

Implementation of the CMS-EXO-22-024 analysis in the MADANALYSIS 5 framework

S. Ajmal¹, O. Panella¹, and M. Presilla²

¹INFN Sezione di Perugia, Via A. Pascoli, I-06123, Perugia, Italy

²Texas Tech University, Department of Physics and Astronomy, Lubbock, TX 79409, USA

Abstract

This paper presents the implementation and validation of the CMS-EXO-22-024 analysis within the MADANALYSIS 5 framework. The analysis searches for signatures of new physics in high-mass diphoton final states produced in proton-proton collisions at a center-of-mass energy of $\sqrt{s} = 13$ TeV. The dataset, corresponding to an integrated luminosity of 138 fb^{-1} , was collected by the CMS detector at the LHC during the 2016–2018 data-taking period. The validation of our implementation is performed by comparing our results against the official cutflow tables provided in the analysis note.

1 Introduction

The Standard Model is incomplete, with the hierarchy problem [1, 2], requiring extreme fine-tuning of the Higgs mass, motivating BSM physics. Proposed solutions include large extra dimensions (ADD), where gravity weakens by propagating in extra dimensions, and the warped Randall and Sundrum (RS) model [3, 4], where gravity is exponentially suppressed. Both predict Kaluza-Klein graviton excitations that decay into diphotons, either as a continuum excess (ADD) or as distinct resonances (RS). The continuum clockwork mechanism [5] offers an alternative with on-shell KK modes yielding a continuum-like signature. Additionally, high-mass diphoton events can probe other BSM scenarios, such as heavy spin-0 resonances from extended Higgs sectors [6, 7, 8]. Unlike the spin-2 graviton signatures, these spin-0 resonances would appear as narrow peaks in the diphoton mass spectrum, with production and decay governed by their scalar couplings to Standard Model particles.

In this paper, we present the implementation and validation of the CMS-EXO-22-024 analysis [9] within the MADANALYSIS 5 framework [10, 11, 12]. The original search, performed by the CMS Collaboration, looks for BSM physics in high-mass diphoton events using proton-proton collision data at $\sqrt{s} = 13$ TeV collected during the LHC Run 2 period (2016–2018), corresponding to an integrated luminosity of 138 fb^{-1} . Resonant signals from RS gravitons and heavy spin-0 resonances are probed using a data-driven functional fit to the diphoton spectrum, while nonresonant signals from ADD and clockwork models are investigated using NNLO QCD background calculations and data-driven jet-to-photon misidentification estimates from control samples. The analysis improves upon the previous CMS result through enhanced photon identification and calibration, improved background modelling, and advanced statistical inference techniques. Our implementation is validated by comparing our results against the official cutflow tables provided in the analysis note.

2 Description of Analysis

Photon candidates are reconstructed from ECAL energy deposits grouped into superclusters using an algorithm that preserves energy from bremsstrahlung and conversions, ensuring high efficiency for high- p_T photons. The same algorithm is applied to $Z \rightarrow e^+e^-$ events, which serve as a control sample to evaluate the photon energy scale, resolution, and selection efficiency.

Jet and electron backgrounds are suppressed by applying stringent photon identification criteria based on shower shape and isolation. The hadronic energy captured along the photon direction in the hadron calorimeter (HCAL) must be less than 5% of the total energy of the photon. Isolation variables are constructed using the total transverse momentum of particles reconstructed within a cone of size $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} < 0.3$ around the photon candidate. Two different isolation variables are defined, summing the transverse momenta of either charged hadrons or photons. The charged-hadron isolation is required to be less than 5.0 GeV, while the photon isolation is required to be less than 2.75 GeV in the ECAL barrel (EB) and less than 2.0 GeV in the ECAL endcaps (EE). The photon isolation is corrected by the average energy density in the event to account for contributions from the photon itself and other interactions within the isolation region. The electromagnetic shower shape is characterized using the second-order moment of the log-weighted crystal energies of the photon candidate in the η direction, denoted $\sigma_{i\eta i\eta}$. Also a selection on shower shape variable is imposed $R_9 > 0.8$, though these two variables are not currently available in MADANALYSIS 5. The reconstruction efficiency is 90% (EB) and 82% (EE), with weak p_T dependence.

In events with more than two photon candidates, the two photons leading in p_T are chosen for further analysis. The trigger photon candidates are required to have a ratio of hadronic to electromagnetic energy of less than 0.15. With these selections, the trigger efficiency is close to 100% for two photons reconstructed offline with $p_T > 125$ GeV, ensuring that the analysis is not limited by trigger inefficiencies in the high-mass signal region.

Two diphoton event categories are considered in this analysis. The first category consists of events where both photons are reconstructed in the fiducial region of the EB with $|\eta| < 1.44$ (denoted EBEB). The second category consists of events with one photon in the EB and the other photon in the fiducial region of the EE, defined as $1.57 < |\eta| < 2.50$ (denoted EBEE). Photons in the overlap region between $1.44 < |\eta| < 1.57$ are not used due to low reconstruction efficiency. The EBEE category typically provides an increase of approximately 10% to the signal sensitivity compared to using only the EBEB category. Events where both photons are in the EE are not considered because this category is dominated by SM backgrounds and has negligible sensitivity to the signal models under study. The minimum reconstructed invariant mass of the diphoton system is required to be $m_{\gamma\gamma} > 500$ GeV, defining the signal region. Photon pairs must additionally satisfy $\Delta R_{\gamma\gamma} > 0.45$ to be consistent with the theoretical calculations used for the SM diphoton background.

The energy of the photons is typically corrected by approximately 0.5%, with variations depending on the p_T and η of the photon. The additional Gaussian smearing needed in simulation to match the measured energy resolution in data varies between 0.8% and 1.5% for photon candidates in the EB region and between 2% and 2.5% for photon candidates in the EE region.

The primary interaction vertex is selected using a multivariate classifier that considers the p_T^2 sum of associated tracks, the p_T balance between the diphoton system and tracks, and photon conversion information when available. For signal events with $m_{\gamma\gamma}$, the correct vertex is assigned with approximately 90% efficiency, ensuring minimal degradation of the diphoton mass resolution.

3 Validation

3.1 Event Generation

Signal predictions are computed at leading order (LO) using the PYTHIA 8.316 program [13] with the NNPDF3.1 [14] LO parton distribution functions and the CP2 [15] underlying event tune. For the resonant search, we consider a spin-0 ($J = 0$) signal corresponding to a heavy SM-like Higgs boson, with a narrow width hypothesis of $\Gamma_X/m_X = 1.4 \times 10^{-4}$, where Γ_X and m_X are the natural width and mass of the resonance, respectively. Signal events are generated for mass points ranging from 1000 GeV to 5000 GeV in steps of 500 GeV, in both the EBEB and EBEE categories, with 100,000 events per point. The cross section values obtained from our implementation are compared with those reported in the analysis note, and are found to be in agreement at the 99% level.

3.2 Comparison with official results

To validate our MadAnalysis 5 (MA5) reimplementation of the CMS-EXO-22-024 analysis, we compare the selection efficiencies obtained from MA5 (version 1.9.60) with the official CMS results for resonance masses of 1000, 2000, 3000, 4000, and 5000 GeV. The efficiency ϵ is defined as

$$\epsilon = \frac{N_{selected}}{N_{generated}} \quad (1)$$

where N is the number of event count, and the relative difference δ^{rel} between the MA5 and CMS values is given as

$$\delta^{rel} = \left| 1 - \frac{\epsilon_i^{MA5}}{\epsilon_i^{CMS}} \right| \times 100 \quad (2)$$

where i corresponds to the cut number.

Cuts	1000 GeV			2000 GeV		
	CMS	MA5	δ^{rel}	CMS	MA5	δ^{rel}
$p_T > 125$ GeV	0.868	0.81718	5.85%	0.903	0.86967	3.69%
Photon ID	0.616	0.5747121	6.70%	0.648	0.5989702	7.57%
EBEB-EBEE	0.611	0.5695817	6.78%	0.637	0.5880571	7.68%
$m_{\gamma\gamma} > 500$ GeV	0.611	0.5694934	6.79%	0.637	0.5876026	7.75%
$\Delta R_{\gamma\gamma}$	0.611	0.5694934	6.79%	0.637	0.5876026	7.75%

Table 1: Comparison of selection efficiencies for 1000 and 2000 GeV resonance masses obtained from our MadAnalysis 5 reimplementation (MA5) and the CMS official results. Efficiencies are defined in Eq. 1, while the relative difference δ^{rel} between the two is computed using Eq. 2.

Due to limitations in the current MA5 release, our implementation does not include the shower shape variables $\sigma_{i\eta i\eta}$ and R_9 . The selection applied in our analysis therefore consists of the following cuts: $p_T > 125$ GeV, photon identification (excluding shower shape variables), EBEB-EBEE categorisation, $m_{\gamma\gamma} > 500$ GeV, and $\Delta R_{\gamma\gamma}$.

Cuts	3000 GeV			4000 GeV		
	CMS	MA5	δ^{rel}	CMS	MA5	δ^{rel}
$p_T > 125$	0.906	0.88281	2.56%	0.883	0.86545	1.99%
Photon ID	0.657	0.6017637	8.41%	0.64	0.5920917	7.49%
EBEB-EBEE	0.643	0.5877038	8.60%	0.625	0.575712	7.89%
$m_{\gamma\gamma} > 500$ GeV	0.643	0.5850308	9.02%	0.625	0.5569519	10.89%
$\Delta R_{\gamma\gamma}$	0.643	0.5850308	9.02%	0.625	0.5569519	10.89%

Table 2: The same efficiency comparison as in Table 1, but for resonance masses of 3000 and 4000 GeV.

Cuts	CMS (official)	MA5	δ^{rel}
$p_T > 125$ GeV	0.798	0.76645	3.95%
Photon ID	0.572	0.5328847	6.84%
EBEB-EBEE	0.558	0.5185625	7.07%
$m_{\gamma\gamma} > 500$ GeV	0.557	0.439922	21.02%
$\Delta R_{\gamma\gamma}$	0.557	0.439922	21.02%

Table 3: Efficiency comparison for a resonance mass of 5000 GeV, following the same format as in Table 1

The comparison is presented in Tables 1, 2, and 3. For the 1000 and 2000 GeV mass points, the MA5 and CMS efficiencies agree within 3.69–7.75% across all cuts, with the best agreement observed for the $p_T > 125$ GeV. For the 3000 and 4000 GeV masses, the agreement remains good, with relative differences below 12% for most cuts. For the 5000 GeV mass point, the first three cuts agree within 7.1%, but a 21% discrepancy appears after the $m_{\gamma\gamma} > 500$ GeV selection.

Moreover, In Fig. 1 the event selection efficiency ϵ is compared between the CMS and Ma5 configurations for the EBEB (red) and EBEE (blue) final-state categories, as a function of the signal resonance mass m_X (ranging from 1000 to 3000 GeV). The efficiencies are shown for the narrow-width hypothesis ($\Gamma_X/m_X = 1.4 \times 10^{-4}$) and spin-parity $J = 0$. The overall discrepancy stays below 15%, indicating good agreement between the two configurations.

Despite the absence of trigger and shower shape variables, the good agreement between MA5 and CMS results across all mass points validates our implementation, confirming its suitability for recasting the CMS-EXO-22-024 search.

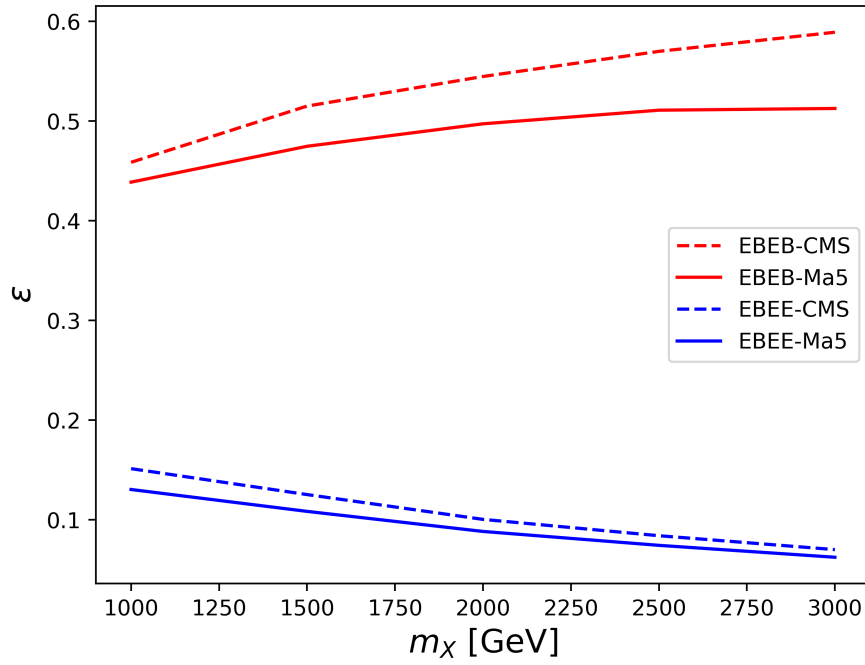


Figure 1: Event selection efficiency ϵ as a function of the signal resonance mass m_X (GeV) for the narrow-width hypothesis ($\Gamma_X/m_X = 1.4 \times 10^{-4}$) and spin-parity $J = 0$. Efficiencies are shown for the EBEB (red) and EBEE (blue) final-state categories. The solid curves correspond to the Ma5 configuration, while the dashed curves correspond to the CMS official values.

References

- [1] Steven Weinberg. Implications of dynamical symmetry breaking. *Phys. Rev. D*, 13:974–996, Feb 1976.
- [2] Leonard Susskind. Dynamics of spontaneous symmetry breaking in the weinberg-salam theory. *Phys. Rev. D*, 20:2619–2625, Nov 1979.
- [3] Lisa Randall and Raman Sundrum. Large mass hierarchy from a small extra dimension. *Phys. Rev. Lett.*, 83:3370–3373, Oct 1999.
- [4] Lisa Randall and Raman Sundrum. An alternative to compactification. *Phys. Rev. Lett.*, 83:4690–4693, Dec 1999.
- [5] Gian F. Giudice and Matthew McCullough. A clockwork theory. *Journal of High Energy Physics*, 2017(2), February 2017.
- [6] G.C. Branco, P.M. Ferreira, L. Lavoura, M.N. Rebelo, Marc Sher, and João P. Silva. Theory and phenomenology of two-higgs-doublet models. *Physics Reports*, 516(1-2):1–102, 2012.
- [7] Nathaniel Craig, Jamison Galloway, and Scott Thomas. Searching for signs of the second higgs doublet, 2013.
- [8] T. D. Lee. A theory of spontaneous t violation. *Phys. Rev. D*, 8:1226–1239, Aug 1973.
- [9] Aram Hayrapetyan et al. Search for new physics in high-mass diphoton events from proton-proton collisions at $\sqrt{s} = 13$ TeV. *JHEP*, 08:215, 2024.
- [10] Eric Conte and Benjamin Fuks. Confronting new physics theories to lhc data with madanalysis 5. *International Journal of Modern Physics A*, 33(28):1830027, 2018.
- [11] B. Dumont, B. Fuks, S. Kraml, S. Bein, G. Chalons, E. Conte, S. Kulkarni, D. Sengupta, and C. Wymant. Toward a public analysis database for lhc new physics searches using madanalysis 5. *The European Physical Journal C*, 75(2), February 2015.
- [12] Eric Conte, Benjamin Fuks, and Guillaume Serret. Madanalysis 5, a user-friendly framework for collider phenomenology. *Computer Physics Communications*, 184(1):222–256, January 2013.
- [13] Christian Bierlich, Smita Chakraborty, Nishita Desai, Leif Gellersen, Ilkka Helenius, Philip Ilten, Leif Lönnblad, Stephen Mrenna, Stefan Prestel, Christian T. Preuss, Torbjörn Sjöstrand, Peter Skands, Marius Uthmeim, and Rob Verheyen. A comprehensive guide to the physics and usage of pythia 8.3, 2022.
- [14] Richard D. and Ball, Valerio Bertone, Stefano Carrazza, Christopher S. Deans, Luigi Del Debbio, Stefano Forte, Alberto Guffanti, Nathan P. Hartland, José I. Latorre, Juan Rojo, and Maria Ubiali. Parton distributions for the lhc run ii. *Journal of High Energy Physics*, 2015(4), April 2015.
- [15] Extraction and validation of a new set of cms pythia8 tunes from underlying-event measurements. *The European Physical Journal C*, 80(1), January 2020.