Session: C21 Poster: T247A

Exascale data archive: A new way of processing high-resolution climate data

Frederic Laliberte[+]; Paul Kushner
[+] University of Toronto, Canada
Leading author: frederic.laliberte@utoronto.ca

Query-based data analysis for current and planned distributed CMIP archives will need to satisfy an increasingly diverse array of research applications with very different data handling requirements. These will range from straightforward assessment of climatological metrics on standard pressure surfaces to more involved calculations that require an extensive knowledge of the underlying models. For the latter, using data converted from the model's grid onto a regular pressure/lat-lon grid is often too lossy and inaccurate. Therefore, each model's data needs to be handled differently, with each model having its own optimal way of completing a given mathematical query. As part of the development of the ExArch project, we propose a server-side computational approach that allows users to request data in nonstandard format directly and that allows each modeling group to design a processing algorithm that is the most consistent with the model's specifications. To illustrate the benefits of this approach, we describe a recent climatological diagnostic that computes the zonal and temporal mean mass flux joint distribution in terms of moist and dry entropy (potential temperature). Because this diagnostic serves to capture midlatitudes eddies, it requires daily sampled three dimensional data, although the size of the output represents a two order of magnitude reduction from the input. The diagnostic in question was already computationally intensive on the CMIP3 archive and it will become even more challenging for the CMIP5 and beyond. We show that the computation of this diagnostic can be optimized by the ExArch query system and we compare the query system efficiency with a simpler approach where all the data are downloaded to a local server and processed locally. Because many other diagnostics require a similar level of accuracy and consistency in the data, it is expected that the query would simplify climate research areas. Monsoon variability, cyclone counting, snow cover sensitivity, just to name a few, all would benefit from being processed using techniques that take full advantage of the model design.