

## 〈ノート〉

## 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

松本 隼<sup>1)</sup>, 山原 裕之<sup>1)</sup>, 横川 勝也<sup>1)</sup>, 毛受 卓<sup>1)</sup>  
黒川 太<sup>1)</sup>, 横山 雄<sup>2)</sup>

<sup>1)</sup> 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,  
hiroyuki.yamahara@toshiba.co.jp, katsuya.yokokawa@toshiba.co.jp,  
takashi.menju@toshiba.co.jp, futoshi.kurokawa@toshiba.co.jp)

<sup>2)</sup> 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運転の自動化・最適化に向け、高精度な残留塩素濃度予測が求められている。複数の塩素消費物質の化学反応に基づく残留塩素濃度予測モデル構築は従来実施されておらず、水質値の手作業測定に伴うリアルタイム性の欠如や、未知の影響因子や機特性を考慮できないことが課題である。本論文では、機械学習とのハイブリッド化により上記欠点を補うモデル構築を実施し、シミュレーション評価により、浄水残留塩素濃度実測値平均0.440 mg/Lに対し予測値絶対誤差平均0.048 mg/Lを達成した。

キーワード：残留塩素, 予測, 機械学習, ハイブリッド

原稿受付 2019.6.10 原稿受理 2019.8.22

EICA: 24(2・3) 40-47

## 1. はじめに

国内の上下水道事業における、運用や維持管理の広域化や共同化、維持管理の効率化というニーズに応えるため、プラント運転の自動化・最適化が課題である。特に塩素注入運転は機特性や季節や天候などによる原水水質特性の変化に応じて、運転員が過去の知見や独自のノウハウに基づき管理目標値以上かつ注入過多とならないよう細かく微調整し注入率を決定している場合が多く自動化・最適化実現のニーズがある。本論文では、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、高精度な残留塩素濃度予測手法について着目する。

予測モデル作成にあたっては一般的に、ホワイトボックスモデル、ブラックボックスモデルに大きく分類される。ホワイトボックスモデルは各処理の反応式など明示的に定式化したモデルであり、プラント定常時だけでなく降雨やダム放流に伴う高濁水の流入など非定常時を含むモデル入力情報に対しても、化学反応に基づき予測値出力が可能である。ただし、モデル入力に必要な水質データの一部が手作業によるオフライン測定により得られるデータ（以下、手分析データ）である場合、リアルタイムに予測できない。また、機特性や外乱など、化学反応モデル上に実装されていない未知の因子を考慮することが困難である。一方で、ブラックボックスモデルは、入力値・出力値の実サン

プルデータセットを用いた機械学習モデルであり、オンラインで得られるデータのみを用いて予測結果をリアルタイムに計算することが可能である。また、内部解析をせずに予測値出力が可能であり、未知の特性に起因する予測結果への影響についても対応可能である。ただし、モデルの過学習や学習データセットの外挿範囲のデータ入力時の予測精度悪化が課題である。

ホワイトボックスモデルの従来研究として文献1), 2) が挙げられる。これらは鉄やマンガンなど各塩素消費物質について論じているが、全物質が混在する中での塩素消費について言及していない。そこでこれらを参考に数値解析に化学反応過程を実施することで複数種類の物質が混在する環境下での塩素消費について定義する。また、ブラックボックスモデルについて、浄水場への機械学習適用事例として文献3), 4) が挙げられるが、いずれも運転計画に関するもので処理水質に関する事例とは異なる。

本論文では、浄水場プラントにおける混和池から浄水池までのプロセスを対象に、混和池入口での前次亜塩素酸注入に対する浄水池入口での残留塩素濃度（ただし、後次亜塩素酸注入前、中次亜塩素酸は注入無し）を実施する。非定常時も予測可能なホワイトボックスモデルを化学反応モデルとしてベースに、未知の機特性や外乱に強いブラックボックスモデルを誤差補正モデルに適用することで、両者の優位点を備えるハイブリッドモデルによる残留塩素濃度予測手法を提案、本モデルの有効

性についてシミュレーション評価した結果を報告する。

## 2. ハイブリッドモデル

### 2.1 全体構成

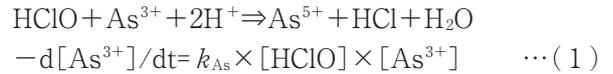
Fig. 1 に提案する残留塩素濃度予測モデルの全体構成、Table 1 に本論文で用いる変数・関数記法一覧を示す。化学反応モデルをベースに、手分析データ推定モデルによりオフラインでしか入手できない手分析水質項目をリアルタイム化し、誤差補正モデルにより未知の因子に対する影響を考慮する。

### 2.2 化学反応モデル

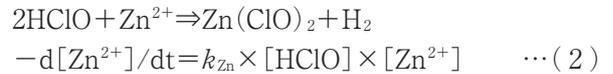
化学反応モデルは、次亜塩素酸 (HClO) に係る化学反応式に基づき、水質変化を数値解析手法により推定する演算モデルである。本論文では、塩素消費への影響が大きく考慮すべき物質として、以下の化学反応式および反応速度定数に基づく各物質の濃度変化を定

義する。 $[x]_t$ は時刻  $t$  における物質  $x$  のモル濃度を示す。 $[x]$  は物質  $x$  のモル濃度を、 $k_x$ は物質  $x$  に関する反応速度定数示す。

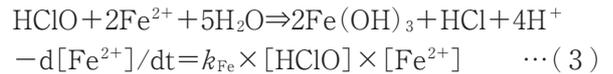
・ヒ素(As)酸化反応



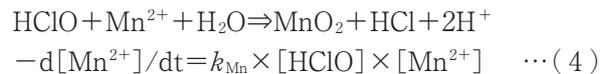
・亜鉛(Zn)酸化反応



・鉄(Fe)酸化反応



・マンガン(Mn)酸化反応



・亜硝酸( $\text{NO}_2^-$ )酸化反応

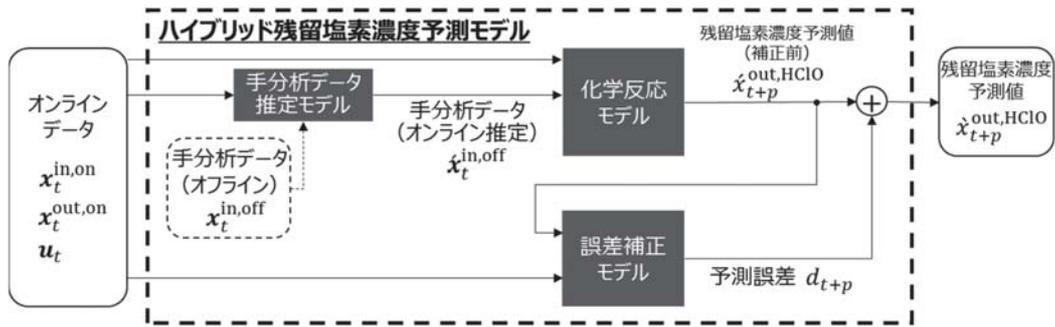


Fig. 1 Residual chlorine prediction model

Table 1 Notations

変数名	種類	説明
$\mathbf{x}_t^{\text{in}} = [\mathbf{x}_t^{\text{in,on}}, \mathbf{x}_t^{\text{in,off}}]^T$	ベクトル	時刻 $t$ の流入水質
$\mathbf{x}_t^{\text{out}} = [\mathbf{x}_t^{\text{out,on}}, \mathbf{x}_t^{\text{out,off}}]^T$	ベクトル	時刻 $t$ の浄水水質
$\mathbf{x}_t^{\cdot,\text{on}}$	ベクトル	時刻 $t$ のオンライン測定水質
$\mathbf{x}_t^{\cdot,\text{off}}$	ベクトル	時刻 $t$ のオフライン測定水質
$\mathbf{u}_t$	ベクトル	時刻 $t$ の流入水質および浄水水質以外のオンライン測定項目
$\hat{\mathbf{x}}_t^{\text{in,off}}$	ベクトル	時刻 $t$ の予測流入水質 (手分析データ)
$\hat{\mathbf{x}}_t^{\text{out}} = [\dots, \hat{x}_t^{\text{out,HClO}}, \dots]^T$	ベクトル	時刻 $t$ の補正前予測浄水水質 (次亜塩素酸濃度を含む)
$\hat{x}_t^{\text{out,HClO}}$	スカラー	時刻 $t$ の補正前浄水残留塩素濃度
$\hat{x}_t^{\text{out,HClO}}$	スカラー	時刻 $t$ の補正後浄水残留塩素濃度
$\mathbf{d}_t^{\text{out},\cdot}$	ベクトル	時刻 $t$ の浄水水質予測誤差
$f_p$	関数	化学反応モデルによる $p$ 時間先水質変化関数
$\mathbf{A}_p$	行列	$p$ 時間先水質変化関数の線形近似モデルにおける 1 次係数
$\mathbf{B}_p$	ベクトル	$p$ 時間先水質変化関数の線形近似モデルにおける 0 次係数
$l$	スカラー	手分析データ推定モデル内の過去参照時間
$m$	スカラー	オンライン水質項目数
$n$	スカラー	オフライン水質項目数
$p$	スカラー	予測先時間
$k_i$	スカラー	化学反応モデル上の水質 $i$ に関する反応速度定数

- ・アンモニア (NH<sub>3</sub>) 酸化反応  
 $\text{HClO} + \text{NH}_3 \Rightarrow \text{NH}_2\text{Cl} + \text{H}_2\text{O}$   
 $-\text{d}[\text{NH}_3]/\text{dt} = k_{\text{NH}_3} \times [\text{HClO}] \times [\text{NH}_3]$  ... (6)
- ・クロラミン (NH<sub>2</sub>Cl) 酸化反応  
 $\text{HClO} + \text{NH}_2\text{Cl} \Rightarrow \text{NHCl}_2 + \text{H}_2\text{O}$   
 $-\text{d}[\text{NH}_2\text{Cl}]/\text{dt} = k_{\text{NH}_2\text{Cl}} \times [\text{HClO}] \times [\text{NH}_2\text{Cl}]$  ... (7)
- ・ジクロラミン (NHCl<sub>2</sub>) 酸化反応  
 $\text{HClO} + \text{NHCl}_2 \Rightarrow \text{NCl}_3 + \text{H}_2\text{O}$   
 $-\text{d}[\text{NHCl}_2]/\text{dt} = k_{\text{NHCl}_2} \times [\text{HClO}] \times [\text{NHCl}_2]$  ... (8)
- ・TOC 酸化反応  
 $\text{HClO} + \langle \text{TOC} \rangle \Rightarrow \langle \text{TOC} \rangle\text{-Cl} + \text{OH}^-$   
 $-\text{d}[\langle \text{TOC} \rangle]/\text{dt} = k_{\text{TOC}} \times [\text{HClO}] \times [\langle \text{TOC} \rangle]$  ... (9)

式 (1) から式 (9) に示される化学反応に基づき、初期濃度の入力に対する  $p$  時間後の水質濃度を数値解析により出力する関数を水質変化関数  $f_p$  として定義する。

Fig. 2 に水質変化関数  $f_p$  を用いた浄水水質推定の概要を示す。本来、浄水場は複数の槽で構成され、薬品の注入や水の流れは連続的に処理されるが、本モデルでは 1 時間単位の流入水についてそれぞれが独立したブロックとみなし前後時間のブロック間で混和することがないものとして処理する。

すなわち、時刻  $t$  から  $p$  時間先浄水水質予測値 (補正前)  $\hat{x}_{t+p}^{\text{out}}$  は、時刻  $t$  の流入水質  $x_t^{\text{in}}$  および水質変化関数  $f_p$  を用いて、

$$\hat{x}_{t+p}^{\text{out}} = f_p(x_t^{\text{in}}) \quad \dots (10)$$

により得られ、 $x_{t+1}^{\text{in}}$  や  $x_{t-1}^{\text{in}}$  による影響を受けないものとする。

### 2.3 手分析データ推定モデル

Fig. 3 に手分析データ推定モデルの概要図を示す。プラント内に設置されたセンサーや外部気象観測・予測機関から取得されるオンラインデータ、月 1 回あるいは年 1 回手作業によるオフライン測定で得られる手分析データ、化学反応モデルを元に、本来オフラインデータであった手分析データをオンライン推定する。非線形である化学反応モデルを、手分析データを元に線形モデルに簡易化し、本線形モデル上でオンライン測定水質に基づき数理計画問題として定式化、多数の解候補の中から妥当な解を手分析データとしてオンライン推定し導出する。

数理計画問題をオンラインで実施するために、化学反応モデルを線形モデルに簡易化する。化学反応モデルは連続関数であるため偏微分可能であり、手分析データを基準とする線形近似により簡易モデルを生成する。本論文では、オフライン測定データを平均とする乱数を生成、最小二乗法により下記近似式の係数  $A_p$  および  $B_p$  を決定する。

$$\hat{x}_{t+p}^{\text{out,on}} + \hat{x}_{t+p}^{\text{out,off}} \approx A_p(x_t^{\text{in,on}} + x_t^{\text{in,off}}) + B_p \quad \dots (11)$$

Lasso 回帰や Ridge 回帰等の別の線形近似モデルを用いて算出してもよいが、サンプルに対する線形近似式を求める目的において、最小二乗法が最もサンプル

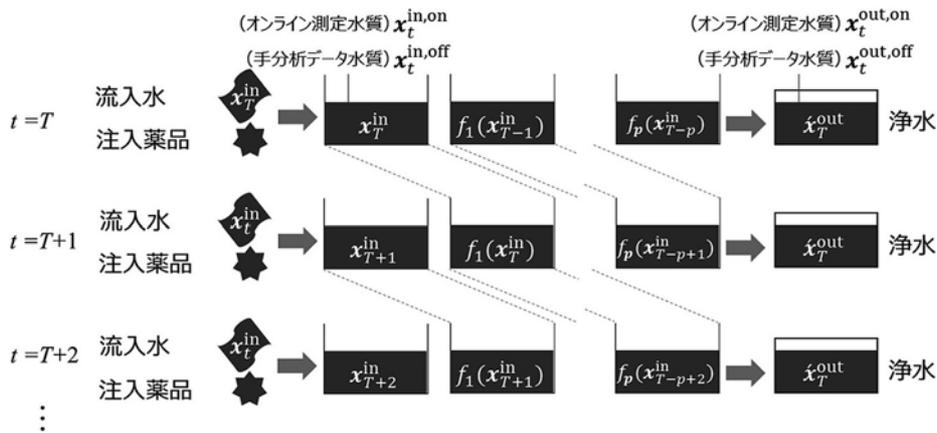


Fig. 2 Overview of the residual chlorine prediction with water quality response function  $f_p$

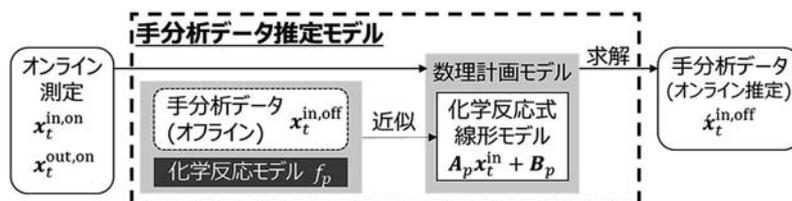


Fig. 3 Manual-sampling data estimation model

と近似値の誤差が小さくなるため、精度の面から最小二乗法を選定した。Lasso 回帰を用いた場合、不要な入力変数の係数が明示的にゼロとなるため出力に影響しない入力変数が含まれる場合に有効であるが、本モデルでは全ての入力変数（塩素消費物質濃度）が残留塩素濃度に影響することが事前に分かっており、考慮する必要がない。また Lasso 回帰や Ridge 回帰は、いずれも過学習を抑える効果があるが、ここでは手分析データに基づく定点近傍の近似式を求める目的でノイズの無い化学反応モデルの出力値を用いているため、考慮する必要がない。

Fig. 2 における化学反応モデルを式 (1) に置き換えたとき、 $l$  の連続する過去参照時間 ( $t=T-l, \dots, T-1, T$ ) について、以下の式が得られる。

$$\left. \begin{aligned} \mathbf{x}_{T-l+p}^{\text{out,on}} + \mathbf{x}_{T-l+p}^{\text{out,off}} &= \mathbf{A}_p \mathbf{x}_{T-l}^{\text{in,on}} + \mathbf{A}_p \mathbf{x}_{T-l}^{\text{in,off}} + \mathbf{B}_p + \delta_{T-l+p} \\ &\vdots \\ \mathbf{x}_{T-1+p}^{\text{out,on}} + \mathbf{x}_{T-1+p}^{\text{out,off}} &= \mathbf{A}_p \mathbf{x}_{T-1}^{\text{in,on}} + \mathbf{A}_p \mathbf{x}_{T-1}^{\text{in,off}} + \mathbf{B}_p + \delta_{T-1+p} \\ \mathbf{x}_{T+p}^{\text{out,on}} + \mathbf{x}_{T+p}^{\text{out,off}} &= \mathbf{A}_p \mathbf{x}_T^{\text{in,on}} + \mathbf{A}_p \mathbf{x}_T^{\text{in,off}} + \mathbf{B}_p + \delta_{T+p} \end{aligned} \right\} \dots (12)$$

$\delta_t$  は時刻  $t$  の推定値と測定値の誤差（ベクトル）である。ただし、オフラインデータは推定値と測定値の比較ができないため、

$$\delta_t = [\delta_t^1, \dots, \delta_t^m, 0, \dots, 0]^T$$

とする。未知の項を左辺、既知の項を右辺に整理すると、

$$\left. \begin{aligned} -\mathbf{A}_p \mathbf{x}_{T-l}^{\text{in,off}} + \mathbf{x}_{T-l+p}^{\text{out,off}} + \delta_{T-l+p} &= \mathbf{A}_p \mathbf{x}_{T-l}^{\text{in,on}} + \mathbf{x}_{T-l+p}^{\text{out,on}} + \mathbf{B}_p \\ &\vdots \\ -\mathbf{A}_p \mathbf{x}_{T-1}^{\text{in,off}} + \mathbf{x}_{T-1+p}^{\text{out,off}} + \delta_{T-1+p} &= \mathbf{A}_p \mathbf{x}_{T-1}^{\text{in,on}} + \mathbf{x}_{T-1+p}^{\text{out,on}} + \mathbf{B}_p \\ -\mathbf{A}_p \mathbf{x}_T^{\text{in,off}} + \mathbf{x}_{T+p}^{\text{out,off}} + \delta_{T+p} &= \mathbf{A}_p \mathbf{x}_T^{\text{in,on}} + \mathbf{x}_{T+p}^{\text{out,on}} + \mathbf{B}_p \end{aligned} \right\} \dots (13)$$

ここで、未知である左辺の各項について下記仮定に基づき数理計画モデルを定式化する。

・第1項： $\mathbf{x}_t^{\text{in,off}} (t=T-l, \dots, T-1, T)$  のバラつきは小さい

$l$  が小さく短時間の場合、急激に流入水質が変化しないものと仮定する。

・第2項： $\mathbf{x}_t^{\text{out,off}} (t=T-l, \dots, T-1, T)$  のバラつきは小さい

$l$  が小さく短時間の場合、定常の運転範囲内において浄水水質についても短時間に急激に変化しないものと仮定する。

・第3項： $\delta_t (t=T-l, \dots, T-1, T)$  は小さい

本線形近似モデルによる推定結果は測定値と十分に誤差が小さいものと仮定する。

上記仮定に基づき、目的関数および制約条件を以下の通り数理計画モデルとして定義し、最も妥当な解として未知の項を推定、 $p$  時間先浄水水質予測値（補正前） $\hat{\mathbf{x}}_{t+p}^{\text{out,off}}$  に含まれる次亜塩素酸濃度を残留塩素濃度予測値（補正前） $\hat{x}_{t+p}^{\text{out,HClO}}$  として出力する。

### 目的関数

Minimize:

$$r_1 \frac{1}{n} \sum_{i=m+1}^{m+n} D_i(\mathbf{x}_i^{\text{in},i}) + r_2 \frac{1}{n} \sum_{i=m+1}^{m+n} D_i(\mathbf{x}_i^{\text{out},i}) + r_3 \frac{1}{l} \sum_{i=0}^l \|\delta_i\|_1 \dots (14)$$

### 制約条件

式 (13)

ここで  $D_i$  はインデント  $i$  についての平均偏差を出力する関数、 $r_1, r_2, r_3$  は各項の重要度に基づく重み付け係数である。第1項、第2項について、標準偏差でなく平均誤差を採用する理由は、LP（線形計画問題）で解導出可能とするためである。同様の理由で、第3項について2次ノルムではなく1次ノルムを採用する。係数  $r_1, r_2, r_3$  は各項の重要度に基づき、値が大きいほど重要であることを示す。各仮定の信頼度を鑑み、 $r_3 > r_1 > r_2$  となるよう設定する。予備シミュレーションの結果に基づき、本論文では  $l=5$  時間、 $r_1=3.0, r_2=1.0, r_3=9.0$  として実施する。

## 2.4 誤差補正モデル

手分析データ推定値に基づき化学反応モデルより出力される残留塩素濃度予測値（補正前）と、実際の測定値には誤差があり、その要因として、化学反応モデル上で定義されていない物理的・化学的反応や機特性などによる影響が考えられる。そこで機械学習を用いて誤差補正することにより、予測精度を高める。

Fig. 4 に誤差補正モデルの概要を示す。化学反応モデルより出力される残留塩素濃度予測値（補正前）お

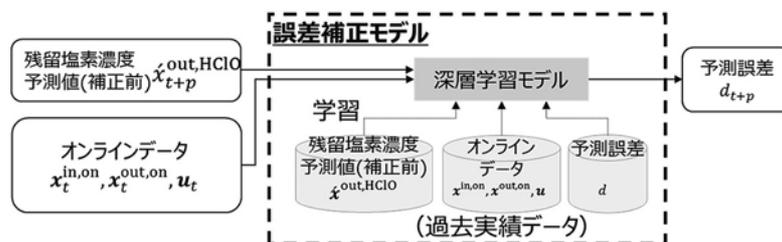


Fig. 4 Error correction model

よびオンラインデータを入力として、実際の浄水水質値との誤差を推定し出力する機械学習モデルを構築する。全てのオンラインデータについて現在時刻までの過去4時間分、オペレータによる操作項目については $h$ 時間先までの計画値、気象観測データについては $h$ 時間先までの予報値が含まれる。学習は上記入出力データの過去実績データを用いて実施する。

### 3. シミュレーション評価

#### 3.1 シミュレーション条件

Fig. 1 に示す水質予測モデルを Python 3.6 上に実装 (OS: Windows 7 64 bit, CPU: Intel Core i5 3.20 GHz, Memory: 4.0 GB) し、国内の某浄水場施設における2017年1年間分の1時間単位帳票データおよび気象観測データをオンラインデータとして、浄水残留塩素濃度の予測を実施した。本浄水場施設は、粉末活性炭処理を備えた凝集沈殿および砂ろ過の処理フローを持ち、配水量は平均約 30,000 m<sup>3</sup>/日、最大約 60,000 m<sup>3</sup>/日の処理能力を有する。平均取水流量および槽容量に基づき本浄水場施設における平均滞留時間を算出し、 $h=4$ 時間とする。Table 2 に本論文で用いたオンラインデータ項目一覧を示す。手分析データについては、1カ月毎に同浄水場にて採水された原水水質分析結果の内2.2節に示す水質項目データを用いる。ただし対

Table 2 Item list of online data

項目番号	項目名
1	原水 pH
2	原水濁度
3	原水アルカリ度
4	原水水温
5	原水流量積算
6	PAC 注入率
7	活性炭注入率
8	消石灰注入率
9	前次亜塩注入率
10	沈澱水 pH
11	沈澱水濁度
12	沈澱水水温
13	総ろ過流量積算
14	ろ過水濁度
15	ろ過水残留塩素
16	ろ過水水温
17	後次亜塩注入率
18	浄水池水位
19	浄水池流量積算
20	浄水水温
21	浄水残留塩素
22	雲量
23	日射量
24	気温
25	降水量
26	日照時間

Table 3 List of reaction rate constants

反応速度定数	値
$k_{As}$	$1.0 \times 10^2$
$k_{Zn}$	$1.0 \times 10^2$
$k_{Fe}$	$1.7 \times 10^4$
$k_{Mn}$	$6.4 \times 10^{-4}$
$k_{NO2}$	$1.0 \times 10^4$
$k_{NH3}$	$1.3 \times 10^4$
$k_{NH2Cl}$	$1.2 \times 10^2$
$k_{NHCl2}$	$1.0 \times 10$
$k_{TOC}$	1.0

Table 4 Parameter of deep learning for the error correction model

項目	値
学習回数	1000~4000
レイヤ数	3~5
ニューロン数	200~300
活性化関数	Sigmoid, Tanh, ReLU
最適化手法	Adam, Momentum SGD
学習率	0.005~0.02

象水質項目の検査が年毎に実施の場合は、本値を用いる。NH<sub>3</sub>は手分析測定項目に無いため、全国水質統計5) および対象機場での実測を参考に 0.20 mg/L を手分析データとして与える。

化学反応モデルにおける水質変化関数は MATLAB/Simulink 上に実装し、ステップ時間を1秒として数値解析を実施する。文献6)~8) らを参考に、Table 3 に示す反応速度定数を用いる。反応速度定数は水温および pH の影響を受けるが、本論文ではそれらの変動に関わらず一定値として検証する。誤差補正モデルは各月25日までを訓練データ (モデル自体の学習)、26日以降を検証データ (学習結果の評価) としてモデル構築を実施し、Table 4 に示す各パラメータを自動チューニングする。

シミュレーションは手分析データ推定あり/なし、誤差補正あり/なしの組み合わせについて実施し、それぞれ (1) 手分析データのオンライン推定による予測精度改善効果、(2) 化学反応モデル上で実装されていない未知の機場特性に対する誤差補正による予測精度改善効果を評価する。手分析データ推定なしの場合、取得した手分析データ値を一定値としてオンラインデータ化し、化学反応モデルの入力として用いる。評価指標として、予測値と実測値との符号付き誤差、絶対誤差それぞれの平均および標準偏差、散布図上でのばらつきを評価する。

### 3.2 評価結果

#### 3.2.1 手分析データのオンライン推定による予測精度改善

Fig. 5 および Table 5 に各シミュレーション条件および結果を示す。浄水残留塩素濃度実測値平均 0.440 mg/L, 標準偏差 0.236 mg/L に対し、Case A の化学

反応モデル単体での予測における絶対誤差平均が0.578 mg/Lと大きくなった。その理由として、天候や季節の変化に伴う手分析水質値の大きな変動に対し追従できていないこと、紫外線照射や機場特性に起因する本論文で対象とする塩素消費物質による化学反応以外による影響が挙げられる。Case A と Case C, Case B と Case D のそれぞれの比較において、後者の方が符号付き誤差、絶対誤差それぞれの平均が小さい値を示していることより、手分析データのオンライン

推定により予測精度が改善していることがわかる。

符号付き誤差平均を参照すると、Case A と Case B はそれぞれ 0.562 mg/L, 0.164 mg/L 実測値より大きく予測しており、手分析データ値を一定値としてオンラインデータ化した場合に、実際の手分析データのリアルタイムな変動よりも少なく見積もっているところが原因と考えられる。

Fig. 6 に手分析データ推定モデルによる各水質項目 (Fe, Mn, NH<sub>3</sub>, TOC) の推定結果の分布を示す。その他の水質値は、実際の手分析測定において検出下限値以下のため本推定モデルの出力は常にゼロ (推定濃度 0 mg/L の頻度 100%) となったため図示を省略した。実際の手分析測定値が Fe, Mn, TOC から順に、0.060 mg/L, 0.012 mg/L, 1.080 mg/L であり、これらと NH<sub>3</sub> の 0.243 mg/L を併せてオンライン推定値と比較すると、Fe, Mn, NH<sub>3</sub> については高く、TOC については低く推定している。手分析データ推定値が高いほど、化学反応モデルによる残留塩素濃度の推定値は低く出力されるが、特に Fe, NH<sub>3</sub> は反応速度が速く、浄水池に到達するまでに反応が終了するため、その影響は大きい。そのため、Fe, NH<sub>3</sub> の手分析データ値が実際の値よりも低かったことが Case A の予測精度悪化の原因と考えられ、実際の手分析データのリアルタイムな変動よりも少なく見積もっていたとした前段落の仮定と一致する。

3.2.2 誤差補正による予測精度改善

Case A と Case B, Case C と Case D のそれぞれの比較において、後者の方が符号付き誤差、絶対誤差それぞれの平均および標準偏差いずれも小さい値を示していることより、化学反応モデル上で実装されていない未知の機場特性に対する誤差補正により予測精度が改善していることがわかる。手分析データのオンライン推定および誤差補正による改善効果において、両因子の交互作用はみられず、Case D が全評価指標において最も優れる。

Case B と Case C を比較すると符号付き誤差、絶対誤差ともに、平均については誤差補正が、標準偏差については手分析データ推定が、より改善効果が高いことがわかる。これは、機械学習モデルが学習データ範囲の内挿となる入力データに対して有効であるが、外挿の入力データに対して予測精度が悪化することが一

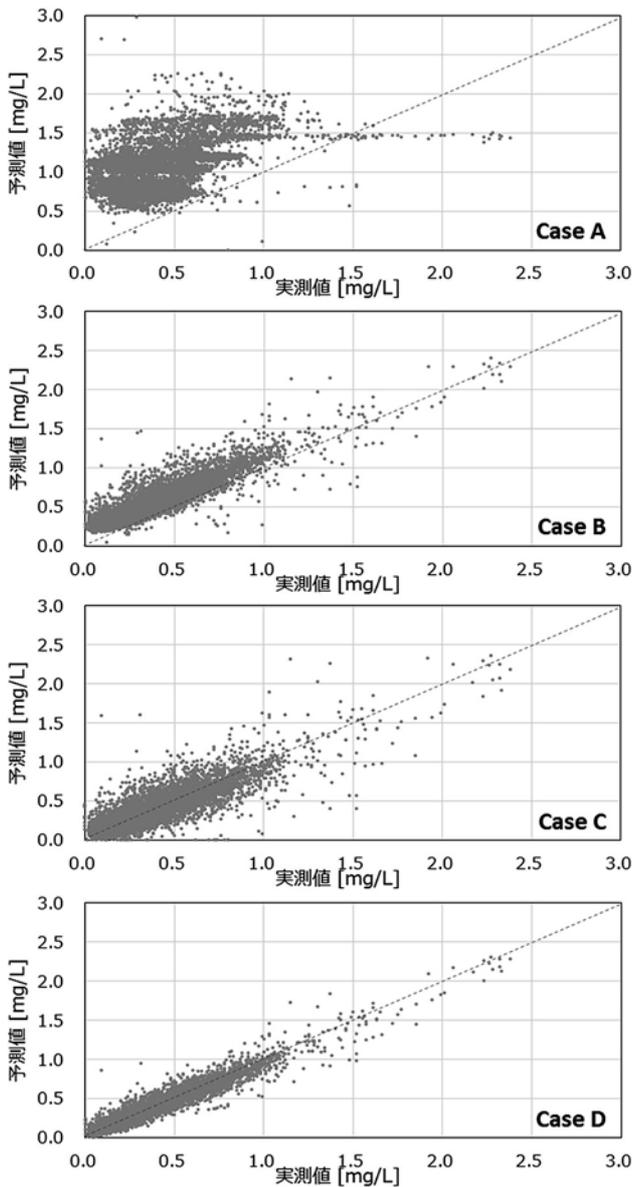


Fig. 5 Simulation results in prediction accuracy

Table 5 Simulation results

	Case A	Case B	Case C	Case D
手分析データ推定	なし	なし	あり	あり
誤差補正	なし	あり	なし	あり
符号付き誤差平均	0.562 mg/L	0.163 mg/L	-0.002 mg/L	-0.001 mg/L
符号付き誤差標準偏差	0.296 mg/L	0.106 mg/L	0.133 mg/L	0.080 mg/L
絶対誤差平均	0.578 mg/L	0.172 mg/L	0.093 mg/L	0.048 mg/L
絶対誤差標準偏差	0.276 mg/L	0.098 mg/L	0.117 mg/L	0.058 mg/L

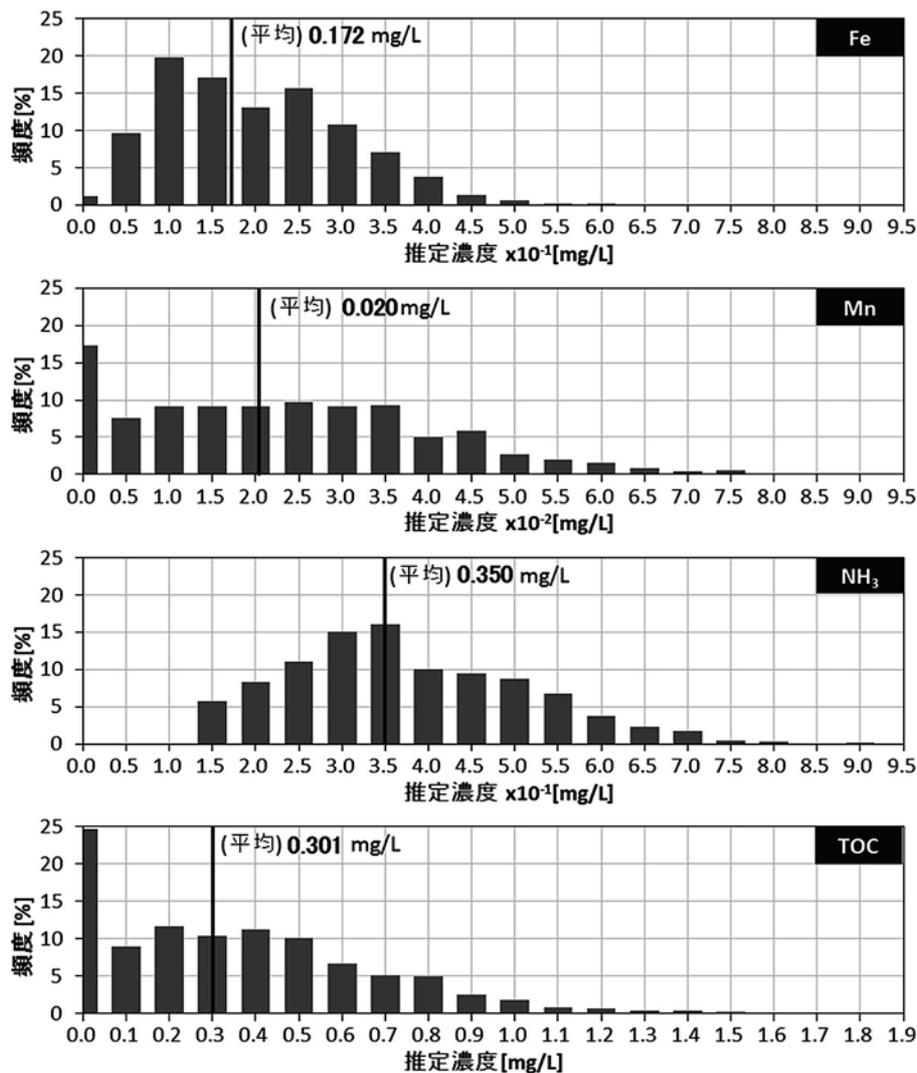


Fig. 6 Distributional results of manual-sampling estimation

因と考えられる。

ただし、手分析データ推定モデルの仕様上、実際の手分析測定値から離れた値の推定についてはその手分析データの推定精度が悪化することが懸念される。Fig. 6における推定結果はマイナス値の不在を除き正規性が見られるが、実際には降雨やダム放流による高濁時や雪解け水の流入による著しい水質変化が伴うことが想定され、それらに起因する特異値が存在していても本結果からは確認できない。本論文における手分析データ推定モデルは実際に手作業で測定された手分析測定値を元にその近傍における妥当解を出力する仕様のため、特異事象に起因する外れ値の推定には不適であると考えられる。本手法による手分析データのオンライン推定の適用範囲や係数 $r_1$ ,  $r_2$ ,  $r_3$ の調整による精度改善の検討については、実採水データを用いた追加検証が必要である。また、紫外線照射等による物質以外による塩素消費への影響についても今後検討が必要である。

#### 4. ま と め

本論文では、非定常時も予測可能なホワイトボックスモデルを化学反応モデルとしてベースに、未知の機場特性や外乱に強いブラックボックスモデルを誤差補正モデルに適用することで、両者の優位点を備えるハイブリッドモデルによる残留塩素濃度予測手法を提案した。本モデルの有効性についてシミュレーション評価した結果、手分析データのオンライン推定および誤差補正それぞれにおいて予測精度改善効果を確認し、両者の組み合わせにより浄水残留塩素濃度予測において絶対誤差平均 0.048 mg/L を達成した。実採水データを用いた手分析データのオンライン推定値の妥当性評価や、紫外線照射分解など化学反応以外の塩素消費要因について今後検討する必要がある。

## 参考文献

- 1) 福崎智司：次亜塩素酸ナトリウムを用いた洗浄・殺菌操作の理論と実際，調理食品と技術，Vol. 16, No. 1, pp. 1-14 (2010).
- 2) 田中利昭，増子敦，尾崎勝，佐藤親房：クロラミン類生成に関する新たな知見と浄水処理での低減化への適用，水道協会雑誌，Vol. 79, No. 6, pp. 2-12 (2010).
- 3) 圓佛伊智朗，馬場研二，松崎晴美，依田幹雄：ニューラルネットワークを用いたプラント運転ルールの抽出に関する研究，電気学会論文誌 D，Vol. 111, No. 1, pp. 20-28 (1991).
- 4) 松木洋，藤本康孝：SVMおよびPCAを用いた運転ノウハウの抽出による上水送水系の運転支援システム，電気学会論文誌 D，Vol. 132, No. 10, pp. 990-996 (2012).
- 5) JWWA, 水道水質データベース 平成 28 年度水道統計水質分布表(原水) 平均値, pp. 1-21 (2016).
- 6) O. Hao, A. P. Davis, and P. H. Chang: Kinetics of Manganese (II) Oxidation with Chlorine, Journal of Environmental Engineering, Vol. 117, No. 3 (1991).
- 7) Z. Qiang and C. Adams: Determination of Monochloramine Formation Rate Constants with Stopped-Flow Spectrophotometry, Journal of Environmental science & technology, Vol. 38, pp. 1435-1444 (2004).
- 8) L. K. Folkes, L. P. Candeias, and P. Wardman: Kinetics and Mechanisms of Hypochlorous Acid Reactions, Archives of biochemistry and biophysics, Vol. 323, pp. 120-6 (1995).

## Residual Chlorine Prediction with Hybridization of Machine Learning and Chemical Reaction model

Jun Matsumoto<sup>1)</sup>, Hiroyuki Yamahara<sup>1)</sup>, Katsuya Yokokawa<sup>1)</sup>, Takeshi Menju<sup>1)</sup>  
Futoshi Kurokawa<sup>1)</sup> and Suguru Yokoyama<sup>2)</sup>

<sup>1)</sup> Infrastructure Systems Research & Development Center, Toshiba Infrastructure Systems & Solutions Corporation

<sup>2)</sup> Social Systems Division, Toshiba Infrastructure Systems & Solutions Corporation

† Correspondence should be addressed to Jun Matsumoto :

(Infrastructure Systems Research & Development Center, Toshiba Infrastructure Systems & Solutions Corporation  
E-mail : jun8.matsumoto@toshiba.co.jp)

### Abstract

In water purification plant, highly-accurate prediction of residual chlorine concentration has gained much importance for the automation and optimization of chlorine injection. There is no conventional residual chlorine prediction model based on the chemical response of multiple chlorine consuming materials. The residual chlorine prediction model has the challenges : it is not in real-time since some of essential inputs are obtained by manual measuring, and it cannot take into account unknown influential factors or plant characteristics which are not implemented on the model. This paper proposes a hybridization approach using machine learning in order to perform the above tasks with the mean absolute error of 0.048 mg/L relative to the measured mean of 0.440 mg/L in the performance evaluation of prediction accuracy.

**Key words :** residual chlorine, prediction, machine learning, hybridization