

정보계측기법을 이용한 생태조사자료의 종다양도 분석

정남수 · 이정재* · 박승기 · 김웅

공주대학교 생물산업공학부 · *서울대학교 조경 · 지역시스템공학부

Species Diversity Analysis of Ecosystem Survey Data Using Total Information

Jung, Nam-Su, · Lee, Jeong-Jae* · Park, Seung-Kie · Kim, Woong

Dept. of Bio-Industry Engineering, Kongju, Nat'l Univ.

**Dept. of Landscape Architecture and Rural Systems Engineering, Seoul Nat'l Univ.,*

ABSTRACT : Shannon and Simpson indexes are used for species diversity analysis of ecosystem. In species diversity analysis of ecosystem, not only frequency of each species but also survey size have to be considered. In this study, total information composed with knowledge and ignorance was suggested as a species diversity analysis method for ecosystem survey. To apply developed method, flora in the Sangachun river valley was sampled with 19 sites and 198 species. In applying results, Shannon index shows more reasonable results than Simpson index by the variance of sample size but has difficulties of determining the relation of surveying species number and sample site number. Suggested total information can overcome this difficulty by the relation of knowledge and ignorance.

Key words : Diversity Index, Ecosystem, Information Measure

1. 서 론

비오톱과 같은 생태자원은 농촌을 구성하는 중요한 요소이므로 농촌자원조사에서 필수적으로 조사해야 하는 항목이며, 생태마을, 환경친화형 마을, 자연형 하천 등 농촌계획에서도 그 위상이 계속 높아지고 있다(Jung, 2006).

현재 많은 식물, 수목, 어류, 조류, 곤충 등 많은 조사가 진행되고 있으나, 조사해야 하는 규모에 비하여 조사 방법 및 자료처리의 결과가 일부 전문가들만이 수행할 수 있도록 극히 난해한 형편이다. 특히, 식생조사의 경우 종다양성 평가 보다는 환경부에서 지정한 특정생물군이 존재하느냐의 여부에 따라 중요도가 결정되어 희귀종의 보호를 목적으로 하는 식생조사에는 유리하나 농촌의 생태적 건전성을 파악하여 계획에 이용하기에는 어려운 실정이다(Lee et al, 2002).

현재 조사된 자료를 이용하여 다양도를 평가할 수 있는 방법은 Simpson지수와 Shannon지수가 가장 많이 이용되고 있으나, 개체의 출현확률을 고려하여 계산된 결과이며 조사규모를 결정할 수 있는 객관적 기준을 설정하기가 곤란하다. 이 중에서 Shannon지수는 정보이론에 따라 도출한 것이며 불확실성의 개념을 기초로 하여 다양도가 높으면 불확실성이 높고 다양도가 낮으면 불확실성이 낮다는 개념에서 이용되고 있다(Kim, 2006).

Gell-Mann and Lloyd (1996)는 복잡계의 해석을 위하여 정보계측에서 Shannon의 불확실성과 더불어 출현요소가 주는 지식(knowledge)을 같이 고려할 것을 제안하였으며, Jung(2003)은 농촌어메니티와 같은 불확실한 개념에 대한 가치평가를 위하여 이를 총정보량식으로 개발하였다.

총정보량식은 지식과 불확실성의 합으로 구성되는데 지식이란 출현한 요소들을 나타내기 위한 최소한의 정보량이며 불확실성은 출현한 요소들이 내포하는 엔트로피이므로 이들을 이용하면 다양성평가와 같이 실체를 완전하게 접근하기 힘들고 샘플링을 통해서 접근해야 하는

Corresponding author : Lee, Jeong-Jae

Tel.: 02-880-4581

E-mail: ljj@snu.ac.kr

문제에 대한 해결이 가능할 것으로 판단된다.

본 연구에서는 총정보량을 구성하는 지식과 불확실성을 수식으로 전개함으로써 그 물리적 의미를 살펴보고 2004년도 충남 예산군 덕산면 상가천을 대상으로 조사한 식생의 빈도분석 결과에 적용하여 기존의 Simpson과 Shannon의 지수와 비교함으로써 현재의 생태계다양성평가의 문제점과 총정보량식의 적용가능성을 보이고자 한다.

II. 총정보량에 관한 이론적 고찰

정보는 수학적 접근을 통하여 정량화 될 수 있다. Shannon(1948)은 정보를 정량화 하기 위하여 다음의 식을 제안하였다.

$$s = -k \sum_r P_r \log P_r, \quad r \in \text{Set } A \quad (1)$$

where, s : entropy,
 k : the number of trials for selection,
 P_r : the probability of selecting element r in set A

식 (1)에서 Shannon(1948)이 제시한 정보량이란 집합 A 에서 특정한 사상 r 을 선택하기 위해 필요한 선택의 횟수를 전체요소에 대해 합한 값을 나타내며 이는 집합 A 를 통신매체를 통하여 옮기고자 할 때 필요한 비트로 볼 수 있다.

Shannon의 식에 근거하여 Gell-Mann and Lloyd (1996) 등은 실제의 가치를 나타내는 총정보량을 다음과 같은 식으로 제안하였다.

$$I(A) = \epsilon + s = K(A) + \sum_r P_r \log \frac{1}{P_r} \quad (2)$$

여기서, $I(A)$: the total information of set A ,
 ϵ : the knowledge,
 s : the ignorance (entropy),
 $K(A)$: algorithmic complexity of set A .

식 (2)에서 Gell-Mann(1996)은 Shannon(1948)이 제시한 정보량을 집합 A 가 나타낼 수 있는 불확실성으로 보고, 집합 A 를 설명하기 위한 최소한의 복잡성(Algorithm Complexity)를 지식으로 간주하여 이의 합을 총정보량으로 정의하였다.

여기서 집합 A 를 설명하기 위한 최소한의 복잡성

(Knowledge)은 기술(description)의 정도, 기술하기 전 이해의 정도, 기술에 사용되는 언어, 언어를 비트로 전환하는 방법 등에 의해 달라질 수 있다고 하였다. Algorithm Complexity에 대한 연구는 현재까지도 많은 연구가 진행 중이나, 몇몇 수학적 성질에 대한 증명에 그치고 있으며 (Li, 1997) 현실적으로 이용하기 힘든 형편이다. jung(2003)은 이를 불확실성과 같은 요소의 수준에서 정의하기 위하여 식 (3)을 제안하였다.

$$\epsilon = \left(\sum_r (-1/\log(\frac{r}{1+n})) \right) s_{relative} \quad (3)$$

여기서, $\log N(r)$: the logarithmic number of element r in set A .

$s_{relative}$: relative entropy

예를들어, 집합 A 에서 특정사상 r 이 1회 출현하였다고 하면, 이를 표현하기 위한 정보량은 단위비트 1이 된다. 그러나, Crutchfield(1994)가 지적한대로, 집합 A 에서 추출의 횟수가 많아지면, 즉 추출한 요소들의 Randomness(임의성)이 증가하면 시스템의 복잡성은 꾸준히 증가하다 일정시점을 지나면서 줄어들게 된다. 이는 단위요소를 표현하기 위한 최소한의 복잡성 또한 일정하지 않고 집합 내에서의 단위요소의 출현확률에 따라 변화한다는 것을 의미한다. jung(2003)은 이를 로그함수로 근사하였고, 0이하의 정보량을 없애기 위해 밑수에 1을 증가시켰다. 또한, 집합 A 를 구성하는 요소의 수가 증가할수록 요소들 내부에 생성되는 리턴던시(여분의 정보량)가 증가하므로 그 효과를 제거하기 위하여 구해진 복잡성에 구성요소 증가에 따른 상대적 엔트로피를 곱하였다.

또한, jung(2003)은 샘플자료의 시물레이션을 통하여 개발된 식이 정보량의 필요조건을 만족시킴을 증명하였으므로 본 연구에서 이 부분은 다루지 않았다.

III. 총정보량의 중다양성 평가에의 적용

1. 자료조사

대상하천은 충남 예산군 덕산면 상가리의 상가천이며 덕산면 소재지에서 북서쪽에 위치하고 있다. 상가천의 유역경계는 해발 약 678m의 가야산을 중심으로 석문봉(635m), 옥양봉, 일낙산, 원효봉이 있어 비교적 험준한 산지 지형을 이루고 있으며, 북쪽으로는 예산군 봉산면과 서산시 운산면이, 서쪽으로는 서산시 해미면이, 남쪽으로는 예산군 덕산면 대치리가 위치하며 유역도는 그림

1과 같다.

조사대상지는 행정자치부 국립방재연구소에서 소하천 정비공법개발을 위하여 선정한 전국 8개소의 소유역중 하나이며, 선정된 8개 구역에서는 호안공법적용에 따른 수리, 수분변화, 하상변동, 생태계 모니터링 등을 수행하였다. 대상지는 산지부와 농경지가 복합된 구역의 특성을 반영하기 위하여 선정하였으며, 소하천정비 공사구간 내 하천에 자생하는 식물체는 모두 제거되며, 공사 완공 후 회복되는 과정을 규명하기 위한 기초자료로 활용하기 위하여 본 조사를 실시하였다.

상가천은 산지지형을 이루고 있는 구역의 북서쪽에서 발원하여 평야지로 이어지는 구역의 동남쪽으로 유하하여 옥계저수지를 이루고 덕산천을 통하여 삼교천 본류에 합류된다.

하천 식물상 조사지점은 2003년 2004년과 같은 총 19 개소이며 자연형하천 공사구간 상류지역 3개소, 공사구간 10개소 및 공사 하류지역 6개소로 정하여 실시하였다.

2. 조사결과의 평가

다양성 지수는 풍부도와 균등도의 조합으로 나타낼 수 있다. 종 다양성 지수는 Simpson(1949)과 Shannon (1948)의 지수가 유용하게 사용되어 왔다. Simpson지수는 우점도(dominance)로부터 출발하는 것으로 최고의 풍부성을 지닌종을 희소종보다 중요하게 인정한다. 따라서 이 지수의 수치는 종풍부도와 개체들의 균등도를 포함한다. 한 군집으로부터 두 개체를 랜덤하게 추출하였을 때 두 개체가 같은 종에 포함될 확률은 식 (4)로 나타낼 수 있다. (Kim, 2006)

$$C = \frac{\sum n_i(n_i - 1)}{N(N-1)} \quad (4)$$

여기서, N = total number, n_i :the number of each species

여기에서 C는 Simpson의 우점도지수이다. 한 군집의 우점도 지수가 높으면 다양도 지수가 낮게 된다. 그러므로 Simpson의 다양도지수(Ds)는 전체 값 1에서 우점도지수를 뺀 값으로 식 (5)와 같이 나타낼 수 있다. (Kim, 2006)

$$D_s = 1 - C \quad (5)$$

조사결과를 요약해보면 본 논문에서는 이 중에서 Shannon의 다양성 지표를 이용하여 조사된 내용 중 식물종을 표 1과 같이 정리하였다.

표 1. 대상하천의 조사구간별 출현종

No.	출 현 종	조 사 지 점 명					
		1	...	10	11	...	19
1	물봉선	○	○	○	○	○	
2	사위질빵	○	○	○			
3	며느리밑씻개	○					
4	갯버들	○				○	
5	달뿌리풀	○	○	○	○	○	○
6	환삼덩굴	○	○	○	○	○	○
7	바랭이	○		○	○	○	
8	자귀풀	○					
9	쇠무릎	○			○		
...
198	황새냉이			○	○	○	
지점별 출현종수		15	23	60	48	50	35

※ 지면상 일부만으로 편집하였음

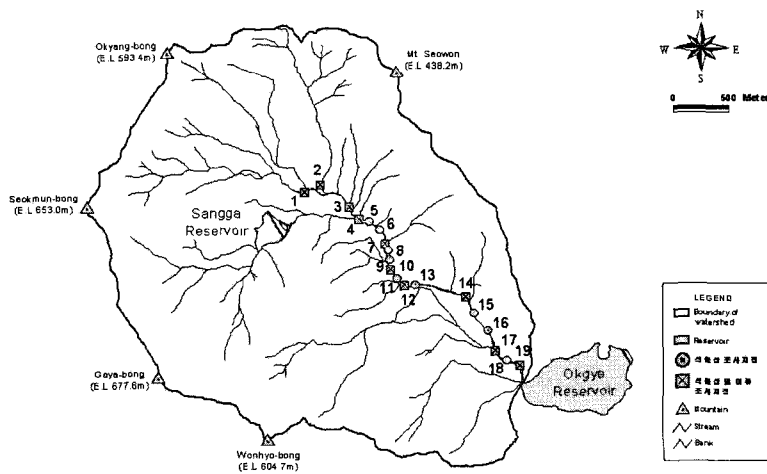


그림 1. 대상하천 유역도 및 식물상 조사지점

정리된 자료를 바탕으로 식물다양성을 평가해 보면 표 2와 같다.

표 2. 조사구간별 다양성 지수

조사구간	1-10	11-19	1-19
조사종수	135	157	198
조사개체수	372	356	728
Shannon지수	6.6319	7.0073	7.1298
Simpson지수	0.9897	0.9933	0.9913

Simpson 다양성 지수 산출 결과에서 알 수 있듯이, 총 개체수에서 종수의 종류가 많을수록 다양성 지수가 높아 짐을 알 수 있었다. 반면 Shannon지수에서는 조사구간이 1에서 10까지 11에서 19까지 나누어져 계산된 지수값보다 1에서 19구간을 모두 한꺼번에 처리한 결과가 높게 나타나 Simpson에서 11에서 19구간이 가장 높게 나타나는 결과보다 좀더 합리적인 결과로 이해할 수 있었다.

그러나 여기서 몇 가지 문제를 제기할 수 있다. 그것

은 첫째 다양성 지표는 분류하고자 하는 종수가 많아지면 많아질수록 높아진다는 것이었다. 즉, 조사자가 알고 있는 식물종이 1000종이나, 2000종이나에 따라 다양성지수를 다르게 나타낼 수 있다는 것이다. 두 번째 총 4km에 달하는 현장 중에서 몇 번이나 조사해야 대상지역을 대표할 수 있는지에 대한 기준이 없다는 것이다.

3. 총정보량의 적용

만들어진 기준을 이용하여 조사된 식물자료에 적용해 본 결과 그림 2와 같이 조사에 대한 Knowledge과 Ignorance 모두 계속적으로 증가하고 있음을 알 수 있다. 이는 사이트의 증가에 대해 지속적으로 새로 추출되는 식물종이 증가함으로써 이러한 현상을 나타내고 있다.

따라서, 파악하는 종수를 40개로 제한하고 계산을 다시 수행해 보면 그림 3과 같이 조사사이트가 7이 넘어가는 지점부터 조사된 요소들의 ignorance가 조사된 요소들로부터 파악할 수 있는 knowledge를 넘어섰으며, 안정적인 값에 도달함을 알 수 있었다. 본 연구는 그 범위를 총정보량을 생태계조사에 이용할 수 있는지에 한정하였

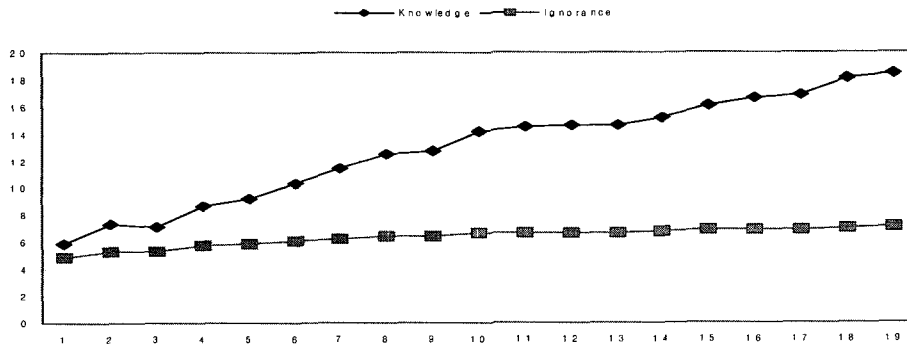


그림 2. 조사지점의 증가에 대한 정보량의 변화

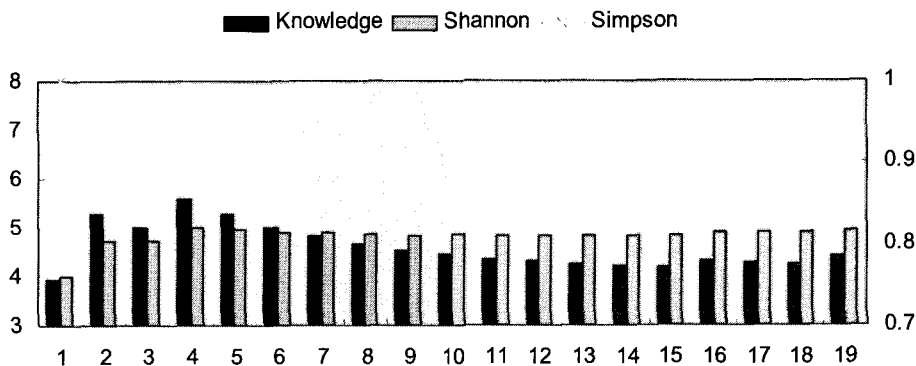


그림 3. 출현종을 40종을 제한한 상태에서 지표변화

기 때문에 40이라는 숫자는 결과를 잘 보여주기 위해 임의로 선정한 값으로 특별한 의미를 두지 않았다.

즉, 이 같은 결과로부터 두 가지 결론에 도달할 수 있다. 첫째, 생태조사를 통해 다양성을 평가하기 위해서는 평가에 필요한 종수를 정해 놓아야 한다는 것과 둘째, 조사는 자료처리를 통하여 조사항목들의 ignorance가 knowledge보다 많아질 때까지 조사가 이루어져야 한다는 것이다.

IV. 결 론

생태계의 다양도를 평가할 수 있는 방법은 Simpson과 Shannon의 지수가 가장 많이 이용되고 있으나, 개체의 출현확률을 고려하여 계산된 결과이며 조사규모를 결정할 수 있는 객관적 기준을 설정하기가 곤란하다.

본 연구에서는 총정보량을 구성하는 지식과 불확실성을 수식으로 전개함으로써 그 물리적 의미를 살펴보고 2004년도 충남 예산군 덕산면 상가천을 대상으로 조사한 식생의 빈도분석 결과에 적용하여 기존의 Simpson과 Shannon의 지수와 비교함으로써 유용성을 보이고자 하였다.

그 결과 Gell-mann이 제시한 지식(Algorithm Complexity)을 정의할 수 있는 증명 가능한 수식을 구성하기 어려웠으며 다만 jung(2003)이 제시한 통계적으로 근사할 수 있는 수식을 통하여 구성할 수 있었다. 이를 이용하여 최초의 조사된 198종을 모두 적용하여 다양성 평가를 수행해 본 결과 추출된 샘플을 설명하기 위한 최소한의 복잡성(Knowledge)는 추출된 샘플이 구성하는 불확실성보다 항상 높게 나왔다.

반면 조사종수를 40종으로 제한하여 다양성을 평가해 본 결과 샘플지점 7을 지나면서부터 샘플이 구성하는 불확실성이 샘플을 설명하기위한 최소한의 복잡성보다 커지게 되어 안정적인 값에 도달함을 알 수 있었다.

본 연구는 그 범위를 총정보량을 생태계조사에 이용할 수 있는지에 한정하였기 때문에 40이라는 숫자는 결과를 잘 보여주기 위해 임의로 선정한 값으로 특별한 의미를 두지 않았다. 다만 현재의 다양성 지표가 지표를 구성하기 위해 필요한 샘플의 수를 결정하지 못한다는 문제점을 지적하였으며 이를 해결하기 위해서는 생태계

조사종수를 한정된 상태에서 조사가 이루어져야 하며 총정보량을 구성하는 지식과 불확실성이 동시에 고려해야 함을 밝혔다.

향후 생태계 다양성 평가에서 지역별 표준조사종이 제안된다면 본 연구에서 개발한 방법을 이용하여 다양성 평가에 필요한 샘플지점의 수를 결정할 수 있을 것으로 기대된다.

참고문헌

1. Crutchfield, J.P., 1994. The Calculi of Emergence: Computation, Dynamics, and Induction. Physica D, Vol. 75 : 11-54.
2. Gell-Mann, M., Lloyd, S., 1996. Information measures, effective complexity, and total Information. Complexity, 2(1) : 44-52.
3. Jung, N.S., 2003, Development of an Information Measure Technique for Rural Facilities Location and Amenities Assessment, Seoul National University. (in Korean)
4. Jung, N.S., 2006, Development of Surveying Tables and Items for Rural Amenity Resources Assessment, Journal of Korean Society of Rural Planning 11(4) : 9-15.
5. Kim, J.G., 2006, Method in Ecology, Bomundang
6. Lee, W.B., Jeon, Y.M., Choi, C.I., 2002, Flora in the Dong River Valley, Korea., Korean J. Limnol. 35(5) : 396-414.
7. Li, M., Vitányi, P., 1997, An Introduction to Kolmogorov Complexity and Its Applications, Springer.
8. Shannon, C.E., 1948, A Mathematical Theory of Communication., The Bell System Technical Journal, Vol. 27 : 379-423.
9. Simpson, E.H., 1949, Measurement of diversity, Nature, pp. 163-688

* 접수일 : 2006년 8월 21일

■ 3인 익명 심사필