PROCEEDINGS OF SCIENCE

Development of novel experimental techniques to improve the understanding of the Higgs sector by the ATLAS experiment

Jiggins, Stephen[a,*] on behalf of the ATLAS Collaboration

[a] *Deutsches Elektronen-Synchrotron DESY,*
Notkestr. 85, 22607 Hamburg, Germany

E-mail: stephen.jiggins@cern.ch

With the full Run-2 (2015-2018) proton-proton collision data collected by the ATLAS detector at the Large Hadron Collider, precise measurements of Higgs boson properties in an array of production and decay modes are now possible. To maximise the scientific value of the recorded data, novel experimental techniques were developed. The following article reviews a representative selection of such techniques, which includes: multi-class machine learning classification optimisation algorithms, experimental uncertainty regression, input variable invariant adversarial neural networks, object embedding, and multi-dimensional likelihood re-weighting techniques designed to maximise the statistical precision of Monte Carlo predictions.

41st International Conference on High Energy physics - ICHEP2022
6-13 July, 2022
Bologna, Italy

*Speaker

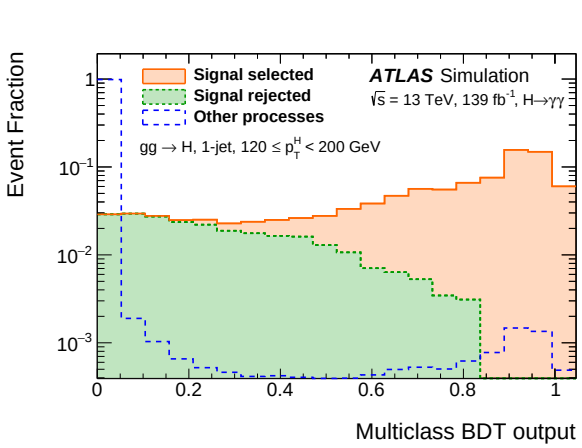
PoS(ICHEP2022)526

1. Introduction

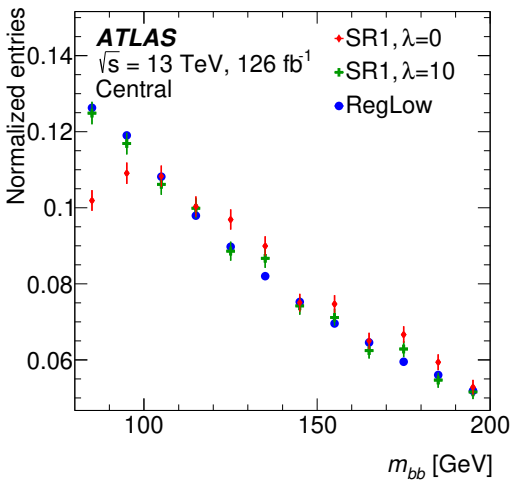
Following the discovery of the Higgs boson [1–4] at the Large Hadron Collider [5] (LHC) in 2012 by the ATLAS [6] and CMS Collaborations [7, 8], precision measurements of the Higgs boson and its properties have become increasingly important as a probe of the Standard Model (SM). This migration towards precision measurements has consequently lead to significant effort in the field of data analysis, with particular emphasis on analysis techniques that maximise information extraction from the recorded proton-proton collision data (e.g. machine learning [9]). The following article will therefore summarise five use cases of data analysis techniques in the sub-fields of machine learning classification (Section 2) and regression (Section 3), object embedding (Section 4), and multi-dimensional re-weighting (Section 5).

2. Machine Learning: Classification

$H \rightarrow \gamma\gamma$: Ever since the discovery of the Higgs boson [7] the di-photon Higgs decay mode ($H \rightarrow \gamma\gamma$) has been a key decay channel for measuring the Higgs boson properties. The latest analysis studying this decay mode, Ref. [10], concentrates on precisely measuring production cross-sections of the Higgs boson within the simplified template cross-section (STXS) paradigm [12]. With 44 STXS cross-section categories split according to production mode, di-jet invariant mass (m_{jj}), jet multiplicity, and transverse momentum (p_{T}) of the Higgs/vector bosons, maximising analysis sensitivity is achieved via a D-optimality (determinant) criterion technique.



(a) $H \rightarrow \gamma\gamma$ [10]



(b) VBF $H \rightarrow b\bar{b}$ [11]

Figure 1: Sub-figure 1a shows the multi-class BDT output score ($\max_{1\leq i\leq 44}(w_i \cdot z_i)$) for a single STXS category following the D-optimality criterion procedure, showing the correctly assigned (signal selected) and incorrectly rejected (signal rejected) signal, in addition to other signal processes accepted into the STXS category as a result of the BDT categorisation [10]. Sub-figure 1b shows the m_{bb} of the reconstructed $H \rightarrow b\bar{b}$ decay for the VBF $H \rightarrow b\bar{b}$ analysis [11] in the signal regions with/without adversarial training (SR $\lambda = 10/0$), and the control region (RegLow).

Specfically, Higgs signal events are distributed across the STXS categories via a multi-class boosted decision tree (BDT), trained with 44 class labels (y_i, where $i \in [1, 44]$) using categorical cross-entropy. The BDT assigns STXS class labels (z_i) to each event based on the maximum predicted class label, $\max_{1 \leq i \leq 44}(z_i)$. However, as the statistical power of each STXS bin is a combination of the signal to noise ratio, statistical error, and theoretical/experimental uncertainties, optimising the analysis BDT output based on the aforementioned statistical features would lead to maximum information retention. Therefore, the BDT output scores are transformed by a scalar quantity w_i, according to $\max_{1 \leq i \leq 44}(w_i \cdot z_i)$, which are then optimised by minimising the D-optimality criterion:

$$D_{\text{opt.}} = \frac{1}{2} \cdot \log\left(\frac{|C_{\text{exp}} + C_{\text{theo}}|}{|C_{\text{exp}}|}\right), \tag{1}$$

where $C_{...}$ represents the covariance matrix of the experimental (exp) and theoretical (theo) uncertainties obtained by single-bin asimov fits (signal+background MC) in each of the 44 STXS categories. Using the Powell algorithm for multi-variable function minimisation [13], the weights of each class label w_i are iteratively optimised from starting values of 1.0 in such a way that D-optimality criterion is minimised. In essence, as the inverse of the determinant of the covariance matrix is a measure of retained information, minimising the D-optimality criterion results in maximum information retention. Figure 1a shows for the $gg \rightarrow H$ production mode with 1-jet in the $120 \leq p_{\text{T}}^{H} \leq 200$ GeV region the true signal that is accepted and rejected based on the transformed BDT output, and the remaining STXS signal components incorrectly assigned to the aforementioned category.

VBF $H \rightarrow b\bar{b}$: Measurement of the vector boson fusion (VBF) production mode of the Higgs boson in which the Higgs decays to bottom quarks, as in the recent VBF $H \rightarrow b\bar{b}$ analysis (Ref. [11]), has two background processes: non-resonant QCD multi-jet, and resonant $Z(b\bar{b})$ + jets . Separating these backgrounds for the signal is achieved by an adversarial neural network (ANN), in order to decorrelate the classifier score $p(y_i)$ of an event with the invariant mass of the Higgs candidate, or more precisely the di-jet invariant mass m_{bb}. The neural network is constructed using a multi-layer perceptron (MLP) as a classifer trained using cross-entropy with two class label:

$$L = -\frac{1}{N}\sum_{i}^{N}\sum_{j}^{c} y_{i,j}\log(p(y_{i,j}))), \tag{2}$$

where $y_{i,j}$ is the class label of the i^{th} event and j^{th} class, and $p(y_{i,j})$ is the predicted probability of class y_i. The adversary is trained to predict from the classifier output score the corresponding m_{bb} bin in which the event is placed, where a total of ten bins in the range of 70 – 200 GeV are constructed. This is achieved by training the adversary on categorical cross-entropy with 10 classes ($c \in [1, 10]$, one class per m_{bb} bin). The total loss of the combined system is then defined as $L = L_{\text{class.}} - \lambda \cdot L_{\text{adv.}}$, where $L_{\text{class.}}$ is the classifier loss, and $L_{\text{adv.}}$ is the adversary regularised by a hyperparameter λ. The aim of ANN is to reduce sculpting of the m_{bb} distribution when cutting on the ANN score to define signal/control regions. Figure 1b shows for the central region of the analysis the m_{bb} spectrum of the signal region (SR1) with ($\lambda = 10$) and without ($\lambda = 0$) the adversary MLP, in addition to the control region (RegLow). As the RegLow is defined by a cut on

the ANN score of < 0.56 and SR1 is defined by an ANN cut of > 0.86 it can be seen that that the adversarial setup successfully decorrelates the ANN score from the m_{bb}.

3. Machine Learning: Quantile Regression

$H \rightarrow Z^*Z$: The Higgs decay to two Z-bosons ($H \rightarrow Z^*Z$) has since 2012 [7] been a key decay channel when measuring properties of the SM, with particular interest placed on the Higgs mass precision as given by the most recent iteration of the analysis, Ref. [10]. The analysis strategy employs an unbinned profile likelihood fit with the generic form:

$$\mathcal{L}(m_H \mid \vec{x}) = \mathcal{L}(m_H \mid m_{4l}, D, \sigma) = \prod_{i}^{N_{evt.}} P(m_{4l}^i, D^i, \sigma^i \mid m_H), \tag{3}$$

where m_H is the Higgs boson mass, m_{4l} is the reconstructed invariant mass of the four lepton system, σ is the resolution of the four lepton system, D is the analysis neural network score used to discriminant signal from SM backgrounds, and the index i runs over all events $N_{evt.}$. In practice constructing a 3-dimensional probability density function (pdf) to estimate m_H is infeasible, therefore the pdf's are broken down into conditional components with the signal pdf model taking the form:

$$P_s(m_{4l}^i, D^i, \sigma^i \mid m_H) = P(m_{4l} \mid D, \sigma, m_H) \cdot P(D \mid \sigma, m_H) \cdot P(\sigma \mid m_H). \tag{4}$$

The pdf of the signal in the reconstructed analysis discriminant m_{4l}, $P(m_{4l} \mid D, \sigma, m_H)$, is expressed as a double sided crystal ball (DCB) function which is dependent on the four lepton system resolution. Since the pdf is conditional on the four lepton system, and given the asymmetric nature of the DCB function, the 1σ variance of the individual lepton resolutions does not translate into a $1\sigma_{\text{DCB}}$ interval, and the lepton momentum resolution errors are underestimated by approximately 10-20%.

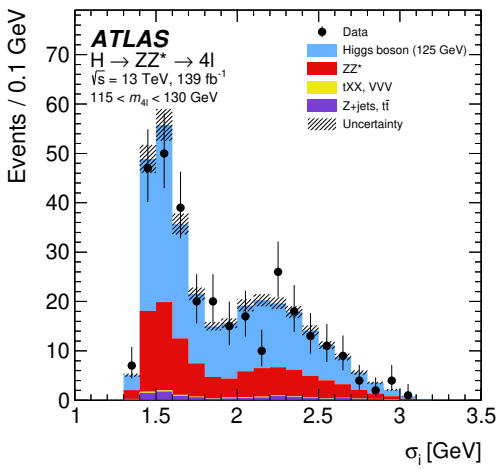


Figure 2: Shows the per event m_{4l} predicted by the QRNN for data (points), signal (Higgs Boson) and the estimated background [10].

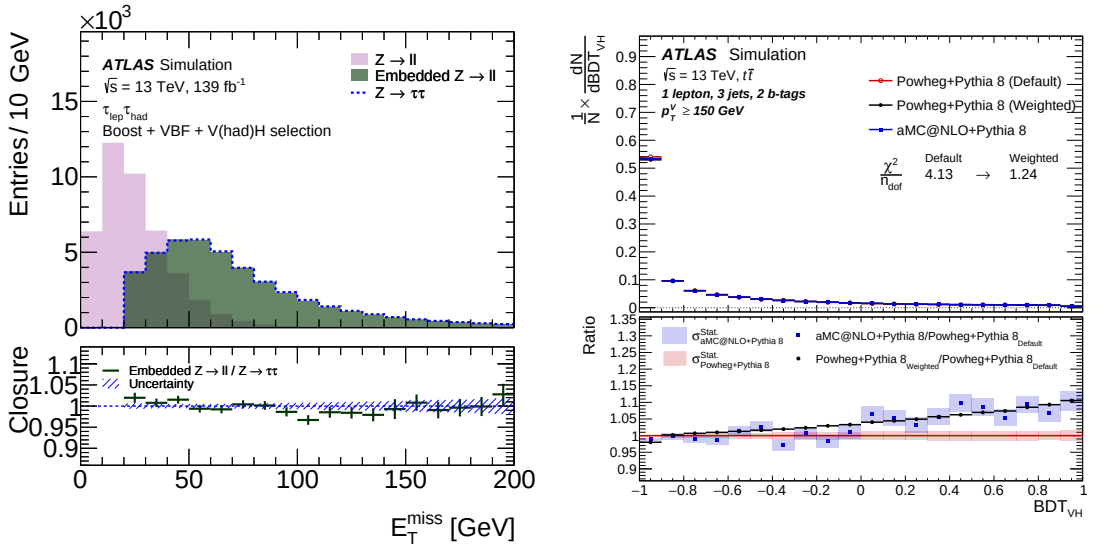
To combat this a per event resolution, σ^i, is estimated using a quantile regression neural network that estimates an error quantile for the $O = |m_{4l}^{\text{const.}} - m_{4l}^{\text{truth,born}}|$ observable. The neural network architecture (QRNN) is constructed from two MLP blocks and a recurrent neural network (RNN) block. The first MLP is fed the transverse momentum of the constrained four lepton system ($p_{\text{T},4l}^{\text{const.}}$) and $\sigma_{m_{4l}}^{\text{const.}}$, which in tandem with the RNN block that is fed the four vector information of the four leptons ordered by transverse momentum, creates a latent space representation that is aggregated by the second MLP to predict the target quantile. This is achieved by the minimisation of the loss function:

$$\text{Loss} = \sum_{i}^{N} max\{q \cdot (O_{\text{true}} - O_{\text{predicted}}), (1-q) \cdot (O_{\text{true}} - O_{\text{predicted}})\} + \frac{1}{w} \sum_{i}^{N} \left(\frac{\sigma_{m_{4l}}^{\text{predicted}}}{\sigma_{m_{4l}}^{\text{const.}}} - \mu \right)^2, \quad (5)$$

where q is the target quantile (0.683 = 68.3% quantile), $O_{\text{true/predicted}}$ is the true/predicted difference of the constrainted four lepton system mass difference between reconstructed and born truth level defintion, and w/μ are the hyper-parameters set during network optimisation. The QRNN is then used to assign each event a four lepton mass resolution value corresponding to the 68.3% quantile 1σ error definition. Figure 2 shows the final σ_i predictions for signal, background, and data, used in the analysis.

4. Object Embedding

VBF $H \to \tau\tau$: For the analysis of vector boson fusion (VBF) production of a Higgs boson that decays to τ-leptons (Ref. [14], referred to as VBF $H \to \tau\tau$), the irreducible $Z(\tau\tau)$+jets background MC prediction suffers from modelling inaccuracies due to the simulation of τ-decays inside the ATLAS detector volumne. To correct this a *kinematic object embedding* procedure was adopted, akin to that of the b-jet embedding procedure of Ref. [11]. In this instance however, instead of replacing electrons/muons in $Z(ee/\mu\mu)$+jets data events with simulated τ-decays, the four vector of the electrons/muons are scaled by a scalar quantity. This scalar quantity is estimated from simulated



(a) VBF $H \to \tau\tau$ [14]

(b) VH $H \to b\bar{b}$ [11]

Figure 3: Sub-figure 3a shows as a function of $E_{\text{T}}^{\text{miss}}$ simulated $Z(l^{\pm}l^{\mp})$ + jets and $Z(\tau\tau)$ + jets processes against simulated $Z(l^{\pm}l^{\mp})$ + jets with kinematically embedded τ-decay emulation [14]. Sub-figure 3b shows the MC prediction of top-quark pair production using Powheg+Pythia 8 and MadGraph5_aMC@NLO+Pythia 8 as a function of the $VH(b\bar{b})$ [15] analysis BDT discriminant. The multi-dimensionally mapped Powheg+Pythia 8 (weighted) matches within statistical errors the target MadGraph5_aMC@NLO+Pythia 8 distribution.

τ-decays, with the intention of emulating invisible τ-decay components such as neutrinos. The missing energy as a result of the invisible decay products are then introduced into the event topology as a missing transverse energy ($E_{\mathrm{T}}^{\mathrm{miss}}$) component to ensure energy-momentum conservation within the event. Figure 3a shows the emulated $E_{\mathrm{T}}^{\mathrm{miss}}$ differential cross-section for simulated $Z(l^{\pm}l^{\mp})$+jets , $Z(\tau\tau)$+jets , and the embedded $Z(l^{\pm}l^{\mp})$+jets sample demonstrating via the closure of the embedded $Z(l^{\pm}l^{\mp})$ + jets with the dedicated $Z(\tau\tau)$ + jets simulation the successful emulation of $E_{\mathrm{T}}^{\mathrm{miss}}$.

5. Multi-dimensional Re-weighting: Statistical Precision

VH H → bb̄ : Within the frequentist statistical paradigm (e.g. profile likelihoods [16]) the statistical precision of an expectation value is often dictated by Monte Carlo sample sizes. The production of MC predictions involves both a *nominal* expectation used in a likelihood fit, in addition to *alternative* samples used to evaluate MC model uncertainties. These alternative samples however are often generated with a smaller effective luminosity due to limited CPU resources. Consequently, differential cross-sections predicted by alternative MC samples often suffer from a reduced statistical precision. To compensate for this, often a re-weighting approach is adopted in which the nominal MC configuration is *weighted* event-by-event to match the prediction of the alternative. This is achieved by a mapping function $f(\vec{x})$ defined by the left hand side of equation 6:

$$f(\vec{x} \mid \vec{\theta}) = \frac{P^{\text{alt.}}(\vec{x} \mid \vec{\theta}_{\text{alt.}})}{P^{\text{nom.}}(\vec{x} \mid \vec{\theta}_{\text{nom.}})} = \frac{P^{\text{alt.}}(s(\vec{x}) \mid \vec{\theta}_{\text{alt.}})}{P^{\text{nom.}}(s(\vec{x}) \mid \vec{\theta}_{\text{nom.}})} = f(s(\vec{x}) \mid \vec{\theta}), \tag{6}$$

where $P^{\text{nom./alt.}}(\vec{x} \mid \vec{\theta})$ is the probability density function of the nominal/alternative MC prediction for a phase space point defined by the vector $\vec{x}$, conditioned on the parameters of the model $\vec{\theta}$. Obtaining this mapping function however is problematic for a p-dimensional input space $\vec{x} \in \mathbb{R}^p$, due to the curse of dimensionality in which the statistical precision of the sampled space via Monte Carlo methods is diluted. To address this problem, a density ratio estimator in the form of a calibrated BDT was employed in the analysis of a Higgs boson produced in association with a SM vector boson for $H \to b\bar{b}$ (Ref. [15]). Utilising a classification BDT trained on the nominal and alternative samples, the mapping function $f(\vec{x})$ is obtained using the equivalence principle shown by the right hand side of equation 6, where $s(\vec{x})$ is the BDT output classifier score. The above density ratio is then calibrated using a binned technique similar to that outlined in Ref. [17].

Figure 3b illustrates the application of the multi-dimensional mapping function $f(s(\vec{x}) \mid \theta)$ to top quark pair production ($t\bar{t}$). The nominal MC sample for $t\bar{t}$ produced via Powheg+Pythia 8 is mapped using the above technique to match the alternative matrix element (ME) MC configuration given by MadGraph5_aMC@NLO+Pythia 8 , as shown by the weighted histogram. The analysis discriminant is shown as it aggregates 11-13 input variables in a BDT to classify the $VH(b\bar{b})$ signal against the SM backgrounds, thereby demonstrating the multi-dimensional mapping power of the technique.

6. Conclusion

Data analysis techniques in the fields of machine learning classification and regression, object embedding, and multi-dimensional probability density weighting have been shown across an array

of ATLAS analyses aimed at studying the Higgs boson. The techniques are designed to maximise the information extracted from LHC proton-proton collision data, with the specific goal of maximising analysis measurement precision or discovery potential in the Higgs sector [9].

References

[1] P. W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*, Phys. Rev. Lett. **13** (1964) 508–509.

[2] F. Englert and R. Brout, *Broken Symmetry and the Mass of Gauge Vector Mesons*, Phys. Rev. Lett. **13** (1964) 321–323.

[3] P. W. Higgs, *Broken symmetries, massless particles and gauge fields*, Physics Letters **12** (1964) 132 – 133.

[4] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, *Global Conservation Laws and Massless Particles*, Phys. Rev. Lett. **13** (1964) 585–587.

[5] A. Breskin and R. Voss, *The CERN Large Hadron Collider: Accelerator and Experiments*. CERN, Geneva, 2009. https://cds.cern.ch/record/1244506.

[6] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, JINST **3** (2008) S08003.

[7] ATLAS Collaboration, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, Physics Letters B **716** (2012) 1 – 29.

[8] CMS Collaboration, *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*, Physics Letters B **716** (2012) 30 – 61.

[9] A. Radovic, M. Williams, D. Rousseau, M. Kagan, D. Bonacorsi, A. Himmel, A. Aurisano, K. Terao, and T. Wongjirad, *Machine learning at the energy and intensity frontiers of particle physics*, Nature **560** (2018) 41–48.

[10] ATLAS Collaboration, *Measurement of the Higgs boson mass in the $H \to ZZ^* \to 4\ell$ decay channel using 139 fb^{-1} of $\sqrt{s}$ = 13 TeV pp collisions recorded by the ATLAS detector at the LHC*, arXiv:2207.00320 [hep-ex].

[11] ATLAS Collboration, *Measurements of Higgs bosons decaying to bottom quarks from vector boson fusion production with the ATLAS experiment at $\sqrt{s}$ = 13 TeV*, Eur. Phys. J. C **81** (2021) 537, arXiv:2011.08280 [hep-ex].

[12] LHC Higgs Cross Section Working Group, *Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector*, arXiv:1610.07922 [hep-ph].

[13] M. J. D. Powell, *An efficient method for finding the minimum of a function of several variables without calculating derivatives*, The Computer Journal **7** (1964) 155–162, https://academic.oup.com/comjnl/article-pdf/7/2/155/959784/070155.pdf.

[14] ATLAS Collaboration, *Measurements of Higgs boson production cross-sections in the $H \to \tau^+\tau^-$ decay channel in pp collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector*, JHEP **08** (2022) 175, arXiv:2201.08269 [hep-ex].

[15] ATLAS Collaboration, *Measurements of WH and ZH production in the $H \to b\bar{b}$ decay channel in pp collisions at 13 TeV with the ATLAS detector*, Eur. Phys. J. C **81** (2021) 178, arXiv:2007.02873 [hep-ex].

[16] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, *Asymptotic formulae for likelihood-based tests of new physics*, Eur. Phys. J. C **71** (2011) 1554, arXiv:1007.1727 [physics.data-an], [Erratum: Eur.Phys.J.C 73, 2501 (2013)].

[17] K. Cranmer, J. Pavez, and G. Louppe, *Approximating Likelihood Ratios with Calibrated Discriminative Classifiers*, arXiv:1506.02169 [stat.AP].

PoS(ICHEP2022)526