韓國水資源學會論文集
第40卷 第9號 · 2007年 9月
pp. 677~685

다목적 유전자알고리즘을 이용한 Tank 모형 매개변수 최적화(Ⅰ): 방법론과 모형구축

Optimization of Tank Model Parameters Using Multi-Objective Genetic Algorithm (I): Methodology and Model Formulation

김 태 순* / 정 일 원** / 구 보 영*** / 배 덕 효****
Kim, Taesoon / Jung, Il Won / Koo, Bo Young / Bae, Deg Hyo

Abstract

The objective of this study is to evaluate the applicability of multi-objective genetic algorithm (MOGA) in order to calibrate the parameters of conceptual rainfall-runoff model, Tank model. NSGA-II, one of the most imitating MOGA implementations, is combined with Tank model and four multi-objective functions such as to minimize volume error, root mean square error (RMSE), high flow RMSE, and low flow RMSE are used. When NSGA-II is employed with more than three multi-objective functions, a number of Pareto-optimal solutions usually becomes too large. Therefore, selecting several preferred Pareto-optimal solutions is essential for stakeholder, and preference-ordering approach is used in this study for the sake of getting the best preferred Pareto-optimal solutions. Sensitivity analysis is performed to examine the effect of initial genetic parameters, which are generation number and population size, to the performance of NSGA-II for searching the proper paramters for Tank model, and the result suggests that the generation number is 900 and the population size is 1000 for this study.

keywords : Multi-objective genetic algorithm, Tank model, Preference ordering, Pareto optimal solution, Sensitivity analysis

요 지

본 연구의 목적은 개념적인 강우-유출모형인 Tank 모형의 매개변수를 산정하기 위한 다목적 유전자알고리즘의 적용성을 평가하는 것이다. 다목적 유전자알고리즘 기법으로는 최근에 가장 많이 사용되는 기법중의 하나인 NSGA-II를 채택하여 Tank 모형과 결합하였으며, 4가지 목적함수(유출용적오차, 평균제곱근 오차, 고수유량 평균제곱근 오차 및 저수유량 평균제곱근 오차)값을 최소화하는 형태의 목적함수를 적용하였다. NSGA-II는 목적함수의 개

* 연세대학교 사회환경시스템공학부 BK21연구교수
BK21 Lecturer, School of Civil and Environmental Engineering, Yonsei Univ., Seoul 120-749, Korea
(e-mail: chaucer@yonsei.ac.kr)
** 세종대학교 토목환경공학과 박사과정
Ph.D Candidate Student, Dept. of Civil and Environmental Engrg., Sejong Univ., Seoul 143-747, Korea
(e-mail : bobilwon@paran.com)
*** 남원건설엔지니어링 사원
Employee, Namwonkeonseol Engineering Co., Ltd., Kyungido Anyang 431-060, Korea
(e-mail : boyoungkoo@nate.com)
**** 세종대학교 물자원연구소 · 토목환경공학과 부교수
Associate Prof., Dept. of Civil and Environmental Engrg., Sejong Univ., Seoul 143-747, Korea
(e-mail : dhbae@sejong.ac.kr)

수가 많아지면 한 번의 실행에 의해 굉장히 많은 수의 파레토최적해를 구하는 단점을 가지고 있기 때문에 구해진 파레토최적해 중에서 어떤 해가 최우선해 인지를 결정해야 할 필요가 있으며, 이러한 고차원적인 의사결정을 위하여 선호적순서화(preference ordering) 기법을 적용하였다. NSGA-II를 이용하여 Tank 모형의 매개변수를 추정할 때 초기조건이 최적화과정에 미칠 수 있는 영향을 최소화하기 위해 세대수(generation number)와 개체군의 크기(population size)에 대한 민감도분석을 수행하였다. 분석결과 Tank 모형의 매개변수 최적화를 위한 세대 수와 개체군 크기의 초기값을 각각 900번과 1000개로 선정하는 것이 적합한 것으로 나타났다.

핵심용어 : 다목적 유전자알고리즘, 탱크모델, 선호적순서화, 파레토 최적화, 민감도 분석

1. 서 론

강우-유출모형(rainfall-runoff model)의 목적 중에 하나는 강우-유출이라는 자연현상을 이해하고, 이를 통해 다양한 수자원 문제에 요구되는 신뢰성 높은 유출량 자료를 산정하는데 있다. 이와 같은 유출모형은 복잡한 실세계를 단순한 형태로 개념화하여 표현하는데, 이러한 과정에서 많은 매개변수들이 필요하게 되며, 정확하고 신뢰성 있는 모형의 매개변수 산정은 강우-유출모형을 이용한 유출분석에 있어서 매우 중요한 과정이라 할 수 있다.

매개변수를 결정하는 방법에는 크게 수동보정법(manual calibration)과 자동보정법(automatic calibration)이 있다. 일반적으로 수동보정법은 전문적인 지식과 숙련된 기술이 요구되며 사용자의 임의성이 개입될 여지가 많은 반면에, 자동보정법은 수치적인 알고리즘을 이용하여 유출모형의 매개변수를 비교적 효율적이고 객관적으로 추정하기 때문에 최근에는 여러 가지 자동보정법을 이용한 다양한 연구가 진행되고 있다(Nelder and Mead, 1965; Gupta et al., 1998; Duan, 2002; Tolson and Shoemaker, 2007).

자동보정법은 수정보정법에 비해서 효율성과 객관성이라는 장점을 지니고 있지만, 탐색구간이나 허용오차와 같은 초기조건이 어떻게 주어지느냐에 따라서, 국부최적해(local optimum)로 수렴할 수 있다는 점과 유출모형의 적합도를 결정하는 목적함수로 하나의 목적함수를 사용할 경우, 모형 보정에 사용한 사상(event)과 성질이 상이한 다른 사상에 산정된 매개변수를 사용할 경우 수문사상을 모의하는데 한계가 있을 수 있다.

앞서 언급한 자동보정법의 단점을 해결하기 위해서 여러 가지 형태의 연구들이 진행되고 있는데, 국부최적해로의 수렴 문제는 유전자알고리즘(Genetic Algorithm, GA; Holland, 1975)이나 SCE-UA(Shuffled Complex Evolution University of Arizona; Duan, 1991) 등과 같은 전역 최적화기법들을 이용하여 해결하려는 연구들이 진행되어 왔으며, 이러한 전역 최적화방법들은 최적화기법을 통해서 산출되는 해가 국부최적해로 수렴되지 않도록 하면서 동시에 전역최적해(global optimum)를 비교적 잘 추정한다고 알려져 있다(Vrugt et al., 2003). 또한, 단일 목적함수만을 사용함으로 인해서 생기는 문제를 해결하기 위하여 Gupta et al.(1998)과 Madsen (2000) 등은 다중목적함수를 적용하는 방법을 제안하기도 하였으며, Vrugt et al.(2003)은 SCE-UA 방법을 이용하여 매개변수의 다목적 최적화를 수행할 수 있는 MOSCEM(Multi-Objective Shuffled Complex Evolution Metropolis) 기법을 제안하였고, Khu and Madsen (2005)은 다목적 유전자알고리즘(Muti-Objective Genetic Algorithm, MOGA)을 이용하여 비선점해를 구한 후, 다시 선호적순서화(preference ordering) 방법을 이용하여 더 강한 순위를 부여하는 방법으로 다수의 최적매개변수 셀 중에서 유출분석에 보다 적합한 해를 선택할 수 있는 방법을 제안하였다.

4단 탱크모형은 국내에서 장기유출분석기법의 하나로 널리 이용되고 있지만, 단순한 모형구조에 비해 매개변수의 수가 많아서 매개변수 보정이 어렵고 따라서 유출결과의 불확실성이 크며, 유출량의 계절별 편중이 심하기 때문에 유출모형의 적합도를 하나의 목적함수로만 고려해서는 평가하기 어렵다는 문제점을 가지고 있다. 본 연구에서는 이를 해결하기 위한 방안의 하나로써, 다목적 유전자알고리즘 기법중의 하나인 NSGA-II (Non-dominated Sorting Genetic Algorithm-II; Deb et. al., 2002)와 선호적순서화 방법을 적용하여 Tank 모형의 매개변수를 추정하는 방법론을 제시하고 적용 결과를 분석하고자 한다.

본 연구의 목적은 4단 Tank 모형의 매개변수를 보다 신뢰성 있게 추정하기 위한 방법을 제시하고 적용성을 평가하는 것으로 내용에 따라, (a) 매개변수 추정 방법론 구축과 NSGA-II의 세대수 및 개체군 크기 결정을 위한 민감도 분석 (b) 소양강댐 유역에서의 적용 및 타 최적화기법과의 비교평가, 두 개의 논문으로 구성하였

다. 본 논문은 그 첫 번째 논문으로, 2장에서는 본 연구에 사용된 이론들을 정리하였으며, 3장에서는 Tank 모형 매개변수 추정 방법론, 4장에서는 민감도 분석에 의한 NSGA-II 기법의 세대수 및 개체군 크기 분석을 제시하였다.

2. 적용 이론

2.1 유전자 알고리즘(Genetic Algorithm)과 NSGA-II

유전자 알고리즘은 1960년대에 Holland에 의하여 창안되었고 Holland와 그의 학생 및 동료들에 의하여 1960년대와 1970년대 미시간대학에서 학문으로써의 체계가 정립되었다(Holland, 1975; Goldberg, 1989). 최근에는 하나의 목적함수에 대해서만 최우수 개체를 찾는 방법에서 탈피하여 여러 가지 목적함수를 한 번의 실행으로 최적화 할 수 있는 형태의 다목적 유전자알고리즘이 많이 개발되고 있으며, 각각의 목적함수값을 그대로 사용하고 산정된 목적함수간의 순위(rank)를 정하여 전체적인 목적함수의 효율성을 따지는 파레토지배(Pareto-dominance) 이론을 적용한 기법들이 많이 개발되고 있다(Van Veldhuizen and Lamont, 2000).

본 연구에서 적용한 다목적 유전자알고리즘 기법인 NSGA-II는 기존 NSGA(Srinivas and Deb, 1994)의 단점을 보완한 기법으로, 비선점 정렬방법(non-dominated sorting)의 복잡도(complexity)가 감소하였으며, 사용자가 sharing parameter를 정의할 필요가 없이 군집거리(crowding distance)를 도입하여 각각의 자원을 보다 효율적으로 배분(niching)되도록 하였고, 현재 세대(genaration)의 최적해를 다음 세대로 넘겨주는 엘리티즘(elitism)이 적용되었다. 엘리티즘 적용으로 인해 수렴속도를 높였고, 군집거리 개념을 적용하여 순위를 결정하는 속도가 다른 다목적 유전자알고리즘 기법들보다 빠르다(Deb, 2002). 또한 목적함수의 형태나 연속성 등에 제약을 받지 않고, 국부최적해로 수렴할 가능성 또한 상당히 줄일 수 있는 기법으로 알려져 있다. Fig. 1은 NSGA-II의 개념도를 나타낸 것으로, 초기 개체군이 결정되면 종결조건을 만족하는지의 가부를 결정하게 된다. 만일 만족하지 못할 경우 토너먼트 선택과 재생산을 통해 새로운 자손세대를 발생시키게 된다. 자손세대는 교배나 돌연변이를 통해 조합되고, 각각의 자손들에 대한 파레토 순위를 결정하게 된다. 군집거리 계산에 의해 현 세대의 정보를 다음 세대로 전달하는 과정을 통해 새로운 개체군을 생성하게 된다. 새로운 개체군은 초기 개체가 되어 종결조건을 만족하는 지를 평가하는 과정을 계속적으로 거치게 된다.

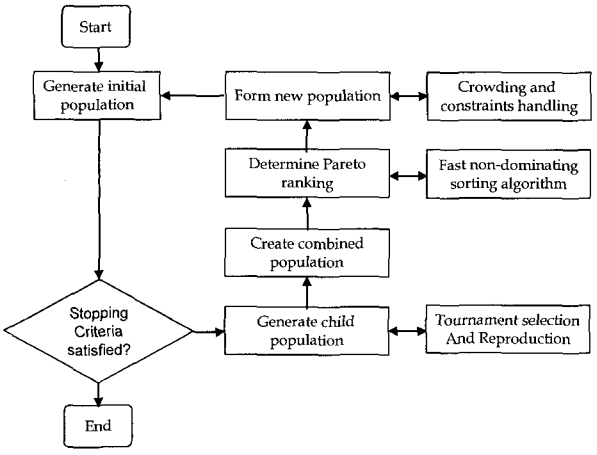


Fig. 1. Schematic diagram of NSGA-II (Khu and Madsen, 2005)

2.2 선호적순서화 (preference ordering)

파레토최적화 기법을 사용할 경우 목적함수의 개수가 많아질수록 필연적으로 파레토최적해의 개수가 많아지는 단점이 발생하게 된다. 이런 현상은 파레토지배가 가지는 원리를 생각하면 당연한 것으로, 목적함수가 많아지면 그 만큼 서로 비선점(non-dominated) 관계에 있는 해의 개수가 늘어나게 되는 것이다. 따라서 다수의 파레토최적해 중 사용자가 원하는 파레토최적해를 선택하는 과정이 필요하게 되며, 이를 위해서 본 연구에서는 선호적순서화(Das, 1999)를 적용하였다. Khu and Madsen(2005)은 선호적순서화와 NSGA-II를 강우-유출모형인 MIKE11/NAM 모형의 매개변수의 추정에 적용하여, 390여개의 파레토 최적해 중에서 최종적으로 3개의 우수한 파레토 최적해를 추출하였다.

선호적순서화 방법론은 다음과 같다. Table 1은 세 개의 목적함수에 대한 선호적순서화 결과를 나타낸 것이다. 여기서 [2,1]과 같은 괄호 내의 숫자들은 각각 조합된 목적함수의 개수와 적용 순서의 효율성에 따라 결정되는 조합의 순번을 나타낸다. Table 1에서 매개변수셀 A는 B에 비해, 최소화문제라고 가정한 경우, 목적함수 1(OB1)과 목적함수 2(OB2) 측면에서는 우수하나 목적함수 3(OB3) 측면에서는 B가 더 좋은 결과를 보이고 있으므로 A와 B 모두 파레토 최적해가 된다. 다른 매개변수셀 또한 이와 같은 관점에서 모두 서로 비선점 관계에 있음을 알 수 있다. 따라서 이 경우 총 12개의 매개변수셀이 추정되는 것이다.

이러한 매개변수셀 중에서 더 적합한 셀을 선정하기 위해 세 가지 목적함수를 각 두 개의 목적함수 조합으로 만들고 이에 대한 파레토최적해를 산정한다(Table 1). Fig. 2는 두 개씩의 목적함수 조합에 따른 매개변수셀들을 2차원 평면에 도시한 것이다. 목적함수 1과 목

적함수 2의 조합으로부터 H, C, A의 파레토프론트(Pareto-front)가 3차원 조합과 공통적인 파레토최적해를 갖는 것으로 나타났다. 다른 두 조합인 목적함수 2와 목적함수 3의 조합은 각각 8개(J, I ,H, G, F, E, B, A)와 4개(H, K ,L, J)가 파레토최적해인 것으로 나타났다. 이러한 분석을 통하여 다음과 같은 결론을 내릴 수 있다.

① m차원에서의 모든 파레토최적해가 m-1차원의 조합에서도 파레토최적해는 아니다.

② m차원의 특정 해가 하나의 m-1차원의 조합에서 파레토최적해의 조건을 갖추었다 하더라도 반드시 다른 m-1차원에서도 파레토최적해가 되는 것은 아니다. 예를 들어 점 C는 목적함수 1과 목적함수 2의 조합에서는 파레토최적해이지만 다른 두 개의 조합에서는 최적해가 아닌 것을 알 수 있다.

③ 점 A, H, J와 같이 2개 이상의 m-1차원에서 파레토최적해가 되는 해가 존재한다.

따라서 m차원뿐 아니라 모든 m-1차원에서도 파레토최적해로 나타나는 매개변수셀 H가 최우선해(most preferred solution)가 되는 것으로 이러한 절차를 통해 최우선해를 선정하는 기법을 선호적 순서화라 한다. 선호적 순서화에 관한 보다 자세한 절차에 관심이 있을 경우, Das(1999)의 연구가 좋은 참고자료가 될 수 있을 것이다.

2.3 Tank 모형

Tank 모형은 복잡한 강우-유출 현상을 계산하는 알고리즘이 다른 강우-유출 모형에 비해 간단하지만 비교적 정확한 유출모의를 수행하는 장점을 가지고 있다(Yokoo et al., 2001). 일반적으로 장기유출분석에 적용되는 4단 직렬탱크(Fig. 3)의 매개변수는 총 16개 이며, 이 중 최적화기법을 적용하여 추정해야 하는 매개변수는 각 단 탱크의 저류고를 제외한 12개이다. 이러한 매개변수들은 유출분석시 내부적으로 서로 영향을 미치고 있어 매개변수 추정이 어렵다.

Tank 모형에서 상단의 두 탱크는 빠른 강우-유출응답을 표현하며, 하단의 두 탱크는 느린 응답을 모의한다. Fig. 3에서 A(1/dat)는 유출공의 계수이며, B(1/day)는 하단탱크로의 침투공 계수를 나타낸다. 또한, H(mm)

Table 1. Objective function (OF) values and application result of the Pareto preference ordering mehtod

Parameters set	OF1	OF2	OF3	Pareto-optimal point			
				[1,2,3]	[2,1]	[2,2]	[2,3]
A	7.32	10.89	4.08	○	○	○	×
B	15.79	12.99	2.65	○	○	×	×
C	2.74	14.9	4.43	○	○	×	×
D	7.84	15.72	3.16	○	×	×	×
E	15.57	17.46	1.76	○	○	×	×
F	3.45	18.98	1.52	○	○	×	×
G	13.09	19.61	1.27	○	○	×	×
H	0.32	35.00	0.75	○	○	○	◎
I	9.96	52.47	0.44	○	○	×	×
J	3.71	70.77	0.25	○	○	○	×
K	1.17	72.78	0.74	○	○	×	×
L	2.12	96.08	0.45	○	○	×	×

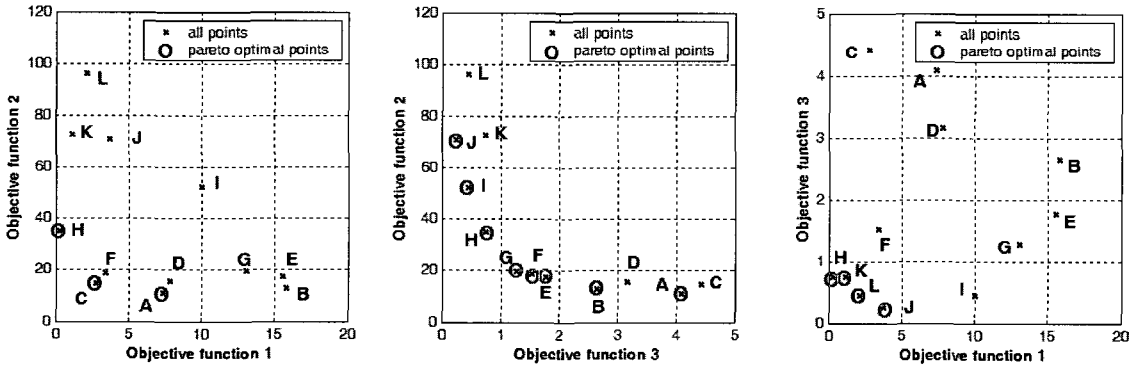


Fig. 2. Two-dimensional Pareto plots for the three-objective function problem (Khu and Madsen, 2005)

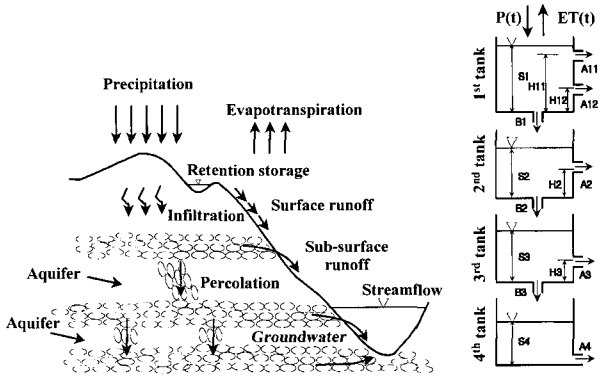


Fig. 3. Schematic of the Tank model used in this study

는 각 탱크의 저면으로부터 유출공까지의 높이이며, S(mm)는 저류고를 의미한다. 일반적으로 각 단의 A, B 및 Z는 유출계산에서 상수로 적용되며, S는 계산 과정에 따른 저류량의 변화를 나타내는 것으로 모형계산 시 초기값만을 지정하게 되어 있다. 초기저류고는 저류탱크의 저류고가 안정화되는 기간을 약 2년 정도로 가정하여 유출결과의 분석 시 최초 2년 기간을 제외하는 것이 일반적이다(건설교통부, 2000). Eq. (1)~Eq. (4)는 Tank 모형의 계산 알고리즘을 나타낸 것이다.

$$R(x,n) = A(x)(S(x,n) - H(x)) \quad (S(x,n) > H(x)) \\ = 0 \quad (S(x,n) \le H(x)) \tag{1}$$

$$I(x,n) = B(x)S(x,n) \tag{2}$$

$$S(x,n+1) = S(x,n) - R(x,n)\triangle t - I(x,n)\triangle t + P(n+1)\triangle t \quad (x = 1) \\ = S(x,n) - R(x,n)\triangle t - I(x,n)\triangle t + P(x-1,n)\triangle t \quad (x \neq 1) \tag{3}$$

$$Q(n) = \sum_{x=1}^{4} R(x,n) \tag{4}$$

여기서 x는 상부로부터의 탱크 개수, n는 유출계산 시작일로부터 현재까지의 일 수, Δt는 유출계산간격, A(x)는 x단 탱크의 유출공계수, B(x)는 x단 탱크의 침투공계수, H(x,n)는 n번째 일의 x번째 저류고, I(x,n)는 n번째 일의 x번째 탱크 침투고, P(n)는 n번째 일의 강우량, Q(n)는 n번째 일의 총유출량, R(x,n)는 n번째 일의 x단 탱크 유출량, Z(x) = x단 탱크의 유출공 높이를 나타낸다.

3. Tank 모형 매개변수의 최적화 방법론

본 연구에서는 다목적 유전자알고리즘인 NSGA-II를 이용하여 Tank 모형의 매개변수를 추정하기 위한 방법을 Fig. 4와 같이 구축하였다. 우선 Tank 모형의 소스코드를 수정하여 NSGA-II 기법과 결합한다. 그 다음 다양한 수문사상을 모의할 수 있는 매개변수의 추정이 가능하도록 적합한 목적함수들을 결정한다. 일반적인 유전자알고리즘은 초기조건인 교배 및 돌연변이 발생 확률이나 세대수(generation number) 및 개체군 크기(population size)에 따라 최적화 결과에 미치는 영향이 크다. 따라서 세대수와 개체군 크기에 따른 민감도분석을 통해 전역최적해를 추정할 수 있는 적절한 값들을 결정해야 한다. 세대수와 개체군 크기가 결정되면 실제 적용을 통해 파레토최적해들을 추정하게 된다. 추정된 최적해들은 선호적순서화기법을 통해 최적해들의 수를 줄여나가게 되며, 최종적으로 최우선해를 선정하고 Tank 모형을 검정하게 된다.

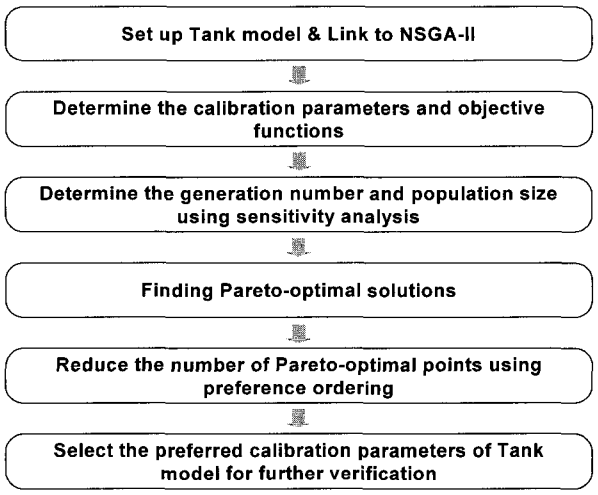


Fig. 4. The calibration algorithm of Tank model parameters with NSGA-II

3.1 제약조건 설정

본 연구에서는 NSGA-II를 이용하여 추정된 Tank 모형의 매개변수들이 물리적인 값의 범위를 벗어나지 않도록 제약조건을 설정하였다. NSGA-II 뿐만 아니라 모든 유전자 알고리즘 기법이 난수 발생(random number generation)에 의존하므로 발생된 난수가 유출

Table 2. The upper and lower boundary of Tank model parameters

Parameter	Description	Upper boundary	lower boundary
A11	1st upper side outlet	0.5	0.1
A12	1st lower side outlet		
B1	1st bottom outlet		
A2	2nd side outlet	0.1	0.03
B2	2nd bottom outlet	0.1	0.01
A3	3rd side outlet	0.01	0.005
B3	3rd bottom outlet		
A4	4th side outlet	0.01	0.0005
H11	Height of 1st upper side outlet	5	60
H12	Height of 1st lower side outlet		
H2	Height of 2nd side outlet	0	50
H3	Height of 3rd side outlet	0	30

모의에 부적합한 해를 산정하지 않도록 하여야 한다. 일본의 Sugawara(1995)는 Tank 모형을 일본의 많은 유역에 적용하여 각 매개변수가 가지는 물리적인 범위를 제시하였다.

이는 일본과 지형 혹은 기후에 상당한 유사성을 가지고 있는 국내 유역에 대하여서도 적용성이 높을 것으로 판단된다. 따라서 Table 2에서와 같이 Sugawara (1995)가 제시한 Tank 모형의 매개변수 범위를 난수 발생시에 상·하한계 값으로 설정하였다. 또한 유출공 계수가 두 개인 1단 탱크에서는 Fig. 3에서 보는 바와 같이 상부유출공의 높이를 나타내는 H11이 하부유출공 높이인 H12보다 항상 큰 값을 가져야 한다. 따라서 항상 H11이 H12에 비해 크도록 제약조건을 적용하였다.

3.2 목적함수의 선정

본 연구에서는 Tank 모형이 관측수문곡선을 전체적으로 잘 모의하도록 하기 위한 목적함수로 유출용적오차(Volume Error, VE), 평균제곱근오차(Root Mean Square Error, RMSE), 저수유량 평균제곱근오차(Low Flow RMSE, L_RMSE), 고수유량 평균제곱근 오차(High Flow RMSE, H_RMSE) 등 4개를 선정하였다. 유출용적오차는 유출량에 대한 모의 정확도를 향상시키기 위해 선정하였으며, 평균제곱근오차는 전체기간에서의 관측치와 모의치의 오차를 줄이고 적합성을 높이기 위해 채택하였다. 또한 홍수기에 발생하는 큰 고수유량(peak flow) 모의와 갈수기 동안의 저수유량(low flow)에 대한 모의 정확도를 향상시키고자 저수 및 고수유량 평균제곱근 오차를 선택하였다.

본 연구에서 적용한 저수 및 고수유량 평균제곱근 오차는 Madsen(2000)이 제시한 사상(event)을 고려하는 방법을 적용하였다. Madsen(2000)은 저수 및 고수유량을 구분하는 기준으로 특정 유량값을 이용하지 않고, 저수유량이 지속되는 사상 혹은 고수유량이 발생하는 사상들을 고려한 방법을 Eq. (7) ~ Eq. (8)과 같이 제시하였다. 이것은 저수 혹은 고수유량 사상들의 평균적인 적합성을 나타내는 기준이다. 다음은 각 목적함수의 수식을 나타낸 것이다. 여기서 $Q_{obs,i}$와 $Q_{sim,i}$는 각각 시간 i에서의 관측유량과 모의유량이다. n은 보정기간 전체의 자료개수, M_P는 고수유량 사상(event)의 수, M_L은 저수유량 사상의 수, n_j는 j번째 고수 혹은 저수 자료의 개수, θ는 모형 매개변수셀을 나타낸다.

$$VE(\theta) = \left| \sum_{i=1}^{n} [Q_{obs,i} - Q_{sim,i}(\theta)] \right| / \left(\sum_{i=1}^{n} Q_{obs} \right) \quad (5)$$

$$RMSE(\theta) = \left[\frac{1}{n} \sum_{i=1}^{n} [Q_{obs,i} - Q_{sim,i}(\theta)]^2 \right]^{1/2} \quad (6)$$

$$L_RMSE(\theta) = \frac{1}{M_P} \sum_{j=1}^{M_P} \left[\frac{1}{n_j} \sum_{i=1}^{n_j} [Q_{obs,i} - Q_{sim,i}(\theta)]^2 \right]^{1/2} \quad (7)$$

$$H_RMSE(\theta) = \frac{1}{M_L} \sum_{j=1}^{M_L} \left[\frac{1}{n_j} \sum_{i=1}^{n_j} [Q_{obs,i} - Q_{sim,i}(\theta)]^2 \right]^{1/2} \quad (8)$$

4. 민감도 분석에 의한 세대 수 및 개체군 크기 분석

유전자 알고리즘은 전역최적해를 찾을 수 있는 기법이지만 교배나 돌연변이발생 확률, 초기 매개변수의 무작위 시작 수(random seed), 세대수 (generation number) 및 하나의 개체군에 몇 개의 개체를 지니고

있는가를 나타내는 개체군 크기 (population size) 등의 초기조건에 따라 최적화 결과가 상이하게 나타나는 특성이 있다. 특히 이 중에서도 세대수와 개체군 크기에 특히 민감한 변화를 보이는데, 이러한 세대수와 개체군 크기를 크게 하여 계산 반복(iterarion) 횟수를 증가시킨다 할지라도 반드시 그 결과가 향상되지 않는 특징을 가지고 있다. 이것은 실제 자연현상의 진화(evolution)에서 차세대 개체가 반드시 우성을 가지는 것은 아닌 것으로 쉽게 설명할 수 있다. 대부분의 기존 유전자 알고리즘 연구에서 이러한 초기조건을 명시하지 않거나 임의적으로 선택한 값을 제시하였는데, 전역최적해를 추정하기 위해서는 이러한 특성에 대한 분석이 선행되어야만 한다고 판단되므로, 본 연구에서는 초기조건 중 영향이 크다고 판단되는 세대수 및 개체군 크기에 대한 민감도분석을 수행하였다.

Fig. 5는 NSGA-II의 목적함수로 단일 목적함수인 RMSE를 사용하여 수행한 최적화의 결과를 도시한 것으로, 세로축은 목적함수인 RMSE를 나타내며, 삼각형 표시들은 각 세대수 및 개체군 크기에 따른 목적함수 값을 나타낸다. 세대수 및 개체군 크기 모두를 100부터 1000까지 100개의 단위로 그 조합을 달리하며 Tank 모형의 평균제곱근 오차의 변화를 살펴본 결과, 대체적으로 세대수 및 개체군 크기의 증가에 따라 목적함수가 점차적으로 감소하는 결과를 나타내었다. 그러나 초기조건에 최대의 값을 부여한 세대수 1000과 개체군의 크기 1000의 모의 결과와 세대수 900과 개체군의 크기 1000의 모의에 의한 결과값을 비교하면, 세대수 900에 의한 RMSE가 오히려 약 0.4×10^{6}의 차이로 더 우수한 것으로 나타났다.

각 개체군 크기 및 세대수에 따라 나타나는 영향을 좀 더 상세히 살펴보기 위하여 Fig. 5의 민감도 분석

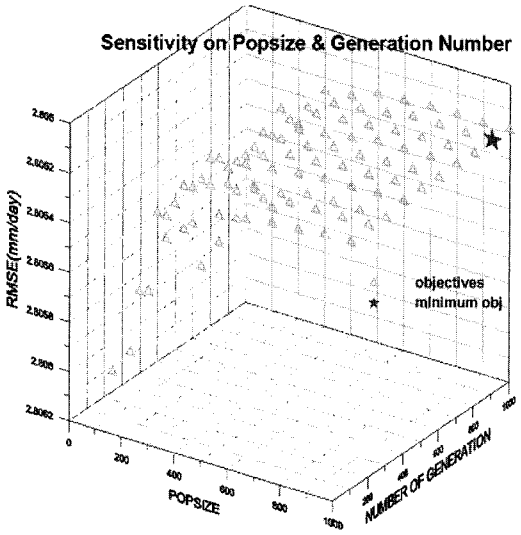
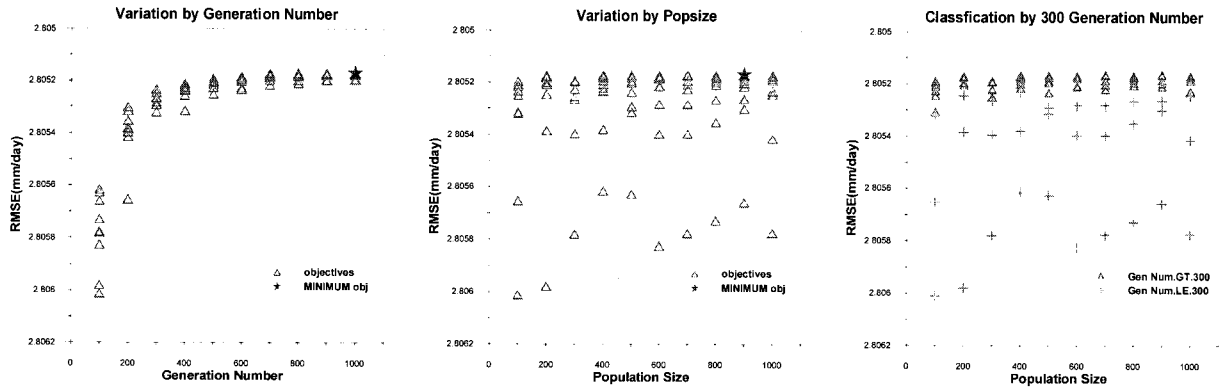


Fig. 5. Change of objective function value according to generation number and population size (3-D)

결과를 각각의 초기조건과 목적함수의 2차원 평면에 도시하였다(Fig. 6). Fig. 6에서 (a)의 가로축은 세대수를 나타내며 세로축은 목적함수를 나타낸다. 동일 세대수에 있는 목적함수들은 각각의 개체군 크기에 따라 분포하는데 개체군 크기가 커질수록 목적함수 값이 작게 나타나는 것을 볼 수 있다. 세대수가 500 이상일 때에는 비교적 군집하는 분포를 보이는데 이것은 세대수를 500 이상으로 적용하였을 때에 개체군 크기에 따른 영향이 줄어들어 목적함수들의 편차가 현저하게 줄어드는 것으로 나타났다. 또한 Fig. 6(b)와 같이 개체군 크기에 따른 변화를 살펴본 결과, 세대수의 증가에 따라 나타나는 목적함수의 뚜렷한 군집성에 비하여 일정한 경향이 나타나지는 않았지만, 개체군 크기가 클수록 보다 개선되는 경향을 나타냈다. Fig. 6(c)와 같이 같은 개체군



(a) RMSE & generation number (b) RMSE & population size (c) RMSE & population size under generation number (=300)

Fig. 6. Change of objective function value according to generation number and population size (2-D)

크기에서 세대 수가 300 이하일 때에는 정도가 비교적 좋지 못한 해를 산정하는 것으로 나타났다. 결과적으로 세대수 500 이상, 개체군 크기 400 이상일 경우 NSGA-II에서 산정되는 목적함수의 차이는 미세한 것으로 나타났다.

따라서, 본 연구에서는 단일목적함수인 RMSE값을 최소화 시키는 것을 기준으로, 세대수 900회와 개체수 1000개인 경우를 채택하였다. 유전자 알고리즘이 해를 탐색하는데 있어서 적절한 세대수와 개체수를 결정하는 것은 매우 중요한 절차로써, 특히 Tank 모형과 같이 모형의 구조가 비교적 단순한 경우에는 민감도분석과 같은 절차를 수행하여 가장 적절한 개수의 세대수와 개체수를 결정하는 절차가 반드시 필요하다고 할 수 있다.

다목적 유전자알고리즘의 경우 세대수는 그 크기를 증가시킨다고 해서 반드시 해의 정확도가 항상 향상되는 것은 아니며, 개체수의 크기를 증가시킬 경우에는 계산시간이 급격하게 늘어나며 해의 수렴성 또한 떨어지게 되는 것이 일반적이다. 결과적으로 이러한 결정은 공학적으로 유리하지 못한 선택이 될 수도 있으므로 실제 최적화 수행 전에 이러한 유전자 알고리즘의 특성에 대한 고찰이 본 연구에서와 같이 충분히 선행되어야할 것이다.

5. 결 론

본 연구에서는 다목적 최적화방법인 NSGA-II를 이용하여 개념적 유출모형인 Tank 모형의 매개변수를 보다 신뢰성 있게 추정할 수 있는 방법을 제안하였다. 관측유량에 대한 모의유량의 적합도를 향상시키기 위해 유출용적오차(VE)와 평균제곱근오차(RMSE)를 목적함수로 선정하였으며, 홍수기 첨두유량 모의와 갈수기의 저수유량 모의 정확도 개선을 위해 저수유량 및 고수유량 평균제곱근오차(L & H_RMSE)도 목적함수로 고려하였다. NSGA-II는 목적함수의 수가 증가될수록 더욱 많은 수의 파레토최적해를 추정하므로, 구해진 파레토 최적해 중에서 어떤 해가 가장 우수한 해인지를 결정하기 위해 선호적순서화 기법이 적용되었다.

NSGA-II의 Tank 모형 매개변수 추정 시 초기조건이 전역최적해 추정에 미칠 수 있는 영향을 최소화하기 위해 세대수와 개체군 크기에 대한 민감도분석을 수행한 결과 계산횟수가 늘어날수록 즉 세대수와 개체군 크기를 증가시킬수록 목적함수가 개선되는 것으로 나타났다. 그러나 전역최적해 산정의 가능성을 향상시키기 위해 세대수와 개체군 크기를 계속해서 증가시킬 경우, Tank 모형의 최적화 시에 해의 개선되는 정도에 비하여 과다한 계산 시간이 소요될 수 있으므로 최단의 계산시간을 통해서 최고의 효율을 얻을 수 있는 초기조건을 결정하는 절차가 필요하다. 본 연구에서는 이와 같은 다목적 유전자알고리즘의 초기조건으로 최적 세대수 900개 및 개체군 크기 1000개를 선정하였다.

감사의 글

본 연구는 21세기 프론티어연구개발사업인 수자원의 지속적 확보기술개발사업단의 연구비지원(1-9-3)에 의해 수행되었습니다.

참 고 문 헌

건설교통부 (2000). **수자원장기종합계획(water vision 2020) 보고서.**

Das, I. (1999). "A preference ordering among various Pareto optimal alternatives", *Structural and Multidisciplinary Optimization*, Vol. 18(1), pp. 30-35.

Deb, K., Pratap, A., Agarwal, S., and Meyarivan, T. (2002). "A fast and elitist multiobjective genetic algorithm: NSGA-II", *IEEE Transactions on Evolutionary Computation*, Vol. 6(2), pp. 182-197.

Duan, Q. (1991). *A global optimization strategy for efficient and effective calibration of hydrologic models*, PhD dissertation, University of Arizona, Tucson.

Duan, Q. (2002). "Global optimization for watershed model calibration", *Calibration of Watershed Models*, Q. Duan, H. V. Gupta, S. Sorooshian, A. N. Rousseau, and R. Turcotte, eds., American Geophysical Union.

Goldberg, D.E. (1989). *Genetic algorithms in search, optimization & machine learning*, Addison Wesley, Reading, Massachusetts.

Gupta, H.V., Sorooshian, S., and Yapo, P.O. (1998). "Toward improved calibration of hydrologic models: Multiple and noncommensurable measures of information", *Water Resources Research*, Vol. 34(4), pp. 751-763.

Holland, J.H. (1975). *Adaptation in natural and artificial systems*, University of Michigan Press.

Khu, S. T., and Madsen, H. (2005). "Multiobjective calibration with Pareto preference ordering: An application to rainfall-runoff model calibration",

Water Resources Research, Vol. 41(3), W03004, doi:10.1029/2004WR003041.

Madsen, H. (2000). "Automatic calibration of a conceptual rainfall-runoff model using multiple objectives", *Journal of Hydrology*, Vol. 235(3-4), pp. 276-288.

Nelder, J.A., and Mead, R.A. (1965). "Simplex method for function minimization", *Computer Journal*, Vol. 7, pp. 308-313.

Srinivas, N., and Deb, K. (1994). "Multiobjective optimization using nondominated sorting in genetic algorithms", *Evolutionary Computation*, Vol. 2(3), pp. 221-248.

Sugawara, M. (1995). "Tank model", *Computer Models of Watershed Hydrology*, Singh, V.J., ed., Water Resources Publications, Highlands Ranch, CO.

Tolson, B.A., and Shoemaker, C.A. (2007). "Dynamically dimensioned search algorithm for computationally efficient watershed model calibration", *Water Resources Research*, 43(1), W01413, doi:10.1029/2005WR004723.

Van Veldhuizen, D.A., and Lamont, G.B. (2000). "Multiobjective evolutionary algorithms: Analyzing the state-of-the-art", *Evolutionary Computation*, Vol. 8(2), pp. 125-147.

Vrugt, J.A., Gupta, H.V., Bastidas, L.A., Bouten, W., and Sorooshian, S. (2003). "Effective and efficient algorithm for multiobjective optimization of hydrologic models", *Water Resources Research*, Vol. 39(8), 1214, doi:10.1029/2002WR001746.

Yokoo, Y., Kazama, S., Sawamoto, M., and Nishimura, H. (2001). "Regionalization of lumped water balance model parameters based on multiple regression", *Journal of Hydrology*, Vol. 246(1-4), pp. 209-222.

(논문번호:07-24/접수:2007.03.22/심사완료:2007.07.31)