

# 유전자 알고리즘을 이용한 강우강도식의 매개변수 추정에 관한 연구

## Parameter Estimation of Intensity-Duration-Frequency Curve using Genetic Algorithm

신주영\*, 김수영\*\*, 김태순\*\*\*, 허준행\*\*\*\*

Juyoung Shin, Sooyoung Kim, Taesoon Kim, Jun-Haeng Heo

### 요 지

본 연구에서는 강우강도식의 매개변수를 보다 효율적으로 산정하기 위해서 유전자알고리즘을 적용한 매개변수 산정법을 제시하였으며, 지속기간의 장, 단기간에 따른 매개변수의 변화를 고려하기 위하여 다목적 유전자알고리즘을 적용하여 매개변수를 추정된 후 기존의 강우강도식에 의한 결과와 비교해 보았다. 매개변수는 지점빈도해석을 통해 산정된 확률강우량을 사용하여 추정하였고, 유전자 알고리즘의 목적함수로는 Nash & Sutcliffe Index, Root Mean Square Error(RMSE), Relative Root Mean Square Error(RRMSE), 결정계수, 평균들을 사용하여 가장 효율적인 형태의 목적함수를 구성하였다. 그 결과 기존의 매개변수 추정 방법들에 비해 유전자알고리즘을 이용한 경우에 더 정확한 강우량값을 산정할 수 있었고, 특히 다목적 유전자 알고리즘을 사용할 경우 장기간과 단기간에 걸쳐서 동시에 정확도를 향상시킬 수 있는 매개변수를 추정할 수 있는 것으로 나타났다.

**핵심용어:** 유전자알고리즘, 강우강도식, 매개변수추정

### 1. 서 론

강우강도식이란 임의의 재현기간과 지속기간에 해당하는 강우강도를 산정하는 식으로 우리나라에서는 Talbot, Sherman, Japanese형과 같이 비교적 간단한 방법으로 매개변수를 산정할 수 있는 강우강도식과 이원환(1993), 허준행(1999)이 개발한 것과 같이 보다 향상된 정확도를 얻기 위해서 다수의 매개변수를 사용하는 강우강도식이 개발되어 있다. 이 중에서 비교적 간단한 형태를 가지는 Talbot, Sherman, Japanese형의 강우강도식은 손쉽게 매개변수를 추정할 수 있기 때문에 실무에 널리 사용되고 있으나, 이원환(1993), 허준행(1999)이 개발한 강우강도식은 매개변수 추정상의 어려움 때문에 실무에 쓰이기에는 어려움이 있다고 할 수 있다.

따라서, 본 연구에서는 최적화기법 중 하나인 유전자 알고리즘을 사용하여 기존의 강우강도식의 매개변수를 추정하고, 추정된 강우강도식의 매개변수의 적용성을 판단하여 강우강도식의 활용도를 높이고자 한다. 또한, 일반적으로 장기간 및 단기간으로 구분하여 산정되어지는 매개변수를 다목적 유전자알고리즘을 활용하여 한번에 구하는 방법을 제시하였으며, 산정된 매개변수를 활용한 강우강도식으로부터 구한 확률강우량과 지점 빈도해석으로 구한 확률강우량값을 비교하여 활용성을 평가하였다.

\* 정회원 · 연세대학교 대학원 토목공학과 석사과정 · E-mail: ausran@yonsei.ac.kr

\*\* 정회원 · 연세대학교 대학원 토목공학과 박사과정 · E-mail: sykim79@yonsei.ac.kr

\*\*\* 정회원 · 세종대학교 토목공학과 BK21교수 · 공학박사 · E-mail: chaucer@yonsei.ac.kr

\*\*\*\* 정회원 · 연세대학교 사회환경시스템공학부 토목환경공학과 교수 · E-mail: jheo@yonsei.ac.kr

## 2. 유전자 알고리즘의 개요

### 2.1 유전자 알고리즘(Genetic Algorithm)

유전자 알고리즘은 자연선택의 법칙인 자연진화 이론으로부터 나온 탐색기법이다(Goldberg, 1989). 유전자 알고리즘은 초기에 무작위로 초기 모집단(Population)을 형성하고 이들을 부모세대(parent)로 하여 선택(Selection), 교배(Crossover), 돌연변이(Mutation) 등의 연산과정을 거쳐 부모세대보다 진화한 새로운 자식세대(offspring)를 생성하게 된다. 유전자알고리즘의 방법으로는 여러 가지 방법이 있으나, 본 연구에서는 NSGA-II를 사용하였다. NSGA-II는 가장 많이 활용되고 있는 다목적 유전자알고리즘 기법으로써, 엘리티즘(elitism)을 적용하였으며, 순위를 결정하기 위한 복잡도(order of complexity)가 감소하였고, 해의 다양성(diversity)을 확보하기 위한 sharing 기법에서 사용자가 임의로 설정하는 매개변수를 없앤 군집거리방법(crowding distance)을 적용한 기법이다(Deb, 2002).

### 2.2 목적함수 선정

유전자 알고리즘을 사용하여, 강우강도식의 매개변수를 최적화시키기 위해서는 적절한 목적함수의 선정이 필요하다. 따라서 본 연구에서는 Nash & Sutcliffe Index, Root Mean Square Error(RMSE), Relative Root Mean Square Error(RRMSE), 결정계수, 평균을 목적함수으로써의 적용성을 확인해보았다(Nash, 1988). 결정계수의 경우 선형회귀분석에서만 사용이 가능하여 적용성이 떨어지는 단점이 있었고, 평균의 경우 매개변수의 변화에 따른 모형의 적합도(performance criteria)를 나타내는 능력이 현저히 부족한 것으로 나타났다. 강우-유출 모형의 적합도를 나타내는데 널리 사용되는 Nash & Sutcliffe Index의 경우 적용은 가능하나 결과값이 RMSE를 적용하였을 때와 같은 결과값이 나오므로, 본 연구에서는 RMSE와 RRMSE를 유전자알고리즘의 목적함수로 사용하였다. RMSE, RRMSE는 식 (4)~(5)와 같다.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{Q}_i - Q_i)^2} \quad (4)$$

$$RRMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n \left( \frac{\hat{Q}_i - Q_i}{Q_i} \right)^2} \quad (5)$$

$\hat{Q}_i$ 는 i번째 추정값,  $Q_i$ 는 i번째 관측값, n은 전체 자료수이다.

## 3. 강우강도식의 매개변수 추정

### 3.1 지점빈도해석

본 연구에서는 1916년부터 1999년까지 대구지점의 강수량 자료를 이용하여 빈도해석을 실시하였고 유의 수준 5%에 대한 확률가중모멘트(Probability Weighted Moments)방법을 매개변수 추정방법으로 사용하였다. 대구지점의 적정확률분포형으로는 김경덕 등(1995)의 연구결과를 인용하여 GEV 분포형으로 선정하였다. 지속기간 1, 2, 6, 9, 12, 15, 18, 24, 48시간의 강우자료를 이용하여 재현기간 2, 3, 5, 10, 20, 30, 50, 70, 80, 100, 200, 300, 500년에 대하여 확률강우량을 산정하였다.

### 3.2 적용 CASE 선정

본 연구에서는 총 3개의 CASE로 매개변수를 추정하였고 각각의 경우에 대한 특징은 표 1과 같다. CASE 1은 허준행식에서 b와 n이 비교적 확률강우량의 크기의 영향을 많이 주는 매개변수으로써 b와 n만을 재현기간

별로 추정하기 위해 적용하였다. CASE 2의 경우 이원환 식에서  $c$ 와  $n$ 값이 다른 매개변수들에 비해 확률강우량 값에 영향을 많이 주는 매개변수로서  $c$ 와  $n$ 만을 재현기간별로 추정하기 위해 적용하였다. CASE 3은 통합형의 매개변수를 추정하기 위하여 적용하였다.

표 1. 적용 CASE 설명

CASE	특 정
CASE 1(허준행식)	모든 재현기간에 대하여 $a, c, d$ 을 추정, 재현기간별로 $b, n$ 를 추정
CASE 2(이원환식)	모든 재현기간에 대하여 $a, b$ 를 계산, 재현기간별로 $c, n$ 를 추정
CASE 3(통합형식)	재현기간별로 $a, b, n$ 을 추정

#### 4. 결과 분석

그림 1은 대구 지역의 지점빈도해석으로 산정한 확률강우량과 목적함수를 RMSE와 RRMSE로 사용하였을 때 유전자 알고리즘을 통해 산정된 확률강우량을, 각각의 지속시간별로 모든 재현기간에 대해서 RMSE를 평균한 값을 나타낸 것이고, 그림 2는 RRMSE를 평균한 값을 나타낸 것이다.

그림 1에서 파란색 막대는 RMSE를 목적함수로 사용하였을 때의 RMSE 평균값을 나타내는 것이고, 빨간색 막대는 RRMSE를 목적함수로 사용하였을 때 RMSE 평균값을 나타낸 것이다. 그림 1을 좀더 자세히 살펴보면, 지속기간 24, 48시간의 RMSE 평균값에서 빨간색 막대가 파란색 막대보다 큰 값을 나타내고 있다. 즉 지속기간 24, 48시간에서는 RMSE를 목적함수로 사용하여 추정된 매개변수가 RRMSE를 목적함수로 사용하여 추정된 매개변수보다 높은 정확도를 나타내는 것을 알 수 있다. 같은 방법으로 그림 2의 지속기간 1, 2시간에서의 RRMSE 평균값을 보면 파란색 막대가 빨간색 막대보다 크게 나타내고 있는 것을 확인 할 수 있다. 이것은 지속기간 1, 2시간에서는 RRMSE를 목적함수로 사용하여 추정된 매개변수가 RMSE를 사용하여 추정된 매개변수보다 높은 정확도를 나타낸다는 것을 의미한다. 따라서, 지속기간 1, 2시간과 같은 짧은 지속기간에서는 RRMSE를 이용한 목적함수가 더 높은 정확도를 내는 매개변수를 산정할 수 있으며, 지속기간 24, 48시간과 같이 비교적 긴 지속기간을 가지는 경우에는 RMSE를 이용한 목적함수를 사용하는 것이 더 높은 정확도를 얻을 수 있는 방법이라고 말할 수 있다.

본 연구에서는 이와 같은 목적함수의 특성을 활용하여 장기간과 단기간의 지속기간을 모두 고려한 형태의 강우강도식을 제시하고자 하였다. RMSE와 RRMSE의 두 가지 목적함수를 사용하는 다목적 유전자알고리즘을 수행한 후에 계산된 파레토프론트 상에서, 각각 RMSE를 최소로 하는 매개변수와 RRMSE를 최소로 하는 매개변수를 선택한 후 이를 이용하여 확률강우량(A)을 구하고, 다시 지점빈도해석을 이용하여 구한 확률강우량(B)과 비교하여 A와 B사이의 절대편의를 지속기간별로 계산한 후, RMSE로 구한 절대편의가 RRMSE로 구한 절대편의보다 작은 값을 가지는 재현기간이 많아지는 시간을 장단기간의 구분점으로 산정하여, 구분점 앞은 RRMSE가 최소인 매개변수를 적용하고 구분점 이후는 RMSE가 최소인 매개변수를 적용하였다(COMBI1). 또 다른 방법(COMBI 2)은 파레토프론트 상에서 RMSE와 RRMSE를 최소로 하는 매개변수를 하나씩 선택한 후, 이 두 매개변수로 구한 확률강우량 곡선이 만나는 지점을 구분점으로 하여 구분점보다 짧은 지속기간에서는 RRMSE를 구분점보다 긴 지속기간에서는 RMSE를 최소로 하는 매개변수를 적용하였다.

그림 3은 위에 설명한 두 가지 방법(COMBI1과 COMBI2)을 이용하여 구한 RMSE, RRMSE값과 앞서 설명한 CASE1에서 구한 파레토프론트를 비교한 그림이다. 그림 3을 살펴보면 다목적 유전자알고리즘을 이용하여 추정된 파레토프론트 상의 최적해들보다 COMBI1과 COMBI2로 구한 최적해의 값이 더 우월한(dominate) 최적해임을 알 수 있으며, 이는 COMBI 1과 COMBI 2를 이용하여 구한 확률강우량의 값이 단순히 다목적 유전자알고리즘을 사용하여 추정된 경우보다는 더 좋은 결과를 나타내는 것을 의미한다고 할 수 있다.

그림 4는 재현기간 200년일 때의 COMBI 1, COMBI 2, 기존 매개변수로 구한 확률강우량, 지점빈도해석으로 구한 확률강우량을 비교한 것이다. At site는 지점빈도해석을 통하여 산정된 확률강우량이고, HEO는

기존 매개변수로 구한 허준행식의 확률강우량, LEE는 이원환식으로 구한 확률강우량고, UNI는 통합형으로 구한 확률 강우량이다(정중호, 1999;이원환, 1980). COMBI 1과 COMBI 2를 사용하여 산정된 확률강우량이 기존 강우강우식으로 구한 확률강우량보다 지점빈도해석을 통해 산정된 확률강우량을 더 잘 추정하고 있는 것을 확인 할 수 있다.

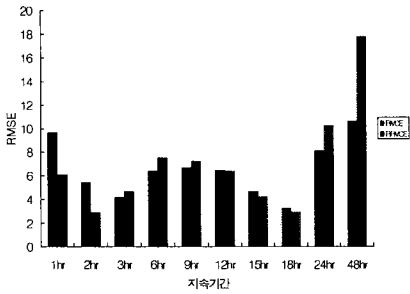


그림 1. 지속기간별 RMSE 평균

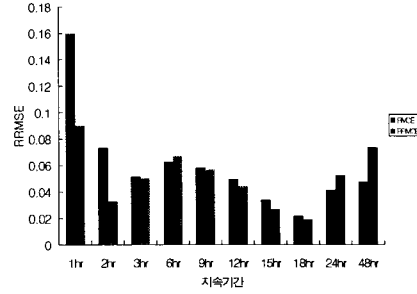


그림 2. 지속기간별 RRMSE 평균

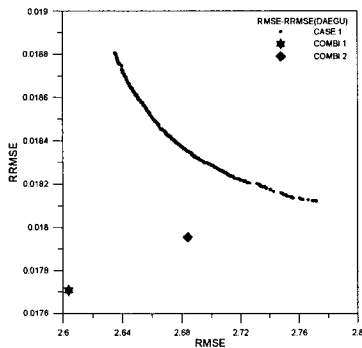


그림 3. RMSE, RRMSE 비교 (CASE 1)

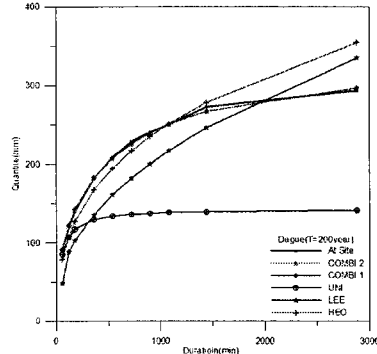


그림 4. 확률강우량 비교

## 5. 결론

강우강도식의 매개변수 추정방법으로 유전자 알고리즘을 적용하였다. 본 연구에서 적용한 3가지 CASE 중에서 CASE 1이 가장 사용하기 적합한 방법으로 나타났으며, 목적함수로 사용된 RMSE는 장기간의 지속기간에서 높은 정확도를 보였고, RRMSE는 단기간의 지속기간에서 높은 정확도를 보였다. 장기간과 단기간의 지속기간에 대한 강우강도식의 매개변수 산정을 위해서 RMSE와 RRMSE를 사용한 다목적 유전자알고리즘을 적용하여 결과를 살펴보았으며, 이 경우 기존의 강우강도식에 의한 결과보다 정확도를 향상시킬 수 있는 것으로 나타났다.

## 감사의 글

이 연구는 건설교통부가 출연하고 한국건설교통기술평가원에서 위탁시행한 2003년도 건설핵심기술연구개발사업(03산학연C03-01)에 의한 도시홍수재해관리기술연구사업단의 연구성과입니다.

## 참 고 문 헌

1. 김경덕, 허준행, 조원철(1995). 지속기간별 강우자료의 분포형에 관한 연구, 대한토목학회 학술발표회 논문집, pp. 25-28.
2. 이원환(1980). 도시 하천 및 하수도 개수계획상의 계획 강우량 설정에 관한 추계학적 해석, 대한토목학회 논문집, 제28권 제4호, pp. 81-94.
3. 이원환, 박상덕, 최성열(1993). 한국대표확률강우강도식의 유도, 대한토목학회 논문집, 제13호 제1호, pp. 115-120.
4. 정중호, 윤용남(1999). 수자원설계실무, 구미서관.
5. 허준행, 김경덕, 한정훈(1999). 지속기간별 강우자료의 적정분포형 선정을 통한 확률강우강도식의 유도, 한국수자원학회 논문집, 제32권 제3호, pp. 247-254.
6. Nash, J. E., Sutcliffe, J. V.(1970). River flow forecasting through conceptual models, Journal of Hydrology, Vol.3 NO.3, pp. 282-290.
7. Deb, K., Anand, A., Joshi, D.(2002). A computational efficient evolutionary algorithm for real-parameter optimization, Evolutionary Computation, Vol.10 NO.3, pp. 371-395.
8. Goldberg, D. E.(1989). Genetic algorithms in search, optimization & machine learning, Addison Wesley, Massachusetts.