

Interactive comment on “Finding the Goldilocks zone: Compression-error trade-off for large gridded datasets” by Jeremy D. Silver and Charles S. Zender

Anonymous Referee #2

Received and published: 30 August 2016

This paper describes a variant of lossy encoding which leverages the multi-dimensional nature of many scientific datasets that have greater data variances along different axes. The axes with small variations in data values are labeled "thin dimensions" and the axes with large variations in values are labeled "thick dimensions". The datasets are then "layer packed" with a linear scaling algorithm in the thin dimensions, recording a scale & offset value for each coordinate in the thick dimension.

I think the insights into the "thick" and "thin" dimensions are the primary value of this paper, with the actual compression algorithm and results being less important, overall. Applying the idea of thick & thin dimensions appropriately to other compression methods (such as the JPEG-2000 algorithm used in GRiB2) would be more valuable than

C1

just the idea of the simple scale & offset compression chosen.

Near the bottom of page 6, "for simplicity will have" should be corrected to "for simplicity we have".

Interactive comment on Geosci. Model Dev. Discuss., doi:10.5194/gmd-2016-177, 2016.

C2