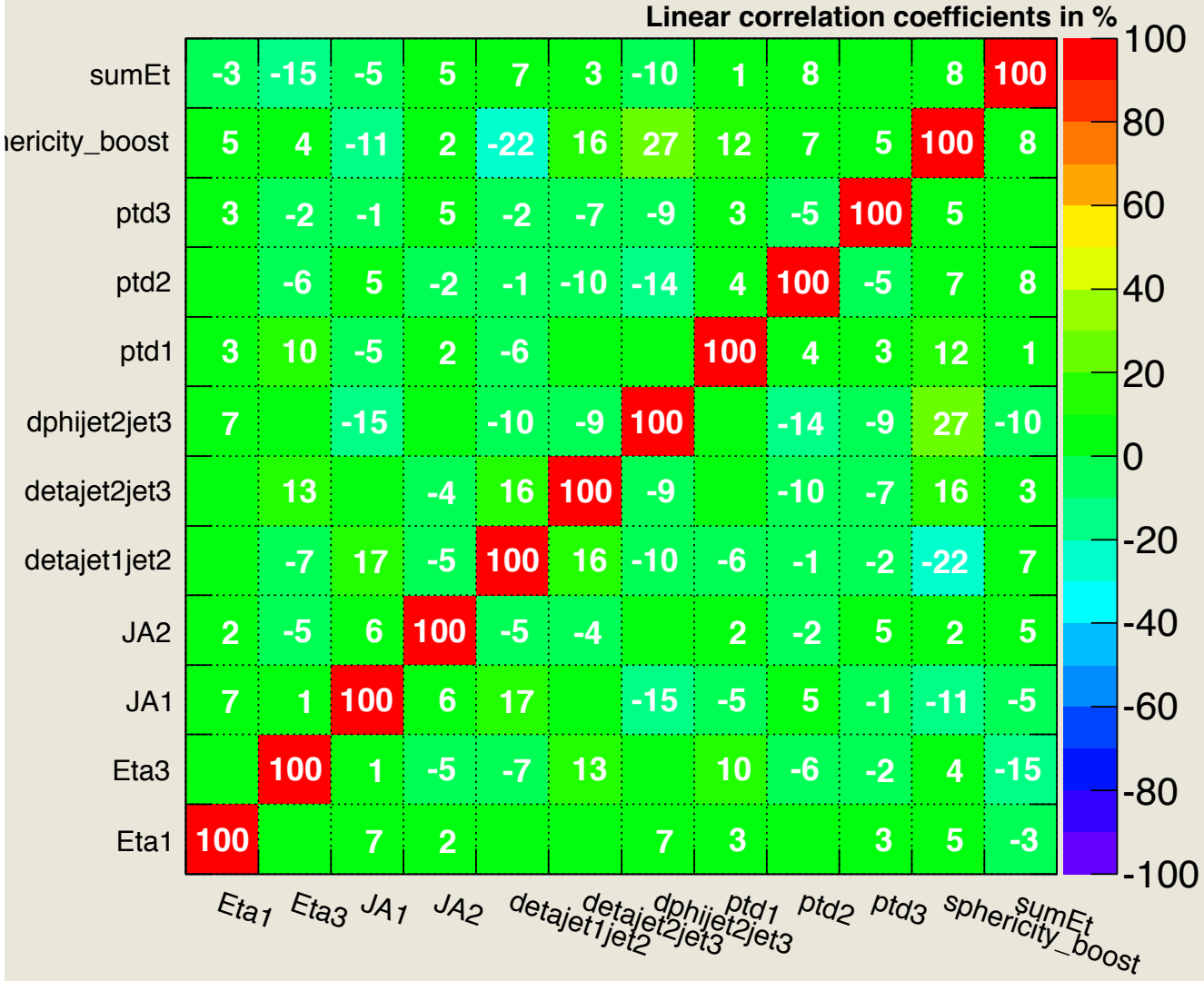
Last Week

- Completed regression exercise
- Moved on to classification in TMVA
  - S/B response to variable input
  - Working with current selected variables
- My task:
  - Manually adjust BDT settings to verify automated results
    - Number of trees used
    - Training sample size

Variable Selection

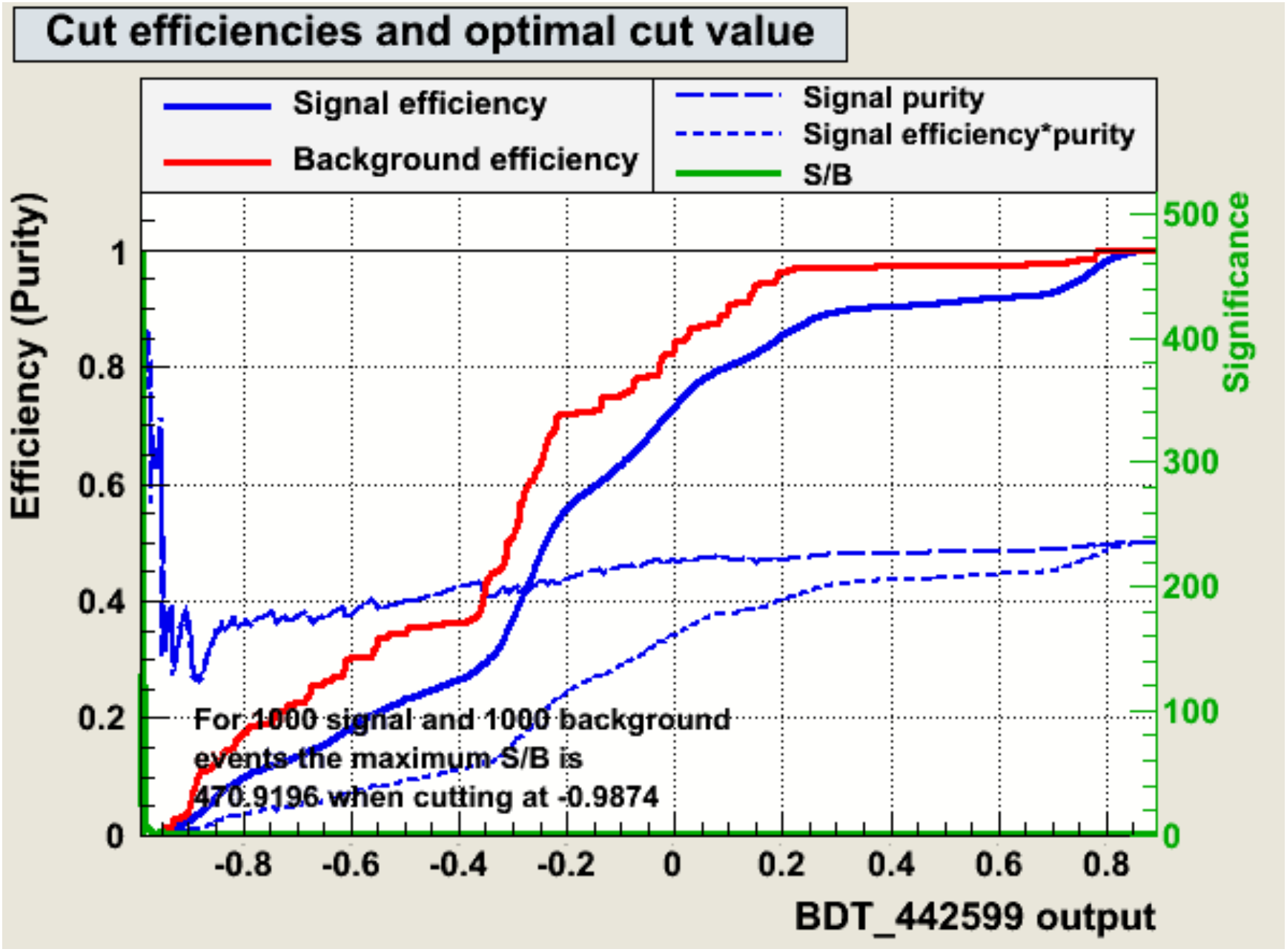
- Variables for cuts have been determined via an optimization script in TMVA.
- Works for Montecarlo, somewhat less effective for real data

Correlation Matrix (background)



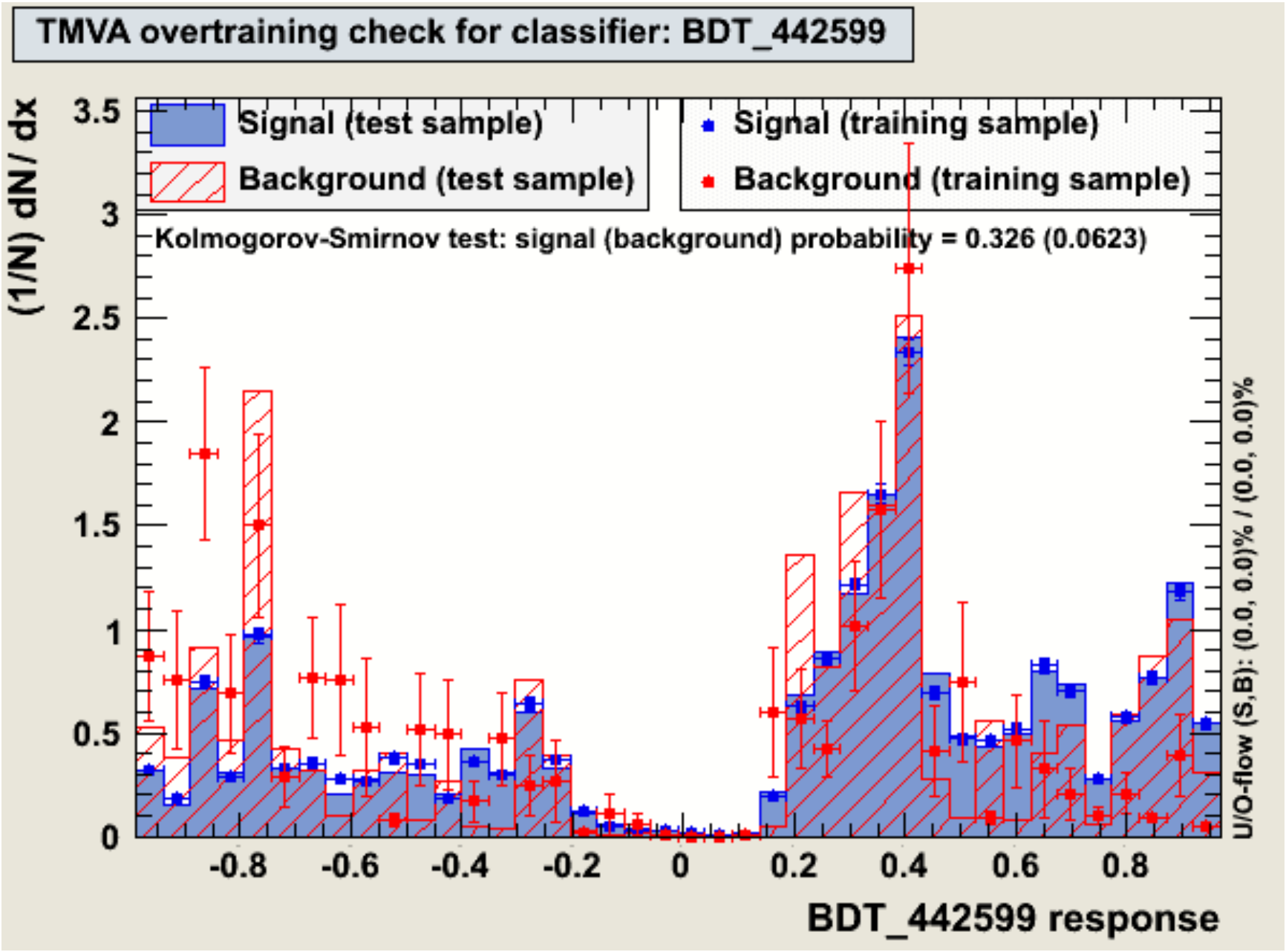
Control Searches

- Control search for Higgs Signal
  - Reference point for areas where we are looking for Higgs Signal
- Must determine optimal cuts for
  - Low Higgs signal
  - Clean signal to background ratio


Cut efficiencies and optimal cut value
Signal efficiency
Background efficiency
Signal purity
Signal efficiency*purity
S/B
Efficiency (Purity)
Significance
BDT_442599 output
For 1000 signal and 1000 background
events the maximum S/B is
470.9196 when cutting at -0.9874

Signal Separation

- Training Steps
  - Signal/Background ratio response to input variables
  - Seeking optimal Signal to Background ratios


TMVA overtraining check for classifier: BDT_442599
Signal (test sample)
Background (test sample)
Signal (training sample)
Background (training sample)
Kolmogorov-Smirnov test: signal (background) probability = 0.326 (0.0623)
(1/N) dN / dx
3.5
3
2.5
2
1.5
1
0.5
0
-0.8
-0.6
-0.4
-0.2
0
0.2
0.4
0.6
0.8
BDT_442599 response
U/O-flow (S,B): (0.0, 0.0)% / (0.0, 0.0)%

This week

- Determine discrepancies between outputs
- Attempt to achieve optimal signal separation for Monte Carlo
- Move on to data analysis