

降水観測がもたらす情報量とその観測網配置計画への応用

THE INFORMATION CONTENT YIELDED BY THE OBSERVATIONS OF
PRECIPITATION AND ITS APPLICATION TO THE
OPTIMAL GAGING SCHEME

高 棹 琢 馬*・池 淵 周 一**

By Takuma TAKASAO and Shuichi IKEBUCHI

1. はし が き

水文資料の観測・収集は、元来地味であり、かつ長年月を要するが、得られたデータの豊富さ、あるいは精度のよさは、治水・利水計画を問わず、実際の設計・操作問題においても直接的にその合理性・経済性を大きく左右するものであり、その必要性・重要性はいうまでもない。

水問題で最も基本となる水文資料は降水であり、その空間的・時間的不確定構造を考えると、降水資料の観測・収集に関しては、最適な観測網体系の確立が最重要課題であろう。

ところで、流域にいかなる観測網を設けるかという問題は以下のような理由により、いまだ十分とは解明されていない。すなわち、1) 降水の気象的・地形的因果関係が明確にされていない、2) 観測データの情報価値が定量化されていない、3) 流出解析においては、降水の地域的分布は機械的に面積雨量に変換されることが多く、それに基づく計算ハイドログラフと実測ハイドログラフとの合致精度に解析の重点がおかれがちである。

もちろん、降水変動は流域で平滑化されて流量となるので、3) の理由はいくぶんゆるめられるが、流出システムは本来分布定数系であり、今後の流出解析が分布型で展開されていく以上、降水に関しても面積雨量を求め空間スケールの設定、地域的分布特性の導入は不可欠であり、その意味からも降水観測網の確立が必要となる。

従来、この分野でなされた研究には、面積雨量誤差と観測点数、面積雨量誤差と雨量分布の標準偏差・観測点数の関係から目的にあった誤差範囲内で観測点数を決める方法^{1),2)}、および Fiering³⁾、Matalas⁴⁾ の研究があ

る。しかし、いずれも既設観測所で得られたデータのみに基づく解析であり、加えて前者においては、既設観測所のすべての観測所を精密な観測網と考え、観測所を減らす方向で議論しており、そこには経済効果や観測所のもつ時間的価値（観測の長短）、降水の空間的従属性などが組み込まれていない。一方、後者は雨量平均値の推定精度に関する情報量を Fisher 流の情報量、すなわち分散の逆数で表現し、分散和最小（総情報量最大）を最適化基準として、観測所の継続・廃止を 0-1 整数計画法で展開した注目すべき研究であるが、対象が通常正規分布を示す年降水量だけであり、しかも分散既知とした平均値推定精度のみに議論が集中している。

本論文では、従来の研究成果をふまえるとともに、観測の行動目的を直接、未知パラメータの推定精度向上におかず、確率的性格の強い日単位以上の降水量の空間的・時間的実現値を知ることにおいて、その情報量を定式化し、予算の許す限り観測所をふやす、過去のデータが内蔵する情報蓄積を最大限利用する、観測所の継続・廃止・新設を同時に考慮した O.R. 手法を導入する、といった立場を重視して、情報量概念に基づく降水観測網の最適配置基準を設定しようとするものである。すなわち、今後の観測のもたらす情報量は、その観測による実現値を知る以前にわれわれがもつ不確かさの減少の尺度と解釈して Shannon 流の情報量で定義するとともに、理論の展開にあたっては事前分布を特徴とするベイズ論的立場をとったために、あいまいであった情報価値なる概念も、ここに定義した情報量によって適切かつ計量的に評価できると考えている。

以下、本論文の内容を概説すると、まず 2. 章では情報量を不確かさの測度を減少してくれるものと規定し、その数式的表現を展開するとともに、事前分布とベイズ論的立場を述べ、3. 章では既設観測所を継続するとき、既設観測所を廃止し、基幹観測所から今後の降水量を推定するとき、および観測所を新設するときに行われる情

* 正会員 工博 京都大学教授 工学部土木工学科

** 正会員 工博 京都大学助教授 工学部土木工学科

報量を観測資料が正規分布および指数分布にしたがう場合について算定する。4. 章ではこれらの情報量の定義および式に基づき、観測網の配置計画の基本方針および最適化への手順を示し、最後に 5. 章ではこれらの理論的アプローチを琵琶湖流域に適用し、その結果について考察する。

2. 情報量の一般的表現法

情報という概念は非常に広義なものとして理解されているが、本論文では不確かさの測度を減少してくれるものと規定したい。つまり、情報は何か不確かなものに関する情報としてのみ意味をもつものであり、確実に生起することがわかっているものに関しては情報というものとは存在しないはずである。確率的性格の強い降水現象に対して、その観測がもたらす情報量はまさにこの規定そのものであり、ある降水事象を知ることにより得られる情報量として、

$$[\text{情報量}] = [\text{事前の不確かさ}] - [\text{事後の不確かさ}] \quad \dots\dots\dots(2.1)$$

なる式が成立することが理解できよう。

(1) 情報量の理論的表現法

一般に、不確かさといっても、その程度にいろいろの差があるものと直感的に考えられるので、不確かさの程度を量的に規定するために、われわれは不確かさということ、ある確率変数 x の実現値が何になるかあらかじめわからないという意味に解釈し、その不確かさの測度 $H(x)$ を、

$$H(x) = - \int f(x) \log f(x) dx \quad \dots\dots\dots(2.2)$$

なる関数で表現することにする。ここに、 $f(x)$ は x の生起確率密度関数であり、 $H(x)$ は通常、エントロピーとよばれている。

次に、確率変数 x, y の結合確率密度関数を $f(x, y)$ とすると、この結合確率密度関数に関する結合エントロピーは式 (2.2) と同様に、

$$H(x, y) = - \iint f(x, y) \cdot \log f(x, y) dx dy \quad \dots\dots\dots(2.3)$$

で表わされる。

最後に、 x の実現値を知ったときに y の実現値が何になるかあらかじめわからないという不確かさ、すなわち条件付エントロピーは、条件付確率密度関数

$$f(y|x) = f(x, y) / f(x) \quad \dots\dots\dots(2.4)$$

を用いて、

$$H(y|x) = - \iint f(x, y) \log f(y|x) dx dy \quad \dots\dots\dots(2.5)$$

で与えられる。

以上のようにして、不確かさの測度としてのエントロピーが求められると、式 (2.1) より情報量が算定されることになるが、これらは Shannon 流の情報量である。情報量には、この他に Kullback 情報量、その極限表示である Fisher 情報量があるが、いずれも未知パラメータの推定精度に関する情報量表現であり、本論文の観測目的とやや異なるので、以下では Shannon 流情報量を用いることにする。

まず、 x の実現値を知ることにより得られる情報量を求める。この場合、 x の実現値を知る以前の x に関する事前の不確かさは式 (2.2) の $H(x)$ で与えられ、事後の不確かさは x の実現値を知ったのであるから 0 となる。したがって、 x の実現値を知ることにより得られる情報量 $I(x)$ は、

$$I(x) = H(x) \quad \dots\dots\dots(2.6)$$

で与えられる。

次に、 y の実現値はわからないが、 x の実現値を知ることにより y に関して得られる情報量を求める。この場合、 x の実現値を知る以前の y に関する事前の不確かさは $H(y)$ であり、 x の実現値を知ることにより y に関して残る事後の不確かさは、式 (2.5) の $H(y|x)$ で与えられる。したがって、 x の実現値を知ることにより y に関して得られる情報量 $I(y|x)$ は、

$$I(y|x) = H(y) - H(y|x) \quad \dots\dots\dots(2.7)$$

で与えられる。

(2) 事前分布とベイズ論的立場

式 (2.6), (2.7) からわかるように、実際に情報量を算定するためには、確率変数 x, y の周辺分布あるいは条件付周辺分布を知らなければならない。ところで、分布にはそれを特徴づけるパラメータ (母数) があり、われわれはその未知パラメータの推定に関心がある。伝統的にはその推定に最尤法や積率法が用いられるが、これらはいずれも過去のデータのみに基づいているので推定誤差はまぬがれない。そこで、この誤差を少しでも減らしうるように、データ以外の利用できる情報をも付加しようとするのがベイズ論的立場であり、この付加的な情報が事前分布に凝縮されている。

本論文が対象とする降水量においても、このベイズ論的立場にたつならば、過去に観測データがある場合にはこの情報蓄積効果が考慮でき、まったく観測データがない場合にも、すでに得られている流域、地域あるいは世界的規模でのデータ、理論的成果および研究者の経験・判断などに基づいた事前分布が導入できれば、未知パラメータの精度が観測によってつぎつぎと高められていく効果をも期待できよう。

以上の観点から、本論文ではベイズ論的立場にたつて情報量を算定する。いま、未知パラメータ θ の事前確率分布を $\xi(\theta)$ とすると、観測 x_1 が与えられての θ の事後確率分布 $\xi(\theta|x_1)$ は、ベイズの定理により次式で与えられる。

$$\xi(\theta|x_1) = \xi(\theta) \cdot f(x_1|\theta) / \int \xi(\theta) \cdot f(x_1|\theta) d\theta \dots\dots\dots(2.8)$$

次に、 x_2 が観測されたならば、 $\xi(\theta|x_1)$ を事前確率分布とみなして、 x_1, x_2 が観測された後の事後確率分布 $\xi(\theta|x_1, x_2)$ は、

$$\begin{aligned} \xi(\theta|x_1, x_2) &= \xi(\theta|x_1) \cdot f(x_2|\theta) / \int \xi(\theta|x_1) \cdot f(x_2|\theta) d\theta \\ &= \xi(\theta) \cdot f(x_1|\theta) \cdot f(x_2|\theta) / \int \xi(\theta) \cdot f(x_1|\theta) \cdot f(x_2|\theta) d\theta \dots\dots\dots(2.9) \end{aligned}$$

となる。ただし、 x_1, x_2 は互いに独立であるとし、 $f(x|\theta)$ は θ が真のパラメータの値であるとしたときの x の確率分布であり、既知であるものとする。

このようにして、 x_1, x_2, \dots, x_n が観測されるごとに、事前確率分布を事後確率分布に修正していけば、観測数が大きくなるにしたがって、たとえ事前確率分布が正確な形で与えられなくても、事後確率分布は真のパラメータの値に確率1を与えるように近づいていく。

したがって、将来観測にあたっては、過去にデータが豊富であれば、式(2.9)を繰返すことによって算出される事後確率分布が事前確率分布となり、それから導かれる周辺確率密度関数 $f(x)$ にしたがう確率変数 x のいずれかが実現することを観測するわけである。データがない場合でも、なんらかの形で事前確率分布を与えれば、同様の考えで確率変数 x を観測することになる。

3. 観測所を継続する・廃止後回帰推定する・新設することによって得られる情報量

ところで、観測所配置計画にあたっては、1) 既設観測所を継続する、2) 既設観測所を廃止し、基幹観測所あるいは周辺観測所から推定する、3) 観測所を新設する、のいずれかの行動を選択し、その行動に伴って得られる情報量とその行動に要する費用とのバランス化が指向されよう。

本章では前章で述べたことを基礎とし、それぞれの行動に伴って得られる情報量をベイズ論的立場から算定する。さらに、情報源から生じる情報、つまり観測資料が正規分布および指数分布にしたがうときに得られる情報量を求める。なお、以下では情報源から情報が発生する過程をエルゴード的独立過程と仮定している。

(1) 既設観測所を継続することにより得られる情報量

a) 一般的定義式

既設観測所において、これまでの観測により得られている情報(観測値)を $\mathbf{X}(x_1, x_2, \dots, x_N)$ とすると、情報 \mathbf{X} を得た後の未知パラメータ θ の事後確率分布 $\xi(\theta|\mathbf{X})$ は、式(2.9)からわかるように、

$$\begin{aligned} \xi(\theta|\mathbf{X}) &= \xi(\theta) \cdot f(x_1|\theta) \cdot f(x_2|\theta) \dots f(x_N|\theta) / \int \xi(\theta) \cdot f(x_1|\theta) \cdot f(x_2|\theta) \dots f(x_N|\theta) d\theta \dots\dots\dots(3.1) \end{aligned}$$

で与えられる。 $\xi(\theta|\mathbf{X})$ が求められると、 \mathbf{X} および将来観測値 $\mathbf{X}^*(x_{N+1}, x_{N+2}, \dots, x_{N+n-1})$ が与えられての x_{N+n} の周辺密度関数 $f(x_{N+n}|\mathbf{X}, x_{N+1}, \dots, x_{N+n-1})$ は、

$$\begin{aligned} f(x_{N+n}|\mathbf{X}, x_{N+1}, \dots, x_{N+n-1}) &= \int \xi(\theta|\mathbf{X}) \cdot f(x_{N+1}|\theta) \dots f(x_{N+n}|\theta) d\theta / \int \xi(\theta|\mathbf{X}) \cdot f(x_{N+1}|\theta) \dots f(x_{N+n-1}|\theta) d\theta \dots\dots\dots(3.2) \end{aligned}$$

で表わされ、さらに \mathbf{X} が与えられての $\mathbf{X}^*(x_{N+1}, \dots, x_{N+n})$ の結合密度関数 $f(x_{N+1}, \dots, x_{N+n}|\mathbf{X})$ は、

$$\begin{aligned} f(x_{N+1}, \dots, x_{N+n}|\mathbf{X}) &= \int \xi(\theta|\mathbf{X}) \cdot f(x_{N+1}|\theta) \dots f(x_{N+n}|\theta) d\theta \dots\dots\dots(3.3) \end{aligned}$$

で与えられる。ここに、式(3.2)、(3.3)より周辺密度関数 $f(x_{N+n}|\mathbf{X}, x_{N+1}, \dots, x_{N+n-1})$ と結合密度関数 $f(x_{N+1}, \dots, x_{N+n}|\mathbf{X})$ との間には、次の関係が成立する。

$$\begin{aligned} f(x_{N+n}|\mathbf{X}, x_{N+1}, \dots, x_{N+n-1}) &= f(x_{N+1}, \dots, x_{N+n}|\mathbf{X}) / f(x_{N+1}, \dots, x_{N+n-1}|\mathbf{X}) \dots\dots\dots(3.4) \end{aligned}$$

さて、 $\mathbf{X}(x_1, \dots, x_N)$ および $\mathbf{X}^*(x_{N+1}, \dots, x_{N+n-1})$ を知ったとき、 x_{N+n} のもつエントロピー、 $H'(x_{N+n}|\mathbf{X}, x_{N+1}, \dots, x_{N+n-1})$ は式(2.2)より、

$$\begin{aligned} H'(x_{N+n}|\mathbf{X}, x_{N+1}, \dots, x_{N+n-1}) &= - \int f(x_{N+n}|\mathbf{X}, x_{N+1}, \dots, x_{N+n-1}) \cdot \log f(x_{N+n}|\mathbf{X}, x_{N+1}, \dots, x_{N+n-1}) dx_{N+n} \dots\dots\dots(3.5) \end{aligned}$$

となるが、現時点においては $\mathbf{X}^*(x_{N+1}, \dots, x_{N+n-1})$ が観測されていないから、 x_{N+n} のもつエントロピーは期待値 $H(x_{N+n}|\mathbf{X}, x_{N+1}, \dots, x_{N+n-1})$ として、

$$\begin{aligned} H(x_{N+n}|\mathbf{X}, x_{N+1}, \dots, x_{N+n-1}) &= - \int \dots \int f(x_{N+1}, \dots, x_{N+n-1}|\mathbf{X}) \cdot f(x_{N+n}|\mathbf{X}, x_{N+1}, \dots, x_{N+n-1}) \cdot \log f(x_{N+n}|\mathbf{X}, x_{N+1}, \dots, x_{N+n-1}) dx_{N+1} \dots dx_{N+n} \\ &= - \int \dots \int f(x_{N+1}, \dots, x_{N+n}, \mathbf{X}) \cdot \log f(x_{N+n}|\mathbf{X}, x_{N+1}, \dots, x_{N+n-1}) \end{aligned}$$

$$\cdot dx_{N+1} \cdots dx_{N+n} \cdots \cdots \cdots (3.6)$$

が得られる。

したがって、 $\mathbf{X}(x_1, \dots, x_N)$ および $\mathbf{X}^*(x_{N+1}, \dots, x_{N+n-1})$ を知った後に、 x_{N+n} を知ることにより得られる情報量 $I^{(1)}(x_{N+n}|\mathbf{X}, x_{N+1}, \dots, x_{N+n-1})$ は式 (2.6) より、

$$\begin{aligned} I^{(1)}(x_{N+n}|\mathbf{X}, x_{N+1}, \dots, x_{N+n-1}) \\ = H(x_{N+n}|\mathbf{X}, x_{N+1}, \dots, x_{N+n-1}) \cdots \cdots (3.7) \end{aligned}$$

で与えられ、さらに将来 n 個観測することにより得られる総情報量 $\tau I^{(1)}(x_{N+1}, \dots, x_{N+n}|\mathbf{X})$ は、

$$\begin{aligned} \tau I^{(1)}(x_{N+1}, \dots, x_{N+n}|\mathbf{X}) \\ = \sum_{k=1}^n I^{(1)}(x_{N+k}|\mathbf{X}, x_{N+1}, \dots, x_{N+k-1}) \\ = - \int \cdots \int f(x_{N+1}, \dots, x_{N+n}|\mathbf{X}) \\ \cdot \log f(x_{N+1}, \dots, x_{N+n}|\mathbf{X}) dx_{N+1} \cdots dx_{N+n} \\ \cdots \cdots \cdots (3.8) \end{aligned}$$

で与えられる。なお、以下の議論では、サフィックス N で正規情報を、 E で指数情報を表わすことにする。

b) 正規情報により得られる情報量

一般に、年降水量は正規分布に、月降水量は対数正規分布にしたがうことが多い⁵⁾。ここでは、a) にしたがって、得られる情報が正規分布をする場合の情報量を定式化する。

いま、平均値 θ が未知、分散 v_x が既知である正規確率変数 x の密度関数は、

$$f(x|\theta) = (2\pi v_x)^{-1/2} \cdot \exp\{- (x-\theta)^2/2v_x\} \cdots \cdots \cdots (3.9)$$

で与えられるから、情報 $\mathbf{X}(x_1, \dots, x_N)$ を観測した場合の平均値 θ の事後確率分布 $\xi(\theta|\mathbf{X})$ は、 $\xi(\theta)$ を正規分布 $N(m_\theta, v_\theta)$ と仮定すると、式 (3.1) より、

$$\begin{aligned} \xi(\theta|\mathbf{X}) = (2\pi \cdot v_{\theta|x(N)})^{-1/2} \\ \cdot \exp\{- (\theta - m_{\theta|x(N)})^2/2v_{\theta|x(N)}\} \\ \cdots \cdots \cdots (3.10) \end{aligned}$$

となる。ここに、 $m_{\theta|x(N)} = (N' \cdot m_\theta + N \cdot \bar{x}) / (N' + N)$ 、 $v_{\theta|x(N)} = v_x / (N' + N)$ 、 $\bar{x} = \sum_{i=1}^N x_i / N$ 、 $N' = v_x / v_\theta$ である。

$\xi(\theta|\mathbf{X})$ がわかれば、正規情報 $\mathbf{X}^*(x_{N+1}, \dots, x_{N+n})$ の結合密度関数 $f(x_{N+1}, \dots, x_{N+n}|\mathbf{X})$ および $\mathbf{X}(x_1, \dots, x_N)$ 、 $\mathbf{X}^*(x_{N+1}, \dots, x_{N+n-1})$ が与えられての x_{N+n} の周辺密度関数 $f(x_{N+n}|\mathbf{X}, x_{N+1}, \dots, x_{N+n-1})$ が求まるので、観測を継続することにより得られる情報量 $I_N^{(1)}(x_{N+n}|\mathbf{X}, x_{N+1}, \dots, x_{N+n-1})$ は式 (3.7) より、

$$\begin{aligned} I_N^{(1)}(x_{N+n}|\mathbf{X}, x_{N+1}, \dots, x_{N+n-1}) \\ = - \int \cdots \int f(x_{N+1}, \dots, x_{N+n}|\mathbf{X}) \\ \cdot \log f(x_{N+n}|\mathbf{X}, x_{N+1}, \dots, x_{N+n-1}) \end{aligned}$$

$$\begin{aligned} \cdot dx_{N+1} \cdots dx_{N+n} \\ = 1/2 \cdot (1 + \log 2\pi v_x) \\ + 1/2 \cdot \log \{ (v_x + n v_{\theta|x(N)}) / v_x \\ + (n-1) v_{\theta|x(N)} \} \cdots \cdots \cdots (3.11) \end{aligned}$$

となる。

また、 $\mathbf{X}^*(x_{N+1}, \dots, x_{N+n})$ を将来観測することにより得られる総情報量 $\tau I_N^{(1)}(x_{N+1}, \dots, x_{N+n}|\mathbf{X})$ は式 (3.8)、(3.11) より、

$$\begin{aligned} \tau I_N^{(1)}(x_{N+1}, \dots, x_{N+n}|\mathbf{X}) \\ = \sum_{k=1}^n I_N^{(1)}(x_{N+k}|\mathbf{X}, x_{N+1}, \dots, x_{N+k-1}) \\ = n/2 \cdot (1 + \log 2\pi v_x) \\ + 1/2 \cdot \log (1 + n/v_x \cdot v_{\theta|x(N)}) \cdots \cdots \cdots (3.12) \end{aligned}$$

で与えられる。なお、 θ の事前分布が広い範囲にわたって一様分布しているとみなしうる場合、あるいは N が大きい場合には、 $v_{\theta|x(N)} = v_x / N$ となるので、上式はそれぞれ、

$$\begin{aligned} I_N^{(1)}(x_{N+n}|\mathbf{X}, x_{N+1}, \dots, x_{N+n-1}) \\ = 1/2 \cdot (1 + \log 2\pi v_x) \\ + 1/2 \cdot \log \{ (N+n) / (N+n-1) \} \cdots (3.13) \end{aligned}$$

$$\begin{aligned} \tau I_N^{(1)}(x_{N+1}, \dots, x_{N+n}|\mathbf{X}) \\ = n/2 \cdot (1 + \log 2\pi v_x) + 1/2 \cdot \log (N+n/N) \\ \cdots \cdots \cdots (3.14) \end{aligned}$$

となる。

以上の理論からわかるように、継続情報量は分散 v_x 、 $\xi(\theta)$ の事後分散 $v_{\theta|x(N)}$ および将来観測数 n の関数になっており、 v_x の増加とともに増大する。また、 v_x が一定であれば現在までの観測数 N が多いほど $v_{\theta|x(N)}$ が小さくなるので、将来得られる情報量は少なくなり、 n の増加に伴う情報量増加率も減少する。

以上は分散 v_x を既知とした場合の展開式であったが、実際には観測数の増加とともに平均値だけでなく、 v_x の精度も向上するのであるから、厳密にはパラメータ θ 、 v_x 両者を未知として解析する必要がある。この場合には、事後確率分布 $\xi(\theta, v_x|\mathbf{X})$ は正規ガンマ分布、 $\xi(\theta|\mathbf{X})$ は一般化されたステューデント分布で与えられる。しかし、これをそのまま使ったのでは以後の計算が非常に煩雑になるので、

$$S_x^2 = \sum_{i=1}^N (x_i - \bar{x})^2 / N - 1$$

を v_x の推定値とし、 x_1, x_2, \dots, x_N は既知の分散 $v_x = S_x^2$ をもって正規分布をするのみなすことにする。したがって、 $\xi(\theta|\mathbf{X})$ はやはり正規分布となり、その平均 $m_{\theta|x(N)}$ 、分散 $v_{\theta|x(N)}$ が次式で与えられる。

$$\left. \begin{aligned} m_{\theta|x(N)} &= (N' m_\theta + N \bar{x}) / (N' + N) \\ v_{\theta|x(N)} &= S_x^2 / (N' + N) \\ N' &= S_x^2 / v_\theta \end{aligned} \right\} \cdots (3.15)$$

この手順は N が大きいと $\xi(\theta|\mathbf{X})$ の十分よい近似になっているので、 $v_x = S_x^2$ として上述の理論式を使っていくことが近似的に許されるはずである。なお、正規-ガンマ分布を用いた解析解も見出しているが、非常に複雑で実用的とはいえない。いくつかのデータ例について実際に求めたところ、 $v_x (= S_x^2)$ 既知とした場合の情報量の値と大きく異なる結果を得ており、近似的には上述の理論式で十分であろうと思われる。

c) 指数情報により得られる情報量

降水資料はその時間単位を小さくしていくと非対称度が強まり、日降水量などは一般に指数分布にしたがうことが多い^{6),7)}。ここでは指数分布をする情報に関して、観測を継続することにより将来得られる情報量を算定する。

パラメータを θ とすると、指数分布をする確率変数 x の密度関数は、

$$f(x|\theta) = \theta \cdot \exp(-\theta x) \dots\dots\dots (3.16)$$

で与えられる。この θ の事前確率分布を $\xi(\theta)$ とすると、観測 $\mathbf{X}(x_1, \dots, x_N)$ による θ の事後確率分布 $\xi(\theta|\mathbf{X})$ は、

$$\begin{aligned} \xi(\theta|\mathbf{X}) &= \xi(\theta) \cdot A^{N+1}/N! \cdot \theta^N \cdot e^{-A \cdot \theta} \\ &\cdot \int \xi(\theta) A^{N+1}/N! \cdot \theta^N \cdot e^{-A \cdot \theta} d\theta \dots\dots\dots (3.17) \end{aligned}$$

となる。ここに、 $A = \sum_{i=1}^N x_i$ である。いま、

$$f(\theta|A) = (A^{N+1}/N!) \theta^N e^{-A \cdot \theta} \dots\dots\dots (3.18)$$

とおくと、 θ は平均値 $N+1/A$ 、分散 $N+1/A^2$ のアーラン分布にしたがうことになり、 N が大きいと（日単位を考えているので、一般に N は大きい）、 $\theta = N+1/A$ の近傍で鋭く突き出し、 $\theta = N+1/A$ の小さい近傍の外ではほとんど 0 に近くなるので、 $\xi(\theta)$ が尖鋭でない限り、 $\xi(\theta) \cdot f(\theta|A)$ は $f(\theta|A)$ に比例すると考えてよく、 $\xi(\theta|\mathbf{X})$ は

$$\xi(\theta|\mathbf{X}) = (A^{N+1}/N!) \cdot \theta^N e^{-A \cdot \theta} \dots\dots\dots (3.19)$$

で十分近似できる。

$\xi(\theta|\mathbf{X})$ がわかると、以下は a) で述べた手順により、観測を継続することにより得られる情報量

$$I_E^{(1)}(x_{N+n}|\mathbf{X}, x_{N+1}, \dots, x_{N+n-1})$$

および総情報量 $T I_E^{(1)}(x_{N+1}, \dots, x_{N+n}|\mathbf{X})$ は十分よい近似で、

$$\begin{aligned} I_E^{(1)}(x_{N+n}|\mathbf{X}, x_{N+1}, \dots, x_{N+n-1}) \\ = 1 + \log \bar{x} + \log(N+n/N+n-1) \dots (3.20) \end{aligned}$$

$$\begin{aligned} T I_E^{(1)}(x_{N+1}, \dots, x_{N+n}|\mathbf{X}) \\ = n \cdot (1 + \log \bar{x}) + \log(N+n/N) \dots\dots (3.21) \end{aligned}$$

となり、正規情報の場合とよく似た式が得られることになる。

(2) 既設観測所を廃止後、基幹観測所との回帰推定により得られる情報量

a) 一般的定義式

これまで観測を継続してきた既設観測所を廃止しても、観測中止によりまったく情報が得られなくなるわけではない。なぜならば廃止観測所周辺の継続観測所あるいは基幹観測所から、これまでの観測資料をもとにして、両観測所間の相関関係を知ることにより、廃止観測所の観測値を推定できるからである。ここでは線形回帰モデルを用いて推定した場合に得られる廃止観測所の情報量を、ベイズ論的立場から算定する。

さて、廃止後、回帰推定される観測所のこれまでの観測値を $\mathbf{Y}(y_1, \dots, y_M)$ 、それに対応する基幹観測所の観測値を $\mathbf{X}(x_1, \dots, x_M)$ とすれば（一般には $N \geq M$ ）、 β_1, β_2 を回帰係数とし、確率変数 ε を攪乱項とする線形回帰モデルは、

$$y_i = \beta_1 + \beta_2 \cdot x_i + \varepsilon_i \dots\dots\dots (3.22)$$

で与えられる。ところで、ベイズ論的立場にたてば、観測継続のときと同様に、式 (3.22) のパラメータ β_1, β_2 はある事前確率分布 $\xi(\beta_1, \beta_2)$ をもち、 \mathbf{X}, \mathbf{Y} の観測とともにその精度が向上し、最終的にはその事後確率分布 $\xi(\beta_1, \beta_2|\mathbf{X}, \mathbf{Y})$ は真の β_1, β_2 の値に確率 1 で近づいていく。

3.(1) で述べたと同様に、 \mathbf{X} および \mathbf{Y} が観測された後の $\xi(\beta_1, \beta_2|\mathbf{X}, \mathbf{Y})$ は、

$$\begin{aligned} \xi(\beta_1, \beta_2|\mathbf{X}, \mathbf{Y}) \\ = \xi(\beta_1, \beta_2) \cdot f(y_1|\beta_1, \beta_2, x_1) \dots f(y_M|\beta_1, \beta_2, x_M) \\ \cdot \int \xi(\beta_1, \beta_2) \cdot f(y_1|\beta_1, \beta_2, x_1) \dots f(y_M|\beta_1, \beta_2, x_M) \\ \cdot d\beta_1 d\beta_2 \dots\dots\dots (3.23) \end{aligned}$$

で与えられる。ここに、 $f(y|\beta_1, \beta_2, x)$ は真のパラメータ β_1, β_2 および x が与えられたときに、 y のしたがう確率分布である。 $\xi(\beta_1, \beta_2|\mathbf{X}, \mathbf{Y})$ が求められると、 $\mathbf{X}(x_1, \dots, x_M), \mathbf{Y}(y_1, \dots, y_M)$ および x_{M+n} が与えられた y_{M+n} の周辺密度関数 $f(y_{M+n}|\mathbf{X}, \mathbf{Y}, x_{M+n})$ は、

$$\begin{aligned} f(y_{M+n}|\mathbf{X}, \mathbf{Y}, x_{M+n}) \\ = \iint \xi(\beta_1, \beta_2|\mathbf{X}, \mathbf{Y}) \\ \cdot f(y_{M+n}|\beta_1, \beta_2, x_{M+n}) d\beta_1 d\beta_2 \dots\dots (3.24) \end{aligned}$$

で表わされる。

さて、 \mathbf{X}, \mathbf{Y} および x_{M+n} を知ったとき、 y_{M+n} に関して残る不確かさ、つまりエントロピー、 $H'(y_{M+n}|\mathbf{X}, \mathbf{Y}, x_{M+n})$ は式 (2.6) より、

$$\begin{aligned} H'(y_{M+n}|\mathbf{X}, \mathbf{Y}, x_{M+n}) \\ = - \int f(y_{M+n}|\mathbf{X}, \mathbf{Y}, x_{M+n}) \\ \cdot \log f(y_{M+n}|\mathbf{X}, \mathbf{Y}, x_{M+n}) dy_{M+n} \dots (3.25) \end{aligned}$$

で与えられ、現時点においては x_{M+n} が観測されていないから、 y_{M+n} のもつエントロピーの期待値は、 x_{M+n} の生起確率 $f(x_{M+n}|X, x_{N+1}, \dots, x_{M+n-1})$ 、さらに、 $X^*(x_{M+1}, \dots, x_{M+n-1})$ の結合確率 $f(x_{M+1}, \dots, x_{M+n-1}|X)$ を乗じることにより、最終的には、

$$H(y_{M+n}|X, Y, x_{M+n}) = - \int \int f(x_{M+n}|X) \cdot f(y_{M+n}|X, Y, x_{M+n}) \cdot \log f(y_{M+n}|X, Y, x_{M+n}) dx_{M+n} \cdot dy_{M+n} \dots\dots\dots(3.26)$$

で表わされる。

一方、回帰推定される観測所においては、 $Y(y_1, \dots, y_M)$ しか観測情報が得られていないから、 x_{M+n} を知る以前の y_{M+n} に関する事前の不確実さ $H(y_{M+n}|Y)$ は、

$$H(y_{M+n}|Y) = - \int f(y_{M+n}|Y) \cdot \log f(y_{M+n}|Y) dy_{M+n} \dots\dots\dots(3.27)$$

となり、ここに、 $f(y_{M+n}|Y)$ は、

$$f(y_{M+n}|Y) = \int \xi(\theta_x|Y) \cdot f(y_{M+n}|\theta_y) d\theta_y \dots\dots(3.28)$$

で与えられる。

以上で、回帰分析による y_{M+n} の事前および事後の不確実さが求められたから、観測所を廃止後、回帰推定することによって得られる情報量 $I^{(2)}(y_{M+n}|X, Y, x_{M+n})$ および総情報量 $T I^{(2)}(y_{M+1}, \dots, y_{M+n}|X, Y, x_{M+1}, \dots, x_{M+n})$ は式 (2.7) より、

$$I^{(2)}(y_{M+n}|X, Y, x_{M+n}) = H(y_{M+n}|Y) - H(y_{M+n}|X, Y, x_{M+n}) \dots\dots\dots(3.29)$$

$$T I^{(2)}(y_{M+1}, \dots, y_{M+n}|X, Y, x_{M+1}, \dots, x_{M+n}) = \sum_{k=1}^n I^{(2)}(y_{M+k}|X, Y, x_{M+k}) \dots\dots\dots(3.30)$$

で与えられることになる。

b) 正規情報により得られる情報量

X, Y 間の線形回帰モデルとして式 (3.22) が成立するものと考え、攪乱項 ϵ_i は互いに独立に正規分布 $N(0, v_\epsilon)$ にしたがうものとし、さらに ϵ_i の分散 v_ϵ が既知であるとする。

さて、 ϵ_i は $N(0, v_\epsilon)$ にしたがうから、従属変数 y_i は $N(\beta_1 + \beta_2 \cdot x_i, v_\epsilon)$ にしたがうことになり、 x_i が与えられ β_1, β_2 が真のパラメータの値であるとき、 y_i の確率密度関数 $f(y_i|\beta_1, \beta_2, x_i)$ は、

$$f(y_i|\beta_1, \beta_2, x_i) = (2\pi v_\epsilon)^{-1/2} \exp\{- (y_i - \beta_1 - \beta_2 x_i)^2 / 2 v_\epsilon\} \dots\dots\dots(3.31)$$

で与えられる。ゆえに、 X および Y が観測されての β_1, β_2 の事後確率分布 $\xi(\beta_1, \beta_2|X, Y)$ は、事前確率分布

$\xi(\beta_1, \beta_2)$ が正規分布で、その平均が $m = (m_1, m_2)'$ 、分散-共分散行列が $V = \|v_{ij}\|$ 、 $i, j=1, 2$ である場合、やはり正規分布となり、その平均および分散-共分散行列はそれぞれ次式で与えられる⁸⁾。

$$\left. \begin{aligned} \begin{bmatrix} m_1' \\ m_2' \end{bmatrix} &= h^{-1} \begin{bmatrix} h_{11} \cdot m_1 + h_{12} \cdot m_2 + \sum_{i=1}^N y_i / v_\epsilon \\ h_{21} \cdot m_1 + h_{22} \cdot m_2 + \sum_{i=1}^N x_i \cdot y_i / v_\epsilon \end{bmatrix} \\ \begin{bmatrix} v_{11}' & v_{12}' \\ v_{21}' & v_{22}' \end{bmatrix} &= h^{-1} \\ V^{-1} &= \|h_{ij}\| \\ h &= \begin{bmatrix} h_{11} + N/v_\epsilon & h_{12} + \sum x_i/v_\epsilon \\ h_{21} + \sum x_i/v_\epsilon & h_{22} + \sum x_i^2/v_\epsilon \end{bmatrix} \end{aligned} \right\} \dots\dots\dots(3.32)$$

$\xi(\beta_1, \beta_2|X, Y)$ が上式で求められると、 X, Y および x_{M+n} が与えられての y_{M+n} の周辺密度関数 $f(y_{M+n}|X, Y, x_{M+n})$ は、

$$f(y_{M+n}|X, Y, x_{M+n}) = \{2\pi(v_{11}' + 2v_{12}' \cdot x_{M+n} + v_{22}' x_{M+n}^2 + v_\epsilon)\}^{-1/2} \cdot \exp\{- (y_{M+n} - m_1' - m_2' \cdot x_{M+n})^2 / 2(v_{11}' + 2v_{12}' \cdot x_{M+n} + v_{22}' \cdot x_{M+n}^2 + v_\epsilon)\} \dots\dots(3.33)$$

で与えられるから、将来正規情報 x_{M+n} が観測され、回帰分析により y_{M+n} が推定された場合、 y_{M+n} に関する事後の不確実さ $H_N(y_{M+n}|X, Y, x_{M+n})$ は式 (3.26) より求まる。また、 x_{M+n} を知る以前の y_{M+n} に関する事前の不確実さ $H_N(y_{M+n}|Y)$ は式 (3.13) より求まるから、既設観測所を廃止し、基幹観測所より推定する場合に得られる情報量 $I_N^{(2)}(y_{M+n}|X, Y, x_{M+n})$ および総情報量 $T I_N^{(2)}(y_{M+1}, \dots, y_{M+n}|X, Y, x_{M+1}, \dots, x_{M+n})$ は、

$$I_N^{(2)}(y_{M+n}|X, Y, x_{M+n}) = 1/2 \cdot \{ \log 2\pi \cdot v_y - \log 2\pi(v_{11}' + v_\epsilon - v_{12}'^2/v_{22}') \} + 1/2 \cdot \log \{ (v_y + n v_{\theta_1 y(M)}) / (v_y + (n-1)v_{\theta_1 y(M)}) \} - 1/2 \cdot \{ 1/\tau^2 \cdot 2\delta - 1 \cdot 3/2 \tau^4 \cdot (2\delta)^2 + 1 \cdot 3 \cdot 5/3 \tau^6 \cdot (2\delta)^3 \dots \} \dots\dots\dots(3.34)$$

$$T I_N^{(2)}(y_{M+1}, \dots, y_{M+n}|X, Y, x_{M+1}, \dots, x_{M+n}) = \sum_{k=1}^n I_N^{(2)}(y_{M+k}|X, Y, x_{M+k}) \dots\dots\dots(3.35)$$

で与えられる。なお、上式の誘導においてはマクローリン展開を用いている。ここに、

$$\begin{aligned} \tau^2 &= (v_{11}' v_{22}' + v_{22}' v_\epsilon - v_{12}'^2) / v_{22}'^2 \\ 2\delta &= v_x + (n-1)v_{\theta_1 x(N)} / v_x (v_x + n \cdot v_{\theta_1 x(N)}) \end{aligned}$$

である。

ところで、上式は事前分布 $\xi(\theta_x), \xi(\theta_y)$ の分散要素、 $\xi(\beta_1, \beta_2)$ の分散-共分散行列要素および過去の両観測所のデータ構造要素をすべて含んでおり、要素間の大小関係により推定情報量は変化する。しかし、各要素が同じように情報量増減に寄与するのではなく、事前分布

の分散が大きい、いわゆる不確かさが大きいほど情報量は大きくなるわけであり、その場合の情報量を議論しておく。

いま、 $\xi(\theta_y)$ および $\xi(\beta_1, \beta_2)$ の変動が観測データの個数に比して十分に小さく、あたかも一様分布であるかのように計算を進めることが許される場合を考えると、上式はより簡単に、

$$=1/2 \cdot (\log 2 \pi v_y - \log 2 \pi v_x) + 1/2 \cdot \log M(M+n)/(M+1)(M+n-1) - 1/2 \{M(N+n)/(N+n-1)(M+1)(M-1) - 1 \cdot 3/2 \cdot M^2(N+n)^2/(N+n-1)^2 \cdot (M+1)^2 \cdot (M-1)^2 + \dots\} \dots \dots \dots (3.36)$$

となり、 M, n が大きいと近似的に、

$$=1/2 \cdot (\log 2 \pi v_y - \log 2 \pi v_x) \dots \dots \dots (3.37)$$

総情報量も $n/2 \cdot (\log 2 \pi v_y - \log 2 \pi v_x)$ で近似される。

もちろん、厳密には式 (3.34) を計算するわけであるが、 $1/2 \cdot (\log 2 \pi v_y - \log 2 \pi v_x)$ の値と比較すると、その他の項はかなり小さくなるものと思われ、実用的には式 (3.37) で十分近似できよう。

c) 指数情報により得られる情報量

観測値 X, Y が指数分布をするときも、同様の手順で回帰分析による推定情報量を記述することができる。しかし、この場合には x が与えられ、 β_1, β_2 が真のパラメータの値であるとき、 y の確率密度関数 $f(y|\beta_1, \beta_2, x)$ が、

$$f(y|\beta_1, \beta_2, x) = 1/\beta_1 \cdot \exp\{- (y + \beta_2 x)/\beta_1\} \cdot I_0\{2(\beta_2/\beta_1)^2 \cdot x \cdot y\}^{1/2} \dots \dots \dots (3.38)$$

で与えられることからわかるように、0 次の変形ベッセル関数が積分項に含まれることになり、情報量算定は解析的にも数値計算的にも困難である。ただ、パラメータ推定に最尤法や積率法を用いれば、二重積分の形で数値計算だけは可能である。

(3) 観測所を新設することにより得られる情報量

既設観測所を継続する場合、あるいは既設観測所を廃止し、基幹観測所から回帰モデルにより推定する場合、いずれにおいても、その地点においてこれまで観測 $X(x_1, \dots, x_N), Y(y_1, \dots, y_M)$ がなされており、したがってその X, Y を用いて θ あるいは β_1, β_2 の事後確率分布が求められ、それが将来観測に対する事前分布となった。

ところが、観測所を新設する場合には、観測資料がないために将来の観測情報に対する事前確率分布は利用できるデータから判断される $\xi(\theta)$ そのものとなり、継続情報量の理論式から明らかなように、正規情報に関する新設情報量および総情報量は、

$$I_N^{(3)}(x_n) = 1/2 \cdot (1 + \log 2 \pi v)$$

$$+ 1/2 \cdot \log \{v + n v_0 / v + (n-1) v_0\} \dots \dots \dots (3.39)$$

$$T I_N^{(3)}(x_1, \dots, x_n) = n/2 \cdot (1 + \log 2 \pi v) + 1/2 \cdot \log (1 + n/v \cdot v_0) \dots \dots \dots (3.40)$$

で与えられる。また、指数情報に関してもいままでの理論から近似的に、

$$I_E^{(3)}(x_n) = 1 + \log \bar{x} \dots \dots \dots (3.41)$$

$$T I_E^{(3)}(x_1, \dots, x_n) = n \cdot (1 + \log \bar{x}) \dots \dots (3.42)$$

で与えられよう。

ところで、問題は新設地点の月降水量に関する v_0 および v 、日降水量の \bar{x} の与え方であり、特に量的効果としては v, \bar{x} の推定が重要となる。既存データがある場合には、1) 近接地点の値で代表する、2) 既設観測所の値から描いた等値線から内・外挿する、3) 地形・気象因子との回帰構造から推定する、などの方法が考えられ、データが豊富なほど後の方法に依存するであろう。なお、対象区域にほとんど既存データがない場合には、4) 類似の気候帯・地形帯での降水情報および従来の経験・判断情報をふまえて推定せざるをえないであろう。

(4) マルコフ系列の情報量

以上の議論は、情報源から情報が発生する過程をエルゴード的独立過程と仮定した場合のものであった。しかし、降水量が日単位以下の時間スケールとなると、この仮定が満足されないこともあり、マルコフ過程としての理論も必要である。ただ、豪雨現象の観測にこうした確率論的アプローチのみで十分かどうかについては検討の余地があろう。

さて、指数情報に関する解は前述したようにベッセル関数を含み求められないので、ここでは単純正規マルコフ系列に限定して、いままでの理論式から許される範囲内で情報量算定式を提示しておく。

いま、 x, y 系列が単純正規マルコフ系列であるとすると、情報量算定には 3 つの線形回帰モデルが必要である。すなわち、

$$x_{i+1} = \beta_{1x} + \beta_{2x} \cdot x_i + \epsilon_{ix} \dots \dots \dots (3.43)$$

$$y_{i+1} = \beta_{1y} + \beta_{2y} \cdot y_i + \epsilon_{iy} \dots \dots \dots (3.44)$$

$$y_i = \beta_{1xy} + \beta_{2xy} \cdot x_i + x_{ixy} \dots \dots \dots (3.45)$$

それぞれの回帰式に対して式 (3.24) から、周辺密度関数 $f(x|X), f(y|Y)$ および $f(y|X, Y, x)$ が求められるので、継続情報量 $I_N^{(1)}(x|X)$ および推定情報量 $I_N^{(2)}(y|X, Y, x)$ は前述の展開と同様にして、事前分布の一様性および N が大きいと仮定した場合、

$$I_N^{(1)}(x|X) = 1/2 \cdot (1 + \log 2 \pi v_{\epsilon x}) \dots \dots \dots (3.46)$$

$$I_N^{(2)}(y|X, Y, x)$$

$$= 1/2 \cdot (\log 2 \pi v_{xy} - \log 2 \pi v_{\epsilon_{xy}}) \dots (3.47)$$

与えられ、将来観測数 n の総情報量も近似的に上式の値をそれぞれ n 倍したもので与えられよう。ここに、 $v_{\epsilon_x}, v_{\epsilon_y}, v_{\epsilon_{xy}}$ は式 (3.43), (3.44), (3.45) の $\epsilon_x, \epsilon_y, \epsilon_{xy}$ の誤差分散である。

以上、本章では既設観測所の継続情報量、推定情報量および新設情報量を理論的に算定したが、次章では一つの応用例として、これら情報量の大小関係を用いて降水観測所の配置計画を考察しよう。

4. 降水観測網配置計画への応用

不確実な降水現象を的確に把握するためには、それ相当地に密なる観測網を配置すればよいのであろうが、現実には観測に要する設備・人員・費用などに限度があり、いたずらに観測点をふやすことは経済的にも技術的にも得策とはいえない。本章では前章までの情報量概念を基礎に、最適な観測網配置とはどのようなものであるかを考察する。

(1) 観測網配置計画に関する基本方針

これまで述べてきたように、観測により得られる情報量 I は、

$$I = f(v, N, n) \dots (4.1)$$

の形で表現されよう。いま、計画年の観測数 n を一定として考えると、分散 v の増加に伴い得られる情報量は増加し、これまでの観測数 N の減少とともに I は増加する。したがって、 v, N の大小いずれにおいても、今後得られる情報量が大きなる観測所において観測することが、より有効な観測方法といえよう。

しかしながら、得られる情報量が最大なる観測の配置といっても、ある地域では観測所が接近しすぎており、一方他の地域ではその間隔がまったくかけ離れている、というような配置になっては意味がない。そこで、観測所はある程度一様に分布していることが必要である。ここで、ある程度一様にと述べたことは、各観測所においてその占める支配面積は異なり、またその支配面積を定量的に正確に把握することは現在のところ困難であるからである。ただ、従来の相関分析⁹⁾によれば、平野部ではその観測所の占める支配面積は広く、山岳地帯では狭くなるのが一般に認められており、したがって平野部では観測所を粗に、山岳部では密にすることが必要であろう。

次に、推定したときに得られる情報量が大きければ、推定でも十分役に立つと考えられ、したがって、そうした観測所は廃止し推定することにより、その観測所の継続に必要な費用を他の観測所新設にまわすことができ

る。つまり、より多くの情報量を得ることができる。

各観測所のもつ支配面積が定量的に表現できない現段階では、以上述べたことより、

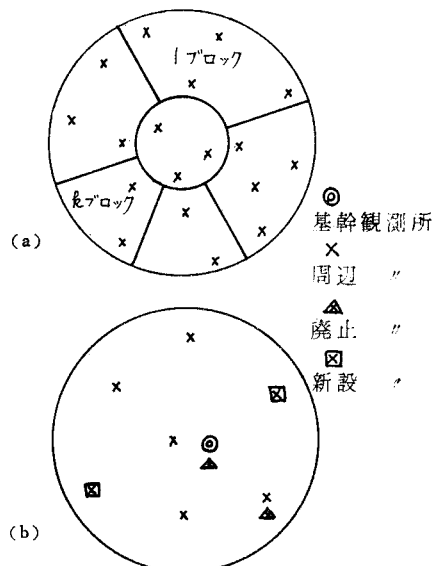
- (1) 観測所が一様に分布するように観測所を新設する。
- (2) 得られる情報量が大きなる地域では観測所を密に、小なる地域では粗に配置する。
- (3) 基幹観測所との相関が高い観測所では、廃止しても推定によってそれ相当の情報が得られる。
- (4) 以上のもので、観測網全体として得られる情報量が最大となる。

といった点を考慮しながら、一定の予算制約のもとで、観測所の継続・廃止・新設を決定し、最適な観測網配置計画をたてる必要がある。

(2) 情報量による観測網配置計画

ここでは、先に述べた方針にしたがって、限定された投資額のもとで、観測所の最適な配置計画作成の手順を述べる。

手順 1: まず、観測網配置計画における対象区域を図一 (a) のように k 個のブロックに分割する。なお、分割方法としては各ブロック内にある既設観測所を継続した場合に得られる総情報量が類似している区域を一つのブロックとする。さて、前述したように、情報量が大きく得られる区域では、観測所を密に、またそのブロックの面積が広い場合にはより多くの観測所設置が要求される。したがって、全投資額 C を、そのブロックの平均情報量と面積の積に比例するように分割し、各ブロックへの投資額 $C_i (i=1, \dots, k)$ を決定する。この手順は前



図一 観測所配置図

述の条件(2)を考慮することになる。

手順 2: 各ブロック内において、その地域の中心的位置にあり、過去の観測数が多くしかも欠測のない精度の高い観測所を選び、そのブロックの基幹観測所とする。次に、各ブロック内において観測所が一樣に分布するように、すなわち 図-1 (b) のように観測所間隔が離れすぎている場合は観測所を新設し、また観測所が異常に接近している場合は廃止する。廃止する場合、いずれを廃止するかといえば、

$$I_{k,l}^{(1)} + I_{k,l+1}^{(2)} > I_{k,l}^{(2)} + I_{k,l+1}^{(1)} \dots\dots(4.2)$$

ならば第 $l+1$ 番目の、その逆ならば l 番目の観測所を廃止する。ここに、 $I_{k,l}^{(1)}$ 、 $I_{k,l}^{(2)}$ は第 k ブロックの l 番目の観測所における継続情報量および基幹観測所との推定情報量である。この手順は前述の条件(1)を考慮することになる。

手順 3: 手順 2 により、各ブロック内の観測所が一樣に分布されたわけであるが、もしこの配置された観測所に必要な費用が、そのブロックへの投資予定額よりも小であれば問題はなく、その差額を他のブロックへの投資額にあてたり、より密な観測所新設にもっていきける。しかし、その逆の場合は既設観測所を廃止するか、あるいは新設の観測所を減らす必要がある。その場合の基準としては、

$$(I_{k,l}^{(1)} - I_{k,l}^{(2)}) / I_{k,l}^{(1)} < \epsilon \dots\dots(4.3)$$

が考えられ、式(4.3)を満足する観測所は廃止してもよからう。ここに、 ϵ は主に予算条件から定まる棄却レベルである。この手順は条件(3)を考慮することになる。

手順 4: 前述の条件(4)を次式で考慮する。

$$\sum_{l=1}^{m_k} \{ I_{k,l}^{(1)} \cdot \delta_{k,l} + I_{k,l}^{(2)} \cdot (1 - \delta_{k,l}) \} + \sum_{l=m_k+1}^{m_k+m_k'} I_{k,l}^{(3)} \cdot \delta_{k,l} \rightarrow \max \dots\dots(4.4)$$

$$\sum_{l=1}^{m_k} \{ C_{k,l}^{(1)} \cdot \delta_{k,l} + C_{k,l}^{(2)} (1 - \delta_{k,l}) \} + \sum_{l=m_k+1}^{m_k+m_k'} C_{k,l}^{(3)} \cdot \delta_{k,l} \leq C_k \dots\dots(4.5)$$

ここに、 m_k 、 m_k' : 第 k ブロック内の既設観測所および新設観測所の総数、 $I_{k,l}^{(3)}$: 新設情報量、 $C_{k,l}^{(1)}$ 、 $C_{k,l}^{(2)}$ 、 $C_{k,l}^{(3)}$: 観測継続・廃止推定・新設に要する費用、 $\delta_{k,l}$: 決定変数で継続なら 1、廃止なら 0、である。

上式は費用の制約条件式(4.5)のもとで、得られる情報量を最大にする目的関数式(4.4)を満足する決定変数 $\delta_{k,l}$ を求める 0-1 整数計画モデルを意味している。ただし、基幹観測所の $\delta_{k,l}$ は 1 であり、実際には手順 1 で選び出した推定可能な観測所間で式(4.4)、(4.5)を解くことになる。

なお、手順 3 で選び出した観測所をすべて廃止し推定することにしても、まだ投資額が不足する場合は、新設観測所において得られる情報量が小さい観測所から順次投資額を満足するまでその新設を中止することにする。

以上の手順により、前述の条件(1)、(2)、(3)を満足し、かつ(4)を満たす最適な観測網配置ができたことになる。

5. 琵琶湖流域への適用と考察

本章では、前章までの理論的アプローチを琵琶湖流域の降水観測網に適用し、各観測所で得られる情報量の地域的・時間的分布特性から、観測網配置計画に関する二、三の考察を試みる。

(1) 流域概要と降水観測資料

本流域は広い沖積盆地のまん中に広大な琵琶湖をもち、琵琶湖東南部は鈴鹿山系から発する野州川・日野川・愛知川などを中心に広大な沖積平野が発達し、西岸は比叡・比良の山系が湖岸に迫り、北東部は姉川を主とする平地が開けている。東岸の彦根・西岸の北小松を結ぶ線より北は、冬期の北西季節風による降雪地帯であ

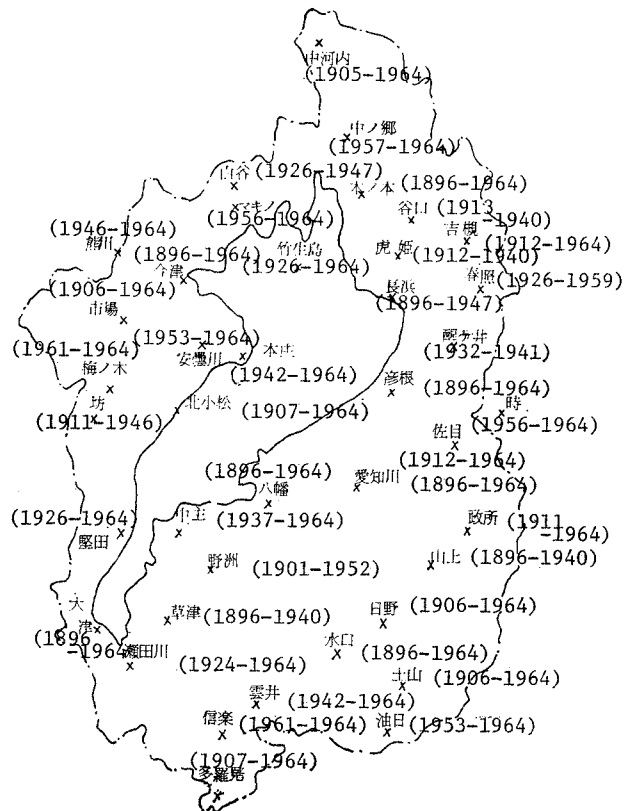


図-2 琵琶湖流域の観測所名と観測期間

て、気候は裏日本型である。

解析に用いた降水資料は、琵琶湖開発の基礎資料として整備されている図-2 に示す 40 個所の観測所での日降水資料および月降水資料である。同図には観測期間も併記してある。なお、以下の計算は京都大学大型計算センター FACOM-230-75 によった。

(2) 観測継続により得られる情報量

一般に、日降水量・月降水量はそれぞれ指数分布および対数正規分布にしたがうといわれているので、ここでは指数情報として日降水量を、正規情報として対数変換

した月降水量を考え、観測継続により得られる情報量を算出した。

まず、情報源がエルゴードの独立過程とみなせるように、一年を冬期(12~2月)、春期(3~5月)、夏期(6~8月)、秋期(9~11月)に分割し、情報量単位はすべてビット単位とした。なお、図-2 からわかるように、過去のデータ数 N は日単位にあっては、最も短い観測期間であっても 4 年 ($N=360$) であり、月単位にあっては信楽・梅ノ木を除いては 9 年以上 ($N=27$) とかなり大きく、事前分布の与え方にあまり影響されない結果となっているので、以下の図上の値は信楽・梅ノ木を除い

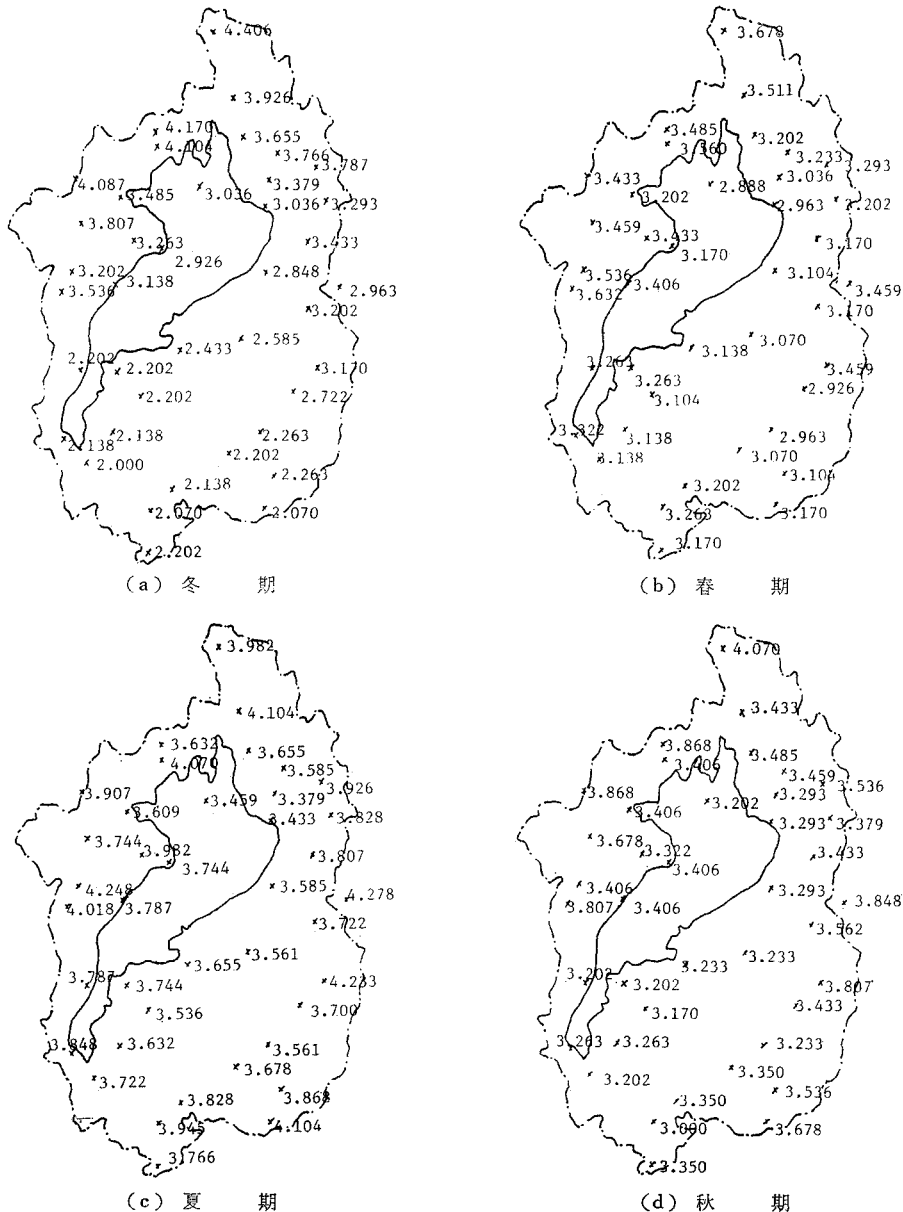


図-3 継続情報量の地域分布図(日降水量)

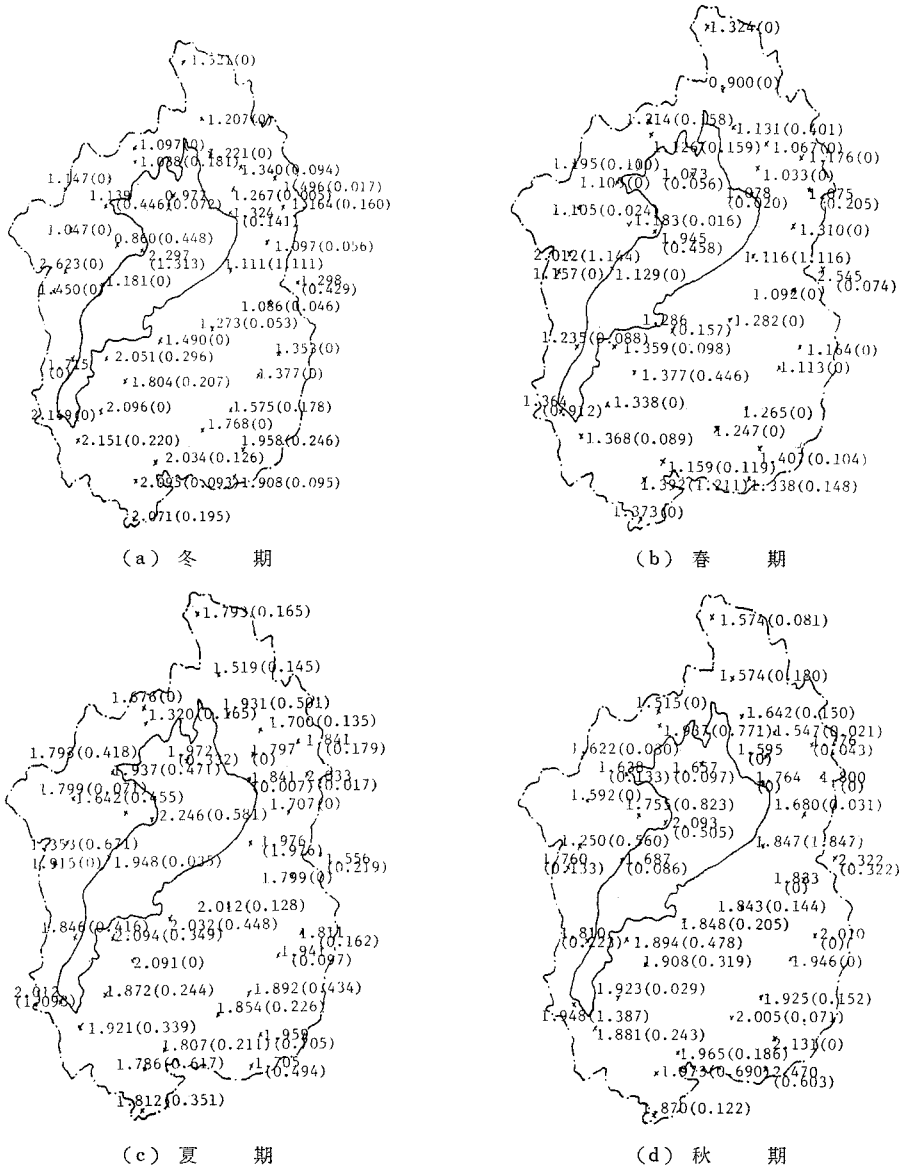


図-4 継続情報量(月降水量)と推定情報量(彦根基幹観測所)の地域分布図

て式 (3.13) および (3.20) から算出したものである。信楽・梅ノ木については 1960 年までの周辺観測所データから得た事前分布に基づいた値を示してある。いずれにしても本流域での事前分布効果はきわめて小さかった。なお、総情報量はこれらの値をほぼ 2 倍すればよろう。

図-3 は日降水量に関する継続情報量である。まず、冬期では情報量の地域的分布特性が他の季節に比べて顕著であり、北へいくほど、それも県境付近で情報量は大きく、東南部は小さい。春期はかなり一様で、しかもその値は 3.0~3.5 とかなり大きく、夏期は湖岸部に比べて県境山地周辺部が大きく、値そのものもほぼ 3.5 以上

となっている。秋期はすべて 3.0 以上であるが、湖周辺部は県境山地部より値は小さい。

次に、図-4 は月降水量に関する各季節の情報量分布図である。ただし、括弧内は次に述べる推定情報量の値である。正規化するために対数変換しているの値そのものは小さいが、やや異なった地域分布特性を示している。すなわち、冬期では湖南周辺部・山地部が大きく、それに隣接する部分および北部県境付近がついで大きく、それらに囲まれる地域は 1.5 以下の値が多い。春期はほとんどが 1.0~1.4 の範囲にあり、特に湖南山地部が大きく、夏期はほぼ一様に分布し、しかもその値は 1.8~2.0 と大きい。秋期は北小松・彦根を結ぶ線より北が 1.8

以下、南が1.8以上とほぼ二分されるような分布をしている。なお、本庄の冬・春期、梅ノ木の夏・秋期は他地点に比べて異常に大きかったり、小さかったりしており、その原因が局所性にあるのか、データに問題があるのか目下検討中である。

こうした分布特性は、日単位では、地域的には平地部より山地部が、季節的には夏・秋期に変動が大きいことを意味しており、月単位では降雪地帯の量的安定性が特徴で、夏・秋期の梅雨・雷雨・台風などは量も大きいとその変動も大きく、その特性がかなり南部のほうに大きく出ているといえよう。

(3) 回帰推定により得られる情報量

指数情報に関しては、ベッセル関数を含み計算が困難であるので、ここでは正規情報すなわち月降水量をいくつかの基幹観測所から推定した場合に得られる情報量を算出し、その結果について考察する。

図—4の括弧内の数値は一例として彦根気象台を基幹観測所とした場合の各観測所の推定情報量の地域分布図である。他の基幹観測所の結果とも併せて考察すると以下のようである。

1) 推定情報量は基幹観測所から遠ざかるにつれて減少するが、その減少特性は季節により、また基幹観測所の位置によって異なる。すなわち、冬期ではその減少は急激であり、推定可能範囲はきわめて狭く、その特性は山地近くの基幹観測所で特に顕著である。また平地部の基幹観測所は推定情報量の大きい観測所をいくつか包含するが、山地に接近すると急激に減少する。平地部基幹観測所としては、大津・瀬田川・八幡・彦根・木の本・今津・北小松などが考えられるが、この中でも瀬田川・木ノ本は狭く、その他においても推定可能範囲は湖岸沿いに広がる傾向を示している。さらに山地近くの基幹観測所として多羅尾・土山・日野・政所・吉槻・市場が考えられるが、推定可能範囲はきわめて狭く、推定情報量自体も小さく、相対的にも西側山地部より、北・東・南部の山地部のほうがその範囲は狭い。

2) ある基幹観測所周辺には各季節とも相当大きな推定情報量が得られる観測所がある。たとえば、大津(基幹)―瀬田川、瀬田川(基幹)―草津であるが、季節ごとになると推定可能な観測所もふえる。

3) それ以外では一般に大きな推定情報量が得られても、その観測所の継続情報量が圧倒的に大きく、観測がいかにも多くの情報を提供するかがわかる。

(4) 観測網配置計画に関する二、三の考察

以上の継続・推定情報量の地域分布状況を考えると、琵琶湖流域の降水観測網配置計画としては次の点が指摘

できよう。

日単位の場合は、推定情報量が見出されていないので一般的なことしかいえないが、月降水量よりもその支配面積が狭いと考えられるので、より多くの観測所が必要となり、各季節とも湖岸平地部より県境山地部に情報量の大きい観測所が多いので、予算に余裕があれば、県境山地部に観測所を新設すべきであろう。

月単位の場合は、季節・基幹観測所によりやや異なった継続・推定情報量分布をとっているもので、それに応じた観測網計画がたてられねばならないが、季節により決定が変動するのは現実的ではないので、総じていけば次のようになろう。すなわち、月単位の場合もどちらかといえば湖岸平地部よりも県境山地部に新設観測所を設けるべきであるが、東南部にも情報量の大きい観測所があり、山地近くの基幹観測所の推定可能範囲の狭さ・観測所間隔の一樣性をも考えあわせると、野州・水口・日野・八幡で囲まれる部分、さらに予算に余裕があれば、野州・草津・雲井・水口で囲まれる中央部、土山・日野・政所・県境で囲まれる部分、吉槻・中河内・県境で囲まれる部分などに観測所が新設されるべきであろう。なお、信楽は多羅尾、雲井との回帰推定によってもかなりの情報量が得られるが、観測開始が最近であるので継続してみるべきであろうし、草津は瀬田川との回帰分析によって十分推定できるので廃止してもよからう。事実、草津は1940年に廃止されている。

6. む す び

本論文は、降水観測がもたらす情報量を Shannon 流情報量で定義し、その展開にベイズ論的立場を導入して、既設観測所を継続する場合、既設観測所を廃止して基幹観測所から推定する場合、および観測所を新設する場合に得られる情報量を算定するとともに、これら情報量概念を用いて観測網最適配置計画の基本フレームを提供したものである。得られた結果を要約すると、以下のようである。

(1) 正規情報に関する継続情報量は、厳密には過去のデータのもつ分散 v_x 、事前分布 $\xi(\theta)$ の分散 v_θ および将来観測数 n の関数で表現されるが、 v_θ 、 n の効果はそれほど小さくなく、特に $\xi(\theta)$ の一様分布あるいは N の大なる場合には、極限情報量として、 $1/2 \cdot (1 + \log 2\pi v_x)$ および $n/2 \cdot (1 + \log 2\pi v_x)$ で近似でき、分散 v_x の増加とともに情報量は増大する。なお、指数情報に関しては一般に N が大きいので、正規情報と類似の式で近似できる。

(2) 正規情報に関する推定情報量は、 $\xi(\theta_x)$ 、 $\xi(\theta_y)$ の分散要素、 $\xi(\beta_1, \beta_2)$ の分散―共分散行列要素および

過去のデータ構造要素の関数となるが、過去のデータ構造要素、特に推定する観測所の分散 v_y および回帰誤差 v_e の効果が大きく、さらにデータ数 N が大きい場合には $1/2 \cdot (\log 2\pi v_y - \log 2\pi v_e)$ および $n/2 \cdot (\log 2\pi v_y - \log 2\pi v_e)$ で近似でき、相関推定の悪い観測所間では推定情報量は小さい。なお、指数情報に関しては、0 次の変形ベッセル関数を含むので解析解の導出は困難である。

(3) 新設情報量においては、新設地点の平均値、分散がなんらかの形で推定できれば、継続情報量の理論式で表現できる。

(4) 基幹観測所のごく近傍以外は、継続情報量が推定情報量よりも圧倒的に大きく、観測がいかにも多くの情報を提供するかがわかる。

(5) 観測網の最適配置計画は、継続・廃止推定・新設情報量を用いて、予算制約下および情報量最大化基準によって 0-1 整数計画問題におきかえることができる。

本論文では、事前分布関数を導入して解を得たが、実際の適用にあたっては過去のデータ数 N が大きいため、事前分布効果はかなり小さなものであった。今後は、こ

うした効果もきいてくる地域への適用をはじめ、費用関数を実際に組み入れて最適化基準に応じた観測網配置計画論を組み立てるとともに、新設地点における平均値、分散の推定にみられるような降水と気象的・地形的因果関係との回帰構造をより明確な形でとらえるよう努力し、本論文で提示した観測網計画にのっとり降水資料を蓄積・整備していきたい。

参考文献

- 1) 中央気象台：磯原地区雨量調査，1952.
- 2) 小林 勉：昭和 32 年夏期只見川上流域水文調査報告，1959.
- 3) Fiering M.B. : An Optimization Scheme for Gaging, Water Resources Research, Vol. 1, No. 4, 1965.
- 4) Matalas, N.C. : Optimum Gaging Station Location, the Progress of Hydrology, IHD, 1969.
- 5) 鈴木栄一：気象統計学，地人書館，pp. 1~12, 1968.
- 6) 角屋 睦：水文統計論，土木学会水工学シリーズ 64-02, 1964.7.
- 7) 石原藤次郎・池淵周一：日降水量の空間的・時間的確率構造とそのシミュレーションに関する研究，土木学会論文報告集 197 号，1972.1.
- 8) 宮沢光一：情報・決定理論序説，岩波書店，pp. 248~250, 1971.11.
- 9) 石原藤次郎・池淵周一：降水観測網の統計的検討，京大防災研究所年報第 13 号 B, 1970.3.

(1974.7.17・受付)