Probing the *CP* nature of the Higgs coupling in *t̄th* events at the LHC

S. Amor dos Santos,[1] M. C. N. Fiolhais,[1,2] R. Frederix,[3] R. Gonçalo,[4] E. Gouveia,[5]
R. Martins,[5] A. Onofre,[5] C. M. Pease,[5] H. Peixoto,[6] A. Reigoto,[5] R. Santos,[5,7,8] and J. Silva[6]

[1]*LIP, Departamento de Física, Universidade de Coimbra, 3004-516 Coimbra, Portugal*
[2]*Department of Science, Borough of Manhattan Community College,*
City University of New York, 199 Chambers St, New York, New York 10007, USA
[3]*Physik Department T31, Technische Universität München,*
James-Franck-Str. 1, D-85748 Garching, Germany
[4]*LIP, Av. Elias Garcia, 14-1, 1000-149 Lisboa, Portugal*
[5]*LIP, Departamento de Física, Universidade do Minho, 4710-057 Braga, Portugal*
[6]*Centro de Física, Universidade do Minho, Campus de Gualtar, 4710-057 Braga, Portugal*
[7]*Instituto Superior de Engenharia de Lisboa—ISEL, 1959-007 Lisboa, Portugal*
[8]*Centro de Física Teórica e Computacional, Faculdade de Ciências, Universidade de Lisboa,*
Campo Grande, Edifício C8 1749-016 Lisboa, Portugal

The determination of the *CP* nature of the Higgs coupling to top quarks is addressed in this paper, using $t\bar{t}h$ events produced in $\sqrt{s} = 13$ TeV proton-proton collisions at the LHC. Dileptonic final states are employed, with two oppositely charged leptons and four jets, corresponding to the decays $t \rightarrow bW^+ \rightarrow b\ell^+\nu_\ell$, $\bar{t} \rightarrow \bar{b}W^- \rightarrow \bar{b}\ell^-\bar{\nu}_\ell$, and $h \rightarrow b\bar{b}$. Pure scalar ($h = H$), pure pseudoscalar ($h = A$), and *CP*-violating Higgs boson signal events, generated with MADGRAPH5_aMC@NLO, are fully reconstructed through a kinematic fit. We furthermore generate samples that have both a *CP*-even and a *CP*-odd component in the $t\bar{t}h$ coupling in order to probe the ratio of the two components. New angular distributions of the decay products, as well as *CP* angular asymmetries, are explored in order to separate the scalar from the pseudoscalar components of the Higgs boson and reduce the contribution from the dominant irreducible background, $t\bar{t}b\bar{b}$. Significant differences between the angular distributions and asymmetries are observed, even after the full kinematic fit reconstruction of the events, allowing to define the best observables for a global fit of the Higgs couplings parameters.

I. INTRODUCTION

In July 2012, the discovery of a Higgs boson, predicted by the electroweak symmetry breaking mechanism [1] of the Standard Model (SM) of particle physics, with a mass close to 125 GeV, was announced by both the ATLAS [2] and CMS [3] collaborations. Since then, studying the Higgs boson's properties has motivated many physics analyses at the LHC. So far, the measured properties of the Higgs boson have shown remarkable consistency with those predicted by the SM [4]. Nevertheless, it is by now clear that the SM cannot explain all of the observed physical phenomena. One of the best known examples is that it fails to explain the matter-antimatter asymmetry of the Universe, for which new sources of *CP*-violation beyond the SM (BSM) are required. One possibility would be to introduce *CP* violation in the Higgs sector. This is allowed in BSM models, such as supersymmetry and two-Higgs doublet models, where the Higgs boson(s) have no definite *CP* quantum number resulting in a Yukawa coupling with two components, one *CP*-even and one *CP*-odd (see, for instance, Ref. [5]).

Analyses focusing on the Higgs boson decays to photons, ZZ, and WW, as well as on the VH ($V = W$, Z) associated production have been conducted to measure its spin and parity quantum numbers [6–8]. All of the results are consistent with a SM-like spin-0, parity-even boson, while the pure pseudoscalar scenario has been excluded at the 99.98% confidence level (C.L.). However, the possibility of a *CP* admixture manifestation in the Yukawa couplings remains to be probed directly. So far, only *CP*-odd components of the Higgs couplings to the weak gauge bosons were shown to be very small. Within all fermions, the top quark is expected to have the largest Yukawa coupling. Currently, this coupling can be measured indirectly from loop effects in $gg \rightarrow h$ and $h \rightarrow \gamma\gamma$, which suffer from large systematic uncertainty and require the assumption of no BSM contributions to the loops. This motivates the interest in associated production of the Higgs boson with a top quark pair ($t\bar{t}h$) [9], which allows for a direct measurement of the top Yukawa coupling and provides sensitivity to its *CP* nature, through the rich kinematics of the events.

The main background contaminating $t\bar{t}h$ searches at the LHC is $pp \rightarrow t\bar{t} + $ jets. In particular, if the dominant Higgs decay channel ($h \rightarrow b\bar{b}$) is analyzed, $t\bar{t}b\bar{b}$ is a challenging irreducible background. Several $t\bar{t}h$ decay channels have

been studied [10–15]. The very complex final states, together with the huge backgrounds, make it a particularly difficult Higgs process to study at the LHC. Nevertheless, both the ATLAS and CMS collaborations have reached remarkable sensitivities, with expected upper limits at 95% C.L. for the $t\bar{t}H$ signal strength, μ, below 2 in the background-only scenario. The best-fit values obtained for μ were 1.7 ± 0.8 by ATLAS [10] and 2.8 ± 1.0 by CMS [14]. Combined results from both collaborations and from the various Higgs analyses were used to fit the signal strengths of five Higgs production processes, while assuming SM-like Higgs branching ratios [16]. The best-fit value obtained for $\mu(t\bar{t}H)$ was $2.3^{+0.7}_{-0.6}$.

In the present work, we address the dileptonic final state of $t\bar{t}$ with the Higgs boson decaying through $h \rightarrow b\bar{b}$. The two leptons in the final state make it a fairly clean channel, with the advantage that they preserve spin information from the top quarks. We investigate possible departures from the SM nature of the Higgs boson by comparing the kinematics of $t\bar{t}h$ signal samples with the SM Higgs boson ($h = H$ and $J^{CP} = 0^+$) to samples of the $t\bar{t}h$ signal with a pure pseudoscalar Higgs boson ($h = A$ and $J^{CP} = 0^-$). Furthermore, we use a general Yukawa coupling for the top quark defined as

$$\mathcal{L} = \kappa y_t \bar{t}(\cos\alpha + i\gamma_5 \sin\alpha)th, \tag{1}$$

where y_t is the SM Higgs Yukawa coupling and α represents a CP phase. This approach allows us to probe the mixing between the CP-even and the CP-odd components of the top quark Yukawa coupling to the 125 GeV Higgs. Note that with this Lagrangian h has no definite CP quantum number. The SM interaction is recovered for $\cos\alpha = \pm 1$, while the pure pseudoscalar is obtained by setting $\cos\alpha = 0$.

Several observables in $t\bar{t}h$ events, sensitive to the CP nature of the top Yukawa coupling, have been proposed from which we will study in detail the ones presented in Refs. [17–19] (other proposals including observables probing the CP nature of the $\tau^+\tau^-h$ coupling were also discussed in Refs. [20,21]). While some rely on leptons in the dileptonic final state, more general observables are obtained from the particles at production (t, $\bar{t}$, and h), which are only accessible experimentally through a reconstruction algorithm.

A full kinematical reconstruction method is applied to recover the four-momenta of the undetected neutrinos from the W-boson decays, and a large set of new angular observables is presented. We will show that the information that is present in the matrix elements partially survives parton showering, detector simulation, event selection, and event reconstruction. It has been suggested [22,23] that the different CP states of h in signal and g in the $t\bar{t}b\bar{b}$ background (g being a gluon which splits into $b\bar{b}$) can be exploited for background discrimination, through

differences in angular distributions. In Ref. [19], we presented a set of interesting observables for that effect, and we will demonstrate similar discriminating power for some of the observables introduced here. Even though we start by considering only the irreducible $t\bar{t}b\bar{b}$ background, without a highly optimized event reconstruction method, we present results with a complete set of SM backgrounds and argue that our findings are also valid in a more general and realistic case. For other observables in this set, two signal samples—one with a scalar Higgs H and another with a pseudoscalar Higgs A—are also differently distributed, suggesting that the observables can be used to probe the CP nature of the top Yukawa coupling.

II. EVENT GENERATION, SIMULATION, AND RECONSTRUCTION

The $t\bar{t}h$ signal events as well as the dominant background process ($t\bar{t}b\bar{b}$) were generated at next-to-leading order (NLO) in QCD, using MADGRAPH5_AMC@NLO [24] with the NNPDF2.3 PDF sets [25]. The SM signal was generated using the default sm model in MADGRAPH_AMC@NLO. The samples in which the Higgs has a nonzero CP-odd component were generated using the HC_NLO_X0 model, described in Ref. [26]. Signal samples were generated for values of $\cos\alpha$ ranging from -1 to 1 (in steps of 0.1). The model also allows the adjustment of effective couplings between the Higgs boson and vector bosons. Since $t\bar{t}h$ associated production with subsequent $h \rightarrow b\bar{b}$ decay is considered, those were all set to 0 (with the exceptions of $H_{\gamma\gamma}$, $A_{\gamma\gamma}$, $H_{Z\gamma}$, and $A_{Z\gamma}$). For this analysis, contributions from both the dominant background ($t\bar{t}b\bar{b}$) and other SM processes were taken into account. Samples of $t\bar{t} + \text{jets}$ (where jets stands for up to three additional c jets or light-flavored jets), $t\bar{t}V + \text{jets}$ (where V can be either Z or $W^{\pm}$ and jets can go up to one additional jet), single top quark production (t-channel, s-channel, and Wt with up to one additional jet), diboson ($WW, WZ, ZZ + \text{jets}$ with up to three additional jets), $W + \text{jets}$ and $Z + \text{jets}$ (with up to four additional jets), and $Wb\bar{b} + \text{jets}$ and $Zb\bar{b} + \text{jets}$ (with up to two additional jets) were generated at LO with MADGRAPH5_AMC@NLO [24]. While the $t\bar{t} + \text{jets}$ sample was normalized to the QCD next-to-next-to-leading-order (NNLO) cross section with next-to-next-to-leading-logarithmic resummation of soft gluons [25,27–30], the single top quark production cross section was scaled to the approximate NNLO theoretical predictions [31,32], assuming the NNPDF2.3 PDF sets and scaled according to the generated top quark mass, following the prescription defined in Ref. [33].

The full spin correlations information of the $t \rightarrow bW^+ \rightarrow b\ell^+\nu_\ell$, $\bar{t} \rightarrow \bar{b}W^- \rightarrow \bar{b}\ell^-\bar{\nu}_\ell$, and $h \rightarrow b\bar{b}$ decays, with $\ell^\pm \in \{e^\pm, \mu^\pm\}$, is preserved by using MADSPIN [23] to perform the decay chain of top quarks and Higgs bosons. All events were generated for LHC pp

collisions, with a center-of-mass energy of 13 TeV, with nonfixed renormalization and factorization scales set to the sum of the transverse masses of all final-state particles. The masses of the top quark (m_t), the W boson (m_W), and Higgs bosons (for both scalar, m_H, and pseudoscalar, m_A) were set to 173, 80.4, and 125 GeV, respectively.

The events were then passed through PYTHIA6 [34] for parton shower and hadronization. Matching between the generator and the parton shower was performed using the MLM [35] scheme for LO events and the MC@NLO [36] matching for NLO events. The DELPHES [37] package was then used for a fast simulation of a general-purpose collider experiment, using the default ATLAS parameter card. During detector simulation, jets and charged leptons are reconstructed, as well as the transverse missing energy. The efficiencies and resolutions of the detector subsystems are parametrized in segments of p_T (or E) and η. Particle tracking only occurs in the $|\eta| \leq 2.5$ region, and its efficiency for a particle with $p_T = 1$ GeV is, at least, 85% for charged hadrons and 83% (98%) for electrons (muons). The momentum resolution of a track is at most 5%. Calorimeters are segmented in (η, ϕ) rectangular cells. In the region with $|\eta| \leq 2.5$, the cells have dimensions $(\eta, \phi) = (0.1, 10°)$, and for $2.5 < |\eta| \leq 4.9$, their size is $(\eta, \phi) = (0.2, 20°)$. Electron and muon identification efficiencies are 95% in the central region $|\eta| \leq 1.5$, 85% in the intermediate region $1.5 < |\eta| \leq 2.5$ (2.7 for muons), and zero for $|\eta| > 2.5$ (2.7 for muons) or $p_T < 10$ GeV. Energy resolution for an electron with $E = 25$ GeV and with $|\eta| \leq 3.0$ is 1.5%, and it drops asymptotically to 0.5% for higher energies. The muon momentum resolution is worse for higher p_T and higher $|\eta|$, with its maximum at 10%, for $p_T > 100$ GeV and $1.5 < |\eta| \leq 2.5$. Jet reconstruction uses the anti-k_t algorithm [38] with the R parameter set to 0.6. The efficiency for b-tagging is given separately for b jets and c jets, as an asymptotically increasing function of p_T. For b jets (c jets), the b-tagging efficiency is limited to 50% (20%) in the $|\eta| \leq 1.2$ region and to 40% (10%) in the $1.2 < |\eta| \leq 2.5$ region. It is zero for jets with $p_T \leq 10$ GeV or $|\eta| > 2.5$. For any other jet, a constant b-tagging misidentification rate was set to 0.1%.

The analysis of the generated and simulated events was performed with MADANALYSIS 5 [39] in the expert mode [40]. Events are selected if at least four reconstructed jets and exactly two oppositely charged leptons with transverse momentum $p_T \geq 20$ GeV and pseudorapidity $|\eta| \leq 2.5$ are present. After selection, 16% (17%) of $t\bar{t}H$ ($t\bar{t}A$) signal events are accepted. No cuts are applied to the events' transverse missing energy ($\not{E}_T$). The full kinematic reconstruction of the four-momenta of the undetected neutrinos is performed by imposing energy-momentum conservation and mass constraints to signal and background events [19]. Mass values are randomly generated for the intermediate particles W^+, W^-, t, and $\bar{t}$, using probability density functions (p.d.f.s) obtained from the

corresponding generator-level mass distributions. First, a two-dimensional p.d.f. for m_t and $m_{\bar{t}}$ is used to generate random mass values for the top quarks. Second, m_{W^+} and m_{W^-} are generated from the two-dimensional p.d.f.s of (m_t, m_{W^+}) and $(m_{\bar{t}}, m_{W^-})$, respectively, such that possible correlations are preserved in the reconstruction. The following mass constraints are then applied to the $t\bar{t}$ system:

$$(p_{\ell^+} + p_\nu)^2 = m_{W^+}^2, \tag{2}$$

$$(p_{\ell^-} + p_{\bar{\nu}})^2 = m_{W^-}^2, \tag{3}$$

$$(p_{W^+} + p_b)^2 = m_t^2, \tag{4}$$

$$(p_{W^-} + p_{\bar{b}})^2 = m_{\bar{t}}^2. \tag{5}$$

The p_b and $p_{\bar{b}}$ correspond to the four-momenta of the two b jets, respectively, from the t and $\bar{t}$ decays. The p_{ℓ^+} and p_{ℓ^-} (p_ν and $p_{\bar{\nu}}$) correspond to the four-momenta of the positive and negative charged leptons (neutrino and antineutrino), respectively, from the decaying W^+ and W^-, which in turn have momenta p_{W^+} and p_{W^-}. In order to reconstruct the neutrino and antineutrino four-momenta (six unknowns, since we set $m_\nu = m_{\bar{\nu}} = 0$), we assume they fully account for the missing transverse energy, i.e.,

$$p_x^\nu + p_x^{\bar{\nu}} = \not{E}_x, \tag{6}$$

$$p_y^\nu + p_y^{\bar{\nu}} = \not{E}_y. \tag{7}$$

The $\not{E}_x$ and $\not{E}_y$ represent the x and y components of the transverse missing energy. If a solution is not found for the particular choice of top quark and W-boson masses, the generation of mass values is repeated, up to a maximum of 500, until at least one solution is found. If still no solution is found, the event is discarded as not compatible with the topology under study.

The kinematic reconstruction based on Eqs. (2)–(7) may result in more than one possible solution for a particular event and choice of masses. We calculate, for each solution, the likelihood ($L_{t\bar{t}h}$) of it being consistent with a $t\bar{t}h$ dileptonic event. This likelihood is computed as the product of one-dimensional p.d.f.s built from p_T distributions of the neutrino, antineutrino, top quark, antitop quark, and $t\bar{t}$ system [$P(p_{T\nu})$, $P(p_{T\bar{\nu}})$, $P(p_{Tt})$, $P(p_{T\bar{t}})$, and $P(p_{Tt\bar{t}})$, respectively], all obtained from fits to the corresponding parton-level distributions. The two-dimensional p.d.f. of the top quark masses, $P(m_t, m_{\bar{t}})$, and the one-dimensional p.d.f. of the Higgs candidate mass, $P(m_h)$, are also included. The latter is obtained at reconstruction level, using a ΔR criterion[1] to match jets to the truth-level b and $\bar{b}$ partons from the h decay.

---

[1]$\Delta R = \sqrt{\Delta\Phi^2 + \Delta\eta^2}$, where $\Delta\Phi$ ($\Delta\eta$) corresponds to the difference in Φ (η) between two objects.

$$L_{t\bar{t}h} \sim \frac{1}{p_{T\nu} p_{T\bar{\nu}}} P(p_{T\nu}) P(p_{T\bar{\nu}})$$

$$\times P(p_{T_t}) P(p_{T\bar{t}}) P(p_{T t\bar{t}}) P(m_t, m_{\bar{t}}) P(m_h). \quad (8)$$

The momenta of the neutrino and antineutrino must accommodate any energy losses in the event (QCD radiation, as well as detector effects) in order to reconstruct the top quarks and W boson masses. This may result in larger estimated neutrino and antineutrino p_T after reconstruction, relative to their p_T at parton level. In order to compensate for this effect, the factor $1/(p_{T\nu} \times p_{T\bar{\nu}})$ is introduced in the likelihood, thus favoring solutions with lower neutrino and antineutrino p_T that better match the parton level. The solution with the largest value of $L_{t\bar{t}h}$ is chosen as the correct one. A solution is found for 70% of truth-matched $t\bar{t}H$ and $t\bar{t}A$ signal events.

At reconstruction level (without truth-match), the number of combinations of jets available to reconstruct the top and antitop quarks, together with the Higgs boson, can be overwhelming. Choosing one of the wrong combinations of jets for reconstructing signal events gives rise to combinatorial background, which is one of the main challenges of this analysis. To reduce the number of possible combinations only the six highest p_T jets are used (it was confirmed that in more than 95% of all signal events, for both $t\bar{t}H$ and $t\bar{t}A$, jets produced from the top quarks and Higgs boson decays are within the six highest p_T jets). Furthermore, the jet combinations were required to verify $m_{\ell^+ b_t} < 150$ GeV, $m_{\ell^- \bar{b}_{\bar{t}}} < 150$ GeV and $50\,\text{GeV} \le m_{b_H \bar{b}_H} \le 200\,\text{GeV}$, where

b_t and $\bar{b}_{\bar{t}}$ refer to the jets assigned in reconstruction to the hadronization of the b and $\bar{b}$ quarks from the t and $\bar{t}$ decays, respectively. The motivation for these cuts results from the mass distribution shapes observed after the effects of detector simulation and selection cuts, which allows to avoid contributions from the tails.

At the reconstruction level (without truth-match), in order to preferentially pick the correct combination among the ones surviving the previous requirements, several multivariate methods were trained, using TMVA [41]. The correct and wrong jet combinations were labeled, respectively, signal and combinatorial background in the following procedure. Nine parton-level variables were used as input for the methods: ΔR, and the lab-frame angles $\Delta\theta$ and $\Delta\Phi$ between the particle pairs (b_t, ℓ^+), $(\bar{b}_{\bar{t}}, \ell^-)$, and $(b_H, \bar{b}_H)$. The invariant masses of the systems composed of these pairs were also included, but were computed at the reconstruction level with truth-match, to take into account detector resolution effects. A sample of $t\bar{t}h$ events (with $h = H$) was used to create both the signal and combinatorial background samples for this training and testing. For the signal sample, the variables were computed once per event, using the correct jet combination. For the combinatorial background sample, three different variable entries took place per event, each one corresponding to a wrong permutation of the four b and $\bar{b}$ partons. These three permutations are chosen such that all the variables computed in each permutation are different from the ones in any other, including the correct one. In Figs. 1 and 2 (left),
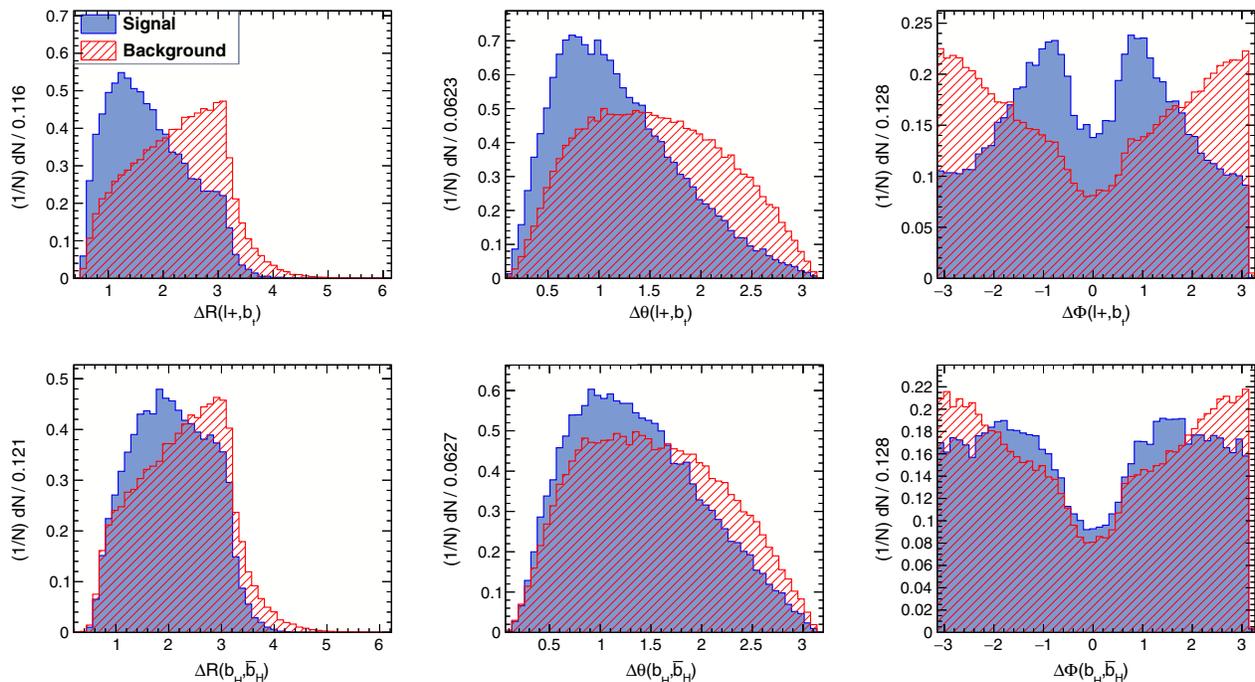
FIG. 1. Distributions of TMVA input variables for right (filled blue, labeled "Signal") and wrong combinations (red shaded, labeled "Background") of jets and leptons from the same parent decaying particle: $\Delta R(\ell^+, b_t)$ (top left) and $\Delta R(b_H, \bar{b}_H)$ (bottom left); $\Delta\theta(\ell^+, b_t)$ (top middle) and $\Delta\theta(b_H, \bar{b}_H)$ (bottom middle); $\Delta\Phi(\ell^+, b_t)$ (top right) and $\Delta\Phi(b_H, \bar{b}_H)$ (bottom right). See text for details.
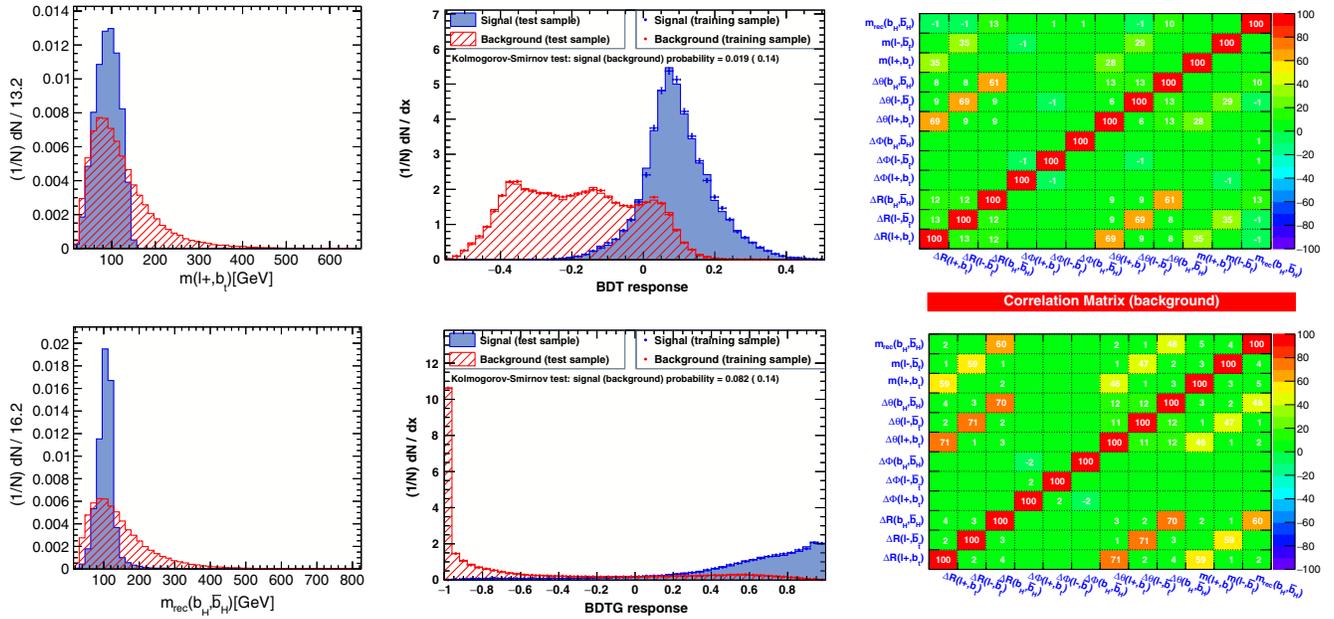
FIG. 2. Mass distributions (left) for right (filled blue, signal) and wrong (red shaded, background) combinations of jets and leptons from the same parent decaying particle: (upper-left) the $m(\ell^+, b_t)$ and (lower-left) $m(b_H, \bar{b}_H)$; (middle top) BDT and (middle bottom) BDTG TMVA methods response for signal and background; (right-top) TMVA input variable correlations for signal and (right-bottom) background.

distributions of the input variables are shown for the signal and combinatorial background training samples. The correlations between variables are shown in Fig. 2 (right), for the signal (top) and combinatorial background (bottom) samples. Two boosted decision trees showed the best performance, one with an adaptive boost (BDT) and the other with a gradient boost (BDTG). The latter being slightly better, it was used in the full kinematic reconstruction of events in order to increase the correct jet assignment. Figure 2 (middle column) shows the distributions of the BDT (top) and BDTG (bottom) discriminants for the signal and for the combinatorial background, for both the training and test samples. The jet combination chosen is the one returning the highest value of the BDTG discriminant, maximizing signal purity. After event selection, 62% (61%) of $t\bar{t}H$ ($t\bar{t}A$) signal events are successfully reconstructed. In 31% (34%) of the $t\bar{t}H$ ($t\bar{t}A$) signal events, the reconstruction without truth-match results in the same jet combination as the truth-matched one. Figure 3 shows, after $t\bar{t}H$ reconstruction without truth-match, two-dimensional p_T distributions of the W^+ (top left), the top quark (top right), the $t\bar{t}$ system (bottom left), and the Higgs boson (bottom right). The correlation between the parton-level p_T distributions (x axis) and reconstructed ones without truth-match (y axis), is clearly visible. The neutrino reconstructed p_T is compared with the parton level at NLO + Shower in Fig. 4 (left) and the distribution of the reconstructed Higgs boson mass is shown in Fig. 4 (right). In spite of the wider spread of values in the neutrino p_T distribution (which is a direct

consequence of the reconstruction of two neutrinos in each of the events), good correlation between the NLO + Shower distribution and the reconstructed neutrino p_T is observed. The distribution of the Higgs mass has a rms of order 20 GeV. Although reconstruction could be improved by using more elaborate methods, this stays outside the scope of the paper.

III. $t\bar{t}H$, $t\bar{t}A$, AND $t\bar{t}b\bar{b}$ ANGULAR DISTRIBUTIONS

As was done in Ref. [19], we define θ_Y^X as the angle between the direction of the Y system in the rest frame of X and the direction of the X system, in the rest frame of its parent system. For the reconstruction of the signal angular distributions, we consider the decay chain that starts with the $t\bar{t}h$ system, labeled (123), and goes through successive two-body decays, i.e., $(123) \to 1 + (23)$, $(23) \to 2 + (3)$, and $(3) \to 4 + 5$. Three families of observables are constructed: $f(\theta_1^{123})g(\theta_4^3)$, $f(\theta_1^{123})g(\theta_3^{23})$, and $f(\theta_3^{23})g(\theta_4^3)$, with $f, g = \{\sin, \cos\}$. The (123) system momentum direction is measured with respect to the laboratory frame. Particles 1 to 3 can either be the t or the $\bar{t}$ quarks, or even the Higgs boson, without repetition. Particle 4 can be any of the products of the decay of the top quarks and the Higgs boson, including the intermediate W bosons. The boost of particle 4 to the center of mass of particle 3 can be performed in two different ways: (i) using the laboratory four-momentum of both particles 3 and 4 (*direct* boost), or (ii) boosting particles 3 and 4 sequentially through all intermediate center-of-mass systems until particle 4 is
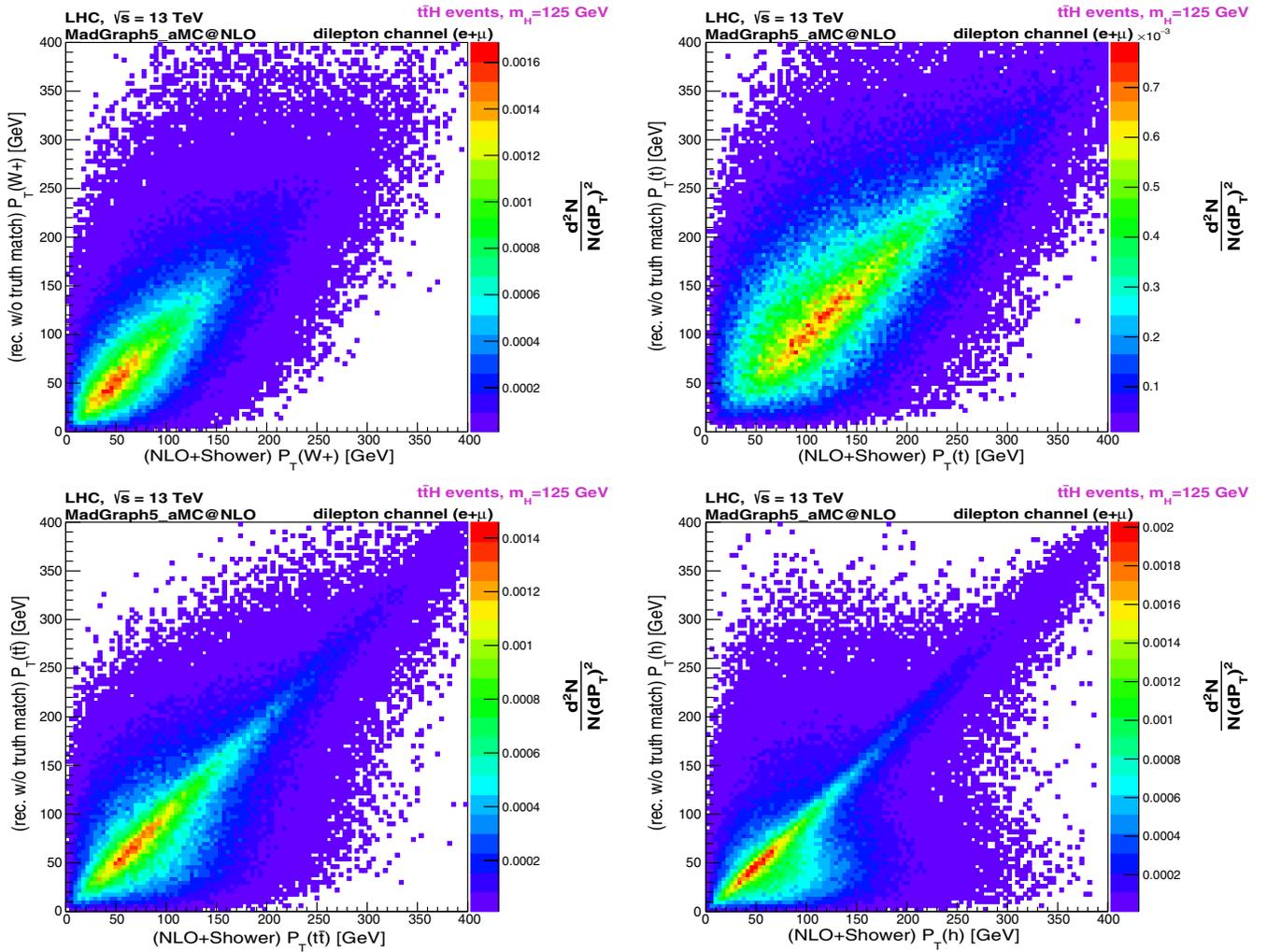
FIG. 3. Two-dimensional distributions of p_T in $t\bar{t}H$ events. The horizontal axes represent variables recorded at NLO + Shower, and the vertical axes represent the corresponding variables recorded at reconstruction level without truth-match. Upper-left: Distribution for W^+. A similar distribution is obtained for W^-, but is not shown here. Upper-right: Distribution for t. A similar distribution is obtained for $\bar{t}$, but is not shown here. Lower-left: Distribution for $t\bar{t}$. Lower-right: Distribution for H.

evaluated in the center-of-mass frame of particle 3 (*sequential* boost or *seq.* boost). Due to Wigner rotations, the directions of particle 4 resulting from each of these boosting procedures are different. The observables addressed in this work were studied using both the sequential and direct prescriptions.

A. NLO versus LO comparison

The impact of NLO corrections on the angular distributions is shown in Fig. 5 (left), by comparing with the LO $x_Y = \cos(\theta_H^{\bar{t}H})\cos(\theta_{\ell^-}^H)$, at the parton level (including shower effects) without any cuts, both for the SM $t\bar{t}H$ signal and $t\bar{t}b\bar{b}$ background events. NLO (LO) corrections with the impact of shower effects are labeled NLO + Shower (LO + Shower) throughout the text. The same distributions are shown for the $t\bar{t}A$ signal in Fig. 5 (middle), with the exception that the sequential prescription was used

for the ℓ^-. Clear differences are visible between the direct and sequential prescriptions, in particular for the background. Figure 5 (right) shows a comparison between $t\bar{t}H$, $t\bar{t}A$, and $t\bar{t}b\bar{b}$ at NLO + Shower, where the different possible natures of the signal ($t\bar{t}H$ or $t\bar{t}A$) do not seem to significantly affect the shape of the distribution. In the bottom plots, the corresponding distributions with two bins are shown, displaying the differences in forward-backward asymmetries.

B. $t\bar{t}H$ and $t\bar{t}A$ signals at NLO + Shower

Exploring kinematic differences between $t\bar{t}H$ and $t\bar{t}A$ is of utmost importance in order to find a set of good discriminating variables that may be sensitive to the nature of the top quark Yukawa coupling. In fact, differences between the scalar and pseudoscalar are visible through angles between particle directions (t, $\bar{t}$ and h), already at
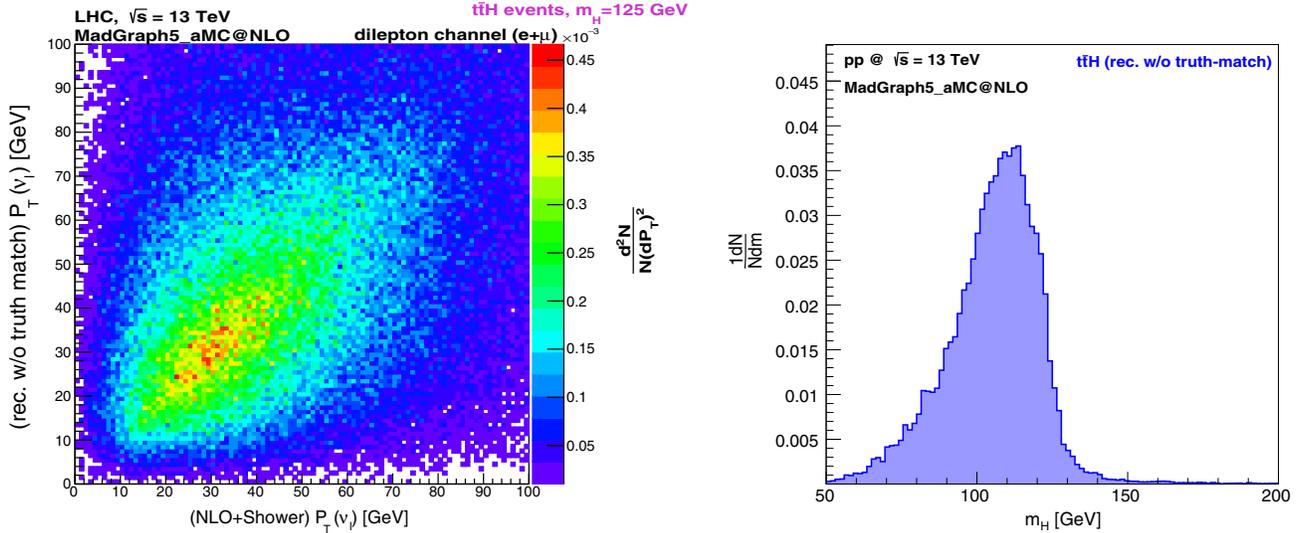
FIG. 4.   Two-dimensional distribution of the neutrino p_T in $t\bar{t}H$ events. Left: The NLO + Shower p_T (x axis) against the reconstructed p_T without truth-match (y axis) is shown. Right: Distribution of the reconstructed Higgs boson mass without truth-matched jets in $t\bar{t}H$ events.

production. Figure 6 (left) shows, at NLO + Shower, the angle between the top quark and Higgs boson directions (x axis) versus the angle between the antitop quark and Higgs boson directions (y axis), all evaluated in the $t\bar{t}H$ center-of-mass system. The same distribution is shown for the pseudoscalar signal $t\bar{t}A$ in Fig. 6 (right). In Fig. 7, the angle between the top quark direction in the $t\bar{t}h$ center-of-mass frame and the $t\bar{t}h$ direction in the lab frame (y axis),

is plotted against the angle between the Higgs direction, in the $\bar{t}h$ rest frame, and the direction of three decay products, all boosted to the h rest frame (x axis): (left) b quark from Higgs boson, (middle) ℓ^+ from the top quark, and (right) ℓ^- from $\bar{t}$. In the top (bottom) row, the $t\bar{t}H$ ($t\bar{t}A$) signal is shown, without any cuts. Differences between the scalar and pseudoscalar signals are clearly visible.
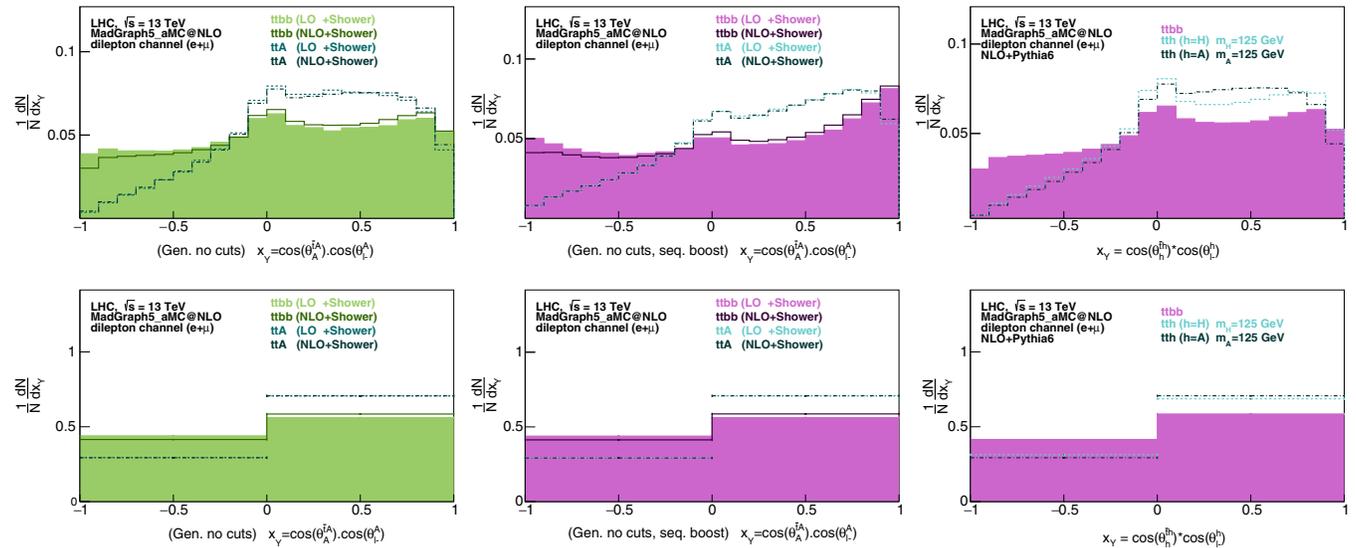
FIG. 5.   NLO + Shower versus LO + Shower behavior of the distribution of $x_Y = \cos(\theta_H^{\bar{t}H}) \cos(\theta_{\ell^-}^H)$ at the parton level with shower effects, without any selection cuts or reconstruction, for the SM signal $t\bar{t}H$ (left) and for the $t\bar{t}A$ signal (middle), each one against the main background $t\bar{t}b\bar{b}$. Notice that, for the middle plots, the sequential boost prescription was employed for the lepton. The differences between the LO + Shower and NLO + Shower angular distributions are shown on top. Asymmetries around $x_Y = 0$ are visible in the two binned distributions (bottom). The $t\bar{t}H$, $t\bar{t}A$, and $t\bar{t}b\bar{b}$ angular distributions at NLO are compared (top right), and the corresponding two binned distributions show the asymmetries (bottom right).
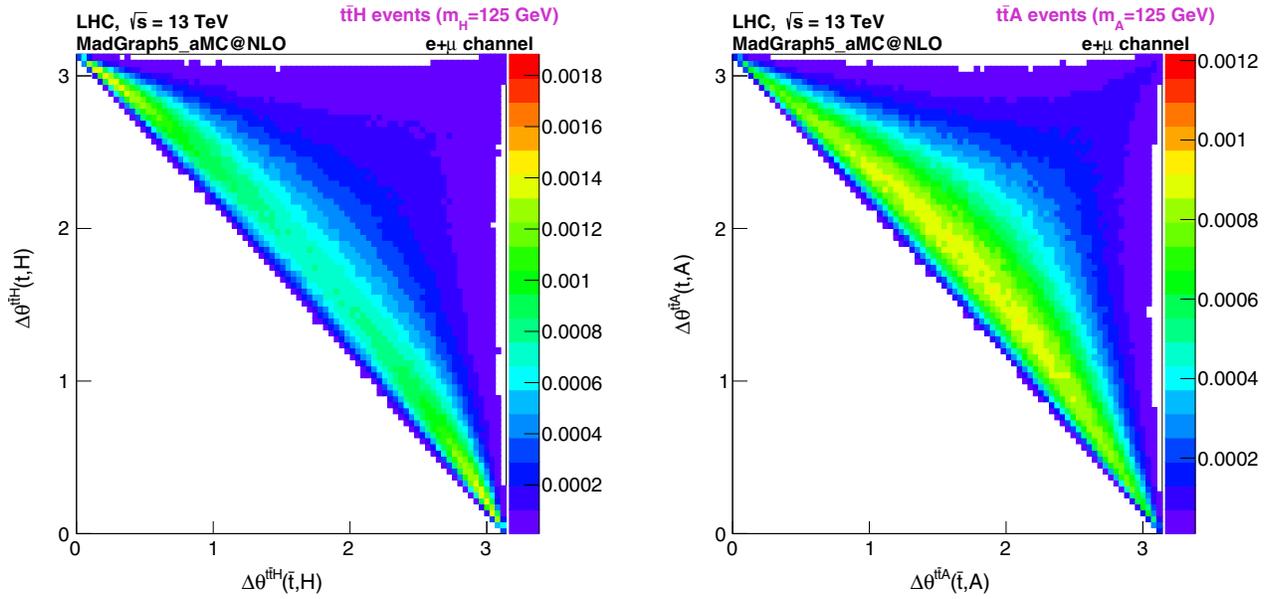
FIG. 6. Angle between the t quark and Higgs boson (x axis) at NLO + Shower effects plotted against the angle between the $\bar{t}$ quark and Higgs boson (y axis), in the $t\bar{t}H$ center-of-mass system. The SM Higgs boson (H) distribution (left) and the pure pseudoscalar Higgs boson (A) distribution (right) are shown.
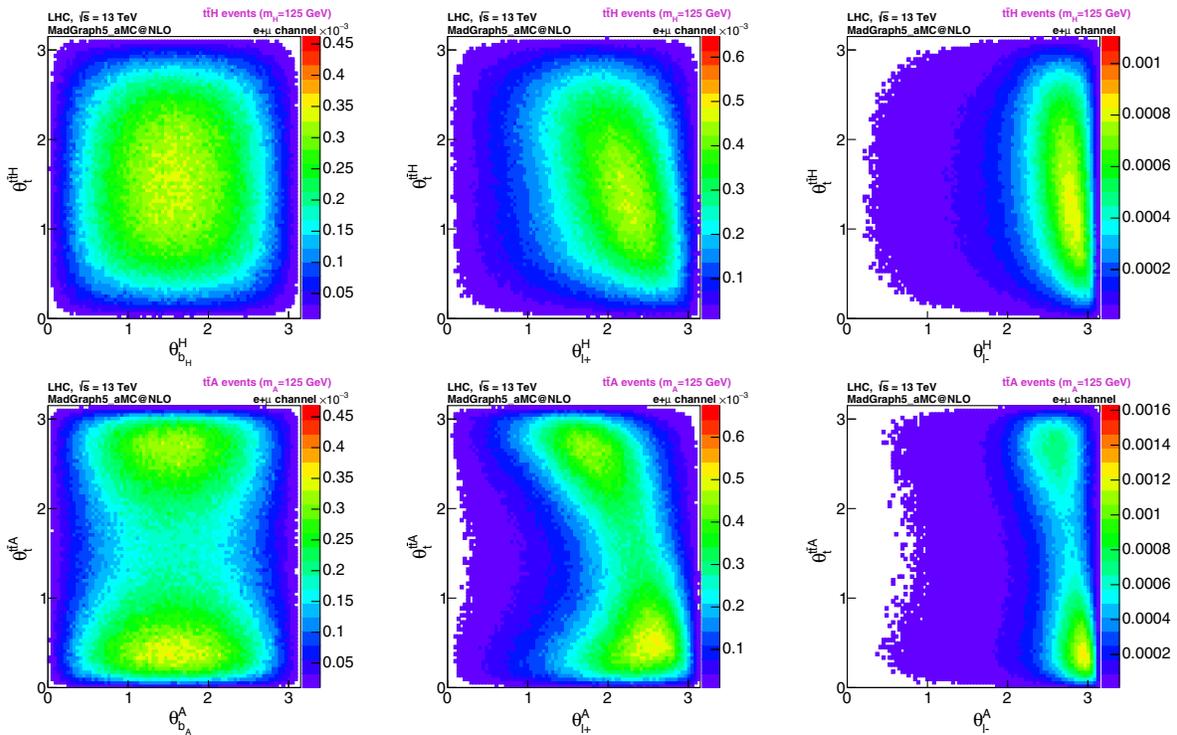
FIG. 7. Two-dimensional distribution at NLO + Shower of the angle between the top quark, in the $t\bar{t}h$ center-of-mass frame, and the $t\bar{t}h$ direction in the lab frame (y axis) plotted against the angle between the Higgs direction, in the $\bar{t}h$ rest frame, and the direction of several decay products (all boosted to the Higgs center of mass): (left) b quark from h, (middle) ℓ^+ from the top quark, and (right) ℓ^- from $\bar{t}$. The top (bottom) distributions correspond to $t\bar{t}H$ ($t\bar{t}A$), without any cuts.
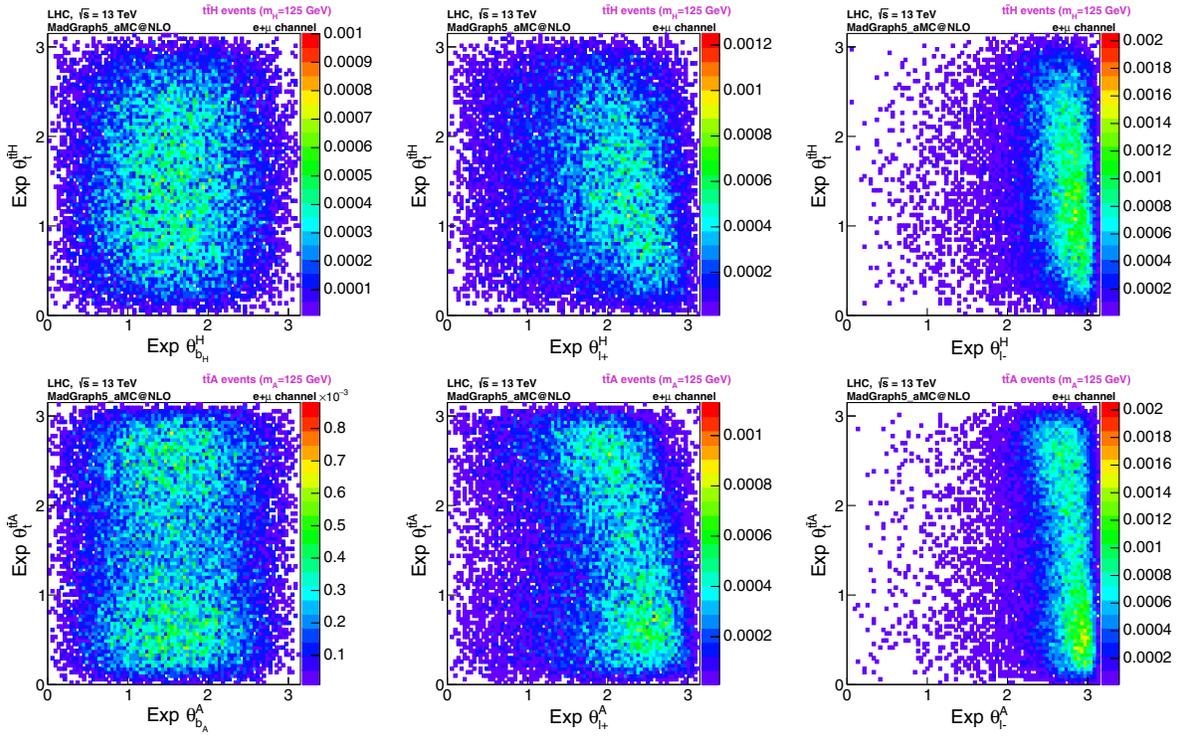
FIG. 8.   The same as Fig. 7, after all selection cuts and full kinematic reconstruction.

C. Angular distributions after reconstruction

Signal distributions are distorted due to cuts from the necessary selection criteria applied to events and the kinematic fit. The shape of the distributions, although affected by the significant reduction on the total number of events, is nevertheless largely preserved. In Fig. 8 the same angular distributions as those shown in Fig. 7 are represented, after selection cuts and full kinematic
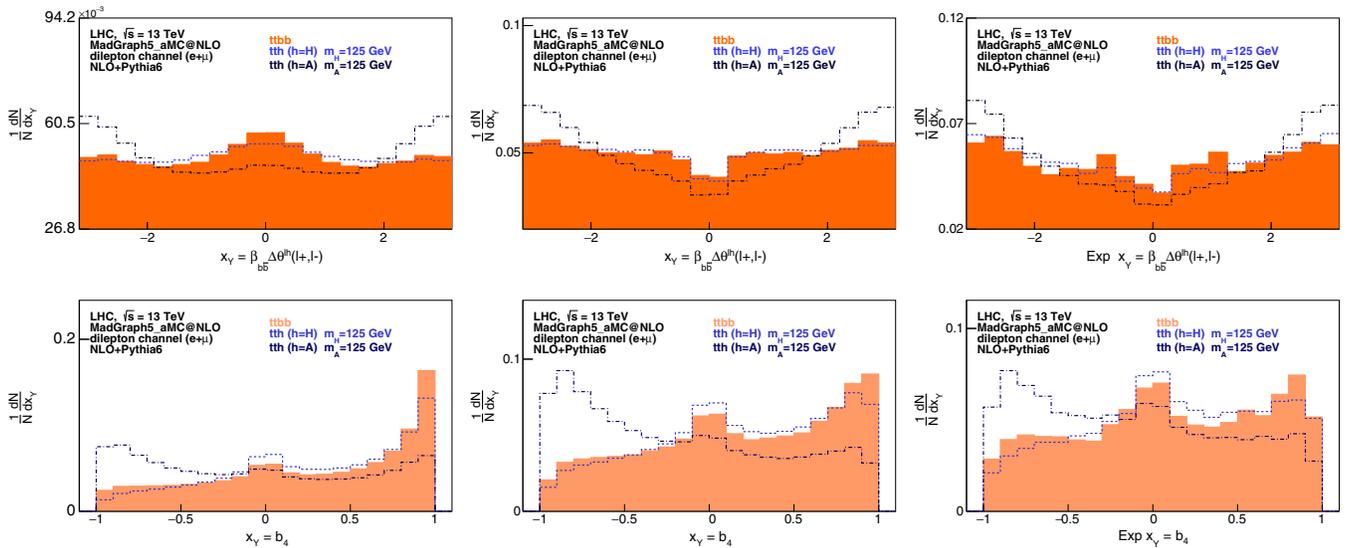
FIG. 9.   Normalized $\beta \Delta \theta^{\ell h}(\ell+, \ell-)$ distributions at NLO + Shower without cuts (top left), with cuts (top middle), and after cuts and full kinematic reconstruction (top right). The NLO + Shower b_4 distribution is also shown at the parton level without cuts (bottom left), with cuts (bottom middle), and after cuts and full kinematic reconstruction (bottom right). The dashed line represents the $t\bar{t}h$ SM model signal ($h = H$ and $CP = +1$) and the dashed-dotted line corresponds to the pure pseudoscalar distribution $t\bar{t}h$ ($h = A$ and $CP = -1$). The shadowed region corresponds to the NLO + Shower $t\bar{t}b\bar{b}$ dominant background.
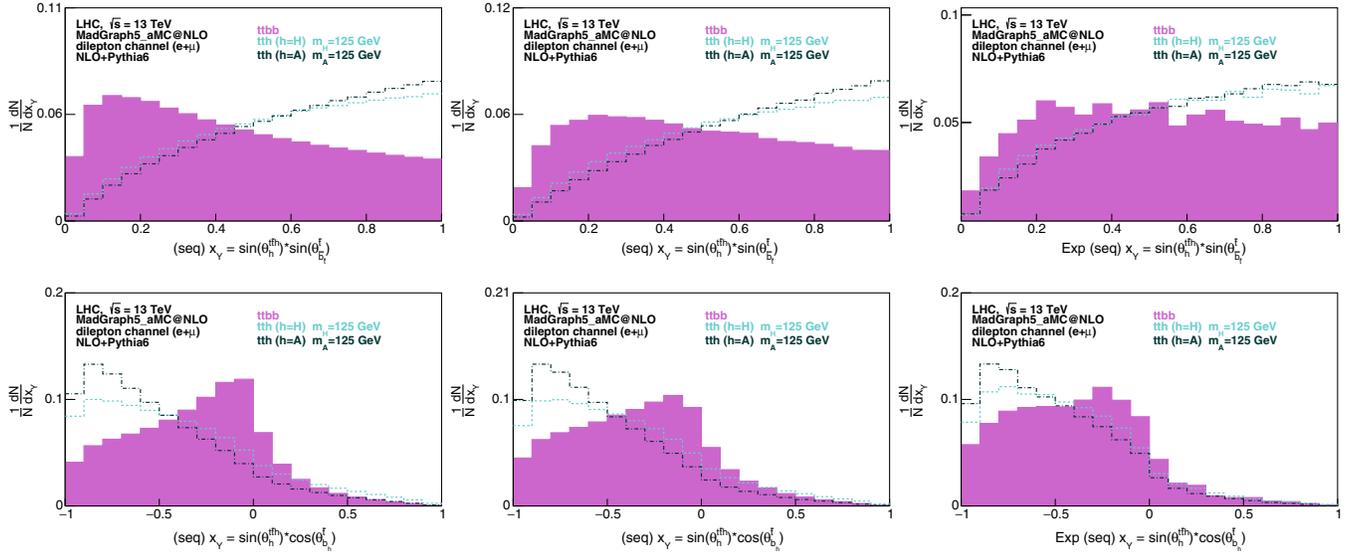
FIG. 10.   Distributions of $x_Y = \sin(\theta_h^{t\bar{t}h})\sin(\theta_{\bar{b}_{\bar{t}}}^{\bar{t}})$ (top) and $x_Y = \sin(\theta_h^{t\bar{t}h})\cos(\theta_{b_h}^{\bar{t}})$ (bottom). The distributions at NLO + Shower (left), after cuts (middle), and after cuts and full kinematic reconstruction (right) are shown. The dashed line represents the $t\bar{t}h$ SM model signal ($h = H$ and $CP = +1$) and the dashed-dotted line corresponds to the pure pseudoscalar distribution $t\bar{t}h$ ($h = A$ and $CP = -1$). The shadowed region corresponds to the NLO + Shower $t\bar{t}b\bar{b}$ dominant background. The laboratory four-momentum of b quarks is boosted sequentially to the Higgs center-of-mass system (see text for details).

reconstruction. The density of points shows a similar pattern to that in Fig. 7. Even after kinematic reconstruction, clear differences between the different signal natures are visible.

Forward-backward asymmetries associated to each of the observables under study were defined according to [19]

$$A_{FB}^{Y} = \frac{\sigma(x_Y > 0) - \sigma(x_Y < 0)}{\sigma(x_Y > 0) + \sigma(x_Y < 0)}, \tag{9}$$

where $\sigma(x_Y > 0)$ and $\sigma(x_Y < 0)$ correspond to the total cross section with x_Y above and below zero, respectively. The asymmetries are evaluated at NLO + Shower and after the kinematic fit, for different choices of the variable x_Y (found to provide a significant difference between the signals and dominant background):

$\cos(\theta_h^{t\bar{t}h})\cos(\theta_{\ell^-}^{h})$ for $A_{FB}^{\ell^-(h)}$,

$b_4 = (p_t^z \cdot p_{\bar{t}}^z)/(|\vec{p}_t| \cdot |\vec{p}_{\bar{t}}|)$, as defined in Ref. [17], for $A_{FB}^{b_4}$,

$\sin(\theta_h^{t\bar{t}h})\sin(\theta_{\bar{b}_{\bar{t}}}^{\bar{t}})$ for $A_{FB}^{\bar{b}_{\bar{t}}(\bar{t})}$ (seq. boost),

$\sin(\theta_h^{t\bar{t}h})\cos(\theta_{b_h}^{\bar{t}})$ for $A_{FB}^{b_h(\bar{t})}$ (seq. boost),

$\sin(\theta_t^{t\bar{t}h})\sin(\theta_{W^+}^{h})$ for $A_{FB}^{W^+(h)}$ (seq. boost),

$\sin(\theta_{\bar{t}}^{t\bar{t}h})\sin(\theta_{b_h}^{h})$ for $A_{FB}^{b_h(h)}$ (seq. boost), and

$\sin(\theta_h^{t\bar{t}h})\sin(\theta_{\bar{t}}^{t\bar{t}})$ for $A_{FB}^{\bar{t}(t\bar{t})}$.

The angular distributions from which each asymmetry was computed are represented in Figs. 5 and 9–11. In Table I we show the NLO + Shower values of the asymmetries without any selection applied and after full kinematic reconstruction.

IV. OBSERVABLES SENSITIVE TO THE *CP* NATURE OF THE TOP YUKAWA COUPLING

In the previous sections, we identified angular observables for which the distributions of $t\bar{t}b\bar{b}$ events and signal ($t\bar{t}H$ and $t\bar{t}A$) events show important differences. For many such observables, the distributions of the $t\bar{t}H$ and $t\bar{t}A$ samples are very similar (see the plot on the right of Fig. 5, as an example). These observables are ideal for implementing a search for (or set limits on) the total $t\bar{t}h$ production cross section, since they have the desirable feature of being insensitive to the *CP* nature of the Higgs-top coupling. However, within the set of new angular observables, many result in incompatible distributions between $t\bar{t}H$ and $t\bar{t}A$ samples at the reconstruction level without truth-match. This suggests that they are useful for experimentally measuring (or setting limits on) a pseudoscalar component of the top Yukawa coupling.

Observables in $t\bar{t}h$ events with this same purpose have been previously proposed, for example, in Refs. [17,22,23]. The observables proposed in those works, for the $t\bar{t}H$ and $t\bar{t}A$ signal samples as well as for the $t\bar{t}b\bar{b}$ background, were studied in reconstructed events. For brevity, we show results for two of the most compelling observables. The authors of Ref. [18] proposed the observable $\beta_{b\bar{b}}\Delta\theta^{\ell h}(\ell^+, \ell^-)$, where $\theta^{\ell h}(\ell^+, \ell^-)$ is the angle between the ℓ^+ and ℓ^- directions, projected onto the plane perpendicular to the h direction in the lab frame, and β is defined as the sign of $(\vec{p}_b - \vec{p}_{\bar{b}}) \cdot (\vec{p}_{\ell^-} \times \vec{p}_{\ell^+})$ (b and $\bar{b}$ result from the t and $\bar{t}$ decays, respectively). The other observable is b_4, already introduced in the previous section, and first proposed in Ref. [17]. An
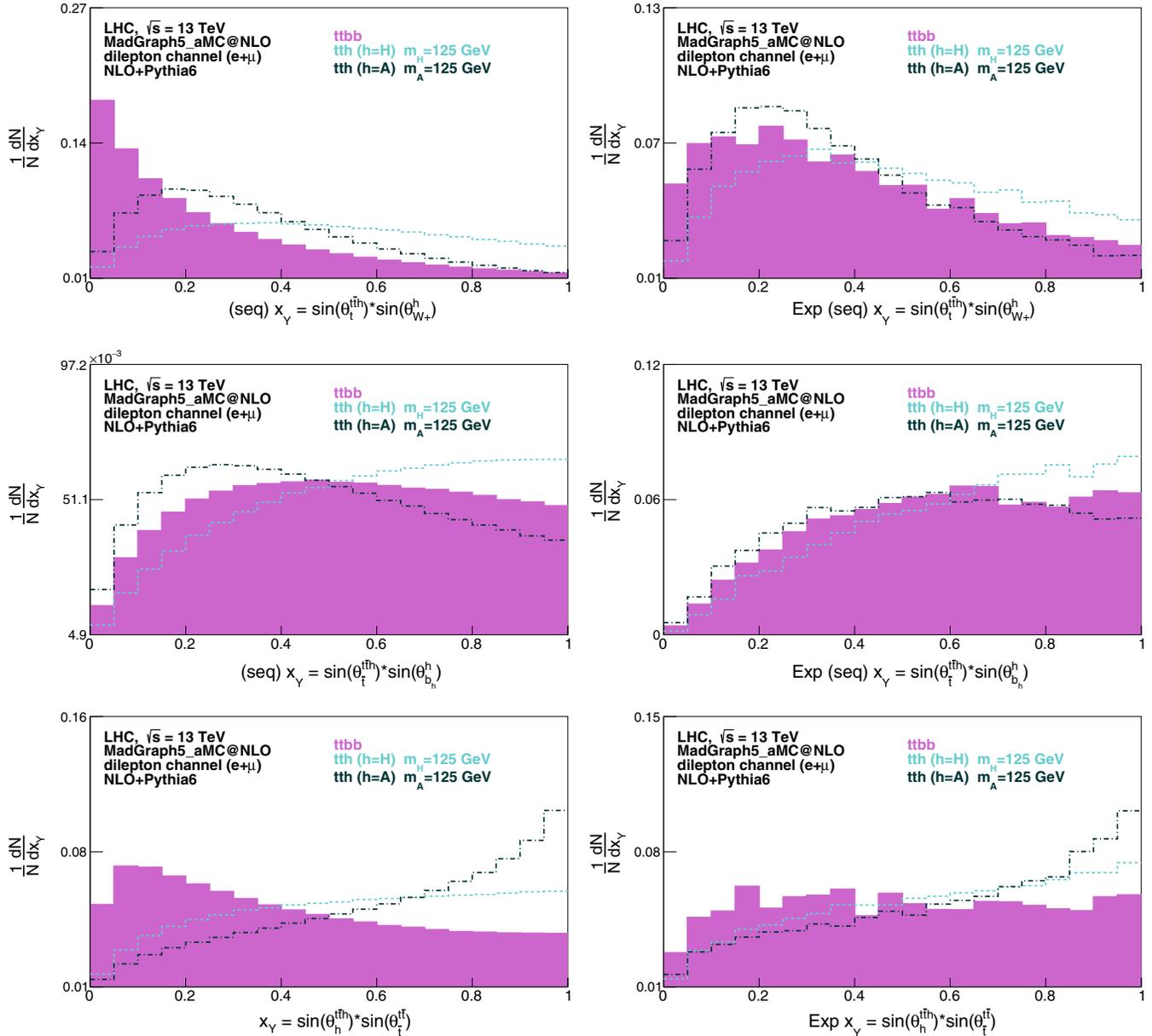
FIG. 11. NLO + Shower angular distributions at the parton level before selection cuts (left) and after all cuts and full kinematic reconstruction (right) of (top) $x_Y = \sin(\theta_t^{t\bar{t}H})\sin(\theta_{W+}^H)$, (middle) $x_Y = \sin(\theta_{\bar{t}}^{t\bar{t}H})\sin(\theta_{b_H}^H)$, and (bottom) $x_Y = \sin(\theta_H^{t\bar{t}h})\sin(\theta_{\bar{t}}^{t\bar{t}})$. The dashed line represents the $t\bar{t}h$ SM model signal ($h = H$ and $CP = +1$), the dashed-dotted line corresponds to the pure pseudoscalar distribution $t\bar{t}h$ ($h = A$ and $CP = -1$), and the shadowed region corresponds to the NLO + Shower $t\bar{t}b\bar{b}$ dominant background (see text for details).

important remark is that b_4, like many other observables in the referred publications, requires the reconstruction of the t and $\bar{t}$ four-momenta, which is only achievable through a kinematic fit such as the one used in this work. In Fig. 9, distributions are presented for $\beta_{b\bar{b}}\Delta\theta^{\ell h}(\ell^+, \ell^-)$ (top) and b_4 (bottom), for $t\bar{t}H$, $t\bar{t}A$, and $t\bar{t}b\bar{b}$ samples. They are shown at NLO + Shower without cuts (left), with cuts (middle), and at reconstruction level without truth-match, after additionally requiring at least three b-tagged jets and $|m_{\ell\ell} - m_Z| > 10$ GeV (right). While it is evident that detector simulation

and reconstruction degrade the discriminating power of these observables, the most dramatic effect on the distribution shapes comes from applying the acceptance cuts. After these cuts, the distributions at NLO + Shower already exhibit roughly the same behavior as the distributions after reconstruction. Optimization of the selection criteria is thus quite important, but stays largely outside the scope of this paper.

In Fig. 10, distributions of $\sin(\theta_h^{t\bar{t}h})\sin(\theta_{\bar{b}_{\bar{t}}}^{\bar{t}})$ (top) and $\sin(\theta_h^{t\bar{t}h})\cos(\theta_{b_h}^{\bar{t}})$ (bottom) are shown. These are among the

TABLE I. Asymmetry values for $t\bar{t}H$, $t\bar{t}A$, and $t\bar{t}b\bar{b}$ at NLO + Shower (without any cuts) and after applying the selection criteria and kinematic reconstruction.

Asymmetries	NLO + Shower (no cuts applied)		After selection and reconstruction	
	$t\bar{t}H/t\bar{t}A$	$t\bar{t}b\bar{b}$	$t\bar{t}H/t\bar{t}A$	$t\bar{t}b\bar{b}$
$A_{FB}^{\ell-(h)}$	$+0.37/+0.41$	$+0.17$	$+0.42/+0.39$	$+0.24$
$A_{FB}^{b_4}$	$+0.35/-0.10$	$+0.33$	$+0.16/-0.17$	$+0.12$
$A_{FB}^{b_l(\bar{t})}$ (seq. boost)	$+0.28/+0.33$	-0.17	$+0.25/+0.28$	$+0.03$
$A_{FB}^{b_h(\bar{t})}$ (seq. boost)	$-0.65/-0.77$	-0.62	$-0.78/-0.83$	-0.76
$A_{FB}^{W+(h)}$ (seq. boost)	$-0.03/-0.46$	-0.60	$+0.17/-0.06$	-0.04
$A_{FB}^{b_h(h)}$ (seq. boost)	$+0.25/-0.08$	$+0.07$	$+0.37/+0.16$	$+0.23$
$A_{FB}^{\bar{t}(t\bar{t})}$	$+0.16/+0.37$	-0.21	$+0.23/+0.31$	$+0.01$

investigated angular observables for which the $t\bar{t}b\bar{b}$ background sample was least compatible with both $t\bar{t}H$ and $t\bar{t}A$ samples. The distributions are represented at NLO + Shower without cuts (left), after selection cuts (middle), and after full kinematic reconstruction and the additional requirements of $|m_{\ell\ell} - m_Z| > 10$ GeV and at least three b-tagged jets (right). The dashed line represents the $t\bar{t}H$ distribution and the dashed-dotted line corresponds to $t\bar{t}A$. The shadowed region corresponds to the $t\bar{t}b\bar{b}$ dominant background.

Figure 11 shows distributions of three angular observables among the ones for which the $t\bar{t}H$ and $t\bar{t}A$ samples were least compatible at the reconstruction level without truth-match. They are represented at NLO + Shower without cuts (left) and at the reconstruction level without truth-match, after the previously mentioned cuts on the b-tag multiplicity and $m_{\ell\ell}$ (right). Distributions of $t\bar{t}b\bar{b}$ events are also included for completeness. The discriminating performance of these observables is comparable to that of those proposed in the literature. Computing the angular observables also requires full reconstruction of t and $\bar{t}$. Again, applying the acceptance cuts, detector simulation and kinematic reconstruction visibly degrades the discrimination between $t\bar{t}H$ and $t\bar{t}A$ samples.

V. ANALYSIS AND RESULTS

In order to estimate the experimental sensitivity of an analysis employing the observables under study, further selection criteria was applied, as mentioned previously. Depletion of the $Z + $ jets background is accomplished by selecting events with a dilepton invariant mass $m_{\ell\ell}$ such that $|m_{\ell^+\ell^-} - m_Z| > 10$ GeV. This selection was applied in all dilepton flavor categories (ee, $\mu\mu$, and $e\mu$). Most backgrounds, notably $t\bar{t} + $ jets, are then mitigated by selecting events with at least three b-tagged jets.

Table II shows the expected effective cross sections in fb, at several levels of the event selection, for dileptonic signal and SM backgrounds. The $t\bar{t}A$ pseudoscalar signal was scaled to the $t\bar{t}H$ scalar cross section for comparison purposes.

In Fig. 12, the expected number of events from the different SM processes are shown, including the Higgs signal, for a luminosity of 100 fb^{-1} at the LHC, for events with at least three b jets (left) and at least four b jets (right). As expected, the composition of backgrounds changes quite significantly after event selection.

The fake data points correspond to one particular pseudoexperiment randomly created from the expected Standard Model $t\bar{t}H$ signal and background distributions. Its purpose is only to guide the reader through the total number of expected events and related statistical uncertainties, after event selection and full reconstruction.

Several kinematic properties of the events, including the new angular distributions introduced in this paper, were tested with several multivariate methods. A BDTG has the best performance among the methods investigated. Its output was used to test the analysis sensitivity to probe the scalar versus pseudoscalar component of the top-Higgs couplings, as a function of $\cos\alpha$. From the long set of variables tried, the 15 best ranked by the multivariate method, after reconstruction, were the b_4 and Higgs mass ($m_{b\bar{b}}$); the angular distributions with

TABLE II. Expected cross sections (in fb) as a function of selection cuts, at 13 TeV, for dileptonic signal and background events at the LHC.

	$N_{\text{jets}} \geq 4$	Kinematic	m_Z	N_b	N_b
	$N_{\text{lep}} = 2$	Fit	cut	≥ 3	≥ 4
$t\bar{t} + c\bar{c}, t\bar{t} + 1f$	2160	1300	1110	4.78	0.06
$t\bar{t} + b\bar{b}$	87.1	51.9	44.5	2.91	0.27
$t\bar{t} + V(V = Z, W)$	7.9	4.5	3.9	0.09	0.01
Single t	54	26	23	0.12	0.00
$V + $ jets $(V = W, Z)$	2700	1200	200	0.00	0.00
$V + b\bar{b}(V = W, Z)$	570	280	20	0.00	0.00
Diboson	130	53	14	0.00	0.00
Total back.	5700	2900	1410	7.90	0.34
$t\bar{t}H$	4.04	2.49	2.15	0.26	0.033
$t\bar{t}A$	4.43	2.69	2.36	0.31	0.041
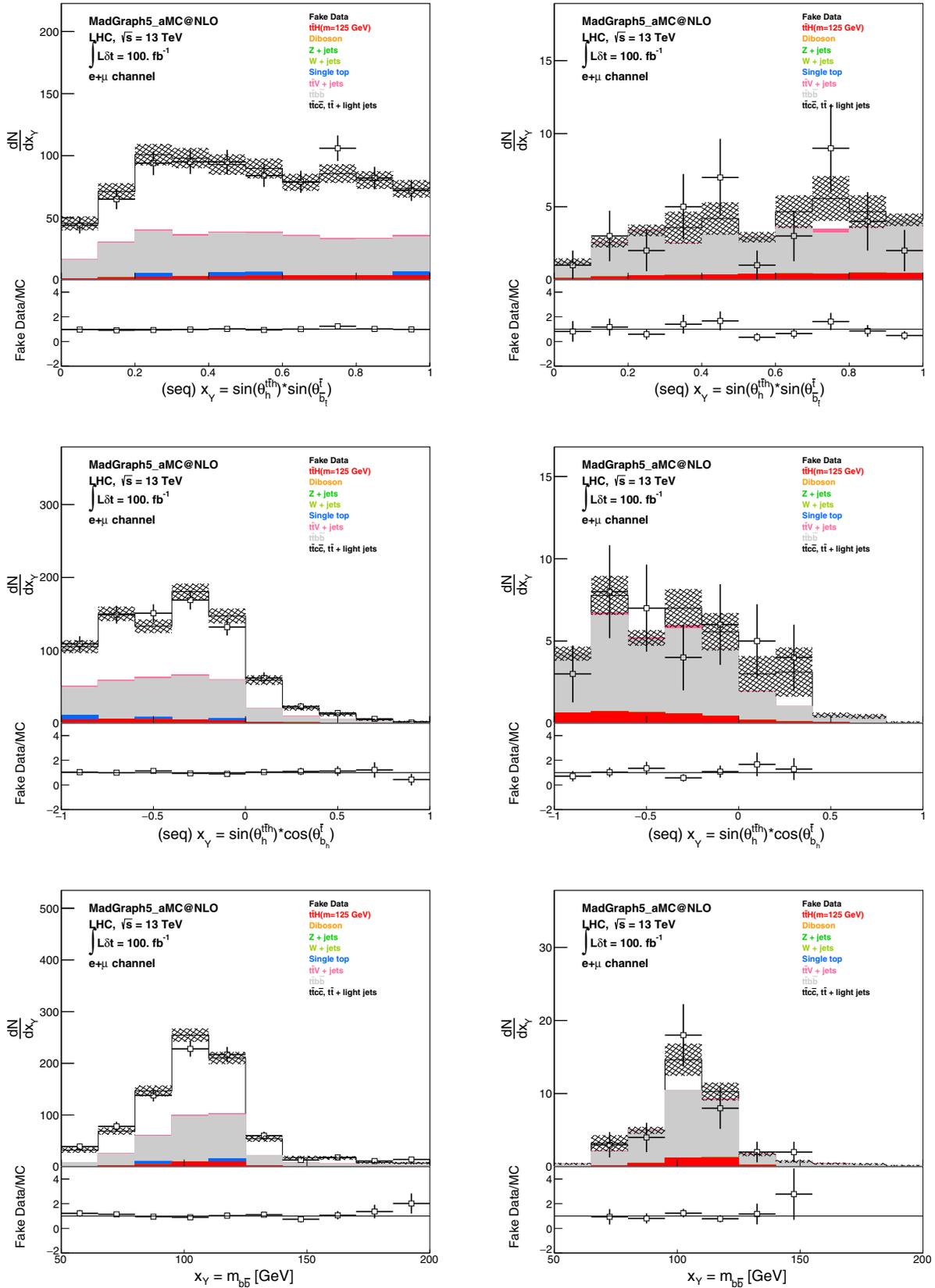
FIG. 12.   Distributions of $x_Y = \sin(\theta_H^{t\bar{t}H})\sin(\theta_{\bar{b}_{\bar{t}}}^{\bar{l}})$ (top), $x_Y = \sin(\theta_H^{t\bar{t}H})\cos(\theta_{b_H}^{\bar{l}})$ (middle), and $x_Y = m_{b\bar{b}}$ (bottom) after final selection at 13 TeV for 100 fb^{-1}. The distributions on the left (right) correspond to events with at least three (four) jets from the hadronization of b quarks.
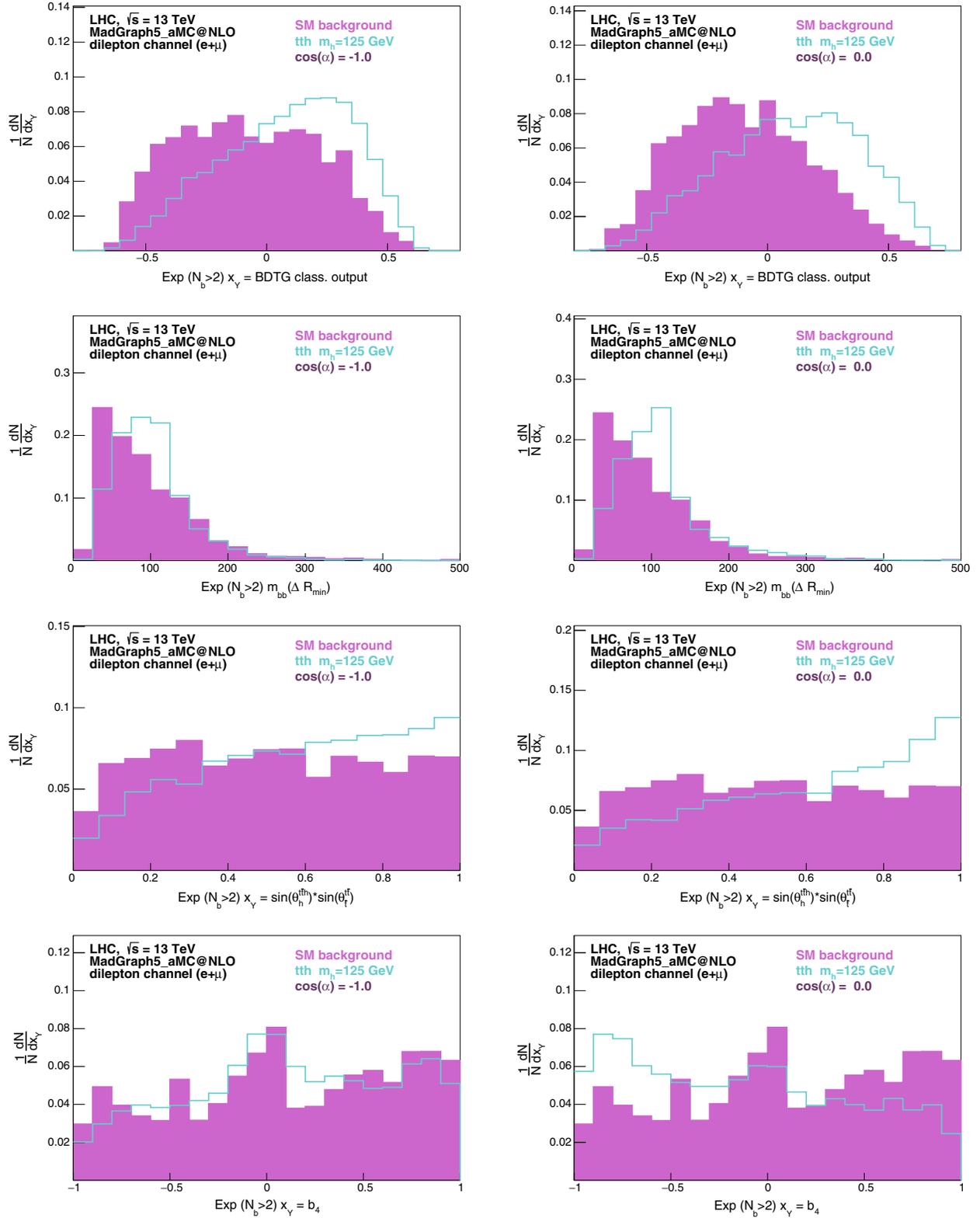
FIG. 13. Normalized distributions of the BDTG output discriminant variable (first row), the invariant mass of the two *b*-tagged jets with minimum ΔR ($m_{bb}^{\min \Delta R}$) (second row), the $\sin(\theta_h^{t\bar{t}h})\sin(\theta_t^{t\bar{t}})$ (third row) and the b_4 variable (fourth row), after final selection at 13 TeV. The distributions on the left (right) correspond to pure scalar (pseudoscalar) Higgs bosons.

direct boost, i.e., $\cos(\theta_h^{\bar{t}h})\cos(\theta_{\ell^-}^h)$, $\sin(\theta_h^{t\bar{t}h})\sin(\theta_{\bar{t}}^{t\bar{t}})$ and the variables with sequential boost $\sin(\theta_{\bar{t}}^{t\bar{t}h})\sin(\theta_{b_h}^h)$(seq), $\sin(\theta_h^{t\bar{t}h})\cos(\theta_{b_h}^{\bar{t}})$(seq), $\sin(\theta_h^{t\bar{t}h})\sin(\theta_{\bar{b}_{\bar{t}}}^{\bar{t}})$(seq), $\sin(\theta_t^{t\bar{t}h})\times\sin(\theta_{W^+}^h)$(seq); the $\Delta\eta$ between the jets with maximum

$\Delta\eta$ ($\Delta\eta_{jj}^{\max\,\Delta\eta}$) and the invariant mass of the two b-tagged jets with lowest ΔR ($m_{bb}^{\min\,\Delta R}$); the ΔR between the Higgs candidate and the closest ($\Delta R_{hl}^{\min\,\Delta R}$) and farthest ($\Delta R_{hl}^{\max\,\Delta R}$) leptons; the ΔR between the b-tagged jets
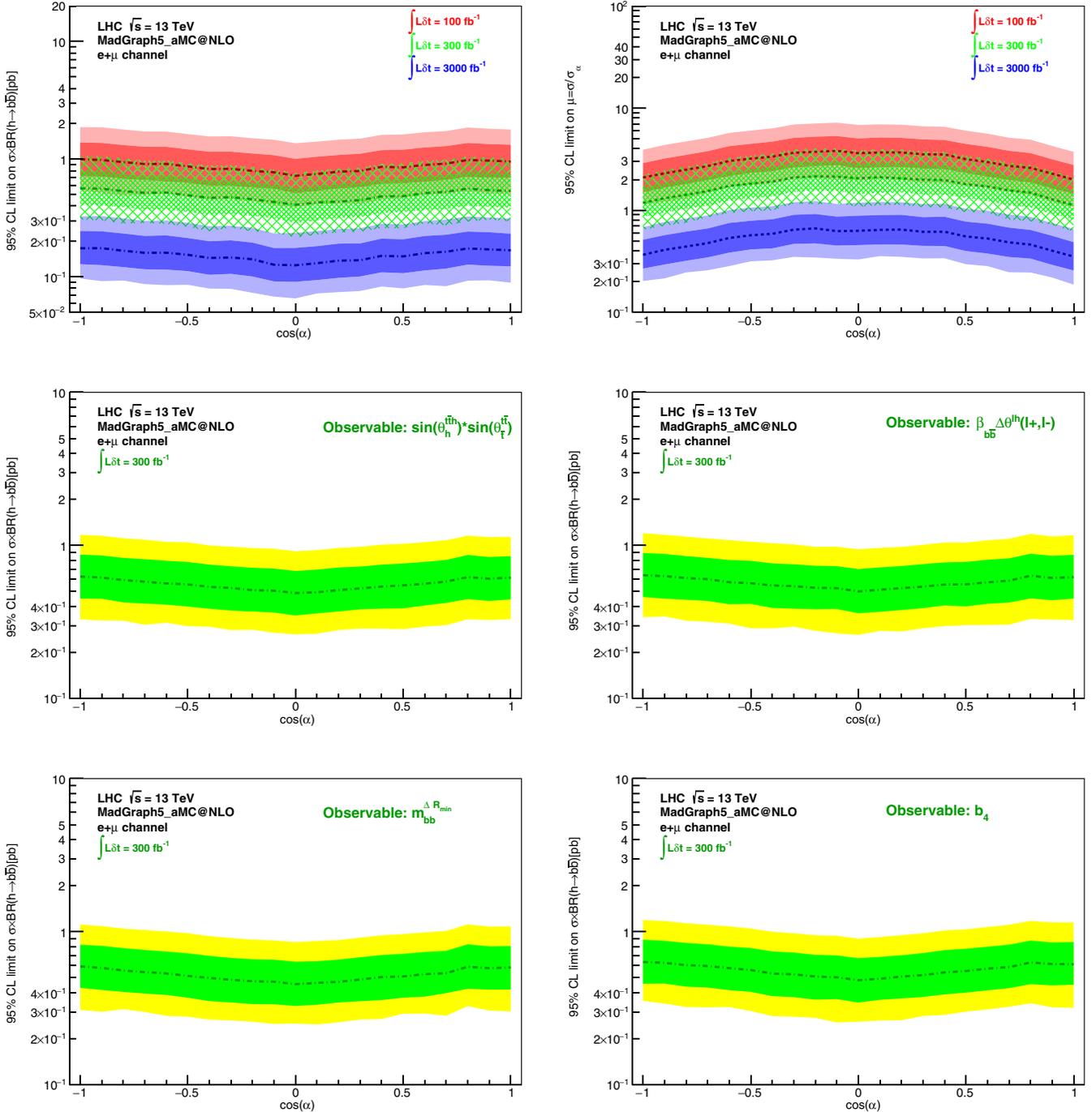
FIG. 14.   Expected limits at 95% C.L. in the background-only scenario, as a function of $\cos(\alpha)$. Limits on $\sigma\times\mathrm{BR}(h\to b\bar{b})$ (top left) and μ (top right) obtained with the BDTG output discriminant for integrated luminosities of 100, 300, and 3000 fb^{-1} are shown. The lines correspond to the median, while the narrower (wider) bands correspond to the 1σ (2σ) intervals. Limits on $\sigma\times\mathrm{BR}(h\to b\bar{b})$ at 300 fb^{-1} are also shown, using the individual observables $\sin(\theta_h^{t\bar{t}h})\sin(\theta_{\bar{t}}^{t\bar{t}})$ (center left) and $\beta_{b\bar{b}}\Delta\theta^{\ell h}(\ell^+,\ell^-)$ (center right), $m_{bb}^{\min\,\Delta R}$ (bottom left), and b_4 (bottom right).

with highest p_T ($\Delta R_{bb}^{\max p_T}$) and the invariant mass of the two jets with the closest value to the Higgs mass ($m_{jj}^{\text{closest to 125 GeV}}$); and the jets aplanarity.

In Fig. 13, normalized distributions of the BDTG output classifier (first row) and three of the input variables (remaining rows) used in the multivariate method are shown for the pure scalar (left plots) and pseudoscalar

(right plots) Higgs bosons. It should be noted that the BDTG used for the limit extraction at a given $\cos(\alpha)$ has been trained on a signal sample generated with the same value for $\cos(\alpha)$. This justifies the different SM background shapes between the left and right plots of the first row in Fig. 13. The invariant mass of the two b-tagged jets with minimum ΔR ($m_{bb}^{\min \Delta R}$) (second row), the $\sin(\theta_h^{\bar{t}lh}) \sin(\theta_{\bar{t}}^{t\bar{t}})$
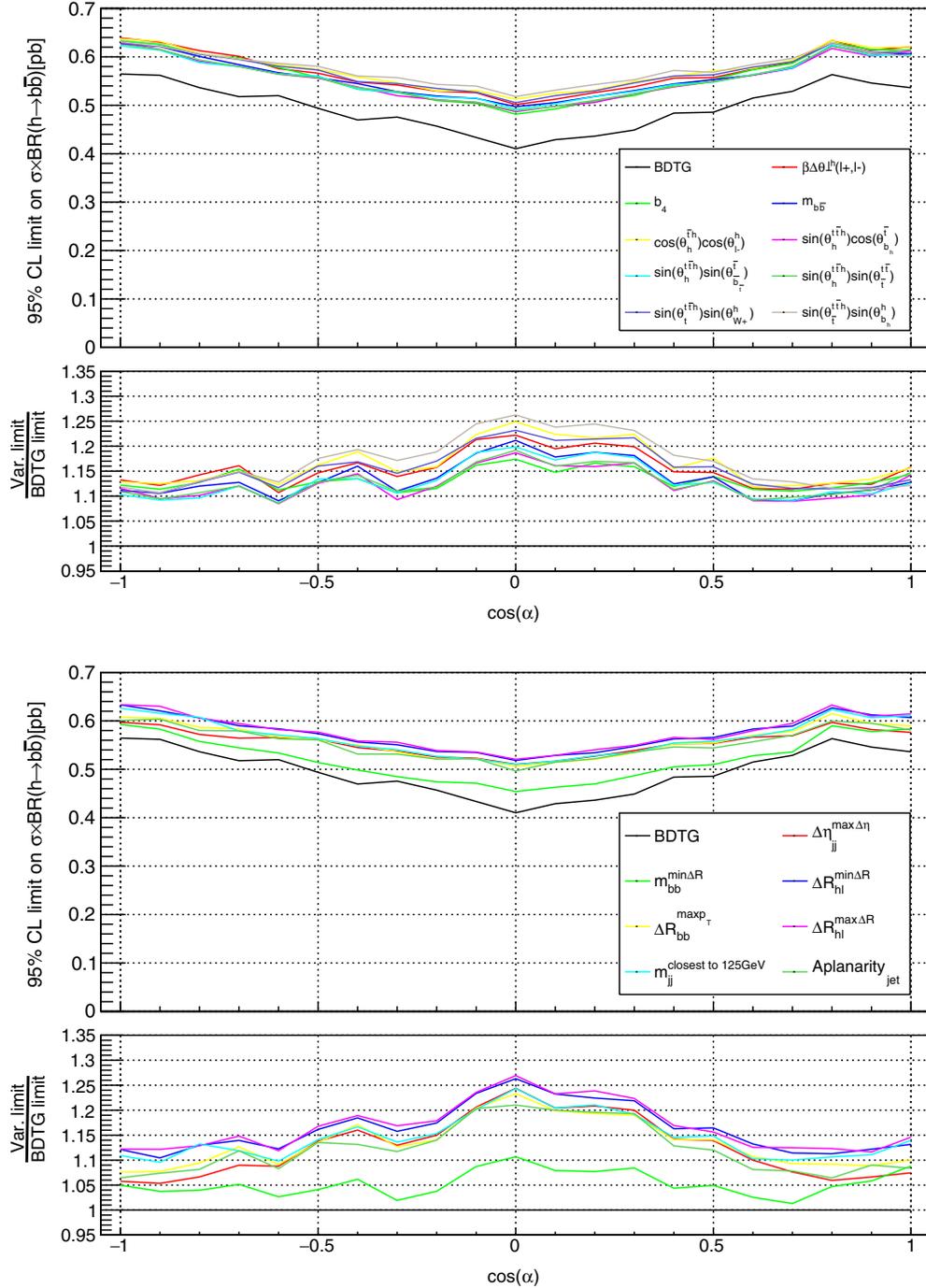
FIG. 15.   Comparison between limits on $\sigma \times \text{BR}(h \to b\bar{b})$ at 300 fb^{-1} obtained from each of the individual distributions used in the BDTG [$\beta_{b\bar{b}}\Delta\theta^{\ell h}(\ell^+, \ell^-)$, b_4, $m_{b\bar{b}}$] and angular distributions (top), and remaining distributions used as input for the BDTG (bottom). The ratios with respect to the limit obtained from the BDTG distribution are also represented.

(third row) and the b_4 variable (fourth row) are also shown for completeness. Shape differences between signal and background are clearly visible, and they are different for the scalar and pseudoscalar cases. In these figures, the line corresponds to the signal distribution and the shaded region corresponds to the full SM background at the LHC.

Expected limits at 95% C.L. for $\sigma \times \mathrm{BR}(h \to b\bar{b})$ and for signal strength μ, in the background-only scenario, were extracted, using the BDTG output distribution. Several signal samples were used, with values of $\cos(\alpha)$ ranging from -1 to 1 (in steps of 0.1). The first row of Fig. 14 shows these limits, for integrated luminosities of 100, 300, and 3000 fb^{-1}. Although data taking for large values of luminosity is expected to occur with $\sqrt{s} = 14$ TeV, we show the results at 3000 fb^{-1} for comparison. Sensitivity to SM $t\bar{t}H$ production at $\mu = 1$ should be attained shortly after the 300 fb^{-1} milestone, using this channel alone. Combining the dileptonic channel with other decay channels should allow to decrease significantly the luminosity necessary to probe the structure of the top quark Yukawa couplings to the Higgs boson. The second and third rows of Fig. 14 show limits on $\sigma \times \mathrm{BR}(h \to b\bar{b})$ at 300 fb^{-1}, obtained from fits to the following individual distributions: $\sin(\theta_h^{t\bar{t}h}) \sin(\theta_{\bar{t}}^{t\bar{t}})$ (center left), $\beta_{b\bar{b}}\Delta\theta^{\ell h}(\ell^+,\ell^-)$ (center right), $m_{b\bar{b}}^{\min\,\Delta R}$ (bottom left), and b_4 (bottom right). The results show that the different distributions used as input to the BDTG, although with the same general dependence on $\cos(\alpha)$, can have different sensitivities. A common feature of all of the variables is a better 95% C.L. limit on $\sigma \times \mathrm{BR}(h \to b\bar{b})$ as we approach the pure pseudoscalar region. In Fig. 15, a comparison is shown between limits on $\sigma \times \mathrm{BR}(h \to b\bar{b})$, at 300 fb^{-1}, obtained from each of the individual distributions used in the BDTG multivariate discriminant. Additionally, the limits corresponding to the BDTG itself are shown, as well as those from the $\beta_{b\bar{b}}\Delta\theta^{\ell h}(\ell^+,\ell^-)$ distribution, which is not included in the BDTG, and is only shown for completeness. Figure 15 (top) includes the limits from the angular observables, $\beta_{b\bar{b}}\Delta\theta^{\ell h}(\ell^+,\ell^-)$, b_4, and $m_{b\bar{b}}$. Figure 15 (bottom) shows the limits from all of the other individual observables used as input for the BDTG method. The ratios with respect to the limit obtained from the BDTG distribution are also represented. While most individual angular variables result in limits 15–20% worse than the BDTG method for the pseudoscalar case, the other variables tend to be in the 20–25% region, with the exception of $m_{bb}^{\min\,\Delta R}$, which clearly shows a better discriminating power (as expected from the plots in Fig. 13).

VI. CONCLUSIONS

In this paper, studies of $t\bar{t}h$ production for scalar and pseudoscalar Higgs bosons at a center-of-mass energy of 13 TeV at the LHC were considered for different luminosities. Dileptonic final states from $t\bar{t}h$ decays ($t \to bW^+ \to b\ell^+\nu_\ell$, $\bar{t} \to \bar{b}W^- \to \bar{b}\ell^-\bar{\nu}_\ell$, and $h \to b\bar{b}$)

were fully reconstructed by means of a kinematic fit that reconstructs the four-momenta of the undetected neutrinos. New angular distributions and asymmetries were proposed to allow better discrimination between signals of different nature (scalar or pseudoscalar) and backgrounds at the LHC. Using fully reconstructed $t\bar{t}h$ events, it is possible to obtain relevant information about the *CP* states of signal and background processes, through the measurements of new angular distributions and asymmetries. Even after event selection and full kinematical reconstruction, the spin information is largely preserved, opening a window for spin measurements and a better understanding of the nature of the top-Higgs Yukawa coupling and $t\bar{t}h$ production at the LHC. Expected limits at 95% C.L. were extracted on the $\sigma \times \mathrm{BR}(h \to b\bar{b})$ and signal strength μ using a boosted decision tree. A comparison between the sensitivities of the individual variables as a function of $\cos(\alpha)$ was also performed, showing that a multivariate method combining all the variables can improve the individual limits up to 25%. It is worth mentioning that the asymmetry measurements should be possible soon after the eventual discovery of the process with the expected accumulation of luminosity. Some of the angular distributions investigated in this work were used in addition to the kinematical distributions commonly discussed in the literature, yielding at least the same sensitivity to the nature of the top quark Yukawa coupling to the Higgs boson, if not better. The fact that the expected limits do not exhibit a too strong dependence on the particular choice of the *CP* phase (α) makes the analysis of the SM Higgs case (*CP*-even) a good starting point for any other case, where mixtures with *CP*-odd contributions are probed. Also, it was found that the invariant mass distribution of the two *b*-tagged jets with the lowest ΔR between them shows a particularly interesting behavior. All results presented so far were obtained using the dileptonic final states of $t\bar{t}h$ events alone. These are expected to be improved when other decay channels are combined, using fully reconstructed final states.

[1] P. W. Higgs, Phys. Lett. **12**, 132 (1964); Phys. Rev. Lett. **13**, 508 (1964); Phys. Rev. **145**, 1156 (1966); F. Englert and R. Brout, Phys. Rev. Lett. **13**, 321 (1964); G. S. Guralnik, C. R. Hagen, and T. W. Kibble, Phys. Rev. Lett. **13**, 585 (1964).

[2] G. Aad *et al.* (ATLAS Collaboration), Phys. Lett. B **716**, 1 (2012).

[3] S. Chatrchyan *et al.* (CMS Collaboration), Phys. Lett. B **716**, 30 (2012).

[4] G. Aad *et al.* (ATLAS Collaboration), Phys. Lett. B **726**, 88 (2013); **734**, 406(E) (2014); Phys. Lett. B **726**, 120 (2013); J. High Energy Phys. 04 (2015) 117; V. Khachatryan *et al.* (CMS Collaboration), Eur. Phys. J. C **75**, 212 (2015); S. Chatrchyan *et al.* (CMS Collaboration), Nat. Phys. **10**, 557 (2014).

[5] D. Fontes, J. C. Romão, R. Santos, and J. P. Silva, J. High Energy Phys. 06 (2015) 060.

[6] V. Khachatryan *et al.* (CMS Collaboration), Phys. Rev. D **92**, 012004 (2015).

[7] V. Khachatryan *et al.* (CMS Collaboration), Phys. Lett. B **759**, 672 (2016).

[8] G. Aad *et al.* (ATLAS Collaboration), Eur. Phys. J. C **75**, 476 (2015); **76**, 152(E) (2016).

[9] J. N. Ng and P. Zakarauskas, Phys. Rev. D **29**, 876 (1984); Z. Kunszt, Nucl. Phys. **B247**, 339 (1984); W. J. Marciano and F. E. Paige, Phys. Rev. Lett. **66**, 2433 (1991); J. F. Gunion, Phys. Lett. B **261**, 510 (1991); J. Goldstein, C. S. Hill, J. Incandela, S. Parke, D. Rainwater, and D. Stuart, Phys. Rev. Lett. **86**, 1694 (2001); W. Beenakker, S. Dittmaier, M. Krämer, B. Plümper, M. Spira, and P. M. Zerwas, Phys. Rev. Lett. **87**, 201805 (2001); Nucl. Phys. **B653**, 151 (2003); L. Reina and S. Dawson, Phys. Rev. Lett. **87**, 201804 (2001); S. Dawson, L. H. Orr, L. Reina, and D. Wackeroth, Phys. Rev. D **67**, 071503 (2003); **68**, 034022 (2003); S. Dittmaier, M. Kramer, and M. Spira, Phys. Rev. D **70**, 074010 (2004); R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, R. Pittau, and P. Torrielli, Phys. Lett. B **701**, 427 (2011); M. V. Garzelli, A. Kardos, C. G. Papadopoulos, and Z. Trócsányi, Europhys. Lett. **96**, 11001 (2011); H. B. Hartanto, B. Jäger, L. Reina, and D. Wackeroth, Phys. Rev. D **91**, 094003 (2015); S. Frixione, V. Hirschi, D. Pagani, H. S. Shao, and M. Zaro, J. High Energy Phys. 09 (2014) 065; Y. Zhang, W. G. Ma, R. Y. Zhang, C. Chen, and L. Guo, Phys. Lett. B **738**, 1 (2014).

[10] G. Aad *et al.* (ATLAS Collaboration), J. High Energy Phys. 05 (2016) 160.

[11] G. Aad *et al.* (ATLAS Collaboration) Phys. Lett. B **740**, 222 (2015).

[12] G. Aad *et al.* (ATLAS Collaboration), Phys. Lett. B **749**, 519 (2015).

[13] G. Aad *et al.* (ATLAS Collaboration), Eur. Phys. J. C **75**, 349 (2015).

[14] V. Khachatryan *et al.* (CMS Collaboration), J. High Energy Phys. 09 (2014) 087; 10 (2014) 106(E).

[15] V. Khachatryan *et al.* (CMS Collaboration), Eur. Phys. J. C **75**, 251 (2015).

[16] G. Aad *et al.* (ATLAS and CMS Collaborations), J. High Energy Phys. 08 (2016) 045.

[17] J. F. Gunion and X. G. He, Phys. Rev. Lett. **76**, 4468 (1996).

[18] F. Boudjema, R. M. Godbole, D. Guadagnoli, and K. A. Mohan, Phys. Rev. D **92**, 015019 (2015).

[19] S. P. A. dos Santos *et al.*, Phys. Rev. D **92**, 034021 (2015).

[20] S. Berge, W. Bernreuther, and S. Kirchner, Eur. Phys. J. C **74**, 3164 (2014); S. Berge, W. Bernreuther, and H. Spiesberger, arXiv:1208.1507; S. Berge, W. Bernreuther, B. Niepelt, and H. Spiesberger, Phys. Rev. D **84**, 116003 (2011); S. Berge, W. Bernreuther, and J. Ziethe, Phys. Rev. Lett. **100**, 171605 (2008); S. Khatibi and M. M. Najafabadi, Phys. Rev. D **90**, 074014 (2014).

[21] G. Brooijmans *et al.*, arXiv:1405.1617.

[22] J. Ellis, D. S. Hwang, K. Sakurai, and M. Takeuchi, J. High Energy Phys. 04 (2014) 004; S. Biswas, R. Frederix, E. Gabrielli, and B. Mele, J. High Energy Phys. 07 (2014) 020; F. Demartin, F. Maltoni, K. Mawatari, B. Page, and M. Zaro, Eur. Phys. J. C **74**, 3065 (2014).

[23] P. Artoisenet, R. Frederix, O. Mattelaer, and R. Rietkerk, J. High Energy Phys. 03 (2013) 015.

[24] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H.-S. Shao, T. Stelzer, P. Torrielli, and M. Zaro, J. High Energy Phys. 07 (2014) 079.

[25] R. D. Ball *et al.*, Nucl. Phys. **B867**, 244 (2013).

[26] P. Artoisenet *et al.*, J. High Energy Phys. 11 (2013) 043.

[27] M. Czakon and A. Mitov, Comput. Phys. Commun. **185**, 2930 (2014).

[28] M. Botje *et al.*, arXiv:1101.0538.

[29] A. D. Martin, W. J. Stirling, R. S. Thorne, and G. Watt, Eur. Phys. J. C **64**, 653 (2009).

[30] J. Gao, M. Guzzi, J. Huston, H.-L. Lai, Z. Li, P. Nadolsky, J. Pumplin, D. Stump, and C.-P. Yuan, Phys. Rev. D **89**, 033009 (2014).

[31] N. Kidonakis, Phys. Rev. D **81**, 054028 (2010).

[32] N. Kidonakis, Phys. Rev. D **83**, 091503 (2011).

[33] M. Czakon, P. Fiedler, and A. Mitov, Phys. Rev. Lett. **110**, 252004 (2013).

[34] T. Sjostrand, S. Mrenna, and P. Z. Skands, J. High Energy Phys. 05 (2006) 026.

[35] J. Alwall *et al.*, Eur. Phys. J. C **53**, 473 (2008).

[36] S. Frixione and B. R. Webber, J. High Energy Phys. 06 (2002) 029.

[37] J. de Favereau, C. Delaere, P. Demin, A. Giammanco, V. Lemaître, A. Mertens, and M. Selvaggi (DELPHES 3 Collaboration), J. High Energy Phys. 02 (2014) 057.

[38] M. Cacciari, G. P. Salam, and G. Soyez, J. High Energy Phys. 04 (2008) 063.

[39] E. Conte, B. Fuks, and G. Serret, Comput. Phys. Commun. **184**, 222 (2013).

[40] E. Conte, B. Dumont, B. Fuks, and C. Wymant, Eur. Phys. J. C **74**, 3103 (2014).

[41] A. Hoecker, P. Speckmayer, J. Stelzer, J. Therhaag, E. von Toerne, and H. Voss, TMVA: Toolkit for multivariate data analysis, *Proc. Sci.*, ACAT2007 (2007) 040, arXiv:physics/0703039.