

공간자기상관 산출을 위한 인접성 정의 방법 비교

박재문 · 황도현 · 윤홍주[†]
(부경대학교)

A Comparison of Neighborhood Definition Methods for Spatial Autocorrelation

Jae-Moon PARK · Do-Hyun HWANG · Hong-Joo YOON[†]
(Pukyong National University)

Abstract

For the identifying of spatial distribution pattern, Moran's Index(I) which has the range of values from -1 to $+1$ is common method for the spatial autocorrelation measurement. When I is close to 1 , all neighboring features have close to the same value, indicating clustered pattern. Conversely, if the spatial pattern is dispersed, I is close to -1 . And I closing to 0 means spatially random pattern. However, this index equation is influenced by how defining the neighboring features for target feature. To compare and understand the difference of neighborhood definition methods, fixed distance neighboring method and Gabriel Network method were used for I . In this study, these two methods were applied to two marine environments with water quality data. One is Gwangyang Bay which has complex geometric coastal structure located in South Sea of Korea. Another is Uljin area adjacent to open sea located in east coast of Korea. The distances between water quality observed locations were relatively regular in Gwangyang Bay, however, irregular in Uljin area. And for the fixed distance method popular Arc GIS tool was used, but, for the Gabriel Network, Visual Basic program was developed to produce Gabriel Network and calculate Moran's I and its Z -score automatically. According to this experimental results, different spatial pattern was showed differently for some data with using of neighboring definition methods. Therefore there is need to choose neighboring definition method carefully for spatial pattern analysis.

Key words : Spatial autocorrelation, Moran's I , Gabriel network

I. 서론

GIS에서는 공간분석을 이용해 공간패턴을 알 수 있게 한다. 공간 패턴 파악은 데이터 속성의 분포 특성을 보다 잘 이해할 수 있게 하고, 어떤

현상의 모니터링과 서로 다른 피쳐 군을 비교할 수 있게 하며, 공간분포의 변화도 추적가능하게 한다(Mitchell, 2005). 공간분석을 하기 위해서는 통계기법을 이용하게 되는데 일반적인 통계방법은 비 공간 데이터를 대상으로 상관관계를 도출

[†] Corresponding author : 051-629-6656, yoonhj@pknu.ac.kr

* 이 논문은 국토해양부 공간정보 전문인력 양성사업의 지원을 받아 수행된 연구임.

해 내기 때문에 공간적인 특성이 내재된 공간 데이터를 적용하기엔 다소 무리가 따른다. 하지만 공간통계방법은 데이터의 공간 요소를 고려하여 상관관계를 계산한다. 본 논문은 “모든 것은 다른 모든 것들과 관련되어 있지만, 가까운 것이 멀리 있는 것보다 더 관련되어 있다.”라는 지리학의 제 1 법칙(Tobler, 1970)과 같이 속성 값의 분포가 공간에 종속적이라는 개념을 표현한 공간 자기상관(spatial autocorrelation) 통계방법을 이용하였다.

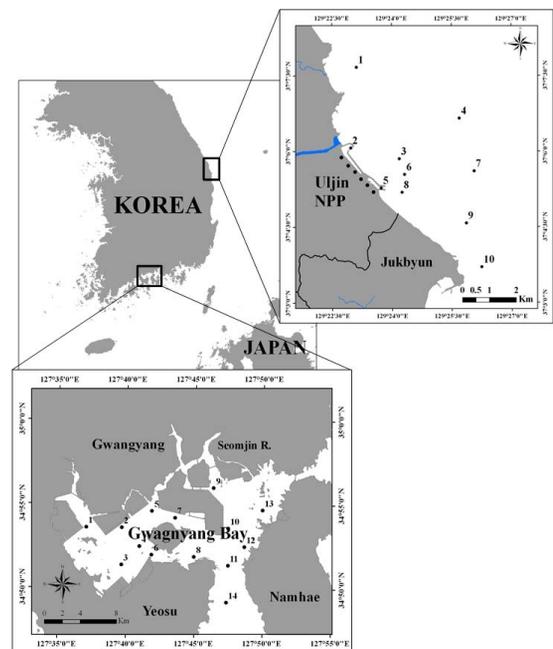
공간자기상관을 파악하기 위한 척도인 Moran's *I*(Moran, 1948; Cliff and Ord, 1973) 값을 사용하여 공간분포 패턴을 정량화 시킬 수 있다. Moran's *I* 값을 사용함에 있어 데이터 정점의 이웃(neighborhood)을 결정하는 것이 중요한 요소 중에 하나이다. 관측 정점들은 각각의 위치와 주변의 지리적인 요소에 대해 서로 영향을 받는다. 각각의 관측 정점과 인접해 있는 이웃 정점과의 관계 정의에 따라 전체적인 상관관계에 영향을 줄 것이다. 따라서 본 연구에서는 정점들 사이에 있는 장애물에 의해 공간 자기상관 지수의 변화에 미치는 영향을 파악할 수 있는 광양만 자료와 정점 간의 불규칙한 거리에 의해 공간 자기상관 지수의 변화에 미치는 영향을 파악할 수 있는 울진 해역 자료를 두 개의 인접성 정의 방법(Gabriel Network, Fixed Distance)을 이용하여 관측 지점들 사이의 관계를 고려한 방법을 Moran's *I* 값을 이용하여 차이점을 비교하였다.

II. 자료 및 방법

1. 연구 지역 및 자료

연구지역은 특별관리 해역으로 지정된 남해의 광양만과 울진원자력발전소 주변해역이다. 광양만은 [Fig. 1]과 같이 해안선이 복잡하고 가운데 섬이 있는 지역이며 연안의 특징을 갖고 있다. 자료는 한국해양연구원에서 2001년 6월부터 2002

년 12월까지 격월로 관측한 표층의 일반 해양수질(pH)과 영양염(Phosphate), 식물플랑크톤 현존량 자료(14개 정점)를 이용하였다(한국해양연구원, 2003). 울진원전 주변해역은 동해안의 전형적인 특징처럼 단순한 해안선을 이루고 있으며 외해의 특징을 갖는다. 자료는 한국전력공사에서 1987년부터 2010년까지 계절별로(년 4회) 모니터링 한 일반 해양수질(DO, Total-N) 데이터(10개 정점)의 2000년부터 2001년까지의 데이터를 이용하였다(한국전력공사, 2000-2001). 두 곳의 연구지역은 지리적으로 뚜렷하게 다른 특징을 갖고 있다. 두 지역 관측 지점의 경위도 좌표를 ArcGIS(ESRI 社) 소프트웨어의 좌표체계 중에서 지리좌표체계(Geographic Coordinate System: GCS)의 UTM-WGS 1984를 기반으로 점(point) 데이터로 만들었다. 그리고 점 데이터를 미터로 된 좌표체인 투영좌표체계(Projected Coordinate System: PCS)의 UTM-WGS 1984 #52로 변환하여 공간분포도를 제작하였다.



[Fig. 1] Study area and observation stations

2. 연구방법

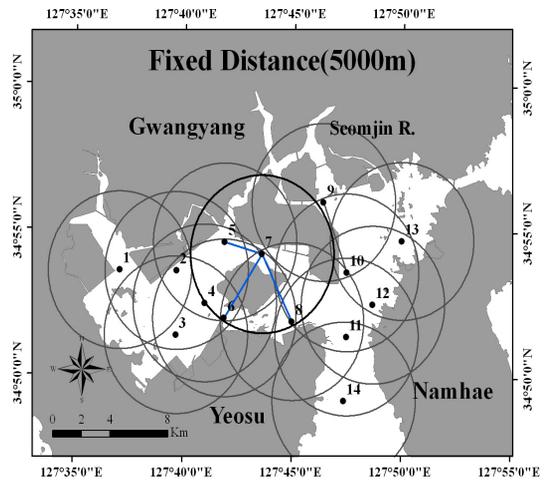
연구지역에서 관측된 데이터를 이용하여 데이터의 군집 패턴을 파악하기 위해 공간자기상관을 파악하였다. 공간 자기상관은 Moran's *I* 값을 산출하여 파악 가능하다. Moran's *I* 값은 정점의 평균과 목표 정점, 목표 정점의 이웃 정점과의 관계를 계산을 통하여 값을 구한다. 일정한 간격으로 관측된 정점을 이용할 경우 한 정점의 인접한 정점을 정의하는 것은 무의미할 수 있다. 하지만 정점 간 불규칙한 거리를 형성하고 연구지역에서 지형적으로 정점들 사이가 막혀 있는 데이터를 이용하여 Moran's *I* 값을 측정하게 되면 정점 간 이웃을 설정하는 방법이 결과에 영향을 미친다. Moran's *I*는 다음과 같이 식 (1)로 구하며 특징은 다음과 같다.

$$I = \frac{n \sum_i \sum_j W_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_i \sum_j W_{ij} \sum_i (x_i - \bar{x})^2} \quad (1)$$

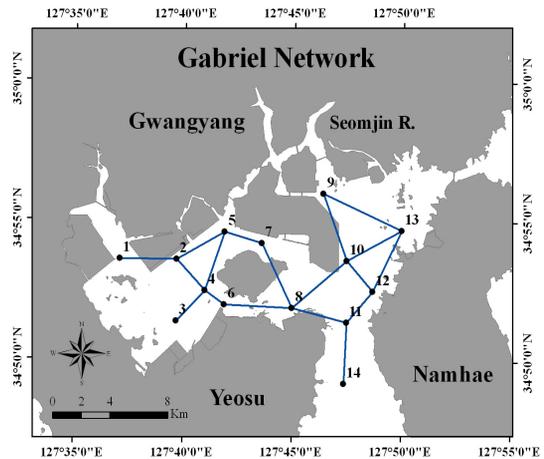
이 식에서 *n*은 관측 정점 수, x_i 는 target, x_j 는 neighborhood, \bar{x} 는 관측 정점의 평균값이다. W_{ij} 는 공간 가중치로 두 정점 *i*와 *j*가 인접하면 1, 인접하지 않으면 0값으로 설정하였다. Moran's *I* 값은 -1부터 +1사이의 값으로 설명할 수 있다. 값이 -1에 가까울수록 데이터의 공간패턴이 분산 패턴(dispersed)을 의미하고, 0에 가까울수록 임의적 패턴(random)을 의미하고, +1에 가까울수록 군집 패턴(clustered)을 의미한다.

본 연구에서는 Moran's *I* 값을 구하기 위한 인접성 정의 방법 중에서 Fixed Distance와 Gabriel Network 두 가지 방법을 사용하였다. Fixed Distance는 [Fig. 2]와 같이 각각의 정점을 기준으로 설정한 거리를 반지름으로 하는 원안에 포함되는 모든 정점을 이웃으로 정의하여 공간자기상관 지수인 Moran's *I* 값을 계산하는 것이다. 광양만에서 관측한 정점에서 인접한 두 정점 간 최대 거리는 약 4700m이고, 울진원전 주변해역에서 인접한 두 정점 사이의 최대 거리는 약 2970m이

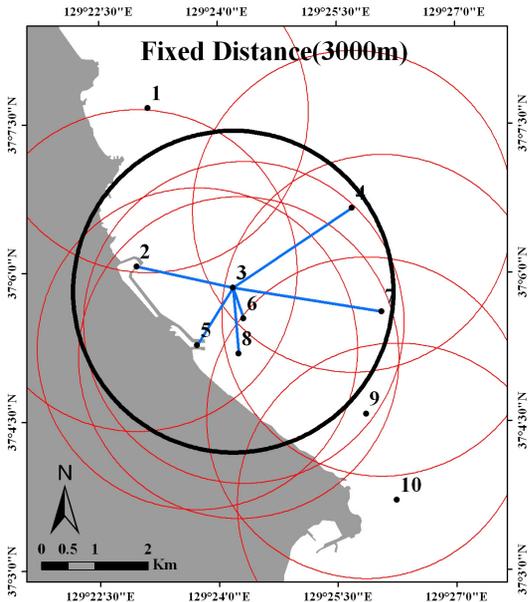
다. 따라서 Fixed Distance를 사용하여 Moran's *I* 값을 구하기 위해 각 정점이 최소한 한 개의 정점을 포함하는 임계 거리는 광양만의 경우 5000m, 울진의 경우 3000m로 정하였다. 이 방법은 ESRI사의 ArcGIS의 tools 중에서 spatial statistics tools의 spatial autocorrelation(Morans *I*)을 사용하였다. 세부설정 중에서 공간관계 설정은 Fixed Distance Band, 거리 경계 값은 임계거리 값을 사용하였다.



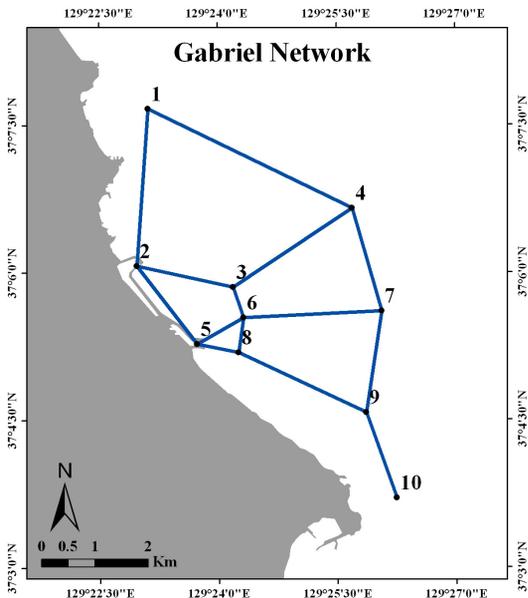
[Fig. 2] Fixed distance method(distance band 5000m) in Gwangyang bay



[Fig. 3] Gabriel network in Gwangyang bay



[Fig. 4] Fixed distance(distance band 3000m) in Uljin

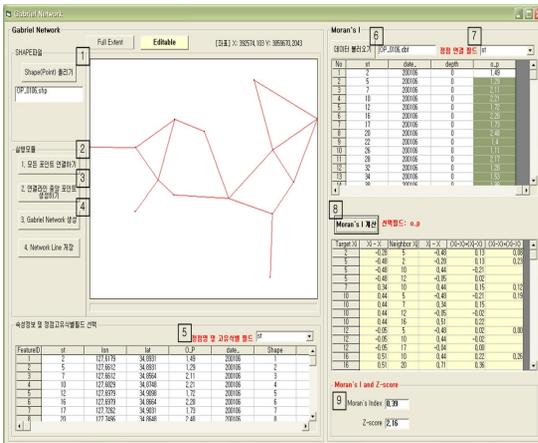


[Fig. 5] Gabriel network in Uljin

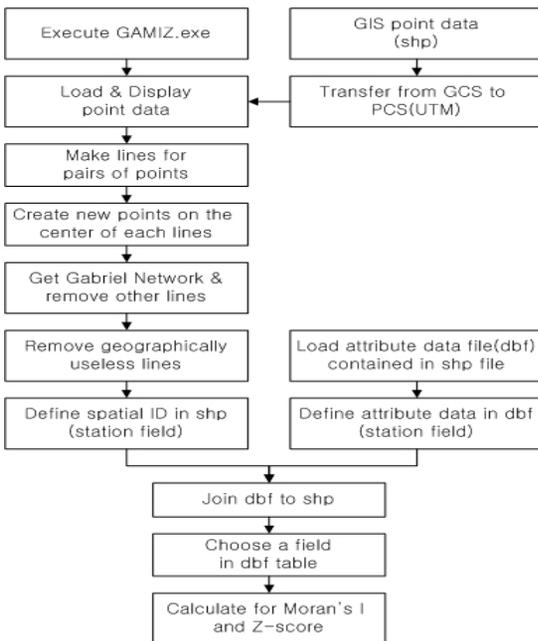
Gabriel Network는 정점 두 개를 한 쌍으로 정하여 두 정점 사이의 거리를 지름으로 하는 원을 만

든다. 이 원안에 두 점 이외의 다른 점이 포함되지 않으면 두 점을 이어서 서로 간의 이웃으로 정한다. 이 과정을 반복하여 network를 제작하였다.

인접성 정의 방법 중 Gabriel Network를 이용하여 Moran's *I* 값을 산출하기 위해서는 우선 정점들 사이에 network를 설정한다. 그 다음, 식 (1)에 각각의 정점(target), 정점과 연결된 이웃의 값 등을 대입하여 수작업으로 Moran's *I* 값을 계산한다. 정점 개수가 많은 경우에는 각각의 정점이 만드는 network가 복잡해져서 network 관계 설정 과정과 Moran's *I* 값을 계산하는 과정에서 실수가 발생할 확률이 높아지게 된다. 그래서 network 관계 설정과 Moran's *I* 값을 계산하는 과정에서 에러를 줄이고 시간을 단축하기 위해 Visual Basic을 이용하여 GAMIZ(Gabriel Network, Moran's *I* and Z-score)라는 실행 프로그램을 제작하였다([Fig. 6]). 이 프로그램은 [Fig. 6]에 표시된 1번부터 9번까지의 단계로 실행된다. [Fig. 6]과 같이 GAMIZ를 실행하면 ArcGIS에서 point 데이터로 만들어진 shape 파일을 불러온다. 이때 point 데이터의 좌표계는 도, 분, 초로 된 경위도 좌표계가 아닌 미터(m)로 된 TM좌표 또는 UTM 좌표로 되어있어야 한다. 그 다음, 두 점을 한 쌍으로 하여 모두 연결한다. 연결된 선을 지름으로 하는 원을 만들기 위해 연결된 모든 선의 중앙에 점을 만든다. Gabriel Network를 만드는 방법을 통해 Gabriel Network를 만든다. point 파일의 식별 ID를 설정한다. 다음, Moran's *I* 값을 구할 데이터 파일을 불러와 point 파일과 연결시킬 식별 ID를 설정했을 때, 일치하면 다음 단계로 넘어가고 일치하지 않으면 에러 메시지가 뜬다. 그리고 데이터 파일에서 Moran's *I* 값을 구할 항목을 선택 한 후 계산 버튼은 누르면 Moran's *I* 값과 유의성 검증에 사용되는 Z-score가 계산되어 표시된다. 그림 7은 GAMIZ의 진행 순서이다.



[Fig. 6] Visual basic program named GAMIZ for gabriel network, Moran's I and Z-score



[Fig. 7] Workflow of GAMIZ

Ⅲ. 결과 및 고찰

1. 공간자기상관 계산 결과

광양만에서 2001년 6월부터 2002년 12월까지 격월로 관측한 일반 해양수질(pH)과 영양염

(Phosphate), 식물플랑크톤 현존량 자료와 울진원 전 주변해역에서 2000년부터 2001년까지 2년 동안 봄, 여름, 가을, 겨울 4계절별로 관측한 해양수질 데이터 중 DO와 Total-N 데이터를 이용하여 Moran's I 값을 구하였다. Moran's I 값과 Z-score를 통한 신뢰구간에 따른 유의성 검증 결과는 <Table 1>, <Table 2>와 같다. 유의성 검증을 위한 Z-score는 유의수준 0.1, 0.05, 0.01에 의해 1.65, 1.96, 2.58이상이면 군집의 Moran's I 값이 유의하다고 볼 수 있다.

아래의 두 표를 보면 Moran's I 값이 전체적으로 방법에 따라 비슷하게 나타나나 부분적으로 차이가 발생하는 것을 볼 수 있다. 이것이 인접성 정의 방법에 의해 값에 영향을 미치는 요인으로 본다. 그리고 두 가지 방법으로 산출된 Moran's I 값의 차이가 0.20 이상 되거나 군집패턴이 차이가 나타나는 경우를 <Table 1>과 <Table 2>에 음영 표시를 하였다.

2. 인접성 정의 방법 비교

본 연구에 이용된 광양만의 경우는 정점들 사이에 섬들이 위치하고 있다(Fig. 1). 지리적 조건을 고려하지 않고 일정한 거리로 이웃을 설정한 Fixed Distance 방법과 지리적 조건을 고려하여 이웃을 설정한 Gabriel Network 방법의 Moran's I 값의 차이는 <Table 3>과 같이 인접성 정의 방법에 의해 정점 개수가 변화하기 때문이다. 예를 들어, 광양만 연구 지역에서 지리적 조건인 섬을 고려하지 않고 정점 6번을 target으로 정하여 경계 거리를 5000m로 한 Fixed Distance를 사용하면 2, 3, 4, 5, 7, 8번 6개 정점이 이웃이 되고, Gabriel Network를 사용하면 4, 7, 8번 3개 정점이 이웃이 되지만 지리적 조건(가운데 위치한 섬)을 고려하여 7번과의 network 선을 지워 4, 8번 2개 정점만 남게 된다. 이로써 지역적으로 관계가 있는 정점들과 Moran's I 값을 계산할 수 있게 된다. 울진지역의 경우는 정점 사이

<Table 1> Moran's I and Z-score of Fixed Distance(FD) and Gabriel Network(GN) in Gwangyang Bay

	200106	200108	200110	200112	200202	200204	200206	200208	200210	200212
pH(FD)	0.49***	0.10[^]	0.16 [^]	0.07[^]	0.65***	-0.14	0.18 [^]	0.30**	0.34**	0.13 [^]
pH(GN)	0.67***	0.78***	0.00	0.30*	0.60***	-0.23	0.13 [^]	0.41**	0.16[^]	0.11 [^]
Phosphate(FD)	0.16[^]	0.01	0.13[^]	-0.08	-0.05	-0.15	-0.11	0.55***	0.31**	0.29**
Phosphate(GN)	0.39**	0.13 [^]	0.30*	0.19[^]	-0.10	0.11 [^]	0.24[^]	0.59***	0.44**	0.34*
Phyto(FD)	ND	ND	ND	ND	0.01 [^]	-0.06	0.83***	0.35**	0.06[^]	0.05 [^]
Phyto(GN)	ND	ND	ND	ND	0.03	-0.16	0.68***	0.14[^]	0.28	0.11 [^]

<Table 2> Moran's I and Z-score of Fixed Distance(FD) and Gabriel Network(GN) in Ulsjin

	200002	200004	200007	200010	200102	200105	200108	200111
DO(FD)	-0.06	-0.13	-0.11	0.01[^]	-0.21 [^]	-0.25[^]	-0.10	0.08[^]
DO(GN)	-0.04	-0.17	-0.48[^]	0.44**	-0.02	-0.08	0.07	0.47**
Total-N(FD)	0.21**	-0.16	-0.05	-0.03	-0.10	0.00[^]	-0.07	-0.30 [^]
Total-N(GN)	0.19 [^]	-0.11	0.09 [^]	-0.08	0.25[^]	0.39**	0.02	-0.27

Z-score : >2.58 ***, >1.96 **, >1.65 *, one tailed >1.65 ^

의 거리가 불규칙하게 배치되어있다([Fig. 1]). 정점 간의 거리를 고려하지 않고 일정한 거리로 이웃을 설정한 Fixed Distance 방법과 지리적 조건을 고려하여 이웃을 설정한 Gabriel Network 방법 사이의 Moran's I 값의 차이는 <Table 4>와 같이 인접성 정의 방법에 의해 정점 개수가 변화하기 때문이다. 예를 들어, 정점 간의 거리를 고려하지 않고 정점 3번을 target으로 정하여 경계 거리를 3000m로 한 Fixed Distance를 사용하면 2, 4, 5, 6, 7, 8번 6개 정점이 이웃이 되고, Gabriel Network를 사용하면 2, 4, 6번 3개 정점이 이웃이 된다. 울진지역의 정점간의 최소거리는 약 610m로 경계거리를 3000m로 설정하게 되면 서로 가까이 있는 정점들은 값의 특징이 없어질 수 있지만 Gabriel Network를 사용하면 값의 특징을 고려할 수 있게 된다. 이렇게 함으로써 불규칙하게 위치하고 있는 정점들과 Moran's I 값을 계산할 수 있게 된다.

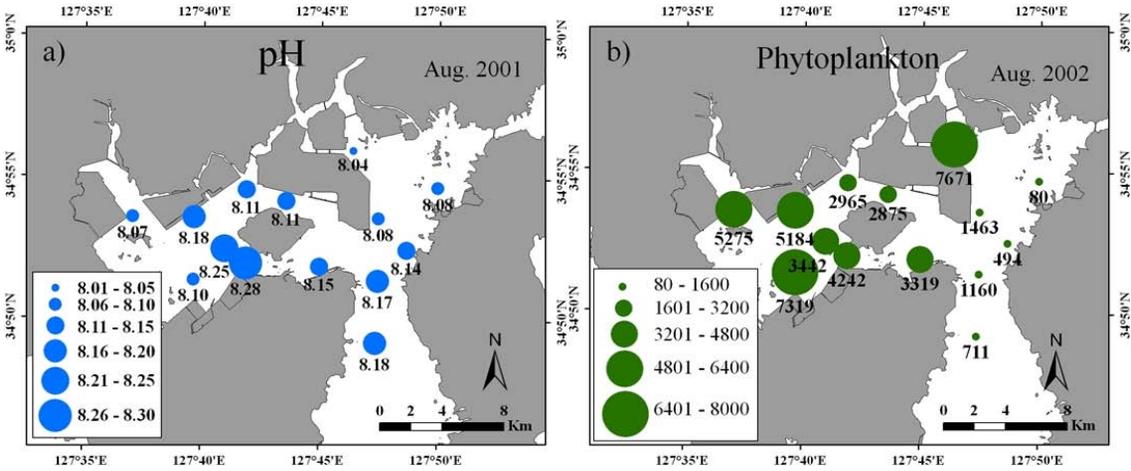
<Table 3> Neighborhood stations of target for neighboring definition method in Gwangyang bay

Target	Neighborhood stations	
	Fixed Distance	Gabriel Network
1	2	2
2	1, 3, 4, 5, 6	1, 4, 5
3	2, 4, 6	4
4	2, 3, 5, 6	2, 3, 5, 6
5	2, 4, 6, 7	2, 4, 7
6	2, 3, 4, 5, 7, 8	4, 8
7	5, 6, 8	5, 8
8	6, 7, 10, 11	6, 10, 11
9	10	10, 13
10	8, 9, 11, 12, 13	8, 9, 12, 13
11	8, 10, 12, 14	8, 12
12	10, 11, 13	10, 11, 13
13	10, 12	9, 10, 12
14	11	11

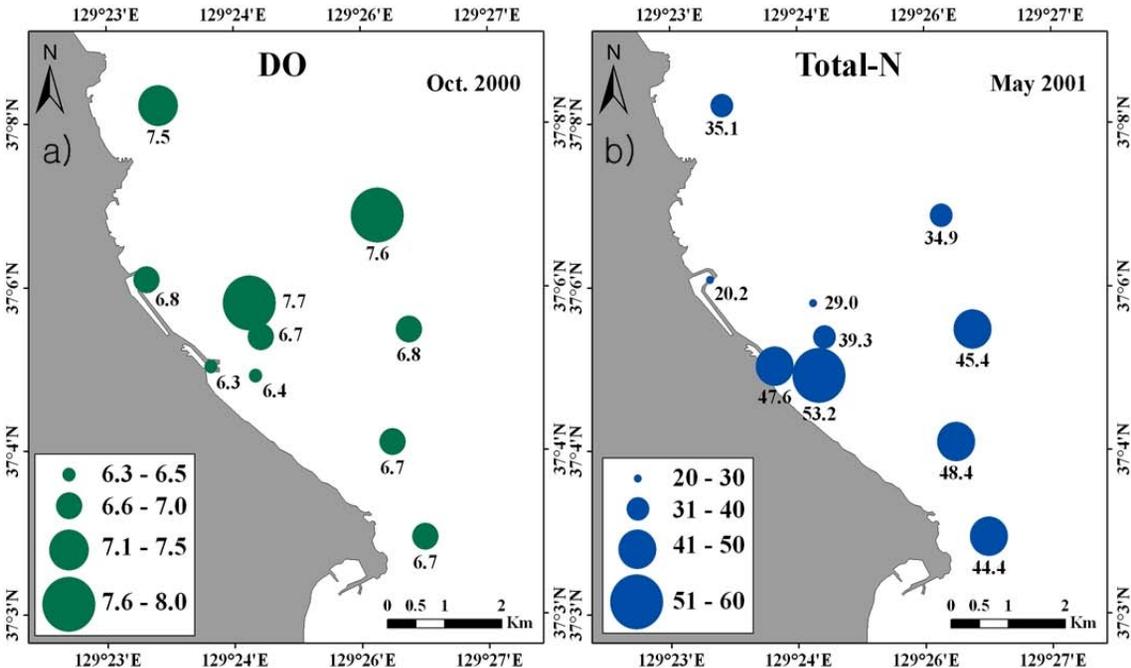
두 가지 방법별 Moran's I 값을 비교한 <Table 1>과 <Table 2>에서 가장 큰 값의 차이와 공간 패턴의 차이를 보이는 자료(2010. 08. pH, 2002.

08. 식물플랑크톤, 2000. 10. DO, 2001. 05. Total-N)를 선택하였다. 그리고 인접성 정의 방법 별 Moran's *I* 값과 점 데이터를 이용하여 제작한 공간분포도를 비교하였다. [Fig. 8-A] 값의 분포는 가운데 부분부터 아래 부분으로 넓게 군집된 형태로 분석할 수 있으며, Moran's *I* 값은 Fixed

Distance 0.03, Gabriel Network 0.78 로 산출되었다. [Fig. 8-B] 값의 분포는 만의 왼쪽과 오른쪽 안에 군집의 형태로 분석할 수 있으며, Fixed Distance는 0.35, Gabriel Network는 0.14로 Moran's *I* 값이 나왔다. [Fig. 9-A] 값의 분포는



[Fig. 8] Spatial distribution map of pH(a) and Phytoplankton(b) in Gwangyang bay



[Fig. 9] Spatial distribution map of DO(a) and Total-N(b) in Ulsan

<Table 4> Neighborhood stations of target for neighboring definition method in Uljin

Target	Neighborhood features	
	Fixed Distance	Gabriel Network
1	2	2, 4
2	1, 3, 5, 6, 8	1, 3, 5
3	2, 4, 5, 6, 7, 8	2, 4, 6
4	3, 6, 7	1, 3, 7
5	2, 3, 6, 8	2, 6, 8
6	2, 3, 4, 5, 7, 8, 9	3, 5, 7, 8
7	3, 4, 6, 8, 9	4, 6, 8, 9
8	2, 3, 5, 6, 7, 9	5, 6, 9
9	6, 10, 7, 8	8, 7, 10
10	9	9

위와 아래가 분리된 군집 형태로 분석할 수 있으며, Fixed Distance는 0.01, Gabriel Network는 0.44 인 Moran's I 값이 나왔다. [Fig. 9-B] 값의 분포는 위에는 작은 값, 아래에 큰 값이 군집을 이루고 있는 형태로 분석할 수 있으며, Moran's I 값은 Fixed Distance 0.00, Gabriel Network 0.39 로 산출되었다. 그러므로 복잡한 또는 단순한 해안선의 형태와 상관없이 Gabriel Network 방법이 Fixed Distance 방법보다 공간 자기상관을 더 잘 나타내었다.

IV. 결론

본 연구는 지리적 특징과 정점 간 불규칙한 거리와 같은 특별한 형태의 해양환경데이터를 이용하여 해양환경의 공간분포 특성을 파악하기 위해 공간자기상관을 이용하였다. 공간자기상관을 파악하기 위한 Moran's I 값을 데이터의 특징을 고려할 수 있는 인접성 정의 방법을 통해 구하였으며, 그 결과는 다음과 같다.

해양환경의 공간분포를 분석하기 위해 공간통계를 사용한다. 해양환경 데이터의 특성상 상호간 독립적인 것이 아니라 지리학 제 1법칙에 따라 가까운 것이 멀리 있는 것보다 더 영향을 미치게 되므로 정점에 대해 인접성을 정의하는 방

법은 중요한 요소 중에 하나이다.

해양에서 관측된 데이터의 정점들이 균일한 간격으로 위치하기 않고, 지리적인 영향에 의해 분리될 수 있다. 이런 경우, 공간분포를 파악하는 지수인 Moran's I 는 인접성 정의 방법에 의해 값이 영향을 받는다. 인접성 정의 방법으로 Gabriel Network와 Fixed Distance를 사용하였다.

두 가지 방법을 이용하여 Moran's I 값과 Z-score를 구하여 공간 군집 패턴을 파악하고 유의성을 검증하였다. Gabriel Network와 Fixed Distance로 측정된 Moran's I 값을 비교하여 차이가 약 0.20 이상 되고 군집패턴의 차이가 보이는 것을 중심으로 연구하였다. 연구결과와 경계 거리를 설정한 거리 안에 포함되는 모든 정점을 이웃으로 설정하는 Fixed Distance 방법과 지리적 조건과 정점 간 거리를 고려한 Gabriel Network 방법으로 산출된 Moran's I 값을 공간분포도와 비교하면 두 개의 방법은 몇 개의 자료의 Moran's I 값에서 큰 차이를 보이며 후자의 방법으로 산출된 Moran's I 값이 관측 정점들의 공간적인 분포 특성을 좀 더 정확하게 설명하였다.

참고 문헌

한국전력공사 전력연구원(2001). 울진원자력 발전소 주변 일반 환경조사 및 평가보고서(2000년 보), 332.
 한국전력공사 전력연구원(2002). 울진원자력 발전소 주변 일반 환경조사 및 평가보고서(2001년 보), 275.
 한국해양연구원(2003). 남해 특별관리 해역의 환경오염 관리모델 연구(1) 광양만 중심 연구, BSPE 836-001577-7, 823.
 Amaud, J.F., L. Madec, A. Guller and A. Bellido(2001). Spatial analysis of allozyme and microsatellite DNA polymorphisms in the land snail *Helix aspersa* (Gastropoda: Helicidae), *Molecular Ecology* 10:1563~1576.
 Choi, H.W., K.H. Kim and C.Y. Lee(2007). Application of spatila autocorrelation for the

- spatial distribution pattern analysis of marine environment, Journal of the Korean Association of Geographic Information Studies 10:60~74.
- Cliff, A.D. and J.K. Ord(1973). Spatial autocorrelation, London Pion.
- Gabriel, K.R. and R.R. Sokal(1969). A new statistical approach to geographic variation analysis, Systematic Zoology 18:259~270.
- Legendre, P. and L. Legendre(1998). Developments in environmental modeling, Numerical ecology.
- Moran, P.A.P(1948). The international of statistical maps, *J. Roy. Stat. Soc. Ser. B (methological)* 10:243~251.
- Schweiger, O., M. Frenzel and W. Durka(2004). Spatial genetic structure in a metapopulation of the land snail *Cepaea nemoralis*(Gastropoda: Helicidae), *Molecular Ecology* 13:3645~3655.
- Smith, M.J, M.F Goodchild and P.A Longley (2007). Geospatial Analysis(A Comprehensive Guide to Principles, Techniques and Software Tools), Winchelsea Press 394.
-
- 논문접수일 : 2011년 06월 19일
 - 심사완료일 : 1차 - 2011년 07월 15일
2차 - 2011년 08월 03일
 - 게재확정일 : 2011년 08월 14일