$HH \to b\bar{b}b\bar{b}$ or How I Learned to Stop Worrying and Love the QCD Background

Sean Gasiorowski

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2021

Reading Committee:

Anna Goussiou, Chair

Jason Detwiler

Shih-Chieh Hsu

David Kaplan

Henry Lubatti

Thomas Quinn

Gordon Watts

Program Authorized to Offer Degree:
Physics

University of Washington

Abstract

$HH \to b\bar{b}b\bar{b}$ or How I Learned to Stop Worrying and Love the QCD Background

Sean Gasiorowski

Chair of the Supervisory Committee:
Professor Anna Goussiou
Physics

Insert abstract here

TABLE OF CONTENTS

Page

List of Figures iii

Glossary iv

Chapter 1: The Standard Model of Particle Physics 1
1.1 Particles and Fields 1

Chapter 2: Beyond the Standard Model 3

Chapter 3: Experimental Apparatus 4

Chapter 4: Simulation 5

Chapter 5: Reconstruction 6

Chapter 6: The Anatomy of an LHC Search 7
6.1 Object Selection and Identification 7
6.2 Defining a Signal Region 7
6.3 Background Estimation 7
6.4 Uncertainty Estimation 7
6.5 Hypothesis Testing 7

Chapter 7: Search for pair production of Higgs bosons in the $b\bar{b}b\bar{b}$ final state 8
7.1 Data and Monte Carlo Simulation 9
7.2 Triggers and Object Definitions 9
7.3 Analysis Selection 9
7.4 Background Estimation 9
7.5 Uncertainties 15

7.6 Statistical Analysis 20
7.7 Results 20

LIST OF FIGURES

Figure Number Page

GLOSSARY

ARGUMENT: replacement text which customizes a LaTeX macro for each particular usage.

ACKNOWLEDGMENTS

As anyone who has written a Ph.D. thesis will probably tell you, it's been a journey. We laughed, we cried, we bled occasionally (though nothing too serious). A pandemic happened, I learned how to make sourdough (see the appendix for more details). I learned how to ski, discovered a love for hiking, and ate large amounts of cheese. The list of people who I have met and shared deep and memorable experiences with is long – I fear to list you all here in case I miss someone! – but please do know that I treasure you. This is the beauty and tragedy of doing a Ph.D. half in Seattle, half at CERN: it allows you to build strong friendships with a large group of people, and then scatters you all across the globe. So to the Seattle friends, to the CERN friends, to the friends from undergrad, and high school, and even earlier, and to everyone in between, thank you for being a part of my life, and I hope to see you soon.

Of course, a thank you to my family for their continuing support, vacationing adventures, and for trying their best to learn physics along with me (my dad re: ATLAS – "This is pretty complicated isn't it?").

And finally a huge thank you to my group: Anna, for your guidance and support, and for always caring about me as a person in addition to me as a physicist. And Jana, for guidance and support, of course, but also for looking at/giving comments on almost literally every single one of my talks (even if you didn't always get my jokes).

DEDICATION

To life!

Chapter 1

THE STANDARD MODEL OF PARTICLE PHYSICS

"I may be bad, but I'm perfectly good at it" - Rihanna re: the Standard Model (SM), or so I've been told

The Standard Model of Particle Physics (SM) is a monumental historical achievement, providing a formalism with which one may describe everything from the physics of everyday experience to the physics that is studied at very high energies at the Large Hadron Collider (Chapter 3). In this chapter, we will provide a brief overview of the pieces that go into the construction of such a model. The primary focus of this thesis is searches for pair production of Higgs bosons decaying to four b-quarks. Consequently, we will pay particular attention to the relevant pieces of the Higgs Mechanism, as well as the theory behind searches at a hadronic collider.

1.1 Particles and Fields

What is a particle? The Standard Model describes a set of fundamental, point-like, objects (shown in Figure SM FIGURE). These objects have distinguishing characteristics (e.g., mass and spin). These objects interact in very specific ways. The set of objects and their interactions result in a set of observable effects, and these effects are the basis of a field of experimental physics.

The effects of these objects and their interactions are familiar as fundamental forces: electromagnetism (photons, electrons), the strong interaction (quarks, gluons), the weak interaction (neutrinos, W and Z bosons). Gravity is not described in this model, as the

weakest, with effects most relevant on much larger distance scales than the rest. However, the description of these other three is powerful – verifying and searching for cracks in this description is a large effort, and the topic of this thesis.

The formalism for describing these particles and their interactions is that of quantum field theory. Classical field theory is most familiar in the context of, e.g., electromagnetism – an electric field exists in some region of space, and a charged point-particle experiences a force characterized by the charge of the point-particle and the magnitude of the field at the location of the point-particle in spacetime. The same language translates to quantum field theory. Here, each particle is represented by a quantum field describing its influence on a region of spacetime. Particles also have charges which describe the forces they experience when interacting with other particles (other quantum fields). Most familiar is electric charge – however this applies to e.g., the strong interaction as well, where particles have an associated *color charge* describing behavior under the strong force.

Particles are observed to behave in different ways under these different forces. These behaviors respect certain *symmetries*, which are most naturally described in the language of group theory. The respective fields, charges, and generators of these symmetry groups are the basic pieces of the SM Lagrangian, which describes the full dynamics of the theory. In the following, we will build up the basic components of this Lagrangian.

1.1.1 Quantum Electrodynamics

1.1.2 The Weak Force

1.1.3 Quantum Chromodynamics

1.1.4 The Higgs mechanism

Chapter 2

BEYOND THE STANDARD MODEL

The job of the experimental particle physicist is two-fold: observation and measurement of the various predictions of the Standard Model are important and exciting – the

Chapter 3

EXPERIMENTAL APPARATUS

Chapter 4

SIMULATION

Chapter 5

RECONSTRUCTION

Chapter 6

THE ANATOMY OF AN LHC SEARCH

In this thesis so far, we have set the theoretical foundation for the work carried out at the LHC. We have described how one may translate between this theoretical foundation and what we are actually able to observe with the ATLAS detector. We have further stepped through the process of simulating production of specific physics processes and their appearance in our detector, allowing us to describe how a hypothetical physics model would be seen in our experiment. The question then becomes: all of these pieces are on the table, what do we do with them? This chapter attempts to answer exactly that, setting up a roadmap for assembling these pieces into a statement about the universe.

6.1 Object Selection and Identification

As described in Chapter 5, there is a complicated set of steps for going from electrical signals in a detector to physics objects.

6.2 Defining a Signal Region

6.3 Background Estimation

6.4 Uncertainty Estimation

6.5 Hypothesis Testing

Chapter 7

SEARCH FOR PAIR PRODUCTION OF HIGGS BOSONS IN THE $b\bar{b}b\bar{b}$ FINAL STATE

This chapter presents two complementary searches for pair production of Higgs bosons in the final state. Such searches are separated based on the signal models being considered: resonant production, in which a new spin-0 or spin-2 particle is produced and decays to two Standard Model Higgs bosons, and non-resonant production, which is sensitive to the value of the Higgs self-coupling λ_{HHH}. Further information on the theory behind both channels can be found in Chapter *TODO: Fill in theory chapter.*

While the searches face many similar challenges and procede (in broad strokes) in a very similar manner, separate optimizations are performed to maximize the respective sensitivities for these two very different sets of signal hypotheses. More particularly, resonant signal hypotheses are (1) very peaked in values of the mass of the HH candidate system near the value of the resonance mass considered and (2) considered across a very broad range of signal mass hypotheses. The resonant searches are therefore split into resolved and boosted topologies based on Lorentz boost of the decay products, with the resolved channel as one of the primary focuses of this thesis. Further, several analysis design decisions are made to allow for sensitivity to a broad range of masses – in particular, though sensitivity is limited at lower values of m_{HH} relative to other channels *TODO: Combination, bbyy* due to the challenging background topology, retaining and properly reconstructing these low mass events allows the $b\bar{b}b\bar{b}$ channel to retain sensitivity up until the kinematic threshold at 250 GeV.

In contrast, non-resonant signal hypotheses are quite broad in m_{HH}, and have a much more limited mass range, with Standard Model production peaking near 400 GeV, and the majority of the analysis sensitivity able to be captured with a resolved topology. Even for

Beyond the Standard Model signal hypotheses, which may have more events at low m_{HH}, the non-resonant nature of the production allows the $b\bar{b}b\bar{b}$ channel to retain sensitivity while discarding much of the challenging low mass background. Such freedom allows for decisions which focus on improved background modeling for the middle to upper HH mass regime, resulting in improved modeling and smaller uncertainties than would be obtained with a more generic approach.

Both searches are presented in the following, with emphasis on particular motivations for, and consequences of, the various design decisions involved for each respective set of signal hypotheses.

7.1 Data and Monte Carlo Simulation

Both the resonant and non-resonant searches are performed on the full ATLAS Run 2 dataset, consisting of $\sqrt{s} = 13\,\text{TeV}$ proton-proton collision data taken from 2016 to 2018 inclusive. The

7.2 Triggers and Object Definitions

7.3 Analysis Selection

7.3.1 Resonant Search

7.3.2 Non-resonant Search

7.4 Background Estimation

fter the event selection described in Section **??** there are two major backgrounds, QCD and $t\bar{t}$, with relative proportions *TODO: Fill in appropriate ggF and VBF background composition*. Following the approach used for the resonant analysis, a fully data-driven estimate is used here. This is warranted due to the flexibility of the estimation method, as well as the high relative proportion of QCD background, and allows for the use of machine learning methods in the construction of the background estimate. However, it sacrifices an explicit treatment of the $t\bar{t}$ component. Performance of the background estimate on the $t\bar{t}$ component is thus

checked in Appendix *TODO: update ttbar studies/do for VBF*. The approach and its benefits over other methods are discussed below.

TODO: Do small background checks, text here kept from resonant Contributions of single Higgs processes to the background are checked in Appendix **??**. ZZ and HH backgrounds are checked in Appendix **??**. All are found to be negligible.

The foundation of the background estimate lies in the derivation of a reweighting function which matches the kinematics of events with exactly two b-tagged jets to those of events with four or more b-tagged jets. The reweighting function and overall normalization are derived in the control region. Systematic bias associated with extrapolating to the signal region in the Higgs candidate mass plane is assessed in the validation region. All derivation is done after the full event selection.

7.4.1 The Two Tag Region

Events in data with exactly two b-tagged jets are used for the data driven background estimate. The hypothesis here is that, due to the presence of multiple b-tagged jets, the kinematics of such events are similar to the kinematics of events in higher b-tagged regions (i.e. events with three and four b-tagged jets, respectively), and any differences can be corrected by a reweighting procedure. The region with three b-tagged jets is split into two b-tagging regions, with the $3b + 1$ loose region used as an additional signal region (see Section *TODO: Add ref*). The lower tagged $3b$ component ($3b + 1$ fail, as described in Section **??**) is reserved for validation of the background modelling procedure. Events with fewer than two b-tagged jets are not used for this analysis, as they are relatively more different from the higher tag regions.

The nominal event selection requires at least four jets in order to form Higgs candidates. For the four tag region, these are the four highest p_T b-tagged jets. For the three tag regions, these jets are the three b-tagged jets, plus the highest p_T jet satisfying a loosened b-tagging requirement. Similarly, and following the approach of the resonant analysis, the two tag region uses the two b-tagged jets and the two highest p_T non-tagged jets to form Higgs candidates.

Combinatoric bias from selection of different numbers of b-tagged jets is corrected as a part of the kinematic reweighting procedure through the reweighting of the total number of jets in the event. In this way, the full event selection may be run on two tagged events.

7.4.2 Kinematic Reweighting

The set of two tagged data events is the fundamental piece of the data driven background estimate. However, kinematic differences from the four tag region exist and must be corrected in order for this estimate to be useful. Binned approaches based on ratios of histograms (*TODO: Cite either 15+16 or full Run 2 VBF*) have been previously considered, but are limited in their handling of correlations between variables and by the "curse of dimensionality", i.e. the dataset becomes sparser and sparser in "reweighting space" as the number of dimensions in which to reweight increases, limiting the number of variables used for reweighting. This leads either to an unstable fit result (overfitting with finely grained bins) or a lower quality fit result (underfitting with coarse bins).

Note that even machine learning methods such as Boosted Decision Trees (BDTs), may suffer from this curse of dimensionality, as the depth of each decision tree used is limited by the available statistics after each set of corresponding selections (cf. binning in a more sophisticated way), limiting the expressivity of the learned reweighting function.

To solve these issues, a neural network based reweighting procedure is used here, following the success of the method used for the resonant search [**Abbott:2708605**]. This is a truly multivariate approach, allowing for proper treatment of variable correlations. It further overcomes the issues associated with binned approaches by learning the reweighting function directly, allowing for greater sensitivity to local differences and helping to avoid the curse of dimensionality.

Neural Network Reweighting

Let $p_{4b}(x)$ and $p_{2b}(x)$ be the probability density functions for four and two tag data respectively across some input variables x. The problem of learning the reweighting function between two

and four tag data is then the problem of learning a function $w(x)$ such that

$$p_{2b}(x) \cdot w(x) = p_{4b}(x) \tag{7.1}$$

from which it follows that

$$w(x) = \frac{p_{4b}(x)}{p_{2b}(x)}. \tag{7.2}$$

This falls into the domain of density ratio estimation, for which there are a variety of approaches. The method considered here is modified from [**NNloss**, **NNloss1**], and depends on a loss function of the form

$$\mathcal{L}(R(x)) = \mathbb{E}_{x\sim p_{2b}}[\sqrt{R(x)}] + \mathbb{E}_{x\sim p_{4b}}[\frac{1}{\sqrt{R(x)}}]. \tag{7.3}$$

where $R(x)$ is some estimator dependent on x and $\mathbb{E}_{x\sim p_{2b}}$ and $\mathbb{E}_{x\sim p_{4b}}$ are the expectation values with respect to the 2b and 4b probability densities. A neural network trained with such a loss function has the objective of finding the estimator, $R(x)$, that minimizes this loss. It is straightforward to show (Appendix **??**) that

$$\arg\min_R \mathcal{L}(R(x)) = \frac{p_{4b}(x)}{p_{2b}(x)} \tag{7.4}$$

which is exactly the form of the desired reweighting function.

In practice, to avoid imposing explicit positivity constraints, the substitution $Q(x) \equiv \log R(x)$ is made. The loss function then takes the equivalent form

$$\mathcal{L}(Q(x)) = \mathbb{E}_{x\sim p_{2b}}[\sqrt{e^{Q(x)}}] + \mathbb{E}_{x\sim p_{4b}}[\frac{1}{\sqrt{e^{Q(x)}}}], \tag{7.5}$$

with solution

$$\arg\min_Q \mathcal{L}(Q(x)) = \log\frac{p_{4b}(x)}{p_{2b}(x)}. \tag{7.6}$$

Taking the exponent then results in the desired reweighting function.

Variables and Results

The neural network is trained on a variety of variables sensitive to two vs. four tag differences. To help bring out these differences, the natural logarithm of some of the variables with a

large, local change is taken. The set of training variables for both ggF and VBF channels is shown in Table 7.1. *TODO: Update with final configuration/optimization when ready*

As mentioned, this number of jets variable is used to address the combinatoric bias of the two-tag region. The neural network used for the ggF reweighting has three densely connected hidden layers of 50 nodes each with ReLU activation functions and a single node linear output. The same applies to the neural network used for the VBF reweighting. These configuration demonstrates good performance in the modelling of a variety of relevant variables, including m_{HH}, when compared to a range of networks of similar size.

In practice, a given training of the reweighting neural network is subject to variation due to training statistics and initial conditions. An uncertainty is assigned to account for this (Section **??**), which relies on training an ensemble of reweighting networks [**DeepEnsembles**]. To increase the stability of the background estimate, the median of the predicted weight for each event is calculated across the ensemble. This median is then used as the nominal background estimate. This approach is indeed seen to be much more stable and to demonstrate a better overall performance than a single arbitrary training. Each ensemble used for this analysis consists of 100 neural networks, trained as described in Section **??**.

The training of the ensemble used for the nominal estimate is done in the kinematic Control Region. The prediction of these networks in the Signal Region is then used for the nominal background estimate. In addition, a separate ensemble of networks is trained in the Validation Region. The difference between the prediction of the nominal estimate and the estimate from the VR derived networks in the Signal Region is used to assign a systematic uncertainty associated with extrapolating in the Higgs Candidate mass plane. Further details on this systematic uncertainty are shown in Section **??**. Note that although the same procedure is used for both Control and Validation Region trained networks, only the median estimate from the VR derived reweighting is used for assessing the extrapolation uncertainty – no additional "uncertainty on the uncertainty" from VR ensemble variation is applied.

Each reweighted estimate is normalized such that the reweighted $2b$ yield matches the $4b$

Table 7.1: Set of input variables used for tthe $2b$ to $4b$ reweighting in the ggF and VBF channels respectively.

ggF	**VBF**
1. $\log(p_T)$ of the 4th leading Higgs candidate jet	1. $\log(p_T)$ of the 4th leading Higgs candidate jet
2. $\log(p_T)$ of the 2nd leading Higgs candidate jet	2. $\log(p_T)$ of the 2nd leading Higgs candidate jet
3. $\log(\Delta R)$ between the closest two Higgs candidate jets	3. $\log(\Delta R)$ between the closest two Higgs candidate jets
4. $\log(\Delta R)$ between the other two Higgs candidate jets	4. $\log(\Delta R)$ between the other two Higgs candidate jets
5. Average absolute value of Higgs candidate jet η	5. Average absolute value of Higgs candidate jet η
6. $\log(p_T)$ of the di-Higgs system.	6. $\log(p_T)$ of the di-Higgs system.
7. ΔR between the two Higgs candidates	7. ΔR between the two Higgs candidates
8. $\Delta\phi$ between the jets in the leading Higgs candidate	8. $\Delta\phi$ between the jets in the leading Higgs candidate
9. $\Delta\phi$ between the jets in the subleading Higgs candidate	9. $\Delta\phi$ between the jets in the subleading Higgs candidate
10. $\log(X_{Wt})$, where X_{Wt} is the variable used for the top veto	10. $\log(X_{Wt})$, where X_{Wt} is the variable used for the top veto
11. Number of jets in the event	11. Number of jets in the event

yield in the corresponding training region. Note that this applies to each of the networks used in each ensemble, where the normalization factor is also subject to the procedure described in Section **??**. As the median over these normalized weights is not guaranteed to preserve this normalization, a further correction is applied such that the $2b$ yield, after the median weights are applied, matches the $4b$ yield in the corresponding training region. As no preprocessing is applied to correct for the class imbalance between $2b$ and $4b$ events entering the training, this ratio of number of $4b$ events ($n(4b)$) over number of $2b$ events ($n(2b)$) is folded into the learned weights. Correspondingly, the set of normalization factors described above is near 1 and the learned weights are centered around $n(4b)/n(2b)$ (roughly 0.01 over the full dataset). This normalization procedure applies for all instances of the reweighting (e.g. those used for validations in Section **??**), with appropriate substitutions of reweighting origin (here $2b$) and reweighting target (here $4b$).

Note that, though there are different trigger and pileup selections during each year, the statistically limited VBF channel benefits from combining all years together in performing the reweighting. The training input is the combined dataset from the years 2016 - 2018 with all configurations remaining unchanged except for the addition of the year of the given event as an input variable on top of the reweighting variables listed in Table 7.1. More detailed studies comparing the nominal split-year training and training on all years together for the VBF reweighting is presented in *TODO: ref appendix*. For the ggF channel, a similar approach was explored *TODO: ref appendix*, but was found to have minimal benefit over reweighting in each year separately. As the trigger selections for each year significantly impact the kinematics of each year, and thus categorizing by year is expected to reflect groupings of kinematically similar events and to provide a meaningful degree of freedom in the signal extraction fit, the split-year approach is kept by the ggF analysis.

7.5 Uncertainties

A variety of uncertainties are assigned to account for known biases in the underlying methods, calibrations, and objects used for this analysis. The largest such uncertainty is associated

with the kinematic bias inherent in deriving the background estimate away from the signal region. However, a statistical biasing of this same estimate has an effect of a similar magnitude. Additionally, due to the use of Monte Carlo for signal modelling and b-tagging calibration, uncertainties related to mismodellings in simulation must also be accounted for. These components, and their impact on this analysis, are described here in detail. Relative magnitudes of the uncertainties for each year are shown in Tables **??**, **??**, and **??** along with an estimate of the impact of the statistics of 4b data in the signal region on the total error. Note that, while the Poisson error (from 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.

7.5.1 Statistical Uncertainties and Bootstrapping

There are two components to the statistical error for the neural network background estimate. The first is standard Poisson error, 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 two tag and four tag datasets, it is the statistical uncertainty due to the four tag 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 $\lim_{n\to\infty} \text{Binomial}(n, 1/n) = \text{Poisson}(1)$, 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. 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 variation from this bootstrapping procedure is used to assign a bin-by-bin uncertainty which is treated as a statistical uncertainty in the fit. Due to practical constraints, a procedure for approximating the full bootstrap error band is developed which demonstrates good agreement with the full bootstrap uncertainty. This procedure is described below.

Calculating the Bootstrap Error Band

The standard procedure to calculate the bootstrap uncertainty would proceed as follows: first, each network trained on each bootstrap replica dataset would be used to produce a histogram in the variable of interest. This would result in a set of replica histograms (e.g. for 100 bootstrap replicas, 100 histograms would be created). The nominal estimate would then be the mean of bin values across these replica histograms, with errors set by the corresponding standard deviation.

In practice, such an approach is inflexible and demanding both in computation and in storage, in so far as we would like to produce histograms in many variables, with a variety of different cuts and binnings. This motivates a derivation based on event-level quantities. However, due to non-trivial correlations between replica weights, simple linear propagation of event weight variation is not correct.

We therefore adopt an approach which has been empirically found to produce results (for this analysis) in line with those produced by generating all of the histograms, as in the standard procedure. This approach is described below. Note that, for robustness to outliers and weight distribution asymmetry, the median and interquartile range (IQR) are used for

the central value and width respectively (as opposed to the mean and standard deviation).

The components involved in the calculation have been mentioned in Section **??** and are as follows:

1. Replica weight (w_i): weight predicted for a given event by a network trained on replica dataset i.

2. Replica norm (α_i): normalization factor for replica i. This normalizes the reweighting prediction of the network trained on replica dataset i to match the correponding target yield.

3. Median weight (w_{med}): median weight for a given event across replica datasets, used for the nominal estimate. Defined (for 100 bootstrap replicas) as

$$w_{med} \equiv \text{median}(\alpha_1 w_1, \ldots \alpha_{100} w_{100}) \tag{7.7}$$

4. Normalization correction (α_{med}): normalization factor to match the predicted yield of the median weights (w_{med}) to the target yield in the training region.

As mentioned in Section **??**, the *nominal estimate* is constructed from the set of median weights and the normalization correction, i.e. $\alpha_{med} \cdot w_{med}$.

For the bootstrap error band, a "varied" histogram is then generated by applying, for each event, a weight equal to the median weight (with no normalization correction) plus half the interquartile range of the replica weights: $w_{varied} = w_{med} + \frac{1}{2}\,\text{IQR}(w_1, \ldots, w_{100})$.

This varied histogram is scaled to match the yield of the nominal estimate. To account for variation of the nominal estimate yield, a normalization variation is calculated from the interquartile range of the replica norms: $\frac{1}{2}\,\text{IQR}(\alpha_1, \ldots, \alpha_{100})$. This variation, multiplied into the nominal estimate, is used to set a baseline for the varied histogram described above.

Denoting $H(\text{weights})$ as a histogram constructed from a given set of weights, $Y(\text{weights})$

as the predicted yield for a given set of weights, the final varied histogram is thus:

$$H(w_{med}+\frac{1}{2}\,\mathrm{IQR}(w_1,\ldots,w_{100}))\cdot\frac{Y(\alpha_{med}w_{med})}{Y(w_{med}+\frac{1}{2}\,\mathrm{IQR}(w_1,\ldots,w_{100}))}+\frac{1}{2}\,\mathrm{IQR}(\alpha_1,\ldots,\alpha_{100})\cdot H(\alpha_{med}w_{med}) \tag{7.8}$$

where the first term roughly describes the behaviour of the bootstrap variation across the distribution of the variable of interest while the second term describes the normalization variation of the bootstrap replicas.

The difference between the varied histogram and the nominal histogram is then taken to be the bootstrap statistical uncertainty on the nominal histogram.

Figure *TODO: include figure* demonstrates how each of the components described above contribute to the uncertainty envelope for the 2017 Control Region and compares this approximate band to the variation of histograms from individual bootstrap estimates. The error band constructed from the above procedure is seen to provide a good description of the bootstrap variation.

7.5.2 Background Shape Uncertainties

To account for the systematic bias associated with deriving the reweighting function in the Control region and extrapolating to the Signal region, an alternative background model is derived in the Validation region. In contrast to previous work, this is done on the entire background model, due to the fully data-driven nature of the background used here. The alternative model and the baseline are consistent with the observed data in their regions and with each other. Differences between the alternative and baseline models are used to define a shape uncertainty on the m_{HH} spectrum, with a two-sided uncertainty defined by symmetrizing the difference about the baseline.

This uncertainty is split into two components to allow two independent variations of the m_{HH} spectrum: *TODO: HT vs quad splitting*

7.6 Statistical Analysis

7.7 Results