

週期性 自己回歸-移動平均(PARMA) 模型의 媒介變數 最適 算定 Parameter Optimization of Periodic Autoregressive- Moving Average(PARMA) Model

고석구*, 배종찬**, 이광만***, 윤재홍****

1. 序 論

단기간의 수문자료로부터 추계학적 모형을 통하여 장기간의 자료로 확충하는 방법은 지난 20여년동안 꾸준히 연구되어 왔다. 하천유량과 같은 장기간의 수문시계열 자료는 수공구조물의 계획, 설계 및 효과적인 저수지 운영 방안 수립 등에 중요한 입력자료로 활용된다. 수문현상을 과거기록자료의 통계적 특성을 분석하고 시간 및 공간적 상관성을 나타내는 매개변수를 산정하여 모형화하는 추계학적 방법은 사용의 간편성과 모형구조의 단순성으로 인해 널리 이용되어 왔는데, 그 대표적인 것이 AR(autoregressive) 모형^{1), 2), 3)}과 ARMA(autoregressive-moving average) 모형^{4), 5)}이다.

ARMA모형은 AR모형에 비해 보다 더 적은 수의 매개변수로써 모형을 구성할 수 있다는 장점을 가지고 있다⁶⁾. ARMA모형의 일반적인 매개변수 산정 방법은 두단계로 나누어지는데 첫째는 Yule-Walker 방정식^{7), 8)}에 의해 AR성분의 매개변수를 산정하고 AR모형에 의한 시계열과 原시계열(original series)의 差로부터 새로운 시계열을 구성한 후 MA성분의 매개변수를 산정하는 예비산정 단계이다. 둘째는 예비산정 단계에서 구해진 매개변수값들을 토대로 最尤法(maximum likelihood method)등으로 모형의 精度를 보다 더 향상시키기 위한 최종산정 단계이다. 이러한 매개변수 산정은 추계학적 모형화에 있어서 매우 중요한 부분이다.

본 연구에서는 시간에 따라 변하는 매개변수를 갖는 週期性 ARMA(periodic ARMA with time varying coefficients)모형 중 사용자가 선택적으로 사용할 수 있는 컴퓨터 Program을 개발하였으며 여기에 직접탐색기법(direct search)을 연결하여 매개변수 최적산정을 실시하였다.

2. PARMA 模型의 基本理論

ARMA모형은 연간유입량 자료와 같이 Constant 매개변수를 갖는 Annual 모형과 월간유입량 자료와 같이 週期性 매개변수를 갖는 Periodic 모형으로 구분할 수 있다. Periodic 모형은 또다시 常數(constant) 係數와 非常數(time varying) 係數를 갖는 모형으로 구분할 수 있는데 본 연구에서는 후자를 이용하였으며 편의상 PARMA모형이라 칭한다.

이 모형의 수학적 구성은 다음과 같이 표시할 수 있다⁹⁾.

$$y_{v,\tau} = \mu_{\tau} - \sigma_{\tau} \cdot z_{v,\tau} \text{ -----(1)}$$

$$z_{v,\tau} = \sum_{j=1}^p \phi_{j,\tau} z_{v,\tau-j} - \sum_{i=1}^q \theta_{i,\tau} \varepsilon_{v,\tau-i} + \varepsilon_{v,\tau} \text{ -----(2)}$$

여기서, $y_{v,\tau}$ 는 정규분포를 가지며, 週期系列 $\tau=1, \dots, w$ 와 연간 계열 v 에 대한 週期的 水文 시계열이다. w 는 1년內에 들어있는 시간단위의 數이며, p 는 AR모형의 차수(order)를, q 는 MA(moving-average) 모형의 차수를 나타내는 數이다. μ_{τ} 와 σ_{τ} 는 각각 $y_{v,\tau}$ 의 週期的 평균값과 표준편차를 나타낸다. $z_{v,\tau}$ 는 표준화된 定常時系列(stationary time series)을 표시하며 평균값 0, 표준편차 1을 갖는다. $\phi_{j,\tau}$ 와 $\theta_{i,\tau}$ 는 각각 AR계수, MA계수를 나타낸다. $\varepsilon_{v,\tau}$ 는 평균값이 0, 분산이 σ_{τ}^2 인 무작위 독립 시계열이다. 식(1)과 (2)에서 모형의 매개변수의 집합은 $(\mu_{\tau}, \sigma_{\tau}, \phi_{j,\tau}, \theta_{i,\tau}, \sigma_{\tau}^2; j=1, \dots, p, i=1, \dots, q, \tau=1, \dots, w)$ 이다. 여기서 μ_{τ} 와 σ_{τ} 는 標本집단(sample)으로부터 계산하거나 Fourier Series Fitting에 의해 계산된 값을 사용하며 그외의 매개변수들은 그림 1과 같은 흐름도에 의해 일반적으로 산정된다.

* 한국수자원공사, 조사계획처장, 공학박사
** 한국수자원공사, 수자원연구소, 연구원
*** 한국수자원공사, 수자원연구소, 연구원, 중앙대 박사과정 수료
**** 한국수자원공사, 수자원연구소, 선임연구원

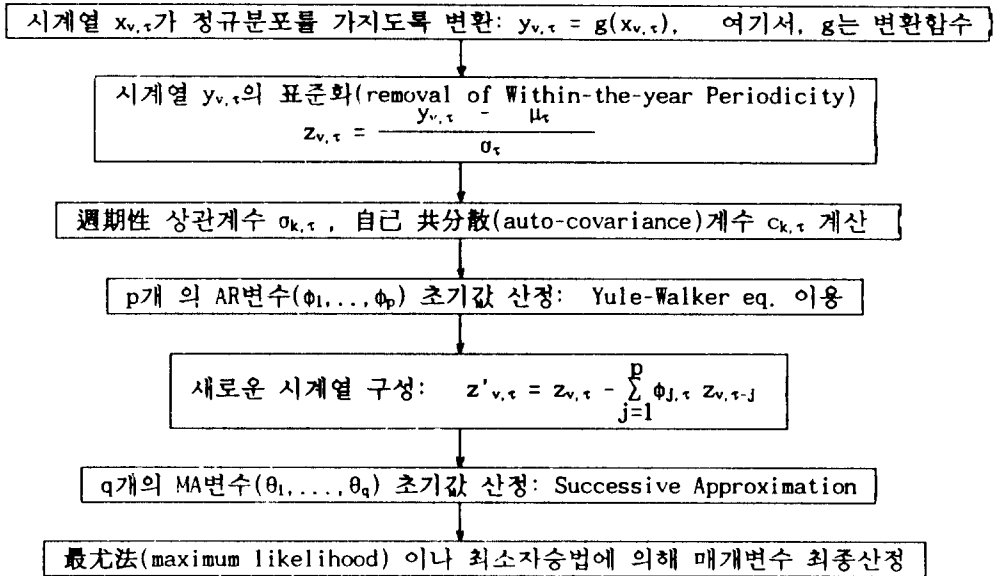


그림 1. PARMA 모형의 매개변수 산정 흐름도

3. 媒介變數 最適化 技法

3.1 最適化 技法의 選擇

복잡한 강우-유출 모형의 매개변수 자동보정을 위해 많은 최적화 기법들이 개발되어 왔는데 이러한 기법들은 평가기준(evaluation criterion)함수를 매개변수에 대해 1계 혹은 2계 미분을 실시하여 해석하는 Gradient기법과 평가기준 함수에 대한 수학적인 기초 지식없이 일종의 시행착오법에 의해 최적점을 찾아가는 직접탐사기법(direct search method)으로 대별할 수 있다.

실제 문제에 많이 사용되는 직접탐사기법으로는 Rotating Directions 기법¹⁰⁾, Pattern Search 기법¹¹⁾, Simplex 기법¹²⁾, Adaptive Random Search 기법¹³⁾과 Shuffled Complex Evolution 기법¹⁴⁾등이 있다. 이러한 기법들 중 Pattern Search와 같은 기법은 주어지는 초기값으로부터 순차적으로 최적점을 찾아가는 방법으로서 최종결과는 주어지는 초기값에 의존하므로 응답공간(response space)이 복잡하고 다수의 극대, 극소점이 존재할 경우 局部 최적점(local optima)에 빠지기 쉽다. 이러한 단점을 보완하기 위해 무작위 탐색기법(random search method)들이 연구되어 왔다.

본 연구에서는 상기에서 소개된 탐색기법들 중 Pattern Search 기법과 이 기법과는 접근방법이 대조적인 Shuffled Complex Evolution(SCE) 기법을 대표적으로 선택하여 PARMA모형의 매개변수 산정에 이용하였으며, 분석결과가 비교 검토 되었다.

3.2 Pattern Search 技法

Hooke과 Jeeves¹¹⁾에 의해 개발되고 Monro¹⁵⁾에 의해 개량된 Pattern Search 기법은 局部이동(Local Excursion, LE)과 標本이동(Pattern Move, PM)의 두과정을 거치면서 반복적으로 평가기준값을 계산하면서 최적해를 찾아 가는 방법으로서 표본이동과정이 이 기법의 두드러진 특징이라 할 수 있다. 이 기법은 사용자에게 주어지는 초기변수값들을 가지고 시작하는데, 우선 국부이동과정을 거치면서 평가기준값을 향상시키는 탐색방향이 결정되며, 표본이동에서는 국부이동에서 결정된 탐색방향과 이동규모에 따라 변수들을 조정하고 평가 기준값을 계산하면서 계속해서 최적해를 찾아가간다. 표본 이동에 있어서 변수들의 조정치를 결정하는 이동규모는 다음과 같은 식에 의해서 결정된다.

$$\begin{aligned} \xi^1(I) &= N(i) * \text{DELTA}(I) \\ N(i) &= n_1(I) - n_2(I) \end{aligned} \quad \text{----- (3)}$$

여기서, $\xi^1(I)$: i 번째 표본이동시 I 번째 변수의 조정치
 $\text{DELTA}(I)$: I 번째 변수의 국부이동시 조정치

$n_1(I)$: I번째 변수에 있어서 연속된 양(陽)의 국부이동 횟수
 $n_2(I)$: I번째 변수에 있어서 연속된 음(陰)의 국부이동 횟수

국부이동과 표본이동이 평가기준값이 더 이상 향상되지 않을 때까지 交互的으로 계속되며, 다음과 같은 특이한 경우가 발생할 수 있다. 표본이동후 평가기준값이 향상되지 않고 또한 그곳에서의 국부이동에서도 향상되지않는 경우가 있는데 이러한 경우는 표본이동에 의한 변수조정치가 너무 크다는 것을 의미하므로 표본이동에 의해 변수를 조정하지않고 국부이동을 행하여 변수를 조정하고 평가기준값을 산정한다. 다음은 표본이동에 의해 평가기준값이 향상되지만 후속 국부이동시에는 그렇지 않는 경우가 있는데 이때에는 국부이동에 의한 변수 조정치를 반감시켜 평가 기준값을 산정한다. 주어진 횟수만큼 변수 조정치를 반감시켜도 평가기준값이 더 이상 향상되지 않을 경우에는 최적점에 수렴하였다고 본다.

이 기법은 주어지는 초기 변수값에서부터 출발하여 최적점을 찾아가기 때문에 분석결과도 초기 변수값에 영향을 많이 받는다. 따라서 분석결과에 대한 신뢰도를 높이기 위해서는 여러개의 초기 변수값을 대상으로 탐색을 실시하여야 한다.

3.3 Shuffled Complex Evolution 技法

Shuffled Complex Evolution(SCE) 기법은 원래 복잡한 강우-유출 모형에 사용되는 매개변수들의 자동 검정시 기준에 개발된 방법으로는 높은 신뢰도를 가지고 Global Optima를 찾을 수가 없기 때문에 이러한 점을 개선하기 위해 Duan 등¹⁴⁾에 의해 개발되었다. 이 기법은 無作爲 탐색 기법의 일종으로서 Nelder와 Mead¹²⁾의 Simplex 기법에 Price¹⁶⁾의 Controlled Random Search 기법, Holland¹⁷⁾의 경쟁적 진화(competitive evolution)이론과 Duan 등¹⁴⁾의 집단 再編개념(complex shuffling)의 장점들을 결합하여 개발되었다.

이 기법은 우선 매개변수들의 妥當域(feasible space)에서 무작위로 점들을 추출하여 하나의 母集團(population of points)을 형성한다. 이 모집단을 몇개의 集團(complex)으로 분할하고 각 집단에서는 Simplex 기법을 이용하여 평가기준값이 향상되는 점들을 탐색해 나간다. 그리고 각 집단에 소속된 점들을 모아 새롭게 모집단을 만들고 다시 집단으로 분할하여 평가기준값이 향상되는 점들을 탐색해 나가는데 이러한 과정을 최적점에 수렴할 때까지 주기적으로 반복한다. 이러한 과정을 도식적으로 나타내면 그림 2와 3과 같다.

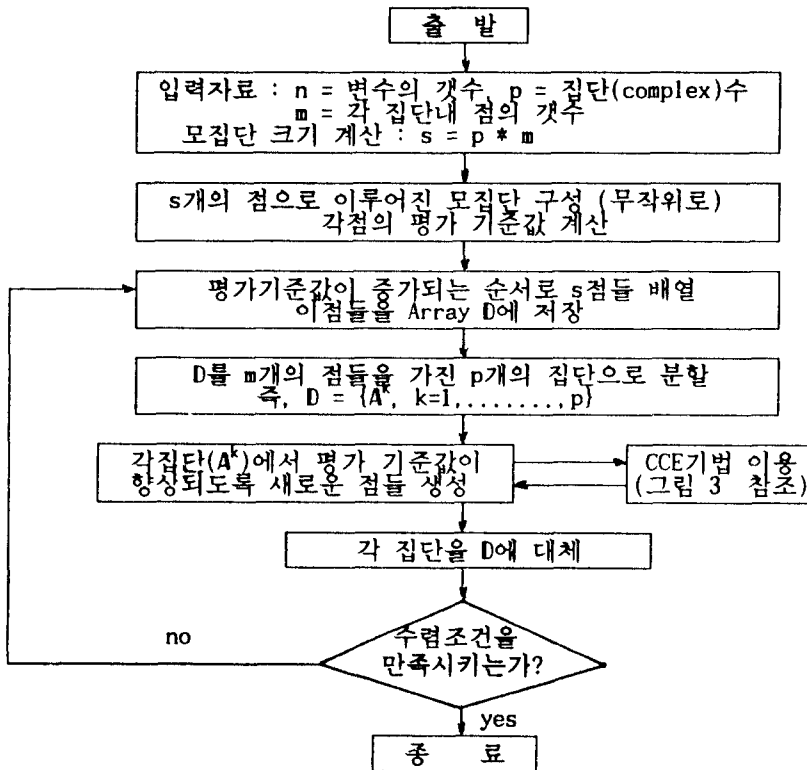


그림 2. SCE 기법의 흐름도 (참고문헌(14)에서 인용)

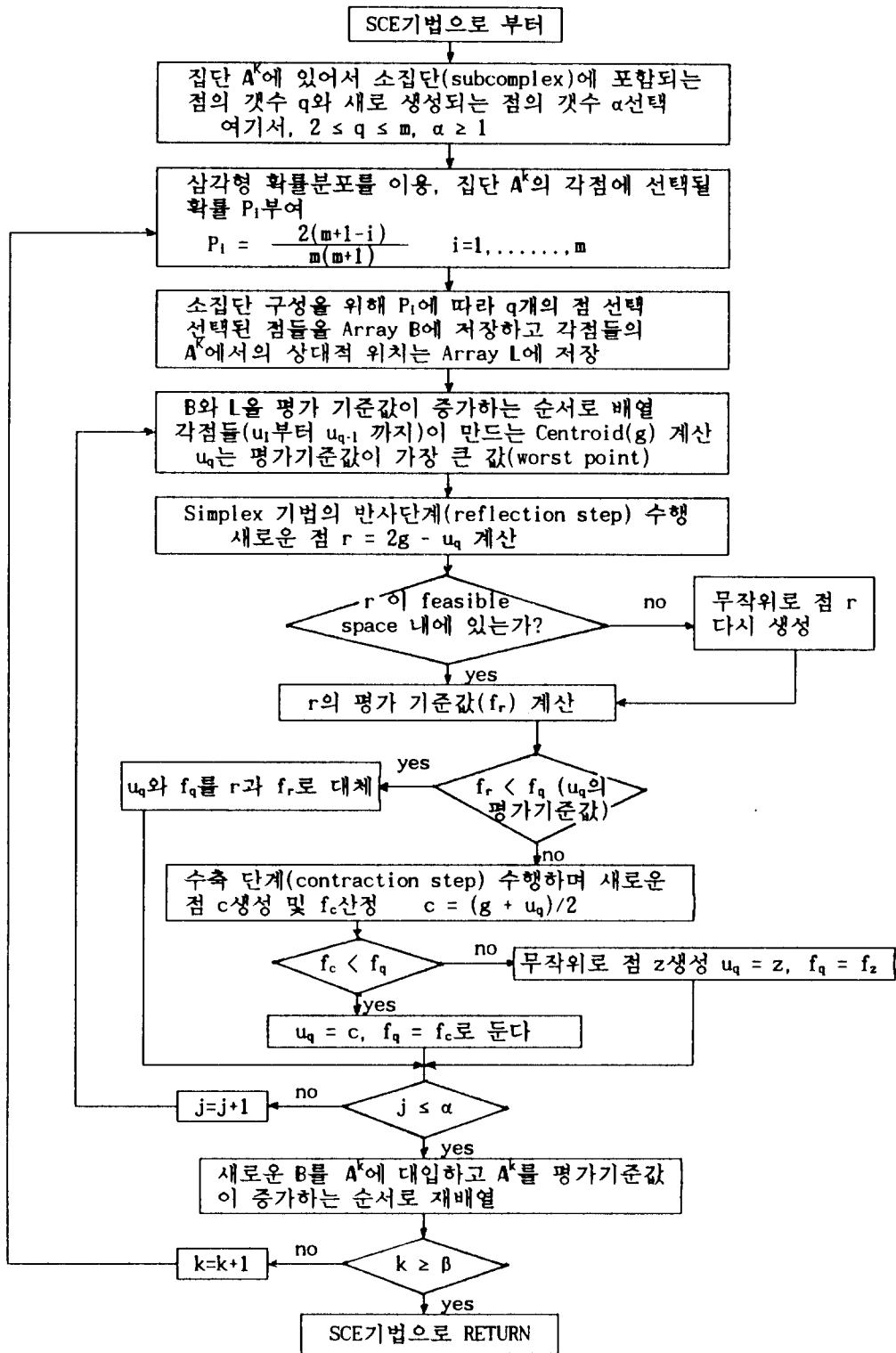


그림 3. CCE(competitive complex evolution) 기법 흐름도 (참고문헌(14)에서 인용)

4. 適用結果 및 考察

4.1 適用概要

본 연구를 위해 PARMA모형은 PARMA(1,1), PARMA(1,2), PARMA(2,1), PARMA(2,2)를 개발하여 하나의 package로 만들었으며, 사용자가 각각 다른 차수를 가진 모형을 선택적으로 사용할 수 있게 하였다. PARMA 모형과 연결된 Pattern Search 기법과 SCE기법의 활용성을 다음의 두가지 경우를 가지고 시험하였는데 첫째는 그림 1과 같은 흐름도에 따라 산정된 초기값을 가지고 최종산정시 이용하는 경우이고 둘째는 매개변수의 초기값을 산정하지 않고 任意로 0의 값을 준 후 Search기법을 이용하여 매개변수를 최종산정하는 방법이다.

탐색시 매개변수값의 上下限값은 각각 +1과 -1이 주어졌으며, 평가기준값(evaluation criterion)은 殘差의 自乘和(SSR, sum of squares of residuals)으로 주어지고 SSR의 값이 최소가 되는 매개변수들의 값을 탐색하게 된다. 殘差는 아래식과 같이 나타난다.

$$\varepsilon_{v,t} = z_{v,t} - \sum_{j=1}^p \hat{\phi}_{j,t} z_{v,t-j} + \sum_{i=1}^q \hat{\theta}_{i,t} \varepsilon_{v,t-i} \quad (4)$$

여기서, $z_{v,t}$ 는 표준화된 시계열, $\hat{\phi}_j$ 와 $\hat{\theta}_i$ 는 각각 산정된 AR계수 및 MA계수이다.

4.2 適用結果 檢討

본 연구에서는 PARMA(2,2)모형에 탐색기법을 연결하여 매개변수를 산정하였으며, 週期的시간은 月단위를 가진다. 월별유입량 자료는 충주지점의 50년간(1917-1940;1956-1981)자료를 이용하였다. 표 1은 탐색기법에 의해 최종탐색된 월별 매개변수값의 집합을 보여주고 있으며, 표 2는 계산된 월별 SSR값을 보여주고 있는데, 두 탐색기법이 매개변수 초기값이 주어질 때나 주어지지 않을 때나 공히 비슷한 결과를 나타내고 있다.

일반적으로 Pattern Search 기법은 응답공간(response space)이 복잡해지면 주어지는 초기값에 따라 局地최적점(local optima)에 빠지기 쉬운데 여기서는 초기값에 관계없이 일관성 있게 좋은 Optima를 탐색하고 있음을 알 수 있다. 이 분석에 소요된 컴퓨터의 계산시간은 PC386, 33MHz 기종을 사용하였을 때 Pattern Search는 약 55초, SCE기법은 약 4분 20초가 소요되어 Pattern Search를 이용한 방법이 약 4.5배 빨리 수렴되었다.

표 1. 최종탐색된 월별 매개변수값

month	condi.	Pattern Search				SCE			
		phi1	ph2	theta1	theta2	phi1	ph2	theta1	theta2
1	w/ init.	0.6243	0.1907	-0.0001	-0.0001	0.6240	0.1909	0.0001	-0.0001
	w/o init.	0.6240	0.1910	0.0000	0.6240	0.6241	0.1908	0.0001	-0.0000
2	w/ init.	0.5078	0.2430	-0.0660	0.0001	0.5076	0.2433	-0.0658	0.5076
	w/o init.	0.5077	0.2430	-0.0660	0.0000	0.5071	0.2437	-0.0683	-0.0001
3	w/ init.	0.2629	0.2472	-0.0879	0.0581	0.2621	0.2476	-0.0880	0.0583
	w/o init.	0.2618	0.2485	-0.0883	0.0590	0.2618	0.2499	-0.0888	0.0557
4	w/ init.	0.2233	-0.0121	0.0695	0.0133	0.2288	-0.0131	0.0676	0.0111
	w/o init.	0.2250	-0.0099	0.0750	0.0000	0.2287	-0.0131	0.0605	0.0153
5	w/ init.	0.6045	0.1057	0.0281	0.0144	0.6047	0.1057	0.0280	0.0143
	w/o init.	0.6048	0.1055	0.0283	0.0145	0.6052	0.1059	0.0284	0.0113
6	w/ init.	0.1777	-0.0057	0.1470	-0.1464	0.1856	-0.0086	0.1418	0.1856
	w/o init.	0.1750	-0.0055	0.1500	-0.1500	0.1854	-0.0086	0.1425	0.1854
7	w/ init.	0.1274	0.0110	-0.0130	-0.0083	0.1278	0.0111	-0.0132	0.1278
	w/o init.	0.1275	0.0110	-0.0130	-0.0085	0.1278	0.0111	-0.0134	-0.0087
8	w/ init.	-0.0381	0.0001	0.0162	0.1054	-0.0344	-0.0003	0.1393	0.0885
	w/o init.	-0.0500	-0.0001	0.1250	-0.0500	-0.0453	-0.0005	0.1250	0.0785
9	w/ init.	0.3453	-0.0120	0.1160	0.0051	0.3416	-0.0113	0.1125	0.3416
	w/o init.	0.3408	-0.0115	0.1133	0.3408	0.3404	-0.0112	0.1141	0.3404
10	w/ init.	0.4915	0.2161	-0.0843	-0.0506	0.4951	0.2132	-0.0992	-0.0392
	w/o init.	0.4883	0.2198	-0.0850	-0.0385	0.4915	0.2187	-0.0882	0.4915
11	w/ init.	0.6879	0.0893	-0.1049	-0.1258	0.6768	0.1045	-0.1039	-0.1328
	w/o init.	0.6875	0.0899	-0.1048	-0.1253	0.6837	0.0998	-0.1080	0.6837
12	w/ init.	0.7760	-0.0155	-0.0507	-0.0483	0.7726	-0.0114	-0.0462	-0.0585
	w/o init.	0.7750	-0.0143	-0.0503	-0.0510	0.7740	-0.0133	-0.0467	-0.0556

표 2. 계산된 월별 SSR값

month	1		2		3		4		5		6	
condi.	P. S.	SCE	P. S.	SCE	P. S.	SCE	P. S.	SCE	P. S.	SCE	P. S.	SCE
w/ init.	17.354	17.354	22.683	22.683	38.045	38.045	45.969	45.969	30.163	30.162	44.729	44.729
w/o init.	17.354	17.354	22.683	22.671	38.045	38.042	45.969	45.969	30.160	30.165	44.731	44.696
diff. (%)	0.000	0.000	0.000	0.053	0.000	0.008	0.000	0.000	0.010	-0.010	-0.004	0.074
month	7		8		9		10		11		12	
condi.	P. S.	SCE	P. S.	SCE	P. S.	SCE	P. S.	SCE	P. S.	SCE	P. S.	SCE
w/ init.	49.032	49.029	48.361	47.421	43.539	43.639	32.865	32.865	20.396	20.158	18.116	18.076
w/o init.	49.031	49.031	47.649	47.773	43.610	43.593	32.865	32.865	20.405	20.311	18.102	18.091
diff. (%)	0.002	-0.004	1.472	-0.742	-0.163	0.105	0.000	0.000	-0.044	-0.759	0.077	-0.083

註) diff.는 Pattern Search 결과를 기준했을 때 상대오차

5. 結論

PARMA(2,2) 모형에 Pattern Search 기법이나 SCE기법을 연결하여 매개변수를 최적산정해 본 결과 다음과 같은 결론을 도출해낼 수 있었다. 첫째, Pattern Search 기법이나 SCE기법 모두 PARMA모형의 매개변수 최적산정에 활용할 수 있다. 둘째, 두기법共に 비슷한 精度的 매개변수 값을 탐색하였으나 계산시간면에서 장점이 있는 Pattern Search 기법이 더욱 효과적인 것으로 판단된다. 셋째, 분석결과에서 보는 바와 같이 초기값을 산정한 후 탐색하는 것이나 임의적으로 0 을 초기값으로 준 후 탐색하는 것이나 최종결과는 비슷하므로 Pattern Search 기법을 사용하더라도 초기값 산정 단계를 생략할 수 있어 효과적이다. 넷째, PARMA(2,2)에서는 ϕ_1 , ϕ_2 , θ_1 , θ_2 모두 네개의 매개변수가 직접 탐색기법으로 최적 산정되었으므로 매개변수가 이보다 적은 PARMA(1,1), PARMA(1,2), PARMA(2,1) 모형에 직접 탐색기법을 사용해도 좋은 결과를 기대할 수 있다.

6. 參考文獻

1. Thomas, H. A. and Fiering, M. B., 1962. Mathematical synthesis of streamflow sequences for the analysis of river basins by simulation. In Design of Water Resources Systems, A. Mass et al., pp. 459-493, Harvard University Press, Cambridge, Massachusetts.
2. Yevjevich, V., 1963. Fluctuations of wet and dry years. Part I. Research data assembly and mathematical models. Hydrology Paper 1, Colorado State University, Fort Collins, Colorado.
3. Matalas, N. C., 1967. Mathematical assessment of synthetic hydrology. Jour. Water Resour. Res., 3, 4, pp. 937-945.
4. Carlson, R. F., McCormick, A. J. A., and Watts, D. G., 1970. Application of linear models to four annual streamflow series. Jour. Water Resour. Res., 6, 4, pp. 1070-1078.
5. O'Connell, P. E., 1971. A simple stochastic modeling of Hurst's law. In Mathematical Models in Hydrology, Warsaw Symposium, (IAHS Publ. 100, 1974), 1, pp. 169-187.
6. Salas, J. D., Boes, D. C., and Smith, R. A., 1980. ARMA modeling of seasonal hydrologic series. Paper submitted for publication to the Jour. of Hyd. Div., ASCE.
7. Yule, G. U., 1927. On a method of investigating periodicities in disturbed series with special reference to Wolfer's sunspot numbers. Phil. Trans., A, 226, pp. 267.
8. Walker, G. T., 1931. On periodicity in series of related terms. Proc. Roy. Soc., A, 131, pp. 518.
9. Tao, P. C., and Delleur, J. W., 1976. Seasonal and nonseasonal ARIMA models in hydrology. Proc. Am. Soc-Civil Engrs., Jou. of Hydr. Div. 102, HY10, pp. 1541-1559.
10. Rosenbrock, H. H., 1960. An automatic method of finding the greatest of least values of a function, Comput. J., pp. 175-184.
11. Hooke, R., and Jeeves, T. A., 1961. Direct search solutions of numerical and statistical problems. J. Assoc. Comput. Math., 8(2), pp. 212-229.
12. Nelder, J. A., and Mead, R., 1965. A simplex method for function minimization, Comp. J., 7(4), pp. 308-313.
13. Pronzato, L., Walter, E., Venot, A., and Lebruchec, J. F., 1984. A general-purpose global optimizer: Implementation and applications, Math. Comput. Simul., 26, pp. 412-422.
14. Duan, Q., Sorooshian, S., and Gupta, V., 1992. Effective and efficient global

optimization for conceptual rainfall-runoff models. J. of Water Resour. Res. 28(4), pp. 1015-1031.

15. Monro, J. C., 1971. Direct search optimization in mathematical modeling and a watershed model application. Tech. Memo NWS HYDRO-12, NOAA, U.S. Dep. of Commerce, Silver Spring, Md.
16. Price, W. L., 1987. Global optimization algorithms for a CAD workstation. J. of Optim. Theory Appl., 55(1), pp. 133-146.
17. Holland, J. H., 1975. Adaptation in natural and artificial system, University of Michigan Press, Ann Arbor.