

전역 최적화방법을 이용한 매개변수 추정

Estimation of Tank Model Parameters using Global Optimization Method

○강신욱¹⁾, 강경석²⁾, 이동률²⁾, 윤석영³⁾

1. 서론

유역내 강우와 유출간의 복잡한 물리적인 현상을 모의하는 유출모형은 유역의 유출특성을 나타내는 몇 개의 매개변수들로 구성된다. 이 매개변수들의 대부분은 이론적으로 계산해 낼 수 없는 미지의 값이다. 그러므로 유출모형의 적용에는 무엇보다도 적정 매개변수의 추정이 필수적이며, 이를 위하여 최적화 방법을 사용해 왔다. 최적화 방법은 지역최적화(local optimization), 전역최적화(global optimization) 방법으로 나눌 수 있으며, 최근 유전자 알고리즘, SCE-UA 방법들과 같은 전역최적화 방법을 많이 사용하고 있다.

본 연구에서는 개념적 강우-유출 모형의 전역해(global optimum)를 구하기 위한 방법으로 생물의 진화원리로부터 착상된 확률적 탐색기법인 유전자 알고리즘(genetic algorithm)과 Duan (1991)이 제시한 SCE-UA(Shuffled Complex Evolution-University Arizona)법을 이용하여 직렬 4단 탱크모형의 적정 매개변수 추정에 이들 방법들을 적용하고 그 결과를 비교 분석하였다.

유전자 알고리즘 및 SCE-UA를 적용한 직렬 4단 탱크모형을 구성하였고 인위적으로 생성한 유출량을 기준으로 두 모형을 비교하였다. 적용 대상유역으로 대청댐 지점을 선정하고 관측 수문 자료를 수집, 정리하여 적정 매개변수를 추정하였다. 그 결과 SCE-UA 방법이 유전자 알고리즘을 적용한 방법보다 적합성이 높게 나타났으며, 탱크모형의 매개변수 추정에 SCE-UA 방법을 사용하는 것이 적정 매개변수 추정방법으로 가능성이 높은 것으로 나타났다.

2. 전역 최적화방법

2.1 유전자 알고리즘

유전자 알고리즘(genetic algorithm)은 생물의 진화과정, 즉 자연선택(natural selection)과 유전법칙을 모방한 확률적 탐색기법이다. 이 기법의 가장 큰 특징은 Newton-Raphson 법과 같은 고

1) 한국건설기술연구원 수자원환경연구부 위촉연구원
2) 한국건설기술연구원 수자원환경연구부 선임연구원
3) 한국건설기술연구원 수자원환경연구부 수석연구원

전적 최적화기법과 같은 이웃해 탐색기법이 하나의 해를 운용하는데 반해, 유전자 알고리즘은 복수개의 잠재해들로 이루어진 해의 집단(population)을 운용하는 것이다. 이러한 해집단에 자연선택(selection)과 교배(cross), 진화(mutation)와 같은 유전법칙의 과정을 적용하여 세대(generation)를 진행시키면서 해 공간(solution space)을 탐색해 간다. 유전자 알고리즘은 좋은 해의 이용능력과 해 공간의 탐색 능력이 적절히 조화되어야 한다. 좋은 해의 이용만을 강조하면 조기에 수렴하여 부분 최적에 빠질 수 있고 해 공간의 탐색만을 강조하면 임의 탐색에 가까워 좋은 해를 찾아가지 못하게 된다.

유전자 알고리즘은 그 개념과 이론이 단순하고, 해의 탐색능력이 우수하여 여러 분야의 최적화 또는 의사결정문제에 다양하게 적용되고 있다. 특히 복잡한 해 공간의 탐색능력이 우수하여 변수와 제약이 많은 문제를 푸는데 적합한 기법이고, 또한 모형에 대한 유연성이 높아 제약식이나 목적함수의 변경이 용이하다는 장점이 있다.

2.2 SCE-UA 알고리즘

SCE-UA 알고리즘은 미국 국립 기상국의 NWSRFS(National Weather Service River Forecasting System)의 보정모듈 중 하나이며, 최적해를 탐색하기 위해 4가지 중요한 개념을 포함하고 있다(Duan et al., 1994); 1)확정론적이고 확률론적인 접근의 결합, 2)컴플렉스의 구조적 진화, 3)경쟁 진화(competitive evolution), 4)컴플렉스 혼합(complex shuffling). SCE-UA 알고리즘은 초기에 해가 존재하는 전체 매개변수 공간에서 표본을 생성하고 컴플렉스들로 분할한다. 각 컴플렉스는 downhill simplex 알고리즘(Nelder와 Mead, 1965)을 사용하여 독립적으로 진화된다. 진화를 거친 컴플렉스들은 다시 혼합되고, 새로운 컴플렉스들이 생성된다. 진화와 혼합의 과정은 수렴조건이 만족될 때까지 반복된다. 자세한 과정은 Duan(1991)의 연구를 참고할 수 있다.

3. 전역 최적화방법의 필요성

일반적으로 자동최적화는 지역탐색(local search) 방법과 전역탐색(global search) 방법으로 나눌 수 있다. Newton Raphson 법, Simplex 방법과 같은 지역탐색 방법은 초기값에서 시작하여 더 나은 값으로 이동하여 최적해를 찾는다. 반면 유전자 알고리즘, SCE-UA 알고리즘과 같은 전역최적화 방법은 해가 존재하는 전 영역을 확률적으로 탐색하여 최적해를 구한다.

그림 1과 그림 2는 직렬 4단 탱크모형의 응답표면(response surface)이다. 이때 목적함수는 오차제곱합을 최소화하는 함수를 사용하였다. 그림 1은 탱크모형의 매개변수 중 1단 탱크의 하단 유출공 승수 A11과 상단 유출공 높이 H12의 응답표면을 나타낸 것이고, 그림 2는 1단 탱크의 상단 유출공 승수 A12와 1단 탱크의 침투공 승수 B1의 응답표면을 나타낸 것이다. 두 그림에서 보는 바와 같이 무수히 많은 지역최적해가 존재하는 것을 볼 수 있으며, 초기값에 의존하는 최적화 방법을 사용한다면 조기에 수렴할 수 있으나 지역해에 빠질 가능성이 많다. 반면 전역최적화 방법은 전 영역에서 임의로 표본을 추출하여 최적화 과정을 반복하여 최적해를 구하므로 전역해를 구할 수 있을 가능성이 높다.

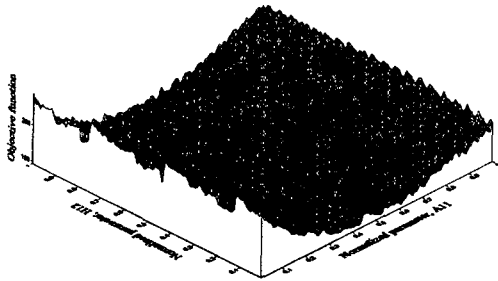


그림 1. 매개변수 A11과 H12의 응답표면

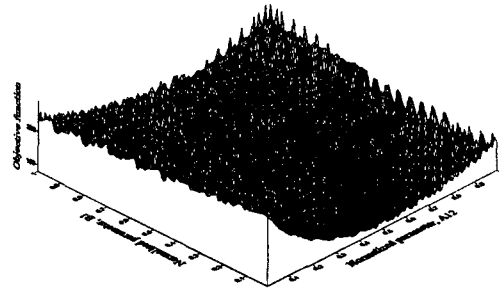


그림 2. 매개변수 A12과 B1의 응답표면

4. 전역 최적화방법에 의한 매개변수 추정

4.1 입력자료

유전자 알고리즘과 SCE-UA 방법의 수렴양상과 수렴정도를 알아보기 위해 자료를 구성하였다. 탱크모형의 주입력자료는 강수량과 증발산량이다. 증발산량의 산정은 증발계 증발량으로부터 잠재증발산량을 환산하는 방법(허유만, 1992)을 사용하였다. 증발계 증발량은 대전 기상관측소 지점의 실측 소형증발계 증발량을 사용하였고, 강수량 자료는 수자원장기종합계획(건설교통부, 2001)과 관련한 대청댐 유역의 면적강수량을 사용하였다(1981년 1월 1일 ~ 1988년 12월 31일).

4.2 매개변수 수렴 비교

매개변수 수렴 비교를 위해 유전자 알고리즘과 SCE-UA 방법을 직렬 4단 탱크모형에 연결하여 모형을 구성하였다. 탱크모형의 매개변수 상·하한값은 산업기지개발공사(1985)의 문헌에서 제시한 값으로 하였다. 그리고 수렴 비교의 객관성을 위하여 매개변수는 임의로 가정한 값을 사용하였고, 가정한 매개변수로 유출량을 모의하였다. 모의된 유출량을 실측 유출량이라 가정하여 두 알고리즘을 비교하였다. 구성된 모형들을 독립적으로 20 회씩 수행하였다.

두 알고리즘의 수행으로 매개변수가 수렴하는 과정을 나타낸 것이다(그림 3~6). 그림 3과 그림 4는 1단 탱크의 하단 유출공 승수 A11이 유전자 알고리즘과 SCE-UA 방법에 의해 수렴하는 과정을 나타낸 것이다. 그림에서 보면 두 방법 모두 초기값에 무관하게 수렴하고 있으나, 유전자 알고리즘에 의한 경우는 평균 4.7 %의 오차가 발생하였으며 SCE-UA 방법에 의한 경우는 평균 0.87 %의 오차가 발생하였다. 또한 유전자 알고리즘에 의한 수렴은 약 26,000 번의 반복에 의해 수렴이 되지만 SCE-UA의 경우는 약 10,000 번의 반복으로 얻어졌다. 그림 5와 그림 6은 1단 탱크의 침투공 승수 B1이 유전자 알고리즘과 SCE-UA 방법에 의해 수렴하는 과정을 나타낸 것이다. 역시 두 방법 모두 초기값에 무관하게 수렴하고 있으나, 유전자 알고리즘에 의한 경우는 평균 27 %의 오차가 발생하였으며 SCE-UA 방법에 의한 경우는 평균 3.6 %의 오차가 발생하였다.

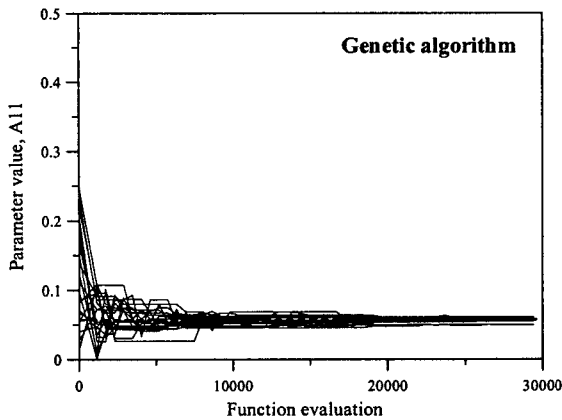


그림 3. 유전자 알고리즘에 의한 매개변수 A11의 수렴과정

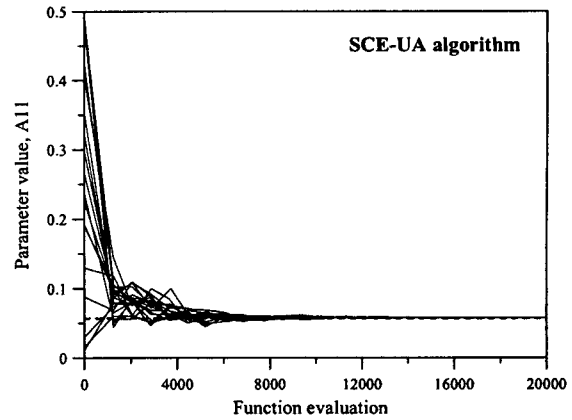


그림 4. SCE-UA 알고리즘에 의한 매개변수 A11의 수렴과정

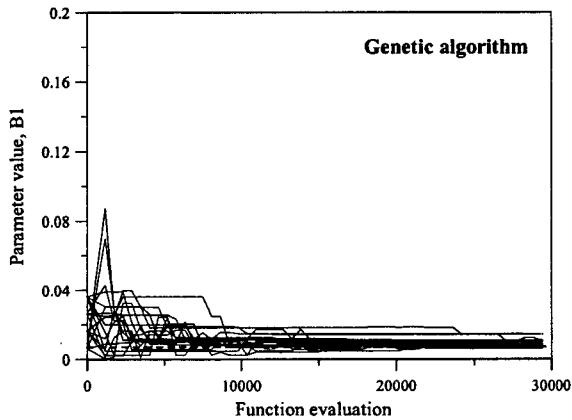


그림 5. 유전자 알고리즘에 의한 매개변수 B1의 수렴과정

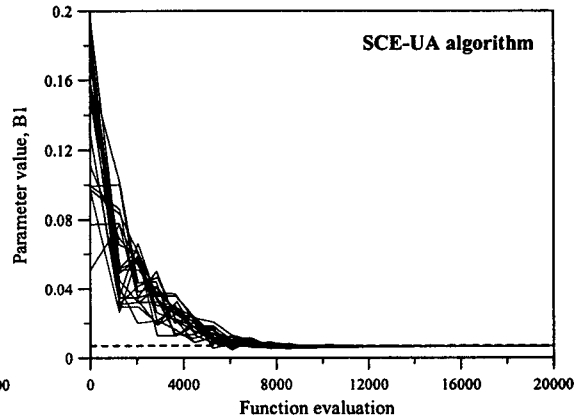


그림 6. SCE-UA 알고리즘에 의한 매개변수 B1의 수렴과정

표 1. 유전자 알고리즘과 SCE-UA 방법에 의한 매개변수별 성공횟수

매개변수	GA	SCE-UA	매개변수	GA	SCE-UA
A11	93	100	B2	1	100
A12	100	100	B3	11	91
A2	28	94	H11	47	100
A3	18	79	H12	99	100
A4	23	87	H2	13	100
B1	25	100	H3	17	84

각각의 알고리즘에 의한 모형을 100번씩 독립적으로 수행하여 최종 수렴값이 오차범위 10 % 내에 들면 성공으로 간주하여 성공한 횟수를 매개변수별로 표 1에 나타내었다. 표에서 보면 SCE-UA 방법은 1단 탱크의 매개변수 A11, A12, B1, H11, H12와 2단 탱크의 매개변수 B2, H2에 대해 모두 성공하였으며, 다른 매개변수들에 대해서도 좋은 결과를 보였다. 하지만 유전자 알고리즘에 의한 경우는 매개변수 A11에 대해서만 모두 성공하였고 1단 탱크의 매개변수 A11, A12, B1, H11, H12에 대해서는 좋은 결과를 보였다. 두 알고리즘 모두 3단과 4단 탱크의 매개변수는

좋지 않은 결과를 보였다.

4.3 대청댐 유역의 매개변수 추정

대청댐 유역에 대해 유전자 알고리즘과 SCE-UA 방법을 이용하여 직렬 4단 탱크모형의 매개변수를 추정하였다. 목적함수는 오차제곱합을 최소화하는 것을 사용하였다. 그리고 두 알고리즘을 비교하기 위한 비교 지표로 평균편차의 비율(PBIAS)과 Nash-Sutcliffe 통계량(NS)을 사용하였다. PBIAS는 '0'에 가까울수록 모의값과 관측값의 차이가 없는 것이고, NS는 '1'에 가까울수록 관측값과 모의값의 오차가 적은 것이다.

$$PBIAS = \frac{\sum_{i=1}^N (q_i^{obs} - q_i^{sim})}{\sum_{i=1}^N q_i^{obs}} \times 100 \quad (1)$$

$$NS = 1 - \frac{\sum_{i=1}^N (q_i^{sim} - q_i^{obs})^2}{\sum_{i=1}^N (q_i^{obs} - q^{mean})^2} \quad (2)$$

여기서, N 은 자료의 수, q_i^{obs} 는 관측유량, q_i^{sim} 은 모의유량, q^{mean} 은 관측유량의 평균값이다.

그림 7과 8은 대청댐의 추정 매개변수에 의한 유량수문곡선이다. 유전자 알고리즘에 의한 경우 PBIAS는 -2.8 %, NS는 0.71, SCE-UA 방법에 의한 것은 PBIAS가 0.53 %, NS는 0.71이었다. PBIAS는 SCE-UA에 의한 결과가 좋았으나 NS는 두 경우 모두 동일한 값을 나타내었다.

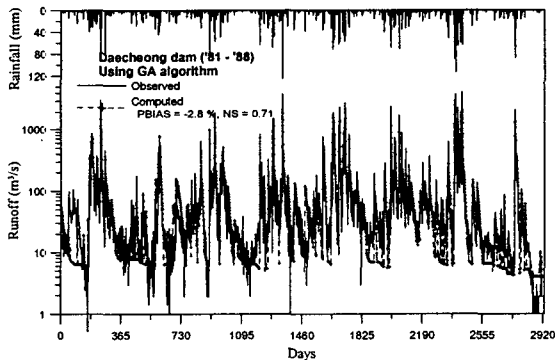


그림 7. 추정 매개변수에 의한 대청댐 유역 유량수문곡선(유전자 알고리즘)

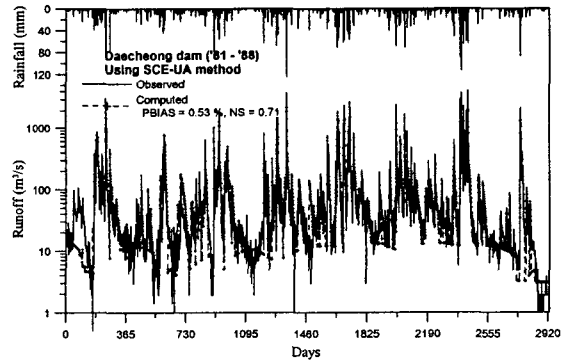


그림 8. 추정 매개변수에 의한 대청댐 유역 유량수문곡선(SCE-UA 방법)

표 2. 대청댐 유역에 대한 유전자 알고리즘과 SCE-UA 방법의 비교

년도	구분	유전자 알고리즘				SCE-UA 알고리즘			
		갈수기		홍수기		갈수기		홍수기	
		PBIAS	NSE	PBIAS	NSE	PBIAS	NSE	PBIAS	NSE
1981		-4.56	0.631	-6.66	0.784	-1.65	0.638	-4.80	0.791
1982		-22.9	0.662	-5.76	0.916	-18.6	0.706	-4.20	0.931
1983		-5.23	0.764	-5.06	0.881	-1.92	0.775	-3.13	0.879
1984		-7.54	0.648	4.61	0.893	-6.59	0.647	4.89	0.896
1985		-0.069	0.730	2.78	0.765	-1.35	0.731	1.88	0.759
1986		5.99	0.628	-3.16	0.940	4.31	0.628	-6.88	0.943
1987		8.81	0.695	5.33	0.753	6.63	0.696	4.34	0.756
1988		0.749	0.856	4.22	0.987	2.39	0.857	3.07	0.988

두 알고리즘에 의한 모의유량을 년도별로 갈수기와 홍수기로 나누어 PBIAS와 NS를 비교하였다. 갈수기는 1월~5월과 10월~12월의 기간이고 홍수기는 6월~9월 사이의 기간이다. 표에서 음영을 넣은 부분은 유전자 알고리즘과 SCE-UA 알고리즘 중에 우수한 값이다. 전체적으로 SCE-UA 알고리즘에 의한 추정 매개변수로 모의한 것이 유전자 알고리즘에 의한 것보다 우수하였다.

5. 결론

본 연구에서는 개념적 강우-유출 모형의 전역해(global optimum)를 구하기 위한 방법으로 유전자 알고리즘과 SCE-UA 방법을 이용하여 직렬 4단 탱크모형의 매개변수 추정에 이들 방법들을 적용하고 그 결과를 비교하였다.

두 모형을 비교한 결과 매개변수의 수렴양상이 SCE-UA 방법에 의한 것이 유전자 알고리즘보다 좋은 결과를 보였고 대청댐 지점에 대한 적용에서 역시 SCE-UA 방법이 유전자 알고리즘을 적용한 방법보다 적합성이 높게 나타났다.

6. 참고문헌

- 건설교통부 (2001). 수자원장기종합계획.
- 산업기지개발공사 (1985). 탱크 모형방법에 의한 유출해석법.
- 허유만 (1992). 중소유역의 일별 용수수급을 위한 하천망 모형의 개발. 박사학위논문, 서울대학교.
- Duan, Q. (1991). *A Global Optimization Strategy for Efficient and Effective Calibration of Hydrologic Models*, Ph.D. dissertation, Dept. of Hydrol. and Water Resour., Univ. of Arizona, Tucson.
- Duan, Q., Sorooshian, S. and Gupta, V.K. (1994). "Optimal use of the SCE-UA global optimization method for calibrating watershed model." *Journal of Hydrology*, Vol. 158, pp. 265-284.
- Nelder, J.A. and Mead, R. (1965). "A simplex method for function minimization." *Computer Journal*, Vol. 7, pp. 308-313.