

# 洪水予測計算における AI を活用した データ同化の高度化（最上川水系）

あべまさみ たきぐちひろき ひだよしひさ うすいえいすけ  
○阿部真己<sup>1</sup>・滝口大樹<sup>2</sup>・樋田祥久<sup>2</sup>・臼井栄佑<sup>2</sup>

<sup>1</sup> いであ（株） イノベーション戦略本部 AI 総合推進室（〒224-0025 神奈川県横浜市都筑区早渕 2-2-2）

<sup>2</sup> いであ（株） 社会基盤本部 国土保全事業部 流域減災部（〒158-0094 東京都世田谷区玉川 3-14-5）

洪水時において、従来は 6 時間先の水位予測が行われてきたが、ダム の事前放流や事前の広域避難等に利活用するためには、予測精度を確保しつつ長期の予測計算（72 時間）が必要である。最上川水系のデータ同化は、カスケード分割した上流～河口の各観測所を対象とした粒子フィルタ法を採用しており、現行の洪水予測システムでは 36 時間先予測の演算時間に 5 分程かかることから、データ同化の計算を簡便化する代替手法の検討を行った（なお、納品した洪水予測システムに実装しているものではないことに留意）。

代替の提案手法では、まず不確実性の高い降水量の大小関係と全域の河川水位波形の高低の応答関係の線形性を仮定することで、河川水位の同化の問題を単純な多重回帰問題に一次的に近似した。次に多重回帰モデルの組み合わせ最適化を深層学習モデルで解かせることで、非常にシンプルな方法で高速化（数秒）する手法を試行的に実装・試運転を行った。加えて、観測結果に物理モデルの表現力を超える値が入り込んだ場合の最適化の安定性に関する簡単な実験も行った。

本手法の予測結果はあくまで物理モデルによるもので、予測の説明責任も果たせる。内容も簡便であるため、調整対象の変数と河川水位の応答の線形性が仮定できる河川（例えばバックウォーターの影響が小さいなど）に限って、一次近似的として日本全国の大規模な流域への展開に期待が持てると思われる。

**Key Words:** データ同化, 河川水位予測の線形性, 深層学習モデル

## 1. はじめに

激甚化する洪水に対して有効な対策を講じていくため、河川水位予測システムは非常に重要な役割を今後も担っていくことが期待される。洪水時において、従来は 6 時間先の水位予測が行われてきたが、ダム の事前放流や事前の広域避難等に利活用するためには、予測精度を確保しつつ長期の予測計算（72 時間）が必要である。

予測の高度化に深層学習を適用していく試み<sup>1)</sup>も見られる一方、深層学習は未学習の状況での予測が不確実である等の懸念がある。例えば敵対的攻撃（Adversarial Attack）<sup>2)</sup>に代表されるような脆弱性とその対抗策も日々新しい研究論文が報告されており、行政による説明責任を果たせる予測とその情報を活用した意思決定には、依然として物理法則に準じたモデルによる予測は重要である。

物理モデルによる予測は、偏微分方程式（連続式と運動方程式）の時空間的な離散化による誤差、設定値や境

界条件となる観測結果の誤差などが含まれるため、必ずしも計算対象である水位等の観測を完全に再現できるわけではない。特に降水予測結果は、気象予測によって配信されている情報を受け取って河川水位の予測を行うが、この予測の不確実性が高く、河川水位の予測精度はほとんど降水量予測の精度に依存している。また、河川部分の予測は 1 次元不定流モデルが一般に用いられるが、数百 m 間隔の測量成果で扱っており、河川の蛇行などの地形条件は粗度係数などに内包されて表現されることが多い。こういった不確実な情報を含む物理モデルを運用するため、観測結果に整合するように不確実な情報を時々刻々と調整するデータ同化<sup>3)</sup>を行うことが一般的である。

最上川水系の計算対象は 44 の支川を持ち、全域に 33 か所の水位観測所（直轄）がある（図-1 上段）。最上川水系のデータ同化は、カスケード分割した上流～河口の各観測所を対象とした粒子フィルタ法を採用しており、現行の洪水予測システムでは 36 時間先予測の演算時間に 5 分程かかる（図-1 下段左側）。予測を 10 分おきに更

新する場合、72 時間先の予測を実現するためには高速化が必須の課題となる。

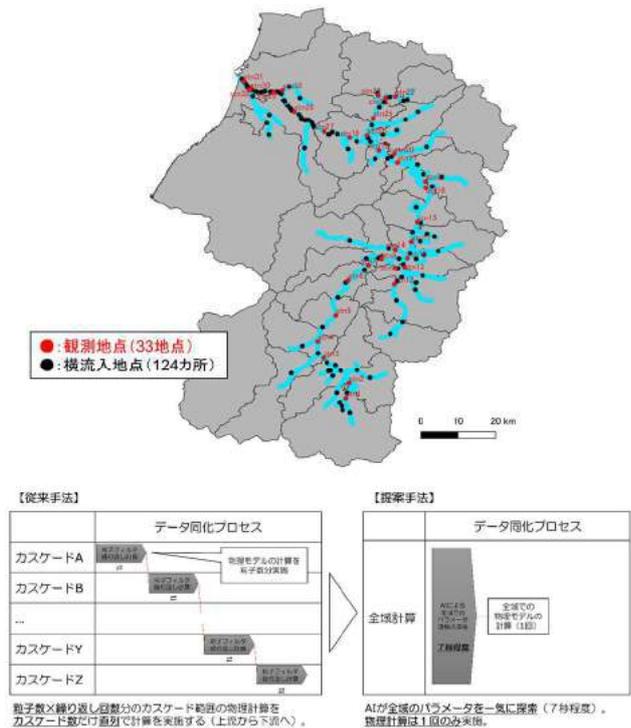


図-1 最上川の地形（上段）とデータ同化プロセス（下段）

カスケード方式の粒子フィルタによる最適化計算では、推定したいパラメータに適切な値を設定し（粒子）、その数だけ実際に物理計算を行い、より観測結果と整合しているものを残していく計算を必要な回数繰り返し実施する計算手法である。そのため、並列化した場合には並列化効率 100%を仮定しても「カスケード内の 1 回分の物理予測に係る時間×繰り返した回数×粒子数÷並列コア数」だけの最適化の計算時間が 1 カスケードにつきかかる。

粒子フィルタに期待される利点としては、①パラメータの確率分布をノンパラメトリックに表現でき、複雑な問題に適切にアプローチできうること、②異常値が設定されてもその粒子は単に棄却されるため異常値などへの頑健性が高いことなどが考えられる。一方で相当の粒子数を用いないとどちらの利点も十分に発揮できない可能性についても懸念があると考えられる。

ここでは、降水量の不確実性のみをデータ同化によって調整（予測された降水量に対して多い・少ないといった倍率を調整）する場合について、現状のカスケード方式・粒子フィルタを代替する効率的な手法について試行的に検討したものである。

まず、降水の多少に対する流域全体の水位の高低の応答特性を整理し、単純な多重線形回帰問題として一次近似可能であることを示す。次に、この多重回帰的なデータ同化問題を観測結果が得られるたびに最適化するため

の深層学習モデルを用いた最適化手法を実装した。最後に本手法の頑健性についても実験を行った。

## 2. 河川水位予測の線形性

### (1) 不良設定 (ill-pose) の解消

流域モデル（分布型モデル）では、流域の降水は 2 次元メッシュで表現された流域内を地形に沿って流下し、124 地点の横流入地点に集積し、1 次元不定流モデルの河川網モデルに受け渡される。1 次元不定流モデルでは 33 か所の水位観測地点において、計算結果と観測結果の誤差を測定している。

本検討では、2次元の流域モデルの同化は行わず、124 地点の横流入地点の大小の調整をデータ同化によって行うこととした。33 地点で観測を行い、124 地点の倍率を推定するため、情報量のバランスからみると不良設定問題となっている可能性がある。感度解析を行ったところ、同じ 2 つの観測地点に挟まれた横流入地点の全域の水位変動に及ぼす影響は高い多重相関性が見られた。以降の問題を解きやすくするため、2 つの同じ水位観測地点に挟まれた横流入地点は同じグループとし、124 の横流入地点を 33 の横流入グループとした。

### (2) 河川水位予測の線形性の仮定

横流入グループの倍率調整による河川水位予測の線形の仮定は以下のように定式化した。非常に簡易な関係式であるが、多重回帰モデルと呼称することとする。

$$H_i(i, \vec{\theta}) = H_{0,i} + \Delta H_i \quad (1)$$

$$\Delta H_i = \sum_{j=1}^M (\theta_j - 1) \times dH_{i,j} \quad (2)$$

$$i \in 1, N \\ j \in 1, M$$

ただし、横流入量地点  $j$  の倍率は  $\theta_j$ 、倍率調整後の水位を  $H$ 、倍率調整前の水位を  $H_0$ 、添え字  $i$  は水位観測地点（1～ $N$ 箇所あり、 $N$ は 33）、添え字  $j$  は横流入グループ（1～ $M$ 箇所あり、 $M$ は 33）、横流入量の倍率調整による任意地点・時点の水位変化量を  $\Delta H$  である。

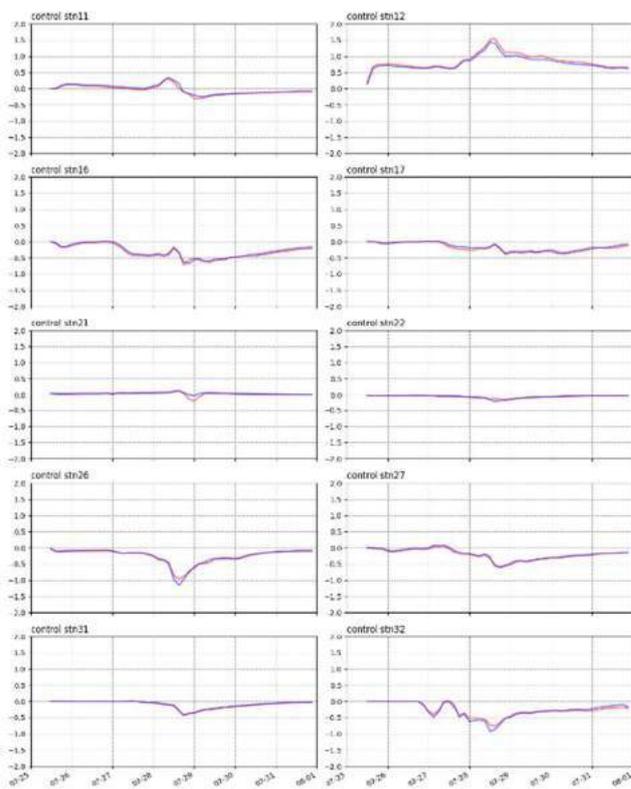
$dH_{i,j}$  は横流入グループ  $j$  の横流入の倍率を 2 倍にした場合の水位観測地点  $i$  の増減水位の時系列データであり、これは横流入グループ 1 つずつ 2 倍に条件設定して予測計算を流して全ての観測地点の水位の変動の差として事前に計測している。

この関係式が一定の精度で成立するということは、1 つの横流入グループの時系列変動だけ把握しておけば、各横流入グループの倍率を任意に変えても、実際に物理

予測モデルを動かしてみるまでもなく、即座に水位が判明するという事となる。

### (3) 横流入地点の流入倍率ランダム変動実験

多重回帰モデルで任意の倍率に調整した場合の河川水位が予測可能かどうかを確認するため、2 期間の洪水期（令和1年10月11～10月15日、令和2年7月25日～8月1日）について、横流入量の倍率をランダムに変えた300 ケースを計算し、多重回帰モデルで再現できるかを確認した。その一例の  $dH$  を図-2 に示した。



— : ランダム実験、— : 多重回帰モデル、単位[m]

図-2 ランダム実験結果に関する多重回帰モデルの  $dH$  再現例

細かい波形も含めて、比較的良好な再現が取れていることが確認でき、一次近似としては十分仮定が成り立つものと考えられる。

なお、波形データは2倍の倍率で測定されているが、1地点のみ0.5倍など小さい倍率にしても、もとの水位に  $dH$  の -0.5 を加算するだけで再現が可能であった。

実際には、本川の水位上昇により、支川は本川の背水影響で水が流れにくくなるといったことが複雑に起こると考えられるが、その影響は水位としてはそれほど大きくないようである。

## 3. 深層学習モデルによる組合せ最適解の探索

### (1) 最適化への深層学習モデルの利用

前節にて、一次近似としては水位のデータ同化の問題

は各地点の水位の差異から各横流入地点の倍率を単純な線形結合として推定するより簡単な問題として置き換え可能であることを示した。ここでは、深層学習モデルでその倍率ベクトルの組み合わせ最適解を推定することを試みた。

なお、解析解として逆行列を計算する方法も考えられるなど、最適化を解くのに深層学習モデルを用いる必然性は低いように感じられるが、①33 地点の時系列データの誤差を扱うため次元が高いこと、②実際の観測データには観測誤差や異常値があり、観測と物理モデルの間にはバイアスとしての誤差も含まれている可能性（どんなに横流入倍率を調整しても到底再現できない水位が含まれる等）などがあるため、損失関数をより自由に表現しやすい生成モデルベースの深層学習モデルを検討した。

ノイズを入力し、全結合層（ノード数 64、ReLU 活性化）を7層連結し、readout として33ノードの平均・分散・バイアスの3種類を出力する Neural Networks を構築した。

本生成モデルで生成した33地点ごとの平均・分散 ( $\theta$ ) を用いて、多重回帰モデル (2式) を計算して、バイアス除去済みの観測水位との尤度を計算し、その最大化問題として倍率 ( $\theta$ ) を最適化した。

損失関数を尤度の最大化としたのは異常値があった場合に柔軟に対応できる効果を期待した設計である。また、(2式)の計算は微分可能であることが重要であるため、深層学習フレームワーク（ここでは TensorFlow2 系）で記述することで、自動微分され生成モデルに損失の勾配が伝わっている。

生成モデルのネットワークのデザイン（層数やアクティベーションの種類最適化等）については検討できていない。損失関数について、損失空間の特性を分析（物理モデルの表現力を超える入力への対応性等の分析）しながら適切なスムージングをかけているが、紙面の都合上記載を省略する。

### (2) 実際の観測データを用いた最適化の適用可能性

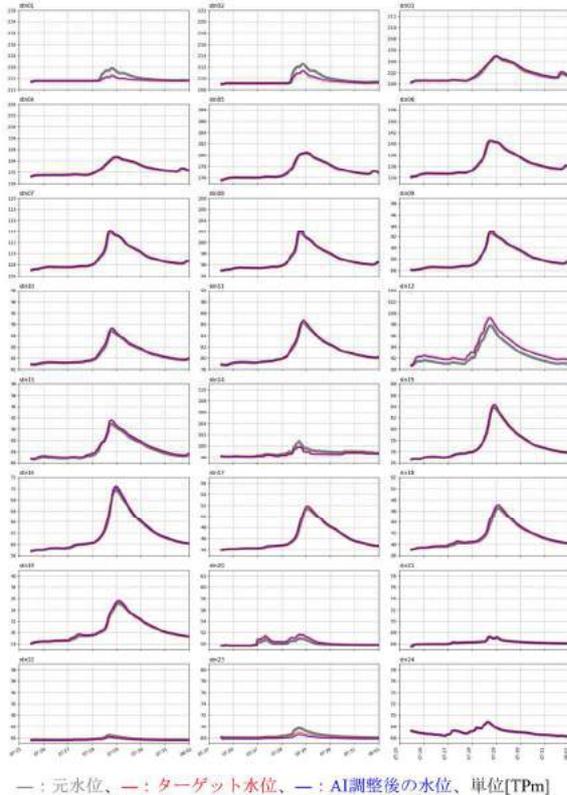
前節のランダム実験では、物理モデルの生成したデータを同じ物理モデルで再現するだけであるので、損失関数の局所解もほとんどない。最適な解も必ず存在し、どれだけ横流入量倍率を調整しても最適化できないデータの入力はないため、最適化問題としては簡単である。結果の一部を図-3(a)に示した。

ここでは、本手法が実測データでも適用できるかを確認し、結果の一例を図-3(b)及び図-3(c)に示した。なお、物理モデルの表現の範囲を示すために前節のランダム実験の結果も実線で同時に示しており、全体の最適化を重視した場合と特定の地点を重視した最適化の結果をそれぞれ示している。

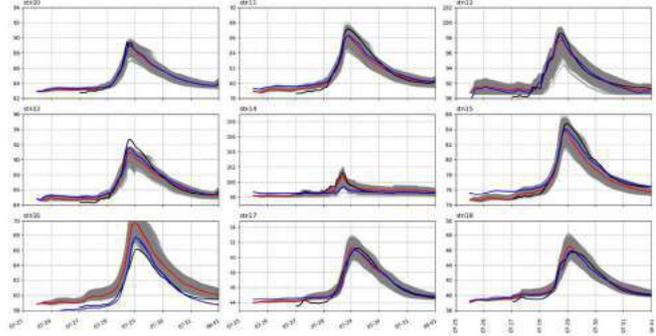
最適化は 1024 粒子で倍率の組合せを 1 万ステップ（数秒程度）探索させ、物理モデルで再計算した。

タや簡易な統計モデル、PINN<sup>9</sup>などに置き換えるなどで更なる高速化が実現できる可能性がある。引き続き検討

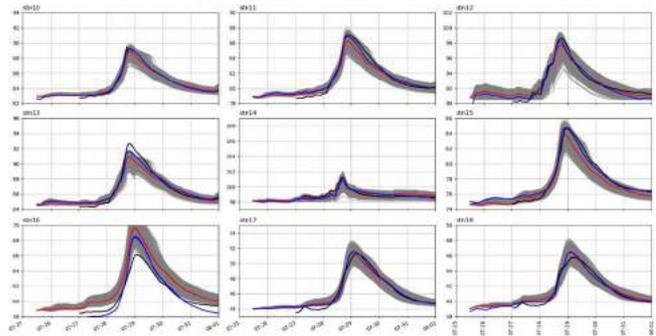
(a) ランダム実験の結果



(b) 実測データの最適化 (全体のバランス重視)



(c) 実測データの最適化 (Stn16重視)



— : 元水位、— : ターゲット水位、— : AI調整後の水位、単位[TPm]

— : 実測、— : ランダム実験、— : 横流入倍率調整前の物理予測、— : 横流入倍率調整後の物理予測、単位[TPm]

図3 本提案手法による 33 地点同時の水位同化の実施例 (令和 2 年 7 月を例に)

実測は物理モデルでは調整できない場所に値が来てしまうことがあり、不運にもそういった値ほど数値が大きい場合も多い。単純な L2 損失などで最適化してしまうと、見かけの損失を小さくすることで、全体の整合性を犠牲にして（低い水位の地点が全体的に水位上昇してしまうなど）、全体としてより不整合が大きくなる方向に最適化されてしまうことがある。

どんな場合でも安定した最適化を実現するのは依然として困難であるが、モデルの調整可能な自由度を適切に調整などしていくことで、データ同化の前処理的な位置づけで活用できないか今後検討していく予定である。

#### 4. おわりに

実際の予測システムには組み込めていないが、水位同化の効率化を検討し、いくつかの課題は残るが一定の高速化を試行できた。

今後はより安定化しつつ、 $dH$  の把握についても今回は物理モデルを 1 回計算しているが、これをエミュレー

していきたい。

#### 5. 参考文献

- 1) 一言正之, 桜庭雅明, 深層ニューラルネットワークと分布型モデルを組み合わせたハイブリッド河川水位予測手法, 土木学会論文集 B1 (水工学), Vol. 73, No. 1, 22-23, 2017
- 2) Goodfellow, Ian J., Jonathon Shlens, and Christian Szegedy. "Explaining and harnessing adversarial examples." arXiv preprint arXiv:1412.6572 (2014)
- 3) 淡路敏之・蒲地政文・池田元美・石川洋一 編著, データ同化 観測・実験とモデルを融合するイノベーション, 京都大学学術出版会, 2009
- 4) Crisan, D. and Doucet, A., "A survey of convergence results on particle filtering methods for practitioners", IEEE Transactions on Signal Processing 50(2002)736-746
- 5) M.Raissi, P.Perdikaris, G.E.Karniadakis, "Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations", Journal of Computational Physics, Volume 378, 1 February 2019, Pages 686-707