

Neyman-Scott Rectangular Pulse 모형의 직접적인 매개변수 추정 Direct Method of Parameter Estimation for Neyman-Scott Rectangular Pulse Model

신주영*, 정창삼**, 허준행***

Juyoung Shin, Chang-Sam Jeong, Jun-Haeng Heo

요 지

Neyman-Scott Rectangular Pulse 모형(NSRPM)은 Poisson process에 기초를 둔 모형으로 수자원분야에서는 강수량 자료를 생성하는데 널리 쓰이고 있다. NSRPM을 구축하기 위해서는 기존에 관측된 강수량 자료를 이용하여 NSRPM의 매개변수를 추정하여야 한다. NSRPM의 매개변수를 추정 시 강수량 자료의 모멘트와 매개변수로 구성된 모멘트식을 비교하여 매개변수를 추정한다. 기존에 사용된 모멘트를 이용한 NSRPM의 매개변수 추정방법의 경우 매개변수로 구성된 모멘트식을 증명하여야지만 NSRPM의 매개변수를 추정할 수 있다. 또한 증명된 모멘트식이 없는 모멘트 값의 경우 매개변수 추정 시 사용하지 못하는 단점이 있다. 이런 한계점으로 인하여 NSRPM의 수정 및 추정이 어려워 NSRPM은 널리 사용되지 못하고 있다. 본 연구에서는 매개변수 추정방법의 다른 한계점을 극복하고자 직접적인 매개변수 추정방법을 제안하였다. 직접적인 매개변수 추정방법은 모멘트 식을 이용하지 않고 생성된 자료를 이용하여 직접적으로 매개변수를 추정하는 방법이다. 본 연구의 대상지점은 금강유역의 대전으로 선정하였으며, 사용된 자료는 기상청에서 운영하는 대전 지상관측소 강수량 자료를 사용하였다. 총 39년의 자료를 이용하여 각 방법을 이용하여 매개변수를 추정하였다. 실험결과 직접적인 추정방법이 기존 매개변수 추정방법보다 더 정확한 매개변수를 추정하는 것을 확인 할 수 있었다.

핵심용어: Neyman-Scott, 구형펄스모형, 매개변수 추정, 유전자알고리즘

1. 서론

세계 곳곳에서 기후변화로 야기되는 이상기후에 의한 피해가 늘고 있다. 우리나라도 이상기후로 인한 피해가 점점 증가하는 추세를 보이고 있다. 이상기후에 대비하기 위해서는 기후변화를 예측한 방재 대책과 수자원 운영 시나리오를 필요로 하는 실정이다. 장기 기후변화 모의에는 전지구 기후모형(Global Climate Model, GCM)을 이용한다. 우리나라의 경우 기상청에서 전지구 기후모형으로 GDAPS(Global Data Assimilation Prediction System)를 사용하고 있다. 전지구 모형을 사용하여 장기 기후변화를 예측 시 결과값은 일자료와 같은 큰 시간 스케일로 예측한다. 전지구 기후모형 수자원 분야에서 사용되는 시간 scale로 변화하기 위해서는 시간을 다운스케일하는 모형이 필요로 하게 된다. Neyman-Scott 구형펄스모형(Neyman-Scott Rectangular Pulse Model,

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

** 정회원 · 인덕대학 건설환경설계학과 전임강사 · E-mail: csjeong@induk.ac.kr

*** 정회원 · 연세대학교 사회환경시스템공학부 토목환경공학전공 교수 · E-mail: jhheo@yonsei.ac.kr

NSRPM)은 점과정(point process)를 이용하여 강우를 생성하는 모형으로써 강우를 생성하는 모형으로 널리 쓰이고 있다. NSRPM를 이용할 경우 강우 시계열 자료의 다운스케일을 용이하게 할 수 있다. NSRPM을 구축하기 위해서는 총 5개의 매개변수를 추정하여야 한다. 일반적으로 사용되는 모멘트법의 경우 매개변수를 추정 시 사용되는 목적함수의 증가에 따라 추정되는 매개변수의 결과가 평탄해 지거나, 강수일수와 무강수일수를 잘 모의하지 못한다. 또한 목적함수를 추가하거나 조정하기 위해서는 복잡한 수식을 다시 계산해야 하는 단점이 있다. 본 연구에서는 NSRPM의 매개변수 추정의 정확도와 사용성을 높이고자 유전자알고리즘(Genetic Algorithm, GA)을 이용한 직접적인 매개변수 추정방법을 적용하였다. 금강유역의 대전지점의 기상청 지상관측소의 시자료를 이용하여 NSRPM의 매개변수를 추정하였다. 기존의 모멘트를 이용하여 추정한 매개변수로 생성된 강우자료와 직접적인 방법을 이용하여 추정한 매개변수로 생성된 강우자료의 정확도를 비교하였다.

2. Neyman-Scott 구형펄스모형(NSRPM)

NSRPM은 Rodriguez-Iturbe 등이 수자원분야에 강수모형으로 처음 적용하였다(Rodriguez-Iturbe et al., 1987). NSRPM은 구형펄스모형(Rectangular Pulse Model)의 강우의 자기상관구조를 표현하지 못하는 단점을 극복하고자 Neyman-Scott 과정(Jerzy Neyman and Elizabeth L. Scott, 1958)을 고려하도록 만들어졌다(Rodriguez-Iturbe et al., 1987).

2.1 NSRPM의 과정

NSRPM은 총 5개의 확률분포형으로 이루어져 있다. 우선 점과정(point process)를 이용하여 강우 군집의 시작점(origin)을 선정한다. 이 때 λ 라는 매개변수가 사용된다. 선정된 강우 군집의 시작에 대하여 기하함수(geometric function)을 이용하여 각 강우 군집의 강우 세포(cell)의 개수를 정한다. 이 때 기하함수(geometric function)은 μ_c 를 매개변수로 갖는다. Exponential 분포를 이용하여 각 강우 세포의 위치를 정한다. 각 강우 세포의 위치를 정할 때 기준이 되는 위치를 강우 군집의 시작점으로 보느냐 아니면 바로 앞 강우세포로 보느냐에 따라 Neyman-Scott 과정과 Bartlett-Lewis 과정으로 나뉜다. 강우의 위치를 정하는 지수 분포는 β 를 매개변수로 갖는다. 강우 cell들의 위치가 정해지면 각 강우 cell들의 지속시간을 지수 분포를 이용하여 정한다. 지속시간을 정하는 분포의 매개변수로는 η 가 사용된다. 지속시간이 정해지면 지수 분포를 이용하여 각 강우 세포들의 강우강도를 정한다. 이때는 ξ 가 사용된다. 이렇게 정해진 강우 세포들의 강우강도들을 중첩하여 시간에 따른 강우를 생성한다(G. Calenda and F. Napolitano, 1999).

2.2 Second order property

NSRPM의 매개변수는 위의 2.1에서 설명했듯이 λ , μ_c , β , η , ξ 총 5개가 사용된다. NSRPM의 매개변수를 추정하기 위해서는 NSRPM의 모멘트를 알아야 한다. NSRPM의 모멘트는 Rodriguez-Iturbe등의 1987년 연구에 의하여 정리 되었으며 그 식은 아래 와 같다(Rodriguez-Iturbe et al., 1987). 식 (1)은 h시간동안의 강수량의 평균을 나타내고 식(2)은 h시간동안의 강수량의 분산을 나타내고 식 (3)은 h시간 동안의 강수량의 상관계수를 나타낸다.

$$E(Y_i^{(h)}) = \frac{\lambda \mu_c h}{\eta \xi} \quad (1)$$

$$Var(Y_i^{(h)}) = \frac{\lambda}{\eta^3}(\eta h - 1 + e^{-\eta h}) \left(\frac{4\mu_c}{\xi^2} + \frac{2\mu_c(\mu_c - 1)\beta^2}{\xi^2(\beta^2 - \eta^2)} \right) - \frac{\lambda(\beta h - 1 + e^{-\beta h})2\mu_c(\mu_c - 1)}{\xi^2\beta(\beta^2 - \eta^2)} \quad (2)$$

$$Cov(Y_i^{(h)}, Y_i^{(h+1)}) = \frac{\lambda}{\eta^3}(\eta h - 1 + e^{-\eta(k-1)h}) \left(\frac{2\mu_c}{\xi^2} + \frac{1\mu_c(\mu_c - 1)\beta^2}{\xi^2(\beta^2 - \eta^2)} \right) - \frac{\lambda(1 - e^{-\beta h})^2 e^{-\beta(k-1)h} 2\mu_c(\mu_c - 1)}{\xi^2\beta(\beta^2 - \eta^2)} \quad (3)$$

3. 매개변수 추정방법

본 연구에서는 NSRPM의 매개변수를 추정하기 위한 최적화 방법으로 유전자 알고리즘을 적용하였다. 모멘트를 이용하여 매개변수를 추정할 경우 통계치는 맞으나 현실에서는 일어날 수 없는 매개변수를 추정하는 경향이 있다. 또한 각 분포형을 변화시키고자 할 때 매개변수로 구성된 복잡한 모멘트식을 적용하여 많은 사람들이 NSRPM를 조정하고 사용하기 어렵게 만드는 요인이 되고 있다. 따라서 직접적인 추정방법을 사용할 경우 이러한 문제들을 해결할 수 있다.

3.2 Moment를 이용한 매개변수 추정

NSRPM의 매개변수를 추정하는 방법으로는 많은 연구가 진행되어 왔다. 그중 모멘트를 이용하여 매개변수를 추정하는 방법이 가장 일반적으로 사용되고 있다. NSRPM의 모멘트에 대해서는 2.2절에서 설명하였다. 각 분포의 매개변수로 표현된 식 (1)~(3)으로 계산된 moment와 실제 관측값의 moment와 차가 가장 적게하는 매개변수를 추정한다. 식 (4)는 P.S.P Cowperwait 등이 제안한 식으로 가장 일반적으로 사용되는 목적함수이다(P.S.P. Cowperwait et al., 1996). 식 (4)의 분자 부분은 관측된 강우의 모멘트이고 분모부분은 식 (1)~(3)을 이용하여 계산된 모멘트이다.

$$Minimize = \left(1 - \frac{E(\widehat{Y_i^{(h)}})}{E(Y_i^{(h)})}\right)^2 + \left(1 - \frac{Var(\widehat{Y_i^{(h)}})}{Var(Y_i^{(h)})}\right)^2 + \left(1 - \frac{Cov(\widehat{Y_i^{(h)}}, \widehat{Y_i^{(h)}})}{Cov(Y_i^{(h)}, Y_i^{(h)})}\right)^2 \dots\dots\dots (4)$$

3.3 직접적인 추정방법

직접적인 추정방법은 매개변수로 구성된 모멘트식을 이용하지 않고 NSRPM의 결과값과 관측값의 통계치를 비교하여 매개변수를 추정하는 방법이다. 컴퓨터 성능의 발달로 기존에는 사용하지 못하였지만 현재에는 NSRPM의 결과를 가지고 직접적으로 매개변수를 추정할 수 있다. 이 경우 결과값을 이용하여 목적함수를 계산하기 때문에 모멘트식을 이용할 때와는 달리 굉장히 직관적으로 목적함수를 구성할 수 있으므로 보다 정확한 매개변수를 추정할 수 있다.

4. 결과 분석

본 연구에서는 금강유역에 위치하고 있는 대전지점을 선정하였다. 금강유역은 우리나라의 중앙에 위치하여 우리나라의 평균적인 기상특성을 보인다. 금강유역의 대표적인 기상청 관측지점인 대전지점은 우리나라 전체에 대한 강수 특성을 잘 대표할 수 있을 것이라 생각되어 대전지점을 선정하였다. 매개변수 추정을 위해 사용된 자료로는 기상청 유인 지상관측소의 관측 강수량 자료가 사용되었다.

NSRPM의 매개변수를 추정 및 강수를 생성 시 Press 등(1992)총 네 개의 무작위변수 생성 알고리즘을 사용하였다. 각 월별 관측 시강수량 자료를 이용하여 NSRPM의 매개변수를 추정하였다.

매개변수 추정 시 목적함수에 적용된 통계량으로는 평균과 1시간, 6시간, 24시간 분산과 1시간, 6시간, 24시간 자기상관계수가 사용되었다. 모멘트를 이용한 방법을 이용할 경우 관측 강수량에 맞는 목적함수 구성과 매개변수 범위를 나누어 매개변수를 추정한다. 본 연구에서는 각 추정방법의 매개변수 추정능력을 검증하기 위해서 매개변수 범위를 구분하거나, 목적함수를 바꾸지 않고 고정된 매개변수 범위와 목적함수를 사용하였다. 표 1은 추정된 매개변수로 생성된 강수량과 관측된 강수량과의 정확도를 나타낸 표이다. 표 1에서 Rel. Err.은 절대상대오차를 나타내고 Err.은 절대오차를 Direct는 직접적인 추정방법을 이용하여 추정된 매개변수로 생성된 강수량의 정확도를 Moment는 모멘트를 이용하여 추정된 매개변수로 생성된 강수량의 정확도를 나타낸다.

각 추정방법의 정확도를 비교해 본 결과 직접적인 추정방법은 절대상대오차가 18.2~35.9%로 절대오차는 5.027~397.499로 나타났으며 모멘트를 이용한 추정방법은 절대상대오차가 30.3%~47.7%로 절대오차는 2.778~407.915로 나타났다. 절대상대오차평균의 경우 직접적인 추정방법이 24.4%로 38.5%인 모멘트를 이용한 방법보다 정확한 값을 추정하는 것을 확인할 수 있다. 절대상대오차의 경우에서도 직접적인 추정방법이 135.797로 149.562인 모멘트를 이용하여 추정한 매개변수 보다 실제 강수를 잘 모의하는 것으로 나타났다.

표 1. 추정된 매개변수의 정확도 비교

Month	Direct		Moment	
	Rel. Err.	Err.	Rel. Err.	Err.
Apr.	22.2%	9.138	30.3%	5.021
May	19.3%	5.309	39.0%	53.774
Jun.	22.3%	45.310	34.3%	38.036
Jul.	33.8%	230.719	42.9%	244.810
Aug	35.9%	257.574	47.7%	294.598
Sp	19.2%	397.499	37.2%	407.915
Oct	18.2%	5.027	38.5%	2.778
Ave.	24.4%	135.797	38.5%	149.562

5. 결론

본 연구는 대전지검 관측 강수량 자료를 이용하여 NSRPM의 새로운 매개변수 추정방법인 직접적인 추정방법의 사용가능성과 적용적합성을 평가하였다. 직접적인 추정방법의 사용가능성과 적용적합성을 평가하기 위하여 기존에 널리 사용되었던 모멘트를 적용한 매개변수 추정방법과의 모의 성능을 비교하였다. 비교결과 직접적인 매개변수 추정방법은 모멘트를 이용한 매개변수 추정방법보다 높은 적용적합성을 보였다. 이런 결과로 미루어 보아 직접적인 방법은 NSRPM의 매개변수를 추정하는 한 방법으로서의 적용이 가능함을 알 수 있다. 향후 비교 연구를 통하여 직접적인 매개변수 추정방법의 적용가능성과 여러 장점들을 증명하는 연구가 필요하다.

감사의 글

본 연구는 21세기 프론티어연구개발사업인 수자원의 지속적 확보기술개발사업단의 연구비지원(과제번호: 1-6-3)에 의해 수행되었습니다.

참고문헌

1. G. Calenda, and F. Napolitano. (1999). Parameter estimation of Neyman–Scott processes for temporal point rainfall simulation, *Journal of Hydrology*, Vol. 225, pp. 45–66.
2. Goldberg, D.E. (1989). Genetic algorithm in search, optimization & machine learning,, *addison Wesley*, Massachusetts.
3. I. Rodriguez–Iturbe, D. E. Cox, and Valerie Isham. (1987). Some Models for Rainfall Based Stochastic Point Process, *Proceedings of the Royal Society of London, Series A*, Vol. 410, No. 1839, pp. 269–288.
4. Jerzy Neyman, and Elizabeth, L. Scott. (1958). Statistical Approach to Problem of Comology, *Journal of the Royal Statistical, Society Series B*, Vol. 20, No. 1, pp.1–43.
5. Press, W. H., Teukolsky, S. A., Vetterling, W. T. and Flannery, B. P. (1992), Numerical Recipes in C The Art of Scientific Computing Second Edition, *CAMBRIDGE UNIVERSITY PRESS*, United States of America, pp. 274–286.
6. P. S. P. Cowpertwait, P. E. O'Connell, A. V. Metcalfe, and J. A. Mawdsley (1996). Stochastic point process modelling of rainfall. I. Single–site fitting and validation, *Journal of Hydrology*, Vol. 177, pp. 17–46.