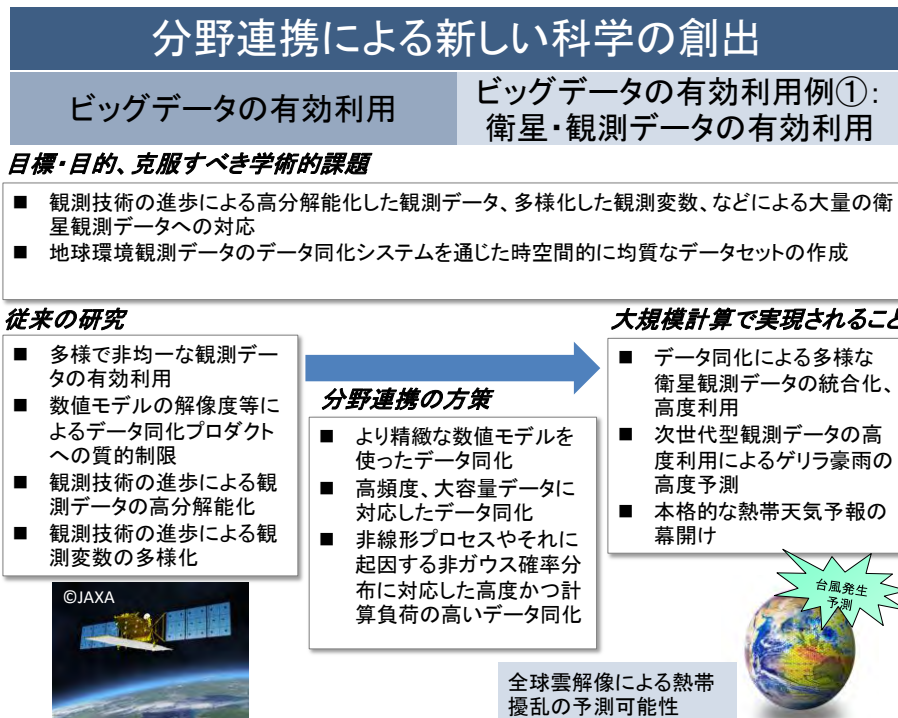


(2) ビッグデータの有効利用例①：衛星・観測データの有効利用

大気、海洋、陸域の物理・化学および生物環境に関する観測データは、環境変動の監視・検出や影響予測のための基礎データとして重要な役割を果たしており、日々の天気予報から地球規模の気候変動まで幅広い環境問題に適切に対処することに貢献している。これらのデータは、現場における直接観測に加え、人工衛星からのリモートセンシングなどさまざまな方法により観測が行われ、その利用分野も広い範囲に及んでいる。ただし、それぞれの観測データは異なる場所、時間で得られており、また観測される物理量変量や精度等もさまざまであるため、単に既存の観測データをまとめただけではそのまま実際に利用することは難しい。そこで、データ同化と呼ばれる、数値モデルを用いて異なった観測データを統合する手法により、より使いやすいデータセットに加工されている。

大規模なデータ同化は、これまで気象・気候分野での技術的発展が顕著であった。しかし、最近では、設計制御、石油掘削、分子シミュレーションなど他の分野においても適用が図られてきており、各アプリケーション分野のコミュニティに閉じていたデータ同化技術は、コミュニティを越えて共通化する動きが起きている。また、観測技術の進歩による高分解能化した観測データ、多様化した観測変数などによる大量の衛星観測データおよびそのデータを用いたシミュレーションの双方の桁違いの大容量化に対応するために、ビッグデータを扱うデータ同化技術のイノベーションが起ころつつある。

データ同化技術の進化にともなって、シミュレーションが高分解能化・精緻化することにより、ゲリラ豪雨や熱帯天気予報と言った社会的インパクトの高い局所的な気象現象の予測などに正面から取り組むことが可能となっていく。



### 3 分野連携による新しい科学の創出

衛星・観測データの有効利用において今後、必要となる計算機性能を下表に示す。

課題	要求性能 (PFLOPS)	要求メモ リバンド 幅 (PB/s)	要求ファ イル/O 性能 (TB/s)	メモリ量/ ケース (PB)	ストレ ージ量/ ケース (PB)	計算時間 /ケース (hour)	ケース数	総演算量 (EFLOP)	概要・計算手法	問題規模	備考
ゲリラ豪雨・熱 帯気象の高度 予測	100	350	100	0.64	10	220	2	160,000	LETKF		NICAM3.5km 100メン バー 1ヶ月
統合地球環境 再解析	8	46	100	0.01	0.6	18	80	42,000	4次元変分法		

※ 本見積もりは、9月末日での見積もりである。未だ精査の余地があり、最終版では、より精度の高い数値を記載する予定である。

#### (3) ビッグデータの有効利用例②：ゲノム解析

ヒトゲノム計画は、分子生物学におけるマイルストーンの一つであり、たった一人のDNA解読に世界の多くの分子生物学者が関わり長い年月を要した。しかし、近年になって次世代DNAシーケンサーと呼ばれる超高速かつ低コストでDNAを解読する装置が開発され、個人のゲノム情報を含むさまざまな細胞のゲノム解析が容易に行うことが可能となった。

分子生物学では、細胞内の大量の遺伝子発現を同時計測可能なDNAマイクロアレイなどの登場により、実験で得られるデータの量が飛躍的に増加している。それらの大量データ、いわゆるハイスループットデータから計算科学を用いて生物学的な発見を行おうとする研究手法（＝バイオインフォマティクス）が行われるようになっており、計算科学との連携が必要な重要な1分野として認識されている。

生命科学におけるデータ解析では、ゲノム配列データだけではなく遺伝子発現データやDNA修飾のデータ（エピゲノム）、タンパク結合など多種多様で膨大なデータを組み合わせる解析が行われる。今後開発される観測技術によってより多くのデータが蓄積されることが予想され、観測技術の進展にしたがって多様な解析ソフトウェアが組み合わされて利用されていくことになる。

本分野の研究の進展により、ゲノム情報の検索と効率的な解析法が開発され、現在は極めて高価な治療である個人のゲノム情報に基づく最適な医療（個人ゲノム医療）が、一般的な治療へとコストダウンを図ることができるようになる。