

ダム流入量予測の精度向上を実現する予測学習

天方 匡純¹・石井 明¹・宮崎 利行²・宮本 崇³

¹正会員 八千代エンジニアリング株式会社 技術開発研究所 (〒111-8648 東京都台東区浅草橋 5-20-8)

E-mail: amakata@yachiyo-eng.co.jp (Corresponding Author)

²非会員 八千代エンジニアリング株式会社 技術開発研究所 (〒111-8648 東京都台東区浅草橋 5-20-8)

³正会員 山梨大学 工学部土木環境工学科

E-mail: tmiyamoto@yamanashi.ac.jp

従来、ダム流入量予測モデルは確定性のある観測データに基づきパラメータ等が最適化され実運用投入されてきた。一方、実運用時のモデルへの入力是不確定性のある予測雨量とするのが一般的である。観測データと予測データは特性が異なり、従来プロセスでは予測雨量が精度向上するまでダム流入量予測の精度向上を期待できない。そこで、モデル構築時と運用時のデータ特性差を小さくするため不確定性のある予測データを含めてモデル学習・構築する。この行為を本稿では「予測学習」と呼ぶ。運用時と同じ不確定性のあるデータ条件下の「予測学習」が運用時のダム流入量予測の高精度化に寄与することを示す。また、予測データに土木的解釈を加えて AI モデルを簡素化できること、「予測学習」では 1~6 時間先のダム流入量予測精度が劣化しないこと、を示す。

Key Words: prediction learning, dam inflow prediction, LSTM, precipitation short-term forecast

1. はじめに

大気中の温室効果ガス濃度の増加に伴い地球上のほぼ全域で気温が上昇するのに対し、降水量は増加する地域と減少する地域があるとされる。そして、日本では、短時間強雨の発生回数が全ての地域及び季節で増加すると予測されている¹⁾。実際、これを証明するかのようになり、平成 25 年 10 月の大島町土石流災害、平成 26 年 8 月の広島土砂災害、平成 27 年 9 月の関東・東北豪雨、平成 29 年 7 月の九州北部豪雨、平成 30 年 7 月の西日本豪雨、令和元年 10 月の東日本台風、令和 2 年 7 月の熊本豪雨と、近年、従来経験したことが無い大雨の多発により、多くの水災害・土砂災害が発生している。

短時間強雨の発生回数だけでなく、無降水日の増加も予測されるなか、私達の生活は様々な水リスクに晒されることになるため、これを平準化する仕組みが必要である。従来、高水と低水の流況格差が大きい我が国において、効率的な水運用を目指して建設されたのがダムである。ダムの貯水能力は一度建設されると柔軟に変えられるものではないが、貯水池運用の工夫により目的毎の効果の向上が期待できる。現在も洪水時に貯水機能を最大限に活用するため、事前放流操作や異常洪水時防災操作といった運用規定が定められている。

しかし、これらの運用は、各自経験に基づく暗黙知に支えられた人的資源に全面的に依存したものであり、団塊の世代の引退等に伴う人員不足とその後の人口減少、更には、前述の従来とは異なる外力の発生頻度の増加に伴い、適切なダム管理を継続するに当たってのウィークポイントとなりつつある。

洪水時に、ダム管理者の意思決定や適切なダム運用を支援するツールとしてダム流入量予測モデルがある。一般的にダム流入量予測に利用される手法は、タンクモデル、貯留関数モデル、分布型流出解析モデルといった物理現象を簡便化したモデルを主体としており、既往洪水に対してモデルを検証してモデルパラメータを最適化する。しかし、入力条件やモデル条件等には数多くの不確定性が含まれ、リアルタイムの流入量再現も難しい現実がある。そこで、昨今、その不確定性を解消する手法として、ベイズ理論に基づく粒子フィルタ²⁾³⁾⁴⁾による状態方程式のデータ同化が適用され始め、予測精度向上が期待される。

一方、上下流の水文量の相関関係からダム流入量を予測する手法として、単相関、重相関、ニューラルネットワークといった手法が存在し、最近では、DNN (Deep Neural Network) を活用した研究も見られる。DNN は、現在、画像認識、音声認識、自然言語処理の分野で積極

的に研究が進み、産業界への具体的な適用も始まっているところである。各分野で適応される DNN のモデル構造は様々であり、洪水予測の分野でも 4 層の DNN を活用した事例⁹⁾がある。また、自然言語処理のように前後の言葉（データ）の繋がりを認識する必要がある時系列データ分野の代表的モデルとして LSTM（Long Short-Term Memory）⁹⁾が挙げられ、これらを活用した事例⁸⁾も見られる。

しかし、いずれも革新的な予測精度向上には繋がっておらず、実運用では高々数時間先の予測精度を確保するに留まる例が多い。一般的に下流域河川では避難のために少なくとも 6 時間先予測が求められ、ダム管理では異常時洪水時防災操作のため少なくとも 6 時間先予測が求められる。本稿では、これらのニーズに対応する成果として、従来と異なる予測フレームワークによりダム流入量予測を高精度に予測できること等を示す。

2. ダム流入量予測の課題

ダム流入量予測に関する既往研究を整理すると共にその課題を整理する。

(1) 既往研究の整理

気候変動による外力変化が想定されるなか、下流域の被害最小化のためにダムの治水・利水ポテンシャルの向上は不可欠であり、ハード整備と共にソフト整備の一層の効果発現が期待される。しかし、ダム下流域で利用される河川水位予測に関わる研究と比較して、ダム流入量予測に関わる研究数は少ない。

ダム流入量予測に関わる研究テーマは大きく 2 つに分けられる。1 つ目は、ダム流入量の予測精度に関わるテーマ、2 つ目は運用に関わるテーマである。前者のダム流入量の予測精度に関わるテーマは、流出解析モデルのパラメータ最適化に関わる研究^{10,11)}、流出解析モデル以外の手法も含めた抜本的な予測手法そのものに関わる研究^{12,13)}に分けられる。一方、運用に関わるテーマは、超過洪水¹⁴⁾、気候変動モデル^{15,16)}、アンサンブル予測雨量^{17,18)}等の外力に応じて被害最小化を目指した運用の研究が進められている。

(2) ダム流入量予測の課題

上記の既往研究の特徴は、実績データ、あるいは、シミュレーションデータといった確定データに対して予測精度や運用に関する研究を進めている点である。予測精度向上を課題とする研究では、当然、モデル構築時とモデル検証時と異なるデータを用意する等して、実運用時の未経験データを疑似的に再現する。しかし、実運用時

はあらゆる面で実績データやシミュレーションデータと異なる条件下となる。例えば、実運用時に 6 時間先のダム流入量を予測する場合は、通常はダム上流域で観測されている水位データも雨量データも役に立たない。その時、6 時間先のダム流入量を予測するために利用できるデータは気象庁が提供する降水短時間予報等の不確定性を有する予測データのみである。

しかし、実運用の予測行為の中で必ず生まれる現象不確定性を想定した研究は限定的であり、予測雨量が実測雨量と同じ精度まで向上した前提での研究が大半である。このような研究内容では実運用時のダム流入量予測の精度向上はいつまでも実現しない。予測が対象とするのは不確定性を持つ未来が大前提である。

3. ダム流入量予測のフレームワーク

実運用場面において、ダム流入量を予測するためには、基本的には予測システムへの入力値として予測雨量が必要となる。しかし、予測雨量は不確定性を有し扱いにくい。現状、一般的にはダム流入量予測モデルの構築作業時には予測雨量を利用せず、不確定性を考慮しないモデルを構築する。つまり、既に観測された雨量とダム流入量の関係からモデルを構築することになる。この結果、予測雨量の不確定性がある程度低減して初めて、構築時に期待した予測精度を実運用時に確保できるダム流入量予測モデル、あるいは、システムが出来上がる。

このフレームワークでは、予測雨量の精度が向上するまでダム流入量予測は実効力を持たないことになり、気候変動による外力増大が顕著化するなか、様々なステークホルダーの QOL（Quality of Life）向上が先送りとなる。そこで、本稿では予測雨量の不確定性までを考慮に入れ

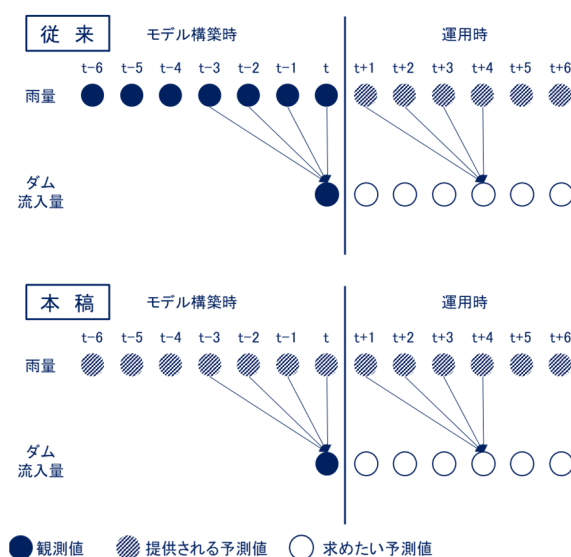


図-1 従来と本稿のダム流入量予測フレームワーク

たダム流入量予測モデルを構築する。具体的には図-1のフレームワークの通りである。図-1の上図は従来の予測モデルの在り方を示している。可能な限り既往現象の再現性の高いモデルを構築するため、モデル構築時の入力値と出力値は両方とも観測値である。この結果、過去に発生した出水に対しては適合度の高いモデルが構築されるが、実運用時に入力値となる不確実性のある予測雨量に対する出力精度が保証されていないモデルが構築される。一方、図-1の下図が本稿での提案フレームワークである。モデル構築時から予測雨量の不確実性と向き合うことで実運用時の予測精度を確保しようとするものである。

具体的には、実運用時と同様に不確実性のある予測雨量を入力値とし、観測値であるダム流入量を教師としてモデルを構築する。このため、入力値が観測値の時と比較してモデル構築時のモデル精度は低下する可能性がある。しかし、モデル構築時から予測雨量を活用したモデル精度確保を目指すため、モデル構築時と実運用時のデータ環境格差が小さくなり、モデル構築時と実運用時の予測精度を同様にできる可能性が高まる。このように不確実性を有する予測データを含めてモデル学習・構築する行為を本稿では「予測学習」と呼ぶ。

ここでは、入力値に予測雨量を用い、これを流出解析モデルを介して予測データを算定する。ただし、入力値である予測雨量と受け皿である流出解析モデルに不確実性が発生するため、これらの不確実性を低減するため以下を考慮する。

- ①1 kmメッシュ単位の予測雨量を単純平均して5 kmメッシュ単位とし流出解析モデルの入力値とする
- ②流出解析モデルの出力値をLSTMネットワークを介して出力し、教師データとなる観測値との誤差を調整する
- ③LSTM ネットワークに入力する際に予測する時刻に対して冗長的に入力する

更に、不確実性のある予測データに土木的解釈を含めた場合のモデルへの影響程度を評価する。具体的には、LSTM ネットワークに対して、土木的解釈を具現化した流出解析モデルの出力値を入力する。そして、土木的解釈程度の違いによって生成されるLSTM ネットワークの複雑さの違いを確認する。LSTM ネットワークの複雑さは、層数、LSTM ブロック数で判断する。

図-2に示す通り、LSTM ネットワークへの入力値となる予測データは、一つは気象庁が全国配信する土壤雨量指数（以下、システムAと呼ぶ）、もう一つは個別流域特性を反映したタンクモデルの計算結果（以下、システムBと呼ぶ）とする。何れも予測雨量（降水短時間予報）

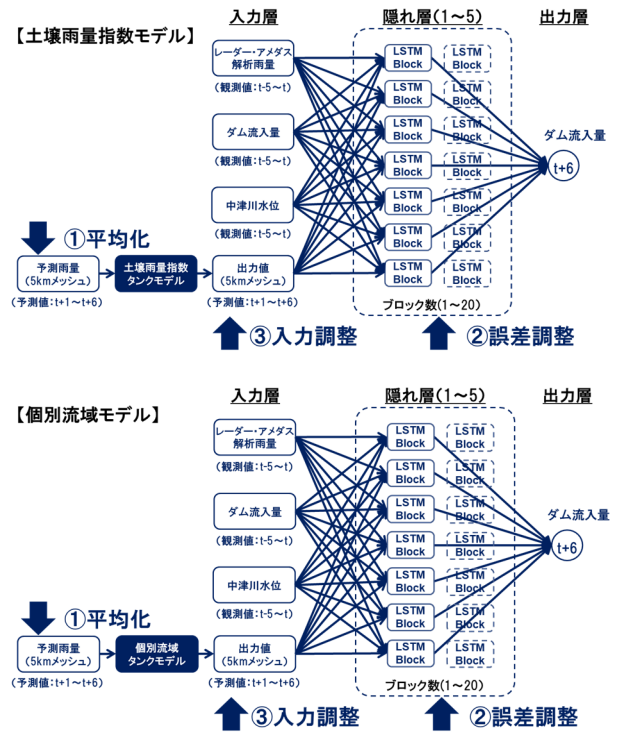


図-2 不確実性を考慮したフレームワーク

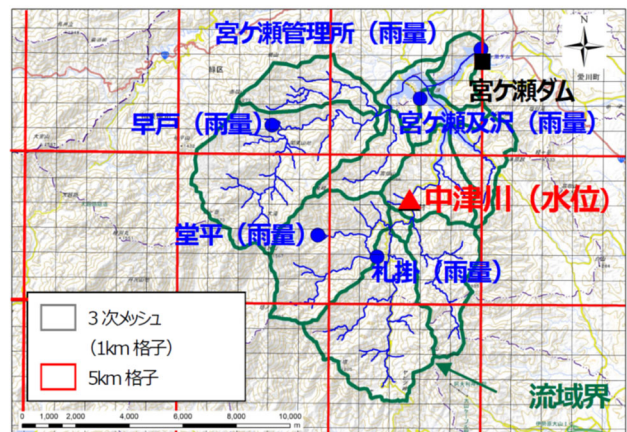


図-3 宮ヶ瀬ダム上流域概要図

を加工して土木的解釈を加えたデータとなる。土壤雨量指数と個別流域タンクモデル計算結果のデータ上の違いは後述する。

4. 対象流域及び利用データ

(1) 対象流域

本稿で取り扱う対象流域は、神奈川県を流れる一級水系である相模川水系中津川に位置する宮ヶ瀬ダムの上流域である（図-3）。宮ヶ瀬ダムの貯水池は同一水系内の相模川、道志川と相模導水で連携接続されているが、これらを含まない宮ヶ瀬ダム単独流域面積は101.4km²である。図-3に示す通り、対象流域はほぼ円形の放射状流

表-1 本稿で活用した水位・雨量データ

	データ名称	データ種類	空間解像度	観測/予測	時間単位	本稿でのデータ利用期間
1	中津川水位	水位	ポイント	観測	時間	2009/12/01~2020/12/31
2	ダム流入量	流量	ポイント	観測	時間	2009/12/01~2020/12/31
4	レーダー・アメダス解析雨量	雨量	5kmメッシュ	観測	時間	2009/12/01~2020/12/31
5	土壌雨量指数	貯留高	5kmメッシュ	予測	時間	2009/12/01~2020/12/31
6	タンクモデル計算結果	流出高	流域	予測	時間	2009/12/01~2020/12/31

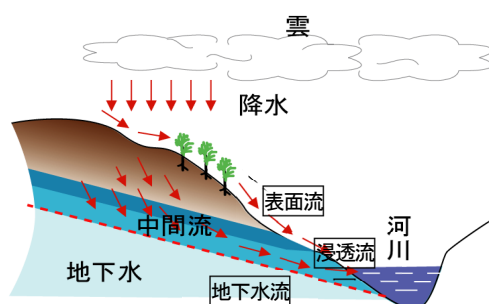


図-4 流出機構の概要

域であり、水位観測所が1箇所、ダム流入量観測所が1箇所、雨量観測所が5箇所ある。ダム地点までの洪水到達時間は既往データ整理より約1時間未満である。

(2) 利用データ

図-3に示す通り、流域内には複数観測所が存在するが、本稿で用いる観測所データは水位データとダム流入量データのみである。他には、レーダー・アメダス解析雨量及び予測雨量（降水短時間予報）を入力値とする土壌雨量指数及び個別流域タンクモデル計算値を用いる。本稿で活用するデータ諸元を表-1に示す。なお、土壌雨量指数及び個別流域タンクモデル計算値については5章及び6章にて詳説する。データ蓄積年数が最も短い土壌雨量指数のデータ存在期間を踏まえ、2009年12月～2020年12月までを本稿の検討対象期間とする。更に、対象とする洪水は宮ヶ瀬ダムの洪水量 $100\text{m}^3/\text{s}$ を超過した出水とし29洪水とした。ただし、2020年には水位観測所の欠測のため適切なデータセットが作れなかった。また、洪水量超過期間の前12時間と後6時間を加えた期間を1洪水期間とした。

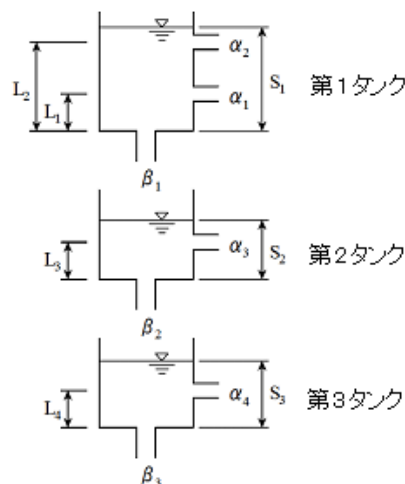


図-5 本稿で用いるタンクモデル構造²⁰⁾

表-2 タンクモデルの各パラメータ値²⁰⁾

	一段目	二段目	三段目
流出孔の高さ (mm)	$L_1=15$	$L_3=15$	$L_4=15$
	$L_2=60$		
流出係数(1/hr)	$\alpha_1=0.1$	$\alpha_3=0.05$	$\alpha_4=0.01$
	$\alpha_2=0.15$		
浸透係数(1/hr)	$\beta_1=0.12$	$\beta_2=0.05$	$\beta_3=0.01$

5. 土壌雨量指数の概要

4章で紹介した土壌雨量指数の概要及び本稿で用いるデータ特性を記す。

(1) 土壌雨量指数の概要

土壌雨量指数は、降雨による土砂災害危険度の高まりを把握するための指標であり、各地の気象台が発表する大雨警報（土砂災害）や土砂災害警戒情報等の判断基準に用いられている。大雨に伴い発生する土砂災害（がけ崩れ・土石流）は、現在の降雨だけでなく、これまでの降雨による土壌中の水分量が深く関係している。土壌雨量指数は、降雨が土壌中に水分量としてどれだけ溜まっているかをタンクモデルを用いて数値化したものである。石井ら¹⁹⁾は土壌雨量指数の予想値を活用して 100km^2 未満流域の6時間先の水位予測の飛躍的な精度向上を達成している。これを踏まえ、本稿では図-2に示す通り、土壌雨量指数をダム流入量予測に活用する。上記タンクモ

デルは日本国内の陸上を対象に地表面を約5km四方の領域に分割して、それぞれの領域毎に図-5の直列3段タンクモデルを対応させている。ただし、流出孔の高さ、流出係数、浸透係数の各パラメータは表-2の全国一律の値²⁰⁾である。土壌雨量指数は2009年12月から30分間隔で実測降雨（レーダー・アメダス解析雨量）による現況値のほか、降水短時間予報を用いた6時間先までの予想値が配信されている。宮ヶ瀬ダム上流域での約5km四方の格子分割は図-3の通りであり約5km四方格子が4つ含まれる。

(2) 土壌雨量指数の計算

土壌雨量指数の計算には図-5に示すタンクモデルが用いられている。3段の各タンクの側面には水が周りの土壌に流れ出す流出孔が、底面には水がより深い土層に

染み込む浸透孔がある。図-4 に示す流出機構に対応して、第1タンク側面の流出孔からの流出は表面流出を、第2タンク側面の流出孔からの流出は表層での浸透流出を、第3タンク側面の流出孔からの流出は地下水流出を表現している。また、第1タンクへの流入は降水に、第2タンクへの流入は第1タンクの浸透孔からの流出に、第3タンクへの流入は第2タンクの浸透孔からの流出に対応している。入力値はレーダー・アメダス解析雨量や降水短時間予報であり、5 kmメッシュの土壤雨量指数を推定計算している。タンクモデルの計算結果は、単位面積当たりの高さ表記で整理され、降雨量の高さ表記に合わせて、流出高、貯留高、浸透高のように表現される。なお、本稿で扱う計算結果は、土壤雨量指数と同じく、降水短時間予報を入力とした場合の、第1タンク、第2タンク、第3タンクの貯留高を合計した数値である。

6. 個別流域タンクモデルの概要

4章で紹介した個別流域タンクモデル計算値の概要及び本稿で用いるデータ特性を記す。

(1) タンクモデルの概要

土壤雨量指数の活用は、自らが計算せずとも予測データが入手可能であること、全国で用いることが可能である等の利点がある。一方、個別流域の流出形態を反映していない、貯留高しか利用できない、といった欠点もある。つまり、土木的解釈をより加えるのであれば、流域特性に合致したタンクモデルを構築し、ダム流入量と流出形態上の特性が類似した流出高を用いることが合理的である。なお、様々な流出解析モデルが存在するなか、タンクモデルを選択した理由は、土壤雨量指数に利用されている手法であり比較がしやすいため、また、本稿では流出現象の精密な再現よりも土木的解釈の違いを表現したいため、である。

以上から宮ヶ瀬ダム流域特性に合致したタンクモデルを構築し、流出高を算出する。タンクモデルの構造自体は図-5の3段タンクモデルとし、宮ヶ瀬ダム上流域を1つのタンクモデルで表現する。そして、図-2に示した宮ヶ瀬ダム上流域内の1km四方の降水短時間予報を流域平均してタンクモデル入力値とした。

(2) タンクモデルの計算

2009年～2019年の宮ヶ瀬ダム地点100m³/s以上の29洪水に対して、2009年～2018年の26洪水で学習してタンクモデルのパラメータを調整し、2019年の3洪水でモデル検証した。ただし、タンクモデルへの入力値は全て降水短時間予報とし、その計算結果を検証するダム流入量

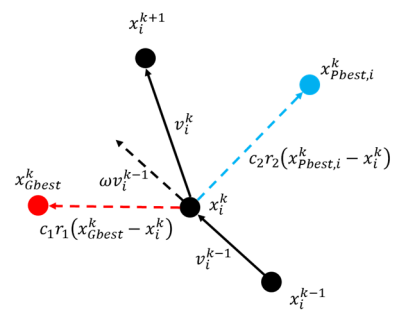


図-6 PSOの探索点の決定方法

は実測値とした。タンクモデルのパラメータ調整はPSO (Particle Swarm Optimization, 粒子群最適化法) を用いた。

PSOは、鳥や魚など群れをなして移動する生物の行動パターンから着想を得た集団探索に基づく確率的最適化手法であり、粒子群が情報を共有しながら解空間を探索する手法である。PSOを用いた理由は、並列処理との親和性が高く、目的関数の自由度が高く、流出解析の最適化計算に用いられる例も多いためである。 n 個の決定係数を持つ目的関数 f を最大化する場合、 n 次元の変数空間中の位置ベクトル x を考え、目的変数 f は位置ベクトル x の関数として表されるものとする。ここで、独立した位置座標 x_i を有する m 個の粒子($i = 1 \sim m$)を与える。全ての粒子は群全体が過去において経験した最良の位置 x_{Gbest} を共有し、また個々の粒子は自身の経験した最良の位置 $x_{Pbest,i}$ を有している。そのうえで、個々の粒子は、群全体の最良位置に近づこうとするベクトル、自身の最良位置に戻ろうとするベクトル、および自身が有している慣性ベクトルの重み付き相加平均によって次回の位置を決定する。 k 回目の探索において、粒子 i の位置 x_i^k と速度 v_i^k を用いて、 $k+1$ 回目の位置 x_i^{k+1} と速度 v_i^{k+1} は次式で更新される。

$$x_i^{k+1} = x_i^k + v_i^k \quad (1)$$

$$v_i^{k+1} = \omega v_i^k + c_1 r_1 (x_{Pbest,i}^k - x_i^k) + c_2 r_2 (x_{Gbest}^k - x_i^k) \quad (2)$$

ここで、 $x_{Pbest,i}^k$ は粒子 i の k 回目の探索までの最良位置、 x_{Gbest}^k は群全体の k 回目までの最良位置、 ω は粒子の慣性パラメータ、 c_1 は自身の既往最良位置へ戻ろうとする強度パラメータ、 c_2 は群の最良位置に近づこうとする強度パラメータ、 r_1, r_2 は[0,1]の一樣乱数である。PSOのアルゴリズムを以下に示す。

- [Step 1] 個体の位置 x_i と速度 v_i を乱数で初期化
- [Step 2] 最小化したい目的関数値 $f(x_i)$ を評価
- [Step 3] 個体のそれまでの最良値 $f(x_{Pbest,i})$ と比較、

表-3 予測雨量を入力とした予測時間先毎のタンクモデルの各パラメータ値

	1時間先用			2時間先用			3時間先用		
	一段目	二段目	三段目	一段目	二段目	三段目	一段目	二段目	三段目
流出孔の高さ (mm)	$L_1=9.509$	$L_3=79.742$	$L_4=24.477$	$L_1=18.757$	$L_3=12.281$	$L_4=59.983$	$L_1=15.417$	$L_3=45.943$	$L_4=72.77$
	$L_2=60.938$			$L_2=100$			$L_2=84.67$		
流出係数(1/hr)	$\alpha_1=0.056$	$\alpha_3=0.406$	$\alpha_4=0.504$	$\alpha_1=0.071$	$\alpha_3=0.59$	$\alpha_4=0.058$	$\alpha_1=0.082$	$\alpha_3=0.79$	$\alpha_4=0.038$
	$\alpha_2=0.184$			$\alpha_2=0.166$			$\alpha_2=0.143$		
浸透係数(1/hr)	$\beta_1=0.104$	$\beta_2=0.043$	$\beta_3=0.009$	$\beta_1=0.107$	$\beta_2=0.41$	$\beta_3=0.008$	$\beta_1=0.08$	$\beta_2=0.207$	$\beta_3=0.011$

	4時間先用			5時間先用			6時間先用		
	一段目	二段目	三段目	一段目	二段目	三段目	一段目	二段目	三段目
流出孔の高さ (mm)	$L_1=57.654$	$L_3=10.195$	$L_4=84.93$	$L_1=65.597$	$L_3=5.185$	$L_4=23.412$	$L_1=32.398$	$L_3=0.003$	$L_4=35.458$
	$L_2=100$			$L_2=68.685$			$L_2=79.18$		
流出係数(1/hr)	$\alpha_1=0.055$	$\alpha_3=0.293$	$\alpha_4=0.03$	$\alpha_1=0$	$\alpha_3=0.284$	$\alpha_4=0.022$	$\alpha_1=0.089$	$\alpha_3=0.059$	$\alpha_4=0.953$
	$\alpha_2=0.188$			$\alpha_2=0.14$			$\alpha_2=0.112$		
浸透係数(1/hr)	$\beta_1=0.175$	$\beta_2=0.207$	$\beta_3=0.008$	$\beta_1=0.164$	$\beta_2=0.203$	$\beta_3=0.011$	$\beta_1=0.082$	$\beta_2=0.042$	$\beta_3=0.047$

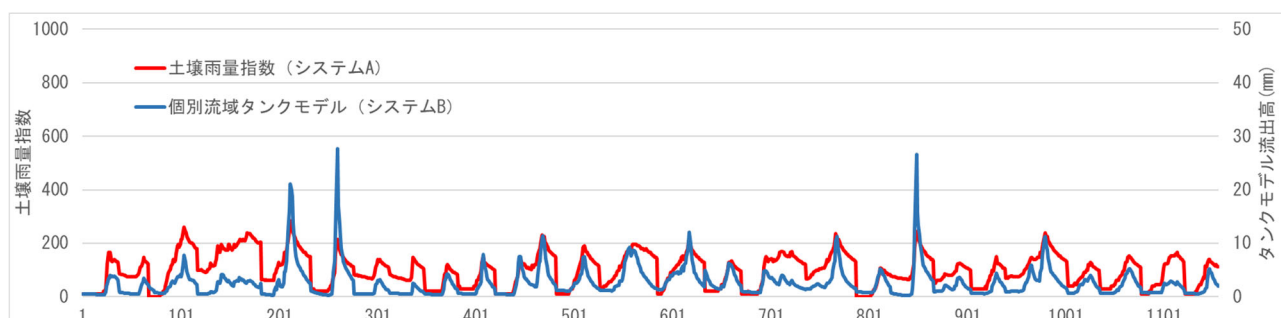


図-7 土壌雨量指数 (1時間先予測, 流域内任意メッシュ) と個別流域タンクモデル (1時間先予測) のアウトプットの違い

もし「 $f(x_i) < f(x_{Pbest,i})$ 」ならば「 $f(x_{Pbest,i}) = f(x_i)$ 」, 「 $x_{Pbest,i} = x_i$ 」と更新

- [Step 4] 群れ全体の最良値 $f(x_{Gbest})$ と比較, もし「 $f(x_i) < f(x_{Gbest})$ 」ならば, 「 $x_{Gbest} = x_i$ 」と更新
- [Step 5] (2)式で速度を更新し, (1)式で位置を更新
- [Step 6] Step 2に戻り, 収束条件を満たすまで繰り返す

タンクモデルへの入力値である予測雨量データの不確定性を考慮し, 予測する1時間先~6時間先のダム流入量毎にタンクモデルを構築する. 本稿でのパラメータ設定主旨が厳密解を求めるものではないこともPSOを選択した理由でもある. なお, 計算初期値は表-2の値を用い, 粒子数 $m = 100$ とした. タンクモデルのパラメータ探索の結果は表-3の通りである. 本章の個別流域タンクモデル計算結果は, 土壌雨量指数と異なり, 降水短時間予報を入力とした場合の, 第1タンク, 第2タンク, 第3タンクの流出高を合計した数値とする.

システムAとシステムBが取り扱う予測データとなる土壌雨量指数と個別流域タンクモデル計算結果の違いを図-7に示す. システムAとシステムBのアウトプットの位相は類似しているが, タンクモデルの機能特性を反映し, 入力である降雨に対応して貯留高の動きは緩やかで流出高の動きは鋭さを持つ. この結果, これらのアウトプットは異なるデータ特性であることを確認できる.

7. 予測精度と土局的解釈の確認方法

本章では, 図-1に示すフレームワークによるダム流入量予測精度を確認する. 更に, 土壌雨量指数 (貯留高) (システムA) と個別流域タンクモデル計算結果 (流出高) (システムB) といった土局的解釈程度の異なる入力値に対して, 教師データとの誤差調整役となるLSTMネットワークの複雑さの違いを確認する. 入出力関係の複雑さは, LSTMネットワークの複雑さを判定の手掛かりとする. LSTMネットワークは, 複数層と複数LSTMブロックから構成され, 層が追加される毎に活性化関数を介してネットワーク全体の非線形性が高まり, 教師データが求める情報抽出力を高めていく. すなわち, 層内のLSTMブロック数を増やすことで線形の表現力を高め, 層数を増やすことで非線形の表現力を高めることになる.

(1) ダム流入量の予測精度の確認

予測データを学習する際に, 不確定性を低減するため図-1の①~③を考慮する. ①は入力データの加工, ②は誤差分布の学習, ③は不確定データの処理, に関する考慮である. ①は所定方法に従い処理する内容であるため, ここでは②と③を考慮した検討を進める.

②の対応として, LSTMネットワークの中間層数と

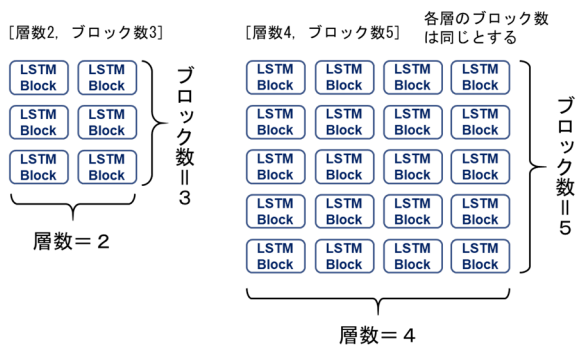


図-8 LSTMネットワークの設定方法

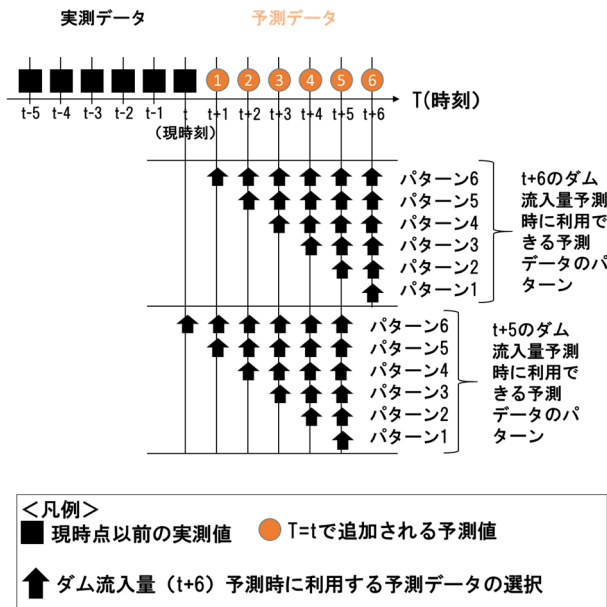


図-9 予測データの利用パターン

LSTM ブロック数の組合せに対して誤差調整の可能性を確認する。利用データが時系列データであるため、ネットワークモデルとして自然言語処理等で利用される LSTM (Long Short-Term Memory) を用いた。ネットワークの複雑さは、層数と LSTM ブロック数で判定する。ネットワーク探索は、層数が 1~5、ブロック数が 1~20 の組み合わせ総当たりとし、図-8 に示す通り、各層のブロック数は同一とする。例えば、層数=3、ブロック数=4 の場合は、中間層が 3 層で各層のブロック数が 4 個のネットワークとなる。この組み合わせは、層数が 1~5、ブロック数が 1~20 の網羅的探索の中で最小 loss の [層数、ブロック数] の組合せが選定される。

なお、層と層の間には活性化関数 relu を配置し、最終中間層と出力層の間のみ恒等関数を配置した。最適化アルゴリズムは Adam ($\alpha = 0.001, \beta_1 = 0.9, \beta_2 = 0.999, \epsilon = 1e - 08$) とし、学習回数は 200 回とした。学習回数 200 回未満で過学習傾向を確認した場合は、その直前で定まるネットワーク条件を最適とした。

③の対応として、不確定データの処理方法を数パター

ンに分けて処理する。本稿では予測データとして土壌雨量指数及び個別流域タンクモデル計算結果を扱う。これらの入力値となる降水短時間予報は 30 分毎に 6 時間先までの 1 時間雨量予測値を提供しているが、本稿では 1 時間毎の値を使った。これに合わせて土壌雨量指数も個別流域タンクモデル計算結果も現時刻 t より先の予測データとして $t+1 \sim t+6$ の 6 時間先までをカバーする。この結果、現時刻 t において $t+6$ のダム流入量を予測しようとした場合、現時刻 t 以前のデータ及び $t+1 \sim t+6$ の予測データを活用できることになる。そこで、現時刻 t にて $t+6$ のダム流入量を予測する際に、 $t+6$ から遡って活用するデータパターンを考慮した。具体的には図-9 に示す通り、パターン 1 からパターン 6 までのデータ形式 6 例を検討する。なお、現時刻 t 以前のシステム A もシステム B もレーダー・アメダス解析雨量を入力とした実測データとして整備した。

②と③のパターンを考慮した上でダム流入量予測の計算ケース毎に層数、ブロック数、index (非線形指数)、RMSE を整理したものを表-4 に示した。非線形 index は layer が増えることによる非線形増を重視し、ネットワークの非線形程度を blocks 数/layer 数と定義したものである。また、ダム流入量予測結果をグラフ化したものを図-10 と図-11 に示す。ただし、図-10 と図-11 は紙面の都合上、システム A とシステム B の 2 種類のフレームワークについて、予測時間毎に最小 loss の [層数、ブロック数] の組合せ結果についてのみ図示した。

(2) 土木的解釈の確認

入力値と出力値の関係の複雑さは LSTM ネットワークの複雑さで判定する。LSTM ネットワークは層と層の間に非線形の活性化関数を介することでネットワークの非線形性を表現する。このため、層数が増えるとネットワークの非線形性が高まる。そして、層内のブロック数が増えると線形の表現力が高まる。

入力情報の土木的解釈性を高めることで入力と出力との関係性は簡便になると想定される。そこで、今回、システム A とシステム B の LSTM ネットワークモデルの複雑さを [層数、ブロック数] で表-4 に整理し、その見える化として図-12、図-13 にグラフ整理する。

8. 結果の考察

(1) 予測学習による予測精度

洪水予測やダム流入量予測のためのモデル構築では現象再現性が要求される。これはこれまでリアルタイムの現場で利用できるデータが限定的であり、予測に活用できるデータ精度が未熟であったためである。この結果、

現象再現性を追求し、そこで構築したモデルを予測に活用するプロセスが主となった。このプロセスは予測で活用するデータと現象再現で活用するデータが同質・同等であれば合理的である。

しかし、実現象を刻印した観測値と気象シミュレーション等で表現される予測値とでは、そもそもデータとしての特性が大きく異なる。つまり、予測と現象再現で活用するデータの同質性・同等性等は保持されない。このため、機械学習分野でいう所の汎化性の確保が難しい。機械学習分野ではこの汎化性を確保することに最大限の努力を払い、汎化性を確保するための様々な工夫が考えられている。ただし、その本質的な所は、学習データ（既にあるデータ）とテストデータ（オペレーション中のデータ）を同質・同等にするということである。そのような観点では、ダム流入量予測のリアルタイムの現場でも予測に活用できるデータのみを使って学習を進めるべきである。幸い、年々、降水短時間予報等の予測精度は向上し、未来予測に不可欠な予測データの積極的活用が可能となり始めている。その成果の一つが例えば図-10と図-11である。図-10と図-11は現時刻 t で得られるレーダー・アメダス解析雨量、中津川水位、ダム流入量を活用し、更に、降水短時間予報を入力した土壤雨量指数、あるいは、個別流域タンクモデル計算結果を予測データとして活用している。図-10と図-11の特徴的な所は、次のようにとりまとめることができる。

〔個別比較〕

- ①RMSEから全体的にシステムAよりもシステムBの予測精度が高い
 - ②システムAの出力値の不確実性が高く、予測結果に揺らぎがある
- 〔以下は両者（システムAとシステムB）共通〕
- ③1時間先予測でさえもピーク値を捉えていない
 - ④1時間～6時間先まで大きく予測精度が劣化しない
 - ⑤6時間先予測でもハイドログラフ形状は捉えている
 - ⑥小規模出水のハイドログラフ形状を適切に捉えられていないものがある
 - ⑦現行運用では6時間先の予測精度に求める仕様は存在しないが6時間先情報量としては十分と考えられる

(2) 土木の解釈の有効性

図-10にシステムAの成果、図-11にシステムBの成果を示す。システムA（土壤雨量指数）は日本全国一律の標準モデルであり、かつ、貯留高をアウトプットとする。このため、土壤雨量指数は、宮ヶ瀬ダム上流域を対象に構築されたシステムBの流出高アウトプットと比較すると土木の解釈力が小さい。このような解釈性の異なる2つのデータを入力値とした場合の違いを整理したものが、図-10、図-11、図-12、図-13となる。

図-10と図-11から分かることは、異なる入力値にも関わらずダム流入量予測の精度には大きな差が無いということである。すなわち、LSTMネットワークの入出力誤差補正が適切に行われているということである。その代わりに、土木の解釈の低いシステムAのLSTMネットワークが複雑な構成になっている。ここでは、layerが増えることによる非線形増を重視し、前述したindexとして表-4に整理した。システムAとシステムBではindexが大きく異なり、より土木の解釈が組み込まれたシステムBの方がLSTMネットワークの複雑化を避けることができると言える。ネットワークの簡便化には利点が多い。一つは、ネットワーク構造を理解しやすくなることである。決してAI技術の専門家でない私達土木技術者にとって、ネットワーク構造の複雑化は避けたい所であり、簡便なネットワークによる意味解釈性の高いモデルを扱える利点は大きい。もう一つは、汎化性が得やすいということである。モデル、ネットワークのパラメータ数が増えると過学習と呼ばれる学習データに過剰適合したモデルが作られやすい。特にデータ数が少ないと過学習が発生しやすく、可能な限りパラメータ数が小さいネットワークを構築することが好ましい。

(3) 不確実性の影響

現時刻 t から先の予測を進めるためには基本的には予測データが必要となる。予測データを算出するプロセスは様々であるが、例えば、1時間毎の予測を積み重ねて逐次プロセス的に $t+1$ ～ $t+6$ の予測を行ったとしても、その逐次プロセスが実現象を再現できていない場合、その順序でさえ不確実性を有する。本稿で活用した予測データの大元となる降水短時間予報も経験的知見や物理学的知見に基づき一連の逐次プロセスの中で提供されているものであるが、そのプロセスは不確実性を有する。このため、得られた予測データを逐次的に扱うのではなくランダムに扱う方法も1つの不確実性対応方法である。しかし、本稿では図-9に示す通り、経験的知見や物理学的知見に基づく逐次性には留意しつつ、数値1つ1つの不確実性に配慮し、データパターンで対応することとした。その結果を図-12、図-13に整理した。

図-12は、横軸にどのくらい先を予測するかを示すリードタイム、縦軸の第一軸にシステムBのRMSEと非線形index、そして、システムAのRMSEを、第二軸にシステムAの非線形indexを取った。第一軸と第二軸を分けたのは、システムAとシステムBの非線形indexが大きく異なったためである。グラフは $t+1$ ～ $t+6$ に対応した6つの予測データ（土壤雨量指数 or 個別流域タンクモデル計算結果）のうち何個を使ったかでパターン分けし、各リードタイムのRMSE及び非線形indexでその影響度を確認した。RMSEについては、どのデータパターンに

表-4 網羅的探索結果

[土壌雨量指数]

リード タイム	selected time range																							
	1				2				3				4				5				6			
	layers	blocks	index	RMSE	layers	blocks	index	RMSE	layers	blocks	index	RMSE	layers	blocks	index	RMSE	layers	blocks	index	RMSE	layers	blocks	index	RMSE
1時間先	3	14	2744	68.2	4	8	4096	64.0	3	11	1331	57.2	4	8	4096	60.4	4	14	38416	68.0	4	9	6561	58.9
2時間先	1	17	17	110.7	1	8	8	98.0	1	20	20	88.8	2	13	169	83.3	2	14	196	98.8	2	19	361	74.2
3時間先	3	6	216	122.4	1	12	12	112.7	3	15	3375	67.4	2	18	324	81.5	2	4	16	96.0	1	20	20	90.4
4時間先	3	6	216	126.2	3	14	2744	77.6	2	18	324	76.3	4	5	625	70.2	2	13	169	74.0	2	19	361	75.2
5時間先	4	17	83521	108.6	3	19	6859	110.0	2	12	144	89.6	4	8	4096	92.5	2	12	144	88.4	3	15	3375	70.4
6時間先	5	3	243	120.6	2	14	196	134.1	3	4	64	120.9	2	15	225	106.6	2	11	121	100.0	3	15	3375	88.6

[個別流域タンクモデル]

リード タイム	selected time range																							
	1				2				3				4				5				6			
	layers	blocks	index	RMSE	layers	blocks	index	RMSE	layers	blocks	index	RMSE	layers	blocks	index	RMSE	layers	blocks	index	RMSE	layers	blocks	index	RMSE
1時間先	1	19	19	79.0	2	5	25	57.5	1	11	11	58.5	2	8	64	54.5	2	10	100	61.3	2	20	400	53.9
2時間先	5	1	1	104.9	1	2	2	102.5	1	18	18	82.2	2	6	36	95.0	2	9	81	84.8	1	11	121	93.0
3時間先	1	7	7	94.8	1	12	12	103.6	1	6	6	92.3	2	5	25	106.2	2	4	16	95.4	2	14	14	94.7
4時間先	1	11	11	77.2	2	2	4	64.3	1	5	5	67.5	2	5	25	68.6	1	13	13	84.4	1	8	8	87.5
5時間先	2	11	121	74.0	2	11	121	77.7	2	4	16	78.7	1	16	16	73.1	2	4	16	68.0	1	4	4	84.3
6時間先	2	9	81	76.6	2	9	81	82.1	1	1	1	87.6	2	1	1	104.5	2	1	1	99.5	1	7	7	101.7

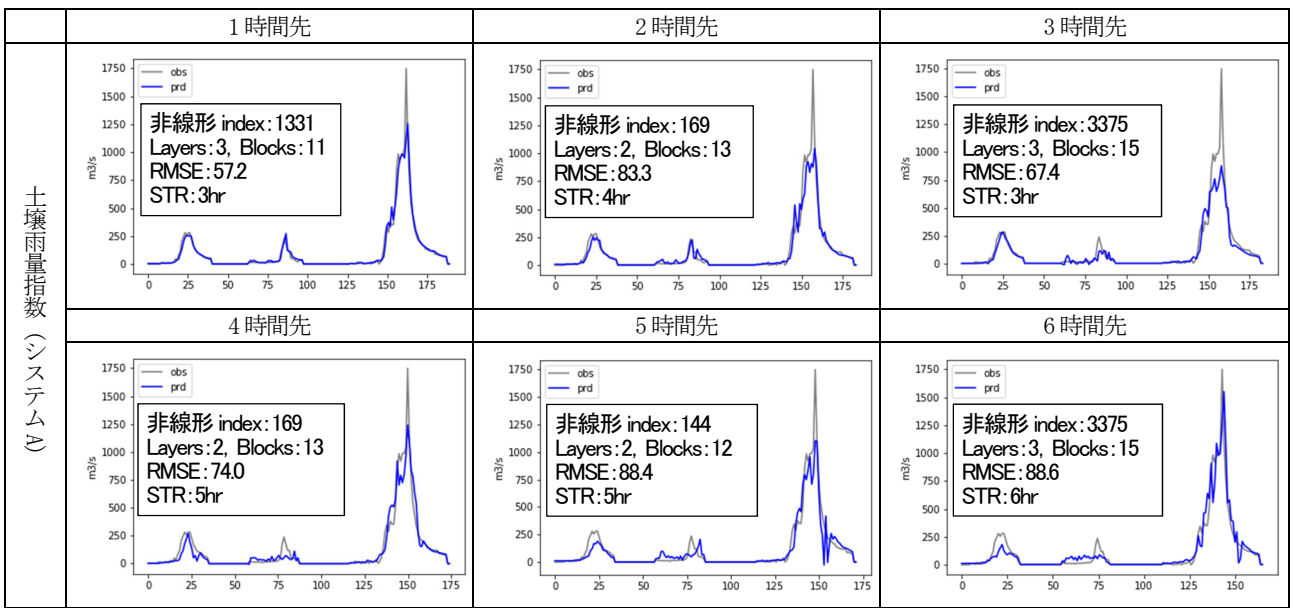


図-10 予測時間毎のダム流入量予測

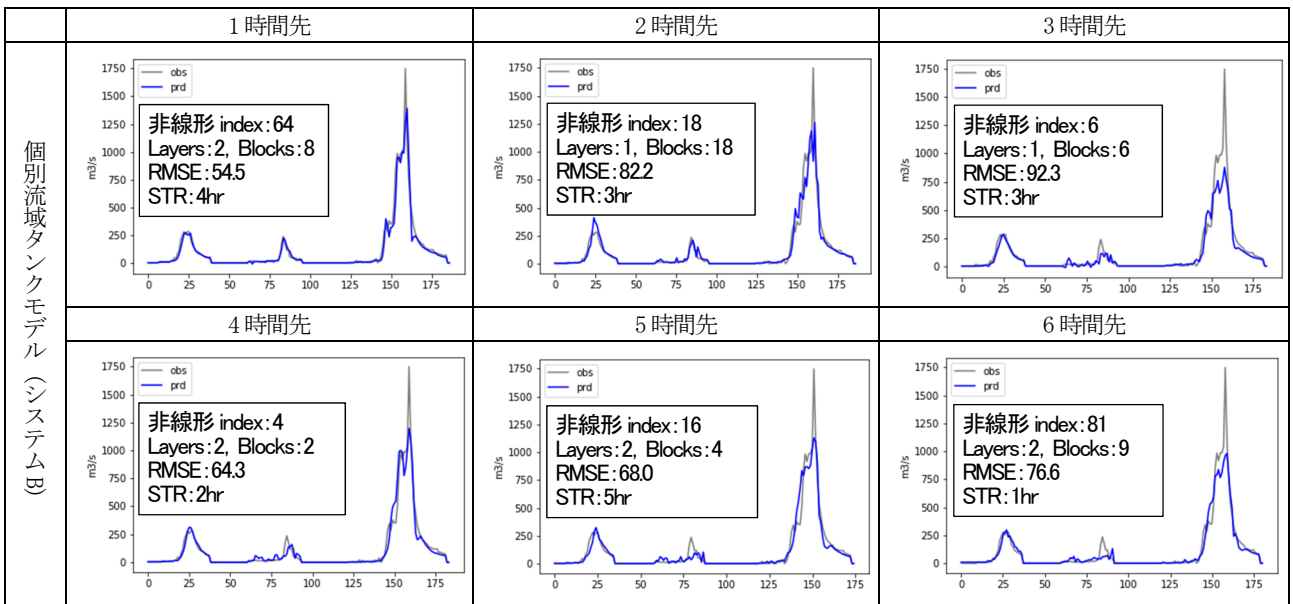


図-11 予測時間毎のダム流入量予測

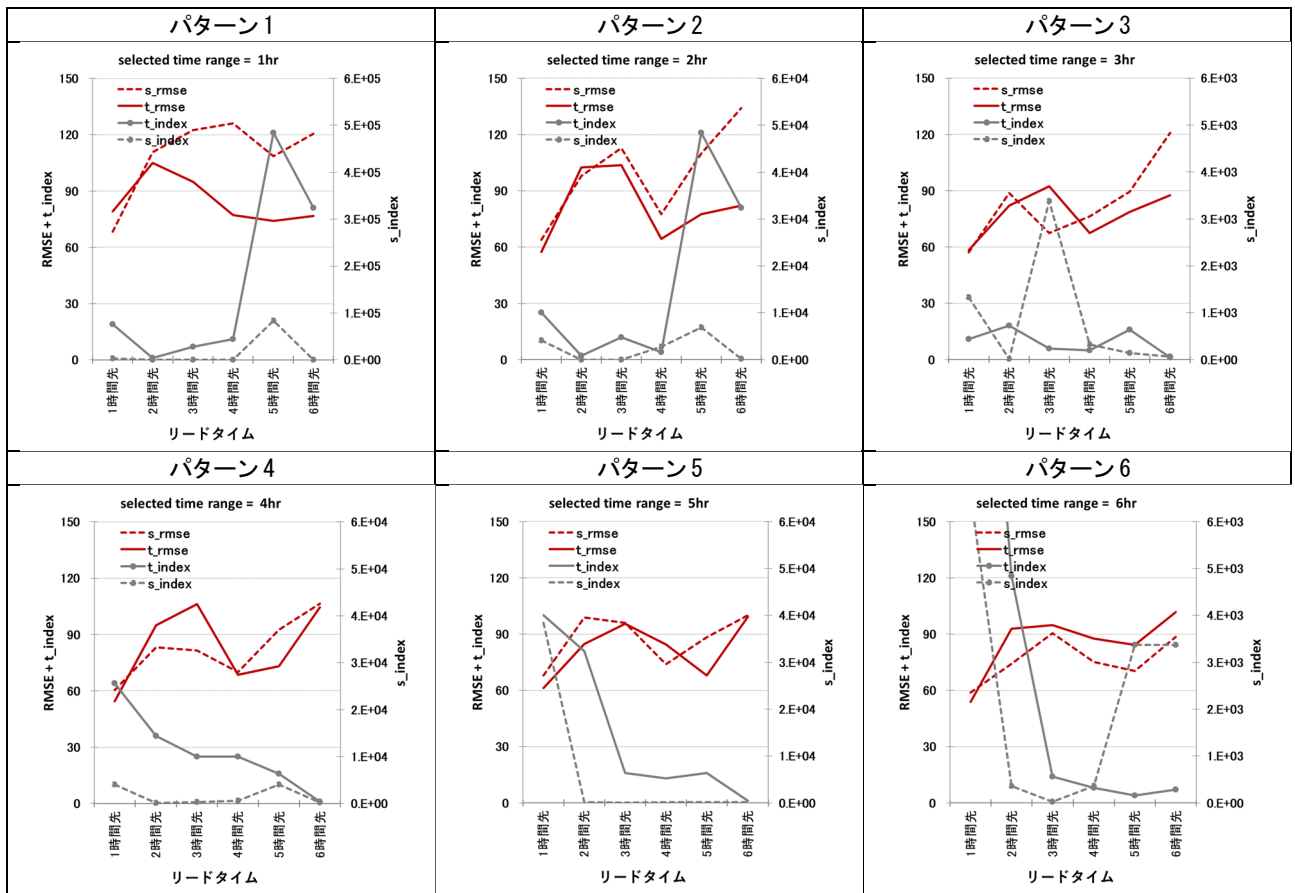


図-12 データパターン毎 (図-9) のリードタイムと RMSE, 非線形 index の関係

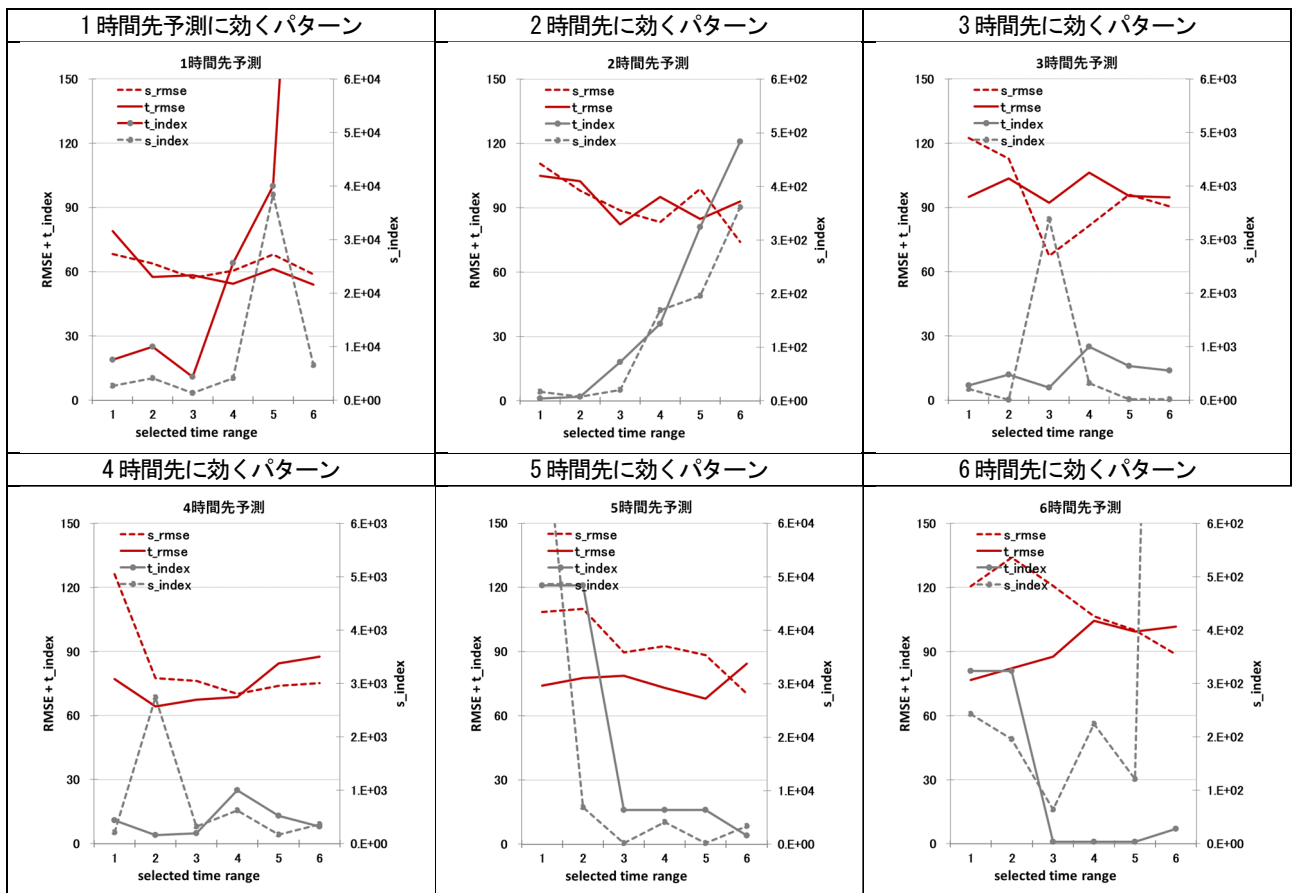


図-13 各リードタイムの予測精度に関するデータパターン (図-9) と RMSE, 非線形 index の関係

においても各リードタイムを通して強い傾向は出ていない。辛うじてリードタイムが1~3時間と進むにつれてRMSEが大きくなり、そこから一旦RMSEが下がり、再びリードタイムが4~6時間と進むにつれてRMSEが大きくなる傾向がみられる。これは、1~3時間先と4~6時間先で降水短時間予報モデル条件を変えている影響と推察されるが詳細は不明である。更には、予測データの利用率と各リードタイムのLSTMネットワークの複雑さの関係傾向を確認することができる。具体的には、特にシステムBで顕著である。予測データを1つ、2つだけ使う場合、数時間先予測には浅いLSTMネットワークで対応しているが、それ以上先の予測には深いネットワークで対応している。一方、予測データを5つ、6つ使う場合、4時間先以上の予測には浅いLSTMネットワークで対応し、数時間先予測には深いLSTMネットワークで対応してしまうことが確認できる。このことは、予測データの順序が重要であることも示唆しており、提供された予測データをランダムに活用した場合は予測精度への悪影響があると想定される。

図-13は、横軸に何個の予測データを使うかを示し、縦軸は図-12と同じである。グラフ全体としては、リードタイム毎の予測に際し、何個予測データを使えば予測精度が良くなるのか、あるいは、ネットワークが簡便になるのか、を示している。1時間先予測を行う場合は、予測データを1つ使っても6つ使っても飛躍的な予測精度向上は見られない。一方、LSTMネットワークは数個の予測データを使った方が簡便化できる。6時間先予測を行う場合も、予測データを1つ使っても6つ使っても飛躍的な予測精度向上は見られない。一方で、LSTMネットワークは予測データを多く使う方が簡便化できる。

図-12からは予測精度と汎化性の観点で、リードタイム相当の予測データ時間分を使うと良いと考えられる。図-13からは予測データを幾つ使うかは予測精度向上の観点よりもLSTMネットワークの簡便化という観点で効果が高いことが確認できる。

9. おわりに

不確定性を有する予測データをモデル学習時に活用する「予測学習」により洪水到達時間1時間未満の流域において6時間先の高精度なダム流入量予測が可能となった。ただし、他流域や他期間での検証が不可欠である。また、予測データに土木的解釈を加えることで誤差補正の役割を果たすLSTMネットワークを簡便化できることを示した。ネットワークの簡便化によりパラメータ数が減り過学習傾向が低減し汎化性が高まるため、その効果による予測精度向上の可能性について検討が必要である。更に、予測データの順序に留意し、リードタイム相当の

予測データを活用することが適切であることを示した。

土木的解釈を有するモデルと誤差補正するネットワークは両方とも関数でありパラメータを有する。そこで、End to End 学習でパラメータ最適化が進むとより予測精度の高いモデルが構築されるであろう。

参考文献

- 1) 環境省, 気象庁, 国土交通省, 農林水産省, 文部科学省: 気候変動の観測・予測及び影響評価統合レポート 2018 ~日本の気候変動とその影響~, 環境省地球環境局, 2018.
- 2) 工藤亮治, 近森秀高, 永井明博: 粒子フィルタを用いた河川流域における実時間洪水予測, 農業農村工学学会論文集, No.259, 2009.
- 3) 立川康人, 須藤純一, 椎葉亮晴, 萬和明, キムスンミン: 粒子フィルタ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, 第55, 2011.
- 4) 高崎忠勝, 河村明, 天口英雄, 石原成幸: 粒子フィルタを用いた都市貯留関数モデルによる実時間洪水流出予測特性, 土木学会論文集 B1 (水工学), Vol.268, No.4, 2012.
- 5) 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), Vol.72, No.4, 2016.
- 6) Hochreiter, S. and Schmidhuber, J: Long short-term memory, Technical Report FKI-207-95, Fakultät für Informatik, Technische Universität München, 1995.
- 7) Hochreiter, S. and Schmidhuber, J: Long Short-Term Memory, Neural Computation, 9(8), 1997.
- 8) 天方匡純, 梁田信河, 藤井純一郎: LSTM (Long Short-Term Memory) を活用したダム流入量予測に関する研究, 電力土木, No.405, 2020.
- 9) 天方匡純, 安野貴人, 藤井純一郎, 島本ゆり, 大久保純一: IoT を想定したダム流入量予測精度向上に関する研究, 水文・水資源学会誌, 第32巻, 第6号, 2019.
- 10) Ishihara, Y. and S. Kobatake: Runoff Model for Flood Forecasting, Bull.D.P.R.I., Kyoto Univ., 29, 1979.
- 11) 杉原成満, 福田慎哉, 倉本和正, 荒木義則, 朝位孝二, 古川浩平: SCE-UA 法を用いたタンクモデルの構築とそれを用いた土砂災害発生危険基準線の設定, 土木学会論文集 F6 (安全問題), Vol.67, No.1, 2011.
- 12) 天方匡純, 藤井純一郎, 梁田信河: Gradient Boosting の適用によるダム流入量予測の精度向上, ダム工学, 30巻, 1号, 2020.
- 13) 田村和則, 加納茂紀, 三浦心, 山脇 正嗣, 金子拓史: ダム流入量長時間予測への深層学習の適用 -ダム防災操作の効率化を目指して-, 土木学会論文集 B1 (水工学), 74巻, 5号, 2018.
- 14) 沖岳大, 中津川誠: 超過洪水による氾濫被害を軽減するための治水施設の総合的な効果分析手法の研究, 土木学会論文集 B1 (水工学), 73巻, 4号, 2017.
- 15) 川村一人, 中津川誠, 杉原幸樹: 気候変動による利水への影響を踏まえたダム貯水池群の最適操作に関する研究, 土木学会論文集 B1 (水工学), 68巻, 4号, 2012.
- 16) 野原大督, 鈴木俊亮, 佐藤嘉展: 将来の河川流量変化に伴うダム利水操作への影響の評価と適応策に関する基礎的検討, 土木学会論文集 B1 (水工学), 74

- 卷, 5号, 2018.
- 17) 猪股広典, 川崎将生, 工藤俊: アンサンブル予測雨量を用いた洪水時ダム操作における操作決定方法に関する研究, 水文・水資源学会誌, 34巻, 1号, 2021.
- 18) 猪股広典, 川崎将生, 後藤祐輔, 増田有俊, 荒木千博, 荒木孝之: アンサンブル予測雨量のダム洪水調節操作への適用に関する研究, 水文・水資源学会誌, 28巻, 6号, 2015.
- 19) 石井明, 天方匡純, 宮崎利行: 土壌雨量指数の予測学習を活用した洪水予測, 土木学会論文集 B1 (水工学), 2021 (投稿中).
- 20) 気象庁, 土壌雨量指数, <https://www.jma.go.jp/jma/kishou/known/bosai/dojoshisu.html> (最終閲覧日: 2021年6月30日)

(Received June 30, 2021)
(Accepted August 31, 2021)

PREDICTION LEARNING REALIZING THE IMPROVEMENT OF DAM INFLOW PREDICTION PRECISION

Masazumi AMAKATA, Akira ISHII, Toshiyuki MIYAZAKI and
Takashi MIYAMOTO

We have optimized the parameters of dam inflow prediction models based on certain observation data and have thrown them into real sites until now. On the other hand, we have generally thought the input data into models as uncertain predictive rain given by Japan Meteorological Agency. There are differences in data character between observation and predictive data, so we cannot expect the improvement of dam inflow prediction precision until the advance of predictive rain precision. Therefore, we offer to construct and learn models including predictive data to minimize data differences between building and operating models. In this thesis, we call this process Prediction Learning. We indicate that Prediction Learning under uncertain circumstances of operating conditions enables dam inflow prediction in operation to predict in high precision. And we show that civil engineering interpretation adding to predictive data allows the AI model to be easy, and dam inflow prediction precision one to six hours ahead is not decreasing.