Reviewer #1: The authors presented a revised version of their original manuscript in which they went through the suggestions provided in the first review.
In the revised version the authors introduced a new method for filtration of POCA points, as the original method was found not performing well on realistic simulations taking into account the finite spatial resolution of their tomograph detector. They presented results with the new algorithm. Although we certainly appreciate the efforts put by the authors in following the comments given in the first review and in adding details and preliminary data from their detector,
There are still some aspects requiring clarification from the authors and a second round of revision. This second revision should mainly focus on clarifying the two thresholding methods, which are still not clearly presented in the text.

== METHOD 1 - CHAUVENET ==
- The description of the first method (Chauvenet) is still misleading in some parts. From the text we understood (with a certain difficulty) the following but we need confirmation from the authors. Threshold is computed from the histogram of voxel POCA counts. The bulk of the distribution is centred on small counts (false positives) while outliers (larger counts) is the signal (corresponding to large peaks in Fig. 14 & 15). The mean and standard deviation of this histogram is computed. The threshold is set at the first counts larger than (mean + 2 sigma). All voxels with counts below this threshold are rejected all at once as false positives. Is our understanding correct? If so, however, we do not understand the text saying "This condition makes sure that all the local maximas that lie even within the 2 sigma range but [...] will also get removed". This sentence (together with the green lines shown in Fig. 14 & 15) seems to suggest that the 2 sigma confidence range is coming from the histogram of number of voxels (mean and sigma of histograms shown in Fig. 14 & 15) and not from the histogram of POCA counts as written in the text. If this is the case the method is flawed and cannot handle detection of threats (one or more) placed at different positions in the inspected volume. If however the method is based on the POCA voxel counts, please remove the misleading text and green lines from the plots.

A. Thanks a lot for correcting us. As you had already correctly pointed out that the histograms show in figure 14 & 15 corresponds to the histogram of Voxels. Corrected the same in the text. Sorry for the confusion, i think i was not able to correctly convey the idea through the text mentioned in the paper. I will certainly expand the text to clearly convey the idea.

The detail of the histograms presented in Fig. 14 & 15 are as follows. Assume, that we divide the volume to be inspected in “Nx” equal parts along X axis, “Ny” equal parts along Y axis and “Nz” equal parts along Z axis, ie. the inspected volume is divided into “Nz” number of 2D slices, each of this 2D slice is having “Nx X Ny” number of voxels. Each of these voxels in 3D is assigned a unique index starting from index 0. These Voxel index is placed along the X axis of histogram, ie. the bins number in histogram corresponds to index of voxel in the voxelized region. Count in each bin corresponds to the PoCA point count in that voxel. The histogram show the peak for the voxel indices where the material is placed. If the PoCA point count in a voxel is less than the selected threshold then the voxel will be considered as noisy voxel and the containing PoCA point will be removed from the reconstructed image. This method works pretty well with exact hit point data. We have tested the method on various scenarios by placing the multiple object at different position of the inspected volume. In the result section also we have presented the results for multiple object placed at different locations. The corresponding text regarding histogram name is modified accordingly to avoid any confusion.

- The authors say also that repeated application of the Chauvenet criterion is not recommended, but they actually excluded a large portion of false positive voxels all at once. Our impression is that the used method is not the Chauvenet criterion (which would lead to exclude one or few outliers) but rather a thresholding method. Suggest to rename the method in a different way (e.g. "POCA counts thresholding method" as opposed to "POCA weighted counts thresholding method") and remove any reference to Chauvenet criterion in the text.

A. Yes we are trying to excluding as large portion of false positive voxel as possible at once. As suggested the method is renamed as “Thresholding based on PoCA count”, and the references to the chauvenet’s are removed. Thanks a lot for the valuable suggestion.

== METHOD 2 - WEIGHTED COUNTS ==
It is not clear from the text and from the pseudo code how the second method (standard deviation weighting) actually works. The weighted counts WC in a given voxel is computed as C x SD, where C is the number of POCA points in that voxel and SD the standard deviation of the scattering angles of POCA points. Voxels with a negative WC are considered as spurious and filtered. How can this product be negative? SD is positive by definition of standard deviation, while C is >=0 (=0 if there are no POCA points in voxel but in this case there are no false detections) so WC>=0. We are probably misunderstanding the method. Could the authors clarify this and eventually improve the description of the method?

A. WC is alway positive ie. WC >=0. But it can also be a real number between 0 and 1. After calculation the WC is converted to integer value, and if this integer value is greater than 0 then the corresponding Voxel is signal voxel (containing true positives) otherwise its considered as voxel containing false positives and is removed from list of reconstructed PoCA points

== EXPERIMENTAL DATA ==
We appreciated that the authors are carrying on their work with the muon detector prototype presenting some preliminary data in the paper. The authors say at pag. 4 that "The variation of counts in pixels is due to the geometric acceptance of the detector.". This does not seem to be the case in Fig. 6 (c) and (d). As the figure presents coincidence hits we would have expected 2d distributions similar to Fig. 10. From Fig. 6 (a) and (b) it seems that the variation of counts in pixels is rather due to detector strip inefficiencies or noise (e.g. strips firing on noise and not on signal)? As the data are still preliminary this goes beyond the scope of the paper. Suggest the authors to modify the sentence as "The variation of counts in pixels, expected to follow the geometrical acceptance of the detector, is distorted by strip inefficiencies and noise."

A. Changes incorporated as suggested

== CONTENT CORRECTNESS ==
Some sentences seem conceptually wrong (or unclear):

- Pag. 7: "These clustering and outlier detection algorithm are supervised and needs some input information like [...]". Clustering algorithms (like K-means or DBSCAN) are actually unsupervised algorithms. By supervised algorithms we usually mean those algorithms requiring some training data as input (e.g. neural networks, etc) and a target to learn, not those algorithms requiring some input parameters. Clustering algorithms require some input parameters but this does not mean they are supervised. They are unsupervised because they do not require training data to be learnt to extract clusters. Please correct the text, for example saying that "unsupervised algorithms, like DBSCAN or k-Means, where considered as possible filtration method but not deeply investigated in this work as requiring a certain number of parameters to be tuned. The method proposed in this work is instead parameter-free.".   :

A. Changes Implemented

- Pag. 10-11: We are not sure if the term "algorithm overfitting" is actually correct to denote the rejection of some positives along the false positives after applying the threshold. We usually denote as overfitting those situations in which a statistical model is lacking generalization capabilities with respect to a new sample of data to be described, e.g. the model is able to accurately describe the fitted data but poorly describing new data. A well-known example of overfitting occur when you train a neural network too much on a given training sample. One gets high modeling performances on this sample but poor performance on a new sample (test or validation sample. Please remove the misleading text.

A. Changes incorporated, section heading “Analysis of degree of overfitting” is changed to more appropriate term “Analysis of the degree of True Positives Reduction and selection of optimal voxel size” . The text of the section also modified to reflect the same.

== FIGURE & TABLES ==
- Please improve the captions of some figures, so that the reader does not need to go through the text to know some details. For example in Fig. 18 & 19 add that the first figure is referring to ideal simulated data, while the second figure is referring to smeared simulated data. Add what is the distance between planes used in Fig. 18.

A. Figure 18 modified to show TPDR and FPDR plots using ideal simulated data for the setup with vertical distance of 30 cm. The corresponding caption also modified to reflect the same.

- Consider to present the results shown in Table 1 and 2 directly in Fig. 18 & 19 (eventually removing the tables), so that in the same plot one can see both the TPDR and FPDR (or 1-FPDR for better graphical visibility) and the optimal voxel size for both RPC distances. Is it possible to add to both plots the TPDR and FPDR obtained without using any threshold method. In this way one directly compares the benefit of using the threshold methods.

A. Figure 19 modified to show the information presented in Table 1 and Table 2. It now shows different plots of TPDR, FPDR and True positives reduction to select the optimal voxel size. Table 1 and Table 2 are removed.

== ENGLISH STYLE ==
The english/scientific style is still to be improved throughout the paper. Some representative examples of unclear text or text not sounding as correct english:
- Pag. 2; "The setup consists of a cosmic Hodoscope, which is a setup used to" DONE
- Pag. 7: "This implies that by using some good algorithm if the envelope of false-positive POCA points can be removed ..." : DONE
- Pag. 8/9 (+ in other sections of the paper): "... that POCA ..."
- Pag. 7: "Its better than K-means because it does not ask the user to supply ...". ask-->require? DONE
- Pag. 9: "get rid false positives" --> "get rid of false positives"? DONE
- Pag. 9: "With the detector of this much resolution, the smearance ..." DONE
- Pag. 3: "10000 reconstruction events" --> "10000 reconstructed events" DONE

Editor's comment:

(1) The language needs major improvements

(2) NIM-A is printed b/w, please put (colour online) into the caption of Fig. 10