



ICEPP
The University of Tokyo



LHC-ATLAS 実験データを用いた VH 生成過程における ヒッグスボソンの性質の精密測定

田中碧人,

石野雅也^A, 江成祐二^A, 奥村恭幸^A, 廣瀬茂輝^B, 増渕達也^A
(東大理, 東大素セ^A, 筑波大数理^B)



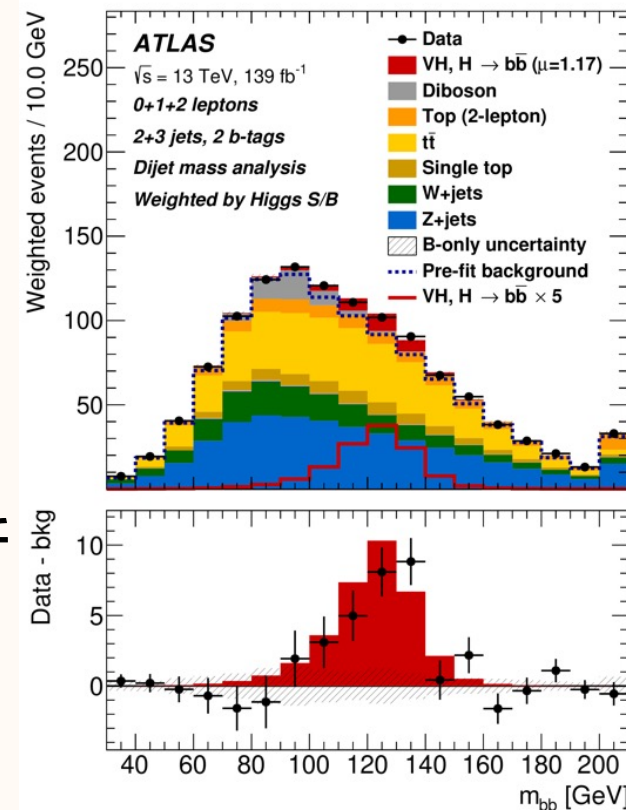
東京大学
THE UNIVERSITY OF TOKYO

LHC-ATLAS実験 と ヒッグス精密測定

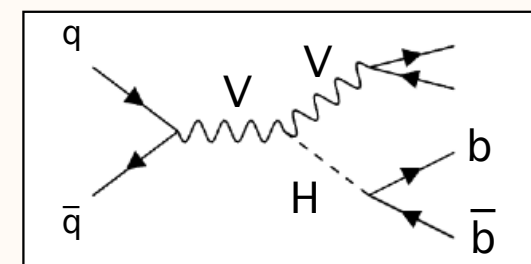
- LHC-ATLAS 実験の第2期実験 (2015 ~ 2018) のデータセットを解析に使用
 - 重心系13 TeV で陽子バンチ衝突
 - 積分ルミノシティ140.1 fb⁻¹
- ベクターボソンと随伴して生成されたヒッグスから2つのボトム(b)クォーク対へ崩壊する事象 “VH(H→bb)” を初めて有意に観測した
 - Hbb の崩壊割合は 58 %

■ VH(H→bb) 過程を更に精密測定することの意義 :

- ① ヒッグス生成過程で生じるベクターボソン運動量微分断面積が、高運動量領域で変化する
 - ・ 有効場理論 (EFT) で解釈し直した際の感度
- ② H→bb の結合の強さは、新物理の効果により標準模型からの差異が ~O(1) %生じると示唆されている
- 新物理の証拠を発見し理論モデルを絞り込むことは、エネルギーフロンティアの有力なアプローチ



[Eur. Phys. J. C 81 (2021) 178]



LHC-ATLAS実験 と ヒッグス精密測定

■ ① ヒッグス系の再構成の最適化

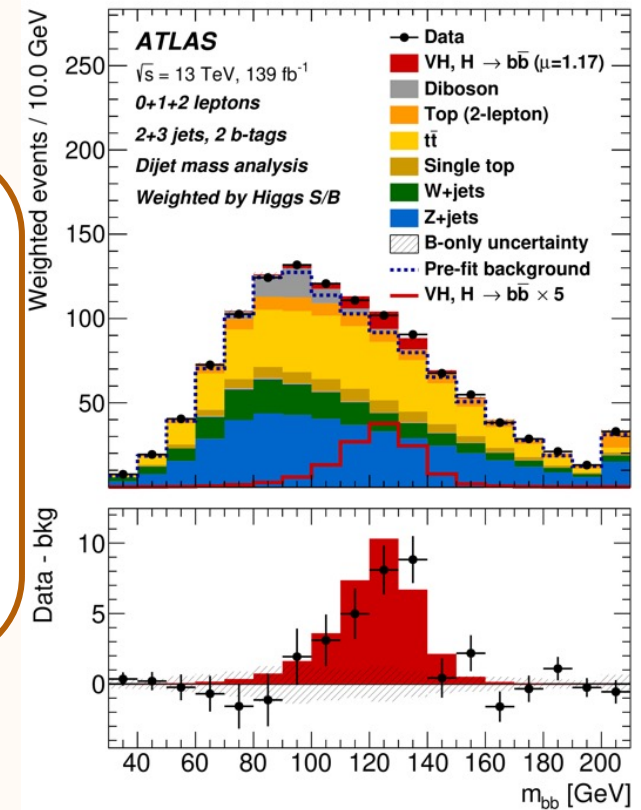
- 再構成に改良の余地があり、領域間の事象分別の正確性向上が可能

■ ② 理論系統誤差の正確な見積もり直し

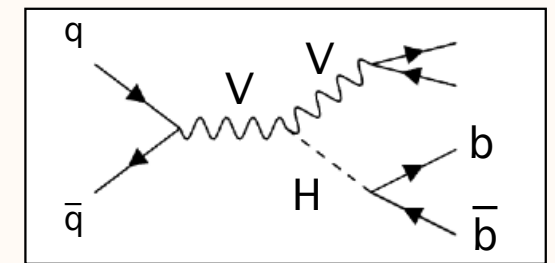
- VH(H→bb) の信号強度測定において、系統誤差が ~13 % で支配的
- 様々な過大評価の可能性をそれぞれ精査し、地道に必要な系統誤差を適応することが肝要

■ VH(H→bb) 過程を更に精密測定することの意義：

- ① ヒッグス生成過程で生じるベクターボソン運動量微分断面積が、高運動量領域で変化する
 - ・ 有効場理論 (EFT) で解釈し直した際の感度
- ② H→bb の結合の強さは、新物理の効果により標準模型からの差異が ~0(1) % 生じると示唆されている
- 新物理の証拠を発見し理論モデルを絞り込むことは、エネルギーフロンティアの有力なアプローチ



[Eur. Phys. J. C 81 (2021) 178]

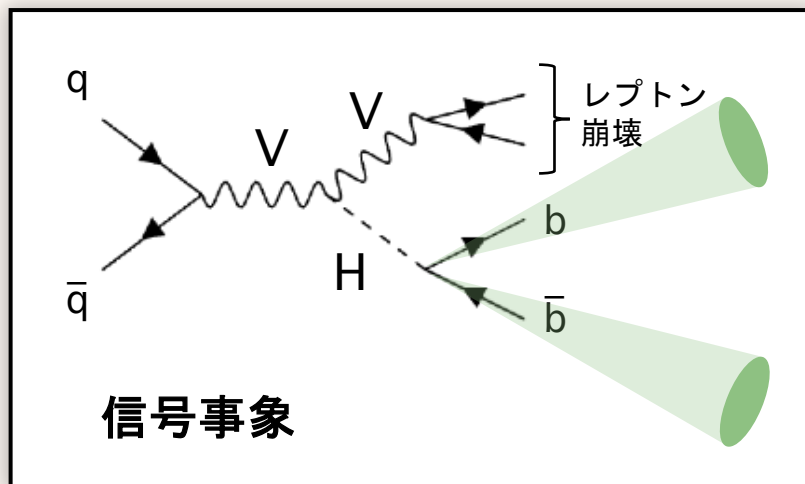


① ヒッグス系の再構成の最適化

微分断面積測定の領域間事象分別の正確性向上へ



注目する信号事象と解析手順



- 2つの b ジェット* からヒッグス粒子系を再構成
- 2つのレプトンからベクターボソンを再構成
 - 崩壊した荷電レプトン数ごとに分けて解析
 - Z \rightarrow $\nu\nu$ (0-lepton)
 - W \rightarrow lv (1-lepton)
 - Z \rightarrow ll (2-lepton)**

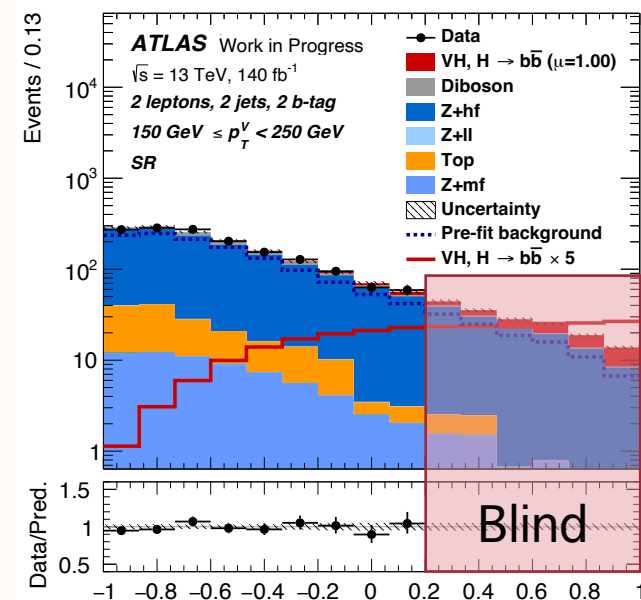
■ 2-lepton 解析の事象選別

- 2つの荷電レプトン($ee / \mu\mu$)の不変質量が Z ボソン付近
- ジェットが 20 GeV 以上の運動量
- b ジェットを2つ同定

■ MCシミュレーションと実験データを照合 (“Fit”)

- BDTを使用して、複数の運動学変数を入力し

信号事象(~ 1) と背景事象(~ -1) を分ける**最終分離変数を定義**



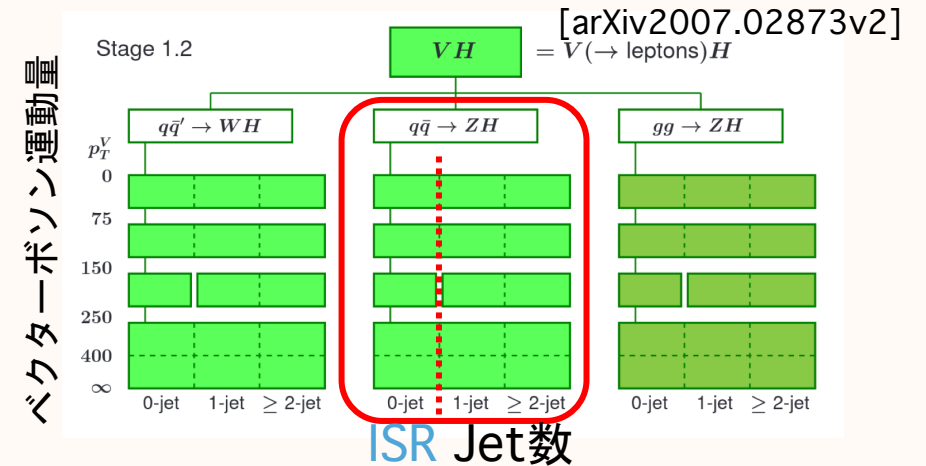
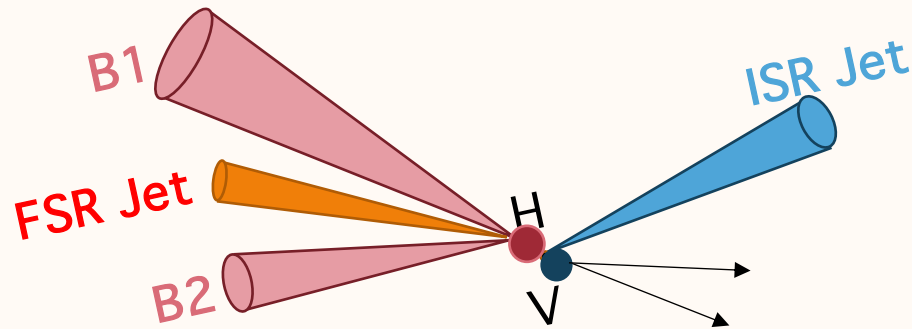
MC と実験データの
最終分離変数の Fit

*ジェット：QCD の閉じ込め効果によりクォーク/グルーオンの飛来方向に観測されるハドロン



信号事象の再構成の最適化余地 ~FSR recovery~

- Bottom quark 以外が由来の jet の可能性
 - ・ Final State Radiation (FSR): ヒッグスから2b-jetsへ崩壊したときのグルーオン放射
 - ・ Initial State Radiation (ISR): ヒッグス生成前の粒子からのグルーオン・クオーク放射
- ヒッグス生成微分断面積測定では、ヒッグス崩壊を考慮せず、
Truthレベルの“ベクターボソンの運動量”と“ISR Jet数”により領域分けがなされている
 - ・ VH 系で起きる新物理の寄与の抽出のために、領域を最適化している [arXiv1605.04692]
- VHbb 事象の解析レベルでも、同様の領域分けを行っている (i.e. ISR “0 jet” = VHbb “2 b jet”)
- 解析レベルでは FSR / ISR を区別しておらず、“0-jet” / “ ≥ 1 -jet”の領域間でイベントが混在してしまい
 微分断面積の精密測定に影響がある



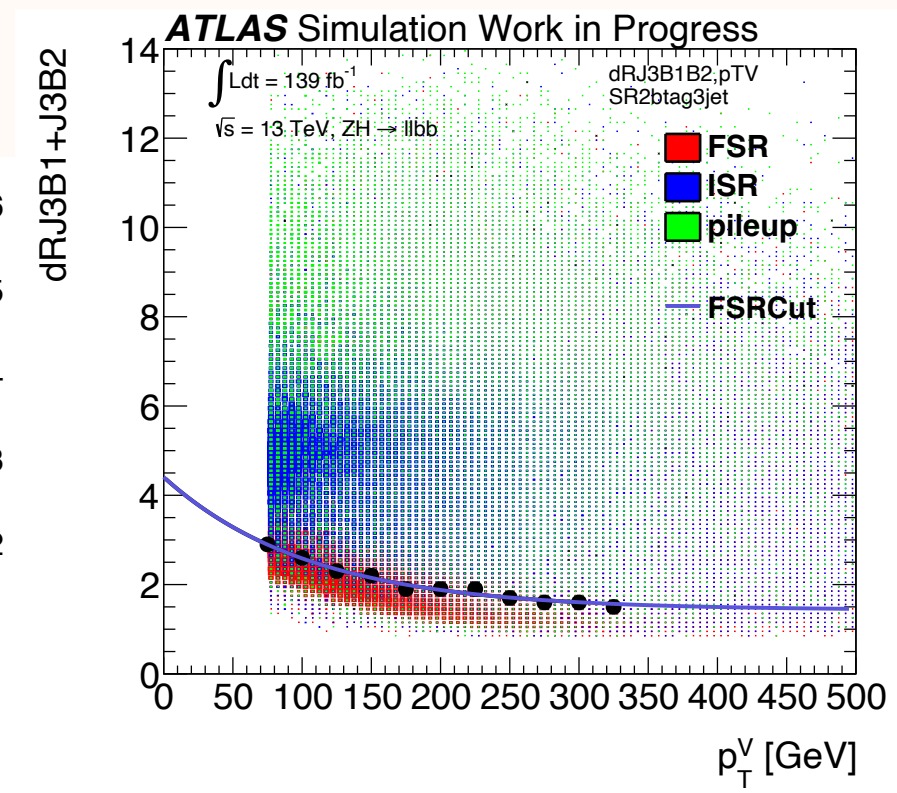
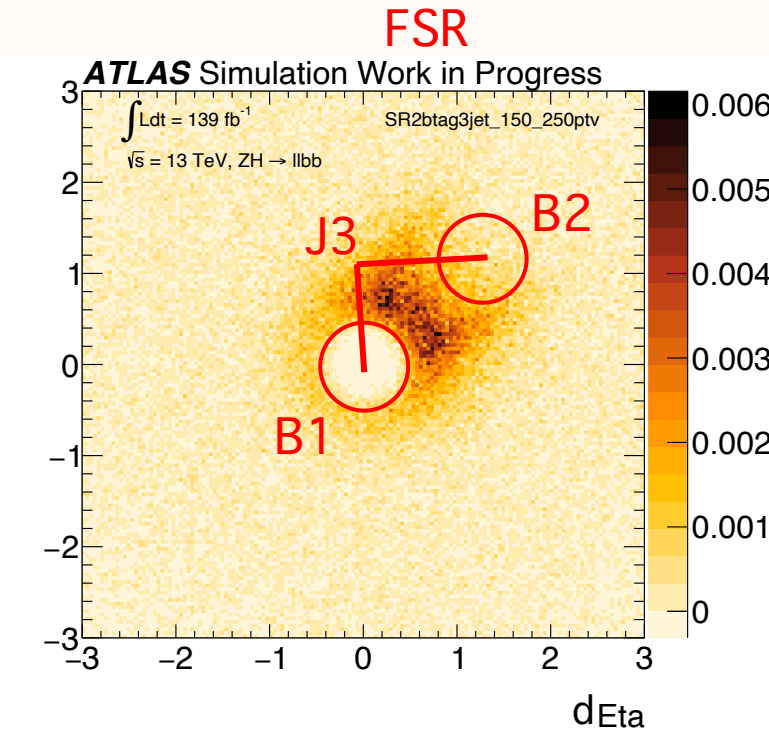
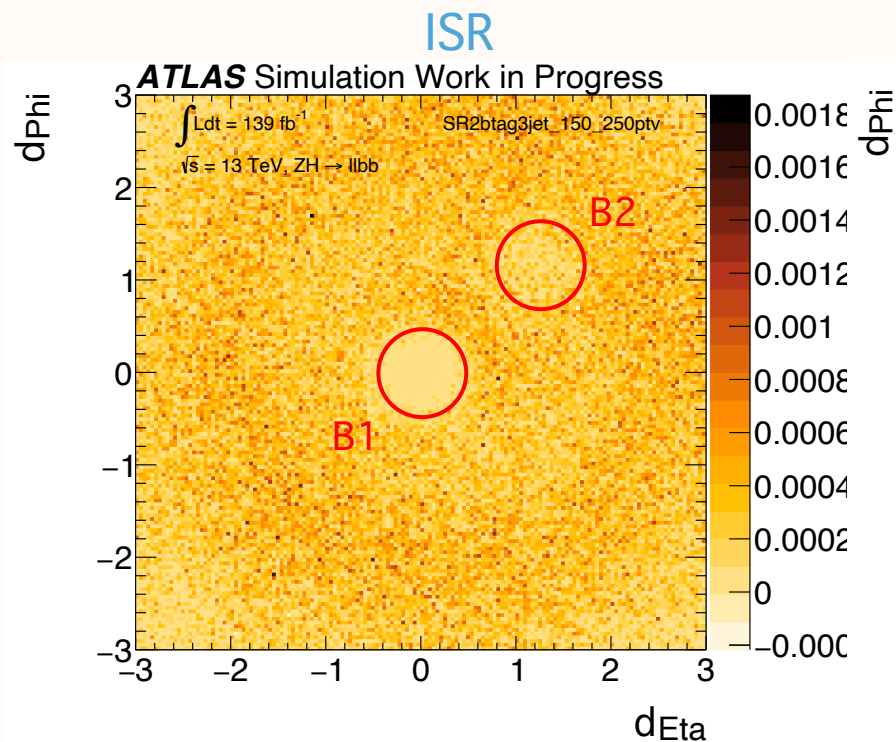
- 本研究では FSR Jet を正しく同定しヒッグス粒子系に戻す手法 (FSR recovery) の開発を行った



FSR jet 同定手法

- FSR は 2b jet 系周辺に発生し、ISR等は 2b jet 系とは無関係に発生する特性を考慮
 - Bottom quarks 間の QCD カラー交換により、2つの b-jet の間にグルーオン放射が発生しやすいと予測
- $dR(J-B1) + dR(J-B2)$ と p_{TV} の Cut 関数を定義
- Cut 関数以下 (2b jet 系周辺) を FSR jet と同定

➡ FSR jet が同定された事象を正しい領域へ (≥ 1 jet \rightarrow 0jet) 戻す

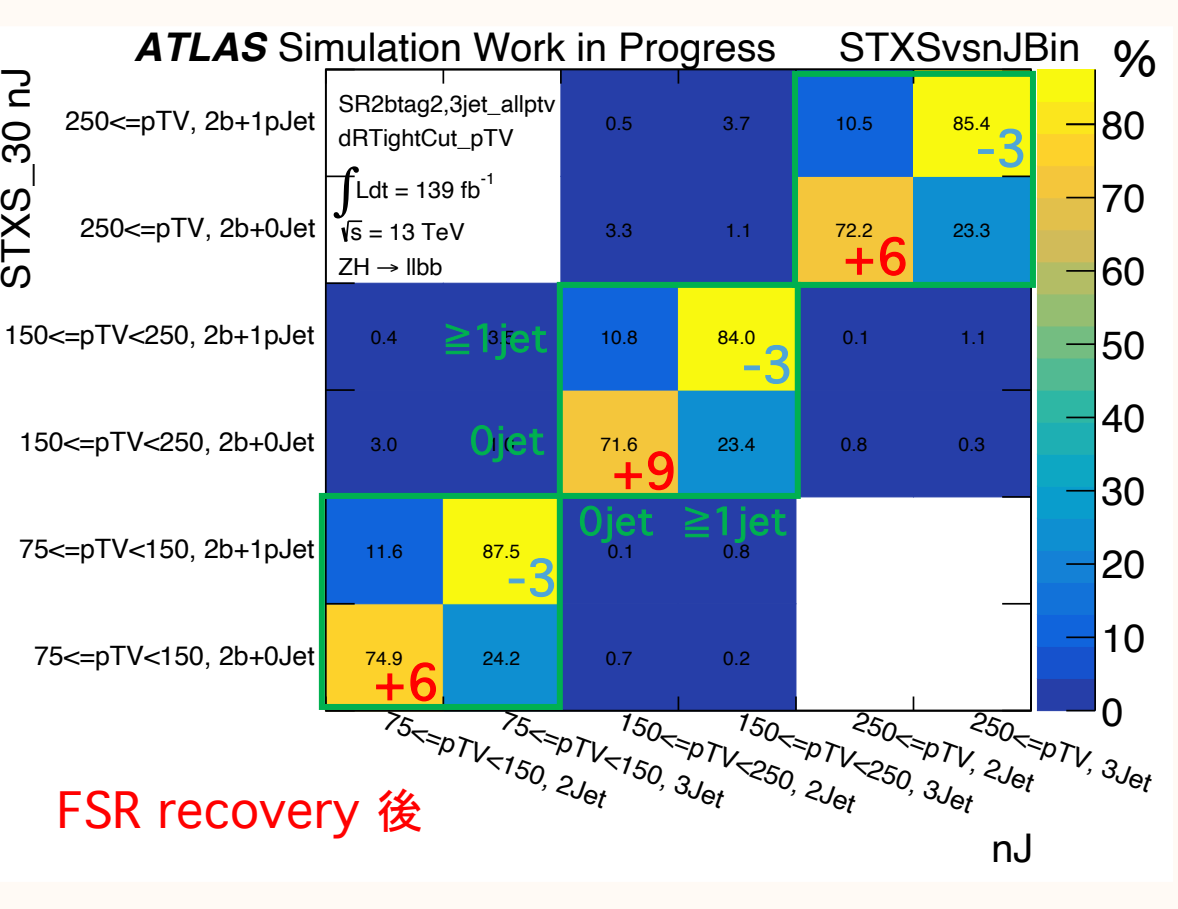
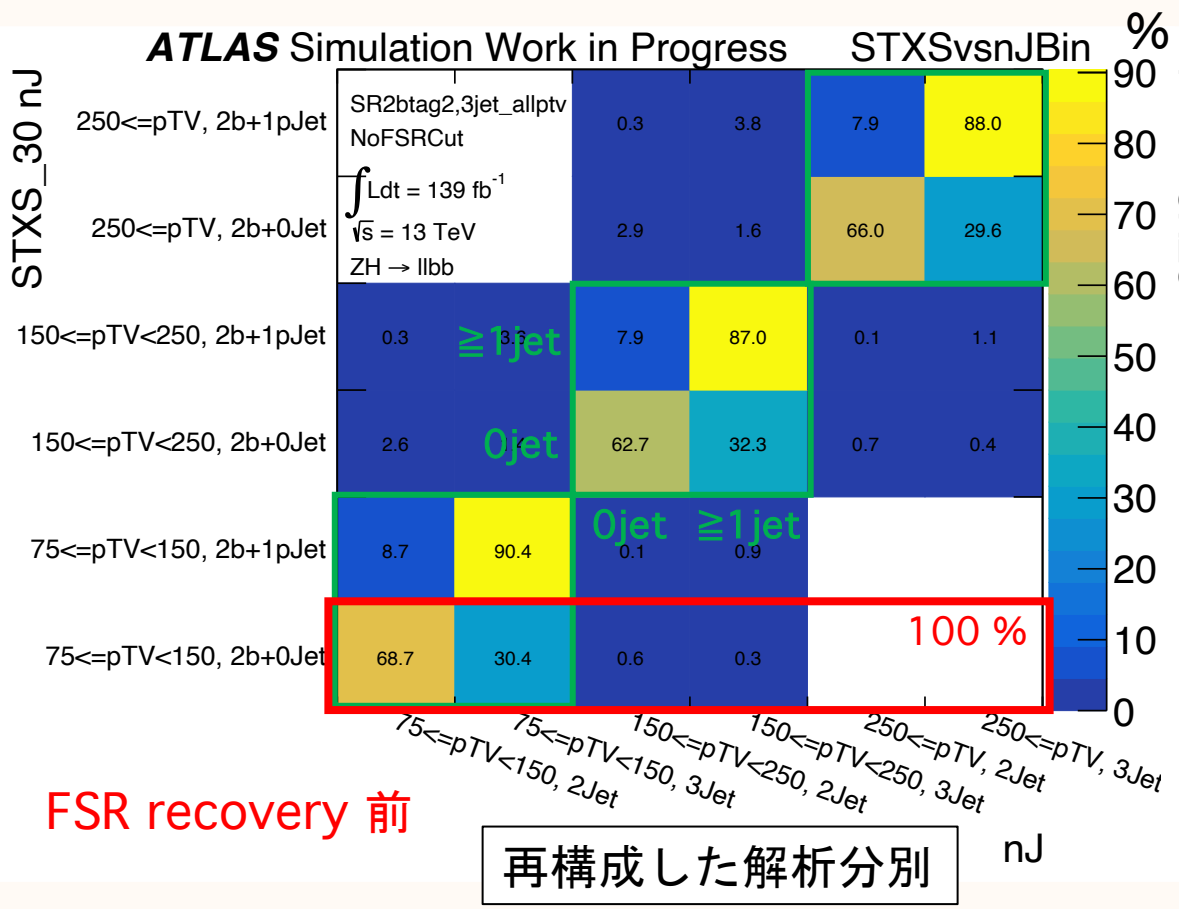


[2bjet と non-b jet の相対位置]



信号事象における領域間の事象分別能力評価

シミュレーションで想定しているTruth分別

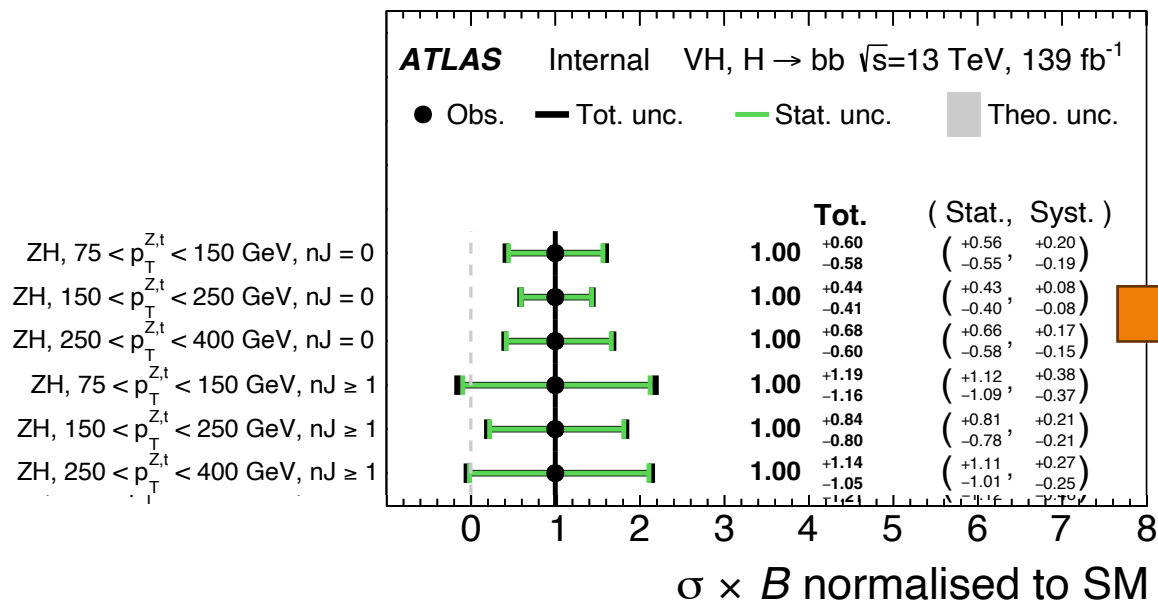


- FSR recovery により解析事象の ≥ 1 jet 領域 \rightarrow 0 jet 領域移動を行った
- 対角成分が全体的に良くなっている；Truth通りに解析で領域分けができた
- FSR recovery をしてFSR / ISRの分離がよくなった

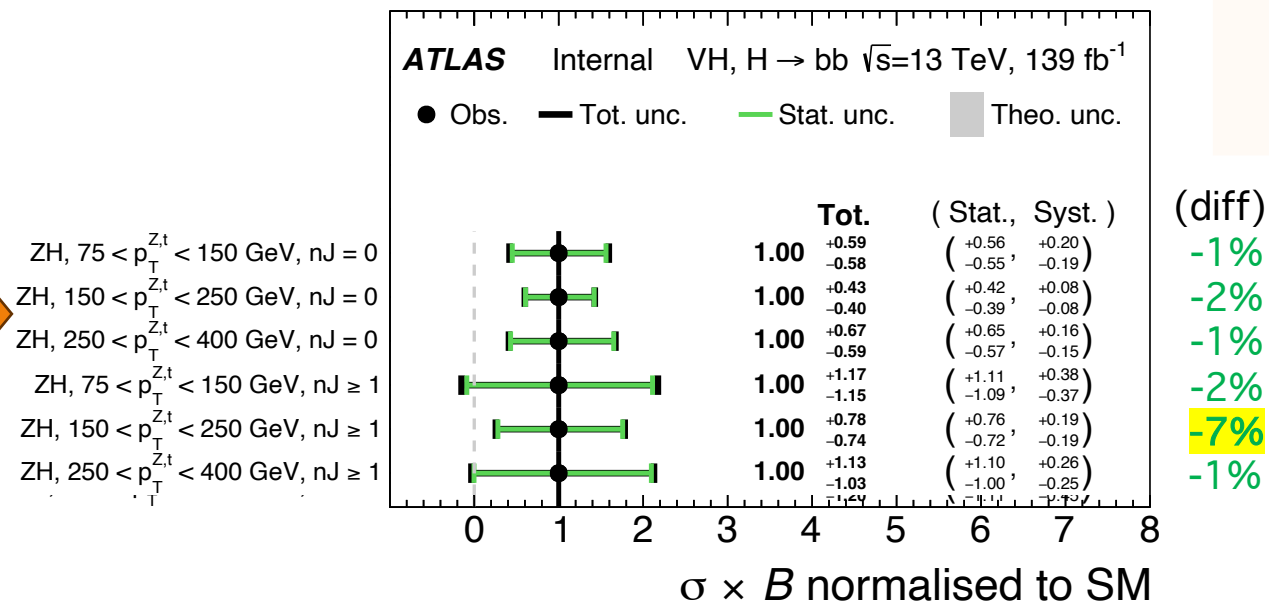


微分断面積測定の精度向上

- 全ての領域を同時に Fit
 - 事象数の変化による測定精度の変動をみるため、統計誤差のみ考慮
 - 微分断面積は SM 理論値に固定
- ISR jet 数 (n_J) ≥ 1 の領域で特に**精度向上**
- 今後、系統誤差も考慮した上で、微分断面積を固定せず測定予定



FSR recovery 前



FSR recovery 後

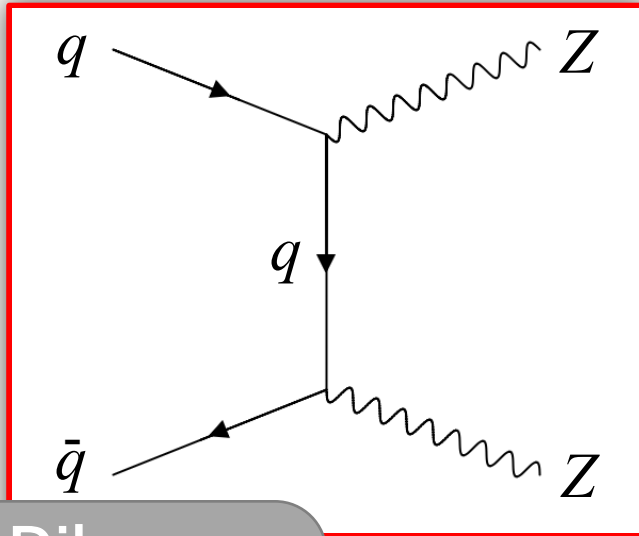


② 理論系統誤差の正確な 見積もり直し

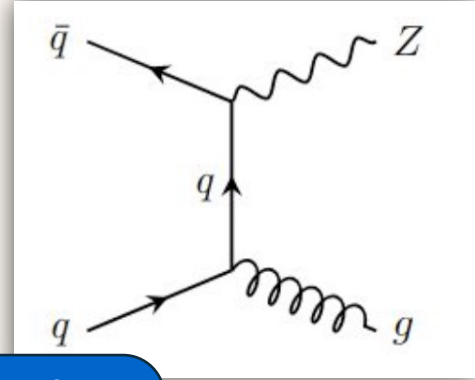
H→bb 信号強度測定の理論系統誤差の減少を模索して



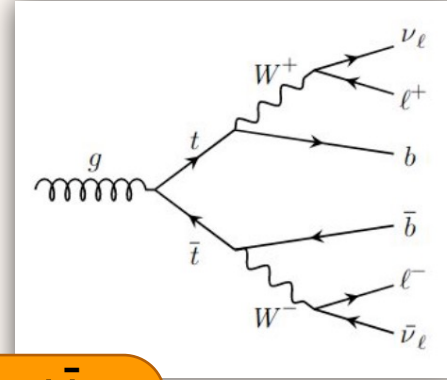
誤差評価改善余地がある事象 : Diboson 生成過程



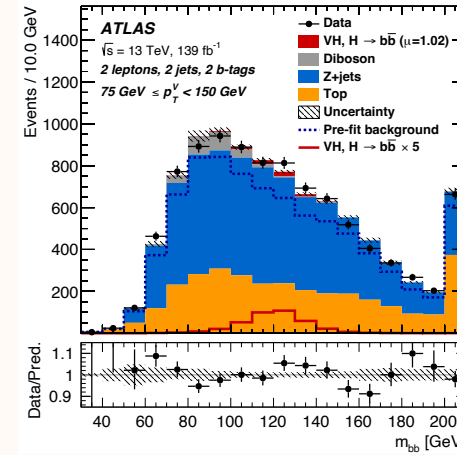
Diboson



Z+jets



$t\bar{t}$



2つのbジェットの不変質量分布

■ Diboson 生成過程

- 本解析の重要な背景事象の一つ
- 信号事象と似た終状態を持ち、信号領域で観測される
- 2つのbジェットの不变質量ピークがZボソン付近
- 理論系統誤差が2番目に大きい
- 解析レベルで正しく評価しきれてない誤差が存在する
- 他背景事象の見積もりの妥当性確認のため、Dibosonの信号強度も測定

既存のヒッグス信号強度測定の誤差

全体誤差	0.177
統計誤差	0.115
系統誤差	0.134

主な実験系統誤差

Jets	0.043
b-tagging (b-jets)	0.045
Luminosity	0.016

主な理論系統誤差

信号事象 $VH(H \rightarrow b\bar{b})$	0.072
Z + jets	0.032
$t\bar{t}$	0.021

Diboson	0.033
MCサンプル統計数	0.031

➔ Diboson 背景事象は正しい系統誤差評価が特に要求される

[Eur. Phys. J. C 81 (2021) 178]



Diboson の理論モデル間の系統誤差の過大評価

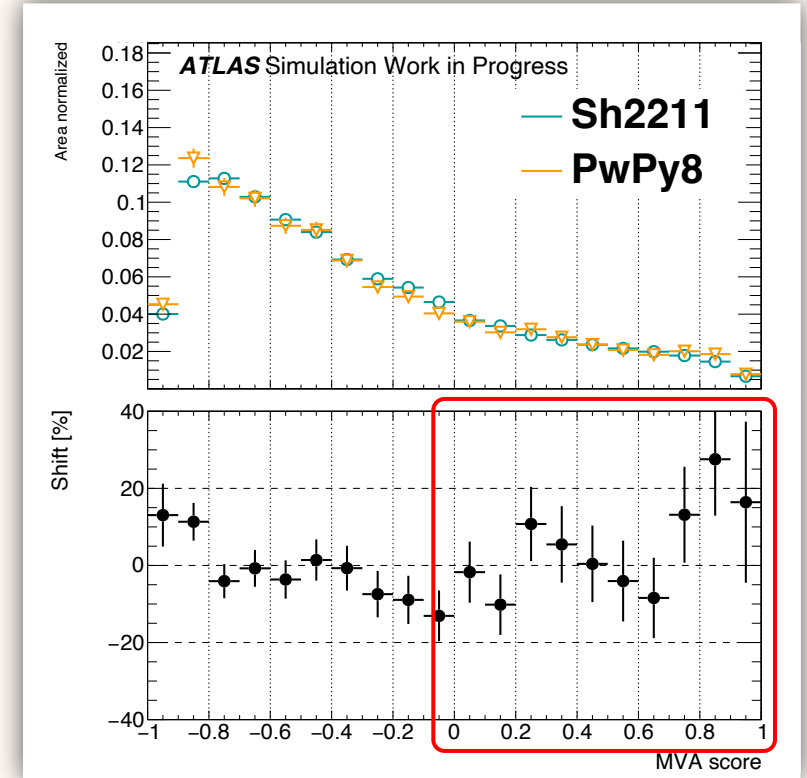
■ 理論モデル間の系統誤差の評価方法

- 理論モデル間の最終分離変数の形の差異を評価
- Diboson の評価に使用する MC シミュレーションサンプル
 - 基準モデル : Sherpa 2.2.11 (サンプル数 : 8.7×10^5)
 - 代替モデル : Powheg+Pythia8 (サンプル数 : 5.6×10^4)
 - ・ Matrix Element や Parton Shower の計算の代替手法用いる
 - ・ 様々な運動学変数で差異が見られる

■ 過大評価の可能性

- コンピュータリソースの制約により代替モデルのサンプルは低統計
- 低統計によりモデル間の差異にふらつき
 - ・ 特に信号感度の高い領域で顕著
- ふらつきによる系統誤差の過大評価を抑えるため、

正しくモデル間の差異を見積もる戦略が必須



モデル間の最終分離変数の差異



理論モデル間の系統誤差評価の戦略と手法改善

■ 基本戦略 :

- 基準モデルサンプル $f_N(\mathbf{x})$ を補正し代替モデルサンプル $f_A(\mathbf{x})$ の形を再現
- イベントごとに推定された重み付け $w(\mathbf{x})$ で補正し再現 ($w(\mathbf{x}) * f_N(\mathbf{x}) \rightarrow f_A(\mathbf{x})$)
- 高統計の代替モデルサンプルとして取り扱い、最終分離変数でモデル間の形の誤差評価

□ [Old] 最終分離変数に一番寄与する一つの運動学変数 (m_{bb}) の差異だけで $w(\mathbf{x})$ を推定 [Eur. Phys. J. C 81 (2021) 178]

□ [New] 寄与する全運動学変数の差異を DNN で学習し次元変数に落とし込んで $w(\mathbf{x})$ を推定
(arXiv:1506.02169v2)

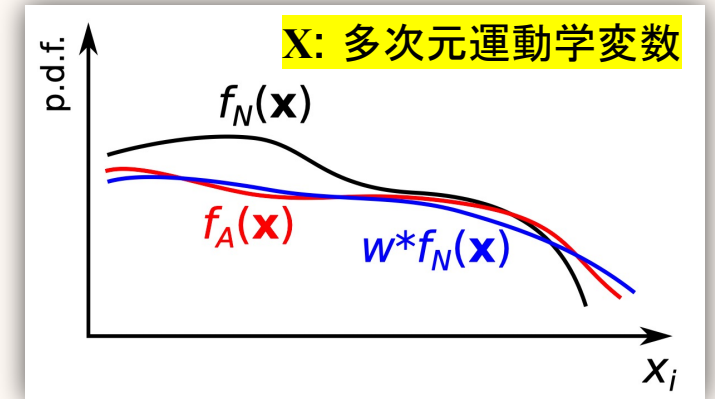
■ [New]実装の目的・理想

- 最終分離変数の入力となる全ての運動学変数の差異を考慮し
代替モデルの変数を再現するための重み付けを算出

■ 困難 : 運動学変数間の相関を正確に取り扱うため、

多次元確率分布関数 で表現した $w(\mathbf{x})$ を用意しモデル間を評価する必要があるが、

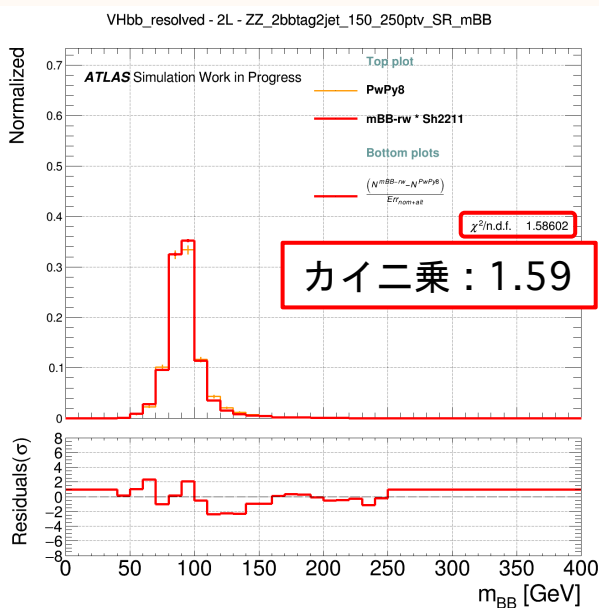
統計量が限られる場合の正確な推定や実装が困難



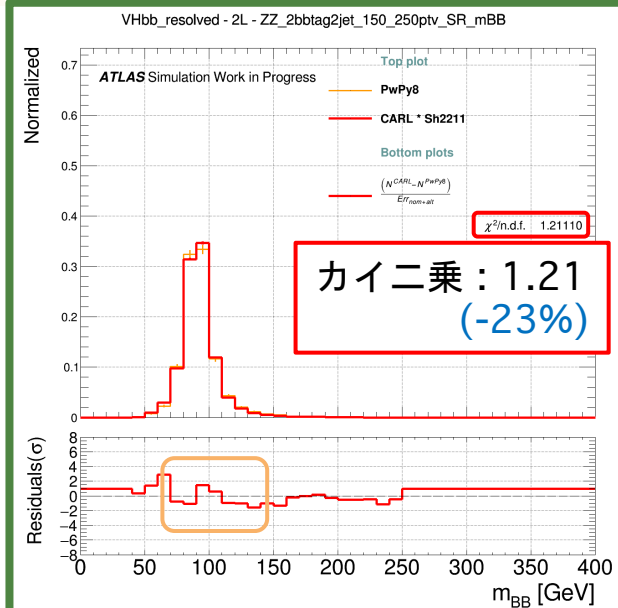
代替モデルの再現精度向上 ~運動学変数~

■ DNN による重み付け手法を使用することで、多くの運動学変数の代替モデルの再現精度が向上

- ・ m_{bb} 以外の運動学変数でも代替モデルの形をよく再現している
- ・ 全体的な分布の一致具合を表すカイ二乗が小さくなる

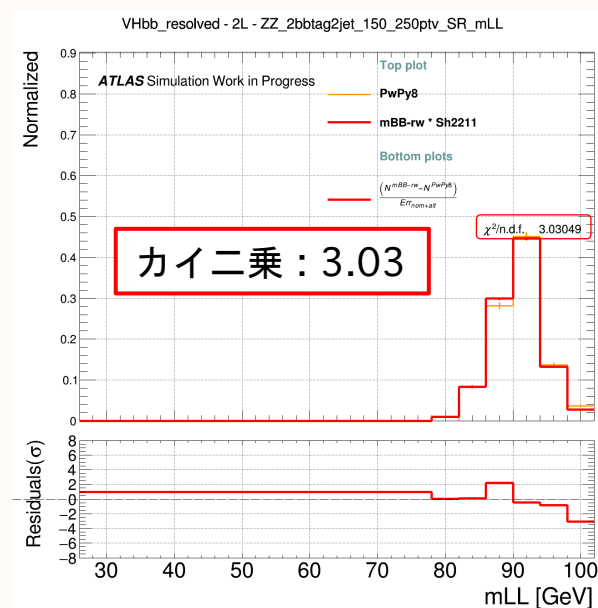


昔の手法 (m_{bb} 重み付け)

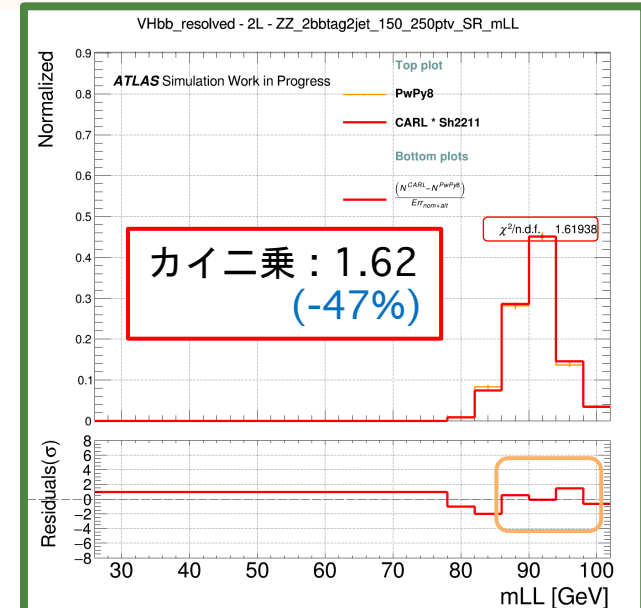


新しい手法 (DNN 重み付け)

2つの b-jet の不変質量 (m_{bb}) 変数分布



昔の手法 (m_{ll} 重み付け)



新しい手法 (DNN 重み付け)

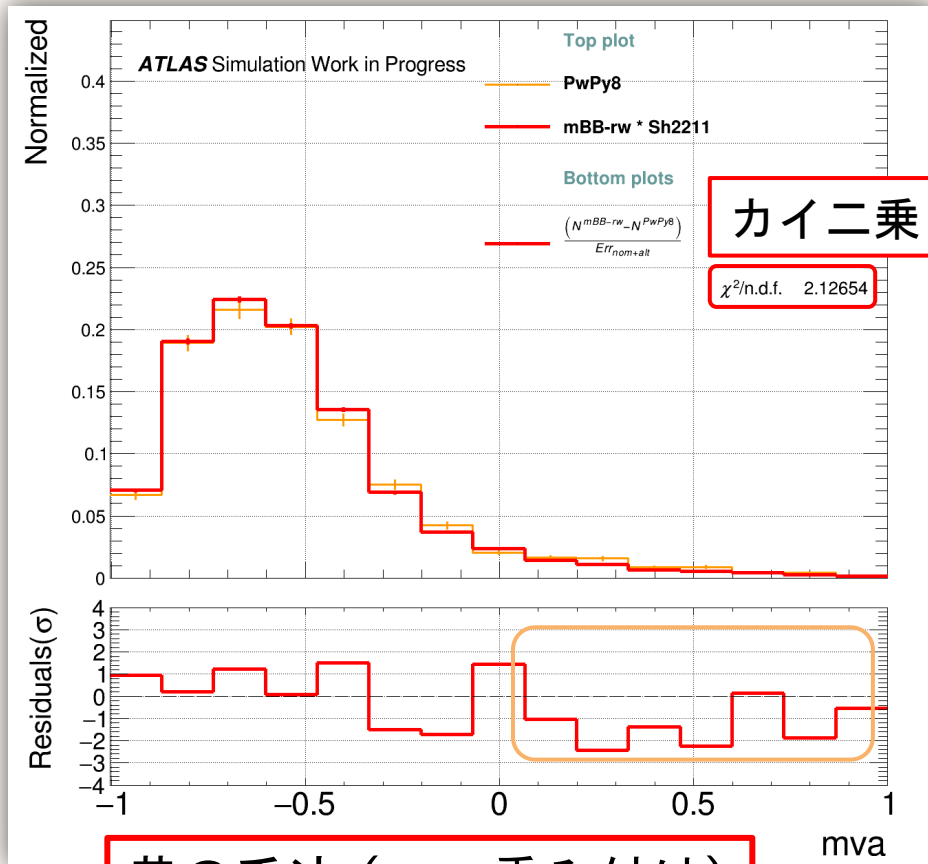
2つのレプトンの不変質量 (m_{ll}) 変数分布



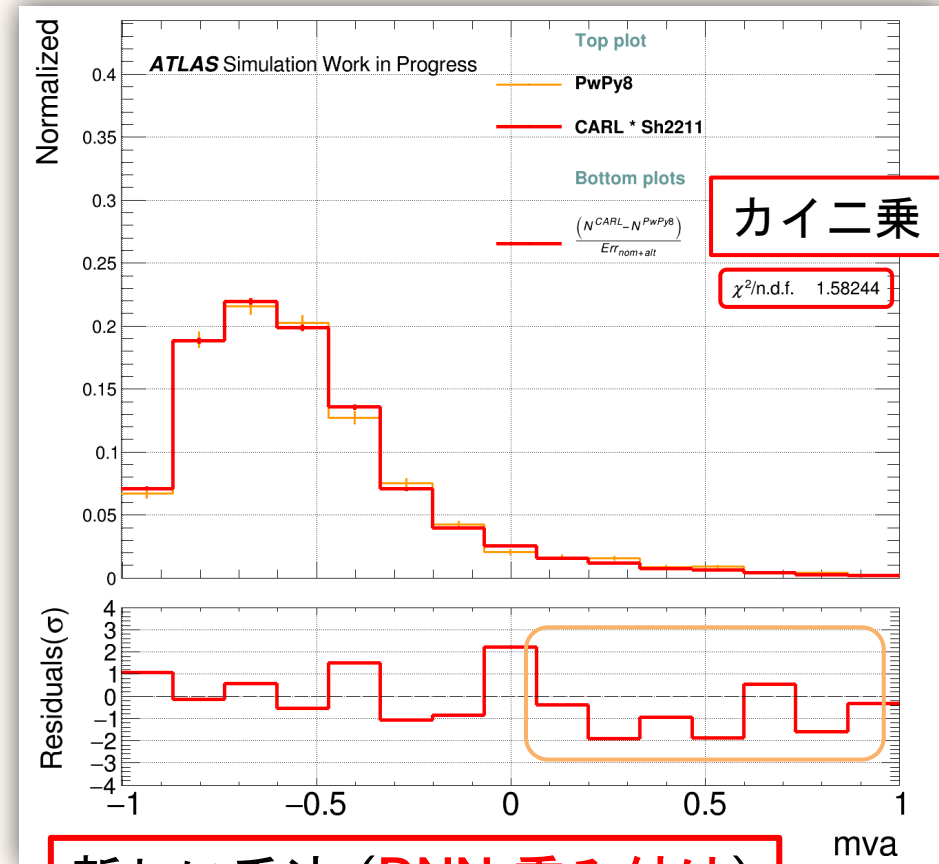
代替モデルの再現精度向上 ~最終分離変数~

■ DNN 重み付け手法を使用することで、最終分離変数の代替モデルの再現精度も向上

- ・ ヒッグス信号強度測定で重要な最終分離変数の値域 [0,1] でも一致
- ・ 全体的な分布の一致具合を表すカイ二乗が小さくなる



昔の手法 (m_{bb} 重み付け)

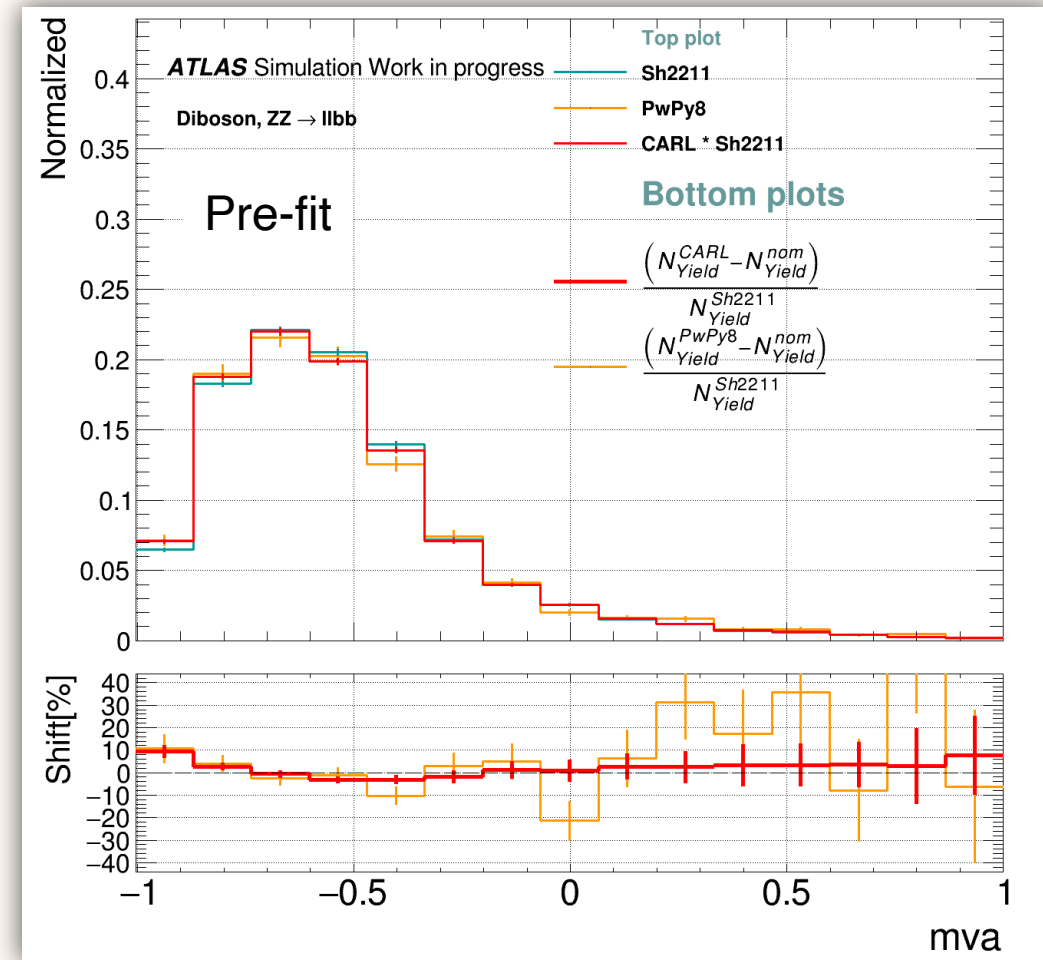


新しい手法 (DNN 重み付け)



実験データとの Fit 時の Diboson 系統誤差評価

- 低統計の代替モデルとの系統誤差を直接見た時より、
重み付けした再現分布で系統誤差を見積もる方が、
ふらつきを抑えた上で系統誤差を評価可能
- 実験データと Fit、ただし信号強度を SM 値に固定 ($\mu=1$)
- 重み付けの手法を改善した結果、
モデル間の差異が小さく見積もられており、
Diboson 系統誤差は大きく減少した
- 背景事象の見積もり手法の妥当性を検証するために、
信号事象と幾何学的性質が似ている
Diboson生成過程の信号強度も測定 (今回はSM値固定)
 - 今後、信号強度を固定せず測定する予定 $\mu = ? \text{ soon...}$



代替モデルとの差異 (オレンジ) と
DNN手法で重み付けした再現分布との差異 (赤)

Diboson 系統誤差	m_{bb} 重み付け	DNN 重み付け
ヒッグス信号強度($\mu=1$)	0.014	0.010 (-29%)
Diboson 信号強度($\mu=1$)	0.123	0.123 (-0%)



まとめと今後の展望

- LHC-ATLAS実験の第2次実験データセットを使用して、ヒッグス粒子の性質の精密測定を行っている
 - 信号事象:ベクターボソンと随伴して生成されたヒッグスから2つのボトム(b)クォーク対へ崩壊する事象
- ヒッグス系の再構成の最適化
 - FSR jet と ISR jet を解析で区別してないため、Truthレベルでの事象分別を解析レベルで行うのは難しい
 - FSR recovery: FSR jet を同定しヒッグス系に組み直し、Truthレベルで正しい解析領域へ分別し直す
 - 領域毎のヒッグス生成微分断面積測定の精度を最大 ~7 % 向上させた
- Diboson生成過程の理論系統誤差の見積もり直し
 - 最終分離変数のモデル間差異を見積もるとき、最も寄与する一変数の差異しか考慮されていない
 - DNN を用いて寄与する全変数の差異を考慮し、モデル間差異を理想的に見積もり直すことができた
 - Diboson の理論系統誤差は ~29 % 改善した
- [Done] Diboson の全断面積理論誤差や解析領域間イベント数誤差等、全ての理論系統誤差を見積もり直した
- [Next] 信号領域の実験データを開示し、信号強度と微分断面積を固定せずに測定する
- [Next] 最終的な物理結果をまとめる

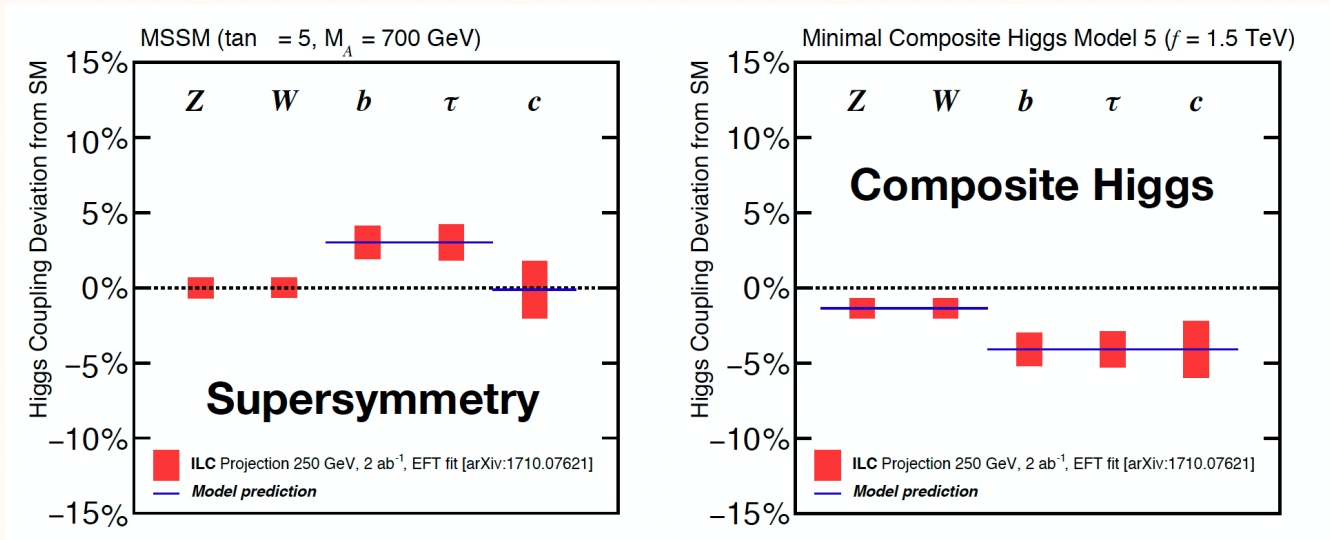


Backup

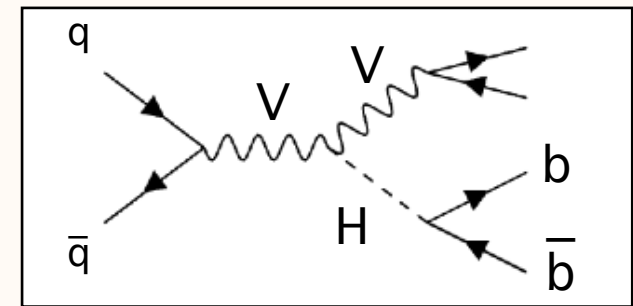


Physics motivation

- $H \rightarrow bb$ Yukawa coupling constant may deviate $\sim O(1)$ % from the SM by the effect of BSM physics model
- The differential cross section of VH production in high momentum region has sensitivity to some BSM physics model with Effective Field Theory interpretation



K. Fujii, et al. arXiv:1710.07621 (2018)



◆ Jet

A hadron containing quarks is observable, where a quark is unable to be observed as a single particle, because of the confinement effect of quantum chromodynamics. As a quark decays with velocity, a lot of hadrons are produced in the flying direction of the quark due to the process of fragmentation (see Fig 4). Then the detector observes a bunch of the hadron showers, what is called “Jet”.

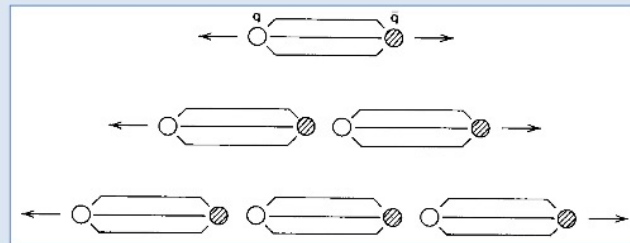


Fig 4, Fragmentation

解析の戦略

ANA-HIGG-2020-20 ([Glance](#))

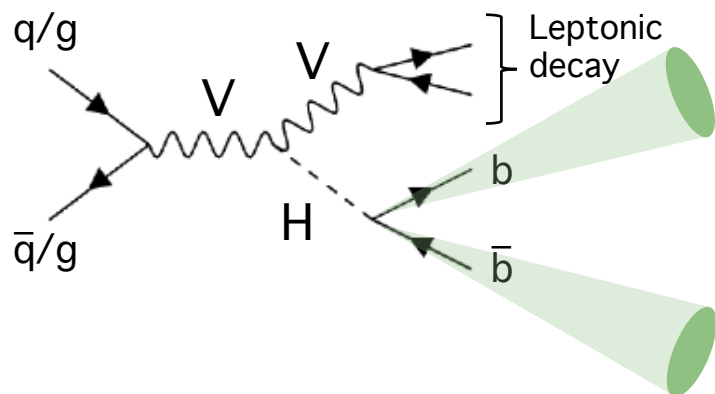
VH(bb)

VH(cc)

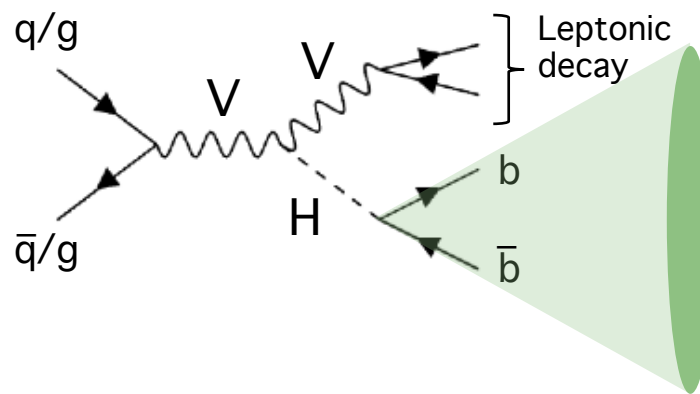
Resolved
75 GeV < pTV < 400 GeV

Boosted
400 GeV < pTV

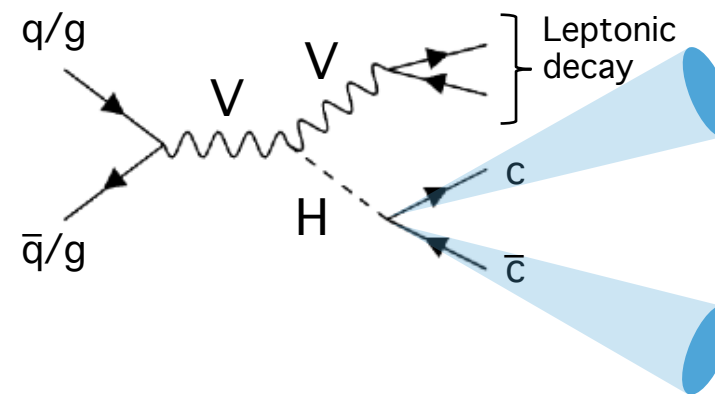
Resolved
75 GeV < pTV



jet R = 0.4



jet R = 1.0



jet R = 0.4

STXS測定

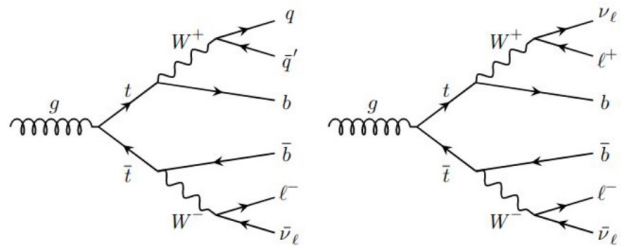
Leptonic decay : $Z \rightarrow \nu\nu$ (0-lepton), $W \rightarrow lv$ (1-lepton), $Z \rightarrow ll$ (2-lepton)
* $l = e, \mu$ (τ in 1-lepton)

* Lepton の分解能が良いことから、pTH の proxy として pTV を使用

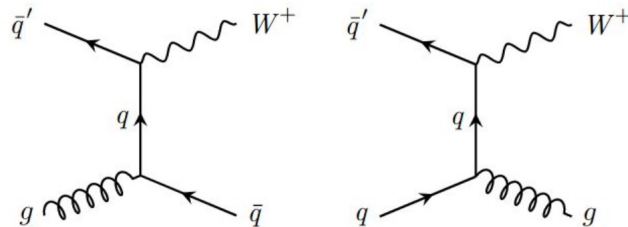


主な背景事象

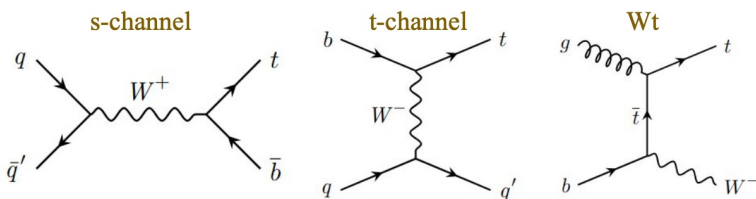
ttbar 0,1L



W+jets 0,1L

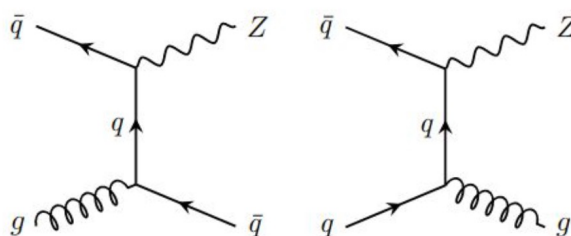


single top 0,1L

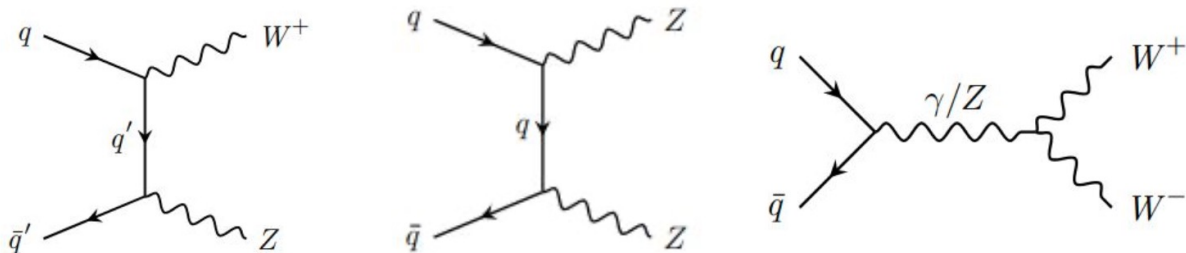


* top in 2L は data-driven

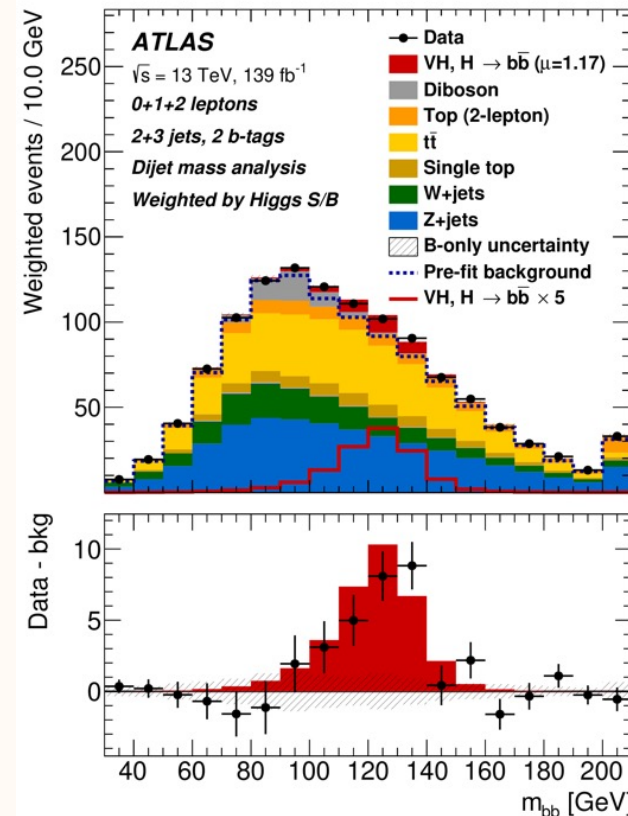
Z+jets 0,2L



Diboson (VZ) 0,1,2L



[arXiv2007.02873v2]

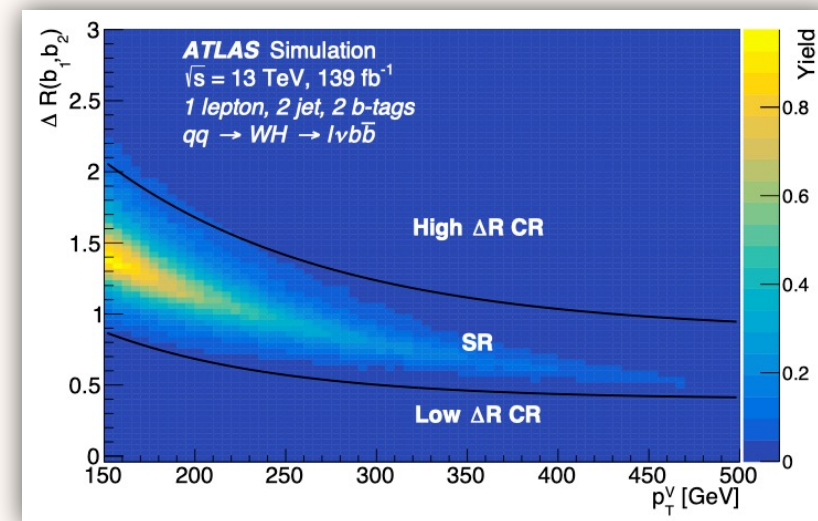
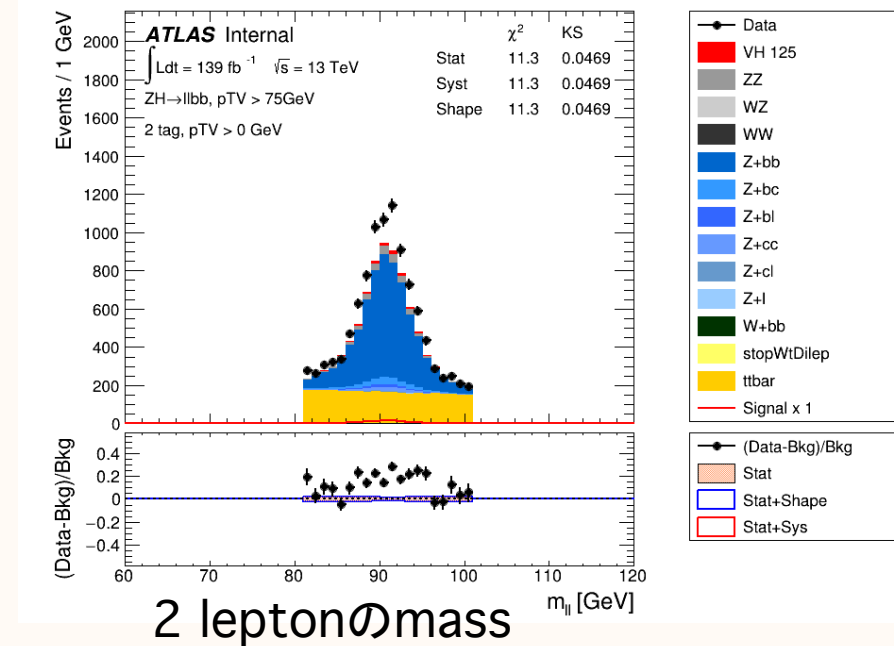


2つのbジェットの
不変質量分布

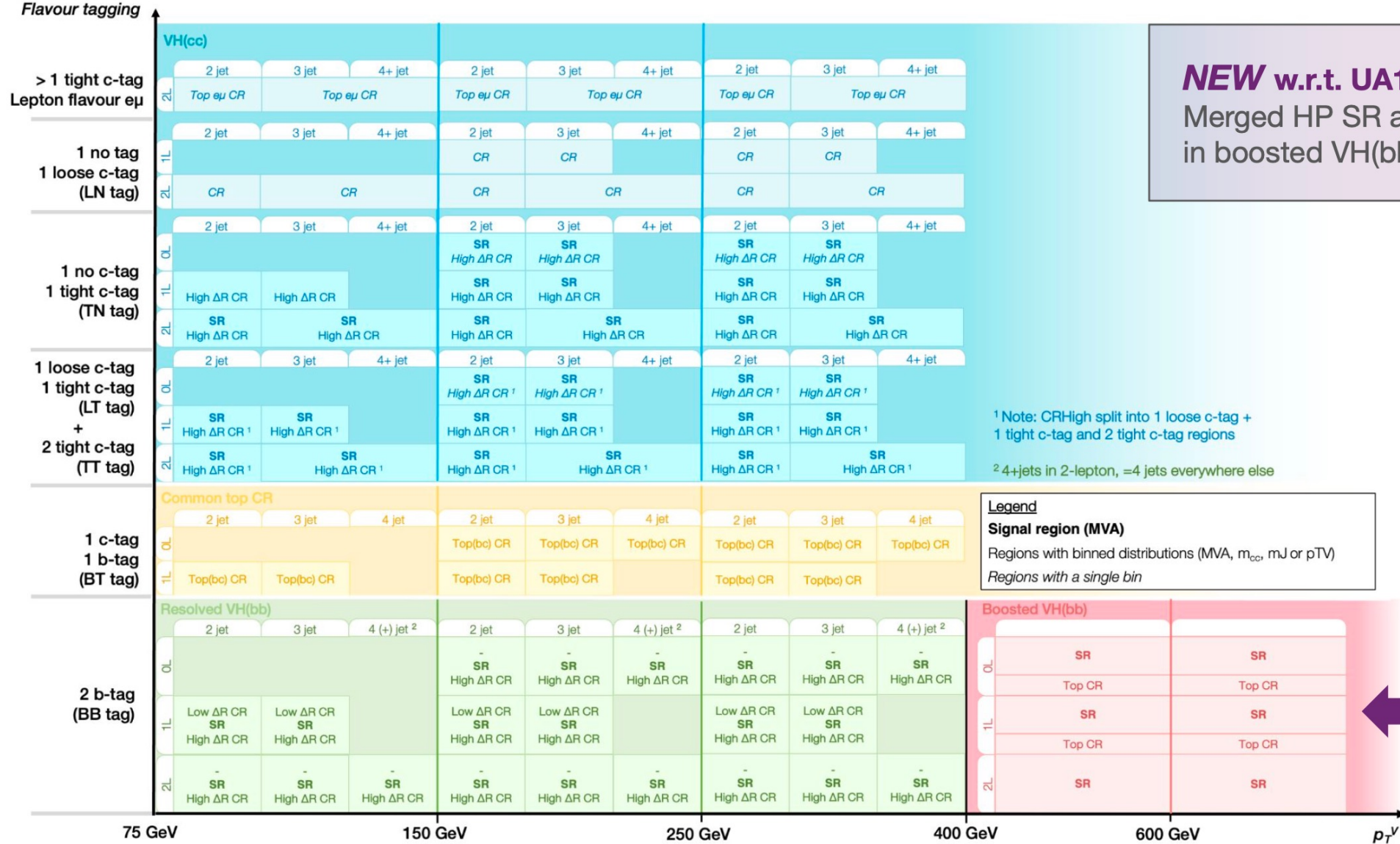


VHbb 解析での事象選択

- Jet :
 - > 2 signal jets
 - 2 b-tagged jets
 - Leading b-jet $p_T > 45$ GeV
 - Sub-leading b-jet $p_T > 20$ GeV
 - Other non-Higgs candidate jet $p_T > 30$ GeV
- Lepton :
 - V から出る2つのlepton
 - 0-lepton: MET triggers
 - 1-lepton: single electron/muon trigger & MET trigger
 - 2-lepton: ee/mumu で 不変質量が m_Z 付近
- Signal領域/Control領域 :
 - $p_{TV}(Z)$ と b-jet間の距離(dR)に従って signalが支配的な領域とBGが支配的な領域にある



Fit region overview



NEW w.r.t. UA1 talk:
Merged HP SR and LP SR
in boosted VH(bb)



最終分離変数

- 決定木 (BDT) を使用
- 信号事象と背景事象を分けるように MC サンプルを使用して学習
- それぞれの解析領域で最適化された運動学的変数が入力変数
- 75 % を学習サンプル, 25 % をテストサンプルとして過学習を防止
- 信号事象を $VH(H \rightarrow bb)$ とするモデルと、信号事象を $VZ(Z \rightarrow bb)$ とするモデルを用意する
- 学習したモデルの出力 score で分離
 - $[-1, 1]$ で、-1 が背景、1 が信号を表現
- Fit では、MC サンプルと実験データの出力 score の合いをみる
 - 学習に使用された MC サンプルも Fit に使用される

Variable	Name	$VH, H \rightarrow b\bar{b}/c\bar{c}$ Resolved			$VH, H \rightarrow b\bar{b}$ Boosted		
		0-lepton	1-lepton	2-lepton	0-lepton	1-lepton	2-lepton
$m_{j_1 j_2}$ or m_J	mBB / mJ	×	×	×	×	×	×
$m_{j_1 j_2 j_3}$	mBBJ	×	×	×			
$p_T^{j_1}$ or $p_T^{j_1, \text{trk}}$	pTB1 / pBTrkJ1	×	×	×	×	×	×
$p_T^{j_2}$ or $p_T^{j_2, \text{trk}}$	pTB2 / pBTrkJ2	×	×	×	×	×	×
$p_T^{j_3, \text{trk}}$	pBTrkJ3				×	×	×
$\sum_{i \neq 1, 2} p_T^i$	sumPtAddJets	×	×	×			
$\Delta R(j_1, j_2)$ or $\Delta R(j_1, \text{trk}, j_2, \text{trk})$	dRBB / deltaRbTrkJbTrkJ	×	×	×	×	×	×
$ \Delta\eta(j_1, j_2) $	dEtaBB	×					
$\text{binDL}_{1r}(j_1)$	bin_btagB1 / bin_bTagBTrkJ1	×	×	×	×	×	×
$\text{binDL}_{1r}(j_2)$	bin_btagB2 / bin_bTagBTrkJ2	×	×	×	×	×	×
p_T^V	pTV	$\equiv E_T^{\text{miss}}$	×	×	$\equiv E_T^{\text{miss}}$	×	×
E_T^{miss}	MET	×	×		×	×	
$E_T^{\text{miss}}/\sqrt{S_T}$	METSig			×			
$ \Delta y(V, H_{\text{cand}}) $	dYVBB/deltaYVJ		×	×		×	×
$ \Delta\phi(V, H_{\text{cand}}) $	dPhiVBB/absdeltaPhiVJ	×	×	×	×	×	×
$\min[\Delta\phi(\ell, b \text{ or } c)]$	dPhiLBmin		×				
m_{eff}	MEff	×					
m_T^W	mTW		×				
m_{top}	Mtop		×				
$m_{\ell\ell}$	mLL			×			
$\cos\theta(\ell^-, Z)$	cosThetaLep			×			×
$(p_T^{\ell_1} - E_T^{\text{miss}})/m_T^W$	lepPtBalance					×	
p_T^ℓ	pTL					×	
$N(\text{track-jets in } J)$	NAdditionalCaloJets				×	×	×
$N(\text{add. small } R\text{-jets})$	NMatchedTrackJetLeadFatJet				×	×	×
Colour	Colour				×	×	×
$\min\{\Delta R(b, j)\}$	minDRBjets	×	×				

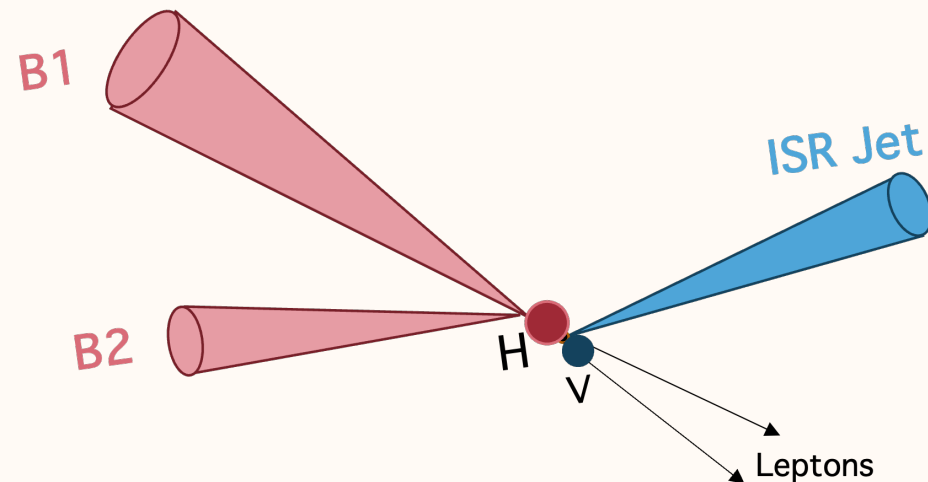
Table 1: MVA variables used for the 0-, 1- and 2-lepton channels in the resolved and boosted topologies for the $VH, H \rightarrow b\bar{b}$ and $VH, H \rightarrow c\bar{c}$ analyses. The $VH, H \rightarrow c\bar{c}$ analysis only probes the resolved topology. The Higgs candidate system, composed of the 2 small- R jets ($b\bar{b}$ or $c\bar{c}$) for the resolved regime or of the leading large- R jet (J) of the event for the boosted regime, is sometimes denoted H_{cand} .

最終分離変数の input 変数一覧

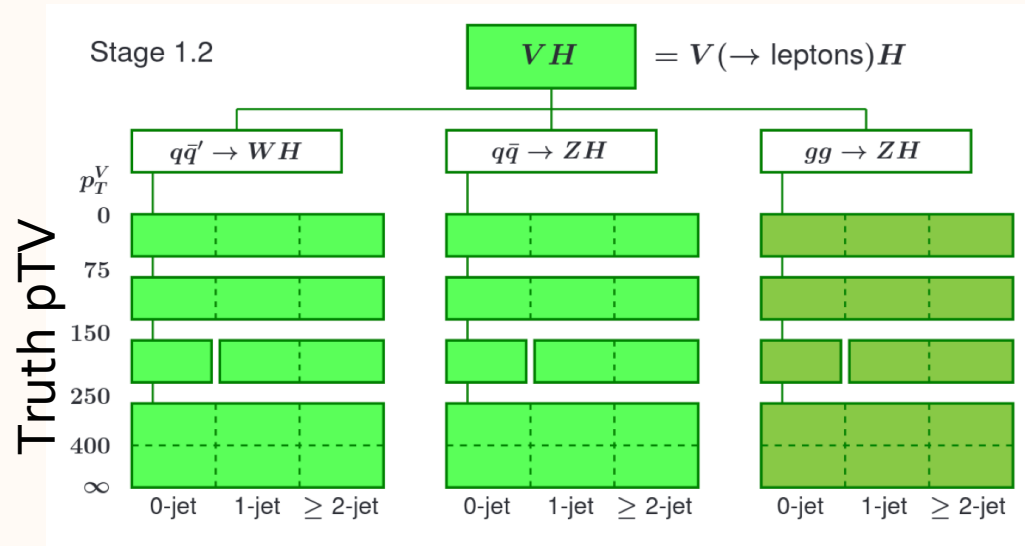


STXS framework と VHbb 解析でのご利益と実装

- Simplified Template Cross-Section (STXS) [Twiki](#)
 - ・ Differential cross-section を精密測定
 - ・ Theoretical dependency を最小化した測定となるよう Truth レベルで領域分けを最適化
 - ・ EFT で解釈し直してBSMの感度を見る測定へと展開



- Truth pTV で bin 分け ([Paper](#))
 - ・ EFT解釈などでHigh pT 領域での BSM に対する感度を持つ
- Truth ISR nJet で bin 分け (**NEW**)
 - ・ QCD scale variation による nJ 間の migration等の違いを考慮せずに断面積測定ができる

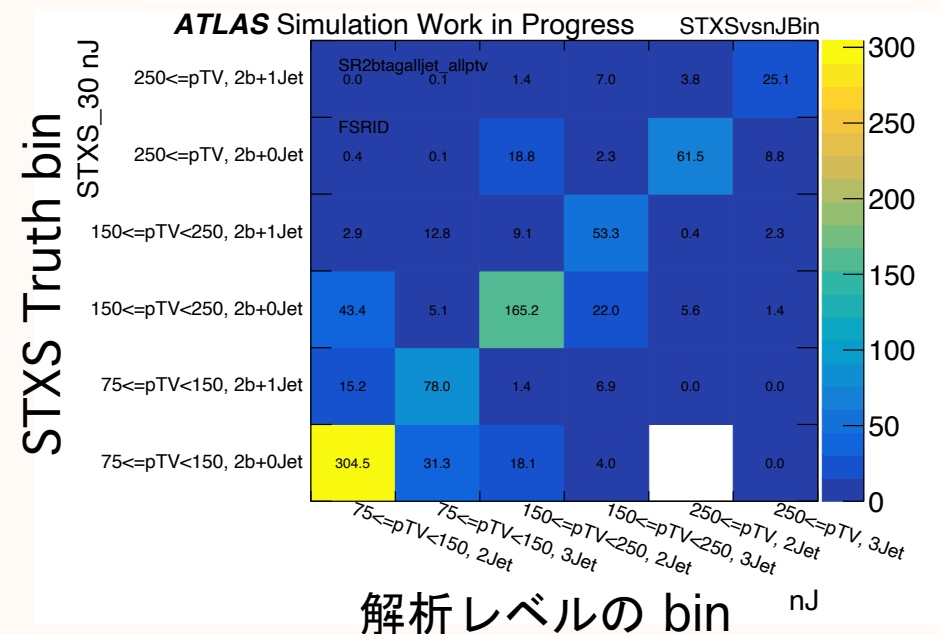
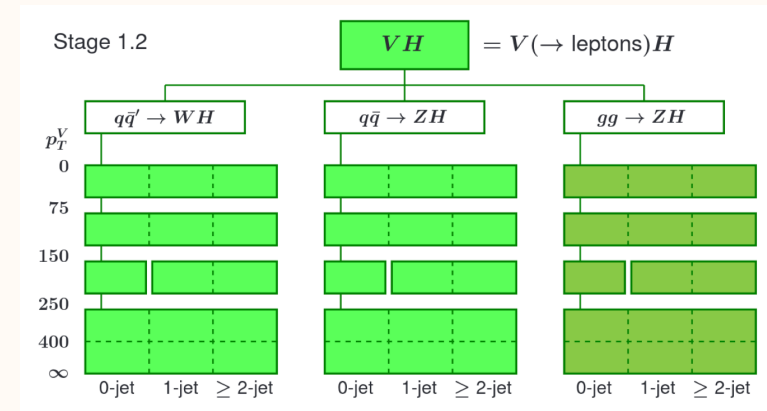


NEW Truth **ISR** nJets



VHbb 解析での STXS 測定の実装

- 解析レベルも同じ bin 分けを適用
- 各 STXS bin の微分断面積の測定方法
 - ・ 解析レベルの各binで同時Fitを行う
 - ・ MC から BG estimation をして、signal の断面積測定
 - ・ MC Template の Migration Matrix で測定した signal を Truth レベルに焼き直す (“Unfold”)
 - ・ MC signal model の中で
解析レベルと Truth レベルが紐づいている

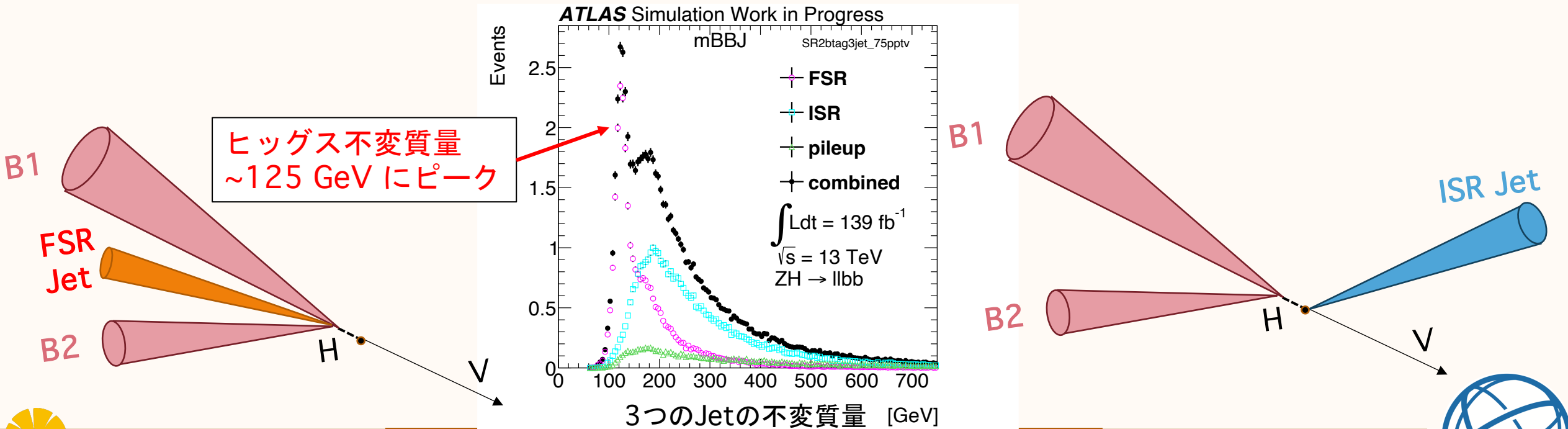


MC signal で定義した migration matrix



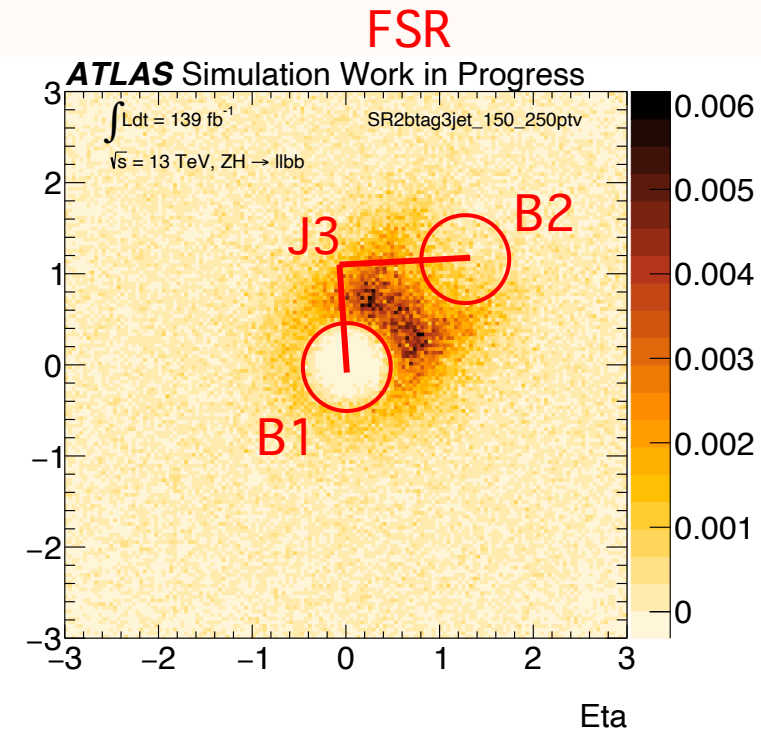
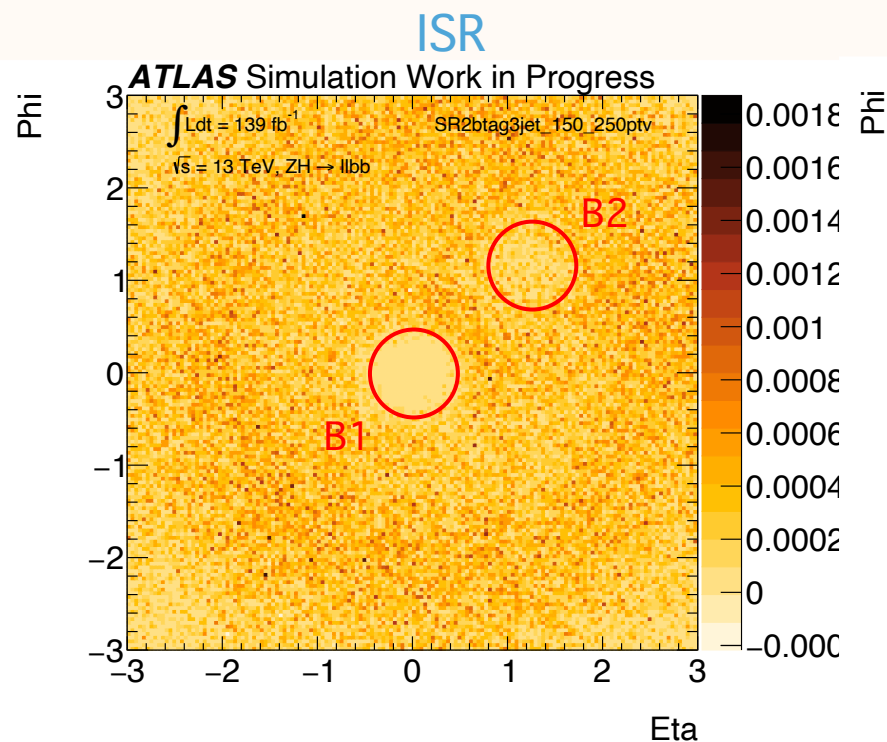
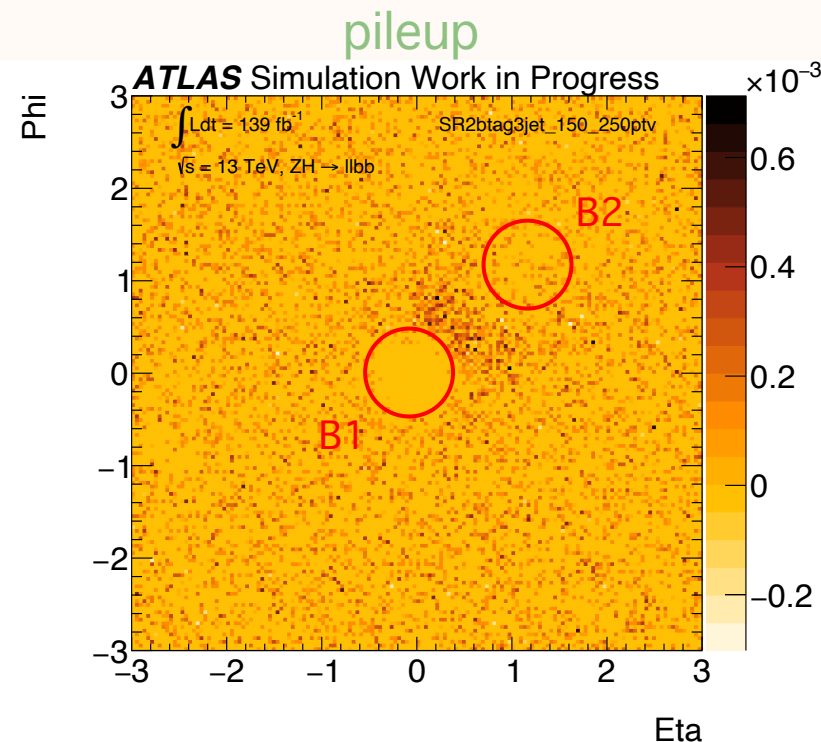
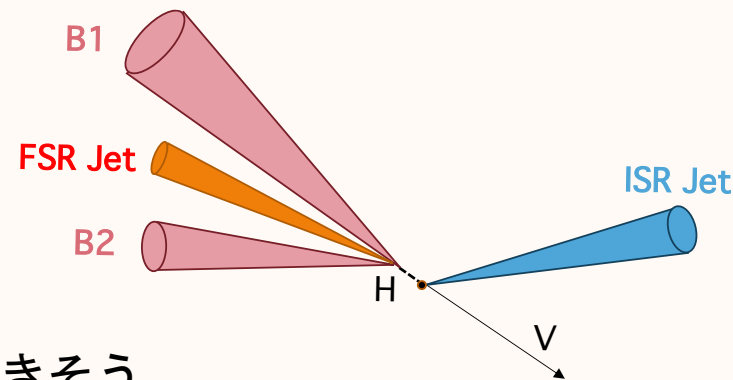
FSR 同定解析の必要性 (1/2) ~ヒッグス不変質量の再構成~

- ヒッグスの性質の測定精度向上には、ヒッグス粒子系の再構成の最適化が必要である
- 現在の解析では、ヒッグス由来の b Jet 以外の Jet は区別無く扱われている
 - Final State Radiation (FSR) ; ヒッグス崩壊由来の終状態輻射
 - Initial State Radiation (ISR) ; ヒッグス以外由来の始状態輻射
 - pileup ; 陽子バンチ衝突時の非弾性衝突による輻射
- ヒッグスの不変質量が **FSR 由来の Jet** 分だけ損失
- 単純な2つの b Jet 系ではなく、**FSR Jet** を含む3つの Jet の系として再構成の最適化が必要



FSR Jet の分離可能な特徴の精査

- 以下の戦略の元、b Jet とそれぞれの Jet の相対位置を見比べた
 - ・ ISR、pileup は 2 b Jet 系とは無関係に発生する
 - ・ FSR は 2 b Jet 系の周辺に発生する
- FSR Jet が 2 b Jet の間に集中して分布している
 - ・ $dR(B1J3) + dR(B2J3)$ が、FSRとそれ以外のジェットの区別に利用できそう



[2bjet と non-b jet の相対位置]

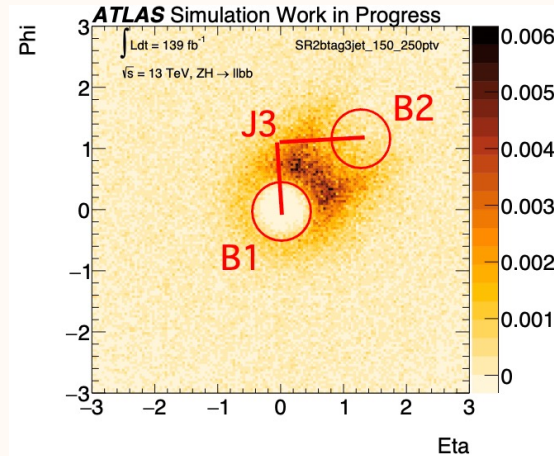


FSR 同定で重要になる変数と同定方法

➤ MCサンプルを再構成して

解析レベルで使用する変数を使用

- $dR(J3B1)+dR(J3B2)$
- ベクターボソンの運動量



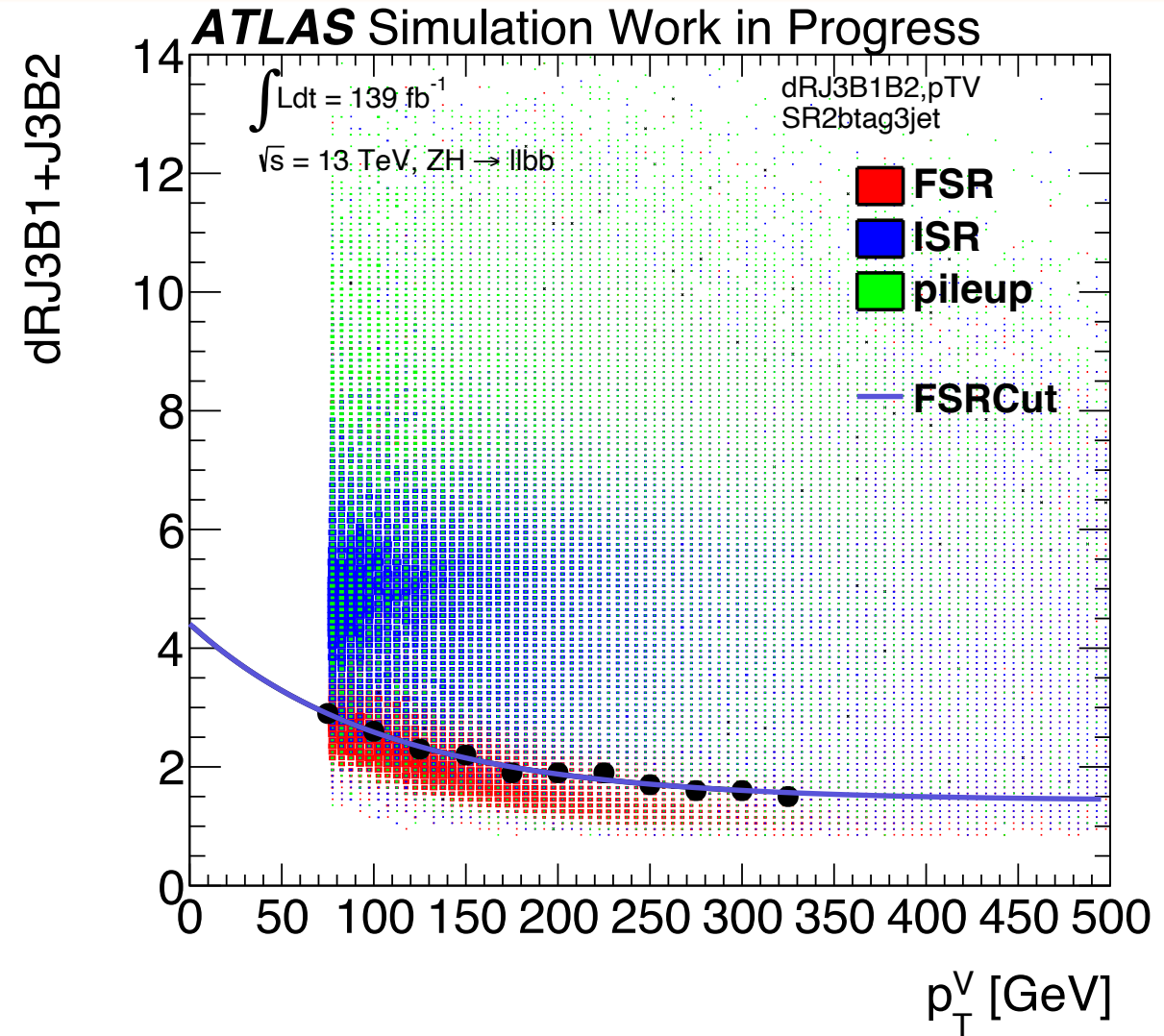
➤ 同定方法

- Cutを掛ける関数を定義
- $FSR / \sqrt{ISR + pileup}$

が最大値になる点でCut関数Fit

- Cut関数以下のdR値となった

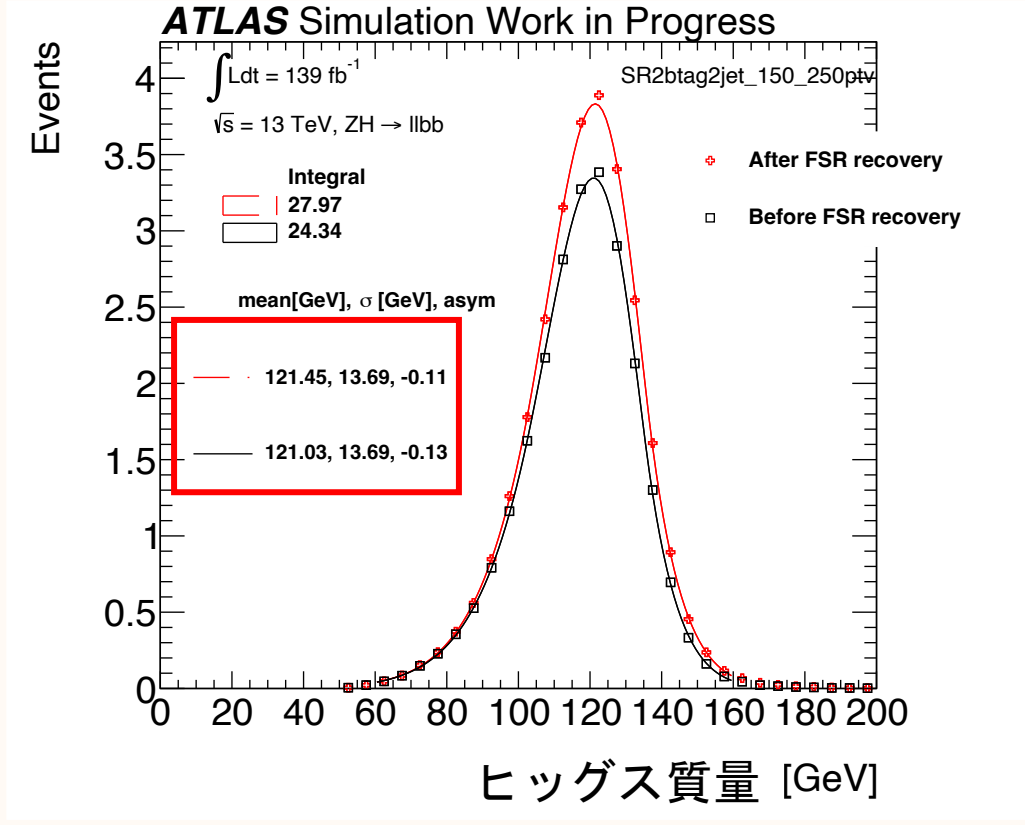
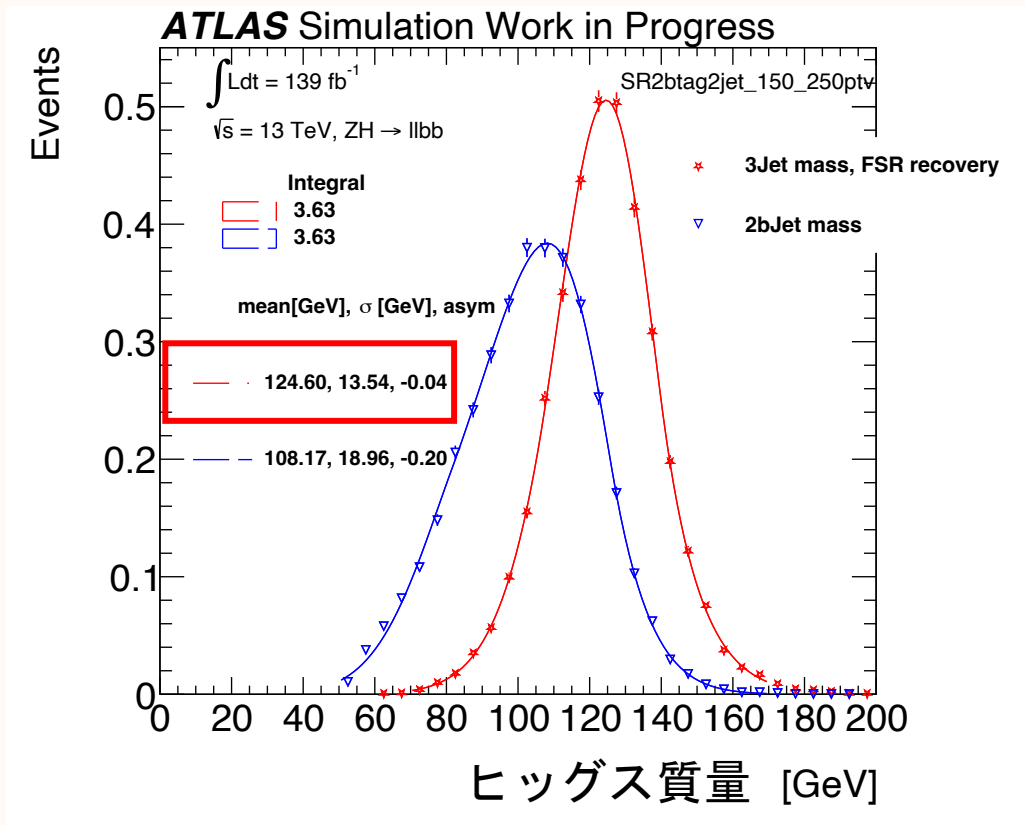
3 jet 事象に対して、FSR recovery の操作をする



Cut関数 : $dR=1.44 + 4.10e-4 * \exp(-1.05e-2 * p_T^V + 8.98)$



① ヒッグス質量分解能の改善 ~2jet~

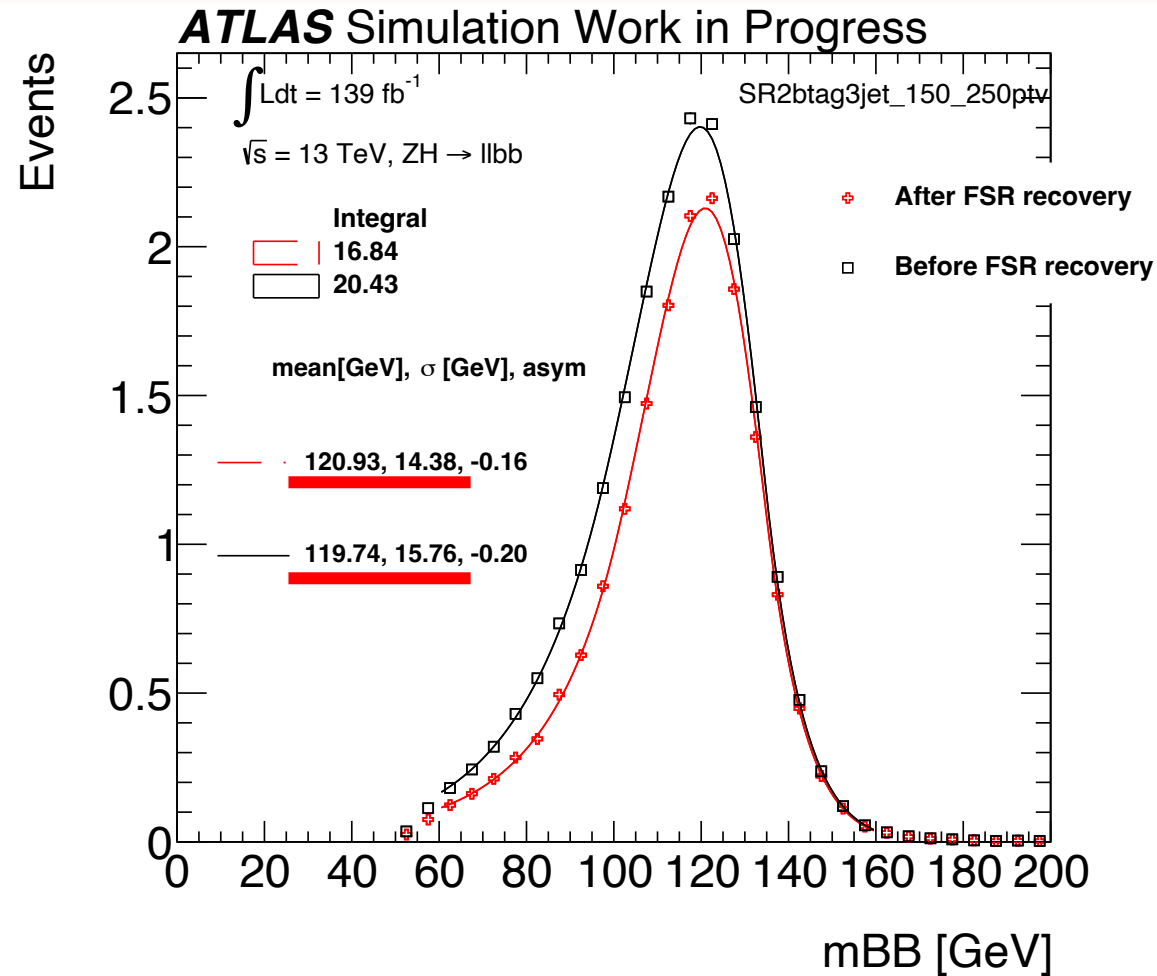


- 3 jet から 2 jet へ移した事象の中で **86 %** と高い割合で正しい FSR を同定した
- FSR recovery を行った事象を 3 jet 系としてヒッグス質量の再構成を行い、
~125 GeVのピークを持つ分布ができた
- FSR recovery を行ったことで、2jet事象全体として**分解能を悪くせず、ピークを125 GeVに近づけた**



① ヒッグス質量分解能の改善 ~3jet~

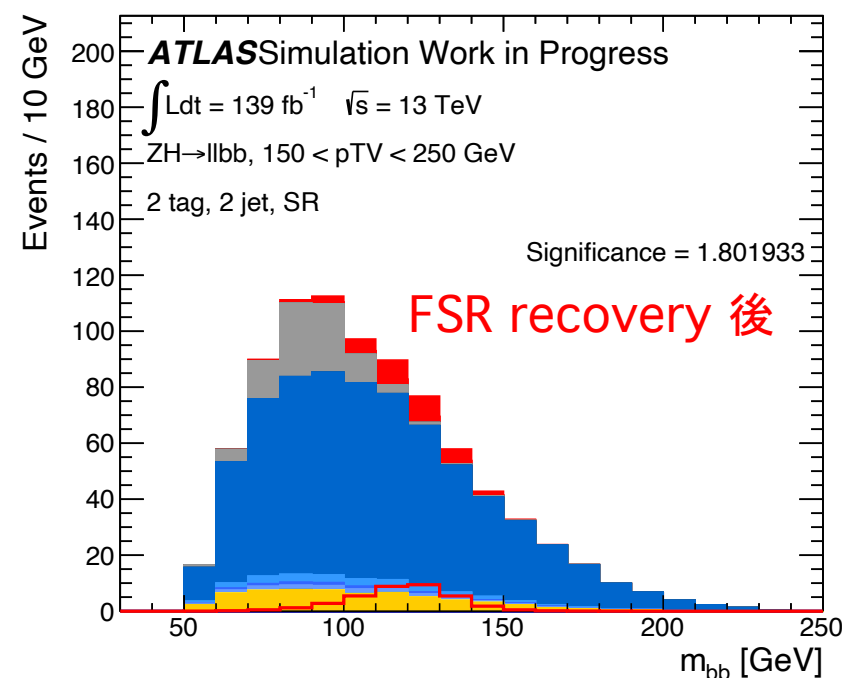
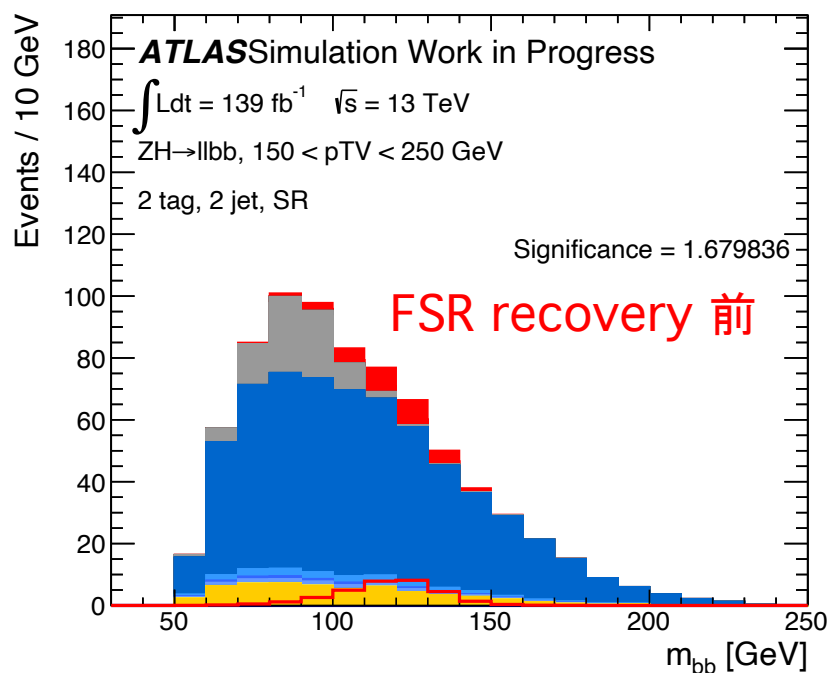
- FSR と同定した Jet を回復させ、改めて3 jet 事象での ヒッグス不変質量分布を計算
- 質量分解能を10%弱改善し、ピークを~125 GeV に近づけた



③ ヒッグス不変質量の信号感度の評価

- 全領域の信号感度は、FSR recovery前と大差は無く、改善は見られなかった
 - ・ ベクターボソン高運動量の解析領域では、3jetの背景事象を多く2jetに移したため、信号感度が少し落ちた
 - 高運動量領域でのFSR Cut関数の最適化が必要
 - ・ 移した背景事象(主にttbar)の分布が125 GeVでピークを持ち、信号感度が上がらなかった
 - 他運動学的変数を追加した FSR 同定 の最適化に挑戦

信号感度	FSR recovery 前	FSR recovery 後
2+3jet, 全運動量領域	3.35	3.33



過去の微分断面積測定結果

[*Eur. Phys. J. C* 81 (2021) 178]

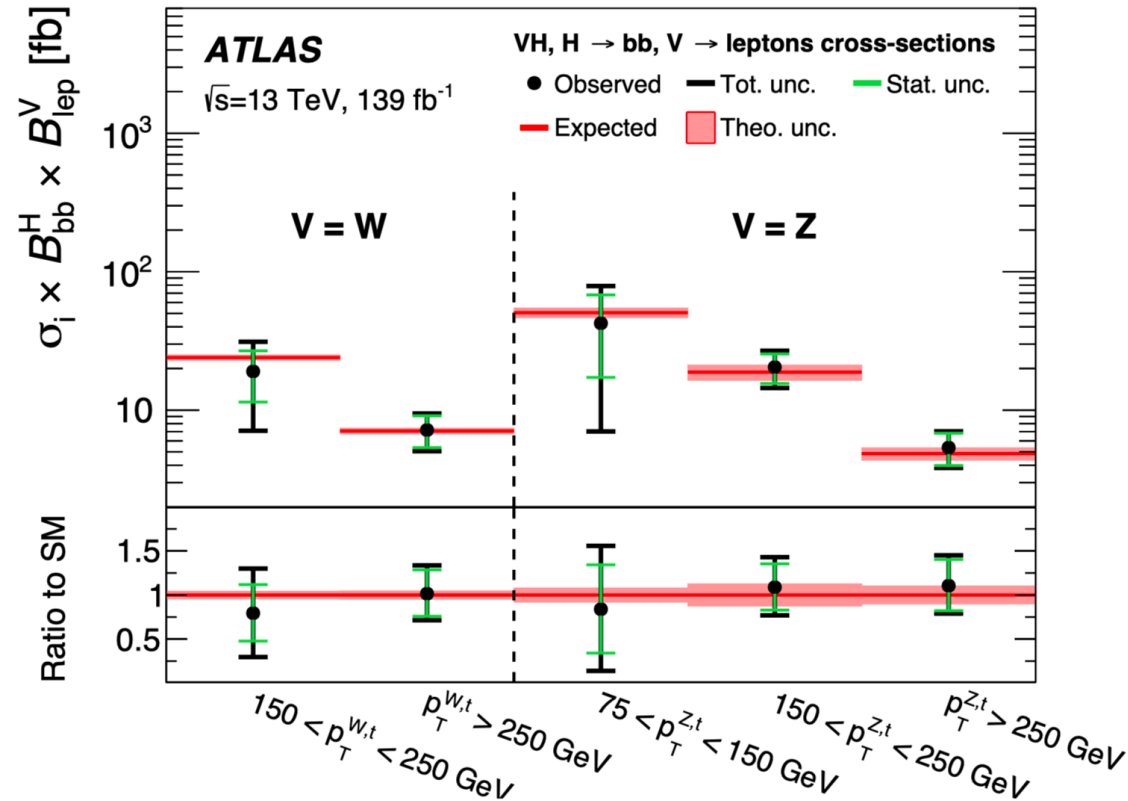


Figure 7: Measured VH , $V \rightarrow$ leptons cross-sections times the $H \rightarrow b\bar{b}$ branching fraction in the reduced STXS scheme.

信号強度測定の詳細

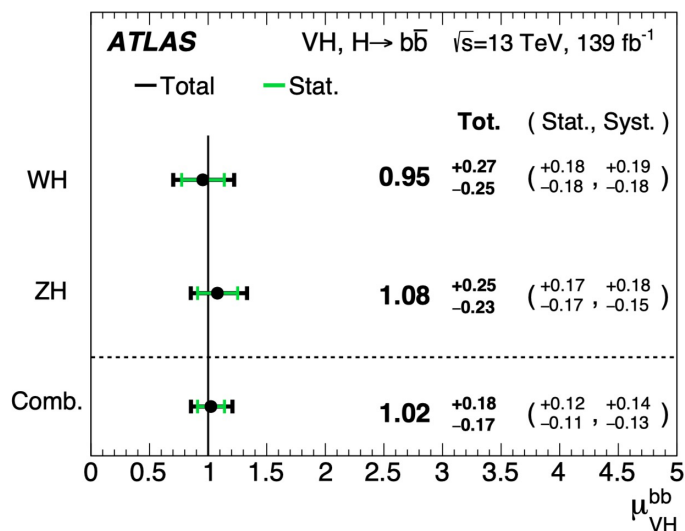


Figure 4: The fitted values of the Higgs boson signal strength μ_{VH}^{bb} for $m_H = 125$ GeV for the WH and ZH processes and their combination. The individual μ_{VH}^{bb} values for the $(W/Z)H$ processes are obtained from a simultaneous fit with the signal strength for each of the WH and ZH processes floating independently. The probability of compatibility of the individual signal strengths is 71%.

[Eur. Phys. J. C 81 (2021) 178]

Source of uncertainty	σ_μ		
	VH	WH	ZH
Total	0.177	0.260	0.240
Statistical	0.115	0.182	0.171
Systematic	0.134	0.186	0.168
Statistical uncertainties			
Data statistical	0.108	0.171	0.157
$t\bar{t} e\mu$ control region	0.014	0.003	0.026
Floating normalisations	0.034	0.061	0.045
Experimental uncertainties			
Jets	0.043	0.050	0.057
E_T^{miss}	0.015	0.045	0.013
Leptons	0.004	0.015	0.005
b -tagging	b -jets	0.045	0.025
	c -jets	0.035	0.068
	light-flavour jets	0.009	0.004
Pile-up	0.003	0.002	0.007
Luminosity	0.016	0.016	0.016
Theoretical and modelling uncertainties			
Signal	0.072	0.060	0.107
Z + jets	0.032	0.013	0.059
W + jets	0.040	0.079	0.009
$t\bar{t}$	0.021	0.046	0.029
Single top quark	0.019	0.048	0.015
Diboson	0.033	0.033	0.039
Multi-jet	0.005	0.017	0.005
MC statistical	0.031	0.055	0.038



Preliminary な最新の信号強度測定の影響

Group of NPs	Impact
Total	± 0.127
Data Stat	± 0.095
Data Stat only	± 0.088
Floating norm	± 0.029
Top $e\mu$ CR stat	± 0.011
Full Systematics	± 0.085
VH modelling	± 0.021
Background modelling	± 0.068
W+jets	± 0.036
Z+jets	± 0.036
ttbar	± 0.011
Single top	± 0.008
Multi-jet	± 0.007
Diboson	± 0.020

Group of NPs	Impact
Experimental systematics	± 0.035
Lepton	± 0.004
MET	± 0.007
JET	± 0.026
Pile-up	± 0.005
Lumi	± 0.007
FTAG PFlow (b-jet)	± 0.015
FTAG PFlow (c-jet)	± 0.008
FTAG PFlow (l-jet)	± 0.003
FTAG VR (b-jet)	± 0.004
FTAG VR (c-jet)	± 0.001
FTAG VR (l-jet)	± 0.001
MC Stat	± 0.020



Diboson 系統誤差全体像

■ Normalization uncertainty

- ・ 全解析領域における、
理論モデルやイベント生成機のパラメータの違いによるイベント数の誤差を評価

■ Acceptance uncertainty

- ・ 解析領域間のイベント数の違いを予め計算し、
ある背景事象が支配的な制御領域から、多くの事象が混在する信号領域での振る舞いを推定
- ・ 理論モデルやイベント生成機のパラメータの違いによる推定の誤差を評価
- ・ Jet number bins / pTV bins / lepton channels

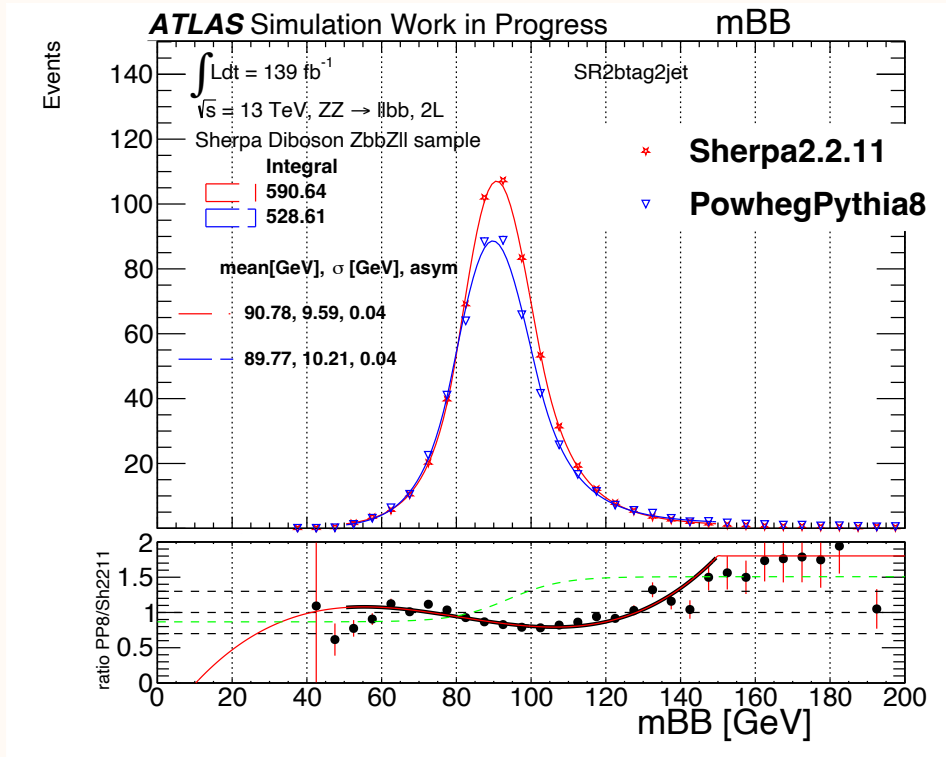
■ Shape uncertainty (当発表で議論しているモデル間系統誤差)

- ・ CARL / mBB reweight

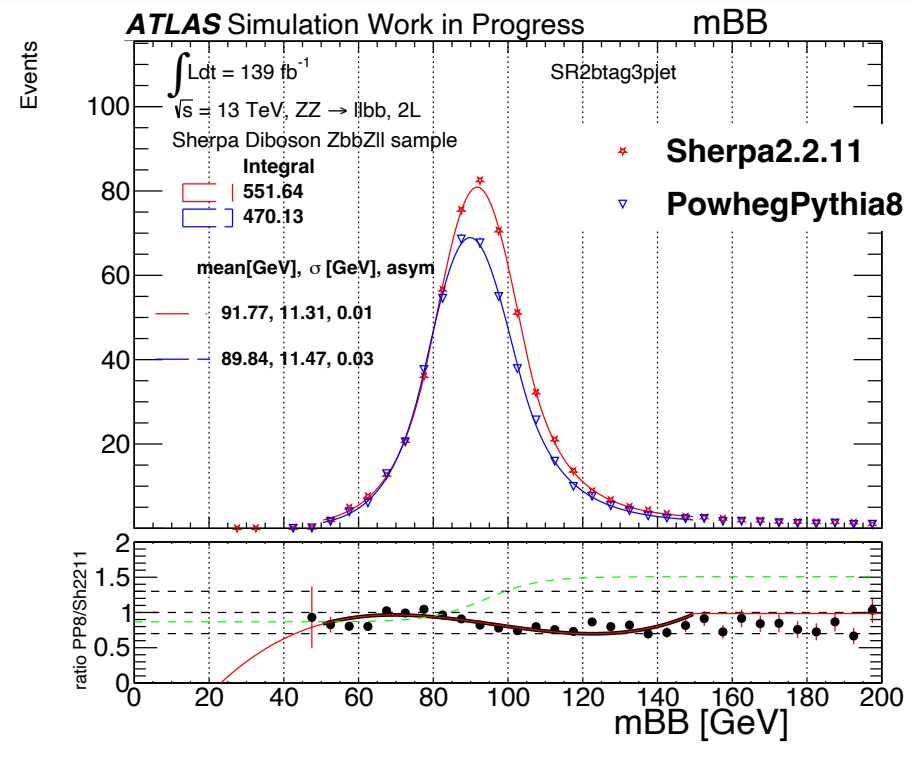


mBB のみを使った重み付け

- ・ mBB shape の 基準モデルと代替モデルの差を関数化
 - ・ 50 – 150 GeV で関数化し、他の領域は定数化
 - ・ 3次関数: (2jet) $y = -6.4e-1 + 7.7e-2 x - 1.1e-3 x^2 + 4.4e-6 x^3$
 - ・ イベント数の違いも考慮して関数化
 - ・ CARL でも、各イベントの重み(イベント数差に対応)も考慮しモデル間の差を training
- ・ 算出した関数に従って、最終分離変数 (mva, mvadiboson) や他の入力運動学変数 に重み付け



2jet 領域



3jet 領域



深層学習を導入した新しい重み付け手法 (CARL)

■ CARL の導入

- 全運動学変数を一次元変数に落とし込むDNN model を準備
 - 全ての運動学変数を学習し、基準モデルと代替モデルの二値分類を出力
- 重み付け $w(\mathbf{x})$ はベイズの定理により、ある運動学変数の代替モデルらしさから近似
- DNNの出力結果(代替モデルであるらしい確率: $D(\mathbf{x})$)から、近似的重み付け $w^*(\mathbf{x})$ を表記

$$w(\mathbf{x}) = \frac{f(\mathbf{x} | y = 1) dx}{f(\mathbf{x} | y = 0) dx} = \frac{\mathcal{P}(\mathbf{x} | y = 1)}{\mathcal{P}(\mathbf{x} | y = 0)} = \frac{\mathcal{P}(y = 1 | \mathbf{x})}{\mathcal{P}(y = 0 | \mathbf{x})} = \frac{\mathcal{P}(y = 1 | \mathbf{x})}{1 - \mathcal{P}(y = 1 | \mathbf{x})} \approx \frac{D(\mathbf{x})}{1 - D(\mathbf{x})} \equiv w^*(\mathbf{x}) \quad (\text{arXiv:1506.02169v2})$$

→ 多次元Likelihood で表現する $w(\mathbf{x})$ の導入を回避

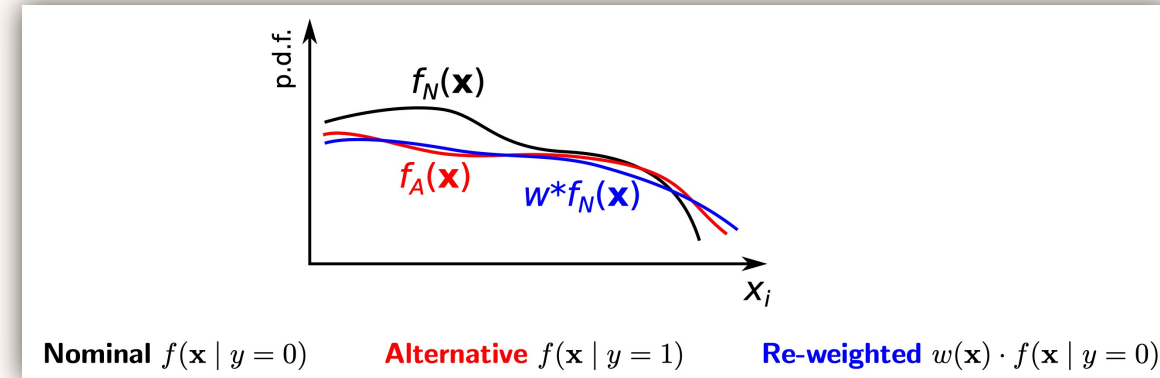
→ 変数間の相関を含んだ形で、基準モデルサンプルから代替モデルを再現する重み付けをイベントごとに推定できるようになった

■ 工夫

- DNN の訓練で、変数間の相関を学ぶための最適化
 - ・ 領域毎の訓練、データセットを変更しての訓練

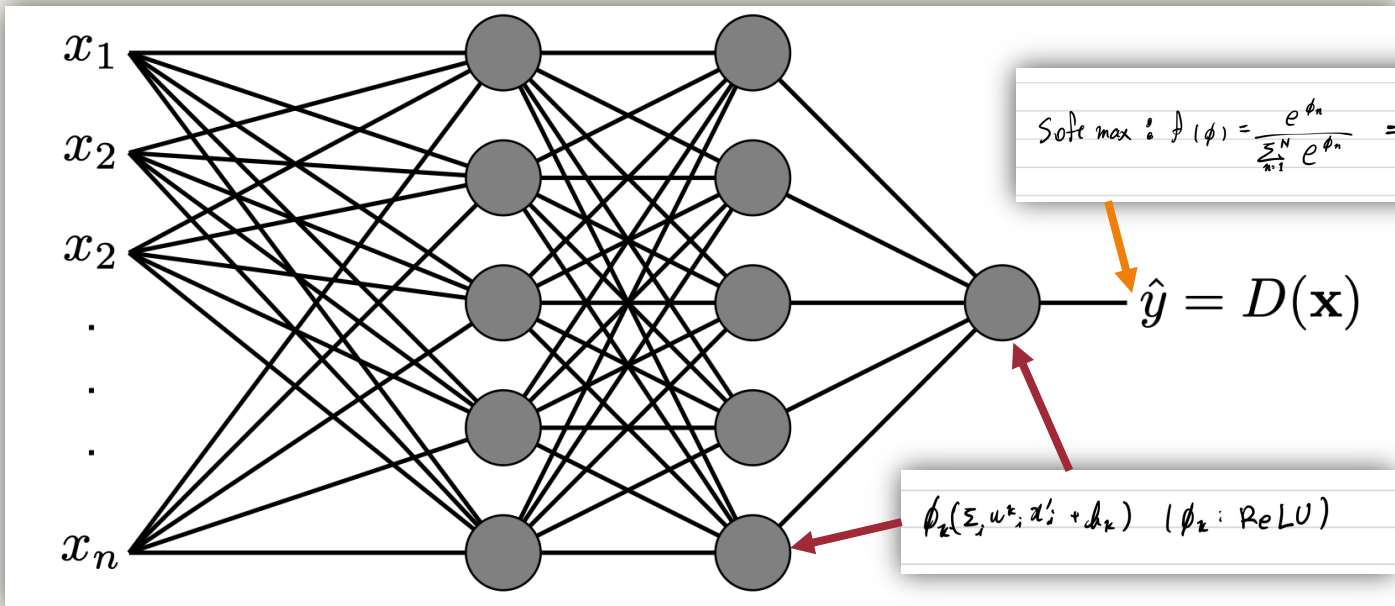
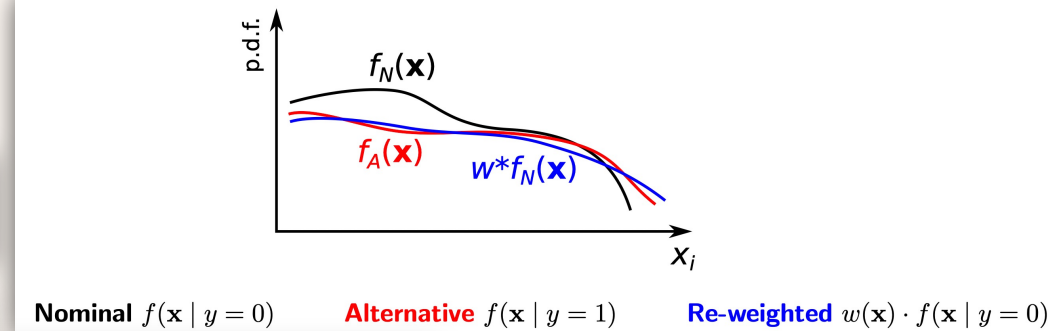
■ 想定されるご利益

- 代替モデルの最終分離変数の分布の再現精度が向上
- 低統計による分布のふらつきを抑制して、正確なモデル間の差異を評価



CARL 詳細 (arXiv:1506.02169v2)

$$w(\mathbf{x}) = \frac{f(\mathbf{x} | y = 1) dx}{f(\mathbf{x} | y = 0) dx} = \frac{\mathcal{P}(\mathbf{x} | y = 1)}{\mathcal{P}(\mathbf{x} | y = 0)} = \frac{\mathcal{P}(y = 1 | \mathbf{x})}{\mathcal{P}(y = 0 | \mathbf{x})} = \frac{\mathcal{P}(y = 1 | \mathbf{x})}{1 - \mathcal{P}(y = 1 | \mathbf{x})} \approx \frac{D(\mathbf{x})}{1 - D(\mathbf{x})} \equiv w^*(\mathbf{x})$$



Soft max: $f(\phi) = \frac{e^{\phi_n}}{\sum_{n=1}^N e^{\phi_n}} = D(\mathbf{x})$

The small graph shows a sigmoid function $f(\phi)$ on the y-axis against ϕ_n on the x-axis, ranging from 0 to 1.

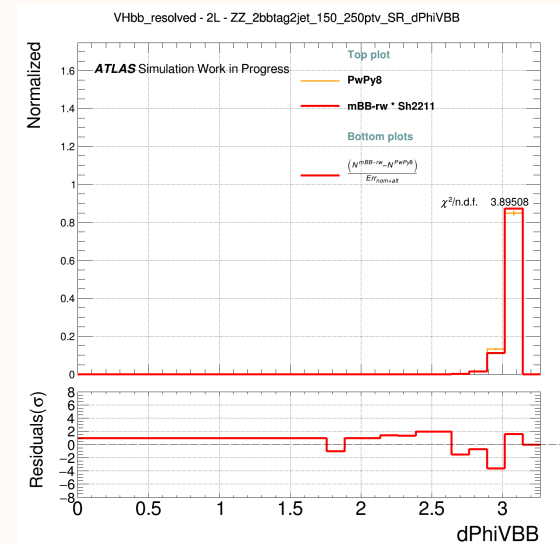
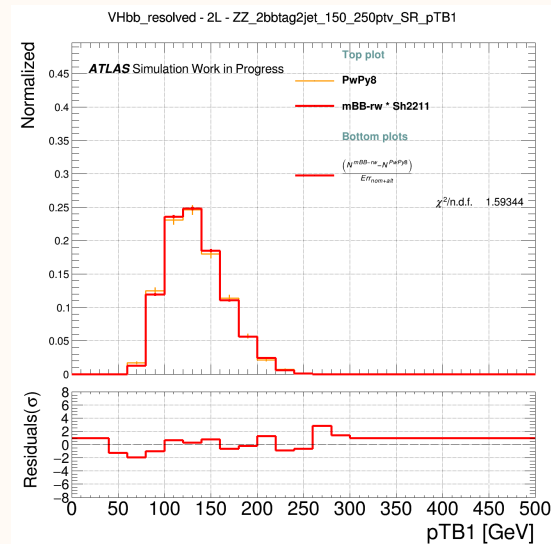
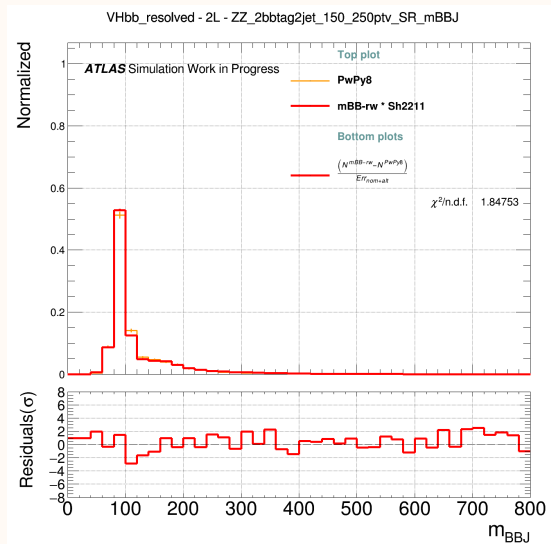
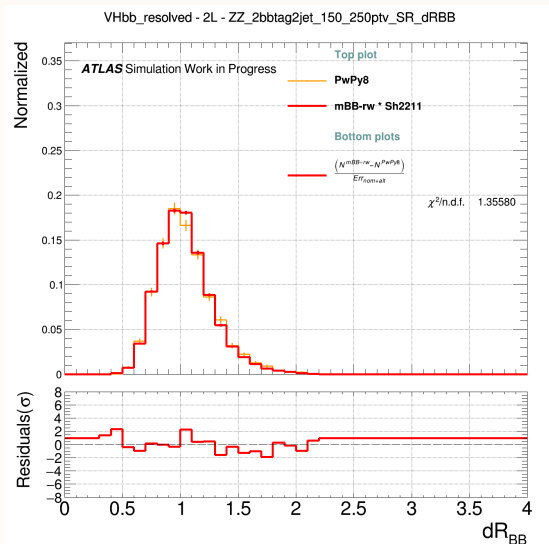
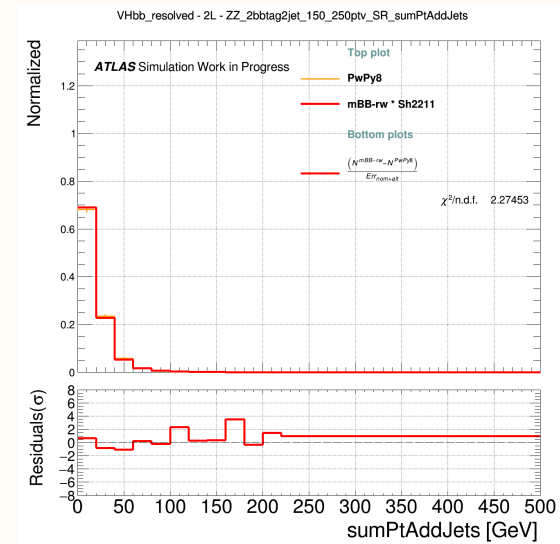
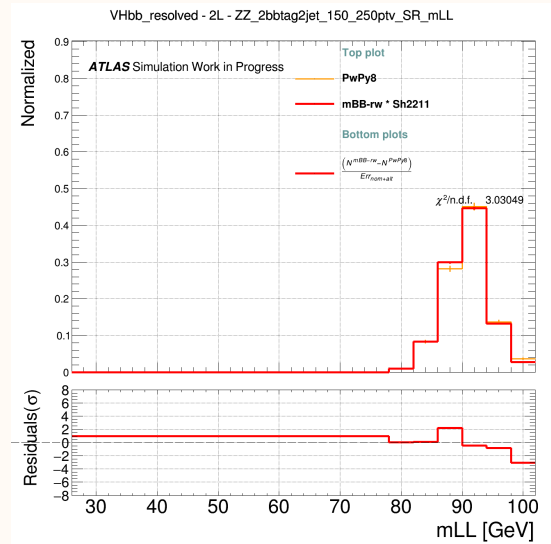
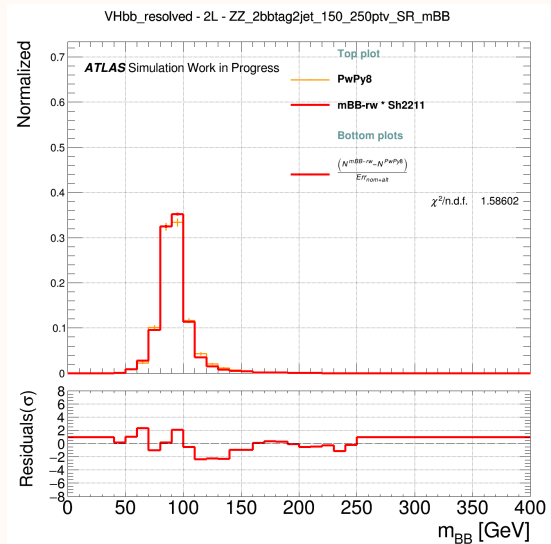
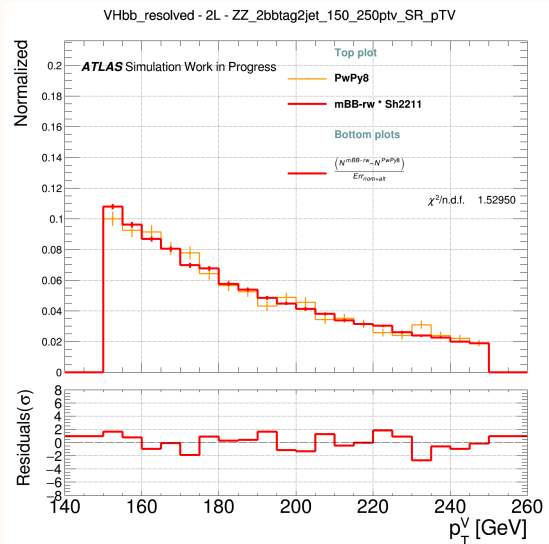
Likelihood function.

$$\begin{aligned} \mathcal{L}_N &= -\ln \left(\prod_{n=0}^N \mathcal{P}(y_n | x_n) \right) \\ &= -\ln \left(\prod_{n=0}^N D(x_n)^{y_n} (1 - D(x_n))^{1 - y_n} \right) \\ &= -\sum_{n=0}^N (y_n \ln D(x_n) + (1 - y_n) \ln (1 - D(x_n))) \end{aligned}$$

- Bayes' theorem により \mathbf{x} 条件下での $y=1$ の確率で $w(\mathbf{x})$ を表現できるため、下記の $D(\mathbf{x})$ を求めるだけで近似的な $w(\mathbf{x})$ を推定できる
- 単純な多層パーセプトロン (MLP) で 基準モデル vs 代替モデル の二値分類を実施
- 非線形活性化関数：隠れ層は ReLU関数、 $D(\mathbf{x})$ の出力関数は softmax関数
- Meta-info: PyTorch, input [0,1] minmax, epoch 300, 正則化(L1,L2)無し
- Softmax関数 [0,1] により、**代替モデルらしさを表す確率 $D(\mathbf{x})$** を近似的に表現
- Likelihoodが最大となる DNN model で、 $P(y=1|\mathbf{x}) \sim D(\mathbf{x})$

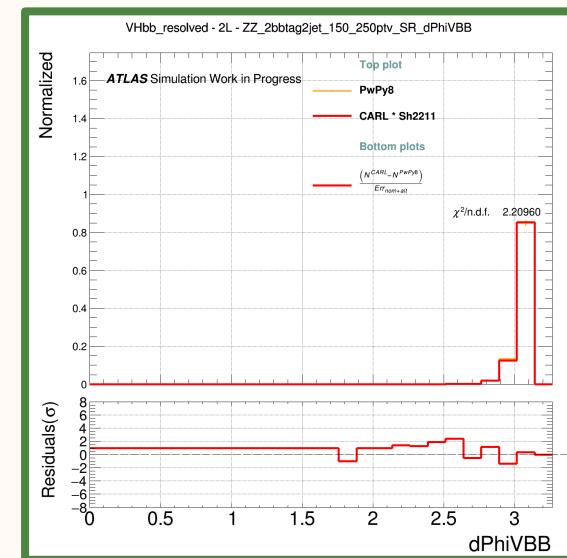
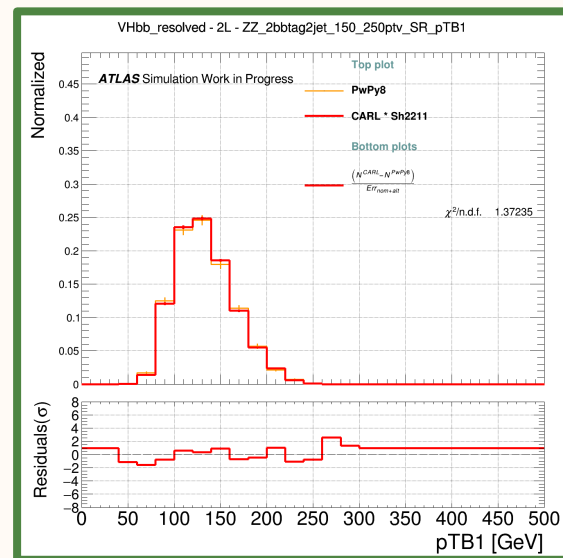
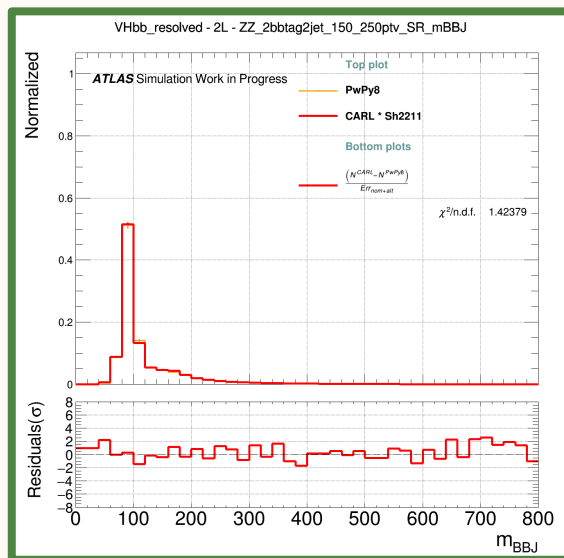
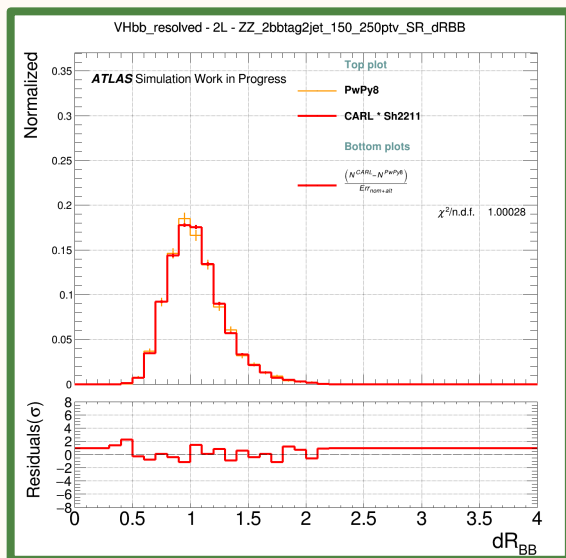
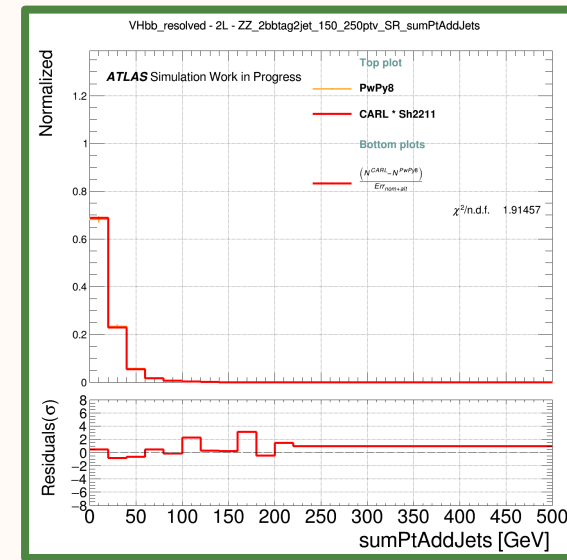
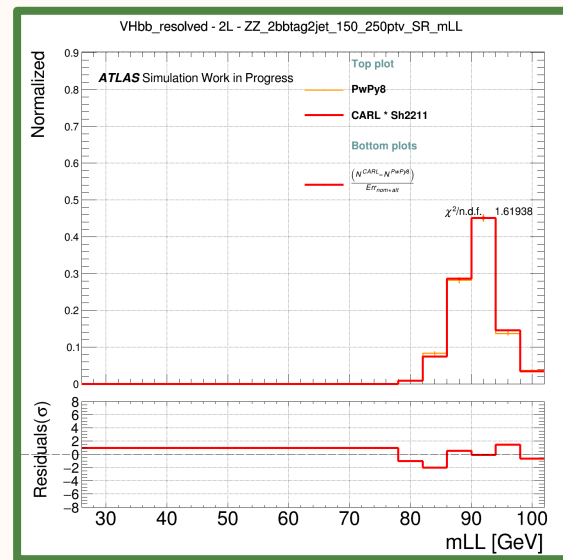
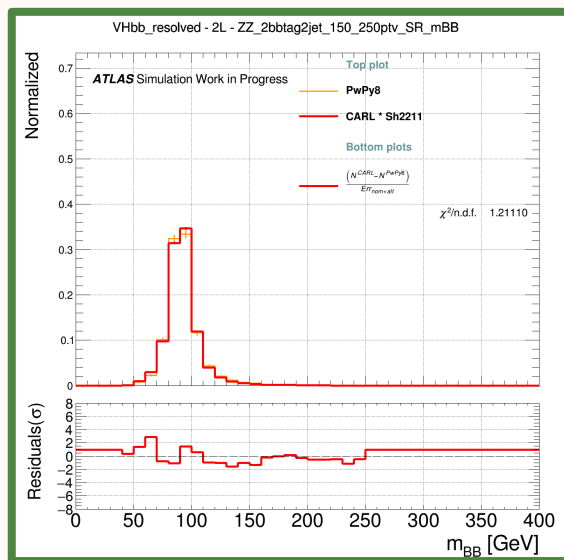
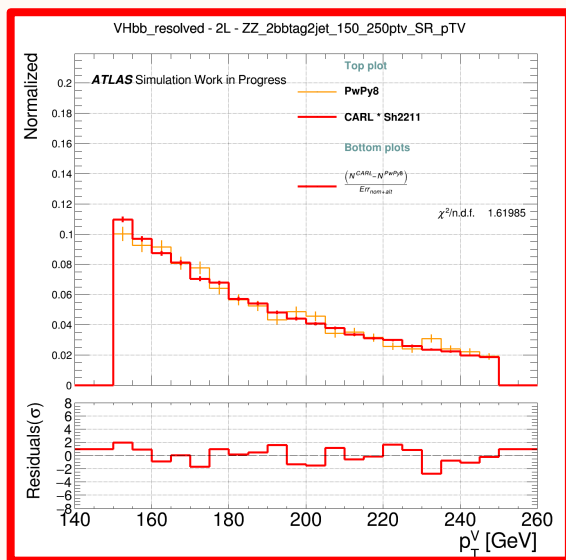


mBB reweight による各運動学変数分布の再現



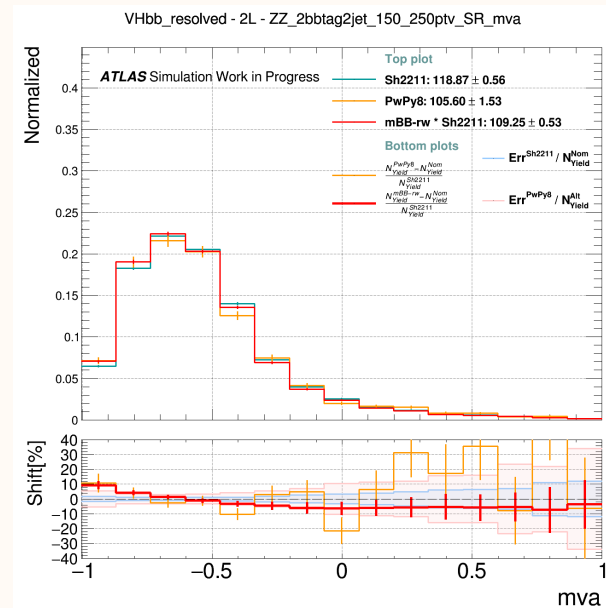
CARLによる各運動学変数分布の再現

赤:再現微量悪化
緑:再現向上



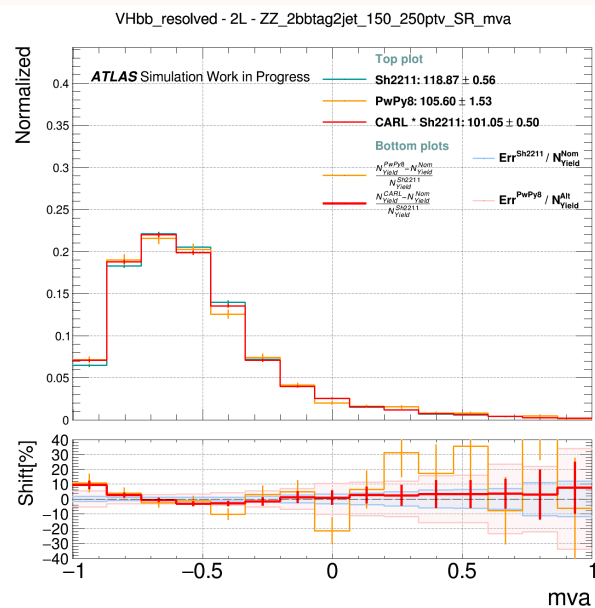
モデル間系統誤差の低統計によるふらつきを抑制

- 重み付けを行う両手法で、代替モデルの直接使用によるモデル間系統誤差のふらつき(赤)を抑えて、スムーズなモデル間系統誤差を評価できるようになった
- CARL の重み付け手法により、代替モデルを正確に再現できるようになっている
 - ヒッグス信号強度用MVA scoreでは、CARL によりモデル間系統誤差が小さく見積もられている
 - Diboson 信号強度用MVA scoreでは、CARL によるモデル間系統誤差が若干大きくなっている

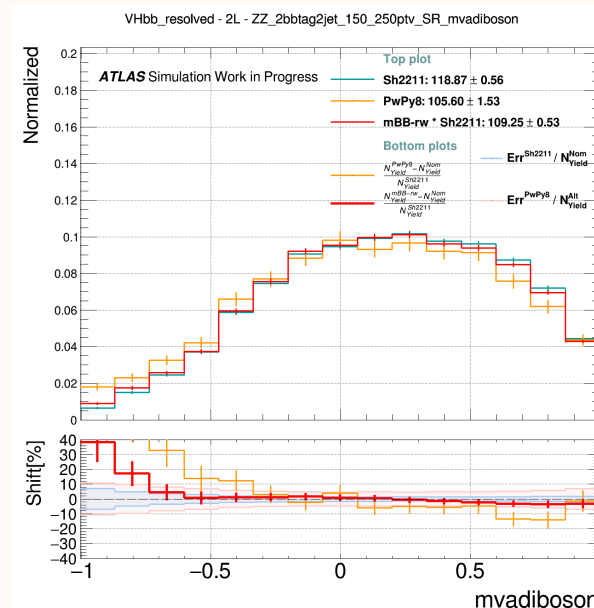


mBB reweight

ヒッグス信号強度測定用最終分離変数

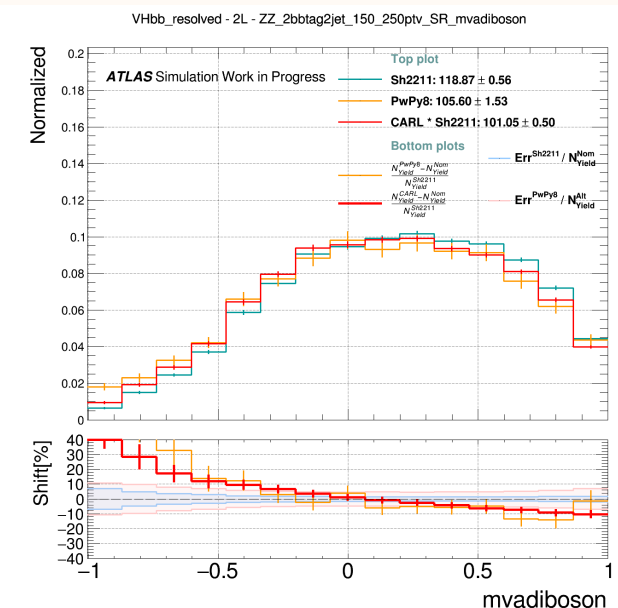


CARL reweight



mBB reweight

Diboson 信号強度測定用最終分離変数

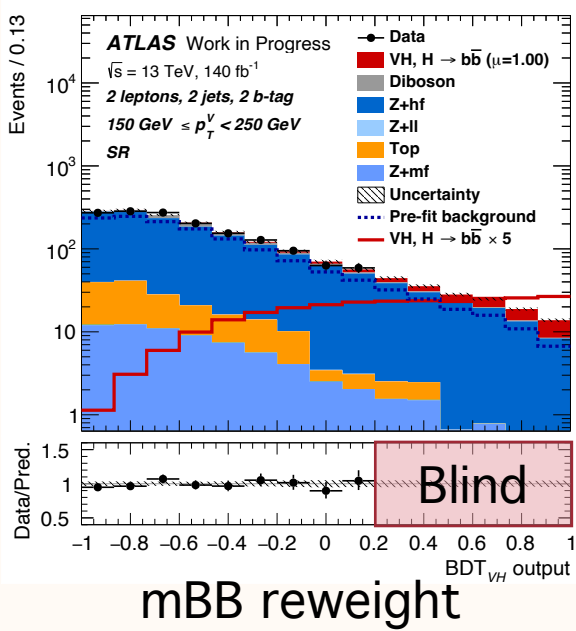


CARL reweight

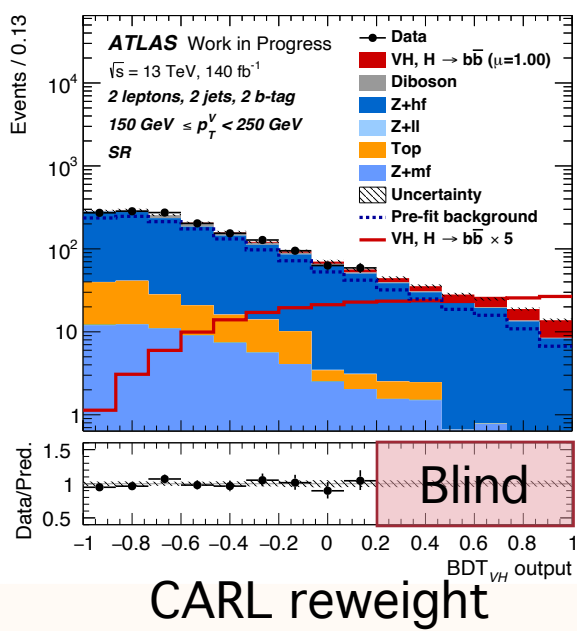


Post-fit mva plots

- 重み付けの方法を変えたことによる、実験データとMC サンプルの合いの変化はない
- ヒッグス信号強度測定の結果に関して：
 - Diboson は全体での貢献が比較的小さい背景事象のため、影響も限定的
 - モデル間系統誤差が小さく見積もられた影響で、Diboson 全体の系統誤差も小さくなった
- Diboson 信号強度測定の結果に関して：
 - Diboson は信号事象として ($BDT_{VZ} \sim 1$: Diboson-like) 扱われるため、影響が色濃く出る
 - モデル関係等誤差の大きな変化はなかったため、Diboson 全体の系統誤差もそこまで変化なかった



ヒッグス信号強度測定用最終分離変数



Diboson 信号強度測定用最終分離変数

