



Uncertainty components in profile likelihood fits

Andrés Pinto^{1,2,a}, Zhibo Wu^{1,3,b}, Fabrice Balli^{1,c}, Nicolas Berger^{4,d}, Maarten Boonekamp^{1,2,e}, Émilien Chapon^{1,f}, Tatsuo Kawamoto^{3,g}, Bogdan Malaescu^{5,h}

¹ Département de Physique des Particules, Université Paris-Saclay, CEA, 91191 Gif-sur-Yvette, France

² Institut für Physik, Universität Mainz, Staudingerweg 7, 55128 Mainz, Germany

³ Department of Modern Physics, University of Science and Technology of China, Hefei, Anhui, China

⁴ LAPP, Univ. Savoie Mont Blanc, CNRS/IN2P3, Annecy, France

⁵ LPNHE, Sorbonne Université, Université Paris Cité, CNRS/IN2P3, 75252 Paris, France

Received: 13 November 2023 / Accepted: 26 April 2024

© The Author(s) 2024

Abstract When a measurement of a physical quantity is reported, the total uncertainty is usually decomposed into statistical and systematic uncertainties. This decomposition is not only useful for understanding the contributions to the total uncertainty, but is also required to propagate these contributions in subsequent analyses, such as combinations or interpretation fits including results from other measurements or experiments. In profile likelihood fits, widely applied in high-energy physics analyses, contributions of systematic uncertainties are routinely quantified using “impacts,” which are not adequate for such applications. We discuss the difference between impacts and actual uncertainty components, and establish methods to determine the latter in a wide range of statistical models.

Contents

1 Introduction	1
2 Example: Higgs boson mass in the di-photon and four-lepton channels	2
3 Uncertainty decomposition in covariance representation	3
4 Equivalence between the covariance and nuisance parameter representations	4
5 Uncertainty decomposition from shifted observables	5

6 Examples	6
6.1 Combination of two measurements	6
6.2 W -boson mass fits	6
Simulation	6
Uncertainty decomposition	6
6.3 Use of decomposed uncertainties in subsequent fits or combinations	6
7 Conclusion	7
References	7

1 Introduction

Measurement results are usually reported quoting not only the total uncertainty on the measured values but also their breakdown into uncertainty components—usually the statistical uncertainty and one or more components of systematic uncertainty. A consistent propagation of uncertainties is of utmost importance for global analyses of measurement data, for example, for determining the anomalous magnetic moment of the muon [1] or the parton distribution functions of the proton [2], and for the measurement of Z boson properties at LEP1 [3], the top-quark mass [4], or the Higgs boson properties [5] at the LHC. In high-energy physics experiments, different techniques are used for obtaining this decomposition, depending on (but not fundamentally related to) the test statistic used to obtain the results.

The simplest statistical method consists in comparing a measured quantity or distribution to a model, parameterized only in terms of the physical constants to be determined. Auxiliary parameters (detector calibrations, theoretical predictions, etc.) on which the model depends are fixed to their best estimates. The measured values of the physical constants result from the maximization of the corresponding

^a e-mail: andres.elay.pinto.pinoargote@cern.ch

^b e-mail: zhibo.wu@cern.ch

^c e-mail: fabrice.balli@cern.ch

^d e-mail: nicolas.berger@cern.ch

^e e-mail: Maarten.Boonekamp@cern.ch (corresponding author)

^f e-mail: emilien.chapon@cern.ch

^g e-mail: tatsuo.kawamoto@cern.ch

^h e-mail: Bogdan.Malaescu@cern.ch

likelihood. The curvature of the likelihood around its maximum is determined only by the expected fluctuations of the data and yields the statistical uncertainty of the measurement.¹ Systematic uncertainties are obtained by repeating the procedure with varied models, obtained from the variation of the auxiliary parameters within their uncertainty, one parameter at a time [6]. Each variation represents a given source of uncertainty. The corresponding uncertainties in the final result are usually uncorrelated by construction, and are summed in quadrature to obtain the total measurement uncertainty.

When using this method, different measurements of the same physical constants can be readily combined. When all uncertainties are Gaussian, the best linear unbiased estimate (BLUE) [7, 8] results from the analytical maximization of the joint likelihood of the input measurements, and unambiguously propagates the statistical and systematic uncertainties in the input measurements to the combined result.

An improved statistical method consists in parameterizing the model in terms of both the physical constants and the sources of uncertainty [9, 10], and has become a standard in LHC analysis. In this case, the maximum of the likelihood represents a global optimum for the physical constants and the uncertainty parameters, and determines their best values simultaneously. The curvature of the likelihood at its maximum reflects the fluctuations of the data and of the other sources of uncertainty, therefore giving the total uncertainty in the final result.

The determination of the statistical and systematic uncertainty components in numerical profile likelihood fits is the subject of the present note. Current practice universally employs so-called impacts [11–13], obtained as the quadratic difference between the total uncertainties of fits including or excluding given sources of uncertainty. But while impacts quantify the increase in the total uncertainty when including new systematic sources in a measurement, they cannot be interpreted as the contribution of these sources to the total uncertainty in the complete measurement. Impacts do not add up to the total uncertainty, and do not match usual uncertainty decomposition formulas [8] even when they should, i.e., when all uncertainties are genuinely Gaussian.

These statements are illustrated with a simple example in Sect. 2. Sections 3 and 4 summarize parameter estimation in the Gaussian approximation. Sources of uncertainty can be entirely encoded in the covariance matrix of the measurements (the “covariance representation”), or parameterized using nuisance parameters (the “nuisance parameter

representation”). The equivalence between the approaches is recalled, and a detailed discussion of the fit uncertainties and correlations is provided. A new and consistent method for the decomposition of uncertainties in profile likelihood fits is proposed in Sect. 5. The method is general, as it results from a Taylor expansion of the likelihood, and a proof that it yields consistent results in the Gaussian regime is given. The different approaches are illustrated in Sect. 6 with examples based on the Higgs and W -boson mass measurements and combinations, which are usually dominated by systematic uncertainties and where the present discussion is of particular relevance. Concluding remarks are presented in Sect. 7.

In the following, we understand the statistical uncertainty in its strict frequentist definition, i.e., the standard deviation of an estimator when the exact same experiment is repeated (with the same systematic uncertainties) on independent data samples of identical expected size. Similarly, a systematic uncertainty contribution should match the standard deviation of the estimator obtained under fluctuations of the corresponding source within its initial uncertainty. Measurements (physical parameters, cross sections, or bins of a measured distribution) and the corresponding predictions will be denoted as \vec{m} and \vec{t} , respectively, and labeled using Roman indices i, j, k . The predictions are functions of the physical constants to be determined, referred to as parameters of interest (POIs), denoted as $\vec{\theta}$ and labeled p, q . Sources of uncertainty are denoted as $\vec{\alpha}$, and their associated nuisance parameters (NPs), $\vec{\alpha}$, are labeled r, s, t .

2 Example: Higgs boson mass in the di-photon and four-lepton channels

Let us consider the first ATLAS Run 2 measurement of the Higgs boson mass, m_H , in the $H \rightarrow \gamma\gamma$ and $H \rightarrow 4\ell$ final states [14]. The measurement results in the $\gamma\gamma$ and 4ℓ channels have similar total uncertainty, but are unbalanced in the sense that the former benefits from a large data sample but has significant systematic uncertainties from the photon energy calibration, while the latter is limited to a smaller data sample but benefits from excellent calibration systematic uncertainties:

$$\begin{aligned} -m_{\gamma\gamma} &= 124.93 \pm 0.40 (\pm 0.21 \text{ (stat)} \pm 0.34 \text{ (syst)}) \text{ GeV}; \\ -m_{4\ell} &= 124.79 \pm 0.37 (\pm 0.36 \text{ (stat)} \pm 0.09 \text{ (syst)}) \text{ GeV}. \end{aligned}$$

The uncertainties in the $\gamma\gamma$ and 4ℓ measurements can be considered as entirely uncorrelated for this discussion. In the BLUE approach, the combined value and its uncertainty are then obtained considering the following log-likelihood:

$$-2 \ln \mathcal{L} = \sum_i \left(\frac{m_i - m_H}{\sigma_i} \right)^2, \quad (1)$$

¹ The curvature of the likelihood around its maximum only provides a lower bound on the standard deviation of the estimator in the general case (Cramér–Rao inequality). Much of the discussion in this paper will be about the maximum likelihood estimator, which is asymptotically efficient, i.e., for which the equality is reached.

where $i = \gamma\gamma, 4\ell$, and $\sigma_{\gamma\gamma}$ and $\sigma_{4\ell}$ are the total uncertainties in the $\gamma\gamma$ and 4ℓ channels, respectively. The combined value m_{cmb} and its total uncertainty σ_{cmb} are derived solving

$$\left. \frac{\partial \ln \mathcal{L}}{\partial m_H} \right|_{m_H=m_{\text{cmb}}} = 0, \quad \frac{1}{\sigma_{\text{cmb}}^2} = \left. \frac{\partial^2 \ln \mathcal{L}}{\partial m_H^2} \right|_{m_H=m_{\text{cmb}}} \quad (2)$$

The solutions can be written in terms of linear combinations of the input values and uncertainties:

$$m_{\text{cmb}} = \sum_i \lambda_i m_i, \quad \sigma_{\text{cmb}}^2 = \sum_i \lambda_i^2 \sigma_i^2 \quad (3)$$

with

$$\lambda_i = \frac{1/\sigma_i^2}{1/\sigma_{\gamma\gamma}^2 + 1/\sigma_{4\ell}^2}, \quad \lambda_{\gamma\gamma} + \lambda_{4\ell} = 1. \quad (4)$$

And the weights λ_i minimize the variance in the combined result, accounting for all sources of uncertainty in the input measurements. Since the total uncertainties have statistical and systematic components, i.e., $\sigma_i^2 = \sigma_{\text{stat},i}^2 + \sigma_{\text{syst},i}^2$, the corresponding contributions in the combined measurement are simply

$$\sigma_{\text{stat,cmb}}^2 = \sum_i \lambda_i^2 \sigma_{\text{stat},i}^2, \quad \sigma_{\text{syst,cmb}}^2 = \sum_i \lambda_i^2 \sigma_{\text{syst},i}^2. \quad (5)$$

In the profile likelihood (PL) approach, or nuisance-parameter representation, the corresponding likelihood reads

$$-2 \ln \mathcal{L} = \sum_i \left(\frac{m_i + \sum_r (\alpha_r - a_r) \Gamma_{ir} - m_H}{\sigma_{\text{stat},i}} \right)^2 + \sum_r (\alpha_r - a_r)^2, \quad (6)$$

where α_r is the nuisance parameter corresponding to the source of systematic uncertainty r , and Γ_{ir} its effect on the measurement in channel i . Knowledge of the systematic uncertainty r is obtained from an auxiliary measurement, of which the central value, sometimes called a global observable, is denoted as a_r . The parameters α_r and a_r are defined in units of the systematic uncertainty $\sigma_{\text{syst},r}$, and a_r is often conventionally set to 0. In this example, since the $\sigma_{\text{syst},r}$ are specific to each channel and do not generate correlations, $\Gamma_{ir} = \sigma_{\text{syst},r} \delta_{ir}$. The combined value m_{cmb} and its total uncertainty are obtained from the absolute maximum and second derivative of \mathcal{L} as above; in addition, the PL yields the estimated value for α_r . One finds that m_{cmb} and σ_{cmb} exactly match their counterparts from Eq. (3) (see also the discussion in Sect. 4).

In PL practice, however, the statistical uncertainty is usually obtained by fixing all nuisance parameters to their best-fit value (maximum likelihood estimator) $\hat{\alpha}_r$, maximizing the likelihood only with respect to the parameter of interest. With fixed α_r , the second derivative of Eq. (6) becomes equivalent

Table 1 Uncertainty components of m_H in the $\gamma\gamma$ and 4ℓ channels, and for the combined measurement. The combined uncertainties are given according to the BLUE result (Eq. (5)) and using impacts (Eq. (7))

Measurement	σ_{stat}	σ_{syst}	σ_{tot}
$\gamma\gamma$	0.21	0.34	0.40
4ℓ	0.36	0.09	0.37
Combined, decomposed	0.22	0.16	0.27
Combined, impacts	0.18	0.20	0.27

to that of Eq. (1), changing only σ_i for $\sigma_{\text{stat},i}$ in the denominator, giving

$$\sigma_{\text{stat,cmb}}^2 = \sum_i \lambda_i'^2 \sigma_{\text{stat},i}^2, \quad \lambda_i' = \frac{1/\sigma_{\text{stat},i}^2}{1/\sigma_{\text{stat},\gamma\gamma}^2 + 1/\sigma_{\text{stat},4\ell}^2} \quad (7)$$

which this time differ from Eqs. (4), (5): here, the coefficients λ_i' are calculated from the statistical uncertainties only, and the combined uncertainty is optimized for this case. The statistical uncertainty is thus underestimated relative to Eq. (3)). The systematic error, estimated from the quadratic subtraction between the total and statistical uncertainty estimate, is overestimated.

For completeness, numerical values are given in Table 1. The “impact” of a systematic uncertainty on a measurement with only statistical uncertainties differs from the contribution of this systematic uncertainty to the complete measurement. In the impact procedure, the estimated measurement statistical uncertainty is actually the total uncertainty of a measurement without systematic uncertainties, i.e., of a different measurement. In other words, it does not match the standard deviation of results obtained by repeating the same measurement, including systematic uncertainties, on independent data sets of the same expected size.

Finally, extrapolating the $\gamma\gamma$ and 4ℓ measurements to the large data sample limit, statistical uncertainties vanish, and the asymptotic combined uncertainty should intuitively be dominated by the 4ℓ channel and close to 0.09 GeV. A naive estimate based on impacts instead suggests an asymptotic uncertainty of 0.20 GeV.

We generalize this discussion in the following, and argue that a sensible uncertainty decomposition should match the one obtained from fits in the covariance representation, and can be also obtained simply in the context of the PL. The Higgs boson mass example is further discussed in Sect. 6.1.

3 Uncertainty decomposition in covariance representation

This section provides a short summary of standard results which can be found in the literature (see e.g. [15]). Gaussian

uncertainties are assumed throughout this section. The general form of Eq. (1) in the presence of an arbitrary number of measurements m_i and POIs $\vec{\theta}$ is

$$-2 \ln \mathcal{L}_{\text{cov}}(\vec{\theta}) = \sum_{i,j} \left(m_i - t_i(\vec{\theta}) \right) C_{ij}^{-1} \left(m_j - t_j(\vec{\theta}) \right), \quad (8)$$

where $t_i(\vec{\theta})$ are models for the m_i , and C is the total covariance of the measurements:

$$C_{ij} = V_{ij} + \sum_r \Gamma_{ir} \Gamma_{jr}, \quad (9)$$

where V_{ij} represents the statistical covariance, and the second term collects all sources of systematic uncertainties. In general, V_{ij} includes statistical correlations between the measurements, but is sometimes diagonal, in which case $V_{ij} = \sigma_i^2 \delta_{ij}$. Γ_{ir} represents the effect of systematic source r on measurement i (see Eq. (6)), and the outer product gives the corresponding covariance.

Imposing the restriction that the models t_i are linear functions of the parameters of interest, i.e., $t_i(\vec{\theta}) = t_{0,i} + \sum_p h_{ip} \theta_p$, according to the Gauss–Markov theorem (see e.g. Refs. [7, 16, 17]), the POI estimators with smallest variance are found by solving $\partial \ln \mathcal{L}_{\text{cov}} / \partial \theta_p |_{\vec{\theta} = \hat{\vec{\theta}}} = 0$, and the corresponding covariance is obtained from the matrix of second derivatives, $\partial^2 \ln \mathcal{L}_{\text{cov}} / \partial \theta_p \partial \theta_q |_{\vec{\theta} = \hat{\vec{\theta}}}$. The solutions are

$$\hat{\theta}_p = \sum_i \lambda_{pi} (m_i - t_{0,i}), \quad (10)$$

$$\text{cov}(\hat{\theta}_p, \hat{\theta}_q) = \sum_{i,j} \lambda_{pi} C_{ij} \lambda_{qj}, \quad (11)$$

where the weights λ_{pi} are given by

$$\lambda_{pi} = \sum_q \left(h^T \cdot S \cdot h \right)_{pq}^{-1} \cdot \left(h^T \cdot S \right)_{qi}, \quad (12)$$

$$S_{ij} = \sum_k V_{ik}^{-1} (\mathbb{I} - \Gamma \cdot Q)_{kj}, \quad (13)$$

$$Q_{ri} = \sum_s \left(\mathbb{I} + \Gamma^T V^{-1} \Gamma \right)_{rs}^{-1} (\Gamma^T V^{-1})_{si}. \quad (14)$$

In particular, using Eq. (9), the contribution to the uncertainties in the POIs of the statistical uncertainty in the measurements, and of each systematic source r , is given by

$$\text{cov}^{[\text{stat}]}(\hat{\theta}_p, \hat{\theta}_q) = \sum_{i,j} \lambda_{pi} V_{ij} \lambda_{qj}, \quad (15)$$

$$\text{cov}^{[r]}(\hat{\theta}_p, \hat{\theta}_q) = \sum_{i,j} \lambda_{pi} (\Gamma_{ir} \Gamma_{jr}) \lambda_{qj}. \quad (16)$$

We note that the BLUE averaging procedure, i.e., the unbiased² linear averaging of measurements of a common phys-

ical quantity, is just a special case of Eq. (8) where the measurements are direct estimators of the POIs. In the case of a single POI, $t_i = \theta$ ($t_{0,i} = 0$, $h = 1$).

A detailed discussion of template fits and of the propagation of fit uncertainties was recently given in Ref. [20]. While the above summary is restricted to linear fits with constant uncertainties, Ref. [20] also addresses nonlinear effects and uncertainties that scale with the measured quantity, i.e., $\Gamma_{ir} \propto m_i$.

4 Equivalence between the covariance and nuisance parameter representations

Similarly, still assuming Gaussian uncertainties, the general form of Eq. (6) is

$$\begin{aligned} -2 \ln \mathcal{L}_{\text{NP}}(\vec{\theta}, \vec{\alpha}) = & \sum_{i,j} \left(m_i - t_i(\vec{\theta}) - \sum_r \Gamma_{ir} (\alpha_r - a_r) \right) V_{ij}^{-1} \\ & \times \left(m_j - t_j(\vec{\theta}) - \sum_s \Gamma_{js} (\alpha_s - a_s) \right) \\ & + \sum_r (\alpha_r - a_r)^2. \end{aligned} \quad (17)$$

The optimum of \mathcal{L}_{NP} can be found by first minimizing Eq. (17) over $\vec{\alpha}$, for fixed $\vec{\theta}$ (i.e., *profiling* the nuisance parameters $\vec{\alpha}$); substituting the result into Eq. (17) (thus obtaining the *profile likelihood* $\ln \mathcal{L}_{\text{NP}}(\vec{\theta}, \hat{\vec{\alpha}}(\vec{\theta}))$); and minimizing over $\vec{\theta}$. The profiled nuisance parameters are given by

$$\hat{\alpha}_r(\vec{\theta}) = \sum_i Q_{ri} \left(m_i - t_i(\vec{\theta}) \right) + a_r, \quad (18)$$

where Q_{ri} was defined in Eq. (14). The expression for the covariance is

$$\text{cov}(\hat{\alpha}_r, \hat{\alpha}_s)(\vec{\theta}) = \left(\mathbb{I} + \Gamma^T V^{-1} \Gamma \right)_{rs}^{-1}. \quad (19)$$

Substituting Eq. (18) back into Eq. (17), and after some algebra, the profile likelihood can be written as

$$-2 \ln \mathcal{L}_{\text{NP}}(\vec{\theta}, \hat{\vec{\alpha}}(\vec{\theta})) = \sum_{i,j} \left(m_i - t_i(\vec{\theta}) \right) S_{ij} \left(m_j - t_j(\vec{\theta}) \right), \quad (20)$$

Footnote 2 continued

of the input covariance matrix (see e.g. Chapter 7 of Ref. [16]). Indeed, such covariances generally carry uncertainties themselves, because the size of the systematic uncertainties and their correlations are never really measured, but rather estimated. The existence and relevance of such uncertainties on the uncertainties and on their correlations has been noted in the context, for example, of α_s fits from jet cross section data [18]. See also the related work in Ref. [19] concerning the uncertainties on uncertainties.

² The word “unbiased” employed here needs to be interpreted with care, as it actually involves several implicit assumptions about the knowledge

where S_{ij} was defined in Eq. (13). Moreover, it can be verified that

$$\sum_k V_{ik}^{-1} (\mathbb{I} - \Gamma \cdot Q)_{kj} = \left(V_{ij} + \sum_r \Gamma_{ir} \Gamma_{jr} \right)^{-1}, \text{ i.e.} \quad (21)$$

$$S_{ij} = C_{ij}^{-1}, \quad (22)$$

so that Eqs. (20) and (8) are in fact identical. In other words, $\mathcal{L}_{\text{cov}}(\vec{\theta})$, in covariance representation, can be seen as the result of maximizing $\mathcal{L}_{\text{NP}}(\vec{\theta}, \vec{\alpha})$ over $\vec{\alpha}$ for fixed $\vec{\theta}$: it is the profile likelihood. Consequently, the best values for the POIs are still given by Eq. (10) and their uncertainties by Eq. (11), and the error decomposition of Sect. 3 applies.

The observation above is not new, and has to the authors' knowledge been discussed in Refs. [21–27] for diagonal statistical uncertainties, and in Refs. [28, 29] in the general case. It is also briefly mentioned in Ref. [20]. The equivalence between the covariance and nuisance parameter representations is recalled here to insist that profile likelihood fits should obey the uncertainty decomposition usual from fits in the covariance representation.

For any value of $\vec{\theta}$, the estimators of the nuisance parameters and their covariance are given by Eqs. (18) and (19). The estimator $\hat{\alpha}$ is given by the product of the differences between the measurements and the model, $m_i - t_i(\vec{\theta})$, and a factor Q determined only from the initial systematic and experimental uncertainties. This factor can be calculated from the basic inputs to the fit. Nuisance parameter pulls ($\hat{\alpha}_r$) and constraints ($\sqrt{\text{cov}(\hat{\alpha}_r, \hat{\alpha}_r)}$) can thus also be calculated *a posteriori* in the context of a POI-only fit in covariance representation, without explicitly introducing $\vec{\alpha}$, \vec{a} in the expression of the likelihood, from the same inputs as those defining C .

This procedure can be repeated, first minimizing over $\vec{\theta}$ for given $\vec{\alpha}$, substituting the result into Eq. (17), and minimizing the result over the nuisance parameter $\vec{\alpha}$. This yields the NP covariance matrix elements as

$$\text{cov}(\hat{\alpha}_r, \hat{\alpha}_s) = \left[\mathbb{I} + (\zeta \cdot \Gamma)^T V^{-1} (\zeta \cdot \Gamma) \right]_{rs}^{-1}, \quad (23)$$

with

$$\zeta_{ij} = \sum_p h_{ip} \rho_{pj} - \delta_{ij}, \quad (24)$$

$$\rho_{pj} = \sum_q (h^T \cdot V^{-1} \cdot h)_{pq}^{-1} (h^T \cdot V^{-1})_{qj}, \quad (25)$$

while the covariance between the NPs and POI is given by

$$\text{cov}(\hat{\alpha}_r, \hat{\theta}_p) = - \sum_s \left[\mathbb{I} + (\zeta \cdot \Gamma)^T V^{-1} (\zeta \cdot \Gamma) \right]_{rs}^{-1} (\rho \cdot \Gamma)_{ps}. \quad (26)$$

Equations (11), (23), and (26) determine the full covariance matrix of the fitted parameters.

Importantly, Eq. (26) can be further simplified to

$$\text{cov}(\hat{\alpha}_r, \hat{\theta}_p) = - \sum_i \lambda_{pi} \Gamma_{ir}, \quad (27)$$

which directly provides the systematic uncertainty decomposition. The inner product of Eq. (27) with itself gives the systematic covariance, Eq. (16), and the statistical uncertainty can be obtained by subtracting the result in quadrature from the total uncertainty in $\hat{\theta}_p$. In other words, the contribution of every systematic source to the total uncertainty is directly given by the covariance between the corresponding NP and the POI.

5 Uncertainty decomposition from shifted observables

While it is a common and relevant approximation, probability models are in general not based on Gaussian uncertainty distributions. Small samples are treated using the Poisson distribution, and the constraint terms associated to nuisance parameters can assume arbitrary forms. The best-fit values of the POI are however always functions of the measurements and the central values of the auxiliary measurements, i.e., $\hat{\theta}_p = \hat{\theta}_p(\vec{m}, \vec{a})$. Assuming no correlations between these observables, the uncertainty in $\hat{\theta}_p$ then follows from linear error propagation:

$$\text{cov}(\hat{\theta}_p, \hat{\theta}_p) = \sum_i \left(\frac{\partial \hat{\theta}_p}{\partial m_i} \cdot \sigma_i \right)^2 + \sum_r \left(\frac{\partial \hat{\theta}_p}{\partial a_r} \cdot 1 \right)^2, \quad (28)$$

where σ_i is the uncertainty in m_i , the uncertainty in a_r is 1 by definition of a_r and α_r (Sect. 2), and $\frac{\partial \hat{\theta}_p}{\partial m_i}$, $\frac{\partial \hat{\theta}_p}{\partial a_r}$ are the sensitivities of the fit result to these observables. The first sum in Eq. (28) reflects the fluctuations of the measurements, i.e., the statistical uncertainty (each term of the sum represents the contribution of a given m_i , measurement, or bin), and the second sum collects the contributions of the systematic uncertainties.

The contribution of a given source of uncertainty can thus be assessed by varying the corresponding measurement or global observable by one standard deviation in the expression of the likelihood, and repeating the fit otherwise unchanged. The corresponding uncertainty is obtained from the difference between the values of $\hat{\theta}_p$ in the varied and nominal fits.

This statement can be verified explicitly for the Gaussian, linear fits discussed in the previous section. Now allowing for correlations between the measurements, varying m_k within its uncertainty yields the following likelihood:

$$\begin{aligned}
& -2 \ln \mathcal{L}_{m_k}(\vec{\theta}, \vec{\alpha}) \\
& = \sum_{i,j} \left(m_i + L_{ik} - t_i(\vec{\theta}) - \sum_r \Gamma_{ri}(\alpha_r - a_r) \right) V_{ij}^{-1} \\
& \quad \times \left(m_j + L_{jk} - t_j(\vec{\theta}) - \sum_s \Gamma_{sj}(\alpha_s - a_s) \right) \\
& \quad + \sum_r (\alpha_r - a_r)^2, \tag{29}
\end{aligned}$$

where L results from the Cholesky decomposition $L^T L = V$ and represents the correlated effect on all measurements m_i of varying m_k within its uncertainty. In the case of uncorrelated measurements, $L_{ik} = \sigma_i \delta_{ik}$, and only m_k is varied, as in Eq. (28). After minimization, the difference between the varied and nominal fit results is

$$\Delta \hat{\theta}_p^{[m_k]} \equiv \hat{\theta}_p^{[m_k]} - \hat{\theta}_p = \sum_i \lambda_{pi} L_{ik}. \tag{30}$$

Similarly, the uncertainty in a_t can be obtained from the following likelihood:

$$\begin{aligned}
& -2 \ln \mathcal{L}_{a_t}(\vec{\theta}, \vec{\alpha}) \\
& = \sum_{i,j} \left(m_i - t_i(\vec{\theta}) - \sum_r \Gamma_{ri}(\alpha_r - a_r) \right) V_{ij}^{-1} \\
& \quad \times \left(m_j - t_j(\vec{\theta}) - \sum_s \Gamma_{sj}(\alpha_s - a_s) \right) \\
& \quad + \sum_r (\alpha_r - a_r - \delta_{rt})^2, \tag{31}
\end{aligned}$$

resulting in

$$\Delta \hat{\theta}_p^{[a_t]} \equiv \hat{\theta}_p^{[a_t]} - \hat{\theta}_p = - \sum_i \lambda_{pi} \Gamma_{it}, \tag{32}$$

as in Eq. (27). The differences between the varied and nominal values of $\hat{\theta}_p$ match the expressions obtained above for the corresponding uncertainties. In particular,

$$\sum_k \Delta \hat{\theta}_p^{[m_k]} \Delta \hat{\theta}_q^{[m_k]} = \sum_{i,j} \lambda_{pi} V_{ij} \lambda_{qj} \tag{33}$$

reproduces the total statistical covariance in Eq. (15), and

$$\Delta \hat{\theta}_p^{[a_t]} \Delta \hat{\theta}_q^{[a_t]} = \sum_{i,j} \lambda_{pi} (\Gamma_{it} \Gamma_{jt}) \lambda_{qj} \tag{34}$$

is the contribution of systematic source t to the systematic covariance in Eq. (16).

As in Sect. 4, the total uncertainty in the NPs can be obtained by minimizing the likelihood with respect to $\vec{\theta}$ for fixed $\vec{\alpha}$, replacing $\vec{\theta}$ by its expression, and minimizing the result with respect to $\vec{\alpha}$. The contribution of the measurements to the uncertainty in $\vec{\alpha}$ is

$$\Delta \hat{\alpha}_r^{[m_k]} \equiv \hat{\alpha}_r^{[m_k]} - \hat{\alpha}_r = \sum_i \tilde{Q}_{ri} L_{ik}, \tag{35}$$

where

$$\begin{aligned}
\tilde{Q}_{ri} & = - \sum_s \left[\mathbb{I} + (\zeta \cdot \Gamma)^T V^{-1} (\zeta \cdot \Gamma) \right]_{rs}^{-1} \\
& \quad \times \left[(\zeta \cdot \Gamma)^T \cdot V^{-1} \right]_{si}; \tag{36}
\end{aligned}$$

and the systematic contributions are given by

$$\Delta \hat{\alpha}_r^{[a_t]} \equiv \hat{\alpha}_r^{[a_t]} - \hat{\alpha}_r = \left[\mathbb{I} + (\zeta \cdot \Gamma)^T V^{-1} (\zeta \cdot \Gamma) \right]_{rt}^{-1}. \tag{37}$$

Summing Eqs. (35) and (37) in quadrature recovers the total NP covariance matrix in Eq. (23), as expected.

Finally, the covariance between the NPs and POIs can be obtained analytically by summing the products of the corresponding offsets, obtained from statistical and systematic variations, that is,

$$\begin{aligned}
& \sum_k \Delta \alpha_r^{[m_k]} \Delta \theta_p^{[m_k]} + \sum_t \Delta \alpha_r^{[a_t]} \Delta \theta_p^{[a_t]} \\
& = - \sum_s \left[\mathbb{I} + (\zeta \cdot \Gamma)^T V^{-1} (\zeta \cdot \Gamma) \right]_{rs}^{-1} (\rho \cdot \Gamma)_{ps}, \tag{38}
\end{aligned}$$

which again matches the expression for $\text{cov}(\hat{\alpha}_r, \hat{\theta}_p)$ in Eq. (26).

The identities (33), (34), (37), and (38) can be obtained analytically only for linear fits with Gaussian uncertainties, but the uncertainty decomposition through fits with shifted observables only assumes the Taylor expansion of Eq. (28) and is therefore general. The covariance and NP representations are equivalent for Gaussian fits, but this equivalence breaks down for fits with non-Gaussian uncertainty distributions, and curvatures at the maximum of the likelihood no longer provide reliable estimates for the variance of the parameters. Such fits can however still rely on Eq. (28) to obtain a consistent uncertainty decomposition where each component directly reflects the propagation of the uncertainty in the corresponding source. In this way, uncertainty components preserve a universal meaning, regardless of the statistical method used for a given measurement.

In practice, the uncertainty can be propagated using one-standard-deviation shifts in m and a as above, or using the Monte Carlo error propagation method, where m or a are randomized within their respective probability density functions, and the corresponding uncertainty in the measurement is determined from the variance of the fit results.³ The latter method makes the correspondence between uncertainty contributions and the effect of fluctuations of the corresponding sources (cf. Sect. 1) explicit. It is also more general, and

³ In order to perform the uncertainty propagation in a linear regime, one can also apply shifts of less than one standard deviation, followed by a rescaling of the resulting propagated uncertainty. For effectively probing possible nonlinear effects impacting the tails of the uncertainty distributions, one can perform a scan of the shifts by e.g. 1, 2, ..., 5 standard deviations.

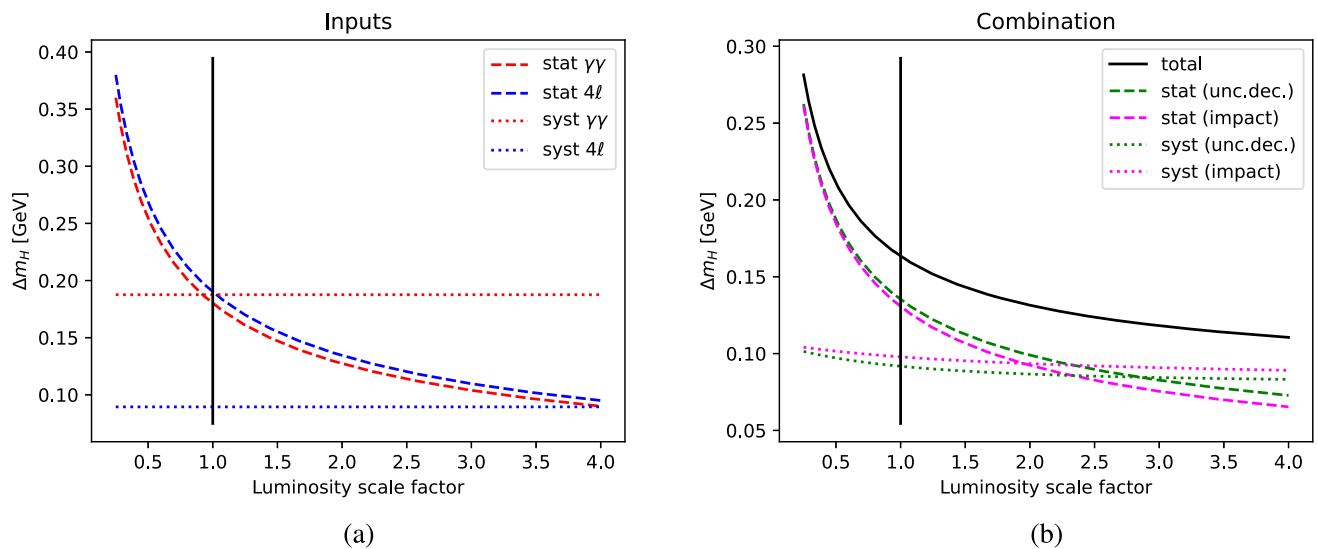


Fig. 1 Uncertainty decomposition as a function of a luminosity scaling factor, using CMS Run 2 results [31]. Left: size of the statistical (stat) and systematic (syst) uncertainties for $\gamma\gamma$ and 4ℓ . Right: decomposition of uncertainties on the combination using either the uncertainty decomposition or impacts approach

gives more precise results in the case of significant asymmetries or tails in the uncertainty distributions. In addition, it can be more efficient when simultaneously estimating the variance contributed by a large group of sources of uncertainty. Similarly, the present method can be generalized to unbinned measurements using data resampling techniques for the extraction of statistical uncertainty components [30].

6 Examples

6.1 Combination of two measurements

Let us consider again the concrete case of the Higgs boson mass m_H described in Sect. 2, which will serve as a simple example with only one parameter of interest (m_H) and two measurements. We will further assume that both the statistical and systematic uncertainties are uncorrelated between the two channels, which is not unreasonable given that they correspond to different events and that the dominant sources of systematic uncertainty are indeed uncorrelated. We will take numerical values from the actual ATLAS [14] and CMS [31] Run 1 and Run 2 measurements, as well as from an imaginary case exaggerating the numeric features of the ATLAS Run 2 measurement.

For each case, the decomposition of uncertainties between statistical and systematic components will be compared between the two approaches—uncertainty decomposition and impacts. In addition, this is done as a function of a luminosity factor k , which is used to scale the statistical uncertainty of the inputs by $1/\sqrt{k}$ (while systematic uncertainties are kept unchanged). The published results in the example

under consideration are for $k = 1$. Though not shown on the plots, we have also checked numerically that the uncertainty decomposition (as usually done in covariance representation methods or BLUE) can be reproduced from a profile likelihood fit with shifted observables (Sect. 4), while the impacts (as usually done in profile likelihood fits) can also be recovered from the BLUE approach, simply by using the statistical uncertainties alone to compute the combination weights λ'_i as in Eq. (7) (i.e., repeating the combination without systematic uncertainties). In addition, both approaches have been checked to yield to the same total uncertainty in all cases.

CMS results We first study the combination of CMS Run 2 results [31]: $\text{stat}_{\gamma\gamma} = 0.18 \text{ GeV}$, $\text{syst}_{\gamma\gamma} = 0.19 \text{ GeV}$; $\text{stat}_{4\ell} = 0.19 \text{ GeV}$, $\text{syst}_{4\ell} = 0.09 \text{ GeV}$. The results of our toy combination are shown in Fig. 1. This figure, as well as the following ones, comprises two panels: the inputs to the combination on the left, and statistical and systematic uncertainties as obtained in either the uncertainty decomposition or impact approaches on the right. The actual published numbers [31] correspond to $k = 1$ (black vertical line).

With this first simple case, where the two measurements have relatively comparable uncertainties, little difference is found between the two approaches, though the uncertainty decomposition gives a larger statistical uncertainty than the impact one, as expected. The difference becomes larger for higher values of the luminosity factor.

ATLAS results We are now considering the ATLAS Run 2 results [14]: $\text{stat}_{\gamma\gamma} = 0.21 \text{ GeV}$, $\text{syst}_{\gamma\gamma} = 0.34 \text{ GeV}$; $\text{stat}_{4\ell} = 0.36 \text{ GeV}$, $\text{syst}_{4\ell} = 0.09 \text{ GeV}$. As shown in Fig. 2, differences between the two uncertainty decompositions are now more evident, already for the nominal uncertainty but

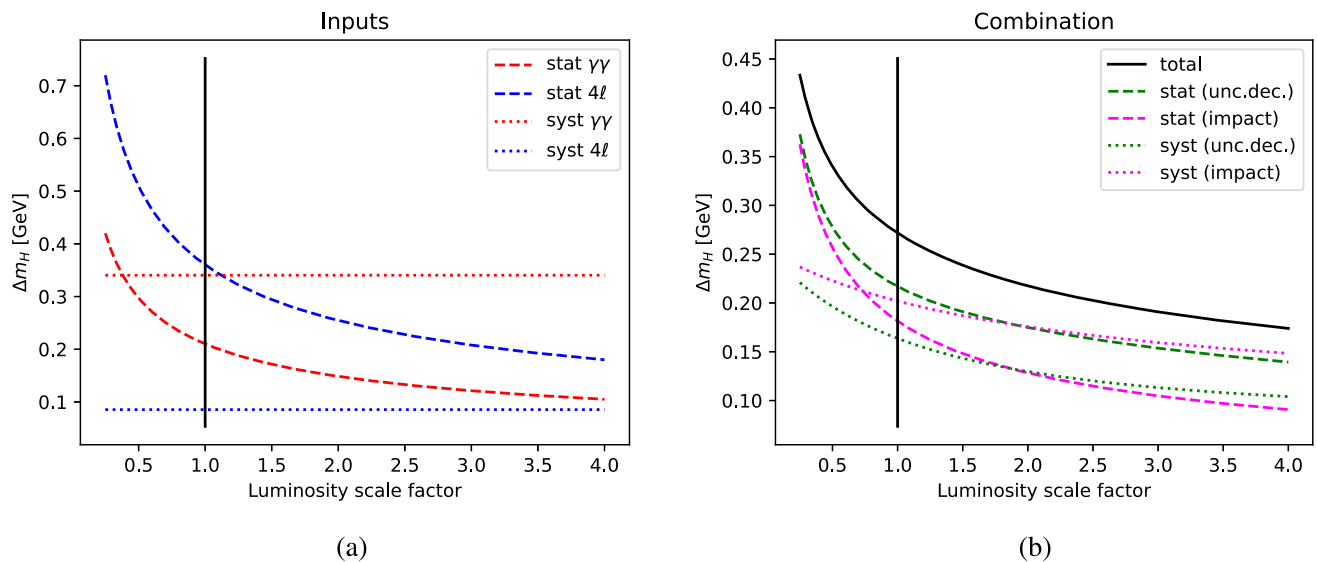


Fig. 2 Uncertainty decomposition as a function of a luminosity scaling factor, using ATLAS Run 2 results [14]. Left: size of the statistical (stat) and systematic (syst) uncertainties for $\gamma\gamma$ and 4ℓ . Right: decom-

position of uncertainties on the combination using either the uncertainty decomposition or impacts approach

even more when extrapolating to larger luminosities (smaller statistical uncertainties). Again, the uncertainty decomposition gives a larger statistical uncertainty than the impact one.

Imaginary extreme case Finally, we consider an extreme case, such that $\text{stat}_{\gamma\gamma} = 0.1 \text{ GeV}$, $\text{syst}_{\gamma\gamma} = 0.5 \text{ GeV}$; $\text{stat}_{4\ell} = 0.5 \text{ GeV}$, $\text{syst}_{4\ell} = 0.1 \text{ GeV}$, exaggerating the features of the ATLAS combination (i.e., combining a statistically dominated measurement with a systematically limited one). Dramatic differences between the two approaches for uncertainty decomposition are observed in Fig. 3: for the nominal luminosity, while uncertainty decomposition reports equal statistical and systematic uncertainties, the impacts are dominated by the systematic uncertainty.

6.2 W-boson mass fits

The uncertainty decomposition discussed above is further illustrated with a toy measurement of the W-boson mass using pseudo-data, where the results obtained from the profile likelihood fit and from the analytical calculation are compared. Since the measurement of W mass is a typical shape analysis, in which the fit to the distributions is parameterized by both POI and NPs, the conclusions drawn from this example can in principle be generalized to all kinds of shape analyses. While the effect of varying the W mass is parameterized by the POI, three representative systematic sources of a W mass measurement at hadron colliders [32–35] are parameterized by NPs in the probability model: the lepton momentum scale uncertainty, the hadronic recoil (HR) resolution uncertainty, and the p_T^W modeling uncertainty. The W mass is extracted from the p_T^ℓ or m_T spectra, since measure-

ments based on these two distributions have very different sensitivities to certain types of systematic uncertainties.

Simulation

The signal process under consideration is the charged-current Drell–Yan process [36] $pp \rightarrow W^- \rightarrow \mu^- \nu$ at a center-of-mass energy of $\sqrt{s} = 13 \text{ TeV}$, generated using Madgraph, with initial and final-state corrections obtained using Pythia8 [37,38]. Detailed information regarding the event generation is listed in Table 2.

Kinematic distributions for different values of the W mass are obtained in simulation via Breit–Wigner reweighting [39]. The systematic variations of p_T^W are implemented using a linear reweighting as a function of p_T^W before event selection, then taking only the shape effect on the underlying p_T^W spectrum.

At the reconstruction level, the p_T of the bare muon is smeared by 2% following a Gaussian distribution. A source of systematic uncertainty in the calibration of the muon momentum scale is considered. The hadronic recoil \vec{u}_T is taken to be the opposite of \vec{p}_T^W and smeared by a constant 6 GeV in both directions of the transverse plane. The second source of experimental systematic uncertainty is taken to be the uncertainty in the calibration of the hadronic recoil resolution. The information about the W mass templates and the systematic variation is summarized in Table 3.

Both the detector smearing and the event selections listed in Table 4 are chosen to be similar to those of a realistic W mass measurement. The reconstructed muon p_T and m_T spectra in the fit range after the event selection are shown

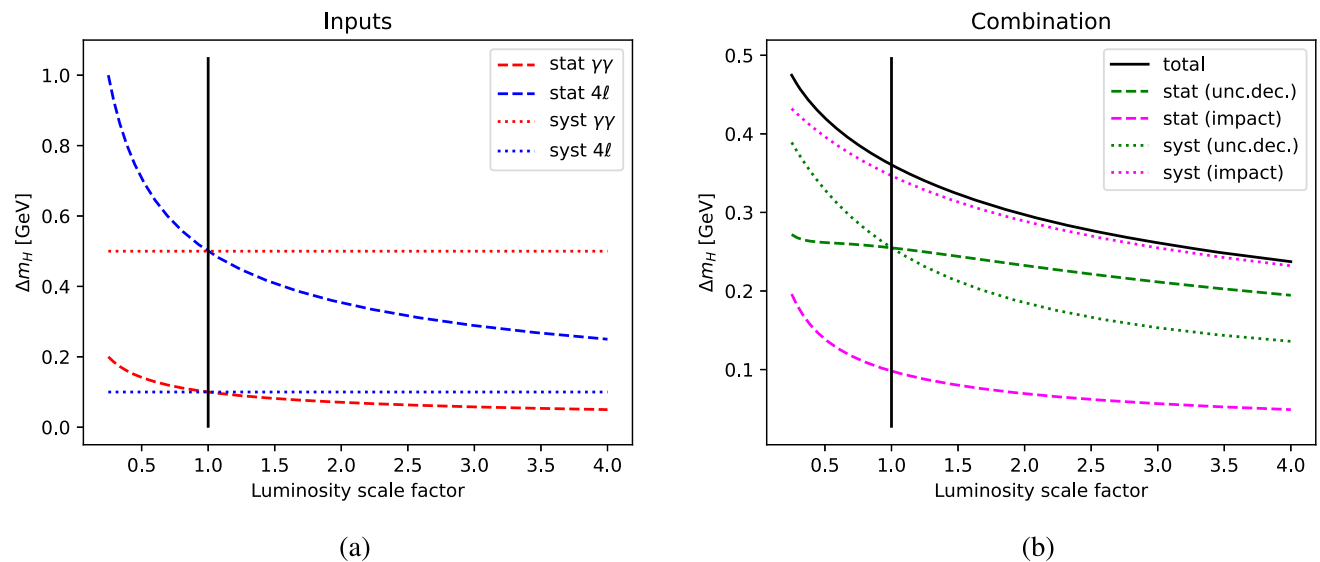


Fig. 3 Uncertainty decomposition as a function of a luminosity scaling factor, using $\text{stat}_{\gamma\gamma} = 0.1 \text{ GeV}$, $\text{syst}_{\gamma\gamma} = 0.5 \text{ GeV}$; $\text{stat}_{4\ell} = 0.5 \text{ GeV}$, $\text{syst}_{4\ell} = 0.1 \text{ GeV}$. Left: size of the statistical (stat) and systematic

(syst) uncertainties for $\gamma\gamma$ and 4ℓ . Right: decomposition of uncertainties on the combination using the uncertainty decomposition or impact approach

Table 2 Madgraph+Pythia8 [37,38] event generation for MC samples. Events with an off-shell boson are excluded in the event generation at the parton level, leading to a total cross section of 6543 pb

Event generator $pp \rightarrow W^- \rightarrow \mu^- \nu_\mu$ at $\sqrt{s}=13 \text{ [TeV]}$	
Number of events	10,000,000
Matrix elements	Madgraph at LO
Input m_W	80.419 [GeV]
Input Γ_W	2.0476 [GeV]
Parton shower & QED FSR	Pythia8

Table 3 W mass templates and systematic variations for the Madgraph+Pythia8 samples

Templates and systematic variations	
W mass templates	$\pm 50 \text{ MeV}$ by Breit–Wigner reweighting
μ calib.	Muon momentum scale $\pm 0.5\%$
HR calib.	Recoil resolutions $\pm 5\%$
p_T^W model	$w(p_T^W) = 0.96 + 8 \cdot 10^{-4} \cdot p_T^W \text{ [GeV]}$
(Reweighting)	

in Fig. 4, along with the relevant templates and systematic variations.

Uncertainty decomposition

The profile likelihood fit is performed using HistFactory [40] and RooFit [41]. Its output includes the fitted central values and uncertainties for all the free parameters. The uncertainty

Table 4 Detector smearing and event selection for Madgraph+Pythia8 samples. The cut-flow efficiency of the event selection is about 29%

Detector smearing	
Lepton p_T resolution	2%
Nominal recoil resolutions	6 [GeV]
Event selection	
η_ℓ selection	$[-2.5, 2.5]$
p_T^ℓ selection	$> 25 \text{ [GeV]}$
E_T^{miss} selection	$> 25 \text{ [GeV]}$
m_T selection	$> 50 \text{ [GeV]}$
u_T selection	$< 25 \text{ [GeV]}$

components of the profile likelihood fit results are obtained by repeating the fit to bootstrap samples obtained by resampling the pseudo data used to compute the results, or those of the central values of the auxiliary measurements; then, computing the spread of offsets in the POI, the analytical solution of the fit can be calculated following the procedures in Sect. 5. For this exercise, the pseudo-data are chosen to be the nominal simulation, but with the statistical power of the data. The effect of changing the luminosity scale factor is emulated by repeating the fit with an overall factor multiplied by all the reconstructed distributions. The setups of the fits for the validation are summarized in Table 5.

Figures 5 and 6 present the uncertainty decomposition as a function of a luminosity scale factor used to scale the statistical precision of the simulated sample. The error bars for the uncertainty decomposition for the profile likelihood fit reflect the limited number of toys. In general, the uncertainty

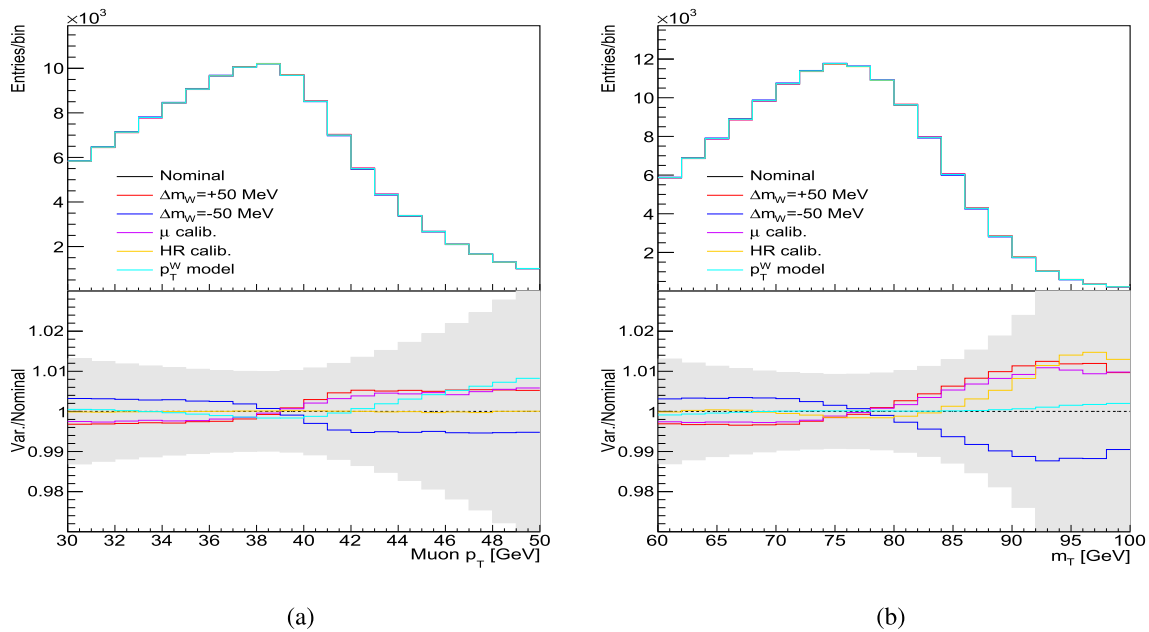


Fig. 4 Reconstructed muon p_T and m_T distributions of the Madgraph+Pythia8 samples. Top: kinematic spectra. Bottom: the variation to nominal ratio with statistical uncertainty indicated by the error band

Table 5 Configuration of the m_W fits. The luminosity scale factor of 1.0 corresponds to $76.42 \text{ [pb}^{-1}\text{]}$

Nuisance Parameters	Full model : μ calib, HR calib, p_T^W model Stat only (for impacts) : None
Luminosity factors	0.1, 0.4, 0.7, 1.0, 2.0, 5.0, 10.0
Fit ranges	$30 < p_T^\ell < 50 \text{ [GeV]}$ $60 < m_T < 100 \text{ [GeV]}$

components derived from the numerical profile likelihood fit and the analytical solution match each other within the error bars. The discrepancy at certain points can be assigned to the numerical stability of the PL fit, which shows up when the uncertainty components becomes too small (typically $< 2 \text{ MeV}$). The uncertainty decomposition is summarized in Table 6, where the total uncertainty is broken down into data statistic and total systematic uncertainties using the shifted observable method, and compared with the results using the conventional impact approach for PL fit. With 10 times higher luminosity, the statistical uncertainty of the impact approach decreases by exactly a factor of $\sqrt{10}$, while that of the shifted observable approach introduced in this study decreases more slowly.

Table 7 shows the analytical systematic uncertainty decomposition for the m_T and p_T^ℓ fits with nominal luminosity, together with the NP-POI covariance matrix elements obtained from the numerical profile likelihood fit. This confirms that the systematic uncertainty components can be directly read from the PL fit covariance matrix, as discussed

around Eq. (27). Finally, Fig. 7 compares the post-fit NP uncertainties between the numerical profile likelihood fit and the analytical calculation. The two methods agree at the 0.1 per-mil level.

6.3 Use of decomposed uncertainties in subsequent fits or combinations

Uncertainty decompositions obtained with the present method are meaningful only if the results can be used consistently in downstream applications, such as measurement combinations or interpretation fits in terms of specific physics models. In particular, uncertainty components that are common to several measurements generate correlations which should be properly evaluated. This happens when measurements are statistically correlated or when they are impacted by shared systematic uncertainties.

As a final validation of the proposed method, we test the combination of profile likelihood fits of the same observable. Such a combination can be performed either using the decomposed uncertainties, or in terms of the PL fit outputs, i.e., the fitted values of the POIs and NPs and their covariance matrix.

The combination is performed starting from Eq. (8), which as noted in Sect. 3 can be applied to linear measurement averaging by adapting the definition of $t(\vec{\theta})$. In the case of a single combined parameter, $t_i = \theta$, for a simultaneous combination of several parameters, $t_i = \sum_p U_{ip} \theta_p$, where U_{ip} is 1 when measurement i is an estimator of POI p , and 0 otherwise [8].

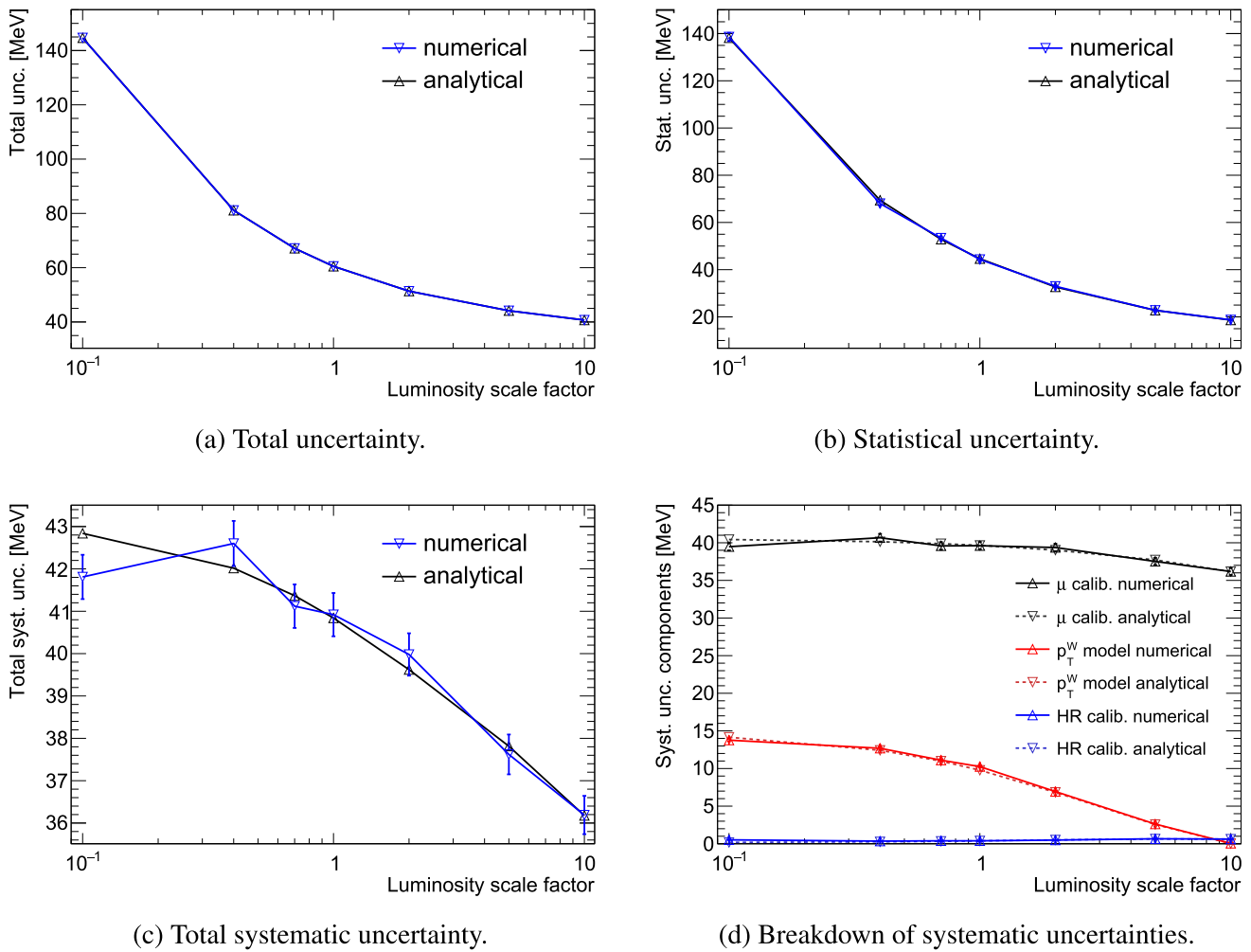


Fig. 5 Uncertainty decomposition for the muon p_T fit compared between the numerical and the analytical PL fit. The total systematic uncertainty of the profile likelihood fit is the quadratic sum of the three components

This gives

$$-2 \ln \mathcal{L}_{\text{cmb}}(\vec{\theta}) = \sum_{i,j} \left(m_i - \sum_p U_{ip} \theta_p \right) C_{ij}^{-1} \times \left(m_j - \sum_p U_{jp} \theta_p \right), \quad (39)$$

which can be solved as in Sect. 3.

As an illustration, we use the m_W fits using the p_T^ℓ and m_T distributions described in the previous section. In the case of a combination based on the uncertainty decomposition, there are two measurements (the POIs of the p_T^ℓ and m_T fits), one combined value, and the covariance C is a 2×2 matrix constructed from the decomposed uncertainties using Eq. (9).

For a combination based on the PL fit outputs, there are in this example eight measurements (one POI and three NPs in the p_T^ℓ and m_T fits), four combined parameters, and C is an 8×8 matrix. The diagonal 4×4 blocks are the post-fit covariance matrices of each fit (p_T^ℓ and m_T). The off-

diagonal blocks reflect systematic and/or statistical correlations between the p_T^ℓ and m_T fits, and can be obtained analytically following the methods of Sect. 5. For two fits f_1 and f_2 , the covariance matrix elements are

$$\begin{aligned} \text{cov}(\theta_p^{f_1}, \theta_q^{f_2}) &= \sum_k \Delta \theta_p^{[m_k], f_1} \Delta \theta_q^{[m_k], f_2} \\ &\quad + \sum_t \Delta \theta_p^{[a_t], f_1} \Delta \theta_q^{[a_t], f_2} \\ \text{cov}(\alpha_r^{f_1}, \alpha_s^{f_2}) &= \sum_k \Delta \alpha_r^{[m_k], f_1} \Delta \alpha_s^{[m_k], f_2} \\ &\quad + \sum_t \Delta \alpha_r^{[a_t], f_1} \Delta \alpha_s^{[a_t], f_2} \\ \text{cov}(\alpha_r^{f_1}, \theta_p^{f_2}) &= \sum_k \Delta \alpha_r^{[m_k], f_1} \Delta \theta_p^{[m_k], f_2} \\ &\quad + \sum_t \Delta \alpha_r^{[a_t], f_1} \Delta \theta_p^{[a_t], f_2} \\ \text{cov}(\theta_p^{f_1}, \alpha_r^{f_2}) &= \sum_k \Delta \theta_p^{[m_k], f_1} \Delta \alpha_r^{[m_k], f_2} \end{aligned}$$

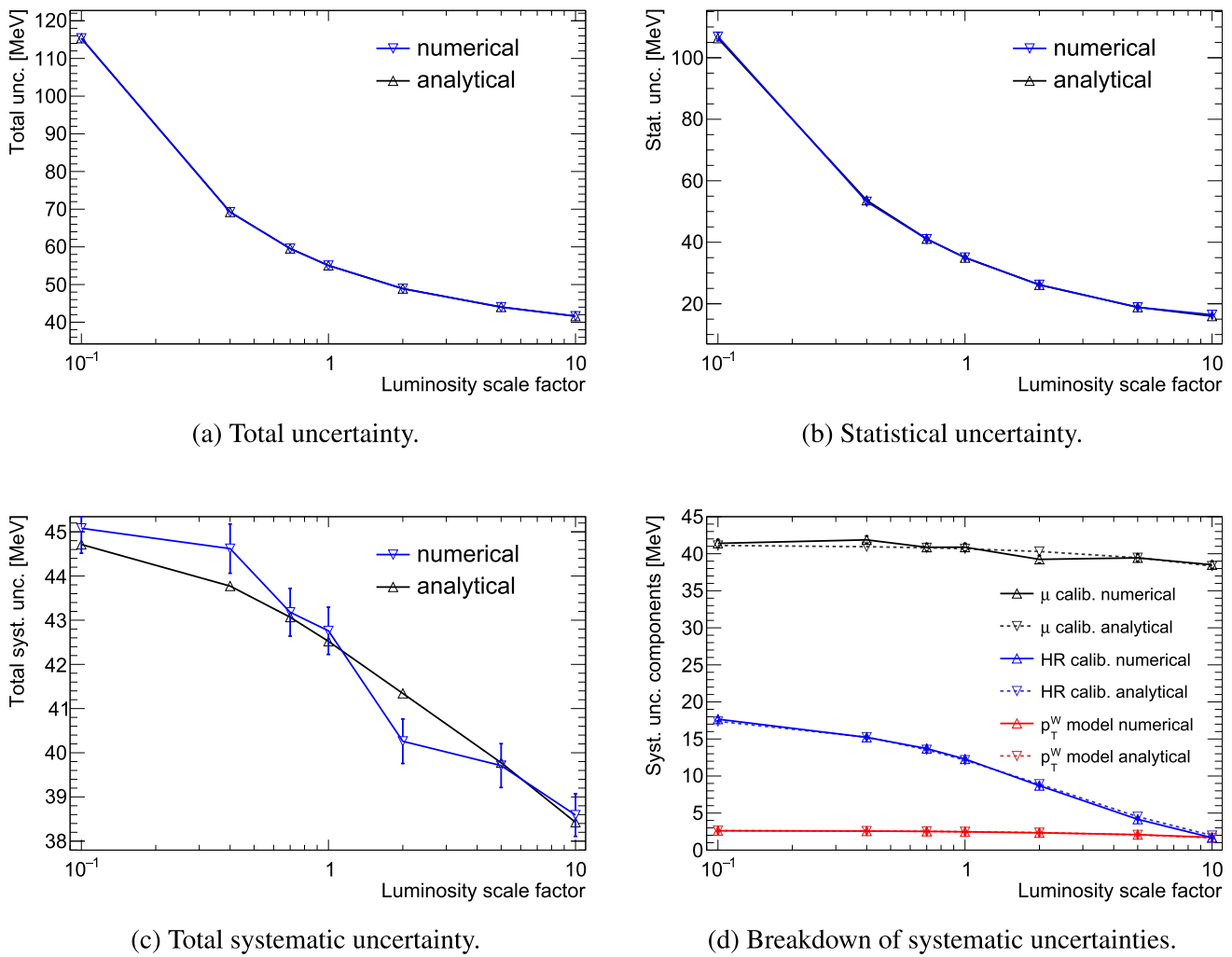


Fig. 6 Uncertainty decomposition for the m_T fit compared between the numerical and the analytical PL fit. The total systematic uncertainty of the profile likelihood fit is the quadratic sum of the three components

Table 6 Uncertainty decomposition for the muon p_T^ℓ and m_T fits, for two different values of the luminosity scale factor, using the shifted observable method and the impact method for PL fit. The errors arise from the limited number of bootstrap toys. The baseline luminosity is $76.42 \text{ [pb}^{-1}\text{]}$

Lumi	Method	p_T^ℓ fit unc. [MeV]			m_T fit unc. [MeV]		
		σ_{stat}	σ_{syst}	σ_{tot}	σ_{stat}	σ_{syst}	σ_{tot}
$\times 1$	Shifted obs.	44.0 ± 0.6	41.0 ± 1.0	60.5	34.4 ± 0.4	39.1 ± 1.0	55.1
$\times 1$	Impact	43.7	41.9	60.5	33.6	43.6	55.1
$\times 10$	Shifted obs.	18.8 ± 0.2	36.2 ± 0.5	40.7	15.7 ± 0.2	38.6 ± 0.5	41.6
$\times 10$	Impact	13.8	38.3	40.7	10.6	40.2	41.6

$$+ \sum_t \Delta \theta_p^{[a_r], f_1} \Delta \alpha_r^{[a_r], f_2} \quad (40)$$

For each matrix element, the first sum is statistical and typically occurs when the fitted distributions are projections of the same data, as is the case for the p_T^ℓ and m_T distributions in m_W fits. The second sum represents shared systematic sources of uncertainty.

Results of this comparison are presented in Fig. 8 and Table 8, which summarize the fit precision as a function of the assumed luminosity. The uncertainty decomposition method and the combination of the PL fit results agree to better than 0.1 MeV. For completeness, the result of a direct joint fit to the two distributions is shown as well; slightly more precise results are obtained in this case, as expected, especially for

Table 7 Left: list of systematic uncertainty contributions and the total uncertainty, in MeV, for the m_T and p_T^ℓ fits performed in covariance representation. Center, right: post-fit covariance among the three NPs

Uncertainty	m_T	p_T^ℓ
μ calib.	40.67	39.65
HR calib.	12.15	0.39
p_T^W model	2.49	9.77
Total	55.06	60.50

$$\begin{pmatrix} 0.99 & -0.01 & 0.00 \\ -0.01 & 0.70 & 0.00 \\ 0.00 & 0.00 & 0.99 \end{pmatrix} \begin{pmatrix} -40.67 \\ -12.15 \\ -2.49 \end{pmatrix}$$

$$\begin{pmatrix} 0.99 & 0.00 & -0.03 \\ 0.00 & 1.00 & 0.01 \\ -0.03 & 0.01 & 0.74 \end{pmatrix} \begin{pmatrix} -39.65 \\ -0.39 \\ -9.77 \end{pmatrix}$$

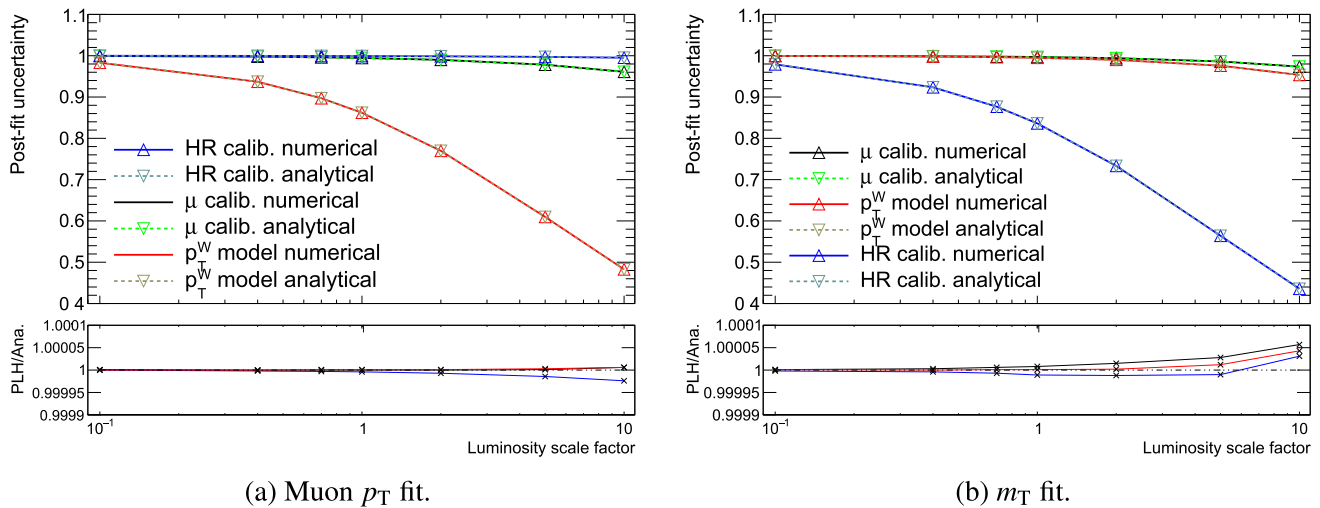


Fig. 7 Post-fit NP uncertainties at different values of the luminosity scale factor. The results of the numerical and the analytical PL-fits are compared in the ratio panel

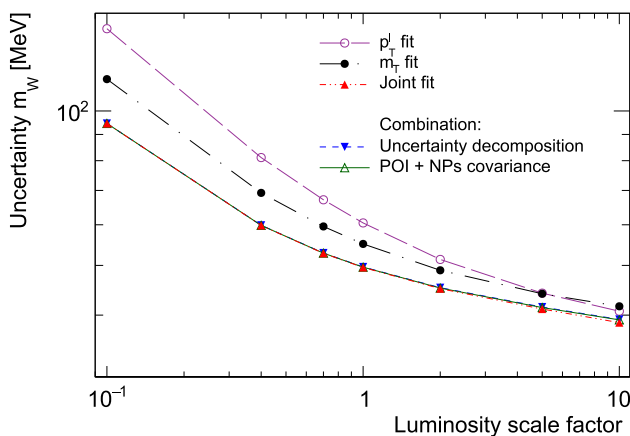


Fig. 8 Summary of m_T and p_T^ℓ PL fit results. Combinations are produced using the uncertainty decomposition method and using the covariance of the PL fit results

highly integrated luminosities where systematic uncertainties dominate.

We note that a combination of PL fit results based on the nuisance parameter representation, Eq. (17), as proposed in Ref. [42], seems difficult to justify rigorously. The principal reason is that Eq. (17) explicitly relies on the absence of

associated to these systematic uncertainties and the POI, for the profile likelihood fits to the m_T and p_T^ℓ distribution, respectively

correlations, prior to the combination, between the sources of uncertainty encoded in the covariance matrix V and the uncertainties treated as nuisance parameters. Since the input measurements result from PL fits, the POI of each input measurement is in general correlated with the corresponding NPs. One possibility would be to add terms to Eq. (17) that describe these missing correlations. It could also be envisaged to diagonalize the covariance of the inputs and perform the fit in this new basis, but this would work only if all measurements can be diagonalized by the same linear transformation, which is generally not the case.

7 Conclusion

We have studied the decomposition of fit uncertainties in two often-used statistical methods in high-energy physics, namely, fits in covariance representation and the profile likelihood. We recalled the equivalence between the two methods in the Gaussian limit and gave a complete set of expressions for the fit uncertainties in the parameters of interest, the nuisance parameters and their correlations. A direct correspondence was established between the standard uncertainty

Table 8 Summary of m_T and p_T^ℓ PL fit results. Combinations are produced using the uncertainty decomposition method, and using the covariance of the PL fit results

Luminosity scale factor	Total uncertainty in m_W [MeV]			Combinations	
	PL fits			Unc. decomp.	POI+NP covariance
	m_T	p_T^ℓ	Joint		
0.1	115.4	144.7	94.5	94.5	94.5
0.4	69.2	81.1	59.8	59.8	59.9
0.7	59.6	67.1	52.8	52.8	52.9
1.0	55.1	60.5	49.5	49.6	49.6
2.0	48.9	51.4	45.0	45.2	45.2
5.0	44.0	44.1	41.1	41.5	41.4
10.0	41.6	40.7	38.7	39.2	39.2

decomposition in covariance representation and the (POI, NP) covariance matrix elements in nuisance representation.

Numerical profile likelihood analyses generally define statistical and systematic uncertainty components from the results of statistical-only fits and systematic impacts, but this identification does not hold. The uncertainty of statistical-only fits underestimates the statistical uncertainty of fits including systematics, and systematic impacts correspondingly overestimate the genuine systematic uncertainty contributions. Impacts cannot be used as inputs to subsequent measurement combinations or interpretation fits.

We have introduced a set of analytical and numerical methods to remove this shortcoming. In Gaussian approximation, a consistent uncertainty decomposition can be directly extracted from the PL fit covariance matrix. For general (non-Gaussian or nonlinear) profile likelihood fits, a consistent uncertainty decomposition can be rigorously obtained from fits using shifted observables. We have illustrated these points by means of simple examples and have shown that profile likelihood fit results with properly decomposed uncertainties can be used consistently in downstream combinations or fits.

Data Availability Statement This manuscript has no associated data. [Author's comment: Data sharing not applicable to this article as no datasets were generated or analysed during the current study.]

Code Availability Statement This manuscript has no associated code/software. [Author's comment: The code developed for the present studies is available from the authors on request.]

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copy-

right holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Funded by SCOAP³.

References

1. M. Davier, A. Hoecker, B. Malaescu, Z. Zhang, Eur. Phys. J. C **71**, 1515 (2011). <https://doi.org/10.1140/epjc/s10052-012-1874-8> [Erratum: Eur. Phys. J. C **72**, 1874 (2012)]
2. S. Amoroso et al., Acta Phys. Polon. B **53**(12), A1 (2022). <https://doi.org/10.5506/APhysPolB.53.12-A1>
3. S. Schael et al., Phys. Rep. **427**, 257 (2006). <https://doi.org/10.1016/j.physrep.2005.12.006>
4. A. Tumasyan et al., Eur. Phys. J. C **83**(7), 560 (2023). <https://doi.org/10.1140/epjc/s10052-023-11587-8>
5. G. Aad et al., JHEP **08**, 045 (2016). [https://doi.org/10.1007/JHEP08\(2016\)045](https://doi.org/10.1007/JHEP08(2016)045)
6. D. van Dyk, L. Lyons (2023). <https://doi.org/10.48550/arXiv.2306.05271>
7. L. Lyons, D. Gibaut, P. Clifford, Nucl. Instrum. Meth. A **270**, 110 (1988). [https://doi.org/10.1016/0168-9002\(88\)90018-6](https://doi.org/10.1016/0168-9002(88)90018-6)
8. A. Valassi, Nucl. Instrum. Meth. A **500**, 391 (2003). [https://doi.org/10.1016/S0168-9002\(03\)00329-2](https://doi.org/10.1016/S0168-9002(03)00329-2)
9. W.A. Rolke, A.M. Lopez, J. Conrad, Nucl. Instrum. Meth. A **551**, 493 (2005). <https://doi.org/10.1016/j.nima.2005.05.068>
10. S. Schael et al., Phys. Rep. **532**, 119 (2013). <https://doi.org/10.1016/j.physrep.2013.07.004>
11. R. Barlow, R. Cahn, G. Cowan, F. Di Lodovico, W. Ford, G. Hamel de Monchenault, D. Hitlin, D. Kirkby, C. Le Diberder, F.G. Lynch, F. Porter, S. Prell, A. Snyder, M. Sokoloff, R. Waldi, Recommended Statistical Procedures for BABAR. BABAR Analysis Document 318 (2002). <https://babar.heprc.uvic.ca/BFROOT/www/Statistics/Report/report.pdf>
12. R.A. Fisher, Earth Environ. Sci. Trans. R. Soc. Edinb. **52**(2), 399–433 (1919). <https://doi.org/10.1017/S0080456800012163>
13. R.A. Fisher, J. Agric. Sci. **11**(2), 107–135 (1921). <https://doi.org/10.1017/S0021859600003750>
14. M. Aaboud et al., Phys. Lett. B **784**, 345 (2018). <https://doi.org/10.1016/j.physletb.2018.07.050>
15. O. Behnke, L. Moneta, Parameter Estimation (Wiley, 2013), chap. 2, pp. 27–73. <https://doi.org/10.1002/9783527653416.ch2>
16. G. Cowan, Statistical Data Analysis (Oxford University Press, Oxford, 1998)
17. K. Nakamura et al., J. Phys. G **37**, 075021 (2010). <https://doi.org/10.1088/0954-3899/37/7A/075021>

18. B. Malaescu, P. Starovoitov, Eur. Phys. J. C **72**, 2041 (2012). <https://doi.org/10.1140/epjc/s10052-012-2041-y>
19. M. Schmelling, Phys. Scr. **51**, 676 (1995). <https://doi.org/10.1088/0031-8949/51/6/002>
20. D. Britzger, Eur. Phys. J. C **82**(8), 731 (2022). <https://doi.org/10.1140/epjc/s10052-022-10581-w>
21. L. Demortier, Equivalence of the best-fit and covariance-matrix methods for comparing binned data with a model in the presence of correlated systematic uncertainties. CDF Note 8661 (1999). https://www-cdf.fnal.gov/physics/statistics/notes/cdf8661_chi2fit_w_corr_syst.pdf
22. G.L. Fogli, E. Lisi, A. Marrone, D. Montanino, A. Palazzo, Phys. Rev. D **66**, 053010 (2002). <https://doi.org/10.1103/PhysRevD.66.053010>
23. D. Stump, J. Pumplin, R. Brock, D. Casey, J. Huston, J. Kalk, H.L. Lai, W.K. Tung, Phys. Rev. D **65**, 014012 (2001). <https://doi.org/10.1103/PhysRevD.65.014012>
24. R.S. Thorne, J. Phys. G **28**, 2705 (2002). <https://doi.org/10.1088/0954-3899/28/10/314>
25. M. Botje, J. Phys. G **28**, 779 (2002). <https://doi.org/10.1088/0954-3899/28/5/305>
26. A. Glazov, AIP Conf. Proc. **792**(1), 237 (2005). <https://doi.org/10.1063/1.2122026>
27. R. Barlow, Nucl. Instrum. Meth. A **987**, 164864 (2021). <https://doi.org/10.1016/j.nima.2020.164864>
28. B. List, Decomposition of a covariance matrix into uncorrelated and correlated errors. Alliance Workshop on Unfolding and Data Correction, DESY. <https://indico.desy.de/event/3009/contributions/64704/> (2010)
29. G. Aad et al., JHEP **05**, 059 (2014). [https://doi.org/10.1007/JHEP05\(2014\)059](https://doi.org/10.1007/JHEP05(2014)059)
30. B. Efron, *Bootstrap Methods: Another Look at the Jackknife* (Springer New York, New York, 1992), pp. 569–593. https://doi.org/10.1007/978-1-4612-4380-9_41
31. A.M. Sirunyan et al., Phys. Lett. B **805**, 135425 (2020). <https://doi.org/10.1016/j.physletb.2020.135425>
32. V.M. Abazov et al., Phys. Rev. Lett. **108**, 151804 (2012). <https://doi.org/10.1103/PhysRevLett.108.151804>
33. M. Aaboud et al., Eur. Phys. J. C **78**(2), 110 (2018). <https://doi.org/10.1140/epjc/s10052-017-5475-4> [Erratum: Eur. Phys. J. C **78**, 898 (2018)]
34. R. Aaij et al., JHEP **01**, 036 (2022). [https://doi.org/10.1007/JHEP01\(2022\)036](https://doi.org/10.1007/JHEP01(2022)036)
35. T. Aaltonen et al., Science **376**(6589), 170 (2022). <https://doi.org/10.1126/science.abk1781>
36. S.D. Drell, T.M. Yan, Phys. Rev. Lett. **25**, 316 (1970). <https://doi.org/10.1103/PhysRevLett.25.316>
37. J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Matelaer, H.S. Shao, T. Stelzer, P. Torrielli, M. Zaro, JHEP **07**, 079 (2014). [https://doi.org/10.1007/JHEP07\(2014\)079](https://doi.org/10.1007/JHEP07(2014)079)
38. T. Sjostrand, S. Mrenna, P.Z. Skands, Comput. Phys. Commun. **178**, 852 (2008). <https://doi.org/10.1016/j.cpc.2008.01.036>
39. D. Bardin, A. Leike, T. Riemann, M. Sachwitz, Phys. Lett. B **206**(3), 539 (1988). [https://doi.org/10.1016/0370-2693\(88\)91627-9](https://doi.org/10.1016/0370-2693(88)91627-9)
40. K. Cranmer, G. Lewis, L. Moneta, A. Shibata, W. Verkerke, HistFactory: A tool for creating statistical models for use with RooFit and RooStats. Tech. rep. (New York University, New York, 2012). <https://cds.cern.ch/record/1456844>
41. W. Verkerke, D.P. Kirkby, in *Proceedings of the 13th International Conference for Computing in High-Energy and Nuclear Physics (CHEP03)* (2003). <http://inspirehep.net/record/634021>. [eConf C0303241, MOLT007]
42. J. Kieseler, Eur. Phys. J. C **77**(11), 792 (2017). <https://doi.org/10.1140/epjc/s10052-017-5345-0>