

観測流量データを用いた多変量自己回帰モデル の適用による河川流量の逆推定に関する研究

A STUDY ON INVERSE ESTIMATION OF RIVER DISCHARGE USING
MULTIVARIATE AUTO REGRESSIVE MODEL WITH OBSERVED DISCHARGE

小山 直紀¹・山田正²
Naoki KOYAMA and Tadashi YAMADA

¹学生会員 修(工) 中央大学理工学研究科 (〒112-8551 東京都文京区春日1-13-27)

²フェロー会員 工博 中央大学教授 理工学部都市環境学科 (同上)

Generally, runoff analysis is using models based on physical processes. Those models are calculated using rainfall as input data. However, in developing countries, there are not enough rainfall observation stations in a basin. Therefore, this study evaluated the reproducibility and prediction of hydrograph using multivariate autoregressive model with only discharge assuming a basin that poor hydrological data. This method is able to reproduce hydrograph, even if there are just several observation station in the basin. By implementing parameters which is estimated by the historical flood events, this method can be applicable to predict future flood events.

Key Words : Observed Discharge, Multivariate Auto Regressive Model, Inverse Estimation

1. はじめに

一般に降雨流出解析は物理過程に基づいたモデルを用い、流域の降雨をインプットデータとし、河川流量や水位の計算をすることがほとんどである。そのため、降雨データの精度が計算結果に影響を及ぼす。一般的に降雨には、地上雨量計や気象レーダによって観測されたものが用いられているが、新興国のような降雨観測の整備が不十分な流域では地上雨量計や気象レーダの他にGsMap (Global Satellite Mapping of Precipitation)のような人工衛星による観測雨量が用いられている場合もある。しかし、GsMapによる降雨は過小評価される傾向があり¹⁾、水文データのほとんどない地域では非常に有効なものであるが、依然として地上雨量計が降雨データを得るための代表的な観測手段である。既往の研究より流域平均雨量を算出する際に使用する雨量計のデータ数と洪水ピーク流量の関係を青木ら²⁾は示しており、洪水ピーク流量を10%以内にするためには雨量計一つ当たりの支配面積(流域面積/雨量計の個数)を10km²に一つの雨量計データが必要という結論を示している。また、雨量計一つ当たりの支配面積が100km²の場合には、5%~50%程度の洪水

ピーク流量に幅があることを示している。そこで、水文データの乏しい新興国と十分に整備されている日本においてどの程度降雨観測の整備に違いがあるのかを簡単に述べる。まず、図-1に日本が管理する(気象庁アメダス、国土交通省)地上雨量計の観測網を載せる。日本の本土面積(約362,215km²)をそれぞれの地上雨量計の合計(約4100個)で割ると、地上雨量計1台が占める雨量計の支配面積は、約88km²(約9km×9km)に1台の地上雨量計がある。また、図-2に新興国のベトナムが管理する地上雨量計による観測網を示す。ベトナムの本土面積(約330,215km²)を地上雨量計の個数(755台)で割ると、約437km²(20km×20km)に1台の地上雨量計があることになり、日本と比較すると、約5倍の違いがある。このように日本とベトナムでは地上雨量計の観測密度が大きく異なっており、流域によって地上雨量計の数は異なるが、そこから算出する流域平均雨量には大きな不確実性があることが予測される。

新興国のような水文データの乏しい地域において洪水災害は毎年のように起きており、大きな経済・人的被害が発生している。洪水被害を防ぐために、降雨データの不十分な流域においても重要な地点における水位・流量

を予測し、水位情報等を住民に提供することが必要である。近年では、ニューラルネットワークを用いた洪水予測が多くされており³⁾、多くのものは雨量と水位を用いて、洪水予測が行われている。水位のみによる洪水予測の既往の研究では、槻山ら⁴⁾が、ニューラルネットワークによる洪水時の水位予測を行っているが、再現性は良好であるものの水位上昇部の適合性が低いという結論を出しており、雨量データを入力値に加えることで改良することにより、予測精度の向上を確認している。

本研究は、降雨データの精度が十分に確保できない流域を想定し、流量データのみから時系列解析の多変量自己回帰モデルを適用し、流域内にある観測地点での流量の再現精度、及び予測を行った。

2. 対象流域

対象流域は、国土交通省の管理する利根川八斗島観測所を下流端とする利根川上流域を対象とした。対象流域は過去の水文データが十分整っているため、精度検証やモデルを構築する上での条件の検討には適しているおり、本流域を対象としている。図-5に利根川上流域における観測所の位置と河川網を記している。流域面積は5100 km²であり、流域内に21地点の流量観測所があり、支川を含め河川の主要な地点で観測されている。また、流量データは、1時間ごとのデータを用いた。

3. 解析手法

時系列解析モデルを用いて、流量ハイドログラフの再現精度の検証を行った。時系列解析は統計モデルの1つであり、現象を過去の動きと関連付け、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。

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

(1)式は流量時系列 q_n を過去の流量時系列 q_{n-i} にパラメータ a_i をかけたものとホワイトノイズ ε_n 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 q_{n-i} の重みであることから、 a_i の大きさによって、 q_n に対してどの程度過去の観測値が影響しているかを表している。

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係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。

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

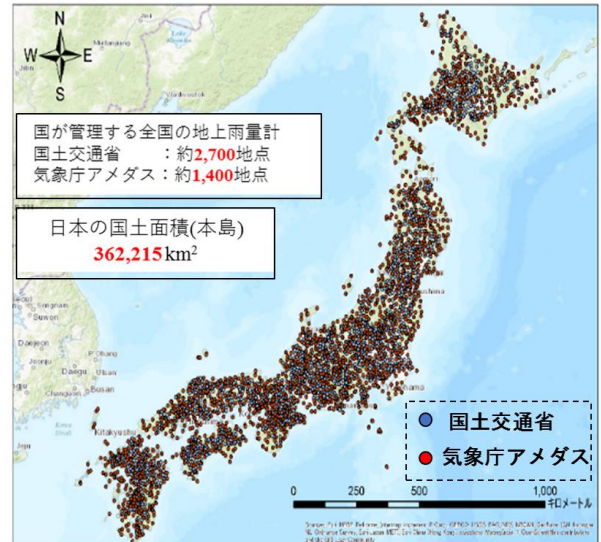


図-1 日本における地上雨量計による観測網

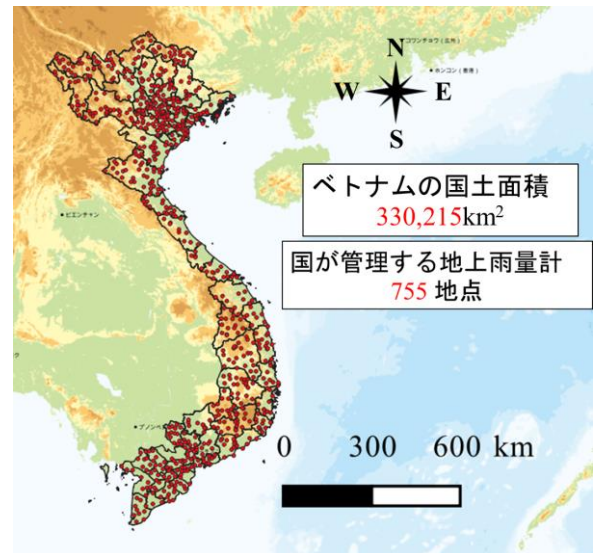


図-2 ベトナムにおける地上雨量計による観測網

(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。

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

(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。ここで、 N :モデルの次数、 P :流量観測所数、 $\mathbf{q}_n^p$: n 時刻における p 地点流量、 $\mathbf{a}_{pl}^i$: i 時刻前における p 地点流量に対する l 地点流量の影響。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、経済学²⁾で用いられており、工学分野では赤池・中川⁴⁾が、セメントキルンプロセスの制御において先駆的に用いている。この(3)式より、ある時刻における流量を求めるときに自身の流量だけでなく、他の流量データを用いて求めることができる。また、それぞれの地点における流量が含まれた式になっているため、上流にある流量から下流の流量を推定するだけではなく、下

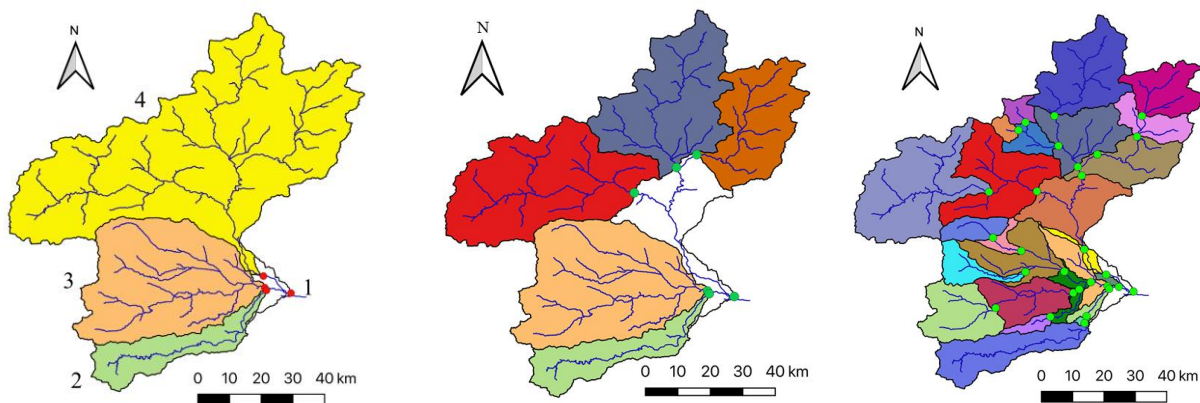


図-5 対象流域における計算に用いた流量観測所数(4地点, 6地点, 21地点)によって分割した流域図

流の流量データを用いて上流の流量を推定できる。

この(3)式のモデルを構築するためには、使用する観測所数 P と過去の観測データをどこまで考慮するかの次数 N を決定しなければならない。そこで、使用した観測所数による再現精度の確認を行った。使用した観測所数は、対象地域である利根川上流域の下流端である八斗島を含む4地点(大きな支川のみを考慮する場合)、6地点(中程度の支川も考慮する場合)、21地点(全ての観測所)の3つのパターンで計算を行った。それぞれ使用した観測所数で流域分割をしたものが図-5である。

次に、3つのパターンにおいて、洪水イベントの再現精度の検証を行った。パラメータの推定に用いた洪水イベントは、2015年9月に起きた関東・東北豪雨のものをを用いた。本イベントは、八斗島地点において水防団待機水位を超えており、ピーク流量は約 $4000\text{m}^3/\text{s}$ である。そのイベントを含む1ヶ月分のデータをパラメータ推定のデータとした。図-6が推定したパラメータを用いて再現計算をした結果である。使用した観測所数のモデルが構築できるが、代表として流域の下流端である八斗島を代表として載せている。また、図-6のそれぞれにおける青線、緑線、赤線は、次数 N を変更した結果である。また、それぞれの観測所における流量を流域面積で割り、流出高に変換し計算を行った。使用した観測所数が4地点の場合は、どの次数においても、傾向をとらえられていることが分かる。しかし、ピーク付近でどの次数においても、多少のばらつきがある。そのばらつきは、ピーク流量において約 $\pm 8\%$ であり、大きな違いはないことから十分に再現できることが分かった。次に使用した観測所数が6地点の場合には、次数 N を変化させても洪水の立ち上がりからピーク付近、また逓減期において実測値との差がほとんどないことが分かり、十分に実測値を再現できていることが分かる。最後に最も多く観測所数を使用した場合には、次数 N が10, 24の場合には、どの洪水期においても、良く推定されていることが分かる。しかし、次数 N が30のときには、実測値の傾向が捕らえられていることには変わりはないが、数値振動している。このこ

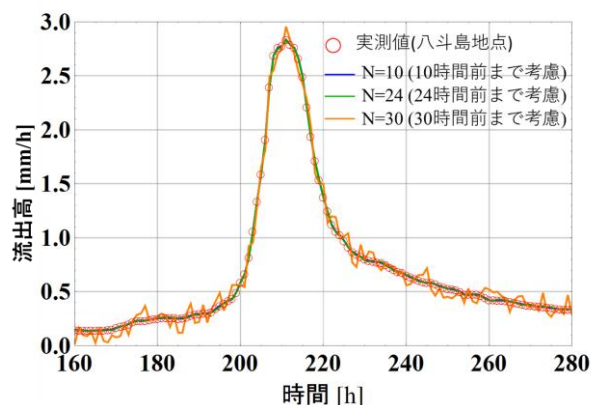
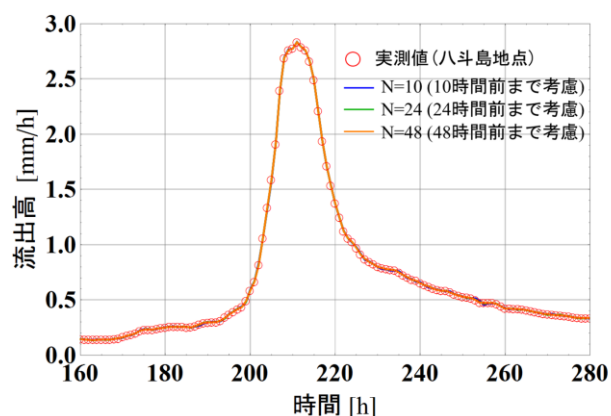
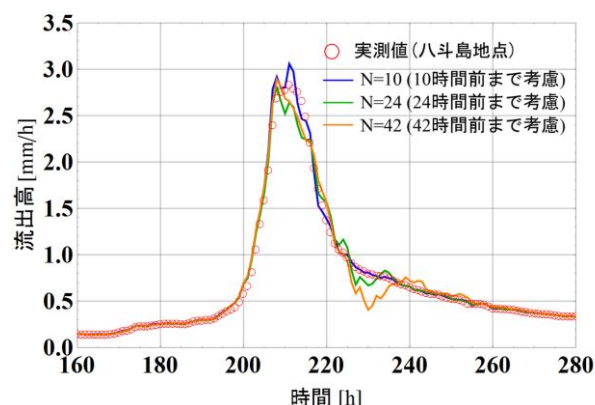


図-6 洪水イベントの再現結果
(使用した観測所数は、上から4, 6, 21地点)

とから、モデルを構築する上で必要な観測所数は、多ければ良いというわけではないことがわかる。観測所数 P が21で次数 N が30の場合は、ある地点でのモデルにおけるパラメータ数は、 $21 \times 30 = 630$ 個のパラメータがあるため、モデルの過剰適合が原因だと考えられる。これらのことから、降雨をインプットとする物理モデルにおいては、降雨の精度が要求されるが、本手法においては、流量のみでも流域内に数点以上の観測データがあれば精度の高い再現を行うことが可能である。

モデルを構築する上では、使用する観測所数 P の選択だけではなく、次数 N も決定させる必要がある。そこで次数の選択方法を説明する。

(3)式における次数 N の選択方法は、赤池⁴⁾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(4)式で表される。

$$AIC = -2\log(L) + 2K \quad (4)$$

ここで、 L ：最大対数尤度、 K ：モデルのパラメータ数である。八斗島地点における次数とAICの関係が図-7である。使用した観測所数が4, 6, 21地点のどれにおいても同様の結果のため、本論分では使用した観測所数が6のものを代表として載せた。この図より、AICの値は次数10より大きくしたとしても、AICの値は大きく変わらないことがわかる。このことより、次数 N の値を10として今後の計算を進めていく。また、次数は過去の値をどの程度まで考慮すべきことを表しているため、10時間前までのデータが八斗島地点の流量に大きな影響を与えていることがわかる。このように次数 N から流域特性である洪水到達時間を推定できることを示している。

以上のことをまとめると(3)式のモデルを構築する上で、利根川上流域程度の流域面積ならば、数点以上の観測データがあれば、十分な再現精度を持つモデルを構築することができることがわかった。また、使用する観測所数が6点を使用する場合には、実測値とほとんど差がない。このことから、流域内の主要な支川における観測データを使用することで、精度の高い再現を行うことができることが推察される。今後の計算においては、使用する観測所数 $P=6$ 、次数 $N=10$ として計算を行う。

本イベントにおいて構築したモデルについて詳しく見ていく。図-8は、下流端である八斗島地点を再現したモデルのそれぞれの観測所のパラメータである。パラメータには正負の値があるが、影響度の大きさのみに注目すると、流域3が最も大きく、次に大きいものが流域2, 4, 5であることが分かる。また、流域6に関しては、ほとんどの次数においてパラメータが小さいことがわかる。下流端である八斗島の流量は、上流の降雨量により決まるので、流域ごとの降雨量とパラメータの影響度は関係していると考えられる。そこで、本降雨イベントにおける流域の累積降雨量図が図-9である。それぞれの流域にかかる色が累積降雨量を表しており、黄色は、100mm～

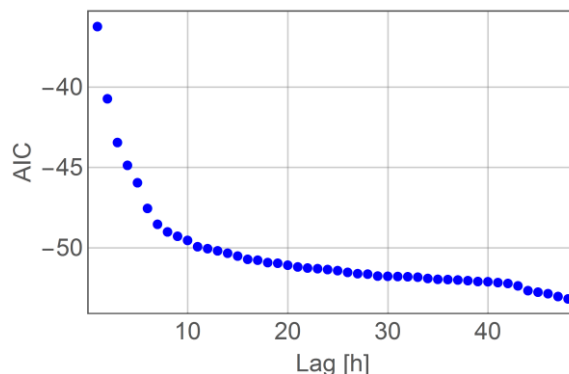


図-7 モデルの次数とAICの関係(観測所数：6)

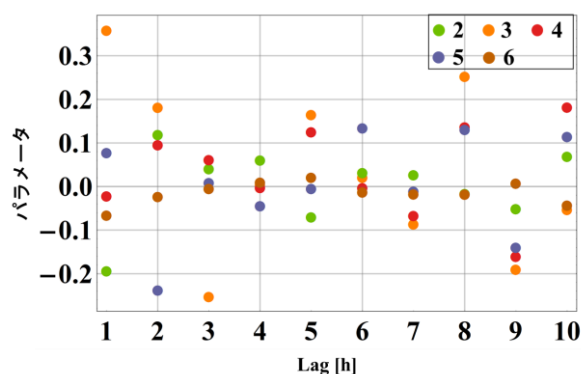


図-8 推定したパラメータの結果

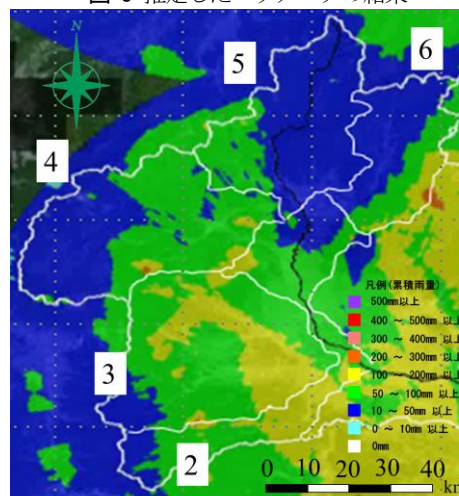


図-9 降雨イベントにおける流域の累積降雨量

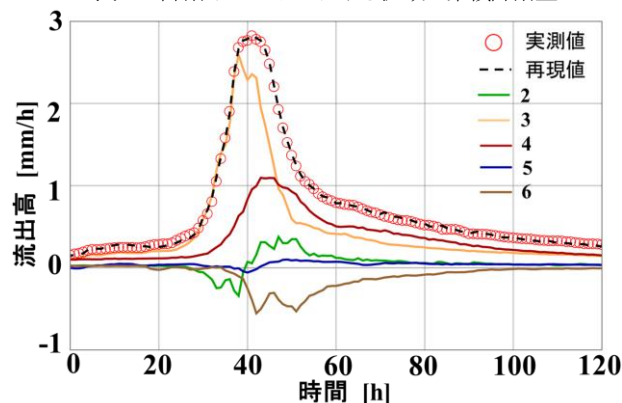


図-10 それぞれの流域ごとに分離した流出高

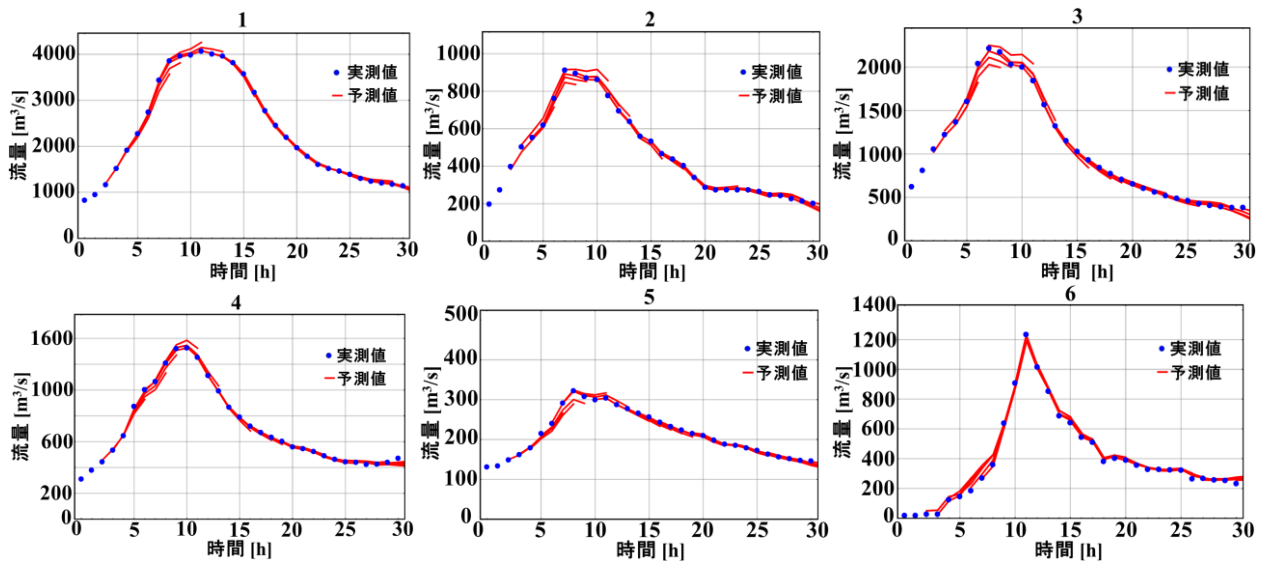


図-11 洪水イベントを含む1ヶ月間からパラメータ推定したモデルの予測値（5時間後まで）

200mm, 緑色は, 50mm ~ 100mm, 青色は, 10mm ~ 50mmを表している. この図より, 利根川上流域の南側である流域2, 3に多くの雨が降っていることが分かる. また, パラメータの影響度が大きい流域3であることから, パラメータの影響度と降雨量は関係していると考えられる. しかし, 流域2において, パラメータの大きさは, 流域内で累積降雨量が多いにも関わらず, 流域4, 5との明確な違いはわからない. 同様に流域4, 5は, 流域2より, 累積降雨量は少ないが, 流域3と同程度のパラメータ値となっている. この要因としては, 上流にダムが存在するため, 降雨が流出に影響していないことや地形や地質, 植生等による流域の流出特性の可能性がある. そのため, これらのパラメータを詳細に分析することにより, 流域特性の比較等を行える可能性がある. 今後検証を行っていく.

次に, (3)式から分かるように, ある観測所の流量を再現するために他の観測所地点が含まれているので, それぞれの流域ごとに分離した流量を算出することができる. それぞれの流域ごとに流出高を分離したものが図-10である. 赤丸は, 実測値を表しており, 黒破線は, モデルによって算出されたそれぞれの流量を足したものである. この図より, 八斗島地点におけるハイドログラフの形成に関する流域の影響度が分かる. 流域3が立ち上がり部分のハイドロ形成の大部分を担っており, その後, 流域番号4の流量が八斗島地点に到達することでピーク値が形成されていることが分かる. 流域2, 5, 6では, 負の流量が計算上で出てくるが, 勾配を考慮すると利根川上流域において逆流が起きているとは考えられない. これは, 統計モデルを用いているため, 実測値に合うようにパラメータ推定をしているためである. しかし, 図-9に累積降雨量図が示すように降雨量が多い流域がハイドログラフの形成に大きく影響を及ぼしていることから, 多変量自己回帰モデルで算出した値には流出特性が含まれてい

ることを示唆していると考えられる.

4. 解析手法の予測への適用

前章において, モデル構築に必要な観測所数 P , 次数 N を求めた. その結果より, 過去の洪水イベントを十分に再現できることが分かった. そこで, 本手法を予測に適用した. (3)式は, 現在の流量を求めるために過去のデータを求めることができることから, 現在の値を用いれば, 1時間先の値を求められる. そこで(3)式を現在の値までを含めて1時間先を予測式にしたものが(4)式である.

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

さらに x 時間後までの予測を行うとするならば, (4)式となる.

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

(6)式を用いれば, 過去のデータから予測を行うことができる. 本論分では, 洪水時の避難等の情報に役立てることを考え, 5時間後までの予測を行った. その結果が図-11である. 青点は実測値であり, 赤線は5時間後先までの予測値である. パラメータ推定に用いたデータは, 本降雨イベントを含む1ヶ月分の流量データであり, 6地点のそれぞれにおける結果を載せた. この結果より, どの観測地点においても立ち上がりからピーク流量において精度良く予測できていることが分かる. しかし, パラメータ推定に用いたデータが, 1ヶ月間という短い期間であり, 本洪水イベントのみに着目したものであることから次にパラメータ推定に用いるデータを本洪水イベントを含む, 5月~11月の6ヶ月間に長期間にしたものから予測を行ったものを図-12に載せる. この結果より, パラメータ推定に用いたデータ期間が1ヶ月よりも精度は

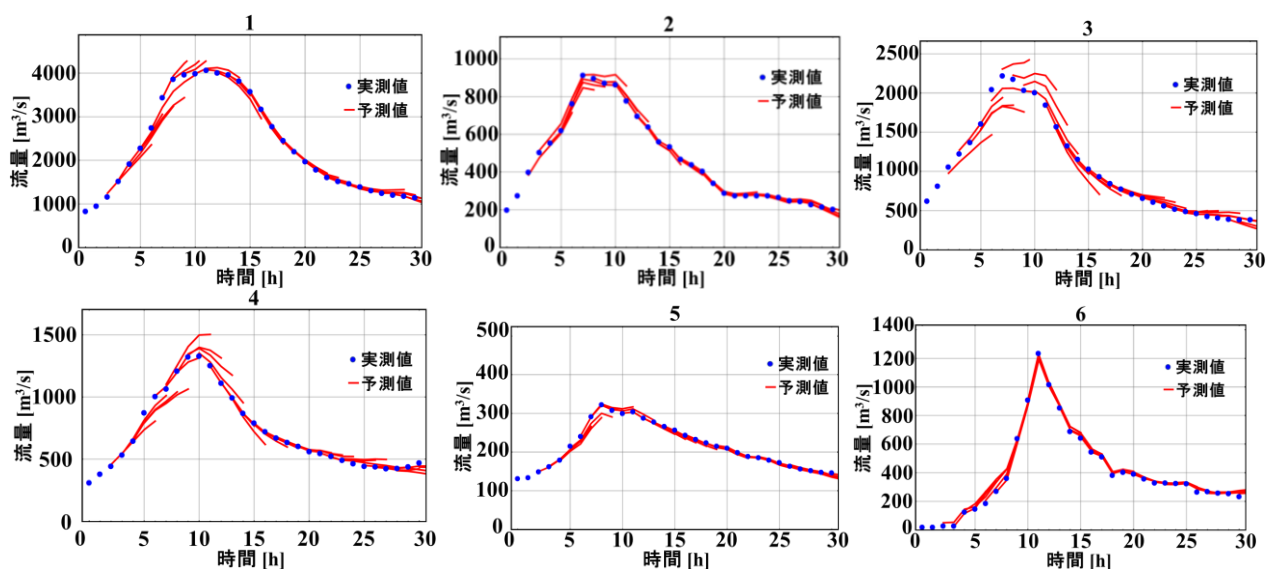


図-12 洪水イベントを含む6ヶ月間でパラメータ推定したモデルの予測値（5時間後まで）

悪くなっているが、避難等の情報に役立てるには、十分に精度良く予測できていることが分かる。このことから、洪水イベントを含むデータからパラメータを推定したモデルは、予測にも適用できることが分かった。

5. まとめ

本研究では、水文データの乏しい地域を想定し、時系列解析の手法である多変量自己解析モデルを適用させ、流量データのみによる洪水イベントの再現精度検証及び予測を行った。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の構築を行った。その結果、本研究の対象流域程度の大きさ（5100km²）ならば、流域内に数点あればピーク流量に対して、約±5%以内の再現性を持つモデルを構築することができる。より再現精度の高いモデルを構築するためには、主要な支川における観測地点を含めたモデルを構築することで可能となることが分かった。
- (2) 構築した多変量自己回帰モデルのパラメータは累積降雨量と関係があることが推測される。このことから、推定したパラメータから下流地点に及ぼす影響の大きさを推測できる。さらに、推定されたパラメータを詳細に分析することで、流域特性を推測できる可能性があるため、今後の課題である
- (3) 本手法による洪水イベントが十分に精度良く再現で

きることから、洪水予測に適用を行った。その結果、過去の洪水イベントを含むデータからパラメータ推定したものは、どの地点においても予測が可能であることが分かった。しかし、既往最大を超える洪水に対して、どこまで本手法により適応できるかを今後検討していきたい。

参考文献

- 1) 例えば、瀬戸心太，芳村圭，沖大幹：高分解能衛星降水マップによる日本域の洪水検出能力，水工学論文集，第52巻，pp.355-360，2008
- 2) 青木啓祐，干場希乃，諸岡良優，寺井しおり，山田正，吉見和紘：山地及び都市流域における流域特性が洪水ピーク流量に与える影響，第46回土木学会関東支部技術研究発表会，2018
- 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，pp.187-192，2016
- 4) 槻山ら：ニューラルネットワークによる阿武隈川洪水予測の基礎的検討，河川技術論文集，第9巻，pp.173-178，2003
- 5) 例えば，植田英三郎：時系列解析による株価予測，大阪府立大学経済研究，第53巻，2017
- 6) 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
- 7) Hirotugu Akaike, A New Look at the Statistical Model Identification, IEEE Trans. Automat. Contrl., NO.6, pp.716-723, 1974

(2019. 4. 2受付)