

## ***Interactive comment on “Evaluating Lossy Data Compression on Climate Simulation Data within a Large Ensemble” by Allison H. Baker et al.***

### **Anonymous Referee #2**

Received and published: 19 September 2016

#### General comments

The paper analyses in detail how lossy data compression affect climate simulation data. After choosing meticulously the level of reduction, they added 3 members in the CESM Large Ensemble project (2 with data loss compression and 1 without) and proposed to scientists to use their own tools to analyse these additional members. 4 different evaluations are summarized: CVDP climate variability diagnostic package, climate characteristics, ensemble variability patterns and coherent structures. They also proposed to compare in detail compressed with original data. 3 different analyses are summarized: climate extremes, causal signatures and AMWG-DP atmosphere working group diagnostics package. During the ensemble data evaluation, reduction factor for one variable has been detected and needed to be changed (from fpzip-16 to fpzip-24). Choices for reduction factor still an important step and required meticulous

C1

action. During the original and reconstructed data analyses, nothing special appears regarding these 3 additional members. This very good news should help to increase the confidence of climate community for lossy and adequate compression. Each analysis is very well described and very clear. The choice of the reduction factor still critical and has to be analysed in detail for a new kind of simulation like the grid resolution, output frequency and numerical precisions as mentioned by authors. After these very complete analyses, the lossy data compression appears as a very useful step to reduce drastically the amount of data produced during climate simulation, especially for large ensemble simulation. It should be considered by all modelling group involved in CMIP6 or, at least, later in CMIP7.

This article is very useful and analyses described should help to convince climate community to use data lossy and adequate compression in their analyses to reduce storage required by large simulations like CMIP.

#### Specific comments

Fpzip tool for compression could be replaced by other smart tools. Lessons learned ie relationship between variables, detectable versus consequential, individual treatment of variables are very important and must be shared through CMIP modelling groups through the WIP: WGCM (Working Group on Climate Models) Infrastructure Panel.

#### Technical corrections

Paragraph 4.2.1 : a figure with the differences between the maximum and minimum monthly averaged temperature is probably missing.

Figures 2 3 4 5 6 7 8 9 10 11 and 13 should include 31-c and 33-c in legend instead of 31 and 33 to help understanding.

Figures titles should be completed:

Figure 7 : add a reference to fpzip-24

C2

âĀĀ Figure 9 : add a reference to fzip-16

âĀĀ Figure 17 : add a reference to the 4 different variables showed

---

Interactive comment on Geosci. Model Dev. Discuss., doi:10.5194/gmd-2016-146, 2016.