

機械学習手法に基づく 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

Study of Water Level Prediction for Continuous Flood of Tokoro River in August 2016 Based on Machine Learning Method

室蘭工業大学 ○学生員 岡崎亮太 (Ryota Okazaki)
 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 谷口陽子 (Yoko Taniguchi)

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。道東の常呂川流域では、下流に位置する上川沿水位観測所において、計画高水位を超過し、観測史上最大の水位を記録した。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大きく影響するため、LT が長く、ピーク水位について精度の良い河川水位予測が望まれる。

河川水位の計算手法としては、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、洪水流出モデルで流量を計算し、水位-流量曲線(H-Q 曲線)から水位に換算する方法がある。大塚ら¹⁾は釧路川を対象とした水位相関法による水位予測が、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。本研究ではその考え方を拡張し、下流地点の水位に関する説明変数の抽出に機械学習を用い、水位予測の精度向上を試みた。これを「関連要因相関法」と呼称する。

関連要因相関法は、対象河川のピーク水位から ΔT 時間前の上流部の河川水文データを使用することにより、水位予測を行うものである。ここで、ΔT 時間前の上流部の河川水文データは水位予測における説明変数と考えられるが、多様な説明変数の中から対象河川の水位予測に適している説明変数を見つけ出すのは容易ではない。

一言ら²⁾は深層学習法に基づく洪水予測手法を提案しており、関連要因が水位予測に反映され精度のよい結果が得られるが、どの説明変数が結果に影響するかの分析ができない。そこで本研究では、説明変数の寄与度を算定できる機械学習法の一つである Random Forest 法（以下、RF 法）によって水位予測を試みる。さらに、現業での実用性を鑑み、寄与度の高い説明変数を用いて関連要因相関法による水位予測を行い、説明変数の選別、LT の延長といった面で検討を行った。以下に研究手順を示す。

- 1) RF 法を用いて、2016年8月の洪水事例を学習させ、同事例の予測モデルを作成した。ここでは、従来の洪水流出モデルによる水位予測法に対して、RF 法による水位予測モデルの精度検証を行った。
- 2) 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、2016年8月の洪水事例を予測し



図-1 対象河川の概略図

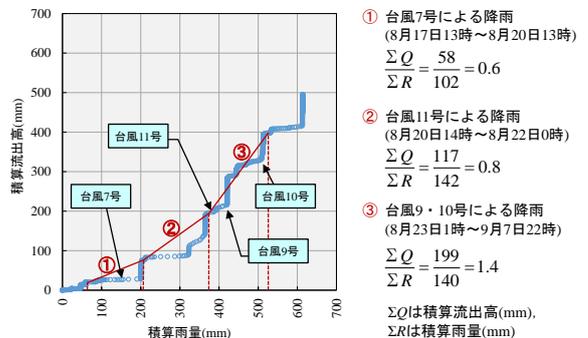


図-2 積算雨量-積算流出高曲線(R-Q 曲線)

た。ここでは、過去のデータのみで 2016 年 8 月の洪水事例を予測検証した。

- 3) 2)において算出された寄与度から、上位の説明変数を用いて重回帰式を算定し、LT 延長の検証を行った。

2. 研究方法

2.1 台風災害の概要

対象河川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地点とした。図-1 に対象河川の概略図を示す。この上川沿地点では、北見市街中心部付近の地点に位置しており、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、少量の雨量でも水位が上昇しやすい状態が続いた。図-2 に今回の災害においての上川沿地点における積算雨量-積算流出高曲線(R-Q 曲線)を示す。これより求めた各台

風の期間における流出率は図-2 に示すようになった。
 ③の台風期間では、積算流出高が積算雨量を上回る形になっており、流出率は 1.0 を超えていることが分かる。これは、上川沿地点における流出量が降雨による表面流出に加え、土壌水分状態が関係しているためと考えられる。よって本研究では、土砂災害発生リスクを判定する指標として用いられる土壌雨量指数(SWI)も水位を予測する上で重要な要素として考慮することにした。

また、本研究ではこの2016年8月4日~9月14日までの台風接近期間を対象事例とし、上川沿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。

2.2 予測に関係する要因

水位予測を行う上で必要な要素として以下の3つの要素が挙げられる。表-1 に用いた 62 次元の関連要因の概要を示す。

1) 水位データ

水位は、国土交通省所管のテレメータ観測所のものを使用した。

2) 時間雨量

時間雨量は、上川沿、北見、置戸地点の各観測所においては、国土交通省のテレメータ観測所の雨量を、常呂川流域平均雨量ではレーダ解析雨量をそれぞれ使用した。

3) 表層土壌雨量指数 SWI_s

土壌水分状態の評価手法として、気象庁が土砂災害発生の警報・注意報発令基準に採用している土壌雨量指数³⁾ (Soil Water Index, 以下, SWI) を用いた。SWIは時間降雨量を図-3 に示す直列3段タンクモデルの第1タンクに入力し、3つのタンクの貯留高の合計値をSWIとして求めている。本研究では表面流出に関係があると考えられる第1タンクの貯留高 S₁を「表層土壌雨量指数 (Soil Water Index surface)」とし、以下「SWI_s」と称する。なお、使用した時間降雨量はレーダ解析雨量による常呂川流域平均雨量とした。

2.3 RF法による水位予測

RF法⁴⁾とは、多数の決定木を組み合わせることで性能を高める機械学習アルゴリズムである。図-4 にRF法の概略図を示す。決定木とは、全データの集合に対して、複数の変数を組み合わせた分割をデータから自動的に生成したものである。また、元データの分布を問わず外れ値に対して頑健な予測ができる特徴を持ち、非線形なデータ分析に多用される。RF法では、学習ごとに m 本の決定木による予備実験によって i 個の最適な値を決定し、1つの決定木の作成に全説明変数から i 個ずつランダムに重複を許して選出を行うことで、n 本の決定木を作成する。その後、各決定木の目的変数の推定値を平均して、最終的な目的変数の値を決定する。

決定木は、各説明変数の寄与度を分割の良さを表す Gini 係数を用いて算出できる。したがって、RF法も各決定木の寄与度を統合して全説明変数の寄与度を分析できる。本研究では、m = 200, n = 2,000 として、表-1 の説明変数を使用することで RF法による上川沿の水位予測及び説明変数の寄与度の算定を行った。

2.4 関連要因相関法

前節の RF法では、入力した説明変数を全て使用して

表-1 関連要因の概要

| 観測項目 | 地点・流域 | データ概要 | 避難に必要な時間 (LT) |
|------------------------------------|-------------------------------------|---|---|
| 水位 (m) | 上川沿 太茶苗 忠志 北見 上常呂 置戸 | 地点観測水位 (テレメータ水位) | 6~10時間前の水位 |
| | 上川沿 北見 置戸 | 地点観測雨量 (テレメータ雨量) | 6~10時間前の雨量 |
| 時間雨量 (mm) | 常呂川流域 | 流域平均雨量 (レーダ解析雨量) | 6~10時間前の流域平均雨量 |
| SWI _s [※] (mm) | 常呂川流域 | 流域平均雨量(レーダ解析雨量)を入力した3段タンクモデルの1段目貯留高S ₁ | 6~12,24,36,48,60,72時間前のSWI _s |

※表層土壌雨量指数(Soil Water Index Surface)

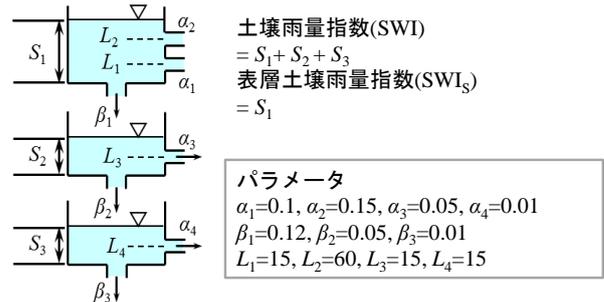


図-3 土壌雨量指数計算の模式図

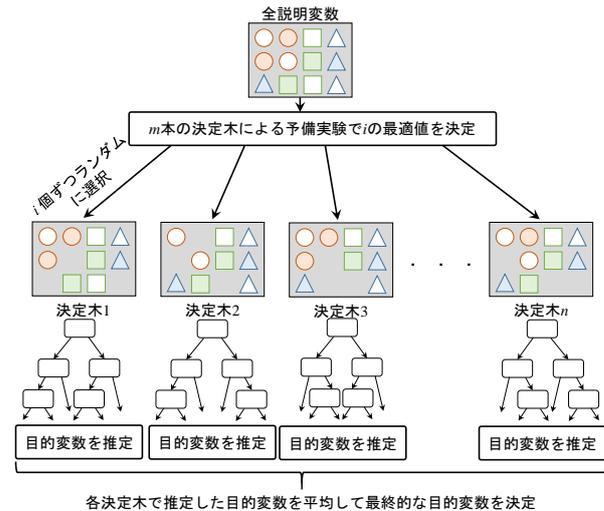


図-4 RF法の概略図⁴⁾

表-2 RF法で学習させた過去の洪水事例

| 事例期間 | ピーク水位 | 総雨量 |
|-----------------------------------|-------|---------|
| 事例1 2001/09/06 1:00 - 09/20 24:00 | 8.48m | 186.8mm |
| 事例2 2006/10/03 1:00 - 10/17 24:00 | 8.41m | 151.3mm |
| 事例3 2011/08/30 1:00 - 09/13 24:00 | 7.22m | 215.0mm |

水位予測を行っている。また、表-1 に示すように説明変数は当該時刻の6時間前のデータから使用しているため、LTは6時間となる。現業の中で水位予測を行う場合、表-1 に示す説明変数を全て使用するのは困難であり、LTも可能な限り延長することが望ましい。したがって、RF法において表-2 に示す過去3事例の洪水を学習させることにより算定した説明変数の寄与度から、上位10個の説明変数を抽出して重回帰式を作成した。その後、過去3事例の洪水の平均回帰係数を用いることで、

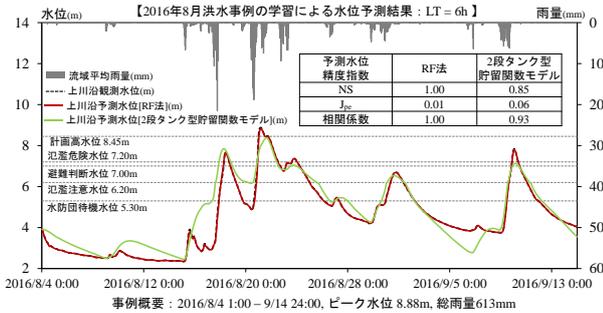


図-5 RF法(クローズドテスト)と洪水流出モデルによる水位予測結果の比較

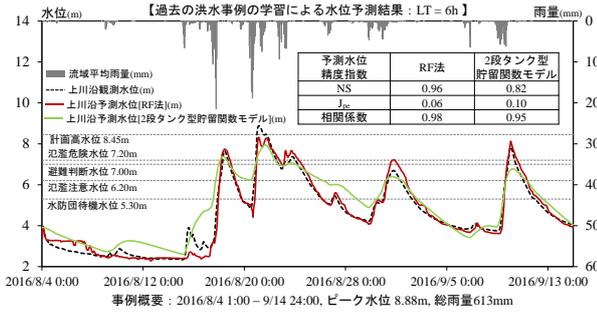


図-6 RF法(オープンテスト)と洪水流出モデルによる水位予測結果の比較

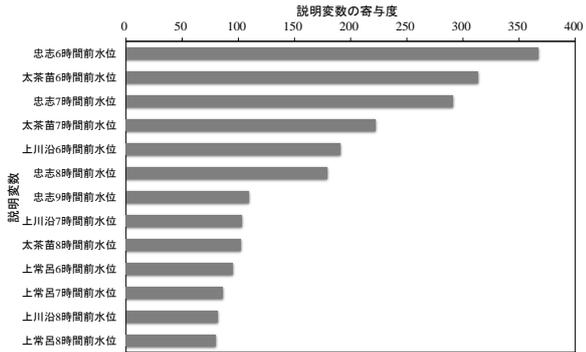


図-7 RF法(クローズドテスト)により算出した寄与度

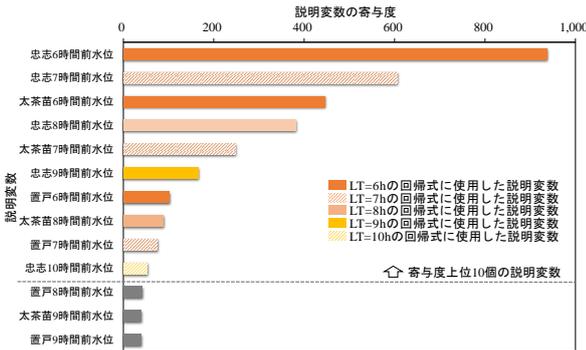


図-8 RF法(オープンテスト)により算出した寄与度

2016年8月洪水事例における上川沿での水位予測を行った。

3. 結果と考察

3.1 2016年8月洪水事例の学習による水位予測結果

RF法において、2016年8月の洪水事例を学習させ、その値を予測した結果を赤線で図-5に示す。これは、学習したデータを予測するクローズドテストの結果であ

り、予測モデルが「知っている」データに対する上限性能を求める評価になる。図-5よりクローズドテストによる対象事例の良好な予測結果は、RF法の適用性が的確に担保されていることを示すものである。また、本研究で用いた精度指標は以下に示す Nash-Sutcliffe 係数(以下、NS 係数)、ピーク水位相対誤差(以下、J_{pc})の2つを用いて行った。

$$NS = 1 - \frac{\sum_{i=1}^N (H_{oi} - H_{ci})^2}{\sum_{i=1}^N (H_{oi} - H_{ave})^2} \quad (1), \quad J_{pc} = \frac{|H_{op} - H_{cp}|}{H_{op}} \quad (2)$$

ここで、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_{ave}は観測水位平均値(m)、H_{op}は観測ピーク水位(m)、H_{cp}は計算ピーク水位(m)、Nは総降雨時間数である。NS係数は、1.0に近づくほどモデルの精度が高く、0.7以上でモデルの精度が高いとされる。

また、星ら⁵⁾が提案した「2段タンク型貯留関数モデル」は、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洪水流出モデルの一つであり、それを用いて流量を計算し、水位-流量曲線(H-Q曲線)から水位に換算した結果を図-5に緑線で示す。これは、対象事例の最適化パラメータを用いて水位を再現した結果であるが、LT6時間をもったRF法の予測結果よりもピーク水位を過小評価している。図-5に示す精度指数結果からも、RF法による水位予測が高精度であることが分かる。

3.2 過去の洪水事例の学習による水位予測結果

RF法において、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と精度指数を図-6に示す。これは、学習していないデータを予測するオープンテストであり、予測モデルが「知らない」データに対する性能評価である。図-6よりRF法の予測結果は、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。同様に、常呂川流域において過去の事例からパラメータを設定⁶⁾した2段タンク型貯留関数モデルによる水位予測の結果を図-6に緑線で示すが、ピーク水位においてRF法が安全側に予測していることが分かる。

これより、対象事例のような今までに経験したことのない洪水事例に対しても、RF法において過去のデータを学習することで予測が可能であることが示された。

3.3 RF法の寄与度算出結果

図-7に3.1節でのクローズドテストで得られた寄与度を、図-8に3.2節でのオープンテストで得られた寄与度の算出結果を示す。図には説明変数の寄与度の大きい順に、全説明変数の上位20%を示している。図-7、図-8ともに上位に各観測所の水位情報が設定されており、時間雨量、SWI_sよりも寄与度が大きいことが分かった。

3.4 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鑑み、図-8の寄与度上位10個から説明変数を抽出し、関連要因相関法によりLTの延長を試みた。この際、なるべく説明変数を少なくし、LTを6時間~10時間までとるように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の左側に示す。

表-3 関連要因相関法に用いた説明変数と重回帰式

| LT | パターン | 重回帰式に用いた説明変数 | | | | 重回帰式 | 予測水位の精度指数結果 | | |
|-----|------|--------------|-----------|----------|-----------------------|--|-------------|---------------|------|
| | | a_1 | a_2 | a_3 | a_4 | | NS | J_{pe}^{**} | 相関係数 |
| 6h | 1 | 忠志6時間前水位 | 太茶苗6時間前水位 | 置戸6時間前水位 | | $y=0.07a_1+0.7a_2+1.64a_3-349.22$ | 0.91 | 0.27 | 0.99 |
| | 2 | 忠志6時間前水位 | 太茶苗6時間前水位 | 置戸6時間前水位 | 12時間前SWI _s | $y=-0.47a_1+0.97a_2+1.44a_3+0.043a_4-298.42$ | 0.92 | 0.32 | 0.98 |
| 7h | 3 | 忠志7時間前水位 | 太茶苗7時間前水位 | 置戸7時間前水位 | | $y=-0.16a_1+0.6a_2+1.81a_3-387.82$ | 0.89 | 0.30 | 0.98 |
| | 4 | 忠志7時間前水位 | 太茶苗7時間前水位 | 置戸7時間前水位 | 12時間前SWI _s | $y=-0.5a_1+0.95a_2+1.56a_3+0.055a_4-323.75$ | 0.89 | 0.37 | 0.98 |
| 8h | 5 | 忠志8時間前水位 | 太茶苗8時間前水位 | | | $y=0.49a_1+0.71a_2-11$ | 0.89 | 0.23 | 0.97 |
| | 6 | 忠志8時間前水位 | 太茶苗8時間前水位 | | 12時間前SWI _s | $y=-0.29a_1+1.1a_2+0.064a_3-0.43$ | 0.86 | 0.34 | 0.96 |
| 9h | 7 | 忠志9時間前水位 | | | | $y=1.52a_1-24.79$ | 0.83 | 0.37 | 0.94 |
| | 8 | 忠志9時間前水位 | | | 12時間前SWI _s | $y=1.47a_1+0.027a_2-23.95$ | 0.79 | 0.48 | 0.94 |
| 10h | 9 | 忠志10時間前水位 | | | | $y=1.52a_1-24.73$ | 0.81 | 0.37 | 0.94 |
| | 10 | 忠志10時間前水位 | | | 12時間前SWI _s | $y=1.45a_1+0.033a_2-23.42$ | 0.78 | 0.48 | 0.93 |

※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8.88m(2016年8月21日23時, 台風11号上陸期間)に対する値である。

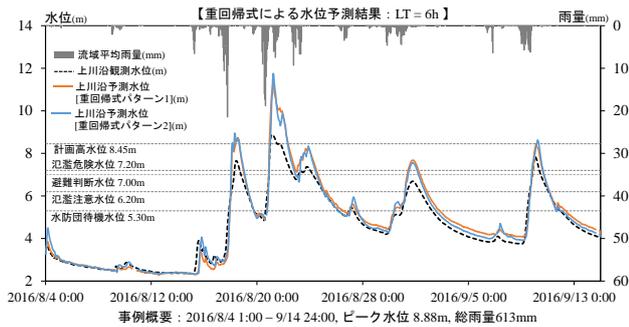


図-9 LT=6hの関連要因相関法による水位予測結果

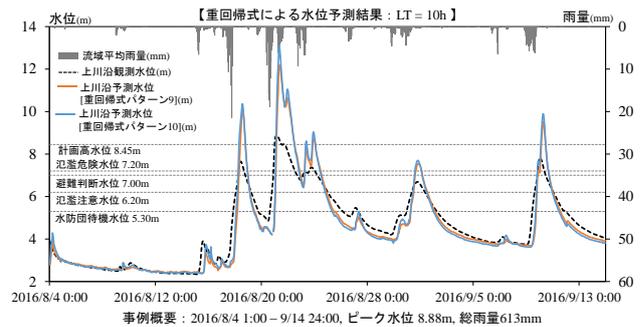


図-10 LT=10hの関連要因相関法による水位予測結果

なお、図-8に示す寄与度の高い説明変数にSWI_sは入っていないが、本研究では説明変数SWI_sの有無による水位予測精度の検証を行うため、比較的寄与度が大きかった12時間前のSWI_sを加えたパターンも考えた。

図-9にLT6時間の予測結果を、図-10にLT10時間の予測結果を示す。図-9よりパターン1、パターン2の重回帰式を用いることで、RF法よりも説明変数を少なく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。また図-10よりLTを10時間まで保持しても、ピーク水位を過小評価することなく予測できることが分かった。表-3の右側に関連要因相関法の精度指数を示す。これより、全パターンでNS係数0.7以上、及び相関係数0.9以上という良好な結果を得た。ただし、ピーク水位に対する精度指数である J_{pe} は、安全側に評価しているものの観測値よりも3m以上大きな値をとるパターンもあった。

また、図-9、図-10において、橙線はSWI_s無しのパターン、青線はSWI_s有りのパターンの水位予測結果を示しているが、両者に大きな差は見られなかった。これは、上流の水位情報にはSWI_sの影響も含まれているためと考えられる。ただし、上流に水位情報がない場合の予測に雨量やSWI_sが有用とも考えられるため、そのようなケースの検討は今後の課題としたい。

4. 結論

本研究において得られた結果を以下にまとめる。

- RF法のクロズドテストにおいて、精度の良い水位予測結果が得られたことから、従来の物理過程を考慮した手法のみでなく、統計的手法のRF法による水位予測の可能性が示された。
- RF法のオープンテストにおいて、水位の予測結果

に良好な結果が得られたことで、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も予測可能であることが示された。

- RF法を組み込んだ関連要因相関法の水位予測結果より、LTを10時間まで延長してもピーク水位を安全側に評価することが可能であることが分かった。

謝辞：本研究を進めるにあたり、国土交通省北海道開発局網走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、(一財)日本気象協会の臼谷友秀氏にはレーダ解析雨量の提供等で協力して頂いた。ここに記して謝意を表す。

本研究は、土木学会2016年8月北海道豪雨災害調査団の調査報告の一部である。

参考文献

- 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.
- 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.
- 気象庁; 土壌雨量指数: <http://www.jma.go.jp/jma/kishou/known/bosai/dojoshisu.html>(閲覧日 2016/9/12).
- 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.
- 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.
- (財)北海道河川防災研究センター・研究所: 実践流出解析ゼミ 講義テキスト編, 第12回, 2006.