

〈研究発表〉

ランダムフォレストを用いた原水水質のリアルタイム推定

松本 隼¹⁾, 毛受 卓²⁾, 小原 卓巳³⁾, 横山 雄⁴⁾

¹⁾ 東芝インフラシステムズ(株) インフラシステム技術開発センター
(〒183-8511 東京都府中市東芝町1番地 E-mail: jun8.matsumoto@toshiba.co.jp)

²⁾ 東芝インフラシステムズ(株) インフラシステム技術開発センター
(〒183-8511 東京都府中市東芝町1番地 E-mail: takashi.menju@toshiba.co.jp)

³⁾ 東芝インフラシステムズ(株) インフラシステム技術開発センター
(〒183-8511 東京都府中市東芝町1番地 E-mail: takumi.obara@toshiba.co.jp)

⁴⁾ 東芝インフラシステムズ(株) 社会システム事業部
(〒212-8585 神奈川県川崎市幸区堀川町72番地34 E-mail: suguru.yokoyama@toshiba.co.jp)

概要

浄水場における塩素注入の自動化・最適化に向け、高精度な残留塩素濃度予測の開発を進めている。塩素注入率に対する残留塩素濃度を化学反応に基づき推定する場合、入力情報として原水中の各反応物質の濃度が必要となる。浄水場におけるオンライン測定項目はその一部に限定されるため、筆者らは他のオンライン測定項目から原水水質を推定する手法を提案した。本稿では、ランダムフォレストを用いて推定を実施し、前回報告した線形計画に基づく手法と比較、その有効性について評価する。

キーワード：残留塩素、水質推定、ランダムフォレスト、AI

原稿受付 2022.7.11

EICA: 27(2・3) 85-89

1. はじめに

浄水場における塩素注入運転の自動化・最適化に向けた要素技術として、浄水プロセスにおける残留塩素の消費・分解を予測するためのモデル構築を進めている。筆者らは、各処理の化学反応式など明示的に定式化したホワイトボックスモデルと、オンラインの測定データとオフラインの手分析データを併用したデータセットを用いた機械学習に基づくブラックボックスモデルのハイブリッド化による残留塩素濃度予測手法を提案、評価した¹⁾。化学反応式への入力情報として用いる水質項目のうち、オンラインで計測されていない採水分析による水質データを、線形計画法を用いてリアルタイム推定する手法を提案し原水水質推定モデルを構築、パラメータ調整最適化により定常時に、原水Fe, Mn, TOC濃度をそれぞれ16.9%, 15.8%, 5.1%の絶対誤差平均で推定可能なことを示した²⁾。ただし本手法は、現在の沈澱池出口残留塩素濃度と滞留時間分過去の前次亜塩素酸注入率を用いて原水水質値を推定するため、原水水質値の推定値が得られるのが滞留時間分遅れる、という課題があった。また、本モデルは手法の特性上、高濁度原水流入などの非定常時への適用が困難であり、シミュレーション評価においても非定常時の推定性能が著しく悪化することを確認している。そこで、既報と異なる手法として、AI手法の一つラ

ンダムフォレストを推定対象時刻と同時刻のオンラインデータを入力として適用し、時間遅れがなくかつ非定常データに対応可能な原水水質推定モデルを構築する。また、シミュレーション評価を実施し、非定常時に適用する際に有効なモデル構築手法を検討した結果を報告する。

2. ランダムフォレストを用いた手分析データ推定モデル

長野市犀川浄水場を対象機場として、本浄水場内に設置されたセンサから取得されるオンラインデータ、年1回以上の実施が想定される手作業によるオフライン測定で得られる手分析データを元に、未測定期間における手分析データの推定モデルをランダムフォレストにより構築した。**Table 1**に本報告で構築する手分析データ推定モデルの入出力項目の一覧を示す。オンラインデータは犀川浄水場に設置された計測機器などから取得されるデータ、オフラインデータは同浄水場において3回にわたり採水した約24時間分の採水分析データを用いる。モデル入力は出力と同一時刻の値のみ用いる。

採水分析データは各セット12または24点のみであり学習用および検証用として点数が少ない。そこで3次エルミート補間手法³⁾を用いて、5分間隔データに

Table 1 Inputs and Outputs of Estimation Model of Intake Water Quality

項目	単位	プロセス	データ取得	入出力
原水流量	m ³ /h	原水	オンライン	入力
原水 pH	-	原水	オンライン	入力
原水濁度	mg/L	原水	オンライン	入力
原水アルカリ度	-	原水	オンライン	入力
原水水温	℃	原水	オンライン	入力
原水アンモニア濃度	mg/L	原水	オンライン	入力
原水導電率	μS/cm	原水	オンライン	入力
PAC 注入率	mg/L	沈澱池	オンライン	入力
前次亜塩素酸注入率	mg/L	沈澱池	オンライン	入力
沈澱水 pH	-	沈澱池	オンライン	入力
沈澱水濁度	mg/L	沈澱池	オンライン	入力
沈澱水残留塩素濃度	mg/L	沈澱池	オンライン	入力
気温 (犀川気温)	℃	気象	オンライン	入力
雨量 (犀川雨量)	mm/h	気象	オンライン	入力
原水 Mn 濃度	mg/L	原水	オフライン	出力
原水 Fe 濃度	mg/L	原水	オフライン	出力
原水 TOC 濃度	mg/L	原水	オフライン	出力

補間し、データ点数を疑似的にかさ増しする。本手法は線形補間やスプライン補間と異なり、サンプル値前後での連続性が保たれ、かつオーバーシュートと呼ばれるサンプル値前後に発生する振動が発生しにくい特性を持つため補間手法として適用した。

3. シミュレーション評価

3.1 評価手法

Table 2 に示す学習およびテストデータ割合で手分析データ推定モデルを構築し、各水質項目の絶対誤差を評価する。「データセット」の内、Set1 (2019/8/26 ~ 27 採水)、Set2 (2019/9/10 ~ 11 採水) を「定常」データ、Set3 (2019/10/12 ~ 13 採水) を「非定常」データとして用いる。「補間前」のデータは、実際に測定した値であり、定常、非定常それぞれ 36 点、24 点となる。「補間後」のデータは、前節で述べた 3 次エルミート補間による 5 分間隔での補間データを含み、それぞれ 542 点、275 点となる。「データセット」項目値はデータの使用割合を示す。入力項目はモデル入力に用いる項目数を示し、「全部」は **Table 1** に示す 14 の入力項目を全てモデル入力として用い、「一部」は原水および気象に係る 9 の入力項目のみモデル入力として用いる。

シミュレーション 1-1, 2 は、入力項目として全部あるいは一部を使用した場合の推定精度への影響を検証した (シミュレーション 2 以降は、その検証結果として「一部」のみに限定して実施)。シミュレーション 2-1, 2 は、1-2 との比較によりデータ補間有無によるモデル精度への影響を検証した。シミュレーション 3-1 は、定常データをテストデータとすることにより、

Table 2 Simulation specifications.

シミュレーション番号		使用データセット				入力項目	
		1, 2 (定常)		3 (非定常)		全部	一部
		補間前	補間後	補間前	補間後		
1-1	学習			75%		○	
	テスト			25%			
1-2	学習			75%			○
	テスト			25%			
2-1	学習				75%		○
	テスト				25%		
2-2	学習				75%*		○
	テスト				25%*		
3-1	学習				100%		○
	テスト		100%				
3-2	学習		75%		75%		○
	テスト		25%		25%		

*2-2 は、最初の 75% 連続区間を学習データ、残り 25% 連続区間をテストデータ。2-2 以外はランダム。

非定常モデルの他データ区間への適用可能性を検証した。シミュレーション 3-2 は、シミュレーション 3-1 の比較検討用として、定常および非定常を含む全データについて学習データまたはテストデータに分離して検証した。

各シミュレーション 100 回試行し、試行毎に異なる乱数を元にしたランダムフォレストを用いてモデルを生成する。

3.2 評価結果

入力項目について

Fig. 1 に全シミュレーション結果を示す。各シミュレーション 100 回の試行結果を実測値と共に示している。**Fig. 1**-(a), (b) に示すシミュレーション 1-1, 2 結果より、いずれの水質項目についても実測値と同様の増減トレンドを示す推定結果となった。シミュレーション 1-1 と 1-2 を比較すると、誤差平均については前者より後者の方が最大で 9% (原水 Mn 濃度) の改善がみられ、入力項目の違いによる推定精度への影響と考えられる。既報^{1,2)}の手分析データ推定モデルにおいて沈澱池残留塩素値を用いているが、本報告では推定対象時刻と同時刻のデータのみ入力の対象としているため後プロセスにあたる本説明因子は不要であり、これらを除外することにより推定モデルの学習精度向上につながったと考えられる。上記考察より、入力項目として沈澱池に係る項目については除外し、原水および気象に係る項目のみを用いてシミュレーション 2 および 3 を実施した。

補間による精度改善について

Table 3 に示すシミュレーション 2-1 の結果より、絶対誤差平均は、原水 Fe 濃度で 0.010 mg/L、原水 Mn 濃度で 0.0014 mg/L、原水 TOC 濃度で 0.0257 mg/L

であり、それぞれ実測値平均を基準に正規化し、4.2%, 3.7%, 1.8%の推定誤差であった。学習データ点数以外同条件であったシミュレーション1-2と比較し、推定精度の平均について原水 Fe 濃度が 88.1%, 原水 Mn 濃度が 88.6%, 原水 TOC 濃度が 87.7% 改善して

いる。データ点数を補間によりかさ増しすることで推定モデルの精度向上が可能であることを示している。**Fig. 1-(c)** に示すトレンドをみると、各水質項目について値が低く水質が急変する前の時間帯である 10/12 17:00 頃までの推定精度が高く、それ以降の各水質項

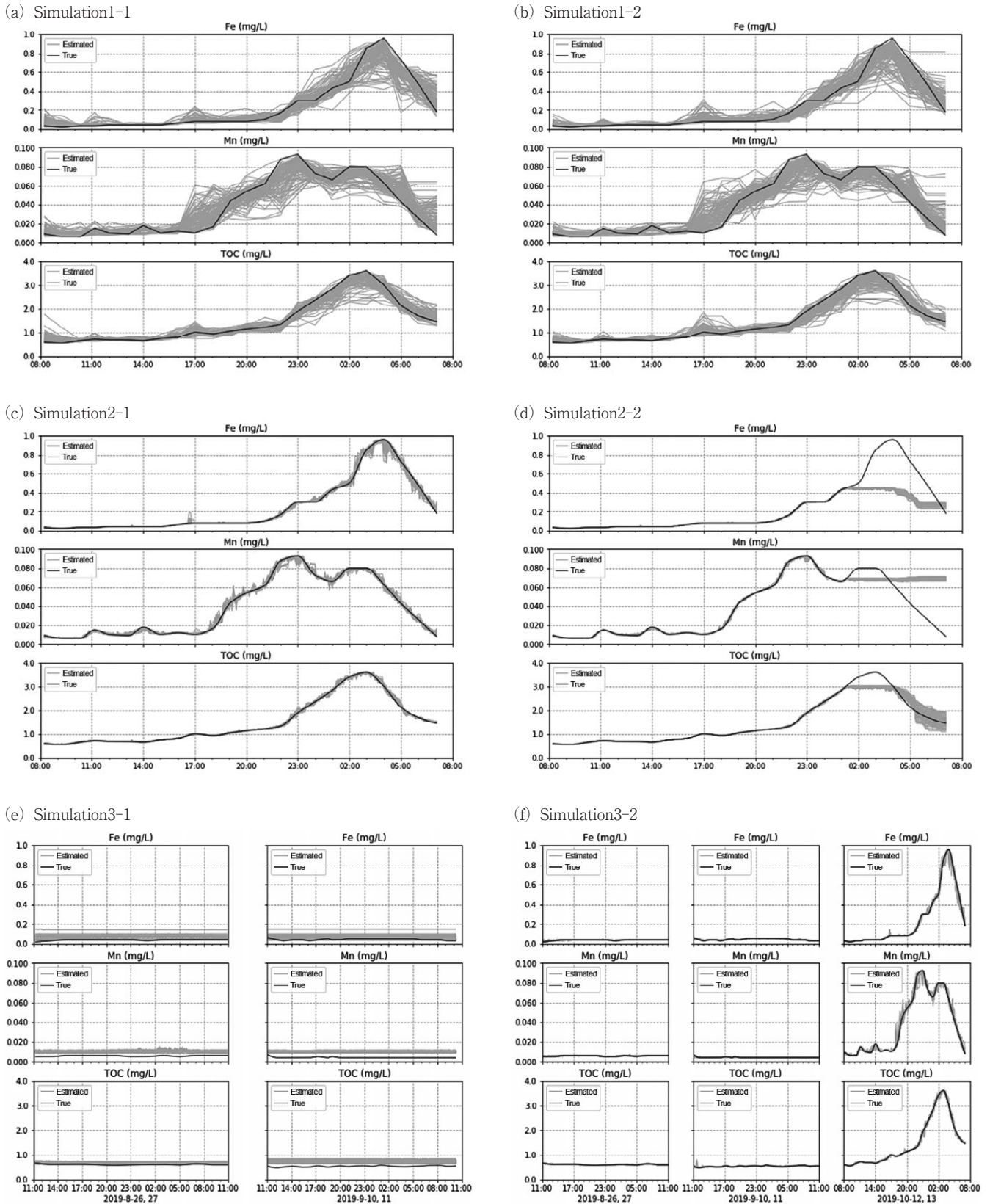


Fig. 1 Simulation results.

Table 3 Summary of simulation results.

シミュレーション番号	原水 Fe 濃度	原水 Mn 濃度	原水 TOC 濃度	
1-1	0.0856	0.0134	0.2106	
1-2	0.0839	0.0122	0.2083	
2-1	0.0100	0.0014	0.0257	
2-2	0.2775	0.0226	0.2937	
3-1	0.0116	0.0053	0.1266	
3-2	定常	0.0014	0.0002	0.0061
	非定常	0.0095	0.0015	0.0263

単位は全て mg/L

目値が高くなる時間帯において推定精度が低い。また、従来手法で発生した数時間の滞留時間分の時間遅れが発生していないことが確認できる。

本結果より、データサンプル数が少ない場合であっても、データ補間によりソフトセンシングのモデル構築およびその推定精度向上が可能であることがわかった。採水分析には多くの手間や費用を要するため、適切なサンプリング間隔とすることでそれらのコスト削減を図りつつモデル構築を実施することができる。

非定常データの複数サンプルの必要性

本報告における非定常データは1つのデータ区間のみであり、本データを用いて学習した推定モデルが他の非定常時についても適用可能かどうか検討ができない。そこで本報告では、本データ区間を前部と後部に分け疑似的に2つの非定常データとしてシミュレーション2-2を、定常時を非定常時と異なる状況の一つと捉え疑似的な非定常時データとしてシミュレーション3を実施した。

Fig. 1-(d) に示すシミュレーション2-2のトレンドをみると、学習データとして用いた75%区間末尾となる10/13 1:00頃を過ぎてからの推定精度が急激に悪化している。原水 Fe 濃度および原水 TOC 濃度は、それぞれ10/13 4:00, 3:00頃に迎えるピークに追従できていない。また、原水 Mn 濃度はテストデータ区間についてほぼ一定値となった。原水 Mn 濃度のピークは10/12 23:00頃であり、出力値として最小と最大の値がいずれも学習データ区間に含まれているにもかかわらず、テストデータ区間については正しく推定が行えなかった。

また、Fig. 1-(c) および (e) に示すシミュレーション2-1 および 3-1 の結果を比較すると、10/12 17:00頃までの比較的値の小さいデータ区間の推定精度について、シミュレーション3-1の方が悪化している。上記より、10/12 17:00頃までのデータ区間も定常ではなく非定常として分類すべき異なる状況下であり、異なる状況下のデータ区間に対しては推定モデルの精度が悪化する。その確認として、Fig. 1-(f) に示すシ

ミュレーション3-2の結果より、定常、非定常いずれも学習データとすることでいずれにも対応可能な推定モデルの作成が可能であることがわかる。また、シミュレーション3-2（非定常データ区間）と2-1の各水質項目の推定誤差平均を比較すると、前者の方が精度が高く、定常データであっても学習データに含むことで非定常データ区間の推定精度が改善可能であることを示している。

上記2つの考察より、本報告における1つの非定常データ区間を用いて学習した推定モデルを他の非定常時について適用することは困難であると想定される。複数の非定常データを学習させることにより、汎用性の向上が見込まれる。また、定常・非定常に関わらずデータ区間が増えることにより推定精度改善が見込まれる。

4. ま と め

本報告では、手分析データ推定について、従来手法が適用できなかった非定常時の推定を可能とするため、ランダムフォレストを用いたモデル構築を実施した。シミュレーション評価によりその適用可能性について検証した結果、下記の知見が得られた。

- ・推定対象時刻と同時刻のオンラインデータを入力とするランダムフォレストにより、時間遅れなく推定可能である。その推定精度は、原水 Fe 濃度で4.2%、原水 Mn 濃度で3.7%、原水 TOC 濃度で1.8%の絶対誤差平均であり、前報で検討した方法よりも精度が改善した。
- ・手分析データ推定モデルの入力項目としては原水関連項目および気象項目のみで推定可能であった。
- ・データ補間によりソフトセンシングの推定精度が約88%向上した。データ補間を前提として採水サンプリング間隔を広くすることにより、人的・金銭的コストを抑えることが可能である。
- ・1つの非定常データ区間を用いて学習した推定モデルを他の非定常時についても適用することは困難である。また、定常データであっても学習データに含むことで非定常データ区間の推定精度が改善可能である。

謝 辞

本研究は、JWRC 研究課題公表による実証研究 (A-IDEA) のテーマ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として採択された、「AIを活用した最適化技術（塩素注入の最適化）の開発」プロジェクトの成果です。本研究の実証先として長野市上下水道局様にご協力いただき、浄水場内の計測データならびに検証用水質データの採取にもご協力いただきまし

た。長野市上下水道局様に、この場をお借りして御礼申し上げます。

参考文献

- 1) 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誌, Vol.24, No.2/3, pp.40-47 (2019)
- 2) 松本隼, 山原裕之, 横川勝也, 毛受卓, 黒川太, 横山雄: 原水水質のリアルタイム推定, 環境システム計測制御学会誌, Vol.25, No.2/3, pp.33-36 (2020)
- 3) Fritsch, F. N. and R. E. Carlson: Monotone Piecewise Cubic Interpolation, SIAM Journal on Numerical Analysis, Vol. 17, 1980, pp. 238-246.