

jh250060

数値シミュレーションで読み解く最深宇宙の構造形成

西澤 淳（岐阜聖徳学園大学）

概要

1. 共同研究に関する情報

(1) 共同利用・共同研究を実施している拠点名

東京大学 情報基盤センター

大阪大学 D3 センター

(2) 課題分野

大規模計算科学課題分野

データ科学・データ利活用課題分野

(3) 参加研究者一覧と役割分担

- 西澤淳（岐阜聖徳学園大学・准教授）：研究の総括、機械学習の開発
- 長峯健太郎（大阪大学・教授）：銀河モデル開発、数値シミュレーション実行補助
- 奥裕理（浙江大学・研究員）：数値シミュレーションの実行補助
- 中島光一郎（名古屋大学・大学院生）：数値シミュレーションの実行・エミュレータ開発
- 大里健（千葉大学・特任助教）：ダークマターシミュレーションの実装、実行
- 永井 佑紀（東京大学・准教授）：機械学習関連のアドバイザー
- 飯野 孝浩（東京大学・准教授）：研究全般に関するアドバイザー

2. 研究の目的と意義

本研究課題「数値シミュレーションで読み解

く深宇宙の構造形成」は、宇宙大規模構造の形成と進化を理論・数値・データ科学の統合的手法により解明することを目的とする。特に、ダークマター、バリオン、そして中性水素ガスの分布とその進化を一貫した枠組みで理解し、観測量へと直接結びつける理論基盤の構築を目指している。

宇宙構造形成は、初期揺らぎの重力不安定成長によって進行するが、その非線形発展は極めて複雑であり、解析的手法のみでは十分な理解が困難である。特に、小スケール領域では、ガス力学、加熱・冷却過程、星形成やフィードバックといったバリオン物理が重要な役割を果たす。このため、精密な理解には数値シミュレーションが不可欠である。しかしながら、完全な物理現象を取り入れたシミュレーションは膨大な計算資源を必要とし、宇宙論パラメータ空間の広範な探索や観測データ解析において大きな制約となっている。

本研究では、この課題に対して以下の三つの戦略を統合的に推進する。第一に、大規模 N 体シミュレーションおよび流体シミュレーションによる高精度な構造形成の再現である。これにより、ダークマターおよびガス分布の非線形進化を詳細に記述する。第二に、観測量（特にライマン α の森フラックス）を直接再現するための理論モデリングの高度化である。速度場による赤方空間効果や光学

的深さの計算精度を向上させることで、観測との比較可能性を高める。第三に、機械学習を用いたエミュレータの構築により、計算コストを大幅に削減しつつ高速かつ高精度な予測を可能にする。

このアプローチにより、本研究は従来の数値シミュレーション研究の制約を克服し、以下のような新しい研究展開を可能にする。

- 宇宙論パラメータに対する観測量の迅速かつ高精度な予測
- 非線形スケールにおける構造形成の統計的理解の深化
- 観測データ解析における理論的不確定性の低減

特にライマン α の森は、高赤方偏移における物質分布を探る強力なプローブであり、小スケールの構造形成やダークマターの性質に敏感である。本研究により、その理論予測精度が向上することで、ウォームダークマターなどの非標準宇宙論モデルの検証が可能となる。

さらに、本研究は今後の大規模観測プロジェクト (DESI、Subaru/PFS 等) に対して重要な理論基盤を提供する。観測データの高精度化に伴い、理論予測の誤差が支配的となる可能性が指摘されているが、本研究で開発するエミュレータはそのボトルネックを解消する鍵となる。

また、本研究の意義は宇宙論にとどまらず、高次元データ解析や機械学習の応用という観点でも広がりを持つ。特に、物理シミュレーションと機械学習の融合は、他分野 (流体力学、気候科学など) への展開可能性を有しており、計算科学全体への波及効果も期待される。

以上のように、本研究は宇宙構造形成の理解を飛躍的に進展させるとともに、観測宇宙論および計算科学の双方に対して重要な貢献をもたらすものである。

3. 当拠点の公募型共同研究として実施した意義

本研究を JHPCN の公募型共同研究として実施した意義は、「大規模計算資源の戦略的活用」「分野横断型研究の推進」「次世代計算手法の創出」という観点から極めて大きい。

第一に、本研究は本質的に大規模計算資源を必要とする課題である。宇宙構造形成の数値シミュレーションでは、数億~数十億粒子規模の計算が必要であり、さらに時間発展を追跡するためには多数のスナップショット出力が求められる。また、パラメータ依存性を評価するためには複数の初期条件や宇宙論モデルに対する計算が不可欠である。このような要求は、単一の研究機関の計算環境では対応が困難であり、JHPCN が提供する高性能計算基盤の活用が不可欠であった。本研究では、これらの資源を活用することで、大規模シミュレーションの実行からデータ解析、さらには機械学習モデルの訓練に至るまで、一貫した計算環境を構築することができた。

第二に、公募型共同研究の枠組みは、分野横断的な連携を促進する上で重要な役割を果たした。本研究は、宇宙論シミュレーション、観測データ解析、機械学習という複数の専門分野にまたがる学際的テーマである。JHPCN の枠組みを通じて、これらの分野の研究者が協働することで、単独研究では達成困難な成果を実現することが

できた。例えば、シミュレーションデータの生成と解析、機械学習モデルの設計と最適化、観測量との比較といった各プロセスにおいて、専門的知見の融合が重要な役割を果たしている。また、共同研究を通じたコードやデータの共有により、研究の再現性と効率性も大きく向上した。

第三に、本研究は高性能計算環境を前提とした新しい計算手法の開発を推進した点でも意義が大きい。従来宇宙論研究では、フル物理シミュレーションに依存した解析が主流であったが、本研究では機械学習を活用したエミュレーション技術を導入し、計算コストを大幅に削減しつつ高精度な予測を実現した。このような手法は、大規模観測データの解析が求められる今後の宇宙論研究において不可欠であり、JHPCN の計算資源を活用することで初めて実用的な形で実装することが可能となった。

さらに、JHPCN の枠組みは、単なる計算資源の提供にとどまらず、研究成果の波及効果を高める役割も担っている。本研究で得られたシミュレーションデータや解析手法は、将来的にコミュニティへの公開や他分野への応用が期待される。特に、機械学習エミュレータは、他の宇宙論プローブや異なる物理系への応用可能性を有しており、計算科学全体への貢献が見込まれる。

また、若手研究者の育成という観点でも、本研究の意義は大きい。大規模シミュレーションの実行、データ解析、機械学習モデルの構築といった一連のプロセスを通じて、高度な計算科学スキルを習得する機会が提供されている。これは、将来的に高性能計算を活用する研究者の育成に寄与するものであり、JHPCN の目的とも合致する。

以上のように、本研究は JHPCN の公募型共同研究として実施することで、大規模計算資源の有効活用、分野横断的研究の推進、新規計算手法の開発、さらには人材育成に至るまで、多面的な意義を有する取り組みとなっている。

4. 前年度までに得られた研究成果の概要 「該当なし」

5. 今年度の研究成果の詳細

本年度は、宇宙大規模構造形成の理解に向けて、「高精度数値シミュレーションの実行」「観測量の理論的再現」「機械学習による高速予測基盤の構築」という三つの柱に沿って研究を推進し、それぞれにおいて具体的かつ統合的な成果を得た。

まず、大規模 N 体シミュレーションおよびデータ解析に関しては、ダークマター密度場の非線形進化を高解像度で追跡する計算を実施し、複数の赤方偏移におけるスナップショットデータを体系的に取得した。これらのデータを用いて、パワースペクトルおよび 2 点相関関数の精密評価を行うとともに、密度場の非ガウス性に起因する高次統計量の振る舞いについても解析を進めた。特に、シミュレーションデータの取り扱いにおいては、ファイル分割された大規模データを逐次的に読み込みながら密度場を再構築する手法を確立し、メモリ使用量を抑えつつ高効率に解析を行うワークフローを構築した。この技術的基盤により、従来は困難であった大規模データセットの実用的な解析が可能となった。

次に、ライマン α の森の理論予測に関する研究では、ガス分布および吸収線形成の物理過

程を数値的に再現するためのアルゴリズムの高度化を行った。具体的には、粒子ベースのガス分布からグリッド上の物理量を構築する際のメッシュアサイン手法を精査し、空間分解能と数値安定性のバランスを改善した。また、従来の実装では十分に考慮されていなかった速度場の寄与を導入し、視線方向のドップラーシフトを考慮した光学的深さの計算を実装した。この改良により、赤方偏移空間におけるフラックスの異方性を再現することが可能となり、観測データに直接対応するフラックスパワースペクトルの理論予測精度が向上した。さらに、スキューワ（視線方向サンプル）の生成と統計解析を組み合わせることで、観測量と比較可能な形での統計量抽出を実現した。

機械学習エミュレータの開発においては、複数の手法を並行して検討し、それぞれの特性を活かした予測基盤を構築した。まず、3次元CNNを用いたマップ・トゥ・マップ推定では、ダークマター密度場からガス分布を直接推定するモデルを構築した。入力および出力はともに3次元ボクセルデータであり、 $(256^3 \sim 512^3)$ の高次元データを扱う必要があるため、パッチベースの学習手法を採用した。これにより、GPUメモリ制約を回避しつつ、大域的構造を保持した学習が可能となった。また、データ拡張（回転・反転）を積極的に導入することで、単一シミュレーションからの学習における過学習を抑制し、汎化性能の向上を実現した。

さらに、統計量予測に特化したエミュレータとして、PCAと回帰モデルを組み合わせた手法を開発した。パワースペクトルなどの高次元出力を主成分空間に射影することで次元圧縮を行い、その後k近傍法(kNN)や多層パーセプトロン(MLP)を用いて回帰を行う構成を採用した。このアプローチにより、計算コ

ストを削減しながら高精度な予測を実現した。さらに、ハイパーパラメータ探索を体系的に行い、モデルの性能評価と最適化を実施したことで、異なるモデル間の性能比較が可能となり、用途に応じた最適な手法選択の指針を得た。本成果については、目標とする予測性能に到達しておらず、一部は論文として未公開であるが、翌年度以降の改善課題とする。

これらの技術開発を通じて、ダークマターシミュレーションから観測量への写像を効率的に構築する統合的なフレームワークが整備された点は、本年度の重要な成果である。このフレームワークは、従来のフル物理シミュレーションに依存しない高速予測を可能にし、今後の大規模観測データ解析において大きな役割を果たすと期待される。

研究成果の発信について、本年度はシミュレーション画像(密度場マップ)から別の密度場マップを推定するmap to map inferenceに関する査読付き論文を1報、中性水素の温度揺らぎからダークマター質量などの物理情報を抽出する論文を2報、CROCODILEシミュレーションを用いたFast Radio BurstへのAGNフィードバック効果を検証した論文を1報報告した。また、学会発表では、ライマン α の森のエミュレータに関する学会・国際会議での発表を3件行った。これらの論文や発表は、理論予測の精度向上と計算効率の両立という点で国際的にも意義のある成果であり、関連分野における基盤的研究として位置づけられる。また、これらの成果は相互に関連しており、個別の研究としてだけでなく、統合的な研究プログラムとしての一貫性を保持している。

以上のように、本年度は数値シミュレーション、理論モデル、機械学習手法の融合により、

宇宙構造形成の理解と観測との接続に向けた実質的な進展を達成した。

また、本課題に関連し、JHPCN の旅費支援を受け、すばる望遠鏡の分光観測装置 PFS の観測開始に伴う理論モデル整備を目的として、数値シミュレーション専門家の会合を9月に東大 IPMU で開催した。約 20 名が参加し、課題や独自技術に関する情報交換を行い、その後の共同研究へと発展している。本共同研究の成果については、今後の報告で詳述する予定である。

6. 進捗状況の自己評価と今後の展望

本研究の進捗状況については、当初の研究計画と比較して総合的に評価すると、「主要目標の一部は達成しており、一部未達成な部分はあるものの、計画を上回る進展を達成した部分もある」と評価する。以下に、研究項目ごとに定量的評価を含めて詳細に述べる。

まず、大規模数値シミュレーションの実行とデータ整備に関しては、当初計画では複数の宇宙論パラメータに対するシミュレーションデータの生成を段階的に進めることを想定していた。本年度は、そのうち代表的なパラメータセットについて高解像度計算を完了し、解析可能なデータセットを構築した。計画していたパラメータ空間全体の網羅には至っていないものの、基盤となるデータ生成および解析手法は確立されており、達成度は約 50%と評価される。特に、大規模データの逐次処理や効率的 I/O 設計に関する技術的課題を解決した点は、今後のスケールアップにおいて重要な進展である。

次に、ライマン α の森に関する理論モデル構築に関しては、光学的深さ計算への速度場導入や赤方空間効果の再現といった課題を達成した。この結果、観測データとの直接比較

が可能な段階のうち 60%程度が達成されたと評価される。

機械学習エミュレータに関しては、当初計画では概念実証レベルのモデル構築を目標としていたが、本年度はそれを上回る成果を達成した。3次元 CNN によるマップ推定、PCA+kNN/MLP による統計量予測など、複数のアプローチを実装し、それぞれの性能を比較評価した点は、計画比で 120%以上の進展と評価できる。ただし、性能面では現状の宇宙論モデル推定に要求されるレベルに到達しておらず、更なる改良を策定している状況である。また、異なる宇宙論モデルへの適用や外挿性能の検証も十分ではなく、これらは今後の重要課題として残されている。

研究成果の公表に関しては、査読付き論文 4 報を発表しており、当初目標（2~3 報程度）を上回る成果を達成した。これにより、本研究の国際的な認知度および学術的インパクトは着実に向上している。

以上を総合すると、本研究の全体的な進捗率は約 70~90%と評価される。特に、基盤技術の確立と主要成果の創出という観点では、計画をほぼ達成している。一方で、パラメータ空間の網羅的探索や観測データへの本格的適用といった次段階の課題が明確化している。

今後の展望としては、以下の三点を重点的に推進する。

第一に、エミュレータの高度化と汎化性能の体系的評価である。現在構築したモデルを異なる初期条件や宇宙論パラメータに適用し、その再現性と限界を定量的に評価する必要がある。また、物理法則に基づく制約を組み込んだ学習手法の導入により、外挿領域での信頼性向上を図る。

第二に、観測データとの統合解析の実施である。特に、ライマン α の森のフラックスパワ

ースペクトルを用いた宇宙論パラメータ推定に本研究のエミュレータを適用し、計算効率と推定精度の両立を検証する。この際、観測ノイズや系統誤差の影響を適切に取り扱うための統計的枠組みの構築も重要となる。第三に、計算基盤およびデータ共有の強化である。多パラメータ空間でのシミュレーションデータベースを作成するためには、シミュレーションそのものの高効率化が必須である。次年度課題として、シミュレーションの高速化を目的とした、GPU 化（GPU と CPU のハイブリッド）を進める。また、本研究で開発した解析コードやエミュレータを再利用可能な形で整備し、共同研究者間での共有を促進するとともに、将来的にはコミュニティへの公開も視野に入れる。これにより、本研究の成果を広く活用可能な形で展開する。さらに中長期的には、本研究で構築した枠組みを他の観測プローブ（銀河分布、弱重力レンズなど）へ拡張することも視野に入れている。これにより、異なる観測手法を統合した宇宙論解析が可能となり、宇宙構造形成の理解がより包括的に進展すると期待される。以上のように、本研究は当初計画を踏まえつつ着実に進展しており、現在は基盤構築段階から応用・展開段階へと移行しつつある。今後は観測との直接比較およびパラメータ推定への応用を通じて、宇宙論研究における実質的な貢献をさらに深化させていく予定である。

※7. 研究業績はウェブ入力です