

地球シミュレータ上における大気大循環モデル(AFES)が生み出す大規模データの処理

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

吉岡 真由美[1], 新宮 哲[1], 淵上 弘光[2], 山田 将志[2], 大淵 濟[3], 山根 省三[3]

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

[1] 地球シミュレータ, [2] NEC 情報システムズ, [3] 地球フロンティア

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

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

可能な限り細かい解像度で物理現象のシミュレーションをコンピュータを利用した数値計算で再現したいという願望は、常に計算機による数値シミュレーションを行っていく上で限りがない。平成14年3月から運用を開始したベクトル型超高速並列計算機「地球シミュレータ」により、これまでにない規模での数値シミュレーションが可能となった。「地球シミュレータ」は、1台あたり8個のベクトル演算プロセッサ(AP)と16GBの共有メモリから成る640台の計算ノード(PN)を単段クロスバで結合し、システム全体として、5120台のAP、主記憶10TBをもち、そのピーク性能40Tflopsに達する。実際、地球シミュレータ用に開発された全球大気大循環モデルを用いて、水平空間解像度T639(赤道上格子点間隔20.8km)の全球規模の大気のシミュレーションが実現した。これは、従来気象分野の研究対象として汎用的に用いられてきた空間解像度T42(赤道上格子点間隔314km)のものよりもはるかに細かい。

しかし、このようなシミュレーション用モデルの高解像度化に伴い、付随するデータが大規模化し、データ処理の問題が浮上している。例えば、出力データを可視化することを考えても、大部分のOSや、ユーザインタフェースとなるマシンで動いている可視化ソフトでは1ファイル2GBを越えたデータを取り扱うのが困難である。実際、前述の水平解像度T639の全球大気大循環モデルで鉛直256層をとったシミュレーションでは、1変数、1snapshotで約1.88GB近い画像用データが生成されているため、限界にきている。この規模では、従来のユーザが慣れ親しんだソフトで、アニメーションなどを行う機能があっても実現不可能である。可視化画像を作るのに、ユーザが実行前に考える必要がある。

モデルの高解像度化は、結果の可視化だけでなく、データの解析に必要な生データについては更なる深刻な問題である。従来用いられてきた空間解像度が低解像度(T42)のものでも、シミュレーション結果を6時間ごとに1年間出力させれば、6変数程度であっても、約4.8GBといった膨大なデータ量となる。気候学的目的で50年積分を行い、10変数のアンサンブル実験を行うと約2.4TBとなる。また、気温・風速・比湿といった基本的な物理量5変数を取り、5日毎の統計量(平均、分散、共分散)を計算した場合、1年分のデータは約800MBとなる。50年積分で10変数のアンサンブル実験を行うと約400GBとなる。

同様のシミュレーションされた物理変数のデータの解析処理を行おうとする場合、高解像度化にともない、間引き・圧縮・切り出しなどの加工を否応なく選択しなければならない。この出力データの加工方法は、解像度と目的とする現象によって変わってくるものである。気象学上の目的からすると、空間解像度について中程度(T213, 赤道格子間隔62km)のものであれば、時間方向については統計量をジョブを実行しながら処理し、解析に使う加工データのみを残し、空間方向にはそのままの解像度で残すことで、データの巨大化を抑える選択となるであろう。そのためには、必要となる変数の絞込み、計算にあたっての統計に用いる時間の決定が事前に必要となる。T639程度の高解像度のものであれば、空間的には見たい現象が起こっている領域に絞り領域の切り出しを行い、時間間隔は細かいまま処理したいという方式になるであろう。またシミュレーションしている領域全体と切り出した領域との関係について調べたいという場合には、2つの時間間隔でデータを出力するようにソフトウェアの出力部を変更するか出力データを全領域と切り出し領域とにわけ、2つの時間間隔で処理して別々のファイルに収めるように工夫したポスト処理が必要となる。この場合、必要となる変数の絞込み、切り出す領域の大きさや位置の決定、適正な時間間隔の決定が事前に必要となる。また、中・高解像度どちらの場合にも、処理するシステムの容量について熟慮した適切なファイルサイズとなるように出力データの1ファイルサイズを決めなければならない。

数値計算の巨大化、計算機システムの巨大化に伴う大規模データの取り組みに真剣にならなければならない時代は現実となった。これまでのように、誰でもが利用している実用可能な解像度でとりあえず目的の現象をシミュレーションすることを試みるため計算機にジョブを放り込んで、終了してからデータの処理を行う、といった計画性がない数値シミュレーションではやっては行けなくなっている。ユーザは、計算開始前に計算後の処理手順を考え、きちんとしたビジョンをもって実行することを要求されるということである。

発表では、地球シミュレータ上での大気大循環モデルによるシミュレーションの一連の作業の流れを例にとり、

地球シミュレータを用いたデータ処理の現状と、大規模データに関する対処法、今後の問題点などについて述べていく予定である。