

研究報告書

「次世代データ同化:自動モデル化と情報フロー抽出技術開発」

研究タイプ:通常型

研究期間:平成21年10月～平成25年3月

研究者:上野 玄太

1. 研究のねらい

大規模データから有用な知識を得るためには、よいモデルを立てること、さらにモデルを通して得た情報の中から必要なものを抽出する必要がある。本提案は、データとモデルを組み合わせるデータ同化手法を基盤とし、

研究テーマ A「省コストのデータ同化手法の開発」

研究テーマ B「モデルが捉えた情報の流れの抽出技術」

研究テーマ C「モデルの高性能化(予測・推定)技術」

を系統的に開発するものである。これらにより、従来は専門家の経験が必要とされていたモデルのチューニングや実行環境に応じた再モデル化、入力データ量を凌ぐ大量のモデル出力の動的な解析が可能になる。

2. 研究成果

(1)概要

研究テーマ A「省コストのデータ同化手法の開発」

データ同化の中心部であるシミュレーションモデルは、微視的な物理・化学法則に基づいて構成されていることが多く、データ同化のために多数のメンバーからなるアンサンブルを構成することは、計算機能力の観点から難しい。アンサンブルメンバーが少ないことによる難点は、メンバー間の相関を求めた場合に、見かけ上の偽の相関値が得られるという点である。そこで、そういった偽相関を抑える正則条件を組み込んだデータ同化手法を開発することとした。ここでの正則条件として、シミュレーションモデルの変数間の依存関係に着目し、アンサンブルメンバー間の相関行列の逆行列に明示的に構造を入れることとした。この構造をもとにして、逆行列を基準としたデータ同化手法(グラフベース・アンサンブルカルマンフィルタ)を開発した。

研究テーマ B「モデル変数の感度計算法の開発」

データ同化が完了しても、具体的にどの変数がどのように機能しているかを見出す作業は依然として容易ではなく、専門家の知識が必要となるのが現実である。しかし、変数の関係はすべて既知であるのだから、原理的にはモデル内に解答があるはずである。そこで、データ同化による推定値に対するモデル変数の感度を得る方法を開発することとした。微分係数のアンサンブル近似の定式化を行い、各時間ステップでの微分係数値を求める方法を考案した。巨大次元の変数への対処が課題であったが、特異値分解等の線形計算を組み合わせることで、アンサンブルメンバー数だけの手間で済むことを示した。

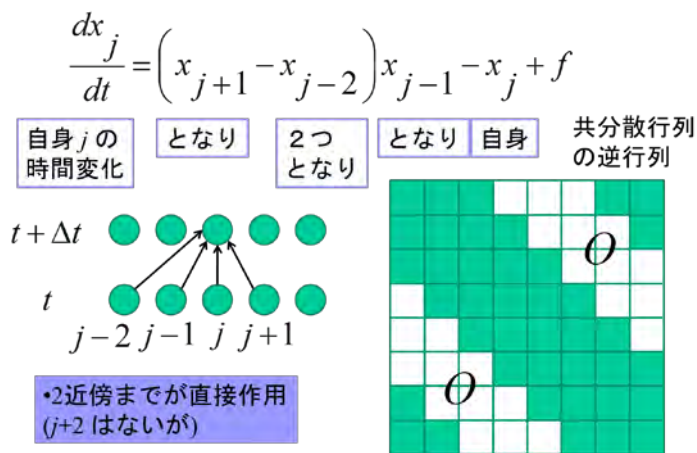
(2)詳細

研究テーマ A「省コストのデータ同化手法の開発」

データ同化におけるシステムモデルの核となるのは、モデル変数の時間更新を表現する部分である。すなわち、あるタイムステップにおける変数を与えたもとの、次ステップでの変数値を推定する部分である。この時間更新モデルには、自己回帰モデルなどの応用分野をそれほど意識しないものも使われるが、応用分野での知見にもとづき開発が進められてきたシミュレーションモデルがあるならば、その分野における有用性や予測能力という見地から、そういったシミュレーションモデルを時間更新部分としてそのまま用いることが望ましい。しかし、それらのモデルは微視的な物理・化学法則に基づいて構成されていることが多く、データ同化のために多数のメンバーからなるアンサンブルを構成することは、計算機能力の観点から難しくなる。

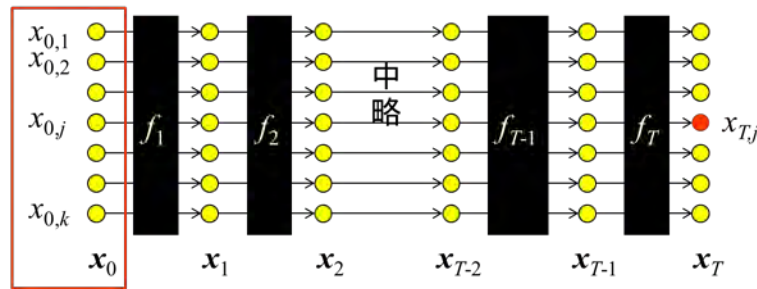
アンサンブルメンバーが少ないことによる難点は、メンバー間の相関を求めた場合に、見かけ上の偽の相関値が得られるという点である。そこで、そういった偽相関を抑える正則条件を組み込んだデータ同化手法を開発することとした。従来は、モデルの構造には着目せず、単に偽相関をテーピングすることで抑える方法がとられているが、テーピングの強度を調整するパラメータの選択に難点があることが知られていた。

そこで、正則条件として、シミュレーションモデルの変数間の依存関係に着目し、アンサンブルメンバー間の相関行列の逆行列に明示的に構造を入れる、新しいデータ同化手法を開発した。この方法により、モデル構造を直接反映した正則化が可能になり、従来のアドホックなテーピングが不要となった。グラフ構造を仮定したアンサンブルカルマンフィルタととらえることができるので、グラフベース・アンサンブルカルマンフィルタ (Graph-based ensemble Kalman filter, GEnKF) と呼ぶこととする。小規模なシミュレーションモデル(40 変数)をもとに、GEnKFの実装を行い、従来の同化手法(アンサンブルカルマンフィルタ、粒子フィルタ)よりも少ないメンバー数でも真値に近い推定値が得られることを確認した。



研究テーマ B「モデル変数の感度計算法の開発」

データ同化の実施により、モデルの枠組みで現象を理解することになる。これは、モデルに含まれる変数の関係が既知であるという点でデータのみの状況よりも大きく前進しているが、具体的にどの変数がどのように機能しているかを見出す作業は依然として容易ではなく、専門家の知識が必要となるのが現実である。しかし、変数の関係はすべて既知であるのだから、原理的にはモデル内に解答があるはずである。



問題 最終状態 x_T の j 成分 ($x_{T,j}$ と書く) に多く寄与するのは、初期状態 x_0 のどの成分か (もしくは成分群か) ?

そこで、データ同化による推定値に対するモデル変数の感度を得る方法を開発し、モデル内の情報の流れを可視化することを考える。モデルは通常、巨大次元である上に数多くの演算をし、非線型のプロセスを含むため、単純な数値微分では有効な感度を得ることができない。アジョイント法もしくは自動微分と呼ばれる計算技術が一部で使われているが、そのコーディングと保守の困難さが問題である。

そこで、複数の異なる入力変数を与えたモデルからなるアンサンブルを用いることで、感度の近似計算法を開発した。基本的なアイデアは、状態のアンサンブルを用いて数値微分を各時間ステップで計算し、それを微分のチェーンルールに照らして掛け合わせていくというものである。具体的には、微分係数のアンサンブル近似の定式化を行い、各時間ステップでの微分係数値を求める方法を考案した。巨大次元の変数への対処が課題であったが、特異値分解等の線形計算を組み合わせることで、アンサンブルメンバー数だけの手間で済むことを示した。目標となる最終時点 T の特徴的な変数 (図で示す $x_{T,j}$) から、時間をさかのぼる形式 ($T-1, T-2, \dots, 2, 1, 0$) で感度が逐次的に求められる。課題Aと同じ小規模な 40 変数のシミュレーションモデルに対して検証を行い、8 メンバー程度のアンサンブルでも妥当な感度が得られることを示した。

3. 今後の展開

研究テーマ A, B で開発したグラフベースアンサンブルカルマンフィルタ (GEnKF)、アンサンブル感度計算法を組み込んだ、大規模なシミュレーションモデルをもとにしたデータ同化システムを構築する。一つは、気象庁による数値予報モデル (非静力学モデル) を用いたシステムを考えている。計算負荷が比較的小さい (50 万変数) 設定での、温帯性低気圧の予報から始めたい。もう一つは、気候変動のアンサンブル予測である。GEnKF により、少ないメンバーでの分布表現が可能になるため、頻度は小さいが被害が甚大な事象の予測を効率的に行うことができるとともに、アンサンブル感度計算法により、温暖化の原因変数の推定を少ない手間で行うことができる。

研究途中段階であるテーマ C については、モデルで説明できない部分と入力変数のモデル化について、引き続き考察と実験を進めていきたい。

4. 自己評価

研究テーマ A、B に関しては、GENKF およびアンサンブル微分法の開発により、解決されたものと考えている。

研究テーマ C については、もとより容易ではない課題であると認識していたが、研究途上段階で期間の終了を迎えてしまった。ただし、その過程で得られた、モデルの高性能化のために、説明すべき変数と入力変数を明確に抽出できたこと、それらのモデル化が必要であるという認識が得られたことは有意義であると考えている。データそのもの、モデル出力そのものを考えるのではなく、誤差の影響も考慮したそれらの組み合わせから得られる変数を出発点とした、第 2 ステージのモデル化対象を明確にできたものと考えている。

5. 研究総括の見解

データとモデルを組み合わせることにより、モデルの高性能化(予測・推定)技術、モデル作成の省コスト化技術、モデルが捉えた情報の流れの抽出技術を系統的に開発するという研究課題であった。「省コストのデータ同化手法の開発」、「モデルが捉えた情報の流れの抽出技術」、「モデルの高性能化技術」の3テーマに分けて取り組んでいる。

「省コストのデータ同化手法の開発」と「モデルが捉えた情報の流れの抽出技術」については、シミュレーションモデルをもとにしたデータ同化システムやアンサンブル微分法等の開発により解決している。「モデルの高性能化技術」については課題達成に至らなかったが、その過程で得られた、モデルの高性能化のために、説明すべき変数と入力変数を明確に抽出できたこと、それらのモデル化が必要であるという認識が得られており、有意義な成果と考えられる。今後の継続研究に期待したい。

6. 主な研究成果リスト

(1) 論文(原著論文)発表

- | |
|--|
| 1. Ueno, G. and N. Nakamura, Iterative algorithm for maximum likelihood estimation of observation error covariance matrix for ensemble-based filters, Quarterly Journal of the Royal Meteorological Society, in press. |
| 2. 上野玄太、データ同化、システム制御情報学会誌、印刷中。 |
| 3. 上野玄太、データ同化における大規模グラフィカルモデルの推定について、統計数理、印刷中。 |
| |
| |

(2) 特許出願

研究期間累積件数: 0 件

(3) その他の成果(主要な学会発表、受賞、著作物、プレスリリース等)

1. Ueno, G., and T. Tsuchiya, Regularization of error covariance with the Gaussian graphical model, 14th Symposium on Integrated Observing and Assimilation Systems



- for the Atmosphere, Oceans, and Land Surface (IOAS-AOLS), The 90th American Meteorological Society Annual Meeting (Atlanta), January 20, 2010.
2. Ueno, G., An iterative algorithm for estimating the observation error covariance matrix for ensemble-based filters, EGU General Assembly 2011 (Vienna), April 8, 2011.
 3. Ueno, G., Derivative-Free Estimation of the Observation Error Covariance Matrix for Ensemble-Based Filters, 16th Symposium on Integrated Observing and Assimilation Systems for the Atmosphere, Oceans, and Land Surface (IOAS-AOLS), The 92nd American Meteorological Society Annual Meeting (New Orleans), January 25, 2012.
 4. Ueno, G., A Bayesian approach for estimating stable observation noise covariance, EGU General Assembly 2012 (Vienna), April 26, 2012.
 5. Ueno, G., Stable Estimation of Observation Error Covariance Matrix by Bayesian Method, 17th Symposium on Integrated Observing and Assimilation Systems for the Atmosphere, Oceans, and Land Surface (IOAS-AOLS), The 93rd American Meteorological Society Annual Meeting (Austin), January 8, 2013.