

# 研究終了報告書

## 「データ同化モデリングの自動化原理開発によるハイレベル予測発見手法の構築」

研究期間：2017年10月～2021年3月

研究者：中村 和幸

### 1. 研究のねらい

データ同化は、シミュレーションと計測データを融合して新たな知見を発見したり予測したりすることにつながる有効な手法である一方、その適用には、問題ごとにオーダーメイドで実施する、「データ同化を行うためのモデル化・定式化」や「アルゴリズムの選択」が不可欠であった。このことは、諸分野の個別問題にデータ同化を適用する際の敷居の高さを生む原因となっていた。その中でも、問題の特性に応じたアルゴリズムの選択、モデルの修正、不確かさの計量をどうすべきであるかという点は、職人芸的な部分が大きく、特に高いハードルとなっていた。

本研究では、数理学・統計科学・データ同化の3つの分野における手法を融合した手法を開発することで、これらの問題を解決する枠組構築を目指す。これにより、各分野の研究者が障害を感じることなく、時空間計測データから高度なデータ同化を実施できるようになることを目的とする。具体的には、

1. 数理とデータ同化の手法を融合した局所ノイズ感受性解析手法によるアルゴリズム選択
2. 数理学と統計科学の手法を融合した機能モデル・エミュレータ構築
3. 統計とデータ同化の手法を融合した不確かさ解析手法による事前分布デザイン

という各分野の融合でありかつ定量的な評価を伴った3手法をデータ同化のための要素技術として再構成する。これにより、解析において目指す精度を決めると、手法を定量的な基準に基づいて適用・選択可能となるところまで一般化できることとなり、上記の問題の解決に寄与できる。

さらに、実問題では計算速度と解析精度の間のトレードオフ調整の必要があるため、上記3手法を調整しながら高次に統合する仕組みの構築も目指す。得られた手法は、共同研究という形で展開する。

以上の開発の実現により、時空間計測が得られた際に、諸分野におけるデータ同化において、ほぼ自動的に高度なデータ同化を実施可能となることを目指す。これは、従来データ同化の適用へのハードルが高かった分野に対して、自由にデータ同化を適用できるようになり、対象となる分野が爆発的に広がることにつながる。すなわち、新たな計測データを得られたさまざまな分野での潜在知識発見や予測を可能とすることになり、諸学問の発展に寄与し限界を突破するようなブレークスルーを与えることになる。

## 2. 研究成果

### (1) 概要

数値シミュレーションと計測データを融合するデータ同化の自動化原理開発に向け、local translation error (LTE) 解析のデータ同化手法への展開、データからの機能モデル構築、不確かさの設計・事前分布の検討、ベイズ統計手法による各要素項目の統合、実問題への応用について進めた。

LTE 解析については、LTE 解析によるシステム分析と、そのデータ同化への応用について検討を進めた。シミュレーションモデルを用いた検討の結果、LTE 解析により時間局所的な非線形性を得られるということ、推定における安定性の情報を得られることの確認ができた。このことは、観測ノイズレベルも考慮したデータ同化アルゴリズムの選択・切り替えにおいて、LTE 解析の結果が有効であることを意味している。これにより、将来的なデータ同化アルゴリズム切り替えのためのツールの一つとなることの示唆が得られた。

機能モデル構築については、時空間局所性を有効に活用した、データからの機能モデル構築を行った。機能モデルは、データ同化に必要なシミュレーションモデルを陽に構築しにくい場合に、代替となるシステムモデルをデータから構築し、その後のデータ同化に用いるためのモデルである。時間局所的に定常で、空間局所的に一様な仮定をした場合の機能モデルである、locally and spatially uniform linear operator construction with the Kalman filter (LSLOCK) を構築した。数値実験の結果、短時間予測について、提案した機能モデルを用いることの有効性を確認できた。また、オペレータの推定結果についても検討した結果、適切な時空間パラメータの設定により、ダイナミクスを抽出できることが確認できた。

事前分布の検討については、生命科学モデルのような不確実性の高いモデルにおける設計と適用結果の検討を行った。その結果、モデルの不確実性を十分に表現できるシステムノイズの設定が必要であること、すなわち、不確実性を十分表現できないシステムノイズでは、定性的な状況の再現はできるが、定量的には不適切な推定結果となることが確認できた。

これらの要素項目のベイズ統計による統合については、整備手法の LSLOCK の実時間性について、問題のサイズを変更しながら検討を行い、問題スケールを変えながらその実時間性を確認した。

実問題への応用については、気象レーダデータ、生命科学モデル、高速 AFM データへの適用を進めるとともに、領域内外のデータ同化モデルへの、特に、気象レーダデータや擬似 AFM データへの適用において、これまでに整備した手法により、適切なフィルタ推定値が得られることの確認ができた。

### (2) 詳細

#### 研究テーマ1「Local translation error によるシステム解析手法のデータ同化への展開」

Local Translation Error(LTE)とは、時系列システムにおける非線形解析手法の一つであり、観測データのみに基づくシステム解析が可能である。そのため、データ同化アルゴリズム選択指針として、LTE 解析による分析が適用可能であると考え、検討を進めた。

LTE 計算ライブラリならびにこれを取り込んだデータ同化プログラムライブラリの作成を進め、データ同化分野で広く用いられる Lorenz 96 モデルに対して適用することで、データ同化への適用性について検討した。具体的には、観測ノイズレベルを変動させた場合の時間局所

的な非線形性の変化に対する LTE 計測結果について、数値実験により検討を進めた。その結果、Lorenz 96 モデルに対する時間局所的な非線形性の推定において、観測ノイズレベルを変えたときに、LTE 値が極大となる時点の変動することが確認できた。このことは、事後的・人工的に加える観測ノイズのレベルを変動させることで、データ同化の状態推定における安定性についての情報を得ることができることを意味している。また、観測ノイズのレベルが変動するようなシステムにおいて、ノイズレベルに適応的なアルゴリズム選択を実現できること、LTE 解析がシステムそのものの特性の推定につながることを示唆している。これにより、将来的なデータ同化アルゴリズム切り替えのためのツールの一つとなることの示唆が得られた。

以上の結果について国際学会において発表した。当初目標としたアルゴリズム選択の具体的な実装については、性能評価も含めて進めている段階である。また、これらとは別に LTE 解析の応用における有効性について共同研究を進め、成果について共著論文として発表した。

## 研究テーマ2「非線形情報解析と Singular Spectrum Analysis による機能モデル・エミュレータ構築」

機能モデル・エミュレータ構築については、当初 Singular Spectrum Analysis (SSA) と非線形情報解析を利用した手法による開発を進め、1次元時系列に対する非線形情報解析における時間遅れと非線形性をコントロールするパラメータ化ルールの検討、SSA による解析の検討を進めた。しかし、実際のデータ同化における機能モデル構築においては、その使用場面を考えた場合にリアルタイム性が重要であることから、数理的には SSA と同様の情報を持った機能モデルを構築できる手法であり、かつ時間局所性とオンライン学習の枠組みを有効活用した linear operator construction with the Kalman filter (LOCK)ならびにその拡張手法により、機能モデル・エミュレータ構築を行うこととした。拡張手法として、グローバルな空間一様性を用いた spatially uniform LOCK (SLOCK)、空間局所性を用いた local LOCK (LLOCK)、この両者の考え方を組み合わせた局所空間一様性を用いた LSLOCK を開発した。

開発したアルゴリズムの有効性について、時空間局所的に物体の移動が発生する人工データ、ならびに局所非線形性のある同心円状の波伝播の人工データに対して、LSLOCK を用いた状態ベクトル・機能モデルの同時推定を行って検討した。その結果、得られたフィルタ推定値について、定常なカルマンフィルタ推定よりも状態推定結果が得られた。

さらに、機能モデルの推定も得られることから、シミュレーションデータに対して、LSLOCK によって得られた機能モデルをもとにした短期予測を行った。その結果、定常カルマンフィルタ推定よりも良い予測精度が得られただけでなく、同様のタスクにおいて用いられることが多い深層学習ベースの手法である、Convolutional LSTM と比較しても、良い予測精度が得られることが分かった。以上の成果についてとりまとめ、査読付き国際会議録に採録された。

さらに、得られた機能モデルをもとに、局所的なダイナミクスの情報抽出も可能であることが確認できている。これにより、単に機能モデル・エミュレータを構築するだけでなく、その時間空間的な構造の抽出も可能となった。

## 研究テーマ3「データ同化のための誤差モデルの構築」

データ同化における誤差モデルの考え方を整理し、網羅的な適用を通じてどの分布が有効

であるか、定量的な予測精度の観点から、モデルの特性と合わせて検証することを予定としていた。検討の過程において、具体的な実問題を通じて検討することとし、糖代謝モデルに対するシステムノイズの影響に関する評価を進めた。検討の結果、代謝モデルにおいてデータ同化を行う際に、代謝中間体生成の過程に大きな不確実性があることから、モデルの不確実性を十分表現できるシステムノイズの設定が必要であること、すなわち、不確実性を十分表現できないシステムノイズでは、定性的な状況の再現はできるが、定量的には不適切な推定結果となることが確認できた。

また、シカ個体数推定モデルにおける事前分布構築の問題について、本項目の検討のための基礎となるとことから検討を進めた。シカ個体数推定モデルは、状態ベクトルの次元が数百程度である一方で、不確実性が大きく、また局所的な空間相関が強いモデルであるという特徴がある。検討の過程において、事前分布デザインと関連する推定アルゴリズムについて、ハミルトニアンモンテカルロ法による推定の有効性が明らかになったが、推定アルゴリズムの性質上、離散値パラメータの推定が困難であること、事前分布に工夫が必要であり、二項モデルをポアソンモデルで置き換えることでも有効な推定が得られるという結果を得た。本結果の一部について、査読付き国際会議録に採択された。

#### **研究テーマ4「研究テーマ1から3のベイズ統計学による融合」**

本テーマでは、各要素項目のベイズ統計による統合について、その計算時間や推定精度を通じて融合することを目的としていた。今回、要素モデルの進捗状況により、統合モデルとしての整備までは至らなかったものの、代替として、要素手法のLSLOCKの実時間性について、問題のサイズを変更しながらその実行時間検討を行った。検討の結果、LSLOCKの計算時間上の有意性が確認できた。

#### **研究テーマ5「従来の応用分野への展開、ならびに研究領域内の共同研究テーマへの展開」**

実問題への応用については、気象レーダデータ、生命科学モデル、高速AFMデータへのデータ同化手法の適用を進めるとともに、領域内外のモデルへの知見の展開を進めた。

まず、気象レーダデータについては、LLOCKならびにLSLOCKを適用し、適切な推定結果と想定されるケースがあることの確認が得られた。また、糖代謝モデルに適切な事前分布デザインを行うことで、数理モデルのパラメータの特徴抽出が可能であること、高速AFMデータの擬似データに対して適用して、有効である可能性を得た。さらに、領域内の複数のテーマとの連携を実施し、データ同化手法や関連する時系列解析手法の提案・展開を実現できた。

以上のことから、研究目的の達成状況については、機能モデル構築については、当初の想定とは異なる形ではあるが、データ同化のための機能モデル・エミュレータ構築を実現でき、さらに、ダイナミクス抽出にも成功しており、当初目標以上の成果を達成した。共同研究テーマへの展開についても、当初の想定以上に幅広く進めることができた。LTEによるシステム解析、事前分布デザインについては、今後のデータ同化の自動化原理構築のための基盤となる知見が蓄積できたことから、一定程度目的は達成できたが、アルゴリズム切り替えにおいて課題が残った。統合モデル構築に関しては達成に至らず、今後の検討課題として残った



### 3. 今後の展開

LTE 解析によるデータ同化アルゴリズムの選択基盤、ならびに事前分布デザインについては、すでに基礎的な開発が終わっていることから、今後有効性の検証を進めるとともに、テストモデル以外のモデルへの展開を進めていく。

LSLOCK を中心とした機能モデル構築については、レーダーデータや高速 AFM データの実データといった、実時間性とナウキャストや短時間予測が重要な対象に対する適用が考えられる。これにより、現場の計測研究者にとって有用な計測情報を与えることができるようになると考えられるため、実データへの適用性の検討・展開を進めていく。

時間局所定常性ならびに局所空間一様性をコントロールするチューニングパラメータがある。当該チューニングパラメータは、本来エミュレート対象としているシステムの時空間スケールに合わせて設定されるものである。しかしながら、これを逆に能動的にコントロールすることで、システム解析のための道具、すなわち現象の時空間スケールの階層性抽出の道具として使用することが考えられる。今後この点について検討していきたい。

### 4. 自己評価

本研究のテーマである「データ同化の自動化原理開発」について、今後の展開につながる基礎的な内容の部分については、個別テーマ毎に達成度に差異があるものの、一定程度達成できたと考えている。その一方で、最終的な目標として提案した統合原理の開発までは至らなかった。また、本さがけ研究の目的の一つとして掲げた、データ同化の領域内外への研究展開に関しては、当初目標以上に達成できたと考えている。

機能モデルの構築に関しては、当初の想定とは異なる方向に進んだものの、データ同化のためにデータから簡易的なエミュレータモデルを構築する枠組みができたこと、さらにそこからのデータ同化だけでなく、システム解析分野への展開が見込まれることから、当初の目標以上の成果が得られたと考えている。

データ同化手法の領域内外への研究展開のうち、領域内については、本さがけ領域内での連携研究者が複数になったことに加え、さがけ・CREST 複合領域として、CREST 領域内チームとの共著の成果を実現できたこと、領域外では、他のさがけ領域との研究者との連携や、データ同化に関する企業との連携につながったことから、当初の想定以上に達成できたと考えている。

LTE 解析ならびに事前分布デザインについては、今後のデータ同化モデリング自動化原理のための基礎的な知見の積み重ねまで実現できた。当初の目標まではいたらなかったが、今後の展開のための知見が得られたと考えている。

当初目標とした統合モデルの最終的な実現までには至らなかったことから、当初目標の設定については、より適切かつ精緻な検討の必要性があったと考えている。

研究実施体制ならび研究費執行については、一定の研究補助を得ながら進められた一方、コロナ禍による影響により、国内外での情報交換の機会が著しく減ったこと、ならびに研究補助に関しても、補助学生の状況に困難な点があったことから、当初想定以下の実質的な体制と状況となった。今回のコロナ禍のような特殊な状況下での進め方・マネジメントに課題が残ったと考えている。

研究成果の科学技術及び社会・経済への波及効果については、基礎的な研究であり、今後の

展開次第と考えている。特に、時空間データからの機能モデル・エミュレータ構築については、データ同化のためのモデル構築だけでなく、昨今の深層学習ベースの方法とは異なるアプローチとしての予測における有効性が得られていることや、推定におけるチューニングパラメータの積極的活用といった新たな数値的研究の切り口が見込まれることから、さきがけ研究の目的とする新たなシーズの創出につながったと考えている。

## 5. 主な研究成果リスト

### (1) 代表的な論文(原著論文)発表

研究期間累積件数:4件

1. Tsuyoshi Ishizone, Kazuyuki Nakamura, “LSLOCK: A method to estimate state space model by spatiotemporal continuity,” Proc. 14th APCA International Conference on Automatic Control and Soft Computing(CONTROLO 2020), Lecture Notes in Electrical Engineering, 2020, 695, pp. 342–351.

時空間データに対する、線形ガウス状態空間モデルにおける実時間性を持った状態・モデル同時推定手法について導入した。ここでは、時間局所化と局所空間一様性という2つのアイデアを用いることで、パラメータ数を削減するとともに実時間性の確保に成功した。論文では、推定手法の内容を示すとともに、数値シミュレーションによって、ノイズ除去と短期予測性能の観点での有効性を示した。

### (2) 特許出願

なし

### (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)

・中村和幸,「データ同化のこれまでとこれから」シミュレーション, Vol. 38, No. 1, pp. 21–27, 2019.

・中村和幸,「時空間強度計測データに対するダイナミクス推定と予測」, RIMS 共同研究(グループ型 A)「モデル駆動とデータ駆動の協同によるデータ数理科学に向けた基盤研究の新展開」, 2019年8月21日(日本、京都)。(招待講演)

・Kazuyuki Nakamura, “Data assimilation with local translation error analysis,” Joint Statistical Meetings, 2019年7月30日(アメリカ、コロラド)。

・Tomoro Kimura, Kazuyuki Nakamura, “Bayesian Estimation of Deer Population Dynamics Using Hamiltonian Monte Carlo Algorithm,” Proceedings of the 50th ISCIE International Symposium on Stochastic Systems Theory and Its Applications, pp. 19–24, 2019.

・Shintaroh Kubo, Suguru Kato, Kazuyuki Nakamura, Noriyuki Kodera, Shoji Takada, “Resolving the data asynchronicity in high-speed atomic force microscopy measurement via the Kalman Smoother,” Scientific reports, 10, 18393, 2020.