The 3rd Pass Back Rejection Analysis using V7R3P4 (repo)

1) Training sample: Bkg: Runs 1-10000 (SLAC) v7r3p4 (redo)
(7518 Sec.)
AG: 50k from final 103k/2M run v7r3p4(redo) first 2M

2) All "CTcut": Excludes the following

ACD Veto: (AcdActiveDist3D > 0 || AcdRibbonActDist > 0) && Tkr1SSDVeto < 2

Corners: AcdCornerDoca > -5 && AcdCornerDoca<50 && CTBTkrLATEdge < 100

Require:

Probs: CTBBestEnergyProb > .1 && CTBCore > .1

3) Exclude from Training Sample: e^+ with $\cos(\theta) < -.2$ (MM-Shield&ThrmBlnk Conversions)

Goal to meet SRD: ~ .035Hz * 7518 sec = 263 events at 50% CT - Prob.
(in Training Sample) (Including Blanket & MMS Conversions)

Discussion of Pre-Selection Cuts

The "CTcuts"

1) ACD Veto: **(AcdActiveDist3D > 0 || AcdRibbonActDist > 0) && Tkr1SSDVeto < 2**
This is the very minimum ACD requirement – The reconstructed trajectory "hits" a Tile or Ribbon and there are isn't a full Tracker Layer to back it up.

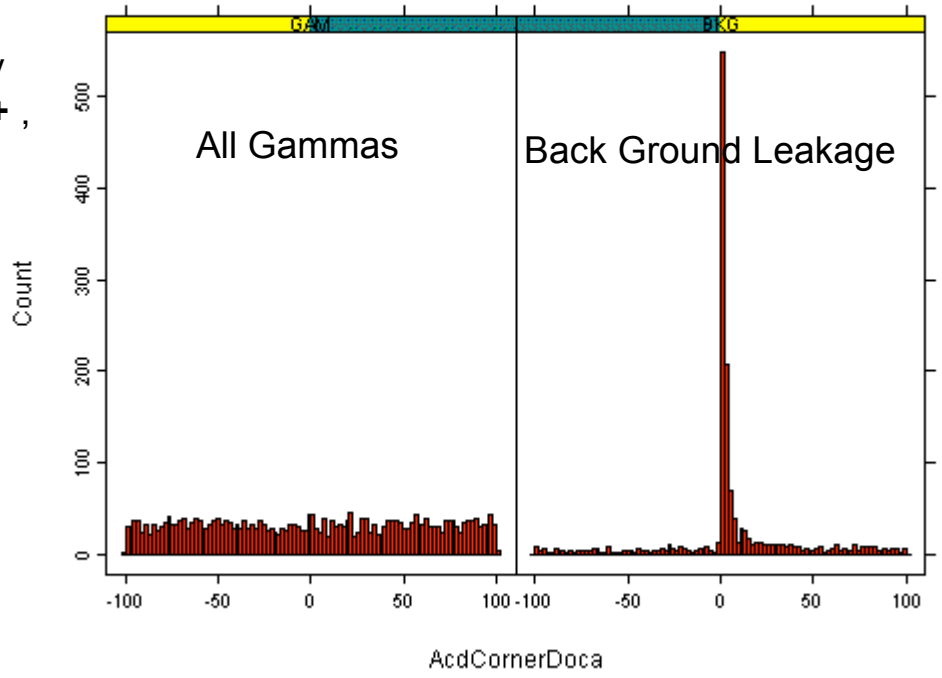
2) Corner leakage: No ACD Ribbons running vertically – Cut out a small piece of Phase Space

AcdCornerDoca: signed DOCA to vertical edge of corner. Signing by "handedness": counter-clockwise **+** , clockwise **–**

Require Track to start within 100 mm of the edge.

AcdCornerDoca > -5 &&
AcdCornerDoca < 50 &&
CTBTkrLATEdge < 100

Cost: .6% loss of All Gammas

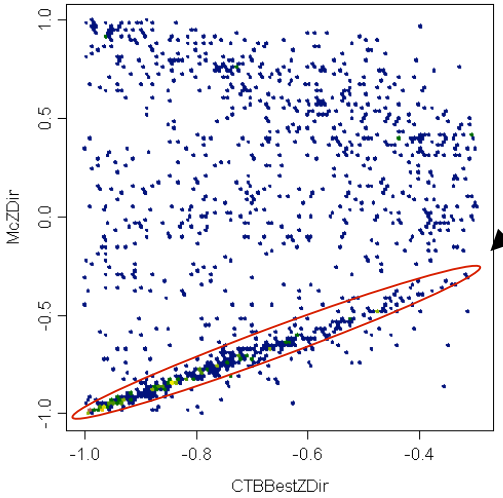


3) Require minimal quality Recon for both Energy & PSF:
CTBCORE > .1 && CTBBestEnergyProb > .1

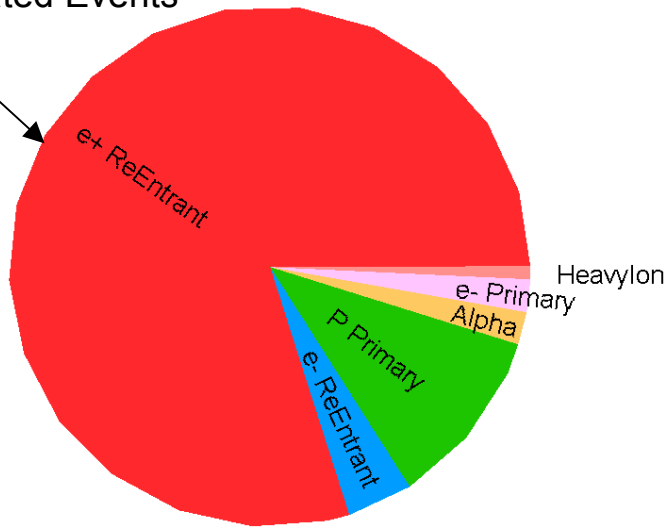
Remove Irreducible Background from CT Training Sample

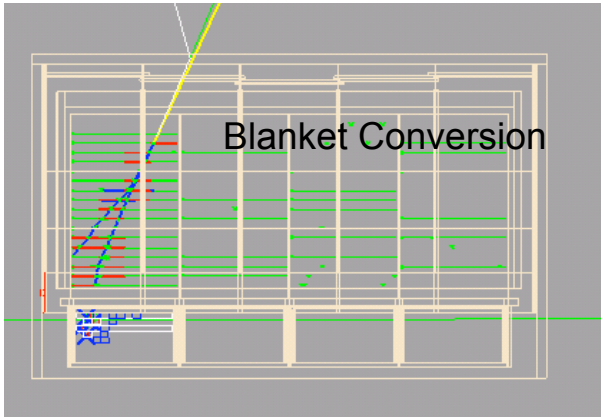
A large portion of the residual backgrounds are unavoidable. These are photons produced in the material ***outside*** the ACD. These pollute the training sample with "signal" in the background sample.

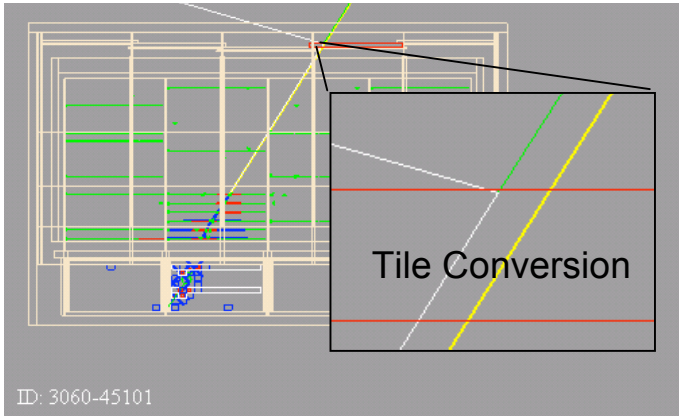
Residual Background



Direction Correlated Events







Remove All e+ with McZDir < -.2 from Training Sample

Analysis Bins

Divide and Conquer:

2 Topological Bins: VTX (VtxAngle > 0) and 1Tkr (VtxAngle == 0)

6 Bins in CalEnergyRaw: < 100, 100-350, 350-1000, 1000-3500, 3500-10000, > 10000 MeV
Note: Previously there was another bin 10-35 GeV – its been eliminated

2 Bins in Conversion Location (lowest Energy Bin only) : Thick Layers (Tkr1FirstLayer < 6)
Thin Layers (Tkr1FirstLayer > 5)

This results in 14 separate analysis bins

Strategy:

In each Bin, identify obvious cuts – reduce background by $\geq 10x$

Follow this by a Classification Tree analysis

- Min. Statistic Req.: Node must have > 20 events to be split and and resulting leaves must have > 7 events
- Use Ensembles of Trees: when possible grow > 3 trees (typ. 5) and average results (this is similar to Random Forests)
- Adjust AG & Bkg. sample sizes to result in Trees with appropriate rejection power at the same Prob. levels across all bins.
- Try Trees based on the full set of variables as well as a reduced set of the most powerful variables

The largest Back Ground Bin

Bin 1, 1Tkr, Thick Radiators

Pre-FIlter

Reject if

(AcdTileCount + AcdRibbonCount) > 0 |
CalEdgeEnergy > 10 | CalELayer7 > 10 |
CalBkHalfRatio > .3 |

} "Hermitically Seal" LAT

20% of remaining Bkg lives here

CalTrackDoca > 250 |
abs(CalLongRms) > 75 | CalTransRms > 60 |
CalMIPRatio > 1. | CalLyr0Ratio > .95 |

} CAL Pattern

	Input	Output
Training Sample	Levels: GAM: 3088 BKG: 139573	Levels: GAM: 1024 BKG: 3678
All	Levels: GAM: 3088 BKG: 176651	Levels: GAM: 1024 BKG: 3879

TkrSurplusHitRatio > 1 | TkrUpstreamHC > 5 |
Tkr1ToTFirst > 4 | Tkr1ToTTrAve > 2.2 |
Tkr1FirstChisq > 20 | TkrNumTracks > 2 |

} Tkr Pattern

(CTBBestEnergyProb + 1.5*CTBCORE) < 1.

} If you have to loose – loose bad ones

Classification Agreement

Input Node - 1Tkr CT All Vars (615)

Training		Predicted		Totals
		BKG	GAM	
Observed	BKG	3612	66	3678
	GAM	293	307	600
Totals		3905	373	4278

	Observed		Overall
	BKG	GAM	
% Agree	98.2%	51.2%	91.6%

Positive Category - GAM

Recall	Precision	F-Measure
51.2%	82.3%	63.1%

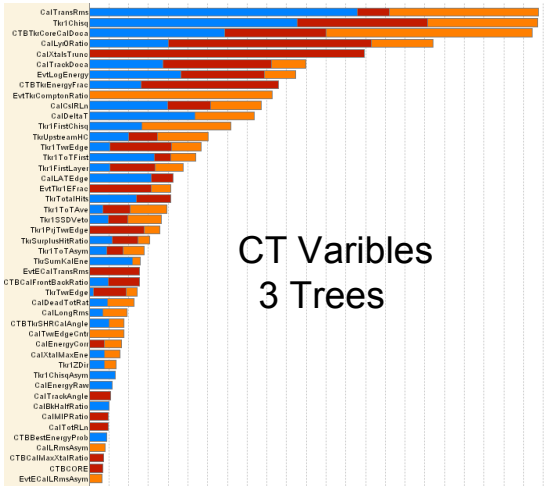
Input Node - CT: BKG 1Tkr E1Thick (721)

All		Predicted		Totals
		GAM	BKG	
Observed	GAM	461	563	1024
	BKG	122	3757	3879
Totals		583	4320	4903

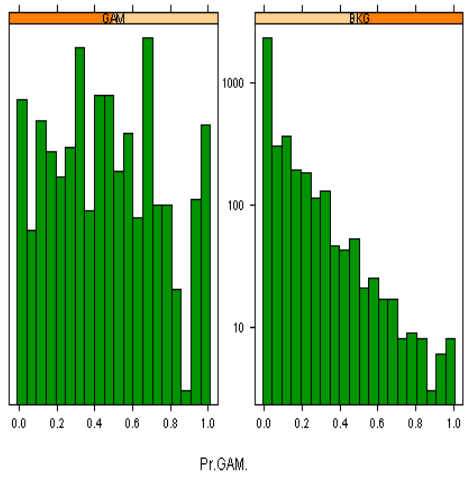
	Observed		Overall
	GAM	BKG	
% Agree	45.0%	96.9%	86.0%

Positive Category - GAM

Recall	Precision	F-Measure
45.0%	79.1%	57.4%



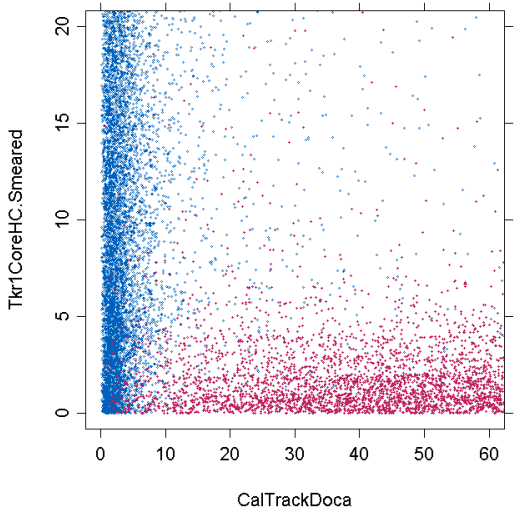
Prob. Distributions

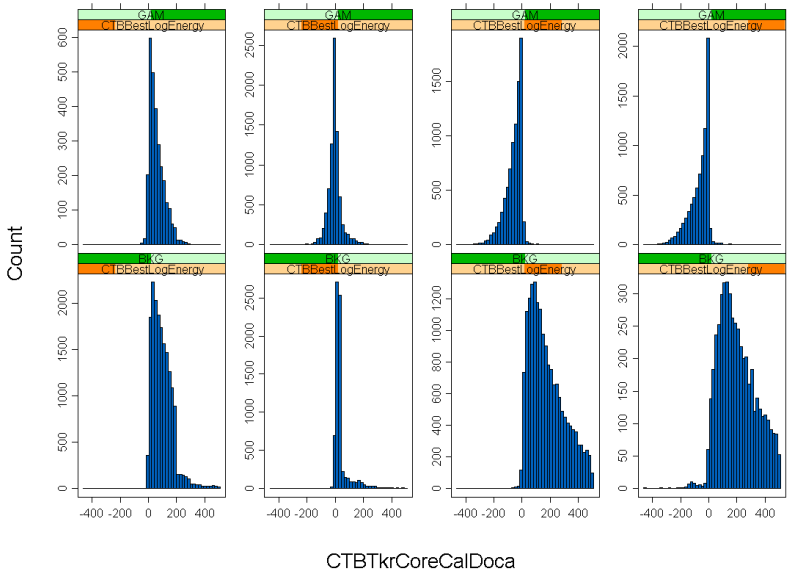


Some Linear Combinations of Variables for Back Ground Rejection

CTBTkrCoreCalDoca = CalTrackDoca – 2.5*Tkr1CoreHC

CTBBestEnergy > 3500

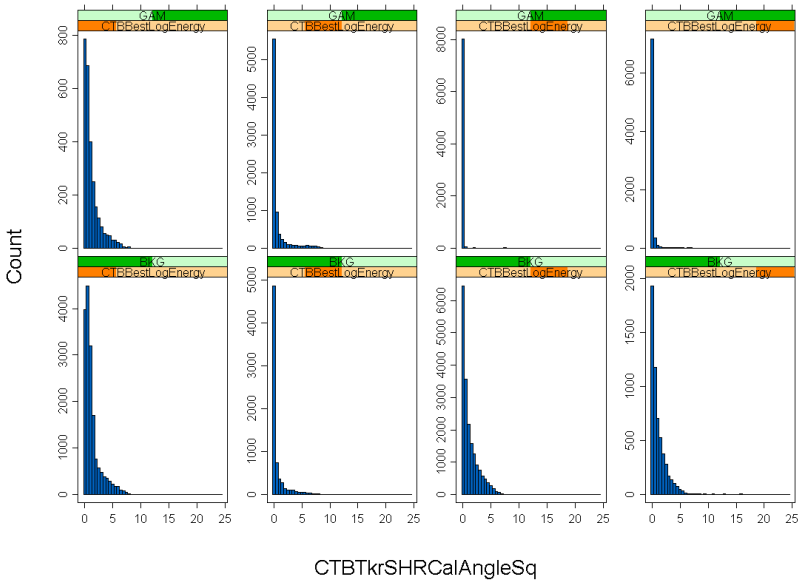




Another example:

CTBTkrSHRCalAngleSq =
(CalTrackAngle - .2*TkrSurplusHitRatio)^2

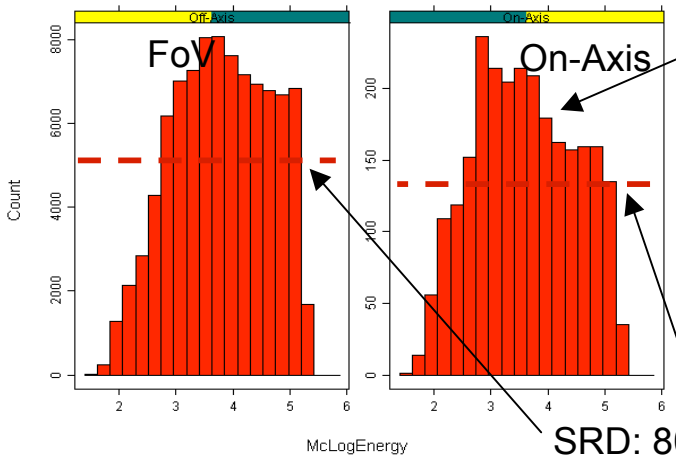
also...
CTBCalTransTCCD =
CalTransRms + .1*CTBTkrCoreCalDoca



The Last Bin: CalEnergyRaw > 10000 MeV

Issue: Self Veto

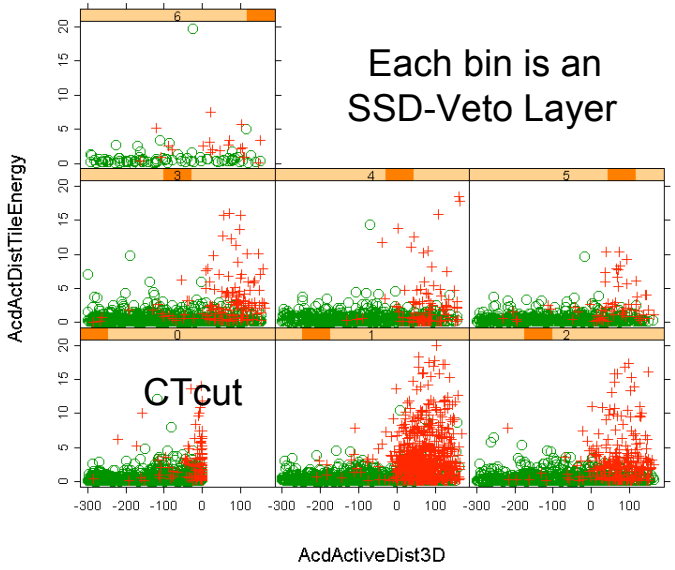
Pass 1 Results:

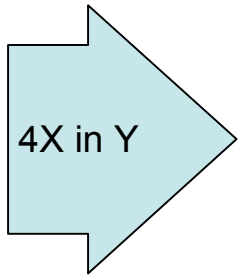


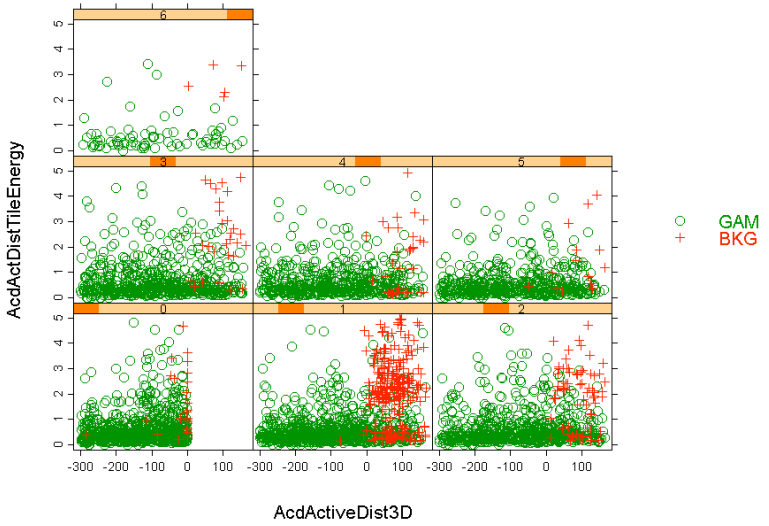
High Energy Fall Off in A_{eff} Persists

SRD: 8000 cm^2 (On-Axis)
1.6 m^2-str A_{eff} x $\Delta\Omega$

Look at AcdActiveDist3D vs AcdActDistTileEnergy







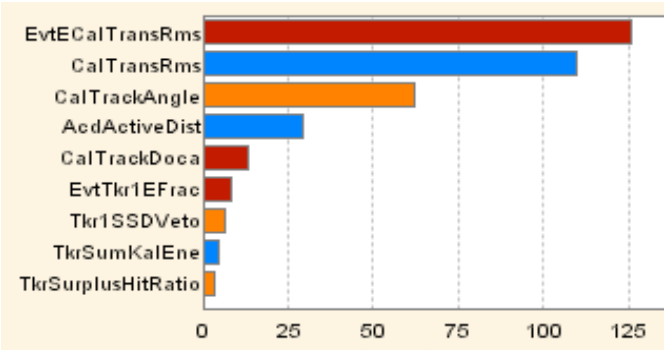
A page from my analysis note book*

PASS 3

1Tkr, Bin 6

PreFIlter: Exclude

Levels:
GAM: 12327
BKG: 18634




```
(AcdActiveDist3D > 0 | AcdRibbonActDist > 0) & Tkr1SSDVeto < 3 |
(AcdActiveDist3D > -10  & AcdActDistTileEnergy > 1 ) |
(AcdActiveDist3D > -50  & AcdActDistTileEnergy > 2 & Tkr1SSDVeto == 0 ) |
CTBTkrCoreCalDoca > 10
```

Levels:
GAM: 11245
BKG: 172

Bkg. Rej.: 108X
AG Eff: 91%

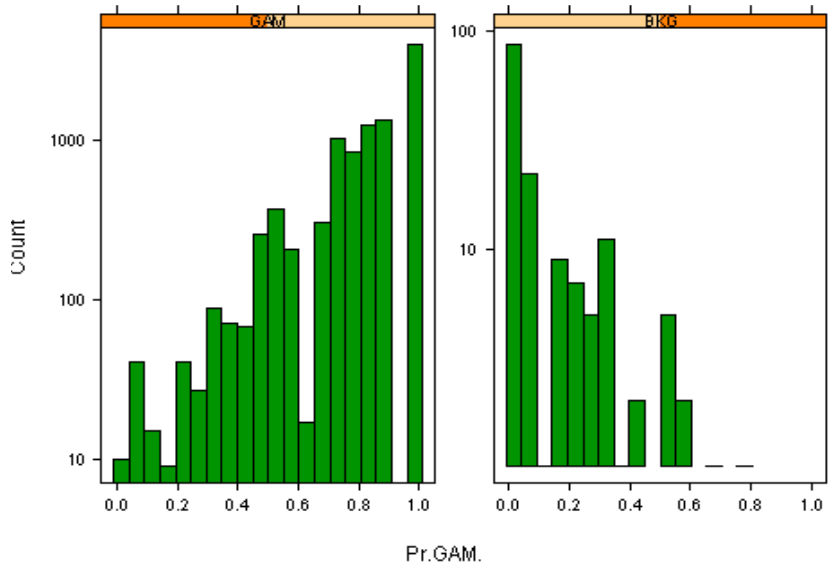
Input Node - Predict: 1Tkr CT All Vars (772)

		Predicted		Totals
		GAM	*BKG*	
Observed	*GAM*	10548	697	11245
	BKG	9	163	172
Totals		10557	860	11417

	Observed		Overall
	GAM	*BKG*	
% Agree	*93.8%*	*94.8%*	*93.8%*

Positive Category - GAM

Recall	Precision	F-Measure
93.8%	*99.9%*	*96.8%*

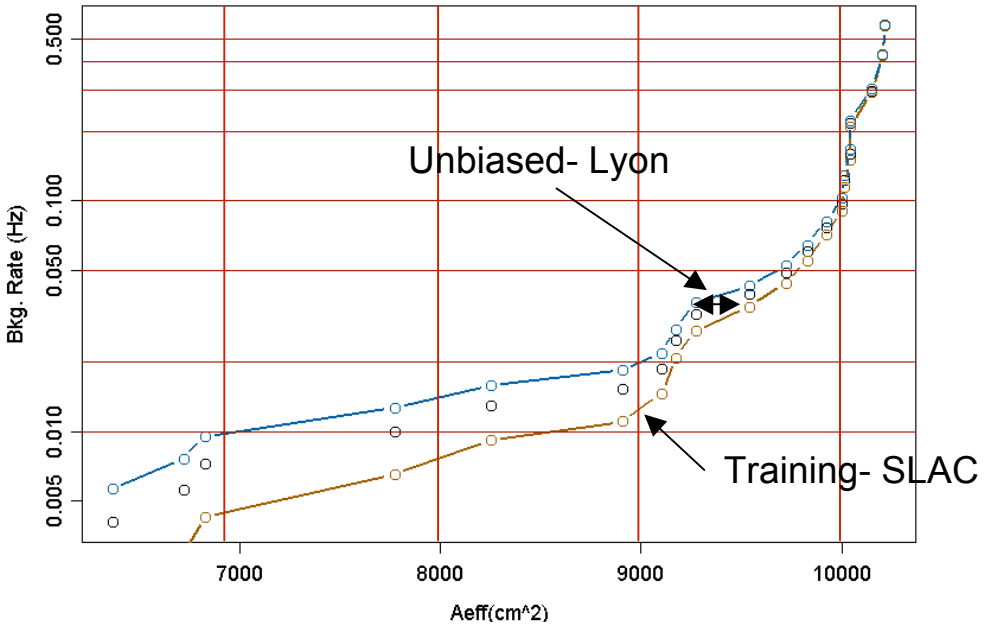


* These will be posted to a confluence page

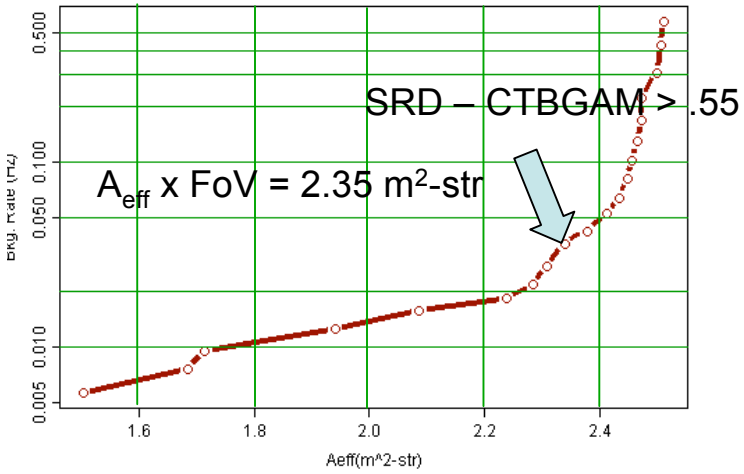
Two Effects giving lower A_{eff} Results

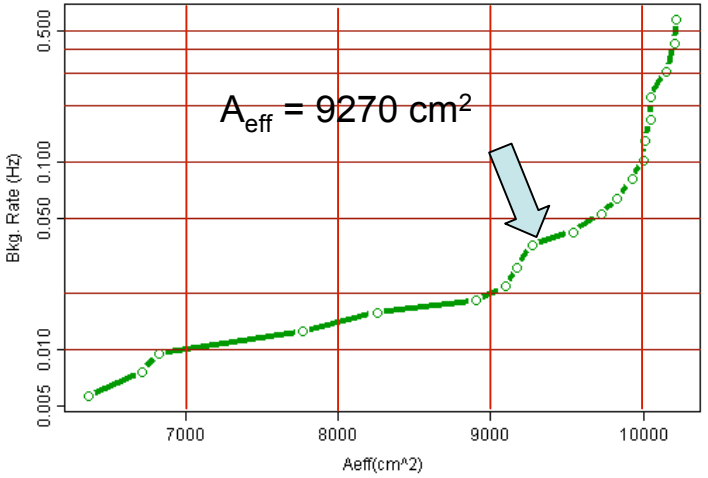
1) Normalization Error in 2M AG set (1.91M vs 2.0M) - 5%

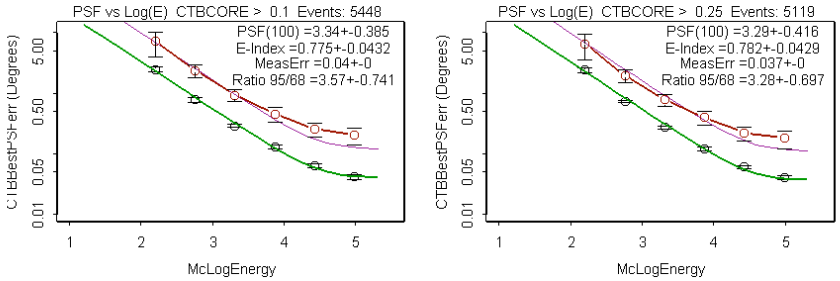
2) Training bias (see Plot) - 3%

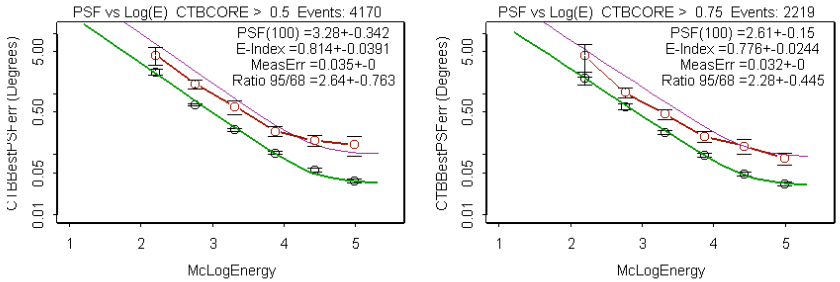


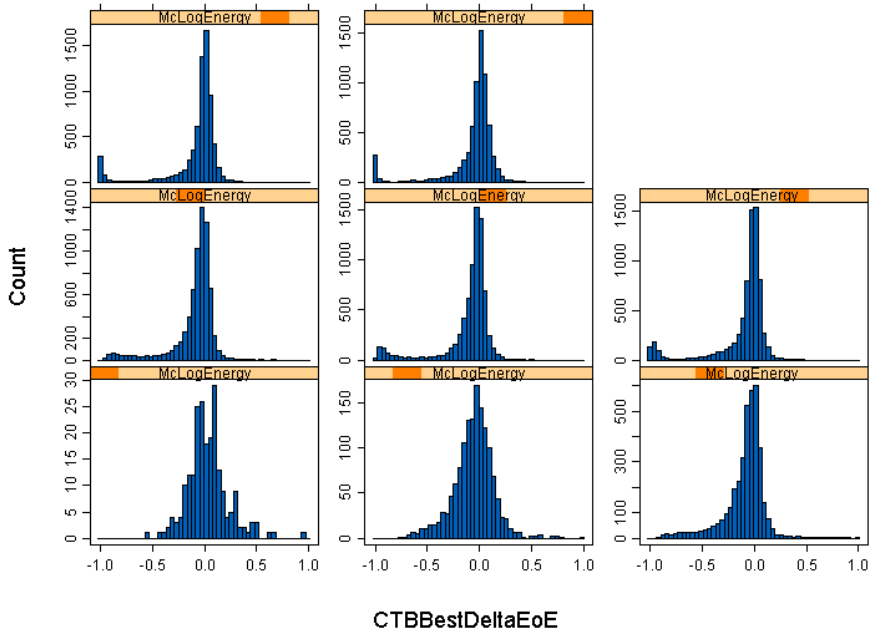
Lyon Data Sets Results:

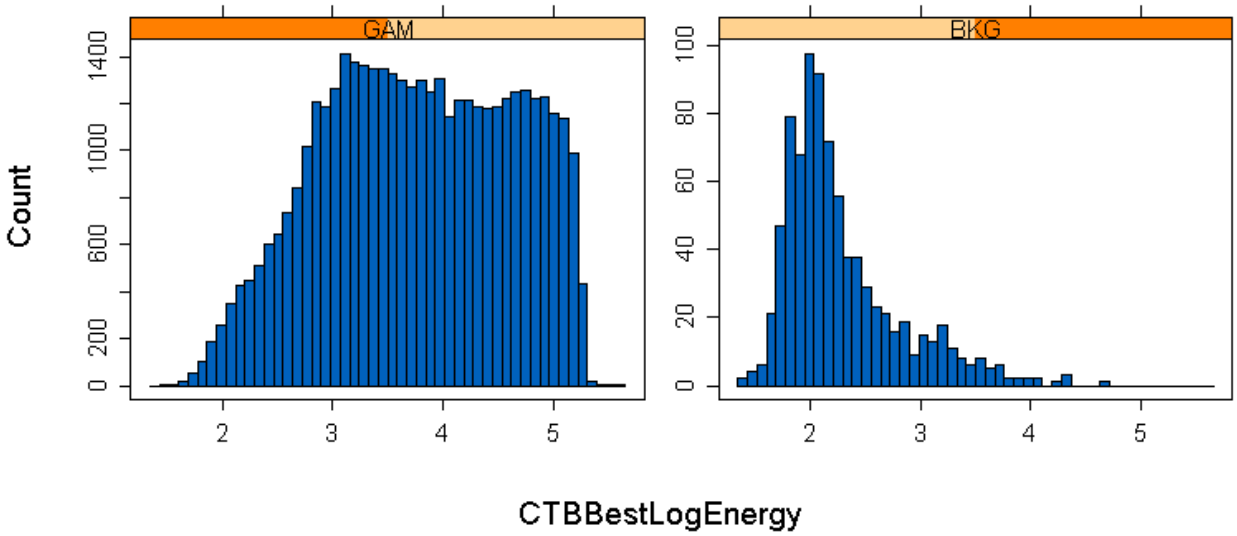












Base Class 2:

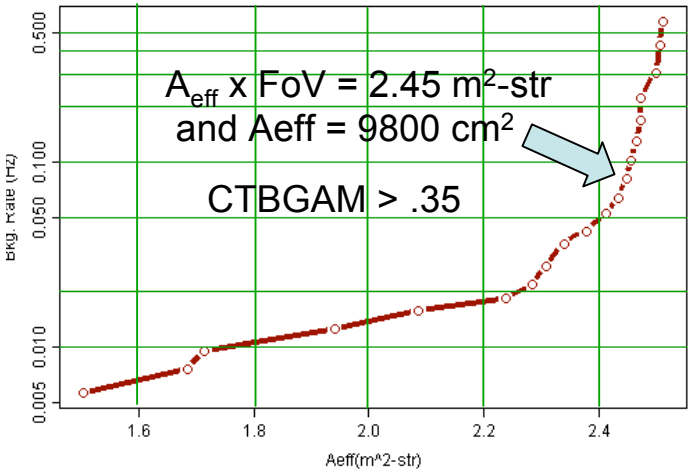
CTBGAM > .55
CTBCORE > .1
CTBBestEnergyProb > .1

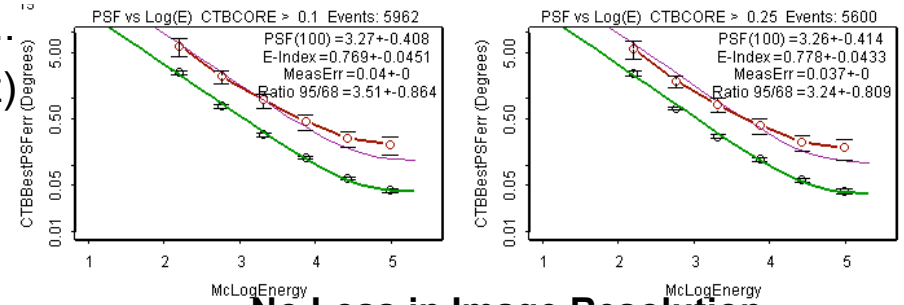
Analysis Classes

We easily meet the SRD. That's not necessarily the best science!

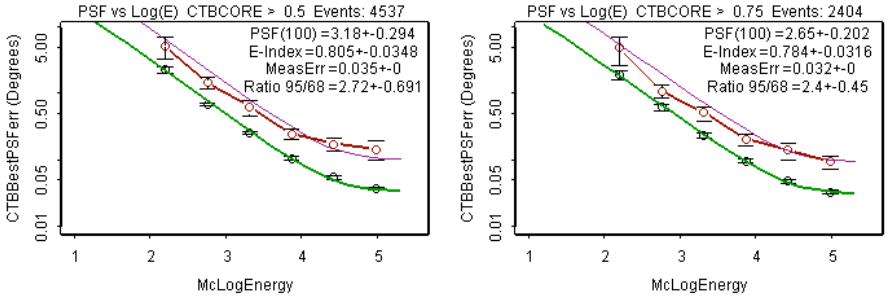
Back Off CTBGAM until no further gains ..

At CTBGAM > .35 we're at 2X SRD (.07 Hz)

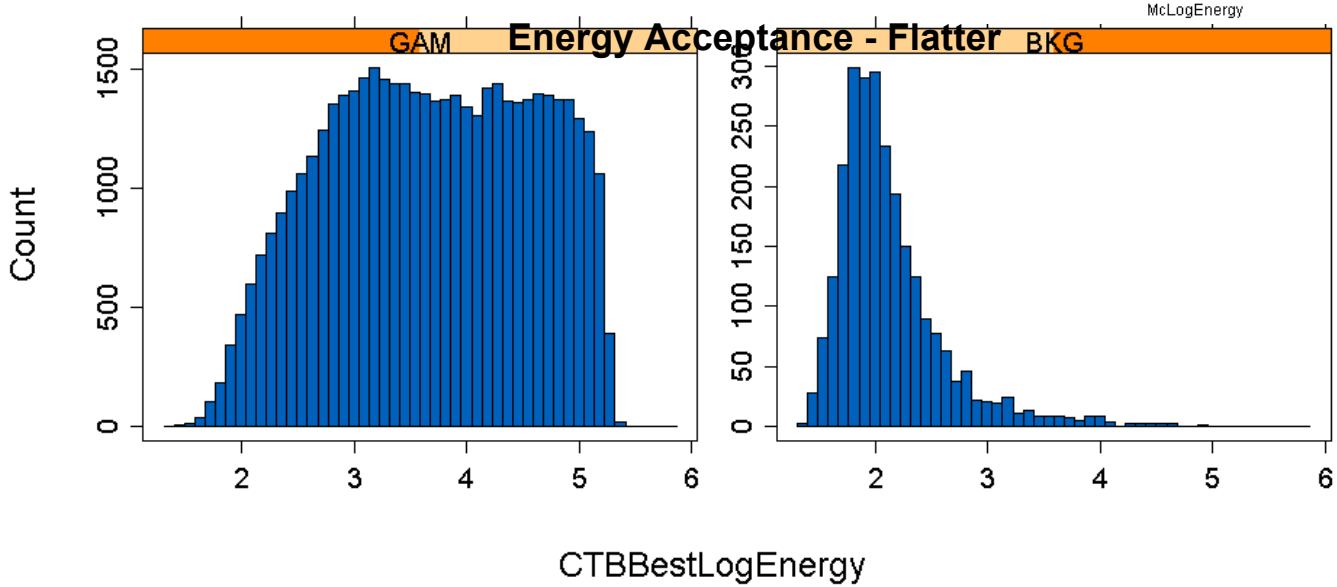




No Loss in Image Resolution



Energy Acceptance - Flatter



Base Class 1:

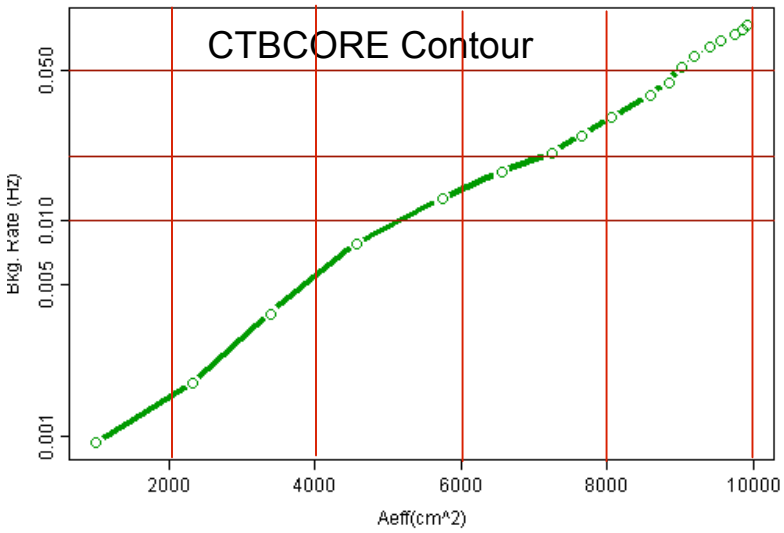
CTBGAM	> .35
CTBCORE	> .1
CTBBestEnergyProb	> .1

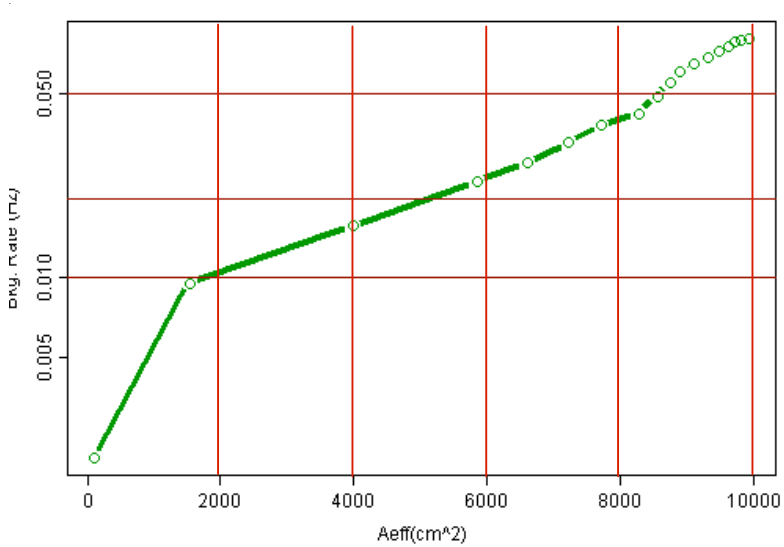
GLAST

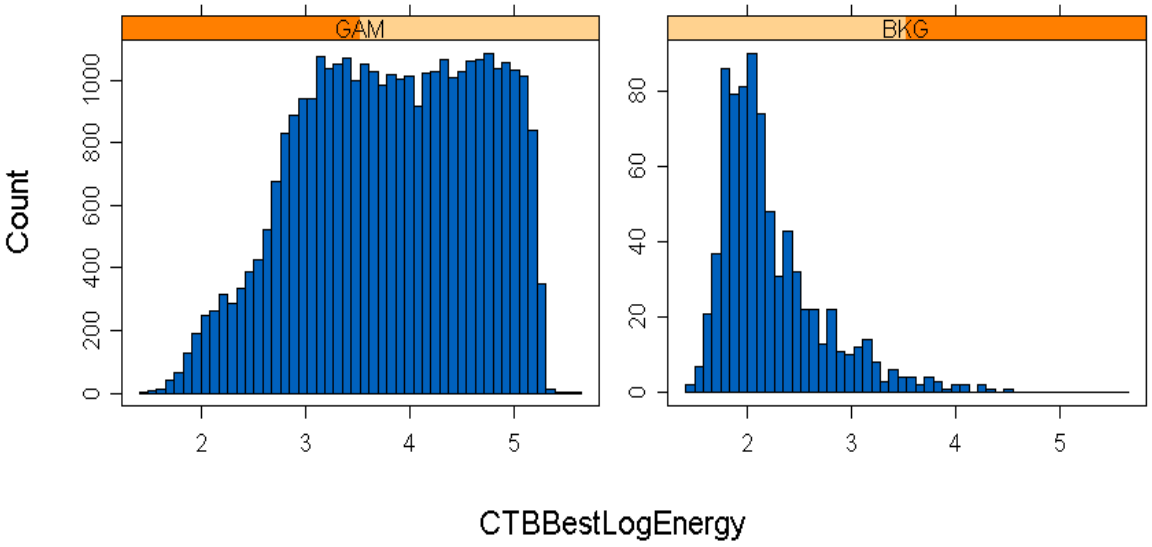
Other end of Knob Space

Go as far in Energy & Image as necessary to get to SRD Background Rejection

After considerable searching I arrived at:







Rate of gain in both Image and Energy Res. slow and similar.

Finally settle on:

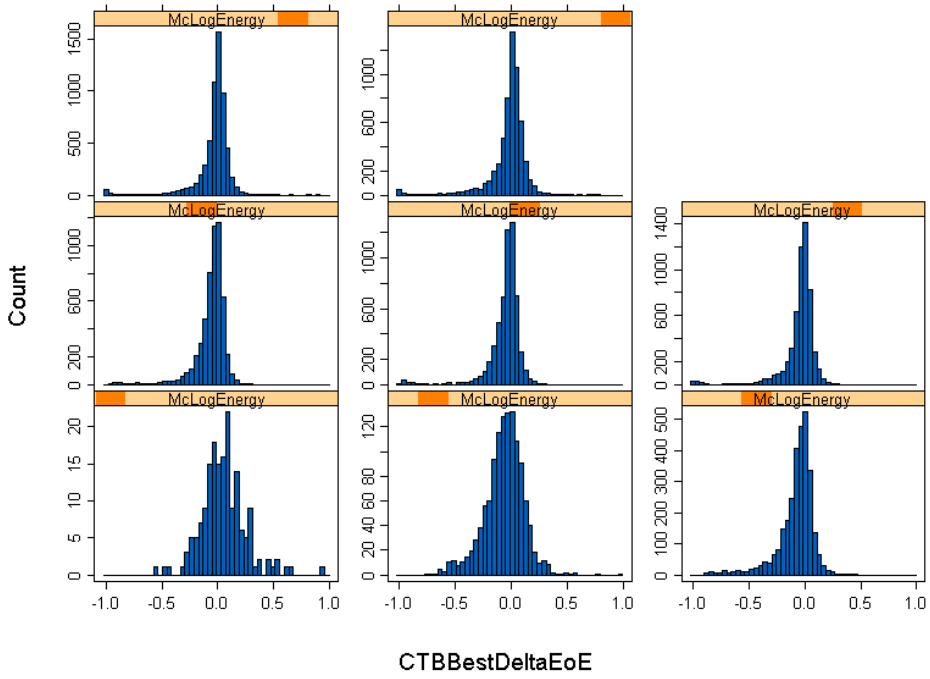
Base Class 3:

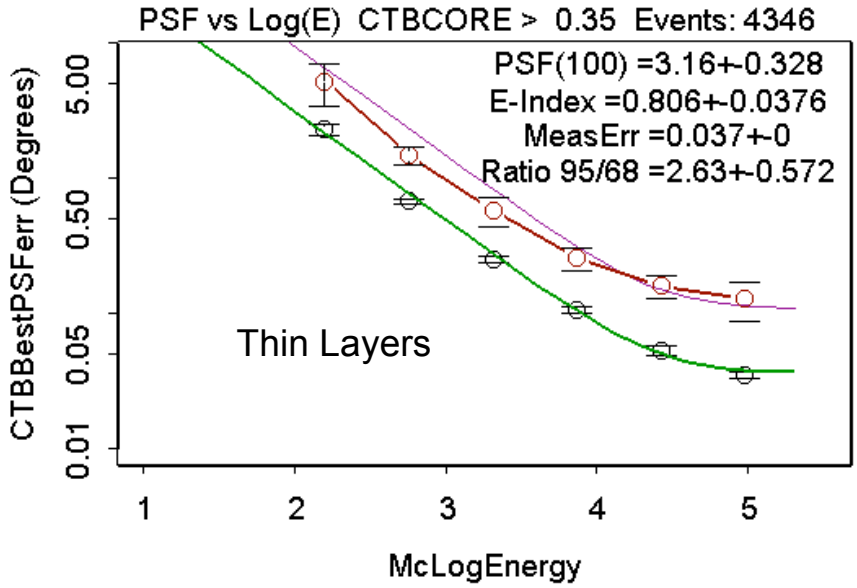
CTBGAM > .50

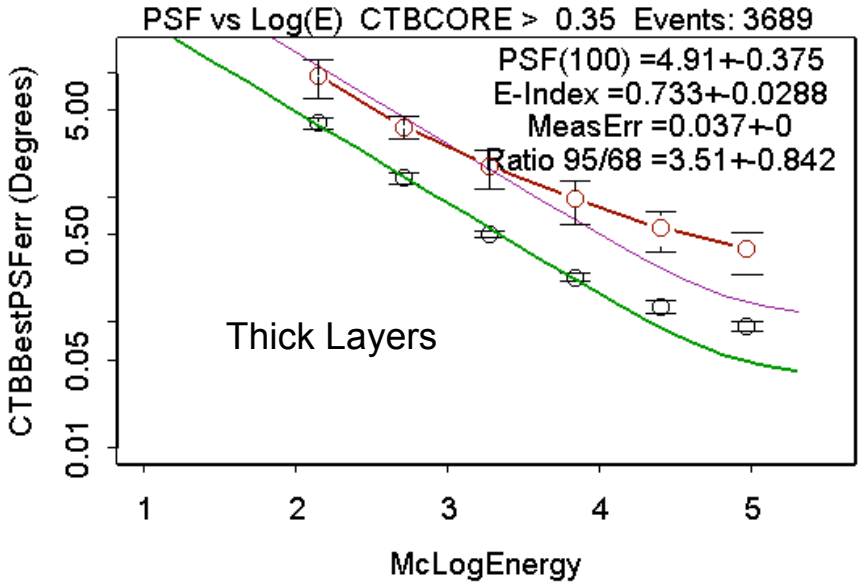
CTBCORE > .35

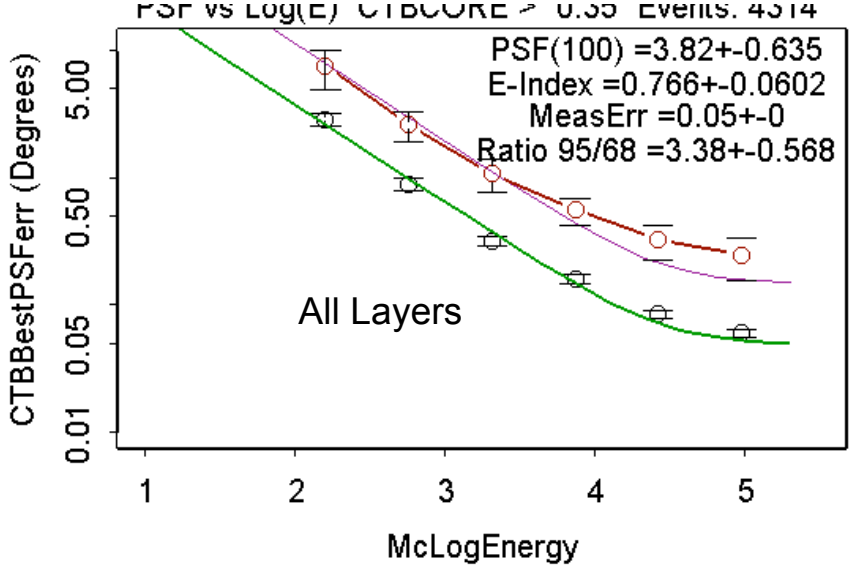
CTBBestEnergyProb > .35

A_{eff} = 8200 cm^2; A_{eff} x FoV = 1.72 m^2-str


McLogEnergy
Count
CTBBestDeltaEoE


PSF vs Log(E) CTBCORE > 0.35 Events: 4346
PSF(100) =3.16+-0.328
E-Index =0.806+-0.0376
MeasErr =0.037+-0
Ratio 95/68 =2.63+-0.572
Thin Layers
CTBBestPSFerr (Degrees)
McLogEnergy


PSF vs Log(E) CTBCORE > 0.35 Events: 3689
PSF(100) =4.91+-0.375
E-Index =0.733+-0.0288
MeasErr =0.037+-0
Ratio 95/68 =3.51+-0.842
Thick Layers
CTBBestPSFerr (Degrees)
McLogEnergy


PSF(100) =3.82+-0.635
E-Index =0.766+-0.0602
MeasErr =0.05+-0
Ratio 95/68 =3.38+-0.568
All Layers
CTBBestPSFerr (Degrees)
McLogEnergy

Summary:

To start the discussion: 3 Base Classes Proposed

	A_{eff} (cm^2)	A_{eff} x FoV (m^2-str)	PSF_{100}	PSF 95/68	$\Delta E/E$ Noise%	Bkg/Diffuse
Base Class 1	9800	2.45	3.3	3.5	11	.2
Base Class 2	9270	2.35	3.3	3.3	11	.1
Base Class 3	8200	1.72	3.2	2.6	3.5	.1

Comments:

At high energy – need additional rejection for high energy electrons – This was incorporated in this analysis – more later

At low energy – limitations imposed by large gaps between CAL modules. Does not allow for hermetic sealing from below.

Hand scan (not discussed here) suggested problems in ACD digi. / analysis.

More then 50% remaining events are IRREDUCIBLE – They're gammas d### it!