Making space for climate simulations

July 29 2016

A statistics-based data compression scheme cuts data storage requirements for large-scale climate simulations by as much as 98 percent.

By accounting for the specific statistical structure of global climate simulation data, Marc Genton at King Abdullah University of Science and Technology (KAUST), Saudi Arabia, partnered with a U.K. researcher to develop a novel and very effective data compression scheme for large-scale climate simulations. The method promises to significantly reduce data storage requirements and accelerate the capacity for climate research.

The ever-increasing power of supercomputers has allowed scientists to become more ambitious when it comes to simulations. In fields such as climate science, researchers are embarking on global-scale simulations at spatial resolutions that would have been unthinkable just a few years ago. Although this has expanded research possibilities, it has also put a huge strain on data storage requirements—a single large-scale simulation can produce many terabytes of data that must be stored and accessed for it to be useful.

"The volume of data produced by these simulations is becoming so large that it is not possible to store it in computational facilities without incurring high costs," explained Genton. "We see this not just in climate science but also in astronomy, where new telescopes are capable of capturing large images at unprecedented resolution, and in engineering applications, where computer simulations are performed at very high-

resolution in both space and time."

With almost no prior research on compression methods for climate data, Genton, in collaboration with Stefano Castruccio from Newcastle University (U.K.), worked to formulate compression principles that would be effective for these types of simulations. Using previously unexplored statistical methods, the pair identified a compression framework that departs significantly from the approach of existing generic compression methods.

"The statistics of the output data provide flexible and useful parameters for compressing the very large data sets that arise in climate science," said Castruccio. "We essentially gave up on the idea of compressing information bit-by-bit and instead developed a method to compress the physics of the climate model itself."

The researchers considered a typical annual three-dimensional model of the global temperature field (see image). By explicitly accounting for the space–time dependence and gridded geometry of the data, the compression scheme reduces the data to a set of estimated parameters that summarize the essential structure of the simulation results. Optimizing the compression scheme for each type of simulation can compress the data to as little as 2 percent of its original size.

"The next step for us is to provide the climate community with easy-to-use compression software that does not require technical knowledge in statistics," Genton said.