

# 통신 빅데이터를 활용한 국가하천 친수지구의 이용등급, 상세유형화 및 친수거점지구 지정 방법 개발\*

A Study on Classification and Characterization of Water-Friendly Space for  
the Smart River Space Management Using the Mobile Big Data

이종소 Lee Jongso\*\*, 이상은 Lee Sangeun\*\*\*

## Abstract

For national and local policies related to urban development and regeneration, welfare, culture and tourism, it is very important for river managers to designate and operate water-friendly river spaces in accordance with residents' demands. Using mobile big data verified by the previous study (Lee, Lee and Choi 2019), this study aims to propose scientific ways to characterize water-friendly river spaces. First, it suggests clear criteria regarding how frequently a river space is visited by residents: river spaces are classified to five grades over the log-normal distribution in the number of visits per area. Second, it infers residents' purposes to visit a river space: the purposes are characterized to five groups, which largely depends on whether a river space is used as a place for neighborhood-convenience facilities, or as one for culture and tourist facilities for regional revitalization. Third, this study suggests a step-wise method for determining water-friendly hub spaces deliberately using all results. This study concludes that river managers should apply more scientific ways to designate and operate water-friendly river spaces while monitoring residents' visits closely with mobile big data.

Keywords: Water-Friendly River Space, River Space Characterization, Mobile Big Data, River Management

## I. 서론

2007년 「하천법」 개정 이후부터 하천기본계획을 수립할 때 주민들이 하천공간을 적절하게 활용할 수 있도록 친수지구를 지정해 오고 있다. 관리청이 친수지

구를 지정하는 행위는 지역의 건전한 하천이용을 지원하기 위한 행정적인 근거를 확보한다는 의미를 지닌다. 우선, 하천관리청은 하천공사를 통해 광장, 산책로, 자전거도로, 경관시설 등 하천이용을 위한 기반을 확충하며, 이러한 친수시설의 기능을 유지하는 데

\* 본 연구는 2018년에서 2020년까지 국토연구원에서 진행 중인 국토교통부의 '통신 빅데이터를 활용한 국가하천관리 효율성 제고방안 연구'로부터 연구비와 기초자료를 지원받았으며, 본 논문의 작성 및 투고에 대해서는 연구의 감독기관과 사전에 협의되었음.

\*\* 국토연구원 책임연구원(제1저자) | Assistant Research Fellow, Korea Research Institute for Human Settlements | Primary Author | [jslee@krihs.re.kr](mailto:jslee@krihs.re.kr)

\*\*\* 국토연구원 수자원·하천연구센터장(교신저자) | Director of Water Resources and River Research Center, Korea Research Institute for Human Settlements | Corresponding Author | [selee@krihs.re.kr](mailto:selee@krihs.re.kr)

필요한 유지·보수비의 일부를 관할 지자체에게 지원하고 있다. 친수지구 관리에는 주민들의 하천이용이 중요한 요소라 할 수 있으므로 관리청은 지역주민들이 요구하는 이용객 편의시설, 레저·문화시설 등의 설치 규제도 하천법령과 점용허가 세부기준의 범위 내에서 완화하게 된다(국토교통부 2018).

문제는 하천기본계획 수립 시 지역주민의 공간활용 방식에 대한 객관적이고 면밀한 수요조사 방법이 마련되어 있지 않다는 점이다. 이러한 가운데 2014년부터는 근린친수와 친수거점으로 테마를 부여하는 등으로 특성화를 시도하기 때문에 여러 가지 부작용이 발견된다(국토교통부 2015; 2016). 국가하천 정비 과정에서 대규모로 지정하여 공원화한 뒤 이용객이 저조하여 친수지구의 효용에 대한 사회적 비판을 받는다는 점, 정비 후 그리 오랜 시간이 소요되지 않았음에도 지역·주민 요청에 의해 공간 내 시설을 변경하는 일이 빈번하게 발생하고 있다. 친수거점지구를 지정한 뒤 랜드마크 조성을 위해 많은 예산이 투입되었으나 실제로는 배후지역 주민 위주의 낮은 이용도를 보이는 경우도 있다. 하천기본계획을 자주 지정하거나 변경할 수 없으므로 사전 수요예측을 기초로 친수지구를 지정할 수밖에 없겠지만, 초기 지정된 친수지구는 이른바 ‘예정자로 간주하는 것이 바람직할 것이다. 실제로 하천을 대규모로 인공화하거나 근린친수나 친수거점으로 특성화하는 것은 지역주민의 하천이용 방식, 즉, 이용도와 공간활용의 유형을 면밀히 관찰하면서 조심스럽게 판단하는 것이 필요하다. 수요예측의 실패는 주민복지와 동떨어진 방식으로 하천을 조성·관리하게 되므로 불필요한 시설 설치와 부적절한 규제완화로 하천 난개발 등의 부작용뿐만 아니라, 장기적으로 유지관리 예산의 효율성 저하를 초래하기 때문이다.

본 논문에서는 이러한 문제점들을 인식하고 통신 빅데이터를 하천공간에 활용하여 그 해결책을 강구하

고자 하였다. 통신 빅데이터는 다양한 분야에서 활용되고 있는데 최근에 들어 맞춤형 행정과 시민의 수요에 따라 공원의 관리(오장근, 허학영, 심규원, 김태근 외 2017)나 교통의 수요관리(성지은, 박기량 2014; 김혜주 2017; 이광섭, 엄진기, 성명언, 유소영 외 2017; AirSage 2017) 분야에서 활발히 이용되고 있다. 또한 지역의 인구와 인구밀도 추정에 관련된 연구를 살펴보면 김종학, 고용석, 김준기, 김동한(2014)은 통신 빅데이터를 활용하여 인구의 활동을 사공간적의 변화에 따라 분포시켜 공간정책 활용방안을 제시하였으며, De Jonge, van Pelt and Roos(2012)는 통신 빅데이터로 주간 인구 밀도와 통근자 이동성을 추정하여 공식 통계화의 가능성을 탐색하였다. Deville, Linard, Martin and Gilbert et al.(2014)는 위치 데이터를 셀 단위로 하여 회귀분석을 실시한 뒤 인구밀도를 추정하는 연구를 수행하였으며, Douglas, Meyer, Ram and Rideout et al.(2015)는 독립변수를 이동전화 통화량과 토지피복데이터로 하여 인구 추정 회귀분석을 실시하였다. 이와 같이 통신 빅데이터는 다양한 분야에서 적용범위가 증가하고 있다.

본 논문은 친수지구 지정 후 하천공간을 체계적으로 특성화하는 방법에 초점을 두고 다음의 세 가지 목적을 가지고 수행되었다. 첫째, 친수지구의 이용도를 판단하기 위해 이용등급을 부여하는 기준을 제시한다. 이용등급은 상대적인 이용객 수에 대한 것으로 친수지구 운영의 효과와 직결되기 때문에 유지·보수비 예산 지원의 차등화 등에 활용될 수 있을 것이다. 둘째, 해당 친수지구가 지역주민의 근거리에서 도시공원이나 생활체육시설로 활용되는지, 아니면 문화·관광시설과 같이 지역 활성화를 위한 장소로 활용되는지에 대한 유형을 구분한다. 이러한 유형구분은 기반시설을 조성하거나 규제완화를 검토할 때 장소의 특수성에 대한 판단의 기준을 제공할 수 있을 것이다.

셋째, 친수거점지구를 지정하기 위한 체계적인 방법을 제안하고자 한다. 친수거점지구는 광역권의 명소로 레저, 문화 등 이용수요가 풍부하고 시설물의 집중설치가 필요한 지역이므로 하천의 고유 가치를 훼손하지 않는 범위 내에서 최소한으로 지정해야 한다. 따라서 친수거점지구를 지정하는 취지상 이용객이 상대적으로 많고 지역 활성화를 위해 필요한 곳인지를 반드시 확인하는 것이 바람직하다. 이에 본 논문에서는 친수지구 이용도 조사를 통해 이용등급과 공간활용의 유형을 고려하여 단계적으로 지정하는 방법과 기준을 제안하고자 한다.

본 논문은 이종소, 이상은, 최진영(2019)의 후속연구에 해당된다. 선행연구에서는 국가하천에 위치한 297개 친수지구의 이용도 조사를 위해 통신 빅데이터 자료의 유용성을 검증하였으며, 통신 빅데이터로부터 추정 가능한 정보를 토대로 다양한 이용지표를 개발한 바 있다. 본 논문은 선행연구에서 제시한 이용지표를 자료원으로 활용하여 하천관리 정책과 실무 차원에서 지역의 하천이용 수요를 잘 이해하고, 이로 인해 친수지구의 계획, 유지관리, 인·허가에 있어서 합리적인 결정을 하는 데에 크게 기여할 수 있을 것으로 예상된다.

## II. 방법론

본 논문에서는 친수지구 이용지표를 활용하여 친수지구 이용등급 및 상세유형화를 실시하고 친수거점지구를 지정하는 방법을 <Figure 1>과 같이 제시하고자 한다.

### 1. 통신 빅데이터를 활용한 친수지구 이용지표 산정

이종소, 이상은, 최진영(2019)은 국가하천 297개 친수지구를 대상으로 통신 빅데이터의 활용성을 검토하고 이용지표를 산정하여 친수지구를 이용하는 정도와 방식을 객관적으로 설명한 바 있다. 현장조사를 통해 직접 측정한 친수지구 이용객 수와 통신 빅데이터로 산정한 친수지구 이용객 수를 동일 날짜, 동일 시간대에 대해 비교하여 활용성을 검토하였다. 이어서, <Table 1>과 같이 친수지구의 이용도 조사를 위해 면적당 이용객 수, 하천이용 침투율, 월 최대 이용시기, 주중 이용률, 배후지역 이용빈도, 이용객의 성비, 이용객의 우점연령층, 1인당 이동거리 등 8가지 이용지표를 정의하고 2017년도 기간에 대해 시범적으로 지표값을 산정하였다. 본 논문에서는 이종소, 이상은, 최진영

Figure 1 \_ Flow Chart

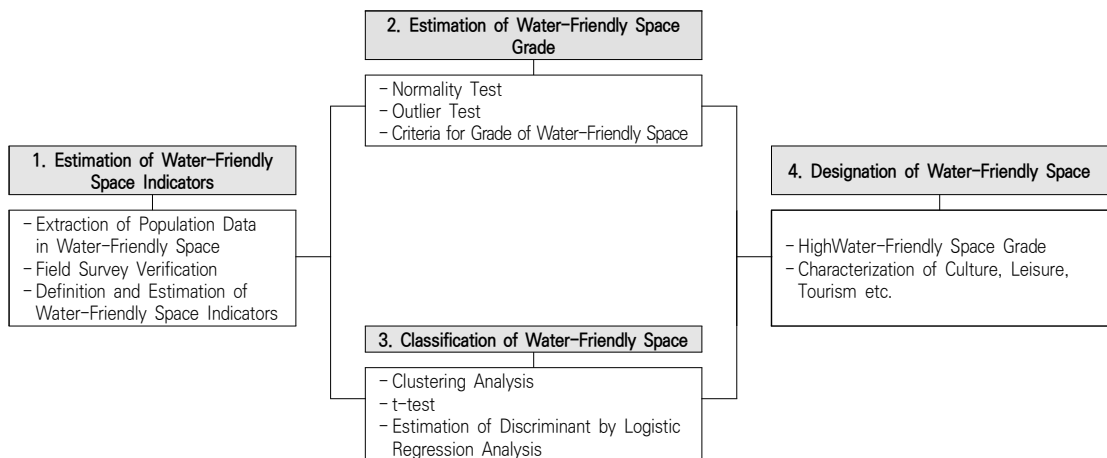


Table 1 \_ Definition of Water-Friendly Space Indicators Source

| Characteristics | Indicator                                           | Definition                                                               | Use in This Study |
|-----------------|-----------------------------------------------------|--------------------------------------------------------------------------|-------------------|
| Spatial Use     | Number of Visitors per Area(people/m <sup>2</sup> ) | Number of Visitors per Water-friendly Space Area(people/m <sup>2</sup> ) | ✓                 |
|                 | Peak Visit Rate(%)                                  | Maximum Number of Visitors/Minimum Number of Visitors                    | ✓                 |
|                 | Maximum Use Month(-)*                               | Month with the Most Visitors                                             |                   |
|                 | Weekly Visit Rate(%)                                | Weekly Visitors/Weekend Visitors                                         | ✓                 |
|                 | Residents' Visit Frequency(times)                   | Average Number of Visits of Residents                                    |                   |
|                 | Travel Distance per Visitors(km)                    | Travel Distance per Visitors                                             | ✓                 |
| Visitors        | Gender Ratio(%)                                     | Ratio of Male/Female Visitors                                            |                   |
|                 | Maximum Age Group(-)*                               | 10s or Less, 20s, 30s, 40s, 50s, 60s or More                             |                   |

Source: Lee, Lee and Choi 2019.

(2019)에서 제안한 다양한 이용지표 가운데 이용등급 산정과 친수저점지구 판별에 불필요한 월 최대 이용 시기, 이용객의 우점연령층, 이용객의 성비 등을 제외하고 면적당 이용객 수, 하천이용 침투율, 주중 이용률, 1인당 이동거리 등의 네 가지 지표를 자료원으로 활용하였다.

## 2. 친수지구 이용등급 산정

### 1) 정규성 검정

친수지구 이용도를 등급화하기 위해 전체 이용객 수, 면적당 이용객 수, 전체 이용객 수와 면적당 이용객 수의 조합 등 여러 가지 방법이 존재한다. 이 가운데, 이용객의 방문이 활발한 정도를 상대 비교하기 위해 친수지구 면적당 이용객 수를 토대로 등급화를 시도하기로 하였다. 2017년도 297개 친수지구의 면적당 이용객 수를 히스토그램으로 나타낸 결과 로그노멀(lognormal) 확률밀도함수와 유사한 형태를 관찰할 수 있었으므로 로그노멀 확률밀도함수에 대한 정규성을 만족하는지를 확인해 보고자 하였다.

먼저 자료의 정규성을 검정하기 위해 표본수가 많을 때 적합한 검정인 Kolmogorov-Smirnov 검정을 적

용하였다. Kolmogorov-Smirnov 검정은 누적 분포함수(Cumulative Distribute Function)의 개념을 이용하며 귀무가설  $H_0 : F(x) = F_0(x)$ 는 '표본의 분포는 정규성을 만족한다', 대립가설  $H_1 : F(x) \neq F_0(x)$ 는 '표본의 분포는 정규성을 만족하지 않는다'로 정의된다. 여기서,  $F_0(x)$ 는 가정된 이론 확률분포의 누가확률분포이며,  $F(x)$ 는 표본자료의 누가확률분포이다. 통상 검정결과가 유의수준 5%에서 p-value가 0.05 이상이 되면 귀무가설을 기각할 수 없기 때문에 정규성을 만족한다고 할 수 있다.

### 2) 이상치 검정 및 친수지구 이용등급 기준 산정

자료의 정규성 검정과 함께 이상치 검정을 실시하였다. 자료의 이상치(Outlier)는 어떤 집단으로부터 추출한 무작위 표본에서 다른 값들에 비해 비정상적으로 크거나 작은 값을 말하며, 이상치가 존재하면 계산된 통계량에 편중된 결과를 초래하게 된다. 본 논문에서는 Tukey(1977)가 제안한 이상치 판정 방식 가운데 제1사분위수( $Q_1$ ), 제3사분위수( $Q_3$ ),  $IQR(Q_3 - Q_1)$ 을 기준으로  $Q_1 \pm 1.5 \times IQR$ 를 적용해 면적당 이용객 수에 대한 이상치의 상·하한을 판정하였다.

다음으로는 친수지구의 이용등급 기준을 산정하기

위해 면적당 이용객 수를 기준으로 다섯 가지 등급으로 구분하고자 하였다. 로그로 변환된 면적당 이용객 수의 확률밀도 함수를 동일한 면적의 5개의 구간, 즉, 이용도가 매우 활발한 I등급부터 이용도가 매우 저조한 V등급으로 이용등급 기준을 구분하고 다시 면적당 이용객 수로 환원하여 등급별 기준을 제시하였다.

### 3. 친수지구 상세유형화를 위한 판별식 도출

#### 1) 군집분석

본 논문에서는 성격이 유사한 친수지구들을 특정 유형으로 묶기 위해 군집분석을 실시하고자 하였다. 군집분석은 데이터가 어떤 군집에 속해 있는지 알 수 없는 상태에서 거리값을 이용해 가까운 거리에 있는 것들끼리 묶어 분류하는 비지도 학습(Unsupervised Learning) 종류 중 하나로 기본 원리는 군집 내 응집도를 최대화하고 군집 간 분리도를 최대화하여 유사한 속성을 지닌 대상끼리 묶어 분류하는 것이다. 군집분석은 크게 계층적 군집(Hierarchical Clustering)과 분할적 군집(Partitional Clustering)으로 나뉘는데, 본 논문에서는 계층적 군집의 종류 가운데 Ward 연결법을 적용하고자 하였다. Ward 연결법은 군집 내의 오차제곱합(Error Sum of Square)에 기초하여 군집을 수행하며, 병합된 군집의 오차제곱합은 이전 각 군집의 오차제곱합보다 커지게 되는데, 그 증가량을 최소로 하는 방향으로 군집을 형성해 나가는 방법으로 다른 방법에 비해 노이즈나 이상치에 덜 민감한 장점이 있다(Ward 1963).

자료를 몇 개의 군집으로 분류할 것인가는 군집분석에서 매우 중요한 요소 가운데 하나이다. Charrad, Ghazzali, Boiteau and Niknafs(2014)는 최적의 군집 개수를 찾아내기 위해 Software R의 패키지인 NbClust를 개발하였다. NbClust는 30개의 최적 군집의 수 결정

방법과 군집의 수, 거리 측정 값, 군집 방법 등의 모든 조합 결과를 통해 최적의 군집개수를 제안하며, 이에 본 논문에서는 Charrad, Ghazzali, Boiteau and Niknafs(2014)가 개발한 NbClust를 활용하여 최적의 군집개수를 찾아내고자 하였다. 군집분석을 통해 친수지구의 유형을 판단하는 데는 <Table 1>의 8가지 이용지표 가운데 월단위 침투율, 주중 이용률, 1인당 이동거리 총 세 가지를 적용하였다. 또한 전체 이용객이 매우 적거나 월평균 이용객이 0~1명에 불과한 친수지구는 이용객 1인에 대한 지표 영향이 지나치게 커 이상치를 야기하므로 해당 친수지구를 제외하고 군집분석을 실시하였다.

#### 2) 독립표본 t검정

군집분석을 통해 분류된 유형 간의 특성을 통계적으로 분석하고 특징을 파악하기 위해 세 가지 지표 각각에 대해 독립된 두 모집단 평균차이 검증에 사용되는 t검정을 실시하고자 하였다. 특정 유형과 타 유형 간에 이용지표에 있어서 차이가 있을 것이라는 연구 가설하에 귀무가설  $H_0 : \mu_i^j = \mu_{other}^j$ 와 대립가설  $H_1 : \mu_i^j \neq \mu_{other}^j$ 을 설정하였다. 여기서  $\mu_i^j$ 는 1번째 유형의  $j$ 번째 지표의 평균값을 의미하며,  $\mu_{other}^j$ 는 타 유형들의  $j$ 번째 지표의 평균값을 의미한다. 또한 군집분석 결과에 따라  $i$ 는 유형의 개수를  $j$ 는 군집분석에 사용된 월단위 침투율, 주중 이용률, 그리고 1인당 이동거리 등 세 가지 이용지표를 각각 의미한다. 통상 유의수준 5%에서 p-value가 0.05 이상이 되면 귀무가설을 기각할 수 없게 되므로 차이가 있다고 할 수 없다.

#### 3) 로지스틱 회귀분석

로지스틱 회귀분석(Logistic Regression)은 회귀분석의

한 종류로서, 종속변수가 이분화된 범주형 변수일 경우 주로 사용된다. 종속변수를 확률변수인 로짓(Logit)으로 변환하여 선형적인 다중회귀모형을 구성한 뒤 모형의 회귀계수 등을 산정하여 종속변수와 독립변수들 간의 관계를 추정할 수 있다. 또한 로지스틱 회귀분석은 다변량 정상분포 가정을 요구하지 않으므로 이분화된 범주형 예측변수의 사용이 쉽다는 장점이 있기 때문에 본 연구에 적합하다.

로지스틱 회귀방정식의 회귀계수를 추정할 때에는, 최소자승법(Least Square Estimation)이 아닌 최대우도법(Maximum Likelihood Method)을 활용한다. 이는 준거변수가 이분변수(0과 1만 존재)일 경우 이분변수의 특성상 이분산의 가능성이 매우 높기 때문에 최소자승법의 가정인 오차의 등분산성을 만족시키기 어렵기 때문이다. 또한 준거변수 값의 빈도가 0일 경우 로짓의 값이 정의될 수 없어 음의 값이 예측되거나 부호 자체가 달라지는 등 추정값이 불안정해지기 때문이다(박광배 1999; 2000; Mertler and Vannatta 2001).

$p$ 개의 설명변수  $(x_1, x_2, \dots, x_p)$ 에 대하여 1을 가질 확률을  $P(Y=1|x_1, x_2, \dots, x_p)$ 라고 표시할 때 다음 <식 1>과 같이 계산된다.

$$P(Y=1|x_1, x_2, \dots, x_p) = \frac{\exp(\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p)}{1 + \exp(\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p)} \quad \text{<식 1>}$$

여기서  $\beta_0$ 는 절편을 의미하고,  $\beta_i (i=1, 2, \dots, p)$ 들은 회귀계수들을 의미한다. 이때  $\beta_0, \beta_1, \beta_2, \dots, \beta_p$ 는 최대우도법에 의해 추정된다. 로지스틱 회귀분석은 정규성, 선형성, 등분산성의 가정을 필요로 하지 않으며, 연속형 및 범주형 설명변수가 모두 포함된 모형으로 분석할 수 있다(계묘진 2012).

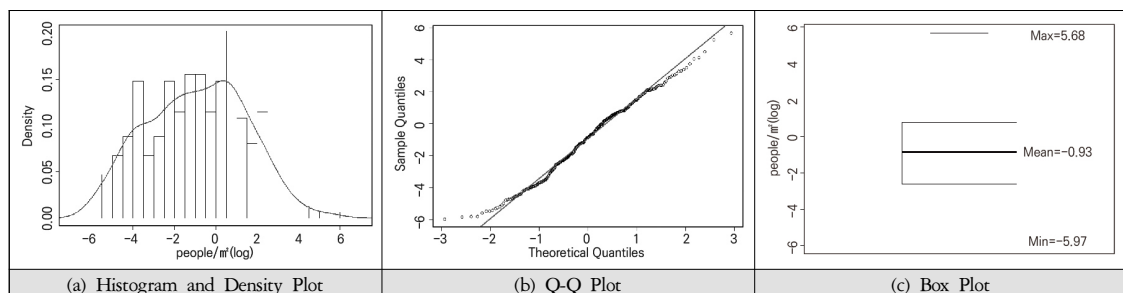
### III. 결과 및 검토

#### 1. 친수지구 이용등급 산정

lognormal 확률밀도함수로 변환된 면적당 이용객 수의 자료는 다음 <Figure 2>의 (a) 히스토그램과 밀도함수 그리고 <Figure 2>의 (b) Q-Q plot과 같이 정규성을 보이고 있음을 확인하였다. 또한 Kolmogorov-Smirnov 검정 결과, 양측검정에 대한 p-value는 0.38로 유의수준 5%에서 귀무가설인  $H_0 : F(x) = F_0(x)$ 를 기각할 수 없어 정규성을 만족한다고 할 수 있다. 이상치 검정을 위해 기초통계량을 분석한 결과, 1분위수( $Q_1$ ) = -2.60, 중간값(Median) = -0.86, 평균값 = -0.93, 3분위수( $Q_3$ ) = 0.79로 분석되었으며, 이상치의 상·하한은  $Q_1 - 1.5 \times IQR = -7.69$ ,  $Q_3 + 1.5 \times IQR = 5.87$ 로 <Figure 2>의 (c) Box Plot과 같이 이상치가 없는 것으로 나타났다.

친수지구의 이용등급 기준을 산정하기 위해 검정

Figure 2\_ Statistical Chart of Lognormal Probability Density Function(people/m<sup>2</sup>)



을 마친 로그로 변환한 면적당 이용객 수를 동일한 면적의 5개의 구간으로 이용등급기준을 구분하고 다시 면적당 이용객 수로 환원하였다. <Table 2>와 같

이 2.928명/m<sup>2</sup>, 0.950명/m<sup>2</sup>, 0.207명/m<sup>2</sup>, 0.033명/m<sup>2</sup>의 분류기준을 얻었으며, 등급별로 약 59개의 친수지구를 포함하는 것으로 분석되었다.

수계별로 볼 때 영산강 수계의 경우, 46개의 친수지구 가운데 15개(32.6%)지구가 이용도가 매우 저조한 V등급 구간으로 분류되었으며, 섬진강 수계 역시 5개(29.4%)의 지구가 V등급 구간에 분류된 것을 확인할 수 있었다(<Table 3> 참조). 전체적으로 낙동강 중류와 금호강, 한강 합류점(<Figure 3> 참조), 일부 도시구간에 있는 친수지구가 이용도가 매우 활발한

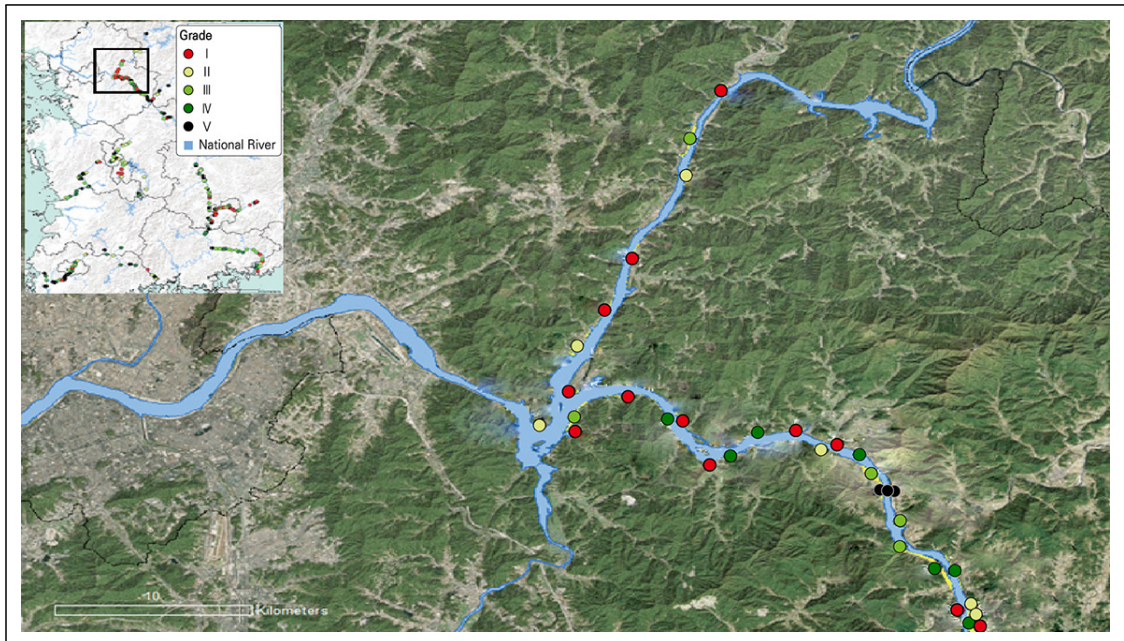
**Table 2 \_ Criteria for Grade of Water-friendly Space**

| Grade         | Grade Criteria                                                         |
|---------------|------------------------------------------------------------------------|
| I (Very High) | (people/m <sup>2</sup> ) > 2.928/m <sup>2</sup>                        |
| II (High)     | 0.950/m <sup>2</sup> < (people/m <sup>2</sup> ) < 2.928/m <sup>2</sup> |
| III (Normal)  | 0.207/m <sup>2</sup> < (people/m <sup>2</sup> ) ≤ 0.950/m <sup>2</sup> |
| IV (Low)      | 0.033/m <sup>2</sup> < (people/m <sup>2</sup> ) ≤ 0.207/m <sup>2</sup> |
| V (Very Low)  | (people/m <sup>2</sup> ) ≤ 0.033/m <sup>2</sup>                        |

**Table 3 \_ Grade of Water-friendly Space by River Basin**

| Basin          | Grade I   | Grade II  | Grade III | Grade IV  | Grade V   | Total    |
|----------------|-----------|-----------|-----------|-----------|-----------|----------|
| Han River      | 16(22.9%) | 14(20.0%) | 12(17.1%) | 13(18.6%) | 15(21.4%) | 70(100%) |
| Nakdong River  | 22(23.4%) | 13(13.8%) | 23(24.5%) | 20(21.3%) | 16(17.0%) | 94(100%) |
| Geum River     | 11(15.7%) | 22(31.4%) | 13(18.6%) | 15(21.4%) | 9(12.9%)  | 70(100%) |
| Yeongsan River | 8(17.4%)  | 7(15.2%)  | 10(21.7%) | 6(13.0%)  | 15(32.6%) | 46(100%) |
| Seomjin River  | 2(11.8%)  | 3(17.6%)  | 2(11.8%)  | 5(29.4%)  | 5(29.4%)  | 17(100%) |
| Total          | 59        | 59        | 60        | 59        | 60        | 297      |

**Figure 3 \_ Water-friendly Space Grade: Han River Basin Example**



반면, 그 외 많은 구간에서는 이용도가 저조한 경향을 보이는 것으로 분석되었다.

## 2. 친수지구 상세유형화를 위한 판별식 도출

### 1) 친수지구 상세유형화

군집분석을 위해 표준화된 세 가지 지표는 <Figure 4>와 같이 서로 간의 상관성은 매우 낮은 것으로 분석되었다. 친수지구의 유형을 분류하기 위해 Ward 연결법으로 군집을 분석하고 NbClust를 활용하여 최적의 군집개수를 탐색한 결과, 최적 군집의 개수는 <Figure 5>와 같이 5개로 분석되었으며, 273개소 친수지구(세 가지 지표의 이상치를 보인 친수지구 24개소 제외)를 5개의 유형으로 군집분석한 결과는 <Figure 6>과 같이 유형I, II, III, IV, V가 각각 59개소(21.6%), 78개소(28.6%), 81개소(29.7%), 41개소(15%), 14개소(5.1%)인 것으로 나타났다.

군집분석을 통해 분류된 유형 간의 특성을 분석하기 위해 독립된 두 모집단 평균차이 검증에 사용되는

t검정을 실시한 결과는 <Table 4~6>과 같다. 이를 해석하면, 유형I은 타 유형에 비해 월단위 침투율이 약간 낮고 주중 이용률은 높으며 1인당 이동거리는 매우 짧은 수준인 것으로 나타났다. 엑스포수상공원, 유등체육공원 등과 같이 근거리에서 주중에 방문하는 비중이 높으며 연중 이용추이는 완만한 특징이 관찰되었다. 유형II는 타 유형에 비해 월단위 침투율이 매우 낮고 주중 이용률은 매우 높은 수준이며 1인당 이동거리는 높은 수준인 것으로 나타났다. 주로 운동장 시설로 활용되는 친수지구가 다수 분포하였으며 비교적 원거리에서 주중에 방문하는 비중이 높고 연중 이용추이는 완만한 특징이 관찰되었다. 유형III은 타 유형에 비해 월단위 침투율이 낮고 주중 이용률도 낮으며 1인당 이동거리는 짧은 것으로 나타났다. 즉, 비교적 근거리에서 주말에 방문하는 비중이 높고 연중 이용추이는 완만한 특징이 관찰되었다. 유형IV는 타 유형에 비해 월단위 침투율이 높고 주중 이용률은 낮으며 1인당 이동거리가 매우 긴 것으로 나타났다. 두물지구, 평사리공원, 송호국민관광지 등과 같이 주로 원거리에서 주말에 집중적으로 방문하며 연중 이용추이

Figure 4 \_ Correlation Coefficients between Variables.

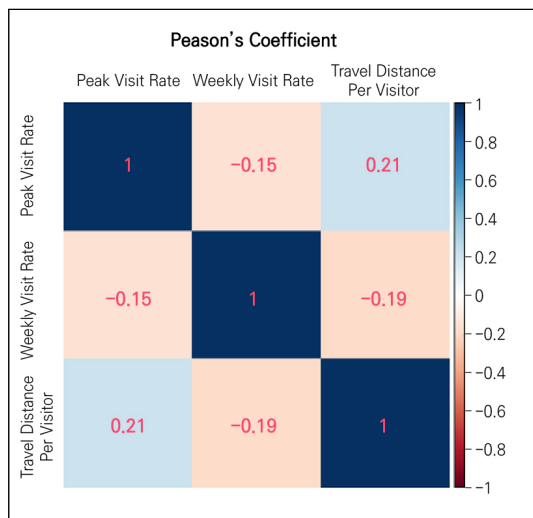
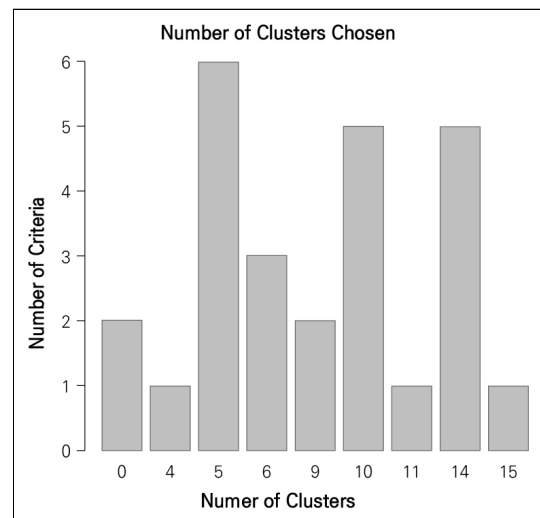
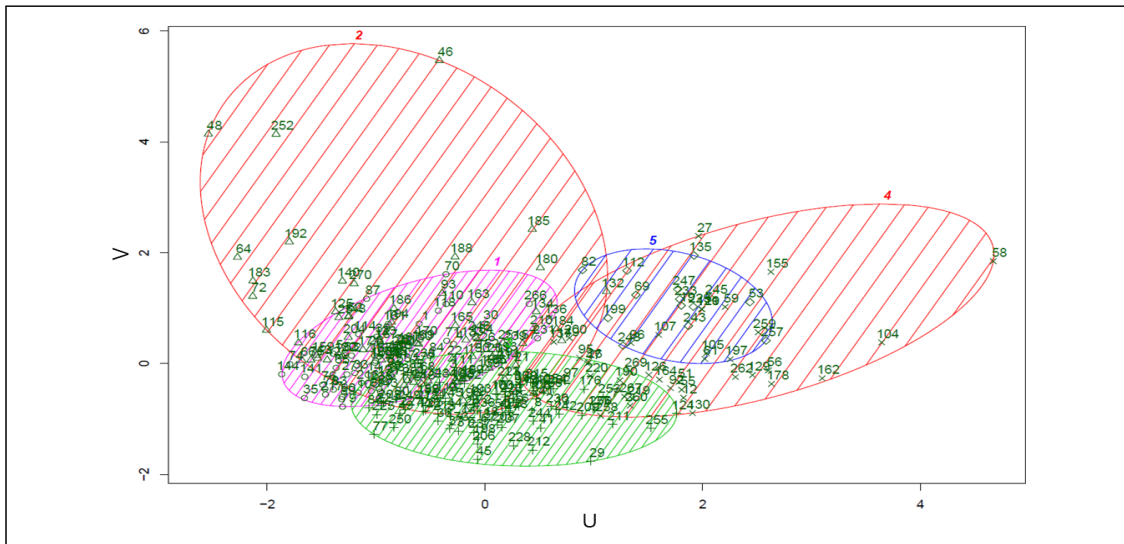


Figure 5 \_ Optimization of Number of Clusters



**Figure 6\_** Result of Clustering Analysis: Location of Principal Components U and V



**Table 4\_** Result of Independent Sample t-test: Peak Visit Rate

| Test Object       | Independent Sample t-test |                   |                 | 95% Confidence Interval |        | Equal Variance Assumption(Levene) |
|-------------------|---------------------------|-------------------|-----------------|-------------------------|--------|-----------------------------------|
|                   | t-value                   | p-value(Two Side) | Mean Difference | Lower                   | Upper  |                                   |
| Type I × Others   | -2.175                    | 0.031             | -0.934          | -1.782                  | -0.086 | -                                 |
| Type II × Others  | -6.116                    | ≒ 0               | -2.470          | -3.261                  | -1.671 | -                                 |
| Type III × Others | -4.490                    | ≒ 0               | -1.602          | -2.304                  | -0.899 | -                                 |
| Type IV × Others  | 5.990                     | ≒ 0               | 3.593           | 2.412                   | 4.773  | ✓                                 |
| Type V × Others   | 21.797                    | ≒ 0               | 11.041          | 9.973                   | 12.110 | -                                 |

**Table 5\_** Result of Independent Sample t-test: Weekly Visit Rate

| Test Object       | Independent Sample t-test |                   |                 | 95% Confidence Interval |        | Equal Variance Assumption(Levene) |
|-------------------|---------------------------|-------------------|-----------------|-------------------------|--------|-----------------------------------|
|                   | t-value                   | p-value(Two Side) | Mean Difference | Lower                   | Upper  |                                   |
| Type I × Others   | 6.3185                    | ≒ 0               | 0.180           | 0.124                   | 0.236  | -                                 |
| Type II × Others  | 9.942                     | ≒ 0               | 0.441           | 0.353                   | 0.529  | -                                 |
| Type III × Others | -11.881                   | ≒ 0               | -0.352          | -0.411                  | -0.294 | -                                 |
| Type IV × Others  | -8.324                    | ≒ 0               | -0.288          | -0.357                  | -0.220 | -                                 |
| Type V × Others   | -4.880                    | ≒ 0               | -0.207          | -0.295                  | -0.119 | -                                 |

**Table 6\_** Result of Independent Sample t-test: Travel Distance per Visitors

| Test Object       | Independent Sample t-test |                   |                 | 95% Confidence Interval |         | Equal Variance Assumption(Levene) |
|-------------------|---------------------------|-------------------|-----------------|-------------------------|---------|-----------------------------------|
|                   | t-value                   | p-value(Two Side) | Mean Difference | Lower                   | Upper   |                                   |
| Type I × Others   | -15.532                   | ≒ 0               | -25.020         | -28.192                 | -21.848 | -                                 |
| Type II × Others  | 3.226                     | 0.001             | 7.608           | 2.958                   | 12.259  | -                                 |
| Type III × Others | -3.690                    | ≒ 0               | -7.901          | -12.117                 | -3.685  | -                                 |
| Type IV × Others  | 11.502                    | ≒ 0               | 36.745          | 30.327                  | 43.164  | -                                 |
| Type V × Others   | -1.281                    | 0.201             | -7.299          | -18.521                 | 3.921   | ✓                                 |

에서 성수기의 특징이 관찰되었다. 마지막으로 유형V는 타 유형에 비해 월단위 침투율이 매우 높고 주중 이용율은 낮으나 1인당 이동거리는 타 유형과의 차이가 특별히 두드러지지 않는 것으로 나타났다. 이포보 오토캠핑장, 대성1친수공원 등과 같이 근거리와 원거리에서 주말에 방문하는 비중이 매우 높고 연중 이용 추이에서 성수기의 특징을 보이는 것으로 관찰되었다. 이용지표에 대한 유형의 통계적 특징과 유형에 속한 친수지구의 성격을 전체적으로 종합해 보면 유형I, II, III은 전반적으로 도시공원 또는 체육시설의 성격을 지니고 있어 근린친수지구로 간주가 가능하며, 유형IV, V는 전반적으로 주말을 이용한 문화·레저·생태교육시설 또는 지역행사 장소로서의 성격을 지니고 있어 친수거점지구로 간주할 수 있는 것으로 판단된다.

## 2) 로지스틱 회귀분석을 통한 판별식 도출

친수지구의 특성을 분류하기 위해서 매번 자료의 표준화와 군집분석을 실시하는 것은 복잡하고 매우 비효율적이라 할 수 있다. 이에 본 논문에서는 군집분석을 통해 분류된 근린친수지구의 성격을 갖는 유형I, II, III과 친수거점지구의 성격을 갖는 유형IV, V를 이분화하기 위해 로지스틱 회귀분석을 실시하였다.

유형I, II, III을 “0”, 유형IV, V를 “1”로 이분화하여

종속변수를 설정하였으며, 독립변수로는 군집분석에서 사용된 월단위 침투율, 주중 이용율, 1인당 이동거리의 원자료를 사용하였다. 개발된 로지스틱 회귀식은 <Table 7>과 <식 2>와 같으며,  $\hat{\pi} \leq 0.5$ 이면 유형I, II, III, 즉, 근린친수지구에 속하고  $\hat{\pi} > 0.5$ 이면 유형IV, V, 즉, 친수거점지구에 속하는 것으로 해석할 수 있다.

위 식의 정확도는 유형IV, V의 55개 가운데 46개를 구분해 내었고 유형I, II, III의 218개 가운데 214개를 구분해 내  $(46+214)/273=0.952$  수준으로 나타났다. 분류해내지 못한 친수지구의 경우는 대부분 군집 간 경계부근에 있어 근린친수지구와 친수거점지구의 특징을 명확히 정의하기 힘든 것으로 나타났다. 친수거점지구 성격이 있음에도 근린친수로 간주하게 되는 경우가 다소 발견된 것이 사실이다. 그러나 친수거점지구를 보수적으로 판별하는 것은 난개발을 방지하고 예산을 효율적으로 활용하기 위해 친수거점지구를 최소한으로 지정해야 한다는 점을 감안할 때 정확도에 문제가 있다고 보기는 힘들다.

## 3. 친수거점지구 지정을 위한 단계별 판단기준 제안

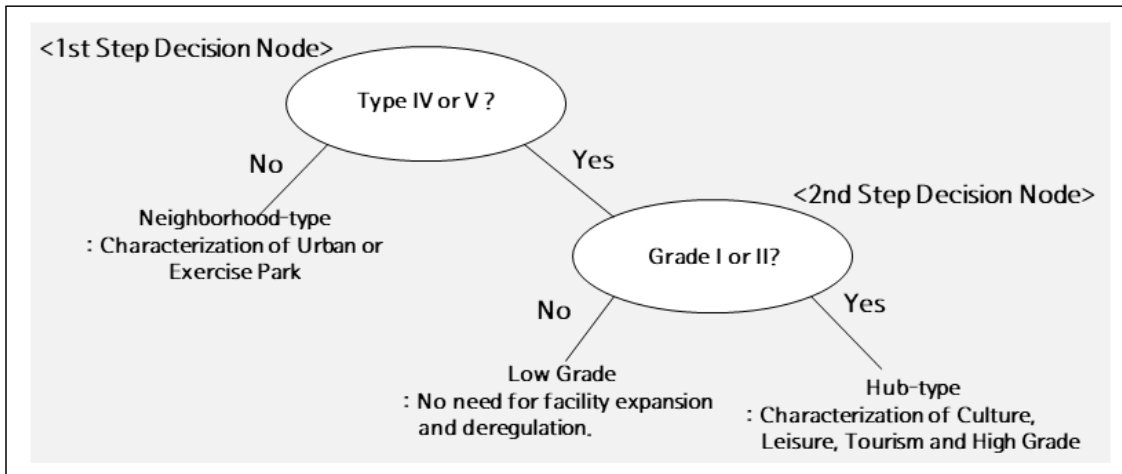
친수거점 지구는 하천의 고유 가치를 훼손하지 않는 범위 내에서 최소한으로 지정해야 하며, 광역권의 명

$$\hat{\pi}(x) = \frac{\exp(-14.480 + 0.984x_1 - 5.879x_2 + 0.170x_3)}{1 + \exp(-14.480 + 0.984x_1 - 5.879x_2 + 0.170x_3)} \quad \text{<식 2>}$$

Table 7\_ Result of Discriminant by Logistic Regression Analysis

| Variable                              | Estimate  | Std.Error | z value | Pr(>  z ) |
|---------------------------------------|-----------|-----------|---------|-----------|
| Intercept                             | -14.48008 | 3.10618   | -4.662  | ≒0        |
| Peak Visit Rate( $x_1$ )              | 0.98376   | 0.19453   | 5.057   | ≒0        |
| Weekly Visit Rate( $x_2$ )            | -5.87905  | 1.37154   | -4.286  | ≒0        |
| Travel Distance per Visitors( $x_3$ ) | 0.16981   | 0.03685   | 4.609   | ≒0        |

Figure 7 \_ Decision Tree for Classification of Water-friendly Space



소로 레저·문화 등의 성격을 갖고 이용수요가 풍부해야 한다. 그러기 위해서는 현재처럼 친수거점지구를 객관적이고 명확한 기준 없이 먼저 지정하기보다는 친수지구를 조성한 뒤 중장기적인 모니터링을 실시하여 친수거점지구인지를 세분화해야 할 것이다.

중장기적인 모니터링 실시 후 본 논문의 결과를 활용하여 로지스틱 회귀분석을 통해 유형IV, V에 속하는지를 <식 2>을 통해 판단한 뒤, <Table 2>에 따라 해당 친수지구의 이용등급이 I 또는 II 등급에 속하게 되면 주말을 이용한 문화·레저·생태교육시설 또는 지역행사장소로서의 성격을 지니고 있다고 판단할 수 있을 것이다. 즉 로지스틱 판별식의 결과가 0.5 이상이고 면적당 이용객 수가 0.950명/m<sup>2</sup> 이상이 된다면 향후 친수거점지구 지정을 고려해 볼 수 있을 것이다.

#### IV. 결론

본 논문은 친수지구를 지정한 이후 하천공간을 특성화하는 방법에 초점을 두고 이용도를 판단하기 위한 이용등급 산정, 이용의 특성을 분류하기 위한 상세유형화 그리고 친수거점지구 지정을 위한 체계적인 방

법을 제안하고자 하였으며, 연구결과는 다음과 같이 요약할 수 있다.

첫째, 친수지구의 이용등급을 산정하기 위해 면적당 이용객 수의 지표를 lognormal 확률밀도함수로 변환하고 정규성과 이상치 검정을 실시하였다. 그 결과 등급별로 약 59개의 친수지구를 포함하는 것으로 분석되었는데 영산강과 섬진강 수계에 이용등급이 저조한 친수지구가 많았으며, 전체적으로 낙동강 중류와 금호강, 한강 합류점 및 일부 도시구간에 있는 친수지구의 이용이 활발한 것으로 분석되었다.

둘째, 친수지구 상세유형화를 위해 군집분석을 실시하여 총 5개의 유형으로 분류하였으며, 통계검정을 통해 각 유형 간의 특징을 분석하였다. 또한 향후 친수지구의 성격을 분류하기 위해 매년 표준화 및 군집분석을 실시하는 것은 복잡하고 비효율적일 수 있기 때문에 실무에서 간편하게 활용될 수 있도록 로지스틱 회귀분석을 통해 근린친수지구와 친수거점지구의 유형을 이분화하였다. 유형I, II, III은 전반적으로 도시공원 또는 체육시설의 성격을 지니고 있어 근린친수지구로 간주가 가능한 것으로 나타났으며, 유형IV, V는 전반적으로 주말을 이용한 문화·레저·생태교육

시설 또는 지역행사 장소로서의 성격을 지니고 있어 친수거점지구로 간주가 가능한 것으로 나타났다. 로지스틱 회귀분석을 통해 근린친수지구와 친수거점지구를 분류할 때 95.2% 수준의 정확도 높은 판별식을 도출할 수 있었으며, 이 판별식은 다소 보수적으로 친수거점지구를 분류한다는 점을 확인할 수 있었다.

셋째, 본 논문에서 수행한 연구결과들을 종합하여 단계적으로 친수지구를 지정하는 방법과 기준을 제안하였다. 친수거점지구는 이용객이 상대적으로 많고 광역권의 명소로 레저·문화 등의 성격을 지니고 있기 때문에 상세유형결과의 유형Ⅳ와 Ⅴ에 속하며, 이용등급이 I, II등급에 속하게 된다면 친수거점지구의 지정을 고려해볼 수 있을 것이다. 중장기적인 모니터링 실시 후 본 논문의 결과를 활용하여 로지스틱 판별식의 결과가 0.5 이상으로 유형Ⅳ, Ⅴ에 속하는지는 판단한 뒤, 면적당 이용객 수가 0.950명/m<sup>2</sup> 이상이 되어 이용등급이 I 또는 II 등급에 속하게 되면 주말을 이용한 문화·레저·생태교육시설 또는 지역행사장소로서의 성격을 지니고 있다고 판단할 수 있을 것이다.

2007년 「하천법」 개정 이후 지역주민의 공간활용 방식에 대한 객관적이고 면밀한 수요조사 방법이 정립되어 있지 않은 상태에서 국가하천 정비 시 대규모로 친수거점지구로 지정하여 공원화하였다. 그 결과, 몇몇 지구에 대해서는 이용객 저조에 따른 친수지구의 효용에 대한 사회적 비판, 하천 난개발의 부작용, 유지관리 예산의 효율성 저하 등의 부작용을 초래해왔다. 이에 근린친수지구나 친수거점지구로 특성화하는 것은 지역주민의 하천이용 방식과 공간 활용의 유형을 면밀히 관찰하면서 조심스럽게 판단하는 것이 필요할 것으로 생각되며, 먼저 세분화하여 지정하기 보다는 친수지구를 지정한 이후 중장기적인 모니터링과 객관적인 분석을 통해 근린친수지구와 친수거점지구로 지정해야 할 것이다.

본 논문의 이용등급 결과를 통해 친수지구의 효율적인 운영과 유지·보수비 예산 지원 차등화 등에 활용될 수 있을 것이며, 상세유형화 및 친수거점지구 지정 방법은 친수지구계획, 기반시설 조성, 인·허가에 있어 합리적인 결정을 하는 데 크게 기여할 수 있을 것으로 판단된다.

## 참고문헌 •••••

1. 계묘진. 2012. Lasso를 기반으로 한 로지스틱회귀모형 연구. 석사학위논문, 계명대학교.  
Kye Myojin. 2012. *A Study on Logistic Regression Model based on Lasso*. M.D. diss., Keimyung University.
2. 국토교통부. 2015. 4대강 수변 친수공간 관리 개선방안. 세종: 국토교통부.  
Ministry of Land, Infrastructure and Transport(MOLIT). 2015. *Improvement of the Management of Water-friendly Space in 4 Rivers*. Sejong: MOLIT.
3. \_\_\_\_\_. 2016. 하천 유지관리 평가 및 개선방안 연구. 세종: 국토교통부.  
\_\_\_\_\_. 2016. *Study on River Maintenance Evaluation and Improvement Plan*. Sejong: MOLIT.
4. \_\_\_\_\_. 2018. 하천 점용제도 개선 및 표준 매뉴얼 개발 연구. 세종: 국토교통부.  
\_\_\_\_\_. 2018. *A Study on the Improvement of River Space Use Permission and Development of Standard Manual*. Sejong: MOLIT.
5. 김종학, 고용석, 김준기, 김동한. 2014. 스마트 셀 기반 활동 인구의 공간정책 활용방안 연구. 세종: 국토연구원.  
Kim Jonghak, Ko Yongseok, Kim Joonki and Kim Donghan. 2014. *The Application of Smart Cell in Space Policy*. Sejong: Korea Research Institute for Human Settlements.
6. 김혜주. 2017. 통신 유동인구 빅데이터로 교통혁명 시대를 열다. 월간교통 229호, 11-15. 세종: 한국교통연구원.  
Kim Hyeju. 2017. Opening the traffic revolution in communication big data. *Monthly KOTI Magazine on Transport* 229, 11-15. Sejong: The Korea Transport Institute.
7. 박광배. 1999. 변량분석과 회귀분석. 서울: 학지사.  
Park Kwangbae. 1999. *Analysis of Variance and Regression*.

- Seoul: Hakjisa
8. \_\_\_\_\_. 2000. 다변량분석. 서울: 학지사.  
\_\_\_\_\_. 2000. *Multivariate Analysis*. Seoul: Hakjisa
  9. 성지은, 박기량. 2014. 빅데이터를 활용한 정책 사례 분석과 시사점. 과학기술정책 24권, 2호: 94-106.  
Seong Jieun and Park Kiryang. 2014. Policy case analysis and implications using big data. *Science and Technology Policy* 24, no.2: 94-106.
  10. 오장근, 허학영, 심규원, 김태근, 최진영. 2017. 통신 빅데이터를 활용한 국립공원 탐방객 실태조사 및 이용패턴 분석검증 연구. 원주: 국립공원관리공단.  
Oh Jangeun, Heo Hakyong, Sim Gyuwon, Kim Taegeun and Choi Jinyoung. 2017. *A Study on National Park Visitor Survey and Pattern Analysis Using Mobile Big Data*. Wonju: Korea National Park Service.
  11. 이광섭, 엄진기, 성명연, 유소영, 민재홍, 이준. 2017. 통신량 빅데이터 활용성에 관한 연구: 의정부시 내·외부 활동 및 이동패턴과 도시철도망 연계성을 중심으로. 국토계획 52권, 5호: 113-130.  
Lee Kwangsub, Eom Jinki, Seong Myeongeon, You Soyoung, Min Jaehong and Lee Jun. 2017. Usability of mobile phone big data: Focusing on the activity and mobility patterns and urban railway network in the Uijeongbu City. *Journal of Korea Planning Association* 52, no.5: 113-130.
  12. 이종소, 이상은, 최진영. 2019. 국가하천 친수지구 공간관리를 위한 통신 빅데이터 활용성 검토: 자료검증과 이용지표 선정. 국토연구 101권: 3-18.  
Lee Jongso, Lee Sangeun and Choi Jinyoung. 2019. Using the mobile big data for the smart river space management: Data validation and water-friendly space indicators. 2019. *The Korea Spatial Planning Review* 101: 3-18.
  13. AirSage. 2017. Innovations in data collection: The power of where and when. <http://tnmug.utk.edu/wp-content/uploads/sites/47/2017/06/AirSageTNMUG.pdf> (accessed May 22, 2019).
  14. Charrad, M., Ghazzali, N., Boiteau, V. and Niknafs, A. 2014. NbClust: An R package for determining the relevant number of clusters in a data set. *Journal of Statistical Software* 61, no.6: 1-36.
  15. De Jonge, E., van Pelt, M. and Roos, M. 2012. Time patterns, geospatial clustering and mobility statistics based on mobile phone network data. In *Proceedings of The Federal Committee on Statistical Methodology Research Conference*. January 10-12, Washington D.C.: Washington Convention Center.
  16. Deville, P., Linard, C., Martin, S., Gilbert, M., Stevens, F. R., Gaughan, A. E. and Blondela, V. D. et al. 2014. Dynamic population mapping using mobile phone data. *Proceedings of the National Academy of Sciences* 111, no.45: 15888-15893.
  17. Douglass, R. W., Meyer, D. A., Ram, M., Rideout, D. and Song, D. 2015. High resolution population estimates from telecommunications data. *EPJ Data Science* 4, no.1: 1-13.
  18. Mertler, C. A. and Vannatta, R. A. 2001. *Advanced and Multivariate Statistical Methods: Practical application and interpretation*. Los Angeles: Pyrczak Publishing.
  19. Tukey, J. W. 1977. *Exploratory Data Analysis*. Boston: Addison-Wesley.
  20. Ward Jr., J. H. 1963. Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association* 58: 236-244.
- 
- 논문 접수일: 2019. 6. 11.
  - 심사 시작일: 2019. 7. 18.
  - 심사 완료일: 2019. 9. 11.

---

## 요약

주제어: 친수지구, 친수지구 유형화, 모바일 통신 빅데이터, 하천관리

도시의 개발, 재생, 복지, 문화 관광 등의 지역 정책과 하천이용의 수요에 맞춰 하천 공간을 지정하고 운영하는 것이 한층 더 중요해졌다. 이에 본 논문은 체계적이고 객관적으로 하천 공간을 지정·운영하기 위해 선행연구에서 검증된 통신 빅데이터 이용지표를 사용하여 친수지구를 특성화하는 방법을 제안하고자 한다. 첫째, 친수지구의 이용도를 판단하기 위해 면적당 이용객 수를 로그노멀 분포로 변환하여 5등급으로 제시하였다. 둘째, 여러 가지 통계분석을 통해 친수지구를 5개의 유형으로 구분하고 특징을 분석하

였으며, 친수지구가 근린친수와 친수거점의 어디에 해당하는지 구분하였다. 셋째, 이용등급과 공간 활용의 유형을 고려하여 친수지구를 단계적으로 지정하는 방법과 기준을 제안하였다. 향후 중장기적인 모니터링과 객관적인 분석을 통해 근린친수지구와 친수거점지구로 지정해야 할 것이며, 본 논문의 결과는 친수지구의 계획, 유지관리, 인·허가에 있어서 합리적인 결정을 하는 데 크게 기여할 수 있을 것으로 예상된다.