

一級水系代表観測点における流量年表データベースの信頼性について

首都大学東京	都市基盤環境コース	学生会員	○岡元 宏薫
首都大学東京	都市基盤工学専攻	学生会員	梶井 剛
首都大学東京	都市基盤工学専攻	正会員	河村 明
首都大学東京	都市基盤工学専攻	正会員	天口 英雄

1. はじめに

現在、日本における 109 の一級水系では、その幹川および支川の複数個所で流量観測が行われ、河川流況を分析するために必要なデータは整っていると言える。このデータは（社）日本河川協会が毎年発行している流量年表¹⁾などで容易に取得することができる。しかしながら、デジタルデータとしての日流量データベースは今まで公には存在しておらず、各研究者が独自に必要とする日流量を流量年表より抽出し電子化していた。現在、流量年表のデータは河川協会よりこれらをデジタルデータ化して「雨量・流量年表データベース」²⁾（以後、DVD 流量年表と記す）として販売されている。しかしながら、DVD 流量年表から抽出したデータより流況曲線を作成すると、明らかに異常なデータが存在し、流量年表¹⁾と比較したところ数多くの相違点が見つかった。そこで、本研究では全国一級水系代表観測点における DVD 流量年表データを流量年表のそれと比較・精査することにより、DVD 流量年表データの入力ミスを抽出するとともに、流量年表自体のデータについても精査を行った。

2. 用いたデータ

本研究では各水系の幹川における代表観測点のデータを抽出しており、その選定方法は以下のようなものである。

1. 河川ハンドブック 2007³⁾の中で一級水系における幹川の基準点として挙げられている観測点のデータを採用する。
2. 基準点よりも 5 年以上データ存在期間が長い観測点が他にある場合はそちらを採用した。また、基準点よりも 5 年未満しか長くない場合はより流域面積が大きい方を採用する。

観測点は河川改修や設置場所の変更と共にその名前が変更されていることも多く、その実態を把握することは必要不可欠である。しかし、名称の異なる同一観測点を見つけ出すのは容易ではないため、本研究ではある観測点の河口からの距離と対象とする流域面積が共に 5%以

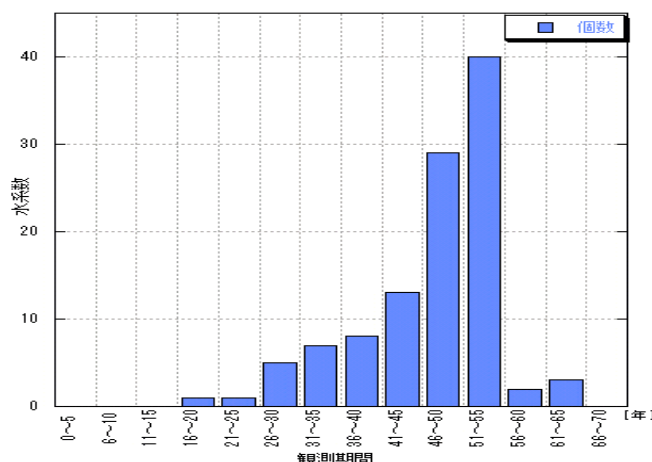


図-1 データ存在期間別ヒストグラム

内の差であった場合、同じ観測点であると見なした。また例外として観測点の観測期間の都合上、九州の六角川水系(佐賀県に河口を持つ)においては幹川の六角川ではなく、その支川である牛津川の妙見橋観測点のデータを採用した。さらに、相模川水系(神奈川県)における代表観測点は、京浜河川事務所より提供をうけた相模大橋観測点を選定し、流量年表には記載は無いがその流量データを採用した。選定した代表観測点におけるデータの記載開始年から 2003 年までの日流量データを解析のために用いた。選定した 109 一級水系代表観測点における観測期間のヒストグラムを図-1 に示す。この図より、観測期間は長くても 65 年であり、46 年～55 年の観測期間を持つ観測点が多いことがわかる。全水系の観測期間平均は 45.64 年間であった。

3. DVD 流量年表データの精査

DVD 流量年表の日流量データを精査するために流量年表と比較した。流量年表には日流量データのみならず月総流量、年間総流量が記載されているので、DVD 流量年表の日流量データを累積し、算出した年間総流量がそれと一致していればその年の日流量データは全て流量年表と同じであるとみなした。

キーワード：流量年表，一級水系，精査，入力ミス，日流量データ

連絡先：〒192-0397 東京都八王子市南大沢 1-1 首都大学東京 都市基盤環境コース

E-mail : okamoto-hironobu@ed.tmu.ac.jp

一方、年間総流量が一致しなかった場合、異なった原因の日流量データを検出するため各月総流量で比較し、最終的に流量年表とどの日流量データが異なっているのかを見つけた。この作業の結果、発見したDVD 流量年表の流量年表との相違点の詳細は以下のものであった。

A) 観測点名の入力ミス

単純にDVD 流量年表に登録してある観測点名の漢字入力ミスが発見された。例えば、荒川(新潟県)において1970～1978年までの期間のみ観測点名が「葛籠山」となっているのに対しそれ以外の期間は「葛籠山」ということになっていた。また、那賀川(徳島県)において1956年のみ観測点名が「古庄」(正しくは「古庄」)となっていた。これらの入力ミスは、DVD 流量年表から日流量データの抽出をする際に、観測点名を入力しなければならないので重要な事項となる。

B) カンマ記号について

DVD 流量年表データの出力形式には表形式とデータ登録形式の2つあり、前者は観測年や観測所名などの諸データを含んでおり、抽出に非常に時間がかかるため、データ登録形式にて抽出を行い、抽出した日流量データの確認のため流況曲線を作成した。図-2にDVD 流量年表から抽出した2003年の石狩川水系石狩大橋の日流量データの流況曲線の一例を示す。これを見ると明らかに流量がほぼ0 m³/sに近い流量データが存在

するものの、その前後では大きな流量となっている箇所が多く見られる。そこで、入力されたデータを調べてみた結果、図-3a)のようにDVD 流量年表データの各数値間の区切りにはカンマが使用されており、CSV形式(カンマ区切り)で読み込むことができるが、実はDVD 流量年表データの千の位の後にも位取りのカンマが入力されていることが分かった。すなわち数値間のカンマと位取りのカンマの区別がされなかったため、図-3b)で示すように、例えば、1964.10 m³/sが1 m³/sと964.10 m³/sの2つのデータとして取り扱われていることが判明した。そこで、DVD 流量年表を用いる場合、位取りで使われているカンマを消去するなどなんらかの処理が必要である。

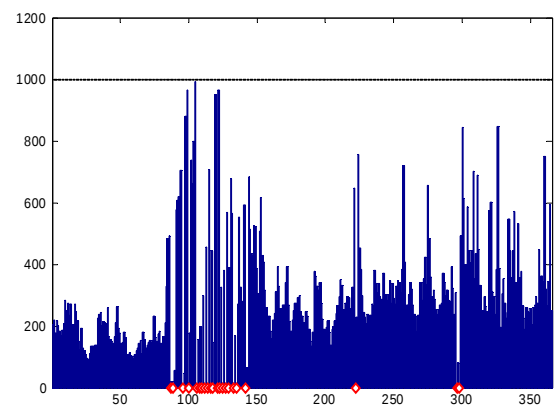


図-2 データ登録形式で抽出した日流況曲線の例

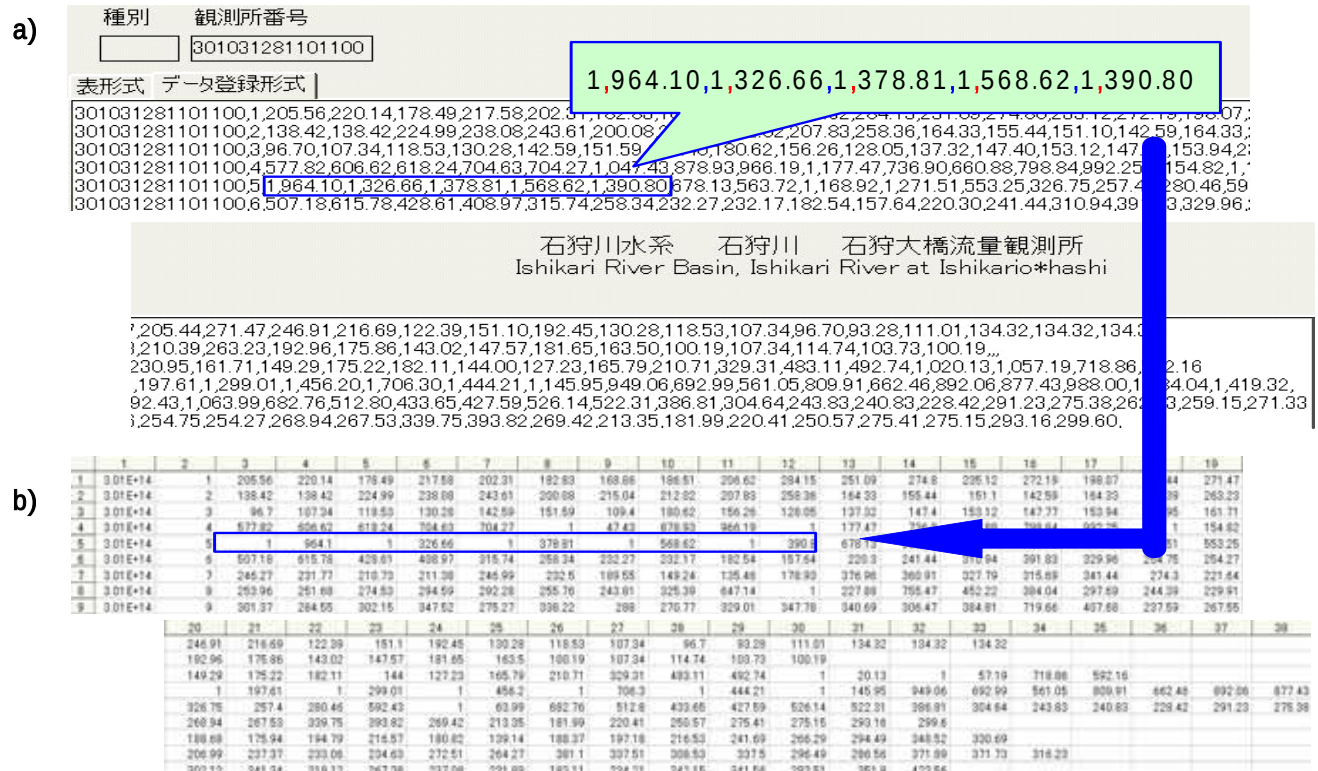


表-1 DVD 流量年表の入力ミス並びに流量年表記載ミスの個数

観測点 番号	水系名	観測点名	観測期間 (年)	全日データ個 数(1977年 除く)	DVD流量年表の入力ミス※				流量年表の記載ミス※			
					α	β	γ	小計	α	β	γ	小計
1	天塩川	菅平	51	18263				0		1	2	3
3	湧別川	開盛	51	18263			2	2				0
5	網走川	美幌	50	17898				0			1	1
6	留萌川	大和田	43	15341				0	1			1
7	石狩川	石狩大橋	51	18263			3	3			2	2
9	後志利別川	今金	49	17533		1	15	16			4	4
10	鶴川	鶴川	50	17898			1	1			1	1
11	沙流川	平取	46	16437		1		1	1		1	2
12	釧路川	標茶	49	17533				0			1	1
13	十勝川	茂岩	51	18263				0	1			1
14	岩木川	五所川原	52	18628				0			2	2
16	馬淵川	剣吉	42	14976				0			1	1
17	北上川	孤禅寺	54	19359				0	1			1
18	鳴瀬川	三本木	52	18628				0			4	4
21	米代川	ニッ井	47	16802		1		1			4	4
22	雄物川	樺川	59	21185		2		2	1		3	4
24	最上川	高屋	46	16437				0			1	1
25	赤川	熊出	38	13515				0			1	1
26	荒川	葛籠山	34	12054			2	2				0
27	阿賀野川	馬下	54	19359				0	3		2	5
28	信濃川	小千谷	54	19359				0			3	3
29	関川	高田	35	12419				0			1	1
31	黒部川	宇奈月	53	18994				0			3	3
32	常願寺川	瓶岩	44	15706			1	1				0
33	神通川	神通大橋	47	16802		1	4	5			6	6
34	庄川	大門	50	17898				0	1			1
35	小矢部川	津沢	46	16437				0			1	1
36	手取川	中島	52	18628				0			1	1
37	梯川	埴田	33	11689				0			1	1
38	久慈川	山方	47	16802			1	1				0
41	荒川	寄居	54	19359			17	17			1	1
42	多摩川	石原	53	18994			3	3				0
43	鶴見川	亀の子橋	34	12054				0			1	1
45	富士川	清水端	53	18994				0			4	4
46	狩野川	太仁	53	18994				0	1		2	3
47	安倍川	手越	43	15341				0			1	1
48	大井川	神座	46	16437				0	1		1	2
49	菊川	加茂	51	18263				0			1	1
50	天竜川	鹿島	63	22646		13	1	14		1	1	2
51	豊川	石田	61	21916				0			1	1
52	矢作川	岩津	65	23377				0	1		2	3
54	木曾川	犬山	54	19359		1	5	6			2	2
55	鈴鹿川	高岡	51	18263				0	1			1
56	雲出川	大仰	44	15706			3	3		1	1	2
57	櫛田川	両郡橋	41	14611			2	2		1	2	3
59	九頭竜川	中角	46	16437				0		1		1
61	由良川	福知山	51	18263				0			5	5
62	淀川	枚方	53	18994				0	1			1
64	円山川	上小田	40	14245				0		1	1	2
65	加古川	国包	37	13150			1	1				0
67	紀の川	船戸第二	53	18994				0			1	1
68	新宮川	相賀	53	18994			11	11	1			1
69	千代川	行徳	51	18263			1	1				0
72	斐伊川	大津	48	17167			1	1			1	1
73	江の川	尾関山	48	17167				0			1	1
75	吉井川	御休	48	17167			18	18		1	1	2
76	旭川	牧山	41	14611			28	28				0
77	高梁川	日羽	41	14611			2	2				0
78	芦田川	山手	39	13880		1	18	19			2	2
80	小瀬川	両国橋	37	13150			1	1				0
81	佐波川	新橋	51	18263				0	1		1	2
82	吉野川	岩津	48	17167				0			1	1
83	那賀川	古庄	49	17533			5	5				0
85	重信川	出合	50	17898				0			1	1
86	肱川	大洲	52	18628	1		1	2		1		1
87	物部川	深淵	44	16336			2	2				0
88	仁淀川	伊野	53	18994			3	3	1			1
89	四万十川	具同	52	18628		1		1				0
90	遠賀川	日の出橋	54	19359			4	4			3	3
91	山国川	下唐原	51	18263	1		1	2				0
92	筑後川	瀬ノ下	55	19724		1	13	14	1	1	1	3
94	松浦川	牟田部	42	14976			1	1			1	1
96	嘉瀬川	川上	36	12784			4	4				0
98	菊池川	玉名	46	16437			2	2				0
99	白川	代雑橋	48	17167			2	2	1			1
100	緑川	城南	31	10958			1	1				0
101	球磨川	人吉	54	19359			7	7				0
102	大分川	明鏡橋	52	18628			10	10			1	1
103	大野川	白滝橋	55	19724				0			2	2
104	番匠川	番匠橋	51	18263			20	20				0
105	五ヶ瀬川	三輪	60	21550				0			1	1
107	大淀川	柏田	44	16336				0			1	1
108	川内川	斧瀬	50	17898				0			2	2
109	肝属川	俣瀬	52	18628			7	7			2	2

注 α：小数点位置の入力 β：数字入れ替わり・抜け・加わり γ：数字の読み間違い

C) 1977 年のデータについて

図 - 4 に利根川における 1977 年の DVD 流量年表および流量年表の日流量データを用いた流況曲線を示している。この図より二つの流況は全く違っていることが分かる。精査の結果 DVD 流量年表の日流量データが間違いであることが分かった。また、本研究で対象とした 109 水系の代表観測点の中でデータ記載開始年が 1977 年以前のは 106 水系あるが、そのうち 101 水系で 1977 年の日流量データは全く違う値が入っていた。何故このような数値がこれほどまで多くの観測点において入力されたのかは不明である。

D) 数値入力ミス

前項の B), C) 以外にも日流量データの相違点が数多く見つかった。具体的には α) $1.56 \text{ m}^3/\text{s}$ が $15.6 \text{ m}^3/\text{s}$ のように小数点の位置を付け間違えた入力ミス, β) $25.96 \text{ m}^3/\text{s}$ が $25.69 \text{ m}^3/\text{s}$ のように数字が入れ替わっている入力ミス, または $111.50 \text{ m}^3/\text{s}$ が $11.50 \text{ m}^3/\text{s}$ のように抜けてしまっているもしくは逆に余計な数字が加わっている入力ミス, γ) 3 と 8 や 1 と 7 のように似通った数字の読み違いによる入力ミスが多数存在していることが分かった。これら DVD 流量年表データの数値入力ミスの個数を表 - 1 に示す。なお表 - 1 には入力ミスのあった水系, もしくは後述する流量年表の記載ミスがあった水系のみ挙げている。この表を見ると数値入力のミスがあったのは 43 河川あり 特に中国地方の水系に多く存在していることが分かった。また, 数字の読み違いによる入力ミスが一番多く見られた。

4. 流量年表自体の精査

流量年表には日流量データの数値の表のみならず, その流量による時系列グラフも図示されている。そのグラフに対応する流量年表の日流量データを比較精査した結果, 流量年表そのものにも明らかに誤りと考えられる数値が記載されていることが分かった。これらの記載ミスとして, DVD 流量年表と同様に, 記載時での α) 小数点のつけ間違い, β) 数字の入れ替わり・抜け・加わり, γ) 単純な数字の読み間違いが考えられる。これら流量年表自体の記載ミスの個数も表 - 1 に併記している。この表を見ると記載ミスの存在する水系は 65 河川あり 場所に関わらず全国のあらゆる河川に記載ミスは見られた。また, DVD 流量年表データと同様に, 数字の読み間違いによる記載ミスが一番多く存在するという結果が得られた。

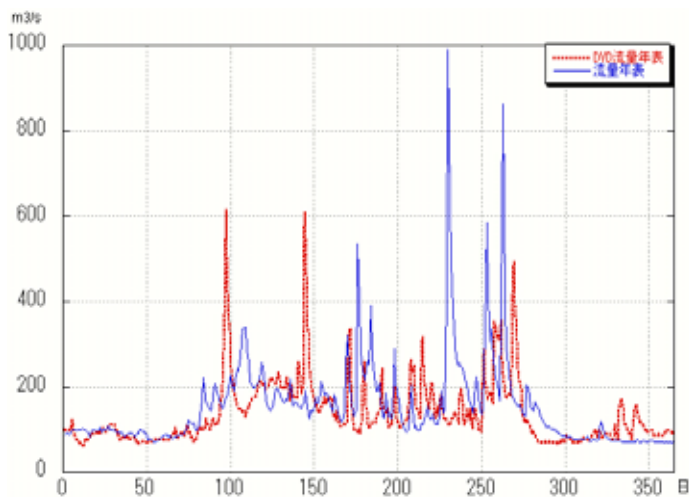


図 - 4 利根川における 1977 年日流量データの相違

5. むすび

本研究では一級水系の代表観測点における DVD 流量年表および流量年表の日流量データの信頼性について検討した。その結果, DVD 流量年表には観測点名の入力ミス, 位取りカンマ記号による流量年表との相違, 1977 年の多くの水系における入力ミス, 単純な数値入力ミスがあることが分かった。また, 流量年表自体についても, 小数点のつけ間違い, 数字の入れ替わり・抜け・加わり, 数字の読み間違いだと考えられる記載ミスが存在することが分かった。なお, DVD 流量年表および流量年表の日流量データの信頼性について検討するに伴い, 一級水系の代表観測点においてデータベースを再構築も行った。

通常, 水文研究者および実務者はそのまま与えられたデータを用いて解析することも多いが, 本研究で示したように, まずはデータの確認・精査は解析を行うにあたって重要であると言える。

謝辞

本研究を遂行するに当たり, 流量年表に関する数多くの問い合わせにお答え下さった日本河川協会の担当者の方や, 対象河川に関する諸データを提供して下さいた各地方整備局の担当者の方から多大な協力を得ました。ここに謝意を表します。

参考文献

- 1) 国土交通省(建設省)河川局編: 昭和 13 年度~平成 15 年度流量年表, 日本河川協会, 1938-2003。
- 2) 「雨量・流量年表データベース」, 日本河川協会, 2007。
- 3) 2007 河川ハンドブック, 日本河川協会, 2007。