한국방재안전학회 논문집 2015년 6월
제8권 제1호, pp. 15-19

(사) 한국방재안전학회
KSDS Korean Society of Disaster & Security

비매개변수 핵밀도함수와 강우-유출모델의 합성곱(Convolution)을 이용한 수학적 해석

이태삼*

Convolution Interpretation of Nonparametric Kernel Density Estimate and Rainfall-Runoff Modeling

Taesam Lee*

접수일자: 2015년 5월 20일/심사완료일: 2015년 6월 12일/게재일자: 2015년 6월 30일

요 약 수문학에서 사용되는 강우-유출 모델의 경우 선형적인 시스템을 기반으로 유효강수량으로부터 시간적 지연을 통해서 유출량이 결정되는데 그 양은 강우량의 선형적인 비로 표현되어서 결국 합성곱을 통해 해석되게 된다. 또한 자료에 대한 확률론적 분석에 많이 이용되는 비매개변수 핵밀도함수의 경우, 핵(Kernel)의 의미자체가 합성곱에서 나온 것으로서 개개의 자료를 바탕으로 핵을 통해 매끄러운 확률밀도함수를 구하게 된다. 본 연구에서는 합성곱을 바탕으로 강우-유출 모델과 비매개변수 확률밀도함수를 해석하는 방법에 대해서 되짚어 보고 그 공통적인 특성과 다른 점을 수학적으로 나타내 줌으로써 사용되는 합성곱 함수의 유용성에 대해서 논하였다.

핵심용어 강우유출모델, 핵밀도함수, 합성곱, 비매개변수

ABSTRACT In rainfall-runoff models employed in hydrological applications, runoff amount is estimated through temporal delay of effective precipitation based on a linear system. Its amount is resulted from the linearized ratio by analyzing the convolution multiplier. Furthermore, in case of kernel density estimate (KDE) used in probabilistic analysis, the definition of the kernel comes from the convolution multiplier. Individual data values are smoothed through the kernel to derive KDE. In the current study, the roles of the convolution multiplier for KDE and rainfall-runoff models were revisited and their similarity and dissimilarity were investigated to discover the mathematical applicability of the convolution multiplier.

KEYWORDS rainfall-runoff model, kernel density estimate, convolution multiplier, nonparametric

1. 소 개

수문학에서는 자료를 분석하거나 수문시스템의 모델링을 위해 수학이나 통계학에서 사용되는 기법들을 적용하여 그 해석이나 분석을 시행하게 된다.

수문자료 분석에는 자료의 확률을 수학적 식으로 나타해주는 확률밀도 함수를 많이 사용하여 극치사상에 대한 확률적 의미를 표현해 주게 된다. 이를 위해 매개변수를 바탕으로 하는 Normal, Log-normal, Gamma, GEV, Weibull 분포형등이 많이 사용되어지고 있으나(Kottegoda and Rosso, 2008; Rao and Hamed, 2000) 최근에는 그 수학적 모형의 형상을 데이터를 통해 대변해 주는 비매개변수적 방법들도 널리 사용되어지게 되었으며 이를 비매개변수 핵밀도함수(Nonparametric kernel density estimate)라고 한다(Adamowski, 1996; Moon and Lall, 1994; Moon et al., 1993; Salas and Lee, 2010; Silverman, 1986). 이러한 비매개변수 핵밀도 함수는 현재 수문학에서 홍수분석이나 가뭄분석 자료의 특성을 나타내 주는 데 다양하게 사용되고 있다(Adamowski, 1996; Adamowski, 2000; Adamowski and Feluch, 1990; Lall et al., 1993; Lee et al., 2010; Lee, 2008;

*경상대학교 토목공학과 ERI, 부교수(E-mail: tae3lee@gnu.ac.kr)

Moon and Lall, 1994; Salas and Lee, 2010). 여기서 비매개변수 핵밀도 함수에서 사용되어지는 핵(‘Kernel’)을 왜 사용하였는지 그 의미가 모호할 수 있다. 따라서 이 논문에서는 핵밀도함수를 합성곱(Convolution)을 통하여 유도하고 핵(‘Kernel’)에 대한 의미를 정확히 짚어보고자 한다.

그리고, 수문시스템에서 사용되어지는 단위 유량도에 의한 강우-유출모델을 Bras(1990)는 수학적 함수인 합성곱(Convolution)을 통해 표현하였는데 핵밀도함수에서 사용되어지는 합성곱과 비교하였다. 두 다른 합성곱의 수문학적 응용의 유사성과 차이점을 보여줌으로써 한 차원 높은 수학적인 모델의 이해를 돕도록하였다.

2. 합성곱(Convolution)과 핵함수(Kernel function)의 정의

합성곱을 설명할 때 해당 시스템의 가장 큰 특징은 선형조건이다. 즉 가산성(additivity)과 비례성(proportionality)을 다음과 같이 만족한다는 것이다. 상수 m, n와 변수 x_1과 x_2에 대해

$$f(mx_1+nx_2)=f(mx_1)+f(nx_2)=mf(x_1)+nf(x_2) \tag{1}$$

로 표현되어진다. 식 (1)의 두 번째 식은 가산성을 보여주고 있고 세 번째 식은 비례성을 보여주고 있다.

합성곱은 어떤 특정 함수g에 대해 다른 함수 f가 옮겨감에 따라서 그 포함하는 양을 표현하는 적분을 의미하는 것으로써 다시 말하면 하나의 함수에 다른 함수를 혼합시키는 것이다. 이는 보통 점원(point source)에 대한 해석을 알고 있는 경우 유한한 크기의 source 를 핵함수(kernel function)와 복합시켜 그 해답을 찾게 하는데 사용된다. a와 b의 합성곱은 ⊗ 부호를 통해 다음과 같이 수학적으로 표현된다.

$$a(t)\otimes b(t)=\int_{-\infty}^{\infty} a(t)b(\tau-t)dt \tag{2}$$

여기서 $a(t)$는 대상함수(solution function)이고 $b(t)$는 핵함수(kernel function) 또는 가중함수(weighting function)라고도 불린다. 이 핵함수는 대상함수의 첨두를 완만하게 하고 가파른 경사를 감소시키는 역할을 하여 합성곱은 일반적으로 필터링(filtering)이라고 간주하기도 한다.

합성곱은 일반적으로 상미분 방정식을 풀기위한 라플라스변환에서 많이 사용되어지는데 그 이유는 식 (2)에서의 합성곱이 라플라스 변환시 a와 b 곱으로 나타내어져서 그 풀이가 쉽게 되기 때문이다. 즉,

$$\Im(a\otimes b)=\Im(a)\Im(b) \tag{3}$$

여기에서 $\Im(a)=\int_{-\infty}^{\infty} e^{-st}a(t)dt$ 로 나타내어 지는데 보통은 t가 시간의 함수가 되어 $t\geq 0$ 부분만 고려하게된다. 특정함수 $a(t)$에 대한 합성곱의 항등함수(identity)는 디랙델타함수로,

$$\delta\otimes a=\int_0^t \delta(t-\tau)a(\tau)dt=a(t) \tag{4}$$

표현되어지며 여기서 디랙델타함수는

$$\delta(t-\tau)=\begin{cases}\infty & \text{for } t=\tau\\ 0 & \text{for } t\neq\tau\end{cases} \tag{5}$$

로 정의되며 $\int_{-\infty}^{\infty}\delta(\tau)d\tau=1$ 의 특성도 지니고 있다. 이것을 라플라스 변환식으로 나타내면 그 특성이 명확히 다음과 같이 나타난다.

$$\Im(\delta\otimes a)=\Im(\delta)\Im(a)=\Im(a) \tag{6}$$

3. 비매개변수 핵밀도함수(Kernel Density Estimate, KDE)

임의의 변수 x에 대한 핵밀도함수(Kernel Density Estimate, KDE) 는 다음과 같이 정의 할수 있다.

$$\hat{f}(x)=\frac{1}{N}\sum_{i=1}^{N}K_h(x-x_i) \tag{7a}$$

또는,

$$=\frac{1}{Nh}\sum_{i=1}^{N}K\left(\frac{x-x_i}{h}\right) \tag{7b}$$

여기서 $x_i(i=1,.., N)$는 관측값을 의미하며 N은 데이터의 갯수를 의미한다. 그리고 h는 핵밀도함수의 부드러움을 결정하는 광역폭 인자이며 $K(.)$는 가중치함수 또는 핵함수라고 하며, $K_h(.)$는 스케일된 핵함수(scaled kernel)이다. 핵함수 $K(.)$의 종류에는 Rectangular, Gaussian, Epanechnikov, Cauchy 등이 사용되고 있다(Silverman, 1986).

본 연구에서는 이 핵밀도함수 (식 7a)에 대한 정의를 다음과 같이 합성곱 식으로 유도하였다.

$$\frac{1}{N}\sum_{i=1}^{N}K_h(x-x_i)=\sum_{y_j=a}^{b}K_h(x-Y_j)\frac{n_j}{N} \tag{8a}$$

여기에서 $\frac{n_j}{N}$는 $f(y_i)dy$로, 즉 $dF(y_i)$ 표현되어질 수 있다.

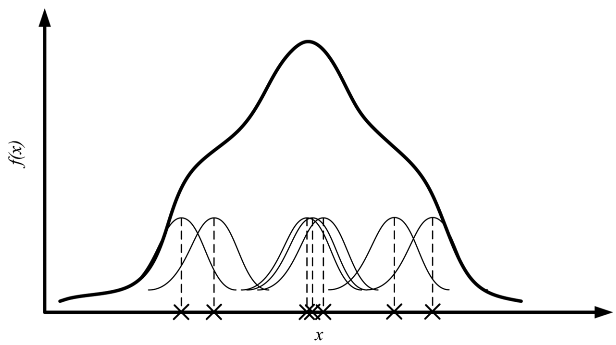

Fig. 1. Kernel density estimate using Gaussian kernel from 7 observed data

따라서

$$\sum_{y_j=a}^{b} K_h(x-Y_j)\frac{n_j}{N} \approx \int_a^b K_h(x-y)f(y)dy \tag{8b}$$

그리고 여기서 $a = -\infty$ and $b = \infty$가 되면 결국

$$\hat{f}(x) = \int_{-\infty}^{\infty} K_h(x-y)f(y)dy \tag{9}$$

식 (9)는 결국 합성곱 식이 되어 식 (3)과 같은 형태가 된다. 그림 1은 7개의 관측자료를 바탕으로 핵밀도 함수를 나타낸 것이며 Gaussian핵함수를 사용하였다. 그림 1에서 보여지는 바와 같이 핵밀도 함수는 관측치들에 놓여져 있는 융기(bumps)들의 합이며 핵함수는 이러한 융기들의 모양을 결정하게 됨을 알 수 있다. 관측자료에 대해 Gaussian 핵함수가 적용되어짐으로써 전체적으로 완만한 형태의 확률밀도 함수가 구해짐을 볼 수 있다. 그리고 핵함수의 부드러움의 정도를 나타내는 광역폭 h가 영으로 접근할 경우에는 관측치들의 디랙델타 함수(Dirac delta function) 첨두들의 합이 된다. 따라서 식 (9)는 식 (4)에서 나타나는 바와 같이 다음으로 정리될수 있다.

$$\hat{f}(x) = \int_{-\infty}^{\infty} K_h(x-y)f(y)dy = f(y)$$

여기서 $f(y)$는 결국 식 (8a)에 나타나는 바와 같이 $\frac{n_j}{N} = f(y_i)dy$에서의 $f(y_i)$와 같은 값이다. 더불어, h가 커지면 밀도함수의 세부적인 변화가 사라져서 완만한 형태의 밀도함수가 된다.

4. 강우-유출모델의 합성곱 해석

수문 시스템해석에서 가장 중요한 것 중의 하나인 강우-유출 모델이 있는데 이 중에서 집중형 모델(Lumped Model)은 일반적으로 단위유량도(Unit Hydrograph)를 가지고 강우-유출 사상을 선형적인 결합함을 그 기본으로 하고 있다(윤용남, 2007). 이러한 단위유량도를 기반으로 강우사상을 선형적으로 결합함으로써 유출량을 표현하는 것을 Bras(1990)는 수학적 함수인 합성곱(Convolution)을 통해 표현하였다.

수문 시스템에서 강우량과 유출의 관계성을 설명하는 것은 매우 중요한 연구중의 하나이다. 일반적으로 특정단위시간 동안 균일한 강도로 유역전반에 걸쳐 균등하게 내리는 단위유효우량(1 cm)으로 인한 직접유출량 수문곡선인 단위 유량도(Unit Hydrograph, UH)를 바탕으로 강우량과 유출 특성을 도출하게 된다. 즉 유역에 설정된 단위 유량도를 바탕으로 강우량과 유출의 선형시스템을 가정하여, 강우량에 대한 지체유출량을 축적시키는 방법으로 유량을 구하게 된다. 이는 강우강도의 시간적 분포를 $I(\tau)$라고 하면 유출량은 수학적으로 합성곱 적분으로 다음과 같이 표현할 수 있다(Bras, 1990).

$$Q(t) = \int_{-\infty}^{\infty} I(\tau)h(t-\tau)d\tau \tag{10}$$

여기에서 $Q(t)$는 시스템의 출력 값이고 $h(t)$는 순간 단위유량도(Instantaneous Unit Hydrograph, IUH) 라고도 하는데 단위우량이 대상유역에 순간적으로 적용되었을 때의 반응 값을 의미하며 디랙델타함수(dirac delta function, $\delta(t)$)를 가지고 다음과 같이 표현 되어진다.

$$h(t) = \int_{-\infty}^{\infty} \delta(t-\tau)h(\tau)d\tau \tag{11}$$

유출에 대한 식 (10)에서 수문학적 시스템에서 음수의 시간은 존재하지 않으며 미래 강우 입력에 대한 고려는 하지 않으므로 다음과 같이 식이 변형되게 된다.

$$Q(t) = \int_{0}^{t} I(\tau)h(t-\tau)d\tau \tag{12}$$

이러한 사항을 그림 2에서 보여주고 있다. 즉, 강우강도 $I(\tau)$에 대한 유출의 합성곱 관계를 순간단위 유량도를 통해 표현되어지는 것을 보여주고 있다.

5. 합성곱 해석을 이용한 유출모델과 핵밀도함수

단위 유량도의 합성곱 해석은 수문학적 시스템을 연구하는 과정에서 나온 산물이며 이에 반하여 핵밀도 함수는 통계학자들에 의해 개발되어 수문 데이터를 분석하는데 이용되는 방법이다. 이 다른 두 가지의 방법은 그 기반이 합성곱 해석에 기반을 두는 방법으로서 그림 1에서 보는 바와 같이 핵밀도 함수의 경우 다양한 핵함수가 임의의 점

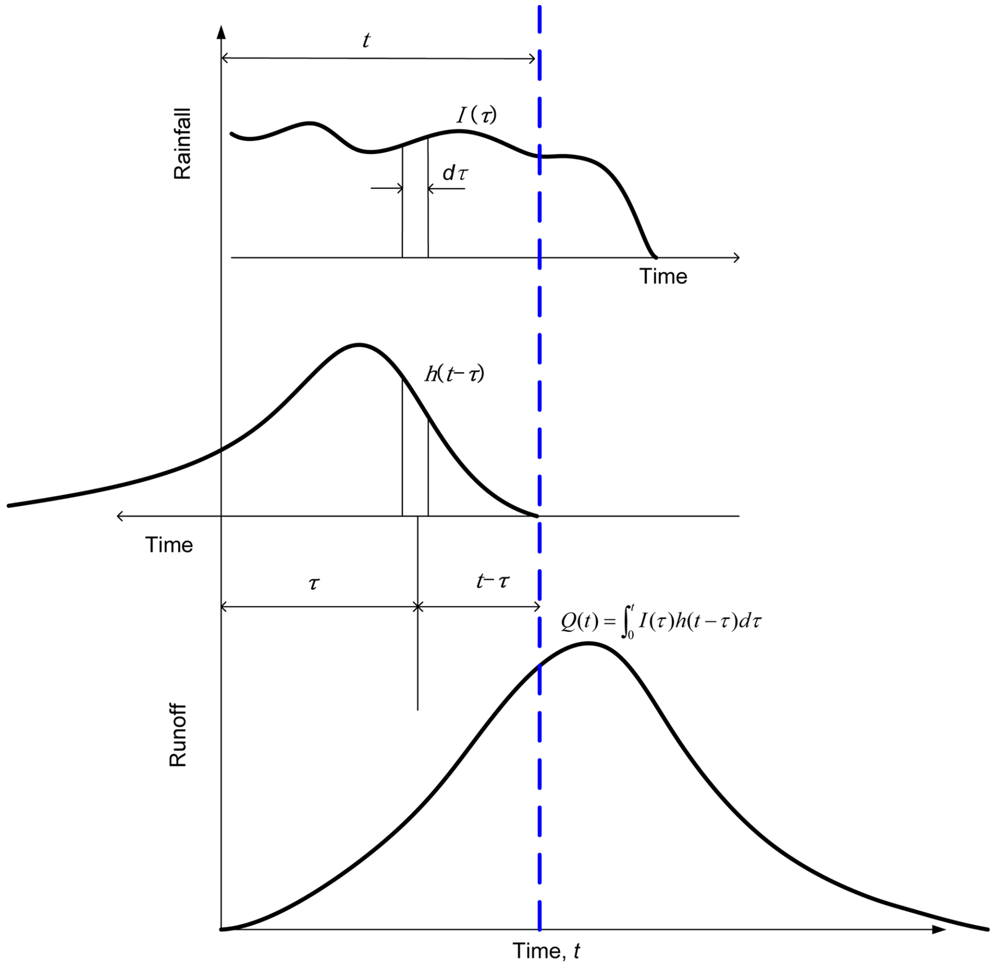

Fig. 2. Representation of convolution concept: rainfall (top panel), IUH (middle panel), runoff (bottom panel)

에 미치는 영향을 합산하여 임의의 점의 확률을 주변의 확률과 매끄럽게 이어져 나갈수 있도록 하는 역할을 하게 된다. 그리고 식 (9)에서 보는 바와 같이 각 핵함수가 어느 위치에 있든 임의의 점에 영향을 미치게 된다. 단위 유량도의 경우는 순간 단위유량도를 핵(Kernel)으로 놓고 강우량을 합성곱 해석하여 유출량을 해석하였다. 여기에서 음수의 시간이 존재하지 않고 미래강우에 대한 입력이 고려되지 않으므로 그 범위가 0에서 특정시간 t에 대한 고려만 이루어진다. 단위유량도와 핵밀도함수 두 가지의 합성곱 해석은 기본적인 방법론에서는 같다고 볼 수 있고 다른 부분은 임의의 점에서 고려하는 핵(Kernel)의 범위가 다르다는 것이다.

6. 결 론

본 연구에서는 일반적으로 유출해석에서 사용되는 유출모델과 핵밀도함수의 수학적 기초가 되는 합성곱 해석을 되짚어 보고 두 모델의 공통점과 다른 점을 살펴보았다.

감사의 글

본 연구는 국민안전처 자연재해저감기술개발사업단(자연피해예측및저감연구개발사업)의 지원으로 수행한 ‘기후변화 적응을 위한 연안도시지역별 복합원인의 홍수 취약성 평가기술 개발 및 대응 방안 연구’ [MPSS-자연-2015-77]과제의 성과입니다.

참고문헌

1. Adamowski, K. (1996) Nonparametric estimation of low-flow frequencies. Journal of Hydraulic Engineering-Asce, 122(1), 46-49.
2. Adamowski, K. (2000) Regional analysis of annual maximum and partial duration flood data by nonparametric and L-moment methods. Journal of Hydrology, 229(3-4), 219-231.
3. Adamowski, K. and Feluch, W. (1990) Nonparametric flood frquency analysis with historical information. Journal of

Hydraulic Engineering-Asce, 116(8), 1035-1047.
4. Bras, R.L. (1990) Hydrology: an introduction to hydrologic science. Addison-Wesley Publishing Company, Reading, MA, 643 pp.
5. Kottegoda, N.T. and Rosso, R. (2008) Applied Statistics for Civil and Environmental Engineers. Wiley-Blackwell, West Sussex, United Kingdom, 736 pp.
6. Lall, U., Moon, Y.I. and Bosworth, K. (1993) Kernel Flood Frequency Estimators - Bandwidth Selection and Kernel Choice. Water Resources Research, 29(4), 1003-1015.
7. Lee, T., Salas, J.D. and Prairie, J. (2010) An Enhanced Nonparametric Streamflow Disaggregation Model with Genetic Algorithm. Water Resources Research, 46, W08545.
8. Lee, T.S. (2008) Stochastic simulation of hydrologic data based on nonparametric approaches, Ph. D. Dissertation, Colorado State University, Fort Collins, CO., USA, 346 pp.
9. Moon, Y.I. and Lall, U. (1994) Kernel Quantile Function Estimator for Flood Frequency-Analysis. Water Resources Research, 30(11), 3095-3103.
10. Moon, Y.I., Lall, U. and Bosworth, K. (1993) A Comparison of Tail Probability Estimators for Flood Frequency-Analysis. Journal of Hydrology, 151(2-4), 343-363.
11. Rao, A.R. and Hamed, K.H. (2000) Flood Frequency Analysis. CRC Press, New York, 350 pp.
12. Salas, J.D. and Lee, T. (2010) Nonparametric Simulation of Single-Site Seasonal Streamflows. Journal of Hydrologic Engineering, 15(4), 284-296.
13. Silverman, B.W. (1986) Density Estimation for Statistics and Data Analysis : Monographs on Statistics and Applied Probability. Chapman and Hall, London, 175 pp.