

Interactive comment on “A note on precision-preserving compression of scientific data” by Rostislav Kouznetsov

Seth McGinnis (Referee)

mcginnis@ucar.edu

Received and published: 11 September 2020

GENERAL COMMENTS

This paper describes some undesirable effects of the lossy compression algorithms currently implemented in NCO, and presents a correction that avoids these problems. This paper is well-structured, clearly written, logically sound, and of interest to the GMD community. The scientific significance, quality, and reproducibility are all very good; the only recommendations I have to make are minor revisions regarding presentation quality. Good work!

SPECIFIC COMMENTS

- 1) The subject is a little dry; I think the paper would benefit from a brief discussion in
C1

the Introduction of the real-world use-case that led to the discovery of the two-point distortion, which would help the reader to follow the analysis using the structure function and understand why the issue matters.

- 2) The figure placement is a bit off. The figures should come shortly after their first mention in the text. This could be a problem that will be solved when the article goes to press, in which case please disregard this comment.

- 3) Figures 3 and 4 need axis labels. I would also recommend placing the legend outside the panels.

- 4) Also in Figures 3 and 4, what does "bg=300" mean in the title of each panel? If it's not necessary, it should be omitted; if it is needed, put it only in one place, perhaps in the caption.

- 5) In the captions for Figures 3 and 4, mention that the structure function is defined in section 4 of the text.

TECHNICAL CORRECTIONS

Line: change from this -> to this

2: magnitude keeping -> magnitude while keeping

5-6: use of a lossy compression -> use of lossy compression

8: twice higher precision -> double the precision

17: only few (or even few tens -> only a few (or even a few tens

18: As a result application -> As a result, application

19: dataset size since -> dataset size, since

20: high entropy, that -> high entropy, which

22,27,28,37,38,94,118: precision trimming -> precision-trimming

26: 2020) work efficiently -> 2020) to work efficiently
28: It has been quickly noticed hat -> It was quickly noticed that
28: implementation of a precision -> implementation of precision
33: algorithms and inaccuracies -> algorithms and the inaccuracies
33: Our analysis revealed -> Our analysis reveals
34: distortion two-point structure -> distortion to the two-point structure
34: fields, caused -> fields caused
35: twice higher accuracy -> double the accuracy
36: Besides that, -> In addition,
37: a method ... is suggested -> we suggest a method
41: extrapolated on -> extrapolated to
51: numbers set transform it so, that -> numbers, transform it so that
53: for N most-significant -> for the N most-significant
53: mantissa, that we let to store -> mantissa, which we use to store
54: "tail bits" the remaining -> "tail bits" for the remaining
60: ones -> one
63: rounding of mantissa -> round mantissa
64: have been implemented -> were implemented
64: and described well -> and are described well
66: floating-pint -> floating-point
76: kept in mantissa -> kept in the mantissa

C3

85: methods of above -> methods above
85: correspondingly -> respectively
86: full quantum -> a full quantum
86: with mean absolute -> with a mean absolute
87: alternate -> alternately
88: at most half-quantum -> at most a half-quantum
89: Same applies to the he -> The same applies to the
90: have equal chance -> have an equal chance
91: margins for an error -> margins of error
93: method -> methods
93: twice smaller error -> half the error
96: correspondingly -> respectively
97: respect to others -> respect to the others
97: in the scale -> on the scale
98: keep bits -> keep-bits
98: well seen -> easily seen
104: get a positive bias in average -> be positively biased on average
106: to compensate -> to compensate for
107: This procedure, however -> However, this procedure
109: a way to half -> a method to halve
110: and to remove -> and remove

C4

113: consider normalized -> consider the normalized
119: and a random -> and random
120: whereas, the latter -> whereas the latter
120: with high stochastic -> with a large stochastic
120: array is of -> array is
123: summary of NRMSE -> summary of the NRMSE
123: trimming precision -> precision-trimming
124: Table. 1: -> Table 1:
126: average applies -> average is applied
130: Notable, that -> It is notable that
131: signal that has -> signal, which has
135: sensible -> sensitive
136: slightly differs form -> differs slightly from
137: equal to corresponding -> equal to the corresponding
138: the structure functions -> a structure function
141: increase of the number -> increase in the number
142: reference one, however -> reference, but
144: that can be -> which can be
145: due to its the -> due to its
148: twice smaller -> half the
148: error than a -> error of the

C5

148: 2016), that has -> 2016) which has
151: two-point statistics -> two-point statistics in

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

C6