

갯벌 미생물 다양성 모니터링 시료 채취 개수 및 간격 선정을 위한 지구통계학적 기법과 데이터 마이닝 적용 연구

Application of Geo-Statistic and Data-Mining for Determining Sampling Number and Interval for Monitoring Microbial Diversity in Tidal Mudflat

양지훈 · 이재진 · 유근제 · 박준홍[†]

Yang, Jihoon · Lee, Jaejin · Yoo, Keunje · Park, Joonhong[†]

연세대학교 사회환경시스템공학부

School of Civil and Environmental Engineering, Yonsei University

(2010년 11월 10일 접수, 2010년 12월 16일 채택)

Abstract : Tidal mudflat is a reservoir for diverse microbial resources. Microbial diversity in tidal mudflat sediment can be easily influenced by various human activities. It is necessary to take representative samples to monitor microbial diversity in tidal mudflat sediments. In this study, we analyzed the microbial diversity and chemical characteristics of vegetation and non-vegetation tidal mudflat regions in the Kangwha tidal mudflat using geo-statistics and data-mining. According to the geo-statistical analysis, most correlation range values for the vegetation region were smaller than those for the non-vegetation region, which suggested that the shorter number and interval of sampling are required for the vegetation tidal mudflat environment due to its higher degree of chemical and biological complexity and heterogeneity. The data-mining analysis suggested that the organic content and nitrate were the major environmental factors influencing microbial diversity in the vegetation region while pH and sulfate were the major influencing factors in the non-vegetation region. Using the geo-statistical and data-mining integration approach, we proposed a guideline for determining the sampling interval and number to monitor microbial diversity in tidal mudflat.

Key Words : Tidal Mudflat, Microbial Diversity, Number of Sample, Geo-statistic, Data-mining.

요약 : 갯벌 퇴적토는 미생물 다양성이 매우 높다고 알려져 있다. 하지만 인위적인 교란에 의해 갯벌 퇴적토 내 미생물 다양성이 달라질 수 있다. 지속적인 퇴적토 내 미생물 다양성 모니터링을 위해서는 대표성을 지닌 시료의 채취가 중요하다. 본 연구에서는 강화도 여차리 갯벌을 대상으로 식생이 있는 지역과 식생이 없는 지역의 미생물 다양성과 이화학적 특성치를 지구통계학적으로 비교분석 하였다. 갯벌 시료에서 측정된 미생물 다양성과 다양한 이화학적 특성치를 지구통계학적으로 분석한 결과, 식생이 존재하는 지역에서의 상관거리가 식생이 존재하지 않는 지역에 비하여 대체로 짧은 것을 알 수 있었다. 이는 식생이 존재하는 지역의 높은 생태학적 및 이화학적 복잡성과 이질성으로 인해 식생이 존재하는 지역에서는 식생이 존재하지 않는 지역에서 보다 비교적 좁은 간격으로 시료를 채취해야 한다는 것을 의미한다. 데이터 마이닝 기법을 사용하여 미생물 다양성에 영향을 주는 주요 환경영향 인자를 분석한 결과, 식생이 존재하는 지역에서는 유기물 함량과 질산염이온, 식생이 존재하지 않는 지역에서는 pH와 황산염이온이 미생물 다양성에 영향을 끼친다는 것을 알 수 있었다. 이러한 지구통계학 및 데이터 마이닝 분석 결과들을 활용해서 갯벌 퇴적토 내 미생물 다양성 측정을 위한 시료 채취 간격 및 개수 선정 지침을 본 연구에서 제안하였다.

주제어 : 갯벌, 미생물 다양성, 시료채취개수, 지구통계학, 데이터 마이닝

1. 서론

갯벌 지역은 만조 때 해수에 잠기고 간조 때 육지처럼 드러나면서 바다 속 해저가 하루에 두 번씩 육지가 되는 환경 조건의 변화가 큰 지역이다.¹⁾ 갯벌이 잘 발달하기 위해서는 조수간만의 차이가 크고, 얕은 바다에 평평한 바닥이 넓게 분포되어 있어야 하며, 바다로 퇴적물이 지속적으로 공급되어야 한다. 이와 같은 갯벌의 형성 조건으로 인해 갯벌은 육상생태계에서 유래된 유기 물질이 해양 생태계로 진입하는 마지막 관문으로 알려져 있다. 또한 갯벌은 두 생태계 사이에 있는 전이 지대로써 퇴적토의 입도나 식생 유·무에 따라 다양한 특성을 지니고 있기 때문에, 미생물 다양성이 매우 높다고 알려져 있다.²⁾ 갯벌의 높은 미생물 다양

성은 갯벌 퇴적토가 다양한 생태학적 기능을 보유하고 있음을 의미한다.³⁾ 따라서 미생물 다양성이 높은 갯벌은 생태학적으로 건강하다고 말할 수 있다. 하지만 해양 유류 유출 사고, 무단 방류된 유해물질 유입 등의 인위적인 교란에 의해 갯벌 퇴적토 내 미생물 종의 다양성 및 그 기능 발현 여부가 상이해 질 수 있다. 그러므로 지속적인 퇴적토 내 미생물 다양성 모니터링을 수행하여 현재 갯벌의 생태학적 상태를 확인할 필요가 있다. 갯벌의 생태학적 중요성은 이미 많은 국·내외 연구진들이 인식하고 있으며 갯벌 퇴적토의 미생물 다양성을 모니터링하기 위한 선행연구가 이루어졌다.⁴⁻⁷⁾ 갯벌 미생물 다양성 모니터링에서 대표성 있는 시료 채취의 개수는 모니터링 비용 산정에 결정적 요인이다. 그럼에도 불구하고 갯벌 퇴적토의 시료 채취 간격 결정을 위

[†] Corresponding author E-mail: parkj@yonsei.ac.kr Tel: 02-2123-5798 Fax: 02-346-5300

한 체계적인 기초 연구가 미미하다.⁸⁻¹⁰⁾ 육상 토양 시료 채취 개수와 간격에 대한 환경부 지침은 제시되어 있는 반면,¹¹⁾ 갯벌 시료 채취에 대한 국내 연구는 없다. 또한 미국 EPA의 경우 식생이 있는 갯벌에서는 그 생태적 복잡성(Complexity)으로 인하여 식생이 없는 지역에서보다 상대적으로 더 많은 시료를 채취해야 한다는 정성적인 지침은 제시하고 있으나, 실증적인 결과에 근거하여 시료 채취 간격과 개수를 산정하는 구체적인 지침이나 방법에 대한 제시는 없다.¹²⁾

본 연구에서는 갯벌 미생물 다양성 모니터링을 위한 대표성을 지닌 시료의 채취를 위하여 시료 채취 간격 및 개수를 선정하는 지침을 제시하려는 취지로 국내 갯벌을 대상으로 식생이 없는 지역과 식생이 있는 지역의 미생물 다양성과 이화학적 특성치를 측정하였다. 측정값을 사용하여 지구통계학적 비교분석을 수행하고, 데이터 마이닝 분석을 이용해서 미생물 다양성에 영향을 미치는 환경인자를 도출하여, 지구통계학과 데이터 마이닝을 접목한 갯벌 미생물 다양성 모니터링을 위한 시료 채취 간격 및 개수를 산정하는 지침을 제안한다.

2. 연구방법

2.1. 연구 지역 및 시료 채취

갯벌 퇴적토는 2010년 1월 인천광역시 강화군 화도면 여차리에 위치한 갯벌에서 채취하였다. 여차리 갯벌은 조수 간만의 차이가 4 m 이상인 대조차 환경의 갯벌이며, 외해를 향하여 열린 상태로 발달한 개방형 갯벌이다. 대상 지역은 연안에 설치된 방조제를 따라 좁은 면적의 갈대 군락이 형성되어 있으며, 시료는 다양한 환경 조건을 고려하기 위해 갈대 군락지역과 갯벌 지역에서 각각 49개씩 채취되었다. 또한 파랑의 수평·수직 방향 시료를 모두 확보하기 위하여 Fig. 1(b)와 같이 다양한 방향으로 총 98개의 시료를 채취하였다. 시료는 갯벌 표면으로부터 5~10 cm 깊이에서 채취하였으며 채취 후 50 mL falcon conical tube에 옮겨 담아 혐기성 조건에서 4℃를 유지하여 실험실까지 운반하였으며 시료 분석 전까지 같은 상태를 유지하였다.

2.2. 시료 특성 분석

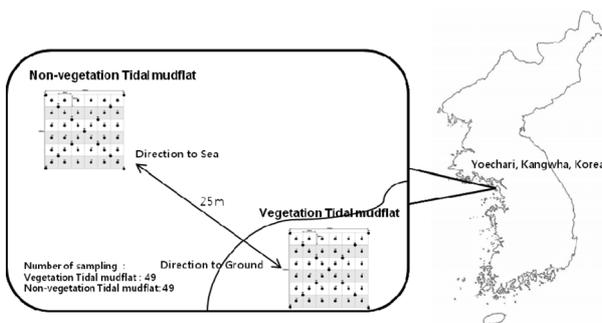


Fig. 1(a). Sampling site.

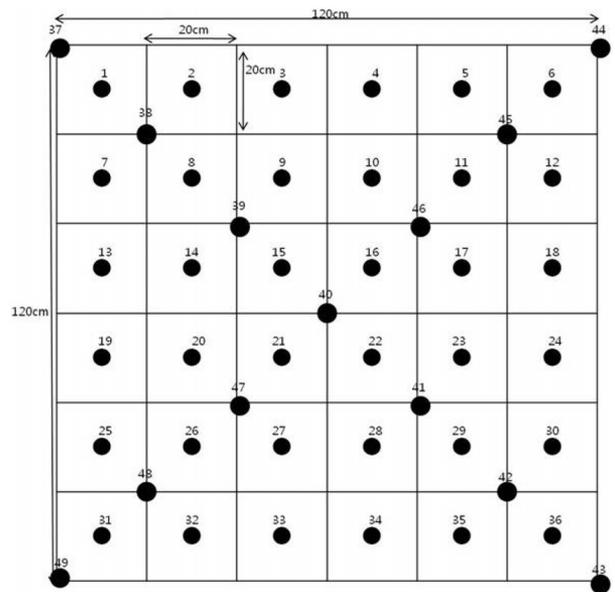


Fig. 1(b). Sampling method for tidal mudflat soil.

2.2.1. 이화학적 분석

시료의 공간별 이화학적 특성을 확인하기 위하여 pH, 수분 함량, 유기물 함량, 음이온 농도를 측정하였다. pH는 퇴적토와 증류수를 1:2(g/mL)의 비율로 혼합하여 만든 슬러리를 원심분리 한 후, 상등액을 취하여 pH meter (Orion 3Star, Thermo)로 측정하였다. 수분 함량은 초기 질량과 105℃에서 24시간 건조시킨 후 질량을 비교하여 측정하였다. 유기물 함량은 강열감량 시험법을 이용하여 600℃에서 24시간 강열한 후의 질량을 이용하여 측정하였다. 음이온 농도는 퇴적토와 증류수를 1:2(g/mL)의 비율로 혼합하여 만든 슬러리를 원심분리 한 후, 0.2 μm nylon 필터로 거른 상등액을 이온크로마토그래피(ICS-2000, DIONEX)로 측정하였다.

2.2.2. 미생물 다양성 생물학적 분석

시료의 공간별 생물학적 특성을 확인하기 위하여 T-RFLP (Terminal Restriction Fragment Length Polymorphism) 분석을 수행하였다. 채취한 퇴적토 시료로부터 genomic DNA를 추출한 뒤(PowerSoil® DNA Isolation Kit, MO-BIO), 6-FAM-labeled 27F, 1492R primer set을 이용하여 16S rRNA 유전자를 대상으로 PCR (PTC-1148, BIO-RAD)을 수행하였다. PCR 산물을 DNA purification kit (QIAGEN, USA)를 사용하여 정제한 후, 제한 효소 Hha I를 이용하여 염기서열의 특정 부분을 절단하였으며 DNA 절편을 에탄올 침전법으로 농축 및 정제 한 후, T-RF sequencer를 이용하여 분석하였다.¹³⁾ 획득한 T-RF pattern을 T-RFLP Phylogenetic Assignment Tool (<http://secure.limnology.wisc.edu/trflp>)로 분석하여 미생물 다양성 지수(Shannon Index, H')를 산출하였다.

2.3. 지구통계학 분석

시료의 특성을 나타내는 측정값들 간의 공간적인 상관성을 분석하기 위하여 지구통계학 기법을 사용하였다. 일반

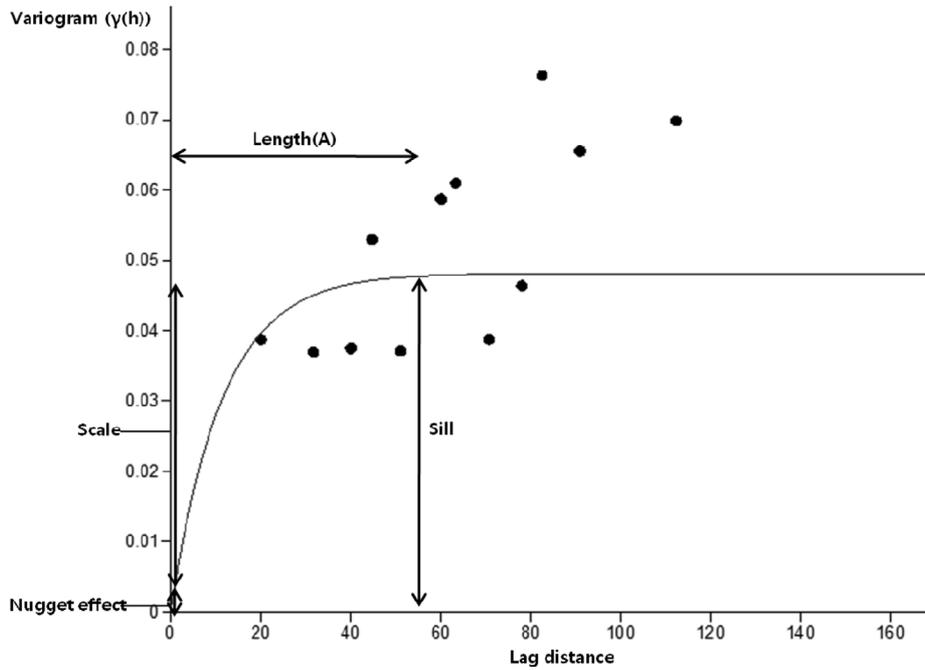


Fig. 2. Conceptual diagram of geostatistics.

적으로 지구통계학 기법은 측정값을 통해 미지의 값을 예측하는 데 사용된다.¹⁴⁾ 하지만 본 연구에서는 일정 거리 간격으로 측정된 퇴적토 특성의 반베리오그램(semivariogram) 값, $\gamma(h)$ 를 식 (1)과 같이 분석하여 시료 특성의 공간적 자기상관성이 존재하는 상관거리를 분석하는데 사용하였다.

$$r(h) = \frac{1}{2n} \sum_{i=1}^n (Z(X_i) - Z(X_{i+h}))^2 \quad (1)$$

여기서 n은 전체 시료의 개수를, X_i , X_{i+h} 는 시료의 위치를 의미한다. $Z(X_i)$, $Z(X_{i+h})$ 는 각 위치에서의 퇴적토 특성의 측정값을 나타낸다. 반베리오그램은 Surfer 8.0 (Golden software, USA) 프로그램을 사용하여 분석하였으며, 본 연구에서 사용한 각 변수 값이 가장 잘 구분되는 모델인 지수 모델을 사용하였다. 반베리오그램의 대표도는 Fig. 2와 같다. Fig. 2의 가로축은 시료간의 간격(분리거리), 문턱값(Sill)은 분리거리가 증가하여 일정한 거리가 되면 자료 값들간의 유효관계가 나타내지 않게 되어 반베리오그램이 일정한 값을 나타내게 되는데, 이 때의 반베리오그램 값을 의미한다. 또한 Length (A)는 상관거리로써 문턱값이 나타날 때의 최대 분리거리를 의미한다. Scale과 Nugget은 각각 실제 반베리오그램 값, 분리거리가 0일 때 분석되는 반베리오그램 값(짧은 분리거리에서 자료의 불확실성)으로써 Scale과 Nugget 값의 합은 Sill과 같다. 본 연구에서는 이와 같이 얻어진 상관거리를 시료 채취 간격 결정에 사용하였다.

2.4. 데이터 마이닝 분석

갯벌 퇴적토 내 미생물 다양성을 결정하는 주요환경인자를 분석하기 위하여 데이터 마이닝 기법인 의사결정나무

(Decision Tree) 분석을 수행하였다. 의사결정나무는 비모수적 방법(Nonparametric method)으로써 가정사항이나 모형에 대해 제약을 받지 않기 때문에 어떠한 자료의 형태나 구조에서도 사용이 가능하며, tree 구조로 모형이 표현되어 해석을 쉽게 할 수 있기 때문에 본 연구와 같이 갯벌 미생물 다양성과 환경인자들간 상관관계에 대한 선행지식이 없는 경우에 사용하기 적당하다. 또한 의사결정나무 분석은 이상치(Outlier)나 결측치(Missing value)에 영향을 받지 않으며 목표변수(Target variable)에 영향을 미치는 설명변수(Input variable)의 상호관계를 파악할 수 있는 장점을 가지고 있다. 그리고 의사결정나무는 데이터 마이닝 방법 중에서 상대적으로 데이터 수가 적어도 모형화가 가능하며, 분류성능도 우수한 것으로 알려져 있다.¹⁵⁾ 의사결정나무 분석은 SAS Enterprise Miner v9.1를 사용하여 수행하였다.

일반적으로 환경 내에서 미생물 다양성을 결정하는 인자는 Na^+ , Mg^{2+} 등의 양이온 농도, PO_4^{3-} , NO_3^- 등의 음이온 농도, 수분 함량, 유기물 함량, pH, 온도 등으로 알려져 있다.¹⁶⁾ 하지만 본 연구에서는 퇴적토를 대상으로 하였기 때문에 토양 내 토착 미생물의 다양성을 예측하는데 좋은 지표라 알려진 pH와 수분 함량을 미생물 다양성 결정인자로 고려하였다.^{17,18)} 또한 퇴적토 내 식생 존재 유·무에 따라 발생하는 유기물 농도의 차이가 미생물 다양성에 미치는 영향을 비교하기 위하여 유기물 함량을 미생물 다양성 결정인자로 고려하였다.¹⁹⁾ 갯벌 퇴적토는 조력에 의하여 하루에 두 번 호기성 및 혐기성 환경에 교차 노출 되는 특성을 가지고 있다. 퇴적토가 혐기성 조건을 유지할 경우, 황환원 작용이 일어나기 때문에 황산염이온(SO_4^{2-})의 농도가 미생물 다양성에 영향을 미칠 것으로 예상할 수 있다.²⁰⁾ 반면에 퇴적토가 호기성 조건을 유지할 경우, 질산화 작용이 일어

나기 때문에 질산염이온(NO₃⁻)의 농도가 미생물 다양성에 영향을 미칠 것으로 예상할 수 있다.²¹⁾ 따라서 본 연구에서는 갯벌 퇴적토의 미생물 다양성 결정인자로 황산염이온, 질산염이온을 선정하였다. 또한 갯벌 퇴적토는 육상 토양에 비하여 긴 시간 동안 해수에 노출되기 때문에 염소이온(Cl⁻) 이 미생물 다양성에 영향을 미칠 것이라 예상하여 염소이온을 미생물 다양성 결정인자로 고려하였다. 본 연구에서는 이와 같이 선정된 6개의 미생물 다양성 결정인자들(pH, 수분 함량, 유기물 함량, SO₄²⁻, NO₃⁻, Cl⁻)을 input variable로 사용하였으며, 미생물다양성지수(Shannon Index, H')를 target variable로 사용하였다. Training set(70%)과 testing set(30%)으로 자료 구분(Data Partition)을 하여, training set데이터는 목표변수를 예측하는 알고리즘 구축을 위하여 사용되고, testing set데이터는 학습된 예측 모델이 어느 정도의 예측능이 있는지를 평가하는데 활용되었다. 일반적으로 모형의 분류성능 평가는 testing set데이터와 예측치 간의 분류성능 결과를 이용하여 결정한다. 본 연구에서는 구축된 모형의 재현성을 확보하기 위하여 목표변수와 설명변수를 무작위(Random)방법으로 10번씩 training set과 testing set으로 구분하여 학습하였고, 학습에서 얻어진 모형의 성능은 testing set의 분류성능 값들의 평균으로 결정하였다.²²⁾

3. 결과 및 고찰

3.1. 식생 유·무에 따른 이화학적 특성 및 미생물 다양성 분석 결과

식생이 존재하지 않는 갯벌과 식생이 존재하는 갯벌에서의 퇴적토 시료의 이화학적 특성 및 미생물 다양성 분석 결과는 Table 1과 같다.

식생 유·무에 따라 측정값에서 10% 이상의 큰 차이를 보인 것은 유기물 함량, 질산염이온(NO₃⁻), 염소이온(Cl⁻) 그리고 미생물 다양성 지수였고, 식생 유·무에 관계없이 적은 차이를 보인 것은 pH, 수분함량, 황산염이온(SO₄²⁻)이었다. 식생 유·무에 따라 10% 이상 차이를 보인 항목들 중, 질산염이온은 식생이 존재하는 갯벌에서 낮은 농도를 보인 반면, 유기물함량, 염소이온 그리고 미생물 다양성 지수는 식생이 존재하지 않는 지역에서 높게 나타났다. 이와 같이

Table 1. Measured characteristic properties of tidal mudflat sediments from vegetation and non-vegetation area

Characteristic properties	Vegetation area	Non-vegetation area
pH	6.16(±0.21)	6.21(±0.14)
Moisture content (%)	30.20(±2.33)	30.00(±1.04)
Organic content (%)	2.67(±0.55)	2.35(±0.20)
NO ₃ ⁻ (mg/L)	0.04(±0.02)	0.15(±0.11)
SO ₄ ²⁻ (mg/L)	26.60(±4.91)	25.15(±5.09)
Cl ⁻ (mg/L)	0.30(±0.17)	0.20(±0.12)
Shannon Index (H')	4.09(±0.72)	2.76(±0.61)

퇴적토의 이화학적 특성 및 미생물 다양성은 일정한 경향이 없기 때문에, 어느 특성을 기준으로 시료 채취 간격 및 개수를 결정해야 하는지와 어떠한 이화학적 특성이 미생물 다양성에 영향을 미치는지 명확하게 파악하는 것이 불가능했다. 따라서 퇴적토 특성의 공간적인 상관성을 분석하여 상관거리를 도출할 수 있는 지구통계학 분석을 수행하였다.

3.2. 지구통계학 분석 결과

이화학적 특성과 미생물 다양성 지수(Shannon Index, H') 측정값을 input variable data로 사용한 지구통계학 분석 결과는 Table 2와 같다.

대체적으로 식생이 존재하는 지역에서의 상관거리가 식생이 존재하지 않는 지역에서의 상관거리보다 짧은 경향을 확인할 수 있었다. 상관거리가 짧다는 점은 퇴적토 특성 간의 공간적인 상관성이 비교적 낮다는 것을 의미하며, 이로부터 식생이 존재하는 지역의 비균질성은 식생이 존재하지 않는 지역에 비하여 높다는 것을 추론할 수 있었다. 또한 식생이 존재하는 지역의 시료를 채취하는 경우 조금 더 짧은 간격으로 시료를 채취해야 한다는 것을 추론 할 수 있었다. 이러한 결론은 미국 EPA의 갯벌 미생물 다양성 모니터링 지침과 동일하나, 본 연구에서는 특정 갯벌 지역을 대상으로 실증적인 자료에 근거하여 구체적이고 정량적인 비교가 수행된 것이 차별적이다.¹²⁾

상관거리 분석에서는 질산염이온이 식생이 존재하는 지역 및 존재하지 않는 지역 모두에서 가장 긴 상관거리를 나타내었으며, 유기물 함량의 상관거리는 양 쪽 모두에서 가장 짧은 상관거리를 나타내었다. 유기물 함량의 상관거리가 짧은 것은 갯벌 퇴적토의 비균질성이라는 특성으로 설명이 가능하다. 즉 갯벌 퇴적토에 고정된 유기물 함량은 갯벌 퇴적토 비균질성의 영향으로 거리 별로 차이가 많은 것으로 추론할 수 있다.²³⁾ 반면 퇴적토 내에 용존되어 있는 이온들의 상관거리는 다른 인자들에 비해 상대적으로 긴데, 이 중에서도 질산염이온의 상관거리가 염소이온이나 황산염이온의 상관거리보다 긴 것으로 확인되었다. 이는 갯벌 퇴적토가 해수에 잠기는 기간 동안 혐기성 상태를 유지하게 되어 암모니아나 유기 질소를 생성하지만, 갯벌 퇴적토가 공기 중으로 노출되는 시간 동안에는 산소의 공급이

Table 2. Geostatistic analysis results for vegetation and non-vegetation tidal mudflat samples

	Vegetation area	Non-vegetation area
Characteristic properties	Correlation range (cm)	Correlation range (cm)
pH	15.01	66.37
Moisture content	19.67	611.38
Organic content	1.0	2.54
NO ₃ ⁻	1,034	3,454
SO ₄ ²⁻	1.25	8.04
Cl ⁻	48.63	68.76
Shannon Index (H')	1.0	21.87

원활하여 호기성 상태에서 질산화가 활발히 일어나기 때문이라 사료된다.²⁴⁾ 식생이 존재하는 갯벌 지역 미생물 다양성 상관거리의 경우, 식생이 없는 지역에서 보다 짧은 것을 확인할 수 있었다. 이로부터 식생이 존재하는 지역은 식생이 존재하지 않는 지역에 비해 유기물 함량이 높을 뿐만 아니라 최종전자수용체로 사용할 수 있는 음이온들이 많이 존재하기 때문에 다양한 종의 미생물이 존재할 수 있다는 것을 유추할 수 있다. 또한 식생의 존재로 인해 높아진 퇴적토 성상의 비균질성 역시 식생이 없는 지역에 비하여 높다고 해석할 수 있다. 이러한 공간적 비균질성은 미생물 종의 다양성을 풍부하게 할 수 있다.²⁵⁾

지구통계학 분석을 통하여 획득한 상관거리를 시료 채취 간격에 사용하여 보았으나, 각 환경인자 별로 상관거리가 서로 상이하여 어떠한 상관거리를 선택하는 것이 합리적인지가 불분명하였다. 또한 이화학적 특성 분석에서 추론한 바와 같이 퇴적토 환경 속에는 미생물의 다양성에 영향을 끼칠 가능성이 있는 인자와 그렇지 않은 인자가 함께 존재하기 때문에 미생물 다양성을 결정하는 주요 환경 인자를 도출하기 위하여 아래와 같이 데이터 마이닝 분석을 실시하였다.

3.3. 데이터 마이닝 분석에 의한 주요 환경 인자 도출

모형의 분류성능 평가는 testing set의 결과를 이용해 결정하였다. 본 연구에서 구축된 모형의 신뢰성을 확보하기 위하여 목표변수와 설명변수를 무작위 방법으로 10번씩 training set과 testing set으로 구분하여 학습한 결과, 분류성능은 40% 정도, 표준편차는 10%로 산정되었다. 환경 데이터의 경우 일반적으로 분류성능이 30~40%의 결과를 가지기 때문에 본 연구에서 구축한 모형의 결과가 타당하다고 판

단하였다.²⁶⁾

의사결정나무 분석 결과는 Fig. 3과 4와 같다. Root node는 의사결정나무의 나무 구조가 시작되는 마디로서 분석대상이 되는 전체 자료로 이루어져 있다. Child node는 하나의 마디로부터 분리되어 나간 2개 이상의 마디들을 의미하고, Terminal node는 각 나무 줄기의 끝에 위치하고 있는 마디를 의미한다.²⁷⁾ 미생물 다양성 지수(Shannon Index, H') 히스토그램을 작성하여 33.3%로 3등급화 하였으며, Shannon index (H') class는 각 등급을 의미한다. Training cases는 각 등급에 속하는 데이터의 개수, %는 전체 데이터 중 해당 등급에 속하는 데이터의 비율을 의미한다.

데이터 마이닝 분석 결과, 식생이 없는 지역에서는 퇴적토의 이화학적 특성 중 pH와 황산염이온의 농도가 미생물 다양성 지수에 각각 1순위, 2순위로 영향을 미치는 것으로 나타났으며, 식생이 존재하는 지역에서는 유기물 함량과 질산염이온이 각각 1순위, 2순위로 나타났다. 일반적으로 퇴적토의 미생물 다양성을 예측하는데 가장 좋은 지표 중 하나가 pH라고 알려져 있다.¹⁷⁾ 하지만 퇴적토에 식생이 존재하는 경우에는 퇴적토 내 영양 염류의 농도가 높아지고 이로 인하여 미생물 다양성이 높아지게 된다.¹⁹⁾ 선행 연구 사례 분석을 통해 식생이 존재하는 지역과 존재하지 않는 지역에서 각각 유기물 함량과 pH가 주요 환경 인자로 선정된 이유를 유추할 수 있었다.

갯벌은 조력에 의하여 물에 잠기는 시간과 공기 중으로 노출되는 시간이 각각 존재한다. 갯벌 퇴적토가 물에 잠기는 시간 동안 퇴적토는 혐기성 조건이 유지된다. 이 경우 황환원 미생물이 우점종을 이루게 되어 황산염이온의 농도가 미생물의 성장에 주요한 요인이 된다.²⁰⁾ 반면에 갯벌 퇴적

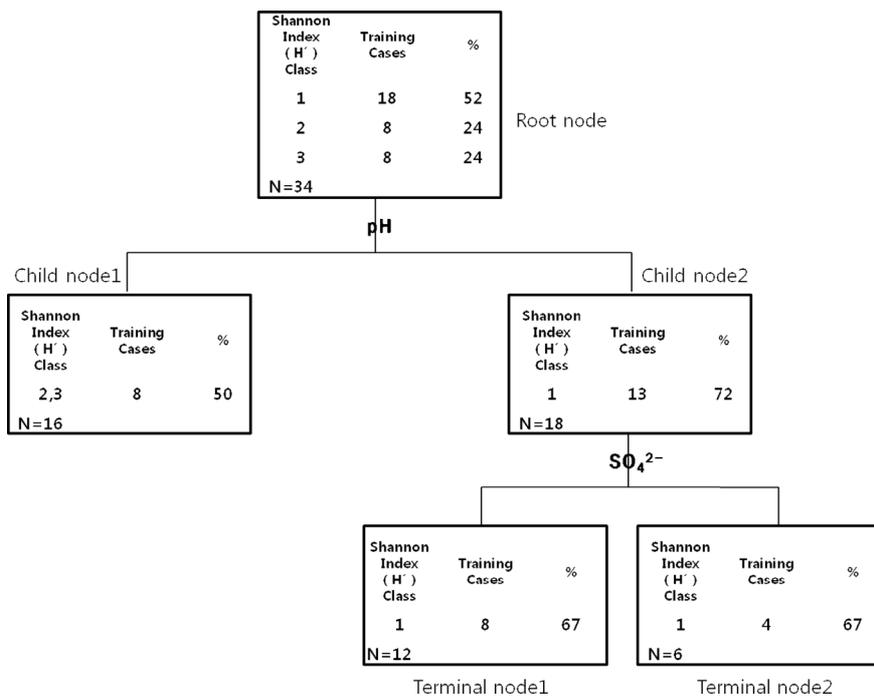


Fig. 3. Decision tree result for non-vegetation region.

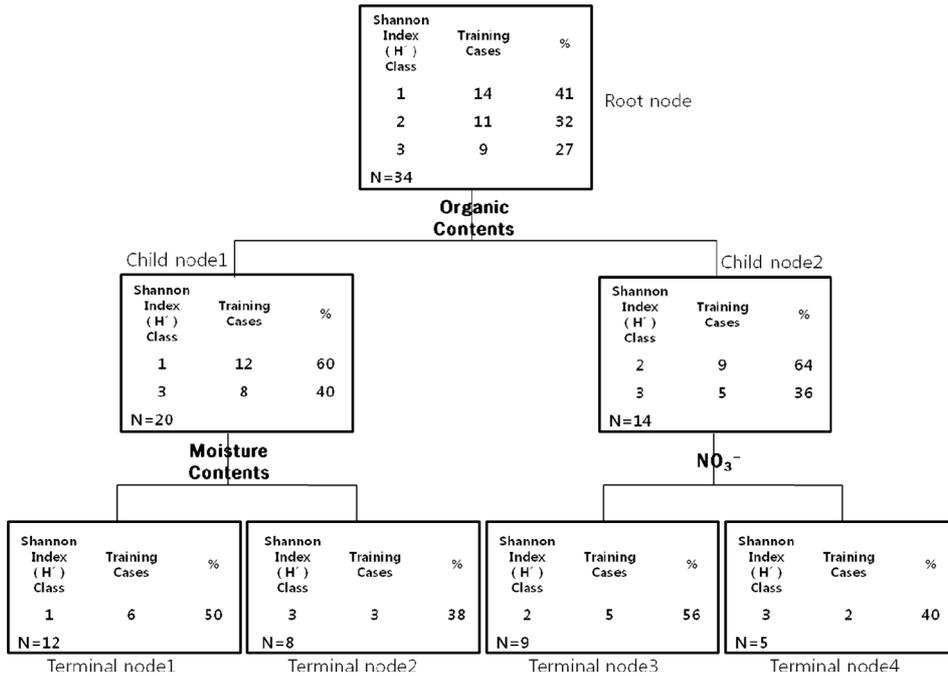


Fig. 4. Decision tree result for vegetation region.

토가 공기 중으로 노출되는 시간 동안에는 호기성 조건이 유지되며 질산화 미생물이 우점종을 이루게 되어 질산염이온이 미생물의 성장에 주요한 요인이 된다.²¹⁾ 이를 통해 식생이 존재하는 지역과 존재하지 않는 지역에서 각각 질산염이온, 황산염이온이 주요 환경 인자로 선정된 이유를 설명할 수 있다.

데이터 마이닝 분석을 이용하여 주요 환경 인자를 도출한 결과, 식생이 존재하지 않는 경우에는 미생물의 생존과 관계 깊은 pH가 미생물 다양성 지수에 영향을 크게 미친다는 것을 알 수 있었으며, 식생이 존재하는 경우에는 미생물의 성장과 관계 깊은 유기물 함량이 미생물 다양성 지수에 크게 영향을 미친다는 것을 알 수 있었다.

3.4. 시료 채취 간격 및 개수 결정 방안 제시

실측된 이화학적 특성 및 미생물 다양성 자료를 사용하여 지구통계학 및 데이터 마이닝 분석을 수행하여 얻은 결과를 토대로 시료 채취 간격 및 개수를 결정하는 방안을 Fig. 5와 같이 논의하였다. 토양특성 파악 단계는 토양의 이화학적 특성 및 미생물 다양성, 분석 대상 토양의 면적을 조사하는 단계이다. 다양한 방법을 이용한 시료채취개수 산정 단계 중 n₁은 환경부 고시 방법에 의하여 조사 대상 토양 면적에 따라 시료 채취 개수를 산정하는 방법이다.²⁸⁾ 환경부 지침에서는 토양 면적 별로 시료 채취 지점 수 및 격자 간격을 제시하고 있다. 토양 면적이 500 m² 이하인 경우 10~14 m의 간격으로 총 5개의 시료를 채취하는 것을 추천하고 있으며, 2,000 m² 이상인 경우 1,000 m² 마다 약 20~26 m 간격으로 2개의 시료를 추가로 채취하는 것을 추천하고 있다. n₂는 지구통계학 분석을 통해 획득한 미생물 다양성의

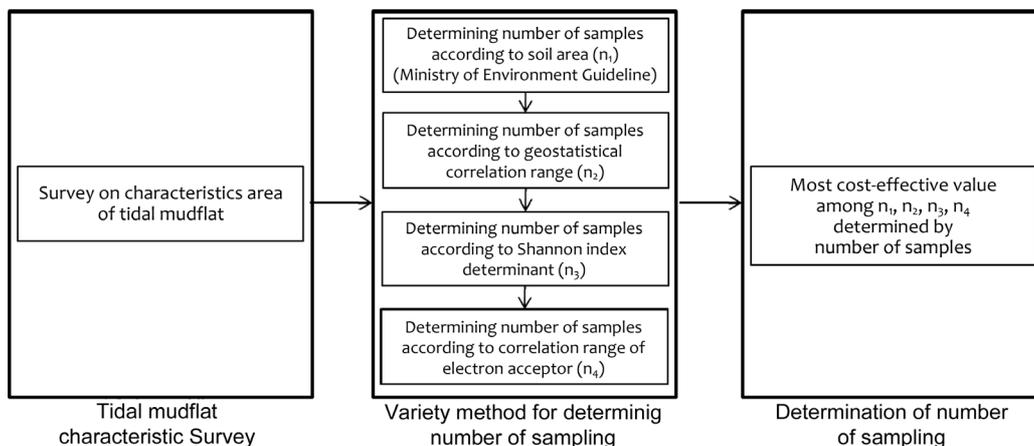


Fig. 5. Scheme of our proposed guideline for determining interval and number of samples using geostatistic and data-mining combined approach.

상관거리를 시료 채취 간격으로 사용하여 산정한 시료 채취 개수이다. n_3 는 데이터 마이닝 분석을 통해 선정된 미생물 다양성 주요 영향 인자의 지구통계학적 상관거리를 시료 채취 간격으로 사용하여 시료 채취 개수를 산정하는 것이며, n_4 는 토양 미생물 군집의 형성에 영향을 미칠 것으로 예상되는 최종전자 수용체의 상관거리를 이용하여 시료 채취 개수를 산정하는 방법이다. 각각의 방법으로 선정된 시료 채취 간격을 사용하여 환경부 지침과 같은 토양 조사 면적에서 획득 가능한 시료의 개수를 산정하였다. 식생이 있는 갯벌은 식생이 없는 갯벌에 비하여 비균질성이 높기 때문에,²³⁾ 본 연구에서는 시료 채취 간격 및 개수 결정의 용이성을 위하여 식생이 없는 갯벌만 고려하여 시료 채취 개수 선정 지침을 제안하였다.

첫 번째 접근방법은 지구통계학 분석을 통해 산정한 미생물 다양성 지수의 상관거리를 사용하여 시료 채취 개수를 도출하는 것이다. 식생이 없는 지역에서 산정한 미생물 다양성 지수의 상관거리는 약 20 cm이다. 하지만 시료간의 간격이 너무 가까울 경우 지구통계학적 상관성 분석 결과를 신뢰할 수 없을 뿐만 아니라,²⁹⁾ 시료간의 연계성 및 시료 채취의 중복성 문제가 발생할 수 있다. 또한 예상한 시료 채취 예산을 초과하는 경우가 발생할 수 있다. 따라서 본 연구에서는 산정된 상관거리의 5배의 거리마다 시료를 채취하는 것이 독립적인 시료를 획득할 수 있으면서도 지구통계학적으로 대표적인 시료의 채취에 적합하다고 판단하여 미생물 다양성 측정을 위한 시료 채취 간격을 약 1 m로 산정하였다. 이를 근거로 조사면적 별 시료 채취 개수를 산정한 결과, Fig. 6과 같이 환경부 고시 방법에 의한 시료 채취 개수에 비해서 면적 당 약 500배 많은 시료 채취가 필요함을 알 수 있었다. 본 방법을 통해 산출된 시료 채취 개수를 사용하는 것은, 시료 채취의 주된 목적이 미생물 다양성 모니터링인 경우에 가능하다. 조밀한 간격으로 시료를 채취하여 정밀한 미생물 다양성 분석이 가능하기 때문이다. 하지만 현실적으로 너무 많은 수의 시료 채취에 소모되는 비용 및 인력 측면에서 비경제적이라는 것이 한계이다.

두 번째 접근방법은 식생이 존재하지 않는 지역에서 데이터 마이닝 분석을 통하여 미생물 다양성 지수에 가장 크게 영향을 미치는 주요 환경 인자로 도출된 pH의 상관거리를 시료 채취 간격으로 사용하여 시료 채취 개수를 산정하는 것이다. 지구통계학 분석을 통해 측정된 pH의 상관거리의 평균치는 약 368 cm이다. 본 연구에서 제안한 방법과 같이 pH의 상관거리의 5배를 시료 채취 간격으로 산정하였다. 이를 근거로 조사면적 별 시료 채취 개수를 산정하면, Fig. 6과 같이 조사면적이 약 3,000 m² 이내에서는 pH의 상관거리를 시료 채취 간격으로 사용한 경우의 시료 채취 개수가 환경부 고시 방법에 의한 시료 채취 개수에 비해서 적거나 유사한 반면, 3,000 m² 이상에서는 환경부 고시 방법에 의한 시료 채취 개수에 비하여 많아지는 것을 확인할 수 있었다. 이 방법은 데이터 마이닝 분석으로 산정한 미생물 다양성에 영향을 끼치는 주요 환경 인자의 상관거리를 사용하여 시료 채취 개수를 산출한 결과라는 점에서 통계적인 유의성을 지니는 장점이 있다. 하지만 3,000 m² 이상의 면적에서는 시료 채취 개수가 증가하여 환경부 고시 지침에 비해 비경제적이라는 한계가 있다.

마지막은 미생물 군집의 형성에 영향을 미칠 것으로 예상되는 전자수용체(electron acceptor)의 상관거리를 사용하여 시료 채취 개수를 결정하는 방법이다. 갯벌 퇴적토 표층 미생물 다양성 분석을 위하여 질산염이온을 주요 전자수용체로 가정하여 이의 측정된 상관거리를 적용하는 방안을 고려하였다. 데이터 마이닝 분석 결과 식생이 존재하지 않는 갯벌의 경우 전자수용체 중 황산염이온이 미생물 다양성 결정에 영향을 미치는 두 번째 주요 환경 인자였고, 질산염이온 역시 식생이 존재하는 갯벌에서 두 번째 주요 환경 인자로 선별되었다. 그럼에도 불구하고 본 연구에서 황산염이온 대신 질산염이온을 고려한 이유는 황산염이온의 경우에 깊이에 따라 그 농도변화가 민감한 반면,²⁰⁾ 질산염이온은 표층에서 농도변화가 민감하지 않기 때문이다.²⁴⁾ 또한 표층 미생물에게 황산염이온보다는 질산염이온이 열역학적으로 좋은 전자수용체로 알려져 있기 때문이다.³⁰⁾ 따라서 질산염이

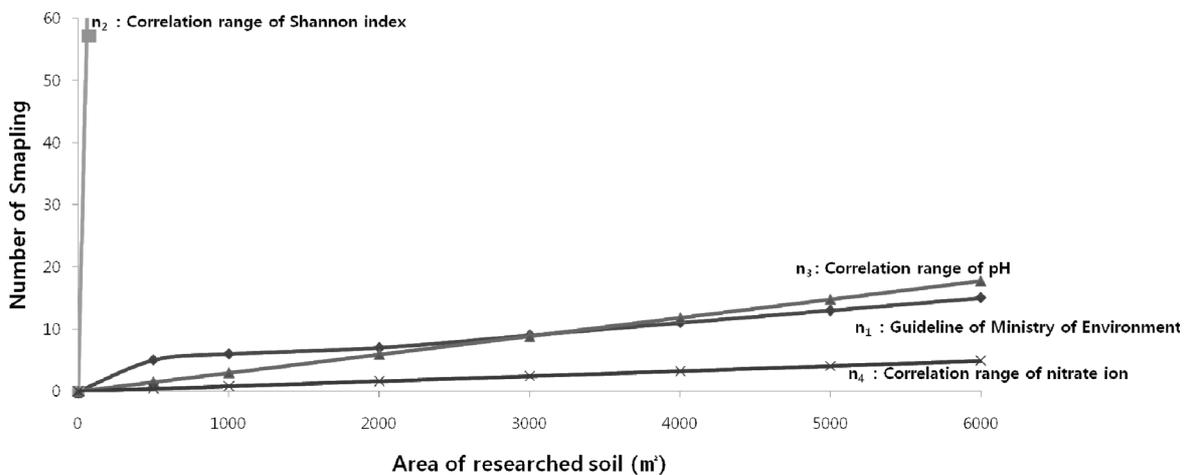


Fig. 6. Comparison of sampling numbers estimated by various approach.

온의 상관거리를 이용하여 약 17 m의 시료 채취 간격을 산정하였다. 이를 근거로 조사면적 별 시료 채취 개수를 산정해 보면, Fig. 6과 같이 전체적으로 환경부 고시 방법에 의한 시료 채취 개수보다 적은 결과를 확인할 수 있다. 위 접근방법은 시료 채취 비용이 절감된다는 측면에서 경제적이라는 장점과 표층 미생물생태의 주요 영향 인자인 질산염이온의 상관거리를 이용한다는 논리적인 장점이 있는 반면, 획득 가능한 시료의 개수가 적어 미생물 다양성 정보를 보다 정확하게 확인하기 힘들다는 한계점이 있을 수 있다.

본 연구 결과에 따라, 시료 채취의 주된 목적이 갯벌 퇴적토 미생물 다양성의 정밀 조사인 경우에는 미생물 다양성 지수의 상관거리를 시료 채취 간격으로 사용하여 매우 조밀한 간격으로 시료를 채취하는 것을 고려할 수 있다. 하지만 시료 채취 예산이 한정적이거나, 시료 채취의 주된 목적이 미생물 다양성과 환경 인자 간의 상호작용 평가인 경우 pH의 상관거리를 시료 채취 간격으로 사용하는 것을 고려할 수 있다. 또한 시료 채취의 주된 목적이 표층 미생물의 분포 및 다양성 평가인 경우 질산염이온의 상관거리를 시료 채취 간격으로 사용하는 것을 고려할 수 있다.

4. 결론

본 연구에서는 갯벌 미생물 다양성 모니터링을 위한 대표성을 지닌 시료의 채취를 위하여, 국내 갯벌에서 체계적으로 획득된 실증자료를 지구통계학 및 데이터 마이닝 기법으로 분석하여 적절한 시료 채취 간격과 개수를 선정하는 지침을 제안하였다. 이 과정에서 식생이 존재하는 지역과 존재하지 않는 지역을 비교분석한 결과, 식생이 존재하는 갯벌의 미생물 다양성 및 이화학적 특성을 분석할 경우에는 식생이 존재하지 않는 지역에서보다 보다 많은 시료의 개수가 필요하다는 것을 지구통계학적 분석으로 알 수 있었다. 이는 식생이 존재하는 지역이 생태학적으로, 이화학적으로 보다 더 복잡하고 이질적이기 때문에 예상할 수 있었던 정성적인 결과이다. 본 연구에서는 데이터 마이닝을 이용해서 특정 갯벌의 미생물 다양성에 영향을 주는 이화학적 환경인자를 도출하고, 이를 지구통계학적 기법과 접목하여 특정 갯벌 미생물 다양성 모니터링을 위한 시료 채취 간격 및 개수를 선정하는 새로운 방법을 제안하였다. 하지만 본 연구에서 사용한 지구통계학 및 데이터 마이닝 방법은 특정 환경에서 획득한 데이터를 기반으로 도출된 결과이기 때문에 환경 조건이 다른 현장에 적용할 경우 별도의 데이터 처리가 필요하다는 한계점이 있다.³¹⁻³³⁾ 그럼에도 불구하고 본 연구는 확보한 일정량의 현장 측정 데이터를 이용하여 갯벌 퇴적토의 미생물 다양성 분석을 위한 적절한 시료채취 개수 및 간격을 설정하는 지침을 제시했다는 점에서 의미가 있다.

KSEE

참고문헌

1. 전승수, 우한준, 제종길, 신상호, 김종관, “갯벌총서 #1 우리나라 갯벌 -자연 생태의 특성,” 해양수산부, 시그마프레스, 서울, pp. 6~11(2005).
2. Kim, B., Oh, H., Kang, H., Park, S. and Chun, J., “Remarkable bacterial diversity in the tidal flat sediment as revealed by 16S rDNA analysis,” *Microb. Biotechnol.*, **14**(1), 205~211(2004).
3. Horowitz, N. H., “The Origins of Molecular Genetics- One Gene, One Enzyme,” *BioEssays*, **3**(1), 37~39(1985).
4. Lee, J., Nishijima, W., Mukai, T., Takimoto, K., Seiki, T., Hiraoka, K. and Okada, M., “Factors to Determine the Functions and Structures in Natural and Constructed Tidal Flats,” *Water Res.*, **21**(9), 2601~2606(1998).
5. Kim, B., Oh, H., Kang, H. and Chun, J., “Archaeal Diversity in Tidal Flat Sediment as Revealed by 16S rDNA Analysis,” *J. Microbiol.*, **43**(2), 144~151(2005).
6. Stevens, H., Brinkhoff, T., Rink, B., Vollmers, J. and Simon, M., “Diversity and abundance of Gram positive bacteria in a tidal flat ecosystem,” *Environ. Microbiol.*, **9**(7), 1810~1822(2007).
7. Haynes, K., Hofmann, T. A., Smith, C. J., Ball, A. S., Underwood, G. J. C. and Osborn, A. M., “Diatom-derived carbohydrates as drivers of bacterial community composition in estuarine sediments,” *Appl. Environ. Microbiol.*, ASM, Online published(2007).
8. Wilms, R., Sass, H., Kopke, B., Koster, J., Cypionka, H. and Engelen, B., “Specific Bacterial, Archaeal, and Eukaryotic Communities in Tidal-Flat Sediments along a Vertical Profile of Several Meters,” *Appl. Environ. Microbiol.*, **72**(4), 2756~2764(2006).
9. Kim, B., Kim, B. K., Lee, J., Kim, M., Lim, Y. W. and Chun, J., “Rapid phylogenetic dissection of prokaryotic community structure in tidal flat using pyrosequencing,” *J. Microbiol.*, **46**(4), 357~363(2008).
10. Miletto, M., Loch, R., Antheunisse, A. M., Bodelier, P. L. E. and Laanbroek, H. J., “Response of the Sulfate-Reducing Community to the Re-establishment of Estuarine Conditions in Two Contrasting Soils : a Mesocosm Approach,” *Microb. Ecol.*, **59**(1), 109~120(2010).
11. 정승우, 안윤주, “토양위해성평가를 위한 합리적 토양조사 방안 연구,” 한국지하수토양환경학회지, **12**(1), 36-43(2007).
12. Gibson, G. R., Bowman, M. L., Gerritsen, J. and Snyder, B. D., “Estuarine and Coastal Marine Water : Bioassessment and Biocriteria Technical Guidance,” EPA 822-B-00-024, U.S. Environmental Protection Agency, Office of Water, Washington, DC, pp. 20~31(2000).
13. Crosby, L. D. and Criddle, C. S., “Understanding Bias in Microbial Community Analysis Techniques due to rrn Operon Copy Number Heterogeneity,” *BioTechniques*, **34**(4), 2~9(2003).
14. 최종근, 지구통계학. 시그마프레스, 서울, pp. 136~161(2007).
15. Breiman, L., Friedman, J. H., Olshen, R. A. and Stone, C. J., “Classification and regression trees,” Chapman and Hall/

- CRC, U.S.A, pp. 10~37(1984).
16. Mitchell, R. J., Hester, A. J., Campbell, C. D., Chapman, S. J., Cameron, C. M., Hewison, R. L., Potts, J. M., "Is vegetation composition or soil chemistry the best predictor of the soil microbial community?," *Plant Soil*, **333**(1), 417~430(2010).
 17. Fierer, N. and Jackson, R. B., "The diversity and biogeography of soil bacterial communities," *PNAS*, **103**(3), 626~631(2006).
 18. Grayston, S. J., Wang, S., Campbell, C. D. and Edwards, A. C., "Selective Influence of Plant Species On Microbial Diversity in the Rhizosphere," *Soil Biol. Biochem.*, **30**(3), 369~378(1998).
 19. Torsvik, V. and Overas, L., "Microbial diversity and function in soil : from genes to ecosystems," *Curr. Opin. Microbiol.*, **5**(3), 240~245(2002).
 20. Shin, W. S., Pardue, J. H. and Jackson, A., "Oxygen demand and sulfate reduction in petroleum hydrocarbon contaminated salt marsh soils," *Water Res.*, **34**(4), 1345~1353 (2000).
 21. Sun, G. and Austin, D., "Completely autotrophic nitrogen-removal over nitrite in lab-scale constructed wetlands : evidence from a mass balance study," *Chemosphere*, **68**(6), 1120~1128(2007).
 22. Kim, K. and Park, J., "A survey of applications of artificial intelligence algorithms in Eco-environmental modeling," *Environ. Engineer. Res.*, **14**(2), 102~110(2009).
 23. Junkins, R., Kelaher, B. and Levinton, J., "Contributions of adult oligochaete emigration and immigration in a dynamic soft-sediment community," *J. Exp. Mar. Biol. Ecol.*, **330**(1), 208~220(2006).
 24. Rocha, C. and Cabbal, A. P., "The influence of tidal action on porewater nitrate concentration and dynamics in intertidal sediments of the Sado estuary," *Estuaries and Coasts*, **21**(4A), 635~645(1998).
 25. Ekschmitt, K. and Griffiths, B. S., "Soil biodiversity and its implications for ecosystem functioning in a heterogeneous and variable environment," *Appl. Soil Ecol.*, **10**(3), 201~215(1998).
 26. Pal, M. and Mather, P. M., "An assessment of the effectiveness of decision tree methods for land cover classification," *Remote Sensing of Environ.*, **86**(4), 554~565(2003).
 27. 강현철, 한상태, 최종후, 이성건, 김은석, 엄익현, 김미경, "고객관계관리(CRM)를 위한 데이터마이닝 방법론," 자유아카데미, 경기, pp. 75~82(2006).
 28. 환경부고시 제 2001-202호 "토양환경평가 지침," 환경부, 대한민국, pp. 14(2001).
 29. Franklin, R. B. and Mills, A. L., "Multi-scale variation in spatial heterogeneity for microbial community structure in an eastern Virginia agricultural field," *FEMS Microbiol. Ecol.*, **44**(3), 335~346(2003).
 30. Lide, D. R., CRC Handbook of Chemistry and Physics, 90st Edition, ed, CRC Press Inc., Boca Raton, FL, pp. 5~4(2009).
 31. 박노옥, 지광훈, 권병두, "지구통계학적 시뮬레이션을 이용한 원격탐사 화상 분류 결과의 공간적 불확실성 분포의 추정," 2004 GIS/RS 공동 춘계학술대회 논문집, 한국지형공간정보학회, 서울, pp. 463~468(2004).
 32. Shim, B. O., Chung, S. Y., Kang, D. H., Kim, G. B. and Park, H. Y., "Geostatistical Analysis for the Groundwater Database of the Youngsan and Sumjin River Basins," *J. Engineer. Geol.*, **10**(2), 131~142(2000).
 33. Berry, M. J. and Linoff, G. S., Data Mining Techniques, WILEY publishing Inc., Indiana, USA pp. 165~209(2004).