Geophysical Research Abstracts
Vol. 14, EGU2012-13136, 2012
EGU General Assembly 2012

Can the ExArch framework facilitate the computation of stormtracks statistics in a petabyte archive?

F. Laliberté (1), P. Kushner (1), M. Juckes (2), and S. Denvil (3)

(1) University of Toronto, Physics, Toronto, Canada (frederic.laliberte@utoronto.ca,paul.kushner@utoronto.ca), (2) NCAS British Atmospheric Data Centre, Harwell Oxford, England, UK (martin.juckes@stfc.ac.uk), (3) Institut Pierre Simon Laplace, Paris, France (sebastien.denvil@ipsl.jussieu.fr)

The identification of cyclones and the analysis of their associated dynamical characteristics is a computationally intensive process requiring high-frequency data (Wernli and Schwierz, 2006; Ulbrich et al., 2009). In a petabyte archive (the CMIP5, for example) large data requirements will impose severe restrictions on the efficient inter comparison of cyclone statistics. To overcome this hurdle and encourage innovative diagnostic tools, the ExArch project is developing a data processing framework that could make such advanced diagnostics processable on petascale and exascale archives.

To illustrate the benefits of adopting a server-side processing approach to climate diagnostics, we derive an estimate for the total time taken for the computation of our diagnostics, including both the data retrieval and the data processing steps. We show that the computation is generally dominated by the transfer time and we demonstrate that server-side processing eliminates this bottleneck.

To benchmark the server-side processing tools developed by the ExArch team, we compute cyclone diagnostics in the CMIP5 archive by employing a cyclone identification algorithm from Wernli and Schwierz (2006). The algorithm, based on contouring techniques, identifies minima in daily sea level pressure and creates a cyclone mask that we then use to compute the cyclone-averaged precipitation, surface potential temperatures, column-integrated transient EKE and column-integrated static energies. These last two quantities require full three-dimensional variables at a daily frequency and are therefore well suited for the benchmarking of a distributed processing framework. In this work, we show that by using a server-side processing cyclone statistics can be computed in a reasonable amount of time on a petascale and even an exascale archive.