

***Interactive comment on “Bit Grooming:
Statistically accurate precision-preserving
quantization with compression, evaluated in the
netCDF Operators (NCO, v4.4.8+)” by
Charles S. Zender***

C. S. Zender

zender@uci.edu

Received and published: 30 July 2016

(Reviewer's comments/questions are in *italics* and my responses are interspersed in plain text.)

I thank the Reviewer for their comments. The Reviewer's questions indicate that the submitted manuscript did not give enough background on the precision and range of the methods employed besides Bit Grooming, and did not adequately intercompare the trade-offs of Bit Grooming with the trade-offs of the other methods, linear Packing in particular. The revised manuscript, which is attached as a PDF to this response,

C1

includes a fuller explanation of the range and precision of linear Packing. This made it easier to describe the trade-offs between compression ratio and precision incurred by Bit Grooming in comparison to Packing.

The core contribution of this paper appears to be the level of compression achieved while retaining a high degree of dynamic range, as well as statistical properties of resulting data. This contribution, compared to other methods, is only clearly articulated in the sentence spanning pages 9–11.

Agreed. The revised manuscript now presents more clearly the trade-offs between size and precision for linear Packing in Section 2.2, and the trade-offs for Bit Grooming in Section 2.3. While these trade-offs are still noted in the results in Section 3.3, the inter-comparison of these trade-offs (which the Referee suggests is the core contributions of the paper) is now concentrated in the newly created Section 4.1 "Comparison of Lossy Compression Techniques".

Tables 4–7, with some interpretation, are good at conveying the relative resultant size after applying the algorithms examined. This may be a good place to bring together, and highlight, the interplay between the data size and precision achieved at that size.

Agreed. As described above, we now bring together the discussion of interplay between size, range, and compression in the newly created Section 4.1. In addition, Tables 4–7 are now easier to interpret. Separate columns clarify the lossless compression method (column LLC), the quantization method (column Qnt), and the overall compression method (column Method).

The number of significant digits is already presented for the Bit Groomer methods. Could this be added for the other methods, either in theory or on a particular data set?

The NSD column in Tables 4–7 now includes the precision for all methods.

Additionally can the dynamic range, or number of bits remaining in the mantissa?

Precision and dynamic range are, for floating point values, determined by bits in the

C2

mantissa and exponent, respectively. This is the case for Bit Grooming. Since packed data are integers and have no exponent, their integer bits determine both their unpacked precision and their dynamic range. Hence we interpret the Referee's questions as asking whether both precision and dynamic range can be added to the Tables. The short answer is Yes, and the revised manuscript includes this information in Tables 4–7.

The new Range column specifies the dynamic range for each method. And the NSD column has been completely filled-in to show the number of significant digits for all methods, not just Bit Grooming. The number of bits retained (in contrast to digits) is described in Section 2.2 and 2.3 for Packing and Bit Grooming, respectively.

Errata:

1. During editing I identified and corrected a numerical mistake. The original manuscript erroneously multiplied the exponents rather than the mantissas by two to estimate the dynamic range of IEEE SP and DP from their maximal values. The correction in the revised manuscript changes

“The dynamic ranges of SP and DP numbers are $\sim 10^{74}$ and $\sim 10^{616}$, respectively, whereas data packed linearly into two-byte and four-byte integers have dynamic ranges of $\sim 10^5$ and $\sim 10^{10}$, respectively.”,

to

“The dynamic ranges of SP and DP numbers are $\sim 10^{37}$ and $\sim 10^{308}$, respectively, whereas data packed linearly into two-byte and four-byte integers have dynamic ranges of $\sim 10^5$ and $\sim 10^{10}$, respectively.”.

Changing the exponents by a factor of two does not qualitatively alter the results or conclusions of the manuscript.

2. The revised manuscript now contains the citable references to the four datasets that are intercompared, and the datasets have been made available online (figshare.com) as described in the Supplement.

C3

3. The revised manuscript contains many minor wording changes that characterize the precision and dynamic range of linear Packing, as requested by the Referee, and how these characteristics compare to those for Bit Grooming.

Please also note the supplement to this comment:

<http://www.geosci-model-dev-discuss.net/gmd-2016-63/gmd-2016-63-AC2-supplement.pdf>

Interactive comment on Geosci. Model Dev. Discuss., doi:10.5194/gmd-2016-63, 2016.