ATLAS Note

ANA-HDBS-2021-23-INT1

April 17, 2023

Draft version 0.95

Search for $X \to SH$ model in the final states of two photons and multiple leptons using 139 fb^{-1} of proton-proton collision data at $\sqrt{s}$ = 13 TeV recorded with the ATLAS detector at the LHC

Yaquan Fang[a], Kaili Zhang[a], Zhijun Liang[a], Bo Liu[a], Xinchou Lou[a,c], Xin Shi[a], Qiyu Sha[a], Shuiting Xin[a], Wei-Ming Yao[b], Yesenia Hernandez Jimenez[d], Fangyi Guo[a]

[a] *Institute of High Energy Physics, Chinses Academy of Sciences, Beijing, China*
[b] *Lawrence Berkeley National Laboratory, USA*
[c] *University of Texas at Dallas, USA*
[d] *University of Stony Brook University, USA*

This note presents a search for a new heavy scalar particle X decaying into a Standard Model Higgs boson and a singlet scalar particle S using 139 fb^{-1} of proton-proton collision data at the centre-of-mass energy of 13 TeV recorded with the ATLAS detector at LHC. The explored X mass range varies from 300 GeV to 1000 GeV, with the corresponding S mass range being from 170 GeV to 500 GeV. This search uses the event signature of two photons from the Higgs boson decay and one or two leptons (e or μ) coming from the process of $S \to WW/ZZ$. The observed (expected) upper limits at the 95% confidence level on the cross-section for $gg \to X \to Sh$ assuming the decay of S following the SM prediction is between X fb (167 fb) and Y fb (710 fb).

List of contributions

Yaquan Fang	Contact editor, optimization, supervision of IHEP students
Kaili Zhang	Contact editor, optimization, plots production, signal/background estimation, systematics and limit setting
Zhijun Liang	Analysis Optimization for BDT and supervision of Bo Liu
Bo Liu	Analysis Strategy and editor of the note
Xinchou Lou	Statisics and supervisiion of Shuiting Xin
Xin Shi	Optimization of the analysis and supervisiion of Kaili Zhang
Qiyu Sha	Signal sample preparation, plotting
Shuiting Xin	Signal sample preparation
Wei-Ming Yao	the optimiazation of BDT analysis and editor of the note
Yesenia Hernandez Jimenez	Editor of the note
Fangyi Guo	Editor of the note

Contents

Changelogs

0.1

Updated time: 19th Sep. 2021.

- Refine the MC ratio to make MC and sideband data more consistent.
- Use BDT instead of BDTG. better consistence between data sideband and MC.
- More description for analysis strategy.

0.2

Updated time: 30th Nov. 2021.

- Updating systematics implementation. Major NPs studied.
- More description for analysis strategy.

0.3

Updated time: 1st Dec. 2021.

- 5.2 Adding the definations for BDT-High and BDT-Middle regions and the BDT cut values.
- 6 Adding the statements for ZZ2l and WW1e1m. Replace plots.
- 6.2 and 7.2.2. How the background is modelled and How the lepton dependence systematics implemented.

0.4

Updated time: 3rd Dec. 2021.

- Replying the questions from convenors.

0.5

Updated time: 6th Dec. 2021.

- Replying the questions from CDS comments.
- Lepton efficiencies explained in p.19 Section 4.2.
- Cross validation method the ks test plots are replaced by the individual folds plots to avoid confusion. See Section 5.1.2
- Alternative cut solution discussed.
- Better text all over the draft.

0.6

Updated time: June 2022, aiming for 2nd EB circulation.

- Section 7: Add details about syst.
- Add Appendix C: visualized the cutflow and moved the numbers in appendix.
- Add Appendix G: diphoton vertex check for SH samples

0.7

Updated time July 2022, aiming for the 2nd EB meeting in August.

- Section 6: Add the spurious signal study.
- Add Appendix I: The inject test is added.
- Add Appendix H: The toylimit is added.
- Add Appendix E: The detail of reweighted method is added.

0.8

Updated time November 2022, aiming for the unblinding approval in December.

- Modify Section 7: Theoretical uncertainties calulation.
- Fill Appendix H, A: Complete Limits, pulls, rankings, toy limits for all mass points.
- Add Appendix E: Adjust description.
- Modify Section 5.2, change to at least 2 sideband event, the impact of different threshold events on limits, change related tables and plots.

0.9

Update for version 0.8.1. aiming for unblinding approval.

- Modify Section 7: Theoretical uncertainties calulation.
- Fill A: Complete Limits, pulls, rankings, toy limits for all mass points.
- Modify Section 5.2, add more scans.

0.10

Update for version 0.8.5. aiming for answering all the commentes received.

-

1 Introduction

The discovery of the Higgs boson (H) by the ATLAS and CMS Collaborations in 2012 [1, 2] is a monumental achievement in the particle physics. Extensive studies of its properties suggest that the observed particle's behavior is consistent with the predictions of the Standard Model (SM) prediction. However, this does not preclude the possibility of new physics, such as the presence of a heavy scalar boson that couples to two Higgs bosons or mixtures of additional scalars that couple to the SM Higgs boson.

In the former scenario, the introduction of extended Higgs sectors could significantly enhance Higgs boson pair production. One of the most common methods to explore this possibility is to search the heavier CP-even component of the two Higgs-doublet model (2HDM) decaying into two Higgs bosons. Both ATLAS and CMS have conducted exhaustive searches for a heavy Higgs boson using Run-1 and Run-2 data, with no notable excess observed [3–5].

Another possibility to explore new physics beyond the SM, is to extend the model by introducing a new scalar singlet, S, through the decay process $X \to SH$ [6, 7], where X is the heavy CP-even scalar particle predicted by the 2HDM model, and H denotes the SM Higgs boson. In this model, the heavy scalar X has Yukawa couplings to other SM particles and is associated with Electroweak Symmetry Breaking(EWSB), where the Higgs field obtains mass through to the Vacuum expectation value(VEV). The scalar singlet S is assumed to have the similar behavior as a Higgs-like particle. At the generation level, S will interact with X and the SM Higgs boson, and for the certain mass, S will have decay modes and branching ratios that are the same as the Higgs-like particle predicted by the SM.

This enables the production of S through the process of $X \to SH, SS$. The production processes are illustrated with Feynman diagrams presented in Figure 1 where X is produced dominantly through gluon fusion (ggF), and has the dominant decay modes $X \to SS$ and $X \to SH$.

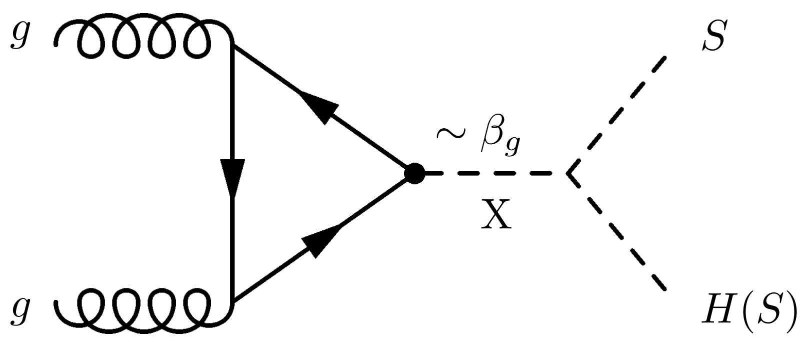


Figure 1: Representative diagrams that contribute to $X \to SH, SS$ via the gluon fusion process.

The ggF production cross-section is considered to be SM-like as a function of m_X, which is also modifiable by a free parameter β_g, a dimensionless constant that multiplies the effective g-g-X coupling in the BSM Lagrangian. A more concrete statement would be that the production cross section is given by:

$$\sigma_X = \beta_g^2 \sigma_{\mathrm{SM}}^{\mathrm{ggF}}(m_X). \tag{1}$$

The relative decay rates of these two X decay modes are considered as a free parameter, and can be controlled by the ratio a_1 of their branching ratios (BRs):

$$a_1 = \frac{\mathrm{BR}(X \to SS)}{\mathrm{BR}(X \to SH)}. \tag{2}$$

It is convenient to have the assumption of $\mathrm{BR}(X \to SH) = 1$ by considering the lack of sensitivity to probe a_1 directly. In addition, the cross-section of X from ggF production is assumed to be 1 pb as a benchmark, and X mass has been chosen from 300 to 1000 GeV. The BSM model often treats S as a portal to the Dark Matter (DM) interactions through the decay $S \to \chi\chi$ where χ is a DM candidate. However, the DM interpretation completely decouples from this search and $\mathrm{BR}(S \to \chi\chi)$ can be set to 0 without any loss of generality. Moreover, in this study, S is assumed to share the same decay BRs as the SM Higgs like particle[8, 9], as documented in the Higgs Yellow Report 4[10].

In addition, for the convenience of the theoretical interpretations, the upper limits assuming 100% decaying to $W^{\pm}W^{\mp}$ or ZZ will also be addressed in the note.

For the chain decay of X, on the one hand, S masses are assumed to be higher than H, ranging from 170 GeV up to 500 GeV. Thus, we focus on the dominant decays of S, i.e. $S \to W^{\pm}W^{\mp}$ and $S \to ZZ$ in this study. The sum of these two decay modes counts more than 80% of the total branch ratio of S decay. The rest is mostly occupied by $t\bar{t}$ decay in particular if S mass is higher than $t\bar{t}$ mass threshold. Meanwhile, the lower bound of 170 GeV is chosen to ensure the 2 W boson particles are on-shell. Z bosons can be off-shell at 170 GeV but its contribution is relatively small compared to the WW channels.

In order to effectively reject huge QCD backgrounds, one or two lepton the final states from $S \to W^{\pm}W^{\mp} or ZZ$ are required in this analysis. These requirements lead to this search mainly in final states of $X \to SH \to \gamma\gamma + 1, 2$ leptons.

On the other hand, $\gamma\gamma$ decay of the H is chosen to take the advantage of the excellent di-photon mass resolution and smooth sideband when estimating backgrounds as well as to further suppress the QCD background.

Finally, b-jet veto is implemented to be orthogonal with process $X \to SH \to bb\gamma\gamma$, also leading to the significant reduction of $X \to SH \to t\bar{t}\gamma\gamma$ process.

This note provides supporting material for the search of $X \to SH$ model in the final states of di-photon and multi-leptons using 139 fb^{-1} of proton-proton collision data at $\sqrt{s} = 13$ TeV. It is organized as follows. In Section 2, the current data set and the MC samples relevant for this analysis are described. Section 3 defines the objects such as photon, lepton, jet etc. used in this analysis. In Section 4, the event selections are summarized. Section 5 describes the BDT method to optimize the analysis. The estimations of the signal and various backgrounds are discussed in Section 6. In Section 7, the systematic uncertainties are presented. Section 8 documents the statistical procedure used to extract the sensitivity of the analysis. Section 9 includes the upper limits and the extrapolations to the whole X-S plane. Finally, Section 10 provides the summarizes the results and conclusions of the study in this note.

2 Data and Monte Carlo samples

This section will introduce the current data set and the MC samples relevant for this analysis. As this study is heavily rely on the photon performance, all those samples are reconstructed into `HIGG1D1` derivation by the `HGamAnalysisFramework`, which is maintained by the ATLAS HGam group. For this specific analysis, the `HGamCore` release tag is `v1.10.33-h026`, the samples used are classfies as h026. The corresponding Athena AnalysisBase release tag is 21.2.131.

2.1 Data samples

The data samples used in this analysis correspond to the data collected by ATLAS during 2015-2018 at the center of mass $\sqrt{s} = 13$ TeV, which sums up to an integrated luminosity of $(139.0 \pm 2.4)\,\mathrm{fb}^{-1}$ [11, 12] after the data quality requirement [13]. The Good Run List (GRL) is presented in the following:

- Year 2015: data15_13TeV.periodAllYear_DetStatus-v89-pro21-02_Unknown_PHYS_StandardGRL_All_Good_25ns.xml
- Year 2016: data16_13TeV.periodAllYear_DetStatus-v89-pro21-01_DQDefects-00-02-04_PHYS_StandardGRL_All_Good_25ns.xml
- Year 2017: data17_13TeV.periodAllYear_DetStatus-v99-pro22-01_Unknown_PHYS_StandardGRL_All_Good_25ns_Triggerno17e33prim.xml
- Year 2018: data18_13TeV.periodAllYear_DetStatus-v102-pro22-04_Unknown_PHYS_StandardGRL_All_Good_25ns_Triggerno17e33prim.xml

Events are selected by the diphoton trigger, requiring transverse energy thresholds of 35 GeV and 25 GeV for leading and sub-leading photon candidates. Loose photon isolation requirements [14] are applied by this diphoton trigger in 2015-2016, for `HLT_g35_loose_g25_loose` and are tightened in 2017 to cope with a higher instantaneous luminosity, for `HLT_g35_medium_g25_medium_L12EM20VH`.

2.2 Monte Carlo samples

Signal samples, SM single Higgs and SM di-Higgs events are estimated with Monte Carlo (MC) simulated samples that are documented in this section, while the continuum photon background of the SM processes with di-photon and multiple leptons is determined with the data in sideband[1] with the data-driven method.

Nevertheless, several relevant MC samples, vector boson production associated with photons ($V + \gamma\gamma$, $VV + \gamma\gamma$), $t\bar{t} + \gamma\gamma$ processes, and multi-jet processes associated with photons ($\gamma\gamma$+jets) are used to check and validate the performance of TMVA modeling. Note that these samples are not used to estimate the background. These samples, In addition to the 3 $\gamma\gamma$ samples which to quantify the lepton dependance uncertainties, are also described in this section.

There are three MC campaigns used for each simulated process, mc16a, mc16d and mc16e, which correspond to different assumptions on the distribution of the number of interactions per branching crossing in 2015-2016, 2017 and 2018 periods, respectively. To match the number of interactions in data, the Monte Carlo samples are reweighted to the observed distribution using the `PileupReweightingTool` [15].

[1] The sideband is defined as $m_{\gamma\gamma} \in [105, 120] \cup [130, 160]$ GeV orthogonal to the signal region defined in Section 4.

These multiple overlaid proton-proton collisions are simulated with the soft QCD processes of Pythia 8.186 [16] using the A2 set of tuned parameters [17] and the MSTW2008LO PDF set.

2.2.1 MC samples for signals

For $X \to SH$ signal production, the event generation is performed at the leading-order (LO) accuracy with Pythia8 [16] for matrix element calculation. Parton showering and hadronization are also simulated using the Pythia8 generator with the A14 tune [18] and using the NNPDF 2.3 LO PDF set [19]. The EvtGen [20] program is used for b- and c-hadron modeling. Detector effects are simulated using AltFastII(AF2) [21], in which the calorimeter response is fast simulated.

Two scalars, X and S, are assumed to have a narrow width with respect to the experimental resolution. Technically, their decay widths are set to 10 MeV in the event generation. The heavier boson X is constrained to decay only to S and H, with S only decaying to a pair of W or Z bosons. Then, the lepton filter, named `MultiLeptonFilter`, is assigned to the generator. In the filter level, the lepton is defined as the electron or the muon. Taus are not included in this analysis. `MultiLeptonFilter` requires the truth leptons with $p_{\mathrm{T}} > 7$ GeV and $|\eta| < 3$. For WW channels, the samples are generated by requiring the lepton filter to have exactly one lepton for semi-leptonic decays, and at least two leptons for full leptonic decays, which is $\gamma\gamma + WW(\ell\nu q\bar{q}')$ and $\gamma\gamma + WW(\ell\nu\ell\nu)$. For ZZ case, the filter requires exactly two leptons in the truth level.[2] So in principle, $ZZ(\ell\ell q\bar{q})$, and also $ZZ(\ell\ell\nu\nu)$, are included in the sample. The contribution from $ZZ(\ell\ell\ell\ell)$ final state is relatively small and not included in this study.

The SM H boson from X, is required to decay into a pair of photons. Thus three dedicated samples with the following final states are produced: $\gamma\gamma + WW(\ell\nu q\bar{q}')$, $\gamma\gamma + WW(\ell\nu\ell\nu)$ and $\gamma\gamma + ZZ(\ell\ell q\bar{q}/\ell\ell\nu\nu)$. Each of these 3 samples has 200k events for one specific X-S mass point.

In order to perform an efficient search on $X \to S + H$ process, a total of 20 signal MC simulated samples corresponding to various combinations of m_X and m_S hypotheses for each of the three final states are generated to cover the most interesting phase space. The mass grid is shown in Figure 2. The signal sample lists of all three channels $\gamma\gamma + WW(\ell\nu q\bar{q}')$, $\gamma\gamma + WW(\ell\nu\ell\nu)$ and $\gamma\gamma + ZZ(\ell\ell q\bar{q}/\ell\ell\nu\nu)$ are presented in Table 2, Table 3 and Table 4, respectively.

Taking $\gamma\gamma + WW(\ell\nu q\bar{q}')$ in $m(X, S) = (1000, 170)$ GeV as an example, the value of the cross-section in the first row of Table 2 is calculated by the following formula:

$$
\begin{aligned}
&\sigma(X \to SH) \times Br_{(S_{170} \to WW)} \times Br_{(H_{125} \to \gamma\gamma)} \times k_{Eff_{Filter}} \\
&= 1pb \times 96.28\% \times 0.0228\% \times 38.74\% \\
&= 8.50 \times 10^{-4} pb
\end{aligned}
\tag{3}
$$

One example jobOption file to generate signal samples with Pythia8 is in Appendix D. For $\gamma\gamma + WW(\ell\nu q\bar{q}')$ channel, the S paricle is forced to decay to W bosons. For $\gamma\gamma + WW(\ell\nu\ell\nu)$ channel, both W bosons are forced to decay leptonically at the generator level, which means an extra $Br_{(WW \to \ell\nu\ell\nu)} = 10.71\%$ counted. For $\gamma\gamma + ZZ(\ell\ell q\bar{q}/\ell\ell\nu\nu)$ channel, the S paricle is forced to decay to Z bosons. process, supposed to be 20% of $\gamma\gamma + ZZ \to \gamma\gamma\ell\ell qq$, is included at the generator level.

In this study, the expected branching ratios of the decay from S are assumed to be the same as the heavy Standard Model-like Higgs boson, as well as 100% decay to WW/ZZ. In either case, the Parameter Of

[2] $\bar{q}'$ refers to the different flavor quark with respect to q.

Interest(POI) of this study will be the $\sigma(pp \to X) \times Br(X \to Sh)$. In former case, the branching ratios of S decay to WW and ZZ are fixed at the SM prediction, as shown in Table 1.

m_S[GeV]	$BR(S \to WW)$	$BR(S \to ZZ)$
170	96.28%	2.44%
200	73.90%	25.68%
300	69.12%	30.72%
400	57.65%	26.90%
500	54.09%	25.86%

Table 1: The branching ratios of WW and ZZ from S following the decay of a Higgs-like particle for different masses from [22]

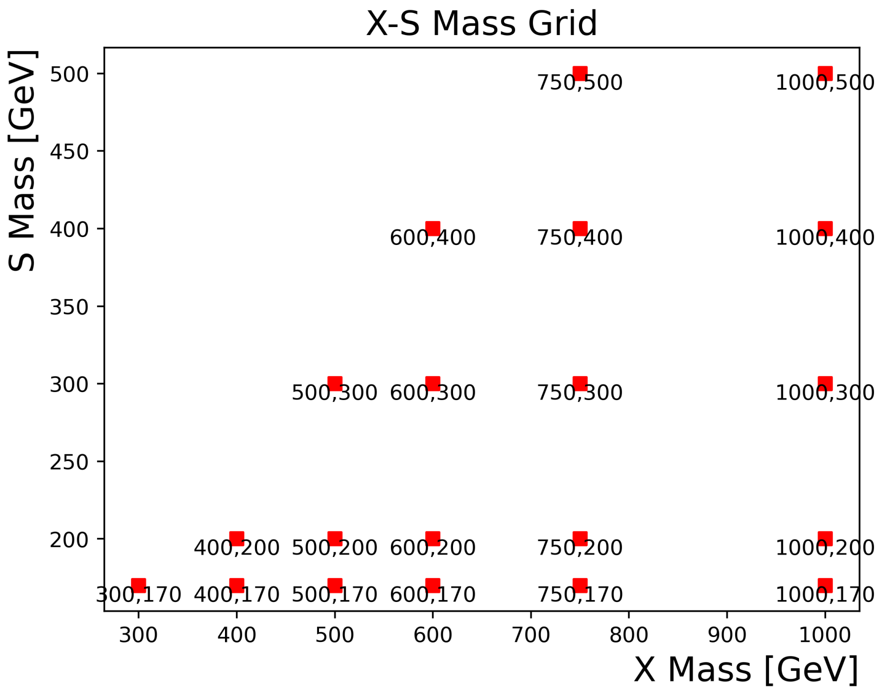


Figure 2: The m_X and m_S grid for the generated signal samples.

2.2.2 MC samples for SM single Higgs and Di-Higgs backgrounds

Simulated samples for SM single Higgs backgrounds are produced to investigate their contributions in $m_{\gamma\gamma}$ peak around 125 GeV. The SM single Higgs backgrounds considered here are produced via several production modes: ggH, VBF, WH, ZH, ttH, bbH, tHjb and tWH, where H represents the 125 GeV Higgs boson.

The MC simulated events for the ggH process are produced with the Powhegv2 generator at the next-to-leading-order (NLO) accuracy and interfaced to Pythia8 for Parton showering (PS). The PDF4LHC15 PDF set is used for incoming parton description in the matrix element (ME) calculation, and the CTEQ6L1 [23] set is used for parton showering calculation with the AZNLO tuned parameters for hadronization and factorization [24]. The EvtGen program is used for b- and c-hadrons modeling.

Kinematic distributions are also reweighted to the next-to-next-to-leading-order (NNLO) and next-to-next-to-leading-logarithmic (NNLL) calculation to have a better prediction of the Higgs boson p_{T} and rapidity [25]. The Parton shower simulation is also at the NNLO level, called NNLOPS [26]. Events are normalized to the cross-section calculated at the next-to-next-to-next-to-leading-order (N^{3}LO) QCD accuracy with the NLO EW corrections [27].

The VBF production events are generated with the Powhegv2 generator at the NLO accuracy and interfaced to Pythia8 for Parton showering. The PDF4LHC15 set is used for ME calculation, and the CTEQ6L1 set is used for PS with AZNLO tuned parameters incorporated with EvtGen for b- and c-hadrons modeling. Generated events are normalized to the cross-section calculated at the NLO QCD accuracy with the NNLO QCD and NLO EW corrections applied.

Events corresponding to the Higgs boson production in association with the vector boson (WH and ZH) are generated with Powhegv2 for ME and interfaced to Pythia8 for PS. The PDF4LHC15 set is used for ME calculation and the CTEQ6L1 set is used for PS with the AZNLO tune. The EvtGen program is used for b- and c-hadron modeling. All samples except for $gg \to ZH$ are generated at the NLO QCD accuracy while events for $gg \to ZH$ are produced at the LO QCD accuracy. Events are finally normalized to corresponding higher-order cross-sections respectively.

DSID	m_X[GeV]	m_S[GeV]	Cross Section [pb]	Filter efficiency	N_{events}
800938	1000	170	0.0008504143	38.74%	200000
800939	1000	200	0.0006439764	38.22%	200000
800940	1000	300	0.0006131967	38.91%	200000
800941	1000	400	0.0005170928	39.34%	200000
800942	1000	500	0.000494904	40.13%	200000
800943	300	170	0.0008488777	38.67%	200000
800944	400	170	0.0008657806	39.44%	200000
800945	400	200	0.0006562763	38.95%	200000
800946	500	170	0.0008574389	39.06%	200000
800947	500	200	0.0006796967	40.34%	200000
800948	500	300	0.0005700161	36.17%	200000
800949	600	170	0.0008302186	37.82%	200000
800950	600	200	0.0006372367	37.82%	200000
800951	600	300	0.000611148	38.78%	200000
800952	600	400	0.0005153841	39.21%	200000
800953	750	170	0.00081134	36.96%	200000
800954	750	200	0.0006402696	38.00%	200000
800955	750	300	0.0006046866	38.37%	200000
800956	750	400	0.0005139382	39.10%	200000
800957	750	500	0.0004878745	39.56%	200000

Table 2: Signal samples for $\gamma\gamma + WW(\ell\nu q\bar{q}')$ final state.

Both ttH and bbH events are generated with the POWHEGV2 generator and interfaced to PYTHIA8 for PS. The PDF4LHC15 set is used for ME calculation and the NNPDF2.3 set [28] is used for PS calculation with the A14 tune [18]. The EVTGEN program is used for b- and c-hadron modeling. Events are normalized to cross-section calculation at the NLO QCD accuracy with the NLO EW correction applied.

Samples for single Higgs boson production in association with single top-quark are generated with the MADGRAPH5_AMC@NLO generator for ME calculation at the NLO QCD accuracy and interfaced to PYTHIA8 for Parton showering. Two final state samples are considered: tHW and $tHbj$ in this analysis. The NNPDF3.0 set is used for ME calculation and the NNPDF2.3 set is used for PS with A14 tune incorporated with EVTGEN for b- and c-hadron modeling.

SM ggF Di-Higgs processes are generated with POWHEG-Box-V2 generator at NLO and interfaced to HERWIG7.1 for Parton shower. The PDF4LHC15 set is used for ME calculation. A set of lepton filters are applied targeting $\gamma\gamma$+multi-lepton final states in this analysis, requiring the lepton kinematic at $p_{\mathrm{T}} > 7GeV$ and $|\eta| < 3$. No kinematic cut for the photon is required at the generator level.

These samples are simulated using the full ATLAS simulation and reconstruction chain. The mass of the SM Higgs boson is set to 125 GeV.

A summary for all the sample used is listed in Table 5. Note that for all the PDF matrix elements calculations, PDF4LHC15 set is used.

DSID	m_X[GeV]	m_S[GeV]	Cross Section [pb]	Filter efficiency	N_{events}
800958	1000	170	0.0001352752	57.56%	200000
800959	1000	200	0.0001054183	58.44%	200000
800960	1000	300	9.95782e-05	59.02%	200000
800961	1000	400	8.38138e-05	59.56%	200000
800962	1000	500	7.81892e-05	59.22%	200000
800963	300	170	0.0001246525	53.04%	200000
800964	400	170	0.0001272142	54.13%	200000
800965	400	200	9.83111e-05	54.50%	200000
800966	500	170	0.0001303869	55.48%	200000
800967	500	200	9.93753e-05	55.09%	200000
800968	500	300	9.63557e-05	57.11%	200000
800969	600	170	0.0001313035	55.87%	200000
800970	600	200	0.0001013776	56.20%	200000
800971	600	300	9.62207e-05	57.03%	200000
800972	600	400	8.14497e-05	57.88%	200000
800973	750	170	0.0001331366	56.65%	200000
800974	750	200	0.0001035423	57.40%	200000
800975	750	300	9.76548e-05	57.88%	200000
800976	750	400	8.22799e-05	58.47%	200000
800977	750	500	7.78723e-05	58.98%	200000

Table 3: Signal samples for $\gamma\gamma + WW(\ell\nu\ell\nu)$ final state.

2.2.3 MC samples for continuum backgrounds

Several continuum backgrounds are used in this study: $\gamma\gamma + multijets$, $V(VV) + \gamma\gamma$ and $t\bar{t}\gamma\gamma$. In addition to these samples, dedicated $\gamma\gamma$+ 0ℓ, $\gamma\gamma$ + $\ell\nu$ jj, and $\gamma\gamma$ + $\ell\nu\ell\nu$ are generated to demonstrate the consistency of $m_{\gamma\gamma}$ distributions with the above three samples. These samples are used to quantify the systematic uncertainty in background modelling, which called lepton dependance uncertainty in the following section. $\gamma\gamma + multijets$ samples, generated with Sherpa 2.2.4, describe the continuum background shape for the diphoton spectrum. The main processes are dressed up by ME NNPDF30 and PDF4LHC15 in NNLO when the number of jets is equal to 0 or 1. For events with more than two jets, the accuracy is NLO. Moreover, the diphoton mass spectrum has been constrained to between 90 GeV to 175 GeV in LO. Note that multijets samples do not have real leptons in the final states, so these samples can be treated as a fake background since the lepton originates from a misidentified jet. As the misidentified component, the yield and the shape is not so consistent with the data behavior. So, multijets sample is only used in the BDT trainging but never in the background estimation.

$V(VV) + \gamma\gamma$ samples are generated with generator Sherpa in version 2.2.4 and the basic accuracy is also NNLO in NNPDF30. Additionally, the photon pT is required to be larger than 17 GeV and diphoton mass larger than 80 GeV in LO accuracy. with the different final states, those samples are separated as $ee + \gamma\gamma$, $\mu\mu + \gamma\gamma$, $\tau\tau + \gamma\gamma$, $e\nu + \gamma\gamma$, $\mu\nu + \gamma\gamma$, $\tau\nu + \gamma\gamma$, and $\nu\nu + \gamma\gamma$. Those processes had one or two real leptons in the final states and share the similar kinematics as our signals, so they make important contributions the background, especially in 2-lepton cases.

DSID	m_X[GeV]	m_S[GeV]	Cross Section [pb]	Filter efficiency	N_{events}
800978	1000	170	7.8942e-06	14.19%	200000
800979	1000	200	8.74157e-05	14.93%	200000
800980	1000	300	0.000109405	15.62%	200000
800981	1000	400	9.66592e-05	15.76%	200000
800982	1000	500	9.55755e-05	16.21%	200000
800983	300	170	7.2488e-06	13.03%	200000
800984	400	170	7.4936e-06	13.47%	200000
800985	400	200	8.32587e-05	14.22%	200000
800986	500	170	7.7106e-06	13.86%	200000
800987	500	200	8.95236e-05	15.29%	200000
800988	500	300	0.0001054826	15.06%	200000
800989	600	170	8.011e-06	14.40%	200000
800990	600	200	8.44297e-05	14.42%	200000
800991	600	300	0.0001033814	14.76%	200000
800992	600	400	9.11394e-05	14.86%	200000
800993	750	170	7.7551e-06	13.94%	200000
800994	750	200	9.0636e-05	15.48%	200000
800995	750	300	0.0001070936	15.29%	200000
800996	750	400	9.45126e-05	15.41%	200000
800997	750	500	9.38656e-05	15.92%	200000

Table 4: Signal samples for $\gamma\gamma + ZZ(\ell\ell q\bar{q}/\ell\ell\nu\nu)$ final state.

DSID	Generator	Prod. Mode	Events in AOD
343981	NNLOPS + Pythia8	ggH	18.3M
346214	Powheg + Pythia8	VBF	7M
345318	Powheg + Pythia8	$W^{+}H$	0.6M
345317	Powheg + Pythia8	$W^{-}H$	0.6M
345319	Powheg + Pythia8	$qq \rightarrow ZH$	1.5M
345061	Powheg + Pythia8	$gg \rightarrow ZH$	0.15M
346525	Powheg + Pythia8	ttH	7.8M
345315	Powheg + Pythia8	bbH	0.299M
346188	MGMCatNLO + PYTHIA8	tHbj	0.4M
346486	MGMCatNLO + PYTHIA8	tHW	0.208M
364352	SHERPA2 (ME@NLO+PS)	$\gamma\gamma + jets$	506.6M (AF2)
364862	Sherpa 2.2.4	$ee + \gamma\gamma$	0.6M
364865	Sherpa 2.2.4	$\mu\mu + \gamma\gamma$	0.5M
364868	Sherpa 2.2.4	$\tau\tau + \gamma\gamma$	0.5M
364874	Sherpa 2.2.4	$e\nu + \gamma\gamma$	0.6M
364877	Sherpa 2.2.4	$\mu\nu + \gamma\gamma$	0.6M
364880	Sherpa 2.2.4	$\tau\mu + \gamma\gamma$	0.5M
364871	Sherpa 2.2.4	$\nu\nu + \gamma\gamma$	0.1M
345868	MGMCatNLO + PYTHIA8	$t\bar{t}\gamma\gamma$ (noallhad)	1.94M
345869	MGMCatNLO + PYTHIA8	$t\bar{t}\gamma\gamma$ (allhad)	1.6M
600542	Powheg + Herwig7	SM Dihiggs $\gamma\gamma$+0L	0.1M
600543	Powheg + Herwig7	SM Dihiggs $\gamma\gamma$+1L	0.5M
600544	Powheg + Herwig7	SM Dihiggs $\gamma\gamma$+2L	0.5M
507017	MadGraph+ PYTHIA8	$\gamma\gamma$+jj	500k
504650	MadGraph+ PYTHIA8	$\gamma\gamma$+lνjj	200k
507018	MadGraph+ PYTHIA8	$\gamma\gamma$+ll$\nu\nu$	200k

Table 5: Summary of all the MC samples used in the analysis, including nominal continuum backgrounds, single Higgs, dihiggs and di-photon background samples used in the analysis.

Another sample used in this analysis is the top-pair production in association with two photons where both top-quarks decay hadronically or one of them decay leptonically: $t\bar{t}\gamma\gamma$ (noallhad) and $t\bar{t}\gamma\gamma$ (allhad). Events for such processes are generated with MadGraph5_aMC@NLO generator and interfaced to Pythia8 for Parton showering.

Although the MCs have been simulated to describe the continuum backgrounds, only the data sideband is reliable. The consistency study between MC and data and rescaling of the MC corresponding to the data sideband will be done in Section 5.

3 Object definition

This section outlines the photon, lepton, jet, and E_T^{miss} selections used in this analysis.

3.1 Photons

The photon is reconstructed by using the supercluster method with the energy deposits in the EM calorimeter. The detailed photon performance for Run-2 analyses can be found in Ref. [29]. A photon candidate is required to have $p_{\mathrm{T}} > 25$ GeV and $|\eta| < 2.37$. Photon inside the crack region $1.37 < |\eta| < 1.52$ is rejected. The photon candidate is also required to pass the *Tight* cut-based photon identification selection which is based on the longitudinal and transverse shower profiles measured in the calorimeter. In addition, the photon candidate has to be isolated and passes both calorimeter-based isolation topoEtCone20$< 0.065 \times p_{\mathrm{T}}$ and track-based isolation ptcone20$< 0.05 \times p_{\mathrm{T}}$. A candidate event is required to have at least two good isolated photons. To match the trigger threshold, the leading photon is required to have $p_{\mathrm{T}} > 35$ GeV and the subleading photon with $p_{\mathrm{T}} > 25$ GeV.

3.2 Jets and b-jets

The jet used in this analysis is reconstructed by the anti-k_t algorithm with radius parameter $R = 0.4$ from the particle-flow objects. The particle-flow (PFlow) algorithm provides a list of tracks and a list of topo-clusters containing both the unmodified topo-clusters and a set of new topo-clusters resulting from the energy subtraction procedure. The algorithm attempts to match each track to a single topo-cluster in the calorimeter. The expected energy deposited in the calorimeter (based on topo-cluster position and the track momentum) is subtracted cell by cell from the set of matched topo-clusters. If the remaining energy is consistent with the expected shower fluctuations of a single particle's signal, the topo-cluster remnants are removed [30].

To increase the efficiency of primary vertex identification in the presence of photons, HGam analyses are using a Neural Network relying on the tracks as well as the di-photon system [31]. The corresponding efficiency check for this study has been documented in Appendix G. The reconstructed jet collection is called `AntiKt4PFlowCustomVtxHggJets` [29] and is used as default in all analyses as well as in this document unless stated differently. Technical details on the collection used are shown in Table 6.

Collection name:	`AntiKt4PFlowCustomVtxHggJets`, `AntiKt4EMPFlowJets`
Configuration file:	`JES_MC16Recommendation_Consolidated_PFlow_April2019_Rel21`
Calibration sequence:	`JetArea_Residual_EtaJES_GSC_Smear[_Insitu]`
Calibration area version:	00-04-82

Table 6: PFlow jet calibration recommendations. The `_Insitu` calibration is applied on data while the jet energy resolution `_Smear` is applied on MC.

The jet selection used for this analysis is:

- $p_{\mathrm{T}} > 25$ GeV.
- anti-kt R = 0.4.

- $|\eta| < 2.5$ (for central jets).
- $|y| < 4.4$
- Jet-Vertex Tagger (JVT) WP:Tight
- Jet cleaning WP:LooseBad

The flavour tagging algorithm used to determine the flavour of the jet is a high-level algorithm based on a deep neural network that uses the output of "recurrent neural network impact parameter" (RNNIP) as input. DL1r outputs three different probabilities (p_{b}, p_{c} and p_{u}) that are combined to define a final discriminant. DL1r algorithm has been re-optimized in 2019 in order to maximize the performance on the jet collections recommended for use in ATLAS, PFlow jets, and VR jets and to extend the algorithm performance to very high jet p_{T} [32],[33]. The b-tagging working point with a 77% efficiency is chosen, such efficiency is measured from $t\bar{t}$ MC samples and dedicated $t\bar{t}$ data. The associated SFs are taken into account.

3.3 Leptons

The selection of a lepton uses the official working point of identification and isolation.

- Electrons: Electrons are reconstructed by matching the energy deposits from the EM calorimeter to the track in the inner detector. It requires $p_{\mathrm{T}} > 10$ GeV, $|\eta| < 1.37$ or $1.52 < |\eta| < 2.37$, Medium LH ID, $|d_0\text{significance}| < 5$, $|\Delta z_0 \sin\theta| < 0.5$ mm, and the Isolation working point is FCLoose.
- Muons: Muons are reconstructed by using the information of the Muon spectrometer and the Inner detector. The candidates should pass the medium working point, and should $p_{\mathrm{T}} > 10$ GeV, $|\eta| < 2.7$, $|d_0\text{significance}| < 3$, $|\Delta z_0 \sin\theta| < 0.5$ mm and PflowLoose_FixedRad isolation criteria.

3.4 Missing transverse energy

The $E_{\mathrm{T}}^{\mathrm{miss}}$ involves all the reconstructed and calibrated objects described above. Compared to the general definition, τ leptons are treated as normal hadronic jets here and do not change the performances [29]. The Track-based Soft Term (TST) is the chosen approach in all HGam analyses to compute the $E_{\mathrm{T}}^{\mathrm{miss}}$ soft term and is therefore used here. Compared to the usual computation, this term is derived with respect to the chosen di-photon vertex instead of the usual hardest vertex.

3.5 Overlap removal

Since objects are reconstructed with different algorithms in parallel, one needs to implement a set of rules to remove objects close to each other to avoid double counting. This overlap removal is done just after full object definitions and two loose photons so that in the samples of reverse ID or reverse isolation, overlap removal is also implemented. The rules are defined below. More details can be found in Ref [34].

- The two leading photons are always kept.
- Electrons with $\Delta R(e, \gamma) < 0.4$ are removed.
- Jets with $\Delta R(jet, \gamma) < 0.4$ are removed.

- Jets with $\Delta R(jet, e) < 0.2$ are removed.
- Muons with $\Delta R(\mu, \gamma) < 0.4$ or $\Delta R(\mu, jet) < 0.4$ are removed
- Electrons with $\Delta R(e, jet) < 0.4$ are removed.

4 Event selection

4.1 Selections

The event selection procedure identifies two photons and then applies requirements on the existence of one or two leptons in order to increase the signal purity and background rejection. The event selection for the analysis starts with the full di-photon selection from the $h \to \gamma\gamma$ analysis in Run-2 to select two isolated photons. The selections listed below are similar to the other HGamma analyses, except for optimizations of the b-veto working point and the BDT selection that will be described in Sec 5.

- **Trigger**: di-photon trigger with two reconstructed photons with E_{T} larger than 35 and 25 GeV passing *loose* (2015/2016) and *medium* (2017/2018) requirements based on the energy leakage in the hadronic compartment and on the shower shape in the second layer of the electromagnetic calorimeter are used for the analysis.
  - HLT_g35_loose_g25_loose (2015/2016)
  - HLT_g35_loose_g25_medium_L12EM20VH (2017/2018)

  For the trigger efficiency, one simple test is done removing the trigger requirement, and the yield deviation after the selections shows the efficiency is higher than 99.999%.
- **Good Run List and Detector Quality**: For real data, events must belong to the luminosity blocks specified in the Good Run Lists as Sec 2 shows. Events with data integrity errors in the calorimeters and incomplete events where some detector information is missing are rejected, as well as events which are corrupted due to power supply trips in the tile calorimeter.
- **Primary Vertex**: The primary vertex is selected using the neural network (NN) algorithm from HGam group. The photons' four momenta, JVT and track isolation are corrected with respect to this origin, and the mass of the di-photon system is accordingly recalculated. In the Appendix G the comparison between NN method and the default hardest vertex is shown.
- **2 tight isolated photons**: At least two tight isolated photons with $E_{\mathrm{T}} > 35\,\mathrm{GeV}$ for the leading photon and $E_{\mathrm{T}} > 25\,\mathrm{GeV}$ for the subleading photon. A further p_{T} selection recommended by the $H\to\gamma\gamma$ analysis [29] is applied to photon candidates with $p_{\mathrm{T}}/m_{\gamma\gamma} > 0.35$ (0.25) for the leading (subleading) photon. Furthermore, the mass range of diphoton mass is required to be between $[105, 160]$ GeV.
- **Number of leptons**: Exactly one lepton (muon or electron) for $\gamma\gamma + 1\ell$ analysis and at least two leptons with opposite charges for $\gamma\gamma + 2\ell$ analysis.
- **Number of jets**: At least two central jets for 1 lepton, and ZZ2l analysis.
- **b-veto**: In order to suppress backgrounds with top quarks and keep orthogonal with b-jet related searches, the event is rejected if there is any b-jet. The b-tagger is DL1r with a b-tagging efficiency of 77%.
- **Tight mass window**: The tight mass window ($120\,\mathrm{GeV} < m_{\gamma\gamma} < 130\,\mathrm{GeV}$) is used to define the final signal region which is blinded till the background estimation is consolidated. In the final fit on the background shape of $m_{\gamma\gamma}$ in Section 6, the events in the whole region, $[105, 160]$ GeV are used.
- **Classification of events**: Events are classified into 4 different categories:

- **WW1l**: An event with only one lepton and two central jets, also noted as 1lepton category.
- **WW1e1m**: An event having one muon and one electron with opposite charges.
- **WW2l**: An event with two same-flavor leptons but failing $|m_{\ell\ell} - m_Z| < 10$ GeV.
- **ZZ2l**: An event with two same-flavor leptons where $|m_{\ell\ell} - m_Z| < 10$ GeV and two central jets;

4.2 Selection efficiencies

The efficiencies of signal event selection are visualized in Figure 3 - 5. The exact numbers are listed in Appendix C.

These efficiencies are derived from signals of the simulated samples. After the selection of the two photons, the signal efficiencies range from 30.4% to 52.1%, while after the additional selection on the jets and the leptons , the signal efficiencies range from 10% up to 30%, for (X, S) mass grid from (300, 170) to (1000, 500) GeV. These selection efficiencies are consistent with previous studies for $WW^*\gamma\gamma$ analysis [35, 36].

Even though in the sample generation level, 3 different samples are produced, $\gamma\gamma + WW(\ell\nu q\bar{q}')$, $\gamma\gamma + WW(\ell\nu\ell\nu)$ and $\gamma\gamma + ZZ(\ell\ell q\bar{q}/\ell\ell\nu\nu)$, but these samples are mixed together with the corrsponding weights to check the cutflow. With such mixing, the migration effect is automatically considered. Including those migrated events, 1lepton category will be enhanced from 15% to 30% for different mass points, which is not negligible.

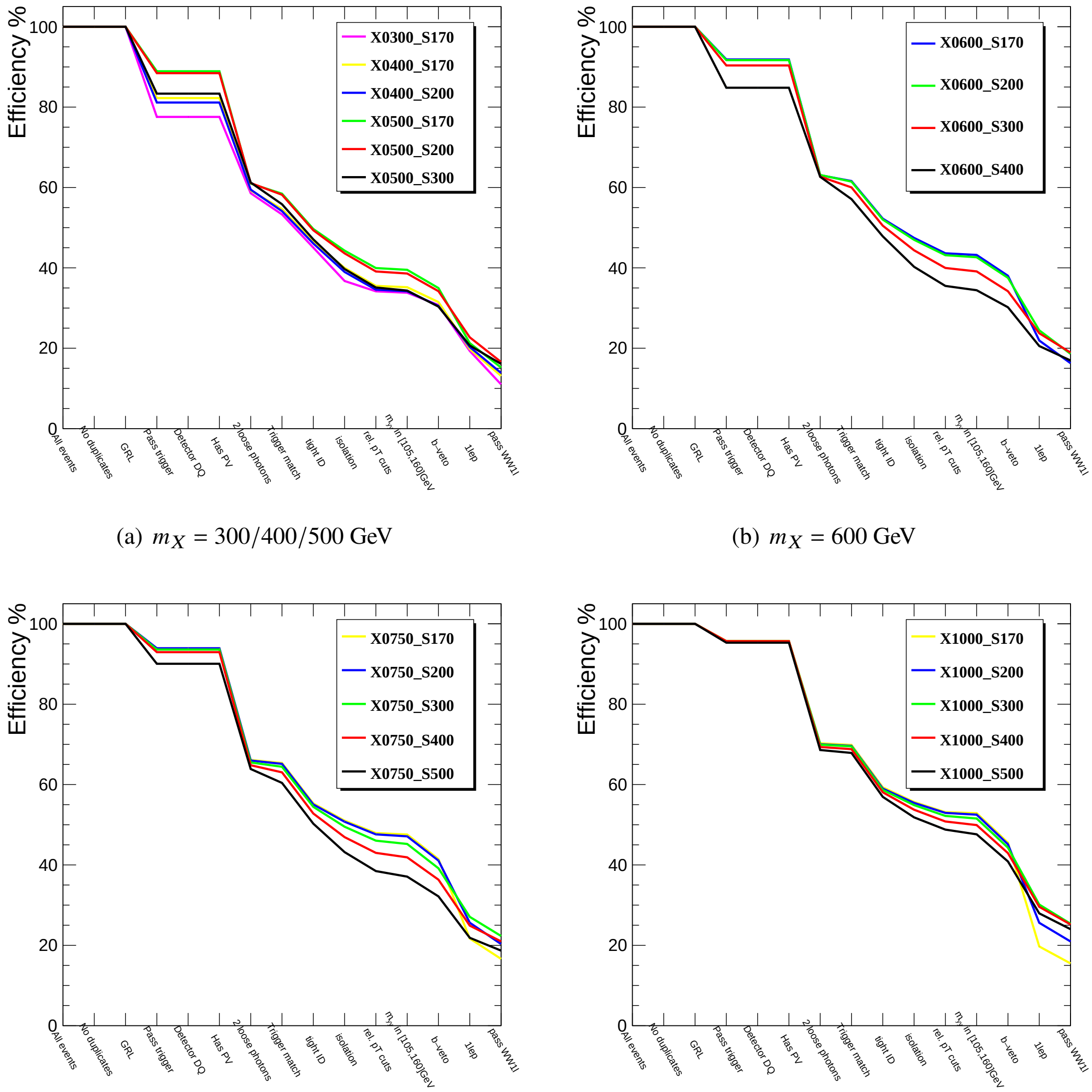
$\gamma\gamma + WW(\ell\nu\ell\nu)$ have 2 leptons in the truth level but in the reconstruction, the migration rate into 1lepton category is high, about one third for $\gamma\gamma + WW(\ell\nu\ell\nu)$ events. In HGam studies, when requiring 2 tight photons, the reconstruction efficiency for 1 real lepton is about 60%. So this one third ratio is in the expectation.

So, in Figures 4 and 5, the contaminations for falling to other categories are specified in the plot and the tables.

Same selections are applied on SM single Higgs and di-Higgs background samples. Table 7 lists the selection efficiency for each process up to the selection of 2 tight photons, and the efficiency is around 35.0%, which agrees with other diphoton analyses. The event yield after the photon selections are listed in the Table 8 and Table 9 in Section 6.

	ggh	VBF	W^+h	W^-h	$qqZh$	$ggZh$	$t\bar{t}h$	di-Higgs
All Events	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Pass Trigger	59.9	61.0	55.6	60.4	57.1	67.7	73.5	78.7
2 loose photons	49.8	50.8	43.2	48.1	46.0	56.2	59.1	60.5
Trig Match	49.7	50.7	43.1	48.0	45.9	56.0	58.6	52.8
Tight ID	43.2	43.8	37.0	41.1	39.4	48.1	48.6	48.5
Isolation	38.6	39.6	32.6	36.2	34.7	43.2	40.3	42.3
Rel.Pt cuts	35.8	36.0	29.9	33.1	31.7	39.4	36.7	38.7
$105 < m_{\gamma\gamma} < 160$ GeV	35.8	36.0	29.7	33.0	31.6	39.2	36.3	38.2

Table 7: Selection efficiencies in percent for SM single Higgs and double Higgs processes up to the selection of 2 tight photons.



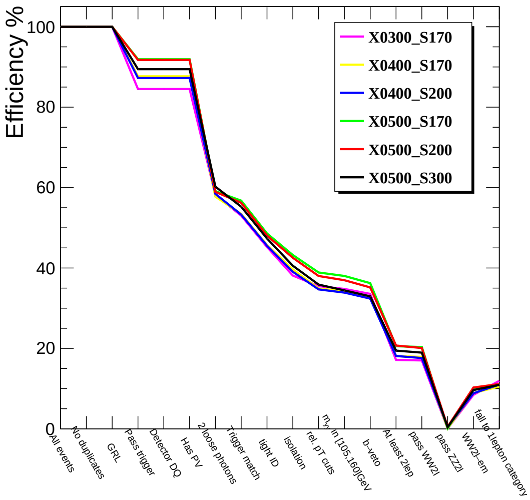
(a) m_X = 300/400/500 GeV

(b) m_X = 600 GeV

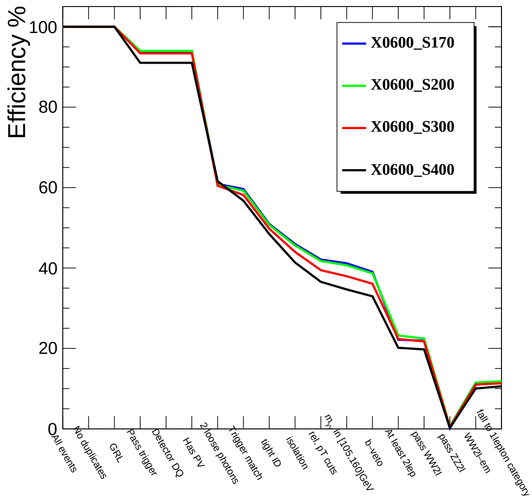
(c) m_X = 750 GeV

(d) m_X = 1000 GeV

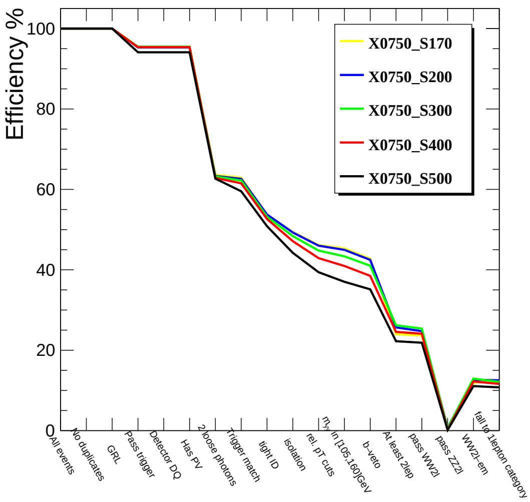
Figure 3: Selection efficiency for $\gamma\gamma + WW(\ell\nu q\bar{q}')$ samples. All 20 mass points are separated to 4 groups based on m_X.



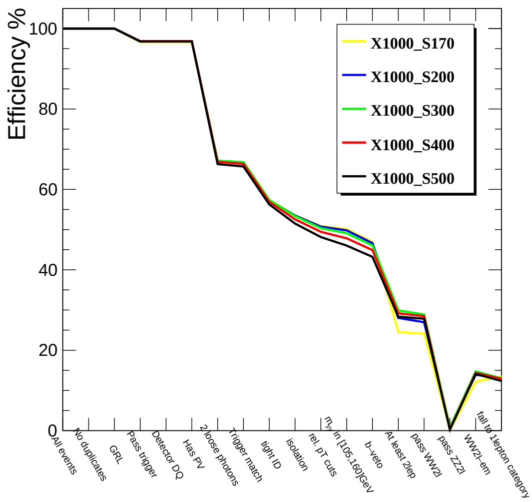
(a) m_X = 300/400/500 GeV



(b) m_X = 600 GeV

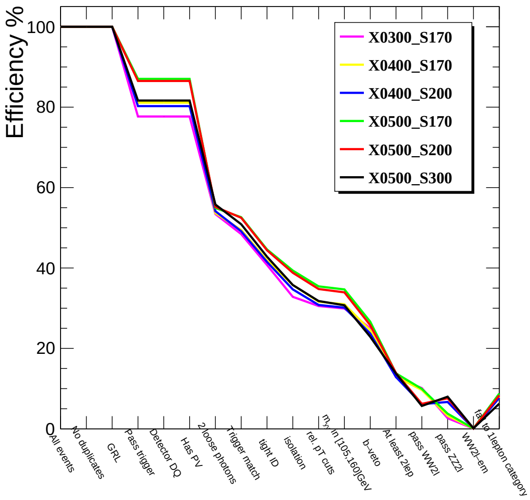


(c) m_X = 750 GeV

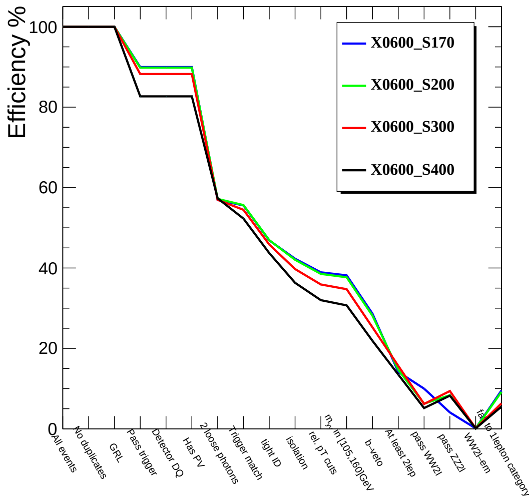


(d) m_X = 1000 GeV

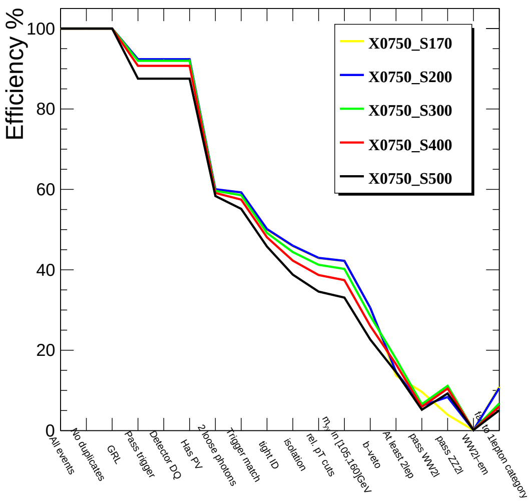
Figure 4: Selection efficiency for $\gamma\gamma + WW(\ell\nu\ell\nu)$ samples. All 20 mass points are separated to 4 groups based on m_X.



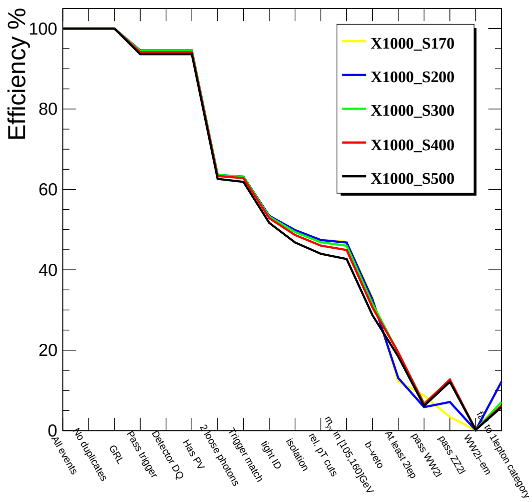
(a) m_X = 300/400/500 GeV



(b) m_X = 600 GeV



(c) m_X = 750 GeV



(d) m_X = 1000 GeV

Figure 5: Selection efficiency for $\gamma\gamma + ZZ(\ell\ell q\bar{q}/\ell\ell\nu\nu)$ samples. All 20 mass points are separated to 4 groups based on m_X.

m_X	300	400	500	600	750	1000	400	500	600	750	1000
m_S	170						200				
1 lepton region											
Signals	10.9469	12.0659	13.9792	18.9271	18.0051	17.9718	8.5875	14.3348	15.7590	11.0111	21.7103
ggF	0.0191						0.0134				
VBF	0.0048						0.0046				
VH	0.6836						0.6257				
ttH	0.4410						0.3915				
tH+bbH	0.0770						0.0703				
di-Higgs	0.0948						0.0832				
2 lepton region											
Signals	2.7433	8.9129	4.0167	17.2679	5.2911	4.9245	5.9601	5.3006	5.1519	5.8852	6.2947
ggF	0.0000						0.0000				
VBF	0.0001						0.0000				
VH	0.1493						1.1642				
ttH	0.0822						0.1355				
tH+bbH	0.0087						0.0125				
di-Higgs	0.0427						0.0460				

Table 8: Signal, SM Higgs as well as SM di-Higgs yields after the event selection. The signal yields correspond to $m_S <= 200 GeV$ assuming the cross section $\sigma(gg \to X) \times \mathrm{BR}(X \to SH)$ of 1 pb, with the integrated luminosity of 139 fb^{-1}. For the SM di-Higgs, $\sigma(gg \to hh) = 31.05$ fb.

The contributions from signals, SM Higgs and di-Higgs are estimated with MC statistics. The expected signal yields after the event selection of resonant X scalar as a function of (m_X, m_S) with the assumption of $\sigma(gg \to X) \times \mathrm{BR}(X \to SH) = 1$ pb are listed in Table 8 and Table 9. Corresponding SM Higgs yields as well as SM di-Higgs contributions with $\sigma(gg \to hh) = 31.05$ fb are listed as well.

For all the SM single Higgs processes, contributions from ggF and VBF processes are small due to the requirements on one or two leptons in the selections. Usually VH channels have the largest yields left.

m_X	500	600	750	1000	600	750	1000	750	1000
m_S	300				400			500	
1 lepton region									
Signals	0.5737	8.0312	32.5921	23.8881	0.5564	10.8044	22.0978	4.5707	23.4183
ggF	0.0049				0.0042				
VBF	0.0015				0.0009				
VH	0.4077				0.3194				
ttH	0.2597				0.1926				
tH+bbH	0.0400				0.0296				
di-Higgs	0.0445				0.0314				
WW 2 lepton region									
Signals	4.0596	7.7640	6.2101	7.5156	2.3114	4.6149	6.1617	3.6571	5.8629
ggF	0.0000				0.0000				
VBF	0.0000				0.0000				
VH	0.6371				0.1100				
ttH	0.1241				0.0903				
tH+bbH	0.0123				0.0090				
di-Higgs	0.0301				0.0152				

Table 9: Signal, SM Higgs as well as SM di-Higgs yields. The Signal yields correspond to $m_S >= 300GeV$ assuming the cross section $\sigma(gg \to X) \times \mathrm{BR}(X \to SH)$ of 1 pb, with the integrated luminosity of 139 fb^{-1}. For the SM di-Higgs, $\sigma(gg \to hh) = 31.05$ fb.

5 BDT optimization for the analysis

Various optimization strategies are used for different channels, as shown in Table 10. For WW1l and WW2l channels, a Boosted Decision Tree (BDT) is employed to reach better performances; While for WW1e1m and ZZ2l channels, events are directly counted after straightforward selections due to the low statistics and high signal-over-background ratios.

For the two channels using BDT, multiple discriminating variables separating signal and background are chosen as inputs for the BDT training. The BDT output, which reflects an optimal combination of these input variables to separate signal and background, is used to define the signal regions to have a good significance. Such signal regions are defined according to the expected significance as described in Eq.(4) and the procedure is mentioned in Section 5.2. When doing the optimization, the cross sections for signals is assumed to be 1 pb as mentioned in the Section 1. To blind the signal region, which is defined as the mass window $m_{\gamma\gamma} \in (120, 130)$GeV, the sideband data is used to estimate the yields for the continuum background in the signal region.

Channel	Definition	Optimization strategy
WW1l	1lepton, 2 central jets	BDT
WW2l	2lepton, same flavor, $\lvert m_{\ell\ell} - m_Z \rvert$ >10 GeV	BDT
WW1e1m	1 electron 1 muon	Cut based
ZZ2l	2lepton, 2 central jets, same flavor, $\lvert m_{\ell\ell} - m_Z \rvert$ <10 GeV	Cut based

Table 10: Definition of the four channels and the corresponding optimization strategies.

5.1 BDT training and testing

The Toolkit for Multi-Variant Data Analysis (`TMVA`) package [37] is used to perform the BDT training. The training is applied on the partial amount of the events (called training sample) and the results are tested and evaluated with the remaining events (called test sample). In order to avoid possible biases, moreover, the Cross Validation method[38] is applied with 4 folds to improve the performance. This feature can be seen as all the events are automatically both trained and tested.

As the events are split by the event ID, it is possible to trace the corresponding fold for each training event.

5.1.1 Input variables

The major production and decay process for SH signal in this study is $X \to SH, H \to \gamma\gamma$ and $S \to W^+W^-$. In the WW1l channel, one W boson decays leptonically ($W \to l\nu$) and the other goes with hadronic decay ($W \to q\bar{q}$). While in the WW2l channel, both W bosons decay leptonically. Several kinematic variables regarding different objects can be constructed as listed respectively in Table 11 and Table 12 for WW1l and WW2l channels. In the tables, the separation value is demonstrated between the MC backgrounds and the X1000S400 sample.

Several assumptions are made to construct these input variables. Firstly, in 1lepton category, all $E_{\mathrm{T}}^{\mathrm{miss}}$ items are regarded as one neutrino. And, considering this neutrino from the $W \to \ell\nu$ decay, it is possible to use W invariant mass as one contriant to obtain the p_z information of this neutrino. With this implementation, it is possible to construct the variable $\Delta R(jj, l\nu)$. For 2lepton category, there are two neutrinos, but still all the $E_{\mathrm{T}}^{\mathrm{miss}}$ items are regarded as one neutrino, and its p_z can be obtained from the W mass constrain with the leading lepton.

Variable	Definition	Separation
Regarding particle X		
$\Delta R(\gamma\gamma, l\nu jj)$	Angular difference between diphoton system (H) and $l\nu jj$ system (S)	0.048
Regarding particle S		
$\Delta R(jj, l\nu)$	Angular difference between dijet system (W_{had}) and $l\nu$ system (W_{lep})	0.089
$p_T^{l\nu jj}$	Transverse momentum of $l\nu jj$ system (S)	0.373
Regarding SM Higgs boson		
$p_T^{\gamma\gamma}$	Transverse momentum of diphoton system (H)	0.484
$\Delta\Phi(\gamma\gamma, l)$	Polar angle difference between di-photon system (H) and signal lepton	0.026
Regarding single W boson from S		
$\Delta R(j, j)$	Angular difference between two jets (W_{had})	0.171
p_T^{jj}	Transverse momentum of di-jet system (W_{had})	0.181
$m_{jj}(m_W)$	Invariant mass of di-jet system whose mass is closest to $m_W(W_{had})$	0.119
$\Delta R(l, E_{\mathrm{T}}^{\mathrm{miss}})$	Angular difference between lepton and $E_{\mathrm{T}}^{\mathrm{miss}}$ (W_{lep})	0.108
$E_{\mathrm{T}}^{\mathrm{miss}}$	Missing transverse momentum	0.248
p_T^l	Transverse momentum of the single lepton	0.203
$m_T(l\nu)$	Transverse mass of $l + E_{\mathrm{T}}^{\mathrm{miss}}$ system (W_{lep})	0.044

Table 11: Variables used for BDT training in WW1l channel and their separation powers.

Variable	Definition	Separation
Regarding particle X		
$\Delta R(\gamma\gamma, ll + E_{\mathrm{T}}^{\mathrm{miss}})$	Angular difference between diphoton system (H) and $ll + E_{\mathrm{T}}^{\mathrm{miss}}$ system (S)	0.031
Regarding particle S		
$\Delta R(l_1 + E_{\mathrm{T}}^{\mathrm{miss}}, l_2)$	Angular difference between leading lepton + $E_{\mathrm{T}}^{\mathrm{miss}}$ (W_{l1}) and l_2	0.038
Regarding SM Higgs boson		
$p_T^{\gamma\gamma}$	Transverse momentum of diphoton system (H)	0.621
$\Delta\Phi(\gamma\gamma, l_1)$	Polar angle difference between di-photon system (H) and the leading lepton	0.079
Regarding single W boson from S		
$p_T^{l_1}$	Transverse momentum of the leading lepton	0.415
$E_{\mathrm{T}}^{\mathrm{miss}}$	Missing transverse momentum	0.638
$p_T^{l_1+E_{\mathrm{T}}^{\mathrm{miss}}}$	Transverse momentum of the leading lepton and $E_{\mathrm{T}}^{\mathrm{miss}}$ system	0.533
$m_T(l_1 + E_{\mathrm{T}}^{\mathrm{miss}})$	Transverse mass of leading lepton and $E_{\mathrm{T}}^{\mathrm{miss}}$ system	0.362
m_{ll}	Invariant mass of di-lepton system	0.358

Table 12: Variables used for BDT training in WW2l channel and their separation powers.

As there are 20 different mass points, it is complicated to train samples 20 times. And, if use only 1 BDT in the training, the phase space limited region is suffered since the sensitivity would be low. Finally, 4 BDTs are chosen to simplify the training processes in 20 mass hypotheses of X and S, the parameterized BDT method [39] is applied in this analysis.

This method targets the classification tasks with uncertain parameters in the training sample, e.g. the hypothesis of new particle mass m_X. These parameters are given to the classifier as the extension of input event-level features. The machine-learning algorithm is expected to learn from these parameters, thus the trained can have good performance by specifying the parameter in different cases. Considering the BDT may not have smooth interpolation for several unknown parameters in this method, m_X is treated as the only input parameter and signal processes are assigned to 4 groups with m_S: m_S = 170 GeV, m_S = 200 GeV, m_S = 300 GeV, and $m_S \geq$ 400 GeV. Which is, for the S170 BDT training, From X300 to X1000, 6 different signal samples are grouped together since they all have m_S = 170 GeV, but their input variable m_X hold their truth information. For background processes, in the BDT trainning, they are randomly assigned with the discrete m_X values with the same fractions as signal did. But, when applying the BDT back to the backgrounds, they will use the same m_X value as the signal.

Based on this method, it is possible to obtain in total 20 different signal and background BDT response distributions for different mass points with 4 BDT trains.

This procedure is performed in both WW1l and WW2l channels, so in total 8 different BDTs are trained. In each channel, the SH processes are treated as the signal during the BDT training, with the input variables described above and the truth m_X as a parameter.

The background training samples includes the continuum background processes, single Higgs processes and SM dihiggs processes. Real data is not used in the trainning due to the low statistics. Considering there is an obvious difference between continuum background and sideband data, the reweighting for MC continuum background samples has been applied for both shape and yield.

To ensure a better agreement between the data and MC, the continuum backgrounds (Sherpa $\gamma\gamma + jets$, $V(VV)\gamma\gamma$ and $t\bar{t}\gamma\gamma$) are reweighted to the sideband data as a function of the transverse mass of leptonic decayed W boson m_T^{W1}. Before this procedure, an individual normalization factor based on the data and MC ratio is implemented on the MC for each channel.

Here, the reweighting is applied on the W transverse mass. Firstly, the ratio for 3 different processes are fixed and then the deviations between MC and sideband data are considered bin by bin for each 10 GeV. The factor for these reweighting is in 10% level.

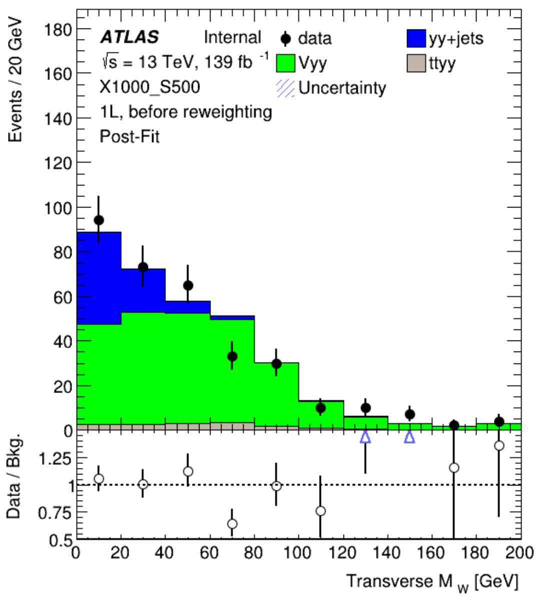
The distribution before and after reweighting is shown in Figure 6. Decent consistencies between sideband data and continuum MC for the disbributions of the discriminating variables can be observed with this method. Appendix E describes the method in detail and presents the comparisons of the other distributions for the kinematic variables before and after the reweighting.

Hyperparameters used in the training are summarized in Table 13.

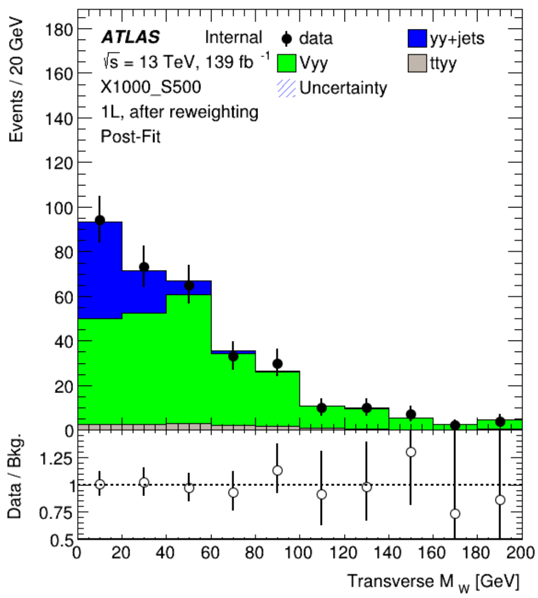
The distributions of input variables after the reweighting in WW1l channels for signals, backgrounds and sideband data, are shown in Figure 7 - 9 with corresponding signal mass $(m_X, m_S) = (1000, 500)$ GeV. To easily compare the shape, the yield of the signals are normalized to backgrounds, with the red dashed line. Reasonable consistencies between sideband data and MC for the discriminating variables are observed. The corresponding distributions of the input variables for WW2l channels are shown in Figure 10 - 11. For all those distrbutions, to be noted that, the ideas for those plots are demonstrating the kinematics for the distributions of signals and backgrounds. For signals and SM Higgs processes, most of them are in the region $m_{\gamma\gamma} \in (120, 130)$ GeV while for continuum MC and data only the sideband events are included for the plots.

Parameters	Value
`BoostType`	AdaBoost
`AdaBoostBeta`	0.5
`NTrees`	850
`MinNodeSize`	2.5%
`UseBaggedBoost`	True
`BaggedSampleFraction`	0.5
`SeparationType`	GiniIndex
`nCuts`	20
`MaxDepth`	3
`NegWeightTreatment`	Ignore
`UseCrossValidation`	True
`Nums of Folds`	4

Table 13: Summary of hyper-parameters used in BDT training.



(a)



(b)

Figure 6: m_T^{W1} distribution for $\gamma\gamma + jets$, $V\gamma\gamma$ and $t\bar{t}\gamma\gamma$ processes and the sideband data before(a) and after(b) background reweighting. The relative ratio between the 3 MC processes is fixed to the SM prediction.

557 The correlation matrix between input variables along with $m_{\gamma\gamma}$ are shown in Figure 12 and 13. Relevant
558 plots for the other mass points are displayed in Appendix F.

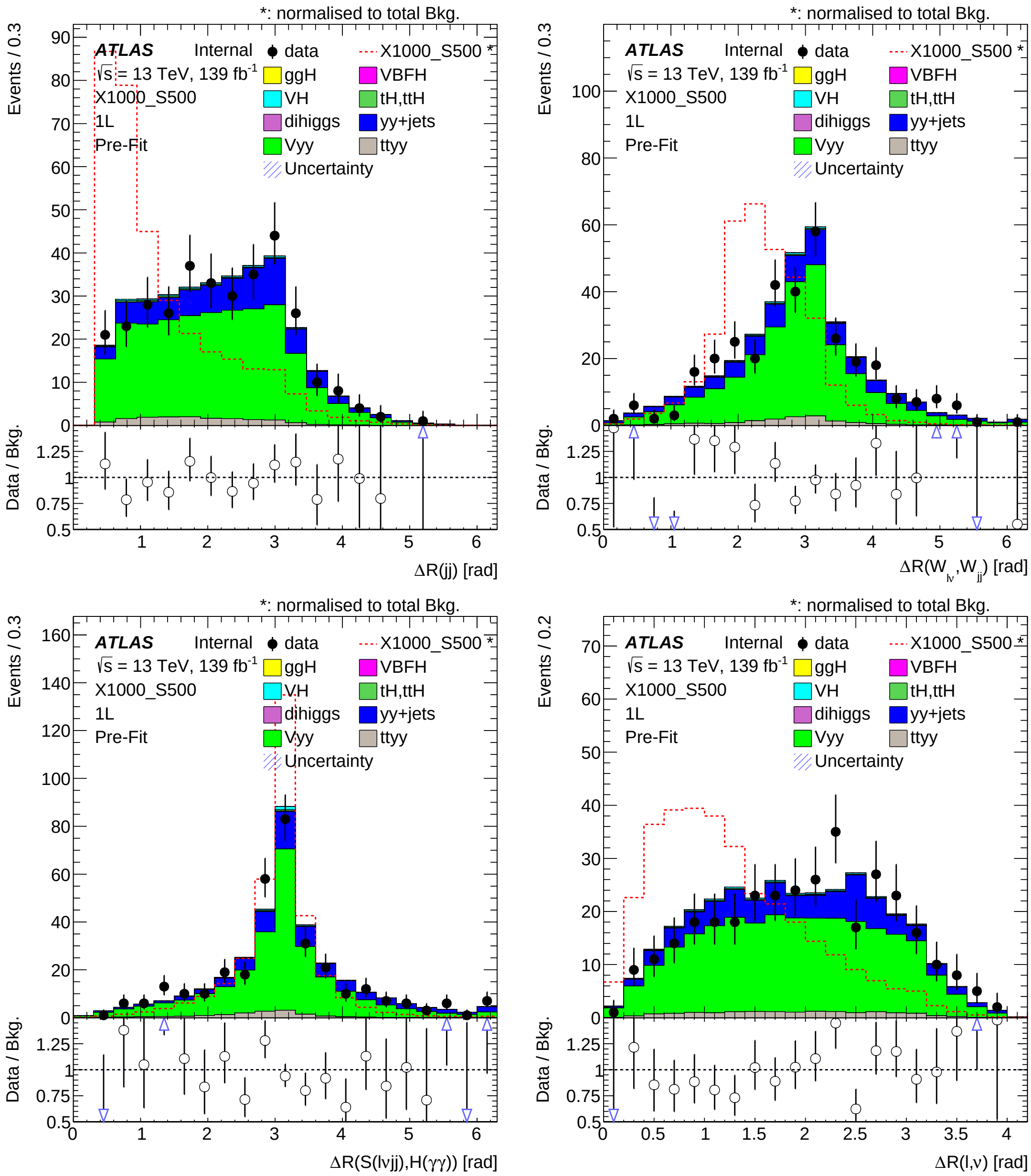


Figure 7: $\Delta R(j, j)$, $\Delta R(W1, W2)$), $\Delta R(S, H)$, $\Delta R(\ell, \nu)$) distributions for the WW1l channel. The events of the continuum MC and data are from the sideband region. For the single Higgs, dihiggs and signal samples, events are from the mass window between 120 and 130 GeV.

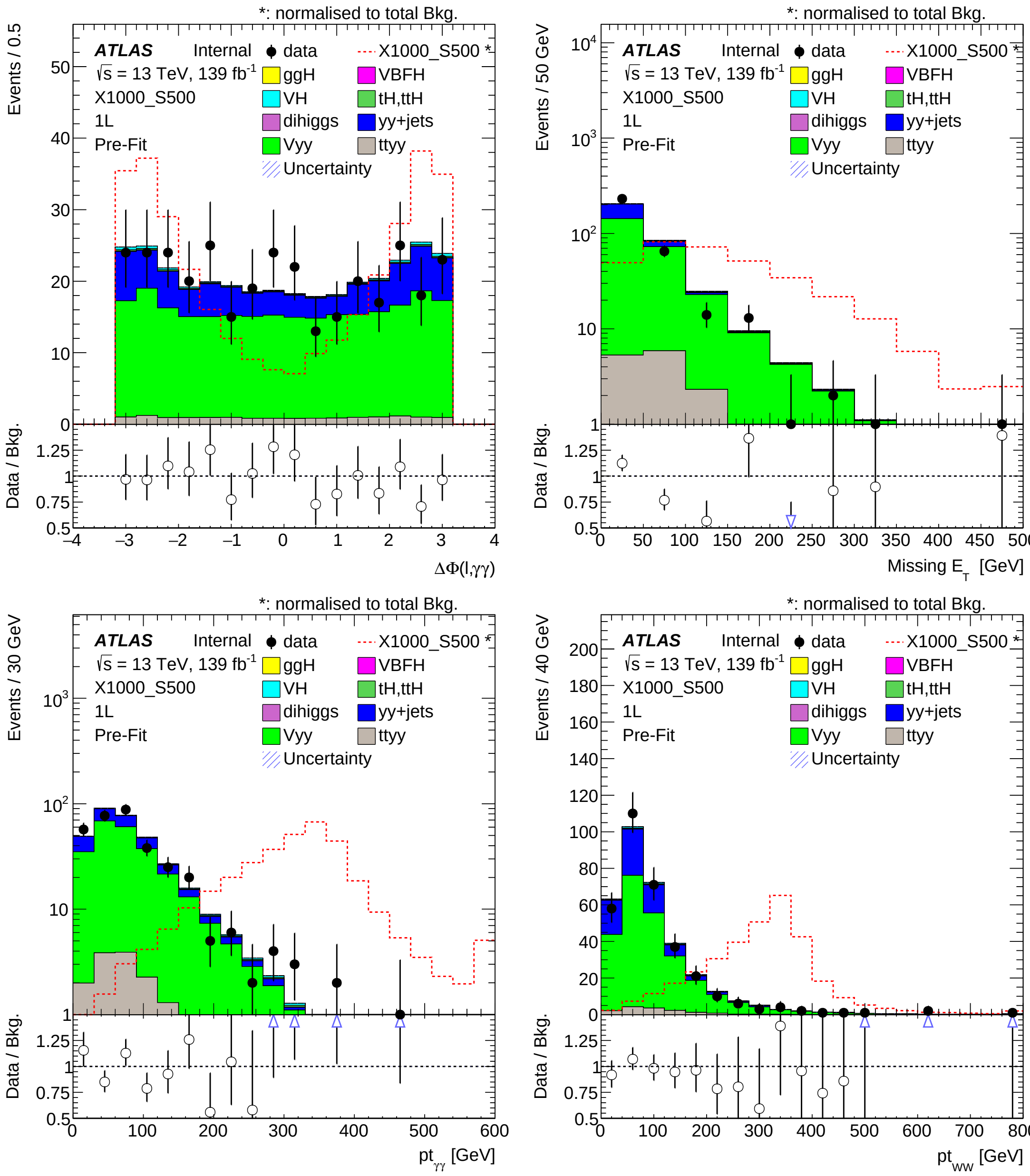


Figure 8: $\Delta\Phi(\gamma\gamma, \ell_1)$, $E_{\mathrm{T}}^{\mathrm{miss}}$, $p_{\mathrm{T}}(\gamma\gamma)$ and $p_{\mathrm{T}}(WW)$ distributions for the WW1l channel. The events of the continuum MC and data are from the sideband region. For the single Higgs, dihiggs and signal samples, events are from the mass window between 120 and 130 GeV.

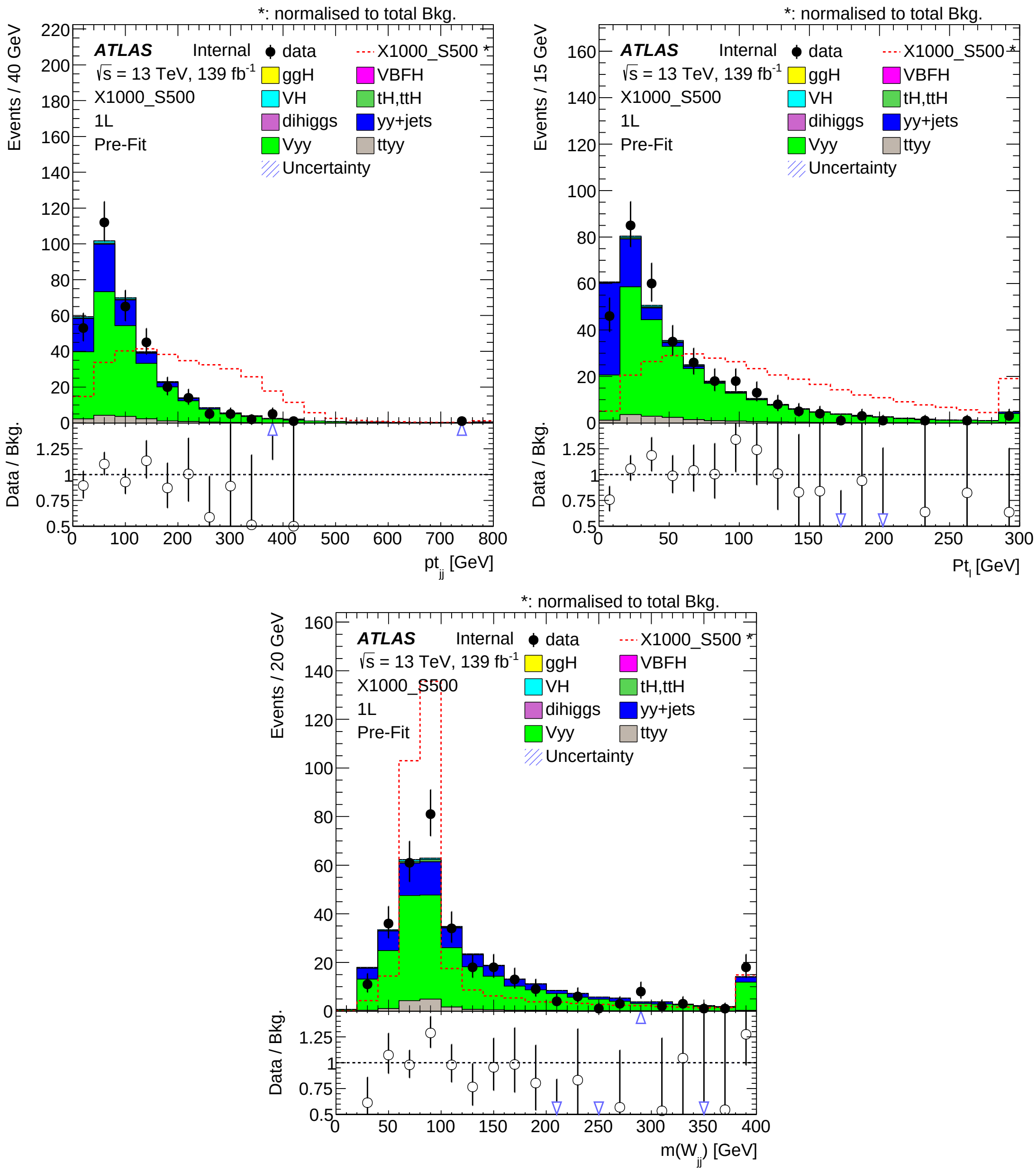


Figure 9: $p_{\mathrm{T}}(jj)$, $p_{\mathrm{T}}(\ell 1)$ and $M_{W(jj)}$ and distributions for the WW1l channel. The events of the continuum MC and data are from the sideband region. For the single Higgs, dihiggs and signal samples, events are from the mass window between 120 and 130 GeV.

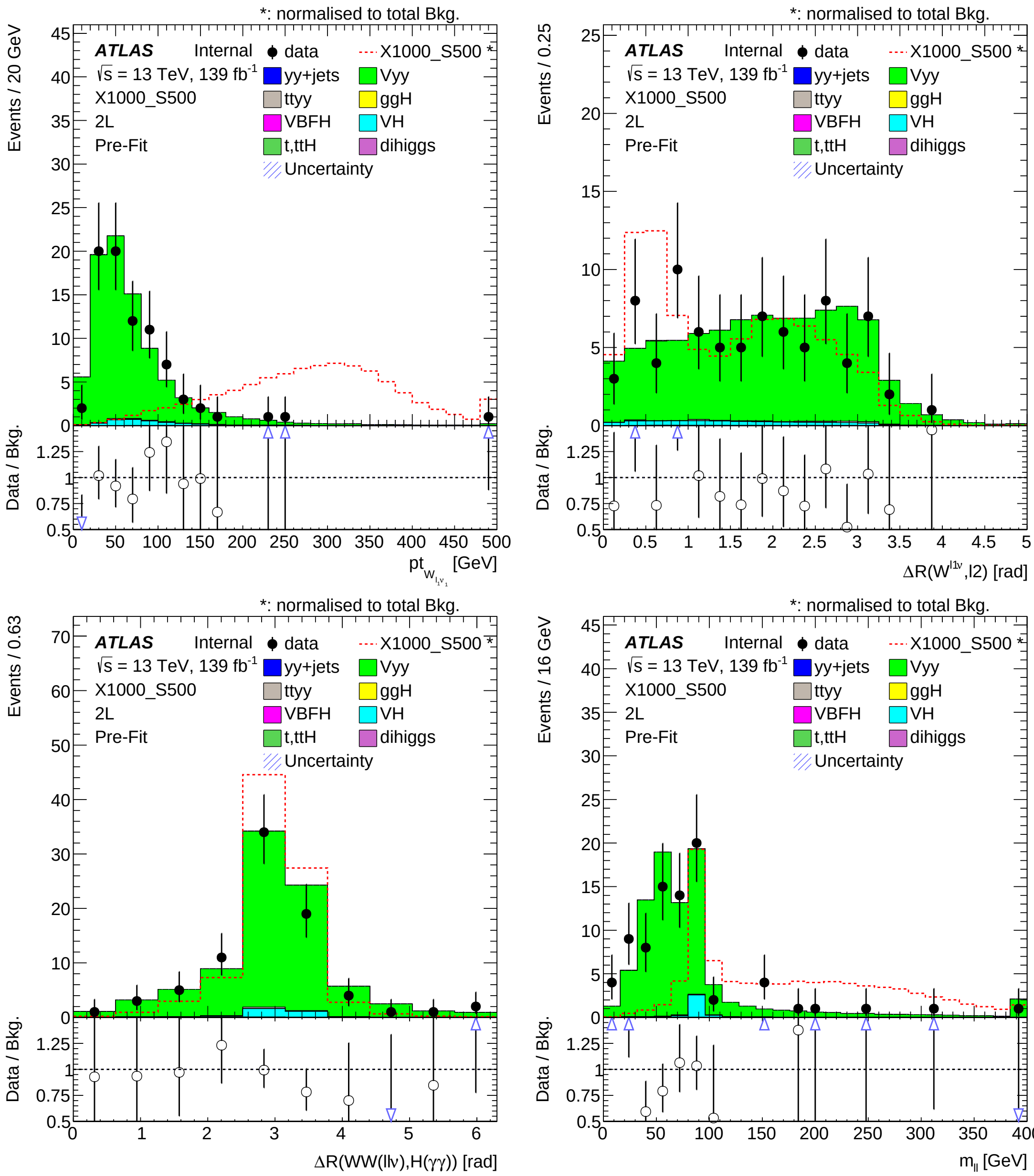


Figure 10: $p_{\mathrm{T}}(\ell\nu)$, $\Delta R(W1, W2)$), $\Delta R(S, H)$, m_{ll} distributions for the WW2l channel. The events of the continuum MC and data are from the sideband region. For the single Higgs, dihiggs and signal samples, events are from the mass window between 120 and 130 GeV.

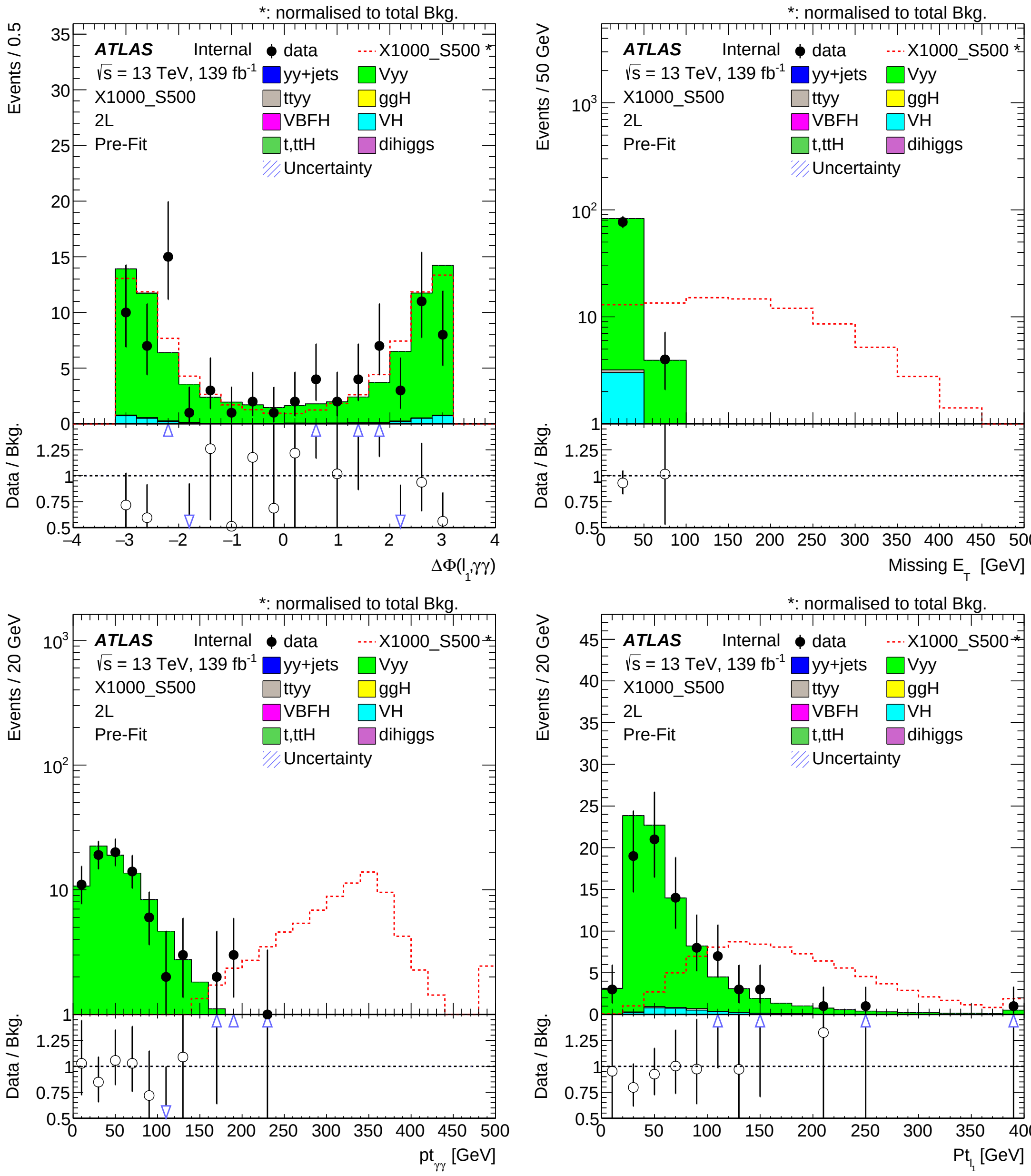
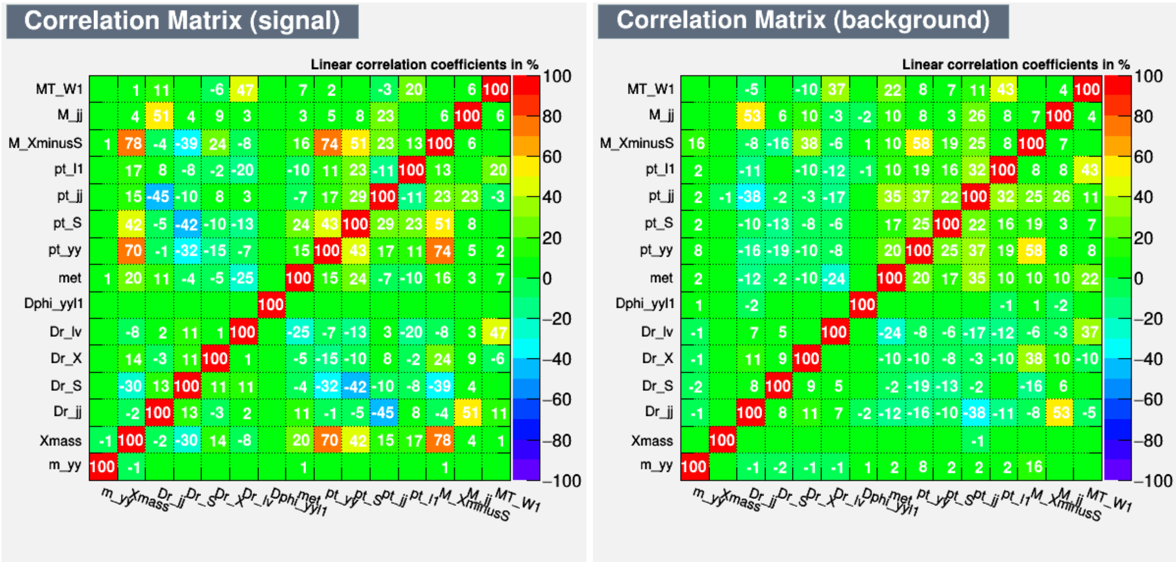
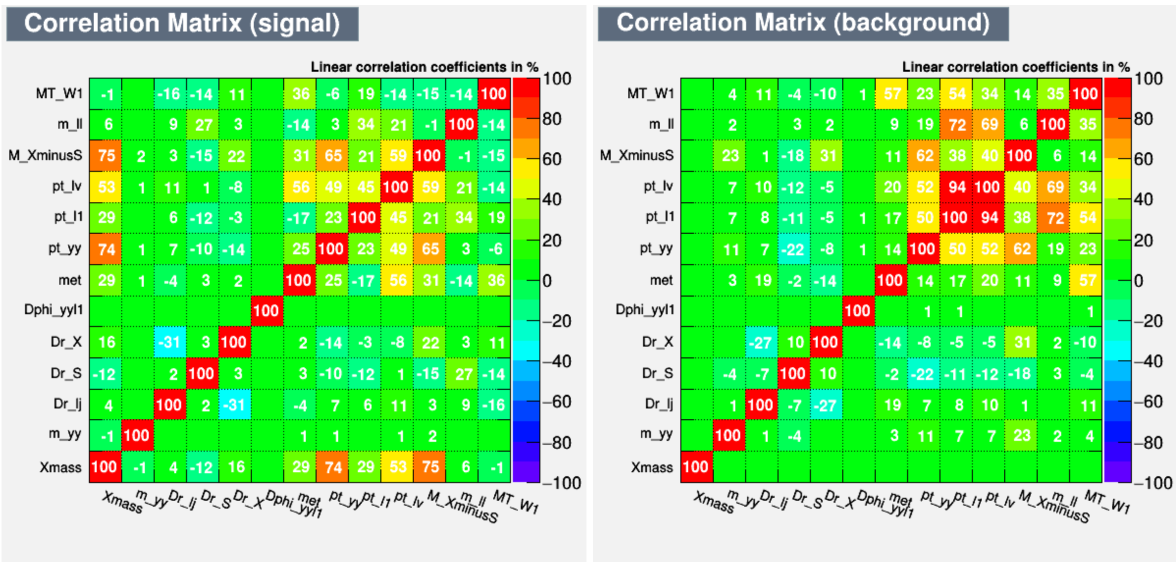


Figure 11: $\Delta\Phi(\gamma\gamma, \ell_1)$, $E_{\mathrm{T}}^{\mathrm{miss}}$, $p_{\mathrm{T}}(\gamma\gamma)$ and $p_{\mathrm{T}}(\ell 1)$ distributions for the WW2l channel. For the continuum MC and data events, are from the sideband region. For the single Higgs, dihiggs and signal samples, are in the mass window.



(a) (b)

Figure 12: Linear correlation matrix between input variables and $m_{\gamma\gamma}$ for signal $(m_X, m_S) = (1000, 500)$ GeV and background in the WW1l channel.



(a) (b)

Figure 13: Linear correlation matrix between input variables and $m_{\gamma\gamma}$ for signal $(m_X, m_S) = (1000, 500)$ GeV and background in the WW2l channel.

5.1.2 Training results

Figure 14 presents the BDT training result and the agreement between training and test samples in 4 folds in the WW1l channel. Figure 15 shows the signal efficiency vs background rejection curve.

Similar information for WW2l channels is shown in Figure 16 and 17. For (m_X, m_S) = (1000, 500) GeV training, the variable importance in these 2 BDTs, which is the rank of the TMVA training, is listed in Table 14 and 15.

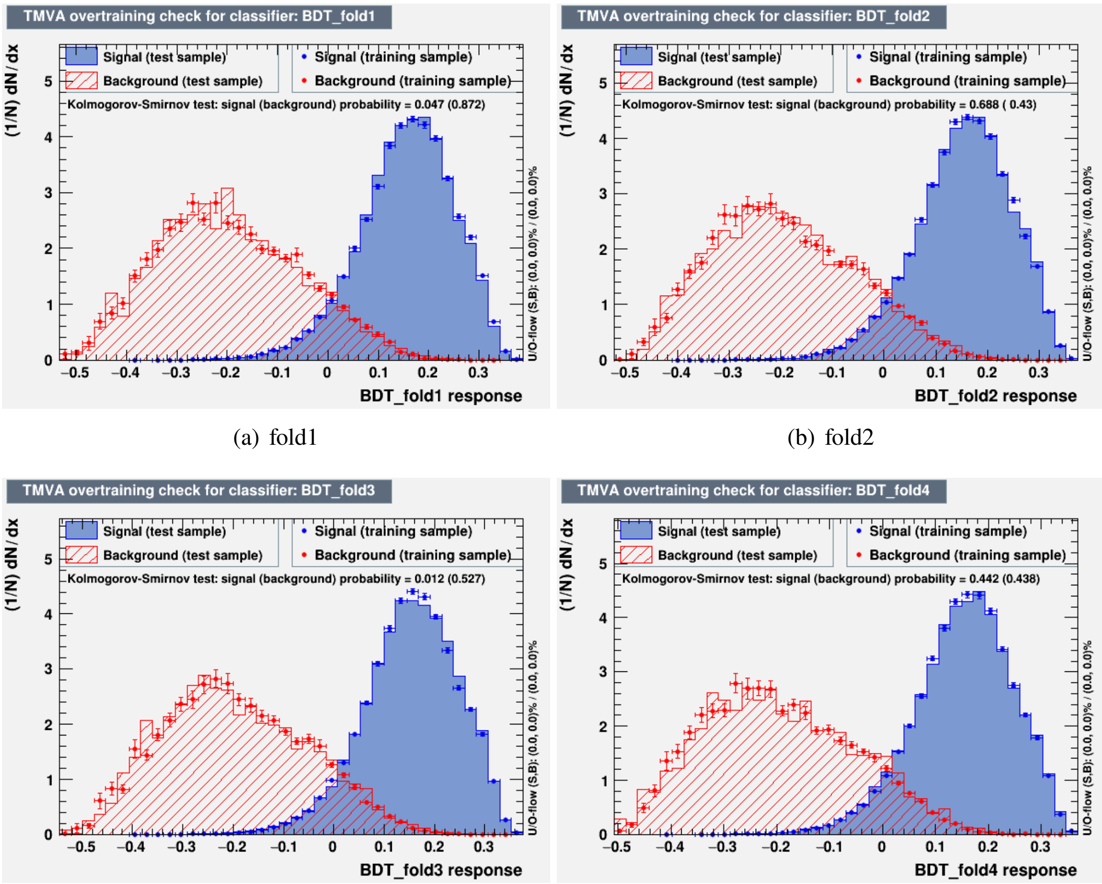
Figure 18 shows the trained BDT response for all processes and sideband data with m_S = 500 GeV hypothesis in two categories. The agreement of the BDT distribution between sideband data and MC is reasonable considering the complex phase space in this process, the known imperfect description in MC and the limited statistics in sideband data.

Ranking	Variable	Importance
1	$p_T^{\gamma\gamma}$	0.1017
2	p_T^{l}	0.0741
3	$E_{\mathrm{T}}^{\mathrm{miss}}$	0.0732
4	$\Delta R(j, j)$	0.0727
5	$\Delta\Phi(\gamma\gamma, l)$	0.0726
6	$\Delta R(l, E_{\mathrm{T}}^{\mathrm{miss}})$	0.0704
7	$\Delta R(jj, l\nu)$	0.0674
8	m_{jj}	0.0633
9	$DeltaR(\gamma\gamma, l\nu jj)$	0.0616
10	p_T^{jj}	0.0594
11	$m_T(l\nu)$	0.0529
12	$p_T^{l\nu jj}$	0.0421

Table 14: Variable importance in WW1l BDT for (m_X, m_S) = (1000, 500) GeV.

Ranking	Variable	Importance
1	$p_T^{\gamma\gamma}$	0.1077
2	$E_{\mathrm{T}}^{\mathrm{miss}}$	0.1038
3	$\Delta\Phi(\gamma\gamma, l_1)$	0.0939
4	$\Delta R(l1, l2)$	0.0885
5	m_{ll}	0.0874
6	$\Delta R(\gamma\gamma, ll + E_{\mathrm{T}}^{\mathrm{miss}})$	0.0782
7	$p_T^{l+E_{\mathrm{T}}^{\mathrm{miss}}}$	0.0739
8	$p_T^{l_1}$	0.0686
9	$m_T(l_1 + E_{\mathrm{T}}^{\mathrm{miss}})$	0.0609

Table 15: Variable importance in WW2l BDT for (m_X, m_S) = (1000, 500) GeV.



(a) fold1

(b) fold2

(c) fold3

(d) fold4

Figure 14: The overtraining plots with ks test values for 4 individual folds in 1 lepton $m_S \geq 400$ GeV group.

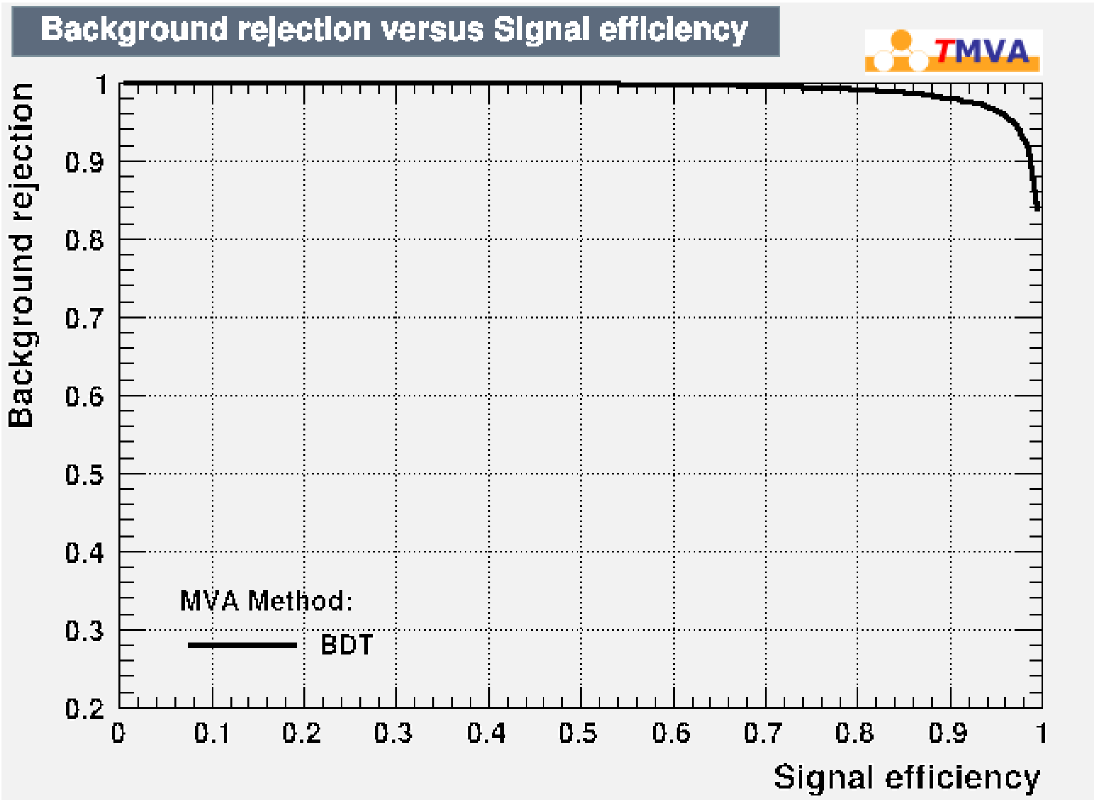
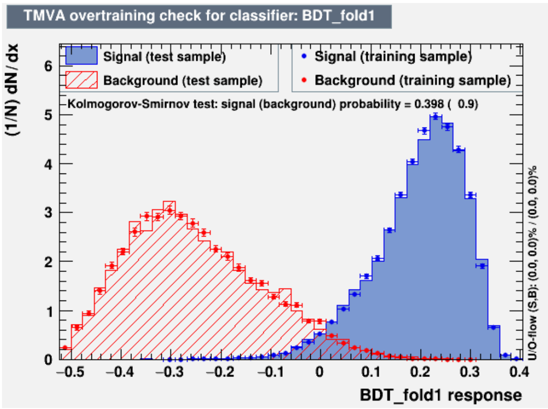
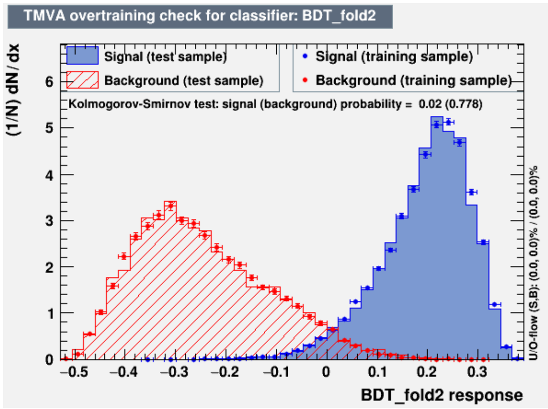


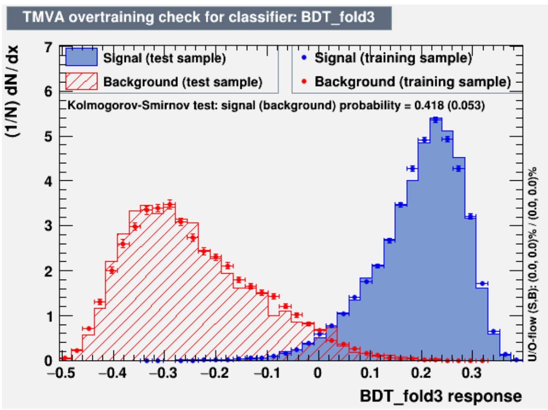
Figure 15: The signal efficiency vs background rejection curve (ROC curve) in WW1l channel for $m_S \geq 400$ GeV group.



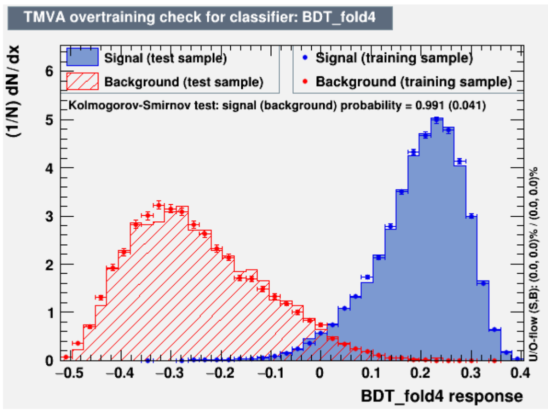
(a) fold1



(b) fold2



(c) fold3



(d) fold4

Figure 16: The overtraining plots with ks test values for 4 individual folds in 2 leptons $m_S \geq 400$ GeV group.

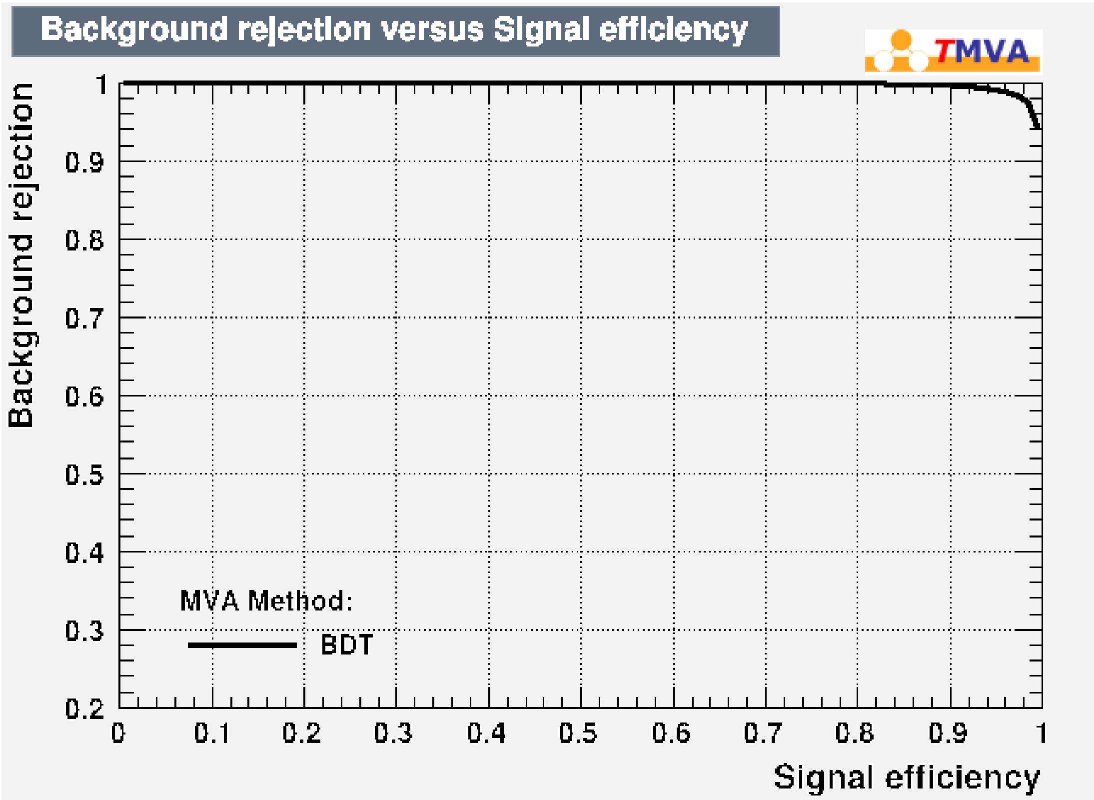


Figure 17: The signal efficiency vs background rejection curve (ROC curve) in WW2l channel for $m_S \geq 400$ GeV group.

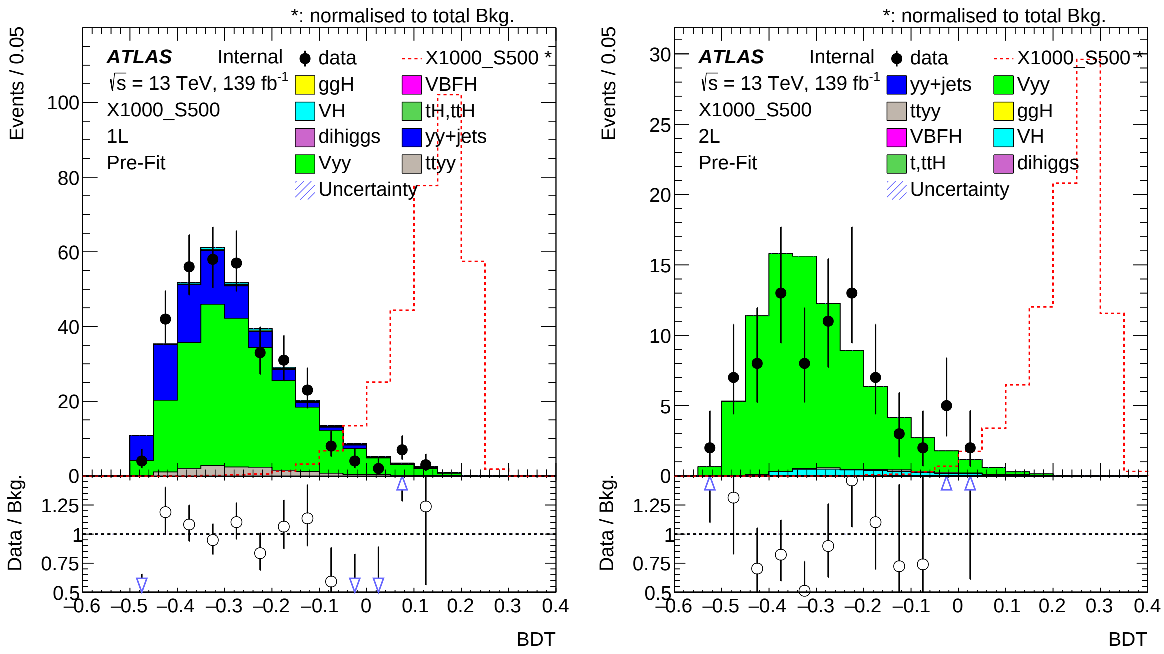


Figure 18: BDT outputs for WW1l and WW2l channels for signal, background MC and sideband data. Here, the signal and SM Higgs events are for the whole mass region $m_{\gamma\gamma} \in (105, 160)$GeV, and the continuum MC and data events, the blind region (120, 130)GeV is excluded.

1l	data	signal	2l	data	signal
X1000_S170	-6.08%	-0.24%	X1000_S170	-7.27%	-0.74%
X1000_S200	0.66%	-0.40%	X1000_S200	3.90%	0.87%
X1000_S300	-0.31%	0.55%	X1000_S300	1.82%	1.62%
X1000_S400	3.95%	0.83%	X1000_S400	4.88%	2.22%
X1000_S500	3.55%	-0.57%	X1000_S500	4.32%	2.59%
X0300_S170	2.22%	-0.97%	X0300_S170	-0.22%	-1.70%
X0400_S170	5.44%	1.03%	X0400_S170	4.50%	3.59%
X0400_S200	5.46%	0.08%	X0400_S200	4.17%	4.79%
X0500_S170	5.93%	0.20%	X0500_S170	0.59%	1.25%
X0500_S200	2.57%	-0.82%	X0500_S200	-3.54%	0.75%
X0500_S300	6.40%	1.91%	X0500_S300	4.17%	4.65%
X0600_S170	6.11%	0.41%	X0600_S170	4.02%	4.19%
X0600_S200	8.06%	0.16%	X0600_S200	0.91%	1.92%
X0600_S300	6.66%	-0.12%	X0600_S300	-2.77%	0.81%
X0600_S400	6.66%	-0.12%	X0600_S400	-2.77%	0.81%
X0750_S170	7.19%	1.52%	X0750_S170	4.29%	3.07%
X0750_S200	6.41%	1.93%	X0750_S200	3.94%	3.32%
X0750_S300	8.64%	2.00%	X0750_S300	0.72%	2.46%
X0750_S400	9.13%	-0.09%	X0750_S400	-3.01%	1.10%
X0750_S500	9.13%	-0.09%	X0750_S500	-3.01%	1.10%

Table 16: The correlation between BDT output value and diphoton mass, for signal and data sideband events.

Table 16 records the correlation between the BDT output value and diphoton mass variable, both our signal and sideband data event shows the correlations are small, for both 1 lepton and 2 lepton channel, which indicates the selection on the BDT value will not affect too much for the diphoton spectrum. In the following section, events will be divided into BDT tight and BDT loose region, in each tight or loose region, the diphoton spectrum can be described in one distribution.

5.2 Optimization of the analysis with BDT output

Figure 18 shows the BDT output distributions for the signal with mass (m_X, m_S)=(1000,500) GeV, background and sideband data. Based on this, events are categorized into two regions, namely loose and tight, by optimizing the signal significance[40]:

$$Z = \sqrt{2 \times [(S+B) \times (\ln \frac{S+B}{B}) - S]} \tag{4}$$

where S is the signal yield and B is the background yield in each category.

For WW1l and WW2l channels, BDT thresholds dividing events into loose and tight regions are obtained by scanning the BDT cut value to reach the highest Z in the combined BDT category significance. Specifically, the scan is done with a step size 0.005, and the maximization of $Z_{combined}$is chosen:

$$Z_{combined} = \sqrt{Z_{tight}^2 + Z_{loose}^2} \tag{5}$$

Following previous experience of $HH \to bb\gamma\gamma$ analysis[41], at least 2 sideband events from data are required to exist in the BDT tight region for the optimization of the event categoriazation.

To investigate the impact of this requirement on the sensitivity of the analyses, different thresholds for the events left in sideband data within the BDT tight regions have been tested. As shown in Table 17, the ratios of the expected limits for 1l channel with the mass of (m_X, m_S) = (1000, 400) GeV vary within 15% while scanning the thresholds from 1 to 10 events, indicating reasonable stabilities of the results. The consistencies have been further verified with the masses of (m_X, m_S) = (1000, 170), (750, 200) GeV as Table 18 and Table 19 show.

Data Entry Threshold	1	2	3	5	6	7	8	9	10
Limit w.r.t 1 data entry	100%	103.7%	105.2%	108.1%	109.8%	110.3%	111.5%	113.4%	114.4%
BDT Cut value	0.12	0.115	0.1	0.095	0.08	0.07	0.065	0.055	0.05

Table 17: Relative limit change for different sideband data entry threshold in (m_X, m_S) = (1000, 400) GeV WW1l channel, with respect to one data entry in sideband for the BDT tight region.

Data Entry Threshold	1	2	3	4	5	6	7	8	9	10
Limit w.r.t 1 data entry	100%	103.1%	105.3%	106.4%	110.2%	109.6%	112.8%	112.7%	114.0%	115.3%
BDT Cut value	0.165	0.15	0.135	0.1	0.09	0.06	0.05	0.045	0.025	0.15

Table 18: Relative limit change for different sideband data entry threshold in (m_X, m_S) = (1000, 170) GeV WW1l channel, with respect to one data entry in sideband for the BDT tight region.

In WW1L the channel for (m_X, m_S) = (750, 200) GeV, the data entry threshold of 1 event doesn't lead to the best sensitivity. A threshold of 4 events can cause 21% improvement on the upper limit. So for this mass point, data entry threshold for 4 events is chosen as an optimized one.

Data Entry Threshold	1	2	3	4	5	6	7	8	9	11
Limit w.r.t 1 data entry	100%	106.9%	99.5%	78.7%	80.4%	83.1%	90.5%	93.3%	92.4%	98.4%
BDT Cut value	0.195	0.185	0.165	0.105	0.09	0.08	0.075	0.07	0.065	0.055

Table 19: Relative limit change for different sideband data entry threshold in (m_X, m_S) = (750, 200) GeV WW1L channel, with respect to one data entry in sideband for the BDT tight region.

For some WW2l cases, different thresholds result in sharp cut value variations, leading to even larger differences on the subchannel limits. However, their impacts to combined upper limits are small.

In conclusion, based on the stability check of different threshold requirements for the events of the sideband data, at least 2 events in the sideband data are required in the tight BDT regions for both WW1l and WW2l. No more than 22% variations for the relative limits with alternative choices of the events are expected.

Furthermore, the extra toy MC studies are done in Appendix H, which shows the fit model is stable in these limited statistics situations.

The two categories are named with `WW1l_Tight` (`WW2l_Tight`) and `WW1l_Loose` (`WW2l_Loose`). The individual thresholds on BDT determined for signals with different masses are summarized in Table 20.

X Mass [GeV]	S Mass [GeV]	WW1L BDT Cut	WW2L BDT Cut
300	170	0.085	0.14
400	170	0.08	0.08
400	200	0.03	0.1
500	170	0.125	0.12
500	200	0.095	0.11
500	300	0.025	0.09
600	170	0.16	0.09
600	200	0.115	0.065
600	300	0.045	0.09
600	400	0.035	0.06
750	170	0.185	0.025
750	200	0.155	0
750	300	0.11	0.035
750	400	0.07	0.065
750	500	0.035	0.065
1000	170	0.155	-0.02
1000	200	0.15	-0.1
1000	300	0.145	-0.03
1000	400	0.115	0.015
1000	500	0.115	0.015

Table 20: BDT thresholds to divide events into tight and loose regions, concerning the signals with different mass points. At least 2 sideband events in tight region are guaranteed.

6 Signal and background estimations

In this analysis, the statistical result is obtained from a binned signal + background fit $m_{\gamma\gamma}$ distribution in data. The shape of the SH signal is derived from MC simulation histogram, analytical function is not used to describe the signal for this binned model. The continuum $\gamma\gamma$ background is modeled using data in 0ℓ control region, following the strategy of previous $X \to hh \to WW\gamma\gamma$ analysis with 36 fb^{-1} data in ATLAS[35]. An analytical function is fitted to side band data, and the choice of the function form is determined from so-called spurious signal approach.

6.1 Models of signals, SM Higgs and SM Higgs pair backgrounds

6.2 Continuum background estimation

The continuum background is expected to be a smoothly falling shape that can be modeled with an analytical function. So this background yield under the signal peak can be modeled with a functional form that is largely constrained by the mass sidebands. Due to the low statistics of sideband data with 1/2 leptons in the final state, the function is determined using the $m_{\gamma\gamma}$ shape in 0 lepton + jets control region data, where the control region here is defined as the events with at least two loose photons, and can not pass two tight photon selections, and the mass is ranged between (105, 120) and (130, 160) GeV. This sample has enough statistics and it is the ideal model to estimate the background shape. The uncertainty from this estimation is then covered by one systematic uncertainty term which is derived from MC discussed in Section 7.2.2. This strategy was used in previous analysis[35].

Two different kinds of control regions are adapted based on the phase space of different channels. Without real lepton, the jet in the control region is then faked as the lepton to simulate the behavior of lepton. Therefore, the control region for the 1-lepton case (WW1l channel) is defined as $\gamma\gamma + 0l + 1j$. And the control region for the 2-lepton case (WW2l, ZZ2l, $WWe\mu$) is defined as $\gamma\gamma + 0l + \geq 2j$.

In Section 5 the BDT is constructed independently with $m_{\gamma\gamma}$, as shown in Table 16, the correlation between diphoton mass and BDT cut value is small. So in the certain range, the shape of diphoton mass is consistent for the BDT cut value. Then finally the mass shape is fitted separately in the BDT tight region and BDT loose region. Specifically, to apply the BDT training on the control region data events, the jet information is fakely used as the lepton kinematics instead. As 20 mass points, 2 BDT region and 2 lepton channels, finally 80 different shapes are fitted from the sideband data.

The analytical function used to describe the model is chosen from the spurious signal test. Following functions are considered in this analysis:

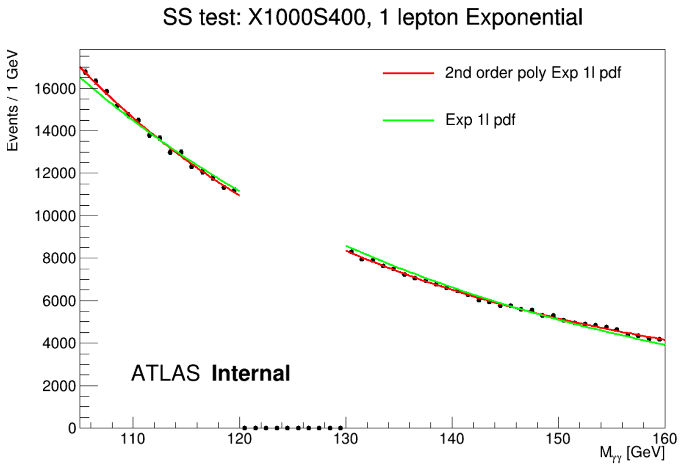
- Exponential: $e^{c \cdot m_{\gamma\gamma}}$
- Exponential of 2^{nd} Order Polynomial: format as $e^{c_1 \cdot m_{\gamma\gamma}^2 + c_2 \cdot m_{\gamma\gamma}}$
- Chebyshev polynomial of order N: N=1-5.

Some other candidate functions, e.g. Bernstain polynomials showing weird fitting behavior, are dropped after the tests.

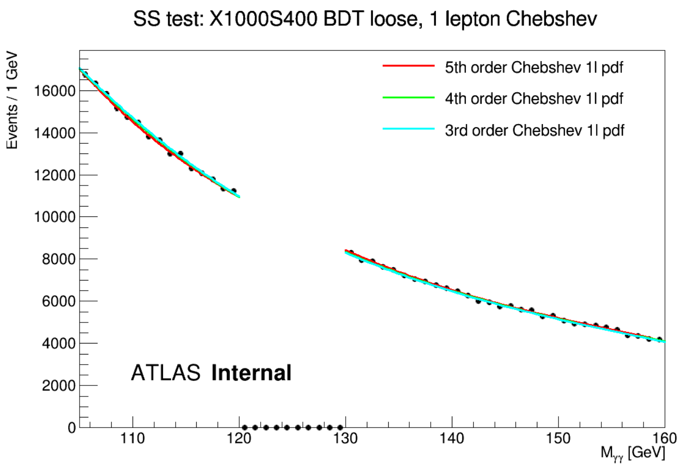
For the convenience, the 2^{nd} order exponential polynomial function is adjusted to:

$$f(x) = e^{c_1 \cdot (m_{\gamma\gamma} - 125)^2 / 125^2 + c_2 \cdot (m_{\gamma\gamma} - 125)/125} \tag{6}$$

Taking $m_X = 1000\,\text{GeV}, m_S = 400\,\text{GeV}$ as the example, in 1-lepton channel BDT loose region, the continuum background distribution from the CR 0l+1j sideband data, is fitted with different functions and shown in Figure 19. Figure 20 displays the $m_{\gamma\gamma}$ distributions from BDT tight and BDT loose regions.



(a)

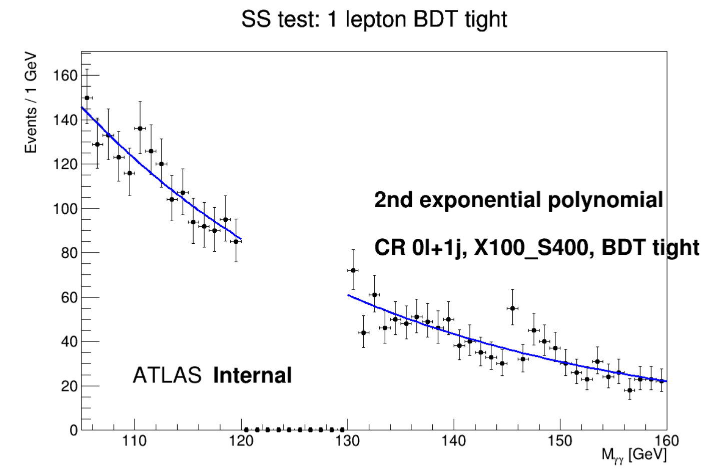


(b)

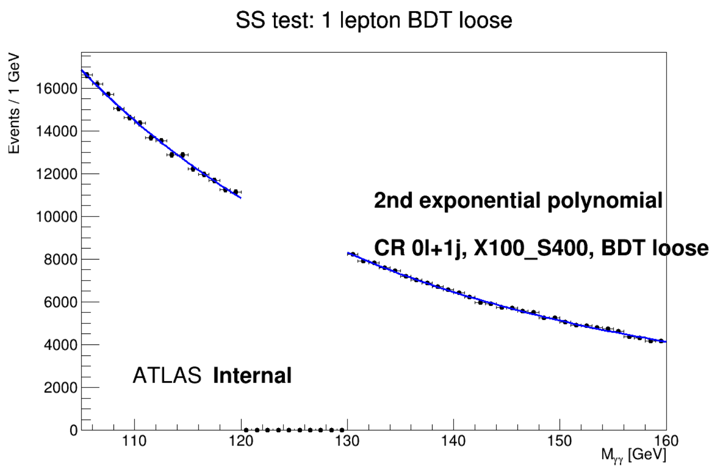
Figure 19: (a) Exponential and second order exponential fitted $m_{\gamma\gamma}$ distributions for 1l BDT loose region, (b) 3,4,5 order Chebyshev polynomial fitted $m_{\gamma\gamma}$ distributions for 1l BDT loose region

The main criterion used to select the functional form in each category is a bias test performed by fitting the control region data using a model with free parameters for both the signal and background event yields. The potential bias due to the mis-modeling of the background $m_{\gamma\gamma}$ distribution is estimated from the fitted signal yield, i.e. the spurious signal.

In the spurious signal test, the cross-section of the signal process is fixed to the expected upper limit with 95% CL instead of the initial one, namely 1 pb. The absolute value of the fitted signal yield $|S_{spur}|$ is considered the potential bias. The background shape function used in the test satisfying at least one of the following conditions is considered acceptable, named as the relaxed template in the previous HGam analysis:



(a)



(b)

Figure 20: (a) second order exponential fitted $m_{\gamma\gamma}$ distributions for 1l BDT tight region, (b) second order exponential fitted $m_{\gamma\gamma}$ distributions for 1l BDT loose region

- $S_{spur} < 10\% N_{s,exp}$ where $N_{s,exp}$ is the expected number of signal yields in that category ($\mu_{sp} = S_{spur}/N_{s,exp}$),
- $S_{spur} < 20\%\sigma_s$, where σ_s is the statistical uncertainty on the fitted signal yields when fitting the signal+background model to a background-only Asimov dataset ($Z_{sp} = S_{spur}/\sigma_{signal}$).
- $P(\chi^2) > 5\%$. This is the intergral is from χ^2 value to infinity on the Cauchy distribution of with its degree of freedoms.

80 spurious signals tests are performed for 20 mass points with four regions for each. The values and the plots for all the tests are documented in the Appendix B.

Then, the one with the smallest degree of freedom is chosen if the multiple functions pass the criterion.

Table 21 lists this spurious signal test results in the WW1l case, taking $m_X = 1000$ GeV, $m_S = 300$ GeV, BDT loose as the example. The same tests are applied to all the mass points. After the spurious signal test, the $\mu_{sp}[\%]$ is counted as the uncertainty due to the mis-modelling of the background.

Among all the mass points and for differnt categories, it is found that, all of the functions formats can pass the $P(\chi^2)$ criteria. However, the Chebyshev polynomials are easily failed in the μ_{sp} tests. On the other hand, the exponential and the second order exponential show the similar performances during the tests. Among 78 out of 80 shapes, the second order exponential function is chosen because of smaller μ_{sp}, and the exponential function is chosen for the left 2 shapes.

For μ_{sp} test results, none of them are larger than 3% in current study and for high mass point BDT regions, the μ_{sp} is found to be smaller than 1%. So it turns out that the uncertainty from spurious signal is small in this study.

1-lepton case					
Function	Ndof	$\mu_{sp}[\%]$	$Z_{sp}[\%]$	$P(\chi^2)[\%]$	Passed SS test
Exp	1	0.47	9.1	39.06	Yes
ExpPoly2	2	0.39	8.2	41.27	Yes
Cheb3	3	10.2	19.3	18.46	No
Cheb4	4	8.8	21.2	27.23	No
Cheb5	5	6.31	23.3	24.13	No

Table 21: The spuirous signal test result for 1 lepton channel in $m_X = 1000GeV, m_S = 300GeV$, BDT loose region.

Finally, the background renormalization factor is determined by the sideband data in each category. The di-photon mass spectrum can be found in Figure 21——24.

For each mass point, as discussed in Section 5, 6 different regions are defined. Among them, MVA methods are used in 4 regions, named as 1/2 lepton BDT tight/loose region. Cut based methods are used for the other 2 regions, named WW1e1m and ZZ2l, due to the limited sideband events. The background estimation paramaters and spuirous signal tests values for these two channels share the same as those for 2 lepton tight region.

These two channels only have 1 or 2 sideband data events after the event selection, so they are free from the constraint of the sideband threshold scans discussed in Section 5.2.

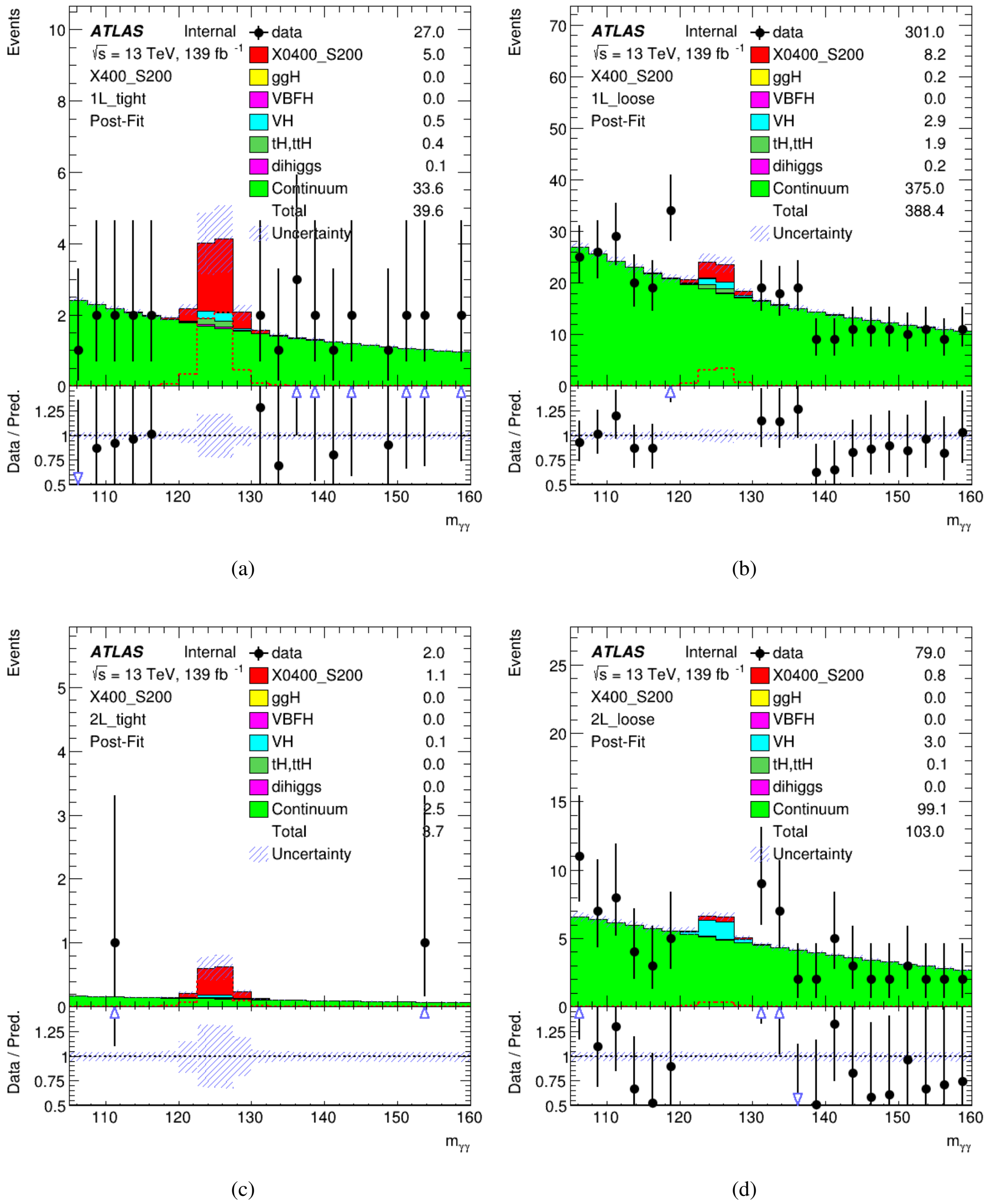
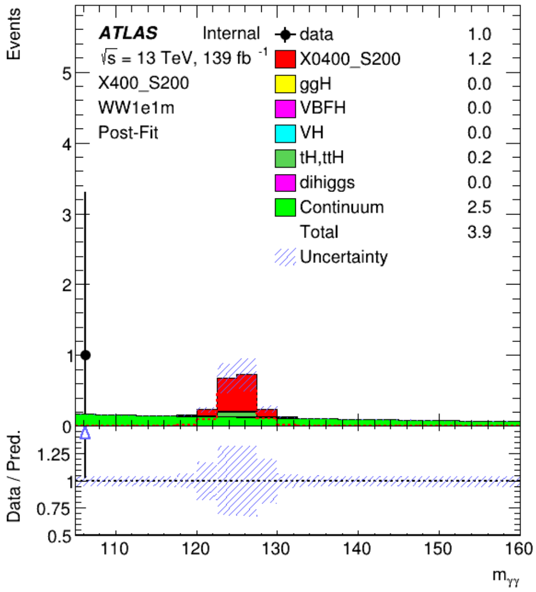
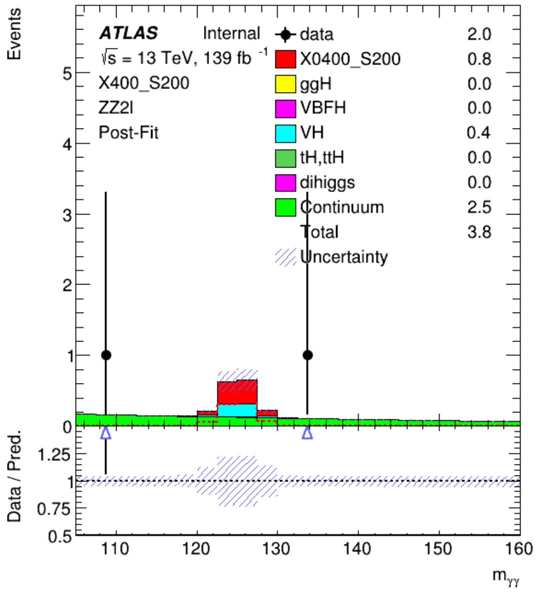


Figure 21: (a) $m_{\gamma\gamma}$ distributions for 1L BDT tight region, (b) $m_{\gamma\gamma}$ distributions for 1L BDT loose region, (c) $m_{\gamma\gamma}$ distributions for 2L BDT tight region, (d) $m_{\gamma\gamma}$ distributions for 2L BDT loose region, Events pass selections (mX,mS) =(400,200) GeV defined in signal region, with scaled smooth continuum background shapes.

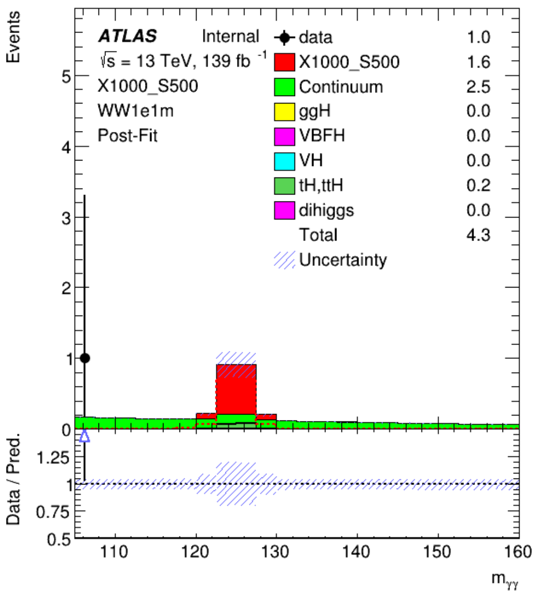


(a)

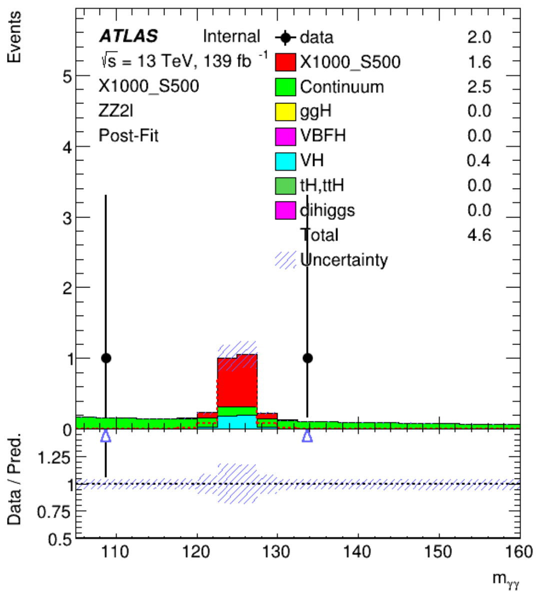


(b)

Figure 22: (a) $m_{\gamma\gamma}$ distributions for WW1e1m region, (b) $m_{\gamma\gamma}$ distributions for ZZ2l region, Events pass selections (mX,mS) =(400,200) GeV defined in signal region, with scaled smooth continuum background shapes.



(a)



(b)

Figure 23: (a) $m_{\gamma\gamma}$ distributions for WW1e1m region, (b) $m_{\gamma\gamma}$ distributions for ZZ2l region, Events pass selections (mX,mS) =(1000,500) GeV defined in signal region, with scaled smooth continuum background shapes.

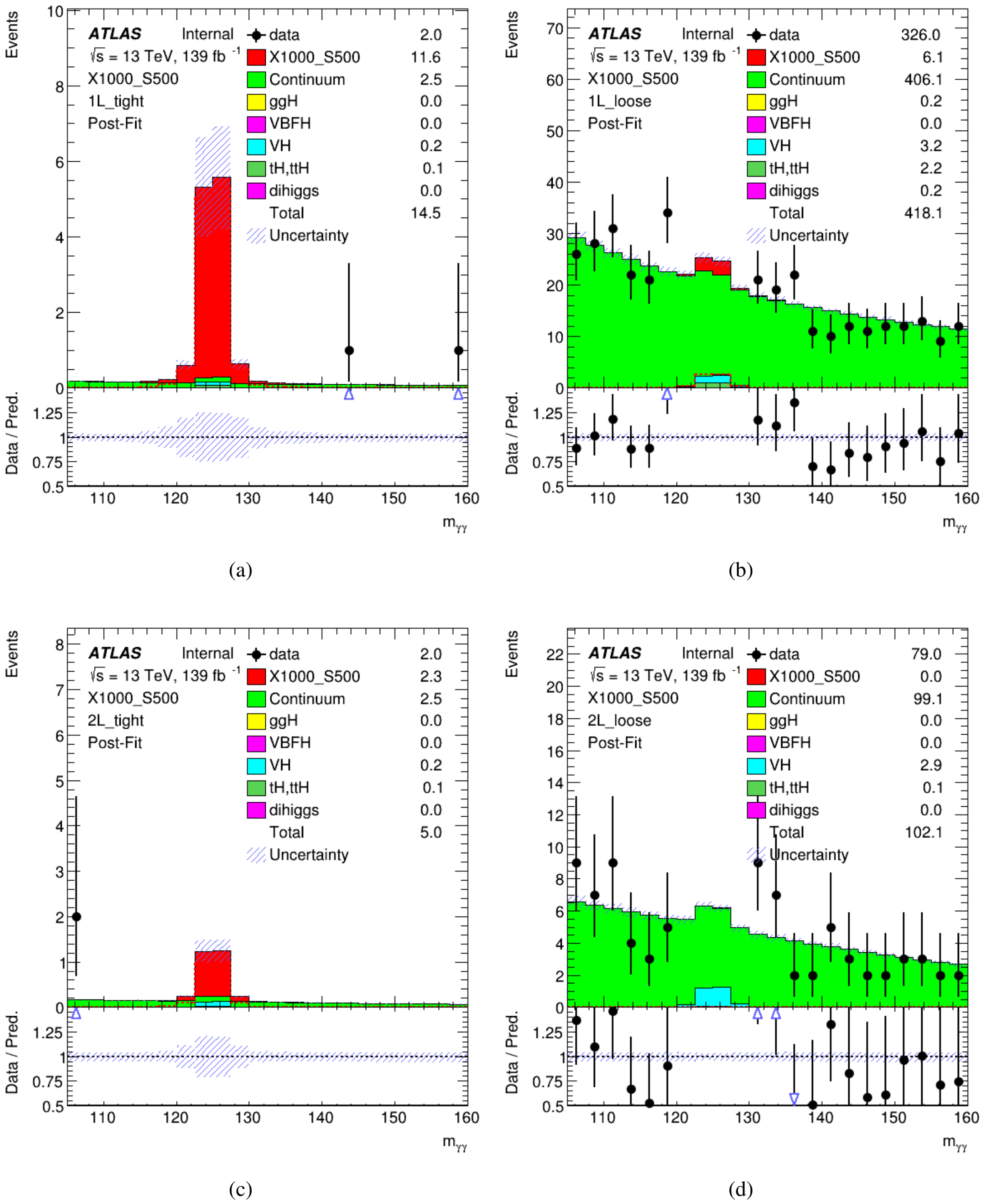


Figure 24: (a) $m_{\gamma\gamma}$ distributions for 1L BDT tight region, (b) $m_{\gamma\gamma}$ distributions for 1L BDT loose region, (c) $m_{\gamma\gamma}$ distributions for 2L BDT tight region, (d) $m_{\gamma\gamma}$ distributions for 2L BDT loose region, Events pass selections (mX,mS) =(1000,500) GeV defined in signal region, with scaled smooth continuum background shapes.

679 # 7 Systematic uncertainties

680 ## 7.1 Theoretical uncertainties

This search relies on the calculation of predicted event numbers for all the SM processes, the uncertainties arise from the imperfect knowledge of:

- the parton density functions (PDFs) and the value of the strong coupling constant, α_s;
- the QCD effects in the soft and collinear regime, hadronization and multi-parton interactions;
- the parton shower behavior varied from different algorithms Pythia and Herwig.

In current ATLAS analysis framework, all those impacts are quantified and stored in the event weights in the sample generation step. These event weights, mostly calculated by LHAPDF [**Butterworth:2014efa**], have the different models to estimate the impacts for these uncertainties, by comparing the deviation of the nominal treatment and the alternative treatment.

In HGam analysis, all these theoretical uncertainties are obtained in the mcEventweights variables for the different variations weights. The noiminal event weights are extracted from the PDF set 90400, named as PDF4LHC15_nol_30_pdfas, where the central value $\alpha_S(M_Z) = 0.118$. With the following set from 90401 to 90430, the PDF Hessian symmetric eigenvectors weights are stored. For set 90431 and 90432, the $\alpha_S(M_Z)$ values are set to 0.1165 and 0.1195, respectively.

In this case, for each sample, α_S theoretical uncertainties can be obtained as:

$$\delta_{\alpha_S} = \frac{|\delta(\alpha_S^{down}) - \delta(\alpha_S^{up})|}{2} \tag{7}$$

And with 30 eigenvectors and the nominal weight δ^0, the PDF uncertainties can be obtained as:

$$\delta_{PDF} = \sqrt{\sum_{k=1}^{30}(\delta^k - \delta^0)^2} \tag{8}$$

The variation of the renormalisation and factorisation scales are used to estimate the uncertainty due to missing higher order corrections. In LHAPDF, both the renormalisation factor μ_r and the factorisation scale μ_f are provided. In this case, 7 matrix elements are provided:

$$\{\mu_r, \mu_f\} \times \{0.5, 0.5\}, \{1, 0.5\}, \{0.5, 1\}, \{1, 1\}, \{2, 1\}, \{1, 2\}, \{2, 2\}. \tag{9}$$

Excluding the edge element $\{0.5, 2\}$ and $\{2, 0.5\}$, it provides full matrix element information for QCD scale variations. Following the recipe from Physics Modelling group, the signed maximum/minimum variations are chosen as the uncertainties:

$$\delta_{QCD_{up}} = max\big(\delta_{\mu_{r_i},\mu_{f_i}} - \delta_{\mu_{r_1},\mu_{f_1}}\big)\delta_{QCD_{down}} = min\big(\delta_{\mu_{r_i},\mu_{f_i}} - \delta_{\mu_{r_1},\mu_{f_1}}\big) \tag{10}$$

In the calculation, all the events passing the selections will be included in the event weight. For statistical limited channels, the events are not enough to estimate the uncertainty. So, in this analysis, for each signal mass point and each MC process, only one theoretical uncertainty value is provided by adding all the events passing event selections. The discrepancy of theoretical uncertainty between 1 lepton and 2 lepton channels are relatively small, while the difference are large between hardonic channels and channels with leptons.

To study the impact of parton shower uncertianty, 2 truth level samples, MadGraph with parton shower algorithm from Pythia8 and the same generator but showered with Herwig7, are produced with 10k events for each. The reason to choose truth level sample is that the reconstruction process do not affect the event weight calulated by LHAPDF. It is noted that, one specific model name SM_loop_twoscalar is assigned to Herwig7.

To compute the uncertainty, the preselections including trigger match, the truth lepton and truth photon satisfing p_t cut and isolation requirement have to be implemented to the sample. The uncertianty of parton shower can be expressed as the following:

$$\delta_{PartonShower} = \frac{\frac{\sum_{after}^{Herwig} w}{\sum_{before}^{Herwig} w}}{\frac{\sum_{after}^{Pythia} w}{\sum_{before}^{Pythia} w}} - 1 \tag{11}$$

where $\sum_{before}^{Herwig/Pythia} w$ refers to the sum of event weights for total 10k events without any selections, and $\sum_{after}^{Herwig/Pythia} w$ refers to the sum of weights for events which passed the above preselections. With this method, Parton Shower uncertainty values varied from 2% to 5% for different mass points.

7.2 Experimental Uncertainties

7.2.1 Uncertainties for Combined Performance nuisance parameters

In ATLAS experiments, experimental uncertainties are obtained in the nuisance parameters(NP). Specific combined performance groups study the impact of these uncertainties and provide the recommendation working points for the anlysis group to use, and the uncertainties are quantified with these nuisance parameters. In this analysis, the histogram of the up and down variations for those nuisance paramaeters are taken into account in the TRExFitter, so both the shape and the yields deviations are considered, for both signal and single Higgs process samples.

The uncertainties on signal yields can come from the measured integrated luminosity, the pileup reweighting, the spurious signal, and object (mainly from photons in this analysis) reconstruction as well as particle identification criteria.

The uncertainty due to the combined 2015-2018 integrated luminosity is 1.7%. It is derived, following the methodology documented in Ref [11], from a preliminary calibration of the luminosity scale using x-y beam-separation scans performed from 2015 to 2018. This uncertainty is applied to the signal, SM Higgs as well as the di-Higgs process.

84 experimental systematic sources are taken into account, listed in the Table 24 and 25.

All those nuisance parameters are officially provided by `HGamFramework`, and the name conventions are following the Dihiggs group scenarios. So for egamma related nuisance parameters, the simplified correlation scheme is used, for egamma scale and egamma resolution, only 1 nuisance parameter is used

1 lepton region				
Uncertainties	α_S (%)	PDF (%)	QCDup(%)	QCDdown (%)
ggF	3.39	3.67	25.89	-15.97
VBF	0.98	7.96	0.97	-0.52
WmH	0.83	5.78	2.77	-3.20
WpH	0.86	4.99	2.45	-3.08
qqZH	0.88	6.08	3.64	-3.68
ggZH	1.14	3.08	25.83	-19.45
ttH	2.00	5.21	7.39	-9.51
tHbj	0.00	17.01	8.38	-8.63
tHW	0.00	8.30	2.51	-2.02
2 lepton region				
Uncertainties	α_S (%)	PDF (%)	QCDup(%)	QCDdown (%)
ggF	4.37	4.39	38.96	-16.38
VBF	0.70	0.11	1.19	-1.12
WmH	1.02	4.83	1.87	-3.37
WpH	1.17	4.26	5.16	-4.12
qqZH	1.00	5.78	4.67	-3.91
ggZH	1.03	2.97	25.77	-19.41
ttH	1.92	4.87	6.63	-9.59
tHbj	0.00	30.57	9.91	-8.77
tHW	0.00	8.07	5.45	-6.46

Table 22: SM Higgs theoretical uncertainties for QCD, α_S and PDF variations.

1 lepton region					
Uncertainties	Parton Shower (%)	α_S (%)	PDF (%)	qcdup(%)	qcddown (%)
di-Higgs	2.51	0.93	3.87	13.22	-12.47
2 lepton region					
Uncertainties	Parton Shower (%)	α_S (%)	PDF (%)	qcdup(%)	qcddown (%)
di-Higgs	-2.85	0.93	3.94	13.17	-12.48

Table 23: Di-Higgs theoretical uncertainties for Parton shower, QCD, α_S and PDF variations.

Variable	Description
Run2_LUMI	Luminosity
PRW_DATASF	Pileup reweighting scale factor towards data
EG_SCALE_ALL	Photon Scale
EG_RESOLUTION_ALL	Photon Resolution
PH_EFF_ID_Uncertainty	Photon Identification
PH_EFF_ISO_Uncertainty	Photon Isolation
PH_EFF_TRIGGER_Uncertainty	Photon Trigger
EL_EFF_ID_TOTAL_1NPCOR_PLUS_UNCOR	Electron Identification
EL_EFF_Iso_TOTAL_1NPCOR_PLUS_UNCOR	Electron Isolation
EL_EFF_Reco_TOTAL_1NPCOR_PLUS_UNCOR	Electron Reconstruction
MET_SoftTrk_Scale	MET
MUON_EFF_ISO_STAT	Muon Isolation Statistical
MUON_EFF_ISO_SYS	Muon Isolation Systematic
MUON_EFF_RECO_STAT	Muon Reconstruction Statistical
MUON_EFF_RECO_STAT_LOWPT	Muon Reconstruction Statistical low p_T (< 15 GeV) case
MUON_EFF_RECO_SYS	Muon Reconstruction Systematic
MUON_EFF_RECO_SYS_LOWPT	Muon Reconstruction Systematic low p_T (< 15 GeV) case
MUON_EFF_TTVA_STAT	Muon Track-To-Vertex-Association Statistical
MUON_EFF_TTVA_SYS	Muon Track-To-Vertex-Association Systematic
MUON_ID	Muon Inner Detector track resolution
MUON_MS	Muon Muon Spectrometer track resolution
MUON_SAGITTA_RESBIAS	Muon Sagitta bias correction
MUON_SAGITTA_RHO	Muon Sagitta bias correction on ρ
MUON_SCALE	Muon Scale
FT_EFF_Eigen_B_0	B quark flavor
FT_EFF_Eigen_B_1	B quark flavor
FT_EFF_Eigen_B_2	B quark flavor
FT_EFF_Eigen_C_0	C quark flavor
FT_EFF_Eigen_C_1	C quark flavor
FT_EFF_Eigen_C_2	C quark flavor
FT_EFF_Eigen_C_3	C quark lavor
FT_EFF_Eigen_Light_0	Light quark flavor
FT_EFF_Eigen_Light_1	Light quark flavor
FT_EFF_Eigen_Light_2	Light quark flavor
FT_EFF_Eigen_Light_3	Light quark flavor
FT_EFF_Eigen_Light_4	Light quark flavor
FT_EFF_extrapolation	Flavor extrapolation
FT_EFF_extrapolation_from_charm	Flavor extrapolation from charm quark

Table 24: Nuisance Parameters Variables used in this analysis, including luminosity, EGamma, leptons, MET, and flavors.

Variable	Description
JET_EffectiveNP_Detector1	Jet energy scale on detector
JET_EffectiveNP_Detector2	Jet energy scale on detector
JET_EffectiveNP_Mixed1	Jet energy scale mixed
JET_EffectiveNP_Mixed2	Jet energy scale mixed
JET_EffectiveNP_Mixed3	Jet energy scale mixed
JET_EffectiveNP_Modelling1	Jet energy scale modelling
JET_EffectiveNP_Modelling2	Jet energy scale modelling
JET_EffectiveNP_Modelling3	Jet energy scale modelling
JET_EffectiveNP_Modelling4	Jet energy scale modelling
JET_EffectiveNP_Statistical1	Jet energy scale statiscal
JET_EffectiveNP_Statistical2	Jet energy scale statiscal
JET_EffectiveNP_Statistical3	Jet energy scale statiscal
JET_EffectiveNP_Statistical4	Jet energy scale statiscal
JET_EffectiveNP_Statistical5	Jet energy scale statiscal
JET_EffectiveNP_Statistical6	Jet energy scale statiscal
JET_BJES_Response	B jet energy scale response
JET_EtaIntercalibration_Modelling	Jet eta calibration modelling
JET_EtaIntercalibration_NonClosure_2018data	Jet eta calibration based on data 2018
JET_EtaIntercalibration_NonClosure_highE	Jet eta calibration based on high energy case
JET_EtaIntercalibration_NonClosure_negEta	Jet eta calibration based on negative eta case
JET_EtaIntercalibration_NonClosure_posEta	Jet eta calibration based on positive eta case
JET_EtaIntercalibration_TotalStat	Jet Eta calibration Statitical
JET_JER_EffectiveNP_1	Jet energy resolution
JET_JER_EffectiveNP_2	Jet energy resolution
JET_JER_EffectiveNP_3	Jet energy resolution
JET_JER_EffectiveNP_4	Jet energy resolution
JET_JER_EffectiveNP_5	Jet energy resolution
JET_JER_EffectiveNP_6	Jet energy resolution
JET_JER_EffectiveNP_7	Jet energy resolution
JET_JER_EffectiveNP_8	Jet energy resolution
JET_JER_EffectiveNP_9	Jet energy resolution
JET_JER_EffectiveNP_10	Jet energy resolution
JET_JER_EffectiveNP_11	Jet energy resolution
JET_JER_EffectiveNP_12restTerm	Jet energy resolution
JET_Flavor_Composition	Jet flavor composition
JET_Flavor_Response	Jet flavo response
JET_JER_DataVsMC_AFII	Jet data and MC deviation on AFII(optinal)
JET_JvtEfficiency	Jet vertex tagger
JET_Pileup_OffsetMu	Jet pileup offset for μ
JET_Pileup_OffsetNPV	Jet pileup offset for number of primary vertices
JET_Pileup_PtTerm	Jet pileup for p_{T}
JET_Pileup_RhoTopology	Jet pileup for ρ topology
JET_PunchThrough_AFII	Jet punch through for AFII(optinal)
JET_RelativeNonClosure_AFII	Jet relative for AFII(optinal)
JET_SingleParticle_HighPt	Jet correction for high p_{T} single pariticle
JET_fJvtEfficiency	Forward jet vertex tagger

Table 25: Jet related nuisance parameters variables used in this analysis.

(SCALE_ALL, RESOLUTION_ALL). Although egamma systematics play a significant role among all the NPs, their impact remains minimal, and this analysis does not demonstrate high sensitivity to photon behavior across various eta bins.

To calculate these systematics, HGamFramework is used to generate the specific systematic variation MxAODs. For one event in each systematic variation, the specific objects 4-momentum and event weight information is produced. The varied diphoton mass is chosed as the obseved variable, and the accumulated histogram including weight information for both up and down variation is produced. By comparing the nominal and the alternative histogram, both shape and yield information of variations are taken into account.

Almost all the nuisance parameters their impacts are small and can not pass the TRExFitter threshold, which is 0.5%. One reason is that, The way to calculate these

Each value of the systematics is computed as the relative difference from nominal signal MC samples with $\pm 1\sigma$ variation:

$$\delta n_{\mathrm{c}}^{\pm 1\sigma} = \frac{n_{\mathrm{c}}^{\pm 1\sigma}}{n_{\mathrm{c}}^{\mathrm{nom}}} - 1 \tag{12}$$

Systematic uncertainties are computed for each individual category c. All the systematic sources are implemented in the fit with asymmetric constraints since up and down variations can have different values. A threshold of 0.5% on the variation value is applied when implementing these nuisance parameters to suppress trivial contributions and simplify the computing processes. After this selection, there are 32-39 terms left depending on categories.

7.2.2 Uncertainty on continuum background estimation

As described in Section 6.2, the continuum background models for different categories are determined by fitting the di-photon mass on the sideband data in the region of $\gamma\gamma + 0 - lepton$. The shape differences between this region and $\gamma\gamma + \ell\nu jj$ (1L) and $\gamma\gamma\ell\nu\ell\nu$ (2L) could introduce additional uncertainties in the estimations of background yields for individual categories. This method and corresponding uncertainties can only be evaluated by the MC, due to the low statistics in data. In practice, the region of $\gamma\gamma + 0 - lepton + jets$ as introduced in Section 2.2.3 is adapted to mimic the event topologies for 1-lepton and 2-lepton categories. The mass distributions for di-photon are fitted with the selected function obtained in Section 6.2. The variations between $\gamma\gamma + 0 - lepton + jet$ and $\gamma\gamma + \ell\nu jj/\gamma\gamma\ell\nu\ell\nu$ are computed bin-by-bin, in the region of $m_{\gamma\gamma} \in [105, 160]$GeV with bin width equal to 1 GeV. The average variation for all bins is chosen as the uncertainty of the background modeling.

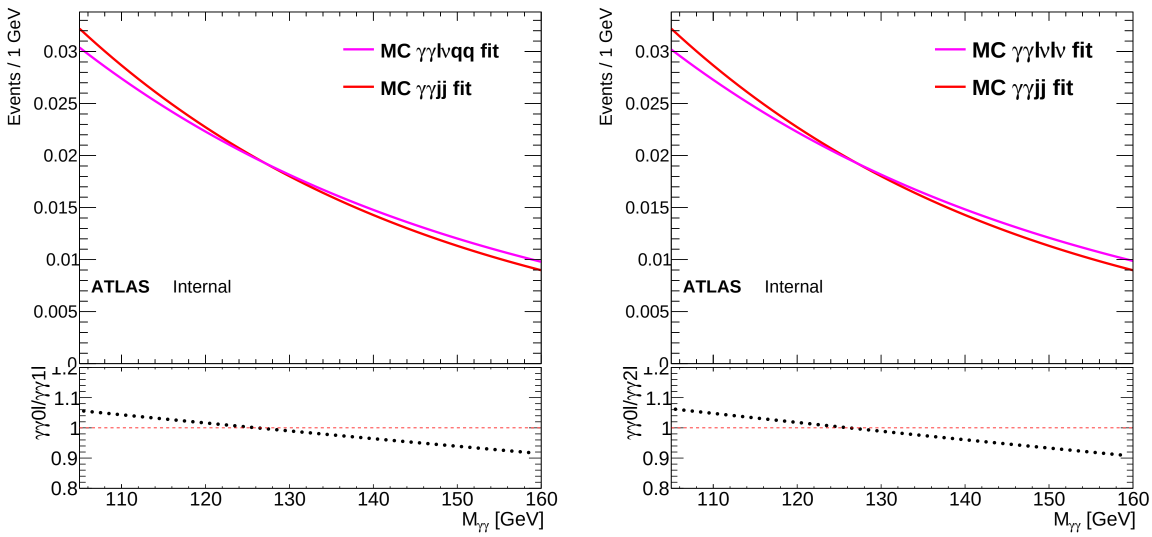


Figure 25: Deviations of di-photon mass distributions between $\gamma\gamma + 0L$ and $\gamma\gamma + 1L$ ($\gamma\gamma + 2L$).

8 Statistical interpretation

8.1 Statistical model

The statistical model is built up with a binned likelihood function. The model is constructed in the following form.

The signal extraction from data is based on the statistical model of binned likelihood estimation. This binned likelihood is built from Poisson distribution in each bin as follows:

$$\mathcal{L} = \prod_{c\in\text{channels}} \prod_{b\in\text{bins}} \text{Poisson}(n^{\text{obs}}_{c,b} | n^{S}_{c,b}, n^{B}_{c,b}) \times \prod_{s\in\mathbb{S}} \mathcal{G}(0|\theta_s, 1), \tag{13}$$

where c stands for channel index, and b is the bin index for each channel. The number of events observed in each bin is shown as $n^{\text{obs}}_{c,b}$, and the expected numbers of signal and background in such bin are $n^{S}_{c,b}$ and $n^{B}_{c,b}$, respectively. $n^{S}_{c,b}$ can be written as a function of the cross-section of X production: σ, which is shared among different bins and channels. This parameter is treated as the Parameter-of-Interest (POI). The number of background includes contributions from the SM resonant backgrounds n_{SM} and continuum backgrounds n_{cont}. The shape of the continuum background is determined by sideband fit and the normalization is floating during the fit. The systematic uncertainty is shown as marked as s for each term and constraint with a Gaussian distribution: $\mathcal{G}(0|\theta_s, 1)$, where θ_s is the nuisance parameter (NP) in fit.

8.2 Upper limit setting

A likelihood ratio-based test statistic is used in the statistical analysis. It is defined as follows:

$$\tilde{q}_\sigma = \begin{cases} -2\ln \frac{\mathcal{L}(\sigma,\hat{\hat{\boldsymbol{\theta}}}(\sigma))}{\mathcal{L}(0,\hat{\hat{\boldsymbol{\theta}}}(0))} & \text{if } \hat{\sigma} < 0 \\ -2\ln \frac{\mathcal{L}(\sigma,\hat{\hat{\boldsymbol{\theta}}}(\sigma))}{\mathcal{L}(\hat{\sigma},\hat{\boldsymbol{\theta}})} & \text{if } 0 \leq \hat{\sigma} \leq \sigma \\ 0 & \text{if } \hat{\sigma} > \sigma \end{cases} \tag{14}$$

where $\mathcal{L}$ stands for the likelihood function for the statistical model of the analysis, $\boldsymbol{\theta}$ is a set of nuisance parameters through which the systematic uncertainties are introduced, and the parameter of interest σ is the cross-section of resonant X production. Single hat stands for unconditional fit and double hat for conditional fit, i.e., POI σ is fixed to a certain value. With this test statistic, the upper limits of the cross section X production at 95% confidence level can be derived by using the CL_s method [42] under the asymptotic approximation [43].

9 Results

The simultaneous binned fit on $m_{\gamma\gamma}$ sharing the same set of parameters for the background modeling is performed on different categories with `TRExFitter`. One advantage of binned fit is that the variations of the nuisance parameters can be easily incorporated into the content for each bin. Thus, it is easier for the estimation of some systematics in particular those for shape uncertainties.

Configurations of `TRExFitter` are listed in Table 26. The binning width for di-photon mass is chosen to be 2.5GeV which is a bit larger than the di-photon mass resolution from Higgs in ATLAS.

Parameters	Value
`MCstatThreshold`	0.005
`SystPruningNorm`	0.005
`SystPruningShape`	0.005
`BlindSRs`	FALSE
`FitType`	SPLUSB
`FitRegion`	CRSR
`LimitType`	ASYMPTOTIC
`Observed Variable`	$m_{\gamma\gamma}$
`Variable Range`	$(105, 160)$GeV
`Blind Range`	$(120, 130)$GeV
`Numebr of bins`	22
`Bin width`	2.5 GeV

Table 26: Summary of configurations used in `TRExFitter`.

9.1 Expected results

The expected 95% C.L. upper limits on cross-section are derived for X separately assuming $S \to WW$ with 100%, S decaying 100% to ZZ, or S decays to WW and ZZ following the SM prediction. Figure 26 - 28 show the results with the 3 scenarios above. The upper limits in other m_X values are extrapolated to a plane through the existing results. $m_X = 1000GeV, m_S = 400GeV$ provides the best limit among all the mass points with the assumption of the decay of S 100% to WW. The exact values of the limits in detail are summarized in Table 27, 28 and 29.

In order to compute the limits for the $\sigma(pp \to X \to S(\to ZZ/WW)H)$, WW1l, WW2l , WW$e\mu$ and ZZ2l channels are combined in which WW1l dominates, as Figure 28 and Table 29 show. The results are obtained with Asymptotic fits on the Asmovi data and a cross check with throwing toy MC to extract the limit which is documented in Appendix H. As Table 32 shows, these two methods agree with each other. Moreover, the signal injection test has been done and fitted signal strengths are consisent with the injected ones as Appendix I shows.

To extrapolate the existing results to the whole X-S plane, linear fits are implemented to obtain the limits.

m_X [GeV]	$m_S[GeV]$	$+2\sigma$ [pb]	$+1\sigma$ [pb]	Median [pb]	-1σ [pb]	-2σ [pb]
300	170	1.327	0.863	0.578	0.417	0.310
400	170	1.166	0.788	0.535	0.385	0.287
400	200	1.126	0.769	0.524	0.378	0.281
500	170	0.762	0.499	0.333	0.240	0.179
500	200	0.831	0.557	0.376	0.271	0.202
500	300	0.981	0.673	0.460	0.332	0.247
600	170	0.648	0.420	0.280	0.202	0.150
600	200	0.573	0.380	0.256	0.184	0.137
600	300	0.615	0.416	0.282	0.203	0.151
600	400	0.796	0.538	0.364	0.263	0.196
750	170	0.564	0.356	0.235	0.169	0.126
750	200	0.450	0.293	0.195	0.140	0.105
750	300	0.450	0.298	0.200	0.144	0.108
750	400	0.466	0.300	0.199	0.144	0.107
750	500	0.776	0.523	0.355	0.256	0.191
1000	170	0.410	0.254	0.167	0.120	0.089
1000	200	0.326	0.202	0.133	0.096	0.071
1000	300	0.280	0.175	0.115	0.083	0.062
1000	400	0.272	0.172	0.113	0.081	0.061
1000	500	0.309	0.196	0.129	0.093	0.069

Table 27: Upper limits at the 95% confidence level for the cross-section of the gluon fusion production of the resonance $X \to SH$ and the S particle is assumed to decay 100% to WW.

9.2 Pull, ranking and correlation matrix

For the process of the fit, if the impacts of the systematic uncertainties are less than 0.5%, the corresponding NPs are dropped by the fit tool to simplify the procedure. The pull and ranking distributions for survived NPs are shown in Figure 29 for the signal with $m_X = 1000$ GeV and $m_S = 500$ GeV. Correlations between major NPs can be found in Figure 30 for this signal. No obvious abnormal behaviors from the pulls and their constraints have been observed in this analysis during the Asimov fit and the most significant impacts of the uncertainties on the extracted signal yield turn out to be from egamma systematic uncertainties.

9.3 Interpolation

Based on the available mass point results, it is possible to extend the estimation to cover the entire X-S mass plane. Figure 28 demonstrates that in the log view, the limit is nearly linear for the same S value across different X mass points. Furthermore, for a certain S mass, the expected limit at a larger X mass is anticipated to be lower due to the larger phase space of the X particle. Similarly, for a certain X mass, a higher S mass suggests a better limit result due to the larger phase space of bosons. These observations suggest that extended interpolations can be conducted to cover the entire X-S mass plane.

Figure 31 shows the initial 20 mass points, and the entire X-S plane is divided into grids for each X mass grid of 20 GeV and each S mass grid of 10 GeV. For a known X mass or S mass as a variable x, and the

m_X [GeV]	$m_S[GeV]$	$+2\sigma$ [pb]	$+1\sigma$ [pb]	Median [pb]	-1σ [pb]	-2σ [pb]
300	170	5.059	3.230	2.119	1.527	1.137
400	170	3.875	2.469	1.620	1.167	0.870
400	200	2.597	1.656	1.088	0.784	0.584
500	170	3.075	1.953	1.280	0.923	0.687
500	200	2.187	1.392	0.915	0.659	0.491
500	300	2.142	1.366	0.900	0.648	0.483
600	170	2.725	1.726	1.130	0.814	0.606
600	200	1.926	1.223	0.803	0.579	0.431
600	300	1.753	1.113	0.732	0.527	0.393
600	400	2.036	1.297	0.854	0.615	0.458
750	170	2.537	1.606	1.050	0.757	0.564
750	200	1.825	1.155	0.758	0.547	0.407
750	300	1.417	0.898	0.589	0.425	0.316
750	400	1.532	0.972	0.639	0.460	0.343
750	500	1.749	1.112	0.731	0.527	0.393
1000	170	2.384	1.499	0.983	0.708	0.527
1000	200	1.999	1.260	0.825	0.594	0.443
1000	300	1.216	0.766	0.502	0.362	0.270
1000	400	1.209	0.762	0.500	0.360	0.268
1000	500	1.272	0.803	0.526	0.379	0.282

Table 28: Upper limits at the 95% confidence level for the cross-section of the gluon fusion production of the resonance $X \to SH$ and the S particle is assumed to decay fully to ZZ.

expected upper limit as variable y, the slope and intercept for these two points can be derived.

$$k = \frac{\log_{10} y_2 - log_{10} y_1}{x_2 - x_1} \tag{15}$$

$$b = \log_{10} y_2 - k \times x_2 \tag{16}$$

So for the certain mass with x to interpolate, the expect limit can be calculated as:

$$y = 10^{k \times x + b} \tag{17}$$

To perform bisect linear interpolation within the line segment, one starts at the endpoints of the segment and moves inward towards the midpoint by halving the distance, and the expected limit is interpolated between two neighboring points. This process is repeated until the grid precision of X mass grid of 20 GeV and S mass grid of 10 GeV is achieved.

During the interpolation, the initial 20 mass points are first connected since these results are the most reliable. Subsequently, the square area is second-order interpolated from these first-order derivatives. Two important considerations are made during this process. Firstly, the mass constraint is upheld to ensure that $m_X > m_S + m_H$. Secondly, in the scenario where S decays 100% into ZZ, the initial result for $m_S = 170$ GeV is not used due to the ZZ being off-shell at this mass point. In contrast, the result for $m_S = 180$ and 190 GeV is backward interpolated from $m_S = 200$ GeV and $m_S = 300$ GeV.

m_X [GeV]	$m_S[GeV]$	$+2\sigma$ [pb]	$+1\sigma$ [pb]	Median [pb]	-1σ [pb]	-2σ [pb]
300	170	1.360	0.896	0.604	0.435	0.324
400	170	1.130	0.762	0.517	0.372	0.277
400	200	1.636	1.138	0.786	0.567	0.422
500	170	0.756	0.493	0.330	0.238	0.177
500	200	1.020	0.683	0.462	0.333	0.248
500	300	0.666	0.974	0.666	0.480	0.357
600	170	0.651	0.420	0.280	0.201	0.150
600	200	0.745	0.496	0.335	0.241	0.180
600	300	0.842	0.571	0.389	0.280	0.209
600	400	1.248	0.844	0.574	0.414	0.308
750	170	0.567	0.366	0.243	0.175	0.130
750	200	0.572	0.369	0.246	0.177	0.132
750	300	0.631	0.429	0.290	0.209	0.155
750	400	0.820	0.547	0.370	0.266	0.198
750	500	1.020	0.687	0.466	0.336	0.250
1000	170	0.421	0.267	0.176	0.127	0.094
1000	200	0.401	0.254	0.168	0.121	0.090
1000	300	0.404	0.260	0.173	0.125	0.093
1000	400	0.471	0.303	0.201	0.145	0.108
1000	500	0.565	0.365	0.243	0.175	0.130

Table 29: Upper limits at the 95% confidence level for the cross-section of the gluon fusion production of the resonance $X \to SH$ and the S particle is assumed to decay to WW/ZZ following the SM prediction.

Using the derived interpolations, Figure 32 through 34 are generated to demonstrate the expected limit across the entire plane, based on the nominal assumption that S decays in a Higgs-like way, and for two scenarios in which S decays 100% into W or Z pairs for the convenience of theorists. During the bisect linear interpolations, the maximum and minimum results are always at the endpoints of each segment, so the expected limit for the best result is kept the same.

To validate the interpolation, a blind test is performed by comparing the interpolated limit and the original limit. e.g. for point $m_X = 750GeV, m_S = 300GeV$, the original expected limit is 290 fb, while the interpolated limit from $(m_X, m_S) = (600, 300), (1000, 300)$ two points is 291.417 fb. Similarly, for same X mass but $m_S = 200GeV$, the original expected limit is 264fb while the interpolated limit is 270.235 fb. Two results are well consistent with each other. The largest deviation is found in the phase-space limited cases where $m_X = 600$ GeV and $m_S = 400$ GeV, with a deviation of about 20%. In conclusion, the interpolation method is valid and robust for the entire plane.

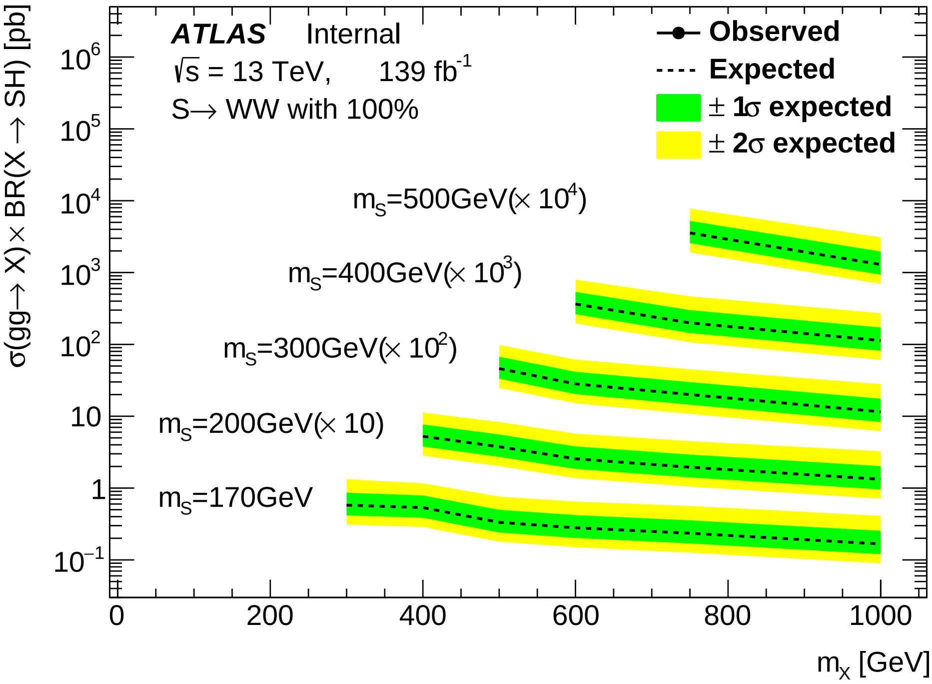


Figure 26: Expected 95% C.L. upper limits on $\sigma(pp \to X \to SH)$ for resonance as a function of the mass of the heavy scalars X and S, assuming $S \to WW$ with 100%.

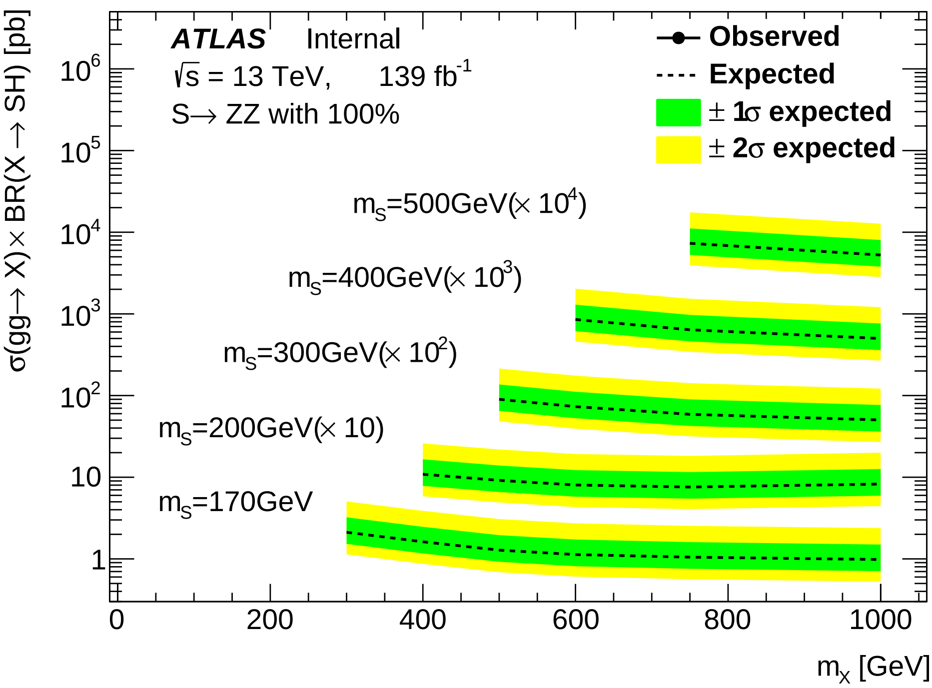


Figure 27: Expected 95% C.L. upper limits on $\sigma(pp \to X \to SH)$ for resonance as a function of the mass of the heavy scalars X and S, assuming $S \to ZZ$ with 100%.

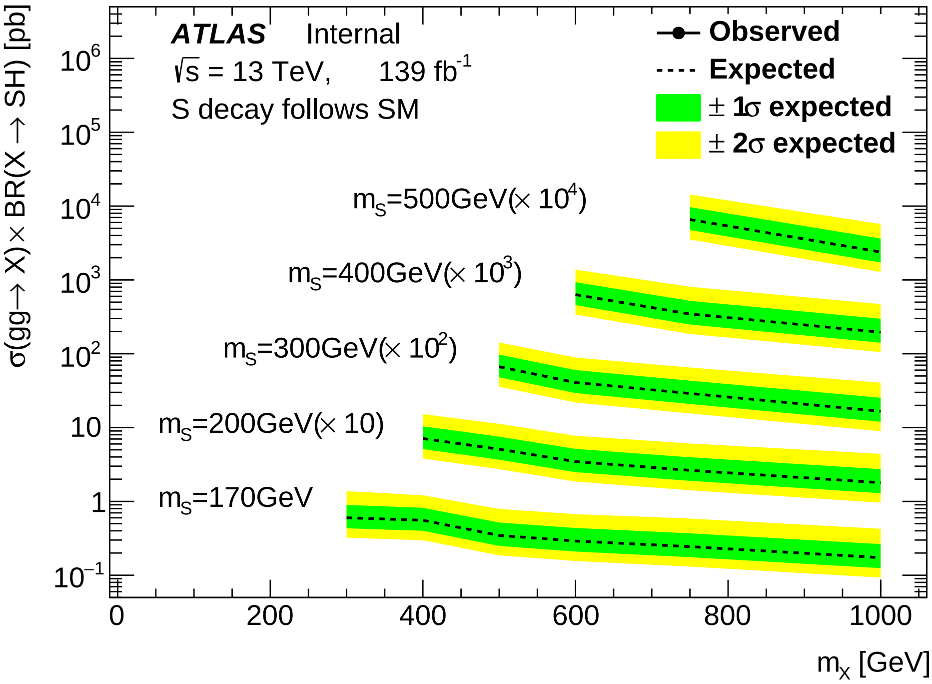


Figure 28: Expected 95% C.L. upper limits on $\sigma(pp \rightarrow X \rightarrow SH)$ for resonance as a function of the mass of the heavy scalars X and S, where the branching ratio of S to WW/ZZ is assumed to be SM-like.

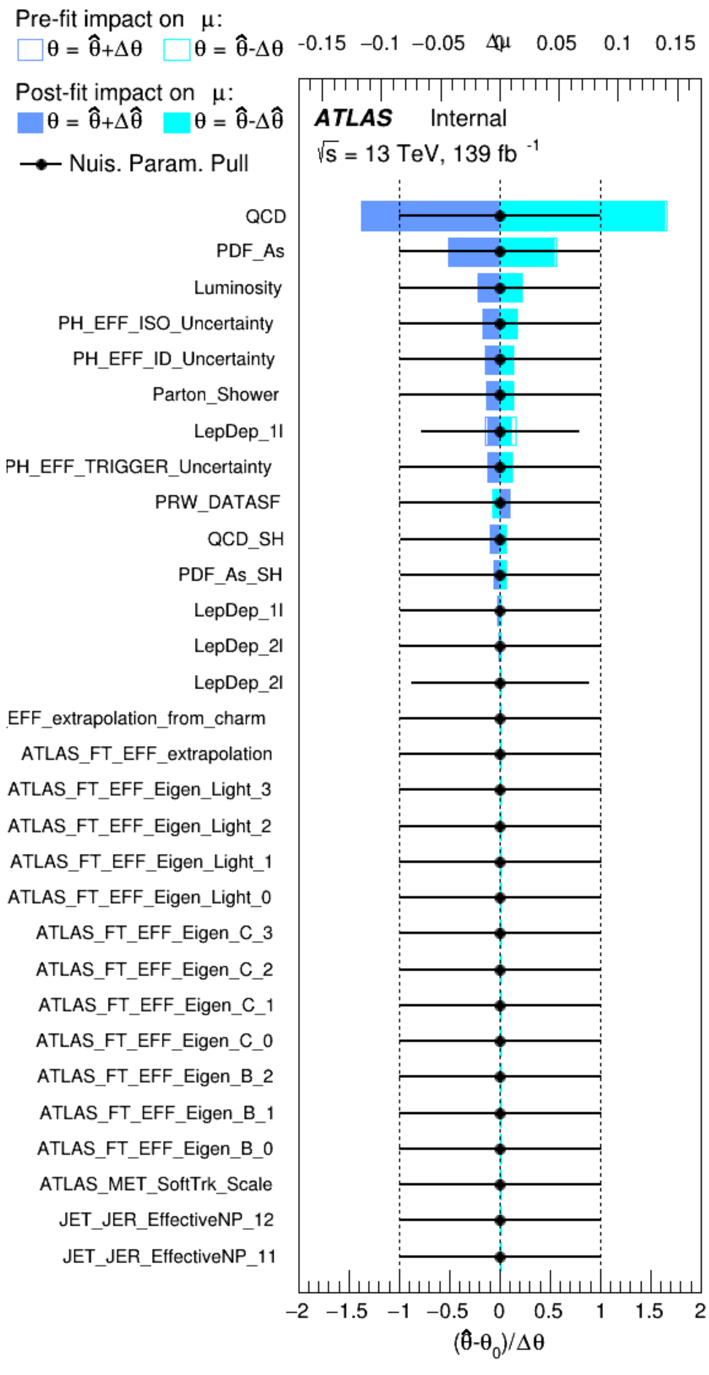


Figure 29: NPs ranking and pull distrbutions in (m_X1000 GeV, m_S = 500 GeV) fit.

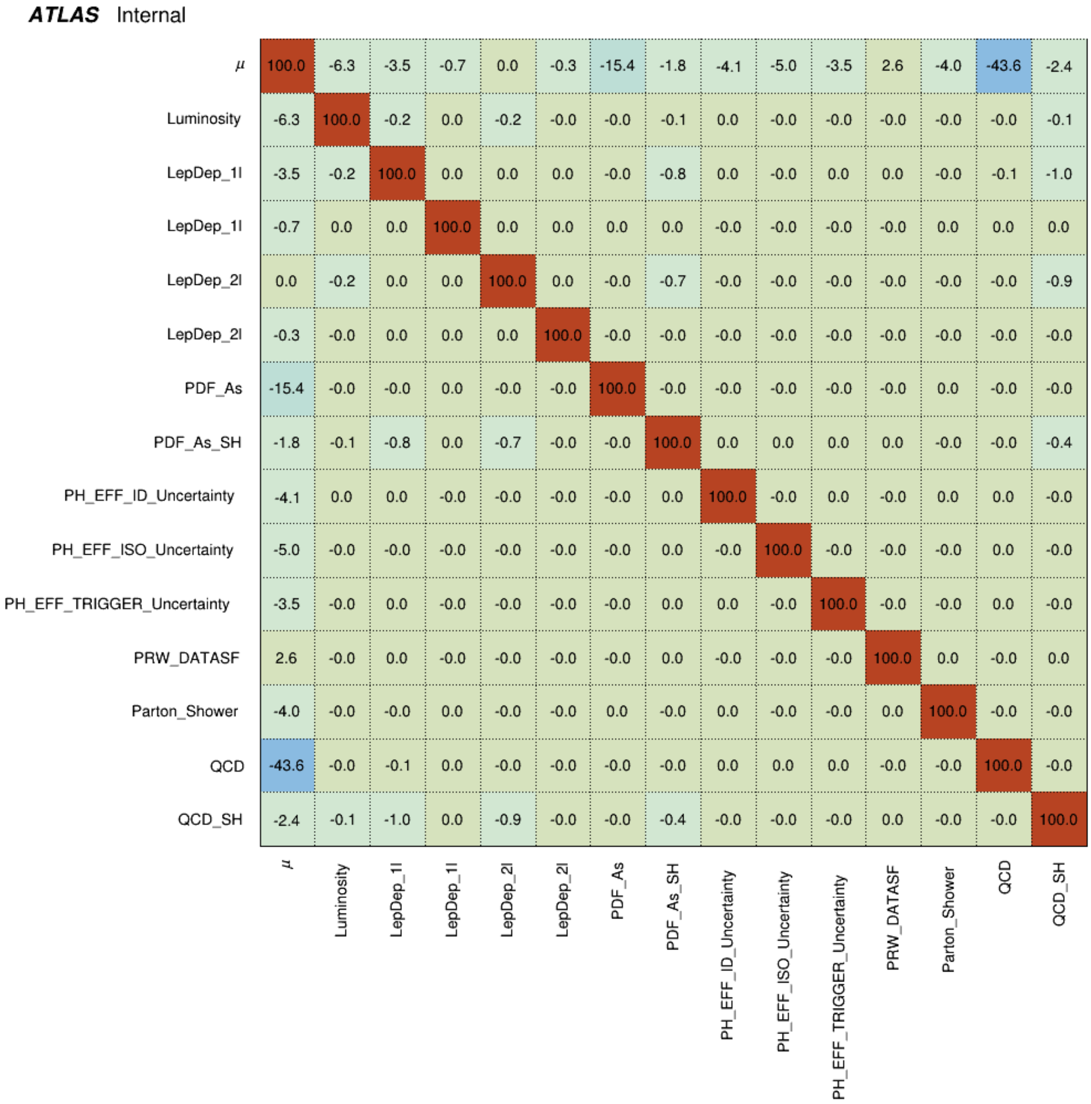


Figure 30: Major NP correlation in (m_X1000 GeV, m_S = 500 GeV) fit.

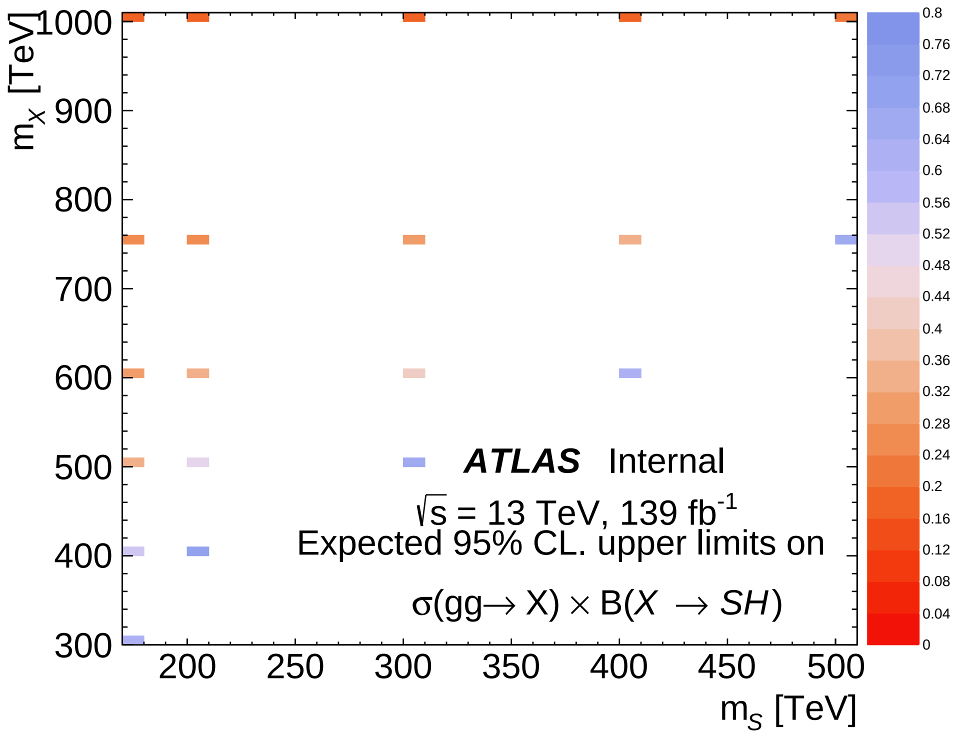


Figure 31: Expected 95% C.L. upper limits on $\sigma(pp \to X \to SH)$ for resonance as a function of the mass of the heavy scalars X and S, shown in a 2D mass plane, for initial 20 mass points.

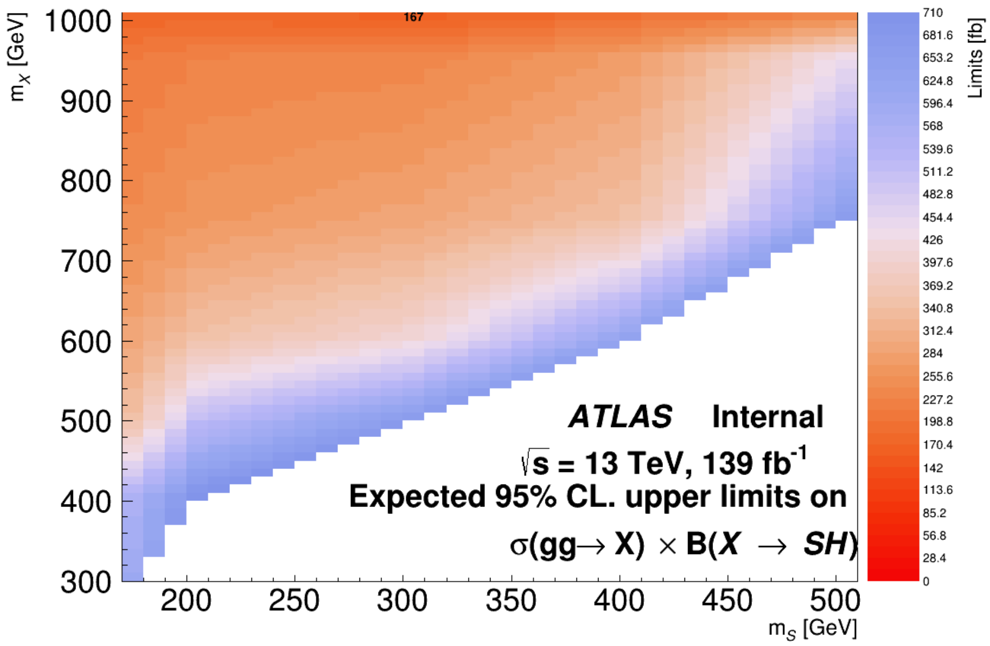


Figure 32: Expected 95% C.L. upper limits on $\sigma(pp \to X \to SH)$ for resonance as a function of the mass of the heavy scalars X and S, shown in a 2D mass plane.

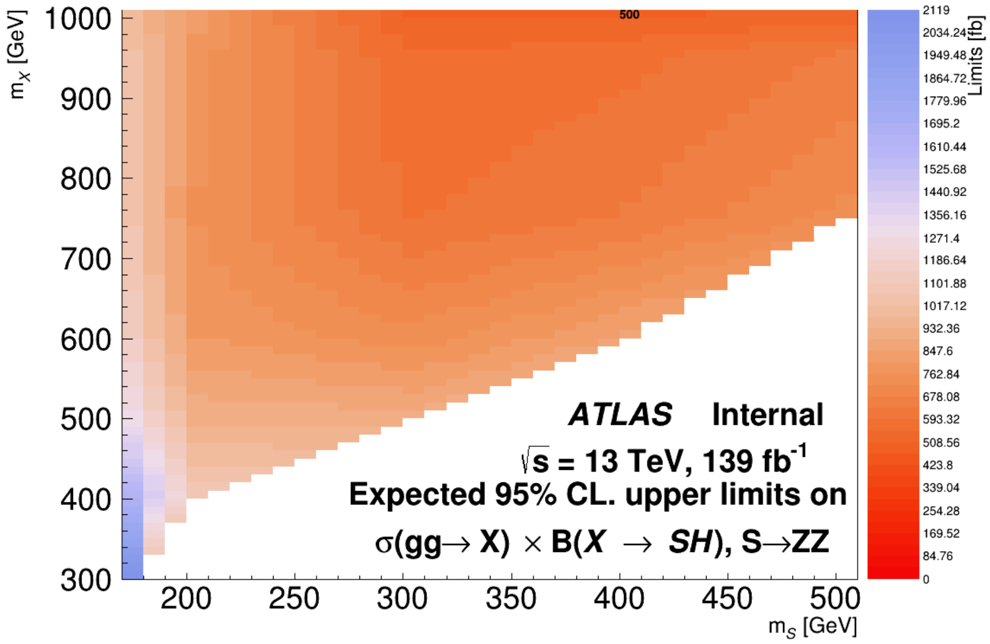


Figure 33: Expected 95% C.L. upper limits on $\sigma(pp \rightarrow X \rightarrow SH)$ for resonance as a function of the mass of the heavy scalars X and S, in assumption S 100% decaying into Z pairs, shown in a 2D mass plane.

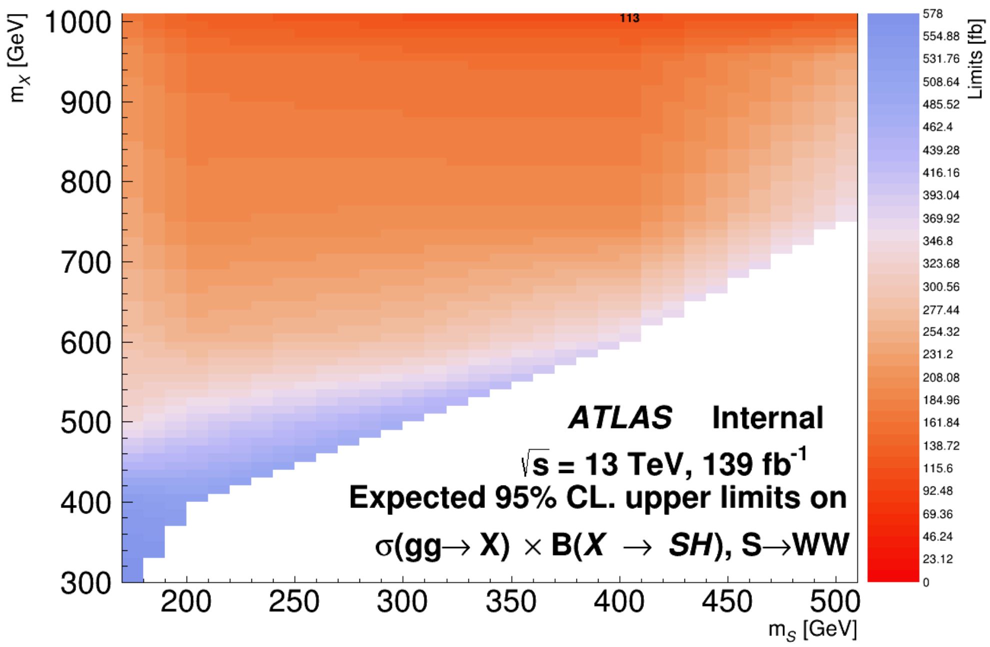


Figure 34: Expected 95% C.L. upper limits on $\sigma(pp \rightarrow X \rightarrow SH)$ for resonance as a function of the mass of the heavy scalars X and S, in assumptio S 100% decaying into W pairs, shown in a 2D mass plane.

10 Summary

In this note, a search for heavy resonance X decaying into a new scalar S and a SM Higgs boson with subsequently this Higgs boson decaying to two photons and S decaying to WW or ZZ is performed. In this analysis, both fully leptonic and semileptonic decays of WW bosons and semileptonic decays of ZZ bosons are explored. Analysis selections are optimized separately for different final states based on their dedicated event topologies. In order to improve significance, the BDT method is performed based on reconstructed discriminating variables. An optimized threshold on the BDT output divides events into tight and loose regions to maximize the significance and the signal contribution is extracted from a binned fit to $m_{\gamma\gamma}$ distribution from 105 to 160 GeV. The signal Higgs boson contribution is determined from SM predictions and the non-resonant background is estimated from $m_{\gamma\gamma}$ sideband fit with analytic function in the $\gamma\gamma + 0L$ control region. The binned likelihood fit is performed by combining all channels as well as assuming S 100% decays to WW or ZZ to extract signal contributions with different m_X and m_S hypotheses.

The observed (expected) upper limits at the 95% confidence level on the cross-section for $gg \to X \to Sh$ assuming the decay of S following the SM prediction is between X fb (167 fb) and Y fb (710 fb).

References

[1] ATLAS Collaboration, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, Phys. Lett. B **716** (2012) 1, arXiv: 1207.7214 [hep-ex] (cit. on p. 8).

[2] CMS Collaboration, *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*, Phys. Lett. B **716** (2012) 30, arXiv: 1207.7235 [hep-ex] (cit. on p. 8).

[3] ATLAS Collaboration, *Searches for Higgs boson pair production in the* $hh \rightarrow bb\tau\tau, \gamma\gamma WW^{*}, \gamma\gamma bb, bbbb$ *channels with the ATLAS detector*, Phys. Rev. D **92** (2015) 092004, arXiv: 1509.04670 [hep-ex] (cit. on p. 8).

[4] ATLAS Collaboration, *Combination of searches for Higgs boson pairs in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, (2019), arXiv: 1906.02025 [hep-ex] (cit. on p. 8).

[5] CMS Collaboration, *Combination of Searches for Higgs Boson Pair Production in Proton–Proton Collisions at $\sqrt{s} = 13$ TeV*, Phys. Rev. Lett. **122** (2019) 121803, arXiv: 1811.09689 [hep-ex] (cit. on p. 8).

[6] S. von Buddenbrock et al., *Phenomenological signatures of additional scalar bosons at the LHC*, Eur. Phys. J. **C76** (2016) 580, arXiv: 1606.01674 [hep-ph] (cit. on p. 8).

[7] S. von Buddenbrock et al., *Constraints on a 2HDM with a singlet scalar and implications in the search for heavy bosons at the LHC*, Journal of Physics G: Nuclear and Particle Physics **46** (2019) 115001, ISSN: 1361-6471, URL: http://dx.doi.org/10.1088/1361-6471/ab3cf6 (cit. on p. 8).

[8] S. Buddenbrock et al., *The emergence of multi-lepton anomalies at the LHC and their compatibility with new physics at the EW scale*, JHEP **10** (2019) 157, arXiv: 1901.05300 [hep-ph] (cit. on p. 9).

[9] S. von Buddenbrock et al., *Multi-lepton signatures of additional scalar bosons beyond the Standard Model at the LHC*, J. Phys. G **45** (2018) 115003, arXiv: 1711.07874 [hep-ph] (cit. on p. 9).

[10] *CERN Yellow Report 4*, https://twiki.cern.ch/twiki/bin/view/LHCPhysics/LHCHWG (cit. on p. 9).

[11] ATLAS Collaboration, *Luminosity determination in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector at the LHC*, ATLAS-CONF-2019-021, 2019, URL: https://cds.cern.ch/record/2677054 (cit. on pp. 10, 54).

[12] G. Avoni et al., *The new LUCID-2 detector for luminosity measurement and monitoring in ATLAS*, Journal of Instrumentation **13** (2018) P07017, URL: https://doi.org/10.1088/1748-0221/13/07/p07017 (cit. on p. 10).

[13] ATLAS Collaboration, *ATLAS data quality operations and performance for 2015–2018 data-taking*, (2019), arXiv: 1911.04632 [hep-ex] (cit. on p. 10).

[14] M. Aaboud et al., *Measurement of the photon identification efficiencies with the ATLAS detector using LHC Run 2 data collected in 2015 and 2016*, The European Physical Journal C **79** (2019), ISSN: 1434-6052, URL: http://dx.doi.org/10.1140/epjc/s10052-019-6650-6 (cit. on p. 10).

[15] ATLAS Internal, *ExtendedPileupReweighting*, 2016, URL: https://twiki.cern.ch/twiki/bin/view/AtlasProtected/ExtendedPileupReweighting (cit. on p. 10).

[16] T. Sjöstrand et al., *An introduction to PYTHIA 8.2*, Comput. Phys. Commun. **191** (2015) 159, arXiv: 1410.3012 [hep-ph] (cit. on p. 11).

[17] ATLAS Collaboration, *Summary of ATLAS Pythia 8 tunes*, ATL-PHYS-PUB-2012-003, 2012, URL: https://cds.cern.ch/record/1474107 (cit. on p. 11).

[18] ATLAS Collaboration, *ATLAS Pythia 8 tunes to* 7 *TeV data*, ATL-PHYS-PUB-2014-021, 2014, URL: https://cds.cern.ch/record/1966419 (cit. on pp. 11, 14).

[19] R. D. Ball et al., *Parton distributions for the LHC Run II*, JHEP **04** (2015) 040, arXiv: 1410.8849 [hep-ph] (cit. on p. 11).

[20] D. J. Lange, *The EvtGen particle decay simulation package*, Nucl. Instrum. Meth. A **462** (2001) 152 (cit. on p. 11).

[21] The ATLAS Collaboration, *The simulation principle and performance of the ATLAS fast calorimeter simulation FastCaloSim*, (2010), ATL-PHYS-PUB-2010-013, URL: http://cdsweb.cern.ch/record/1300517 (cit. on p. 11).

[22] *LHC Higgs Cross Section Working Group*, URL: https://twiki.cern.ch/twiki/bin/view/LHCPhysics/CERNYellowReportPageBR#Higgs_2_fermions (cit. on p. 12).

[23] P.M. Nadolsky, et al., *Implications of CTEQ global analysis for collider observables*, tech. rep., 2008 013004, URL: %7Bhttp://prd.aps.org/abstract/PRD/v78/i1/e013004%7D (cit. on p. 13).

[24] ATLAS Collaboration, *Measurement of the Z/γ^* boson transverse momentum distribution in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector*, JHEP **09** (2014) 145, arXiv: 1406.3660 [hep-ex] (cit. on p. 13).

[25] G. Bozzi, S. Catani, D. de Florian, and M. Grazzini, *Transverse-momentum resummation and the spectrum of the Higgs boson at the LHC*, Nucl. Phys. B **737** (2006) 73, arXiv: hep-ph/0508068 (cit. on p. 13).

[26] K. Hamilton, P. Nason, E. Re, and G. Zanderighi, *NNLOPS simulation of Higgs boson production*, JHEP **10** (2013) 222, arXiv: 1309.0017 [hep-ph] (cit. on p. 13).

[27] J. R. Andersen et al., *Handbook of LHC Higgs Cross Sections: 3. Higgs Properties*, (2013), ed. by S. Heinemeyer, C. Mariotti, G. Passarino, and R. Tanaka, arXiv: 1307.1347 [hep-ph] (cit. on p. 13).

[28] R. D. Ball, V. Bertone, S. Carrazza, C. S. Deans, L. Del Debbio, et al., *Parton distributions with LHC data*, Nucl. Phys. **B867** (2013) 244, arXiv: 1207.1303 [hep-ph] (cit. on p. 14).

[29] I. Nomidis et al., *Event selection, performance and background estimation in the $H \to \gamma\gamma$ channel with Run-2 data*, tech. rep., CERN, 2020, URL: https://cds.cern.ch/record/2718255 (cit. on pp. 17, 18, 20, 245).

[30] ATLAS Collaboration, *Jet reconstruction and performance using particle flow with the ATLAS Detector*, Eur. Phys. J. C **77** (2017) 466, arXiv: 1703.10485 [hep-ex] (cit. on p. 17).

[31] J. Adelman et al., *Measurement of fiducial and differential cross sections in the $H \to \gamma\gamma$ decay channel with 140fb^{-1} of 13 TeV proton–proton collision data with the ATLAS detector*, tech. rep. ATL-COM-PHYS-2019-035, CERN, 2019, URL: https://cds.cern.ch/record/2654897 (cit. on p. 17).

[32] A. Collaboration, *ATLAS Flavor-Tagging Calibration Results with 139 ifb*, 2019, URL: http://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PLOTS/FTAG-2019-004/ (cit. on p. 18).

[33] A. Collaboration, *Expected performance of the 2019 ATLAS b-taggers*, 2019, URL: http://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PLOTS/FTAG-2019-005/ (cit. on p. 18).

[34] ATLAS Collaboration, *Supporting Note: Selection and performance for the $H \to \gamma\gamma$ and $H \to Z\gamma$ analyses*, (2015), URL: https://cds.cern.ch/record/2196102 (cit. on p. 18).

[35] M. Aaboud et al., *Search for Higgs boson pair production in the $\gamma\gamma WW^*$ channel using pp collision data recorded at $\sqrt{s} = 13$ TeV with the ATLAS detector*, The European Physical Journal C **78** (2018), ISSN: 1434-6052, URL: http://dx.doi.org/10.1140/epjc/s10052-018-6457-x (cit. on pp. 21, 46).

[36] G. Aad et al., *Searches for Higgs boson pair production in the $hh \to bb\tau\tau, \gamma\gamma WW^*, \gamma\gamma bb, bbbb$ channels with the ATLAS detector*, Phys. Rev. **D92** (2015) 092004, arXiv: 1509.04670 [hep-ex] (cit. on p. 21).

[37] A. Hoecker et al., *TMVA - Toolkit for Multivariate Data Analysis*, 2007, arXiv: physics/0703039 [physics.data-an] (cit. on p. 27).

[38] A. Hoecker, P. Speckmayer, J. Stelzer, J. Therhaag, E. von Toerne, and H. Voss, "TMVA 4 (Toolkit for Multivariate Data Analysis with ROOT) Users Guide," 2009 (cit. on p. 27).

[39] P. Baldi, K. Cranmer, T. Faucett, P. Sadowski, and D. Whiteson, *Parameterized neural networks for high-energy physics*, Eur. Phys. J. C **76** (2016) 235, arXiv: 1601.07913 [hep-ex] (cit. on p. 28).

[40] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, *Asymptotic formulae for likelihood-based tests of new physics*, Eur. Phys. J. C **71** (2011) 1554 (cit. on p. 43).

[41] G. Aad et al., *Search for Higgs boson pair production in the two bottom quarks plus two photons final state in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, Phys. Rev. D **106** (5 2022) 052001, URL: https://link.aps.org/doi/10.1103/PhysRevD.106.052001 (cit. on p. 44).

[42] A. L. Read, *Presentation of search results: The CL(s) technique*, J.Phys. **G28** (2002) 2693 (cit. on p. 60).

[43] G. Cowan et al., *Asymptotic formulae for likelihood-based tests of new physics*, Eur. Phys. J. **C71** (2011) 1554, arXiv: 1007.1727 [physics.data-an] (cit. on p. 60).

Appendices

A Limits, pulls, rankings for all mass points

The 20 mass points are put here to show the stucture of the whole results be presented.

990 ### A.1 (m_X, m_S)=(300, 170) GeV

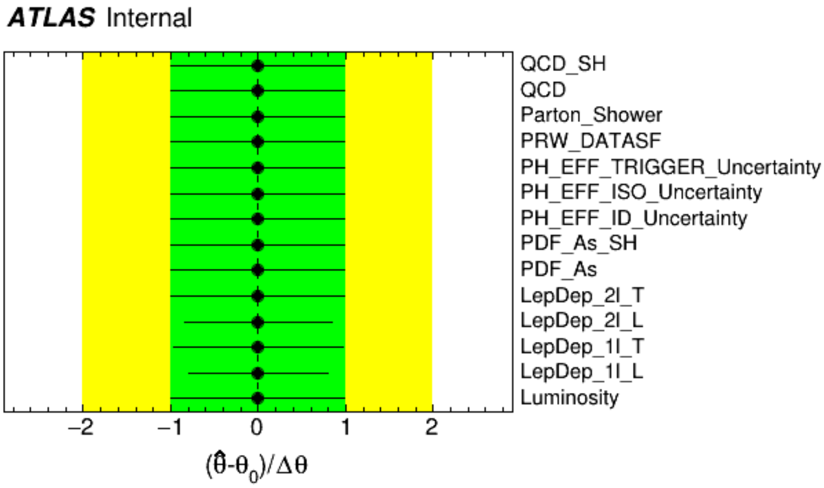


Figure 35: Nuisance parameters pull plot for (m_X, m_S)=(300,170) GeV mass

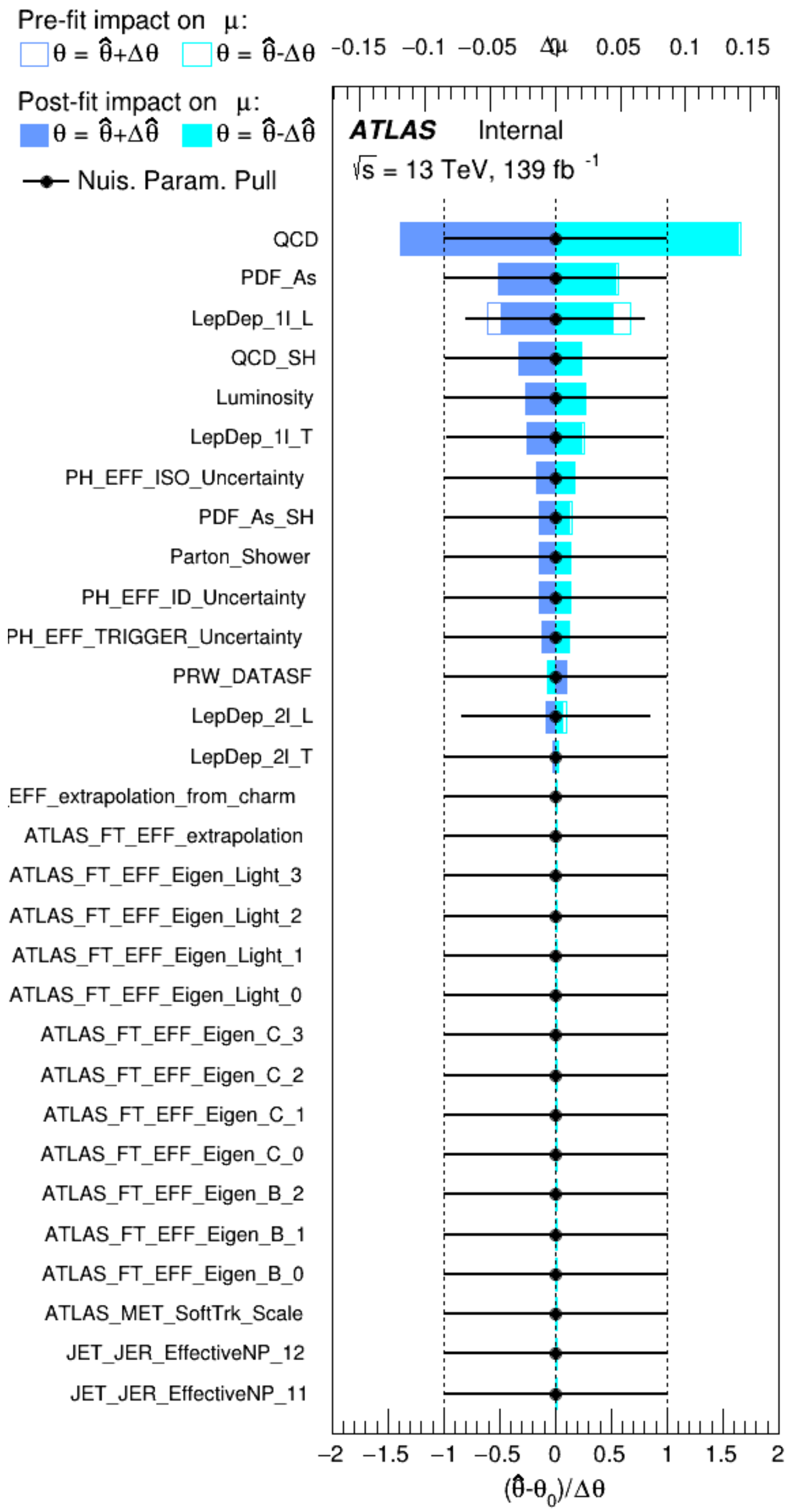


Figure 36: Nuisance parameters rankings for (m_X, m_S)=(300,170) GeV mass

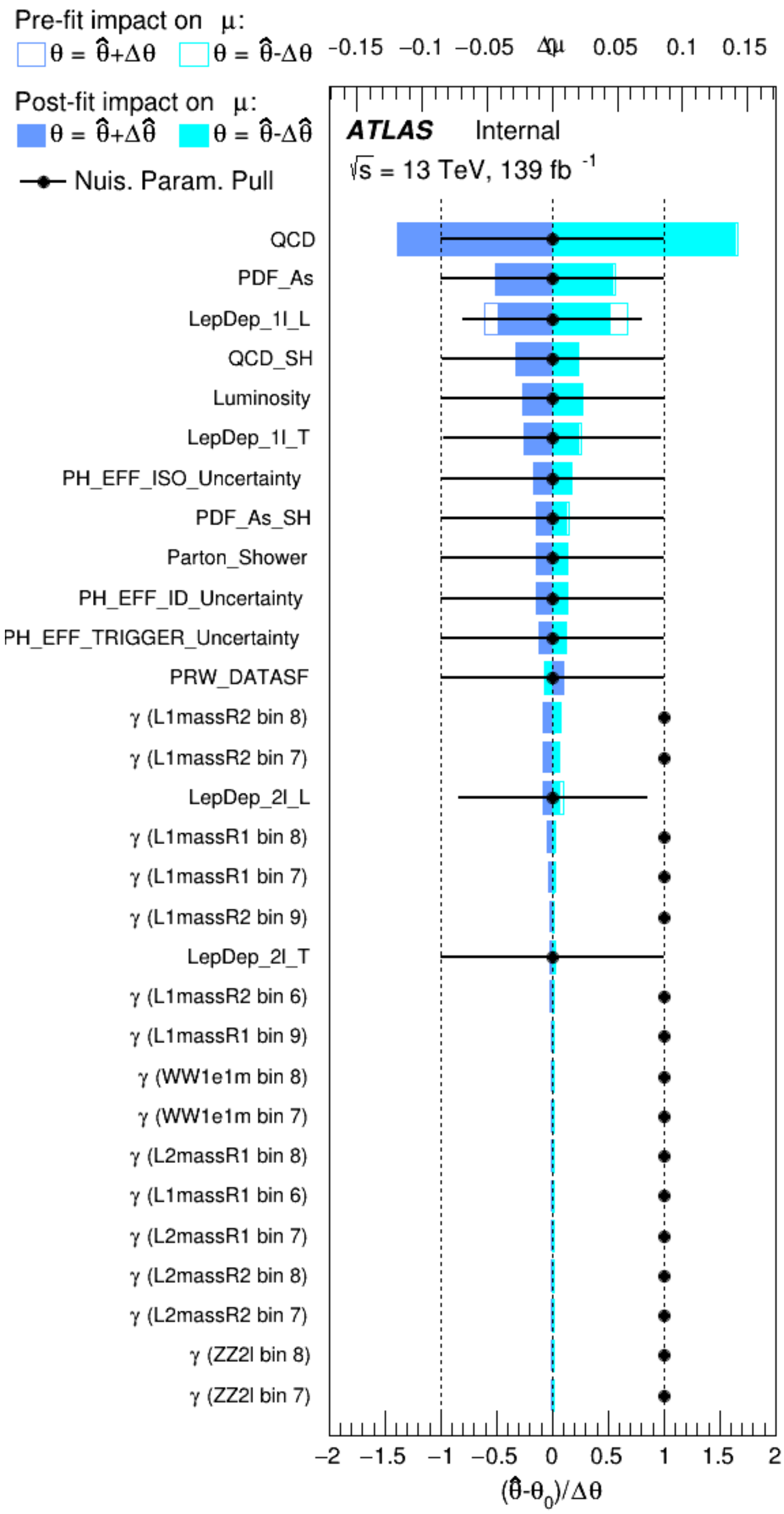


Figure 37: Rankings including systematics and gammas for (m_X, m_S)=(300,170) GeV mass

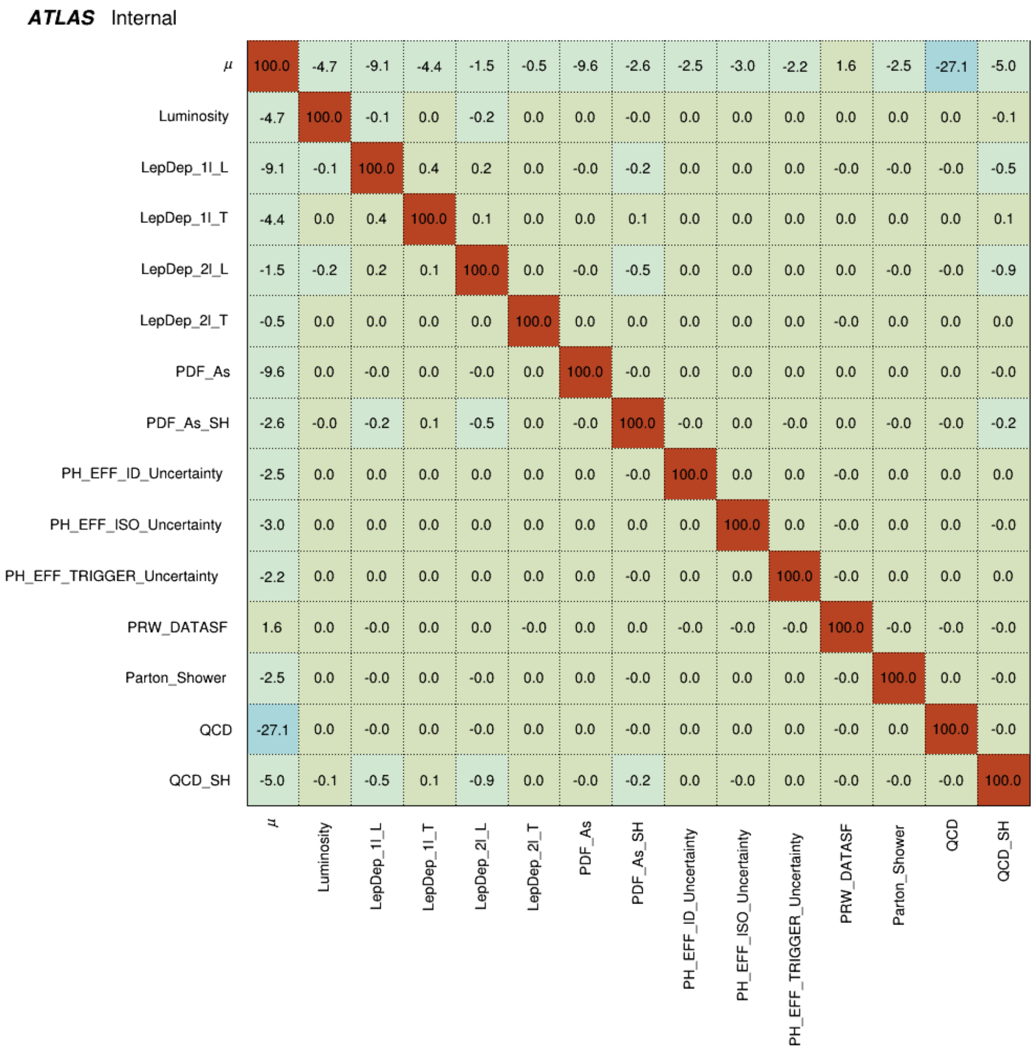


Figure 38: Correlation Matrix for (m_X, m_S)=(300,170) GeV mass.

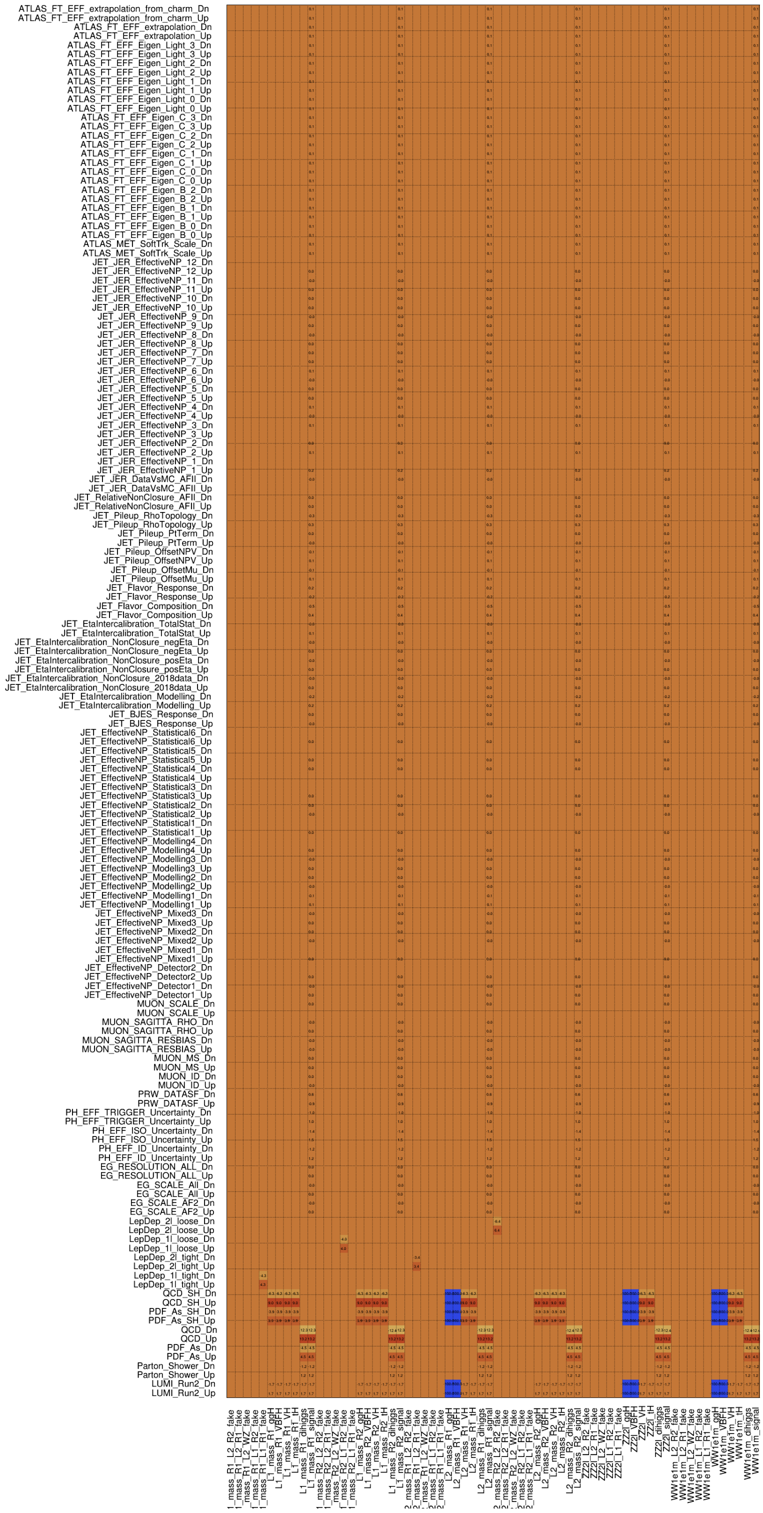


Figure 39: Normalization Factors for Nuisance parameters among all the regions for (m_X, m_S)=(300,170) GeV mass.

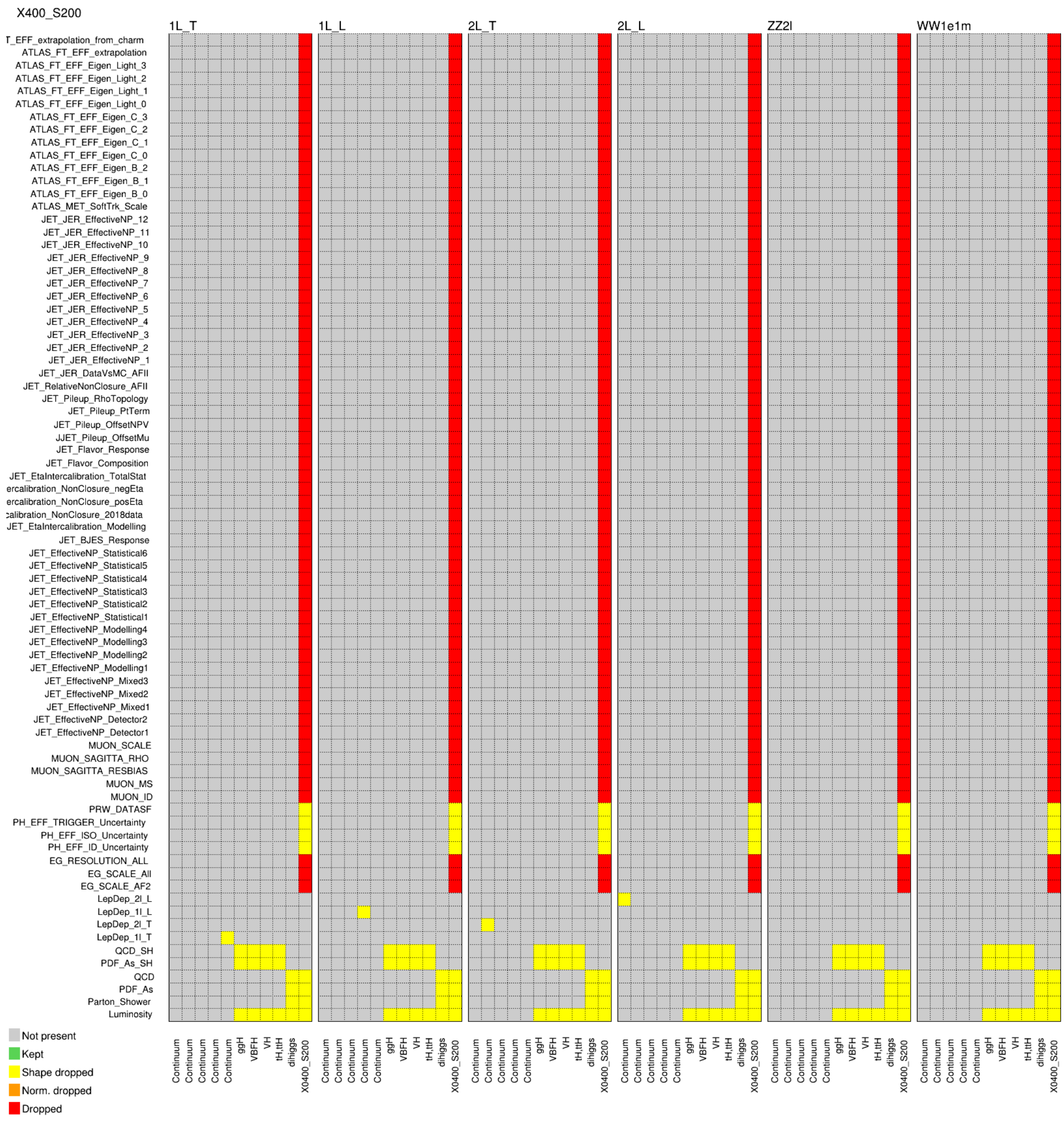


Figure 40: Pruning situation among all the regions for (m_X, m_S)=(300,170) GeV mass.

991 A.2 (m_X, m_S)=(400, 170) GeV

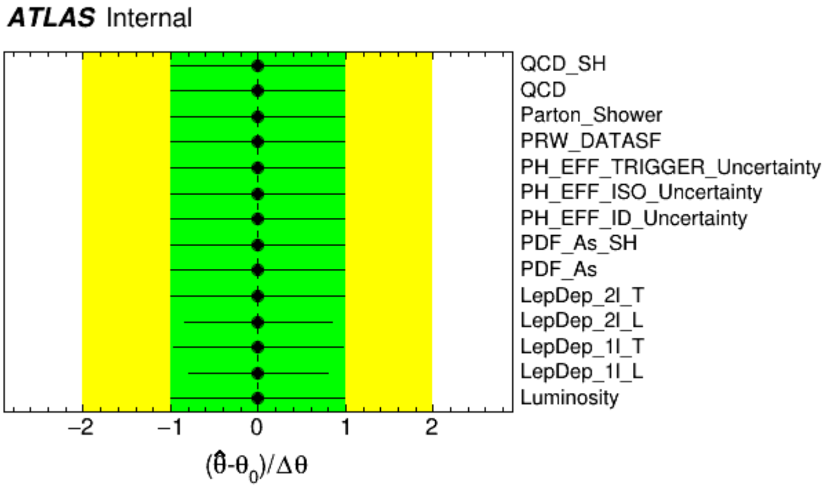


Figure 41: Nuisance parameters pull plot for (m_X, m_S)=(400,170) GeV mass

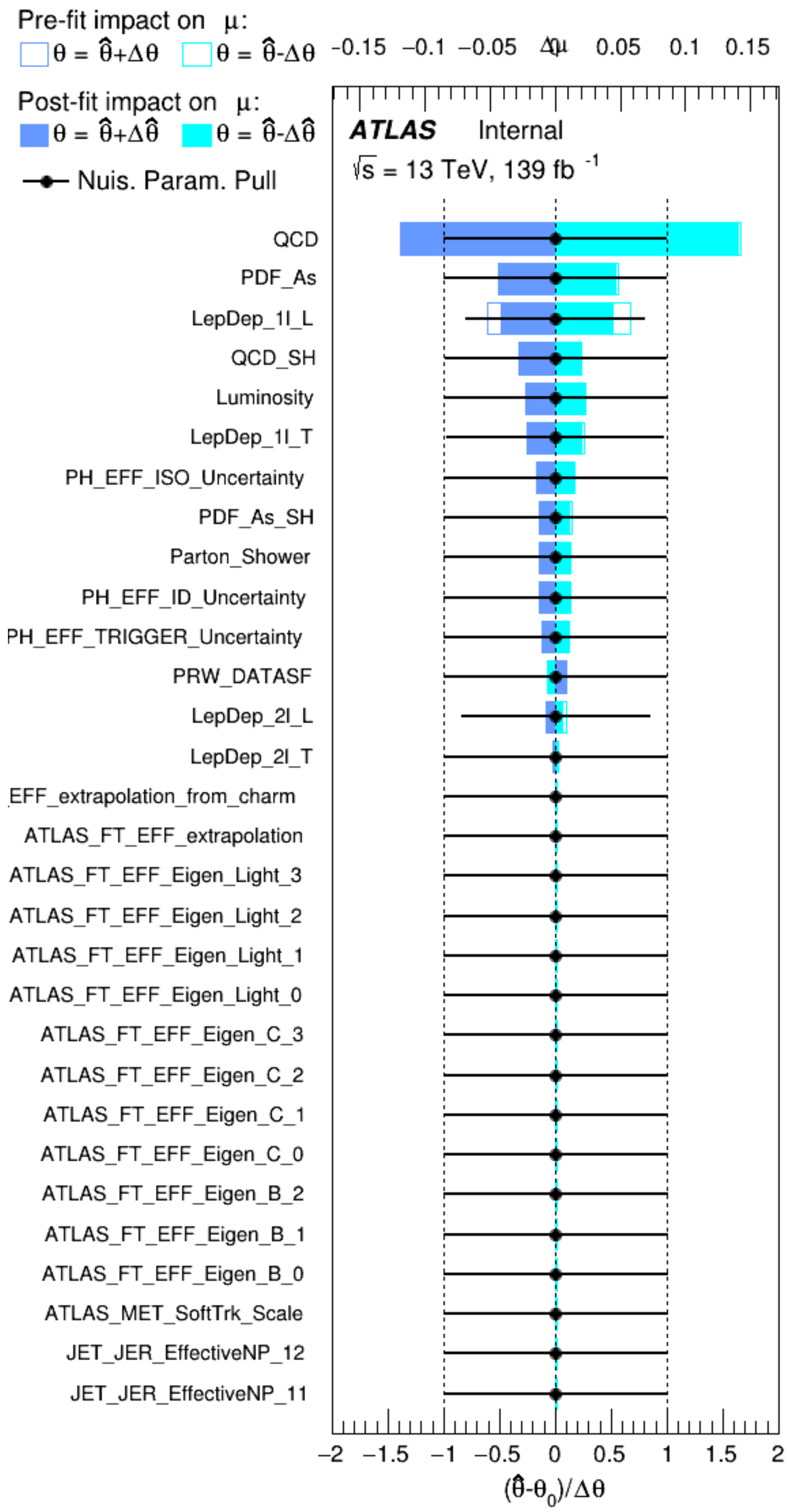


Figure 42: Nuisance parameters rankings for (m_X, m_S)=(400,170) GeV mass

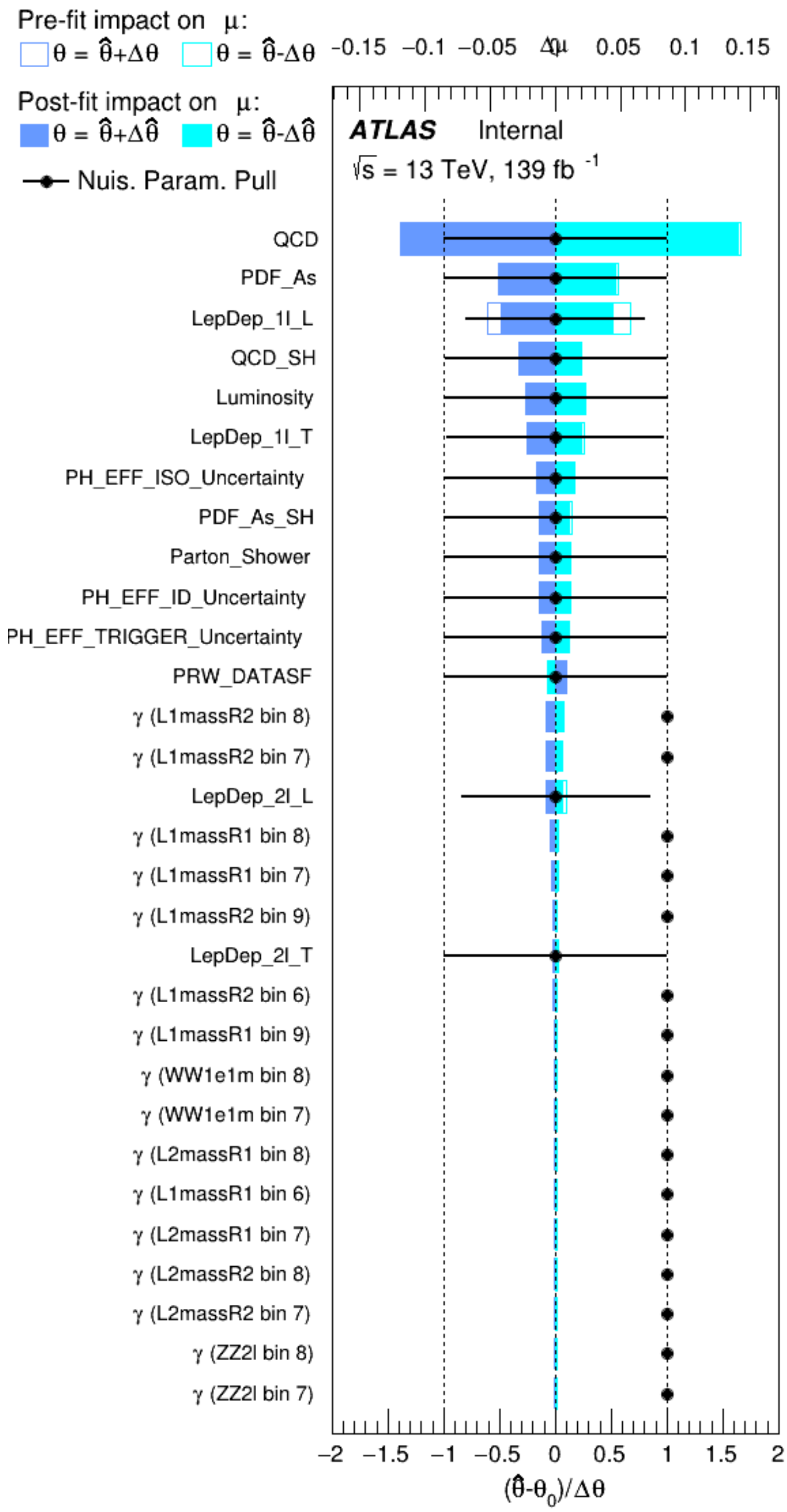


Figure 43: Rankings including systematics and gammas for (m_X, m_S)=(400,170) GeV mass

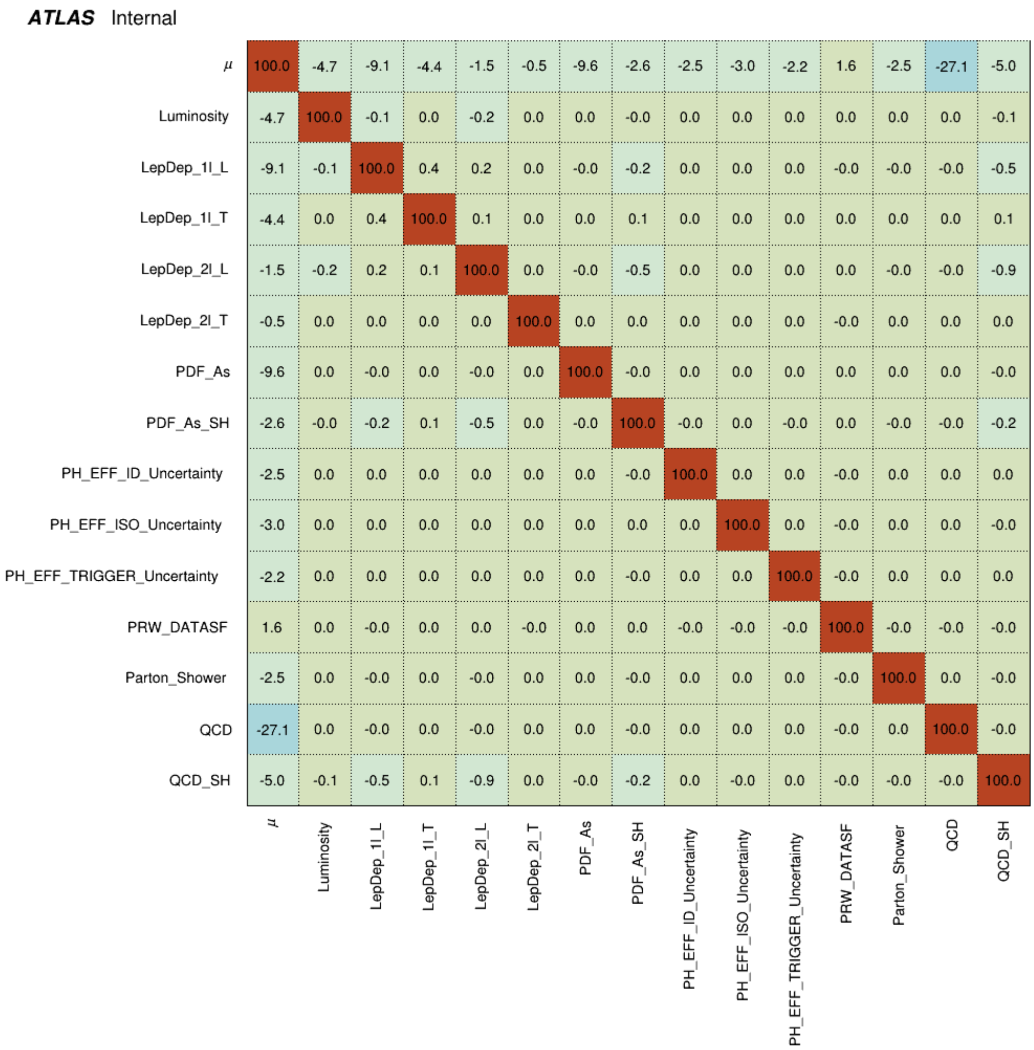


Figure 44: Correlation Matrix for (m_X, m_S)=(400,170) GeV mass.

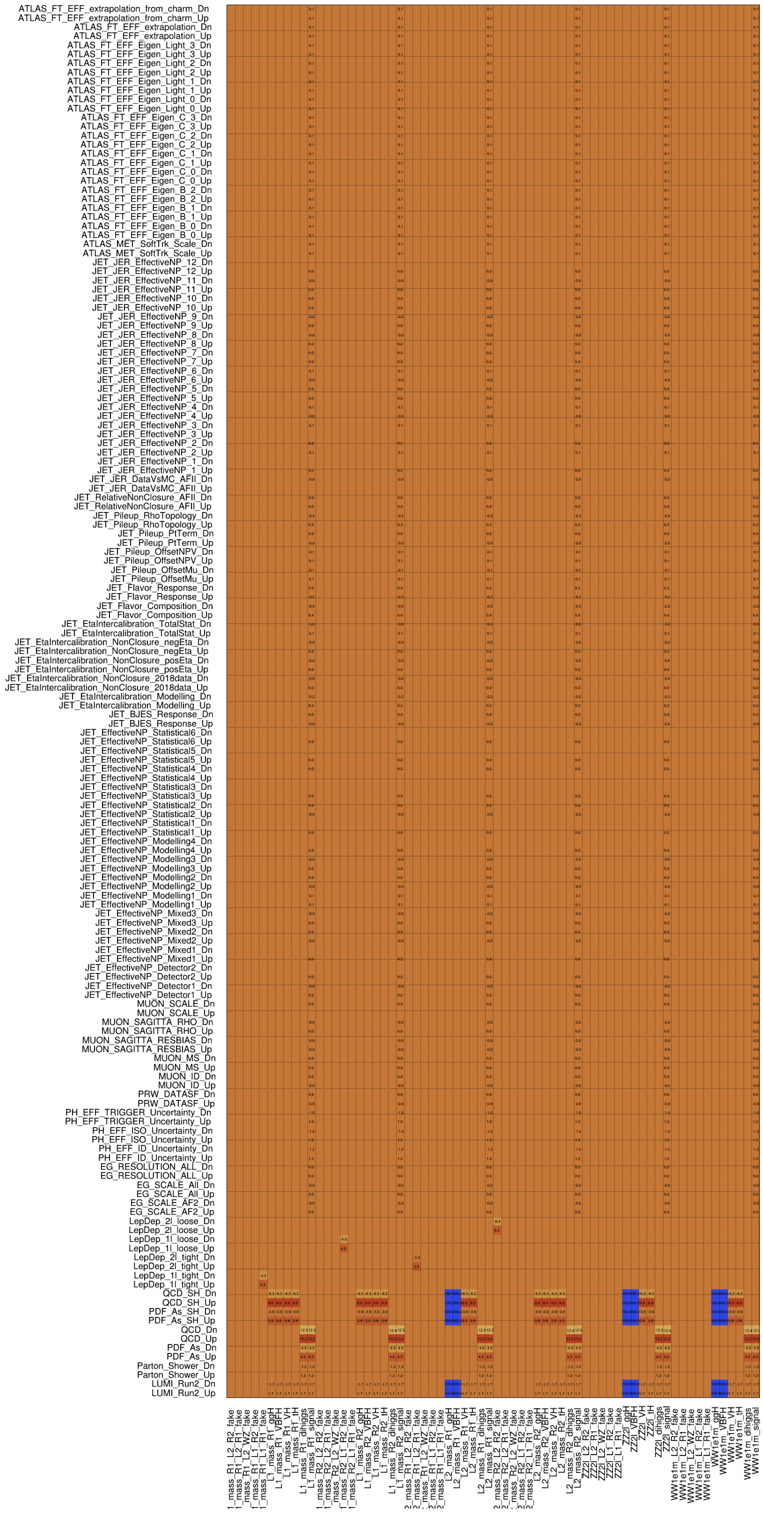


Figure 45: Normalization Factors for Nuisance parameters among all the regions for (m_X, m_S)=(400,170) GeV mass.

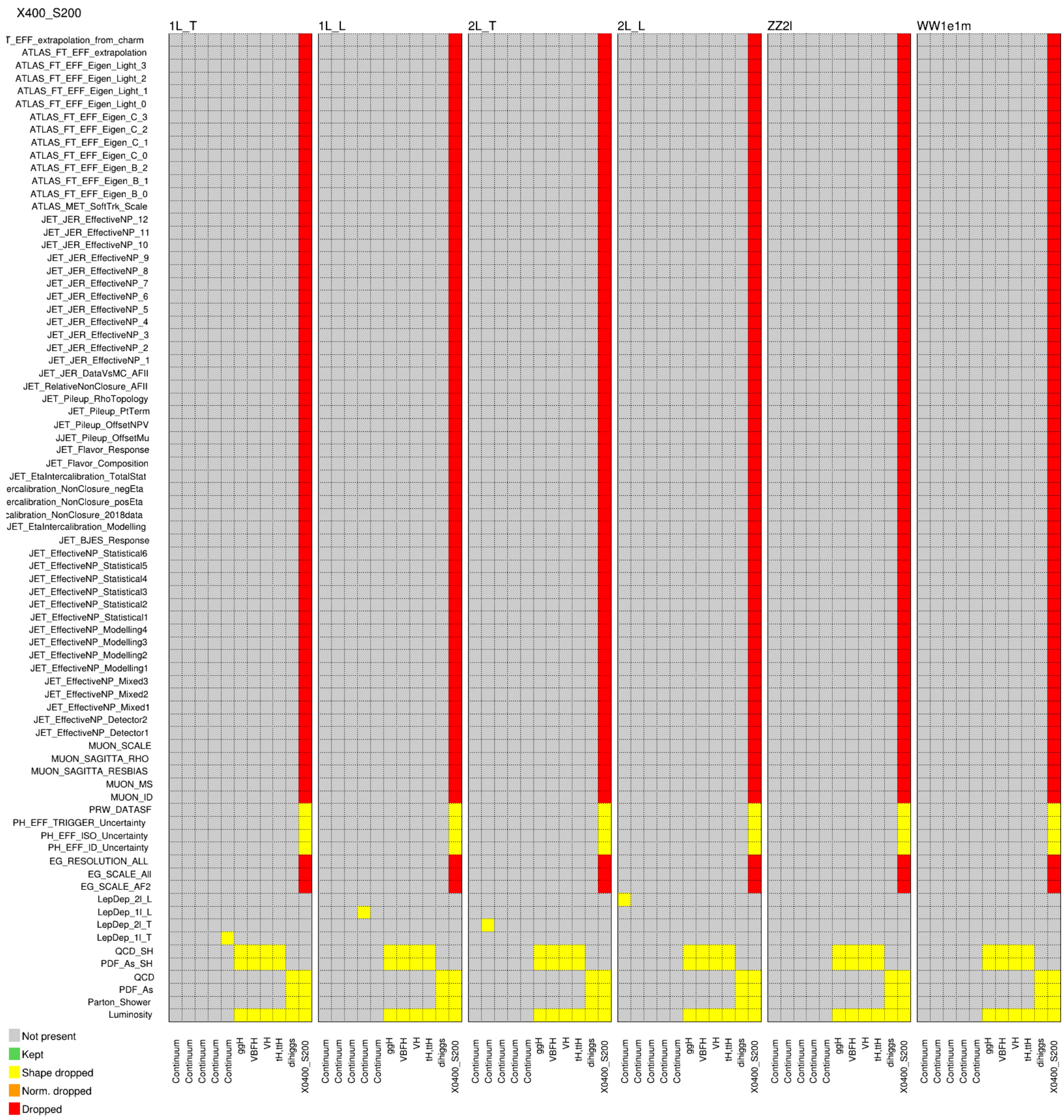


Figure 46: Pruning situation among all the regions for (m_X, m_S)=(400,170) GeV mass.

992 A.3 (m_X, m_S)=(400, 200) GeV

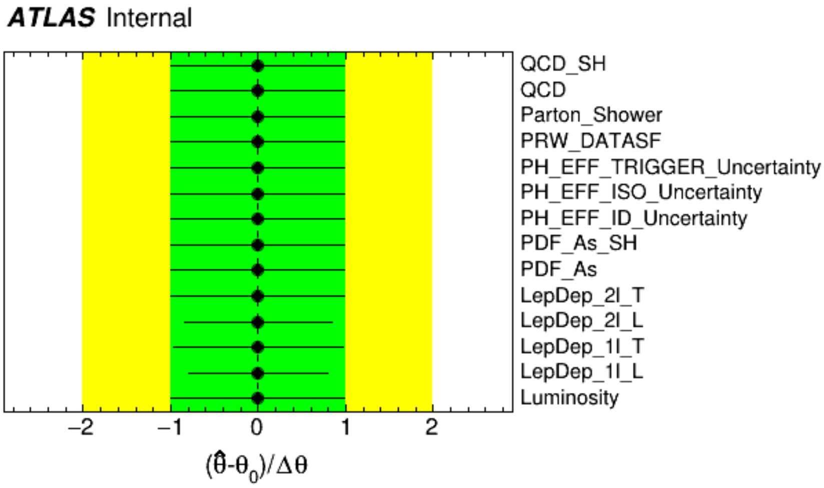


Figure 47: Nuisance parameters pull plot for (m_X, m_S)=(400,200) GeV mass

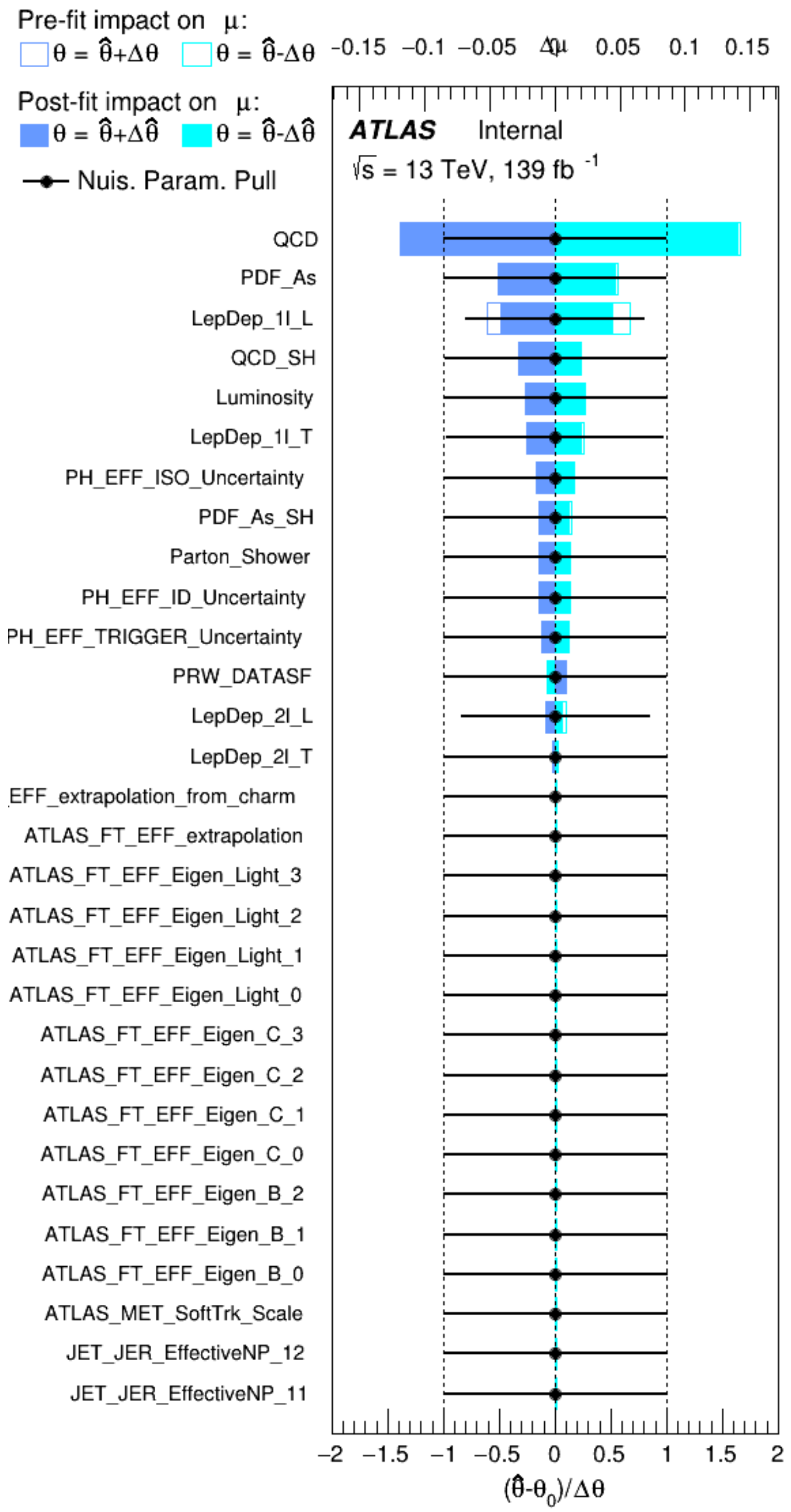


Figure 48: Nuisance parameters rankings for (m_X, m_S)=(400,200) GeV mass

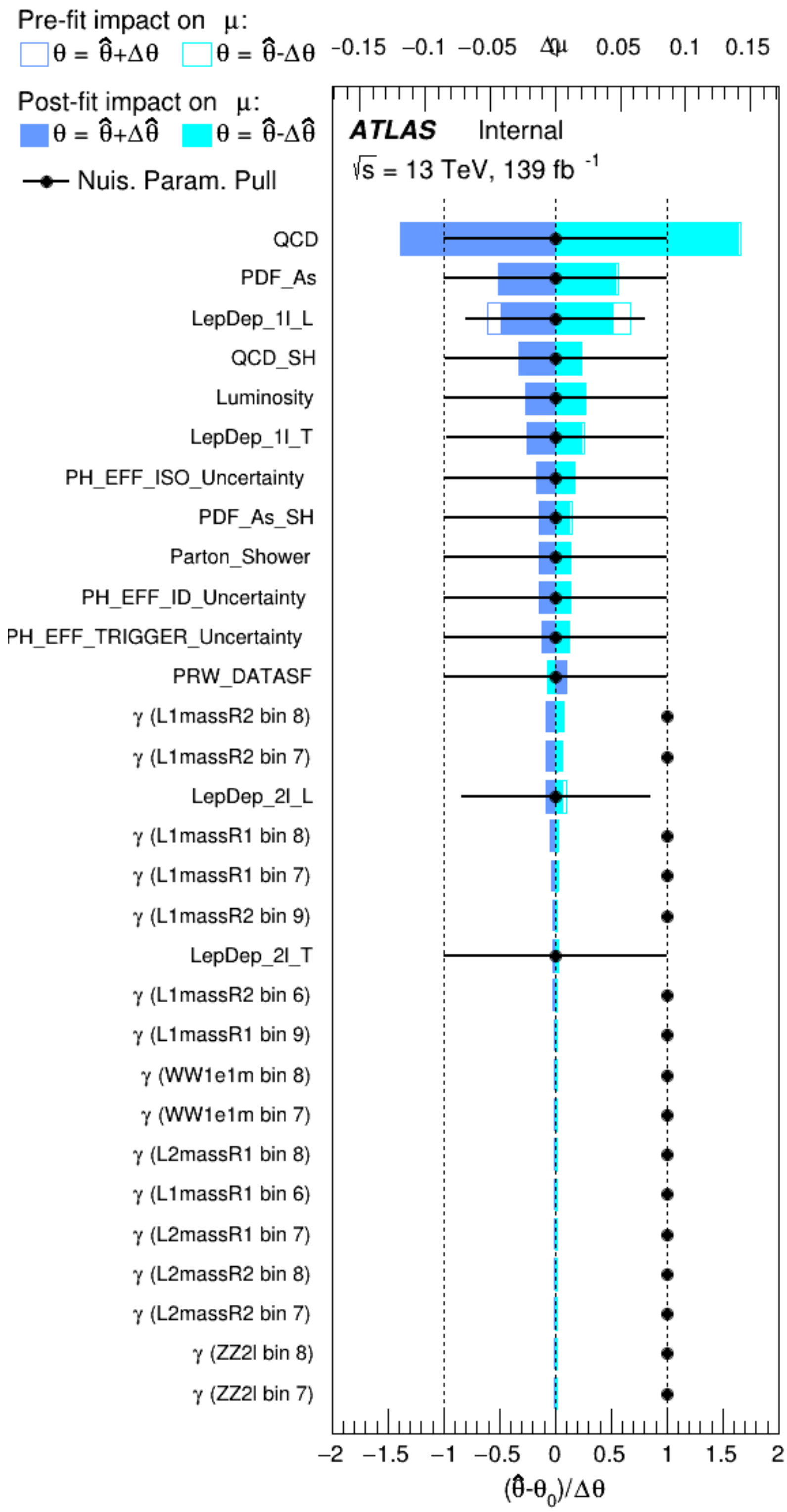


Figure 49: Rankings including systematics and gammas for (m_X, m_S)=(400,200) GeV mass

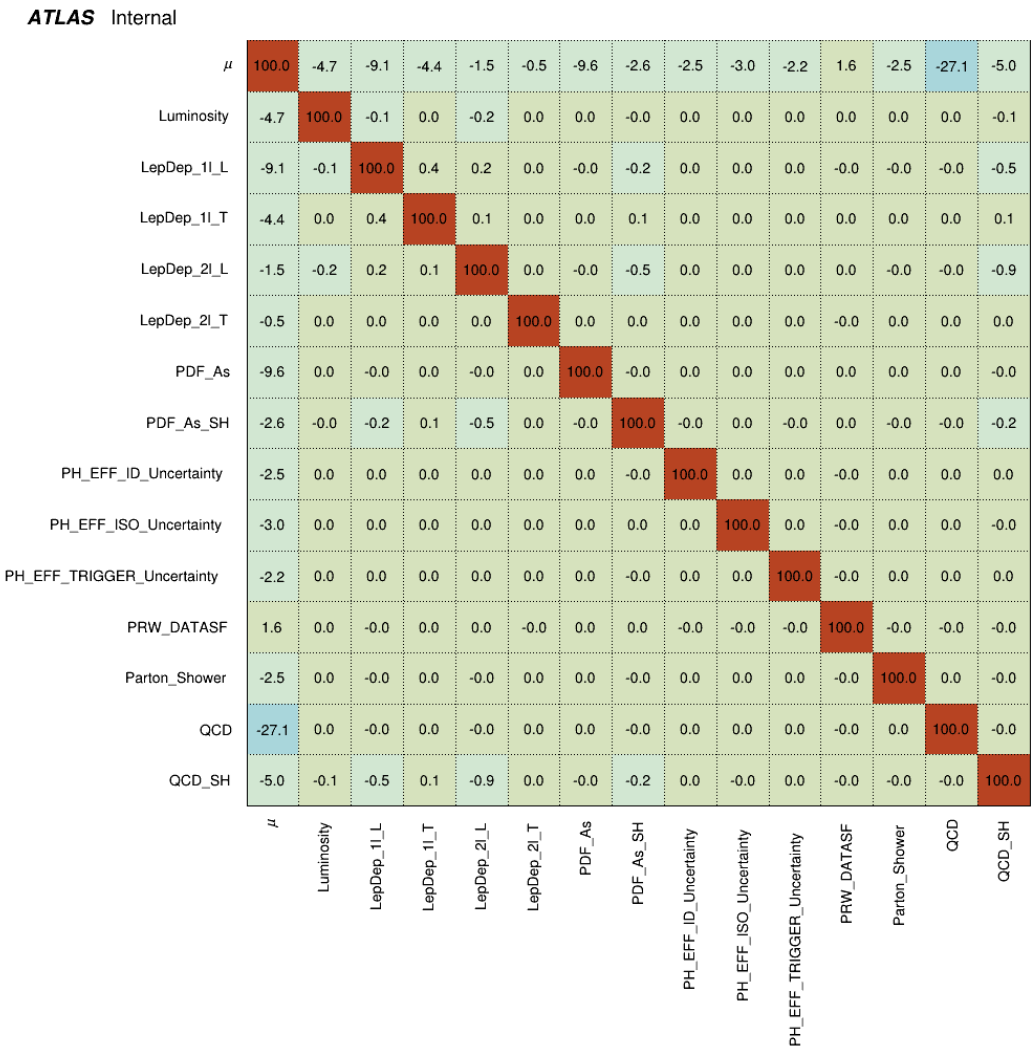


Figure 50: Correlation Matrix for (m_X, m_S)=(400,200) GeV mass.

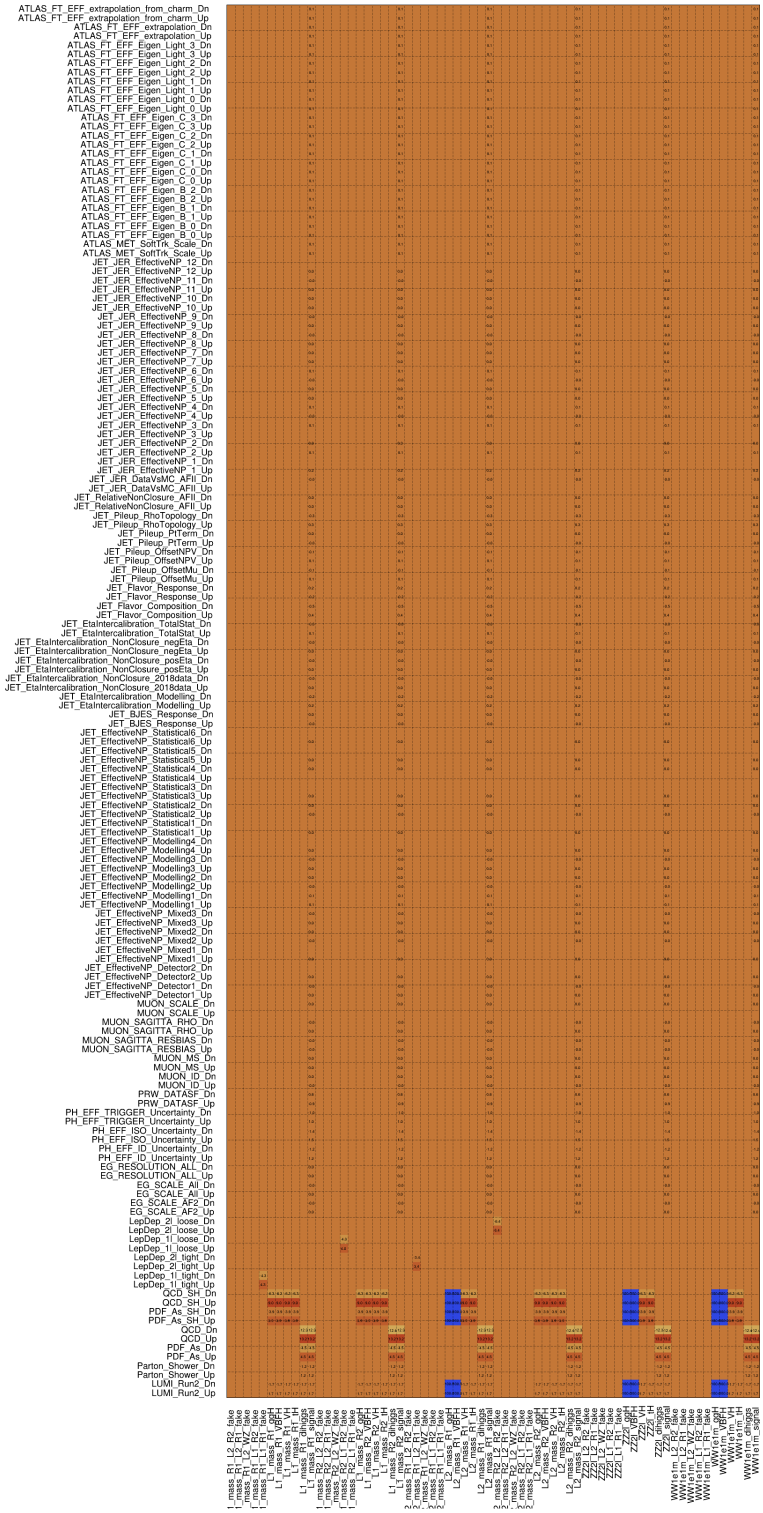


Figure 51: Normalization Factors for Nuisance parameters among all the regions for (m_X, m_S)=(400,200) GeV mass.

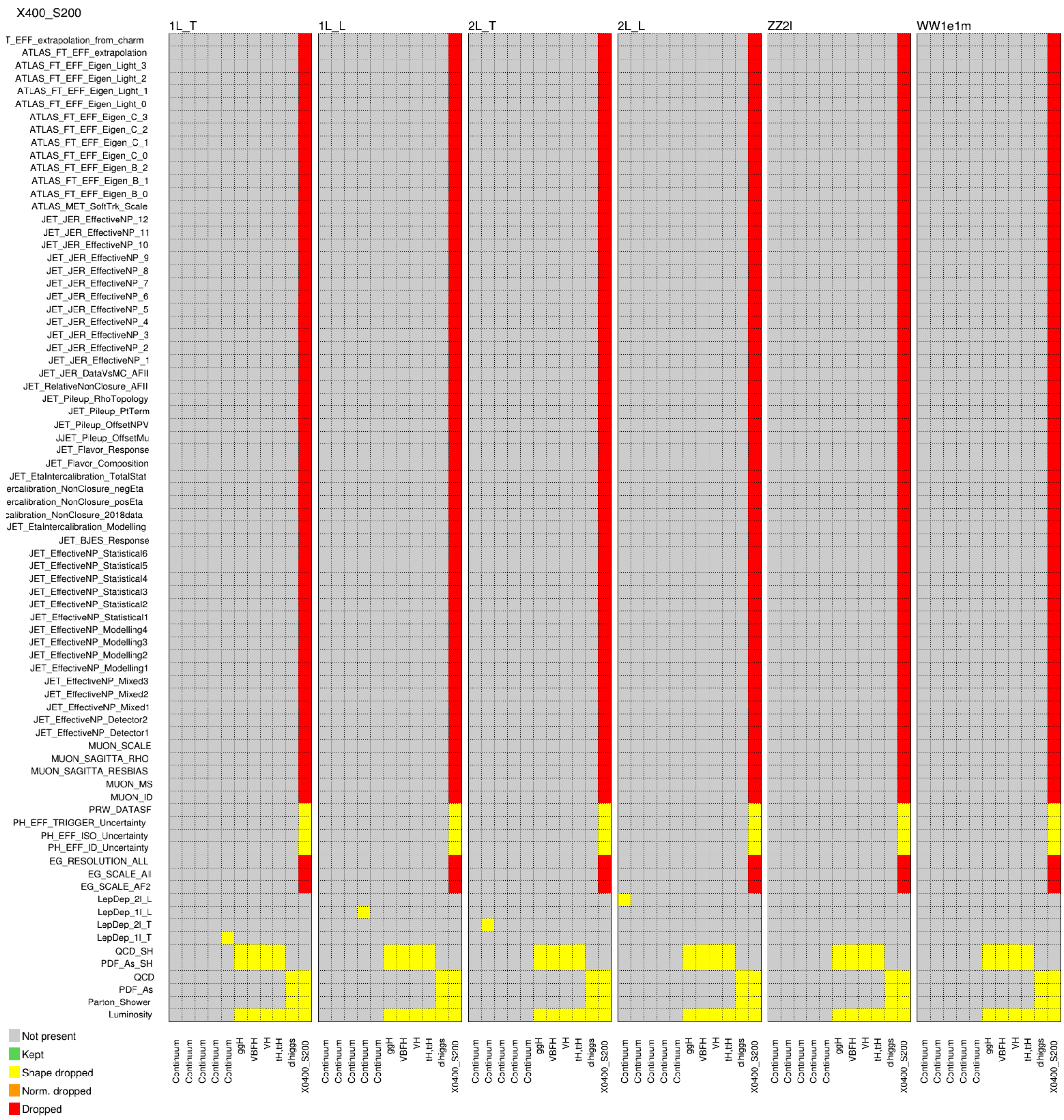


Figure 52: Pruning situation among all the regions for (m_X, m_S)=(400,200) GeV mass.

993 ## A.4 (m_X, m_S)=(500, 170) GeV

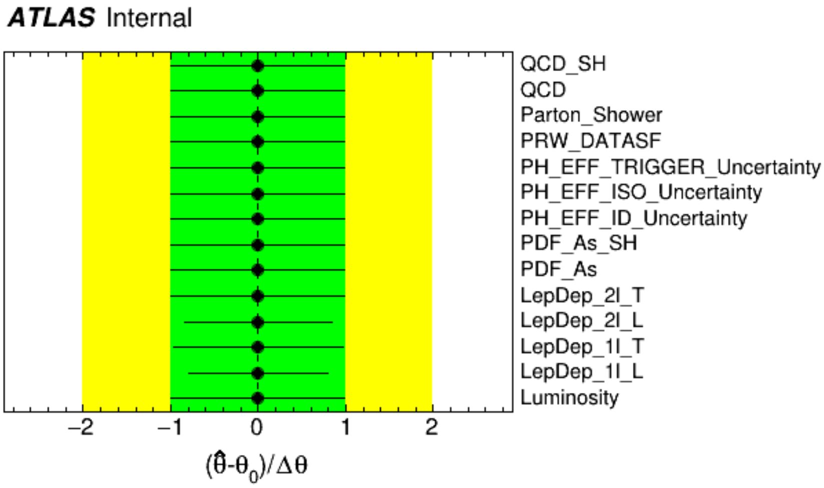


Figure 53: Nuisance parameters pull plot for (m_X, m_S)=(500,170) GeV mass

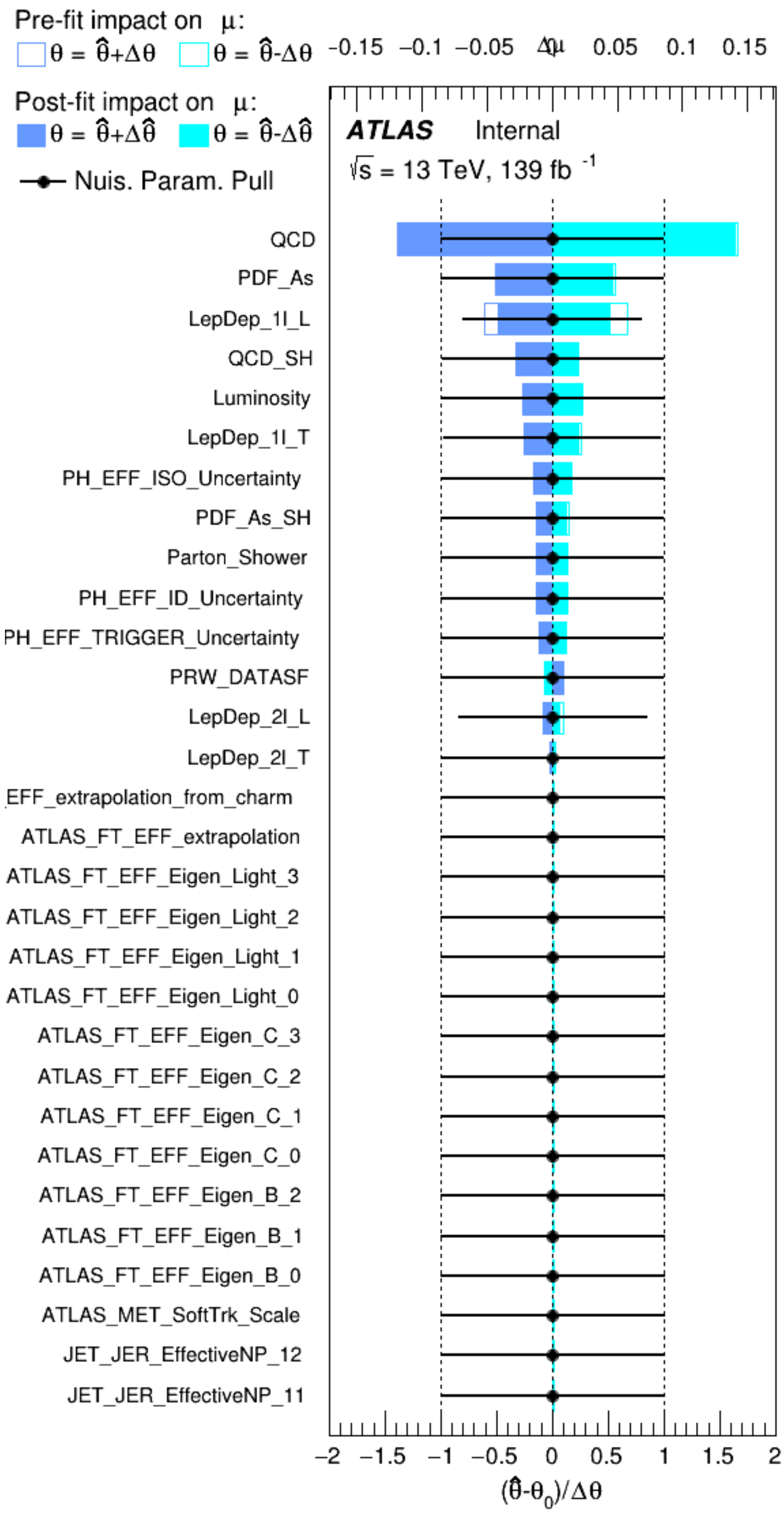


Figure 54: Nuisance parameters rankings for (m_X, m_S)=(500,170) GeV mass

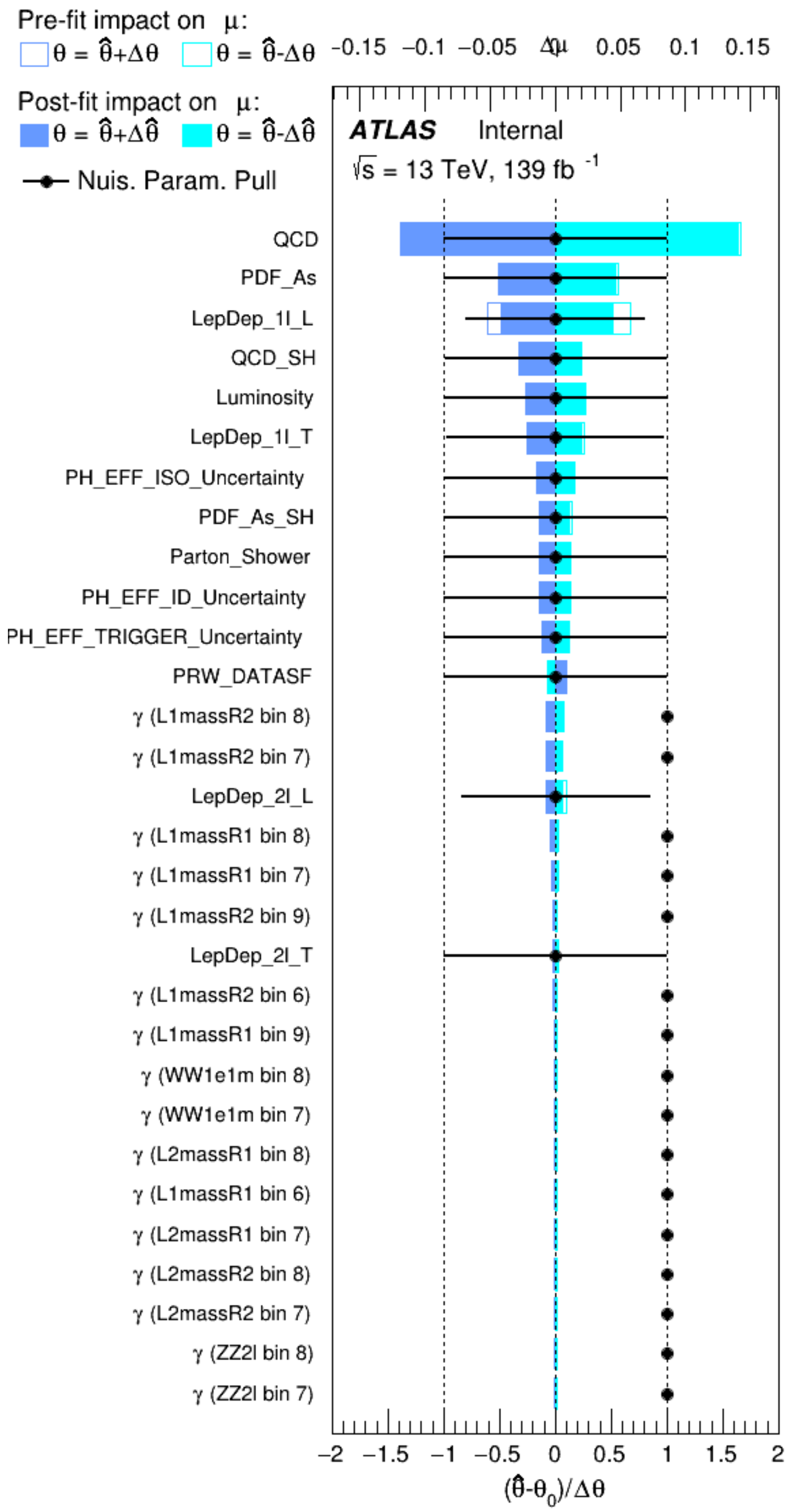


Figure 55: Rankings including systematics and gammas for (m_X, m_S)=(500,170) GeV mass

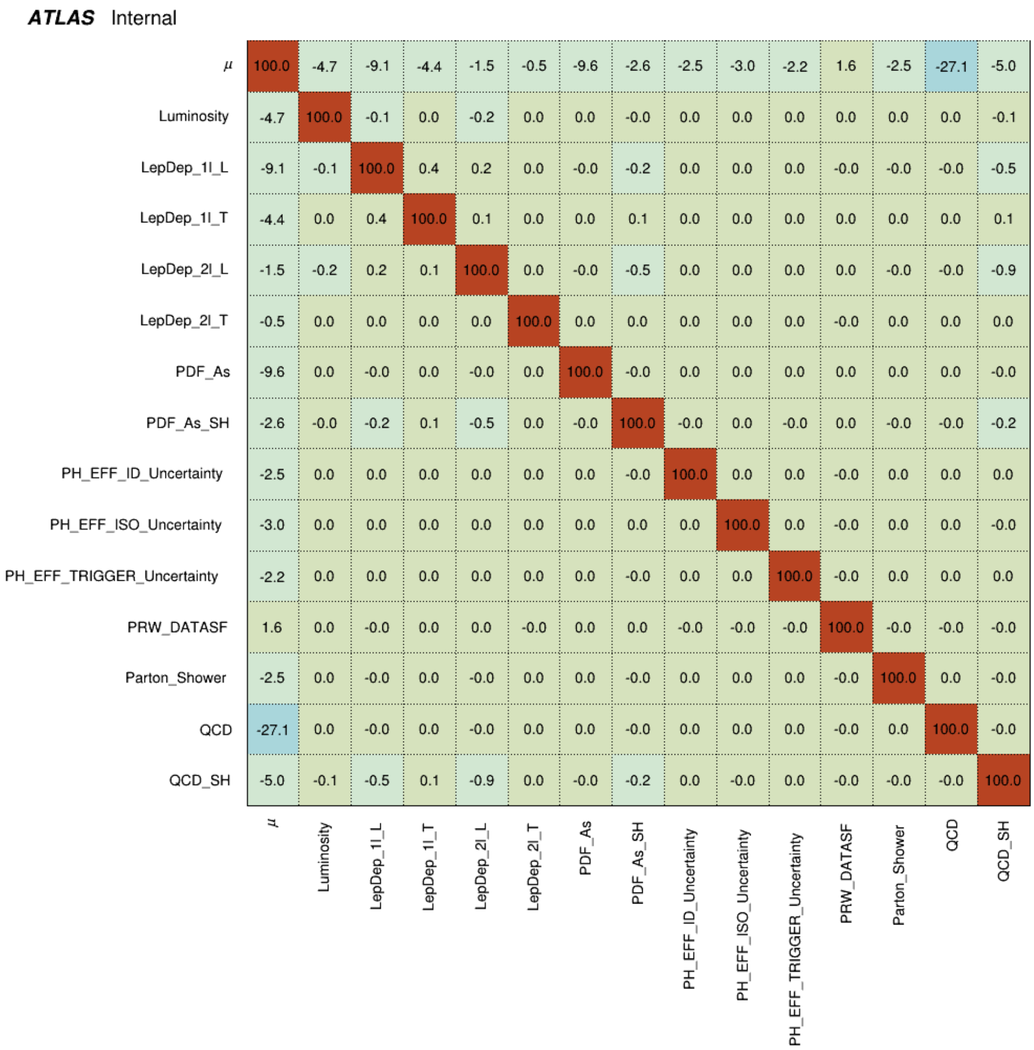


Figure 56: Correlation Matrix for (m_X, m_S)=(500,170) GeV mass.

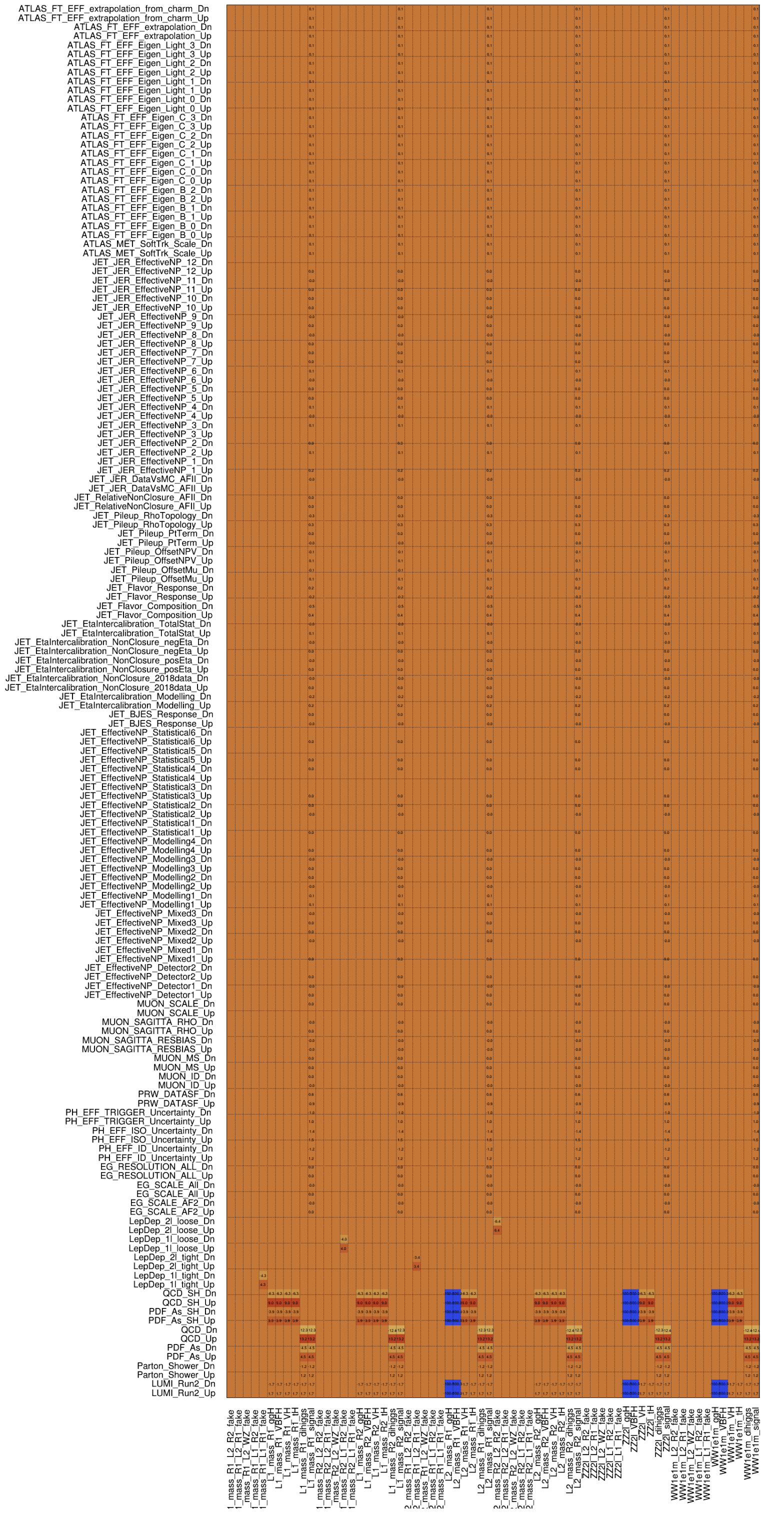


Figure 57: Normalization Factors for Nuisance parameters among all the regions for (m_X, m_S)=(500,170) GeV mass.

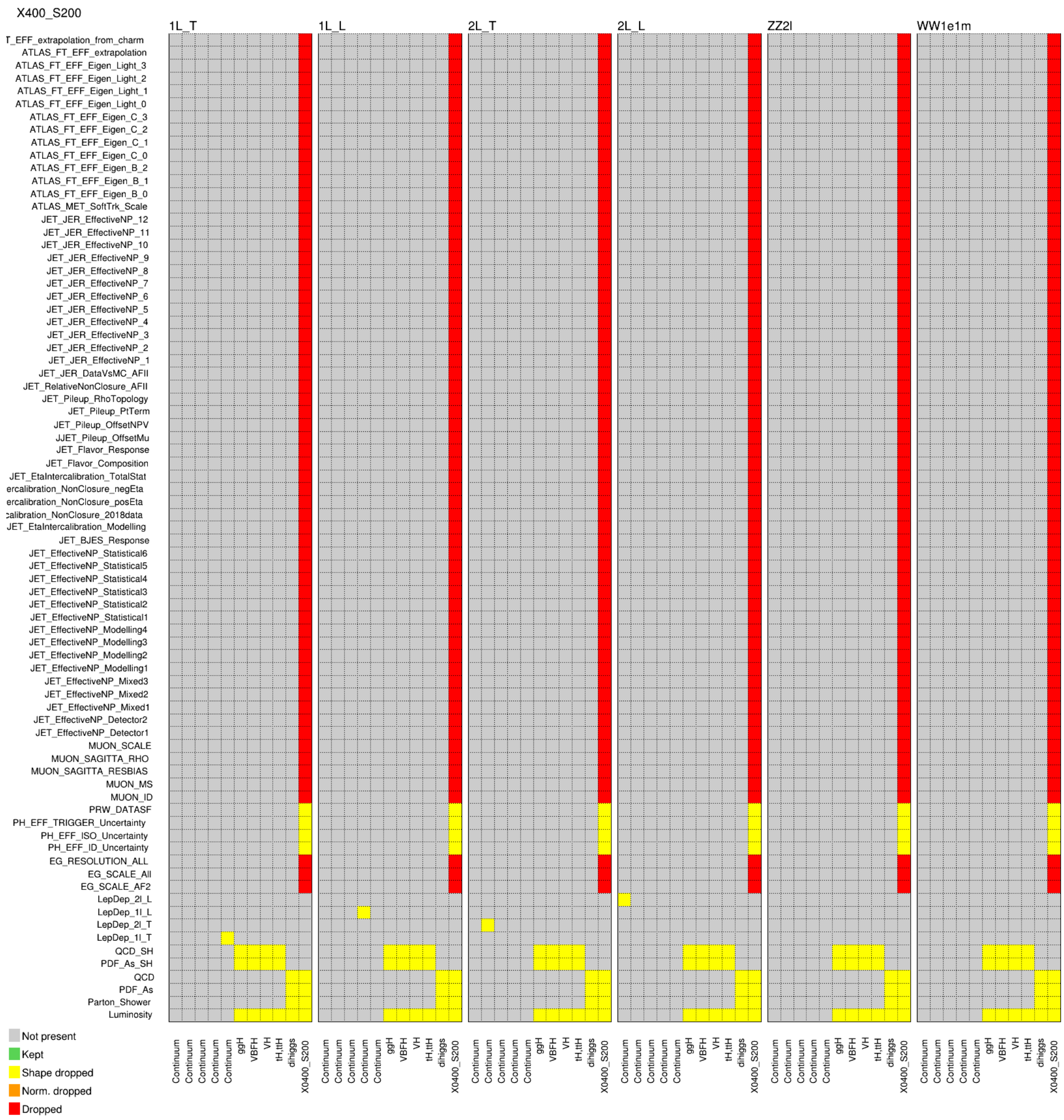


Figure 58: Pruning situation among all the regions for (m_X, m_S)=(500,170) GeV mass.

994 ## A.5 (m_X, m_S)=(500, 200) GeV

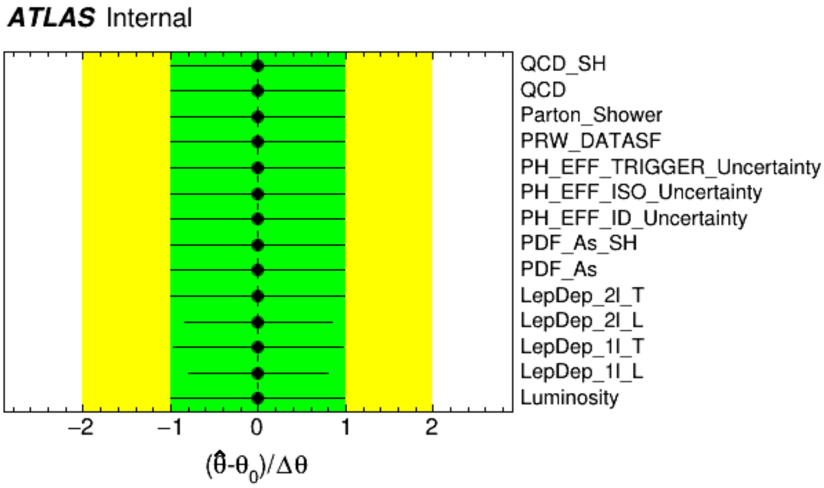


Figure 59: Nuisance parameters pull plot for (m_X, m_S)=(500,200) GeV mass

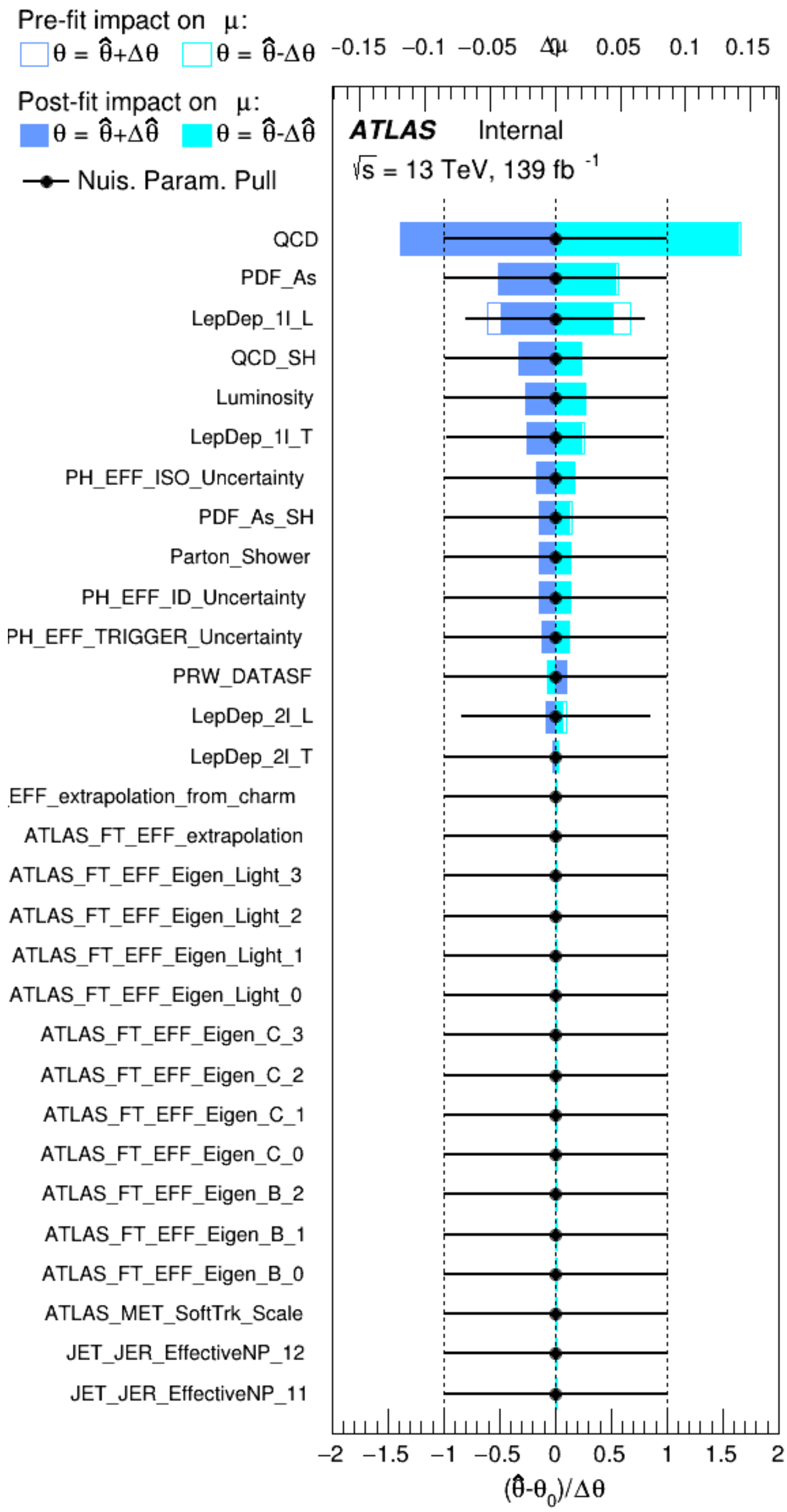


Figure 60: Nuisance parameters rankings for (m_X, m_S)=(500,200) GeV mass

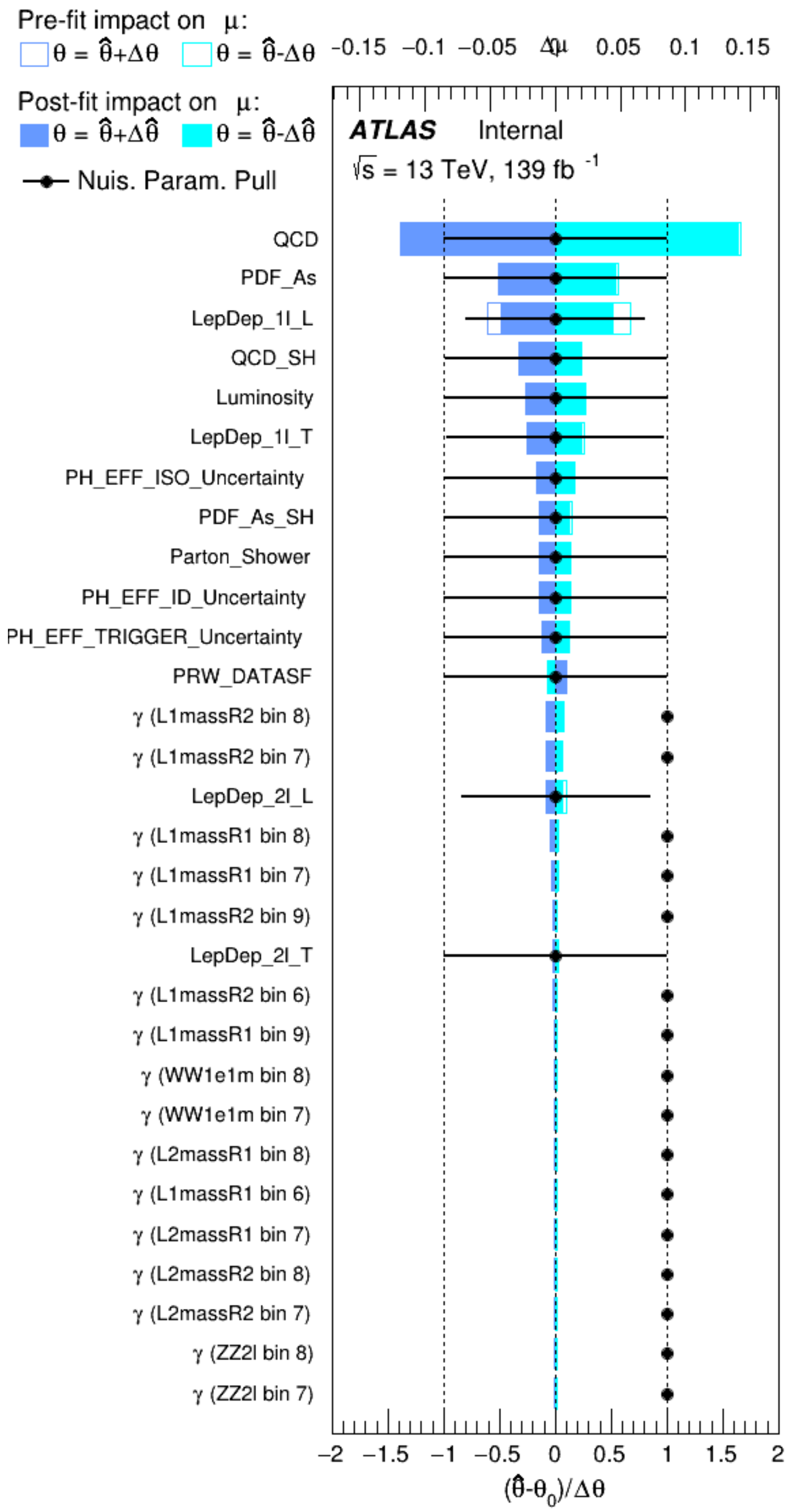


Figure 61: Rankings including systematics and gammas for (m_X, m_S)=(500,200) GeV mass

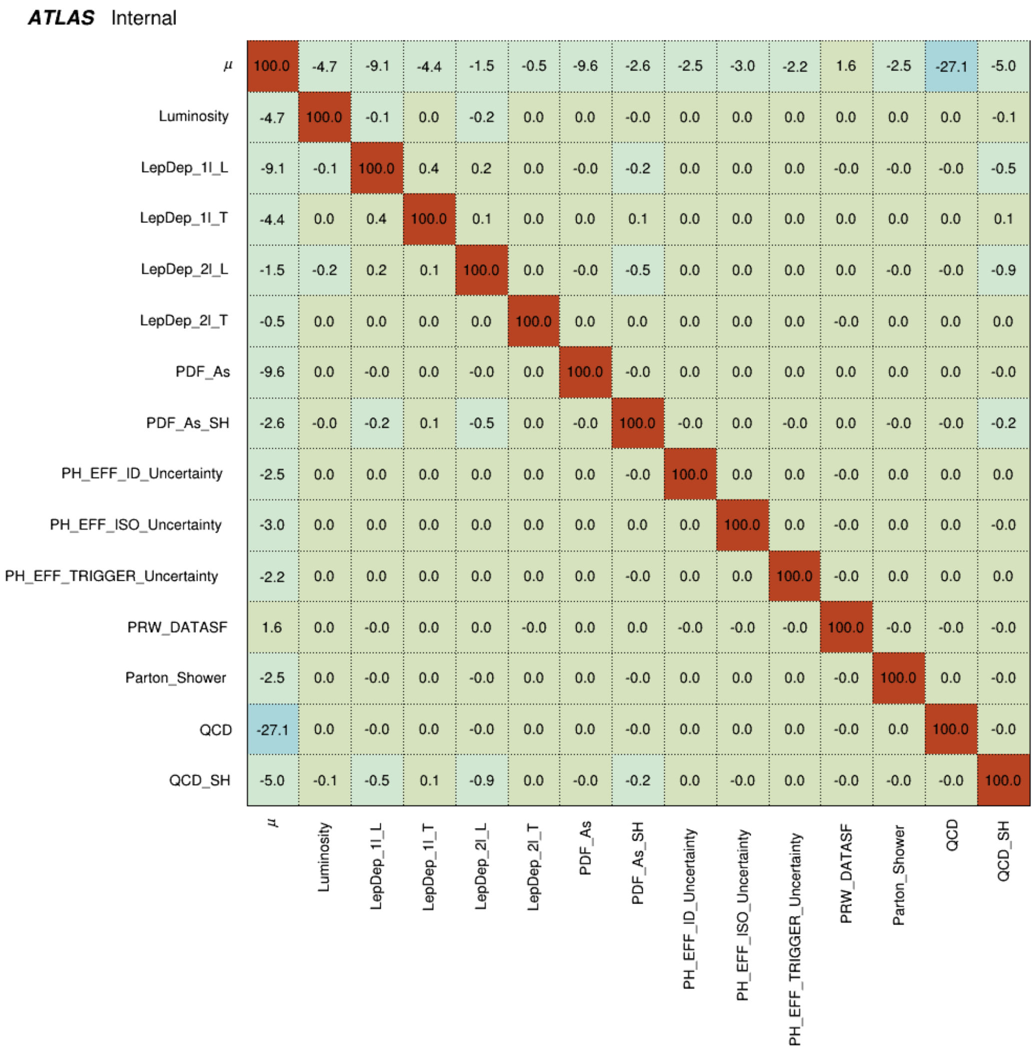


Figure 62: Correlation Matrix for (m_X, m_S)=(500,200) GeV mass.

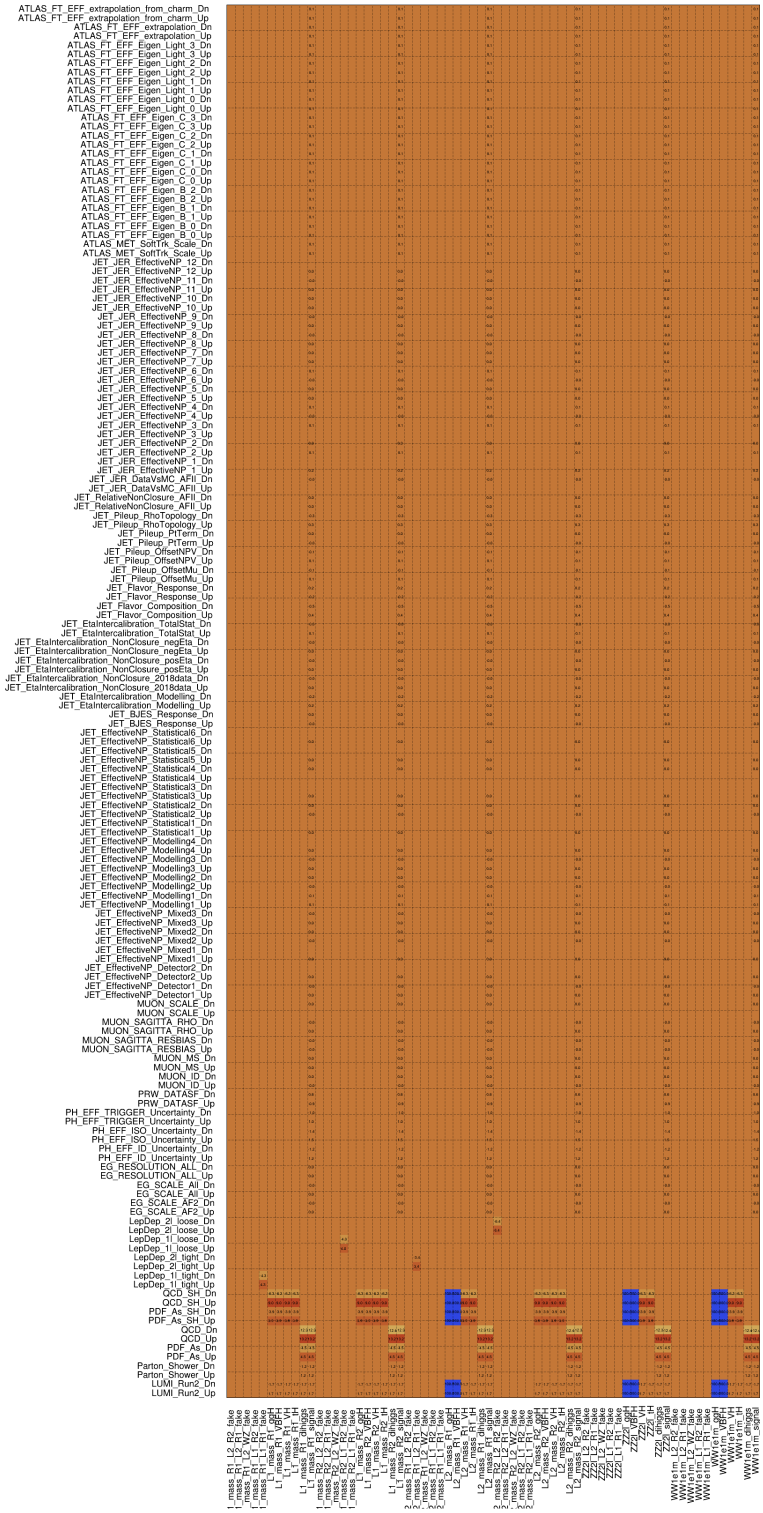


Figure 63: Normalization Factors for Nuisance parameters among all the regions for (m_X, m_S)=(500,200) GeV mass.

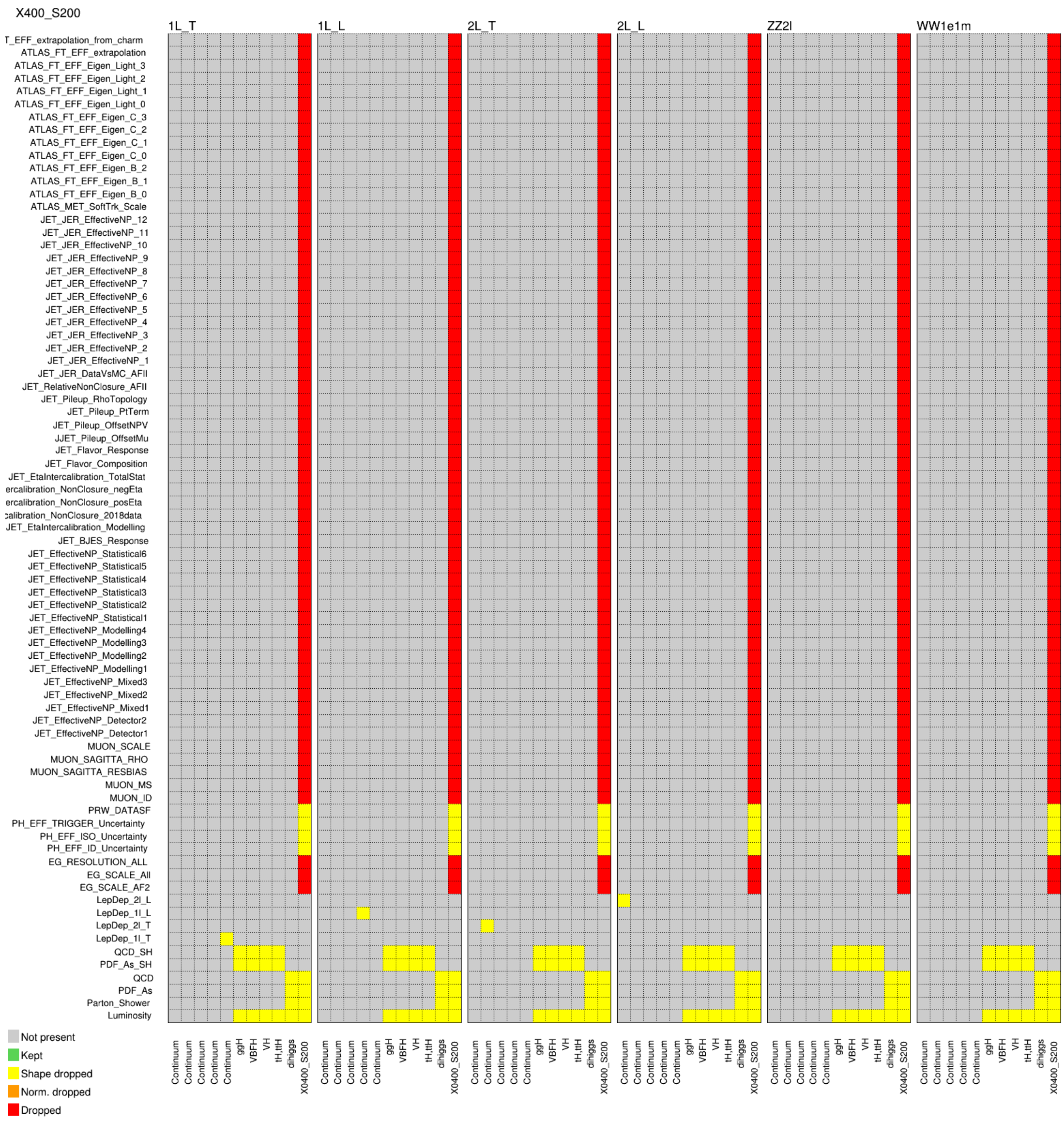


Figure 64: Pruning situation among all the regions for (m_X, m_S)=(500,200) GeV mass.

995 ## A.6 (m_X, m_S)=(500, 300) GeV

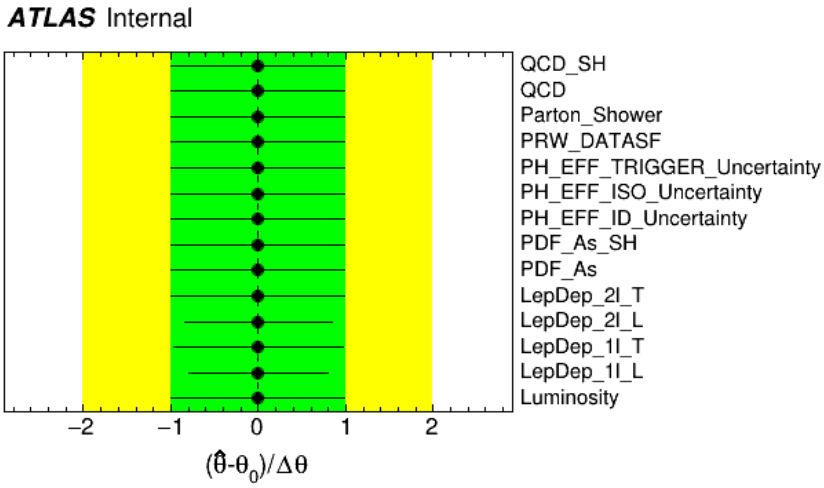


Figure 65: Nuisance parameters pull plot for (m_X, m_S)=(500,300) GeV mass

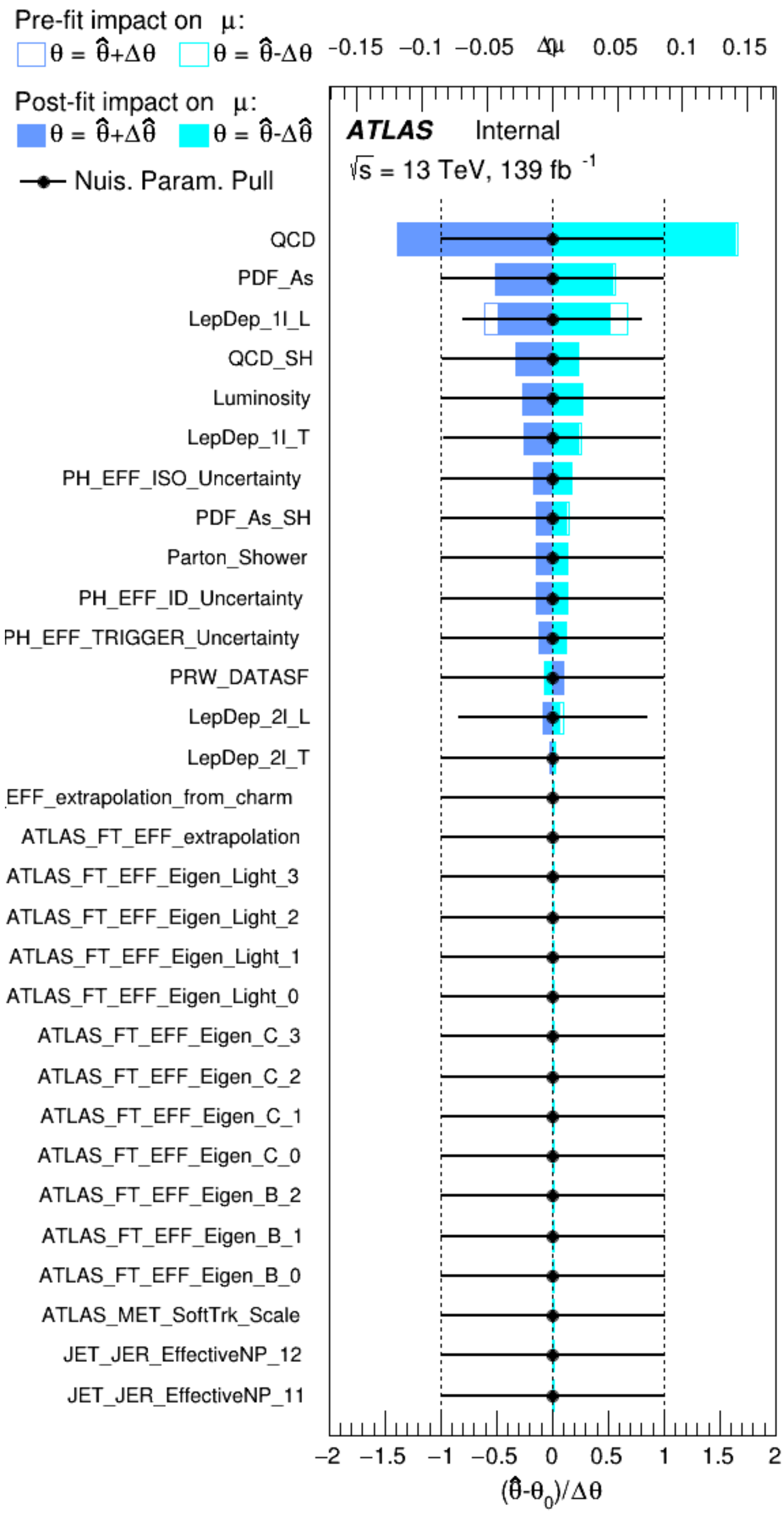


Figure 66: Nuisance parameters rankings for (m_X, m_S)=(500,300) GeV mass

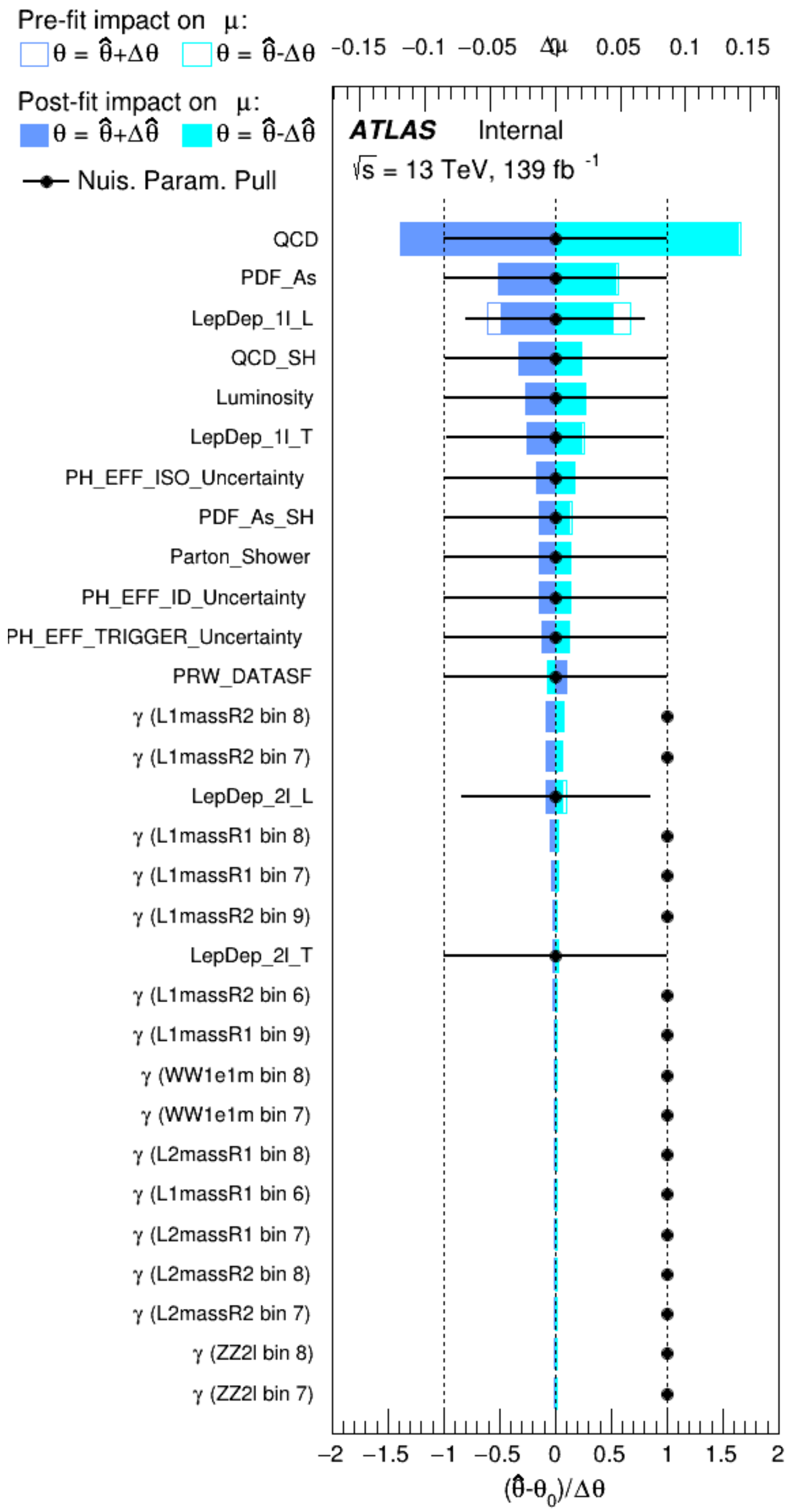


Figure 67: Rankings including systematics and gammas for (m_X, m_S)=(500,300) GeV mass

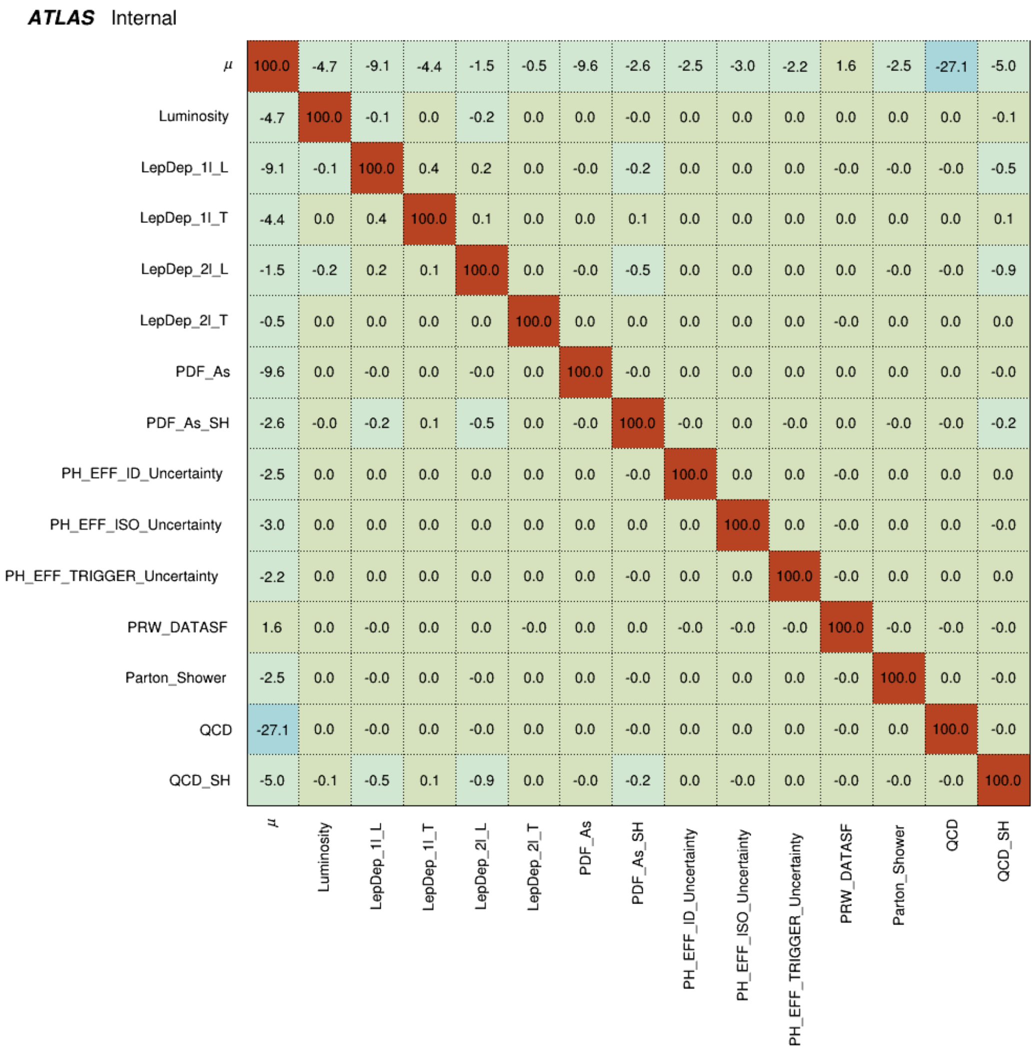


Figure 68: Correlation Matrix for (m_X, m_S)=(500,300) GeV mass.

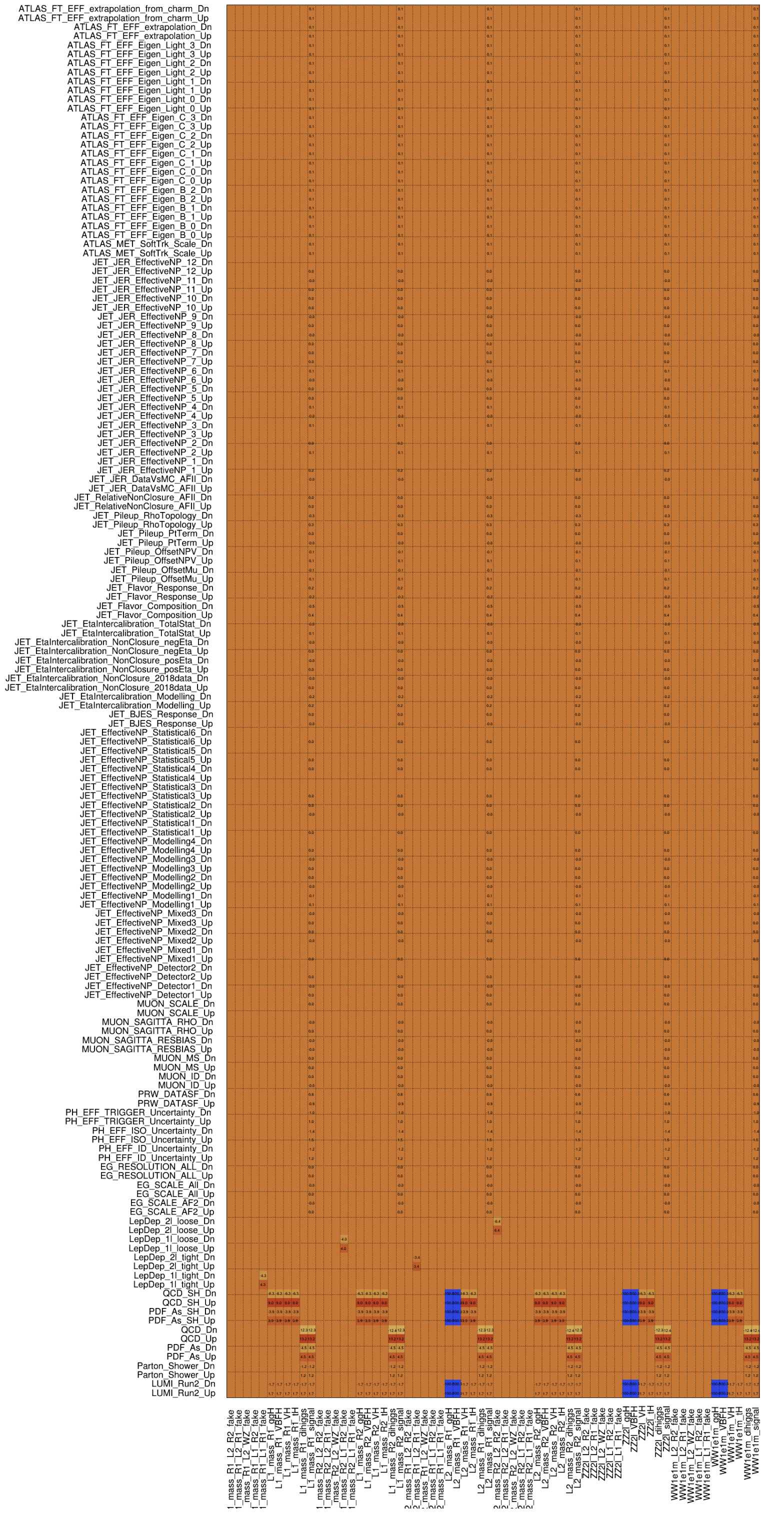


Figure 69: Normalization Factors for Nuisance parameters among all the regions for (m_X, m_S)=(500,300) GeV mass.

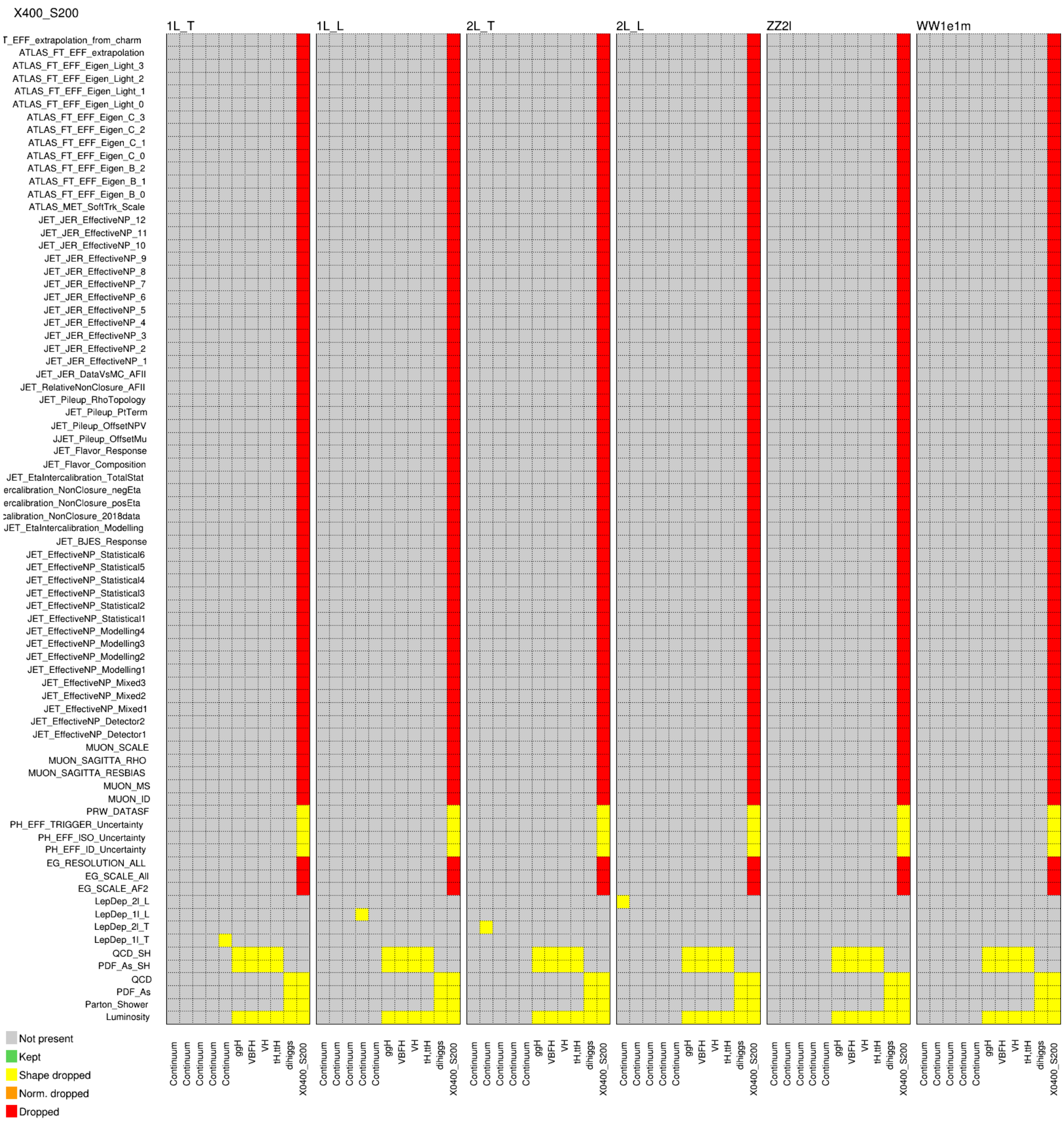


Figure 70: Pruning situation among all the regions for (m_X, m_S)=(500,300) GeV mass.

996 A.7 (m_X, m_S)=(600, 170) GeV

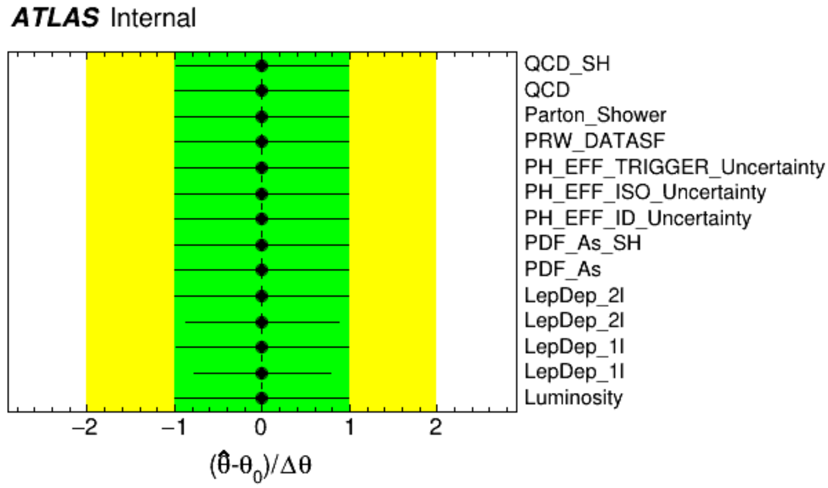


Figure 71: Nuisance parameters pull plot for (m_X, m_S)=(600,170) GeV mass

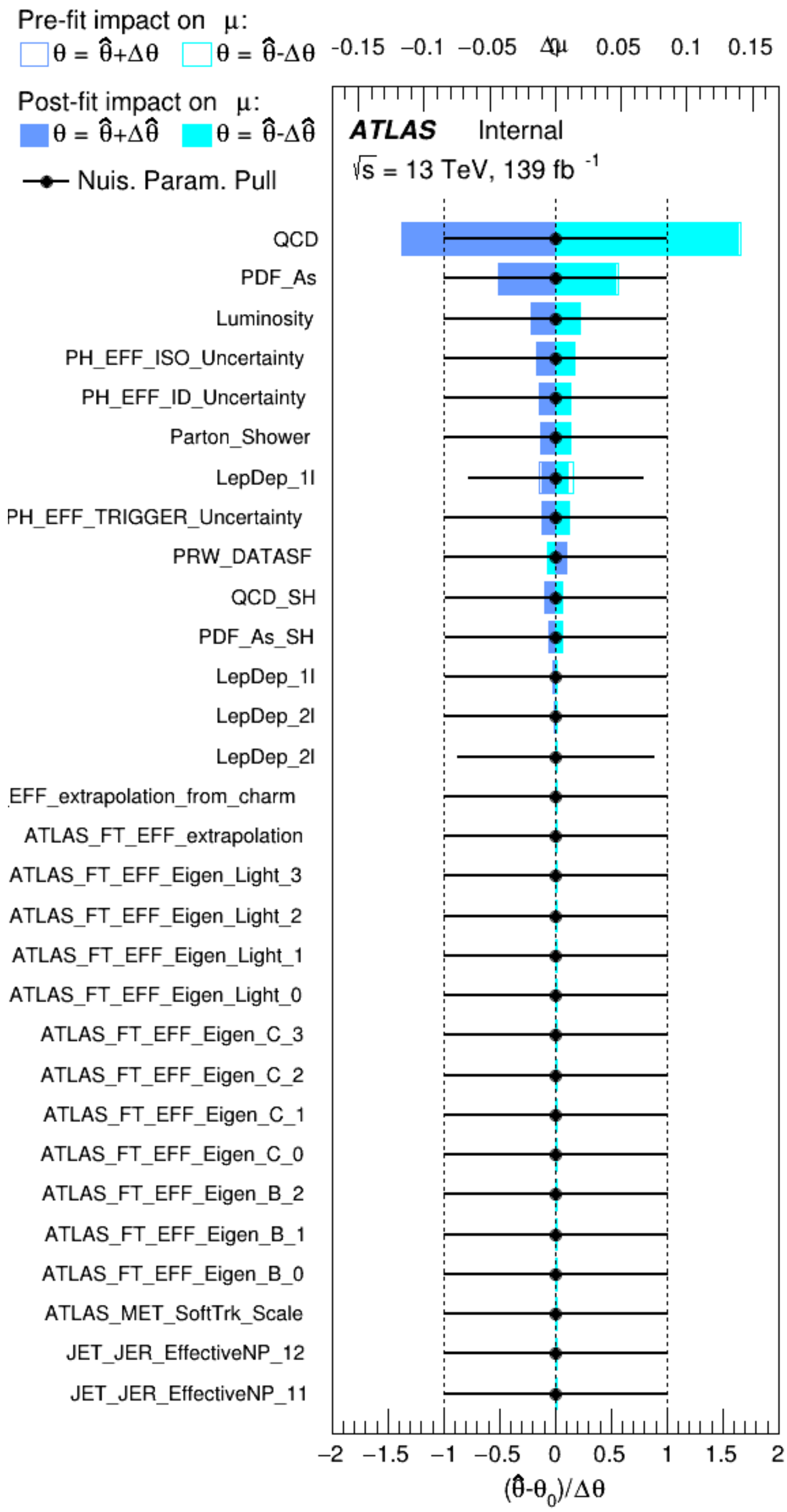


Figure 72: Nuisance parameters rankings for (m_X, m_S)=(600,170) GeV mass

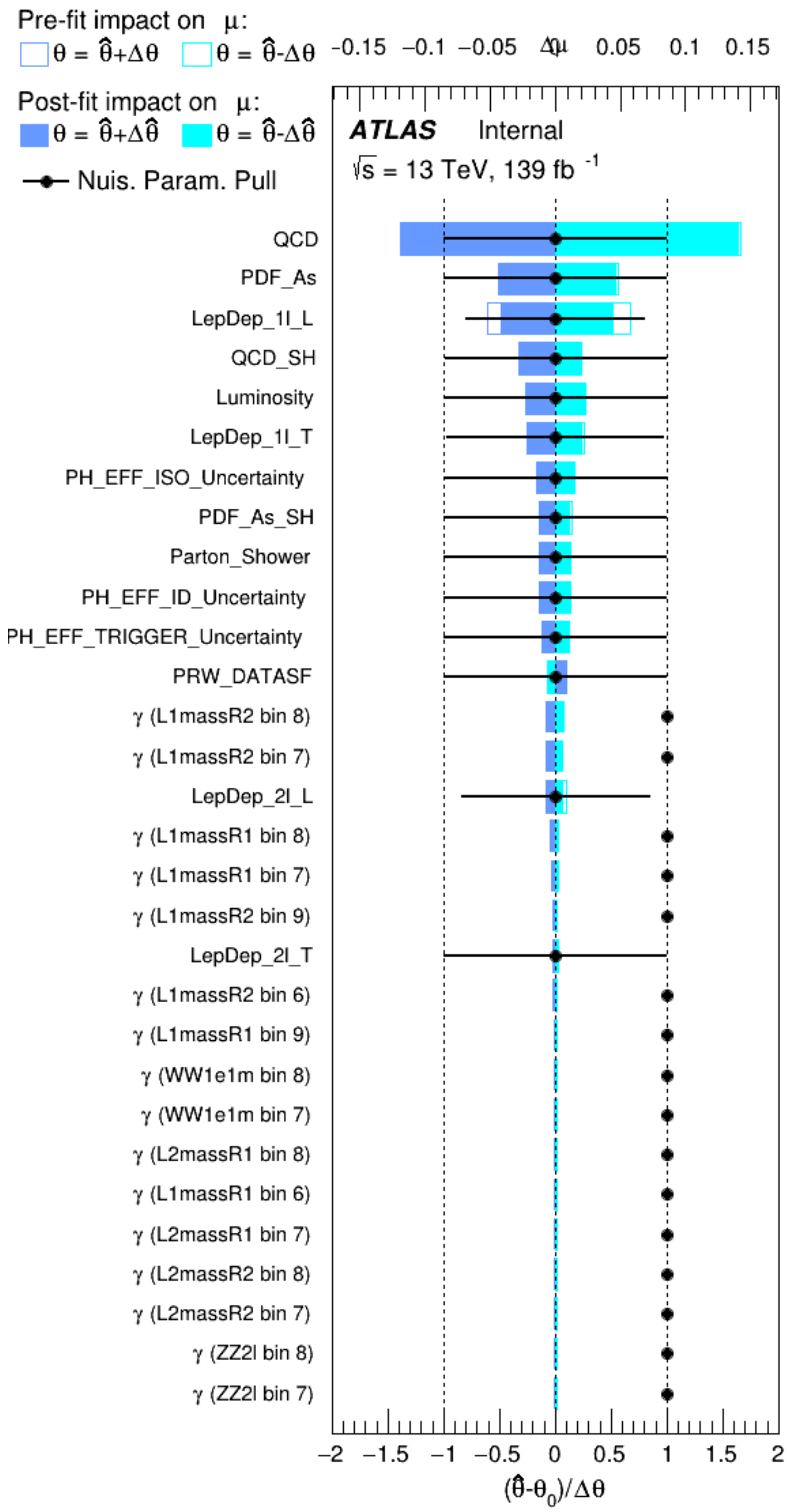


Figure 73: Rankings including systematics and gammas for (m_X, m_S)=(600,170) GeV mass

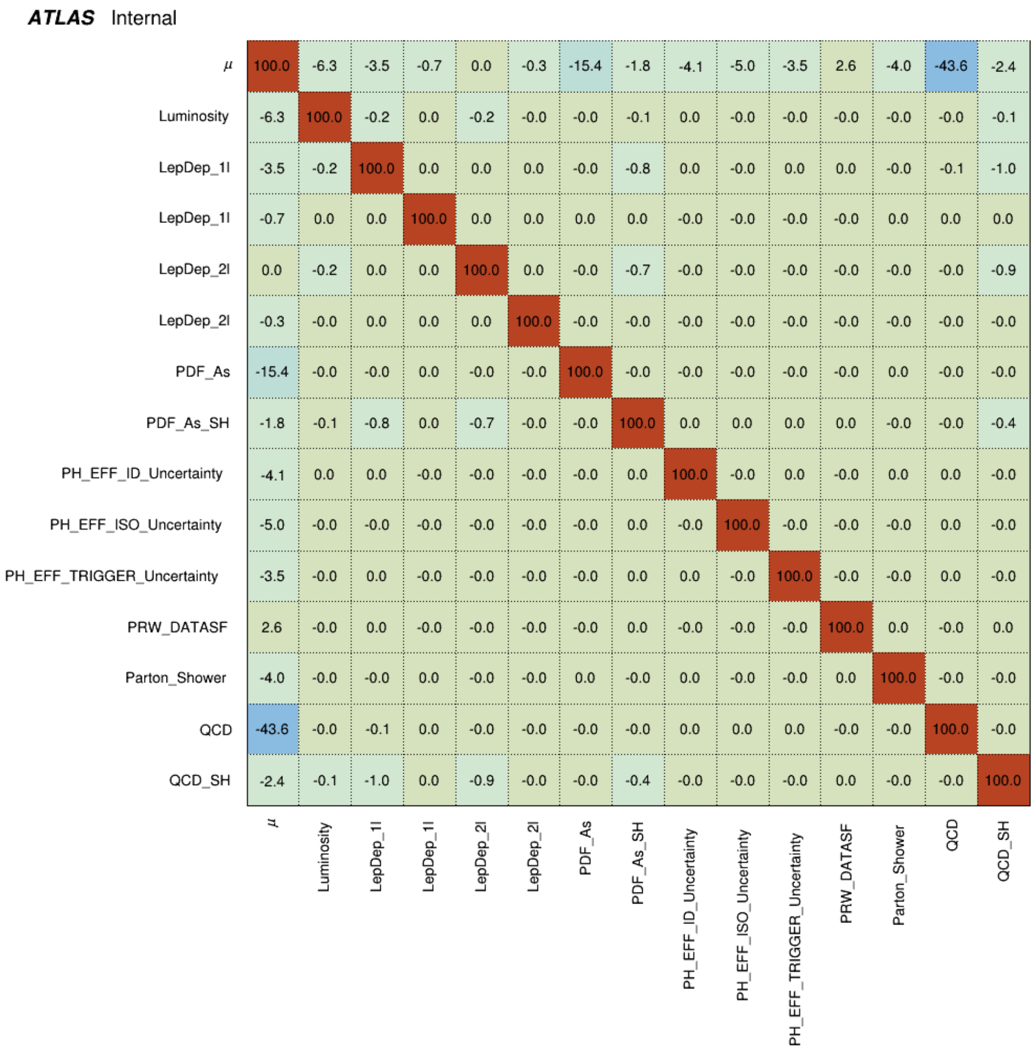


Figure 74: Correlation Matrix for (m_X, m_S)=(600,170) GeV mass.

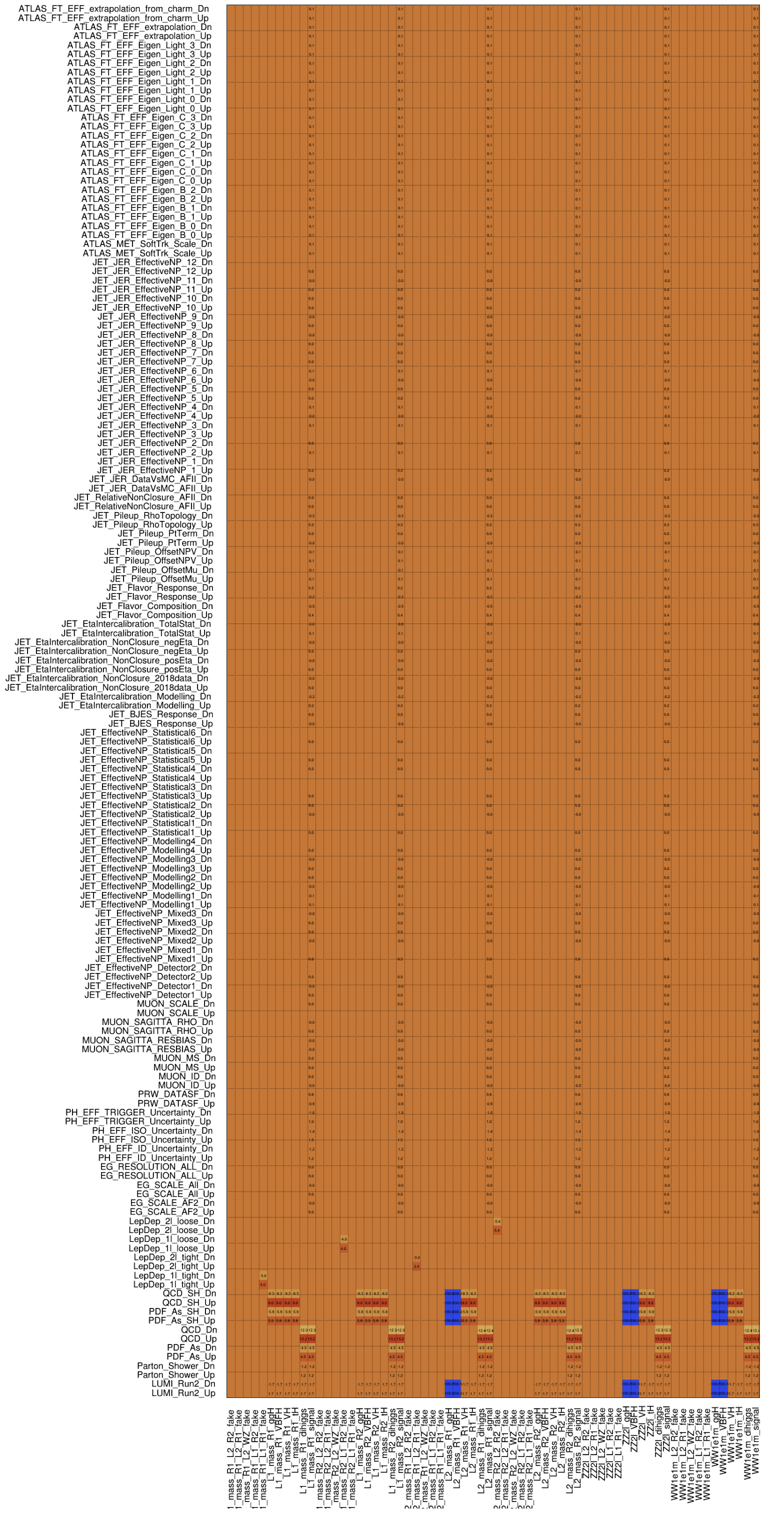


Figure 75: Normalization Factors for Nuisance parameters among all the regions for (m_X, m_S)=(600,170) GeV mass.

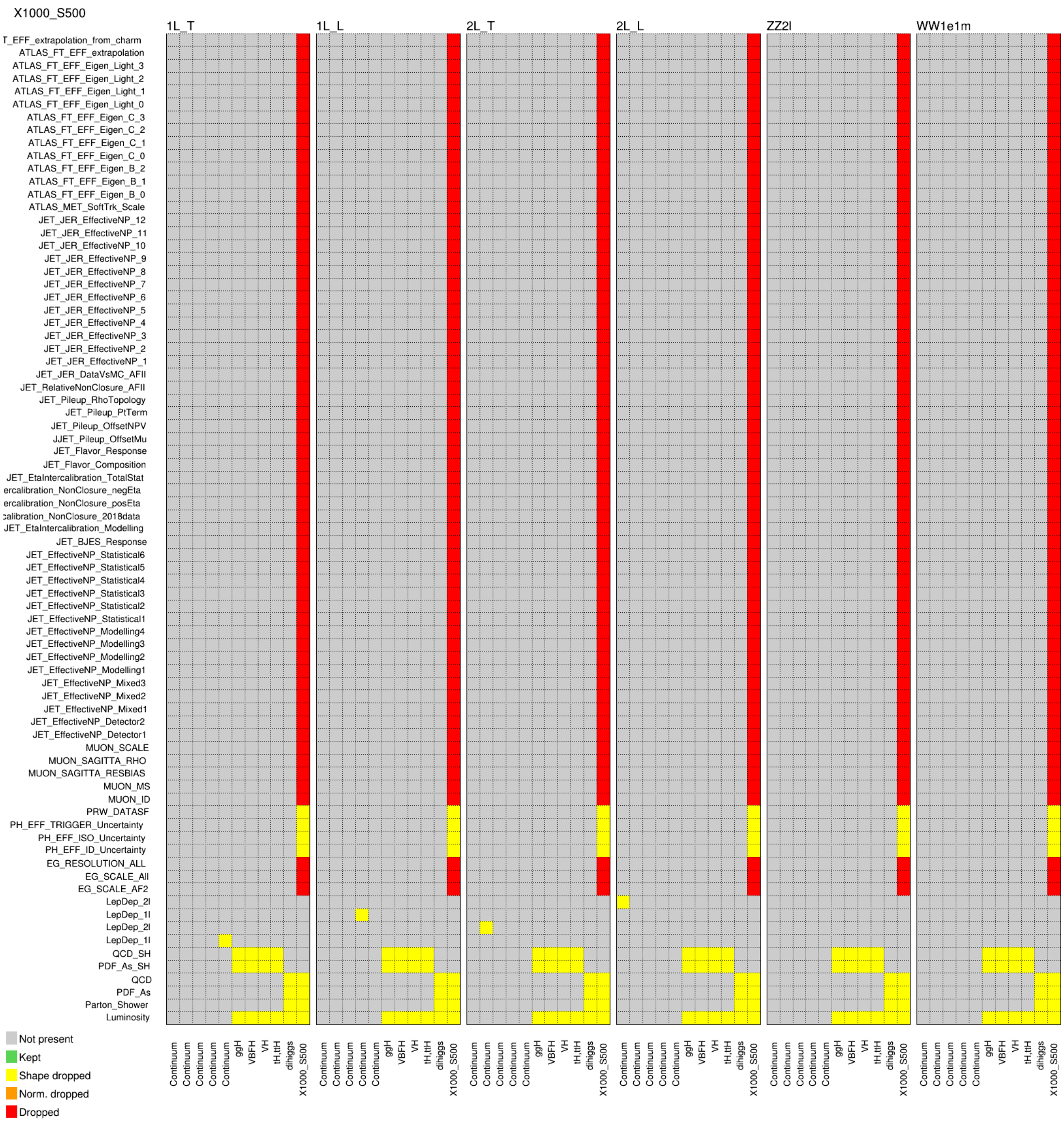


Figure 76: Pruning situation among all the regions for (m_X, m_S)=(600,170) GeV mass.

997 ## A.8 (m_X, m_S)=(600, 200) GeV

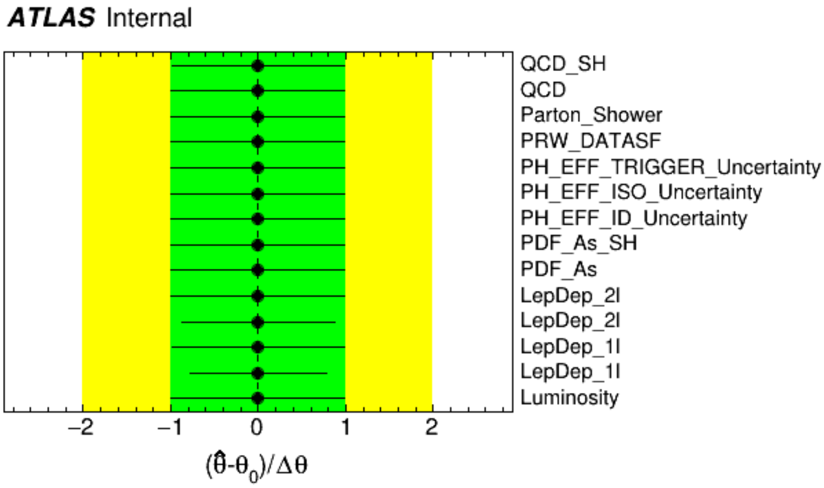


Figure 77: Nuisance parameters pull plot for (m_X, m_S)=(600,200) GeV mass

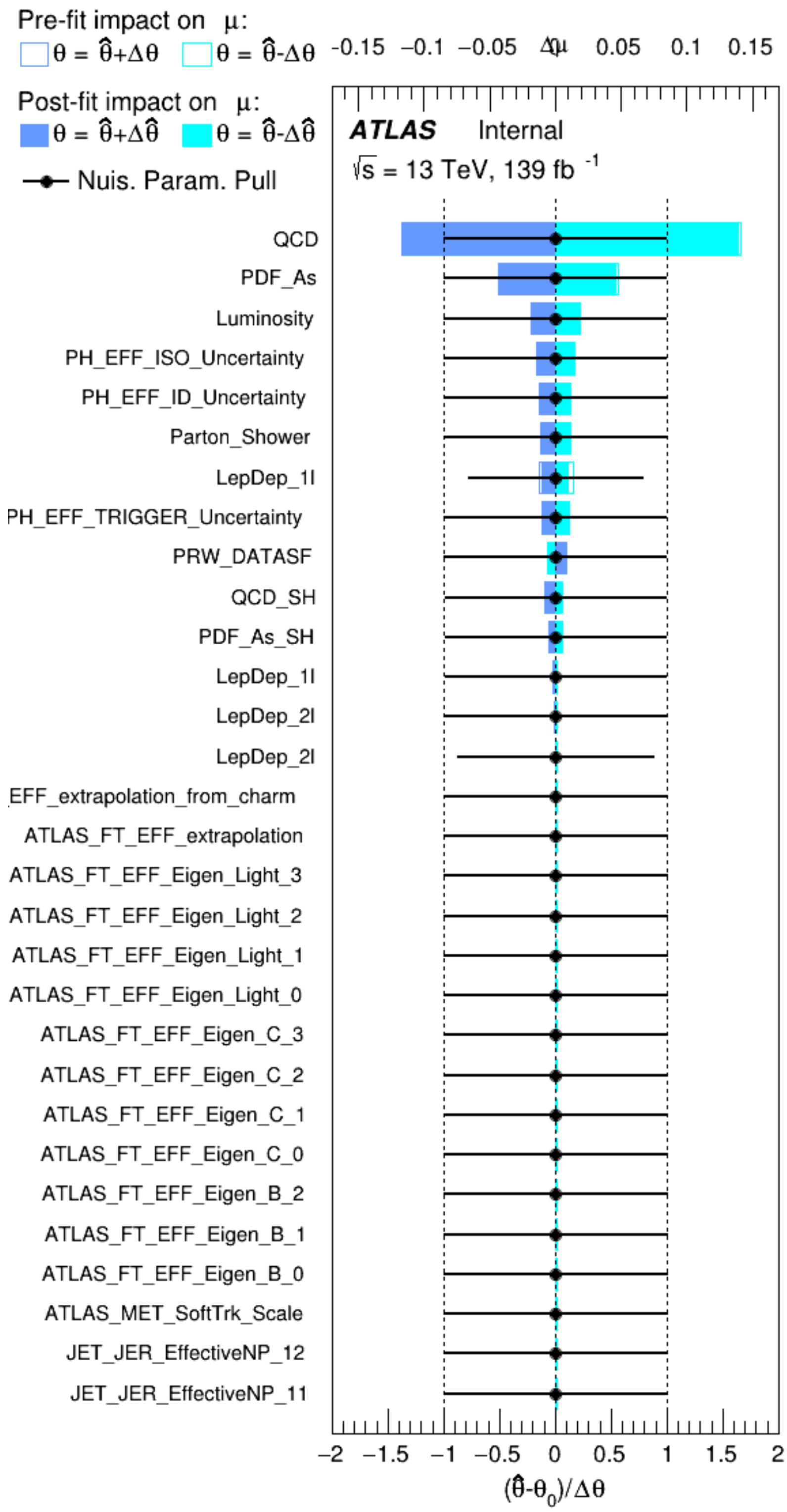


Figure 78: Nuisance parameters rankings for (m_X, m_S)=(600,200) GeV mass

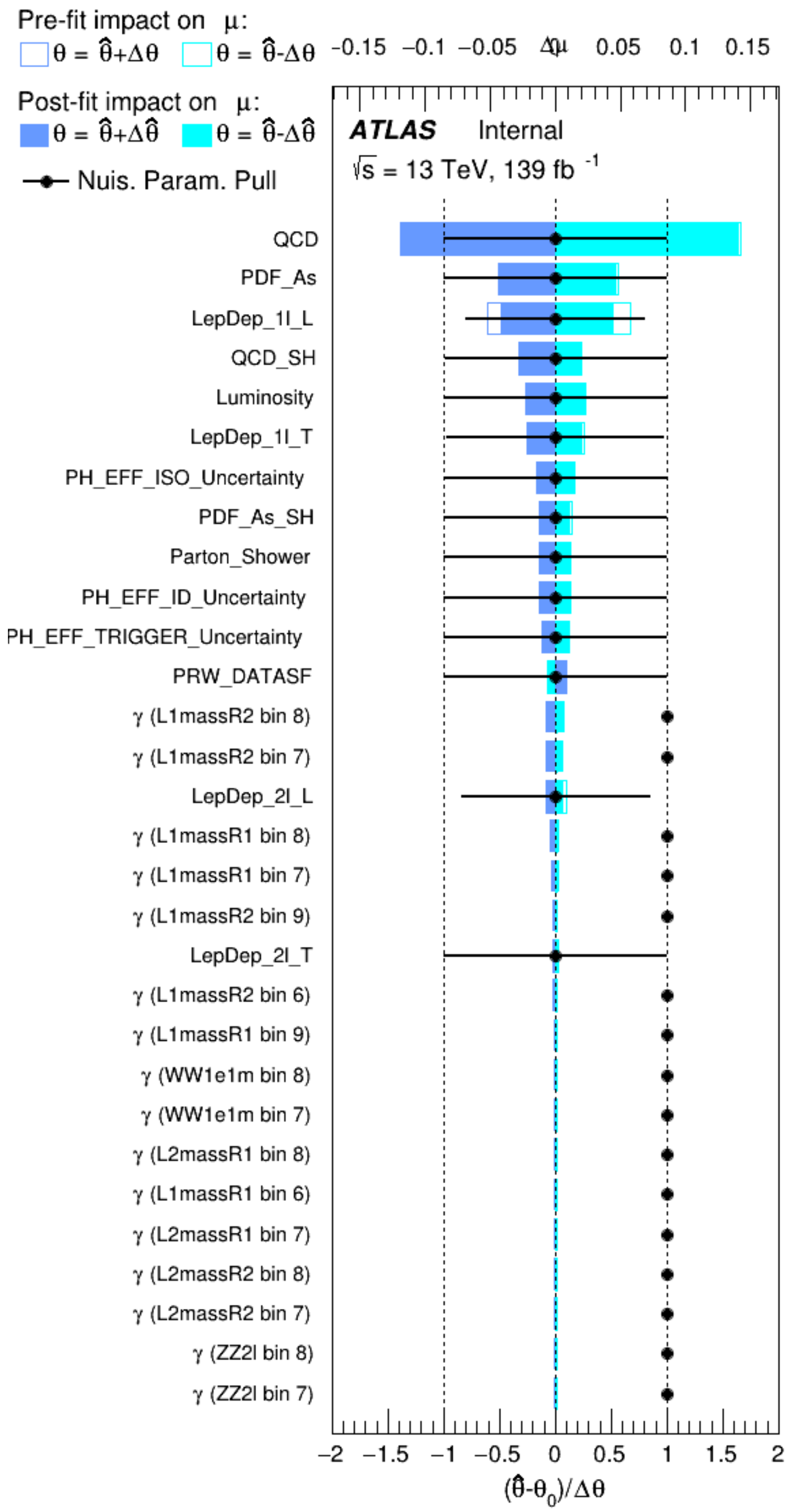


Figure 79: Rankings including systematics and gammas for (m_X, m_S)=(600,200) GeV mass

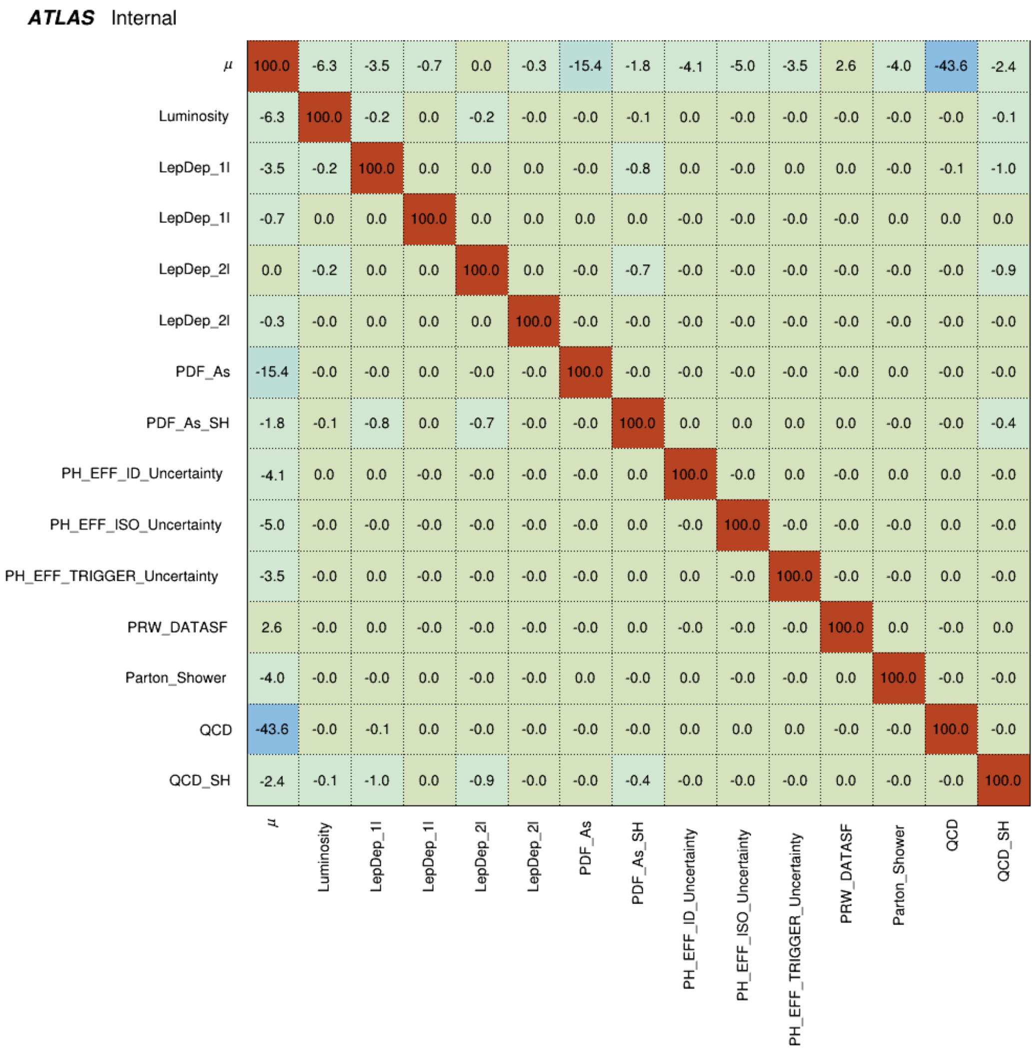


Figure 80: Correlation Matrix for (m_X, m_S)=(600,200) GeV mass.

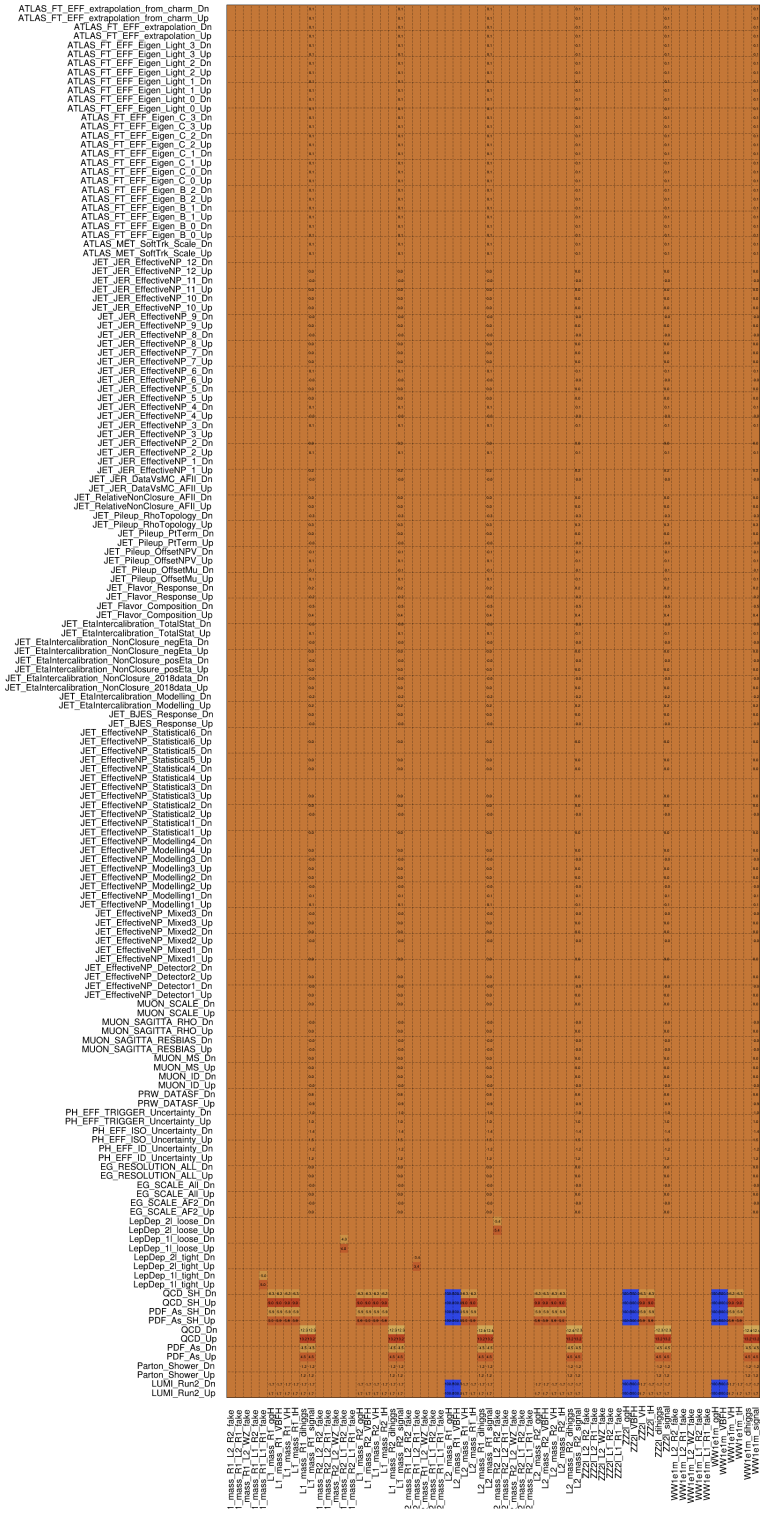


Figure 81: Normalization Factors for Nuisance parameters among all the regions for (m_X, m_S)=(600,200) GeV mass.

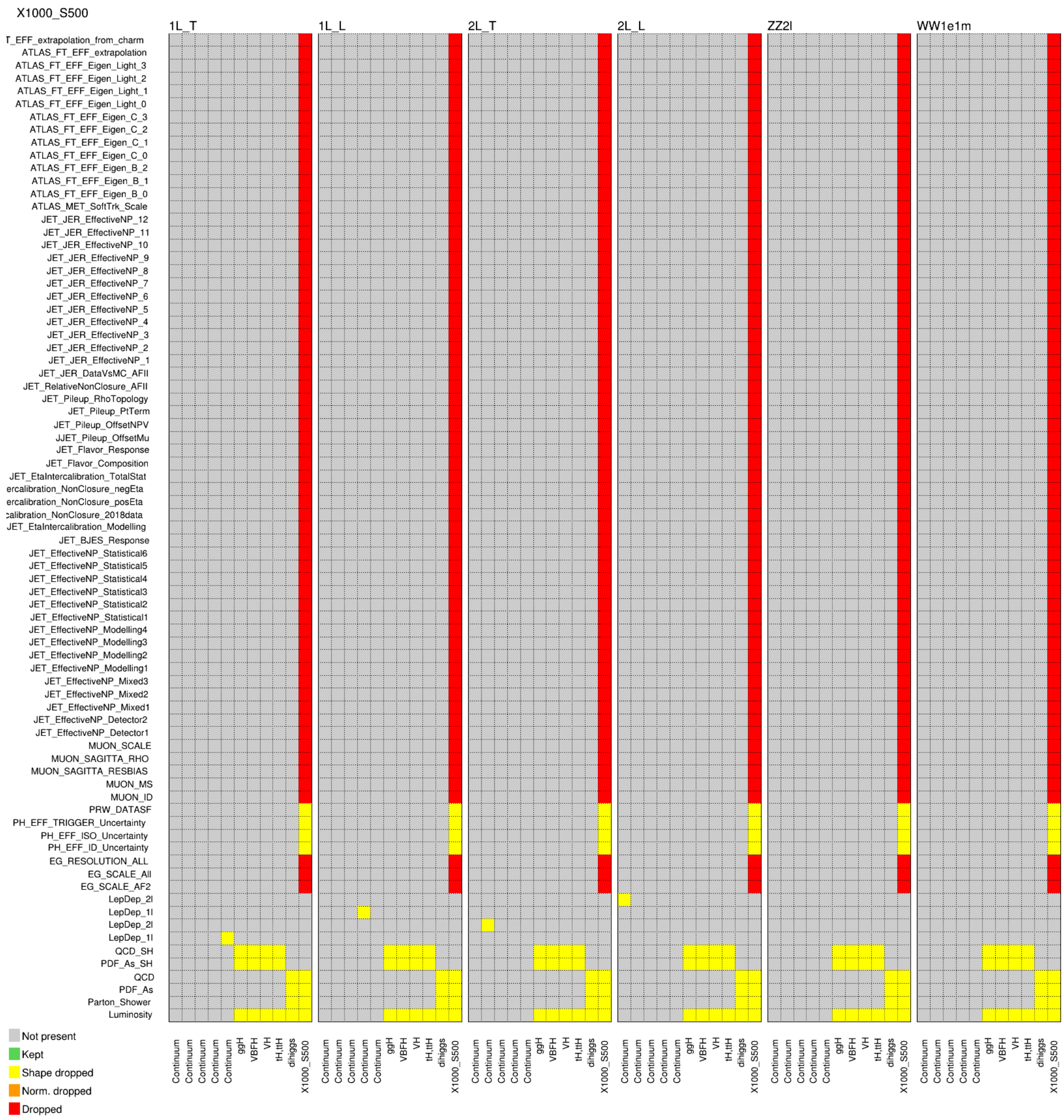


Figure 82: Pruning situation among all the regions for (m_X, m_S)=(600,200) GeV mass.

998 A.9 (m_X, m_S)=(600, 300) GeV

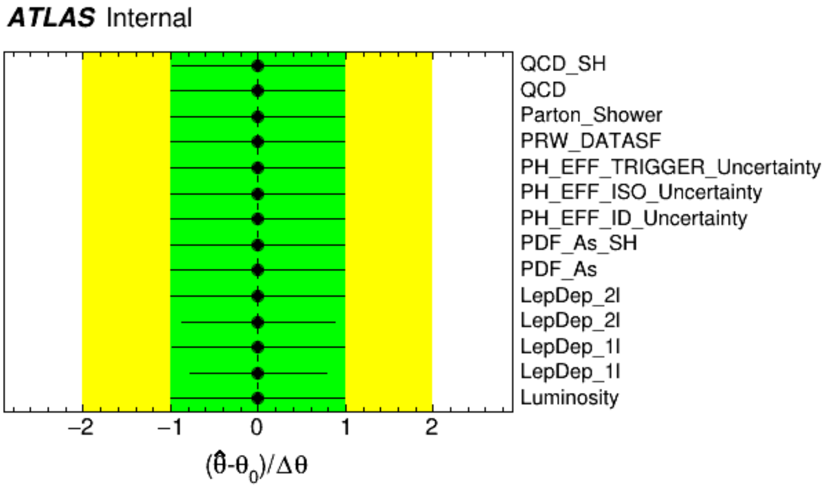


Figure 83: Nuisance parameters pull plot for (m_X, m_S)=(600,300) GeV mass

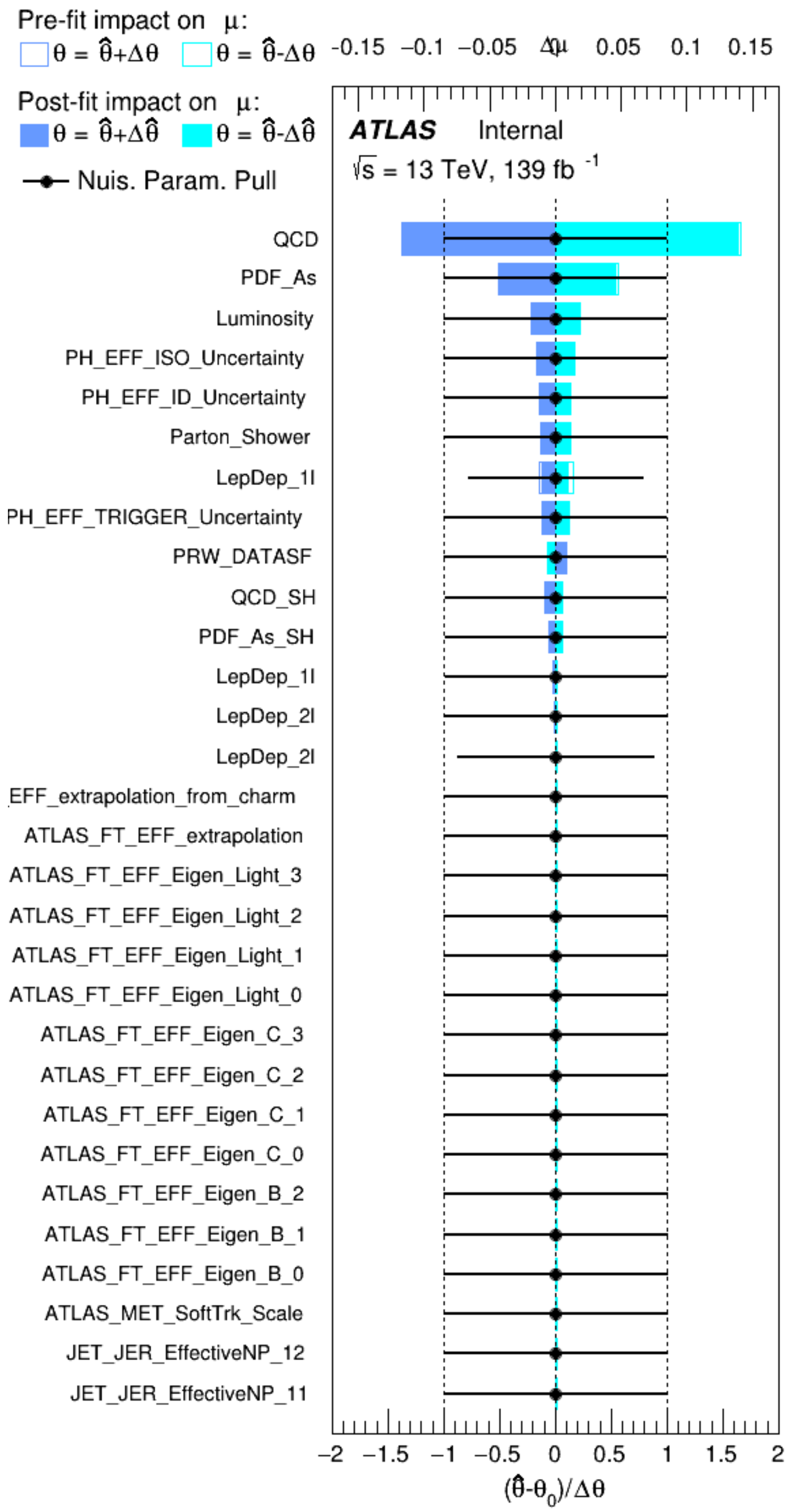


Figure 84: Nuisance parameters rankings for (m_X, m_S)=(600,300) GeV mass

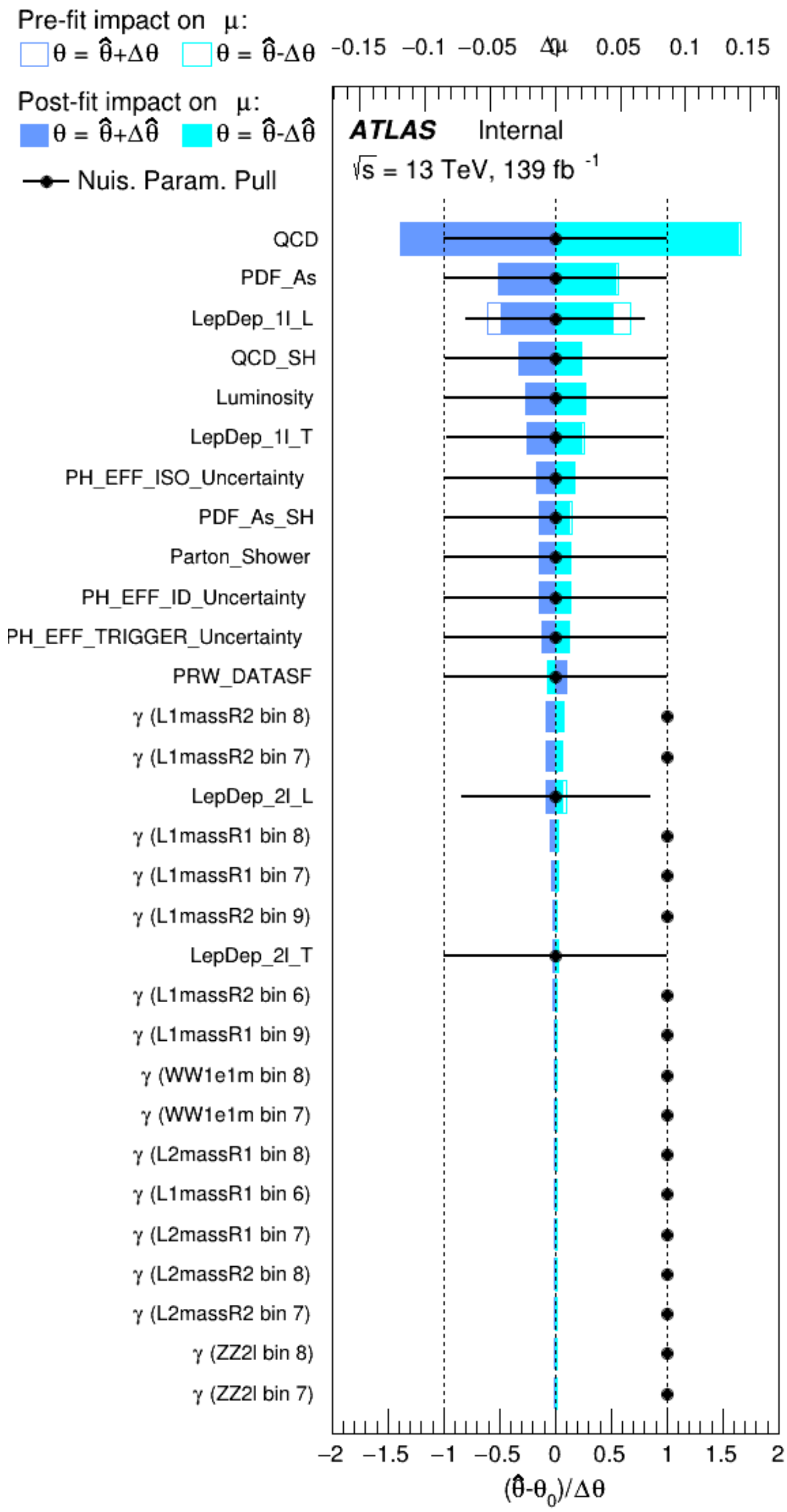


Figure 85: Rankings including systematics and gammas for (m_X, m_S)=(600,300) GeV mass

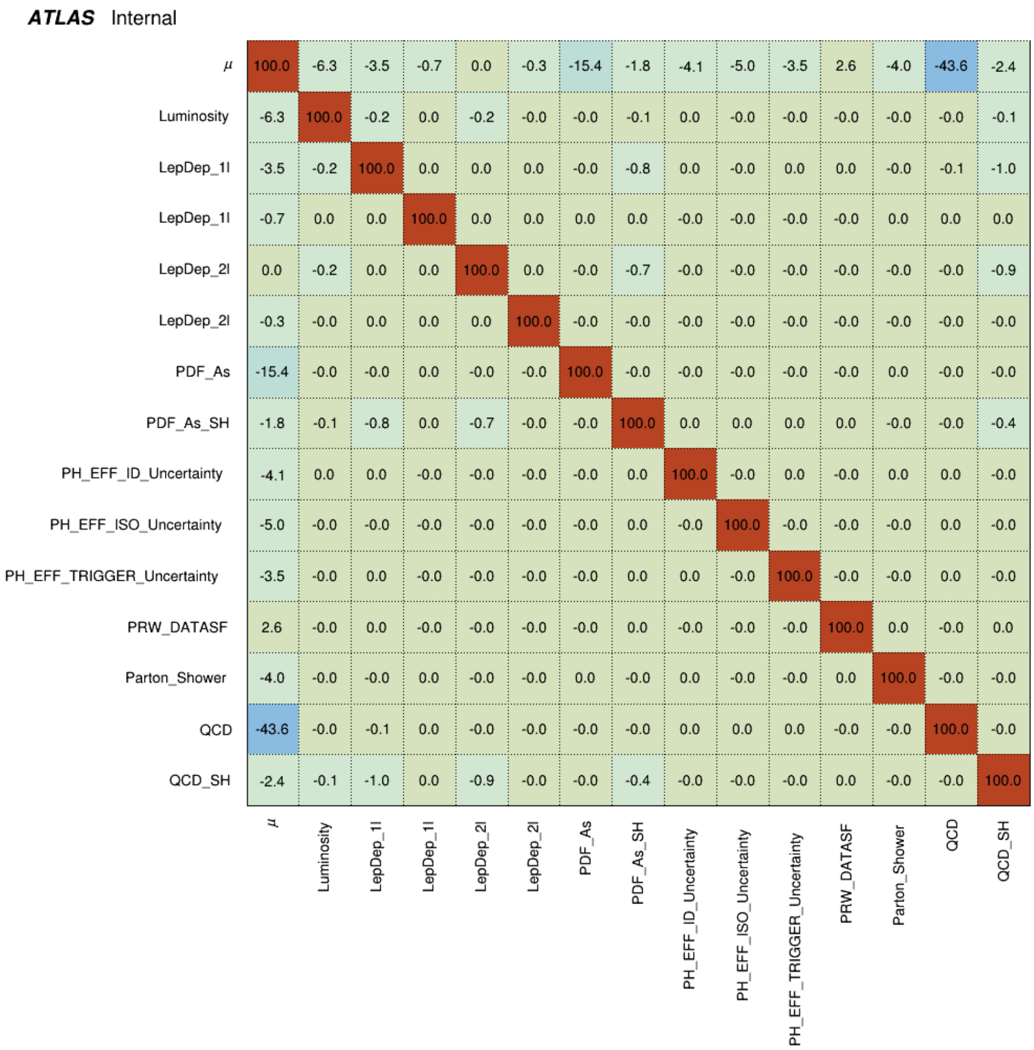


Figure 86: Correlation Matrix for (m_X, m_S)=(600,300) GeV mass.

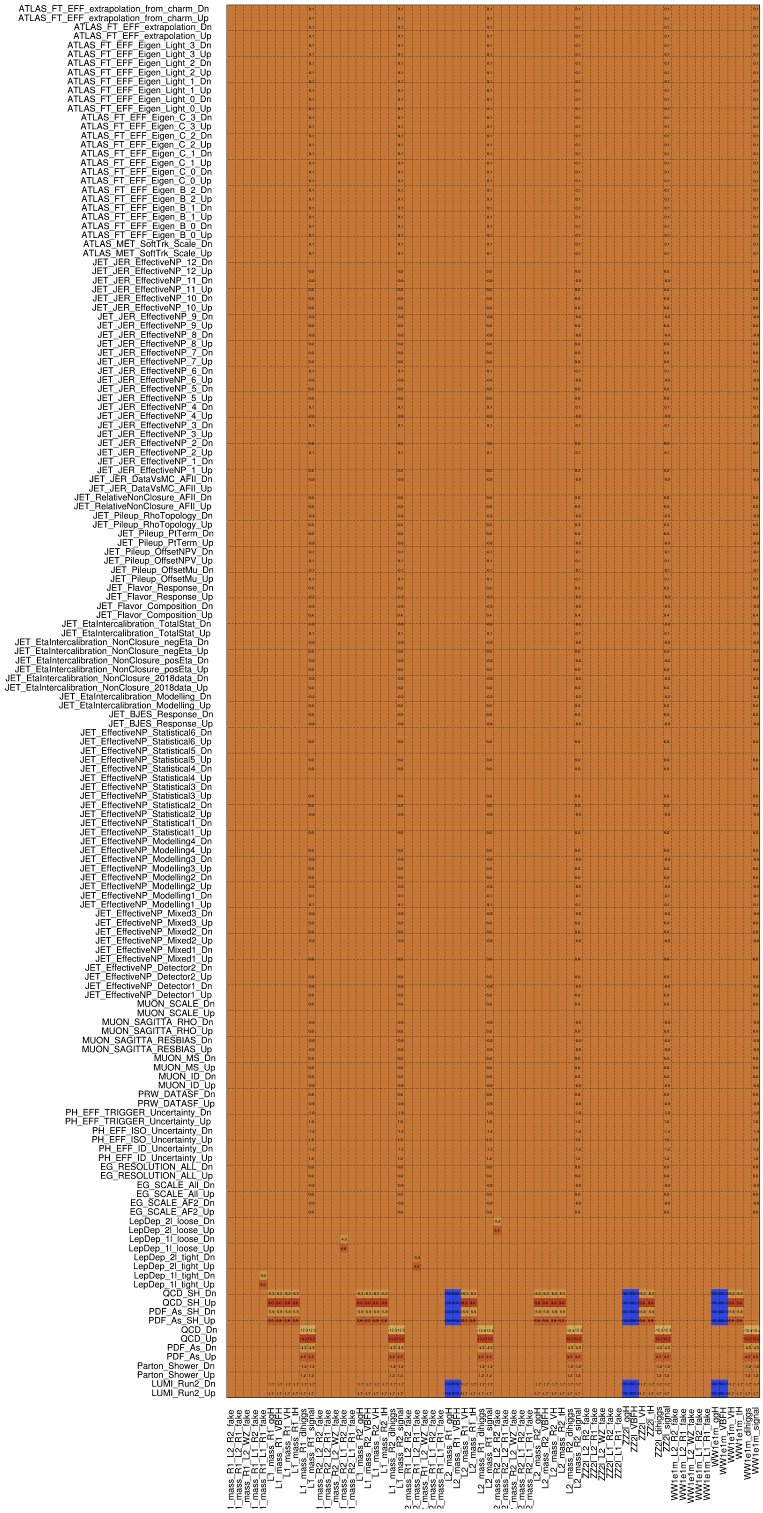


Figure 87: Normalization Factors for Nuisance parameters among all the regions for (m_X, m_S)=(600,300) GeV mass.

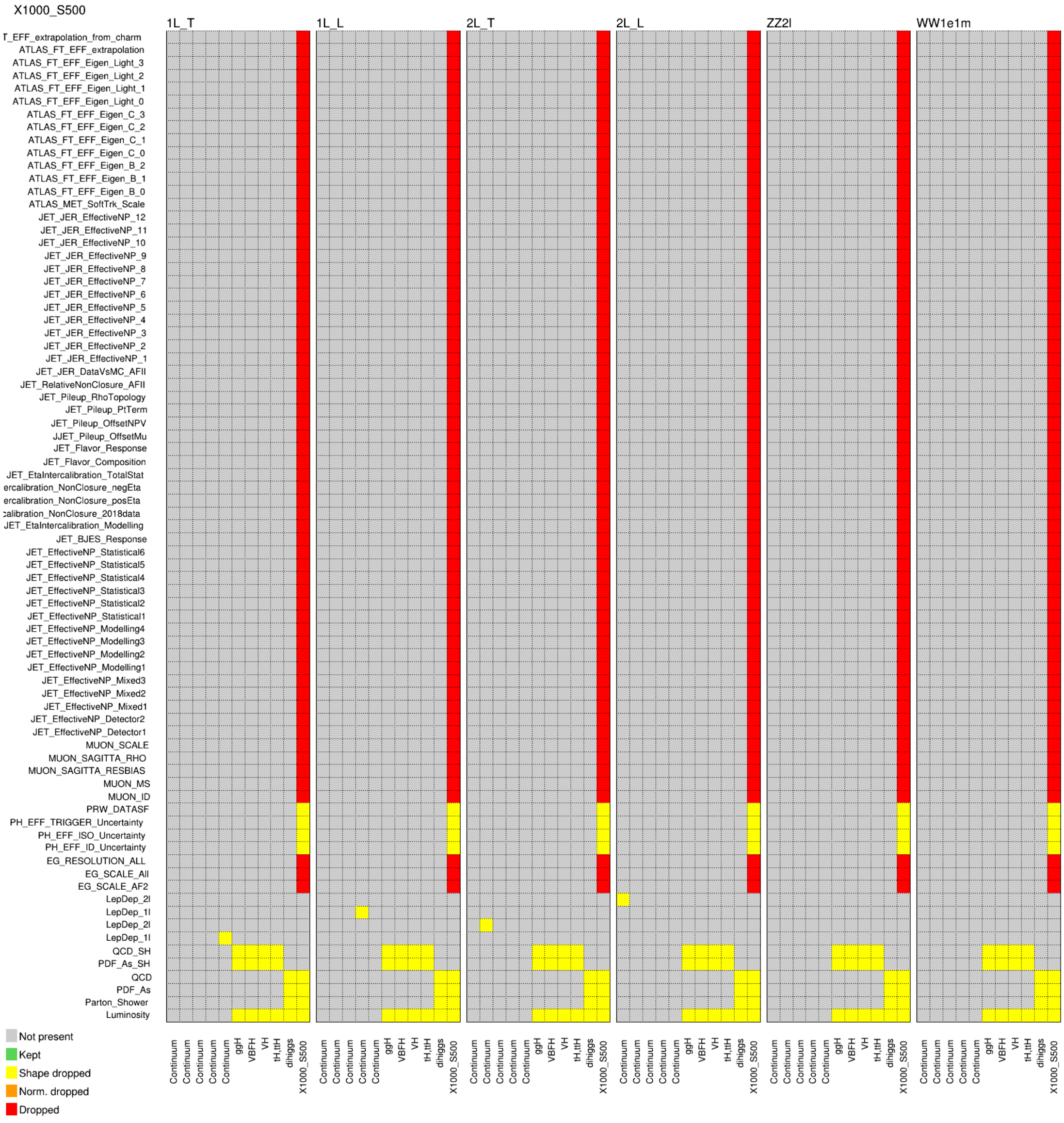


Figure 88: Pruning situation among all the regions for (m_X, m_S)=(600,300) GeV mass.

999 ## A.10 (m_X, m_S)=(600, 400) GeV

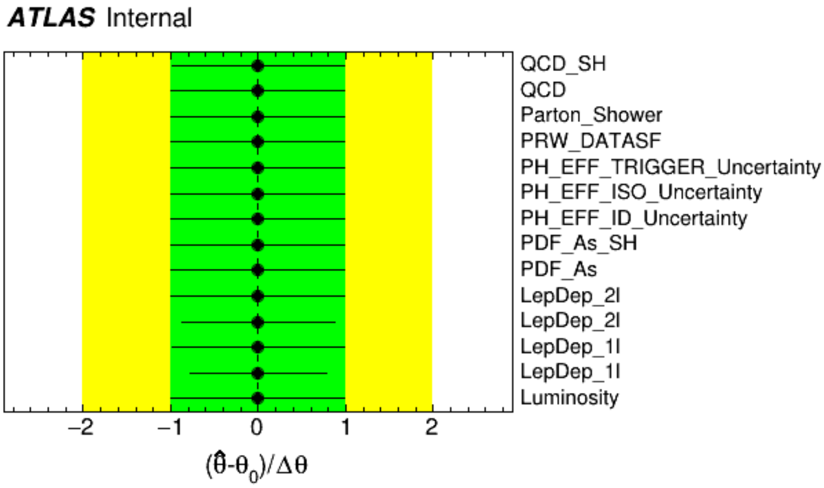


Figure 89: Nuisance parameters pull plot for (m_X, m_S)=(600,400) GeV mass

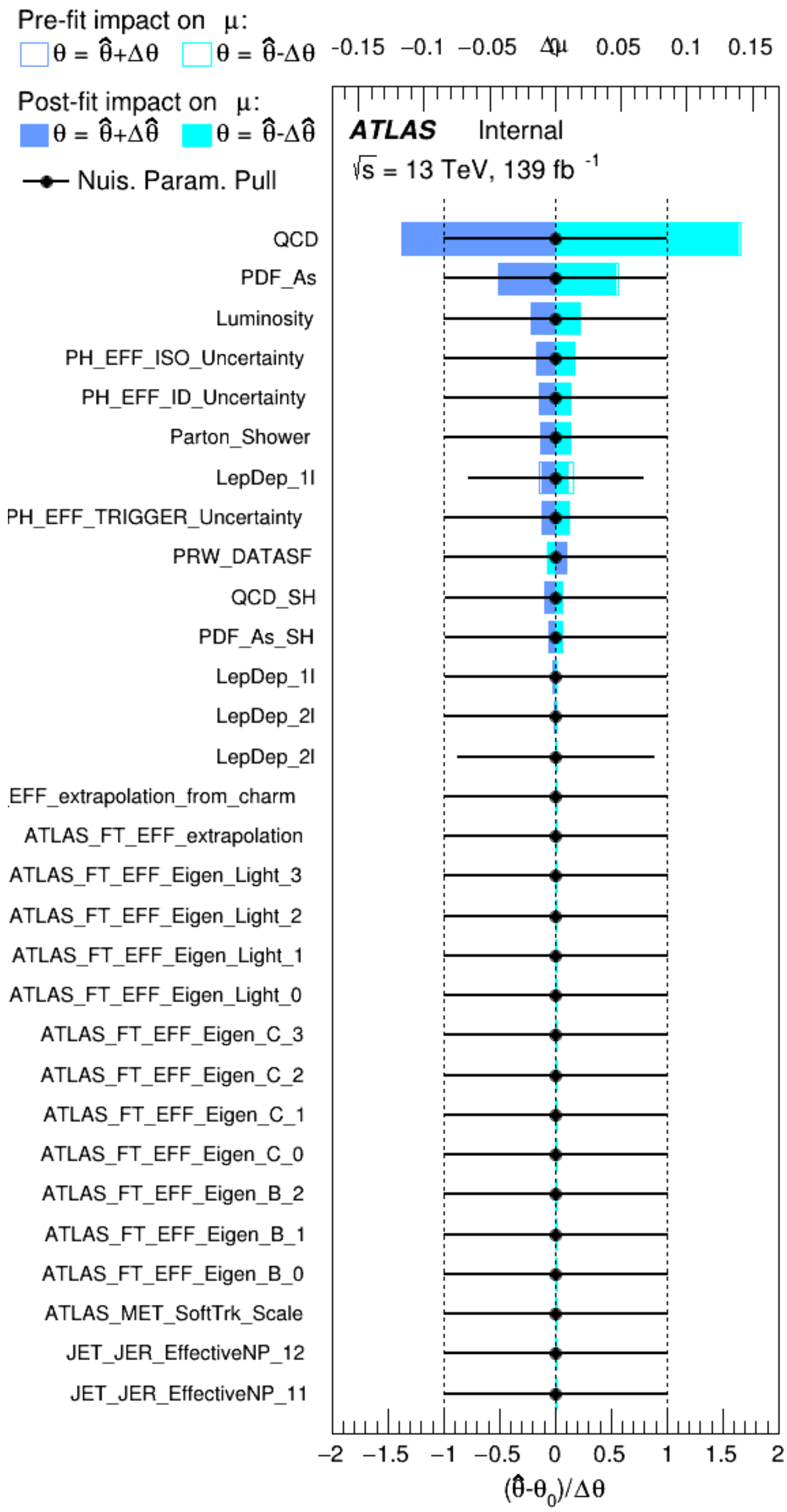


Figure 90: Nuisance parameters rankings for (m_X, m_S)=(600,400) GeV mass

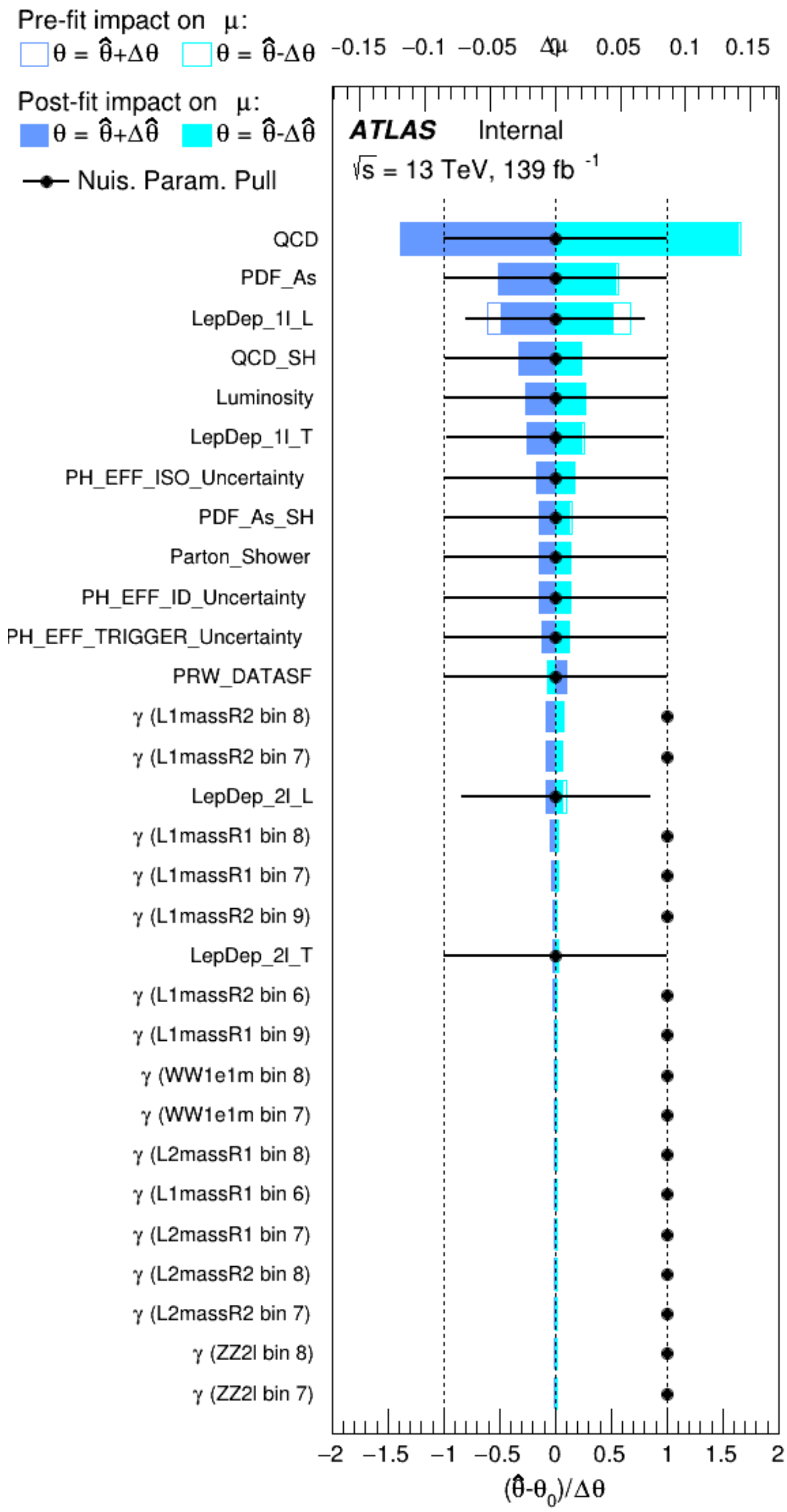


Figure 91: Rankings including systematics and gammas for (m_X, m_S)=(600,400) GeV mass

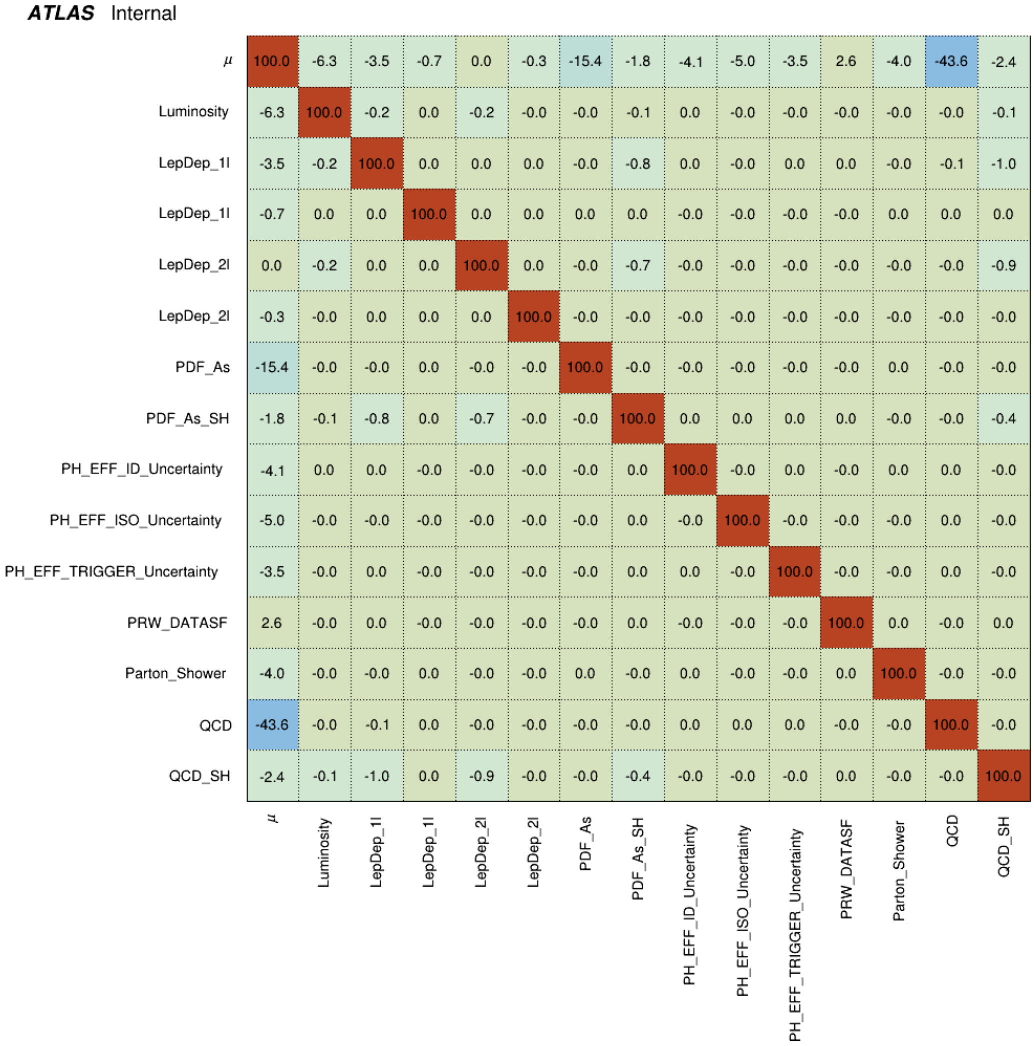


Figure 92: Correlation Matrix for (m_X, m_S)=(600,400) GeV mass.

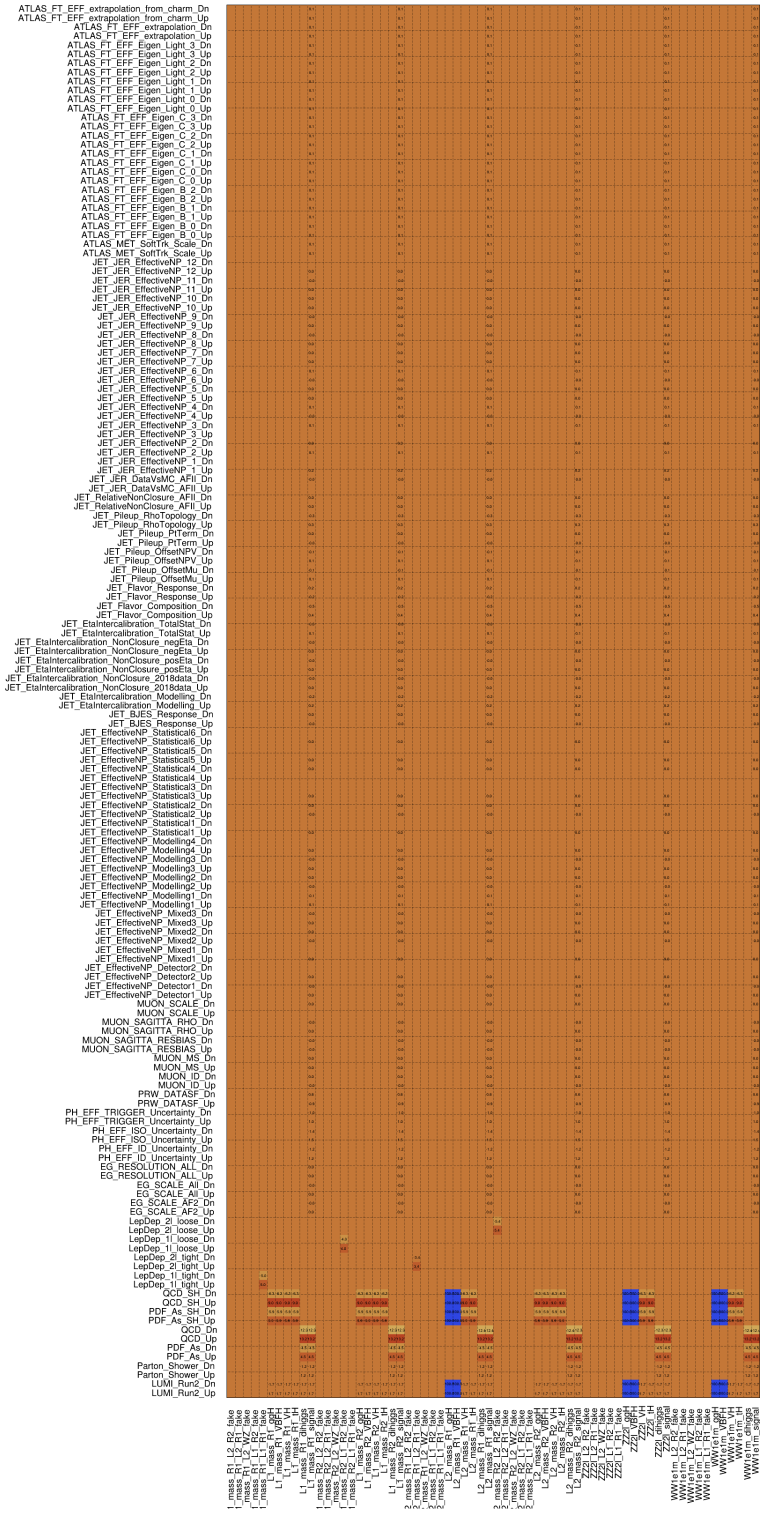


Figure 93: Normalization Factors for Nuisance parameters among all the regions for (m_X, m_S)=(600,400) GeV mass.

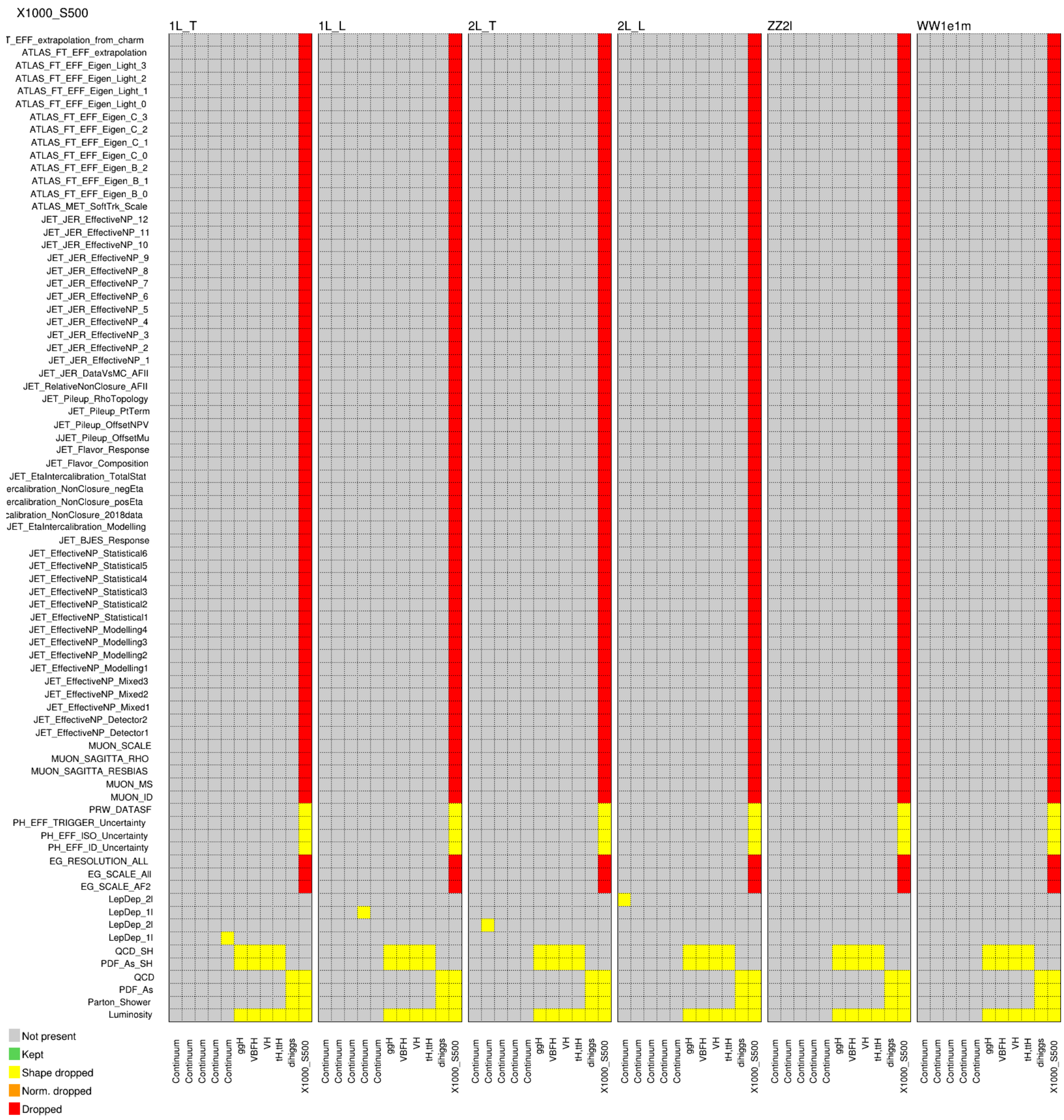


Figure 94: Pruning situation among all the regions for (m_X, m_S)=(600,400) GeV mass.

1000 ## A.11 (m_X, m_S)=(750, 170) GeV

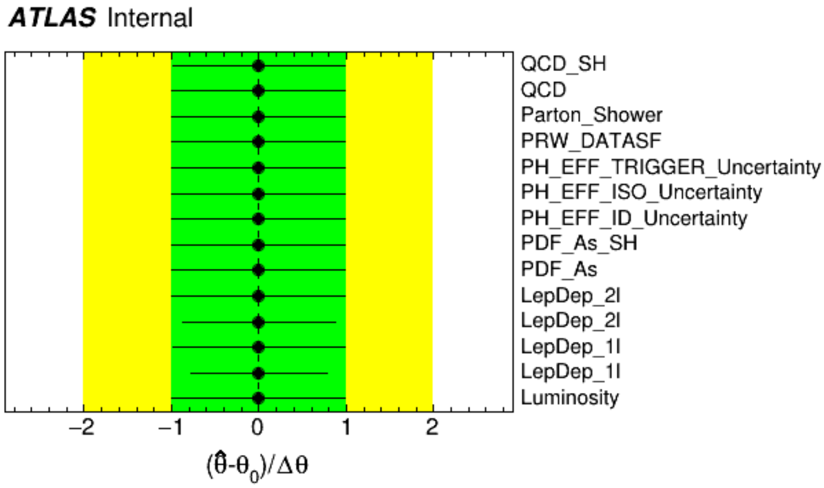


Figure 95: Nuisance parameters pull plot for (m_X, m_S)=(750,170) GeV mass

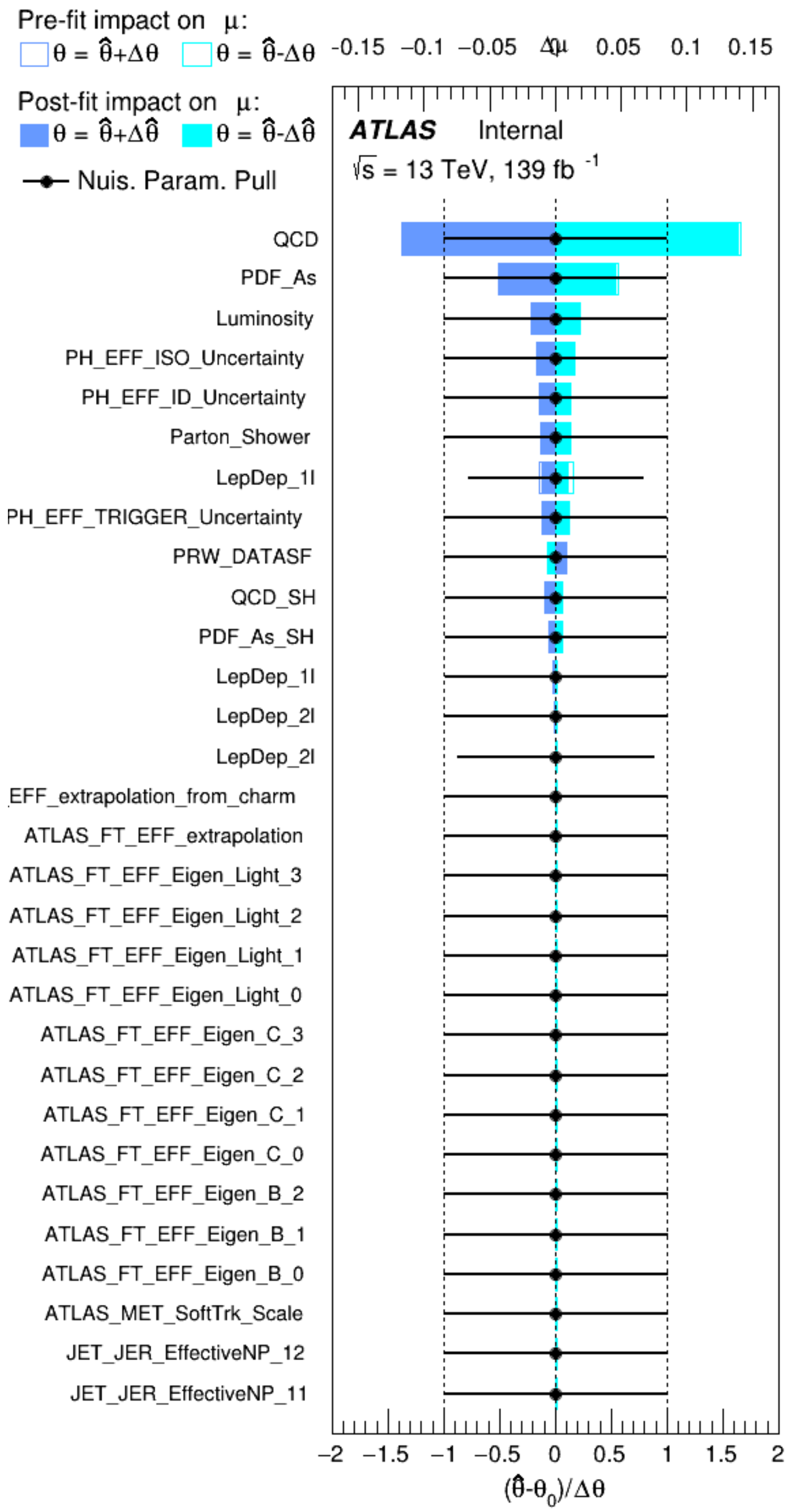


Figure 96: Nuisance parameters rankings for (m_X, m_S)=(750,170) GeV mass

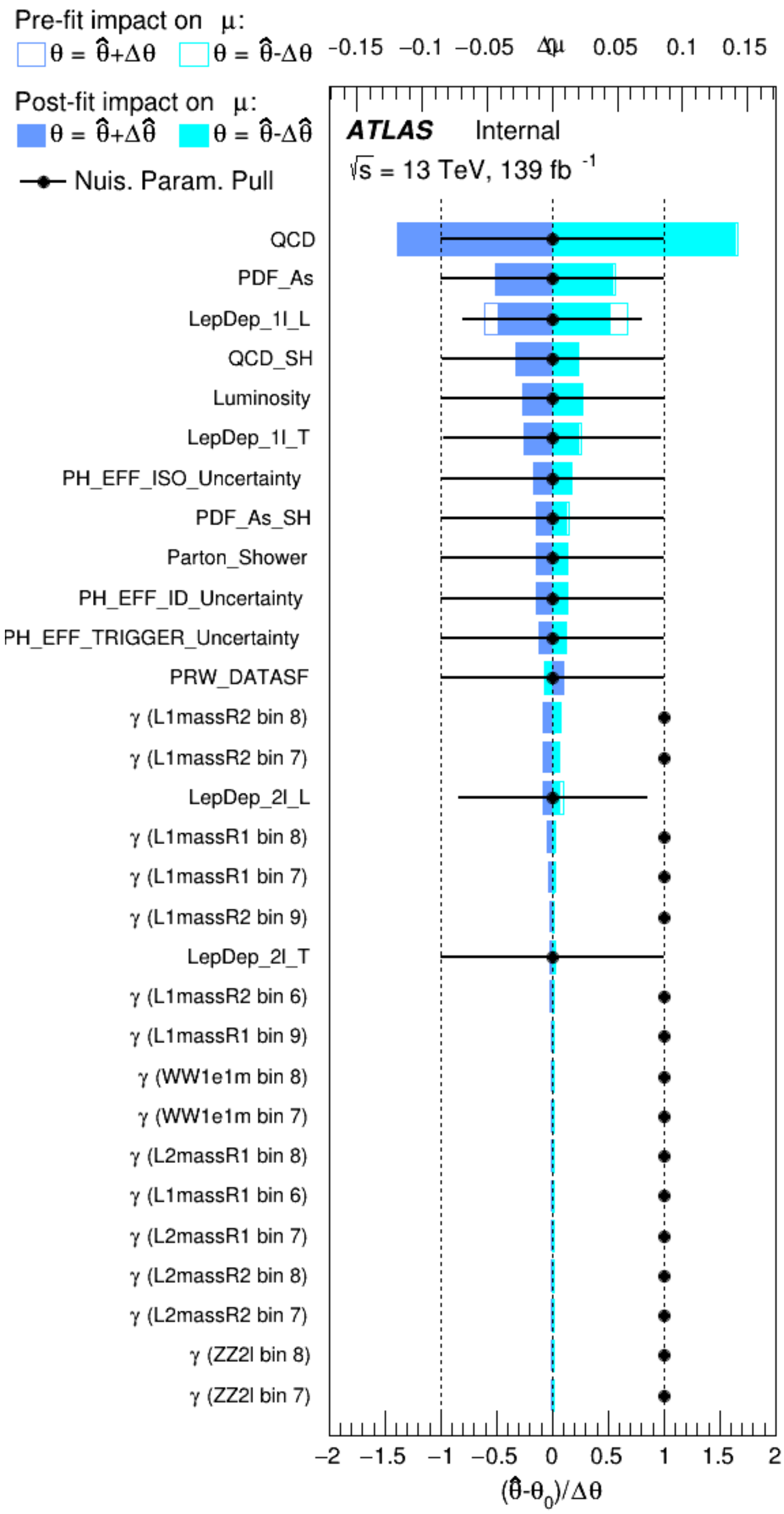


Figure 97: Rankings including systematics and gammas for (m_X, m_S)=(750,170) GeV mass

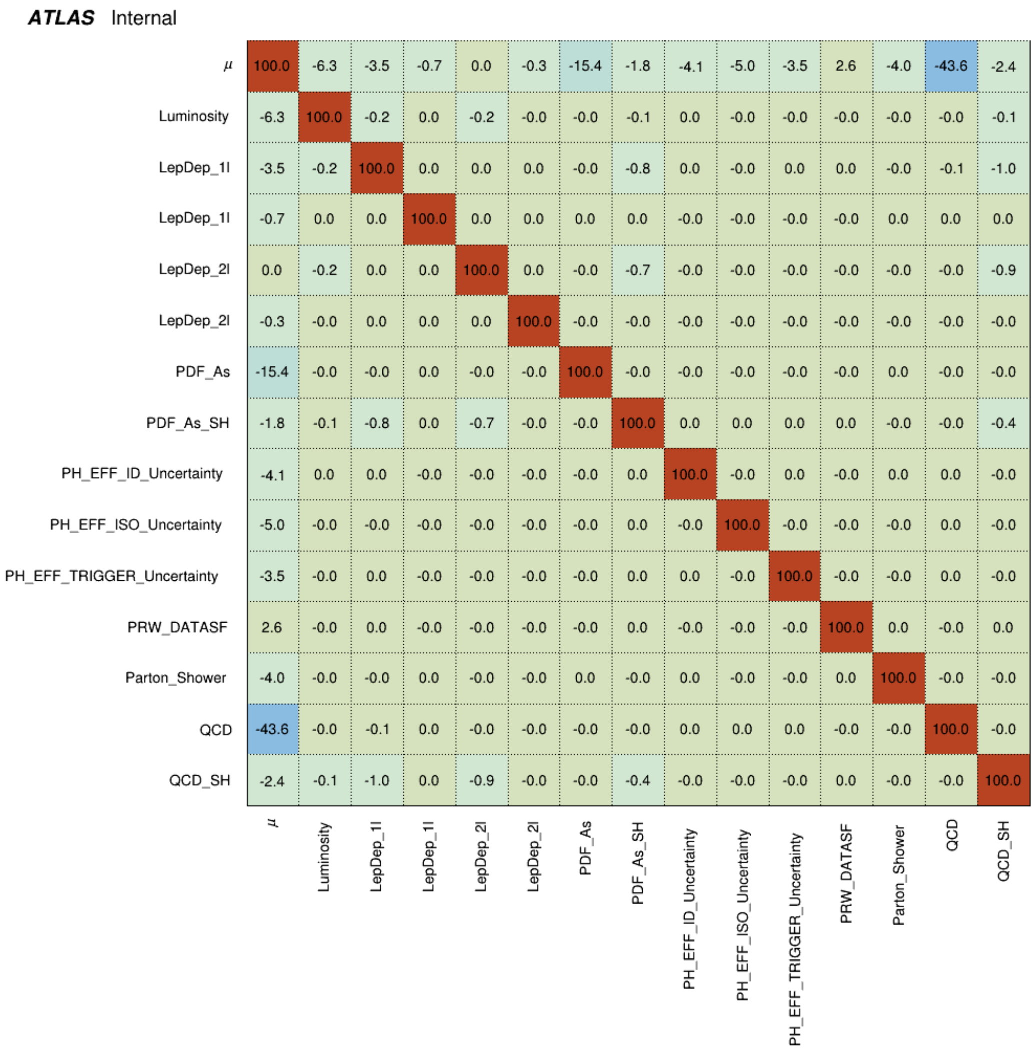


Figure 98: Correlation Matrix for (m_X, m_S)=(750,170) GeV mass.

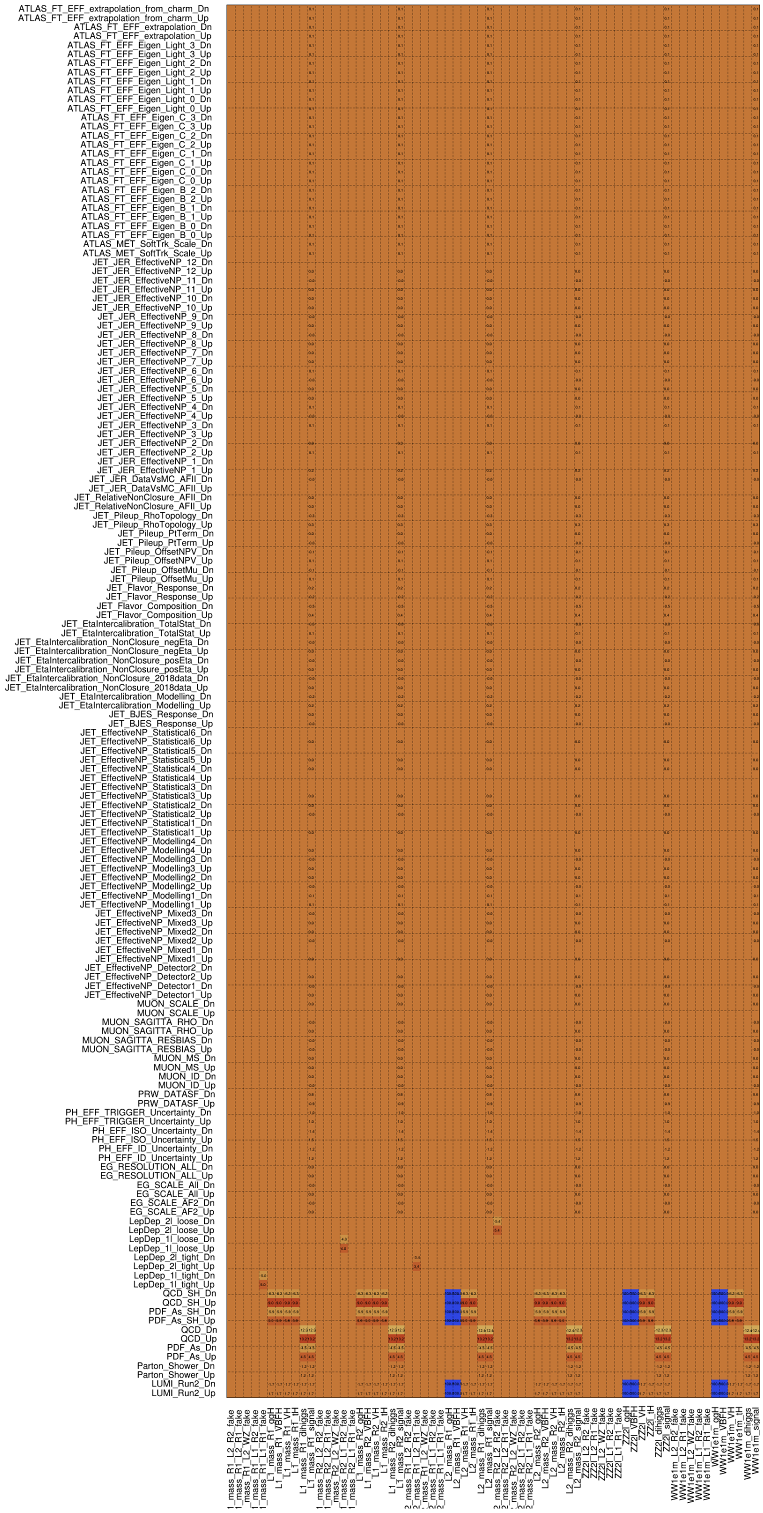


Figure 99: Normalization Factors for Nuisance parameters among all the regions for (m_X, m_S)=(750,170) GeV mass.

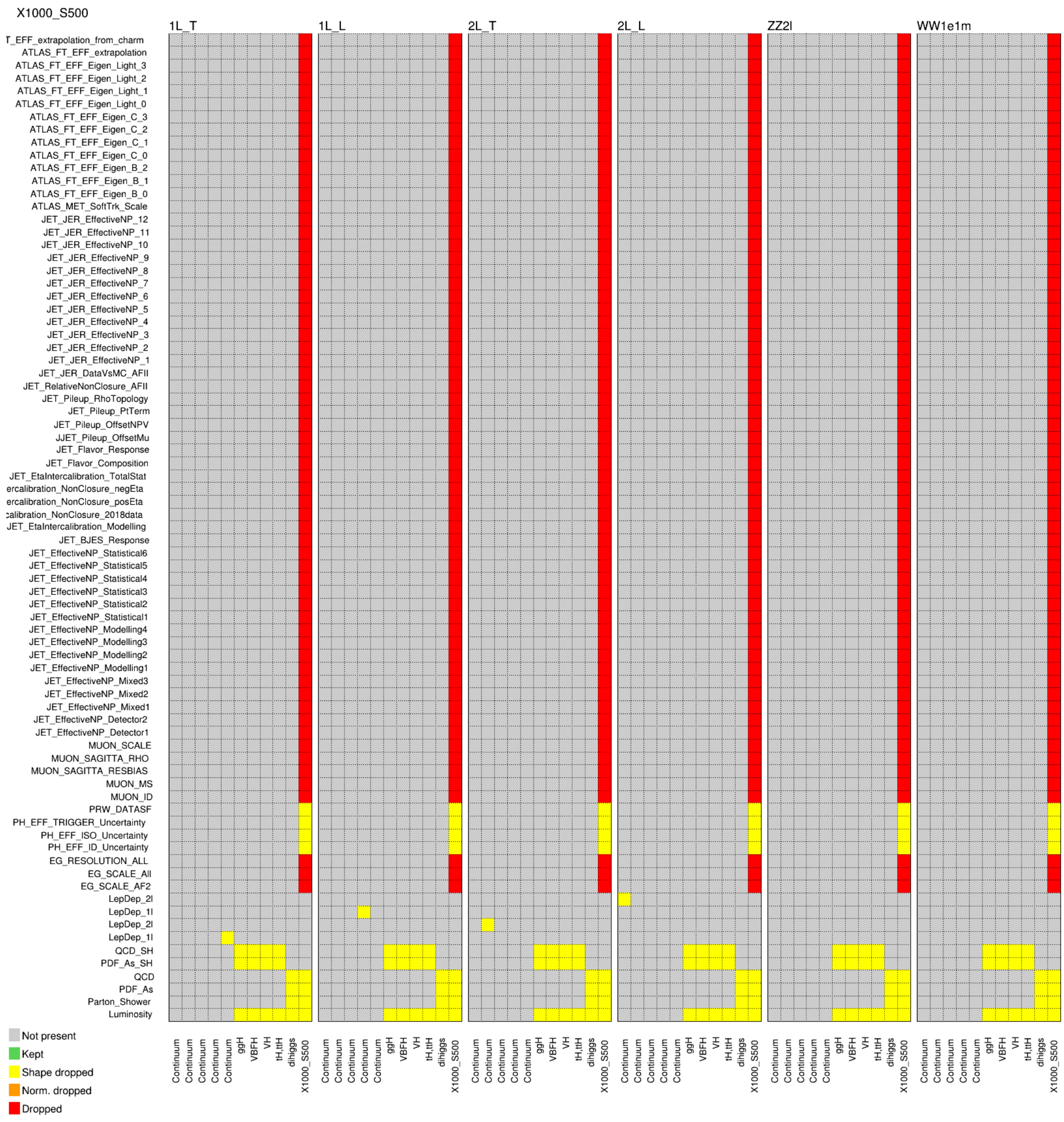


Figure 100: Pruning situation among all the regions for (m_X, m_S)=(750,170) GeV mass.

1001 ## A.12 (m_X, m_S)=(750, 200) GeV

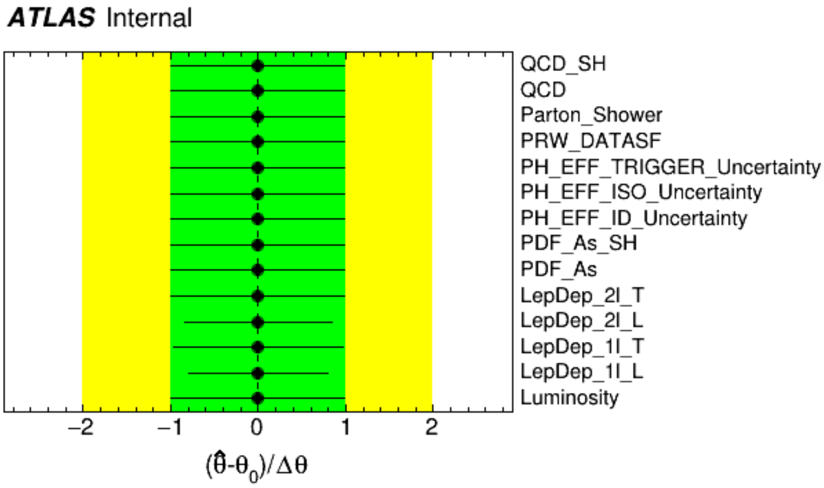


Figure 101: Nuisance parameters pull plot for (m_X, m_S)=(750,200) GeV mass

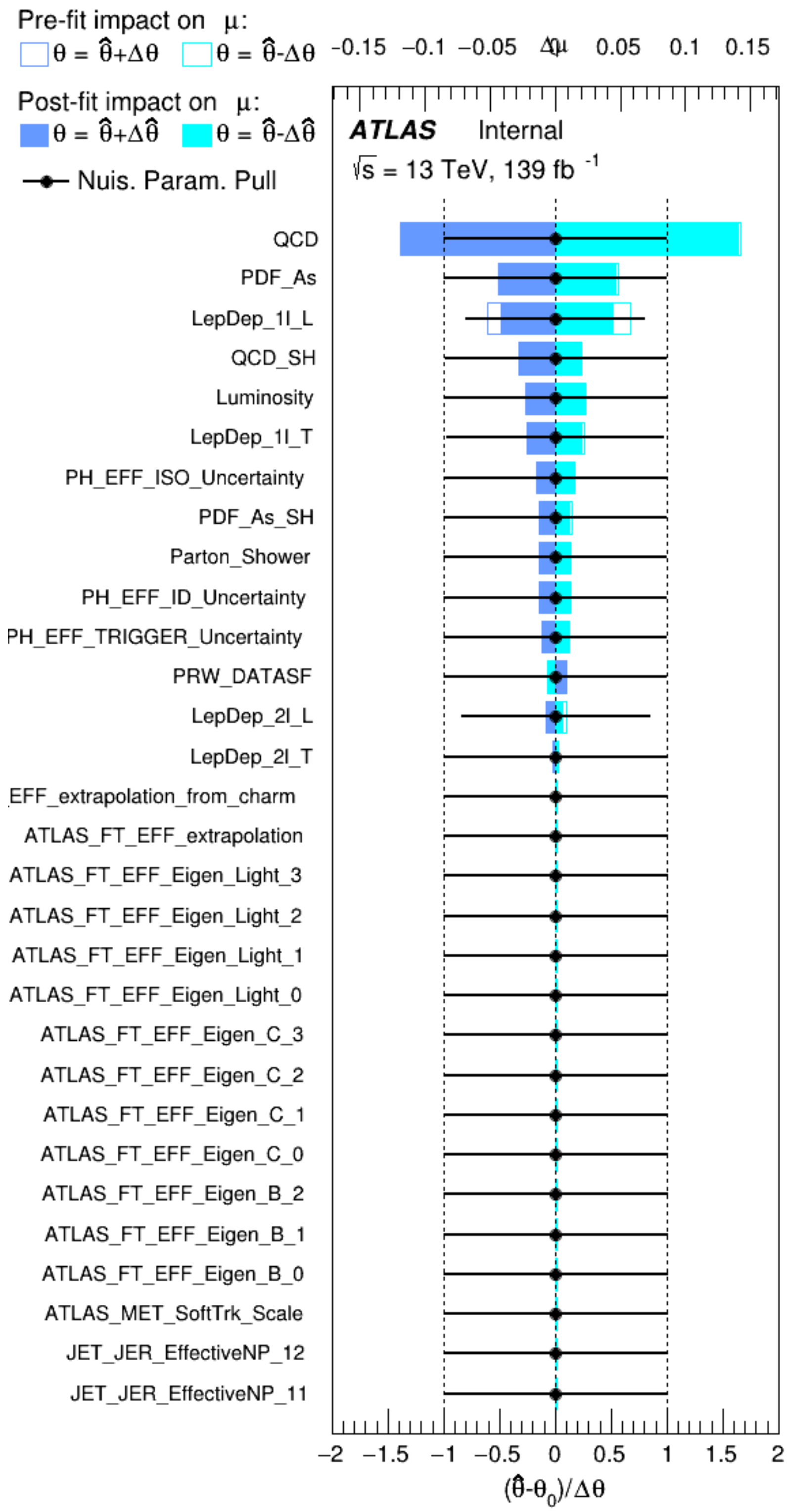


Figure 102: Nuisance parameters rankings for (m_X, m_S)=(750,200) GeV mass

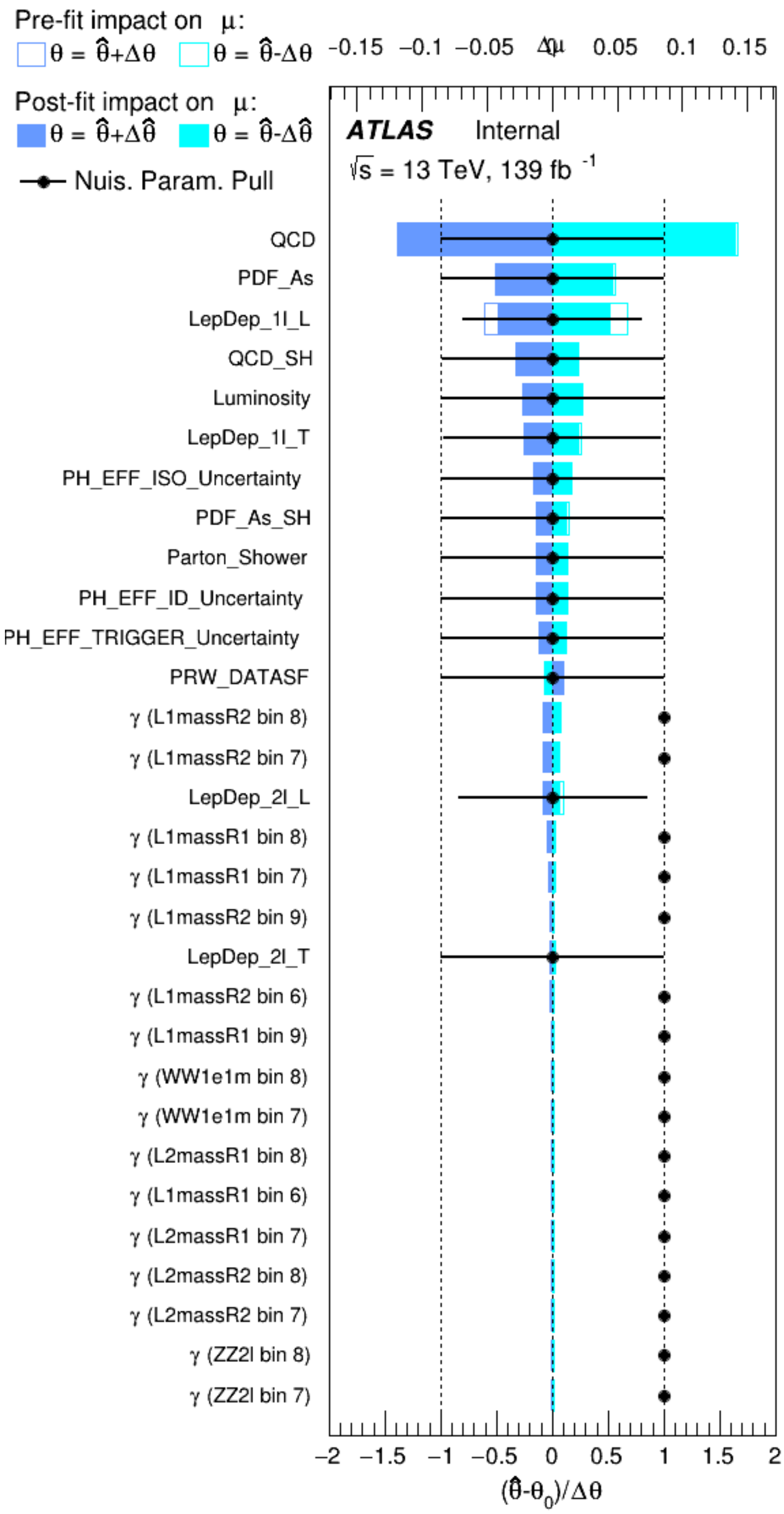


Figure 103: Rankings including systematics and gammas for (m_X, m_S)=(750,200) GeV mass

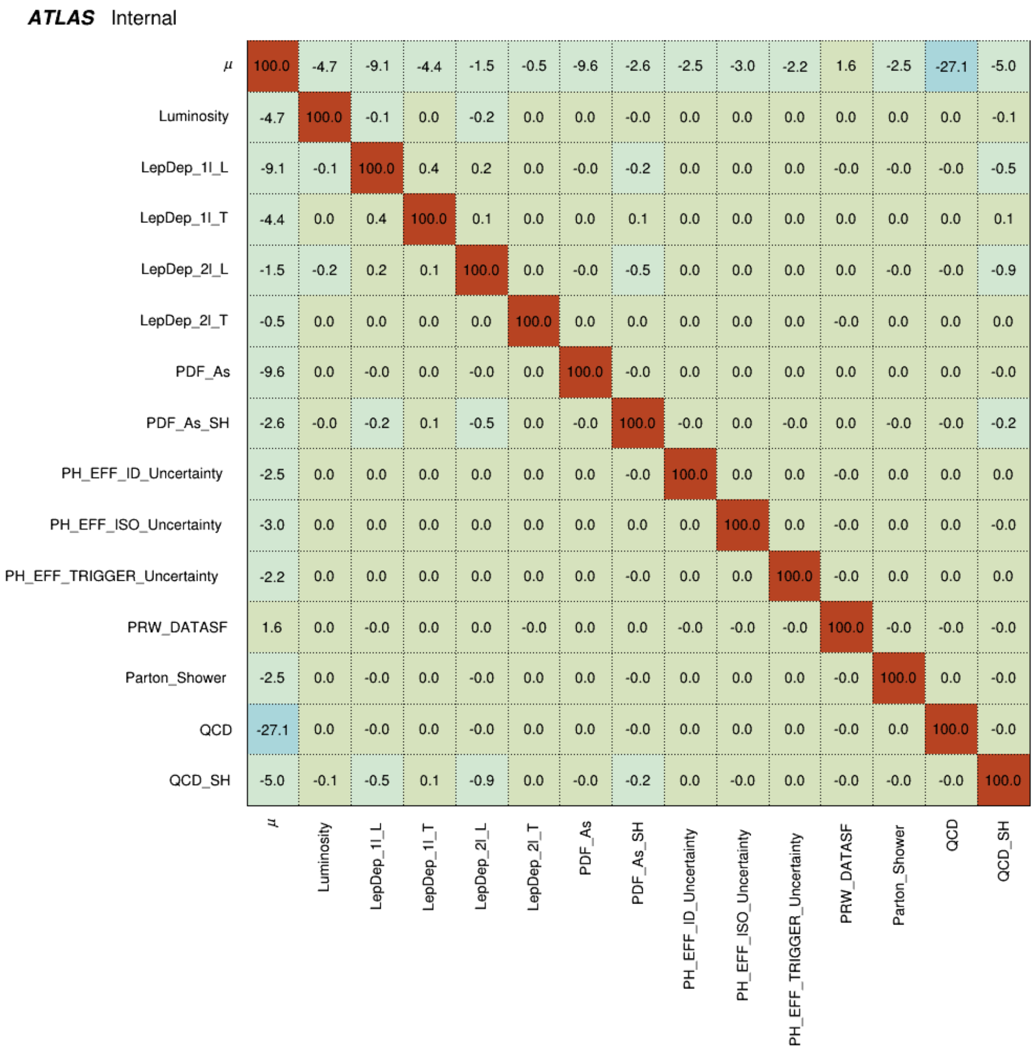


Figure 104: Correlation Matrix for (m_X, m_S)=(750,200) GeV mass.

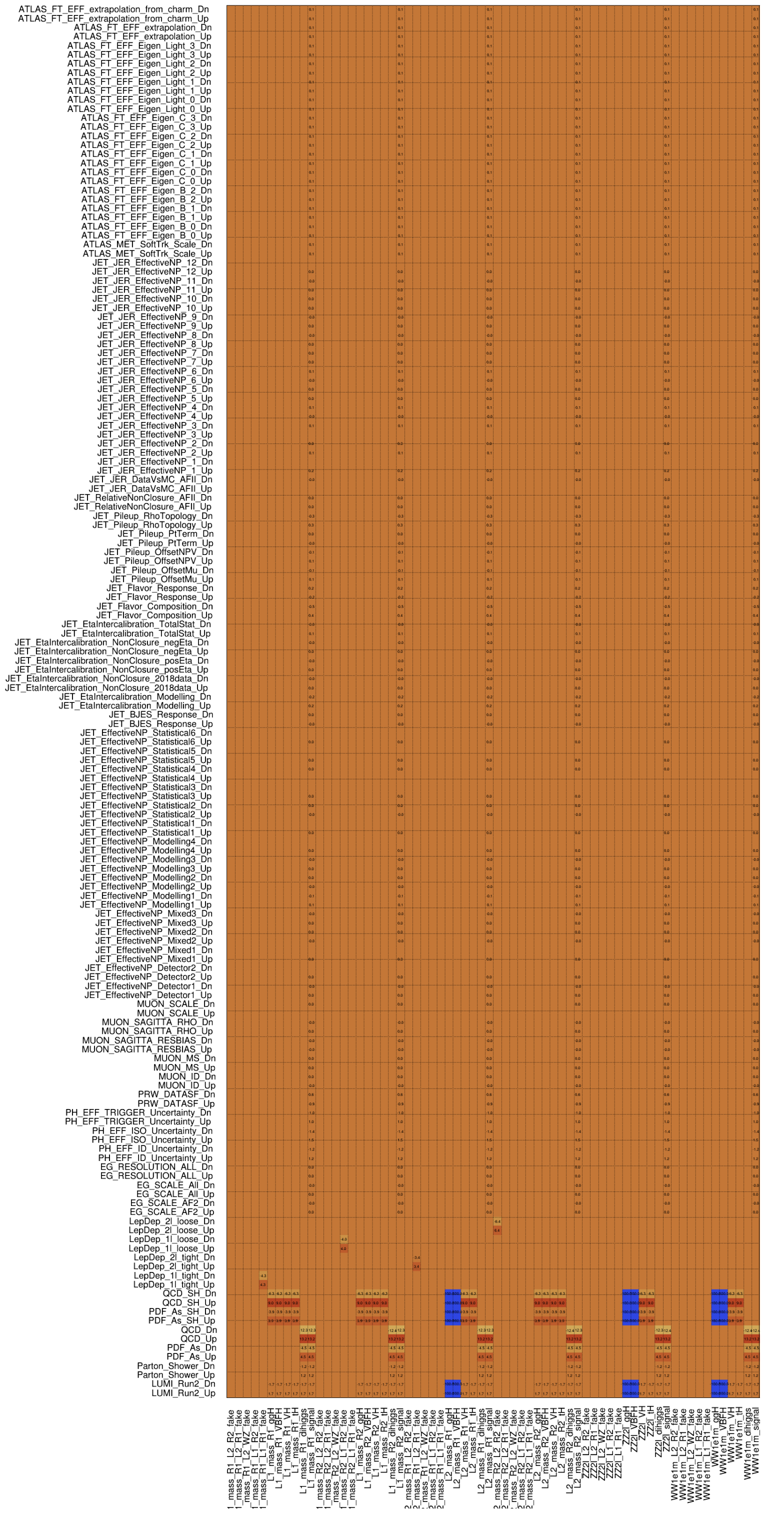


Figure 105: Normalization Factors for Nuisance parameters among all the regions for (m_X, m_S)=(750,200) GeV mass.

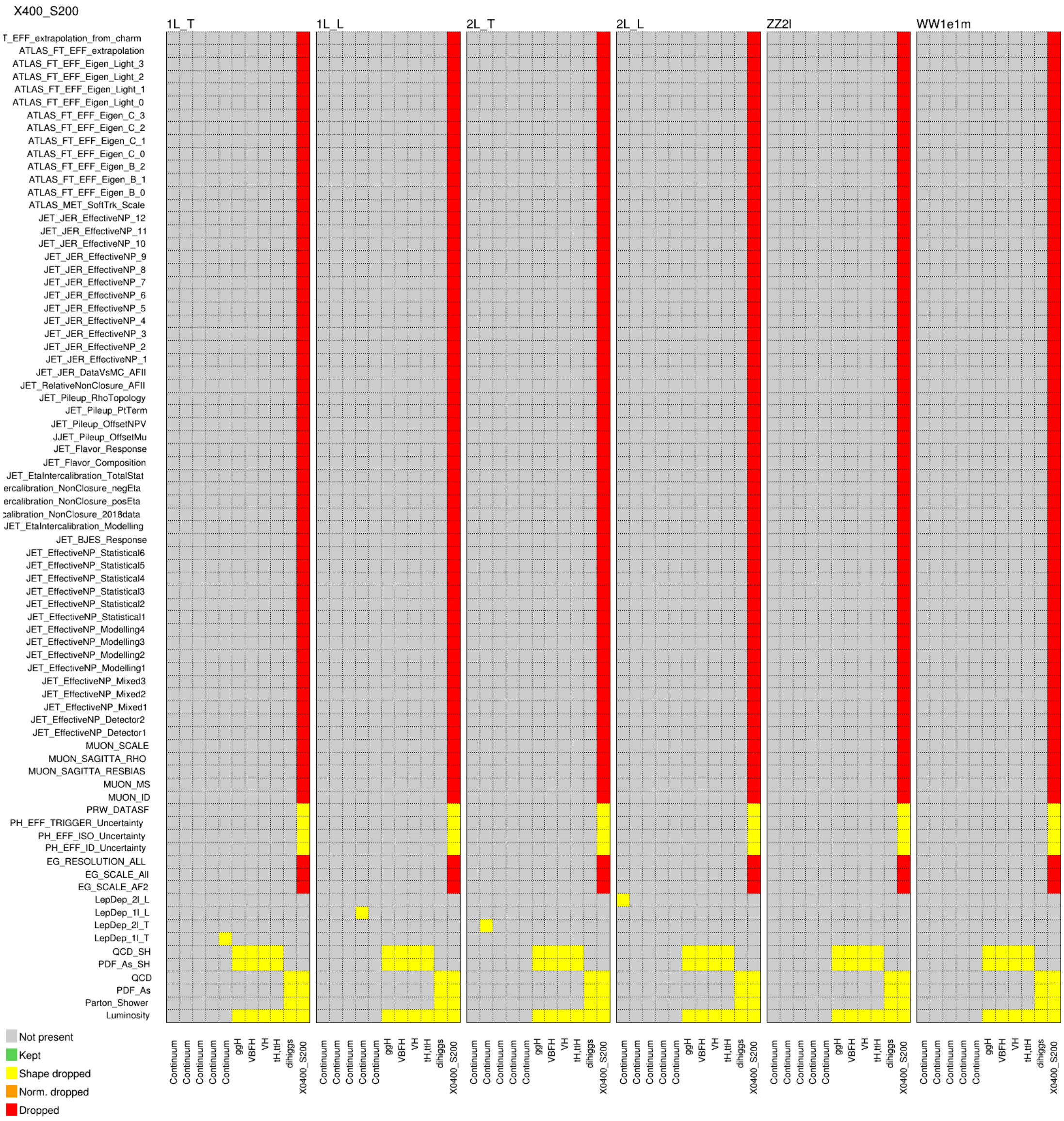


Figure 106: Pruning situation among all the regions for (m_X, m_S)=(750,200) GeV mass.

1002 ### A.13 (m_X, m_S)=(750, 300) GeV

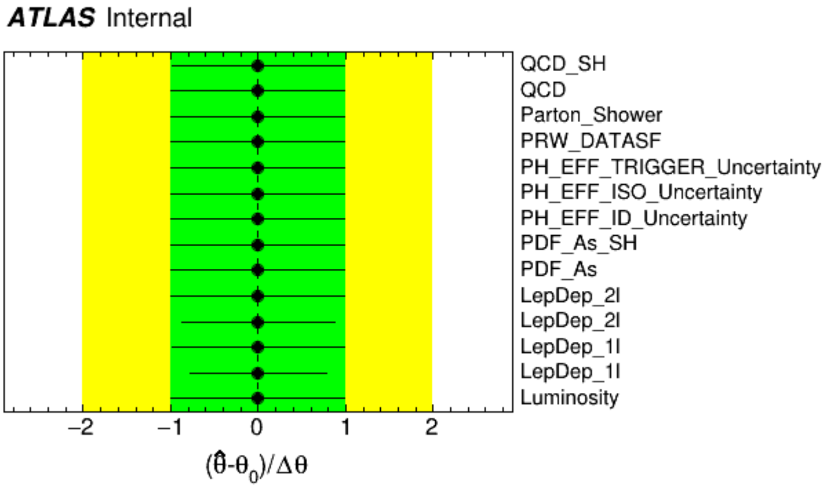


Figure 107: Nuisance parameters pull plot for (m_X, m_S)=(750,300) GeV mass

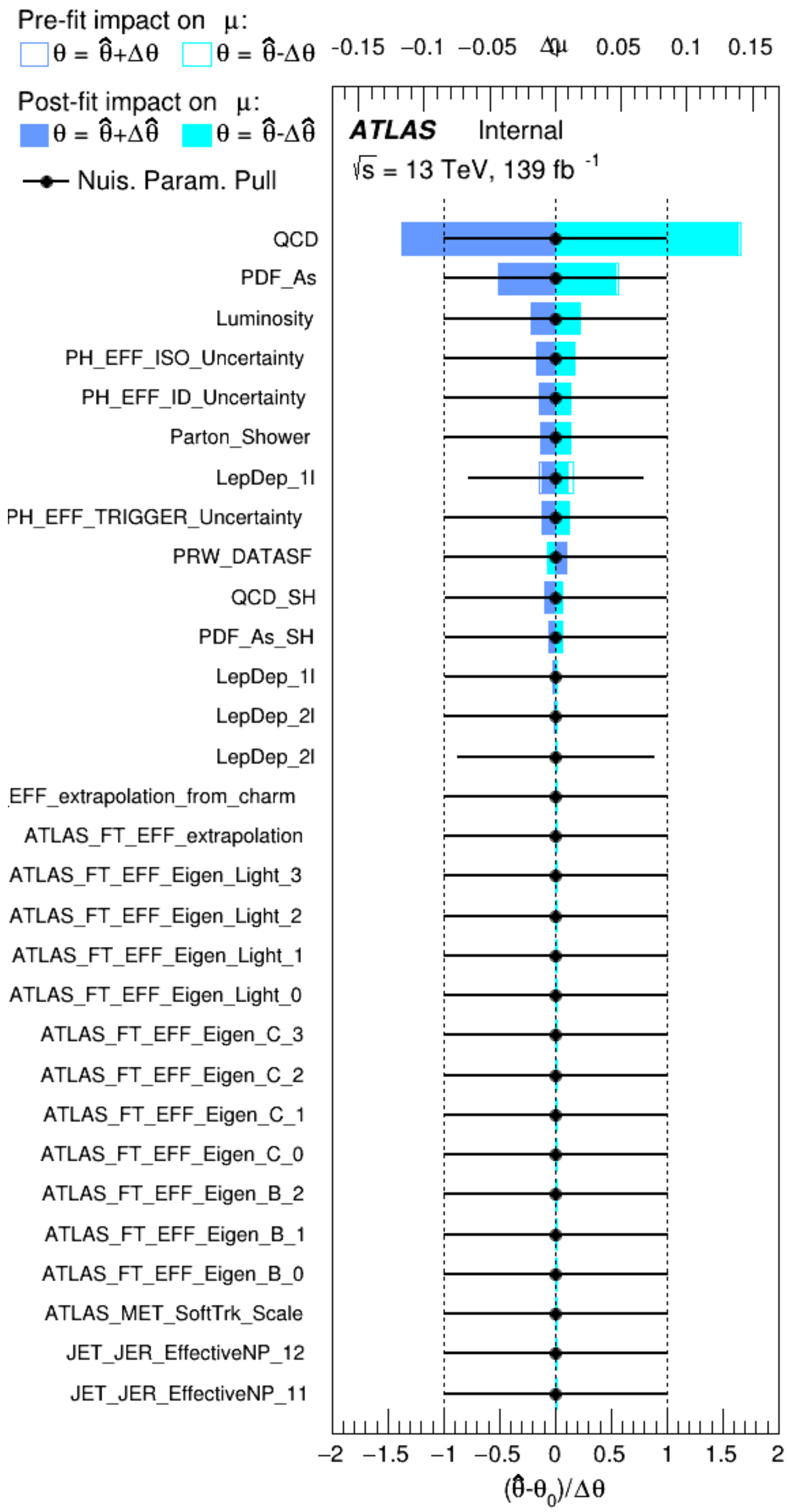


Figure 108: Nuisance parameters rankings for (m_X, m_S)=(750,300) GeV mass

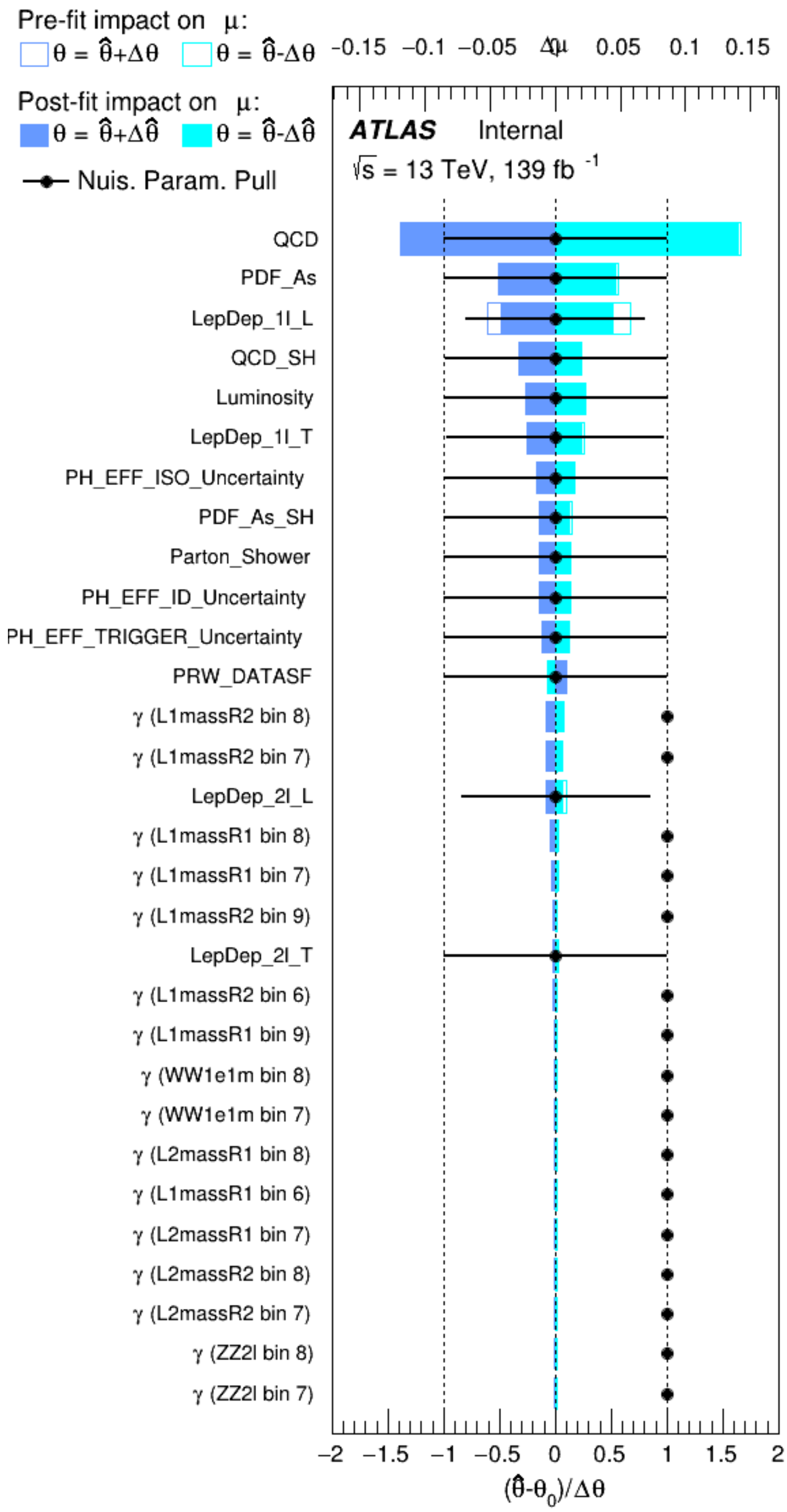


Figure 109: Rankings including systematics and gammas for (m_X, m_S)=(750,300) GeV mass

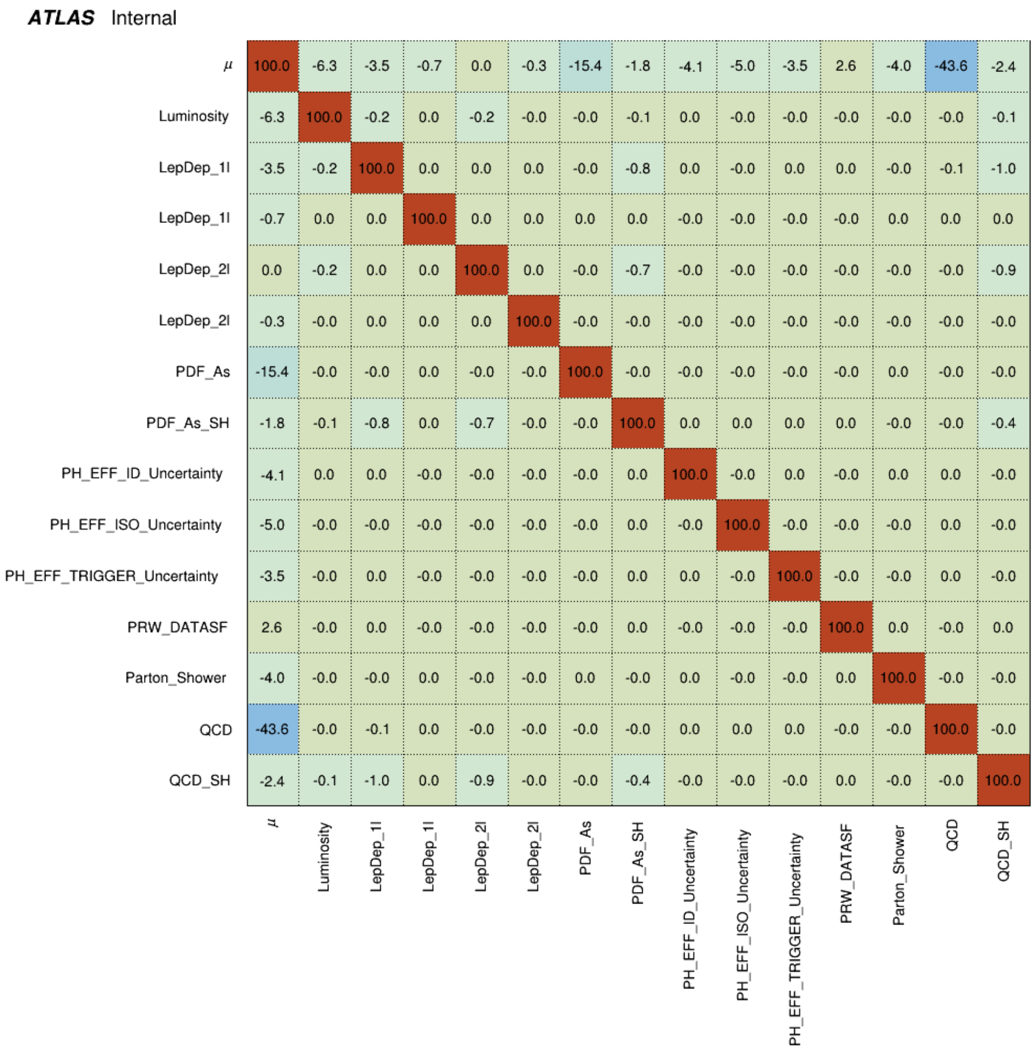


Figure 110: Correlation Matrix for (m_X, m_S)=(750,300) GeV mass.

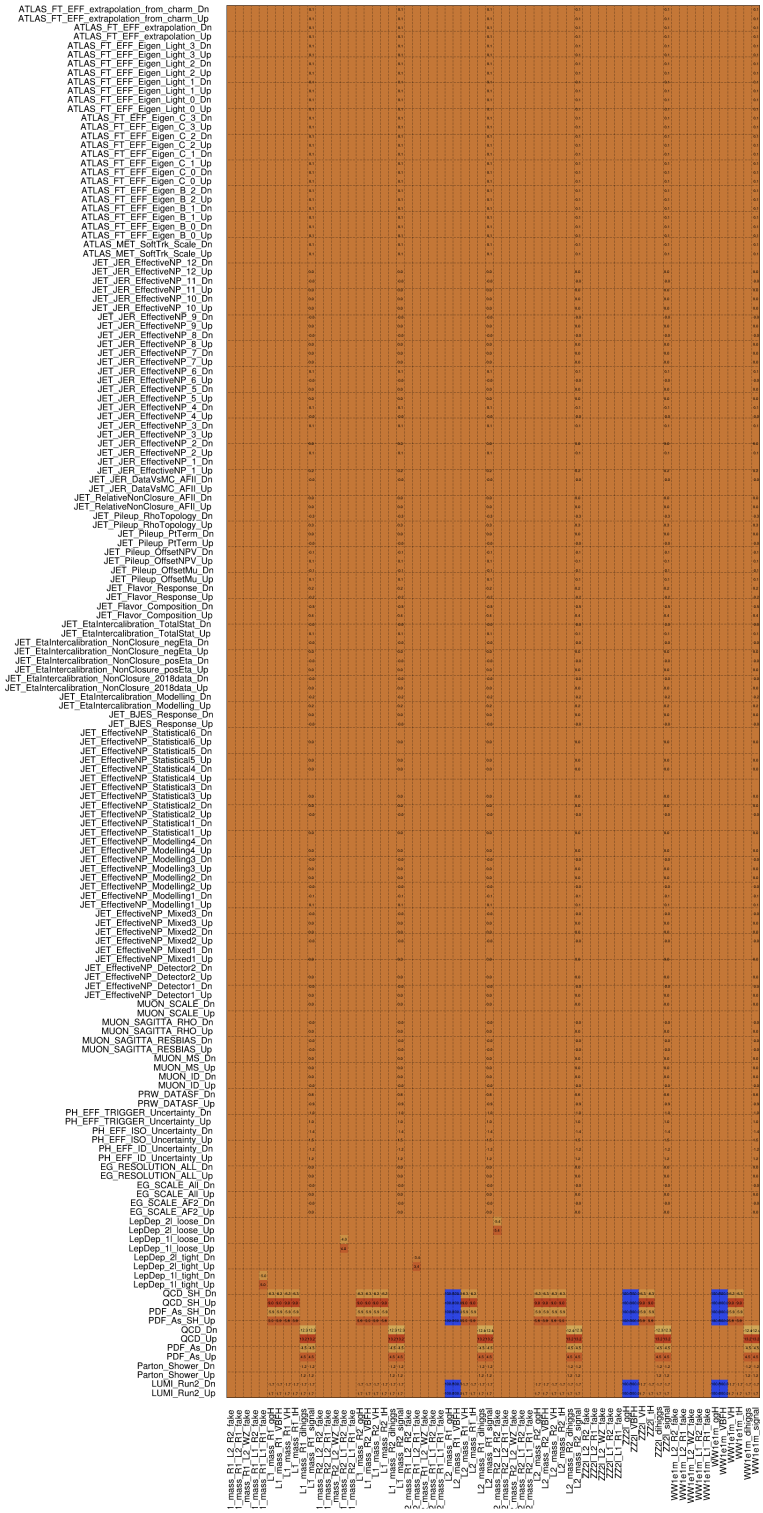


Figure 111: Normalization Factors for Nuisance parameters among all the regions for (m_X, m_S)=(750,300) GeV mass.

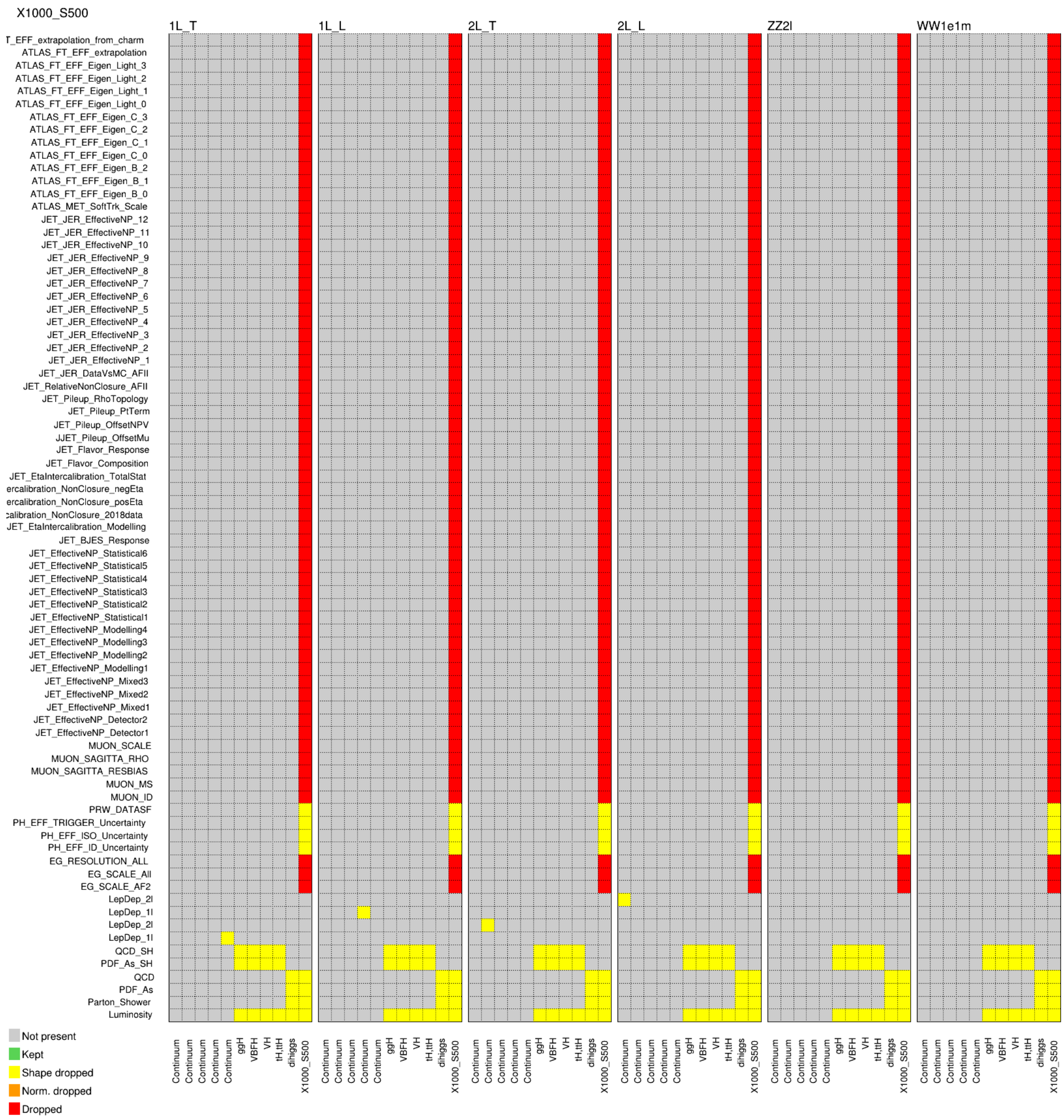


Figure 112: Pruning situation among all the regions for (m_X, m_S)=(750,300) GeV mass.

1003 ## A.14 (m_X, m_S)=(750, 400) GeV

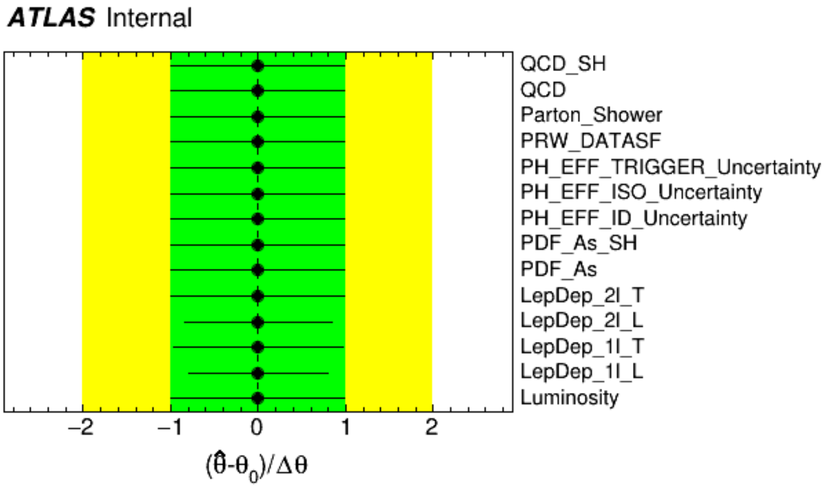


Figure 113: Nuisance parameters pull plot for (m_X, m_S)=(750,400) GeV mass

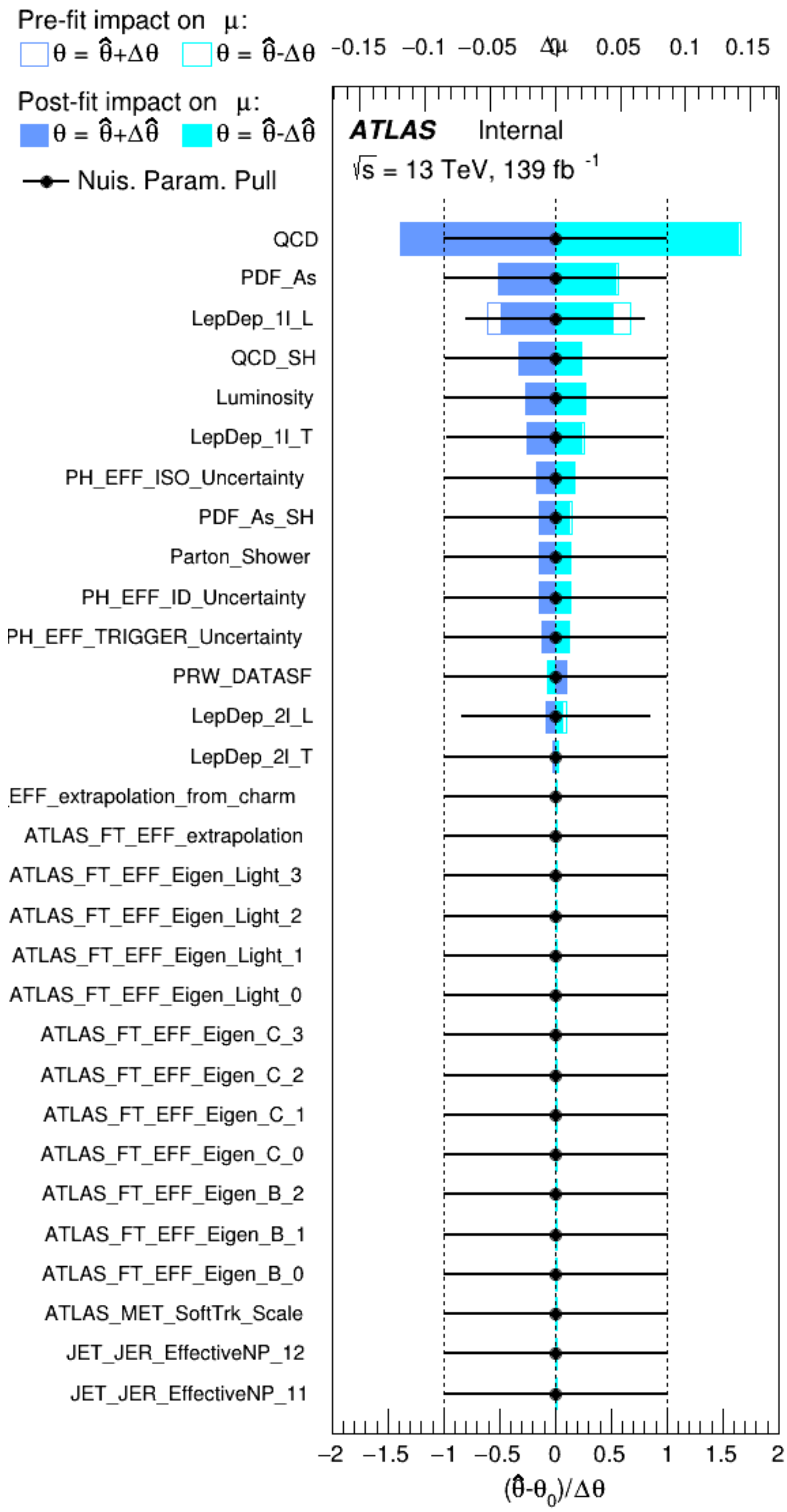


Figure 114: Nuisance parameters rankings for (m_X, m_S)=(750,400) GeV mass

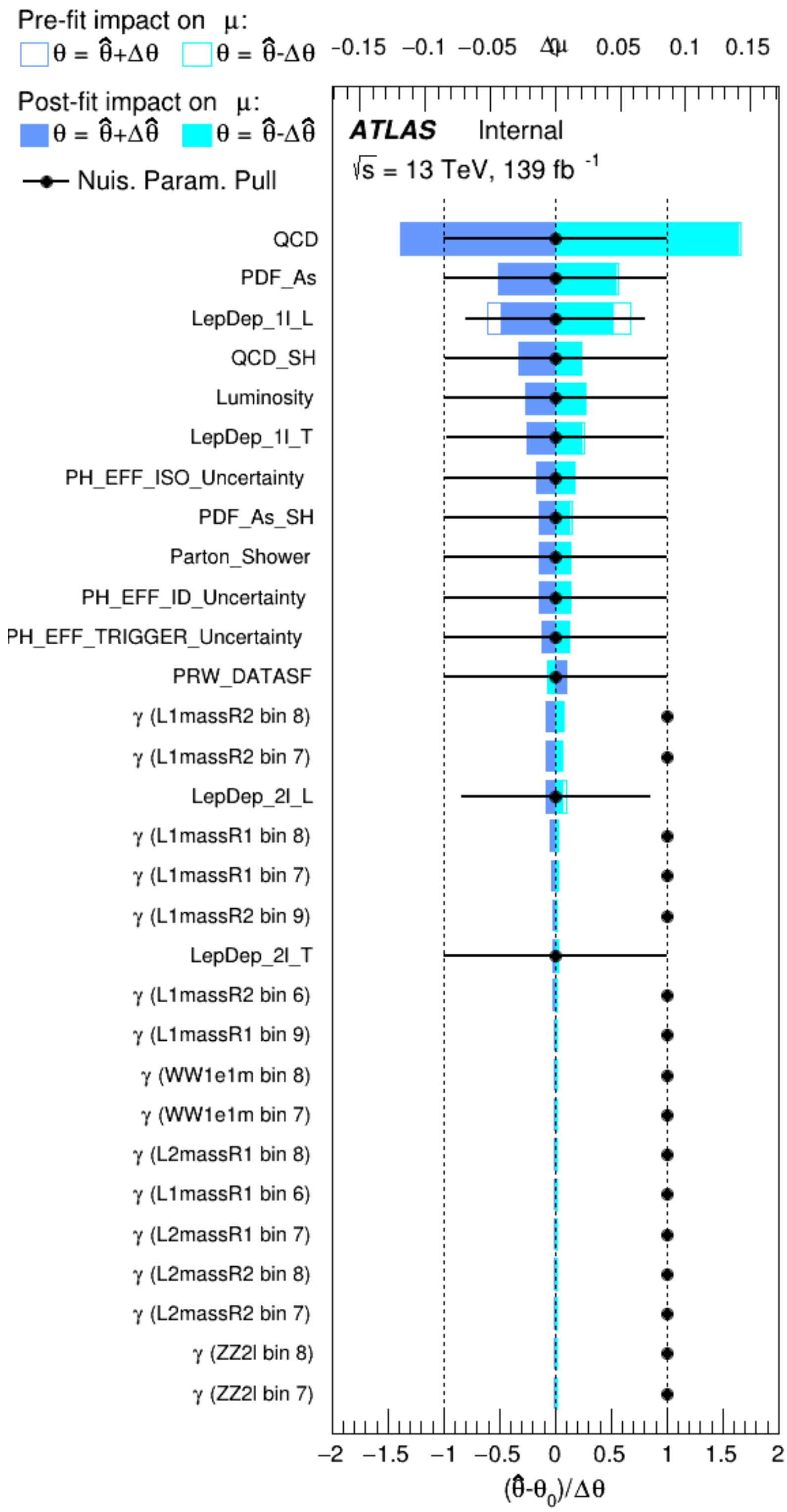


Figure 115: Rankings including systematics and gammas for (m_X, m_S)=(750,400) GeV mass

ATLAS Internal

	μ	Luminosity	LepDep_1l_L	LepDep_1l_T	LepDep_2l_L	LepDep_2l_T	PDF_As	PDF_As_SH	PH_EFF_ID_Uncertainty	PH_EFF_ISO_Uncertainty	PH_EFF_TRIGGER_Uncertainty	PRW_DATASF	Parton_Shower	QCD	QCD_SH
μ	100.0	-4.7	-9.1	-4.4	-1.5	-0.5	-9.6	-2.6	-2.5	-3.0	-2.2	1.6	-2.5	-27.1	-5.0
Luminosity	-4.7	100.0	-0.1	0.0	-0.2	0.0	0.0	-0.0	0.0	0.0	0.0	0.0	0.0	0.0	-0.1
LepDep_1l_L	-9.1	-0.1	100.0	0.4	0.2	0.0	-0.0	-0.2	0.0	0.0	0.0	-0.0	-0.0	-0.0	-0.5
LepDep_1l_T	-4.4	0.0	0.4	100.0	0.1	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.1
LepDep_2l_L	-1.5	-0.2	0.2	0.1	100.0	0.0	-0.0	-0.5	0.0	0.0	0.0	0.0	-0.0	-0.0	-0.9
LepDep_2l_T	-0.5	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	-0.0	0.0	0.0	0.0
PDF_As	-9.6	0.0	-0.0	0.0	-0.0	0.0	100.0	-0.0	0.0	0.0	0.0	0.0	0.0	0.0	-0.0
PDF_As_SH	-2.6	-0.0	-0.2	0.1	-0.5	0.0	-0.0	100.0	-0.0	0.0	-0.0	0.0	-0.0	-0.0	-0.2
PH_EFF_ID_Uncertainty	-2.5	0.0	0.0	0.0	0.0	0.0	0.0	-0.0	100.0	0.0	0.0	-0.0	0.0	0.0	0.0
PH_EFF_ISO_Uncertainty	-3.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	-0.0	0.0	0.0	-0.0
PH_EFF_TRIGGER_Uncertainty	-2.2	0.0	0.0	0.0	0.0	0.0	0.0	-0.0	0.0	0.0	100.0	-0.0	0.0	0.0	0.0
PRW_DATASF	1.6	0.0	-0.0	0.0	0.0	-0.0	0.0	0.0	-0.0	-0.0	-0.0	100.0	-0.0	-0.0	-0.0
Parton_Shower	-2.5	0.0	-0.0	0.0	-0.0	0.0	0.0	-0.0	0.0	0.0	0.0	-0.0	100.0	0.0	-0.0
QCD	-27.1	0.0	-0.0	0.0	-0.0	0.0	0.0	-0.0	0.0	0.0	0.0	-0.0	0.0	100.0	-0.0
QCD_SH	-5.0	-0.1	-0.5	0.1	-0.9	0.0	-0.0	-0.2	0.0	-0.0	0.0	-0.0	-0.0	-0.0	100.0

Figure 116: Correlation Matrix for (m_X, m_S)=(750,400) GeV mass.

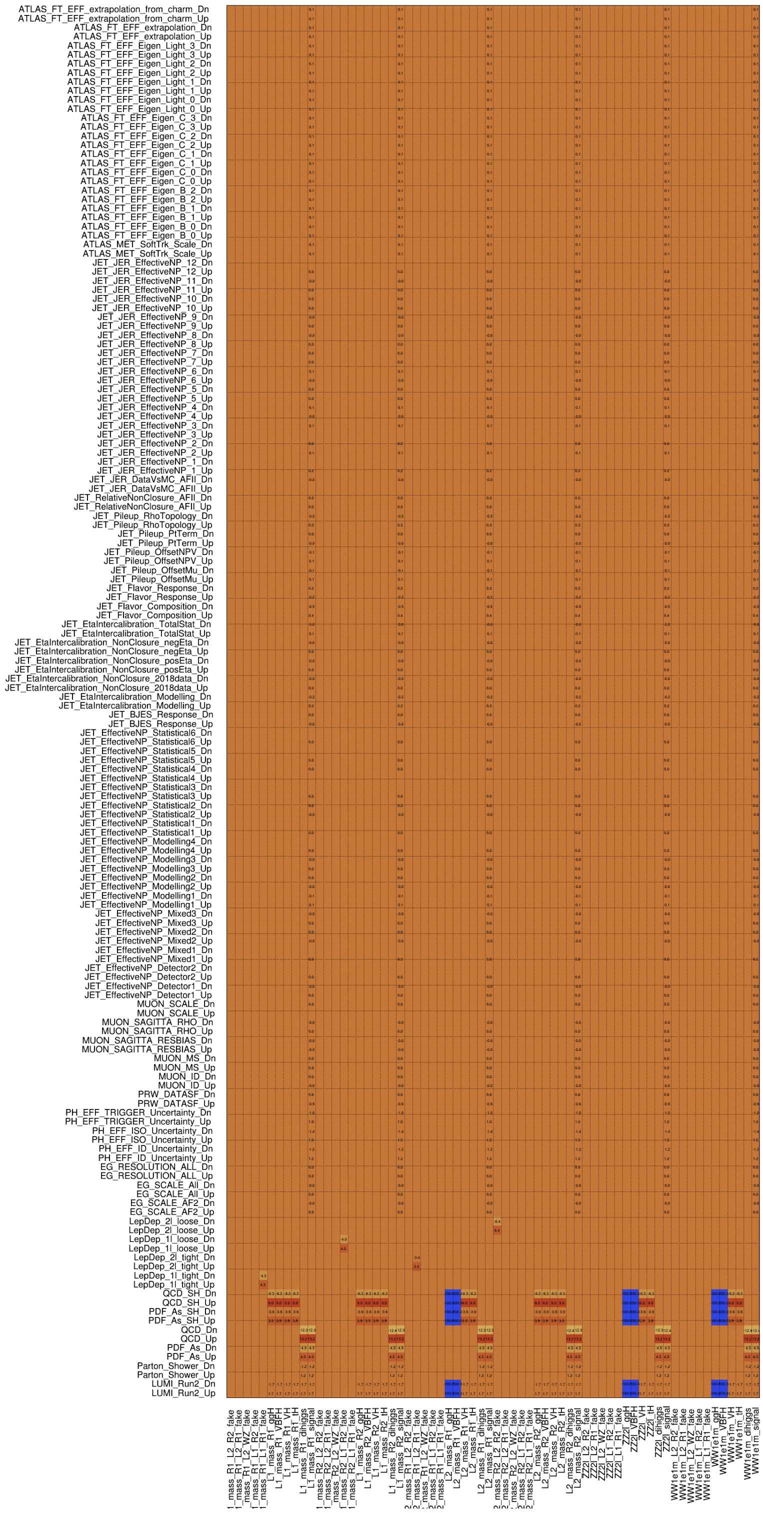


Figure 117: Normalization Factors for Nuisance parameters among all the regions for (m_X, m_S)=(750,400) GeV mass.

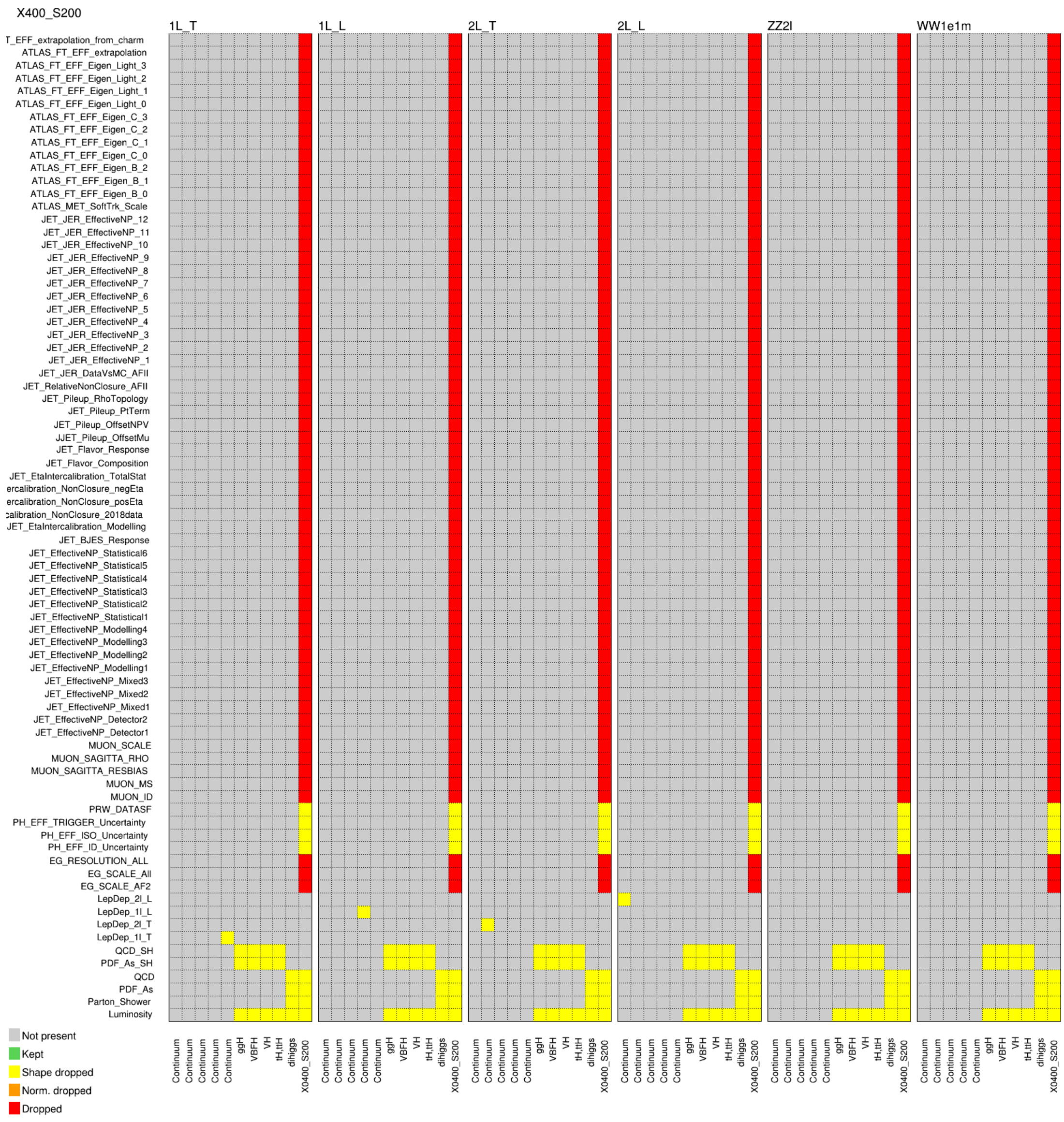


Figure 118: Pruning situation among all the regions for (m_X, m_S)=(750,400) GeV mass.

1004 ## A.15 (m_X, m_S)=(750, 500) GeV

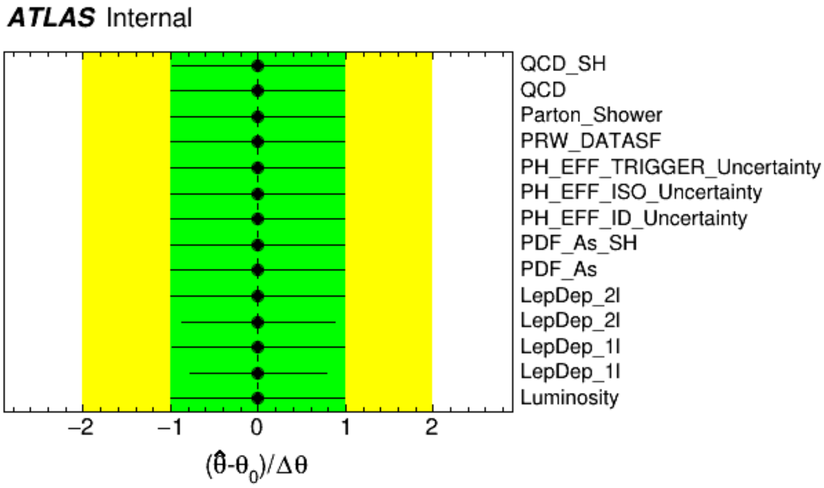


Figure 119: Nuisance parameters pull plot for (m_X, m_S)=(750,500) GeV mass

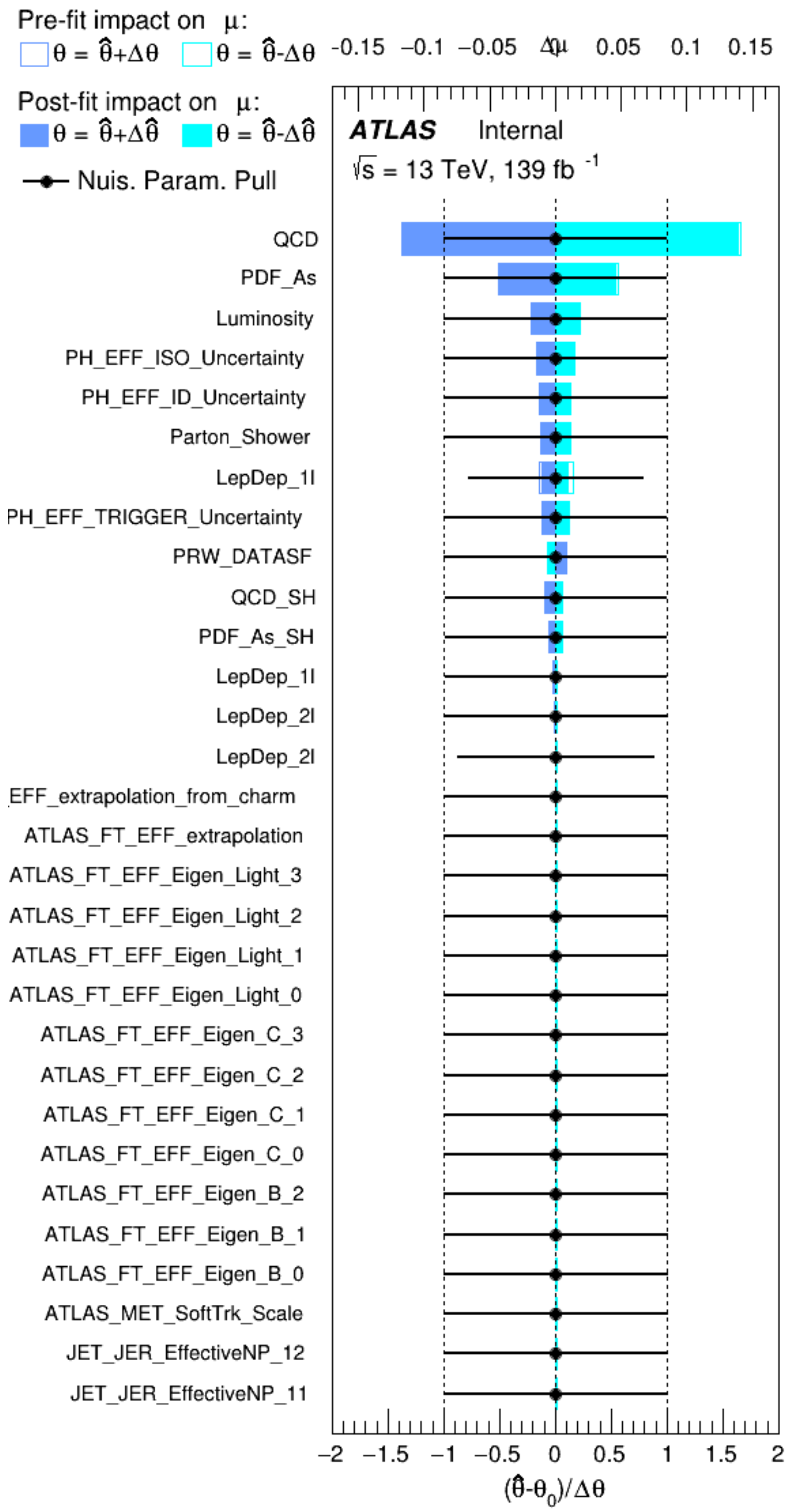


Figure 120: Nuisance parameters rankings for (m_X, m_S)=(750,500) GeV mass

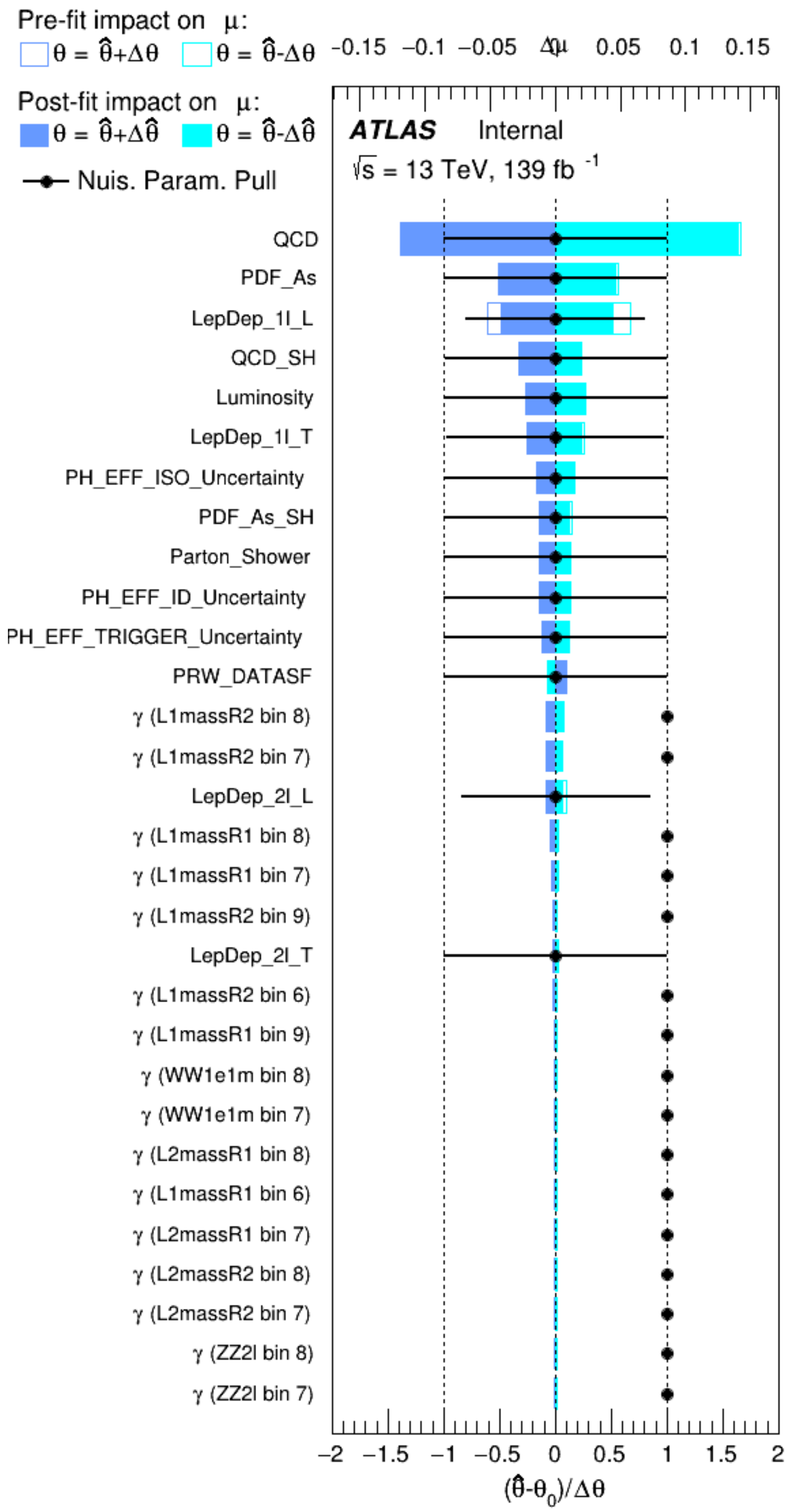


Figure 121: Rankings including systematics and gammas for (m_X, m_S)=(750,500) GeV mass

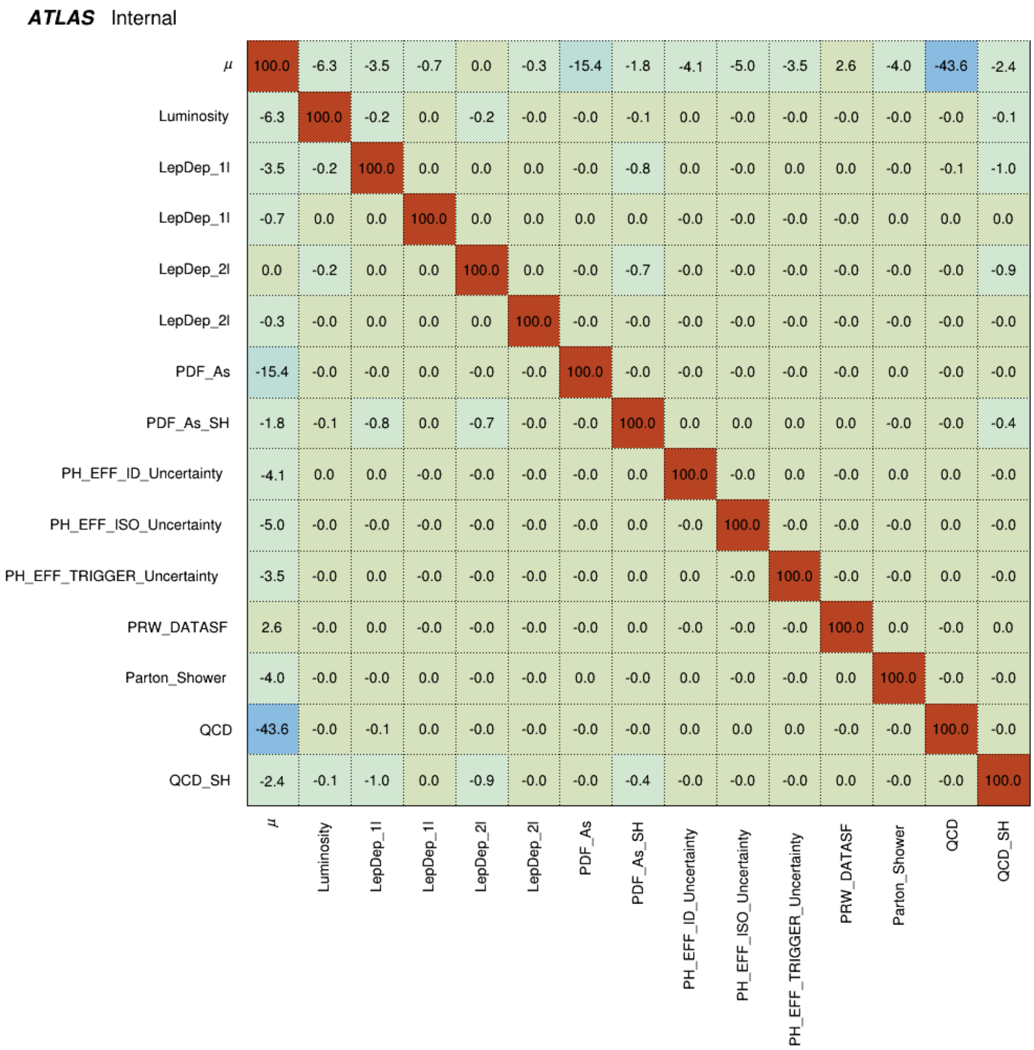


Figure 122: Correlation Matrix for (m_X, m_S)=(750,500) GeV mass.

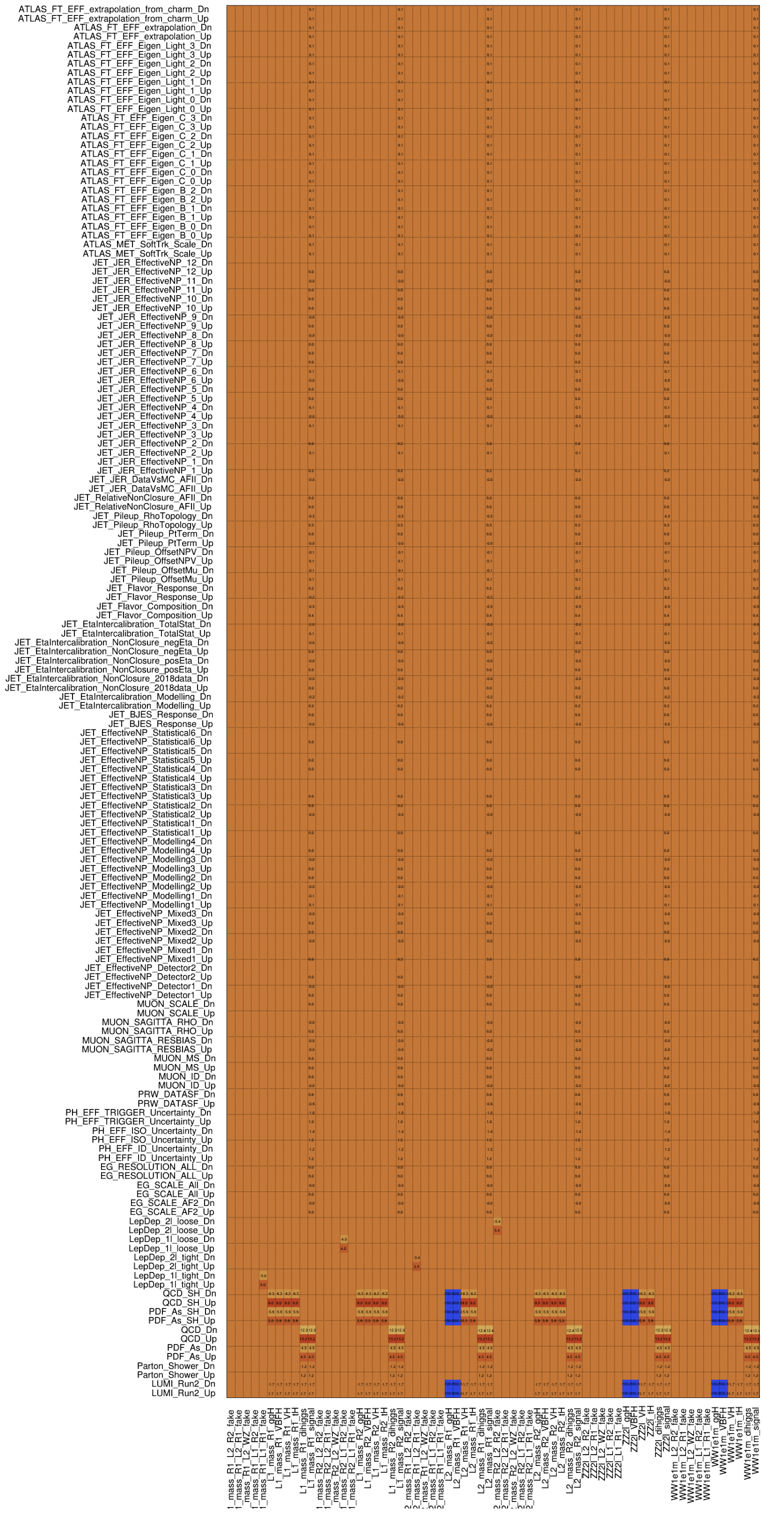


Figure 123: Normalization Factors for Nuisance parameters among all the regions for (m_X, m_S)=(750,500) GeV mass.

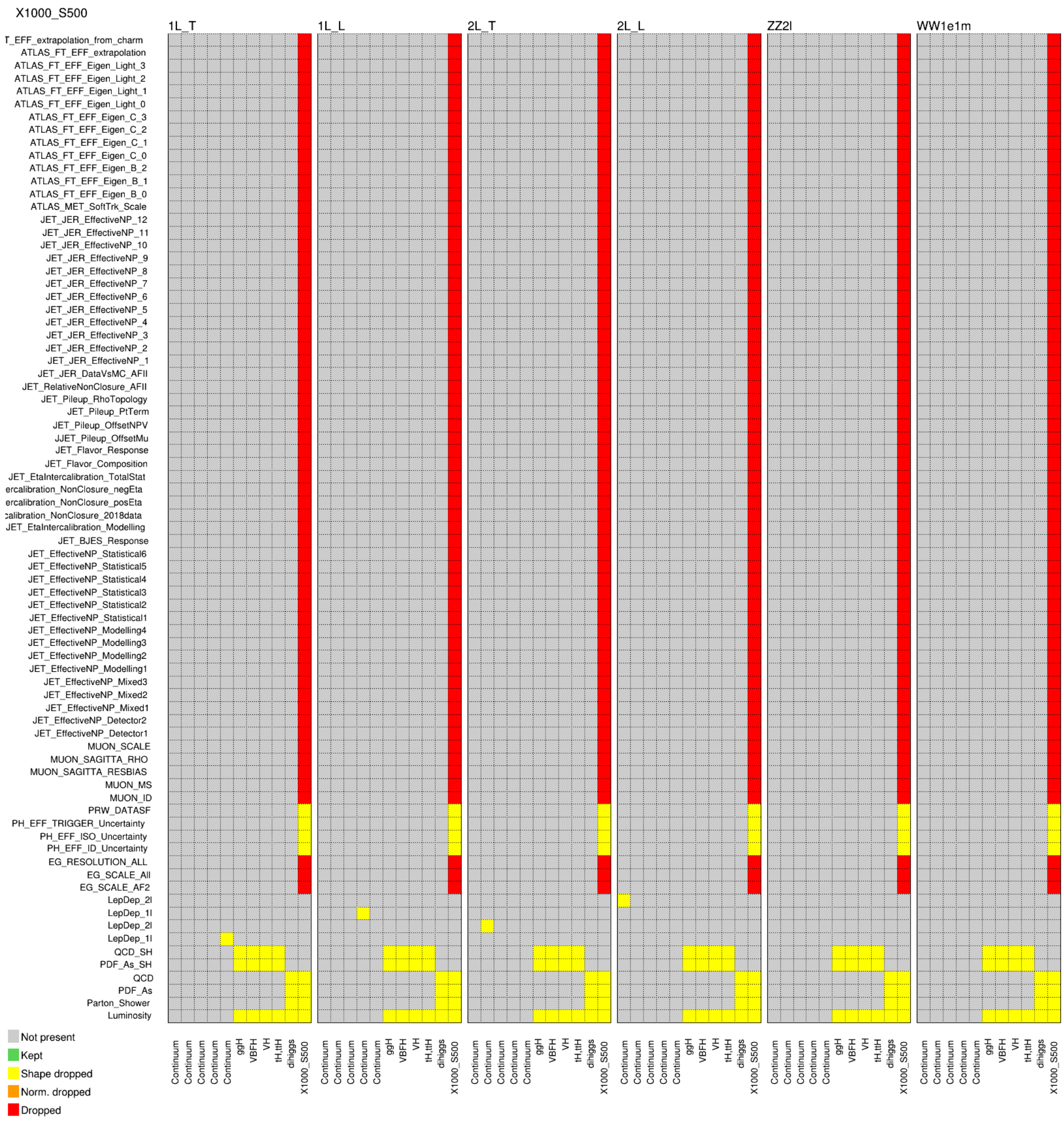


Figure 124: Pruning situation among all the regions for (m_X, m_S)=(750,500) GeV mass.

1005 ## A.16 (m_X, m_S)=(1000, 170) GeV

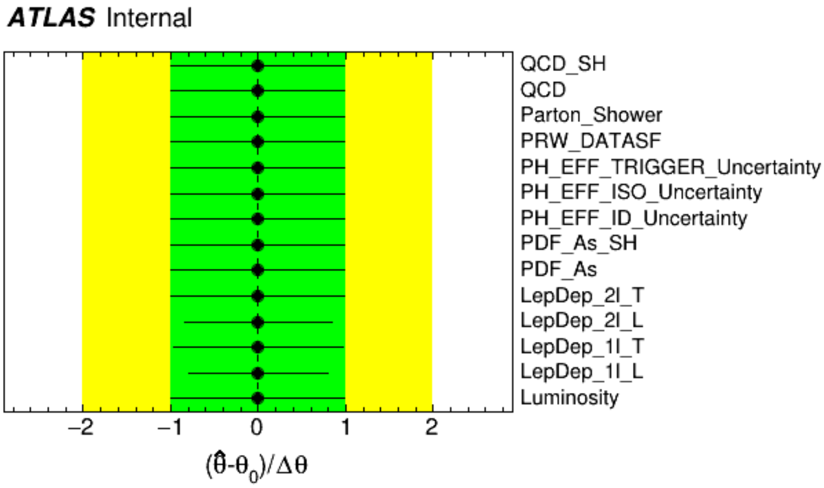


Figure 125: Nuisance parameters pull plot for (m_X, m_S)=(1000,170) GeV mass

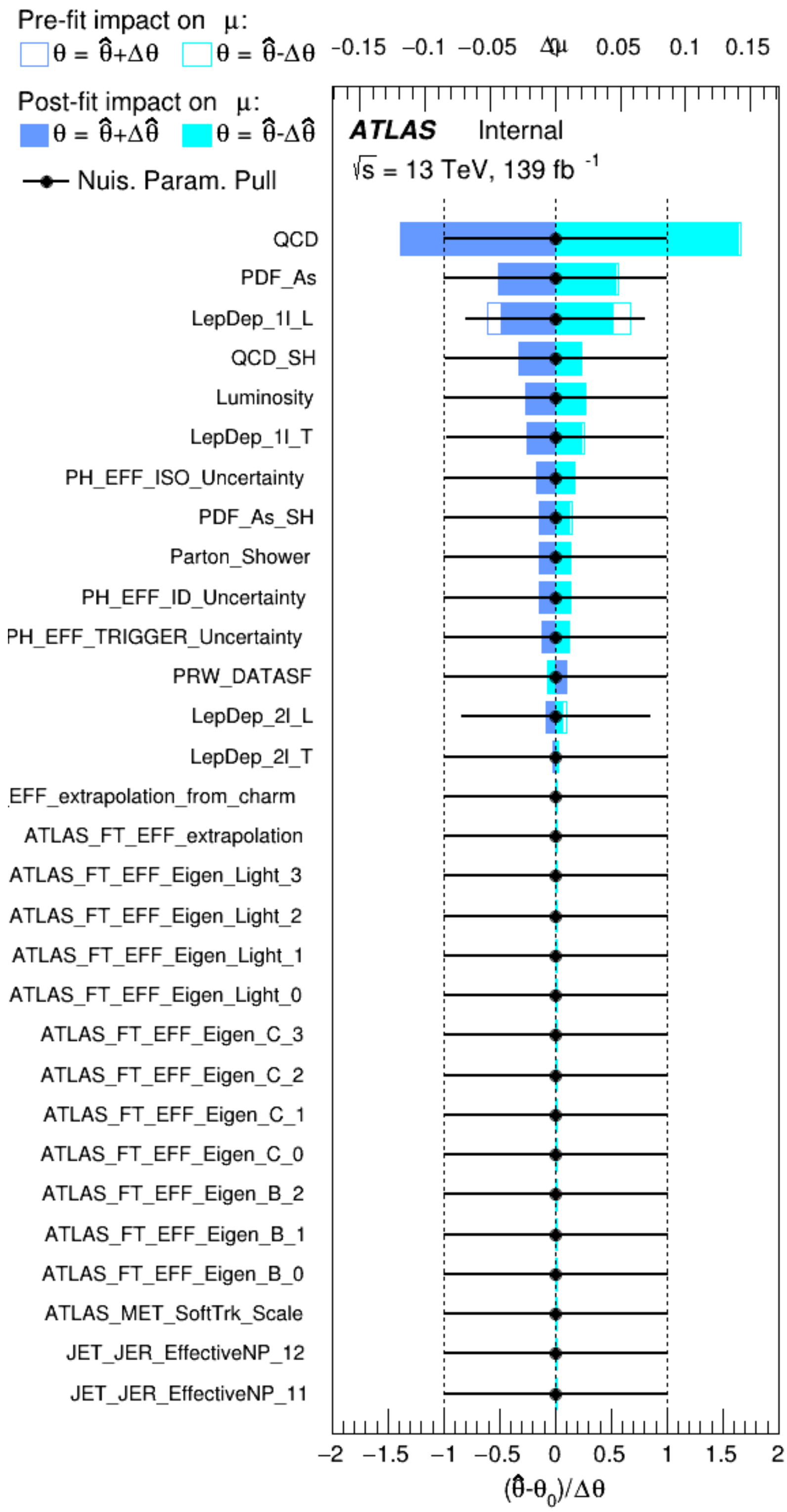


Figure 126: Nuisance parameters rankings for (m_X, m_S)=(1000,170) GeV mass

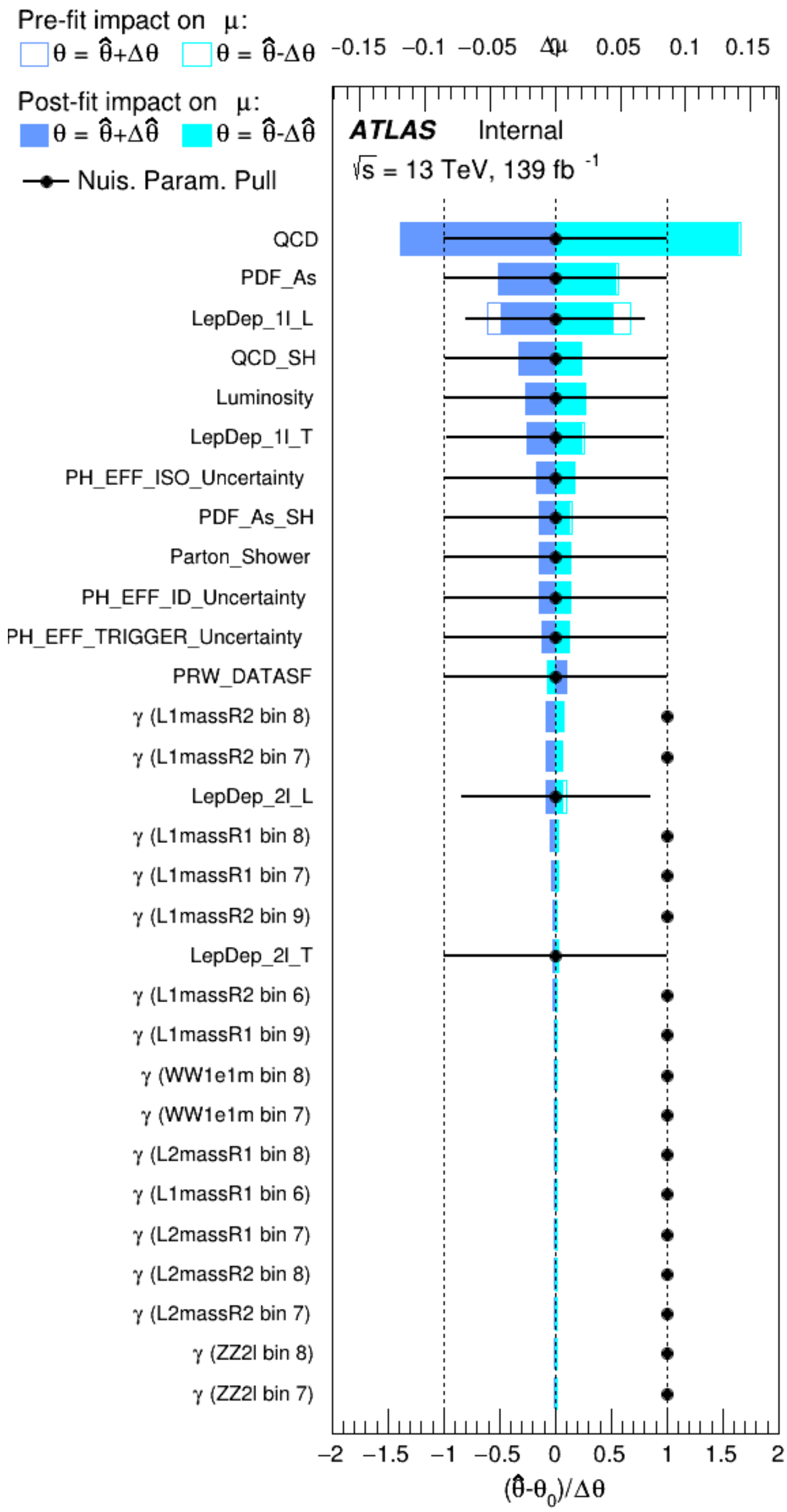


Figure 127: Rankings including systematics and gammas for (m_X, m_S)=(1000,170) GeV mass

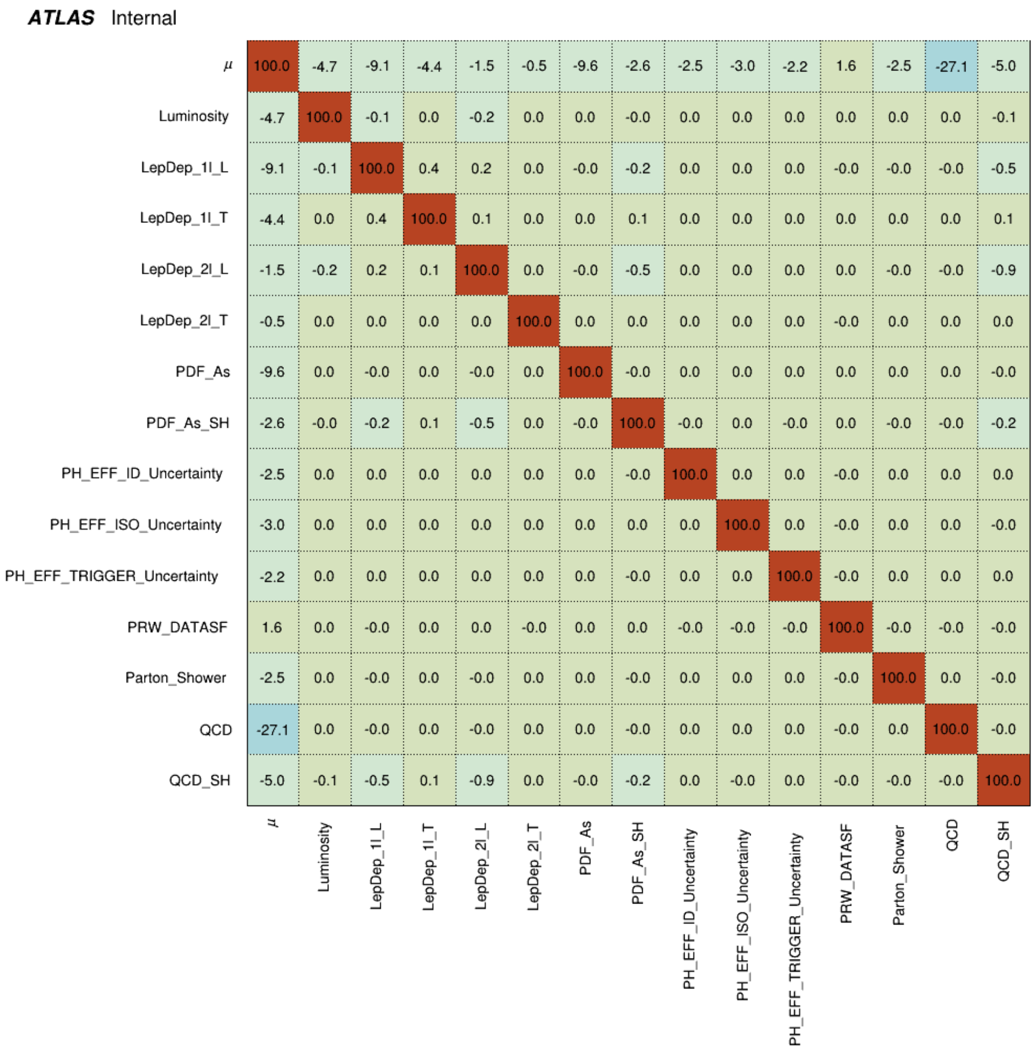


Figure 128: Correlation Matrix for (m_X, m_S)=(1000,170) GeV mass.

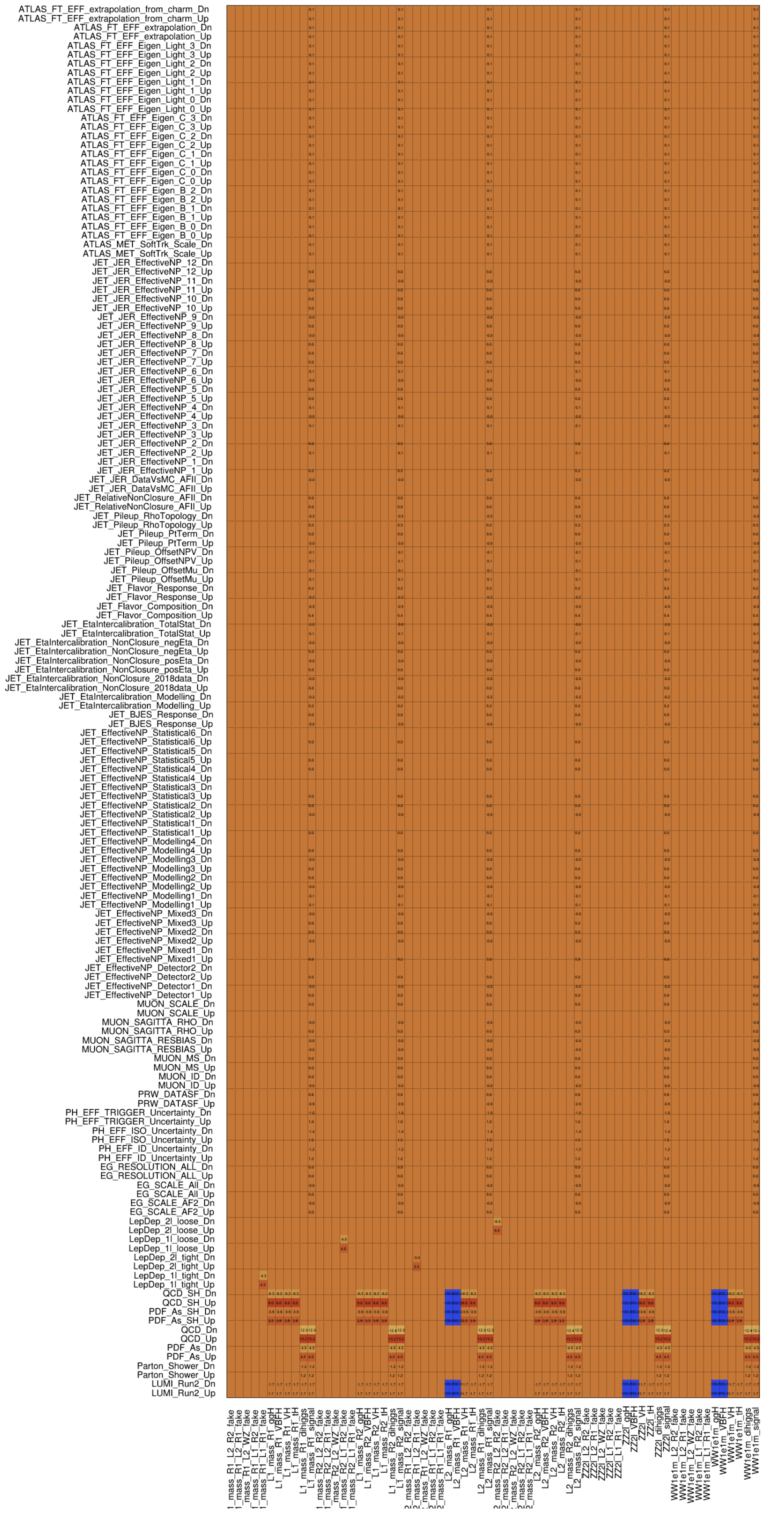


Figure 129: Normalization Factors for Nuisance parameters among all the regions for (m_X, m_S)=(1000,170) GeV mass.

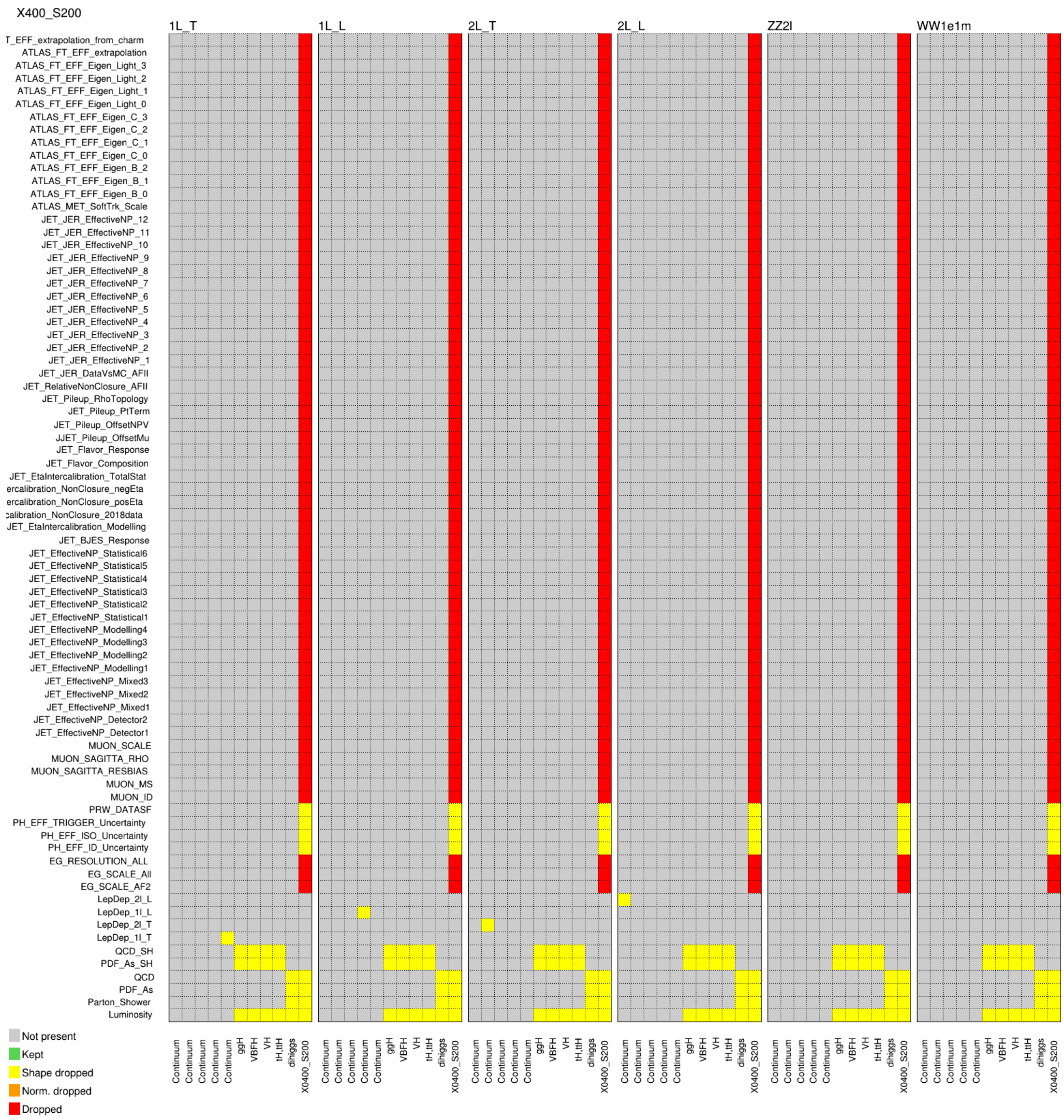


Figure 130: Pruning situation among all the regions for (m_X, m_S)=(1000,170) GeV mass.

1006 A.17 (m_X, m_S)=(1000, 200) GeV

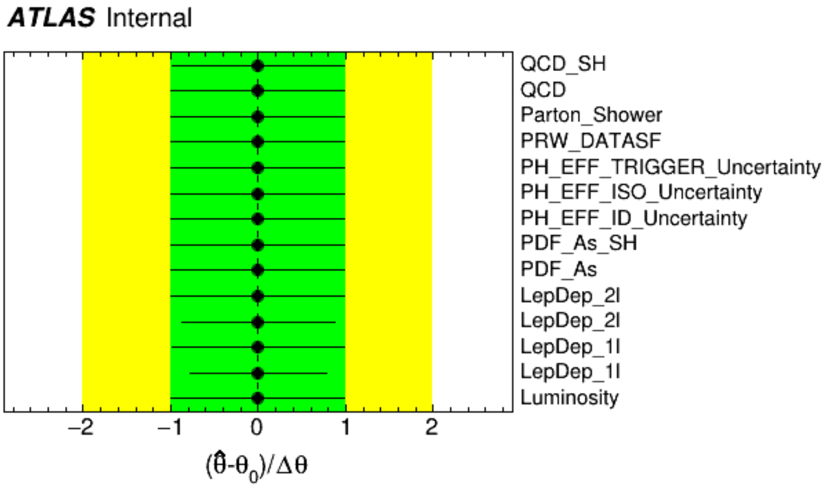


Figure 131: Nuisance parameters pull plot for (m_X, m_S)=(1000,200) GeV mass

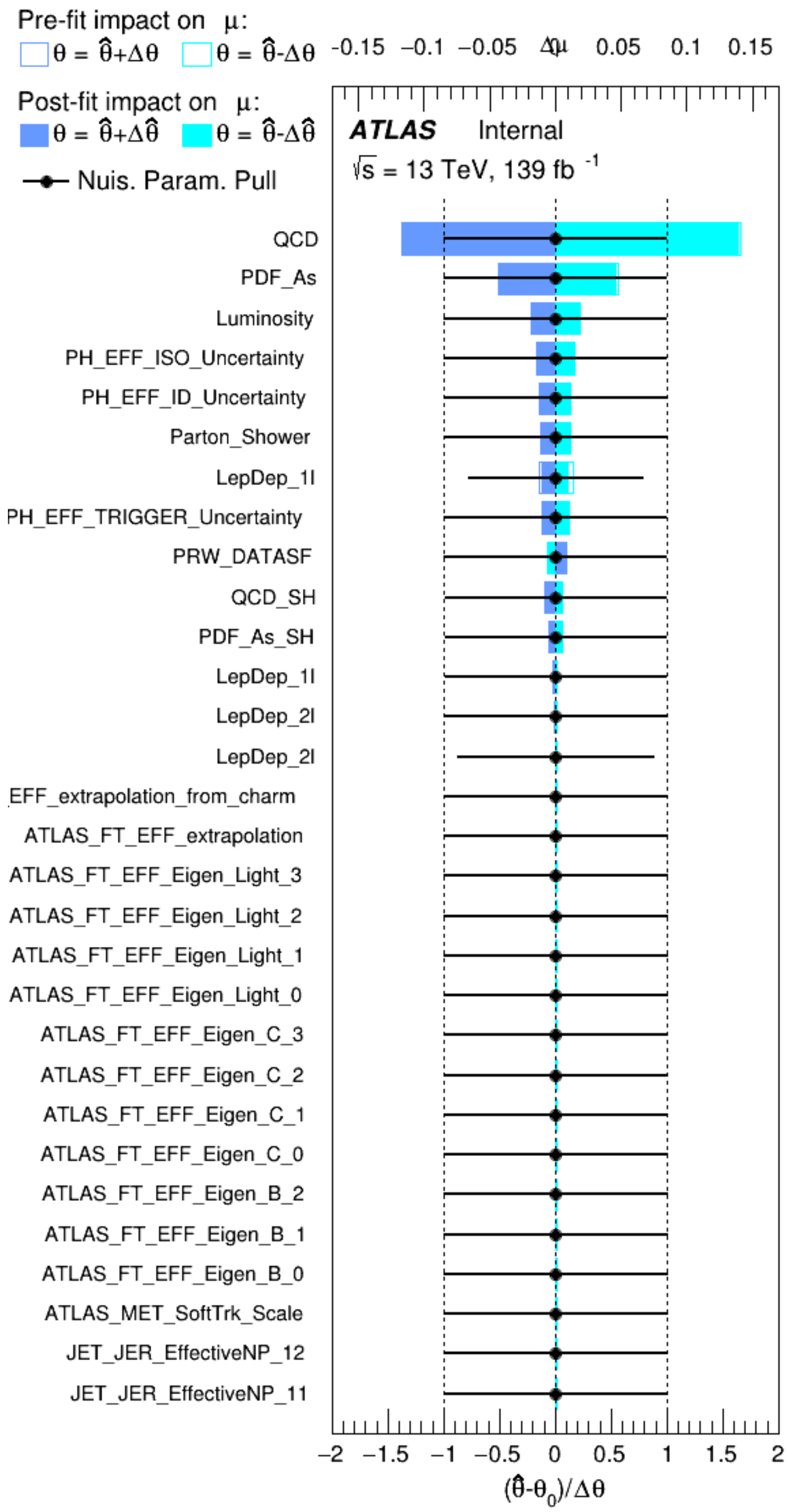


Figure 132: Nuisance parameters rankings for (m_X, m_S)=(1000,200) GeV mass

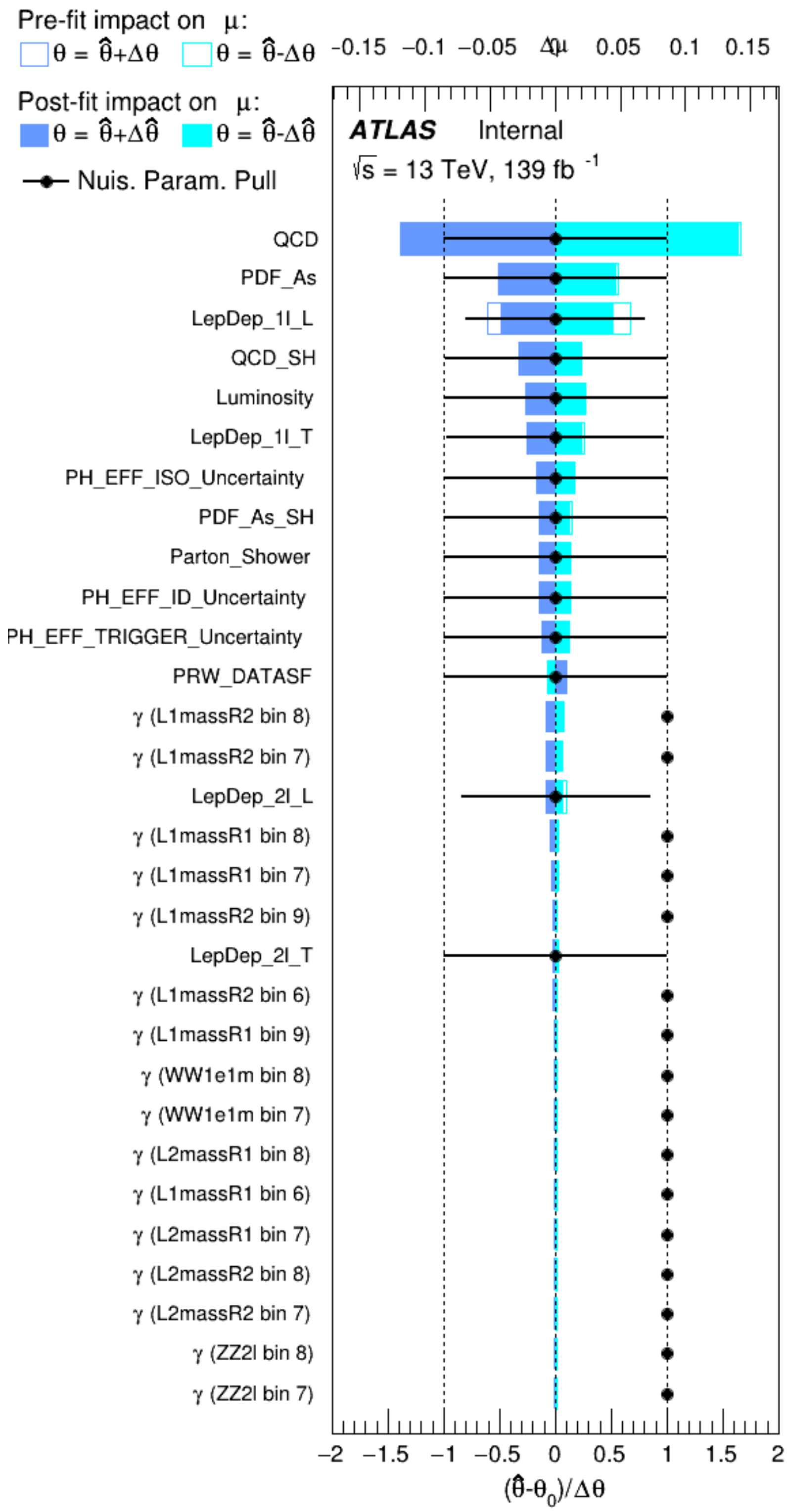


Figure 133: Rankings including systematics and gammas for (m_X, m_S)=(1000,200) GeV mass

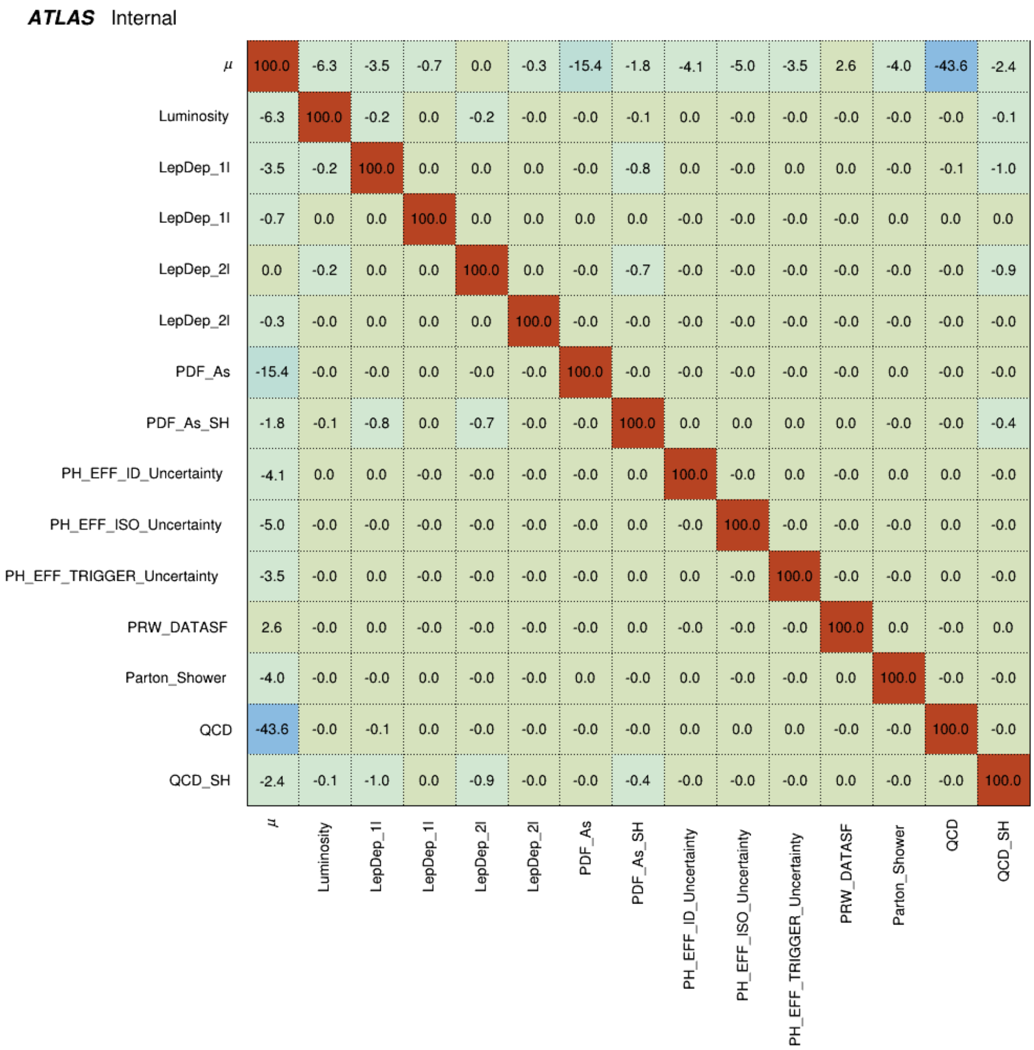


Figure 134: Correlation Matrix for (m_X, m_S)=(1000,200) GeV mass.

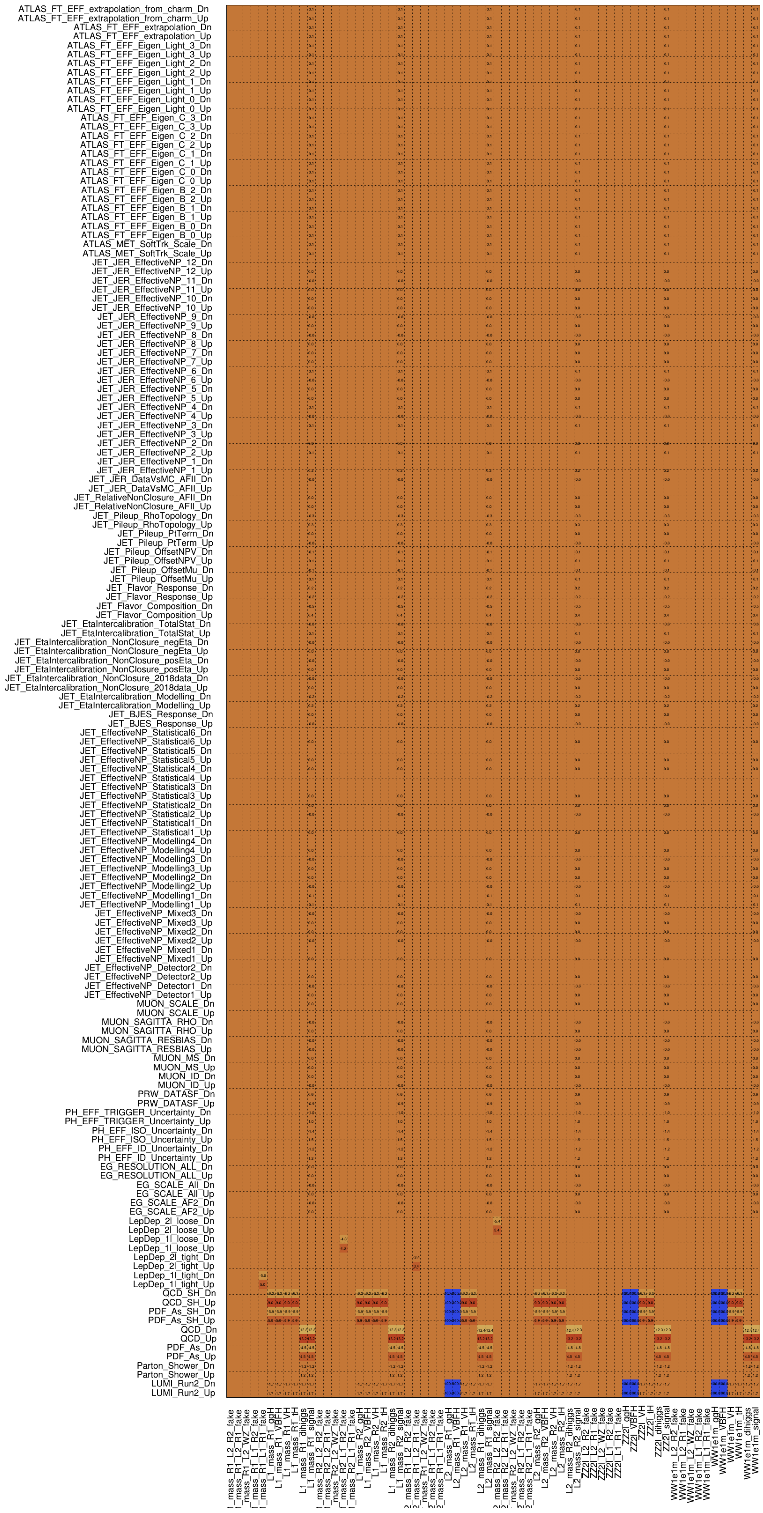


Figure 135: Normalization Factors for Nuisance parameters among all the regions for (m_X, m_S)=(1000,200) GeV mass.

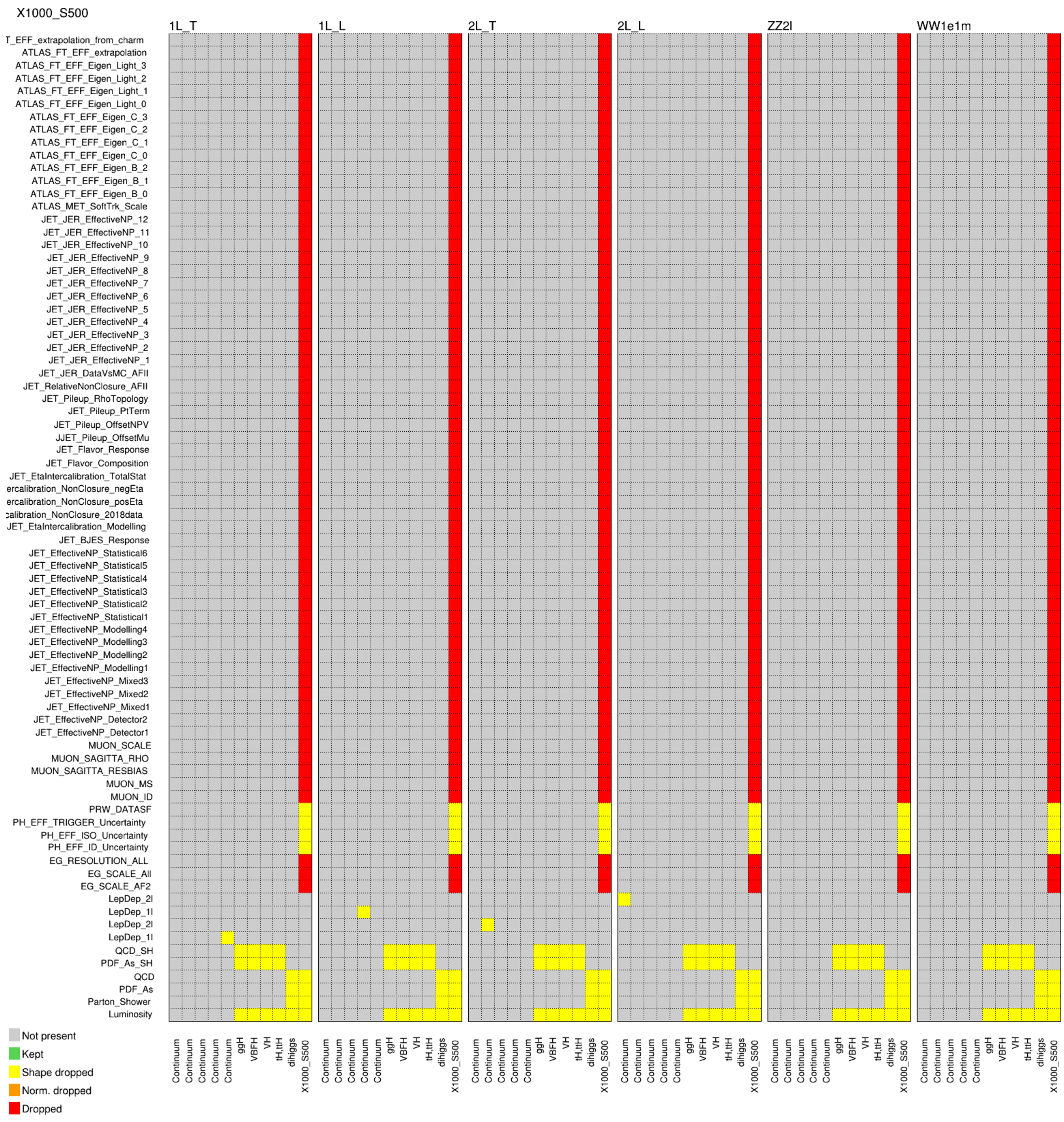


Figure 136: Pruning situation among all the regions for (m_X, m_S)=(1000,200) GeV mass.

1007 ### A.18 (m_X, m_S)=(1000, 300) GeV

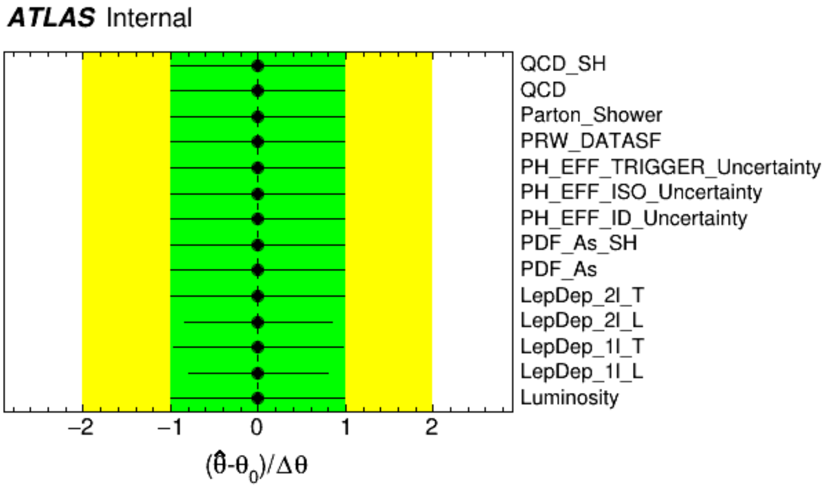


Figure 137: Nuisance parameters pull plot for (m_X, m_S)=(1000,300) GeV mass

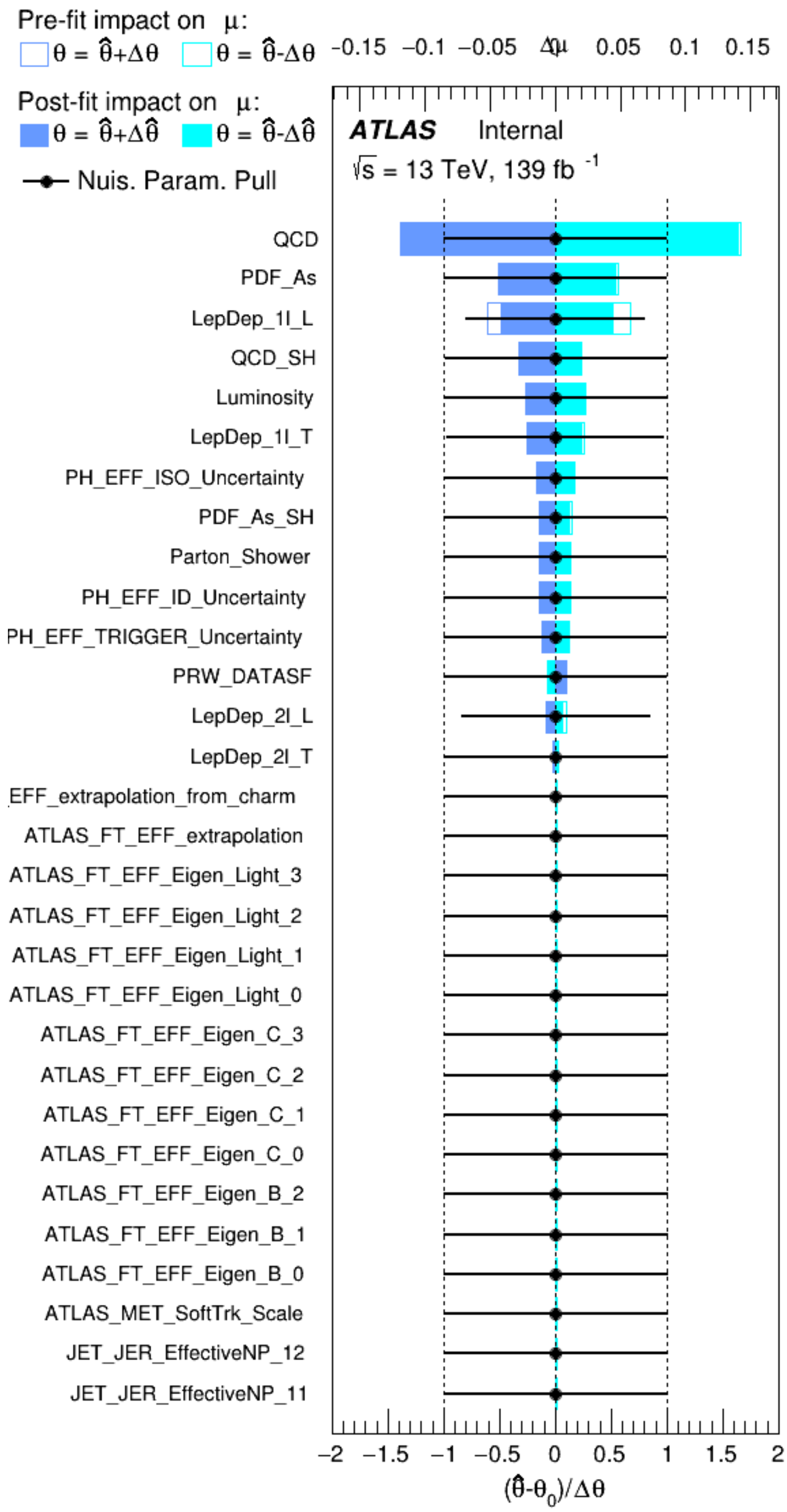


Figure 138: Nuisance parameters rankings for (m_X, m_S)=(1000,300) GeV mass

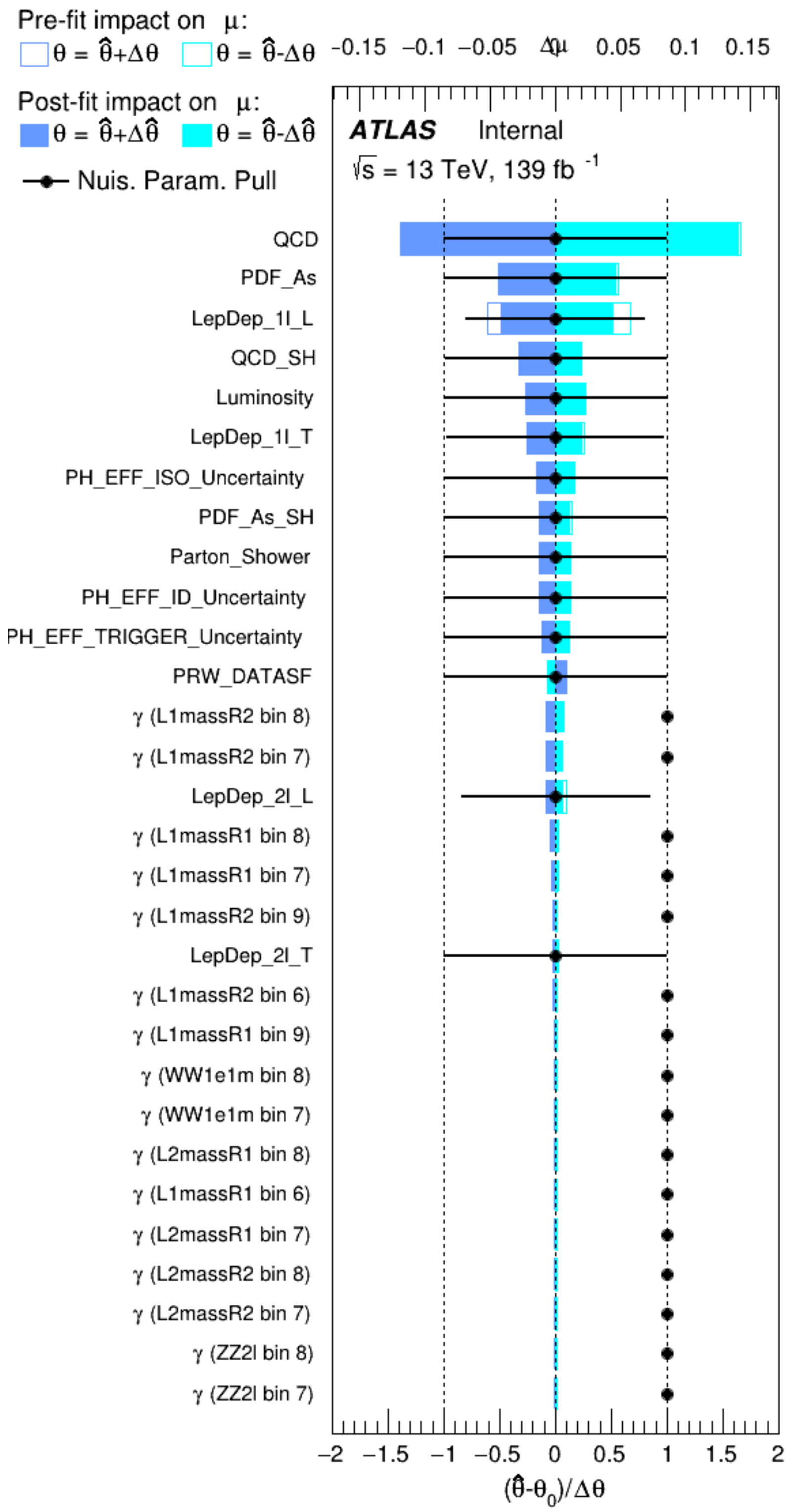


Figure 139: Rankings including systematics and gammas for (m_X, m_S)=(1000,300) GeV mass

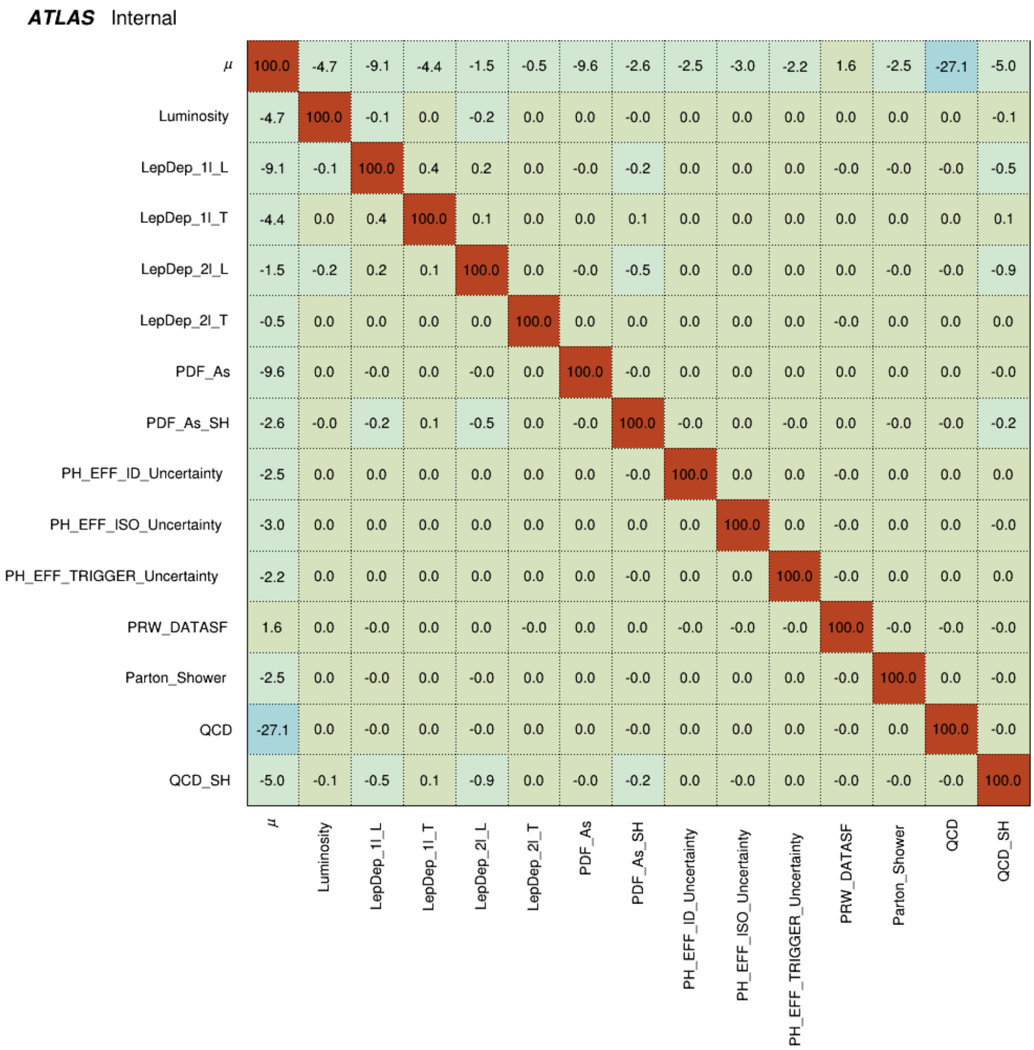


Figure 140: Correlation Matrix for (m_X, m_S)=(1000,300) GeV mass.

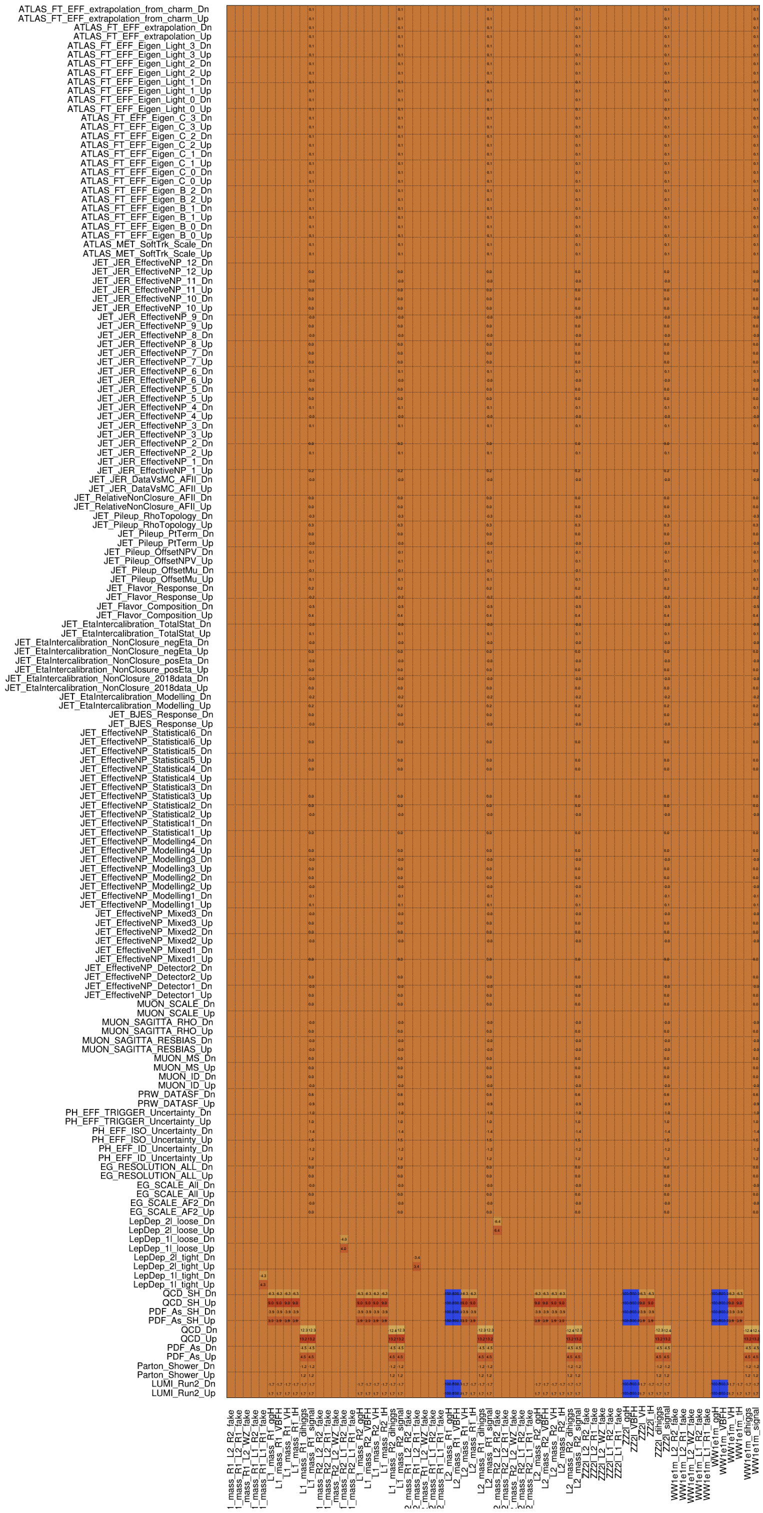


Figure 141: Normalization Factors for Nuisance parameters among all the regions for (m_X, m_S)=(1000,300) GeV mass.

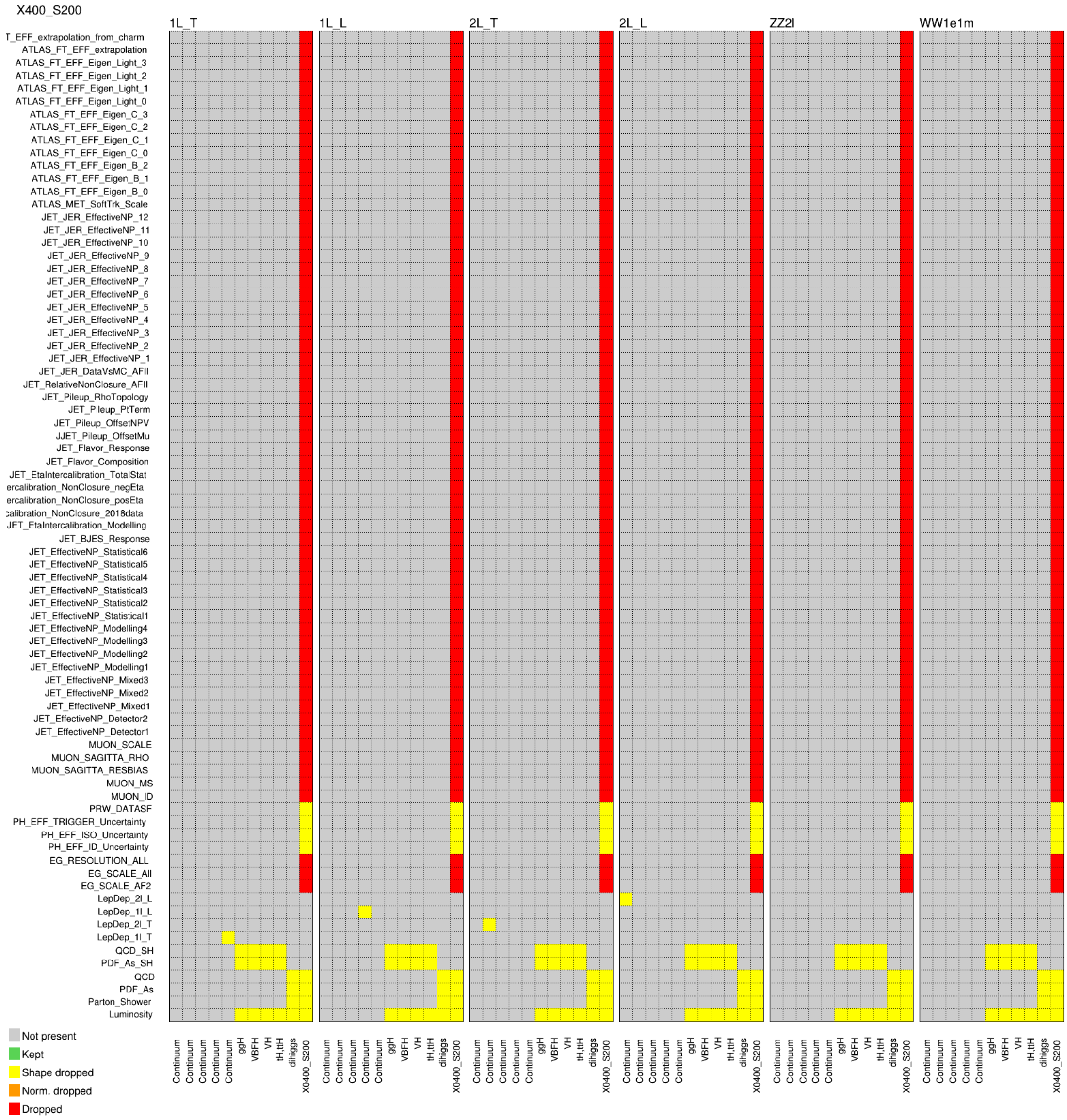


Figure 142: Pruning situation among all the regions for (m_X, m_S)=(1000,300) GeV mass.

1008 A.19 (m_X, m_S)=(1000, 400) GeV

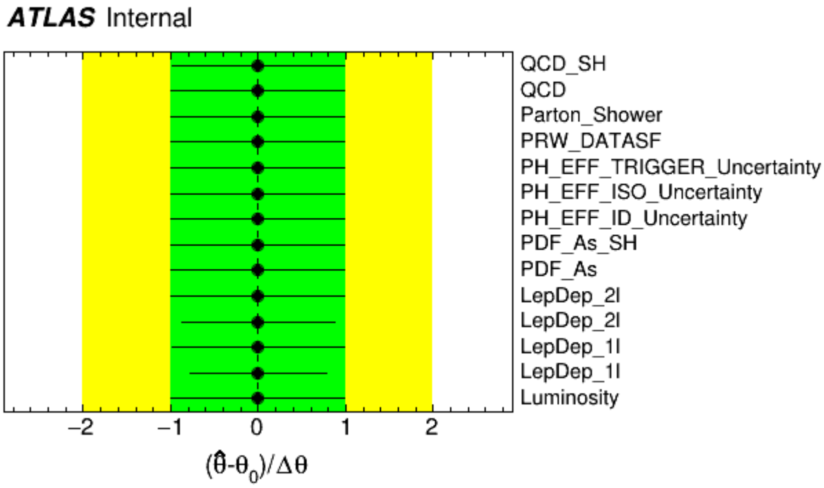


Figure 143: Nuisance parameters pull plot for (m_X, m_S)=(1000,400) GeV mass

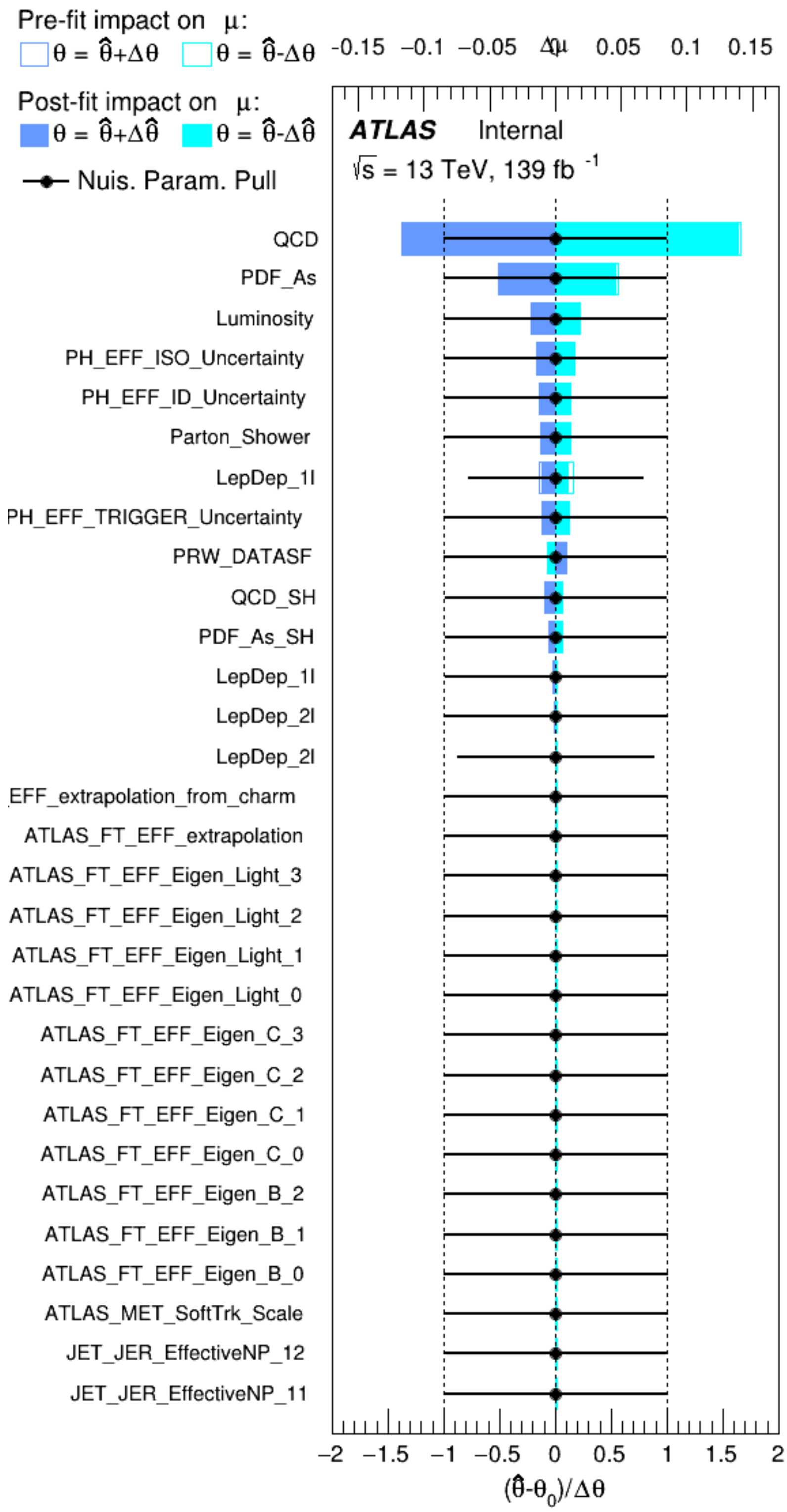


Figure 144: Nuisance parameters rankings for (m_X, m_S)=(1000,400) GeV mass

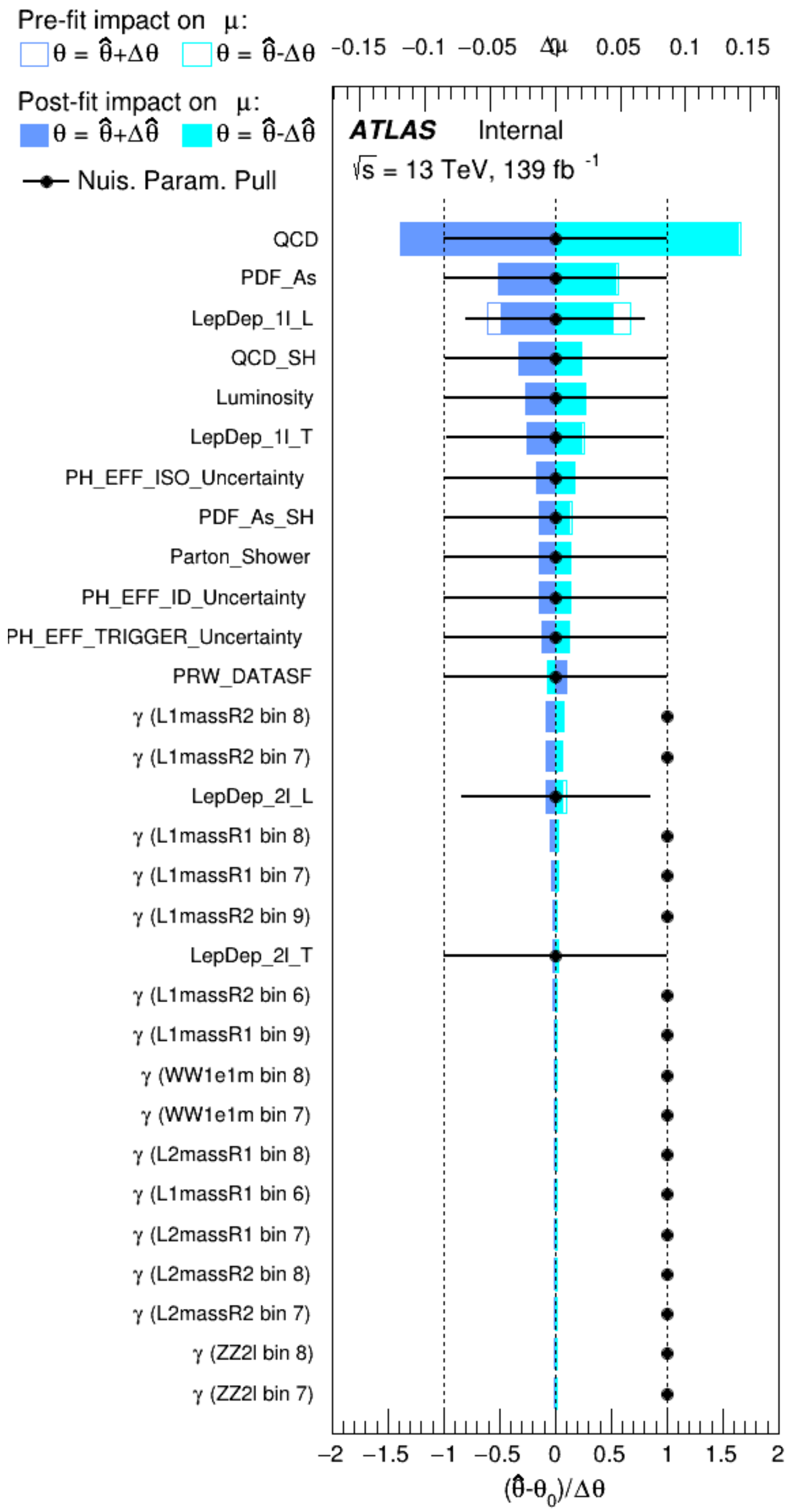


Figure 145: Rankings including systematics and gammas for (m_X, m_S)=(1000,400) GeV mass

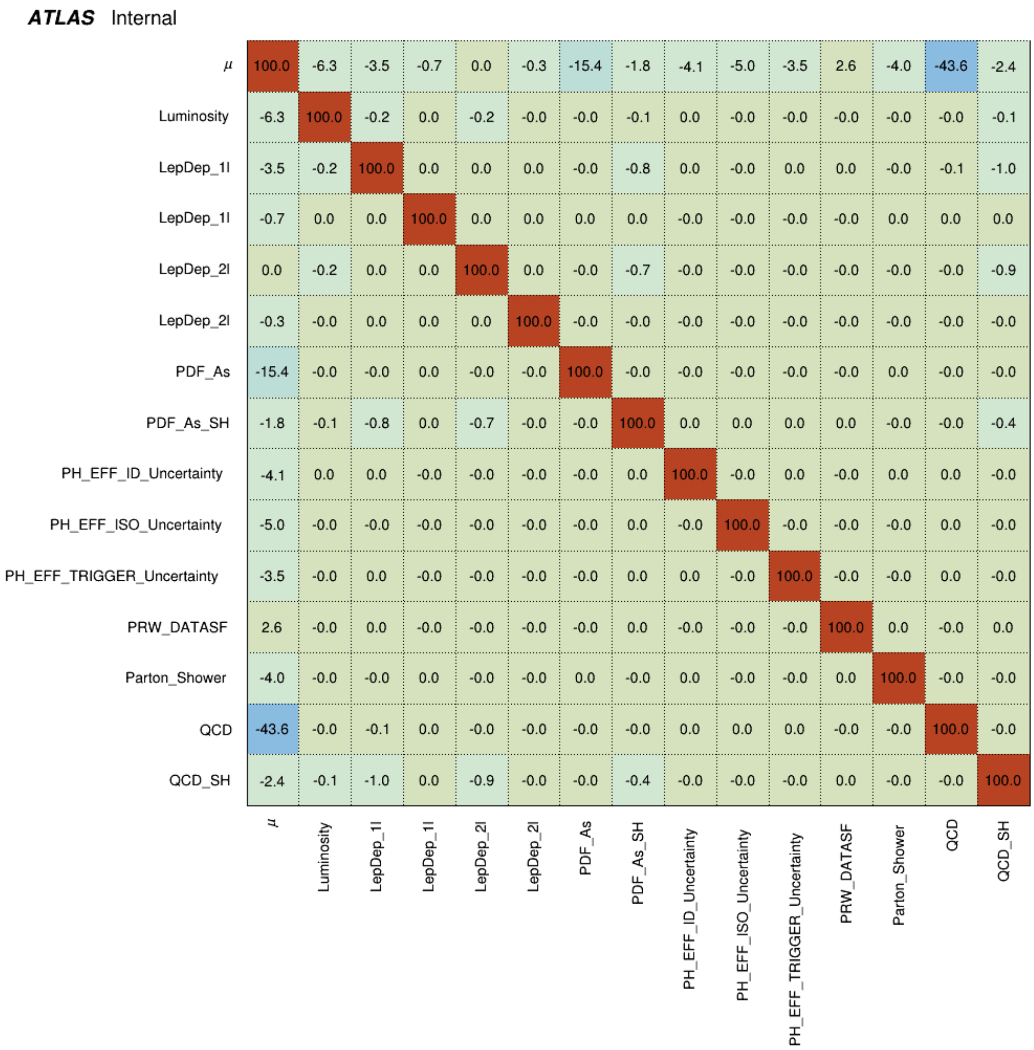


Figure 146: Correlation Matrix for (m_X, m_S)=(1000,400) GeV mass.

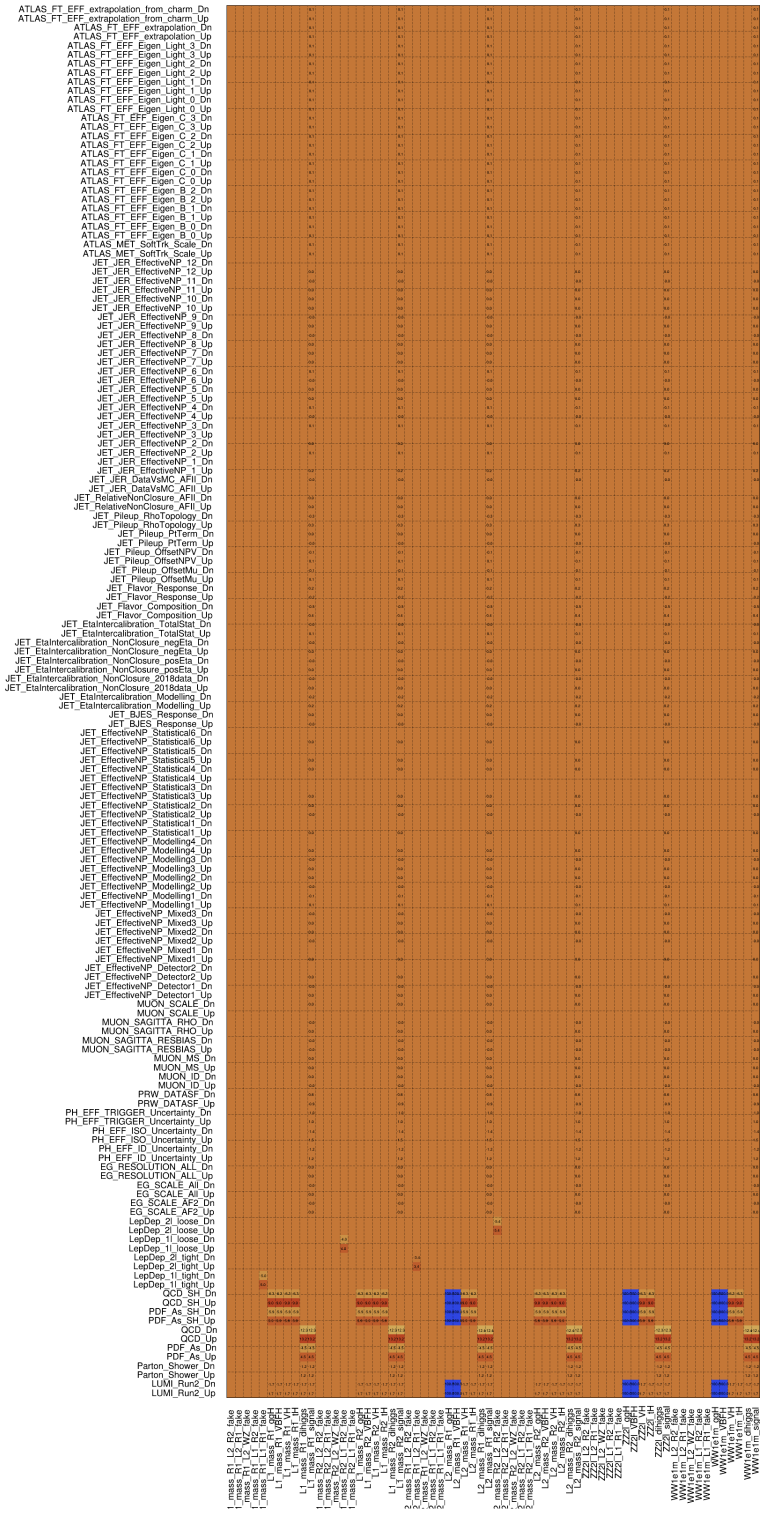


Figure 147: Normalization Factors for Nuisance parameters among all the regions for (m_X, m_S)=(1000,400) GeV mass.

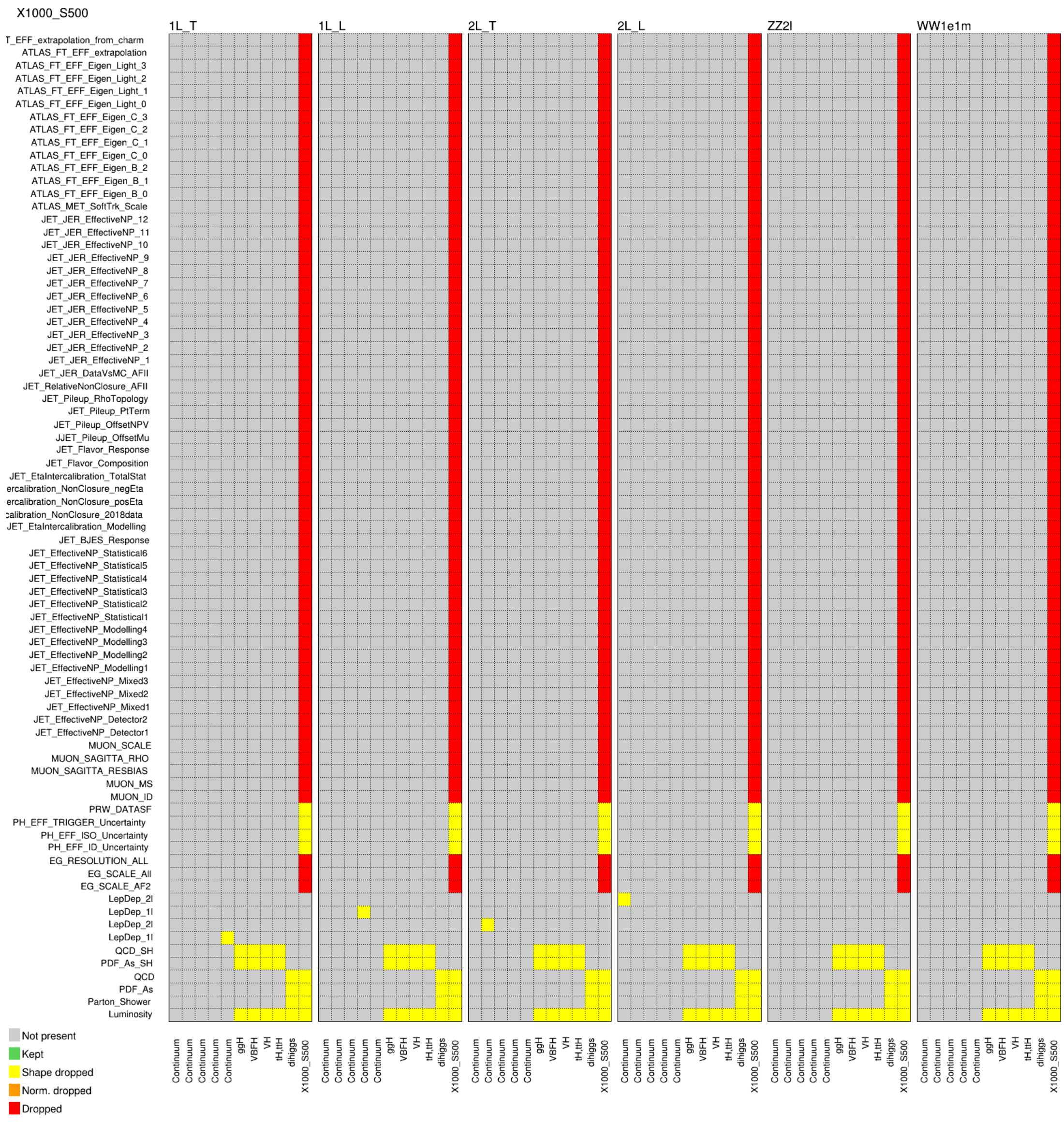


Figure 148: Pruning situation among all the regions for (m_X, m_S)=(1000,400) GeV mass.

1009 A.20 (m_X, m_S)=(1000, 500) GeV

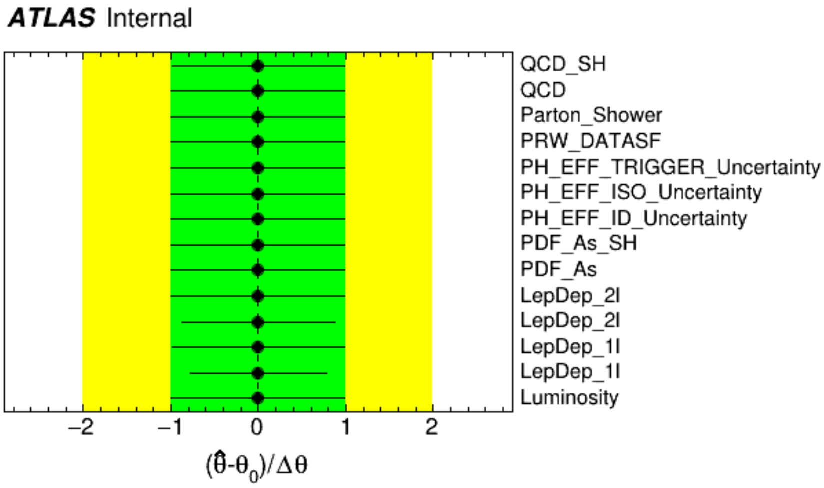


Figure 149: Nuisance parameters pull plot for (m_X, m_S)=(1000,500) GeV mass

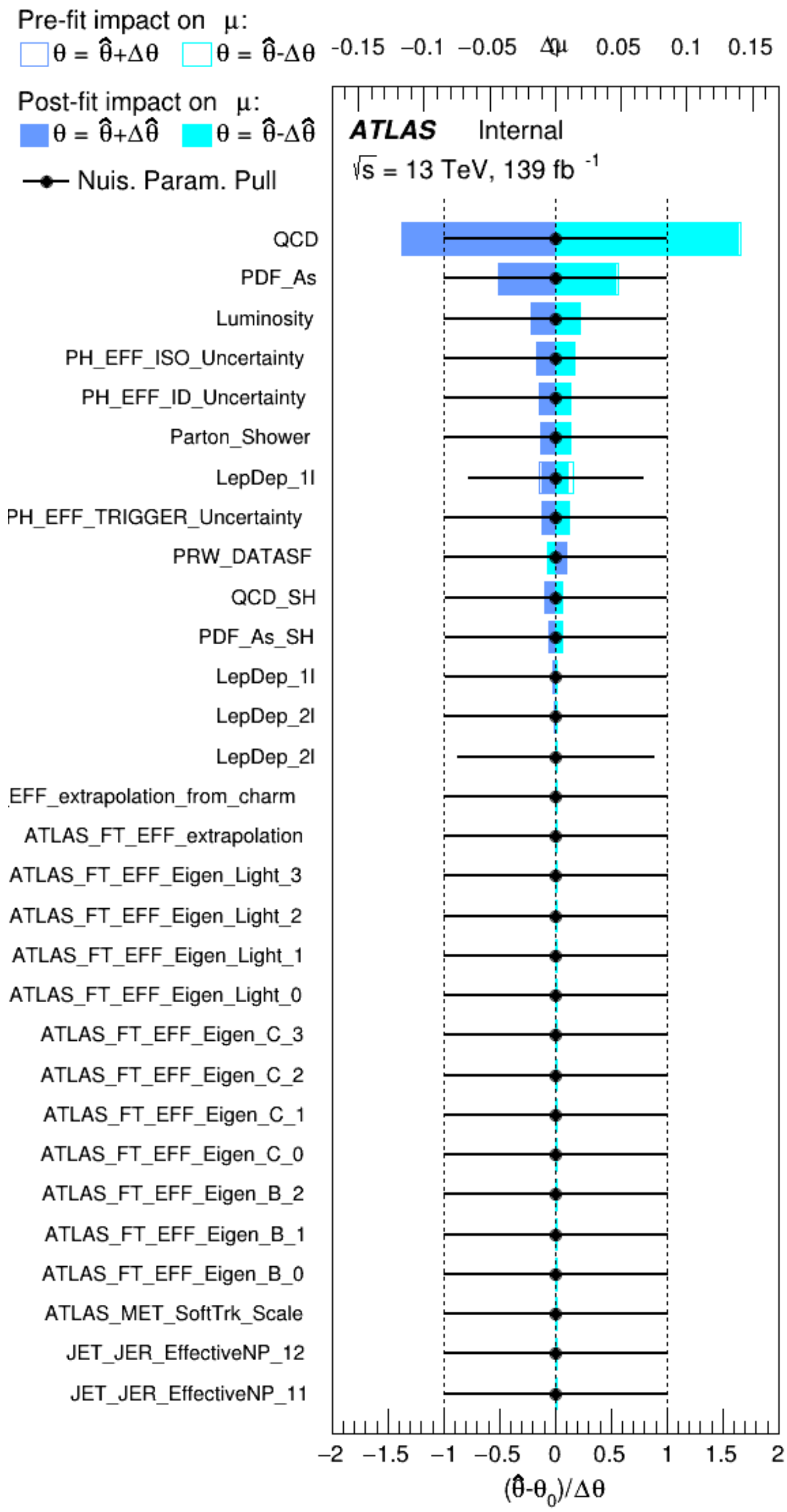


Figure 150: Nuisance parameters rankings for (m_X, m_S)=(1000,500) GeV mass

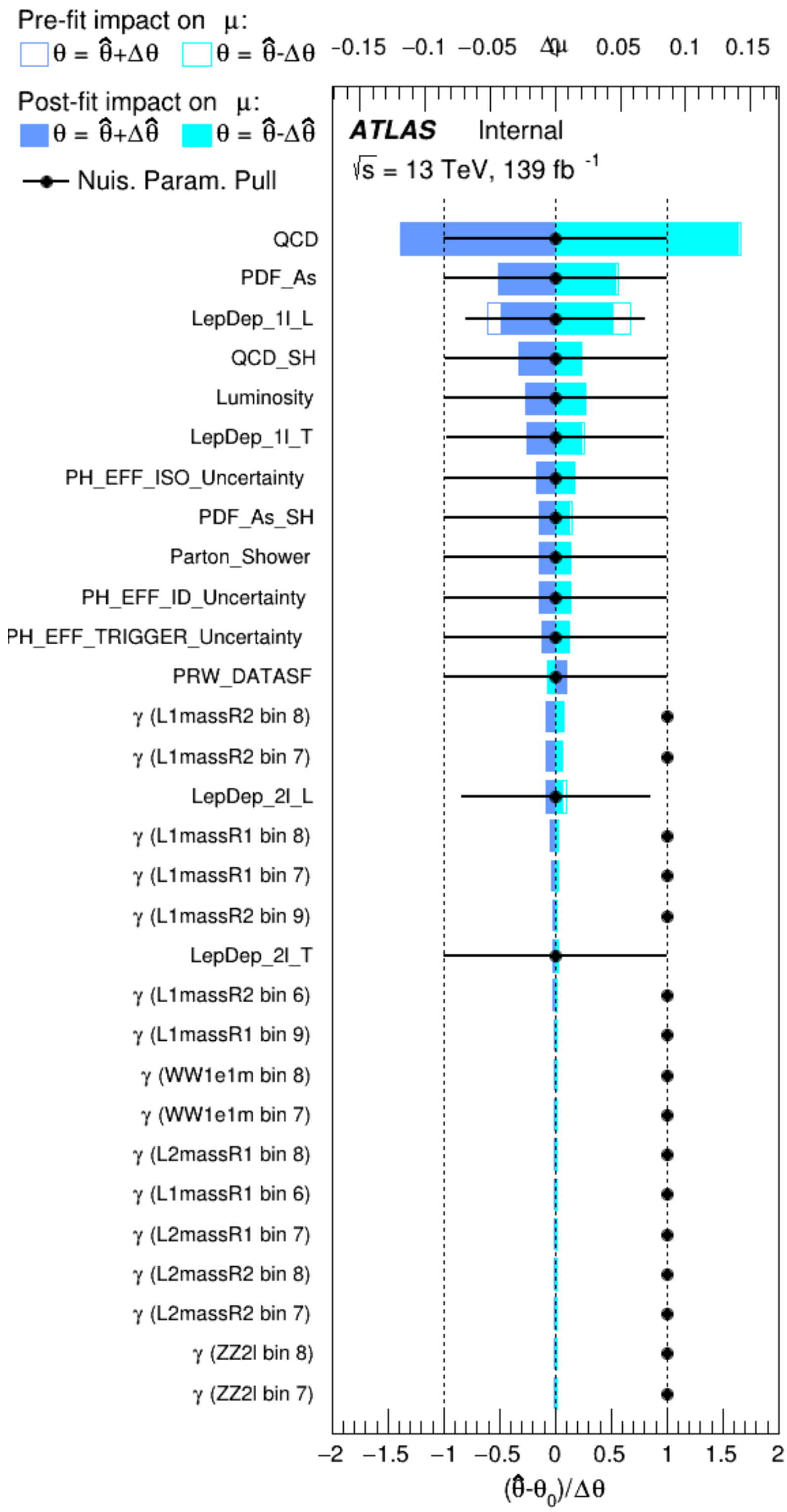


Figure 151: Rankings including systematics and gammas for (m_X, m_S)=(1000,500) GeV mass

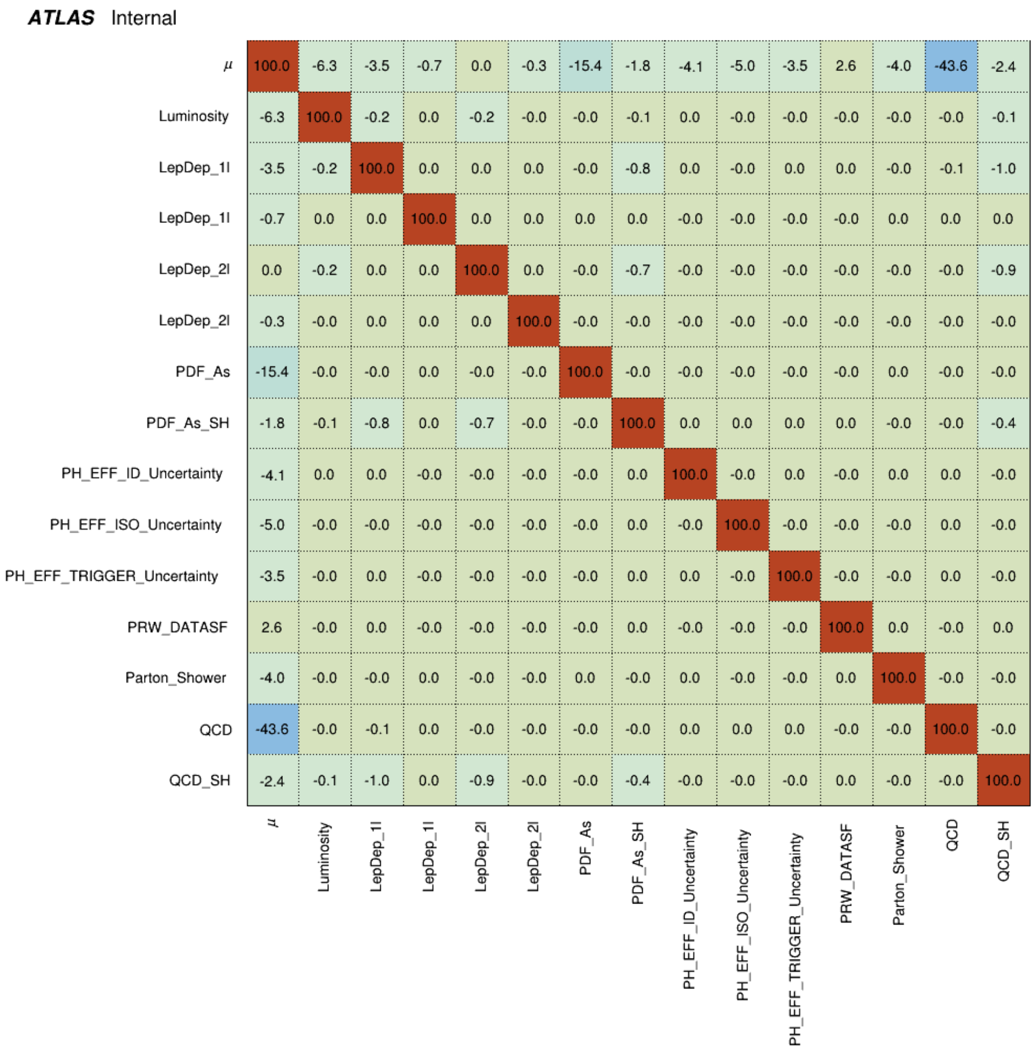


Figure 152: Correlation Matrix for (m_X, m_S)=(1000,500) GeV mass.

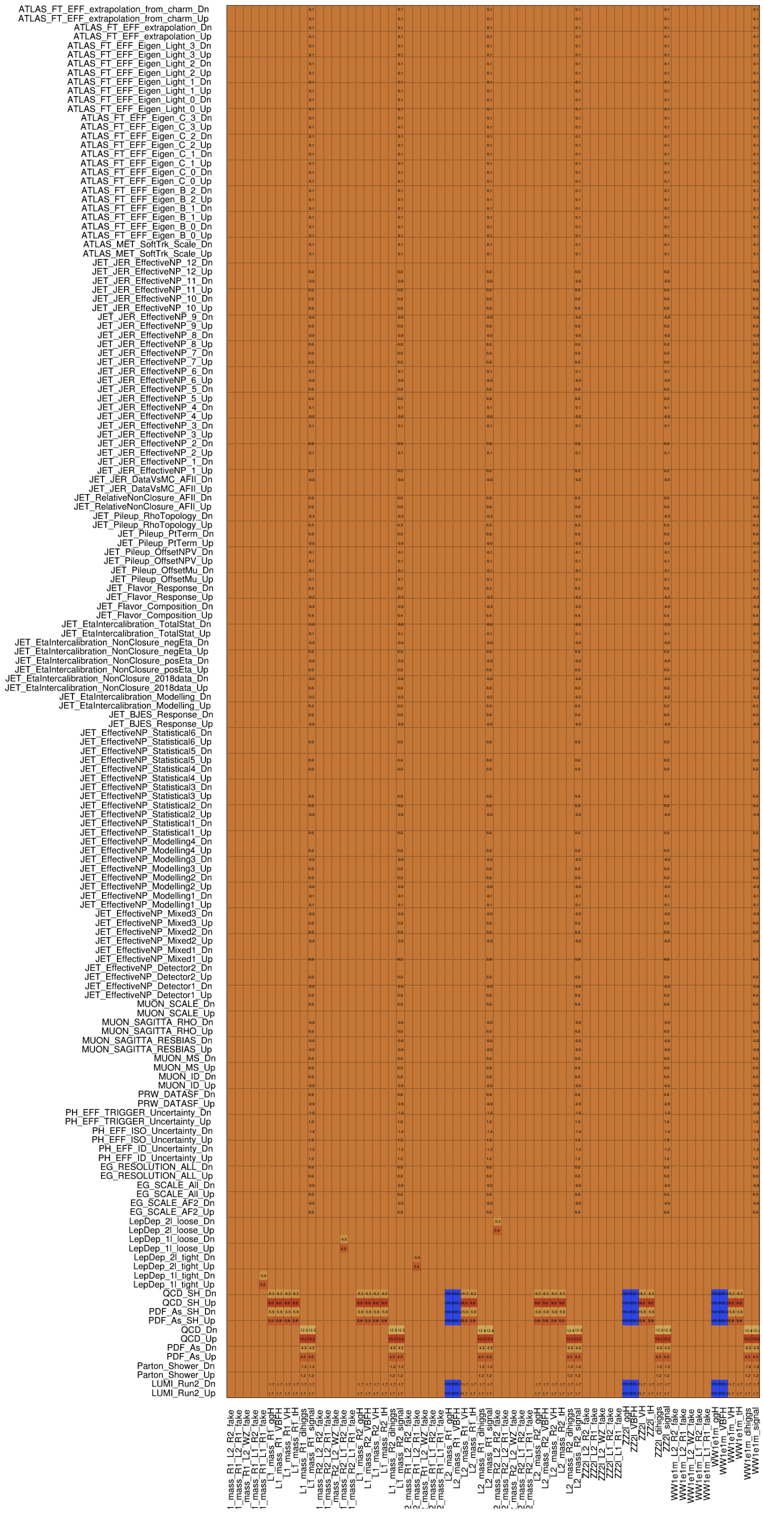


Figure 153: Normalization Factors for Nuisance parameters among all the regions for (m_X, m_S)=(1000,500) GeV mass.

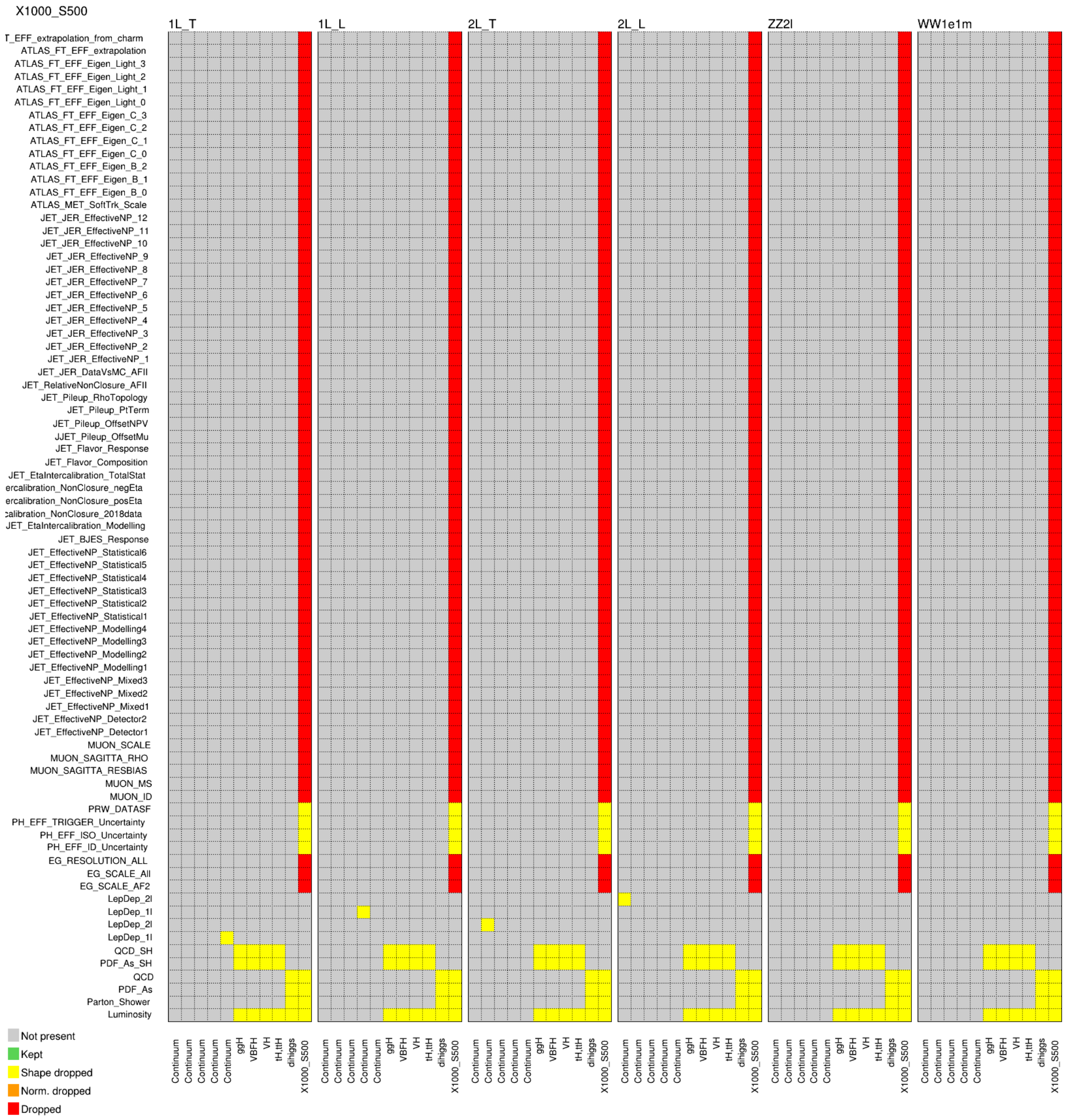


Figure 154: Pruning situation among all the regions for (m_X, m_S)=(1000,500) GeV mass.

B Continuum background modeling

All the spurious signal tests recorded in this Appendix.

C Cutflow of signal samples

m_X	300	400	400	500	500	500	600	600	600	600
m_S	170	170	200	170	200	300	170	200	300	400
WW1l, DSID	800943	800944	800945	800946	800947	800948	800949	800950	800951	800952
All events	100	100	100	100	100	100	100	100	100	100
No duplicates	100	100	100	100	100	100	100	100	100	100
GRL	100	100	100	100	100	100	100	100	100	100
Pass trigger	77.56	82.25	81.14	88.92	88.45	83.35	91.86	91.69	90.37	84.82
Detector DQ	77.56	82.25	81.14	88.92	88.45	83.35	91.86	91.69	90.37	84.82
Has PV	77.56	82.25	81.14	88.92	88.45	83.35	91.86	91.69	90.37	84.82
2 loose photons	58.57	59.26	59.46	60.98	61.10	61.23	63.06	63.11	62.69	62.65
Trigger match	53.29	54.64	54.12	58.45	58.19	55.83	61.60	61.46	60.02	57.10
tight ID	45.04	46.47	46.07	49.67	49.36	47.09	52.24	52.01	50.49	47.84
isolation	36.73	39.99	38.98	44.28	43.61	39.70	47.45	46.98	44.37	40.26
rel. pT cuts	34.16	35.51	34.57	39.95	39.10	35.11	43.62	43.15	39.97	35.50
m_{yy} in [105, 160]GeV	33.81	35.18	34.16	39.51	38.59	34.34	43.21	42.64	39.11	34.43
b-veto	30.75	31.49	30.52	34.97	34.19	30.35	38.03	37.52	34.18	30.19
At least 1lep	19.32	19.81	20.30	21.27	22.71	20.67	21.92	24.46	23.73	20.55
pass WW1l	11.01	13.12	13.85	15.20	16.58	16.11	16.27	18.60	18.92	16.95
WW2l, DSID	800963	800964	800965	800966	800967	800968	800969	800970	800971	800972
All events	100	100	100	100	100	100	100	100	100	100
No duplicates	100	100	100	100	100	100	100	100	100	100
GRL	100	100	100	100	100	100	100	100	100	100
Pass trigger	84.51	87.68	87.25	91.92	91.75	89.45	93.83	93.99	93.39	91.03
Detector DQ	84.51	87.68	87.25	91.92	91.75	89.45	93.83	93.99	93.39	91.03
Has PV	84.51	87.68	87.25	91.92	91.75	89.45	93.83	93.99	93.39	91.03
2 loose photons	58.15	57.75	58.43	59.10	58.94	60.23	60.91	60.68	60.45	61.59
Trigger match	53.03	53.45	53.28	56.75	56.22	55.25	59.63	59.21	58.15	56.72
tight ID	45.22	45.74	45.55	48.59	48.07	47.34	50.94	50.69	49.77	48.45
isolation	38.12	39.65	39.06	43.23	42.55	40.57	45.96	45.60	43.99	41.37
rel. pT cuts	35.49	34.90	34.68	38.88	38.01	35.87	42.10	41.73	39.45	36.56
m_{yy} in [105, 160]GeV	34.79	34.22	33.88	38.02	36.96	34.43	41.17	40.68	37.95	34.67
b-veto	33.56	32.74	32.39	36.23	35.17	32.95	39.05	38.65	36.08	32.96
At least 2lep	17.14	18.06	18.12	20.55	20.72	19.45	22.16	23.23	22.35	20.15
pass WW2l	17.01	17.90	17.60	20.34	20.09	18.95	21.90	22.44	21.73	19.76
pass ZZ2l	0.07	0.10	0.43	0.13	0.52	0.40	0.14	0.65	0.46	0.23
WW2l-em	8.46	8.91	8.85	10.17	10.27	9.64	10.96	11.50	11.02	10.03
fall to 1lepton category	11.93	10.51	10.99	11.07	11.14	10.93	11.54	11.82	11.33	10.58
ZZ2l, DSID	800983	800984	800985	800986	800987	800988	800989	800990	800991	800992
All events	100	100	100	100	100	100	100	100	100	100
No duplicates	100	100	100	100	100	100	100	100	100	100
GRL	100	100	100	100	100	100	100	100	100	100
Pass trigger	77.68	81.12	80.26	87.03	86.52	81.65	89.98	89.82	88.24	82.69
Detector DQ	77.68	81.12	80.26	87.03	86.52	81.65	89.98	89.82	88.24	82.69
Has PV	77.68	81.12	80.26	87.03	86.52	81.65	89.98	89.82	88.24	82.69
2 loose photons	53.42	53.66	54.13	54.92	55.21	55.79	56.92	57.19	56.96	57.33
Trigger match	48.43	49.46	49.14	52.63	52.49	50.88	55.57	55.65	54.47	52.29
tight ID	40.75	41.91	41.52	44.61	44.41	42.81	46.87	46.93	45.85	43.71
isolation	32.83	35.78	34.68	39.36	38.81	35.77	42.31	42.05	39.74	36.32
rel. pT cuts	30.54	31.61	30.81	35.43	34.77	31.78	38.92	38.53	35.90	32.00
m_{yy} in [105, 160]GeV	29.93	30.98	30.14	34.67	33.94	30.73	38.17	37.69	34.73	30.70
b-veto	25.04	24.53	23.71	26.65	25.77	22.89	28.65	28.14	25.25	21.89
At least 2lep	12.82	12.97	12.87	13.79	13.86	13.75	14.16	14.69	15.67	13.50
pass WW2l	10.13	9.66	6.13	9.94	6.21	5.69	9.99	6.26	6.18	5.16
pass ZZ2l	2.64	3.24	6.68	3.77	7.60	7.99	4.08	8.35	9.40	8.25
WW2l-em	0.07	0.09	0.08	0.09	0.09	0.13	0.11	0.11	0.14	0.11
fall to 1lepton category	8.52	8.06	7.69	8.76	8.32	6.26	9.56	9.20	6.33	5.53

Table 30: Efficiencies in percent for event selection for signals.

X	750	750	750	750	750	1000	1000	1000	1000	1000
S	170	200	300	400	500	170	200	300	400	500
WW1l, DSID	800953	800954	800955	800956	800957	800938	800939	800940	800941	800942
All events	100	100	100	100	100	100	100	100	100	100
No duplicates	100	100	100	100	100	100	100	100	100	100
GRL	100	100	100	100	100	100	100	100	100	100
Pass trigger	93.90	93.92	93.60	92.95	90.06	95.70	95.56	95.69	95.69	95.31
Detector DQ	93.90	93.92	93.60	92.95	90.06	95.70	95.56	95.69	95.69	95.31
Has PV	93.90	93.92	93.60	92.95	90.06	95.70	95.56	95.69	95.69	95.31
2 loose photons	66.12	65.98	65.50	64.81	63.87	70.26	70.03	69.93	69.35	68.60
Trigger match	65.35	65.15	64.43	63.04	60.41	69.85	69.62	69.50	68.81	67.87
tight ID	55.30	55.05	54.46	52.80	50.27	59.30	58.99	58.67	58.08	56.94
isolation	51.01	50.75	49.51	46.92	43.19	55.65	55.44	54.89	53.74	51.84
rel. pT cuts	47.92	47.61	46.04	43.00	38.47	53.12	52.95	52.20	50.81	48.79
m_{yy} in [105, 160]GeV	47.57	47.12	45.22	41.89	37.09	52.87	52.51	51.55	49.92	47.62
b-veto	41.45	41.08	39.16	36.33	32.17	45.53	45.14	44.30	42.94	40.86
1lep	21.69	25.62	27.08	24.90	21.87	19.71	25.60	30.14	29.57	27.95
pass WW1l	16.62	20.31	22.32	20.98	18.68	15.49	20.95	25.37	25.19	24.01
WW2l, DSID	800973	800974	800975	800976	800977	800958	800959	800960	800961	800962
All events	100	100	100	100	100	100	100	100	100	100
No duplicates	100	100	100	100	100	100	100	100	100	100
GRL	100	100	100	100	100	100	100	100	100	100
Pass trigger	95.43	95.31	95.49	95.37	94.10	96.61	96.77	96.84	96.88	96.82
Detector DQ	95.43	95.31	95.49	95.37	94.10	96.61	96.77	96.84	96.88	96.82
Has PV	95.43	95.31	95.49	95.37	94.10	96.61	96.77	96.84	96.88	96.82
2 loose photons	63.62	63.33	63.21	62.85	62.67	67.11	67.07	67.16	66.86	66.30
Trigger match	62.94	62.62	62.21	61.51	59.53	66.77	66.70	66.76	66.37	65.70
tight ID	53.72	53.75	53.21	52.64	50.85	57.43	57.28	57.33	56.91	56.27
isolation	49.43	49.29	48.39	47.13	44.22	53.56	53.49	53.36	52.56	51.50
rel. pT cuts	46.18	46.00	44.78	42.90	39.40	50.90	50.75	50.38	49.46	48.15
m_{yy} in [105, 160]GeV	45.32	44.96	43.35	40.95	37.02	50.16	49.81	49.12	47.84	46.02
b-veto	42.79	42.47	41.01	38.49	35.14	46.94	46.63	46.05	44.92	43.22
At least 2lep	23.96	25.67	26.20	24.56	22.24	24.54	28.04	29.91	29.17	28.33
pass WW2l	23.65	24.75	25.38	24.11	21.87	24.14	26.97	28.89	28.49	27.84
pass ZZ2l	0.17	0.74	0.63	0.29	0.16	0.22	0.86	0.80	0.42	0.22
WW2l-em	11.90	12.65	12.93	12.22	11.08	12.22	13.90	14.75	14.40	14.12
fall to 1lepton category	12.27	12.52	12.14	11.62	10.76	13.54	13.04	12.99	12.88	12.34
ZZ2l, DSID	800993	800994	800995	800996	800997	800978	800979	800980	800981	800982
All events	100	100	100	100	100	100	100	100	100	100
No duplicates	100	100	100	100	100	100	100	100	100	100
GRL	100	100	100	100	100	100	100	100	100	100
Pass trigger	92.35	92.38	91.98	90.73	87.54	94.41	94.56	94.45	94.11	93.63
Detector DQ	92.35	92.38	91.98	90.73	87.54	94.41	94.56	94.45	94.11	93.63
Has PV	92.35	92.38	91.98	90.73	87.54	94.41	94.56	94.45	94.11	93.63
2 loose photons	59.89	60.04	59.63	59.11	58.35	63.46	63.58	63.61	63.33	62.62
Trigger match	59.27	59.25	58.55	57.46	55.16	63.09	63.17	63.14	62.80	61.87
tight ID	49.98	50.14	49.17	48.08	45.78	53.39	53.43	53.22	52.85	51.68
isolation	45.90	46.00	44.42	42.29	38.80	49.83	49.88	49.41	48.66	46.80
rel. pT cuts	42.91	42.97	41.25	38.70	34.58	47.48	47.40	46.85	46.02	43.98
m_{yy} in [105, 160]GeV	42.22	42.23	40.22	37.43	33.09	46.95	46.78	45.99	44.92	42.68
b-veto	30.66	30.57	28.53	26.06	22.68	33.10	32.64	31.58	30.64	28.72
At least 2lep	13.68	14.63	17.85	16.51	14.55	12.22	13.08	19.37	19.46	18.42
pass WW2l	9.64	6.18	6.58	5.93	5.20	8.72	5.85	6.76	6.58	6.17
pass ZZ2l	3.95	8.36	11.15	10.48	9.25	3.39	7.12	12.48	12.72	12.14
WW2l-em	0.11	0.13	0.17	0.18	0.16	0.12	0.10	0.20	0.21	0.21
fall to 1lepton category	10.88	10.59	6.77	5.98	5.06	12.07	12.17	7.12	6.21	5.84

Table 31: Efficiencies in percent for event selection for signals.(Continued)

D Pythia8 for signals

Here, the script used for generating SH model WW 1lepton channel at the mass point of X 400 GeV, S 200 GeV is given. The scriptes for other mass points, and other decays, are basically the same.

```
include("Pythia8_i/Pythia8_A14_NNPDF23LO_EvtGen_Common.py")
evgenConfig.generators = ["Pythia8", "EvtGen"]
evgenConfig.process     = "gg->X->SH->WW+yy, 1 lepton"
evgenConfig.description = "Generation of gg > X > SH where S decays to W+W- with 1 lepton
evgenConfig.keywords = ["BSMHiggs"]

genSeq.Pythia8.Commands += ['Higgs:useBSM = on',
                            'ParticleDecays:mSafety = 0.0',
                            'HiggsBSM:gg2A3 = on',
                            'HiggsA3:parity = 1',
                            'Higgs:clipWings = off',
                            '36:m0 = 400.0',
                            '36:mWidth = 0.01',
                            '36:doForceWidth = yes',
                            '36:addChannel = 1 1 100 25 35',
                            '36:onMode = off',
                            '36:onIfMatch = 25 35',
                            '36:mayDecay = on',
                            '35:mMin = 50.0',
                            '25:mMin = 50.0',
                            '35:m0 = 200.0',
                            '35:mWidth = 0.01',
                            '35:doForceWidth = yes',
                            '25:onMode = off',
                            '25:onIfMatch = 22 22',
                            '35:onMode = off',
                            '35:onIfMatch = 24 -24',
                            ]


from GeneratorFilters.GeneratorFiltersConf import MultiLeptonFilter
filtSeq += MultiLeptonFilter("LepOneFilter")
filtSeq.LepOneFilter.NLeptons = 1
filtSeq.LepOneFilter.Ptcut = 7000
filtSeq.LepOneFilter.Etacut = 3

filtSeq += MultiLeptonFilter("LepTwoFilter")
filtSeq.LepTwoFilter.NLeptons = 2
filtSeq.LepTwoFilter.Ptcut = 7000
filtSeq.LepTwoFilter.Etacut = 3
filtSeq.Expression = "LepOneFilter and not LepTwoFilter"
```

1057 # E MC Reweighting

1058 The discrenpancy between continuum MC and data sideband data includes both yields and shape
1059 difference.

1060 For the yields difference, to simplify the situation, the ratio between 3 MC processes are fixed and these
1061 are scaled by 1.69. In the same way, 2 lepton channels are scaled with 1.04. With the scaling, the yields
1062 consistency between MC and sideband data is confirmed.

1063 To further mimic the deviation, the bin by bin reweighting on m_T^{W1} distribution is done, as described in
1064 Figure 6. Other kinematic distributions are shown in the followiing Figure 155 to 166.

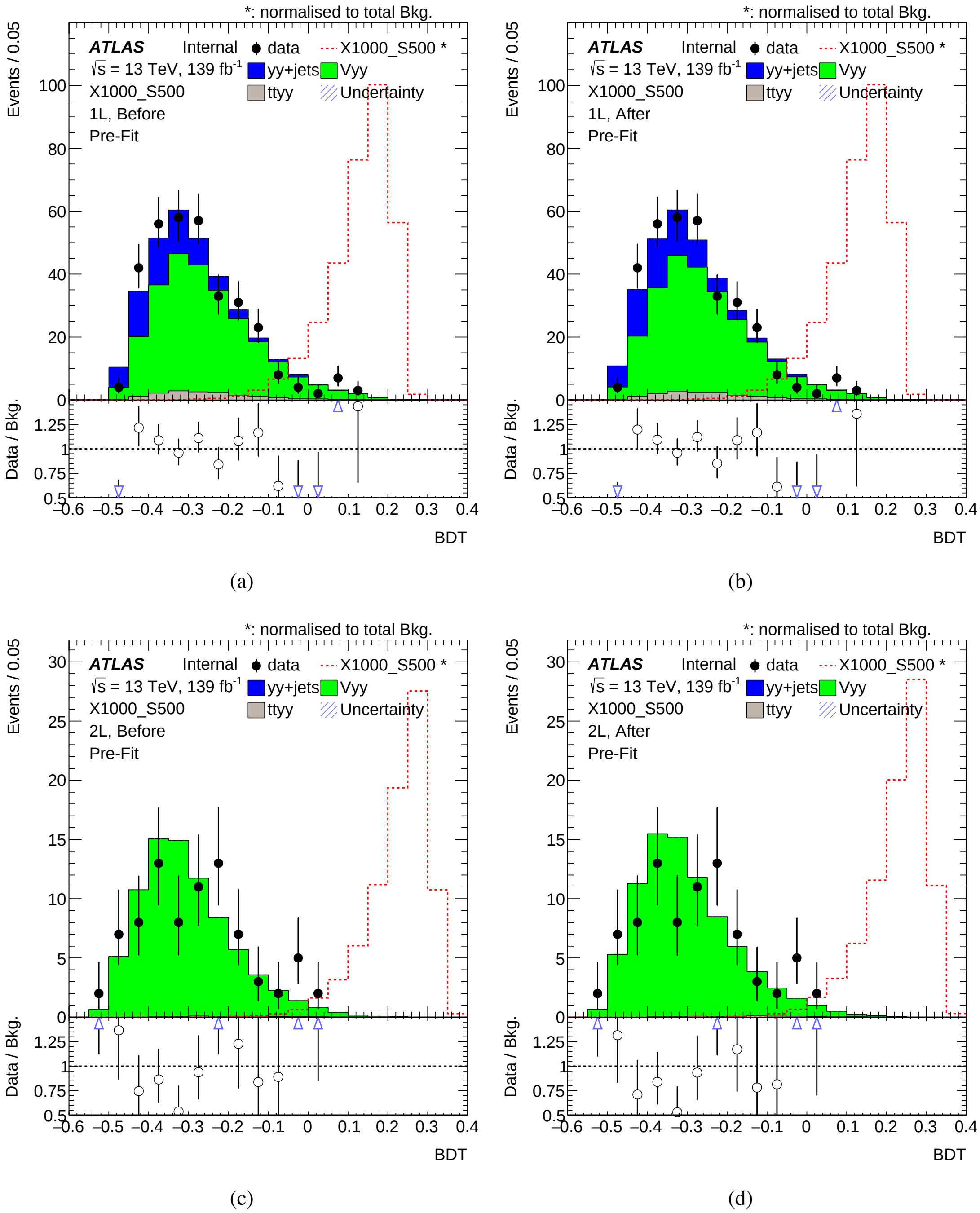
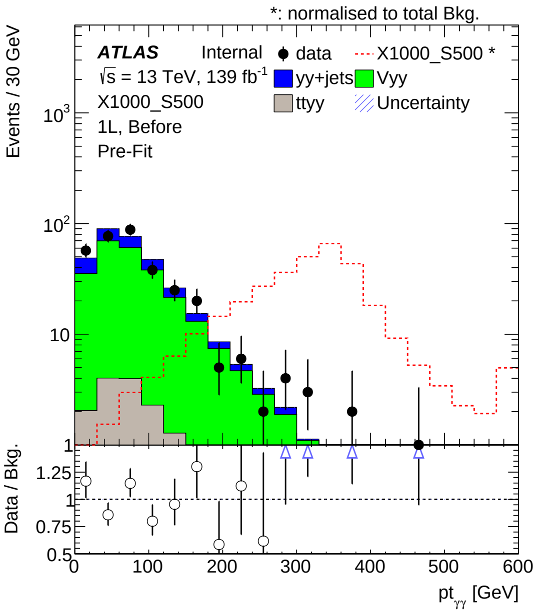
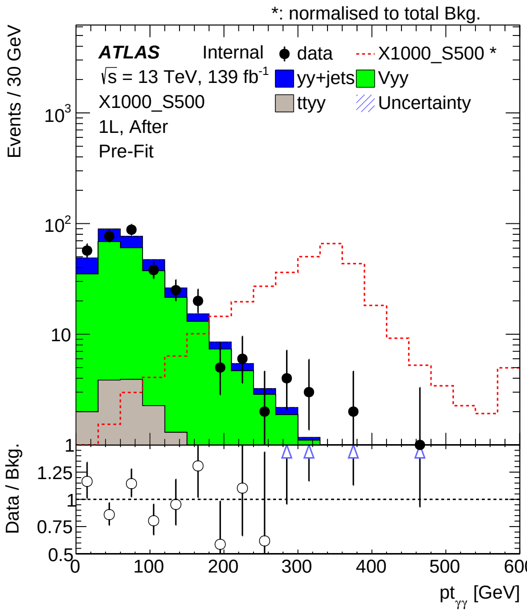


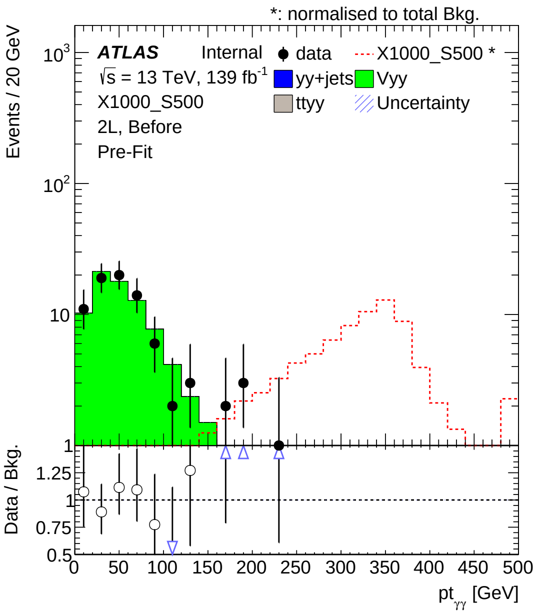
Figure 155: BDT distribution for continuum MC and the sideband data before(left) and after(right) background reweighting.



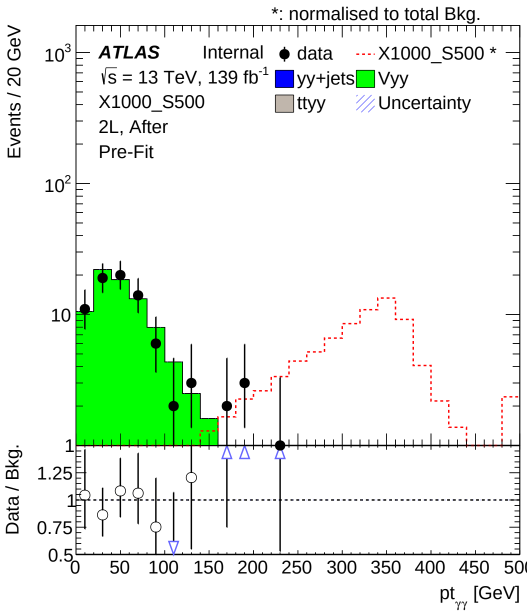
(a)



(b)

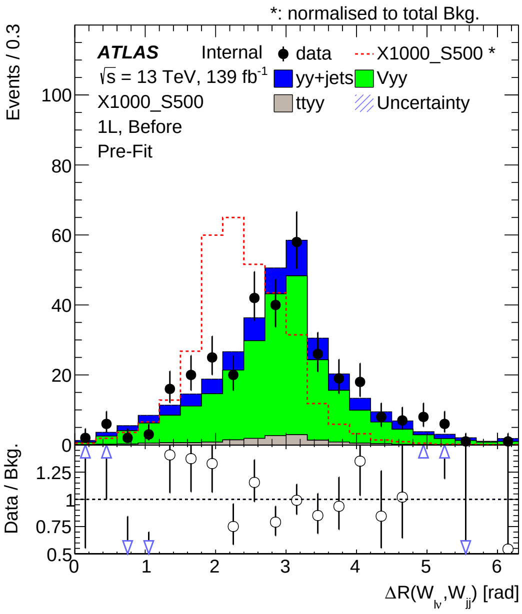


(c)

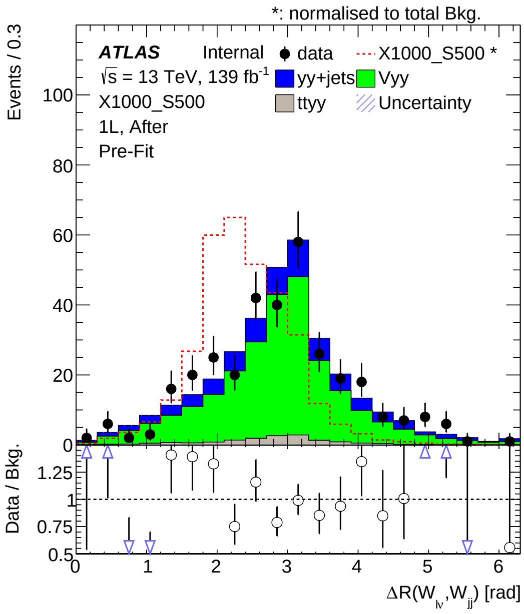


(d)

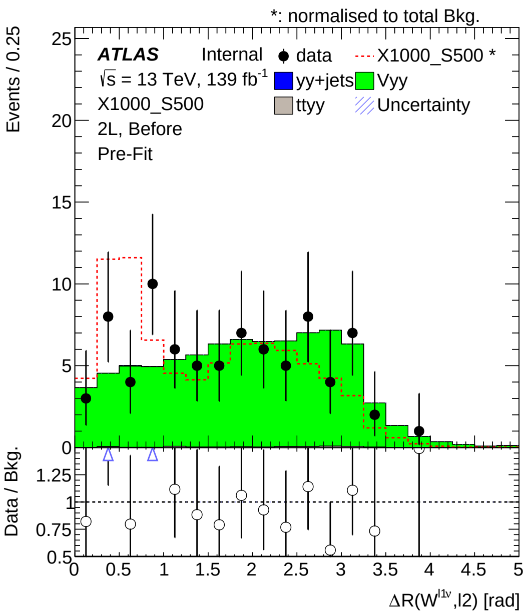
Figure 156: $p_{\mathrm{T}}(\gamma\gamma)$ distribution for continuum MC and the sideband data before(left) and after(right) background reweighting.



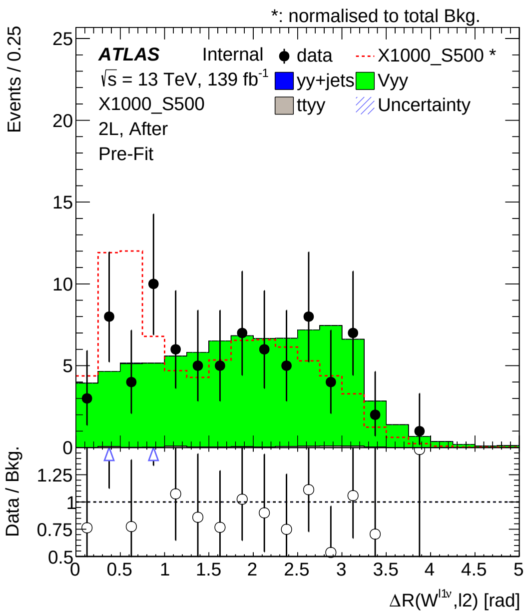
(a)



(b)

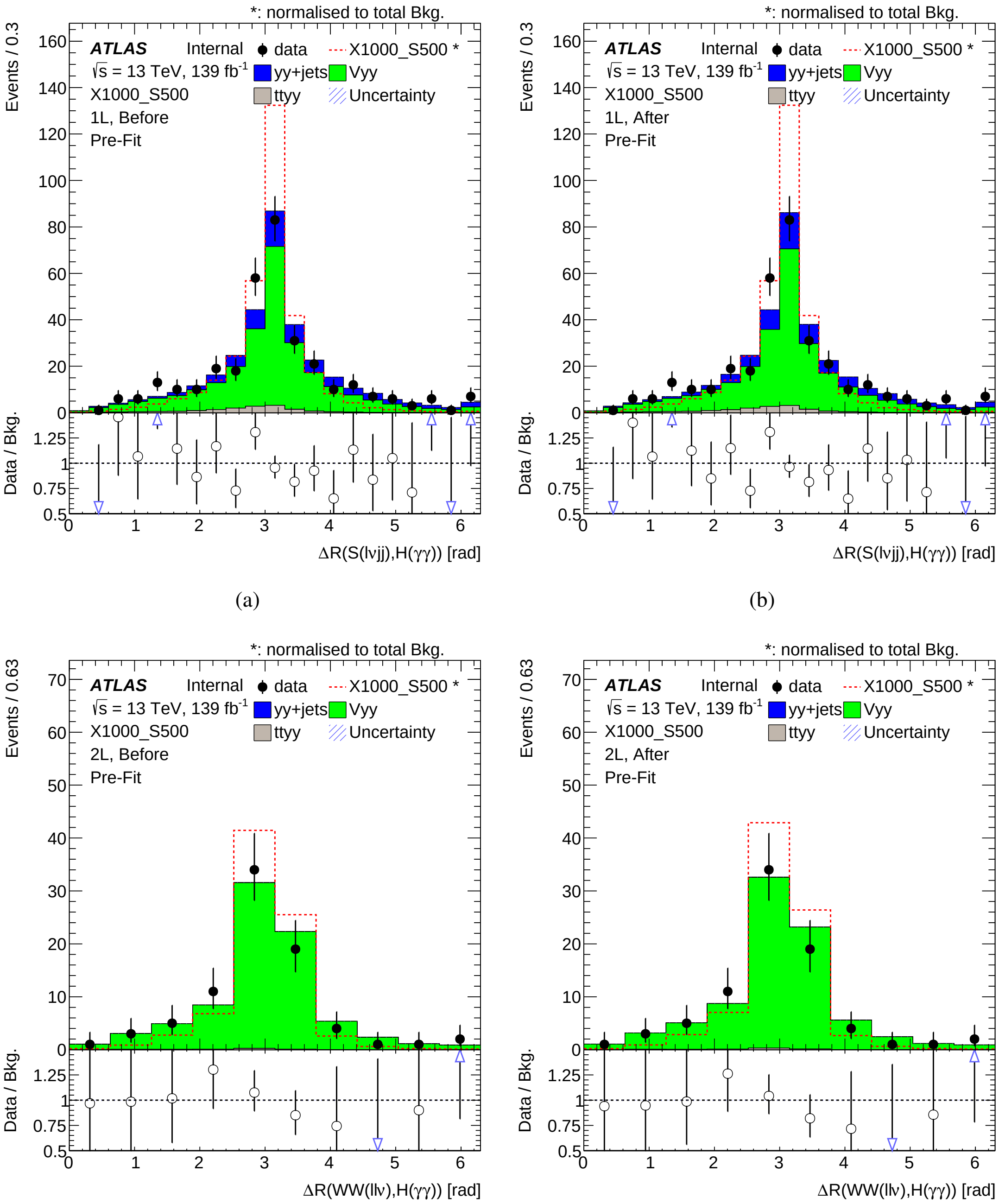


(c)



(d)

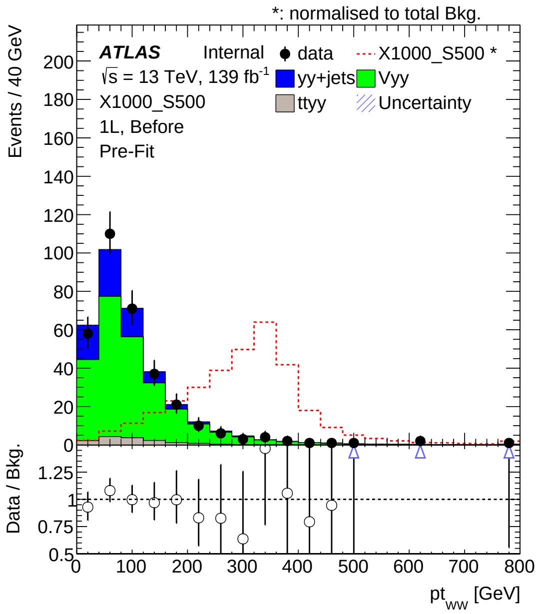
Figure 157: $\Delta R(W1, W2)$ distribution for continuum MC and the sideband data before(left) and after(right) background reweighting.



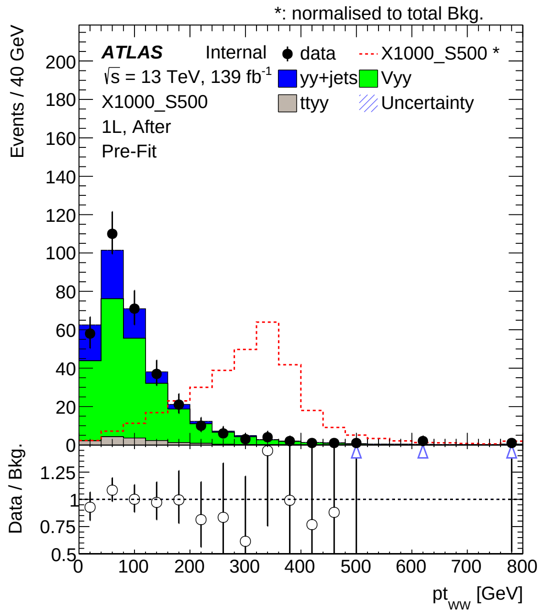
(a) (b)

(c) (d)

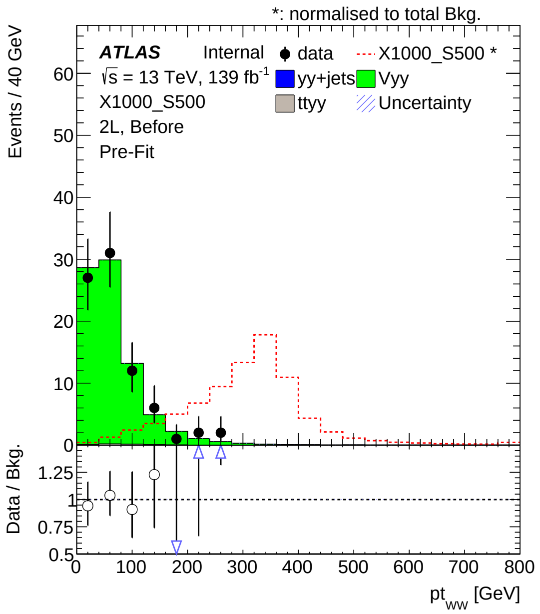
Figure 158: $\Delta R(WW, H)$ distribution for continuum MC and the sideband data before(left) and after(right) background reweighting.



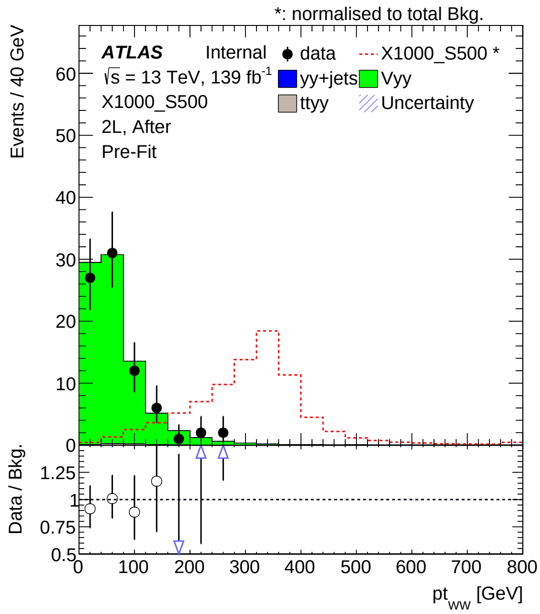
(a)



(b)

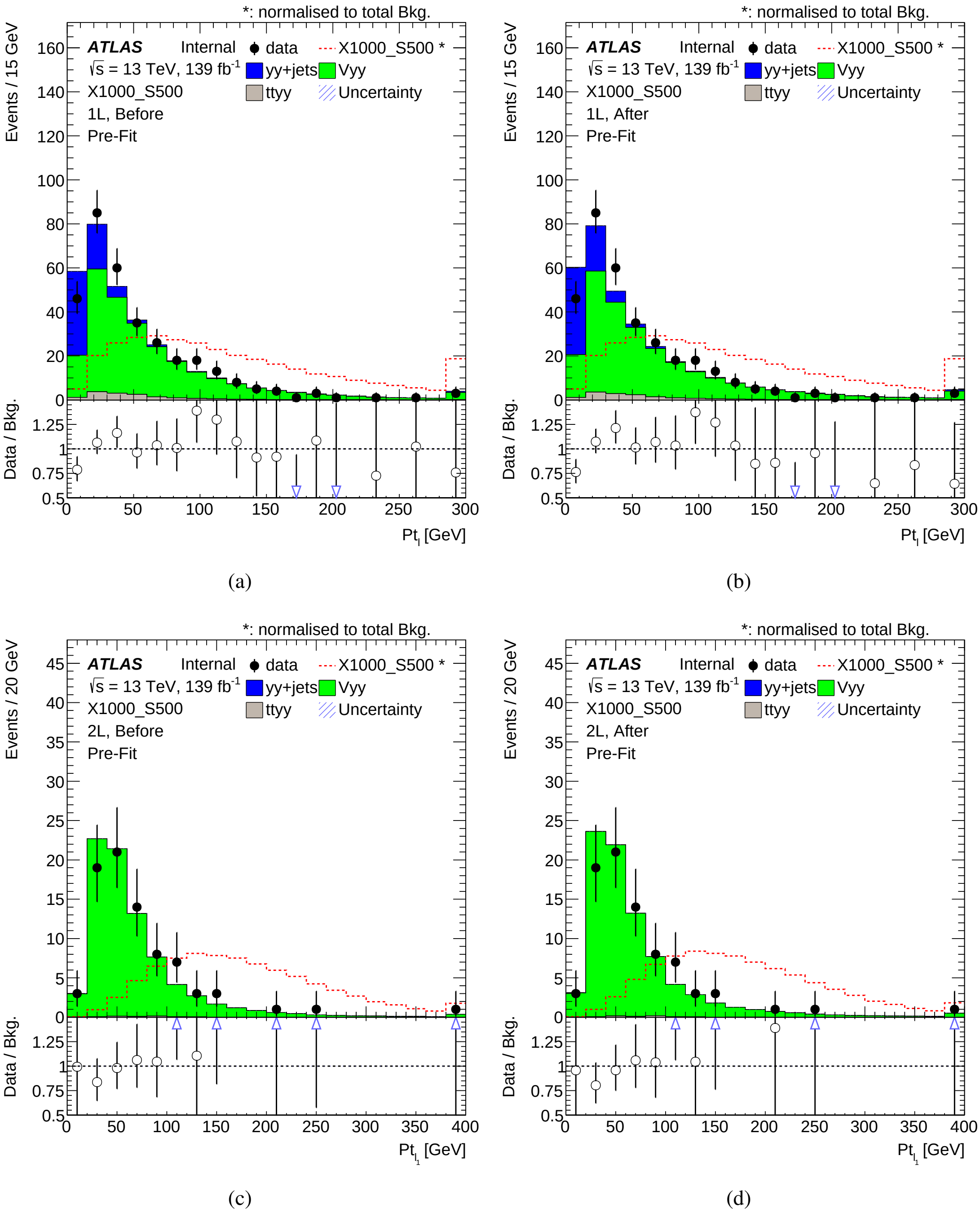


(c)



(d)

Figure 159: $p_{\mathrm{T}}(WW)$ distribution for continuum MC and the sideband data before(left) and after(right) background reweighting.



(a) (b) (c) (d)

Figure 160: $p_{\mathrm{T}}(l1)$ distribution for continuum MC and the sideband data before(left) and after(right) background reweighting.

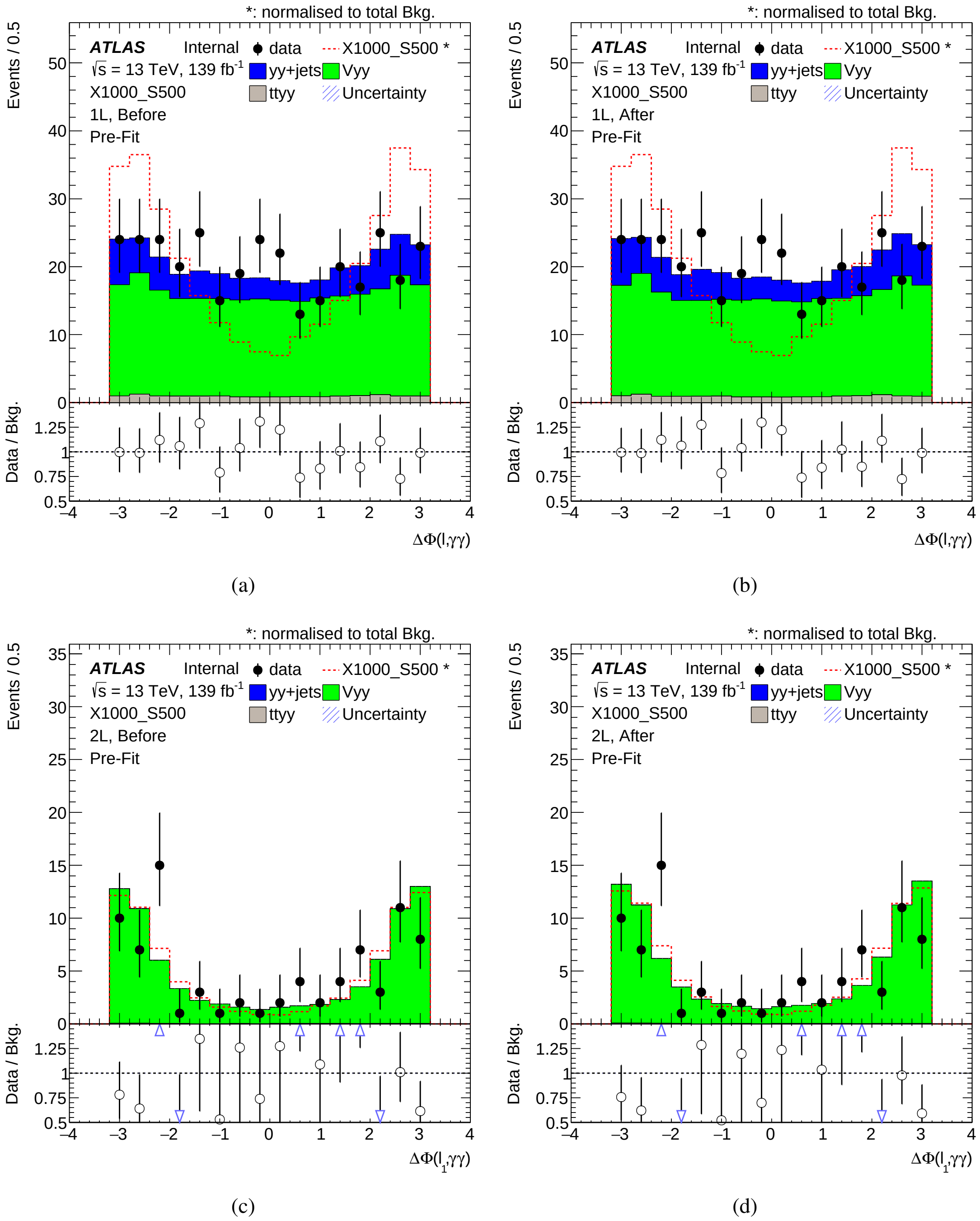
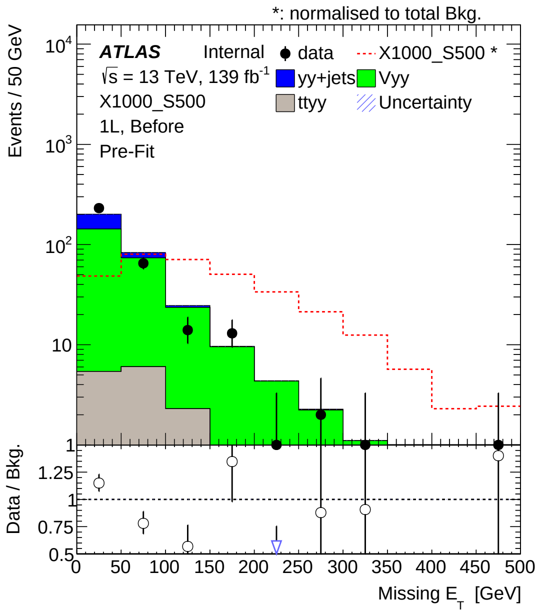
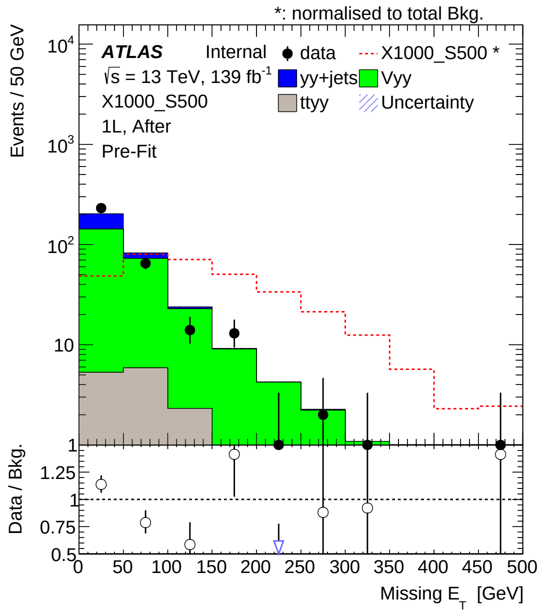


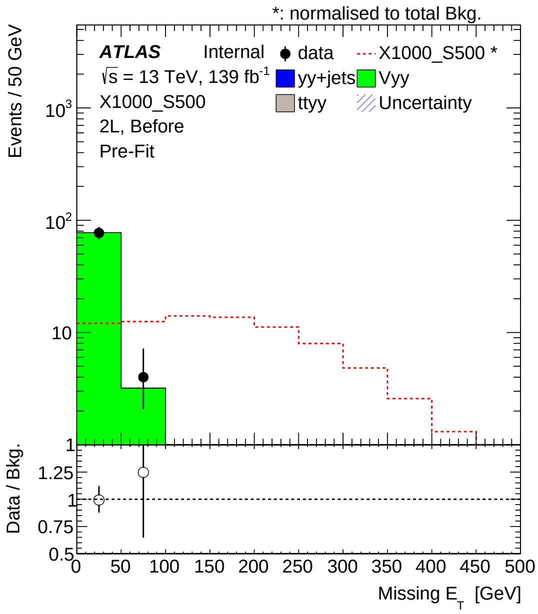
Figure 161: $\Delta\Phi(\gamma\gamma, \ell_1)$ distribution for continuum MC and the sideband data before(left) and after(right) background reweighting.



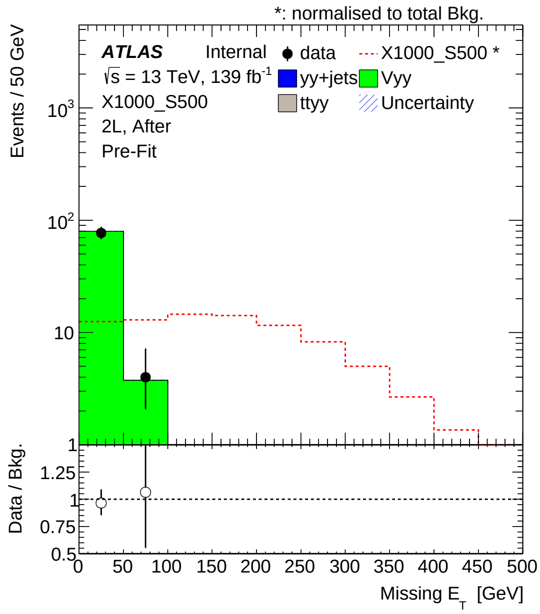
(a)



(b)



(c)



(d)

Figure 162: $E_{\mathrm{T}}^{\mathrm{miss}}$ distribution for continuum MC and the sideband data before(left) and after(right) background reweighting.

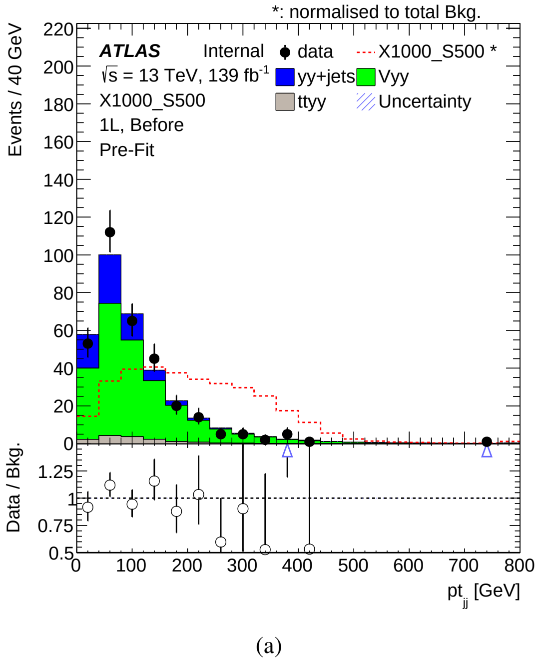


Figure 163: $p_{\mathrm{T}}(jj)$ distribution for 1l continuum MC and the sideband data before(left) and after(right) background reweighting.

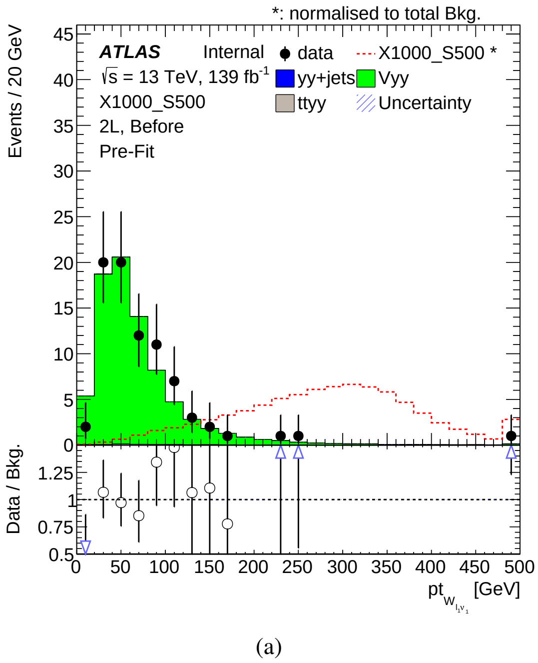
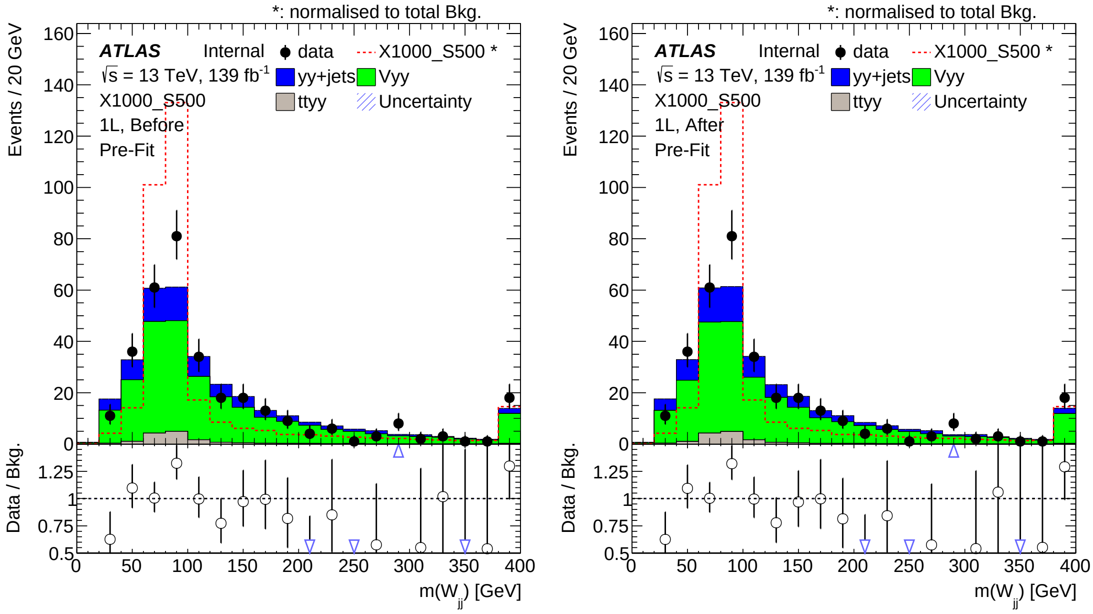
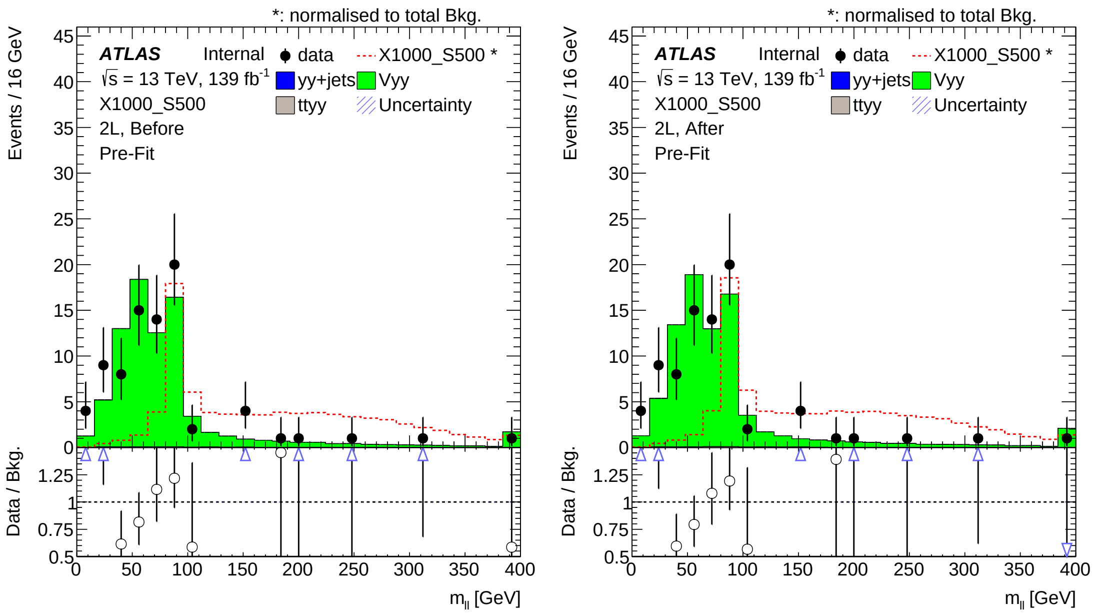


Figure 164: $p_{\mathrm{T}}(l\nu)$ distribution for 2l continuum MC and the sideband data before(left) and after(right) background reweighting.



(a) (b)

Figure 165: $M(jj)$ distribution for 1l continuum MC and the sideband data before(left) and after(right) background reweighting.



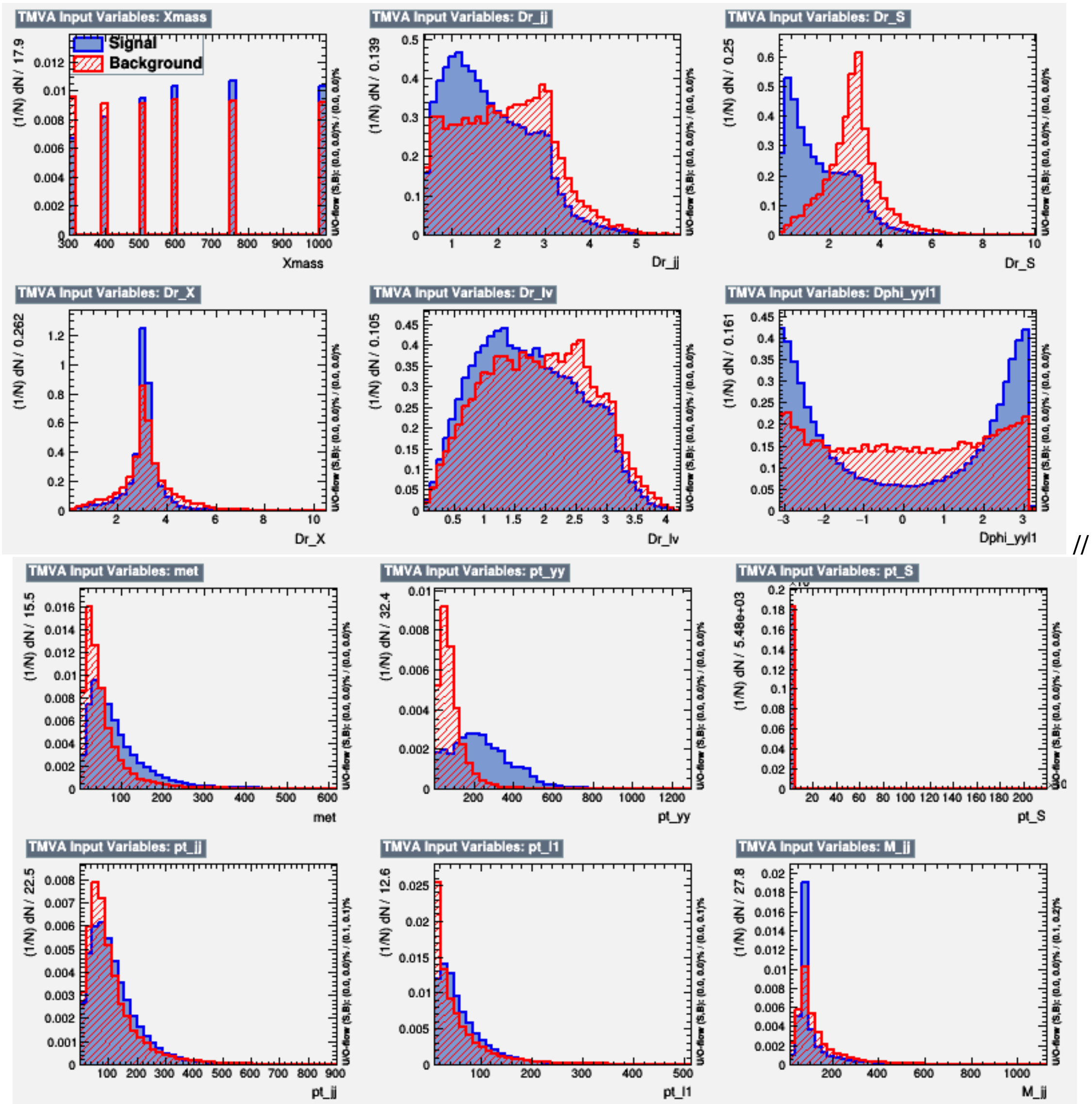
(a) (b)

Figure 166: $m(ll)$ distribution for 2l continuum MC and the sideband data before(left) and after(right) background reweighting.

F BDT training in all mass points

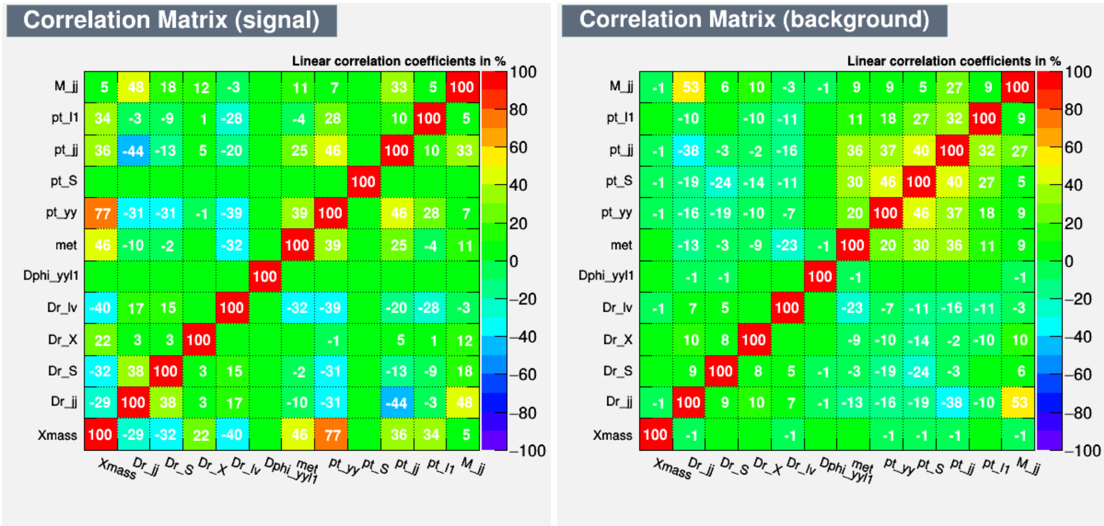
F.1 WW1l category

F.1.1 m_S = 170 GeV



//

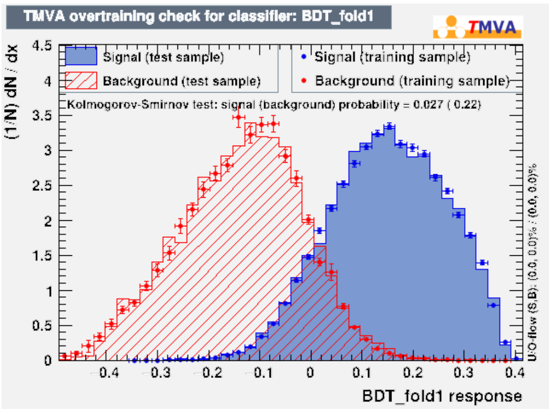
Figure 167: Input variables for BDT in WW1l category.



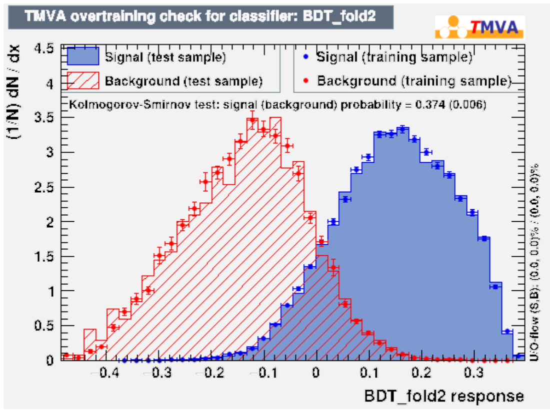
(a) Signal

(b) Background

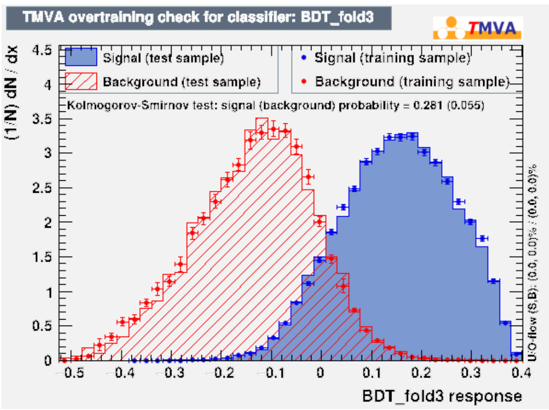
Figure 168: Input variables' correlation in WW1l category.



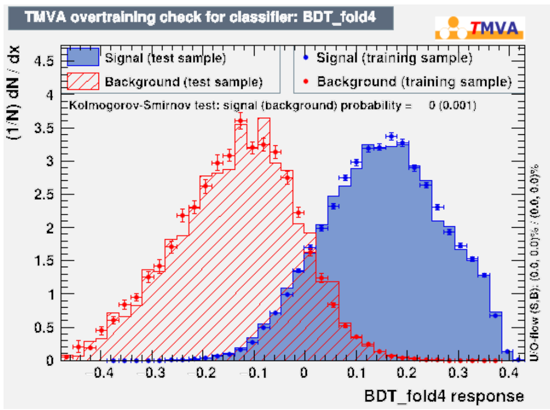
(a) fold1



(b) fold2

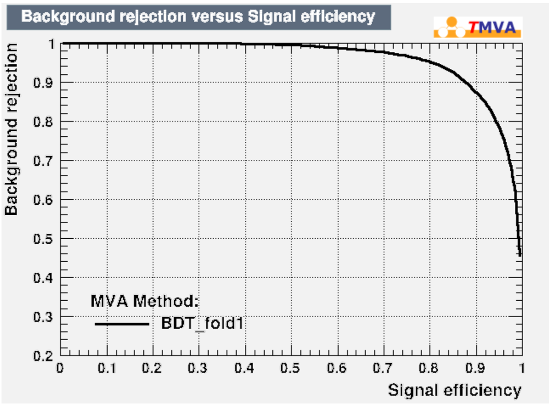


(c) fold3

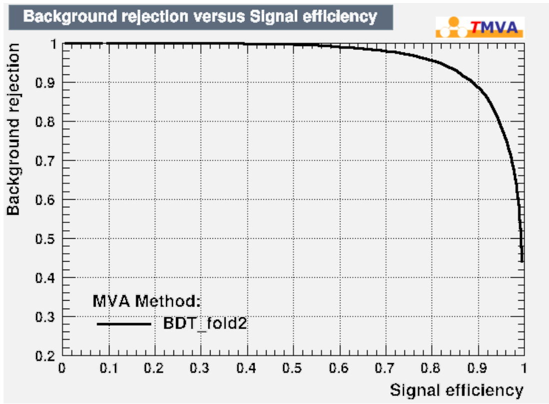


(d) fold4

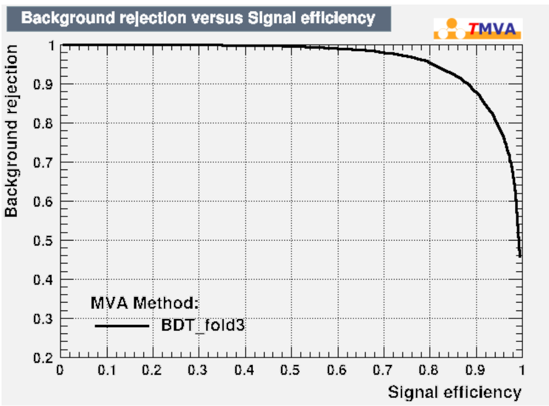
Figure 169: BDT distributions for training and test templates in 4 folds



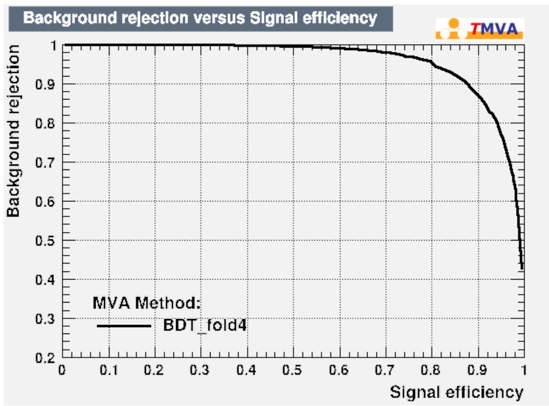
(a) fold1



(b) fold2



(c) fold3



(d) fold4

Figure 170: Trainned BDT ROC in 4 folds

1068 F.1.2 m_S = 200 GeV

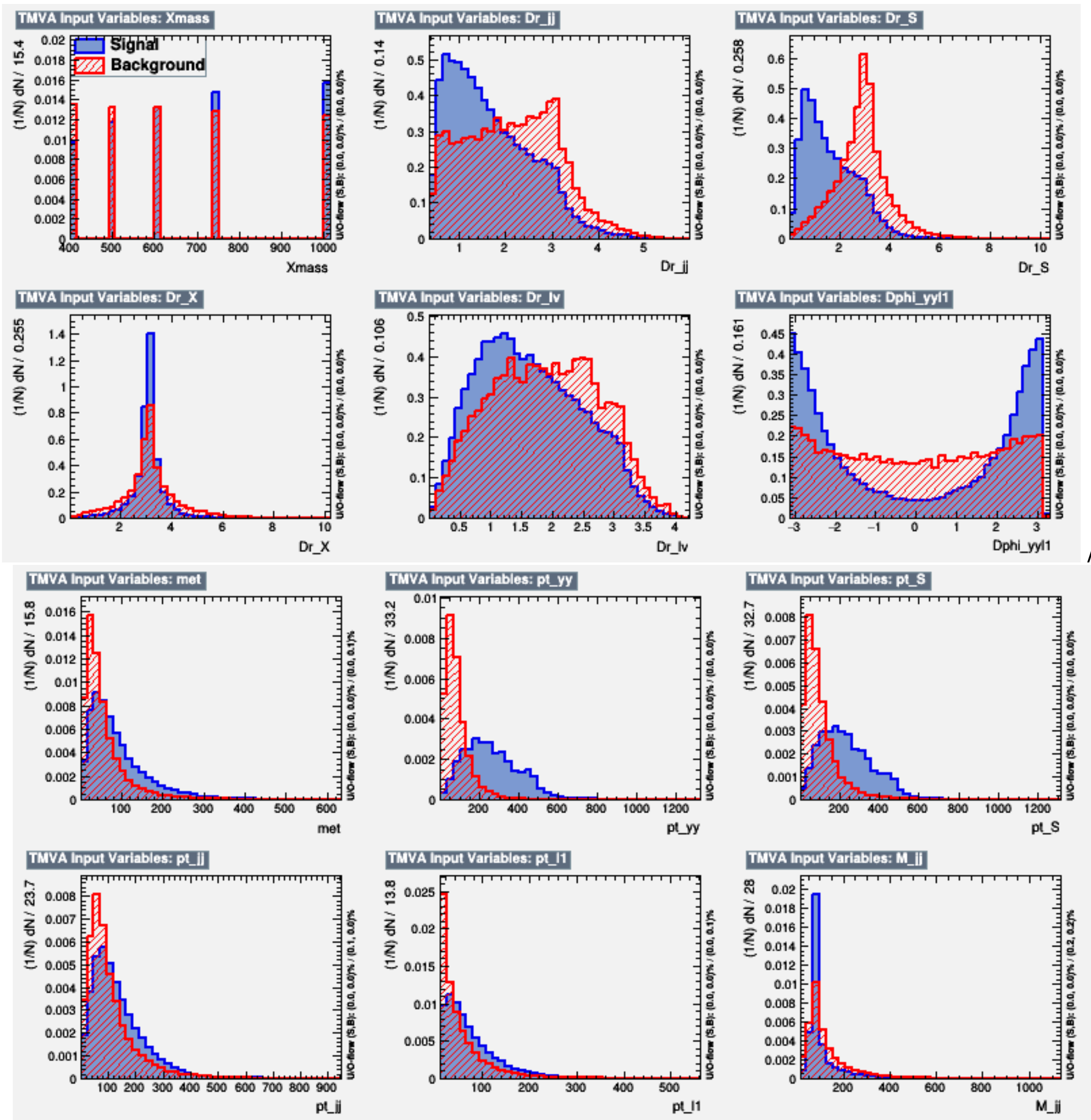
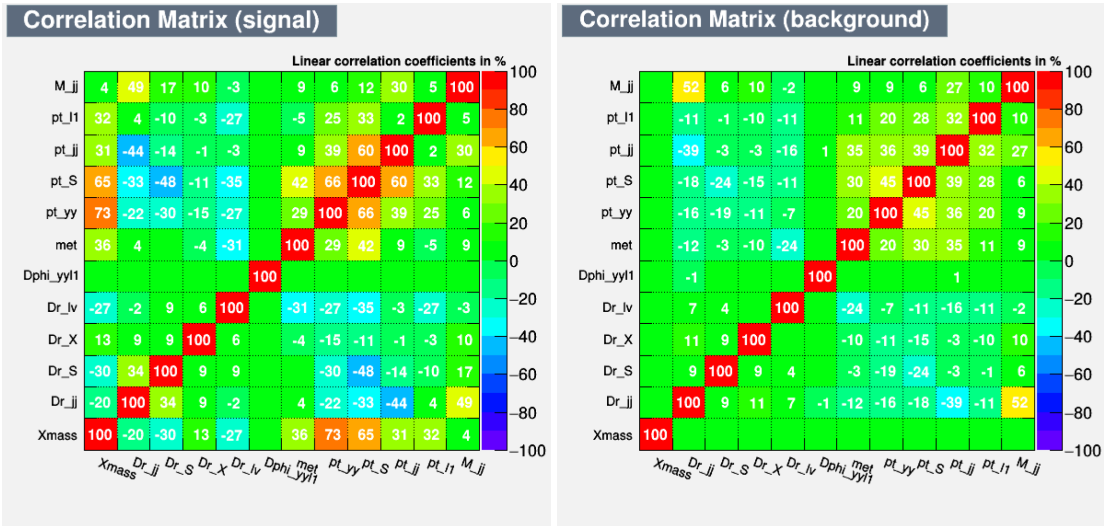


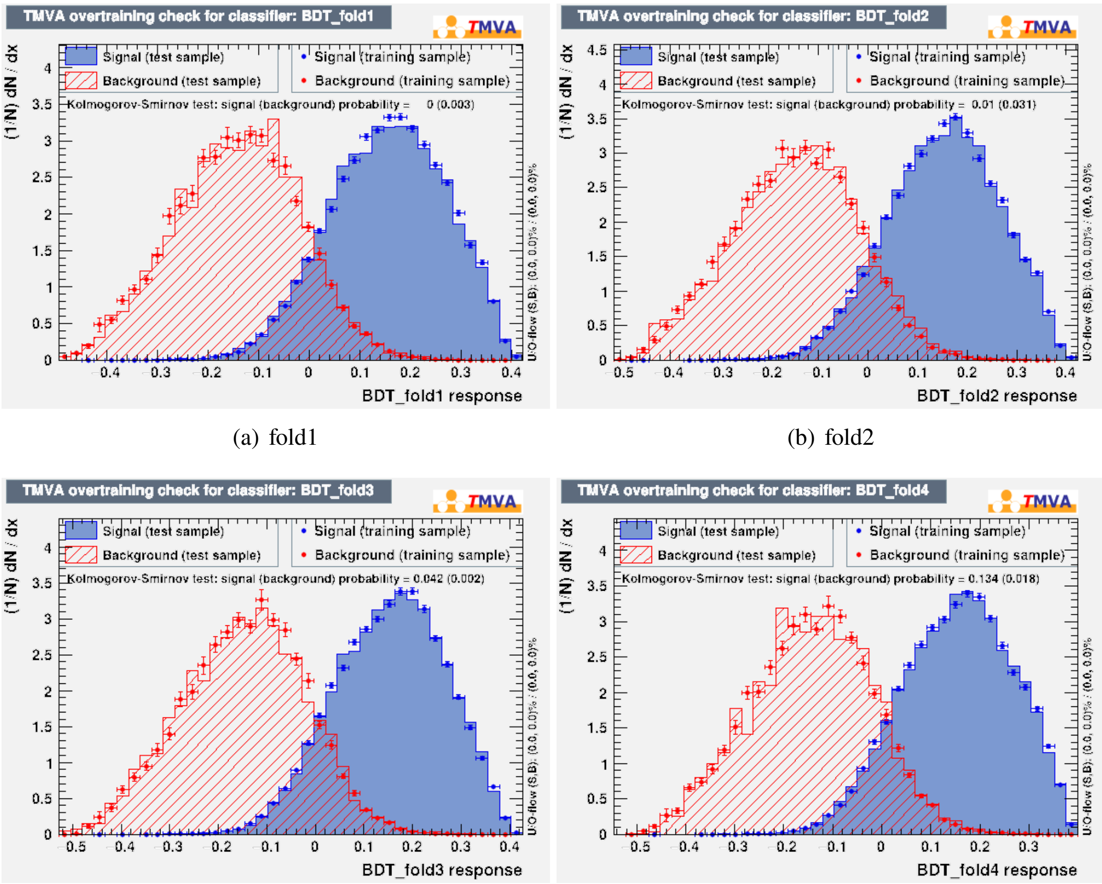
Figure 171: Input variables for BDT in WW1l category.



(a) Signal

(b) Background

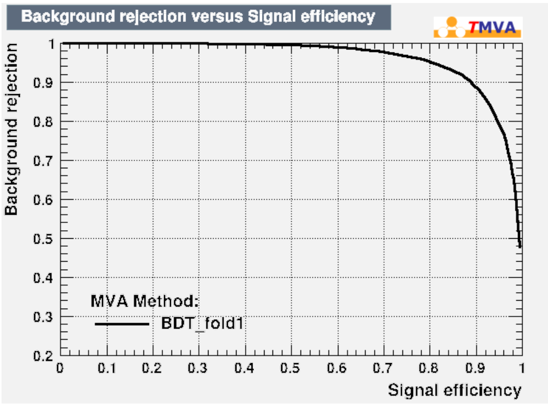
Figure 172: Input variables' correlation in WW1l category.



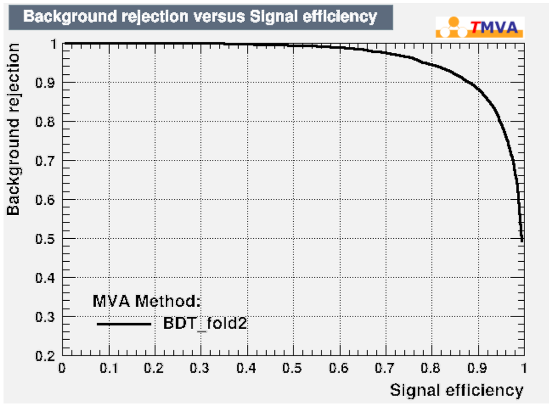
(a) fold1 (b) fold2

(c) fold3 (d) fold4

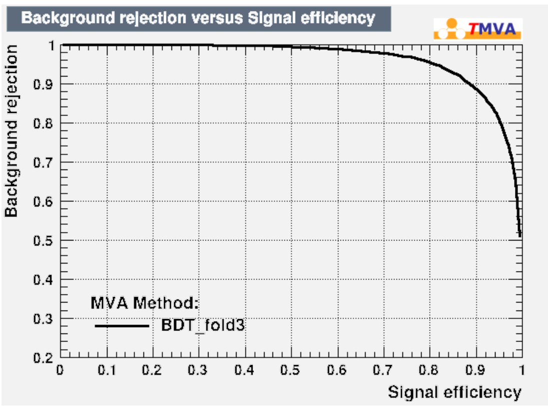
Figure 173: BDT distributions for training and test templates in 4 folds



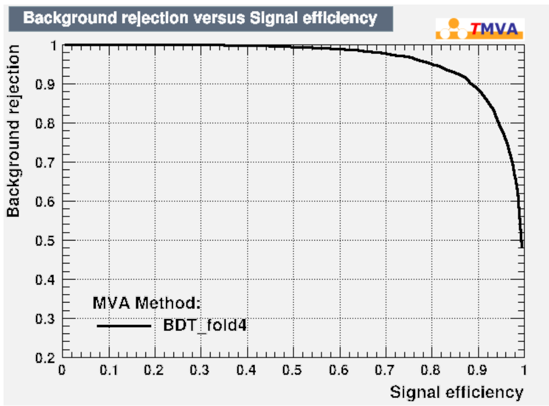
(a) fold1



(b) fold2



(c) fold3



(d) fold4

Figure 174: Trainned BDT ROC in 4 folds

F.1.3 m_S = 300 GeV

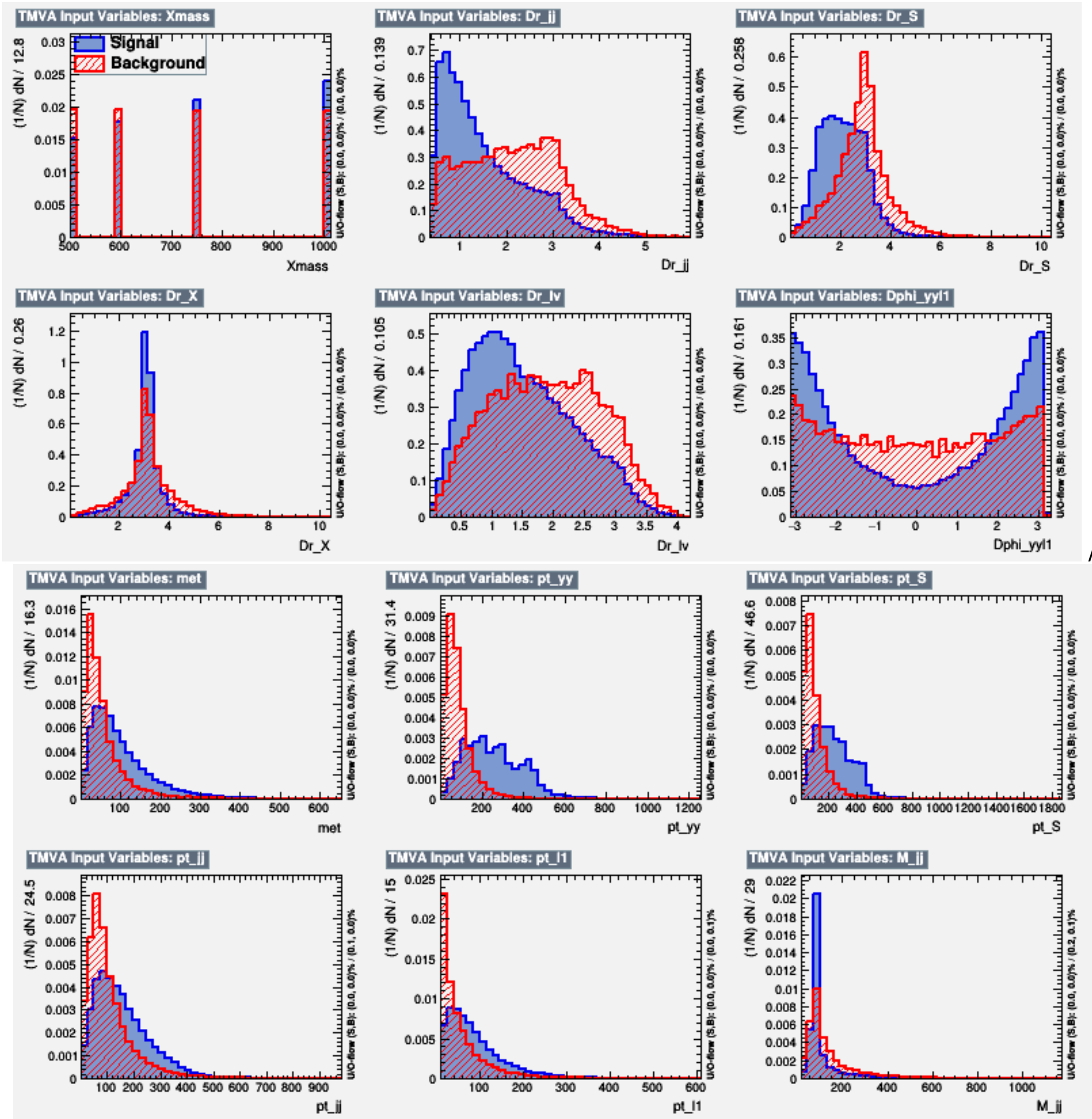
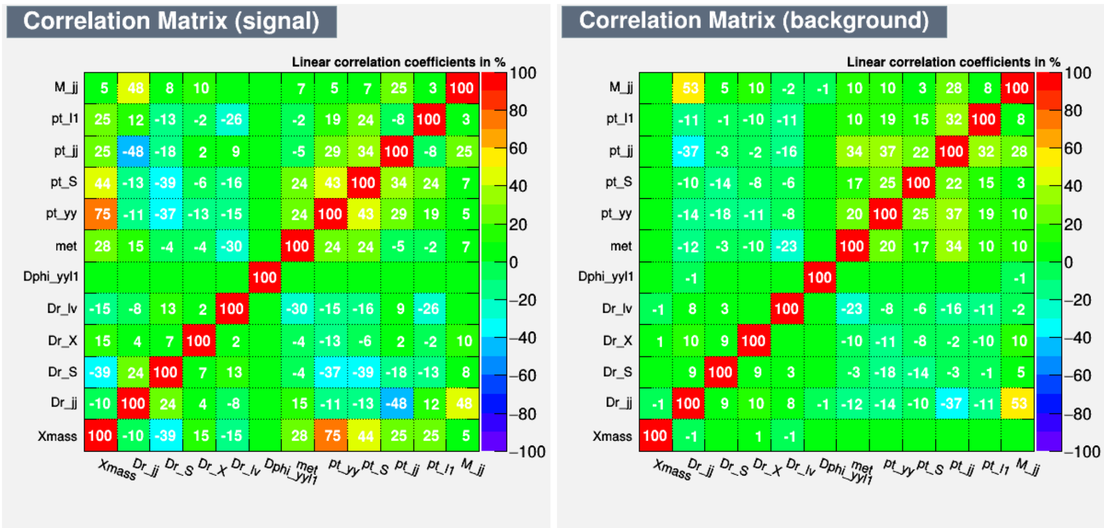


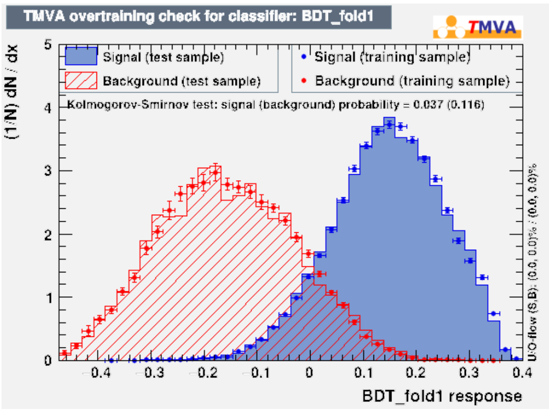
Figure 175: Input variables for BDT in WW1l category.



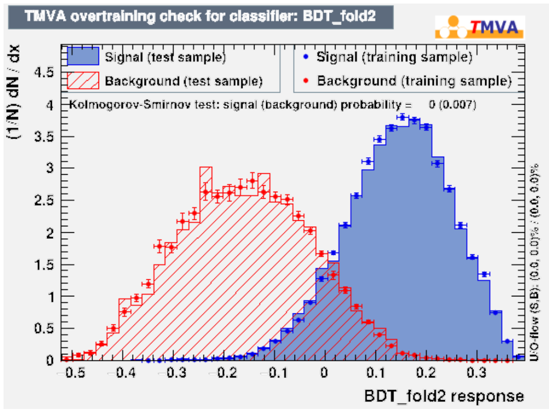
(a) Signal

(b) Background

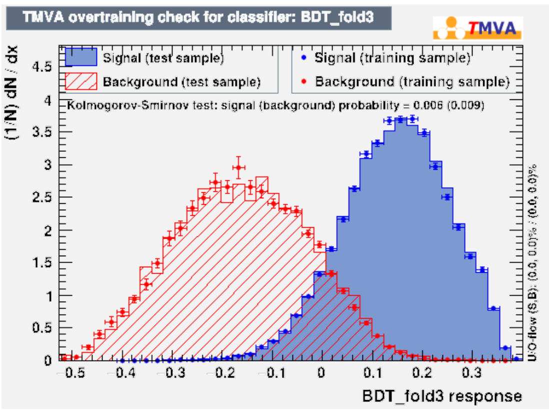
Figure 176: Input variables' correlation in WW1l category.



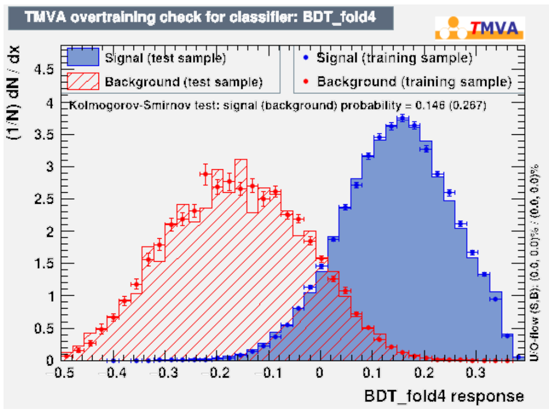
(a) fold1



(b) fold2

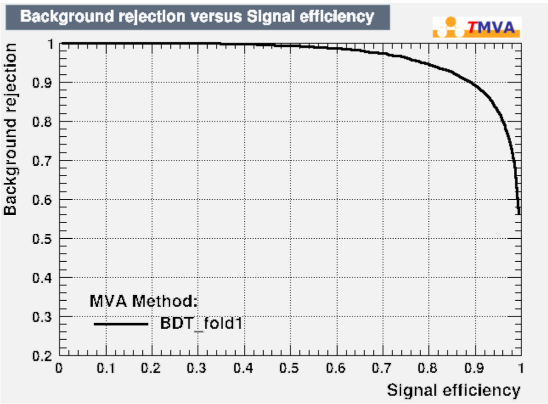


(c) fold3

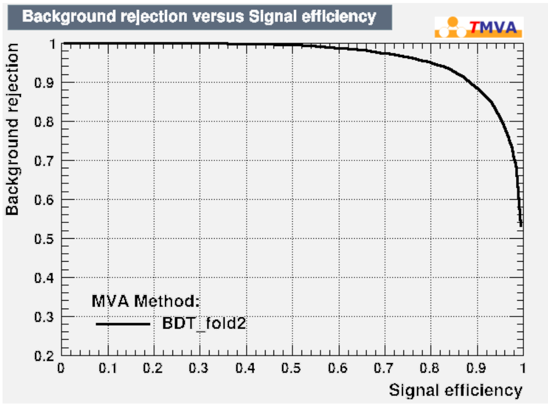


(d) fold4

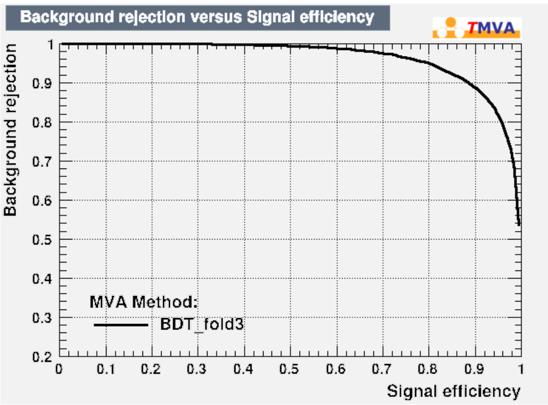
Figure 177: BDT distributions for training and test templates in 4 folds



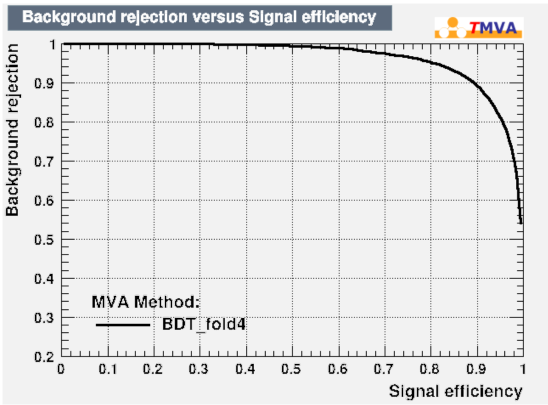
(a) fold1



(b) fold2



(c) fold3



(d) fold4

Figure 178: Trainned BDT ROC in 4 folds

F.1.4 $m_S \geq 400$ GeV

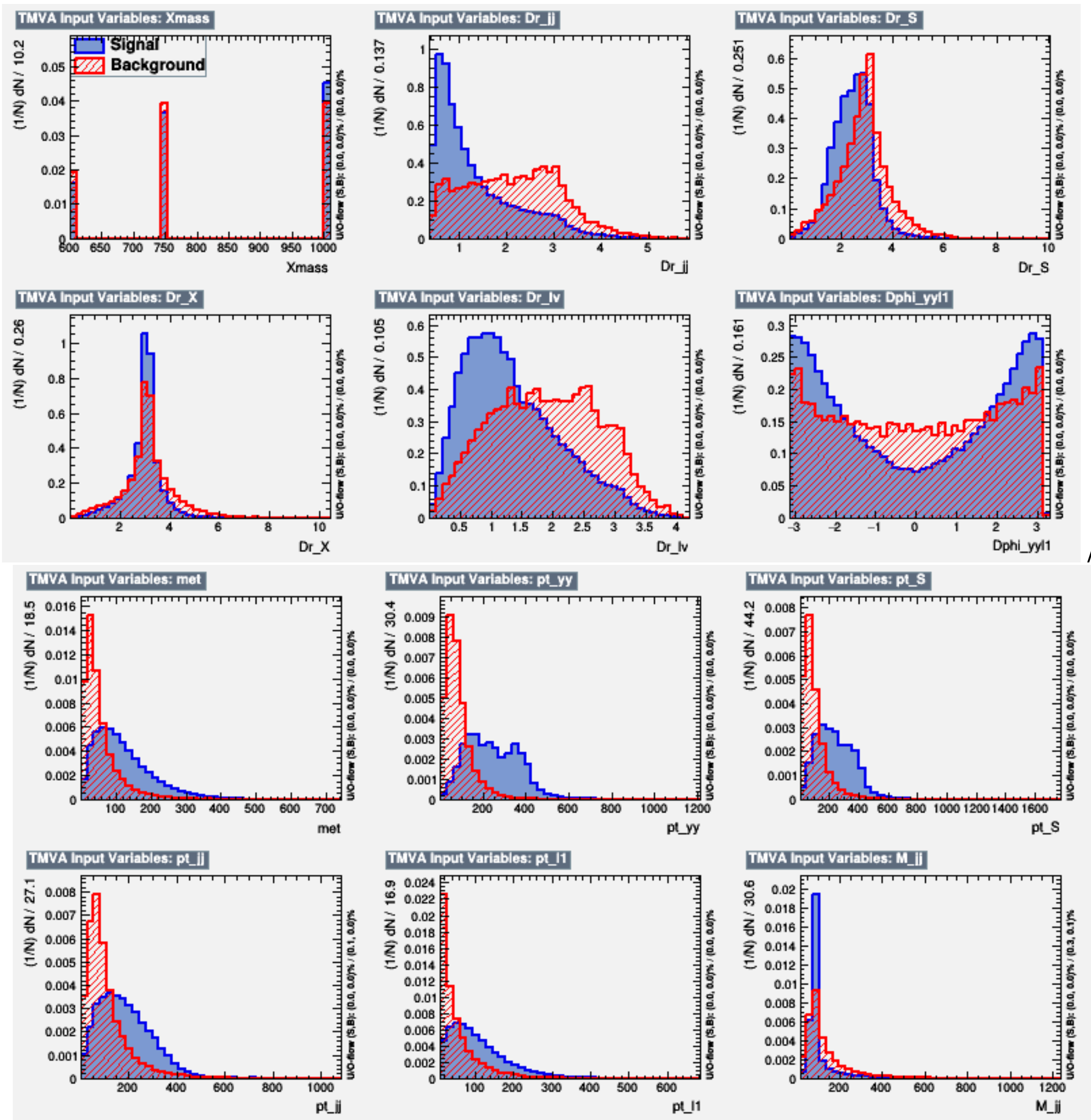
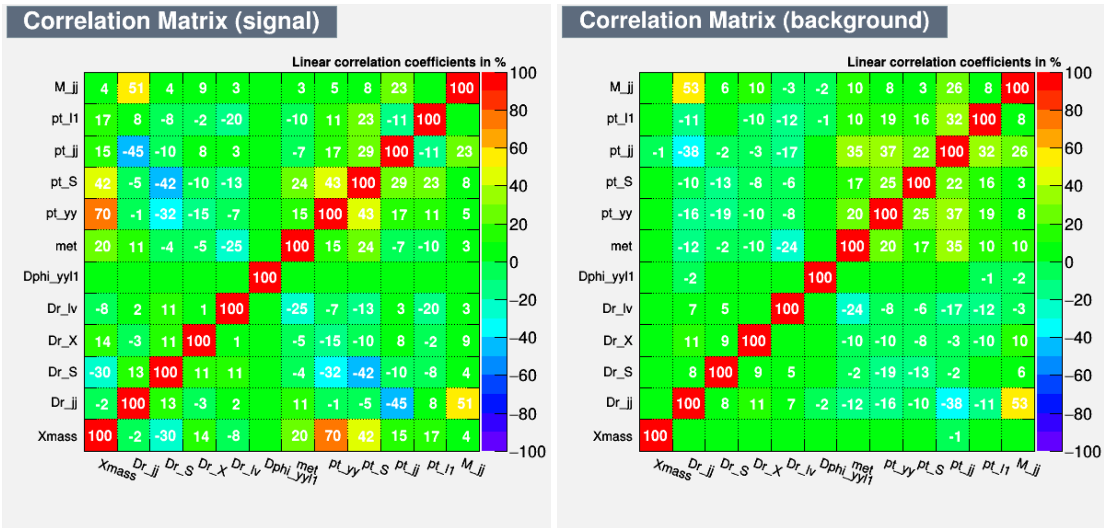


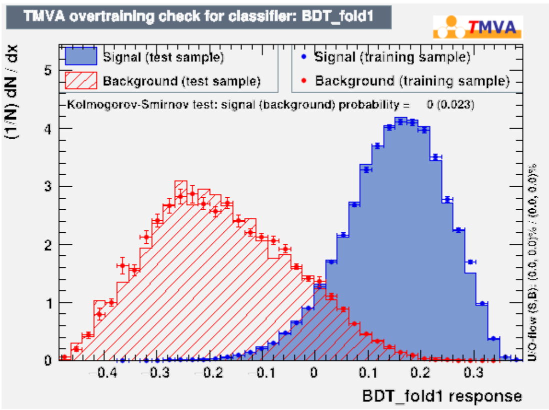
Figure 179: Input variables for BDT in WW1l category.



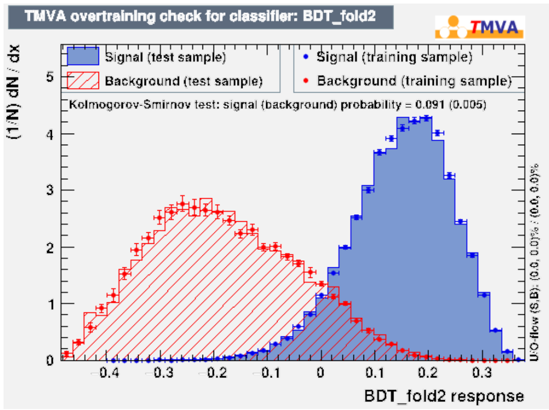
(a) Signal

(b) Background

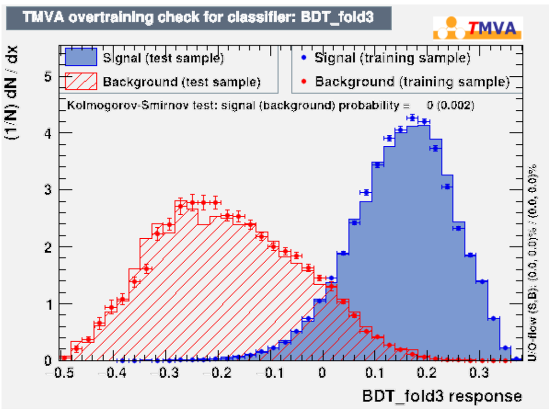
Figure 180: Input variables' correlation in WW1l category.



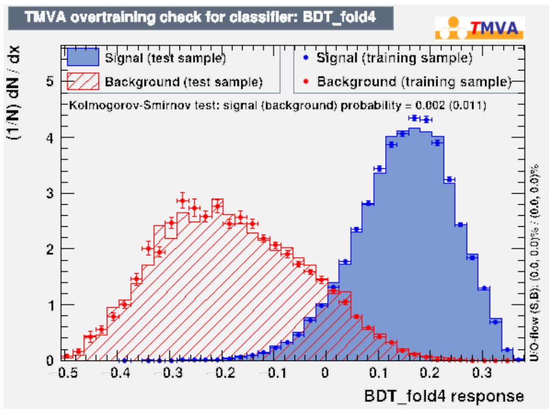
(a) fold1



(b) fold2

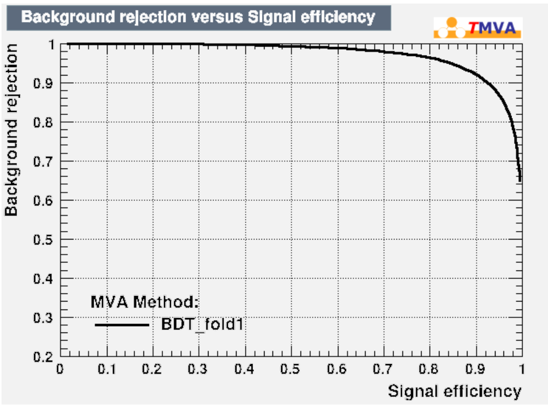


(c) fold3

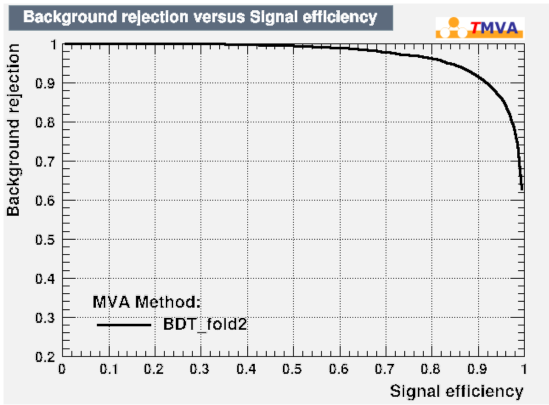


(d) fold4

Figure 181: BDT distributions for training and test templates in 4 folds

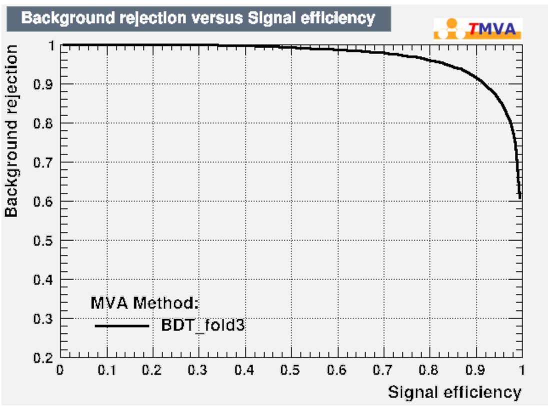


(a) fold1

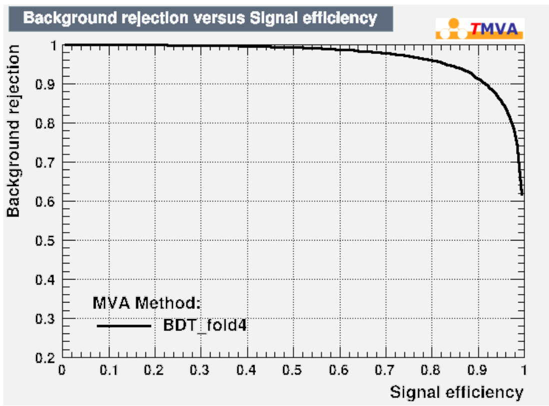


(b) fold2

//



(c) fold3



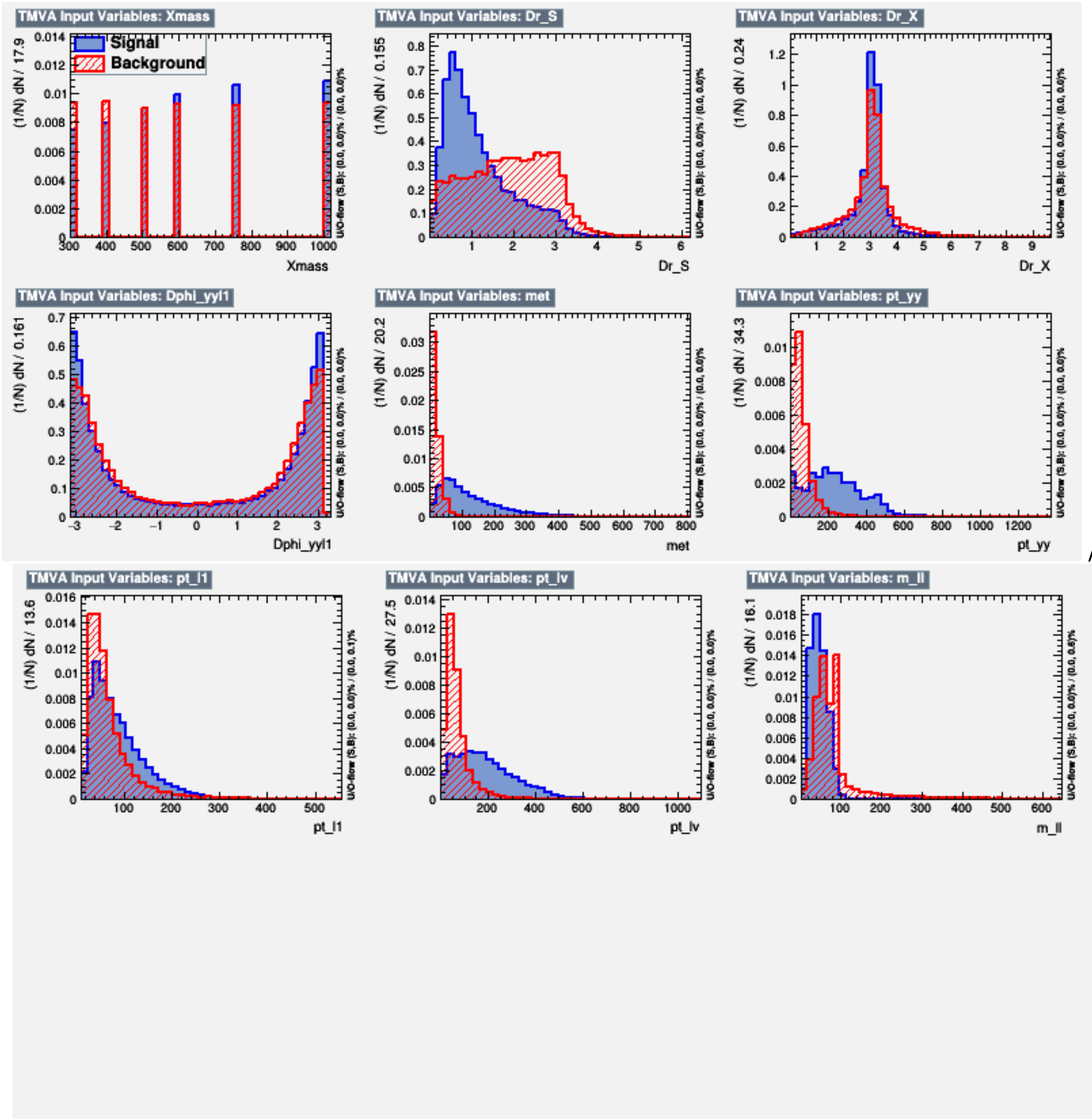
(d) fold4

//

Figure 182: Trainned BDT ROC in 4 folds

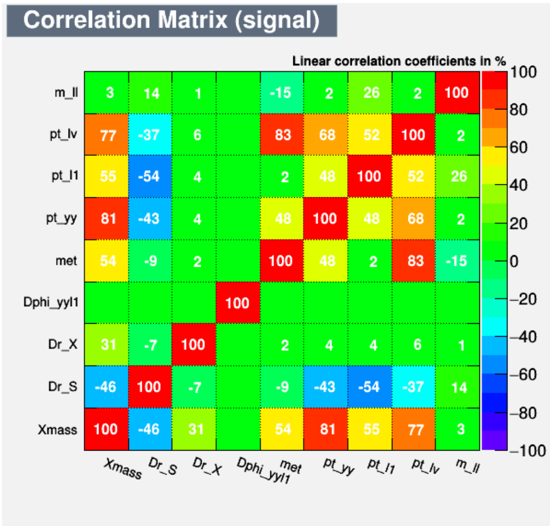
F.2 WW2l category

F.2.1 m_S = 170 GeV

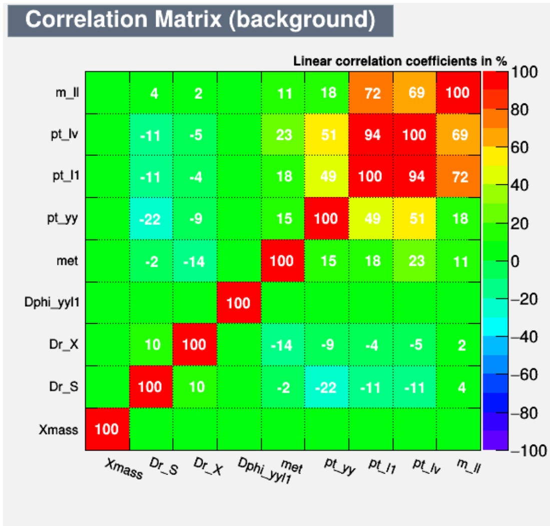


//

Figure 183: Input variables for BDT in WW1l category.

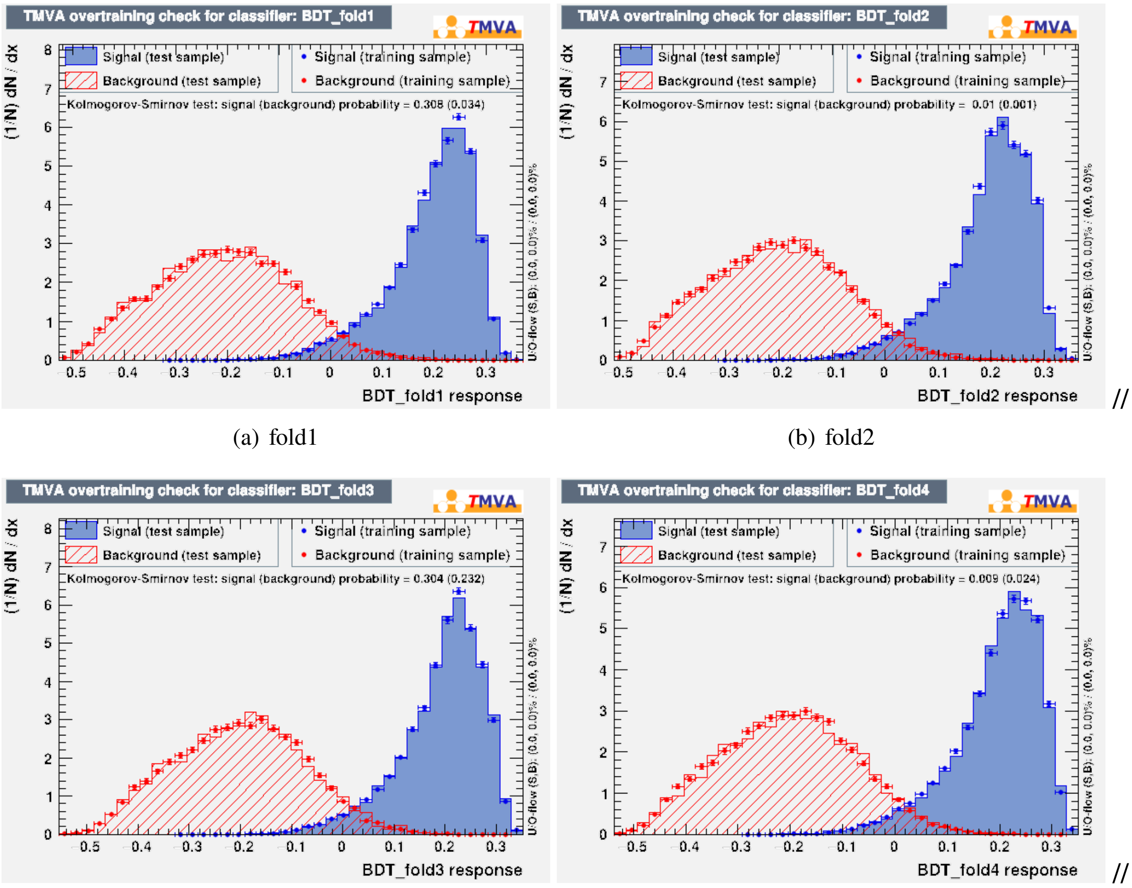


(a) Signal



(b) Background

Figure 184: Input variables' correlation in WW1l category.



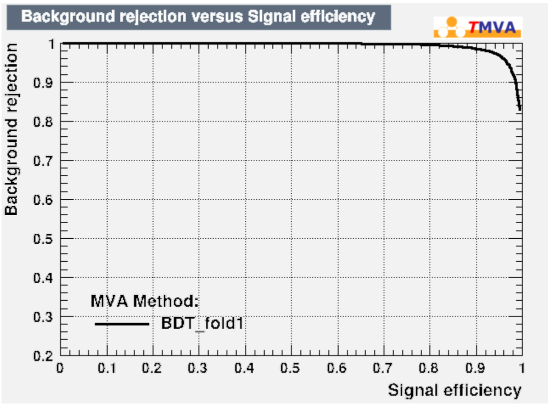
(a) fold1

(b) fold2

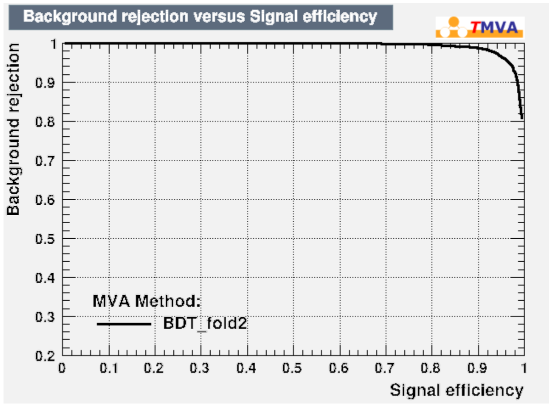
(c) fold3

(d) fold4

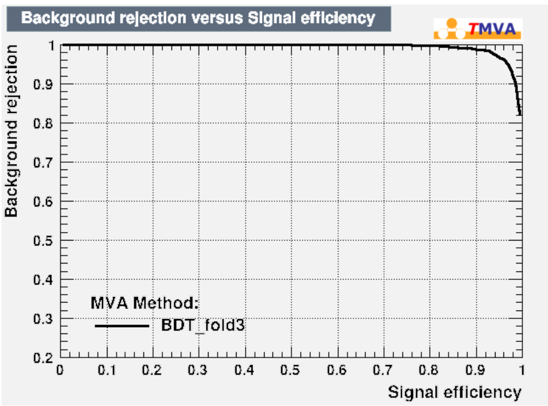
Figure 185: BDT distributions for training and test templates in 4 folds



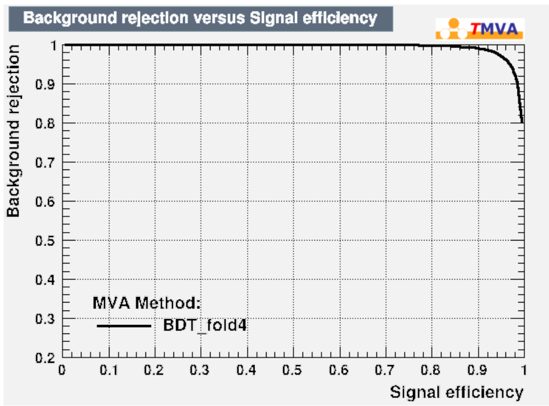
(a) fold1



(b) fold2



(c) fold3



(d) fold4

Figure 186: Trainned BDT ROC in 4 folds

F.2.2 m_S = 200 GeV

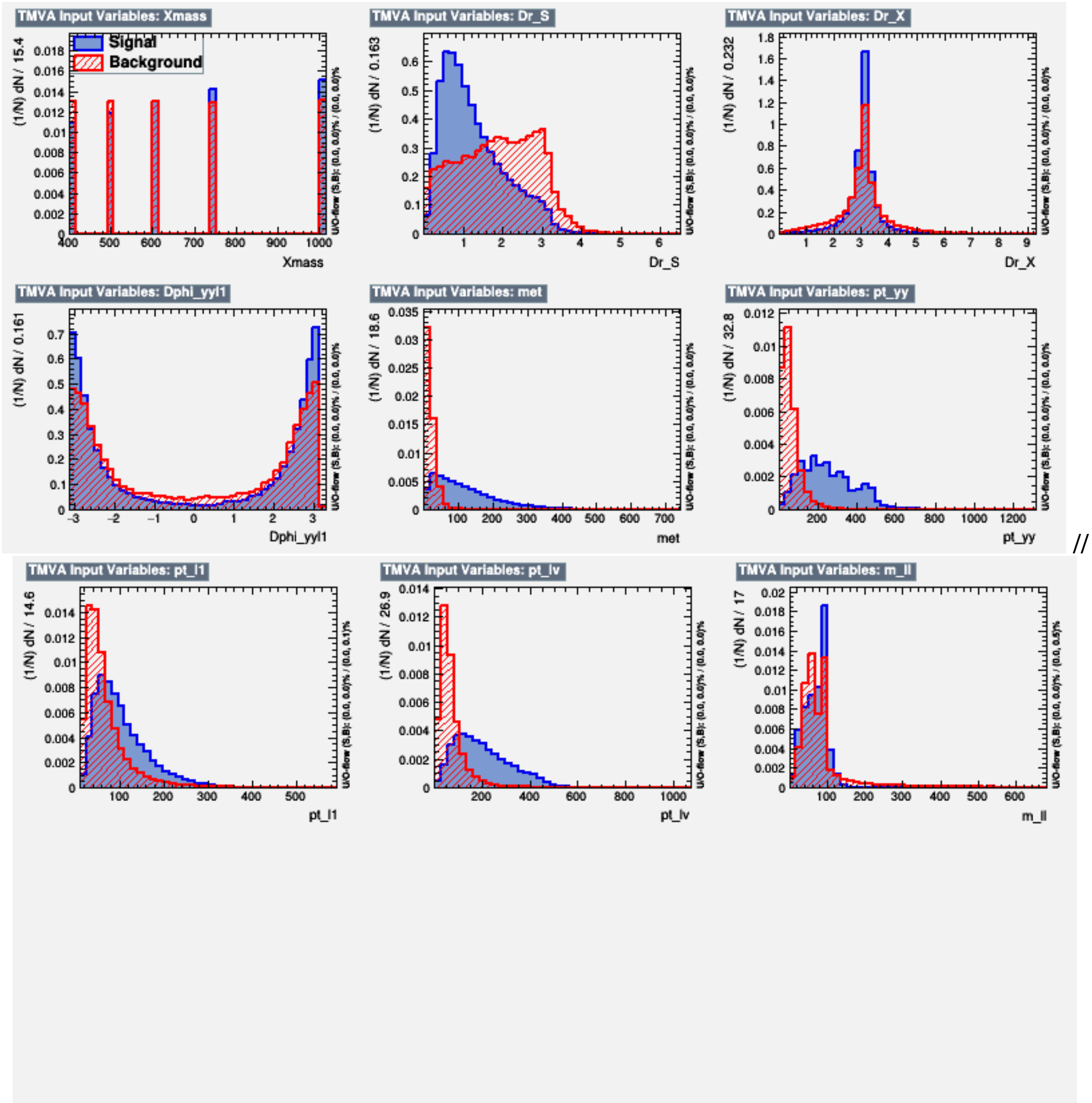
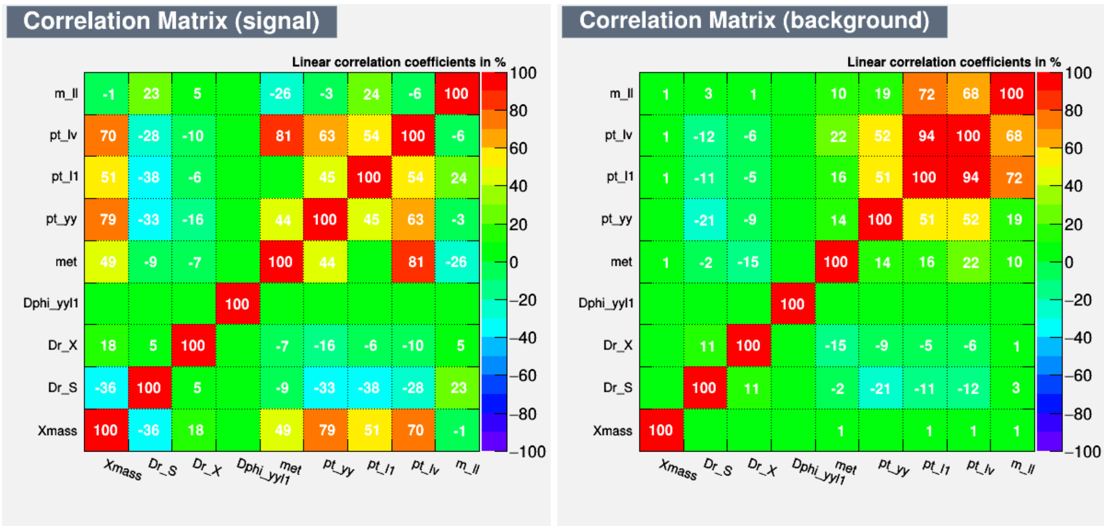


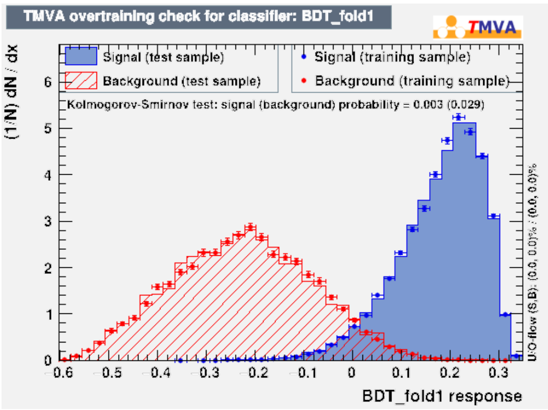
Figure 187: Input variables for BDT in WW1l category.



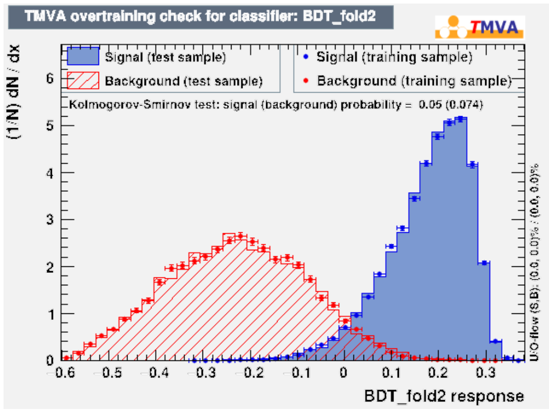
(a) Signal

(b) Background

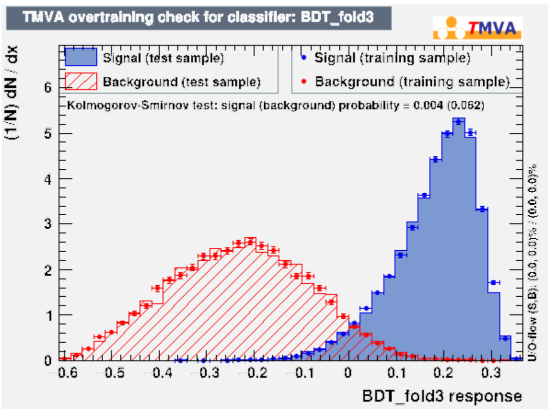
Figure 188: Input variables' correlation in WW1l category.



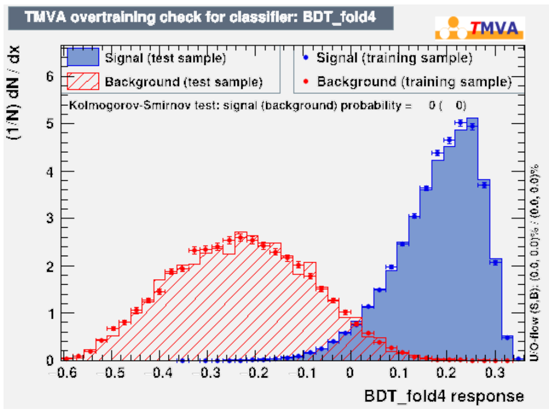
(a) fold1



(b) fold2

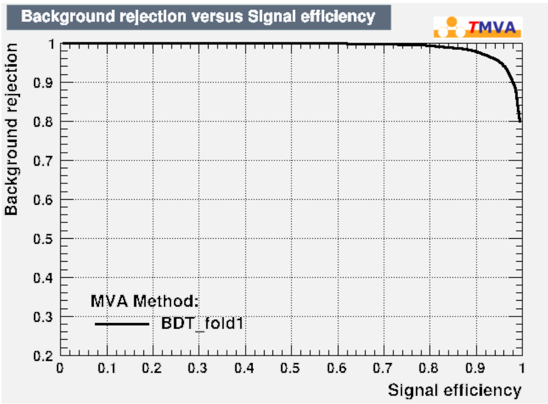


(c) fold3

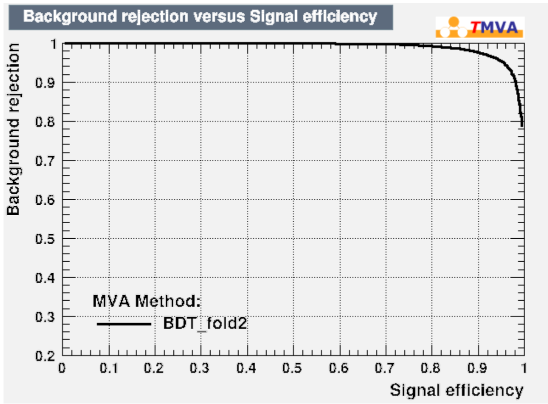


(d) fold4

Figure 189: BDT distributions for training and test templates in 4 folds

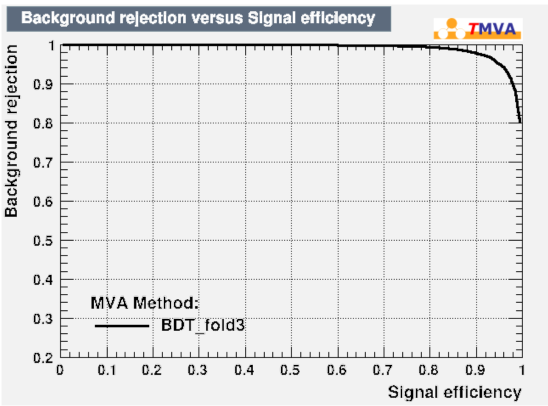


(a) fold1

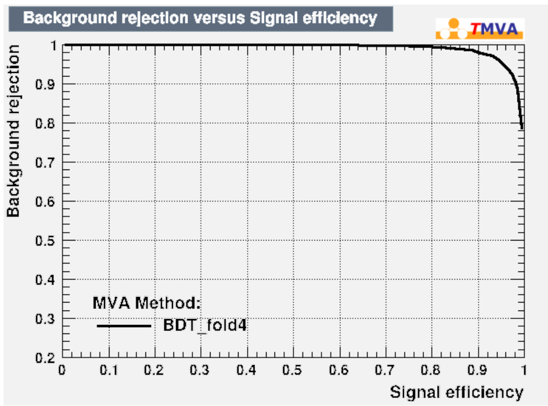


(b) fold2

//



(c) fold3



(d) fold4

//

Figure 190: Trainned BDT ROC in 4 folds

F.2.3 m_S = 300 GeV

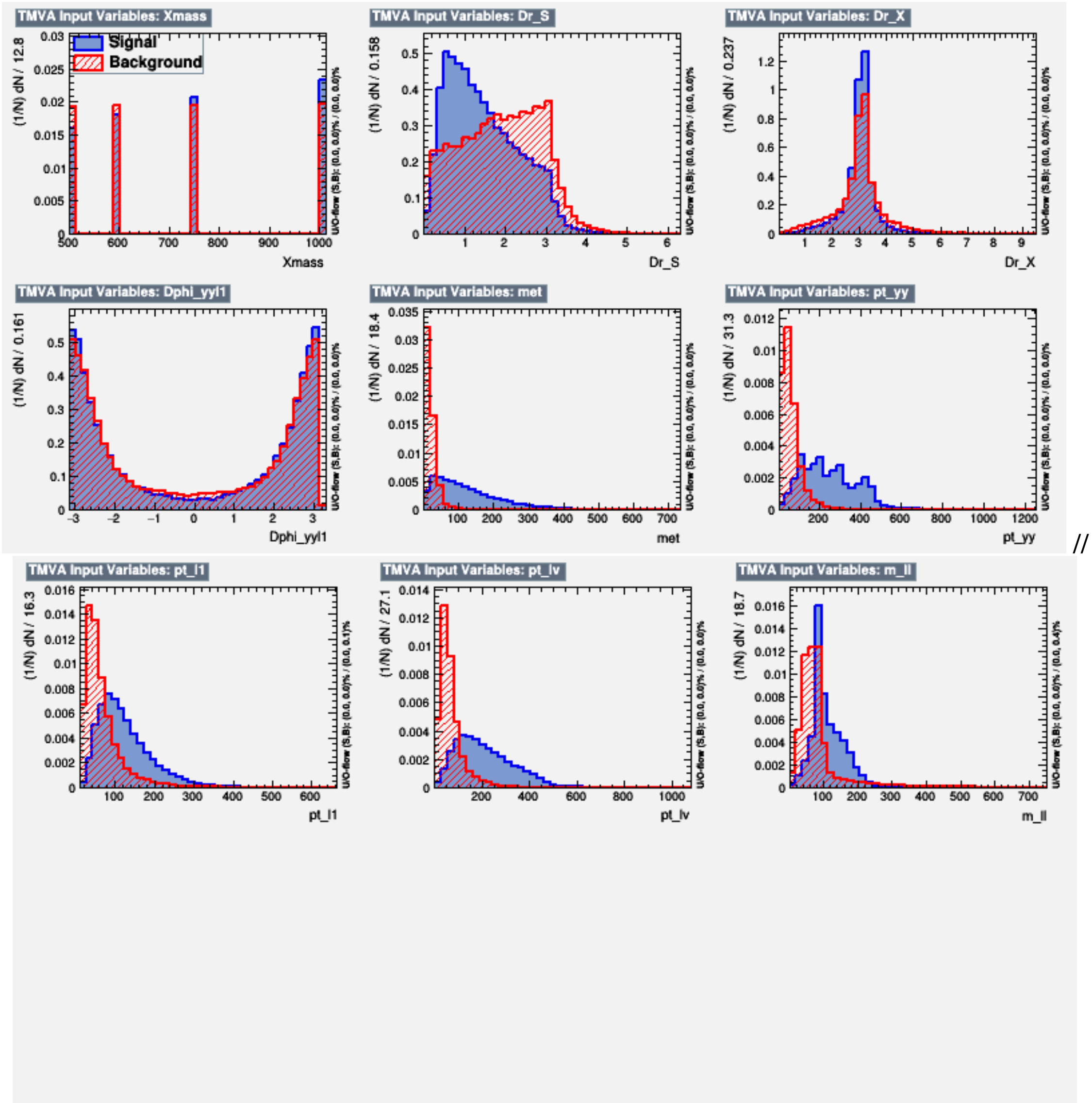
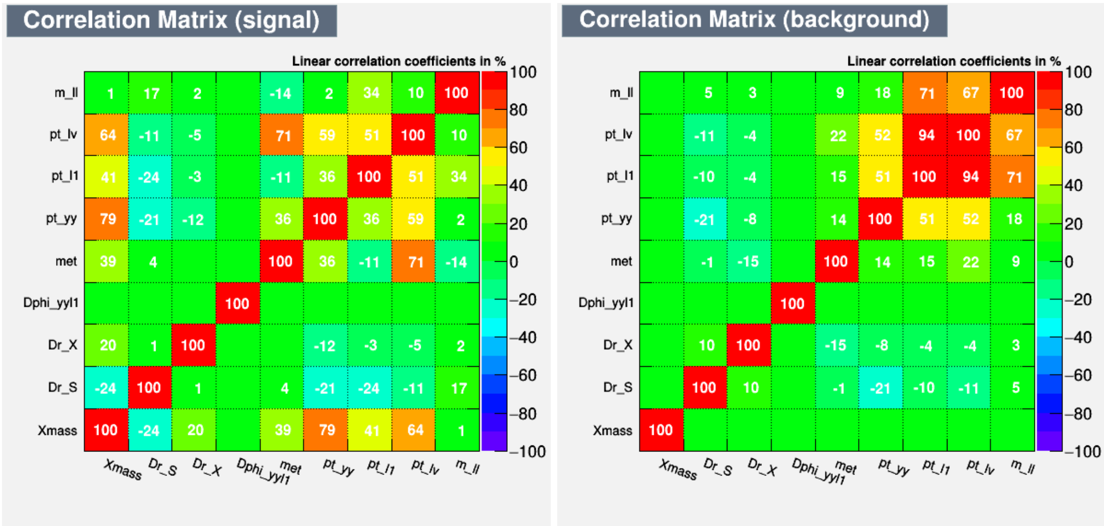
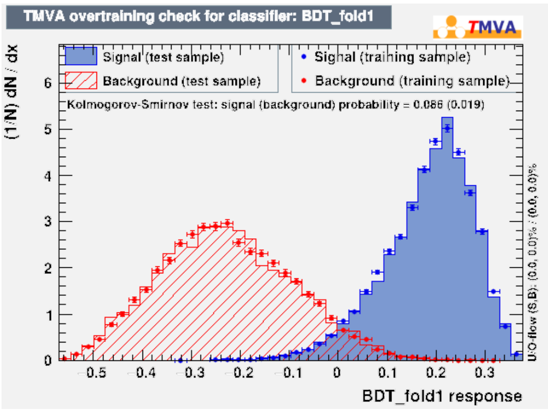


Figure 191: Input variables for BDT in WW1l category.

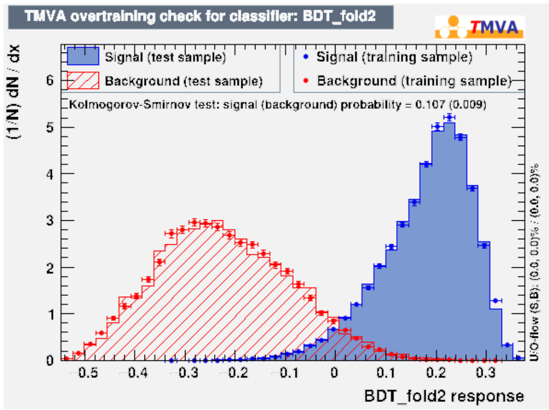


(a) Signal (b) Background

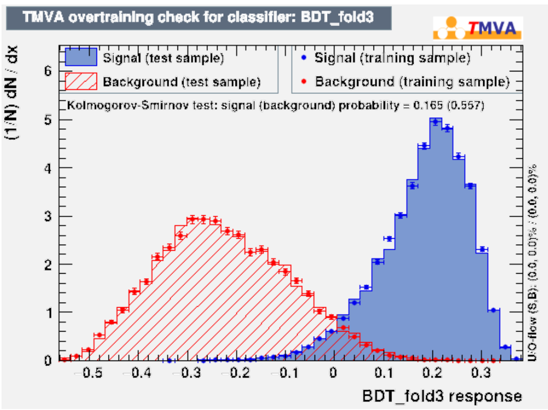
Figure 192: Input variables' correlation in WW1l category.



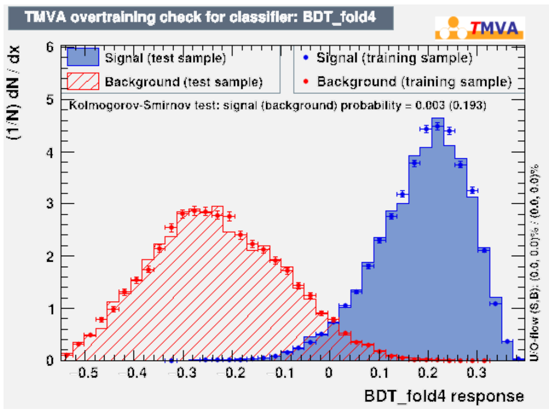
(a) fold1



(b) fold2

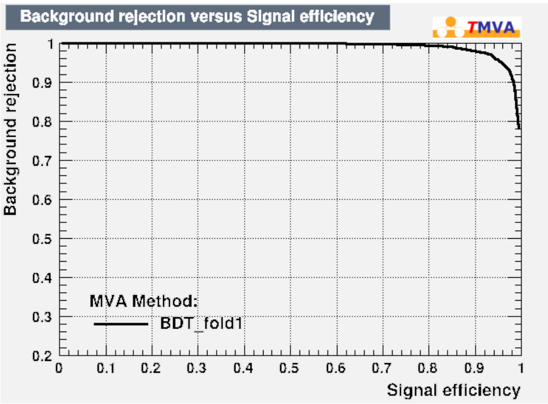


(c) fold3

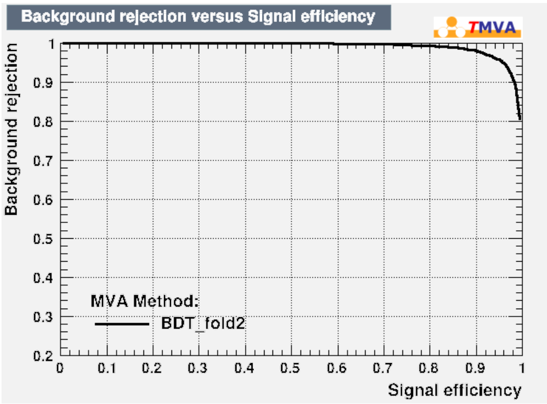


(d) fold4

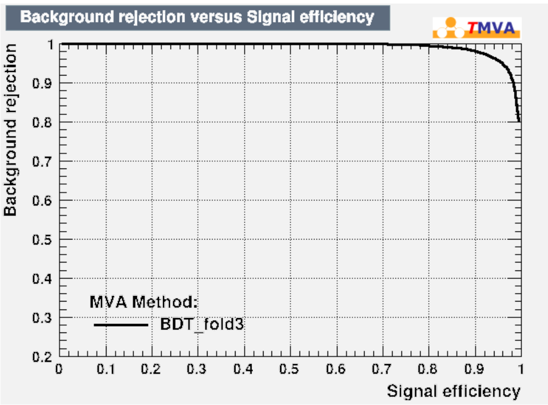
Figure 193: BDT distributions for training and test templates in 4 folds



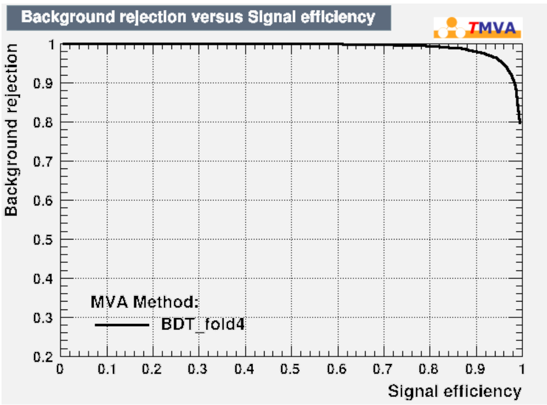
(a) fold1



(b) fold2



(c) fold3



(d) fold4

Figure 194: Trainned BDT ROC in 4 folds

F.2.4 $m_S \geq 400$ GeV

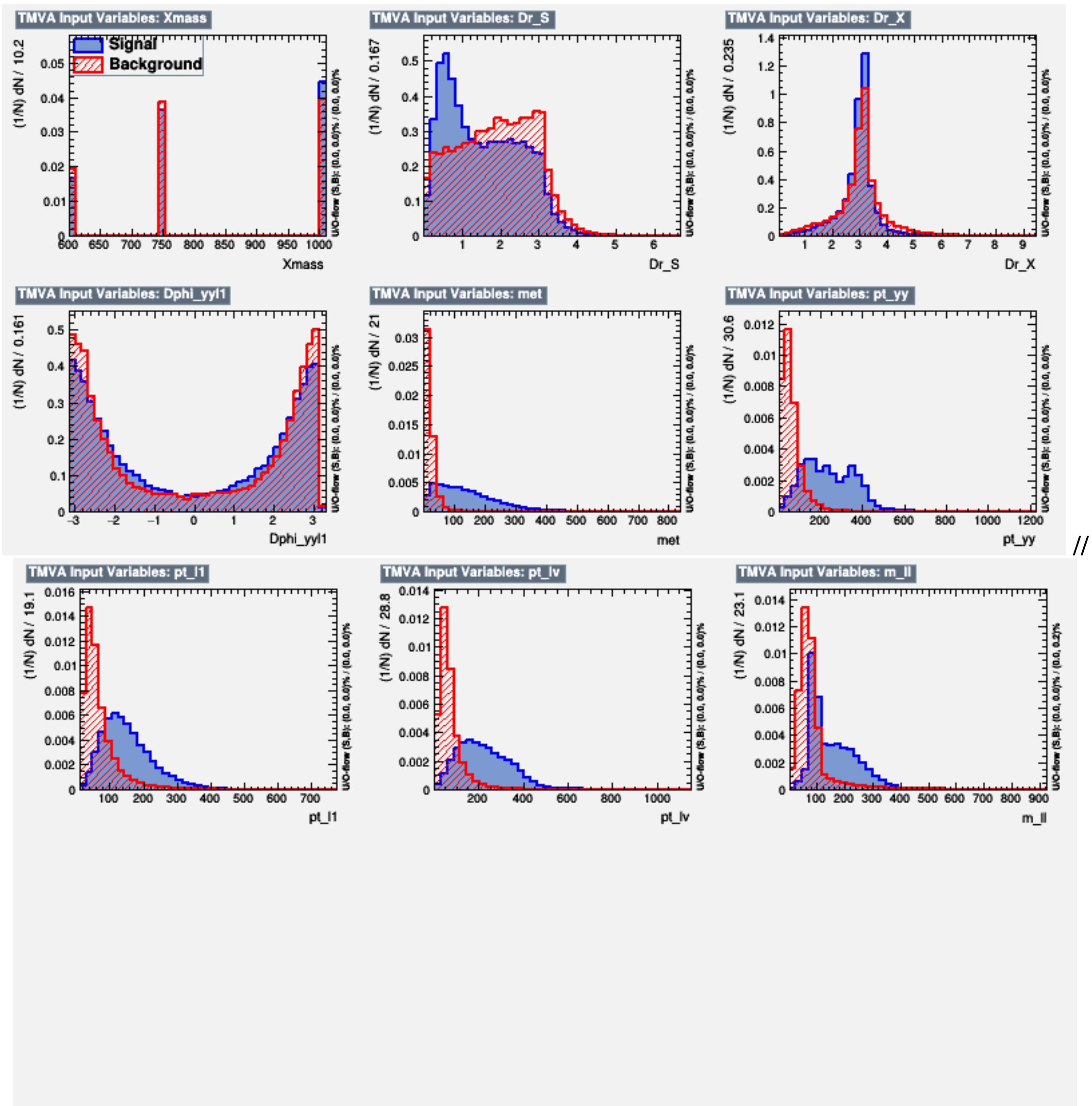
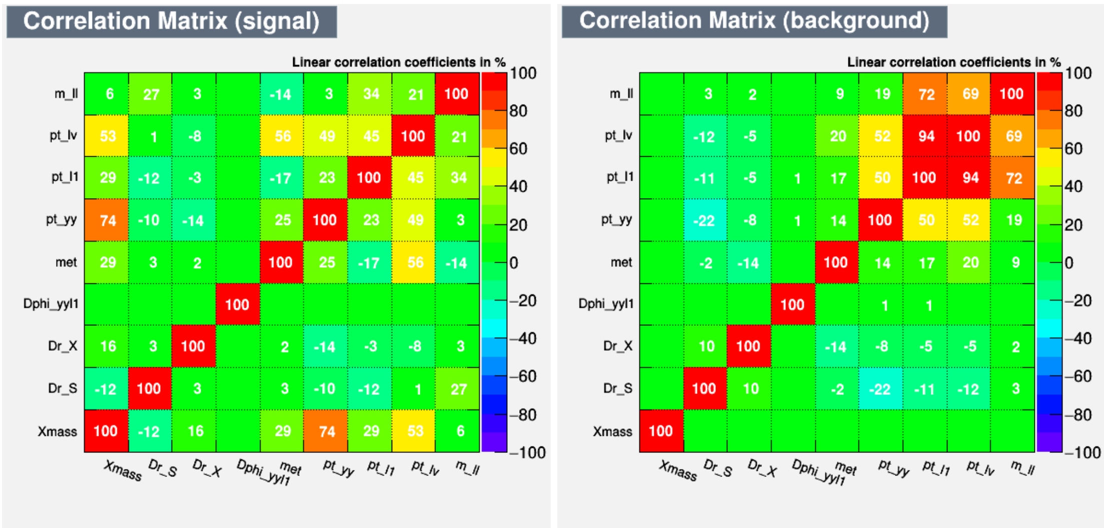
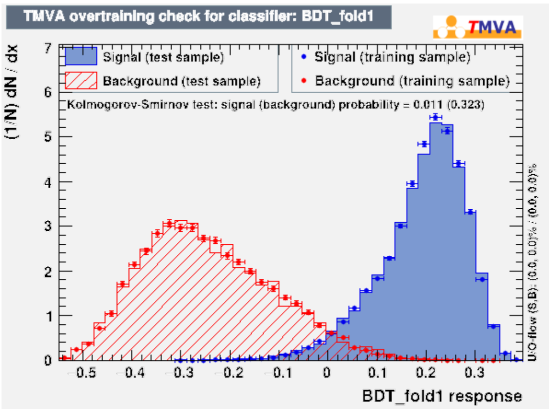


Figure 195: Input variables for BDT in WW1l category.

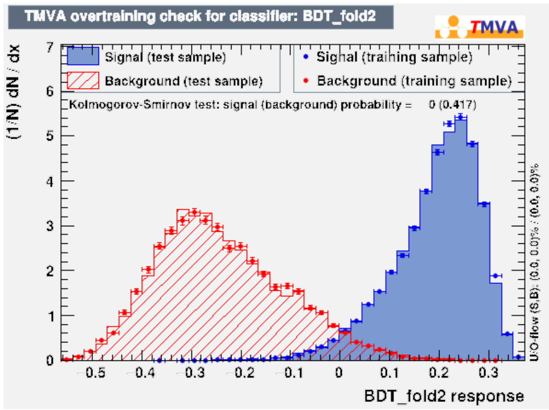


(a) Signal (b) Background

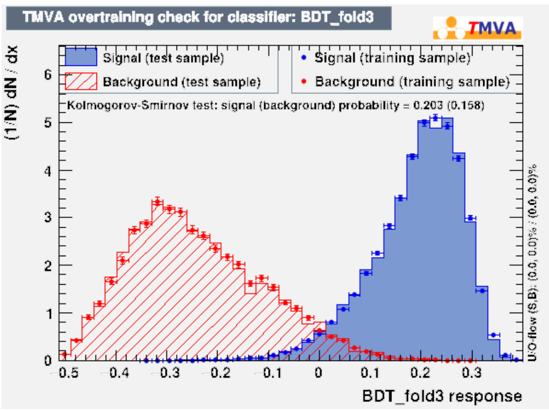
Figure 196: Input variables' correlation in WW1l category.



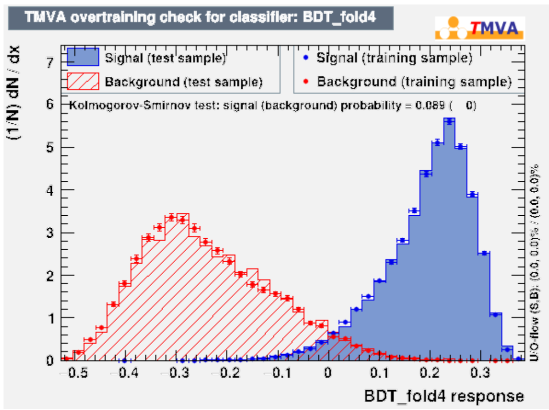
(a) fold1



(b) fold2

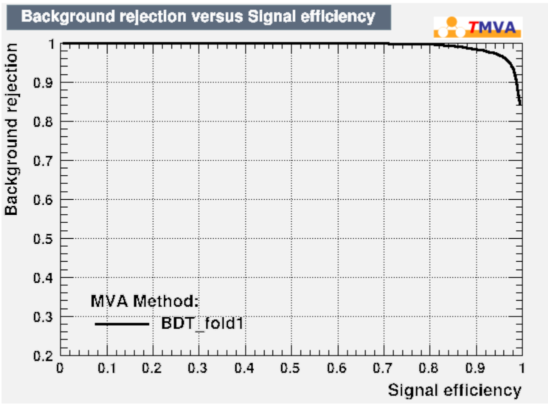


(c) fold3

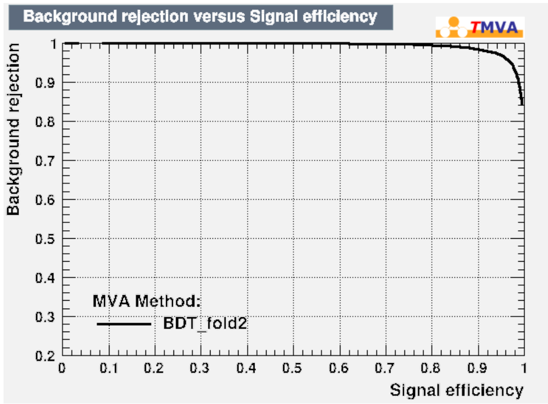


(d) fold4

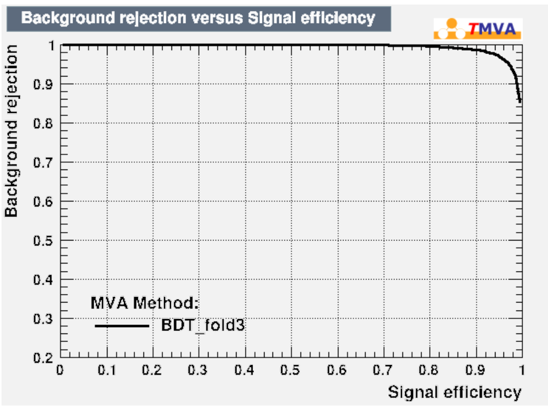
Figure 197: BDT distributions for training and test templates in 4 folds



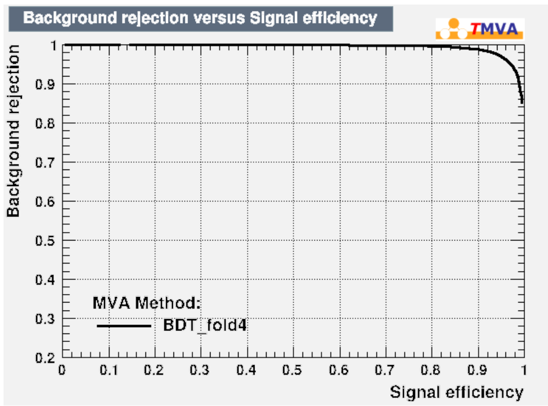
(a) fold1



(b) fold2



(c) fold3



(d) fold4

Figure 198: Trainned BDT ROC in 4 folds

G Diphoton vertex efficiency in SH signal samples

The efficiency for selecting the correct vertex is studied in $H \rightarrow \gamma\gamma$ analysis [29]. It varies between 60% to 100% depending on the diphoton p_{T}, the number of primary vertices in the event, the Σp_T and Σ_T^2 of the hard scattering vertex. And the selection efficiency using an artificial neural network is usually higher than that using the hardest vertex in SM Higgs processes, except for the $t\bar{t}h$ process. The same procedure is repeated in the SH process for validation. 3 mass points of (m_X, m_S) = (1000, 170), (1000, 500) and (600, 400) GeV are considered and merged together, covering a wide $p_T^{\gamma\gamma}$ region from 0 to 600 GeV. For each mass point sample, 20000 events in $\gamma\gamma + l\nu l\nu$ channel are used with a mixture of mc16 a, d, e. These events are only required to pass the default HGam selections, i.e. the first 4 criteria listed in Sec. 4. The other selections on leptons and jets are supposed to not influence the diphoton vertex. Figure 199 shows the distribution of $p_T^{\gamma\gamma}$, number of primary vertex and $log(\Sigma p_T^2)$ in the merged samples. In $p_T^{\gamma\gamma}$ plot, the 3 peaks are the signature of 3 mass points. Figure 200 shows the efficiencies of selecting the correct primary vertex as a function of the above variables Figure 201 shows the fraction of events that selected vertex by 2 approaches (hardest vertex and NN) are the same one.

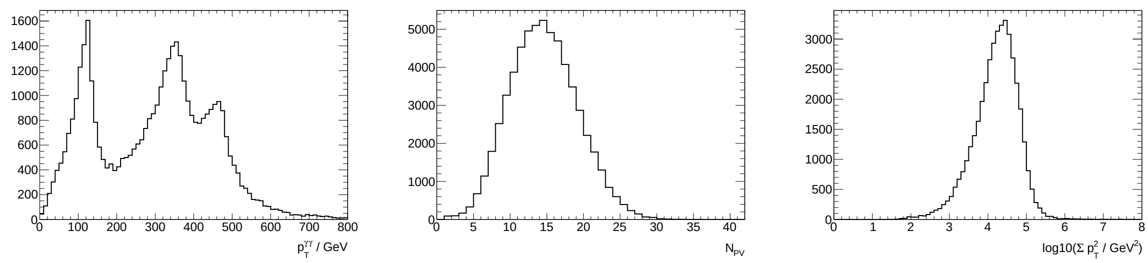
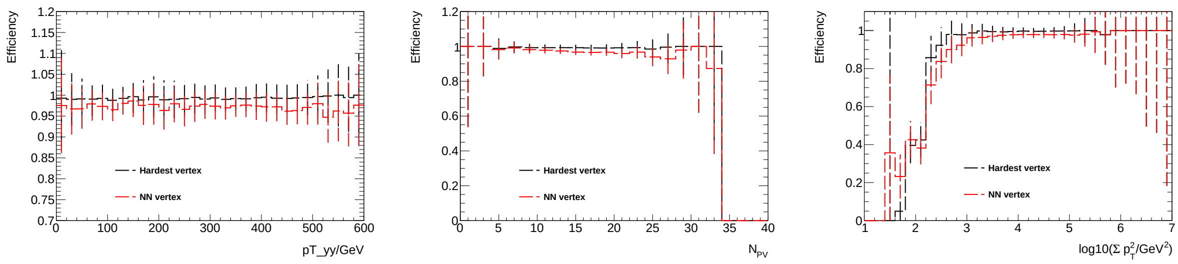
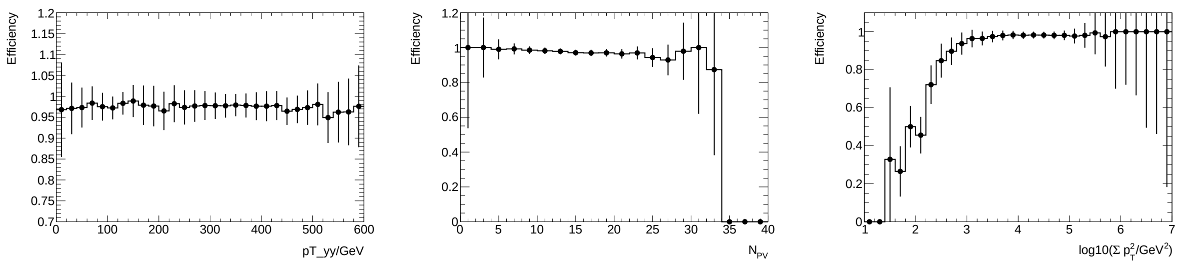


Figure 199: Distributions of $p_T^{\gamma\gamma}$, N_{PV} and $log(\Sigma p_T^2)$ in $SH \rightarrow \gamma\gamma + l\nu l\nu$ process, with mass points (m_X, m_S) = (1000, 170), (1000, 500) and (600, 400) GeV. Distributions of these variables in SM Higgs sample can be found in Ref [29].



(a) $p_T^{\gamma\gamma}$ (b) N_{PV} (c) $log(\Sigma p_T^2)$

Figure 200: Efficiency of selecting a correct vertex by two approaches as a function of variables.



(a) $p_T^{\gamma\gamma}$ (b) N_{PV} (c) $log(\Sigma p_T^2)$

Figure 201: Fraction of the event that selected vertex by 2 approaches are the same one as a function of variables.

1090 # H Toy Limits results

1091 In the previous selection, the limits are calculated by the asymptotic method. To validate the assumption
1092 that asymptotic should be performed with enough statistics, the limits with the toy are computed. For a
1093 first-step check, the result of mass point m_X = 1000 GeV, m_S = 300 GeV. The μ is scanned from 0 to 0.3,
1094 having 5000 toys SplusB and Bonly. The result documented in Table 32 shows the difference between
1095 asymptotic and toy is acceptable .The issue at $+2\sigma$ band is due to the wrong setup of the scanned range of μ.
1096 Results with 1000 toys where the range of μ is from 0 to 0.4: 0.310 ($+2\sigma$), 0.224 ($+1\sigma$), 0.152 (Median),
1097 0.112 (-1σ), 0.067 (-2σ) , giving the confidence that the issue can be fixed with a wider POI range.

	$+2\sigma$	$+1\sigma$	Median	-1σ	-2σ
Asymptotic	0.348	0.227	0.152	0.109	0.081
toy	10.00	0.215	0.153	0.116	0.097

Table 32: The expected limits of the search m_X = 1000 GeV, m_S = 300 GeV with asymptotic and toy.

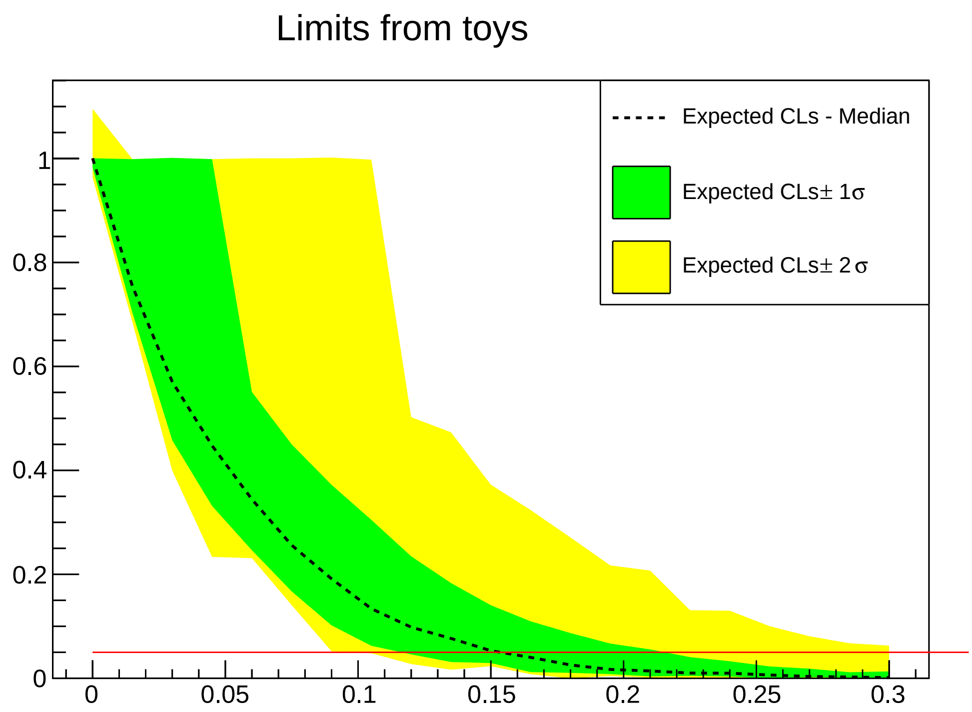


Figure 202: The expected limits of the search m_X = 1000 GeV, m_S = 300 GeV with toy.

I Signal Injection Test

To test the robustness of the model used in the binned likelihood fit, a signal injection test is applied. In the test, the signal plus background model is fitted to various signal plus background toy datasets in which the signal strengths are set to 1, 1.5 and 2 respectively. The fitted signal strength distributions with 5000 toys for these three cases are shown in Figure 203. It is obvious that the fitted signal strengths peaked at the nominal values following mostly the Gaussian distributions and the deviations are relatively small and stable. The σ of the signal strength is about 0.2, so it is in expected considering the left signal plus background events usually less than 10 in the BDT tight region.

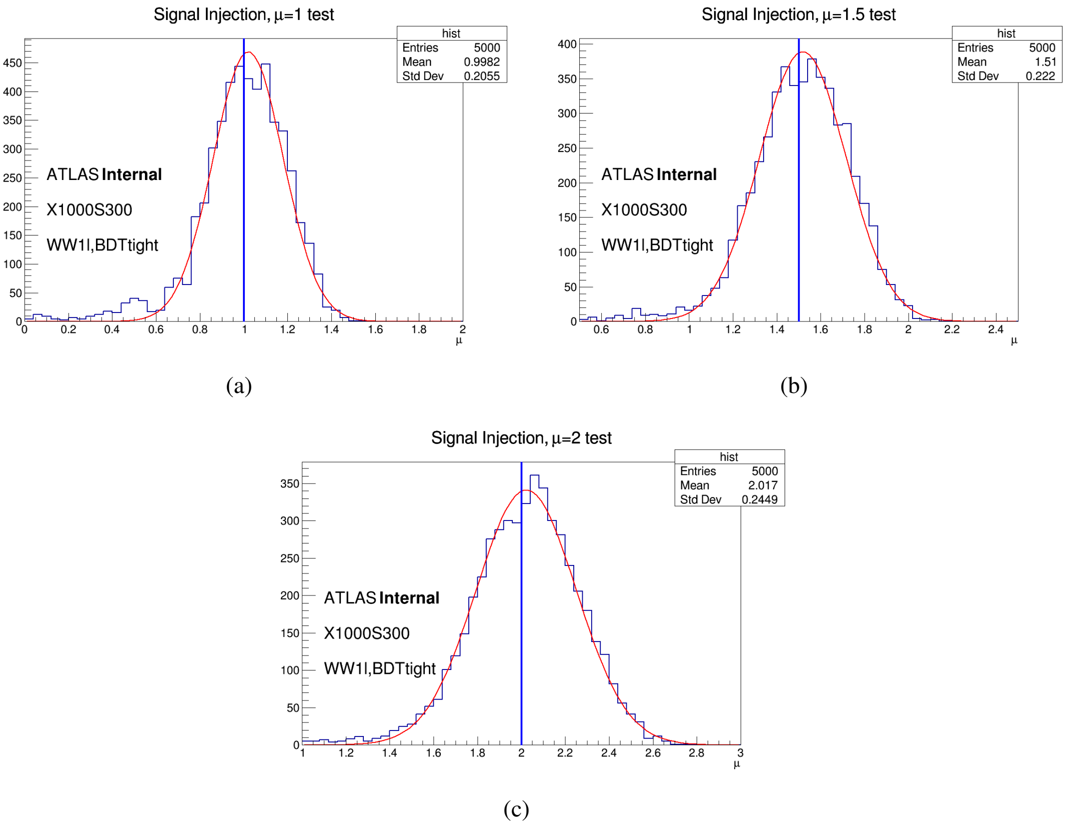


Figure 203: Signal strength distributions in the test of the signal injection with $\mu = 1, 1.5$ *and* 2. Tests are done with the mass point (mX,mS) =(1000,300) and 1 lepton BDT tight region is chosen.