J005-003 Room: C513 Time: May 28 9:30-9:45

A management for enormous data output by AGCM (AFES) on the Earth Simulator

Mayumi Yoshioka[1], Satoru Shingu[1], Hiromitsu Fuchigami[2], Masayuki Yamada[2], Wataru Ohfuchi[3], Shouzou Yamane[3]

[1] ESC, [2] NIS, [3] IGCR/FRSGC

//www.es.jamstec.go.jp/

It is a big dream for people who want to know physical phenomena by computer simulation to calculate with a numerical model in high-resolution as much as possible. A large-scale supercomputer system is required for such a simulation: the Earth Simulator (ES) started service in March 2002 is one of the computer systems to realize an advanced simulation. The ES consists of 640 processor nodes (PN) with a cluster of 16GB shared memory parallel vector processors composed of 8 vector arithmetic processors (AP) connected by a single-stage full crossbar switch. The peak performance of each AP is 8 Gflops and the total number of processors is 5120, resulting 40 Tflops in total peak performance and 10 TB main memory capacity. AFES (AGCM For Earth Simulator) is a spectral-method atmospheric general circulation model developed for high resolution simulations on the ES. AFES has achieved a high-resolution simulation of T639 (20.8 km horizontal mesh at the equator), which is much higher resolution than major climate AGCM resolution of T42 (314 km horizontal mesh).

A highly resolved numerical simulation has actually come true on the ES, however, it is also a nightmare which leads to produce huge output data after calculation. In fact, an AFES simulation of T639L258 (528 vertical layers) produces a 1.8 GB single file of graphics for one variable as a snapshot. It is close to the limit size to draw a figure, and oversize for making an animation file using several snapshots because most of OS and graphical software on popular computer environments have limit to treat a single file up to 2 GB. Moreover, it is also hard to analyze huge output data of several variables. A climate simulation of T42L20 outputs data of about 4.8 GB for 6 variables at an interval of 6 hours for a year. If we make a data of 50-years ensemble simulation of T42L20, AFES produces 2.4 TB output data for 10 variables. Much greater size of output data will be produced in a higher resolution of AFES, so only a few users can use completely huge output data of highly resolved simulation result for the same output time interval or members of variable.

We encounter difficulties when we manage the data produced in a higher resolution simulation. It is reasonable to think that we reduce data by averaging, skipping or selecting some focus regions in a highly resolved simulation. It depends on the model resolution what reduction method is applied. To take an example, the following approach will be employed in the simulation of T213 (62 km horizontal mesh). One of the data reduction ways for T213 simulation is to save the data as statistically arranged variables in time for each grid to original 3-dimensional mesh. In this case, we have to decide variables to be reduced in output and time interval before starting a calculation. The way to save output variables for some selected focus regions will be employed for the highly resolved simulation of T639 without time reducing. Otherwise, 2 types of time series of output data will be made during calculation: One is the output of variables for global 3-dimensional mesh at a long time interval, the other is for a regional mesh at short time interval without any data reduction. Post processing of output data is also necessary for both of middle and high-resolution simulation by considering output file size.

We have to make a plan carefully for the management of output data before executing simulation in high-resolution. The problem we have to consider is a strategy for managing processing sequences of enormous output data produced by a high-resolution numerical model on a huge computer system such as the ES. In the presentation, we would like to show how much enormous data is going to be produced by AFES and how to manage such enormous data on the ES.