

機械学習を用いた沿岸海域における赤潮予測

金 相曄¹・藤井 直紀²・濱田 孝治³

¹正会員 島根大学 エスチュアリー研究センター (〒690-8504 島根県松江市西川津町 1060)

E-mail: kim@soc.shimane-u.ac.jp

²非会員 佐賀大学 農学部 (〒852-852 佐賀県佐賀市本庄町 1)

E-mail: medusae@cc.saga-u.ac.jp

³正会員 株式会社 シャトー海洋調査 (〒111-0053 東京都台東区浅草橋 5-23-9)

E-mail: hamada-chateau@to.email.ne.jp

沿岸海域で発生する赤潮は、水産業に甚大な悪影響を与えている。その赤潮の発生と消失には、海況や気象条件、河川からの流入量など様々なプロセスがかかわっている。諫早湾を含む有明海では、赤潮が頻繁に発生しているため、多年にかけて水質の連続観測が行われているものの、赤潮の生成消失のメカニズムの把握や、その予測は容易ではない。本研究では、有明海と諫早湾で得られたビックデータを用いて、機械学習のSVR、RF、GBRTから赤潮予測モデルを構築するとともに性能検証を行った。その結果、データを学習と予測データに分割する際にランダムで行った場合は、性能に問題がなかった。しかし、時系列でデータを分割した場合は、予測データのChl-a変動特性が学習データに含まれていなかったため、再現性が劣ることが確認された。

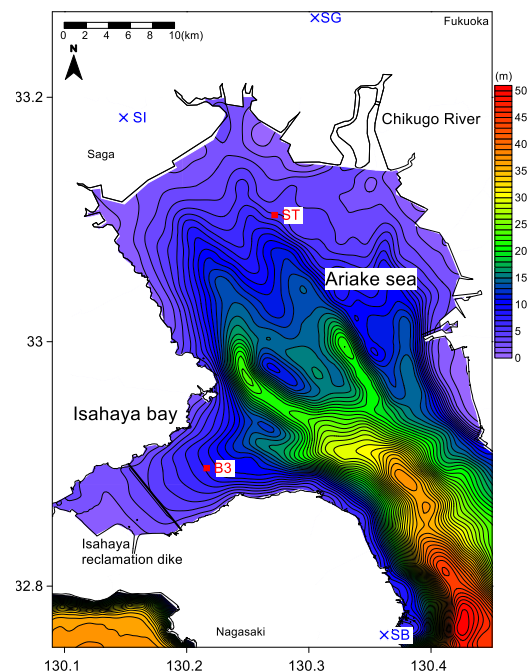
Key Words: Machine Learning, Red Tide, Ariake sea, Isahaya Bay, Random Forest

1. はじめに

近年、陸域からの栄養塩負荷の増大や地球温暖化などに伴い、世界各地で有害有毒赤潮が問題となっている。有明海においては特に、*Chattonella* 属による有害赤潮と、珪藻赤潮の発生が注目されている¹⁾。*Chattonella*は魚の鰓上で粘性物質の放出と細胞崩壊を起こして魚を窒息させ、養殖魚のへい死をもたらす²³⁾。有明海ではブリ・ハマチの養殖が行われていないため*Chattonella*赤潮による漁業被害は表面化していないが、二枚貝(アサリ、タイラギ、マテガイ、コケガラス等)、魚類(クロダイ、コノシロ、スズキ等)、甲殻類(シャコ、石ガニなど)のへい死にかかわっていることが疑われている⁴⁾。珪藻は栄養塩をめぐってノリと競合関係にあり、有明海の冬の珪藻赤潮はしばしば養殖ノリの色落ち被害をもたらす⁵⁾。

赤潮は、その死骸が海底に沈降し分解する際に酸素を消費することにより、貧酸素水塊をもたらす原因ともなっている。有明海では二枚貝等ベントスが生態系に占める役割が大きく、底生生物に打撃となる貧酸素水塊は大きな問題である。

赤潮の発生と消失には様々なプロセスがかかわっている。たとえば、河川からの出水は海域に栄養塩をもたらす、赤潮を引き起こす。一方、出水は赤潮の制限要因と



図—1 データの位置 (■: 水質, ×: 気象)

もなり得る。表層塩分が10を下回るほどの大きな出水があった場合、*Chattonella* 属は赤潮を形成できないことが報告されている⁶⁾。栄養塩は河川からだけでなく、外

洋や海底からも供給される。有明海奥部では秋から冬にかけて河川流入に匹敵する栄養塩が海底からもたらされている⁷⁾。ただし、その詳細は明らかでない。

移流・拡散プロセスもまた重要である。有明海では、出水に伴い諫早湾に発生した *Chattonella* 赤潮が、橘湾まで輸送された例が報告されている⁸⁾。しかし、有明海奥部の海水の挙動は地球自転、潮汐、風の影響を受け複雑に変動しており⁹⁻¹²⁾、赤潮の発生・消失との関係を理解することは容易ではない。

こうした背景から、有明海における赤潮の発生・消失のメカニズムを明らかにし、また、その予測を行うことには大きな意義があるといえる。本研究では、有明海の支湾である諫早湾を対象海域とし、沿岸域で連続的に計測されている水質・気象データなどのデータを用いて、機械学習により諫早湾における赤潮（クロロフィル a の量）予測モデルを構築した。また、構築したモデルに基づき各指標量の寄与について考察した。

2. 解析に用いたデータの詳細

本研究では、水質データとして、九州農政局が計測を行っている観測櫓（B3, B4, B5, S6, S1, S6）と佐賀大学有明海観測櫓（ST）で毎正時、海面下 0.5m 層の水温、塩分、溶存酸素濃度（以降、DO と略記）およびクロロフィル a（以降、Chl-a と略記）を使用した。観測櫓の位置を図-1 に示す。

河川流量データとしては有明海に注ぐ 8 つの 1 級河川の中で、既往の研究⁹⁻¹²⁾で諫早湾への影響が明らかになった筑後川の流量のみを考慮することとし、国土交通省の水文水質データベース¹³⁾の瀬ノ下観測所（CR）の時刻流量データを使用した。

気象データとしては、諫早湾に最も近い島原アメダス（SB）と筑後川河口部に最も近い白石アメダス（SI）における気温、降水量、風向風速および日照時間、長崎地方気象台（NS）と佐賀地方気象台（SG）における湿度、海面気圧のデータを用いた¹⁵⁾。風向は One-hot-coding で 16 方位へ変換し、カテゴリーデータとして扱った。

上記データのうち、諫早湾の中央に位置する観測櫓（B3）の Chl-a 値を目的変数、佐賀大学観測櫓（ST）の水質データおよび河川流量データ、気象データを説明変数としてモデルを構築した。対象期間は 2007 年から 2012 年である。環境条件が赤潮に反映されるまでのタイムラグを考慮するため、すべての独立変数について、7 日前まで 24 時間ごとにシフトしたデータも独立変数として使用した。変数はタイムラグを含めて 325 個、エラーを除いた各変数におけるデータ数は 20,208 個であった。

なお、観測櫓（B4, B5, B6, S1, S6）については、観測櫓（B3）と近すぎ、また、データに欠測が多いことが

分かったため解析に使用しなかった。

3. 解析方法の概要

機械学習（Machine Learning）は、ビックデータからパターンおよび相関性を統計的に識別するデータマイニングに基づいて、学習アルゴリズムを用いて知識を抽出し、それから予測を行う手法である。学習方法によって、教師あり学習、教師なし学習、強化学習に分類できる。本研究では、特徴量と出力を持つデータから学習して、経験のないデータや未来のデータに対する予測をする教師あり学習のうち、Support Vector Regression（SVR）、Random Forest（RF）、Gradient Boosting Regression Tree（GBRT）を使用した。SVR は、与えられたデータのグループに基づいて、新たなデータがどのカテゴリに入るのかを判断する非線形分類アルゴリズムの SVM（Support Vector Machine）の回帰版である。RF は、決定木アルゴリズムの過学習（Overfitting）問題を解決するため提案された決定木アンサンブルモデルであり、植物プランクトンの時系列予測への適用例がある¹⁶⁾。GBRT は、RF と同じく決定木を用いたモデルであるが、誤差を最小化するように逐次的に決定木を追加していく点が特徴であり、浮魚の漁獲量推定への適用例がある¹⁷⁾。

なお、本研究ではプログラミング言語 Python 上のオープンソース機械学習ライブラリ scikit-learn¹⁸⁾を使用して各モデルを構築した。

表-1 ハイパーパラメータチューニング結果

モデル名	主なハイパーパラメータ
SVR	C=5 gamma=0.0001
RF	n_estimators=190 min_samples_leaf=4 max_depth=4
GBRT	n_estimators=290 min_samples_leaf=3 max_depth=4

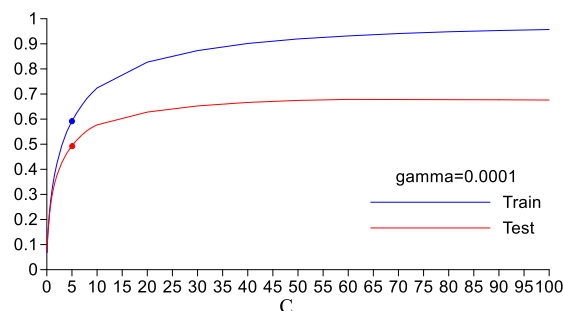


図-2 SVR におけるパラメータ C 調整と性能変化

4. 結果と考察

(1) モデルのチューニングと性能の評価

モデル性能の向上のため、グリッドサーチ法によるハイパーパラメータチューニングを行った。2007 年から 2012 年の中で、80%を既知のトレーニングデータ (train)、残りの 20%をテストデータ (test) としてランダムにデータを分割し、学習と予測の性能差が 0.1 以下の内、予測性能が最も大きい時のハイパーパラメータを選択した。さらに、モデル性能の確認のため、全データと得られたハイパーパラメータを用いて K-Folds Cross Validation Method による交差検定を行った。交差検定の回数(K)は 5 回とした。

表-1 に、チューニングに使用したハイパーパラメータとグリッドサーチ法により選択された値を示す。また、図-2 に、パラメータ設定に伴う変動が最も大きかった SVR について、 $\gamma=0.0001$ (最適値)としたときの、C 調整による train, test に対する R^2 の変化を示す。本研究では C=5 が選択され、このとき Train=0.59, Test=0.49 であった。C が 5 以上になると、train と test の R^2 に開きが生じ、過学習傾向となっていた。

図-3 に、選択されたハイパーパラメータを用いた時の train, test の R^2 と、交差検証の結果 (R^2 の平均値, Average) を示す。各学習モデル中で、GBRT が train, test 共に優秀な性能を示した。交差検証の結果は GBRT=0.68, SVR=0.54, RF=0.44, であった。

(2) 諫早湾の Chl-a 変動に対する特徴量選択

諫早湾の Chl-a の変動に対する重要な影響因子を把握するため、Boruta アルゴリズムを用いて特徴量選択 (Feature Selection) を行った。Boruta は、RandomForest に基いて、有意意味な変数を探して、ランクを付けるアルゴリズムのライブラリーである¹⁹⁾。

図-4 に、Boruta で選ばれた上位 37 個のランクを示す。

1 位~5 位は、①佐賀大学観測槽における Chl-a (ST_Ch1-a)、②筑後川の 3 日前の出水量 (CR_3days)、③佐賀大学観測槽における 3 日前の表層塩分 (ST_SSS)、④筑後川の 2 日前の出水量 CR_2days、⑤賀大観測槽における 2 日前の表層塩分 ST_SSS_2days であった。

①が高い理由は、佐賀大学槽(ST)と諫早湾観測槽(B3)

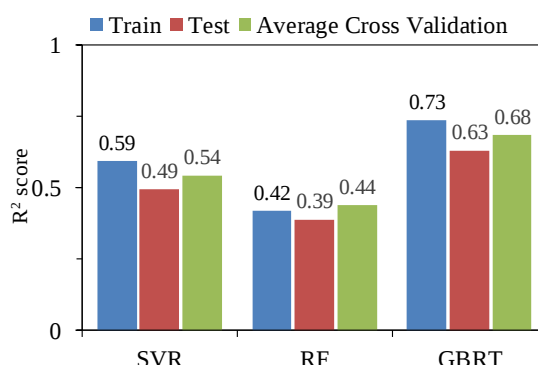


図-3 各モデルの学習と予測性能と交差検証点数 (観測槽 B3 の Chl-a 値)

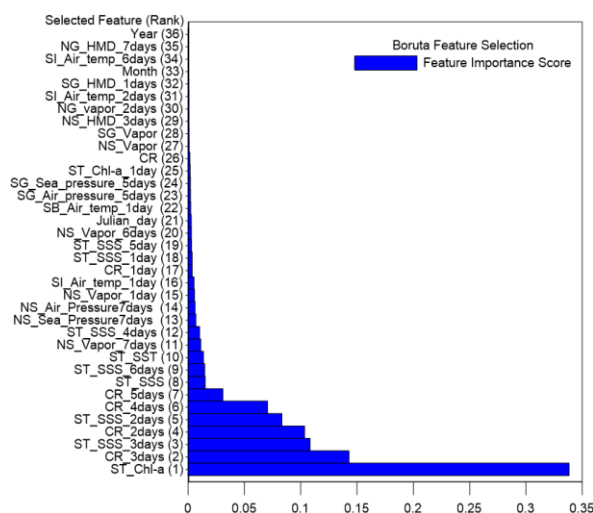


図-4 観測槽 B3 の Chl-a 値に対する特徴量選択結果

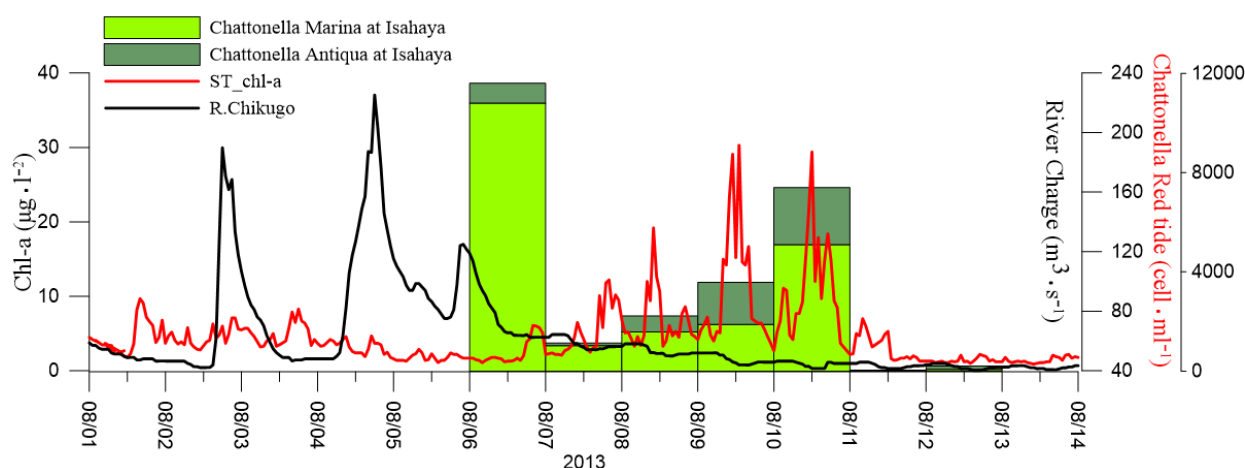


図-5 諫早湾で発生した Chattonella 属赤潮の細胞数と ST の表層 Chl-a および筑後川の出水量の比較

の直線距離が 25km であり、似通った気象条件の下で植物プランクトンの増殖が同期するためと推察される。

②、④と③、⑤は、筑後川(CR)の出水の影響が佐賀大学櫓(ST)を経て諫早湾観測櫓(B3)に及ぶまでのタイムラグを表していると推察される。齋田ら⁹⁾は、出水時の筑後川河川水の運動は、コリオリ力に支配されており、出水後の一潮汐間で佐賀県太良沖まで到達することを示している。また、Kim ら¹⁰⁾は、諫早湾観測櫓データの解析により、筑後川瀬ノ下からの淡水の影響が 1 日から 3 日後に諫早湾に及ぶことを示している。上記の結果はこれらと整合する。

ここまでは Chl-a を赤潮の指標として再現性を見てきたが、ここでは個別の赤潮種に注目することにより現在のモデルの改善すべき点を考察する。図-5に、2013 年の夏季に諫早湾で発生した *Chattonella* 属赤潮の最高細胞数（長崎県総合水産試験場で公表された赤潮速報から作成）、佐賀大学観測櫓の表層の Chl-a、および筑後川瀬の下の流量を示す。諫早湾では 8 月 6 日に *Chattonella* 属赤潮が発生し、8 月 7 日には一旦減少したが 8 月 8 日に再び増加する傾向が見られた。特に、8 月 7 日の 3 日前に筑後川から多量の出水、8 月 7 日～10 日までは有明海奥部（佐賀大学観測櫓）における表層の Chl-a の増加が確認された。この期間の環境因子は、本研究で構築した赤潮予測モデルの特徴量選択結果のランク 1 と 2（図-4）に相当するものであり、2013 年の *Chattonella* 赤潮の挙動に対する整合性の取れるものと判断される。しかし、8 月 6 日に多量に発生した赤潮に対しては、2 日前（図-4 のランク 4）のと 4 日前（図-4 のランク 6）の筑後川からの出水は確認されるものの、ランク 1 の有明海奥部における表層の Chl-a の増加は確認されていない。これは、モデルの構築際に考慮していない環境因子が関係している可能性がある。特に、*Chattonella* 属の赤潮は、密度成層の形成と貧酸素化に伴う海底からの栄養塩溶出が引き金になっていることが指摘されているため²⁰⁾、密度成層と海底からの栄養塩溶出などの環境因子をさらに考慮することによってモデルが改善される可能性がある。このように、機械学習を用いた赤潮予測モデルの性能を向上させるためには、海況条件によって変わる赤潮の種類に対する適切な環境因子の選択が重要である。

(3) Chl-a の時系列の再現性の検証

有明海・諫早湾で発生する赤潮は、魚介類のへい死、貧酸素水塊の発生、養殖ノリの色落ちなどと深くかかわっている。赤潮が時間を追って変化していく様子が表現されれば、機械学習は被害予防のためのツールとして役に立つ。

(1)で行った交差検証では、ランダムに分割したデータを使用したため、時系列の再現性を評価することがで

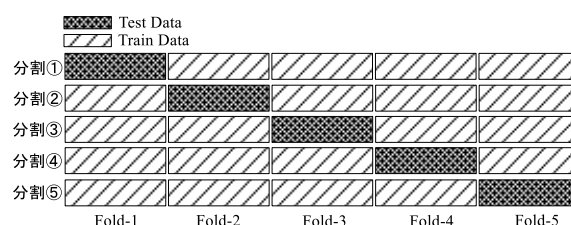


図-6 交差検証における時系列データ分割の概略

表-2 時系列データの交差検証結果

学習機	分割①	分割②	分割③	分割④	分割⑤
SVR	0.16	0.09	0.00	0.119	0.00
RF	0.125	0.107	0.00	0.224	0.104
GBRT	0.336	0.307	0.00	0.245	0.00

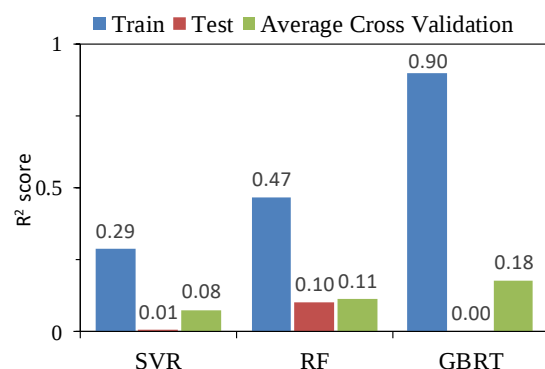


図-7 各モデルの学習と時系列予測性能と交差検証点数

きない。そこで、図-6に示すように全期間を 5 等分し、ある 1 期間を test、残り全てを train として K=5 の交差検証を行った。SVR, RF, GBRT のそれぞれについての各回の R^2 を表-2 に示す。

図-7 に、分割 5 の train, test に対する R^2 と、5 回分の test の平均 (Average) を示す。どのモデルについても、train の R^2 の値は test の R^2 よりかなり小さく、過学習が起こっていることが示された。特に、SVR と GBRT では、 R^2 はほとんど 0 であった。そこで、時系列がどのように再現されているのか、また、 R^2 が低い原因は何か、最も R^2 が高かった RF の結果に基づき考察した。

図-8 に、分割⑤の場合における観測櫓 B3 の Chl-a の実測値と予測値の比較を示す。train, test とともに Chl-a の変動を定性的には再現しているが、短周期かつ大振幅の変動はあまり再現されなかった。特に、夏季における Chl-a の増大を再現できているものの、その大きさは完全に再現されていない。一方、2008 年 12 月から 2009 年 3 月までや 2011 年 10 月から 12 月までのように、冬季における Chl-a の変動幅が小さい期間中はとくに定量的な再現性が劣っているのが確認される。

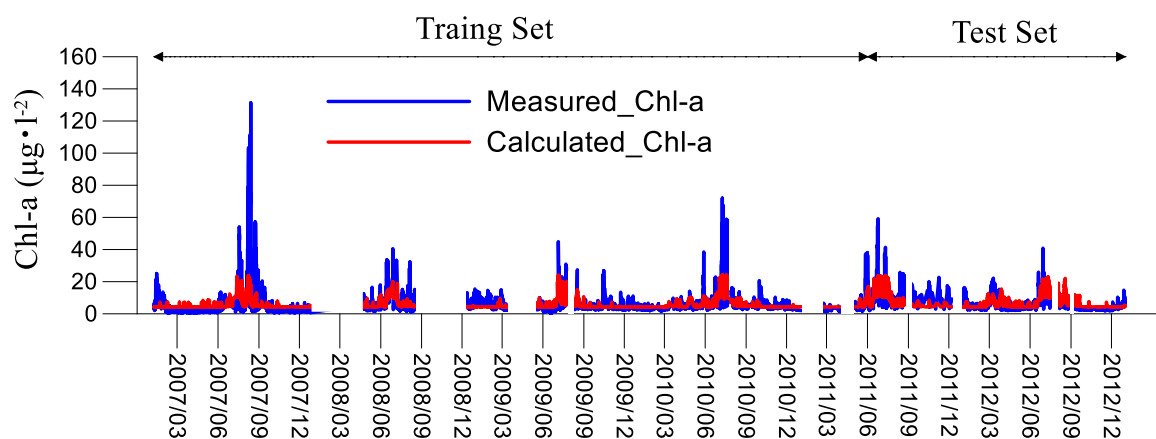
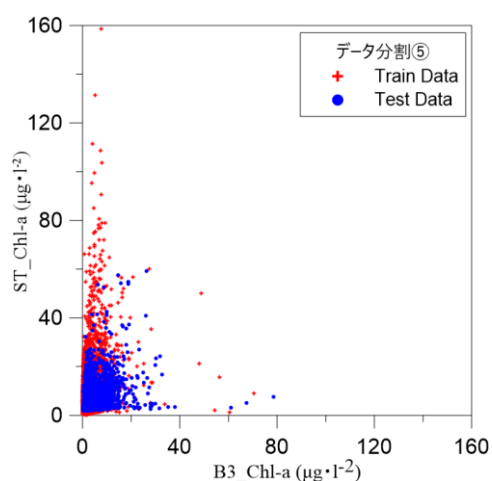
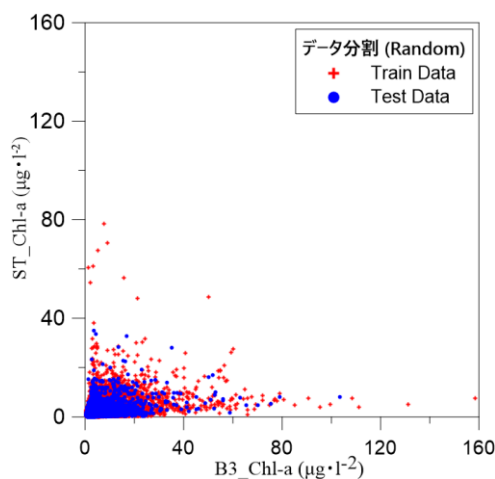


図-8 観測槽 B3における Chl-a 値と予測結果 (RandomForest, 分割⑤)



(a) 分割⑤



(b) ランダム分割

図-9 ST_ChI-a と観測槽 B3 の ChI-a のトレーニングとテストデータの分布特性

図-9に、分割⑤とランダム分割の場合の、諫早湾観測槽(B3)の ChI-a 実測値と、*Boruta* で最も高い寄与があるとされた有明海観測槽(ST)の ChI-a の散布図を示す。分割⑤の場合は、test 期間に属する点が random の場合に比べて偏っていることがわかる。これは、ChI-a の変動のメ

カニズムが年によって異なっており、時系列に基づく分割を行った場合は test 期間の変動特性を学習器が把握できなかったためと推測される。

時系列に基づきデータを分割して機械学習される場合は、データの分布に十分に注意し、特に、様々な赤潮発生パターンを網羅した、十分大きいトレーニングデータセットを用いて解析を行う必要がある。また、ChI-a 値のように時間・空間的にばらつきの多いデータに対しては、カテゴリ化を行うことも有効な対策になると推測される。

5. 結論

本研究では、機械学習のアルゴリズムの中で SVR, RF, GBRT を用いて、沿岸海域における赤潮 (ChI-a) 予測モデルを構築するとともに性能評価を行った。その結果、機械学習による赤潮予測に有効であることが分かった。特に、諫早湾における ChI-a 変動に対する特徴量選択を行った結果、既往の研究と整合する結果が得られ、さらに、学習対象期間外の年度で発生した *Chattonella* 赤潮の挙動にも整合性が取れることが確認された。しかし、学習と予測の期間を時系列で分割した場合は、すべてのモデルから過学習現象が確認された。その原因は、時系列に基づく分割を行った場合はテスト期間の変動特性を学習器が把握できなかったためと推測された。今後、現在構築中である RNN (Recurrent Neural Network) など、時系列の再現に適した、より高性能なアルゴリズムを用いて赤潮予測モデルの性能向上を試みる予定である。

謝辞： 研究を行うにあたり、農林水産省九州農政局から諫早湾の水質、風速・風向および諫早湾調整池からの排水量データを提供頂いた。また、本研究で用いられた佐賀大学観測槽データは、ハブ型ネットワークによる有明海地域共同観測プロジェクト (COMPAS) の一環として取得されたものである。ここに記して深甚なる謝意を表します。

参考文献

- 1) 松原賢：有明海佐賀県海域における赤潮研究－現状と課題－，佐賀県有明海水産振興センター研究報告，28, pp.129-133, 2017.
- 2) 高山晴義, 大内晟, 難波憲二, 村地四郎：魚類におよぼす赤潮の影響について-I. ハマチ *Seriola quinqueradiata* TEMMINCK et SCHLEGEL におよぼすムカシウミミドリムシ *Hemieutreptia antiqua* HADA の影響について，魚病研究，8(2)，1974.
- 3) 松里寿彦, 小林博：海産ミドリムシ赤潮による魚類の斃死機構解明に関する研究，南西海区水産研究所研究報告，7, pp.43-67, 1974.
- 4) 山砥稔文, 坂口昌生, 岩滝光儀, 岡敷充：諫早湾に出現する有害赤潮鞭毛藻 4 種の増殖に及ぼす水温, 塩分の影響，日本水産学会誌，72(2)，pp.160-168, 2006.
- 5) 川口修, 山本民次, 松田治, 橋本俊也：有明海におけるノリと浮遊珪藻の栄養塩競合におよぼす環境諸因子の影響評価，海の研究，14(3)，pp.411-427, 2005.
- 6) Katano T, Yoshino K, Matsubara T, Hayami Y：Wax and wane of *Chattonella* (Raphidophyceae) bloom with special reference to competition between *Skeletonema* (Bacillariophyceae) in the Ariake Sea, Japan, *Journal of Oceanography*, 68(4)，pp.497-507, 2012.
- 7) 渡辺康憲, 川村嘉広, 半田亮司：ノリ養殖と栄養塩ダイナミックス，沿岸海洋研究，42(1)，pp.47-54, 2004.
- 8) Aoki K, Onitsuka G, M. Shimizu, T. Yamatogi, N. Ishida, S. Kitahara, K. Hirano：Chattonella (Paphidophyceae) bloom spatio-temporal variations in Tachibana Bay and the southern area of Ariake Sea, Japan：Interregional displacement patterns with *Skeletonema* (Bacillariophyceae). *Marine Pollution Bulletin*, 99(1-2), 54-60, 2015.
- 9) 齋田範倫, 矢野真一郎, 田井明, 重田真一, 小松利光：筑後川から有明海へ流入する河川水の挙動，水工論文集，第 52 巻, pp.1327-1332, 2008.
- 10) 李洪源, 松永信博：諫早湾における筑後川起源水の風応答特性，土木学会論文集 B1(水工学), Vol. 67, No.2, pp.86-91, 2011.
- 11) 金相曄, 多田彰秀, 田井明, 八百山工祐, 鈴木誠二：2013 年度夏季の諫早湾北部海域における淡水流入の挙動について，土木学会論文集 B1(水工学), Vol.71, No.4, L_667-L_672, 2015.
- 12) Kim S.Y, Tai A, Tada A and Suzuki S：Mechanism of Salinity Stratification Formation Occurring in the Northern Part of Isahaya Bay, Japan, 36th IAHR-APD, 8 pages, 2016.
- 13) 国土交通省 筑後川河川事務所水文・水質速報：<http://www.qsr.mlit.go.jp/chikugo>
- 14) 多田彰秀, 中村康裕, 阿部和也, 田井明, 鈴木誠二, 中村武弘：淡水流入が諫早湾の水質動態に及ぼす影響について，土木学会論文集 B2(海岸工学), Vol.66, No.1, pp.366-370, 2010.
- 15) 気象庁ホームページ：<https://www.jma.go.jp>
- 16) Yajima H, Derot J：Application of the Random Forest model for chlorophyll-a forecasts in fresh and brackish water bodies in Japan, using multivariate long-term databases, *Journal of Hydroinformatics*, 20 (1), 206-220, 2017.
- 17) 濱田孝治, 吉田司, 岡村寛, 原武史, 鈴木輝明：機械学習を利用した内湾における浮魚類群集の空間分布量の推定，土木学会論文集 B2(海岸工学), 印刷中, 2019.
- 18) Scikit-learn Machine Learning in Python. <https://scikit-learn.org/stable/>
- 19) Kursa M., Rudnicki W.,：Feature Selection with the Boruta Package, *Journal of Statistical Software*, Vol. 36, Issue 11, 2010.
- 20) 田中昌宏, 小田切光典：有明海湾奥部の貧酸素水塊形成過程とそのモデル化，土木学会論文集 B2(海岸工学), Vol.66, No.1, pp.1011-1015, 2010.

(Received May 31, 2019)

(Accepted August 7, 2019)

RED TIDE PREDICTION USING THE MACHINE LEARNING IN COASTAL AREA

Sangyeob KIM, Naoki FUJII and Takaharu HAMADA

Red tides occurring in the coastal area has a tremendous negative impact on the fishery industry. The occurrence of red tides is strongly affected by the coastal environment. In the Ariake Sea including Isahaya Bay, red tides occur frequently. Although continuous observation of water quality has been conducted for other years in coastal area, the forecasting model of red tides has not been constructed yet. In this study, using big data measured in Ariake Sea and Isahaya Bay, we tried to make a phytoplankton prediction model from SVR, RF, and GBRT of machine learning. As a result, there was no problem in performance if the data was divided into learning and prediction data at random. However, when the data was divided in time series, the Chl-a fluctuation characteristic in the prediction data was not included in the learning data, so it was confirmed that the reproducibility of the time series model is poor.