

ランダムフォレスト法による 洪水時の水位予測手法の提案

岡崎 亮太¹・中津川 誠²・小林 洋介³

¹学生会員 室蘭工業大学大学院工学研究科 (〒050-8585 室蘭市水元町 27-1)

E-mail:17041010@mmm.muroran-it.ac.jp

²正会員 博士 (工学) 室蘭工業大学大学院工学研究科 (〒050-8585 室蘭市水元町 27-1)

E-mail:mnakatsu@mmm.muroran-it.ac.jp

³非会員 博士 (工学) 室蘭工業大学大学院工学研究科 (〒050-8585 室蘭市水元町 27-1)

E-mail:ykobayashi@csse.muroran-it.ac.jp

本研究の目的は、河川水位の的確な予測を目指し、全国で整備が進められている水文情報の機械学習による予測手法を提案することである。2016年8月の北海道豪雨など、近年大規模水害が頻発しており、住民の避難等に活用するため予測情報の重要性が高まっている。本研究では機械学習手法の一つであるランダムフォレスト(RF)法を用いて予測リードタイムが6時間以上の水位予測を行った。また、RF法から得られる寄与度が大きい説明変数を抽出し、それらと目的変数の関係を回帰式で表す方法を関連要因相関法として提案し、予測を行った。以上の結果から、ピーク水位の予測に関して過小評価しない安全側の結果が確認でき、情報工学的手法に基づく実用的な水位予測を提案することができた。

Key Words: machine learning, Random Forest method, prediction of river water level, related factor correlation technique, continuous floods

1. はじめに

2016年8月、北海道において4つの台風が連続して上陸及び接近し、広域にわたって甚大な被害をもたらした。特に、北海道東部の常呂川では、下流部で、計画高水位を超過し、観測史上最大の水位を記録した¹⁾。このような洪水を事前に予測し被害を軽減させるためには、水防活動、避難場所の開設、住民の避難等といった防災・減災対策に要する時間(リードタイム、以下、LTと称す)の確保が望まれる。特に市街地を流れる河川では洪水時における早めの適切な判断が住民の安全に大きく影響する。したがって、LTを長く確保させるため、河川水位の増減やピークの予測精度の向上が望まれる。

水位予測の手法として実用化されているものに、上流部の河川水位との相関をとる方法(以後、水位相関法)や、洪水流出モデルに予測雨量を与えて流量を計算し、水位-流量曲線(H-Q曲線)から水位に換算する方法²⁾がある。さらに計算された値を観測値で修正するカルマンフィルタ³⁾や粒子フィルタ⁴⁾といったフィードバック手法も提案されている。そのような中、大塚ら⁵⁾は釧路

川を対象として上流地点の水位や雨量との重相関に基づく水位予測が、洪水流出モデルによる水位予測と同等またはそれ以上の精度があることを示した。本研究では、大塚ら⁵⁾の予測地点の上流地点との重相関による分析を拡張し、近年急速に発展している情報工学分野の機械学習手法による水位予測を提案する。

これまでも水位予測に機械学習が利用されてきた例はいくつかある。Holger⁶⁾らはニューラルネットワークによる洪水予測手法を提案しているほか、一言ら^{7), 8)}は深層学習法に基づく洪水予測手法や深層ニューラルネットワークと物理モデルである分布型モデルを組み合わせた手法を提案している。これらの手法では、水位に関連する多数の要因が予測に反映され、精度の良い予測結果が得られている。しかし、予測結果にどの要因が影響するのか分析できず、因果関係の把握が困難であった。そこで本研究では、説明変数となる各要因の寄与度も算定できる機械学習手法の一つであるランダムフォレスト法(以下、RF法)による水位予測を試みる。さらに、現業での実用性を鑑み、RF法から得られる寄与度の高い説明変数を用いた相関式による水位予測を行う方法

表-1 学習データの概要

予測対象地点	上川沿地点(LT6h)				上川沿地点(LT12h)				北見地点(LT6h)				標茶地点(LT6h)			
観測項目地点	観測地点数	データを抽出する時間	抽出するデータ個数	説明変数の個数	観測地点数	データを抽出する時間	抽出するデータ個数	説明変数の個数	観測地点数	データを抽出する時間	抽出するデータ個数	説明変数の個数	観測地点数	データを抽出する時間	抽出するデータ個数	説明変数の個数
水位 (時間値, m)	7	$t-4 \sim t$	5	35	7	$t-4 \sim t$	5	35	5	$t-4 \sim t$	5	25	4	$t-4 \sim t$	5	20
観測雨量 (累積値, mm)	1	$t-6 \sim t$	7	7	1	$t-6 \sim t$	7	7	1	$t-6 \sim t$	7	7	1	$t-6 \sim t$	7	7
予測雨量 (みなし予測) (累積値, mm)	1	$t \sim t+5$	6	6	1	$t \sim t+11$	12	12	1	$t \sim t+5$	6	6	1	$t \sim t+5$	6	6
ダム放流量 (時間値, mm/h)	1	$t-4 \sim t$	5	5	1	$t-4 \sim t$	5	5	1	$t-4 \sim t$	5	5	-	-	-	-
水位差分 (時間値, m)	1	$t-1 \sim t$	1	1	1	$t-1 \sim t$	1	1	1	$t-1 \sim t$	1	1	1	$t-1 \sim t$	1	1
全説明変数 n	54				60				44				34			
学習データ数*	12事例				12事例				11事例				7事例			
計画高水位(m)	8.45				8.45				60.81				23.50			



※鑑別地点水位(開発局)は2014年から閉局のため鑑別川地点水位(北海道)と相関をとって補完した。

図-1 対象河川の概略図

(関連要因相関法)を提案し、説明変数の選別、LTの延長といった観点で検討を行った。以下に研究ケースを示す。

- 1) RF法を用いて、過去の洪水事例を学習させて作成した予測モデルより、北海道で豪雨災害が起きた2016年8月洪水を予測対象とした水位予測
- 2) RF法から算出される寄与度によって抽出された上位の説明変数を用いて回帰式を求める関連要因相関法による予測
- 3) RF法の寄与度上位の説明変数のみから再度RF法によって水位予測を行った。

2. 研究方法

(1) 災害の事例と対象河川の概要

* t は現在時刻を意味する。

*予測雨量(みなし予測)：実測雨量を予測雨量とみなしたもの。

*学習データ：過去の洪水事例を選出し学習データとした。

対象地点は、北海道東部に位置する一級河川常呂川における下流の上川沿(かみかわぞい)地点および北見地点とし、同じく北海道東部の太平洋側に位置する一級河川釧路川における下流の標茶(しべちや)地点とした。図-1に対象河川の概略図と諸元を示す。河口からの距離をKPで表すと上川沿地点はKP10.9km、北見市中心部の北見地点はKP49.4km、標茶地点はKP46.1kmに位置している。2016年8月の洪水時は、水位が下がりきらないうちに次々と台風が連続して接近したため、少量の雨でも水位が上昇しやすい状態が続いたため、従来の予測方法では十分な精度で予測を行うことができない例がみられた。

よって本研究ではこの2016年8月4日から9月14日までの台風による大雨が降った期間を対象事例とし、上川沿、北見、標茶地点でのピーク水位の予測精度向上およびLTの延長を目指した。

(2) 予測に係る要因

水位予測を行う上で必要な要素として以下の5つの要素を考えた。表-1に用いた学習データの概要を示す。

「水位」として国土交通省所管のテレメータ観測所のデータを使用した。また、「観測雨量」として国土交通省所管のテレメータ観測所のデータを使用し、ティーセン法により流域平均雨量を求め、累積雨量にしたデータを使用した。「予測雨量」に関しては実測の流域平均雨量を予測雨量に見立てて使用し、累積雨量にしたデータを使用した(以下、みなし予測と称する。)。「ダム放流量」は常呂川上流部にある国土交通省所管の鹿ノ子ダムの放流量を使用した。「水位差分」は、上川沿、北見、標茶自身の1時間前と現在時刻の水位を使用した。これら5つの要素を取得できるものは水文・水質データベース⁹⁾より取得し、現在閉局となっている釧路川に位置する鑑別(とうべつ)地点の水位は国土交通省北海道開発局釧路開発建設部、鹿ノ子(かのこ)ダムの放流量は、国土交通省北海道開発局網走開発建設部から提供いただいた。学習データは、1997年から2015年の年最大水位19事例の中から夏期(8月、9月、10月)に最大値が観

表-2 予測ケース表

ケース	予測方法	予測地点	リードタイム	目的変数 (予測値)	説明変数
1	RF法	上川沿	6h	水位(時間値) $t+6$	水位(時間値) $t-4 \sim t$ 観測雨量(累積値) $t-6 \sim t$ 予測雨量* (累積値) $t \sim t+5$ ダム放流量(時間値) $t-4 \sim t$ 水位差分(時間値) $(t-1)-t$
			12h	水位(時間値) $t+12$	水位(時間値) $t-4 \sim t$ 観測雨量(累積値) $t-6 \sim t$ 予測雨量(累積値) $t \sim t+11$ ダム放流量(時間値) $t-4 \sim t$ 水位差分(時間値) $(t-1)-t$
		北見	6h	水位(時間値) $t+6$	水位(時間値) $t-4 \sim t$ 観測雨量(累積値) $t-6 \sim t$ 予測雨量(累積値) $t \sim t+5$ ダム放流量(時間値) $t-4 \sim t$ 水位差分(時間値) $(t-1)-t$
		標茶	6h	水位(時間値) $t+6$	水位(時間値) $t-4 \sim t$ 観測雨量(累積値) $t-6 \sim t$ 予測雨量(累積値) $t \sim t+5$ 水位差分(時間値) $(t-1)-t$
2	関連要因 相関法*	上川沿	6h	水位(時間値) $t+6$	忠志地点水位(時間値) t , H_1 $H=1.23H_1-19.07$
			12h	水位(時間値) $t+12$	忠志地点水位(時間値) t , H_1 $H=1.17H_1-17.94$
		北見	6h	水位(時間値) $t+6$	上常呂地点水位(時間値) t , H_1 $H=0.88H_1-0.0045H_2-20.82$ 北光社水位(時間値) t , H_2
		標茶	6h	水位(時間値) $t+6$	鑑別地点水位(時間値) t , H_1 $H=0.12H_1+0.79H_2+0.012R-7.51$ 標茶水位(時間値) t , H_2 観測雨量(時間値) $t-6$, R
3	修正RF 法*	北見	6h	水位(時間値) $t+6$	上常呂地点水位(時間値) t 北光社水位(時間値) t 観測雨量(累積値) $t-6 \sim t$ 予測雨量(累積値) $t \sim t+5$

* t は現在時刻を意味する

*予測雨量：当日出された予測値を使用

*関連要因相関法：ケース1のRF法の寄与度から抽出された説明変数との回帰式で予測する方法

*修正RF法：ケース1のRF法の寄与度から抽出された説明変数に再度RF法を適用して予測する方法

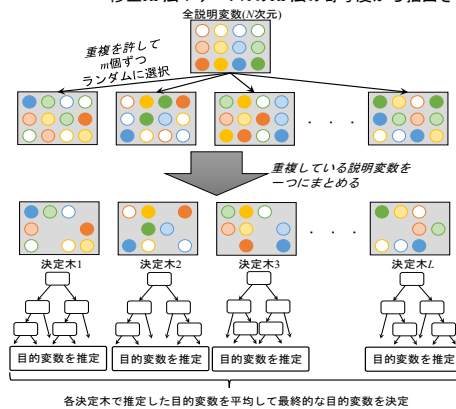


図-2 RF法の概略図

測されている年の事例を抽出した。

(3) RF法による水位予測

RF法^{10)・11)}とは、多数の決定木を組み合わせることによって汎化能力を高める方法である。図-2にRF法の概略図を示し、以下の方法で計算を行う。

- 1) 全説明変数 n 個から重複を許しランダムに1つの説明変数を抽出する作業を m 回行う。 n 及び m の値が十分大きい場合 n の約64% $(=1-(1-1/n)^m)$ の説明変数が抽出される。
- 2) 抽出された説明変数を用いて1本の決定木を作成する。決定木の構造は以下に示す平均二乗誤差（以下、MSE）を用いて作成する。

$$MSE = \frac{1}{N} \sum_{i=1}^N (x_i - x)^2 \quad (1)$$

ここで、 N は分割時の時系列データ数、 x_i は説明変数の値、 x は説明変数の分割値である。これを全てのパターンで計算して、MSEの値が最も小さくなる x で説明変数を分割するような決定木を作成し、目的変数を推定する。

- 3) 1), 2)の作業を L 回行い L 本の決定木を作成する。各決定木で推定した目的変数を平均し、最終的な

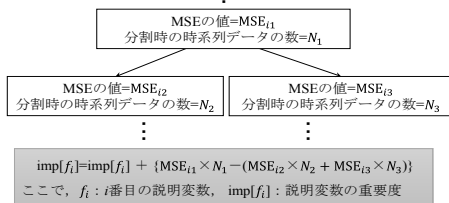


図-3 寄与度算出方法の概略図

目的変数を決定する。

- 4) 本研究では、全説明変数 n は表-1に示す。抽出回数 m は n と同数($m=n$)とした。決定木の本数 L は全て $L=2,000$ とした。

また、図-3に示すような変数重要度（以下、 $imp[f_i]$ ）を決定木の分割ごとに計算を行い、全ての $imp[f_i]$ の総和で各説明変数の $imp[f_i]$ を除して正規化したものが寄与度となる。よって寄与度の値は0～1の範囲となる。この寄与度は、個々の説明変数と目的変数の相関係数を求めた値と異なり、説明変数内の従属性の影響を排除したロバストな意味をもつ指標となる。

3. 結果と考察

(1) 検討ケース

水位予測を実施し結果の検証を行う。本研究では表-2のケースで2016年8月の事例を対象に水位予測を行った。予測雨量は当日に予測された雨量を使用した。予測雨量は気象台（6h先まで）及び日本気象協会（24h先まで）のデータを用いた。RF法により推定した水位の予測結果に対する誤差評価には、次式に示す平均二乗誤差平方根（以下RMSE）を用いる。

$$RMSE = \sqrt{\frac{\sum (H_{pre} - H_{obs})^2}{N}} \quad (2)$$

ここで、 H_{pre} は予測水位（m）、 H_{obs} は観測水位（m）、

N はデータ数である。また、既存手法として国土交通省北海道開発局の現業で用いられている手法の結果を比較対象とした。これは、貯留関数モデルで計算された流量と観測流量にカルマンフィルターを作用させて更新した流量を初期値として流量を計算し、これをH-Q式で水位に換算した結果を予測値とするものである。（以下、既存手法の予測結果と称する。）

(2) RF 法による水位予測結果

過去の洪水事例を学習させ、2016年8月の洪水事例をRF法で予測した結果をケース1として図-4に示す。この結果、上川沿地点でのLT6の予測結果は、ピーク水位に関して若干誤差があるものの、全体的に観測水位の挙

動をよく予測している。北見地点でのLT6のRF法の予測結果は、ピークのタイミングは一致しているものの低水位が過大に予測されており、ピーク値も過小評価されている。このことから北見地点では精度の改善が求められる。また、標茶地点でのLT6のRF法の予測結果は、低水位に関しては、観測水位の挙動をよく予測しているものの、ピーク水位は過小評価されており精度の改善が求められる。

(3) RF 法の寄与度算出結果

図-5に3節で得られた上川沿地点、北見地点、標茶地点の寄与度の算出結果を示す。図には寄与度の大きい順に、上位10位までの説明変数を示している。上川沿地点、北見地点はともに、上位に各観測所の水位情報が設定されており、流域平均雨量、予測雨量、ダム諸量、水位差分よりも寄与度が大きいことが分かった。また標茶地点は他の二つの地点と同様に、上位に各観測所の水位情報が設定されていることがわかる。しかしながら、流域平均雨量が上位に入っていることから他の寄与度の算出結果とは異なった結果となった。このことから、標茶地点を予測する際には、観測雨量データが必要であることがわかる。精度指標としてRMSEを表-3にまとめる。

(4) 関連要因相関法による予測結果

現業における実用性に鑑み、図-5の寄与度上位から色付けした説明変数を抽出し、関連要因相関法により予測精度向上を試みた。表-2のケースのとおり説明変数との回帰式を作成し予測を行った。結果（ケース2）を図-6に示す。（以下、グラフ上では重回帰式と称する。）

結果より上川沿地点ではRF法よりも説明変数を限定した関連要因相関法においても、ピーク水位を安全側に予測する結果を得た。ただし、北見地点の結果から低水

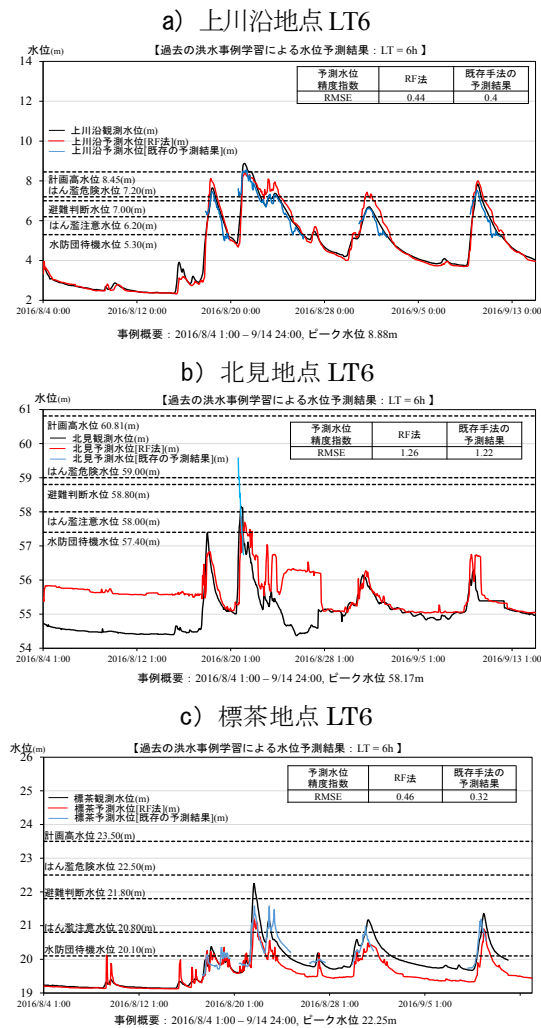


図-4 ケース1による予測結果

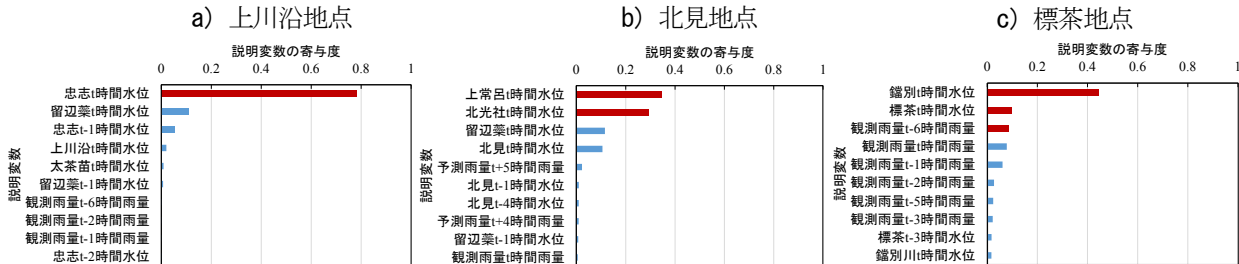


図-5 ケース1のRF法により算出された寄与度

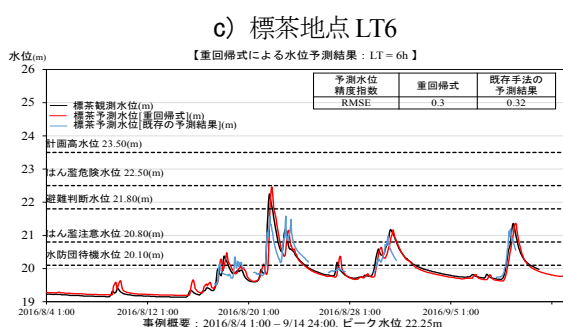
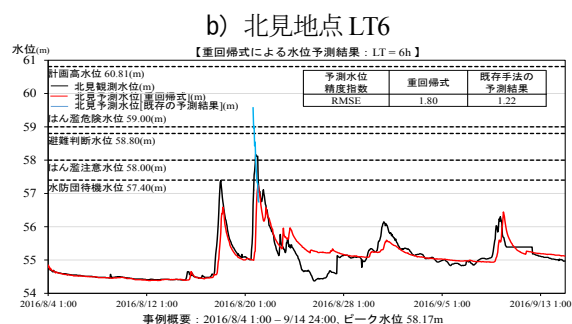
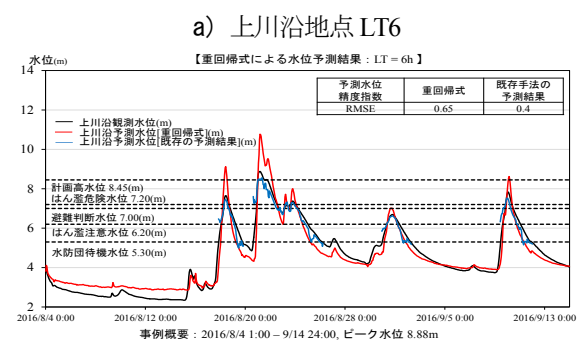


図-6 ケース2による予測結果

位に関しては、観測水位の挙動をよく予測しているものの、ピーク水位が過少評価されていることがわかる。このことから精度の改善が求められる。標茶地点での LT6 の RF 法の予測結果として全体的に観測水位の挙動をよく予測している。同様に既存手法の予測結果を青線で示すが、ピーク水位において RF 法が安全側に予測しており、予測精度が高いことがわかる。精度指標として RMSE を表-3 に示す。標茶地点が大幅に改善され既存手法の予測結果よりも精度向上がなされた。

(5) 修正 RF 法による予測結果

図-4、図-6 の結果から北見地点の予測精度に改善が求められることから、算出された寄与度の上位である上常呂地点と北光社地点及び流域平均雨量、予測雨量を学習データとし、RF 法による再計算を行った。これを修正 RF 法と称する。結果（ケース 3）を図-7 に示す。図-7 から他のケースとは異なり、ピーク水位に着目すると、実測水位に対し予測水位も同様の挙動を示し、低水位に関しても改善がみられていることがわかる。精度指標として RMSE を表-3 に示す。既存手法の予測結果と同等の精度を得ることができた。

(6) LT の延長による予測結果

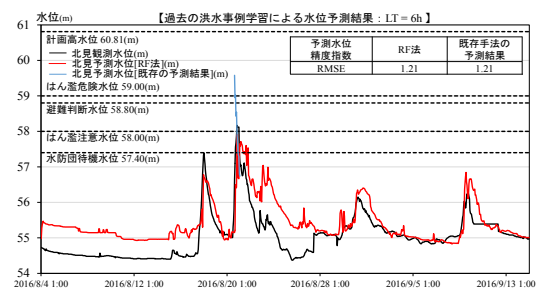


図-7 ケース3による予測結果（北見 LT6）

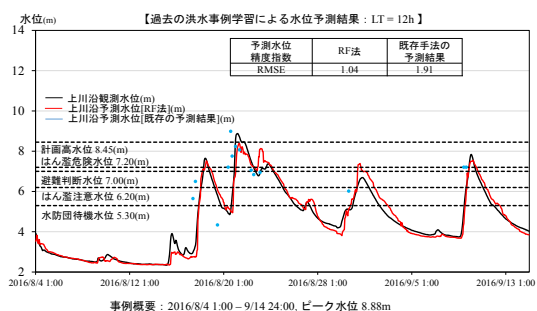


図-8 ケース1による予測結果（上川沿 LT12）

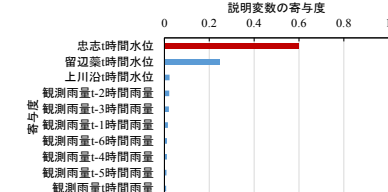


図-9 ケース1の RF 法により算出された寄与度（上川沿 LT12）

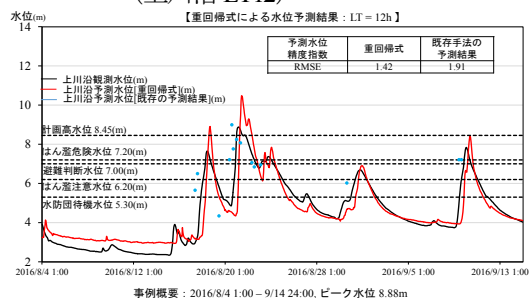


図-10 ケース2による予測結果（上川沿 LT12）

夜間の避難回避のため、半日以上の LT をもつ予測ができることが望ましい。そこで LT が 12h とした予測を RF 法で行った。LT6 と同様に、過去の洪水事例を学習させ、2016 年 8 月の洪水事例を予測した結果（ケース 1）を図-8 に示す。上川沿地点での LT12 の RF 法の予測結果は、LT6 と同様に、ピーク水位に関して若干誤差があるものの、全体的に観測水位の挙動をよく予測している。これにより、リードタイムを伸ばしたとしても十分な予測ができることを示した。

次に LT12 での RF 法により算出された寄与度を図-9 に示す。LT6 で得られた結果と同様に、上位に各観測所の水位情報が設定されており、上川沿地点では水位データが重要であることがわかる。

以上に基づき図-10（ケース2）に上川沿での関連要因相関法による LT12 時間の予測結果を示す。この結果から RF 法よりも説明変数を少なくした関連要因相関法に

においても、LTを12hとしてもピーク水位を安全側に評価していることがわかる。精度指標としてRMSEを表-3に示すが、既存手法の予測結果よりも精度が良い結果となった。

4. 結論

本研究で得られた結果を以下に記す。

- 1) RF法において過年度の情報から2016年8月のような経験のない洪水時の水位の予測が可能であることが示された。
- 2) 寄与度を算出することによって各説明変数の度合いが明確になることを示した。
- 3) 寄与度に基づく説明変数を限定することで関連要因相関法、修正RF法といった方法から予測結果の改善が図られた。
- 4) LTを12hに延長してもRF法及び関連要因相関法である程度の水位予測が可能であることを示した。

謝辞:本研究を遂行するにあたり、北海道開発局網走開発建設部、釧路開発建設部および株式会社ドーコンにはデータ提供等で協力して頂いた。ここに記して謝意を表す。

参考文献

- 1) 国土交通省北海道開発局：平成28年8月20日からの大雨及び台風第10号による出水の概要，<http://www.hkd.mlit.go.jp/ab/tisui/v6dkjr000000021d-att/v6dkjr000000023w.pdf>（閲覧日：2017/9/20）
- 2) 立川康人，永谷言，寶馨：飽和・不飽和流れの機構を導入した流量流積関係式の開発，水工学論文集，第48巻，pp. 7-12, 2004.
- 3) 片山直樹，星清，橋本識秀：一般化貯留関数法を用いた洪水予測手法の改良，土木学会北海道支部，第61号，II-11, 2005.
- 4) 立川康人，須藤純一，椎葉充晴，萬和明，キムスンミン：粒子フィルターを用いた河川水位の実時間予測手法の開発，水工学論文集，Vol. 67, No. 4, I_511-I_516, 2011.
- 5) 大塚康平，中津川誠，谷口陽子：釧路川を対象とした洪水流出再現精度向上の研究，土木学会北海道支部，第72号，B-23, 2015.
- 6) Maier, H. R. and Dandy, G. C. : Neural networks for the prediction and forecasting of water resources variables: a review of modelling issues and applications, *Environmental Modelling & Software*, Vol. 15, 2000.
- 7) 一言正之，桜庭雅明，清雄一：深層学習を用いた河川水位予測手法の開発，土木学会論文集 B1（水工学），Vol. 72, No. 4, I_187-I_192, 2016.
- 8) 一言正之，桜庭雅明，清雄一：深層ニューラルネットワークと分布型モデルを組み合わせたハイブリッド河川水位予測手法，土木学会論文集 B1（水工学），Vol. 72, No. 1, 22-33, 2017.
- 9) 国土交通省：水文水質データベース，<http://www1.river.go.jp>（閲覧日：2017/09/26）
- 10) 速水悟：事例+演習で学ぶ機械学習 ビジネスを支えるデータ活用のしくみ，第5章，森北出版株式会社，2016.
- 11) 武藤佳恭：超実践アンサンブル機械学習，第3章，近代科学社，2016.

(2017.9.29 受付)

A PROPOSED METHOD FOR FLOOD WATER LEVEL PREDICTION USING THE RANDOM FOREST METHOD

Ryota OKAZAKI, Makoto NAKATSUGAWA and Yosuke KOBAYASHI

This study proposes a method for accurately predicting river water levels by using hydrological information and machine learning. The provision of hydrological information has been promoted in wide areas of Japan. In recent years, large-scale flood damage has frequently occurred, including that from heavy rainfall in Hokkaido in August 2016. The importance of flood prediction, which is used for planning the evacuation of residents, has been increasing. In this study, water levels of previous floods were predicted with lead times of 6 hours or longer by using the Random Forest (RF) method, a machine learning method. Explanatory variables with high contribution were extracted from all the RF explanatory variables used for predicting previous floods, and prediction was done by developing a regression, that is the related factor correlation method to express the relationship between the extracted variables and the objective variables. The prediction of the peak water level was found to be evaluated on the safety side, and the authors were able to propose a practical water level prediction method based on the information engineering technique.