

Domain Adversarial Learning for Reducing Training Bias in ttH(bb) search at ATLAS

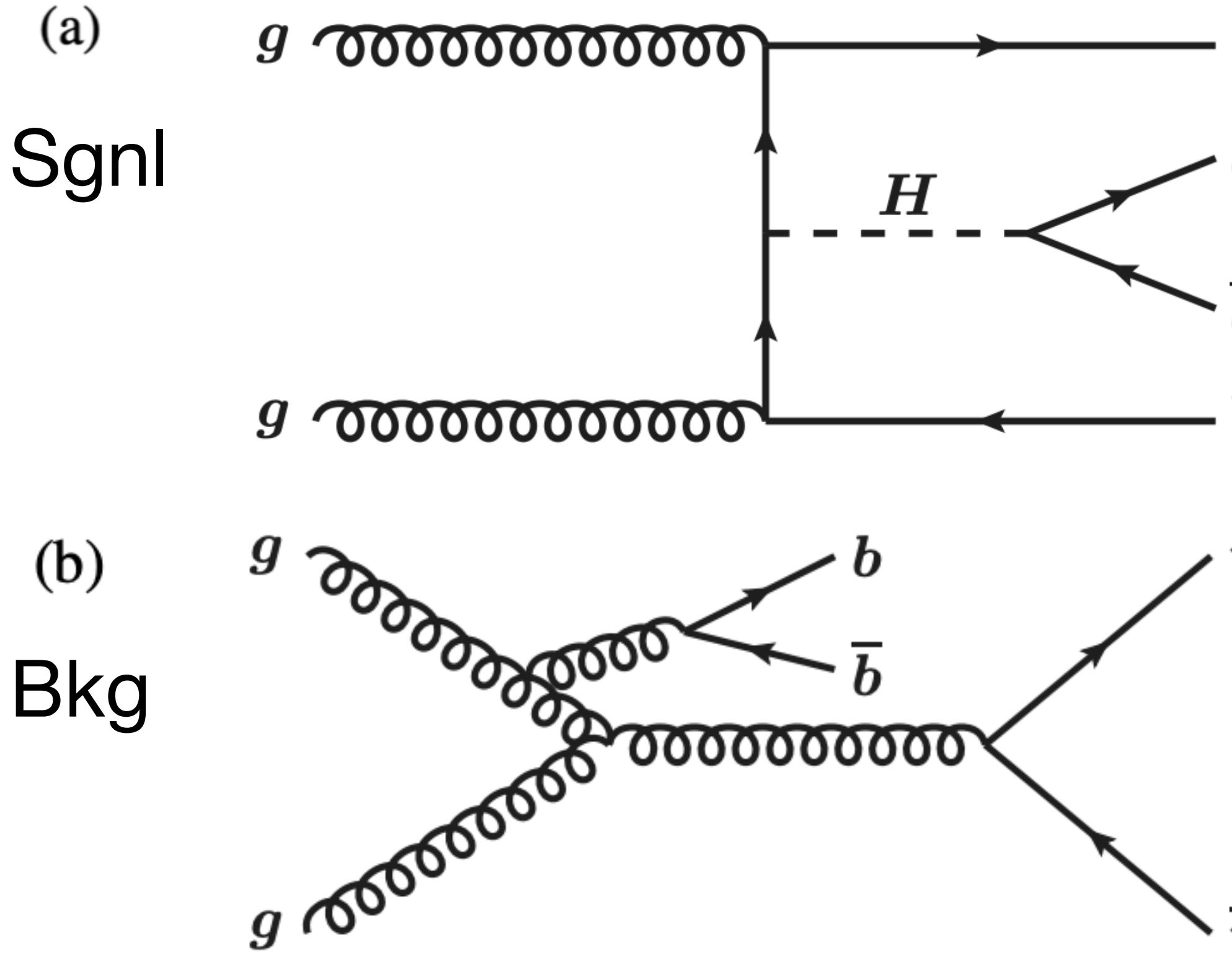
Fatkullin I. F.
Supervisors: Glaysher P., Katzy J., Pollard C.

ATLAS, DESY
2019

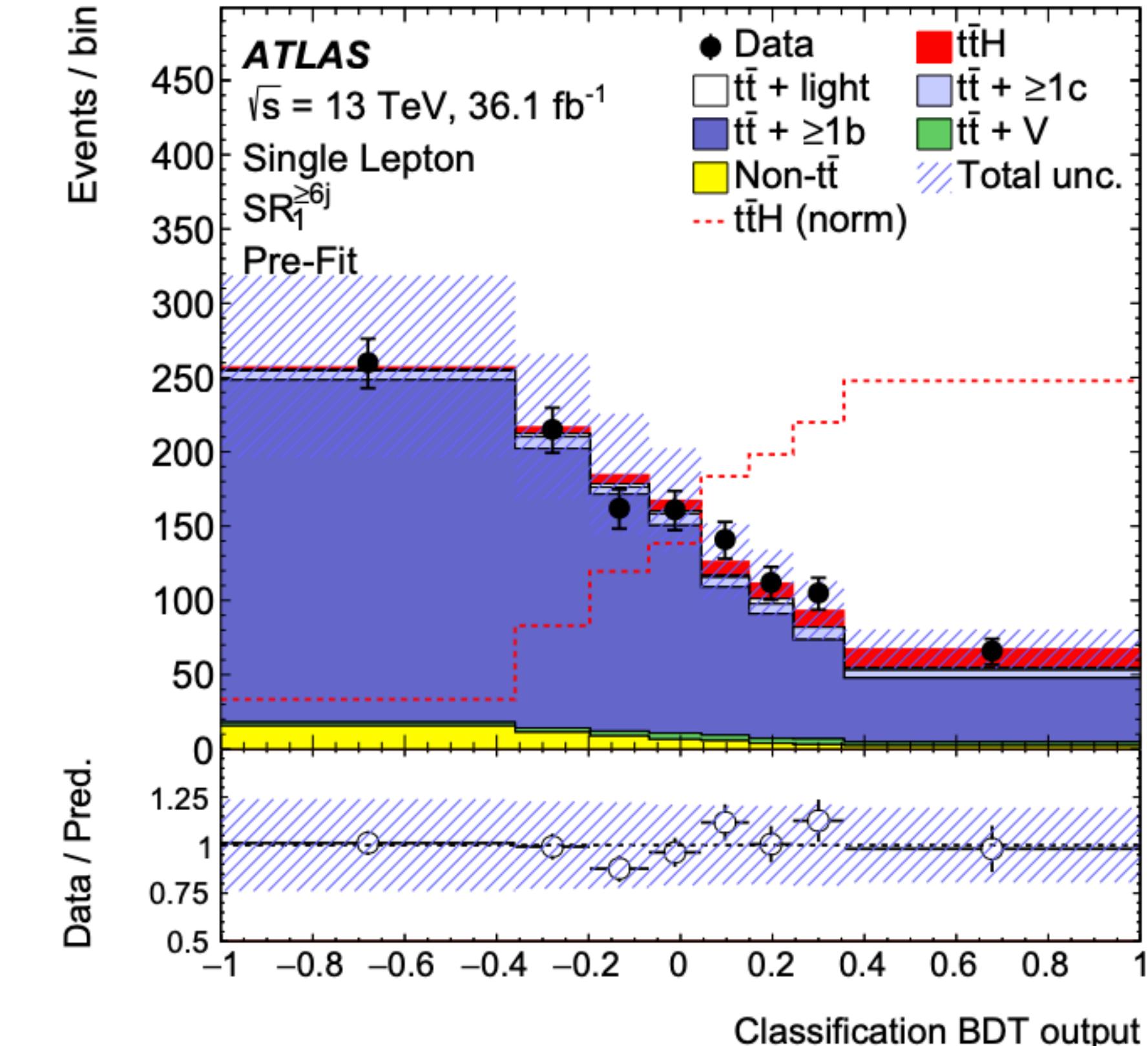
Outline

- Introduction
- Classification of $t\bar{t}H(bb)$ vs $t\bar{t}bb$ bkg
- Idea of Adversarial Domain Adaptation
- Results

ttH(bb) Production Search

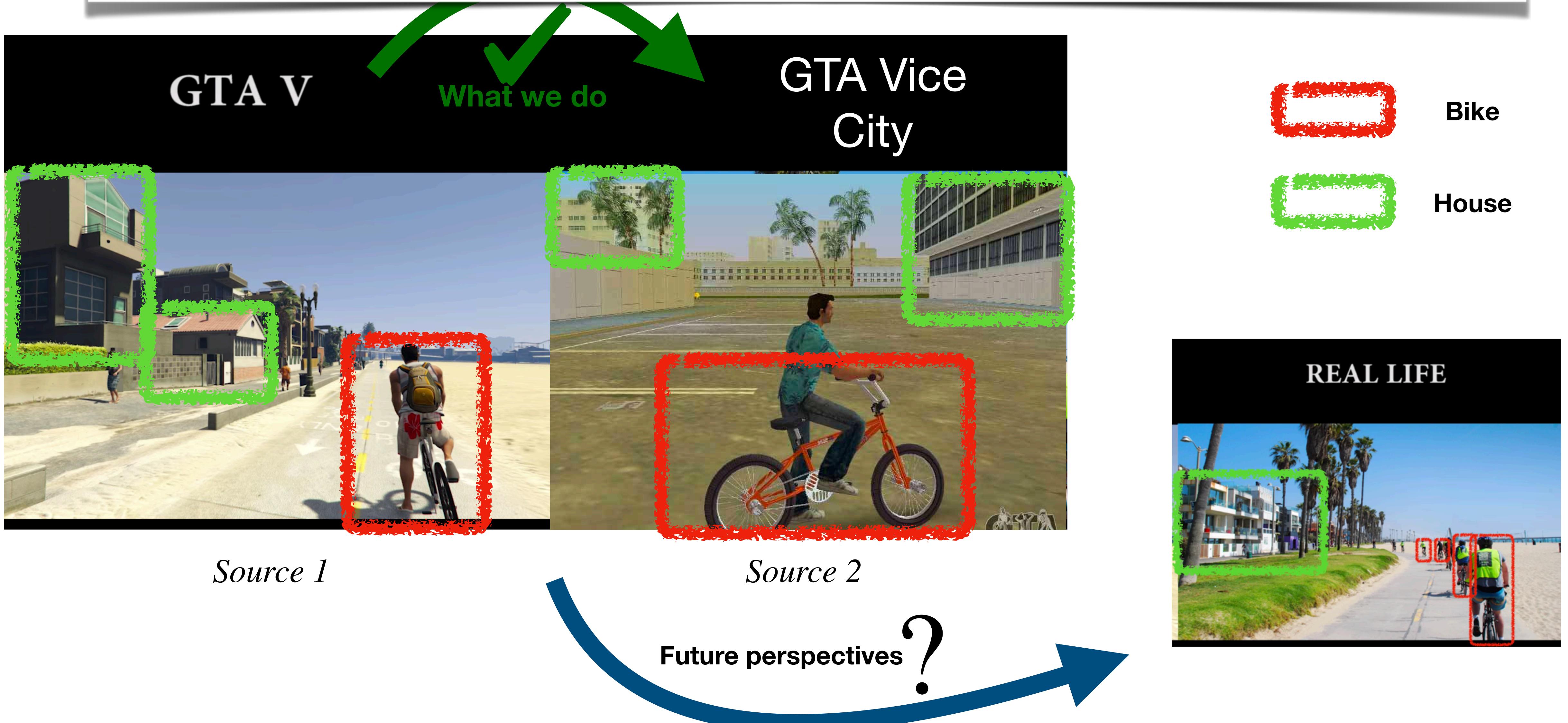


Feynman diagrams for a) ttH(bb) production b) tt+bb background

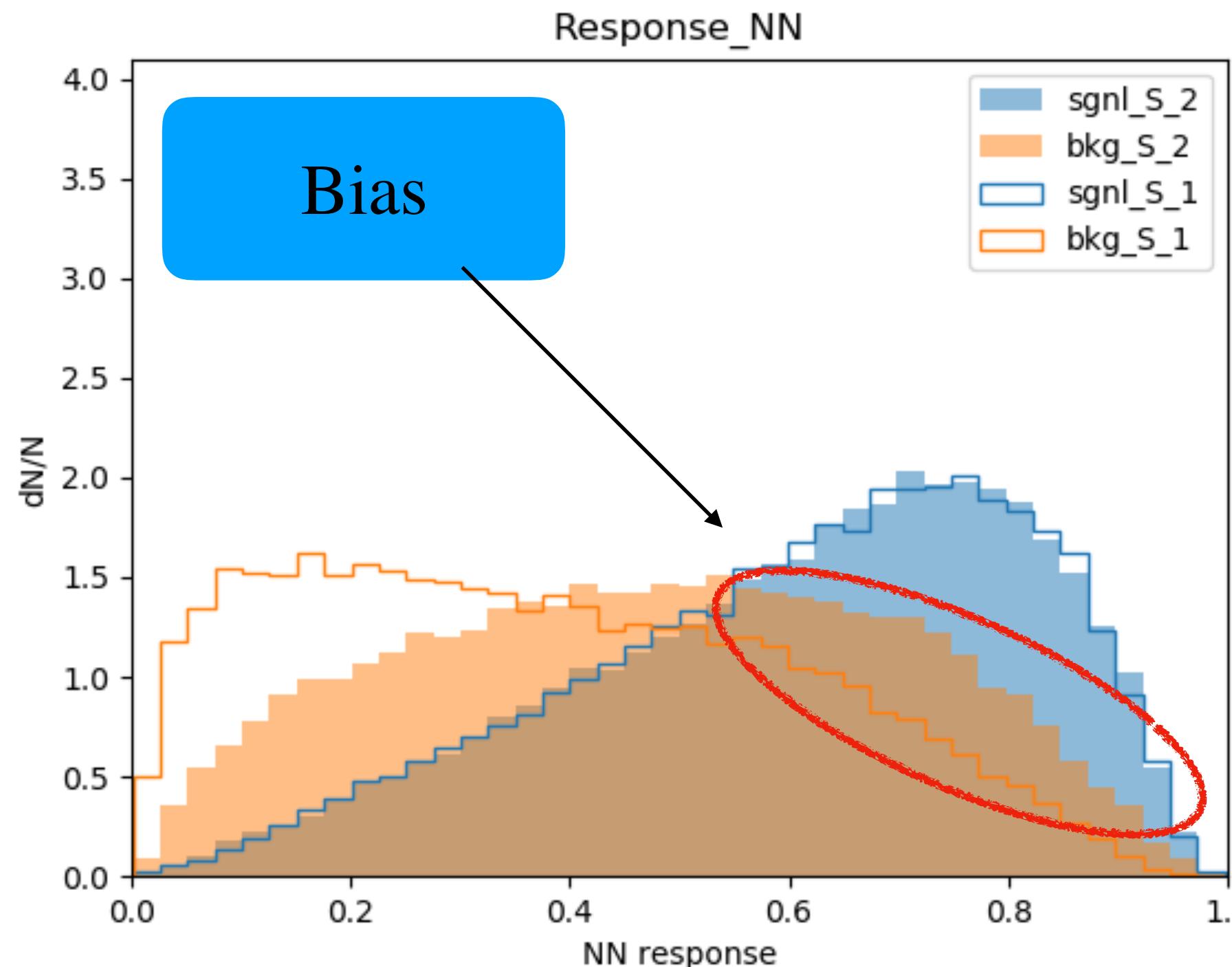


- Systematic uncertainties come because of different simulations
- The goal is to minimise this uncertainty while preserving good classification performance

Domain Adaptation Explained by Analogy



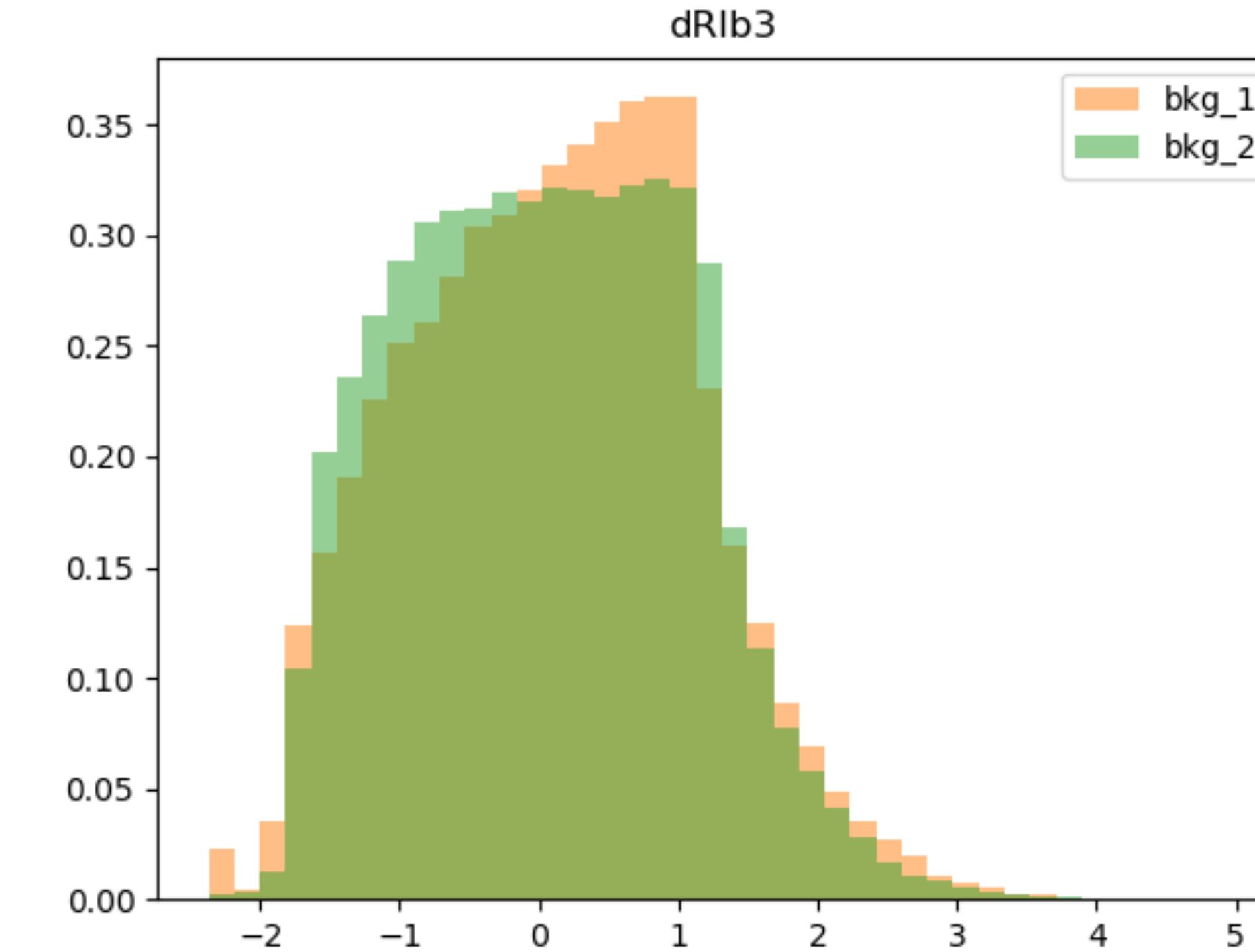
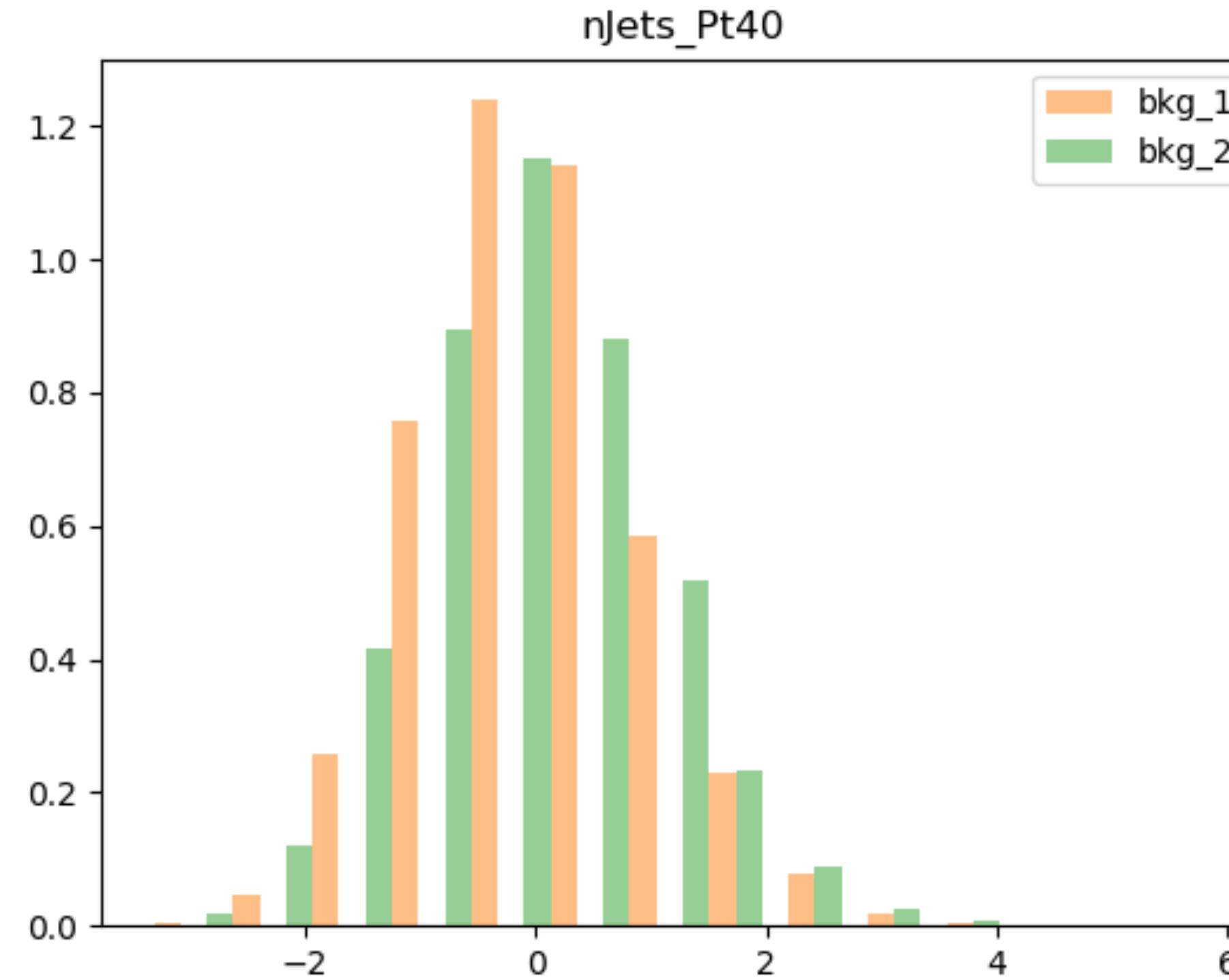
Motivation



Performance of baseline Feed Forward NN with 3 hidden layers

- We have one signal and two different background simulations
- It's been noted that if we train NN classifier on data with one background simulation and apply it to the test data with another background, its performance is reduced significantly
- This happens because backgrounds are slightly different
- Domain adaptation techniques [1] are hoped to reduce the training bias of NN and show good performance

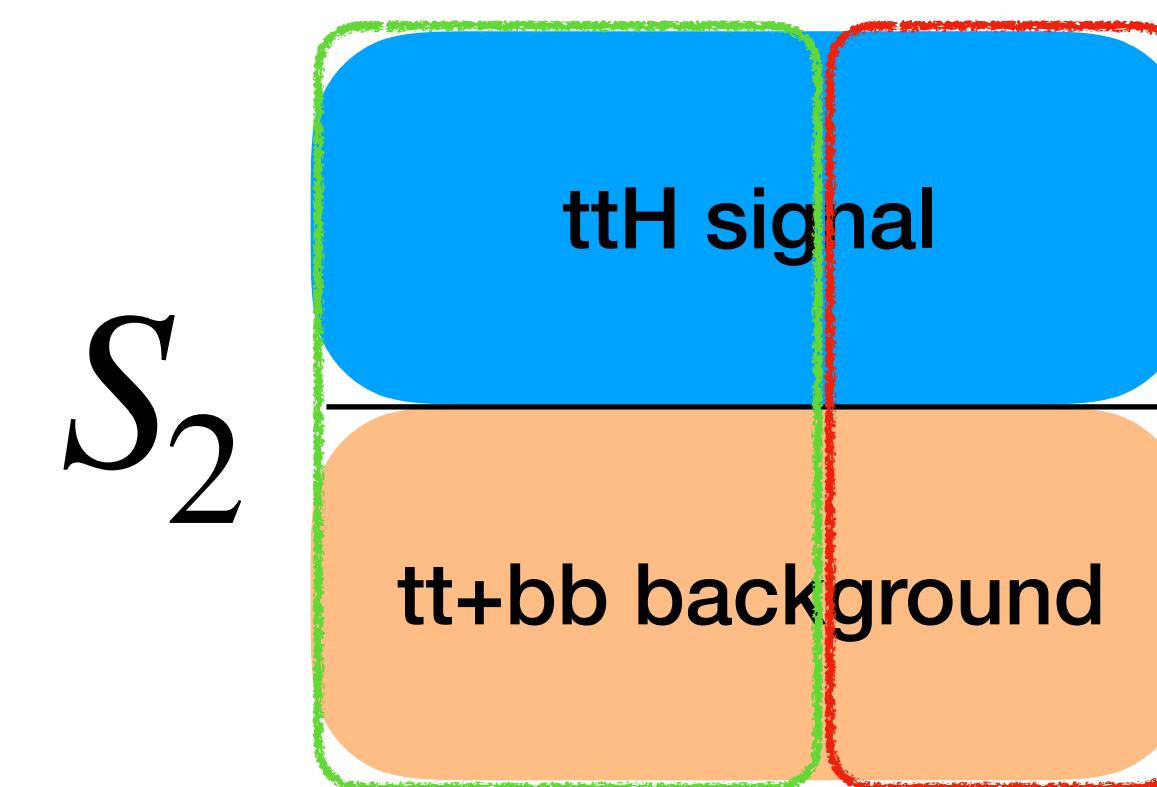
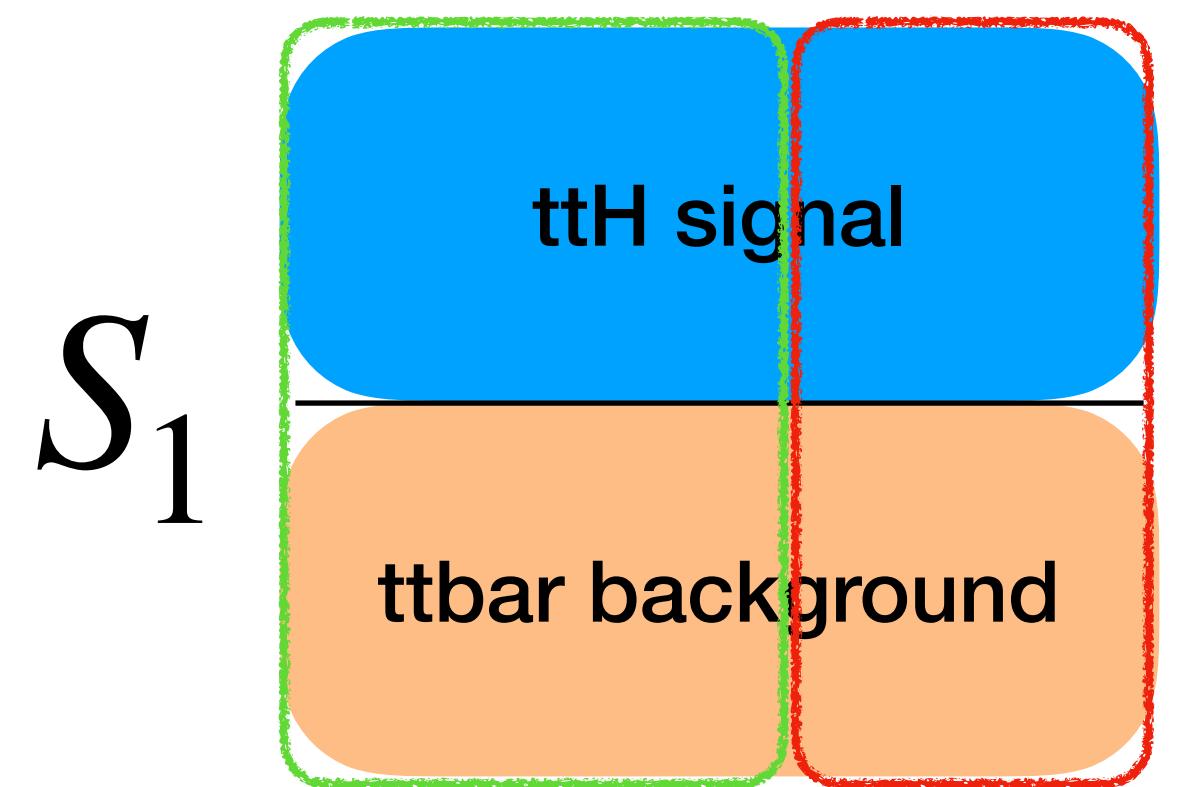
Background distributions



- $bkg_1 = ttbar$ background: *MadGraph / Pythia6*
- $bkg_2 = tt+bb$ background: *Powheg Pythia8*
- *Two background simulations are similar in distribution shapes, but there is some difference.*

Notations

- \mathcal{X} an input space (40 variables)
- $\mathcal{Y} = \{0,1\}$ a set of classes (if an event comes from signal or background)
- We define domain as a pair $\langle \mathcal{D}, f \rangle$ consisting of distribution \mathcal{D} on \mathcal{X} and a labelling function f on $\mathcal{X} \times \mathcal{Y}$.
- S_1 and S_2 denote two datasets obtained from domains Source 1 $\langle \mathcal{D}_{S_1}, f_{S_1} \rangle$ and Source 2 $\langle \mathcal{D}_{S_1}, f_{S_1} \rangle$

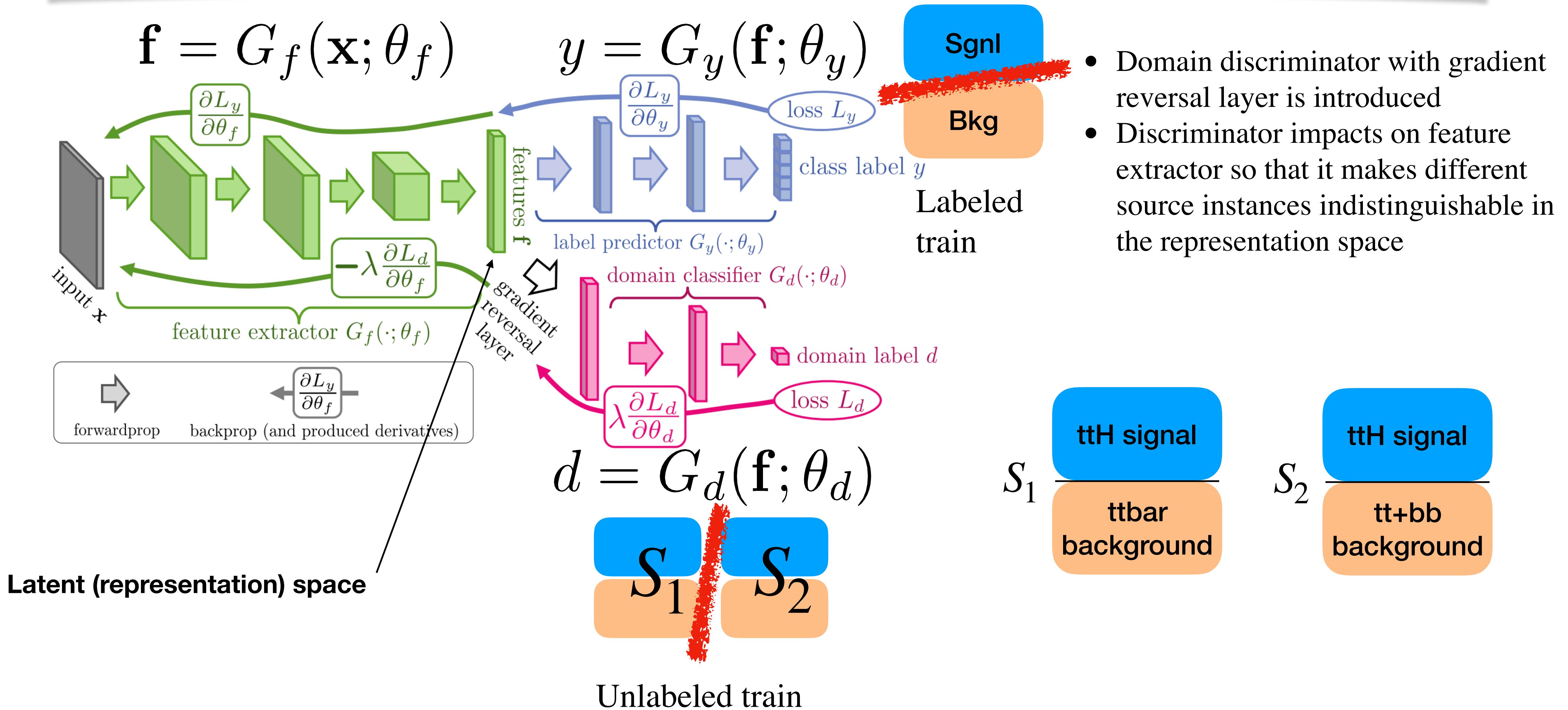


Source 1

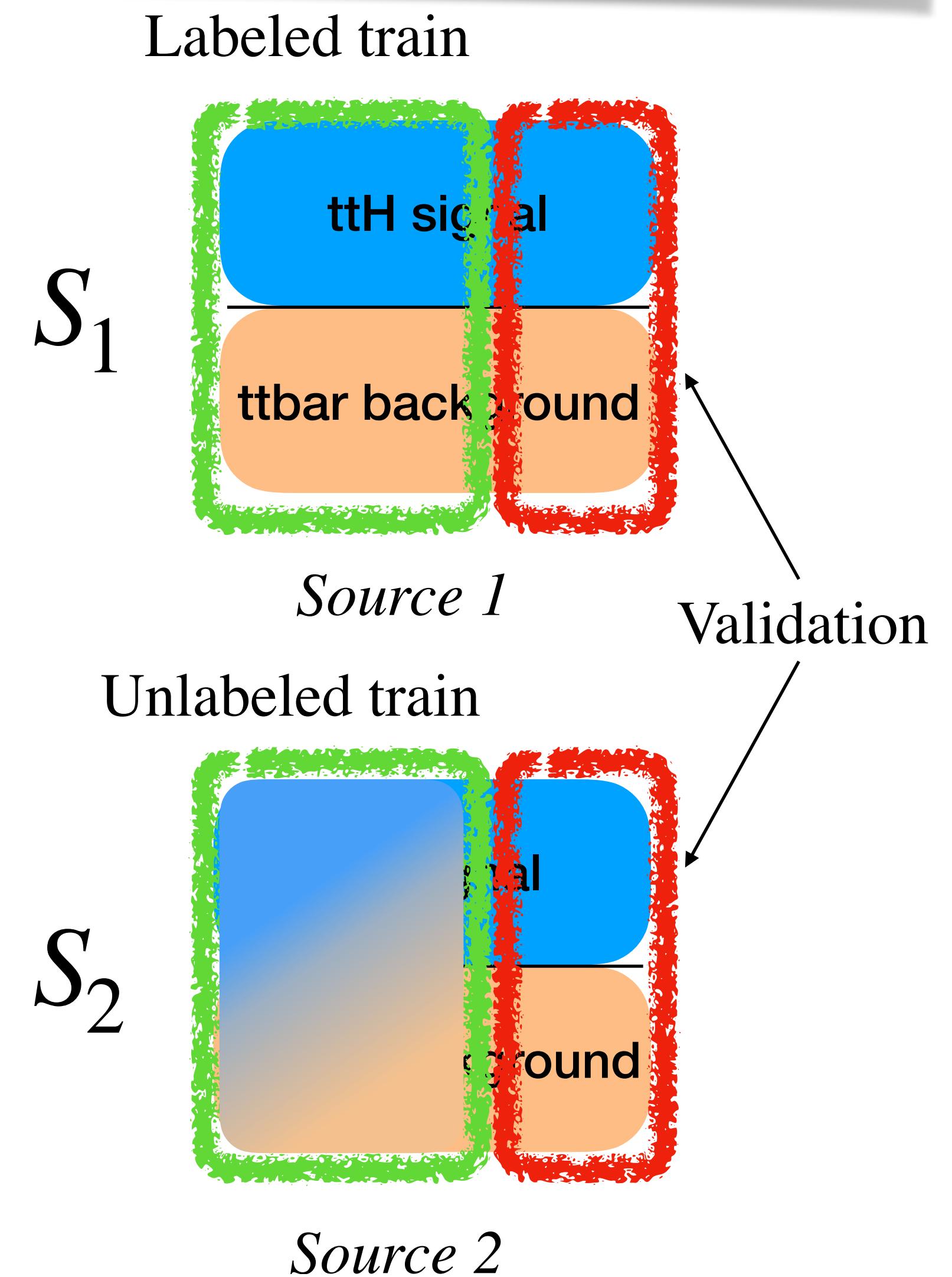
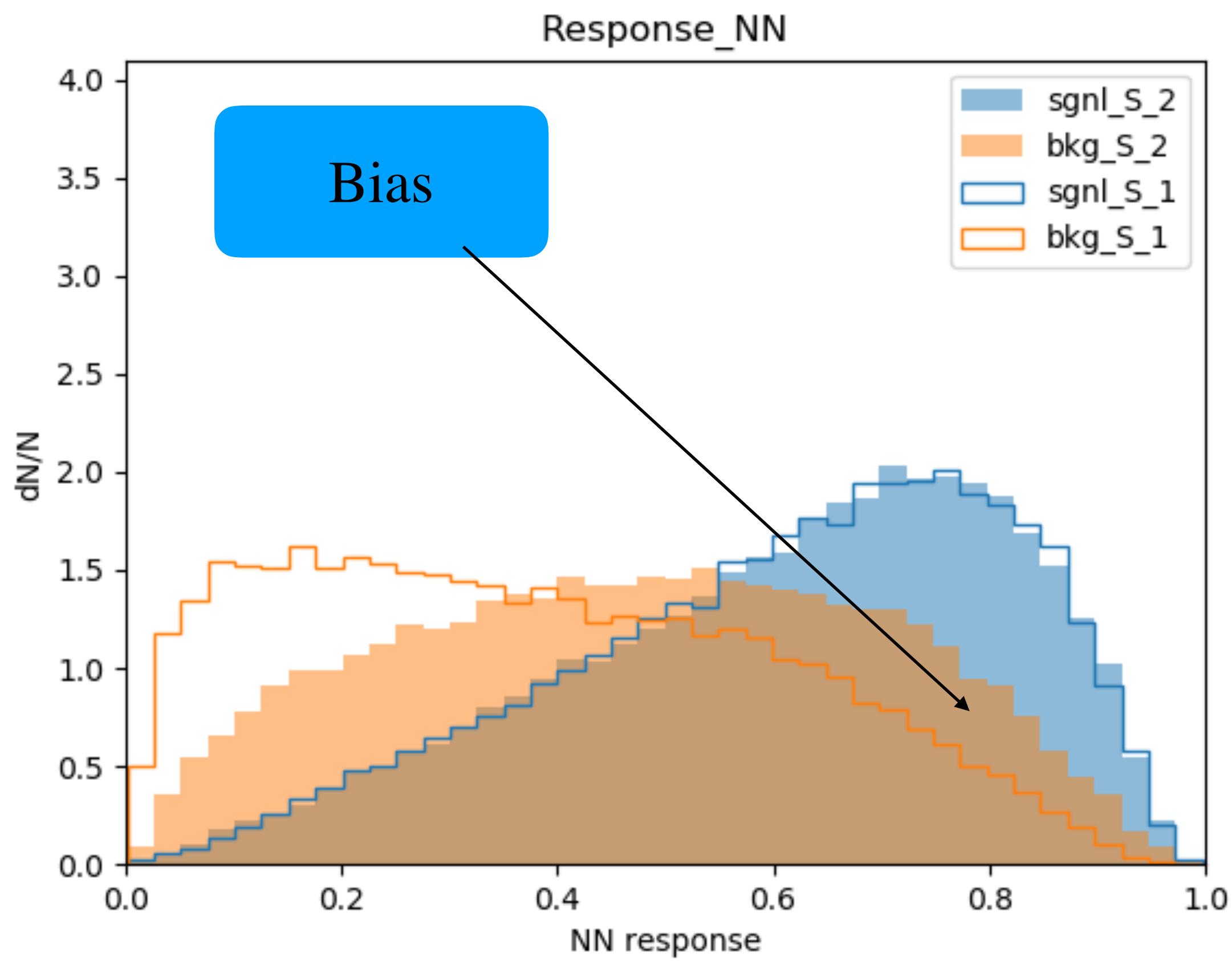
Source 2

- ttH signal: MadGraph/Herwig6
- ttbar background: MadGraph/Pythia6
- tt+bb background: Powheg Pythia8

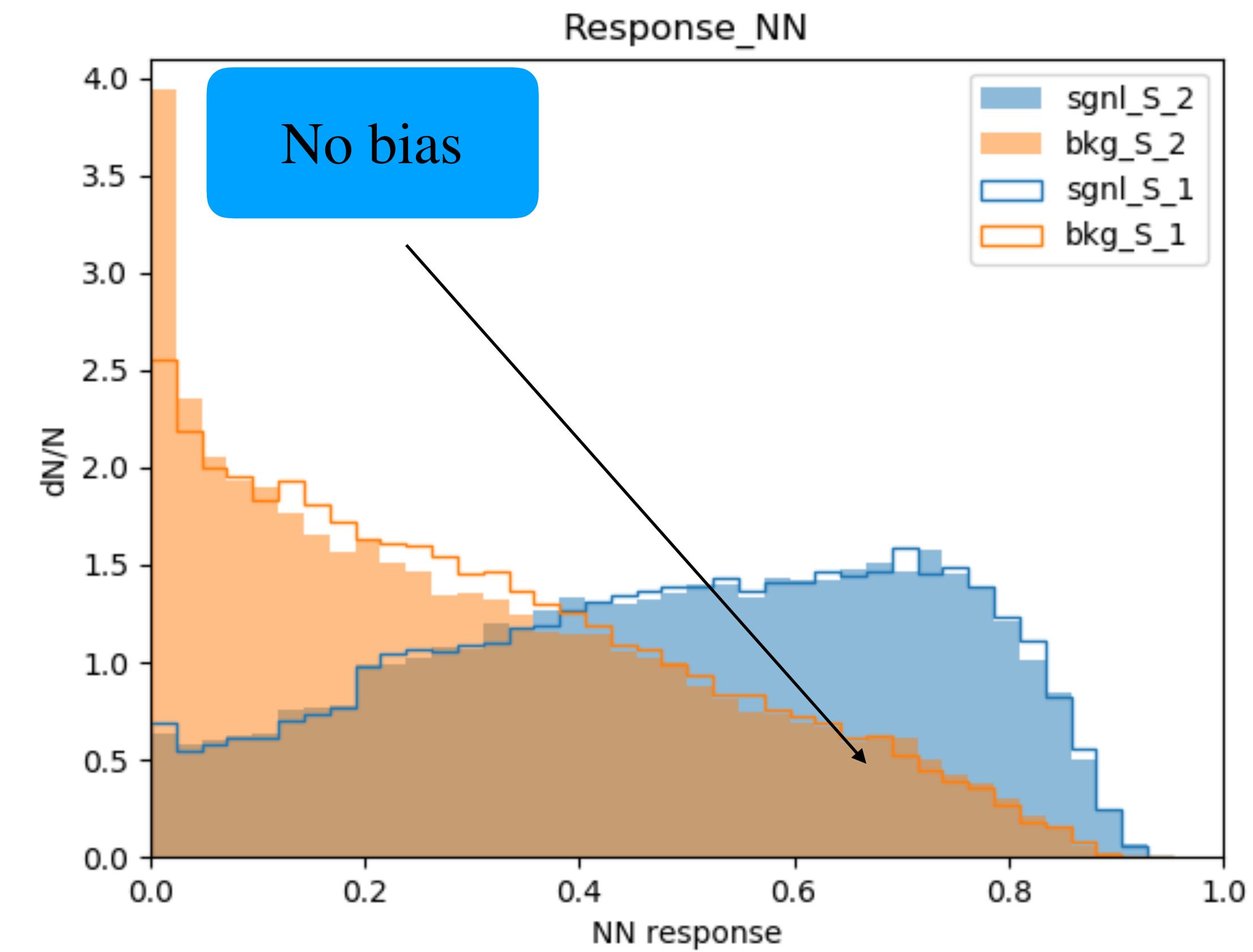
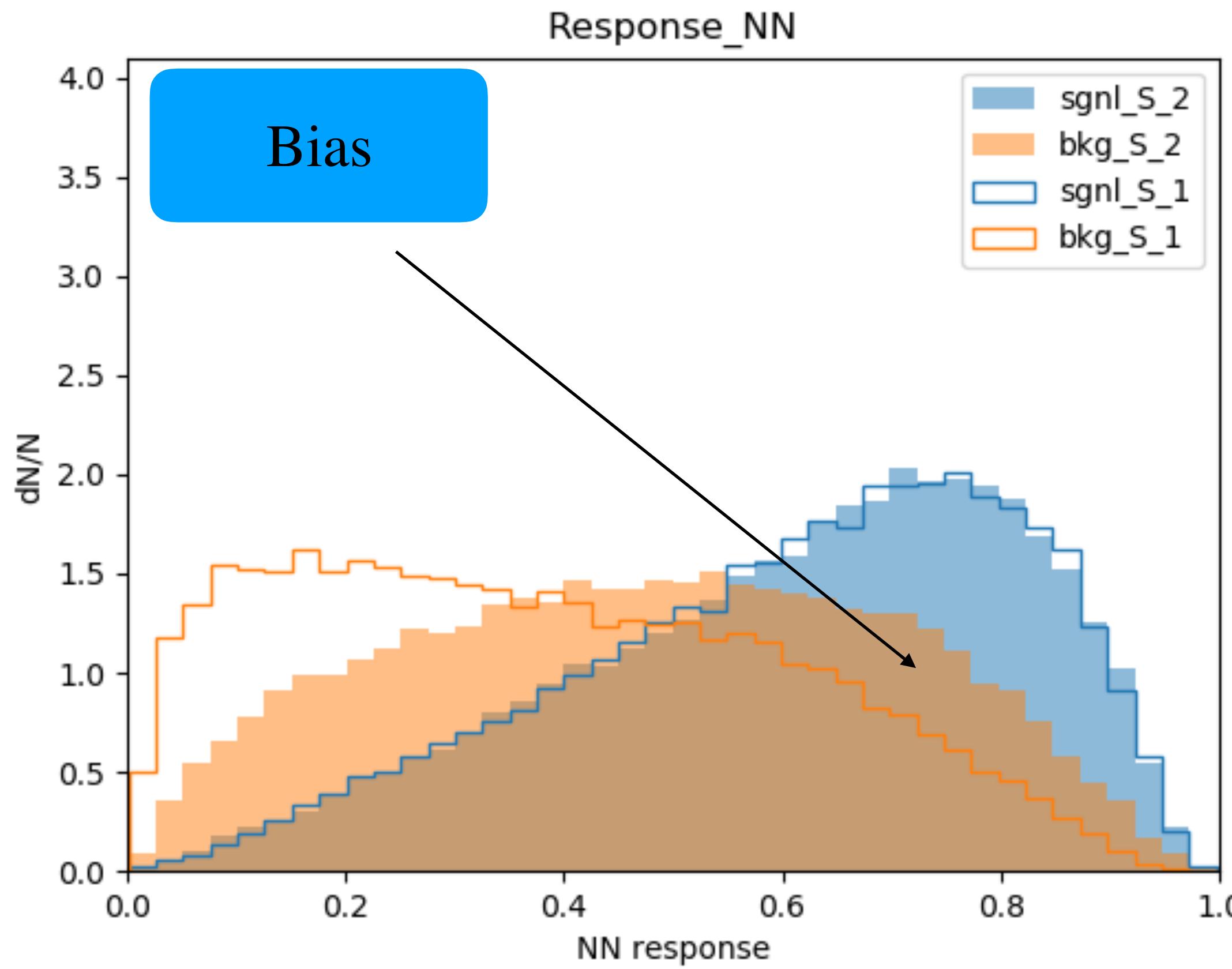
Main Idea



Results



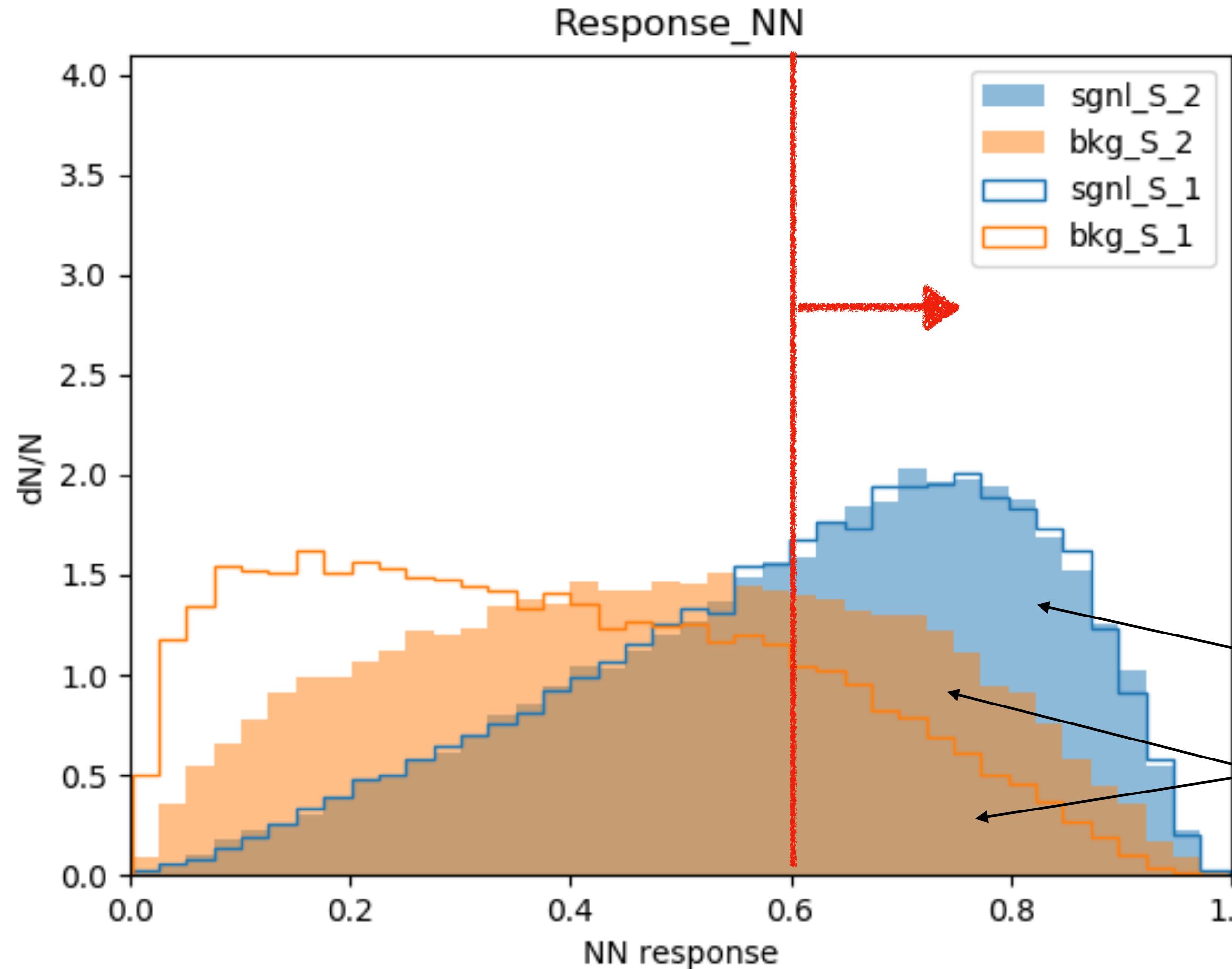
Results



- Discriminator importance parameter $\lambda = 0$

- Discriminator importance parameter $\lambda = 1.3$

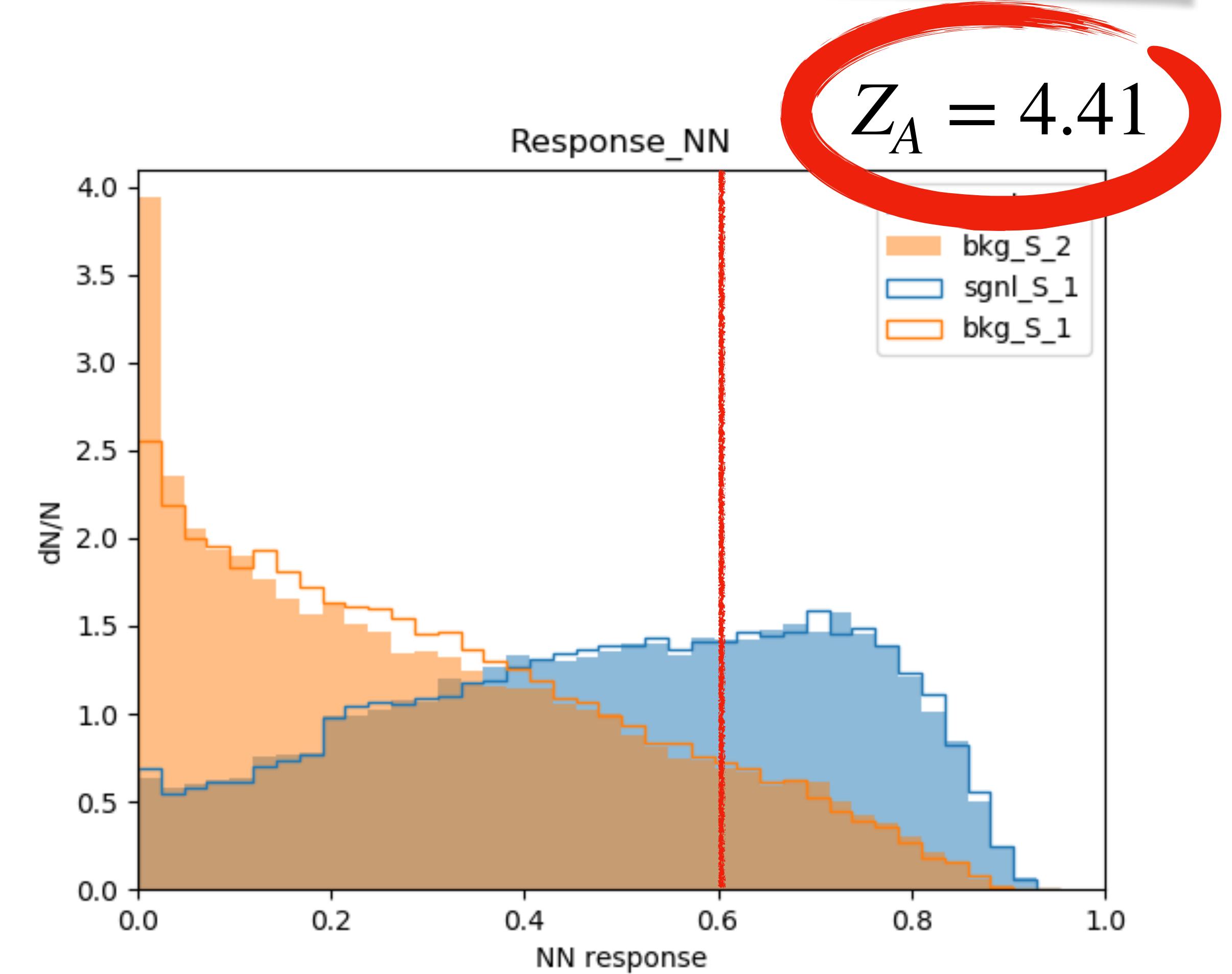
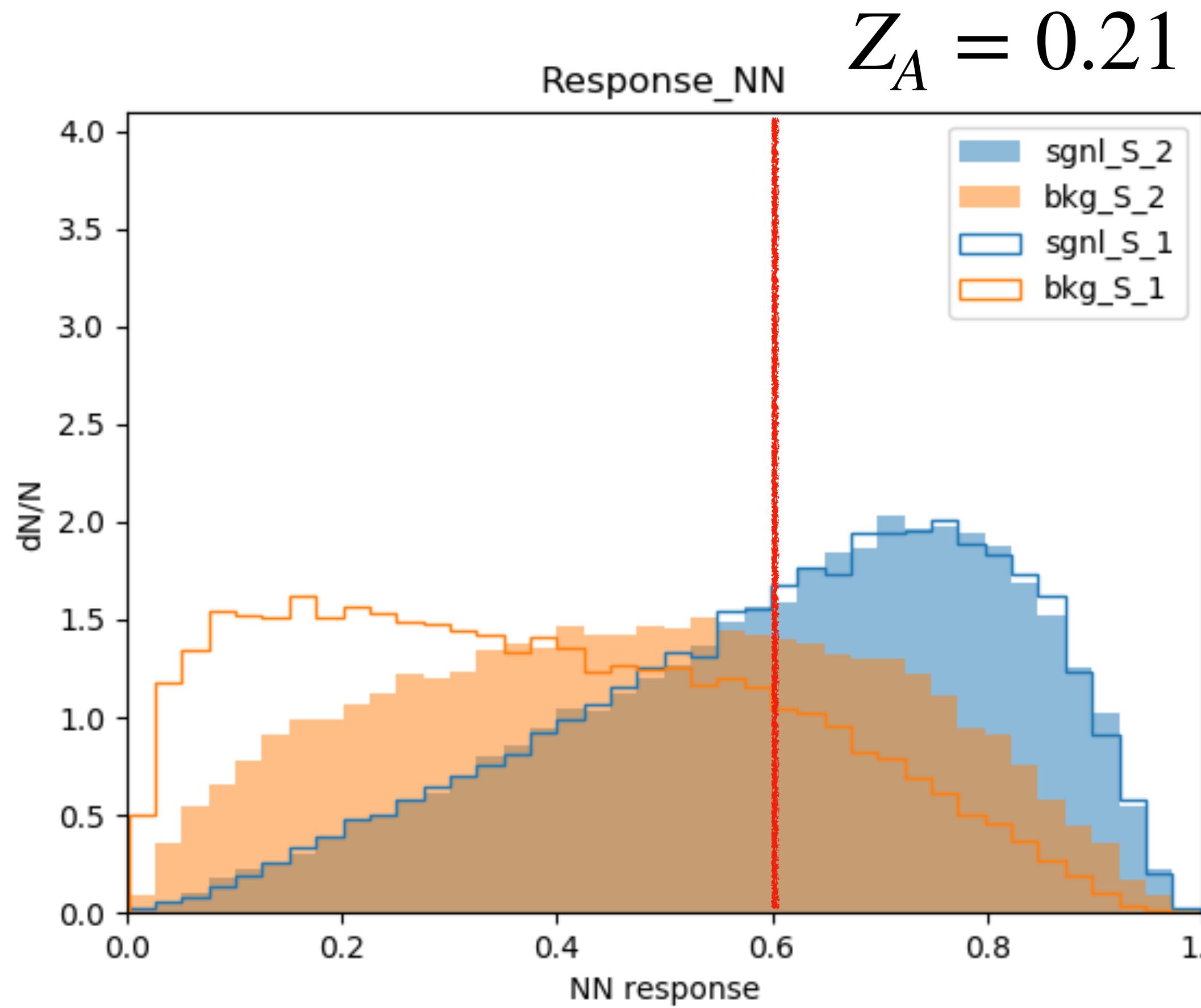
Significance Calculation



$$Z_A = \frac{s}{\sqrt{b + \sigma_b^2}}$$

- s - signal events in the cut
- b - test bkg events in the cut
- σ_b^2 - difference btw test and train bkg

Significance and Response

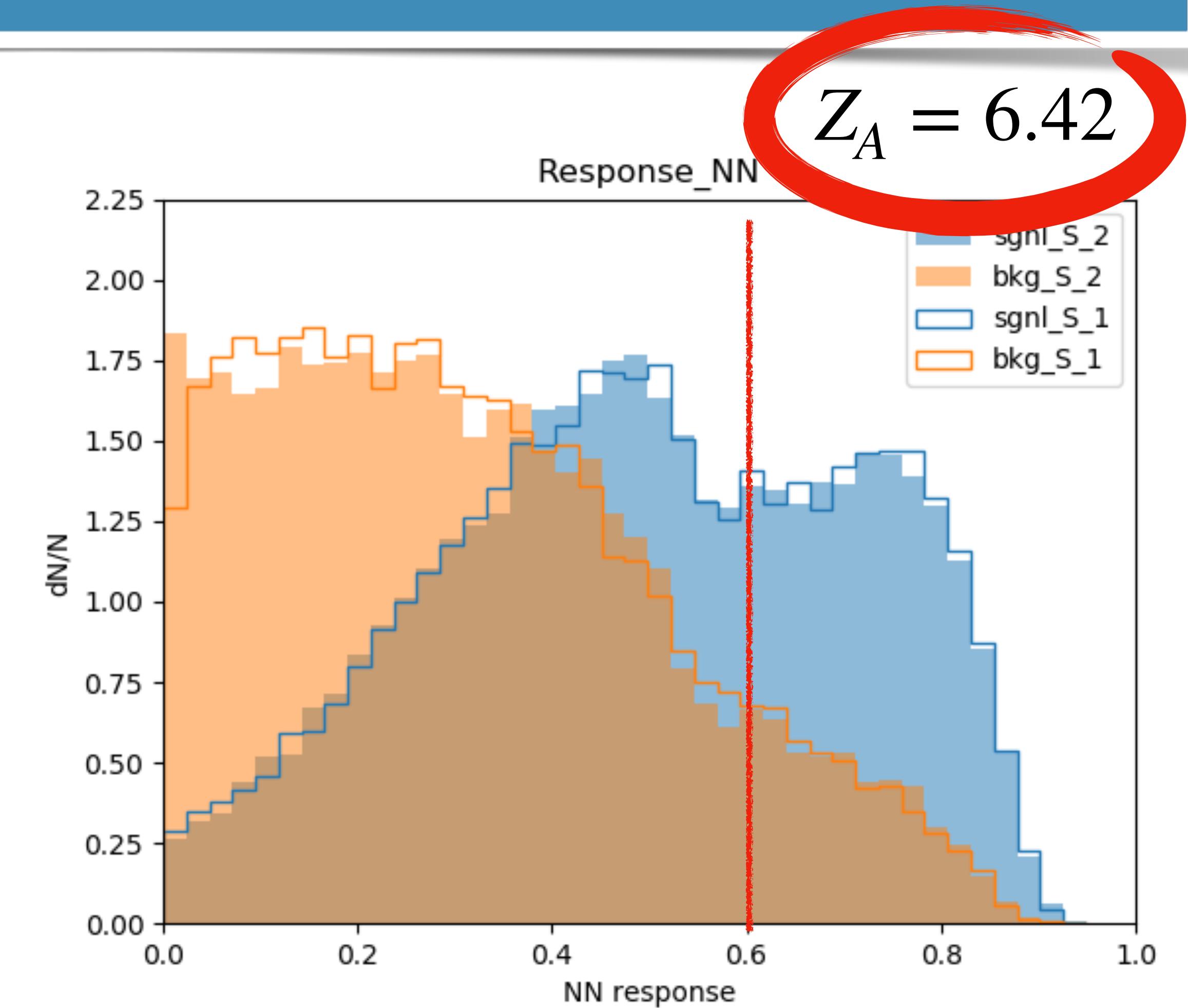
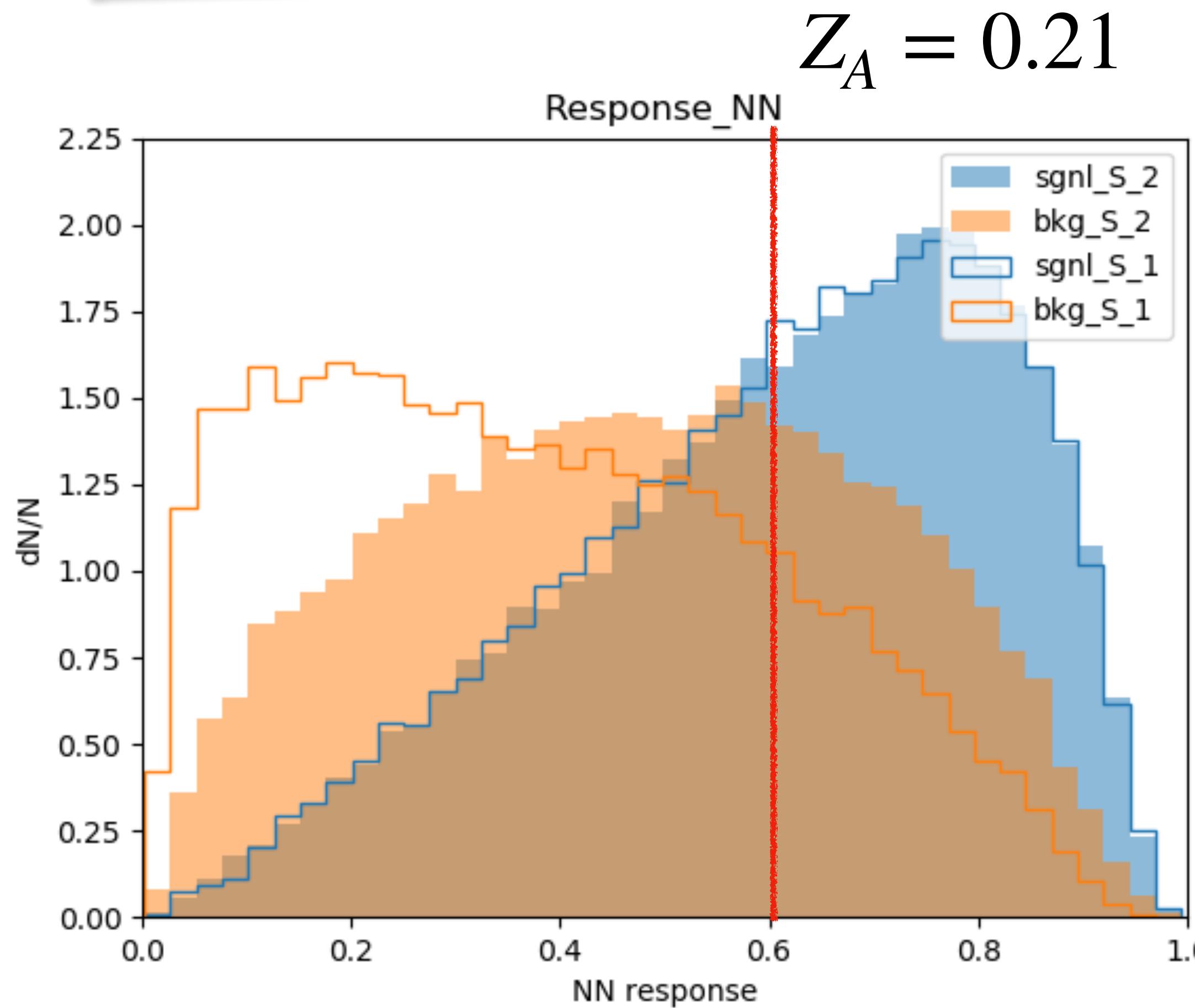


- Discriminator importance parameter $\lambda = 0$

- Discriminator importance parameter $\lambda = 1.3$

Testing on different NN settings

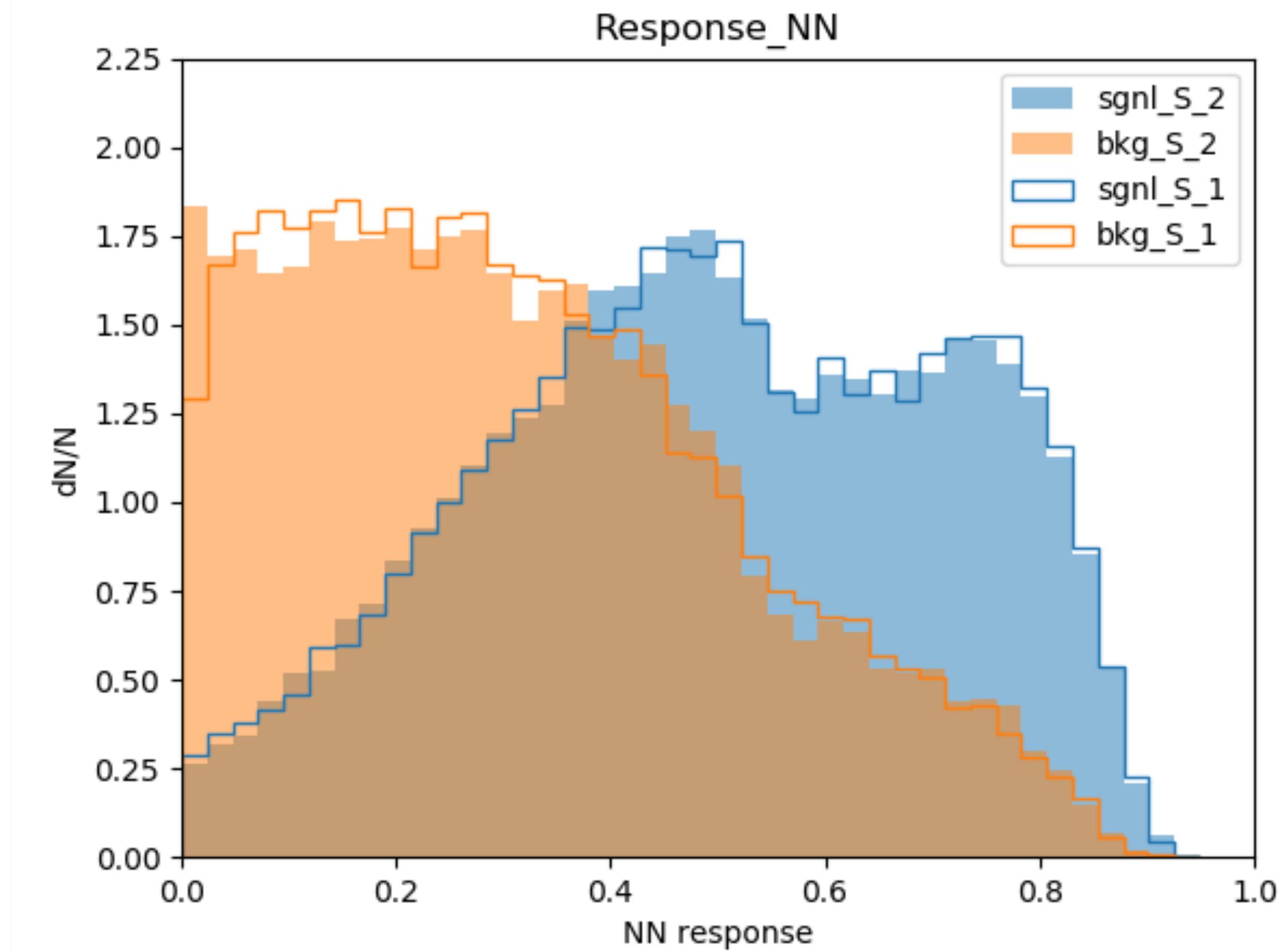
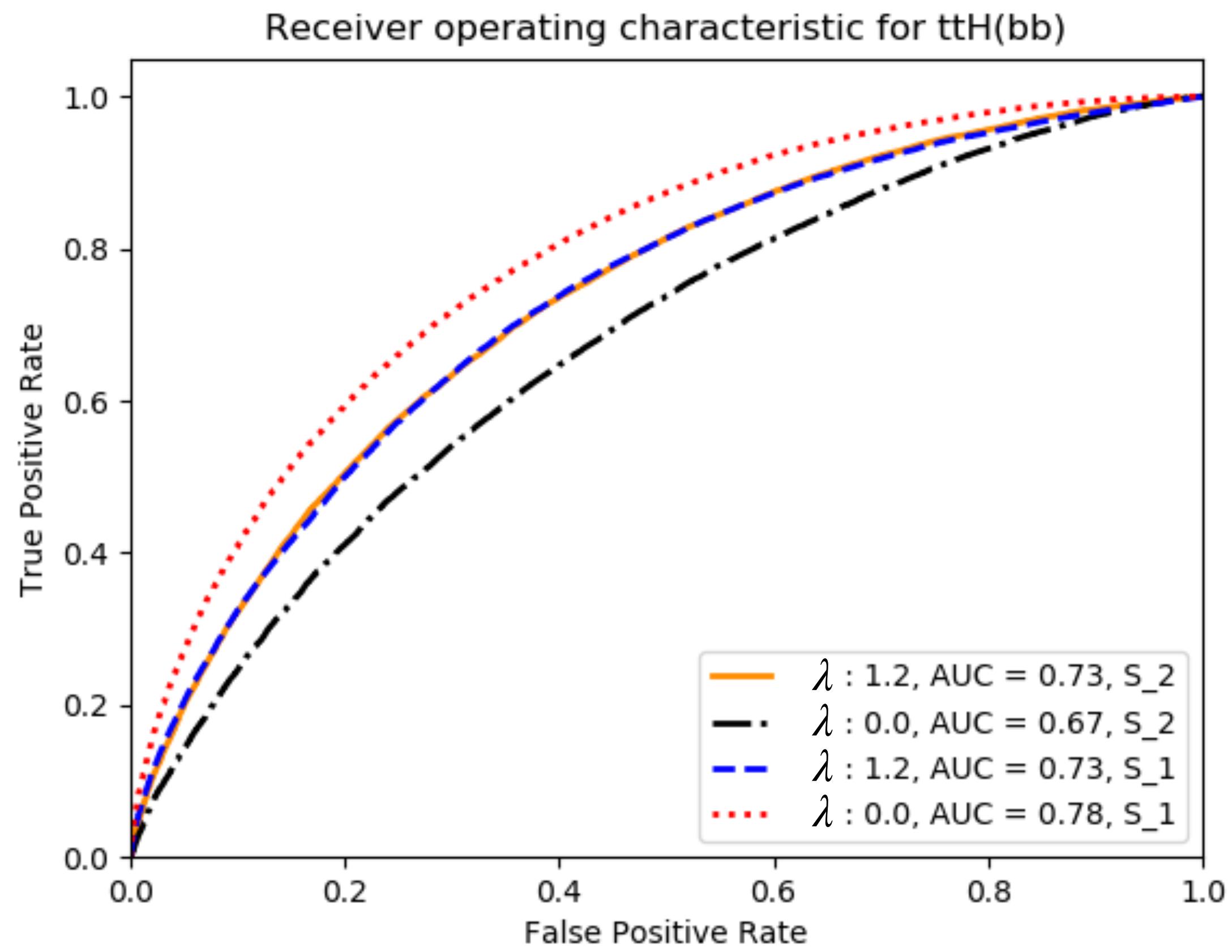
Significance and Response



- Discriminator importance parameter $\lambda = 0$

- Discriminator importance parameter $\lambda = 1.2$

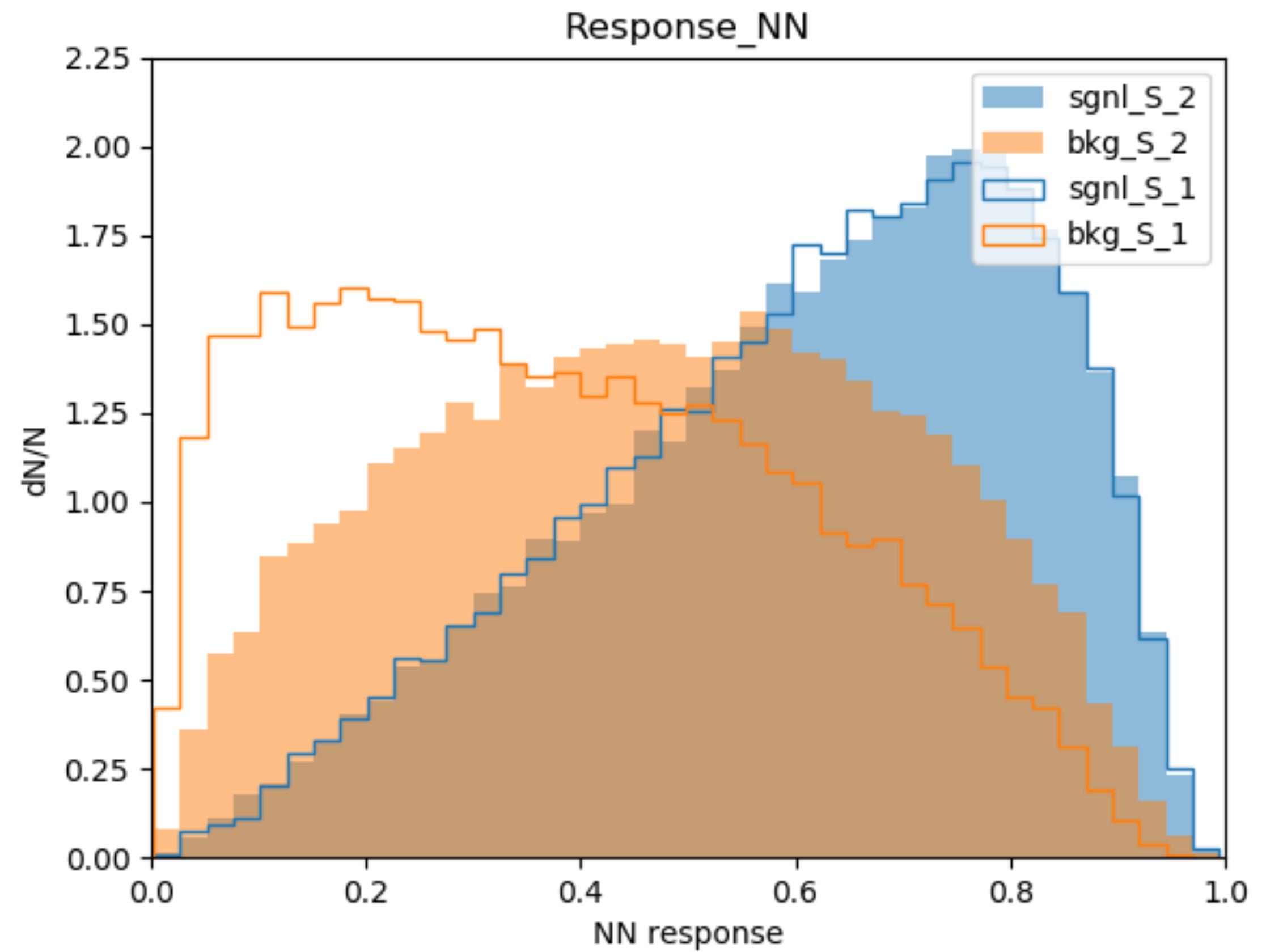
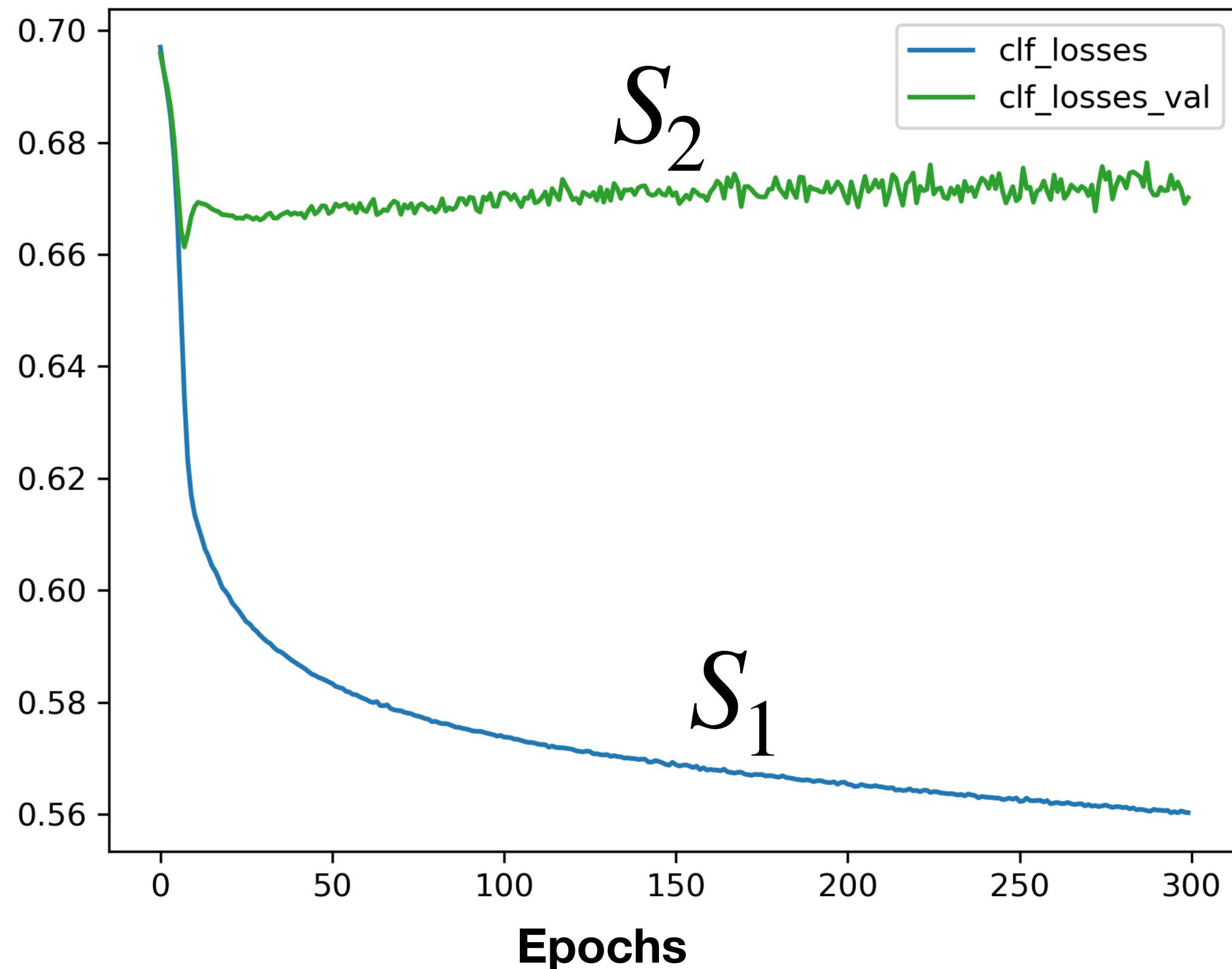
Results



Training

- $\lambda = 0$

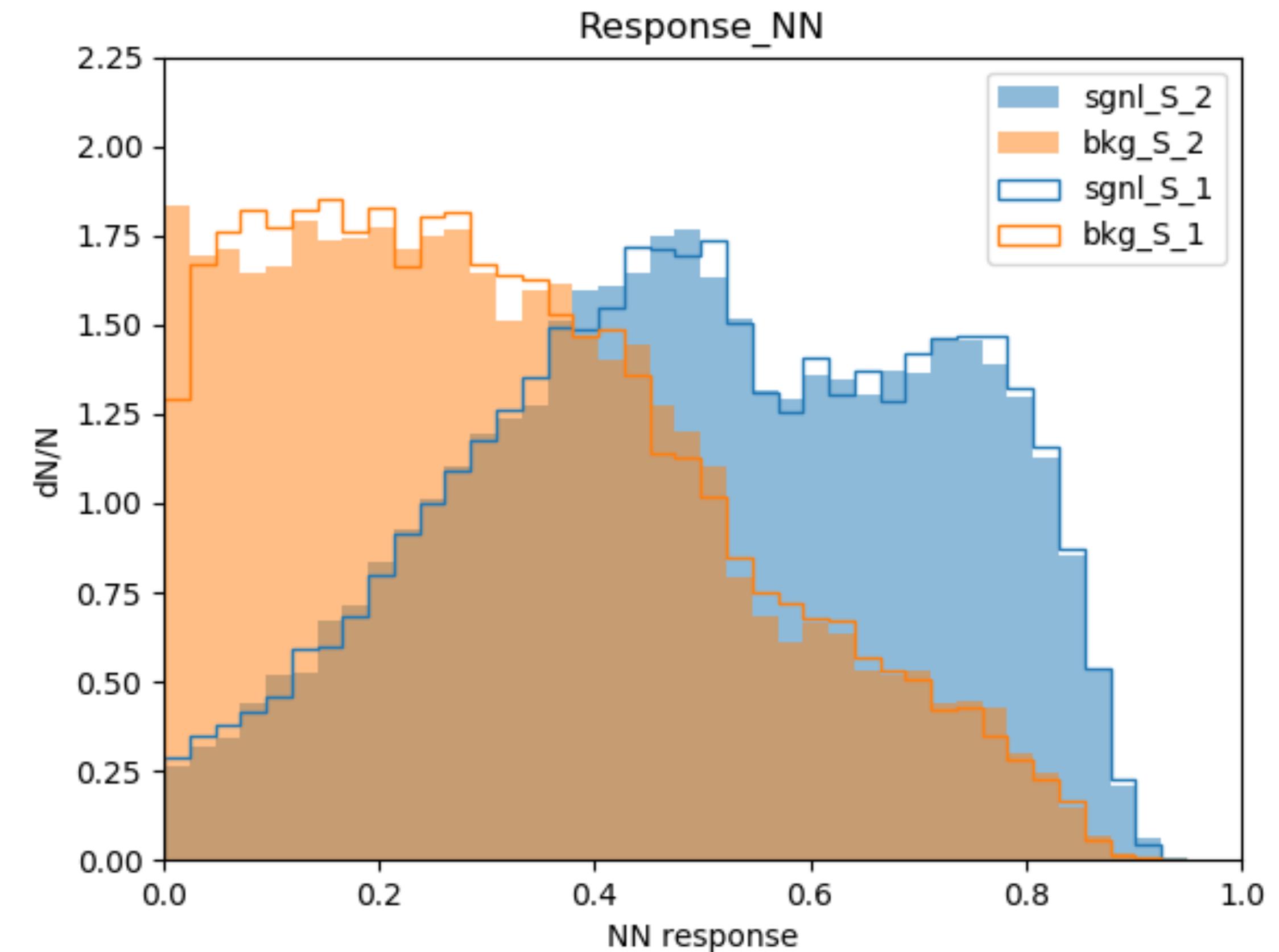
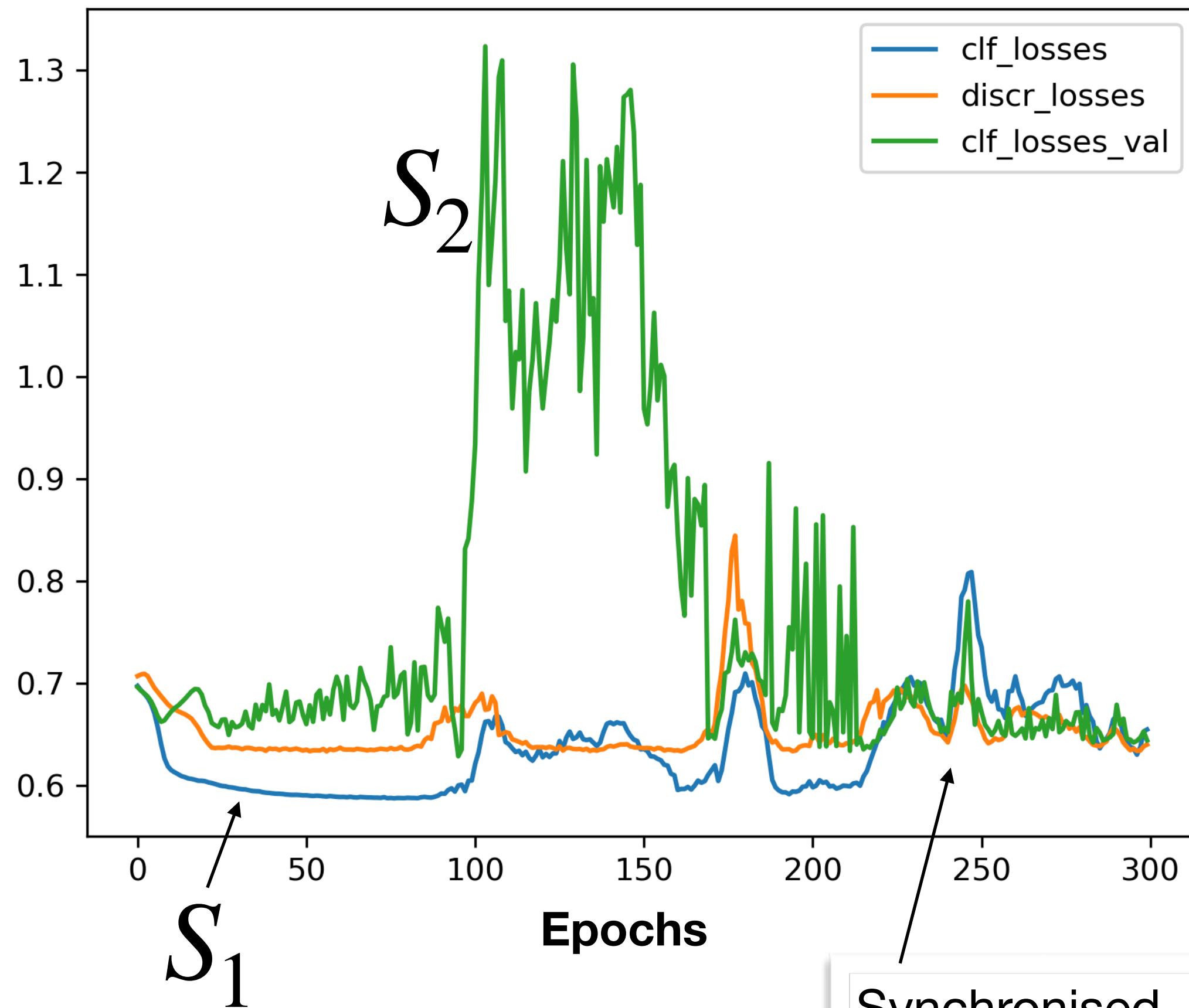
Training plot



Training

- $\lambda = 1.20$

Training plot



Results

Table 1

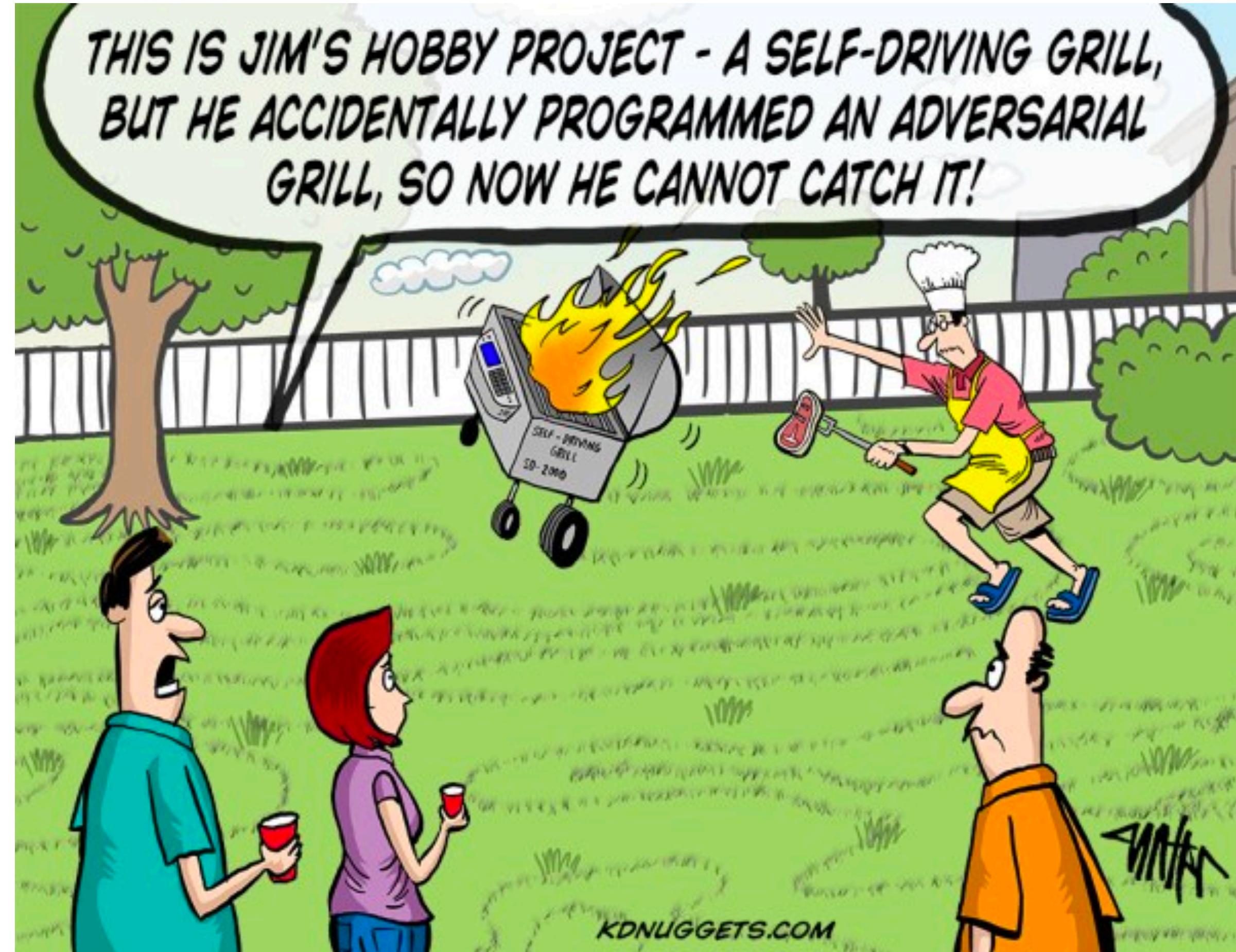
#Neurons in hidden layer	\lambda	Accuracy (test)	AUCROC (test)	Significance
[35, 30, 15]	0.000	0.621	0.668	0.212
[35, 30, 15]	1.300	0.652	0.729	<u>4.408</u>
[40, 25, 25]	0.000	0.623	0.671	0.215
[40, 25, 25]	1.245	0.642	0.729	<u>3.953</u>
[45, 30, 25]	0.000	0.622	0.671	0.214
[45, 30, 25]	1.201	0.649	0.729	<u>6.417</u>

- Adadelta optimiser with 300 epochs and batch size of 5000 events

Future work

- Develop an appropriate stopping criterion
- Understand how the network architecture and discriminator importance parameter λ result on the performance
 - Add additional layers to discriminator part
- Play with signal/bkg ratio

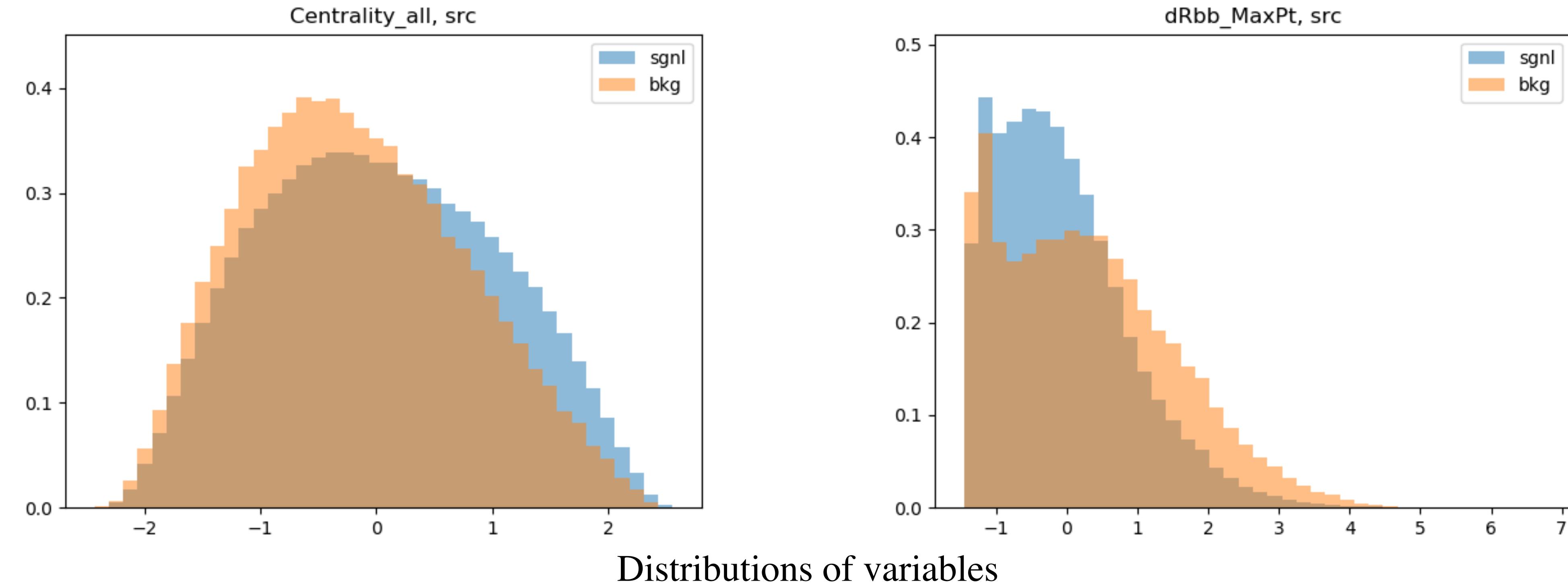
Questions?



Variables description

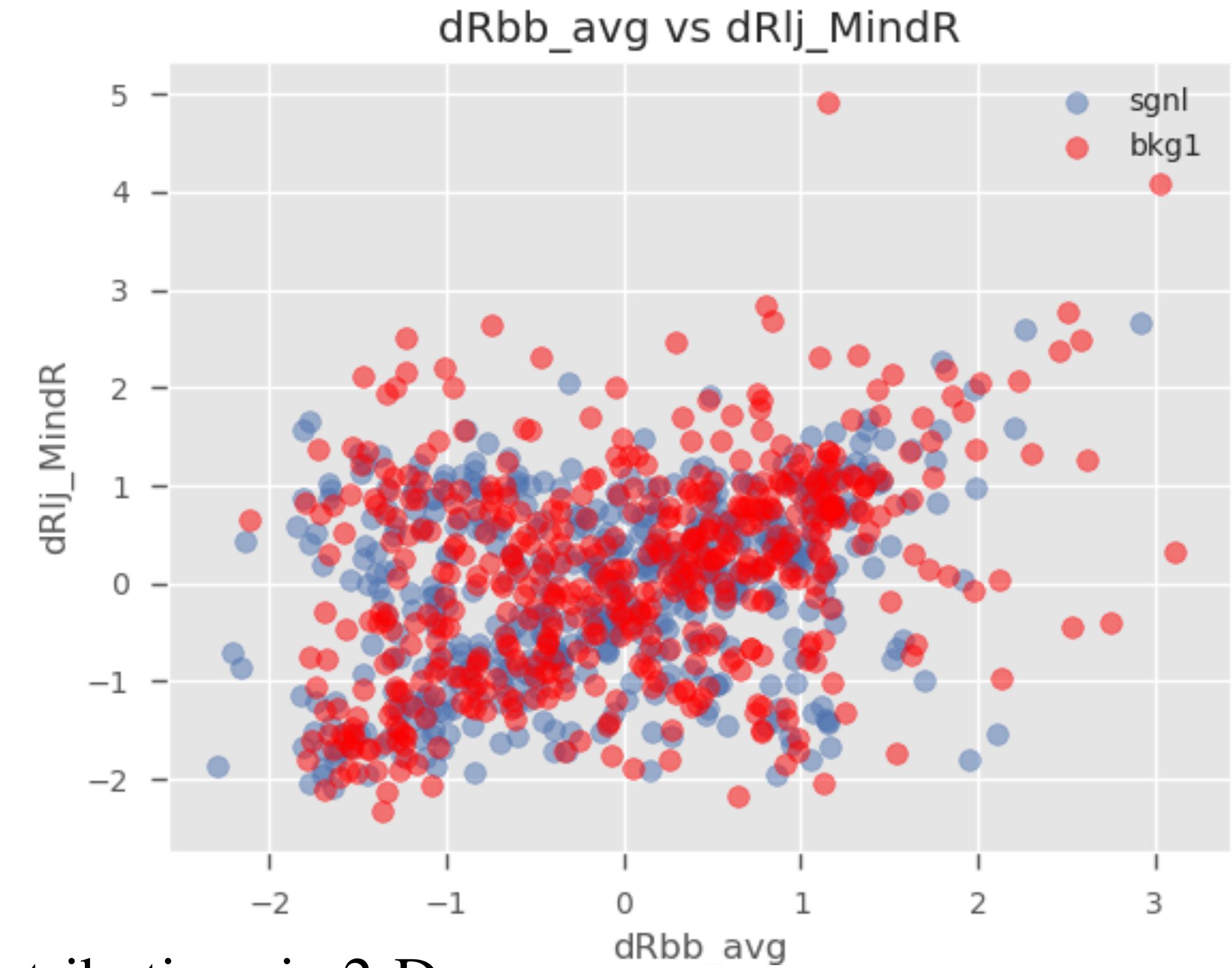
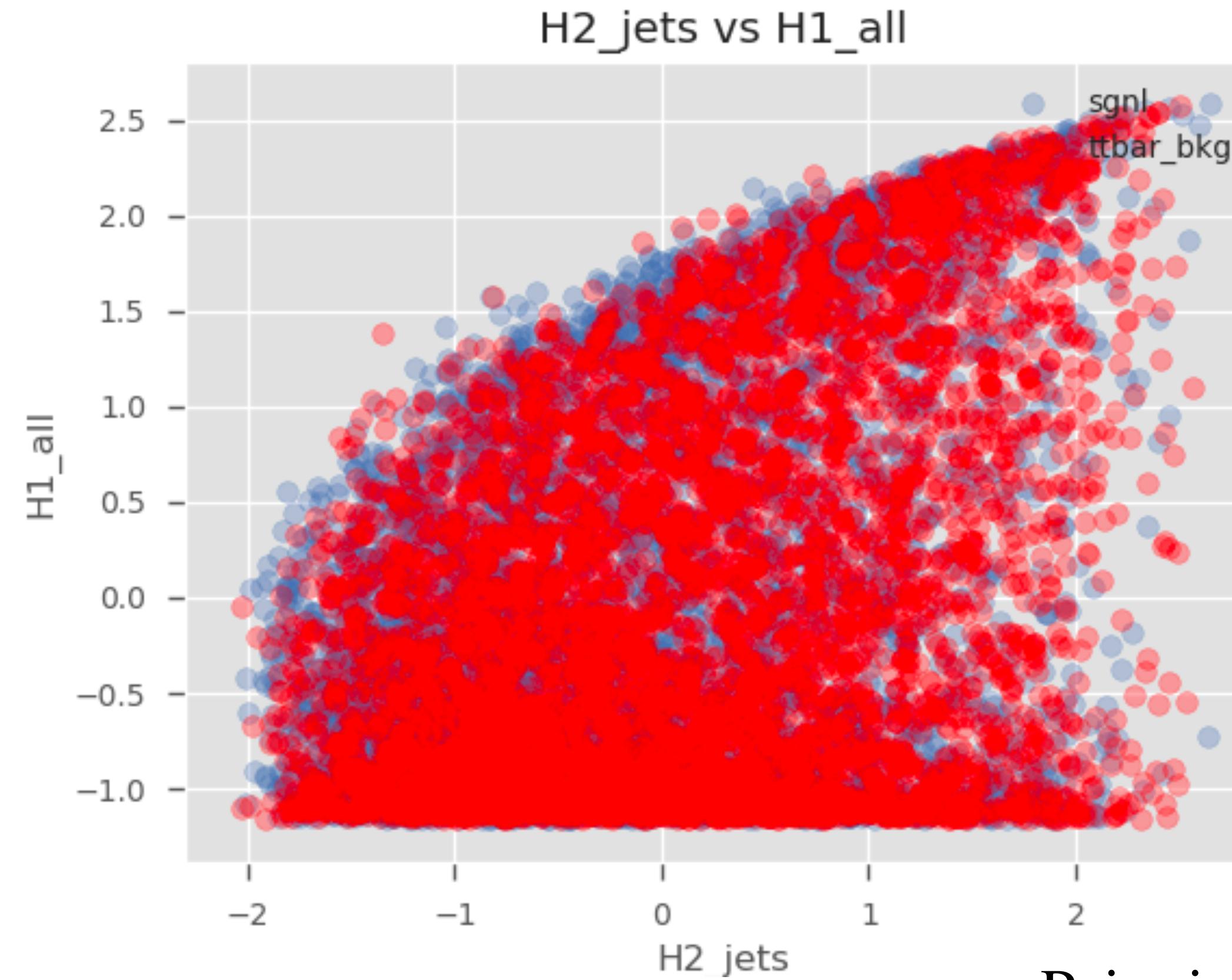
dRbb_avg	average dR of all b -jet pairs
dRbb_MaxPt	dR of the b -jet pair with the highest sum of p_T
dRbb_MaxM	dR of the b -jet pair with the highest invariant mass
dRlb1-dRlb3	dR of the charged lepton and the b -jet with the 1st-3rd largest p_T
dRlbb_MindR	dR of the charged lepton and total b -jet pair system which has the smallest dR
dRlj_MindR	minimum dR between the charged lepton and any jet
Mbb_MaxM	maximum invariant mass of any b -jet pair
Mbb_MindR	invariant mass of b -jet pair which has the smallest dR
Mbj_MaxPt	invariant mass of two jets with the largest p_T sum, where exactly one of the jets is a b -jet
Mjjj_MaxPt	invariant mass of any three jets with the largest p_T sum
pT_lep	transverse momentum of the charged lepton
HT_jets	sum of transverse momentum of all jets
HT_all	sum of transverse momentum of all jets and the charged lepton
nJets_Pt40	number of jets with $p_T \geq 40$ GeV
nbTag	number of b -jets
nHiggsbb30	number of b -jet pairs with an invariant mass within 30 GeV of the Higgs boson mass of 125 GeV
MET	missing transverse energy
dEtajj_MaxdEta	largest difference in longitudinal angle η of any two jets
Centrality_all	ratio of momentum sum over the energy sum of all objects
Hi_all, H2_jets	1st-5th Fox Wolfram transverse moment [9] of all objects

Signal vs Background classification is hard task



- The problem is hard in terms of classification
- Distributions most of the variables look similar for signal and background
- One cannot separate them neither in one nor in two dimensions

Signal vs Background classification is hard task



Pairwise distributions in 2-D

- One cannot separate signal vs background in two dimensions as well
- Principal Component Analysis (PCA) and Linear Discriminate Analysis (LDA) transforms into 2D were unsuccessful