

深層学習を活用した階層・分散型データ同化システム

大西 領¹,
森 航哉², 荒木文明¹, 板倉憲一¹, 川原慎太郎¹,
中田亮太², 高橋紀之², 高橋桂子¹

¹J海洋研究開発機構 地球情報基盤センター

²NTT未来ねっと研究所

2017/9/25

第2回理研データ同化ワークショップ
理化学研究所 計算科学研究機構

Motivation

『地球情報基盤センターは、「海洋地球インフォマティクス (Marine Earth Informatics)」という新しい分野を切り拓き、**海洋地球科学のデータと情報をイノベーションへつなげる挑戦**をしていきます。』(高橋桂子センター長)

Challenge:

超スマート社会実現のための予測情報インフラの構築

Construct the prediction information infrastructure for future smart society

Objective:

- ▶ IoT気象観測ビッグデータ、および、人工知能／機械学習 を活用した、新しい気象予測システムを開発する。
- ▶ To develop a smart weather prediction system that utilizes IoT weather big-data and AI/ML technology.

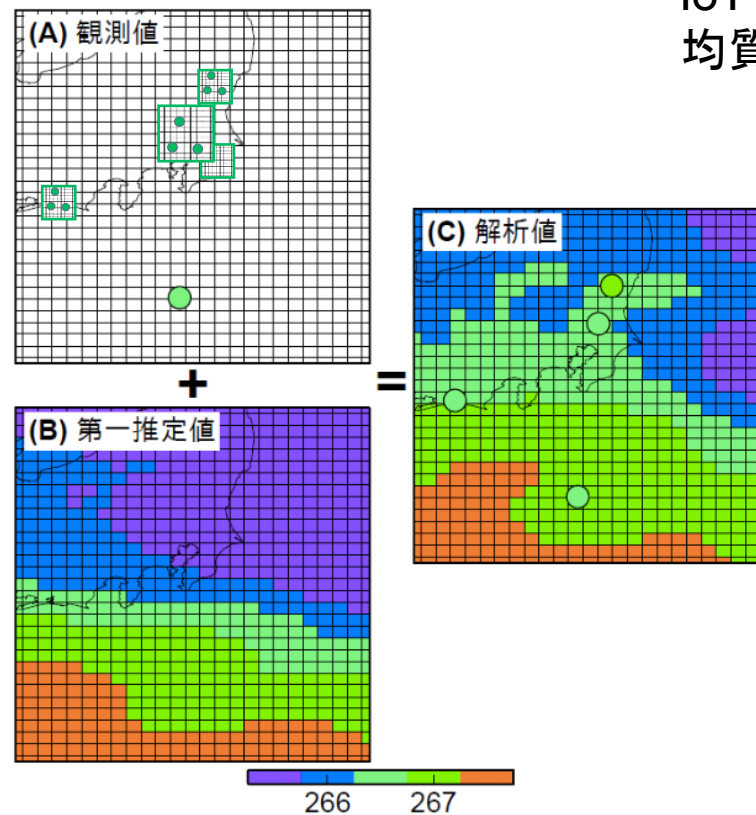
IoTセンサーデータの特徴、問題点

- ▶ 大量にある、しかし、地表付近に集中しがち
- ▶ 不均質、代表性が小さい
 - ▶ 例:アメダスによる気象データは数kmの領域を代表するのに対し、身近なIoTセンサーは数10～100m程度の領域しか代表できない。
- ▶ センサーが移動するかも

これら全てを解決出来るシステムを提案する

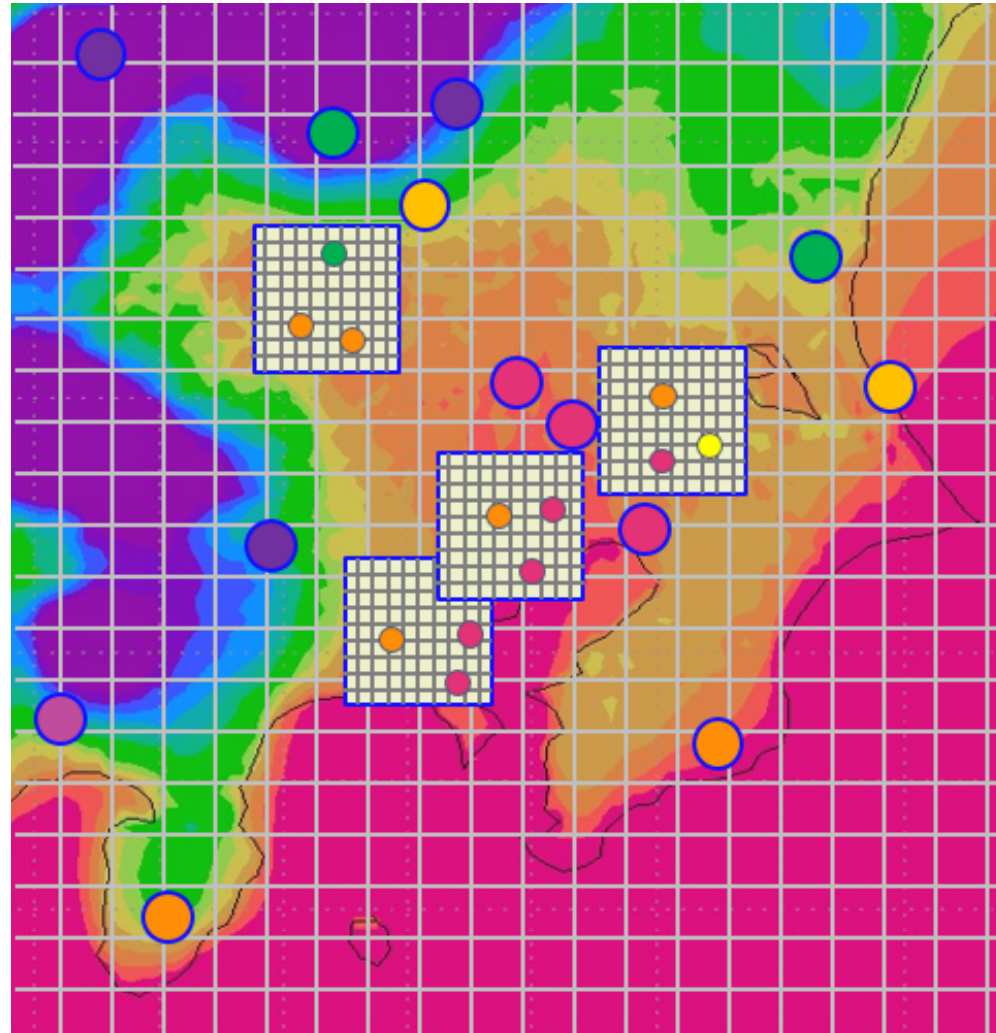
データ同化

IoTセンサーデータは不均質、代表性が小さい



www.jma.go.jp/jma/kishou/books/nwptext/45/1_chapter3.pdf

Multi-scale Data Assimilation (MDA)



IoT気象データの活用

IoT(モノのインターネット)センサーによる気象データ

- ✓ 大量 (big)
- ✓ 不均質 (variety)
- ✓ 移動 (mobility)

アメダスデータ、衛星データとは性質がまったく異なる

制御、検定、調整 ↑ ↓ 同化

エッジコンピューティングによる超高解像度、狭域(県~市単位)、短期予測

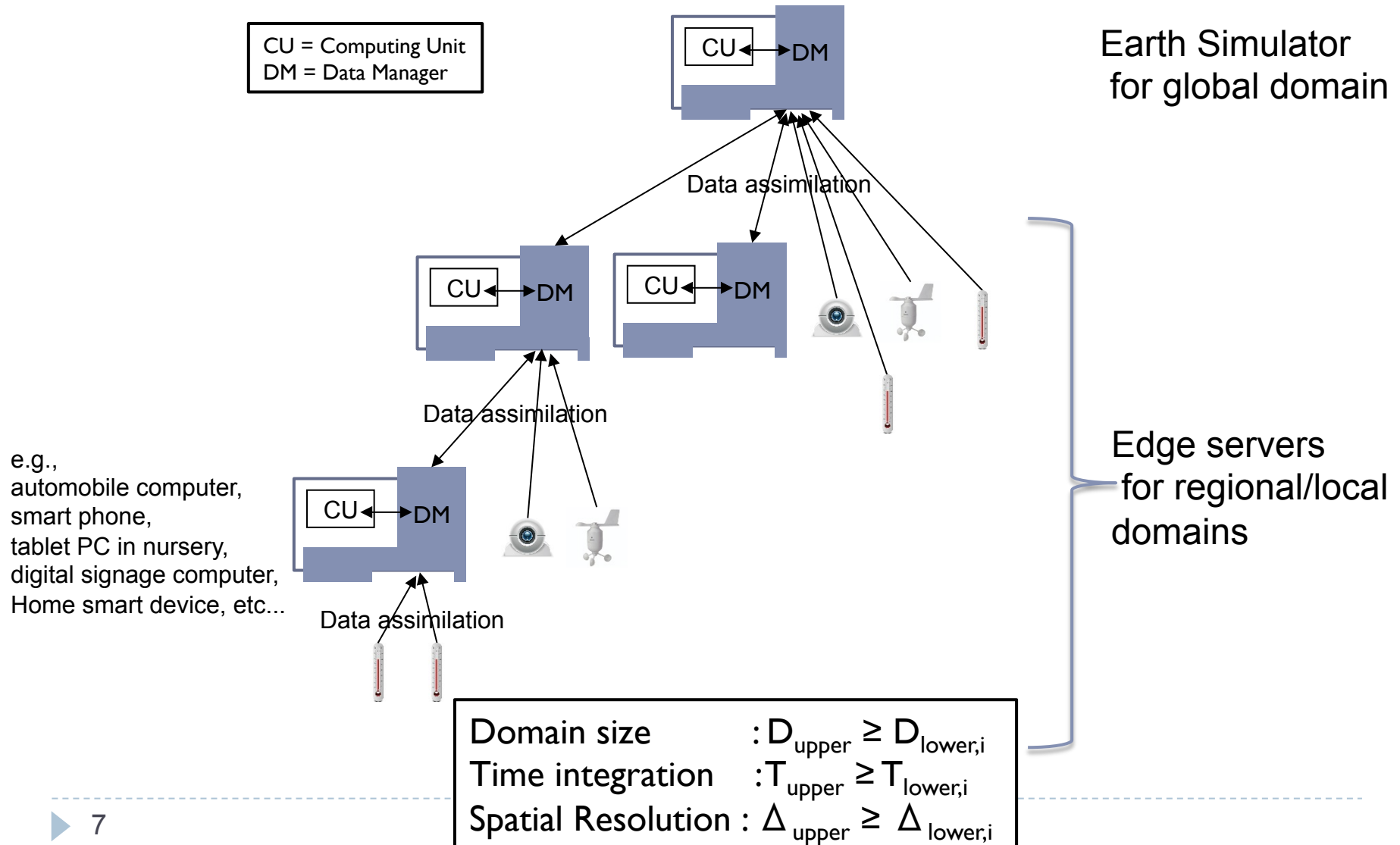
境界条件 ↑ ↓ 加工された準均質データ

地球シミュレータによる高解像度、広域(全球~日本全域)、長期予測

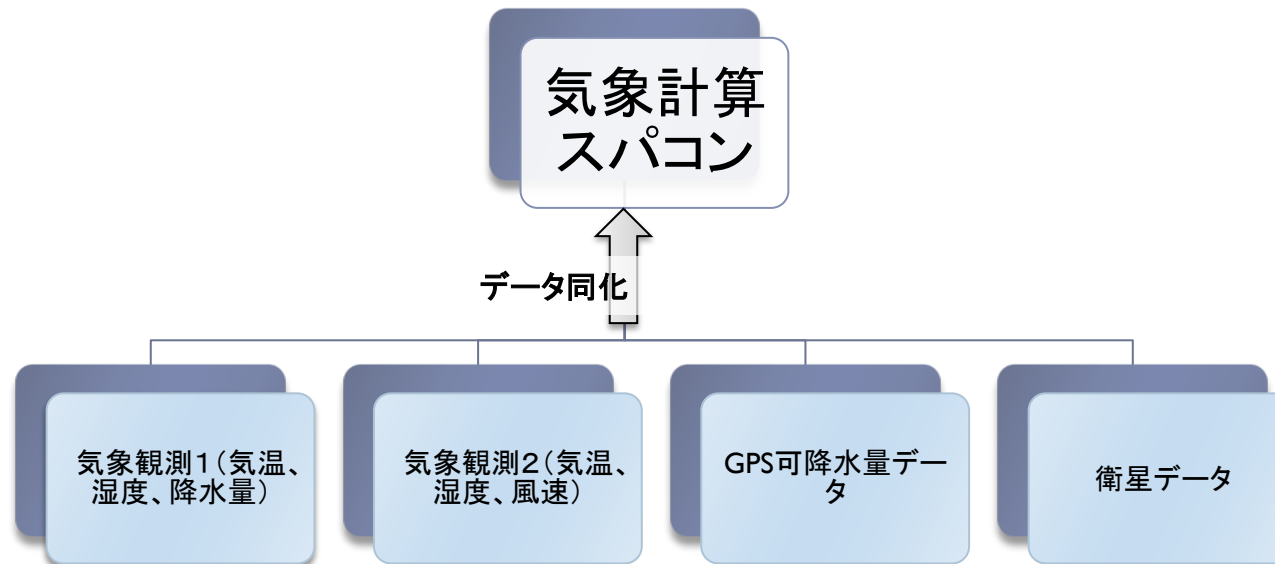
データ爆発
直接活用するのは困難

エッジコンピューティング層をバッファ(緩衝)層として利用し、データ爆発を抑えつつ、大量のIoTデータを活用する。

Multi-scale Data Assimilation (MDA)



一般的なデータ同化システム



最新研究例:

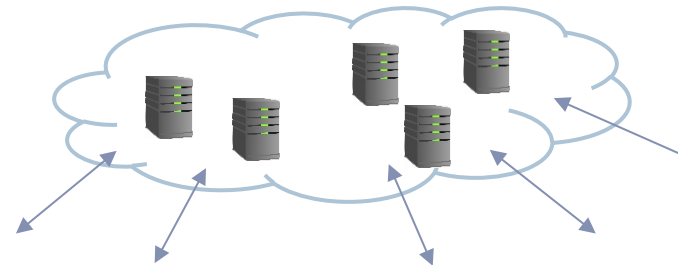
- アメダスデータ、衛星データをデータ同化に利用(気象庁の現業気象予報システム)。
- ひまわり8号のデータを活用したデータ同化システム(気象庁気象研究所の最新研究)
- 理研三好2016/8/9プレスリリース:『「京」と最新鋭気象レーダを生かしたゲリラ豪雨予測ー「ビッグデータ同化」を実現、天気予報革命へー』(http://www.data-assimilation.riken.jp/jp/research/research_press.html)、スーパーコンピュータ「京」と最新鋭気象レーダを生かした「ゲリラ豪雨予測手法」を開発。

新しいHPCの形

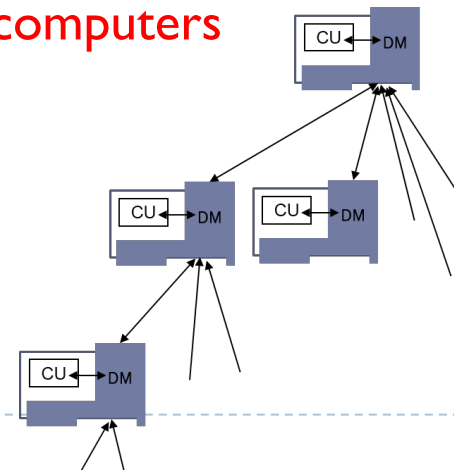
- ▶ HPC with single strong supercomputer



- ▶ HPC with cloud computers



- ▶ HPC with layered-in-network multiplatform computers



深層学習を活用した、階層・分散型データ同化システム (MDA: Multi-scale Data Assimilation)

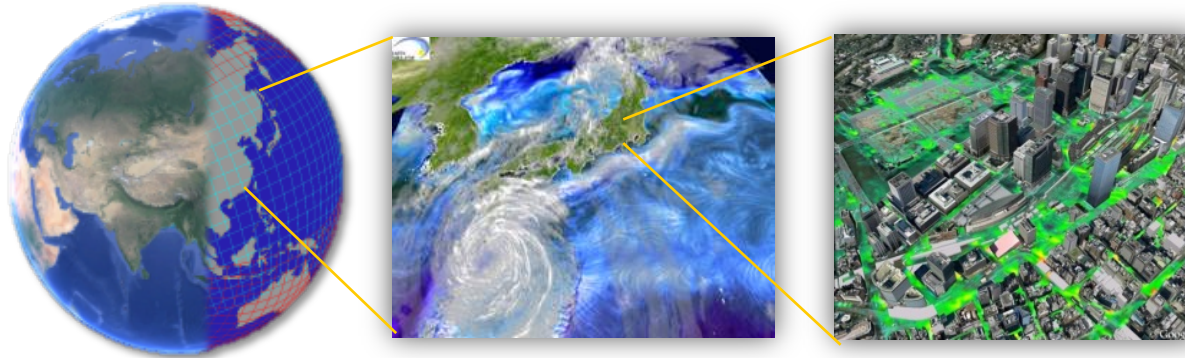
Main features:

- ▶ 大量、不均質、可動のIoTセンサーデータを活用できる。
→ 予測性能向上
- ▶ エンドユーザーに、きめ細かいオンデマンド情報を提供できる。
→ サービス向上

- ▶ 計算機資源とネットワーク資源を有効に活用できる。
- ▶ Fail safe (いくつかのユニットが壊れても、システム全体としては稼動可能)
- ▶ Data Managerをsmart化することによって、システム全体がsmart化される。

Multi-Scale Simulator for the Geoenvironment (MSSG)

Takahashi et al. (2013), Onishi&Takahashi (2012)など



地球スケール

O(1~10 km)解像度

領域スケール

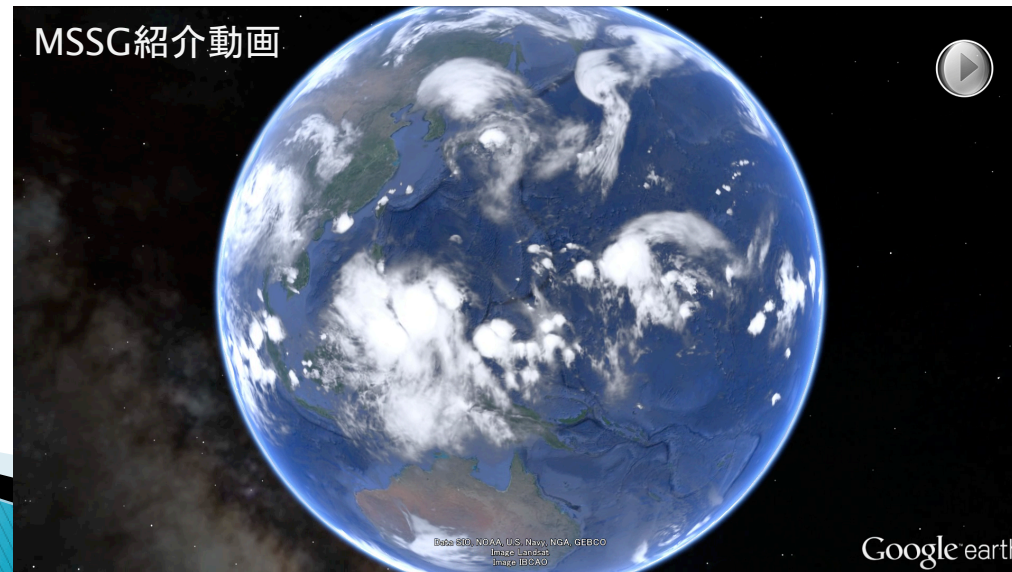
O(100 m~1km)解像度

都市スケール

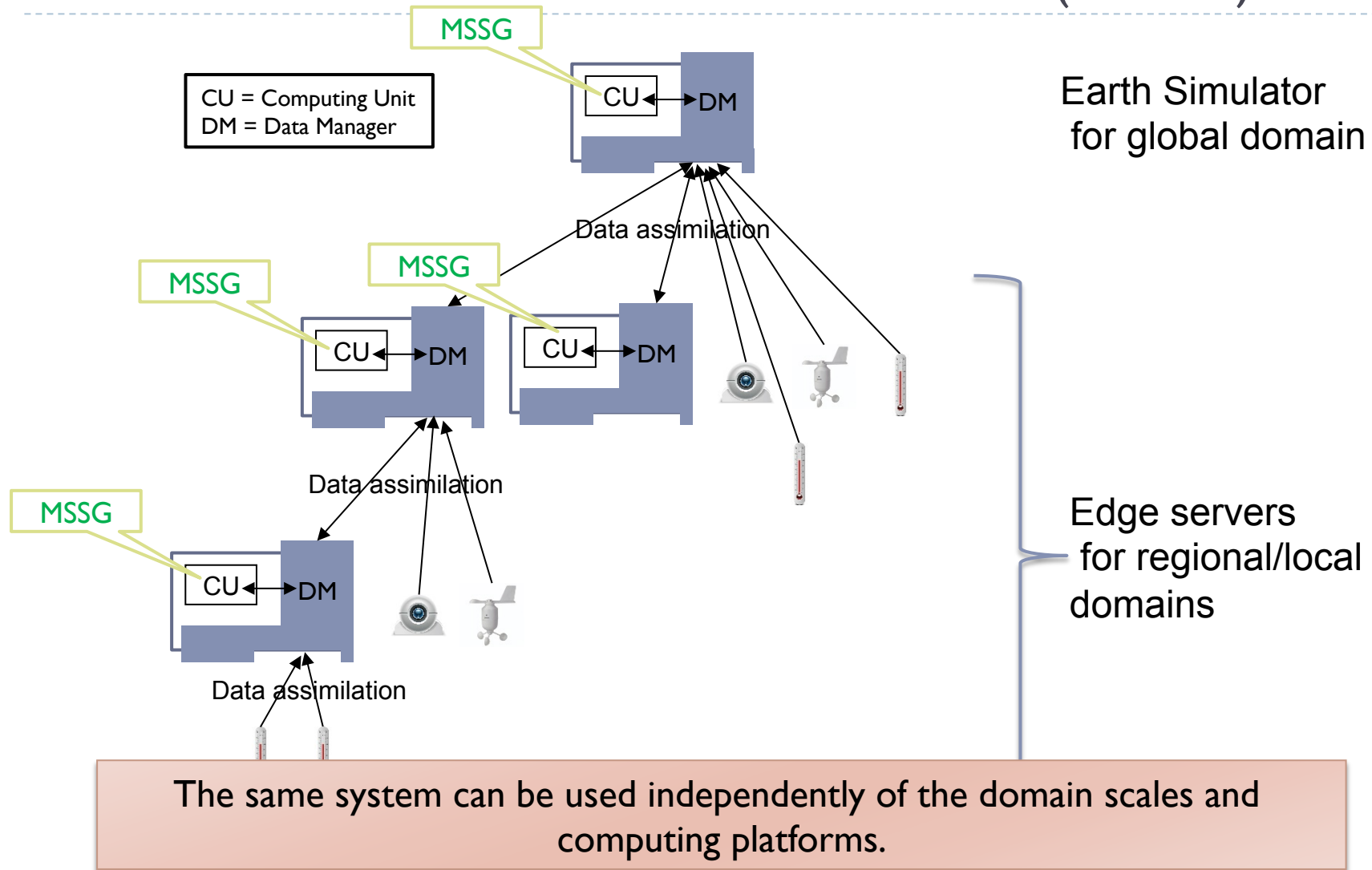
O(1~10 m)解像度

- MSSGは、地球一領域一都市のシームレスなシミュレーションが可能
- 大気、海洋単体計算だけでなく大気・海洋結合計算が可能
- 地球シミュレータに高度に最適化されている

MSSG紹介動画

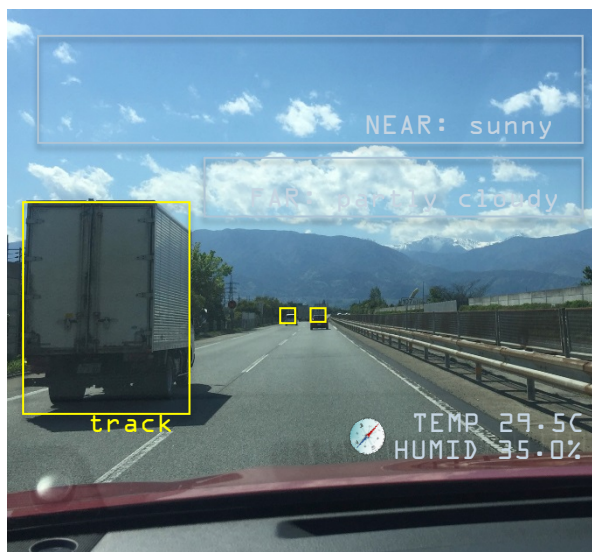


Multi-scale Data Assimilation (MDA)

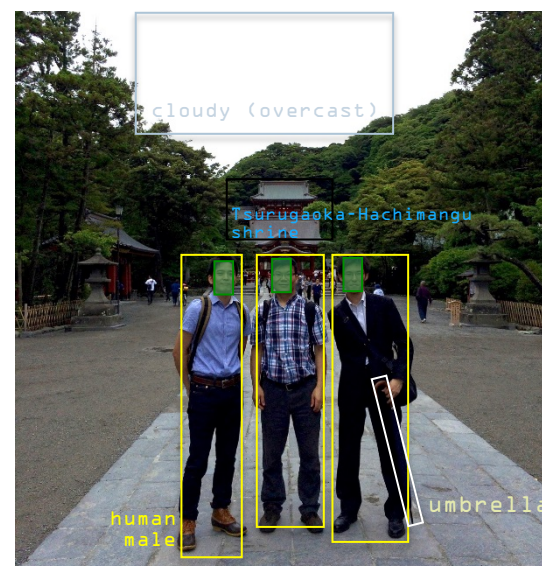


深層学習装置：IoTによる新たな気象観測網

従来型気象センサー（温湿度センサーなど）の利用、だけでなく、
例えば、各種カメラによる膨大な画像情報を活用する



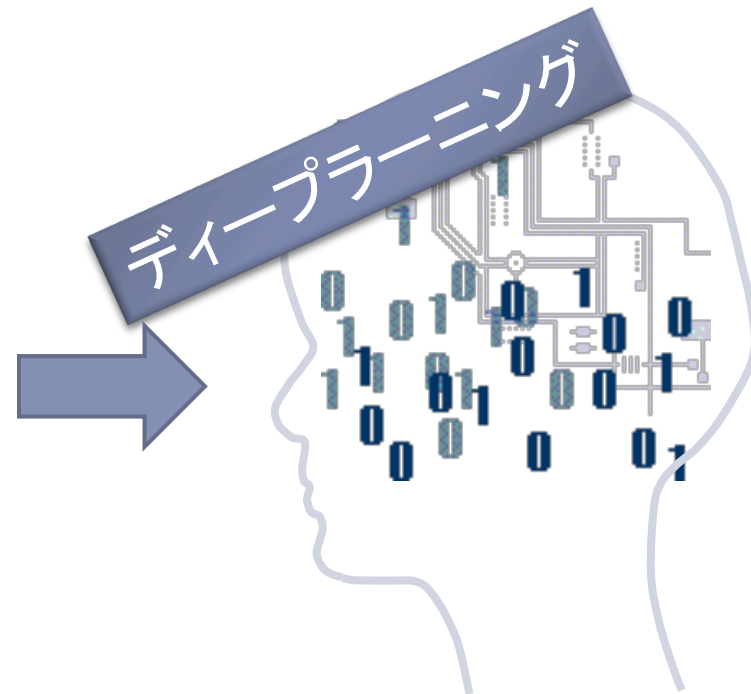
車載カメラ画像に含まれる気象情報



スマホカメラ画像に含まれる気象情報

機械学習（ディープラーニング）を用いた画像認識 ー(部分)雲量推定ー

「雲量」とは、空の全天に占める雲の割合。0(雲なし)～10(全天曇り)まで(日本方式)。



インターネット上から雲画像を自動収集し、それを学習データとした深層学習(ディープラーニング)を実施。カメラ画像から雲量を推定するシステムを開発

14
定点カメラ画像でなくとも、良好な推定精度が得られる

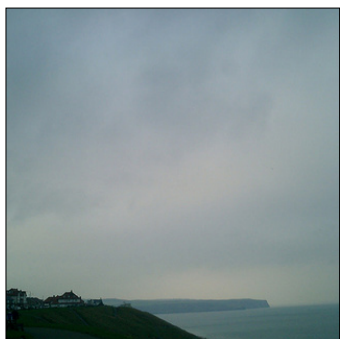
CNNによる(部分)雲量推定結果



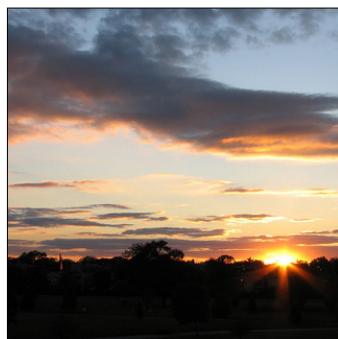
教師ラベル 0
出力 0.5



教師ラベル 6
出力 5.6



教師ラベル 10
出力 7.7



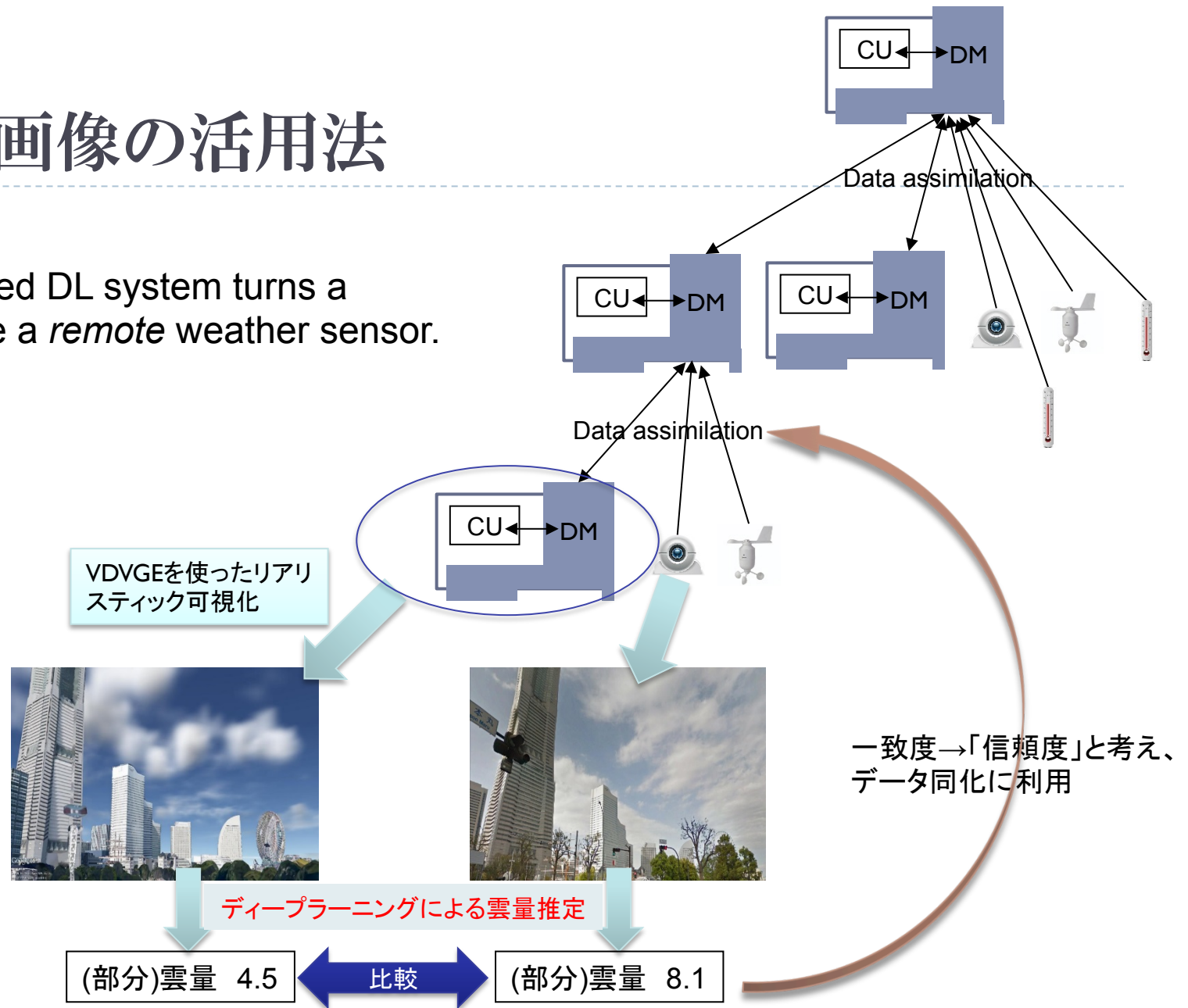
教師ラベル 5
出力 4.5

混同行列
平均2乗誤差 **3.1**

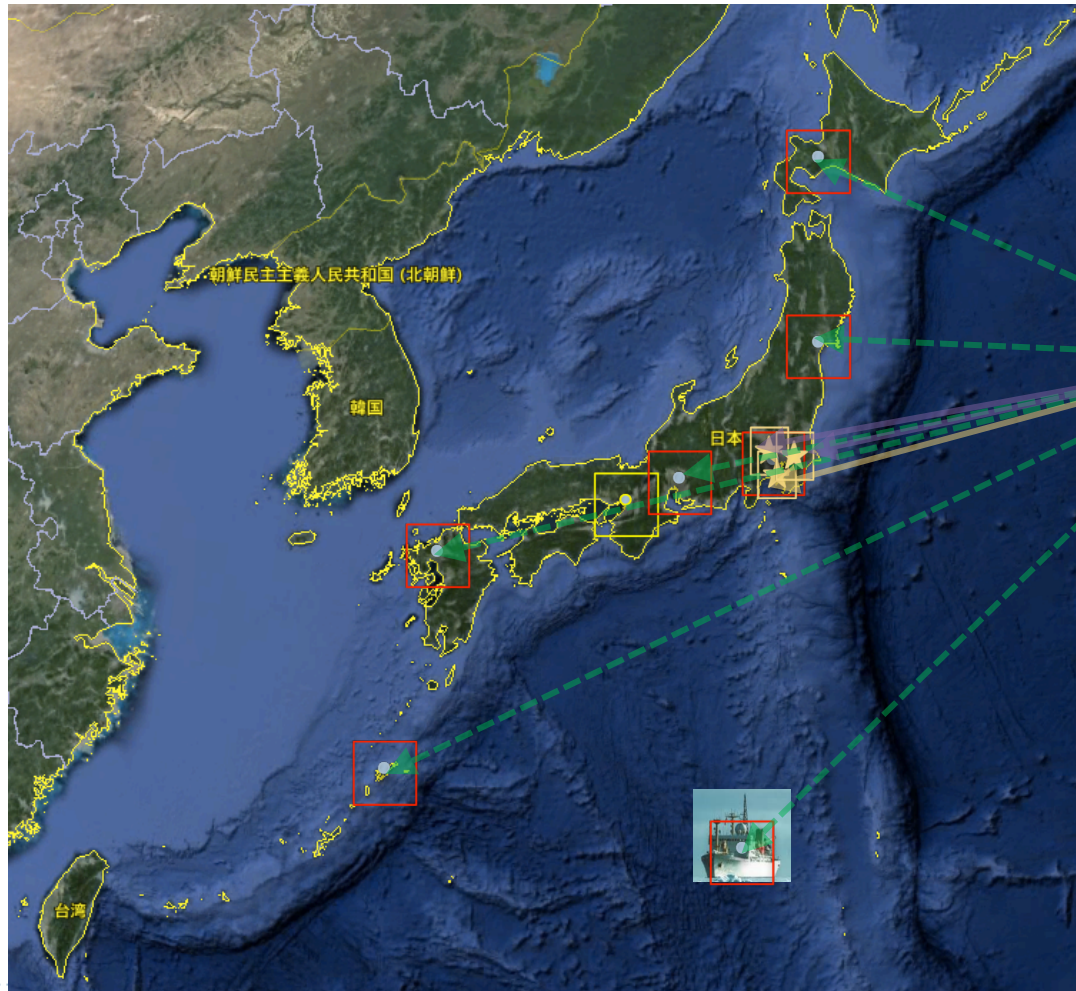
セグメンテーション法よりも大幅に良い推定精度

カメラ画像の活用法

The developed DL system turns a camera to be a *remote* weather sensor.



ESと複数エッジサーバーの連携実験



全球計算

地球シミュレータ



エッジサーバー

まとめ

- ▶ 「スマート気象予測システム」
 - ▶ 階層型データ同化システム (MDA; Multiscale Data Assimilation) を用いることによりIoTビッグデータを扱うことができる。
 - ▶ 多様なコンピュータ、ネットワークで構成されたシステム(新しいHPCの形態)。
 - ▶ マルチスケール、マルチプラットフォーム気象モデルMSSG (Multi-Scale Simulator for the Geoenvironment)を利用することにより、シンプルなシステムを構築できる。
 - ▶ データマネジャー(サイバーマネジャー)部において、**深層学習を活用**することにより、カメラをリモート気象センサー(雲量センサー)として利用できる。
- ▶ 現状、本スマート気象予測システムを用いた地球シミュレータ+エッジサーバの連携運用に成功している。今後、IoTビッグデータを利用することによる予測精度向上の確認を行う。

