



# On new physics searches with multidimensional differential shapes

Felipe Ferreira<sup>a</sup>, Sylvain Fichtet<sup>b,\*</sup>, Veronica Sanz<sup>c</sup>

<sup>a</sup> Departamento de Física, Universidade Federal da Paraíba, Caixa Postal 5008, 58051-970, João Pessoa, Paraíba, Brazil

<sup>b</sup> ICTP-SAIFR & IFT-UNESP, R. Dr. Bento Teobaldo Ferraz 271, São Paulo, Brazil

<sup>c</sup> Department of Physics and Astronomy, University of Sussex, Brighton BN1 9QH, UK

## ARTICLE INFO

### Article history:

Received 20 July 2017

Received in revised form 4 January 2018

Accepted 6 January 2018

Available online 9 January 2018

Editor: G.F. Giudice

### Keywords:

LHC

Statistical analysis

Bayesian statistics

Effective theory

## ABSTRACT

In the context of upcoming new physics searches at the LHC, we investigate the impact of multidimensional differential rates in typical LHC analyses. We discuss the properties of shape information, and argue that multidimensional rates bring limited information in the scope of a discovery, but can have a large impact on model discrimination. We also point out subtleties about systematic uncertainties cancellations and the Cauchy–Schwarz bound on interference terms.

© 2018 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>). Funded by SCOAP<sup>3</sup>.

## 1. Introduction

In modern High Energy Physics, the use of large datasets has become commonplace. In two areas in particular, Particle Physics and Cosmology, the forefront of discoveries and characterisation of new phenomena relies on extraction of information from complex datasets produced by experiments like Planck [1] and the LHC [2]. In both fields, a precise theoretical paradigm is used to interpret the data ( $\Lambda$ CDM and SM, respectively) and the search for new phenomena depends then on identifying subtle deviations within the data, often relying on machine learning techniques. For example, the discovery rare SM processes, like mono-top [3] and Higgs decays to tau-leptons [4], has been achieved using this methodology.

On the theoretical side, these multivariate techniques obscure the physical understanding of which variables drive the analysis, making the re-interpretation of results very difficult and in general hindering the public use of the data. Yet more detailed information, in particular differential rates, is required to advance the programme of searching for a new paradigm beyond the standard one. For example, the use of differential information on Higgs production [5] has proven key to pushing the limits of understanding the impact of possible new phenomena in the Higgs boson properties.

In this paper we investigate the advantages and limitations of multidimensional shape information in searching for new physics

and present two case studies, the new physics search in the context of the SM Effective Field Theory (SMEFT) and the characterisation of the quantum numbers of a new resonance. These case studies, together with the material collected in the Appendix, give examples of how differential distributions can be exploited by theorists. Currently, experiments provide mostly one-dimensional distributions, and only rarely two-dimensional information – a notable exception to this trend is provided by the ATLAS analysis [6] of  $h \rightarrow \gamma\gamma$ , which made public the 2D differential distributions in  $p_T$  of the Higgs and number of jets. This work is also meant as an incentive for experiments to provide systematically differential distributions in an exploitable form.

## 2. Statistical basics

In this section we set the notation of the statistical analysis. We denote phase space by  $\mathcal{D}$ , and consider a binning of  $\mathcal{D}$  in  $d$  dimensions. The bins are set along a dimension  $i \in (1 \dots d)$  and labelled by  $r_i$ , with the coordinates  $(r_1, \dots, r_d)$  of a bin denoted  $r$ , and the associated piece of phase space  $\mathcal{D}_r$ . The observed event number in the bin  $r$  is denoted  $\hat{n}_r$ , and the expected event number for a given value of the underlying parameter  $\theta$  is denoted  $n_r(\theta)$ . Total number of observed events is  $\hat{n} = \sum_r \hat{n}_r$  and the expected total number events is  $n = \sum_r n_r$ .

For further convenience one also introduces the  $d$ -dimensional density of expected events  $f_X(x)$ , where  $X = (X_i)$  denotes the set of binned variables.  $f_X(x)$  is simply the differential event rate

\* Corresponding author.

E-mail addresses: [ff83@sussex.ac.uk](mailto:ff83@sussex.ac.uk) (F. Ferreira), [sylvain@ift.unesp.br](mailto:sylvain@ift.unesp.br) (S. Fichtet), [v.sanz@sussex.ac.uk](mailto:v.sanz@sussex.ac.uk) (V. Sanz).

normalised by the total event rate. The expected event number in a bin  $r$  is then given by

$$n_r = n \int_{\mathcal{D}_r} f_X(x) dx. \quad (1)$$

### 2.1. Likelihood

The likelihood function  $L$  is defined as the conditional probability of obtaining the observed data given a hypothesis, taken as a function of this hypothesis. For a hypothesis  $H$  with a set of parameters  $\theta$ ,

$$L(\theta) \equiv \text{Pr}(\text{data}|H, \theta). \quad (2)$$

The likelihood function can be defined up to an overall constant factor.

The events counted in each of the bins are statistically independent, hence the likelihood factorises as

$$L = \prod_r L_r. \quad (3)$$

The event number in every bin follows a Poisson statistics, so that the likelihood function in the bin  $r$  is given by

$$L_r(\theta) = n_r(\theta) \hat{n}_r e^{-n_r(\theta)}. \quad (4)$$

For a given integrated luminosity  $\mathcal{L}$ ,  $n_r(\theta)$  is given by the event rate on the bin,  $n_r(\theta) = \mathcal{L} \sigma_r(\theta)$ .

The likelihood can be formally factored in a Poisson term  $L_{\text{tot}}$  containing the information about the total rate and a term  $L_{\text{shape}}$  containing the information about the shape of the differential distribution, so that  $L(\theta) = L_{\text{tot}}(\theta) L_{\text{shape}}(\theta)$  with

$$L_{\text{tot}}(\theta) = n(\theta) \hat{n} e^{-n(\theta)} \quad (5)$$

$$L_{\text{shape}}(\theta) = \prod_r \left( \frac{n_r(\theta)}{n(\theta)} \right)^{\hat{n}_r}. \quad (6)$$

We stress this feature remains valid beyond Poisson statistics, when systematic uncertainties are also included in the likelihood. Indeed, it is always possible to split the nuisance parameters into as subset affecting only  $L_{\text{tot}}(\theta)$  and a subset affecting only  $L_{\text{shape}}(\theta)$ .

Finally, in the limit where bin size is small enough so that every bin contains only zero or one event,  $L_{\text{shape}}(\theta)$  tends to the classical “unbinned” likelihood expressed in terms of the continuous probability density function of the events along the previously-binned variable  $X$ ,

$$L_{\text{shape}}^{\text{unbinned}}(\theta) = \prod_{l=1}^{\hat{n}} f_X^\theta(x_l), \quad (7)$$

where the  $x_l$  are the values of  $X$  associated to each of the observed events. The  $f_X^\theta$  has been defined in Eq. (1).

### 2.2. Credible regions and hypothesis testing

We adopt the framework of Bayesian statistics.<sup>1</sup> The model parameters are given an a-priori probability density  $\pi(\theta)$ , called “prior”, that can encode both subjective and objective information. The

<sup>1</sup> The Bayesian framework is consistent with the “likelihood principle”, which states that all experimental information is encoded in the likelihood function. This is not the case of, for example, frequentist  $p$ -values.

“posterior” density is defined as  $p(\theta) \propto L(\theta)\pi(\theta)$ , it provides the preferred regions of  $\theta$  ones data are taken into account. The shape of the prior becomes irrelevant once enough data are accumulated, i.e. when the posterior is data-dominated.<sup>2</sup>

A so-called  $1 - \alpha$  credible region of highest density is defined by the domain  $\Omega^{1-\alpha} = \{\theta | p(\theta) > p_{1-\alpha}\}$ , where  $p_{1-\alpha}$  is determined by the fraction of integrated posterior

$$\frac{\int_{\Omega^{1-\alpha}} d\theta p(\theta)}{\int_{\Omega} d\theta p(\theta)} = 1 - \alpha, \quad (8)$$

$\Omega$  being the whole parameter space. We will use the credible regions associated with  $1 - \alpha = \{68.27\%, 95.45\%, 99.73\%\}$ .<sup>3</sup>

Comparison between two hypotheses  $H_0$  and  $H_1$  is done by means of the Bayes factor

$$B_{01} = \frac{\int_{\Omega_1} L(\theta_1) \pi_1(\theta_1)}{\int_{\Omega_0} L(\theta_0) \pi_0(\theta_0)}, \quad (9)$$

where the  $\pi_{0,1}$  are the priors for hypotheses  $H_{0,1}$  respectively. The Bayes factor is interpreted using the Jeffreys’ scale [7], which associates weak, moderate and strong evidence in favour of  $H_0$  to the threshold values  $\log B_{01} \sim 1, 2.5, 5$  (i.e.  $B_{01} \sim 3, 12, 150$ ).

The Bayes factor framework can be used in the context of new physics searches. In order to assess that the data favour a hypothesis where a parameter  $\theta$  is different from a given value  $\theta_0$  one has to compare the  $H_1$  hypothesis to  $H_0 \equiv H_1 | \theta = \theta_0$  (see also [7], [8]). In the context of effective operators,  $H_1$  can for instance be the SM deformed by higher dimensional operators (the SMEFT), while  $H_0$  is the SM. Defining  $B_0 \equiv 1/B_{01}$ , we have

$$B_0 = \frac{\int_{\Omega} L(\theta) \pi(\theta)}{L(\theta_0)}, \quad (10)$$

that we refer to as the *discovery Bayes factor*. The test assesses that  $\theta \neq \theta_0$  for  $B_0 > 1$ , using the thresholds given above.

### 2.3. Asimov (projected) data

In order to evaluate the sensitivity of a future analysis, measurement, or experiment, one can rely on imaginary, speculative data. That is, instead of introducing actual observed data in the likelihood Eq. (2), one can instead introduce speculative data coming for instance from a simulation of the experiment. We refer to these as *projected data*.

An important subtlety, well discussed in [9], is that an assumption has to be made on the statistical fluctuations present in the projected data. Along this paper, we will simply consider the case where no statistical fluctuations are present in the projected data. A dataset satisfying this condition is sometimes referred to as an “Asimov” dataset [9].

The event numbers in the projected dataset assuming no statistical fluctuations and the presence of an operator with coefficient  $c'$  are then simply given by  $\mathcal{L} \sigma_r(c')$ . In practice, these rates have to be estimated by MonteCarlo simulations, just like the expected ones.

<sup>2</sup> The prior has however to satisfy basic physical conditions, such as keeping an event number positive in order to avoid singularities in the posterior. As a general rule, the posterior should be data-dominated in the limit of many data, otherwise the inference process cannot happen.

<sup>3</sup> Note that confidence regions are not uniquely defined, but the method of highest density is the most commonly used and arguably the most natural.

### 3. Information content of multidimensional differential rates

In this section we study the information content of differential rates and their use in new physics search and model comparison.

The information content of a likelihood function with respect to a parameter  $\theta$  is measured by the observed Fisher information  $I_\theta[L] \equiv -\partial_\theta^2 \log L$ .<sup>4</sup> In this section we focus on a single parameter case, which is enough for our arguments. The likelihood functions we focus on arise from event counting and can always be factored as

$$L(\theta) = L_{\text{tot}}(\theta) L_{\text{shape}}(\theta), \quad (11)$$

where  $L_{\text{tot}}(\theta)$  contains the information on the total rates and  $L_{\text{shape}}(\theta)$  contains information on the shape of differential distributions.<sup>5</sup> The information content from total event number and from the shape ( $I_{\text{tot}} \equiv I[L_{\text{tot}}]$ ,  $I_{\text{shape}} \equiv I[L_{\text{shape}}]$ ) are thus independent from each other.  $I_{\text{shape}}$ , therefore, could be arbitrarily large with respect to  $I_{\text{tot}}$ , i.e. the amount of information contained in the shape could dominate over the amount of information contained in the total rate. It is thus fully justified to systematically take into account the shape information on top of the total rate information.

The information content of  $L_{\text{shape}}$  with respect to the dimensionality of the differential rate distribution is slightly more subtle. For concreteness let us consider the case of one kinematic variable (“1D”) versus two kinematic variables (“2D”). The variables are labelled  $X$  and  $Y$ . Before making general considerations, let us present a concrete example. Let us consider the unbinned shape likelihood shown in Eq. (7). For two variables, the 2D unbinned likelihood takes the form

$$L_{\text{shape}}^{2D}(\theta) = \prod_{I=1}^{\hat{n}} f_{X,Y}(x_I, y_I). \quad (12)$$

In the limit where the underlying process does not lead to correlation between the  $X$  and  $Y$  variables, one has  $f_{X,Y}(x, y) = f_X(x)f_Y(y)$  by definition. Hence the 2D likelihood factorises into the 1D likelihoods associated with  $X$  and  $Y$  respectively,  $L_{\text{shape}}^{2D}(\theta) = L_{\text{shape}}^{1D,X}(\theta)L_{\text{shape}}^{1D,Y}(\theta)$ . On the other extreme, in the limit where the underlying process leads to a full correlation between the two variables, one has a linear relation  $X = \alpha Y$  with  $\alpha \neq 0$ , and the joint density becomes  $f_{X,Y}(x, y) \equiv f_Y(y)\delta(x - \alpha y) = \alpha f_X(x)\delta(x - \alpha y)$ . The Dirac  $\delta$ 's being independent of  $\theta$ , one gets that  $L_{\text{shape}}^{2D}(\theta) \propto L_{\text{shape}}^{1D,X}(\theta) \propto L_{\text{shape}}^{1D,Y}(\theta)$  – the overall constant being irrelevant in the definition of the likelihood. Another example assuming a Gaussian shape for  $f_{X,Y}$  is worked out in Appendix A.

Let us now discuss the information content of the 2D shape likelihood  $L_{\text{shape}}^{2D}$ . This discussion applies to the above example, but the arguments used are general. If the  $X$  and  $Y$  variables are totally correlated, one has  $L_{\text{shape}}^{2D} = L_{\text{shape}}^{1D,X} = L_{\text{shape}}^{1D,Y}$ , and there is no gain in going from 1D distributions to the 2D distribution. On the other extreme, if the two variables  $X, Y$  provide uncorrelated information, the likelihood factorises and the total information is given by

$L_{\text{shape}}^{1D,X} + L_{\text{shape}}^{1D,Y}$ . This is the maximum information possible, thus one obtains that

$$L_{\text{shape}}^{2D} \leq L_{\text{shape}}^{1D,X} + L_{\text{shape}}^{1D,Y}, \quad (13)$$

which implies that the information gain from 1D to 2D cannot be arbitrarily large. The gain from 1D to 2D is maximal when the two variables are uncorrelated and their Fisher information are of the same order of magnitude,  $L_{\text{shape}}^{1D,X} \sim L_{\text{shape}}^{1D,Y} \equiv L_{\text{shape}}^{1D}$ . In such case Eq. (13) tells that the maximal value for  $L_{\text{shape}}^{2D}$  is given by twice  $L_{\text{shape}}^{1D}$ . These considerations hold in presence of both statistical and systematic uncertainties. In the rest of this paper, we will often refer to the information gain obtained from using a 1D differential distribution to using a 2D differential distribution as the “1D/2D” gain.

Pursuing in our general considerations, let us evaluate the impact of the various pieces of information discussed above in the subsequent statistical analyses. In order to proceed, we need to consider statistical tests. We adopt the framework of Bayesian statistics, which allows a unified treatment of discovery and model comparison. For hypothesis testing, the relevant quantity to use is the Bayes factor (see Sec. 2 for definitions).

We assume that the likelihood for each hypothesis can be approximated by a Gaussian with respect to the parameter of interest  $\theta$ . This limit tends to occur once at least  $O(10)$  events are collected.<sup>6</sup> The likelihood function then takes the form

$$L(\theta) \approx L_{\text{max}} \exp\left(-I \frac{(\theta - \bar{\theta})^2}{2}\right), \quad (14)$$

where  $\bar{\theta}$  is the value of  $\theta$  preferred by the data,  $I$  is the Fisher information for  $\theta$ , and the constant  $L_{\text{max}}$  encodes the information about goodness-of-fit<sup>7</sup> between the hypothesis and the data. In this Gaussian limit, the Bayes factors exhibit simple expressions with respect to the Fisher information(s). Moreover, the Fisher information depends linearly on the observed total event number  $n_{\text{obs}}$  to a good approximation.<sup>8</sup> Hence we have  $L_{\text{shape}}^{1D} = \alpha^{1D} n_{\text{obs}}$ ,  $L_{\text{shape}}^{2D} = \alpha^{2D} n_{\text{obs}}$ , with  $\alpha^{2D} \leq \alpha^{1D,X} + \alpha^{1D,Y}$ . Note that the  $\alpha^{2D}$  information coefficient can be at best  $\alpha^{1D,X} \sim \alpha^{1D,Y}$ . This direct link of Fisher information to the event number is crucial to concretely quantify the impact of the various likelihoods.

To characterise discovery, we introduce the *discovery Bayes factor*, which compares a model hypothesis with a free parameter  $\theta$  with the same hypothesis restricted to  $\theta = \theta_0$  (see Sec. 2, and also [7], [8]). The discovery Bayes factor is given by

$$\log B_0 = \alpha n_{\text{obs}} \frac{(\theta_0 - \bar{\theta})^2}{2} - \frac{1}{2} \log\left(\frac{V^2 \alpha n_{\text{obs}}}{2\pi}\right), \quad (15)$$

with  $\alpha = \alpha_{\text{tot}} + \alpha_{\text{shape}}$ . The first term in Eq. (15) encodes the comparison of central values while the second term encodes prior information. The  $V$  parameter, given by  $V = 1/\pi(\theta_0)$ , encodes the information about both the prior volume and its value at  $\theta_0$ . This second term becomes quickly negligible once  $n_{\text{obs}}$  increases, realising explicitly the fact that prior information tends to become irrelevant in presence of data. We note that the constant  $L_{\text{max}}$  does

<sup>4</sup> The expected Fisher information has been recently advocated in Ref. [27] as a way of visualising the information content of differential distributions and the expected constraints on EFT operators. The results presented in this reference are orthogonal to those discussed in the present Letter, where we focus on the relative power of multidimensional analyses and use the observed Fisher information only for qualitative arguments.

<sup>5</sup> This is true for Poisson statistics, and remains also true in presence of systematic uncertainties, as these can always be split into as subset affecting only  $L_{\text{tot}}(\theta)$  and a subset affecting only  $L_{\text{shape}}(\theta)$ .

<sup>6</sup> This behaviour has first been described in Ref. [28].

<sup>7</sup> Here “goodness-of-fit” is meant in the broad sense of how well the model fits the data. The value of  $L_{\text{max}}$  is by itself not enough to reach robust conclusions about the model, which is why better-defined hypothesis tests such as the Bayes factor (see Sec. 2) or frequentists tests (see e.g. [9],) have to be introduced.

<sup>8</sup> In the Gaussian case this is simply because the observed Fisher information equals the inverse of the determinant of the covariance, which is approximately proportional to  $1/n_{\text{obs}}$ .

not appear in the discovery Bayes factor. Comparing the discovery Bayes factor for 1D and 2D distributions, and assuming  $\bar{\theta}_{1D} \sim \bar{\theta}_{2D}$ , we get that

$$\log B_0^{2D} \leq \log B_0^{1D,X} + \log B_0^{1D,Y}. \quad (16)$$

The bound is saturated when the 2D information is maximal,  $\alpha_{\text{shape}}^{2D} = \alpha_{\text{shape}}^{1D,X} + \alpha_{\text{shape}}^{1D,Y}$ , and for  $\alpha_{\text{tot}} \ll \alpha_{\text{shape}}$ . Finally, the information gain from 1D to 2D would be maximal when  $\alpha_{\text{shape}}^{1D,X} \sim \alpha_{\text{shape}}^{1D,Y} \equiv \alpha_{\text{shape}}^{1D}$ , in which case we obtain that  $\log B_0^{2D}$  could be at most twice  $\log B_0^{1D}$ .

This bound on the 2D Bayes factor can be easily translated in terms of sample size and evidence strength. In terms of sample size, the bound can be translated using the fact that the 1D/2D gain amounts to *at most* doubling the  $n_{\text{obs}}$  from the 1D case. In terms of strength of evidence (see Jeffreys' scale), we observe that moving from 1D to 2D can lead to a shift of at most one step in evidence strength. For instance, if the 1D Bayes factor would give moderate evidence ( $\log B_0 = 2.5$ ), the 2D Bayes factor could at most reach strong evidence ( $\log B_0 = 5$ ).

So far we have discussed how the 1D/2D information gain is bounded in the scope of a discovery. But what about model discrimination? Approximating the likelihoods as Gaussians in both hypotheses  $H_1, H_2$ , the Bayes factor comparing  $H_1$  to  $H_2$  reads

$$\log B_{12} = \log \left( \frac{L_{\text{max},1}}{L_{\text{max},2}} \right) - \log \left( \frac{\alpha_1}{\alpha_2} \right). \quad (17)$$

Note that the structure of this Bayes factor is different from the discovery Bayes factor. The first term encodes the relative goodness-of-fit of the models with respect to data, whereas the second term in Eq. (17) is a ratio of Fisher information, and should be understood as a measure of the relative fine-tuning of the two models, see [10].<sup>9</sup> This second, “naturalness” term is independent of  $n_{\text{obs}}$ . In contrast, the ratio of maximum likelihoods depends in general on  $n_{\text{obs}}$ , as goodness-of-fit is in general different in both hypotheses. In fact, in the large sample limit, one expects

$$L_{\text{max},1}/L_{\text{max},2} \sim \exp(\beta n_{\text{obs}}), \quad (18)$$

where  $\beta$  a positive or negative constant. The case  $\beta > 0$  corresponds to the  $H_1$  model being a better fit than  $H_2$ , and conversely. The absolute value of  $\log(L_{\text{max},1}/L_{\text{max},2})$  is thus expected to grow with  $n_{\text{obs}}$ , reducing the relative impact of the naturalness term.

We can now compare a Bayes factor based on a 1D distribution,  $\log B_{12}^{1D}$ , with a Bayes factor based on a 2D distribution,  $\log B_{12}^{2D}$ . Neglecting the naturalness terms (which are however different for 1D and 2D), we are left with comparing the goodness-of-fit terms of the 1D and 2D cases, roughly given by  $\beta_{1D} n_{\text{obs}}$  and  $\beta_{2D} n_{\text{obs}}$ . We have found no bound on the  $\beta_{2D}/\beta_{1D}$  ratio based on general information considerations. This suggests that the 1D/2D information gain can in principle be arbitrarily large in case of model comparison – this gain being set by the relative goodness-of-fit of the models. These features will be demonstrated in concrete examples in the next sections. Such unbounded 1D/2D gain does not happen for discovery because the discovery Bayes factor does not involve goodness-of-fit, i.e.  $B_0$  does not depend on  $L_{\text{max}}$ .

#### 4. Case studies for discovery and characterisation

In this Section we perform multidimensional shape analyses in some realistic scenarios, with the primary focus of illustrating

the properties of shape information described in Sec. 3. Other aspects of the statistical analyses are also discussed, and all details needed to reproduce the simulations and analyses are provided. Two new physics scenarios are considered: operators from the SMEFT, and bosonic resonances, which are treated respectively in subsections 4.2 and 4.3.

##### 4.1. General setup

This subsection collects information about the simulation and statistical setup which is common to all cases treated.

To simulate the conditions in the LHC for different model hypotheses, we use FeynRules to implement the BSM models, and the UFO [11] output to interface with MadGraph5 aMC@NLO platform [12]. The parton events are then passed through Pythia [13] for parton-showering and hadronisation. Finally, the hadrons are reconstructed via the anti- $k_T$  algorithm [14] with an R-parameter set to 0.4 using the FastJet [15] interface of MadAnalysis 5 [16]. A jet is tagged as arising from a  $b$ -quark when a  $B$ -hadron is present within a cone of radius  $R = 0.4$  centered on the jet momentum direction. A private pyROOT script has been developed in order to automatise and monitor the whole analysis in the framework of MadAnalysis.

As our focus is on evaluation and comparison of analyses based on future data prospects, we have introduced “projected” data in our likelihoods, see Sec. 2 for details. No statistical fluctuations in the projected data are assumed (i.e. we assume Asimov data), and these are thus directly given by the expected rates.

Finally, regarding the aspect of shape information, in this section we quantify the “1D/2D gain” using the expression

$$n^{2D}|_{B=150}/n^{1D}|_{B=150} - 1. \quad (19)$$

This measure compares the sample size needed to reach a Bayes factor of 150 in the 1D and 2D cases. In the case of a new physics search, the upper bound on the 2D information (see Eq. (13) and below) translates here as an upper value of 100% for this 1D/2D gain.

##### 4.2. Case I: CP-violating and -conserving effective operators

###### 4.2.1. Scenarios and simulation

In the scenario where new particles are too heavy to be produced on-shell at the LHC, their observable effects are better described by a low-energy effective theory, in the so-called SMEFT framework. For our case study we assume the presence of two characteristic dimension-six operators,

$$\mathcal{L} = \mathcal{L}_{\text{SM}} + c_{HW} O_{HW} + \tilde{c}_{HW} \tilde{O}_{HW}, \quad (20)$$

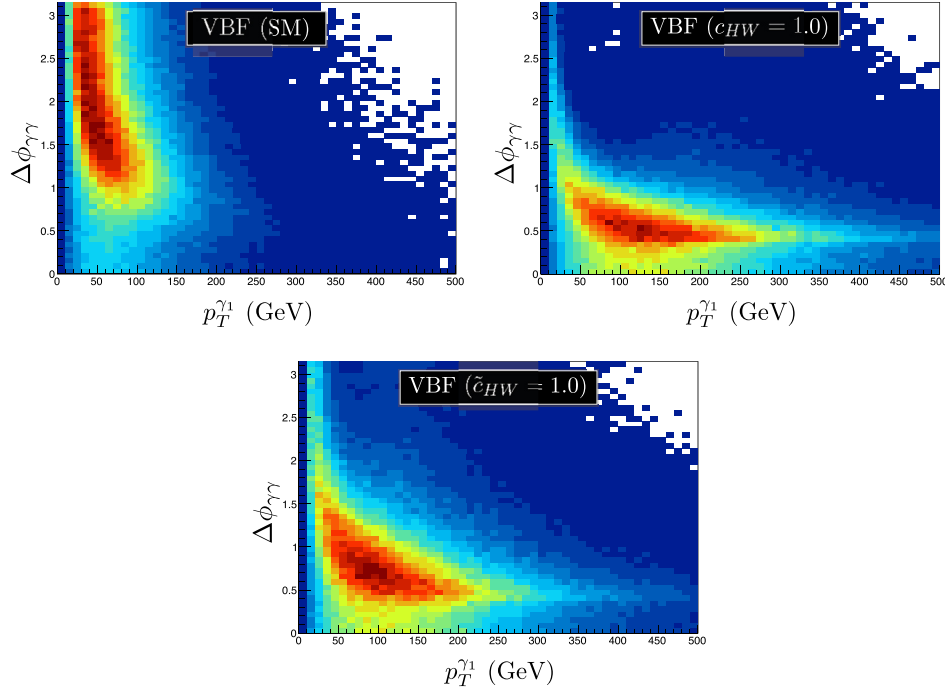
where the operators  $O_{HW}$  and  $\tilde{O}_{HW}$  are defined as

$$O_{HW} (\tilde{O}_{HW}) = \frac{2ig}{m_W^2} [D^\mu \Phi^\dagger T_{2k} D^\nu \Phi] W_{\mu\nu}^k (\tilde{W}_{\mu\nu}^k). \quad (21)$$

Here  $\Phi$  is the Higgs doublet and  $W_{\mu\nu}$  the  $SU(2)_L$  field strength. The Wilson coefficients  $c_i$  are normalised following the SILH basis conventions [17] and their current bounds can be found in Refs. [18,19]. Note we use the implementation of these operators provided in Ref. [20].

As an example of the use of differential information, we consider Higgs production in the Vector Boson Fusion (VBF) mechanism, and generate samples of 450 K events. Basic selection cuts require the presence of two jets with a transverse momentum  $p_T^j > 20$  GeV, pseudo-rapidity  $|\eta_j| < 4.5$ , as well as typical VBF cuts: the Dijet invariant mass is required to be larger than 400 GeV

<sup>9</sup> The likelihood ratio  $L_{\text{max},1}/L_{\text{max},2}$  also appears in frequentist tests. In contrast the second term does not show up in the frequentist approach.



**Fig. 1.** Differential event rate in the  $p_T^{\gamma_1} - \Delta\phi_{\gamma\gamma}$  plane, in the vector boson fusion (VBF) process for the SM (top), CP-conserving (middle) and CP-violating (bottom) hypotheses for new physics in the SMEFT.

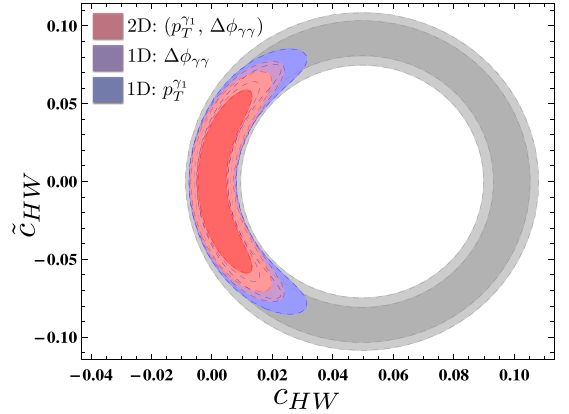
and the jet separation in pseudo-rapidity to be above 2.8. The analysis selects two high-momentum jets  $j_1, j_2$  and two photons  $\gamma_1, \gamma_2$  from the Higgs decay. The indexes 1 and 2 denote the leading and sub-leading particles.

In order to determine the differential rates for arbitrary values of the effective operator coefficients, we use the reconstruction method described in [21–23] – which has been dubbed “morphing” in experimental references. The optimal version of the reconstruction method has been described in [23] and is used in our analyses. The reconstruction provides *estimates* of the various components of the rate on a given bin,  $\hat{\sigma}_r(c) = \hat{\sigma}_r^{\text{SM}} + c\hat{\sigma}_r^{\text{int}} + c^2\hat{\sigma}_r^{\text{BSM}}$  with  $c = c_{HW}, \tilde{c}_{HW}$ . An important subtlety related to the estimation of the interference component in regions with low event rates is described in Appendix B.

The projected data are directly given by the  $\hat{\sigma}(c')$  event rates, where  $c'$  is the value operator coefficient assumed to be present in these projected data. The fact that we use the same rates  $\hat{\sigma}(c)$  for both projected and expected rates leads to an interesting simplification. It turns out that the main Monte Carlo uncertainties cancel out from the likelihood, leaving the maximum likelihood rigorously unchanged (see Appendix C). Rather, the uncertainties are only changing the Fisher information part, and more generally the likelihood line-shape. This simplification implies that in practice, the number of Monte Carlo events to perform the simulation needs to be only mildly larger than the nominal number of events. Having for example  $n_r^{\text{MC}} > 3n_r^{\text{obs}}$  gives a systematic uncertainty of  $\sim 33\%$  on the Fisher information, and thus of  $\sim 16.5\%$  on the projected statistical uncertainty.

#### 4.2.2. Results and discussion

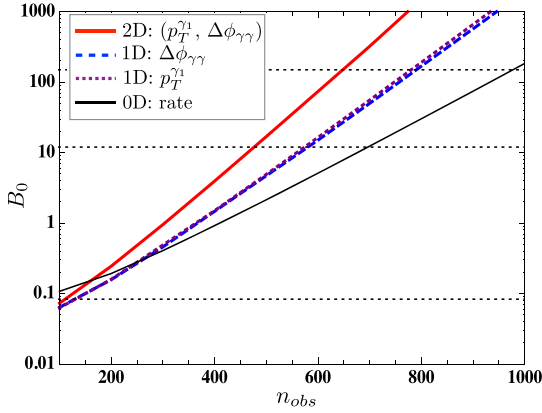
Having described all the aspects of our simulations and likelihoods, let us proceed with the data analysis focusing on 1D and 2D differential distributions and discuss the shape information. We have tested the constraining power of a set of basic kinematic variables including transverse momenta, azimuthal angles and longitudinal rapidity differences between final state objects. Throughout



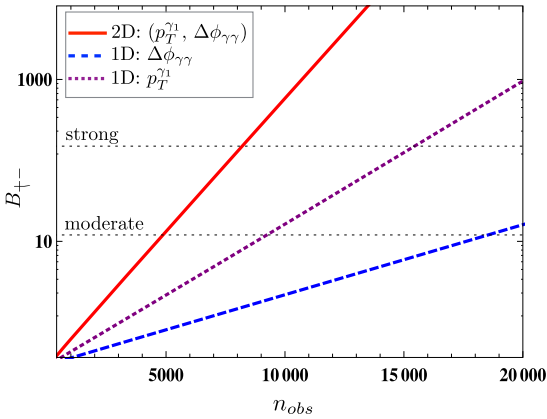
**Fig. 2.** Preferred regions in the  $c_{HW} - \tilde{c}_{HW}$  plane, assuming  $3000 \text{ fb}^{-1}$  of integrated luminosity and no statistical fluctuations. The 95% and 99% credible regions from measurement of the total rate are shown in gray. Regions taking into account 1D differential rate in  $\Delta\phi_{\gamma\gamma} [p_T^{\gamma_1}]$  are shown in blue [purple]. The regions taking into account the 2D differential rate  $(p_T^{\gamma_1}, \Delta\phi_{\gamma\gamma})$  are shown in red. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

our study we found that the pair of variables with the best 1D/2D gain for discovery are  $p_T^{\gamma_1}, \Delta\phi_{\gamma\gamma}$ , hence we present the analysis with respect to these two variables, see Fig. 1. Note, though, that the analysis does not include detector effects which could change the set of optimal variables. For the figure we have chosen  $c_{HW} = 1$  in order to make clear how the distributions change. In the rest of the analysis, smaller values of  $c_{HW}$  are used, which is more realistic.

We first compute the posterior distributions for  $L_{\text{tot}}, L_{1D}, L_{2D}$ , assuming  $c'_{HW} = \tilde{c}'_{HW} = 0$  in the projected data. The preferred regions are shown in Fig. 2. We can see that taking into account the 1D distributions is crucial in order to lift the degeneracy in the  $c_{HW} - \tilde{c}_{HW}$  plane. In contrast, the gain from 1D to 2D differen-



**Fig. 3.** Discovery Bayes factor for the  $O_{HW}$  operator, assuming an underlying value of  $c_{HW} = -0.01$  in the data and no statistical fluctuations. A flat prior over  $[-1, 1]$  is assumed for  $c_{HW}$ . The gray, blue, purple, red lines correspond respectively to total rate (0D), 1D differential rate in  $\Delta\phi_{\gamma\gamma}$ , 1D differential rate in  $p_T^{\gamma_1}$ , 2D differential rate in  $(\Delta\phi_{\gamma\gamma}, p_T^{\gamma_1})$ . (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

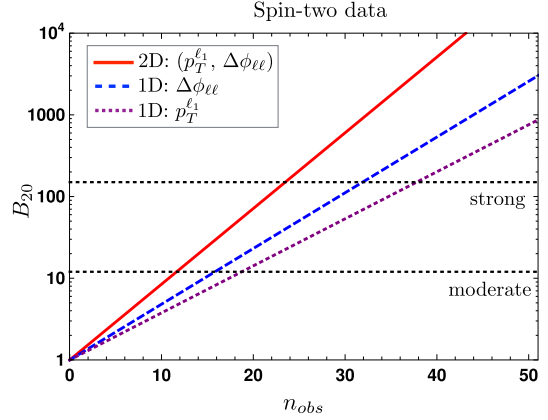


**Fig. 4.** Bayes factor for  $O_{HW}$  versus  $\tilde{O}_{HW}$ , assuming the underlying values  $c_{HW} = -0.01$ ,  $\tilde{c}_{HW} = 0$  in the data and no statistical fluctuations. A flat prior over  $[-1, 1]$  is assumed for  $c_{HW}$ ,  $\tilde{c}_{HW}$ . The gray, blue, purple, red lines correspond respectively to total rate, 1D differential rate in  $\Delta\phi_{\gamma\gamma}$ , 1D differential rate in  $p_T^{\gamma_1}$ , 2D differential rate in  $(\Delta\phi_{\gamma\gamma}, p_T^{\gamma_1})$ . (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

tial information turns out to be mild. This is qualitatively expected from the arguments in Sec. 3 as the 1D/2D gain is bounded.

Assuming that the  $O_{HW}$  operator with  $c'_{HW} = -0.01$  is present in the data, we compute the discovery Bayes factor for  $O_{HW}$  as a function of the sample size, as shown in Fig. 3. A mild gain between  $B_0^{\text{tot}}$  and  $B_0^{1D}$  and between  $B_0^{1D}$  and  $B_0^{2D}$  is observed. That the two 1D Bayes factors have almost the same value is apparently a mere coincidence. The 1D/2D gain for this pair of kinematic variables is the *best* we found among all the kinematic variables considered. A positive result is also obtained when assuming the existence of the operator  $\tilde{O}_{HW}$  instead of  $O_{HW}$ . In all cases, the mild 1D/2D gain observed is in agreement with our general statistical arguments of Sec. 3.

Still assuming the presence of the  $O_{HW}$  in the data, we then use a Bayes factor comparing the  $c_{HW} \neq 0$  hypothesis with the  $\tilde{c}_{HW} \neq 0$  hypothesis, see Eq. (17). The result is shown in Fig. 4. We observe that the 1D/2D gain in this case is much larger than for discovery. For example we can see that the 1D/2D gain in sample size is about 90%, which corresponds to almost doubling the sample size. For comparison, for the discovery, the gain in sample size is of 20%. These features are again in agreement with our general statistical analysis of Sec. 3.



**Fig. 5.** Bayes factor for spin-2 versus spin-0, assuming spin-2 in the data and no statistical fluctuations. The blue, purple, red lines correspond respectively to the 1D differential rate in  $p_T^{\ell_1}$ , 1D differential rate in  $\Delta\phi_{\ell\ell}$ , 2D differential rate in  $(p_T^{\ell_1}, \Delta\phi_{\ell\ell})$ . (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

We should stress that certain kinematic variables such as  $m_{jj}$  have a better discriminating power than the variables we consider but are not as good for discovery, hence we present results based on the  $p_T^{\gamma_1}$ ,  $\Delta\phi_{\gamma\gamma}$  variables in order to have a direct comparison with the discovery Bayes factor. Nevertheless, the large 1D/2D gain persists for these other combinations of variables, the pair  $m_{jj} - p_T^{\gamma_1}$ , for instance, has 1D/2D gain in sample size is about  $\sim 100\%$ .

#### 4.3. Case II: Testing the spin of a resonance

We consider now the discovery of a new resonance with either spin zero or two, and how our analysis would help on the characterisation of the resonance using a final state of a pair of  $Z$  bosons further decaying leptonically,  $pp \rightarrow \phi + X \rightarrow 2\ell^+ 2\ell^- + X$ . The spin-0 and spin-2 resonance behaviour is simulated using existing FeynRules models. Samples of 450K events were generated, following the hadronisation procedure previously described. The two pairs of opposite-sign leptons are required to have an invariant mass close to the  $Z$  boson mass,  $75 \text{ GeV} < m_{\ell\ell} < 105 \text{ GeV}$ , and are sorted by their transverse momentum, with  $\ell_1$  being the hardest lepton. We chose  $p_T^{\ell_1}$ ,  $\Delta\phi_{\ell\ell}$  as kinetic variables for analysis. Widths and production rates of the two resonances are assumed to be the same, so that only differential distributions may be used to distinguish the spin of the resonance.

Following the same approach as for Case I, we assume that the projected data arises from a spin-2 resonance, and compute the Bayes factor  $B_{20}$  comparing the spin-2 hypothesis to spin-0 hypothesis. The 1D/2D gain in sample size is found to be  $\sim 50\%$  for  $p_T^{\ell_1}$  and  $\sim 75\%$  for  $\Delta\phi_{\ell\ell}$ . Thus we observe again a substantial gain in Fisher information when using the 2D distribution instead of individual 1D distributions. Similar results are obtained when assuming a spin 0 resonance and computing  $B_{02}$ . These features are in agreement with the expectations from Sec. 3.

The analysis of Case II gives a concrete idea of how much information gain can be realistically expected when going from 1D to 2D in a situation of observation of resonant signal. The information gain in terms of sample size  $n_{\text{obs}}$  can be read on Fig. 5. For instance, going from the  $\Delta\phi_{\ell\ell}$  distribution to the  $(p_T^{\ell_1}, \Delta\phi_{\ell\ell})$  distribution is equivalent to have a sample of events smaller by a factor  $\sim 0.66$ .

## 5. Conclusions

In the view of future new physics searches and characterisation at the LHC, we investigate the impact of multidimensional differential rates in typical LHC analyses. Through general observations based on Fisher information and Bayes factors, we find that in the occurrence of a discovery, the gain from using 2D differential distributions instead of 1D is fundamentally bounded. In contrast, for model discrimination, no such bound is found, thus the gain from 1D to 2D can be much higher. To illustrate these features and show realistic values of the 1D/2D information gain, we study two new physics scenarios: operators from the SMEFT, and bosonic resonances.

We carried out discovery and discrimination tests in the VBF channel in presence of CP-even and CP-odd operators. The best 1D/2D gain for discovery is found for the combination of variables  $(p_T^{\gamma\gamma}, \Delta\Phi_{\gamma\gamma})$ . As expected, the 1D/2D gain for CP discrimination is found to be much larger than for discovery. This observation also holds for various other choices of variables. In the presence of a heavy bosonic resonance, we evaluate the discrimination power of spin-0 versus spin-2 using the  $(p_T^{\ell\ell}, \Delta\Phi_\ell)$  variables, and observe a 1D/2D gain of about 50%. Overall, none of the 1D/2D gain we have observed exceed 100%. Note that the procedure of adding more differential information saturates as the kinematic information in a given final state is limited. Hence the gain from 2D to higher-dimensional distributions will be restricted due to correlations among the variables involved.

All details needed to reproduce our analysis are provided, and important subtleties generally present in these analyses are pointed out. First, in the reconstruction method of the differential rates in the SMEFT, we point out that in the low-event regions of phase space the Cauchy–Schwarz bound on the interference can be violated by the large systematic uncertainties, resulting in unphysical results. Second, we find that when using projected data, a cancellation between the leading uncertainties of expected and projected rates naturally occurs, implying that the maximum likelihood would remain unchanged and hence the MonteCarlo sample size would only have to be mildly larger than the nominal sample size to provide meaningful results.

We hope this study serves as a guide to experiments to provide differential information to theoretical collaborations, and as how theorists could use this information for model discrimination.

## Acknowledgements

The S.F. work was supported by the São Paulo Research Foundation (FAPESP) under grants #2011/11973 and #2014/21477-2. V.S. is supported by the Science Technology and Facilities Council (STFC) under grant number ST/L000504/1. F.F. is supported by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES).

## Appendix A. Information content of 1D/2D likelihoods, a Gaussian example

Here we illustrate the inequalities between 1D and 2D information contents in an analytic example. It is enough to focus on the shape likelihood associated to a single event, the case of various events can be worked out in a fully similar way. We choose the underlying event rate distribution along the  $X, Y$  variables to be a bivariate normal distribution. We further assume that the central values depend linearly of the parameter of interest  $\theta$ , hence

$$f_{X,Y} = \mathcal{N}(\mu, \Sigma), \quad \Sigma = \begin{pmatrix} \sigma_X^2 & \sigma_X \sigma_Y \rho \\ \sigma_X \sigma_Y \rho & \sigma_Y^2 \end{pmatrix},$$

$$\mu = \begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix} \equiv \begin{pmatrix} a_X + b_X \theta \\ a_Y + b_Y \theta \end{pmatrix}. \quad (\text{A.1})$$

It is also convenient to introduce the conditional density  $f_{X,Y} = f_X f_{Y|X}$  with

$$f_X = \mathcal{N}(\mu_X, \sigma_X^2),$$

$$f_{Y|X} = \mathcal{N}\left(\mu_Y - \frac{\sigma_Y}{\sigma_X} \rho (\mu_X - x), \sigma_Y^2 (1 - \rho^2)\right). \quad (\text{A.2})$$

In the  $\rho \rightarrow \pm 1$  limit, the conditional density

$$f_{Y|X} = \frac{1}{\sqrt{2\pi(1-\rho^2)}\sigma_Y} \exp\left(-\frac{\left(\mu_Y - \frac{\sigma_Y}{\sigma_X} \rho (\mu_X - x)\right)^2}{2\sigma_Y^2(1-\rho^2)}\right) \quad (\text{A.3})$$

tends to zero everywhere except where the argument of the exponential approaches zero. The limit is a Dirac distribution

$$f_{Y|X} \rightarrow \delta\left(y - \mu_Y \mp \frac{\sigma_Y}{\sigma_X} (\mu_X - x)\right), \quad \rho \rightarrow \pm 1 \quad (\text{A.4})$$

This has to be true for any  $\theta$ , hence the relation

$$\frac{b_X}{\sigma_X} = \pm \frac{b_Y}{\sigma_Y}, \quad \rho \rightarrow \pm 1 \quad (\text{A.5})$$

must hold in that limit.

Let us now study the information content of  $f_X, f_Y, f_{X,Y}$  with respect to the  $\theta$  parameter. For the 1D distributions we have simply

$$I_\theta^{1D,X} \equiv -\partial_\theta^2 \log f_X = \frac{b_X^2}{\sigma_X^2}, \quad I_\theta^{1D,Y} = \frac{b_Y^2}{\sigma_Y^2}. \quad (\text{A.6})$$

For the 2D distribution  $f_{X,Y}$ , away from  $\rho \sim 1$  we have in general

$$I_\theta^{2D} = -\partial_\theta^2 \log f_{X,Y} = (b_X \ b_Y) \Sigma^{-1} \begin{pmatrix} b_X \\ b_Y \end{pmatrix}$$

$$= \frac{1}{1-\rho^2} \left( \frac{b_X^2}{\sigma_X^2} + \frac{b_Y^2}{\sigma_Y^2} - 2\rho \frac{b_X b_Y}{\sigma_X \sigma_Y} \right). \quad (\text{A.7})$$

For  $\rho = 0$ , we recover the equality  $I_\theta^{2D} = I_\theta^{1D,X} + I_\theta^{1D,Y}$ . Away from  $\rho = 0$  we recover the upper bound on the 2D information, Eq. (13). Near  $\rho \sim 1$ , the relation Eq. (A.5) must hold, hence the Fisher information of  $f_X$  and  $f_Y$  are equal,  $I_\theta^{1D,X} = I_\theta^{1D,Y} = I_\theta^{1D}$ . Moreover the 2D Fisher information simplifies to

$$I_\theta^{2D} = -\partial_\theta^2 \log f_{X,Y} = \frac{2}{1+|\rho|} \left( \frac{b_{X,Y}^2}{\sigma_{X,Y}^2} \right) = \frac{2}{1+|\rho|} I_\theta^{1D}. \quad (\text{A.8})$$

Hence we have recovered the limit of full correlation for which the 2D distribution does not bring new information with respect to 1D. One can notice that in this simple model, whenever  $b_{X,Y}$  are independent of  $\rho$ , Eq. (A.5) holds for any  $\rho$  and so does Eq. (A.8). When it is the case one has  $I_\theta^{2D} = 2I_\theta^{1D}$  for  $\rho = 0$ , which is the maximal 1D/2D gain possible.

## Appendix B. Violation of the interference's upper bound

The interference component in the true rate  $\sigma_r(c) = \sigma_r^{SM} + c\sigma_r^{\text{int}} + c^2\sigma_r^{\text{BSM}}$  satisfies a bound based on Cauchy–Schwarz inequality,  $|\sigma_r^{\text{int}}| < 2\sqrt{\sigma_r^{\text{SM}}\sigma_r^{\text{BSM}}}$ . From a concrete viewpoint, this bound prevents the cross-section for becoming negative for any value of  $c$ . Now, on bins featuring few events, the uncertainty is large enough so that there is a non-negligible probability that the

Cauchy–Schwarz bound be violated, i.e.  $|\hat{\sigma}_{\text{int}}| < 2\sqrt{\hat{\sigma}_{\text{SM}}\hat{\sigma}_{\text{BSM}}}$  can be false. This then results in a negative rate on the bin for some interval of  $c$ .

If the simulation has been done with enough events compared to the nominal event number, for any bin and any  $c$ , these rare events regions have a negligible impact on the subsequent analysis. Violation of the Cauchy–Schwarz constitutes then only a practical obstruction, and it is convenient to simply remove the bins on which the Cauchy–Schwarz bound is not respected.<sup>10</sup>

### Appendix C. Cancellation of systematic uncertainties

When confronting the expected event numbers to observed event numbers, one has to make sure that the systematic uncertainty on the expected event numbers has to be negligible with respect to statistical uncertainty, for every bin and for both SM and BSM hypotheses (for any relevant value of  $c$ , for example). In practice, this means that the number of MC events has to be larger than the number of observed events in all of these situations.

However, when the analyses involve projected data instead of actual data, a very nice and useful feature appears. It turns out that, provided one uses the same estimates for the projected rates and for the expected ones, the respective systematic uncertainties present in these rates will approximately cancel each other.

Let us detail how this occurs. The uncertainties on the reconstructed rates take the form

$$\hat{\sigma}(c) = \hat{\sigma}_{\text{SM}}(1 + \delta_{\text{SM}}) + c\hat{\sigma}_{\text{int}}(1 + \delta_{\text{int}}) + c^2\hat{\sigma}_{\text{BSM}}(1 + \delta_{\text{BSM}}) \quad (\text{C.1})$$

where the  $\delta$ 's are the nuisance parameters – which are in general correlated (see [23] for their correlation matrix). Let us first adopt the Gaussian limit for simplicity. The likelihood using projected data takes the form  $\exp(-\mathcal{L}(\hat{\sigma}(c') - \hat{\sigma}(c))^2/2\hat{\sigma}(c))$ , in which it is clear that the SM uncertainty in the numerator cancels out exactly, and the other ones are suppressed by  $c - c'$  and  $c^2 - (c')^2$ . As a result, the maximum of the likelihood remains unchanged and the main effect of the uncertainty is to distort the Gaussian and the Fisher information.

But these features turn out to be rigorously valid beyond the Gaussian limit, for Poisson likelihoods with any number of events. To see this, one first combines the three nuisance parameters into a single one. This operation is rigorously defined, and has already lead to useful developments in the context of LHC analyses [24–26]. After combination, the event rate takes the form

$$\hat{\sigma}(c) = \hat{\sigma}_0(c)(1 + \delta\Delta(c)), \quad (\text{C.2})$$

where  $\delta$  is the nuisance parameter and  $\Delta(c)$  controls the relative magnitude of the uncertainty. The marginal Poisson likelihood

with projected data is

$$\bar{L}(c) = \int d\delta \pi(\delta) \frac{e^{-n(c)} n(c)^{n(c')}}{\Gamma(n(c') + 1)}, \quad (\text{C.3})$$

where  $n(c) = \mathcal{L}\hat{\sigma}(c)$  and  $\pi(\delta)$  is the prior for  $\delta$ . Computing the derivative of  $\bar{L}(c)$ , it turns out that the maximum of  $\bar{L}(c)$  still occurs for  $c = c'$ , in spite of the deformations induced by the uncertainties, and for any  $\pi(\delta)$ .

### References

- [1] P.A.R. Ade, et al., Planck, *Astron. Astrophys.* 594 (2016) A13, arXiv:1502.01589 [astro-ph.CO].
- [2] D. de Florian, et al., LHC Higgs Cross Section Working Group, 2016, arXiv:1610.07922 [hep-ph].
- [3] T. Aaltonen, et al., CDF, *Phys. Rev. Lett.* 103 (2009) 092002, arXiv:0903.0885 [hep-ex].
- [4] G. Aad, et al., ATLAS, CMS, *J. High Energy Phys.* 08 (2016) 045, arXiv:1606.02266 [hep-ex].
- [5] J. Ellis, V. Sanz, T. You, *J. High Energy Phys.* 07 (2014) 036, arXiv:1404.3667 [hep-ph].
- [6] G. Aad, et al., ATLAS, *Phys. Lett. B* 753 (2016) 69, arXiv:1508.02507 [hep-ex].
- [7] R. Trotta, *Contemp. Phys.* 49 (2008) 71, arXiv:0803.4089 [astro-ph].
- [8] R.D. Cousins, *Synthese* 194 (2017) 395, arXiv:1310.3791 [physics.data-an].
- [9] G. Cowan, K. Cranmer, E. Gross, O. Vitells, *Eur. Phys. J. C* 71 (2011) 1554, Erratum: *Eur. Phys. J. C* 73 (2013) 2501, arXiv:1007.1727 [physics.data-an].
- [10] S. Fichtel, *Phys. Rev. D* 86 (2012) 125029, arXiv:1204.4940 [hep-ph].
- [11] C. Degrande, C. Duhr, B. Fuks, D. Grellscheid, O. Mattelaer, T. Reiter, *Comput. Phys. Commun.* 183 (2012) 1201, arXiv:1108.2040 [hep-ph].
- [12] T. Sjöstrand, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H.S. Shao, T. Stelzer, P. Torrielli, M. Zaro, *J. High Energy Phys.* 07 (2014) 079, arXiv:1405.0301 [hep-ph].
- [13] T. Sjöstrand, S. Mrenna, P.Z. Skands, *J. High Energy Phys.* 05 (2006) 026, arXiv:hep-ph/0603175.
- [14] M. Cacciari, G.P. Salam, G. Soyez, *J. High Energy Phys.* 04 (2008) 063, arXiv:0802.1189 [hep-ph].
- [15] M. Cacciari, G.P. Salam, G. Soyez, *Eur. Phys. J. C* 72 (2012) 1896, arXiv:1111.6097 [hep-ph].
- [16] E. Conte, B. Fuks, G. Serret, *Comput. Phys. Commun.* 184 (2013) 222, arXiv:1206.1599 [hep-ph].
- [17] R. Contino, M. Ghezzi, C. Grojean, M. Muhlleitner, M. Spira, *J. High Energy Phys.* 07 (2013) 035, arXiv:1303.3876 [hep-ph].
- [18] F. Ferreira, B. Fuks, V. Sanz, D. Sengupta, arXiv:1612.01808 [hep-ph], 2016.
- [19] J. Ellis, V. Sanz, T. You, *J. High Energy Phys.* 03 (2015) 157, arXiv:1410.7703 [hep-ph].
- [20] A. Alloul, N.D. Christensen, C. Degrande, C. Duhr, B. Fuks, *Comput. Phys. Commun.* 185 (2014) 2250, arXiv:1310.1921 [hep-ph].
- [21] L. Brenner, N. Belyaev, V. Bortolotto, C.D. Burgard, M. Duehrssen-Debling, K.M. Ecker, S. Gadatsch, D.S. Gray, A. Kaluza, K. Koeneke, R. Konoplich, S. Kortner, K. Prokofiev, C. Schmitt, W. Verkerke, Report No. LHCHSWG-INT-2016-004, 2016, <http://cds.cern.ch/record/2143180>.
- [22] A Morphing Technique for Signal Modelling in a Multidimensional Space of Coupling Parameters, Tech. Rep. ATL-PHYS-PUB-2015-047, CERN, Geneva, 2015.
- [23] S. Fichtel, P. Rebello Teles, A. Toner, arXiv:1611.01165 [hep-ph], 2016.
- [24] S. Fichtel, G. Moreau, *Nucl. Phys. B* 905 (2016) 391, arXiv:1509.00472 [hep-ph].
- [25] S. Fichtel, *Nucl. Phys. B* 911 (2016) 623, arXiv:1603.03061 [hep-ph].
- [26] A. Arbey, S. Fichtel, F. Mahmoudi, G. Moreau, *J. High Energy Phys.* 11 (2016) 097, arXiv:1606.00455 [hep-ph].
- [27] J. Brehmer, K. Cranmer, F. Kling, T. Plehn, 2016, arXiv:1612.05261 [hep-ph].
- [28] A. Wald, *Trans. Am. Math. Soc.* 54 (1943) 426.

<sup>10</sup> Another way of reconstructing the differential rate is to use the recently implemented option in Madgraph 5 that removes the  $\sigma_{\text{BSM}}$  term from the amplitudes. We found a large amount of bins violating the Cauchy–Schwarz bound when using this prescription, and thus did not pursue with this method.