1 Generic Comments

The paper writing can be improved. Section 2 reads vaguely as a user manual with some explanation of the details of the model. The code already comes with a User Guide and I find the explanations written in the paper on how to call the functions rather repetitive and distracting. The fact that the paper uses symbols and variable names that are different from those used in the code creates unnecessary confusion. For example at page 7, the variable Nr is also closet.N_raysXSA and so on. My suggestion is to dedicate a separate section for the model description (which should be the focus of a GMD paper) where only mathematical symbols and clearly defined variables are used. After, the authors can give some details of the implementation and provide code examples on how it is possible to interact with the software.
The description of the methods can be generically improved. Sometimes, variables are used without being introduced before. For example d_eq is used at page 6 and called "characteristic size", than it is defined for ellipsoids only at page 10 and it seems to be a spherical equivalent diameter, but at page 7 (line 194) it is the maximum dimension and at page 8 the spherical-equivalent diameter is used to define the size of the aggregates and it is called D_agg. The units of the quantities are also not specified, I believe this is because the shapes can be scaled and do not require to be defined into specific units (i.e. if the STL files is defined in meters the resulting Dagg will be meters, V will be m**3 and so on, same for density). I think it will be useful to make this explicit: the authors should say that the units of the output will depend on the units used for the input. On the other hand, there is, at least, one quantity that must hold units and this is the solid angle omega which also appears with numbers in various places of the text and figures, but without units. I also have the feeling that the authors make implicit assumptions sometimes and this does not help understanding the argumentation, I will try to point where this occur in my detailed review.

The model aims at being generically applicable to various collision phenomena in geophysics. However, at least for what concerns snowflake aggregation (my field of expertise), it fails to cite adequate literature and the conspicuous production of models that happened in the past 20 years. Sometimes, the cited literature is also inappropriate as it does not sustain the argumentation in the text. I cannot speak about interstellar dust grains, but the model is not applicable to the problem of snow aggregation. Judging by the results presented in this paper it also seems to not fully capture the properties of volcanic ash aggregates. For these reasons I would suggest the authors to focus on the application they designed the model for and clearly demonstrate the suitability of the model assumptions for that specific problem. If this is not done, the code remains a nice piece of software, but with a limited applicability to geophysical problems and the authors might consider publishing it on a different journal.

The authors stress a lot about the generality of the model. However, I have the feeling that SCARLET might be too generic for having an application in geophysics. To me the aggregation process can be divided into 4 main events: a) the generation of monomers distributions, b) the selection of which particles have to collide, c) the collision trajectory and d) the contact dynamics.
a) Is not treated in SCARLET, while it is a fundamental aspect at least for snowflake aggregation [5, 6].
b) In SCARLET the order of aggregation is again arbitrary. The user decides the sequence of collisions and the design choices prevent from making an aggregate of previously aggregated particles (without considerable effort). Again this is a process that is essential for snow aggregation because it follows the process of collision through differential sedimentation [14]. Considering that volcanic ash fall in the Earth atmosphere I do not see why it should follow a different process, although it is true that differential sedimentation can only act after the initial inertia due to the volcanic explosion is dissipated and probably electric or chemical process also have a contribution in the dynamics. The authors should clarify which physical processes play a role and how SCARLET models them.
c) SCARLET makes a very precise assumption about the collision trajectory. Each colliding particle is attracted by the center of mass (CM) of an evolving core object. This is not the case for snowflakes (the authors cite [4] where this is explicit), I guess this might be true for planet formation due to gravitational attraction, but again I do not see an immediate justification for volcanic ashes (perhaps electrostatic forces?). d) SCARLET explores various orientations of the colliding particle and select the one that minimizes the distance between the CMs of core and colliding particle. Again this is not seen in snowflake formation where the crystal attitude is generally determined by aerodynamic forces, and I do not see a justification for volcanic ashes. It might be true for planet formation, but I suppose that the most important effect of collision would be the break-up of the colliding objects.
Considering my list of aggregation events. SCARLET implements only c) and d). This of course increases the applicability of SCARLET to different problems, but reduces the scope of the model from aggregation of particles to simple collision-collection.

What the paper lacks the most is a clear justification of the design choices made for the model. Since I am not an expert of volcanic ash aggregation I might lack some knowledge that is trivial to volcanologists. Because of that, please correct me if I am wrong.
I suggest the authors to do a major revision of the study and include a model description in their revised manuscript. I honestly do not understand what is the rationale behind the choice of making the aggregation trajectories converging towards the CM of the core particle. What is the dynamical model of ash aggregation? Why should the CM of

the aggregating and the core particle be minimized? Failing of explaining this breaks the link between the model and the physics. The model will indeed produce aggregated particles, but not particles that are resulting from a physical process. If the physics behind ash aggregation is unknown it will be only logical to make simplifications and assumptions, but again this must be clearly stated in the description of the model.

While reviewing this paper I tried to gain some bibliographic information about volcanic ash aggregation. This recent paper [1] lists three processes that are acting in the determination of an aggregation kernel, namely Browninan motion, fluid shear and differential deposition. My experience in snowflake aggregation makes me believe that the discrepancies found in the paper between observed and modeled aggregates are likely to be a result of the lack of a physical model for ash aggregation. This is simply because if one allows the larger object to collect already aggregated particles the packing efficiency of the end process will be much lower. I suggest the authors to take this into account because: 1) The existence of an aggregation kernel is not incompatible with the evidence of having a large object collecting smaller ones; 2) Leveraging on the knowledge taken from snowflake aggregation, it is very likely that the lack of porosity of the modeled aggregates is due to an incorrect aggregation dynamics; 3) The inclusion of an aggregation kernel binds the model to the physics.

2 Major points

1. There are some occasions where the authors mention the necessity of reaching a maximum packing. Why is this the case? I cannot find a reference for that. This is confusing because I first though that the code was designed to reach the maximum packing, but it failed. However what is shown later is that even without enforced packing the code produces aggregates that are too packed with respect to the physical evidence. This is because of a switch in the focus from packing to porosity. I would suggest to provide evidence for the maximum packing theory (or minimum porosity) and shape the manuscript in the sense that it only looks for the packing tau and not porosity (or viceversa).

2. Lines 10, 11, 54, 55, and other parts. I do not see a clear justification for the design choice of a "central oriented collision". This is clearly not the case for snowflakes [14]. The fact that samples of volcanic ash present one block much larger than the others can be explained by a non continuous size distribution of monomers and again the process of differential sedimentation.

3. Line 15. I do not see why the distance of the CM should be minimized. What is the physical force that acts in this process?

4. The cited bibliography in the introduction needs some work. In particular for snowflake aggregation the only paper among those cited that covers aggregation is [4]. Some useful resources might be [14, 6, 5]. Concerning available models there are some resources [7, 10, 11] which are usually referenced using the following papers [6, 9, 2]. All these models are capable of dealing with arbitrary shapes and it does not seem to be difficult to deactivate the aggregation kernel in order to simulate a central aggregation process, this is evident especially for [9] where this effect is fully explored. A comprehensive review of snowflake aggregation models can be found in [13].

5. Line 55. I do not see how the evidence of finding an aggregate composed of a piece much larger than the other constituents is supporting the idea of a central aggregation effect. The same result can be achieved from the differential sedimentation of discontinuous distributions of monomers. This seems physically sound to me if I imagine the initial distribution of monomers as resulting from the explosion of rocks where the internal inhomogeneous distribution of mass and hardness might play a substantial role. Also what would be the physical force that generates this central pole of attraction? I think that these argumentation needs more discussion or the citation of previous work.

6. Line 200-207 (also figures 6 and 7) the approach of having trajectory cones centered around the CM is not convincing to me. It seems to implicitly assume that the CM is always inside the core particle and thus the cone always intersects the particle surface. What happens if the core particle is shaped like a C or a parabola? The cone might not intersect the shape. Alternatively, consider a flat surface. In this case the contact point will always be the same regardless of the angle omega. I know that these are rather exotic cases (having the CM outside the shape is actually rather common for snowflakes), but I still do not know if I should consider this code as rather generic or applicable to only specific geophysical situations.

7. Sections 2.2 and 2.3 The code uses the spehere equivalent representation and avoid the overlapping of the spheres. Somehow it must be possible to get back the STL representation of the aggregates. How is this done? What is the degree of overlapping of the final STL polyhedrons? This might have an impact on the aggregate porosity

8. Line 245-265 This part is quite difficult to follow. The ellipsoids are defined only for by flatness and elongation if one assumes that the generic size is always the same, or scale invariant. Is the size always the same? also the parameter gamma is the ratio between a (a property of the monomers) and R (a property of the aggregates) this means that a is uniquely defined only if the same ellipsoids are used within an aggregate and R will depend on the number of monomers. Am I right?

9. Line 308 I suggest to not refer to this as a time evolution. As a matter of fact the aggregation code does not involve process rates. Time is never a variable here. What the authors can analyze is the evolution with respect to the number of collisions. An "aggregation stage" perhaps. I would also be curious of knowing why this is not observable in field and lab studies. One can envision the possibility of collecting many samples and derive the properties of them as a function of the number of monomers, which is precisely what the code can give.

10. Line 357 The expected mass loss is a new concept for me. Also what is this Na replication actually doing?

11. I disagree with the statement at line 359. The small particles might be individually a second order contribution to the final mass. But collectively they can contribute more than the largest. This is simply because the contribution to the final mass depends also on the number concentration and according to fig 14 there are a lot of small pieces. More importantly, the contribution of small particles to packing is even more important. This is because small pieces are the only one able to fill the gaps in between larger pieces and achieve higher packing values. as packing approaches 1 the importance of having small pieces accounted for is larger and larger.

12. Line 378 this measured values between 80-97%, why are them not reported in the figure 16 and 17 like it was the case for figure 15? I have the feeling that in contrast to what is affirmed here even the loose packing is not capable of reproducing the observed porosities.

13. An effect that was not commented emerges from fig 16. The sequential densely packed do not reach a plateau, but their porosities start decreasing after some number of collisions.

14. Line 425-427. I do not really understand the argumentation here. The authors suggest that there is a violation of the contact comdition. Is this condition the idea that aggregates stick immediately on contact? Also they speculate that liquid bonds might be responsible for this. Is there any evidence of this process? Also, it is not immediate to me understanding how this liquid bond works. Perhpas the particle collide formin a liquid bond, but they bounce. While bouncing the liquid solidifies generating the permanent bond? Is this the process suggested? Shouldn't there be evidence of this in the samples?
    I know I might sound repetitive, but also the implementation of a collection kernel (differential sedimentation, brownian or else) would increase the final aggregate porosity.

15. Line 440. Again, if you let the colliding particle aggregate among them before sticking to the larger one (implement an aggregation kernel) the history would be much different and also the final properties.

3 Implementation details

From what I can understand from the code there are two passages that seems to not fit the what it is stated in the manuscript.

The Euler matrix used at line 209 should explicit the convention used. There are many of them, from what I see from the code implementation it seems the authors used the ZXZ convention, please clarify this. The Euler angles at line 209 are picked from the wrong domain. At line 284-287 of the code SCARLET_v1.m the phi, theta and psi angles are sampled uniformly from the (0, 180), (0, 180) and (0, 360) domains respectively. Actually, for random orientations in space the phi angle should be sampled in the (0, 360) domain and the theta angle should be sampled uniformly in the (0, 1) domain of cos(theta). The current implementation generates orientations that are more dense around the "poles".

This is not a big problem for the results since the angles are only useful for the Monte Carlo minimization. Sampling a non-uniform distribution means that on average more orientations have to be sampled in order to reach the same convergence. However, the Monte Carlo minimization could benefit from a performance boost, and the effects (if any) should be visible in Fig. 9. In general it is better to not make implementation mistakes.

Also the versor v (line 201) components are derived from a normal distribution. This is something different from having the versor in a random direction on the unit sphere which is achieved with the procedure detailed before. In any case this is different from what is written in the paper at line 201. Both what is implemented and what is written in the paper does not provide uniformly distributed colliding trajectories. This might have an impact on the final aggregate properties because if the collision trajectories are not uniformly distributed the resulting aggregate is more likely to have "blobs" of colliding particles.

What I might suggest is to consider the reciprocity relation: having random colliding trajectories is equivalent of having always the same trajectory (let's say the vertical z axis [0, 0, 1]) and the central particle rotated randomly. This simplifies the implementation quite a bit because we already have a function that rotates the particles (with the corrections suggested before) and this avoids unnecessary coding. Also sampling the rays in a cone always centered around the z axis is simpler in my opinion.

4 Minor points

These are mostly a list of suggestions which I hope will be useful to the authors

1. Line 13 (and others). The benefit of 3D printing is not clear. Perhaps this resource demonstrate a scientific application [8]

2. Line 14. The concept of a vector of sizes and shapes is not clear at this stage. I would rather say "a list of colliding shape"

3. Line 14 (and many other places). A spherical equivalent representation of a irregular shape is not necessary for an analytic solution of the collision problem. The solution for a generic polyhedron (such as those represented in STL shapes) is formulated analytically by considering all vertexes and faces instead of simply center and radius of the spheres. This is implemented in [10, 11] for the aggregation of snowflakes.

4. Line 16 (and many others). I personally do not like thinking about functions and parameters as entities that can take responsibilities (at least not in a scientific paper). Instead of saying that those objects are "in charge" of something I think it is more clear to just explicitly say what is their meaning or what happens when changed, invoked, called ...

5. Lines 16, 102, 146, 250, 360, 400, 402, 702, and in connection my generic comments. The angle omega is called solid angle, however in the code it is called "aperture angle". The values of omega does not have units but the range of values are compatible with linear angles in degrees. This means that omega is not a solid angle, but the aperture angle. Please clarify and correct.

6. Line 19. The fact that SCARLET uses STL shapes is not unique (similar approaches in [10, 11, 7]) and does not extend the range of applications. It only means that it can deal with particles of arbitrary shapes. The implemented physics is the main limitation for its applicability.

7. Line 27 The term "minimization" is rather generic and not clear at this stage. I think the authors should refer to the forced compaction of the aggregating particles.

8. Line 39 I think that [5] might be a perfect example of how virtual reality can compensate the lack of direct observations.

9. Line 51-53. The arguments presented is only valid for monodispersed distributions of monomer sizes, but the theory of fractal aggregation is not limited to those cases [14, 12]. Moreover, fractal aggregates are in general the result of the modeling of physical processes and not a model by itself (although there are cases where fractals are used as models of particles). The fractal properties of aggregates emerge in general when the size of an aggregate is much larger than the size of any of its constituents.

10. Line 75. The term opensource does not oppose under-license. The vast majority of opensource software are released under license (including SCARLET).

11. Line 86 I think that the MATLAB software and programming language is written capitalized. The company releases the IDE two times a year and the releases are named for example as R2015a, while the version number follows a different scheme. It seems that R2015 implements either 8.5 or 8.6 version of the language. Also, it is not clear if 2015 means that the package has been only tested with the release R2015 (a or b?) or if it compatible only for R2015 and newer versions.

12. Line 80. I do not see the scientific value of including aggregates of t-rex and LEGO characters. However I find it not much distracting, so they cause no harm (and I personally find it funny). However I am not sure if LEGO must be written making explicit that it is a trademark and if the company is ok with their name being used in this publication.

13. Line 98. "User-defined number" should have a hyphen. At line 102 the hyphen is used correctly. Conversely "User Guide" at Line 110 should not be hyphenated. It might worth checking.

14. Line 112. I think GMD require a fixed DOI for the code. So the zenodo link should always refer to the fixed version 1.0 of the code and not being updated.

15. Line 155. Cancel "as a matter of fact"

16. Line 156 The subject here is the triangles which are uniquely defined by the coordinates of their vertices. The faces are themselve the triangles. The whole first paragraph (line 154-159) is superfluous. You only need to say that STL represents any surface as a set of triangular facets in the 3D space.

17. Line 160. I believe that the *inpolyhedron* function is not built-in but rather a contribution [3] which has been ported in the code

18. Line 160-164 This paragraph does not read very algorithmically. We first place one point, than we check against other points which we did not place yet. We check also if points are inside spheres that are not defined anywhere. Since this looks like an iterative process you can clearly explain it iteratively: Say how to place the first sphere, than say how given N spheres you can find the sphere N+1.

19. Line 167. I think it is better to explain from the beginning what is the effect of using Oext=1, Otherwise, one might think that this takes out some mass. Also, from what I understand of the documentation the function *boundary* this is a convex hull finding algorithm. I think it is clear and simple if you call it convex hull. The idea of "discarding internal spheres" or "those fully covered" is not straightforward.

20. Line 170. I think the word trial has very specific meanings in english (in law, medicine and so on). A better word might be tries, or attempts.

21. Line 197. I though that at this stage the shapes are represented with their sphere-analogues. Why is the maximum dimension of the STL cited here? How it relates to the maximum dimension of the sphereical representation?

22. Line 199. Might be worth making explicit what is the nesting relation among the loops. The fact that Loop 2 is nested inside loop 1 becomes clear while reading, but I do not see the point of having Loop 3 nested in loop 2. Loop 3 could be one of the routines performed inside loop 1 and the same set of rotated colliding particles can be reused for each ray. Potentially it makes possible to parallelize loop 2 and loop 3 independently and increase performances.

23. Line 201 Components cannot be randomly oriented, they are scaler values. Look also at my comments on the implementation.

24. I assume, but it is not written anywhere, that the colliding particles are placed with their CM on the rays. Am I correct?

25. Line 235 Can you please provide a better reference for the algorithm that computes volumen from STL? Like an URL?

26. Line 243 Eq.3 Dagg does not provide additional information with respect to Vext. Would it be better to calculate the final Dmax?

27. Line 313 I think that ash concentration would be overestimated by models if aggregation rates are underestimated not generically an uncorrect parameterization.

28. Line 325 How can the high-speed camera derive the density of the object before they are captured? This is an evergreen problem in snow microphysics.

29. Figure 9 The other of the parameters changes from the figure to the caption. And I believe that also the number of monomers should be a parameter here.

References

[1] S. D. Egan, M. Stuefer, P. W. Webley, T. Lopez, C. F. Cahill, and M. Hirtl. Modeling volcanic ash aggregation processes and related impacts on the april–may 2010 eruptions of eyjafjallajökull volcano with wrf-chem. *Natural Hazards and Earth System Sciences*, 20(10):2721–2737, 2020.

[2] P. Eriksson, R. Ekelund, J. Mendrok, M. Brath, O. Lemke, and S. A. Buehler. A general database of hydrometeor single scattering properties at microwave and sub-millimetre wavelengths. *Earth System Science Data Discussions*, 2018:1–42, 2018.

[3] Sven Holcombe. inpolyhedron - are points inside a triangulated volume? https://www.mathworks.com/matlabcentral/fileexchange/37856-inpolyhedron-are-points-inside-a-triangulated-volume, MATLAB Central File Exchange., 2015. Retrieved February 4, 2021.

[4] Ken ichi Maruyama and Yasushi Fujiyoshi. Monte carlo simulation of the formation of snowflakes. *Journal of the Atmospheric Sciences*, 62(5):1529 – 1544, 2005.

[5] M. Karrer, A. Seifert, C. Siewert, D. Ori, A. von Lerber, and S. Kneifel. Ice particle properties inferred from aggregation modelling. *Journal of Advances in Modeling Earth Systems*, 12(8):e2020MS002066, 2020. e2020MS002066 10.1029/2020MS002066.

[6] J. Leinonen and D. Moisseev. What do triple-frequency radar signatures reveal about aggregate snowflakes? *Journal of Geophysical Research: Atmospheres*, 120(1):229–239, 2015.

[7] Jussi Leinonen. aggregation. https://github.com/jleinonen/aggregation, 2017. Retrieved February 4, 2021.

[8] M. W. McCorquodale and C. D. Westbrook. Trail part 2: A comprehensive assessment of ice particle fall speed parametrisations. *Quarterly Journal of the Royal Meteorological Society*, 147(734):605–626, 2021.

[9] D. Ori, T. Maestri, R. Rizzi, D. Cimini, M. Montopoli, and F. S. Marzano. Scattering properties of modeled complex snowflakes and mixed-phase particles at microwave and millimeter frequencies. *J. Geophys. Res. Atmos.*, 119(16):9931–9947, 2014.

[10] Davide Ori. Sam snow aggregation and melting. https://github.com/OPTIMICe-team/pySAM, 2014. Retrieved February 4, 2021.

[11] Torbjörn Rathsman. Snowflaketoolkit–a toolkit for generating ice particles. https://github.com/milasudril/snowflake-toolkit, 2016. Retrieved February 4, 2021.

[12] C. M. Sorensen. Light scattering by fractal aggregates: A review. *Aerosol Science and Technology*, 35(2):648–687, 2001.

[13] J. Tyynelä and Annakaisa von Lerber. Validation of microphysical snow models using in situ, multifrequency, and dual-polarization radar measurements in finland. *Journal of Geophysical Research: Atmospheres*, 124(23):13273–13290, 2019.

[14] C.D. Westbrook, R.C. Ball, P.R. Field, and A.J. Heymsfield. Universality in snowflake aggregation. *Geophys. Res. Lett.*, 31(15):L15104, 2004.