

1

ZH→llbb search ^Á using a Matrix Element technique

Université catholique de Louvain Camille Beluffi, Roberto Castello, <u>Adrien Caudron</u>, Ludivine Céard, Christophe Delaere, Tristan du Pree, Vincent Lemaitre, Arnaud Pin, Jesus Vizan

Scalar Search and Study in Belgium 23-24 January 2014

B





Adrien Caudron (UCL-CP3)

Motivations

Branching ratios W/W THC HIGGS XS MG bb ττ 10⁻² 10^{-3} 10⁻⁴ – 100 120 140 160 180 200 M_⊔ [GeV] 10 [dd] (X+H ∖s= 8 TeV → H (NNLO+NNLL QCD + NLO EW) 10 σ(pp → qqH (NNLO QCD + NLO EW) + NLO EW ttH (NLO QCD 10⁻¹ 10⁻²∟ 80 120 100 140 160 180 200 M_н [GeV]

- Important to probe SM in the fermionic sector
- H boson coupling to fermions harder to probe:
 - High background for $H \rightarrow bb$ due to QCD
 - Associate production (ZH)
 - Worse mass resolution
- At CP3:
 - IIbb final state: Z+b(b) cross-section
 - ME automatized tool: MadWeight
- Search for Z(II)H(bb)
- Use ME method to discriminate signal process from the background processes

Selection

Object	Z(ll)H(bb) (2011+2012)
Trigger	DiMu + DiEl
Leptons	$p_{\mathrm{T}}(l) > 20 \ \mathrm{GeV}$
(PF muons, electrons)	$ \eta(l) < 2.4, \eta(l) < 2.5$
	isolation criteria
ll-pair	$76 < m(l^+l^-) < 106 { m GeV}$
	$p_{ m T}(ll) > 20~{ m GeV}$
ak5 PF jets	$p_{\rm T}(j_1) > 40, p_{\rm T}(j_2) > 25~{ m GeV}$
	$ \eta(j) < 2.4$
Jet-lepton separation	$\Delta R(l,j) > 0.5$
B-tagging	CSV-MM
	M: <u>CSV > 0.67</u> 9
	(control: CSV-ML)
	(L: CSV > 0.244)
PF MET	MET significance $< 10^{\circ}$

Basic selection -

	inne
IVEA	10115.
U	

U U			
CD.	Signal	bb-pair	2 jets: $2j$: $80 < M(bb) < 150$ GeV
эк:	-		>2 jets: $3j: 50 < M(bb) < 150 \text{ GeV}$
	Control	bb-pair	2j: M(bb) < 80 M(bb) > 150 GeV
CR:			$3j: M(bb) < 50 \mid \mid M(bb) > 150 \text{ GeV}$
	Fit	ll-pair	Control + $61 < m(l^+l^-) < 121 \text{ GeV}$
·		For background estimation	



Matrix Element method

Matrix Element method* provide probability that an experimental event corresponds to a specific process (hypothesis).

$$P(x^{vis}|\alpha) = \frac{1}{\sigma_{\alpha}} \int dx_1 dx_2 f(x_1) f(x_2) \int d\Phi |M(p)_{\alpha}|^2 W(p^{vis}, p)$$

Where :

- p^{vis}: experimental event : {(Pt,eta,phi,E,B-tag,...)_{jet1}; (Pt,eta,phi,E,B-tag,...)_{jet2}; (Pt,eta,phi,E,charge,...)_{lep1}; (Pt,eta,phi,E,charge,...)_{lep2}; (Et,phi...)_{met} }
- **p** : partonic state
- $\triangleright \quad \alpha$: set of parameter defining the theoretical frame (α is fixed in this analysis).
- \rightarrow $|M(p)|^2$ Matrix element @ L.O.
- W(p,p^{vis}) : transfer function. Conditional probability that an observed quantity (p^{vis}) is the evolution of a partonic level one (p).

Weights are defined as : $W = \sigma \times P$

S3Be 24 Jan. 2014



Adrien Caudron (UCL-CP3)

Choice of event categorization

- Problem: events with >= 3 jets
 - Bigger background
 - Worse dijet mass resolution



Solution: separate categories

- 2 categories based on the number of jets
 - $p_T > 20$ GeV: 2 or >= 3 jets categories
- Improves sensitivity
- Special treatment for events with extra jets
 - ISR: transverse boost to LO ME rest frame
 - FSR: use M(bbj) and ΔR(b,j)
- Z pt spectrum found in agreement in each category



Background normalisation

- 2D simultaneous fit in 4 cats:
 - ee & µµ
 - 2 jets & 3 jets
- 2 discriminating observables
 (1) MLP TTbar vs Zbb
 - ME weights: qqZbb, ggZbb, tt
 - (2) CSV product
 - Product of CSV of the 2 b-tagged jets



SF	value 2011 MM	value 2012 MM
SF_Zbb	1.10 ± 0.09	1.12 ± 0.05
SF_Zbx	1.29 ± 0.11	1.27 ± 0.05
SF_Zxx	0.87 ± 0.18	1.08 ± 0.11
SF_tt	1.00 ± 0.10	0.94 ± 0.03

- Considering Zbbx production
- All SFs consistent
- SF_Zbx > 1







Construction of final discriminant



Unblinded ZH search 2011+2012

- Data/simulation of multivariate discriminant at 125 GeV
- Signal: 100*SM





- Observed limit at 125: $1.8 \times \sigma/\sigma_{sm}$
- At 125 GeV
 - Compatible with background only within 1 s.d.
 - Compatible with a SM H boson within 2 s.d.

ZZ search cross-check

2011+2012

Limits

- Expected: 0.8 x $\sigma/\sigma_{\text{CMS Meas}}$
- Observed: 1.4 x $\sigma/\sigma_{CMS Meas.}$
- Signal injection: 1.8 $\sigma/\sigma_{CMS Meas.}$
- Significance
 - Expected: 2.4
 - Observed: 1.5
- Signal strength: µ=0.6±0.4
- Compatible with SM





CMS internal

√s = 8 TeV, L = 19.6 fb

Data

ZZ

Z+bb

Z+bx

Z+xx

tt lvivbb

tt lviibb

ZH M H=125

Summary

- Matrix Element technique was used to perform a search of Z(II)H(bb) process
 - The MW weights are used in NN to discriminate signal and backgrounds
 - Extra jets are included with complementary info
- Presented unblinded result for 2011+2012
 - ZZ compatible with SM
 - ZH just unblinded and results compatible with SM H within 2 s.d.
- Analysis competitive
 - Many ways to improve it (CHS jets, jet energy regression...)
- Outlook
 - $H \rightarrow bb$ properties
 - Exotic searches in scalar sector with IIbb final state

Backup

Documentation

- Arnaud Pin thesis: Technique, validation and first 2011 results (link)
- AN-12-476
 - To be updated with unblinded ZH results

Scale factors for the background estimation

- 2D simultaneous fit of the 4 channels ee $-\mu\mu$ in both jet categorization (2 jets and 3 jets) - Variables used :
 - CSV product: product of the CSV value attached to the 2 b-tagged jets
 - MLP TTbar vs Zbb trained with 2011 or 2012 MC
 - * inputs are the qqZbb, ggZbb and tt MW weights



Fit Region

Each channel: - MLP: left - CSVprod: right

Adrien Caudron (UCL-CP3)

ZZ normalisation fixed to CMS measurements

SF Zbx \rightarrow Zbb 3 jets category + Zbx \rightarrow Zbb+j SF $Zxx \rightarrow$ same in both categories (DY+jets)

SF tt \rightarrow same in both categories

Scale factors: results

 Table 4: The background scale factors as estimated from the 2D fits for 2011 and 2012 data.

SF	value 2011 MM	value 2012 MM	value 2011 ML	value 2012 ML
SF_Zbb	1.10 ± 0.09	1.12 ± 0.05	1.05 ± 0.08	1.06 ± 0.05
SF_Zbx	1.29 ± 0.11	1.27 ± 0.05	1.20 ± 0.10	1.22 ± 0.04
SF_Zxx	0.87 ± 0.18	1.08 ± 0.11	0.88 ± 0.10	1.38 ± 0.04
SF_tt	1.00 ± 0.10	0.94 ± 0.03	0.96 ± 0.09	1.00 ± 0.03

Very good agreement between all four cases, except for SF_Zxx, where the fake rate is affected by pileup in 2012

Strategy

Z(II)H(bb) search (I= e/mu):

- Based on Z(II)+bb cross-section measurement analysis (similar selection).
- b-tag: CSV discriminant at Medium-Medium working point.
- Categorization in jet multiplicity
 - 2 jets: events with Exactly 2 jets identified as b-jets
 - 3 jets: events with at least 3 jets with 2 identified as b-jets.

Considered Backgrounds:

- ttbar: Dileptonic decay channel (ee/mumu)
- DY + jets: Z+bb, Z+bx, Z+xx x=light/c jets
- diboson: Z(II)Z(bb)

Matrix Element Method	Process	Hypothesis	ISR correction	E-P conservation
For each event we test	Higgs	$qq \rightarrow ZH \rightarrow l^- l + b \bar{b}$	without MeT	conserved
- For each event we test	Higgs	$qq ightarrow ZH ightarrow l^- l + b ar{b}$	without MeT	Not conserved
signal and background	$t\bar{t}$	$pp ightarrow t ar{t} ightarrow l^- l + u ar{ u} \ b ar{b}$	with MeT	conserved
hypothesis → 7 weights	$Zb\overline{b}$	$gg ightarrow l^- l + b ar{b}$	without MeT	conserved
per event	$Zb\overline{b}$	$qq ightarrow l^- l + b ar{b}$	without MeT	conserved
	$\mathbf{Z}\mathbf{Z}$	$qq ightarrow ZZ ightarrow l^- l + b ar{b}$	without MeT	conserved
	7.7.	$aa \rightarrow ZZ \rightarrow l^- l + b\bar{b}$	without MeT	Not conserved

Multivariate analysis:

- MLP discriminants use as inputs the M.E. weights + other variables in the 3 jets category.

- Intermediate Discriminants: ZH-tt, ZH-Zbb, ZH-ZZ
- → ZH-BKG discriminant based on the three intermediate ones.

ME TF



ME ISR



21/09/12

120 140

Pt Boost [GeV]

Arnaud Pin

160

140

Pt Boost [GeV]

10

Datasets

Table 2: Data and MC samples used in this analysis. MASS means all masses between 110 and 135 by step of 5 GeV/c2. * is Fall11–PU_S6 for 7 TeV MC and Summer12_DR53X–PU_S10 for 8 TeV MC. Samples with a^{\dagger} are not used in the MVA training. All samples are taken from AOD (Data) and AODSIM (MC) files.

2011		
Data	ElA	/DoubleElectron/Run2011A-08Nov2011-v1/
	ElB	/DoubleElectron/Run2011B-19Nov2011-v1/
	MuA	/DoubleMu/Run2011A-08Nov2011-v1/
	MuB	/DoubleMu/Run2011B-19Nov2011-v1/
MC	DY	/DYJetsToLL_TuneZ2_M-50_7TeV-madgraph-tauola/*-START44_V5-v1/
	DY $p_{\rm T}(Z) > 100$	/DYJetsToLL_PtZ-100_TuneZ2_7TeV-madgraph-tauola/
	TT	/TTJets_TuneZ2_7TeV-madgraph-tauola/*-START44_V5-v1/
	ZZ	/ZZ_TuneZ2_7TeV_pythia6_tauola/*-START44_V5-v1/
	ZH	/ZH_ZToLL_HToBB_M-MASS_7TeV-powheg-herwigpp/
Training	Zbb	/Zbb_4F_7TeV_madgraph/*_START44_V9B-v1/
2012		
Data	ElA	/DoubleElectron/Run2012A-13Jul2012-v1/
	ElB	/DoubleElectron/Run2012B-13Jul2012-v1/
	ElA 06aug	/DoubleElectron/Run2012A-recover-06Aug2012-v1/
	ElC-v1	/DoubleElectron/Run2012C-24Aug2012-v1/
	ElC-v2	/DoubleElectron/Run2012C-PromptReco-v2/
	ElD	/DoubleElectron/Run2012D-PromptReco-v1/
	MuA	/DoubleMu/Run2012A-13Jul2012-v1/
	MuB	/DoubleMu/Run2012B-13Jul2012-v4/
	MuA 06aug	/DoubleMu/Run2012A-recover-06Aug2012-v1/
	MuC-v1	/DoubleMu/Run2012C-24Aug2012-v1/
	MuC-v2	/DoubleMu/Run2012C-PromptReco-v2/
	MuD	/DoubleMu/Run2012D-PromptReco-v1/
MC	DY inclusive	/DYJetsToLL_M-50_TuneZ2Star_8TeV-madgraph-tarball/*_START53_V7A-v1/
	DY $p_{\rm T}(Z) [50 - 70]^{\dagger}$	/DYJetsToLL_PtZ-50To70_TuneZ2star_8TeV-madgraph-tarball/*_START53_V7A-v1/
	DY $p_{\rm T}(Z) [70 - 100]^{\intercal}$	/DYJetsToLL_PtZ-70To100_TuneZ2star_8TeV-madgraph-tarball/*_START53_V7A-v2/
	DY $p_{\rm T}(Z) > 100^{+}$	/DYJetsToLL_PtZ-100_TuneZ2star_8TeV-madgraph/*_START53_V7A-v2/
	DY $p_{\rm T}(Z) > 180^{\dagger}$	/DYJetsToLL_PtZ-180_TuneZ2star_8TeV-madgraph-tarball/*_START53_V7C-v1/
	TT Fully Leptonic	/TTJets_FullLeptMGDecays_8TeV-madgraph/*_START53_V7A-v2/
	TT Semi-Leptonic [™]	/TTJets_SemiLeptMGDecays_8TeV-madgraph/*_START53_V7A_ext-v1/
	ZZ	/ZZ_TuneZ2star_8TeV_pythia6_tauola/*_START53_V7A-v1/
	ZH	/ZH_ZToLL_HToBB_M-MASS_8TeV-powheg-herwigpp/*_START53_V7A-v1/
Training	Zbb	/ZbbToLL_massive_M-50_TuneZ2star_8TeV-madgraph-pythia6_tauola/*_START53_V7A-v1/
	TT inclusive	/TTJets_MassiveBinDECAY_TuneZ2star_8TeV-madgraph-tauola/*_START53_V7A-v1/

PU discussion



- Brown is purely PU events contribution:
 - Both b-tagged jets are unmatched with a gen-jet
- Cut on jets lower: pt>20 and no cut on pt Z

- Same cuts, comparing same DY vs DY events. (~12% of whole sample)
- Raw number of selected events
- Old DY is part of the DY on the right,
- Zbb, Zbx and Zxx used CHS jets

Adrien Caudron (UCL-CP3)

CR: Variables used to perform the SFs fit



Adrien Caudron (UCL-CP3)

Systematics

- Luminosity unc.:
 - 2.2% (4.4%) on signal normalisation at 7 TeV (8 TeV)
- Theoretical unc. on signal cross-section:
 - 4% on signal normalisation
- Lepton Reconstruction and Trigger Efficiency:
 - 4% on the signal normalisation
- ZZ cross-section measurement unc.:
 - 40% (15%) at 7 TeV (8 TeV)
- Background normalisation unc.:
 - Uncorrelated unc. from the fit unc.
- JER unc.:
 - 2-6% normalisation on the signal normalisation

- B-tag unc.:
 - shape unc., b, c SFs unc. and light SFs unc. are taken uncorrelated
- JES unc.:
 - shape unc. (MW weights recomputed)
- MC statistics unc.: shape unc.
 - Allow a bin by bin fluctuation of the MC normalisation according the statistical uncertainties in the 10 most sensitive bins.

Systematics: Background fit correlations

 Diagonalization of covariance Matrix → we obtain transformation Matrix T such as T⁻¹ Cov T = Diag.

error matrix=	Correlated Uncert.	bg1	bg2	bg3
	e1	а	0	0
	e2	0	b	0
	e3	0	0	С

input in data card: unity matrix + Modif_error

	Not correlated	bg1	bg2	bg3	
ModifiedError=T- ¹ Error=	Eprime1	x11	x12	x13	● ePrin
	Eprime2	x23	x33	x23	corre
	Eprime3	x13	x23	x33	

ePrime1, ePrime2, ePrime3 not correlated

0

Systematics breakdown

Systematics	limit	degradation (%)
No	2.07	
All	2.43	
-MC statistical unc.	2.26	7.5
-Zbb	2.30	5.7
-Zxx	2.40	1.3
-Zbx	2.41	0.8
-tt	2.41	0.8
-ZZ	2.42	0.4
-ZH	2.42	0.4
-Background norm.	2.32	4.7
-JER	2.41	0.8
-JES	2.41	0.8
-ZH cross-section	2.41	0.8
-Luminosity	2.41	0.8
-Lepton SFs	2.41	0.8
-Btag b, c-jets SFs	2.41	0.8
-Btag light-jets SFs	2.43	0

Breakdown of systematics
 Show the effect of removing ONE source of systematic each time
 Most important effect from MC stat. uncertainty and from Bkg Normalization SFs



ee+uu

- Signal region defined as M(bb) in [45,115] in the 2 jets category and M(bb) in [15,115] in the 3 jets category.
- Same strategy but only 2 intermediate NNs trained against DY and TTbar
- Difficult training as the background peak around the Z mass



Signal EWK NLO p_T(Z) reweigting



- We use the official values to reweight our signal samples to higher order corrections.
- Signal yields decrease by ~3-4%

Nuisance parameters fit

Comparison of nuisances

nuisance	background fit Δx/σ _{in} , σ _{out} /σ _{in}	signal fit Δx/σ _{in} , σ _{out} /σ _{in}	ρ(μ, θ)
JER	+0.00, 0.99	+0.00, 0.97	-0.00
JER2012	-0.12, 0.99	-0.09, 0.97	+0.00
JES 2011	+0.00, 0.99	+0.01, 0.97	-0.01
ZHunc	+0.00, 0.99	-0.01, 0.97	+0.04
ZHuncEWK	+0.00, 0.99	-0.01, 0.97	+0.05
ZHuncQCD	+0.00, 0.99	-0.00, 0.97	+0.02
ZZunc	+0.04, 0.98	+0.09, 0.96	-0.06
_EEChannelCutlTT-FullLeptstat_binl1	-0.07, 0.98	-0.07, 0.92	-0.01
_EEChannelCut1TT-FullLeptstat_bin12	-0.18, 0.97	-0.18, 0.91	+0.00
_EEChannelCut1TT-FullLeptstat_bin13	-0.12, 0.97	-0.11, 0.91	-0.00
_EEChannelCutlTT-FullLeptstat_bin14	-0.07, 0.98	-0.06, 0.91	-0.01
_EEChannelCutlTT-FullLeptstat_bin15	-0.09, 0.97	-0.08, 0.91	-0.00
_EEChannelCutlTT-FullLeptstat_bin16	-0.04, 0.98	-0.04, 0.92	-0.01
_EEChannelCutlTT-FullLeptstat_bin17	-0.13, 0.95	-0.13, 0.89	-0.01
_EEChannelCutlTT-FullLeptstat_bin18	-0.04, 0.97	-0.03, 0.92	-0.02
EEChannelCut1TT-FullLeptstat_bin19	+0.03, 1.02	+0.12, 0.98	-0.08
_EEChannelCutlTT-FullLeptstat_bin20	-0.00, 0.99	-0.00, 0.44	+0.00
_EEChannelCutlTT-SemiLeptstat_binll	-0.01, 0.97	-0.00, 0.91	-0.00
_EEChannelCut1TT-SemiLeptstat_bin12	-0.04, 0.97	-0.04, 0.91	+0.00
EEChannelCut1TT-SemiLeptstat_bin13	-0.01, 0.95	-0.01, 0.89	-0.00

_JES 2012 +0.43, 0.08 +0.44, 0.08	-0.17
-----------------------------------	-------



Z(II)H(bb) with Matrix Element Method: Unblinding

Nuisance parameters fit



MuMuChannelCut1Zbxstat_bin19	+0.09, 1.07	+0.14, 1.01	-0.06
MuMuChannelCut1Zbxstat_bin20	-0.00, 0.09	-0.00, 0.01	-0.00
MuMuChannelCut1Zxxstat_bin11	+0.51, 0.94	+0.51, 0.88	-0.01
MuMuChannelCut1Zxxstat_bin12	+0.28, 0.87	+0.29, 0.81	-0.01
MuMuChannelCut1Zxxstat_bin13	-0.45, 0.73	-0.45, 0.70	-0.01
MuMuChannelCut1Zxxstat_bin14	+0.25, 0.93	+0.26, 0.86	-0.01
MuMuChannelCut12xxstat_bin15	+0.09, 0.95	+0.11, 0.88	-0.02
MuMuChannelCut1Zxxstat_bin16	-0.17, 0.91	-0.17, 0.87	-0.01
MuMuChannelCut1Zxxstat_bin17	-0.45, 0.83	-0.44, 0.78	-0.00
MuMuChannelCut1Zxxstat_bin18	+0.01, 1.01	+0.01, 0.95	-0.00
MuMuChannelCut1Zxxstat_bin19	+0.18, 1.11	+0.30, 1.02	-0.14
MuMuChannelCut1Zxxstat_bin20	-0.00, 0.08	-0.00, 6.05	-0.05
bgnorml	-0.22, 0.93	-0.20, 0.91	-0.02
bgnorm2	-0.52, 0.96	-0.51, 0.94	-0.02
bgnorm3	+0.34, 0.74	+0.32, 0.73	+0.03
bgnorm4	+2.14, 0.69	+2.08, 0.68	+0.11
bgnorm5	-0.19, 0.94	-0.17, 0.92	-0.01
boostEWK	+0.00, 0.99	+0.00, 0.97	-0.00
boostQCD	+0.00, 0.99	+0.01, 0.97	-0.00
elecSF	+0.00, 0.99	+0.00, 0.97	+0.00
lepunc_ee	+0.00, 0.99	+0.02, 0.97	-0.01
lepunc_mm	+0.00, 0.99	-0.03, 0.97	+0.05
lumi	+0.01, 0.99	+0.02, 0.97	+0.02
lumi2011	+0.00, 0.99	+0.00, 0.97	-0.00
muonSF	+0.00, 0.99	+0.00, 0.97	-0.00
sftt_MM	-0.22, 0.89	-0.20, 0.88	-0.03
sfzbb_MM	-0.19, 0.84	-0.31, 0.84	+0.17
sfzbx_MM	+1.37, 0.86	+1.36, 0.84	+0.02
sfzxx_MM	-1.49, 0.78	-1.40, 0.77	-0.13
signorm	+0.00, 0.99	+0.01, 0.97	-0.00

Limitation and room for improvement

- basing the analysis on the re-reco dataset for 2011 and 2012.
- implementing the Charge Hadron Substraction.
 - Better treatment of PU jets in the event.
- Improving the M_{bb} resolution applying the M_{bb} regression.
- Re-evaluating the M.E. weights with optimized transfer functions.
- Using single Muon trigger.

Timescale: during 2014 (for all of them)