

# Chapter 1

## Background estimation

Although the 4b final state is tantalizingly attractive because of the highest branching ratio fraction for the signal, the fully hadronic final state also means that we have larger backgrounds to cope with. QCD backgrounds are notoriously difficult to simulate from first principles, so we need can't rely on simulation to estimate our background estimate, and instead need to rely on data driven methods to form an estimate of the background in a (blinded) signal region.

There are two main proportions of background that we have: QCD and  $t\bar{t}$ , with the estimated relative proportions shown in Table 1.1(a) and 1.1(b) for ggF and VBF, respectively. Since the  $t\bar{t}$  proportion is only 10% of the background, our background estimate is derived inclusively for the QCD and  $t\bar{t}$  proportions.

	16	17	18	All		16	17	18	All
pre- $X_{Wt}$	19.5%	17.6%	18.2%	<b>18.3%</b>	pre- $X_{Wt}$	10.9%	11.1%	13.3%	<b>12.2%</b>
post- $X_{Wt}$	10.4%	9.5%	10.4%	<b>10.2%</b>	post- $X_{Wt}$	7.5%	6.6%	9.2%	<b>8.1%</b>

(a) ggF 4b

(b) VBF 4b

Table 1.1: Percentage of the data-driven background estimate expected to be composed of  $t\bar{t}$  events for the ggF 4b (left) and VBF 4b categories (right).

The key idea of the background estimate is to reweight the distributions in a lower  $b$ -tag region into a higher  $b$ -tag region by deriving these reweighting maps in dedicated control regions. The task of accurately estimating this background and assigning an appropriate error bar is the key to this analysis, and is the focus of this chapter. Section 1.1 overviews the nominal background estimate strategy, while Section 1.2 shows some of the validation plots that the method is working. the errors that we assign to the nominal estimate are shown in Section 1.3 and the tests of the methodology in different validation regions are discussed in Section 1.4

## 1.1 Reweighting overview

This technique derives a mapping to get an event-by-event weight from a lower tagged region to a higher tagged region (like a generalized ABCD method).

The lower  $b$ -tagged region consists of events with exactly two  $b$ -tags (which will subsequently be referred to as “2bb” events), and we reweight to match the distributions of events with four or more  $b$ -tags (“4bb” events). For “2bb” events, the leading two non- $b$  tagged jets are taken for the other HC jets, and these four jets are still paired into HCs with the  $\min\Delta R_{jj}^{HC1}$  pairing algorithm and pass through the rest of the analysis selection.

The reweighting maps are derived in CR 1 (as shown in [Need to cite figure](#)) since the events in the 4b SR are not observed until the analysis selection is finalized. The reweighting consists of deriving the maps  $w(x)$

$$p_{4b}(x) = w(x) \cdot p_{2b}(x), \quad (1.1)$$

where  $x$  is a set of features characterizing the kinematics of the event, specified in Table 1.2.

Variable description	ggF	VBF
$\log(\Delta R_1)$ : between the closest two HC jets	✓	
$\log(\Delta R_2)$ between the other two HC jets	✓	
$\log(p_T)$ of the 4th leading HC jet	✓	
$\log(p_T)$ of the 2nd leading HC jet	✓	
$\langle  HC\eta  \rangle$ : average absolute value of the HC jets $\eta$	✓	✓
$\log(p_{T,HH})$	✓	
$\Delta R_{HH}$	✓	
$\Delta\phi$ between the jets in the leading HC	✓	
$\Delta\phi$ between the jets in the subleading HC	✓	
$\log(X_{Wt})$	✓	✓
Number of jets in the event	✓	
Trigger bucket index	✓	✓
Year index		✓
Second smallest $\Delta R$ between the jets in the leading HC (out of the three possible pairings)		✓
Maximum di-jet mass out of the possible pairings of HC jets		✓
Minimum di-jet mass out of the possible pairings of HC jets		✓
Energy of the leading HC		✓
Energy of the subleading HC		✓

Table 1.2: Set of input variables used for the  $2b$  to  $4b$  reweighting for the ggF and VBF channels. The variables included in the background estimate are denoted with a checkmark.

To enforce positivity of the weights, we ac NN  $Q(x) = \log w(x)$ , minimizing the loss function:

$$\mathcal{L}(Q(x)) = \mathbb{E}_{x \sim p_{2b}} \left[ \exp \left( \frac{1}{2} Q(x) \right) \right] + \mathbb{E}_{x \sim p_{4b}} \left[ \exp \left( -\frac{1}{2} Q(x) \right) \right] \quad (1.2)$$

results in a model learning  $Q^*(x) = \log w(x)$  (where the loss function is set up in this way to are positive numbers using the loss function);

These mappings are derived in a kinematically similar control region (shown as CR 1 in Figure ??). Then an error on this method from this choice of training region is taken by using an alternative control region (CR 2) and using the difference of the CR1 and CR2 predictions as an error bar.

Applying these reweighting maps in the SR gives us a background estimate of the observed data in a blinded SR.

In practice, there is some amount of noisiness due to the initializations of the NNs. To this extent, the NN for each background estimate is retrained 100 times and the average of the weights is taken as the nominal weight. Additionally, for each of the NN trainings the CR1 training events have a weight sampled from a Poisson distribution with mean 1 to account for the finite statistics of the training dataset.

Since we use different triggers between each of the years, the ggF background estimate has a different background estimate derived for each of the three years. Since VBF has two orders of magnitude less statistics than ggF, it was not as sensitive to this effect, so instead for VBF the training was done inclusively for the years with the year passed as an extra variable.

For ggF, the background estimates were derived before the  $X_{Wt}$  cut, while for the VBF the trainings were derived after the  $X_{Wt}$  cut.<sup>1</sup>

## 1.2 Validation plots

What might be good to include here?

- Let's include a *few* key variables for both ggF and VBF (?)
- And also the unrolled plots!!

### 1.2.1 Marginal distributions

VBF plots ...

### 1.2.2 Validation plots in CR1 in categories

Background modeling in the CR1 and CR2 regions are checked for the ggF and VBF channels. Figure 1.3 shows the CR1 distributions for the three channels, taking 2018 for ggF as representatives.

---

<sup>1</sup>We saw for ggF that we had the same performance whether we trained before or after the  $X_{Wt}$  cut.

The ggF plots for the other years can be found in Appendix ??, as well as the same set of plots in CR2.

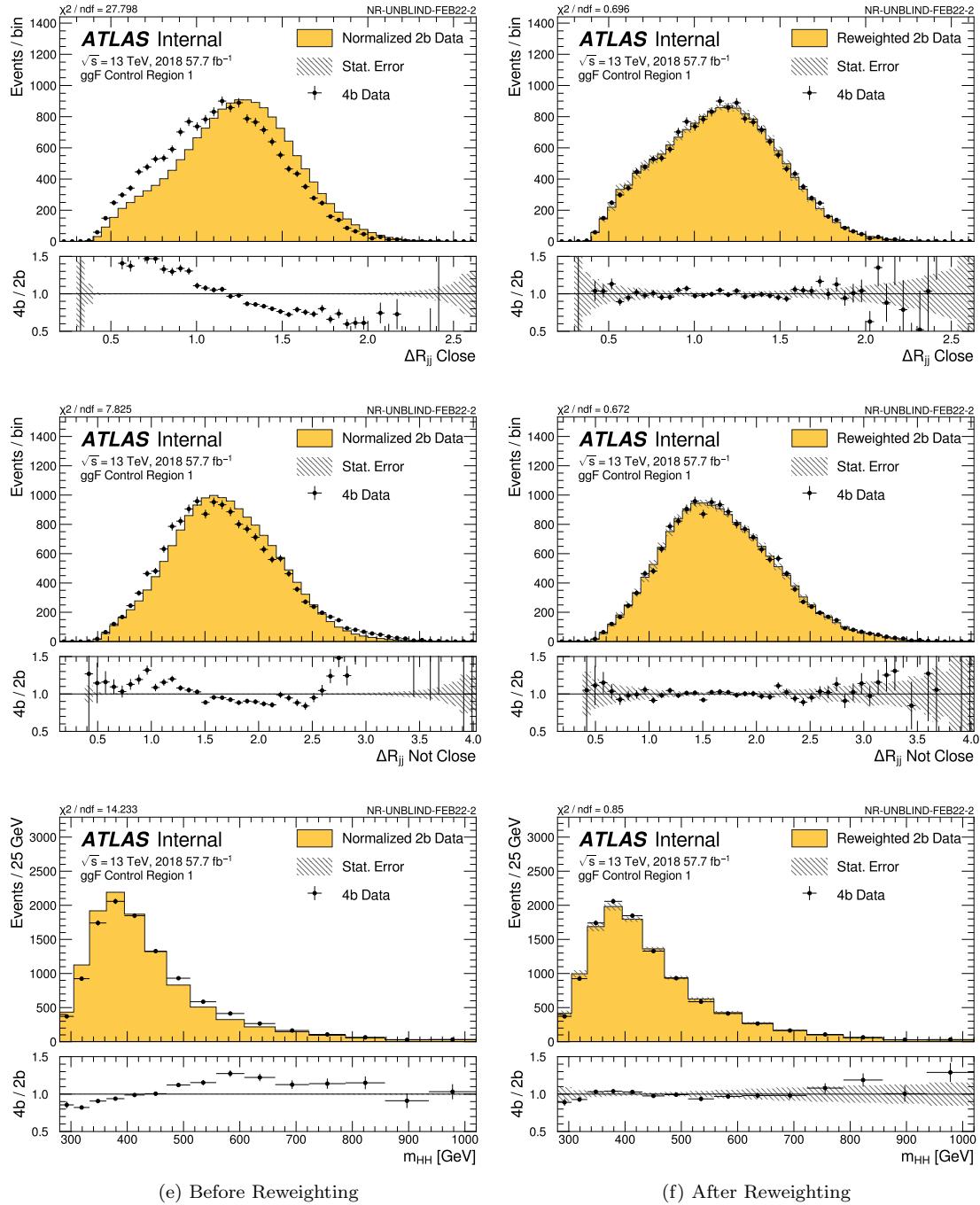


Figure 1.1: Distributions of  $\Delta R$  between the closest Higgs Candidate jets,  $\Delta R$  between the other two (training variables) and the mass of the di-Higgs system (non-training variable) before and after CR 1 derived reweighting for the 2018 Control Region 1.

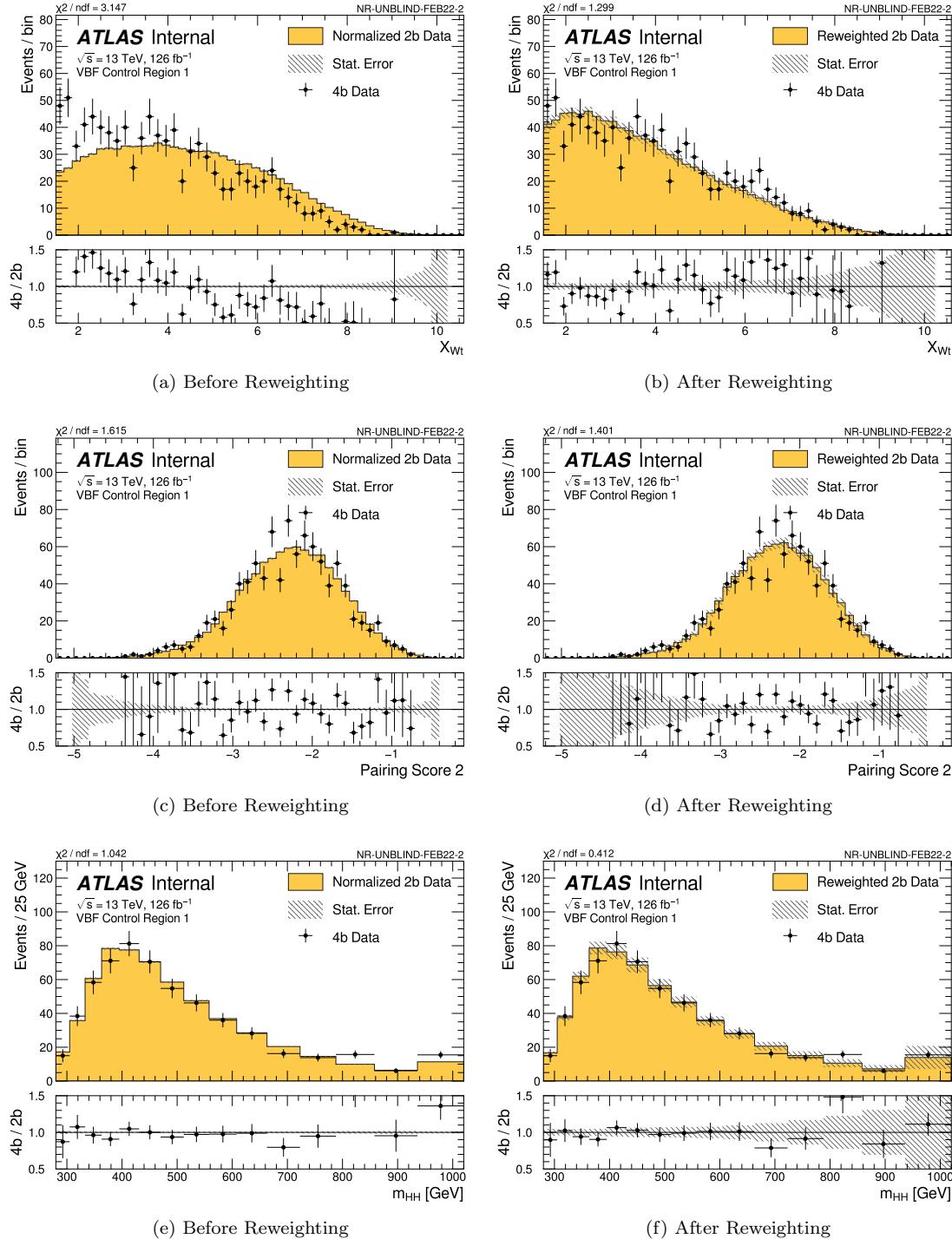


Figure 1.2: Distributions of the top veto variable,  $X_{Wt}$ , the second smallest  $\Delta R$  between the jets in the leading candidate (training variables) and the mass of the leading and subleading Higgs candidates and of the di-Higgs system (non-training variable) before and after CR 1 derived reweighting for the all years inclusive Control Region 1.

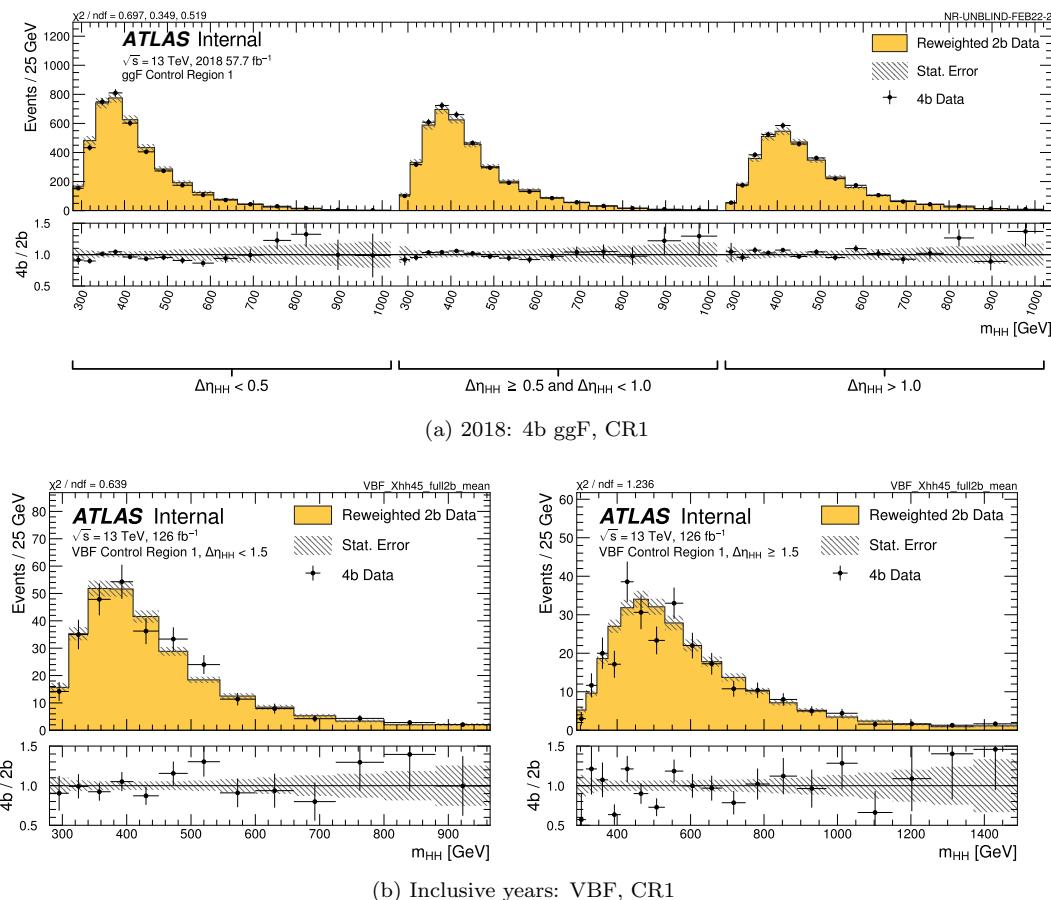


Figure 1.3: Distributions in CR1 after reweighting and categorization for ggF 2018 and VBF inclusive years. Bootstrap and Poisson errors are included.

## 1.3 Background systematics

For our fully data-driven background estimate, we assess several custom background modeling uncertainties:

- NN optimization Section 1.3.1
- choice of CRs for deriving these reweightings Section 1.3.2,
- and (ggF-only) remaining extrapolation uncertainty from testing a reweighting to another (lower)  $b$ -tagged target distribution Section 1.3.3.

We present the tables and plots comparing the relative size of each of these contributions (Section 1.3.3), and conclude with a description of how we characterize these uncertainties in the profile likelihood fit in Section subsec:bkgUnc-stats.

### 1.3.1 Deep ensembles

There are two components to the statistical error for the neural network background estimate. The first is standard Poisson error of the 2b data in the SR, i.e., a given bin,  $i$ , in the background histogram has value  $n_i = \sum_{j \in i} w_j$ , where  $w_j$  is the weight for an event  $j$  which falls in bin  $i$ . Standard techniques then result in statistical error  $\delta n_i = \sqrt{\sum_{j \in i} w_j^2}$ , which reduces to the familiar  $\sqrt{N}$  Poisson error when all  $w_j$  are equal to 1.

However, this procedure does not take into account the statistical uncertainty on the  $w_j$  due to the finite training dataset. Due to the large size difference between the 2b and 4b datasets, it is the statistical uncertainty due to the 4b training data that dominates that on the background. A standard method for estimating this uncertainty is the bootstrap resampling technique [**Bootstrap**]. Conceptually, a set of statistically equivalent sets is constructed by sampling with replacement from the original training set. The reweighting network is then trained on each of these separately, resulting in a set of statistically equivalent background estimates. Each of these sets is below referred to as a replica.

In practice, as the original training set is large, the resampling procedure is able to be simplified through the relation  $\lim_{n \rightarrow \infty} \text{Binomial}(n, 1/n) = \text{Poisson}(1)$ , which dictates that sampling with replacement is approximately equivalent to applying a randomly distributed integer weight to each event, drawn from a Poisson distribution with a mean of 1.

Though the network configuration itself is the same for each bootstrap training, the network initialization is allowed to vary, and this uncertainty. It should therefore be noted that the bootstrap uncertainties implicitly capture the uncertainty due to this variation in addition to the previously mentioned training set variation.

The procedure to calculate the bootstrap uncertainty is as follows: first, each network trained on each bootstrap replica dataset is used to produce a histogram in the variable of interest. This

results in a set of replica histograms (e.g. for 100 bootstrap replicas, 100 histograms are created). The nominal estimate is mean of bin values across these replica histograms, with errors set by the corresponding standard deviation.

The variation from this bootstrapping procedure is used to assign a bin-by-bin uncertainty which is treated as a statistical uncertainty in the fit.

Figure 1.4 demonstrates how an uncertainty envelope is calculated from this procedure by visualizing the  $m_{HH}$  histogram in the Control Region 1 (where the background estimates are derived) for the 2018 ggF (left) and years inclusive VBF (right) background estimates.

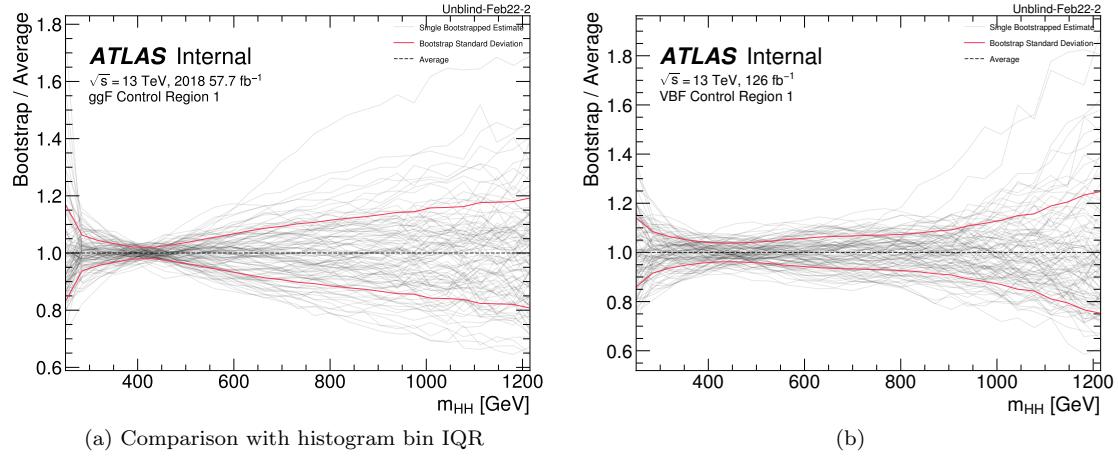


Figure 1.4: Illustration of the bootstrap band procedure, shown as a ratio to the nominal estimate. Each grey line is from the  $m_{HH}$  prediction for a single bootstrap training, and the solid red line shows the standard deviation of histograms for the 2017 ggF background estimate (left) and VBF years inclusive background estimate (right).

### 1.3.2 Choice of control region

To account for the systematic bias associated with deriving the reweighting function in CR1, an alternative background model is derived in CR2. To aid in deriving an uncertainty to cover this bias, the Signal Region is split into quadrants - four sectors of approximately equal area - where the angle of alignment is set to match the quadrants that define the control regions, CR1 and CR2. The four SR quadrants are named for their relative positions in the Higgs Candidate mass plane, where Q is short for quadrant -  $Q_N$  (north),  $Q_S$  (south),  $Q_E$  (east) and  $Q_W$  (west), and shown in Figure 1.5.

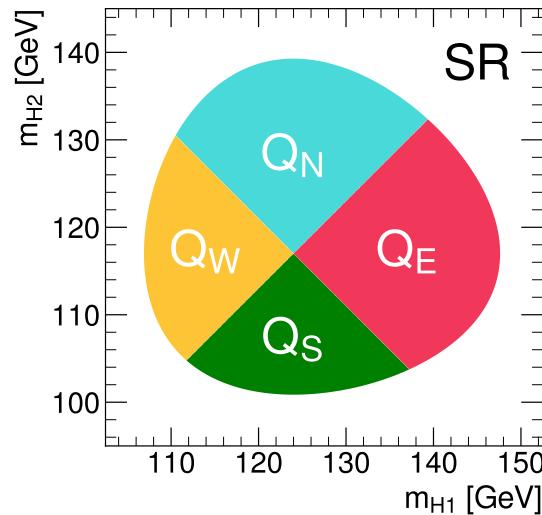


Figure 1.5: SR quadrants chosen to derive the four background variation nuisance parameters.

The nominal background estimate is derived by applying weights derived in CR1 to all four SR quadrants. Four alternative background estimates are derived by applying weights derived in CR1 to three of the SR quadrants and weights derived in the CR2 to the one remaining SR quadrant. For example, one alternative background estimate is derived by applying CR1-derived weights to  $Q_S$ ,  $Q_E$  and  $Q_W$ , and CR2-derived weights to  $Q_N$ . The procedure is to reduce the constraints of a single nuisance parameter (NP) if we were to take the whole CR2 weights as a single uncertainty. It constraint was seen in previous  $36 \text{ fb}^{-1}$  paper and the resonant paper, where certain NP decomposition was also applied. As we believe the QCD background contains multiple sources, this choice of splitting NPs is to introduce more degrees of freedom to the fit. This can be done in various ways. The current way of splitting to four quadrants aligns well with the CR1 and CR2 region definitions, therefore is deemed to be a natural choice.

The difference between the baseline and alternative estimates is used to define a shape uncertainty on the  $m_{HH}$  spectrum, which is made two-sided by symmetrizing the difference around the baseline. Note that both normalization and shape differences are taken into considered.

These SR quadrants align with the four sectors of CR1 and CR2 to give these shape systematics the

flexibility to naturally follow the kinematic similarity of the adjacent CRs. The symmetrized difference between the nominal and alternative background estimates are shown for the ggF discriminant for each year in Figures 1.6–1.8. Complementary comparisons of the nominal and alternative background estimates for the VBF selection are included in Appendix ??.

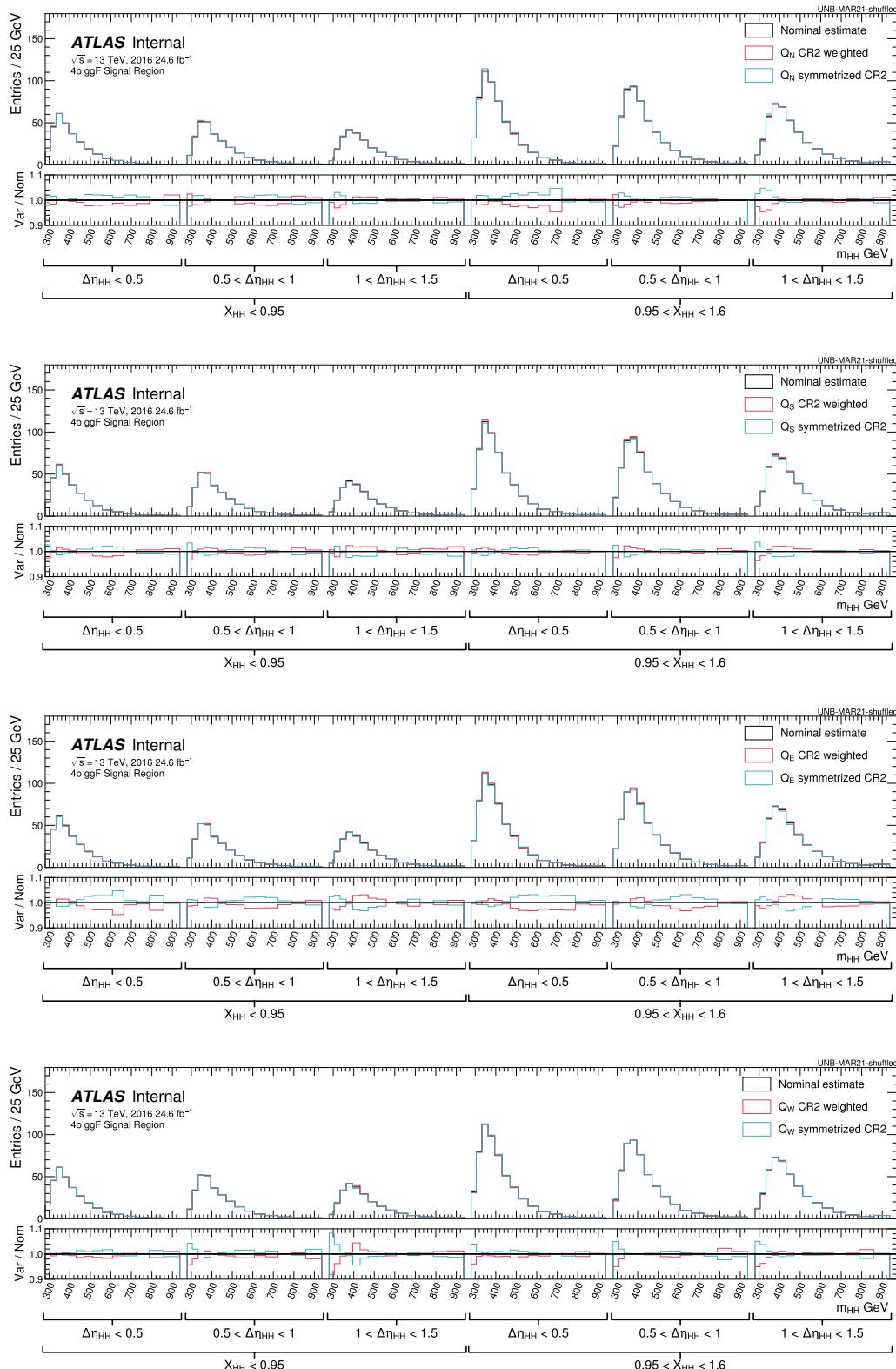


Figure 1.6: Example of variation in the SR NP quadrants for the 2016 ggF discriminant.

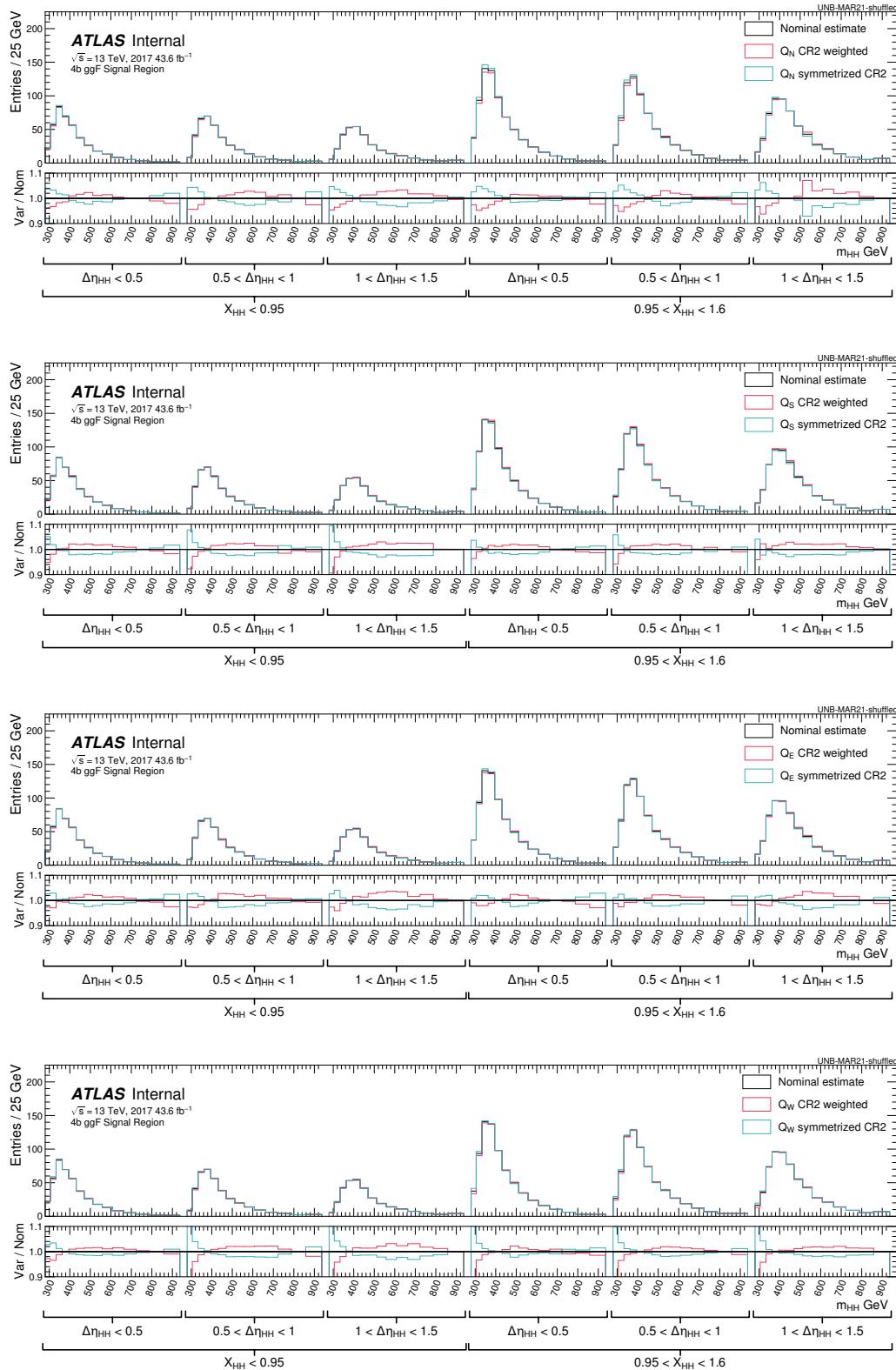


Figure 1.7: Example of variation in the SR NP quadrants for the 2017 ggF discriminant.

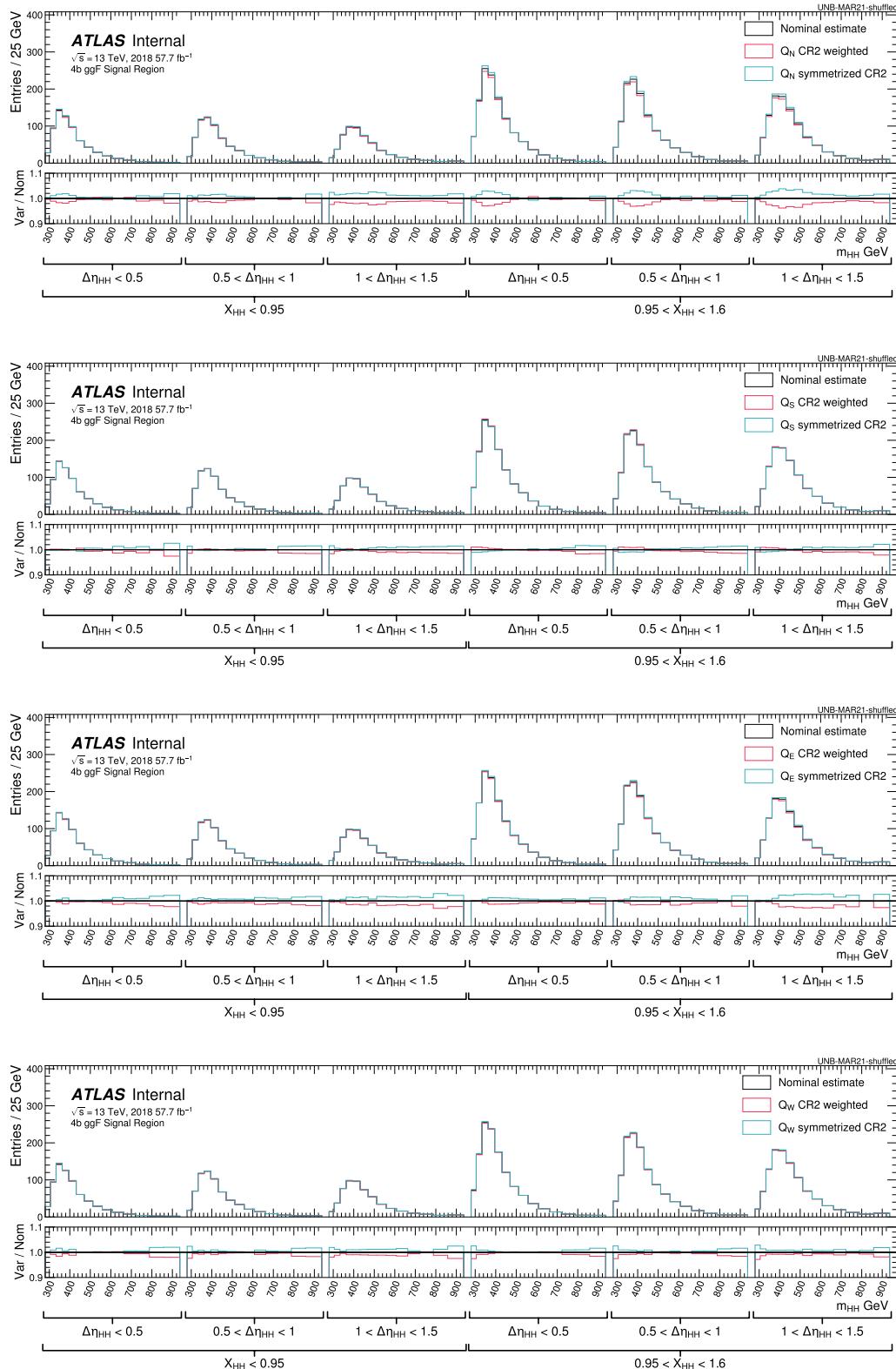


Figure 1.8: Example of variation in the SR NP quadrants for the 2018 ggF discriminant.

### 1.3.3 3b1f non-closure uncertainty

The 3b1f region is defined such that it contains events with three jets that were b-tagged at the 77% working point, but that all other jets failed to be b-tagged at even the loosest 85% working point. The fourth jet required to reconstruct the di-Higgs is taken as the highest  $p_T$  non-b-tagged jet. The standard methodology for deriving the background estimation (Section ??) is performed on this region, with 4b events replaced by 3b1f. The background prediction uncertainties include 2b stat error, bootstrap (Section ??), and CR1 versus CR2 shape uncertainties (Section ??). Detailed studies can be found in Appendix ???. A small deviation between the reweighted 2b and the 3b1f data distributions in the SR was found in the ggF channel but not in the VBF channel. An additional uncertainty due to the 3b1f non-closure is derived for the ggF channel to cover this deviation.

The uncertainty is derived as follows: In each of the analysis categories (Section ??), the ratio of the 3b1f and the reweighted 2b  $m_{HH}$  distributions is computed using the binning described in Table ???. The ratio is compared to unity bin by bin. If the difference is not covered by the quadratic sum of the background prediction uncertainties and the 3b1f data statistical uncertainty in that bin, the residual is taken to form a template. To reduce the statistical fluctuations between bins, the template is smoothed by averaging each bin with its two neighboring bins. The smoothed template is used as a non-closure uncertainty in 4b.

### Summary of Background Modeling Uncertainties

Relative magnitudes of the uncertainties for each year are shown in Table ?? for ggF and Table 1.3 for VBF along with an estimate of the impact of the statistics of 4b data in the signal region on the total error.

Table 1.3: Magnitude of error components for the VBF analysis in the Signal Region. Total statistic error is quadrature of bootstrap and 2b poisson errors. All errors are in the normalisation before categorisation

<b>SR 4b</b> (Bkgd Yield = 354.6) (%)	<b>Counts</b>	<b>Relative</b> <b>Relative(%)</b>
Total Statistic Error	32.3	9.1
Shape $Q_N$	-1.2	-0.35
Shape $Q_W$	-0.8	-0.23
Shape $Q_S$	-0.5	-0.15
Shape $Q_E$	-0.3	-0.095
Total Systematic Error	1.6	0.45
<b>Total Error</b>	<b>32.3</b>	<b>9.1</b>

A visualization of the magnitudes of these templates for the ggF channels is shown in Figure 1.9 and VBF in Figure reffig:bkgRelErr-VBF-4b. The statistical error dominates for high  $m_{HH}$  while the shape systematic from the difference in the CR1 and CR2 contributes more in the moderate  $m_{HH}$  region that drives our analysis sensitivity. Although the stat error (from the limited 2b data statistics) is negligible relative to the bootstrap error in the bulk of the distribution, it becomes relevant in the high  $m_{HH}$  tail. The final statistical uncertainty used for the limit setting is therefore the sum (in quadrature) of these two components.

#### 1.3.4 Choice of background systematic parametrization

OK - I made these plots... I just need to *pop* them in from somewhere else!

The number of categories, fits and all possible background nuisance parameters are described in table 1.4, informed by the fitting strategy described in the previous sub-section. For both VBF and ggF channels, the shape and normalization uncertainties for each category are obtained separately by comparing the CR1 and CR2 models after each category selection.

These nuisance parameters are not necessarily completely uncorrelated, and a choice on the correlation scheme must be made to perform the fit. The main argument for correlating the background shape nuisance parameters across kinematic regions is due to the neural network reweighting. The reweighting is trained inclusively with respect to kinematic categories, per year (for the ggF channel) or for all years together (for VBF). Possible mismodelings occurring due to the neural network's

performance, or, more importantly, its ability to interpolate into the signal region, will therefore have the same source. For the VBF channel, one step ahead is taken, and the fit is performed inclusively for all the years for this reason. This approach, however, was observed to be detrimental to the ggF background modeling performance, due to the kinematic differences between the years, and is not done.

Another important indication that these nuisance parameters should be correlated across kinematic regions is the relative impact they have in the  $m_{HH}$  fitting variable. This can be seen on Figure 1.11, which shows the nuisance parameter relative variations within each quadrant and each year of the ggF analysis. Most variations have very similar impacts on  $m_{HH}$ , which corroborates the hypothesis of correlated variations.

Table 1.4: Magnitude of error components for the ggF analysis in the Signal Region. Total statistic error is quadrature of bootstrap and 2b poisson errors. All errors are in the normalisation before categorisation

<b>SR 16 4b</b> (Bkgd Yield = 3211.1)		<b>Counts</b>	<b>Relative</b>	<b>Relative(%)</b>
	(%)			
Total Statistic Error		225.8		7.0
Shape Q <sub>N</sub>	−25.3		−0.79	
Shape Q <sub>W</sub>	−12.1		−0.38	
Shape Q <sub>S</sub>	16.2		0.50	
Shape Q <sub>E</sub>	7.0		0.22	
3b1f NC	6.8		0.21	
Total Systematic Error		33.9		1.1
<b>Total Error</b>		228.3		7.1
<b>SR 17 4b</b> (Bkgd Yield = 4492.7)		<b>Counts</b>	<b>Relative</b>	<b>Relative(%)</b>
	(%)			
Total Statistic Error		308.9		6.9
Shape Q <sub>N</sub>	−24.7		−0.55	
Shape Q <sub>W</sub>	−4.8		−0.11	
Shape Q <sub>S</sub>	43.0		0.96	
Shape Q <sub>E</sub>	14.8		0.33	
3b1f NC	17.6		0.39	
Total Systematic Error		54.9		1.2
<b>Total Error</b>		313.7		7.0
<b>SR 18 4b</b> (Bkgd Yield = 7720.0)		<b>Counts</b>	<b>Relative</b>	<b>Relative(%)</b>
	(%)			
Total Statistic Error		483.0		6.3
Shape Q <sub>N</sub>	−141.3		−1.8	
Shape Q <sub>W</sub>	−55.8		−0.72	
Shape Q <sub>S</sub>	7.0		0.091	
Shape Q <sub>E</sub>	−87.1		−1.1	
3b1f NC	42.6		0.55	
Total Systematic Error		180.4		2.3
<b>Total Error</b>		515.6		6.7

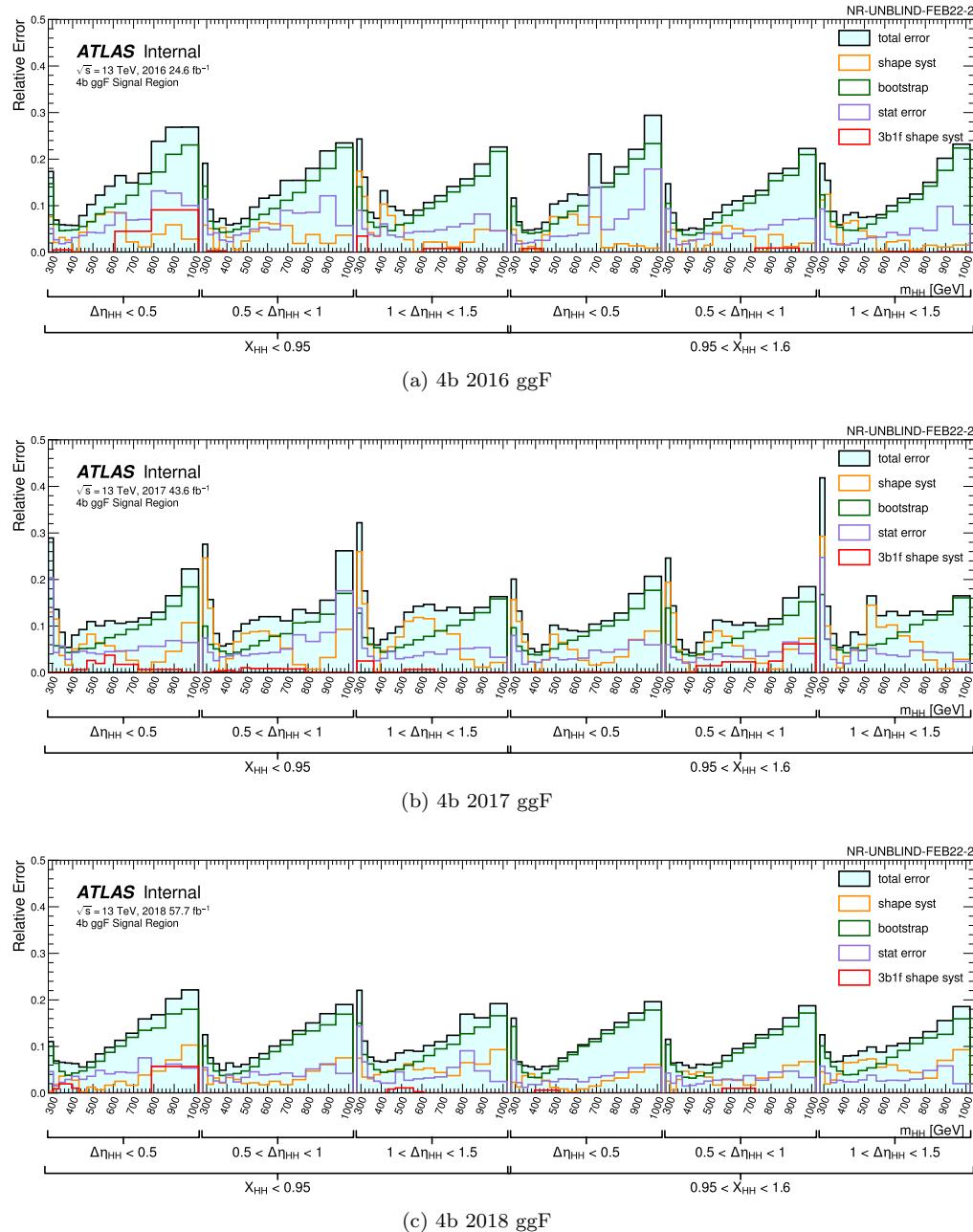


Figure 1.9: Relative error contributions of the background for the 4b ggF discriminant.

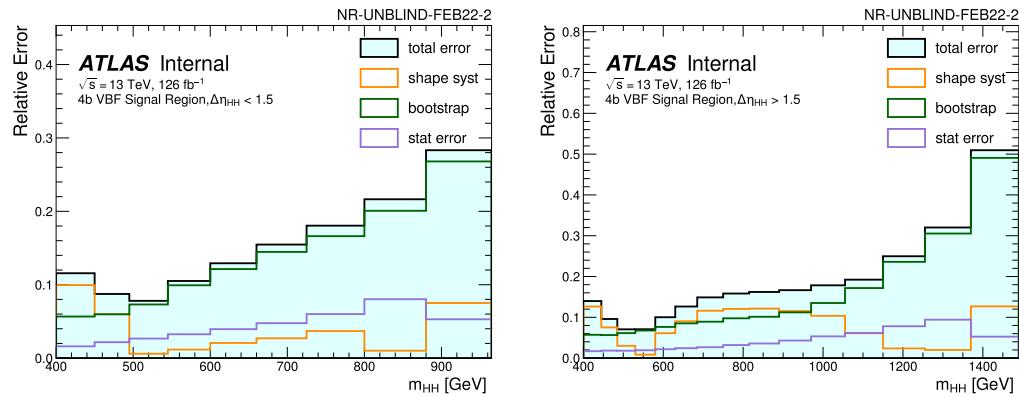


Figure 1.10: Relative error contributions of the background for the 4b VBF discriminant.

	ggF channel	VBF channel
Categories	3 years $\times 3  \Delta\eta_{HH} $ $\times 2 X_{HH}$ $= 18 \text{ categories}$	$2  \Delta\eta_{HH}  \text{ categories}$
Background shape nuisance parameters	3 years $\times (Q_N, Q_E, Q_S, Q_W)$ $= 12$	$(Q_N, Q_E, Q_S, Q_W)$ $= 4$
3b1f non-closure nuisance parameters	18 (1 per category)	0

Table 1.5: Summary of categorization strategy and background-related nuisance parameters in ggF and VBF analyses.

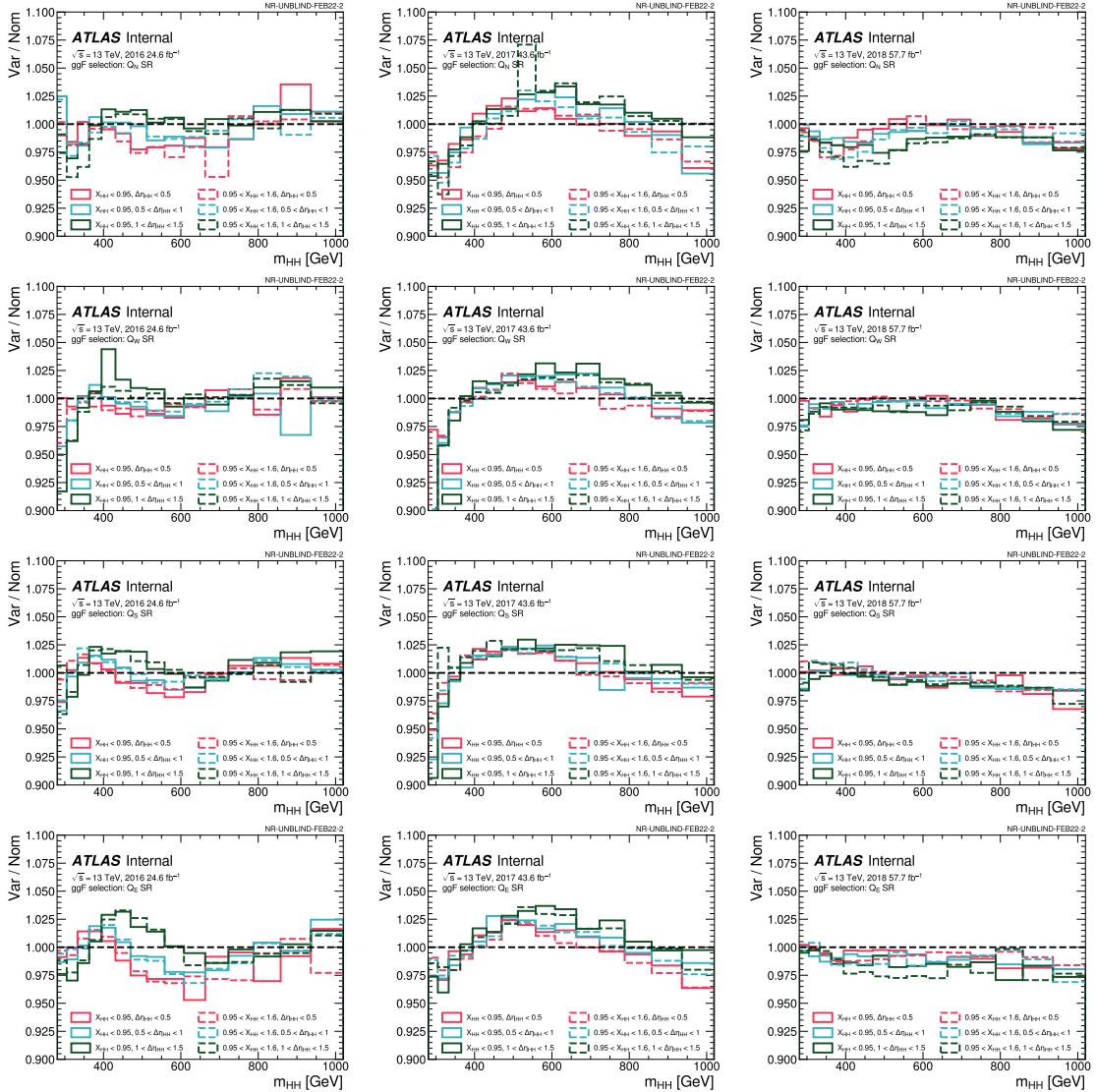


Figure 1.11: Impact of background shape nuisance parameter variation on  $m_{HH}$  in different kinematic categories for the ggF channel. Each column is a different year for the ggF channel templates while the rows show the SR NP quadrants.

Given the arguments above, we choose to correlate the background shape nuisance parameters across kinematic categories, within a year and each quadrant. Therefore, the number of free nuisance parameters in the ggF fit related to the background shape is reduced from 72 (3 years  $\times$  6 categories  $\times$   $(Q_N, Q_E, Q_S, Q_W)$ ) to 12 (3 years  $\times$   $(Q_N, Q_E, Q_S, Q_W)$ ).

For completeness, we also show the templates by category with the different quadrant contributions overlaid in Figure 1.12.

The initial analysis unblinding occurred with a different nuisance parameter correlation scheme, where all parameters were treated as uncorrelated systematics. This choice was made due to the lack of coherent variations in the different validation regions in the analysis, particularly the 3b1f region. Therefore, we believed giving flexibility for the fit to vary in different directions in different categories to be the best option. In addition, it was observed that the expected SR limits with the uncorrelated were more performant across different  $\kappa_\lambda$  hypotheses. Results pertaining to this unblinding, which show clear indication and preference for the correlated scheme described above, can be seen in Appendix ???. This appendix also collects extensive studies on understanding these results.

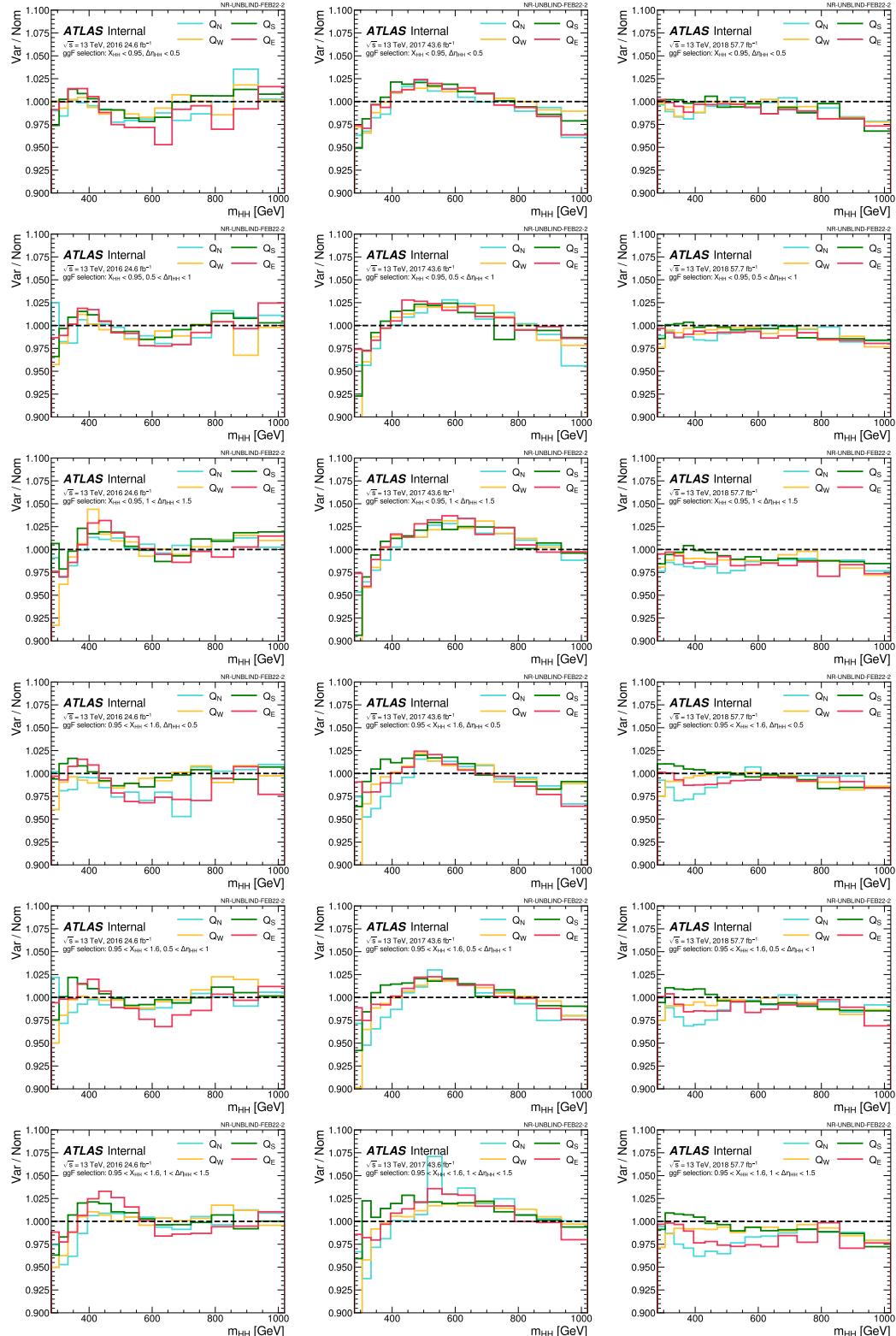


Figure 1.12: Impact of background shape nuisance parameter variation on  $m_{HH}$  in different kinematic categories for the ggF channel. Each column is a different years of the ggF channel templates while the rows show each category with the SR NP quadrants overlaid.

## 1.4 Background validation

- Let's start off by re-showing the analysis graphic that I had before, but add **red text** for the inverted cuts!

... mybib