

〈ノート〉

人工知能による浄水場塩素注入管理の最適化
——トレンドデータの時間間隔の影響——隋 鵬 哲¹⁾, 島 村 和 彰²⁾, 森 次 歩 人³⁾, 山 口 陽 生⁴⁾¹⁾ 水ing(株) 研究開発センター
(〒251-8502 神奈川県藤沢市本藤沢4-2-1 E-mail: sui.pengzhe@swing-w.com)²⁾ 水ing(株) 研究開発センター
(〒251-8502 神奈川県藤沢市本藤沢4-2-1 E-mail: shimamura.kazuaki@swing-w.com)³⁾ (株)水みらい広島
(〒730-0041 広島県広島市中区小町1-25 E-mail: moritsugu.ayuto@mizumirai.com)⁴⁾ (株)水みらい広島
(〒730-0041 広島県広島市中区小町1-25 E-mail: yamaguchi.akio@mizumirai.com)

概 要

浄水場において、消毒のために広く次亜塩素酸ナトリウムが使用されており、確実な残留塩素管理及び水道技術者不足を解消するための技術継承が課題となっている。そこで、本研究は浄水場の残留塩素管理最適化を目的とし、人工知能を用いた次亜注入率モデルの構築を検討した。浄水場の運転データには、1時間間隔や5分間隔など異なる時間間隔のトレンドデータが存在する。そこで、今回AIモデルの予測精度を向上させるために、同じアルゴリズムで構築したAIモデルの机上検証及び現場検証で得られた結果を用いて、トレンドデータの時間間隔が次亜注入AIモデルの予測精度に及ぼす影響を検討した。

キーワード：次亜塩素酸ナトリウム、浄水処理、AI
原稿受付 2022.5.31 原稿受理 2022.8.17

EICA: 27(2・3) 73-79

1. はじめに

浄水場において、消毒のために広く次亜塩素酸ナトリウム（以降、「次亜」と表記）が使用されている。次亜注入率の決定は、ベテラン運転員が随時水質を監視しながら経験をもとに決定しており、確実な残留塩素管理が難しい。また、浄水場の運転において、今後水道技術者不足が予想される中、安定した施設のオペレーションには技術継承が欠かせない。特に、次亜を注入した後のろ過水残留塩素や浄水残留塩素濃度を維持するために、過小及び過剰量とならないような次亜注入量の最適化は、ベテラン運転員をもってしても難しいオペレーションの一つである。従来のろ過水残留塩素や浄水残留塩素濃度によるフィードバックやフィードフォワード制御注入は、タイムラグがあり、制御精度向上やコストダウンはまだ改善の余地がある。そこで、技術継承の取り組みの一つとして、浄水場の残留塩素管理及び次亜注入管理の最適化を目標として、人工知能(AI: Artificial Intelligence)を用いた次亜注入率予測用AIモデル（以降、「次亜モデル」と表記）を構築^{1,2)}し、現場検証³⁾を行った。

さらに、近年AIの急速発展より、AIの水処理への活用は研究開発の焦点になっているが、時系列AI

モデルに対しては、学習したトレンドデータの時間間隔によるモデル予測精度の影響、及び最適な時間間隔に関する検討はまだない。時間間隔の長いトレンドデータは、所定期間のデータ量（データの数）が少ないので、データの詳細や変化傾向などの情報は獲得しにくい。一方、時間間隔の短いトレンドデータは、データ量が多いので左記のような情報は得られやすいものの、逆にランダム変化などのノイズも多くなる。そこで本稿は、AIによる塩素注入の最適化モデルの構築と実装に向けて、まず次亜モデルの予測精度を向上させるために、トレンドデータの時間間隔の影響に関する検討結果を報告する。なお時間間隔以外の検討は引き続き実施しているので後報としたい。

2. データ及び方法

浄水場の塩素処理は、沈澱池より前で次亜を注入する「前塩素処理」、沈澱池とろ過池の中間で注入する「中間塩素処理」、ろ過池以降に添加する「後塩素処理」がある。今回は、A浄水場の中間塩素処理における「中次亜注入率」と、後塩素処理における「後次亜注入率」を対象とした。各次亜モデルは、水量、水温、雨量を共通変数とし、中次亜注入率は沈澱処理水

及びろ過水の水質を加えた10項目、後次亜注入率はろ過水及び浄水の水質を加えた9項目を説明変数とした。AIモデル構築に用いたアルゴリズムは、既報²⁾で①全結合ニューラルネットワーク (FANN, Fully connected Artificial Neural Network) と、②回帰型ニューラルネットワーク (RNN, Recurrent Neural Network) の一種であり時系列データに対して長期依存性を学習可能なことが特徴である長・短期記憶モデル (LSTM, Long Short-Term Memory)⁴⁾の2種類を検討した結果、LSTMの汎用性が良いことが分かったので、今回の検討ではLSTMを利用することとした。AIモデルの構築は、Pythonを利用し、ライブラリー「Keras」を用いた。また、Kerasのバックエンドとして、Google社のTensorFlowを用いた。モデルを評価するために、目的変数の実測値と予測値の相関係数 (Cor, Correlation Coefficient) 及び平均二乗誤差 (MSE, Mean Squared Error) を予測精度の評価指標とした。

次亜モデル構築に用いたデータは約2年分の自動水質計器及び運転データ (2017/12~2019/11) である。トレンドデータの時間間隔は1分、5分、1時間である。構築したモデルの検証は机上検証と現場検証を実施した。机上検証は、1時間間隔のデータを用いたモデルの汎用性を確認するために、新たな運転データ (2019/11~2020/05) を利用し、予測精度を検証した。現場検証は、2020/6~2021/5の一年間、現場でリアルタイムデータをオンライン自動取得し、予測精度をリアルタイムで確認した。なお、現場検証の結果より、予測のタイムラグなどの課題を解消し、また予測精度を向上させるために、1時間間隔のモデルの他に、1時間間隔と同様のデータ期間で、5分間隔のトレンドデータを用いたモデルを再構築し、再度検証した。5分間隔の検証については、1時間間隔モデルと同じ期間のデータを用いた机上検証、及び別途開発した現場検証ソフトのデモモードを利用し、リアルタイム予測をシミュレーションしたバーチャル現場検証とし、1時間間隔モデルの検証と同じ期間の予測精度を確認した。

更に、予測精度が向上した原因を検討するために、1時間間隔データと同様のデータ量で5分間隔のモデルを再構築し、データ量の影響を検討した。

上記全てのモデルの構築において、運転データは適宜異常値の削除や修正等の前処理を行った。データセットを三つに分け、それぞれ学習 (Train) 用、検証 (Validation) 用、及び予測 (実際の応用を想定した将来予測, Prediction) 用とした。

1時間間隔のトレンドデータ数は約1.5万個であった。データセットを時系列順に並べた際の最初の10,000個をモデルの学習及び検証用に使用し、その中

の7,500個を学習データ、2,500個を検証データとした。残りのデータは将来予測データとした。

また、5分間隔のトレンドデータ数は約19万個であった。その内、2018/11/30までの約9.5万個を学習データ、2018/12/1~2019/3/31の約3万個を検証データ、2019/4/1以降の約6.5万個を将来予測データとした。

更に、予測精度向上におけるデータ量の影響を検討するために、1時間間隔のトレンドデータと同様の量 (約1.5万個) の5分間隔トレンドデータを用いて、モデルを再構築した。このとき、最初の7,500個を学習データ、2,500個を検証データとし、残りのデータは将来予測データとした。

3. 結果及び考察

3.1 1時間間隔のトレンドデータで構築した次亜モデルの予測精度

(1) 中次亜モデルの予測精度

中次亜モデルの机上検証結果について、半年の予測期間中、予測値と実測値の相関係数は0.941であり、平均二乗誤差は0.00606であった。机上検討結果より、中次亜の注入率は、全体的に高めの予測傾向であったが、非常に高い精度で予測された。

一年間実施した中次亜モデルの現場検証結果について、予測値と実測値の経時変化を **Fig. 1 (c)** に、数値の比較を **Fig. 1 (a)** に示す。予測値と実測値の相関係数は0.947であり、平均二乗誤差は0.00609であった。現場検証結果より、全てのピークを精度よく予測できたが、時系列のピークの遅延が観察された³⁾。中次亜モデルの予測値の誤差分布について、予測誤差の絶対値は0.2 mg/L以下で、相対誤差の絶対値は20%以下に落ち着いた。また、絶対誤差/相対誤差の分布により計算した中次亜予測値の絶対誤差/相対誤差の出現確率は、絶対誤差が0.07 mg/L以内になる確率は80%以上、相対誤差が8.5%以内になる確率は80%以上であり (データの数は約7,000個)、開発目標である相対誤差が10%以内になる確率80%以上をクリアした。上記中次亜モデルの現場検証結果より、机上検証結果と同様に、中次亜モデルは非常に高い精度を達成した。

(2) 後次亜モデルの予測精度

後次亜モデルの机上検証結果について、半年の予測期間中、予測値と実測値の相関係数は0.691であり、平均二乗誤差は0.00149であった。机上検討結果より、モデルは後次亜注入率の変化傾向をよく予測していたが、実際の注入率が0.3 mg/L以上である場合は正しく予測されていないことが明らかとなった。

後次亜モデルの現場検証結果について、予測値と実

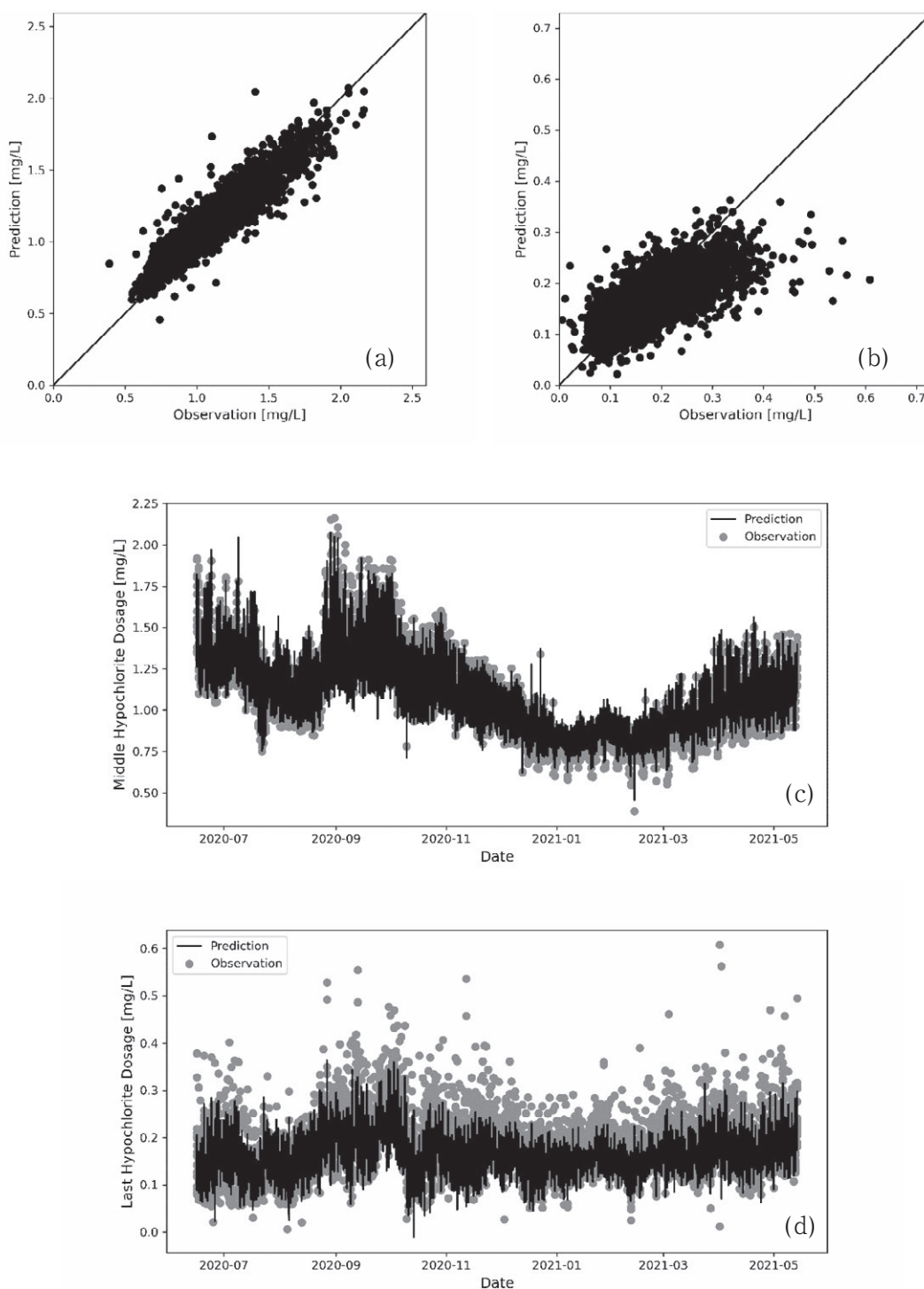


Fig. 1 The onsite calibration results of AI model built by hourly trend data
 (a) and (c) Model for Middle Hypochlorite Dosage; (b) and (d) Model for Last Hypochlorite Dosage

測値の経時変化を **Fig. 1 (d)** に、数値の比較を **Fig. 1 (b)** に示す。予測値と実測値の相関係数は 0.745 であり、平均二乗誤差は 0.00188 であった。机上検証結果と同様に、現場検証結果は後次亜注入率の変化傾向をよく予測していたが、0.3 mg/L 以上の高注入率の場合は、正しく予測されていないことがあった。毎日の取水停止及び再開に伴い、後次亜を高めに注入したことが原因であると推測される。取水再開時は、運転員が安全性の担保のために、今までの経験を基に高めの

注入率としているが、このような経験に基づくオペレーションを踏まえた最適注入率をシミュレーションすることは、今後の検討課題である。後次亜モデルの予測値の誤差分布について、予測誤差の絶対値は 0.1 mg/L 以下で、相対誤差の絶対値は 50% 以下に落ち着いた。また、絶対誤差/相対誤差の分布により計算した後次亜予測値の絶対誤差/相対誤差の出現確率は、絶対誤差が 0.05 mg/L 以内になる確率が 80% 以上、相対誤差が 27.5% 以内になる確率が 80% 以上で

あり（データの数は約 7,000 個）、相対誤差が 10% 以内になる確率 80% 以上の開発目標をクリアできなかった。

3.2 5 分間隔のトレンドデータで構築した次亜モデルの予測精度

上記 1 時間間隔のトレンドデータで構築した次亜モデルの現場検証結果より、特に後次亜注入率の予測精度をさらに向上させる必要があることが分かった。そこで、予測精度向上のために、トレンドデータの時間

間隔が予測精度に与える影響を調査することにした。また、実装の際にも、原水変動に追従するために適した推奨値の提示間隔の把握が必要である。そのため、ここでは、更に時間間隔の短い運転データ（5 分間隔）を用いたモデルと、1 時間間隔の精度を比較することにした。

(1) 中次亜モデルの予測精度

5 分間隔の中次亜モデルのバーチャル現場検証結果（リアルタイム現場検証相当）について、予測値と実測値の経時変化を **Fig. 2 (c)** に、数値の比較を

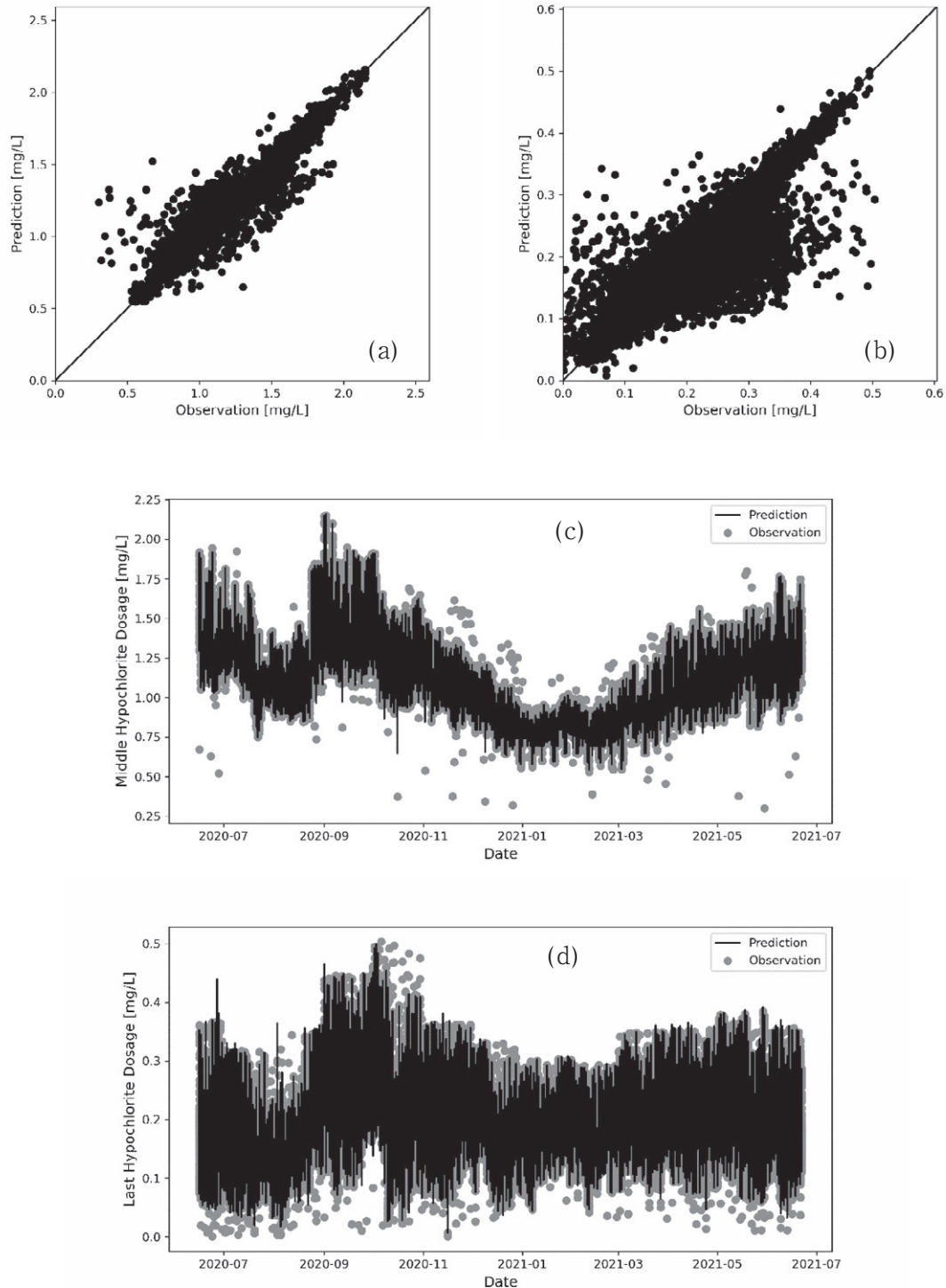


Fig. 2 The virtual onsite calibration results of AI model built by 5-min trend data (a) and (c) Model for Middle Hypochlorite Dosage; (b) and (d) Model for Last Hypochlorite Dosage

Fig. 2 (a) に示す。予測値と実測値の相関係数は0.993であり、平均二乗誤差は0.000690であった。中次亜注入率予測値の誤差分布について、予測誤差の絶対値は0.075 mg/L以下で、相対誤差の絶対値は7.5%以下に落ち着いた。また、絶対誤差/相対誤差の分布により計算した中次亜注入率の予測値の絶対誤差/相対誤差の出現確率は、絶対誤差が0.03 mg/L以内かつ相対誤差が3.0%以内になる確率として90%以上であった(データの数は約14万個)。上記検証結果より、中次亜モデルは非常に高い精度を達成した。

(2) 後次亜モデルの予測精度

5分間隔の後次亜モデルのバーチャル現場検証結果(リアルタイム現場検証相当)について、予測値と実測値の経時変化を**Fig. 2 (d)**に、数値の比較を**Fig. 2 (b)**に示す。予測値と実測値の相関係数は0.959であり、平均二乗誤差は0.000262であった。後次亜注入率予測値の誤差分布については、予測誤差の絶対値は0.04 mg/L以下で、相対誤差の絶対値は20%以下に落ち着いた。また、絶対誤差/相対誤差の分布により計算した後次亜注入率の予測値の絶対誤差/相対誤差の出現確率は、絶対誤差が0.02 mg/L以内かつ相対誤差が9.1%以内になる確率として90%以上であった(データの数は約14万個)。上記検証結果より、5分間隔のトレンドデータで構築した後次亜モデルは、1時間間隔モデルより精度が大幅に向上した。

3.3 予測精度向上の原因検討

1時間間隔及び5分間隔のトレンドデータで構築した中次亜モデルと後次亜モデルにおいて、それぞれのモデルを構築する際の将来予測の精度、机上検証の予測精度、及び現場検証の予測精度を**Table 1**にまとめた。結果より、5分間隔のトレンドデータで再構築した次亜モデルは、1時間間隔の次亜モデルより予測精度が大幅に向上したことが示された。

5分間隔の次亜モデルの予測精度が高い原因については、

- ① 1時間間隔トレンドデータより、5分間隔のデータに含まれる情報量(局所的な変化や傾向など)が多い。
- ② 1時間間隔トレンドデータより、5分間隔のデータの数が多。

という2点が挙げられる。5分間隔データは1時間間隔データよりも水質の変化をより細かに追跡できているため、予測精度が高かったと考えられる。ただし、5分間隔のトレンドデータが持つ情報には、ノイズも1時間間隔トレンドデータより多く含まれることが懸念される。

5分間隔のモデルの予測精度が向上した原因を確認するために、1時間間隔のモデルとほぼ同数のデータ量(約1.5万個)の5分間隔トレンドデータを用い、同じデータ量の場合の両モデルの予測精度を比較した。

Table 1 Comparison of prediction accuracy for AI model with different time span

項目	予測期間	トレンドデータの時間間隔			
		1時間		5分	
		相関係数	MSE*	相関係数	MSE*
<中次亜モデル>					
モデル構築時	2019/03/01 [#] ~ 2019/11/20	0.947	8.27 × 10 ⁻³	0.994	8.72 × 10 ⁻⁴
机上検証	2019/11/20 ~ 2020/05/20	0.941	6.06 × 10 ⁻³	0.993	3.16 × 10 ⁻⁴
現場検証	2020/06/16 ^{**} ~ 2021/05/14	0.947	6.09 × 10 ⁻³	0.993	6.90 × 10 ⁻⁴
<後次亜モデル>					
モデル構築時	2019/03/01 [#] ~ 2019/11/20	0.829	1.65 × 10 ⁻³	0.967	3.37 × 10 ⁻⁴
机上検証	2019/11/20 ~ 2020/05/20	0.691	1.49 × 10 ⁻³	0.956	1.90 × 10 ⁻⁴
現場検証	2020/06/16 ^{**} ~ 2021/05/14	0.745	1.88 × 10 ⁻³	0.959	2.62 × 10 ⁻⁴

注：
[#]: 5分間隔トレンドデータを用いたモデル構築の検証期間は2019/04/01から。
^{**}: 5分間隔トレンドデータを用いたモデル現場検証期間は2021/06/22まで。
 *: MSE: Mean Square Error、平均二乗誤差。

Table 2 The effect of learning data amount on the prediction accuracy (results of desktop simulation)

項目	トレンドデータの時間間隔					
	1時間 (約1.5万)		5分 (約19万)		5分 (約1.5万)	
	相関係数	MSE*	相関係数	MSE*	相関係数	MSE*
中次亜モデル	0.941	6.06 × 10 ⁻³	0.993	3.16 × 10 ⁻⁴	0.974	3.07 × 10 ⁻³
後次亜モデル	0.691	1.49 × 10 ⁻³	0.956	1.90 × 10 ⁻⁴	0.943	2.49 × 10 ⁻⁴

注：
 *: MSE: Mean Square Error、平均二乗誤差。

また、少ないデータ量（約 1.5 万個）で構築した次亜モデルを用いて机上検証を行い、データ量が多い（約 19 万個）場合の次亜モデルの机上検証結果と比較した。以上の結果を **Table 2** にまとめた。また、中次亜モデルの予測値と実測値の経時変化及び数値の比較を **Fig. 3** に示す。

約 1.5 万個の 5 分間隔トレンドデータで再度構築した次亜モデルの机上検証結果より、中次亜モデルと後次亜モデルの予測精度について、実測値と予測値の相

関係数と平均二乗誤差は、それぞれ 0.974 と 0.943、 3.07×10^{-3} と 2.49×10^{-4} であった。約 19 万個の 5 分間隔トレンドデータで構築したモデルより予測精度が低くなったが、1 時間間隔のモデルと比較し、中次亜モデルの平均二乗誤差は約 50%（5 分間隔： 3.07×10^{-3} 、1 時間間隔： 6.06×10^{-3} ）に、後次亜モデルは約 1/6（5 分間隔： 2.49×10^{-4} 、1 時間間隔： 1.49×10^{-3} ）に低減した。この結果より、同じデータの量でも、5 分間隔のトレンドデータで構築したモデルは、1 時間間隔の

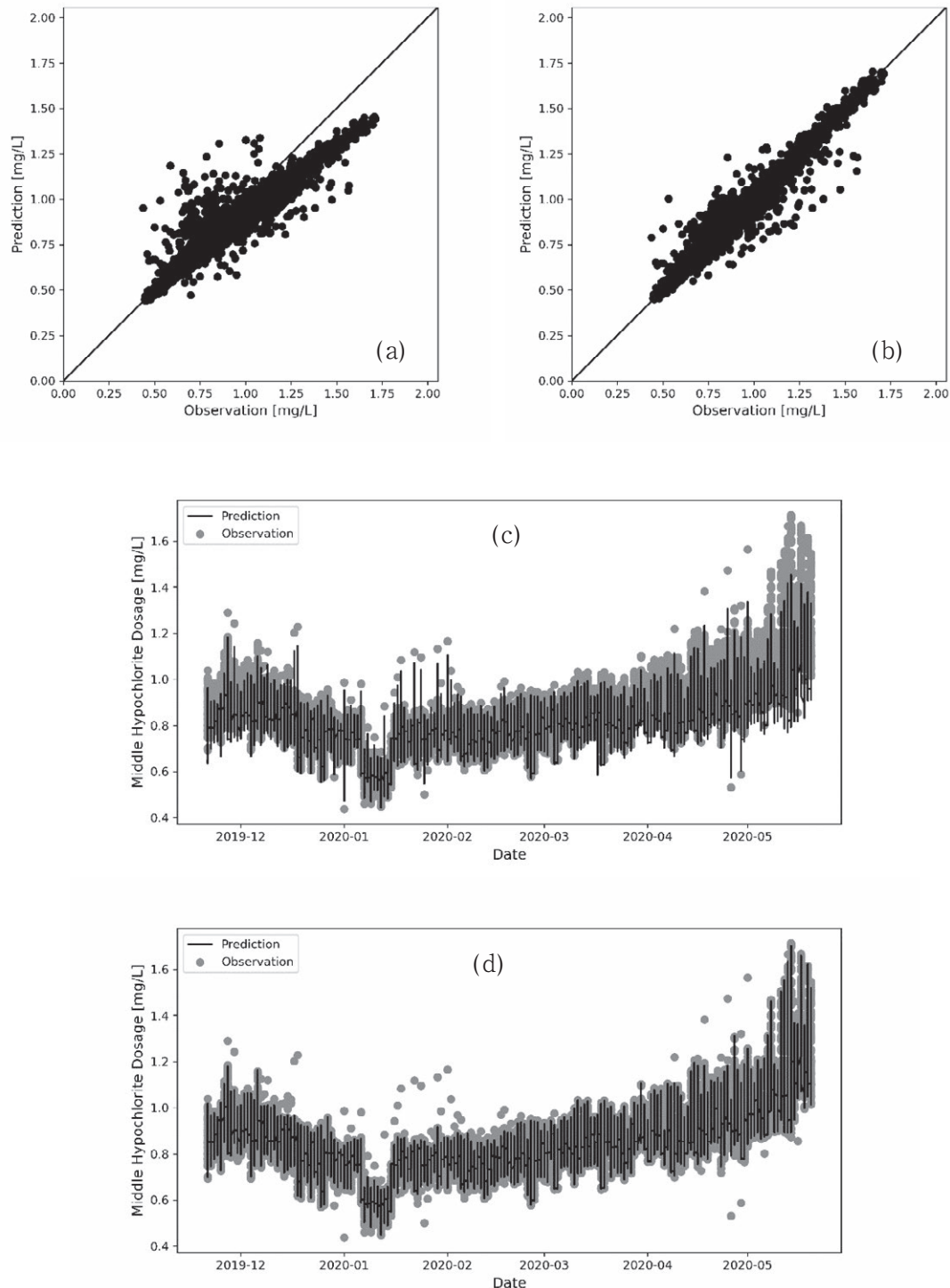


Fig. 3 Comparison of prediction accuracy for AI model trained with different data amount. (Model for Middle Hypochlorite Dosage with 5-min trend data)
 (a) and (c) train data amount of 10 thousand; (b) and (d) train data amount of 140 thousand.

モデルより予測精度が大幅に向上したことが明らかとなった。5分間隔のモデルの高予測精度の要因は、学習したデータの量が単に増えただけではなく、5分間隔のトレンドデータに含まれる局所的な変化などの情報量が多いためであると考えられる。また、5分間隔トレンドデータに含まれるノイズは1時間間隔トレンドデータのノイズより多いものの、予測精度に及ぼす影響は小さいと考えられる。

4. ま と め

本研究は浄水場の残留塩素管理最適化を目的とし、人工知能を用いた次亜注入率モデルの構築を検討した。今回 AI モデルの予測精度を向上させるために、同じアルゴリズムで構築した AI モデルの机上検証及び現場検証で得られた結果を用いて、トレンドデータの時間間隔が次亜注入 AI モデルの予測精度に及ぼす影響を検討した。検討結果より、5分間隔のトレンドデータで構築した次亜モデルは、1時間間隔の次亜モデルより予測精度が大幅に向上したことが示された。また、

精度向上の要因は、単に学習したデータの量が増えたためではなく、5分間隔のトレンドデータが持つ詳細な情報量が多いためであった。

謝 辞

本研究を進めるにあたり、広島県企業局にご協力頂いた。ここに感謝の意を表す。

参 考 文 献

- 1) 隋鵬哲, 鈴木美有, 島村和彰, 坂下真栄: 人工知能による次亜塩素酸ナトリウム注入率の最適化, 令和2年度水道研究発表会講演集, p. 318
- 2) 隋鵬哲, 島村和彰: 人工知能による浄水場塩素注入管理の最適化, 環境浄化技術, 2021年7・8月号, p. 43
- 3) 隋鵬哲, 島村和彰, 森次歩人, 山口陽生: 人工知能による浄水場塩素注入管理の最適化— AI モデルの現場検証, 令和4年度水道研究発表会講演集, (掲載予定)
- 4) Sepp Hochreiter, Jurgen Schmidhuber, 1997, Long short-term memory, Neural Computation, 9 (8): 1735-1780. doi:10.1162/neco.1997.9.8.1735.

Optimization of Hypochlorite Dosage for Drinking Water Treatment by AI: The Effect of Time Span of Real-time Data on Prediction Accuracy

Pengzhe Sui^{1)†}, Kazuaki Shimamura¹⁾, Ayuto Moritsugu²⁾ and Akio Yamaguchi²⁾

1) Research and Development Center, Swing Corporation

2) MIZU MIRAI HIROSHIMA Co., Ltd.

† Correspondence should be addressed to Pengzhe Sui :

Research and Development Center, Swing Corporation

E-mail : sui.pengzhe@swing-w.com

Abstract

Hypochlorite is widely used in drinking water treatment plant for the purpose of disinfection, and the control of total residual chlorine in treated water is one of key issues for the operation and maintenance, especially facing the challenge of technological inheritance caused by the shortage of operators in Japan. Based on such situation, the optimization of hypochlorite dosage by AI was investigated in this study. There are different time spans for the real-time operational data in drinking water treatment plant and their effect on the prediction accuracy of developed AI model for hypochlorite dosage was studied with the results of desktop simulation and onsite verification in order to increase the prediction accuracy.

Key words : hypochlorite dosage, drinking water treatment, AI