Status of PCAMDF XY in arislite

A W Watson | 25 July 2016 | Temple U

What is PCAMDF?

PCA (Principal Component Analysis) is a well-documented mathematical procedure. The process to calculate the covariance matrix can be found at docdb 1518. After we calculate this matrix, we find its eigenvalues and eigenvectors and transform our array of PMT S2 fractions from "feature space" to "pattern space" by moving them into this eigenvector space. We can move from F-space to P-space and vice versa as desired. PCA finds patterns in the data and reduces noise.

MDF (Multi-Dimensional Fitting) involves finding a functional relationship between our independent variables (the pattern space vector) and our dependent variable (either X or Y). This is also well-documented in literature.

I use the ROOT classes TPrincipal and TMultiDimFit to perform my PCAMDF.

Training PCAMDF on MC

12,601 uniformly-distributed (in XYZ) ^{39}Ar events available at:

//ds50srv01/scratch/darkside/awatson/ARISMC.root

The PCA has 0 degrees of freedom:
it's just a step-by-step processing of data analysis.

The MDF has several degrees of freedom:*
- limit the number of terms to avoid over-fitting
- limit the powers of the terms in the fit
- limit the number of terms in the fit
- limit the number of trial fits performed
- weight the radial distribution of training events

Automated parameter tuning:

1. generate $O(10^6\text{-}10^7)$ MDF's with random values selected for these DoF's (within some reasonable ranges)
2. test the PCAMDF on a different event sample (events from the end of the MC file) and find avg. reconstruction error
3. keep new tuning if new average error is smaller than old average error

*See *H. Wind. Function parameterization. Proceedings of the 1972 CERN Computing and Data Processing School, volume 72-21 of Yellow report. CERN, 1972.* for a thorough explanation of the mathematics behind the MDF implemented in TMultiDimFit

drvec = scalar distance from true MC position to reconstructed position
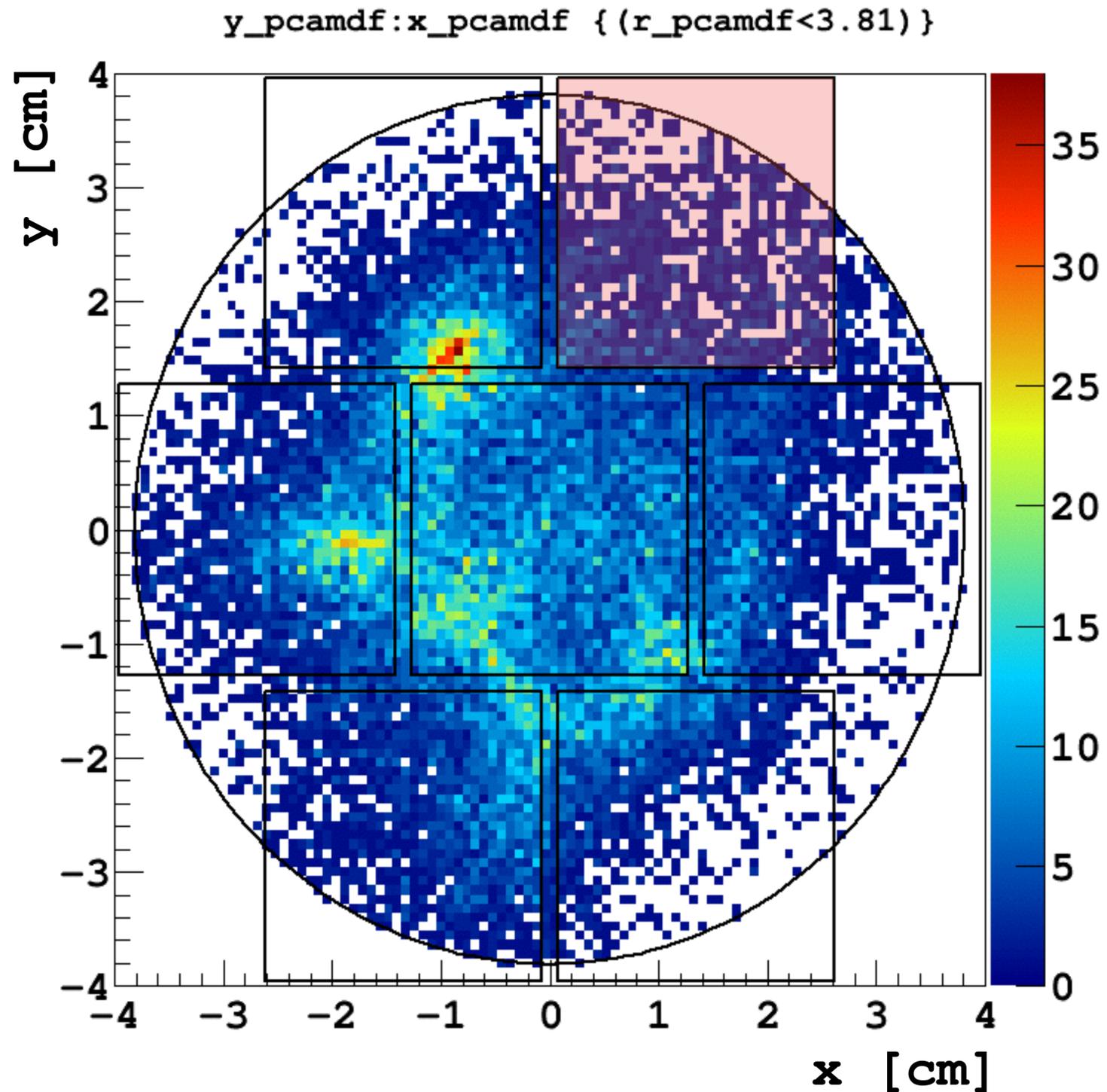
t = theta
r = radius
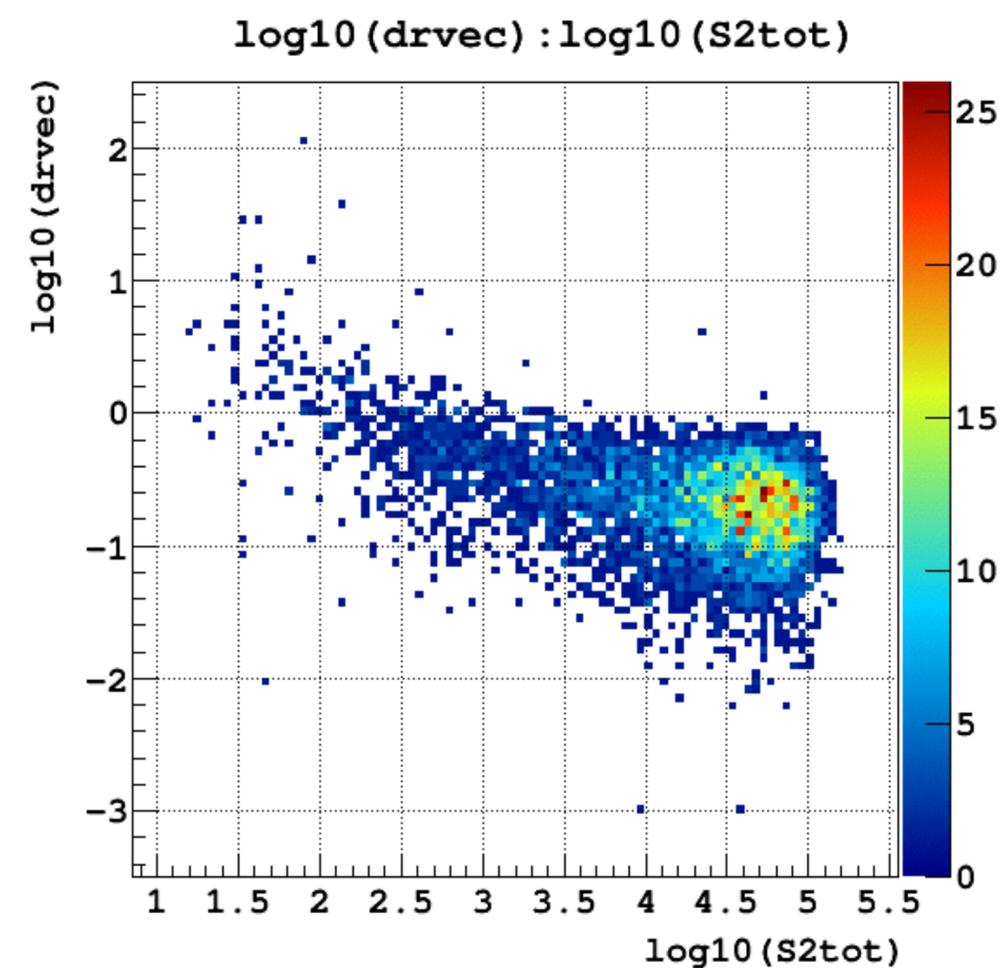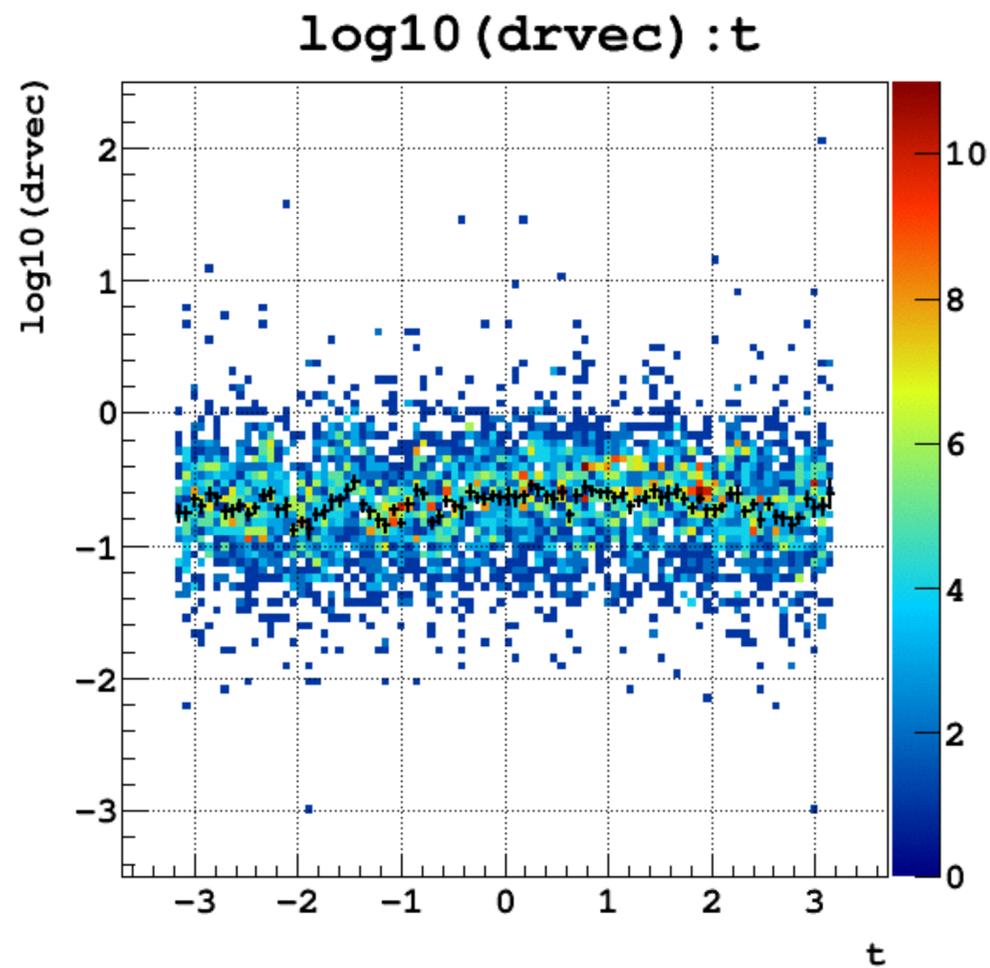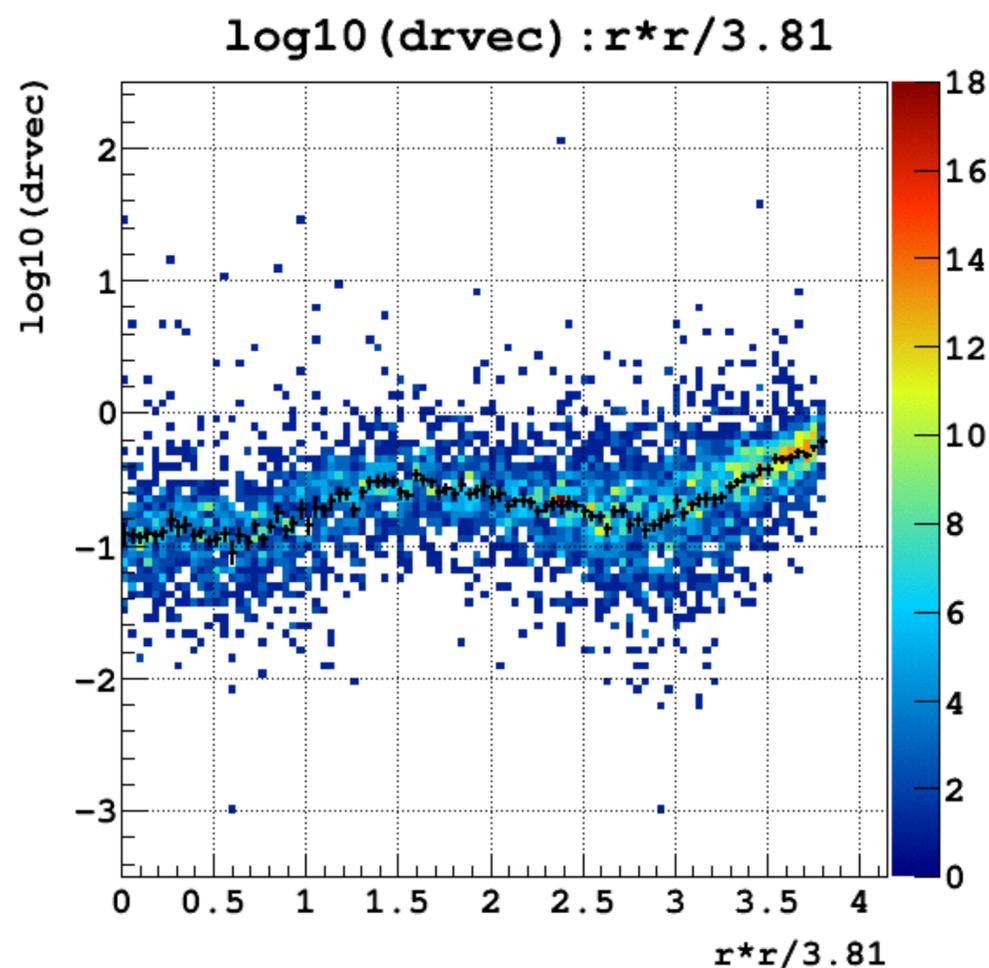
As in DS-50, the PCAMDF XY reconstruction accuracy is largely radially-dependent and energy dependent. In spite of a missing outer-ring PMT, there doesn't appear to be a significant angular dependence in the reconstruction accuracy.

$\langle |\mathbf{x-x'}| \rangle$ = 1.952 mm
$\langle |\mathbf{y-y'}| \rangle$ = 2.140 mm
$\langle |\mathbf{r-r'}| \rangle$ = 2.462 mm
$\langle |\mathbf{t-t'}| \rangle$ = 8.193°
$\langle |\vec{\mathbf{r}}-\vec{\mathbf{r}}'| \rangle$ = 3.232 mm
$\langle \mathbf{drvec} \rangle$

Performance on data

y_pcamdf:x_pcamdf {(r_pcamdf<3.81)}

Status

Of 100,000 events in Run 100764 (57Co run)…

…**68,876** events not reconstructed because ≥ 1 PMT had < 1PE S2 OR S2 was written incorrectly (isnan error)

4,071 events reconstructed outside r = 3.81 cm

68,876 + 4,071 = **72,947** "bad" events

(Compare to ~**69,000** events reconstructed outside r = 3.81 cm with barycenter method)

Challenges

MC is not yet tuned — leads to poor PCAMDF XY.

The red-colored PMT is PMT #3, a Xe PMT, and is not included in the PCAMDF.

We're having some data-processing issues (overshoot, etc.) with the bottom PMT, and so S2 TBA is also not included in the PCAMDF.

Work to be done

MC tuning!
- Needs to better match data to have a good XY
- …but first, we need data with all 8 PMTs running
- …and we need to fix the problems with the 3" PMT

Different PCAMDF for different PMT configurations:
- Cannot "adjust" a PCAMDF tuning to add/remove PMTs
- Need one for 6 top PMTs + 1 bottom PMT
- Need one for all 8 PMTs

Optimization of barycenter, possibly *à la*
http://arxiv.org/pdf/1309.5561v2.pdf

Code is available at…

arislite//modules/include/pcamdf_tuning/
arislite//modules/include/XY_PCAMDF.hh
arislite//modules/src/XY_PCAMDF.cc

get the code with
git clone git@gitlab.in2p3.fr:ARIS/arislite.git

Extra Slides

Barycenter vs. PCAMDF XY for Run 100764