

# 水位及び降雨の観測データを用いた多変量 自己回帰モデルによる水位予測の精度検証

ACCURACY VERIFICATION OF WATER LEVEL PREDICTION BY MAR  
MODEL USING OBSERVATION DATA OF WATER LEVEL AND RAINFALL

小山 直紀<sup>1</sup>・山田正<sup>2</sup>  
Naoki KOYAMA and Tadashi YAMADA

<sup>1</sup>学生会員 中央大学 理工学研究科都市人間環境学専攻 (〒112-8551 東京都文京区春日1-13-27)

<sup>2</sup>フェロー会員 中央大学教授 理工学部都市環境学科 (同上)

The purpose of this study is to verify the accuracy of water level prediction in flood disasters using a multivariate autoregressive model, which is one type of time series analysis. The forecasting information on the water level is the most important information for residents to evacuate. The major problem is to decide how long the lead time should be. Therefore, flood prediction was performed using a multivariate autoregressive model with water level and rainfall data during heavy rains. In the present study, we studied two cases, in the first case, when calculating using only the water level, it was possible to predict the time of concentration at observation station of used observed water level data and the reference station. The second case, when calculating using the water level and rainfall, it can predict with high accuracy several hours ahead compared to using only water level.

**Key Words :** *water level , multivariate auto regressive model, water level prediction in flood*

## 1. はじめに

近年, 毎年のように洪水災害が発生し, 多大な被害が発生している. このような災害対策は, 一般的にハード対策とソフト対策の両面から講じられているが, 気候変動による豪雨の激甚化, 高頻度化が懸念されている. そのため, 従来では, 「水害は施設整備により発生を防止するもの」と考えられてきたが, 今後は, 「施設能力には限界があり, 防ぎきれない大洪水は発生するもの」と社会全体の意識変化の必要がある<sup>1</sup>とされている. そのため, ハード対策では防ぐことができない程の大規模な水害時にはソフト対策で補う必要があり, 今後ソフト対策の重要性がより一層高まっている.

2019年の令和元年台風第19号では, 71河川140箇所堤防が決壊し, 多数の死者を出す災害となった. 牛山<sup>2</sup>によると, この災害の犠牲者となった原因外力のうち, 約50%は河川の氾濫水に起因する犠牲者であるとされており, また, 60年代以上の犠牲者が多いことが報告されている. 河川の氾濫は, 河道内水位の上昇により越水または堤防決壊によって起こるため, 氾濫が発生する前に

避難を行うことができれば災害から逃れることが可能である. また, 現在の避難情報の発令には河川水位が用いられているため, 数時間後の水位が精度良く求めることができれば, 洪水時の急迫した状況の中, 適切な避難判断を行うことができる. そのため, ソフト対策の中でも水位の予測情報は, 住民の命を守るためにも非常に重要な情報である. また, この水位の予測情報をできる限り高精度に先の時間まで求めることができれば, 避難に要するリードタイムを伸ばすことができる.

一般的に水位予測手法は, 流出解析モデルと統計モデルの2つに大別される. 流出解析モデルの中でも物理モデルは, 降雨流出のプロセスを詳細に表現することができる. しかし, 降雨データをインプットデータとするため, 降雨の精度が重要である. 日本においては, 降雨観測レーダによる観測網が十分に整備されており, 日本全国で時空間的に高精度な降雨データを得ることができる. しかし, 新興国の場合, 高精度なデータが得られない, または, データが得られないこともある. さらに, 流出解析モデルは, 解析結果が流出量であるため, 水位に変換する際に, HQ曲線による誤差が含まれることもある. 一方, 統計モデルは, 機械学習を用いた多くの研究<sup>3</sup>

が進められている。この手法は、流出現象の非線形性を考慮できるという利点があるが、多くのデータが存在する場合に精度の高いモデルが構築できると一般的に言われている。それに対し、ARモデルやARMAモデルを代表とする時系列解析は、線形モデルのため、非線形性を表現できないという欠点もあるが、モデル構造が単純なため、データ量が少ない際においても精度が出る人が多いといわれている。そこで、本研究では、水文データが乏しい流域においても適用できる水位予測を目的として、統計モデルの1つである時系列解析の多変量自己回帰モデルを用い、水位及び降雨データを用いた水位予測精度の検証を行った。

## 2. 解析手法

### (1) 水位のみによる多変量自己回帰モデルの水位予測

時系列解析は統計モデルの1つであり、現象を過去のデータと関連付け、現象の予測や制御のために用いられている。まず、最も基本的な時系列解析に用いられている自己回帰モデル式を(1)式に示す。

$$h_n = \sum_{i=1}^N a_i h_{n-i} + \varepsilon_n \quad (1)$$

(1)式は水位時系列 $h_n$ を過去の水位時系列 $h_{n-i}$ にパラメータ $a_i$ を掛けたものとホワイトノイズ $\varepsilon_i$ の線形和で表したものである。 $N$ は次数であり、考慮する過去のデータである。ここでパラメータ $a_i$ について考えてみると、過去の観測値 $h_{n-i}$ の重みであることから、 $a_i$ の大きさによって、 $h_n$ に対する過去の観測値の影響度を表している。

しかし、河川の流量は様々な支川より上流から下流に集っており、それぞれの地点における流量は相互に関連していると考えることができる。そこで、自己回帰モデルを拡張し、それぞれの支川の影響を含めたものを表現した式が(2)式である。

$$\begin{bmatrix} h_n^1 \\ h_n^2 \\ \vdots \\ h_n^p \end{bmatrix} = \sum_{i=1}^N \begin{bmatrix} a_{11}^i & a_{12}^i & \cdots & a_{1p}^i \\ a_{21}^i & a_{22}^i & \cdots & a_{2p}^i \\ \vdots & \vdots & \ddots & \vdots \\ a_{p1}^i & a_{p2}^i & \cdots & a_{pp}^i \end{bmatrix} \begin{bmatrix} h_{n-i}^1 \\ h_{n-i}^2 \\ \vdots \\ h_{n-i}^p \end{bmatrix} + \begin{bmatrix} \varepsilon_n^1 \\ \varepsilon_n^2 \\ \vdots \\ \varepsilon_n^p \end{bmatrix} \quad (2)$$

(2)式はベクトル形式で表されたものだが、テンソル形式で表したものが(3)式である。

$$\mathbf{h}_n^p = \sum_{i=1}^N \sum_{l=1}^p \mathbf{a}_{pl}^i \mathbf{h}_{n-i}^l + \varepsilon_n^p \quad (3)$$

(3)式は(1)式をより一般化したものであり、多変量自己回帰モデルやベクトル自己回帰モデルと呼ばれているものである。ここに、 $N$ : モデルの次数、 $P$ : 用いる水位の数、 $\mathbf{h}_n^p$ :  $n$ 時刻における $p$ 地点水位、 $\mathbf{a}_{pl}^i$ :  $i$ 時刻前における $p$ 地点水位に対する $l$ 地点水位の影響。また、パラメータの推定は、最小二乗法によって行った。多変量自己回帰モデルは、多くの分野で用いられており、工学分野では赤池・中川<sup>4)</sup>が、セメントキルンプロセスの制御

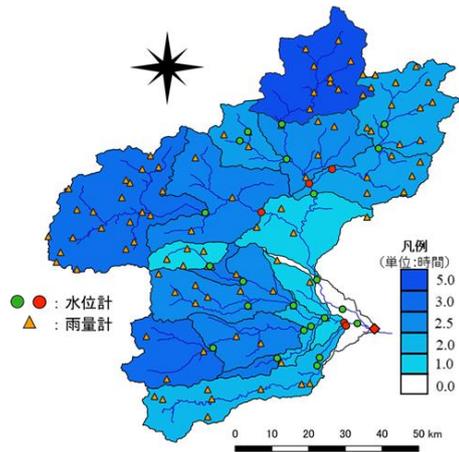


図-1 利根川上流域

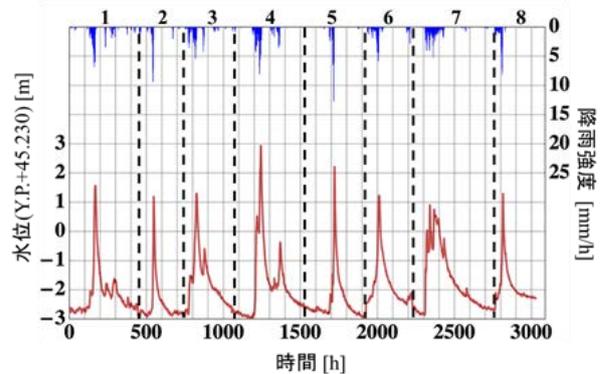


図-2 2002年から2018年までの八斗島地点における水防団待機水位を超えた洪水(全8イベント)

において先駆的に用いている。

### (2) 水位と降雨による多変量自己回帰モデルの水位予測

降雨流出プロセスを考えると、流域に雨が降り、それが土壌に浸透・流出し、川に流れ、ある地点に集まり、ある水位となる。ある観測地点の水位情報は、降雨流出プロセスの中では、最後の情報であり、一番の始まりは、降雨情報である。そこで、水位情報のみを使用するだけではなく降雨情報を追加することにより、水位情報より前の情報をモデルに加えることができ、より長期の水位予測を行えることが考えられる。そこで、(2)式に降雨を追加した式が(4)式である。

$$\begin{bmatrix} h_n^1 \\ \vdots \\ h_n^p \\ r_n^1 \\ \vdots \\ r_n^s \end{bmatrix} = \sum_{i=1}^N \begin{bmatrix} a_{11}^i & \cdots & a_{1p}^i & b_{1s}^i & \cdots & b_{1s}^i \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ a_{p1}^i & \cdots & a_{pp}^i & b_{ps}^i & \cdots & b_{ps}^i \end{bmatrix} \begin{bmatrix} h_{n-i}^1 \\ \vdots \\ h_{n-i}^p \\ r_{n-i}^1 \\ \vdots \\ r_{n-i}^s \end{bmatrix} + \begin{bmatrix} \varepsilon_{qn}^1 \\ \vdots \\ \varepsilon_{qn}^p \\ \varepsilon_{rn}^1 \\ \vdots \\ \varepsilon_{rn}^s \end{bmatrix} \quad (4)$$

$r$ : 降雨データ、 $s$ : 降雨のデータ数である。この式より、ある地点の水位を求めるために水位と降雨を用いていることがわかる。

### (3) 予測式の導出

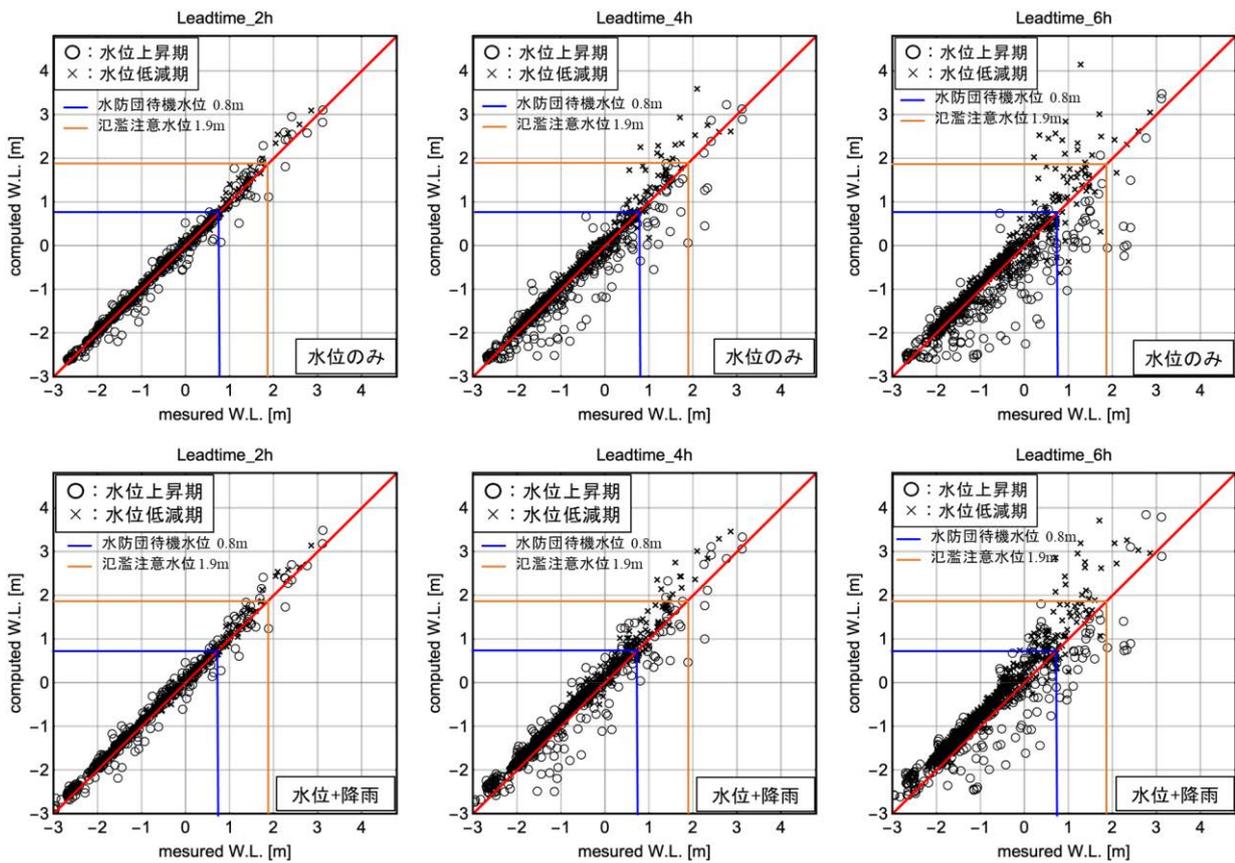


図-3 2, 4, 6時間後の予測値と観測値の比較  
(上図：水位のみを用いた場合，下図：水位と降雨を用いた場合)

(2)式及び(4)式は，現在の水位を過去データから求めている．そのため，この式に現在データを含めることで1つ先の予測を行うことができ，(5)式で表せる．

$$\hat{h}_{n+1}^p = \sum_{i=1}^N \sum_{l=1}^P a_{pl}^i h_{n+1-i}^l + \varepsilon_n^p \quad (5)$$

ここに， $\hat{h}$ ：予測水位を表す．その予測した結果を再帰的に次ステップの予測に用い， $x$ 時間後までの予測を行うとすると(6)式で表すことができる．

$$\hat{h}_{n+x}^p = \sum_{j=1}^{x-1} \sum_{l=1}^P a_{pl}^j \hat{h}_{n+x-j}^l + \sum_{i=1}^{N-x} \sum_{l=1}^P a_{pl}^i h_{n+1-i}^l + \varepsilon_n^p \quad (6)$$

同様の考え方で(4)式も水位予測を行った．また，洪水時の避難情報に役立てるため，8時間後まで予測を行った．

### 3. 対象流域

対象流域は，利根川の基準点である八斗島地点を懸案地点とした利根川上流域である．流域面積は5100km<sup>2</sup>であり，河口から181.45kmの地点に基準点である八斗島がある．図-1は，流域図である．それぞれの丸は水位観測所を表しており，全21箇所である．また，三角は地上雨量計を表しており，約100箇所ある．また，それぞれの水位及び観測データは1時間ごとのデータである．それぞれの観測所ごとに流域を分割しサブ流域に分け，それぞれの色は，過去の洪水イベントデータから求めた八斗島地点とそれぞれの観測地点のピーク水位の時間差を表

しており，濃青色ほどピーク水位と時間差があることを示している．また，それぞれのサブ流域における色分けは，次章に説明するパラメータの推定に用いた洪水イベント(全8イベント)ごとの平均値である．

### 4. 水位予測の精度検証

#### (1) モデルの構築と精度検証の方法について

本研究で使用する(2)，(4)式のモデルを構築する上で，使用する観測所数 $P$ の選択，次数 $N$ を決定させる必要がある．著者らの既往研究<sup>5)</sup>により，対象流域における再現精度の検証を行っており，数点以上の観測値があれば十分再現できることがわかっている．その結果をもとに図-1の赤点の6点を $P$ として使用する．また，(4)式の降雨データに関しては，使用するそれぞれの水位観測所より上流域にある地上雨量観測所のデータをティーセン分割により流域平均降雨とし，水位観測所数 $P$ と同じ6つを入力データとした．

次に，次数 $N$ は，Akaike<sup>6)</sup>によって，考案された赤池情報量基準，AIC(Akaike Information Criterion)を用いた．AICは(7)式で表される．

$$AIC = -2 \log(L) + 2K \quad (7)$$

ここで， $L$ ：最大対数尤度， $K$ ：モデルのパラメータ数である．既往の研究<sup>5)</sup>より八斗島地点における次数と

AICの関係より、半日である12として今後の計算を行う。

パラメータ推定に用いたデータは、2002年から2018年における出水期である5月から11月までの水位を抽出し、その中から八斗島地点における水防団待機水位を超えたものを1つの洪水イベントとした。その期間中に水防団待機水位を超えた洪水イベントは8回であった。図-2は、その8洪水イベントを並べたものである。図中上の数字はそれぞれの洪水イベントを表しており、かっこ内の数字は、水位の立ち上がりからピーク水位に至るまでの時間を表している。予測精度を評価するため、限られたデータを全て用いる分割交差検定で行った。分割交差検定の手順は以下の通りである。1. 対象洪水イベントを図-2のように8個のデータセットに分割する。2. 1つのデータセットを検証データとする。3. 残りの7つのデータをパラメータ推定用のデータセットとする。4. 設定したデータでパラメータ推定・精度検証を行う。これら2., 3., 4.を繰り返し、予測精度の評価を行った。

### (2) 観測値と予測水位の精度検証

図-3は、水位データのみを用いた場合(上図)と水位及び降雨データを用いた場合(下図)に、観測値と予測値を比較した図である。また、洪水イベントは全8イベントあり、それぞれのイベントにおける観測値と予測値を全て重ねたものである。代表として、2, 4, 6時間後の予測を載せた。2時間後の予測値と観測値を比較すると、どちらの方法においても十分に予測できることがわかる。しかし、予測時間が長期になる4, 6時間後の予測値では、どちらの方法においても、観測値からとらばつきが大きくなっていることわかる。また、水位の立ち上がりにおいて、ばらつきが大きくなり、特に水位が低い場合には、観測値と比較して、過小傾向があることがわかる。低減部に関して、水位が低い場合に十分な精度で予測できていることがわかる。しかし、水位が1mを超えたところでは、ばらつきが大きくなっており、観測値と比較して過大傾向にあることがわかる。このことから、本手法の特徴として、水位が-3mから-1mの水位の立ち上がり及び1mから3mのピーク水位から低減期にかけて精度が低いことがわかる。これは、過去のデータから予測をしているため、水位の立ち上がりや水位ピークからの低減期などの変曲点付近ではそれらの情報がないため、精度が落ちることが考えられる。水位のみの場合と降雨を用いた場合の比較より、全体の傾向は同じだが、4, 6時間後では、水位のみと比較して、水位の立ち上がり、ピーク水位から低減期にかけてばらつきが小さくなっている傾向があることがわかる。これは、水位より前の降雨情報を追加したことにより、精度の向上を図ることができたと考えられる。

### (3) ハイドログラフの精度検証

次にハイドログラフの予測精度の評価を行う。避難を

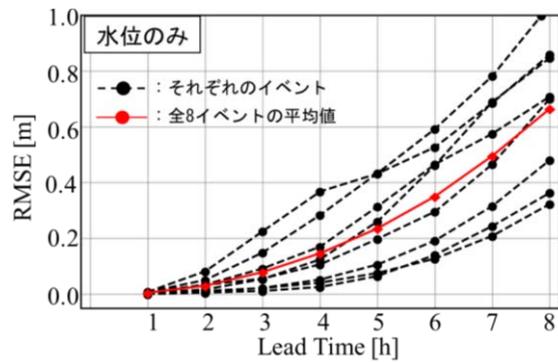


図-4 水位のみを用いた場合の1時間～8時間後のRMSEによる予測精度

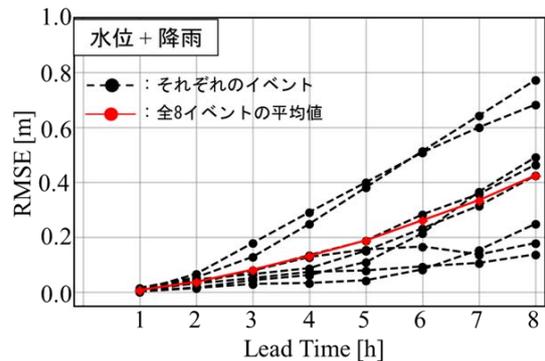


図-5 水位と降雨を用いた場合の1時間～8時間後のRMSEによる予測精度

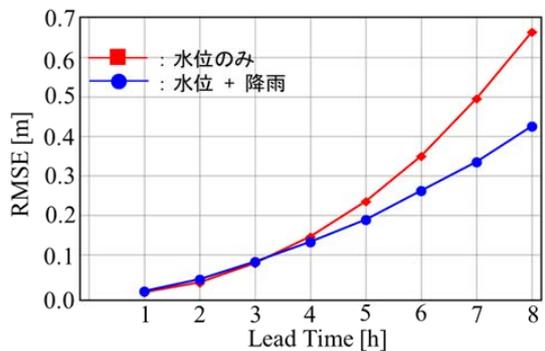


図-6 水位のみの場合と水位と降雨を用いた場合のRMSEによる平均値の比較

行うときに重要な時間であるハイドログラフにおける水位の立ち上がりからピーク水位までを用いて評価を行っ

$$RMSE = \sqrt{\frac{\sum_{n=1}^{N_{peak}} (h_{ob}^i - h_{pre}^i)^2}{N_{peak}}} \quad (8)$$

た。(8)式に評価式を示す。

ここに、 $N_{peak}$ ：水位の立ち上がりからピーク水位までの数、 $h_{ob}^i$ ：観測値、 $h_{pre}^i$ ：予測値である。図-4は、水位のみを用いた場合のRMSEによる予測精度の評価を行った結果である。それぞれのイベントのRMSEである黒破線を見ると、ほとんどのイベントにおいて、4時間を超えたあたりから傾きが変わり、5時間後から傾きが大きくなっている。それぞれのイベントによって、傾きの傾向が異なるのは、水位の立ち上がりからピーク水位までの

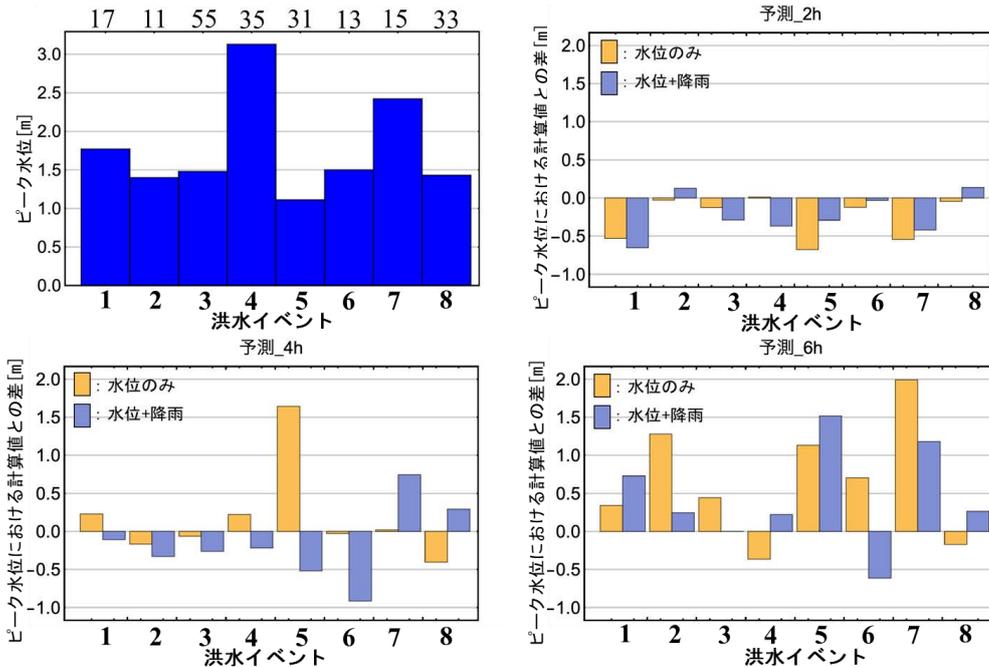


図-7 それぞれのイベントにおけるピーク水位と予測値の比較 (左上：それぞれのイベントのピーク水位(図上に書いてある数字は、水位の立ち上がりからピークまでの時間)、右上、左下、右下：それぞれ、2、4、6時間後の予測値とピーク水位との差)

時間が大きく関係しており、この時間が長いイベントでは、精度の高い予測ができることが分かった。また、それぞれのイベントの平均値である赤線を見ると、5時間後から傾きが大きくなっており、これは、図-1に示す対象流域の洪水到達時間と同等であることから、水位のみを用いて予測を行う場合、高精度に予測を行うことができる時間は、使用する観測地点と基準点の水位の時間遅れ程度であることがわかる。

次に、図-5は水位と降雨を用いた場合のRMSEによる予測精度の評価を行ったものである。水位のみを用いた場合と異なり、多くのイベントにおいて予測時間が5時間を超えても傾きが大きく変化しないことがわかる。

図-6は、水位のみを用いた場合と水位と降雨を用いた場合の全イベントの平均値を比較したものである。この図より、水位のみの場合でも4時間後までの予測値において、精度が変わらないことがわかる。しかし、5時間以降では、降雨情報を追加すると精度が向上することがわかる。これは、流出プロセスの中で最後の情報である水位より前の情報である降雨データを追加したことが精度向上につながったことが考えられる。

#### (4) ピーク水位の精度検証

現在の避難情報の発令基準は河川水位が用いられていることがほとんどである。そのため、河川のピーク水位がどの程度になるかを精度良く予測できるかは、避難判断をする際に重要である。そこで、本手法におけるピーク水位における精度評価を行った。

図-7は、水位のみを用いた場合と水位と降雨を用いた場合のピーク水位と予測値の比較である。代表として、

2、4、6時間後の予測を載せている。どの予測時間に対しても、水位と降雨を使った場合でも水位のみの場合と大きく変わらないか、それより精度の良い予測ができることが多いことがわかる。2時間後の予測では、どちらの場合でも最大で約50cm以内に予測ができていたことがわかる。4時間後の予測値では、イベント5において、水位のみの場合は、予測値が観測値と比較して、大きな値を示していることがわかる。これは、図-2のイベントを見るとわかるが、ハイドログラフが、ピーク付近で、上昇と下降を繰り返しており、複雑な形をしているため、うまく予測ができなかったことが考えられる。しかし、水位と降雨を用いた場合には、60cm程度低めに予測しているが、精度が向上していることがわかる。6時間後の予測では、どちらの手法においても、1m程度大きく予測していることがわかる。立ち上がりからピーク付近までのハイドログラフをRMSEにより評価したときは、水位のみと比較し、水位と降雨を用いた方が精度の良いことがわかった。しかし、ピーク値にのみ注目すると、水位と降雨を用いた場合のほうが精度の良い場合が多いものの、水位のみと大きく値が変わらないことがわかった。この理由として、降雨と水位の関係には時間遅れがあり、ピーク水位の時には、降雨の影響が小さいため、降雨のデータを用いても大きな差が出なかったことが考えられる。また、本手法の特徴として、水位のみと水位と降雨を用いた場合においても6時間以降の予測は、水位を大きく予測する傾向があり、今後、さらなる精度向上を図っていくべきではあるが、避難等にとっては安全側に予測していることがわかった。

### (5) 予測降雨を用いた場合の精度検証

水位と降雨を用いた場合の式(4)を見ると、降雨のデータは、多変量自己回帰モデルにより過去のデータから予測を行っており、このプロセスの中に、気象学的観点は考慮されていない。近年では、レーダ観測網が整備されたこともあり、降雨予測に関する研究が精力的に行われている。そこで、予測降雨を用いて、水位予測をする式を構築し、その精度評価を行った。予測降雨を用いた水位予測式が、以下に示す(9)式である。

$$\begin{bmatrix} h_n^1 \\ h_n^2 \\ \vdots \\ h_n^p \end{bmatrix} = \sum_{i=1}^N \begin{bmatrix} a_{i1}^i & a_{i2}^i & \cdots & a_{ip}^i \\ a_{21}^i & a_{22}^i & & a_{i2}^i \\ \vdots & & \ddots & \vdots \\ a_{p1}^i & a_{p2}^i & \cdots & a_{pp}^i \end{bmatrix} \begin{bmatrix} h_{n-i}^1 \\ h_{n-i}^2 \\ \vdots \\ h_{n-i}^p \end{bmatrix} + \sum_{j=1}^R \begin{bmatrix} C_{11}^i & C_{12}^i & \cdots & C_{1L}^i \\ C_{21}^i & C_{22}^i & & C_{i2}^i \\ \vdots & & \ddots & \vdots \\ C_{L1}^i & C_{L2}^i & \cdots & C_{LL}^i \end{bmatrix} \begin{bmatrix} r_{n-i}^1 \\ r_{n-i}^2 \\ \vdots \\ r_{n-i}^L \end{bmatrix} + \begin{bmatrix} \varepsilon_n^1 \\ \varepsilon_n^2 \\ \vdots \\ \varepsilon_n^p \end{bmatrix} \quad (9)$$

(9)式は、外部入力付き多変量自己回帰モデル(MARX : Multivariate Auto Regressive Model with eXogenous)と呼ばれているものである。この式より、予測水位を求めるとき、水位の自己回帰モデルによる予測項と降雨の外力項が分離しているため、この外力項に予測降雨を入力し、予測を行った。また、予測降雨には、実測降雨を仮想の予測降雨として扱った。パラメータについては、最小二乗法によって求めた。

図-8は、降雨を自己回帰モデルで求め、式(4)により予測した場合と予測降雨を使用し(9)式による予測をRMSEによる予測精度の評価をしたものである。図より、水位予測をした場合でも、予測精度の評価は変わらないことがわかる。このことより、予測雨量データがない地域においても計算時点での降雨データがあれば、予測降雨を用いた場合と同様の精度を持った降雨予測が行えることがわかった。

## 5. まとめ

本研究では、時系列解析の手法である多変量自己回帰モデルを水位予測に適用させ、その精度検証を行った。以下にそのまとめを示す。

- (1) 利根川上流域を対象にして、多変量自己回帰モデルによる水位予測を行った。水位のみを用いた水位予測を行った場合、おおよそ3-4時間後までの水位は、30cm以内に予測が行えることがわかった。このことから、水位のみを用いた場合、高精度な予測を行うことができる時間は、モデルを構築する際に使用する水位観測所と懸案地点における水位の遅れ時間と同等程度であることがわかった。
- (2) 水位と降雨データを用いて予測を行った場合、3時間程度先までは、水位のみと比較し、精度の違いはほとんど見られない。しかし、より長時間を予測する

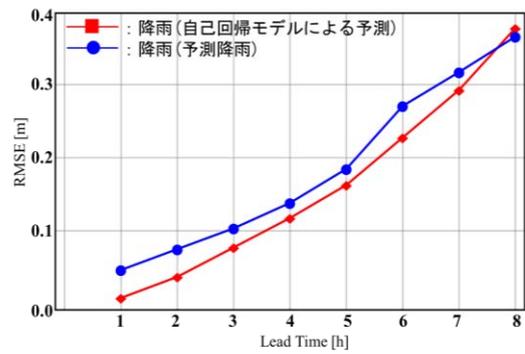


図-8 降雨を自己回帰モデルによる予測を用いた場合と予測降雨を用いた場合のRMSEの比較

場合は、降雨データを含めることでより長期の予測ができることがわかり、長時間の予測を行うほど、その差は顕著となる。

- (3) 水位のみによる場合と水位と降雨を用いた場合において、洪水のピーク水位の精度評価を行ったところ、多くの場合、降雨を用いたほうが、精度が高いことが多い。また、本手法における6時間後の予測では、予測水位と観測値を比較し、大きく予測することから、安全側に予測する傾向があることがわかった。
- (4) 降雨データを用いた多変量自己回帰モデルによる水位予測と予測降雨を用いた場合による外部入力付き多変量自己回帰モデルの水位予測の結果は大きく変わらないことがわかった。そのため、計算時点での降雨データがあれば、予測降雨データがない流域においても同等の精度を持った水位予測が行える。

本論文では、対象流域を利根川上流域における1つの流域にとどまっている。今後は、小規模な流域等の様々な流域において、同様の計算を行い、本手法による適用性について知見を増やしていく。

## 参考文献

- 1) 社会資本整備審議会：大規模氾濫に対する減災のための治水対策のあり方について ～社会意識の変革による「水防災意識社会」の再構築に向けて～、2015。
- 2) 牛山素行：2019台風19号等による人的被害についての調査(10/25大雨を含む速報 2019年12月30日版)、<http://disaster-i.coco-log-nifty.com/blog/files/20191230report.pdf>、(2020.3 参照)
- 3) 例えば、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, 187-192, 2016。
- 4) 赤池弘次、中川東一郎：ダイナミックシステムの統計的解析と制御、サイエンス社、1972
- 5) 小山直紀、山田正：観測流量データを用いた多変量自己回帰モデルによるハイドログラフの再現精度の検証、水文・水資源学会2019年度研究発表会、pp.166-167, 2019。
- 6) Hirotugu Akaike : A New Look at the Statistical Model Identification, IEEE Trans. Automat. Contrl., NO.6, pp.716-723, 1974。

(2020.4.2受付)