Geosci. Model Dev. Discuss.,
https://doi.org/10.5194/gmd-2020-239-RC1, 2020

Geoscientific Model Development Discussions

Interactive comment on "A note on precision-preserving compression of scientific data" *by* Rostislav Kouznetsov

Mario Acosta (Referee)

mario.acosta@bsc.es

Received and published: 7 September 2020

This paper explains the limitations for data compression of one default precision-trimming method used for NetCDF and propose a simple but effective way to improve it using a mantissa-rounding technique, a novel idea which improves significantly the precision results. Moreover, the paper proves that this Bit Grooming algorithm has sub-optimal accuracy and produces substantial artifacts in multipoint statistics. Additionally, they suggest a way to rectify the data already processed with Bit Grooming.

The new technique presented should be interesting for the community and novel enough for publication, being one of the default methods used by NetCDF. I would recommend extending significantly the state of the art and some rewriting of the sec-

tions for a better comprehension before this paper could be published and to support the novelty contribution.

One main question that the author would explain here before publication. How should the rounding bit size would be decided? Is this a parameter so set up through NetCDF?

Will the precision depend on the particular application or not? Apart from the array examples presented, the author should explain more in detail this.

Please find other requirements according to the section of the paper: line 14: Consider 32 bits as float can be confused since this depends on the program language. Actually, some languages define float as double precision, using integer (int) for this declaration. If a more in detail explanation is not done, I would remove float and double. lines 17-23: Have you consider different scientific fields to affirm that level of accuracy (less than 7 bits) or entropy? I agree that in most cases single precision is more than enough but you should enumerate cases where this could not be true and provide a state of the art about this. As an example, I consider that in the inputs for data assimilation of chaotic applications, as weather models, the level of accuracy or entropy to ensure the number of bits used could affect the results. Introduction: I consider that the state of the art should be extended. In order to support the argument explained in the second paragraph, some successful examples should prove that the reduction of precision does not affect to the accuracy. Both of them should be listed, applications using the reduction of precision and the reduction of precision to save data. Introduction: Is precision-trimming the single compression method available? Apart from a default method, is it the most used or one of the best methods for Netcdf files? The author should explain why this particular method and why the comparison done should be interesting for the reader, including some state of the art about this (apart from the single reference provided). The mantissa-rounding technique explained here should highlight if it is novel or there are other works which present similar approaches, or at least, differences among them. Section 2 should be rewritten. Second paragraph is difficult to follow and the explanation about the problem of trimming LSBs difficult to

understand. Five methods are described but only three of them are commented and only two discussed in details without more explanations about the reasons. It is also not clear how these results in Figure 1 are obtained. Moreover, it is not clear if these results and the explanation given is coming from Zender (2016) paper or from Section 3 and 4 of this paper. The author should highlight the novel contribution of this paper for the different sections.

As minor details about the figures, X and Y axes for all of them do not contain units. Figure 2: a) and b) are difficult to see.

Interactive comment on Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2020-239, 2020.