
MURAVES

towards full pythonisation – bug fix

2026-04-07

Alice Biolchini, CP3, UCLouvain

Pic: <https://www.vesuviusnationalpark.it/>

Pythonisation of reconstruction script

- Python version of the main reconstruction scripts is available on [MURAVES GitHub](#):
kept same structure as before.
 - **MURAVES_reco.cpp → MURAVES_reco.py**
 - **ReadEvent.cc and ReadEvent.h → ReadEvent.py**
 - **ClusterLists.cc and ClusterLists.h → ClusterLists.py**
 - **Tracking.cc and Tracking.h → Tracking.py**
 - **EvaluateAngularCoordinates.cc and .h → EvaluateAngularCoordinates.py**

Validation

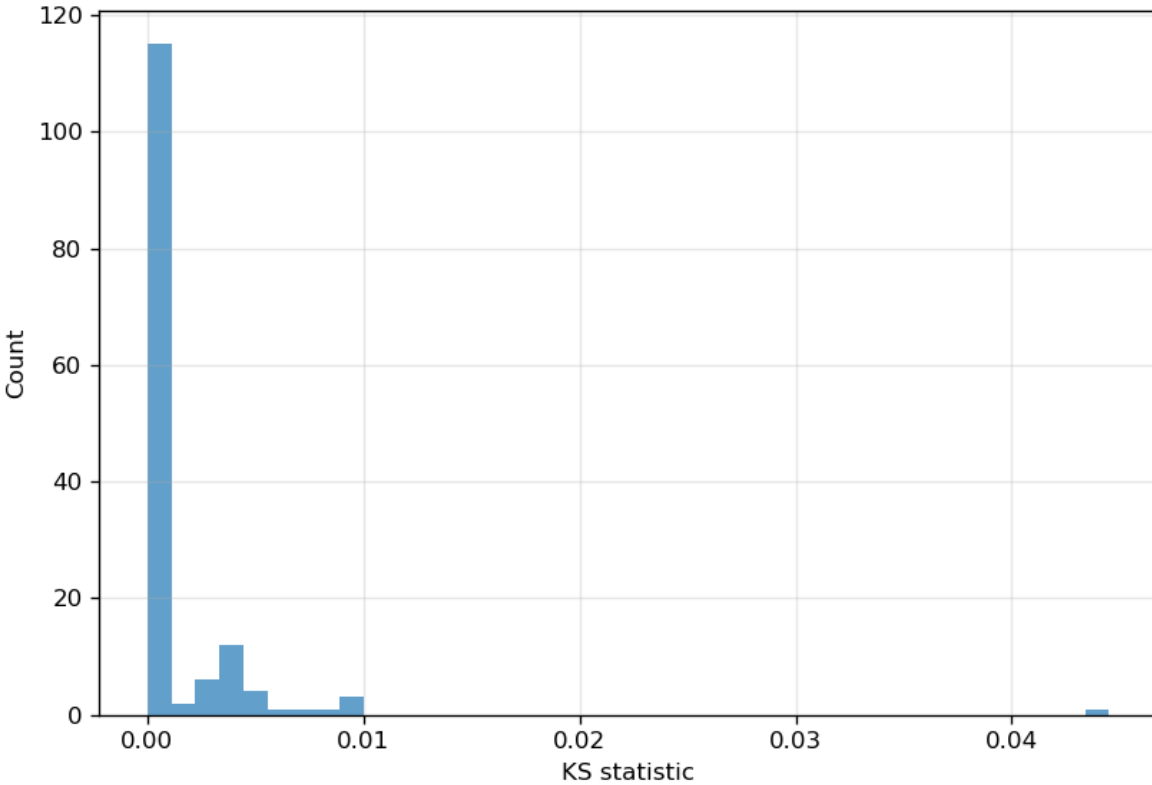
- Comparing root outputs using **run NERO 2546**.
 - Kolmogorov-Smirnov test: unbinned test (larger distance between cumulative distributions).
 - 146 variables in total
 - 1 variables have p-value <5%, *meaning it is very unlikely that the two distributions comes from the same underlying distribution.*
 - Results: Tables available in backup
- In this presentation I will show:
 - Summary histograms of the KS and p-values over the 146 variables
 - Sanity checks: variables that should exactly the same
 - **Strips ID after corrections**
 - Variable have p-value <5%

Summary histograms

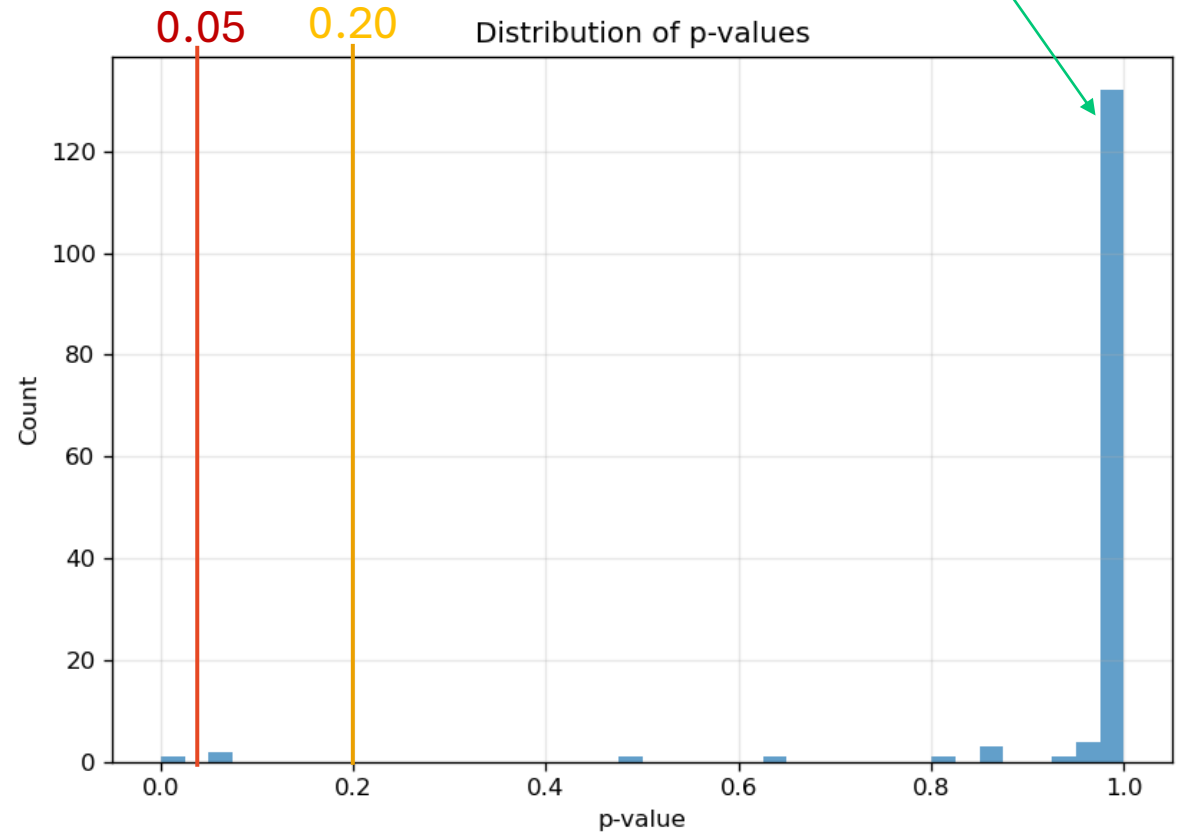
Sanity checks:

Variables that are not subjected to any processing should be identical, KS statistic= 0 and p-value= 1.

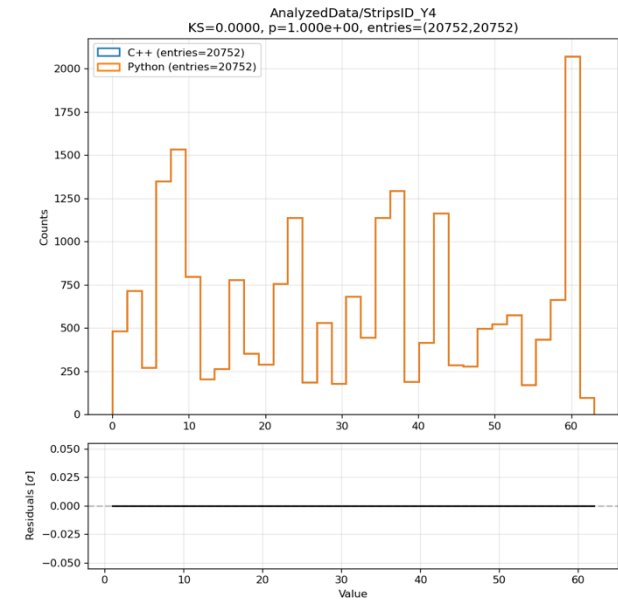
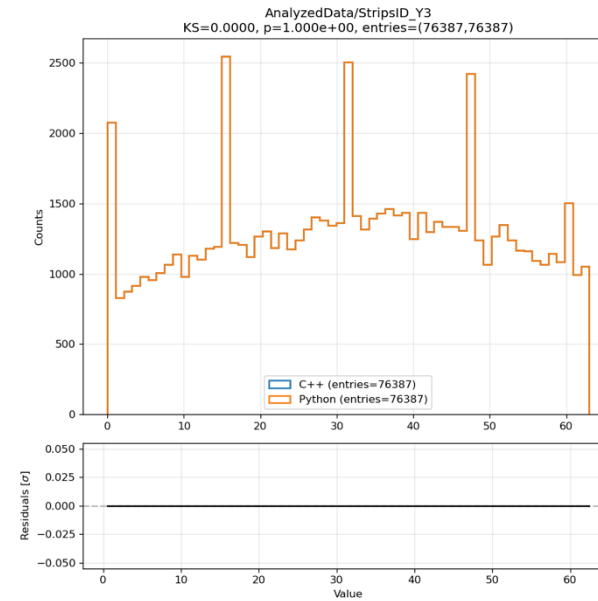
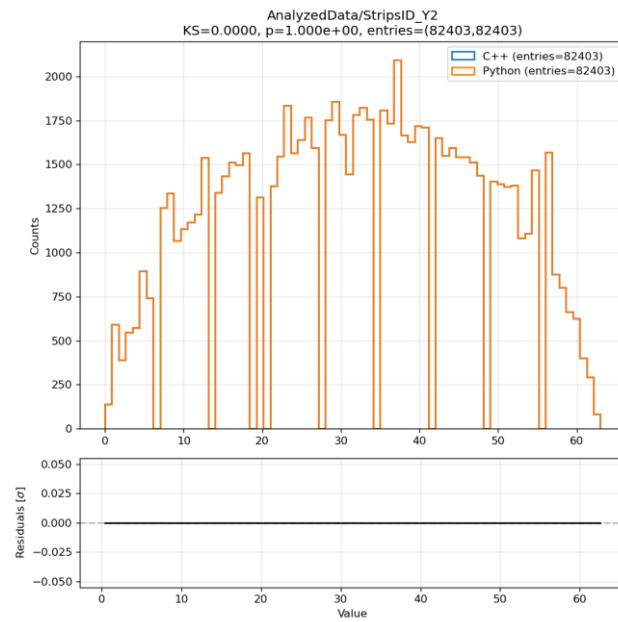
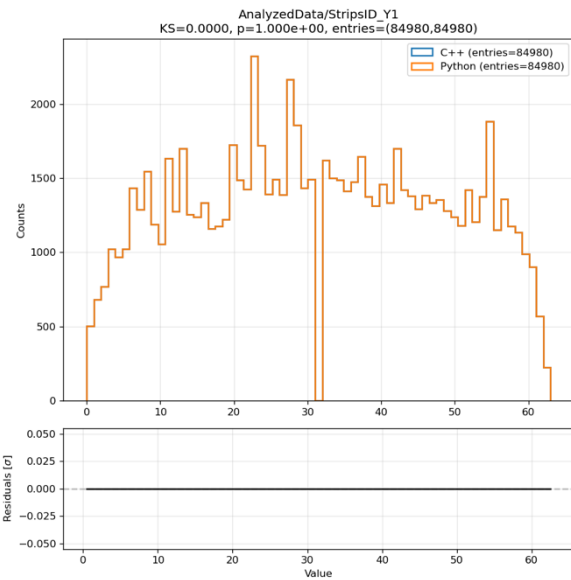
Distribution of KS statistics



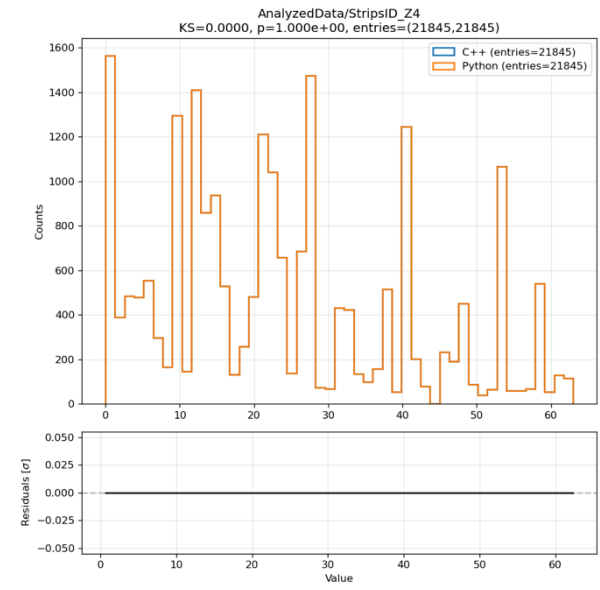
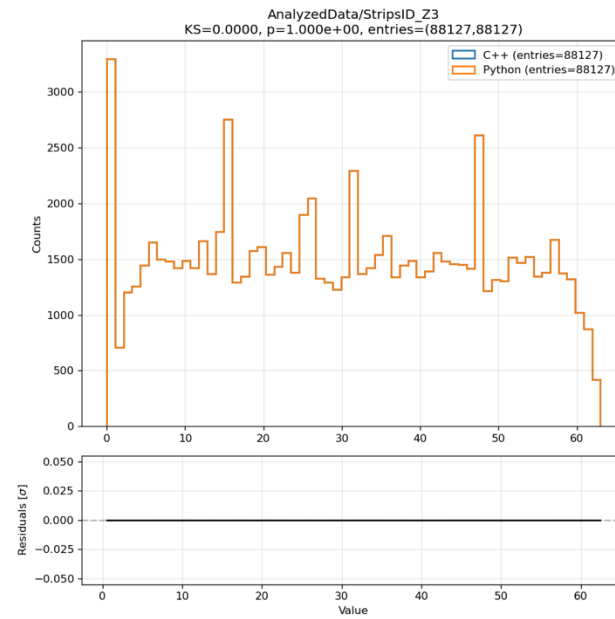
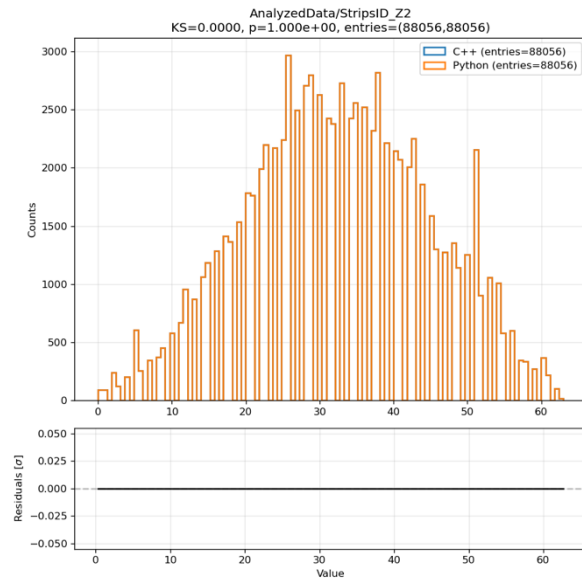
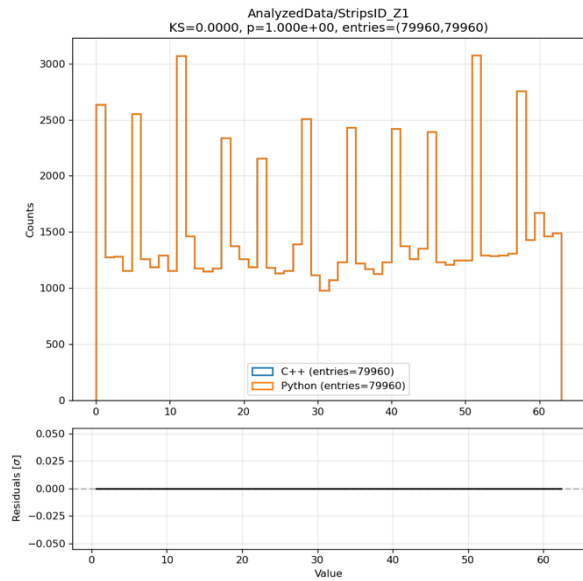
Distribution of p-values



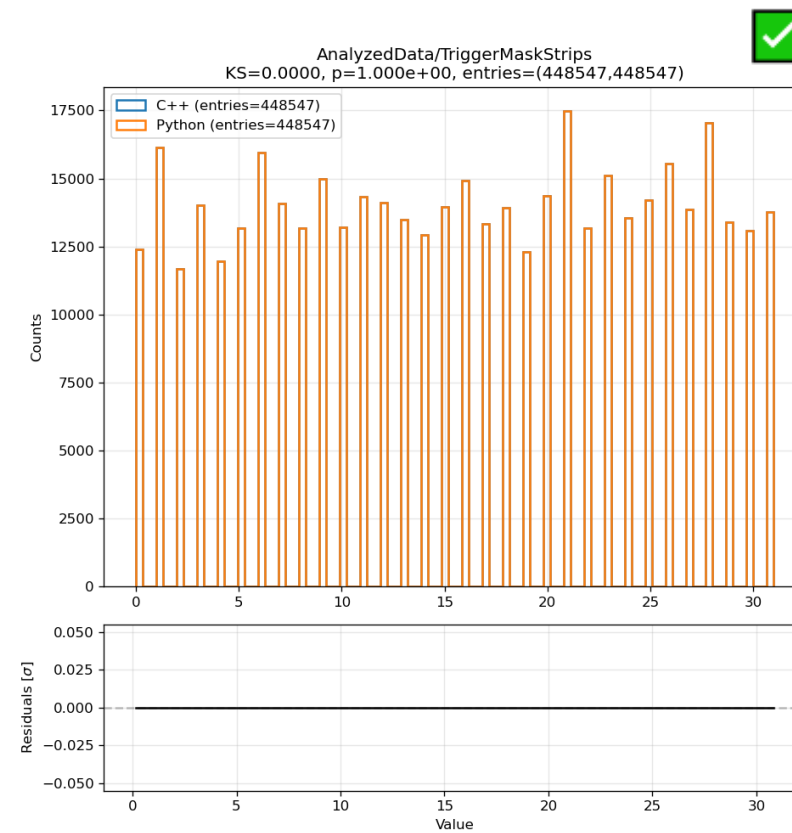
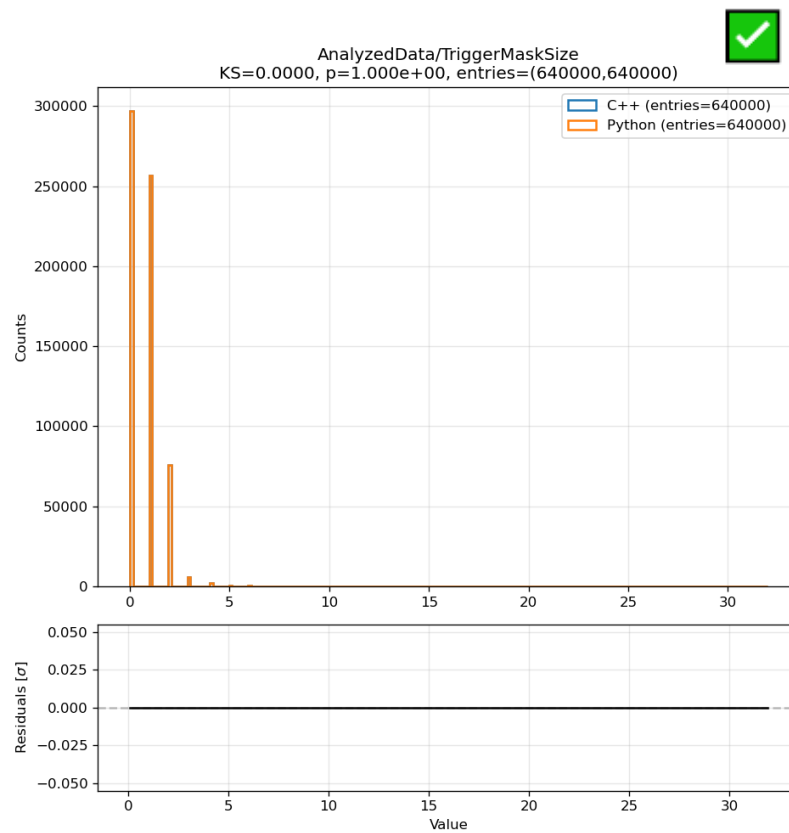
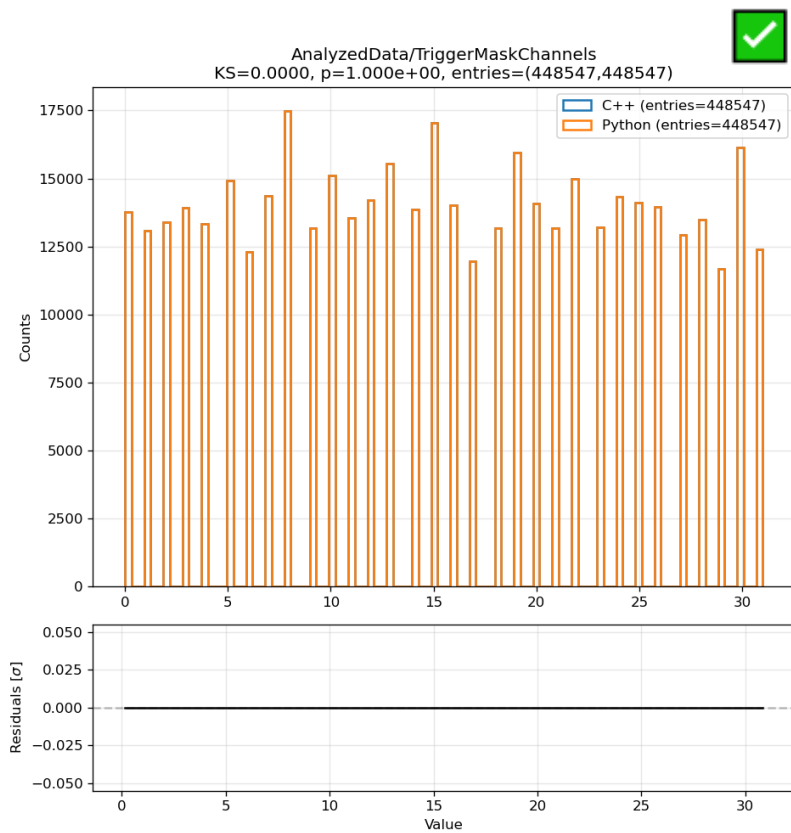
Sanity check: StripsID Y



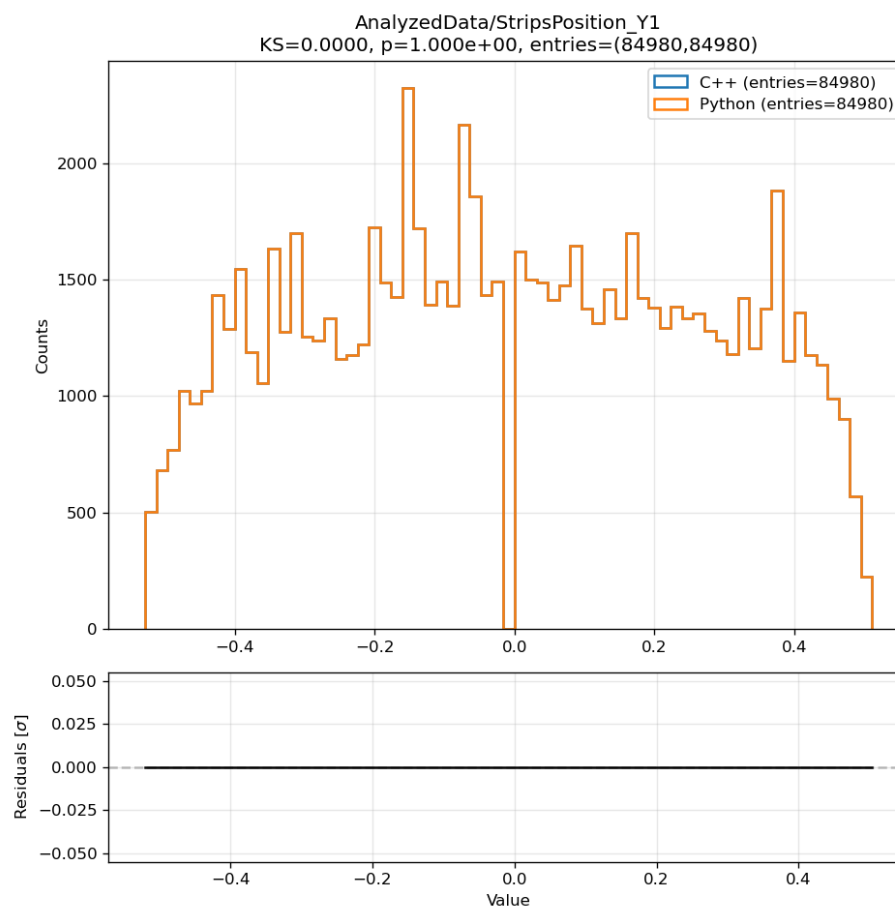
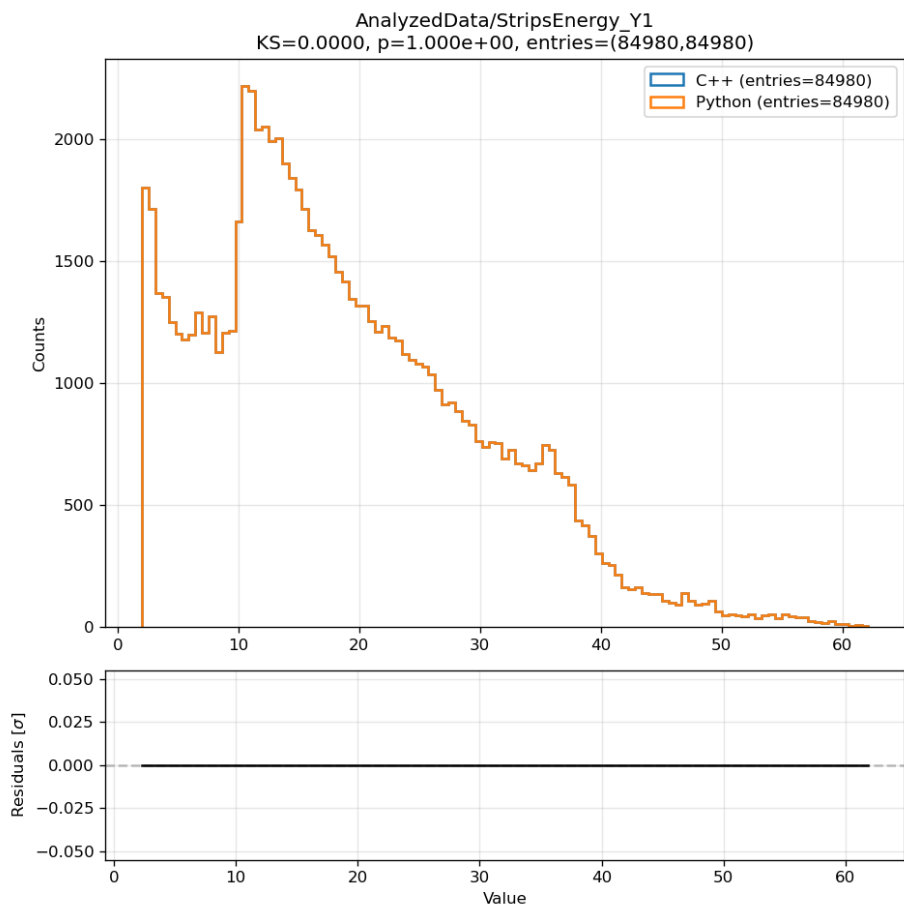
Sanity check: StripsID Z



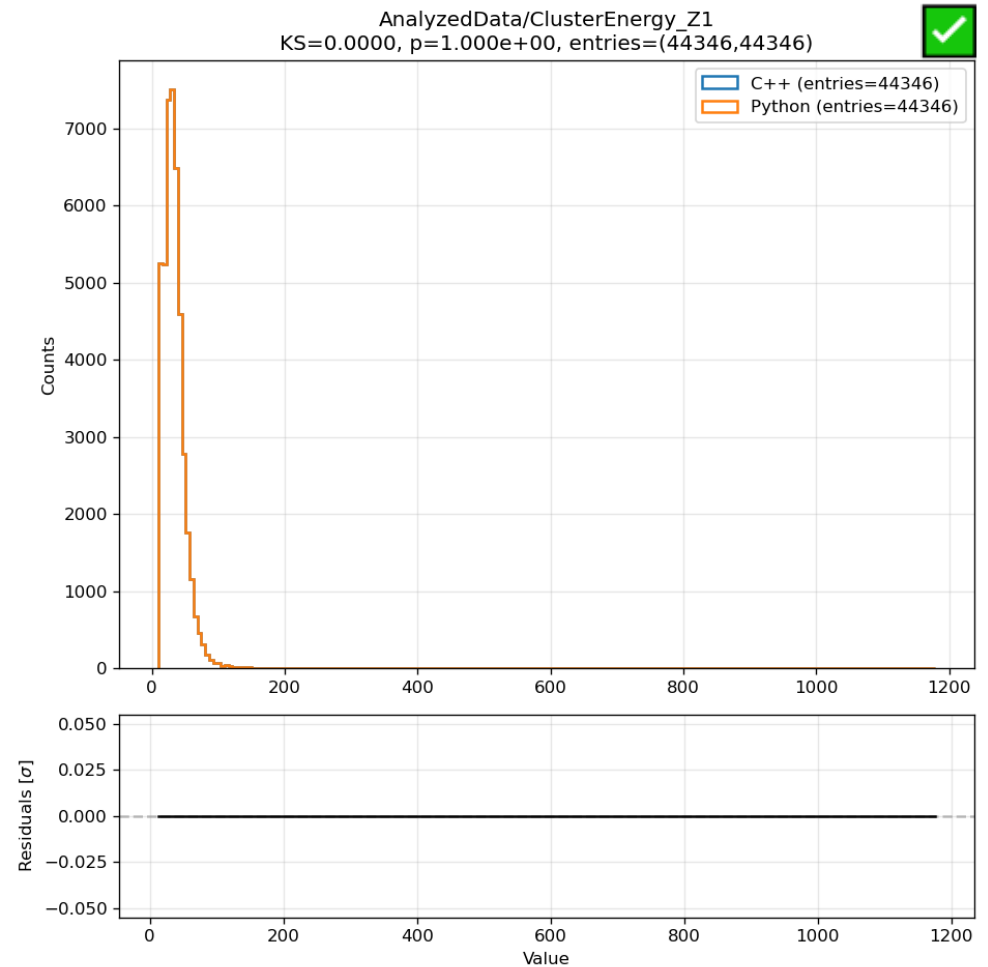
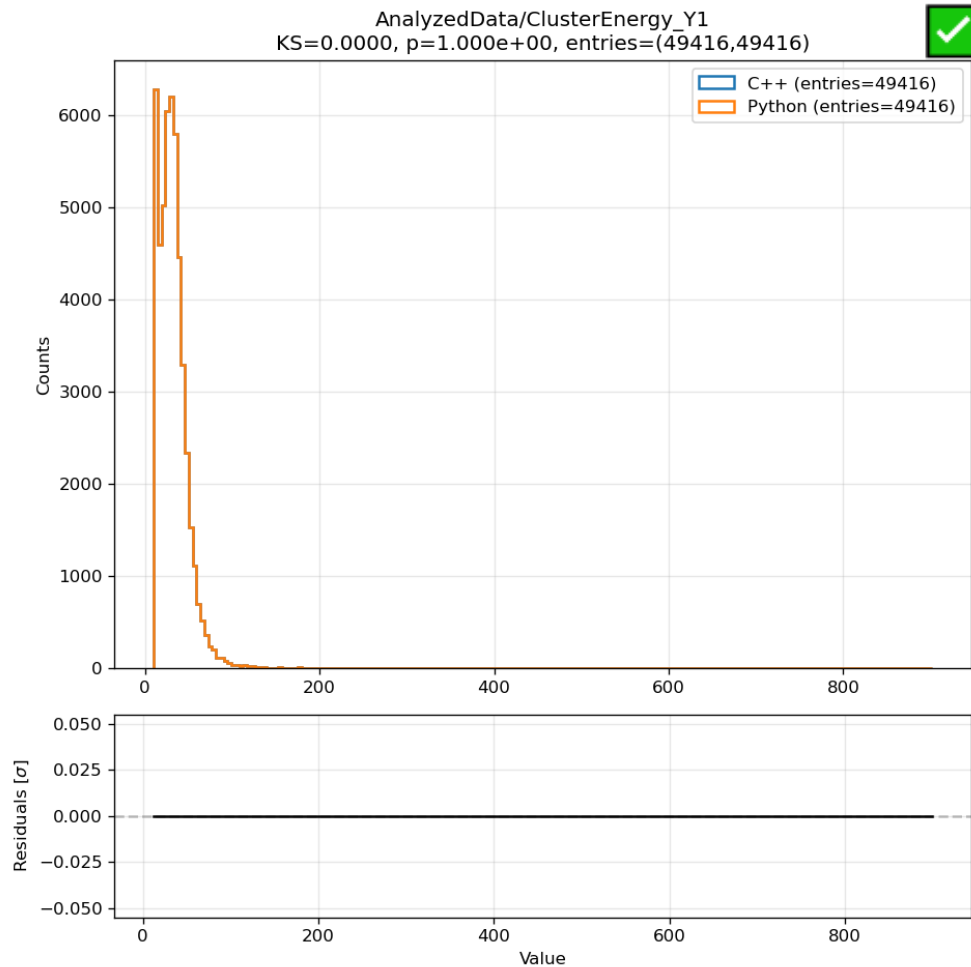
Sanity check: Trigger variables



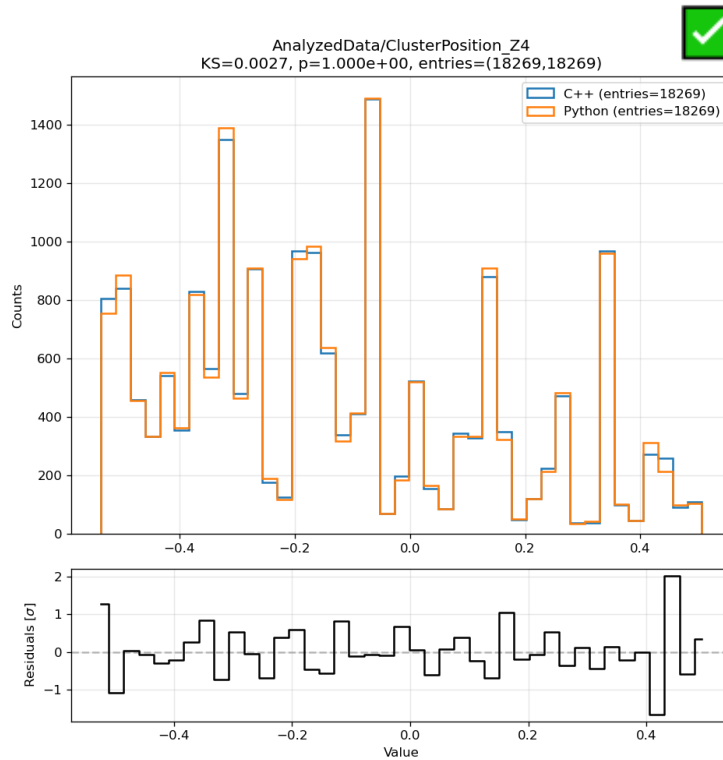
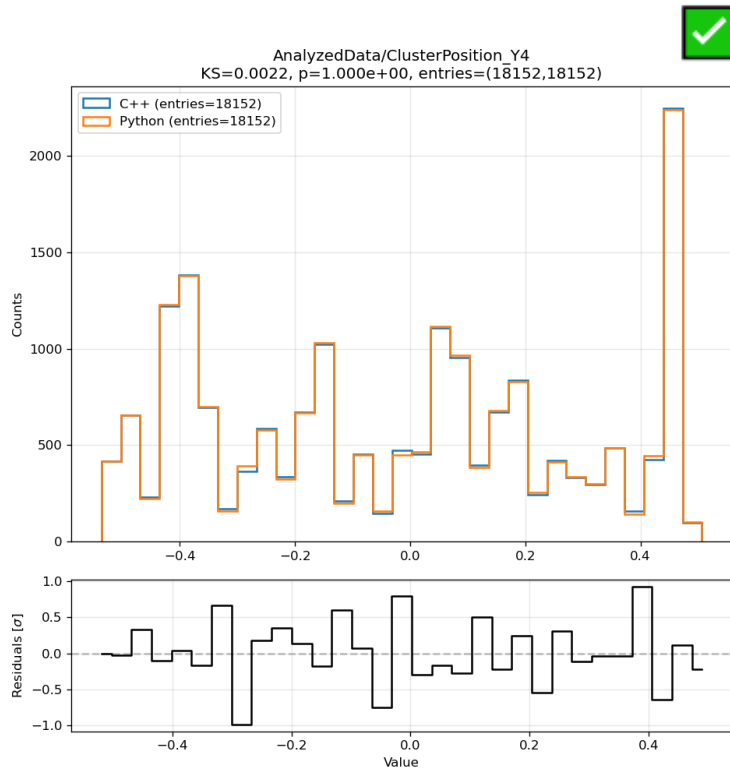
Sanity check: Strips energy and position



Sanity check: Clusters energy



Sanity check: Clusters position



Reason of non-zero residuals:

- random smear applied to 1-strip clusters.

If a cluster has size 1,
a random offset is added
with step 0.001 and range -0.008 to +0.008.
In Python this comes from *SystemRandom*.
In C++ this comes from *random_device*.

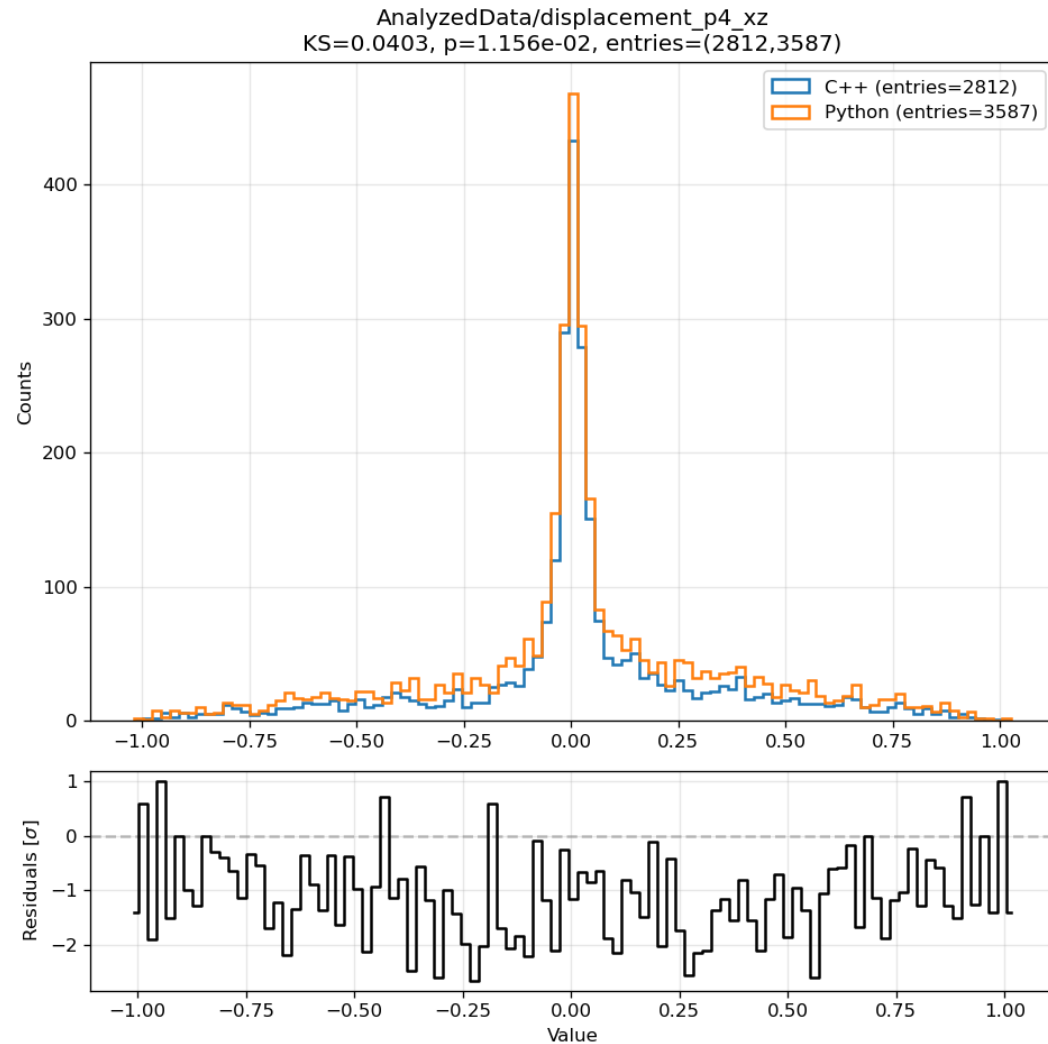
**Histograms here are produced
without fixed seed.**

If necessary, this can be redone.

Variables with **significant differences:**

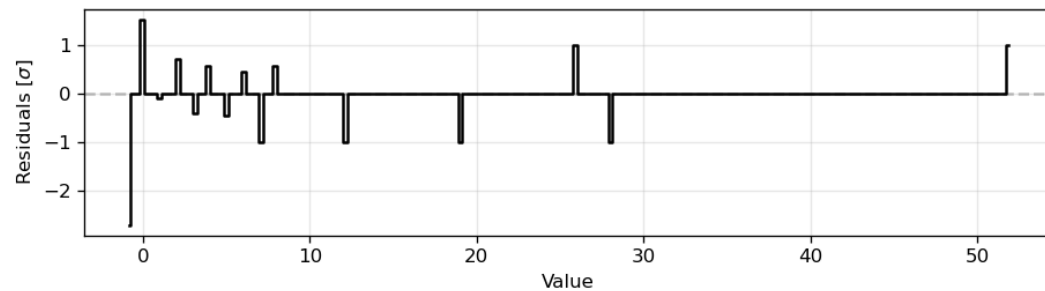
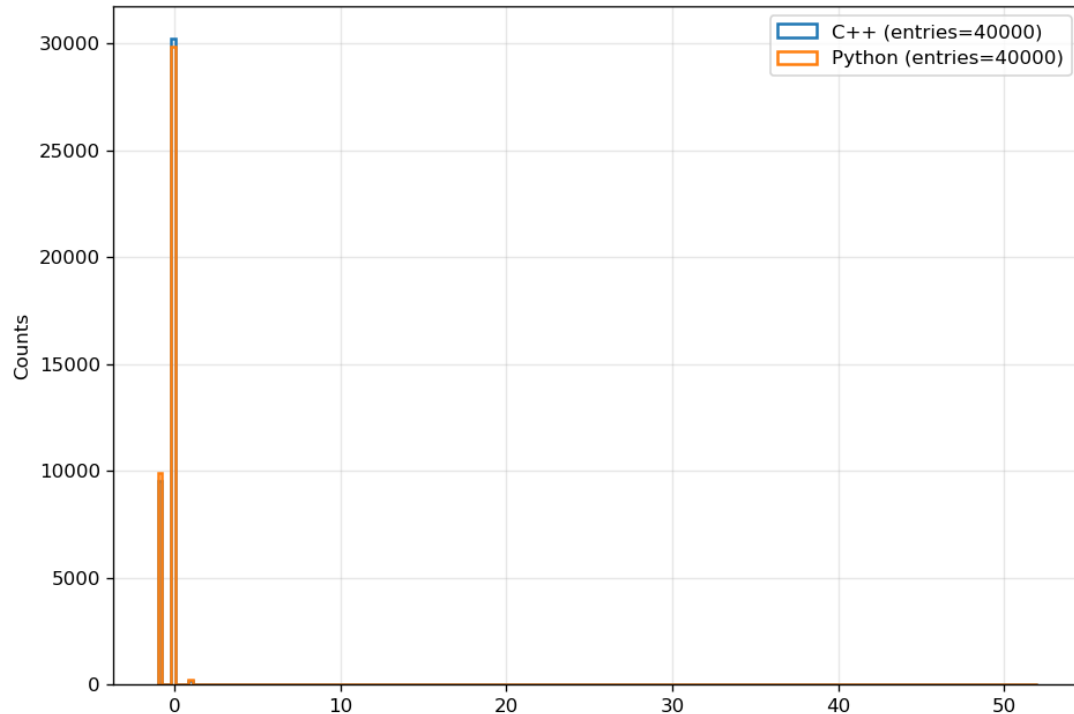
displacement_p4_xz

Only variable post-fitting
showing discrepancies

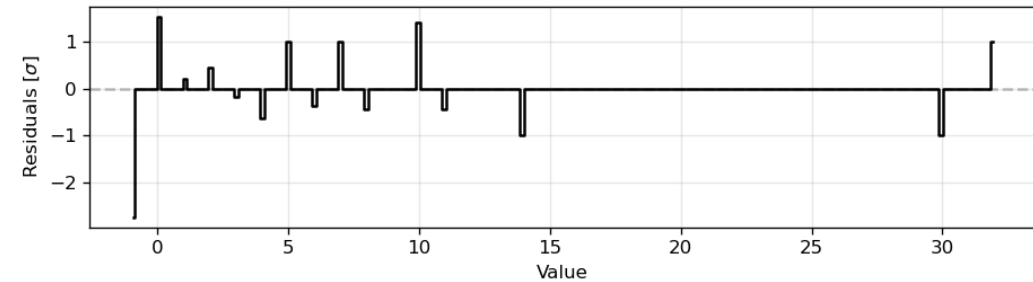
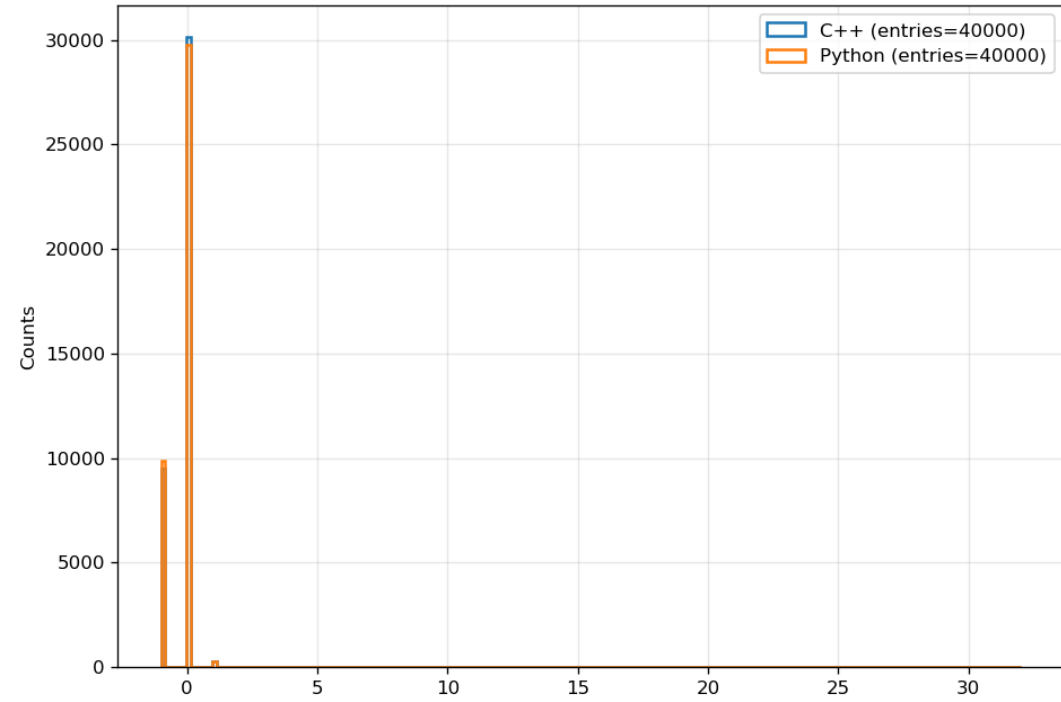


Variables close to significant differences:

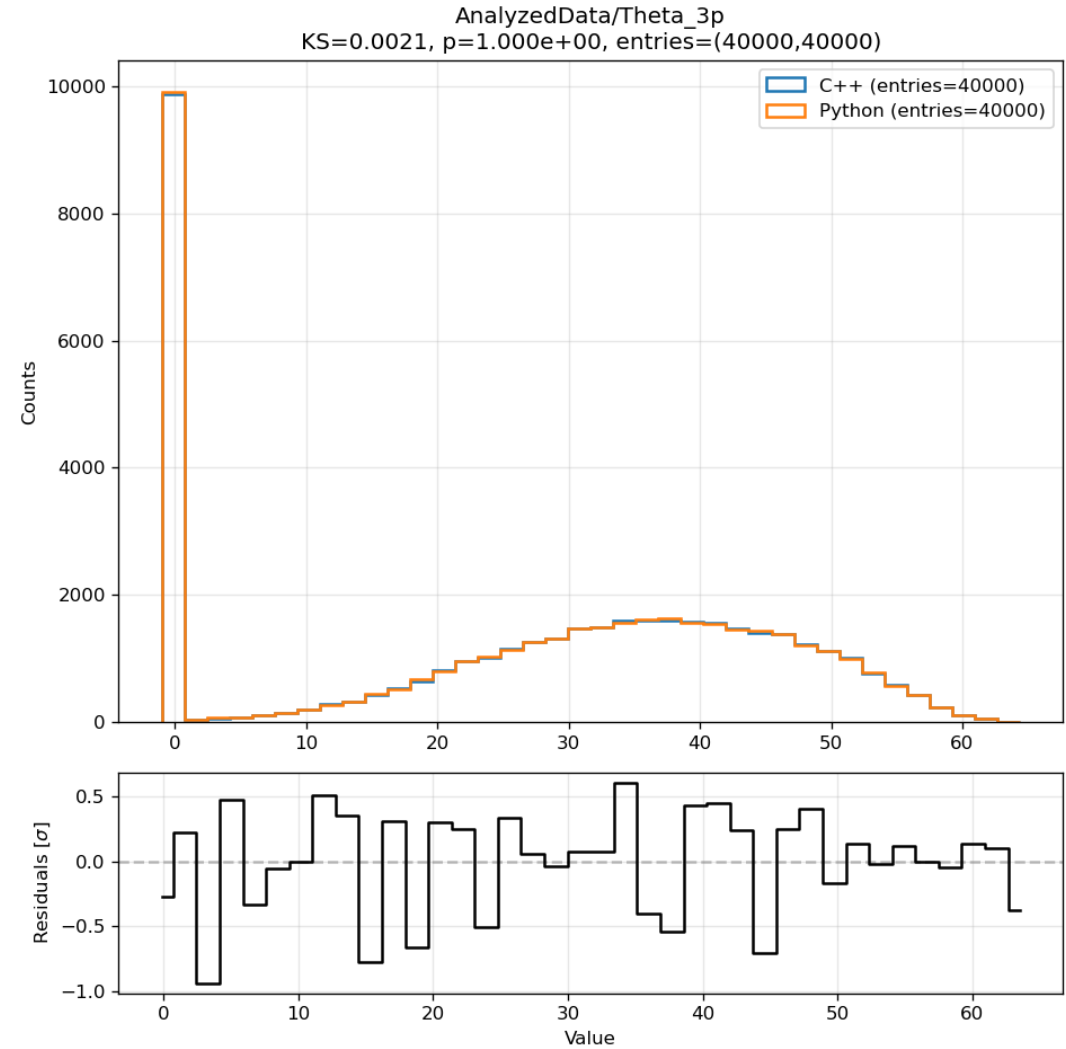
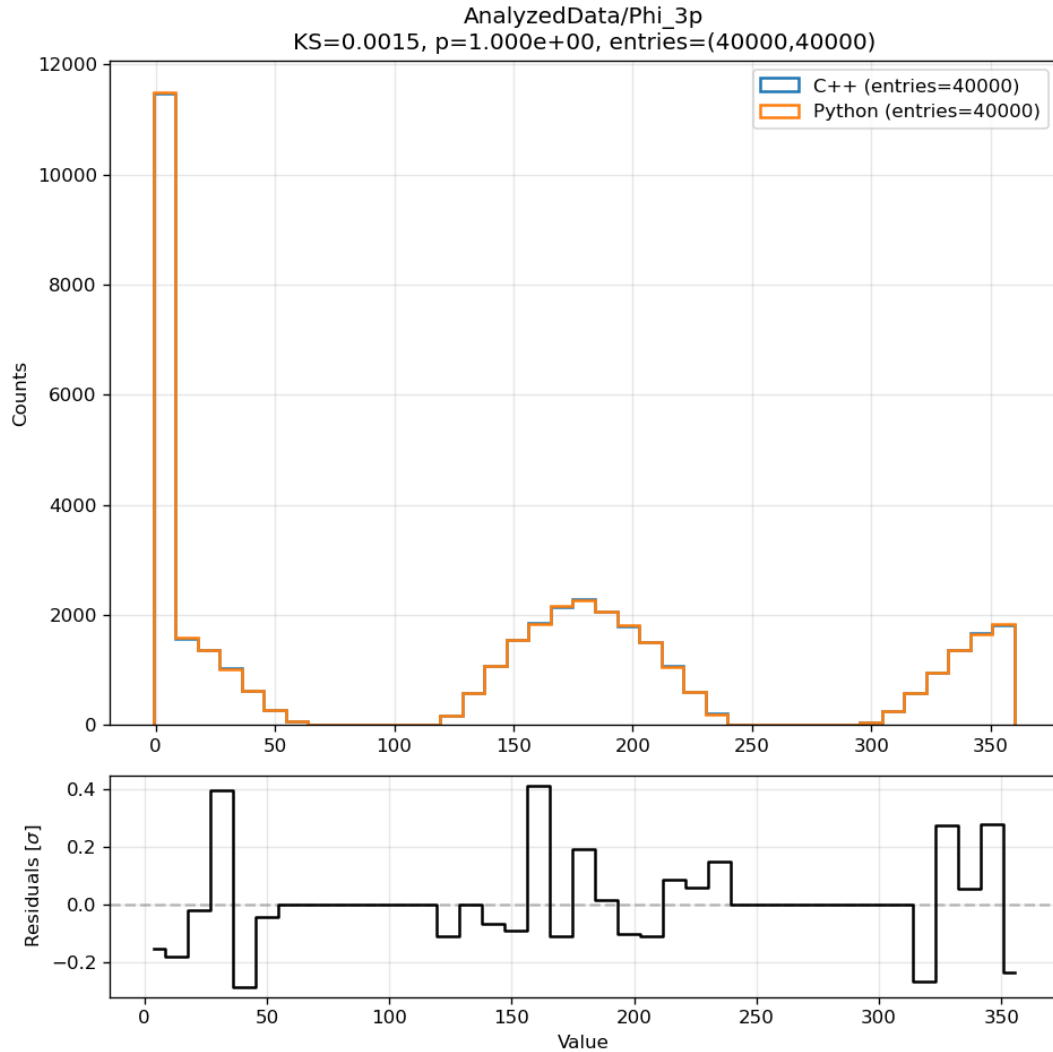
AnalyzedData/BestTrack_3p_xy_index
KS=0.0095, p=5.584e-02, entries=(40000,40000)



AnalyzedData/BestTrack_3p_xz_index
KS=0.0096, p=5.175e-02, entries=(40000,40000)

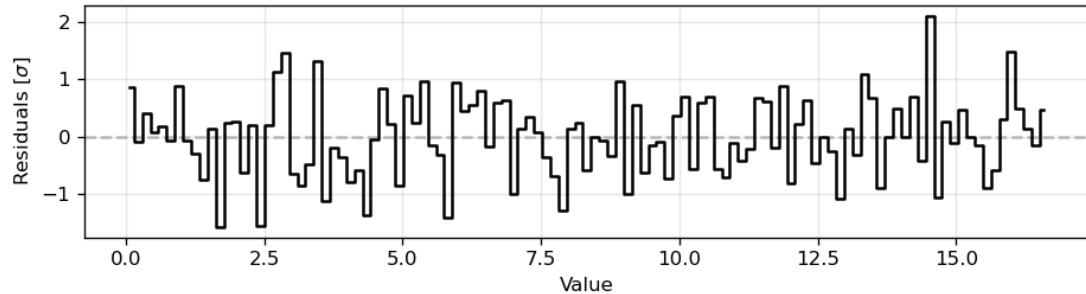
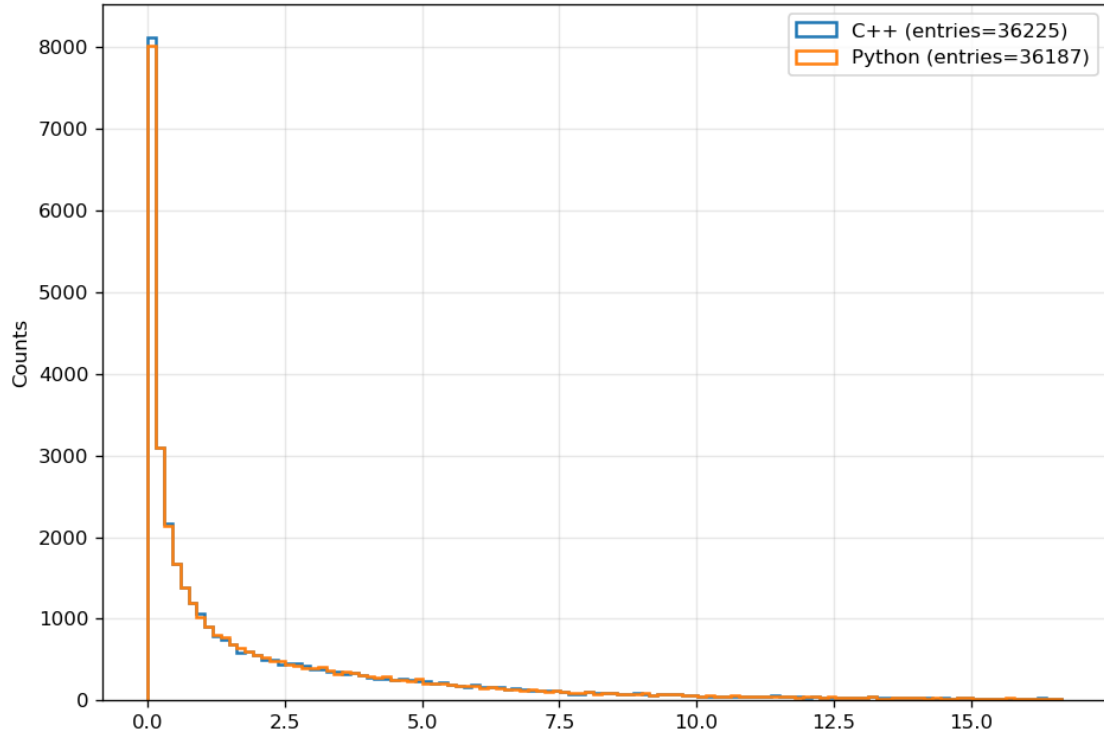


Example of good KS test: angles

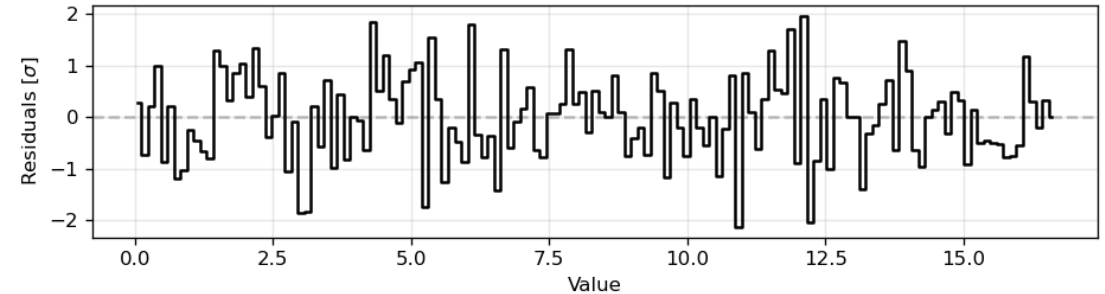
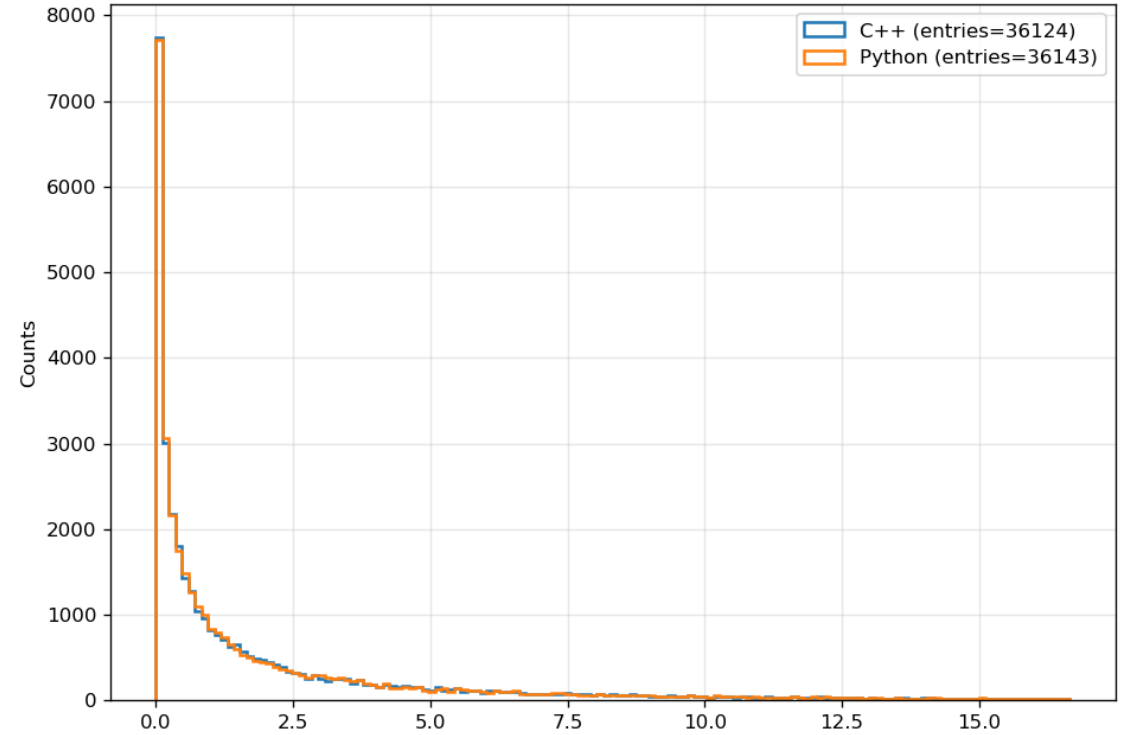


Example of good KS test: chi2, 3plane

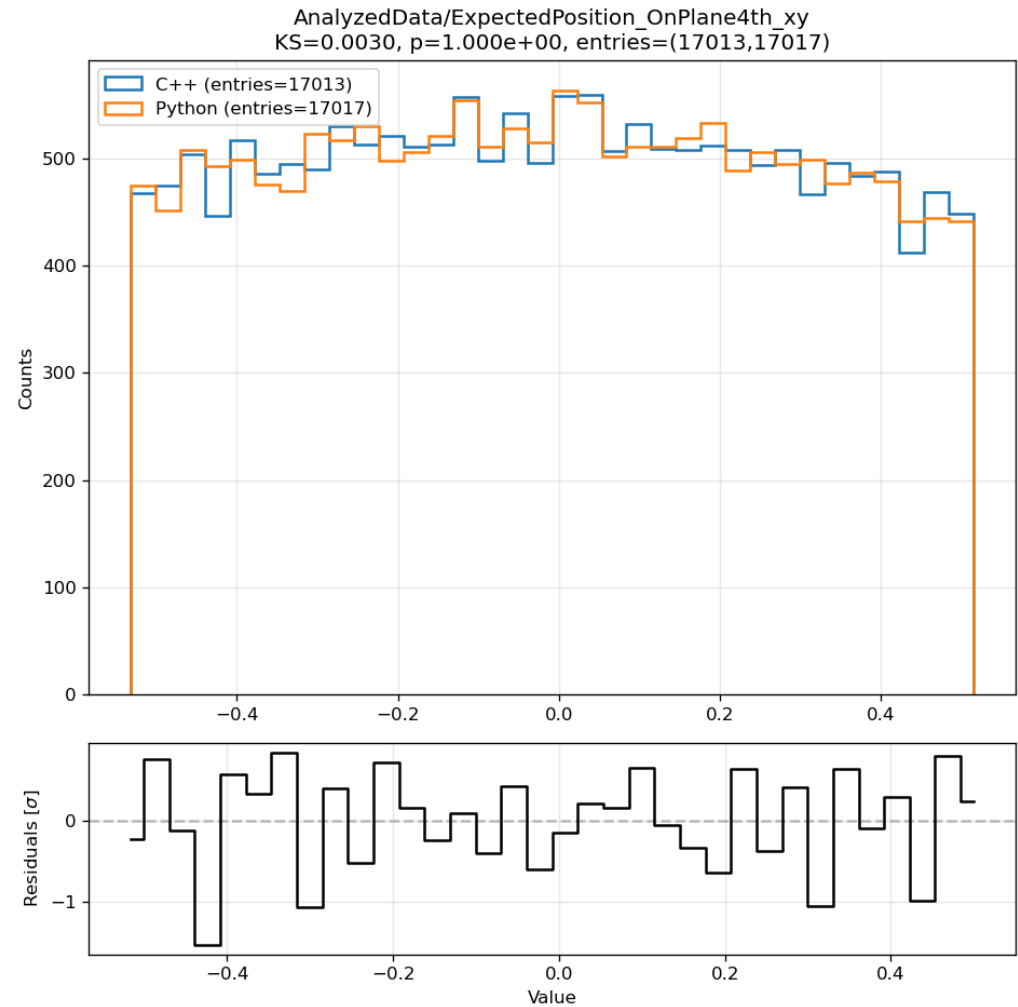
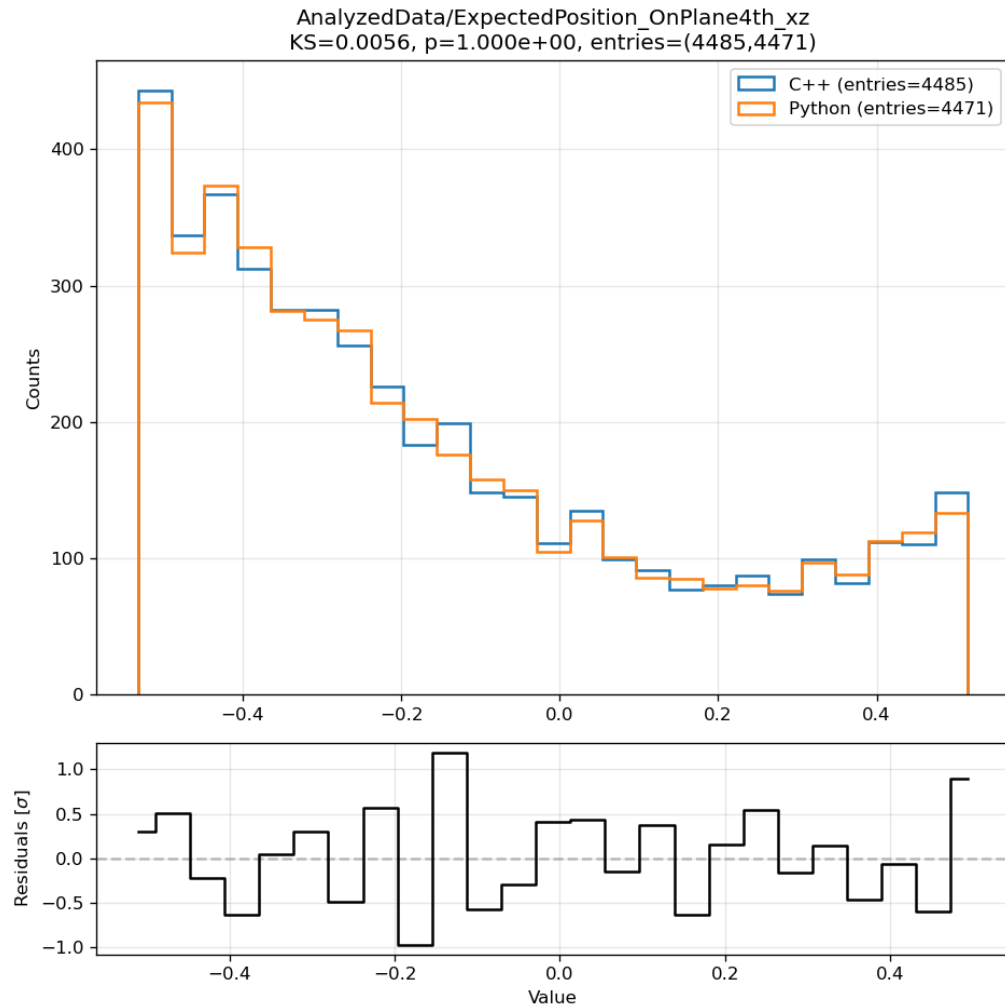
AnalyzedData/chiSquare_3p_xz
KS=0.0049, p=7.835e-01, entries=(36225,36187)



AnalyzedData/chiSquare_3p_xy
KS=0.0046, p=8.369e-01, entries=(36124,36143)



Example of **good KS test**: expected pos on 4th p



Conclusions and next steps

- **Bug:**
 - In ClusterLists.cc script, the variable ClusterStripsID was not cleared like the others. This leads to a carry-over of stripsID between clusters.
 - After adding the line `clusterStripsID.clear();`, the variables StripsID occupancy now match perfectly with the corresponding pythonic version.
 - In the Python version, the lists are reinitialized at each iteration over clusters, so this issue cannot occur by construction.
- **Ongoing tests:** Run Principal Component Analysis (PCA) on output variables: a dimensionality reduction technique used to simplify complex, high-dimensional datasets.
- **Next steps:** homogenise nomenclature of branches. For now, kept as the original to facilitate comparisons.

Backup

- Full table comparison table available here:
`/user/abiolchi/Software/dev/compare_run2546_preStripIDbugfix/report/compare_run2546_StripIDbugfix.csv`
- It will be part of GitHub repository