

水文データにおける変動発見の研究
ON DETECTING CHANGES IN HYDROLOGIC DATA

中央大学 理工学部 正会員 林 泰造
中央大学 大学院 学生員○大橋正和

1. まえがき

実際に観測される水文データには、ランダムな変動成分の他に確定的と考えられる種々の変動成分が含まれている。一般に、それらは傾向変動成分と周期変動成分に分類される。河川においては、森林の伐採、河道改修、ダムの建設、流域の都市化等の原因により、観測される流量データの統計的性質に種々の影響を与えていることが予想される。また、降雨量データにおいても、気候変動等自然的な原因により統計的性質に種々の影響を受けていると思われる。そこで、本研究では、水文データにおける統計的性質に、それらがどのような影響を与えているかを考察し、その変動時点の発見の方法について検討を加えた。ここでは、統計的性質として時系列の平均を考え、ダム、都市化等の影響が表われるのは、年以上の時間単位によるであろうと考え、年データを用い解析を行なった。(図-1、図-2)

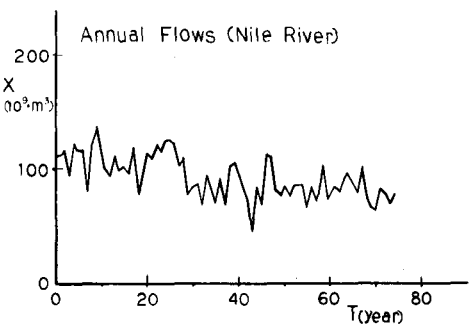


図-1

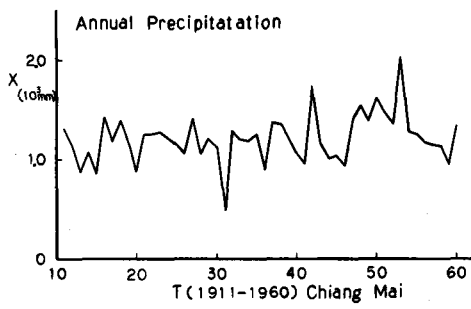


図-2

2. 従来の研究の review

変動が存在した場合の、累積分布関数、平均、分散、歪度、尖度への影響等は、Yevjevich ら (1969) により研究されている。その研究を、ここでは、簡単のため、変動として1段の Constant Jump の場合についての review を行なうこととする。

i) 定義

変動前と変動後の時系列を次式で表わす。

$$X_t = \varepsilon_t \quad (t \leq m) \quad (1)$$

$$X_t = \varepsilon_t + \delta \quad (t > m) \quad (2)$$

X_t : 非等質時系列, ε_t : 独立定常時系列

δ : Constant Jump, m : 変動時点

n : 標本数

図-3に離散的データの例を示す。($t = i$)

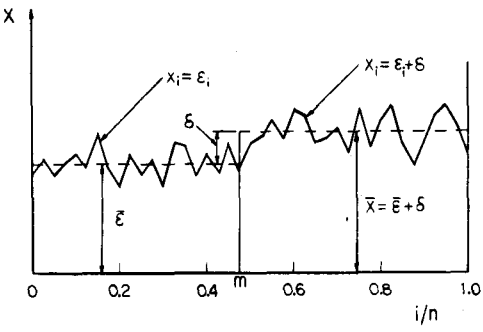


図-3

ii) 累積分布関数への影響

変動前と変動後の時系列が、累積分布関数Gを持つとする。ここで、 m' を変動後の標本数とする

$n = m + m'$ 。そこで次のような p, q を導入する: $p = m/n, q = m'/n, p + q = 1$ 。

全体の累積分布関数 $F(X)$ は、

$$F(X) = p G(X) + q G(X - \delta) \quad (3)$$

G が、 $N(0, 1)$ の正規分布に従い、 $\delta > 0$ と仮定すると、連続データに関する確率密度関数は、

$$f(X) = \frac{p}{\sqrt{2\pi}} e^{-\frac{X^2}{2}} + \frac{q}{\sqrt{2\pi}} e^{-\frac{(X-\delta)^2}{2}} \quad (4)$$

上式を X について微分すると、

$$f'(X) = \frac{1}{\sqrt{2\pi}} e^{-\frac{X^2}{2}} [-pX + q e^{\delta(X-\frac{\delta}{2})} (\delta - X)] \quad (5)$$

上式より $\frac{1}{\sqrt{2\pi}} e^{-\frac{X^2}{2}}$ が、0 にならないことを考慮して、 $f'(X) = 0$ の根を求めると、

$$q e^{\delta(X-\frac{\delta}{2})} (\delta - X) - pX = 0 \quad (6)$$

(6)式は、1根あるいは3根を持ち、解析の結果3根を持つ場合は、2つのピークを持つ確率密度関数となる。

水文データにおいて、2つのピークを持つ分布は、一般には、1) 流域が2つの気候区を持つ場合、2) 支流が合流した場合、3) 標本数が少ない場合で、変動による影響とこれらを混同する恐れがある。なお、3) は、標本数が多くなればピークが1つになるものと考えられる。

iii) 平均への影響

標本平均をスティルテス積分により表わすと、

$$\begin{aligned} \bar{X} &= \int_{-\infty}^{\infty} X dF(X) \\ &= \int_{-\infty}^{\infty} X (1-q) dG(X) + \int_{-\infty}^{\infty} X q dG(X - \delta) \\ &= (1-q) \bar{\epsilon} + q \bar{\epsilon} + q \delta \\ &= \bar{\epsilon} + q \delta \end{aligned} \quad (7)$$

iv) 分散への影響

$$\begin{aligned} \sigma_x^2 &= \int_{-\infty}^{\infty} (X - \bar{X})^2 dF(X) \\ &= \sigma_{\epsilon}^2 + (1-q) q \delta^2 \end{aligned} \quad (8)$$

v) 歪度への影響

$$\begin{aligned} \beta &= \frac{1}{\sigma_x^3} \int_{-\infty}^{\infty} (X - \bar{X})^3 dF(X) \\ &= \left(\frac{\sigma_{\epsilon}}{\sigma_x} \right)^3 \beta_{\epsilon} + (2q-1)(q-1) q \alpha^3 \end{aligned} \quad (9)$$

β : 歪度, β_{ϵ} : 変動前の歪度, $\alpha = \frac{\delta}{\sigma_{\epsilon}}$

$$\begin{aligned} \Delta \beta &= \beta - \beta_{\epsilon} \\ &= \frac{1}{\sigma_x^3} \{ \beta_{\epsilon} (\sigma_{\epsilon}^3 - \sigma_x^3) + (2q-1)(q-1) q \delta^3 \} \end{aligned}$$

正規分布に従うデータは、有意な歪度を持たないという性質 ($\beta = 0$) に着目して、Hipelら (1977) は、正規性をテストするために、歪度テストを行なった。もし変動により、歪度に対する影響が大きい場合には、上記のテストにより、誤った結論を導くことになる。

vi) 尖度への影響

歪度と同様に、尖度の性質 ($r = 3$) は、正規性のテストに用いられる。(Hipelら (1977))
そこで、尖度への変動の影響を求めると

$$\begin{aligned} r &= \frac{1}{\sigma_x^4} \int_{-\infty}^{\infty} (X - \bar{X})^4 dF(X) \\ &= \frac{1}{\sigma_x^4} \{ \sigma_\varepsilon^4 r_\varepsilon + 6(1-q)q\delta^2\sigma_\varepsilon^2 - q\delta^4(q-1)(3q^2-3q+1) \} \end{aligned} \quad (10)$$

r : 尖度, r_ε : 変動前の尖度

$$\Delta r = r - r_\varepsilon$$

$$= \frac{1}{\sigma_x^4} \{ r_\varepsilon(\sigma_\varepsilon^4 - \sigma_x^4) + (1-q)q\delta^2[6\sigma_\varepsilon^2 + (3q^2-3q+1)\delta^2] \}$$

上記により、尖度についても、歪度の時と同様に Constant Jump を含んだデータの尖度を用いて正規性をテストする場合、誤った結論を導く恐れがある。また、その様なデータを用い、種々の解析を行なう場合、原時系列の統計的性質の等質性に留意する必要がある、そのために変動成分の発見が、必要となる。

3. 変動成分の発見方法

降雨量データや流量データに、自然や人工的な原因により統計的性質に変動が存在するか否かを発見するために、観測されたデータを Lee and Heghinian (1977) により提案された方法により解析を行なった。

i) 仮定

$X_1, X_2, \dots, X_n$ を確率変数とし、変動の時点点を τ とすると、

$$X_i = \mu + \varepsilon_i, \quad i = 1, 2, \dots, \tau \quad (11)$$

$$X_i = \mu + \delta + \varepsilon_i, \quad i = \tau + 1, \dots, n \quad (12)$$

ここで、 ε_i は、平均 0、分散 σ^2 の独立な正規分布に従う確率変数とし、 τ と $\tau + 1$ の間で変動が起こったと考える。 μ は、変動前の時系列の平均、 δ は、変動量、 σ を標準偏差とすると、 $1 \leq \tau \leq n-1$ 、 $-\infty < \mu < \infty$ 、 $-\infty < \delta < \infty$ 、 $\sigma > 0$ となり、 $\tau, \mu, \delta, \sigma$ は、変動の時点 τ を未知とすると、全て未知のパラメータで、互いに独立とする。

また、 $\tau, \mu, \delta, \sigma$ の事前分布を次の様に仮定する。

$$1) f(\tau) = \frac{1}{n-1}, \quad \tau = 1, 2, \dots, n-1$$

$$2) f(\mu) \text{ は、平均 } 0, \text{ 分散 } \sigma_\mu^2 \text{ の正規分布に従う}$$

$$3) f(\delta) \text{ は、平均 } 0, \text{ 分散 } \sigma_\delta^2 \text{ の正規分布に従う。}$$

$$4) f(\sigma) \propto \sigma^{-1}$$

ii) τ と δ の結合事後分布

上記の仮定より、観測値、 $x_1, x_2, \dots, x_n$ が、得られた時、Bayes の定理を用い、 τ, δ, σ の結合事後確率関数は Lee and Heghinian (1977) によりつぎのように求められている。

$$f(\tau, \delta, \sigma | X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) \propto \sigma^{-n} \exp \left[-1/(2\sigma^2) \right] \{ Q(\tau, \delta) + \sigma^2 \delta^2 / \sigma^2 \} \quad (13)$$

ここで、 $Q(\tau, \delta) = H(\tau) + \tau(n-\tau)(\delta - \hat{\delta}_\tau)^2 / n$

$$\hat{\delta}_\tau = \bar{x}_{n-\tau} - \bar{x}_\tau$$

$$= \frac{1}{n-\tau} \sum_{i=\tau+1}^n x_i + \frac{1}{\tau} \sum_{i=1}^{\tau} x_i$$

$$H(\tau) = \sum_{i=1}^{\tau} (x_i - \bar{x}_\tau)^2 + \sum_{i=\tau+1}^n (x_i - \bar{x}_{n-\tau})^2$$

$$\frac{\sigma^2}{n} \ll \sigma_\mu^2$$

次に、 $H(\tau)$ を書き換えると、 $H(\tau) = \sum_{i=1}^{\tau} (x_i - \bar{x}_\tau)^2 - \tau(n-\tau) \hat{\delta}_\tau^2 / n$ 。次に $\sigma_\delta^2 \gg \sigma^2$ とすれば、

$$f(\tau, \delta, \sigma | X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) \propto \sigma^{-n} \exp \left[-Q(\tau, \delta) / (2\sigma^2) \right] \quad (14)$$

上式より、 τ と δ の結合事後確率関数を求めると

$$f(\tau, \delta | X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) \propto [Q(\tau, \delta)]^{-(n-1)/2} \quad (15)$$

iii) τ の事後分布

τ の事後確率関数を(15)式より求めると、

$$f(\tau | X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) \propto [n / (\tau(n-\tau))]^{1/2} [H(\tau)]^{-(n-2)/2} \quad (16)$$

上式を書き換えると、

$$f(\tau | X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) \propto [n / (\tau(n-\tau))]^{1/2} [R(\tau)]^{-(n-2)/2} \quad (17)$$

ここで、 $R(\tau) = 1 - \tau(n-\tau) \delta_\tau^2 / (n(n-1) S_n^2)$, $S_n^2 = \sum_{i=1}^n (x_i - \bar{x}_n)^2 / (n-1)$

iv) δ の事後分布

$$f(\delta | \tau, X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) =$$

$$f(\tau, \delta | X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) / f(\tau | X_1 = x_1, X_2 = x_2, \dots, X_n = x_n)$$

と表わせる。よって、 δ の事後分布は、

$$f(\delta | X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) = \sum_{\tau=1}^{n-1} f(\delta | \tau, X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) \cdot f(\tau | X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) \quad (18)$$

よって、(17)式により τ の分布が求まり、(18)式により δ の分布が表わされる。

4. 適用例

以上の方法を、流量データと降雨量データに適用して変動成分の解析を行ない同時に従来の方法との比較を試みた。データとしては、Nile 河年流量データ(1870年-1945年)(図-1)タイ北部 Chiang Mai に於ける年降雨量データ(1911年-1960年)(図-2)を用いた。

a-1) Nile 河年流量データ

エジプトの Aswan における Nile 河の 1870 年 10 月 1 日から 1945 年 10 月 1 日までの 75 年間の年流量データにおける変動成分の発見を試みた。(17) 式により分布 $f(\tau | X_1 = x_1, X_2 = x_2, \dots, X_n = x_n)$ を求めると図-4 の様になる。この結果分布より τ の値を求めると、

分布の平均による : $\tau = 27.899$

分布のピークによる : $\tau = 28.000$

$$(f(\tau | X) = 0.789)$$

95% の信頼区間 : $25.567 \leq \tau \leq 29.815$

次に $f(\delta | X)$ の分布は、図-5 となる。

分布の平均による : $\delta = -27.0 \quad (10^9 \cdot m^3)$

95% の信頼区間 : $-34.4 \leq \delta \leq -19.6$

以上の結果より時系列全体として $\tau = 28$ すなわち 1898 年から 1899 年にかけて流量の平均が $27.0 \quad (10^9 \cdot m^3)$ 減少する変動があったことが発見された。特に、変動点の推定が、分布の平均による値とピークによる値がほぼ同じであり 95% の信頼区間より分布が平均の回りに集中していると考えられるので、この変動は constant jump と考えられる。Nile 河の歴史によれば、1899 年から Aswan においてダムの工事が開始され 1903 年に完成、運用が始まっている。このためこの変動は、人工的な影響によるものと考えられる。

a-2) t-テストによる検証

変動の時点を既知として変動前の時系列と変動後の時系列の平均が等しいか否かを検定する方法として t-テストが従来からよく用いられている。そこで、Nile 河についての上記の結果について検証のため t-テストを行なった。

$$t = \frac{\bar{X}_1 - \bar{X}_2}{S \sqrt{\frac{n_1 + n_2}{n_1 \cdot n_2}}} \quad (19)$$

$$S = \sqrt{\frac{n_1 \sum_{i=1}^{n_1} (X_i - \bar{X}_1)^2 + n_2 \sum_{j=1}^{n_2} (X_j - \bar{X}_2)^2}{n_1 + n_2 - 2}}$$

$$n_1 = \tau, \quad n_2 = n - (\tau + 1)$$

上式により求めた統計量 t は、Student の t 分布に従う。よって、Lee らの方法で求めた τ の値について有意水準を設定し、2つの時系列について平均が等しいという仮説検定を行なう。Nile 河のデータについて、Hipel ら (1975) の研究によると変動の時点を 1902 年から 1903 年の間、すなわちダムの運用が開始された年と考察しているので、2つの変動時点 $\tau = 28$ と $\tau = 32$ についても t-テストを行なっ

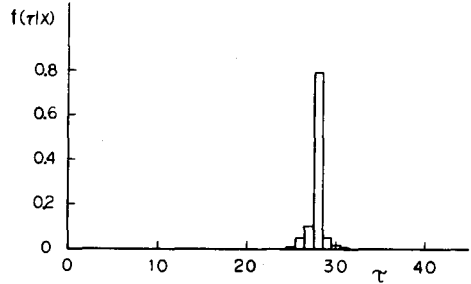


図-4

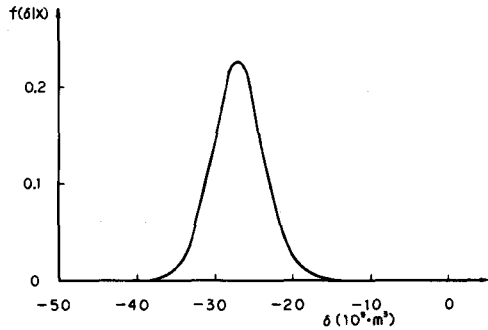


図-5

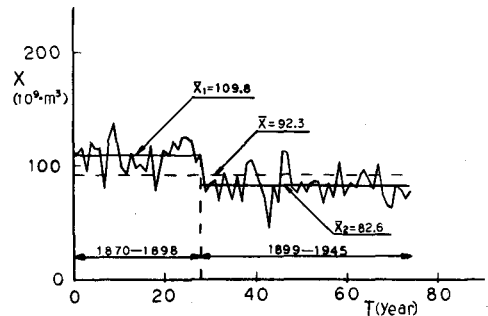


図-6

た。 $\tau = 28$ $t = 8.643$ (有意水準 95% : $t = 1.668$)

$$\bar{X}_1 = 109.8$$

$$\bar{X}_2 = 82.6$$

$$\delta = -27.2$$

$\tau = 32$ $t = 6.725$ (有意水準 95% : $t = 1.668$)

$$\bar{X}_1 = 106.0$$

$$\bar{X}_2 = 82.9$$

$$\delta = -23.1$$

上記の結果より $\tau = 28, 32$ の両方とも仮説は棄却され、平均が異なることを示している。しかし、 t の値を比較すると $\tau = 28$ の方が大きく、変動の時点は、1898年～1899年の間と考えることが妥当であり、変動量についても Lee らの方法によって得られた値とほぼ等しいと考えられる。(図-6)

b-1) Chiang Mai 年降雨量データ

次に変動に対して気象変動等、人工的な影響より自然の影響が大きいと考えられる降雨量データについて解析を行なった。データとしてタイ北部の Chiang Mai にける 50 年間 (1911 年から 1960 年) の年降雨量データを採用した。Lee らの方法による解析の結果、 τ の分布は、図-7 の様になった。 τ の分布より、分布の平均による : $\tau = 30.1$

分布のピークによる : $\tau = 36.0$

$$(f(\tau|X) = 0.194)$$

95% の信頼区間 $2.4 \leq \tau \leq 47.8$

次に δ について分布を求めると図-8 となる。 δ の分布より、分布の平均による : $\delta = 0.144$

95% の信頼区間 $-0.146 < \delta < 0.348$

変動時点 τ について、分布の平均より求めた値と分布のピークにより求めた値とが大きく異なっている。これは、1940 年から 1946 年にかけて緩やかに降雨量が増加したためと思われる。

b-2) t-テスト, Double Mass Curve による

検証

○ t-テスト 上記の結果を参考にして $\tau = 30, \tau = 36$ について、t-テストを行なうと、

$\tau = 30, t = 2.045$ (有意水準 95% $t = 1.679$)

$$\bar{X}_1 = 1.159$$

$$\bar{X}_2 = 1.299$$

$$\delta = 0.140$$

$\tau = 36, t = 2.928$ (有意水準 95% $t = 1.679$)

$$\bar{X}_1 = 1.156$$

$$\bar{X}_2 = 1.366$$

$$\delta = 0.210$$

有意水準 95% では、 $\tau = 30, \tau = 36$ の両時点とも仮説が棄却されるが、有意水準 99% とすると

$t = 2.410$ となり、 $\tau = 36$ の場合のみ仮説が棄却

される。よって、変動の時点は、 $\tau = 36$ すなわち

1946 年から 1947 年の間と考えるのが適当である。(図-9)

○ Double Mass Curve による解析

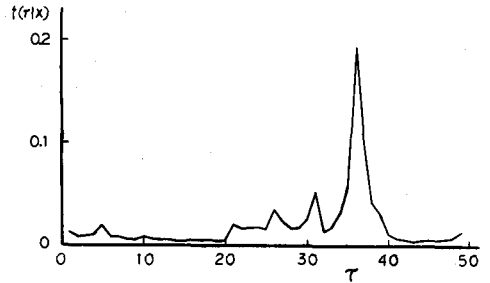


図-7

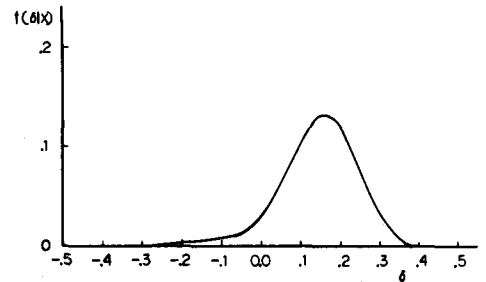


図-8

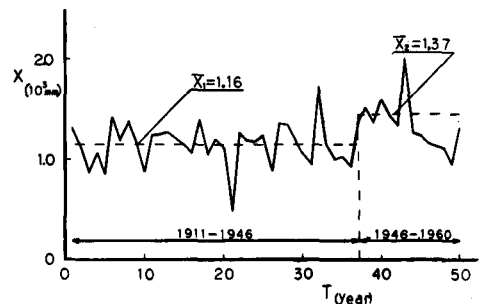


図-9

一般に降雨量データは、一地点で観測されるばかりでなく近くにいくつかの観測点が存在し、複数のデータが得られる場合が多い。そこでChiang Mai の降雨量データと近隣の7つの観測点によるデータの平均についてDouble Mass Curve を求めた。

$x_1, x_2, \dots, x_n$: Chiang Mai の降雨量データ

$y_1, y_2, \dots, y_n$: 他の7つの観測点の降雨量データの平均

$S_1 = x_n, S_2 = x_n + x_{n-1}, \dots, S_n = x_n + \dots + x_1$

$S'_1 = y_n, S'_2 = y_n + y_{n-1}, \dots, S'_n = y_n + \dots + y_1$

上記より求めた $(S'_1, S_1), (S'_2, S_2), \dots, (S'_n, S_n)$

を対としてプロットする。7つの観測点の平均データについては、Leeらの方法とt-テストにより有意な変動が存在しないことを確かめた。(図-10, 11)

Double Mass Curve 図(図-12)より変動成分の解析を行なうと1931年のデータは誤まった観測値であると考えられる以外、大きな変動は発見されなかった。以上の結果、この例では、Double Mass Curve による解析は、誤測データの発見には適切であるが、変動成分の発見方法としては精度がよくないと考えられる。この理由として、降雨量データにおける変動傾向は、1地点に影響を与えるのみでなく、近くの他の地点にも及ぶと考えられる。基準として7つの観測点の平均を用いたため各地点毎の傾向が弱められはしたが、有意でない程度に存在し、そのため互いに干渉し合ったものと思われる。

。この様な干渉を防ぐためには、より多地点のデータを基準データとして用いる必要があると思われる。

5. 結び

Nile 河年流量データとChiang Mai における年降雨量データに種々の方法により変動成分の発見を試みた結果Leeらの方法の精度が高く、未知の変動点の発見が可能であり、変動が緩やかに起こる場合も τ の分布の広がりにより発見可能であると考えられる。しかし、原データに周期的変動が存在する場合周期の振幅を変動と見なし

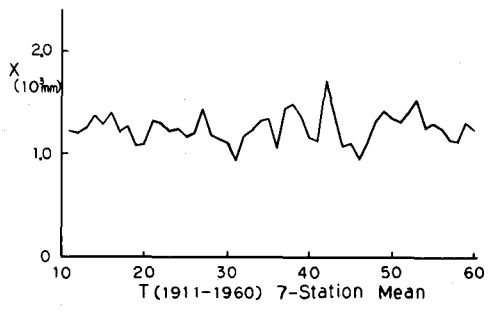


図-10

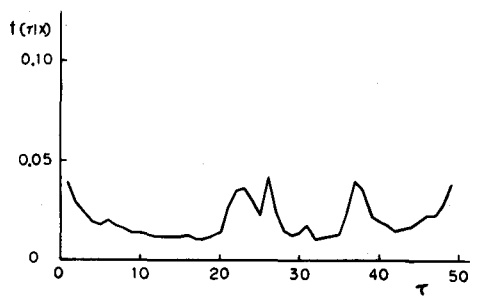


図-11

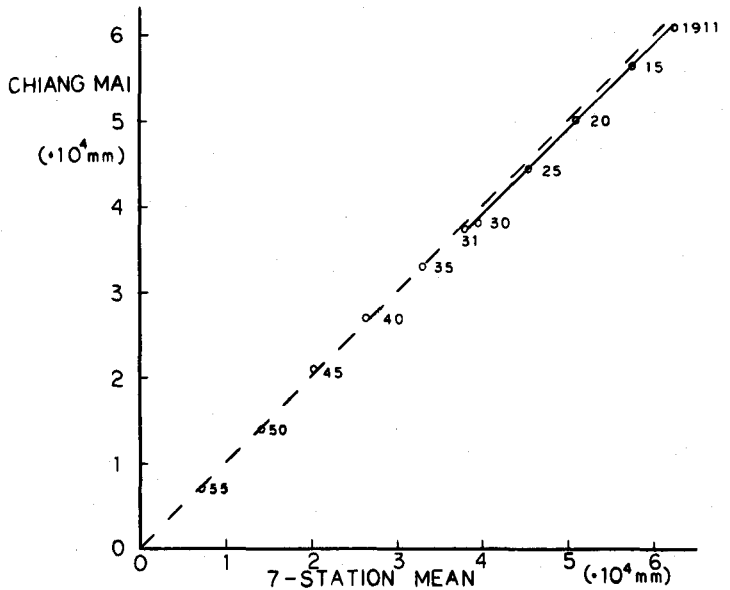


図-12

てしまう可能性もあり、テストを行なう前に周期的な変動成分が存在するか否かを確かめる必要があると思

われる。また、変動の時点を決断するためには、Leeらの方法を軸として種々のテスト方法を併用し、人工的影響等も含めた歴史的考察も加えつつ解析を行なう必要があると思われる。逆に、河道改修や、ダム等を建設した場合、流量データにどのような影響を及ぼしたか、あるいは都市化等による降雨量データへの影響等を調査する事も可能である。特に、定常時系列解析モデルを用いて種々の予測等を行なう場合は、原データの統計的等質性について十分な考察を加える必要があると思われる。将来、このような研究を基礎として原データの変動成分を総合的に把握し、欠測データや誤測データの補間や、種々の予測における精度向上に役立てたいと考えている。

。謝辞

本研究において、資料を提供して頂き、適切な御指導を頂いた、コロラド州立大学統計学科 D.C. Boes 準教授、前中央大学数学科 高島 己千雄教授に心からの謝意を表する。

参考文献

- 1) Yevjevich, V. and Jeng, R. I. : Properties of Non-Homogeneous Hydrologic Time Series, Colorado State University Hydrology Paper No. 32, April, 1969.
- 2) Yevjevich, V. : "Stochastic Processes in Hydrology", Water Resources Publications, 1972.
- 3) Lee, A. F. S. and Heghinian, S. M. : A Shift of The Mean Level in A Sequence of Independent Normal Random Variables - A Bayesian Approach -, Technometrics, Vol. 19, No. 4, 503 - 506 November, 1977.
- 4) Hipel, K. W., Lennox, W. C., Unny, T. E. and McLeod, A. I. : Intervention Analysis in Water Resources, Water Resources Research, Vol. 11, No. 6, 855-861, December, 1975.
- 5) Hipel, K. W., McLeod, A. I. and Lennox, W. C. : Advances in Box - Jenkins Modeling I. Model Construction, Water Resources Research, Vol. 13, No. 3, 567-575, June, 1977.
- 6) Hinkley, D. V. : Inference about The Change-Point in a Sequence of Random Variables, Biometrika, 57, 1, 1-17, 1970.
- 7) Chang, M. and Lee, R. : Objective Double-Mass Analysis, Water Resources Research, Vol. 10, No. 6, 1123-1126, December, 1974.
- 8) Sen, A. and Srivastava, M. S. : On Tests for Detecting Change in Mean, The Annals of Statistics, Vol. 3, No. 1, 98-108, 1975.
- 9) Boes, D. C. : Detection of Changes in Hydrologic Data, Lecture Notes for the Computer Workshop in Statistical Hydrology, Colorado State University, July, 1978.
- 10) Box, G. E. P. and Tiao, G. : A Change in Level of a Non-Stationary Time Series, Biometrika, 52, 1 and 2, 181-192, 1965.
- 11) Kitanidis, P. K. and Bras, R. L. : Real Time Forecasting of River Flows, MIT Ralph M. Parsons Lab. Report No. 235, June, 1978.
- 12) 林 泰造・大橋正和 : 水文時系列における季節変動成分の解析について, 第6回関東支部年次研究発表会, 1979.