambiguities in the interpretation of these numbers will be discussed case by case.

²⁶ In the following, to adopt compact notation, we will omit the c_V, c_f arguments of the likelihood function when no ambiguity is possible.

6. Combining the Higgs rate uncertainties

In this section we shall combine the Higgs rate uncertainties that will be used in the marginal likelihood studied in Section 7. The most clear and rigorous statistical context for the marginalisation procedure is arguably the one of Bayesian statistics. In particular, the nuisance parameters are treated on the same ground as the variables of interest and are thus automatically given a probability distribution (see for instance Ref. [47]). For that reason we focus in this section on the error combinations within the *Bayesian* context. The resulting likelihood involving the combined errors will be formally treated within both the Bayesian and frequentist marginalisations in Section 7.

As we have described in Section 2.2.1, the Bayesian marginalisation procedure eliminates the dependence of the likelihood on the nuisance parameters through an integration. For the Higgs likelihood Eq. (5.11), this integration reads

$$L(c_V, c_f) = \int \left(\prod_{n,n',X,Y} d\delta_X^n d\delta_Y^{n'} \right) \pi_0(\delta_X^n, \delta_Y^{n'}) L_0(c_V, c_f; \delta_X^n, \delta_Y^{n'}), \quad (6.1)$$

where π_0 is the joint prior of all the nuisance parameters. Recall that this prior factorises when parameters are independent. More explicitly, this marginal likelihood reads

$$\begin{aligned} L(c_V, c_f) &= \int \left(\prod_{n,n',X,Y} d\delta_X^n d\delta_Y^{n'} \right) \pi_0(\delta_X^n, \delta_Y^{n'}) \exp \left[-\frac{1}{2} \sum_{i,j} (\mu_i^{\text{th}}[c_V, c_f] - \mu_i^{\text{ex}}(1 + \delta_i^\mu \Delta_i^\mu)) \right. \\ &\quad \left. \times C_{ij}^{\text{ex}-1} (\mu_j^{\text{th}}[c_V, c_f] - \mu_j^{\text{ex}}(1 + \delta_j^\mu \Delta_j^\mu)) \right]. \end{aligned} \quad (6.2)$$

The theoretical uncertainties $\delta_i^\mu \Delta_i^\mu$ on each signal strength μ_i are expressed in terms of the uncertainties on cross section $\delta_X^n \Delta_X^n$ and partial decay width $\delta_Y^{n'} \Delta_Y^{n'}$ through Eqs. (5.7) to (5.10).

In the following subsections, starting from Eq. (6.2), we will combine all the sources of uncertainty step-by-step, following the combination formalism established in Section 3. The aim of this section is to provide a clear and exhaustive treatment of all the Higgs theoretical uncertainties.

6.1. Combining the PDF and α_s uncertainties

Let us first discuss the errors on QCD predictions for the Higgs production cross sections at the proton level. Those are induced by the uncertainties on the parton Probability Density Functions (PDF) inside the proton. First, one may distinguish between two distinct origins to the PDF uncertainties: an experimental source – as the PDF are reconstructed from collider data – and the choice of a specific PDF set (MSTW, CT/CTEQ, NNPDF...).

Second, we consider simultaneously the parametric uncertainty coming from the strong coupling constant, α_s . We consider both PDF and α_s uncertainties simultaneously because they contribute in an intricate way to the cross section, as α_s enters both in the hard process matrix element and the PDF themselves.

• **MODELING THE UNCERTAINTIES:**

The uncertainties from α_s and the collider data are modelled by the nuisance parameters δ^{α_s} , δ^{data} and constitute independent sources of uncertainty (hence with factorisable priors). The relative uncertainties on α_s and the PDF data can be parametrised as

$$\alpha_s(1 + \delta^{\alpha_s} \Delta^{\alpha_s}), \quad \text{data}(1 + \delta^{\text{data}} \Delta^{\text{data}}). \quad (6.3)$$

The α_s error enters in the cross section in two different ways. On one hand, α_s is used in the fit of the data aimed at determining the PDF themselves. On the other hand, α_s is also involved in the hard subprocess that is convoluted with the PDF to obtain the final cross section. These two contributions to the cross section uncertainty, named here as $\Delta^{\alpha_s, \text{fit}}$ and $\Delta^{\alpha_s, \text{hard}}$, are not available in the literature. However, we will show that the knowledge of these two separate contributions is not necessary either. Rather, provided that the relative errors $\Delta^{\alpha_s, \text{fit}}$ and Δ^{data} are small enough to be linearised, only the sum $\Delta^{\alpha_s, \text{hard}} + \Delta^{\alpha_s, \text{fit}}$ is needed. This sum can typically be inferred from the literature.

In order to understand the interplay among the α_s and the data uncertainties, it is instructive to write explicitly how they enter into the cross section. One should start with the form

$$\sigma_X^{\text{SM}}[f_{\text{PDF}}[\alpha_s, \text{data}], \alpha_s], \quad (6.4)$$

where the first argument corresponds to the PDF input, while the second argument represents the α_s -dependence coming from the partonic process. From this general form, one then introduces the δ^{α_s} and δ^{data} nuisance parameters, and expand the expression at first order,²⁷

$$\begin{aligned} \sigma_X^{\text{SM}}[f_{\text{PDF}}[\alpha_s(1 + \delta^{\alpha_s} \Delta^{\alpha_s}), \text{data}(1 + \delta^{\text{data}} \Delta^{\text{data}})], \alpha_s(1 + \delta^{\alpha_s} \Delta^{\alpha_s})] \\ = \sigma_X^{\text{SM}}[f_{\text{PDF}}[\alpha_s, \text{data}], \alpha_s] \left(\left(1 + \delta^{\alpha_s} (\partial_1 f_{\text{PDF}} \partial_1 \sigma_X^{\text{SM}} \Delta^{\alpha_s}) + \delta^{\text{data}} (\partial_2 f_{\text{PDF}} \partial_1 \sigma_X^{\text{SM}} \Delta^{\text{data}}) \right. \right. \\ \left. \left. + \delta^{\alpha_s} (\partial_1 f_{\text{PDF}} \partial_2 \sigma_X^{\text{SM}} \Delta^{\alpha_s}) + O(\Delta^2) \right) \right). \end{aligned} \quad (6.5)$$

The terms in the last two lines represent the errors propagated to the cross section at first order in Δ , expressed as partial derivatives of σ_X^{SM} , and correspond precisely to the relative errors on the cross section,²⁸

$$\delta^{\alpha_s} \Delta^{\alpha_s, \text{fit}} + \delta^{\text{data}} \Delta^{\text{data}} + \delta^{\alpha_s} \Delta^{\alpha_s, \text{hard}}. \quad (6.6)$$

It appears clearly that only the sum $\Delta_X^{\alpha_s, \text{hard}} + \Delta_X^{\alpha_s, \text{fit}}$ is needed. Fortunately, this is what is provided in the literature. This sum $\Delta_X^{\alpha_s} \equiv \Delta_X^{\alpha_s, \text{hard}} + \Delta_X^{\alpha_s, \text{fit}}$ can be read for example from Ref. [20]. Note also that the nuisance parameter δ^{α_s} is common to any production mode, *i.e.* it does not carry the index X . In contrast, the nuisance parameter δ_X^{data} carries an index X because each production mode potentially involves different initial states. These initial states correspond to different PDF, which are fitted from different data sets.

Finally, one should check the validity of the error propagation at linear order in the cross sections (*i.e.* that the $O(\Delta^2)$ in Eq. (6.5) is well negligible). From Eqs. (6.5)–(6.6), one can see

²⁷ The $\partial_{1,2}$ represents derivative with respect to the first and second argument of the function respectively, $\partial_1 f = \partial f(x, y)/\partial x$, $\partial_2 f = \partial f(x, y)/\partial y$.

²⁸ Note that the Δ 's in Eq. (6.6) can be negative as they are identified from the partial derivatives in Eq. (6.5). In the rest of the paper however, the Δ 's are taken positive by convention. Different signs for the Δ 's would correspond to a negative correlation, that is instead included at the level of the δ 's in the rest of the paper.

that at linear order, for any fixed value of α_s (*i.e.* fixed value of δ^{α_s}), the error bar on σ_X^{SM} induced by the data uncertainty (obtained from varying δ^{data} , e.g. in $[-1, 1]$) should have the same size. A change with α_s of this bar size could thus come only from higher order terms such like

$$\delta^{\alpha_s} \delta^{\text{data}} (\partial_1 \partial_2 f_{\text{PDF}} \partial_1^2 \sigma_X^{\text{SM}} \Delta^{\alpha_s} \Delta^{\text{data}}).$$

In the Figs. 57–59 of Ref. [20] for the various Higgs production reactions at the 8 TeV LHC, we see that the change of this bar size (vertical bar there) is small with respect to the shift (*i.e.* $\Delta_X^{\alpha_s}$) of the bar central values. We conclude that one can restrict the expansion Eq. (6.5) to linear order in a good approximation.

Notice that a customary way to write these uncertainties is by splitting between the overall PDF error and the hard subprocess error, $\delta_X^{\text{PDF}} \Delta_X^{\text{PDF}} + \delta^{\alpha_s} \Delta_X^{\alpha_s, \text{hard}}$, with $\delta_X^{\text{PDF}} \Delta_X^{\text{PDF}} = \delta^{\alpha_s} \Delta_X^{\alpha_s, \text{fit}} + \delta_X^{\text{data}} \Delta_X^{\text{data}}$. The trouble when using this form is that the δ_X^{PDF} and $\delta^{\alpha_s, \text{hard}}$ contributions are correlated via α_s . Combining these uncertainties then requires to know such a correlation coefficient, which is fixed by $\Delta_X^{\alpha_s, \text{fit}}$, as well as $\Delta_X^{\alpha_s, \text{hard}}$. We emphasise that the use of this intermediate parametrisation brings unnecessary complications, and we recommend thus to avoid it.

Hence according to Eq. (6.6), the parametric uncertainties from α_s are cast into a single error $\Delta_X^{\alpha_s}$, and add up with the statistical error from the data as

$$\delta_X^{\text{data}} \Delta_X^{\text{data}} + \delta^{\alpha_s} \Delta_X^{\alpha_s}. \quad (6.7)$$

Using this approach, one deals directly with the elementary sources of uncertainty. These two sources of error have no intrinsic relation and are thus independent, meaning that δ^{data} and δ^{α_s} have factorisable priors.

Similarly, the uncertainty from the choice of a specific PDF set, modelled by δ^{set} , can be added up linearly to the errors of Eq. (6.7) in a good approximation. The linear approximation can be justified from Fig. 57 in Ref. [20]. There one can see that the size of the data error bars as well as the shifts induced by α_s depend only weakly on the PDF set choice. The δ^{set} error is also independent from the δ_X^{data} , δ_s^{α} errors and in turn possesses its own prior distribution. All those errors induce three terms in the sum of theoretical errors entering Eq. (5.9). These terms can be cast into a global PDF uncertainty,

$$\delta_X^{\text{PDF}+\alpha_s} \Delta_X^{\text{PDF}+\alpha_s} \triangleq \delta^{\text{set}} \Delta_X^{\text{set}} + \delta_X^{\text{data}} \Delta_X^{\text{data}} + \delta^{\alpha_s} \Delta_X^{\alpha_s}. \quad (6.8)$$

We recall that $X = \{\text{ggF}, \text{VBF}, \text{VH}, \text{ttH}\}$ and that the Δ 's are relative errors, which are chosen by convention to correspond to one standard deviation. Those are related to the 1σ absolute errors on the SM Higgs cross section through *e.g.*

$$\Delta_X^{\text{data}} \triangleq \frac{\Delta \sigma_X^{\text{data}}}{\sigma_X^{\text{SM}}}.$$

• COMBINING THE THREE UNCERTAINTIES:

Here we combine the three sources of theoretical uncertainty described in Eq. (6.8). We will add up more and more errors progressively in the following subsections. These three independent sources of error are associated with three priors π^{α_s} , π_X^{data} , π^{set} . These nuisance parameters appear in Eq. (6.2), where they are integrated over. We now proceed to combine these errors following the analysis of Section 3, starting from Eq. (3.2). In practice, for the discussion, it will be convenient to combine only two errors at a time. One then finds a likelihood of the type (6.2) depending only on the nuisance parameter $\delta_X^{\text{PDF}+\alpha_s}$. The distribution of this nuisance parameter comes with a 1σ width $\Delta_X^{\text{PDF}+\alpha_s}$ given by

$$(\Delta_X^{\text{PDF}+\alpha_s})^2 = (\Delta_X^{\text{set}})^2 + (\Delta_X^{\text{data}})^2 + (\Delta_X^{\alpha_s})^2. \quad (6.9)$$

The nuisance parameter $\delta_X^{\text{PDF}+\alpha_s}$ obeys a new prior $\pi_X^{\text{PDF}+\alpha_s}$, obtained via two successive convolutions of the initial priors (as in Eqs. (3.3)–(3.5)),

$$\bar{\pi}_X^{\text{PDF}+\alpha_s} = \bar{\pi}_X^{\text{set}} \star \bar{\pi}_X^{\text{data}} \star \bar{\pi}_X^{\alpha_s}, \quad (6.10)$$

where $\bar{\pi}_X^{\text{PDF}+\alpha_s}(x) = \pi_X^{\text{PDF}+\alpha_s}(x/\Delta_X^{\text{PDF}+\alpha_s})$ and the variable x corresponds to the relative error $\delta_X^{\text{PDF}+\alpha_s}/\Delta_X^{\text{PDF}+\alpha_s}$. For the initial priors one has for example $\bar{\pi}_X^{\alpha_s}(x) = \pi^{\alpha_s}(x/\Delta_X^{\alpha_s})$. The Eq. (6.9) and then (6.10) are justified in details in the rest of this subsection.

• DETAILS ON THE DATA AND α_s ERROR COMBINATIONS:

We emphasise that the Bayesian combination of the 1σ widths, as here in Eq. (6.9), is *independent* of the shapes of the prior distributions. This combination only depends on the possible correlations among individual errors [cf. Section 3.2]. In the present case, there is no correlation between the δ_X^{data} and $\delta_X^{\alpha_s}$ parameters, as explained right below Eq. (6.7). This leads to the sum in quadrature of the 1σ errors $(\Delta_X^{\text{data}})^2 + (\Delta_X^{\alpha_s})^2$ in Eq. (6.9).

Let us comment about those uncertainties. First, the error associated to π_X^{data} originates mainly from measurements: it is mainly induced by the limited accuracy of data points used to perform the fit for reconstructing PDF. Hence this error is mostly of statistical nature. There exists of course systematic errors as well, but it has been checked by several groups that the final π_X^{data} distribution can be reasonably taken as Gaussian [18].

Second, the uncertainty on α_s originates mainly from lattice calculation errors (mainly theoretical) and especially from perturbative truncation errors [48].²⁹ Indeed the α_s determination from lattice methods (most accurate one in Ref. [48]) represents today the most precise determination and hence essentially dictates the final world average error [30]. The FLAG Working Group on lattice calculations has estimated a more conservative uncertainty on α_s , which is increased by a new QCD perturbative error estimation [49], thus still leading to a dominant theoretical uncertainty.

At this level, a comment is needed on the link between the 1σ errors and the uncertainty magnitudes provided in literature. To remain conservative we use $\Delta_X^{\alpha_s} = \Delta_X^{\alpha_s}|_0$ for the 1σ error, where $\Delta_X^{\alpha_s}|_0$ is the error provided by Refs. [17,20]. There is indeed a somewhat arbitrary choice for the relation between $\Delta_X^{\alpha_s}$ and $\Delta_X^{\alpha_s}|_0$, due to the theoretical (QCD) nature of the uncertainty. The origin of this arbitrariness is the fact that the QCD errors are just estimated by varying the renormalisation and factorisation scales on arbitrary intervals. We present a similar discussion in the beginning of next Section 6.2 for Δ_X^{scale} . Concerning the 1σ error from data, one can adopt $\Delta_X^{\text{data}} = \Delta_X^{\text{data}}|_0$ ($\Delta_X^{\text{data}}|_0$ being read from Refs. [17,20]). Indeed, the probability distribution for the uncertainty induced by the experimental data can be safely described by a Gaussian, as described above, so that the errors provided by Refs. [17,20] can reasonably be interpreted as 1σ errors.

Let us now discuss the convolution between $\bar{\pi}_X^{\text{data}}$ and $\bar{\pi}_X^{\alpha_s}$ that appears in Eq. (6.10). For that purpose, we first need to discuss the form of the $\pi_X^{\alpha_s}$ distribution. The shape of $\pi_X^{\alpha_s}$ can be taken as flat since the uncertainty on α_s originates mainly from theoretical uncertainty, as mentioned above. However, the choice of the prior for a theoretical uncertainty is often controversial, so that

²⁹ The only source of experimental error is, m_{η_c}, m_{η_b} , and is minor – as can be read from the Table IV of Ref. [48].

we will also consider the case of a non-flat π^{α_s} distribution.³⁰ Let us bring a precision on the treatment here. Once a hypothesis (possibly motivated by physical considerations) is chosen for the shape of the prior distribution, it is then rigorously considered as correct from the statistical point of view.

Finally, the convolution of the Gaussian prior, $\bar{\pi}_X^{\text{data}}$, with a flat prior, $\bar{\pi}_X^{\alpha_s}$, gives rise to a Gaussian distribution, $\bar{\pi}_X^{\text{data}} \star \bar{\pi}_X^{\alpha_s}$, in a good approximation for the various Higgs production modes. The justification is that the $\bar{\pi}_X^{\alpha_s}$ width, $\Delta_X^{\alpha_s}$, is systematically smaller or of the same order as Δ_X^{data} ,³¹ in which case the convolution leads to an almost pure Gaussian prior. This will be demonstrated explicitly in Fig. 2 for other priors.

• DETAILS ON THE COMBINATION WITH THE PDF SET ERROR:

The various PDF estimations provided by the different fitting groups reflect several sources of error [50–52]. Indeed, these groups make different choices/hypotheses about the numbers of free parameters used to model the PDF,³² the statistical methods adopted to fit the data,³³ the number of independently parameterised PDF (in particular regarding (anti-) strangeness), the collider results exploited, the matching methods applied to include heavy-quark mass effects in the flavour number scheme and the variable- or fixed-flavour number scheme. All these sources of uncertainty are synthesised in the 1σ error on the Higgs production rates noted Δ_X^{set} . To remain conservative, we assume $\Delta_X^{\text{set}} = \Delta_X^{\text{set}}|_0$, where $\Delta_X^{\text{set}}|_0$ is the error reads from Figs. 57–59 of Ref. [20]. $\Delta_X^{\text{set}}|_0$ can be estimated by taking half the interval obtained by using the various PDF sets which lead to a finite number of predictions for the Higgs rate central values. Of course, this determination of $\Delta_X^{\text{set}}|_0$ is probably underestimated as (i) the hypotheses made by the groups provide illustrative examples which do not necessarily indicate the extremal values of the PDF, and, (ii) the effects of the various sources of error listed above can potentially compensate each other. We comment on this point in the following paragraph.

In Eq. (6.9), the sum in quadrature between the Δ_X^{set} error and the data and α_s errors is justified because these are independent uncertainties. Nevertheless, in practice, for our numerical applications, we use the so-called envelope method³⁴ to determine $\Delta_X^{\text{PDF}+\alpha_s}$ as done in Refs. [20,53]³⁵ and calculated by the LHCHWG [17]. Note that the envelope method overestimates the com-

³⁰ To be consistent throughout the paper, concerning the initial priors, we will assume a flat shape for the distributions whose shape is unknown (uncertainties from QCD, parametrisation...).

³¹ For the ggF example, our conservative treatment of the errors provided in Fig. 59 of Ref. [20] gives an half absolute width, $W/2 \triangleq \sqrt{3}\Delta\sigma_X^{\alpha_s} = \sqrt{3}\Delta\sigma_X^{\alpha_s}|_0 \simeq 0.5$ pb, which is indeed comparable to, $\Delta\sigma_X^{\text{data}} = \Delta\sigma_X^{\text{data}}|_0 \simeq 0.5$ pb. In the alternative case (see the analogous discussion at the start of Section 6.2), one has instead, $W/2 \triangleq \sqrt{3}\Delta\sigma_X^{\alpha_s} = \Delta\sigma_X^{\alpha_s}|_0 \simeq 0.3$ pb, which is clearly smaller than, $\Delta\sigma_X^{\text{data}} \simeq 0.5$ pb, so that the Gaussian approximation for the final convolution would be even better because this case would tend to a situation where the non-Gaussian error becomes negligible.

³² The infinite-dimensional problem of representing a space of functions is reduced to a finite-dimensional form, in order to be manageable, by introducing a parametrisation of the PDF.

³³ There exist mainly two classes of methodology currently used to determine a confidence interval represented in the space of functions: some variations of the Hessian approach (multi-Gaussian probability distributions) and the Monte Carlo approach. Both types of methods have their own limitations.

³⁴ This “envelope method” corresponds precisely to the uncertainty combinations in the bias approach, see Section 3.5. What we call envelope method in the present paper is rather described in Section 2.2.2.

³⁵ In the envelope method used in this reference, the whole uncertainty interval is found by searching at the minimum and maximum rates (considering the various PDF sets, α_s values and including the possibility to move along the data-error bars). Then dividing by two this interval gives an estimation of the combined error as well as a central value for the rate.

binned errors, compensating somehow for the underestimation of the PDF set error. For the ggF mechanism, the $\Delta\sigma_{\text{ggF}}^{\text{PDF}+\alpha_s}$ error derived in this way has to be reduced by $\sim 40\%$ to recover the quadrature summation of Eq. (6.9), and the decrease is smaller for the other Higgs production reactions. Hence, we conclude that the use of the envelope method to determine the global PDF uncertainties gives rise to a substantial overestimation of these errors.

We finally discuss the shape of the prior of the final combination $\pi_X^{\text{PDF}+\alpha_s}$. Most of the sources of error taken into account in Δ_X^{set} are of theoretical nature and all the errors have unknown distributions. The shape of π_X^{set} is therefore assumed to be flat. The convolution of $\bar{\pi}_X^{\text{set}}$ (see Eq. (6.10)) with the nearly Gaussian distribution $\bar{\pi}_X^{\text{data}} \star \bar{\pi}_X^{\alpha_s}$ leads in a good approximation to a final Gaussian prior, $\pi_X^{\text{PDF}+\alpha_s}$.³⁶ Once more, this is guaranteed by the fact that for any Higgs production mode at the LHC, Δ_X^{set} is smaller or comparable to the combination of Δ_X^{data} and $\Delta_X^{\alpha_s}$ (see for instance Ref. [20]).

6.2. Scale and EFT errors: the amplitude uncertainties

• SCALE ERROR:

There exists another major type of error, this time at the parton level, on the QCD prediction for Higgs production cross sections. It originates from the lack of knowledge on the higher order contributions to the amplitude in the perturbative expansion, and can be recast into the dependence on the QCD renormalisation and factorisation scales. We note δ_X^{scale} the nuisance parameter representing this “scale uncertainty”.

There are no strong arguments to choose the shape for π_X^{scale} . As for many other theoretical uncertainties, the choice of the prior is typically a subject of controversy. Here we choose π_X^{scale} to be flat. Concerning the magnitude of the scale uncertainty Δ_X^{scale} , it is also not clear to which width exactly corresponds the provided value, noted Δ_X^0 here, that is found in Refs. [17, 18, 25]. It is reasonable to expect Δ_X^{scale} to be of order Δ_X^0 . To be more precise, we could make the two different assumptions, $\Delta_X^0 \hat{=} \Delta_X^{\text{scale}}$ or $\Delta_X^0 \hat{=} \mathcal{W}/2$ where \mathcal{W} is defined as the support of the distribution,³⁷ with *e.g.* in the case of a flat distribution on an interval with size \mathcal{W} : $2\Delta_X^{\text{scale}} \hat{=} \mathcal{W}/\sqrt{3} = 2\Delta_X^0/\sqrt{3}$. In order to be conservative in the choice of Δ_X^{scale} , we choose the former hypothesis throughout this paper: $\Delta_X^{\text{scale}} = \Delta_X^0$.

It is remarkable that recently [25], the calculation for the ggF mechanism has been pushed up to the complete N³LO order in perturbative QCD. This has allowed a reduction of the symmetrised³⁸ scale error from $\Delta_{\text{ggF}}^0 \simeq 7.51\%$ (with the renormalisation/factorisation scale $\mu_0 = m_H/2$ to absorb some of the soft-gluon resummation corrections [54]) [17, 18], down to $\Delta_{\text{ggF}}^0 \simeq 4.16\%$ (with $\mu_0 = m_H$)³⁹ [25]. The error was obtained in both cases by spanning the interval $[\mu_0/2, 2\mu_0]$, for the renormalisation/factorisation scale $\mu = \mu_R = \mu_F$, at an energy $\sqrt{s} = 8$ TeV and for $m_H \simeq 125.2$ GeV.

³⁶ Given that there are several sources of errors contained in the PDF set uncertainty, one may expect the π_X^{set} prior to be somehow peaked. This feature improves even more the Gaussian approximation of $\pi_X^{\text{PDF}+\alpha_s}$.

³⁷ Recall that the support of a distribution is the domain where this distribution is not zero-valued.

³⁸ Symmetrised over the positive and negative errors as, $\Delta = [(\Delta_+^2 + \Delta_-^2)/2]^{1/2}$.

³⁹ Choosing instead, $\mu_0 = m_H/2$, could be motivated by a faster convergence of the perturbative series [25]. However, since it would lead to a significantly smaller uncertainty, $\Delta_{\text{ggF}}^0 \simeq 2.13\%$, we stick to the central choice, $\mu_0 = m_H$, in order to remain conservative.

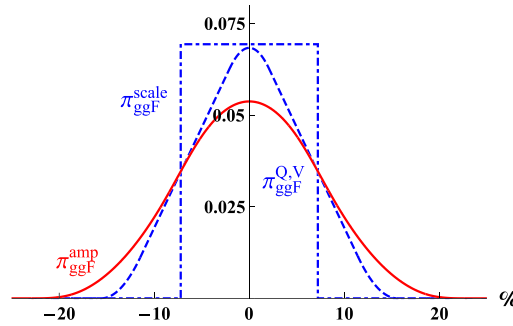


Fig. 1. Probability density distribution, $\pi_{\text{ggF}}^{\text{amp}}(x/\Delta_{\text{ggF}}^{\text{amp}})$ (plain red line), involving the relative error x (in %) of the ggF cross section, as derived through the convolution of the $\pi_{\text{ggF}}^{\text{Q,V}}$ (dashed blue line) and $\pi_{\text{ggF}}^{\text{scale}}$ (dot-dashed blue line) priors. The quantity $\Delta_{\text{ggF}}^{\text{amp}}$ represents the relative 1σ error on the Higgs production rate (see text). The distributions are normalised. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

• EFT ERROR:

In the specific case of the ggF mechanism, another source of error arises in the amplitude of the Higgs production [55], that we describe now. The evaluation of this amplitude beyond the NLO level is possible within the Effective Field Theory (EFT) approach, where the particles running in the triangle loop are assumed to be much heavier than the produced Higgs boson to integrate out the heavy particles.

For the top quark exchange, the infinite mass assumption, $m_t \gg m_H$, induces a negligible error on the ggF amplitude [27,56]. In contrast, the EFT approach is clearly not valid for the other significant ggF contribution: the bottom quark exchange [25]. This inappropriate use of the EFT limit introduces some non-negligible error mainly through the interference between the bottom and dominant top quark loops (this error being smaller at the Tevatron than at the LHC) [57].

A similar uncertainty originates from the mixed QCD-EW corrections to the ggF process [27]. Those have been calculated at NNLO via the EFT approach based on the simplifying but unrealistic assumption, $M_{W,Z} \gg m_H$. For all the EFT errors, some approximative estimations can be computed at NNLO (using K -factors obtained at NLO and NNLO for the top loop) [26,56].

A related uncertainty comes from the freedom in the choice of a renormalisation scheme for the bottom quark mass, involved in the ggF amplitude (on-shell scheme, $\overline{\text{MS}}$ scheme...). The error from the renormalisation scheme dependence can be approximately estimated at NLO [56].

These three sources of theoretical uncertainty, namely the two kinds of EFT assumptions (on the heavy quark masses, m_Q ($Q = b, t$), and vector boson masses, M_V ($V = W, Z$)) and the m_b scheme dependence, are independent and their respective priors are unknown. We assume these priors to be flat. To be conservative, we take the three 1σ errors to be equal to the numbers estimated in Refs. [26,56], for the 8 TeV LHC. Summing those in quadrature gives rise to the relative rate error, $\Delta_{\text{ggF}}^{\text{Q,V}} \triangleq \Delta\sigma_{\text{ggF}}^{\text{Q,V}}/\sigma_{\text{ggF}}^{\text{SM}} \simeq 5.6\%$. The convolution of the three flat priors (accordingly to Eq. (3.5)) leads to the blue distribution, $\pi_{\text{ggF}}^{\text{Q,V}}$, shown in Fig. 1, which already resembles a Gaussian shape as predicted by the central limit theorem.

• COMBINING THE $\Delta_{\text{ggF}}^{\text{scale}}$ AND $\Delta_{\text{ggF}}^{\text{Q,V}}$ ERRORS:

The theoretical scale and EFT uncertainties on the ggF mechanism are of different nature and are thus independent. The combined ggF 1σ error is in turn given by

$$(\Delta_{\text{ggF}}^{\text{amp}})^2 = (\Delta_{\text{ggF}}^{\text{scale}})^2 + (\Delta_{\text{ggF}}^{\text{Q,V}})^2. \quad (6.11)$$

This error constitutes the characteristic width of the $\pi_{\text{ggF}}^{\text{amp}}$ distribution obtained by convoluting the $\bar{\pi}_{\text{ggF}}^{\text{scale}}$ and $\bar{\pi}_{\text{ggF}}^{\text{Q,V}}$ priors, as performed in Fig. 1 (see the final red curve). Remarkably, this distribution,

$$\bar{\pi}_{\text{ggF}}^{\text{amp}} \equiv \bar{\pi}_{\text{ggF}}^{\text{scale}} \star \bar{\pi}_{\text{ggF}}^{\text{Q,V}}, \quad (6.12)$$

derived from four purely flat priors, is Gaussian in a good approximation. This can be also seen in Fig. 3 where $\pi_{\text{ggF}}^{\text{amp}}$ is plotted together with a pure Gaussian distribution (blue curves). Recall that $\bar{\pi}_{\text{ggF}}^{\text{amp}}(x) = \pi_{\text{ggF}}^{\text{amp}}(x/\Delta_{\text{ggF}}^{\text{amp}})$ and the variable x corresponds to $\delta_{\text{ggF}}^{\text{amp}} \Delta_{\text{ggF}}^{\text{amp}}$.

6.3. Combination of the PDF and amplitude errors

For the various Higgs production modes – except the ggF process that will be discussed separately below, one has to combine the PDF and scale errors to determine the final uncertainty on the whole cross section. The scale error adds up to the PDF error of Eq. (6.8), according to Eq. (5.9), defining the total uncertainty on the cross section,

$$\delta_X^\sigma \Delta_X^\sigma = \delta_X^{\text{PDF}+\alpha_s} \Delta_X^{\text{PDF}+\alpha_s} + \delta_X^{\text{scale}} \Delta_X^{\text{scale}}. \quad (6.13)$$

These errors being independent, the 1σ widths add-up in quadrature,

$$(\Delta_X^\sigma)^2 = (\Delta_X^{\text{PDF}+\alpha_s})^2 + (\Delta_X^{\text{scale}})^2, \quad (6.14)$$

as dictated by Section 3.2, *i.e.* irrespective of the $\pi_X^{\text{PDF}+\alpha_s}$ and π_X^{scale} shapes. Recall that Δ_X^σ is the 1σ width of the resulting $\bar{\pi}_X^\sigma$ distribution. The prior π_X^σ of this total uncertainty is then given by (see Eq. (3.5))

$$\bar{\pi}_X^\sigma \equiv \bar{\pi}_X^{\text{PDF}+\alpha_s} \star \bar{\pi}_X^{\text{scale}}, \quad (6.15)$$

with $\bar{\pi}_X^\sigma(x) = \pi_X^\sigma(x/\Delta_X^\sigma)$ and x corresponding to $\delta_X^\sigma \Delta_X^\sigma$.

Let us discuss the form of the π_X^σ function, as generated through Eq. (6.15). The shape of π_X^{scale} being unknown, we assume a flat π_X^{scale} distribution. Remind that this error is simply obtained by varying the QCD scale, so that no favoured value is predicted for the cross section. It is therefore a sensible choice to assign equal probabilities to all the values of δ_X^{scale} (or equivalently of the Higgs cross section) inside a certain range. On the other hand, we have seen in Section 6.1 that $\pi_X^{\text{PDF}+\alpha_s}$ is approximatively Gaussian. Given the relative values of $\Delta_X^{\text{PDF}+\alpha_s}$ and Δ_X^{scale} for each process X – which are systematically such that either $\Delta_X^{\text{PDF}+\alpha_s} > \Delta_X^{\text{scale}}$ or $\Delta_X^{\text{PDF}+\alpha_s} \approx \Delta_X^{\text{scale}}$ ⁴⁰ – a Gaussian $\pi_X^{\text{PDF}+\alpha_s}$ and a flat π_X^{scale} lead in a good approximation to a final Gaussian π_X^σ . This combination is shown in Fig. 2 for ZH production, for which $\Delta_{\text{ZH}}^{\text{PDF}+\alpha_s} \simeq 2.5\%$ and $\Delta_{\text{ZH}}^{\text{scale}} = \Delta_{\text{ZH}}^0 \simeq 3.1\%$ (at $\sqrt{s} = 8$ TeV with $m_H \simeq 125.2$ GeV) [17].

• THE ggF REACTION:

In the case of Higgs production via the ggF mechanism, the PDF error has to be combined with the whole amplitude error studied previously in Section 6.2. The resulting total error on the cross section is

⁴⁰ Whatever is the prescription: $\Delta_X^{\text{scale}} = \Delta_X^0$ or $\Delta_X^{\text{scale}} = \Delta_X^0/\sqrt{3}$.

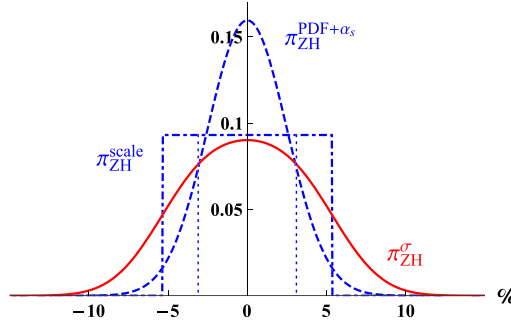


Fig. 2. Probability density distribution, $\pi_{ZH}^\sigma(x/\Delta_{ZH}^\sigma)$ (plain red line), involving the relative error x (in %) of the ZH production cross section, as derived through the convolution of a Gaussian $\pi_{ZH}^{PDF+\alpha_s}$ (dashed blue line) and a flat π_{ZH}^{scale} (dot-dashed blue line) priors. The quantity, Δ_{ZH}^σ , represents the relative 1σ error on the Higgs production rate. The distributions are normalised. The 1σ band for the π_{ZH}^{scale} distribution is indicated as well (vertical dotted blue lines). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

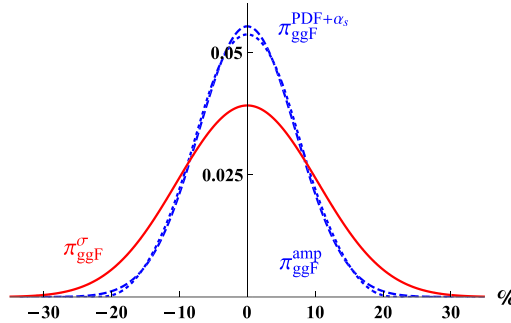


Fig. 3. Probability density distribution, $\pi_{ggF}^\sigma(x/\Delta_{ggF}^\sigma)$ (plain red line), involving the relative error x (in %) of the ggF cross section, as derived through the convolution of a Gaussian $\pi_{ggF}^{PDF+\alpha_s}$ prior (dashed blue line) and the π_{ggF}^{amp} distribution (dotted blue line) obtained in Fig. 1. The quantity, Δ_{ggF}^σ , represents the relative 1σ error for the ggF rate. The distributions are normalised. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

$$\delta_{ggF}^\sigma \Delta_{ggF}^\sigma = \delta_{ggF}^{PDF+\alpha_s} \Delta_{ggF}^{PDF+\alpha_s} + \delta_{ggF}^{amp} \Delta_{ggF}^{amp}. \quad (6.16)$$

These two errors being independent, their widths add-up in quadrature,

$$(\Delta_{ggF}^\sigma)^2 = (\Delta_{ggF}^{PDF+\alpha_s})^2 + (\Delta_{ggF}^{amp})^2, \quad (6.17)$$

and their priors are convoluted following

$$\tilde{\pi}_{ggF}^\sigma \equiv \tilde{\pi}_{ggF}^{PDF+\alpha_s} \star \tilde{\pi}_{ggF}^{amp}. \quad (6.18)$$

This convolution (6.18) is performed in Fig. 3, using the π_{ggF}^{amp} distribution obtained in Fig. 1 and the value $\Delta_{ggF}^{PDF+\alpha_s} \simeq 7.20\%$ (at $\sqrt{s} = 8$ TeV with $m_H \simeq 125.2$ GeV) [17]. Both priors π_{ggF}^{amp} , $\pi_{ggF}^{PDF+\alpha_s}$ being nearly Gaussian, the final distribution is almost Gaussian.⁴¹

⁴¹ Recall the convolution of two Gaussian distributions gives rise to a Gaussian distribution.

6.4. The production contamination

There are several production mechanisms for the Higgs boson (recall that $X = \{\text{ggF}, \text{VBF}, \text{WH}, \text{ZH}, \text{ttH}\}$). The cross section for each of these production modes is associated with a theoretical uncertainty, that has been obtained through subsections 6.1 to 6.3. In fact, one may note that the uncertainties of these various cross sections are potentially correlated, as they partly arise from common sources like the α_s parametric error. Therefore the δ_X^σ follow a common distribution π^σ , which does not necessarily factorise into $\pi_{\text{ggF}}^\sigma \pi_{\text{VBF}}^\sigma \times \dots$. The aspect of correlations among the cross section errors will be further discussed in Section 7.1. Here we shall proceed using the most general prior π^σ , and we denote the resulting correlation matrix as $\rho_{XX'}^\sigma$.⁴²

The contribution from the cross sections errors in a given detection channel can be read from Eq. (5.8). Let us first adopt a more compact notation,

$$\frac{\sum_X \epsilon_X^i \sigma_X^{\text{SM}} \delta_X^\sigma \Delta_X^\sigma}{\sum_{X'} \epsilon_{X'}^i \sigma_{X'}^{\text{SM}}} \triangleq \sum_X \delta_X^\sigma \Delta_{X,i}, \quad (6.19)$$

where the $\delta_X^\sigma \Delta_X^\sigma$ are defined in Eqs. (6.13), (6.16). The Higgs detection channels have been designed to select predominantly a certain mode of production. That is, for a given channel i , the experimental cuts are profiled so that typically the efficiency ϵ_X^i for one of the production modes X (see Eq. (4.6)) is much larger than for the others, implying a hierarchy among the $\Delta_{X,i}$. We can therefore use the leading moment approximation, developed in Section 3 and Appendix A, to proceed to the combination of the errors. Applying the leading moment approximation amounts to treat the contaminations as a small perturbation of the uncertainty from the leading production mode. The cross section uncertainties propagate in a given detection channel as (P stands for production)

$$\delta_{X_i}^P \Delta_i^P = \delta_{\text{ggF}}^\sigma \Delta_{\text{ggF},i} + \delta_{\text{VBF}}^\sigma \Delta_{\text{VBF},i} + \delta_{\text{ZH}}^\sigma \Delta_{\text{ZH},i} + \delta_{\text{WH}}^\sigma \Delta_{\text{WH},i} + \delta_{\text{ttH}}^\sigma \Delta_{\text{ttH},i}. \quad (6.20)$$

Here the label of the combined nuisance parameter $\delta_{X_i}^P$ is chosen to be the label of the dominant production mode in the i channel. Note that X_i should be understood as $X(i)$. This naming refers to the fact that the shape of the combined nuisance parameter prior corresponds approximatively to the shape for the dominant uncertainty, see Eq. (3.14). For example, if the production mode ggF dominates in the channel i , one has

$$\delta_{X_i}^P = \delta_{\text{ggF}}^P. \quad (6.21)$$

The various nuisance parameters δ_X^P are potentially correlated. They should thus follow a joint prior distribution, π^P , generating a correlation matrix $\rho_{XX'}^P$.

Assuming generic correlations $\rho_{XX'}^\sigma$ among the various cross section errors, the magnitude of the combined production uncertainty in a channel i is given exactly by

$$(\Delta_i^P)^2 = \sum_{XX'} \rho_{XX'}^\sigma \Delta_{X,i} \Delta_{X',i}. \quad (6.22)$$

The leading moment approximation then dictates (see Eqs. (3.16)–(3.19)) that

$$\pi^P \approx \pi^\sigma. \quad (6.23)$$

⁴² In Section 7.1, the assumptions adopted for $\rho_{XX'}^\sigma$ will allow us to express π^σ in terms of the π_X^σ .

Equation (6.23) implies that the correlations among the δ_X^P are approximatively the same as the ones between the δ_X^σ , *i.e.*

$$\rho_{XX'}^P \approx \rho_{XX'}^\sigma. \quad (6.24)$$

This fact can be understood as follows. Consider only two detection channels, i and j . If the same production mode $X \triangleq X_i = X_j$ dominates in both channels, they are nearly 100% correlated, so that they are described by a single nuisance parameter δ_X^P , which is equivalent to say that $\rho_{XX}^P \approx 1$. Note that one has $\rho_{XX}^\sigma = 1$ by definition, so that $\rho_{XX}^P \approx \rho_{XX}^\sigma$. Besides, if two different production modes $X_i \neq X_j$ dominate respectively in the i and j channels, the uncertainties in both channels are respectively described by $\delta_{X_i}^P$ and $\delta_{X_j}^P$. These two nuisance parameters inherit the correlation from the leading production modes X_i and X_j , which is given by $\rho_{X_i X_j}^\sigma$. Therefore one recovers Eq. (6.24).

Finally, notice that for certain kinematical cuts selecting the ttH mode in the diphoton decay channel [31], even additional production modes can slightly contribute, like the bbH , tHW and $tHbq$ productions. These production modes participate in the contamination and have thus been included in the combination of production modes in Eq. (6.20).

6.5. The uncertainties on branching ratios

Two sources of error affect the Higgs signal strengths: the production and the decay rate uncertainties (see Eq. (4.6)). The latter is often not considered in the Higgs fits. Still following our approach of step-by-step combinations, one should start with the signal strength error Eq. (5.8), where all uncertainties on production modes have been already combined (Eq. (6.20)). The uncertainties on production and decay rates combine thus as, up to an irrelevant global sign,

$$\delta_{X_i}^\mu \Delta_i^\mu = \delta_{X_i}^P \Delta_i^P + \delta_{Y_i}^B \Delta_{Y_i}^B \quad \text{with} \quad \Delta_{Y_i}^B = \frac{\Delta B_i^{\text{SM}}}{B_i^{\text{SM}}}, \quad B_i^{\text{SM}} = \frac{\Gamma_{Y_i}^{\text{SM}}}{\Gamma_{\text{tot}}} \quad (6.25)$$

where $\Gamma_{Y_i}^{\text{SM}}$ is the SM partial decay width for the detection channel i . In this equation, we apply the leading moment approximation to treat the branching ratios errors as perturbations of the leading error from production modes. This is why the $\delta_{X_i}^\mu$ parameters carry the index X_i , which is the index of the dominant production mode in the channel i , as in the previous subsection. For example, if the production mode ggF dominates in the channel i , one has

$$\delta_{X_i}^\mu = \delta_{\text{ggF}}^\mu. \quad (6.26)$$

The relative error $\delta_{Y_i}^B \Delta_{Y_i}^B$ on the SM branching ratio is expressed as in Eq. (5.7), where the decay width uncertainty (5.10) can now be specified in terms of the various sources of error (*cf.* Section 3 of Ref. [56] for a recent overview, and references therein),

$$\delta_Y^\Gamma \Delta_Y^\Gamma = \sum_a \delta_Y^{\text{pu}_a} \Delta_Y^{\text{pu}_a} + \delta_Y^{\text{thu}} \Delta_Y^{\text{thu}} \quad \text{where e.g.} \quad \Delta_Y^{\text{thu}} = \frac{\Delta \Gamma_Y^{\text{thu}}}{\Gamma_Y^{\text{SM}}}. \quad (6.27)$$

The partial decay width errors $\Delta \Gamma_Y^{\text{thu/pu}_a}$ are taken from the LHCHWG [17,18,20]. The $\Delta \Gamma_Y^{\text{thu}}$ denote the theoretical uncertainties due to the limitations of QCD perturbative calculations. The $\Delta \Gamma_Y^{\text{pu}_a}$ represent the parametric uncertainties induced by the experimental errors on the input parameters, labelled by $a \equiv \alpha_s, m_c, m_b, m_t$ (charm, bottom and top quark masses). Typically,

one has $\Delta_{b\bar{b}}^{\text{th}/\text{pu}_a} \gg \Delta_{VV^*}^{\text{th}/\text{pu}_a}$, $\Delta_{\tau\bar{\tau}}^{\text{th}/\text{pu}_a}$ since the QCD corrections to the $h \rightarrow VV^*$, $\tau\bar{\tau}$ decay channels arise only at orders higher or equal to $O(\alpha_s^2)$.

The $\Delta\Gamma_Y^{\text{pu}_a}$ errors are associated to Gaussian distributions, and are thus identified without ambiguity with the errors defined in Ref. [20]. The $\Delta\Gamma_Y^{\text{th}}$ errors are purely theoretical, so that one associates them with flat priors. To adopt a conservative prescription, as in Section 6.3, we interpret the numbers given in [17] as 1σ -widths. These numbers are thus directly identified with the $\Delta\Gamma_Y^{\text{th}}$.

Now inserting Eq. (6.27) into Eq. (5.7) provides the contributions of the theoretical and parametric uncertainties to the branching ratios,

$$\begin{aligned} \delta_{Y_i}^B \Delta_{Y_i}^B &= \sum_{Y,a} \delta_Y^{\text{pu}_a} \Delta_Y^{\text{pu}_a} \left(B_Y^{\text{SM}} - \delta_{Y_i Y} \right) + \sum_Y \delta_Y^{\text{th}} \Delta_Y^{\text{th}} \left(B_Y^{\text{SM}} - \delta_{Y_i Y} \right) \\ &\triangleq \sum_{Y,a} \delta_Y^{\text{pu}_a} \Delta_{Y,i}^a + \sum_Y \delta_Y^{\text{th}} \Delta_{Y,i} , \end{aligned} \quad (6.28)$$

where in the last line one introduces a compact notation for the error magnitudes. The sum over Y here must include all the individual Higgs decay channels (not only the ones effectively detected at colliders), namely $Y \equiv b\bar{b}, c\bar{c}, WW, ZZ, \tau\bar{\tau}, \gamma\gamma, g g \dots$

We stress that the parametric error $\delta_Y^{\text{pu}_a} \Delta_Y^{\text{pu}_a}$ on various decay rates Y arises from the same source (namely, varying the fundamental parameter a). The parametric errors on the various decays are thus fully correlated. Therefore, one could in principle drop the Y index on $\delta_Y^{\text{pu}_a}$. There is however a subtlety, because these errors can be either 100% correlated or 100% anti-correlated. The use of parameters δ^{pu_a} would render the full correlation manifest, but minus signs would have to be included in certain $\Delta_Y^{\text{pu}_a}$. Here instead, we chose positive Δ 's by convention. We have thus to keep the Y index on $\delta_Y^{\text{pu}_a}$, bearing in mind that this Y labels only 100% correlation or anti-correlation. A second subtlety is that these signs are actually not clearly given in the literature. Rather, only the absolute values of the $\Delta_Y^{\text{pu}_a}|_0$ are provided. We adopt a conservative choice by assuming that all these errors are 100% correlated.

We can now apply the leading moment approximation on the combination of Eqs. (6.25)–(6.28), where the leading uncertainty is $\delta_{X_i}^P \Delta_i^P$ and the perturbation is $\delta_{Y_i}^B \Delta_{Y_i}^B$, i.e. $\Delta_i^P \gg \Delta_{Y,i}, \Delta_{Y,i}^a$. The 1σ -width of the global theoretical uncertainty in a channel i is given by

$$(\Delta_i^\mu)^2 = (\Delta_i^P)^2 + \sum_a \left[\sum_Y \Delta_{Y,i}^a \right]^2 + \sum_Y (\Delta_{Y,i})^2 , \quad (6.29)$$

with Δ_i^P given by Eq. (6.22). Regarding the prior distribution of the δ_X^μ , the discussion is exactly the same as the one in Section 6.4. That is, following the leading moment approximation, the joint distribution of the δ_X^μ corresponds to the one of the leading uncertainties δ_X^P , so that

$$\pi^\mu \approx \pi^P . \quad (6.30)$$

This implies in particular that the δ_X^μ inherit the correlations from the δ_X^P , that is $\rho_{XX'}^\mu \approx \rho_{XX'}^P$.

Let us discuss the correlations used to derive Eq. (6.29), which are drawn from Refs. [17,18,20]. First, a given parametric uncertainty associated to $\delta_Y^{\text{pu}_a}$ introduces 100% correlated errors among the various decay modes Y , so that the sum over Y of the $\Delta_{Y,i}^a$ is linear. Recall the parametric correlations are taken to be all positive. There is also a slight correlation between $\delta_{X_i}^P \Delta_i^P$

and $\delta_Y^{\text{pu}_{\alpha_s}} \Delta_{Y,i}^{\alpha_s}$, because δ_X^P also contains a contribution from the α_s error. The α_s contribution being subleading in δ_X^P , its correlation with $\delta_Y^{\text{pu}_{\alpha_s}}$ is expected to be small, so that we can neglect it. All the other sources of uncertainties are independent due to their different origins, so that summations in quadrature appear everywhere else in Eq. (6.29).

Using the definitions of the reduced Δ 's in Eq. (6.28), we finally write explicitly the total theoretical uncertainty on the signal strength of a Higgs detection channel i ,

$$(\Delta_i^\mu)^2 = (\Delta_i^P)^2 + \sum_a \left[\sum_Y \Delta_Y^{\text{pu}_a} \left(B_Y^{\text{SM}} - \delta_{Y_i Y} \right) \right]^2 + \sum_Y \left[\Delta_Y^{\text{thu}} \left(B_Y^{\text{SM}} - \delta_{Y_i Y} \right) \right]^2. \quad (6.31)$$

6.6. Summary

In this section we have assembled step by step all the theoretical uncertainties on the Higgs signal strengths, starting from the Higgs likelihood Eq. (6.2). This combination is made possible by the statistical analysis of Section 3, whose results have been extensively used here. The final Higgs likelihood involving the combined errors reads

$$\begin{aligned} L(c_V, c_f) = & \int \left(\prod_X d\delta_X^\mu \right) \pi^\mu(\delta_X^\mu) \\ & \times \exp \left[-\frac{1}{2} \sum_{i,j} \left(\mu_i^{\text{th}}[c_V, c_f] - \mu_i^{\text{ex}}(1 + \delta_{X_i}^\mu \Delta_i^\mu) \right) C_{ij}^{\text{ex}-1} \right. \\ & \left. \times \left(\mu_j^{\text{th}}[c_V, c_f] - \mu_j^{\text{ex}}(1 + \delta_{X_j}^\mu \Delta_j^\mu) \right) \right]. \end{aligned} \quad (6.32)$$

The only label for the combined nuisance parameters $\delta_{X_i}^\mu$ is X_i , the dominant production mode for a given channel i (see for instance Eq. (6.26)). The prior π^μ is approximately equal to the prior of the production mode uncertainties π^σ , through Eq. (6.23) and Eq. (6.30). In Section 7.1, the assumptions on the correlations among the production modes will allow us to express π^σ in terms of the priors of individual production mode uncertainties π_X^σ (see Eqs. (3.16)–(3.19)).

One of the outcome of the combination procedure followed throughout this section is that the shape of the combined priors π_X^σ appears to be almost Gaussian. This comes partly because some of the priors for the individual sources of uncertainty are Gaussian. However, the main reason is actually that a substantial number of the individual sources of uncertainty are independent and of same order of magnitude. These conditions resemble to the ones of the *central limit theorem*, which predicts that the combination would converge towards a Gaussian distribution. Besides, the small errors from contamination and partial decay widths do not affect either the final prior shape under the leading moment approximation. It follows that the π^μ distribution is close to a multivariate Gaussian distribution. An interesting feature of this resulting Gaussian shape, in contrast with the uniform one, is that the distribution has smooth tails instead of excluding any value of the nuisance parameter outside a finite interval (that would have no clear reason).⁴³

Finally, we stress again that the famous question of the linear versus quadratic summation of individual errors (as the ones used in this section to derive Δ_i^μ in Eq. (6.31)) relies uniquely on

⁴³ Another motivation for the normal distribution originates from the principle of maximum entropy [59], that is the principle of not imposing more information than demanded by the constraints on the expected value and the variance.

the correlations among the errors, and is therefore independent of the shapes of the priors. This general feature holds when uncertainties are combined using Bayesian statistics.

7. Marginalising the Higgs likelihood

7.1. Correlations of the detection channels

In this subsection we focus on the correlations among Higgs detection channels induced by the theoretical uncertainties. These correlations appear whenever a source of uncertainty contributes simultaneously to various channels.

As a preliminary observation, let us recall that these correlations are sometimes not taken into account in the literature. What is typically done in such case is that some amount of error, typically from Refs. [17–20], is added independently to the statistical error of each detection channel. Such combination typically reads $(\Delta\mu_i^{\text{ex}})^2 + (\Delta\mu_i^{\text{th}})^2$ if done in quadrature. From the point of view of nuisance parameters, this combination would correspond to associating one independent $\delta_i^\mu \Delta_i^\mu$ to each detection channel, and thus performing one integration per channel in the marginal likelihood.

The issue with such approach is that the correlations among channels induced by the theoretical uncertainties are lost. As stated in Section 2.2.2, these correlations are crucial because they potentially change the tension among the various channel measurements, which in turn can modify the best-fit regions. As slight modifications of the best-fit regions are expected in presence of new physics, treating correctly the theoretical uncertainties is fundamental.

Taking into account the correlations among channels amounts to consistently propagate the theoretical errors into the different detection channels. This is precisely what is done through the combination procedure of Section 6. Combining the errors together and using the leading moment approximation to treat subdominant errors, only five nuisance parameters δ_{ggF}^μ , δ_{VBF}^μ , δ_{ZH}^μ , δ_{WH}^μ and δ_{ttH}^μ arise (see Eqs. (6.31)–(6.32)). The uncertainty on each channel is described by only one of these δ_X^μ , where the X corresponds to the dominant production mode in this channel. That is, all channels dominated by the same production mode X have the same nuisance parameter δ_X^μ . This implies that these channels are 100% correlated.

In principle, the combination procedure of Section 6 describes the complete distribution for the δ_X^μ , π^μ , including the correlations $\rho_{XX'}^\mu$ among the different δ_X^μ . In practice, a complete knowledge of the correlations among the individual sources of uncertainties is needed to obtain $\rho_{XX'}^\mu$. Here we consider the determination of $\rho_{XX'}^\mu$ as beyond the scope of this paper, since for example one would have to work out clearly the correlations among the Higgs production modes induced by the PDF data uncertainties (δ_X^{data}). Using the information available in the literature we will rather consider some characteristic cases for $\rho_{XX'}^\mu$.

Let us first discuss the typical correlations induced by the PDF uncertainties (originating from the PDF data fit) and the scale uncertainties (cf. Section 6.2) on the production cross sections. From now on, the δ_X^μ are denoted as δ_X for simplicity,

$$\delta_X^\mu \hat{=} \delta_X. \quad (7.1)$$

First, we will set $\delta_{\text{ggF}} = -\delta_{\text{ttH}}$ since an anti-correlation between the corresponding PDF errors is reported in Ref. [58].⁴⁴ Note that in reality, this anti-correlation is not total (its value is -0.6 in

⁴⁴ It is not clear from this reference whether the correlations include as well the whole error from α_s which is 100% correlated between the production modes. Nevertheless this source of error is minor compared to the other ones.

Ref. [58]) and furthermore the other source of error, the scale uncertainty, does not correlate the ggF and ttH cross sections as these come from independent QCD calculations.

The correlation coefficients of the PDF errors – between ~ 0.63 and 0.93 [58] – for the three other production modes motivate us to take $\delta_{\text{VBF}} = \delta_{\text{ZH}} = \delta_{\text{WH}}$. This assumption is further justified by the fact that the PDF error is larger than the scale error (particularly for VBF) and that the scale error most probably correlates the ZH and WH modes.

The correlation coefficients of the PDF errors between ggF and WH (-0.23), ZH (-0.14) or VBF (-0.57) suggest to consider the two extreme cases of vanishing correlation and 100% anti-correlation. The scale uncertainties tend to decorrelate these modes. It is thus coherent to consider the cases of vanishing correlation and 100% anti-correlation as the two extreme cases to study. All these assumptions are summarised as the two following configurations on the nuisance parameters,⁴⁵

$$\delta_{\text{ggF}} = -\delta_{\text{ttH}}, \quad \delta_{\text{VBF}} = \delta_{\text{ZH}} = \delta_{\text{WH}}, \quad (7.2)$$

$$-\delta_{\text{ggF}} = \delta_{\text{ttH}} = \delta_{\text{VBF}} = \delta_{\text{ZH}} = \delta_{\text{WH}}, \quad (7.3)$$

keeping in mind that the realistic situation lies in between these extreme cases.

Regarding the PDF set error, the individual uncertainties giving rise to this error are not available in the literature. Rather, only the global PDF set error is estimated by changing various assumptions at a time. One can at least notice that the PDF set errors can be potentially correlated either negatively or positively, respectively, for the ggF and VBF reactions or the VBF and VH processes, as observed from the relative signs of rate variations in Fig. 57 of Ref. [20] when changing the PDF set.⁴⁶ These correlations are roughly consistent with the ones in Eq. (7.3).

Let us describe how the correlation configurations of Eqs. (7.2)–(7.3) are related to the π^μ appearing in the marginal likelihood (6.32). The prior π^μ is approximately equal to the prior of the production mode uncertainties π^σ (Eq. (6.23) and Eq. (6.30)) which can itself be expressed (according to (3.18)–(3.19)) in terms of the π_X^σ under the assumptions (7.2)–(7.3). One ends up with the two final priors, associated respectively to the correlation configurations of Eqs. (7.2)–(7.3),

$$\pi^\mu(\delta_X) = \pi_{\text{ggF}}^\sigma(\delta_{\text{ggF}}) \delta(\delta_{\text{ggF}} + \delta_{\text{ttH}}) \pi_{\text{VBF}}^\sigma(\delta_{\text{VBF}}) \delta(\delta_{\text{VBF}} - \delta_{\text{ZH}}) \delta(\delta_{\text{VBF}} - \delta_{\text{WH}}), \quad (7.4)$$

$$\pi^\mu(\delta_X) = \pi_{\text{ggF}}^\sigma(\delta_{\text{ggF}}) \delta(\delta_{\text{ggF}} + \delta_{\text{ttH}}) \delta(\delta_{\text{ggF}} + \delta_{\text{VBF}}) \delta(\delta_{\text{ggF}} + \delta_{\text{ZH}}) \delta(\delta_{\text{ggF}} + \delta_{\text{WH}}), \quad (7.5)$$

where $\delta()$ denotes the Dirac distribution.

7.2. The Bayesian analytical likelihood

The π_X^σ priors deduced from the combination of all the cross section errors, in Section 6.3, have been found to be nearly Gaussian distributions. These Gaussian shapes are obtained by choosing flat shapes for all the unknown priors for theoretical uncertainties. As mentioned in Section 6.6, one expects this result to hold approximatively for other choices of initial priors. Nevertheless, in order to take into account in our numerical results the possibility of non-flat initial shapes, we also consider a totally different form of the final prior: we take it as a flat distribution. The choice of these two shapes (Gaussian and flat) provides an estimate of the impact of the prior shape on the final results. The distributions π_X^σ appearing in Eqs. (7.4)–(7.5) are hence defined as

⁴⁵ For consistency, these two configurations are used as well to determine the $\rho_{XX'}^\sigma$, correlation matrix of Eq. (6.22).

⁴⁶ Recall that the Fig. 57 of Ref. [20] is for the 8 TeV LHC.

$$\pi_X^\sigma(\delta_X^\sigma) = \frac{1}{\sqrt{2\pi}} e^{-(\delta_X^\sigma)^2/2}, \quad (7.6)$$

$$\pi_X^\sigma(\delta_X^\sigma) = \begin{cases} 1/2\sqrt{3} & \text{if } \delta_X^\sigma \in [-\sqrt{3}, \sqrt{3}], \\ 0 & \text{otherwise} \end{cases} \quad (7.7)$$

for the Gaussian and flat cases respectively. Recall that the variance of all the δ 's, including δ_X^σ , are chosen to be equal to one for any prior shape. This appears clearly in Eq. (7.6) and implies the $[-\sqrt{3}, \sqrt{3}]$ interval in Eq. (7.7).

For analytical integrations of the final likelihood (6.32), it is convenient to denote by \mathcal{X} a subset of fully correlated production modes, $\{X, X', \dots\}$. We then denote by $\Omega_{\mathcal{X}}$ the subset of channels (labelled by i) dominated by the production modes contained in \mathcal{X} . In presence of anti-correlations, one further divides $\Omega_{\mathcal{X}}$ into two anti-correlated subsets $\Omega_{\mathcal{X}}^+$, $\Omega_{\mathcal{X}}^-$. Finally, the set of all channels is written Ω . Assuming the correlations among production modes follow Eq. (7.2), the set of detection channels is splitted into $\Omega_{\{\text{ggF}, \text{ttH}\}}$ and $\Omega_{\{\text{VBF}, \text{WH}, \text{ZH}\}}$. $\Omega_{\{\text{ggF}, \text{ttH}\}}$ is then splitted into the anti-correlated subsets $\Omega_{\{\text{ggF}, \text{ttH}\}}^+ = \Omega_{\text{ggF}}$, $\Omega_{\{\text{ggF}, \text{ttH}\}}^- = \Omega_{\text{ttH}}$. Assuming the correlations of Eq. (7.3), there is instead a unique set $\Omega = \Omega_{\{\text{ggF}, \text{ttH}, \text{VBF}, \text{WH}, \text{ZH}\}}$. It is splitted into the anti-correlated subsets $\Omega_{\{\text{ggF}, \text{ttH}, \text{VBF}, \text{WH}, \text{ZH}\}}^+ = \Omega_{\text{ggF}}$, $\Omega_{\{\text{ggF}, \text{ttH}, \text{VBF}, \text{WH}, \text{ZH}\}}^- = \Omega_{\{\text{ttH}, \text{VBF}, \text{WH}, \text{ZH}\}}$.

At that point it is also convenient to introduce the following quantities $\zeta_{\mathcal{X}}$ and $\eta_{\mathcal{X}\mathcal{X}'}$ defined as

$$\begin{aligned} \zeta_{\mathcal{X}} &= \sum_{i \in \Omega_{\mathcal{X}}, j \in \Omega} \kappa_i \Delta_i^\mu (\mu_i^{\text{th}} - \mu_i^{\text{ex}}) C_{ij}^{\text{ex}-1} \mu_j^{\text{ex}}, \quad \kappa_i = \begin{cases} 1 & \text{if } i \in \Omega_{\mathcal{X}}^+ \\ -1 & \text{if } i \in \Omega_{\mathcal{X}}^- \end{cases} \\ \eta_{\mathcal{X}\mathcal{X}'} &= \sum_{i \in \Omega_{\mathcal{X}}, j \in \Omega_{\mathcal{X}'}} \kappa_i \Delta_i^\mu \mu_i^{\text{ex}} C_{ij}^{\text{ex}-1} \kappa_j \Delta_j^\mu \mu_j^{\text{ex}}. \end{aligned} \quad (7.8)$$

The overall sign of $\zeta_{\mathcal{X}}$ is irrelevant. Note also that if $\mathcal{X} \neq \mathcal{X}'$ (as may occur in the $\eta_{\mathcal{X}\mathcal{X}'}$ function), there are no theoretical correlations at all between the channels belonging to $\Omega_{\mathcal{X}}$ and $\Omega_{\mathcal{X}'}$.

In the case of a Gaussian prior (Eq. (7.6)), it is noticeable that the most general likelihood (6.32) can be integrated analytically and results in the simple analytical expression⁴⁷

$$L_{\text{B}}^{\text{Gauss}} = L_{\mu} \exp \left[\frac{1}{2} \sum_{\mathcal{X}\mathcal{X}'} \zeta_{\mathcal{X}} (\delta_{\mathcal{X}\mathcal{X}'} + \eta_{\mathcal{X}\mathcal{X}'})^{-1} \zeta_{\mathcal{X}'} \right]. \quad (7.9)$$

Here $\delta_{\mathcal{X}\mathcal{X}'}$ is the Kronecker symbol. L_{μ} is the base likelihood defined in Eq. (5.1), *i.e.* the likelihood before introducing nuisance parameters. One observes that the marginal likelihood takes the form of a product of the base likelihood with a term generated by the theoretical uncertainties. This term, which depends on c_V, c_f through $\zeta_{\mathcal{X}}$, as well as on all theoretical and experimental uncertainties, implements all the deformations and correlations induced by the theoretical uncertainties.

For the case of no experimental correlations between different group of channels of dominant production modes, including the case considered without experimental correlations at all (see Section 5.1), one has $\eta_{\mathcal{X}\mathcal{X}'} = 0$ for $\mathcal{X} \neq \mathcal{X}'$ and

$$\eta_{\mathcal{X}\mathcal{X}} \equiv \eta_{\mathcal{X}} = \sum_{i, j \in \Omega_{\mathcal{X}}} \kappa_i \Delta_i^\mu \mu_i^{\text{ex}} C_{ij}^{\text{ex}-1} \kappa_j \Delta_j^\mu \mu_j^{\text{ex}}. \quad (7.10)$$

⁴⁷ A similar expression can also be obtained for an arbitrary correlation matrix $\rho_{\mathcal{X}\mathcal{X}'}^\mu$. Note one dropped an overall factor, as the likelihood is defined up to a normalisation constant.

The marginal likelihood (7.9) then reduces to,

$$L_B^{\text{Gauss}} = L_\mu \prod_{\mathcal{X}} e^{\zeta_{\mathcal{X}}^2 / 2(\eta_{\mathcal{X}} + 1)}. \quad (7.11)$$

Note that this product is over different \mathcal{X} subsets *i.e.* there are no theoretical correlations among the channels belonging to the different $\Omega_{\mathcal{X}}$ groups.

Note that if one assumes a single independent nuisance parameter per channel, there is no sum in Eq. (7.8), meaning that no correlation among channels is induced.⁴⁸ One can directly verify that in the purely de-correlated case (neither experimental nor theoretical correlations), Eq. (7.11) gives back the primary likelihood (5.1) with a summation in quadrature between the absolute experimental and theoretical errors, $\Delta\mu_i^{\text{ex}}$ and $\mu_i^{\text{ex}}\Delta_i^\mu$.

In the case of the flat prior of Eq. (7.7), there is no simple general form such as Eq. (7.9). However, assuming no experimental correlations among various $\Omega_{\mathcal{X}}$ subsets, the marginal likelihood takes a simple form,

$$L_B^{\text{flat}} = L_\mu \prod_{\mathcal{X}} e^{\zeta_{\mathcal{X}}^2 / 2\eta_{\mathcal{X}}} \left[\text{Erf} \left(\frac{\sqrt{3}\sqrt{\eta_{\mathcal{X}}}}{\sqrt{2}} + \frac{\zeta_{\mathcal{X}}}{\sqrt{2\eta_{\mathcal{X}}}} \right) - \text{Erf} \left(\frac{\sqrt{3}\sqrt{\eta_{\mathcal{X}}}}{\sqrt{2}} - \frac{\zeta_{\mathcal{X}}}{\sqrt{2\eta_{\mathcal{X}}}} \right) \right], \quad (7.12)$$

where Erf is the standard error function.

7.3. The frequentist treatment

7.3.1. The marginal likelihood

In classical frequentist statistics, hypotheses are not associated with probabilities, so that there is no such thing as a prior distribution for a nuisance parameter. In the hybrid frequentist framework however, one can associate a parameter with a “prior” distribution that can be seen as an extra likelihood constraining the nuisance parameter. Pushing forward the analogy with the Bayesian case, we worked out the way to combine uncertainties within frequentist statistics in Section 3.3. One may find however that the Bayesian combination of uncertainties are better defined than the frequentist one.

More pragmatically, frequentist combinations are also more complicated, as the combination of the magnitude of the errors (the Δ ’s) depends on the shape of the frequentist “priors”, contrary to the Bayesian case. These drawbacks can constitute motivations to rather follow the Bayesian approach developed in previous sections. Nevertheless, for completeness we describe here the final part of the frequentist method for the Higgs fit. For that purpose we consider in the following, a generic prior, $\pi^\mu(\delta_X^\mu)$, of width Δ_i^μ , obtained after a first phase of frequentist combination.

Recall that the frequentist marginalisation procedure, also called profiling, consists in maximising over δ_X^μ , instead of integrating as done in Eq. (6.32). Hence the frequentist marginal Higgs likelihood reads

$$L(c_V, c_f) = \max_{\delta_X^\mu} \left[\pi^\mu(\delta_X^\mu) \exp \left[-\frac{1}{2} \sum_{i,j} (\mu_i^{\text{th}}[c_V, c_f] - \mu_i^{\text{ex}}(1 + \delta_{X_i}^\mu \Delta_i^\mu)) C_{ij}^{\text{ex} - 1} \right. \right. \\ \left. \left. \times (\mu_j^{\text{th}}[c_V, c_f] - \mu_j^{\text{ex}}(1 + \delta_{X_j}^\mu \Delta_j^\mu)) \right] \right]. \quad (7.13)$$

⁴⁸ We recall that such a combination should be avoided as it is not realistic.

As often done in practice for the frequentist treatment, one can equivalently minimise the χ^2 distribution, $\chi^2 = -2 \log L$, instead of the maximisation in Eq. (7.13),

$$\chi^2(c_V, c_f) = \min_{\delta_X^\mu} \left[-2 \log \pi^\mu(\delta_X^\mu) + \sum_{i,j} \left(\mu_i^{\text{th}}[c_V, c_f] - \mu_i^{\text{ex}}(1 + \delta_{X_i}^\mu \Delta_i^\mu) \right) \mathcal{C}_{ij}^{\text{ex}-1} \right. \\ \left. \times \left(\mu_j^{\text{th}}[c_V, c_f] - \mu_j^{\text{ex}}(1 + \delta_{X_j}^\mu \Delta_j^\mu) \right) \right]. \quad (7.14)$$

The best-fit point given by the χ^2 minimum in the (c_f, c_V) parameter space is noted (\hat{c}_f, \hat{c}_V) and the best-fit regions are obtained by drawing contour levels of the difference (cf. Section 2.1)

$$\Delta\chi^2(c_f, c_V) = \chi^2(c_f, c_V) - \chi^2(\hat{c}_f, \hat{c}_V) \quad (7.15)$$

at the values given by Eq. (2.8).

7.3.2. The frequentist analytical likelihood

Assuming that the Bayesian and frequentist combinations of the errors lead to analogous shapes for the final priors, we consider both a Gaussian and a flat shape for each π_X^σ prior, as in Eqs. (7.6)–(7.7). In the Gaussian case, the marginal likelihood (7.13) can be computed analytically,

$$L_F^{\text{Gauss}} = L_\mu \exp \left[\frac{1}{2} \sum_{\mathcal{X}\mathcal{X}'} \zeta_{\mathcal{X}} (\delta_{\mathcal{X}\mathcal{X}'} + \eta_{\mathcal{X}\mathcal{X}'})^{-1} \zeta_{\mathcal{X}'} \right], \quad (7.16)$$

where the $\zeta_{\mathcal{X}}$, $\eta_{\mathcal{X}\mathcal{X}'}$ are defined as in Section 7.2. This is precisely the same result as for the Bayesian likelihood of Eq. (7.9), L_B^{Gauss} .

For the case of no experimental correlations between the $\Omega_{\mathcal{X}}$'s, the marginal likelihood with Gaussian prior thus simplifies just like in Eq. (7.11).⁴⁹ In this case, the marginal likelihood with a flat prior also gets an analytical expression,

$$L_F^{\text{flat}} = \prod_{\mathcal{X}} \exp \left[-\frac{1}{2} \sum_{i,j} \left(\mu_i^{\text{th}}[c_V, c_f] - \mu_i^{\text{ex}}(1 + \xi_{\mathcal{X}} \kappa_i \Delta_i^\mu) \right) \mathcal{C}_{ij}^{\text{ex}-1} \right. \\ \left. \times \left(\mu_j^{\text{th}}[c_V, c_f] - \mu_j^{\text{ex}}(1 + \xi_{\mathcal{X}} \kappa_j \Delta_j^\mu) \right) \right] \quad (7.17)$$

with

$$\xi_{\mathcal{X}} = \begin{cases} \zeta_{\mathcal{X}}/\eta_{\mathcal{X}} & \text{if } \zeta_{\mathcal{X}}/\eta_{\mathcal{X}} \in [-\sqrt{3}, \sqrt{3}] \\ \sqrt{3} & \text{if } \zeta_{\mathcal{X}}/\eta_{\mathcal{X}} > \sqrt{3} \\ -\sqrt{3} & \text{if } \zeta_{\mathcal{X}}/\eta_{\mathcal{X}} < -\sqrt{3} \end{cases}, \quad (7.18)$$

where $\zeta_{\mathcal{X}}$, $\eta_{\mathcal{X}}$ are defined as in Eqs. (7.8), (7.10).

⁴⁹ Hence the same likelihood (with a sum in quadrature) as in the Bayesian framework arises, in the case of neither experimental nor theoretical correlations.

7.4. Numerical results

The frequentist marginalisation (likelihood (7.16) for the Gaussian prior or (7.17) for the flat one) is not illustrated here because the frequentist framework may seem slightly less consistent than the Bayesian one and the error combinations are more delicate. For these reasons, we rather recommend to use the Bayesian marginalisation technics for the Higgs fits. In any case, the Bayesian and frequentist approaches are expected to converge as the experimental uncertainties become small relatively to the theoretical ones. This situation will gradually occur in the next LHC Runs due to the decrease of the statistical uncertainties and the expected improvement in the knowledge of the experimental systematic errors. We have described this feature in Ref. [60].

Now as a general remark allowing a better comprehension of the following subsections, let us try to explain in simple words the reason why the presence of nuisance parameters can indeed modify the size and the location of the best-fit domains in c_V – c_f .

For the sake of understanding the impact on the size, it is easier to focus on frequentist marginalisation. Frequentist marginalisation can be seen as an approximation of Bayesian marginalisation, so that the same explanation holds for both. The frequentist marginalisation consists of a maximisation of the nuisance parameter (say δ_X^μ) at any point in the space of the parameters of interest. This means that the value of δ_X^μ at a given point is chosen in order to maximise goodness-of-fit. Now, this improvement of goodness-of-fit is typically larger for the points far away from the best-fit point than for those close by the best-fit point. When this fact is true (which is usually the case), the operation of marginalising tends to enlarge the best-fit regions.

The effect of the nuisance parameters on the location of the best-fit regions in c_V – c_f can be understood as follows. Recall that the nuisance parameters enter in the likelihood as $\mu_i^{\text{ex}}(1 + \delta_{X_i}^\mu \Delta_i^\mu)$ (see Eq. (6.32)), so that they shift the central experimental value of the signal strength. This in turn can induce a change in the location of the best-fit point in c_V – c_f . Such a shift actually occurs if a non-zero value of δ_X^μ is preferred. This happens when a non-zero value for δ_X^μ helps relaxing the tensions (*i.e.* different preferred values of c_V , c_f) among various signal strengths μ_i^{ex} . Notice that this means that the likelihood itself favours a non-zero value for δ_X^μ , even though the prior of δ_X^μ is centered on zero.

7.4.1. The forbidden case: no correlations

Following our overview approach, let us start with the simplest case: the Bayesian marginalisation in the absence of correlations between the theoretical errors of the different Higgs channels. Let us take for instance a Gaussian prior (taking a flat one would not change our conclusions). This case was described in more details in the beginning of Section 7.1 as well as in Section 7.2. In this “de-correlated” case, the likelihood is simply the primary likelihood (5.1) with a summation in quadrature of the absolute experimental and theoretical errors, $(\Delta\mu_i^{\text{ex}})^2 + (\mu_i^{\text{ex}}\Delta_i^\mu)^2$. The best-fit domains in the c_V – c_f plane are derived following the standard procedure described in Section 2, and are shown in Fig. 4. Here and throughout Section 7.4, the priors for c_V , c_f are taken flat, $\pi(c_{V,f}) \propto 1$.

We see on this figure that the theoretical SM prediction ($c_V = c_f = 1$) lies well within the 68% C.L.⁵⁰ region. Physically, this implies that, with such a fit, no physics beyond the SM is required to interpret the 8 TeV LHC measurements of the Higgs rates. The increase of the best-fit domain sizes induced by the existence of theoretical errors is relatively weak, due to the

⁵⁰ The acronym C.L. will stand for Credible Level within the Bayesian framework and for Confidence Level in the frequentist framework.

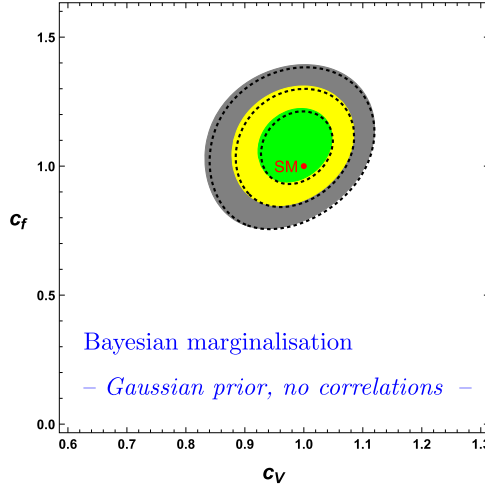


Fig. 4. The best-fit regions in the c_V – c_f plane obtained from Bayesian marginalisation and Gaussian priors for the theoretical uncertainties. The 68%, 95% and 99% credible regions are represented respectively by the green, yellow and grey domains. No theoretical error correlations between the Higgs detection channels are taken into account in this figure. The dashed contours illustrate the case without theoretical uncertainties. The SM prediction is shown by the red point. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

sum in quadrature, as observed when comparing to the best-fit regions obtained with vanishing theoretical errors. The latter regions are superimposed in Fig. 4 for illustration purpose (as the dashed contours) and to ease the comparison with next plots.

However let us recall that the likelihood used here (and leading to the coloured regions of Fig. 4) is not realistic as the correlations among the Higgs channels should not be neglected. We thus do not recommend the use of this likelihood.

7.4.2. Flat prior

From now on we consider the more realistic likelihoods obtained in Section 7.2. These likelihoods contain all the correlations between Higgs channels induced by the theoretical uncertainties. First, we consider the configuration with two independent nuisance parameters (see Eq. (7.2) and Eq. (7.4)). The Bayesian marginalisation over these two nuisance parameters leads to the analytical likelihood (7.12) for flat final priors. Applying the standard Bayesian procedure, described in Section 2, we find the best-fit regions of Fig. 5[left].

By comparing the coloured plots in Fig. 4 and Fig. 5[left], one observes clearly a shift of the best-fit regions. This shift originates from the theoretical correlations that are taken into account in Fig. 5[left]. This shift occurs because the relaxation of the tensions between the individual signal strength measurements (see discussion in the introduction of Section 7.4) is different in the correlated case and in the “de-correlated” one. We emphasise that this shift is a consequence of taking into account the theoretical correlations. Indeed we will see in next subsection that the same effect occurs for a different prior shape. Concerning the region size, a slight increase occurs relatively to Fig. 4. This comparison can be done by looking at the reference case (dashed contours) without theoretical errors at all, which is once more superimposed in Fig. 5[left].

The plot on the right hand side of Fig. 5 is the same as the left plot but for the second correlation configuration, involving a single nuisance parameter (discussed in Eq. (7.3) and Eq. (7.5)).

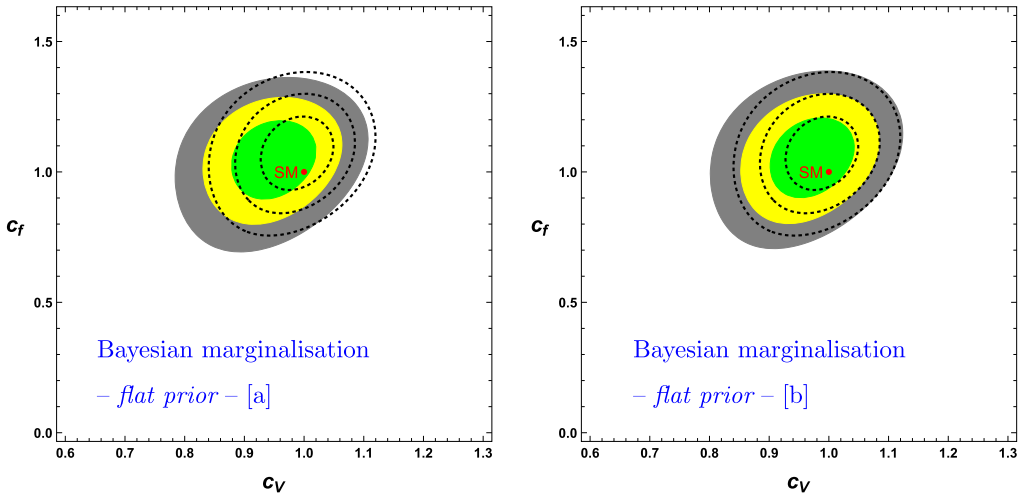


Fig. 5. The best-fit regions in the c_V – c_f plane obtained from Bayesian marginalisation and flat priors for the theoretical uncertainties. The 68%, 95% and 99% credible regions are represented respectively by the green, yellow and grey domains. The [a] and [b] plots correspond, respectively, to the two characteristic correlation configurations described in Eq. (7.4) and Eq. (7.5). The dashed contours illustrate the case without theoretical uncertainties. The SM prediction is shown by the red point. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

The effect of the theoretical correlations (relatively to Fig. 4) appears to be softer than for the left plot: the shift is smaller. This difference between the two coloured regions of Fig. 5 makes clear that the theoretical correlations have an important impact on the fits, and should thus be carefully taken into account.

As described below Eq. (7.3), the most realistic correlation configuration is most probably an intermediate configuration between those adopted in the two plots of Fig. 5. We thus conclude that, with the statistical treatment adopted here, the SM prediction remains in a good agreement (1σ level) with the 8 TeV LHC Higgs data, even once realistic theoretical correlations are taken into account.

7.4.3. Gaussian prior

Fig. 6 illustrates the same case as in Fig. 5 except that the final priors are now Gaussian,⁵¹ which leads to the marginalised Bayesian likelihood of Eq. (7.9) and Eq. (7.11). It appears that there is no substantial difference (neither in location, size nor shape of the best-fit regions) between these two figures. This illustrates the mild impact of the choice of the shape for the prior of the theoretical uncertainties. We conclude that, with the present statistical uncertainties on Higgs data, the recurring question of the exact shape of the prior,⁵² in particular for the errors due to truncated perturbative expansions in QCD, is nearly irrelevant.

However we should stress that this insensitivity to the prior shape occurs because the experimental uncertainties of the current data are typically larger or of the same order as the theoretical

⁵¹ At this stage, we recall that the Gaussian priors are obtained from a combination of all the individual priors, while the flat priors have just been chosen ‘by hand’ to illustrate what happens for completely different distributions.

⁵² Including the details of the form at the boundaries in case e.g. of a flat distribution.

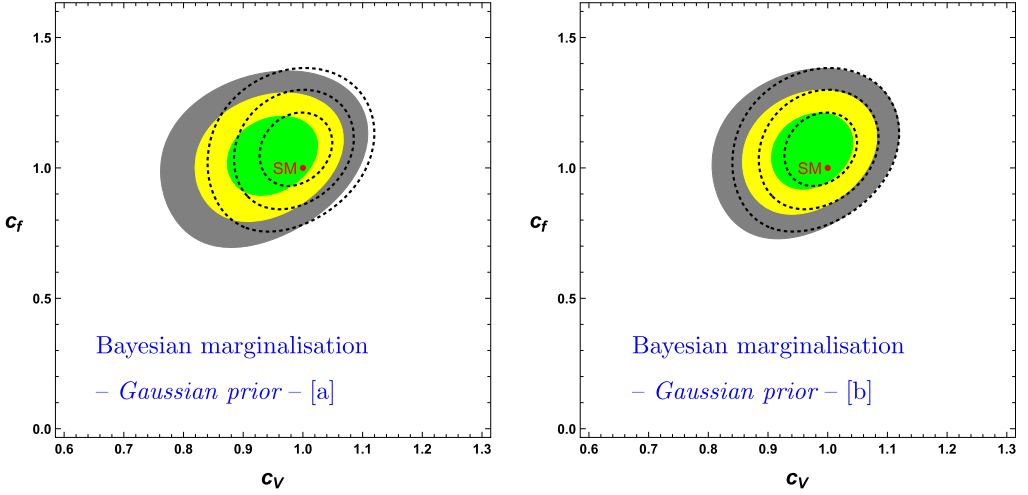


Fig. 6. The best-fit regions in the c_V – c_f plane obtained from Bayesian marginalisation and Gaussian priors for the theoretical uncertainties. The 68%, 95% and 99% credible regions are represented respectively by the green, yellow and grey domains. The [a] and [b] plots correspond, respectively, to the two characteristic correlation configurations described in Eq. (7.4) and Eq. (7.5). The dashed contours illustrate the case without theoretical uncertainties. The SM prediction is shown by the red point (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.).

ones. This situation is expected to change with the upcoming LHC runs, as the statistical uncertainties will decrease with the integrated luminosity.

It might also happen in the future that the data, possibly including new physics effects, induce a significant shift of a parameter of interest with respect to its SM expectation or favour a nuisance parameter value far from its 1σ interval. The latter case would be strictly forbidden by a uniform prior (in contrast with a Gaussian prior), without a deep reason. Hence an “improved” uniform prior with smooth tails could be preferred.

7.4.4. The nuisance parameters favoured by the data

Let us now consider the posterior distribution for the theoretical uncertainties themselves, instead of the posterior for the parameters of interest. Here we shall take the priors associated with the theoretical uncertainties (π_X^σ) as flat and with an *infinite* range. For such choice of prior, the information of the posterior is fully contained in the likelihood (second line in Eq. (7.19)). The interest of this data-dominated posterior is that it allows us to study exclusively the information that the sole Higgs data provide about the theoretical uncertainties, Δ_i^μ .

We first consider the case with a single nuisance parameter δ_{ggF} (*i.e.* the fully correlated case), given in Eq. (7.3), and we present in Fig. 7 the data-dominated posterior for δ_{ggF} ,

$$\begin{aligned}
 p(\delta_{\text{ggF}}|\mu_i^{\text{ex}}) &= \int dc_V dc_f \pi(c_V, c_f) \pi_{\text{ggF}}^\sigma(\delta_{\text{ggF}}) \\
 &\times \exp\left[-\frac{1}{2} \sum_{i,j} \left(\mu_i^{\text{th}}[c_V, c_f] - \mu_i^{\text{ex}}(1 \pm \delta_{\text{ggF}} \Delta_i^\mu) \right) C_{ij}^{\text{ex}-1} \right. \\
 &\times \left. \left(\mu_j^{\text{th}}[c_V, c_f] - \mu_j^{\text{ex}}(1 \pm \delta_{\text{ggF}} \Delta_j^\mu) \right) \right].
 \end{aligned} \tag{7.19}$$

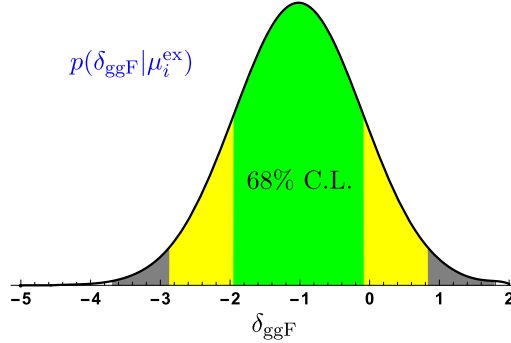


Fig. 7. The data-dominated posterior $p(\delta_{\text{ggF}}|\mu_i^{\text{ex}})$ (Eq. (7.19)). The 68%, 95% and 99% credible domains are indicated respectively by the green, yellow and grey areas. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

This posterior is obtained by integrating the likelihood of Eq. (6.32) (with π^μ given by Eq. (7.4)) over all δ 's but one, chosen to be δ_{ggF} , and marginalising with respect to the c_V, c_f parameters for the class of physical models with $\pi(c_V, c_f) \propto 1$ (considered for simplicity).

It appears in Fig. 7 that the posterior for δ_{ggF} is centred on $\delta_{\text{ggF}} \simeq -1$.⁵³ This means that for each signal strength, the data typically favour a value falling at $\pm 1\sigma$ (*i.e.* at $\pm \Delta_i^\mu$) from the nominal value μ_i^{ex} . In other words, for the correlation configuration of Eq. (7.3) the Higgs data provide a non-trivial indication that the magnitudes of the theoretical errors are reasonably well estimated. Indeed, the theoretical estimations predict the μ_i^{ex} to lie typically within the 1σ interval $\pm \Delta_i^\mu$.

This compatibility suggests that the Δ_i^μ uncertainties, whose estimations rely on quite ad hoc QCD scale variations and on the arbitrariness in the choice of PDF sets, are nevertheless quite robust. On the other hand, one also notices in Fig. 7 that the credible intervals for $p(\delta_{\text{ggF}}|\mu_i^{\text{ex}})$ go beyond -1 . This could be taken as an argument for slightly increasing the overall magnitude of the theoretical uncertainties (see next subsection).

The correlation configuration with two nuisance parameters, given by Eq. (7.2), leads to larger preferred values for the nuisance parameters $\delta_{\text{ggF}} \simeq -2$, $\delta_{\text{VBF}} \simeq -5$. We interpret these very large values as the fact that neglecting totally the correlation between the two nuisance parameters is an unrealistic hypothesis (as already described in Section 7.1). As a matter of fact, if one restored the usual prior for the δ 's (*i.e.* a prior with unit variance, $V[\delta] = 1$), a hypothesis testing would show that the data favour the correlation configuration of Eq. (7.3) with respect to the configuration of Eq. (7.2).

7.4.5. More conservative theoretical errors

Throughout this paper, we have been observing that, among the various origins of theoretical uncertainty involved in the Higgs fit, some are of a nature (see Sections 6.1–6.5) which renders difficult the *exact determination* of the associated 1σ interval. These are the truncation of the perturbative expansion for the QCD calculation of Higgs rates translated into an arbitrary error range for the renormalisation/factorisation scale $\mu = \mu_R = \mu_F$ (affecting the production and decay amplitudes as well as the α_s coupling constant), the choices made (on the statistical method,

⁵³ For comparison, the maximum of $p(\delta_{\text{ggF}}, c_V = c_f = 1|\mu_i^{\text{ex}})$ is reached for $\delta_{\text{ggF}} \simeq -0.7$.

the number of free parameters. . .) in the different PDF sets, and finally the m_b renormalisation scheme and EFT assumptions for the ggF mechanism. These considerations can be taken as a motivation to adopt more conservative theoretical errors.

Moreover, we have seen in the previous subsection (see Fig. 7) that the data tend to prefer theoretical uncertainties that are somewhat larger than the combined 1σ width Δ_i^μ obtained in Section 6, see e.g. the 68% C.L. interval in Fig. 7. Taking seriously this fact, it makes sense to perform the fits with a slight overall increase of the uncertainties. We suggest a rescaling

$$\Delta_i^\mu \rightarrow 1.5 \Delta_i^\mu \quad (7.20)$$

as a reasonable estimation for a most conservative choice of theoretical uncertainties. Notice that the rescaling of Eq. (7.20) is equivalent (cf. Eq. (6.32)) to rescale by 1.5 the axis in Fig. 7. For example, the point $\delta_{\text{ggF}} = -1$ becomes $\delta_{\text{ggF}} = -1.5$.

The best-fit regions with $\Delta_i^\mu \times 1.5$ are shown in Fig. 8 for the two correlation configurations and considering the flat prior case (Eq. (7.12)), keeping in mind that with the current Higgs data, the final prior shape does not affect significantly those best-fit domains. The impact of the increase of the theoretical uncertainties (Eq. (7.20)) on the fit of the current Higgs data can be seen by comparing Fig. 5 and Fig. 8. It turns out that the shift of the preferred regions with respect to the case without theoretical errors gets slightly accentuated. In the correlation configuration of Eq. (7.4), i.e. with two independent δ_X , it even appears (see Fig. 8[left]) that the SM point moves just outside the 68% C.L. region.

The increase of this shift can be understood by recalling that rescaling the Δ_i^μ is equivalent to increase the width of the δ_X prior. It is then clear that more possibilities are opened for the preferred values of δ_X . It turns out that these preferred values move further away from zero, which induces a more pronounced shift of the best-fit regions.

Even though these effects are not statistically significant for the current Higgs data, we stress that the impact of the theoretical errors will increase while more data will be accumulated at the LHC. The ambiguity existing in the theoretical errors estimation deserves thus to be taken into account. For future LHC phenomenological studies, we suggest to take into account, in the same way as proposed in this subsection, the impact on the fits from the lack of knowledge in theoretical errors.

8. Biasing the Higgs likelihood

The principle of bias has been presented in Section 2.2.2. To have a self-consistent section, we recall here the basics of a “biasing” procedure. We distinguish two realisations of the bias principle: the *extremal bias* and the *envelope method*.

The method of extremal biasing consists in drawing the best-fit regions for the parameters of interest for extreme fixed values of the theoretical errors. By the word ‘extreme’, we mean that we set the nuisance parameters δ at ± 1 (corresponding to one-standard deviations with our conventions) in order to obtain a strong impact on the fit. In our Higgs fit, the theoretical uncertainties affect the signal strengths μ_i^{ex} , which in turn modify the preferred value of $\mu_i^{\text{th}}(c_f, c_V)$ and thus the best-fit regions of c_V, c_f . Note that the choice of extreme values $\delta \pm 1$ can be seen as natural, and for that reason will be used in our numerical results, but strictly speaking remains only a choice with a certain degree of arbitrariness.

The envelope method corresponds formally to the continuous version of this extremal biasing. Loosely speaking, this is what one obtains if one does the fit for each fixed value of the

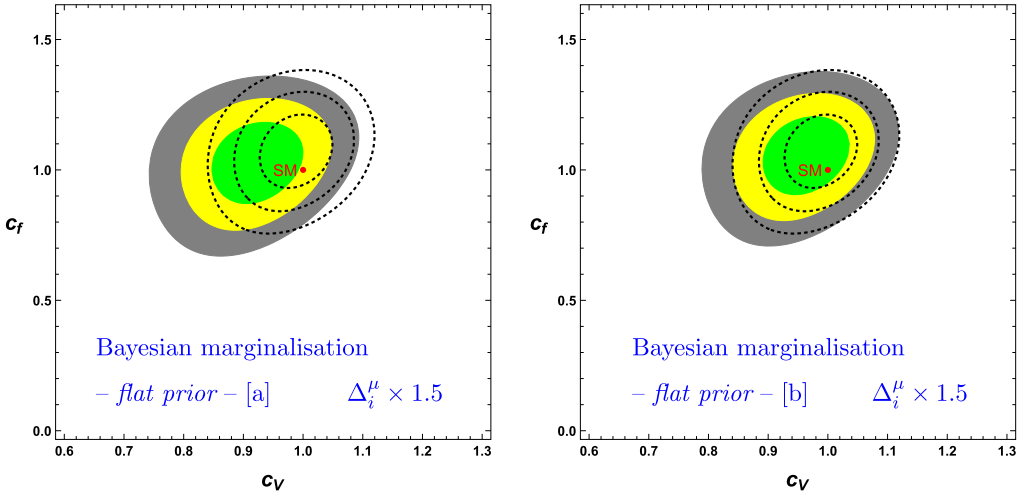


Fig. 8. The best-fit regions in the c_V – c_f plane obtained from Bayesian marginalisation and flat priors for the theoretical uncertainties. The 68%, 95% and 99% credible regions are represented respectively by the green, yellow and grey domains. The [a] and [b] plots correspond, respectively, to the characteristic correlation configurations described in Eq. (7.4) and Eq. (7.5). The dashed contours illustrate the case without theoretical uncertainties. The SM prediction is shown by the red point. The difference with Fig. 5 is the enhancement of the uncertainties, accordingly to $\Delta_i^\mu \rightarrow \Delta_i^\mu \times 1.5$. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

nuisance parameters between the extreme values $\delta = \pm 1$. One expects typically a deformed contour somehow interpolating between the regions of extremal biasing. For a more formal and unified description of these biasing methods, see Section 2.2.2.

What are the motivations for choosing the marginalisation or the bias approaches (extremal bias or envelope method) in the Higgs fits? The lack of knowledge on the shape of the prior⁵⁴ associated to the main QCD uncertainties discussed in Section 6.3 encourages one to apply a bias method, which does not rely on the prior shape – in contrast with the marginalisation.

Besides, the bias is more conservative. Indeed, while in the marginalisation the best-fit domain corresponds roughly to nuisance parameters centered around a preferred δ_X value, in the bias methods δ_X rather spans by construction its $[-1, 1]$ interval without favouring any value. Hence, generally speaking (and this is the case for the Higgs fit), the best-fit regions in the space of the parameters of interest obtained through the bias methods are wider than the ones from marginalising.

In addition, the envelope method allows one to see at a glance the whole best-fit domain in the c_V – c_f plane spanned by varying the nuisance parameters inside their entire $[-1, 1]$ intervals. The price to pay here is maybe a heavier technical approach than in the marginalisation procedure: compare the marginalisation definitions in Eqs. (2.9), (2.10) with the biasing definitions in Eqs. (2.14), (2.17) (see for example Eq. (6.32) and Eq. (8.5) for the application to the Higgs likelihood). It is clear that more operations (either integrations or maximisations) are needed for the envelope method.

⁵⁴ Once again, from the physical point of view only (not in the statistical sense).

8.1. Combining the uncertainties

The starting point is the likelihood (5.1), and then (5.11). Applying the Eqs. (3.21)–(3.24) together with the definition of Eq. (6.28) and,

$$\Delta_{X,i}^P \hat{=} \frac{\epsilon_X^i \sigma_X^{\text{SM}}}{\sum_{X'} \epsilon_{X'}^i \sigma_{X'}^{\text{SM}}} \left(\Delta_X^{\text{amp}} + \Delta_X^{\text{PDF}+\alpha_s} \right), \quad (8.1)$$

which is a new compact notation comparable to Eq. (6.19), we obtain the likelihood depending on a unique nuisance parameter, δ_b ,

$$L_{\text{bias}}(\delta_b) = \exp \left[-\frac{1}{2} \sum_{i,j} \left(\mu_i^{\text{th}}[c_V, c_f] - \mu_i^{\text{ex}}(1 + \delta_b \Delta_i^b) \right) C_{ij}^{\text{ex}-1} \right. \\ \left. \times \left(\mu_j^{\text{th}}[c_V, c_f] - \mu_j^{\text{ex}}(1 + \delta_b \Delta_j^b) \right) \right] \quad (8.2)$$

relying on the combined error,

$$\Delta_i^b = \left| \Delta_{\text{ggF},i}^P - \Delta_{\text{ttH},i}^P \right| + \Delta_{\text{VBF},i}^P + \Delta_{\text{WH},i}^P + \Delta_{\text{ZH},i}^P + \sum_{Y,a} (\Delta_{Y,i}^a + \Delta_{Y,i}), \quad (8.3)$$

or,

$$\Delta_i^b = \left| \Delta_{\text{ggF},i}^P - (\Delta_{\text{ttH},i}^P + \Delta_{\text{VBF},i}^P + \Delta_{\text{WH},i}^P + \Delta_{\text{ZH},i}^P) \right| + \sum_{Y,a} (\Delta_{Y,i}^a + \Delta_{Y,i}), \quad (8.4)$$

for the two configurations of correlations defined in Eq. (7.2)–(7.3), respectively.

The combinations of the errors on the partial decay widths are dictated by the fact that their nuisance parameters are either independent (among them and from the nuisance parameters at the production level) or taken 100% correlated to each other, as discussed in Section 6.5.

In Eq. (8.1), Δ_X^{amp} is either equal to Δ_X^{scale} (see Section 6.2) or taken as $\Delta_{\text{ggF}}^{\text{amp}} = \Delta_{\text{ggF}}^{\text{scale}} + \Delta_{\text{ggF}}^{\text{Q,V}}$, for the ggF channel (instead of Eq. (6.11)) with now, $\Delta_{\text{ggF}}^{\text{Q,V}} \simeq 9\%$, from the linear sum of the three errors originating from EFT assumptions and m_b scheme dependence [26]. These linear summations are all motivated by the fact that these errors are independent.

The $\Delta_X^{\text{PDF}+\alpha_s}$ uncertainty entering Eq. (8.1) is obtained from Refs. [20,53] using an “envelope method”, which corresponds exactly to the combinations in the bias approach presented in Section 3.5. Indeed, this combination is equivalent to a linear sum of the individual errors Δ_X^{set} , Δ_X^{data} and $\Delta_X^{\alpha_s}$, which are independent (cf. Section 6.1). Finally, the linear sum in Eq. (8.1) is justified by the independence of the errors Δ_X^{amp} and $\Delta_X^{\text{PDF}+\alpha_s}$.

The 1σ -errors (Δ 's) are taken to be exactly the symmetrised errors provided by the LHCHWG [17,18,20] in order to be conservative (similar discussion as in Sections 6.3 and 6.5). These errors are consistent with the previous marginalisation framework, so that the results from bias and marginalisation can readily be compared.

8.2. The Bayesian approach

8.2.1. Extremal bias

According to Section 2.2.2, the extremal bias within the Bayesian framework consists in deriving the best-fit regions in the c_V – c_f plane for two fixed values of the nuisance parameters, $\delta_b = \pm 1$, using the likelihood $L_{\text{bias}}(\delta_b)$ of Eq. (8.2). Recall that in the Bayesian case,

the best-fit regions are computed by integrating the posterior density probability, according to Eqs. (2.3)–(2.5). The priors ($\pi(\theta)$) for the parameters of interest (here $\theta \equiv c_V, c_f$) entering Eq. (2.3) are taken flat, *i.e.* $\pi(c_{V,f}) \propto 1$.

Note that, if the two extreme regions have an overlap, one cannot display them together consistently. Instead, one has to follow the rigorous definition of Eq. (2.14), using a discrete domain $\mathcal{D} = \{-1, 1\}$. This equation dictates to use the sum of the posteriors at $\delta_b = -1$ and $\delta_b = 1$, with each posterior separately normalised by its integral over the c_V – c_f plane.

8.2.2. Envelope method

The envelope method corresponds to letting vary continuously δ_b within $[-1, 1]$, *i.e.* this is the continuous version of the extremal bias, as discussed in Section 2.2.2. The corresponding likelihood is

$$\bar{L}_B(c_f, c_V) = \int_{-1}^1 d\delta_b \left[\frac{L_{\text{bias}}(c_f, c_V, \delta_b)}{\int dc_f \int dc_V L_{\text{bias}}(c_f, c_V, \delta_b)} \right]. \quad (8.5)$$

This likelihood is derived by applying Eq. (2.14) with the likelihood $L_{\text{bias}}(c_V, c_f, \delta_b)$ from Eq. (8.2). The best-fit regions are obtained through the standard procedure of Eqs. (2.3)–(2.5). Again, we take the priors for the parameter of interest to be flat, $\pi(c_{V,f}) \propto 1$.

8.3. The frequentist approach

8.3.1. Extremal bias

For the extremal bias in the frequentist framework (see Section 2.2.2), one uses again the likelihood $L_{\text{bias}}(\delta_b)$ (Eq. (8.2)), with δ_b fixed at the two extreme values $\delta_b = \pm 1$. In practice, in order to draw the best-fit regions in c_V – c_f , one can define a χ -squared function difference

$$\Delta\chi^2(c_f, c_V, \delta_b) = \chi^2(c_f, c_V, \delta_b) - \chi^2(\hat{c}_f, \hat{c}_V, \delta_b), \quad \chi^2(c_f, c_V, \delta_b) = -2\log[L_{\text{bias}}(\delta_b)], \quad (8.6)$$

as follows from Eq. (2.6). Remind that $\chi^2(\hat{c}_f, \hat{c}_V, \delta_b)$ stands for the minimum of χ^2 with respect to c_f, c_V for a given δ_b . The best-fit regions are obtained by drawing the contour levels of $\Delta\chi^2$ set at the values given in Eq. (2.8). Once more, the prior for the parameters of interest entering in Eq. (2.6) are taken flat, $\pi(c_{V,f}) \propto 1$.

If the two extreme regions overlap, the same remark as in the Bayesian case holds. To display consistently the two regions together, one has to follow the rigorous definition of Eq. (2.17), using a discrete domain $\mathcal{D} = \{-1, 1\}$. This equation dictates to use the minimum of the two $\Delta\chi^2$, *i.e.* $\min_{\delta_b \in \{-1, 1\}} [\Delta\chi^2(c_f, c_V, \delta_b)]$.

8.3.2. Envelope method

For the envelope method in the frequentist case, one can proceed with the χ^2 introduced in Eq. (8.6) and define

$$\bar{\chi}^2(c_f, c_V) = \min_{\delta_b \in [-1, 1]} \left[\chi^2(c_f, c_V, \delta_b) - \chi^2(\hat{c}_f, \hat{c}_V, \delta_b) \right], \quad (8.7)$$

according to the general definition of Eq. (2.17). This equation is the frequentist analog of Eq. (8.5). In order to draw the best-fit regions in the c_V – c_f plane, one should then define

$$\Delta\bar{\chi}^2(c_f, c_V) = \bar{\chi}^2(c_f, c_V) - \bar{\chi}^2(\hat{c}_f, \hat{c}_V). \quad (8.8)$$

The best-fit regions are obtained by drawing the contour levels of $\Delta\bar{\chi}^2$ set at the values given in Eq. (2.8). Again, the prior for the c_V, c_f parameters entering in Eq. (2.6) are taken flat, $\pi(c_{V,f}) \propto 1$.

Let us finally recall the parallel between Eq. (8.5) and Eq. (8.7). As first explained in Section 2.2.2, the subtracted term in Eq. (8.7) is the frequentist analogy of the ratio over $\int dc_f dc_V L_{\text{bias}}(c_f, c_V, \delta_b)$ in Eq. (8.5). In both cases, the effect of this term is to remove the contribution of δ_b to goodness-of-fit (which avoids favouring specific values of δ_b). Both formulas are analog up to exchanging integration over δ_b with minimisation over δ_b . The fact that the integration/minimisation over δ_b is performed on the whole range $[-1, 1]$, rather than on the discrete domain $\{-1, 1\}$, leads to an envelope in the c_f – c_V plane, instead of two distinct domains as in the extremal bias.

8.4. Numerical results

In this section, we apply both the frequentist and Bayesian versions of the bias method to the Higgs likelihood. We stress that the Higgs likelihood $L_{\text{bias}}(\delta_b)$ is exactly the same in the two statistical frameworks, so that the discrepancies observed among the plots originate solely from the different statistical treatments. These two treatments differ in their definition of the best-fit regions (see Section 2.1) and their realisation of the bias principle (see Eqs. (2.14), (2.17)).

8.4.1. Extremal bias

In Fig. 9, we present the best-fit regions obtained through the Bayesian and frequentist bias methods, respectively described in Sections 8.2.1 and 8.3.1. The likelihood, $L_{\text{bias}}(\delta_b)$ of Eq. (8.2), is used together with one of the two combined errors (8.3)–(8.4) depending on which correlation configuration is considered (Eq. (7.2) or Eq. (7.3) respectively).

The left and right panels of Fig. 9 correspond to the two correlation configurations surrounding the case with realistic correlations. It turns out that the best-fit regions obtained in these two extreme correlation configurations have only mild differences.

Now, compare the two upper plots and lower plots of Fig. 9, corresponding respectively to the frequentist and Bayesian treatments. A small difference appears at the junction of the two set of regions, coming from the different realisation of the bias principle in the two statistical frameworks. Besides, the frequentist best-fit regions are slightly larger than the Bayesian ones, due to the non-equivalent definitions of the Bayesian and frequentist contours. Overall, there is a strong resemblance between the Bayesian and frequentist results. This reflects the weak impact of choosing the Bayesian or frequentist procedure for the extremal bias.

Let us now compare the lower plots of Fig. 9 with the previous Bayesian marginalisation plots obtained in Fig. 5 – considering of course respectively the two correlation configurations used in the left and right plots. One can clearly see that the best-fit regions⁵⁵ obtained from the extremal bias are larger than the ones obtained through marginalisation. This is because the regions in Fig. 5, derived by marginalising, correspond somehow to fix the nuisance parameters to their

⁵⁵ Notice that these best-fit regions include essentially the two extreme sub-domains corresponding to $\delta_b = \pm 1$.

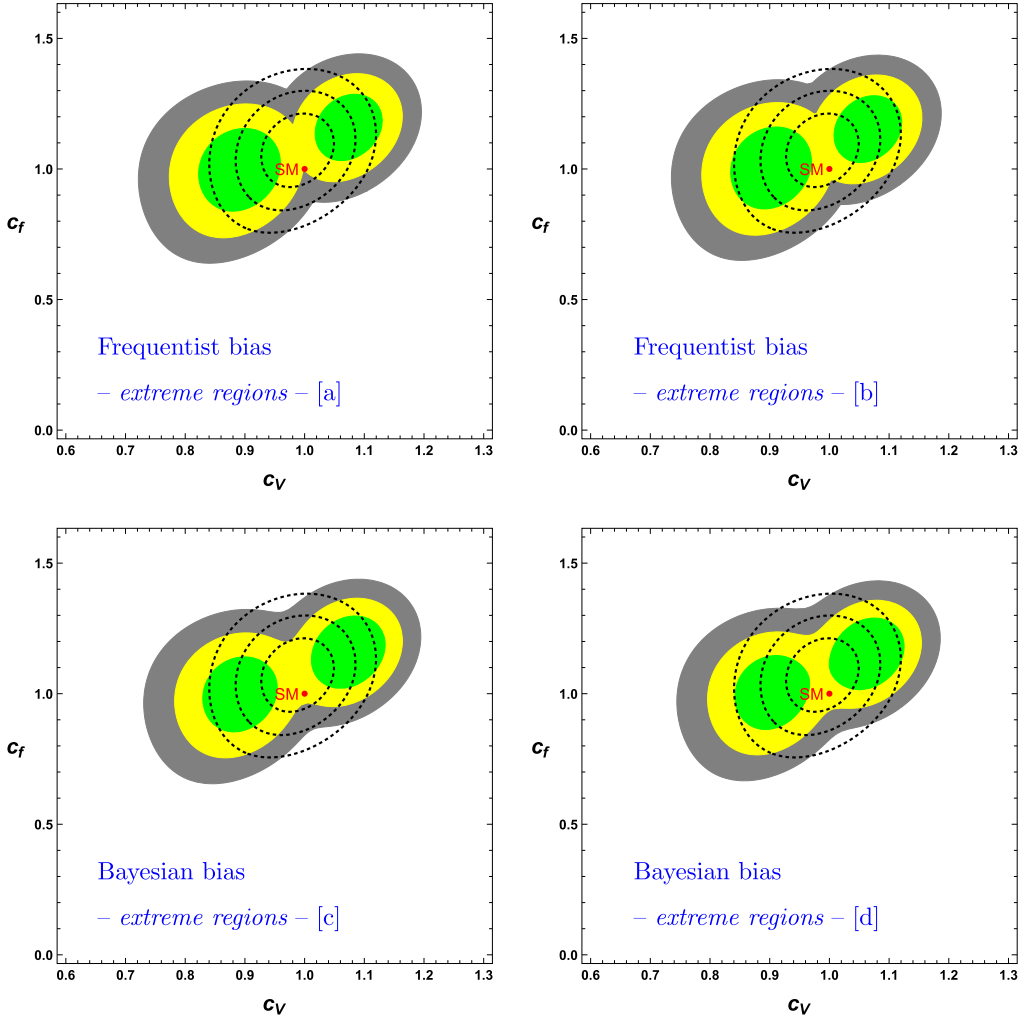


Fig. 9. The best-fit regions in the c_V - c_f plane obtained through an extremal bias. The 68%, 95% and 99% confidence regions are represented respectively by the green, yellow and grey domains. The upper plots illustrate the frequentist approach whereas the two lower ones show the Bayesian approach. The [a], [c] and [b], [d] plots correspond, respectively, to the characteristic correlation configurations described in Eq. (8.3) and Eq. (8.4). The dashed contours illustrate the case without theoretical uncertainties. The SM prediction is shown by the red point. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

values favoured by the fit. For the present Higgs fits, it turns out that these preferred values are close to $\delta \approx -1$. Hence, the regions from the extremal bias (Fig. 9) being obtained for $\delta_b = \pm 1$ (lower left set is for $\delta_b = -1$ ⁵⁶), they clearly cover more space in the c_V - c_f plane than the domains in Fig. 5.

⁵⁶ The dependence of the best-fit region location on the nuisance parameter is induced by the dependence of the likelihood (8.2) on, $\mu_i^{\text{ex}}[1 + \delta_b \Delta_i^b]$.

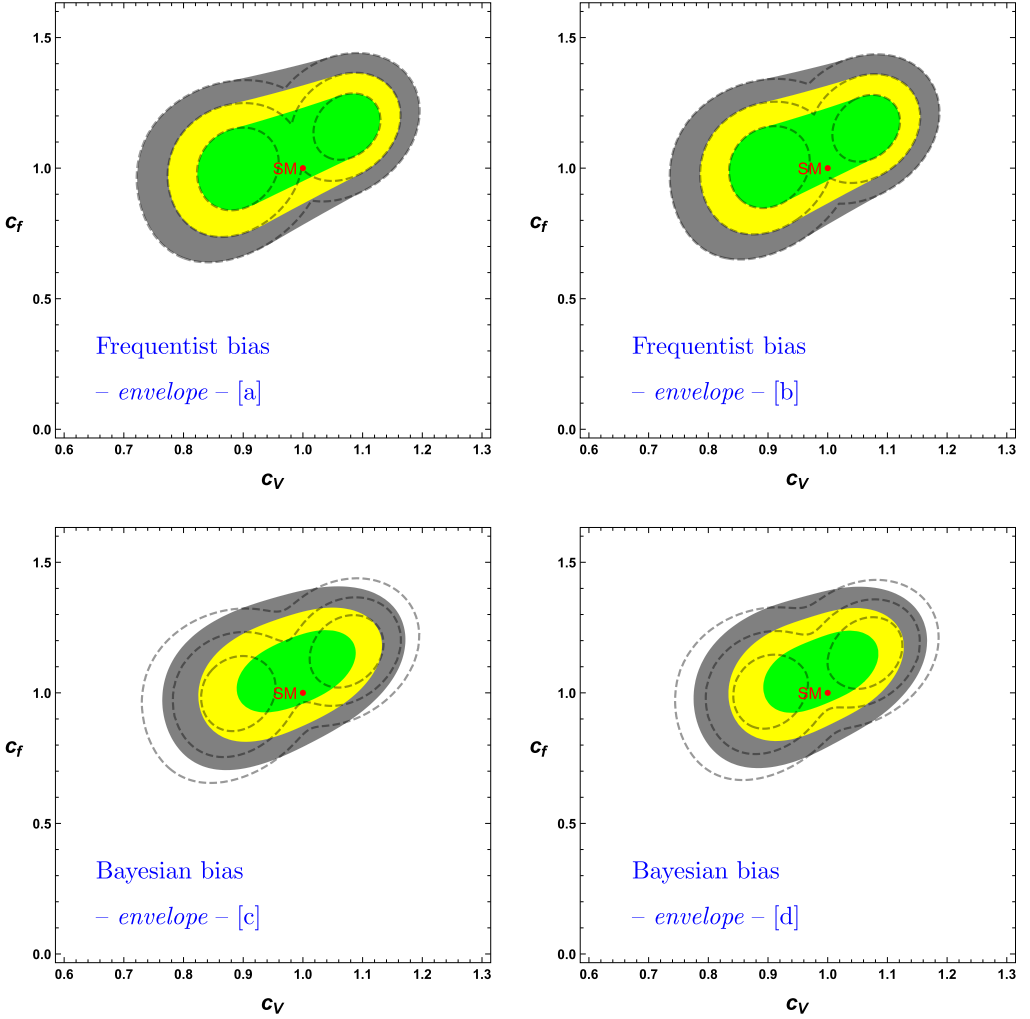


Fig. 10. The best-fit regions in the c_V - c_f plane obtained through the envelope method. The 68%, 95% and 99% confidence regions are represented respectively by the green, yellow and grey domains. The upper plots illustrate the frequentist approach whereas the two lower ones show the Bayesian approach. The [a], [c] and [b], [d] plots correspond, respectively, to the characteristic correlation configurations described in Eq. (8.3) and Eq. (8.4). The dashed grey contours illustrate the best-fit regions at 68% C.L., 95% C.L. and 99% C.L., obtained in Fig. 9. The SM prediction is shown by the red point. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

8.4.2. Envelope method

The four plots of Fig. 10 illustrate the Bayesian and frequentist envelope methods performed accordingly to Sections 8.2.2 and 8.3.2. Again, both correlation configurations, giving rise to the combined errors of Eqs. (8.3)–(8.4), are studied numerically. The two upper and lower plots of Fig. 10 differ due to the direct envelope method being not equivalent within the Bayesian and frequentist cases.

The sets of frequentist envelopes represent the best-fit areas that would be obtained by superimposing the best-fit regions of the extremal bias, but for δ_b spanning continuously the interval

$[-1, 1]$. This correspondence between the envelope method and extremal bias appears clearly when one realises (*cf.* end of Section 2.2.2) that the former is based on the Eqs. (8.7)–(8.8) while the latter can be obtained through the same equations just with a minimisation over the discrete domain $\delta_b \in \mathcal{D} = \{-1, 1\}$ in Eq. (8.7), instead of the continuous range $[-1, 1]$. The correspondence is visible when comparing the envelopes with the extreme sets of best-fit domains at $\delta_b = \pm 1$, obtained previously from the frequentist bias method and also superimposed on upper plots of Fig. 10, as dashed contours: these contours draw exactly the extreme limits of the envelopes.

The two sets of Bayesian envelopes obtained in the two lower plots of Fig. 10 represent less conservative regions with respect to the frequentist envelope. Besides, the envelopes of these plots cover smaller regions than the best-fit domains that would be obtained by superimposing the best-fit regions of the extremal bias, but for δ_b spanning continuously the interval $[-1, 1]$. This appears clearly when comparing those envelopes to the extreme sets of best-fit regions at $\delta_b = \pm 1$, obtained previously from the Bayesian bias method (once more superimposed on the lower plots of Fig. 10, as dashed contours).

Finally, we mention that the SM point belongs to all the 68% C.L. regions of Fig. 10. At this level, we can illustrate one of the potential interests of the bias. Let us consider an hypothetical but plausible situation. For example, suppose that with future LHC data, the SM point would fall outside the 3σ region obtained by marginalising. Such a discrepancy could be interpreted either as an indirect effect of physics underlying the SM on the Higgs sector, or as a shift of the best-fit regions induced by values of the nuisance parameters favoured statistically by the fit. This shift induced by the nuisance parameters would come from the fact that the nuisance parameters and the parameters of interest are determined simultaneously. In contrast, in the envelope method, a SM prediction falling beyond the 3σ region of the parameter space of interest could indicate the presence of new physics (the entire 1σ interval of the nuisance parameters being covered), although of course it would be an indirect new physics signature only if (i) the 1σ intervals for δ 's are not underestimated and (ii) the true value for each nuisance parameter lies well within its 1σ interval (not in the 2σ range for example). This example provides a motivation to apply both bias and marginalisation methods, which are somehow complementary.

9. Conclusions

The main goal of this analysis was to work out a consistent statistical treatment of the theoretical uncertainties in the fits of the Higgs boson rates. We have analysed in a unified formalism both the Bayesian and frequentist approaches to theoretical uncertainties. We systematically analysed how to perform error combinations in a given statistical context and we have introduced a framework to use the bias principle on firm ground.

This analysis has been the opportunity to update the Higgs rate fit based on the latest LHC data at 7 and 8 TeV. In the case of Bayesian marginalisation, we have found that the SM prediction for the Higgs couplings still falls into the 68% C.L. region of the c_V – c_f plane. Bayesian marginalisation benefits from well-defined distributions for the nuisance parameters and from an easier convolution of these error distributions compared to frequentist marginalisation.

We have reviewed all the fundamental sources of the individual theoretical errors involved in the SM Higgs cross sections and branching ratios. Then those errors have been combined in a careful ‘step-by-step’ approach following the Bayesian rules. In this task of combining a significant number of uncertainties (various Higgs production modes, decay channels...), we were

helped by the leading moment approximation – which has been deduced from considerations on the moment-generating function.

This has allowed us to show that the prior of the total uncertainty resulting from the combination of all the theoretical errors (using flat priors for the unknown ones) converges to a nearly Gaussian shape. Besides, it also came out from the numerical results that the precise form of this final theoretical prior is not crucial with respect to the determination of the best-fit regions. This conclusion holds only for the present data, which still have large experimental errors with respect to the theoretical ones.

In contrast, our analysis has shown that the correlations of the theoretical uncertainties among the Higgs detection channels induce a significant shift of the best-fit domains in the space of the parameters of interest. These correlations appear thus to be an unavoidable ingredient of the fits. The Higgs fits were performed in two extreme configurations of theoretical correlations between the various detection channels. The most realistic correlation setup is an intermediate configuration between those two. Such an approach is thus conservative. Besides, considering characteristic configurations have allowed us to derive simple analytic expressions for the marginal likelihood functions.

For future Higgs fits, given the ambiguities inherent to the estimation of the theoretical error magnitudes, we recommend to present an additional analysis with 1σ errors enhanced by a typical factor of 1.5 as a conservative benchmark. Such a factor is consistent with the 1σ theoretical errors preferred by the data. Of course the present degree of arbitrariness in the theoretical error magnitudes could be improved for instance with future higher order QCD calculations or new methods to determine the PDFs.

Finally, we have provided a rigorous statistical framework for the bias principle, which constitutes an alternative to marginalisation. This framework has leads us to define two complementary bias treatments: the extremal bias and the envelope method. The bias principle is more conservative than marginalisation by construction, and does not depend on the shape of the priors of the nuisance parameters, which are not always known. Therefore, a reasonable advice is to apply both the marginalisation and bias methods to the Higgs data. Using the envelope method, we find that the SM prediction belongs to the 68% C.L. region of the c_V – c_f plane.

Beyond parameter inference, one can recall that in the case of appearance of a significant deviation with respect to the SM Higgs couplings in future data, the Bayesian setting would also provide a standard procedure to assess discovery by using a discovery Bayes factor (*i.e.* comparing the Bayesian evidences of the SM and new physics hypotheses).

Acknowledgements

The authors gratefully acknowledge Glen Cowan for his involvement and valuable inputs. The authors also thank Henri Bachacou, Damir Bečirević, Abdelhak Djouadi, Guillaume Drieu La Rochelle, Gero von Gersdorff, Andrea Massironi, Sezen Sekmen, Michael Spira for stimulating and useful discussions, and especially Marumi Kado for helpful exchanges about the bias methods. It is finally a pleasure to thank the organizers of the “Physics at TeV Colliders” Workshop at *Les Houches* (2013) where this work was initiated [60]. SF acknowledges the *Brazilian Ministry of Science, Technology and Innovation* for financial support as well as the LPT at Orsay University for hospitality during a part of this work. The work of GM is partly supported by the *Institut Universitaire de France*, the European ERC Grant “Higgs@LHC” and the European Union FP7 ITN INVISIBLES (Marie Curie Actions, PITN-GA-2011-289442).

Appendix A. The leading moment approximation

Consider a linear combination δ_C of random variables δ_A, δ_B , given by

$$\delta_C \Delta_C = \delta_A \Delta_A + \delta_B \Delta_B, \quad (\text{A.1})$$

with $\Delta_b \ll \Delta_a$ and $E[\delta] = 0$, $V[\delta] = 1$ by convention. The *pdf* of $\delta_A, \delta_B, \delta_C$ are noted respectively π_A, π_B, π_C .

We mainly work in Laplace space, using the moment-generating function

$$\phi_Z(t) = E[e^{Zt}] = \int dz e^{zt} \pi_Z(z). \quad (\text{A.2})$$

If all moments are finite, $\phi_Z(t) = \sum_{n=0}^{\infty} \frac{m_n^Z}{n!} t^n$, where m_n^Z denotes the n -th moment of Z . m_1 being the mean, we have $m_1^A = 0 = m_1^B$. m_2 being the variance, we have $m_2^A = 1 = m_2^B$.

Let us assume in a first place that δ_A, δ_B are independent. This implies that $\pi_{A,B} = \pi_A \pi_B$, that the *pdf* of δ_C is given by a convolution product, and that the moment generating function of δ_C is given by the product

$$\phi_C(\Delta_C t) = \phi_A(\Delta_A t) \phi_B(\Delta_B t). \quad (\text{A.3})$$

Having $\Delta_B \ll \Delta_A$ by assumption, we can use Δ_B/Δ_A has an expansion parameter. At leading order, neglecting the contribution from δ_B to the combination amounts to approximate

$$\phi_B(\Delta_B t) = 1 + O(\Delta_B^2 t^2) \quad (\text{A.4})$$

in the product (A.2). This corresponds to approximating π_B as a Dirac distribution centred on zero.

Going one order further in the expansion leads to keep

$$\phi_B(\Delta_B t) = 1 + \Delta_B^2 \frac{t^2}{2!} + O(\Delta_B^3 t^3). \quad (\text{A.5})$$

This subleading term induces $O(\Delta_B^2/\Delta_A^2)$ corrections to the moments of δ_C . Explicitly one finds

$$\Delta_C^n m_n^C = \Delta_A^n \left(m_n^A + \frac{\Delta_B^2}{\Delta_A^2} m_{n-2}^A N_n \right), \quad (\text{A.6})$$

with $N_n = n!/(2(n-2)!)$. At that point, the corrections to all moments m_n^C should in principle be kept.

We then take a second step in our approximation, by considering that the amount of information relevant for our problem somehow decreases with the order of the moment. As a consequence, the corrections to the first moments are the more relevant. Keeping the next-to-leading corrections up to order p , our approximation scheme thus reads

$$\Delta_C^n m_n^C = \begin{cases} \Delta_A^n \left(m_n^A + \frac{\Delta_B^2}{\Delta_A^2} m_{n-2}^A N_n + O\left(\frac{\Delta_B^3}{\Delta_A^3}\right) \right) & \text{if } 1 \leq n \leq p \\ \Delta_A^n \left(m_n^A + O\left(\frac{\Delta_B^2}{\Delta_A^2}\right) \right) & \text{if } p < n. \end{cases} \quad (\text{A.7})$$

In particular, truncating the corrections at $p = 2$ amounts to take into account only the correction to the variance,

$$\Delta_C^2 = \Delta_A^2 + \Delta_B^2. \quad (\text{A.8})$$

The other details of the shape remaining unperturbed, it follows that

$$\pi_C = \pi_A + O\left(\frac{\Delta_B^2}{\Delta_A^2} \delta^{(3)}\right). \quad (\text{A.9})$$

Here $\delta^{(3)}$ is the third derivative of the Dirac distribution. It comes from the Laplace transform of the t^3 term of the moment-generating function (see also Ref. [61]). This $\delta^{(3)}$ should be understood as the leading functional deformation to π_A . In practice, it appears that keeping only the first leading moment is appropriate when π_A is a one-parameter *pdf*. In that case, the parameter characterising π_C is identified through the combination of variances. For example, taking the normal distribution $\pi_A = \mathcal{N}(0, \sigma_A^2)$ gives $\sigma_C^2 = \sigma_A^2 + \Delta_B^2$ and $\pi_C = \mathcal{N}(0, \sigma_C^2)$.⁵⁷

The approach above also extends to correlated variables. The difference with respect to the uncorrelated case is that the moment-generating functions do not factorise, as δ_A, δ_B now share common moments. For example, truncating the corrections at $p = 2$ gives the correction

$$m_2^C = \Delta_A^2 + \Delta_B^2 + 2\Delta_A \Delta_B \rho, \quad (\text{A.10})$$

where $\rho (= m_1^{AB})$ is the covariance of (δ_A, δ_B) . In the limit of full correlation, one has $\rho = 1$, so that $\Delta_C^2 = (\Delta_A + \Delta_B)^2$. Note that when $\rho > \Delta_B/\Delta_A$ in Eq. (A.10), the contribution from the correlation term becomes larger than the contribution from the square term Δ_B^2 .

Finally, the leading moment approximation also extends to the case of several linear combinations of variables. Here we consider the case with two linear combinations of two variables δ_A, δ_B with correlation ρ . The combinations are defined as

$$\delta_{C1} \Delta_{C1} = \delta_A \Delta_{A1} + \delta_B \Delta_{B1}, \quad (\text{A.11})$$

$$\delta_{C2} \Delta_{C2} = \delta_A \Delta_{A2} + \delta_B \Delta_{B2}. \quad (\text{A.12})$$

The variances are found to be

$$\Delta_{C1}^2 = \Delta_{A1}^2 + \Delta_{B1}^2 + 2\rho \Delta_{A1} \Delta_{B1}, \quad (\text{A.13})$$

$$\Delta_{C2}^2 = \Delta_{A2}^2 + \Delta_{B2}^2 + 2\rho \Delta_{A2} \Delta_{B2}, \quad (\text{A.14})$$

like in the one-combination case described above. In the case $\Delta_{A1} \gg \Delta_{B1}, \Delta_{A2} \gg \Delta_{B2}$, the correlation coefficient ρ_{12} between δ_{C1} and δ_{C2} reads

$$\rho_{12} = 1 - \frac{1}{2} \left(\frac{\Delta_{B1}}{\Delta_{A1}} - \frac{\Delta_{B2}}{\Delta_{A2}} \right)^2 - \rho^2 \left(\frac{\Delta_{B1}}{\Delta_{A1}} + \frac{\Delta_{B2}}{\Delta_{A2}} \right)^2 + O\left(\frac{\Delta_{B1,2}^3}{\Delta_{A1,2}^3}\right). \quad (\text{A.15})$$

In the case $\Delta_{A1} \gg \Delta_{B1}, \Delta_{A2} \ll \Delta_{B2}$, the correlation coefficient is instead

$$\rho_{12} = \rho + \left(\frac{\Delta_{B1}}{\Delta_{A1}} + \frac{\Delta_{A2}}{\Delta_{B2}} \right) (1 - \rho^2) + O\left(\frac{\Delta_{B1,2}^2}{\Delta_{A1,2}^2}\right). \quad (\text{A.16})$$

⁵⁷ It is worth noticing that in the Gaussian case, this identification reproduces exactly the correction to the m_n^C at any order. This is not true for other distributions.

References

- [1] The ATLAS Collaboration, Phys. Lett. B 716 (2012) 1.
- [2] The CMS Collaboration, Phys. Lett. B 716 (2012) 30.
- [3] P. Higgs, Phys. Lett. 12 (1964) 132;
P. Higgs, Phys. Rev. Lett. 13 (1964) 506;
F. Englert, R. Brout, Phys. Rev. Lett. 13 (1964) 321;
G. Guralnik, C. Hagen, T. Kibble, Phys. Rev. Lett. 13 (1964) 585;
S. Weinberg, Phys. Rev. Lett. 19 (1967) 1264.
- [4] The ATLAS Collaboration, CONF-2013-014;
The ATLAS Collaboration, CONF-2013-034;
The ATLAS Collaboration, CONF-2015-007;
The ATLAS Collaboration, arXiv:1507.04548 [hep-ex].
- [5] The CMS Collaboration, PAS-HIG-12-036;
The CMS Collaboration, PAS-HIG-12-045;
The CMS Collaboration, PAS-HIG-14-009;
The CMS Collaboration, arXiv:1412.8662 [hep-ex].
- [6] ATLAS web page, <http://atlas.web.cern.ch/Atlas/Collaboration/>.
- [7] CMS web page, <https://twiki.cern.ch/twiki/bin/view/CMSPublic/PhysicsResults>.
- [8] A. Djouadi, G. Moreau, Phys. Lett. B 660 (2008) 67;
C. Bouchart, G. Moreau, Phys. Rev. D 80 (2009) 095022.
- [9] A. Azatov, J. Galloway, Int. J. Mod. Phys. A 28 (2013) 1330004.
- [10] A. Arbey, M. Battaglia, A. Djouadi, F. Mahmoudi, J. High Energy Phys. 1209 (2012) 107;
A. Arbey, M. Battaglia, A. Djouadi, F. Mahmoudi, Phys. Lett. B 720 (2013) 153;
A. Djouadi, L. Maiani, G. Moreau, A. Polosa, J. Quevillon, V. Riquer, Eur. Phys. J. C 73 (2013) 2650.
- [11] Physics Briefing Book, Input for the Strategy Group to draft the update of the European Strategy for Particle Physics, CERN-ESG-005.
- [12] D. Carmi, A. Falkowski, E. Kuflik, T. Volansky, J. Zupan, J. High Energy Phys. 1210 (2012) 196;
J. Espinosa, C. Grojean, M. Muhlleitner, M. Trott, J. High Energy Phys. 1212 (2012) 045;
P. Giardino, K. Kannike, M. Raidal, A. Strumia, Phys. Lett. B 718 (2012) 469;
J. Ellis, T. You, J. High Energy Phys. 1209 (2012) 123;
T. Corbett, O.J.P. Eboli, J. Gonzalez-Fraile, M.C. Gonzalez-Garcia, Phys. Rev. D 86 (2012) 075013;
F. Bonnet, T. Ota, M. Rauch, W. Winter, Phys. Rev. D 86 (2012) 093014;
A. Alves, et al., Eur. Phys. J. C 73 (2) (2013) 2288;
J. Cao, Z. Heng, J.M. Yang, J. Zhu, J. High Energy Phys. 1210 (2012) 079;
S. Banerjee, S. Mukhopadhyay, B. Mukhopadhyaya, J. High Energy Phys. 1210 (2012) 062;
I. Low, J. Lykken, G. Shaughnessy, Phys. Rev. D 86 (2012) 093012;
M. Klute, R. Lafaye, T. Plehn, M. Rauch, D. Zerwas, Phys. Rev. Lett. 109 (2012) 101801;
T. Plehn, M. Rauch, Europhys. Lett. 100 (2012) 11002;
G. Cacciapaglia, A. Deandrea, G. Drieu La Rochelle, J.-B. Flament, J. High Energy Phys. 1303 (2013) 029;
N. Bonne, G. Moreau, Phys. Lett. B 717 (2012) 409;
G. Bélanger, B. Dumont, U. Ellwanger, J.F. Gunion, S. Kraml, J. High Energy Phys. 1302 (2013) 053;
C. Cheung, S.D. McDermott, K.M. Zurek, J. High Energy Phys. 1304 (2013) 074;
K. Cheung, J.S. Lee, P.-Y. Tseng, J. High Energy Phys. 1305 (2013) 134;
G. Moreau, Phys. Rev. D 87 (2013) 015027.
- [13] J. Ellis, T. You, J. High Energy Phys. 1306 (2013) 103;
T. Alanne, S. Di Chiara, K. Tuominen, J. High Energy Phys. 1401 (2014) 041;
A. Falkowski, F. Riva, A. Urbano, J. High Energy Phys. 1311 (2013) 111;
T. Corbett, et al., arXiv:1505.05516 [hep-ph];
S. Banerjee, T. Mandal, B. Mellado, B. Mukhopadhyaya, arXiv:1505.00226 [hep-ph].
- [14] A. Djouadi, G. Moreau, Eur. Phys. J. C 73 (2013) 9, 2512.
- [15] P.P. Giardino, K. Kannike, I. Masina, M. Raidal, A. Strumia, J. High Energy Phys. 1405 (2014) 046.
- [16] B. Dumont, S. Fichet, G. von Gersdorff, J. High Energy Phys. 1307 (2013) 065.
- [17] L.H.C. Higgs, Cross Section Working Group web page, <https://twiki.cern.ch/twiki/bin/view/LHCPhysics/LHCHSWG>.
- [18] Handbook of LHC Higgs cross sections: 1. Inclusive observables, arXiv:1101.0593 [hep-ph], CERN-2011-002.

- [19] Handbook of LHC Higgs cross sections: 2. Differential distributions, arXiv:1201.3084 [hep-ph], CERN-2012-002.
- [20] Handbook of LHC Higgs cross sections: 3. Higgs properties, arXiv:1307.1347 [hep-ph], CERN-2013-004.
- [21] K. Cranmer, S. Kreiss, D. Lopez-Val, T. Plehn, Phys. Rev. D 91 (2015) 054032.
- [22] M. Cacciari, N. Houdeau, J. High Energy Phys. 1109 (2011) 039;
E. Bagnaschi, M. Cacciari, A. Guffanti, L. Jenniches, J. High Energy Phys. 1502 (2015) 133.
- [23] R.S. Gupta, H. Rzehak, J.D. Wells, Phys. Rev. D 86 (2012) 095001.
- [24] R. Lafaye, T. Plehn, M. Rauch, D. Zerwas, M. Duehrssen, J. High Energy Phys. 0908 (2009) 009.
- [25] C. Anastasiou, C. Duhr, F. Dulat, F. Herzog, B. Mistlberger, Phys. Rev. Lett. 114 (2015) 212001.
- [26] J. Baglio, A. Djouadi, R.M. Godbole, Phys. Lett. B 716 (2012) 203.
- [27] C. Anastasiou, R. Boughezal, F. Petriello, J. High Energy Phys. 0904 (2009) 003.
- [28] J.O. Berger, R.L. Wolpert, The likelihood principle, in: Institute of Mathematical Statistics Lecture Notes – Monograph Series, vol. 6, 1988.
- [29] A. Wald, Tests of statistical hypotheses concerning several parameters when the number of observations is large, Trans. Am. Math. Soc. 54 (3) (1943) 426.
- [30] J. Beringer, et al., Particle Data Group, Phys. Rev. D 86 (2012) 010001;
K.A. Olive, et al., Particle Data Group, Chin. Phys. C 38 (2014) 090001.
- [31] The ATLAS Collaboration, CONF-2013-012;
The ATLAS Collaboration, Phys. Rev. D 90 (2014) 112015;
The ATLAS Collaboration, Phys. Lett. B 740 (2015) 222.
- [32] The ATLAS Collaboration, CONF-2013-012;
The ATLAS Collaboration, Phys. Lett. B 726 (2013) 88;
The ATLAS Collaboration, Phys. Lett. B 738 (2014) 234;
The ATLAS Collaboration, Phys. Rev. D 91 (2015) 012006.
- [33] The ATLAS Collaboration, CONF-2013-030;
The ATLAS Collaboration, CONF-2014-060;
The ATLAS Collaboration, CONF-2015-006;
The ATLAS Collaboration, CONF-2015-005;
The ATLAS Collaboration, arXiv:1412.2641 [hep-ex].
- [34] The ATLAS Collaboration, CONF-2012-170;
The ATLAS Collaboration, CONF-2014-043;
The ATLAS Collaboration, J. High Energy Phys. 1501 (2015) 069;
The ATLAS Collaboration, arXiv:1503.05066 [hep-ex].
- [35] The ATLAS Collaboration, CONF-2012-160;
The ATLAS Collaboration, CONF-2014-061;
The ATLAS Collaboration, J. High Energy Phys. 1504 (2015) 117.
- [36] The CMS Collaboration, PAS-HIG-13-001;
The CMS Collaboration, PAS-HIG-13-029;
The CMS Collaboration, J. High Energy Phys. 1409 (2014) 087.
- [37] The CMS Collaboration, PAS-HIG-13-002.
- [38] The CMS Collaboration, PAS-HIG-13-003;
The CMS Collaboration, PAS-HIG-13-023.
- [39] The CMS Collaboration, PAS-HIG-12-020;
The CMS Collaboration, PAS-HIG-12-035;
The CMS Collaboration, PAS-HIG-13-012;
The CMS Collaboration, PAS-HIG-14-004;
The CMS Collaboration, PAS-HIG-14-010.
- [40] The CMS Collaboration, PAS-HIG-13-004;
The CMS Collaboration, PAS-HIG-13-033.
- [41] CDF web page, <http://www-cdf.fnal.gov>;
D0 web page, <http://www-d0.fnal.gov>.
- [42] CDF Collaboration, D0 Collaboration, CDF Note 10884, D0 Note 6348, arXiv:1207.0449 [hep-ex].
- [43] A. Djouadi, Phys. Rep. 457 (2008) 1.
- [44] S. Fichet, G. von Gersdorff, J. High Energy Phys. 1403 (2014) 102.
- [45] P.R. Archer, M. Carena, A. Carmona, M. Neubert, J. High Energy Phys. 1501 (2015) 060.
- [46] G. Moreau, Phys. Rev. D 87 (2013) 015027.
- [47] R. Trotta, Contemp. Phys. 49 (2008) 71.

- [48] C. Mc Neile, C.T.H. Davies, E. Follana, K. Hornbostel, G.P. Lepage, *Phys. Rev. D* 82 (2010) 034512.
- [49] Review of lattice results concerning low energy particle physics, *Eur. Phys. J. C* 74 (2014) 2890.
- [50] The PDF4LHC working group interim recommendations, arXiv:1101.0538 [hep-ph].
- [51] PDF4LHC Working Group web page, <http://www.hep.ucl.ac.uk/pdf4lh>.
- [52] S. Forte, G. Watt, *Annu. Rev. Nucl. Part. Sci.* 63 (2013) 291.
- [53] A.D. Martin, W.J. Stirling, R.S. Thorne, G. Watt, *Eur. Phys. J. C* 64 (2009) 653.
- [54] C. Anastasiou, S. Buehler, F. Herzog, A. Lazopoulos, *J. High Energy Phys.* 1204 (2012) 004.
- [55] J. Baglio, A. Djouadi, S. Ferrag, R.M. Godbole, *Phys. Lett. B* 699 (2011) 368;
J. Baglio, A. Djouadi, S. Ferrag, R.M. Godbole, *Phys. Lett. B* 702 (2011) 105 (Erratum);
J. Baglio, A. Djouadi, R.M. Godbole, arXiv:1107.0281 [hep-ph] (Extended version of talks given at several winter conferences).
- [56] J. Baglio, A. Djouadi, *J. High Energy Phys.* 1103 (2011) 55.
- [57] J. Baglio, A. Djouadi, *J. High Energy Phys.* 1010 (2010) 64.
- [58] ATLAS Collaboration, CMS Collaboration, The LHC Higgs Combination Group, Procedure for the LHC Higgs boson search combination in Summer 2011, CMS-NOTE-2011-005, ATL-PHYS-PUB-2011-11.
- [59] See for example, D. Sivia, J. Skilling, *Data Analysis, a Bayesian Tutorial*, Oxford Science Publications, 2006.
- [60] G. Brooijmans, et al., Proceedings of the New Physics Working Group of the 2013 Les Houches Workshop, Physics at TeV Colliders, Les Houches, 3–21 June 2013, 201 pages, arXiv:1405.1617 [hep-ph], June 2013.
- [61] S. Fichet, *Int. J. Mod. Phys. A* 30 (2015) 09, 1550039.