Eric Paulz
ECE 4930-004
November 7, 2018

Paper Review: A Methodology for Evaluating the Impact of Data Compression on Climate Simulation Data

Summary
The goal of this paper is to find a way to reduce the amount of storage required to preserve data produced by high-resolution climate simulations. These simulations are used around the world by climate scientists, and we know that current methods will be unsustainable as the resolution/accuracy of simulations increase. The researchers propose the use of compression algorithms to compress the data yet allow the data to be reconstructed with acceptable variability.

Strengths
I liked this paper because it was about something that was easy for me to visualize compared to some of the more obscure concepts we have read about before. I think the breakdown of the paper was really good, and that they spent enough time and care to explain everything without going overboard. The results section was very thorough, and I particularly enjoyed how they presented the data in the figures.

Shortcomings
It was pretty difficult to keep up with all of the different abbreviations throughout the paper.

Improvements
They could have used a few more compression techniques for comparison. Also, since they said they need an algorithm that can be used lossy and lossless, I'm not sure why they even chose to include GRIB2 since lossless isn't even an option.

Question(s) for Presenter
How exactly did they go about customizing each variable as noted in section 5.4?

Additional Questions
- Why study the impact of data compression on climate data?
  - Data from climate simulations is used around the world by climate scientists. As is, the simulations produce a huge amount of data and requires vast amounts of storage. As simulations become more precise and detailed, these numbers will increase and the current methods of storage will not be sustainable.
- What did you think of their selection of compressors?
  - I think they chose good compressors for comparison because, even though none of them fit all of their criteria, they all seemed to have different benefits and drawbacks. This allowed them to really see which of the criteria they specified were the most important for their application.
- What measure of error or error bounding metric is best and why?