

HPC とともに進化する大規模データ同化

樋口知之

情報・システム研究機構 統計数理研究所

[アブストラクト]

複雑な現象の高精度予測のために、時空間観測・計測データの解析と数値シミュレーション計算を統合することにより、シミュレーションの初期値や境界値、パラメータ等を実際の現象をなるべく再現するように決め、時にはシミュレーションモデル自体にも手を加える（リモデリング）一連の計算作業がデータ同化と呼ばれるものである^(1,2)。データ同化は逐次型と非逐次型に大別されるが、我々データ同化研究グループ⁽³⁾はアンサンブルベースの逐次型の研究に力を注いでいる。大規模なデータ同化の実現においても、計算効率を高めるために HPC のハード特性を最初から意識した手法の開発が肝要となる。本講演では、データ同化の概念と基本的なアルゴリズムを概説するとともに、次世代スーパーコンピュータ等の HPC の階層性とネットワーク構造を十分意識した逐次データ同化の研究開発の現況⁽⁴⁾を紹介する。

[キーワード]

データ同化、融合シミュレーション、アンサンブル予測、粒子フィルタ

[講演要旨]

1. はじめに

データ同化は、シミュレーションのような演繹的な推論と、統計科学に代表される帰納的な推論を融合するためのプラットフォームである⁽¹⁾。データ同化の問題は、シミュレーションによるシステムの時間更新をシステムモデル、シミュレーションに内在する諸変数のさまざまな観測法による観測を観測モデルと定式化することで、制御工学、音声解析、統計科学等において 30 年ほど前から研究されてきた状態空間モデルの一般化版として定式化でき⁽²⁾。統計数理研究所は、長年、状態空間モデルにもとづく時系列解析の研究に関して、日本において世界的レベルにある数少ない研究機関の一つであり、データ同化の先端的計算法の研究においても多大な貢献をしてきた。

データ同化には大別して二つの流儀があるが、私たちデータ同化グループ⁽³⁾では、逐次データ同化とよばれる同化手法の研究と応用を行っている。逐次型としてはアンサンブルカルマンフィルタが主に用いられているが、粒子フィルタの適用も少しずつなされてきている。我々データ同化グループでは、このアンサンブルカルマンフィルタ及び粒子フィルタの二

つを基盤的計算技術と位置づけ、逐次データアルゴリズムの高度化とその応用研究を行っている。一方、現業機関（気象庁）では、利用するシミュレーションモデルがあまりにも巨大であるため、非逐次型データ同化法を採用している。非逐次型データ同化手法としては、4次元変分法、とくにアジョイント法がよく知られている。

2. 世界最大規模の計算

実際にデータ同化手法の研究を行うには、具体的なシミュレーションモデルとデータセットの二つ、つまり具体的テーマの選定が必要である。データ同化においては、シミュレーションモデルが含む変数の次元が数百次元から一千万次元程度に、また観測の次元も数十から数十万となるため、私たちは計算の限界に挑戦しなくてはならない。従って計算には世界最先端クラスのスーパーコンピュータが必要なことは言うまでもない。一つの計算に従前のスパコンが必要であるならば、データ同化専用計算機は数千のスパコンが必要となるイメージである。本研究グループは、次世代スーパーコンピュータの研究開発プロジェクトにも、生命体統合シミュレーションのためのデータ同化技術の開発を目標に参画している。

[参考文献]

- (1) 樋口知之、統計数理は隠された未来をあらわにする：ベイジアンモデリングによる実世界イノベーション（監修・執筆）、東京電気大学出版局、2007.
- (2) 中村和幸、樋口知之、最近のベイズ理論の進展と応用[II]ー逐次ベイズとデータ同化ー、電子情報通信学会誌， 92-12， 1062-1067， 2009.
- (3) <http://daweb.ism.ac.jp/>
- (4) S. Nakano and T. Higuchi, A dynamic grouping strategy for implementation of the particle filter on a massively parallel computer, Proceedings of 13th International Conference on Information Fusion, 2010.