大雨災害時の北海道内ダム流入量予測へのスパースモデリングの適応とその評価

室蘭工業大学大学院 工学研究科　　　小　林　洋　介
室蘭工業大学大学院 工学研究科　　　渡　邉　真　也

大雨災害時の北海道内ダム流入量予測への
スパースモデリングの適応とその評価

Evaluation of sparse modeling to predict dam inflow for heavy rain disasters in Hokkaido

小林　洋介[1]・渡邉　真也[2]
Yosuke KOBAYASHI and Shinya WATANABE

[1] 室蘭工業大学 大学院工学研究科 助教
[2] 室蘭工業大学 大学院工学研究科 准教授

要　旨

河川やダムの洪水予測手法は大きく分けて，河川の水位やダムの流入量等を目的変数として直接予測する統計的回帰モデルと，降雨からの流出流量評価を基盤とした流出解析モデルがある．本稿では前者の統計的回帰モデルを用いたダム流入量予測問題を議論する．

河川の水位予測においては，下流の都市部等の水位予測には上流の計測値を説明変数として用いることが有効であることは論を俟たない．一方で，ダム流入量の予測ではダム周辺の観測情報しか説明変数に利用できず，どの変数を用いるのが良いか定めにくい問題がある．この様な問題の解決に，近年発展しているニューラルネットに代表される機械学習手法による多変量非線形回帰モデルが広く用いられている．しかし，本来は学習範囲内に目的変数の取りうる全ての値があることを前提とした内挿問題への最適化であり，防災工学分野で考慮しなければならない未経験の災害に対しての予測性能を保証できないという問題を抱えている．

これらの問題を解決するために，著者らはスパースモデリングの手法の一つであるElastic netをダム流入量予測に用いることを提案している．Elastic netは圧縮センシングと呼ばれるモデルへ入力する説明変数の選択により多数の説明変数からでもより良い予測モデルの学習が行われる．加えて，線形モデルであることから，外挿条件に対してもある程度整合的な解を出力することが期待される．

本稿では，北海道にある金山ダム，札内川ダム，豊平峡ダムの近年の流入量データを元に，ダム流入量予測に用いる学習アルゴリズムとしてElastic net と機械学習手法であるランダムフォレストを用いた線形回帰モデル，ニューラルネットを比較した．その結果，モデル学習に用いた最大流入量を超える外挿条件において，Elastic net はピーク時間の予測誤差が小さく，ピーク流入量差も安全側になることを確認し，提案の有効性を示すことができた．

《キーワード：ダム流入量予測；スパースモデリング；正則化；Elastic net》

1. はじめに

（1）統計的回帰モデルによる洪水予測

椿らによると，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することができると分類している[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．

統計的回帰モデルには，古典的な手法である線形回帰[*1]によるモデル化や，ここ数年の人工知能ブームを牽引する技術であるニューラルネット[*2]による予測（回帰）モデルといった手法がある．統計的な回帰モデルの洪水予測での実装例では，谷岡らによる水位情報と降雨情報を説明変数とした線形回帰モデル[3)]，一言らによるニューラルネットによる水位予測モデル[4)] がある．これらの手法は，提案時にケーススタディとして選択した河川とそれに基づく説明変数が異なり，学習アルゴリズム[*3]も異なるため，全く異なる手法と考えられがちだが，モデルへ入力する説明変数の重み付き線形結合による目的変数への回帰モデル（回帰式）という意味では変わりがない．特に，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[5),6),8)]．

ここで，回帰モデルを式に書き起こすと式（1）であり，目的変数である河川・ダムの水位 $\boldsymbol{y}$ は計測値である観測行列（収集データ）$\boldsymbol{D}$ と説明変数の重みベクトル $\boldsymbol{w}$ の積とモデルのバイアス b で表現される．

$$\boldsymbol{y} = \boldsymbol{w}\boldsymbol{D} + b \tag{1}$$

観測行列 $\boldsymbol{D}$ は，データ計測種数である説明変数の重みベクトル $\boldsymbol{w}$ 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，$M \geq N$ 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 $M < N$ となり，劣決定問題と呼ばれ，基本的な線形代数の問題としては解けない問題となる．

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 $\boldsymbol{w}$ 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，$M \geq K$ 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，スパース性に基づいた統計モデリングをスパースモデリングと呼ぶ．水位直接評価モデルによる洪水予測に置き換えると，多数の計測法・地点による説明変数の重みベクトル $\boldsymbol{w}$ 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，$N \geq K$ 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．この処理は圧縮センシングとも呼ばれる．

例えば，従来の水位直接評価モデル設計において設計者の勘と経験に基づいた下流の水位に基づく変数は上流の水位予測へ影響しないからモデル入力に使わないという判断は，圧縮センシングにおいて下流の水位の重みが0となることと等価である．言い換えるとスパースモデリングの圧縮センシング処理は，設計者の勘と経験に基づく判断の自動化であり，人間が関与することによる主観的な判断による見

[*1] 慣例的に説明変数が 1 変数の場合を単回帰，多変量解析の場合を重回帰と呼ぶが本稿では総称して線形回帰と呼ぶ．

[*2] 2006 年に Geoffrey Everest Hinton らが提案したスタックドオートエンコーダ[2)]以降をディープラーニング（深層学習）に基づいたニューラルネットと区別している例も多い（例えば参考文献[4)]）が，本稿では総称してニューラルネットと呼び，本稿での実装例は近年用いられるディープラーニングに基づいたニューラルネットによる実装のみ考慮した．

[*3] 回帰モデルの重みを最適化する方法の呼称には，推計統計学に基づけば推定手法，ベイズ統計学的な機械学習手法に基づけば学習手法と分野によって呼び名が変わるが，本稿ではモデルの重み最適化アルゴリズムという意味で学習アルゴリズムという語を用いる．

落としを防ぐメリットがある．

河川の各地点における洪水予測であればこのような勘と経験に基づいて上流のみ入力とするモデル設計をすることが可能である．一方で，ダムの貯水位や流入量に関しては，周辺の雨量など様々な観測データが入手可能になった反面，入力とする説明変数に何を選択すれば良いかの基準が河川よりは不明瞭であり，圧縮センシングによる予測モデルの説明変数の選択が必要である．

統計的回帰モデルに基づいたダムの流入量予測での説明変数の選択について，河川やダムでの洪水予測では，計測技術の進化に伴い，様々な観測値を説明変数として用いることができる．さらに，水文水質データベース[9]のように計測データのデータベースが整備されてきており，収集されているデータは1時間単位でありそのデータレコード数Nは十分に多い．

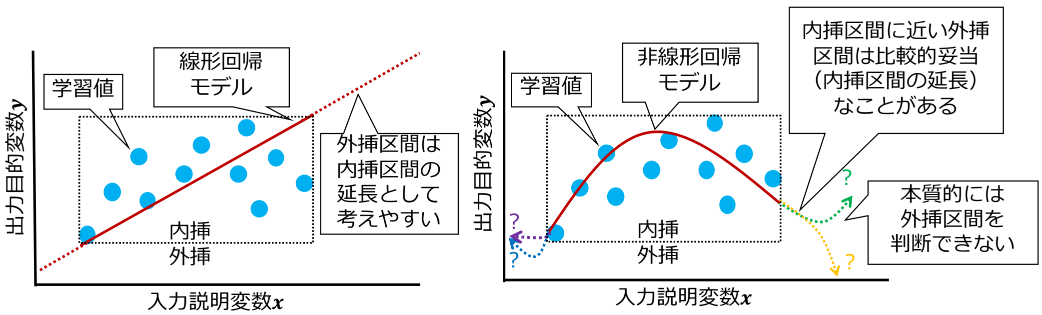
このようなビッグデータを利用し，著者らは時系列を考慮したリカレントニューラルネットワークでの洪水予測の提案した．この提案では，年単位の時系列データを学習し，通常のニューラルネットワークによる水位予測より高精度になることを示した[6]．ただし，多くの洪水予測の研究（例えば先行研究[4]）においては，洪水の事例期間のみ学習することが多く，学習に用いるレコード数Nは十分に多いとはいえない．このような事例期間のみを学習する場合は，前述のスパースモデリングによる洪水予測が有効であることが想定され，著者らも北海道内のダムの洪水事例で十分な予測精度があることを確認した[7),8)]．

（2）内挿問題と外挿問題

以上の検討より，多数のデータレコード数を学習に利用できる場合にはニューラルネットが，少数のデータレコード数であればスパースモデリングが良いという傾向にあることがわかってきた．ここで，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．これまでに述べてきた様に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．統計的な回帰モデルにおいて，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．

図 1 に内挿・外挿問題について，線形モデルと非線形モデル別にイメージを示す．この図は簡単のために2次元とした．線形回帰モデルの代表的学習アルゴリズムがスパースモデリングであり，非線形回帰モデルの代表的学習アルゴリズムはニューラルネットである．

図に示すように，ニューラルネットに代表される非線形回帰モデルは関数近似手法だが，学習データに対する内挿問題を解くことが前提で，外挿区間の予測値は内挿問題である学習データとの関係が説明できないため，計算上求まるが妥当性は保証されない．言い換えると，洪水事例が無い河川やダムの洪水予測モデルはニューラルネットで最適な学習ができず，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．この問題への対策として，例



(a) 線形モデルの内挿と外挿　　(b) 非線形モデルの内挿と外挿

図1　内挿と外挿の例

えば一言らの検討では，目的変数を直接水位とせず水位の差分とすることで，直接水位としては外挿になる場合も水位差分としては内挿となる様な工夫をしている[4]．一方で，経験的にはニューラルネットでも洪水予測に利用可能な程度の予測精度である関数が学習されることが多く，著者らの検討でも学習水位を上回る水位を予測することができている[6]．

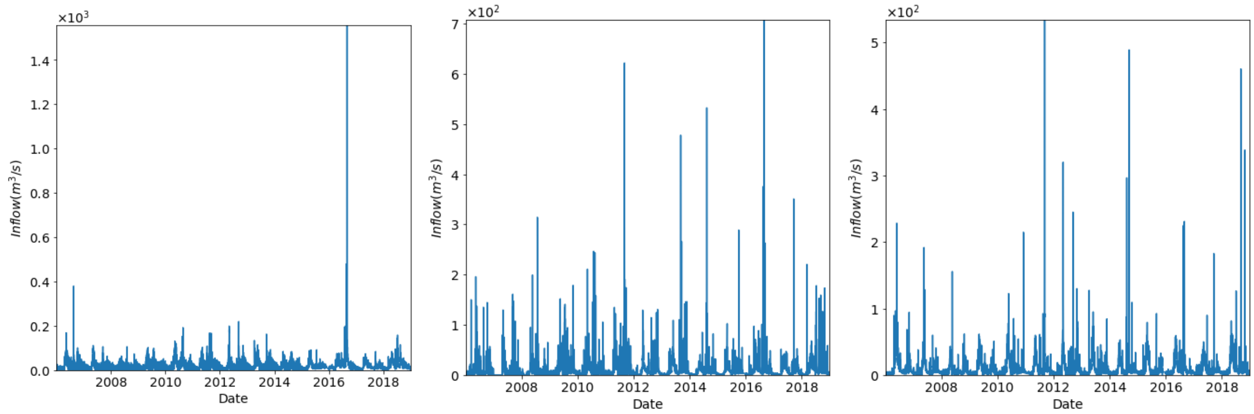
また，スパースモデリングに基づく線形回帰モデルも厳密には外挿条件の予測精度は担保できないものの，**図 1**(a) に示した様に外挿問題に対しては線形補完を学習範囲外に延長する線形外挿（直線外挿）となる．よって，外挿の予測区間の真値が学習データと大きく異なる非線形性を持たない限り，比較的妥当な予測となることが想定される．この際に，多数の説明変数による予測モデルやニューラルネットの様に複雑な重みベクトル $\boldsymbol{w}$ となる予測モデルであれば外挿区間の妥当性を検討することが難しいが，少数の説明変数へと変数選択された線形モデルであれば，個々の変数の外挿区間における妥当性の検討も可能となる．

（3）学習データの制約

本稿では，ニューラルネットでもモデル化可能な多数の説明変数と年単位のデータレコード数を学習したダムの流入量予測モデルについて，内挿・外挿条件での予測精度を複数の学習アルゴリズムを比較検討する．この際に，収集したデータレコード数 N が説明変数の次元数 M に対し $N \geq M$ となる様に学習データを収集する．

ここで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ことを想定する必要がある．これは，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事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学習アルゴリズムとしては望ましい．しかし，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様な学習法が望ましい．**図 2** に本稿のケーススタディで用いる 3 つのダムの流入量を示す．図より 10 年を超える計測期間に対し，洪水自体の発生頻度は数年に 1 度とスパースであり，有意味なデータレコード数 N_m は $N_m < M$ と想定される．

よって，この様な $N_m < M$ が想定される様な学習データに対する洪水予測性能とスパースモデリングと特にニューラルネットによる予測を比較することで最適な回帰モデルの学習法について提案する．特に学習に利用するデータや前処理を各学習方式で全く同一とすることで，アルゴリズムの違いによる結果解釈や予測性能差を明確化することを目的とする．



(a) 金山ダム　(b) 札内川ダム　(c) 豊平峡ダム

図2　本稿のケーススタディで扱うダムごとの流入量

2. 統計的回帰分析手法

本稿で比較に用いる統計的回帰分析の手法について述べる．まずベースとなる古典的な線形回帰モデルについて説明し，次に機械学習的な変数選択法（圧縮センシング）を述べ，その後にスパースモデリングとニューラルネットによる回帰モデルについて説明する．

（1）線形回帰モデル

洪水予測に統計的回帰モデルを用いる場合は，モデルの出力となる評価地点の水位・流量などとなる目的変数の計測値（実測値，真値のうち計測された値）を $\boldsymbol{y}$，予測モデルの出力である予測値を $\boldsymbol{y}_m$ とし，モデルの入力となる各種計測対象である説明変数 $\boldsymbol{x}$ との関係は，説明変数の重みベクトルである $\boldsymbol{w}$ と切片（ノイズ成分も含むバイアス）の b を用いて，以下の式（2）で一般化できる．古典的な線形回帰モデルは，式（2）の説明変数の重みベクトル $\boldsymbol{w}$ と切片の b を最小二乗法（OLS：Ordinary least squares）で推定[*4]し，全ての説明変数から予測に必要な説明変数のみを選択する圧縮センシングのために情報量基準を用いる．

$$\boldsymbol{y}_m = \boldsymbol{w}\boldsymbol{x} + b \tag{2}$$

a）重みの決定：最小二乗法

線形回帰モデルでは，式（2）の説明変数の重みベクトル $\boldsymbol{w}$ と切片の b を最小二乗法で推定する．このモデルの目的変数の計測値ベクトルを $\boldsymbol{y}$，予測値ベクトルを $\boldsymbol{y}_m$ として二乗誤差は式（3）であり，この誤差のデータ数 n の和を最小とする $\boldsymbol{w}$ と b を推定することが最小二乗法である．式（3）の $\boldsymbol{y}$ は誤差計算による $\boldsymbol{w}$ と b の推定に用いる既知の値で，予測値に対する正答として用いられることから，教師（データ）と呼ばれる．また，最小二乗法は線形回帰モデルの学習における損失関数[*5]と呼ぶ．

$$\begin{aligned} J(\boldsymbol{w}, b)_{\mathrm{OLS}} &= \sum_{i=1}^{n} (y_{(i)} - y_{m(i)})^2 \\ &= \sum_{i=1}^{n} \{y_{(i)} - (\boldsymbol{w}x + b)_{(i)}\}^2 \end{aligned} \tag{3}$$

この式（2）を最小化するには，$\boldsymbol{w}$ を変数として偏微分した式を 0 として $\boldsymbol{w}$ の次元数 m 元の連立方程式とし，これを解くことで $\boldsymbol{w}$ の推定値 $\hat{\boldsymbol{w}}$ を得る．切片の b は $\boldsymbol{w}$ の推定値が求まれば，入力変数と説明変数の算術平均を用いて推定できる．

b）従来型の変数選択：情報量基準

線形回帰モデルでは，入力に用いる説明変数ベクトル$\boldsymbol{x}$から予測に必要な説明変数のみを選択し，必要最小な数の変数での予測つまり理想的な統計モデル式の推定を目指す．これは「ある事項を説明するために，必要以上に多くを仮定することはない」というオッカムの剃刀と呼ばれる法則に則っている．不必要に多数の説明変数によるモデルは，式（3）を（たまたま）最小化するためだけの特異な $\boldsymbol{w}$ と b が推定されることが起こりうる．洪水予測モデルでオッカムの剃刀に基づいて変数選択（圧縮センシング）を行うということは，予測モデルに入力する説明変数となる計測地点や計測物理量等の計測種を最適化することであり，予測モデルの解釈性の向上や不要な計測の削減に寄与する．

[*4] $\boldsymbol{y}$ が真値のうち計測された一部の値である限り，真のモデルパラメータである $\boldsymbol{w}$ と b は決定できないため，$\boldsymbol{w}$ と b は推定することしかできない．しかし，実用的には観測データ $\boldsymbol{y}$ から $\boldsymbol{w}$ と b を決定するような処理を行っている．

[*5] 目的関数，評価関数などとも呼ぶ．

変数を選択するための基準として情報量基準があり，赤池の情報量基準（AIC：Akaike's information criterion）やベイズ統計基準（BIC：Bayesian information criterion）などがよく知られている．情報量基準は，統計モデルがどれだけ説明しているか（適合性）と，どれだけ簡潔であるか（簡潔性）とを同時に評価する．AIC と BIC は適合性と簡潔性のバランスの取り方が異なるが基本的な処理は共通する．

しかし，情報量基準は説明変数 $\boldsymbol{x}$ の次元数 M に対し全ての組み合わせ 2^M を評価する必要がある．説明変数の数M は洪水予測問題に置き換えれば計測可能な物理量・地点数の総数であり，1 節で述べたように計測データ数 N との間に $N \geq M$ がなければ情報量基準が求められない．さらに線形回帰モデルでは，後述する説明変数の交互作用について，最低限の 2 次の交互作用まで求めると説明変数は 1 次の変数の数を M_1 とした際に $M = {}_{M_1}C_2 + M_1$ 次元まで増えるため，評価総数はさらに増えてしまう．よって，計測技術が大幅に発展した現在では，情報量基準はデータ数と計算量が共に膨大となり，必ずしも良い変数選択法とならない．

（2）機械学習的手法に基づく変数選択：ランダムフォレストによる説明変数重要度の利用

機械学習アルゴリズムを用いたデータ分析の分野では，変数選択に回帰分析に利用可能な決定木を複数組み合わせるランダムフォレストの重要度を変数選択に用いる実践例がある[10]．**図 3** に決定木の例を示す．決定木は木構造を持ち，出発点となるルート（根）から条件分岐を行うノード（幹）を経由し，最終的な予測値である $\boldsymbol{y}_m$ が入るリーフにたどり着く．一方で，同じ分布に従う学習データを用いても**図 3** に二例示したように異なる決定木が学習される問題[*6]がある．この問題を解決するために，学習時の説明変数の数を木ごとにランダムに選択し多数の決定木の出力を平均するランダムフォレスト[12]や，勾配ブースティングを用いた決定木[13]などが提案されている．

しかし，決定木は**図 3** の予測値 $\boldsymbol{y}_m$ で示すように離散的な値の集合である．加えてこの $\boldsymbol{y}_m$ は学習された値のみであり，教師となる学習データ $\boldsymbol{y}$ に予測結果が全て含まれていないといけない．これは決定木が「観測データが代表する真のデータの分布は大数の法則に基づいて正規分布する」という従来からの統計学の前提に従わず，観測データの分布が真のデータの分布と完全一致し外挿を考慮しないという哲学でアルゴリズムが設計されていることに依拠している[*7]．洪水予測問題に置き換えると，学習データに含まれている洪水事例の最大水位までしか予測できないことと等価であり，未曾有の大災害に相当する水位値は予測結果を出力できないという欠点となる．

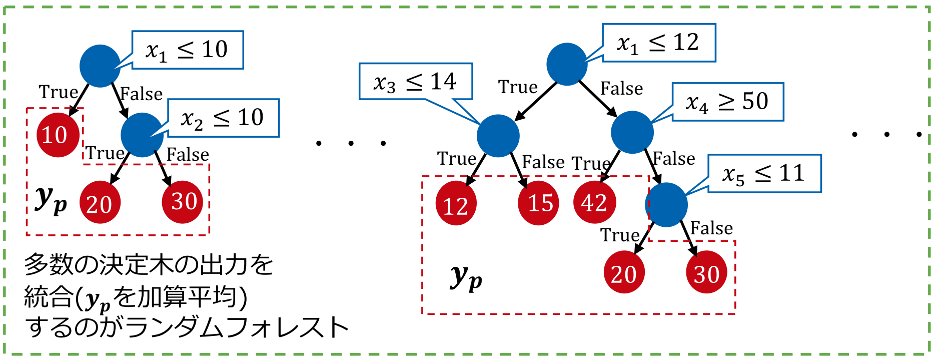


図3　決定木の例：説明変数を用いたルートとノード（青）での分岐の結果，目的変数の値が入るリーフ（赤）にたどり着くのが基本構造だが，二例を示すように同じ分布に従う学習データでも異なる決定木が学習されることがある

*6　決定木の学習についての詳細は参考文献[11]などを参照されたい．

*7　この前提に基づくため，決定木の学習は計算も早く，実用的なデータ分析に多用されるようになった．

一方で，決定木は**図 3** のようにルートとノードでの分割条件でどの程度の学習データが分割されるかを定量化することが可能である．各分割条件は説明変数ごとの不等式であるため，多くのデータを分割する説明変数は重要な説明変数と考えることができる．特にランダムフォレストの計算アルゴリズムでは，観測行列（収集データ）$\boldsymbol{D}$と説明変数ベクトル $\boldsymbol{x}$ をランダムサンプリングした $\boldsymbol{D}_s$ と $\boldsymbol{x}_s$ より決定木を多数生成する．これにより個々の決定木で $\boldsymbol{D}_s$ のデータ数 N_s と $\boldsymbol{x}_s$ の次元数 M_s の間に $M_s \geq N_s$ となり決定木ごとのルートとノードの分割条件が最適化されていると考えられ，分割条件に基づいた説明変数の重要度も最適化されているとみなせる．全決定木の説明変数の重要度を平均すれば，ランダムフォレストの決定木学習においては重要であった説明変数の重要度とみなせる．**図 3** の例で説明すると，二つの決定木のルートの分割条件に使われた説明変数$\boldsymbol{x}_1$はどちらも全データの分割に影響し，複数の決定木で登場することから重要な説明変数となる．

先に述べた様に決定木は外挿条件が予測できないため，重要度で上位となった説明変数を選択して線形回帰モデルを学習する．線形回帰モデルは決定木と異なり，外挿区間に対しては線形補完による線形外挿（直線外挿）となるため，外挿条件となる未曾有の大災害に相当する水位予測値も計算上は出力可能となる．著者らのランダムフォレストの説明変数重要度を用いた変数選択による線形回帰モデル[*8]として学習した河川水位予測モデルの評価では，最大54の説明変数から 3 変数を選択することで十分な予測が可能であることを示した[5]．

（3）正則化に基づく変数選択：スパースモデリング

（2）節で用いた変数選択は最終的に用いる線形回帰モデルと説明変数の選択に用いるランダムフォレストアルゴリズムとに特に関係はなく，説明変数の重要度が計算される別の学習アルゴリズムの結果を利用している実用上計算しやすい手法である．最終的に用いる線形回帰モデルの学習の際に従来の情報量基準と同様の線形回帰モデルを前提とした変数選択を行う手法として，近年着目されているのが正則化を用いたスパースモデリングである．

正則化とは，線形回帰モデルの重みを決定する式 (3) の損失関数である J（$\boldsymbol{w}, b$）$_{\mathrm{OLS}}$ に説明変数の重みベクトル $\boldsymbol{w}$ から計算されるペナルティ項となる正則化項を追加したコスト関数[*9]でモデル学習を行うことである．これにより，最小二乗法による説明変数の重みベクトル $\boldsymbol{w}$ の評価だけでなく，正則化項をも最小化する $\boldsymbol{w}$ が選択される．つまり，特定の説明変数 $\boldsymbol{x}_{(n)}$ の重み $a_{(n)}$ がゼロとなれば，その説明変数は選択されないこととなる．本稿では正則化を用いた回帰手法としてリッジ回帰（Ridge regression）[14] と LASSO 回帰 (Least absolute shrinkage and selection operator regression)[15] を説明し，これらを統合した Elastic net[16] の洪水予測問題への利用を提案する．

a）$L2$ 正則化：リッジ回帰

リッジ回帰で用いる $L2$ 正則化項を用いたコスト関数 J（$\boldsymbol{w}, b$）$_{\mathrm{Ridge}}$ は $L2$ 正則化項を用い，式 (4) で示す．$L2$ 正則化項は説明変数の重みベクトル $\boldsymbol{w}$ の $L2$ ノルム，つまりユークリッド距離を対象とし，式 (4) のコスト関数を最小化することで回帰モデルを推定するのがリッジ回帰である．ここで，λ は損失関数の値を決めるハイパーパラメータ[*10]である．λ の値が大きい場合，コスト関数 J（$\boldsymbol{w}, b$）$_{Ridge}$ を最小

[*8] 文献[5] では関連要因相関法と呼んだが，本稿で検討するスパースモデリングも同様に変数選択による圧縮センシングを利用するため，説明変数の選び方を厳密に表現して呼称することとした．

[*9] 目的関数，コスト関数，評価関数，誤差関数，損失関数はほぼ同じ意味で使われているが，本稿では損失関数に正則化を加えた関数を特にコスト関数と呼ぶこととする．

[*10] ハイパーパラメータは損失関数・コスト関数の計算（学習）で求まらないパラメータで，外的に与えるモデルパラメータである．一般に学習で求めるのは損失関数・コスト関数で求まる説明変数の重みベクトル $\boldsymbol{w}$ と b であり，利用するハイパーパラメータの組み合わせ全てを比較することで最適な学習結果を得る．実務的には最適なハイパーパラメータの探索も含めて学習と呼ばれることが多いが，異なる概念である．

化するには，説明変数の重みベクトル $\boldsymbol{w}$ の二乗和が最小となる必要があり，不要な説明変数の重みがゼロに近い値になることが期待される．これは式（1）の説明変数の重みベクトル $\boldsymbol{w}$ が対角化で必要な値になることに相当し，山の尾根 (Ridge) の様にみえるためリッジ回帰と呼ばれる．

$$
\begin{aligned}
J(\boldsymbol{w},b)_{\text{Ridge}} &= J(\boldsymbol{w},b)_{\text{OLS}} + L2 \\
&= \sum_{i=1}^{n}\{y_{(i)} - (\boldsymbol{w}x+b)_{(i)}\}^2 + \lambda\|\boldsymbol{w}\|_2^2 \\
&= \sum_{i=1}^{n}\{y_{(i)} - (\boldsymbol{w}x+b)_{(i)}\}^2 + \lambda\sum_{j=1}^{m} w_j^2
\end{aligned}
\tag{4}
$$

b）$L1$ 正則化：LASSO 回帰

LASSO 回帰は式 (5) で示す $L1$ 正則化項を用いた損失関数 J（$\boldsymbol{w}, b$）$_{\text{LASSO}}$ を用いた線形回帰モデルである．$L1$ 正則化項は説明変数の重みベクトル $\boldsymbol{w}$ の $L1$ ノルム，つまり絶対値の和であるマンハッタン距離を対象とし，式 (5) のコスト関数を最小化することで回帰モデルを推定するのが LASSO 回帰である．$L1$ 正則化項はハイパーパラメータ λ が大きい場合，$\boldsymbol{w}$ が全てゼロになってしまう[*11]．ただし，説明変数の数 M がデータの総数 N より多い場合，説明変数は N までしか選択されず，説明変数間の相関が高い複数の変数がある場合には片方しか採用されないというデメリットがある．

$$
\begin{aligned}
J(\boldsymbol{w},b)_{\text{LASSO}} &= J(\boldsymbol{w},b)_{\text{OLS}} + L1 \\
&= \sum_{i=1}^{n}\{y_{(i)} - (\boldsymbol{w}x+b)_{(i)}\}^2 + \lambda\|\boldsymbol{w}\|_1 \\
&= \sum_{i=1}^{n}\{y_{(i)} - (\boldsymbol{w}x+b)_{(i)}\}^2 + \lambda\sum_{j=1}^{m} |w_j|
\end{aligned}
\tag{5}
$$

c）L2 正則化と L1 正則化の混合：Elastic net

単に不要な変数を削除する変数選択には LASSO 回帰の方が向いているようにみえるが，データの総数 N に選択される説明変数の次元数が限定されることは多数の計測データが手に入り M が大きくなる傾向にある近年の情報計測をベースとしたモデル化に向かない．一方で，リッジ回帰は全ての説明変数の重みベクトル $\boldsymbol{w}$ がゼロになりにくく，説明変数間の相関が高くても片方のみ採用される様な極端な重み付をされることがない．このように，$L2$ 正則化と $L1$ 正則化はノルム計算法が異なることから，リッジ回帰と LASSO 回帰で選択される説明変数は必ずしも一致しない．よって $L2$ 正則化項と $L1$ 正則化項の両方を持つ線形回帰モデルである Elastic net が提案されている[16)]．

Elastic net のコスト関数を式 (6) で示す．変数 α は $L2$ 正則化項と $L1$ 正則化項の比率であり，$\alpha = 0$ のときにリッジ回帰，$\alpha = 1$ の時に LASSO 回帰となる．この α も λ と同様にハイパーパラメータとして最適化することで，リッジ回帰と LASSO 回帰それぞれで学習された説明変数の重みを用いた線形回帰モデルを推定する．

$$
\begin{aligned}
J(\boldsymbol{w},b)_{\text{Elastic net}} &= J(\boldsymbol{w},b)_{\text{OLS}} + (1-\alpha)L2 + \alpha L1 \\
&= \sum_{i=1}^{n}\{y_{(i)} - (\boldsymbol{w}x+b)_{(i)}\}^2 + \lambda(1-\alpha)\|\boldsymbol{w}\|_2^2 + \lambda\alpha\|\boldsymbol{w}\|_1 \\
&= \sum_{i=1}^{n}\{y_{(i)} - (\boldsymbol{w}x+b)_{(i)}\}^2 + \lambda(1-\alpha)\sum_{j=1}^{m} w_j^2 + \lambda\alpha\sum_{j=1}^{m}|w_j|
\end{aligned}
\tag{6}
$$

[*11] 厳密には $L2$ 正則化項でも同じことが起こりうるが，二乗和である $L2$ ノルムの方がゼロになりにくい．詳細は参考文献[11)] などを参照されたい．

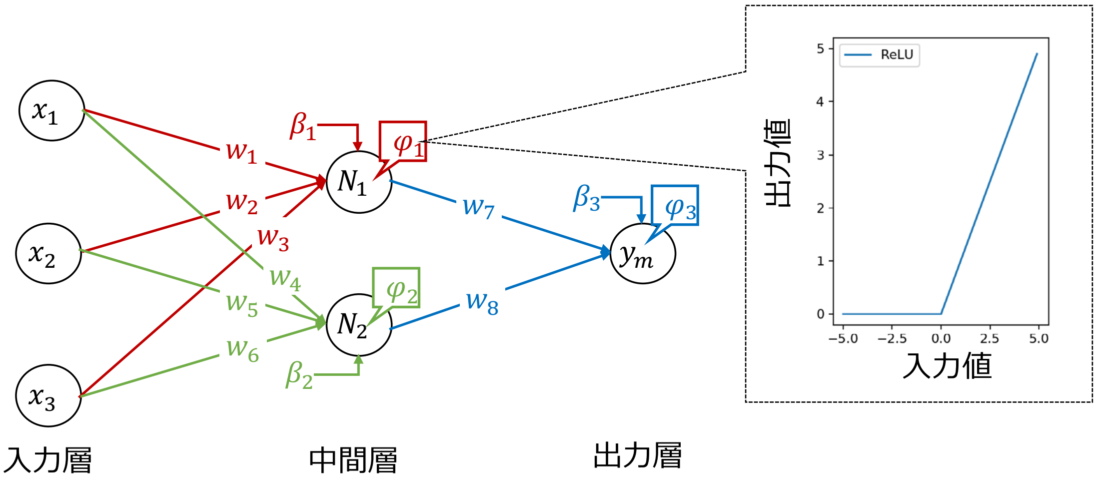


図4　全結合型ニューラルネットの例：ニューラルネットはノード間の重みを全て最適化する．また，各ノードの活性化関数ϕ（ここでは ReLU を例示）を利用することで非線形性を表現する

（4）ニューラルネットによる回帰モデル

ニューラルネットは1節で述べたように近年の人工知能ブームを支えている非線形統計モデルの学習法である．本稿では先行研究[4]でも用いられている全結合型のニューラルネットを利用する[*12]．**図 4** に入力層，中間層，出力層の 3 層の簡易的な全結合型ニューラルネットを示す．

入力層の$\boldsymbol{x}_n$ はこれまでと同様に洪水予測モデルで用いる物理量・計測地点に相当する説明変数であり，この図では3 変数の場合を示す．中間層であるノード N_1 には 3 つの説明変数それぞれの重み $\boldsymbol{w}_1$ から $\boldsymbol{w}_3$ を乗じた線形結合式にバイアス β_1 を加算し，非線形性を表現するための活性化関数 ϕ_1 を乗ずる．活性化関数 ϕ への入力はこれまで説明してきた式（2）の線形回帰モデルと同じ説明変数に重みを乗じた線型結合式である．

活性化関数自体はニューラルネットの語源となった神経細胞（neuron）の動作を模擬した関数であり，従来は単位ステップ関数やシグモイド関数が用いられていたが，最近では**図 4** にも示した ReLU（Rectified Linear Unit）などのランプ関数が用いられる．ランプ関数は，閾値以上のみで出力となる傾斜を持つ関数である．従来型のニューラルネットで用いられていたシグモイド関数は，学習に利用する誤差逆伝播法で必要となる微分係数が 0 に近くなることが多く，多層の学習時に勾配が消失してしまう問題があった．一方で**図 4** に示すランプ関数は入力が 0 以外[*13]で微分係数が 1 か 0 になるので，勾配が消失しにくくなり，深い層の学習が可能になった．

図 4 の二つの中間層は式（7）と（8）となり，出力層の出力である予測値 $\boldsymbol{y}_m$ は式（9）となる．

$$N_1 = \phi_1(w_1x_1 + w_2x_2 + w_3x_3 + \beta_1) \tag{7}$$
$$N_2 = \phi_2(w_4x_1 + w_5x_2 + w_6x_3 + \beta_2) \tag{8}$$
$$y_m = \phi_3(w_7N_1 + w_8N_2 + \beta_3) \tag{9}$$

式（9）は非線形な処理である ϕ_1 と ϕ_2 を含んだ線形結合式[*14]であることから，これまで説明してきた様な統計的回帰モデルと同様の出力$\boldsymbol{y}_m$ を持つ．つまり，ニューラルネットは入力である説明変数と出力

[*12] 著者らは全結合型ニューラルネットよりも時系列情報を適切に表現できるリカレントニューラルネットの河川洪水予測での有効性も確認[6]しており，最近では畳み込み型のニューラルネットも利用例がある．しかし，Elastic net の学習に用いる場合とシーケンスデータの作り方が変わるため，本稿の目指す“同一入力での学習による比較”のため全結合型ニューラルネットを比較対象とした．

[*13] 入力が 0 の時に微分不可だが，実用上は入力 0 のときを 0 か 1 にしている．

[*14] 本稿で扱う様な回帰モデルでは ϕ_3 のみは恒等関数が用いられる．

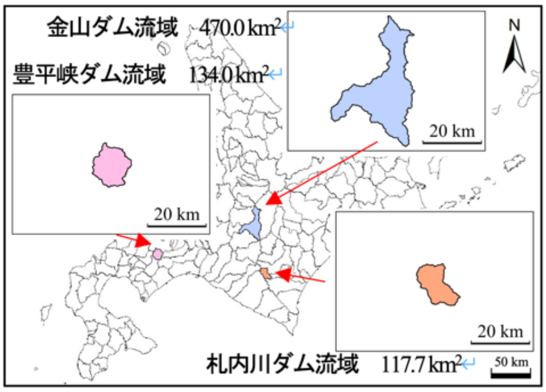
である目的変数の間に非線形な関係があることを前提とした非線形回帰モデルであるといえる．特にランプ関数を活性化関数に用いる場合で考えると，線形結合式とバイアスの値が閾値（**図 4** の ReLU 関数だと 0.0）を超えると恒等関数であることから，複数の線形回帰モデルを活性化関数でオンオフしながら組み合わせた複雑な回帰式であるとみなすことができる．

一般にニューラルネットは学習範囲外の外挿問題にも経験的にうまくいく例が多い．これはニューラルネットは**図 3** で示した決定木の様に値自体を学習するのではなく，線形回帰モデルと同様に関数近似の手法であるため，学習した値に近い外挿的な値に対しては比較的妥当な出力となるためと考えられる．しかし，ニューラルネットは**図 4** の様な簡易なモデルであっても，その重み数は 8 でバイアス数 3 と多く，実用的に使われる複雑なネットワークであれば説明変数間の関係を重みから判断することは難しい．この問題がニューラルネットの予測が評価した範囲では高精度であっても実用化がなかなかされない理由と考えられる．

3. 学習と評価に用いる水文データ

（1）ケーススタディ

本稿では，2 節で述べた，スパースモデリング手法である Elastic net を用いた回帰モデルがダム流入量予測へどの程度利用可能か評価するために，ランダムフォレストによる重要度を用いた線形回帰モデルとニューラルネットによる回帰モデルとの比較で評価する．具体的には**図 5** に示す北海道の金山ダム，札内川ダム，豊平峡ダムの観測値を利用する[*15]．



		金山ダム	札内川ダム	豊平峡ダム
型式		中空重力式ダム	重力式 コンクリートダム	アーチ式 コンクリートダム
水系河川		石狩川水系空知川	十勝川水系札内川	石狩川水系豊平川
堤高	m	57.3	114.0	102.5
堤頂長	m	288.5	300.0	305.0
流域面積	km²	470.0	117.7	134.0
総貯水容量	m³	150,450,000	54,000,000	47,100,000
有効貯水容量	m³	130,420,000	42,000,000	37,100,000
サーチャージ水位	m	345.0	484.0	474.88
洪水貯留準備水位	m	338.5	466.0	458.78
最低水位	m	320.0	447.5	437.68

図5　ケーススタディに用いたダムの位置と諸元

（2）説明変数と目的変数

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/s），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ー雨量を気象予報値として用いた．よって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[9] より収集した．なお後述する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．

[*15] 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．

（3）学習と評価の期間

各ダムの学習評価条件を**表 1** から **3** に示す．表の学習条件はモデル学習に用いる既知条件であり，未学習の期間を目的変数の値が内挿のみの条件と外挿となるピークがある条件とをモデル学習されていない未知の評価条件として設定した．

金山ダムと札内川ダムは 2007 年から 2012 年を学習データとし全データレコードを学習，2013 年から 2015 年を内挿条件の評価期間に，2016 年から 2018 年を外挿条件の評価期間とした．外挿条件の評価期間には，2016 年 8 月の北海道に 4 つの台風が連続して上陸及び接近した平成 28 年 8 月北海道豪雨期間を含む．

豊平峡ダムは 2007 年から 2017 年（2011 年を除く）を学習データとし全データレコードを学習，2018 年を内挿条件の評価期間に，2011 年を外挿条件の評価期間とした．外挿条件の評価期間は平成 23 年 9 月の停滞前線等による大雨の影響が出ている．

また, 各ダムの学習期間の最大流量の 25% を基準として, その値を超えるデータレコードを集計した．その結果，金山ダムは**図 2** にからも明らかな様に，外挿評価期間の 2016 年のみに極端な流入があったたため，学習期間の最大流入量を前提とすると学習に用いる説明変数の次元数を大幅に超えるデータレコードがある．しかし，札内川ダムと豊平峡ダムは数年おきに流入量が多いタイミングがあり，学習期間の最大流入量も相対的に大きく，説明変数の次元数より少ないデータレコード数となった．言い換えれば札内川ダムと豊平峡ダムで大きな流入の発生頻度は学習期間を前提とすれば比較的スパースに発生

表1　学習と評価のデータ収集条件（金山ダム）

条件	学習		内挿評価		外挿評価	
流量の積算時間（時間）	6	12	6	12	6	12
データ収集期間 (年)	2007-2012		2013-2015		2016-2018	
データレコード数 (時間)	52,607		26,279		26,303	
収集説明変数次元数	41	47	41	47	41	47
学習説明変数次元数	861	1,128	861	1,128	861	1,128
期間最大流入量 (m^3/s)	4,065,480	7,943,562	3,316,986	5,980,068	26,264,016	46,217,970
学習最大流入量の 25% 超のデータレコード数 (時間)	3,241	3,416	1,448	1,511	2,179	2,304

表2　学習と評価のデータ収集条件（札内川ダム）

条件	学習		内挿評価		外挿評価	
流量の積算時間（時間）	6	12	6	12	6	12
データ収集期間 (年)	2007-2012		2013-2015		2016-2018	
データレコード数 (時間)	52,607		26,279		26,303	
収集説明変数次元数	41	47	41	47	41	47
学習説明変数次元数	861	1,128	861	1,128	861	1,128
期間最大流入量 (m^3/s)	9,486,486	14,799,690	9,335,286	14,511,942	14,321,124	26,966,160
学習最大流入量の 25% 超のデータレコード数 (時間)	207	350	66	108	113	162

表3　学習と評価のデータ収集条件（豊平峡ダム）

条件	学習		内挿評価		外挿評価	
流量の積算時間（時間）	6	12	6	12	6	12
データ収集期間 (年)	2007-2017 (2011 除く)		2018		2011	
データレコード数 (時間)	87,672		8,759		8,759	
収集説明変数次元数	41	47	41	47	41	47
学習説明変数次元数	861	1,128	861	1,128	861	1,128
期間最大流入量 (m^3/s)	6,515,640	11,780,496	5,678,586	7,699,194	7,853,238	12,566,340
学習最大流入量の 25% 超のデータレコード数 (時間)	220	288	31	44	38	42

するといえる[*16]．金山ダムに関しては，従来の想定を上回る外挿的な流入量を評価するのに適した条件と考えられる．

（4）データの前処理

a）交互作用変数の計算

収集した説明変数は 41 次元または 47 次元だが，これらは単独の計測値である．よって，複数の説明変数が相互に影響した目的変数への効果は表現できず，これらの値のみを学習した回帰モデルの重み w に影響しない．本稿で用いる説明変数で例えると，現時刻の SWI に基づく土壌湿潤状態指標と降雨量は相互に影響しあい目的変数であるダムへの流入量に関与していると想定されるが，SWI や降雨量の単一変数の重みでは単独の変数の重要度しかわからない．

この様な複数の説明変数の交互作用を線形回帰モデルに加えるために，複数の説明変数のアダマール積（要素積）を交互作用に相当する合成変数として線形モデルの説明変数として追加することが行われる．例えば，目的変数を$\boldsymbol{y}_m$，説明変数を $\boldsymbol{x}_n$ とし，重み $\boldsymbol{w}_n$ とバイアス b を用いた説明変数が二つの線形回帰モデルは式 (10) となる．

$$\boldsymbol{y}_m = w_1\boldsymbol{x}_1 + w_2\boldsymbol{x}_2 + b \tag{10}$$

これに説明変数 $\boldsymbol{x}_1$ と $\boldsymbol{x}_2$ の交互作用の合成変数 $\boldsymbol{x}_{12}$ を考慮すると式 (11) となる．

$$\begin{aligned}\boldsymbol{y}_m &= w_1\boldsymbol{x}_1 + w_2\boldsymbol{x}_2 + w_{12}\boldsymbol{x}_{12} + b \\ &= w_1\boldsymbol{x}_1 + w_2\boldsymbol{x}_2 + w_{12}\boldsymbol{x}_1 \circ \boldsymbol{x}_2 + b\end{aligned} \tag{11}$$

ここで，交互作用の合成変数 $\boldsymbol{x}_{12}$ は $\boldsymbol{x}_1$ と $\boldsymbol{x}_2$ の要素積 $\boldsymbol{x}_1 \circ \boldsymbol{x}_2$ という非線形処理だが，式 (11) は線形結合式のままであり，スパースモデリングを含む線形回帰分析で各説明変数の重みwを推定することができる．この計算を最低限の 2 次の交互作用だけとした場合でも説明変数は 1 次の説明変数の数を M_1 とした際に総説明変数 M は以下の式（12）で計算する通り全組み合わせなので膨大な数となるため，本稿では 2 次の交互作用までを利用することとした．

$$M = {}_{M_1}C_2 + M_1 \tag{12}$$

なお，ランダムフォレストなどの決定木を用いるアルゴリズムは**図 3** のノード分岐で交互作用を表現することが可能とされているため，交互作用を一般には考慮しなくて良いとされている．しかし，本稿では決定木学習後に行う線形回帰モデルには交互作用が必要なことからランダムフォレストの学習にも説明変数の交互作用の計算を行った．

b）標準化

収集した水文データは学習期間の目的変数ベクトル$\mathrm{y}_{\mathrm{train}}$ と説明変数ベクトル[*17]$\mathrm{x}_{\mathrm{train}}$ で学習を行う．この際に，$\mathrm{y}_{\mathrm{train}}$ と $\mathrm{x}_{\mathrm{train}}$ をスケーリングするために標準化を行う．標準化は学習に利用するデータベクトルの平均値 μ を 0 に，標準偏差 σ を 1 にする処理である．標準化は最大値と最小値を用いて正規化する場合に比較して異常値（外れ値）に対する情報を持ちながら，ニューラルネットなどで用いる勾配降下法による学習による変数間の尺度差を吸収する作用を持つ．

[*16] 25%という閾値自体は計算上の目安で，洪水に相当する流入量のあった期間の参考として用いた．**図 2** を見ても厳密には各ダムごとに最適な閾値を設定した方が良いと考えられる．しかし，本稿の検討は事例期間のみの学習ではなく全レコード学習であるため，この値自体はデータの分布を示す指標としてしのみ利用する．

[*17] 実際は多次元なので行列となる．

本稿での標準化は式（13）と（14）で目的変数と説明変数に対して行い，学習時に未知となる目的変数と説明変数には式（15）と（16）に示す様にそれぞれの学習条件における平均値 μ と標準偏差 σ を用いた．

$$\mathbf{y}_{\mathrm{train\_norm}} = \frac{\mathbf{y}_{\mathrm{train}} - \mu_{y_{\mathrm{train}}}}{\sigma_{y_{\mathrm{train}}}} \tag{13}$$

$$\mathbf{x}_{\mathrm{train\_norm}} = \frac{\mathbf{x}_{\mathrm{train}} - \mu_{x_{\mathrm{train}}}}{\sigma_{x_{\mathrm{train}}}} \tag{14}$$

$$\mathbf{y}_{\mathrm{test\_norm}} = \frac{\mathbf{y}_{\mathrm{test}} - \mu_{y_{\mathrm{train}}}}{\sigma_{y_{\mathrm{train}}}} \tag{15}$$

$$\mathbf{x}_{\mathrm{test\_norm}} = \frac{\mathbf{x}_{\mathrm{test}} - \mu_{x_{\mathrm{train}}}}{\sigma_{x_{\mathrm{train}}}} \tag{16}$$

また，評価条件の目的変数の値は標準化した値から元の値域に戻すために以下の式（17）の処理を行い，最終的な結果の作図に用いた．

$$\mathbf{y}_{\mathrm{test}} = \sigma_{y_{\mathrm{train}}} \mathbf{y}_{\mathrm{test\_norm}} + \mu_{y_{\mathrm{train}}} \tag{17}$$

なお，交互作用と同様にランダムフォレストなどの決定木を用いるアルゴリズムは特徴量の標準化によるスケーリングは**図 3** のルートやノードにおける分岐条件の絶対値が変わるだけで意味を持たない．しかし，本稿では決定木学習後に行う線形回帰モデルの性能には標準化が影響することから，標準化したのちにランダムフォレストの学習を行った．

4. 評価の設定

（1）ハイパーパラメータの最適化

本稿では，比較に用いるスパースモデリングに基づく Elastic net，決定木を用いたランダムフォレストの重要度に基づいた線形回帰（以下，RF-lr）およびニューラルネットの 3 つの回帰分析法を比較する*18．実装はオープンソースのライブラリを使用した．具体的には Elastic net と RF-lr に用いるランダムフォレストと線形回帰は Pyrhon 用機械学習ライブラリの scikit-learn[18)]，ニューラルネットは tensorflow[17)] の API である Keras[19)] で実装した．

Elastic net の各ハイパーパラメータでの交差検証回数は 10 回とし，ニューラルネットは学習データの 40% をバリデーションデータとした．RF-lr はランダムフォレストのハイパーパラメータを最適化したモデルの重要度から上位 3 変数を選択し，線形回帰モデルを学習した．Elastic net と RF-lr の線形回帰ではすべての重み w が正の値になる様に制約した．

各学習アルゴリズムは，**表 1** から **3** の学習条件でハイパーパラメータを最適化する．ハイパーパラメータの最適化にはハイパーパラメータ自動最適化フレームワークである Optuna[20), 21)] を各学習法で説明する探索範囲で式（18）の平均二乗対数誤差（MSLE : Mean Squared Logarithmic Error）を用いて最適化を設定した．式（18）の n は学習・評価条件のサンプル数であり，$\boldsymbol{y}_{(i)}$ は目的変数の実測値，$\boldsymbol{ym}_{(i)}$ はモデルの予測値である．

$$\mathrm{MSLE} = \frac{1}{n}\sum_{i=1}^{n}\{\log(1+y_{(i)}) - \log(1+y_{m(i)})\}^2 \tag{18}$$

*18 同様の比較は文献[7)]でも行なったが，本項では入力の前処理を共通とし，学習アルゴリズムのみを比較可能となるように評価する．

表4　Elastic net のハイパーパラメータと探索範囲

名称	探索範囲	説明
α	0.0 – 1.0	リッジ回帰と LASSO 回帰の比率
λ	1 – 100	正則化項の重み

表5　RF−lr のハイパーパラメータと探索範囲

名称	探索範囲	説明
決定木数	10 – 100	モデル当たりの最大決定木数
最大深さ	2 – 16	各決定期の最大深さ
最小分割数	2 – 16	ノード当たりの最小分割数
説明変数の最大数	$\sqrt{m}$, $\log_2 m$	決定木あたりの説明変数を最大値 m としてどの様に選ぶか

表6　ニューラルネットのハイパーパラメータと探索範囲

種類	名称	探索範囲	説明
ネットワーク構造	中間層の層数	2 – 30	ネットワークの深さ
	中間層のノード数	10 – 32	中間層のノード数は全ノード共通
	各ノードの活性化関数 ϕ	ReLU, SeLU, eLU	全てランプ関数
学習設定	オプティマイザ	Adam, RMSprop	学習最適化アルゴリズム
	学習時の損失関数	MSE, MAE, logcosh, MSLE	学習時の重みを評価する損失関数（optuna の評価関数とは異なる）
	ミニバッチサイズ	8 – 32	ミニバッチ学習の際のバッチ当たりのサンプル数
	ドロップアウト割合	0.0 – 0.5	学習時にドロップアウト（学習凍結）するノードの割合
	学習エポック数	10 – 32	学習データを何回繰り返し学習させるかの数

MSLE は評価関数としてよく用いられる平均二乗誤差（MSE : Mean Squared Error）に対して，予測値 $\boldsymbol{y}_m$ が大きな値でもその自然対数はあまり大きくならず，誤差の範囲が極端に幅広い場合にも対応できる利点がある．なお，1 を加算するのは，$\boldsymbol{y}$ と $\boldsymbol{y}_m$ が 0 の場合に備えているためだが，本稿で対象とする目的変数のダムへの流入量 $\boldsymbol{y}$ は正の実数であるため，$\boldsymbol{y}_m$ が 0 となる場合を除いて本質的な意味は持たない．

各学習アルゴリズムのハイパーパラメータの設定範囲を表 4 から 6 に示す．探索範囲が整数表記の場合はその範囲の整数型変数，小数表記の場合はその範囲を実数型の変数，固有名詞または数式の場合は表記内容の組み合わせとして Optuna で探索設定した．特にニューラルネットのハイパーパラメータは固有名詞が多いが，全てオープンソースのライブラリである Keras で実装したため，詳細は Keras のオンラインドキュメントを参照されたい[19]．

ハイパーパラメータ探索に用いる Optuna の探索数は予備実験より Elastic net とランダムフォレストは 20 モデルとし，ニューラルネットは同数の 20 モデル（以下，DNN−20）と十分に学習が進んだと考えられる 100 モデル（以下，DNN−100）を設定した．本稿で利用した計算資源の設定を表 7 に示す．ハイパーパラメータ探索について 20 回は研究室レベルのワークステーションで計算したが，ニューラルネットの探索回数 100 回は産総研の AI 橋渡しクラウド（AI Bridging Cloud Infrastructure：ABCI）[22] を用いた．なお，計算に用いる GPU メモリとメインメモリが共に ABCI は誤り訂正のある ECC メモリだが研究室レベルのワークステーションは ECC メモリではなく誤り訂正が無いが，本稿ではこの影響はないものとして比較した．

（2）評価指標

アルゴリズムの比較は学習時間と予測性能の二つの観点で行った．

各学習法による評価結果の評価関数には，表 1 から 3 に示した各評価条件の実測最大流入量を中心に前後 80 時間分の予測結果を切り出して評価した．式 (19) のピーク積算流入量時刻差 ΔT_p と式 (20)

表7　利用した計算資源

探索条件	CPU	GPU	メインメモリ
探索数 20 (ワークステーション)	Intel Core i9-7920 (12 コア 24 スレッド)	NVIDIA Quadro RTX 6000 *2 枚 (GPU メモリ GDDR6 24 GiB * 2)	128 GiB (non ECC)
探索数 100 (ABCI)	Intel Xeon Gold 6148 *2 (20 コア 40 スレッド*2)	NVIDIA Tesla V100 * 4枚 (GPU メモリ HBM2 16 GiB *4)	384 GiB (ECC)

表8　学習アルゴリズム別の学習時間

	Elastic net	RF-lr	DNN-20
学習時間 (秒)	301	**52**	13,846

のピーク時積算流入量相対誤差 J_{pe} で評価する．なお，T_{obs} は実測ピーク時刻，T_{pred} は予測ピーク時刻で V_{Qobs} は期間中のピーク積算流入量の実測値，V_{Qpred} は同予測値である．

$$\Delta T_p = T_{pred} - T_{obs} \tag{19}$$

$$J_{pe} = \frac{V_{Qobs} - V_{Qpred}}{V_{Qobs}} \tag{20}$$

本稿では，ピーク積算流入量時刻が早めに評価される $\Delta T_p \leq 0$，ピーク積算流入量が過大に評価される $J_{pe} \leq 0$ の場合を安全側でより望ましいと定義し，どちらも絶対値がゼロに近いほどより良いと判断する．

5. 評価結果

（1）計算時間の比較

金山ダムの 6 時間流入量の学習時間を 3 つのアルゴリズムで比較した．**表 8** に学習時間を示す．比較に用いた計算機は**表 7** の Optuna の探索モデル数 20 条件だが，DNN−100 は常に DNN−20 より学習時間がかかるのでこの結果は最低限必要なの DNN 最適化時間とみなして良い．結果より RF−lr は学習時間が 1 分未満の 52 秒であり最も高速だった．

ニューラルネットの学習と最適なハイパーパラメータの探索に時間がかかるのは本稿で検討した**表 6** の探索パラメータ数も多く，その探索範囲が比較的広いことも影響している．本稿では，ニューラルネットのハイパーパラメータ探索範囲が一般化できないこともあり，今回の探索範囲は著者らが様々な分野の応用プロジェクトで利用している探索範囲をそのまま利用したことも影響している．一方で，規模の大きいニューラルネットの方が表現力が高く複雑な非線形回帰も可能となるため，計算コストはかかるものの，一般に予測性能は高いものとなる．

（2）線形回帰モデルで学習された重み

図 6 から **11** に Elastic net 学習による重みの絶対値 $|\boldsymbol{w}|$ の大きさによる重要変数および RF−lr に用いたランダムフォレストモデルの変数重要度の上位 20 変数を学習条件別に示す．入力する説明変数を標準化したモデル式の重みの絶対値のため，値の大小は変数の重要さの相対値となる．

図の変数名のうち流量（m^3/s）は Q，流域雨量は Basin，ダムのある地点雨量は Point，SWI，レーダ解析による予測雨量は Pre_Rain とし，現時刻と異なる変数には時間差を _nh として示した．また，交互作用は二つの変数名を表記している．

次節の予測では，Elastic net では重みのある全変数を，RF−lr は各図の上位 3 変数のみを用いた．結果より，二つの手法の順位が異なるものの，概ね Q および Q を含む交互作用が上位にきている．特に現

表9　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10　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11　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

時刻の 1 時間流入量である Q はほぼすべての条件で最重要となっており，これは現在の流入量が予測対象となる将来の流入量の変化の概形となることを示唆している．

（3）予測結果の比較

a）全体の傾向

3 つの学習手法のダム別の結果を表 9 から 11 に示す．表の同一行での ΔT_p と J_{pe} の最良値をボールドで示す．J_{pe} の最良値は，まず安全側の $J_{pe} \leq 0$ を満たした条件があるか判定し，満たしている条件で絶対値が最も小さい結果を最良とし，$J_{pe} \leq 0$ を満たした条件がない場合は単純に絶対値の小ささで選択した．このため，J_{pe} の絶対値は小さくても危険側のため最良としていない条件がある．

また，各条件のハイドログラフを図 12 から 23 に示す．作図の際に，予測値が学習法によって大きく変動することから，同一条件でも縦軸のスケールが異なることに注意されたい．RF−lr 条件のハイドログラフにはランダムフォレストが学習した決定木の結果と重要度から計算した線形回帰モデルの予測値を合わせて示す．なお，ランダムフォレストの予測結果は外挿範囲の出力はないため，内挿条件では RF−lr の線形モデルより良い場合もあるものの，すべての外挿条件で学習最大値で頭打ちする予測モデルとなっている．

結果より，ピーク積算流入量時刻差 ΔT_p はほぼ全てが $\Delta T_p > 0$ と安全側ではないが，金山ダムの 12 時間流入量の内挿条件を除いて Elastic net は最良であり，外挿の 12 時間流入量という最も厳しい条件でも札内川ダムと豊平峡ダムで 1, 3 時間と実用的な性能を示している．唯一最良ではない金山ダムの外挿の 12 時間流入量も最良の RF−lr と比べて 1 時間差であり，全体的にピーク時間の予測性能が良いと言える．加えて，$\Delta T_p \leq 0$ と実測よりも安全側になった札内川ダムの内挿条件の 12 時間流入量の Elastic net による予測結果は，図 17（a）より，ピークが二つ発生予測されており，本来のピーク時刻で予測値が減少している．一方でニューラルネットは最適化回数にかかわらず，外挿条件での 12 時間流入量の ΔT_p が悪く，外挿の未学習条件に弱いこと示唆している．6 時間流入量であれば外挿条件でもそれほど

悪くないことから，図 1 に示した，内挿条件に近い外挿条件であれば比較的妥当な結果を示すというこれまでの経験則と同じ傾向を示唆している．

ピーク時積算流入量相対誤差 J_{pe} については，絶対値としては DNN−20 または DNN−100 が良い傾向にあるものの，安全側の $J_{pe} \leq 0$ となるのは Elastic−net か RF−lr の外挿条件のみである．また，DNN−20 と DNN−100 ではDNN−100 の方が J_{pe} の絶対値が小さくなっていることが多い．

b）ダムごとの結果：金山ダム

金山ダムについては，表 1 に示した様に外挿期間の流入量が学習・内挿期間よりも極端に大きい．しかし，どの学習法でも比較的良好な J_{pe} である．図 6 と 7 より，Elastic net および RF−lr の重み上位の説明変数を確認するとどちらも流入量に関わる説明変数が上位だが，RF−lr は雨量が入らないのに対し，Elastic net は交互作用として他の説明変数にも重み付けされることがわかる．これは金山ダムの流入量は直近の流入の影響が大きいという意味となるが，表 1 に示した様に学習期間におけるピーク流入量に近い値を多数学習していることから，同じ次元の物理量である流入量の変化だけでもある程度予測可能となったと考えられる．

c）ダムごとの結果：札内川ダム

札内川ダムについては，外挿条件において Elastic net の J_{pe} は安全側だが，絶対値としては DNN−20, DNN−100 より大きい．ただし，図 18 と 19 より，ΔT_p の評価値が良かった様に，Elastic net の予測値は低水位からピークへ向けての立ち上がりに関しては実測との整合が良い様に見える．これに関しては立ち上がりの一致度を評価する指標を検討し，より詳細な分析を今後行いたい．また，図 8 と 9 より，Elastic net の重み上位の説明変数を確認すると，金山ダムと同様に現時刻流入量が上位に来るが，予測雨量及びその交互作用, SWI と予測雨量の交互作用などが上位となった．ランダムフォレスト学習では，流入量 Q と予測雨量の相互作用が上位となっている．つまり，札内川ダムの流入量は雨の影響を金山ダムよりも強く受けることを示唆している．

d）ダムごとの結果：豊平峡ダム

豊平峡ダムは図 21 と 23 より DNN−20, DNN−100 を用いた場合の特に 12 時間後流入量予測値の概形が悪く，ΔT_p からわかる様に特に立ち上がりが悪い．一方で外挿条件に限れば，Elastic net が最も概形が近いが，RF−lr は同じ線形モデルであるにもかかわらず，6 時間流入量と 12 時間流入量の概形が異なる．これは図 10 と 11 より，選択された説明変数が，RF−lr では 6 時間では過去の流入量または流入量同士の交互作用，12 時間では流入量と予測雨量の交互作用と異なるのに対し，Elastic net は 6,12 時間のどちらも現時刻および 6 時間前流入量と予測雨量同士の交互作用であり同じ様な説明変数が上位になったことが関係している．つまり，豊平峡ダムの流入量は，予測する 6, 12 時間先共に予測雨量の影響を受けやすいということがいえる．

e）ニューラルネットの結果について

DNN のハイパーパラメータ最適化の回数だが，20 回より 100 回の方が明らかに改善したのは豊平峡ダムのみである．この時の最適化回数を計算ログより確認すると，6 時間で 83 回目，12 時間で 80 回目が最適値となっていた．つまり，DNN−100 は今回探索した中で最良のモデルである．一方で最良とはいえ未学習値に対してピークが内挿と外挿の両条件ともに改善しないこともある．これは学習時の損失関数やハイパーパラメータ探索の評価関数が ΔT_pやJ_{pe} と異なることが理由の一つに考えられる．他分野では損失関数を解くべきタスクに合わせて選択することで，目的に対して最適な学習が行われるといっ

た応用も多い．今後は洪水予測に利用するニューラルネットも防災工学の視点から妥当な損失関数を設計していく必要があると考えられる．

6. まとめと今後の課題

本稿では，河川における洪水を予測する水位直接評価モデルが統計的回帰モデルであることに着目し，外挿条件を考慮することが可能な線形モデルとしてスパースモデリングによる予測について提案した．特に，上流下流が明確な河川と異なり，上流にあるダムは予測モデルの入力となる説明変数に何が有効か議論しにくいことから，スパースモデリングが持つ圧縮センシングと呼ばれる変数選択が有効であると考えた．評価には，入力となる説明変数について相互作用も含めた予測モデルとし，学習前の標準化など，線形回帰モデルの性能をきちんと担保するための処理を丁寧に行った．その結果，近年盛んに研究されている深層学習に基づくニューラルネットと同等以上の予測結果となり，先に述べた圧縮センシングによって選択された説明変数で結果が議論できることを示した．

一方で，ニューラルネットは画像分類等では重み解析の表示など解析的なことが行われているものの，本稿で扱う様な回帰問題においては有効な重み解析・変数重要度を議論する方法でデファクトスタンダードとなる手法はない．将来の予防的な対策の実施には説明変数の重要度は大いに参考となる資料であることも考えると，内挿条件を含む高度な予測と，外挿条件の予測や解析的な予測にはElastic netのような線形モデルも重要である．

また，学習に要する時間もニューラルネットと比べれば大幅に速いこともあり，まだ評価が済んでいないダムや河川への適応もデータの収集次第可能であることから，今後の洪水予測問題での発展が期待される．著者らも2020年の令和２年７月豪雨で被害のあった熊本県球磨川，山形県最上川などでの適応を目指しデータ収集を行っている．

謝辞

本研究の実施に関して，室蘭工業大学の中津川誠教授には有益な情報提供等で貢献いただいた．また，データ収集と一部の作図には室蘭工業大学の大学院生である山洞智弘氏に協力いただいた．ここに記して謝意を表す．

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）Hinton, G. E. and Salakhutdinov, R. R., “Reducing the dimensionality of data with neural networks,” Science 313（5786）, pp. 504-507, 2006.

3）谷岡康, 福岡捷二, 岩永勉, 北川明, “都市域中小河川における洪水位と雨量の直接的関係を用いた洪水解析,” 水工学論文集, 38 巻, pp. 69-74, 1994.

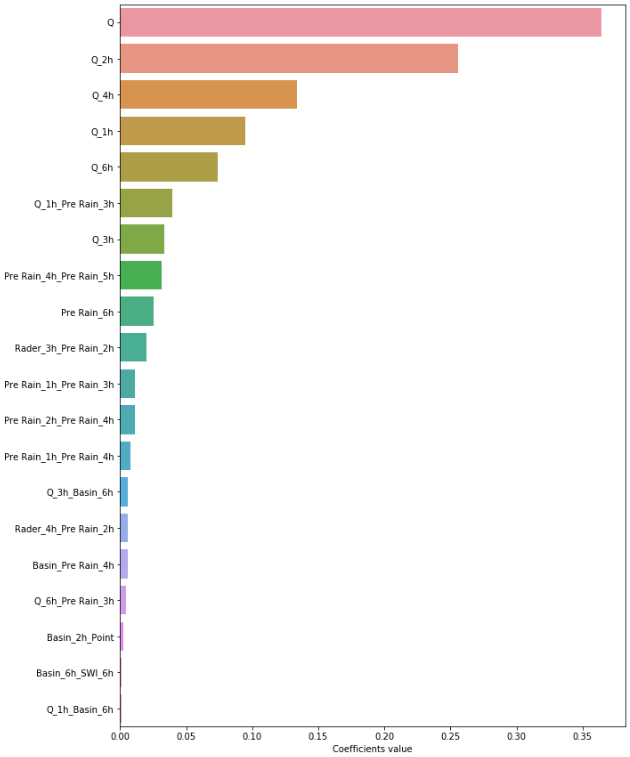
4）一言正之, 櫻庭雅明, 清雄一, “深層学習を用いた河川水位予測手法の開発,” 土木学会論文集B1（水工学）, 72 巻, 4 号, pp. I_187-I_192, 2016.

5）岡崎亮太, 中津川誠, 小林洋介, “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,” 土木学会論文集B1（水工学）, 74 巻, 4 号, pp. I_1459-I_1464, Feb. 2018.

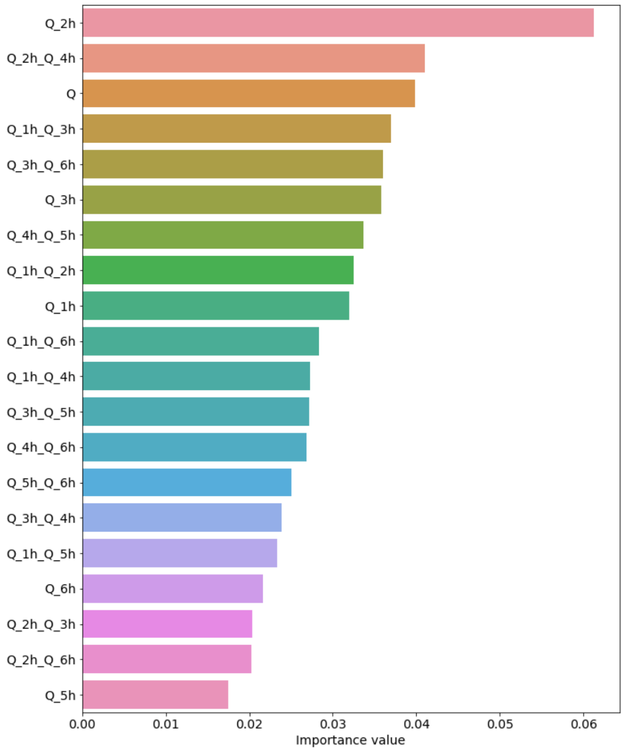
6）山田恒輝, 小林洋介, 中津川誠, 岸上順一, “リカレントニューラルネットワークを用いた2016 年の常呂川洪水事例の水位予測,” 土木学会論文集B1（水工学）, 74 巻, 5 号, pp. I_1369-I_1374, Nov. 2018.

7）坂本莉子, 小林洋介, 中津川誠, “異常洪水時のダム貯水位予測に用いる機械学習手法の比較,” 土木学会論文集B1（水工学）, 75 巻, 2 号, pp.I_85-I_90, Nov. 2019.

8）山洞智弘, 中津川誠, 小林洋介, 坂本莉子, “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, Nov. 2020.
9）国土交通省, “水文水質データベース,” http://www1.river.go.jp/
10）Sebastian Raschka and Vahid Mirjalili, 福島真太朗監訳, “Python 機械学習プログラミング：達人データサイエンティストによる理論と実践第3 版,” インプレス, 2020.
11）鈴木讓, “統計的機械学習の数理100 問with Python,” 共立出版, 2020.
12）Leo Breiman, “Random Forests,” Machine Learning, 45, 1, pp. 5‑32, 2001.
13）Tianqi Chen and Carlos Guestrin. “GBoost：A Scalable Tree Boosting System,” In Proc. KDD 2016, ACM, pp. 785‑794, 2016.
14）Arthur E. Hoerl and Robert W. Kennard, “Ridge Regression：Biased Estimation for Nonorthogonal Problems,” Technometrics, Vol. 12, No. 1, pp. 55−67, 1970.
15）Robert Tibshirani, “Regression Shrinkage and Selection via the Lasso,” Journal of the Royal Statistical Society. Series B（Methodological）, Vol. 58, No. 1（1996）, pp. 267‑288, 1996.
16）Zou, H. and T. Hastie “Regularization and variable selection via the elastic net,” Journal of the Royal Statistical Society, Series B, Vol. 67, No. 2, pp. 301‑320, 2005.
17）tensorflow, https://www.tensorflow.org/
18）scikit−learn Machine Learning in Python, https://scikit−learn.org/
19）Keras Simple. Flexible. Powerful., https://keras.io/
20）Takuya Akiba, Shotaro Sano, Toshihiko Yanase, Takeru Ohta, and Masanori Koyama. “Optuna：A Nextgeneration Hyperparameter Optimization Framework, ” Proc. 25th ACM KDD, 2019.
21）Optuna ハイパーパラメータ自動最適化フレームワーク, https://www.preferred.jp/ja/projects/optuna/
22）産業技術総合研究所, AI 橋渡しクラウドABCI, https://abci.ai/ja/

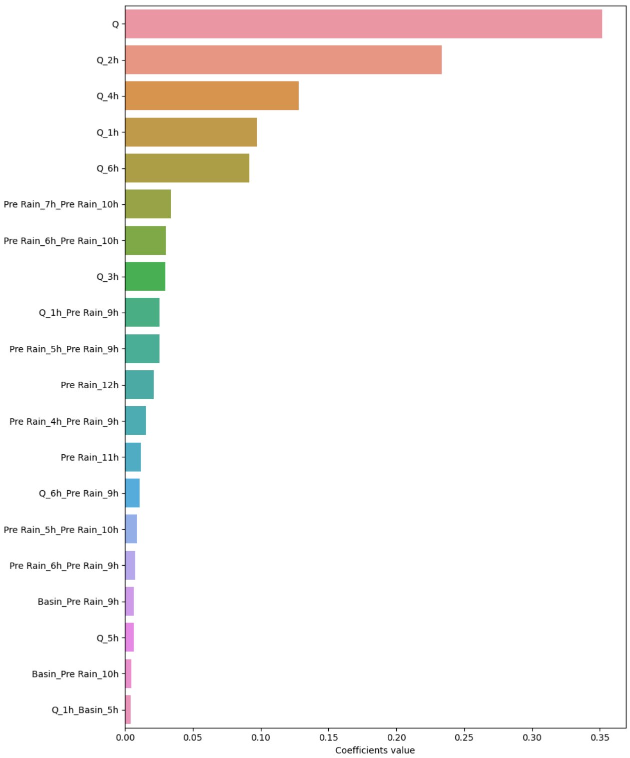


(a) Elastic net で学習された変数重み

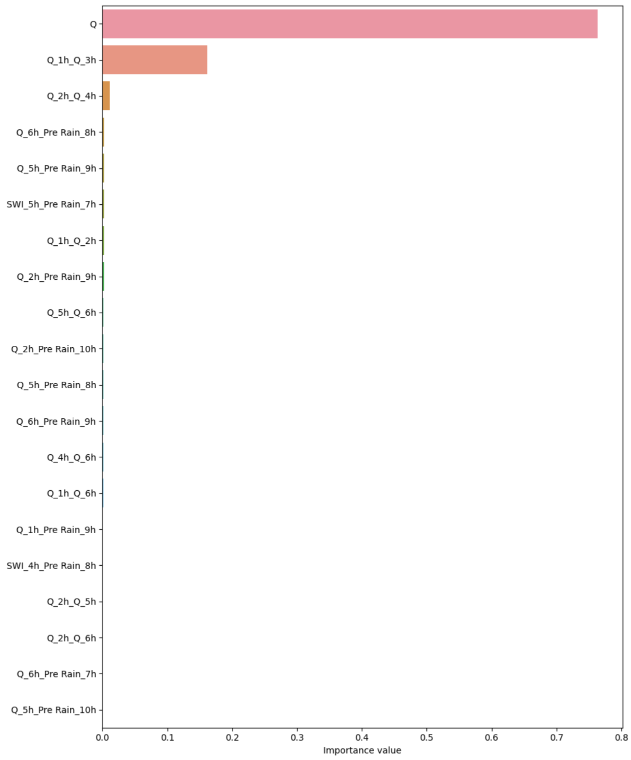


(b) RF 学習による説明変数の重要度

図6　金山ダムの 6 時間流入量における学習条件の重み上位の説明変数

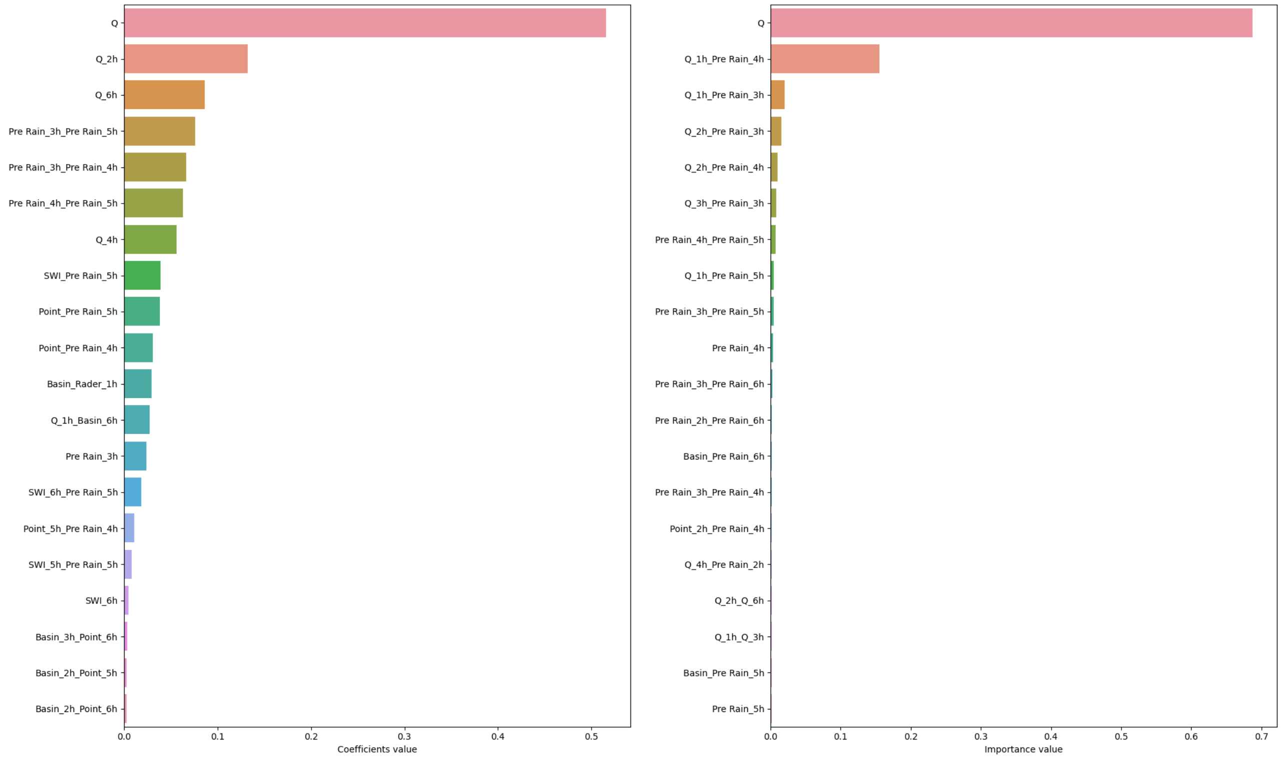


(a) Elastic net で学習された変数重み



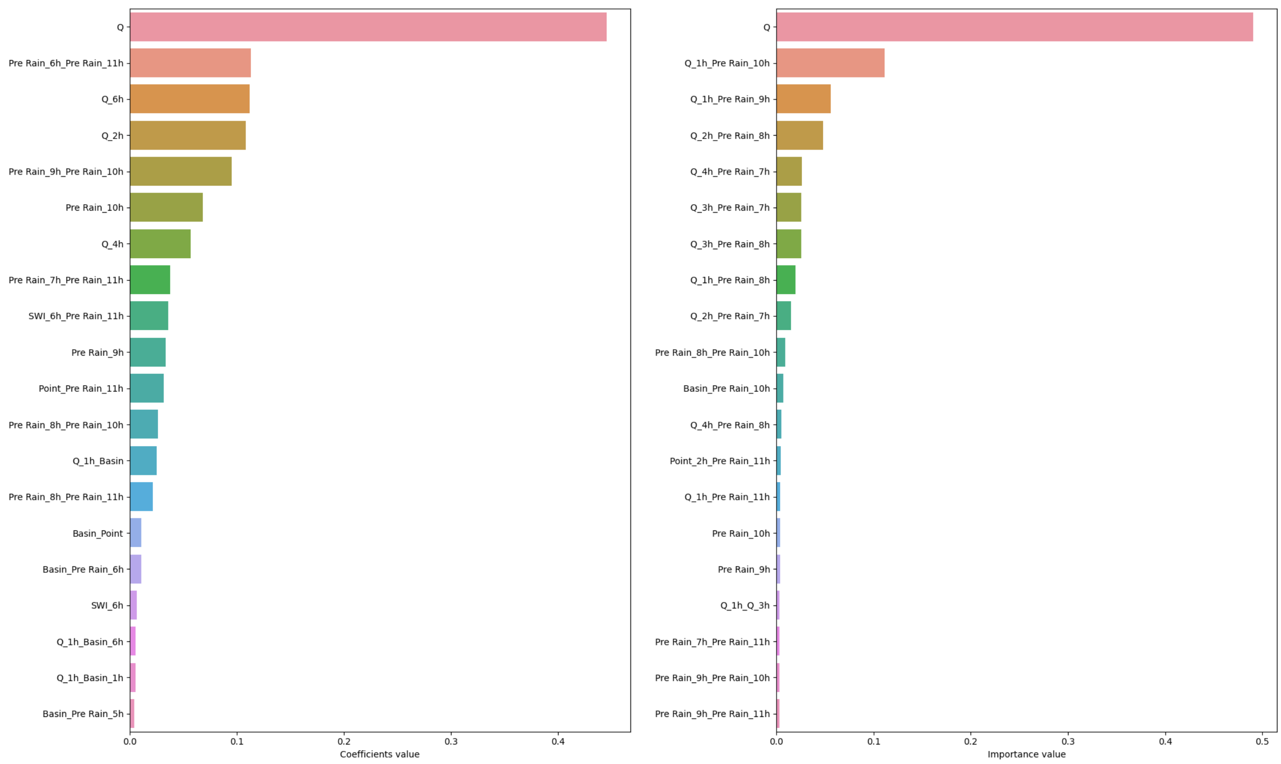
(b) RF 学習による説明変数の重要度

図7　金山ダムの 12 時間流入量における学習条件の重み上位の説明変数



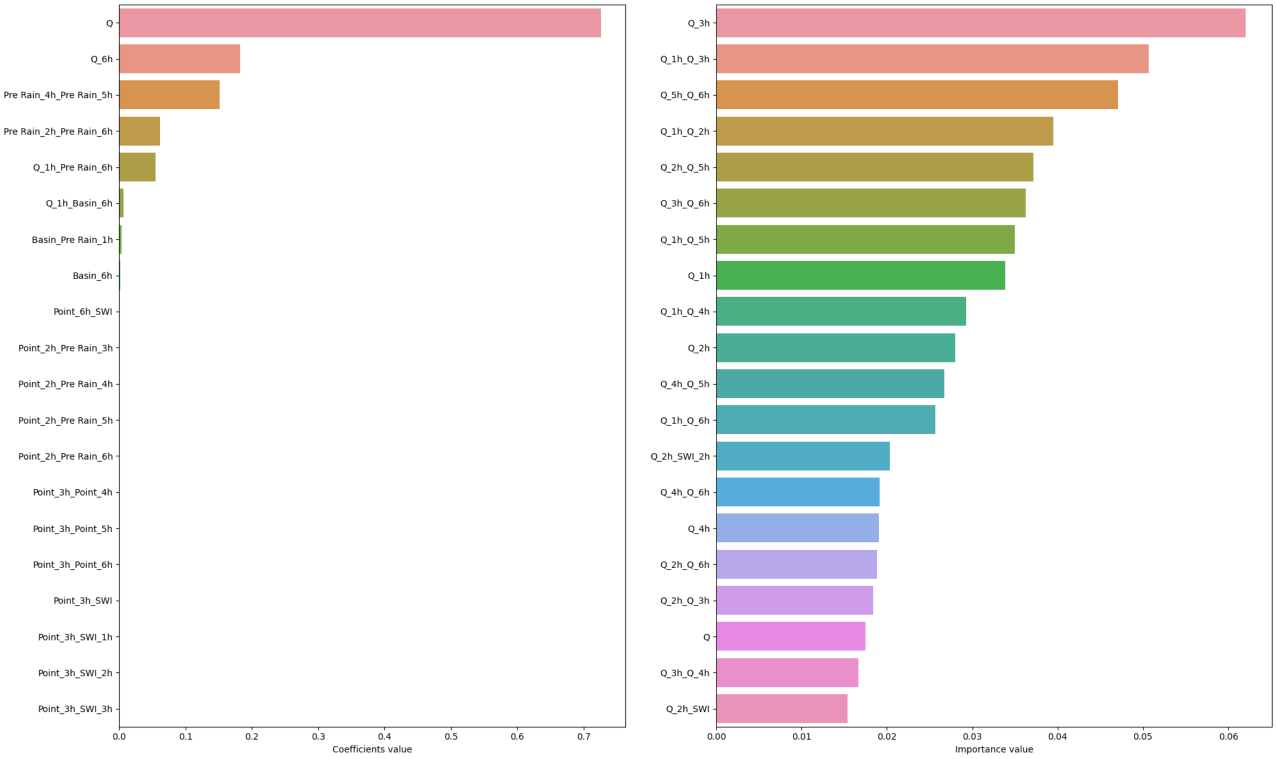
(a) Elastic net で学習された変数重み　　(b) RF 学習による説明変数の重要度

図8　札内川ダムの 6 時間流入量における学習条件の重み上位の説明変数



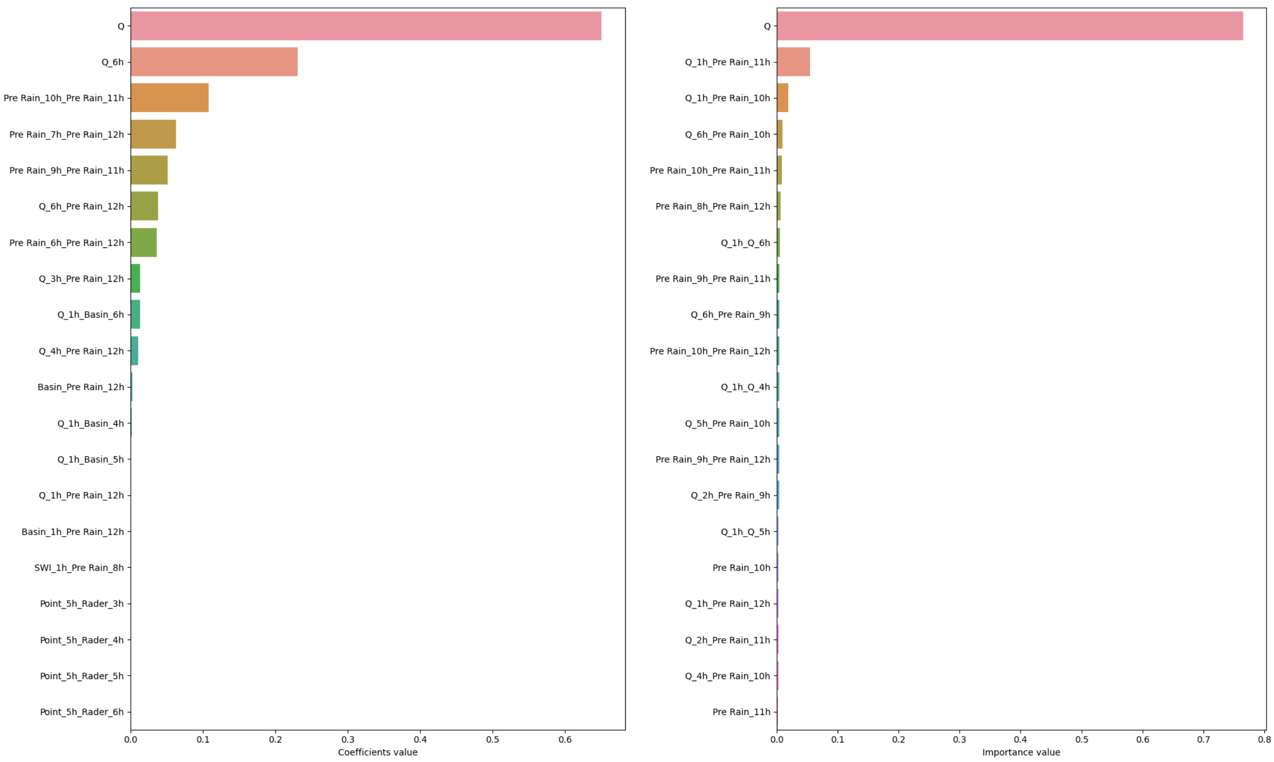
(a) Elastic net で学習された変数重み　　(b) RF 学習による説明変数の重要度

図9　札内川ダムの 12 時間流入量における学習条件の重み上位の説明変数



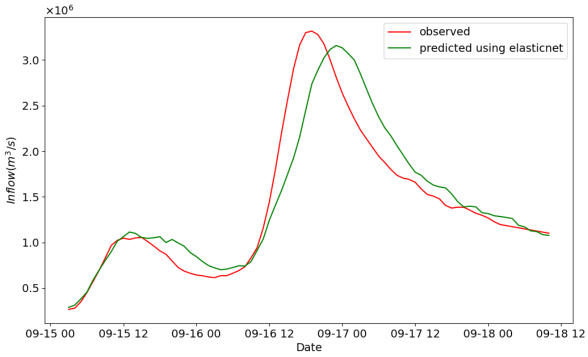
(a) Elastic net で学習された変数重み　(b) RF 学習による説明変数の重要度

図10　豊平峡ダムの 6 時間流入量における学習条件の重み上位の説明変数

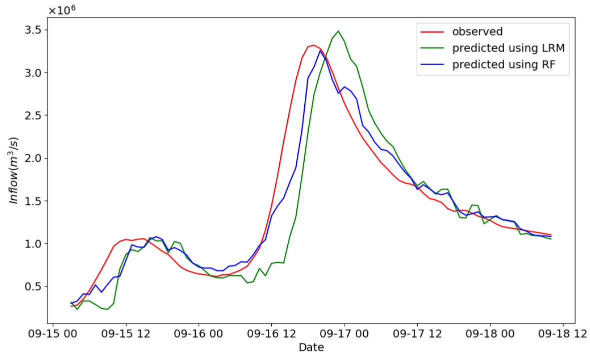


(a) Elastic net で学習された変数重み　(b) RF 学習による説明変数の重要度

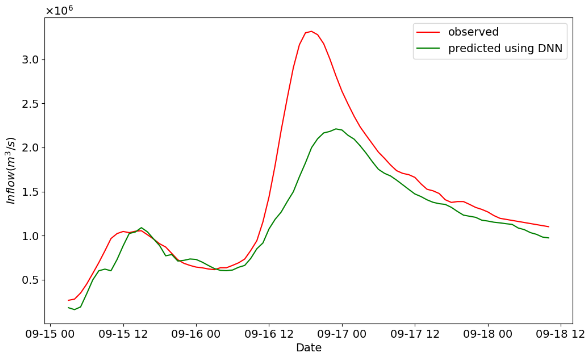
図11　豊平峡ダムの 12 時間流入量における学習条件の重み上位の説明変数



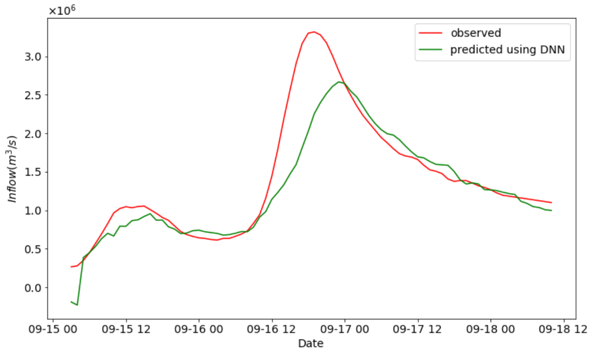
(a) Elastic net



(b) RF-lr

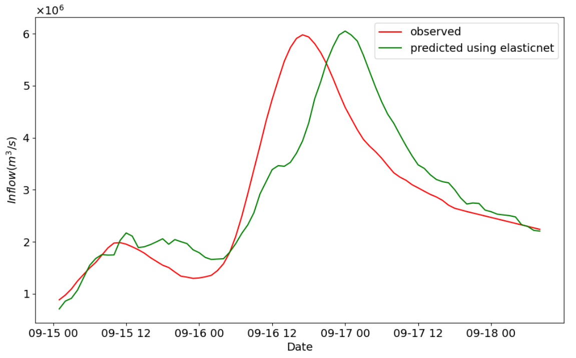


(c) DNN-20

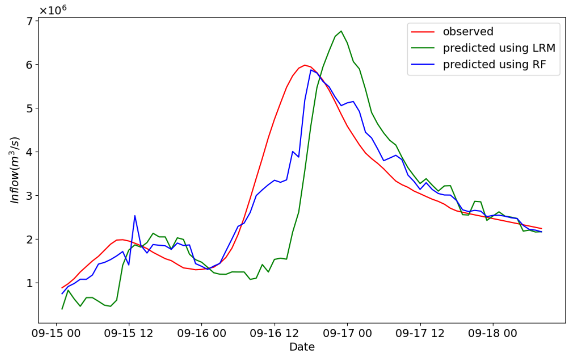


(d) DNN-100

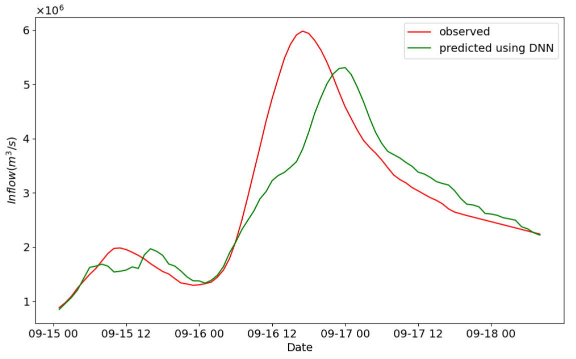
図12　金山ダムの 6 時間流入量における内挿条件の予測結果ハイドログラフ



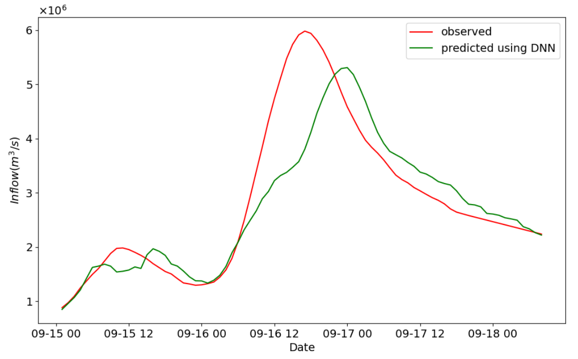
(a) Elastic net



(b) RF-lr

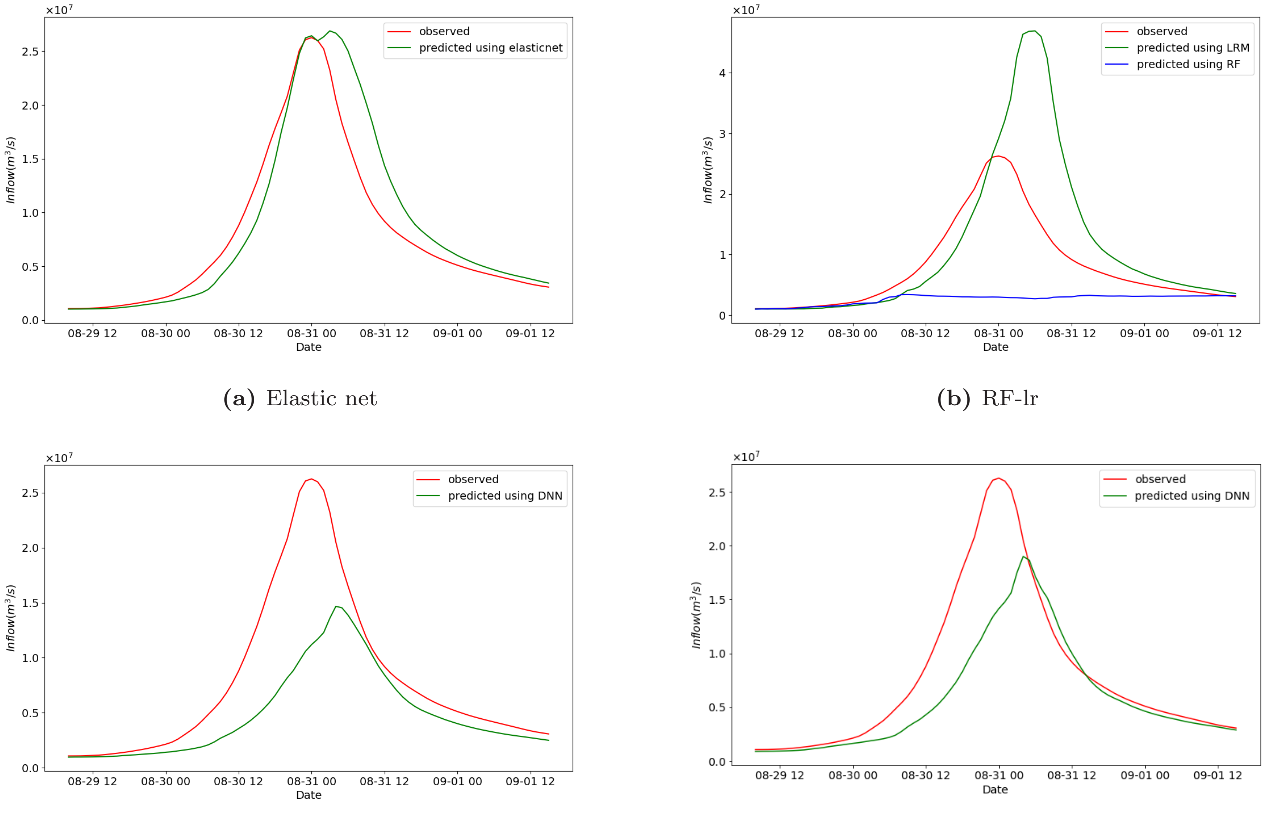


(c) DNN-20



(d) DNN-100

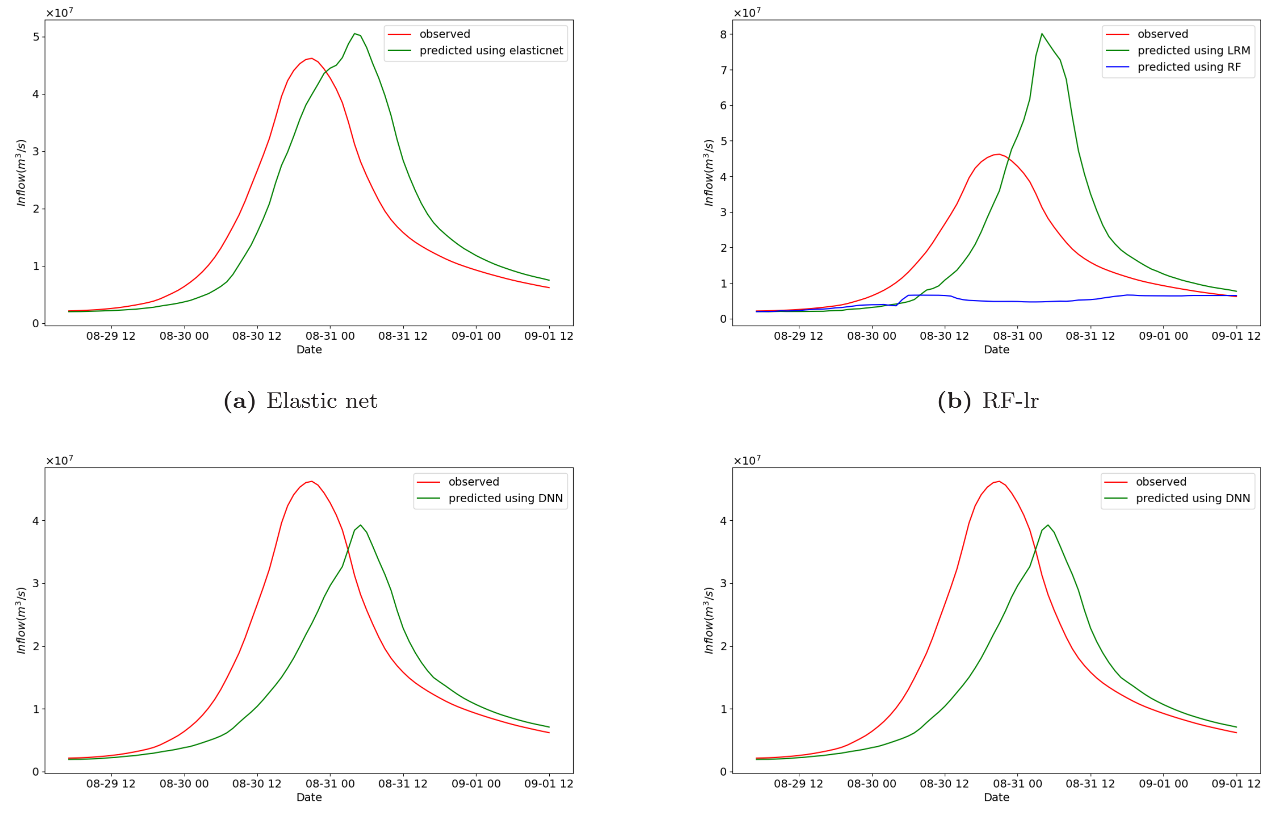
図13　金山ダムの 12 時間流入量における内挿条件の予測結果ハイドログラフ



(a) Elastic net　(b) RF-lr

(c) DNN-20　(d) DNN-100

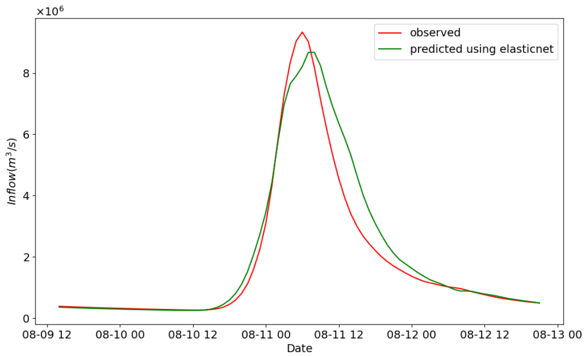
図14　金山ダムの 6 時間流入量における外挿条件の予測結果ハイドログラフ



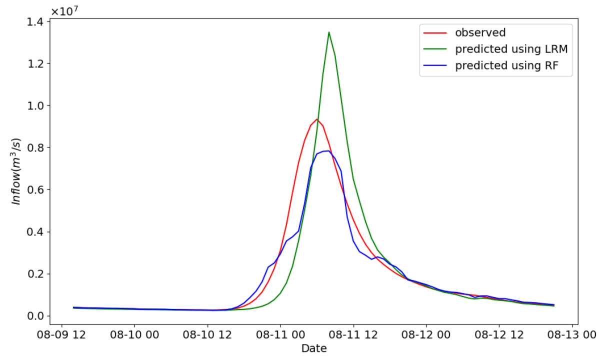
(a) Elastic net　(b) RF-lr

(c) DNN-20　(d) DNN-100

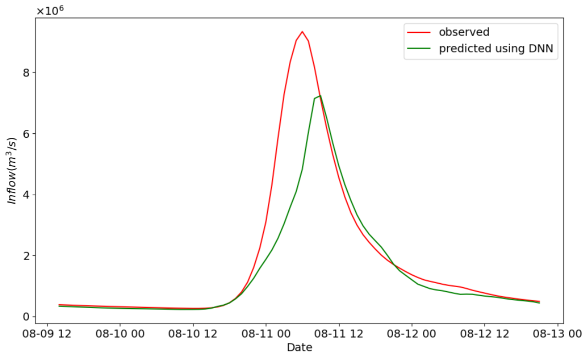
図15　金山ダムの 12 時間流入量における外挿条件の予測結果ハイドログラフ



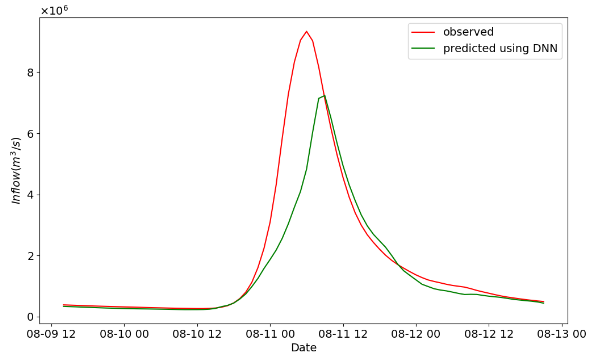
(a) Elastic net



(b) RF-lr

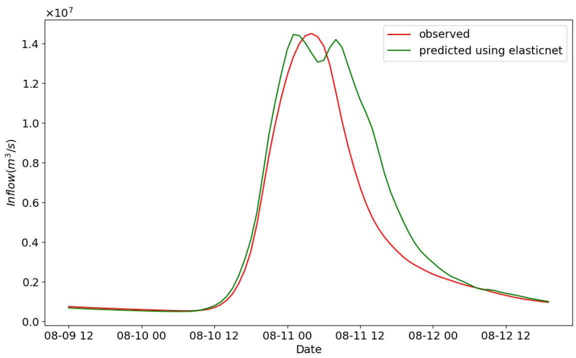


(c) DNN-20

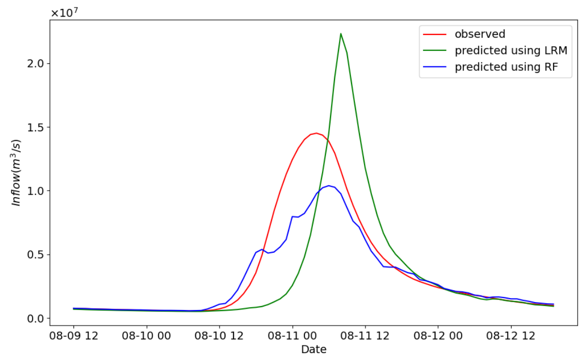


(d) DNN-100

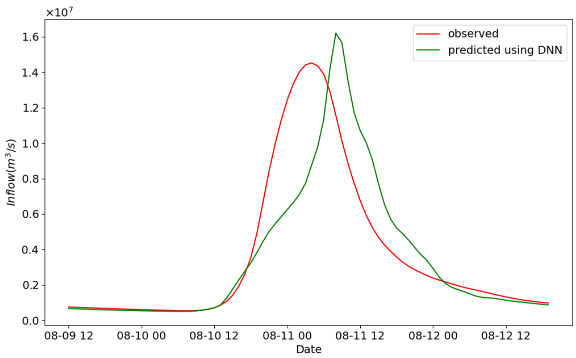
図16　札内川ダムの 6 時間流入量における内挿条件の予測結果ハイドログラフ



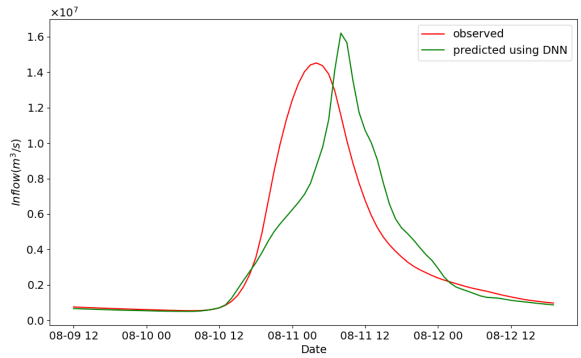
(a) Elastic net



(b) RF-lr

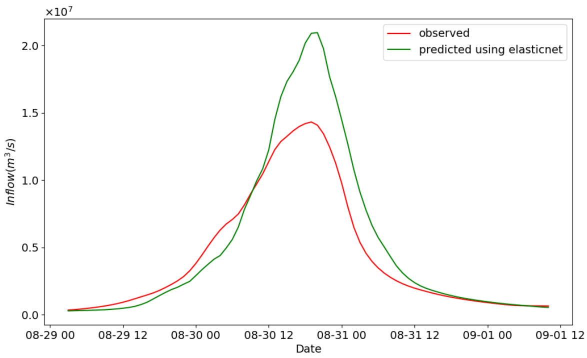


(c) DNN-20

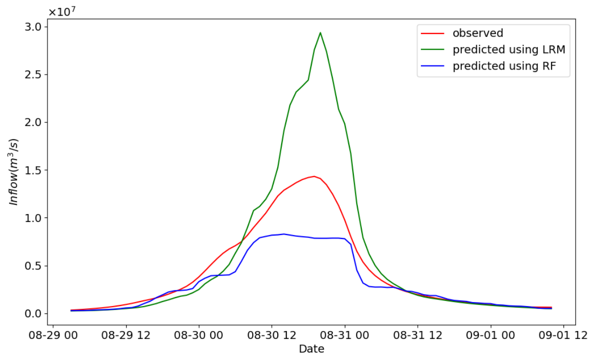


(d) DNN-100

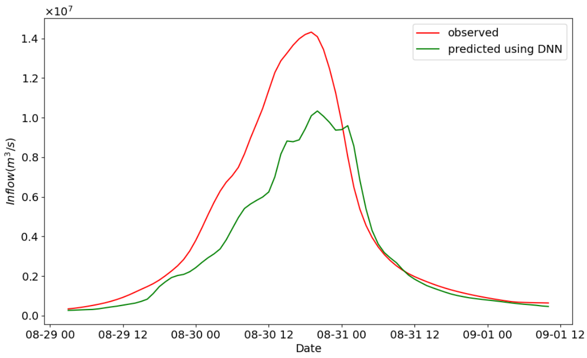
図17　札内川ダムの 12 時間流入量における内挿条件の予測結果ハイドログラフ



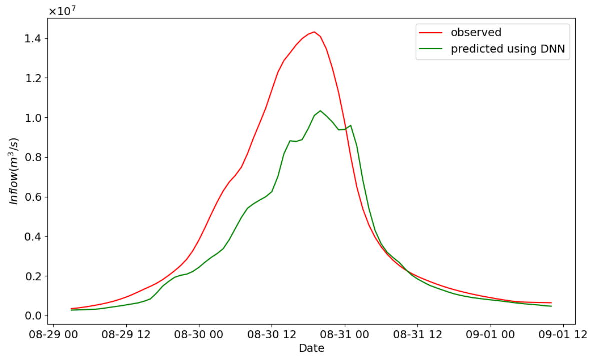
(a) Elastic net



(b) RF-lr

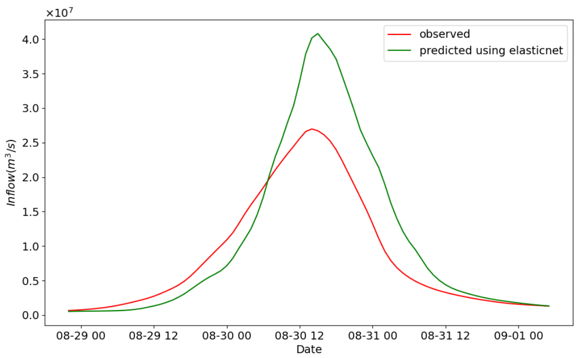


(c) DNN-20

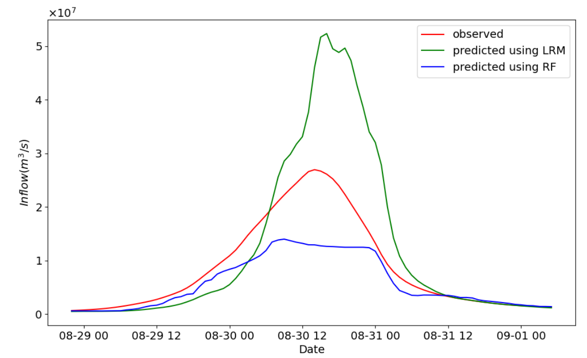


(d) DNN-100

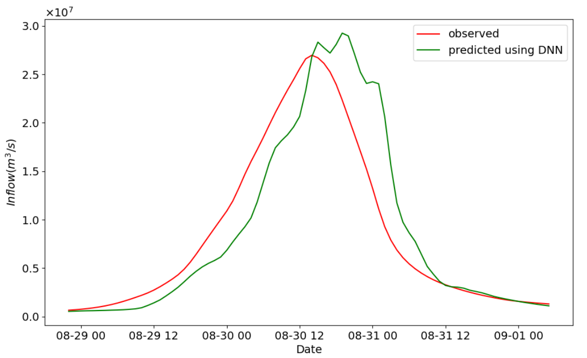
図18　札内川ダムの 6 時間流入量における外挿条件の予測結果ハイドログラフ



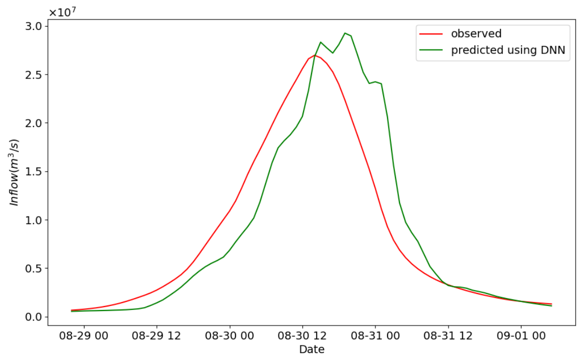
(a) Elastic net



(b) RF-lr

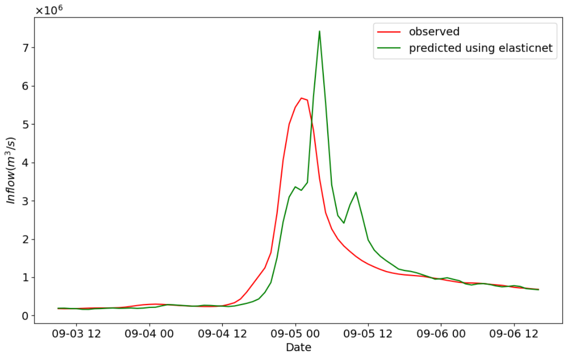


(c) DNN-20

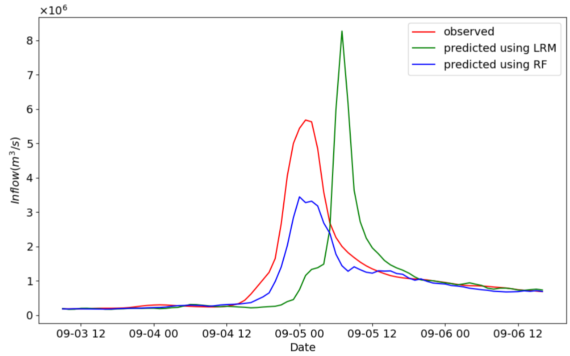


(d) DNN-100

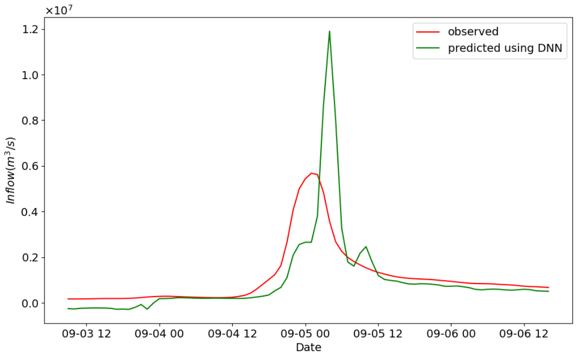
図19　札内川ダムの 12 時間流入量における外挿条件の予測結果ハイドログラフ



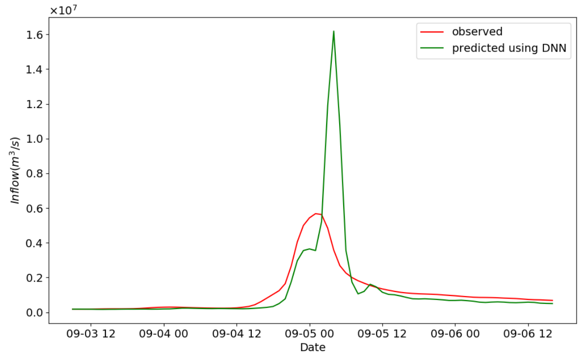
(a) Elastic net



(b) RF-lr

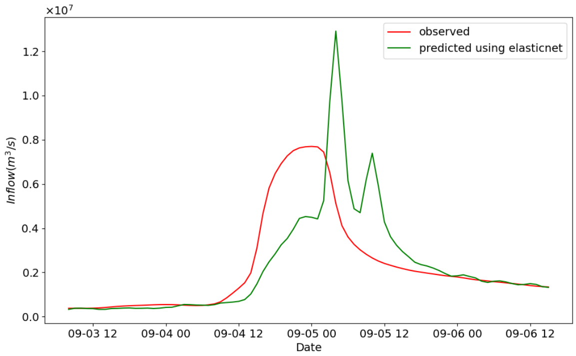


(c) DNN-20

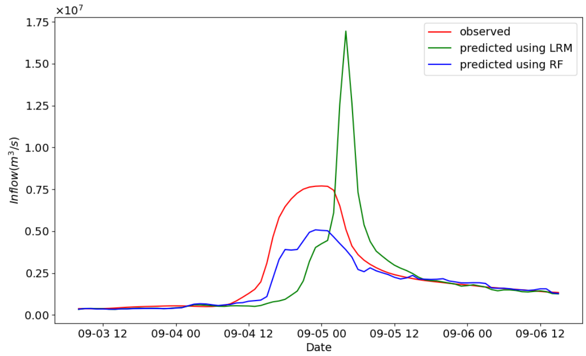


(d) DNN-100

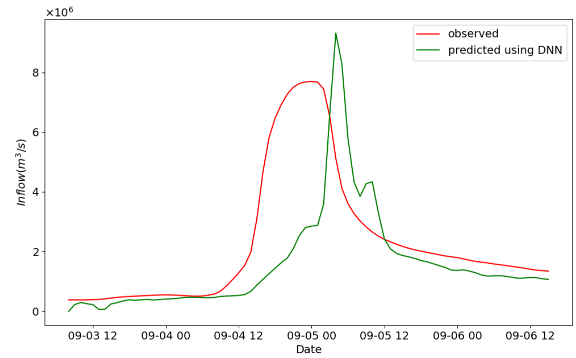
図20　豊平峡ダムの 6 時間流入量における内挿条件の予測結果ハイドログラフ



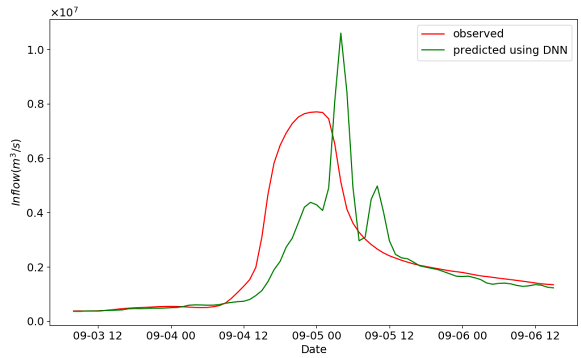
(a) Elastic net



(b) RF-lr

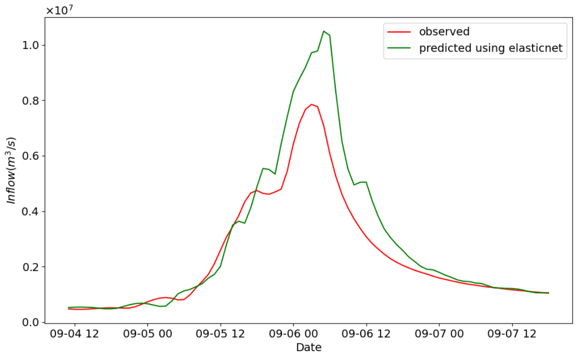


(c) DNN-20

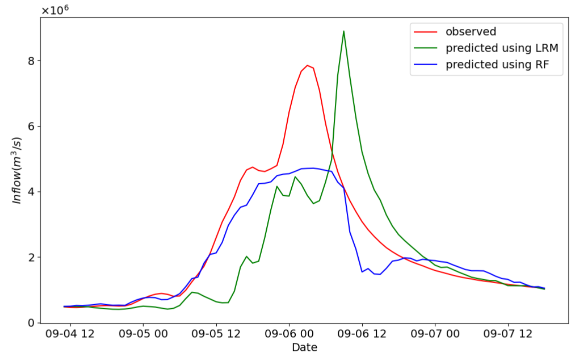


(d) DNN-100

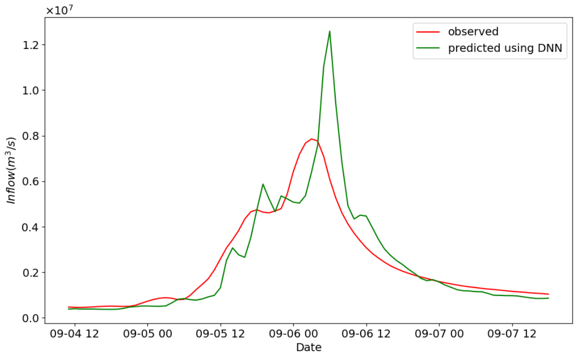
図21　豊平峡ダムの 12 時間流入量における内挿条件の予測結果ハイドログラフ



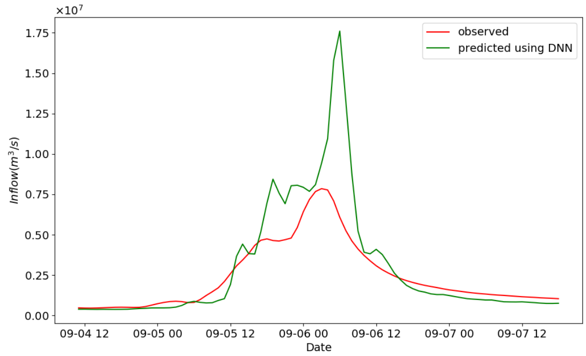
(a) Elastic net



(b) RF-lr

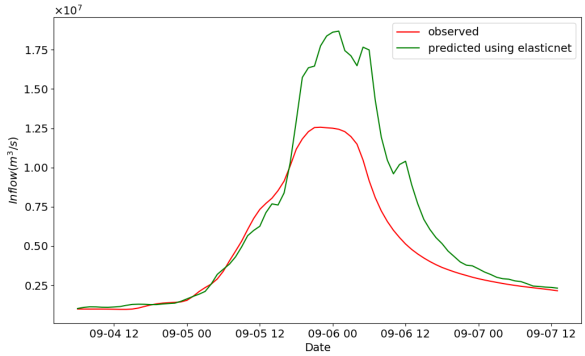


(c) DNN-20

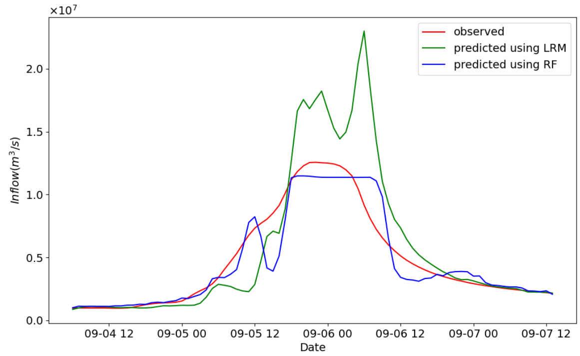


(d) DNN-100

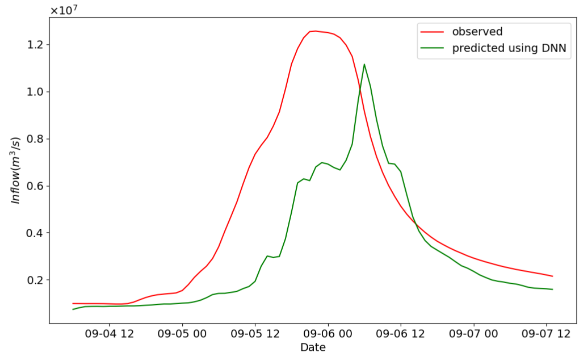
図22　豊平峡ダムの 6 時間流入量における外挿条件の予測結果ハイドログラフ



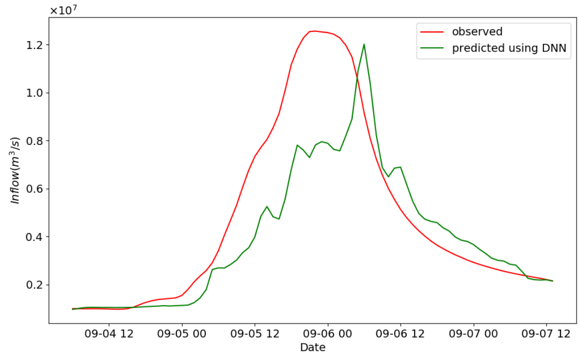
(a) Elastic net



(b) RF-lr



(c) DNN-20



(d) DNN-100

図23　豊平峡ダムの 12 時間流入量における外挿条件の予測結果ハイドログラフ