IL NUOVO CIMENTO **39 C** (2016) 220
DOI 10.1393/ncc/i2016-16220-3

Colloquia: IFAE 2015

Higgs pair production: Choosing benchmarks with cluster analysis

Martino Dall'Osso(1)(2)

(1) *Dipartimento di Fisica e Astronomia and INFN, Sezione di Padova - Padova, Italy*
(2) *CERN - 1211 Geneva 23, Switzerland*

received 7 January 2016

Summary. — New physics theories often depend on a significant number of parameters and the phenomenology of fundamental processes may vary drastically depending on their values. It is advantageous to categorize the resulting final state kinematics such that the investigation of a few typical points yields knowledge on a large area of the parameter space. Here it is showed how this can be done effectively with a cluster analysis, by considering as specific process the non-resonant Higgs boson pairs production in the context of Standard Model extensions to Higgs boson anomalous couplings.

In the present work we consider the non-resonant production of Higgs boson pairs with the extension to any kind of coupling deviation from the SM Higgs sector, resulting in a five-dimensional parameter space to be investigated. Little modifications of some of the couplings lead to drastical changes of the di-Higgs signal topology due to interference effects. This limits the validity of the search to a small neighbourhood of the investigated point without the possibility to extend the result in a wide region of the parameters space. We thus propose an approach based on the identification of a small set of benchmark points which are maximally representative of the largest possible volume of the parameter space. We identify sets of parameters that yield similar final-state kinematics and this simplifies the problem of investigating a large and unconstrained model space. The gluon-gluon fusion (GF) production of Higgs boson pairs at LHC can be generally parametrized with five parameters in a tree-level effective field theory. Written in terms of physical states as in eq. (1), these five parameters are κ_λ, κ_t, c_g, c_{2g}, and c_2. The κ_λ and κ_t are multiplicative factors that parametrize the deviations from the SM value of the Higgs boson trilinear coupling and the top Yukawa interaction, respectively. The contact interactions with gluons and the contact interaction involving two Higgs bosons are instead not predicted by the SM and they are parametrized by the absolute couplings c_g, c_{2g}, and c_2:

$$(1)\quad \mathcal{L}_h = \frac{1}{2}\partial_\mu h\partial^\mu h - \frac{1}{2}m_h^2 h^2 - \kappa_\lambda\,\lambda_{SM} v\, h^3 - \frac{m_t}{v}\left(v + \kappa_t\, h + \frac{c_2}{v}\, h\, h\right)(\bar{t}_L t_R + \text{h.c.}) + \frac{1}{4}\frac{\alpha_s}{3\pi v}\left(c_g\, h - \frac{c_{2g}}{2v}\, h\, h\right) G^{\mu\nu}G_{\mu\nu}\,.$$

We assume the Higgs boson couplings to light fermions to be negligible, and the absence of any other light state in addition to the SM particles. We also ignore CP-violating BSM effects. The following constraints on the parameters are initially set by considering single Higgs boson measurements and absolute cross section limits on inclusive di-Higgs production: $|\kappa_\lambda| < 10$; $0.5 < \kappa_t < 2.5$; $c_g \sim O(1)$. Instead of using a regularly spaced grid to get the initial sampling of the parameters space, we focus on the local minima of the production cross section where the probability densities of the final state observables exhibit the fastest variability with parameter variation. The initial grid includes 1507 points of the five-dimensional parameter space and a generator level information is produced for each of them. The process we are considering is a $2 \to 2$ process at leading order. The two Higgs bosons are produced with identical transverse momenta, and they are back-to-back in azimuth. Thus, the final state can be completely defined by just two kinematical variables: the invariant mass of the di-Higgs system (m_{hh}) and the modulus of the cosine of the polar angle of one Higgs boson with respect to the beam axis ($|\cos\theta^*|$). We proceed with the classification of the initial samples, based on the similarity of the event kinematics, by using a test statistic which exploits these two variables in each event. Due to the numerosity of our generated datasets (20k events per sample) and the small dimensionality of the feature space, that completely defines the final state of the process, we employ as test statistic a simple log-likelihood function constructed as a product of $N_{bin}^{tot} = 50 * 5$ Poisson terms. The maximum likelihood estimate for the expected contents in a given bin i, which the two samples populate respectively with $n_{i,1}$ and $n_{i,2}$ entries, is $\hat{\mu_i} = (n_{i,1} + n_{i,2})/2$. Given this, the two samples log-likelihood can then be written as

$$TS = 2\log\left(\frac{L}{L_S}\right) = -2\sum_{i=1}^{N_{\rm bins}}\left[\log(n_{i,1}!) + \log(n_{i,2}!) - 2\log\left(\frac{n_{i,1}+n_{i,2}}{2}!\right)\right]. \tag{2}$$

The TS value is computed for each sample and used iteratively by the final clustering procedure, producing a grouping of the parameter space points based on the kinematic densities of the corresponding final states. The algorithm we chose allows to univocally identify the sample in each cluster which is the most representative of the set - what we call a *benchmark*. The free parameter of this procedure is the final number of clusters we want to obtain. This is obviously linked to the similarity of the benchmark with each sample of the cluster. In our case study, a good uniformity within each clusters is obtained with twelve final clusters and relative benchmarks. A complete description of the cluster analysis and of the theoretical framework is reported in [1]. This procedure is mainly automatic and it exploits generator level information only, with no necessity of a wide production of fully reconstructed Monte Carlo samples. It allows us to study the kinematic behaviour of the Higgs boson pair inside a five dimensional space and to group the initial 1507 samples in twelve clusters, each one represented by a unique benchmark. A physical analysis based on this restricted number of benchmarks will thus provide information on a wide range of the 5D parameter space. Such a general approach to select physics benchmarks has never been adopted before in a new physics search context and, even if it has been developed *ad hoc* for the particular case study, it could find application in many other searches.

REFERENCES

[1] Carvalho Alexandra *et al.*, *JHEP*, **2016** (2016) 4.