학술기사

비매개변수적 Kernel 가중함수의 수문학적 응용

문 영 일 (서울시립대학교 토목공학과 조교수)

요 약

전통적인 매개변수적 목적함수 추정방법은 관측자료의 모든 영역에 걸쳐 선형 또는 지수함수 형태의 가정을 기본으로 매개변수를 추정하는 반면 비매개변수적 Kernel 가중함수를 이용한 방법은 목적함수의 형태에 대한 가정이 필요 없이 관심 있는 임의의 추정지점에서 이웃하는 자료를 이용하여 목적함수를 국지적으로 근사하는 방법이다. 추계학적 수문학의 전형적인 문제인 "목적함수의 가정"에 의해 발생되는 문제를 줄이려는 노력의 일환으로 비매개변수적 Kernel 가중함수를 이용하는 방법이 연구되었고, 본 지면에서는 Kernel 가중함수를 이용한 비매개변수적 확률밀도함수의 기본이론과 빈도해석, 회귀모형 및 비동질성 천이확률 등의 수문학적 응용에 대하여 살펴보았다.

1. 서론

비매개변수적 Kernel 가중함수를 이용한 수문변수의 추정은 관심 있는 임의의 추정지점에서 이웃하는 자료를 이용하여 목적함수를 국지적으로 근사하는 방법이다. 전통적인 매개변수적 방법은 관측자료의 모든 영역에 걸쳐 선형 또는 지수함수 형태의 가정을 통하여 목적함수에 도달하는 반면 비매개변수적 방법은 목적함수의 형태에 대한 가정이 필요 없이 이웃하는 자료를 이용하여 목적함수를 추정하는 방법이다. 전통적으로 매개변수적 가정들, 즉 예를 들면 한 지점에서 홍수량은 Log-Pearson Ⅲ 분포를 따르고, 수리적 전도는 대수정규분포를 따르고, 연 하천유량은 대수정규나 감마분포를 따르고, 일 강우량은 지수분포를 따르며, 수문학적 공간 자료의 분산도(variogram)는 지수의 법칙을 따른다는 가정 등은 통계적으로 수문변수 추정치를 구하는데 사용되었다. 이런 추계학적 수문학의 전형적인 문제인 "목적함수의 가정"에 의해 발생되는 문제를 줄이려는 노력의 일환으로 비매개변수적 Kernel 가중함수를 이용하는 방법이 연구되었다.

특정 모형에 대한 가정(예를 들면 홍수량는 LP3 분포)을 신봉하는 것과, 매개변수적 추정 과정은 어떻게 하면 자료들에 내재한 구조적 연관성을 잘 표현하는 함수를 가장 잘 추정하는가는 수문통계학자들이 직면하는 기본적 문제들이다. 그러나 대체로 수문학자들은 목적함수의 특정 형태에 대한 물리적·이론적 원리 등은 부족한 편이다(?). 전통적인 연구에서는 미리 정의된 일련의 곡선들 중에서 주어진 자료 사이에 적합한 것을 선택하는 일이 주로 수행되었던 것이 사실이다. 연구자가 관측자료 내부의 구조를 식별하게 된 후 목적함수의 후보곡선들 모두가 적합하지 않는 것을 알았다면 어떠한 선택을 해야 하는지? 매개변수 추정방법에 의한 변화 뿐만 아니라 모형선택의 불확실성까지 감안하기 위해서는 추정된 신뢰폭을 어디까지 어떻게 해석할 것인지? 유한한 양의 자료가 주어졌을 때 모형선택의 문제를 감안하지 않고서 자료에 대한 유용한 해석을 내리는 것이 가능한 것인지? 이러한 질문들이 목적함수에 대한 가정이 필요 없는 비매개변수적 추정방법에 대한 연구를 부채질해 왔다.

지금까지 많은 문헌에서 여러 종류의 비매개변수적 Kernel 가중함수는 소개되었다. 이것들은 추정 효율성, 계산의 선행 조건들, 적용성, 그리고 수학적 형태에 따라 각기 다르게 표현되나 기본적으로 추정값으로 이웃하는 자료에 의해서만 영향을 받는 국지적 목적함수를 가진다는 공통점이 있다. 따라서 비매개변수적 방법은 Kernel 함수를 이용한 자료의 가중이동평균으로 해석될 수 있고 국지화는 추정의 지점으로부터 멀어질수록 가중치를 적게 줌으로써 이루어진다. 지난 10년 이상 비매개변수적 Kernel 가중함수를 이용한 대표적인 문헌들을 살펴보면 Devroye와 Györfi(1985), Silverman(1986), Müller(1988), Eubank(1988), Härdle(1989, 1990), Scott(1992), Green과 Silverman(1994) 그리고 Wand과 Jones(1995)가 있다. 수문학적 응용분야을 살펴보면 빈도해석(Guo, 1991; Moon 등, 1993; Moon과 Lall, 1994; Adamowski, 1996), 시계열분석(Cleveland와 Devlin, 1988; Bradley와 Potter, 1992; Cleveland, 1993; Tarboton, 1994; Rajagopalan 등, 1996), 공간분석(Owosina 등, 1992; Lall과 Ali, 1992; Yakowitz, 1993; Lall 등, 1994)에 이르기까지 연구가 수행중이며 본 연구에서는 지면상 비매개변수적 확률밀도함수의 기본이론, 빈도해석, 회귀분석, 및 비동질성 천이확률 산정방법 등에 관하여 개략적으로 기술하였다.

2. 비매개변수적 확률밀도함수의 기본이론

확률변수(x)의 분포상태를 쉽게 알기 위하여 확률분포를 그래프로 나타내는 것이 일반적이다. 이 때 x에 관한 확률분포는 확률밀도함수(probability density function) f(x)에 의해 표시되며, f(x)는 자료의 무작위의 성향을 표현해 준다. 아마도 가장 오래되고 가장 널리 사용되는 확률밀도함수 추정법은 막대그래프(Histogram)일 것이다. 막대그래프는 시작점(x_0)이 주어진 상태에서, 막대그래프의 구간간격은 양과 음의 정수 m과 양의정수 h를 사용하여 구간 $[x_0+mh, x_0+(m+1)h]$로 정의된다. 이때 막대그래프의 밀도함수는 다음 식(1)과 같이 표현 할 수 있다.

$$f(x) = \frac{1}{nh}\ (\text{x와 같은 구간에 있는 자료의 수}) \quad (1)$$

막대그래프는 가장 오래된 확률밀도함수 추정법이고, 자료의 시각적인 그래픽 표시를 위한 고전적인 비매개변수적 확률밀도함수 추정법이다. 막대그래프는 이해하기가 쉽고 수작업에 의해서도 쉽게 계산 될 수 있는 반면에, 계급구간이 변화는 점에서 불연속적이고 구간간격과 그래프의 시작점의 선택에 따라 확률밀도함수의 모양이 달라진다. 그러므로, 막대그래프는 IMSE(Integrated Mean Squared Error)의 관점에서는 상당히 비능률적이다. 확률변수가 주어졌을 때, IMSE는 막대그래프의 구간간격이 최적으로 선택되어 졌을 때 $n^{-2/3}$에 비례하게 점근선적으로 적어지게 된다고 알려지고 있다(Devroye and Gyorfi, 1985). 그러나, 막대그래프를 만들 때의 발생되는 문제는 아무리 이상적인 동일한 구간간격을 사용할지라도, 다른 시작점의 위치 선택에 따라 같은 자료에서도 매우 다른 확률밀도함수의 결과를 초래한다는 것이다. 이런 고전적인 막대그래프의 저지르기 쉬운 잘못은 많은 연구자들에게 연속적이며 변화하는 히스토그램을 연구하는 동기를 주었다. 따라서 자료가 발생된 각각의 위치에 막대그래프의 박스중앙을 위치하도록 하여, 구간을 이동시킬 수 있는 이동 히스토그램을 발달시켰다.

일반적인 확률밀도 함수의 정의로부터 임의의 변수 x는 다음 식(2)와 같이 확률밀도함수 f(x)로 표현할 수 있다.

$$f(x) = \lim_{h\to 0} \frac{1}{2h} P(x-h < X < x+h) \quad (2)$$

주어진 h에 대하여 $P(x-h < x < x+h)$의 값은 구간 $[x-h, x+h]$ 사이의 있는 자료의 비율에 의해 계산되어질 수 있다. 그래서 본래 추정 식은 작은 숫자 h의 선택으로 다음과 같이 얻을 수 있다.

$$\hat{f}(x) = \frac{1}{2hn}$$

(구간 $[x-h, x+h]$ 사이에 있는 자료의 수) (3)

식(3)을 보다 함수적으로 표현하기 위해서 가장 간단

한 핵함수(kernel function) K(u)를 다음 식(4)와 같이 정의해 보자.

$$K(u) = \frac{1}{2} \text{ if } \ |u| < 1 \qquad (4)$$

이 때 식(4)의 내용을 식(3)에 적용시키면 아래의 식(5)와 같이 표현됨을 알 수 있다.

$$f(x) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{n}K\left(\frac{x-X_i}{h}\right) \qquad (5)$$

여기에서 x는 임의의 값이며, X_1, X_2, …, X_n은 독립적으로 동일하게 분포된 실관측치, K(.)는 핵함수, h는 n이 무한대로 갈 때 0으로 접근하는 값을 갖는 양의 광역폭(bandwidth)이다.

식(5)는 식(1)과 같은 표현으로, 각 관측치에 폭 2h, 높이 $(2nh)^{-1}$의 막대를 각 관측치에 놓고, 그 들의 합계에 의해 그림 1.과 같이 추정 식이 만들어지는 것을 보여주고 있다.

여기서 관측자료가 에너지원에 비유된다면, 자료는 발생 위치에서 최고 값을 갖고 발생위치로부터 멀어질수록 급속히 감소한다고 할 수 있기 때문에 핵함수를 하나의 가중함수로 생각할 수 있다. 따라서 식(5)는 이동할 수 있는 히스토그램을 나타내며 모든 실수 x에 대하여 정의할 수 있다.

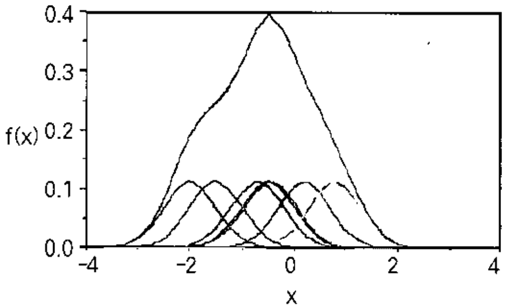
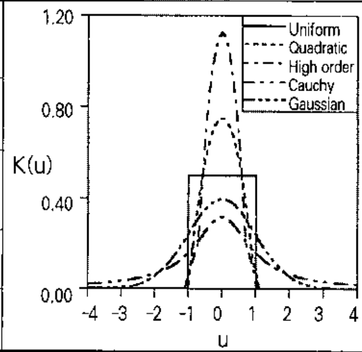


그림 1. 각각의 Gaussian 핵함수에 의한 핵밀도함수 추정

자주 사용되는 핵 함수 K(u)의 종류는 표 1.과 같으며 어느 함수를 사용해도 기본적 개념은 같다. 핵함수는 대개 t=0에서 최대치를 가지며, 연속적이며, 대칭인 방정식의 형태를 가진다. 즉, 핵함수의 면적은 1이고($\int K(u)du=1$) 기대값 0($\int uK(u)=1$)과 유사한 분산값($\int u^2K(u)du=constant$)을 갖는다. 그러나, 때로는 이 특성들을 만족하지 않는 핵함수가 사용되어질 수도 있다. 핵함수의 선택은 Epanchnikov (1969)에 의해 고려되었는데 그는 평균자승오차(mean square error, MSE)에 의하면 포물선 형태의 핵함수가 거의 최적의 결과를 나타낸다는 것을 보여주었다. 그러나, 많은 연구에서 핵함수의 선택은 실제 생각되어진 만큼 중요한 문제가 아니라고 주장되었다. 선택된 함수의 효율성(선택된 핵함수의 MSE/최적의 핵함수의 MSE)은 주어진 핵함수의 특성을 만족하면 거의 1에 가깝다. 그러나, 각각 다른 핵함수들은 사용목적에 따라 점검되어야 한다. 예를 들어 밀도 함수의 연속성과 미분가능이 필요한 경우 제한된 구간을 갖는 핵함수보다는 무한한 구간을 갖는 핵함수를 선택하는 것이 좋을 것이다.

그러나 광역폭 h의 값은 핵함수 선택보다는 상대적으로 중요하다. 너무 큰 h는 큰 편차(bias)와 너무 완만(oversmooth)한 밀도함수의 추정과 정보의 손실을 가져온다. 반면에, 너무 작은 h는 큰 분산(variance)과 들쭉날쑥한 거친(rough) 추정치를 나타낸다. 따라서 객관적인 광역폭 선택을 위하여 해석학적으로 h를 구하는 방법이 연구되었으며 그 중에서 대표적인 것이 IMSE, Maximum Likelihood,

표 1. Kernel 함수

Kernel	Formular		Shape
Uniform	$K(u) = 0.5$	$\|u\| \le 1$	
	$= 0$	$\|u\| > 1$	
Quadratic	$K(u) = 0.75(1-u^2)$	$\|u\| \le 1$	
	$= 0$	$\|u\| > 1$	
High order	$K(u) = 0.375(3-5u^2)$	$\|u\| \le 1$	
	$= 0$	$\|u\| > 1$	
Cauchy	$K(u) = 1/\pi\,(1+u^2)$		
Gaussian	$K(u) = \{\exp(-u^2/2)\}/\sqrt{2\pi}$		

Least Squared, Biased Cross-validation, Plug-in method 등이 소개되었고(Wand와 Jones, 1995) 사용목적에 따라 방법을 선택할 수 있다.

3. 수문학적 응용

3.1 빈도분석

수문학자들은 빈도가 낮은 홍수나 갈수량의 상대빈도를 추정하는 데 자주 초점을 맞추게 된다. 관측자료 $x_i (i=1,2,....,n)$가 독립적이라고 가정하고 동일하게 분포되어지는 단변량 자료가 주어지면 임의의 x_T의 초과확률을 결정하는 것이 주된 목적이 된다. 하나의 분포함수의 가정을 통하여 빈도해석을 하는 전통적인 매개변수적 방법과 다르게 관측자료의 가중이동평균에 의해 분석을 하는 핵밀도함수(kernel density function)에 의한 빈도해석 방법에는 고정핵밀도함수(fixed kernel density function) 방법과 변동 핵밀도함수(variable kernel density function) 방법이 있다. 식(5)에서 h가 변화 없이 일정하면 고정핵밀도함수(fixed kernel density function)라 하고, h가 자료의 상황에 따라 변화하면 이를 변동 핵밀도함수(variable kernel density function)라 한다. 고정 핵밀도 추정법은 군데군데 길게 늘어진 자료분포에 적용될 경우 결점을 갖는다. 왜냐하면, 일정한 bandwidth, h를 전체자료에 적용될 경우 추정치의 꼬리(tail) 부분에서 자료가 드물게 분포할 경우에 추정치의 값이 불연속 또는 거친 밀도함수의 모양을 갖기 때문이다. 따라서 고정 핵밀도함수에 의한 빈도해석보다 변동 핵밀도함수에 의한 빈도해석의 결과가 우수하다(Moon 등, 1993; Moon과 Lall, 1994). Breiman 등(1977)은 고정 핵밀도함수 추정법의 특성에다 자료의 지역적인 밀도를 고려하는 k^{th} nearest neighbor방법을 결합한 변동 핵밀도함수 추정법을 제안하여 자료가 발생된 위치에 놓여지는 핵함수의 폭이 자료의 밀도에 따라 변하게 다음과 같이 정의하였다.

$$f(x) = \frac{1}{nh} \int \frac{1}{d_{j,k}} K\left(\frac{x - X_j}{hd_{j,k}}\right) dx \qquad (6)$$

여기서, k를 양의 정수로 놓고, $d_{j,k}$는 한 개의 자료 x_j에서 그 나머지 자료 (n-1)개 중에서 k번째로 가까운 지점에 있는 자료까지의 거리다. 따라서 자료의 분포가 적은 낮은 밀도지역에서 $d_{j,k}$는 커지고 변동 핵함수의 모양은 넓게 퍼지게 된다. 자료의 분포가 많은 높은 밀도지역에서는 그 반대현상이 일어나, 변동 핵함수는 좁게 밀착된 형태를 가질 것이다.

위와 같이 변동 핵밀도함수 추정법을 빈도해석에 사용시 전통적인 매개변수적 해석방법의 어려운 점인 분포함수의 객관적 선택과 여러가지 원인으로 인한 복합분포(mixed distribution)의 경우 해석의 어려움을 해소할 수 있다. 그러나 변동 핵밀도함수가 빈도해석시 발생되는 문제점 모두를 해결해 주는 것은 아니다. 비매개변수적 빈도해석도 자료 크기보다 매우 큰 재현기간에서는 추정값이 작은 경향을 보이는 문제가 있다. 즉, 관측자료의 범위를 상당히 벗어난 위치에서는 외삽의 기본성향을 이끄는 선행값의 사용에 제한을 받기 때문이다. 따라서 이에 대한 보완방법으로 꼬리 부분이 두꺼운 핵함수의 개발과 최적의 광역폭의 선택, 자료의 대수변환, 또는 준매개변수적 방법(Semi-Parametric Method) 등 여러가지 방법이 제안되고 있다. 이러한 배경하에서 비매개변수적 핵함수 추정방법에 의해 가뭄/홍수량의 빈도분석이 성공적으로 소개되어왔다는 것은 그리 놀라운 것은 아니다.

3.2 비매개변수적 회귀모형

비매개변수적 회귀분석(non-parametric regression)은 매개변수를 통하지 않고 자료로부터 알고자하는 지점의 추정값을 추정한다. 이러한 비매개변수적 회귀분석은 주어진 자료의 특성으로부터 잡음(noise)을 제거 또는 감소시킬 수 있으며, 이로 인하여 자료의 해석에 있어 보다 원자료에 근접하는 회귀모형을 구할 수 있다는 장점을 지닌다. 비매개변수적 회귀분석은 기본 회귀분석이 해석하기 어려운 자연계의 이질적이고, 다중변수, 시간과 공간적인 변수를 지니게 되는 자료들에 대한 유용한 해석방법이라고 할 수 있다.

최근에 많이 사용되는 비매개변수적 회귀방법(non-parametric regression estimator)은 Kernel estimators, Nearest neighbor methods, Smoothing spline 등이 있으며, 이와 같은 회귀모형은 기존의 회귀분석 방법에 비하여 보다 쉽게 적정한 회귀모형을 제시해 준다. Smoothing spline의 경우 관측치에 가까운 추정값를 구할 수 있지만, 자료의 경향성을 나타내는 것이 부족한 것으로 알려져 있다. 이에 비하여 Kernel regression의 경우는 자료의 분포 및 경향성을 근접하게 나타내는 장점이 있다고 할 수 있다. 따라서 기존의 회귀분석에서 해석하기 어려운 자연계의 이질적이고, 다중의 변수를 지니며, 시간과 공간적인 변수를 지니게 되는 자료들에 대한 유용한 해석방법이다.

Kernel regression estimator의 가장 기본적인 형태를 나타내면 다음과 같다.

$$f(x) = \frac{\sum_{j=1}^{n} K\left(\frac{(x - x_j)}{h}\right) y_j}{\sum_{j=1}^{n} K\left(\frac{(x - x_j)}{h}\right)} \tag{7}$$

여기서, h는 광역폭(bandwidth), n는 자료개수, (x_i, y_i)는 주어진 관측자료이며, x는 추정하고자하는 값, $K(\cdot)$ 는 Kernel 함수이다.

식(7)를 연속적인 값으로 나타내기 위해 발전시키면 다음과 같다.

$$\hat{f}(x) = \frac{1}{n} \sum_{j=1}^{n} \int_{s_{j-1}}^{s_j} K\left(\frac{(X - s)}{h} ds\right) Y_j \tag{8}$$

여기서 $s_j = \frac{(X_{j+1} + X_j)}{2}$ 이다.

식(8)은 식(7)와 같은 개념이며, 식(7)에 비하여 상대적으로 bias가 작아지는 경향이 있다. 특히 이 식은 공간적으로 비균일한 자료에 있어서 가장 적합한 추정값을 제시해 주는 것으로 알려져 있다.

Kernel 함수는 관측자료에 가중치를 부여하며, 가중치는 Kernel 함수의 모양에 의해 결정되고, 그 폭은 광역폭에 의해 결정된다.

3.3 비동질성 천이확률 추정

일강우자료를 모의발생 할 경우 필요한 천이확률에 대해 살펴보자. 지금까지 많은 경우 간편함의 이유로 동질성 마코프 연쇄모형을 사용하여 천이확률이 천이발생일(1~365일)에 상관없이 과거의 관측된 자료로부터 크기에 따른 상태벡터를 구성하여 상태별 천이 빈도수에 따라 월별 동질성 천이확률을 사용하였다. 그러나, 천이확률은 날짜에 따라 변하기 때문에 연중에 다르게 나타날 수 있다는 것에 문제가 있다. 따라서 여기서는 관측 강우자료로부터 핵확률밀도함수 개념를 이용하여 비동질성 천이확률 구성방법에 대하여 살펴보았다. 먼저 아래에 강우발생과정을 그림 2.와 같이 나타내었다.

일 지수(t)는 1년 중의 1일부터 365일(윤년은 366일)을 나타낸다. 여기서, 일 강우자료로부터 네 종류의 지수로 나눌 수 있다. ①일 지수 $t_{w_1}, t_{w_2}, \ldots, t_{w_{nw}}$는 nw개의 강우일을 나타내고, ②일 지수 $t_{d_1}, t_{d_2}, \ldots, t_{d_{nw}}$는 nd개의 무강우일을 나타내며, ③일 지수 $t_{wd_1}, t_{wd_2}, \ldots, t_{wd_{nwd}}$는 강우일에서 무강우일로 바뀌는 nwd개의 천이 발생일을 나타낸다(즉, t_{wd_1}, t_{wd_2} 등은 강우일을, $t_{wd_1}+1, t_{wd_2}+1$ 등 무강우일을 나타낸다). ④일 지수 $t_{dw_1}, t_{dw_2}, \ldots, t_{dw_{nwd}}$는 무강우일에서 강우일로 바뀌는 ndw개의 천이발생일을 나타낸다(즉, t_{dw_1}, t_{dw_2} 등은 무강우일을, $t_{dw_1}+1, t_{dw_2}+1$ 등은 강우일을 나타낸다). 여기서,

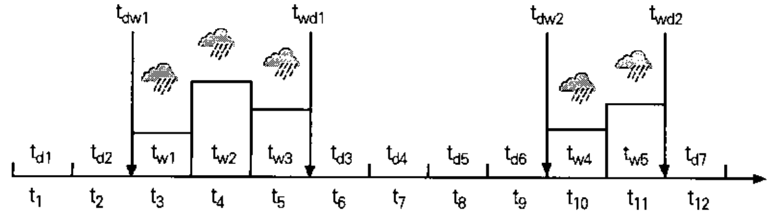


그림 2. 임의의 해에 대한 강수일

역력 t일의 천이확률 $P_{wd}(t)$와 $P_{dw}(t)$은 비매개변수적 핵밀도함수 추정치를 이용하여 자료로부터 구한다. 동질성 마코프 모형에서는 일반적으로 천이확률을 단순하게 월별 과거의 자료에서 천이의 수와 강우 또는 무강우의 수에 대한 비율로써 추정하여 계절변동성을 나타내었으나, 여기서는 핵함수를 이용하여 관심 있는 사상(강우일, 무강우일, 또는 천이일)이 발생된 날짜에 핵함수의 중앙이 위치하도록 하여 관심의 대상이 되는 날의 부근에 일어난 사상(즉, 강우나 무강우일 또는 천이상태)은 가중치를 더 주고, 멀리 떨어져 일어난 사상은 가중치를 덜 줌으로써 천이발생일에 따른 천이확률의 계절 변동성을 나타낼 수 있다. 이에 대한 결과로 천이확률에 대해 핵함수 추정식은 아래와 같다. 나머지 두 개의 천이확률($P_{ww}(t)$와 $P_{dd}(t)$)은 $P_{wd}(t)+P_{ww}(t)=1$과 $P_{dw}(t)+P_{dd}(t)=1$로부터 직접 구할 수 있다.

$$\hat{P}_{wd}(t) = \frac{\sum_{i=1}^{nwd} K\left(\frac{t - t_{wdi}}{h_{wd}}\right)}{\sum_{i=1}^{nw} K\left(\frac{t - t_{wi}}{h_{wd}}\right)} \tag{9}$$

$$\hat{P}_{dw}(t) = \frac{\sum_{i=1}^{ndw} K\left(\frac{t - t_{dwi}}{h_{dw}}\right)}{\sum_{i=1}^{nd} K\left(\frac{t - t_{di}}{h_{dw}}\right)} \tag{10}$$

$$P_{ww}(t) = 1 - P_{wd}(t) \tag{11}$$

$$P_{dd}(t) = 1 - P_{dw}(t) \tag{12}$$

이 때, n_{wd}는 과거 자료에서 강우일로부터 무강우일로 바뀌는 천이의 수, n_{dw}는 과거 자료로부터 무강우일로부터 강우일로 바뀌는 천이의 수, n_d는 과거의 자료에서 무강우일의 수, n_w는 과거의 자료에서 강우일의 수이다.

4. 맺는말

목적함수의 형태에 대한 가정이 필요 없이 관심 있는 임의의 추정지점에서 이웃하는 자료를 이용하여 목적함수를 국지적으로 근사하는 비매개변수적 Kernel 가중함수의 수문학적 응용 방법에 대하여 서술하였다. 관측된 자료가 기존의 정의된 분포형으로 표현되지 못할 경우(특히 관측자료가 bimodal 형태의 밀도분포를 가지고 있을 시) 및 자료의 형태에 따라 특정 목적함수 선정에 어려움이 있는 경우에는 전통적인 매개변수적 방법의 접근에는 한계가 있다. 따라서 추계학적 수문학분야에서 목적함수의 가정이 필요 없는 비매개변수적 핵밀도함수 개념이 유용하게 사용될 수 있을 것이며 앞으로 수문자료 해석에 많은 활용의 계기가 되기를 바란다.

〈참고문헌〉

Adamowski, K.,(1996), Nonparametric Estimation of Low-Flow Frequencies. Journal of Hydraulic Engineering, Vol.122(1), pp. 46-49.

Bradley, A.A. and K.W. Potter,(1992), Flood Frequency Analysis of Simulated flows. Water Resources Research, Vol.28(9), pp. 2375-2386.

Breiman, L., W. Meisel, and E. Purcell,(1977), Variable kernel estimates of multivariate densities. Technometrics, Vol.19(2), pp. 135-144.

Cleveland, W.S. and S.J. Devlin,(1988), Locally Weighted Regression: An Approach to Regression Analysis by Local Fitting. J. Amer. Stat. Assoc., Vol.74(368), pp. 829-836.

Cleveland, W.S.(1993), Visualizing Data. Hobart Press, Summit, NJ.

Devroye and Gyorfi,(1985), Nonparametric density estimation: The L1 view. John Wiley, New York.

Epanechnikov, V.A.(1969), Nonparametric estimation of a multidimensional probability density. Theory Probability and Applications Vol.14, pp. 153-158.

Eubank, R.L.,(1988), Spline Smoothing and

Nonparametrci Regression. Marcel Dekker Inc, New York.
Green, P.J. and B.W. Silverman,(1994), Nonparametric Regression and Generalized Linear Models. Chapman and Hall, New York.
Guo, S. L.(1991), Nonparametric variable kernel estimation with historical floods and paleoflood information. Water Res. Research Vol.27(1), pp. 91-98.
H rdle, W.,(1989), Applied Nonparametric Regression. Cambridge University Press, Cambridge.
H rdle, W.,(1990), Smoothing techniques with implementation in S. Springer Verlag, New York.
Lall, U. and A.I. Ali,(1992), Nonparametric Stratigraphic Interpretation from Drill Log Data, Utah Water Res. Lab., Utah State Univ.
Lall, U., K. Bosworth, A. Owosina, and Y.-I. Moon, (1994), Local Polynomial Estimation of Spatial Surfaces, Utah Water Res. Lab., Utah State Univ.
Moon, Young-Il, U. Lall, and K. Bosworth,(1993), A comparison of tail probability estimators. Journal of Hydrology 151, pp. 343-363.
Moon, Young-Il, and U. Lall,(1994), Kernel Quantile Function Estimator for Flood Frequency Analysis. Water Resources Research, Vol. 30, No. 11, pp. 3095-3103.
M ller, H.-G.,(1988), Nonparametric Regression Analysis of Longitudinal Data. Springer Verlag, Derlin.
Owosina A., U. Lall, T. Sangoyomu, K. Bosworth, and Y.-I. Moon,(1992), Methods for Assessing the Space andTime Variability of Groundwater Data, Utah Water Res. Lab., Utah State Univ.
Rajagopalan, B., U. Lall, and D. Tarboton,(1996), Nonhomogeneous markov model for daily precipitation. Journal of Hydrologic Engineering Vol. 1, No. 1, pp. 33-40.
Scott, D.W.,(1992), Multivariate Density Estimation. John Wiley and sons, New York.
Silverman, B.W.(1986), Density estimation for statistics and data analysis. Chapman and Hall, New York.
Tarboton, D.G.,(1994), The Source Hydrology of Severe Sustained Drought in the Southwestern U.S., J. Hydrology Vol.161, pp.31-69.
Yakowitz, S.,(1993), Nearest Neighbor Regression Estimation for Null-Recurrent Markov Time Series. Stochastic Processes and Their Applications Vol.48, pp. 311-318.
Wand, M.P. and M.C. Jones,(1995), Kernel Smoothing. Chapman and Hall, New York.