

jh220055

## 数値シミュレーションと機械学習との融合による水圏生態系予測

菊地淳（理化学研究所）

### 概要

本研究では、沿岸海洋を対象として、高解像度長期間の数値シミュレーションの出力、現業の気象予報・解析データ、現場観測データを機械学習により統合した魚群の位置・規模を含む水圏エコシステムの面的・空間的予測を実施する。東京大学情報基盤センターの Wisteria/BDEC-01 (Odyssey, Aquarius) システムを使用し、シミュレーション (Simulation) + データ (Data) + 学習 (Learning) (S+D+L) 融合の促進に資する。

### 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 (該当するものを残す)

東京大学 情報基盤センター

- (2) 課題分野 (該当するものを残す)

大規模計算科学課題分野

- (3) 共同研究分野 (HPCI 資源を利用している研究課題のみ、該当するものを残す)

超大規模数値計算系応用分野

超大規模データ処理系応用分野

- (4) 参加研究者の役割分担

研究総括：

菊地淳

数値シミュレーションパート：

中島研吾, 入江政安, 三好建正,

河合直聡, 古川賢, 永野隆紀

機械学習パート：

埴敏博, 黒谷篤之, 古川賢

### 2. 研究の目的と意義

2050 年には 97 億人にも達するとされる人口爆発と資源搾取によって、SDGs やプラネタリーバウンダリー等、環境評価は重要な局面

を迎えている。特に海洋の生態系変動は著しく、サンゴの白化、赤潮発生、海藻類の磯焼け、そして生物多様性の喪失等の深刻な被害があり、複雑な生態系予測には解析技術の高度化が求められる。理化学研究所では、菊地 (本プロジェクト代表者) を中心に、東京大学、大阪大学とも協力のもと、Wisteria Odyssey, Aquarius の利用を念頭において、数値シミュレーションデータ、気象データ (JMA-MSM) [A]、東京湾内での観測データ [B]、河川データ [C] を機械学習で統合した、赤潮をはじめとする水圏エコシステム予測システムの開発を進めている。本研究では、Wisteria Odyssey による数値シミュレーションによって生成された 3 次元のビッグデータを Wisteria Aquarius による大規模計算により学習することにより、高精度な面的・空間的予測システムを構築する。さらに、環境オミクス解析による水圏生態系因子の観測データの追加により魚群を含めた空間的なエコシステム予測の開発を目指す。本研究チームは、数値シミュレーションで必要となる河川・海洋の生物場の観測・分析能力を有しており、それらを用いた生物場の高精度な数値シミュレーションが可能である。本研究は、計算科学・数値シミュレーションに加えて、データ科学、機械学習等の知

見を融合した新しい手法の適用によって、サイバー空間とフィジカル空間の融合を通じた「Society 5.0」実現に大きく貢献する。本研究より得られた知見は、Wisteria/BDEC-01 上での数値シミュレーションから機械学習への、Odyssey と Aquarius の連携事例として、他の様々なアプリケーションへ応用が可能である。

本研究における数値シミュレーションデータの構築では、入江グループで使用している領域海洋モデル(ROMS)[D]を用いる。物理場モデルは、流速、海面高度、温度、塩分などの物理場をシミュレーション変数として含む。生物場・化学場モデルは、窒素・酸素・リン等の循環に基づく低次生態系モデルを採用する。3次元シミュレーションデータと相模湾東京湾内の現場観測データ、そして現業の気象予報・解析データを

Wisteria Aquarius を用いて大規模機械学習で統合し、面的・空間的水圏エコシステム予測手法を構築する。

[A] JMBSC, <http://www.jmbsec.or.jp/jp/online/file/online10200.html>

[B] TBEIC, <https://www.tbeic.go.jp/MonitoringPost/Top>

[C] 小槻峻司ら、灌漑を考慮した陸域水循環モデルの構築. 水工学論文集, 55, 2011, 553-558.

[D] A. F. Shchepetkin and J. C. McWilliams, The Regional Ocean Modeling System: A split-explicit, free-surface, topography following coordinates ocean model, *Ocean Modelling*, **9**, 2005, 347-404.

### 3. 当拠点の公募型研究として実施した意義

本プロジェクトは 2021 年度東京大学萌芽共同研究公募課題「AI for HPC: Society 5.0 実現へ向けた人工知能・データ科学による計算科学の高度化」を経て、この萌芽研究課題での研究を発展させた、東京湾・相模湾での水圏エコシステムの面的・空間的な予測を行うことが目標である。理化学研究所、東京大学情報基盤センターらの計算科学、大規模数

値シミュレーション、データ科学、環境オミクス解析、機械学習の専門家の緊密な協力のもとに「シミュレーション+データ+学習(S+D+L)」融合を目指して実施される学際的なものである。本プロジェクトの実現には、CPU と GPU を用いた並列化や最適化手法の知見や数値シミュレーションで生成したビッグデータを機械学習で利用するため、データの受け渡しが円滑な Wisteria/BDEC-01 (Odyssey, Aquarius) の利用は最適なソリューションである。また、本研究で Wisteria は高い計算能力とデータ解析、機械学習に適したハードウェア、ソフトウェアを有し、「S+D+L」融合のためのプラットフォームとして最適である。

### 4. 前年度までに得られた研究成果の概要

本年度が初年度のため該当しない。

### 5. 今年度の研究成果の詳細

#### (i) 数値シミュレーション

今年度は、コンパイル環境の構築、大阪湾のデータを用いたモデルのテストラン、萌芽研究課題で用いていた入力データを用いた東京湾・相模湾を対象とした数値シミュレーションを実施するための計算条件に関するパラメータの探索を行った。

本プロジェクトでは、萌芽課題で用いていたものとは別開発の ROMS を使用している。東京湾・相模湾領域を対象とした数値シミュレーションの実施にあたって、入力境界条件等に異なる点が多く、東京湾・相模湾領域で用いられていた入力データをチューニングする必要があった。

本プロジェクトで利用する ROMS の実行環境の整備と計算結果の検証を大阪湾を対象として行った。入江グループでは過去大阪湾を対象とした数値シミュレーションを実施していた。大阪湾での数値シミュレーションで用いていたものと同じの入力データを用

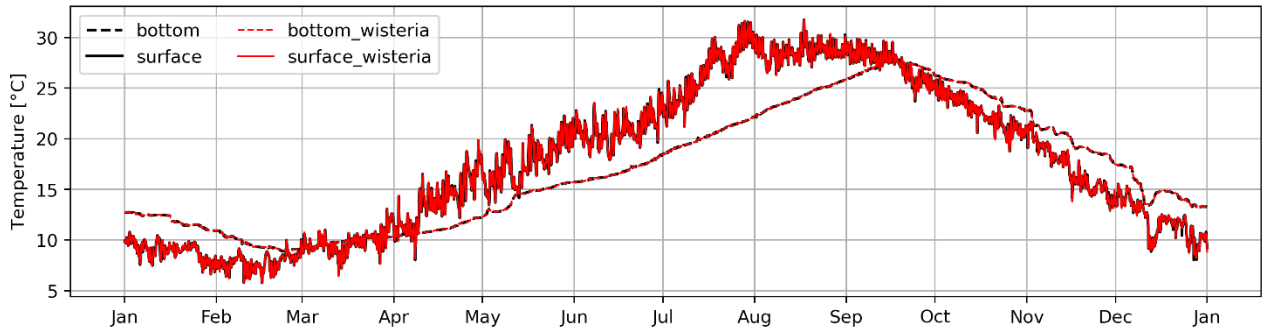


図 2 テストラン結果

大阪湾神戸沖の表層・底層水温の時系列. 赤は wisteria-o, 黒は移行前のマシンの結果.

(mg-O<sub>2</sub>/L)

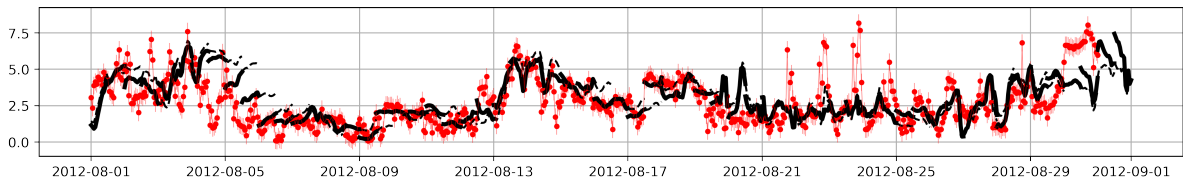


図 2 テストラン結果(データ同化)

4 次元変分データ同化の実行結果. (赤：観測値，黒破線：同化前，黒実線：同化後)

いて Odyssey 上で計算を実行し，計算結果に問題がないことを確認した (図 1).

4 次元変分データ同化の実行テストも同様に大阪湾で実施した. テスト用に内部ループ外部ループ数はそれぞれ 2 回で行ったが，観測値-モデル間のミスフィットが削減できることを確認した (図 2).

次に，東京湾・相模湾領域用の入力データ作成し，東京湾・相模湾を対象とした数値シミュレーションを実施した. しかし，南開境界付近での計算の発散や浅海部の境界付近で計算の異常値が見られ，これらの問題に対処する必要があった.

南開境界での問題では境界条件のデータのチューニングを行った. その結果，計算の発散の問題は解消された. 浅海部の境界付近で計算の異常値の問題は現在検討中である. 必要な場合は計算領域や計算グリッドの格子幅の探索も検討している.

入江グループで利用している計算用サーバと比較して wisteria Odyssey では計算速度が 1.3 倍程度になった. 今後はファイル

IO を最適化することによってさらなる高速化を図る.

#### (ii) 機械学習

RNN ベースのディープラーニングに基づく時系列予測を行った. 東京湾上の川崎人工島での現場観測データ，JMA-MSM による再解析データ，河川流入モデルによる河川データ，UCLA-ROMS (神戸大学内山雄介教授提供) による数値シミュレーションデータを含む計 90 項目の内，任意の項目の時系列データの説明変数とし，川崎人工島での現場観測 24 項目内任意項目を目的変数とするシステムを用いた. 本プロジェクトの前身となる 2021 年度東京大学萌芽共同研究公募課題で利用していた機械学習システムを東京大学の協力の下で改良したものである.

2015 年 1 月から 2017 年 12 月までのデータをトレーニング用，2018 年を検証用，2019 年をテスト用とした. まずは説明変数項目の最適化(次元削減)を行った. ベースモデルは，川崎人工島でのクロロフィル a 濃度(上中下層)，水温(上中下層)，溶存酸素量 DO(上中下

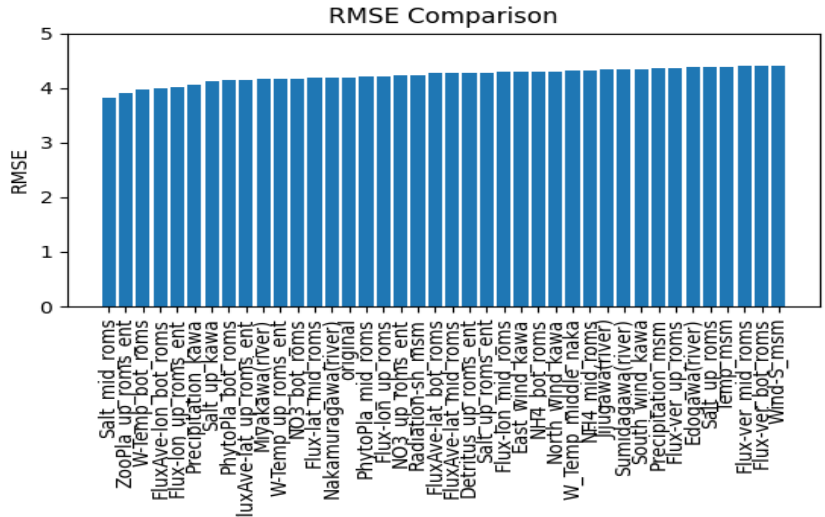


図 3 説明変数の選択と精度(RMSE)の比較

横軸はベースモデルに説明変数として加えた項目を表し、縦軸はそのテストモデルによる予測と検証用データ(2018 年分)の RMSE(クロロフィル a 濃度中層)を示す。original はベースモデルを表す。

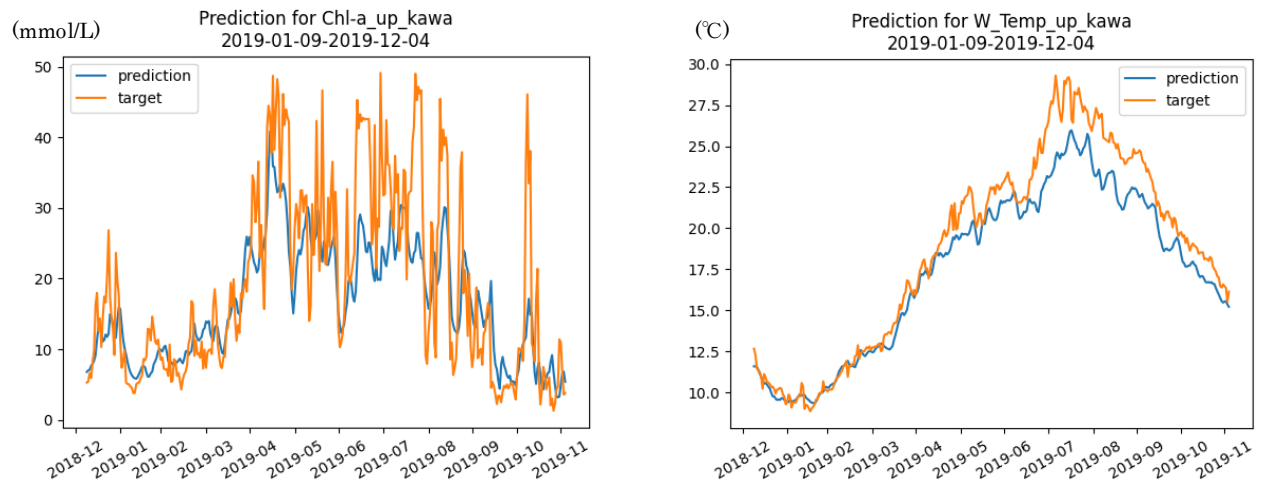


図 4 2019 年 1 月-12 月の川崎人工島のクロロフィル濃度 a 上層(左)と水温上層(右)の観測とディープラーニングによる予測との比較。target は観測，prediction はモデルによる予測を表す。

層), 濁度(上中下層)の 12 項目の観測データを説明変数とし、それら全てを目的変数として時系列を予測するものに設定した。ベースモデルに対して説明変数に上記の 12 項目以外の 78 項目の中から新たに各 1 項目ずつを説明変数として加えたテストモデルを 78 個作成し、学習と予測とテストモデルの性能の比較を行った。性能の指標として中層のクロロフィル a 濃度の誤差(RMSE)を用いた (図 3)。計 14 個のテストモデルでベースモデルより指標の改善が見られた。ROMS による塩分・

動物プランクトン・水温のシミュレーションデータや川崎人工島での降水量・塩分の観測データがディープラーニングによる予測に寄与していることがわかった。なお、Wisteria Aquarius 上での機械学習にあたっては、GPU を利用することで、菊地グループが理化学研究所で利用しているコンピュータでの計算時間より約 100 倍程度高速化した。指標の改善した 14 項目を新たに説明変数としてベースモデルに加えたモデルを作成し、2019 年分の時系列予測を行った。図 4 で川崎

人工島でのクロロフィル a 濃度(上層)と水温(上層)の観測の時系列とモデルによる時系列予測の比較を示す。クロロフィル a 濃度は1-4 月、9-12 月は概ね観測と近い値にとどまるが、変動が大きくなる 5-8 月は実測と符合しなかった。水温に関しては通年で観測と近い値にとどまるが、同じく 5-8 月の夏季には予測と実測との差が大きくなった。

## 6. 進捗状況の自己評価と今後の展望

### (i) 数値シミュレーション

1 年目では、本研究の対象となる相模湾・東京湾は外洋に比べて栄養塩が豊富なため、検証に際して生物場のパラメータチューニングを予定していた。現在の進捗は予定より遅い。その理由は、入力データの作成とその入力データに適した ROMS の内部パラメータの最適化に時間を要したからである。2023 年度では内部パラメータや生物場パラメータのチューニングを行い、4 次元変分法に基づくデータ同化実験を行うことを予定している。これらのシミュレーションデータを機械学習の入力データとして用いることで、水圏エコシステムの将来予測の高精度化を図る。

### (ii) 機械学習

東京湾・相模湾のシミュレーションデータと水圏生態系因子を含む時系列観測データを統合した東京湾・相模湾海洋ビッグデータセットから機械学習用のデータセットへの次元削減と面的・空間的な時系列予測を行う予定であった。現在の進捗は予定より遅い。その理由は、本プロジェクトで用いる機械学習システム的大幅な改良と本プロジェクトへの最適化を行い、これに時間を要したからである。しかしながら、これらの最適化により、システムの柔軟性が向上した。2023 年度は面的予測のための次元削減と、東京湾・相模湾の機械学習による面的・空間的予測を行う。さらに、魚群予測を含む水圏エコシステムの

予測モデル構築へと発展させる。

## 7. 研究業績

### (1) 学術論文 (査読あり)

- Miyamoto, H. and Kikuchi, J.\* “An evaluation of homeostatic plasticity for ecosystems using an analytical data science approach” *Computational and Structural Biotechnology Journal* **21**, 869-878 (2023).
  - Yokoyama, D.\* and Kikuchi, J. “Inferring microbial community assembly in an urban river basin through geo-multi-omics and phylogenetic bin-based null-model analysis of surface water” *Environmental Research* (in press).
- (2) 国際会議プロシーディングス (査読あり)  
 (3) 国際会議発表 (査読なし)  
 (4) 国内会議発表 (査読なし)  
 (5) 公開したライブラリなど  
 (6) その他 (特許, プレスリリース, 著書等)