

II. 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）

「水道の基盤強化に資する技術の水道システムへの実装に向けた研究」

分担研究報告書

ビッグデータに基づく水質変動の早期予測手法の検討

研究分担者 中央大学 山村 寛

研究要旨

塩素注入量は、現在、高度な技術を持つ職員が経験に基づいて注入量を決定しているが、大量のベテラン職員の退職と職員数の減少などから、経験に依存しない、新しい塩素注入量管理手法が求められている。本研究では、浄水場が保有する残留塩素濃度の時系列データに着目し、時系列の濃度変化の傾向を学習することで、数時間先の残留塩素濃度を推測できる予測モデルの構築を目的とした。具体的には、長期短期記憶ネットワーク（LSTM）アルゴリズムにより、3時間、6時間、12時間、24時間先の残留塩素濃度予測モデルの構築を試みた他、モデルの構築に必要な最小限のデータ量を検討した。

モデル構築にあたって最適なブロックは24時間であり、誤差目標値 ± 0.025 mg/L以下に収めるには、予測時間を6時間以下にする必要があることが判明した。また、モデル構築に必要な最小限のデータ量は4月～7月の4ヶ月間の残留塩素濃度低減量であることがわかった。以上より残留塩素濃度の1時間間隔の時系列データが4ヶ月分準備できれば、6時間先の残留塩素濃度をLSTMにより予測できることがわかった。

A. 研究目的

日本では、水道法において蛇口における残留塩素濃度を0.1 mg/L以上に維持することが義務づけられている¹⁾。塩素消毒は高い消毒効果を長時間にわたって持続できる一方で、浄水処理施設で注入された塩素消毒剤が配水管・配水池ならびに給水管を経由して給水栓や一般家庭等の蛇口に到達する間に、水中のフミン質やアンモニア態窒素、配水管の管路壁面や表面に付着した生物膜などとの化学反応により、徐々に残留塩素濃度が希薄化する。給水栓や蛇口で残留塩素濃度0.1 mg/Lを維持するためには、

配水中に消費される残留塩素量を勘案した上で、浄水場での塩素注入量を決定する必要がある²⁻⁴⁾。

塩素消毒剤と水中の有機物が反応することで、トリハロメタンをはじめとする消毒副生成物が生成される。水質管理目標設定項目では残留塩素濃度が1 mg/L以下となるように設定されている他、総トリハロメタン濃度が0.1 mg/L以下となるように水質基準項目が定められている¹⁾。神奈川県内広域水道企業団では、消毒副生成物の抑制も考慮した残留塩素濃度の管理値を各給水地点で季節別に定めており、現状、高度な技



図-1 神奈川県内広域水道企業団が管理する浄水場と給水地点。本研究では相模原浄水場から上和田および西谷給水地点までに低減する残留塩素量を予測するモデルを構築した

術を持つ職員が経験に基づいて適宜、塩素注入量を判断している。日本は、2008年に人口がピークに達した後、徐々に人口が減少する人口減少社会に突入した。浄水場の職員数も徐々に減少しており、2030年までには、2000年比30%程度職員が減少すると予測されている⁵⁾。特に、高度な技術を持つベテラン職員の大量退職を控えており、これらの技術と経験の継承が重要な課題となっている⁵⁾。既存施設を持続的に維持・管理していくためにも、職員の技能や経験に依存しない、新しい浄水場の運転管理手法が求められている。そこで、本研究では配管内の残留塩素濃度の低減量を予測するモデルの構築に挑戦する。予測モデルが構築できれば、高度な技術を持った職員の判断を必要とせず、正確かつ迅速な塩素注入量の設定を自律的に最適化できるようになると期待する。

これまで、配管内の残留塩素予測を目的として、様々な物理モデルが構築されている⁶⁻¹⁰⁾。代表的なものとして、米国EPAが提供するEPANETが世界中の水道事業者にご利用されており、滞留が存在しない配管では比較的正確に残留塩素濃度の予測が可能とされる⁹⁾。一方で、貯水槽や滞留を伴う配管および二次枝管などについては既存モデルの適用が難しいことが指摘されている⁹⁾。Abokifaら⁹⁾は、既存の物理モデルに確率需要発生器を接続することで、水需要の変動による滞留時間の変化を組み込んだ確率モデルを開発したが、既存モデルよりも精度が向上した一方で、塩素注入制御に足る精度には至っていない。

神奈川県内広域水道企業団は、神奈川県内の4事業者（神奈川県営水道、横浜市水道局、川崎市上下水道局、横須賀市上下水道局）へ浄水を給水する特別地方公共団体

表-1 モデル構築に用いたデータセットの概要

データセットA：5分間隔の2年分データ

		全データ	欠損値	異常値	30分にデータを圧縮+ 移動平均算出後	学習データ	テストデータ	平均値	最小値	最大値
相模原	残留塩素	210528	0	129				0.73 mg/L	0.64 mg/L	0.87 mg/L
	水温	210528	0	0	35088	29833	5255	16.7 °C	0 °C	26.5 °C
	電気伝導度	210528	0	130				16.5 mS/m	12.1 mS/m	18.4 mS/m
上和田	残留塩素	210528	0	103				0.67 mg/L	0.49 mg/L	0.80 mg/L
	水温	210528	0	2	35088	29833	5255	17.0 °C	0 °C	30.3 °C
	電気伝導度	210528	0	68				17.6 mS/m	12.4 mS/m	19.0 mS/m
西谷	残留塩素	210528	0	67				0.64 mg/L	0.55 mg/L	0.81 mg/L
	水温	210528	0	1	35088	29833	5255	17.2 °C	0 °C	29.0 °C
	電気伝導度	210528	0	66				16.0 mS/m	12.7 mS/m	19.9 mS/m

データセットB：1時間間隔の10年分データ

		全データ	欠損値	異常値	移動平均算出後	学習データ	テストデータ	平均値	最小値	最大値
相模原	残留塩素	87672	0	19				0.75 mg/L	0.63 mg/L	0.89 mg/L
	水温	87672	0	2	87648	74500	13148	16.32 °C	5.2 °C	27.0 °C
	電気伝導度	87672	0	43				16.6 mS/m	12.5 mS/m	19.2 mS/m
上和田	残留塩素	87672	0	22				0.68 mg/L	0.54 mg/L	0.88 mg/L
	水温	87672	0	3425	87648	74500	13148	16.78 °C	0 °C	30.2 °C
	電気伝導度	87672	0	3508				17.2 mS/m	12.8 mS/m	21.7 mS/m
西谷	残留塩素	87672	0	1701				0.67 mg/L	0.49 mg/L	0.87 mg/L
	水温	87672	0	1485	87648	74500	13148	16.61 °C	4.1 °C	28.9 °C
	電気伝導度	87672	0	1561				17.5 mS/m	12.3 mS/m	22.7 mS/m

データセットC：1時間間隔の3年分データ

		全データ	欠損値	外れ値	移動平均算出後	学習データ	テストデータ	平均値	最小値	最大値
石川	残留塩素	26304	0	0	26280	17520	8760	1.11 mg/L	0.93 mg/L	1.32 mg/L
金武	残留塩素	26304	0	0	26280	17520	8760	0.65 mg/L	0.40 mg/L	0.83 mg/L
東恩納	残留塩素	26304	0	15	26280	17520	8760	0.92 mg/L	0 mg/L	1.10 mg/L
山城	残留塩素	26304	0	2	26280	17520	8760	0.92 mg/L	0.76 mg/L	1.25 mg/L
具志川	残留塩素	26304	0	9	26280	17520	8760	0.79 mg/L	0.45 mg/L	1.47 mg/L

である。浄水処理された水は、42ヶ所の給水地点を経由して、各事業体に供給されており、各給水地点において、基本的な水質項目が連続的に監視されている(図-1)。よって、神奈川県内広域水道企業団は、浄水場の出口に加え、各給水地点において、基本的な水質に関する連続監視データを保有していることになる。これらの膨大なデータセットを活用することで、給水地点における残留塩素濃度を高精度に予測しうるモデルが構築できるものと期待する。

ビッグデータを利用した制御方法として、ニューラルネットワークを用いたモデル構築が挙げられる。計算機の進化に伴って、ニューラルネットワークの中間層を時系列

の前後で接続することで、時系列変化のパターンを学習するリカレントニューラルネットワーク(RNN: Recurrent neural network)が開発され、様々な分野で将来予測に利用されている。Bowdenら¹⁰⁾は、南オーストラリア、アデレード南部の配水システムを対象として、浄水場出口、ポンプ場、給水地点での残留塩素データを使用してRNNによる72時間後の残留塩素レベル予測モデル

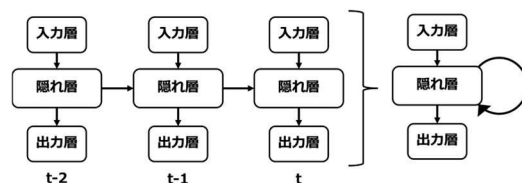


図-2 LSTM ネットワーク概要

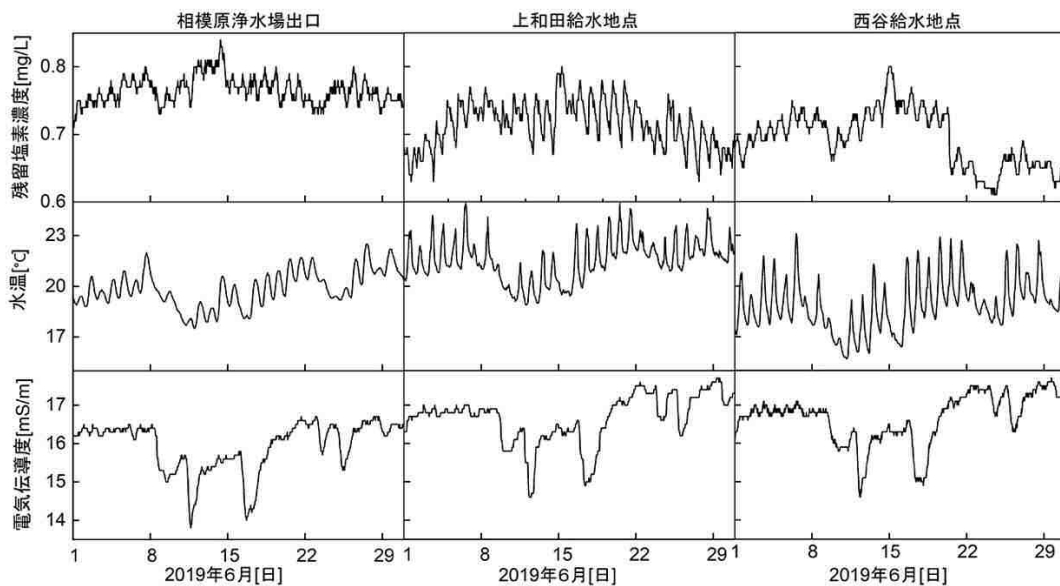


図-3 データセット A、B の 2019 年 6 月における残留塩素濃度，水温，電気伝導度の 1 ヶ月変動(图中，異常値除去後の値を示す)

を構築した結果、 $R^2=0.96$ の精度で予測することに成功している。本モデルは、浄水場出口および給水地点前段の塩素濃度を把握する必要があるため、残留塩素濃度の連続測定装置を多数備えた地域に限定されることが欠点として挙げられる。実用性の高いモデル構築には、さらに簡潔なデータセットによる予測モデルが必要とされる。

近年、特に長期間の予測を目的として、RNN の隠れ層に CEC(constant error carousel) を設置することで、入力ゲート、出力ゲート、忘却ゲートによって、過去から引き継いだデータを必要に応じて取得・修正・消去できる特徴を持つ長短期記憶 (LSTM: Long-short term memory) モデルが開発された (図-2)。Xuan-Hien Le ら¹¹⁾は、LSTM によってベトナムの洪水予測モデル構築に挑戦している。1961 年から 1984 年の 24 年間、上流地点の降雨量と流量を入力値として、下流地点における 3 日後の流量を予測する

モデルを構築した結果、95%以上の予測精度が得られている。LSTM により精確な予測結果を得るには、ブロック (もしくは、タイムステップ) と予測時間を適切に設定する必要がある。ブロックは 1 ユニットの時系列データの範囲を示す値であり、水質変動の周期に合わせて設定する必要がある。水質変動の周期よりブロックが短すぎると変化のパターンが単調になり、特徴を捉えられずに、モデル精度の低下を招く。一方で、水質変動の周期よりブロックが長すぎる場合、変化のパターンが多すぎるため、全てのトレンドを学習するには膨大なデータ量が必要となる。

以上の背景から、本研究では神奈川県内広域水道企業団が保管する浄水場と給水地点における膨大なデータセットを活用して、3~12 時間先の給水地点における残留塩素濃度を予測しうるモデルを構築する。モデル構築にあたって、時系列データの将来予

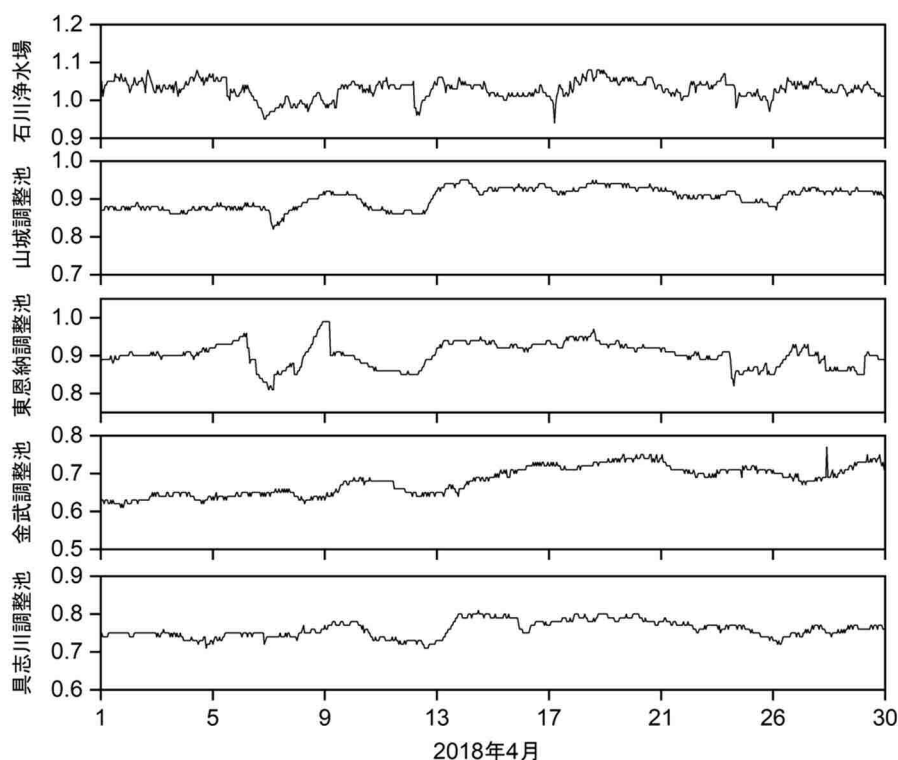


図-4 データセット C の 2018 年 4 月における残留塩素濃度の 1 ヶ月変動
(図中、異常値除去後の値を示す)

測に有効な RNN モデルのうち、長期間の予測モデル構築に優れる LSTM モデルを利用する。具体的には、データの前処理アルゴリズムを検討すると共に、モデルの各種パラメータ、データ種、データ数が予測精度に及ぼす影響について検討する。また、沖縄県企業局が保管する浄水場と調整池地点における膨大なデータセットを活用して、モデルの汎用性を評価した。

B. 研究方法

モデル構築に利用したデータ

本研究で使用したデータは、水道技術研究センターの協力の下、神奈川県内広域水道企業団、および沖縄県企業局より提供頂いた。図-1 に示す通り、神奈川県内広域水道企業団は相模川、酒匂川で取水した水を

6ヶ所の浄水場で浄水している。沖縄県企業局はダムより取水した水を 5ヶ所の浄水場で浄水している。

本研究では、相模川および酒匂川の 2 河川を水源とする相模原浄水場を基点として、約 15km 先に位置する上和田給水地点（平均給水量約 7,500m³/hr）、および矢指調整池を経由して約 20km 先に位置する西谷給水地点（平均給水量約 75,000m³/hr）における残留塩素濃度低減量を予測するモデルを構築する。上記の 2 給水地点は、通常、相模原浄水場からのみの給水となっている。相模原浄水場出口、上和田給水地点、西谷給水地点での 2018 年 7 月 1 日から 2020 年 6 月 30 日（2 年間）における 5 分間隔の残留塩素濃度（mg/L）、水温（℃）、電気伝導度（mS/m）（以下データセット A と表記する）

と、2010年4月1日から2020年3月31日（10年間）における1時間間隔の残留塩素濃度(mg/L)、水温(°C)、電気伝導度(mS/m)（以下データセットBと表記する）を使用した。また、モデルの汎用性の評価には、久志浄水場処理水、漢那・金武・山城ダムを水源とする石川浄水場を基点として、金武調整池、東恩納調整池、山城調整池、および具志川調整池における残留塩素濃度低減量を予測するモデルを構築した。それぞれの調整池での2018年4月1日から2021年3月31日（3年間）における1時間間隔の残留塩素濃度(mg/L)（以下データセットCと表記する）を使用した。使用したデータセットの特徴（平均値、最大値、最小値）を表-1にまとめる。データセットA、Bは、以下の方法で異常値の除去およびノイズを除去したものを学習・検証に使用した。異常値の除去は、箱ひげ図を用いて、上限=95%点+四分位範囲(IQR)×3、下限=95%点-IQR×3として、極端にトレンドから外れた値を除外した後に、前後の値で線形補間した。データセットCは、以下の方法で異常値の除去およびノイズを除去したものを学習・検証に使用した。異常値の除去は、箱ひげ図を用いて、上限=75%点+四分位範囲(IQR)×3、下限=25%点-IQR×3として、極端にトレンドから外れた値を除外した後に、前後の値で線形補間した。表-1中に除外した異常値数を示す。データセットAは、30分間隔に変換することでデータ量を圧縮した。データノイズを除去するため、前後12時間の24時間移動平均を算出することで、平滑化処理を実施した。

データセットA、Bの平滑化処理前の各

項目の2019年6月における1ヶ月変動を図-3、データセットCの平滑化処理前の各項目の2018年4月における1ヶ月変動を図-4に示す。残留塩素濃度と水温は24時間周期で小刻みに変動しながら値が推移した一方で、電気伝導度は大きな変動が数回発生する変動傾向を示した。水温が高い時に残留塩素濃度が低く、水温の低下に伴って残留塩素濃度が上昇する傾向が観察された。本研究では、任意時刻における浄水場出口の残留塩素濃度と給水地点の残留塩素濃度の差を「残留塩素低減量」として、電気伝導度、水温および残留塩素低減量をLSTMの入出力値に用いた。管路内の滞留時間に時間周期性があると仮定し、滞留時間の周期変動も加味した残留塩素低減量を評価する。電気伝導度および水温は、正規分布として平均0、分散1になるように一般標準化したものを入力値として用いた。

LSTMモデルの概念を図-2に示す。LSTMは、現時刻を t としたとき、時刻 $t-L$ から時刻 t までのデータを使用して、時刻 $t+M$ のデータを予測する。このとき、 L をブロック、 M を予測時間と定義する。本研究では、活性化関数にReLU、ネットワーク重みにAdamを使用した。バッチサイズは32、エポック数は50、ノード数は1024とした。学習関数、エポック数、ノード数を複数検討したが、どの組み合わせでもモデル精度に大幅な変化がなかった（データは非表示）。ブロックとして、12、24、48および168時間（7日間）の4条件を検討した。

データセットを目的に応じて任意の割合でモデル構築用の学習データとモデル精度検証用のテストデータに分割した。本研究

で使用した学習データおよびテストデータを表-1 中にまとめる。モデルの構築・検証は、tensorflow 1.13.1 を用いて、CPU : intel core i9 9900KF、GPU : Nvidia GeForce RTX 2080 Ti を搭載したワークステーションにて計算した。

モデル精度の評価方法

残留塩素低減量予測モデルの精度は、式 1 に示す二乗平均平方根誤差 (RMSE) により評価した。

$$RMSE = \sqrt{\frac{1}{n} \sum (\text{実測値} - \text{予測値})^2} \quad (\text{式 1})$$

残留塩素濃度の連続測定装置の検出限界値 0.05 mg/L を勘案して、±0.025 mg/L を許容誤差範囲とし、この範囲内に予測結果の最大値が収まることをモデル構築にあたっての目標とした。

C. 結果と考察

ブロックサイズがモデル精度に与える影響

データセット A 中、入力を残留塩素低減量、電気伝導度および水温とし、出力を 6 時間後の残留塩素低減量とした際に、ブロックサイズがモデル精度に及ぼす影響を検討する。なお、モデル構築には、表-1 に示したデータセット A の全データのうち、学習データとして 2018 年 7 月 1 日 6:00 から 2020 年 3 月 31 日 17:30 までの 29833 個のデータをモデル構築に、テストデータとして 2020 年 3 月 13 日 18:00 から 2020 年 6 月 30 日 17:30 までの 5255 個のデータを精度の評価に使用した。

図-5 に上和田給水地点および西谷給水地点におけるブロックサイズと RMSE の関係

を示す。図-5 を見ると、ブロックが 12~48 時間において、すべてのプロットが目標精度である実測値±0.025 mg/L 以内に収まっていた。特に、ブロックを 24 時間に設定した際に、最も高い精度が得られた。この結果は、図-3 に示すように、対象とした地域では、水質変動パターンが 24 時間周期であることに起因すると考える。本研究では、以後の検討において、モデル構築時のブロックを 24 時間に設定した。

図-5 中、上和田給水地点と西谷給水地点を比較すると、どの条件も上和田給水地点がより低い予測精度を示した。上和田給水地点は浄水場から比較的近く、残留塩素濃度低減量が低いことから測定誤差が大きくなったものと推測する。

上和田給水地点および西谷給水地点のブロックを 168 時間に設定した際、および西谷給水地点のブロックを 48 時間に設定した際に、モデルの出力が出来なかった。これはパラメータが複雑になったことで、モデルが収束しなかったことが原因と考えられる。

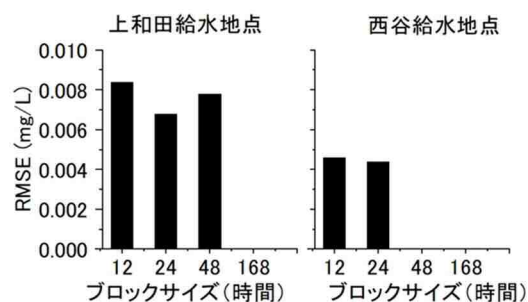


図-5 ブロックサイズと予測誤差の関係

予測期間がモデル精度に与える影響

モデルの出力となる予測期間は、浄水場職員からヒアリングした上で、3、6、12 お

よび 24 時間をそれぞれ検討した。本研究で対象とした給水地点は、平均的な滞留時間が 6 時間程度であることが事前調査で明らかになっている。予測結果に応じて浄水場出口での次亜塩素酸の注入制御を実施するには、6 時間先の予測結果が少なくとも必要となる。

データセット A 中、入力を残留塩素濃度低減量、電気伝導度および水温とし、出力を 3、6、および 12 時間先の残留塩素濃度低減量とした際のモデル精度を図-6 に示す。両方の給水地点共に、予測時間の増加に伴って精度が徐々に悪化する傾向が得られた。予測時間を 12 時間に設定した際に、RMSE は 0.020 mg/L を示したが、実測値と予測値の散布図 (図-7) から、一部のプロットにおいて、実測値 ± 0.025 mg/L から逸脱する点が観察された。予測時間を 6 時間に設定した際 (図-8) に、すべてのプロットが実測値 ± 0.025 mg/L 以内に収まったことから、より高い精度で予測するには、予測時間 6 時間が望ましいことが分かる。また、図-5 と同様に、図-6 中、上和田給水地点は西谷給水地点と比較して、どの条件でもより低い予測精度を示した。以上の結果から、現場で必要とされる 6 時間先の残留塩素濃度低減量を予測できることが示された。

井上ら¹²⁾は、天候、水量、水質および初期残留塩素濃度を入力としたニューラルネットワークにより、予測誤差 0.0806mg/L を達成した。松本ら¹³⁾は、機械学習と化学反応モデルのハイブリッド化により、予測値誤差 0.048mg/L を達成した。本研究で構築した 6 時間先の残留塩素濃度低減量を予測するモデルの予測誤差は 0.008 mg/L であり、

LSTM による予測精度は上述した既存の機械学習モデルと比較してより高いことが示された。

入力項目が予測精度に与える影響

これまで、入力には残留塩素濃度低減量その他、一般的な水質項目である電気伝導度と水温も使用してモデルを構築してきた。続いて、モデル構築に最小限必要となる項目を検討するために、各項目の予測精度に対する影響を分析した。

データセット A を対象として、入力項目中、残留塩素濃度低減量、電気伝導度および水温から、電気伝導度を削除した場合、水温を削除した場合、電気伝導度と水温を削除した場合の 4 条件で 6 時間後の残留塩

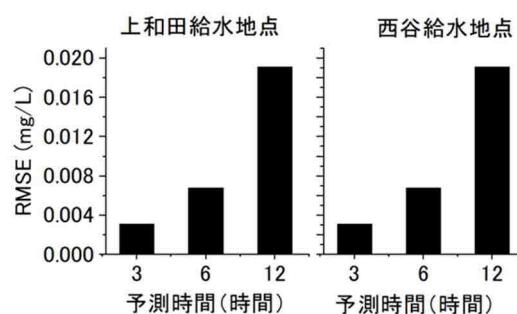


図-6 予測時間と予測誤差の関係

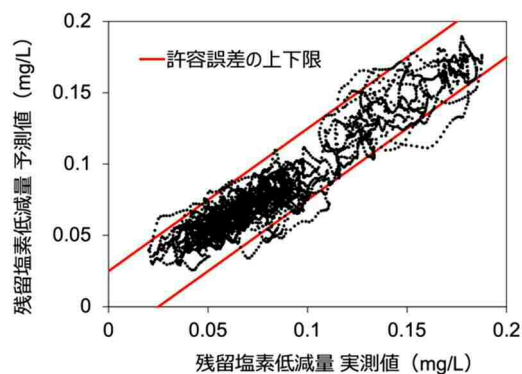


図-7 上和田給水地点における 12 時間後の残留塩素低減量予測モデルの精度

素濃度低減量を予測するモデルを構築し、予測精度を比較した結果を図-9 に示す。

図-9 の通り、入力項目を3項目(残留塩素濃度低減量、水温、電気伝導度)から1項目(残留塩素濃度低減量)に減らした際に、RMSE に大差が見られなかったことから、水温と電気伝導度がモデル精度に及ぼす影響が小さいことが分かる。残留塩素濃度低減量を予測する物理モデルについて検討する既往研究において、水温や電気伝導度を環境因子とする研究^{リ7)}が散見されるが、LSTM により構築したモデルは、これらの環境因子を参照せずに、残留塩素濃度低減量の経時変化のトレンドを捉えることで、将来を予測していると推測される。本研究により、予測モデルの構築にあたって、残留塩素低減量の情報のみで、十分な精度のモデルが構築できることが明らかになった。

データ量が予測精度に与える影響

モデル構築に必要なデータ量(データ蓄積期間)について検討する。データセット B を対象として、入力項目を残留塩素濃度低減量として、6時間後の残留塩素濃度低減量を予測するモデルを構築した。なお、モデル構築には、表-1 に示したデータセット B の全データのうち、学習データとして2014年4月1日00:00から2019年3月31日23:59までのデータをモデル構築に、テストデータとして2019年4月1日00:00から2020年3月31日23:59までのデータを精度の評価に使用した。

モデル構築にあたって、年度毎に使用するデータを12ヶ月分から1ヶ月ずつ減少することで、データ量がモデル精度に及ぼす

影響を検討した。なお、データセット B 中、異常値の出現にあたって、年および月に偏りがないことを確認している。データの減量については、3月から4月にかけて1ヶ月ずつ減らした P1 と4月から3月にかけて1ヶ月ずつ減らした P2 について、それぞれ検討した。2014年から2018年の5年分実施した際の平均値を図-10 および図-11 に示す。

上和田給水地点(図-10) および西谷給水地点(図-11) 共に、3月から前年の4月ま

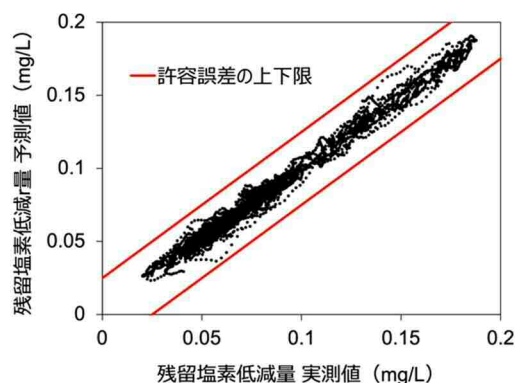


図-8 上和田給水地点における6時間後の残留塩素低減量予測モデルの精度

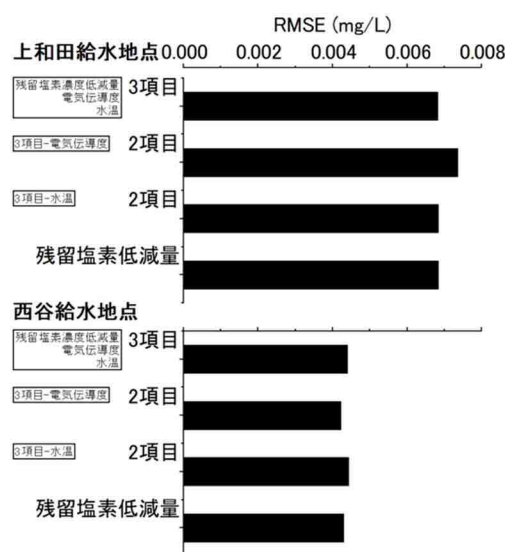


図-9 入力項目が予測精度に及ぼす影響

で1ヶ月ずつ遡って学習データを減少した際(P1)に、4ヶ月分まで同程度の精度を維持した一方で、3ヶ月分になると急激に精度が悪化した。この傾向は、4月から3月まで1ヶ月毎データを減らした際(P2)でも同様の傾向が観察された。これらのことから、本研究で使用したデータについて、最低でも4ヶ月間の学習データが必要となることが示唆された。

西谷給水地点(図-11)において、P1条件下で4ヶ月間(12月~3月)学習した際のモデル精度が他の結果と比較して顕著に低いことがわかる。おそらく12月から3月のデータが他とは異なるトレンドを示したことが原因と考える。これらの結果から、データによってモデル精度が変化することが明らかになったと共に、モデル構築に用いるデータによっては、さらに長期間のデータセットを用いてモデル構築を行う必要性が示された。

学習データの質が予測精度に与える影響

本研究で用いたデータでは、精度の高いモデル構築に4ヶ月間の学習データが必要となることが明らかになった。4ヶ月間の学習データについて、開始月と終了月を変化させてモデル構築することで、精度の悪化を誘発するデータ群を探索した。

1年のデータを4ヶ月毎に区切ることで12パターンの学習データを準備し、モデル構築に用いた。2014年から2018年の5年間について年度毎にモデルを構築し、各パターンにおける精度の平均値を算出した。予測には、入力値、出力値共に残留塩素濃度低減量を用いた。

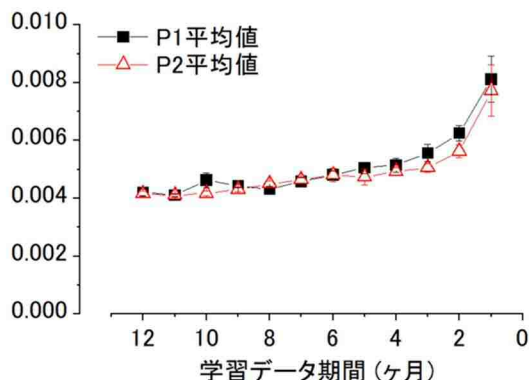


図-10 上和田給水地点における学習データ数が予測精度に及ぼす影響

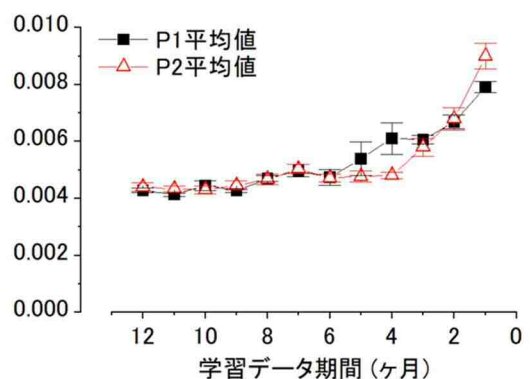


図-11 西谷給水地点における学習データ数が予測精度に及ぼす影響

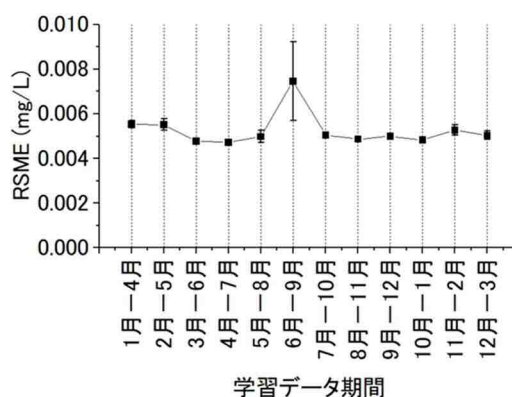


図-12 上和田給水地点における学習データの時期が予測精度に及ぼす影響

上和田給水地点の結果を図-12、西谷給水地点の結果を図-13に示す。4ヶ月間の学習データであっても、用いたデータ期間によって精度が異なることが明らかになった。本研究では、上和田給水地点、西谷給水地点共に、4月から7月の4ヶ月間のデータを学習データとして使用した際に、最も高い精度が得られた。一方で、上和田給水地点では6月から9月にかけて夏期間のデータ、西谷給水地点では11月から2月にかけて冬期間のデータを学習データとして使用した際に、モデル精度が低くなることが明らかになった。

精度に影響する学習データ特性を検討するため、特に精度が高いモデルおよび低いモデル構築に用いた学習データについて、最大値、最小値、中央値、四分位値をそれぞれ算出した結果を図-14に示す。

テストデータとして、本研究では2019年のデータを用いたが、特に精度が高かった4月から7月のデータの分布とテストデータの中央値、四分位値がほぼ一致していることがわかる。一方で、精度が低かった学習データは、中央値、四分位値がテストデータと大きくずれていた。これらの結果から、モデルの適用にあたって、学習データとテストデータの中央値、四分位値を比較することで、モデルの適否を迅速に評価できる可能性が示唆された。

モデル構築アルゴリズムの汎用性評価

LSTM モデル構築アルゴリズムの汎用性を検証するため、沖縄県企業局より取得したデータセットCを用いて、ブロックサイズを24時間、予測時間を6時間として同様

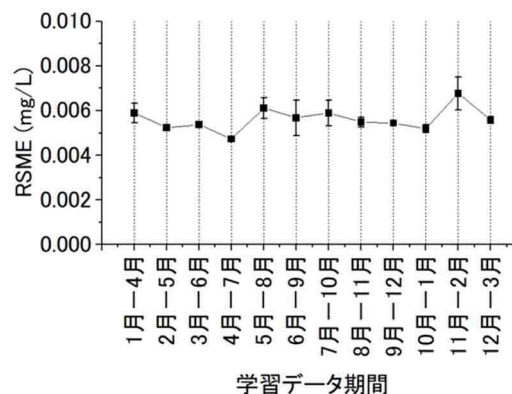


図-13 西谷給水地点における学習データの時期が予測精度に及ぼす影響

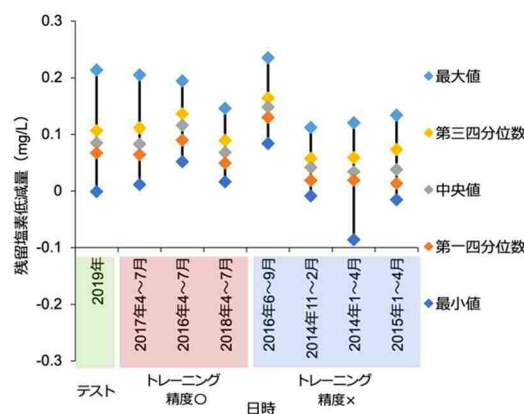


図-14 高精度および低精度モデルの学習データの特性

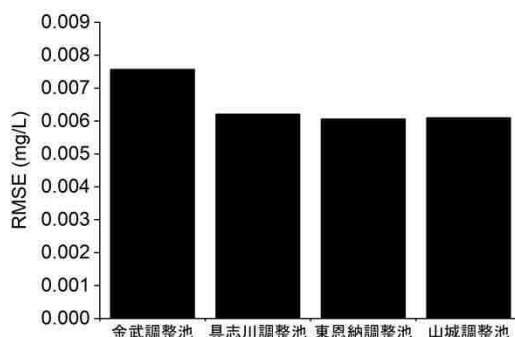


図-15 各調整池における予測誤差

に予測モデルを構築した。

図-15 にそれぞれの調整池地点における RMSE を示す。図-15 を見ると、すべての調整池において目標精度である実測値 ± 0.025 mg/L 以内に収まっていた。図-15 中、金武調整池地点は他の条件より低い予測精度を示した。金武調整池は石川浄水場と久志浄水場の間に位置しており、流量の変化が大きく、予測精度が低くなったものと推測する。

続いて、モデル構築に必要なとなるデータ量（データ蓄積期間）とデータの質が予測精度に与える影響について検討する。データセット C のうち、学習データとして 2019 年 4 月 1 日 00:00 から 2020 年 3 月 31 日 23:59 までのデータをモデル構築に、テストデータとして 2020 年 4 月 1 日 00:00 から 2021 年 3 月 31 日 23:59 までのデータを精度の評価に使用した。モデル構築にあたって、使用するデータを 12 ヶ月分から 1 ヶ月ずつ減少することで、データ量がモデル精度に及ぼす影響を検討した。データの減量については、3 月から 4 月にかけて 1 ヶ月ずつ減らした P1 と 4 月から 3 月にかけて 1 ヶ月ずつ減らした P2 について、それぞれ検討した。P1 の結果を図-16、P2 の結果を図-18 に示す。また、図-16 中、RMSE の最大値を 0.055 mg/L とした結果を図-17 に示す。

図-16、18 より、具志川調整池、および金武調整池において、1 ヶ月ずつ学習データを減少した際（P1、および P2）に、4 か月分まで同程度の精度を維持した一方で、3 ヶ月分になると急激に精度が悪化した。また、山城調整池、および東恩納調整池は 1 ヶ月分になると急激に精度が悪化した。具志川

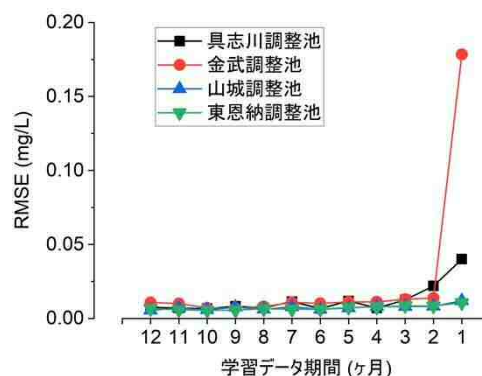


図-16 各調整池における学習データ数が予測精度に及ぼす影響（P1）

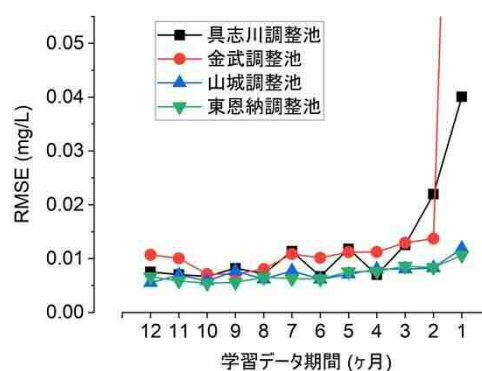


図-17 各調整池における学習データ数が予測精度に及ぼす影響（P1・縦軸拡大）

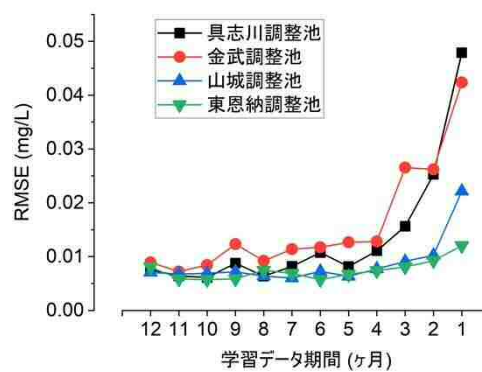


図-18 各調整池における学習データ数が予測精度に及ぼす影響（P2）

調整池、および金武調整池と、山城調整池、および東恩納調整池は異なる傾向が得られた。これらの結果から、神奈川県内広域水道企業団と同様に、データによってモデル精度が変化することが明らかになった。

D. 結論

本研究では、時系列の濃度変化の傾向を長短期記憶ネットワーク (LSTM) アルゴリズムにより学習することで、数時間先の残留塩素濃度を推測できる予測モデルの構築を目的とした。また、他の自治体での適用可能性を示すためにモデルの構築に必要最小限のデータ量を検討した。

モデル構築にあたって最適なブロックは 24 時間であり、誤差目標値 ± 0.025 以下に収めるには、予測時間を 6 時間以下にする必要があることが判明した。

モデル構築に必要最小限のデータ量は 4 月～7 月の 4 ヶ月間の残留塩素濃度低減量であることがわかった。この期間のデータを学習データに用いたとき、高い精度のモデルが構築できた。以上より残留塩素濃度の 1 時間間隔の時系列データが 4 ヶ月分準備できれば、6 時間先の残留塩素濃度を LSTM により予測できることがわかった。

E. 研究発表

1. 論文発表

山村 寛, 石井 崇晃, 小野 一樹, 市川 学, 清塚 雅彦, 長短期記憶ネットワークによる給配水系統における残留塩素濃度低減量の予測. 環境工学研究論文集, 77 巻 7 号 p. III_293-III_301, 11 月, 2021 年.

2. 学会発表

○山村 寛, 石井 崇晃, 小野 一樹, 市川 学, 清塚 雅彦. 長短期記憶ネットワークによる給配水系統における残留塩素濃度低減量の予測. 土木学会環境工学委員会 第 58 回環境工学研究フォーラム, A31, オンライン開催, 2021/11/16-18.

F. 知的所有権の取得状況

1. 特許取得

なし

2. 実用新案登録

なし

3. その他

なし

参考文献

- 1). 公益財団法人水道技術研究センター: 管路内残留塩素濃度管理マニュアル, 1999.
- 2). He G., Zhang T., Zheng F., Li C., Zhang Q., Dong F., Huang Y.: Reaction of fleroxacin with chlorine dioxide in drinking water distribution systems: Kinetics, transformation mechanisms and toxicity evaluations, Chemical Engineering Journal, Vol.374, pp. 1191-1203, 2019.
- 3). Clark R. M., Goodrich J. A., Wymer L. J.: Effect of the distribution system on drinking-water quality, Journal of Water Supply: Research and Technology—AQUA, Vol.42, pp. 30-38, 1993.
- 4). Hrudey S. E.: Chlorination disinfection by-products, public health risk tradeoffs and me, Water Research, Vol.43, pp. 2057-2092, 2009.

- 5). 厚生労働省健康局：新水道 ビジョン ,https://www.mhlw.go.jp/seisakunitsuite/bunya/topics/bukyoku/kenkou/suido/newvision/1_0_suidou_newvision.htm.
(2021年5月10日閲覧)
- 6). 稲員とよの,小泉明,荒井康裕,佐藤親房,尾崎勝,渡辺映一: 送水管路におけるフィールド調査に基づく残留塩素減少モデルに関する研究, 環境工学研究論文集, Vol.45, pp. 605-612, 2008.
- 7). 荒井康裕,稲員とよの,小泉明,赤池陽介,佐藤親房,尾崎勝: ANN を用いた排水管網における残留塩素濃度推定モデルと低減化シミュレーション, 環境工学研究論文集, Vol.45, pp. 595-603, 2008.
- 8). 張裕喆, 遠山忠, Randall A., 高見澤一裕, 菊池慎太郎: 水道管材料が微生物膜形成に及ぼす影響—特に米国における研究例—, 環境浄化技術, Vol.36, pp. 55-58, 2007.
- 9). Abokifa A. A., Yang Y. J., Lo C. S., Biswas P.: Water quality modeling in the dead end sections of drinking water distribution networks, Water Res, Vol.89, pp. 107-17, 2016.
- 10). Bowden G. J., Nixon J. B., Dandy G. C., Maier H. R., Holmes M.: Forecasting chlorine residuals in a water distribution system using a general regression neural network, Mathematical and Computer Modelling, Vol.44, pp. 469-484, 2006.
- 11). Le X. H., Ho H. V., Lee G., Jung S.: Application of Long Short-Term Memory (LSTM) Neural Network for Flood Forecasting, Water, Vol.11, pp. 2019.
- 12). 井上正人,岡隆光,菅原通雅,塚田司郎,萩岡光治,前原俊信, 森田めぐみ, 守屋節男, 渡辺義信: 残留塩素濃度を予測するニューラルネットワークモデル, 社会情報学研究, Vol.6, pp. 149-168, 2000.
- 13). 松本隼, 山原裕之, 横川勝也, 毛受卓, 黒川太, 横山雄: 機械学習と化学反応モデルのハイブリッドによる残留塩素濃度予測, 環境システム計測制御学会, Vol.24 巻, pp. 111-151, 2019.