

# データ非線形性とダム流入量予測精度に関する研究

## Study on the relationship between data non-linearity and dam inflow discharge prediction accuracy

天方匡純<sup>\*1</sup> 安野貴人<sup>\*1</sup> 藤井純一郎<sup>\*1</sup> 嶋本ゆり<sup>\*1</sup> 大久保順一<sup>\*1</sup>  
 Masazumi Amakata Takuto Yasuno Junichiro Fujii Yuri Shimamoto Junichi Okubo

<sup>\*1</sup> 八千代エンジニアリング株式会社  
 Yachiyo Engineering Co., Ltd.

**Abstract:** We aim at improving the predictive precision of the dam inflow discharge using the correlation between the upper stream data and the lower stream data. When we handle the correlation between the upper and lower stream data, we focus on the non-linearity the data have and we have to change the way of modeling according to the degree of the non-linearity. In this paper, we gave the sequential data generated by the distributed flow analysis model the deviation in order to express the non-linearity of the data. After that, we selected LSTM which is a kind of deep learning network and made the predictive model of the dam inflow discharge learning the non-linearity data. As the result, we knew that the deviation which the normal observation values have doesn't influence the complexity of deep learning model.

### 1. はじめに

気候変動による地球温暖化の影響を受け、各地方でこれまで経験もしたことが無い外力や事象に起因する水・土砂災害に苦しんでいる。これまでの河川整備の水準を超える外力規模・頻度の増大により、従来の治水対策の在り方を抜本的に見直さざるを得ない状況も想定され、まずは人命を救うための可能な限り迅速で正確な情報提供が求められる。洪水時のダム管理に当たっても、下流側の資産集積地の被害軽減やダム自体の安全性確保の観点から、洪水関連情報の早期取得・利用は死活問題であり、その一つのツールであるダム流入量予測の精度向上は喫緊の課題である。

### 2. ダム流入量予測の原理

ダム流入量予測には、図1の通り、主に2つの方法があり、雨量から流入量への流出過程を疑似的に再現する方法と、上流側の雨量や水位から下流側の流入量を予測する方法である。前者はモデルの説明責任を果たせる伝統的手法であるがデータ同化等の技術を駆使しても、入力データやモデル構造等のバイアスを低減することが難しい。一方、後者は上流側と下流側

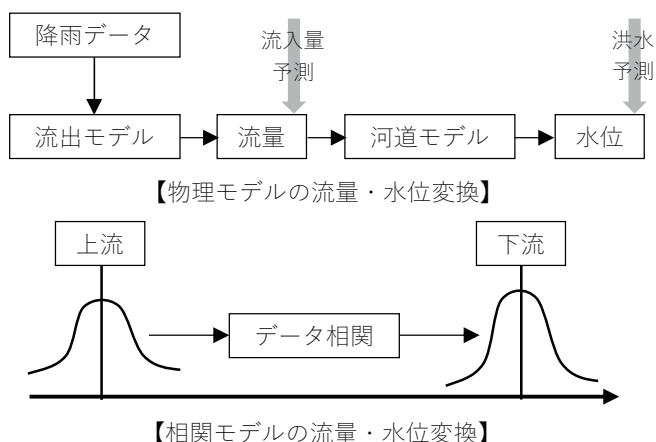


図1 水位・流量予測手法の概要

のデータ相関を活用する手法であり、確保できるリードタイムが短く、また、説明責任を果たせない手法と考えられているが、洪水到達時間内予測精度は前者よりも高い。本稿では、後者の原理に基づくダム流入量予測モデルの留意点の一つであるデータ間の非線形性への対応をとりまとめる。

### 3. 観測データの特徴

ダム流入量と相関性が高い一般的な上流側データは、水位データと雨量データである。水位データは雨量が地面に到達した後の水文過程の結果であるため、よりダム流入量との関係が濃厚である。このため、本稿では検討の簡便化のため上流側データとして水位データのみを扱う。

#### 3.1 水位データの特徴

洪水時の河川は、河道全体に水量が充満し、河道形状等の影響で水面形状は複雑な起伏を持つ(図2参照)。ところが、水位データは通常、一点を捉えた時系列データとなり、洪水時には非常に不確定性の高い振幅を有するデータとなる。

#### 3.2 ダム流入量データの特徴

ダム流入量は、ダム貯水池への接続河道からだけでなく、貯水池総縁辺からの流れ込みにより形成されるため、観測は事実上不可能である。そこで、観測可能な放流量と貯水位から間接的にダム流入量を推定する。ただし、洪水時には大量の水が流れ込むことによる衝撃波等が発生し、貯水池は大きく振動し、貯水位も大きく振動する。この結果、貯水位から間接的に変換したダム流入量も大きく振動する(図3参照)時系列データとなる。



図2 洪水時の水面形状

連絡先: 天方匡純, 技術創発研究所 (RIIPS), 〒111-8648 東京都台東区浅草橋 5-20-8, amakata@yachiyo-eng.co.jp



図3 ダム流入量データ

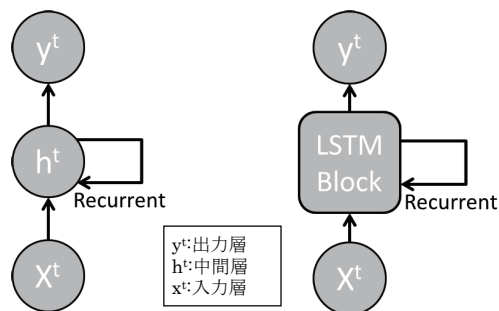


図4 RNN と LSTM の基本構造

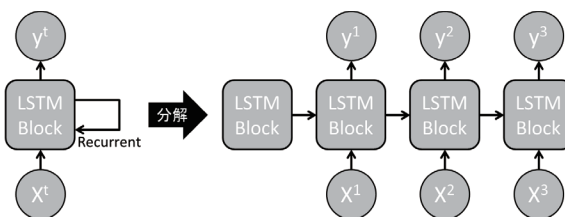


図5 LSTM 構造の分解

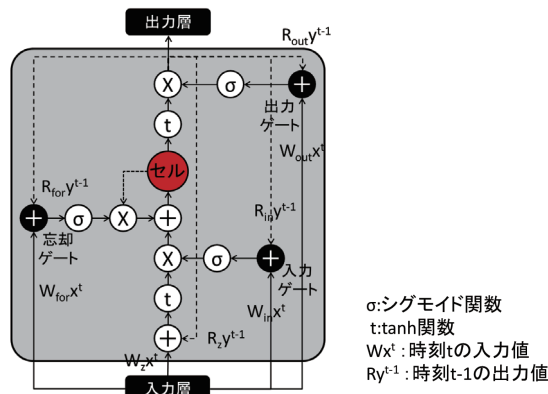


図6 LSTM ブロックの詳細図

#### 4. アーキテクチャーの検討

現時刻以前の情報を保持し、次のステップの予測に役立てることができる DNN (Deep Neural Network) である RNN 系の LSTM (Long Short-Term Memory) を活用し、ダム流入量の予測精度を高めるアーキテクチャーを検討する。なお、扱うデータは前章で記した振幅を有するデータとなるため、それらの雑音を排した汎化性のある出力データを求めることとなる。

##### 4.1 LSTM の概要

DNN のモデル構造は、これまでも洪水予測やダム流入量予測に利用されてきたニューラルネットワークのものと同様している。DNN は、中間層の数を大幅に増やすことができ、非線形性に代表されるより複雑なデータ間の関係を読み解くことが可能である。特に LSTM は、前後関係が重要となる時系列データの特徴量抽出に威力を発揮する。そこで、本稿では LSTM を用いて時系列のハイドログラフの特徴量抽出に適用する。

これまで、我々が洪水予測等に利用してきたニューラルネットワーク[竹村 2013]は、中間層を 1 層(1 層当たりニューロン 7 個)とし、活性化関数にシグモイド関数を用いてきた。しかし、このネットワークは入力データの前後関係を保持する機構を持たず、時系列データを扱う場合は、時刻  $t$  のデータと時刻  $t + \Delta t$  のデータの関係进行学习することにより、時刻  $t$  のデータから時刻  $t + \Delta t$  のデータを予測する。また、入力データに時刻  $t$ 、時刻  $t+1$ 、時刻  $t+2$ 、 $\dots$ 、時刻  $t + \Delta t$  のデータを別々に揃えることで時系列の関係を保持しながら出力データとの関係を構築することも可能であるが、数日先のように予測スパンが長くなると入力データセットが煩雑となる、また、 $X_{rain}$  に代表されるメッシュ雨量等の巨大入力データを利用する際は類似データセットの冗長性は避けたい等を想定すると入力データ構造として限界がある。

RNN 系ネットワークは、入力データの前後関係を保持する機能が整備され、入力データを必要最小限としてデータの前後関係の特徴量を捉え、出力データとの関係を構築することができる。RNN は図 4 のように時刻  $t-1$  以前の中間層を時刻  $t$  の中間

層に紐づけ、入力データの前後関係を保持している。更に、LSTM は、図 4、図 5 のように基本的構造は RNN と同様であるが、中間層をニューロンではなく LSTM ブロックと呼ばれるメモリセルと 3 つのゲートを持つブロックに置き換え、データの前後関係の長期依存性能 (long-term dependencies) [Bengio 1994] を高めている。LSTM ブロックの詳細は図 6 の通りであり、DNN の勾配消失問題をネットワーク構造(図 6 のメモリセルと 3 つのゲートによるデータ制御)により解決している。3 つのゲートは、忘却ゲート、入力ゲート、出力ゲートに分けられ、それぞれシグモイド層によりデータを通す・通さないを判定する。

##### 4.2 取り扱うデータ内容

降った雨が 60~90 分後にダム貯水池に流れ込む流域を想定し、土研モデル[吉野 1990]を活用して分布型流出解析モデル(図 7 参照)を構築した。そのモデルの出力結果である計算流量に人工ノイズを与え、振幅付き観測値を疑似的に再現する(図 8 参照)。ダム上流域に 3 地点の観測所を想定し、3 地点の疑似観測値を活用して LSTM モデルを作成し、ダム流入量を予測する試みとした。図 9 の通り、全 55 出水を扱い、そのうち 52 出水で学習し、3 出水で精度検証した。

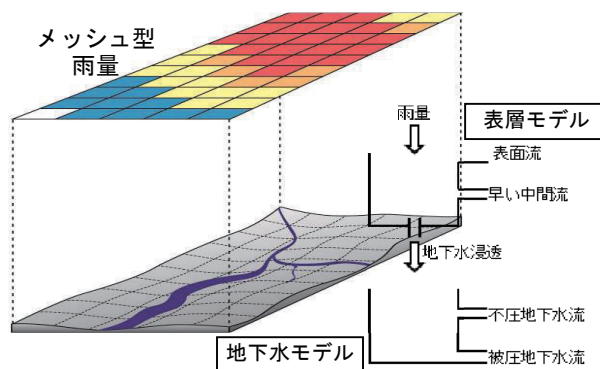


図7 分布型流出解析モデルの概念図

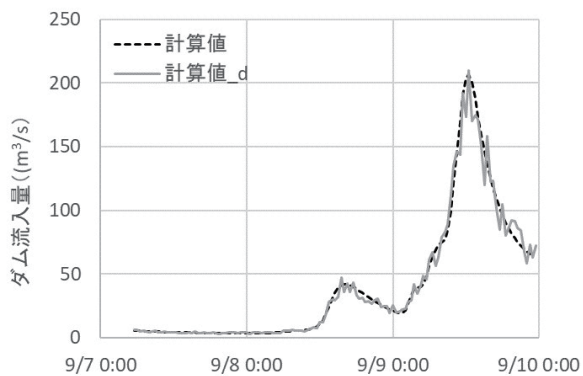


図8 人工ノイズを与えた計算流量（観測値に見立てる）

【学習過程】



【予測過程】

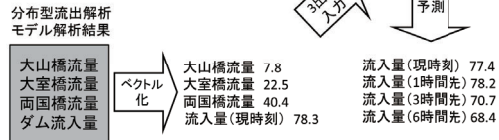


図9 LSTMの学習と予測の過程

表1 LSTMのモデル条件

モデル名	層	ブロック数	出力関数	評価関数	最適手法
LSTM1	1	7	恒等関数	MSE	SGD
LSTM2	1	56	恒等関数	MSE	SGD
LSTM3	2	7×7	恒等関数	MSE	SGD

4.3 アーキテクチャーの設定

学習・予測で用いた LSTM モデルは表 1 の通り. LSTM1 層 (7ブロック)でも他モデルと同精度を確保できることが分かる.

5. おわりに

計算結果にノイズを含めなければ, 1層(7ブロック)の LSTM モデルの計算精度が最良結果となることは別途確認済みである. 本稿では, 表 2~表 4 の結果から観測時に想定されるノイズを含めても LSTM モデルの複雑さが求められるわけではないことを確認した. この結果, ノイズを有する実観測データに対しては, 非線形対応として表現力の高い, 層の深いモデルを選択するよりも平滑化等のノイズ除去の前処理が重要であると考えられる. 一方, 大流域ではノイズとは異なる観測データ間の非線形性が高まると考えられるため, 今後はその非線形性のダム流入量予測精度への影響を確認していく予定である.

参考文献

[竹村 2013]竹村仁志, 眞間修一, 吉田武司, 石上尚:中流域に洪水調節ダムを有する河川における洪水予測システムの精度検証, 河川技術論文集, 土木学会, 2013.  
 [吉野 1990]吉野文雄, 吉谷純一, 堀内輝亮:分布型流出モデルの開発と実流域への適用, 土木技術資料, 土木研究センター, 1990.  
 [Bengio 1994] Yoshua Bengio, Patrice Simoard, and Paolo Frasconi: Learning Long-Term Dependencies with Gradient Descent is Difficult, TRANSACTIONS ON NEURAL NETWORKS, IEEE, 1994.

表2 各 LSTM モデルの計算結果 (H27. 7. 3 洪水)

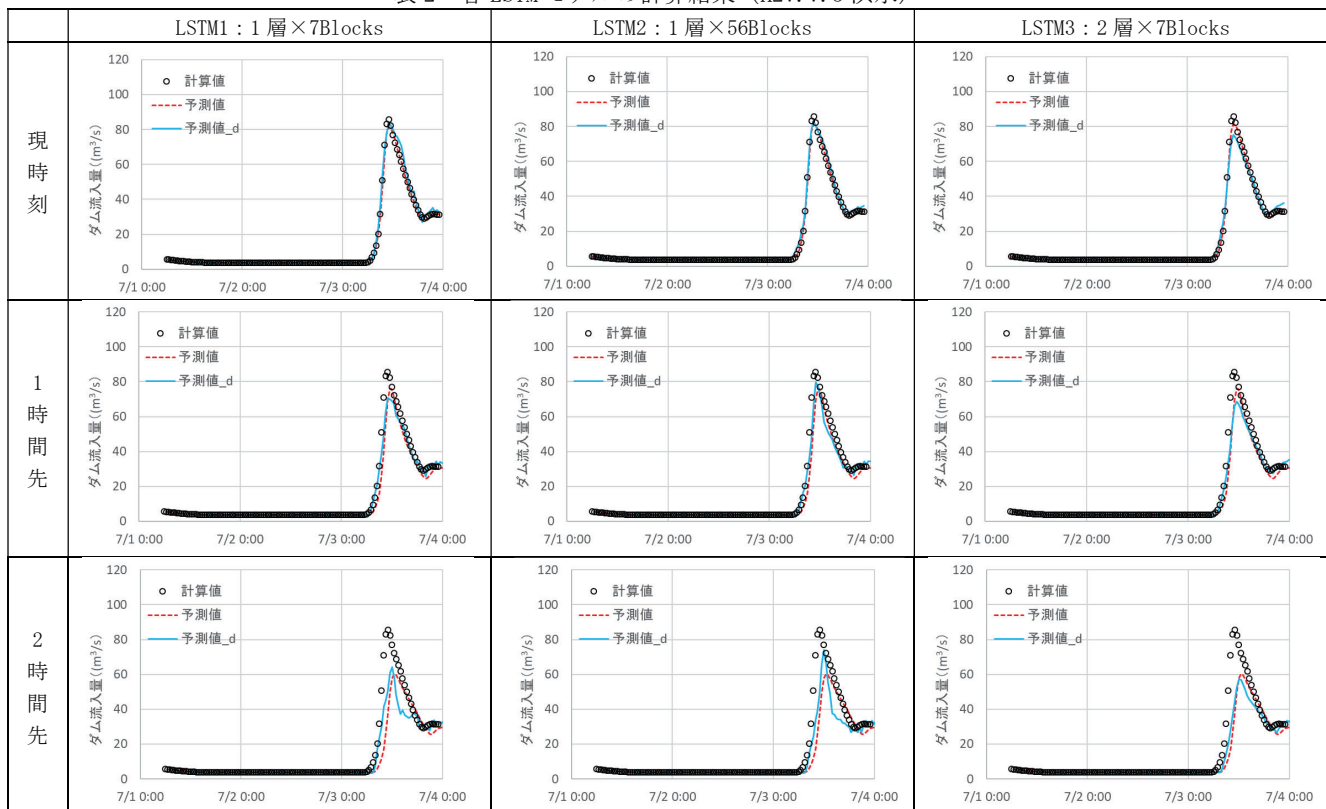


表3 各 LSTM モデルの計算結果 (H27. 7. 16 洪水)

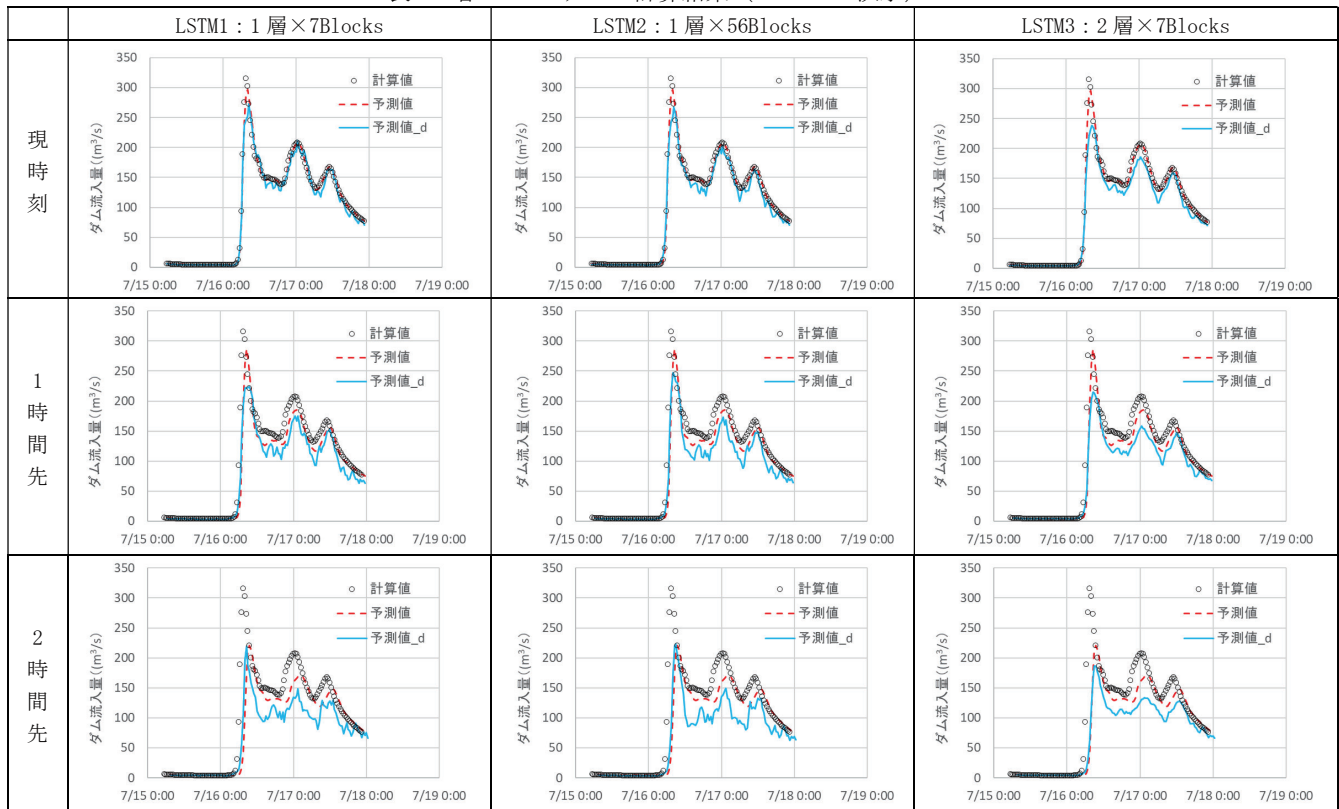
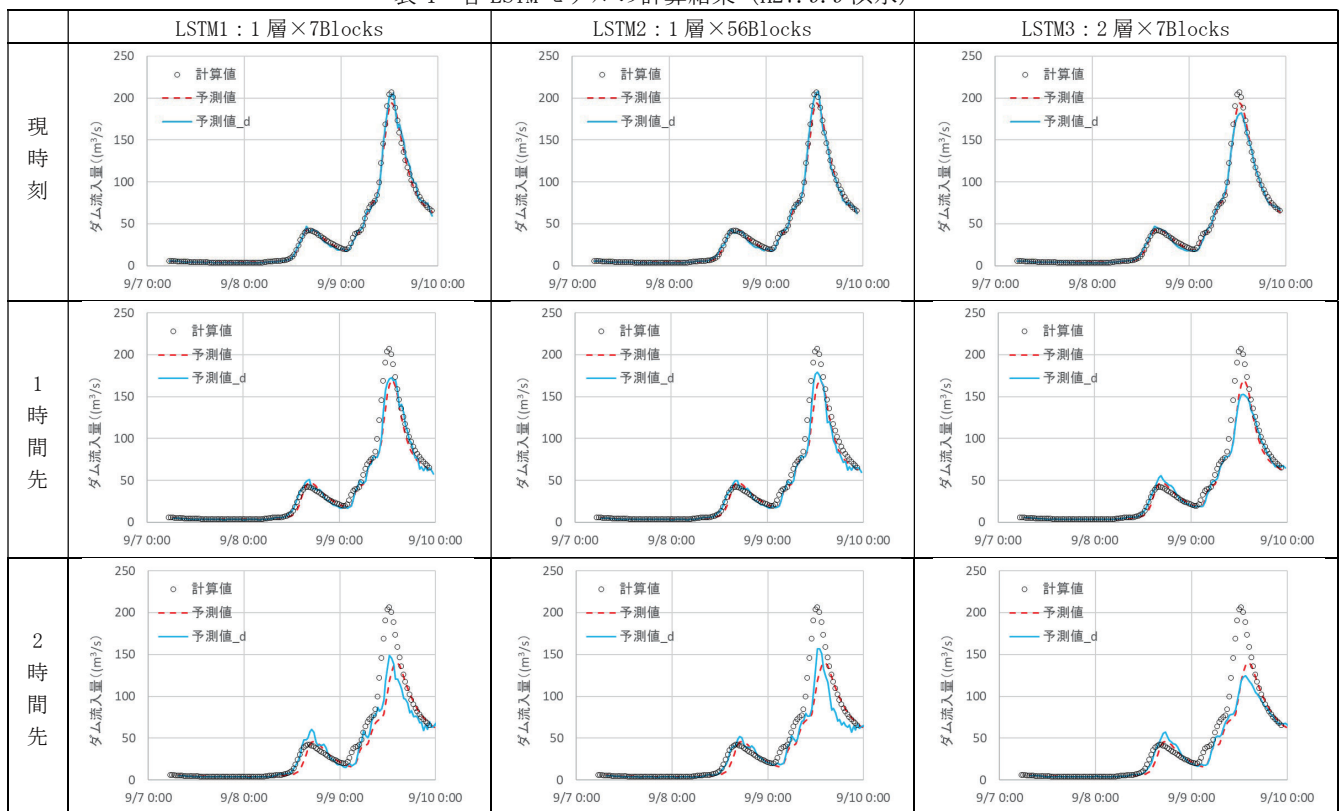


表4 各 LSTM モデルの計算結果 (H27. 9. 9 洪水)



注)表2, 表3, 表4の凡例説明

計算値(○印):分布型流出解析モデルの出力値である真値

予測値(赤点線):真値を基に学習したモデルの出力値(予測値)[どのグラフも1層×7Blocksモデルの計算結果を統一して記載]

予測値\_d(青線):真値にノイズを含めたデータを基に学習したモデルの出力値(予測値)