

タイトル	AICを用いた折れ線による水位流量曲線の自動最適化に関する研究
著者	許士, 達広; Kyoshi, Tatsuhiko
引用	工学研究 : 北海学園大学大学院工学研究科紀要(19): 3-9
発行日	2019-09-30

AIC を用いた折れ線による水位流量曲線の 自動最適化に関する研究

許 士 達 広*

A RESEARCH ON AUTOMATIC OPTIMIZATION OF STAGE-DISCHARGE RATING CURVE

Tatsuhiro Kyoshi*

要 旨

水位流量曲線は日本においては2次曲線の組み合わせが用いられ、水位 H と流量 Q の平方根 \sqrt{Q} の関係に目視で折れ点位置を定め、最小自乗法で折れ線を当てはめて求められている。その結果折れ線本数など形状に個人差が生ずることや、折れ線をつなぐことに労力を要することなどの問題があった。本研究において簡易な折れ点選定法や折れ線をつなぎ方を考案し、本数を AIC で最適化する方法を用いた結果、従来の目視よりも適合度の良い折れ線を数学的に求め、自動化して描くことができた。これにより水位流量曲線を描く際の省力化と精度の向上が実現された。

Key Words : stage discharge rating curve, optimization, polygonal line, AIC, automating

1. はじめに

河川の水位流量観測は通常水位のみが水位計により自動観測されており、流量への換算は各観測所における水位流量曲線 (H-Q 曲線) により行われる。水位流量曲線は日本の基準では(1)式のように2次曲線式で表され、(2)あるいは(3)式のように直線化して折れ線として算出されている。データに最適な折れ線を当てはめるため、現在 \sqrt{Q} は目的変数 H を説明変数として折れ線の各線分についてそれぞれ最小自乗法で定数を定めているが、その作成において折れ点 (折曲点) 候補点の選定、折れ線をつなぐこと、折れ線本数の決定の各段階にいくつかの問題点が存在する。

$$Q = a(H + b)^2 \quad (1)$$

$$\sqrt{Q} = \sqrt{a}(H + b) \quad (2)$$

$$\sqrt{Q} = AH + B \quad (3)$$

H : 水位 Q : 流量

一つ目は \sqrt{Q} -H でプロットした点群に最適な折れ線を当てはめる際に、作業者が目視で地形の変化点を参考にして折れ点を定めていることである。後述するようにこれ自体に大きな精度の問題があるわけではないが、折れ線形状に個人差が生じかつ自動化ができないため、精度が同程度以上であれば数学的に折れ点を選定する方が好ましい。

二つ目はデータを区分して最小自乗法でいくつかの直線とする場合、隣接する2つの線分間に交点が来ないことが多く、2つの線をつなぐのに作業者の判断による煩雑な作業を行っているのが実態である。これに対し省力化の観点から計算で自動的に交点を定める手法が求められる。

三つ目は折れ線の本数については上限がないため、担当者の考え方により細かく区分されたものや1本のみでつながっているものが存在することである。水文観測業務規程ではそれぞれの直線における相関係数0.8以上が目安とされているが、線を細かく分けるほど相関は上がっていき極端な

* 北海学園大学大学院工学研究科建設工学専攻 (社会環境系) 教授・博士 (工学)
Graduate School of Engineering (Civil and Environmental Eng.), Hokkai-Gakuen University

場合隣接する2点を全て結べば各線の相関係数が最大の1となる。一つの水位に対して幾つもの流量があてはまる凹凸の激しい線が最良となるという矛盾が生じるため、相関係数とは別に最適な折れ線本数を定める客観的な判断基準を導入する必要がある。

海外における水位量曲線は(4)式の河床の標高で流量ゼロとなるべき乗曲線を用いることが多い⁽¹⁰⁾⁽¹¹⁾。べき乗曲線は通常両対数を取って直線化して最小自乗法で係数を求め、一つの線で不十分なときは、2本～3本に分割するため、やはり折れ線の最適化問題となる。ここでも折れ点は目視により候補点を決め最小自乗誤差が最小となるものを選んでおり、日本の方式と同じ問題が生じる。

$$Q=a(H-b)^c \quad (4)$$

Q：流量　H：水位標高

a, c：データから定まる定数　b：河床の標高

水位流量曲線を対象にする既往研究としては、洪水時にループを描く問題や水理計算・流出解析を用いた流量精度の向上などがあるが、今回検討する観測されたデータから統計理論的に曲線の最適化を考えるものは極めて少ない⁽⁶⁾。また回帰による折れ線の最適化（データの点と点をつながない）は、数学の世界で見ても実際の計算例は1～2の折れ点にとどまり⁽⁷⁾⁽⁸⁾⁽⁹⁾、5～6本程度に分かれた折れ線に対し回帰による実用的最適化計算を行うものは見当たらない。

筆者らは過去にAICを用いたベイズ型一次スプラインを検討したが⁽⁶⁾、直接関数が示されないため水位から流量を算出する目的では取り扱いにくく、かつ外挿がむずかしいといった問題があった。この状況を踏まえ本研究では折れ線本数の最適化にAICを用い、かつ簡易に折れ点候補点を抽出し直線をつなぐ手法を開発した。これにより最適な回帰折れ線を数学的に自動化して算出し、水位流量曲線の作成において、省力化と精度向上が可能となった。なおここで扱うのは年間を一期として標高H方向に区分する場合であり、期別に分ける場合は別の検討とする。

2. 検討手法

折れ線の最適化は折れ点の最適配置により実現されるが、上述したとおり多本数に対する理論的最適化手法が開発されていない。したがって本研

究では実用的観点から便宜的な近似解を求めることとし、①折れ点の候補点の算出、②区分した線分を接続する、③線の組み合わせによるAIC最小化の3段階の作業で折れ線の最適化を行う。

(1) 折れ点候補点の選出

第一段階として線をつなぐに折れ点の候補点を何点か算出する。手法は水位流量観測データを幾つかに区分し、最小自乗法で直線を当てはめたときのデータとの残差平方和が最小となる区分位置の組み合わせを、繰り返し数値計算で算出する。区分の境目が第一段階の折れ点候補点である。区分される各線分の最小データ点数は3とし、データ間隔の違いから端部のみ最下端は4、最上端で2にしている。またこの段階では折れ線本数は、実際に用いられているH-Qの状況と計算時間の関係から最大5本（折れ点4点）とした。

具体的には下記(5)式の残差平方和 Se の式において水位流量データを水位が小さい順に並べ、直線の勾配 a_i および切片 b_i を算出しながら、区分する点（線分の上端の点） $m_1 \sim m_5$ を Se が最小となるように試行錯誤的に定める。

$$Se = \min \left(\sum_{i=1}^{m_1} [y_i - (a_1 x_i + b_1)]^2 + \dots + \sum_{i=m_4+1}^{m_5} [y_i - (a_5 x_i + b_5)]^2 \right) \quad (5)$$

x_i ：各時点の水位　 y_i ：各時点の流量の平方根

i ：データの下からの番号

$m_1 \sim m_5$ ：折れ点のデータ番号

折れ点候補点の組み合わせは以下の2つの考え方で選定する。選定法Aは折れ線が1から5本のそれぞれのケースについて(5)式で残差平方和最小となる区分点の組み合わせを算出し、それ自体を折れ点組み合わせの候補とするものである。例を示したものが表-1であり、2本から5本の本数に対し区分点すなわち折れ点候補点を表している。番号は水位の高さの順に水位流量データを並べた場合の水位が小さい方からの順位で、例えば3本の場合データ区分を小さい順に1番目から17番目、18番目から22番目、23番目からデータ個数（この場合36個）としてそれぞれの区分で直線を算出し折れ線とする。これに1本を加え、1～5本までの折れ点候補点で5通りの組み合わせが算出される。

選定法Bは選定法Aで算出したすべての折れ点候補点から、下記の条件で候補点を5個取り出

表-1 選定した折れ点候補（鶴川）

折れ点選定法	ケース	折れ点候補点
A	2本	17
	3本	17, 22
	4本	17, 22, 30
	5本	17, 21, 24, 34
B	5点	17, 22, 30, 24, 34

すものである。これにより最大6本の組み合わせの折れ線を比較する。取り出す条件はまず表-1で示すように、選定法 A で得られた折れ点候補点が本数により重複するので、重複個数が多いものを優先する。同じ個数の候補点があれば本数の少ないものの候補点を優先する。また、候補点番号が隣接する場合は、隣接する2点の線分はデータ上端を除き採用しないこととし、ここでは本数の多い方の候補点を除外する。表-1の例では4本重複する17、2本重複する22を選定。残る21、24、30、34のうち21は22に隣接するため除外となり24、30、34が選定され、17、22、24、30、34の5点となる。

なお水文観測業務規程¹⁾ではデータが観測された水位内に河道横断面変化点がある場合、折曲点は変化点から1m以内とされているため、断面変化点から1m以内に存在しない場合は断面変化点に最も近い観測点を候補点に加えるか、算定後1m以内の点に置き換える。このようにして後述するように選定法 B では候補点が5点の場合、折れ線1本から6本まで32通りの組み合わせができる。

表-1の2～5本までが手法 A によって選出されたもの、最下段が手法 B によって選出されたものである。選定法 A と選定法 B の組み合わせは多くが重複するが、選定法 B では選定法 A の候補点が網羅されない場合があり、選定法 A のみあるいは選定法 B のみでどの程度の精度が得られるかを見るため、明らかに重複する1本を選定法 B から除いた合計36通りの組み合わせを精度比較した。

(2) 区分した線をつなぐ

データを区分していくつかの直線とする場合、隣接する2つの線分の中に交点がないことが多く、極端な場合図-1に示すように2つの線が平行に近くなるなど、2つの線をつなぐのに工夫を要

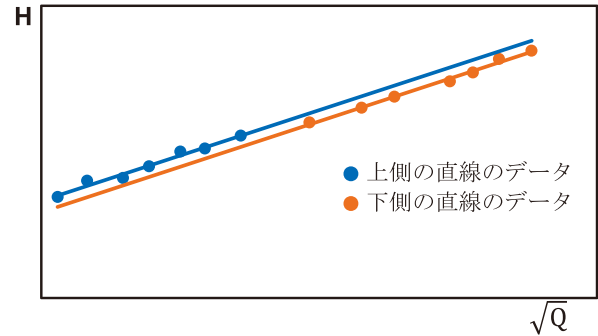


図-1 2つの直線が平行に近い例

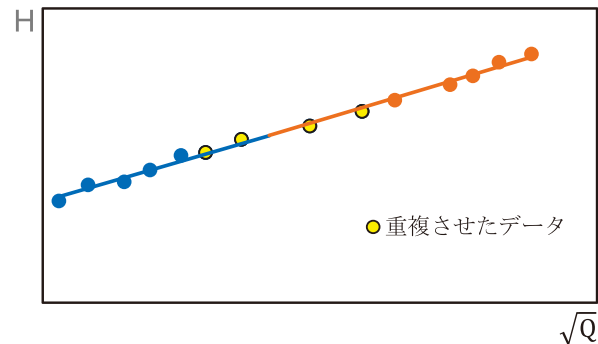


図-2 データを重複させて交点を作る方法

する。このような時は下側のデータの上端部と上側のデータの下端部の幾つかを使って中間に第3の線を作ることや、全体を1本の線にしてしまうなどの方法がとられることがある²⁾。多く用いられるのが図-2のように下側の上端と上側の下端のいくつかのデータを重複して使って直線を求め、データ区分点近くに交点ができるようにすることである。図-1の上部の青のデータと下部の赤のデータのそれぞれの直線が境界で交わらない時、図-2の黄色のデータを重複させ青と黄、赤と黄のデータでそれぞれ直線を算出して交点を作り、図-1の青と赤の間に交点に来るようにする。しかしどのような点を重複させるかは担当者の試行錯誤であり、個人差があると同時にかなりの時間と労力を要する。

今回この線をつなぐ作業の負担を減らすために以下の簡略した方法を用いる。図-3においてデータを下からの番号で1～ m_1 と m_1+1 ～ m_2 の2つに分けて、それぞれ最小自乗で直線を当てはめる。このとき2つの直線の交点がデータの端部の間、すなわち m_1 と m_1+1 の間に来ない時 m_1 と m_1+1 の間で強制的につなぐ方法をとる。

まず図-3-1に示すように区分したデータでそれぞれ回帰線を描き、2直線の交点が区分された

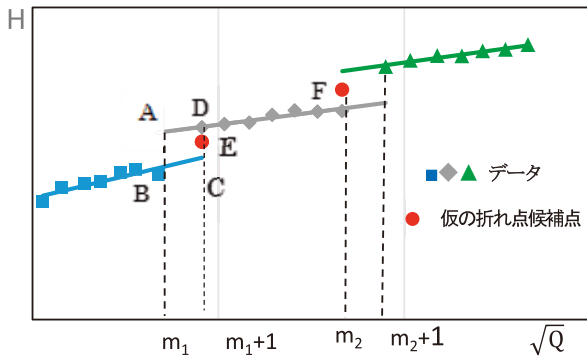


図-3-1 線分をつなぐ手順1

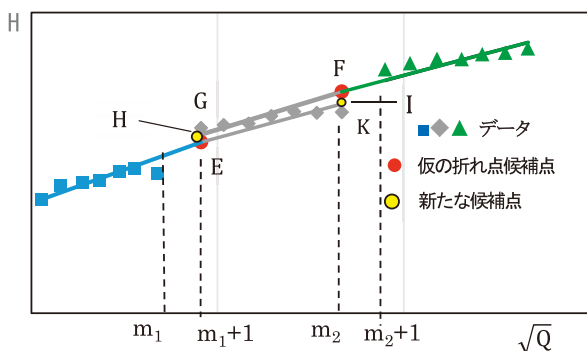


図-3-2 線分をつなぐ手順2

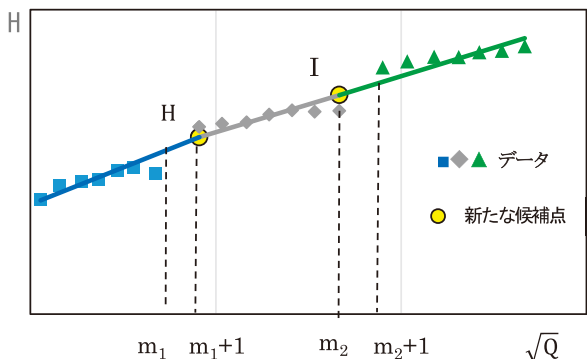


図-3-3 線分をつなぐ手順3

データの両端の点 B, D の間 (m_1, m_1+1 間) がある場合はその交点を、交点がない場合は A と B, C と D の長さを比べ、小さい方の中点 E を仮の折れ点とする。点 F も同様に定める。次に図-3-2 に示すように上で定めた仮の折れ点 E を通る直線 EK, 仮の折れ点 F を通る直線 FG をデータとの回帰で定める。G は仮折れ点 F を通る直線の E と同じ x 軸の位置 (m_1+1) における値であり、K も E を通る回帰直線から同様に定める。

図-3-2 では EK, FG のように候補点からの線が重複するため、E と G および F と K の間に再

度中点 H および I を求め、新たな折れ点候補点とする。折れ線 4 本以上では 3 本の線の端部が重なる位置が生ずるが、そこでは中点は 3 本の端部の点の平均とする。最後に図-3-3 に示すように外側の 2 直線は新たな折れ点候補点 H および I を通る直線を最小自乗法で求め、内側の線は H と I をつないだものとする。この方法は線が結びついていない段階で算定される水位を折れ点の水位候補点として結ぶため、多少誤差が生ずることになるが、後述するように目視で線を引くよりはなおむね良好な値となる。

(3) AIC による分割本数の最適化⁽⁴⁾⁽⁵⁾

線をつないで折れ線にした場合、前述したように折れ線が 1 本から 6 本の範囲であれば表-2 の例のように 36 通りの組み合わせが考えられ、これらから最適な分割を選択するため AIC (赤池情報量基準) を適用する。AIC は

$$AIC = -2 \times (\text{モデルの最大対数尤度}) + 2 \times (\text{モデルの自由パラメータ数}) \quad (6)$$

で表され、AIC を最小にするモデルが最適である。これはモデルの当てはまりを表す最大対数尤度が同程度であれば、パラメータ数 (定数の数) が最も少ないものを選んでおり、節約の原理の一つの具現化となっている。AIC を用いれば例えば直線から 6 次曲線のうち、どれの当てはまりが良いかといった問題の答えを出すことができ、折れ線の場合も同様に最適化できる。回帰モデルに対しては AIC は以下ようになる。

$$AIC = n \log 2\pi + n \log \sigma^2 + n + 2k \quad (7)$$

モデル比較のためには定数部分を無視して

$$AIC = n \log \sigma^2 + 2k$$

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 \quad (8)$$

σ^2 : 誤差分散

n: データ数

Y_i : データ値,

\hat{Y}_i : 回帰直線または曲線による推定値

ここでの log は自然対数である。2k をバイアス補正項と呼び k は式に含まれるパラメータ数で、直線 $y = ax + b$ では a と b の 2 つであるから、折れ線の場合は線の数 $\times 2$ でパラメータ k が増える。AIC はデータ数の違いによる補正式

CICがあるが、ここでは同一データの折れ線本数の違いを見るため、(8)式を採用した。また他にもAICを改良した情報量基準は多数あるが、通常用いられる(8)式とした。

3. 最適化計算

(1) 折れ点組み合わせの最適化

以上のプロセスを全自動化して計算する。今回用いたデータは河川管理者である官庁が当該年度の水位から流量を算出する計算で用いたものを使用している。すなわち2箇年の水位流量データの内、前年度のデータおよび結氷などにより異常な計測の可能性があるデータの選択は、管理者が判断したものを用いており、これにより後述する本

手法と官庁が実施した従来手法の精度比較が可能となる。折れ点候補点をつないだ36の分割パターンに対し、AICを計算すると表-2のようになる。表中の「組み合わせ」とは、折れ点候補点を用いて直線を区分したパターンであり、1～6は折れ点候補点で区分されたデータ点群の下からの番号である。例えば123-456は下から1～3のデータ群の直線と4～6のデータ群の直線で2つに分けられている。右端の順位の欄はAICが小さい順を示しAICのマイナスが最大なものが最適で、表の例では3本の123-45-6となる。順位が1から36位までついているが、この例では選択法Aは全て選定法Bに網羅されている。しかし前出の表-1の例のように、選定法Aの折れ点の中に選定法Bで選択されないものが入ることがあり、選定法Bだけでは不十分である。

表-2 折れ線最適本数結果（多度志）

選定法	本数	組み合わせ	AIC	順位	
A	1	123456	-118.53	33	
	2	1234-56	-151.59	15	
	3	1234-5-6	-168.56	5	
	4	12-3-45-6	-168.34	7	
	5	1-2-3-45-6	-153.23	14	
B	2	1-23456	-126.03	32	
		12-3456	-139.73	27	
		123-456	-153.63	12	
		1234-56	-151.59	16	
	3	12345-6	-115.09	35	
		1-23-456	-150.63	20	
		1-234-56	-148.76	23	
		123-4-56	-145.99	24	
		123-45-6	-172.01	1	
		12-34-56	-151.16	19	
		12-3-456	-151.48	18	
		12-345-6	-148.95	22	
		1-2-3456	-94.68	36	
		1-2345-6	-132.96	30	
	4	1234-5-6	-168.56	4	
		123-4-5-6	-167.81	9	
		1-2-3-456	-139.55	28	
		1-23-45-6	-168.16	8	
		12-34-5-6	-171.00	2	
		1-2-34-56	-136.19	29	
		12-3-4-56	-142.79	25	
		1-234-5-6	-168.68	3	
		1-2-345-6	-117.48	34	
		12-3-45-6	-168.34	6	
		1-23-4-56	-141.86	26	
		12-3-4-5-6	-165.87	10	
		5	1-23-4-5-6	-164.11	11
			1-2-34-5-6	-151.56	17
1-2-3-45-6	-153.23		13		
1-2-3-4-56	-131.24		31		
6	1-2-3-4-5-6	-150.31	21		

(2) 計算精度および従来の方法との比較

今回の手法の精度を評価する上で既往の手法との精度比較が必要となる。管理者からデータ年の水位流量曲線を提供してもらい、既存の $H-\sqrt{Q}$ 成果に対してもAICを算出した。これは観測 \sqrt{Q} 値と既存成果の計算値 \sqrt{Q} から誤差分散 s_e^2 を算出し求めたものである。

表-3は石狩川下流部の26箇所、鶴川、沙流川各1箇所の観測所について、今回の手法A、Bと既往の手法のAICおよび最適折れ線本数を比較したものである³⁾。赤字が最もAICが小さく適合性が高いものを示し、28か所中26か所で今回の手法A、Bいずれかが既往の手法より精度が上回った。下回った場合も差は僅かであり、今回の手法でおおむね精度の向上が期待できる。選定法AとBでは、選定法Bが良い場合が選定法Aがと一致するものも含め23か所あるが、選定法Aが選定法Bより良いものも5か所あり、選定法A、Bの両方を計算する必要がある。

今回の手法と既往の手法の折れ線を図示して比較したものを表-3に示す。折れ線の図化は既存のいろいろな方法で可能であり、最適折れ線が自動化で計算されれば図化まで自動化できる。AICの差が大きい例として図-4-1の幌加内はAICの差が24で、違いの原因は中間の折れ点の位置と本数が違うことによる。図-4-2の千秋は差が23程度で折れ線の本数の違いが原因で低水部に差がある。AICは流量の大小の影響を受けるため、

表-3 折れ線最適本数結果

観測所	選択法A		選択法B		既往成果		既往と本研究の差
	本数	AIC	本数	AIC	本数	AIC	
西川向	5	-215.53	4	-213.16	4	-224.30	-8.77
西越	2	-220.80	2	-220.83	2	-219.71	1.12
幌加内	5	-130.51	3	-140.31	2	-116.29	24.02
石山	5	-200.06	5	-196.07	2	-180.40	19.66
当別川下	3	-260.04	5	-262.47	2	-211.79	50.69
薄別	1	-249.32	2	-249.32	1	-242.78	6.54
望月寒	2	-171.41	2	-173.11	1	-169.86	3.25
オカバルシ	3	-79.40	3	-79.40	2	-63.96	15.45
岩見沢大橋	5	-3.99	4	-9.61	4	-7.17	2.44
千秋	5	-75.96	4	-93.92	1	-70.98	22.94
多度志	3	-168.56	3	-172.01	2	-153.54	18.47
旧雨竜橋	2	-43.66	2	-43.66	2	-41.97	1.70
舞鶴	2	-128.22	2	-128.22	3	-129.07	-0.85
白井	4	-124.55	2	-107.26	2	-106.30	18.26
藤松	4	-285.67	3	-296.24	3	-295.16	1.08
赤平	2	-133.07	2	-137.56	3	-133.60	3.96
納内	2	-131.40	2	-134.31	2	-133.96	0.36
妹背牛橋	3	-97.41	2	-108.89	2	-107.93	0.96
雨竜橋	2	-174.35	2	-174.35	2	-163.29	11.06
円山	2	-209.80	3	-223.37	2	-219.77	3.60
清幌橋	2	-234.15	2	-234.15	1	-213.09	21.06
下島松	3	-317.89	3	-326.42	2	-314.32	12.11
橋本町	2	-118.40	3	-126.39	2	-100.56	25.84
奈井江大橋	5	-168.36	5	-167.26	3	-140.79	27.57
輪厚	3	-309.38	5	-311.61	3	-297.46	14.14
嶮淵	2	-222.58	2	-222.58	2	-220.37	2.21
平取	1	-140.80	3	-138.83	2	-139.06	1.74
鶴川	2	-103.94	2	-106.39	2	-103.65	2.73

図-4-3の奈井江大橋は差が28程度であるがほとんど線が重なって見える。他のAICの差が一桁などの箇所は外見上全く線が重なり、流量に換算した場合も差はほとんどなくなる。このため本研究の方法は現状よりも流量の推定精度は向上するが、既往の方法との流量差は大きなものではない。

今回の方法のメリットはデータを入力すれば図化まで自動的に行われる省力化の面で大きい。計算時間はデータ数により異なりデータ数50程度ならば5分強、70では20分強であるが、計算時間中他の作業をしていけば時間のロスにならない。計算時間の大半は(5)式に費やされる。式(5)で算出する候補点数を現在の4から5に増やすとさらに精度は向上すると考えられるが、計算がデータ数に近い形で増大するため、実用面を考えると(5)式では候補点増は難しい。

4. まとめと今後の課題

H-√Q 曲線について①折れ点候補点の選定、②折れ線をつなぐ、③折れ線本数の決定、それぞれの段階でプログラムによる最適化と効率化を試みた。計算結果は従来の人間の感覚を主とした既存成果に近く、かつほとんどの箇所で精度の向上が見られた。また計算を自動化したことにより、従

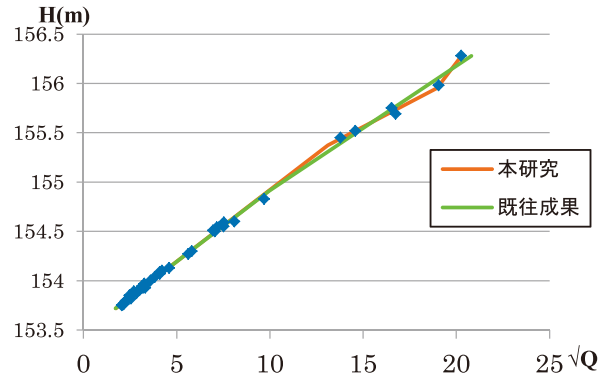


図-4-1 折れ線比較（幌加内）

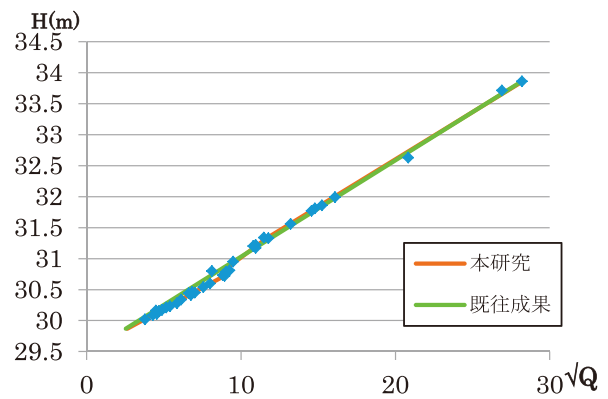


図-4-2 折れ線比較（千秋）

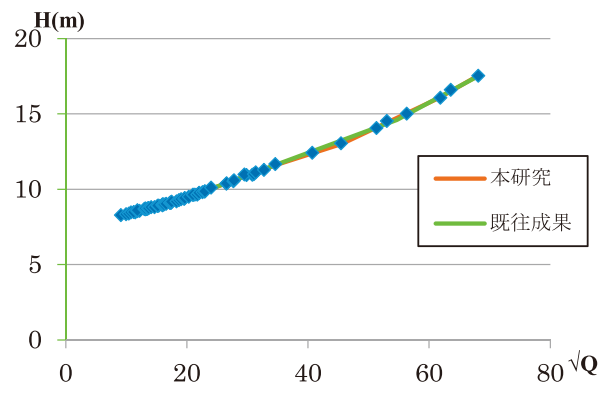


図-4-3 折れ線比較（奈井江大橋）

来の手作業で試行錯誤した部分に対し大幅な省力化が図られた。さらに従来何本が最適であるか算定できなかった多本数の折れ線近似に対し、AICによる最適化理論を導入し、人間の主観で行われていた工程に客観性が導入された。本研究の中では観測データから異常値を外すことや前年度のデータを取り入れる作業は行わず、管理者が選定したデータを用いている。水位流量曲線決定にはデータの判断も大きな比重を占めるが、その部分

は背後にある物理的要因を考える必要があり、単純には自動化できない。

今後の課題としては水位流量曲線が期別に分離される場合の検討がある。水位流量曲線のうち何割かは期別に水位流量関係が異なる枝分かれした線となっている。これについても AIC の適用が可能であるが、標高 H 方向に加え時間方向にもデータを分離するため、計算が複雑となる。また今回一部既往の方法より精度が悪かった点についても原因を考え、実用性の向上を図りたい。

謝辞：本研究に際し情報を提供いただいた国土交通省北海道開発局、同札幌開発建設部、株式会社福田水文センター、ご支援いただいた北海道河川財団、学生として本研究に従事した吉村悟史君、藤林佳汰君、井原拓美君、小川篤人君その他の卒業生諸君に厚く謝意を表します。

参考文献

- 1) 国土交通省河川局：水文観測業務規程関係集，pp. 245～256，財団法人河川情報センター，2005
- 2) 独立行政法人土木研究所：水文観測，pp.221～241，全日本建設技術協会，2002
- 3) 札幌開発建設部：平成 24 年度年度流量観測業務報告書（豊平川・千歳川・夕張川・空知川・雨竜川），2013
- 4) 坂本慶行，石黒真木夫，北川源四郎：情報量統計学 pp.128～138，共立出版株式会社，1983
- 5) 小西貞則，北川源四郎：情報量基準，pp.180～182，朝倉書店，2004
- 6) 高橋啓，許士達広，石黒真木夫：ベイズ型スプライン回帰による水位-流量曲線のフィッティング，土木学会論文集 B1，Vol.71，No.22，1-11，2015
- 7) 大塚雍雄，吉原雅彦：1 ないし 2 の折曲点を持つ折れ線モデルのあてはめ，応用統計学 vol.5，No.1，1975
- 8) Aronov Boris Asano Tetsuo, Katoh Naoki., Mehiorh Kurt, Tokuyama Takeshi: Polyline Fitting of Planar Points under Min-Sum Criteria, Lecture Notes in Computer Science, 3341, Springer, 2005
- 9) Feipeng Zhng., Qunhua Li.: Robust bent line regression, Journal of Statistical planning and inference, vol185, 2017
- 10) Wikipedia: Rating Curve
- 11) DHV Consultants BV & DELFT HYDRAULICS: HOW to establish stage discharge rating curve, Word Bank and Government of Netherlands funded training module, 1999