

## ***Interactive comment on “An approach to enhance pnetCDF performance in environmental modeling applications” by D. C. Wong et al.***

### **Anonymous Referee #1**

Received and published: 23 November 2014

The authors do not provide sufficient information to reproduce the results of their work. What version of pnetcdf was used and what versions of supporting software such as MPI and Lustre were used? What version of the CMAQ models was used and is the pseudo-code used to perform the experiments publicly available? No reference was provided for either netcdf or pnetcdf libraries.

The data aggregation technique that the authors refer to as a novel new approach is in essence the same technique that pnetcdf, MPIIO and Lustre apply at lower levels in the software stack so the question becomes - why does doing this aggregation at a higher level in the software stack work better than it does at a lower level?

Total time to perform parallel IO includes both communication time and IO time but the authors make no attempt to separate these factors. In section 5.3 the authors claim

C2389

that increasing the data size on the processor which is responsible for I/O will translate into higher I/O rates. If this is the case why not just use the original serial approach in which one I/O processor is responsible for all of the data?

The pnetcdf library contains a number of IO interfaces, collective vs independent and synchronous vs asynchronous, and supports several techniques for data aggregation. The authors do not indicate which pnetcdf interfaces they are using, or whether they experimented with others.

Figures 3-8 are too small and too busy to convey any meaningful information, perhaps tables would be better?

---

Interactive comment on Geosci. Model Dev. Discuss., 7, 7427, 2014.