

MVA for $L_c \rightarrow p h h$ offline selection
update
(Will do this in LaTeX next time, just
very tired tonight)

25.07.2012

Recap

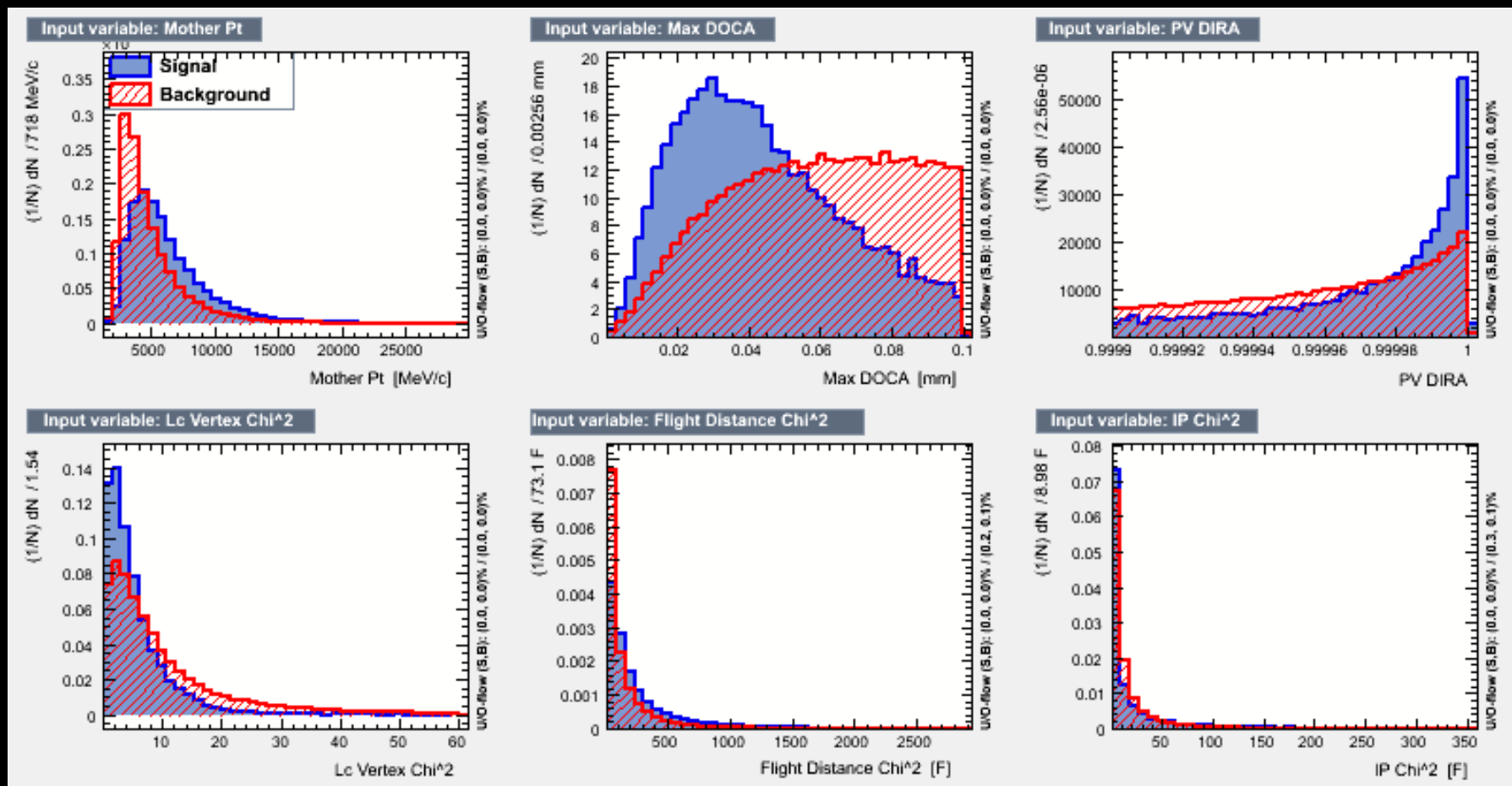
- We have previously examined the use of MVA with MC, and have found that BDTs and MLPs display optimum discrimination between signal and background for the CF Lc mode.
- Now have moved onto looking at MVA responses to sWeighted CF data.

MVA Setup

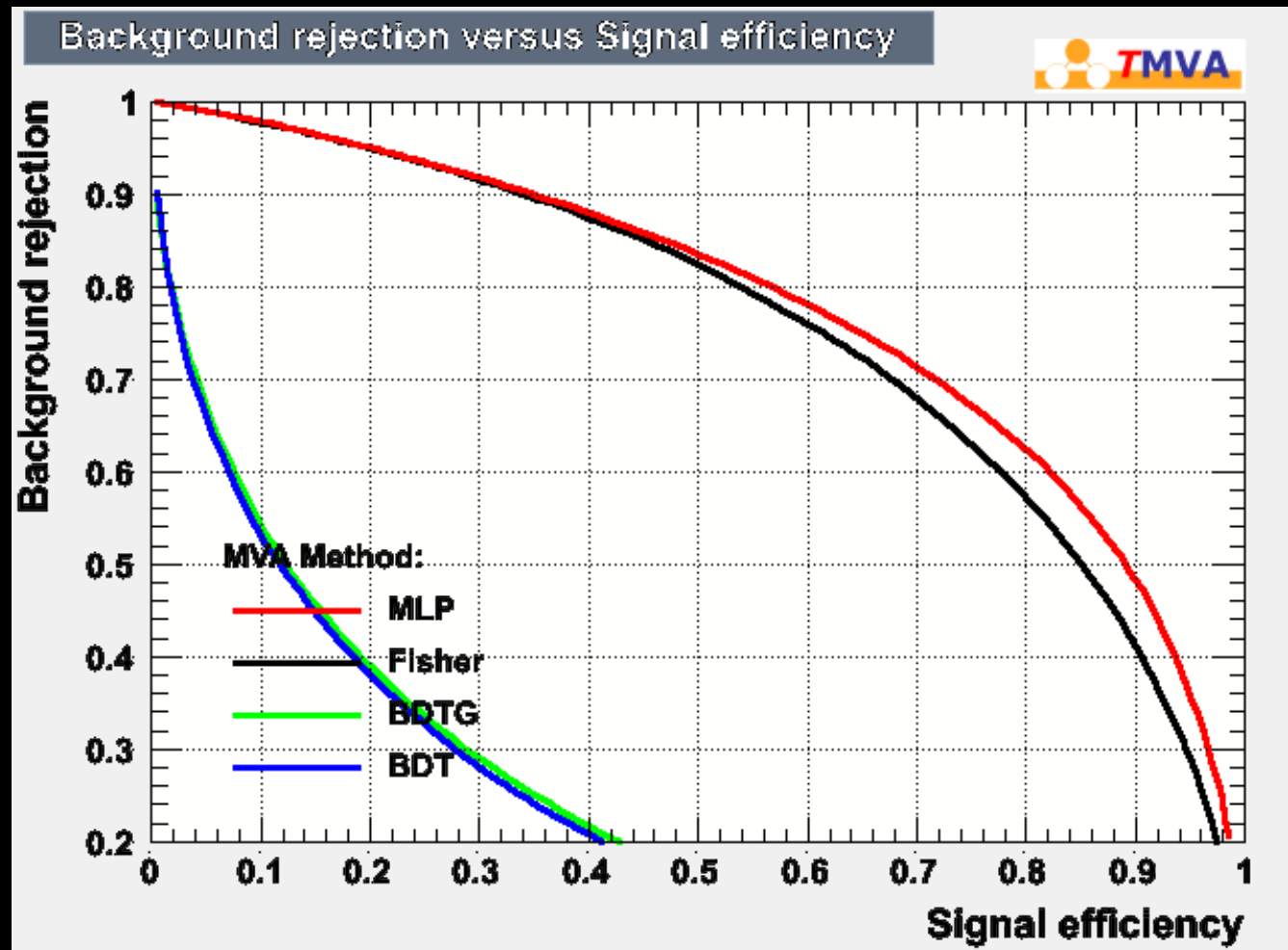
- Keep MVA agnostic to daughter properties for use in all phh modes.
- Use in MVA training:
 - Lc Pt
 - Lc daughter DOCA
 - Lc Vertex χ^2
 - Lc FD χ^2
 - Lc IP χ^2
 - Lc DIRA wrt PV
- Use the following PID cuts globally:
 - $\text{Proton_PIDp} > 5 \ \&\& \ \text{Kaon_PIDK} > 5 \ \&\& \ \text{Pion_PIDK} < 10$
 $\&\& \ (\text{Proton_PIDp} - \text{Proton_PIDK}) > 0$

TIS Input variable Distributions

All info in these slides concerns the TIS sample.

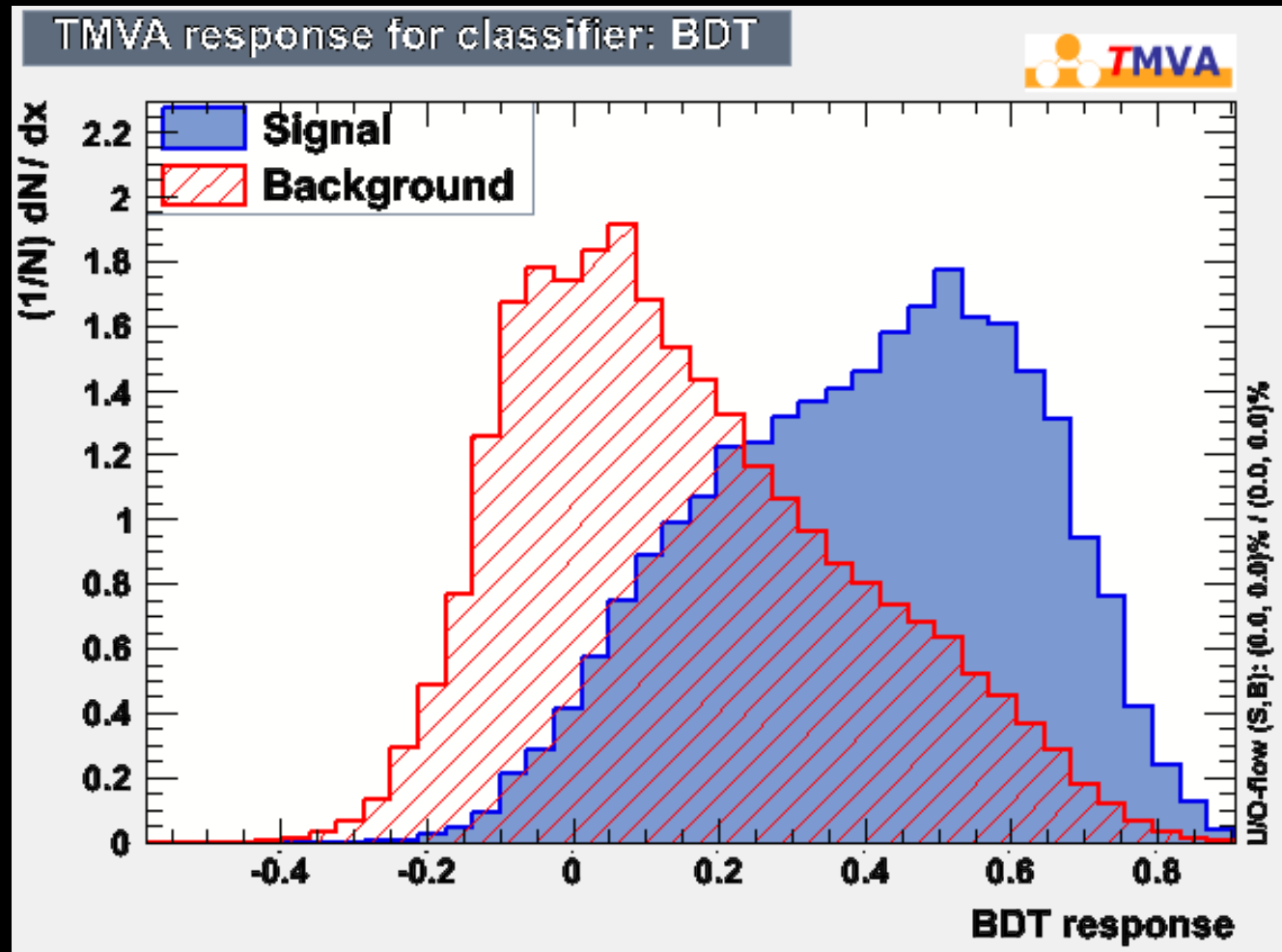


ROC curve



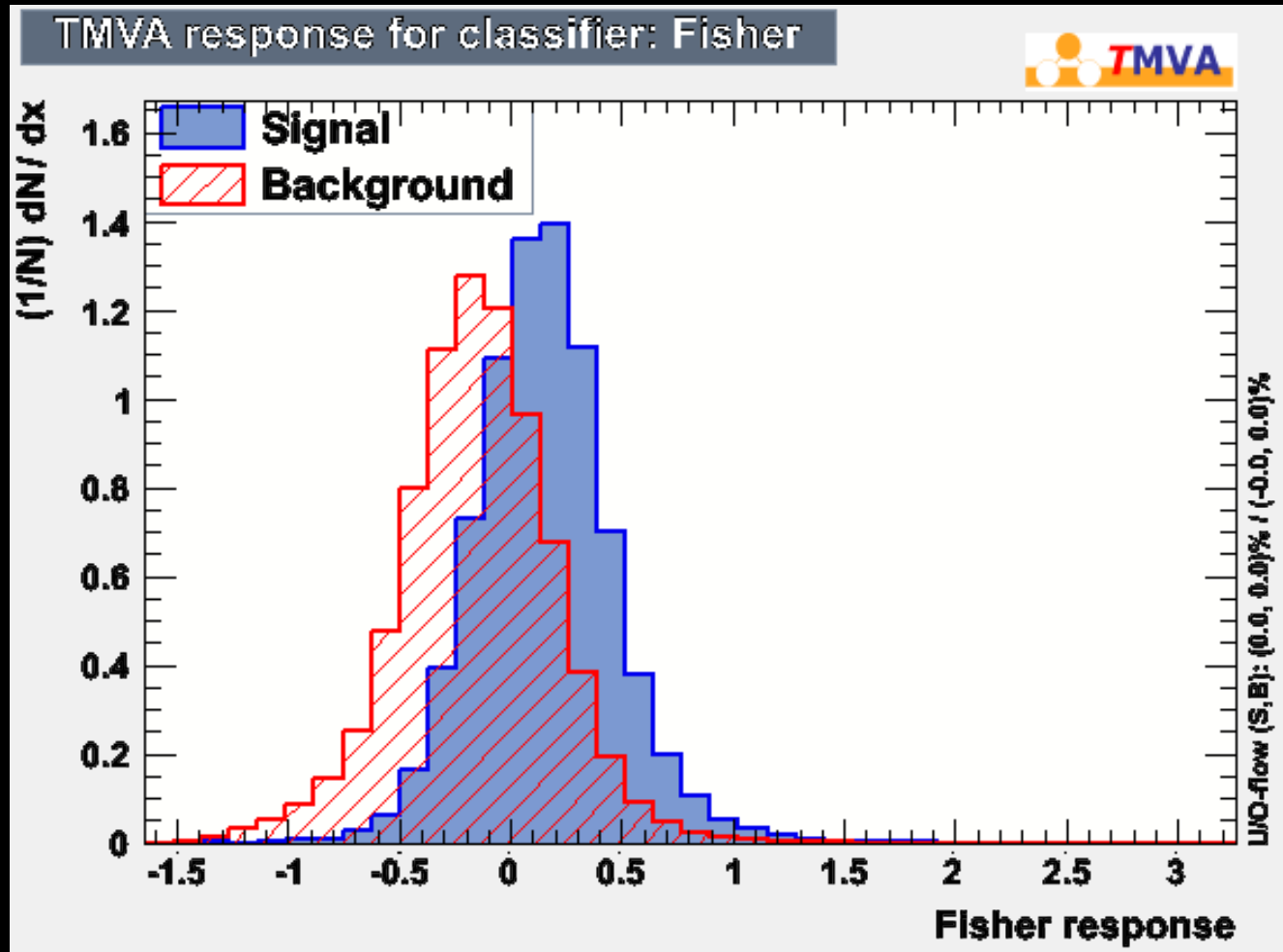
ROC curve indicates BDT is worse than random guessing, but...

BDT response



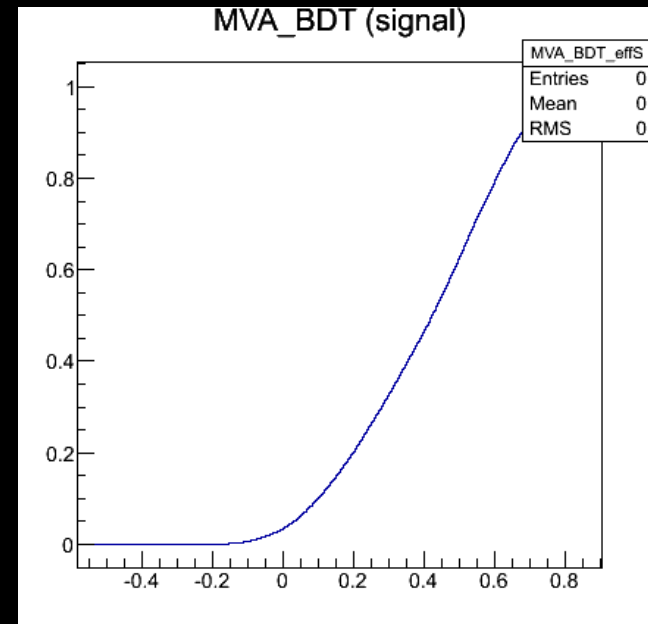
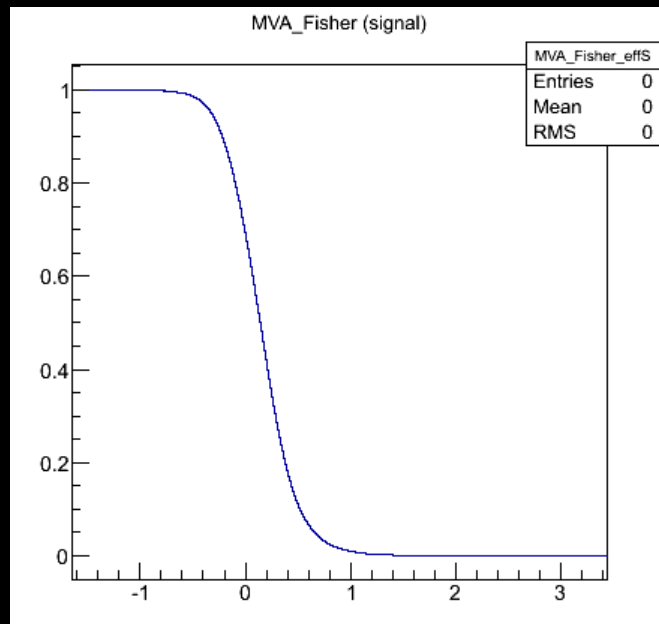
... BDT response displays obvious discriminatory power.

Fisher Response



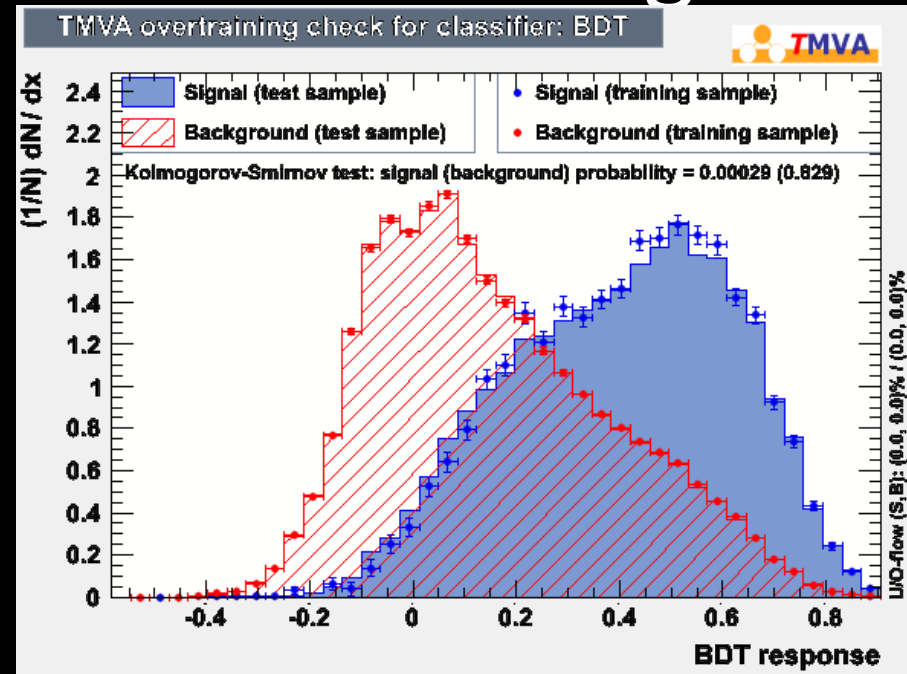
Compared to Fisher, BDT should display better discrimination.

Answer lies in how signal efficiency is handled by the histogram booking in TMVA



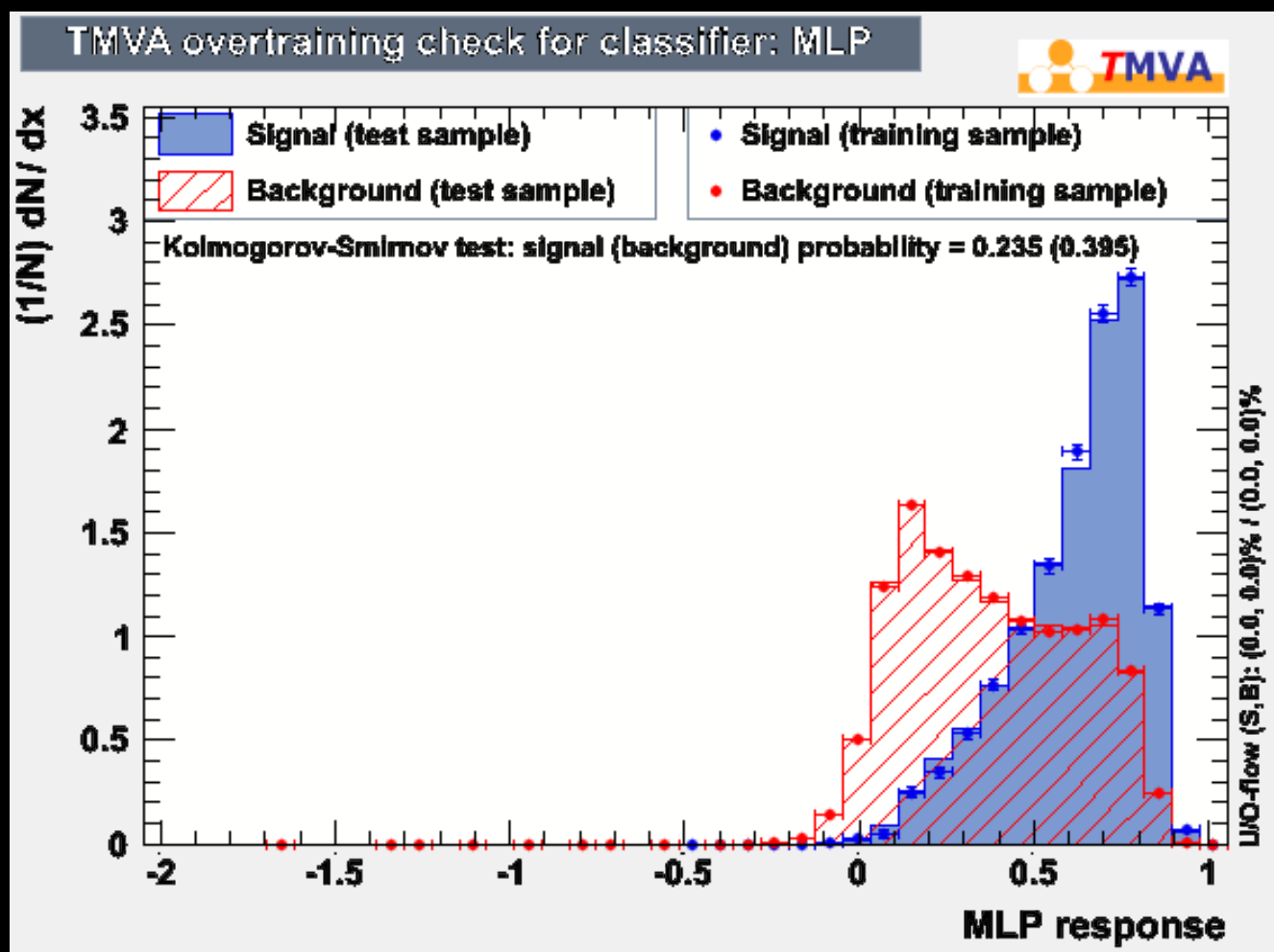
- Shown are signal efficiency histos, used to plot the GUI ROC curves and classifier efficiencies. For all other MVA methods this number means “fraction of signal with MVA value above this value”
- For BDT (for whatever reason) the histo is booked such that it means “fraction of signal with MVA value below this value”
- This leads to the errors in the GUI plots we see.

BDT Overtraining Checks



- So the BDT does display sensitivity. How do the testing and training phases match up?
- Answer: not very well. Strong agreement between bkg training and testing but not for signal.
- Indicates we are overtraining the BDT. I'm currently looking at tree/forest construction parameters to see if we can get improvement.

Meanwhile with MLPs...



Very good classifier efficiencies, but also good agreement in the training and testing phases. Requires very few cross checks of MLP structure to ratify as a valid selection... A tempting option.

What now...

- Will mail TMVA list to see what can be done about BDT bug (if it is a bug and not an obscure training option flagged).
- If not easy to fix will manually generate ROC curves.
- Further alter parameters for BDT construction to try for better signal training and testing agreement.
- While this is going on, can do quick cross checks with MLP, if it looks solid I'll begin a S/route(S+B) optimisation for the CF and DCS modes.
- Can also check what it does to the Dalitz space, which may be a factor in choosing between BDT and MLP (assuming I can get a solid BDT put together)