Reviewer #1: The authors presented a method to improve muon tomography reconstruction imaging with the POCA algorithm. We report below a general review of the paper against NIM A publication policies and major/minor questions/suggestions emerged during the review:

- Relevance to the readers of NIM A
The content of the paper is suitable for NIM A, given the number of works focusing on muon tomography an relative applications already published in the journal.

- Originality, quality, and correctness of the research
In this context, however, the submitted paper, at least in its current form, does not seem to introduce relevant improvements in the muon tomography domain both from an algorithmic/data analysis point of view as well as from an experimental view (e.g. innovative detectors or experimental layout).
For example, the improvements proposed for the POCA (a filter based on Chauvenet's criterion to voxel POCA data) can be more regarded as a pre-processing step rather than a novel tomography reconstruction method.

From the experimental point of view, the authors mention in Section 2 that an RPC based detector is being developed for tomographic applications but, as reported in the Summary section, the design is still ongoing and no preliminary data was reported for the prototype (2 xy planes).

A. Preliminary data from scintillators, figure 2 in section 2, showing the coincidence pattern of scintillators in top and bottom scintillator planes is added. Data from two assembled RPC detectors is also reported in the paper in figure 6 in section 2.3, where strip profile of all the 64 strips of detector as well as 3D hit plot, which show the hit count of each pixel of detector are shown.

Further, the target requirements and application context for the detector in development was not discussed. This makes hard to evaluate how the planned detector compares to existing gas detectors used for tomography.

A. The prototype tomography setup is developed for the discrimination and identification of high Z material like Pb, U, from medium Z and low Z material like Fe and Al. Scope of application is mentioned in the introduction part of the paper.

- Quality of the presentation, e.g., structure, conciseness, understandability, quality of language and figures
The organization of the paper has to be improved in these areas:

Structure: some sections are too verbose (e.g. reporting concepts that are widely known to the audience of the journal) and can be either safely removed or moved to a different part of the paper. Almost all sections briefly discuss at the end what is going to appear soon on the next section. Consider to present a summary of content at the end of the Introduction section to avoid repetitions. Some other section are, on the other hand, not detailed enough, e.g. parameters used in the simulation, filtering method (see major comments below), etc.

A. Repeated and too verbose text is removed. The summary is presented at the end of the introduction section, which shows the organisation of different sections of the paper.

Style: english and scientific style has to be definitely improved throughout the paper
A. English and scientific style is checked and improved throughout the paper.

- Compatibility of text and figures with black and white printing
No particular issue related to color figure. Additional details to be reported in figure captions (e.g. Fig. 5 and Figures presenting the results)

A. Detailed captions are added to the figures.

- Is the relevant literature quoted?
The relevant literature of muon tomography projects, application and simulation works was not extensively cited in the introduction section.

A. Literature is now extensively cited in the introduction, as well as, in the other parts of the paper, wherever required. Also the references are now appended in the reference list.

In its current form we do not recommend the paper for publication according to the evaluation metrics discussed above. The authors may, however, take into consideration the suggestions given below to improve the manuscript and produce a major revision of the paper.

== Major comments/suggestions ==

We recognize the authors' effort to build a real prototype (4 xy planes) to validate their method and we encourage them to carry on their work for the future, as they say in the Summary section. Although complementing the results with experimental data from the prototype would certainly increase the quality of the paper, we agree with the authors that the paper can be self consistent only with GEANT4 simulated data. However, in its current form, we believe the paper requires additional details in the text to be provided and further analysis to be carried out with available simulated data to test the robustness of the method and account for several aspects that were neglected (e.g. finite resolution, ...). We provide here a list of questions and suggestions the authors may want to go through to produce a revised version of the paper:

- Please add in the introduction the context and scope in which software and experimental test setup is developed. For example is the prototype aimed to perform tomography of container cargos or developed for different applications? What are the target technical requirements for the application (e.g. spatial resolution needed, scanning times, etc)?

A. The prototype setup is aimed for disciminations and identification of high-Z material from medium-Z and low-Z material and to do the image reconstruction in the lab. So scanning time is not the very critical factor. In the paper the main focuss is on the filtration algorithm for clear image reconstruction, hence spatial resolution is an important factor. Results are shown using simulated data with the finite pixel resolution of 3cm X 3cm. Algorithms are proposed to do the image reconstruction with this much coarse resolution.

Please extend the list of references. Below we report, as an example, a possible list of relevant works in this area we are aware of.

Pioneer works

E.P. George, Commonwealth Engineer (1955) 455.
L.W. Alvarez et al., Science 167 (1970) 832.
K. Nagamine, et al., Nucl. Instr. and Meth. A 356 (1995) 585.
K.N. Borozdin et al., Nature 422 (2003) 277.
L.J. Schultz, et al., Nucl. Instr. and Meth. A 519 (2004) 687.
L.J. Schultz, et al., IEEE Trans. Image Process. 16 (2007) 1985

Muon Tomography Algorithms

S. Riggi et al., NIM A 728 (2013), 59-68
D. Poulson et al., NIM A 842 (2017), 48-53
M. Benettoni et al., JINST 2013 8P12007
S.Chen et al., JINST 9 (2014) no. 10, C10022

Plastic scintillator/fibre detectors for muon tomography

F. Riggi et al., NIM A 912, 21 (2018) 16-19
A. Clarkson et al, NIM A 745 (2014) 138-149
V. Anghel et al, NIM A 798 (2015) 12-23

Gas detectors/RPC for muon tomography

S. Pesente et al., NIM A 604 (2009), 738-746
P. Baesso et al., 2013 JINST 8 P08006
X. Wang et al, NIM A 784 (2015) 390-393

GEM Detectors for muon tomography

K. Gnanvo, et al., Proc. IEEE NSSMIC 2010, NI19-90, pp. 552-559, arXiv:1011.3231

A. List of references is exteded with the suggested and required references, and citation is done in the required sections

- The method is validated with simulated data that assume perfect spatial resolution. However we do know the finite resolution (not quoted in the paper, is it ~3 cm/sqrt(12)?) is affecting tomography imaging considerably, particularly if based on POCA method. We strongly suggest to include in the revised version the impact of resolution on the results. A possibility could be to use the same simulated GEANT4 data smeared with the same spatial resolution of the experimental setup.

A. Yes, the finite resolution is ~3 cm/sqrt(12), now also added in the paper in section 2.1. This resolution is certainly affecting the imaging with PoCA because it introduces a lot of false positives. The paper shows results using perfect spatial resolution where Chauvenet's based algorithm works pretty well. But with smeared data, Chauvenet's based algorithm doesn't works well, which is describe in detail in the section 5.2.3 . To handle this, a Standard deviation weighting factor based filtration algorithm is added in the paper in section 5.2.3, which show good results for filtering out the false positive and give a coarsely reconstructed image with smeared data.

Q.- Not fully clear how the muon track is reconstructed in the simulation. Is it simply assumed equal to the straight line connecting true hits (both for incoming and outgoing tracks) or is there any tracking algorithm run on hits (3d straight line fitting, Kalman-filter, etc)? Please specify particularly for the case of simulations with finite resolution (see comment above). Tracking is another effect that should be taken into account as affecting POCA estimate.

A. In simulations, we have worked with exact hit data, as well as with smeared data. For the case of exact hit points, the track is simply the straight line connecting the exact hits. When working with smeared data, the track is basically the 3D straight line fit for the center of hitted pixel. Section 3.2 explains the track reconstruction algorithm in details.

Q. - The object detection method based on POCA is not described in Section 4.1. What criterion is used to actually form a tomographic image and detect an object? For example some authors in the literature uses average number of POCA events in voxel (weighted by powers of the muon scattering angle, e.g. theta^2) and apply a threshold on it (e.g. n x sigma significance) to find whether an object is detected or not.

A. The object detection method using Chauvenet's criteria is based on the number of PoCA points that lie within the Voxel. These are considered as signal points and forms the part of the image, if the PoCA point count is more than the threshold selected using Chauvenet's criteria. This algorithm works pretty well with exact hit point data, because the proportion of false positives is less while considering exact hit point. For smeared data, the algorithm does not works well because of exponential increase in false positives. To handle that, a Standard deviation weighting factor based filtration algorithm is proposed in section 5.2.3,

Q. - Section 4.2.2: It is not fully clear from the text and from Fig. 11 and 12 if the threshold is computed assuming knowledge of the object position and/or signal size. It is said in fact that "the threshold value is calculated by considering only those voxels whose POCA point count lie within 2 sigma range as the signal. [...]". If this is the case how the threshold is computed when there is no knowledge of the object position nor signal size (as we expect in the reality)? Or, alternatively, is the threshold dynamically determined through processing of voxel histograms (e.g. find signal/background size through some peak finder or clustering algorithm)?

A. Threshold value is calculated without any assumption of object position. Infact, after the application of threshold, the object position can be determined, based on the voxel numbers which are declared as signal voxels. The threshold value is calculated based on the processing of Voxels. The count of PoCA points in the voxel where material is placed gives the higher peaks at the corresponding voxel number as shown in the fig 14 & 15. The selected threshold value is the maximum of PoCA point counts for the voxels, which lie outside two-sigma range.

Q. Does the threshold set depend on exposure time/number of muon events? In other words, is the threshold dynamically/adaptively computed during the volume scanning time? Or is it fixed?

A. Yes Threshold value depends on exposure time/number of muon events, and is computed dynamically. For different exposure time we get different threshold value.

Q. Does the threshold vary from a scenario to another (for example from the scenario with only lead objects in Fig. 1 to the one with only iron objects in Fig. 19)?

A. Yes, it vary from one scenario to another. For high-Z objects, we get a dense cluster of PoCA point in the voxel where the material is placed, as compared to the voxels where low-Z or medium-Z material is placed

Q. We also expect that POCA counts somehow depends on the detector geometrical acceptance (e.g. distance among planes and inspected volume), e.g. for a given scanning time POCA counts should not be equal for two signal voxels located at the volume center or at the volume upper/bottom (close to the detection plane).

A. Yes, PoCA point count depends on detector geometrical acceptance. Section 4 shows the study of detector geometric acceptance, where the variation of acceptance is shown at three different Z position, which clearly indicates that how the point count varies as we move the scatterer from upper plane towards lower planes. The variation of acceptance along X and Y can also be seen on the individual slices.

Q. Should the threshold also account for this aspect? Indeed we want the algorithm to work on real or simulated data with complete ignorance of the object location and material at given scanning time. This is the goal of tomography reconstruction. As the thresholding method is the main subject of the paper it is quite important that all these aspects are clarified and deeply analyzed.

A. Yes thresholding algorithm is taking care of this aspect. The detector acceptance will only affect the counts of PoCA points in the voxels, but the PoCA points counts for the voxel where object is place will always results in a cluster formation and will show higher peak for the bins corresponding to that voxel number as shown in figure 14& 15, and hence distinguish it from the voxels where scatterer is not placed. Its completely independent of object location, and the thresholding is applied based on the counts in the voxel, and not the object position, Once we remove the PoCA point from the voxels whose count is less than the threshold, the remaining voxel give the location of scatterer under test.

- It would be useful to report in Section 4.2.2 a Figure (before Fig. 11) showing POCA voxel counts distribution to show the population normality assumed in the Chauvenet criterion.

A. Figure 14 & 15 is showing the same thing. i.e. it shows the distribution of PoCA points in Voxels numbers as the bins.

- The results reported in Section 6 does not discuss any metric (e.g. significance above background) used to detect the object (see also comment above). In fact, it would be useful to show in the paper the true/false object detection rate for raw and filtered method and discuss the impact of filtering on the object detection time and in the false detections. Indeed we expect minimum scanning time to detect an object will increase (due to events rejected in the filtering step) when compared to raw POCA. On the other side false detection will decrease due to filtering.

A. TPDR plots 18 & 19, show the variation of true positive detection rate. The first data point represent TPDR for raw PoCA points, whereas the other data points shows how the TPDR improves after application of filtration algorithm as the voxel volume decreases. Section 6.3 discuss another important parameter that will help in chosing the optimum voxel volume to keep a balance in the degree of filtration and reconstructed image quality and time. It also explains how the uncontrolled filtration may result in distorted final image and increase image reconstruction time as a result of overfitting

- TPR and FPR discussed in Section 5.2 seem defined in a different way with respect to the conventional definitions for TPR and FPR in image processing (e.g. FPR is not equal to 1-TPR as in the paper). To avoid misunderstanding we suggest the authors to use a different name or clearly state the difference.

A. To avoid any confusion with the conventional definitions for TPR and FPR in image processing, TPR is now changed to TPDR (True Positive Detection Ratio) and FPR to FPDR (False Positive Detection Ratio). Fig 18 & 19 shows the true and false positive detection ratio as the function of voxel volume. The first data point shows the TPDR for raw PoCA point, and the successive data points show the improvement of TPDR as we decrease the voxel size.

Q- Not clear in section 5.2 why the TPR improves when we decrease the voxel size (Fig. 14) and why this is important for the method. I would rather expect the opposite, e.g. you get almost all POCA events falling in the target voxel (TPR increases up to 1) if you increase the size of the voxel until completely enclosing the object by a large amount. Imagine to have a huge voxel size compared to object size. Then, all POCA will fall in this and you will get almost 100% TPR according to the TPR definition given. It is possible that we misunderstood text here, so please discuss this aspect.

A. TPDR is defined as the ratio of True Positive PoCA Points (the PoCA point which really lie inside scatterer region, and can be easily found in simulation) and Total number of PoCA Points. Here the target is to increase the TPDR by reducing the number of false positives. If we take bigger voxel then it contains not only true positives but also the false positives i.e. the points which lie outside the scatterer region(that we want to remove), hence results in lower TPDR. When the bigger voxels are divided into smaller one, then the region where scatterer is not placed, contains only false positives, where the PoCA points count will be much less as compare to the counts in the voxels where scatterers are placed. This results in the relatively smaller peaks in the voxels of false positives.

- It would be useful for the reader to see in the paper a detailed analysis of threshold computation (see comment above) and object detection for raw and filtered method as a function of scanning time, target materials/shape/location inside volume, volume material background (air or scratch metal?), providing indications of suitable options for real applications

A. Detailed analysis of threshold computation is presented in the paper for both the exact hit point data and smeared data with detector resolution. This is done by showing the variation of TPDR as a function of voxel volume, and showing its effect on the degree of overfitting as shown in section 6.2 & 6.3. It is also shown in figure 19, that how the TPDR varies by varying the vertical distance between the detectors. Also the effect of increased vertical distance on the degree of overfitting is shown in table 1 & table 2.

Q. Did the authors investigate different methods for outlier rejection (besides Chauvenet's criterion)? It would be useful to provide indications or a discussion of the best one to be used as POCA pre-processor for tomography applications

A. Various clustering and outlier detection algorithm like DBSCAN, Kmeans are investigated as mentioned in section 5.2. All these algorithm are supervised and needs some input from the user, which are not available from the experimental data. For example K-means, one need to specify the number of clusters around which the data points needs to be clustered, and also the location of objects, so that the outliers can be rejected. On the other hand the voxelization based algorithm proposed in the paper does not requires any such information to remove the outlier or false positives

Q. Please report results (e.g. in Fig. 14) as a function of the voxel size in cm^3 (which is independent of volume size) rather than as a function of number of voxels (which depends on the volume size inspected).

A. As suggested the results of variation of TPDR are now reported as a function of variation of voxel volume as shown in figure 18 & 19.

== Minor comments/suggestions ==
- Separate units from value throughout the paper, e.g. in Section 2 180cm-->180 cm.
A. As suggested the units are separated throughout the paper.
- Review text to define acronyms used, e.g. NPD (Nuclear Physics Division)?
- "Simulation" in lower case (e.g. beginning of Section 3 and in other parts)
A. Modified as suggested.
- Many typos found, e.g. Pont of Closest Approach (section 4.1), missing punctuations, missing blank spaces after commas, etc.
- Review english style and correct typos (correct verb per subject, etc)
A. Typos are fixed, english style is corrected and punctuation check is done throughout the paper.
- Use same number of significant digits in numbers and relative errors, e.g. m=... , n=2.025 +- 0.016666 in Section 3.
A. Same number of significant digits are now used.
- Remove Fig. 4 and leave only Fig. 5 with fit results (m, n parameters) shown in the plot. Improve the figure caption with more details :
A. Fig 4 Removed, and the figure captions are modified wherever required, throughout the paper.
- Maximum-Likelihood algorithm is mentioned as an improvement to POCA but not really used in the paper. Consider to remove references to it or, better, mention in the paper that an "improved POCA reconstruction" can improve also likelihood-based algorithms, given that they uses POCA as starting step.
A. The use of filtered PoCA points as a good starting guess for likelihood-based iterative algorithm is mentioned as suggested in section 5.
- What is the background volume material used in the simulation? Air or other materials? Please add in the relevant section (e.g. Section 3.1).
A. Air is used as the background material throughout the simulations. This information is added in the sections, wherever it is required.
- Consider to remove the text at the beginning of section 4. (up to text "for individual muons.") as already discussed before.
A. Text removed as suggested.
- Consider to remove Section 2.3 and include it (in a short and concise version) at the beginning of section 2.
A. Section 2.3 is removed and is merged in concise form in section 2
- Consider to add a Figure near Fig. 1 showing RPC + SCINTILLATOR layout with detector sizes reported
A. A figure is added in the section 2, showing RPC + Scintillator layout.
- Consider to remove verbose texts, for example "Using multiple scattering muon tomography [...] atomic number of the material [13]"

A. Verbose text removed as suggested.