

データ科学、AI時代における HPCの新展開

2019年1月21日

理化学研究所 革新知能統合研究センター

NTT コミュニケーション科学基礎研究所

上田 修功

Naonori Ueda

情報化社会から超スマート社会へ

サイバー空間(仮想空間)と**フィジカル空間**(現実空間)を高度に融合させたシステムにより、経済発展と社会的課題の解決を両立する**人間中心の社会**(society)

新たな社会
“Society 5.0”

5.0



1.0
Society 1.0 狩猟

2.0
Society 2.0 農耕

3.0
Society 3.0 工業

4.0

Society 4.0 情報

[内閣府作成]

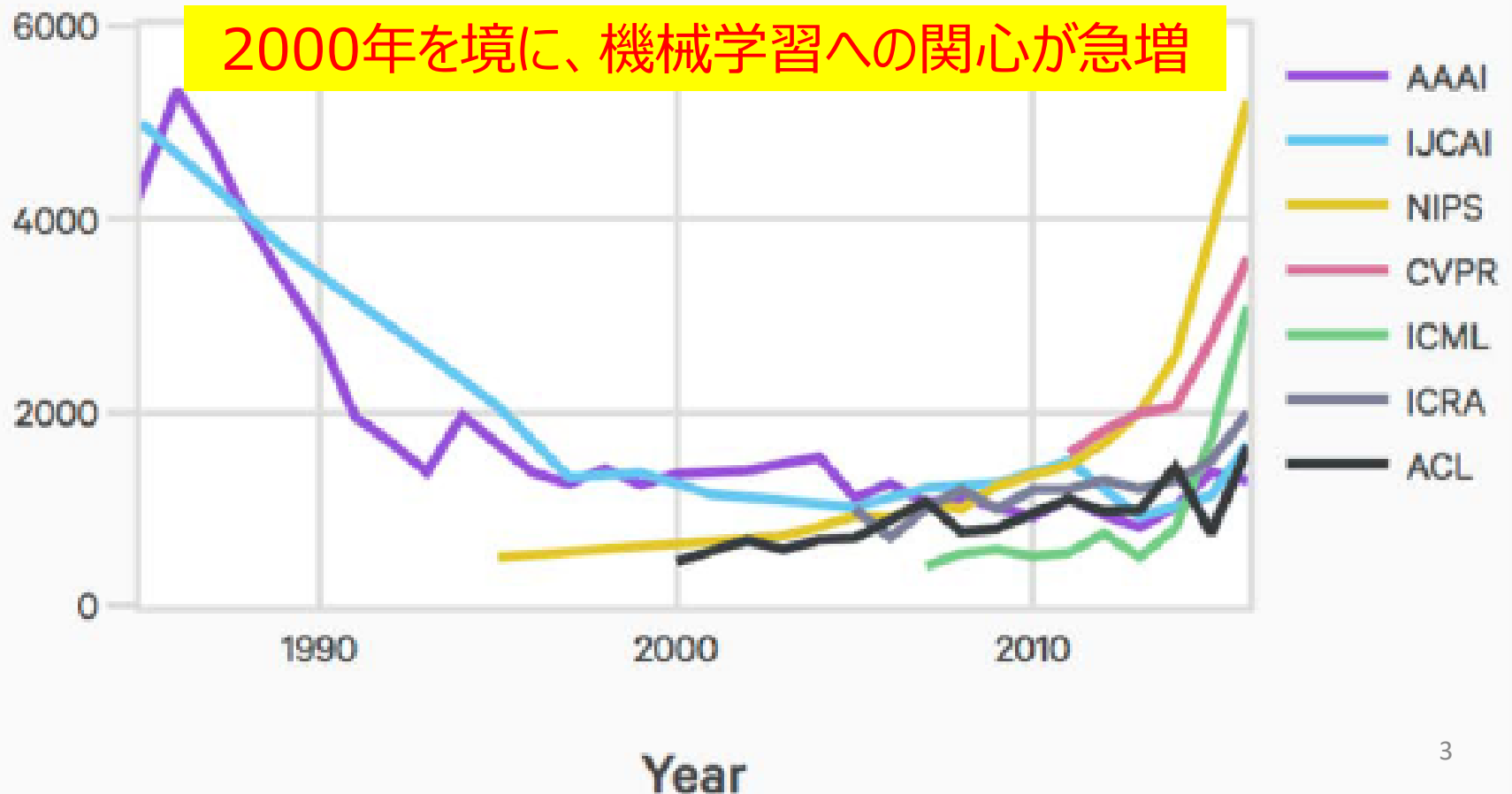
Source: 内閣府

第3次AIブーム

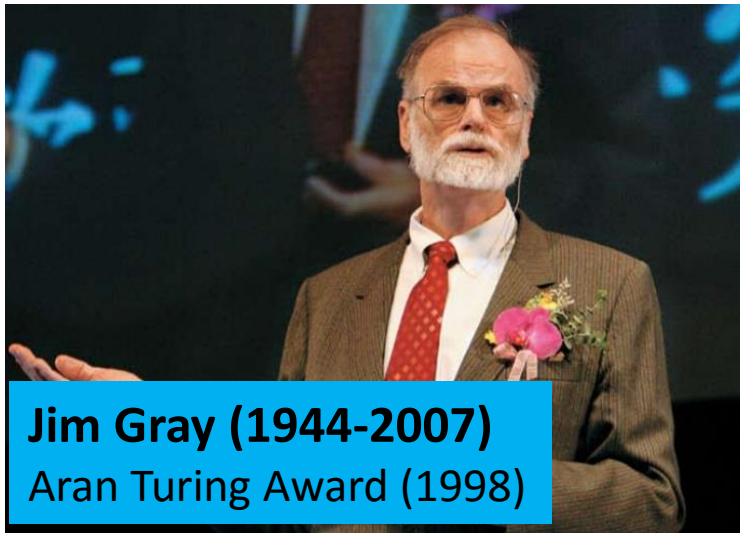
現在の機械学習ブーム = 深層学習ブーム

Large Conference Attendance

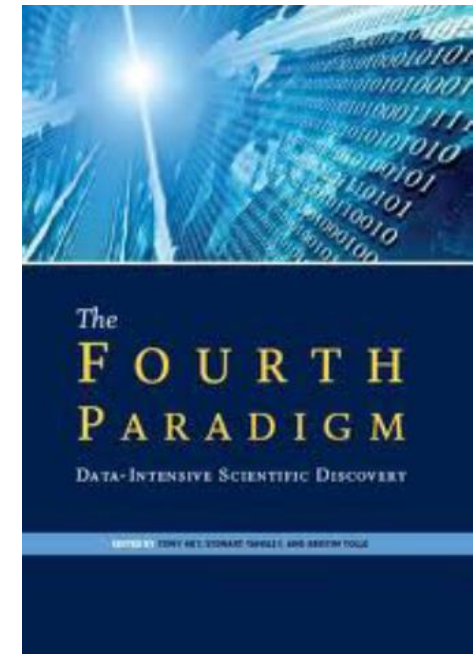
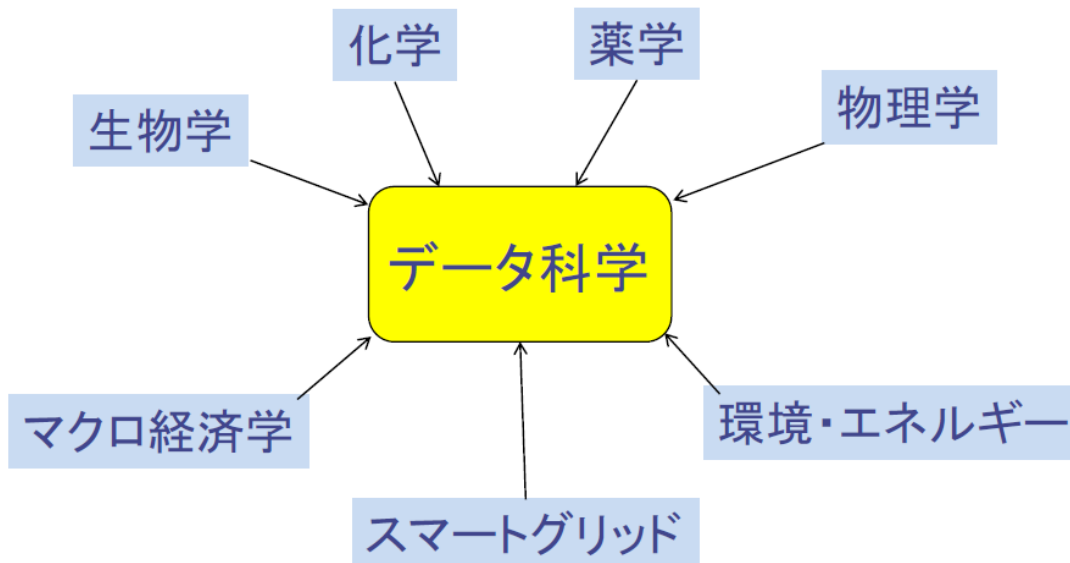
"AI Index Nov. 2017"より



データ駆動型科学 (第4の科学)



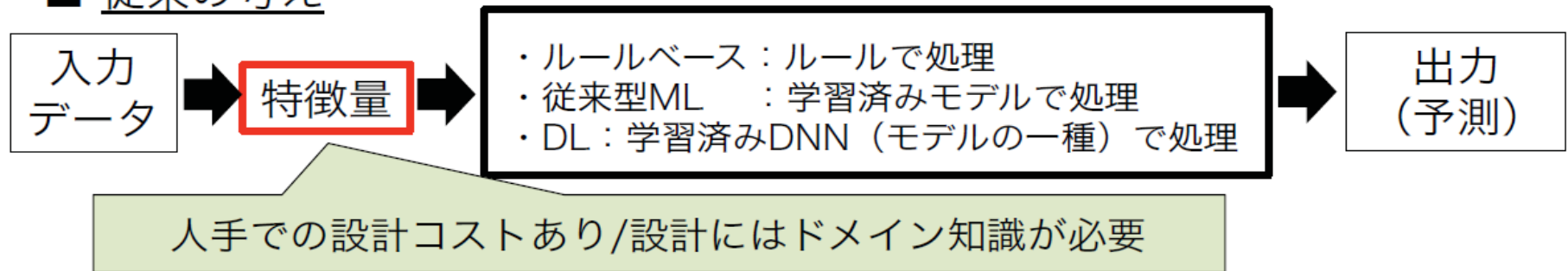
多種多様な分野において
データ科学の重要性を認識



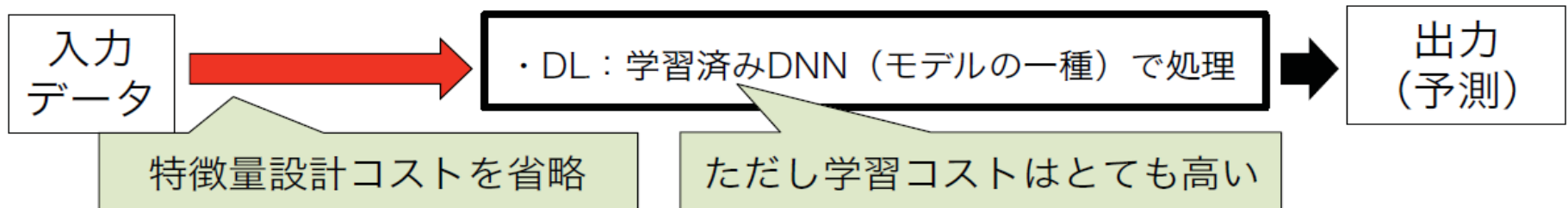
第4の科学を支える深層学習

- ・入力と教師データの対である学習データのみを必要とし、モデルの構築は不要（データ任せ）
- ・プロフェッショナルな知識は不要
- ・ 深層学習は、特徴量の抽出ステップを省略しても高い精度が出せる。

■ 従来の考え



■ 深層学習は下記フローでも精度が出せる（[パラダイムシフト](#)）



深層学習の限界

- ・観測データと教師信号が**大量に与えられた教師有り学習**では **最強**のツール（画像、音声、自然言語処理では、当深層学習は**必須要素技術**）
- ・ただし、**学習データが容易に準備できない応用**（異常現象（例：故障、災害などの希少データからの推定、予測）や、結果に対する**説明が必要な応用**ではDNNの適用は困難



- ・複雑現象の解明には、**専門家の知識（モデル）**の活用が不可欠

科学研究の変遷

1. 実験科学

アリストテレスの天動説（自然哲学）

2. 理論科学

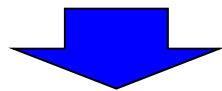
ニュートン（プリンキピア）

3. 計算科学

数値計算, シミュレーション

4. データ集約型科学（第4の科学）

データからの仮説発見



what's next?

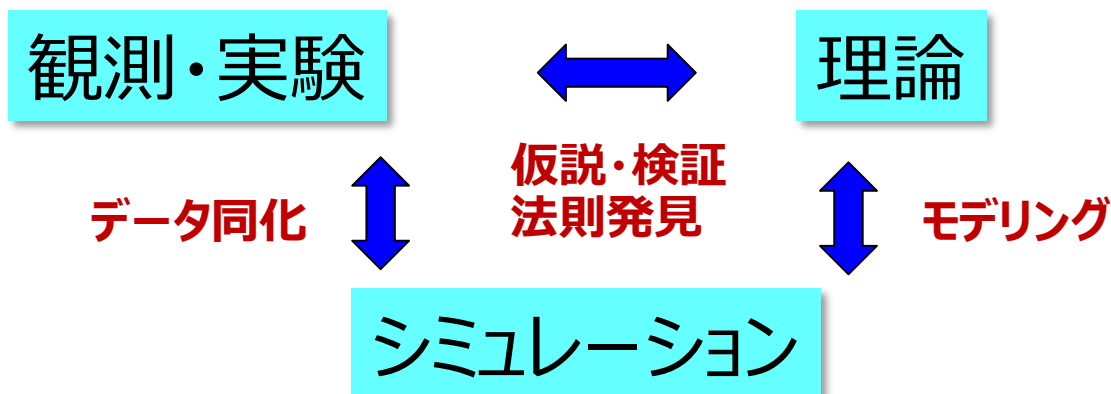
シミュレーション科学

伝統的な科学研究： 人手による仮説策定と修正の試行錯誤

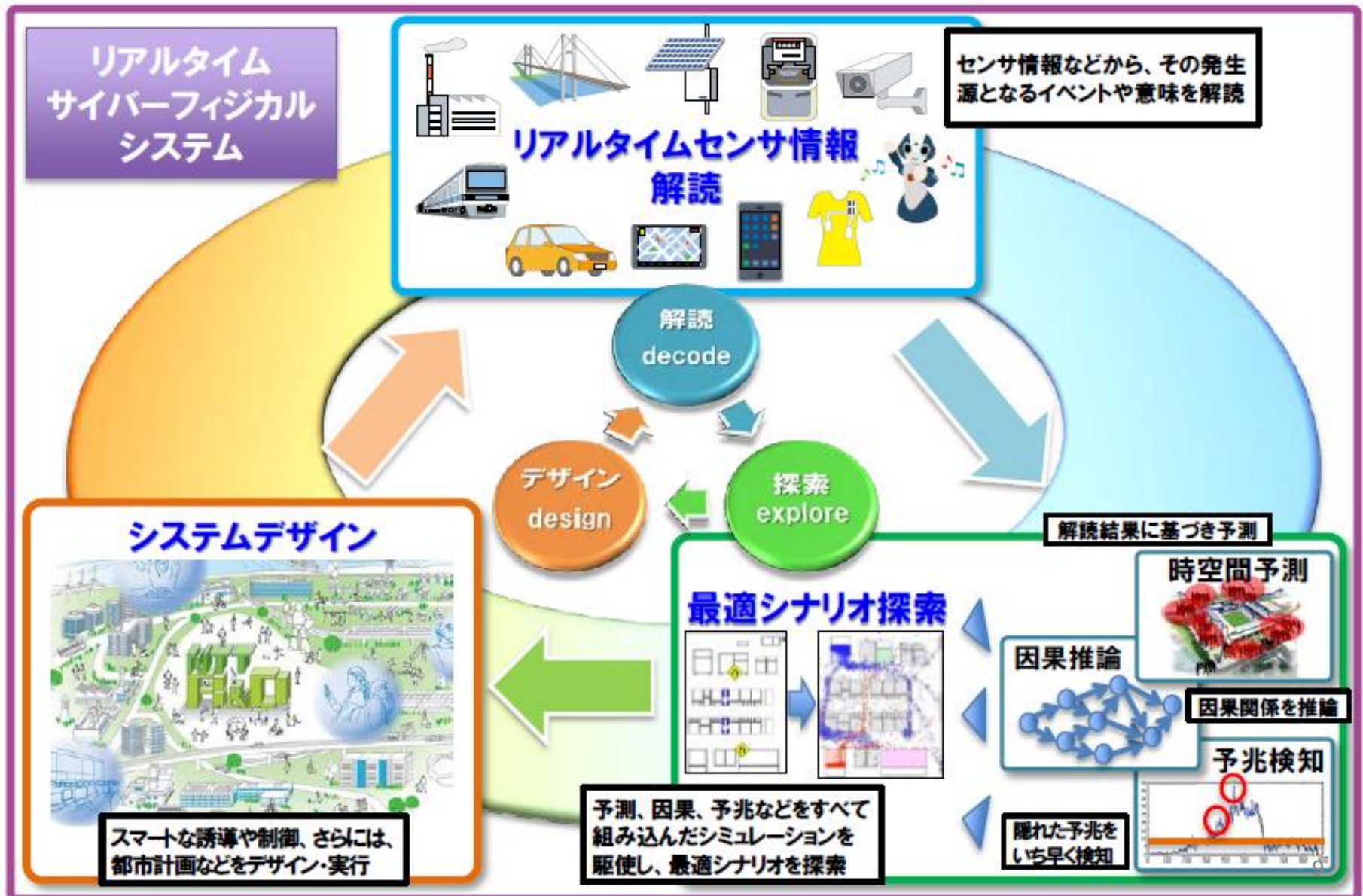


シミュレーションデータ駆動型科学研究：

少数観測データから、シミュレーション技術により
ビッグデータを創出し、新たな知識を自動獲得



IoT-AI社会 : Ambient AI



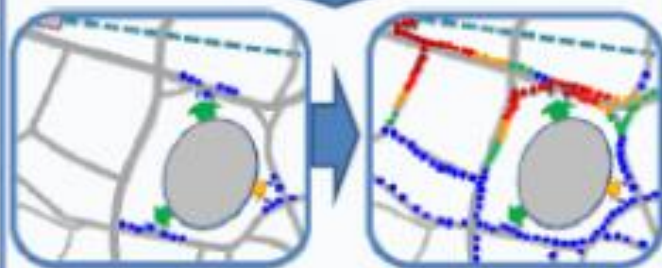
(例) 学習型マルチエージェントシミュレータ(MAS) NTT

従来型MAS*



パラメータ

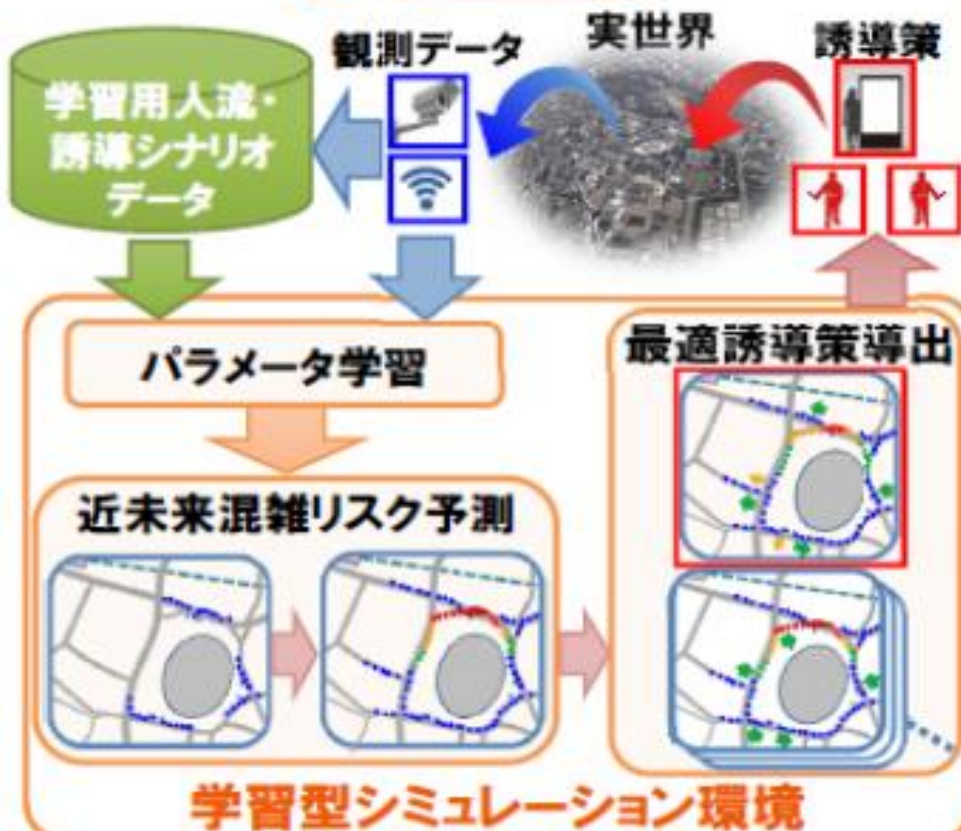
誘導シナリオ



シミュレーション環境

- MASのパラメータや誘導シナリオを手入力
- あらかじめ決められた誘導シナリオのオフライン評価

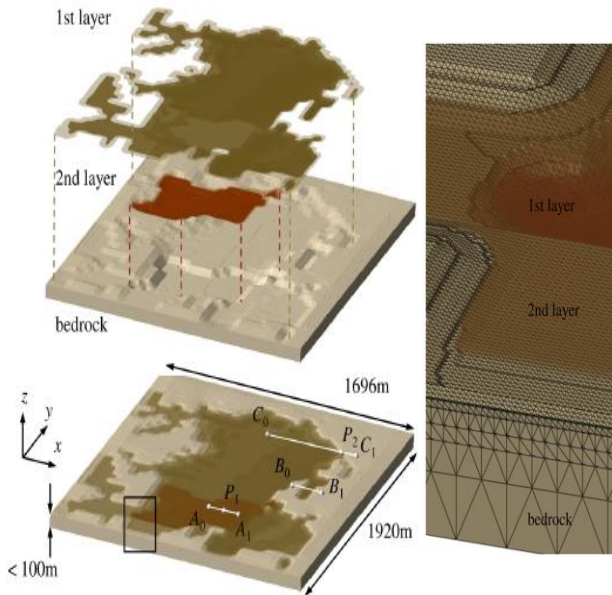
学習型MAS



- MASの**パラメータ**を観測データに基づき**自動学習**
- 事前の誘導・What-ifシナリオを用いた学習により、混雑リスクを回避可能な**最適誘導策**をオンラインで随時自動導出

(例) 地震動予測

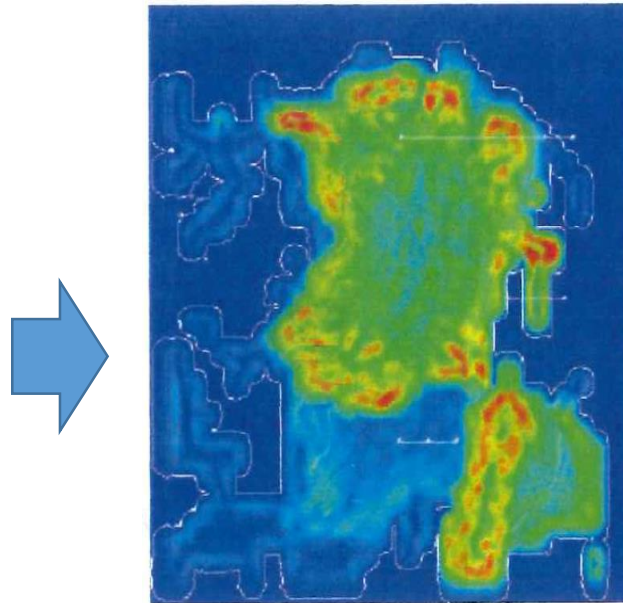
3D地盤モデル



3D分布

- V_P : P-wave speed
- V_S : S-wave speed
- ρ : mass density

地震動シミュレーション



FEM IIによる大規模シミュレーション

$$V_P^2 \Delta \phi \quad \phi = \nabla \cdot \mathbf{u}$$

$$\ddot{\psi} = V_S^2 \Delta \psi \quad \psi = \nabla \times \mathbf{u}$$

地震被害推定



建物、インフラ被害
予測

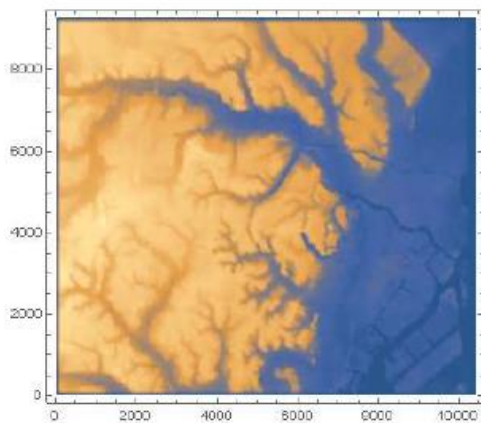
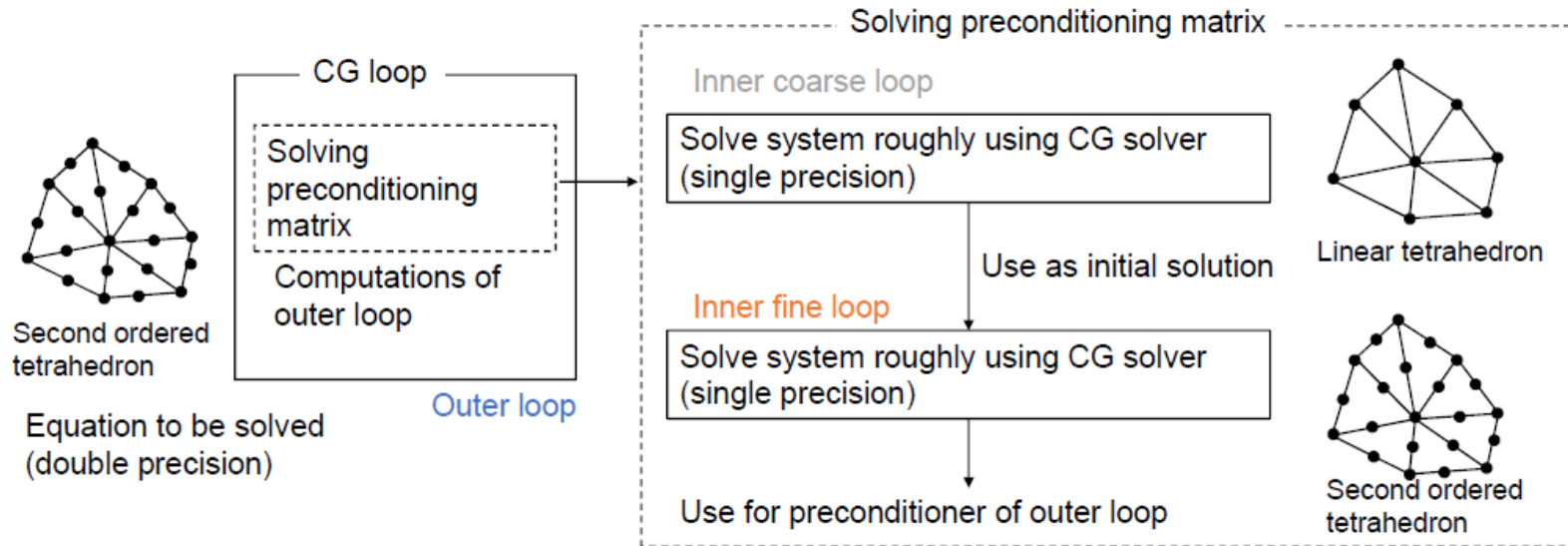
メッシュ間隔
<100 m

市村准教授（理研RCCS/東大地震研）との共同研究

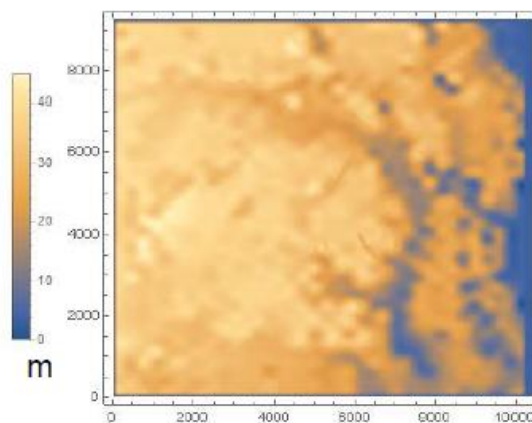
HPC上での物理シミュレーション

HPC手法の開発

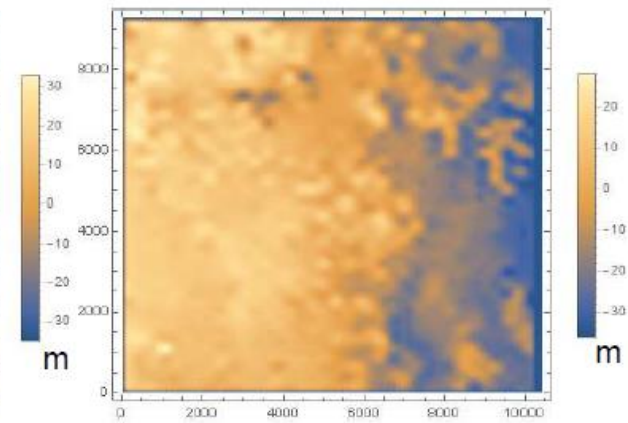
SC15 ゴードンベル賞ファイナリスト^{ref6)}



地表面標高



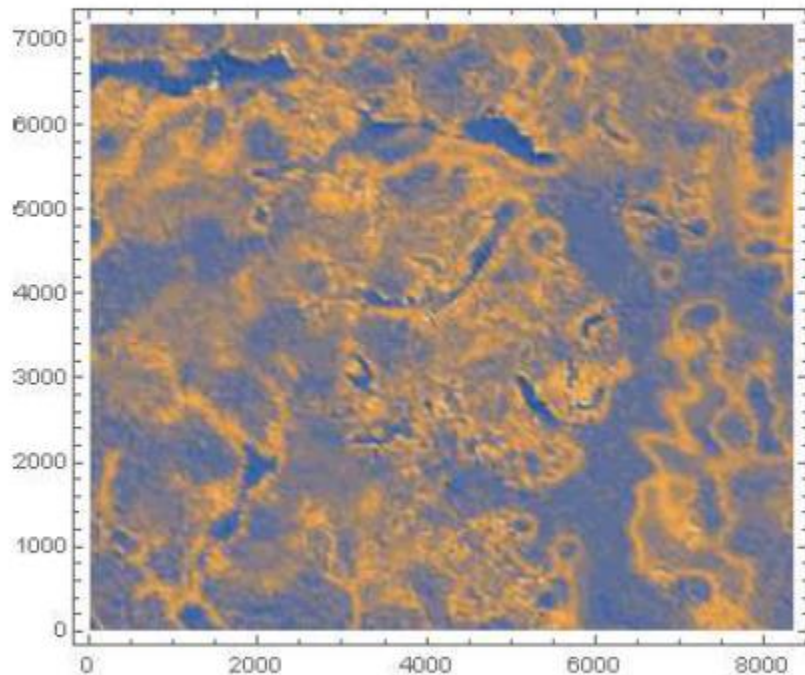
1層目と2層目の層境標高



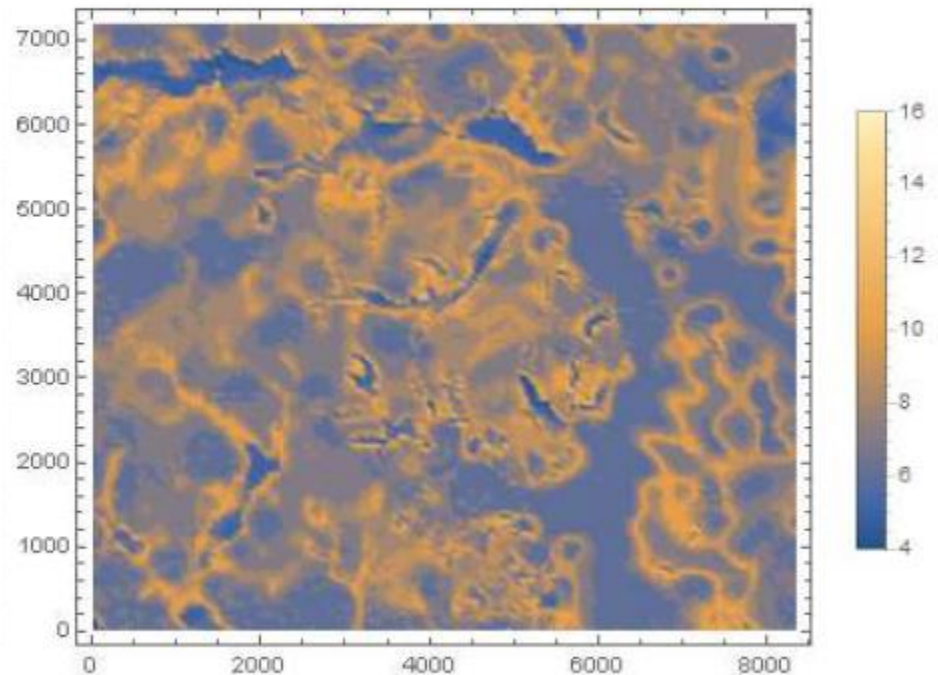
2層目と3層目の層境標高

AI(CNN)による圧倒的な効率化実現

Super-computed dataとAIによる地震動強度分類



HPCによるモンテカルロ
シミュレーションによる
推定結果

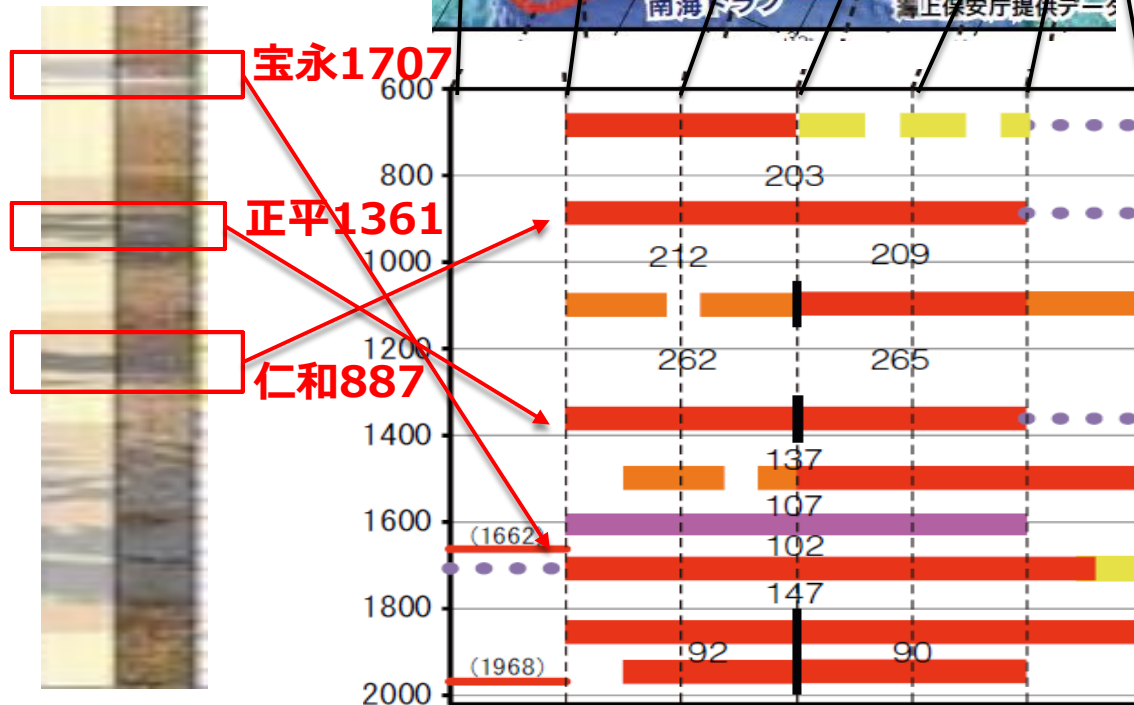


1回のHPCシミュレーション
+AIによる推定結果

SC17 Best Poster Award受賞

シミュレータを用いた南海地震予測

津波堆積物
年代を調べる



過去帳
慰霊碑
手記
などの古文書の記
録を基にする

- 確実な震源域
- 確実視されている震源域
- 可能性のある震源域
- 説がある震源域
- 津波地震の可能性が高い地震
- 日向灘のプレート間地震 (M7クラス)

陸上の湖沼

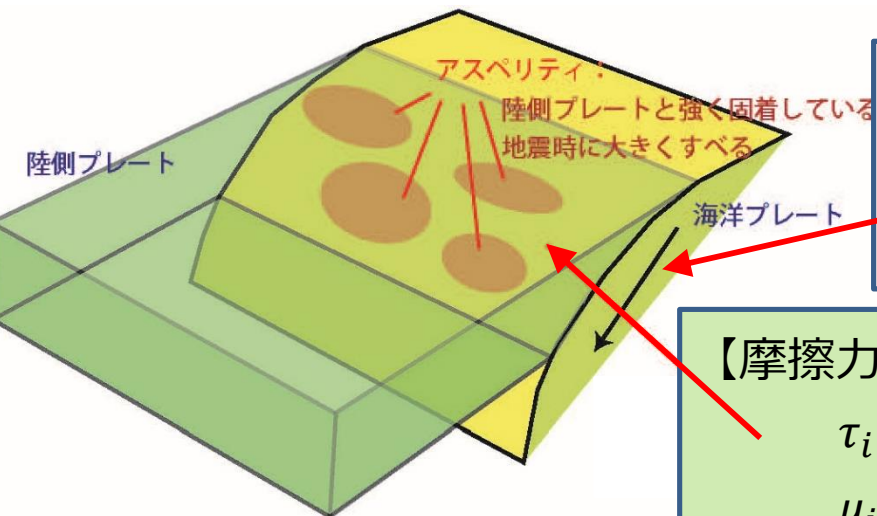
日向灘 南海 東海

地震本部

- 目的：過去の地震を再現する摩擦パラメータを求めて、シミュレータを用いて将来の地震を予測する

プレート型地震発生予測

- 海洋プレートのせん断応力と陸側プレートの摩擦力の運動方程式



【せん断応力（押す力）】

$$\tau_i(t) = \sum_{j=1}^N K_{ij} \left(v_j^{pl} t - u_j(t) \right) - \frac{G}{2\beta} \frac{du_j(t)}{dt}$$

【摩擦力（止める力）】

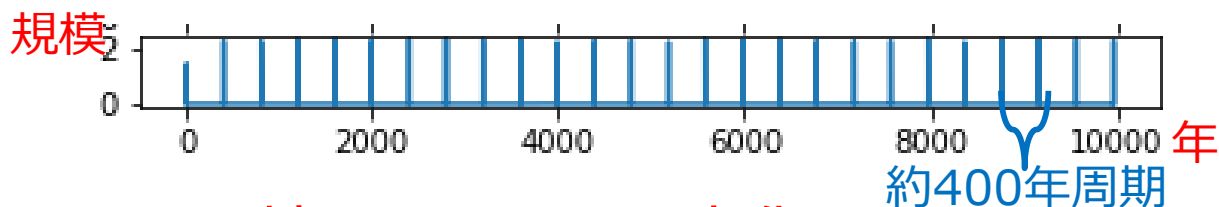
$$\tau_i(t) = \mu_i(t) \sigma_i^{eff}$$

$$\mu_i(t) = \mu_* + a_i \ln \left(\frac{v_i(t)}{v_*} \right) + b_i \ln \left(\frac{v_* \theta_i(t)}{L_i} \right)$$

○：摩擦パラメータ

- 運動方程式を用いて地震の発生サイクルと規模をシミュレーション

南海プレートの日向灘の地震サイクルの例

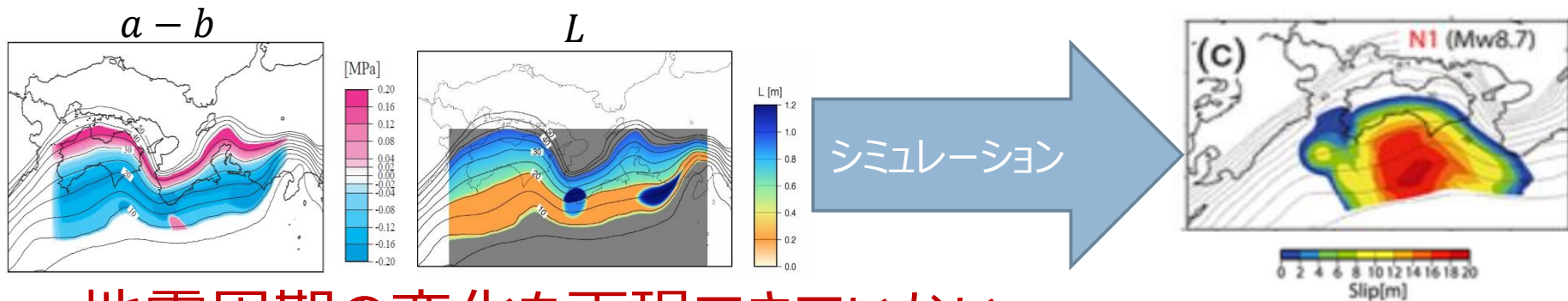


- 摩擦パラメータ a_i 、 b_i 、 L_i により地震のサイクルを変化

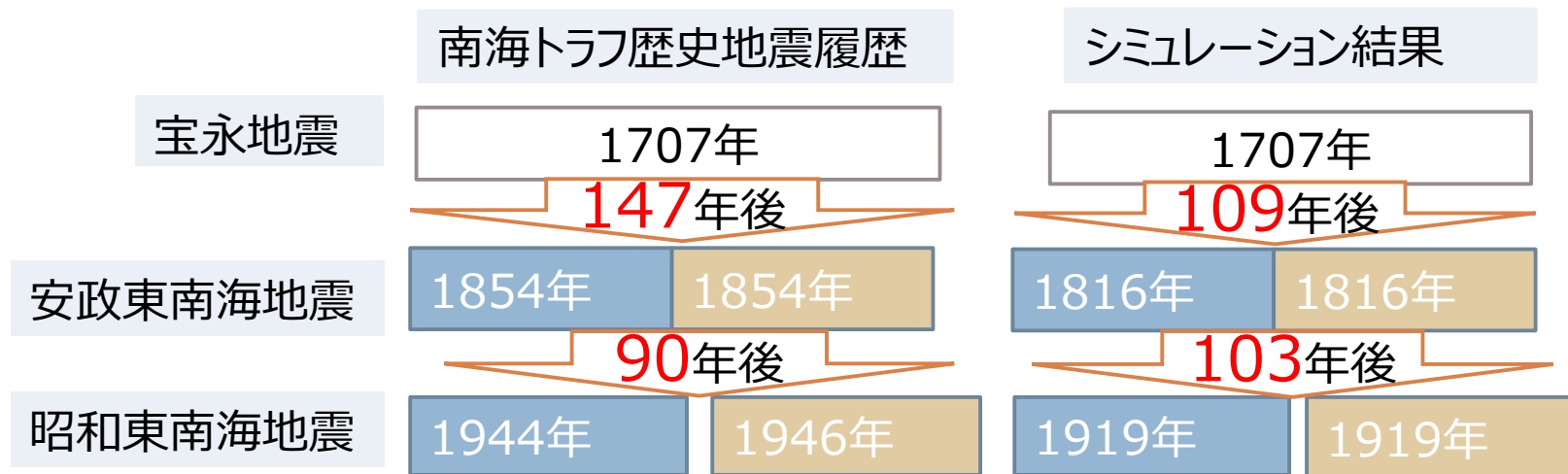
南海トラフ地震予測の現状

Hyodo et al. (2016)

- 温度分布・断層構造に基づき摩擦パラメータを設定



- 地震周期の変化を再現できていない



- 課題：人手の仮説に基づくため限定的な摩擦パラメータのみ探索

南海トラフ地震発生予測

簡易的に南海プレートを8個のセルに分割



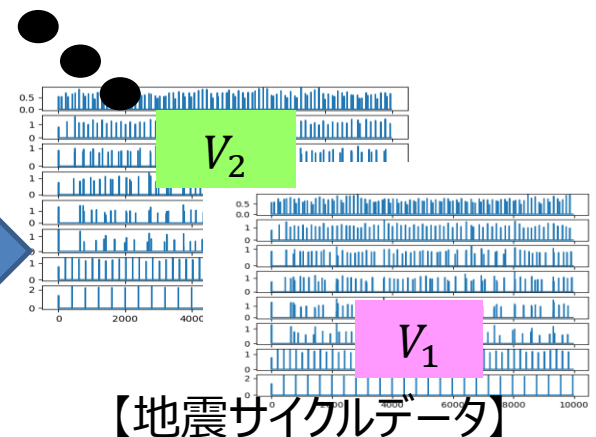
- R1とR2セルの摩擦パラメータ b_1^i と b_2^i を網羅的にサンプリング
 - 他のパラメータは初期値に固定
 - 9000個のデータを作成し、学習と評価

$$(b_1^2 = 0.012, b_2^2 = 0.011)$$

$$(b_1^1 = 0.011, b_2^1 = 0.011)$$

【R1とR2の摩擦パラメータ】

シミュレーション

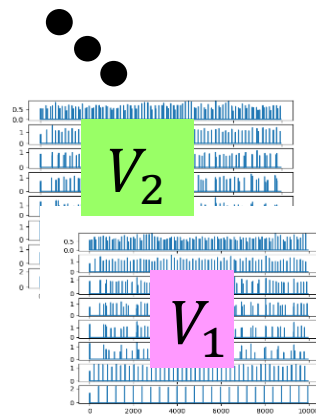


【地震サイクルデータ】

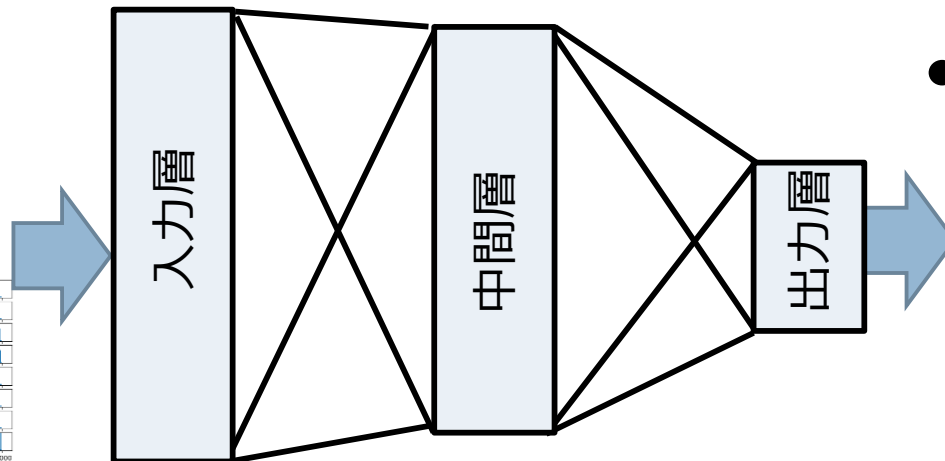
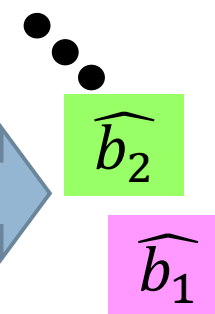
ディープラーニングを用いた逆関数近似

- 多階層のニューラルネットワークを用いて逆関数 $F^{-1}(V)$ を近似

【地震サイクルデータ】



【摩擦パラメータの予測値】

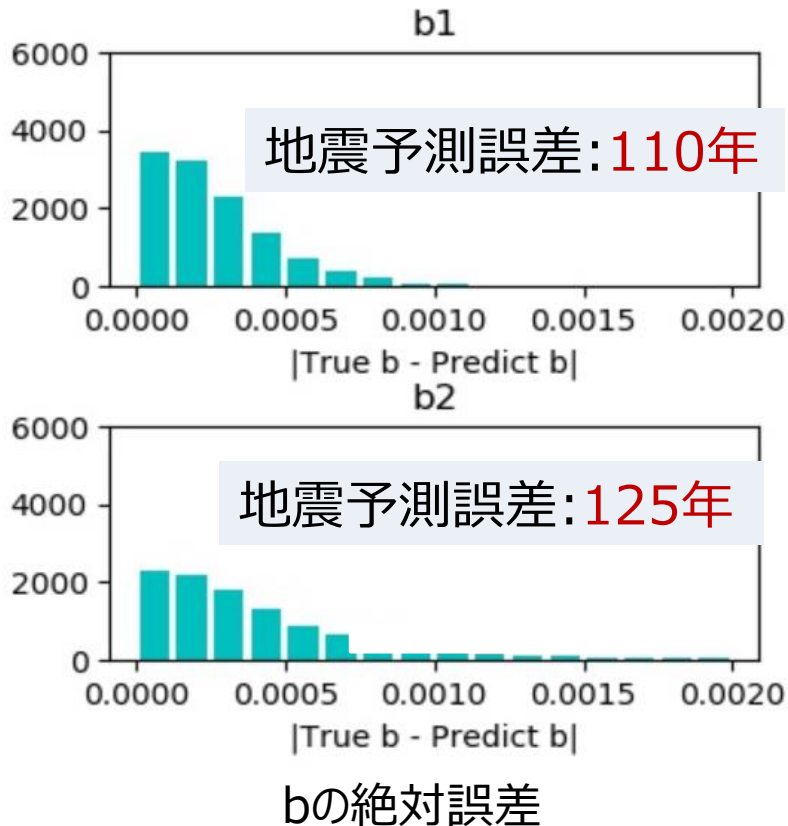


- 摩擦パラメータは小数第5位で変動
- 通常のディープラーニングで微小な変化をとらえるのは困難

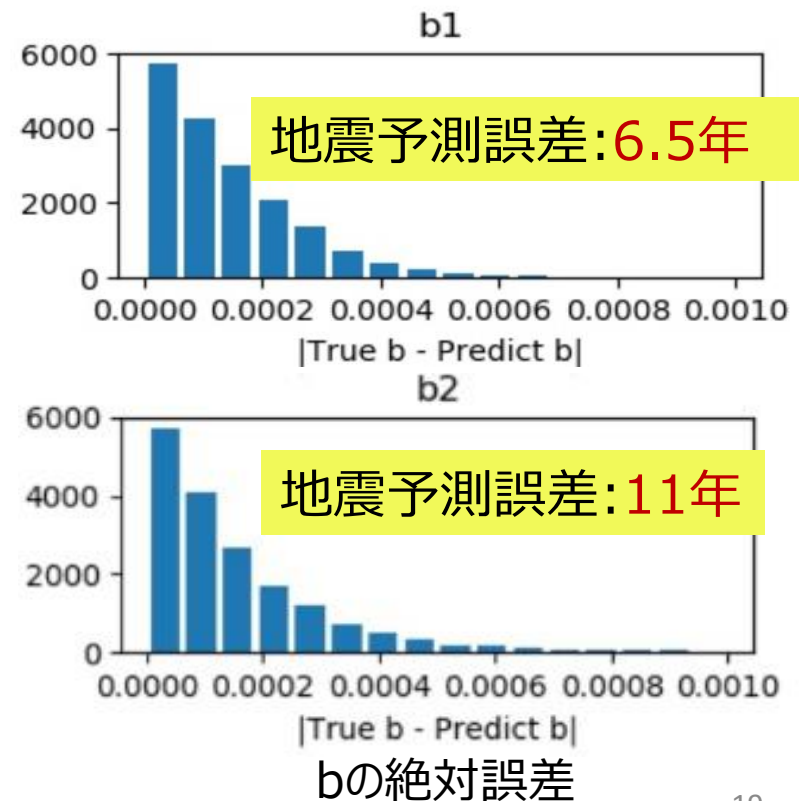
新アプローチ

- 真の摩擦パラメータ値と予測値との誤差の比較
予測誤差を大幅に改善！

通常のディープラーニング



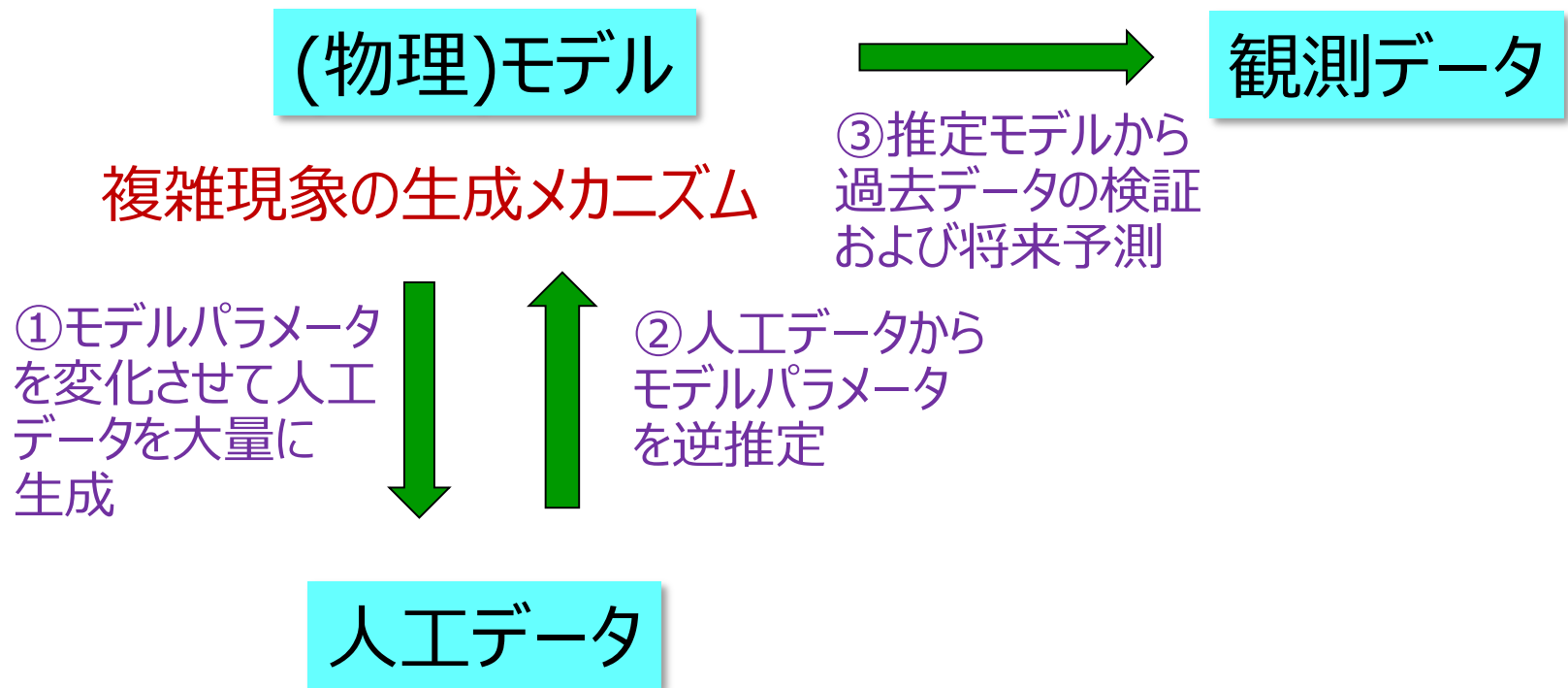
提案法



まとめ

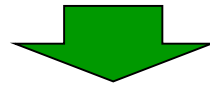
シミュレーション+AI = シミュレーション科学

HPCは、順モデルの可視化のみならず、**シミュレーションデータ**からの**帰納的学習**による**真のモデルの推定**と将来予測に有用



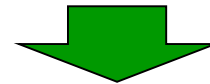
補足：HPCをより普及させるには

大半のデータサイエンティストはPython環境を使用
(∵ 多数の有用なライブラリーがフリーで使用可)



HPC上で同様なインターフェースの整備が必須
コンテナ技術

代表的なコンテナ仮想化(Docker)では、並列処理におけるコンテナ間でのMPI通信がボトルネックとなる



HPCに特化したコンテナの実装がmust