

東京大学国際高等研究所カブリ数物連携宇宙研究機構 (Kavli IPMU)
アリゾナ大学天文学科
京都大学基礎物理学研究所
名古屋大学素粒子宇宙起源研究所
大学共同利用機関法人 自然科学研究機構 国立天文台

AI とスーパーコンピュータで広大な銀河地図を解読 - 宇宙の成り立ちを決める物理量を精密に測定

概要

アリゾナ大学天文学科 小林洋祐 (こばやし ようすけ) 博士研究員 (2021 年まで東京大学国際高等研究所カブリ数物連携宇宙研究機構 (以下カブリ IPMU) 大学院生及び特任研究員)、京都大学基礎物理学研究所 西道啓博 (にしみち たかひろ) 特定准教授 (兼:カブリ IPMU 客員科学研究員)、カブリ IPMU 高田昌広 (たかだ まさひろ) 教授、名古屋大学素粒子宇宙起源研究所 宮武広直 (みやたけ ひろなお) 准教授からなる共同研究チームは、現在世界最大の銀河サーベイであるスローン・デジタル・スカイ・サーベイ (SDSS) から得られた銀河の 3 次元分布 (地球から見た奥行き方向および 2 次元角度方向) のデータと、宇宙の大規模構造^(注1) の理論モデルを比較し、「宇宙論パラメータ」^(注2) と呼ばれる、宇宙の性質を決める基本的な物理量を測定しました。これを行うために、国立天文台のスーパーコンピュータ「アテルイ II」^(注3) を用いて様々な宇宙論パラメータを仮定して宇宙の構造形成シミュレーションを実行し、その大規模データを人工知能 (AI) 技術のひとつであるニューラルネットワーク^(注4) に学習させることで、任意の宇宙論パラメータに対する理論計算を高速かつ高精度に実行できるソフトウェアを開発しました。つまり、今回の解析は銀河地図の観測とあらゆる宇宙論モデル^(注2) のシミュレーションとの比較と同等になります。直接数値シミュレーションを用いてこの操作を行うには、現実的な時間では完了できないほど膨大な計算量が必要です。ニューラルネットワークに基づくモデルを用いることで、世界で初めてこのような解析が可能となりました。その結果、ダークマターの総量、および現在の宇宙の凸凹の度合いを表す宇宙論パラメータを、先行研究を上回る精度で測定することに成功しました。今回の手法は、カブリ IPMU のリードで現在開発が進んでいるすばる望遠鏡超広視野多天体分光装置 Prime Focus Spectrograph (PFS) による広天域銀河サーベイのデータにも適用することができます。本研究成果は、2022 年 4 月 20 日に米国の物理学専門誌「Physical Review D」にオンライン掲載されました。

1. 背景

宇宙全体の性質やその創成と現在までの進化を明らかにしようとする「宇宙論」は、およそ半世紀にわたる観測技術の著しい向上に支えられ、その理論を観測データによって検証できる実証科学として確立しました。宇宙全体を記述する「宇宙論モデル」として、ダークマターやダークエネルギーといった未知の成分が宇宙のエネルギーの大半を占める「 Λ CDM モデル」^(注5)が、現在の宇宙論における標準モデルとされています。このような宇宙論モデルは、ダークマターの総量や宇宙の凸凹の度合いなど、宇宙の具体的な性質を決める物理量を含んでおり、これらは「宇宙論パラメータ」と呼ばれています。宇宙論モデルを検証するために必要不可欠なのが、宇宙論パラメータを観測データから高い精度で測定することであり、これを目指した観測が世界中で展開されています。

広い天域の宇宙分光サーベイによる銀河の3次元地図は、銀河が特徴的なパターンに従って分布していることを明らかにしています。例えば、多数の銀河が密集する銀河団の領域、銀河が紐状に分布するフィラメント構造、あるいは銀河がほとんど存在しないボイドと呼ばれる構造です。すなわち、銀河は宇宙空間で一様に形成したのではなく、空間的に非一様に、また環境に依存して形成してきたことを示しています。この銀河の非一様分布を背後で支配しているのが、それ自体は光らないために直接観測できないダークマターの空間分布による重力と考えられています。銀河の3次元地図データは、宇宙の大規模構造から宇宙論パラメータを測定する強力な手段になっています。

しかし、銀河の分布は完全にダークマターの分布を反映していないことが分かっており、この不確定性は「銀河バイアス不確定性」として認識されています。この理由で、銀河地図を再現する理論モデルを作るのは簡単ではありません。銀河はガス、星の集団、またダークマターからなる系ですが、物理法則に従い、星の形成から銀河全体の形成まで計算するのは、スーパーコンピュータでもまだ不可能です。さらに、銀河地図に必要なのは、多数の銀河の空間分布を再現する理論モデルです。このため、観測データの解析では、銀河形成を完全にはモデル化できないことを認め、まずダークマターの空間分布を再現する理論モデルを構築し、そのあとで銀河とダークマターの空間分布のあいだの不確定性をモデル化します。最終的には銀河バイアスの不確定性のとり得る範囲を考慮し、興味ある宇宙の物理量（例えばダークマターの総量）を測定するという手法が必要になります。これまでの主な理論モデルは、宇宙背景放射の解析で大成功を収めた宇宙構造形成の線形理論、あるいはそれを拡張した解析的理論に現象論的に銀河バイアス不確定性を記述するパラメータを導入したモデルでした。

2. 研究手法・成果

本研究グループは、解析において観測データとの比較に用いる銀河分布の主要な統計的性質を抽出した量である「パワースペクトル」^(注6)の理論計算を、スーパーコンピュータによる数値シミュレーションと人工知能(AI)とを組み合わせた手法によって実現しました。

近年の宇宙論研究では、宇宙の大規模構造をコンピュータ上で再現するシミュレーション「N体シミュレーション」が大きな役割を担っています。N体シミュレーションでは、宇宙の物質の大部分を占めるダークマターを模した大量の粒子を生成し、それらが重力で引き合う効果を計算することで、ダークマターの空間分布が成長していく過程を数値的に再現することができま

す。これまでの研究では、今回のようなパワースペクトルの理論計算には、物質が従う流体の運動方程式を解析的に解くことで、ダークマターの密度ゆらぎ^(注7)の成長を計算するアプローチが採られてきました。ところが、この流体方程式は厳密解が知られておらず、密度ゆらぎが小さいとする近似を用いて解かざるを得ないため、N体シミュレーションの結果を正確には再現できないことが問題とされてきました。

本研究グループは、そうした近似を用いないN体シミュレーションをパワースペクトルの理論計算に直接応用するために、AI技術に着目しました。N体シミュレーションは多数の粒子の運動を解くため、その計算にはスーパーコンピュータを使っても数十時間かかります。観測データと照らし合わせ、確からしい宇宙論パラメータの値を探し出すためには、宇宙論パラメータを変えた理論計算を数百万回も実行する必要があるため、これをそのまま理論計算に用いることは現実的ではありません。そこで、国立天文台のスーパーコンピュータ「アテルイII」を利用し、百万とは言わないまでも、宇宙論パラメータの値を変えた100通りほどの宇宙論モデルの下でN体シミュレーションを実行しました。その上で、ニューラルネットワーク^(注4)を用いてそれらの結果を学習させることで、シミュレーションを実行していない宇宙論パラメータでもシミュレーションと同程度の精度でパワースペクトルを1秒以内に計算するソフトウェアの開発に成功しました(図1が概念図)。このような、計算コストの大きな物理シミュレーションをそれより遥かに小さい計算コストで模倣する技術は「エミュレータ」と呼ばれ、宇宙論以外にもコンピュータサイエンスなどの分野で既に活用されています。

銀河の形成・進化のプロセスは重力以外にも様々な物理作用から成り立っており、このことが、銀河とダークマターの分布の違い(銀河バイアス)を生んでいます。しかし、この銀河の形成・進化に関わる物理作用は実は現在でも十分に解明されていないため、銀河バイアスを完全に計算することはできず、理論的に不確定な要素(不確定性)が残っています。そこで本研究グループでは、大規模構造のなかでダークマターの空間密度が高く銀河ができやすいダークマターハロー(以下ハローという)と呼ばれる領域に着目しました。そして、ハロー内部における銀河の分布をモデル化し、ハローのパワースペクトルから銀河パワースペクトルを計算するという手法を用いました。ハロー内部の銀河分布のモデルには銀河バイアスの不確定性がありますが、これを表すパラメータを広い範囲に渡って動かしながら宇宙論パラメータを推定することで、銀河バイアスの不確定性があっても宇宙論パラメータを正確に推定できることを確かめました。実際に今回の研究では、エミュレータの訓練に用いたものよりも巨大なシミュレーションデータを新たにアテルイIIで作成し、これにエミュレータに基づく解析パイプラインを適用することで、宇宙論パラメータの推定結果が銀河バイアス不確定性に対して頑強であることを厳格に検証しました。

本研究グループは、エミュレータの理論予言を用いて、現在世界最大の銀河地図データであるスローン・デジタル・スカイ・サーベイで得られた銀河パワースペクトルから宇宙論パラメータを測定しました。エミュレータにより、観測された銀河地図データとN体シミュレーションが計算する銀河の分布をいわば直接比較して、観測データに誤差の範囲内で合致する宇宙論パラメータの値を探索することが可能になります(図2)。図3は、今回の解析で得られた、現在の宇宙全体のエネルギーに占める物質の割合および現在の宇宙の大規模構造がもつ凸凹の度合いを示す宇宙論パラメータの値が観測データと合致する範囲を表示しています。宇宙に占める物

質のエネルギーは約 3 割で、残りの約 7 割は宇宙の加速膨脹を引き起こすダークエネルギーで占められているという、これまでの宇宙論研究で明らかになってきた宇宙の姿を、独立な解析として改めて示しています。また、このような銀河地図の解析に従来用いられてきた手法では、図のオレンジ色の領域で示されるように、この 2 つの宇宙論パラメータを同時に決めることができず、物質が多く凸凹が小さい宇宙と、物質が少なく凸凹が大きい宇宙とを区別できませんでした。今回の解析では、エミュレータの高速な計算を利用して、2 つのパラメータが同時に小さい誤差の範囲で測定されました。この測定精度は従来行われてきた大規模構造の解析では達成されなかった精度であり、宇宙論研究におけるエミュレータの有用性を実証しています。

3. 波及効果・今後の展望

本研究は、ニューラルネットワークを用いたシミュレーションの学習によって、宇宙大規模構造のデータ解析で実際に宇宙論パラメータを高い精度で測定できることを世界で初めて示しました。近年行われた複数の大規模な観測計画により、これまで定説と考えられてきた最も単純な宇宙論モデルである Λ CDM モデルに綻びがある可能性が指摘されつつあります。本研究グループは、ニューラルネットワークを用いた新しい解析手法がこの問題に独立な視点から答えを出せることを実証したものとと言えます。今後もシミュレーションや機械学習の技術の向上が期待され、より高精度な理論計算を可能とするエミュレータを開発できる見込みです。今回開発したエミュレータの手法は、カブリ IPMU が中心に進めるすばる望遠鏡超広視野多天体分光装置 Prime Focus Spectrograph による銀河地図にも用いることができ、ダークマターの総量、またダークエネルギーの性質などを解明し、現在の標準的な宇宙の理解を超えた新しい発見へと繋がる可能性を秘めています。

4. 本研究プロジェクトについて

本研究は、科学技術振興機構 CREST (JPMHCR1414、代表：吉田直紀)、科学技術振興機構 AIP 加速課題 (JP20317829、代表：吉田直紀)、科学研究費補助金 新学術領域研究「加速宇宙」総括班 (15H05887、代表：村山斉) 計画研究 (15H05892、代表：宮崎聡、15H05893、代表：高田昌広)、科学研究費補助金学術変革領域「ダークマター」総括班 (20H05850、代表：村山斉) 計画研究 (20H05855、代表：高田昌広、20H05861、代表：安藤真一郎)、科学研究費補助金 基盤研究(A) (19H00677、代表：高田昌広)、科学研究費補助金 若手研究 (B) (17K14273、代表：西道啓博)、科学研究費補助金 基盤研究 (B) (20H01932、代表：宮武広直)、東京大学 Beyond AI、の援助を受けています。

<用語解説>

注 1) 宇宙の大規模構造：宇宙初期に量子論的な効果で生成されたとされる微小な密度揺らぎが重力によって成長してできた、物質（ダークマターなど）や銀河が織りなす泡状の構造。およそ 100 万年以上という巨大なスケールにわたって広がることから大規模構造と呼ばれる。

注 2) 宇宙論モデル・宇宙論パラメータ：宇宙の誕生から現在までを物理法則に従って記述するモデル（宇宙全体の模型）を宇宙論モデルという。宇宙論モデルは、宇宙の性質を決める基本的

な物理量（宇宙の全エネルギーに占めるダークマターなどの諸成分の割合、宇宙の膨張速度、宇宙初期に生成された密度ゆらぎの大きさなど）をパラメータとしてもつ。これらを総称して宇宙論パラメータという。

注3) アテルイII：国立天文台が運用する天文学専用スーパーコンピュータ（Cray XC50）。岩手県奥州市の国立天文台水沢キャンパスに設置され、2018年から稼働を続ける。理論演算性能3.087ペタフロップス（1秒間に3000兆回の浮動小数点演算を行う性能）で、国内外の天文学者がアクセスし利用している。平安時代に奥州市の地域を治め活躍した蝦夷の長「阿弭流為」にちなみ、宇宙の謎に果敢に挑んでほしいという願いをこめて命名された。

注4) ニューラルネットワーク：人工知能（AI）技術の一種である機械学習で用いられるアルゴリズムの一つ。入力と出力との対応関係を複雑なネットワーク状の構造によって学習する数理モデルで、本研究では主に入力となる宇宙論パラメータと出力となるハローのパワースペクトルとの対応関係を精密に学習することでエミュレータの開発が実現した。

注5) Λ CDMモデル：宇宙膨張を加速させるダークエネルギーの候補としてアインシュタインの宇宙定数 Λ （ラムダ）を、宇宙の主な重力源を担うダークマターとして、重力以外の効果を一切及ぼさない「冷たいダークマター」（Cold Dark Matter: CDM）を採用した、最も基本的な宇宙論モデル。長らく多方面の観測データによるテストを通過してきたが、近年になって宇宙の最初期の姿を捉えた宇宙マイクロ波背景放射の観測と、近傍宇宙の天体の明るさと距離の関係に基づく宇宙の膨張速度測定結果を同時に説明できない可能性が指摘され、宇宙論における大きな話題となっている。

注6) パワースペクトル：大規模構造に広がる銀河がどのような分布の仕方をしているかを統計的に表す量の一つで、銀河サーベイで銀河の分布から測定される最も基本的なデータである。本研究ではパワースペクトルの観測データを理論計算と比較することで、宇宙論パラメータを測定した。

注7) 密度ゆらぎ：宇宙に存在するエネルギー成分（物質など）の空間密度の濃淡（凸凹）をさす。現在の宇宙の密度ゆらぎの平均的な大きさを表す宇宙論パラメータが本研究で測定された。

<研究者のコメント>

西道啓博特定准教授のコメント

2015年のプロジェクト立ち上げから、シミュレーションデータベースの構築、機械学習の導入、模擬データを用いた綿密な検証を経て、ようやく本物の宇宙から測定されたデータの分析に辿り着くことができました。感無量です。とは言え、今回提案した手法をもってしても、観測データの全てを解釈するまでには至っていません。宇宙論分野においても、近年機械学習を中心とした人工知能技術の導入が急速に進み、様々な形での応用が検討されています。今後とも、人工知

能と物理学者がそれぞれの得意分野で力を発揮し、協働することで、ダークマター、ダークエネルギーなどの宇宙の根源的な謎に迫っていきたいです。

小林洋祐研究員のコメント

計算コストのかかる数値シミュレーションを機械学習と組み合わせることで、シミュレーションがもたらす精度の高い理論計算を直接的に観測データの解析に使うことができるようになりました。実はこのようなエミュレータの開発自体はこれまでの宇宙論分野でも複数の先行研究がありましたが、観測で考慮しなければならない効果をきちんと取り入れられていないという事情で、実際の銀河サーベイのデータに適用して本格的に宇宙論パラメータの推定まで実行した例はまだほとんどありません。本研究で開発したエミュレータはそうした効果を考慮できるように設計したもので、問題なく実際の観測データに適用できました。今回の成果は、シミュレーションと機械学習を用いた大規模構造データ解析のフロンティアを拓いた研究の一つと言えます。

高田昌広教授のコメント

今回の研究成果を大変うれしく思っています。小林氏の博士論文の研究テーマで、5年間ほどの努力の集大成です。今回の新しい研究手法は、カブリ IPMU がリードしているすばる主焦点多天体分光器 Prime Focus Spectrograph (PFS)が行う銀河地図にも適用できます！それによりダークマターの総量を正確に測定し、さらにニュートリノ質量を決定できると期待しています。

宮武広直准教授のコメント

今回の研究では、人工知能を利用することによって、銀河のパワースペクトルの測定データを最大限に生かした宇宙論解析を可能にしました。同様に重力レンズ効果などの宇宙論的統計量に対しても、人工知能を用いることで精密かつ高精度な解析が可能になります。近い将来、銀河のパワースペクトルと重力レンズ効果を同時に用いたさらに高精度な解析を行うことを楽しみにしています。

<論文タイトルと著者>

タイトル：Full-shape cosmology analysis of SDSS-III BOSS galaxy power spectrum using emulator-based halo model: a 5% determination of σ_8 (エミュレータに基づくハローモデルを用いた SDSS-III BOSS 銀河パワースペクトルのフルシェイプ宇宙論解析: σ_8 を 5%精度で決定)

著者：Yosuke Kobayashi (1,2), Takahiro Nishimichi (3,2), Masahiro Takada (2), Hironao Miyatake (4,2)

著者所属：

1 Department of Astronomy/Steward Observatory, University of Arizona, Arizona, USA

2 Kavli Institute for the Physics and Mathematics of the Universe (WPI), The University of Tokyo, Chiba, Japan

3 Center for Gravitational Physics, Yukawa Institute for Theoretical Physics, Kyoto University, Kyoto, Japan

4 Kobayashi-Maskawa Institute for the Origin of Particles and the Universe (KMI), Nagoya University, Nagoya, Japan

雑誌名 : Physical Review D

DOI : <https://link.aps.org/doi/10.1103/PhysRevD.105.083517> (2022 年 4 月 20 日掲載)

* 論文のアブストラクト (Physical Review D のページ)

<https://journals.aps.org/prd/abstract/10.1103/PhysRevD.105.083517>

* プレプリント (arXiv.org のウェブページ)

<https://arxiv.org/abs/2110.06969>

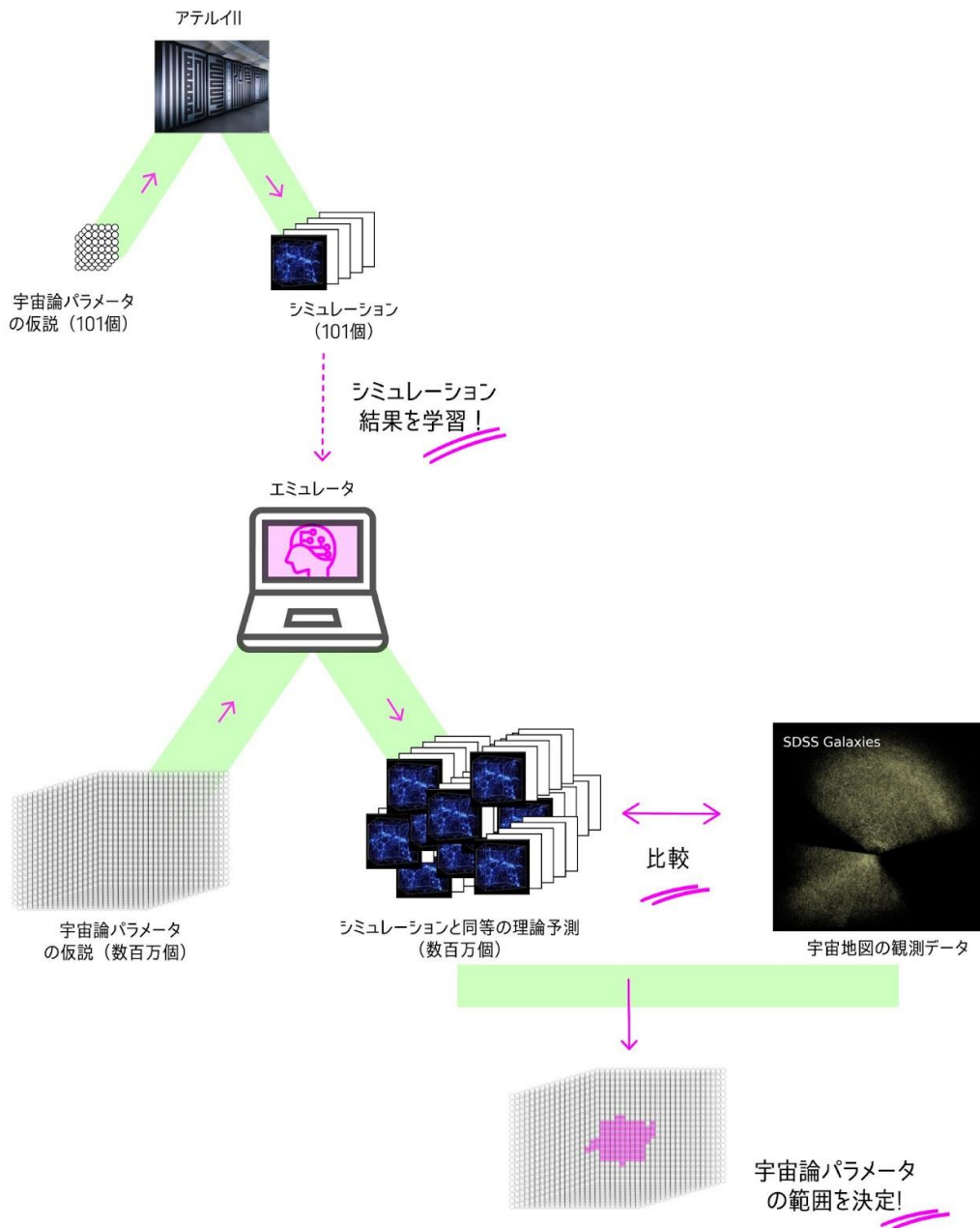


図1. 本研究で開発した、ニューラルネットワークによる宇宙の構造形成の理論模型を高速かつ高精度で計算する手法の概念図。大規模構造の形成を再現するシミュレーション群をデータセットとしてニューラルネットワークに学習させることで、実際にはシミュレーションを行っていない宇宙論パラメータに対しても、大規模構造に広がる銀河分布の性質をシミュレーションと同等の精度で1秒以内に計算できる。(クレジット：Kavli IPMU、国立天文台)

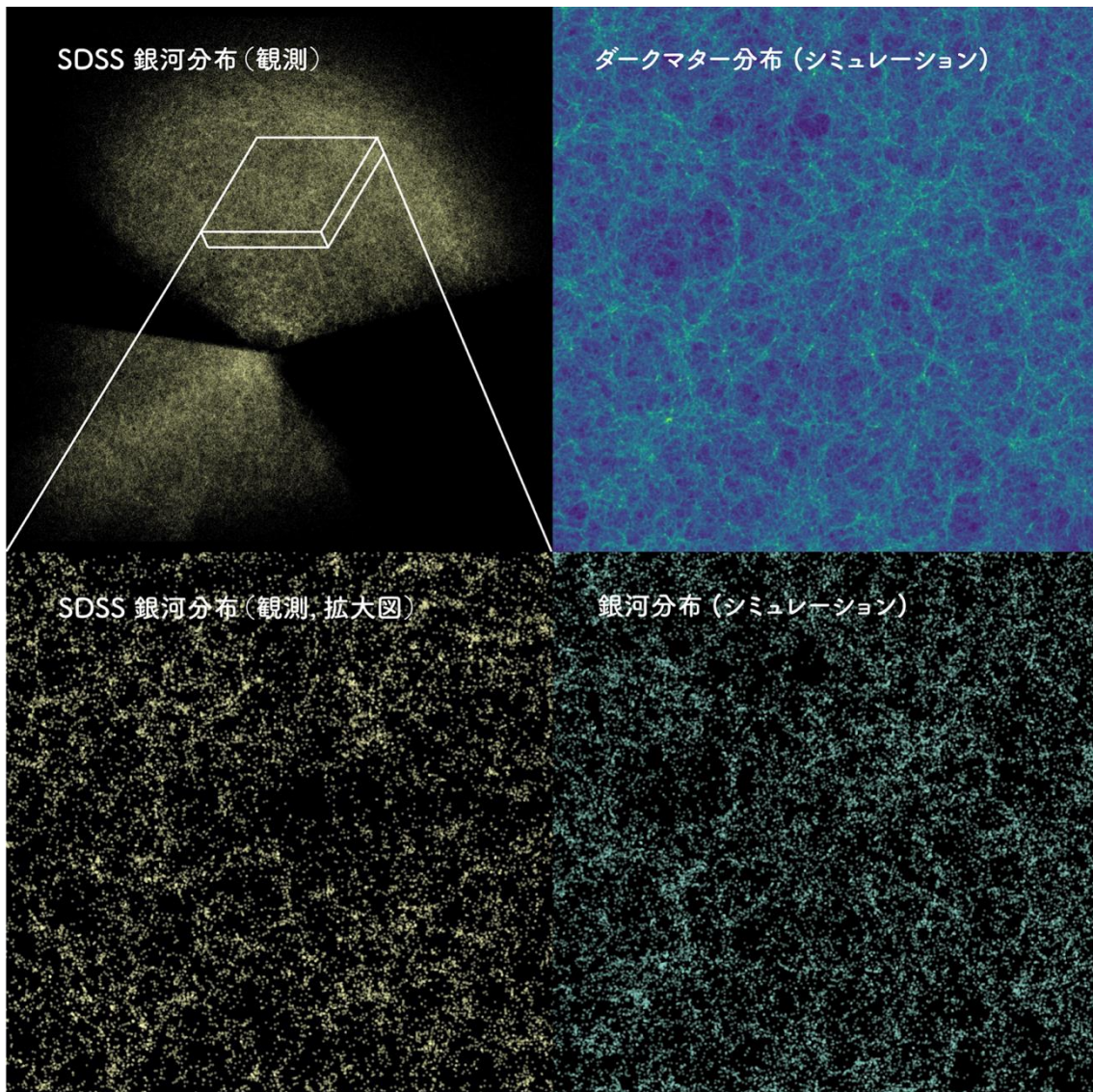


図 2. 本研究に利用した、スローン・デジタル・スカイ・サーベイによる約 100 万個の銀河地図 (左上: 一点一点が一つの銀河を表しており、中心に地球が位置している) と、薄い直方体領域の拡大図 (左下)。左下図と同じ大きさの領域に対して、AI が導き出した宇宙論パラメータを採用した数値シミュレーションから予想されるダークマターの分布 (右上) と、ダークマターが密集した場所に形成される銀河の分布 (右下)。AI と数値シミュレーションが予想する銀河の分布には、実際の観測データとよく似た、銀河団やフィラメント、ボイドなどの特徴的パターンが見られる。(クレジット: 西道啓博)

高解像度版:

https://www2.yukawa.kyoto-u.ac.jp/~takahiro.nishimichi/figs/EmuFullShape_fig2.pdf

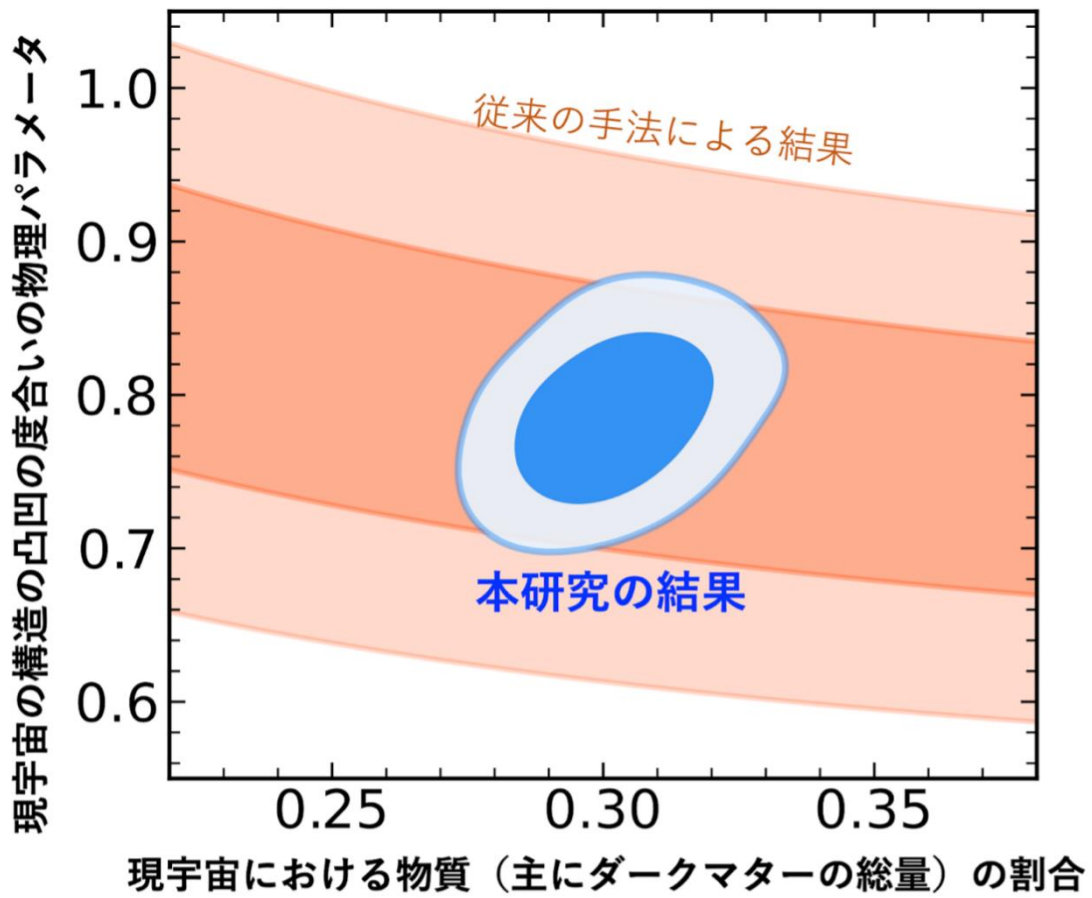


図3. スローン・デジタル・スカイ・サーベイの3次元銀河地図のデータと本研究のニューラルネットワーク（機械学習）に基づく方法の理論模型との比較から得られた宇宙論パラメータの測定の結果。横軸は現宇宙のエネルギーに対する物質（主にダークマター）の割合であり、縦軸は現宇宙の構造の凸凹の度合いを表す物理量（値が大きい宇宙ほど、多くの銀河が存在する宇宙に対応する）。薄い、濃い青色の領域は測定結果の68%、95%の信頼区間で、この領域のなかにそれぞれの確率で本当の宇宙の値が存在することを意味する。オレンジ色の領域は先行研究の結果。（クレジット：小林洋祐）