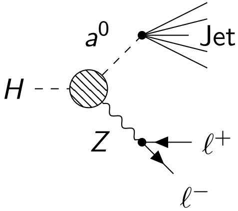
Search for $H \to Za^0 \to \ell\ell j$ with ATLAS

IoP, APP and HEPP Meeting, 2019

Elliot Reynolds

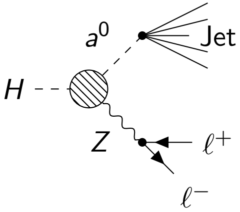


ATLAS EXPERIMENT

UNIVERSITY OF BIRMINGHAM

Aims

- Use full ATLAS Run II dataset (140 fb^{-1}) to perform first search for $\boldsymbol{H \rightarrow Z(\ell\ell)X(\text{had})}$, where $\ell = e$ *or* μ
- Interpret X as $\boldsymbol{J/\psi}$, $\boldsymbol{\eta_c}$, or $\boldsymbol{a^0}$ (BSM) with $\boldsymbol{m_{a^0} < 4}$ **GeV**



Motivation - Charmonium

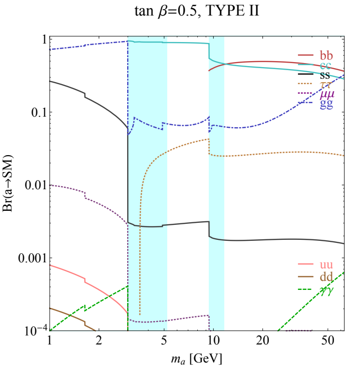
- Higgs boson decay to Z + light resonances unconstrained
- Provides low Q^2 probe of $H \rightarrow ZZ^*$
- Potential limits on charm Yukawa coupling

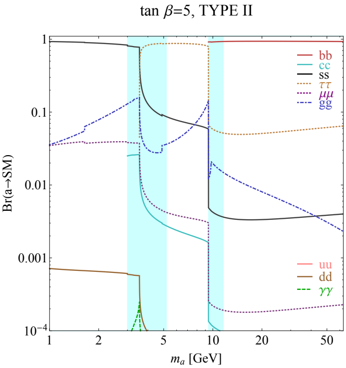
Motivation - BSM

- Many BSM models† predict Higgs boson decays into a Z boson and a light psuedoscalar (a^0) with a large BR to hadrons

†Eur. Phys. J. C (2016) 76: 501

- 2HDM+s is required to provide the masses in NMSSM
- Figures† show **dominant hadronic BR** until $\sim 2m_c$
- BR(had)$\gtrsim$99% for $\tan\beta = \frac{1}{2}$ in Type II, III, & $\tan\beta = 5$ in Type IV†





†Phys. Rev. D 90, 075004

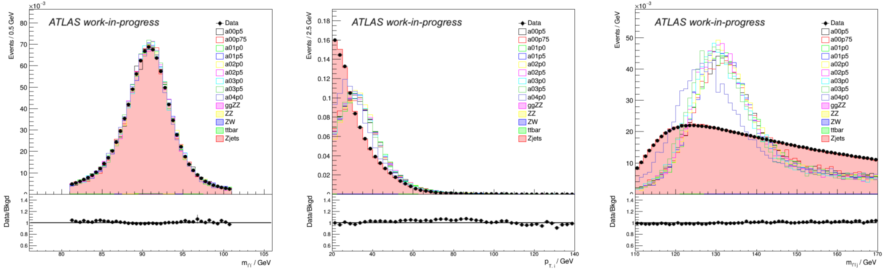
Physics Processes

- Focus on low mass ($<$ **4 GeV**) signals, as higher BR and unique a^0 decay kinematics lead to higher sensitivity
- Search for signals from inclusive Higgs boson production
- The dominant background is $\boldsymbol{Z+}$**jets**, with $< 1\%$ contributions from $t\bar{t}$ and diboson

Simulation

- Signals modelled using POWHEG, PYTHIA8 and EVTGEN
- $Z +$ jets modelled using SHERPA 2.2.1
- Full GEANT4 simulation of the ATLAS detector

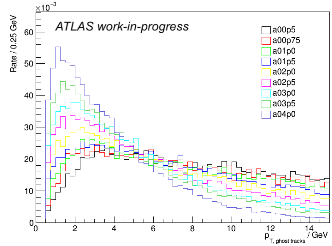
Event-Level Kinematics and Selection



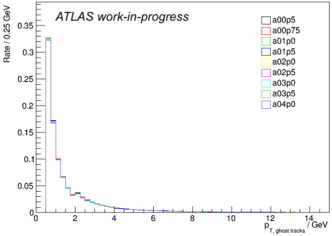
Selection	Details
Triggers	Single lepton triggers $p_{\mathrm{T,\ lead\ lepton}} > 27$ GeV
Leptons	$N_\ell \geq 2$ with $p_\mathrm{T} > 18$ GeV
Z boson	2 SF OS leptons, with $\lvert m_{ll} - m_Z \rvert < 10$ GeV
Jet (a^0)	Anti-k_T $R = 0.4$ jet with $p_{\mathrm{T,\ j}} > 20$ GeV
Pre-Higgs	$m_{\ell\ell\mathrm{j}} < 250$ GeV
Select highest p_T jet as a^0-candidate	
≥ 2 tracks	≥ 2 tracks ghost associated to the calorimeter jet
Higgs SR	120 GeV $< m_{\ell\ell\mathrm{j}} <$ 135 GeV

- Tracks **Ghost-Associated**† to the calorimeter jet used to form input variables for classification MVA
- Loose track quality requirements applied
- $|d_0| < 2$ & $|\Delta z_0 \sin\theta| < 3$ required
- Signal efficiencies of $94 - 96\%$, for a pileup rejection of $\sim 60\%$
- Jets are required to have ≥ 2 tracks surviving these requirements

Tracks from a^0



Tracks not from a^0

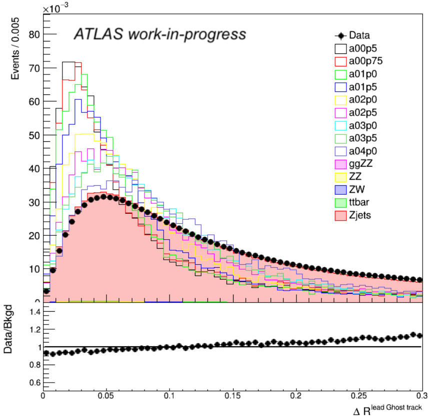


† Phys. Lett. B 659:119-126, 2008

MVA Input Variables

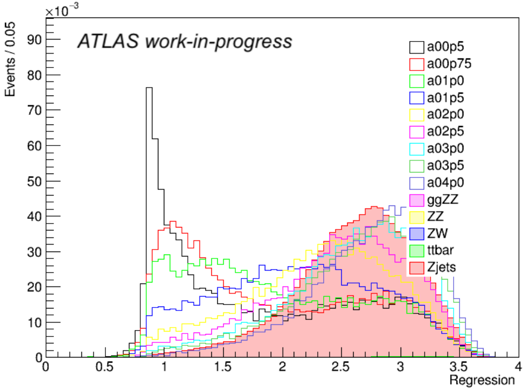
- Input variables:
  1. $\Delta R_{\text{lead track}}$
  2. $p_{\text{T, lead track}} / p_{\text{T, all tracks}}$
  3. τ_2
  4. $U1(0.7)^{\dagger}$
  5. $M2(0.3)^{\dagger}$
  6. $angularity(2)$

- All dimensionless to minimise correlation between MVA output and $m_{\ell\ell j}$

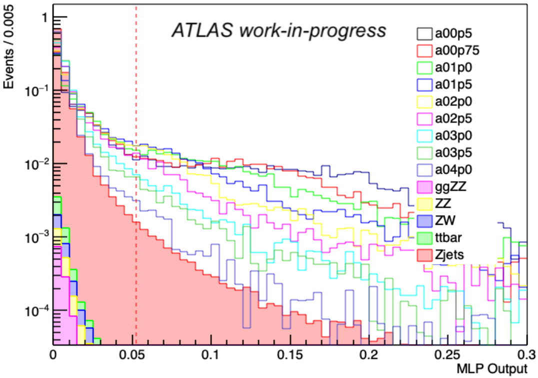


† J. High Energ. Phys. (2016) 2016: 153

- A **Multi-Layer-Perceptron** (MLP) is used to classify signal resonances against background jets
- Not a standard classification problem, due to the spectrum of signals
- This is solved by training a regression MLP to predict m_{a^0}
- The mass hypothesis is input to the classifier, informing it which part of the phase space to consider
- This results in $\sim$ **13%** improvement in the expected $S/\sqrt{B}$



- Cut chosen to optimise the expected $S/\sqrt{B}$, assuming all values of a^0 mass equally likely: 0.052
- MLP background efficiency = **1.0%**

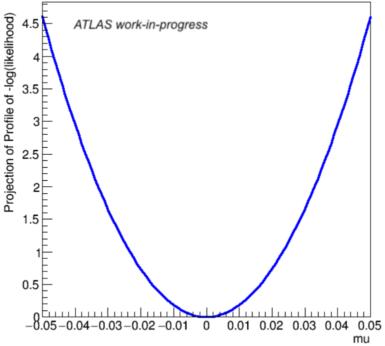


a^0 mass / GeV	0.5	0.75	1	1.5	2	2.5	3	3.5	4
MLP Eff (%)	31	28	26	21	17	11	5.6	4.3	1.7
MLP $S/\sqrt{B}$ Gain	3.1	2.7	2.6	2.1	1.7	1.1	0.56	0.43	0.17

- 2-bin **cut-and-count** analysis strategy adopted
  - Signal region (SR): 120 GeV $< m_{\ell\ell\mathrm{j}} <$ 135 GeV
  - Control region (CR): 100 GeV $< m_{\ell\ell\mathrm{j}} <$ 110 GeV or 155 GeV $< m_{\ell\ell\mathrm{j}} <$ 175 GeV
- Background estimated using MC-based transfer factor: $T = B^{\mathrm{MC}}_{\mathrm{SR}} / B^{\mathrm{MC}}_{\mathrm{CR}}$
- Expected 0.5 GeV a^0 yield for BR=1: **24k**
- Expected background yield: **84k**

Alternative Background Model	Closure
Direct MC estimate	2.2%
MC extrapolation from MLP CR	0.68%
MC corrected ABCD estimate	3.4%

0.5 GeV a^0



- Likelihood fits to Asimov datasets are used to extract expected results in the absence of systematics

- Uncertainties and 95% CL limits estimated for $\mu_{\text{Sig}} = {}^{\sigma(H)\text{BR}(H\to Za^0)}/_{\sigma_{\text{SM}}(H)}$

Expected Stat-Only Results

a^0 mass / GeV	0.5	0.75	1	1.5	2	2.5	3	3.5	4
$\Delta\mu_{\text{Sig}}$ (%)	1.7	1.9	1.9	2.1	2.7	4.0	7.5	10	37
95% CL Limit (%)	3.2	3.8	3.7	4.2	5.2	7.9	15	20	72

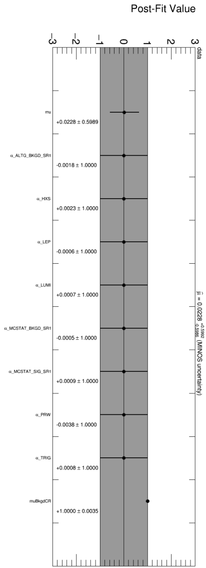
Systematic Uncertainties

Systematic Uncertainty	95% CL Limit (%) for 0.5 GeV a^0
All Systematics	**30**
Background Modelling	19
Background MC Statistics	17
Pileup	15
Leptons	8.1
Trigger	3.4
Higgs Cross Section	3.3
Signal MC Statistics	3.2
Luminosity (2% assumed)	3.2
Stat-Only Fit	3.2

- Background modelling uncertainty on $T = B_{SR}/B_{CR}$ evaluated by comparison with MadGraph
- Jet, tracking and signal modelling uncertainties yet to be added

Validation Region

- Full analysis strategy validated in MLP-sideband validation region (VR): $0.034 < MLP < 0.052$
- 89.8k background events expected in VR
- **89919** events observed in VR
- All systematics included in fit
- $\mu = 0.02 \pm 0.60$



ATLAS work-in-progress

Summary

- Expected 95% CL limit set on $\sigma(H)\mathrm{BR}(H\to Za^0)/\sigma_{\mathrm{SM}}(H)$, in the absence of systematics, at:
  1. **3.2 – 72%** for the a^0 signal samples ($0.5 < m_{a^0} < 4$ GeV)
  2. **19%** for the η_c signal sample
  3. **18%** for the J/ψ signal sample
- Expected 95% CL limit with (most) systematics set for 0.5 GeV a^0 signal sample at: **30%**
- Analysis validated in MLP-based VR

- Expected 95% CL limit set on $\sigma(H)\mathrm{BR}(H\to Za^0)/\sigma_{\mathrm{SM}}(H)$, in the absence of systematics, at:
  1. **3.2 – 72%** for the a^0 signal samples ($0.5 < m_{a^0} < 4$ GeV)
  2. **19%** for the η_c signal sample
  3. **18%** for the J/ψ signal sample
- Expected 95% CL limit with (most) systematics set for 0.5 GeV a^0 signal sample at: **30%**
- Analysis validated in MLP-based VR

Thank you for listening!

Backup Slides

- Likelihood
- MC generators
- a^0 Branching Ratios
- Other MLP input variables
- Alternative background estimates
- Pileup tracks
- Correlation and pull plots

Likelihood (stat-only):
$\mathcal{L} = \text{Pois}(N^{\text{D}}_{\text{SR}}; \boldsymbol{\mu} S^{\text{MC}}_{\text{SR}} + \boldsymbol{T} \times \boldsymbol{\mu_B} \times N^{\text{D}}_{\text{CR}})) \times \text{Pois}(N^{\text{D}}_{\text{CR}}; \boldsymbol{\mu_B} \times N^{\text{D}}_{\text{CR}}))$,
where $S^{\text{MC}}_{CR} \approx 0$

Table below shows MC generators used to model various backgrounds

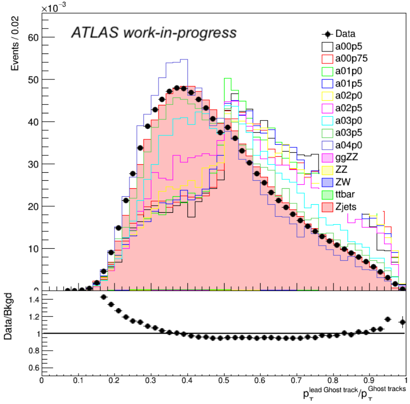
Process	MC Generator
(ggF) $H \rightarrow Za^0$	Powheg+Pythia8+EvtGen
(ggF) $H \rightarrow Z\eta_c$	Powheg+Pythia8+EvtGen
(ggF) $H \rightarrow ZJ/\psi$	Powheg+Pythia8+EvtGen
Z+jets	Sherpa 2.2.1
ZZ	Sherpa 2.2.1
ZW	Sherpa 2.2.1
$t\bar{t}$	Powheg+Pythia8+EvtGen

Table below shows mains decay modes ($BR > 1\%$) for various a^0 mass points, assuming $\Gamma = m/1000$

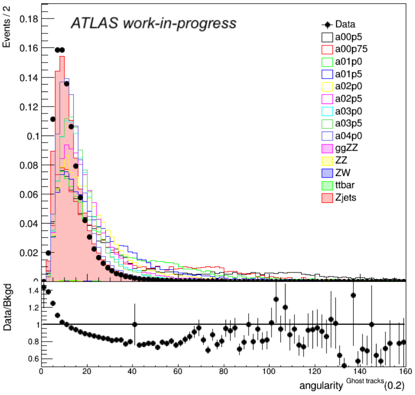
Mass Point / GeV	Main Decay Modes
0.5	gg (92%), $\mu^+\mu^-$ (8%)
1	gg (88%), $\mu^+\mu^-$ (12%)
1.5	gg (76%), $s\bar{s}$ (16%), $\mu^+\mu^-$ (8%)
2	gg (82%), $s\bar{s}$ (13%), $\mu^+\mu^-$ (5%)
2.5	gg (88%), $s\bar{s}$ (8%), $\mu^+\mu^-$ (4%)
3	gg (86%), $s\bar{s}$ (9%), $\mu^+\mu^-$ (4%)
3.5	$c\bar{c}$ (88%), gg (10%), $s\bar{s}$ (1%)
4	$c\bar{c}$ (57%), $\tau^+\tau^-$ (37%), gg (5%)
4.5	$c\bar{c}$ (52%), $\tau^+\tau^-$ (43%), gg (4%)
5	$c\bar{c}$ (50%), $\tau^+\tau^-$ (45%), gg (4%)
8	$\tau^+\tau^-$ (45%), $c\bar{c}$ (40%), gg (14%)
12	$b\bar{b}$ (81%), $\tau^+\tau^-$ (10%), $c\bar{c}$ (7%), gg (2%)

Other MLP Input Variables

$p_{\mathrm{T,\,lead\,track}}/p_{\mathrm{T,\,all\,tracks}}$



angularity(2)



Number of expected background events in various regions

Background Estimation Method	SR	Gap	VR	VR Gap
Direct MC estimate	82400	142000	93800	142000
MC extrapolation from $m_{\ell\ell j}$ CR	84300	145000	89800	136000
MC extrapolation from MLP CR	83700	146000	95400	146000
ABCD estimate	81400	142000	86700	133000

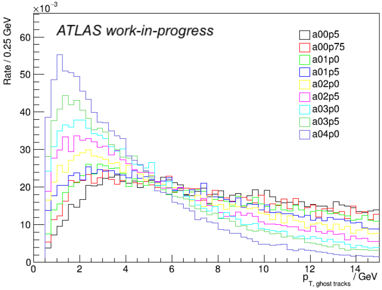
(Gap: the disconnected region between the SR and CR, 110 GeV $< m_{\ell\ell j} <$ 120 GeV and 135 GeV $< m_{\ell\ell j} <$ 155 GeV)

(VR: MLP-based validation region, defined to be as close as possible to SR in MLP, and contain the same amount of background, $0.034 < MLP < 0.052$)

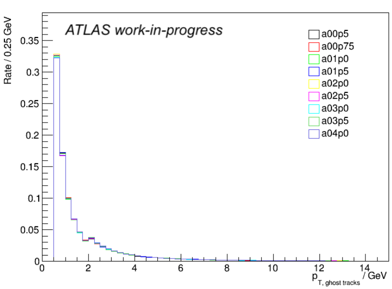
Truth Matching

- Need algorithm to reject pileup
- To access such an algorithm, must know which tracks are from a^0
- Links to truth particles responsible for tracks stored in AOD
- Follow family tree up and record if a^0 is present

Tracks from a^0



Tracks not from a^0

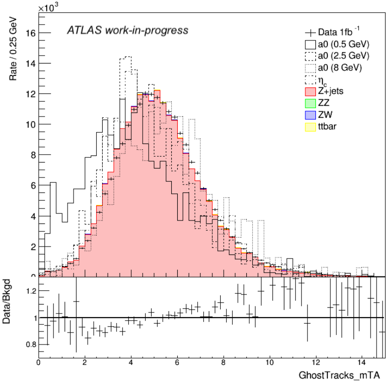


- Two pileup rejection working points considered
- Both include Loose track quality WP (included in efficiencies)
- Loose TTVA benefits from being the standard for R21, so will use that

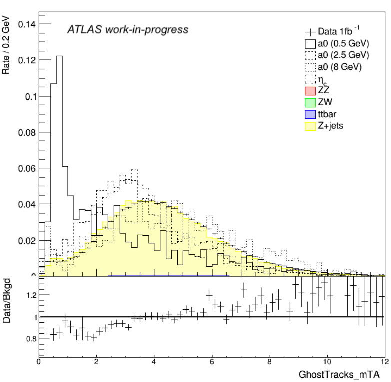
Signal Sample	Loose TTVA ($d_0 < 2,\ \Delta z_0 \sin\theta < 3$)		$d_0 < 2,\ \Delta z_0 \sin\theta < 1,\ p_T > 1$ GeV	
	Signal Efficiency	Pileup Efficiency	Signal Efficiency	Pileup Efficiency
$a^0_{0.5\text{ GeV}}$	$96 \pm 7\%$	$41 \pm 3\%$	$94 \pm 7\%$	$26 \pm 2\%$
$a^0_{2.5\text{ GeV}}$	$95 \pm 6\%$	$38 \pm 2\%$	$90 \pm 5\%$	$24 \pm 2\%$
$a^0_{8\text{ GeV}}$	$94 \pm 6\%$	$47 \pm 3\%$	$83 \pm 5\%$	$32 \pm 2\%$
η_c	$95 \pm 4\%$	$37 \pm 2\%$	$88 \pm 4\%$	$23 \pm 2\%$

- Removing the pileup tracks greatly improved the track-assisted mass
- This could form the basis for the reconstruction of the a^0 mass

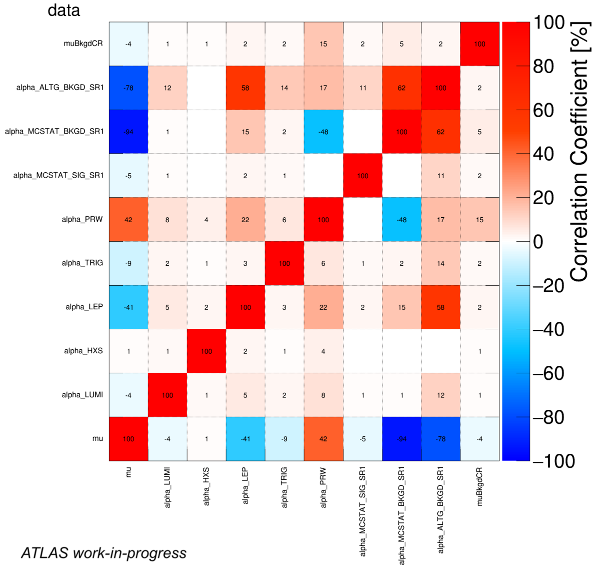
Before

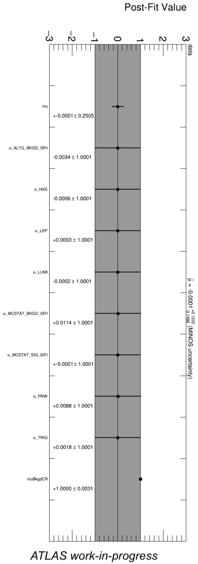


After



Systematic Uncertainty Correlations and Pulls





Validation Region Correlations and Pulls