

実証研究成果報告書

1. 依頼書受付番号 実支 第31-001号

2. 実証研究の名称

ディープラーニングを活用した自律型浄水支援システムの開発

3. 研究成果の概要

本研究は、技術系の人材不足を補い、また、浄水プロセスにおける意思決定を支援するためのシステム開発を目的とした。テーマは2つあり、「1) ビッグデータを活用した原水水質の予測」では、浄水場に蓄積した原水水質等のビッグデータを活用し、数時間後のかび臭濃度を予測すること、「2) 画像解析を用いた凝集判定手法の研究」では、緩速攪拌直後の凝集フロックの画像データから凝集の良否を判定することとし、どちらのテーマも深層学習（ディープラーニング）を使用した。各テーマの結果は以下のとおりである。

1) ビッグデータを活用した原水水質の予測

- ・LSTM（長・短期記憶）を使用することで、1時間毎の測定データから3時間後の原水のかび臭濃度を予測することができた（本文 pp.19～26）。
- ・構築したモデルを別の浄水場に適用する場合、同じ水系には学習時の重みを再利用することはできるが、別の水系には再利用できない（本文 p.30）。
- ・データ収集間隔が1日と長い場合には、1日後のかび臭物質濃度の予測精度は低い（本文 p.34）。
- ・高濃度にかび臭が検出される場合にはバッチ正規化を行うことによりかび臭濃度をほぼ正確に予測することができた（本文 pp.46～49）。
- ・モニター上の数値を使用する場合には、学習時にモニターの数値の桁数を揃えて学習する必要がある（本文 pp.56～57）。

2) 画像解析を用いた凝集判定手法の研究

- ・カメラのレンズ上に付着する異物を削除する前処理法として暗合成-差分法を用いることによってモデルの予測精度が5%程度高まる（本文 p.72）。
- ・水温別にモデルを構築すると全体的に精度が向上する（本文 p.80）。
- ・凝集剤（PAC）の注入率が不足している場合の予測精度は高い（本文 p.81）。
- ・凝集沈澱処理水濁度の予測クラスを3分類ではなく2分類にすることで精度は15%以上向上し、その予測精度は85%程度となる（本文 p.79）。
- ・画像による予測判定クラスモデルに加え、凝集沈澱処理水濁度の安全性を担保する安全性補完モデルを併用することでより安全な凝集沈澱の操作が可

能となる（本文 pp.88～90）。

4. 適用条件

1) ビッグデータを活用した原水水質の予測

- ・入力データは原水のかび臭物質を含む 8 項目を基本とする。
- ・かび臭物質は季節によって変動するため少なくとも 1 年間のデータが必要である。
- ・3 時間後の予測を行うためには、1 時間毎のデータが必要である。

2) 画像解析を用いた凝集判定手法の研究

- ・原水の水温や濁度によって影響されるため、実験を行う場合は、学習データとして 1 年間分のフロック画像のデータが必要になる。
- ・画像は緩速攪拌直後のフロックを対象とする。
- ・カメラ画質はフル HD 以上を推奨する。

5. 研究成果の活用方法

かび臭濃度を事前に予測することで濃度に応じた適正な粉末活性炭が注入できること、凝集フロックにより沈澱濁度の良否を早期に判断することで、PAC 注入率不足や沈澱処理水濁度の異常時に早急に対応することができ、適正な凝集沈澱池の管理が行えること等が考えられる。

令和 6 年 2 月 26 日

企業名 前澤工業株式会社

報告者 根本雄一、凌海、大澤裕志

ディープラーニングを活用した 自律型浄水支援システムの開発



前澤工業株式会社

中央大学

北海道科学大学

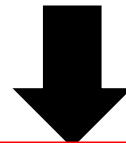


報告内容

1. 研究背景
2. 研究目的
3. 研究内容
4. 研究体制
5. 研究結果 —ビッグデータを活用した原水水質の予測—
6. 研究結果 —画像解析を用いた凝集判定手法の研究—
7. 達成目標と検証
8. 適用条件
9. 研究成果の活用方法

1. 研究背景

- 技術系職員の退職に伴う人材不足や技術継承不足
- 気候変動による原水の水質異常に対する浄水処理の対応が困難
- 浄水場に運転管理データや水質データ等のビッグデータが保存されているが、有効に活用されていない



- 技術系職員がいなくとも水質異常等を簡便かつ迅速に判定できるシステムが開発できないか？
- ビッグデータを有効に活用できないか？

2. 研究目的

技術系職員の不足を補い、浄水処理プロセスにおける意思決定を支援するため、

深層学習（ディープラーニング）を用いて、

- 1) 蓄積された浄水場の水質や運転等のビッグデータを解析し、精度の高い水質予測システムを開発する
- 2) 画像による凝集フロック形状から、早期に異常を発見するシステムを開発する

JWRC 公募概要テーマ③『情報技術の活用による水道事業の基盤強化に資する研究かつ実フィールドでの実証を必要とする研究』に該当

3. 研究内容

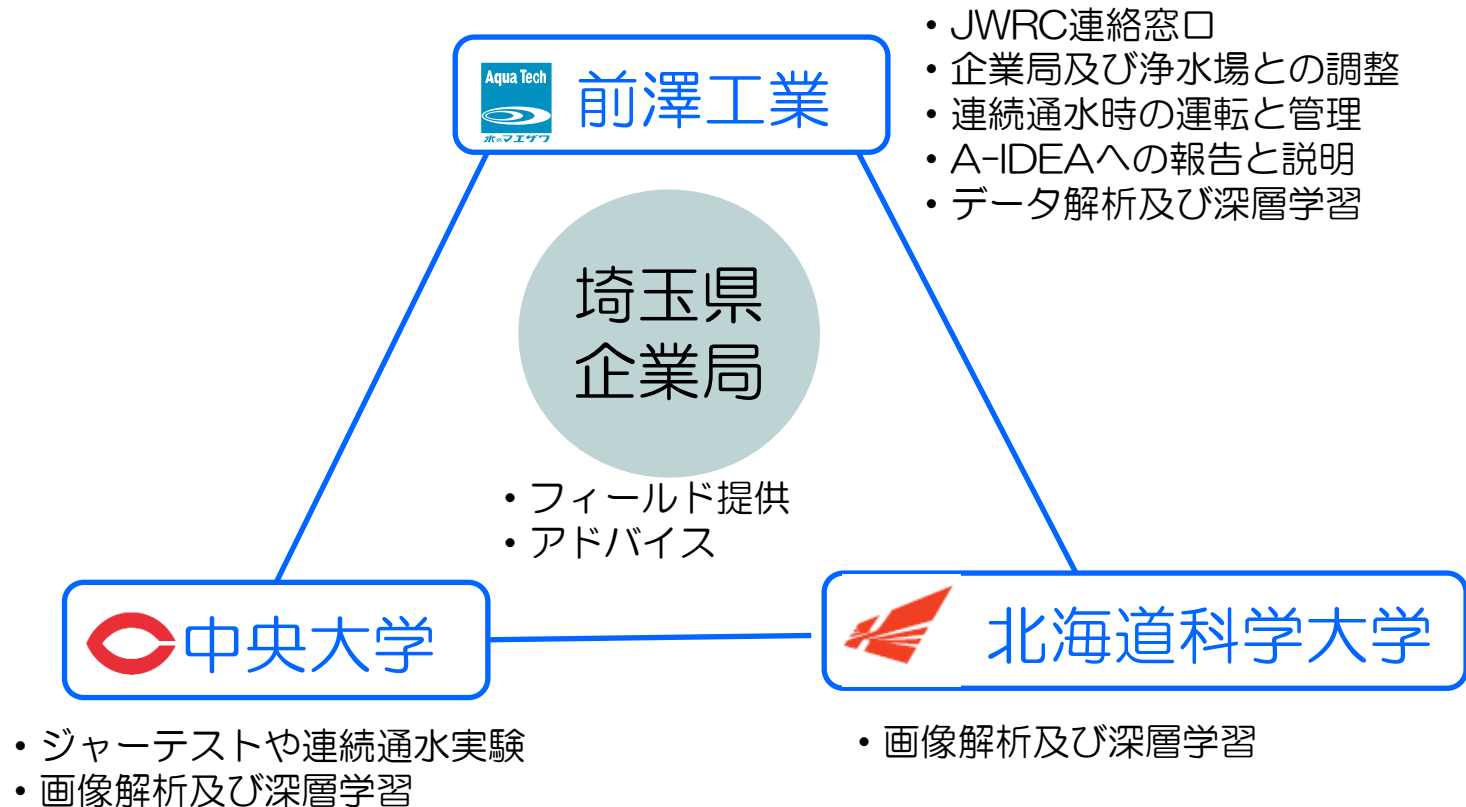
1) ビッグデータを活用した原水水質の予測

原水の水質データ、環境データ等を収集・解析し、**かび臭の濃度を予測**する。

2) 画像解析を用いた凝集判定手法の研究

ビデオカメラを用いて緩速攪拌直後の凝集フロック画像を撮影し、そのフロック形状等から**凝集の適不適を判断**する。

4. 研究体制





5. 研究結果

ビッグデータを活用した
原水水質の予測



埼玉県企業局の3浄水場における かび臭予測モデルと精度

研究背景 (1)

- 平成26年に荒川水系において、かび臭物質(2-MIB)を生成する藻類(*Phormidium autumnale*)の生息を確認
- かび臭物質(2-MIBやジェオスミン)の濃度は水道水質基準より10 ng/L以下と定められている



GC-MS 高価な分析装置 高度な分析スキル ▶ 月に一度程度の定期分析のみ測定



浄水場で収集・保存しているデータから
かび臭濃度を容易に推定する手法はない

研究背景 (2)

吉見浄水場の処理フローと測定している水質

原水

沈砂池

着水井

急速
攪拌池

沈澱池

ろ過池

浄水

○濁度
○ 2-MIB
○ ジェオスミン
○濁度
○pH
○アルカリ度
○導電率
○塩素要求量
○水温

○濁度
○ pH

○ pH
○残塩

○濁度
○ pH
○残塩

○濁度
○ pH
○残塩

○濁度
○ pH
○残塩
○水温
○アルカリ度
○トリハロメタン

- 吉見浄水場では2-MIBとジェオスミンを含む水質データを時間単位で連続測定している

→ 大量のデータがありビッグデータとして有効活用できないか？

これらのビッグデータを活用することで、かび臭濃度を予測するモデルを開発できるのではないか？

実験方法 ～入力データ～

原水

- 濁度
- 2-MIB
- ジェオスミン

沈砂池

- pH
- アルカリ度
- 導電率
- 塩素要求量
- 水温

重要な因子

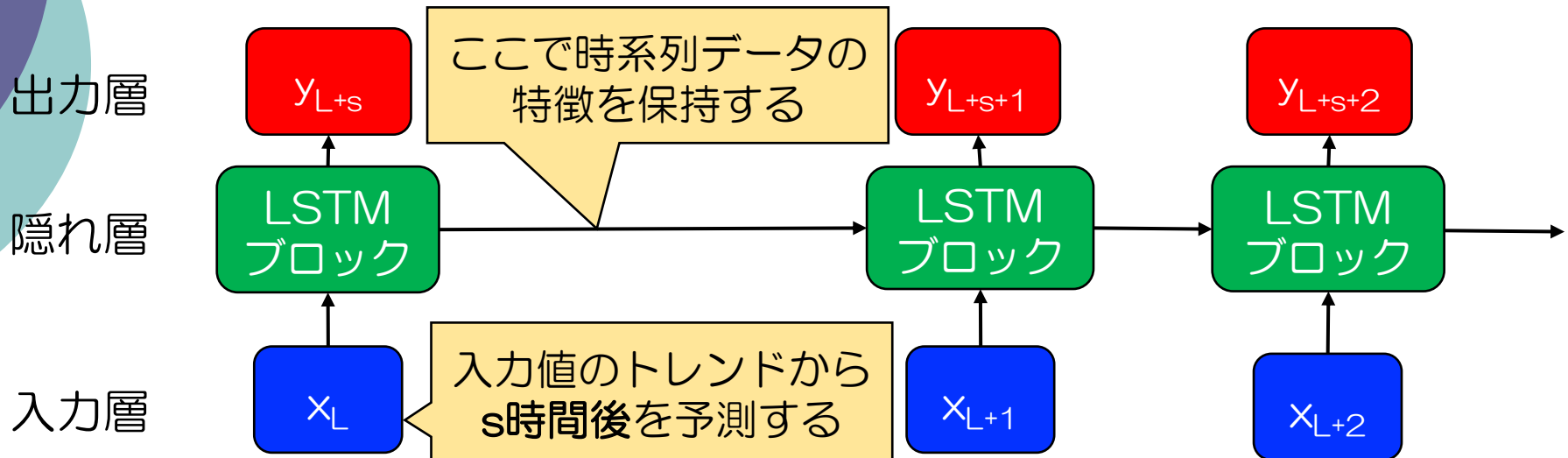
ダム放流, 栄養塩, 水温, アルカリ度

- | | | |
|----------|---|-------------|
| 1. 濁度 | → | 降雨、ダム放流の指標 |
| 2. pH | → | 藻類活動、降雨の指標 |
| 3. アルカリ度 | → | 藻類活動、降雨の指標 |
| 4. 導電率 | → | 栄養塩(N、P)の指標 |
| 5. 塩素要求量 | → | 有機物濃度の指標 |
| 6. 水温 | → | 藻類の活動に影響 |
| 7. 河川水位 | → | 降雨、ダム放流の指標 |

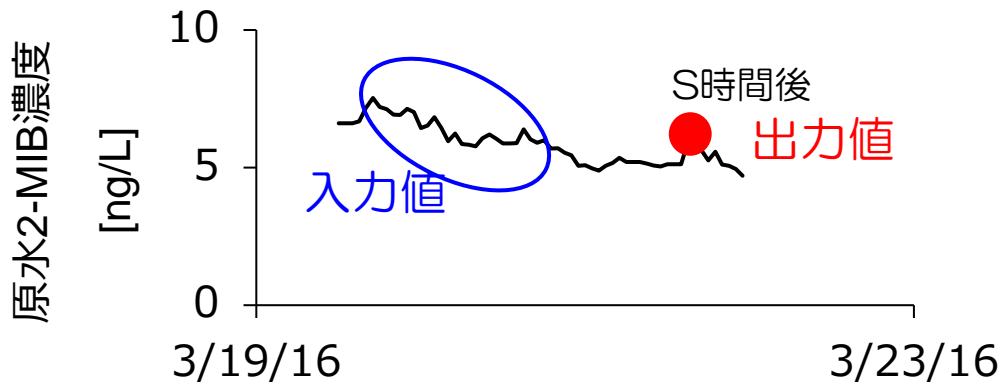
原水のかび臭物質（2-MIBかジェオスミンのどちらか）と重要な因子と考えられる7項目の計8項目を入力データとした

実験方法 ～LSTM（長・短期記憶）～

- 時系列データを扱うことができるニューラルネットワークの発展モデル



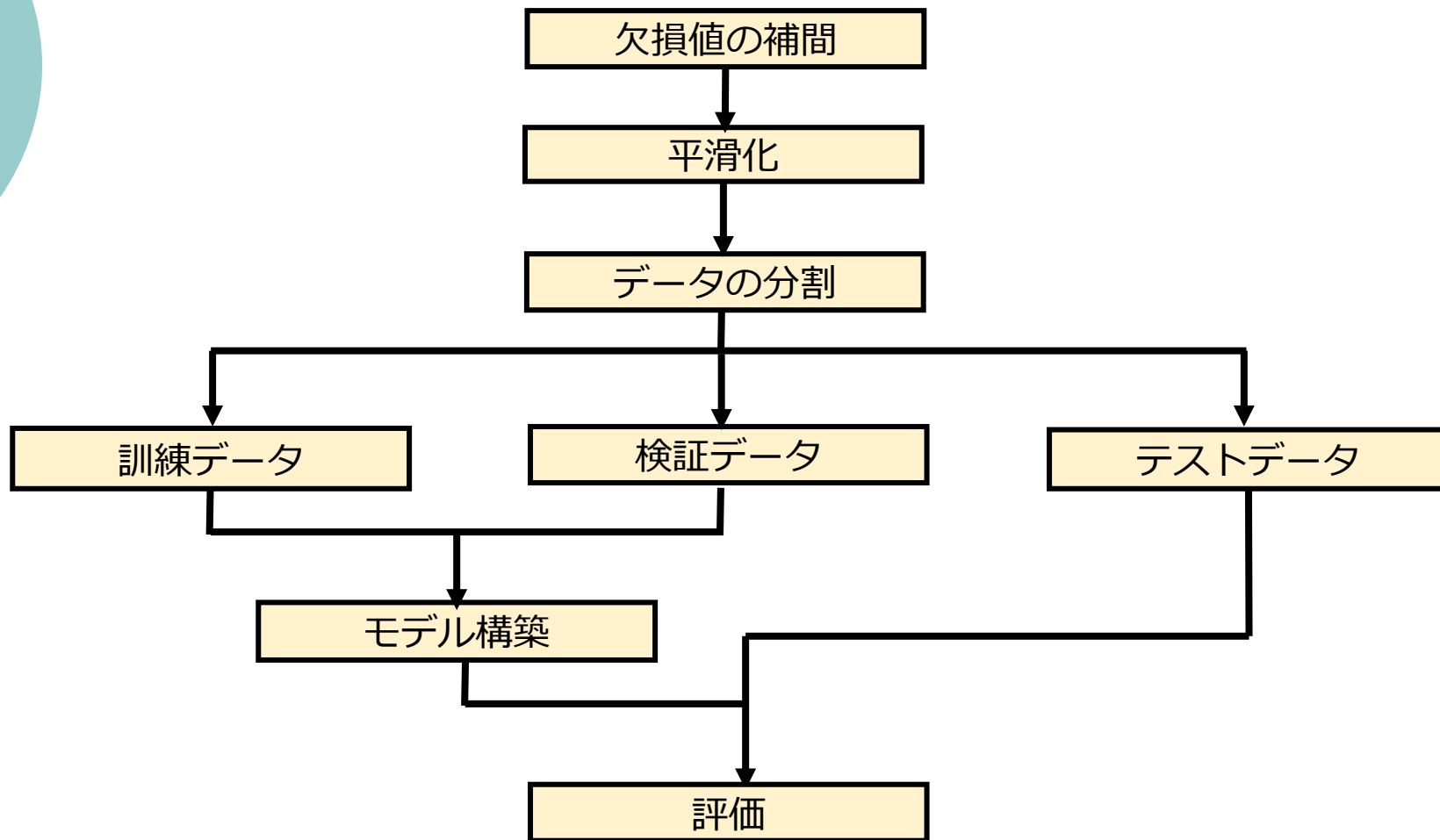
$x_L = \{x_t, x_{t+1}, x_{t+2}, \dots, x_{t+L-1}\}$: 時刻 t から時刻 t+L-1 までの連続データ



ブロック : 24、48時間
 予測時間 : 3~24時間

※上図は一例

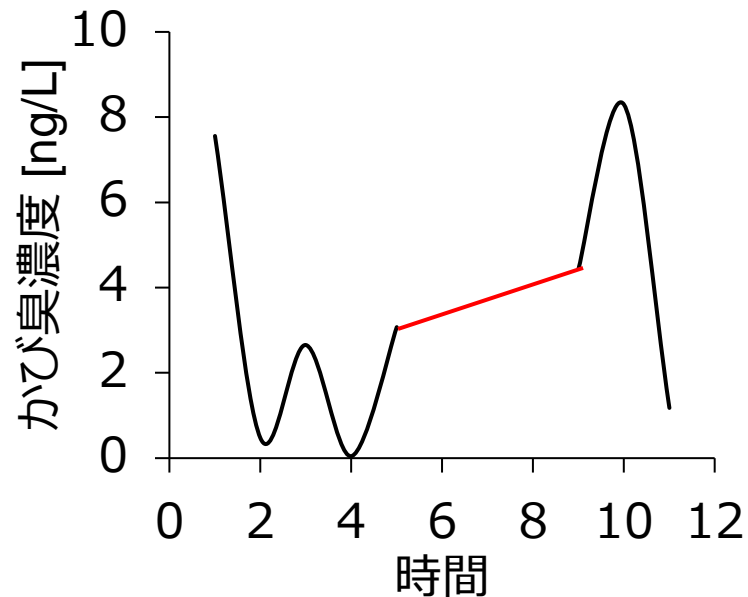
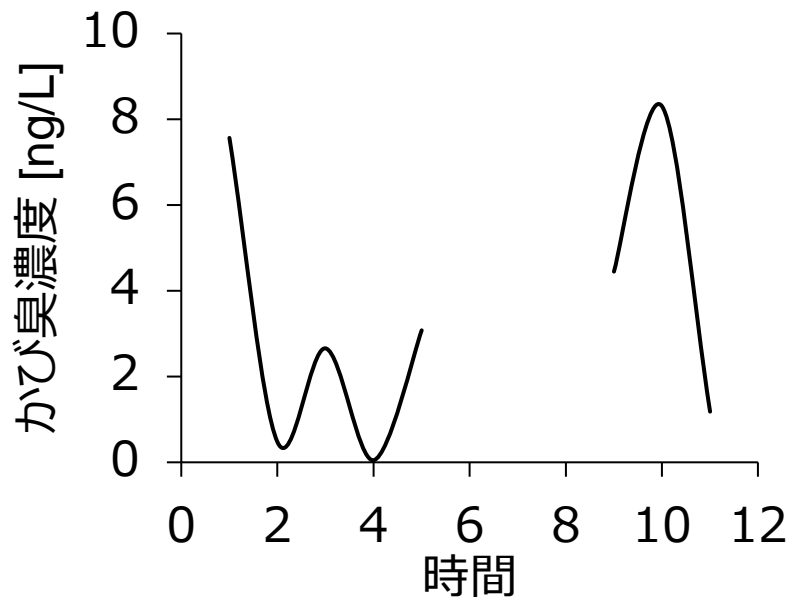
実験方法 ～モデル構築フロー～



実験方法 ～前処理方法～

欠損値：線形補間

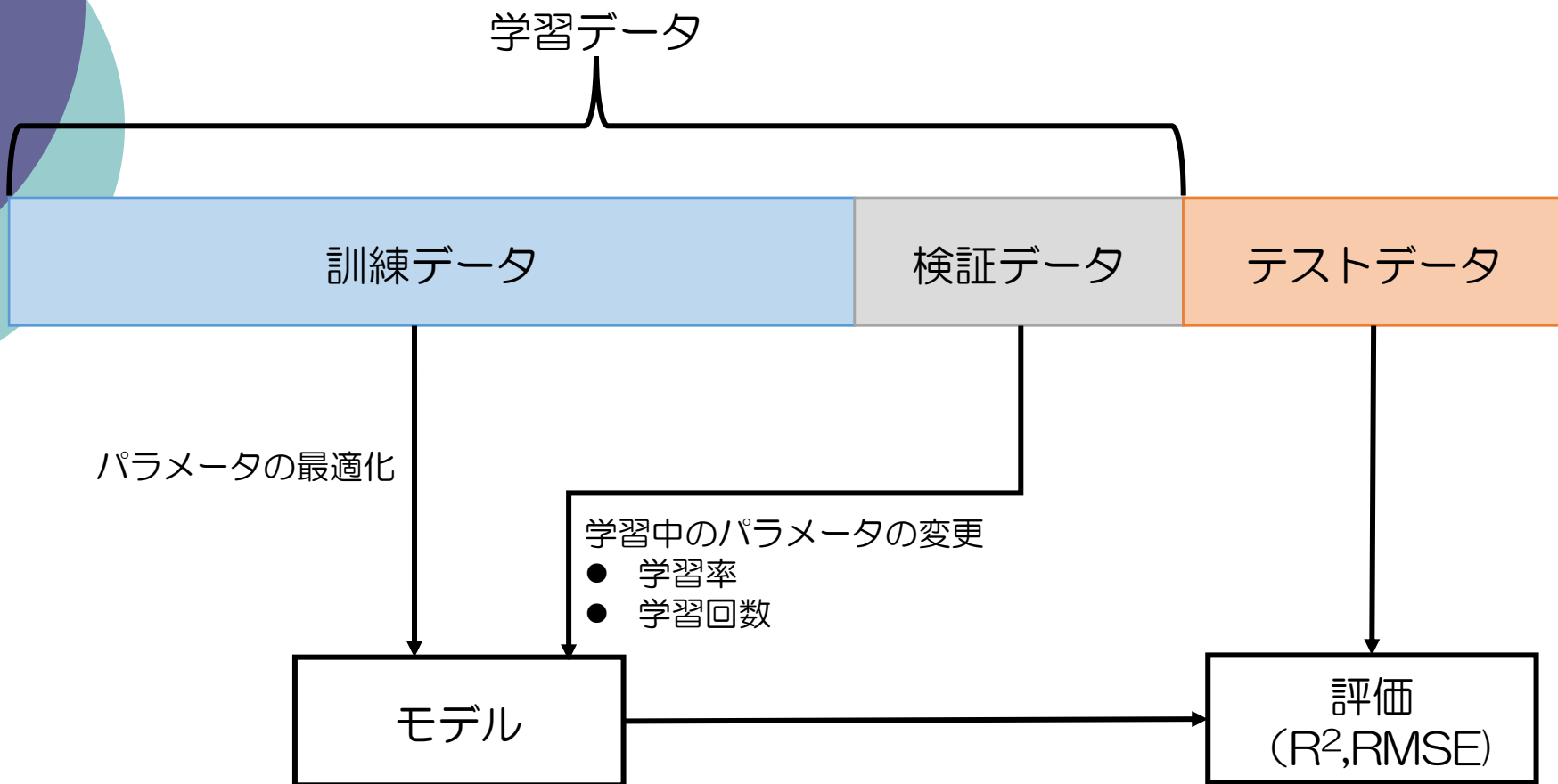
※値は例です。



平滑化

24時間の移動平均を取ることで微小な変化やノイズを抑制

実験方法 ～データの分割と用途～



※RMSE : Root Mean Squared Error $\sqrt{\frac{1}{n} \sum_{i=1}^n (X_i - \widehat{X}_i)^2}$

※R²値

実験方法 ～データセット～

		吉見浄水場	大久保浄水場	行田浄水場
訓練データ	始	2016/3/17	2015/2/4	2018/10/1
	終	2018/5/31	2018/9/30	2019/7/31
検証データ	始	2018/6/1	2018/10/1	2019/8/1
	終	2018/8/31	2019/3/31	2019/9/30
テストデータ	始	2018/9/1	2019/4/1	2019/10/1
	終	2020/7/1	2020/7/1	2020/10/1

2-MIB

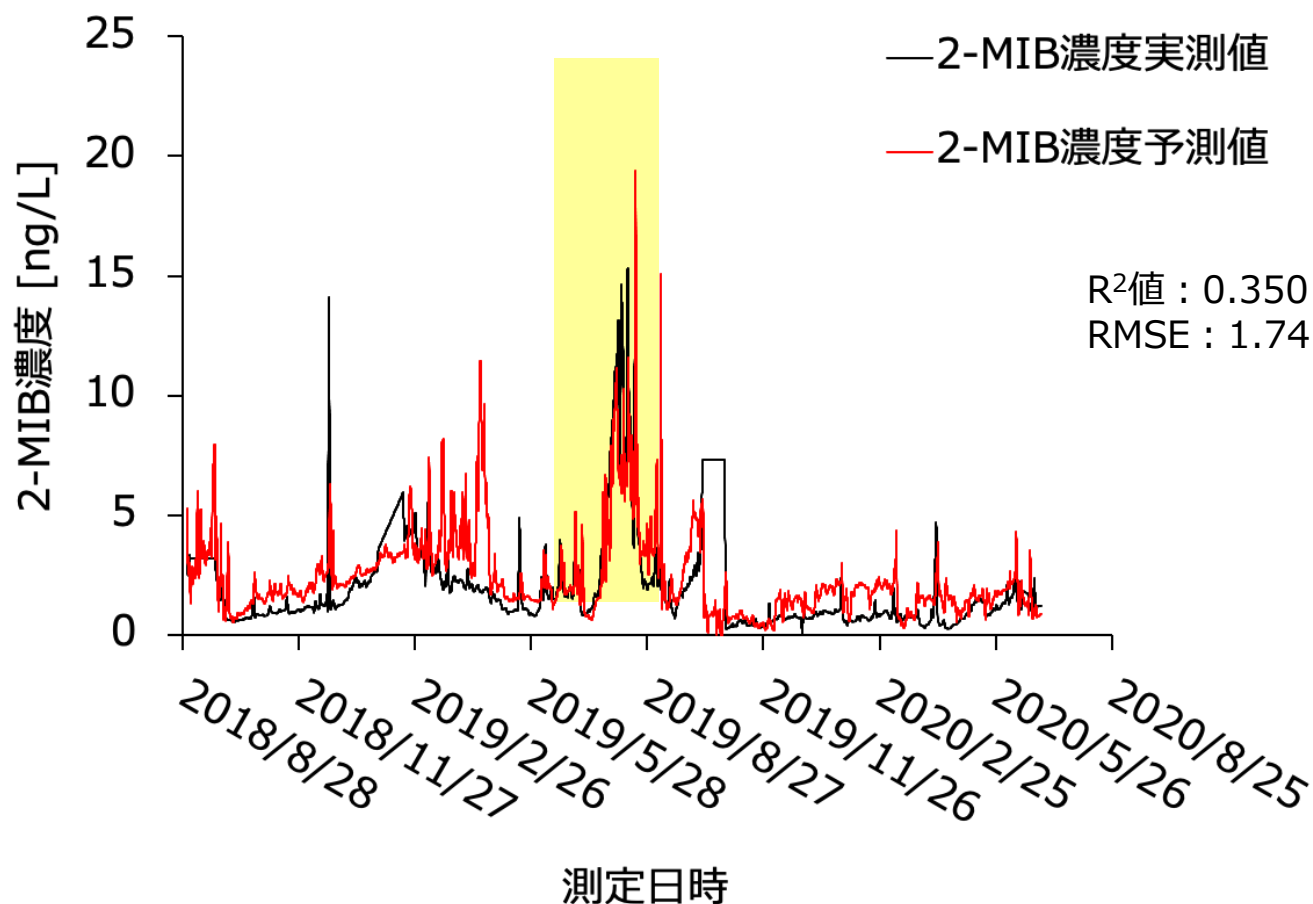
ジェオスミン

対象となるかび臭物質※

※浄水場によって顕著に検出されるかび臭物質の種類が異なる

実験結果

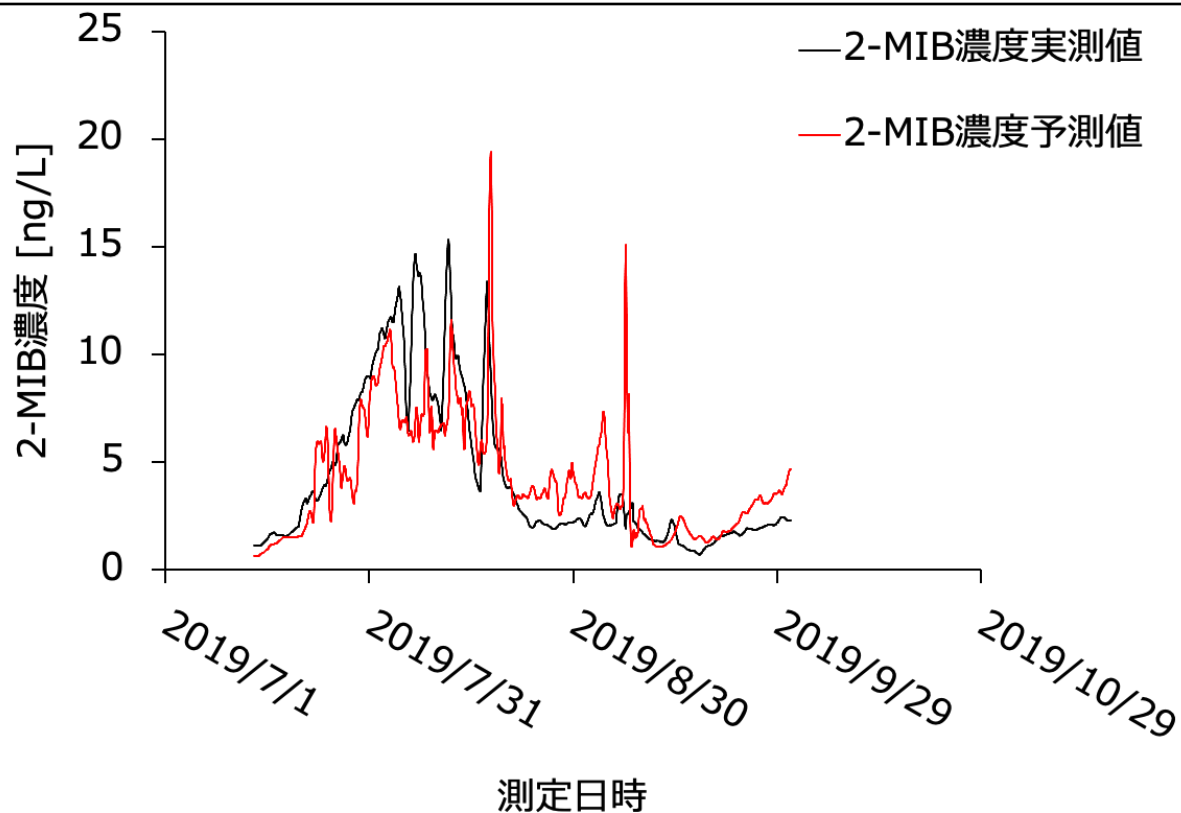
～2-MIB濃度の24時間後の予測（吉見浄水場）～



黄色の部分を拡大したものを次スライドに示す

実験結果

～2-MIB濃度の24時間後の予測（吉見浄水場）～



発生予測はできていたが、24時間後の正確な濃度予測はできていない
⇒予測時間を24時間よりも短い時間で精度を確認した

実験結果 ～2-MIB濃度の予測（吉見浄水場）～

ブロックと予測時間による精度

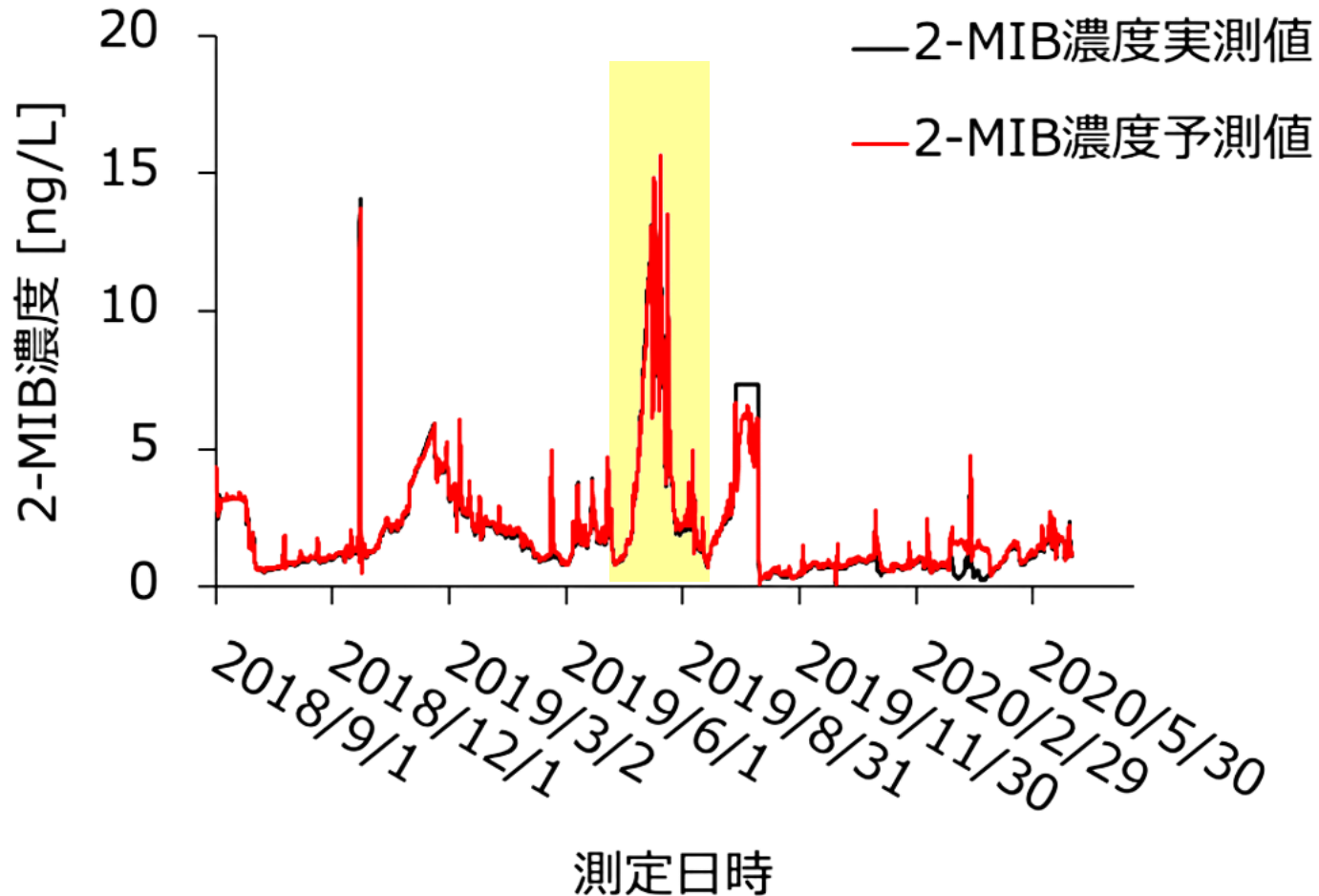
ブロック [時間]	予測時間 [時間]	1回目		2回目		3回目		4回目		5回目	
		R ²	RMSE	R ²	RMSE	R ²	RMSE	R ²	RMSE	R ²	RMSE
24	3	0.896	0.692	0.962	0.401	0.968	0.369	0.962	0.401	0.987	0.247
24	6	0.896	0.692	0.919	0.600	0.923	0.580	0.875	0.733	0.891	0.709
24	9	0.861	0.790	0.772	0.963	0.838	0.888	0.881	0.725	0.877	0.752
24	12	0.024	5.76	0.689	1.13	0.689	1.11	0.764	0.991	0.764	0.983
48	3	0.993	0.192	0.983	0.263	0.980	0.287	0.983	0.261	0.974	0.322
48	6	0.281	0.981	0.476	2.58	0.976	0.314	0.908	0.609	0.945	0.472
48	9	0.893	0.689	0.915	0.585	0.886	0.677	0.812	0.871	0.906	0.614
48	12	0.102	4.75	0.868	0.730	0.004	208	0.761	0.983	0.705	1.09

※表中の赤字は5回の学習における最も高い精度を示す

ブロックによる差は小さくなく、どのブロックも予測時間が小さいほど精度が高まり、3時間後の精度が最も高かった

実験結果

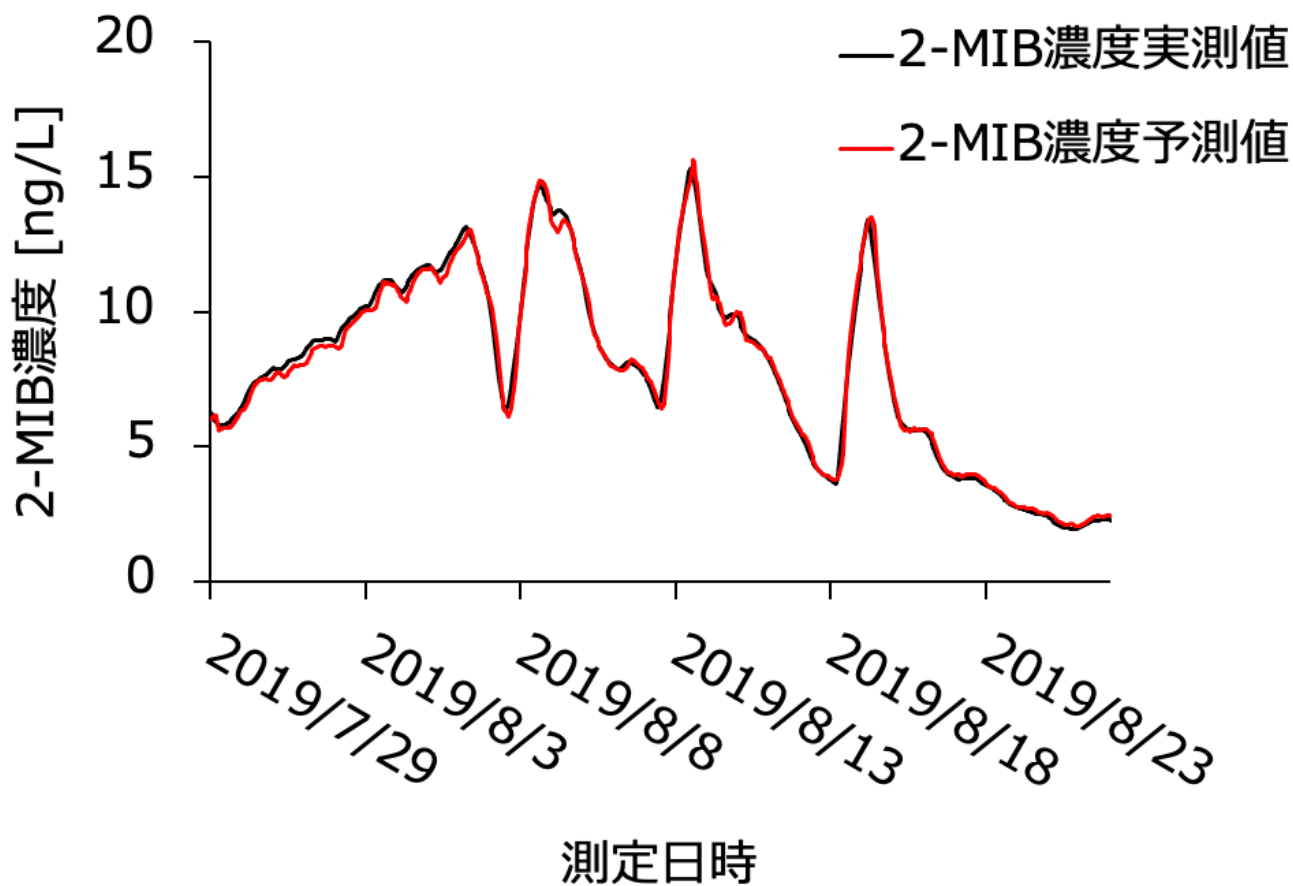
～2-MIB濃度の3時間後の予測（吉見浄水場）～



黄色の部分を拡大したものを次のスライドに示す

実験結果

～2-MIB濃度の3時間後の予測（吉見浄水場）～



3時間後の予測精度は高い

実験結果 ～2-MIB濃度の予測（大久保浄水場）～

ブロックと予測時間による精度

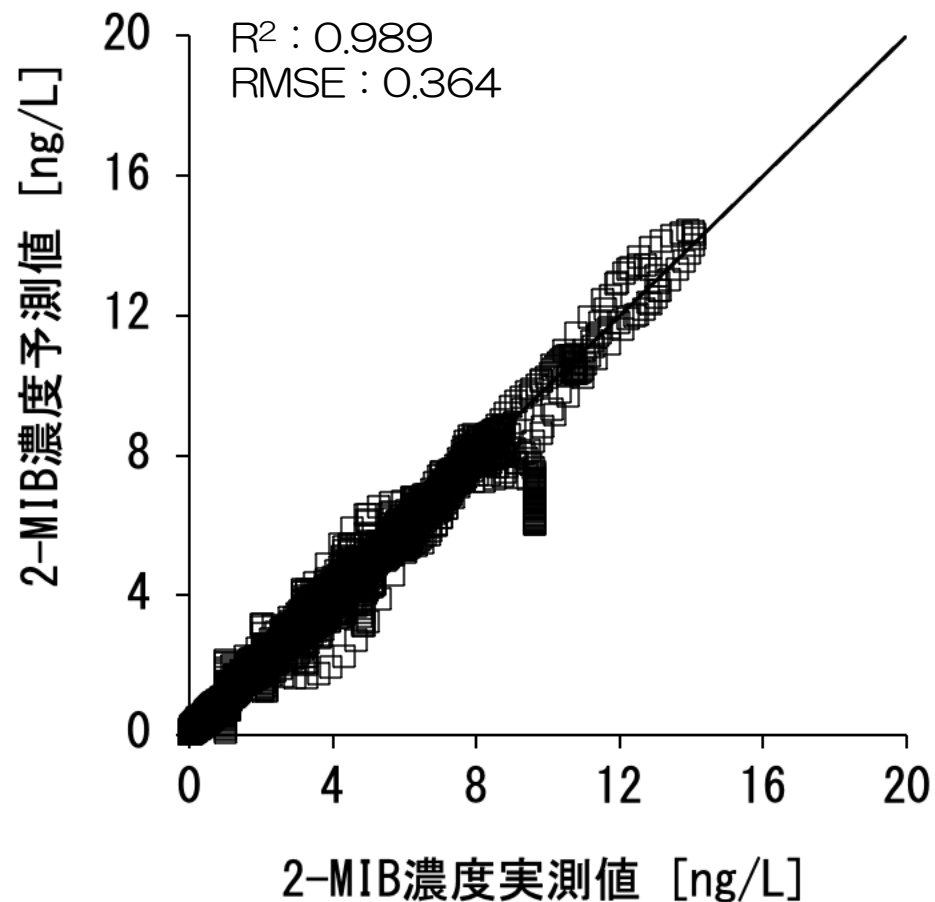
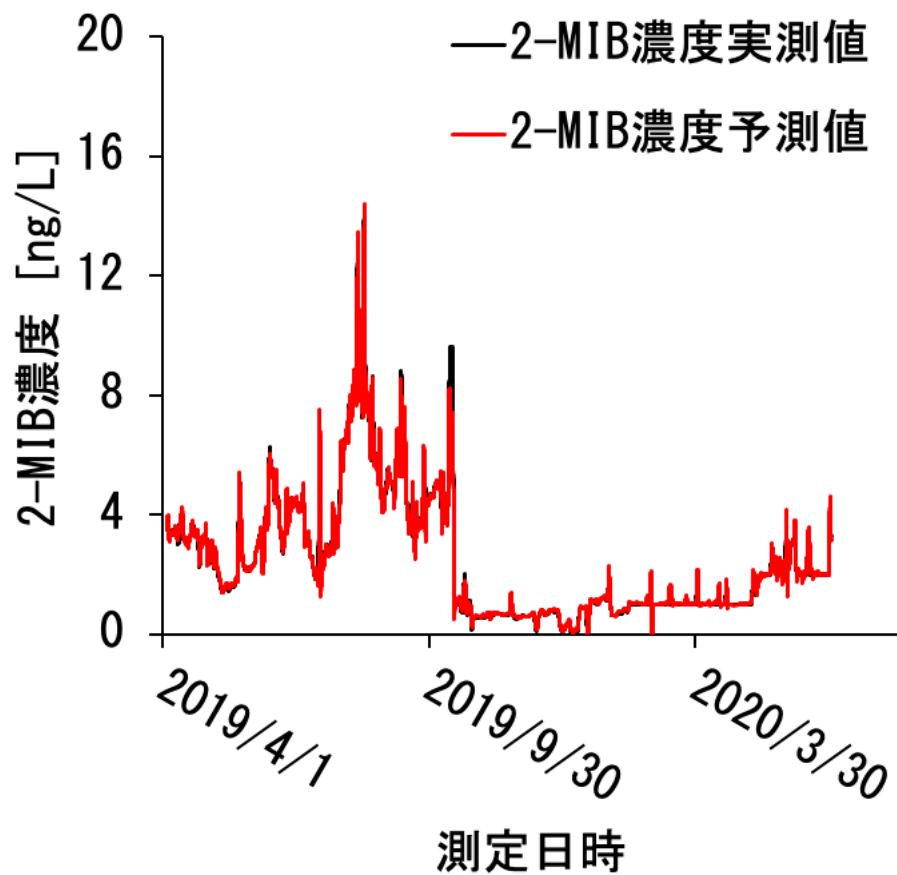
ブロック [時間]	予測時間 [時間]	1回目		2回目		3回目		4回目		5回目	
		R ²	RMSE	R ²	RMSE	R ²	RMSE	R ²	RMSE	R ²	RMSE
24	3	0.989	0.364	0.976	0.331	0.973	0.352	0.951	0.475	0.959	0.436
24	6	0.896	1.08	0.813	0.926	0.959	0.435	0.895	0.696	0.726	1.119
24	9	0.916	0.859	0.629	1.30	0.916	0.621	0.927	0.579	0.900	0.679
24	12	0.886	0.761	0.815	0.921	0.785	0.994	0.807	0.942	0.813	0.927
48	3	0.978	0.422	0.952	0.470	0.042	3.879	0.934	0.550	0.134	1.995
48	6	0.972	0.447	0.958	0.438	0.968	0.384	0.894	0.699	0.963	0.411
48	9	0.212	1.90 × 10 ⁴	0.510	1.50	0.054	2.370	0.686	1.200	0.000	4.47 × 10 ⁷
48	12	0.916	1.10	0.874	0.762	0.754	1.062	0.819	0.913	0.000	7.07 × 10 ⁴

※表中の赤字は5回の学習における最も高い精度を示す

大久保浄水場も吉見浄水場と同様な傾向を示していた

実験結果

～2-MIB濃度の3時間後予測（大久保浄水場）～



実験結果 ～ジェオスミン濃度の予測（行田浄水場）～

ブロックと予測時間による精度

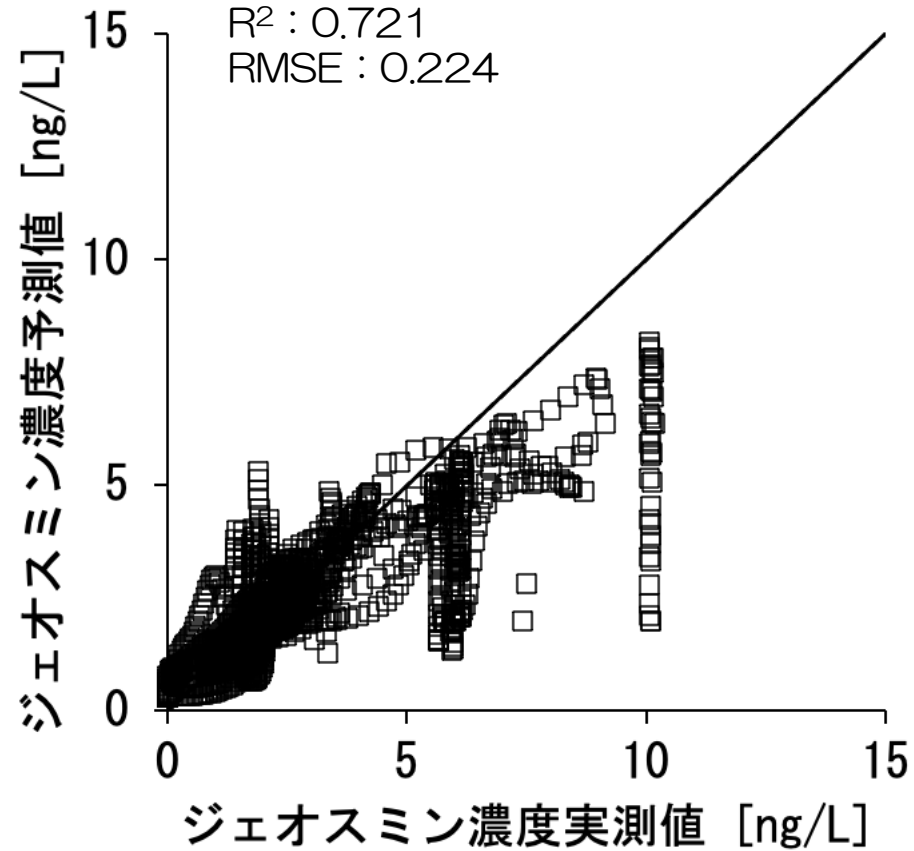
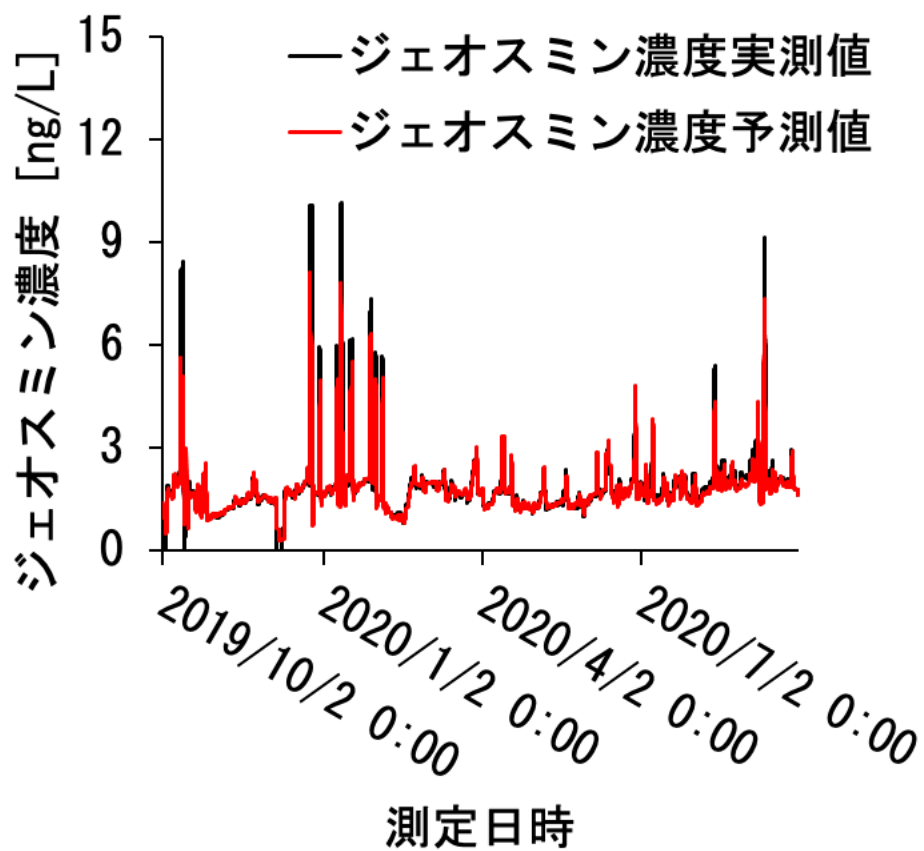
ブロック [時間]	何時間後の 予測か [時間]	1回目		2回目		3回目		4回目		5回目	
		R ²	RMSE	R ²	RMSE	R ²	RMSE	R ²	RMSE	R ²	RMSE
24	3	0.721	0.224	0.054	1.08	0.495	0.261	0.601	1.08	0.068	0.376
24	6	0.569	0.696	0.329	0.869	0.174	0.964	0.177	0.350	0.164	0.320
24	9	0.037	0.550	0.357	0.851	0.478	0.329	0.029	0.594	0.462	0.329
24	12	0.000	0.874	0.001	1.32	0.288	0.419	0.328	0.394	0.120	0.465
48	3	0.593	0.681	0.000	352	0.005	376	0.441	0.824	0.004	3.13 × 10 ⁶
48	6	0.000	16202	0.000	1.20 × 10 ⁵	0.000	44114	0.167	1.23	0.004	293
48	9	0.259	1.01	0.000	12.2	0.002	73.6	0.079	1.63	0.006	1157
48	12	0.005	401	0.210	1.02	0.017	2.08	0.063	2.17	0.120	1.24

※表中の赤字は5回の学習における最も高い精度を示す

行田浄水場は他の2つの浄水場ほどではないがほぼ同様な傾向

実験結果

～ジェオスミン濃度の3時間後予測（行田浄水場）～



実験結果 ～3時間後の濃度予測（まとめ）～

2-MIB

データ地点	R ²	RMSE	学習データ量
吉見浄水場	: 0.993	0.192	(30ヶ月)
大久保浄水場	: 0.897	0.364	(50ヶ月)

ジェオスミン

データ地点	R ²	RMSE	学習データ量
行田浄水場	: 0.721	0.224	(12ヶ月)

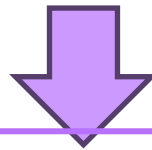
- かび臭物質を含む8項目を入力データとしてLSTMによるかび臭の予測モデルを構築した。その結果、
 - 少なくとも1年間のデータで学習することで3時間後のかび臭濃度を予測することができた。
 - 埼玉県企業局が有する3浄水場ではかび臭予測にLSTMが適用できた。



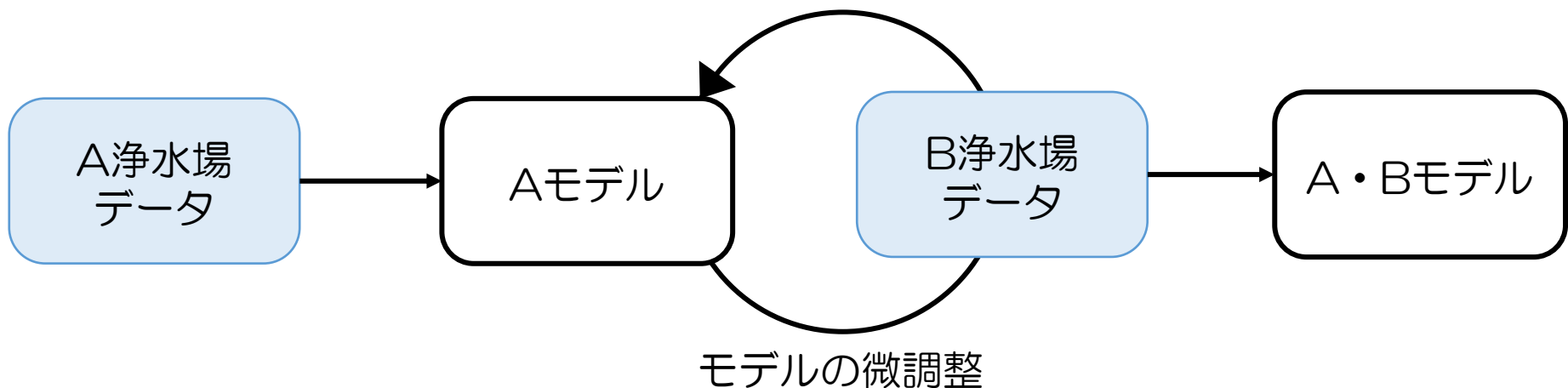
モデルの汎用性 ～モデルの他浄水場への適用～

背景 ～モデルの他浄水場への適用～

それぞれの浄水場のデータを使用することでかび臭の予測モデルが構築できた。



構築したA浄水場のモデルに別のB浄水場のデータを入力データとし、A・Bモデルを作成することはできないか？（下図）

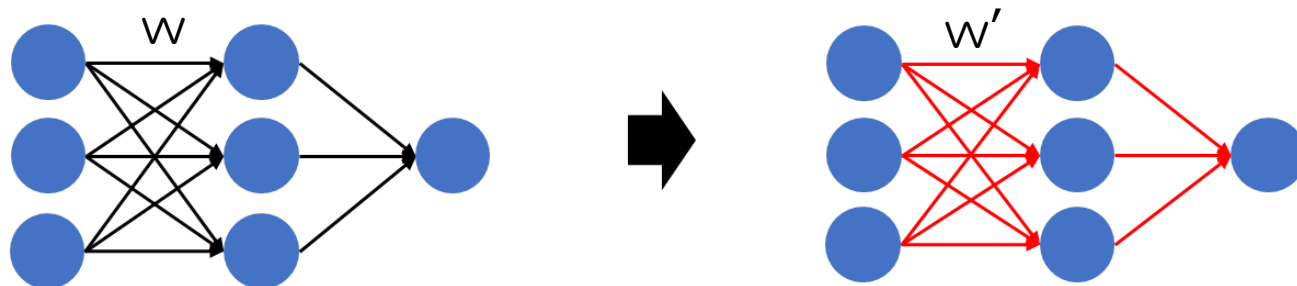


実験方法 ～汎用性の可能性～

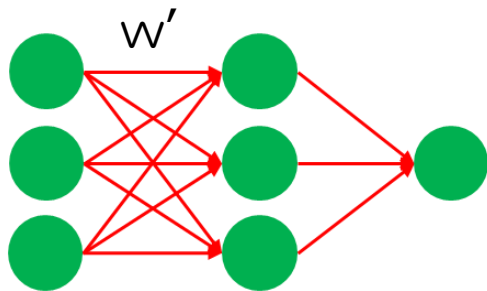
他浄水場の重みを別の浄水場に適用

▶ 重みを再利用する（ファインチューニングと呼ぶ）

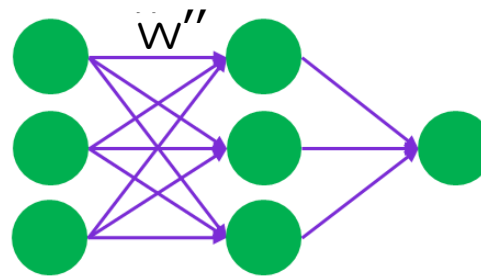
- ① 学習前の重み(w)は乱数 ② 学習することで重み(w)は最適化された w' となる



- ③ A浄水場で最適化された重みを B浄水場に適用



- ④ B浄水場のデータで学習させ、再度重みを最適化する (w'')

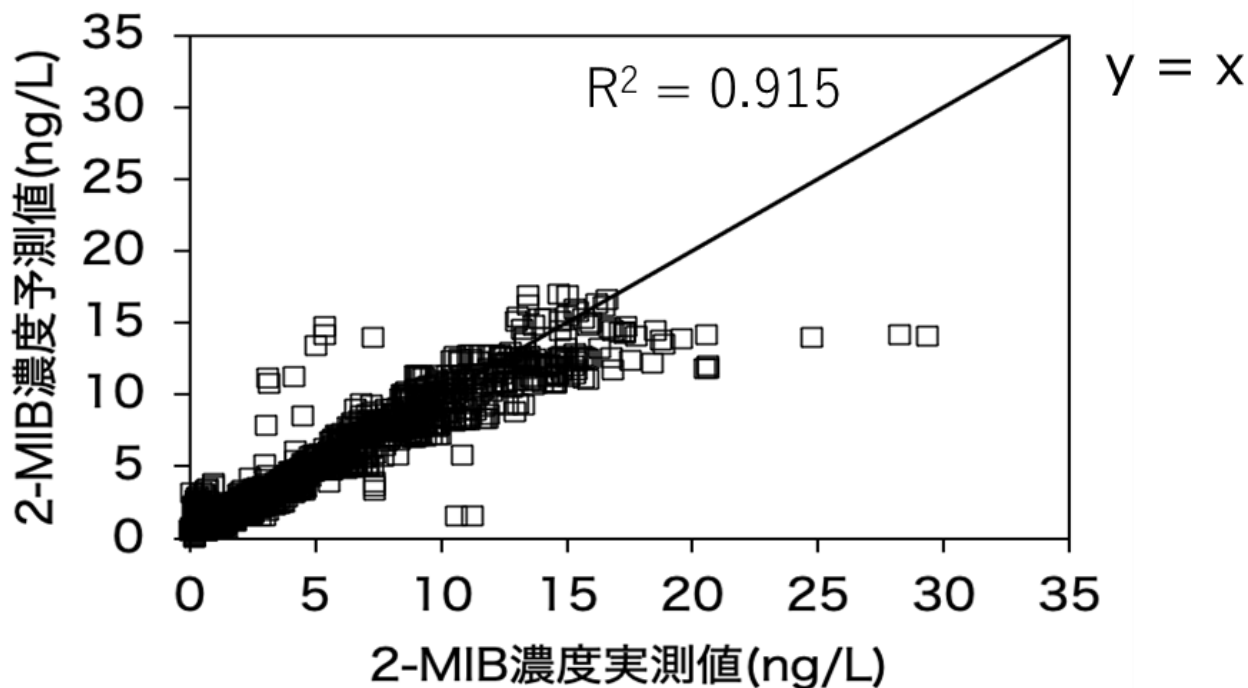


● A浄水場データ

● B浄水場データ

実験結果

～吉見浄水場データに大久保浄水場の重みを再利用～



- 同じ水系である吉見浄水場のデータに大久保浄水場のモデルの重みを再利用し学習することで予測は可能であった。
- 異なる水系の浄水場においても同様の検討を行ったが、吉見浄水場の重みを再利用することはできず、異なる水系には重みの再利用が困難であった。



他浄水場におけるモデルの構築事例

湖沼系A浄水場

河川系B浄水場

湖沼系C浄水場

湖沼系A浄水場 ～入出力データ～

時刻：t-Lからtまで

- 原水濁度
- 沈砂池pH
- 沈砂池アルカリ度(データなし)
- 沈砂池導電率
- 沈砂池塩素要求量
- 沈砂池水温
- 水位
- 原水2-MIB濃度

時刻：t + S

- 原水2-MIB濃度

吉見浄水場では8項目の入力データであったが、A浄水場はアルカリ度を計測していなかったため、入力データは7項目とした

湖沼系A浄水場 ～データ収集間隔と下限値～

- A浄水場と吉見浄水場ではデータ収集間隔が異なる

吉見浄水場での項目(1時間間隔)

- 原水濁度
- 沈砂池pH
- 沈砂池アルカリ度
- 沈砂池導電率
- 沈砂池塩素要求量
- 沈砂池水温
- 河川水位(国土交通省より取得)
- 原水2-MIB濃度

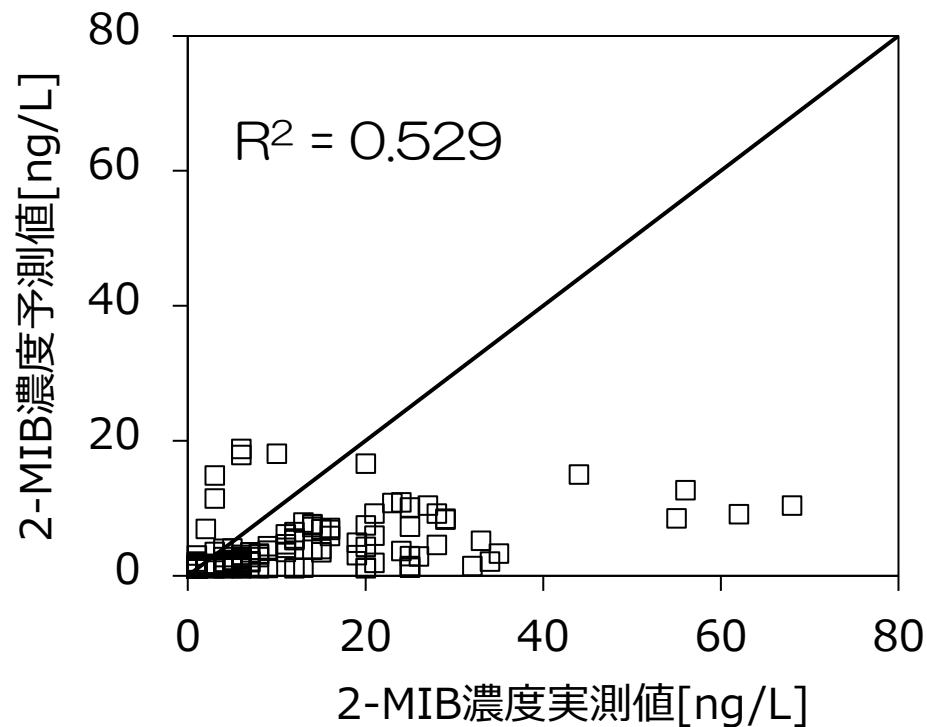
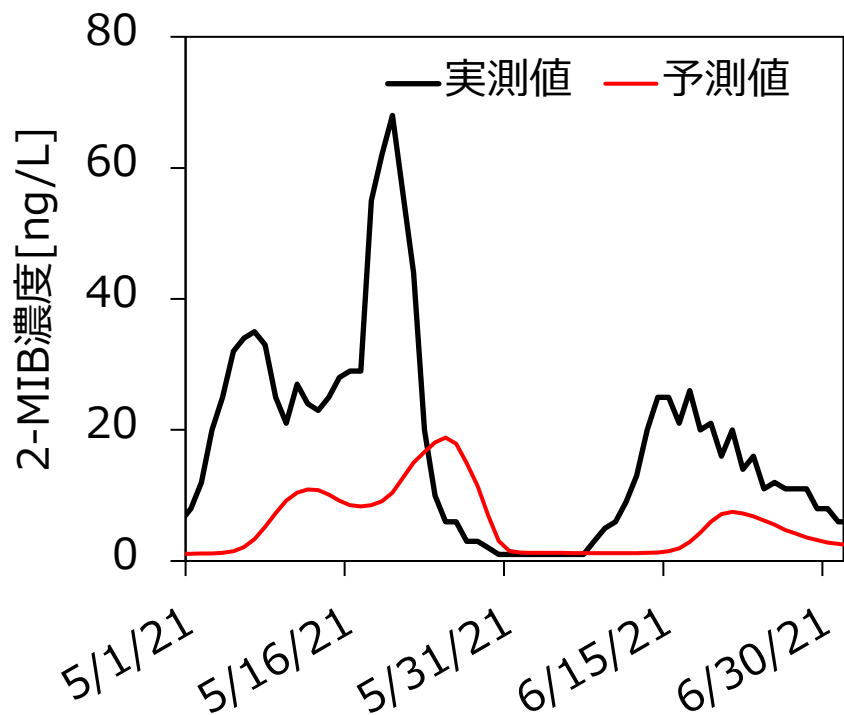
A浄水場での項目(1日間隔)

- ➔ 原水濁度
- ➔ 原水pH
- ➔ 原水アルカリ度
- ➔ 原水導電率
- ➔ 原水アンモニア態窒素
- ➔ 原水水温
- ➔ 水位
- ➔ 原水2-MIB濃度

- かび臭濃度の下限値が2ng/Lであり、<2と表記しているため、今回は1として変換処理

湖沼系A浄水場 ～2-MIB濃度の予測結果～

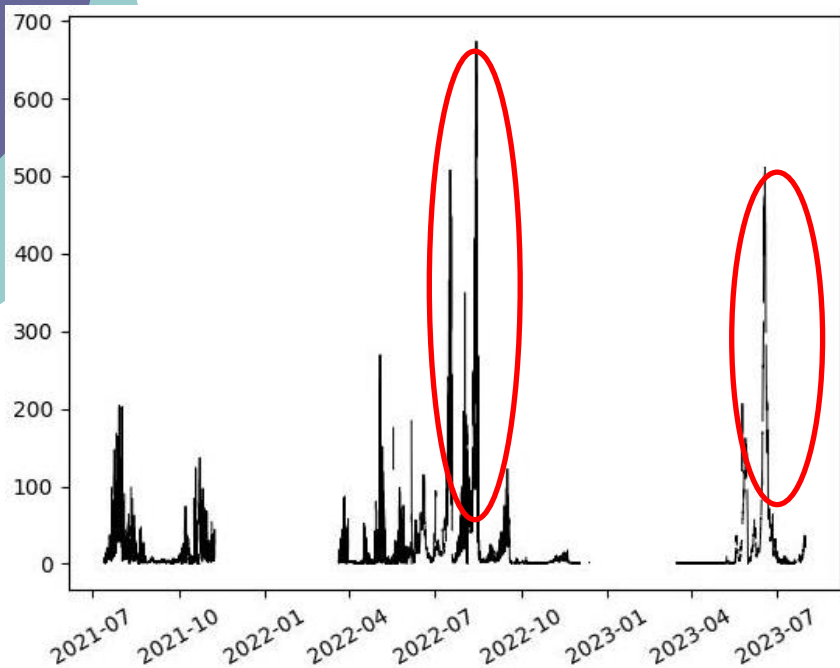
ブロック：7日、1日後の予測



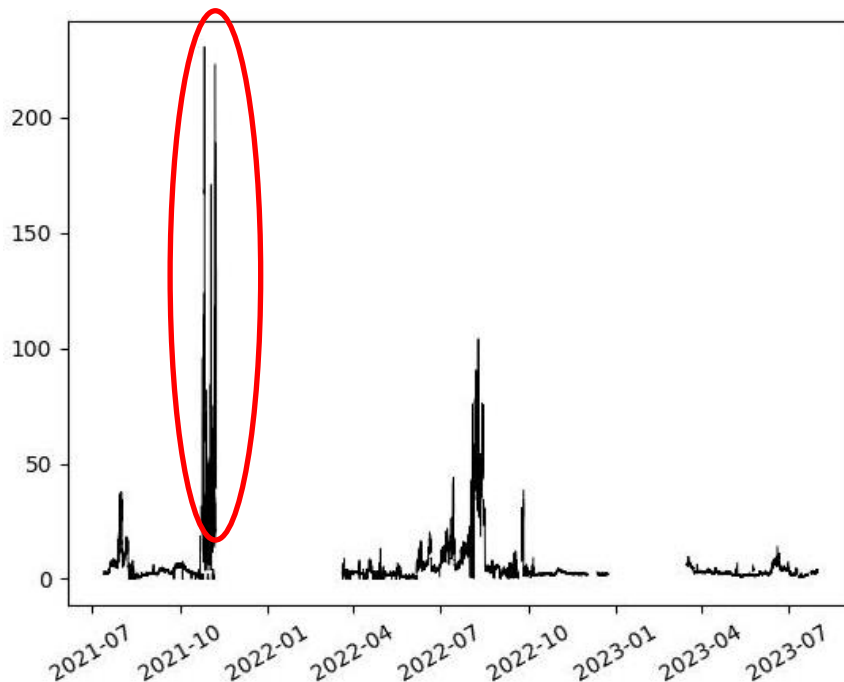
データ収集間隔が1日であったため、1日後の予測を行ったが予測精度は低かった

河川系B浄水場 ～原水のかび臭濃度の経日変化～

2-MIB



ジェオスミン



どちらのかび臭物質とも100ng/L以上の検出頻度が高く、200ng/L以上検出されることもあった（赤丸）。

河川系B浄水場 ～入出力データ～

時刻：t-Lからtまで

- 原水濁度
- 沈砂池pH
- 沈砂池アルカリ度(データなし)
- 沈砂池導電率
- 沈砂池塩素要求量(データなし)
- 沈砂池水温
- 水位
- 原水の2-MIB濃度orジェオスミン濃度

時刻：t + S

- 原水の2-MIB濃度orジェオスミン濃度

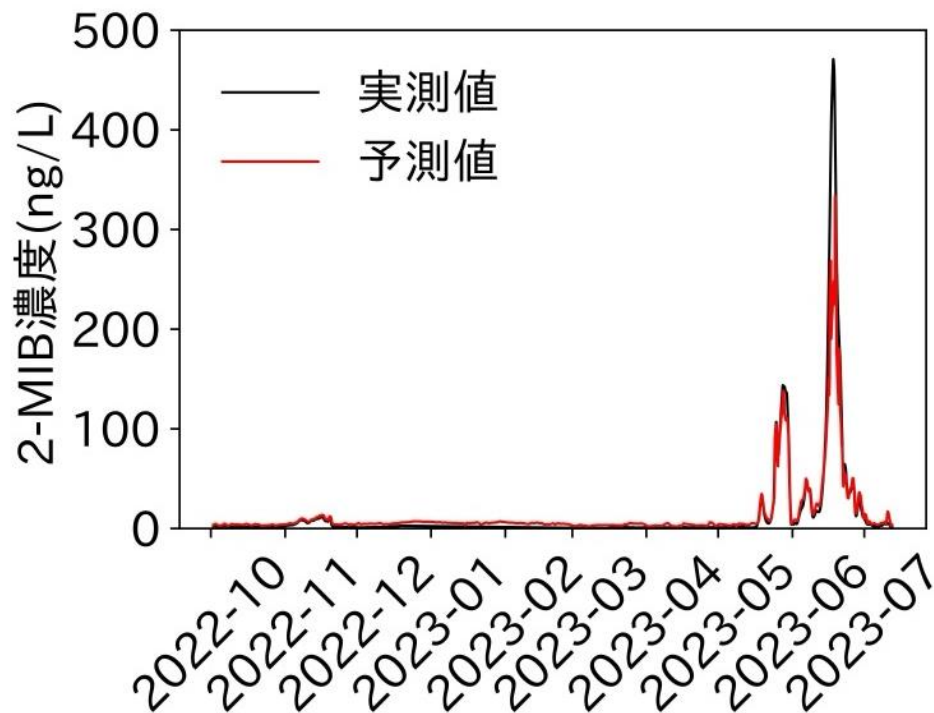
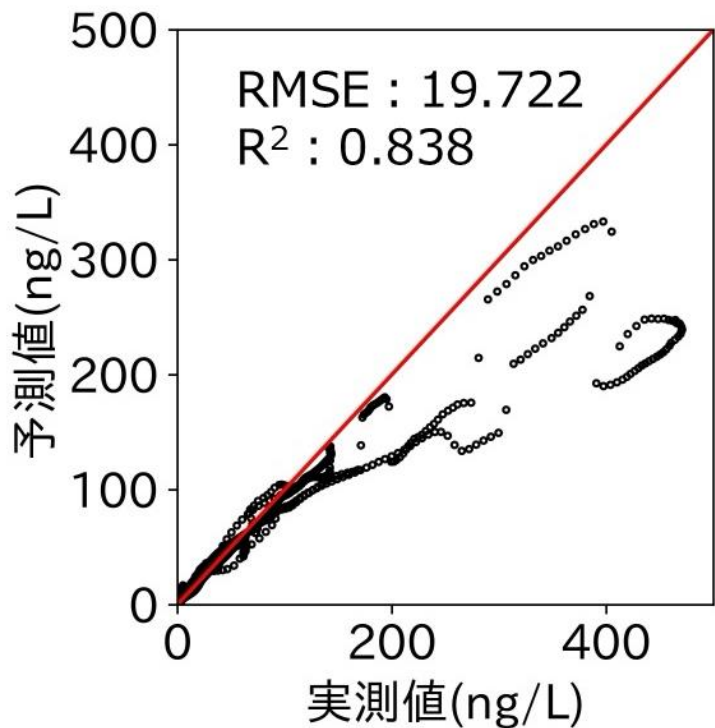
- 吉見浄水場では8項目の入力データであったが、B浄水場はアルカリ度と塩素要求量を計測していなかったため、入力データは6項目とした。
(原水2-MIB濃度と原水ジェオスミン濃度は一緒に入力しない)
- 1時間毎のデータを用いている。

河川系B浄水場 ～データセット～

B浄水場	
学習データ 始	2021/7/1
学習データ 終	2022/6/30
検証データ 始	2022/7/1
検証データ 終	2022/9/30
テストデータ 始	2022/10/1
テストデータ 終	2023/7/12

河川系B浄水場 ～2-MIB濃度の3時間後の予測～

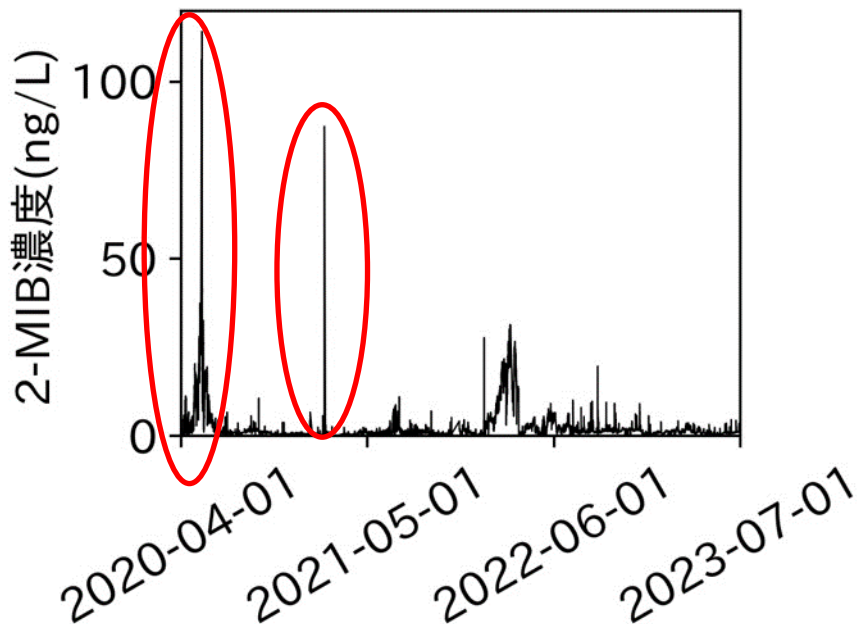
ブロック：24時間、予測時間：3時間後



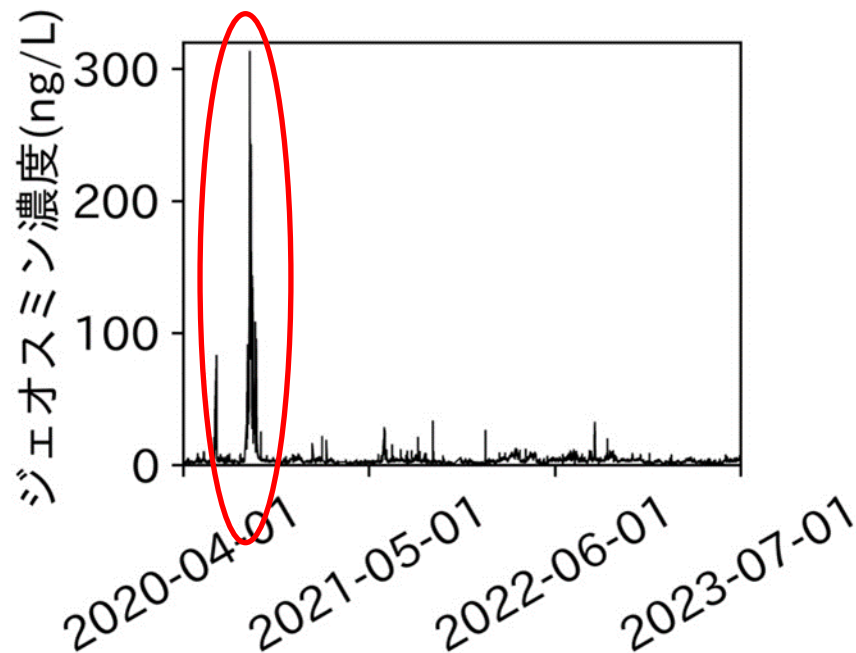
R²値は高く予測精度が比較的高いが、RMSEの値が吉見浄水場に比べて大きい

湖沼系C浄水場 ～原水のかび臭濃度の経日変化～

2-MIB



ジェオスミン



2-MIB濃度が100ng/Lになったり、ジェオスミン濃度が300ng/Lになるなど急激に変化することがあった（赤丸）

湖沼系C浄水場 ～入出力データ～

時刻：t-Lからtまで

- 原水濁度
- 原水pH
- 原水アルカリ度
- 原水導電率（データなし）
- 原水塩素要求量（データなし）
- 原水水温
- 水位（データなし）
- 原水の2-MIB濃度orジェオスミン濃度

時刻：t + S

- 原水の2-MIB濃度orジェオスミン濃度

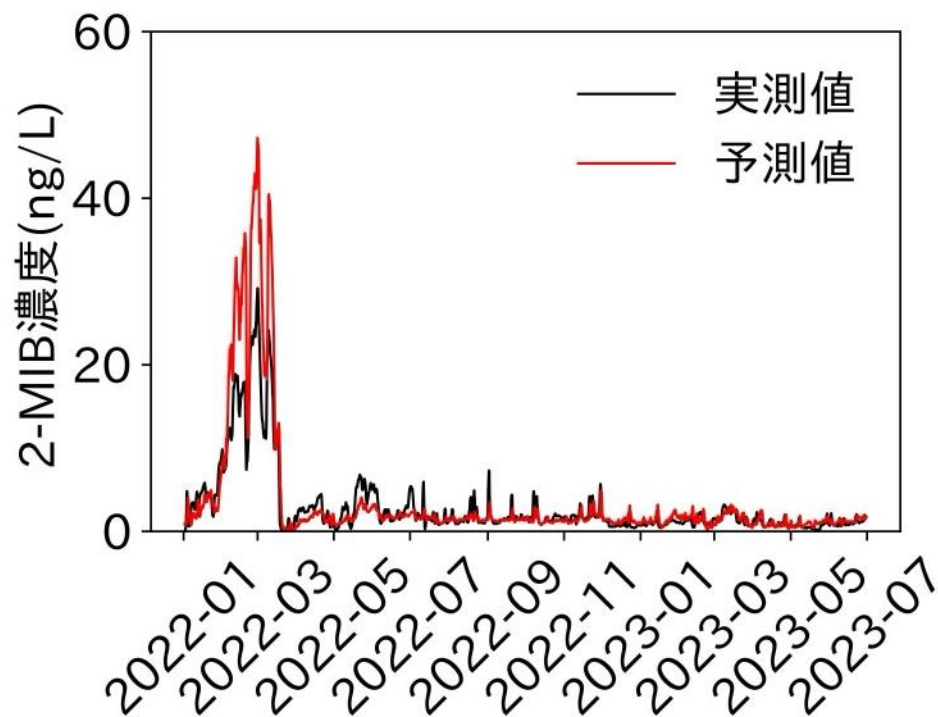
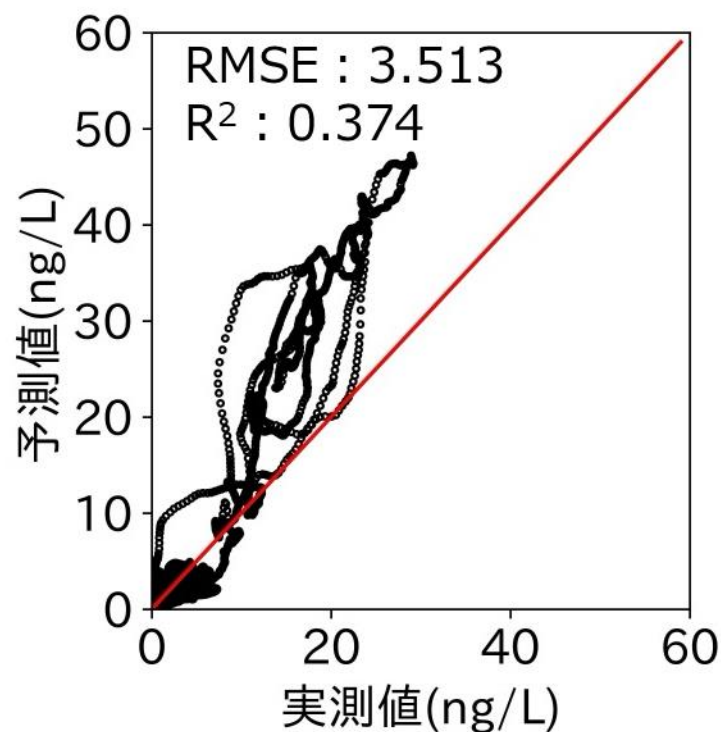
- C浄水場は、導電率、アルカリ度と水位情報もなかったため、入力データは5項目とした。
（原水2-MIB濃度と原水ジェオスミン濃度は一緒に入力しない）
- 1時間毎のデータを用いている。

湖沼系C浄水場 ～データセット～

C浄水場	
学習データ 始	2020/4/1
学習データ 終	2022/9/30
検証データ 始	2022/10/1
検証データ 終	2022/12/30
テストデータ 始	2023/1/1
テストデータ 終	2023/6/30

湖沼系C浄水場 ～2-MIB濃度の3時間後の予測～

ブロック：24時間、予測時間：3時間後



R²値は低く、予測精度は高くなかった



高濃度時の対応とその精度

河川系B浄水場
湖沼系C浄水場

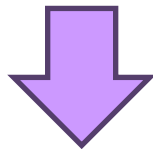
背景 ～高濃度時に予測ができなかった原因～

B浄水場においてもC浄水場においても2-MIB及びジェオスミンの濃度予測の精度は高くなかった



▶かび臭濃度を予測できなかった原因

- LSTMモデルが過学習していた
- 平滑化によるノイズ除去が完全にできていない



▶対策

• バッチ正規化

過学習とノイズの影響を防止するため、入力データの変動が小さい時と大きい時の学習を安定化させる方法

実験方法 ～バッチ正規化～

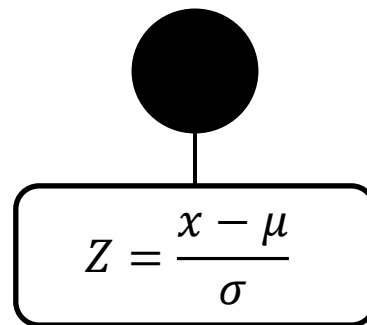
▶バッチ正規化

- バッチ毎に平均を0、分散を1にする

バッチ正規化 $Z = \frac{x - \mu}{\sigma}$

X: 変数の値
 μ : 変数の平均
 σ : 標準偏差

出力層

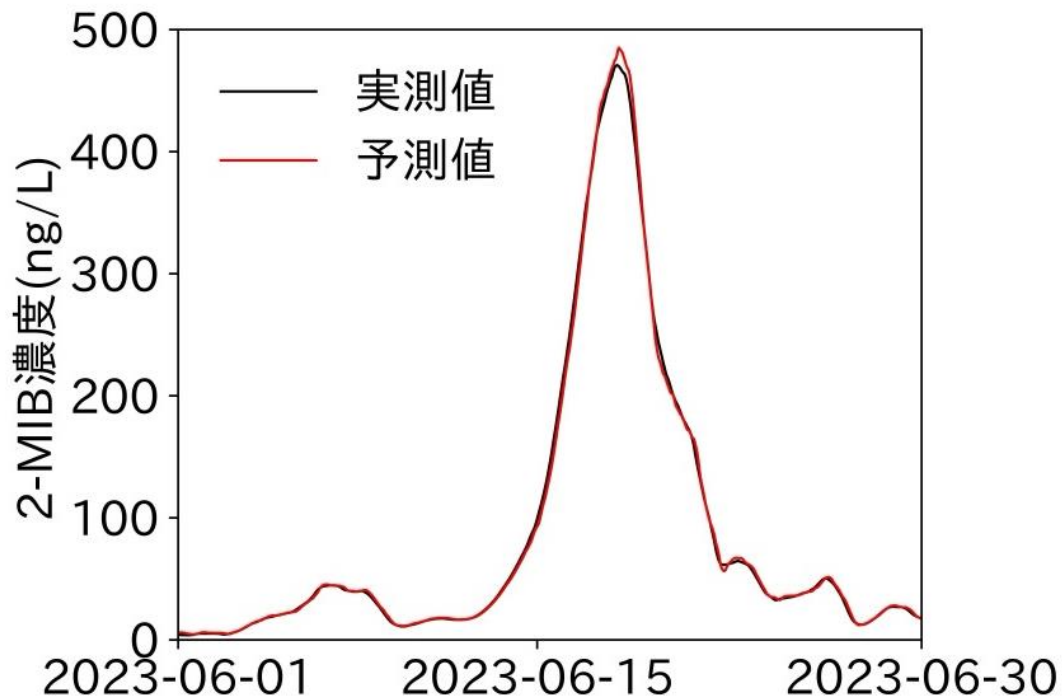
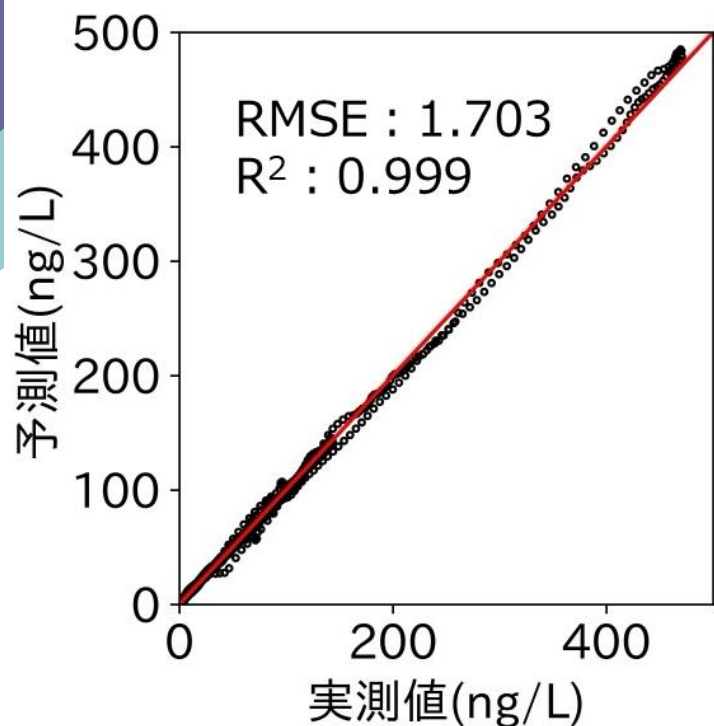


隠れ層

入力層

実験結果 ～バッチ正規化(2-MIB(B浄水場))～

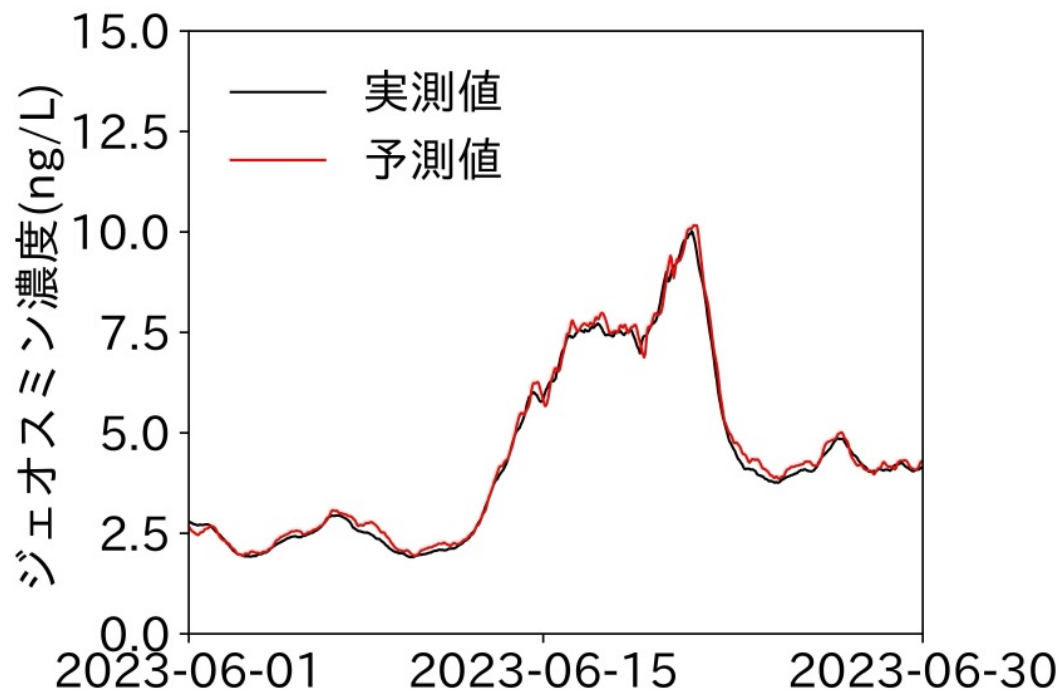
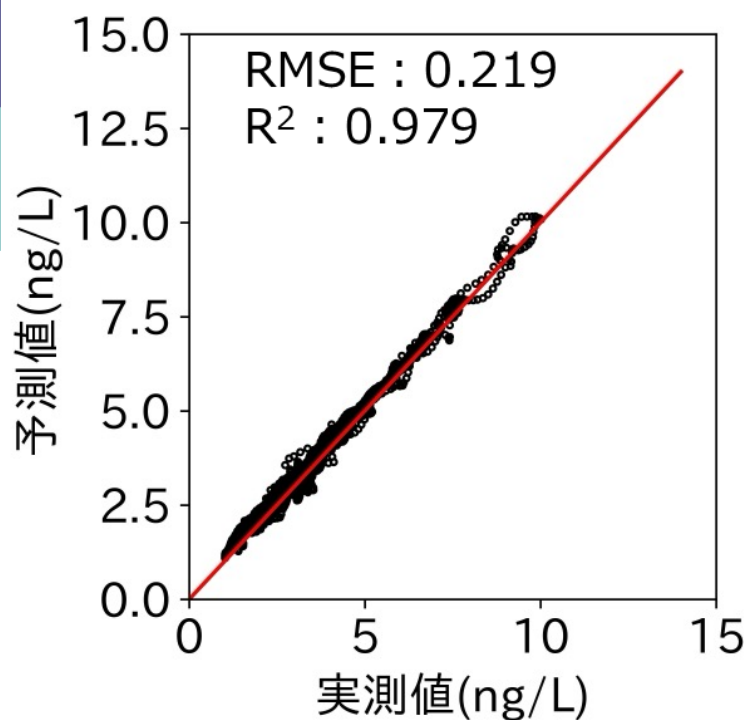
ブロック：24時間、予測時間：3時間後



予測と実測の差が小さくなり予測精度が向上し、ほぼ正確に2-MIB濃度を予測できた

実験結果 ～バッチ正規化(ジェオスミン(B浄水場))～

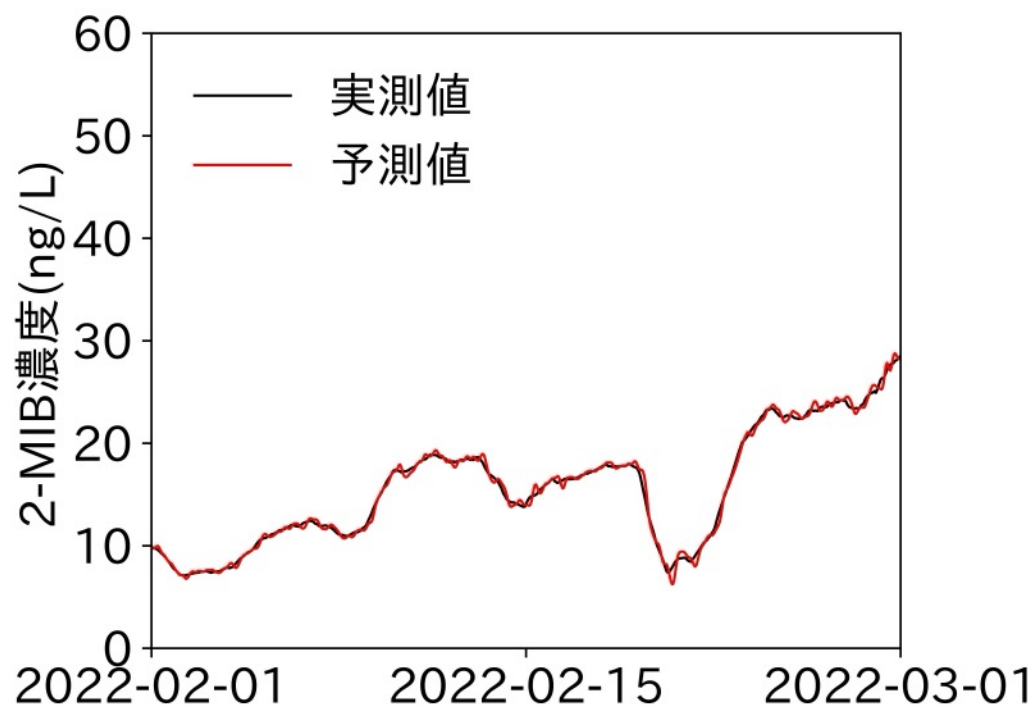
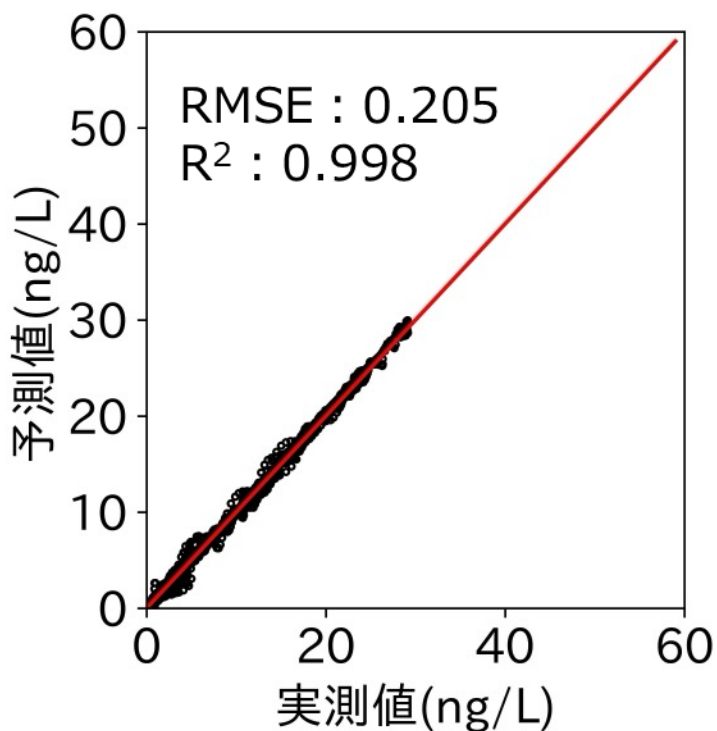
ブロック：24時間、予測時間：3時間後



ジェオスミンにおいても予測と実測の差が小さくなり予測精度が向上し、ほぼ正確に濃度を予測できた

実験結果 ～バッチ正規化(2-MIB(C浄水場))～

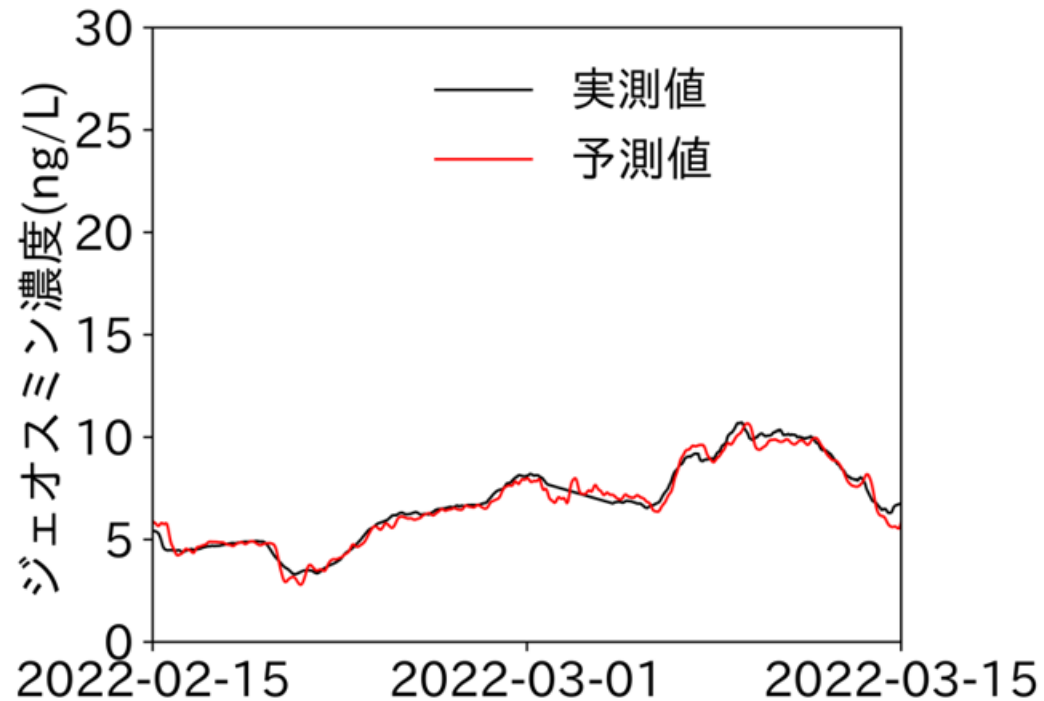
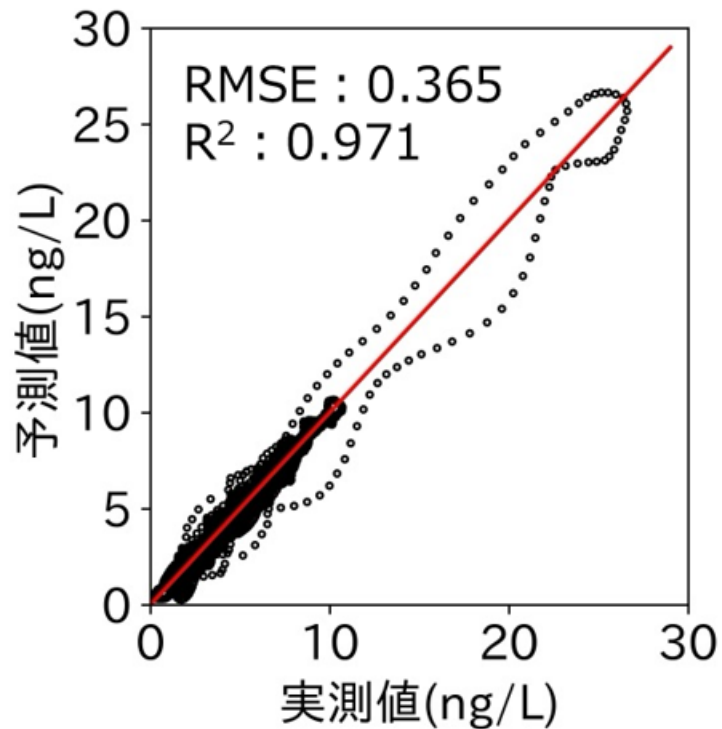
ブロック：24時間、予測時間：3時間後



C浄水場においても予測値と実測値の誤差が小さくなり予測精度が向上し、ほぼ正確に2-MIB濃度を予測できた

実験結果 ～バッチ正規化(ジェオスミン(C浄水場))～

ブロック：24時間、予測時間：3時間後



ジェオスミンにおいても予測値と実測値の誤差が小さくなり予測精度が向上し、ほぼ正確に予測できていた



モニター表示部の数値認識による かび臭予測モデルの検証

実験背景



写真：埼玉県で使用している
水質等を管理している
モニター

実際に構築した予測モデルを運用する場合、

- すべての入力データがPCに直接伝送され
るとは限らない
- 入力されるデータ（数値）が監視室等の
モニターに映し出されることもある



- モニター上に映し出された数値をカメ
ラで読むことができないか？

実験背景 ～データ取得と予測フロー（イメージ）～



①モニター値の数値をカメラにより撮影

吉見浄水場

原水	濁度	18 度	データロガーデータ表示部
2MIB	2.1	ng/L	
ジェオスミン	1.4	ng/L	
沈砂池			
導電率		186	μS/cm

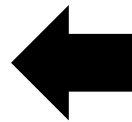
次亜塩素素注入率

1系前次亜	0.50	mg/L
2系前次亜	0.50	mg/L
中次亜	1.60	mg/L
後次亜	0.32	mg/L

②数値の画像処理



3時間後の2-MIBは
3.0 ng/Lです



時刻	濁度	2MIB	..
..
xx:yy:zz	18	2.1	
..

③Excel等のデータに変換

④LSTMモデルによる予測

実験方法 ～数値認識フロー～


1. 撮影画像 (モニター)



画像には

- ・ リアルタイムのモニター (写真右)
- ・ データロガーのモニター (写真左) の数値が表示されている

2. カメラによる画像の撮影



原水	2	個
濁度	1.5	mg/L
2H1B	1.0	mg/L
ジエオスミン	1.0	mg/L
残酸素	2.28	mg/cm
次亜塩素素注入率		
1系前次室	1.00	mg/L
2系前次室	1.00	mg/L
中次室	0.71	mg/L
後次室	0.40	mg/L
毎戸最少送水流量	4.00	m ³ /h
毎戸最少送水圧力	0.90	MPa
送水残塩	0.73	mg/L
送水濁度	0.00	個
送水pH	7.11	



3. 画像処理(二値化)

4. OCR等を使用したモニター数値の判定

実験方法 ～モニター値の留意点（桁数の影響）～

リアルタイムモニターにより取得したデータ

モデル構築のために使用したデータ
(浄水場からの提供データ)

測定日時	濁度	導電率	2-MIB
2022/12/15 11:32	3	228	1.5
2022/12/15 11:37	3	228	1.5
2022/12/15 11:42	3	228	1.5
2022/12/15 11:47	3	228	1.6
2022/12/15 11:52	3	228	1.6
2022/12/15 11:57	3	228	1.6
2022/12/15 12:02	3	228	1.6
2022/12/15 12:07	3	228	1.6

測定日時	濁度	導電率	2-MIB
2017/8/31 8:00	6.48000 0496	213.449 9969	1.32000 0052
2017/8/31 9:00	6.28000 021	213.950 0122	1.77999 9971
2017/8/31 10:00	6.38999 9866	214.449 9969	2.13000 0114
2017/8/31 11:00	6.51999 504	216.550 0031	1.76999 9981
2017/8/31 12:00	6.38000 0114	228.050 0031	2.24000 001

モニター上のリアルタイムの数値とモデル構築のために使用したデータの数値の桁数が異なる場合には桁数の影響が危惧される

実験方法 ～桁数揃え（前処理）～

▶生データ(モデルに使用したデータ)

測定日時	濁度	導電率	水温	2-MIB濃度
7/1 0:00	14.9399996	191	19.9759998	2.72000003
7/1 1:00	16.0900002	191.25	20.0079994	1.88
7/1 2:00	16.0900002	191.600006	20.0359993	1.95000005
7/1 3:00	16.1499996	192.75	20.0919991	1.90999997

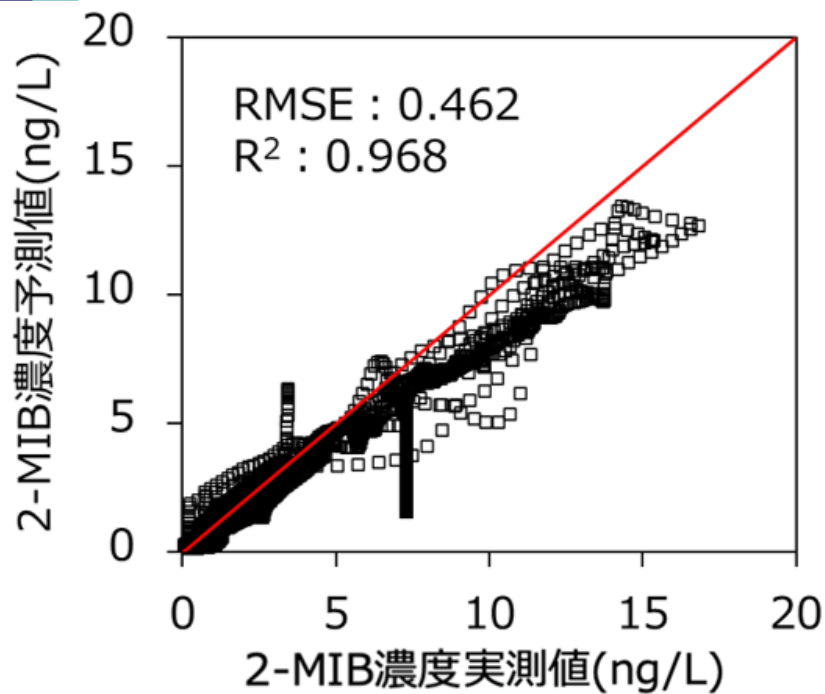


▶前処理後(桁数をモニター値と揃える)

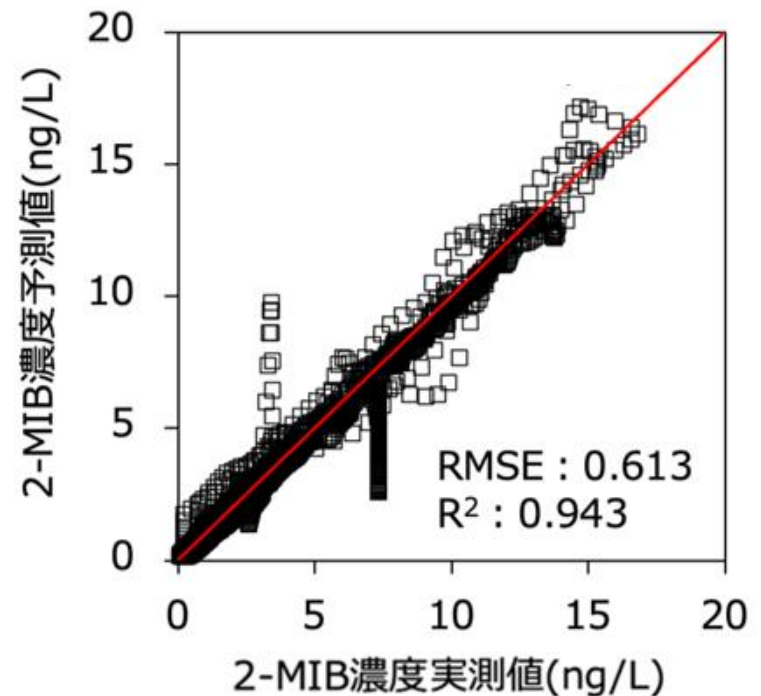
測定日時	濁度	導電率	水温	2-MIB濃度
7/1 0:00	15	191	19.98	2.7
7/1 1:00	16	191	20.01	1.9
7/1 2:00	16	192	20.04	2.0
7/1 3:00	16	193	20.09	1.9

検証結果 ～桁数を揃えた時の精度～

桁数を揃えた場合の精度

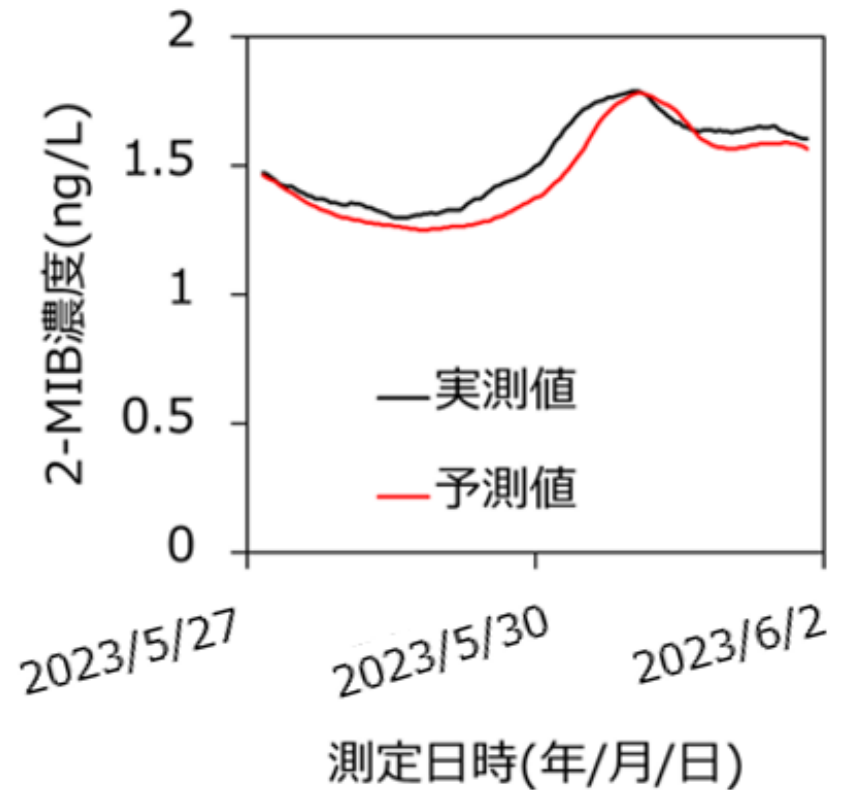
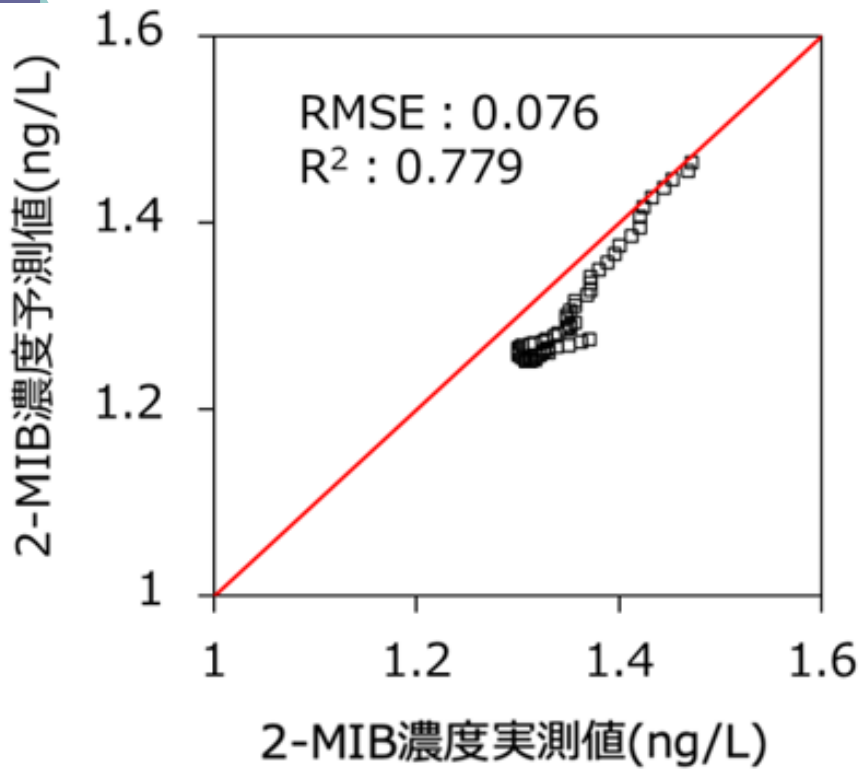


生データを使用したモデルによる精度



桁数を揃えることによって生データ使用時と同等の予測精度であった
⇒モニター値を使用する場合にはモニター値と同じ桁数でモデル構築する
必要がある

検証結果 ～桁数を揃えた時の再検証～



桁数を水総合管理システムと同様にし、外部カメラにて数値を読むことでも予測が可能である

まとめ①～ ビッグデータを活用した原水水質の予測～

- 埼玉県企業局が有する3浄水場では、かび臭物質を含む8項目を入力データとし、LSTMを用いることで3時間後のかび臭濃度を予測することができた。
- 水系が同じ浄水場の場合には、重みの再利用は可能であるが、別の水系では重みの再利用はできず、浄水場が保存しているデータを使用してモデルを構築する。
- A浄水場のようにデータ収集間隔が1日と長い浄水場では1日後の予測精度は低かった。

まとめ②～ ビッグデータを活用した原水水質の予測～

- B浄水場やC浄水場のように高濃度にかび臭が検出される場合にはバッチ正規化を行って学習させることによりかび臭濃度をほぼ正確に予測できた。
- モニター上の値を外部カメラで読み取っても予測できるが、学習時にモニターの数値の桁数と合わせる必要がある。

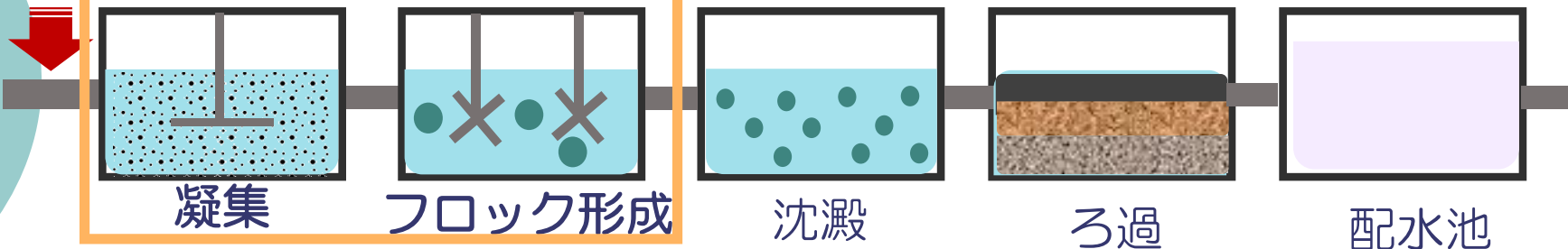


6. 研究結果

画像解析を用いた
凝集判定手法の研究

研究背景

凝集剤



“どのくらい凝集剤を入れればいいのか?”

課題：水道事業の職員減少によって技術継承が困難

演算式・ジャーテスト



- 時間かかる
- 経験者必要



深層学習



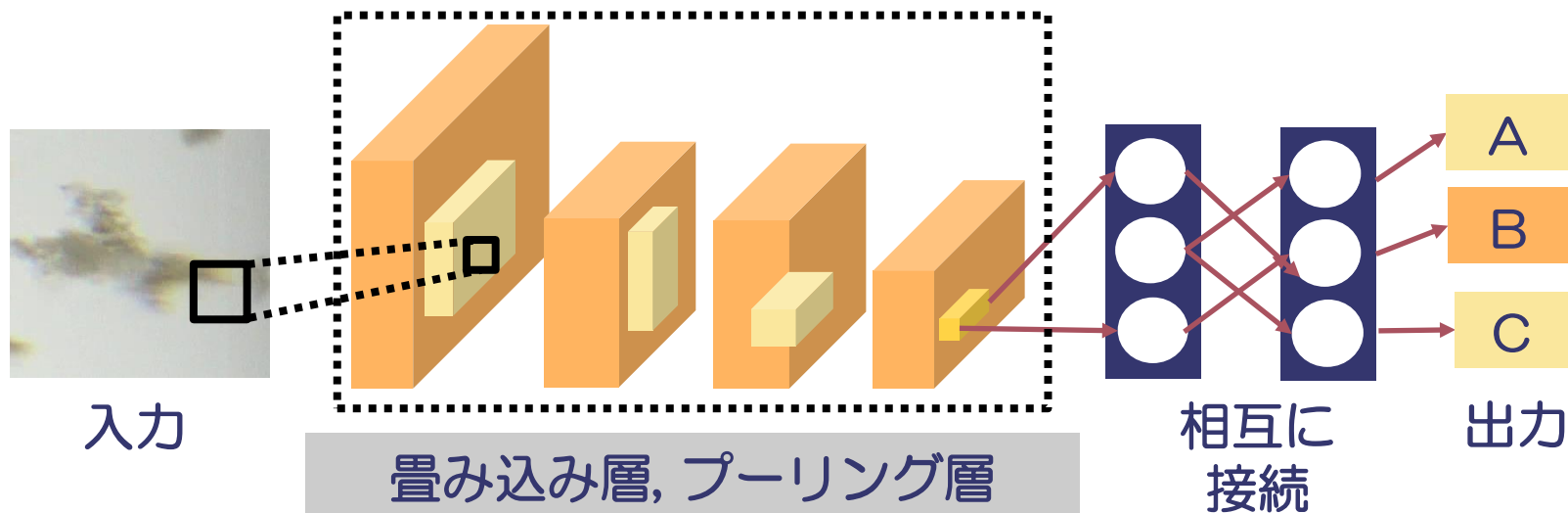
- 迅速
- 初心者でも簡単

凝集処理の運転を支援するシステムの開発を目指す

実験方法 ~CNN~



90%以上の精度で分類できる



畳み込み層とプーリング層によって、画像から重要な「特徴（カーブやエッジなど）」を人の判断を介さずに抽出することができる



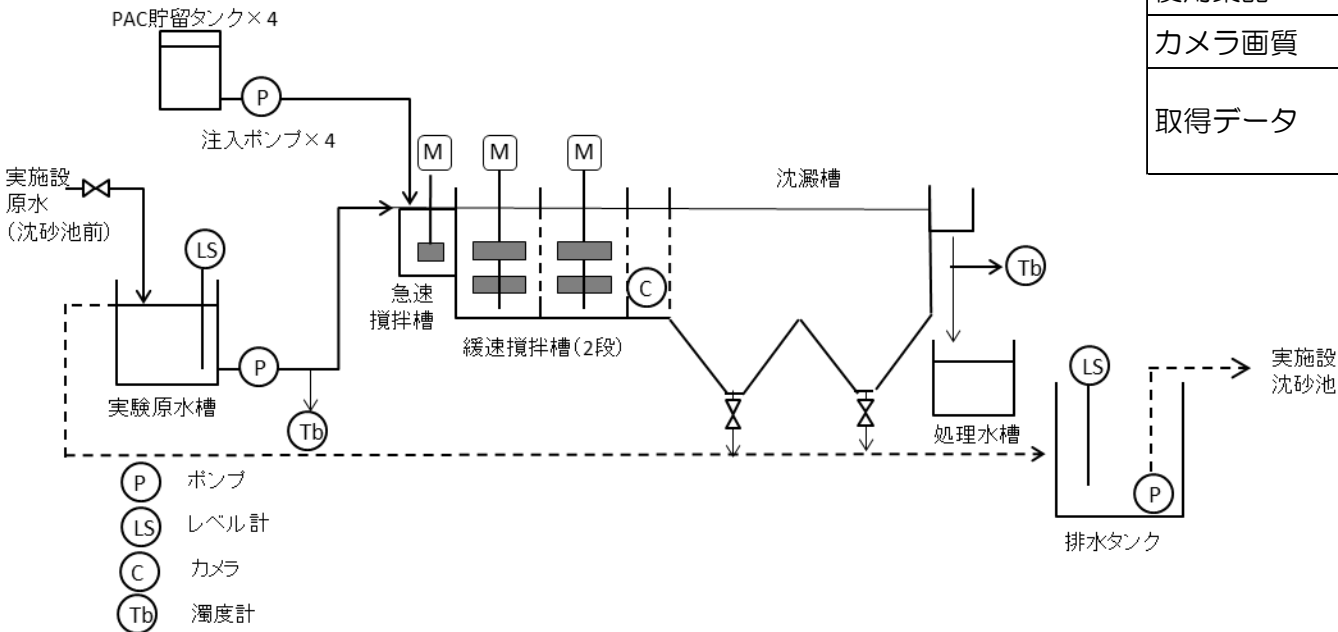
連続通水運転による検証

1) 1年間の学習期間による精度

実験概要 ～設備概略フローと仕様～

- ❑ 実験場所 : 吉見浄水場（埼玉県）
- ❑ 期間 : 2021年9月～2024年1月
- ❑ 仕様等 : 右表参照

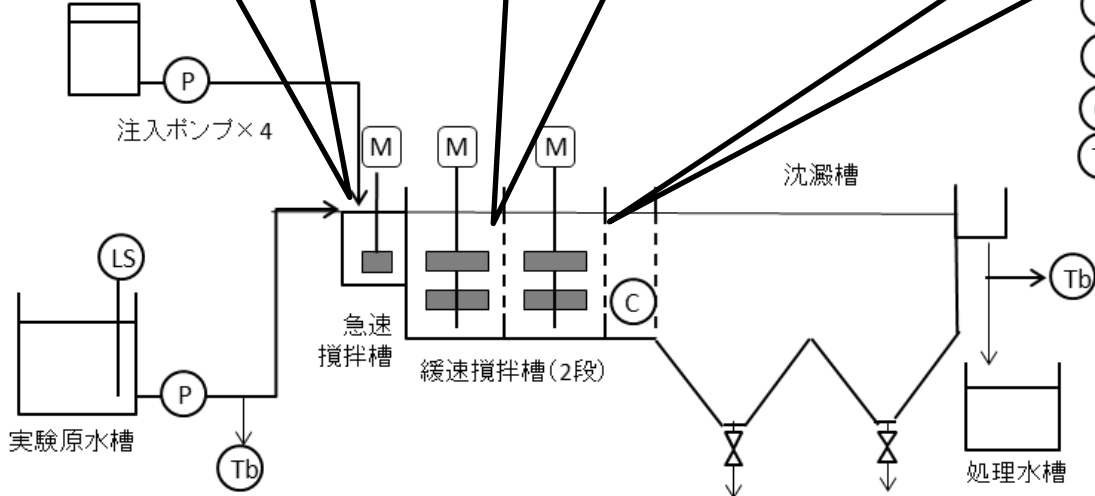
処理水量	15m ³ /日
設置場所	吉見浄水場内
材質	鋼板製
凝集沈澱槽	混和槽 1槽 200rpm
	フロック形成槽 2槽 (20～30rpm)
	中間槽 (フロック形成直後) ⇒水中カメラ設置
	横流式沈澱槽
使用薬品	PAC 4段階で注入
カメラ画質	フルHD (1920×1080ピクセル)
取得データ	フロック画像 (1fps) 原水濁度、沈澱処理水濁度、水温



実験方法 ～実験設備～



PAC貯留タンク×4



- (P) ポンプ
- (LS) レベル計
- (C) カメラ
- (Tb) 濁度計



実験方法 ～凝集剤と注入率～

- 様々なフロックを形成させるために、4つの注入率を12時間ごとに切り替えた

凝集剤

ポリ塩化アルミニウム(PAC)

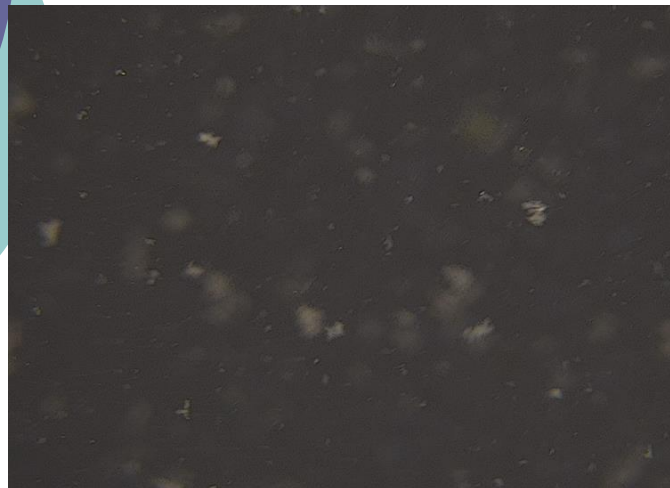
- ・ 銘柄：PAC250A
- ・ Al_2O_3 ：10%
- ・ 比重(20℃)：1.19以上
- ・ 塩基度(%)：47～57

注入率			
1	2	3	4
10 mg/L	20 mg/L	40 mg/L	60 mg/L

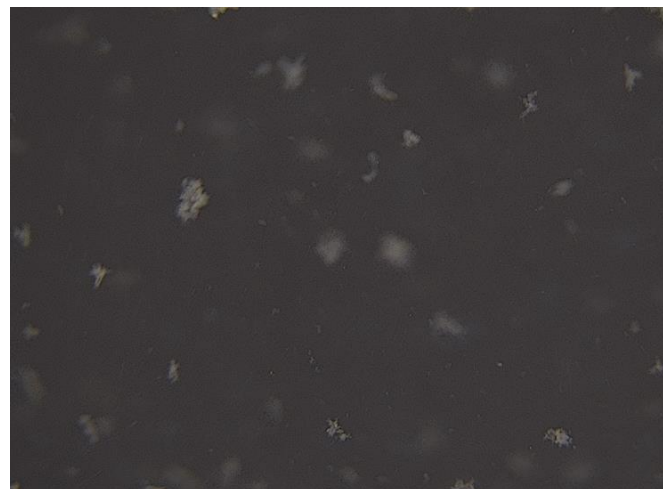
12時間 12時間 12時間 12時間

実験方法 ～画像の例～

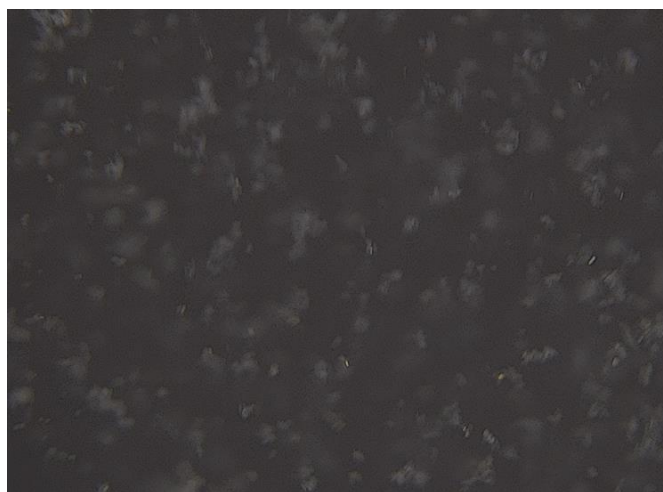
- PAC注入率:10mg/L



- PAC注入率:20mg/L

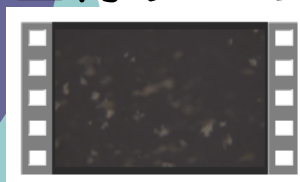


- PAC注入率:60mg/L



実験方法 ～画像の切り出しとデータセット～

● 画像データ



動画

- 10秒に一枚
- サイズ：1920×1080 pixel

python™

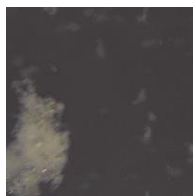


大きい画像（静止画）
800×800 pixel

● データセット

教師データ
(処理水濁度)

<0.5度
Class1



0.5～1度
Class2



>1度
Class3



入力データ
(画像データ)

200×200 pixel にリサイズ

学習データ(1年間)

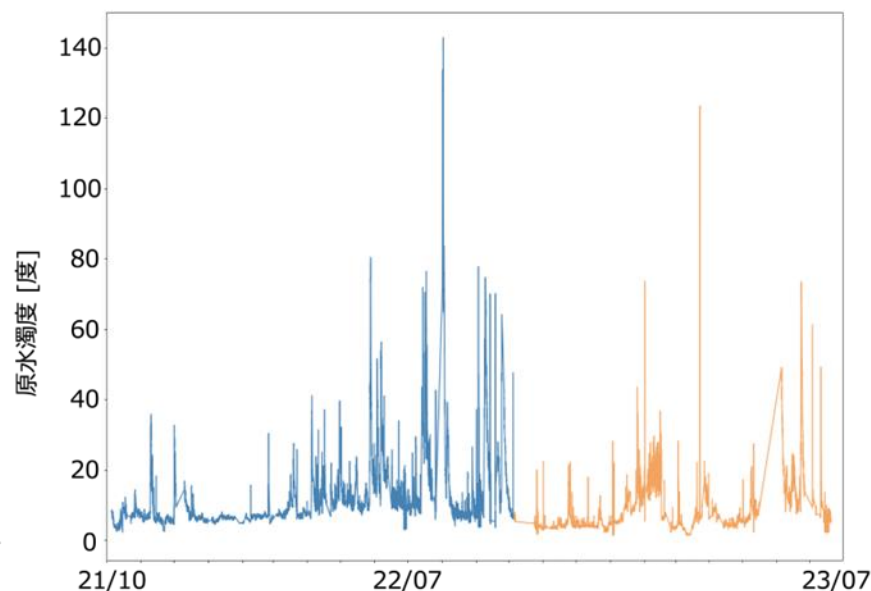
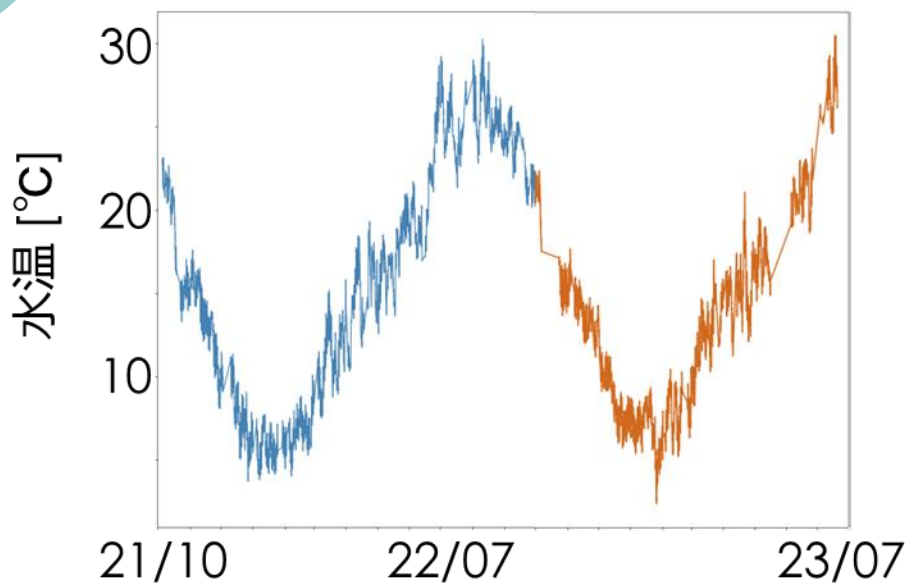
テストデータ

$$\text{テスト精度} = \frac{\text{正解データ数}}{\text{テストデータ数}}$$

実験方法 ～データ数と原水性状～

学習データ：1,562,358 枚
2021年10月5日～2022年10月4日 (1年間)

テストデータ：1,014,860 枚
2022年10月5日～2023年7月20日



実験方法 ～評価方法～

C_{xy} : 正解クラスがxの時、予測クラスがyであるデータ数

A_{xy} : 各正解クラスのうち、予測クラスの割合

正解クラス	3	C_{31}	C_{32}	C_{33}
	2	C_{21}	C_{22}	C_{23}
	1	C_{11}	C_{12}	C_{13}
		1	2	3
		予測クラス		

結果解析用

$$A_{xy} = \frac{C_{xy}}{C_{x1} + C_{x2} + C_{x3}}$$

正解クラス	3	A_{31}	A_{32}	A_{33}
	2	A_{21}	A_{22}	A_{23}
	1	A_{11}	A_{12}	A_{13}
		1	2	3
		予測クラス		

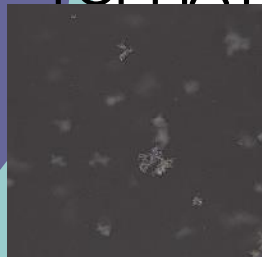
評価指標

$$\text{精度} = \frac{C_{11} + C_{22} + C_{33}}{C_{11} + C_{12} + C_{13} + C_{21} + C_{22} + C_{23} + C_{31} + C_{32} + C_{33}}$$

※以下の**テスト精度**は
全てデータ数で計算したもの

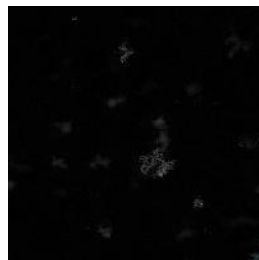
実験方法 ～前処理～

● TOPHAT



元画像

画像を収縮・膨張させ、大きい物体を除去



メリット：大きい白色領域を除去することが可能

デメリット：有効なフロックも削除する可能性がある

● OTSU-差分



元画像

2つの物体の境界となる画素値を自動で算出



メリット：時間変化で同じ位置の画素値があれば除去可能
背景を削除することが可能

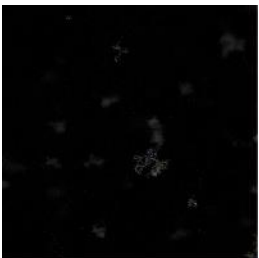
デメリット：フロックの輪郭を削除する可能性がある

● 暗合成-差分



元画像

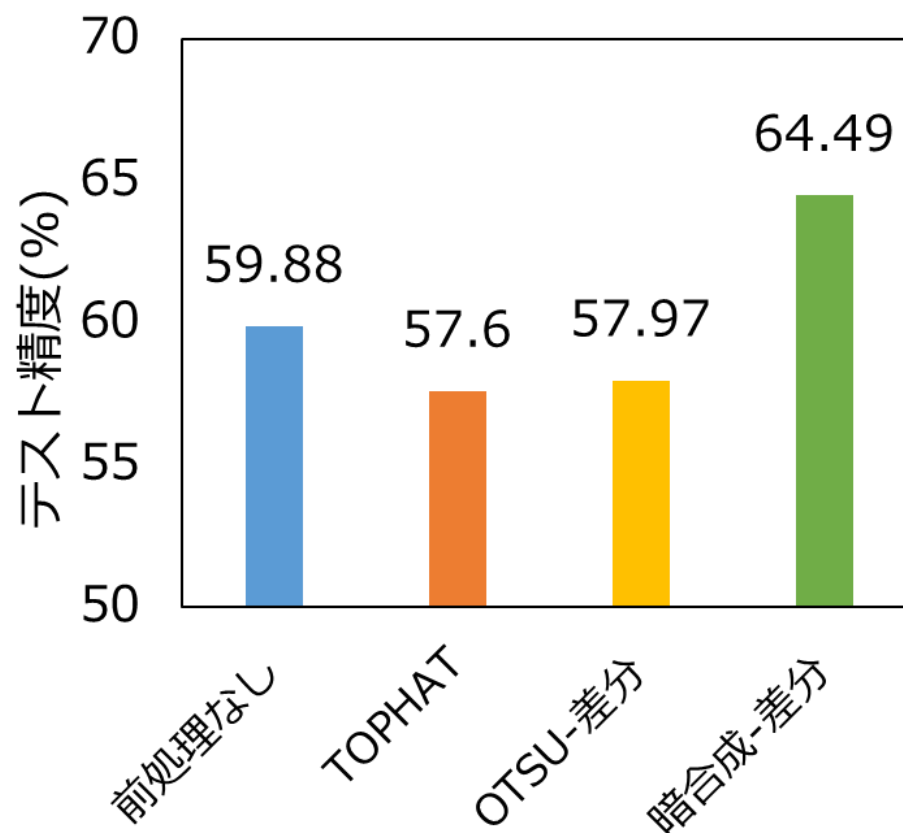
画像の明るさを比較し、画像前後の異なる部分を削除



メリット：時間変化で動かない部分を除去することが可能
汚れ除去と同時にフロックの形を確保

デメリット：フロックから背景の画素値を削除するため色が薄くなる

実験結果 ～前処理比較～



前処理方法として暗合成-差分法を用いることで予測精度が5%程度高まったため、画像の前処理は暗合成-差分法を使用することとした

実験結果 ～1年間のモデルと精度～

1年間の収集画像で構築したモデルで予測精度を求めた



正解クラス	3	6.9 %	12.4 %	80.7 %
	2	43.4 %	39.8 %	16.8 %
	1	86.3 %	12.4 %	1.3 %
		1	2	3
		予測クラス		

テスト精度 68.9 %

1年間の画像によって構築したモデルにおける精度は69%であった。精度が低かったのはクラス2（0.5～1度）であり、特に濁度境界部分が低かった。

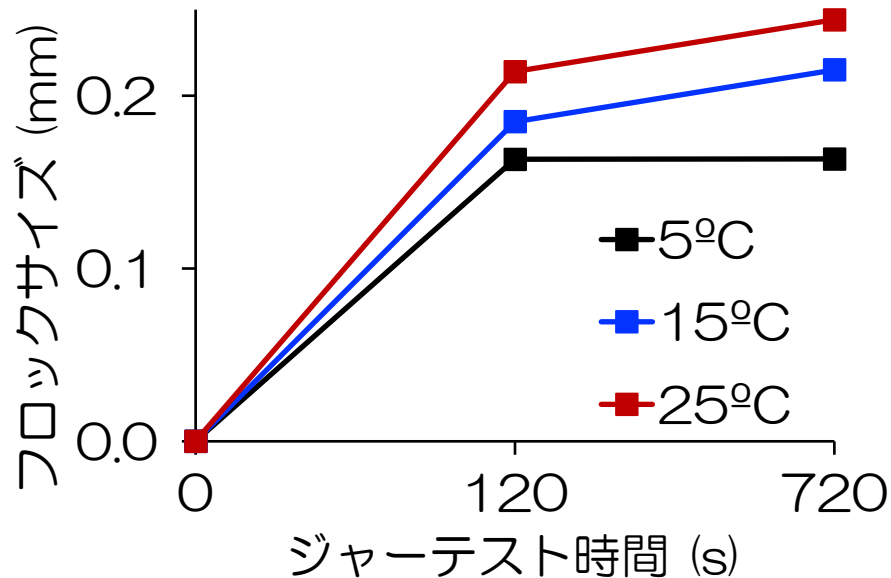


連続通水運転による検証

2) 水温別モデルと精度

実験背景 ～水温別モデルと精度～

ジャーテストの結果とアレニウスの式よりフロック形成は水温に影響される



アレニウスの式

$$\ln(k) = \ln(A) - \frac{E_a}{RT}$$

k = 反応速度定数

A = 頻度因子

E_a = 活性化エネルギー

R = 気体定数

T = 水温

水温別にモデルを構築することとした

実験方法 ～水温別モデルと精度～

水温別モデルのデータ数

▶0-10°Cモデル

訓練データ : 2021-11-25-2023-02-04 (約47万枚)

検証データ : 2021-11-25-2023-02-03 (約7万枚)

テストデータ : 2023-02-04-2023-03-04 (約13万枚)

▶10-20°Cモデル

訓練データ : 2021-10-13-2022-11-13 (約40万枚)

検証データ : 2021-10-13-2022-11-13 (約6万枚)

テストデータ : 2022-11-14-2023-06-24 (約22万枚)

▶20°C以上モデル

訓練データ : 2021-10-05-2022-09-11 (約22万枚)

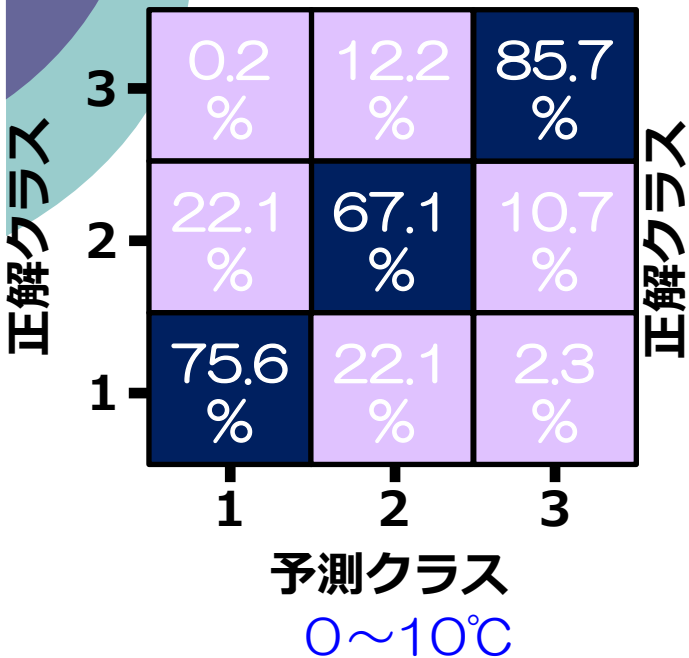
検証データ : 2021-10-05-2022-09-11 (約3万枚)

テストデータ : 2022-09-12-2023-08-31 (約13万枚)

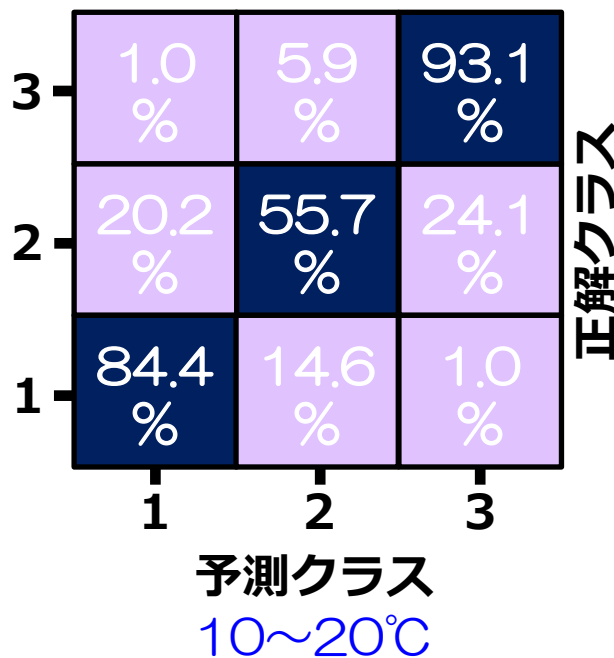
実験結果 ～水温別モデルと精度～

全体のテスト精度68.9%に対して

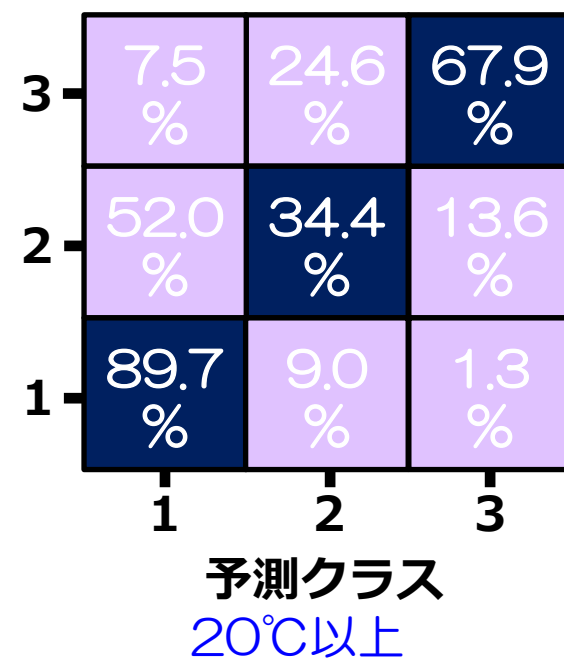
テスト精度 76.2 %



テスト精度 77.7 %



テスト精度 64.0 %



0~10℃、10~20℃の水温モデルではテスト精度が向上したが、20℃以上のモデルでは精度が若干低下した



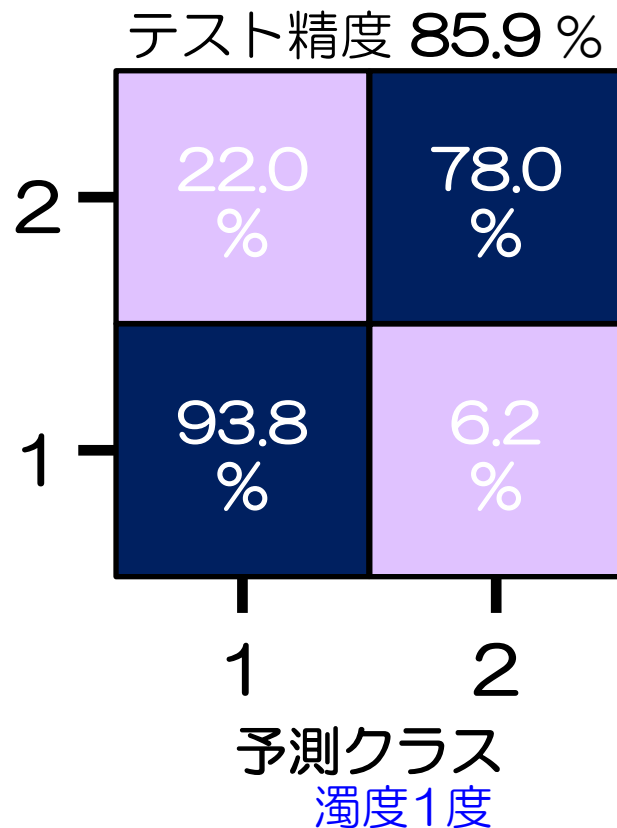
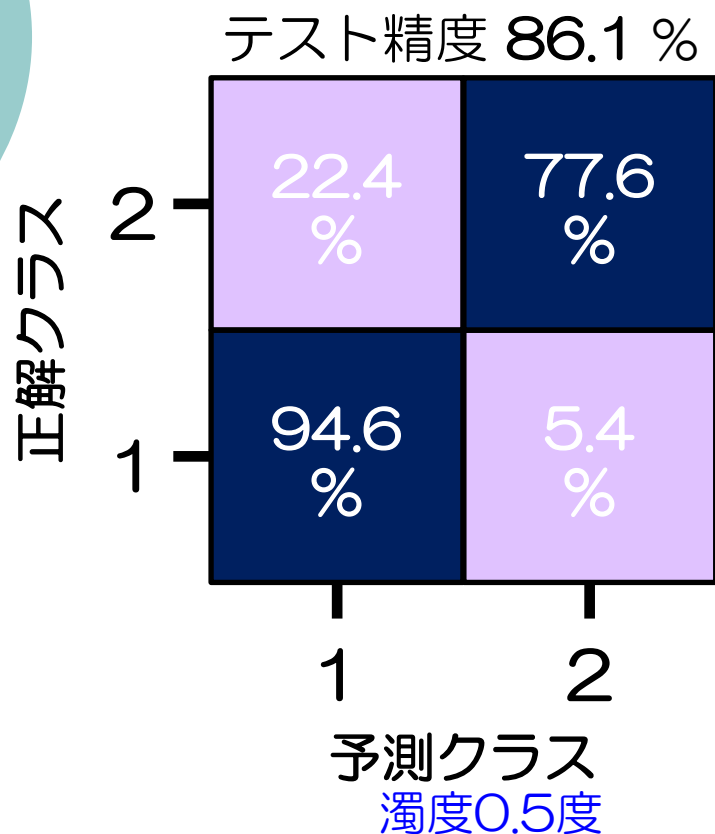
連続通水運転による検証

3) 実用上の運転における精度

(浄水場では沈澱処理水の目標濁度を定めている所も多く、3分類ではなく2分類でまとめた)

実験結果 ～実用上の運転における精度～

3分類ではなく2分類により精度を検証

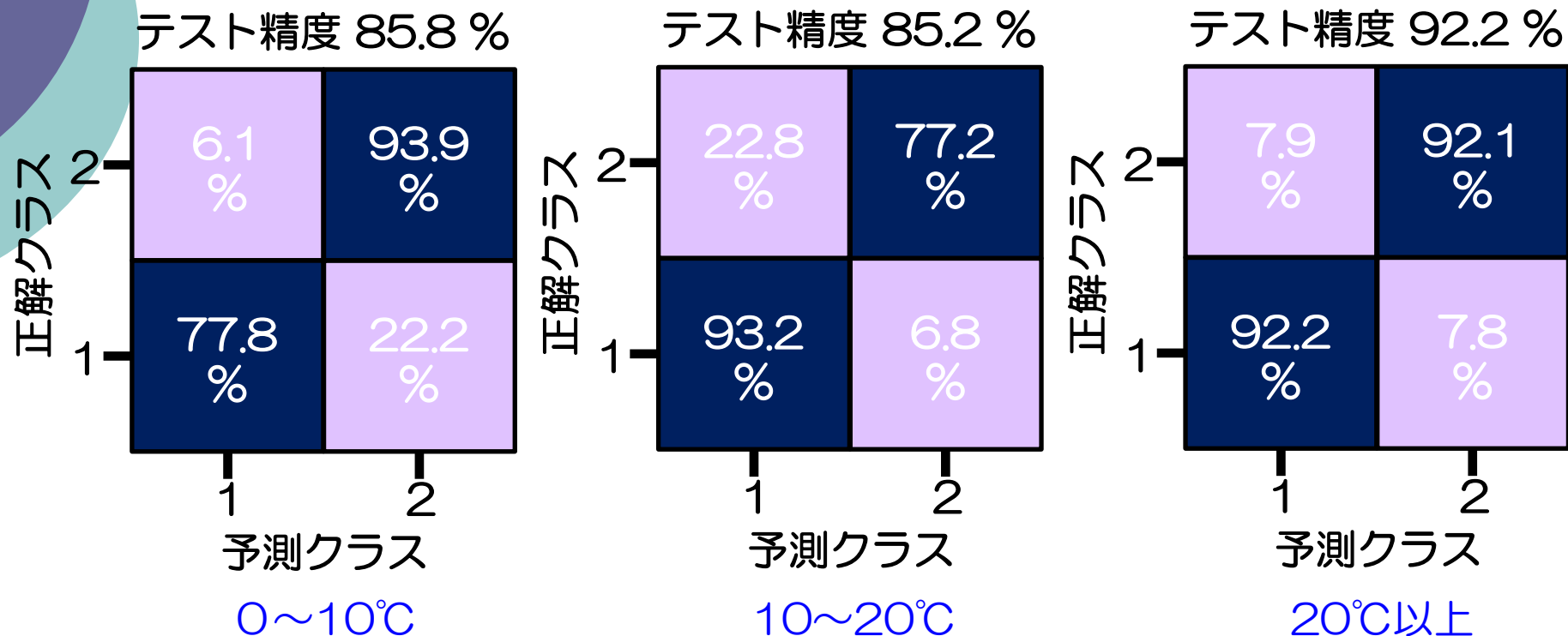


2分類にすることで精度が3分類の15%以上向上し、85%程度となった

実験結果 ～実用上の運転における精度～

水温別モデル（目標濁度1度の場合）

全体のテスト精度85.9%に対して

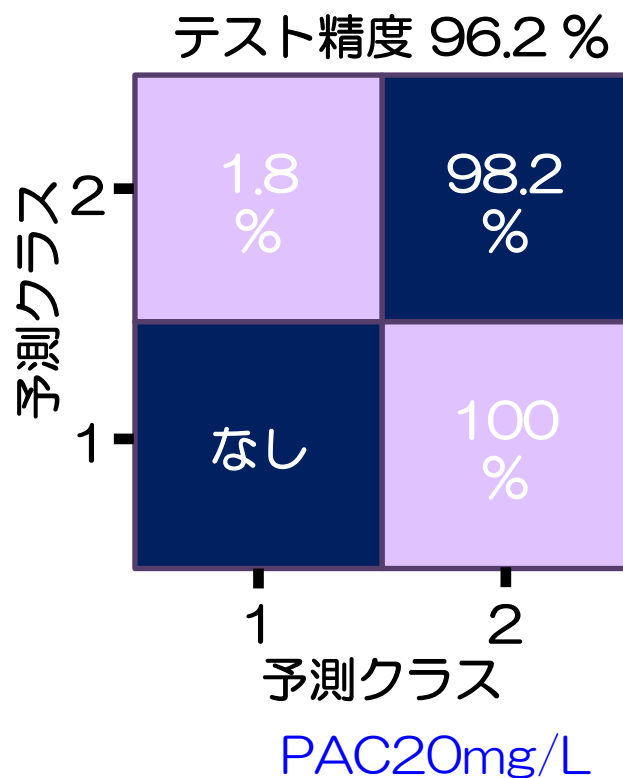
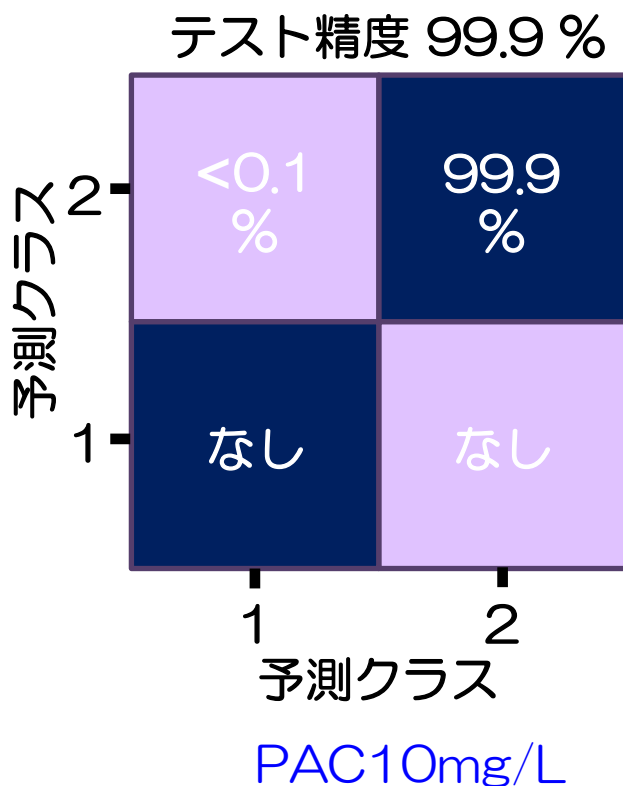


水温別に2分類にするとテスト精度は全体的に向上した

実験結果 ～実用上の運転における精度～

同一注入率による精度（目標濁度1度の場合）

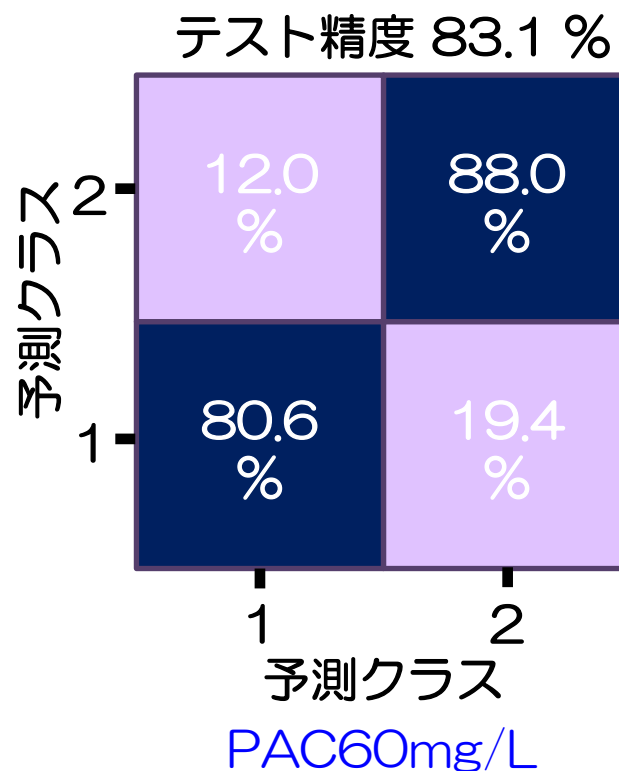
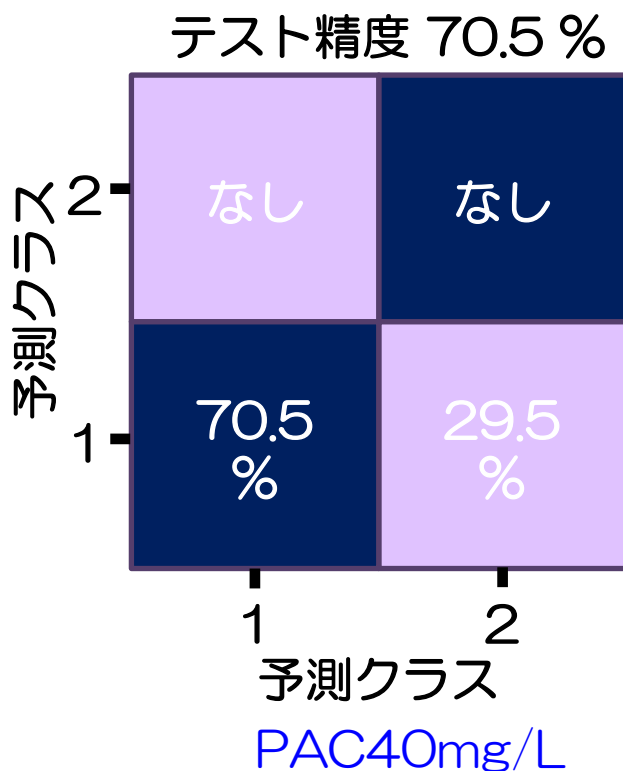
実験条件：およそ一週間一定のPAC注入率にて運転



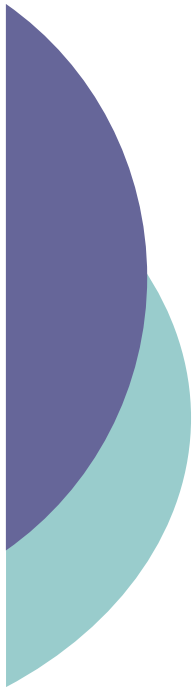
低注入時におけるテスト精度は高かった

実験結果 ～実用上の運転における精度～

同一注入率による精度（目標濁度1度の場合）



2分類においてもPAC40、60mg/Lの場合には精度が若干低下していたが予測はできていた



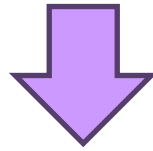
連続通水運転による検証

4) 異なるモデルの構築

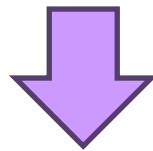
～数値信頼度からの検討～

背景

構築したモデルを浄水場に導入するためには予測精度の确实性と安全性の担保が必要と思われた。



确实性を高めるためには複数の異なる観点から作成されたモデルを組み合わせることが一般的に行われている。



安全性を補完する別なモデル（安全性補完モデル）を構築し検証することとした。

実験方法

～安全性補完モデルのデータセット～

- 1週間のうち週5日間を学習データ、2日間をテストデータとして分割
- 1秒に1枚ずつ動画から画像を切り出し、15秒分を比較明合成した

2021年10月5日～2023年7月30日



学習データ：24,493,193枚、テストデータ：9,159,977枚

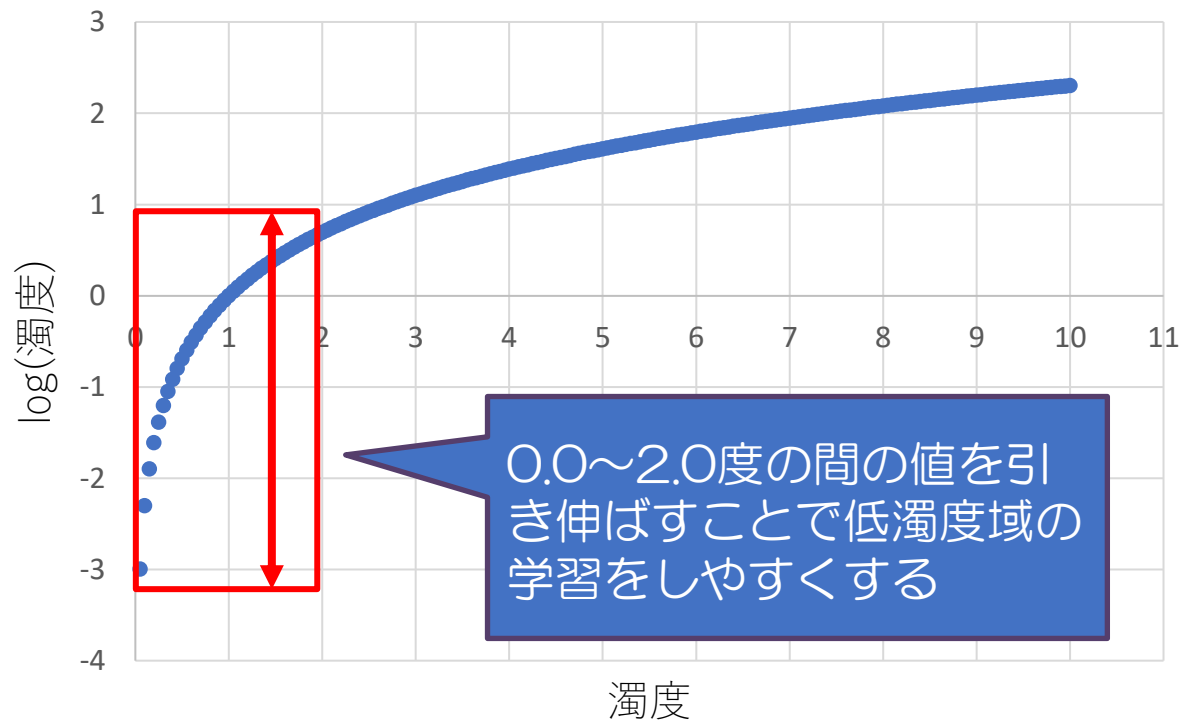
実験方法

～安全性補完モデルのチューニング～

モデル：ResNetR200

出力層：画像から予測後の濁度を回帰予測するように修正

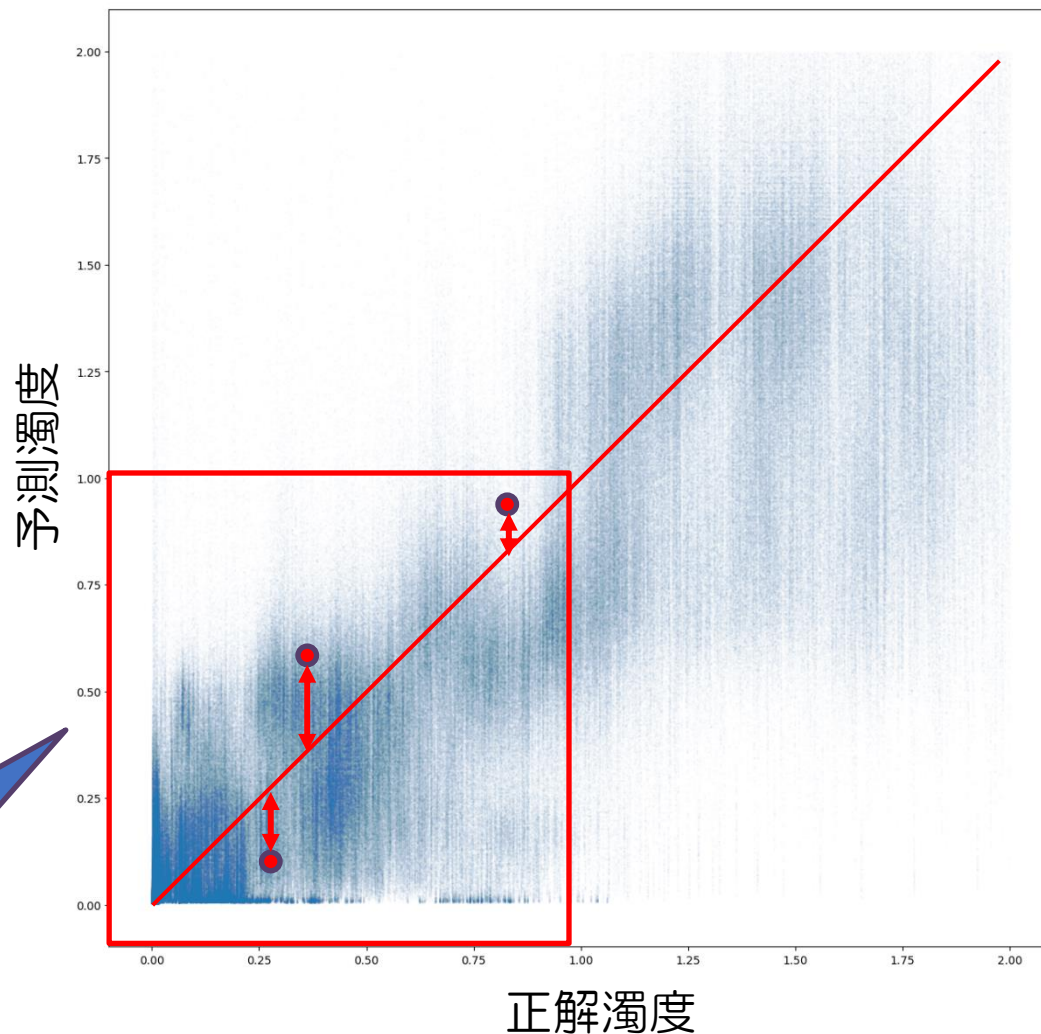
教師信号：凝集沈澱水濁度の自然対数をとった値を使用することで
低濁度域の予測性能を高める



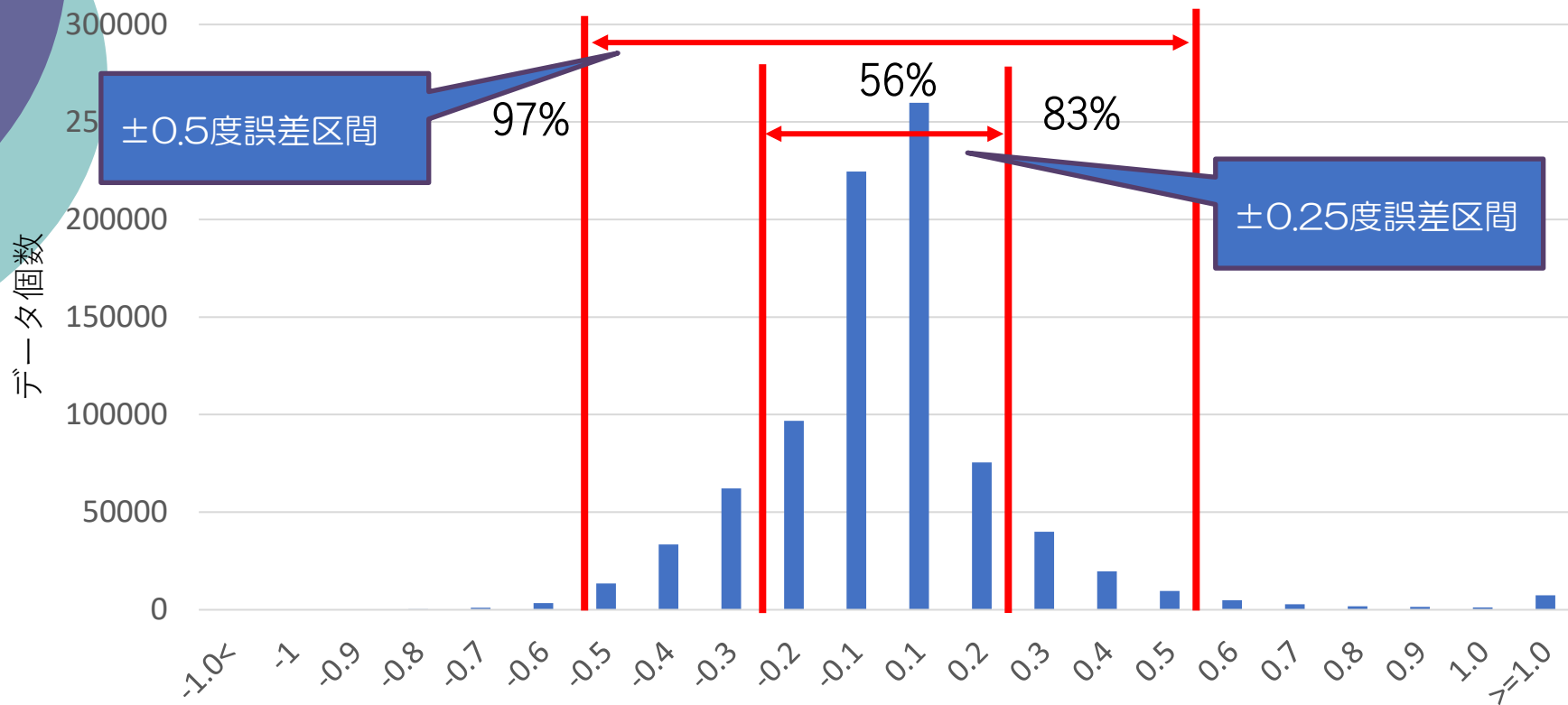
実験方法 ～ヒストグラムの作成～

モデルの結果がどの程度信頼できるのかを把握するために予測1件ごとの誤差からヒストグラムを作成した

1.0度以下のデータについてヒストグラムを作成



実験結果 ~ヒストグラムの作成~

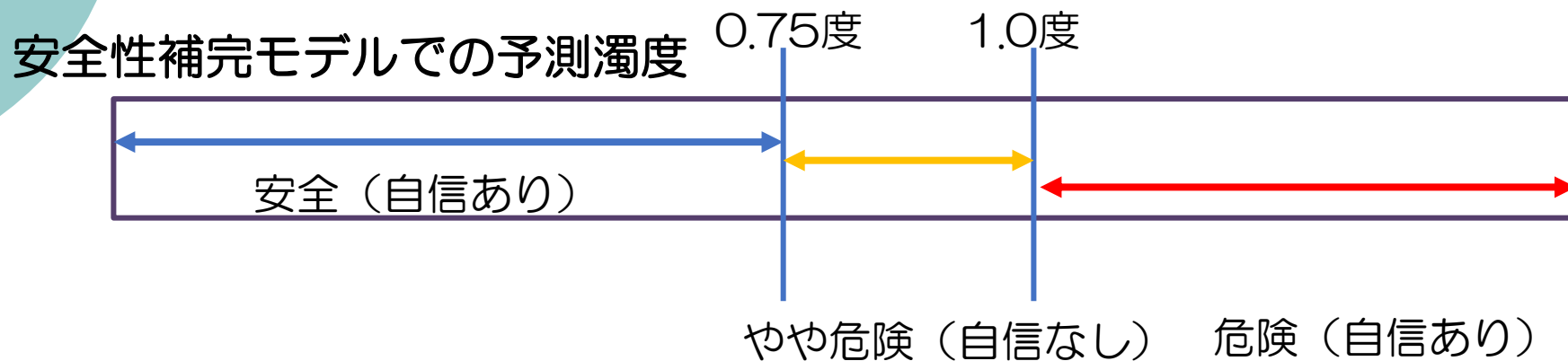


予測結果の97%は0.5度以下の濁度誤差で予測可能
同様に83%は0.25度で予測可能

実験結果

～安全性補完モデルへの応用～

多くの場合で0.25度の誤差で予測できることから・・・



安全性補完モデルの予測濁度から自信度への応用への可能性
0.5度をしきい値にすればより確実に安全性を補完可能

実験結果 ～画像分類モデルと安全性補完モデル～

2つのモデル（画像判断モデルと安全性補完モデル）を組合わせたシステムを構築

		画像判断モデル	
		管理値未 満	管理値以 上
安全性 補完モ デル	管理値未 満	正常	警報
	管理値以 上	警報	警報

どちらかのモデルが管理値以上になった場合に警報をだすことでより安全な凝集沈澱の運転が可能となる。

まとめ ～画像解析を用いた凝集判定手法の研究～

- モデルの前処理に暗合成-差分法を用いることで予測精度は向上した。
- 3分類を2分類にすることにより予測精度は向上し85%程度となった。
- 水温別にモデルを構築することにより全体的に予測精度が向上した。
- 凝集剤不足時の予測精度は非常に高かった。
- 画像判定モデルに安全性補完モデルを併用することでより安全な運転が可能となる。

7. 達成目標と検証①

1) ビッグデータを活用した原水水質の予測

①湖沼系原水等異なる水系による予測モデルを構築し、様々な地域・水域に対応できるモデルを構築する。

⇒少なくとも1年間のデータによりLSTMとバッチ正規化を行うことで、様々な原水における3時間後のかび臭濃度の予測が可能。また、同一水系であればモデルの重みを再利用することも可能であるが、別水系の場合は重みの再利用は困難である。

②吉見や大久保モデルを用いて実浄水場のリアルタイムデータを使用し精度を確認する。

⇒浄水場の1時間毎のリアルタイムデータを使用することでも3時間後のかび臭濃度の予測が可能。

ただし、学習時にはモニター上の数値の桁数を同一にする。

7. 達成目標と検証②

③水質基準を超えるような濃度の高いかび臭の予測精度を向上させるモデルを構築するあるいは高濃度のかび臭の増加予測が可能なモデルを構築する。

⇒高濃度であってもバッチ正規化を行うことによって精度よく予測することができた。

7. 達成目標と検証③

2) 画像解析を用いた凝集判定手法の研究

①画像データだけでなく、原水水質を組み込んだ学習モデルを構築する。

⇒原水水質を用いなくとも、凝集フロックの画像で凝集沈澱処理水の良し悪しを判断することができた。

②実際に構築モデルによりテストを行い、浄水場にとって支援となるような精度の高いモデルとする。

⇒凝集剤の注入不足に対する効果は高かった。また、管理目標濁度を定め2分類にすることで予測精度はさらに高まった。

8. 適用条件①

1) ビッグデータを活用した原水水質の予測

- 入力データは、原水のかび臭物質（2-MIBかジェオスミン）は必須であり、その他7項目（濁度、pH値、水温、導電率、塩素要求量、水温、水位）の計8項目を基本の入力データとする。しかし、測定していない項目がある場合には減らせるケースもある。
- かび臭物質は季節によって変動するため、少なくとも1年間のデータが必要である。
- 3時間後の予測を行うためには、1時間毎のデータが必要である。

8. 適用条件②

2) 画像解析を用いた凝集判定手法の研究

- 凝集は、原水の水温や濁度によって影響されるため、学習データとして1年間分のフロック画像データが必要になる。
- 画像はフロック形成（緩速攪拌）直後のフロックを対象とする。
- カメラ画質はフルHD以上を推奨する。

9. 研究成果の活用方法

1) ビッグデータを活用した原水水質の予測

- 今回の研究により少なくとも3時間後の原水のかび臭濃度が比較的高精度に予測することができた。
かび臭濃度を事前に予測できることによって浄水場での対応に余裕が持てるとともに、粉末活性炭の使用量削減にも寄与できる。また粉末活性炭の調達にも効果がでる。

2) 画像解析を用いた凝集判定手法の研究

- 今回の研究により凝集剤無注入あるいは不足している状態を凝集フロックの状態より判断することができ安全な沈殿池管理ができる。また、凝集沈殿水濁度ではなく、凝集フロックによって沈殿濁度の良否が早期に判断できることにより、高濁度等の突発的に発生する原水の流入時にも早急に対応することができ、適正な沈殿池の管理ができる。



謝辞

本研究を行うにあたり、実験フィールドを提供していただいた埼玉県企業局吉見浄水場並びに数年にもわたる貴重な水質データを提供して頂いた埼玉県企業局とその他の事業体の皆様に感謝申し上げます。