

厚生労働科学研究費補助金

健康安全・危機管理対策総合研究事業

水道事業の流域連携の推進に伴う水供給システムにおける  
生物障害対策の強化に関する研究

令和2年度 分担研究報告書

令和3（2021）年 3月



閉鎖性水域における  
藻類異常発生予測モデルの構築

研究代表者	秋葉	道宏
研究分担者	西村	修
研究協力者	今本	博臣
研究協力者	佐野	大輔



厚生労働科学研究費補助金 (健康安全・危機管理対策総合研究事業)  
「水道事業の流域連携の推進に伴う水供給システムにおける生物障害対策の強化に関する研究」  
分担研究報告書

研究課題：閉鎖性水域における藻類異常発生予測モデルの構築

研究代表者 秋葉 道宏 国立保健医療科学院 生活環境研究部 部長  
研究分担者 西村 修 東北大学大学院 工学研究科 教授  
研究協力者 今本 博臣 水資源機構 シニアアドバイザー  
研究協力者 佐野 大輔 東北大学大学院 環境科学研究科 准教授

研究要旨

日本の水道水源の約 50%はダム湖から供給されている一方で、主に夏に発生する藻類により、景観障害や悪臭、水生生物の大量死、カビ臭など、さまざまな悪影響が及ぼされている。もし事前に藻類の異常発生を予測することができれば、取水位置の変更準備や凝集剤の補充等、被害を最小限に抑えるための工学的対応を取ることが可能となる。そこで本研究では、機械学習手法のひとつであるサポートベクターマシン (Support Vector Machine : SVM) を用いて、ダム湖での藍藻類の発生を事前に予測するための統計モデルを構築した。予測モデルの入力データとして水質データ、気象データ、及び水理学的データを使用し、予測日から 1 週間後に藍藻類が発生するか否かを予測することを試みた。その結果、*Microcystis* 属及び *Anabaena* 属に関し高い予測性能を示すモデルの構築に成功した。*Microcystis* 属の最良モデルでは、全窒素濃度、流入量、最高気温、風速の変数が採用されていたことから、これらの変数が *Microcystis* 属の発生予測において重要な予測因子であると考えられた。*Anabaena* 属に関しては、最良モデルにおいて全窒素濃度、最高気温、風速の変数が採用されており、これらの変数が *Anabaena* 属の発生予測において重要な予測因子であると考えられた。

A. 研究目的

日本においてダム湖は水道水源全体の約 5 割を占める重要な取水源となっている。一方で、夏季を中心にダム湖の表層付近で藍藻類が異常発生することで、水面に藻類群体が集積するアオコ等の水の華が見られるようになる。アオコの発生によってダム湖における景観障害や異臭、水生生物や魚類の大量死など様々な悪影響を及ぼすだけでなく、上水道システムにおいてもカビ臭等の異臭味障害や処理設備の故障による処理効率の低下などをもたらす。加えて、*Microcystis* 属が産生するミクロキスチン等の毒性物質による健康被害も懸念される。もし藻類発生時期を事前に予測することが可能であれば、取水位置の変更や凝集剤使用量の増加に備えた準備等の工学的な対策を前もって施すことが可能となる。

以上の背景のもと、本研究では、機械学習手法の一つであるサポートベクターマシン (Support Vector Machine : SVM) を用いて、ダム湖における藻類異常発生予測モデルの構築を試みた。

B. 研究方法

水資源機構が管理する 4 つのダムを対象とし、それらのダムにおける藻類異常発生予測を行っ

た。使用した藻類細胞数及び放流量等のデータは水資源機構による定期水質調査の結果を用いた。気象データについては対象ダム近傍の気象観測所のデータを気象庁 HP (<http://www.jma.go.jp/jma/index.html>) よりダウンロードして用いた。本研究では 2005 年から 2017 年の期間のデータを収集した。モデルに用いる変数選択においては、正則化回帰手法である Lasso 回帰及び Elastic Net 回帰を用いた。表 1 に示した変数候補は、既往研究及び現場管理者による経験を基に決定した。本研究では 1 週間後の藻類異常発生を予測の目的としたことから、気象データ及び水理学的データについては予測日の過去 7 日間のデータを用いた。

構築する予測モデルは 2 クラス分類を扱うため、各データに対し陽性と陰性の正解ラベルを与える必要がある。本研究では”1 週間後に藻類異常発生する”を陽性、”発生しない”を陰性とした。この際、現場管理者の経験を基に藻類種ごとに藻類異常発生の定義を決定した。*Microcystis* 属については 1ml あたりの藻類細胞数が 10000 個以上の場合を陽性、それ未満を陰性と定義した。*Anabaena* 属については 1mL あたりの藻類細胞数が 100 個以上の場合を陽性、それ未満を陰性とした。

整理されたデータセットは、予測モデル構築用の学習データと、モデルの予測性能検証用の検証データに分割した。データは陽性データ及び陰性データの割合に偏りのある不均衡データであるため、データの偏りを補正する学習データのサンプリング方法としてアンダーサンプリングと Synthetic Minority Over-Sampling Technique (SMOTE) を採用した。

本研究で構築するモデルは、入力データを基にそれぞれのデータに対し藻類異常発生が起きる確率を算出し、発生確率により1週間後に藻類異常発生が起きるかどうかを判定するものである。実際に藻類異常発生が起きた場合に予測に失敗すると藻類による被害が拡大してしまう可能性があることから、実際に藻類異常発生が起きる場合を確実に陽性と予測可能な状況にする必要がある。同時に、実際には藻類異常発生が起きない場合に陽性と予測してしまうことをできるだけ避ける必要がある。そこで、陽性と陰性を分ける閾値を再現率（実際のクラスが陽性であるデータを正しく陽性であるとモデルが予測できた割合）が100%となるように調整し、偽陽性率（False Positive Rate : FPR, 陰性を陽性であるとモデルが予測した割合）が最小となるモデルを最良モデルとした。併せて、各モデルの Area Under ROC Curve (AUC), 正解率, 適合率及びF値も算出した。

## C. 研究結果およびD. 考察

### (1) 変数選択

Lasso 回帰および Elastic Net 回帰を用いた変数選択の結果、室生ダムにおいては風が強いほど *Microcystis* 属が発生しにくいと考えられた。また、リンや窒素といった栄養塩濃度については濃度が高いほど藻類が発生しやすと考えられ、既往研究で得られた知見と一致した。*Anabaena* 属に関しては、リン濃度の回帰係数が0と推定され、室生ダムにおいてリン濃度は *Anabaena* 属の発生に与える影響は弱いという推定結果であった。実際にデータを確認したところ、室生ダムにおける *Anabaena* 属の発生と過去1年間の平均リン濃度の間には明確な関係性が見られなかった。

### (2) 室生ダムにおける藍藻類発生予測結果

室生ダムにおける *Microcystis* 属の発生予測モデルにおける変数選択の結果を表2に、予測結果を表3に示した。最良モデル#5では、AUC : 0.89, FPR : 20.0%, 正解率 : 84.0%, 適合率 : 55.6%と高い予測性能を示した。なお、最良モデル#5はカーネル関数としてガウスカーネル、サンプリング方法に SMOTE を採用して得られたものであった。

室生ダムにおける *Anabaena* 属の予測結果を表4に、予測結果を表5に示した。最良モデル#1では、AUC : 0.83, FPR : 35.0%, 正解率 : 72.0%, 適

合率 : 41.7%と高い予測性能を示した。なお最良モデル#1はカーネル関数としてガウスカーネル、サンプリング方法にアンダーサンプリングを採用して得られたものであった。

### (3) 追加検証

室生ダムにおける予測では高い予測性能が示されたが、本研究で提案する予測モデル構築方法の汎用性を確認するために追加検証を行った。水資源機構の管理するダムの中から阿木川ダム（岐阜県恵那市）、寺内ダム（福岡県朝倉市）、一庫ダム（兵庫県川西市）を追加検証の対象ダム湖として選択し、室生ダムと同様の予測モデル構築方法を適用した。また、室生ダムおよび追加3ダムを全て合わせた合同データによる予測モデル構築も行った。なお、阿木川ダム及び寺内ダムの *Microcystis* 属データについては陽性データがモデル構築を行う上で非常に少なかったため、4ダム合同データを用いた予測においてのみ用いた。

各ダムの最良モデル及びその予測結果を表6、表7、表8に示す。*Microcystis* 属についてはすべてのダムにおいて非常に高い予測性能が得られた。一方で、*Anabaena* 属の予測では寺内ダムを除く3つのダムにおいては高い予測性能が得られた。

寺内ダムにおいて予測性能が低く留まった原因を考察するために、各ダム間でのそれぞれの予測因子についてデータの特徴を確認した。各ダムにおけるデータの平均N/P比を比較すると、寺内ダムは平均51.4と他のダムと比較して高かった。藤本ら<sup>1)</sup>によれば、*Anabaena* 属を含む藍藻類はN/P比が30以上となる場合には優占化しにくい。またSmith<sup>2)</sup>らによる世界各地の17湖沼のデータ解析の結果でも、N/P比が29未満となる場合には優占化しやすくと結論づけられている。これらの知見から、N/P比が30程度を上回ると藍藻類は発生しにくくなると考えられる。しかしながら、寺内ダムでは年に3回程度の *Anabaena* 属の陽性データが確認されていることから、寺内ダムにおいてはN/P比の他に *Anabaena* 属の発生に強い影響を与える因子が存在すると考えられた。

## E. 結論

本研究では、SVMを用いることで、ダム湖における1週間後の藻類異常発生を予測するモデルの開発を行った。*Microcystis* 属に関しては解析した全てのダムにおいて高い予測性能が得られた。各最良モデルにおいては共通して、全窒素濃度、流入量、最高気温、及び風速の変数が用いられ、これらが *Microcystis* 属の予測において重要な予測因子であると考えられた。*Anabaena* 属については、寺内ダムを除いた各ダムにおいて高い予測性能が得られた。各最良モデルにおいては共通して、全窒素濃度、最高気温、及び風速の変数が用いら

れ、これらが *Anabaena* 属の予測において重要な予測因子であると考えられた。寺内ダムにおける *Anabaena* 属の予測結果より、年平均 N/P 比が 30 を大きく超える場合には、本研究では考慮しなかった変数の影響が *Anabaena* 属の発生に関し大きいと考えられた。

F. 健康危険情報

該当なし

G. 研究発表

1. 論文発表

該当なし

2. 学会発表

八島将太, 門屋俊祐, 西村修, 今本博臣, 三浦尚之, 秋葉道宏, 佐野大輔. 浄水障害を引き起こす藻類の水源域における異常発生予測モデルの開発, 第 55 回日本水環境学会年会, 2021.3, オンライン

Sirapat Khodseewong, Gissela Pascual, Munehiro Nomura, Takashi Sakamaki, Osamu Nishimura, Michihiro Akiba. The Mechanism of Aluminium Flocs Formation in Coagulation Process. 第 55 回日本水環境学会年会, 2021.3, オンライン.

H. 知的財産権の出願・登録状況 (予定も含む。)

1. 特許取得

該当なし

2. 実用新案登録

該当なし

3. その他

該当なし

I. 参考文献

1) 藤本尚志, 福島武彦, 稲森悠平, 須藤隆一. 全国湖沼データの解析による藍藻類の優占化と環境因子との関係. 水環境学会誌, Vol. 18, No. 11, pp. 901-908, 1995.

2) V. H. Smith. Low nitrogen to phosphorus ratios favor dominance by blue- green algae in lake phytoplankton. Science, Vol. 221, No. 4611, pp. 669-671, 1983.

表 1. モデル構築における変数候補

データ	単位	説明
cellnumber	cell/mL	モデルの目的変数として、藻類細胞数を用いた。
TP1yp	mg/L	測定日より過去1年間の全リン濃度 (各月1回程度の測定) の平均値を用いた。
TN1yp	mg/L	測定日より過去1年間の全窒素濃度 (各月1回程度の測定) の平均値を用いた。
Inflow7	m <sup>2</sup> /s	測定日より過去7日間の流入量の平均値を用いた。
Discharge7	m <sup>2</sup> /s	測定日より過去7日間の放流量の平均値を用いた。
AveMaxTemp7	°C	測定日より過去7日間の最高気温の平均値を用いた。
AveWind7	m/s	測定日より過去7日間の風速の平均値を用いた。
Sun7	時間	測定日より過去7日間の日照時間の合計を用いた。
Rain7	mm	測定日より過去7日間の降水量の合計を用いた。

表 2. 室生ダムにおける *Microcystis* 属予測モデル

モデル	入力データ
#1	TP1yp, TN1yp, Inflow, Discharge7, AveMaxTemp7, AveWind7, Sun7, Rain7
#2	TP1yp, TN1yp, Inflow, AveMaxTemp7, AveWind7, Sun7, Rain7
#3	TP1yp, TN1yp, Inflow, AveMaxTemp7, AveWind7, Sun7
#4	TN1yp, Inflow, AveMaxTemp7, AveWind7, Sun7
#5	TN1yp, Inflow, AveMaxTemp7, AveWind7

表3. 室生ダムにおける *Microcystis* 属の発生予測結果

	gaussian						linear						polynomial					
	Under Sampling			SMOTE			Under Sampling			SMOTE			Under Sampling			SMOTE		
	AUC	FPR	OC	AUC	FPR	OC	AUC	FPR	OC	AUC	FPR	OC	AUC	FPR	OC	AUC	FPR	OC
#1	0.76	60.0%	47.8%	0.85	45.0%	20.1%	0.66	80.0%	29.8%	0.76	70.0%	7.0%	0.67	80.0%	46.2%	0.77	60.0%	4.8%
#2	0.72	55.0%	42.9%	0.84	40.0%	13.4%	0.72	75.0%	24.2%	0.74	70.0%	4.8%	0.66	80.0%	40.9%	0.78	60.0%	2.8%
#3	0.72	55.0%	46.8%	0.86	40.0%	6.6%	0.67	85.0%	36.4%	0.71	70.0%	7.5%	0.63	90.0%	36.7%	0.74	90.0%	0.0%
#4	0.75	50.0%	34.9%	0.82	20.0%	21.3%	0.65	90.0%	32.4%	0.71	70.0%	9.1%	0.78	75.0%	37.9%	0.75	55.0%	14.9%
#5	0.64	100.0%	0.0%	0.89	20.0%	7.4%	0.63	85.0%	39.1%	0.67	75.0%	9.1%	0.76	70.0%	35.0%	0.70	50.0%	3.9%

表4. 室生ダムにおける *Anabaena* 属予測モデル

モデル	入力データ
#1	TN1yp, Inflow, AveMaxTemp7, AveWind7, Rain7
#2	TN1yp, Inflow, AveMaxTemp7, Rain7
#3	TN1yp, Inflow, AveMaxTemp7

表5. 室生ダムにおける *Anabaena* 属の発生予測結果

	gaussian						linear						polynomial					
	Under Sampling			SMOTE			Under Sampling			SMOTE			Under Sampling			SMOTE		
	AUC	FPR	OC	AUC	FPR	OC	AUC	FPR	OC	AUC	FPR	OC	AUC	FPR	OC	AUC	FPR	OC
#1	0.83	35.0%	24.5%	0.59	75.0%	4.2%	0.54	85.0%	12.6%	0.53	85.0%	2.8%	0.78	80.0%	28.6%	0.79	75.0%	0.0%
#2	0.76	40.0%	33.5%	0.46	100.0%	0.0%	0.55	85.0%	10.5%	0.52	85.0%	2.7%	0.73	80.0%	10.4%	0.66	80.0%	0.0%
#3	0.70	40.0%	31.3%	0.56	80.0%	4.2%	0.51	85.0%	8.1%	0.51	85.0%	1.7%	0.63	80.0%	5.7%	0.64	80.0%	0.1%

表6. 4ダムにおける最良モデル

最良モデルの入力データ		
室生ダム	<i>Microcystis</i>	TN1yp, Inflow, AveMaxTemp7, AveWind7
	<i>Anabaena</i>	TN1yp, Inflow, AveMaxTemp7, AveWind7, Rain7
寺内ダム	<i>Microcystis</i>	-
	<i>Anabaena</i>	TN1yp, AveMaxTemp7, Sun7
一庫ダム	<i>Microcystis</i>	TN1yp, Inflow, AveMaxTemp7, AveWind7
	<i>Anabaena</i>	TN1yp, AveMaxTemp7, AveWind7, Rain7
4ダム合同	<i>Microcystis</i>	TP1yp, TN1yp, Inflow, AveMaxTemp7, Sun7, AveWind7
	<i>Anabaena</i>	TP1yp, TN1yp, Discharge7, Inflow7, AveMaxTemp7, Sun7, AveWind7

表 7. 4 ダムにおける *Microcystis* 属の発生予測結果

	室生ダム	寺内ダム	一庫ダム	4ダム合同
AUC	0.89	-	0.95	0.93
FPR	20.0%	-	10.0%	15.0%
正解率	84.0%	-	92.0%	88.0%
適合率	55.6%	-	71.4	62.5
再現率	100.0%	-	100.0%	100.0%
F値	0.71	-	0.83	0.77

表 8. 4 ダムにおける *Anabaena* 属の発生予測結果

	室生ダム	寺内ダム	一庫ダム	4ダム合同
AUC	0.83	0.60	0.81	0.82
FPR	35.0%	60.0%	25.0%	25.0%
正解率	72.0%	52.0%	80.0%	80.0%
適合率	41.7%	29.4%	50.0%	50.0%
再現率	100.0%	100.0%	100.0%	100.0%
F値	0.59	0.45	0.67	0.67

