

研究課題別事後評価結果

1. 研究課題名： 先端的データ同化手法と適応型シミュレーションの研究

2. 研究代表者名及び主たる研究参加者名 (研究機関名・職名は研究参加期間終了時点):

研究代表者

樋口 知之 (情報・システム研究機構統計数理研究所 副所長・教授)

3. 研究実施概要

シミュレーションなどの数値モデルによる対象状態の時間発展更新と、装置からの部分的な観測量に基づく状態補正の二つを適切に組み合わせる作業はデータ同化と呼ばれる。本研究では先端的なデータ同化手法を開発し、さらにこの技法をもとにシミュレーションモデルを複数走らせ、データ適応的に数値モデルを切り替え、あるいはそれらを統合するようなメタシミュレーションモデルを創出するプラットフォームの創造を最終目標とした。シミュレーションモデルが含む変数の次元は数百次元から 100 万次元程度の超大規模なものであり、データ同化の研究はいろいろな意味で計算限界への挑戦である。データ同化は逐次型と非逐次型のアプローチがあり、逐次型においてはアンサンブルカルマンフィルタなどが主に用いられているが、粒子フィルタの適用もわずかながらなされていた。本研究の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た。

手法の観点からは総括して二つの新しいアルゴリズムを開発した。一つは、融合粒子フィルタと呼ぶ粒子フィルタの変形版であり、大規模なシミュレーションモデルを用いたデータ同化システムに対しても、境界条件や初期条件、パラメータ値等のシステムが内包する不確実性を分布としての確に表現しながら、現実的な計算時間内でデータ同化することが可能となった。二つ目は、モデルの系統的比較の基礎となる尤度計算に関して、大規模なデータ同化システムでも数値的に安定して計算可能となるアルゴリズムである。これにより情報量規準にもとづくモデル比較およびモデルを混合創出するメタシミュレーションの基礎を築けた。

アルゴリズムの開発においては、計算機環境の変化にも常に注視しておくことが大切である。近年、プロセッサのマルチコア化が劇的に進んでおり、GPGPU (General Purpose computation on Graphical Processing Units)、マルチコア等の 100 を超えるスケールの並列プログラミング環境もデスクトップパソコンで実現されつつある。粒子フィルタを高並列計算に適した形で適切に実装することで、次元の呪いにより従来は困難とされている問題を超並列計算機システムでシンプルに解決することにも取り組んだ。その結果、GPGPU から次世代スーパーコンピュータまで、さまざまな計算機アーキテクチャ上で効率よくデータ同化を行うための知見と経験が豊富に得られた。

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。そのテーマは、手法の新展開の観点から未解決の数理的及び数値的課題を内包し、併せて応用分野における知識発見や予測性能向上の観点からも興味深いものでなければならない。この観点から本研究では、大気・海洋、津波、海洋潮汐、宇宙空間、ゲノムデータ同化等のさまざまな領域における具体的テーマに取り組んだ。さらには“ものづくり”やサービスサイエンス等、これまで現場熟練者の経験と勘が特に重要であった分野に対して、生産性向上策の提案をデータ同化により効率的に実現する研究も推進した。ゲノム情報分野において世界で初めてデータ同化の概念と方法論の有効性を明らかにしたことと、今後の“ものづくり”にはデータ同化の視点が欠かせないことを啓発した二点は、応用分野の新規開拓の観点から最も成功した事例と言える。

以上のように本研究では、新規アルゴリズムの開発といった理論研究と、新しいデータ同化研究分野の開拓と概念の啓発となる応用研究をバランス良く遂行できた。

4. 事後評価結果

4-1. 研究の達成状況及び得られた研究成果(論文・口頭発表等の外部発表、特許の取得状況等を含む)

新しく開発した非線形非ガウスデータ同化アルゴリズムにより、大規模なシミュレーションモデルを用いたデータ同化システムに対しても、境界条件や初期条件、パラメータ値等のシステムが内包する不確実性を分布としての確に表現しながら、現実的な計算時間内でデータ同化することが可能となった。また、モデルの系統的比較の基礎となる尤度計算に関して、大規模なデータ同化システムでも数値的に安定して計算可能となる新規アルゴリズムを開発し、情報量規準にもとづくモデル比較および、モデルを混合創出する、メタシミュレーションの基礎を築けた。さらにこれら新しいアルゴリズムの開発とともに、データ同化を様々な応用分野に適用しその有効性を実証した。研究開始時においてはデータ同化の主たる適用分野が大気・海洋中心に限定されていたが、宇宙空間、ゲノムデータおよび“ものづくり”やサービスサイエンスまで応用を拡張し、当初は予想できなかった分野にまでもデータ同化が有効な手法となることを示したことは高く評価できる。

各分野へのデータ同化の適用のためには、その分野の内容をよく把握しなければならず、外部発表にはかなりの準備期間が必要である。そのことを考えると、単独研究グループで構成した研究チームの大きさを鑑みれば論文数や発表件数は非常に良好であると判断する。統計分野だけでなく広い応用分野のレベルの高い雑誌に発表しており、論文のいくつかは研究者コミュニティの大きな注目を浴びている。一例として宇宙・地球科学で最も権威のある雑誌群を刊行している AGU (American Geophysical Union) が毎週報告している **This Week's Most Downloaded Articles** において、16 雑誌中の全論文の No.1 にランクされた。また国内外を問わず招待講演の数が年々増えており、本プロジェクトのデータ同化手法の認知度は増加の一途をたどっている。またデータ同化手法の入門解説書や教科書の刊行やデータ同化の公開講座の開講など、学術コミュニティへの貢献度も大きい。

本プロジェクトはサブチーム(研究グループ)を置かず、研究代表者が全ての研究メンバーを監督するフラットな研究チーム構成をとっており、研究代表者の一元的な研究チーム運営のもとリーダーシップが発揮され、研究の進捗状況に柔軟かつ弾力的に対応しながら研究プロジェクトを推進した。その結果、手法の観点と応用領域の観点の 2 軸を組み合わせた研究体制もうまく機能し非常に広い問題にデータ同化手法を適用できた。

研究費の執行に関しては、主に人件費が主たる経費であり人材育成を中心に力を入れ、研究代表者とその周辺の研究メンバーに留まっていたデータ同化という手法とその応用を各種分野の研究者と研究グループに拡大した。その結果、短期間ながら一部は産業界まで連携する状況になったことは分野の桁違いの拡張と捉えられ高く評価できる。

4-2. 研究成果の科学技術や社会へのインパクト、戦略目標への貢献

世界で初めて、アンサンブルベースの逐次データ同化手法を大気・海洋結合モデルに適用した成功例を示した。磁気圏領域のグローバルモデルにおいて逐次データ同化を適用した成果も他に前例を見ない。また、超高次元の状態空間モデルに対する粒子フィルタを新規に開発し、大規模なシミュレーションモデルを用いた逐次データ同化手法の未踏の領域を拓いた。さらに本研究では逐次データ同化手法の特徴である、実用性と汎用性の高さに早くから注目し、新しい適用分野を先導的に次々と開拓した。特に、生物学的知見にもとづく数値モデル(イン・シリコモデル)を統計モデルに援用することで、統計的推測と生物学的知見への直接的な橋渡しを担う、ゲノム情報分野における基盤技術を創生したことは大変な功績である。“ものづくり”データ同化プロジェクトにおいては、製造業の製品性能の向上や均一化のための作業工程の効率化にデータ同化を用い大きな効果を上げている。

以上に示すように逐次型データ同化手法の高度化により、広い分野のシミュレーション研究に重要な貢献をした。そのことにより、多くの分野においてデータ同化が今後の重要な手法であることを認識させることとなった。

本研究成果のさらなる展開は学術、産業の両方から以下のような動きがあることから大いに期待できる。学

術分野においては、数学・数理科学やシミュレーション科学の新たな飛躍の端緒を求めて招待講演の依頼数が増えている。産業界においては、ものづくり分野を中心としてデータ同化により製品開発プロセスの効率化を向上したいという期待感が多数寄せられている。また、本プロジェクトによるゲノム情報分野へのデータ同化適用の成功を受けて、本研究メンバーの多くは、次世代スーパーコンピュータ上での利用を目指した「次世代生命体統合シミュレーションソフトウェアの研究開発」にも参画しており、本プロジェクトの成果のコアの部分は継続的に発展されていくであろう。

データ同化手法は以下に示すように当領域の戦略目標である原子・分子レベルの現象に基づく精密製品設計・高度治療実現のための次世代統合シミュレーション技術に不可欠であり、貢献することが大いに期待できる。

一般的なシミュレーション技術のアプローチと異なり、初期条件・境界条件・パラメータ等データの最適化と計算資源の集中化の研究を行うことで、逆問題解析の革新的アルゴリズムの開発に貢献した。シミュレーションの解の評価にあたっては、確率分布の視点でシナリオ全体を考えることにより、時間・経費節約を勘案しつつ情報抽出力性能の高い計測（観測）システムの構築策を提案した。このことにより、多品種少量生産を基本とする製品開発現場でのステップの簡略化や、患者一人一人に合った治療サービスの提供と期間の短縮化等の、シミュレーション技術の個人化技術への発展的進化に貢献できると期待している。

4-3. 総合的評価

逐次データ同化手法の研究を行っているのは国内においては本研究チームの数人のみといった状況でプロジェクトを開始したが、日本国内において先導的な立場で研究推進しデータ同化手法シミュレーションの方向付けをしてきたことを高く評価する。

特に粒子フィルタの根源的問題を回避しつつ超高次元の具体的問題への応用の筋道をつけたこと、逐次データ同化手法の未踏の領域を拓き新しい魅力あるデータ同化適用分野を先導的に多数開拓した点など十分評価に値する。

新聞のサイエンス欄でデータ同化手法の鍵となるベイズ統計に関する談話が紹介され、またビジネス情報を提供するテレビ番組のハイライトコーナーにてデータ同化手法の研究内容が紹介されるなど関係する研究者だけではなく一般社会にも本研究の重要性が浸透しつつあり今後のさらなる研究の発展が期待できる。