Electron-proton separation

An update

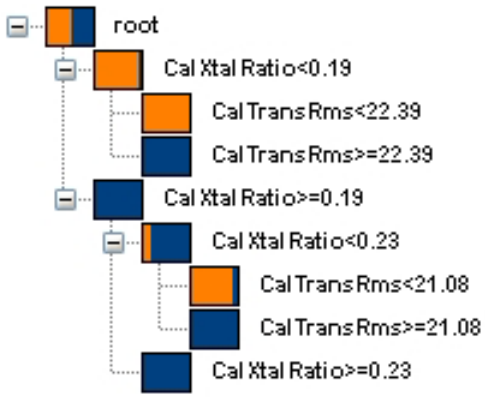
(see Dec 18 for the first results)

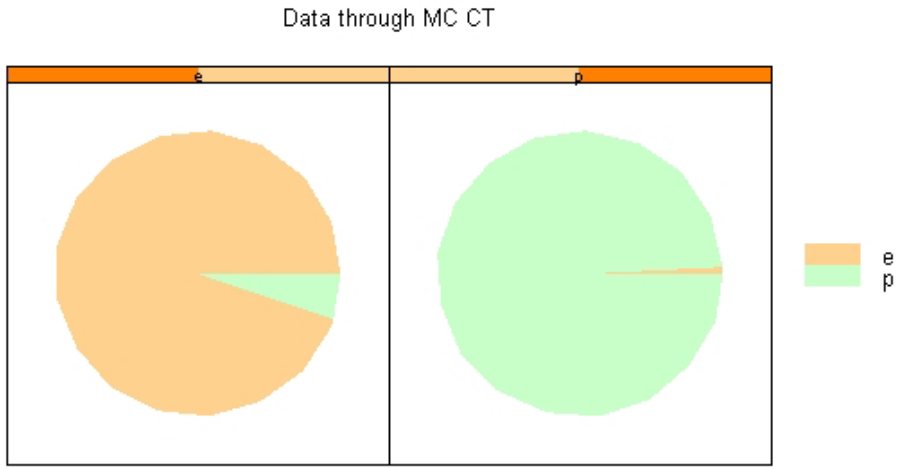
Goals and data sets

- Plan
  - Oversimplify data set using e and p runs at 0 degrees
    - Possibly a very simple dataset to test sensitivity of CT to data/MC discrepancies
  - Scale N-tuple variables used for CT growth according to data/MC discrepancies
  - Re-grow CT using scaled MC sample and reclassify data
- Data set
  - e runs
    - 1922 (280GeV), 1981 (99GeV), 2039 (50GeV)
  - p runs
    - 2363 (100GeV), 2237 (20GeV), 1755 (150GeV)

Selection using state of the art MC

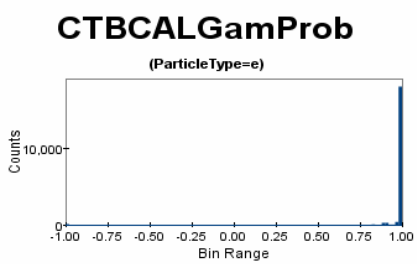
- Used latest BTRelease to generate MC sample
  - With pass5 variables
- Caveat – CT is not optimized
  - Grew CT using a well defined set of variables (Alex prefilter cuts to select high energy electrons + CTBCalGamProb)
- CT is very simple
- 5% of electrons events are misclassified
  - Why?
  - How does this change after scaling variables?

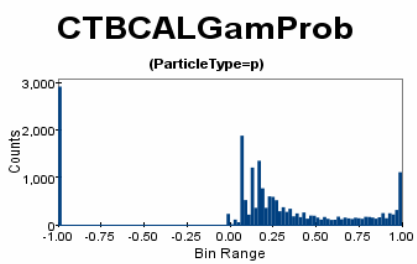


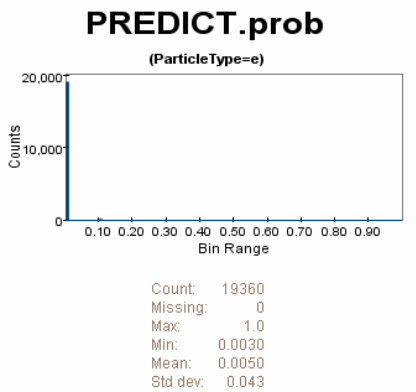


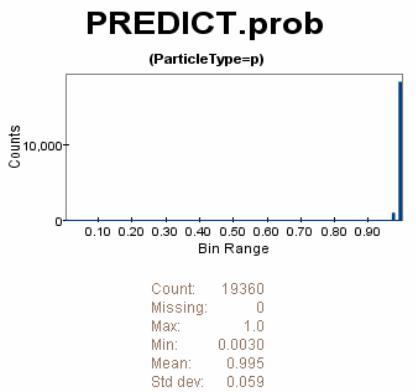
Prediction power

- CTBCalGamProb works very well for electrons
  - But you cannot expect to use just this to classify this specific data set
- Many p events are rejected by Pass5 prefilter



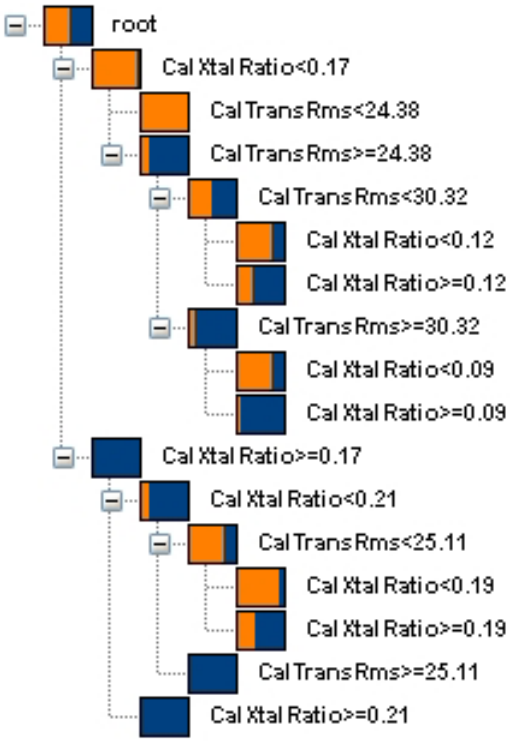


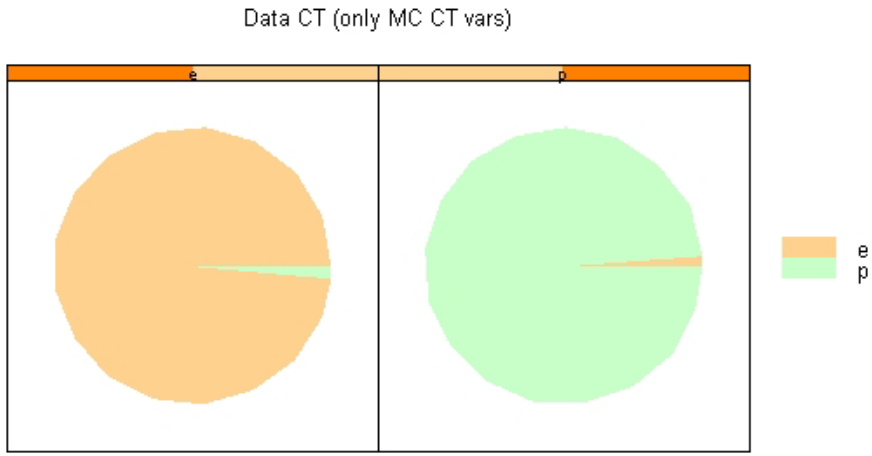




Select using Data CT

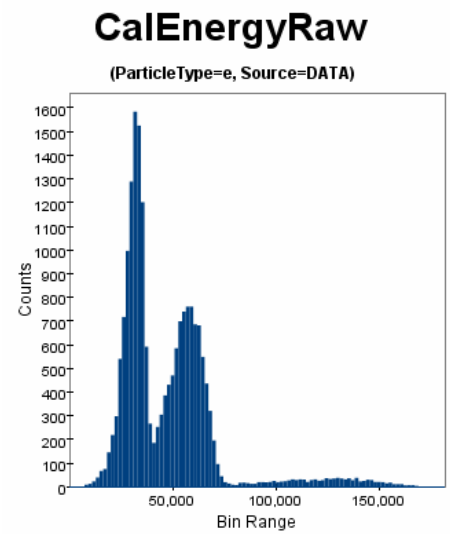
- Marginal misclassification (0.5%) can be achieved by training on data using only the variables selected by MC CT
  - This is always possible, basically defined by CT properties (number of events in basket, entropy threshold)
  - Number of branches increases

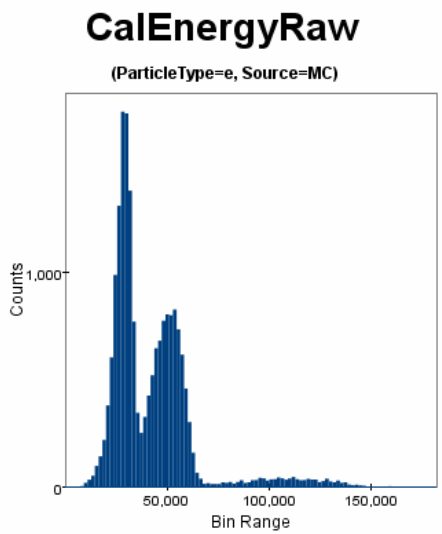


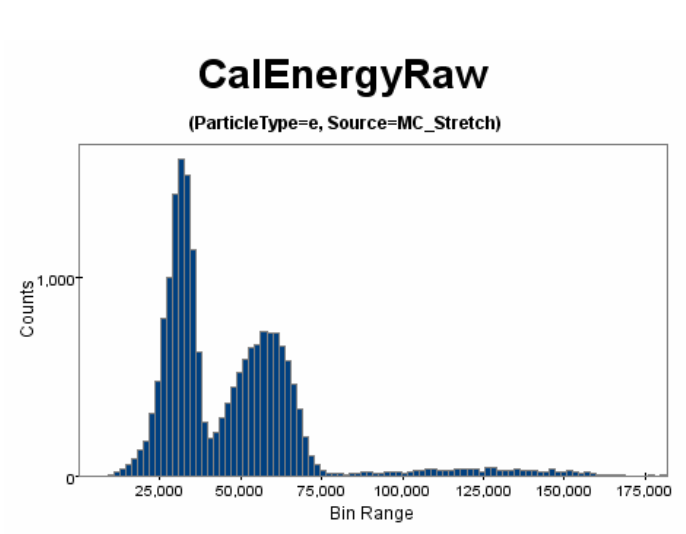


Scaling variables

- Scaled each variable in the CT by ratio <Data>/<MC> over the whole data set
- Some values of CalXtalRatio exceeded 1 and were put at 1 after scaling

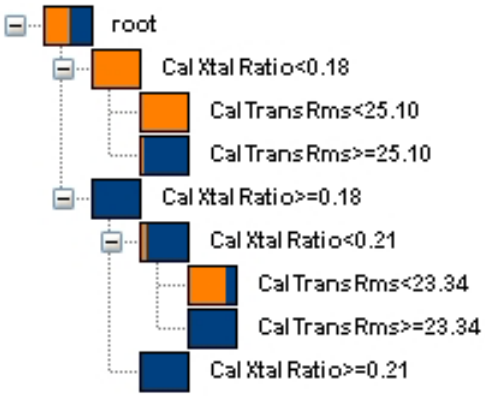


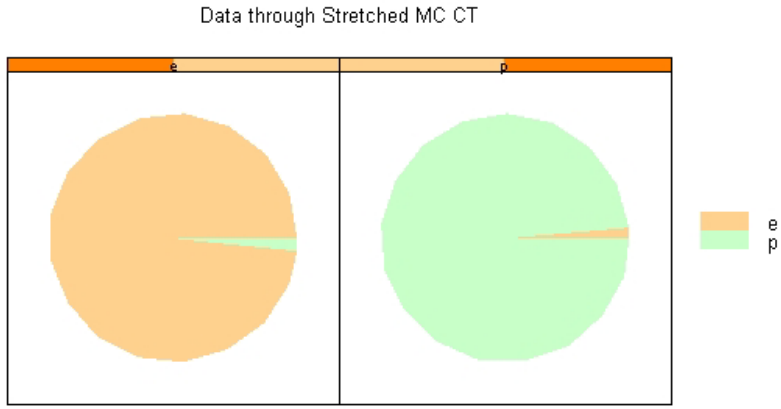




Selection using scaled MC

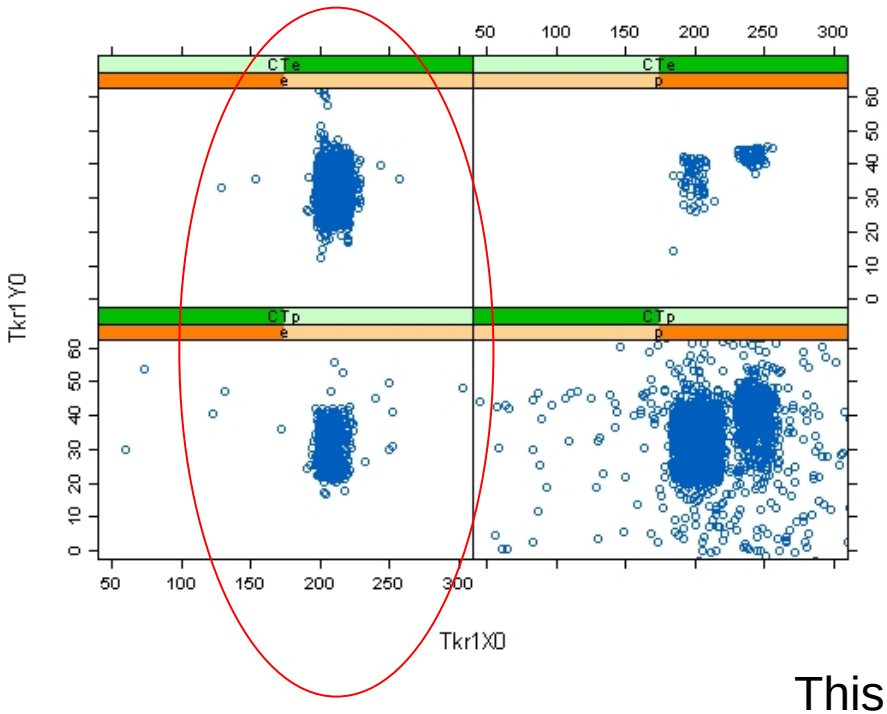
- 5.1% misclassification in e sample reduced to 1.7%
- MCstretch CT is still simpler than data and much alike MC CT with just stretched thresholds in the cuts

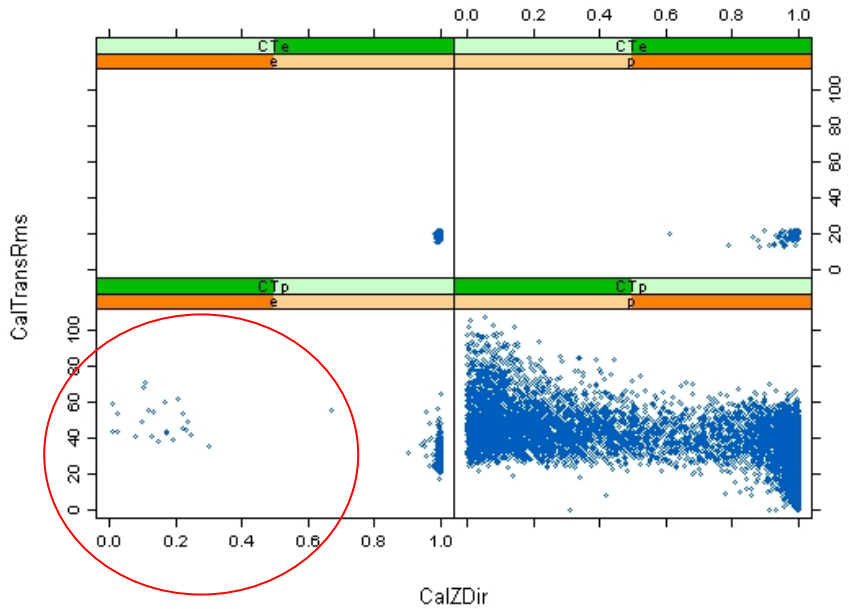




A closer look at misclassified electrons

- What do they have special and different from the others?
- Looked at variables not used for classification, in particular impact point, extrapolated impact point on cal, CalZDir and found no difference with other electrons

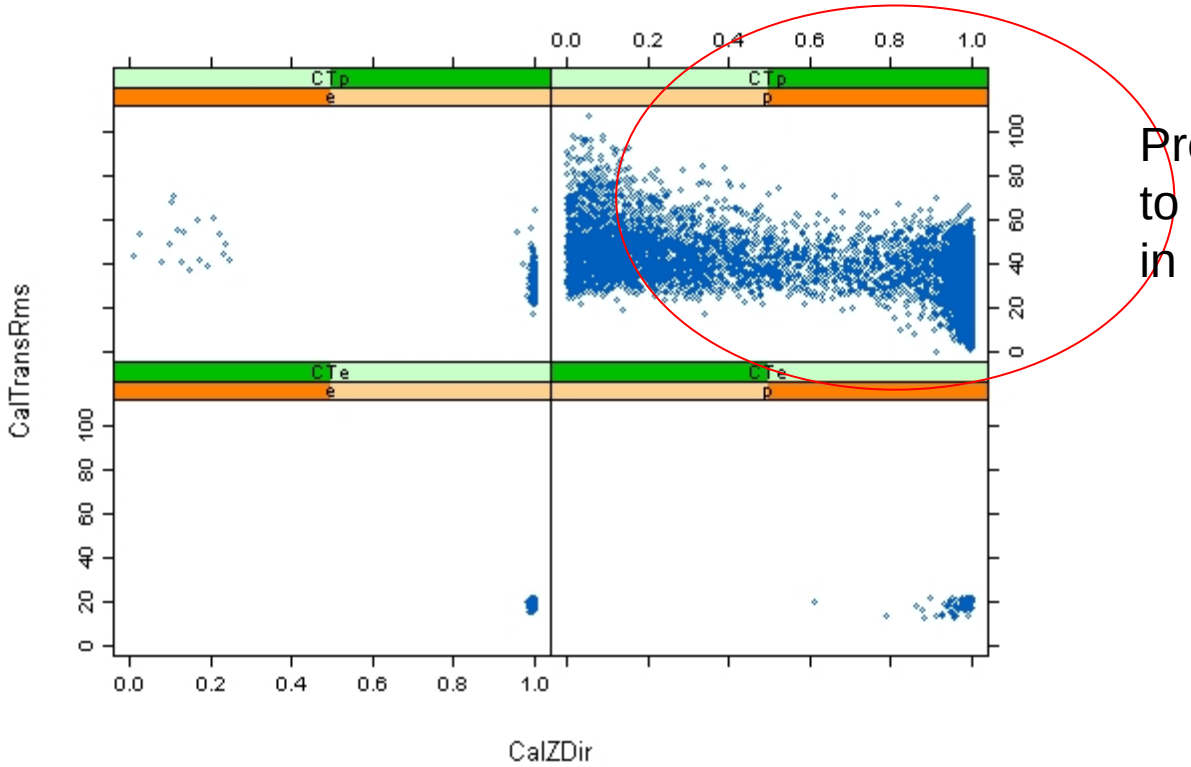




This is different but a small fraction of the events

Adding fiducial cuts

- **Borrowed Philippe fiducial cuts to reject crystals border and plotted the same but found again no difference**



Protons direction is known to be poorly reconstructed in CAL

Conclusions

- Still looking at root cause for misclassifications
- Now able to scale variables and regrow trees
- Will test what happens with scaling on the Pass5 worksheets that Bill provided

Appendix – Prefilter cuts

- Details of cuts applied to data set
- BaseCut (all runs)
  - CalTransRms>0 && CalTransRms>0 && Tkr1ToTTrAve>0 && CalCfpEnergy<500000
- Data
  - BaseCut && (GemDeltaEventTime*50/1000000)>1. && EventGtccFifo==0
- electron runs (basically raw energy < beam energy)
  - 2039
    - CalEnergyRaw> 2000 && CalCfpEnergy> 20000
  - 1981
    - CalEnergyRaw> 5000 && CalCfpEnergy> 5000
  - 1922
    - CalEnergyRaw> 10000 && CalCfpEnergy> 50000