

スパースモデリングによる大雨災害時のダム流入量予測モデル

室蘭工業大学 小林 洋介 大学院工学研究科 しくみ解明系領域 助教
室蘭工業大学 中津川 誠 大学院工学研究科 もの創造系領域 教授

1 洪水予測問題へのスパースモデリングの適応

河川における洪水を予測する手法には、大きく分けて上流下流の水位の相関関係などに基づき評価地点の水位を予測する手法である水位直接評価モデルと、降雨から流出流量を評価する流出解析技術を基盤とする手法である流出解析モデルに分類できる[1]。このうち、水位直接評価モデルは予測対象とする水位を目的変数、入力する情報を説明変数とする統計的な回帰モデルである。計測技術の向上とデータストレージの大容量化で、データ種の多様化と大量のデータの収集が可能となったことが、ニューラルネットに代表される多量のデータを用いることを前提とした学習アルゴリズムによる回帰モデルの性能向上に大きく貢献している。著者らも、機械学習的な手法を用いた河川やダムの洪水予測モデルを提案している[2]。

しかし、防災工学分野で考慮しなければいけない「これまでに経験のない未曾有の大災害」について、安全への余裕度を考える必要がある。統計的な回帰モデルによる洪水予測は、過去の洪水情報を学習するため、学習したデータの説明変数の値域内の未学習値を予測する問題が内挿問題であり、同じ未学習の値でも学習した説明変数の値域の外を予測する問題が外挿問題である。このため、洪水事例が無い河川やダムの洪水予測モデルは学習ができず、学習できたとしても既往最大値までしか予測性能を担保できず、未曾有の大災害の予測精度は理論的には不明である。スパースモデリングに基づく線形回帰モデルは外挿問題に対しては線形補完を学習範囲外に延長する線形外挿（直線外挿）となる。よって、外挿の予測区間の真値が学習データと大きく異なる非線形性を持たない限り、比較的妥当な予測となることが想定される。

ここで回帰モデルを定式化すると以下の式になる。目的変数である河川・ダムの水位 y は計測値である観測行列（収集データ） D と説明変数の重みベクトル w の積とモデルのバイアス b で表現される。

$$y = wD + b$$

観測行列 D は、データ計測種数である説明変数の重みベクトル w の次元数を M 、計測されたデータのレコード数 N とすると $M \times N$ 行列であり、 $M \leq N$ であれば解析的に解くことが可能である。しかし、新たな計測技術が開発された場合、開発以前のデータが手に入らず $M > N$ の劣決定問題となる。劣決定問題を解く方法に、真のモデルにおける説明変数の重みベクトル w のほとんどが0であり、非0の説明変数の数を K とすると、 $K \leq N$ であれば解くことができる。このような、ほとんどの説明変数の成分が0となると期待される性質をスパース性と呼び、このような変数の絞り込みは特に圧縮センシングと呼ばれている。

圧縮センシングの考え方を水位直接評価による洪水予測にあてはめると、多数の計測法・地点による説明変数の重みベクトル w の次元数 M うち、ある地点の予測に真に必要な計測法・地点に相当する説明変数の総数 K で、 $K \leq N$ となる場合、つまり、多数の手に入る観測情報のいくつかだけで予測モデルが説明できる場合にスパース性が成立することとなる。この特性はさまざまな計測を同時に行うことが可能となった最近の河川計測に対し相性が非常に良い。

また、各学習アルゴリズムが予測対象とする目的変数の水位などの値は、従来に起こったこともない洪水であり、統計的には異常値である。単純に水位を予測する問題であれば平時の低水位も洪水の高水位も予測できるのが最も良いモデルとなるので、収集できる限りの全データで学習するのが望ましい。一方で、洪水予測問題で予測したいのは高水位時の予測精度であり、高水位が適切に予測できるような学習が望ましい。以上の観点から、本稿では、まず、研究申請時に計画したダムの流入量予測問題にスパースモデリングと茶の機械学習手法との比較を行い、スパースモデリングの有効性を検証した。次に令和2年7月豪雨災害における被害が大きかった3河川の中流都市部での洪水予測問題にスパースモデリングがどのように適応可能か検証した。

2 スパースモデリングを用いたダムの流入量予測性能評価

2-1 研究対象のダム

本研究では、北海道の金山ダム、札内川ダム、豊平峡ダムの観測値を利用する(金山ダム、豊平峡ダムは制限水位方式であり、7/1~9/30の期間の値となる、札内川ダムはオールサーチャージ方式である)。ダムの位置と諸元を図1に示す。また、図2に各ダムの流入量を示す。金山ダムは2016年の台風に伴う大雨による流入量の極端な増加が見られた、他のダムも流入量が極端に増える頻度は全計測期間に対しスパーズである。

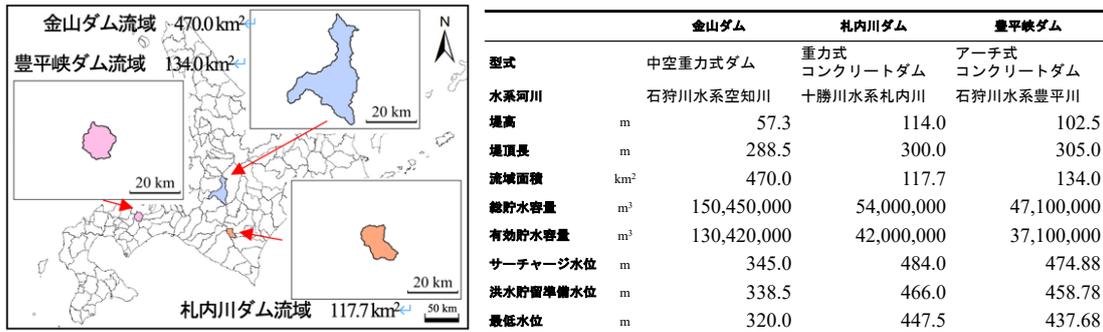


図1 研究対象のダムの諸元

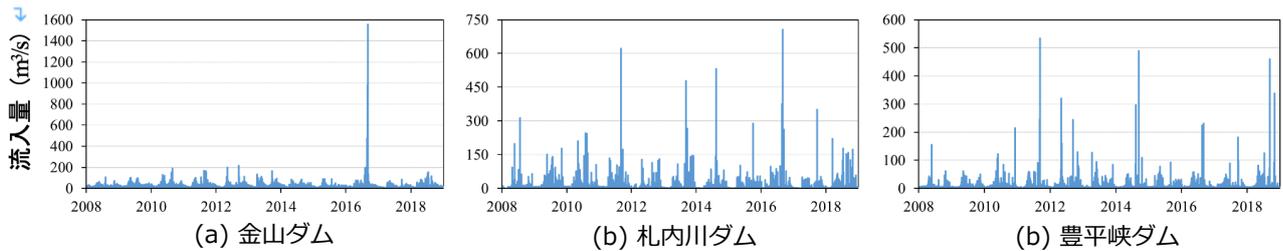


図2 各ダムの流入量

2-2 モデルに利用する変数

予測モデルの目的変数には、6時間後または12時間後までの積算流量(m^3)とした。説明変数には、現時刻から6時間前までの流量(m^3)、流域雨量(mm)、ダムのある地点雨量(mm)、SWI (Soil water index)に基づく土壌湿潤状態指標(mm)、レーダ解析雨量(mm)及び現時刻から予測対象となる6時間後または12時間後までのレーダ雨量を気象予報値として用いた。

収集した説明変数は過去情報5種の計測値に対して時間差が7パターンの35次元に、未来情報となる予測雨量が6時間後予測では6次元、12時間後予測では12次元となり、合計の説明変数は41次元または47次元となる。これらの説明変数に基づく情報のうち、レーダ解析雨量は(一財)気象業務支援センター、それ以外は水文水質データベースより収集した。これらの変数について2次の交互作用の計算を考慮するとモデル自体が学習する説明変数の数は6時間後予測では861次元、12時間後予測では1,128次元となる。

2-3 比較するアルゴリズムと評価指標

学習法の比較は、スパースモデリングの代表手法であるElastic Net、申請者らが過去に検討したランダムフォレストを用いた変数選択による線形回帰(RF-1r)、深層学習によるDNN(他の手法と同一のコンピュータで計算したハイパーパラメータ探索回数20回のDNN-20と産総研ABCIによる探索回数100回のDNN-100の2条件)を内挿条件と外挿条件で比較した。その結果を下記の表1から3に示す。

評価指標は2つあり、ピーク時間差 ΔT_p は観測値 T_{op} と予測値 T_{cp} 、最大流入量誤差 J_{pe} は観測流入量最大値 H_{op} と予測流入量最大値 H_{cp} を用いて、それぞれ以下の式で計算する。

$$\Delta T_p = T_{cp} - T_{op}$$

$$J_{pe} = \frac{H_{op} - H_{cp}}{H_{op}}$$

2-4 比較結果

表1から3にダムごとの結果を示す。太字は同一条件における最良値である。結果より、金山ダムと札内川ダムではおおむねスパースモデリングが最良であった。一方で、豊平峡ダムでは十分な学習時間をとったDNN-100が良くなる場合が見られた。しかし、DNN-100は数十時間の計算時間をかけて見つけた最良パラメータであるのに対し、Elastic netは300秒ほどでこの結果が得られており、未曾有の豪雨などの災害の現場でも予測結果を更新しながら最適な予測ができる可能性を示した。

表1: 金山ダムの方式別予測精度

評価指標		ΔT_p				J_{pe}			
学習方式		Elastic net	RF-lr	DNN-20	DNN-100	Elastic net	RF-lr	DNN-20	DNN-100
内挿	6時間流入量	4	4	4	1	0.17	0.17	0.40	0.33
	12時間流入量	7	6	7	7	0.34	0.40	0.36	0.39
外挿	6時間流入量	3	6	4	7	-0.01	-0.11	0.57	0.70
	12時間流入量	7	7	8	8	0.14	0.22	0.49	0.45

表2: 札内川ダムの方式別予測精度

評価指標		ΔT_p				J_{pe}			
学習方式		Elastic net	RF-lr	DNN-20	DNN-100	Elastic net	RF-lr	DNN-20	DNN-100
内挿	6時間流入量	1	2	3	3	0.12	0.06	0.48	0.45
	12時間流入量	-3	4	4	4	0.07	0.39	0.40	0.44
外挿	6時間流入量	1	1	1	1	-0.46	-0.92	0.30	0.18
	12時間流入量	1	2	5	6	-0.49	-0.71	0.01	0.17

表3: 豊平峡ダムの方式別予測精度

評価指標		ΔT_p				J_{pe}			
学習方式		Elastic net	RF-lr	DNN-20	DNN-100	Elastic net	RF-lr	DNN-20	DNN-100
内挿	6時間流入量	3	6	3	3	0.42	0.80	0.53	0.37
	12時間流入量	4	4	4	4	0.42	0.45	0.63	0.44
外挿	6時間流入量	2	6	3	3	-0.24	0.51	0.19	-0.20
	12時間流入量	3	8	8	8	-0.41	-0.40	0.46	0.38

3 スパースモデリングを用いた令和2年豪雨災害における中流都市部の水位予測

3-1 対象となる河川

a) 最上川

最上川は、源流を山形・福島県境の西吾妻山に発し、米沢市、山形市の両盆地を北流しつつ県内の大小支川を合流させながら、酒田市にて日本海へ注ぐ、幹川流路延長229 km、流域面積7,040 km²の一級河川である。予測は、令和2年7月豪雨で出水被害が報告された付近の大石田(副)観測所における水位予測を行う。

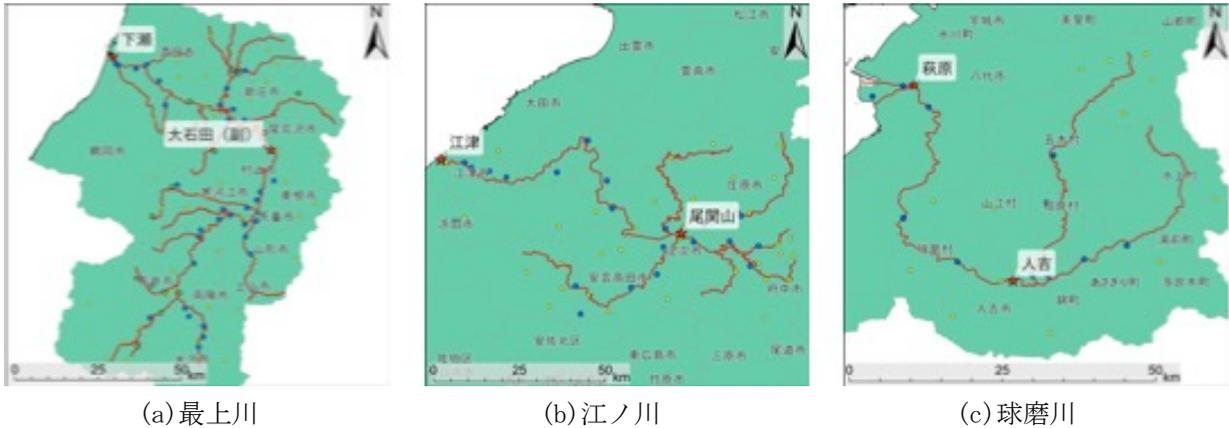
b) 最上川

江の川は、源流を広島県山県郡北広島町阿佐山に発し、広島県三次市で馬洗川、西城川、及び神野瀬川を合流し、島根県江津市にて日本海へ注ぐ、幹川流路延長194 km、流域面積3,900 km²の一級河川である。予測は、令和2年7月豪雨で出水被害が報告された付近で複数支川の合流後すぐの尾関山観測所における水位に対して行う。

c) 球磨川

球磨川は、源流を熊本県球磨郡銚子笠に発し、大小支川を合流させ人吉・球磨盆地をほぼ西に向かって貫流し、八代市にて八代海へ注ぐ、幹川流路延長115 km、流域面積1,880 km²の一級河川である。予測は、令

和2年7月豪雨で出水被害が報告された付近にあたる人吉観測所における水位に対して行う。



(a)最上川 (b)江ノ川 (c)球磨川
図4各河川の情報（青：水位観測所，黄：雨量観測所，赤：予測対象の水位観測所）

3-2 予測結果

ダムの例でも示したように，スパースモデリング手法は学習時間が短く，細かくモデル学習できることを踏まえて， ΔT_p と J_{pe} のリードタイム（予測先時間までのタイムラグで避難などに使える時間）別の精度を評価した。この評価ではスパースモデリング手法のElastic NetのほかにElastic Netの内部で重みつき平均されるLASSO回帰とRidge回帰を個別計算した場合と比較した。結果を図5から7に示す。

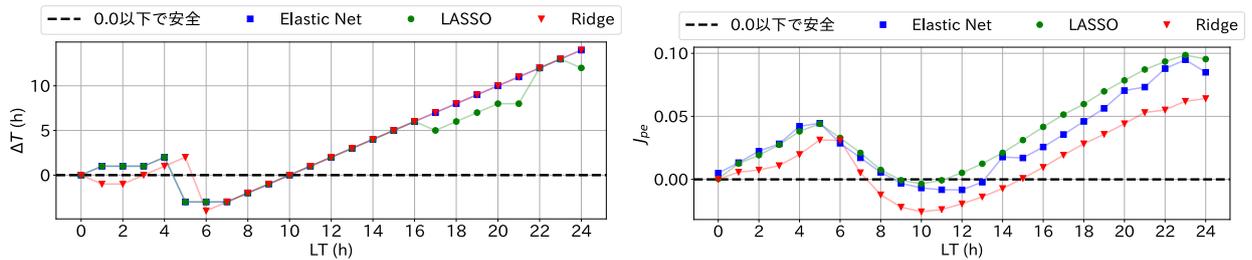


図5 最上川中流の大石田町（大石田副観測所）の ΔT_p と J_{pe} のアルゴリズム比較

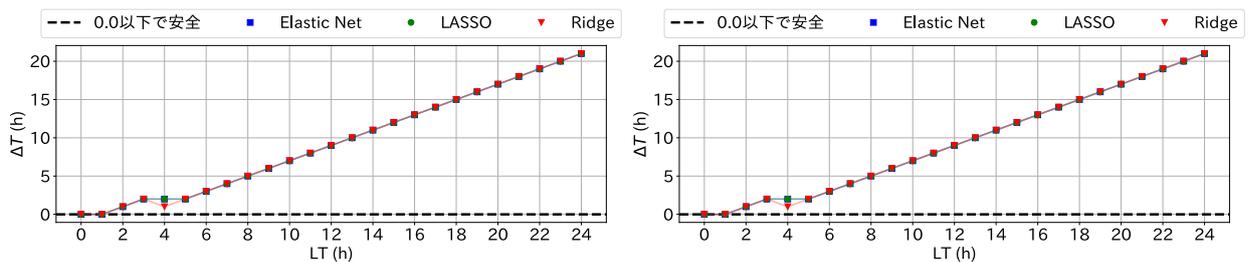


図6 江ノ川中流の三次市（尾関山観測所）の ΔT_p と J_{pe} のアルゴリズム比較

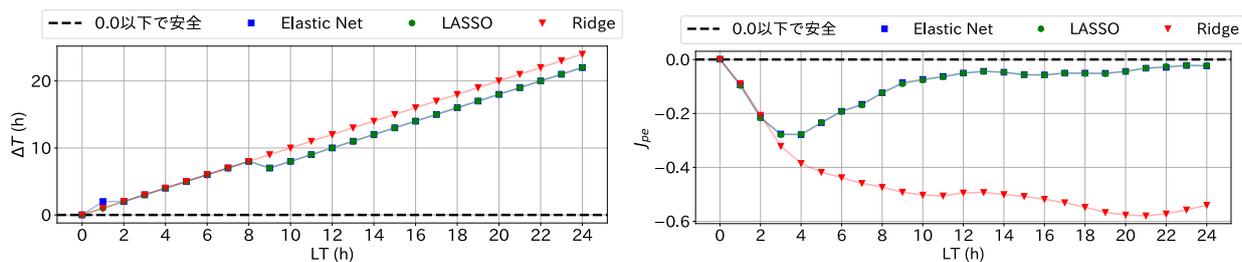


図7 熊川中流の人吉市（人吉観測所）の ΔT_p と J_{pe} のアルゴリズム比較

結果より、 ΔT_p はリードタイムが伸びるほどどのアルゴリズムも悪化するが、最上川の大石田観測所は5から9時間後の結果が安全側である。これは最上川はより上流の方にも都市部が多く観測所数が多いことが影響していると考えられる。一方で、三次市と人吉市の結果ではより上流の観測所が大石田より少ないことが影響していると考えられる。 J_{pe} に関しては大石田と人吉の結果はRidge回帰の方が一定時間を過ぎるとより安全側に出ているが、これはピークに対し過剰評価しているためであり、実測値に近い0付近になるElastic NetとLASSOの方が精度としては良いことがわかった。この結果の詳細は今後も評価・検討していく予定である。

【参考文献】

- [1] 椿良太, 小林健一郎, 内藤正彦, 谷口丞, “洪水予測技術の現状と課題について,” 河川技術論文集, 第19巻, pp. 1-6, 2013.
- [2] 山田恒輝, 小林洋介, 中津川誠, 岸上順一, “リカレントニューラルネットワークを用いた2016年の常呂川洪水事例の水位予測,” 土木学会論文集 B1(水工学), 74巻, 5号, pp. I_1369-I_1374, Nov. 2018.
- [3]

〈発表資料〉

題名	掲載誌・学会名等	発表年月
未経験事例に適用できる Elastic Net による 24 時間先までのダム流入量予測手法の提案	土木学会論文集 B1 (水工学), 76 巻, 2 号, pp. I_835-I_840	2020 年 11 月
スパースモデリング手法に基づくダム流入量予測の一般化の研究	土木学会 AI・データサイエンス論文集, 2 巻, J2 号, pp. 393-399	2021 年 11 月

