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). **(Added hit plots from experimental data)**

A. Preliminary data from RPC detector is now reported in the paper which shows the Strip profile and 3D hit plot of each detector using cosmic muons.

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. To Detect and discriminate high Z material from low Z material.

- 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.

Style: english and scientific style has to be definitely 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)
(Added detailed catption to fig 5. ie. Angular distribution)
(Need to add detailed caption to results)

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

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)?

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.

(S. Riggi)Design of a muonic tomographic detector to scan travelling containers (Muon Portal) (Covers ref num 14 & 17)
(M. Benettoni) Noise reduction in muon tomography for detecting high density objects

(FOR INTRODUCTION)
Pioneer works
E.P. George, Commonwealth Engineer (1955) 455. (1)
L.W. Alvarez et al., Science 167 (1970) 832. (2)
K. Nagamine, et al., Nucl. Instr. and Meth. A 356 (1995) 585. (3)
K.N. Borozdin et al., Nature 422 (2003) 277. (4) (ADDED)
L.J. Schultz, et al., Nucl. Instr. and Meth. A 519 (2004) 687. (5) (ADDED)
L.J. Schultz, et al., IEEE Trans. Image Process. 16 (2007) 1985 (6) (Already quoted in paper)

(FOR ALGORITHM)
Muon Tomography Algorithms
S. Riggi et al., NIM A 728 (2013), 59-68 (7) (ADDED)
D. Poulson et al., NIM A 842 (2017), 48-53 (8)
M. Benettoni et al., JINST 2013 8P12007 (9) (ADDED)
S.Chen et al., JINST 9 (2014) no. 10, C10022 (10) (ADDED)

(FOR HARDWARE)
Plastic scintillator/fibre detectors for muon tomography
F. Riggi et al., NIM A 912, 21 (2018) 16-19 (11)
A. Clarkson et al, NIM A 745 (2014) 138-149 (12)
V. Anghel et al, NIM A 798 (2015) 12-23 (13)

Gas detectors/RPC for muon tomography
S. Pesente et al., NIM A 604 (2009), 738-746 (14) (ADDED)
P. Baesso et al., 2013 JINST 8 P08006 (15) (ADDED)
X. Wang et al, NIM A 784 (2015) 390-393 (16) (Already quoted in paper)

GEM Detectors for muon tomography
K. Gnanvo, et al., Proc. IEEE NSSMIC 2010, NI19-90, pp. 552-559, arXiv:1011.3231 (17)
(Above Muon Portal Ref)

- 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.

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 simulation where we have considered exact hit point, the track is simply the straight line connecting the hit.

One line for incoming track and other for the outgoing track. When we are considering the strip of 3 cm width then the track is basically the straight line fit for the center of hitted pixel.

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. To make it more clear we have added another algorithm which does the filtration based on Standard deviation of scattering angle of PoCA points that lie within each voxel. The algorithm makes use of weight PoCA Point count, which is define as product of PoCA point count in a Voxel multiplied by SD of scattering angle of PoCA points within the Voxel. If the weighted count is greater than zero than the Voxel is considered as signal voxel. The object detection method using Chauvenet's criteria is based on the number of PoCA point that lie within the Voxel. These are considered as signal points that forms the image if the PoCA point count is more than the threshold selected using Chauvenet's criteria. For the exact hit point scenario, there is not much deviation in the results of both the algorithm. But when working with the smeared data, then SD based algorithm work nicely, although the number of reconstruction point are much less when we considered smeared data, because due to smearance the number of false positive increased substantially and results in decrease in the number of true positive. This increase false positive is nicely chopped off by SD based algorithm, by decreasing the voxel size. The decrease in voxel size also removed some proportion of true positive point, but the false positive are removed at much higher rate due to their uncorrelated nature.

But we can't keep on reducing the voxel size because after a point the rate of removal of true positive dominates the rate of removal of false positive and which result in overfitting and we start loosing true positive points, which will affect our reconstruction. In the scenario of four block we have observed this overfitting starts once the size of voxel is reduced after112.5 cm^3. This defines the limit of voxel size, where get a balance trade off between the false positive rejection and reconstruction.

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 applying threshold we are able to tell the object location visually, and it can be programmatically determined by calculating the cluseter and centroid of cluster as the object location. The value is calculated based on the processing of Voxels. The count of PoCA points in the voxel where material is placed gives the longer peaks at the corresponding voxel number as shown in the fig 11 & 12. Now the threshold value is selected as the mean of the voxel count as the function of number of 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 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 dense object you get dense cluster of PoCA point in the voxel where the dense 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. Plot have been added to the paper show the variation of detector geometrical acceptance on the voxel which are close to upper plane, bottom plane and a floating center plane. Another graph showing the PoCA point count variation on floating target along Z direction is also shown. Which clearly indicates that how the point count varies as we move the scatterer from upper plane towards lower planes. The point count starts increasing upto the point when the scatterer reaches the center in the along the Z axis, and then it agains starts decreasing as expected.

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 peak for the bin corresponding to that voxel number as shown in figure 11 & 12, 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 11 & 12 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.

- 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. Done, TPR->TPDR, FPR->FPDR **(MAY BE WE SHOULD ALSO ADD SOME MORE EXPLAINATION)**

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. TPR 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 TPR by reducing the number of false positives. If we take considerable bigger voxel then it will contains not only true positive but also the false positives i.e. the points which lie outside the scatterer region, as a result to this the total number of PoCA points which is basically the sum of True Positive and False Positive increase and will result in bigger denominator and hence lower the TPR value. When we divide the bigger voxels to small one, then some of the smaller voxels contains only false positive whose counts are much less as compare to the counts in the Voxel contains the scatterer, which is show as small peaks in the distribution in figure 11 & 12.

- 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

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. We have investigated various clustering and outlier detection algorithm like DBSCAN, KMeans, ABOD. All these algorithm are supervised and needs input the number of clusters that we want to find, and also the location of objects so that the outlier for the cluster at the specified locations can be rejected. On the other hand the voxelization based algorithm proposed in the paper makes use of counts in the voxel and Chauvenet's criteria to chose the threshold, to determine if the voxel can be considered as signal or background. **(ADD THE CORRESPONDING TEXT IN THE**

PAPER)

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. Done

== Minor comments/suggestions ==
- Separate units from value throughout the paper, e.g. in Section 2 180cm-->180 cm. **(Done)**
- Review text to define acronyms used, e.g. NPD (Nuclear Physics Division)? **(Done)**
- "Simulation" in lower case (e.g. beginning of Section 3 and in other parts) **(Done)**
- Many typos found, e.g. Pont of Closest Approach (section 4.1), missing punctuations, missing blank spaces after commas, etc (Continuous progress)
- Review english style and correct typos (correct verb per subject, etc) (Continuous progress)
- Use same number of significant digits in numbers and relative errors, e.g. m=... , n=2.025 +- 0.016666 in Section 3. **(Done)**
- 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 : **Fig 4 Removed (Done)**
- 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. **(Done , added 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). **(Done, specified air as background volume)**
- Consider to remove the text at the beginning of section 4. (up to text "for individual muons.") as already discussed before. **(Done)**
- Consider to remove Section 2.3 and include it (in a short and concise version) at the beginning of section 2. **(Done)**
- Consider to add a Figure near Fig. 1 showing RPC + SCINTILLATOR layout with detector sizes reported
- Consider to remove verbose texts, for example "Using multiple scattering muon tomography [...] atomic number of the material [13]"