# Climate Data Compression via Conditional Emulation

Joseph Guinness

Numerical climate model simulations run at high spatial and temporal resolutions generate massive quantities of data. As our computing capabilities continue to increase, storing all of the data is not sustainable, and thus it is important to develop methods for representing the full datasets by smaller compressed versions. We propose a statistical compression and decompression algorithm based on storing a set of summary statistics as well as a statistical model describing the conditional distribution of the full dataset given the summary statistics. The statistical model can be used to generate realizations representing the full dataset, along with characterizations of the uncertainties in the generated data. Thus, the methods are capable of both compression and conditional emulation of the climate models. Considerable attention is paid to accurately modeling the original dataset--one year of daily mean temperature data--particularly with regard to the inherent spatial nonstationarity in global fields, and to determining the statistics to be stored, so that the variation in the original data can be closely captured, while allowing for fast decompression and conditional emulation on modest computers.