

유전알고리즘을 이용한 능형회귀모형의 검정 : 빈도별 홍수량의 지역분석을 대상으로

Calibration of the Ridge Regression Model with the Genetic Algorithm: Study on the Regional Flood Frequency Analysis

성 기 원*

Seong, Kee Won

Abstract

A regression model with basin physiographic characteristics as independent variables was calibrated for regional flood frequency analysis. In case that high correlations existing among the independent variables the ridge regression has been known to have capability of overcoming the problems of multicollinearity. To optimize the ridge regression model the cost function including regularization parameter must be minimized. In this research the genetic algorithm was applied on this optimization problem. The genetic algorithm is a stochastic search method that mimic the metaphor of natural biological heredity. Using this method the regression model could have optimized and stable weights of variables.

Keywords: genetic algorithm, ridge regression, regional analysis, flood frequency analysis

요 지

빈도별 홍수량의 지역분석을 위하여 유역의 지형특성을 독립변수로 이용하는 회귀모형을 검정하였다. 그런데 이들 독립변수들간의 상관관계가 존재할 경우 능형회귀모형이 이용되기도 하는데 이 방법은 다중공선성 문제를 극복하는데 적합한 방법으로 알려져 있다. 능형회귀모형을 최적화 하기 위해서는 조정변수가 포함되는 비용함수를 최소화 하여야 한다. 본 연구에서는 이 최적화를 위하여 유전알고리즘을 이용하였다. 유전알고리즘은 자연 생물의 유전 및 진화과정을 모방한 추계학적 탐색방법을 말한다. 이러한 유전알고리즘을 이용하여 지역분석 모형을 검정한 결과 안정된 매개변수의 가중치를 얻을 수 있었다.

핵심용어 : 유전자 알고리즘, 능형회귀, 지역분석, 홍수빈도해석

* 수원대학교 토목공학과 전임강사

1. 서 론

추계학적 최적해 탐색방법인 유전알고리즘(genetic algorithm, GA)을 이용하여 회귀모형의 최적화에 적용하여 보았다. Holland(1975)에 의하여 제안된 GA는 일반적으로 목적함수의 미분이 해석적으로 존재하지 않고 국부최소치(local minimum)가 다수 존재하는 multi modal 문제에서 전체최소치(global minimum)를 효율적으로 찾아주는 모형에 비의존적인 탐색법이다(Goldberg, 1989). GA는 생물의 유전 및 진화과정을 모방하여 매개변수를 최적화하는 기법인데 이 방법은 매개변수들이 염색체(chromosome)를 구성하고 있다고 보고 도태(selection), 교차(crossover)와 변형(mutation)을 반복하여 보다 적합한 최적해를 찾는 방법이다. 최적화 문제의 해결을 위하여 GA를 수공학분야에 적용한 연구사례로는 Wang(1991)의 연구가 있다. 그는 Xinanjiang 강우-유출 모형의 매개변수 검정을 위하여 GA를 이용하였으며 이 방법은 효율적인 수단임을 보여준 바 있다. 그리고 Savic과 Walters(1997)는 최적관망을 구성하는 문제에 적용한 바 있으며 McKinney와 Lin(1994)은 대수층에서의 양수문제에 관한 최적화에 그리고 Cieniawski와 Wayland(1995)는 대수층의 모니터링의 최적화에 GA를 적용하여 좋은 결과를 얻은 바 있다. 그리고 국내에서는 오경두(1994)가 최적 수질 관측망 설계를 위하여 GA를 이용한 바 있다.

본 연구에서는 GA를 이용하여 회귀분석 모형의 최적화에 적용하여 보았는데 최적화 하려는 대상 회귀모형은 빈도별 홍수량의 지역분석을 위한 모형이다. 그런데 빈도별 홍수량의 지역분석을 위해 영역의 지형 및 기상 인자를 독립변수로 이용하는 회귀모형을 수립할 경우 다중공선성(multicollinearity)이라 불리는 독립변수들의 독립성에 연관된 문제에 직면하게 되는 것이 보통이다. 이러한 다중공선성에 관한 문제를 해결하거나 완화시키는 기법은 여러 가지가 있는데 능형회귀(ridge regression)와 주성분회귀(principal component regression) 기법으로 대표되는 편의추정법(biased estimation method)도 그 가운데 한 가지로 분류된다. 이러한 편의추정법의 장점 가운데 한 가지는 다중공선성을 야기한 변수를 제거하지 않고 모형에 포함시킴으로써 변수들간의 상대적인 기여도를 평가할 수 있으며 평균제곱오차(mean squared error, MSE)를 구성하는 분산과 편의간의 교환(trade off)에 관한 정량

치를 확인할 수 있는데 있다. 능형회귀분석의 경우 모형을 최적화하기 위해 비용함수(cost function)를 최소화 하여야 하는데 모형이 다수의 독립변수들로 수립된 경우 전체최소치를 찾는 것은 대단히 곤란한 문제이다. 그래서 이러한 최적화 문제를 효과적으로 해결하기 위하여 GA를 이용하여 보았다.

본 연구에서는 IHP 대표유역 가운데 위천유역에 대한 홍수빈도분석 결과를 바탕으로 유역의 지형특성을 독립변수로 하는 능형회귀모형을 구성하였다. 이러한 능형회귀모형에 관한 연구는 Roy 등(1989)에 의하여 수행된 바 있으며 이 방법은 홍수빈도분석에 적절하다고 언급된 바 있다. 그리고 모형의 최적화를 위해 이용된 GA기법은 도태단계에서는 stochastic universal sampling방법을 교차단계에서는 multipoint crossover 방법을 적용하였다. 그런데 본 연구의 주 목적을 GA의 회귀분석모형에 대한 적용성 검토에 설정하였기 때문에 홍수빈도의 지역분석에 관한 여타의 방법은 비교하지 않았다.

2. 본 론

2.1 유전알고리즘

2.1.1 유전알고리즘의 개요

유전알고리즘(genetic algorithm, GA)은 생물체의 유전이나 진화과정을 은유적으로 모방하여 최적화 모형에 적용하는 추계학적 탐색법(stochastic search method)이다. 그런데 이 방법을 GA라 하는 이유는 최적화 하고자 하는 모형변수의 컴퓨터 비트(bit)내 문자(string)를 개체로 표현하는 것이 염색체(chromosome)와 유전자(gene) 등과 유사성이 있다는 사실에 기인하는 것이며 이 과정에서 목적함수의 형태나 최적화 과정의 수학적 고려는 사실상 없다. 이 방법은 유전이나 진화과정에서 역할을 하는 유전자 개체의 개체군(population of individual)의 형태와 성질을 모방하여 이용하는데 이것은 주어진 최적화 문제에 대한 잠재적인 해나 마찬가지로이다. 그 다음에는 적합도(fitness)를 평가하여 선별된 우수한 개체들의 결합(combining 혹은 breeding)을 통해 보다 적합한 개체를 생성시킨다. 적자생존의 법칙에 따라 보다 좋은 개체들이 생성되는데 이것은 좋은 해가 보다 많이 생겨난다는 의미뿐만 아니라 보다 더 나은 해를 얻을 수 있다는 것도 의미하는 것이다. 이것은 보다 우수한 자식을 생산하기 위해서 부모역할을 하는 개체의 가장 좋은

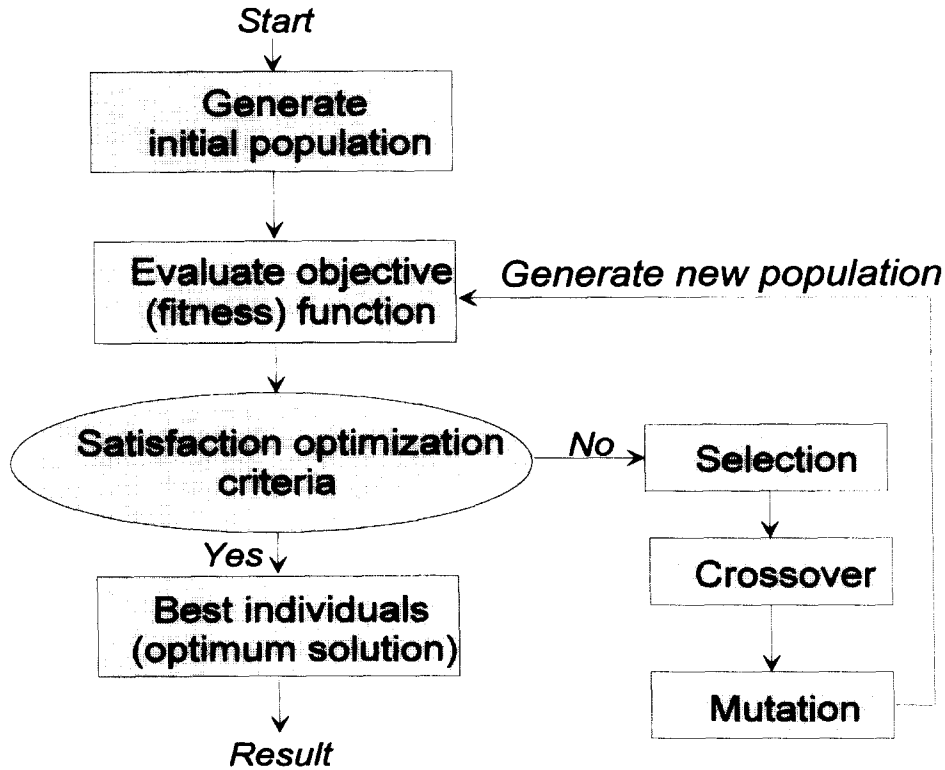


그림 1. 일반적인 GA 구조

특성(trait)들이 결합되기 때문에 가능한 것인데 이러한 결합과정을 생물학적으로는 교차(crossover)라 부른다. 교차를 통한 재결합에 부가하여 이러한 비트 내 문자들의 무작위 변형을 적용하기도 하는데 이 과정은 GA가 보다 좋은 해를 찾도록 하는 역할을 한다. 그러나 이러한 변형이 발생할 확률은 낮게 적용하는 것이 보통이다. 그림 1은 일반적인 GA에 대한 구조를 보여주고 있다. GA를 적용하는 과정에서 이용되는 기법들을 보다 자세히 설명하면 다음과 같다.

(1) 도태(selection)

도태는 결합 혹은 교배(mating)를 위해 어떤 개체들이 선택되며 각 선택된 개체가 얼마나 많은 자식(offspring)을 갖게되는지 결정하는 과정이다. 이 과정에서는 개체들에 대한 적합도를 측정후 이 값에 의하여 결합에 이용되는 부모가 결정되는데 이때 roulette-wheel selection, stochastic universal sampling, local selection, truncation selection 등과 같은 방법이 이용된다.

(2) 결합(combining), 교차(crossover)

결합과정에서는 부모가 갖고 있는 정보를 결합하여 새로운 개체를 생산해내는 과정이다. 그런데 개체를 즉 변수들이 자신의 문자를 표현하는 방식에 따라 real valued combining과 binary valued combining으로 나누는데 real valued combining에서 이용되는 방법으로는 discrete combining, intermediate combining, line combining, extended line combining 등이 있다. 그리고 binary valued combining 방식으로 분류되는 방법으로는 single point crossover, multi point crossover, uniform crossover, shuffle crossover 등이 있다.

(3) 변형(mutation)

보통 결합과정 후에 염색체의 의도적인 변형을 한다. 이 경우 자식 변수들은 낮은 확률로 작은 양의 섭동(perturbation)을 이용하여 변형되게 된다. 일반적으로 변형이 일어날 확률 혹은 변형율은 개체내 변수의 수에 대한 역수 이하의 값을 선택한다(Davis, 1991).

이는 변형율이 높을 경우 변수 선택의 무작위성이 커지기 때문이다.

2.1.2 적용된 유전알고리즘

GA는 생물의 유전과 진화과정을 모방한 알고리즘이지만 알고리즘을 구성하는 도태, 교차, 변형의 단계에서 여러 가지 명칭으로 불리우는 다양한 조작 방법이 있다. 그런데 이러한 조작방법에 따라 최적해에 도달하는 효율성에 있어서 상대적 우열성은 있는 것으로 알려져 있다. 그러나 본 연구에서 최적화 하려는 능형회귀모형에 대하여 적용한 각 단계별 방법은 GA 연구에서 널리 이용되며 안정성이 있다고 알려져 있지만 계산논리는 비교적 단순한 방법을 택하였다.

도태단계에서는 stochastic universal sampling 방법(Baker, 1987)을 택하였다. 이 방법은 GA용어로 이용되는 편이(bias, 개체에 대한 정규화된 적합도와 그것의 재생산확률의 차의 절대값)가 0인 특징이 있고 spread(개체가 만들 수 있는 자식의 수의 가능값의 범위)가 다른 방법에 비하여 작기 때문에 가장 빈번히 이용되는 도태방법인 roulette wheel selection에 비하여 최적화에 도달하는 시간이 가속되는 경향을 보여준다. 이 방법을 그림 2에 나타내었다.

그림 2는 stochastic universal sampling을 설명하기 위하여 임의로 설정한 10개의 개체에 대하여 적합도를 가정하고 이에 따라 정렬한 후 선택확률이 계산된 가상의 예를 들고 이에 대한 stochastic universal

sampling 방법의 적용을 나타낸 것이다. 그림 2에서 볼 수 있듯이 개체들을 한 선의 연속된 구획에 배치 시키는데 이 때 각 구획의 크기는 적합도의 정도에 비례하도록 결정한다. 그리고 이 선에 선택될 개체의 수만큼 위치자(pointer)를 등간격으로 배치한다. 위치자의 수가 결정되면 위치자들 간의 간격은 위치자의 갯수의 역수(1/위치자의 수)로 지정하는데 단 첫 번째 위치자의 위치는 0과 1/위치자의 수 사이에 난수를 발생시켜 무작위로 결정한다. 그림에서는 4개의 개체가 선택되도록 예를 들었다. 따라서 위치자들간의 간격은 0.25(=1/4)이다. 그리고 첫 번째 위치자의 위치 결정을 위해 0과 0.25 사이에 난수를 발생시켜 0.1를 얻었을 경우 그림에서와 같이 선택되는 개체는 1, 2, 4 그리고 6이 된다.

교차단계에서는 multi-point crossover 방법을 이용하였다. 이 방법은 각 개체가 포함하고 있는 변수를 Nvar이라 할 때 Nvar-1개의 교차점 위치를 중복되지 않게 결정하는 방법을 말한다. 그리고 교차점의 위치가 결정되면 이 사이의 부모 변수들이 교환되어 두 개의 새로운 자식변수들이 생성된다. 그림 3에 이와 같은 과정을 도시하였다. 그런데 초기 최적해 탐색과정에서 높은 적합도를 가진 개체와 그 주변으로 계산을 수렴시키는 다른 방법에 비하여 multi-point crossover 방법은 분열성(disruptive nature)이 있기 때문에 탐색공간이 확장되어 탐색결과가 보다 robust한 장점이 있다고 알려져 있다(Davis, 1991).

Number of individual	1	2	3	4	5	6	7	8	9	10
fitness value	2.2	1.8	1.6	1.4	1.0	0.8	0.6	0.4	0.2	0.0
selection probability	0.22	0.18	0.16	0.14	0.10	0.08	0.06	0.04	0.02	0.00

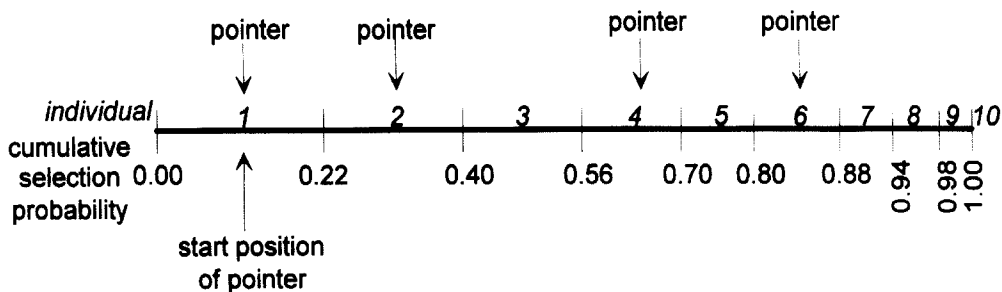


그림 2. stochastic universal sampling 방법의 예시

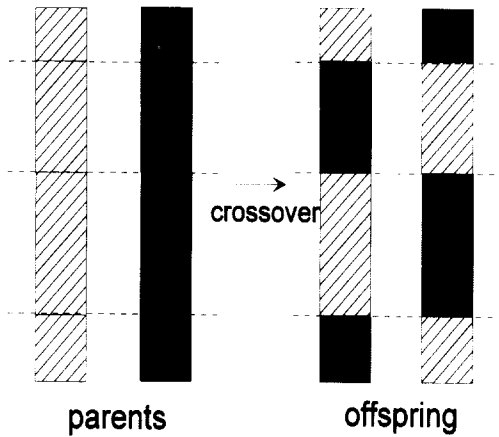


그림 3. multi-point crossover 원리

변형단계는 낮은 확률로 발생시킨 적은 양의 난수값을 지식변수에 부가시키는 단계를 말한다. 변형되는 변수의 확률, 즉 변형률은 일반적으로 개체의 수에 무관하게 결정되는 반면 각 개체가 갖고 있는 변수 개수의 역수에 비례하도록 하는 것이 일반적이다. 그러나 의도적으로 높은 변형률을 가할 경우 최적에 빨리 도달하는 경우도 있다고 알려져 있다. 이것은 탐색공간을 확장시키는 역할을 하기 때문으로 생각되는데 이러한 방법은 그다지 적절치 못하다고 판단된다. 이는 변형단계에서 적은 양의 섭동의 변화를 통해 보다 정밀한 최적해를 찾으려 하는 것이 적절하고 탐색공간의 확장은 선택이나 교차단계에서 결정하는 것이 논리적으로 타당하기 때문이다.

2.2 회귀모형의 수립

2.2.1 홍수량빈도의 지역분석

지역분석은 자료의 보유연한에 차이를 보이는 유역간의 홍수분위수(flood quantile)를 추정하고 이를 바탕으로 수문학적 동질성을 갖는 유역에 대한 홍수량을 추정하는 경험식을 수립하는 것이다. 이러한 경험식은 대상 유역들의 자료에 대한 통계적 특성을 유역의 지형특성을 독립변수로 하는 회귀모형으로 표현하는 것이 일반적이다. 그러나 이 방법에서는 서로 연관된 독립변수가 존재할 경우 발생하는 다중공선성을 평가하고 이를 처리해야 한다. Roy 등(1989)은 홍수량빈도의 지역분석과정에서 능형회귀모형은 다중공선성을 극복하는데 유용하며 특히 자료의 변화나 일부 관측자료의 배제에도 불구하고 회귀계수의 안정성이 우수하다고 밝힌 바

있다. 이에 본 연구에서도 홍수량빈도의 지역분석을 위하여 능형회귀모형을 수립하였다.

2.2.2 배경이론

능형회귀분석은 ill-posed 문제의 해를 추정하기 위하여 조정변수(regularization parameter)라는 여분의 정보를 이용하는 기법을 말한다 (Orr, 1995). 능형회귀분석에서 MSE는 편의(bias)나 분산(variance)의 오차의 합으로 이루어지는데 의도적인 편의를 도입하여 분산을 감소시키고 결과적으로 MSE를 감소하게 하는 기법을 이용한다. 일반적인 다중회귀분석모형의 표준형은 식(1)과 같다.

$$f(x) = \sum_{j=1}^m w_j h_j(x) \quad (1)$$

여기서 $f(x)$ 는 관측치 $y(x)$ 의 선형모형이며 $h(x)$ 는 독립변수 그리고 w 는 미지의 가중치이다. 그리고 관측치 벡터를 $\{(x_i, \hat{y}_i)\}_{i=1}^p$ 라 하면 오차제곱합(sum squared error)은 식(2)와 같다.

$$SSE = \sum_{i=1}^p (\hat{y}_i - f(x_i))^2 \quad (2)$$

그런데 능형회귀분석시 penalty를 SSE에 부가할 경우 다음의 비용함수(cost function)를 얻는다.

$$C = \sum_{i=1}^p (\hat{y}_i - f(x_i))^2 + \lambda \sum_{j=1}^m w_j^2 \quad (3)$$

여기서 λ 는 조정변수이다. 그런데 여기서 능형회귀모형에 대한 최적가중치 벡터는 행렬식으로 나타낼 경우 식(4)와 같다.

$$w = (H^T H + \lambda I_m)^{-1} H^T y \quad (4)$$

여기서 I_m 은 $(m \times m)$ 차원의 단위행렬이고 H 는 설계행렬로 식(5)와 같다.

$$H = \begin{bmatrix} h_1(x_1) & h_2(x_1) & h_3(x_1) & \cdots & h_m(x_1) \\ h_1(x_2) & h_2(x_2) & h_3(x_2) & \cdots & h_m(x_2) \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ h_1(x_p) & h_2(x_p) & h_3(x_p) & \cdots & h_m(x_p) \end{bmatrix} \quad (5)$$

또한 MSE는 식(6)과 같다.

$$MSE = (y(x) - \langle f(x) \rangle)^2 + \langle (f(x) - \langle f(x) \rangle)^2 \rangle \quad (6)$$

여기서 $\langle \rangle$ 는 기대값을 나타낸다. 그런데 식(3)을 최소화 하기 위해서는 모두 $m+1$ 개의 변수에 대한 최적화가 필요한데 본 연구에서는 이를 GA를 이용하여 최

적화 하였다. 그리고 이 최적화의 검증을 위하여 식(1), 식(4) 그리고 식(6)을 이용하여 MSFE를 평가하였다.

3. 결과 및 분석

3.1 대상 유역의 개황 및 분석자료

본 연구를 위한 대상유역은 IHP 대표유역 가운데 한 유역인 위천유역으로 이 지역에는 모두 6곳의 수위관측소와 11곳의 우량관측소가 있다. 따라서 수위관측소를 기준으로 유역을 분할하면 모두 6개의 소유역으로 나누어진다. 이 유역의 빈도별 홍수량의 지역분석을 위하여 유역의 지형특성자료를 수집하여 표 1에 수록하였다. 그리고 이들 지역특성인자들간의 식(7)로 나타내는 상관계수를 계산하여 종속성을 살펴보았다.

$$\rho_{X,Y} = \frac{COV(X,Y)}{Var(X) \cdot Var(Y)} \quad (7)$$

여기서 $Var(X)$ 와 $Var(Y)$ 는 각각 변수 X , Y 의 분산이고 $Cov(X, Y)$ 는 변수 X , Y 의 공분산이다. 유역특성인자 상호간의 상관분석 결과를 표 2에 나타내었다. 그런데 표 2를 살펴보면 특히 A , L , L_c , Z 그리고 D 간에는 상관계수가 0.7을 초과하는 등 일부 특성인자들 간에는 높은 상관관계가 있음을 알 수 있다. 이와 같은 높은 다중공선성이 예상되는 상황에서 모든 지형특성인

자를 독립변수로 이용하는 회귀모형을 수립하는 것은 실용적인 적용성 면에서 바람직하지 않으며 정량적 정확성에 대한 향상도 기대하기 어렵다. 그러나 지역특성인자들의 상대적 기여도를 비교, 평가하는 것은 편이추정법 연구의 중요한 일면이다. 그러나 본 연구에서는 R 과 D 는 다른 특성인자들의 결합으로 얻어진 값이기 때문에 분석에서 제외하였다. 또한 L 과 Z 는 A 와 거의 0.9를 웃도는 상관관계를 보여주기 때문에 제외하는 것이 바람직 하다. 그래서 상관관계가 거의 1과 일치하는 Z 는 분석에서 제외하였다. 그러나 상관관계가 이 보다는 낮은 L 은 능형회귀모형의 독립변수로 이용이 가능하다고 가정하고 분석에 포함하였다. 위천유역의 홍수량의 빈도분석자료는 가장 최근에 이 지역에 대하여 분석한 김남원(1994)의 연구보고서를 참고하였다. 그는 자료년한의 2배에 해당하는 부분계열모형을 채택하고 Wakby 분포를 이용하여 IHP 대표유역에 관한 지역빈도분석을 수행한 바 있다. 이 가운데 자료의 신뢰범위 내에 있다고 판단되는 재현기간 20년 이하의 홍수량을 분석에 이용하였는데 이를 표 3에 수록하였다.

3.2 회귀분석모형의 검증

빈도별 홍수량과 유역특성인자와의 관계를 식(8)과 같이 설정하였다.

표 1. 위천유역의 지형특성인자

소유역명	유역면적 A (km ²)	주하천연장 L (km)	유역표고차 F (m)	중심장 L_c (km)	유출곡선 번호, CN	총하천연장 Z (km)	형상계수 R	하천밀도 D (km ⁻¹)
무 성	472.53	42.83	461.71	20.60	65.80	743.32	0.258	1.573
병 천	302.13	40.65	446.74	17.10	66.90	488.87	0.183	1.618
미 성	171.25	30.90	431.98	18.90	69.20	274.86	0.179	1.605
고 노	109.73	18.90	393.88	9.70	73.10	153.80	0.307	1.402
동 곡	33.63	8.00	324.56	4.00	76.20	39.92	0.525	1.187
효 령	151.03	21.80	677.76	8.90	64.20	225.70	0.318	1.494

표 2. 유역인자들간의 상관계수

	A	L	F	L_c	CN	Z	R	D
A	1.0000	0.9207	0.1913	0.8307	-0.6728	0.9992	-0.5868	0.7033
L		1.0000	0.2142	0.9414	-0.7223	0.9323	-0.8356	0.9002
F			1.0000	0.0870	-0.8269	0.1855	-0.2937	0.4355
L_c				1.0000	-0.5966	0.8439	-0.8642	0.8924
CN					1.0000	-0.6757	0.6571	-0.8070
Z						1.0000	-0.6040	0.7188
R							1.0000	-0.9684
D								1.0000

표 3. 위천유역의 재현기간별 홍수량 (단위 : cms)

소유역 명	재현기간 (년)			
	2	5	10	20
무 성	979.89	1445.09	1813.95	2198.08
병 천	625.19	922.00	1157.34	1402.43
미 성	348.59	514.09	645.31	781.96
고 노	265.16	391.05	490.86	594.81
동 곡	71.96	106.12	133.20	161.41
효 령	338.66	499.44	626.93	759.69

$$Q_T = A^{w_1} L^{w_2} F^{w_3} L_c^{w_4} CN^{w_5} 10^{w_6} \quad (8)$$

여기서 Q_T 는 재현기간 T 년의 홍수량이다. 그리고 식 (8)의 10^{w_6} 은 상수항을 나타낸 것인데 이는 추정하려는 변수들의 염색체를 동일한 크기로 배치하고자 의도적으로 이와 같이 나타낸 것이다. 그리고 식(8)을 식 (1)과 같은 선형다중회귀모형으로 전환하기 위하여 대수변환을 수행하였고 이에 의한 모형을 검정하였다.

본 연구를 위하여 구성한 GA를 위해 모두 200개의 개체의 규모를 갖는 개체군이 설정되었으며 일양난수발생기를 이용 변수의 초기값이 부여되었다. 그리고 개체 즉 염색체는 $w_1 \sim w_6$ 과 λ 등 모두 7개의 변수 혹은 특성(trait)으로 구성되었다. 각각의 변수에 대한 상·하한 경계값은 1과 0으로 지정하였고 변형률은 Davis (1991)가 제시한 방법인 '변수개수의 역수' 즉 1/7 이내로 하였다. 소수점 이하의 유효자리가 5자리 까지 가능하도록 개체내 유전자를 배열하였고 초기의 GA연구에서 이용되던 2진수가 아닌 10진수 배열을 이용하여 기억용량을 줄이고 빠른 계산이 이루어지도록 도모하였다. 이 과정을 통해 얻은 빈도별 홍수량에 대한 능형회귀모형을 최적화한 결과 및 λ 를 표 4에 수록하였다.

표 4를 살펴보면 매개변수의 가중치와 λ 모두 안정적인 관계를 보여주고 있다. GA를 이용한 최적화가 적절하게 이루어 졌는지 살펴보기 위해 최적화 결과인 $w_1 \sim w_6$ 를 고정시키고 λ 를 $10^{-4} \sim 10$ 까지 변화시키면서 MSE의 변화를 살펴보았다. 빈도별 변화경향을 그림 4~7에 수록하였다. 그림에서 '•' 표시점은 GA를 이용하여 찾아낸 최적 λ 로 이와 같은 MSE 분석을 통해 GA는 회귀분석 모형의 검정에 적절히 이용될 수 있음을 간접적으로 알 수 있었다. 그리고, 검정된 빈도별 홍수량이 소유역별로 어느 정도의 일치도를 갖는지 확인하기 위하여 그림 8에 나타내었다.

그림을 살펴보면 고노 유역에서 약간의 편차를 보일 뿐 유역 전반적으로 적절한 검정결과를 보여주고 있다. 따라서 이상과 같은 분석결과를 바탕으로 판단해건데 적절한 최적화가 이루어졌다고 판단된다. 그러나 계산 과정에서 들어난 GA의 문제점도 간과할 수 없었다. 유전 및 진화과정에서 가장 좋은 개체 즉 최적해를 도출하는데 있어서 최적해와 20% 정도의 차이를 나타내는 영역으로 대체로 20회 이내의 유전과정이면 접근하였으나 이 때부터 최적화를 위해 설정한 한계값에 이르기까지는 대단히 느린 접근속도를 보여주었다.

표 4. GA를 이용한 능형회귀모형의 최적화 결과

최적화 결과	재현기간 (년)			
	2	5	10	20
w_1	0.79749	0.80143	0.80262	0.80672
w_2	0.20667	0.20372	0.20286	0.20016
w_3	0.13526	0.15421	0.16638	0.17471
w_4	0.01422	0.01341	0.01380	0.01192
w_5	0.05033	0.09992	0.12871	0.15414
w_6	0.02783	0.04982	0.06281	0.07377
λ	0.01529	0.01482	0.01472	0.01423

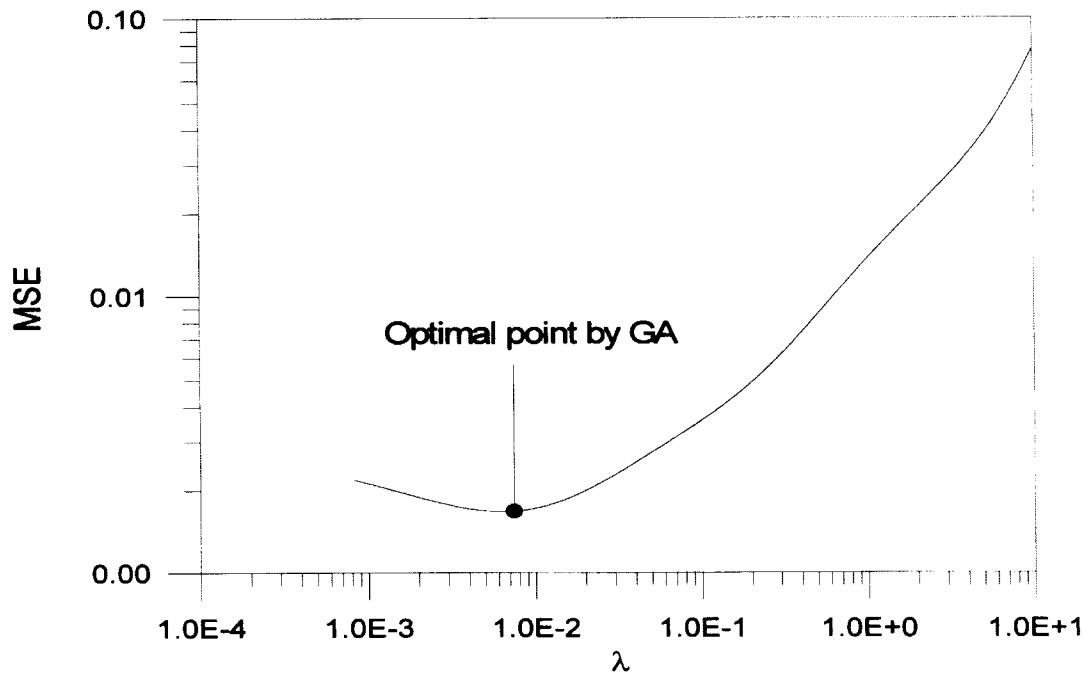


그림 4. 재현기간 2년의 회귀모형에 대한 MSE 분석

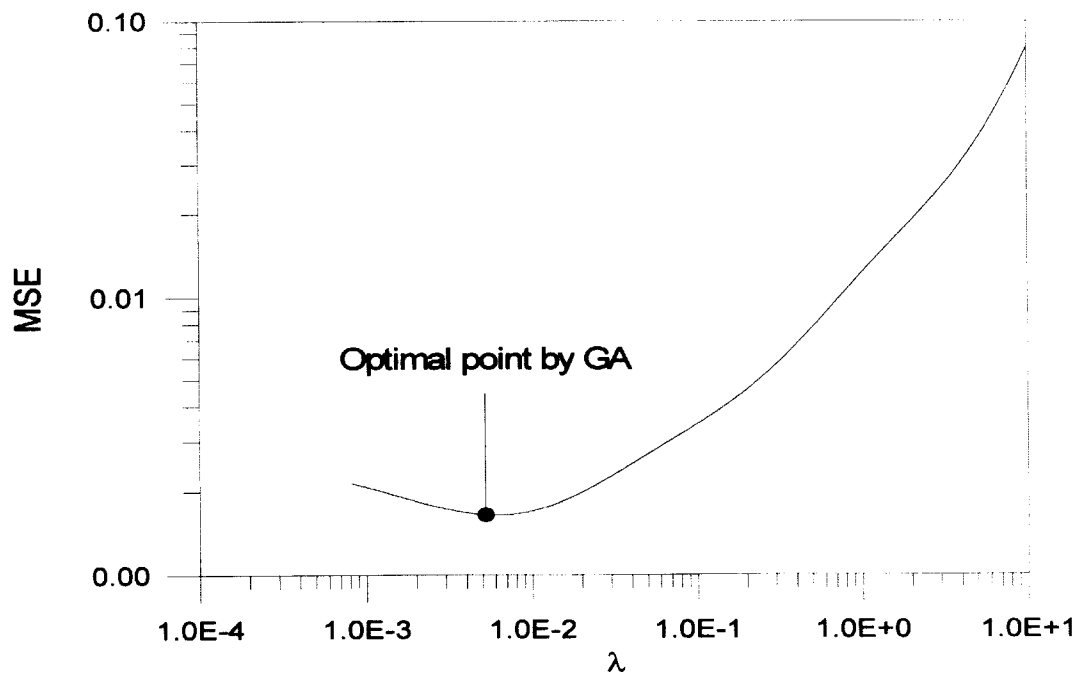


그림 5. 재현기간 5년의 회귀모형에 대한 MSE 분석

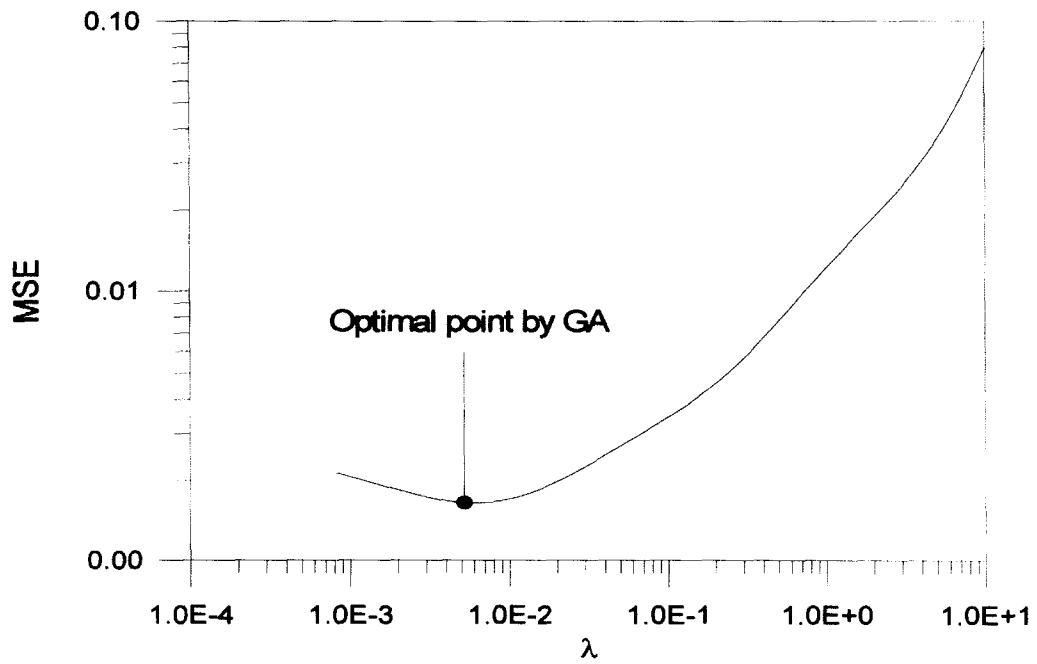


그림 6. 재현기간 10년의 회귀모형에 대한 MSE 분석

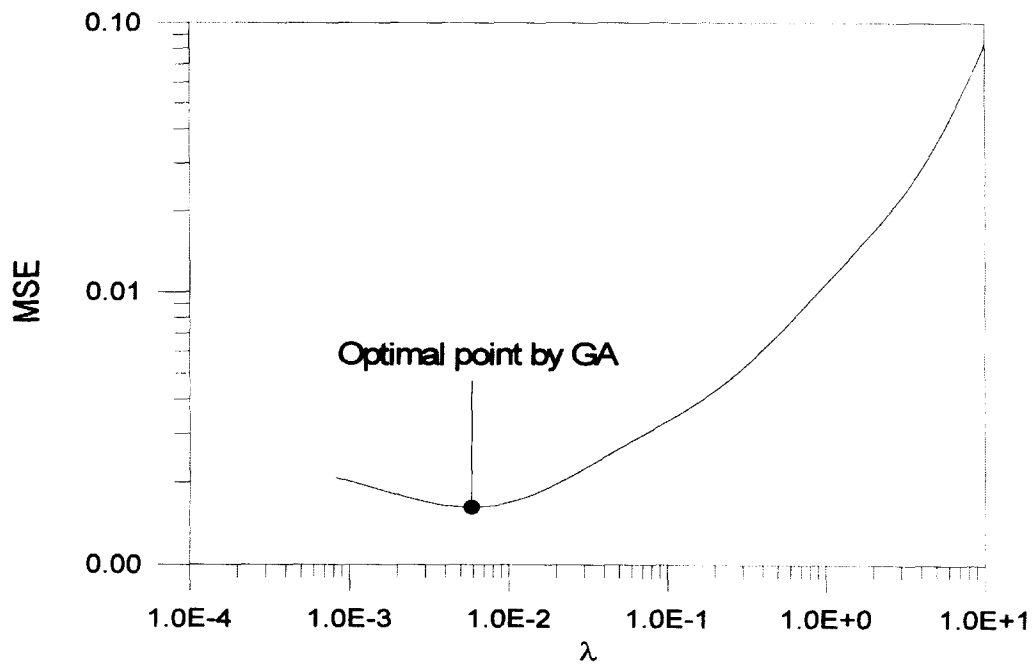


그림 7. 재현기간 20년의 회귀모형에 대한 MSE 분석

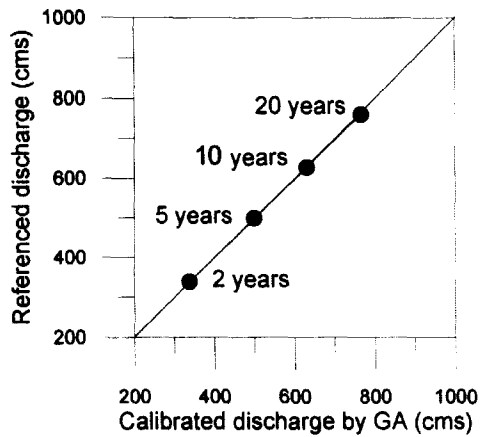
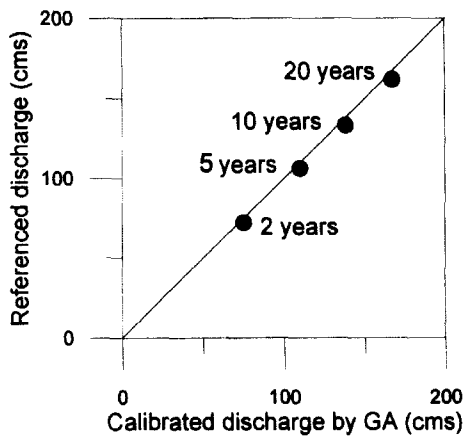
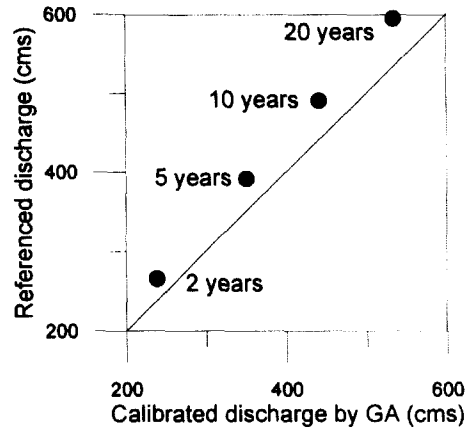
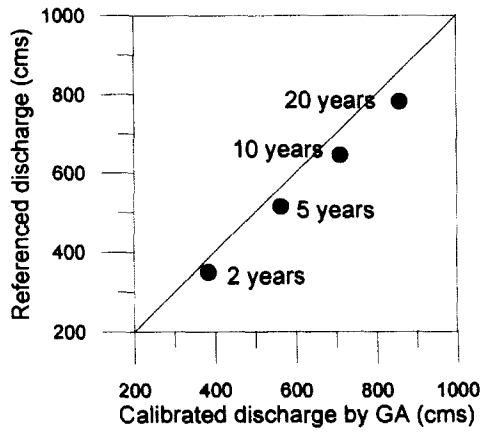
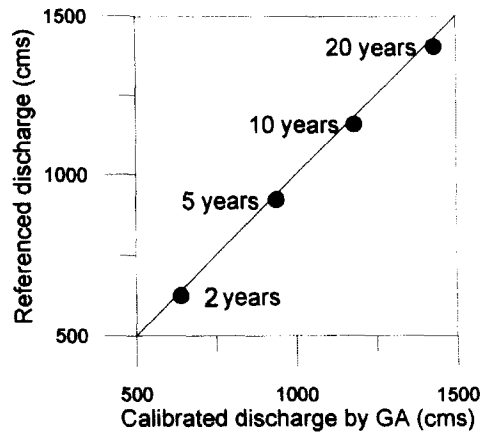
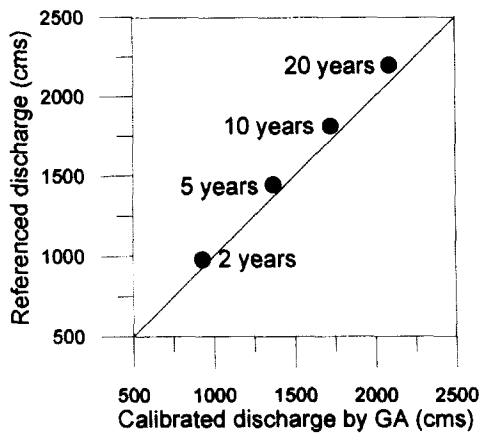


그림 8. 빈도별 홍수량의 검정결과 비교

이 경향이 GA의 일반적 성향임을 이미 Davis(1991)가 지적한 바가 있다. 따라서 이를 개선하기 위한 도태, 교차 그리고 변형과정에 대한 연구가 요구된다.

4. 결 론

빈도별 홍수량의 지역분석 모형으로 적절하다고 제시된 능형회귀 모형을 GA를 이용하여 최적화 하였다. 이를 위한 과정과 결론을 다음과 같이 요약할 수 있다.

(1) 위천유역에 대한 빈도별 홍수량의 지역분석을 위해 편의추정법 가운데 한 가지인 능형회귀모형을 검정하였다.

(2) 회귀모형의 최적화를 위하여 GA를 이용하였으며 이 과정에서 도태단계는 stochastic universal sampling 방법을 교차단계에서는 multi-point crossover 방법을 그리고 변형단계에서는 염색체내 변수의 개수에 의존하는 변형률을 적용하는 방법을 선택하였다.

(3) GA를 이용하여 지역분석 모형을 검정한 결과 추정된 독립변수의 가중치와 조정변수 모두 안정된 값을 보여주었다.

(4) MSE 분석결과 GA는 적절히 최적해를 찾아주고 있었다.

(5) GA의 효율은 최적해가 포함되어 있는 영역을 찾을 때가 정확한 최적해에 이르기 까지보다 높았다.

참 고 문 헌

김남원 (1994). 미계측 유역의 홍수유출 특성에 관한 비교연구." 연구보고서, 건기연94-WR-112, 한국건설기술연구원.
 오경두 (1994). "하천 수질 관측망 최적설계 기법 개발." 국제수문개발계획(IHP) 연구보고서, 건설부, pp. 9.1-9.8.
 Baker, J.E. (1987). "Reducing bias and inefficiency in the selection algorithm." *Proc. 2nd International Conf. of on Genetic Algorithms and their Application*, ICGA, Hillsdale, New Jersey, pp. 14-21.

Cieniawski, S.E. (1995). "Using genetic algorithms to solve a multiobjective groundwater monitor." *Water Resources Research*, Vol. 31, No. 2, pp. 399-409.
 Davis, L. (1991). *The handbook of genetic algorithms*. Van Nostrand Reingold, New York.
 Goldberg, D.E. (1989). *Genetic algorithms in search, optimization, and machine learning*. Addison-Wesley, Mass.
 Holland, J.H. (1975). *Adaption in natural and artificial systems*. The University of Michigan Press, Ann Arbor.
 McKinney, D.C., and Lin, M (1994). "Genetic algorithm solution of groundwater management models." *Water Resources Research*, Vol. 30, No. 6, pp. 1897-1906.
 Orr, M.J.L. (1995). "Regularisation in selection of radial basis function centres." *Neural Computations*, Vol. 3, No. 2, pp. 606-623.
 Savic, D.A., and Walters, G.A. (1997). "Genetic algorithms for least-cost design of water distribution networks." *J. of Water Resources Planning and Management*, ASCE, Vol. 123, No. 2, pp. 67-77.
 Roy, R., Bobee, B., Ashkar, F., and Roberge, F. (1989). "Regional flood frequency analysis using ridge regression." *New directions for surface water modeling, Proc. of Baltimore Symposium of IAHS*, IAHS publ. no. 181, pp. 293-300.
 Wang, Q.J. (1991). "The genetic algorithm and its application to calibrating conceptual rainfall-runoff models." *Water Resources Research*, Vol. 27, No. 9, pp. 2467-2471.
 <최종본 접수일 : 1997년 12월 19일>