

群馬高専 正会員 小本好克

1. はじめに 時系列モデルは、水工計画に必要となる降水量や流量データの模擬発生あるいは将来予測をする上で有用なツールである。ここでは、年単位の流量時系列を念頭に置いて、定常時系列モデルの選定とその次数の決定方法について述べる。

2. ARMA(p, q) モデルとそのパラメータ推定方法 定常過程とみなせる年流量時系列データとは、 p 次自己回帰型 (AR(p))、 q 次移動平均型 (MA(q))、あるいは混合型 (ARMA(p, q)) などのモデル化に供しうる。ARMA(p, q) モデルの一般式は、 $y_t = \mu + z_t + \dots$ (1)、 $z_t = \sum_{j=1}^p \phi_j z_{t-j} - \sum_{j=1}^q \theta_j \varepsilon_{t-j} + \varepsilon_t$ (2) と表わされる。ここで y_t : 時系列値、 μ : y_t の平均値、 ϕ_j : j 番目の自己回帰係数および移動平均係数、 ε_t : 平均 0、分散 σ^2 の正規分布をもつ白色雑音である。なお $\phi_0 = -1$ である。ここで $q = 0$ とすれば AR(p) モデルに、 $p = 0$ とすれば MA(q) モデルとなる。ARMA(p, q) モデルの自己共分散関数は、(2) 式の両辺に z_{t-k} を乗じ期待値をとることによって、 $\gamma_k = \sum_{i=1}^p \phi_i \gamma_{k-i} - \sum_{j=1}^q \theta_j \gamma_{k-j} + \sigma^2 \delta_{k,0}$ (3)、 $\gamma_k = \sum_{i=1}^p \phi_i \gamma_{k-i}$ 、 $k \geq q+1$ (4) となる。 $\gamma_k(k) = \text{Cov}[z_{t-k}, z_t]$ である。 $k=0$ とすると分散 $\text{Var}[z_t] = \sigma^2 + \sum_{i=1}^p \phi_i \gamma_i - \sum_{j=1}^q \theta_j \gamma_{k-j} + \dots$ (5) が得られる。また自己相関係数 $\rho_k = \gamma_k / \gamma_0$ 、 $k \geq q+1$ (6) であり、部分自己相関係数 $\alpha_k(k)$ は、次数 k の AR 過程 $z_t = \alpha_1(k) z_{t-1} + \dots + \alpha_k(k) z_{t-k} + \varepsilon_t$ と Durbin (1960) の手法により解き、次式となる。 $\alpha_1(1) = \rho_1$ 、 $\alpha_1(2) = \frac{\rho_2(1-\rho_1^2)}{1-\rho_1^2}$ 、 $\alpha_k(k) = \{\rho_k - \sum_{j=1}^{k-1} \alpha_j(k-j) \rho_j\} / \{1 - \sum_{j=1}^{k-1} \alpha_j(k-j) \rho_j\}$ (7)。ここで k は、遅延時間を表わす。

ARMA(p, q) モデルのパラメータ $\phi_1 \sim \phi_p$ および $\theta_1 \sim \theta_q$ を推定するには、実測時系列データより自己共分散 C_k を計算し、(4) 式の値を C_k とおきかえることにより、 $\phi_1 \sim \phi_p$ を推定する。次に (2) 式を $z'_t = z_t - \phi_1 z_{t-1} - \dots - \phi_p z_{t-p}$ (8) とし、この z'_t を MA 過程とみなし、この過程の自己共分散 $C'_k = \sigma^2 \sum_{j=0}^{q-k} \theta_j \theta_{k-j}$ 、 $k \leq q$ (9)、 $C'_k = 0$ 、 $k > q$ (10) および分散 $\text{Var}[z'_t] = C'_0 = \sigma^2 \sum_{j=0}^q \theta_j^2$ (11) より、 $\sigma^2 = C'_0 / (1 + \theta_1^2 + \dots + \theta_q^2)$ (12)。および $\hat{\theta}_j = -(\frac{C'_j}{\sigma^2} - \hat{\theta}_1 \hat{\theta}_{j+1} + \dots - \hat{\theta}_{j-1} \hat{\theta}_j)$ (13) とし、両式より $\hat{\theta}_1 \sim \hat{\theta}_q$ を推定する。またパラメータ μ は、 $\mu = \frac{1}{N} \sum_{t=1}^N y_t$ (14) により、 σ^2 は、(2) 式を $z_t = z_{t-1} - \phi_1 z_{t-1} + \dots + \hat{\theta}_q \varepsilon_{t-q}$ とし、初期値 $\varepsilon_0, \varepsilon_1, \dots, \varepsilon_q$ および $z_0, z_1, \dots, z_p = 0$ とおき、 $\varepsilon_1 = z_1$ 、 $\varepsilon_2 = z_2 - \phi_1 z_1 + \hat{\theta}_1 \varepsilon_1$ 、 $\dots$ と導出し、 $\sigma^2 = \frac{1}{N} \sum_{t=1}^N (z_t - \hat{z}_t)^2$ (14) より推定する。ここで N は、 N 年間の時系列データを表わす。

3. モデルの選定および次数決定方法の手順 N 年間の時系列データ y_t を用いて、モデルの選定とその次数を決定する方法の手順は、以下のとおりである。なお解析上のデータとしては、(2) 式の $z_t (= y_t - \mu)$ を用いる。

手順1 z_t の正規性を検証する。これには、 χ^2 検定、 Shapiro-Wilk 検定あるいは、正規確率紙法などが利用できる。非正規の場合には、正規にならざる変換等とし、再度検定する。

手順2 $z_t \sim z_{t-k}$ を時間軸上にプロットしてみる。流出過程において地下水や表面貯留の影響が現れる場合には AR モデルが、無視できない場合には MA モデルあるいは ARMA モデルが適当であろうと考えられるので、図よりこれらの有無を推察し、モデルの選定を試みる。

手順3 z_t の自己相関係数 ρ_k を $\rho_k = \frac{\sum_{t=1}^{N-k} (z_t - \bar{z})(z_{t+k} - \bar{z})}{\sum_{t=1}^N (z_t - \bar{z})^2}$ 、 $0 \leq k < N$ (15) より、また部分自己相関係数 $\alpha_k(k)$ を (7) 式の値を ρ_k とおきかえることにより計算する。なお ρ_k 、 $\alpha_k(k)$ は近似時に平均 0、分散 $1/N$ の正規分布としているので、95% 確率範囲は Anderson (1991) の式 $\rho_k(95\%) = -1 \pm 1.96 / \sqrt{N-k-1} / (N-k)$ によって計算する。これらとすべてプロットする。

手順4 ρ_k が指数的にかつ/あるいは波状に急激に減少し、遅延時間 p 以降が 95% 確率範囲内にあれば、AR(p) モデルと、遅延時間 1 から q の間にくいつかのピーク値を持ちながら減少し、 q 以降が確率範囲内にあ

これはMA(8)モデルと考える。残差が指数的かつ/あるいは緩慢的にゆるやかに減少する場合には、ARMA(p,8)と考える。判然としない場合あるいは次数中、8の決定には、 $\Phi_R(k)$ 図とを参考合わせる。

手順5 上述の手順にて選択されたモデルの次数中、8の近傍の値を種々変化させ、各々の次数ともつモデルの残差、 $\Phi_R(k)$ および σ_e^2 を(6)、(7)および(14)式より算出する。

手順6 時系列データより求めた $\hat{Y}_k$ と種々のモデルから得られた $\hat{Y}_k$ とを比較し、適合性が良く、かつ σ_e^2 の値が最小値を示すモデルを選定する。

4. 適用例 群馬県烏川流域下流端岩鼻観測所における、16年間(1958-1973)の流量データを用い、上述の手順に従い、モデルの選定とその次数の決定を試みる。なお予測データは、流域内の降雨データを用いた単回帰手法にて補完した。解析では、実測流量データ y_k とその平均値を除した値を z_k とし、さらに $z_k = y_k - \bar{y}$ としこの z_k を用いている。これらを表-1に示す。

表-1 岩鼻地点での年流量時系列データ(m³/s)および z_k

* 補完流量

年	1958	1959	1960	1961	1962	1963	1964	1965	1966	1967	1968	1969	1970	1971	1972	1973
流量	15271	17026	5532	10911	6195	5581	9353	8619	*13131	*11018	14936	*7071	8723	10673	11481	5949
z_k	0.573	0.687	-0.452	0.021	-0.386	-0.447	-0.073	-0.146	0.301	0.092	0.440	-0.299	-0.135	0.058	0.138	-0.411

手順1 z_k の歪み係数は、0.422と出る。Snedecor (1967)から83 確率水準 $\alpha=0.02$ にあたる(1-d)確率範囲は±0.504と出るので、このデータは、正規分布をばしていると思われる。

手順2 流量時系列データ z_k と時間軸上にプロットし、図-7に示す。図から時系列間に何らかの関連性が見え出せる。特に後半では、高流量は前年の高流量に依存している様があり、ARMAモデルが考えられるが、この段階でのモデル選定は困難である。

手順3 z_k の $\hat{Y}_k$ 、 $\Phi_R(k)$ および $\hat{Y}_k(95\%)$ の計算結果を表-2と図-2,3に示す。

図-7 年流量時系列データ z_k

表-2 $\hat{Y}_k$ 、 $\Phi_R(k)$ および $\hat{Y}_k(95\%)$

k	95%下限値	$\hat{Y}_k$	$\Phi_R(k)$	95%上限値
1	-0.537	0.030	0.030	0.464
2	-0.576	0.006	0.006	0.433
3	-0.599	-0.125	-0.125	0.445
4	-0.625	-0.246	-0.246	0.458
5	-0.654	-0.270	-0.303	0.472

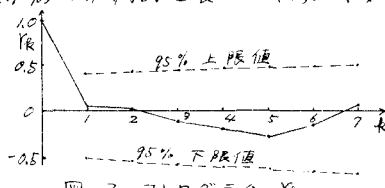


図-2 コレログラム Y_k

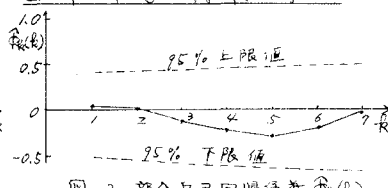


図-3 部分自己相関係数 $\Phi_R(k)$

手順4 z_k のコレログラム Y_k は、 $\hat{Y}_k=1$ から k に急激に減少し、その後、95%確率範囲内に存在する。また $\Phi_R(k)$ のすべての値は、95%確率範囲内にある。これらからAR(1)モデルあるいはARMA(1,1)モデルが考えられる。

手順5 $\Phi_R(1,2)$, $\Phi_R(2,1)$ の組合せの

表-3 各種モデルの $\hat{Y}_k$ 、 $\Phi_R(k)$ および σ_e^2

モデルととりわけ、これらの $\hat{Y}_k$ 、 $\Phi_R(k)$ および σ_e^2 を計算し、結果を表-3に示す。

手順6 すべてのモデルの σ_e^2 は同一値と

した。データの $\hat{Y}_k$ とARMA(1,1)モデルの $\hat{Y}_k$ とが最も良く一致している。よって、ここからはこのモデルを選定する。

	ARMA(0,0)	ARMA(0,1)	ARMA(1,0)	ARMA(1,1)	ARMA(2,0)
	$\Phi=0, \Theta=0$	$\Phi=0, \Theta=-0.028$	$\Phi=0.028, \Theta=0$	$\Phi=0.015, \Theta=0.015$	$\Phi=0.028, \Theta=0.005$
k	$\hat{P}_k$	$\hat{P}_k$	$\hat{P}_k$	$\hat{P}_k$	$\hat{P}_k$
1	0.000	0.000	0.028	0.030	0.028
2	:	0.000	0.028	0.030	0.028
3	:	0.000	0.000	0.000	0.000
4	:	:	:	0.000	0.000
5	:	:	:	0.000	0.000
σ_e^2	0.130	0.130	0.130	0.130	0.130

5. おわりに ここでの適用例は、データ数等の制約から、必ずしも良い例とはいえないが、3.での手順は、年時系列データをモデル化する上で、参考にしようと考えている。なお選定されたモデル(ここではARMA(1,1)モデル)が水工計画に、有意なものがあるかどうかは、さらに、推定されたパラメータの定常性や収束性条件の満足性あるいは信頼性の検討、残差の独立性と正規性の検定およびモデル次数を減少できるかどうかの検討の必要がある。【参考文献】 J.D.Salas, et al, "Applied Modeling of Hydrologic Time Series", WRP, 1970 ; J.D.Salas, et al, "Statistical Computer Programs in Hydrology", CSU, July 1970