

RF 法を利用した石狩川下流部の水位予測の研究

Study of Water Level Prediction for downstream part of the Ishikari River Using the Random Forest Method

室蘭工業大学 ○学生員 坂本莉子 (Riko Sakamoto)
 室蘭工業大学 正員 中津川誠 (Makoto Nakatsugawa)
 室蘭工業大学 非会員 小林洋介 (Yosuke Kobayashi)
 室蘭工業大学 学生員 岡崎亮太 (Ryota Okazaki)

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。このような水害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、住民の避難といった防災・減災対策に要する予測先行時間（リードタイム、以下、LTと称す）の確保が望まれる。特に市街地に近い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影響するため、LTが長く、水位上昇のタイミングやピーク水位について精度のよい水位予測が望まれる。

現在、国土交通省北海道開発局の現業部門で主に用いられているのは予測雨量を入力条件に、星ら¹⁾が提案した貯留関数モデルで流量を計算し、水位に換算する予測手法が主流である。しかし、予測雨量、モデルパラメータ、H-Q式の精度など多くの誤差要因があり、必ずしも満足な結果が得られない場合がある。そのため近年では、機械学習手法により水位そのものを使用し予測する方法が注目されている。一言ら²⁾は深層学習による洪水時の水位予測手法を提案している。しかし、深層学習ではどの説明変数が結果に影響したのか分析を行うことが出来ない。そこで、岡崎ら³⁾は各説明変数の寄与度が算定できる機械学習手法の1つであるRandom Forest法（以下、RF法）による水位予測手法を提案した。また、RF法により各説明変数の寄与度を算出し、それらから回帰式を作成し水位予測を行う関連要因相関法を提案している。本研究では石狩川の水位予測に岡崎ら³⁾の考え方を適用し、RF法と関連要因相関法を用いて、LTの延長、既往最大水位を考慮した予測といった点で検討を行った。

2. 研究方法

2.1 対象地点と対象事例

対象河川は北海道西部に位置する一級河川石狩川とした。その中で最も重要な基準点であり、札幌市に近局的確な防災対策が必要である中下流部の石狩大橋地点を対象に、水位の予測を行った。図-1に予測対象地点の概略図を示す。石狩川は北海道全体面積(83,424km²)の約1/6の流域面積を占める日本でも有数の河川である。そのため石狩川本川のみではなく支川からの影響を考慮し、石狩川支川の主要河川である雨竜川・空知川・幾春別川・夕張川・千歳川からそれぞれ一箇所観測所を選定し予測に用いた。対象事例は2016年8月北海道豪雨を対象とした。また、既往最大水位である1981年8月事例を予測が行えるか検証した。水位予測を行う上で必要な要素として国土交通省所管のテレメータ観測所の水位デ

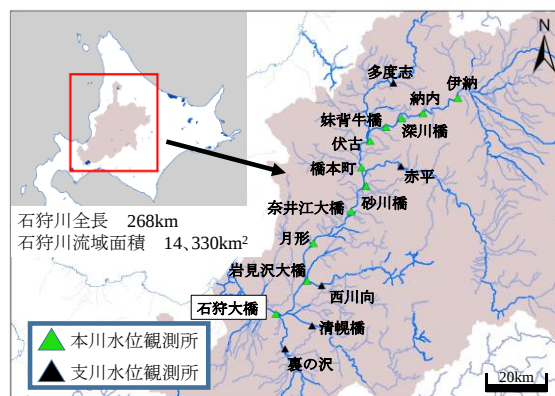


図-1 対象河川の概略図

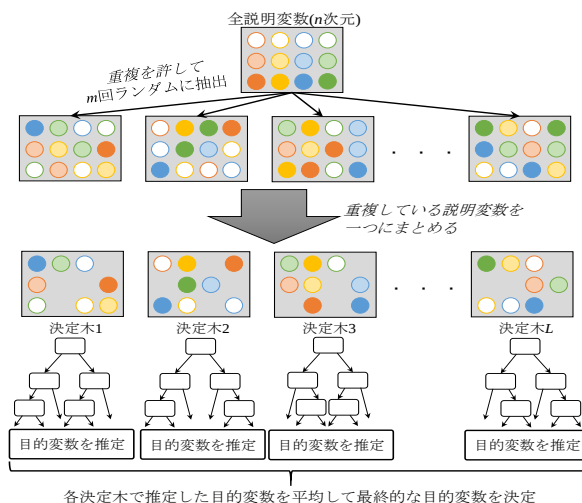


図-2 RF法の概略図

ータを使用した。

2.2 RF法による統計モデル

RF法とは、多数の決定木を組み合わせることで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。

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

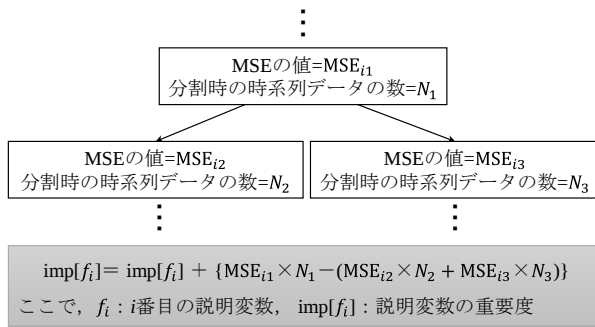


図-3 変数重要度算出方法の概略図

表-1 2016年8月予測のための説明変数の諸元

観測項目	観測地点	データ概要	避難に必要な時間 (LT)
水位 (m)	石狩大橋		
	岩見沢大橋		
	月形		
	奈井江大橋		
	砂川橋		
	橋本町		
	伏古		
	妹背牛橋	地点観測水位	①6~12時間前の水位
	深川橋	(テレメータ水位)	②12~24時間前の水位
	納内		
	伊納		
	千歳川	裏の沢	
	夕張川	清幌橋	
	幾春別川	西川向	
	空知川	赤平	
	雨竜川	多度志	

表-2 RF法で学習させた過去の洪水事例

	事例期間	ピーク値
事例1	2001/09/06 01:00 ~ 2001/09/21 24:00	6.28m
事例2	2005/08/16 01:00 ~ 2005/08/31 24:00	4.03m
事例3	2011/08/30 01:00 ~ 2005/09/14 24:00	4.97m

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。

- 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、最終的な目的変数を決定する。
- 4) 本研究では、全説明変数 n は LT=6h では $n=112$ 、LT=12h では $n=208$ であり、抽出回数 m は n と同数 ($m=n$) とした。決定木の本数 L は双方共に $L=2,000$ とした。

また、図-3 に示すような変数重要度 (以下、 $\text{imp}[f_i]$) を決定木の分割ごとに計算を行い、全ての $\text{imp}[f_i]$ の総和で各説明変数の $\text{imp}[f_i]$ 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変数間の従属性の影響を排除したよりロバストな説明力の指標となる。

2. 3 関連要因相関法による予測

前節の RF 法による統計モデルでは、表-1 で示すように入力した説明変数を全て使用し水位予測を行う。現

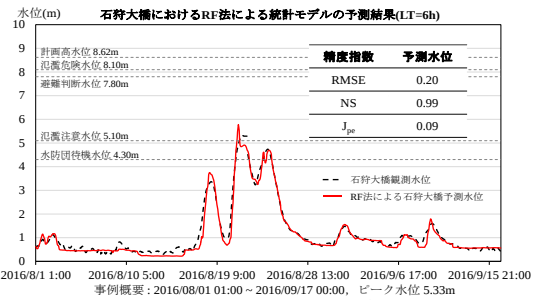


図-4 LT=6h の RF 法による統計モデルの水位予測結果

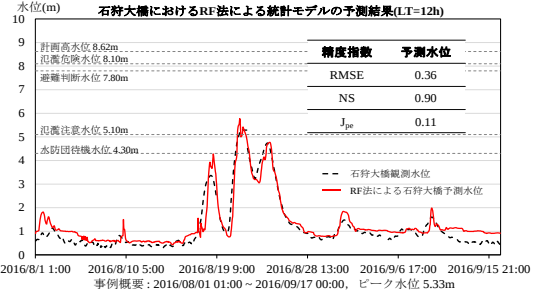


図-5 LT=12h の RF 法による統計モデルの水位予測結果

業の中で水位予測を行う場合、欠測等で表-1 に示す説明変数を全て使用することが困難な場合がある。したがって、表-2 に示す過去 3 事例の洪水を学習させた統計モデルの全説明変数の寄与度から、上位の説明変数を抽出して回帰式を作成した。作成する回帰式は過去 3 事例の平均回帰係数を用いることで、石狩大橋地点での水位予測を行った。

3. 結果と考察

3. 1 RF 法による水位予測結果

RF 法において、表-2 に示す過去の洪水 3 事例を学習させ、2016 年 8 月の洪水事例を予測した結果を LT=6h は図-4 に、LT=12h は図-5 に示す。また、本研究で用いた精度指数は以下に示す平均二乗誤差平方根 (以下、RMSE)、Nash-Sutcliffe 係数 (以下、NS 係数)、ピーク水位相対誤差 (以下、 J_{pe}) の 3 つを用いて行った。

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (H_{oi} - H_{ci})^2} \quad (2)$$

$$\text{NS} = 1 - \frac{\sum_{i=1}^N (H_{oi} - H_{ci})^2}{\sum_{i=1}^N (H_{oi} - H_{oave})^2} \quad (3)$$

$$J_{pe} = \frac{|H_{op} - H_{cp}|}{H_{op}} \quad (4)$$

ここで、 N はデータ数、 H_{oi} は降雨開始 i 時間後の観測水位(m)、 H_{ci} は降雨開始 i 時間後の計算水位(m)、 H_{oave} は期間中の観測水位平均値(m)、 H_{op} は期間中の観測ピーク水位(m)、 H_{cp} は期間中の計算ピーク水位(m)である。RMSE は、値が小さいほどモデルの精度が高いとされ、NS 係数は 1.0 に近づくほどモデルの精度が良く、0.7 以上でモデルの精度が高いとされる。また、ここでの J_{pe} は対象事例の期間中に観測された最大水位に対する値で

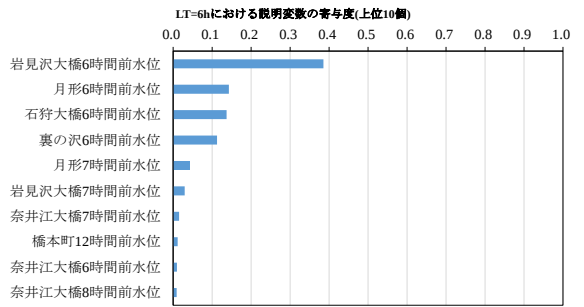


図-6 LT=6hの寄与度算出結果

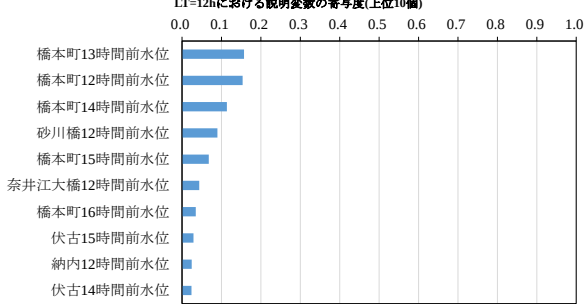


図-7 LT=12hの寄与度算出結果

表-3 関連要因相関法を用いた説明変数と重回帰式

LT	使用した説明変数		重回帰式
	H_1	H_2	
6h	岩見沢大橋6時間前水位	月形6時間前水位	$y = 0.15H_1 + 0.45H_2 - 1.91$
12h	橋本町12時間前水位	砂川橋12時間前水位	$y = -0.14H_1 + 0.76H_2 - 7.13$

ここで、 y ：石狩大橋地点における予測水位(m)

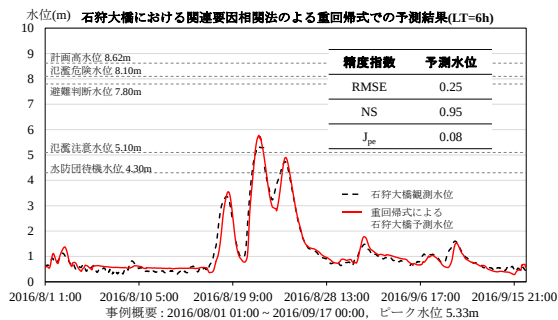


図-8 LT=6hの関連要因相関法による水位予測結果

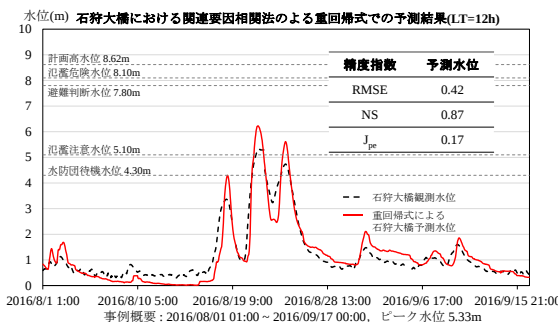


図-9 LT=12hの関連要因相関法による水位予測結果

ある。図-4、図-5に示す精度指数結果からもRF法による水位予測は高精度であることが分かる。

LT=6h, LT=12hそれぞれの寄与度算出結果を図-6、図-7に示す。寄与度の算出結果より、LTを延長することに応じて寄与度が高くなる観測所が石狩大橋観測所から上流にシフトしている。これは石狩川の洪水到達時間に応じて寄与度が高くなる観測所が変わることによると

表-4 1981年8月事例の説明変数の諸元

観測項目	観測地点	データ概要	避難に必要な時間 (LT)
水位 (m)	石狩大橋	地点観測水位 (テレメータ水位)	6~12時間前の水位
	岩見沢大橋		
	月形		
	奈井江大橋		
	橋本町		
	伏古		
	妹背牛橋		
	納内		
	伊納		

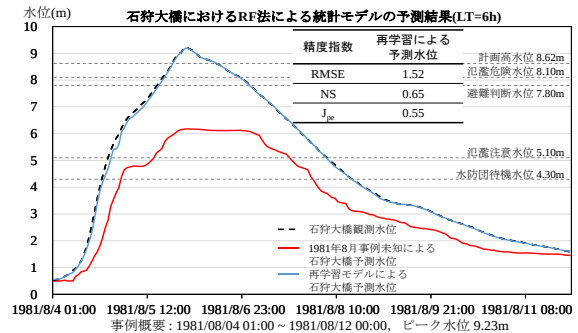


図-10 RF法による統計モデルの水位予測結果

考えられる。また、全説明変数の寄与度の和は1となるため、LT=6hで最も寄与度が高かった岩見沢大橋6時間前水位は水位予測を行う上で重要な役割を果たしていることが分かる。また、LT=12hでは橋本町が石狩大橋との相関が高く、水位予測を行う上で重要な役割を果たしていることが分かる。

3.2 関連要因相関法による水位予測結果

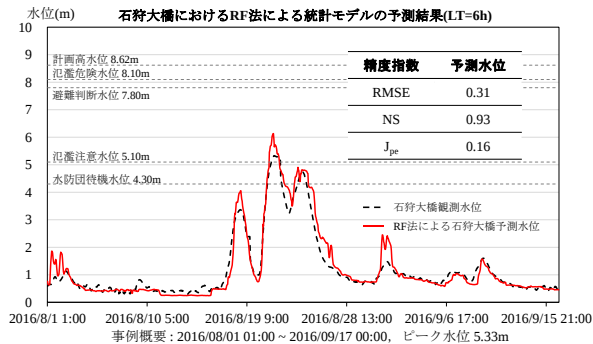
現業における実用性を考え、図-6、図-7の寄与度上位10個から2個の説明変数を抽出し、関連要因相関法による水位予測を行った。LT=6h, LT=12hでそれぞれ抽出した説明変数のパターンと、作成した重回帰式を表-3に示す。図-8、図-9にLT=6h, LT=12hの重回帰式による予測結果を示す。この結果、LT=12hでは低水位の予測精度は良好なものではないが、ピーク水位を安全側に予測する結果を得た。また図-8、図-9に示す精度指数では双方でRMSEの値が小さく、NS係数は0.7以上という高い精度の結果を得た。しかし、双方共に水位の立ち上がりに若干の遅れがあった。

3.3 1981年8月事例の水位予測結果と再学習

前節までは2016年8月事例の水位予測を行い、精度の良い結果を得た。それよりも水位の高い事例で予測が可能かを検証するため、既往最大水位事例である1981年8月事例で予測を行った。学習データは表-2を使用し、使用した説明変数は表-4に示す。説明変数が当該時刻から6~12時間前のデータを使用しているためLT=6hとなり、説明変数の数は $n=63$ となる。また、表-4の1981年8月のデータは表-1に比べ水位観測所が少なく限定的なものである。予測結果を図-10に赤線で示す。これを見ると予測水位は実測値を下回り、RF法による統計モデルでは学習したデータ以上の値を予測することが難しいことが示唆される。そのため、既往最大水位である1981年8月事例も含め再度RF法による学習を行い、よりロバストな統計モデルを作成した。これまでのRF法による統計モデルは表-2の事例データを学習し、学習していない2016年8月及び1981年8月の

表－5 再学習により追加した 1981 年 8 月事例の概要

	事例期間	ピーク値
事例4	1981/08/03 01:00 ~ 1981/08/12 24:00	9.23m



図－11 再学習後の RF 法による統計モデルの水位予測結果

水位を未知のデータとしてテストした。再学習モデルは、1981 年 8 月のデータが既知となるため、既往の最大値が算出される統計モデルになると考えられる。再学習モデルより追加した 1981 年 8 月事例の概要は表－5 に、それによる予測結果を図－10 に青線で示す。結果より、観測水位とほぼ一致するため、RF 法による統計モデルは学習したデータに対しては高い精度での予測が可能であることが示された。

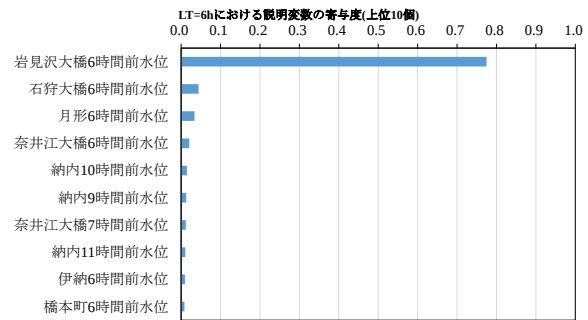
3. 4 再学習モデルによる 2016 年 8 月事例の水位予測結果

次に前節で作成した RF 法による再学習モデルにとって未知となる 2016 年 8 月事例の予測精度を検証する。学習させたデータは表－2 と表－5 を合わせた 4 事例とした。説明変数は表－4 に示しているものを使用した。RF 法による統計モデルの水位予測結果は図－11 に示す。1981 年 8 月事例を使用せずに学習を行った図－4 に比べピーク水位を安全側に予測できていることが分かる。なお、精度指標を比べると図－4 の方が良い評価となっているが、図－11 の結果も RMSE の値が小さく、NS 係数は 0.7 以上であり精度は良好であることが分かる。

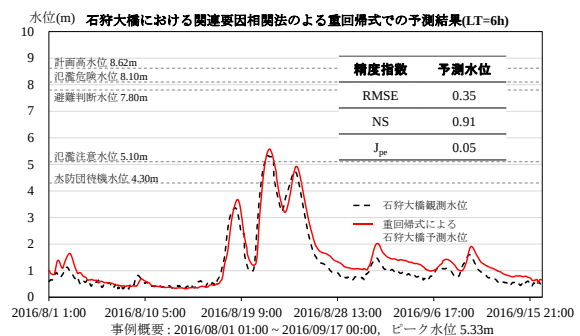
次に再学習モデルでの関連要因相関法による水位予測を行う。RF 法により算出した寄与度を図－12 に示す。1981 年 8 月事例を使用していない寄与度の算出結果（図－6）と比べ、順位に大差はないが岩見沢大橋 6 時間前水位の寄与度の値が大きくなり、重要な役割を果たしていることが分かる。ここから寄与度が高かった岩見沢大橋 6 時間前水位と石狩大橋 6 時間前水位を使用して以下に示す重回帰式で予測を行う。

$$y=0.51H_1+0.19H_2-0.32 \quad (5)$$

ここで、 y は石狩大橋地点での予測水位(m)、 H_1 は当該時刻 6 時間前の岩見沢大橋地点での観測水位(m)、 H_2 は当該時刻 6 時間前の石狩大橋地点での観測水位(m)である。上式を用いて予測を行った結果を図－13 に示す。1981 年 8 月事例を使用せずに関連要因相関法を行った図－8 に比べ、 J_{pe} は図－13 の方が向上しピーク値の予測が改善されている。また RMSE の値は小さく、NS 係数は 0.7 以上となっているため精度のよい予測結果が確認できた。一方、低水位では過大評価となった。また、3.2



図－12 再学習後の RF 法による寄与度算出結果



図－13 再学習後の関連要因相関法による寄与度算出結果

節で課題となっていた立ち上がりの遅れに関しては、多少の遅れはあるが図－8 に比べ既往最大値を取り組んだことで改善されたと考えられる。

4. 結論

本研究において得られた結果を以下にまとめる。

- 1) 石狩川のように長大な河川では上流部の水位データのみで下流水位の予測が可能であることが示された。
- 2) 既往最大水位をそれ以下の情報より予測することは難しいが、学習データとして使用することにより、ピーク水位や立ち上がりの遅れが改善され、より精度の高い結果が得られた。
- 3) RF 法を組み込んだ関連要因相関法による水位予測結果より、少ない説明変数で精度の良い水位予測をすることが可能であり、現業に実装しうる予測方法が提案できた。

謝辞：本研究を進めるにあたり、北海道開発局札幌開発建設部には洪水に関する資料を提供して頂いた。また室蘭工業大学データサイエンス研究室の山田恒揮氏には解析プログラムの提供等、多大なご協力をして頂いた。ここに記して謝意を表す。

参考文献

- 1) 星清，村上泰啓：小流域における総合貯留関数法の開発，第 31 回水理講演会論文集，pp.107-112，1987。
- 2) 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。
- 3) 岡崎亮太，中津川誠，小林洋介，谷口陽子：機械学習法に基づく 2016 年 8 月の常呂川連続洪水を対象とした水位予測の研究，土木学会北海道支部，第 73 号，B-19，2016。