

市販流量年表データベースによる一級水系代表観測点日流量データの検証

首都大学東京 大学院都市基盤環境学域 正会員 ○河村 明
首都大学東京 大学院都市基盤環境学域 非会員 岡元 宏薫
首都大学東京 大学院都市基盤環境学域 正会員 天口 英雄
首都大学東京 大学院都市基盤環境学域 正会員 中川 直子

1. はじめに

河川の水害防御，水資源確保，環境保全などの課題に対処するには，まずその基礎的水文資料，特に河川流域の長期にわたる雨量および河川流量データの解析が必要不可欠となる．現在，日本における109の一級水系では，昭和13年（1938年）以降，国土交通省（当時の内務省土木局）により雨量観測および流量観測が実施され，これらのデータは雨量年表・流量年表として（社）日本河川協会より毎年刊行されている¹⁾．さらに近年，これらの貴重なデータは同協会により平成15年分まで電子化され，DVD媒体の「雨量・流量年表データベース」²⁾として98,000円(税込)で販売されている．これにより，例えば，これまで各研究者が独自に必要な日流量を流量年表より抽出し電子化していたが，この手間が省け作業能率は著しく向上したと言えよう．著者らも本データベースを活用し河川流量解析を行っていたが，その過程で明らかにおかしいと思われるデータが存在することが分かり，冊子体の流量年表¹⁾と比較したところ結構多くの相違点が見つかった．そこで，本研究では，さしあたり全国一級水系代表観測点のみを対象に流量年表データベースの日流量データ（以後，DVD日流量データと記す）を流量年表のそれと比較・精査することにより，DVD日流量データの入力ミス抽出しその成因について考察した．さらに，流量年表自体の日流量データの誤記についても精査・検証を行った．

2. 対象データ

本研究では109一級水系の各幹川における代表観測点として，基本的に河川ハンドブック³⁾の中で基準点として挙げられている観測点を採用している．しかし，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，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%以内の差であった場合同じ観測点とみなした⁴⁾．そして，各水系の代表観測点における日流量データをDVD流量年表より抽出した．

3. DVD日流量データの精査

DVD日流量データを精査するために流量年表のそれと比較した．まず，流量年表には日流量データのみならずそれらを月，年毎に合計した値も記載されているので，ここでは，DVD日流量データを累積した年間合計値が流量年表の年間合計値と一致していればその年のDVD日流量データは全て流量年表と同じであるとみなした．一方，年間合計値が一致しなかった場合，さらに各月合計値として比較し，月合計値が一致していない月についてのみ日流量データを一つ一つ見比べ，どの日流量データが異なっているのかを見つけ出した．この作業の結果，発見したDVD日流量データの流量年表との相違点を以下に述べる．

3-1 観測点名の誤記および位取りのカンマ記号

単純な観測点名の漢字入力ミスと思われる場合が見つかった．例えば，那賀川(徳島県)において1956年のみ「古庄」観測点が「古圧」と誤記されていた．また，大和川（大阪府）において1960年，61年の2年間「柏原」観測点名が「相原」となっていた．これら以外にも5水系において観測点名が年毎に不統一となっている場合が見られた．これらの入力ミスは，観測点名を入力してDVD 流量年表から日流量データを抽出する場合には大きな影響がある．

次に，流量年表データベースの出力形式には表形式とデータ登録形式の2つがあり，前者ではデータ抽出に非常に時間がかかるため，データ登録形式にてデータを出力しテキスト(csv)形式として抽出し，抽出した日流量データの流量時系列を作成したところ，例えば，2003年の石狩川水系石狩大橋の日流量時系列のように，流量が非常に大きい時期に，その前後の日に突如流量がほぼゼロとなる場合が多く出現した．そこで，抽出データを調べたところ，DVD日流量データには各数値間の区切りにカンマが使用され，CSV形式(カンマ区切り)で読み込むことができる一方，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，そのままDVD日流量データをデータ登録形式にて出力してデータを抽出すると，数値間のカンマと位取りのカンマの区別がされず，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．このため，DVD日流量データをデータ登録形式として用いる場合，位取りのカンマを消去するなどなんらかの処理が必要である．

3-2 1977年のデータ入力ミス

図-1に、1977年の利根川（八斗島観測点）における流量年表による日流量データおよびDVD日流量データによる流量時系列を示す。この図よりこの年の二つの河川流況が全く違っていることが分かる。精査した結果、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。本研究で対象とした109水系の代表観測点の中で、1977年以前より流量観測が開始されている観測点は106水系あるが、そのうち101水系で、DVD日流量データの1977年のデータには全く違う値が入っていることが判明した。例えば、石狩川では1977年のDVD日流量データはそれ以外の年に比べオーダーが全く異なるので、入力ミスに気づきやすいが、図-1のような場合は一見ただけではそれがおかしいかどうかは判断できないので注意が必要である。

3-3 単純な数値入力ミス

DVD日流量データを流量年表のそれと比較・精査した結果、上記以外に単純な数値入力ミスが数多く見つかった。入力ミス種別としては、 α)48.20(m³/s)が482.00(m³/s)のように小数点の位置を付け間違えた入力ミス、 β)149.70(m³/s)が19.70(m³/s)のように抜けてしまっている入力ミス、逆に1640.10(m³/s)が10640.10(m³/s)のように余計な数字が加わっている入力ミス、139.21(m³/s)が139.12(m³/s)のように数字が入れ替わっている入力ミス、さらに γ)88.22(m³/s)が38.22(m³/s)や42.14(m³/s)が12.14(m³/s)のように、3と8や1と4のように似通った数字の読み違いによる入力ミスも多く見つかった。これらの数値入力ミス（1977年を除く）は43水系で見つかり、入力ミス種別としては α)が2、 β)が23、 γ)が224（合計249個）で、似通った数字の読み違いによる入力ミスが圧倒的に多く見られた⁴⁾。これらは単純な入力ミスと考えられるが、これをそのまま用いると解析結果には大きな影響が現れることが示唆される。ただし、1977年を除く全データ数は1,809,683であるので、その入力ミス率としては0.014%と極めて小さい。

4. 流量年表日流量データの精査

流量年表には日流量データの数値の表のみならず、その月合計値や年合計値が記載され、さらに日流量時系列グラフも図示されている。そして、記載日流量データより算定した月合計値が記載月合計値と異なる場合が見つかり、この場合、日流量データを日流量時系列グラフと比較精査した。その結果、流量年表の日流量データにも明らかに誤記と考えられる数値が記載されていることが分かった。ここでは誤記かどうかよく分からない場合はそのままとし、明らかに誤記と考えられる場合についてのみ検証した結果、数値入力ミスは65水系で見つかり、記載ミス種別としてはDVD日流量データの場合と同様に分類したところ、 α)が2、 β)が26、 γ)が95（合計123個）となった。これよりDVD日流量データの場合と同様の数値入力ミスの傾向があることが分かった⁴⁾。なお、入力ミス率としては0.0067%である。

5. むすび

本研究では一級水系代表観測点におけるDVD流量年表および流量年表自体の日流量データの検証を行った。その結果、DVD日流量データには、観測点名の誤記や位取りのカンマ記号が挿入されていることが分かった他、1977年の大規模なデータ入力ミスおよび α)小数点のつけ間違い、 β)数字の抜け・加わり・入れ替わり、 γ)数字の読み間違いによる単純な数値入力ミスがあることが判明した。また流量年表自体についても、DVD日流量データ同様、単純な数値入力ミスが存在することが分かった。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、本研究よりまず解析の前にデータの確認・精査を行うことが重要であると言える。

参考文献

- 1) 国土交通省河川局編：昭和13年度～平成15年度雨量年表、流量年表、日本河川協会、1938-2003.
- 2) 日本河川協会：雨量・流量年表データベース、DVD版、2007.
- 3) 日本河川協会：2007河川ハンドブック、2007.
- 4) 岡元宏薫、梶井 剛、河村 明、天口 英雄：一級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、土木学会関東支部研究発表会講演集、CD-ROM版(II-035)、2009.

キーワード：流量年表、データベース、一級水系、日流量データ、検証

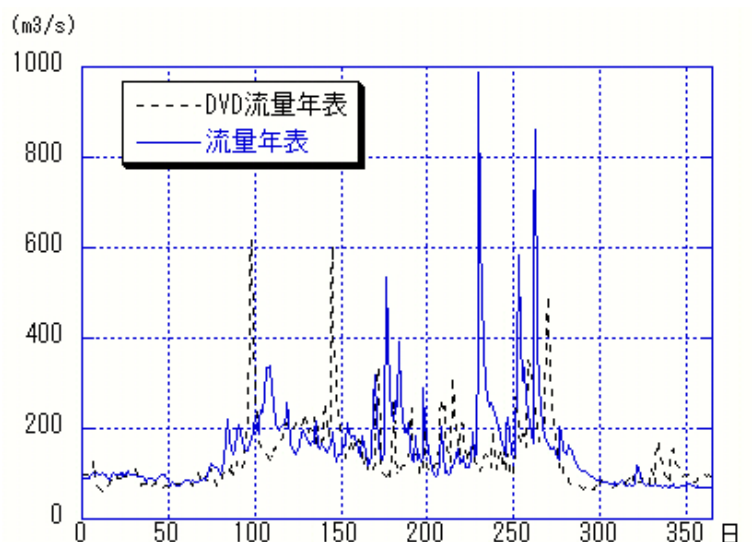


図-1 1977年利根川(八斗島観測点)日流量時系列の相違