

水中の病原ウイルス濃度レベルを予測するソフトセンサーの開発

(研究期間：令和6年度～令和7年度)

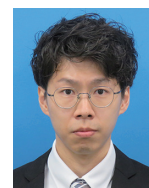
上下水道研究部 浄水処理研究室

研究官 (博士(工学)) 門屋 俊祐* 室長 (博士(工学)) 島崎 大

浄水処理・水道防災
システム研究官
(博士(工学))

増田 貴則

(*:写真掲載者)



(キーワード) ウイルス、濃度レベル、クラス分類予測

3.

快適で安心な暮らしを支える研究

1. 研究背景および目的

気候変動に伴う降雨の激甚化は、合流式下水道越流水や野生動物糞便等の環境水への流入増加を招き、高濁度原水を発生させる。高濁度原水中には病原微生物（細菌、原虫、ウイルス）が含まれる可能性があることから、水道事業体は給水栓水の安全性を維持するために、濁度管理や残留塩素濃度の管理を徹底する必要がある。しかしながら、諸外国とは異なり、我が国ではウイルスに対する水道水の安全性の基準が設けられておらず、検出機器の導入及び人材確保に係る負担もあり、ウイルスの“常時”監視を可能とする水道事業体は現状では少ないと考えられる。

本研究では、水質及び気候条件等の入手が容易な情報をもとにウイルス濃度“レベル”を予測可能なモデルを構築することで、環境水中のウイルスの監視を補助することが可能なツールの開発を目的とした。ウイルス濃度の多寡に応じて分類した“レベル”として予測することで、測定データに含まれる誤差が予測モデルに与える影響を低減することを試みた。

2. 研究方法

(1) データ収集

関東圏の下水道が整備された水域において糞便汚染実態調査を実施し（2021-2023年）、取得したデータをモデル構築に供した。本研究における予測対象は、ヒトへの病原性を示すアイチウイルス（AiV）及び糞便汚染指標であるトウガラシ微斑ウイルス（PMMoV）とした。ウイルス濃度とともに測定された濁度等の水質情報及び降雨量等の気象情報の計12を説明変数とした。

(2) データ生成とウイルス濃度レベル

モデルの汎用性の正確な評価には多数のデータが必要なため、各変数の確率分布から乱数を発生させ仮想データを生成した。主成分分析により、仮想データが降雨や濁度等の環境条件によって分類され（図-1）、元データの特性が仮想データに保持されていることを確認した。ウイルス濃度データは、確率分布のパーセンタイル値をもとにLow (L)、Middle (M)、High (H) の3クラスに分類した（図-2）。

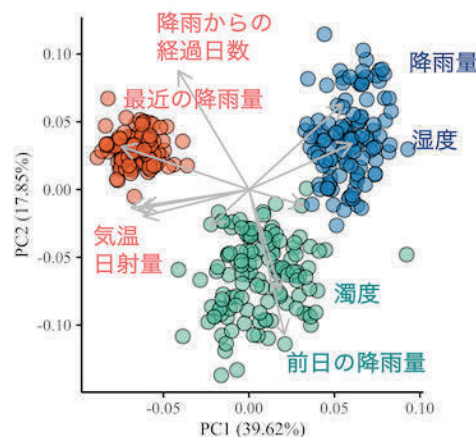


図-1 仮想データセットの特性の可視化

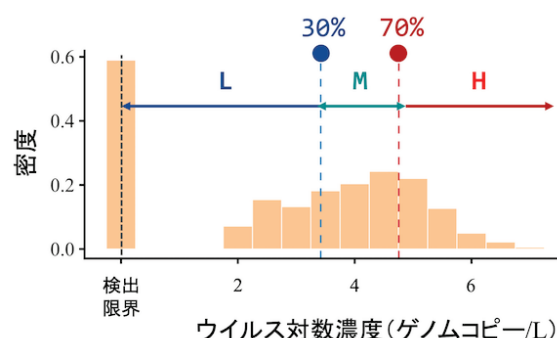


図-2 確率分布に基づくウイルス濃度レベルの3クラス分類

（３）クラス分類予測と性能評価

クラス分類アルゴリズムとして、サポートベクターマシン（SVM）、ランダムフォレスト（RF）、XGBoost（XGB）を適用し、最適なアルゴリズムの特定を試みた。さらに、相互作用項（例：変数a×変数b）の追加により予測精度が向上するか検証した。

モデルのデータへの適合度と予測精度をそれぞれ評価するため、データをモデル構築用データ及び評価データにあらかじめ分割した。本研究では、“高い”濃度レベルを“低い”と誤判定された際（偽陰性）に罰則が課される予測精度指標Recall（1に近いほど高い予測精度を示す）を採用した。

3. 結果および考察

最良な予測モデルの基準は、予測精度指標が訓練及び評価データでともに1に近く（その差が小さいとなお良く）、また、水利用にかかる感染制御という観点から、特にレベルHの予測性能が高いことが望ましい。以上を踏まえると、AiV、PMMoVともに相互作用項を適用したRFが現時点で最良のモデルと言える（表-1、2）。レベルL、Mの評価データに対するRecallの値は0.5～0.6であるが、感染リスクが高いことが懸念されるレベルHの予測に対しては0.7程度と高いRecall値を示したためである。

表-1 AiVに対するRecall値（訓練/評価）

	SVM	RF	XGB
L	0.99/0.55	0.95/0.63	0.92/0.63
M	1/0.47	1/0.52	0.95/0.39
H	1/0.5	1/0.73	0.94/0.48

表-2 PMMoVに対するRecall値（訓練/評価）

	SVM	RF	XGB
L	0.65/0.69	0.95/0.69	0.69/0.67
M	0.71/0.67	0.99/0.69	0.70/0.64
H	0.71/0.67	1/0.71	0.74/0.73

予測精度は説明変数の種類を増やすことにより改善する可能性がある。水道事業体においては常時監視する水質項目を増やすことは困難と考えられるが、地域の感染流行状況等の疫学情報を説明変数として取り込むことは有用かつ実現可能と考えられる。また、より多くのデータ数が利用可能であれば、高度な機械学習アルゴリズムへの適用が可能となり、予測精度の向上が見込まれる。

本研究は環境変数を用いたウイルス濃度“レベル”の予測可能性を評価したものであり、将来予測性については十分な検討はなされていない。しかしながら、気象予報値や各種水質項目を予測するモデルの出力値等と本研究で示したモデル構築方法を組み合わせることで、将来のウイルス濃度“レベル”予測モデルの構築が可能となる。ただし、気候変動影響が年々変化する場合、新たなデータセットを入力するなどモデル更新を定期的に行う必要がある。今後は適切な更新頻度も含めて検討を進めていく。

4. 結論

水質及び気象データを説明変数とした環境水中のウイルス濃度“レベル”予測モデルの構築を行い、RF+相互作用項モデルが現段階で最良の予測精度を示した。予測精度の一層の向上のためには、他に入手が可能な情報を説明変数として利用検討することが第一ステップとなると考えられた。

5. 謝辞

本研究の一部は、クリタ水・環境科学振興財団2024年度国内研究助成萌芽的研究b（24H025）及び令和7年度技術研究開発推進費により遂行された。