# A statistical inference for Neyman-Scott Rectangular Pulse model

Nam Hee  $\operatorname{Kim}^a$  · Yongku  $\operatorname{Kim}^{a,1}$ 

<sup>a</sup>Department of Statistics, Kyungpook National University

(Received May 23, 2016; Revised June 28, 2016; Accepted July 7, 2016)

#### Abstract

The Neyman-Scott Rectangular Pulse (NSRP) model is used to model the hourly rainfall series. This model uses a modest number of parameters to represent the rainfall processes and underlying physical phenomena such as the arrival of a storm or rain cells. In this paper, we proposed approximated likelihood function for the NSRP model and applied the proposed method to precipitation data in Seoul.

Keywords: approximated likelihood, maximum likelihood estimation, Neyman-Scott Rectangular Pulse model, precipitation model

#### 1. 서론

근래 들어 전 세계적으로 이상기후로 인한 자연재해의 빈도와 강도가 심해지고 있다. 우리나라도 이 상기후로 인한 소위 게릴라식 집중호우 또는 태풍으로 인한 돌발홍수가 국지적으로 발생하거나 극심 한 가뭄에 따른 수도권의 주상수원인 소양강댐의 수위가 1973년 댐 준공 이후 역대 최저치까지 떨어 지는 일까지 발생하고 있다. 이러한 자연재해에 대한 중요한 예방대책은 다가올 재해에 대비하여 사 전조치를 취함으로써 피해를 최소화하는 것인데, 이는 과거의 재해특성, 현재의 재해환경, 그리고 예 측되는 미래의 재해에 대비하여 결정하게 된다. 특히 강우예측의 경우에는 강우 분포의 파악이 중요 한데, 일반적으로 강우의 분포는 시간적으로 분에서 연단위를 가지고 공간적으로는 수 킬로제곱미터에 서 수천 킬로 제곱미터를 가지고 있으며 강우발생 요인들 사이의 상호작용에 대한 지식의 한계로 현 재 강우의 예측은 극복해야 할 문제점이 많다. 강우를 생성하는 모형으로 현재 Neyman-Scott 구형펄 스모형(Neyman-Scott Rectangular Pulse model; NSRPM)이 많이 사용되고 있다. NSRPM은 점과 정(point process)을 이용하여 강우를 생성하는 모형으로 강우의 발생, 강우세포(rain cell)의 강우강도 그리고 지속시간의 분포로 표현된다. 특히 이 모형은 구형펄스모형(Rectangular Pulse model)에서 포 함하지 않았던 강우사상의 군집특성을 반영하였다는 장점을 가지고 있다. 기존의 구형펄스모형의 경우 포아송 분포를 이용하여 임의의 강우를 생성하였으며, 강우의 자기상관구조를 표현하지 못한 단점을 가 지고 있었으나, Rodriguez-Iturbe 등 (1987)이 구형펄스모형에 Neyman-Scott 과정 (Neyman과 Scott, 1958)을 고려한 Neyman-Scott 구형펄스모형(Neyman-Scott Rectangular Pulse model; NSRPM)을 소개하였다. 이러한 강우세포의 군집특성에 기초한 모형이 실제 관측된 강우자료를 더 적합하게 모의하

This work is based on part of Nam Hee Kim's Master Thesis.

<sup>&</sup>lt;sup>1</sup>Corresponding author: Department of Statistics, Kyungpook National University, 80 Daehak-Ro, Daegu 41566, Korea. E-mail: kim.1252@knu.ac.kr

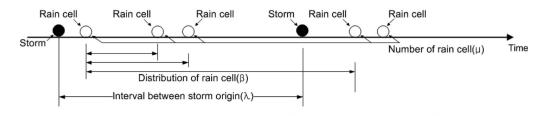
였으며 서로 다른 강우의 총계시간(aggregation time)에 따른 극치값(extreme values)도 잘 고려하였다. 또한 NSRPM의 강우세포의 지속시간을 나타내는 분포형을 추가하여 보다 정확한 강우의 군집특성을 나타내는 연구가 수행되었다 (Rodriguez-Iturbe 등, 1988; Entekhabi 등, 1989).

NSRPM 이외에 점과정을 이용한 강우생성모형으로는 Bartlett-Lewis Rectangular Pulse model (BLRPM) (Rodriguez-Iturbe 등, 1987, 1988)이 있는데, 두 모형은 3개의 독립적인 확률과정으로 구성된다. 첫째는 강우의 사상의 시작위치를 결정짓는 과정이고, 둘째는 각각의 강우 사상에 속해있는 강우세포의 개수를 결정짓는 과정이며 셋째는 강우세포 위치를 결정짓는 과정이다. 이 때, BLRPM과 NSRPM의 차이점은 강우세포의 위치를 결정짓는 기준에서 차이가 있다. 강우세포의 위치를 결정하 는데 있어 기준이 되는 위치를 강우 사상의 시작점으로 보느냐 바로 앞 강우세포로 보느냐에 따라 달 라진다. 그러나 BLRPM은 매개변수를 추정하는데 있어 초기값에 예민하다는 단점을 가지고 있어 NSRPM이 더 적합한 강우모형으로 알려져 있다 (Velghe 등, 1994). 점 과정을 기본으로 한 연속 모형 은 강우강도의 시간적 변화패턴을 표현할 수 있는데, 일반적으로 강우의 관측 자료는 정해진 관측시간 동안에 누적해서 얻어진 자료이기 때문에 강우를 모의하기 위해서는 강우모형의 매개변수를 추정하여 야 한다. 최근 이들 강우모형의 매개변수를 추정하는 연구도 활발히 진행되고 있는데, Cowperwait 등 (1996)은 습윤 지속기간과 건조 지속기간의 변화 확률을 이용한 NSRPM의 매개변수 추정수식을 제시 하였으며 여기에서 나타난 시계열 강우가 lag 1인 자기상관계수를 이용한 시계열 강우보다 더 잘 모의 한다는 것을 보였다. Calenda와 Napolitano (1999)는 강우의 총계시간의 스케일이 작으면 NSRPM의 매개변수 추정에 사용되는 목적함수(objective function)값이 평탄해지므로 NSRPM으로 생성된 강우 의 시간적인 변화를 이용하여 매개변수를 추정하는 연구를 하였다. 또한, NSRPM의 매개변수 추정에 서 최적화 기법을 이용시에는 목적함수가 비선형이거나 미분불가능일 경우, 직접적인 매개변수 추정으 로 Nelder-Mead 기법이나 유전자 알고리즘이 이용된다 (Bunday와 Garside, 1987; Goldberg, 1989). 국내에서도 NSRPM에 대한 연구가 활발히 이루어졌는데, Kim 등 (1998)은 NSRPM을 이용하여 강 우 시계열을 모의하였고, Yu 등 (2001)은 점 과정을 이용하여 개개의 강우사상의 발생의 중첩을 고 려한 모의와 고려하지 않은 모의를 비교하였다. Kum 등 (2001)은 NSRPM의 매개변수를 추정하기 위한 최적화 기법 중 비선형계획법(nonlinear programming; NLP)과 유전자 알고리즘(genetic algorithm; GA)의 정확도 비교 연구를 통해 유전자 알고리즘을 이용한 매개변수 추정이 더 높은 정확도를 가짐을 보였다. Shin 등 (2011)과 Jung 등 (2009)은 직접적인 방법과 moment 방식을 이용하여 얻은 NSRPM의 추정된 매개변수의 적합성을 비교하여 전체적인 정확도에서 직접적인 방법의 매개변수 추정 이 더 높은 정확도를 보였다. 특히 moment 방법을 이용하여 매개변수를 추정할 경우, 목적함수를 추 가하거나 조정하기 위해서는 복잡한 수식을 다시 계산하여야 하는 단점이 있으며, 전체적인 강우의 특 성을 반영하기 어렵고 스케일에 따른 추정값의 변동도 크게 나타난다. 또한 moment를 이용한 추정값 은 추정오차를 구할 수 없기 때문에 신뢰구간을 구할 수 없다는 단점이 있다. 하지만 실제로 관측된 데 이터는 누적강수량이고 또 이 강수량에 대한 분포를 정의하기가 어려우므로 정확한 우도함수(likelihood function)에 의한 매개변수 추정에는 한계가 있다.

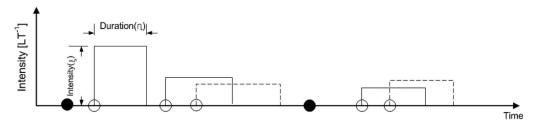
이에 본 연구에서는 누적강수량에 대한 근사적인 우도함수(approximated likelihood function)를 소개하고 이를 통해 NSRPM의 매개변수를 추정하고자 한다. 또한 분석에 사용되는 누적강수량의 시간 스케일에 따른 추정치의 변동성도 함께 알아보고자 한다.

# 2. 추계학적 강우 모형: NSRPM

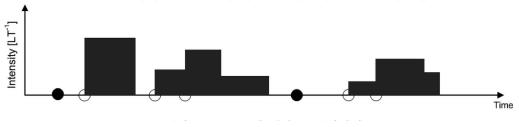
NSRPM이 도입되기 이전에 기존 수문학 분야에서 강우 생성을 위해 Rectangular Pulse model(RPM) 이 주로 사용되었는데, 포아송분포를 이용해 임의의 강우를 생성하는 모형이다. 그러나 각 강우 사상이



(a) Storm의 시작과 Rain Cell의 개수와 시작 위치를 결정한다.



(b) 각각의 Rain Cell의 지속시간과 강우강도를 결정한다.



(c) 각 Rain Cell의 강우를 중첩한다.

Figure 2.1. Constructing Neyman-Scott Rectangular Pulse model.

독립적으로 나뉘어져 있어서, 실제 강우의 군집화된 특성을 표현하기가 어렵고 강우사상들의 자기상관 구조를 표현하는데에도 한계가 있다. 이에 Rodriguez-Iturbe 등 (1987)은 RPM에서 반영하지 못한 강우의 군집특성을 보여주기 위해 Neyman-Scott Rectangular Pulse model(NSRPM)을 도입하여 수공학 분야에서 처음 제시하였다. 이 후 모델의 각 분포형을 조정하여 강우를 보다 더 정확하게 모의하는 연구를 진행하였으며 (Rodriguez-Iturbe 등, 1988; Entekhabi 등, 1989), 현재에는 지속 기간과 강우 강도의 상관성까지 고려한 NSRPM이 연구되고 있다 (Kim과 Kavvas, 2006; Evin과 Favre, 2008).

#### 2.1. NSRPM의 구축과정

NSRPM은 크게 3개의 확률과정으로 이루어져 있고 이들 확률과정은 5개의 매개변수로 구성되어 있다. NSRPM에서는 먼저 강우사상의 시간적 시점이 결정되고 그 후에 각 강우사상에 속하는 강우세포의 개수를 결정한다. 마지막으로 강우 세포의 위치를 결정하는데 이 때 3의 확률과정은 각각 독립이다. 강우사상의 시작 위치, 그에 따른 강우세포의 개수 그리고 위치가 결정되면 각 강우세포는 임의적으로 강우세포의 지속시간과 지속강도를 가진다. 여기에서 각각의 매개변수들은 서로 독립적인 분포를 가진다.

구체적인 NSRPM의 구축과정은 다음과 같다.

먼저 강우사상의 시작 위치(origin of storm)는 모수  $\lambda$ 를 가지는 지수분포를 따른다고 가정한다. 다음 주어진 강우사상의 시작 위치에서 각 강우사상을 이루는 강우세포(rain cell)의 개수를 결정하게 되는데 이때 강우세포의 개수는 모수  $\mu$ 를 가지는 기하분포를 따른다고 가정한다. 일단 강우세포의 개수가 결정되면 각 강우세포의 위치를 정해야 하는데 각 강우세포의 위치에 대하서는 모수  $\beta$ 를 가지는 지수분포를 가정한다. 이 과정은 Figure 2.1(a)과정을 통해 시각적으로 나타낸다. 강우사상의 시작 위치와 그에 따른 강우세포의 개수 및 위치가 결정되면 각 강우세포는 임의적으로 강우세포의 지속시간과 지속강도를 가진다. 이 때 강우세포의 지속시간은 모수  $\eta$ 를 가지는 지수분포를 따르고, 강우세포의 강우강도는 모수  $\xi$ 를 가지는 지수분포를 따른다고 가정하낟. 이 과정은 Figure 2.1의 (b)과정을 통해 시각적으로 보여준다. 마지막으로 강우사상에 속하는 강우세포의 강우강도를 중첩하여 시간에 따른 강우를 생성한다 (Calenda와 Napolitano, 1999). 위의 NSRPM의 구축과정을 정리하여 나타내면 다음과 같다.

- 1. 강우사상의 시작 위치(origin of storm)는 모수  $\lambda$ 를 가지는 지수분포를 따른다.
- 2. 강우세포(rain cell)의 개수는 모수  $\mu$ 를 가지는 기하분포를 따른다.
- 3. 강우세포의 위치는 모수  $\beta$ 를 가지는 지수분포를 따른다.
- 4. 각각의 강우세포의 지속시간은 모수  $\eta$ 를 가지는 지수분포를 따른다.
- 5. 각각의 강우세포의 강우강도는 모수 ξ를 가지는 지수분포를 따른다.
- 6. 정해진 강우세포의 지속기간에 따른 강우강도를 중첩한다.

# 2.2. NSRPM의 통계적 특성

시간 t에서의 h시간 누적 강우량의 평균과 분산을 각각  $E[Y^{(h)}(t)]$ 와  $Var[Y^{(h)}(t)]$ 라고 하고, lag가 k인 h시간 누적 강우량의 공분산을  $Cov[Y^{(h)}(t),Y^{(h)}(t+k)]$ 라고 한다면, Rodriguez-Iturbe 등 (1987)이 제시한 NSRPM의 moment는 다음과 같다.

$$E\left[Y^{(h)}(t)\right] = \lambda \eta^{-1} \mu \xi h, \tag{2.1}$$

$$\operatorname{Var}\left[Y^{(h)}(t)\right] = \lambda \eta^{-3} \left(\eta h - 1 + e^{-\eta h}\right) \left[4\mu \xi^2 + 2(\mu^2 - \mu)\xi^2 \frac{\beta^2}{\beta^2 - \eta^2}\right]$$

$$-\lambda \left(\beta h - 1 + e^{-\beta h}\right) \frac{2(\mu^2 - \mu)\xi^2}{\beta(\beta^2 - \eta^2)}, \tag{2.2}$$

$$\operatorname{Cov}\left[Y^{(h)}(t), Y^{(h)}(t+k)\right] = \lambda \eta^{-3} \left(1 - e^{-\eta h}\right)^2 e^{-\eta(k-1)h} \left[2\mu x i^2 + (\mu^2 - \mu)\xi^2 \frac{\beta^2}{\beta^2 - \eta^2}\right]$$

$$-\lambda \left(1 - e^{-\beta h}\right)^2 e^{-\beta(k-1)h} \frac{(\mu^2 - \mu)\xi^2}{\beta(\beta^2 - \eta^2)}. \tag{2.3}$$

#### 3. 매개변수의 모수 추정

# 3.1. Method of Moments를 이용한 추정

강우 예측을 위한 추계학적 강우 모형인 NSRPM의 매개변수 추정하는 방법으로 많은 연구들이 진행되어 왔다. 그러나 앞서 제시한 NSRPM의 모형에서 볼 수 있듯이 NSRPM은 강우사상과 강우세포의 관련된 매개변수가 존재한다. 즉 실제로 얻어지는 데이터인 강우량만으로 NSRPM의 매개변수에 관한 정

보를 얻을 수 없다. 이에 Rodriguez-Iturbe 등 (1987)이 제시한 NSRPM의 moment를 이용하여 매개 변수를 추정해왔다. 가장 일반적인 추정방법은 Moment를 이용한 방법이다. Moment를 이용하여 매개변수를 추정할 때는 목적함수를 필요로 한다. 다음 식은 가장 일반적으로 사용하는 Cowperwait 등 (1996)이 제시한 목적함수로  $\hat{E}[Y^{(h)}(t)], \widehat{\mathrm{Var}}[Y^{(h)}(t)]$  그리고  $\widehat{\mathrm{Cov}}[Y^{(h)}(t), Y^{(h)}(t+k)]$ 는 실제 관측된 h시간 누적 강우량을 이용한 moment 추정량이고,  $E[Y^{(h)}(t)], \widehat{\mathrm{Var}}[Y^{(h)}(t)]$ 와  $\widehat{\mathrm{Cov}}[Y^{(h)}(t), Y^{(h)}(t)]$ 와  $\widehat{\mathrm{Cov}}[Y^{(h)}(t), Y^{(h)}(t+k)]$ 는 앞서 설명한 NSRPM의 moment이다. 이는 위에서 제시한 NSRPM의 moment와 실제 관측값의 moment의 차가 가장 작게하는 모형변수를 추정한다. 따라서 앞에서 언급한 moment에 기초한 식 (3.1)과 같은 목적함수를 이용하여 매개변수의 모수를 추정할 경우, 추정해야 하는 모수의 개수가 5개이므로 목적함수도 최소한 5개를 필요로 한다.

$$\min \left\{ \left( 1 - \frac{\widehat{E}\left[Y^{(h)}(t)\right]}{E\left[Y^{(h)}(t)\right]} \right)^2 + \left( 1 - \frac{\widehat{\operatorname{Var}}\left[Y^{(h)}(t)\right]}{\operatorname{Var}\left[Y^{(h)}(t)\right]} \right)^2 + \left( 1 - \frac{\widehat{\operatorname{Cov}}\left[Y^{(h)}(t), Y^{(h)}(t+k)\right]}{\operatorname{Cov}\left(Y^{(h)}(t), Y^{(h)}(t+k)\right)} \right)^2 \right\}. \quad (3.1)$$

이에 Kim과 Yoo (2008)는 목적함수를 구성할 때 목적함수의 항이 증가할수록 개별적인 정확도가 떨어지는 경향이 발생하므로 목적함수의 항의 개수를 5개 또는 6개로 하는 것이 적절하다고 제시하였다. 따라서 NSRPM의 moment를 이용하여 매개변수를 추정할 때 목적함수를 위해서 최소한 3개의 항이 필요하다. 이 때 보다 정확한 매개변수 추정을 위해 누적시간 h와 공분산의 lag인 k를 적절하게 구성하여야 하는데, Calenda과 Napolitano (1999)은 누적시간 h의 차이가 클수록 매개변수를 정확하게 추정할수 있다는 것을 보여주었다. 하지만 정확한 추정을 위해 어떤 누적시간 h와 lag인 k를 사용해야 하는지에 대해서는 좀 더 연구가 필요한 것으로 알려져 있다. 따라서 앞서 소개된 Moment를 이용하여 매개변수를 추정할 때, 어떤 lag k를 목적함수에 사용하느냐에 따라서 매개변수의 추정값이 달라질 수 있다는 단점이 있다. 그러므로 특정시간 스케일의 강우만을 이용하여 추정시에는 시간 스케일에 따라 추정값의변동이 크게 나타나 전체적인 강우의 특성을 반영하기 힘들다. 또한 추정값의 추정오차를 구할 수 없으므로 신뢰구간과 같은 불확실성에 대한 설명에는 한계가 있다. 그 이외에 Moment를 이용한 추정의 특성상 복잡한 수식을 계산해야 하는 어려움을 가진다. 따라서 본 연구에서는 근사적인 우도함수를 이용하여 NSRPM 매개변수의 모수를 추정하고자 한다.

# 3.2. 근사적인 우도함수를 이용한 모형모수 추정

여러가지 확률모형의 모수를 추정하기 위해 우도함수를 이용한 방법이 보편적으로 많이 이용되어 왔다. 먼저  $Y_i^{(h)}(t)$ 를 i번째 연도 t번째 시간에 관측된 h시간 누적 강수량이라고 하면  $(i=1,\ldots,n,\ t=1,\ldots,T)$ , NSRPM 모수에 대한 우도함수는  $L(\lambda,\mu,\beta,\eta,\xi)=\prod_{i=1}^n f(y_i^{(h)}(1),\ldots,y_i^{(h)}(T);\lambda,\mu,\beta,\eta,\xi)$ 로 표현할 수 있다. 여기에서  $f(y_i^{(h)}(1),\ldots,y_i^{(h)}(T);\lambda,\mu,\beta,\eta,\xi)$ 는 i번째 연도의 h시간 누적 강수량에 대한 결합확률밀도함수이다. 하지만 NSRPM 가정하에서 누적 강수량의 확률밀도함수는 앞에서 정의된 5가지 매개변수의 분포함수가 복잡하게 섞여있는 구조이다. 따라서 정확한 형태의 확률밀도함수를 표현하기가 어렵고 이를 통해 우도함수로 표현하는데는 한계가 있다. 또한 데이터를 이용한 비모수적 분포추정은 기존의 매개함수와의 연관성을 보여주기가 어렵다는 한계가 있다. 단, NSRPM의 moment는 잘 알려져 있으므로 관측된 h시간 누적 강수량과 이들 moment와의 관계를 이용하여, 근사적인 우도함수를 제시하고 이를 이용한 NSRPM의 모형모수를 추정하고자 한다. 먼저 NSRPM의 매개 변수를 추정하기 위해 제시한 근사적인 우도함수는 관측값의 수가 충분히 많을 때, 중심극한정리에 의해서 관측값을 가지고 추정한 moment(평균, 분산, 공분산 또는 상관계수)가 실제 moment를 중심으로한 근사적인 정규분포 형태를 가진다는 가정에서 시작한다 (Kim과 Kim, 2016). 여기에 공분산의 경우,

각각의 lag가 k인 추정된 공분산은 실제 공분산이 주어졌을 때, lag가 k'인 공분산과는 서로 독립이라는 가정을 바탕으로 NSRPM의 매개변수 추정을 위한 근사적 우도함수를 다음과 같이 제시한다.

$$\tilde{L}(\mu, \beta, \lambda, \eta, \xi) \propto \exp\left(-\alpha_1 \sum_{i=1}^n \frac{\left(\bar{Y}_i^{(h)} - E\left[Y^{(h)}(t)\right]\right)^2}{2\sigma_1^2}\right) \times \exp\left(-\alpha_2 \sum_{i=1}^n \frac{\left(\log\left\{\hat{V}_i^{(h)}\right\} - \log\left\{\operatorname{Var}\left[Y^{(h)}(t)\right]\right\}\right)^2}{2\sigma_2^2}\right) \times \exp\left(-\sum_{i=1}^n \sum_{k=1}^K \alpha_{3,k} \frac{\left(\log\left\{\hat{C}_{i,k}\right\} - \log\left\{\operatorname{Cov}\left[Y^{(h)}(t), Y^{(h)}(t+k)\right]\right\}\right)^2}{2\sigma_{3,k}^2}\right)^2, (3.2)$$

여기에서

$$\begin{split} \bar{Y}_i^{(h)} &= \frac{\sum\limits_{t=1}^T Y_i^{(h)}(t)}{T}, \\ \hat{V}_i^{(h)} &= \frac{\sum\limits_{t=1}^T \left(Y_i^{(h)}(t) - \bar{Y}_i^{(h)}\right)^2}{T-1}, \\ \hat{C}_{i,k} &= \frac{1}{T-k-1} \sum\limits_{t=1}^{T-k} \left(Y_i^{(h)}(t) - \bar{Y}_i^{(h)}\right) \left(Y_i^{(h)}(t-k) - \bar{Y}_i^{(h)}\right). \end{split}$$

이 때,  $\bar{Y}_i^{(h)}$ ,  $\hat{V}_i^{(h)}$  그리고  $\hat{C}_{i,k}$ 는 i번째 연도에 관측된 h시간 누적 강우량을 이용하여 구한 평균, 분산 그리고 공분산 추정값들이고  $E[Y^{(h)}(t)]$ ,  $Var[Y^{(h)}(t)]$  그리고  $Cov[Y^{(h)}(t),Y^{(h)}(t+k)]$ 는 앞서 Rodriguez-Iturbe 등 (1987)이 제시한 NSRPM의 moment들이다. 여기에서  $\sigma_1^2$ ,  $\sigma_2^2$  그리고  $\sigma_{3,k}^2$ 는 각추정값들의 분산으로서 일반적으로 관측값을 이용하여 추정되며,  $\alpha_1,\alpha_2,\alpha_{3,k}$ 는 임의의 상수로 각각의 평균, 분산, 공분산에 대한 가중치를 나타낸다. 또한 기존 연구의 결과를 바탕으로 본 연구에서는 우도 함수에 고려되는 공분산의  $\log$  범위를 K=10으로 제한하였다. 근사적인 우도함수 (3.2)를 이용한 모수추정에서는 표본을 통해 얻은 우도함수를 최대로 하는 모수를 찾는 최우추정법(maximum liklihood estimation; MLE)이 일반적으로 사용되므로, 앞에서 정의한 근사적인 우도함수인  $\tilde{L}(\mu,\beta,\lambda,\eta,\xi)$ 를 최대로 만드는  $\mu,\beta,\lambda,\eta$  그리고  $\xi$  값을 찾으면 이 값들이 NSRPM 매개변수의 모수에 대한 근사적인 최우추정량이 된다.

#### 4. 자료분석

본 논문에서 제안된 NSRPM에 대한 근사적인 우도함수를 실제 데이터에 적용해보기 위하여 기상청에서 제공하는 한강유역의 14개 기상관측소 지점 중 서울지점에서 관측된 데이터에 적용하였다. 기간은 1973년부터 2011년까지 39년 동안의 시간당 시계열 강우 자료이며, 연 강수량의 50-60%의 강우가 여름에 내린다는 기상청의 정보를 통해 각각의 해당 연도의 6, 7, 8월 동안 내린 강우 데이터를 이용하여 모형에 적용시켰다. Table 4.1에서는 매개변수를 추정하기 위해 설정한 매개변수모수의 범위를 나타낸 것으로 Calenda와 Napolitano (1999)가 제시한 범위를 참고하였다. 앞서 언급했듯이 Moment를 이용하여 NSRPM의 매개변수를 추정할 때는 한계점이 있다. 추정하고자 하는 매개변수의 모수가 5개이므

**Table 4.1.** Ranges of NSRPM model parameters  $(\lambda, \mu, \beta, \eta, \xi)$ 

Parameters	$\lambda$	$\mu$	β	$\eta$	ξ
Minimum	0.001	2.0	0.01	0.10	0.3
Maximum	0.050	100.0	0.50	10.00	15.0

**Table 4.2.** Method of Moments estimator of NSRPM model parameters  $(\lambda, \mu, \beta, \eta, \xi)$  using summer precipitation in Seoul under various objective functions

lag $k$ used in $(3.1)$	λ	$\mu$	β	η	ξ
1hr, 2hr, 4hr	0.0138	21.4227	0.3620	7.5294	10.0123
1hr, 3hr, 6hr	0.0121	15.1873	0.1681	4.0743	8.6132
1 hr, 6 hr, 12 hr	0.0140	15.0640	0.3130	2.6480	5.2510
1hr, 9hr, 18hr	0.0101	12.2546	0.1671	2.3253	7.3807
1 hr, 12 hr, 24 hr	0.0094	11.3217	0.1364	2.0217	7.4635

**Table 4.3.** Approximated maximum likelihood estimator of NSRPM model parameters  $(\lambda, \mu, \beta, \eta, \xi)$ 

Parameters	$\lambda$	$\mu$	$\beta$	$\eta$	ξ
Est.	0.0105	10.4022	0.1515	1.3469	4.8471
S.E.	0.0013	2.0180	0.0233	0.4626	0.8976

로 최소 5개 이상의 목적함수를 필요로 하다. 이를 위해서 다양한 방법으로 목적함수를 구성하여 각각의 매개변수 추정값을 알아보았다. Table 4.2에서는 목적함수의 구성에 따라 매개변수 추정 정확도가 변한다는 것을 보여주었다. 참고로 제약조건이 없는 비선형 함수의 문제를 해결하기 위하여 최적화기법으로 Differential Evolution Algorithm을 이용하여 구하였다.

Table 4.3에서는 NSRPM 매개변수의 모수인  $(\lambda, \mu, \beta, \eta, \xi)$ 에 대한 근사적인 우도함수를 이용한 최우추 정량을 각각의 추정오차와 함께 보여주었다. 이를 통해 신뢰구간과 같은 추정값에 대한 불확실성에 대한 접근이 가능하다.

실제로 제안된 NSRPM 매개변수의 모수에 대한 근사적인 최우추정량이 실제 강우의 군집화된 특성 등을 잘 표현하는지를 알아보기 위해서 매개변수의 추정 모수값을 이용한 Rodriguez-Iturbe 등 (1987)이 제시한 NSRPM의 moment 값과 실제 관측값을 이용해 구한 평균, 분산, 공분산 그리고 자기상관계수값들을 비교하였다. Figure 4.1은 서울지역에 1시간 동안 내린 강우량의 데이터를 이용하여 구한 평균, 분산, 공분산 그리고 자기상관계수 값들의 상자그림을 근사적 우도함수를 이용한 추정값을 이용한 NSRPM의 moment 값과 함께 보여주었다.

# 5. 결론

기후변화가 야기하는 요인들의 변화에 대한 관심과 경각심이 고조됨에 따라 일반적인 강우의 변동 특성 분석은 강우사상의 특성을 지속시간에 따른 총량적 개념으로 접근하여 분석하였지만 실제로 강우구조의 변화를 파악하는데 한계가 있었다. Rodriguez-Iturbe 등 (1987)에 의해 NSRPM이 소개되면서 현실적인 반영을 위해 모형 조정과 수정 등 많은 연구가 이루어졌으며 적합한 매개변수를 추정하기 위한 연구도 활발히 진행되어왔다. 그 중 Moment를 이용하여 매개변수를 추정해온 연구가 많았지만 Moment를 이용하여 매개변수 추정시 강우 시간스케일에 영향을 받아 사용된 강우 시간스케일에는 통계적 특성을잘 반영하였지만 다른 시간 스케일에서는 적합하지 못한 단점을 보였다. 따라서 전체적인 강우 특성을 반영하는데는 한계가 있다. 이를 보완하고자 시간당 누적강우량에 대한 근사적인 우도함수를 소개하였

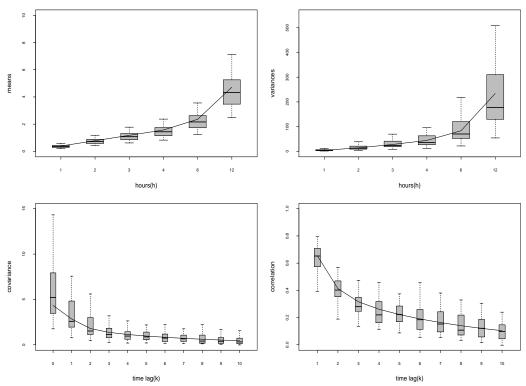


Figure 4.1. Boxplots of observed mean, variance, covariance and correlation using 1 hour precipitations with corresponding estimator; clockwise from top left.

고 이를 이용하여 NSRPM 매개변수의 모수를 추정한 결과 전체적인 통계적 특성을 반영하여 추정시 사용한 시간 스케일 이외에도 다른 시간 스케일에도 매개변수가 잘 적합함을 보여주었다. 또한 추정량의 추정오차를 이용하여 추정값의 신뢰구간을 얻을 수 있었다. 따라서 근사적인 우도함수를 통한 매개변수 추정으로 인한 추계학적 강우 생성모형인 NSRPM의 개선으로 향후 더 발전된 강우특성 분석과 강우예측으로 재해대책에 대한 예방대책의 발전을 기대해본다.

# References

Bunday, B. D. and Garside, G. R. (1987). Optimisation Methods in Pascal, Edward Arnold Publishers, London.

Calenda, G. and Napolitano, F. (1999). Parameter estimation of Neyman-Scott processes for temporal point rainfall simulation, *Journal of Hydrology*, 225, 45–66.

Cowperwait, P. S. P., O'Connell, P. E., Metcalfe, A. V., and Mawdsley, J. A. (1996). Stochastic point process modelling of rainfall. I. Single-site fitting and validation, *Journal of Hydrology*, **175**, 17–46.

Entekhabi, D., Rodriguez-Iturbe, I., and Eagleson, P. S. (1989). Probabilistic representation of the temporal rainfall by a modified Neyman-Scott rectangular pulse model: parameter estimation and validation, Water Resources Research, 25, 295–302.

Evin, G. and Favre, A. C. (2008). A new rainfall model based on the Neyman-Scott process using cubic copulas, *Water Resources Research*, 44, W03433.

- Goldberg, D. (1989). Generic Algorithms, Addison Wesley, Boston.
- Jung, C. S. (2009). Study of direct parameter estimation for Neyman-Scott rectangular pulse model, Journal of Korea Water Resources Association, 42, 1017–1028.
- Kim, J. H., Lee, J. S., Lee, J. J., and Son, K. I. (1998). A modeling of daily precipitation series using the Poisson cluster process, Journal of the Korean Society of Civil Engineers, 18, 231-241.
- Kim, K. W. and Yoo, S. C. (2008). A selection of the point rainfall process model considered on temporal clustering characteristics, *Journal of Korea Water Resources Association*, **41**, 747–759.
- Kim, S. and Kavvas, M. (2006). Stochastic point rainfall modeling for correlated rain cell intensity and duration, *Journal of Hydrologic Engineering*, 11, 29–36.
- Kim, Y. and Kim, D. H. (2016). An approximate liklihood function of spatial correlation parameters, Journal of the Korean Statistical Society, 45, 276–284.
- Kum, J. H., Ahn, J. H., Kim, J. H., and Yoon, Y. N. (2001). Parameter estimation of a point rainfall model, Neyman-Scott rectangular pulses mod. In *Proceedings of Korea Water Resources Association 2001*, 206–211.
- Neyman, J. and Scott, E. L. (1958). Statistical approach to problem of cosmology, *Journal of the Royal Statistical, Society Series B*, **20**, 1–43.
- Rodriguez-Iturbe, I., Cox, D. R., and Isham, V. (1987). Some models for rainfall based stochastic point process. In Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, 410, 269-288.
- Rodriguez-Iturbe, I., Cox, D. R., and Isham, V. (1988). A point process for rainfall: further development. In Proceedings of the Royal Society of London, 417, 283-298.
- Shin, J., Jung, C. S., and Heo, J. H. (2011). A study of new modified Neyman-Scott rectangular pulse model development using direct parameter estimation, Journal of Korea Water Resources Association, 44, 135–144
- Velghe, T., Troch, P. A., De Troch, F. P., and Van de Velde, J. (1994). Evaluation of cluster-based rectangular pulse point process models for rainfall, Water Resource Research, 30, 2847–2857.
- Yu, S. C., Kim, N. W., and Jeong, G. S. (2001). A point rainfall model and rainfall intensity-duration-frequency analysis, Journal of Korea Water Resources Association, 34, 577–586.

# Neyman-Scott Rectangular Pulse Model에 대한 통계적 추론

김남희 $^a$  · 김용구 $^{a,1}$ 

<sup>a</sup>경북대학교 통계학과

(2016년 5월 23일 접수, 2016년 6월 28일 수정, 2016년 7월 7일 채택)

#### 요 약

대표적인 강우생성 모형인 Neyman-Scott 구형펄스모형은 점과정(point process)을 이용하여 강우를 생성하는 모형으로 강우의 발생, 강우세포의 강우강도 그리고 지속시간의 분포로 표현된다. 특히 이 모형은 구형펄스모형(rectangular pulse model)에서 포함하지 않았던 강우사상의 군집특성을 반영하였다는 장점을 가지고 있다. NSRPM의 매개변수를 추정하는데 있어 moment를 이용한 여러가지 최적화 기법들이 연구되어 왔는데, 이러한 방법들은 목적함수를 추가하거나 조정하기 위해서는 복잡한 수식을 다시 계산하여야 하는 단점이 있으며, 전체적인 강우의 특성을 반영하기 어렵고 스케일에 따른 추정값의 변동도 크게 나타난다. 또한 moment를 이용한 추정값은 추정오차를 구할 수 없기 때문에 신뢰구간을 구할 수 없다는 단점이 있다. 이에 본 연구에서는 누적강수량에 대한 근사적인 우도함수(approximated likelihood function)를 소개하고 이를 통해 NSRPM의 매개변수를 추정하고자 한다. 또한 분석에 사용되는 누적강수량의 시간 스케일에 따른 추정치의 변동성도 함께 알아보고자 한다.

주요용어: 강우모형, 근사적 우도함수, Neyman-Scott 구형펄스모형, 매개함수 추정

본 논문은 제1 저자인 김남희의 석사학위 논문 중 일부를 정리한 것임.

<sup>&</sup>lt;sup>1</sup>교신저자: (41566) 대구광역시 북구 대학로 80, 경북대학교 통계학과. E-mail: kim.1252@knu.ac.kr