

Nonlin. Processes Geophys. Discuss., referee comment RC1
<https://doi.org/10.5194/npg-2021-2-RC1>, 2021
© Author(s) 2021. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on npg-2021-2

Anonymous Referee #1

Referee comment on "The blessing of dimensionality for the analysis of climate data" by Bo Christiansen, Nonlin. Processes Geophys. Discuss.,
<https://doi.org/10.5194/npg-2021-2-RC1>, 2021

In the present manuscript, various aspects of the blessing of dimensionality are analysed theoretically and experimentally. Specifically, high-dimensional (HD) spaces are shown to feature the phenomena of the concentration of measures and waist concentration which are explained at an abstract level, although with clear references to previous climatological research. The authors show that independent vectors in high dimensions can be taken to be orthogonal and how sample means can be treated as a constant. The extension to dependent samples is discussed by making a review of the notion of effective dimension and the mixing/decorrelation rates of a time-evolving system. Then, three HD climate data sets are studied and shown to match, generally, the earlier theoretical discussion. The CMIP5 data set was argued to possess somewhat dependent samples and thus, the blessing of dimensionality (in the sense of concentration of measures and waist concentration) did not fully apply in this case. Exploiting the orthogonality of independent vectors in HD, the author finds that the correlation between sample differences is equal to 0.5 and that correlations of sample differences and differences between ensemble mean and ensemble members is equal to $2^{0.5}$. This calculation is reflected in the climate data. The importance of sample size is also discussed in the context of HD data, whereby the authors show that ensemble mean length converges to the true mean at a rate inversely proportional to the sample size. This fact is illustrated in the conceptual cases of Gaussian and Gamma distributions as well as two of the climate data sets.

The paper is concise and well explained, linking theoretical findings with experimental ones. I would encourage the Journal to accept this manuscript if some minor comments are taken into account.

Specific minor corrections and comments:

Line 107: are the vectors a and b sampled from the Gaussian distribution mentioned above or are they generic?

Line 119: The inequalities in Eq. (4)-(5) very much resemble a large deviation principle. Making a reference to large deviations theory will strengthen the connection of high-dimensionality with the statistical mechanics theory presented in the paragraph starting in line 99

Line 185: When presenting the climate data, it would be very useful for the reader to know the dimensions N and sample size K in each case. Certainly, this is done later when analysing each case, but I believe doing it earlier could help the reading.

Line 228: The author mentions that $x^k - \bar{x}$ is orthogonal to \bar{x} in high dimensions. This fact follows from the concentration of measures result presented in section 2.2, so I would encourage the author to make a clear reference to the theory.

Line 240: The reference to the spikes of the unit cube is very helpful and in fact it illustrates the comment done in Line 228. The author might consider making this analogy earlier.

Clearly, the CMIP5 outputs deviated the most from having waist concentration, i.e. it provided non-orthogonal samples. This was attributed to high dependence. To what extent could it be attributed, instead, to low effective dimension of the climatological fields? From a personal perspective, I'd be interested in hearing the author's further views on this general question.

Technical Minor Corrections:

Line 18: "This might seem..."

Figure 1: In the caption there is no reference to the right panel (two references to the "left").

Line 171: "Other measure is..."