

気象データを用いた三河湾有害藻類予測 AI の構築

GCOM-C 衛星由来のクロロフィル a 濃度予測モデルの検討

有馬千晴 高橋信雄 神沼英里

近年、日本では有害藻類ブルームが原因の漁業被害が増えている。本研究では、気象データを入力とした有害藻類ブルームの予測モデル構築を試みる。有害藻類ブルームの事例として、三河湾の赤潮に着目する。三河湾の特定地点でのクロロフィル a 濃度を有害藻類ブルーム観測値として、Random Forest 機械学習回帰モデルで過去気象データから予測を試みた。モデルの学習には、三河湾での 2018 年から 2022 年までの観測前数日の気温・降水量・日照時間・風速、および直前の雲無し過去観測日のクロロフィル a 濃度を特徴量とし、雲無し日の衛星観測クロロフィル a 濃度値を予測した。回帰モデルの特徴量分析から、過去観測のクロロフィル a 濃度値が重要であること、気象データの中では気温と日照時間がクロロフィル a 濃度値の予測に寄与することが示唆された。

キーワード：有害藻類ブルーム、クロロフィル a 濃度、機械学習、GCOM-C 衛星、気象データ

1. はじめに

海や湖、貯水池の水質汚染の問題の一つとして有害有毒な藻類の大量発生に起因するものがあり、有害藻類ブルーム(Harmful Algal Blooms: HABs)と呼ばれている 1)。有害藻類ブルームは、水中の酸素レベルを著しく低下させて魚や他の海洋生物を死滅させる場合がある。例えば、2021 年秋に北海道南東部の赤潮発生でウニ等の 80 億円程の漁業被害となった 2)。また 2022 年夏季には熊本県で過去 2 番目の規模の 19 億円の赤潮被害が発生している。赤潮だけでなく、湖やダム、貯水池にも植物プランクトンが大量に増えて、湖面はアオコと呼ばれる青緑色の粉をまいたような状態になり、飲料水の水質悪化や悪臭の源となっている。2021 年の UNESCO 報告書「GLOBAL Harmful Algal Bloom Status Report 2021」によると、1950 年から 2018 年にかけて地球規模での HABs 発生数は増加している 3)。これは、地球温暖化の影響が示唆されており、今後、発生頻度が増加して水産物の被害が深刻になると言われてい

る。この HABs の発生イベントのモニタリングや発生予測の為に、様々な地球観測衛星を使った研究が報告されている。衛星の場合、センサの種類が問題になるが、HABs 研究報告 420 件のうち、クロロフィル a (Chl-a) 濃度が衛星データのプロキシとして 8 割の論文で採用されていたとの報告がある 4)。クロロフィル a 濃度の HABs スペクトル特性は、植物プランクトンのバイオマス増加との関連から、550nm と 700nm に反射率ピークがあることが知られている 5)。日本を観測対象とした場合は、クロロフィル a 濃度計測衛星として、JAXA の GCOM-C 衛星(しきさい衛星) 6)のデータが活用可能である。

本研究では、GCOM-C 衛星による海域のクロロフィル a 濃度を用いて HABs の AI 分析を行う。HABs 発生地域として、愛知県名古屋市沖の伊勢湾と隣接する三河湾をターゲットとした。三河湾の衛星クロロフィル a 濃度データと気象データを用いて、HABs 発生に関連するクロロフィル a 濃度の予測 AI モデルを構築した。クロロフィル a 濃

度予測 AI モデルの重要変数を求めることで、HABs 発生に寄与する要因を探索する。

2. 研究の方法

2.1 AI 学習データの収集

(1) JASMES しきさい衛星クロロフィル a 濃度

JAXA 地球観測研究センターは、日本国内内湾を対象とした GCOM-C 衛星観測値のクロロフィル a 濃度データを JAXA Satellite Monitoring for Environmental Studies (JASMES) Website 7) の「しきさい内湾モニタ」から提供している。GCOM-C 衛星の搭載光学センサ「多波長光学放射計」の観測値データは、等緯度経度に地図投影した値が公開されている。内湾モニタは、地点毎に 1 日平均でクロロフィル a 濃度・懸濁物質濃度・海面水温の時系列データを提供している。このうちクロロフィル a 濃度(Chl-a)のみをモデル構築に利用する。モニタリング地点は、No.4 の伊勢湾(Area ID=ISE)の 16 観測サイトから三河湾の豊橋寄り奥側の ty01(経度 137.25,緯度 34.75)サイトを選択した。しきさい衛星の観測データは雲が無い日の分のみである。2018 年 1 月 13 日から 2022 年 10 月 31 日までの 371 日分のクロロフィル a 濃度データを利用する事とした。ty01 地点の地図は、ESRI Japan 社提供の背景地図へのマッピング機能を持つ MATLAB R2022b 8) で作成した。

(2) 気象庁過去データベースから気象観測値を収集

国土交通省気象庁ホームページから、過去のアメダス観測地点での気象データ 9) をダウンロードした。アメダス観測地点は、三河湾を囲む渥美半島の愛知県田原市にあるアメダス伊良湖気象観測所を選択した。対象期間は、2018 年 1 月 1 日から 2022 年 10 月 31 日として、伊良湖観測地点の日平均気温(°C)、降水量の日合計(mm)、日照時間(時間)、日平均風速(m/s)を 1,765 日分ダウンロードした。クロロフィル a 濃度の観測値は、371 日分しかないので、前日の平均気温、降水量、日照時間、風速、また前 3 日・同 5 日・同 7 日を平均値を計算した日照時間と風速を AI 予測モデ

ル構築用の入力変数とした。

2.2 クロロフィル a 濃度の予測 AI モデルの構築

三河湾 ty01 地点のクロロフィル a 濃度の機械学習予測モデルの構築は、Python の自動機械学習 (AutoML) ライブラリの“PyCaret ver. 2.3.10” 10) を用いた。回帰木のアルゴリズムを比較して、最も精度が高いモデルを選択した。学習データは訓練とテストの割合を 8:2 に分割した。プログラミング実行環境は、Web ブラウザ経由でクラウド上の計算機を利用できる Google Colaboratory 11) を用いた。また入力変数毎の予測モデルへの重要度は、PyCaret の Feature Importance 機能を用いて計算した。

3. 研究の結果

3.1 伊勢湾流域における HABs 発生状況と Chl-a モニタリング地点

国土交通省中部地方整備局 名古屋港湾空港技術調査事務所が管理している伊勢湾環境データベース 12) に、1993 年から 2017 年までの赤潮の発生件数が、伊勢湾・知多湾・渥美湾の湾別に分類されている。赤潮は海域における有害ブルームの 1 種である。伊勢湾環境データベースでは、発生源別に珪藻類と渦鞭毛類に分けて赤潮発生件数がまとめられている。図-1 に、2 つの発生源別の件数を合算したグラフを示す。

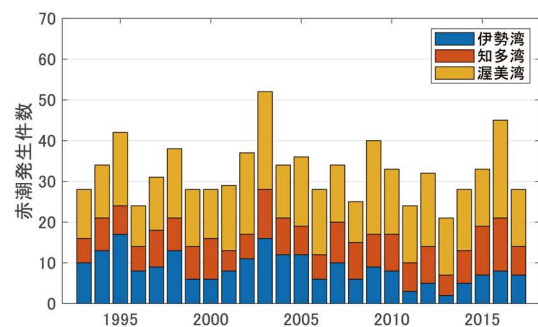


図-1. 伊勢湾流域(伊勢湾、知多湾、渥美湾)における赤潮発生数

伊勢湾流域は、伊勢湾と三河湾に分割できる。三河湾は更に知多湾と渥美湾に分割できる。図-1 では、三河湾の赤潮

発生数は、知多湾（赤色）と渥美湾（黄色）を合わせた部分に相当する。3つの湾をまとめた伊勢湾流域全体では、2017年まで平均して年32回の赤潮が発生している。湾別での赤潮の発生割合は、伊勢湾(27%)と知多湾(25%)に対して、渥美湾(48%)が最も多く、全体の約半分を占めている。この為、AIモデル予測対象のクロロフィルa濃度のモニタリング地点は、渥美湾中の候補から豊橋寄りの ty01 地点を選択した。また三河湾の気象データを抽出するにあたり、ty01 地点に近い気象庁アメダス観測地点として、渥美半島の伊良湖気象観測所のデータを用いる事とした。図-2に、ty01 地点と伊良湖観測所の地点を示す。



図-2. しきさい衛星でのクロロフィルa濃度の観測地点 ty01と、気象観測地点の伊良湖観測所

3.2 ty01 地点の Chl-a データと、伊良湖地点の気象データの特徴

AIモデル構築用の学習データには、2018年1月1日から2022年10月31日までの ty01 地点の Chl-a データと伊良湖観測所の気象データを使用する。通常の時系列データならば、過去の値から未来値を予測する AI モデルを作成可能である。しかし、今回のクロロフィルa濃度データは、衛星の光学センサ由来の為に欠損値が多いので、時系列データとして扱う事が出来ない。実際に、衛星と地上 (ty01 地点)の間に雲がかかって Chl-a が欠損値となった割合は、全日数の8割を占めた。一方で、伊良湖観測所の気象データは、地上センサで観測値の為に、全1,765日に亘り欠損値がない。この様に、Chl-a データは全体の2割

しかデータが無く時系列データとして AI 予測モデルを構築する事ができない。この為、Chl-a を予測するために主として前日の気象データを用いる事とした。

3.3 気象データからの Chl-a 予測機械学習モデルの構築

海域の有害ブルームに関連するクロロフィル a 濃度を気象データから推論する AI モデルを構築する。AIモデルの学習用入力データは、前回観測の Chl-a 値・前日の気温・降水量・日照時間・風速と、前三日、前五日、前七日で平均値を求めた日照時間と風速を追加した。合計で11種類の入力変数となる。予測対象の Chl-a 値は、対数スケールに変換した。

次に、自動機械学習(AutoML)のローコードライブラリである PyCaret を用いて伊良湖地点の気象データを入力変数とし、ty01 地点のクロロフィル a 濃度を出力変数とする三河湾 Chl-a 予測モデルを構築した。前処理で入力変数の標準化を行った。予測モデルの作成には、AutoML ツールである PyCaret を用いて、Chl-a を予測する最適な回帰アルゴリズムを探索した。5分割交差検証を行い、平均2乗誤差 (Mean Square Error, MSE) の評価尺度を用いて比較したところ、Random Forest 回帰 13)の機械学習アルゴリズムが最も低い MSE を示した。図-3にクロロフィル a 濃度の予測値と観測値を示す。

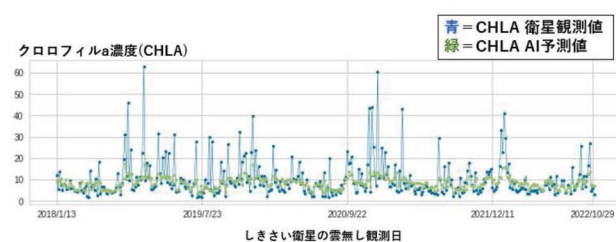


図-3. クロロフィル a 濃度の観測値(青)と予測値(緑)。横軸は雲無し観測日の為、時系列データではない事に注意。

図-3 の予測値と観測値の決定係数 R^2 は、訓練データで0.56、テストデータで-0.027と予測モデルとして精度が低かった。本論文に似た既存研究として、入力変数に気象データに河川流量と水質データを加えた諫早湾の赤潮予測モデルの報告がある 14)。諫早湾の予測モデルも、訓練デ

一タは精度が出るものの、テストデータでは低精度だった。一方で、時系列で過去の Chl-a データを入力情報として機械学習予測モデルを構築する研究では、対象海域サイト毎に精度が大きく変動する結果が出ている 15)。

3.4 Chl-a 予測モデルにおける重要変数の分析

次に、ty01 地点のクロロフィル a 濃度予測に、各入力変数がどの程度影響を与えているかの重要度 (Feature Importance) を計算してランキングを行った (図-4)。横軸の値が大きい程、入力変数として Chl-a 予測モデルへの寄与度が高い。結果として、クロロフィル a 濃度の前回計測値が、気象変数よりも重要である事が判る。一方で、気象変数の中では、降水量が見られないことから Chl-a モデルへの影響が小さい事が判る。また図-4 の重要度のランキング結果から、前日データに限ると気温、日照時間、風速の順番に Chl-a 予測に影響があることが明らかとなった。

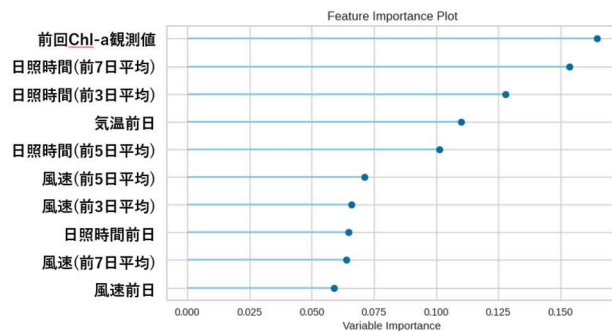


図-4. ty01 地点の Chl-a 予測モデルの特徴量重要度ランキング

一方で、日照時間も風速も、前7日平均や前5日平均の影響が、前日1日分の影響よりも大きかった。今回は気象データを入力してChl-aを推論する回帰モデルを構築したが、過去7日間の日照時間が前回Chl-a観測値に近い重要度を示し、HABs予測に有効な気象属性であることが示唆された。

4. まとめ

本研究では、有害藻類ブルームの AI 予測モデルを構築するために、過去気象データからの三河湾 HABs 予測モ

デルを構築した。HABs の定量化には、GCOM-C 衛星のクロロフィル a 濃度観測値を用いた。予測モデルに影響する特徴量重要度分析から、クロロフィル a 濃度予測には、気温・日照時間・風速の順で影響があり、降水量は影響があまりない事が明らかとなった。一方で、海水での物理化学分析から硝酸塩が HABs に影響する等の報告 16)や、HABs 関連のメタバーコーディング研究 17)も進められている。今後は、衛星観測では網羅が出来ない観測サイトでの計測情報を HABs 予測モデルに取り込み、時系列データを作成していく予定である。

謝辞

本論文で使用した三河湾のしきさい衛星クロロフィル a 濃度データは、宇宙航空研究開発機構より提供を受けました。有害藻類の衛星観測研究については、豊橋技術科学大学 大学院工学研究科の広瀬侑先生にアドバイスを頂きました。厚く、御礼を申し上げます。

参考文献

- 1) 国立科学博物館：有害有毒藻類ブルーム、
<https://www.kahaku.go.jp/research/db/botany/habs>
- 2) Iwataki, *et al.*: Morphological and phylogeny of *Karenia selliformis* in an intensive cold-water algal bloom in eastern Hokkaido, Japan”, *Harmful Algae*, Vol. 114 (2022). pp.102204.
- 3) UNESCO, GLOBAL Harmful Algal Bloom Status Report, IOC/CL-2849 (2021).
- 4) Khan R.M., *et al.*: A Meta-Analysis on Harmful Algal Bloom (HAB) Detection and Monitoring: A Remote Sensing Perspective, *Remote Sensing*, Vol. 13, No. 21 (2021). pp.4347.
- 5) Jensen J.R.: *Remote Sensing of the Environment: An Earth Resource Perspective*. 2nd ed. Pearson Prentice Hall, Upper Saddle River, NJ, USA (2007).

- 6) Hori M *et al.*: GCOM-C Data Validation Plan for Land, Atmosphere, Ocean and Cryosphere, Trans. of JSASS, Vol.16, No.3 (2018). pp.218-223.
- 7) JAXA 地球観測研究センター: JASMES しきさい内湾モニタ, <https://www.eorc.jaxa.jp/JASMES>
- 8) The MathWorks Inc. : MATLAB ver. 9.13.0.2049777, R2022b (2022).
- 9) 国土交通省気象庁: 過去の気象データ・ダウンロード, <https://www.data.jma.go.jp/gmd/risk/obsdl/index.php>
- 10) Ali, M.: PyCaret: An open source, low-code machine learning library in Python, <https://www.pycaret.org>. (2020).
- 11) Bisong, E.: Google Colaboratory, in Building Machine Learning and Deep Learning Models on Google Cloud Platform, Apress Access Books (2019). pp.59-64.
- 12) 国土交通省中部地方整備局 名古屋港湾空港技術調査事務所: 伊勢湾環境データベース, <http://www.isewan-db.go.jp/>.
- 13) Liaw, A. and Wiener, M.: Classification and Regression by randomForest, R News 2 (2002). pp.18-22.
- 14) 金,藤井,濱田: 機械学習を用いた沿岸海域における赤潮予測, 土木学会論文集 B1, Vol.75, No.2 (2019). pp.781-786.
- 15) Baek,S.*et al.*: Deep learning-based algorithms for long-term prediction of chlorophyll-a in catchment streams (2022).
- 16) Mahmudi, M. *et al.*: A count model approach on the occurrences of harmful algal blooms (HABs) in Ambon Bay, The Egyptian Journal of Aquatic Research, Vol.46 (2020). pp.347-353.
- 17) Sildever, S. *et al.* : Monitoring harmful microalgal species and their appearance in Tokyo Bay, Japan, using metabarcoding, Metabarcoding and Metagenomics, Vol.6 (2022). e79471.