creative commons

COMMONS DEED

저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:

저작자표시. 귀하는 원저작자를 표시하여야 합니다.

비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.

변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 이용허락규약(Legal Code)을 이해하기 쉽게 요약한 것입니다.

Disclaimer

dCollection

공학박사 학위논문

하천의 경로를 고려한 지구통계학 기반의 하상퇴적물 중금속 오염농도 예측

Geostatistical prediction of heavy metal concentrations in stream sediments considering the path of stream

2016년 2월

서울대학교 대학원

에너지시스템공학부

김 성 민

초 록

광산지역에 적재되어 있는 폐석이나 광물찌꺼기에서 발생하는 침출수는 인근 수계나 농경지 토양으로 유입되어 환경오염을 유발할 수 있다. 중금속 오염에 의한 피해를 효율적으로 복구하고 방지하기 위해서는 중금속의 오염정도와 범위, 중금속 원소 분포, 이동경로 등을 고려해야 한다. 그러나 중금속 오염은 영향 범위가 넓기 때문에 현장에서 직접 오염양상을 파악하는 것은 어려운 일이다. 또한 충분한 양의 데이터를 취득하는 것이 어렵기 때문에 제한된 데이터로부터 중금속 오염을 효율적으로 예측할 필요가 있다. 이 때 다양한 공간정보를 활용하는 지리정보시스템(GIS, geographic information systems)을 지구통계학과 결합한다면 유용한 결과를 기대할 수 있다. GIS를 이용한 분석은 자료 획득이 쉽고 광범위한 면적에 대한 분석이 용이하기 때문에 중금속 오염 분포 예측에 적합한 방법이라고 할 수 있다.

중금속의 오염 분포를 예측하기 위해서 지구통계학을 적용한 사례가 다수 있으나 수계를 따라 오염물이 이동하는 현상을 고려한 연구는 부족한 실정이다. 일반적인 지구통계학 기반의 예측 기법은 하천을 따라 이동하는 하상퇴적물의 오염이나 수질 오염을 모델링할 때에 샘플 간의 거리를 제대로 고려하지 못하는 한계가 있다. 이러한 예측 방법은 베리오그램 계산 과정에서 유클리디안 거리를 이용하기 때문에 예측의 정확도가 떨어질 수 있다. 본

연구에서는 이를 개선하기 위해 지형적 요인을 반영하여 하천의 경로를 고려한 크리깅 기법을 제안하고 소프트웨어 구축을 통해 전산화된 자료들의 분석이 용이하도록 설계하였다.

본 연구에서 제안하는 예측 기법을 견본 자료에 적용한 결과 하천 네트워크 및 하천 줄기에 따른 유용한 예측 결과를 제공하였으며 교차검증을 통해서 기존의 크리깅 기법에 비해 예측 능력이 향상된 것을 확인하였다. 또한 강원도 삼척시 제2연화광산의 광미댐이 유실되었을 때 취득된 현장자료에 적용하여 일반적인 크리깅 기법의 적용 결과와 비교를 수행하였다. 그 결과 실제 자료에서도 제안한 기법의 예측 능력이 더 우수한 결과를 보였다. 샘플의 집수구역 면적을 부가자료로서 활용하여 집수구역의 면적과 중금속 오염농도 간의 관계를 이용할 경우에는 하나의 하천 줄기 내에서 활용성이 높은 것으로 생각된다.

제안된 예측기법은 하천 네트워크에 적용하여 하천 줄기별 오염 복구우선순위를 산정할 때 유용하게 활용될 수 있을 것이다. 폐광산 지역 뿐만 아니라 목축지나 도심지, 공장지대로 인한 오염 등 다양한 상황에서 범용적으로 활용될 수 있으며 하천의 네트워크가 복잡하고 굴곡이 심할수록 상대적 활용도가 높을 것으로 기대된다.

주요어 : 지리정보시스템, 지구통계, 중금속, 크리깅, 하상퇴적물

학 번 : 2009-23182

목 차

1. 서론.....1

1.1. 연구배경.....1

1.2. 오염 모델링에 관한 연구동향.....3

1.3. 연구의 필요성.....5

1.4. 연구목적.....7

2. 지구통계학 기반의 공간분포 예측에 관한 배경이론.....9

2.1. 공간자료와 자기상관성.....9

2.2. 공간자료의 특성을 반영하는 베리오그램 모델링.....11

2.3. 미지의 값 예측을 위한 크리깅 기법.....16

3. GIS 기반의 수계분석을 결합한 크리깅 기법 개발.....20

3.1. 지형을 고려한 하천경로 분석.....25

3.2. 하천경로를 고려한 샘플 간의 거리분석과 크리깅.....30

3.3. 집수구역을 고려한 다변량 크리깅.....34

3.4. 소프트웨어 구현.....38

4. 예측능력 검증을 위한 견본자료 분석.....43

4.1. 유실량과 집수구역을 고려한 오염농도 시나리오 설계.....44

4.2. 일반적인 크리깅 분석.....51

4.3. 하천경로를 고려한 크리깅 분석.....59

4.4. 집수구역을 고려한 다변량 크리깅 분석.....71

5. 알고리즘의 현장자료 적용 및 예측결과 분석.....79

5.1. 연구대상지역.....79
5.2. 대상지역 자료 검토 및 전처리.....84
5.3. 샘플 간의 거리계산 및 베리오그램 모델링.....90
5.4. 오염농도 예측 및 검증.....95

6. 결론.....107

참고문헌.....110

Abstract.....116

감사의 글.....118

List of Tables

Table 4.1. Cross validation results of ordinary kriging using Euclidean distance for test samples.....58
Table 4.2. Cross validation results of ordinary kriging using distance on the stream for test samples.....64
Table 4.3. Cross validation results of kriging using Euclidean distances and distances on the stream for non-point pollution source.....70
Table 4.4. Cross validation results of SKLM using the correlation of watershed area and concentration for test samples.....78
Table 5.1. Monthly precipitation (mm) of Donghae, Gangwondo from 2005 to 2014.....83
Table 5.2. Heavy metal concentrations of stream sediment samples.....88
Table 5.3. Heavy metal concentrations of farmland soil samples 89
Table 5.4. Euclidean distance and flow distance between samples.....91
Table 5.5. Cross validation results of general kriging and flow kriging for study area.....99

List of Figures

Fig. 2.1. General concept of empirical variogram and theoretical variogram.....13
Fig. 2.2. Classification of predicted total phosphorus concentrations in hydrologic cataloging units of the conterminous United States (Smith et al., 1997).....19
Fig. 2.3. Estimates of the normalised specific 100-year flood Q100N on the stream network of the Mur region (Skøien et al., 2006).....19
Fig. 3.1. Input data and overall procedures of the prediction of heavy metal concentration considering the stream path.....21
Fig. 3.2. Rasterization of (a) unknown area, (b) samples and stream paths, and (c) modification of sample location to stream.....23
Fig. 3.3. Grid-based concept of (a) DEM, (b) maximum downhill slope, (c) flow direction and (d) flow path.....27
Fig. 3.4. Grid-based (a) flow direction, (b) flow accumulation and (c) defined stream.....29
Fig. 3.5. Flowchart for the calculating the distances between samples and predicting the unknown value using kriging.....32
Fig. 3.6. Euclidean distance (a) and flow distance (b) between samples in the stream network.....33
Fig. 3.7. Dilution of heavy metal pollution depending on the

catchment area for each sampling location.....37
Fig. 3.8. Graphical user interface of the distance calculator module.....40
Fig. 3.9. Graphical user interface of the variogram modeling module.....41
Fig. 3.10. Graphical user interface of the kriging analysis module.....42
Fig. 4.1. (a) DEM and samples and (b) stream order and watersheds of the test area.....45
Fig. 4.2. Distribution of (a) LS factor that represents soil loss and (b) simulated heavy metal concentration of test samples.....48
Fig. 4.3. Empirical variogram calculated from Euclidean distances between test samples and theoretical variogram models using (a) linear model, (b) spherical model and (c) gaussian model.....52
Fig. 4.4. Prediction of heavy metal concentrations for the test area using linear model with the correlation length of (a) 3,000 m and (b) 4,500 m.....54
Fig. 4.5. Prediction of heavy metal concentrations for the test area using spherical model with the correlation length of (a) 3,000 m and (b) 4,500 m.....55
Fig. 4.6. Prediction of heavy metal concentrations for the test area using gaussian model with the correlation length of (a) 3,000 m and (b) 4,500 m.....56

Fig. 4.7. Empirical variogram calculated from distances on the stream between test samples and theoretical variogram models using (a) linear model, (b) spherical model and (c) gaussian model.....60
Fig. 4.8. (a) Prediction of heavy metal concentrations for the test area using flow kriging with spherical variogram model and (b) the number of connected samples of each sample.....62
Fig. 4.9. Distribution of (a) non-point pollution source and (b) simulated heavy metal concentration of stream sediment test samples.....66
Fig. 4.10. Empirical variogram and theoretical variogram calculated from (a) Euclidean distances and (b) distances on the stream between test samples for non-point pollution source.....67
Fig. 4.11. Prediction of heavy metal concentrations for the non-point pollution source test area using (a) Euclidean distance and (b) flow kriging.....69
Fig. 4.12. (a) Correlation between watershed areas and heavy metal concentrations of stream sediments and (b) the residual distribution according to watershed area.....73
Fig. 4.13. Empirical variogram and theoretical variogram of residuals calculated from (a) linear regression and (b) log regression.....74
Fig. 4.14. (a) Regressed heavy metal concentrations using the correlation of watershed area and concentration and (b) prediction of concentration by the sum of regressed

concentrations and predicted residuals.....76
Fig. 5.1. Location of study area on the outline map and satellite image (Google Earth).....81
Fig. 5.2. DEM and locations of samples. Sample numbers in brackets indicate stream sediment samples collected from non-polluted stream.....82
Fig. 5.3. Heavy metal concentration of samples. (a) As (b) Cd (c) Cu (d) Ni (e) Pb (f) Zn.....87
Fig. 5.4. Empirical variogram and theoretical variogram for 20 samples considering (a) Euclidean distances and (b) distances on the stream.....93
Fig. 5.5. Empirical variogram and theoretical variogram for 15 samples considering (a) Euclidean distances and (b) distances on the stream.....94
Fig. 5.6. Prediction of Zn concentrations for the study area using (a) Euclidean distance and (b) flow kriging with variogram model for 20 samples.....97
Fig. 5.7. Prediction of Zn concentrations for the study area using (a) Euclidean distance and (b) flow kriging with variogram model for 14 samples.....98
Fig. 5.8. (a) Correlation between watershed areas and heavy metal concentrations of study area and (b) the residual distribution according to watershed area.....101
Fig. 5.9. Empirical variogram and theoretical variogram of residuals.....102

Fig. 5.10. Prediction of Zn concentrations for the study area using flow kriging with (a) spherical variogram model and (b) spherical variogram model combined with nugget model.....105
Fig. 5.11. Variation of predicted Zn concentrations for different methods according to distance from tailings dam.....106

1. 서론

1.1. 연구배경

광산지역에 적재되어 있는 폐석이나 광물찌꺼기에서 발생하는 침출수는 인근 수계나 농경지 토양으로 유입되어 환경오염을 유발할 수 있으며, 특히 여름철 집중호우에 의해 대량의 광물찌꺼기가 적치되어 있는 광미댐이 무너질 경우 주변 환경에 큰 피해를 줄 수 있다. 광산에서 배출되는 갱내폐수 및 광산폐기물에는 다양한 유해금속이 농축되어 있기 때문에 이들 중금속 원소들이 생지구화학적 경로로 유입될 경우 인체에 누적되어 심각한 위험을 초래할 수 있다(박천영 외, 1995).

중금속 오염에 의한 피해를 효율적으로 복구하고 방지하기 위해서는 중금속의 오염정도와 범위, 중금속 원소 분포 패턴, 이동경로 등에 대한 종합적인 분석 방법이 도입되어야 한다. 중금속의 존재형태를 규명하기 위해서 완전분해법, 연속추출법 등의 방법이 사용되고 있으며, pH 측정, XRD(X-ray diffraction) 분석, ICP(inductively coupled plasma)나 XRF(X-ray fluorescence) 등을 이용한 화학분석 등을 통해 오염 현장에서의 중금속 원소 함량과 분포 특성을 파악할 수 있다. 그러나 산악지형이 많은 우리나라의 특성상 접근성이 좋지 않은 경우가 많고, 오염원으로부터 발생하는 중금속의 영향 범위가 넓기 때문에 현장에서 직접 오염양상을 파악하는 것은 어려운 일이다(이평구, 2004). 또한 오염정도를 파악하기 위한 데이터 취득 과정에 많은 노력과 비용이 필요하기

때문에 충분한 양의 데이터를 취득하는 것이 어려운 실정이다. 따라서 제한된 데이터로부터 중금속 오염이 예상되는 지역과 그 오염정도를 보다 정확하게 예측하는 것은 환경오염으로부터 인간의 건강을 보호할 뿐 아니라 복구비용의 절감 측면에서 중요한 과제가 아닐 수 없다.

국내에서는 광산지역에서의 오염 연구를 중심으로 중금속의 함량 측정 및 존재형태 규명 등에 대한 연구(김상현과 전효택, 1993; 안주성 외, 1999; 박천영 외, 2002)가 활발하게 진행되어 왔으나 분산되어 있는 측정 데이터로부터 오염 분포를 예측하는 연구는 상대적으로 미흡한 상황이다. 이때 공간적 또는 시간적으로 분포하는 자료를 해석하는 지구통계학(geostatistics)이 미지의 값을 예측하는데 있어 유용하게 활용될 수 있다. 중금속의 오염 분포를 예측하기 위해서 지구통계학이 사용된 사례는 일부 보고된바 있으나(김경웅과 이현구, 1996), 하천에서 취득한 수질 데이터나 하상퇴적물 데이터를 활용하여 일반적인 토양의 오염 정도를 예측하는 등 오염물이 주로 수계를 따라 이동하는 경향을 고려하지 못한 한계가 있다. 특히 하상퇴적물 내의 중금속은 시간적인 변화가 적고 함량이 높아 중금속에 의한 지속적인 환경영향을 평가하기에 적절하기 때문에(Thornton, 1983) 하천의 흐름을 고려한 오염 분포 예측 연구의 필요성이 대두된다. 지형조건이나 지리조건 등 다양한 공간정보를 활용하는 지리정보시스템(GIS, geographic information systems)을 지구통계학과 결합한다면 지형분석을 통한 유용한 결과를 기대할 수 있다. 또한 GIS를 이용한 분석은 자료 획득이 쉽고 광범위한 면적에 대한 분석이 용이하기 때문에 중금속 오염 분포 예측에 적합한 방법이라고 볼 수 있다.

1.2. 오염 모델링에 관한 연구동향

컴퓨터를 활용한 GIS 기술의 발전과 함께 다양한 연구자들에 의하여 오염물의 이동경로 및 오염 분포를 예측하기 위한 GIS 모델링 기법들이 개발되었고 적용되었다.

지형정보를 가지는 수치표고모델(digital elevation model, DEM)은 GIS 공간분석에 있어 가장 기초적인 자료로서 이를 활용한 지표수 및 오염 모델링에 관한 연구가 다수 보고된바 있다. 8방향 중 한 방향으로 물이 흐른다고 가정하는 D8 알고리즘(O'Callaghan and Mark, 1984; Jenson and Domingue, 1988)은 ArcGIS, IDRISI, GRASS 등의 다양한 GIS 소프트웨어에서 널리 사용되고 있으며 강우의 흐름 분석 및 침전물과 오염물의 이동 경로 분석 등의 분야에서 활용되고 있다. 김성민 등(2011)은 D8 알고리즘을 이용하여 시간에 따른 광산폐기물 침출수의 이동경로를 분석하였으며, 이 기법은 광산지역의 다양한 재해를 분석하는 ArcMine 프로그램에 적용된 바 있다(Kim et al., 2012). Heathwaite et al.(2005)는 도로나 철로 등의 영향을 고려한 D8 알고리즘을 적용하여 농경지에서의 토양오염분포를 예측하였다. Ayad(2006)는 폐광산에서 유출된 산성광산배수(acid mine drainage, AMD)의 오염경로를 분석하기 위해서 GIS를 이용하여 대상지역을 소규모의 집수구역으로 구분하고, pH와 미량원소 분포특성에 따라 집수구역별 복구 우선순위를 산정하였다. Tamas and Kovacs(2003)는 토양침식모델(universal soil loss equation, USLE)과 GIS를 이용하여 광물찌꺼기의 유출방향과 이에 따른

오염물질의 분포를 분석하였다. Yenilmez et al.(2011)은 GIS를 이용하여 지표 위에서 침출수가 흐르는 이동경로를 예측하고 실측한 정밀조사 자료와의 비교를 통해 이를 검증하였다. 그러나 해당 연구들은 실측 데이터를 활용하지 못하거나 오염 모델링 결과에 대한 검증용으로 활용한 경우가 대부분이기 때문에 실제 오염 정도 및 분포를 반영하기 어려운 한계를 가지고 있다.

이러한 문제를 해결하기 위해서 실측 데이터를 이용한 지구통계학 기반의 오염 예측 연구가 수행되었다. Salgueiro et al.(2008)은 지구통계학적 예측기법을 활용하여 포르투갈 Vale das Gatas 광산 인근 하상퇴적물의 화학적 오염정도를 예측하였고, 그 외에도 농경지를 대상으로한 중금속 토양오염 예측 연구가 활발히 수행되고 있다(Steiger et al., 1996; White et al., 1997; Lin et al, 2001; Liu et al., 2006). 지구통계학 기반의 오염 모델링은 실측 데이터에 기반하여 오염 분포를 합리적으로 예측할 수 있다는 장점을 가진다. 그러나 일반적인 지구통계학 기반의 예측 기법은 하천을 따라 이동하는 하상퇴적물의 오염이나 수질 오염을 모델링할 때에 샘플 간의 거리를 제대로 고려하지 못하는 한계가 있다.

수계와 밀접한 관련이 있는 하상퇴적물의 중금속 오염을 예측하기 위해서 Curriero(1996)는 샘플 간의 거리를 구할 때 유클리디안(Euclidean) 거리가 아닌 하천 상의 거리를 이용하는 방안을 제안하였다. Ver Hoef et al.(2006)은 하천 네트워크에서의 오염물 함량을 예측하고 개발한 기법의 통계적 유효성을 검증하였다.

1.3. 연구의 필요성

중금속 오염농도를 예측하기 위한 기존 연구들은 크게 다음과 같은 세 가지 유형으로 구분할 수 있다.

(1) 지구화학 분야 연구에서는 중금속 오염의 영향을 받는 토양이나 퇴적물을 채취하고 분석하여 중금속 원소들의 지구화학적 분산을 밝히고 화학적 형태와 함량, 분포 범위 등이 규명되었다. 이러한 연구방법은 샘플을 채취한 지점에서 상대적으로 정확한 중금속 원소 함량을 구하고 특정 지역의 중금속 오염 현황을 파악하기에 용이하나 샘플 채취가 이루어지지 않은 지점에서의 값을 예측할 때에는 대략적인 경향만 파악하는 한계가 있다.

(2) 이러한 부분을 보완하기 위해서 채취한 샘플의 중금속 오염 측정값을 가지고 지구통계적으로 오염농도 분포를 예측하는 연구가 수행되었다. 대표적으로 잘 알려진 보간법 중 하나인 크리깅(kriging)이 주로 사용되고 있으며 토양을 대상으로 미측정 지점의 중금속 오염농도를 예측할 때 유용하게 활용될 수 있다. 그러나 채취한 샘플들의 위치만 고려하여 분석을 수행하기 때문에 지형적 특성을 반영하기 어렵다는 한계를 가지고 있다. 따라서 하상퇴적물과 같이 주로 하천을 따라 분산되는 경우, 샘플 간의 유클리디안 거리를 이용하는 기존의 크리깅 기법을 그대로 적용하기에 어려움이 있다.

(3) 대상지역의 지형적 특성 및 오염물과 매개체의 물리적 특성 등을 고려한 유동 시뮬레이션 기반의 연구도 활발하게 수행되고 있다. 이러한 연구방법은 특정 지점의 중금속 함량 데이터가 없거나 부족할 경우 유용하게 활용될 수 있다. 그러나 정확한 예측을 위해서는 다양한 변수를 필요로 하며 일반적으로 변수가 많고 복잡하여 필요한 데이터를 취득하기 어렵다. 또한 실측한 오염농도 값을 이용하지 않는 경우 예측값이 실제 오염농도를 반영하지 못하고 비현실적인 값을 나타낼 수도 있다.

지구화학 분석을 통해 중금속 함량을 측정할 수 없는 환경에서는 시뮬레이션을 통한 오염농도 예측이 좋은 방법이 될 수 있다. 반대로 재정과 시간이 허락된다면 모든 지점에 대해서 지구화학적 분석을 수행하여 오염농도를 파악하는 것이 가장 좋은 방법일 것이다. 그러나 현실적으로는 오염분포를 예측하고자 하는 영역에 대한 측정 데이터는 제한되어 있기 때문에 제한된 측정값으로부터 미지의 값을 예측할 필요가 있다. 여름에 강우가 집중되고 수계가 발달한 우리나라에서는 하천을 따라 중금속이 분산되는 경우가 많기 때문에 하천의 경로를 고려한 오염농도 예측 연구가 필요하다.

1.4. 연구목적

본 연구의 목적은 지구통계학에 기반하여 하천을 따라 이동하는 중금속의 오염농도를 예측하는 것이다. 데이터의 위치와 분포의 방향성을 고려하는 기존 크리깅 기법의 단점을 개선하기 위해서 GIS 기법을 결합한 오염농도 예측 알고리즘을 개발하고자 한다. 또한 중금속 오염농도에 영향을 줄 수 있다고 판단되는 샘플의 집수구역을 고려한 다변량 크리깅의 적용 가능성을 확인하고자 한다. 이를 위한 세부적인 연구목표는 다음과 같다.

(1) GIS를 이용하여 지형을 고려한 하천의 경로 및 네트워크를 분석하고 하천에서의 샘플 간 거리를 이용하는 크리깅 알고리즘을 개발한다.

(2) 하천에서의 중금속 오염에 영향을 줄 수 있는 지형적 요인들을 분석하고 다변량 크리깅을 통한 예측 정확도 향상 가능성을 평가한다.

(3) 개발한 알고리즘의 예측 정확도를 평가하기 위한 교차검증을 수행하여 기존 크리깅 기법과의 차이점을 비교한다.

본 연구에서는 가상의 측정값을 가지는 견본자료를 대상으로 분석 기법의 정확도를 평가하고, 실제 현장에서의 적용성을 확인하기 위해 폐광산 지역 하상퇴적물의 중금속 농도 데이터를

이용한 사례를 제시하고자 한다.

2. 지구통계학 기반의 공간분포 예측에 관한 배경이론

본 장에서는 중금속 오염농도를 예측하기 위해 공간상에서 자료들 간의 상관성을 나타내는 척도로서 공분산과 베리오그램을 살펴보고 이들의 관계로부터 유도된 크리깅 예측 기법에 대해서 알아보고자 한다. 본 연구에서 제안하는 예측 기법은 기존의 크리깅 이론에 기반을 두고 있기 때문에 올바른 분석을 수행하기 위해서는 지구통계학에 대한 이해가 필수적이다. 본 장에서 서술하는 배경이론은 최종근(2013)의 저서에 기반을 두고 있다.

2.1. 공간자료와 자기상관성

자연환경에서의 속성값들은 가까이 있는 값끼리는 유사한 값을 가지고 거리가 멀어질수록 유사도가 떨어진다는 원리가 받아들여지고 있다(Tobler, 1970). 이러한 현상의 의미를 해석하면 공간상에서 어떠한 분포가 나타나는 자연적 혹은 인위적 원인이 존재한다는 것을 뜻한다. 주어진 공간자료들을 이용하여 미지의 값을 예측하기 위해서는 먼저 자료들 간의 공간적 상호관계를 파악할 필요가 있다.

어떠한 하나의 변수에 대해서 위치에 따른 동일 변수의 상관관계를 자기상관이라고 하며 임의의 거리만큼 떨어진 자료들이 얼마나 관계가 있는지를 나타내기 위해 자기공분산(autocovariance)을 정의한다(Eqn. 2.1). 자기공분산은 자료 간의

거리가 가까우면 큰 값을 가지며, 자료 간의 거리가 멀수록 작은 값을 가지게 된다.

$$Cov(z_i, z_{i+k}) = E[(z_i - \overline{z_i})(z_{i+k} - \overline{z_{i+k}})] \qquad \text{Eqn. 2.1}$$

$$= E(z_i z_{i+k}) - E(z_i)E(z_{i+k})$$

크리깅을 통해 미지의 값을 예측하기 위해서는 최종적으로 자기공분산을 이용해야 하므로 자기공분산은 지구통계학에 있어 매우 중요한 개념이다. 자기공분산을 구하기 위해서는 공간자료로부터 베리오그램을 모델링하고 베리오그램과 자기공분산의 관계를 이용한다.

2.2. 공간자료의 특성을 반영하는 베리오그램 모델링

베리오그램은 일정한 거리에 있는 자료들의 유사성을 나타내는 값으로서 Eqn. 2.2와 같이 정의된다. 베리오그램 $2\gamma(h)$은 일정한 거리 h만큼 떨어진 두 자료들 간의 차이를 제곱한 것의 평균을 의미한다. 공간적으로 분포하는 자료들은 일반적으로 거리가 가까울수록 그 값들이 비슷한 값을 가지므로 베리오그램은 작아지고, 자료들 간의 거리가 멀어질수록 그 값이 크게 나타나는 성향을 가진다. 수식에서 h는 분리거리로서 두 자료 간에 떨어져 있는 거리를 의미하며 베리오그램과 공분산은 Eqn. 2.3과 같은 관계를 가진다.

$$2\gamma(h) = E[(z(x) - z(x+h))^2] \qquad \text{Eqn. 2.2}$$

$$\gamma(h) = \sigma^2 - Cov(h) \qquad \text{Eqn. 2.3}$$

표본의 분산은 이미 알고 있기 때문에 이론적 베리오그램식을 구한다면 임의의 분리거리에 대하여 공분산 값을 계산할 수 있다. 크리깅 예측을 위해서는 공분산이 필요하기 때문에 자료를 대표할 수 있는 베리오그램을 구하는 과정이 중요하다.

분리거리를 증가시키면서 베리오그램을 구하면 일반적인 베리오그램은 Fig. 2.1과 같이 점점 증가하는 형태로 나타난다. 이때 분리거리가 증가하면서 어느 정도의 거리가 되면 자료들 간의

관계성이 약해지기 때문에 베리오그램이 일정한 값을 나타내는데 이를 문턱값(sill)이라고 하며, 이처럼 자료들이 상관관계를 보이는 최대 분리거리를 상관거리라고 한다. 또한 분리거리가 더 증가하면 상관관계가 약해지고 자료쌍의 개수도 부족하여 일정한 경향없이 진동하는 현상이 나타난다. 이론적으로 분리거리가 0이면 베리오그램도 0값을 가지겠지만 실제 데이터를 이용해 계산한 실험적 베리오그램으로부터 이론적 베리오그램을 찾을 때에는 분리거리 0에서 상수인 너깃(nugget) 값을 가질 수 있다. 너깃값은 짧은 거리에서의 불확실성을 반영하기 위해서 이용된다.

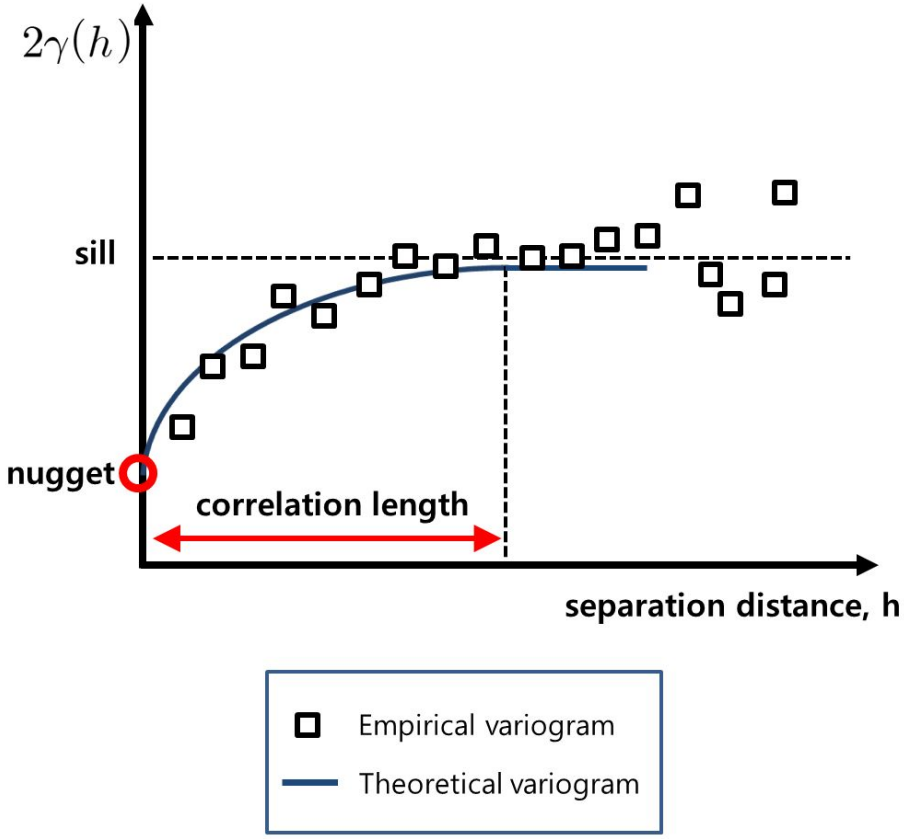


Fig. 2.1. General concept of empirical variogram and theoretical variogram.

실제로 크리깅을 이용한 예측에 활용되는 것은 이론적 베리오그램이다. 따라서 주어진 자료로부터 계산된 실험적 베리오그램을 바탕으로 이를 가장 잘 대표하는 이론적 베리오그램을 찾아내는 것이 매우 중요하다. 이론적 베리오그램은 비교적 단순한 수식을 사용하여 나타내며 특징에 따라 문턱값이 있는 모델, 문턱값이 없는 모델, 주기성을 갖는 모델로 구분할 수 있다. 본 연구에서는 문턱값이 있는 모델인 선형모델, 구형모델, 지수모델, 가우스모델, 너깃모델을 적용하여 이론적 베리오그램을 구하는 과정을 수행하였다. 이 과정에서는 분석자의 주관적인 판단이 결과에 영향을 줄 수 있다. 각 모델의 수식은 Eqn. 2.4부터 Eqn. 2.8과 같다.

(1) 선형모델

$$\gamma(h) = \begin{cases} C_0 \dfrac{h}{a}, & (h \le a) \\ C_0, & (h > a) \end{cases} \qquad \text{Eqn. 2.4}$$

(2) 구형모델

$$\gamma(h) = \begin{cases} C_0 \left[1.5\left(\dfrac{h}{a}\right) - 0.5\left(\dfrac{h}{a}\right)^3 \right], & (h \le a) \\ C_0, & (h > a) \end{cases} \qquad \text{Eqn. 2.5}$$

(3) 지수모델

$$\gamma(h) = C_0 \left[1 - \exp\left(-\frac{3h}{a} \right) \right] \qquad \text{Eqn. 2.6}$$

(4) 가우스모델

$$\gamma(h) = C_0\left[1 - \exp\left\{-3\left(\frac{h}{a}\right)^2\right\}\right] \qquad \text{Eqn. 2.7}$$

(5) 너깃모델

$$\gamma(h) = \begin{cases} C_0, h > 0 \\ 0 \quad , h = 0 \end{cases} \qquad \text{Eqn. 2.8}$$

여기서 C_0는 문턱값, a는 상관거리, h는 분리거리를 의미한다. 실험적 베리오그램을 가장 잘 대표하는 모델을 찾기 위해서는 시각적 판단이 중요하며 가중최소자승법(weighted least square method)을 이용해서 최적화할 수 있다(Cressie, 1985). 토양의 특성 예측, 수질 예측 등 다양한 연구분야에서 구형모델과 지수모델이 주로 이용되고 있으며(Dumanski et al., 1998; Kalivas et al., 2002; Hengl et al., 2004; Bijanzadeh et al., 2014), 작은 분리거리에서 자료들이 강한 상관성을 나타낼 때 가우스모델이 사용된다. 선형모델은 간단하지만 자료의 상관성이 선형적으로 변화한다는 가정의 한계가 있으며 너깃모델은 자료들의 상관관계가 전혀 없는 경우에 사용된다.

2.3. 미지의 값 예측을 위한 크리깅 기법

크리깅은 값을 모르는 관심 지점에서의 특성값을 알기 위해 Eqn. 2.9와 같이 이미 값을 알고 있는 주위값들의 선형조합으로 그 값을 예측하는 지구통계적 기법이다. 여기서 z^*는 관심지점에서의 예측치, z_i는 위치와 값을 알고 있는 주위의 자료값, λ_i는 각 자료의 가중치, n은 크리깅 예측을 위한 자료의 총 개수를 의미한다.

$$z^* = \sum_{i=1}^{n} \lambda_i z_i \qquad \text{Eqn. 2.9}$$

실험적 베리오그램을 잘 대표하는 이론적 베리오그램을 구하면 각 자료들의 가중치를 계산하여 원하는 지점에서의 자료값을 예측할 수 있다. 크리깅의 종류에는 오차분산을 최소로 하는 가중치를 적용하는 단순크리깅(simple kriging), 크리깅 추정식이 편향되지 않으면서 오차분산을 최소로 하는 정규크리깅(ordinary kriging), 원하는 구역에 대하여 하나의 크리깅 방정식을 세워 값을 예측할 수 있는 구역크리깅(block kriging) 등이 있다. 이러한 크리깅 방법들은 자료들이 분포한 위치에 상관없이 평균이 같은 값을 나타내기 때문에 자료가 지역에 따라 변화하는 경우에는 공간적 분포특성을 반영하지 못할 수 있다. 공간적으로 변화하거나 특정한 경향을 반영하여 가중치를 계산하는 크리깅 방법으로 일반크리깅(universal kriging)이 있다.

크리깅은 다양한 분야에서 활발하게 연구되고 활용되고 있으며 다수의 상용 소프트웨어들이 크리깅 분석 기능을 제공하고 있다. 통계 및 지구통계 분석 소프트웨어 SAS (http://www.sas.com), SGeMS (http://sgems.sourceforge.net), GIS 분석 소프트웨어 ArcGIS (https://www.arcgis.com), IDRISI (https://clarklabs.org), 지질모델링 및 채광 관련 소프트웨어 GEMS (http://www.geovia.com), GoCAD (http://www.pdgm.com), ISATIS (http://www.geovariances.com) 등 다수의 소프트웨어들이 목적에 따라 크리깅 기능을 제공하고 있으나 수계를 따라 분포하는 변수를 예측할 수 있는 기능을 제공하지 못한다.

환경 분야 및 지형수문학 분야에서는 수계에서의 변수 예측을 위한 크리깅 분석 사례가 보고되고 있다. Yuan (2004)은 하천에서의 수질을 예측하기 위해서 크리깅 기법을 적용하였고 Smith et al. (1997)은 다수의 수질 관측소 자료를 이용하여 미국 전역의 수질 오염농도를 예측하였다(Fig. 2.2). 그러나 크리깅 적용 과정에서 일반적인 유클리디안 거리를 이용하여 분석을 수행하는 문제점이 있다. Dent and Grimm (1999)은 하천 상에서의 거리를 이용하여 수질의 분포 패턴을 분석하였고, Torgersen et al. (2004)은 물고기의 분포를 분석하기 위해서 하천 상에서의 거리를 이용하였다. 그러나 이를 이용한 크리깅 예측까지는 수행하지 않았다. Little et al. (1997)과 Rathbun (1998)은 하천에서의 거리를 이용하여 수질에 대한 크리깅 예측을 수행하였으나 지형적인 요인을 고려할 수 없고 하천의 네트워크에 대해서는 예측 방법을 적용할 수 없었다. Skøien et al. (2006)은 하천의 네트워크에 대한 미지값을 예측하기 위해 하천의 집수구역을 고려한 크리깅 방법을

제안하였으나 하천의 중심점만을 고려하기 때문에 하천 줄기 내에서의 변화를 예측할 수 없다는 한계가 있다(Fig. 2.3). Ver Hoef et al.(2006)은 유량을 고려한 구역크리깅을 적용하여 하천 네트워크에서의 오염물 함량을 예측하였다. 그러나 지금까지 보고된 수계에 대한 크리깅 적용 사례들은 대부분 선(line) 객체로 이루어진 벡터자료 형태의 하천을 대상으로 분석이 이루어졌다. 따라서 래스터 형태의 수치표고모델로부터 하천을 추출하는 방식에 비해서 지형적 요인을 고려하기 어렵고, 하천에서의 샘플들 간 거리를 개별 객체로서 구축해야 하기 때문에 소프트웨어 개발을 통한 자동화된 분석 사례들은 보고되지 않았다.

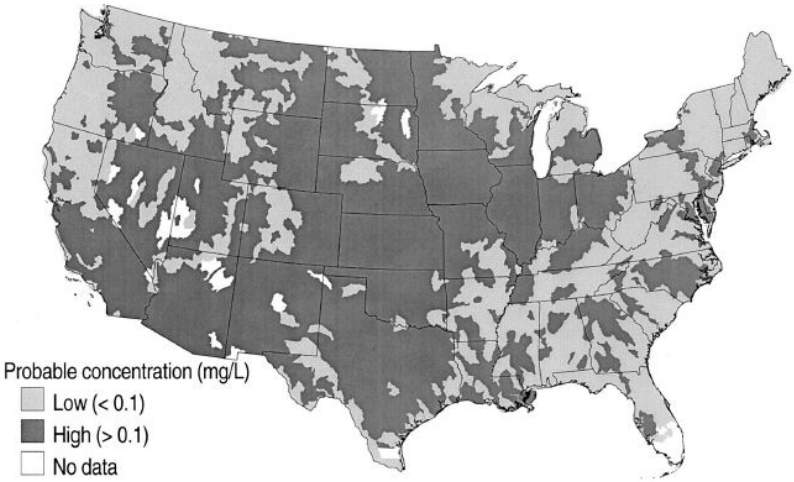


Fig. 2.2. Classification of predicted total phosphorus concentrations in hydrologic cataloging units of the conterminous United States (Smith et al., 1997).

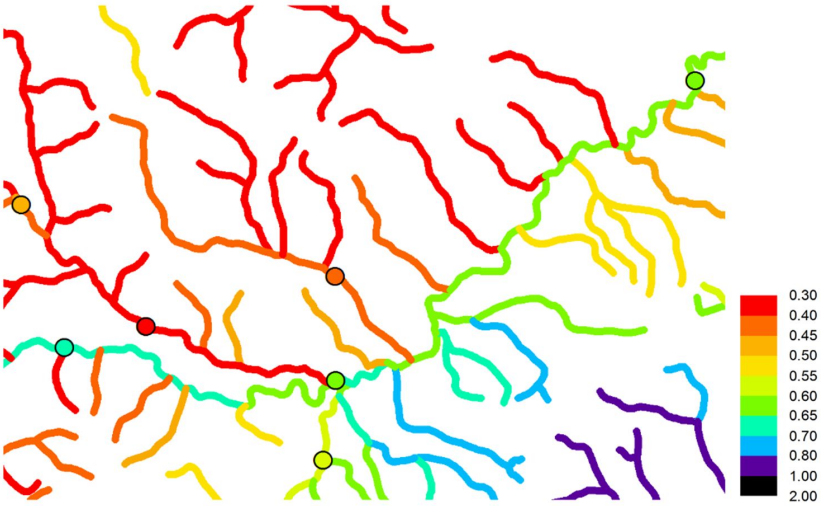


Fig. 2.3. Estimates of the normalised specific 100-year flood Q100N on the stream network of the Mur region (Skøien et al., 2006).

3. GIS 기반의 수계분석을 결합한 크리깅 기법 개발

본 연구에서 제안하는 크리깅 기법은 자료들의 공간적 위치만을 고려하는 기존의 크리깅 기법과 달리 대상지역의 수계흐름을 고려하여 설계되었다. 중금속 오염물이 유실되어 분산되는 과정에서는 주로 물이 흐르는 방향을 따라 오염물이 이동하기 쉽기 때문에 시간적인 변화가 적은 하상퇴적물은 대상지역의 오염경로를 분석하기에 적절한 지구화학적 매개체이다. 본 연구는 하천의 흐름을 오염원의 주요 이동수단으로 가정하기 때문에 이러한 하상퇴적물 샘플을 대상으로 적용하기에 적합하다. 하상퇴적물과 주변 토양은 위치적으로 가깝더라도 오염농도가 상당히 다른 양상을 보일 수 있기 때문에 토양샘플과 하상퇴적물 샘플을 하나의 데이터 세트로 이용할 경우에는 중금속 오염농도 예측 과정에서 실제 오염분포를 왜곡할 가능성이 있다.

Fig. 3.1은 전체적인 분석의 흐름을 보여주고 있으며 각 과정에서 필요한 입력 자료를 표시하였다.

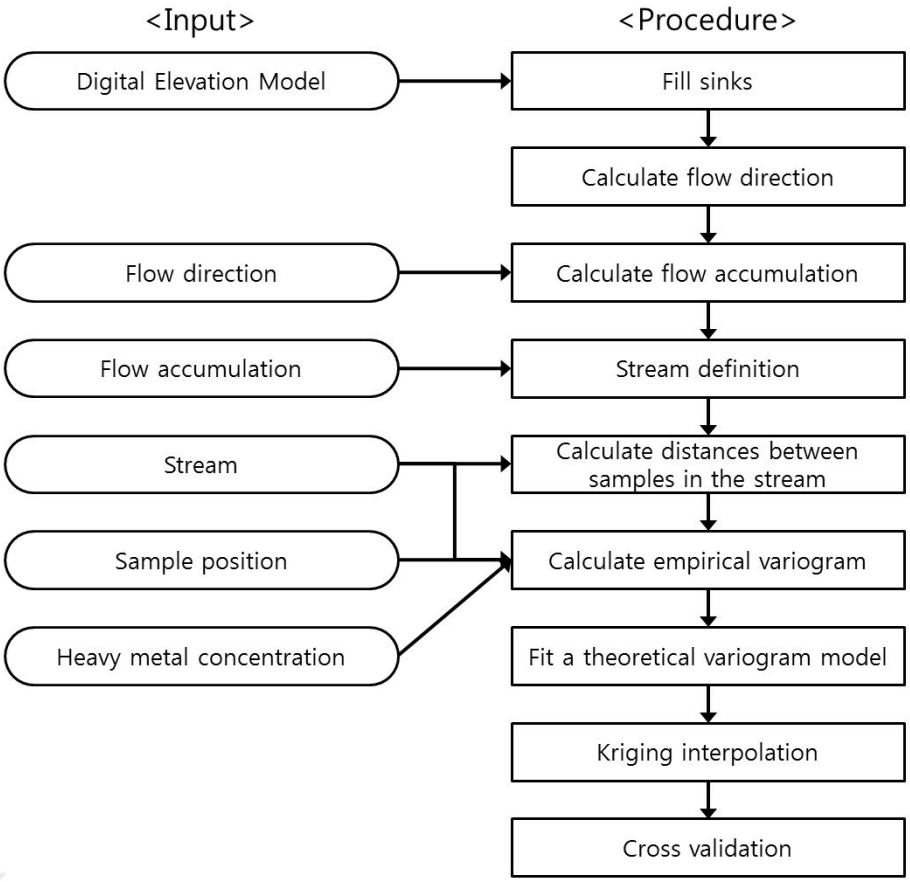


Fig. 3.1 Input data and overall procedures of the prediction of heavy metal concentration considering the stream path.

Fig. 3.2는 실제 하천을 전산자료의 형태로 변환하여 크리깅 분석을 수행할 때 하천의 경로를 고려하는 방법과 일반적인 크리깅을 적용하는 방법의 차이점을 보여준다. Fig. 3.2(a)에서 실선은 중금속 분산에 의해 오염된 하천, 점선은 오염되지 않은 하천을 의미한다. 진한색 원은 오염된 하천에서 취득한 샘플, 흰색 원은 오염되지 않은 하천에서 취득한 샘플을 의미한다. 격자구조는 래스터 형태의 GIS 데이터를 의미하며 진한 격자는 오염농도를 알고자 하는 위치에 해당한다.

Fig 3.2(b)는 이러한 실제의 자연물 및 데이터를 격자형태의 래스터 자료로 변환한 결과를 보여준다. 포인트 자료인 진한색의 원(오염된 샘플 데이터)을 래스터 자료로 변환하면 빗금친 격자와 같이 변환이 되고, 실험적 베리오그램 계산을 위해서 데이터 간의 거리를 구하면, 오염된 샘플과 관심 지역 사이의 거리는 두 샘플의 경우 모두 ($\sqrt{2}$×격자크기)의 값을 가진다. 하지만 하천의 경로를 고려한다면 오른쪽 샘플이 왼쪽 샘플보다 관심지역과 더 가깝기 때문에 오염농도 예측에 있어 더 큰 영향을 줄 것이다.

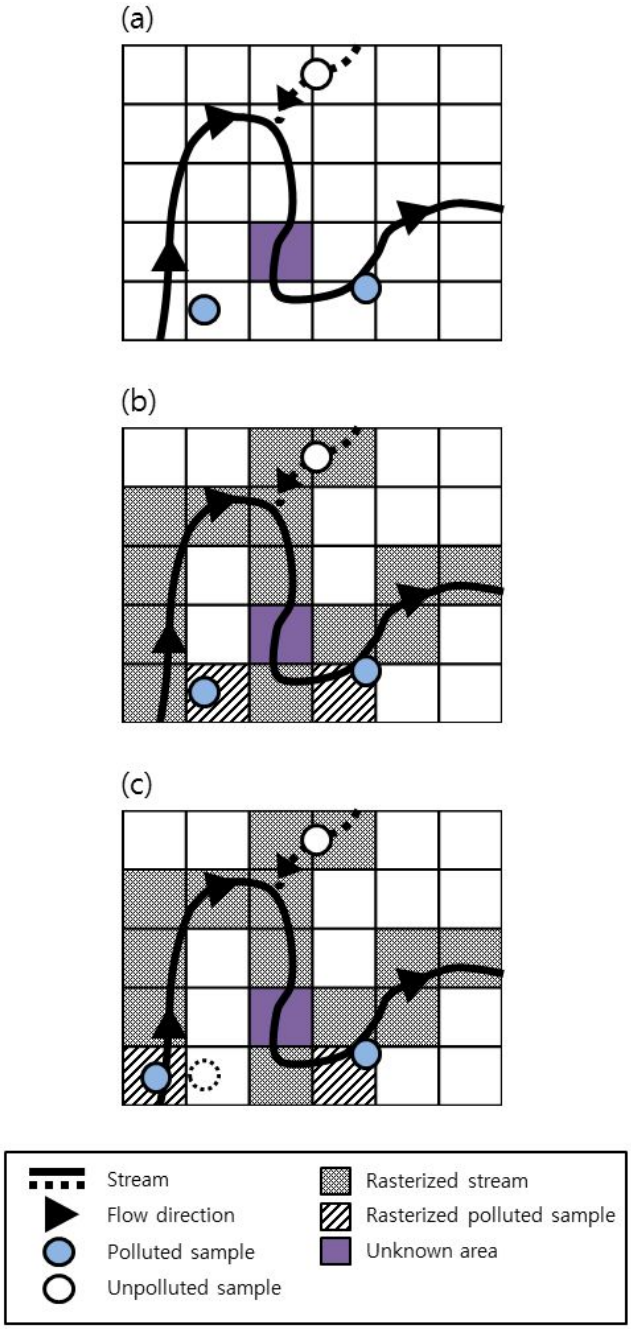


Fig. 3.2. Rasterization of (a) unknown area, (b) samples and stream paths, and (c) modification of sample location to stream.

하천의 흐름에 의한 영향을 고려하기 위해서는 먼저 GIS 공간상에서 하천의 위치와 하상퇴적물 샘플의 위치를 정확하게 대응시키는 과정이 필요하다. GIS 기반의 수계분석을 통해서 하천의 경로를 얻은 후에 샘플의 위치가 이 경로에서 벗어날 경우 흐름방향을 따라 하천을 대표하는 가장 가까운 격자의 위치로 샘플을 이동시킨다(Fig. 3.2(c)). 오염지역의 샘플 위치가 오염경로에서 벗어나는 이유는 다음과 같은 경우가 있을 수 있다. 첫 번째로 샘플의 위치가 정확하게 측정이 되지 않았을 수 있고, 두 번째로 수계분석을 위한 수치표고모델 데이터의 정확도가 낮거나 해상도의 문제가 있을 수 있다. 세 번째로 D8 알고리즘 기반의 분석을 통해 얻은 하천은 한 줄로 나타나기 때문에 격자의 크기가 실제 하천의 너비보다 작을 경우에는 실제의 하천을 모두 포함하지 못하게 된다. 이러한 이유들로 인해 하천의 경로에서 벗어난 샘플은 가장 가까운 하천 격자의 위치로 이동시켜야만 하천 상에서 샘플들 사이의 거리를 구할 수 있다. 모든 샘플들 사이의 거리를 계산한 후에는 실험적 베리오그램을 구할 수 있으며 분석자의 판단 하에 실험적 베리오그램을 잘 대표할 수 있는 이론적 베리오그램을 구하는 과정이 필요하다. 오염농도를 예측하려는 지점과 측정 샘플과의 거리, 이론적 베리오그램을 알고 있다면 크리깅을 적용하여 하천 경로 상의 중금속 오염농도 예측값을 구할 수 있다. 예측의 정확도 향상을 위해서 다른 변수를 함께 고려한 다변량 크리깅을 적용할 수 있으며, 본 연구에서는 GIS 분석을 통한 샘플의 집수구역을 고려하는 방법을 제시한다.

3.1. 지형을 고려한 하천경로 분석

GIS 분석을 통해서 하천을 추출하기 위해서는 먼저 수치표고모델을 이용하여 대상지역의 흐름방향을 정의할 필요가 있다. 이 때 수치표고모델은 데이터 취득과정에서의 부정확성이나 오류, 또는 실제 지형의 특성으로 인해 8개의 모든 주변격자보다 높이가 낮은 함몰격자(Sink)가 생길 수 있다(Raaflaub, 2002). 본 연구에서 적용한 D8 알고리즘의 경우 함몰격자가 있는 지역의 흐름방향을 정의할 수 없기 때문에 채움 연산(Jenson and Domingue, 1988)을 이용하여 전처리 작업을 수행하였다.

8방향 중 한 방향으로 물이 흐른다고 가정하는 D8 알고리즘은 매트릭스 형태의 격자구조에서 중심격자와 중심격자 주변에 있는 8개 격자 사이의 기울기를 각각 계산한 후 그 중 하강하는 방향으로 경사도가 가장 큰 방향을 빗물이 흐르는 방향으로서 정의한다. 경사도를 구하기 위해서는 동, 서, 남, 북 방향의 경우 두 격자 사이의 높이 차이를 격자의 길이로 나누어 구하고, 대각선 방향의 경우 격자길이에 $\sqrt{2}$를 곱한 값으로 높이 차이를 나누어 구한다. 예를 들어 Fig. 3.3(a)와 같은 수치표고모델에서 높이 5의 값을 갖는 중심격자로부터 하강하는 방향은 중심격자보다 낮은 높이 1, 2, 3, 4 격자를 향한 방향이며, 빗물이 흐르는 방향은 하강경사도가 3으로 가장 큰 값을 갖는 남쪽 방향이다(Fig. 3.3(b)). 본 연구에서는 동쪽부터 시계 방향으로 1, 2, 4, 8, 16, 32, 64, 128 값을 흐름방향으로서 할당하는 방법을 적용한다(Fig. 3.3(c)). 흐름방향 분석을 반복 수행하면 연구지역의 모든 격자에 대해서

물이 흐르는 방향을 8방향 중 하나로 분석할 수 있다(Fig. 3.3(d)).

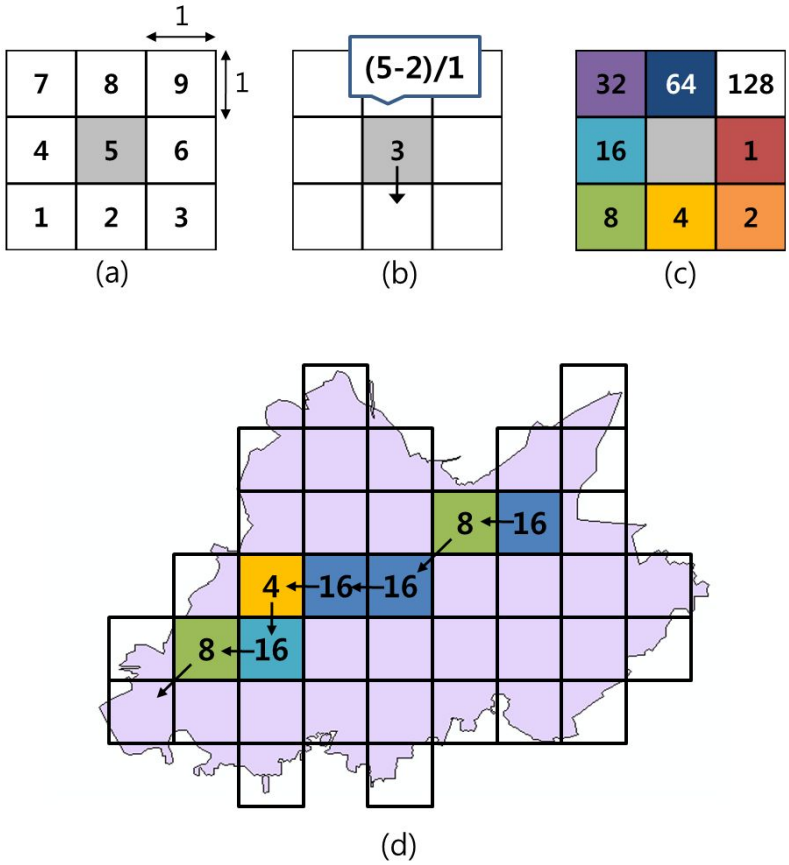


Fig. 3.3. Grid-based concept of (a) DEM, (b) maximum downhill slope, (c) flow direction and (d) flow path.

대상지역의 흐름방향을 알면 이로부터 각 격자의 누적흐름량을 구할 수 있다. 각 격자의 누적흐름량은 해당 격자로 흘러들어오는 격자의 개수를 나타내며 값이 클수록 물의 흐름이 집중되는 지역을 나타낸다고 볼 수 있다. 예를 들어 Fig. 3.4(a)와 같은 수치표고모델이 있다고 하면 누적흐름량은 Fig. 3.4(b)와 같이 계산될 수 있다. 누적흐름량을 계산한 후에는 누적흐름량이 특정값 이상이 되는 지역을 하천으로 정의하여 하천을 추출할 수 있다. Fig. 3.4(b)에서 누적흐름량이 7 이상인 값을 하천이라고 정의한다면 진한 색으로 표시한 격자와 같이 하천이 정의되며, 누적흐름량이 5 이상인 값을 하천이라고 정의한다면 빗금친 격자와 같이 하천이 정의된다(Fig. 3.4(c)). 이러한 방식으로 하천을 추출할 경우 대체로 실제 형성된 하천과 유사한 결과를 나타내지만 실제 하천의 분포와는 차이가 있을 수 밖에 없다. 극단적인 예로 물이 부족한 국가에서는 하천이 있을 것으로 분석된 지역에 물이 전혀 없을 수도 있다. 따라서 정확한 분석을 위해서는 하천을 정의하는 값을 조절하면서 실제 하천의 형성과 유사한 결과를 보이도록 분석을 수행할 필요가 있다.

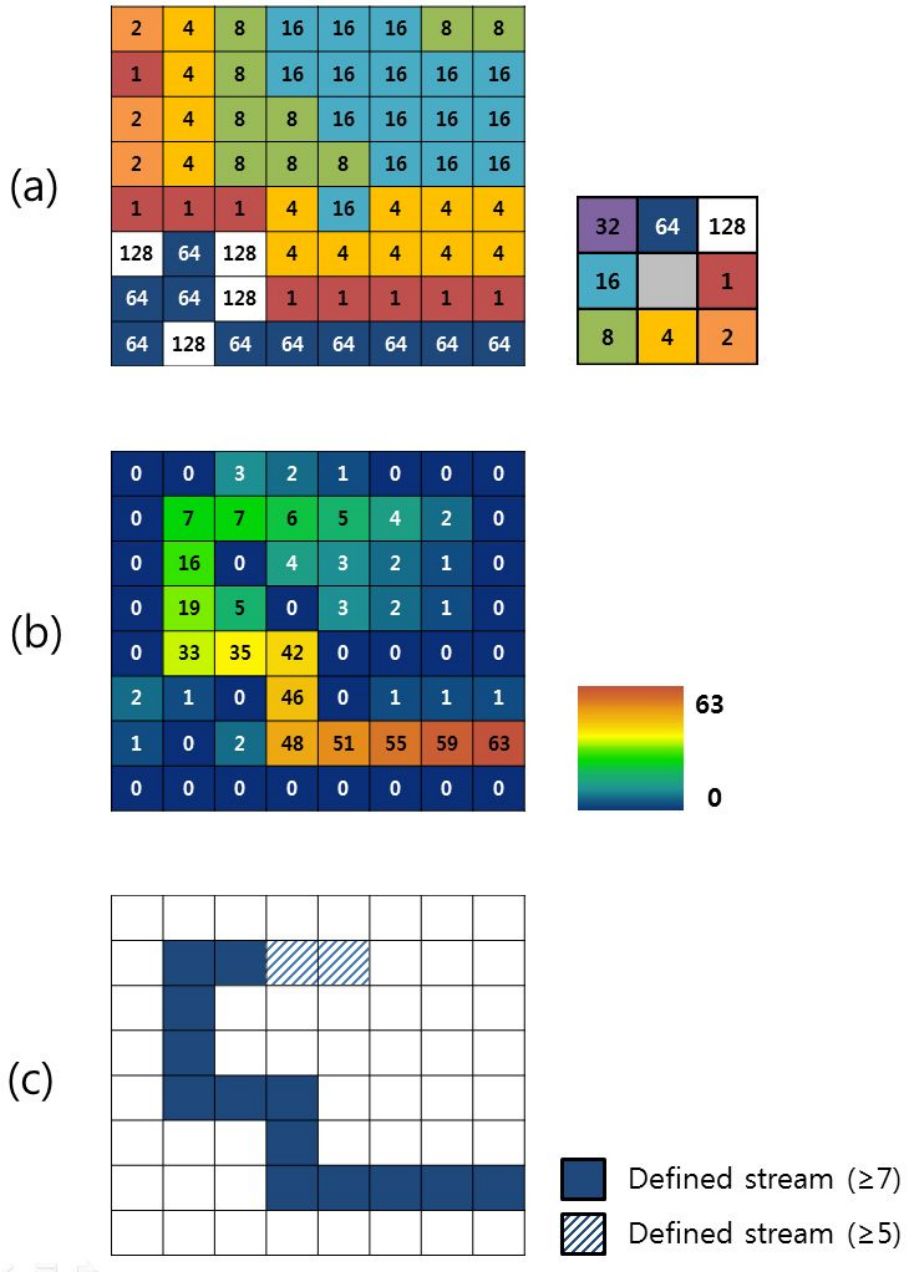


Fig. 3.4. Grid-based (a) flow direction, (b) flow accumulation and (c) defined stream.

3.2. 하천경로를 고려한 샘플 간의 거리분석과 크리깅

크리깅을 적용하여 미지의 값을 예측하기 위해서는 오염농도의 분포를 잘 나타내는 베리오그램을 구해야 하며, 베리오그램을 구하기 위해서는 모든 샘플 사이의 거리를 계산하여야 한다. Fig 3.5는 흐름방향을 고려하여 샘플들 간의 거리를 계산하고 크리깅을 수행하는 전체적인 알고리즘 순서도를 보여준다.

첫 번째 과정으로 샘플들 간의 거리를 계산하기 위해서 대상지역의 흐름방향과 샘플들의 위치를 나타내는 자료가 필요하다. 샘플의 위치가 하천 위에 있다면 흐름방향을 따라 거리를 누적시켜가며 다른 샘플을 만났을 때의 누적된 거리를 계산한다. 일반적인 크리깅을 적용할 경우에는 모든 샘플에 대해서 거리쌍이 존재한다. 예를 들어 Fig 3.6(a)와 같이 4개의 샘플이 존재한다고 하면 자기 자신과의 거리 0까지 고려할 경우 총 16개의 거리쌍이 존재하게 된다. 반면 하천경로를 고려하여 샘플 간의 거리를 계산할 경우에는 하천의 네트워크를 고려하여야 한다. Fig. 3.6(b)와 같이 하천의 네트워크가 구성되어 있을 경우에 1번 샘플의 오염물이 2번 샘플, 3번 샘플에 영향을 줄 수 있기 때문에 크리깅 분석 과정에서 서로의 거리를 계산하는 것이 합리적이다. 반면 4번 샘플은 3번 샘플을 제외한 1, 2번 샘플과는 상관성이 없다고 볼 수 있기 때문에 본 연구에서는 이러한 샘플들 간의 거리를 계산하지 않는다.

두 번째 과정으로 샘플들의 거리와 그에 따른 오염농도 값의 차이를 이용하여 실험적 베리오그램을 구한다. 이 때 분리거리를

어떻게 설정하는지에 따라서 베리오그램의 분포가 달라질 수 있기 때문에 오염분포의 전체적인 경향을 나타낼 수 있는 적절한 분리거리를 설정할 필요가 있다. 본 연구에서는 허용한계를 두어 완화된 분리거리를 이용함으로써 베리오그램 계산에서 제외되는 자료가 없도록 하였다.

다음 단계로 실험적 베리오그램을 이용하여 이론적 베리오그램을 구한다. 이 과정은 분석자의 주관적 판단이 가장 많이 개입되는 과정으로서 예측의 최종결과에 직접적인 영향을 미치므로 실험적 베리오그램을 잘 대표하는 것이 중요하다. 본 연구에서는 선형모델, 구형모델, 지수모델, 가우스모델, 너깃모델을 이용하여 베리오그램 모델링을 수행할 수 있도록 하였다.

최종적으로 오염농도를 모르는 지점에 대한 예측값을 구한다. 이를 위해서 샘플과 예측지점의 거리와 앞서 구한 이론적 베리오그램이 필요하다. 샘플과 예측지점의 거리를 구하는 방법은 첫 번째 단계에서 샘플들 사이의 거리를 구하는 방법을 동일하게 적용하면 된다. 다만 샘플이 예측지점보다 높은 지대에 위치할 경우에는 예측지점에서부터 거리 계산의 누적을 시작할 때 샘플을 만나지 못하므로 거리를 계산할 수 없다. 따라서 1단계로 예측지점에서 거리 측정을 시작하는 과정을 거치고 2단계에서는 1단계에서 만나지 못한 나머지 샘플에서 거리 측정을 시작하는 과정을 거쳐야 한다. 분리거리와 베리오그램으로부터 공분산을 계산함으로써 가중치가 구해지면 크리깅 방정식을 통해 미지의 값을 예측할 수 있다.

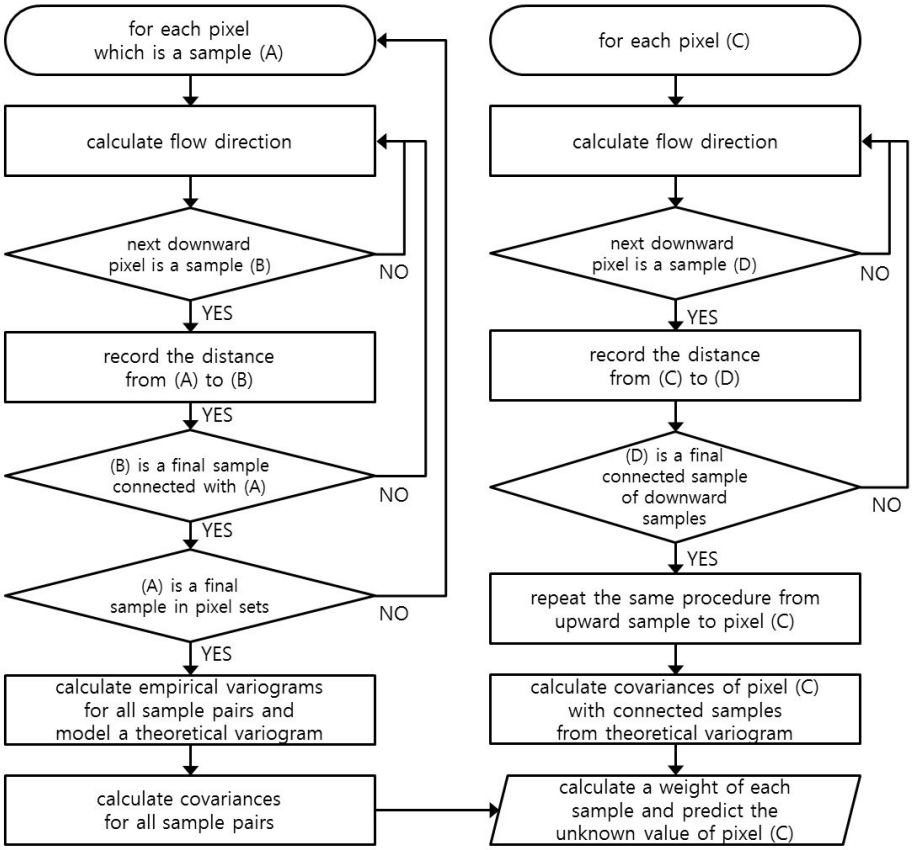


Fig. 3.5. Flowchart for the calculating the distances between samples and predicting the unknown value using kriging.

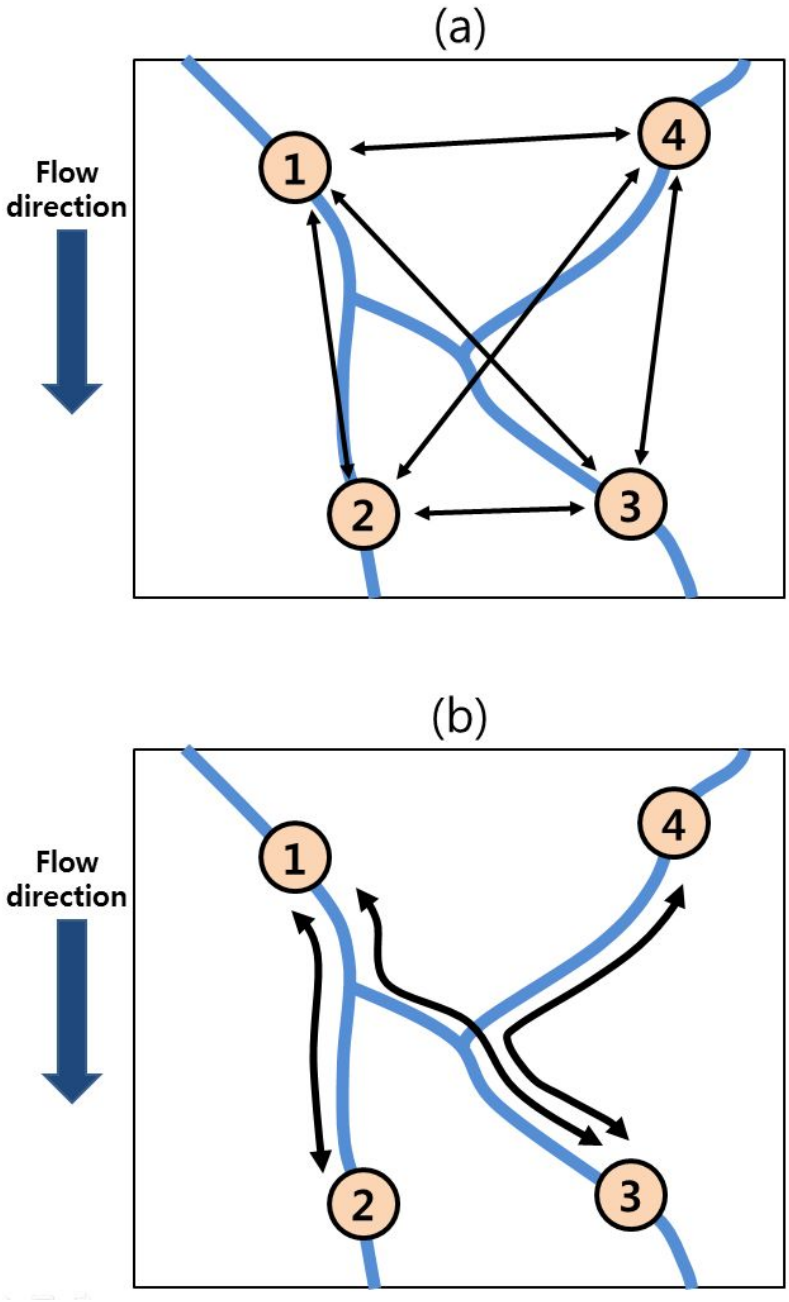


Fig. 3.6. Euclidean distance (a) and flow distance (b) between samples in the stream network.

3.3. 집수구역을 고려한 다변량 크리깅

일반적으로 공간예측을 위해서 사용되는 자료는 소수의 지점에서 취득되기 때문에 예측하고자 하는 속성과 연관성이 높은 부가자료가 있다면 표본이 없는 지점에서의 불확실성을 개선할 수 있다. 예를 들어 대상지역의 모든 위치에 대한 값을 가지는 원격탐사 자료나 지질도 등이 부가자료로서 가치가 높으며, 지하수 수위 자료나 강우 자료 또한 부가자료로 사용될 수 있다(정상용과 이강근, 1995; Goovaerts, 2000; 박노욱과 장동호, 2008). 본 연구에서는 수치표고모델로부터 대상지역 전체에 걸쳐서 계산할 수 있는 집수구역을 고려하여 다변량 크리깅 기법 중 하나인 가변적 지역 평균 크리깅(simple kriging with varying local means, SKLM)을 적용하였다.

SKLM을 적용하기 위한 과정은 다음과 같다. 우선 샘플 위치에서의 오염농도와 해당 위치의 부가자료와의 관계식을 구한다. 이 때 관계식을 구하기 위해서 회귀분석을 적용할 수 있다. 회귀분석을 통해서 얻어진 관계식으로부터 대상지역 전체의 부가자료 값을 오염농도로 변환할 수 있으며 샘플 위치에 대해서는 원래의 오염농도 값과 변환된 값의 차이인 잔차를 계산할 수 있다. 잔차는 샘플이 있는 위치에서만 구할 수 있으며, 이 잔차에 대해서 크리깅을 적용하면 연구 지역 전체에 걸친 잔차를 예측할 수 있다. 최종적인 오염농도 예측값을 구하기 위해서 부가자료로부터 변환한 오염농도 값과 잔차의 크리깅 결과를 더한 값을 이용한다.

본 연구에서는 하상퇴적물 샘플의 집수구역 면적이 오염농도와

연관성이 있다는 가정 하에 집수구역을 부가자료로 활용하여 SKLM의 적용가능성을 확인하고자 한다. 지구화학탐사를 위해 채취된 하상퇴적물 샘플은 어딘가에서 유입된 물질들이 샘플 채취지점에서 퇴적과 침식을 반복한 결과로서 이 때 물질들을 이동시키는 주요인으로는 빗물이나 하천의 흐름을 꼽을 수 있다. 물은 지형을 따라서 고도가 낮은 방향으로 흐르기 때문에 GIS를 이용한 집수구역 분석을 수행하면 대상지점으로 유입되는 물이 어디로부터 흘러왔는지 파악할 수 있다. 채취한 샘플에 영향을 주는 영역 또한 동일한 방식으로 파악할 수 있으며 집수구역의 면적과 하상퇴적물 내 원소 함량의 관계에 대한 연구가 수행된 바 있다(Bonham-Carter and Goodfellow, 1984; Carranza and Hale, 1997). Fig. 3.7은 오염지역 P-1, P-2가 존재하는 지역에 대해 집수구역의 면적과 하상퇴적물 샘플의 함량이 어떠한 관계를 가지는지 설명하기 위한 그림이다. 이 때 채취한 하상퇴적물 샘플은 집수구역에서 침식된 물질들이 오랜 시간에 걸쳐 균질하게 섞여있다고 가정한다. 따라서 동일한 양의 중금속 물질이 유입된다고 해도 집수구역의 면적이 클수록 그 외의 물질이 많이 유입되기 때문에 평균함량은 더 작은 값을 가질 것이다. 샘플 S-1의 경우 집수구역(대각선 빗금패턴)이 상대적으로 작기 때문에 오염지역 P-1에서 침식된 물질이 차지하는 비중이 클 것이고 이에 따라 샘플의 함량이 상대적으로 높은 값을 가질 수 있다. 반면 집수구역(대각선 빗금패턴과 눈금패턴)이 상대적으로 큰 샘플 S-2의 경우 마찬가지로 P-1에서 침식된 물질의 영향을 받지만 비오염 지역의 비율이 상대적으로 커서 S-1에 비해 함량값이 작을 것이다. S-4 샘플의 경우 P-1과 P-2의 영향을 모두 받지만

집수구역(Fig. 3.7의 모든 영역)이 매우 크기 때문에 희석효과가 있을 것이다.

분포를 알 수 없는 복수의 오염지역이 존재할 수 있고, 침식 및 퇴적의 영향을 정확히 알 수 없기 때문에 이러한 경향을 정량화하는 것은 매우 어려운 일이다. 그러나 동일한 오염원의 영향을 받는 구간에 대해서는 집수구역의 면적과 샘플의 중금속 함량 사이에 일반적으로 음의 상관관계가 있을 것이라고 판단할 수 있다.

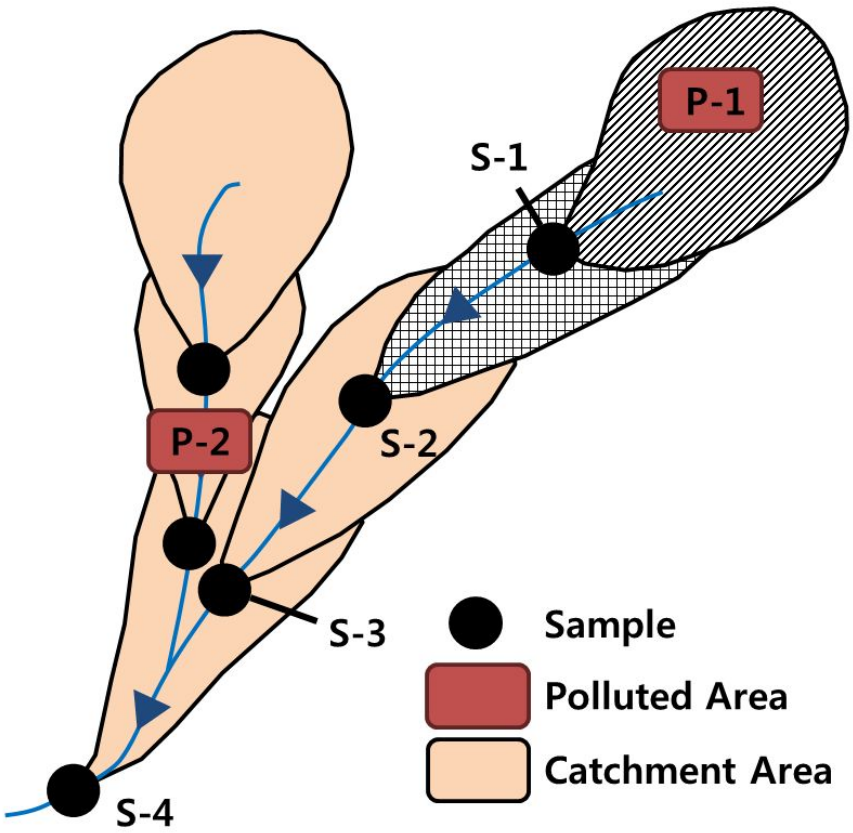


Fig. 3.7. Dilution of heavy metal pollution depending on the catchment area for each sampling location.

3.4. 소프트웨어 구현

본 연구에서는 샘플의 위치로부터 하천경로 상의 거리를 계산하여 베리오그램을 구하고 크리깅을 통한 예측값을 구할 수 있도록 소프트웨어를 구현하였다. 프로그램 코딩은 Microsoft Visual Studio 2013 플랫폼에서 작성하였으며, ESRI社의 상용 소프트웨어인 ArcGIS와 호환이 가능한 Ascii 형식의 래스터 파일로 자료를 입출력 할 수 있도록 하였다.

Fig. 3.8은 샘플 간의 거리를 계산하는 모듈로서 샘플의 위치정보를 가지는 샘플번호 레이어와 흐름방향 레이어, 하천 레이어를 입력하면 하천경로 상에서의 샘플들 사이의 거리가 아래 테이블에 나타나도록 설계하였다. 일반적인 유클리디안 거리도 계산이 가능하도록 설정할 수 있으며 텍스트 파일로 출력하여 엑셀 등의 플랫폼에서 별도의 가공이 가능하게 하였다.

Fig. 3.9는 실험적 베리오그램을 계산하고 이로부터 이론적 베리오그램을 설정 및 저장할 수 있는 모듈이다. 소프트웨어 상단에는 그래프로서 베리오그램을 시각적으로 가시화할 수 있는 기능을 제공하며 분리거리를 조정하여 오염농도의 분포를 잘 반영할 수 있는 실험적 베리오그램을 확인할 수 있다. 이론적 베리오그램을 구할 수 있도록 선형모델, 구형모델, 지수모델, 가우스모델, 너깃모델을 중복 가능하게 제공하며 실험적 베리오그램을 잘 나타내는 이론적 베리오그램을 구한 후에는 다음 단계인 크리깅 모듈에서 활용가능한 형태로 파일을 저장할 수 있다.

Fig. 3.10은 GIS 레이어와 앞서 저장한 베리오그램 파일을

이용하여 미지의 값을 예측하는 크리깅 모듈의 인터페이스를 나타낸다. 예측지점의 좌표값을 입력하면 한 지점에서의 예측값을 확인할 수 있으며, 대상지역 전체에 걸친 예측값을 래스터 형식의 GIS 레이어로 저장할 수 있다. 또한 예측 능력을 정량적으로 검증하기 위해서 교차 검증 기능을 제공한다. 교차 검증은 n개의 표본 자료에서 임의로 1개를 제거한 후에 나머지 (n-1)개의 자료를 이용하여 제거한 지점에서의 값을 유추하는 과정을 n번 반복하는 검증방법이다. 제거한 샘플의 참값과 예측값을 저장하여 비교할 수 있도록 하였으며 예측 능력의 정량적 비교를 위해 평균 오차(mean error, ME), 평균 절대 오차(mean absolute error, MAE), 제곱근 평균 제곱 오차(root mean square error, RMSE)를 제공한다.

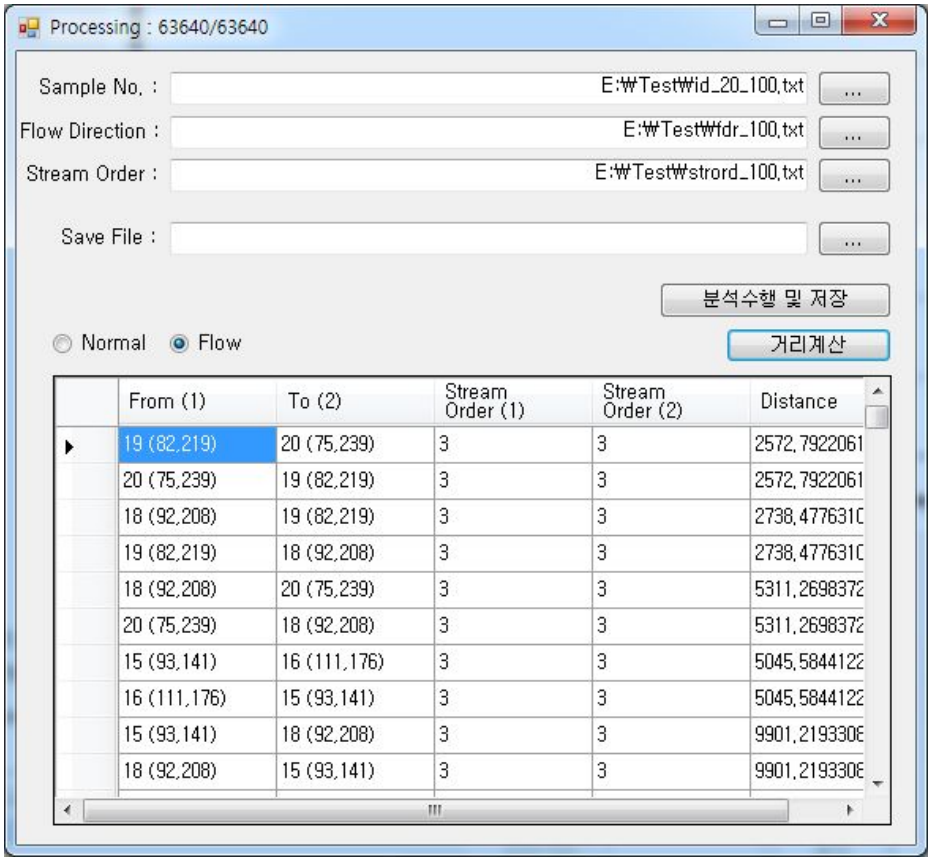


Fig. 3.8. Graphical user interface of the distance calculator module.

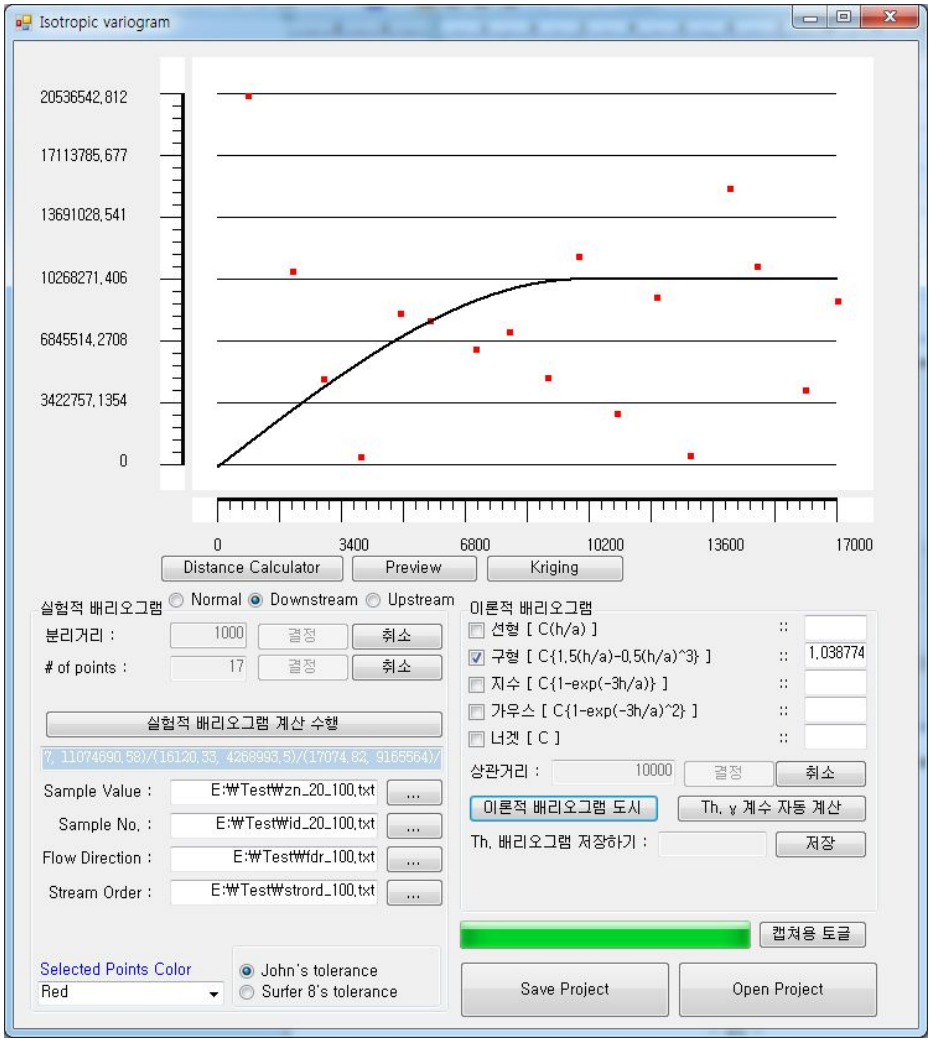


Fig. 3.9. Graphical user interface of the variogram modeling module.

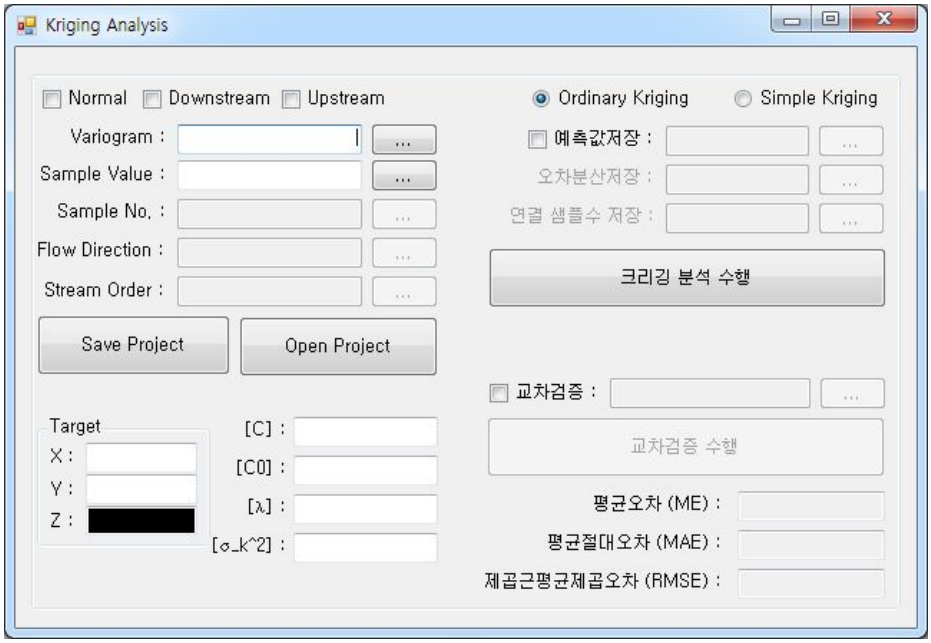


Fig. 3.10. Graphical user interface of the kriging analysis module.

4. 예측능력 검증을 위한 견본자료 분석

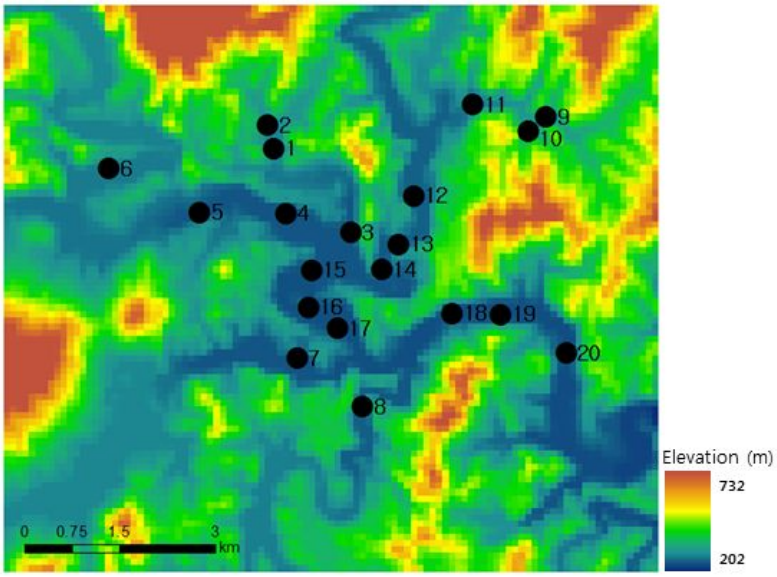
본 연구를 통해 제안하는 중금속 오염농도 예측 기법은 GIS와 지구통계학을 결합하여 하천의 경로를 고려한 크리깅 예측을 수행할 수 있도록 구축되었다.

본 장에서는 이러한 예측 기법을 통해 어떠한 결과가 산출되는지 확인하고 그 결과의 유용성을 검증하기 위한 견본자료 분석을 수행한다. 실제의 자연환경은 변수가 무수히 많기 때문에 자연환경에서 어떠한 결과값을 얻었을 때에 그 결과에 대한 인과관계를 명확히 파악하는 것은 매우 어려운 일이다. 따라서 조건을 자유롭게 변경할 수 있고 이에 따른 결과의 변화를 확인할 수 있는 견본자료 분석은 본 연구에서 제안한 예측기법의 검증을 위해 반드시 필요한 과정이다. 본 연구에서는 가상의 오염농도 데이터를 가지는 실제의 지형을 대상으로 견본자료를 구축하였다. 일반적인 크리깅 기법과 본 연구에서 제안하는 기법을 적용하고 비교 분석하며 가상의 오염농도는 토양의 유실량과 샘플의 집수구역 면적을 고려한 값을 이용한다.

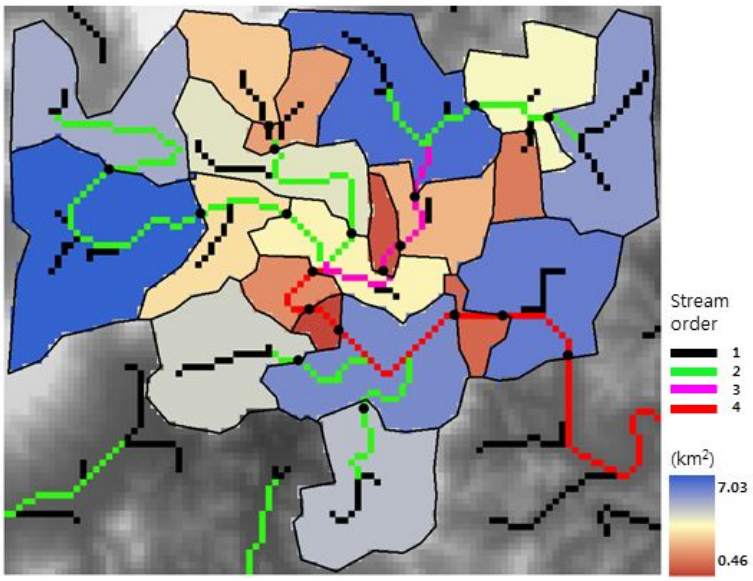
4.1. 유실량과 집수구역을 고려한 오염농도 시나리오 설계

본 연구에서 제안하는 기법의 예측능력을 검증하기 위해선 실제의 오염분포와 유사한 원리로 가상의 오염농도를 설계할 필요가 있다. 예를 들어 가상의 샘플이 아무런 규칙 없이 완전히 무작위한 오염농도 값을 가질 경우, 즉 자료들 간의 상관성이 전혀 없을 경우에는 예측의 의미가 없기 때문이다. 실제 자연환경에서 오염물이 분산되는 원인은 매우 다양하여 그 원인을 모두 알 수는 없기 때문에 본 연구에서는 수치표고모델로부터 얻을 수 있는 토양 유실량과 샘플의 집수구역을 고려하여 오염농도 분포에 대한 가상의 시나리오를 설계하였다.

먼저 Fig. 4.1(a)와 같은 수치표고모델을 가지는 지형에 대해서 검은색 점으로 나타낸 지점에서 샘플을 취득한다고 가정하였다. 이때 분석의 현실성을 높이기 위해 지형은 실제 지형을 대상으로 하여 100 미터 해상도의 수치표고모델을 사용하였다. Fig. 4.1(b)는 이 지형을 대상으로 수계분석을 수행하여 추출한 하천차수와 각 샘플의 집수구역을 보여준다. 하천차수는 하천이 만나서 합쳐지면 숫자를 더해가는 방식으로 값을 부여하기 때문에 숫자가 가장 높은 하천이 가장 집수구역이 넓은 하류의 하천을 의미한다. 면적이 상대적으로 넓은 집수구역은 파란색 계열의 색상으로 나타냈고 면적이 상대적으로 작은 집수구역은 붉은색 계열의 색상으로 나타냈다.



(a)



(b)

Fig. 4.1. (a) DEM and samples and (b) stream order and watersheds of the test area.

기존 연구들에 의하면 일반적으로 하상퇴적물 샘플은 집수구역에서 침식된 물질들이 오랜 시간에 걸쳐 균질하게 섞여있다고 가정한다. 본 장에서는 가상의 오염농도를 설정하기 위해서 Fig 4.2(a)의 타원으로 표시되어 있는 곳에 오염원이 위치하고 있다고 가정하였다.

집수구역의 크기가 동일하더라도 하상퇴적물의 근원이 되는 토양의 유실량에 따라 중금속 함량이 달라질 수 있다. 예를 들어 오염된 토양의 침식이 전혀 없다면 오염원 하류의 하천이 오염되지 않을 것이다. 본 장에서는 오염농도를 설정하기 위해서 토양침식모델인 USLE을 이용해서 각 지점의 유실정도를 고려하였다. USLE은 강우분포, 토양 유형에 따른 침식율, 경사길이와 경사도, 식생분포, 경작지 형태라는 여섯 가지 인자들로 구성되며, 각각의 인자를 곱하는 방식으로 Eqn. 4.1과 같이 구성된다.

$$A = R \times K \times LS \times C \times P \qquad \text{Eqn. 4.1}$$

여기서 A는 연평균 토양 유실량(t/ha/yr), R은 강우침식인자(10^7J • mm/ha/hr/yr), K는 토양침식인자(t/ha/R), LS는 지형인자(dimensionless), C는 식생피복인자(dimensionless), P는 침식조절인자(dimensionless)를 의미한다. 특히 지형인자(LS)는 경사길이와 경사도에 따른 토양 유실량을 결정하는 중요한 인자이다. 본 장에서는 다른 인자는 동일하다고 가정하고 지형인자 값을 계산하여 상대적인 유실량을 견본자료 생성을 위해 이용하였다. 지형인자 값을 산출하는 방법으로는 몇 가지 식들이

제안된 바 있으며 본 연구에서는 McCool(1987)의 제안식(Eqn. 4.2)을 사용하였다.

$$L = (\lambda/22.13)^m \qquad \text{Eqn. 4.2}$$

$$m = \frac{\beta}{1+\beta}$$

$$\beta = \frac{\sin\theta / 0.0896}{3.0(\sin\theta)^{0.8} + 0.56}$$

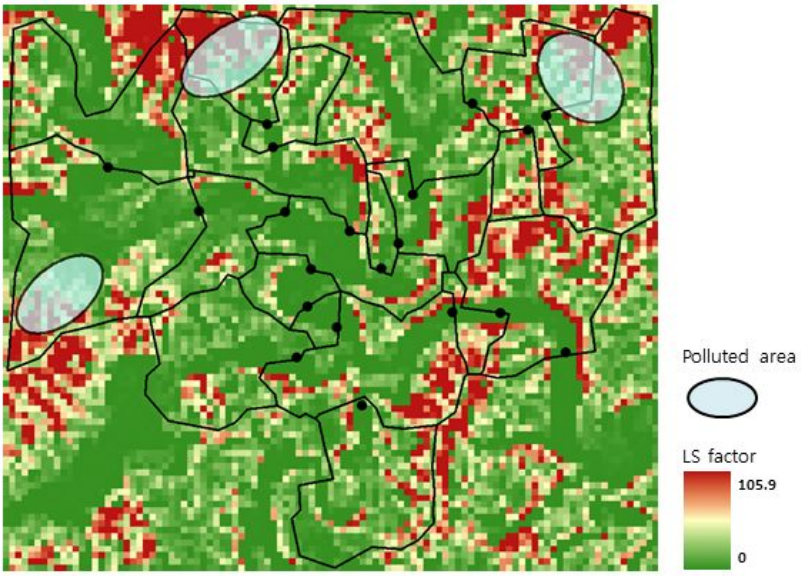
(λ=경사길이(m), θ=경사도(degree))

$$S = 10.8\sin\theta + 0.03 \quad \text{for } S < 9\% \text{ and } \lambda > 15ft$$

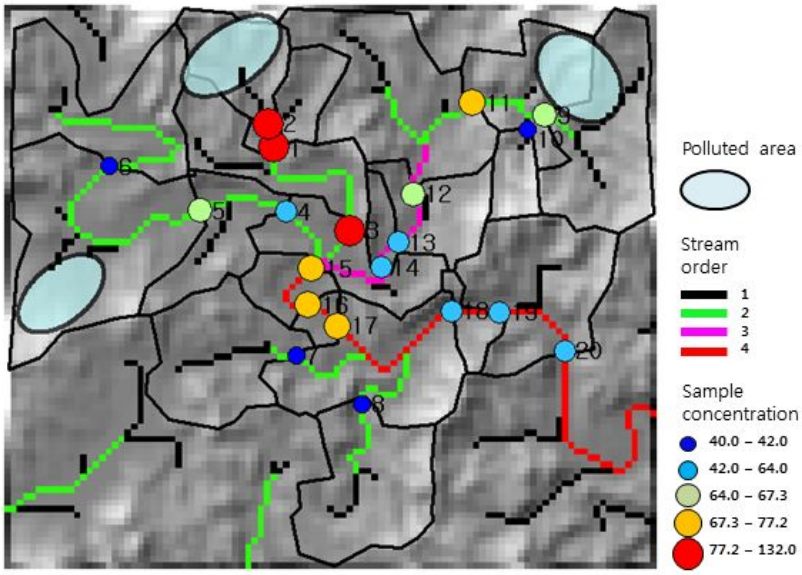
$$S = 16.8\sin\theta - 0.05 \quad \text{for } S < 9\% \text{ and } \lambda > 15ft$$

$$S = 3.0(\sin\theta)^{0.8} + 0.56 \quad \text{for } \lambda \leq 15\text{ft}$$

Fig. 4.2(a)는 견본자료의 지형인자 값을 보여주며 다른 인자는 대상지역 전체에 걸쳐 동일하다고 가정하였기 때문에 각 격자의 상대적인 토양 유실량을 의미한다고 할 수 있다.



(a)



(b)

Fig. 4.2. Distribution of (a) LS factor that represents soil loss and (b) simulated heavy metal concentration of test samples.

본 장에서 샘플들의 가상 오염농도를 생성하기 위해 수행한 전체적인 과정을 요약하면 다음과 같다.

(1) 오염토양의 중금속 농도는 200, 일반 토양의 중금속 농도는 40이라 가정하였다.

(2) 대상지역의 각 격자별 토양유실량 A_i를 구한 뒤 중금속 농도와 곱하면 각 격자에서 발생하는 중금속의 양을 반영한다. 여기서 i는 개별 격자를 의미한다.

(3) 각 샘플의 집수구역과 집수구역의 면적 W_k를 구한다. 여기서 k는 개별 샘플을 의미한다.

(4) 각 샘플의 집수구역 안에 있는 오염토양의 면적을 P_k라고 하면 집수구역 안의 오염 토양과 비오염 토양으로부터 샘플까지 유입되는 중금속의 양은 Eqn. 4.3과 같다고 가정한다. 이 때 집수구역 내 모든 지점에서 유입된 물질이 균질하게 섞여있는 것으로 생각한다.

$$(200 \times \sum_{i}^{P_k} A_i) + (40 \times \sum_{i}^{W_k - P_k} A_i) \qquad \text{Eqn. 4.3}$$

(5) Eqn. 4.3의 결과를 집수구역 W_k에 대한 전체 유실량 $\sum_{i}^{W_k} A_i$으로 나눈 값을 k 샘플의 오염농도로 계산한다.

이와 같은 방식으로 산출한 견본자료의 샘플 오염농도는 Fig. 4.2(b)와 같다. 오염농도가 클수록 샘플의 크기를 크고 붉은 색상으로 나타냈다. 이러한 오염농도가 실제 자연에서의 오염분포를 정확히 반영할 수는 없지만 자연적인 원리에 근거하여 오염원에 의해 하천이 오염되는 현상과 오염원으로부터 멀어질수록 오염농도가 감소하는 경향을 가지는 견본자료를 만들 수 있었다.

4.2. 일반적인 크리깅 분석

본 연구에서 제안하는 하천경로를 고려한 크리깅 예측기법을 적용하기에 앞서 샘플들 간의 유클리디안 거리를 고려하는 일반적인 크리깅 분석을 수행하였다. 오염농도를 설계한 견본자료는 샘플의 수가 20개이므로 자기 자신과의 거리 0까지 고려하면 총 400개의 거리쌍으로부터 실험적 베리오그램을 얻게 된다. Fig. 4.3은 견본자료에 대한 베리오그램을 나타내며 점으로 표시한 것은 실험적 베리오그램, 실선으로 표시한 것은 이론적 베리오그램을 보여준다. 여기서 Fig. 4.3(a), (b), (c)는 각각 선형모델, 구형모델, 가우스모델을 적용한 베리오그램 결과이다. 베리오그램 모델은 서로 조합이 가능하지만 본 연구에서는 각 모델의 비교를 위해 단순화하여 분석을 수행하였다. 가로축의 값이 커질수록 베리오그램 값이 증가하는 것을 알 수 있는데 이는 샘플 간의 거리가 멀어질수록 상관성이 작아지는 것을 나타낸다. 반면 어느 정도 거리가 증가하면서 베리오그램 값이 다시 작아지는 것을 확인할 수 있는데 이는 오염되지 않은 상류의 오염농도 값과 희석으로 인한 하류의 오염농도 값이 비슷하기 때문에 생기는 현상으로 판단된다. 상관거리에 따라 어떠한 결과의 차이가 있는지 확인하기 위해서 상관거리 3,000 미터와 4,500 미터 두 가지 경우에 대해서 분석을 수행하였다. 오염농도 예측을 위해서는 정규 크리깅 기법을 적용하였다.

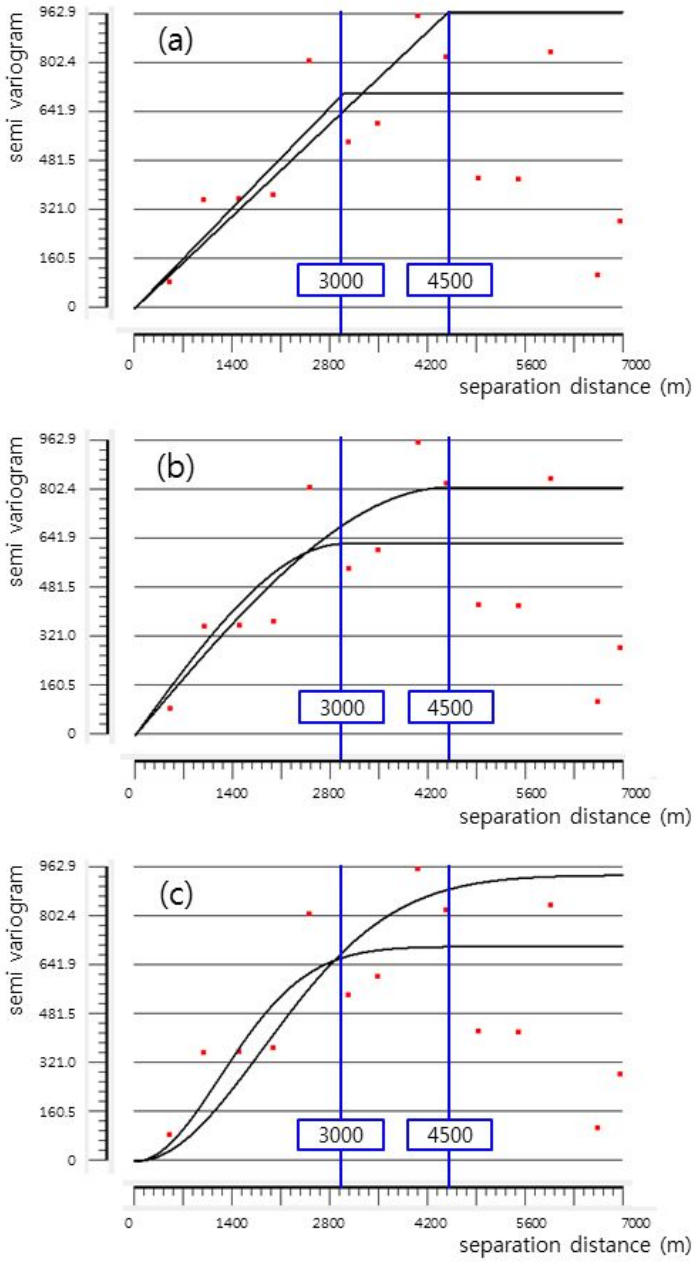
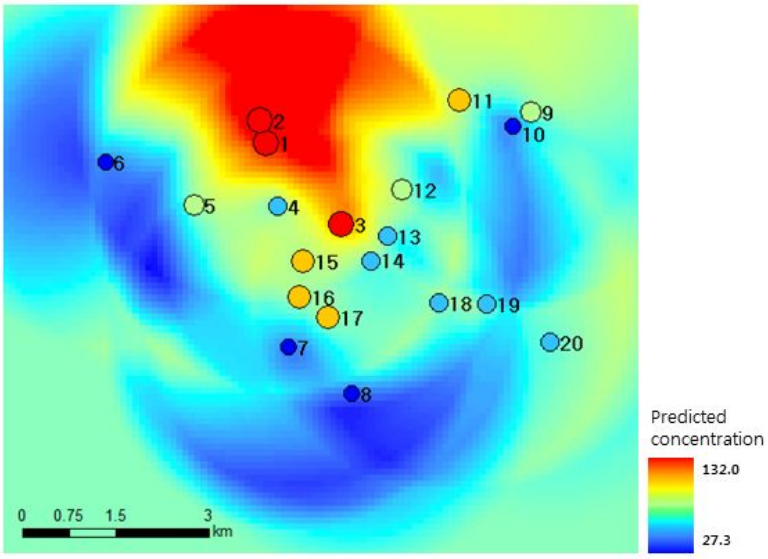


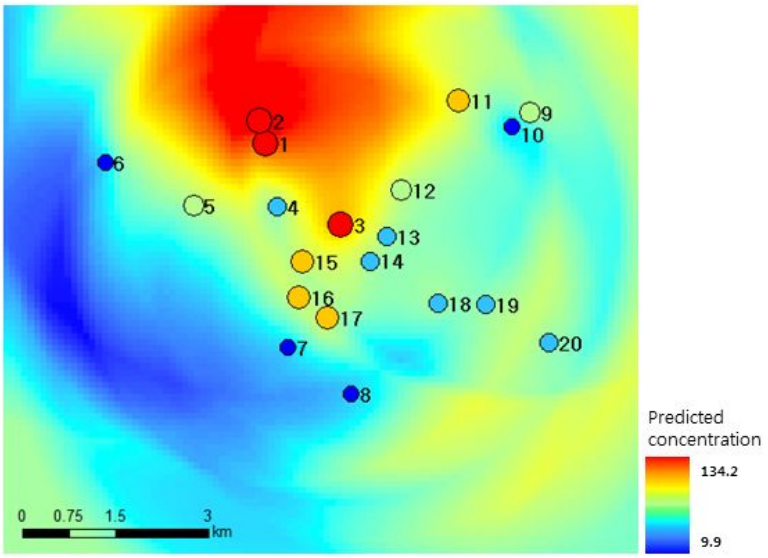
Fig. 4.3. Empirical variogram calculated from Euclidean distances between test samples and theoretical variogram models using (a) linear model, (b) spherical model and (c) gaussian model.

Fig. 4.4는 선형모델에 대해서 두 가지 상관거리를 적용하여 오염농도를 예측한 결과이다. 일반적인 크리깅을 적용할 경우 모든 지점에서의 예측값을 얻을 수 있으나 하상퇴적물을 대상으로 분석을 수행하였기 때문에 다른 토양 지역에 대한 예측값은 신뢰도가 떨어질 것이다. 선형모델을 적용한 경우에 샘플에서부터 거리가 멀어질수록 상관성이 선형적으로 감소하기 때문에 상관거리에 따른 경계가 뚜렷하게 보이는 것을 확인할 수 있다. 상관거리가 상대적으로 긴 Fig. 4.4(b)의 경우가 Fig. 4.4(a)보다 샘플에 의한 영향범위가 길게 나타나는 것을 알 수 있다.

Fig. 4.5과 Fig. 4.6은 각각 구형모델과 가우스모델에 대해서 두 가지 상관거리를 적용하여 오염농도를 예측한 결과이다. 앞서 구한 이론적 베리오그램을 살펴보면 구형모델은 분리거리가 작을 때 베리오그램이 상대적으로 큰 값을 가지므로 가까운 값의 영향을 상대적으로 덜 받을 것이라 예상할 수 있으며, 반면 가우스모델은 분리거리가 작을 때 베리오그램이 상대적으로 작기 때문에 가까운 값들에 더 큰 영향을 받을 것이라 예상할 수 있다. 이러한 경향은 Fig. 4.6(a)의 가우스모델 적용 결과에서 두드러지는 것을 확인할 수 있다. 특히 가우스모델을 적용한 경우 자료가 없는 지점에서의 예측값이 극단적인 값을 가져서 예측값의 오차가 커지는 것을 알 수 있다.

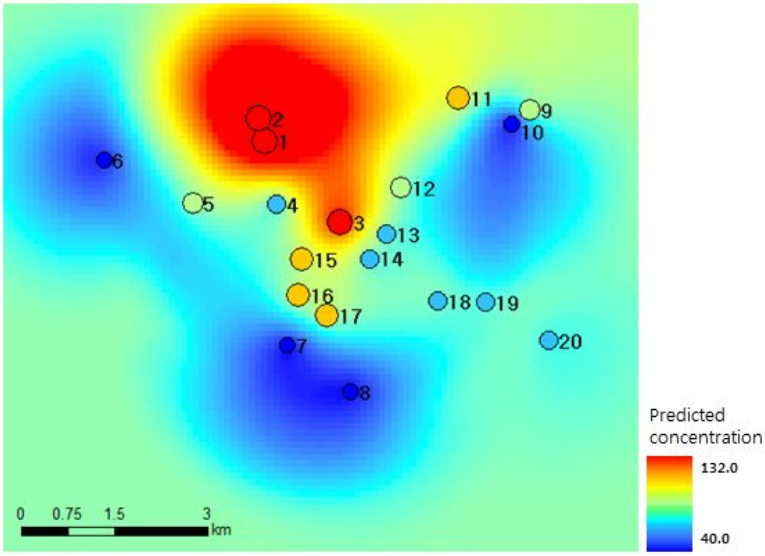


(a)

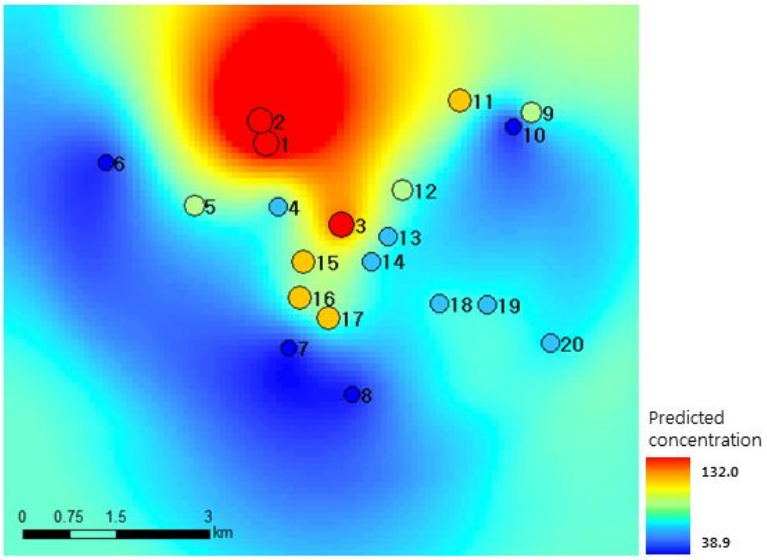


(b)

Fig. 4.4. Prediction of heavy metal concentrations for the test area using linear model with the correlation length of (a) 3,000 m and (b) 4,500 m

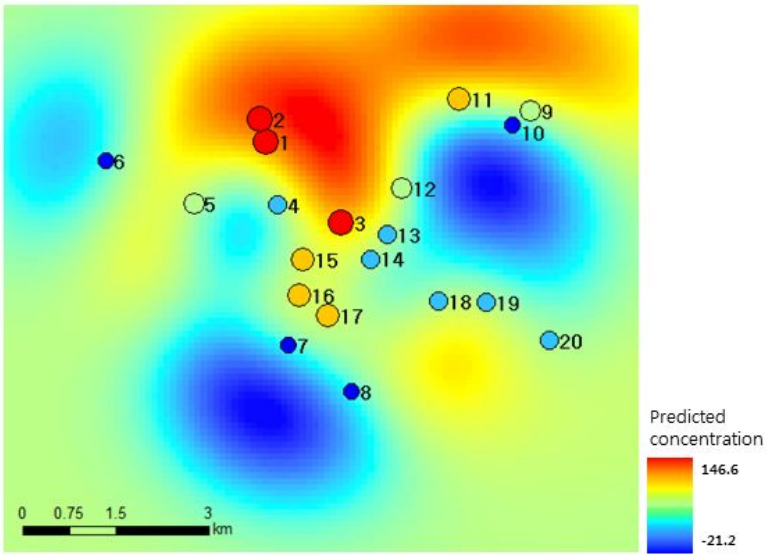


(a)

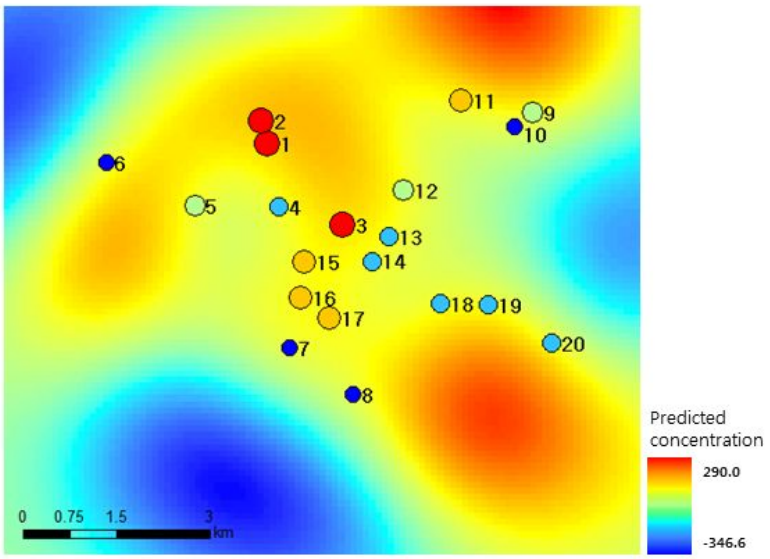


(b)

Fig. 4.5. Prediction of heavy metal concentrations for the test area using spherical model with the correlation length of (a) 3,000 m and (b) 4,500 m



(a)



(b)

Fig. 4.6. Prediction of heavy metal concentrations for the test area using gaussian model with the correlation length of (a) 3,000 m and (b) 4,500 m

적용한 베리오그램 모델과 상관거리에 따른 예측능력을 검증하기 위해서 교차검증을 수행하였다. 크리깅은 자료가 있는 지점의 값을 정확히 예측하는 특징이 있기 때문에 검증을 위해서는 자료를 제거한 뒤에 제거한 자료에 대해 예측한 결과를 원래의 값과 비교할 필요가 있다. 본 연구에서 작성한 프로그램을 이용하여 20개의 샘플에서 1개씩 샘플을 제거하고, 제거한 샘플의 예측값을 구하는 방식을 20번 반복하였다. 20개 샘플에 대한 교차검증 결과를 Table 4.1에 정리하였다. 예측 능력의 정량적 비교를 위해 평균 오차(ME), 평균 절대 오차(MAE), 제곱근 평균 제곱 오차(RMSE)를 비교하면 선형모델과 구형모델은 비슷한 예측 결과를 보이며 가우스모델은 상대적으로 오차가 매우 큰 것으로 판단된다. 상관거리는 선형모델과 구형모델의 경우 4,500 미터에서 더 좋은 예측 결과를 얻은 반면, 가우스모델은 4,500 미터에서 오차가 증가하는 경향을 보였다. 또한 6가지 모델 중 각 샘플과 가장 유사한 값을 회색 음영으로 표시하였다. 그 결과 상관거리를 4,500 미터로 설정한 구형모델이 20개 중 5개의 샘플을 가장 유사하게 예측한 것을 확인할 수 있다. 주목할 점은 주변의 다른 값과 상대적으로 오염농도 값의 차이가 큰 샘플(3, 6, 9, 10)에 대한 예측값은 참값과 상대적으로 차이가 크다는 점이다. 크리깅을 비롯한 예측기법의 특성상 교차검증 과정에서 제거한 값과 주변값의 차이가 크면 오차가 커질 수 밖에 없다. 이를 개선하기 위해서는 샘플의 수를 늘리는 방법이 가장 효과적인 수단이 될 수 있다. 이론적 베리오그램을 설정하는 단계는 분석자의 주관이 크게 작용하는 과정이며 결과에 직접적인 영향을 주기 때문에 신중한 분석이 요구되는 것을 확인할 수 있다.

Table 4.1. Cross validation results of ordinary kriging using Euclidean distance for test samples (shaded number is the most similar value to original value).

sample ID	original value	linear model		spherical model		gaussian model	
		3,000 m	4,500 m	3,000 m	4,500 m	3,000 m	4,500 m
1	117	112.2	110.7	111.8	112.8	112.3	101.1
2	132	110.4	111.8	107.4	110.6	134.7	156.7
3	101	80.1	72.1	68.0	70.1	63.9	64.5
4	62	99.0	89.7	95.8	94.5	76.1	81.2
5	65	55.1	61.8	58.4	62.9	3.4	-106.1
6	42	74.1	75.2	70.4	68.8	104.5	371.5
7	40	65.6	58.0	64.1	58.4	66.2	53.3
8	40	60.3	47.1	57.1	58.4	55.2	104.9
9	66	47.9	40.6	45.1	42.9	46.8	88.3
10	40	62.3	64.3	67.6	65.8	59.2	22.7
11	77	55.5	70.8	51.8	56.4	15.8	-79.5
12	67	74.6	85.0	70.7	78.3	98.5	116.9
13	64	63.8	70.5	70.9	68.7	53.0	44.9
14	63	65.6	74.7	73.6	72.5	76.4	78.4
15	70	74.0	67.5	74.1	75.5	71.0	65.3
16	70	56.3	69.9	60.8	61.7	67.7	74.3
17	70	55.6	58.4	58.6	57.8	55.1	43.0
18	64	81.4	52.0	62.1	56.9	36.8	12.7
19	63	47.8	55.1	60.0	61.8	81.5	118.9
20	61	68.0	73.2	68.2	63.4	48.0	-42.3
ME	-	1.8	1.75	1.2	1.3	-2.5	-0.2
MAE	-	15.8	14.2	15.3	14.3	22.8	60
RMSE	-	18.5	17.1	18.6	17.4	29.6	97.6

4.3. 하천경로를 고려한 크리깅 분석

하천 경로를 고려한 크리깅 분석을 수행하기 위해서는 일반적인 크리깅과 같이 샘플들에 대한 거리쌍을 이용하여 베리오그램을 얻게 된다. 다만 모든 샘플들 간에 거리쌍이 존재했던 일반적 크리깅과 달리 하천 네트워크가 서로 연결되지 않은 샘플들 간에는 거리를 구하지 않도록 설계하였기 때문에 거리쌍의 개수가 더 적게 계산된다. 본 연구에서 사용한 견본자료의 경우에는 총 226개의 거리쌍으로부터 실험적 베리오그램을 얻을 수 있었다. Fig. 4.7은 견본자료에 대한 베리오그램을 나타내며 점으로 표시한 것은 실험적 베리오그램, 실선으로 표시한 것은 이론적 베리오그램을 보여준다. 여기서 Fig. 4.7(a), (b), (c)는 각각 선형모델, 구형모델, 가우스모델을 적용한 베리오그램 결과이다. 대체적인 경향은 일반적인 크리깅의 경우처럼 가로축의 값이 커질수록 베리오그램 값이 증가하다가 다시 감소하는 구간이 있는 것을 알 수 있다. 다만 하천의 경로를 고려하기 때문에 샘플들 간의 최대거리를 살펴보면 유클리디안 거리를 이용한 경우보다 더 큰 값을 가지는 것을 확인할 수 있다. 상관거리 또한 일반적인 크리깅에 비해서 더 길다고 판단하여 본 연구에서는 상관거리 4,500 미터와 5,500 미터 두 가지 경우에 대해서 분석을 수행하였다. 오염농도 예측을 위해서는 일반적인 크리깅 분석과정과 동일하게 정규 크리깅을 적용하였다.

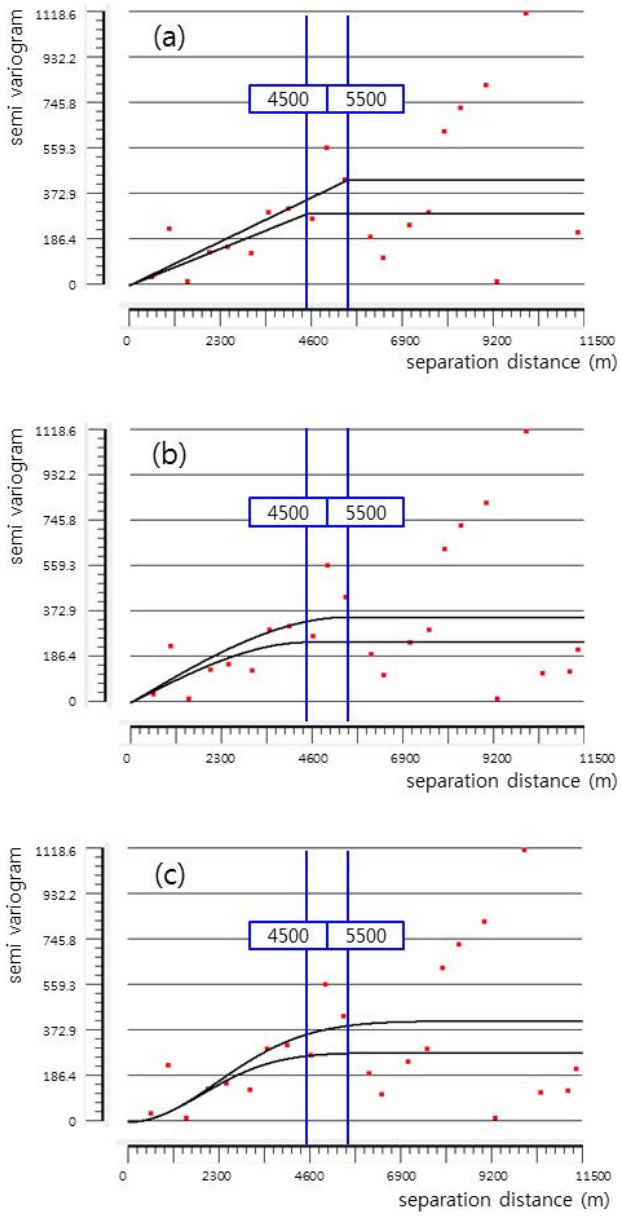
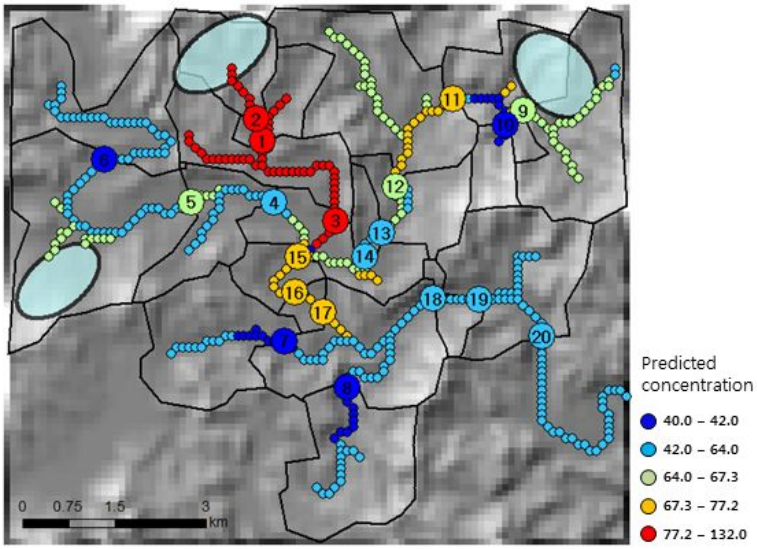
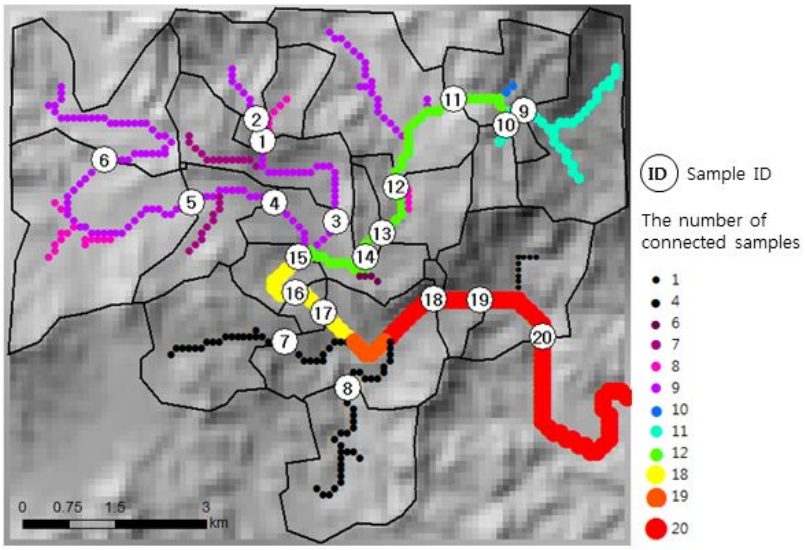


Fig. 4.7. Empirical variogram calculated from distances on the stream between test samples and theoretical variogram models using (a) linear model, (b) spherical model and (c) gaussian model.

Fig. 4.8(a)는 구형모델에 5,500 미터 상관거리를 적용하여 오염농도를 예측한 결과이다. 일반적인 크리깅과 달리 하상퇴적물을 대상으로 분석을 수행하기 때문에 전체 영역에 걸친 예측값을 얻지 않고 하천 경로 상에서만 예측값을 얻는 것을 확인할 수 있다. 샘플의 참값을 알고 있는 지점은 큰 원으로 표시하였으며, 예측지점에 대한 예측값은 작은 원으로 표시하였다. 값의 크기에 따라 색상을 달리하여 등급을 구별하였으며 샘플의 참값과 예측값의 비교를 위해서 등급을 나누는 구간은 동일하게 설정하였다. 일반적인 크리깅 기법과의 가장 큰 차이점은 하천이 만나는 지점에서의 값의 변화를 통해 확인할 수 있다. 일반적인 크리깅 기법의 경우 샘플로부터 멀어질수록 예측값이 샘플의 값으로부터 연속적으로 변하게 된다. 이러한 경향은 전체 영역에 걸쳐서 동일하게 적용된다. 본 연구에서 제안하는 하천을 고려한 크리깅 기법의 경우에도 하나의 하천 줄기 내에서는 동일한 방식으로 예측값을 구하지만, 네트워크가 연결된 샘플만 예측에 사용하기 때문에 하천이 접하는 지점에서 비연속적인 값을 예측할 수 있다. 예를 들어 오염된 하천 줄기와 오염되지 않은 하천 줄기가 근접해 있을 경우에 일반적인 크리깅은 유사한 값을 예측하지만, 하천 경로를 고려할 경우 두 줄기의 예측값에는 큰 차이가 있을 수 있다. Fig. 4.8(b)는 하천 경로 상의 예측값을 구할 때 사용된 샘플의 개수, 즉 예측지점과 하천 네트워크가 연결된 샘플의 개수를 보여준다. 예를 들어 7번 샘플이 위치한 하천 줄기는 7번 샘플을 비롯하여 하류의 18, 19, 20번 샘플 값만 상관성이 있을 것이기 때문에 연결된 샘플의 개수는 3개가 된다.



(a)



(b)

Fig. 4.8. (a) Prediction of heavy metal concentrations for the test area using flow kriging with spherical variogram model and (b) the number of connected samples of each sample.

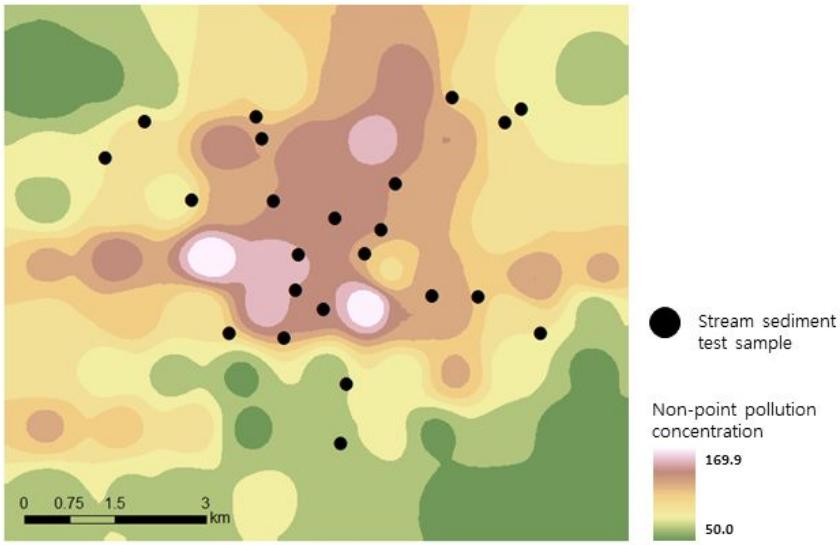
Table 4.2에 나타낸 교차검증 결과를 통해서 각 모델의 예측능력 분석결과를 비교하였다. Table 4.2에서 MAE와 RMSE를 통해서 전체적인 오차를 고려할 경우 상관거리 5,500 미터의 구형모델을 적용한 결과가 가장 우수한 예측능력을 보이는 것을 확인할 수 있으며 상관거리 4,500 미터의 구형모델을 적용한 경우에도 거의 비슷한 예측능력을 보였다. 제외한 샘플을 가장 정확하게 예측한 값이 많은 모델은 상관거리를 4,500 미터로 설정한 구형모델이다. 이 모델은 20개 중 9개의 샘플을 가장 정확하게 예측하였다. 일반적인 크리깅을 적용했을 경우 가장 우수한 모델의 MAE와 RMSE가 각각 14.2, 17.1이었던 것에 비해서 하천 경로를 고려할 경우 11.8, 15.4로 더 우수한 예측 결과를 보이는 것을 확인할 수 있다.

Table 4.2. Cross validation results of ordinary kriging using distance on the stream for test samples (shaded number is the most similar value to original value).

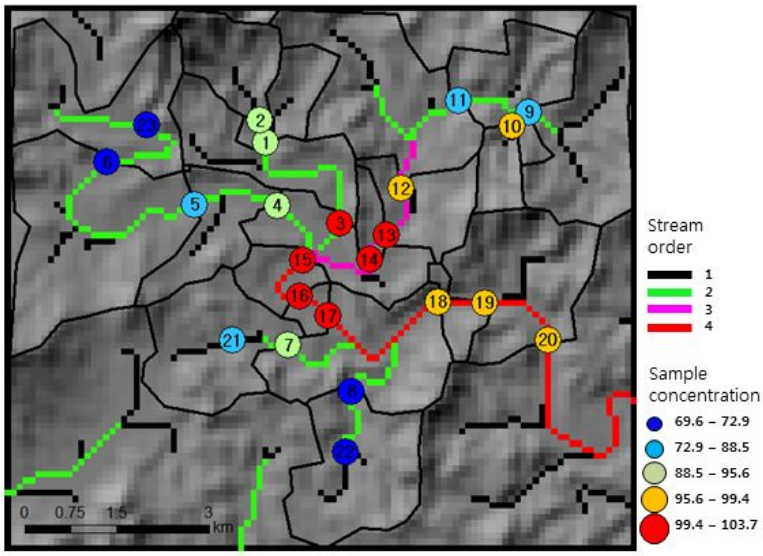
sample id	original value	linear model		spherical model		gaussian model	
		4,500 m	5,500 m	4,500 m	5,500 m	4,500 m	5,500 m
1	117	126.7	127.0	127.6	127.6	153.2	170.1
2	132	116.1	115.8	111.1	112.9	91.5	66.6
3	101	82.7	83.9	78.6	80.4	70.3	71.6
4	62	69.7	69.8	67.1	68.4	70.3	70.8
5	65	54.6	54.1	57.7	56.0	44.5	35.4
6	42	66.8	66.2	65.4	65.7	75.2	115.7
7	40	63.4	64.0	62.8	63.1	63.0	63.5
8	40	63.9	64.0	63.1	63.4	63.5	64.0
9	66	77.8	78.2	73.6	75.0	100.8	154.8
10	40	78.0	78.0	73.8	75.2	99.7	147.9
11	77	48.4	46.5	53.8	52.1	57.5	0.0
12	67	69.8	70.0	72.0	72.1	72.7	77.4
13	64	42.3	64.5	65.5	65.9	62.0	61.2
14	63	45.0	55.9	62.0	66.9	65.1	65.6
15	70	0.0	47.0	82.4	79.0	59.5	61.5
16	70	70.6	79.4	71.2	66.8	73.3	74.2
17	70	70.8	66.3	73.6	69.7	66.0	56.7
18	64	42.5	47.7	55.5	60.3	59.2	37.8
19	63	56.8	59.8	64.7	61.9	72.3	167.0
20	61	117.6	81.8	59.2	58.3	74.0	131.9
ME	-	-0.4	2.4	3.4	3.3	5.9	16.1
MAE	-	20.6	15.1	11.9	11.8	19.3	41.2
RMSE	-	26.8	18.0	15.4	15.4	24.7	53.8

앞서 생성한 견본자료는 특정한 지점에 오염원이 위치한다는 가정 하에 샘플들의 중금속 오염농도를 모델링하였다. 그러나 현실 세계에서는 특정한 지점에 오염원이 위치하기 보다는 전체 지역에 걸쳐 비점오염원의 형태로 오염원이 분포할 수 있다. 비점오염원에 의한 중금속 오염을 예측할 때에 본 연구에서 제안한 기법의 적용 가능성을 확인하기 위해서 Fig. 4.9(a)와 같은 가상의 비점오염원 자료를 사용하였다. 전체 토양에 걸쳐서 오염원이 분포하고 있다는 점을 제외하고, 샘플의 집수구역 내의 토양이 유실되어 하상퇴적물 샘플로 유입되는 과정은 앞서 견본자료를 만드는 과정과 동일하게 적용하였다. Fig. 4.9(b)는 Fig. 4.9(a)와 같은 비점오염원이 분포하고 있을 때 샘플들의 중금속 오염농도 값을 보여준다. 이 때 생성된 샘플의 오염농도는 앞서 특정 오염원으로부터 생성한 오염농도와 관련이 없는 새로운 자료이다.

비점오염원으로부터 생성한 샘플로부터 오염농도를 예측하기 위해서 일반적인 크리깅 기법과 하천의 경로를 고려한 크리깅 기법을 적용하여 비교하였다. Fig. 4.10(a)는 유클리디안 거리를 이용한 베리오그램을 나타내며, Fig. 4.10(b)는 하천의 경로를 고려한 베리오그램을 나타낸다. 이론적 베리오그램 모델은 구형모델을 적용하였으며 유클리디안 거리의 경우 상관거리 4,500 미터, 하천을 고려한 경우 상관거리 5,500 미터를 적용하여 분석하였다.



(a)



(b)

Fig. 4.9. Distribution of (a) non-point pollution source and (b) simulated heavy metal concentration of stream sediment test samples.

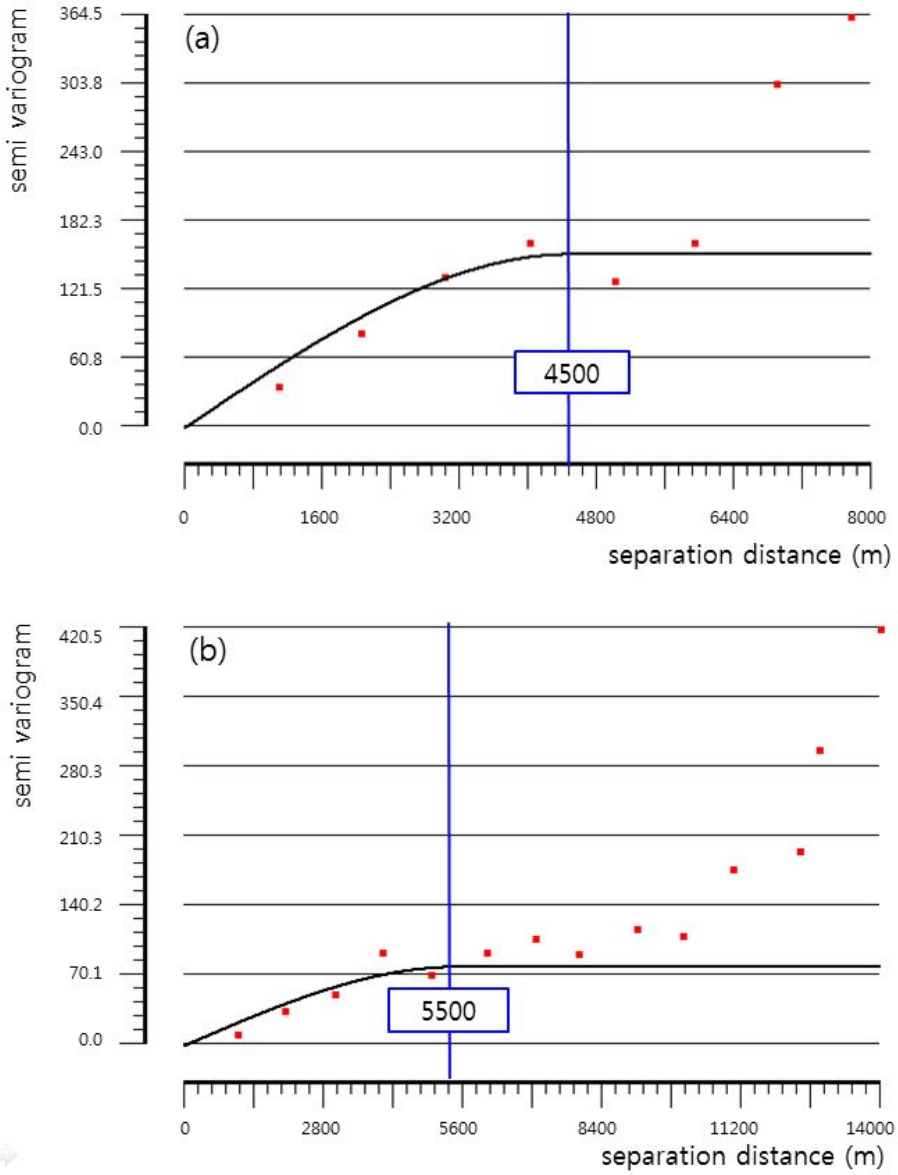
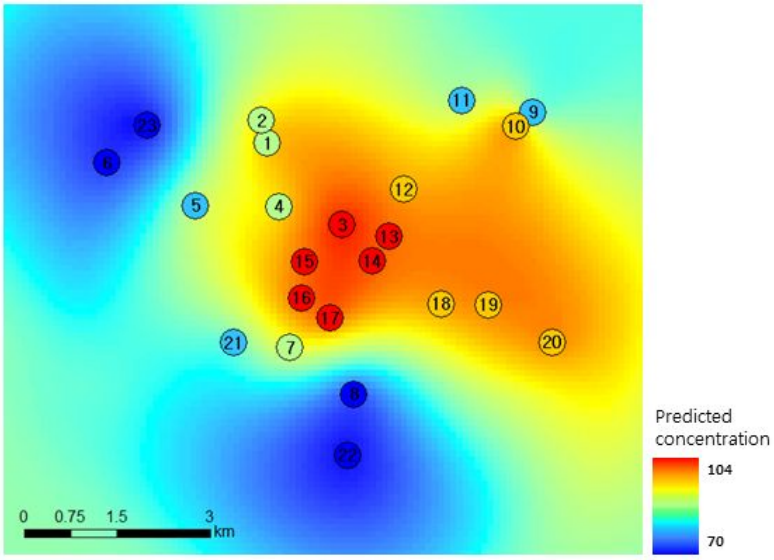


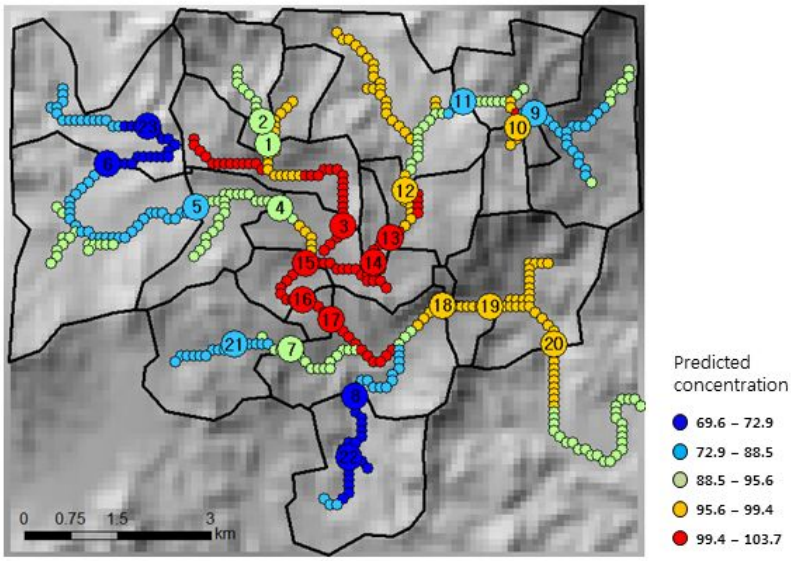
Fig. 4.10. Empirical variogram and theoretical variogram calculated from (a) Euclidean distances and (b) distances on the stream between test samples for non-point pollution source.

Fig. 4.11(a)는 구형모델에 4,500 미터 상관거리를 적용하여 일반적인 크리깅 기법으로 오염농도를 예측한 결과이고 Fig. 4.11(b)는 하천의 경로를 고려하여 5,500 미터 상관거리의 구형모델을 적용한 예측 결과이다. 각 기법을 적용한 차이는 오염원을 특정하였을 때와 유사한 결과를 보인다. 하천의 경로를 고려할 경우 일반적인 크리깅과 달리 하천 경로 상에서만 예측값을 얻는다. 샘플의 참값을 알고 있는 지점은 큰 원으로 표시하였으며, 예측지점에 대한 예측값은 작은 원으로 표시하였다.

Table 4.3의 교차검증 결과를 통해서 두 기법의 예측능력을 비교하였다. MAE와 RMSE를 통해서 전체적인 오차를 고려할 경우 두 기법에 큰 차이는 없었지만 하천의 경로를 고려한 결과가 더 우수한 예측능력을 보이는 것을 확인할 수 있다. 본 연구에서 제안한 기법은 정해진 샘플의 측정값으로부터 오염농도를 예측하기 때문에 적용대상이 하상퇴적물이라면 비점오염원을 대상으로도 적용할 수 있다. 다만 지역의 특성상 하천을 따라서 오염물이 이동하는 경향이 약하다면 제안한 기법의 활용성이 떨어질 것이다.



(a)



(b)

Fig. 4.11. Prediction of heavy metal concentrations for the non-point pollution source test area using (a) Euclidean distance and (b) flow kriging.

Table 4.3. Cross validation results of kriging using Euclidean distances and distances on the stream for non-point pollution source (shaded number is the most similar value to original value).

sample id	original value	kriging using Euclidean distance	flow kriging
1	96	94.1	**95.4**
2	94	91.2	**96.2**
3	104	**100.1**	99.1
4	94	98.7	**94.7**
5	88	83.7	**88.8**
6	72	**73.7**	78.7
7	91	**88.5**	86.3
8	73	83.2	**79.0**
9	84	97.0	**89.5**
10	99	85.2	**89.3**
11	87	93.3	**91.0**
12	97	**99.8**	95.2
13	100	101.0	**99.4**
14	101	**101.7**	100.1
15	100	**100.5**	111.9
16	100	98.2	**100.7**
17	101	94.8	**99.2**
18	98	96.3	**97.3**
19	99	**98.8**	97.8
20	99	91.4	**95.1**
21	83	**90.6**	92.0
22	70	**74.4**	76.1
23	70	**79.7**	81.0
ME	-	1.2	1.4
MAE	-	5.0	4.3
RMSE	-	7.5	5.4

4.4. 집수구역을 고려한 다변량 크리깅 분석

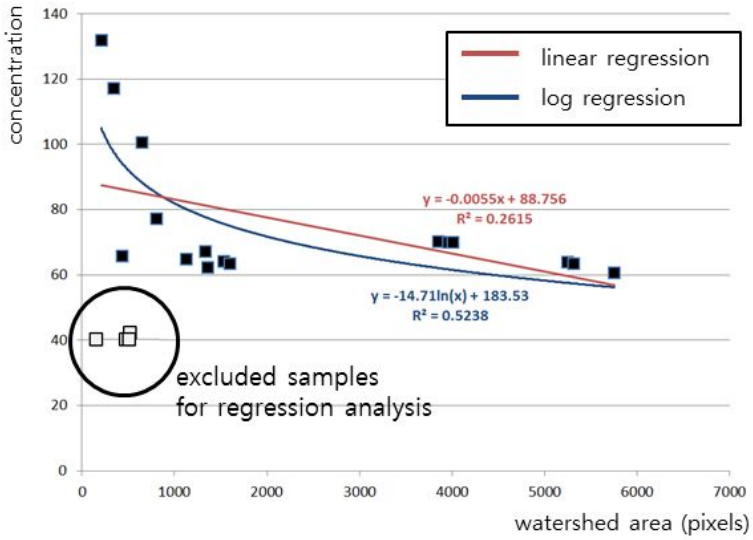
예측하고자 하는 변수와 연관성이 있는 부가자료가 있다면 다변량을 고려한 분석을 통해 부가자료만 있는 지점에서의 불확실성을 개선할 수 있다. 집수구역의 크기와 하상퇴적물의 중금속 농도 사이에는 상관성이 있다고 알려져 있으며 GIS의 수계분석 기능을 이용하면 수치표고모델로부터 전역적으로 집수구역을 구할 수 있다.

따라서 본 연구에서는 부가자료가 전역적으로 얻어진 경우에 적용할 수 있는 가변적 지역 평균 크리깅(SKLM)을 이용하여 오염농도를 예측하고자 한다. 부가자료로서 활용 가능성을 확인하기 위해 각 샘플을 비롯한 전체 대상지역의 집수구역을 구하고, 집수구역의 면적과 오염농도의 관계를 살펴보았다. 샘플의 집수구역의 면적과 오염농도를 각각 가로축과 세로축으로 하는 Fig. 4.12(a)에서 확인할 수 있듯이 두 변수는 대체로 음의 관계를 가지는 것을 알 수 있다. 그러나 집수구역이 작더라도 오염의 영향을 받지 않는 지역은 이러한 경향에서 벗어나기 때문에 샘플 6, 7, 8, 10번은 회귀 분석 대상에서 제외하고 회귀식을 구하였다.

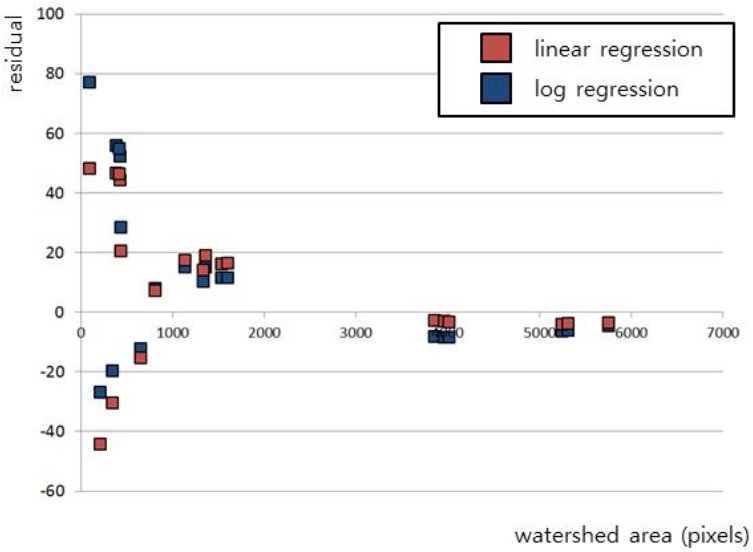
SKLM을 적용하는 과정은 회귀식을 통해 지역 평균을 계산하는 과정과 회귀식을 통한 예측값과 실제값의 잔차를 크리깅하는 과정으로 나눌 수 있다. 최종적으로는 두 값을 합산하여 대상지역에 대한 예측값을 구할 수 있다.

본 연구에서는 먼저 선형 회귀식과 로그 회귀식을 적용하여 각 샘플의 잔차를 구하였다. 이 때 집수구역의 크기에 따른 잔차의

분포는 Fig. 4.12(b)와 같다. 주로 집수구역이 작은 샘플이 많고 오염정도에 차이가 있기 때문에 집수구역이 작을수록 잔차의 편차가 큰 것을 알 수 있다. 회귀식을 각 샘플에 적용하여 구한 잔차에 대한 베리오그램은 선형 회귀식의 경우 Fig. 4.13(a), 로그 회귀식의 경우 Fig. 4.13(b)에 나타내었다. 상관거리는 7,000 미터로 설정하여 선형모델과 구형모델을 적용하였다.



(a)



(b)

Fig. 4.12. (a) Correlation between watershed areas and heavy metal concentrations of stream sediments and (b) the residual distribution according to watershed area.

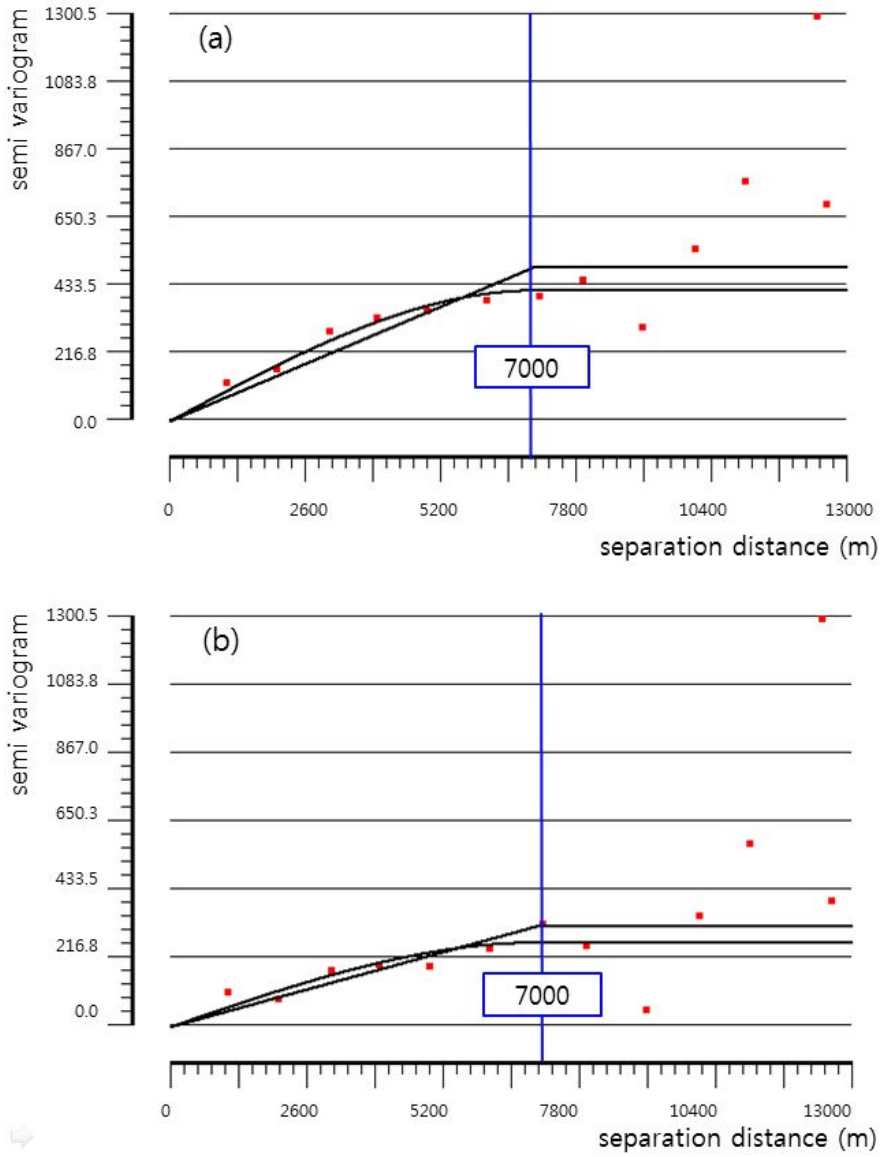
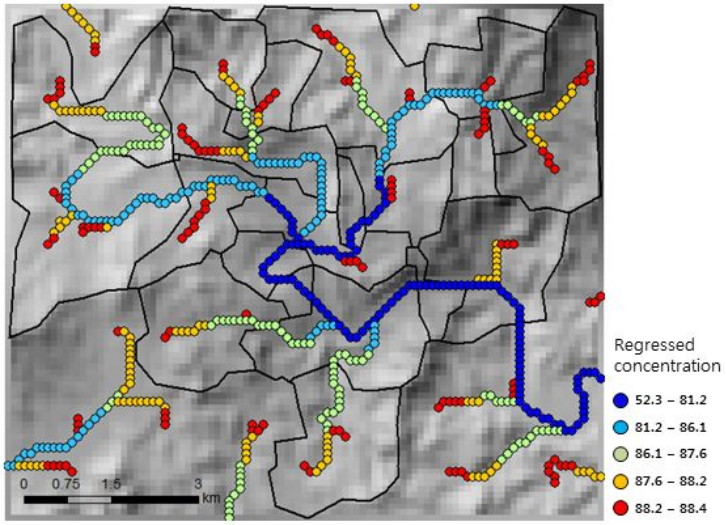


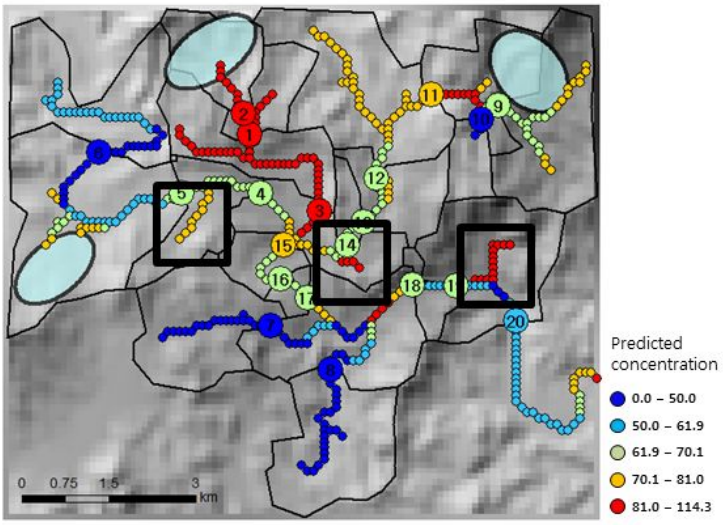
Fig. 4.13. Empirical variogram and theoretical variogram of residuals calculated from (a) linear regression and (b) log regression.

다음 과정으로 회귀 분석을 통해 얻어진 집수구역의 면적과 오염농도 사이의 관계식을 연구 지역 전체에 걸친 집수구역에 적용하여 오염농도로 변환하였다. 이 값은 각 지점의 지역평균으로서 이용된다. 각 격자마다의 집수구역의 면적 값은 누적흐름량을 구하는 것과 동일한 개념이기 때문에 누적흐름량 분석을 통해 얻은 값을 사용하였다. 선형 회귀식을 이용하여 변환한 하상퇴적물의 오염농도는 Fig. 4.14(a)와 같다. 로그 회귀식에 의해 변환된 오염농도 분포도 시각적으로 큰 차이가 없기 때문에 선형 회귀식을 적용한 결과에 대해서만 나타냈다. 집수구역의 면적과 오염농도 사이의 관계는 회귀식을 이용하여 전체지역에 동일하게 적용되기 때문에 하천의 시작점과 같이 집수구역이 작은 지점은 변환된 오염농도 값이 큰 값을 가지는 것을 알 수 있다.

최종적으로 잔차의 크리깅 결과를 집수구역의 면적으로부터 변환된 오염농도 값과 합산하면 Fig. 4.14(b)와 같은 오염농도 예측값을 구할 수 있다. 이 결과는 선형 회귀식과 구형 베리오그램 모델을 적용한 하상퇴적물 오염농도 예측 결과이다. 전체적인 경향은 집수구역을 고려하지 않은 것과 큰 차이가 없으나 사각형으로 표시한 곳과 같이 집수구역이 작은 곳의 오염농도가 과장되는 것을 확인할 수 있다. 이러한 결과는 하천 줄기 안에 측정 샘플이 없는 경우 회귀식의 영향이 크게 작용했기 때문인 것으로 생각된다. 실제로 집수구역이 작으면 오염으로 인한 영향을 크게 받을 수 있기 때문에 측정 샘플이 없는 지역에 대한 보수적 평가를 위해서는 유용하게 활용될 수 있을 것이다.



(a)



(b)

Fig. 4.14. (a) Regressed heavy metal concentrations using the correlation of watershed area and concentration and (b) prediction of concentration by the sum of regressed concentrations and predicted residuals.

SKLM을 적용한 오염농도 예측에 대한 교차검증 결과는 Table 4.4와 같다. 전체적인 오차의 크기는 보조자료를 사용하지 않았을 때보다 더 증가한 것을 알 수 있다. 특히 샘플 6, 7, 8, 10번과 같이 오염되지 않고 집수구역이 작은 상류 지역의 예측값이 전체적으로 높게 계산되는 것을 확인할 수 있으며 선형 베리오그램 모델을 적용한 결과에서는 일부 음수값이 예측된 것을 볼 수 있다. 이처럼 오차가 커진 이유로 다음과 같은 원인을 생각할 수 있다. 첫 번째로 집수구역의 면적이 비슷한 샘플이더라도 오염유무에 따라 오염농도가 확연히 차이가 날 수 있기 때문에 회귀식이 일괄적으로 적용되는 과정에서 오차가 커질 수 있다. 또한 잔차를 크리깅하는 과정에서 잔차들의 공간적 상관성이 약하기 때문에 오차가 커지는 원인이 될 수 있다. 따라서 같은 집수구역 면적을 가지더라도 오염유무에 차이가 있는 전체 하천 네트워크에 일괄적으로 적용하기보다는 오염된 하나의 하천 줄기 내에서의 예측을 할 때 더 적합할 것이라 생각된다.

Table 4.4. Cross validation results of SKLM using the correlation of watershed area and concentration for test samples (shaded number is the most similar value to original value).

sample id	original value	linear regression		log regression	
		linear model	spherical model	linear model	spherical model
1	117	124.9	125.4	**121.0**	121.1
2	132	115.5	114.0	**123.0**	122.1
3	101	94.4	92.8	**98.5**	97.4
4	62	**74.0**	76.1	74.3	76.5
5	65	56.6	**58.5**	54.5	56.5
6	42	**72.4**	81.0	80.7	89.4
7	40	88.0	**87.5**	98.9	97.7
8	40	88.2	**87.8**	98.2	97.3
9	66	**79.7**	82.9	86.2	89.5
10	40	**81.5**	84.5	109.1	112.2
11	77	-23.8	**33.4**	-100.5	8.8
12	67	69.3	58.4	**68.7**	51.4
13	64	65.8	61.5	**64.9**	58.8
14	63	68.3	**62.9**	67.6	58.8
15	70	82.6	148.7	**81.0**	432.6
16	70	63.0	**71.6**	59.3	73.2
17	70	**75.2**	77.1	80.5	81.5
18	64	-192.7	**29.8**	-255.2	20.8
19	63	58.6	66.6	56.2	**66.1**
20	61	32.6	**83.9**	25.8	90.9
ME	-	-58.5	-38.0	-14.1	-22.1
MAE	-	59.3	41.9	63.5	54.5
RMSE	-	87.2	46.7	105.1	85.5

5. 알고리즘의 현장자료 적용 및 예측결과 분석

본 장에서는 제안한 예측 기법을 현장 자료에 적용한 결과를 분석하고 현장의 중금속 오염 분포 현황을 파악하고자 한다. 먼저 대상지역의 위치, 지형적 특성, 중금속 오염의 원인 등 기본적인 정보를 살펴보고 중금속 오염농도 측정 자료를 통해 중금속별 오염현황을 검토하였다. 연구대상지역에 제안한 예측 기법을 적용하기 위하여 수치표고모델을 생성하고 오염농도 정보를 수치화하였다. 개발한 알고리즘을 수치화된 자료에 적용하여 오염농도를 예측하고 기존 기법과의 비교를 통해 제안된 알고리즘의 활용성을 검토하였다.

5.1. 연구대상지역

본 연구에서 제안하는 기법의 현장 적용성을 확인하기 위해 강원도 삼척시 풍곡리에 위치한 제2연화광산 지역을 연구대상지역으로서 선정하였다. 제2연화광산은 광물찌꺼끼 및 폐석, 침출수 등이 유출되는 복합적인 광해가 발생하면서 주변 토양 및 하천에 직접적으로 환경오염을 일으킨 사례가 있다. 제2연화광산에는 댐골 광미장과 직내골 광미장이 위치하고 있으며(Fig. 5.1), 이 곳에는 1969년부터 1996년 폐광시까지 납(Pb)과 아연(Zn) 원석을 채굴하면서 발생한 광물찌꺼기가 적치되어 있다. 특히 댐골 광미장의 하부 광미장은 태풍 ‘루사’와

‘매미’에 의해 광물찌꺼기가 유실되는 피해를 입은 바 있다(한국광해관리공단, 2007).

Fig. 5.2는 연구대상지역의 수치표고모델로서 대상지역의 지형적 특성을 보여준다. 대상지역은 지형적으로 대부분 고지대 산간지역을 형성하고 있으며 서쪽으로 태백산맥이 위치하고 있어 동쪽보다 서쪽의 표고가 높다. 태백산맥이 동쪽으로 급격히 내려오면서 동해안으로 흘러 들어가는 형태를 나타낸다. 주변에 많은 하천이 골고루 발달되어 있으며 큰 줄기는 동해안으로 향하는 특징을 지닌다.

대상지역은 태백산맥과 동해안에 위치하고 있기 때문에 해양성 기후의 특징으로 온난다습한 기후를 가진다. 2005년부터 2014년까지 최근 10년간 평균 연강수량은 1,315.1 mm이며 장마철인 7월과 태풍으로 인한 9월의 강수량이 높은 것을 알 수 있다. 특히 연강수량의 54.8%에 해당하는 평균 720.7 mm의 강수량이 여름철인 7부터 9월 사이에 집중되기 때문에 태풍과 집중호우에 의한 광물찌꺼기의 유실에 주의할 필요가 있다. 10년간 월강수량을 Table 5.1에 정리하였다.

본 연구에서는 Fig. 5.2에 위치가 표시된 샘플들의 오염농도 정보를 이용하여 지구통계학 기법을 활용한 중금속 오염농도 예측을 수행하였다. 샘플링한 시료에 대한 위치정보와 중금속 오염농도 등에 대한 정보는 2009년 7월 집중호우로 인해 발생한 광물찌꺼기 유실 실태파악 결과(한국광해관리공단, 2009)를 활용하였다.

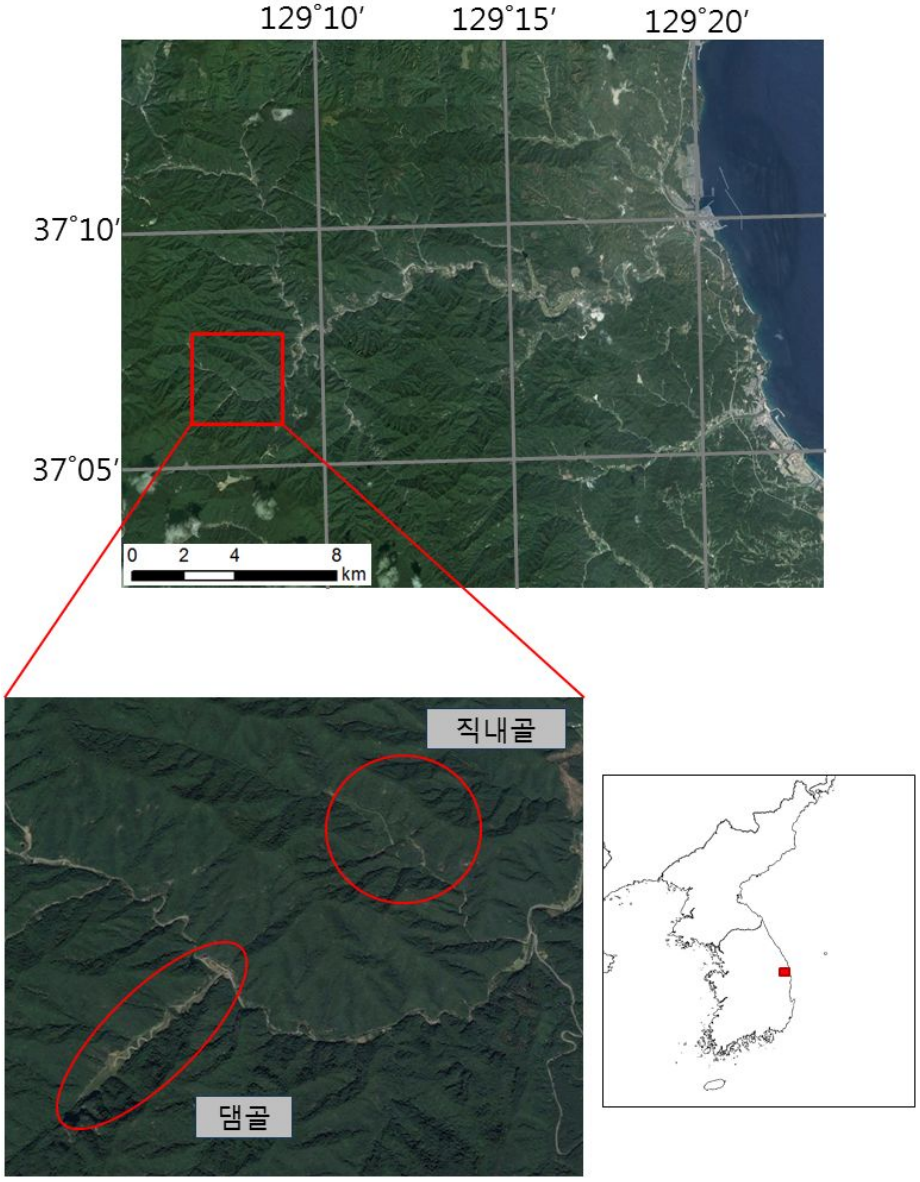


Fig. 5.1. Location of study area on the outline map and satellite image (Google Earth).

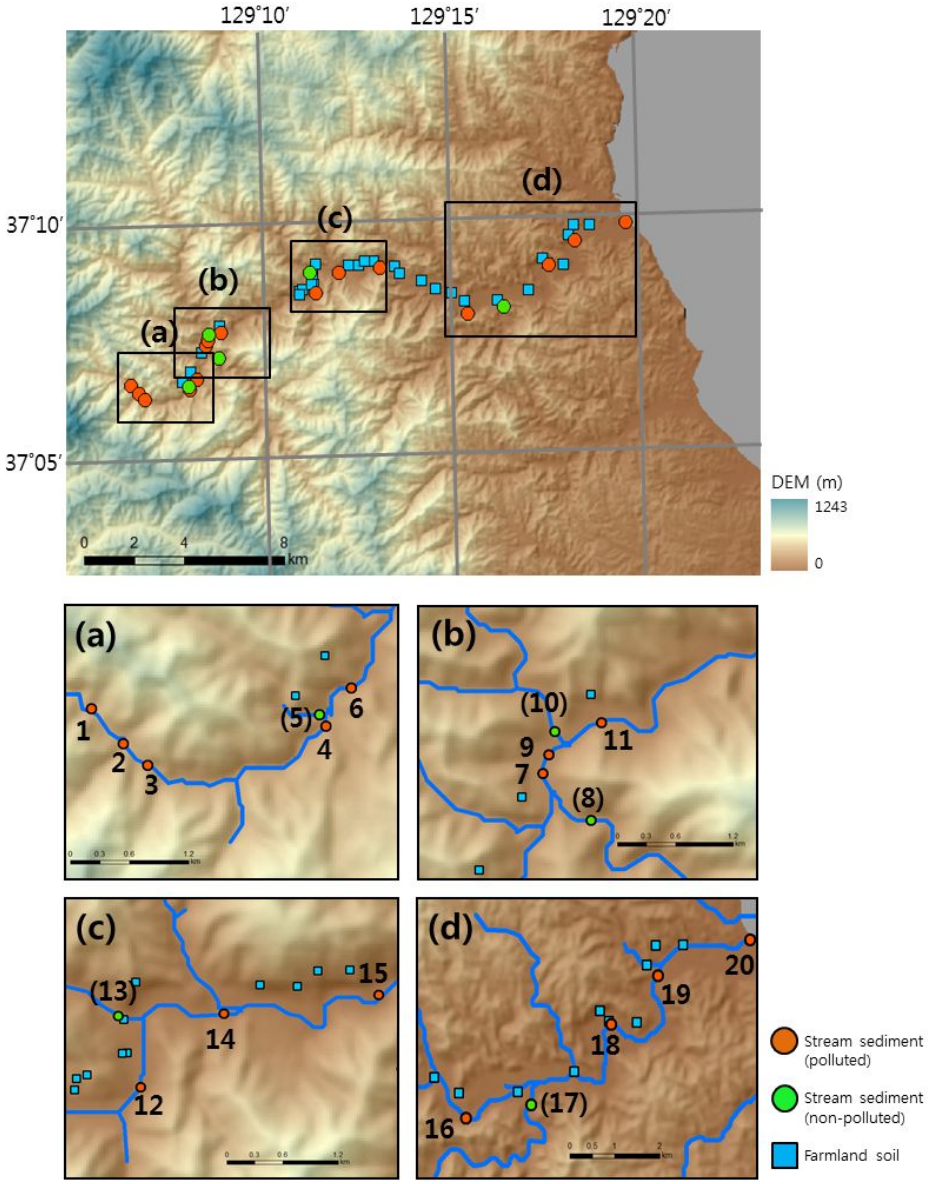


Fig. 5.2. DEM and locations of samples. Sample numbers in brackets indicate stream sediment samples collected from non-polluted stream.

Table. 5.1. Monthly precipitation (mm) of Donghae, Gangwondo from 2005 to 2014 (Korea Meteorological Administration).

	Jan.	Feb.	Mar.	Apr.	May.	Jun.	Jul.	Aug.	Sep.	Oct.	Nov.	Dec.	Sum
2005	49.0	48.3	134.6	30.3	31.5	76.5	239.5	172.8	554.1	90.6	5.5	0.1	1432.8
2006	72.2	13.8	12.7	110.9	98.7	45.7	910.7	261.1	179.0	151.1	77.2	34.0	1967.1
2007	29.8	34.2	157.3	33.1	86.1	79.5	242.0	153.8	605.5	31.4	58.8	3.0	1514.5
2008	73.3	0.1	100.1	36.1	48.5	90.6	252.1	154.3	65.6	61.3	60.6	35.9	978.5
2009	12.0	18.6	61.2	49.5	58.4	83.1	281.5	120.4	65.9	17.1	152.0	21.2	940.9
2010	36.1	110.3	128.3	32.8	108.1	18.1	85.4	148.1	291.1	70.3	6.9	13.6	1049.1
2011	22.1	180.3	42.3	93.6	104.5	333.3	292.8	124.5	231.0	67.0	174.3	100.0	1765.7
2012	37.7	18.8	108.8	96.2	38.9	28.4	287.0	233.6	209.8	14.5	18.9	30.7	1123.3
2013	68.4	22.0	74.1	53.8	85.8	55.6	202.0	63.4	181.7	132.5	4.2	21.5	965.0
2014	63.1	147.4	47.6	247.5	7.0	51.9	66.2	382.4	149.2	154.7	97.2	-	1414.2
Avg.	46.4	59.4	86.7	78.4	66.8	86.3	285.9	181.4	253.3	79.1	65.6	28.9	1315.1

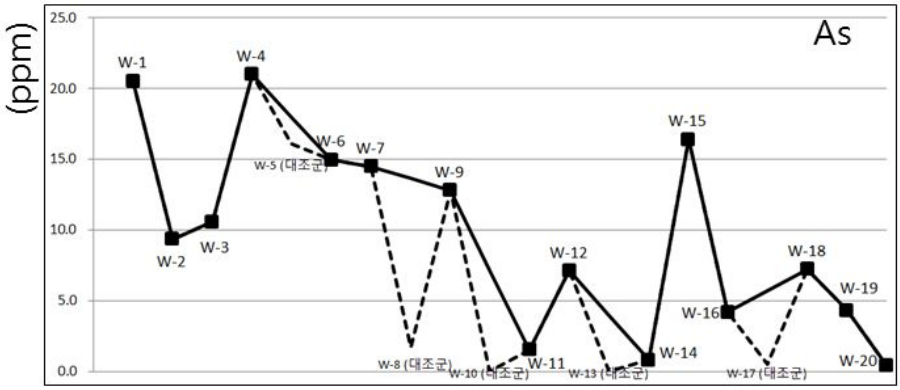
5.2. 대상지역 자료 검토 및 전처리

본 연구에서는 국토지리정보원에서 공개하는 90 미터 해상도의 수치표고모델을 이용하여 공간분석 및 크리깅 예측을 수행하였다. 취득된 중금속 오염 샘플의 공간적 분포는 Fig. 5.2에 나타냈다. 원으로 나타낸 샘플은 하상퇴적물에서 취득한 샘플을 의미하며 사각형으로 나타낸 샘플은 하천 주변 토양에서 취득한 샘플을 의미한다. 파란색 선은 GIS 기반의 수계분석을 통해 추출한 하천을 나타내며 빗물이 경사를 따라 집중되는 지역을 의미한다. 수계분석을 통해 추출한 하천과 샘플들의 위치를 비교해보면 샘플 포인트 자료의 상대적 위치가 대체로 잘 대응되므로 분석결과가 유의미할 것으로 기대할 수 있다. 개발한 알고리즘을 적용하기 위해서는 래스터 형태로 변환한 하천 격자와 샘플 격자의 위치가 일치하여야 하므로 샘플의 위치를 가장 가까운 하천 격자의 위치로 조정하였다.

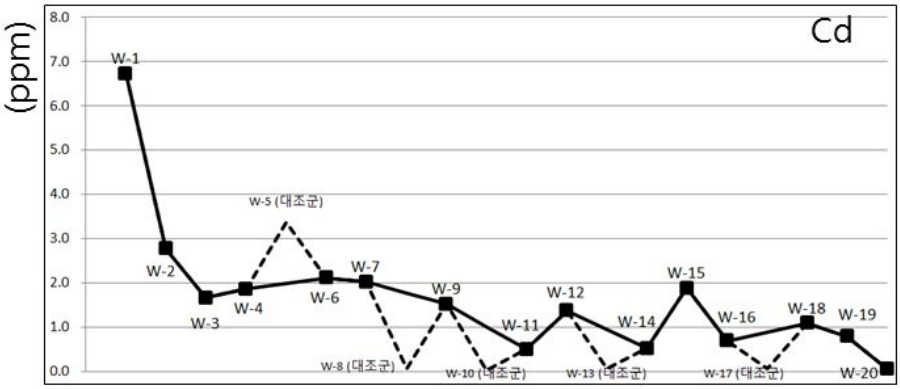
본 연구에서는 2009년 7월 집중호우에 의해 댐골 광미장의 광물찌꺼기가 유실된 후 취득된 샘플을 사용하였다. 20개의 하상퇴적물 샘플 중 5개는 오염되지 않은 대조군으로서 취득한 샘플이며 Fig. 5.2에서 포인트의 색상을 다르게 하고 번호를 괄호로 묶어 구별하였다. 대조군 샘플들은 다른 방향에서 유입되는 하천을 대상으로 샘플을 취득한 것을 확인할 수 있다. 댐골 광미장과 가까운 지점에서는 상대적으로 많은 시료가 취득되었고 광미장에서 멀어질수록 샘플 간의 거리가 멀어지는 것을 확인할 수 있다. 취득된 하상퇴적물 샘플은 비소(As), 카드뮴(Cd), 구리(Cu), 니켈(Ni), 납(Pb), 아연(Zn)에 대한 측정값을 가지고 있으며 샘플에

따른 수치를 Fig. 5.3에 나타냈다. 대조군으로 취득한 샘플은 대부분 주변값들에 비해서 낮은 측정값을 가지는 것을 알 수 있다. 따라서 대조군 외 샘플들이 위치한 하천 줄기가 상대적으로 오염되어 있는 것을 알 수 있다. As와 Zn의 경우 광미장에서 멀리있는 샘플일수록 오염농도가 낮아지는 경향을 보이기 때문에 댐골 광미장 유실의 영향이 있음을 알 수 있다. 그 외의 원소는 특별한 경향은 보이지 않는다. 하상퇴적물 샘플의 오염농도를 정리하면 Table 5.2와 같다. As는 토양오염 우려기준을 초과하는 샘플들이 일부 존재하며 Zn은 대부분이 미국의 하상퇴적물 제거기준을 초과할 정도로 높은 오염농도를 보인다. 따라서 Zn의 농도가 2009년 당시 광물찌꺼기 유실에 의한 영향을 가장 잘 반영하는 지표임을 알 수 있으며 본 연구에서는 Zn을 대상으로 분석을 수행하였다.

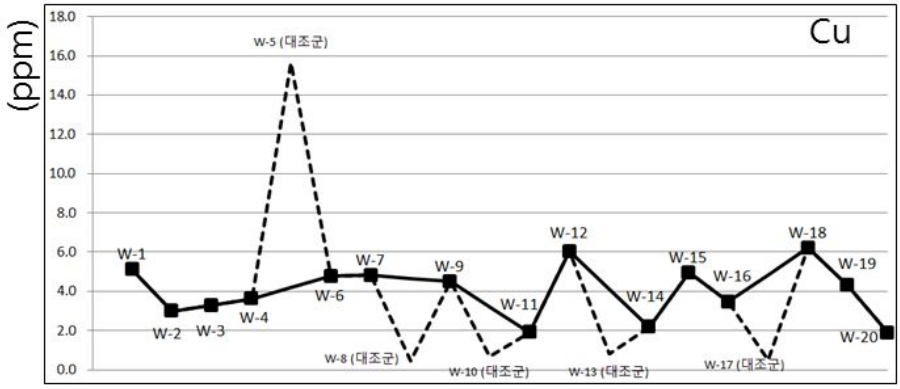
하천 주변 토양으로부터 취득한 오염농도는 Table 5.3과 같다. 하상퇴적물에 비해서 전체적으로 오염농도가 작은 것을 알 수 있으며 일부시료에서 Cd의 농도만 토양오염 대책기준을 초과하는것으로 나타났다. 따라서 하상퇴적물 샘플과 비슷한 위치에 있더라도 오염농도가 훨씬 작기 때문에 하상퇴적물 샘플과 함께 분석하는 것은 비합리적임을 확인할 수 있다. 샘플의 위치에 따라 오염농도가 특별한 경향을 보이지 않기 때문에 하상퇴적물에 비해서 폐광산에 의한 중금속 오염을 반영하지 못하는 것을 알 수 있다.


As
(ppm)
25.0
20.0
15.0
10.0
5.0
0.0
W-1
W-2
W-3
W-4
W-5 (대조군)
W-6
W-7
W-8 (대조군)
W-9
W-10 (대조군)
W-11
W-12
W-13 (대조군)
W-14
W-15
W-16
W-17 (대조군)
W-18
W-19
W-20

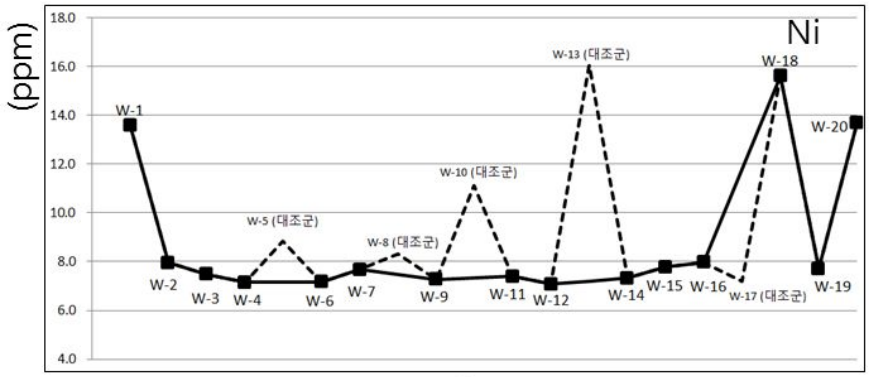
(a)


Cd
(ppm)
8.0
7.0
6.0
5.0
4.0
3.0
2.0
1.0
0.0
W-1
W-2
W-3
W-4
W-5 (대조군)
W-6
W-7
W-8 (대조군)
W-9
W-10 (대조군)
W-11
W-12
W-13 (대조군)
W-14
W-15
W-16
W-17 (대조군)
W-18
W-19
W-20

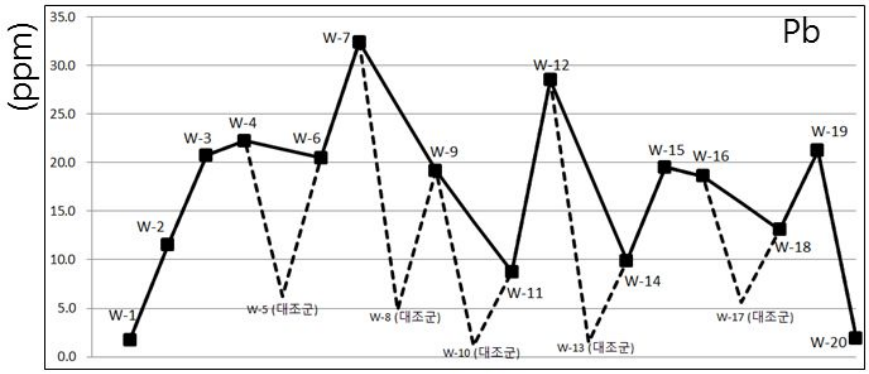
(b)


Cu
(ppm)
18.0
16.0
14.0
12.0
10.0
8.0
6.0
4.0
2.0
0.0
W-1
W-2
W-3
W-4
W-5 (대조군)
W-6
W-7
W-8 (대조군)
W-9
W-10 (대조군)
W-11
W-12
W-13 (대조군)
W-14
W-15
W-16
W-17 (대조군)
W-18
W-19
W-20

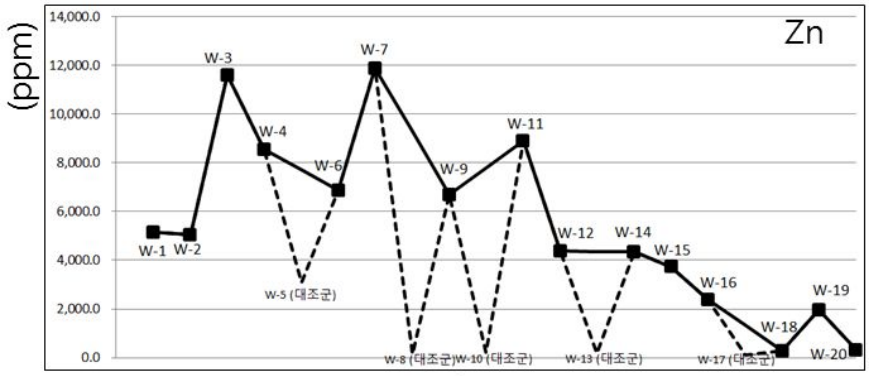
(c)



(d)



(e)



(f)

Fig. 5.3. Heavy metal concentration of samples. (a) As (b) Cd (c) Cu (d) Ni (e) Pb (f) Zn

Table 5.2. Heavy metal concentrations of stream sediment samples (light shaded number exceeds the soil contamination warning standards of korea and dark shaded number exceeds the US standard of sediment removal).

(unit: mg/kg)

		As	Cd	Cu	Ni	Pb	Zn
토양오염대책기준		6	0.5	50	40	100	300
토양오염우려기준		15	4	125	100	300	700
미국 퇴적토 제거기준		93	6.7	390		530	960
sample id	1	20.5	6.7	5.1	13.6	1.7	5157.8
	2	9.3	2.8	3.0	8.0	11.6	5046.3
	3	10.6	1.7	3.3	7.5	20.8	11607.3
	4	21.0	1.9	3.6	7.1	22.3	8558.4
	5*	16.1	3.4	15.7	8.8	6.2	3090.4
	6	15.0	2.1	4.8	7.2	20.5	6859.3
	7	14.5	2.0	4.8	7.7	32.4	11882.3
	8*	1.7	0.1	0.5	8.3	4.9	150.6
	9	12.8	1.5	4.5	7.3	19.2	6684.0
	10*	ND	0.0	0.7	11.1	1.1	228.7
	11	1.5	0.5	1.9	7.4	8.8	8910.7
	12	7.1	1.4	6.0	7.1	28.5	4385.2
	13*	ND	0.1	0.8	16.1	1.5	161.1
	14	0.8	0.5	2.2	7.3	9.9	4346.7
	15	16.4	1.9	5.0	7.8	19.5	3726.2
	16	4.2	0.7	3.5	8.0	18.6	2380.0
	17*	0.5	0.1	0.5	7.2	5.6	108.6
	18	7.2	1.1	6.2	15.6	13.1	279.6
	19	4.3	0.8	4.3	7.7	21.2	1955.6
	20	0.4	0.1	1.9	13.7	1.9	307.1

* control group

Table 5.3. Heavy metal concentrations of farmland soil samples (light shaded number exceeds the soil contamination countermeasure standards of korea and dark shaded number exceeds the soil contamination warning standards of korea).

(unit: mg/kg)

		As	Cd	Cu	Ni	Pb	Zn
토양오염대책기준		6	0.5	50	40	100	300
토양오염우려기준		15	4	125	100	300	700
sample id	1	0.9	0.3	2.0	20.4	27.5	176.6
	2	1.3	0.3	1.6	20.0	21.1	183.1
	4	0.1	0.1	1.0	21.8	17.6	140.1
	5	2.0	0.2	3.8	20.7	8.5	137.5
	6	0.5	0.3	3.5	13.3	15.7	127.8
	7	1.0	0.6	4.8	12.2	15.2	207.8
	7-1	0.3	0.1	2.2	15.5	6.9	77.2
	7-2	1.1	0.2	1.8	16.1	5.7	116.6
	7-3	0.2	0.1	1.5	15.2	4.0	65.2
	7-4	0.5	0.2	3.9	12.6	6.2	102.1
	8	1.5	0.3	2.0	15.5	1.6	152.8
	9	0.7	0.1	1.7	14.4	3.6	110.8
	10	4.7	0.1	1.7	16.3	2.4	72.6
	11	3.4	0.2	2.0	19.7	2.4	158.7
	12	ND	0.1	1.7	16.2	2.8	51.6
	13	0.4	0.1	2.0	8.9	7.5	54.1
	14	1.1	0.1	2.2	10.8	6.6	83.9
	15	1.6	0.2	2.9	10.6	6.2	70.3
	16	3.8	0.5	4.5	17.8	8.9	204.2
	17	1.0	0.2	3.0	13.8	8.7	103.9
	18	1.3	0.6	5.8	13.8	14.6	213.7
	19	0.7	0.1	1.6	9.3	8.1	44.0
	20	2.8	0.2	2.6	16.6	3.6	109.8
	21	2.4	0.1	2.4	12.5	5.7	99.0
	22	1.3	0.5	5.5	15.0	13.9	197.5
	23	2.0	0.2	2.5	21.7	2.8	121.4
	24	1.1	0.2	3.2	23.5	6.2	179.8
	25	0.9	0.1	2.7	16.1	4.0	81.7
	26	3.3	0.3	6.1	20.9	7.3	179.2
	27	1.7	0.1	3.2	21.2	4.9	111.0

5.3. 샘플 간의 거리계산 및 베리오그램 모델링

중금속 오염농도를 예측하기 위해서는 자료들의 거리와 오염농도를 이용하여 실험적 베리오그램을 계산하고 이를 잘 대표하는 이론적 베리오그램을 모델링하여야 한다. 본 연구에서 분석을 수행한 샘플은 20개이기 때문에 자기 자신과의 거리까지 포함하면 400개의 거리쌍이 존재하며 개발한 소프트웨어를 이용하여 각 샘플들의 거리를 계산하였다. 일반적인 크리깅을 위해서 각 샘플들 간의 유클리디안 거리를 계산하였고 본 연구에서 제안한 예측기법의 적용을 위해서 하천의 경로를 따라 샘플들의 거리를 계산하였다. 샘플들 사이의 거리는 Table 5.4와 같으며 모든 쌍의 거리를 기재하지 않고 인접한 샘플들 간의 거리만 기재하였다. 하천의 경로를 고려한 경우 8번 샘플을 제외한 대조군 샘플들은 앞 번호 샘플들과 연결되지 않는 것을 알 수 있다. Fig. 5.2를 보면 8번 샘플이 위치한 하천 줄기가 중심 하천 줄기에 합류한 이후에 7번 샘플이 위치하기 때문에 두 샘플만 예외적으로 연결이 된 것이다. 하천의 경로를 고려한 경우에는 샘플들 간의 곡선 거리를 계산하기 때문에 유클리디안 거리에 비해서 큰 값을 가지는 것을 확인할 수 있다.

Table 5.4. Euclidean distance and flow distance between samples.

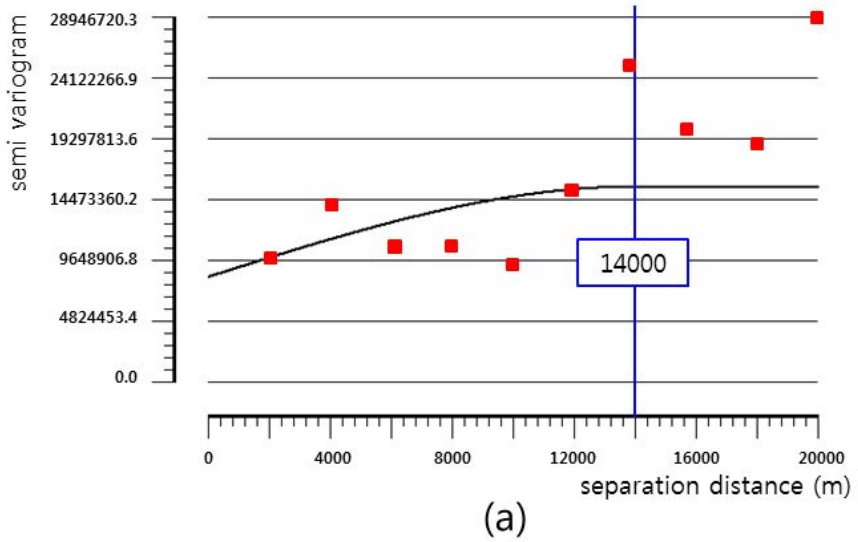
(unit: m)

sample id (A)	sample id (B)	Euclidean distance between (A) and (B)	flow distance between (A) and (B)
1	2	500.0	582.8
2	3	360.6	382.8
3	4	1843.9	2131.4
4	5*	141.4	not connected
5*	6	424.3	482.8
6	7	1360.1	1690.0
7	8*	707.1	765.7
8*	9	806.2	1007.1
9	10*	223.6	not connected
10*	11	412.3	582.8
11	12	4123.1	5062.7
12	13*	824.6	not connected
13*	14	1100.0	1265.7
14	15	1711.7	2090.0
15	16	3935.7	5045.6
16	17*	1529.7	not connected
17*	18	2334.5	3407.1
18	19	1486.6	2738.5
19	20	2119.0	2572.8

* control group

모든 샘플들 사이의 거리와 오염농도 값을 이용하여 실험적 베리오그램을 계산하였다. Fig. 5.4는 Zn 샘플 20개의 오염농도에 대한 베리오그램을 보여준다. Fig. 5.4(a)는 유클리디안 거리를 계산했을 때의 결과이며 Fig. 5.4(b)는 하천의 경로를 고려한 결과이다. 본 연구에서는 분리거리를 2,000 미터 간격으로 늘려가며 베리오그램을 계산하였으며 이론적 베리오그램은 구형모델과 너깃모델을 결합하여 가중최소자승법을 이용하여 모델링하였다. 두 경우 모두 분리거리가 작을 때에도 베리오그램 값이 상당히 높은 값을 가진다. 그 이유는 대조군을 포함하여 모든 샘플에 대해서 분석을 했기 때문에 대조군 샘플과 오염 샘플의 오염농도 차이에 의한 너깃 효과가 크게 나타나기 때문이다. 상관거리는 두 경우 모두 진동이 커지기 시작하는 14,000 미터로 설정하였다.

Fig. 5.5는 대조군 샘플 5개를 제외한 샘플 15개에 대한 베리오그램을 보여준다. 대조군 샘플은 모두 다른 하천 줄기에서 한 개씩만 취득하였기 때문에 분석 대상에서 제외하는 것이 합리적일 수 있다. 베리오그램 결과를 비교해보면 Fig. 5.4에 비해서 전형적인 베리오그램의 형태에 가까운 것을 확인할 수 있다. 이론적 베리오그램은 구형모델과 너깃모델을 결합하여 모델링하였으며 상관거리는 모두 12,000 미터로 설정하였다.



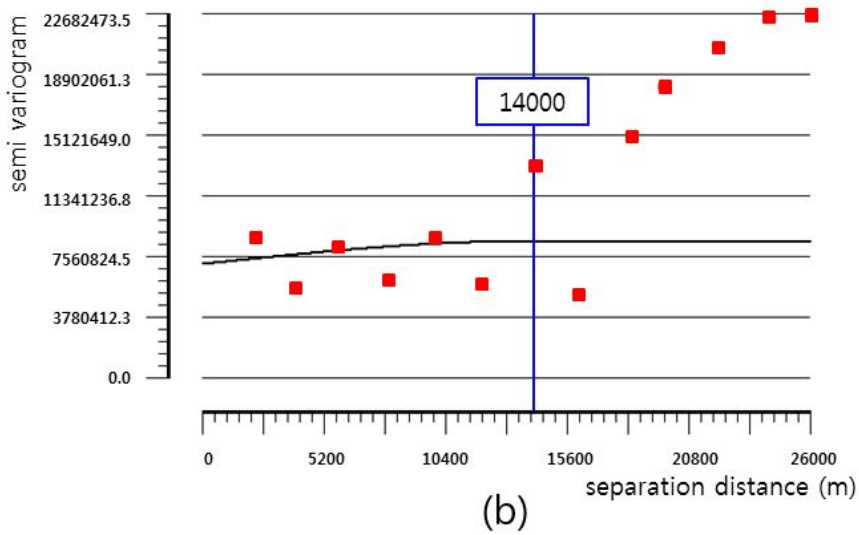


Fig. 5.4. Empirical variogram and theoretical variogram for 20 samples considering (a) Euclidean distances and (b) distances on the stream.

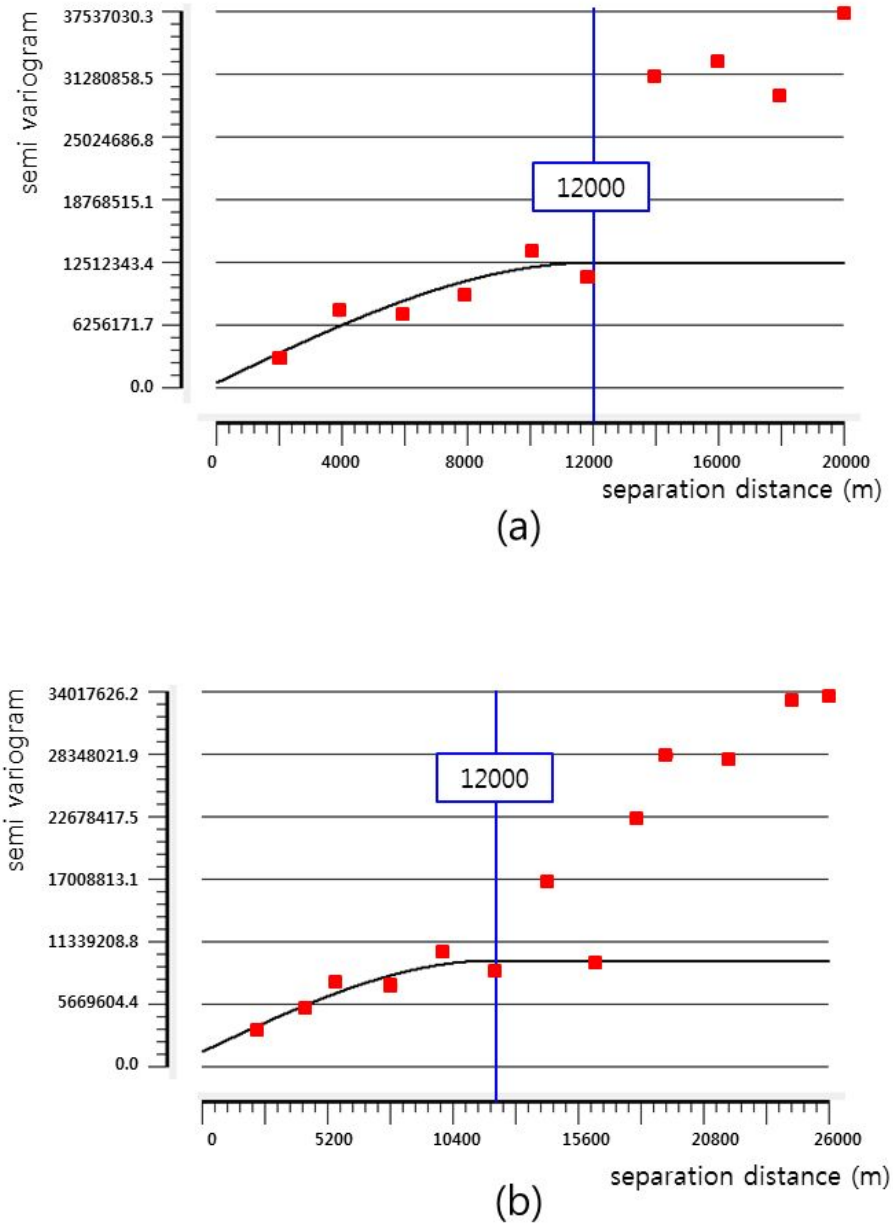


Fig. 5.5. Empirical variogram and theoretical variogram for 15 samples considering (a) Euclidean distances and (b) distances on the stream.

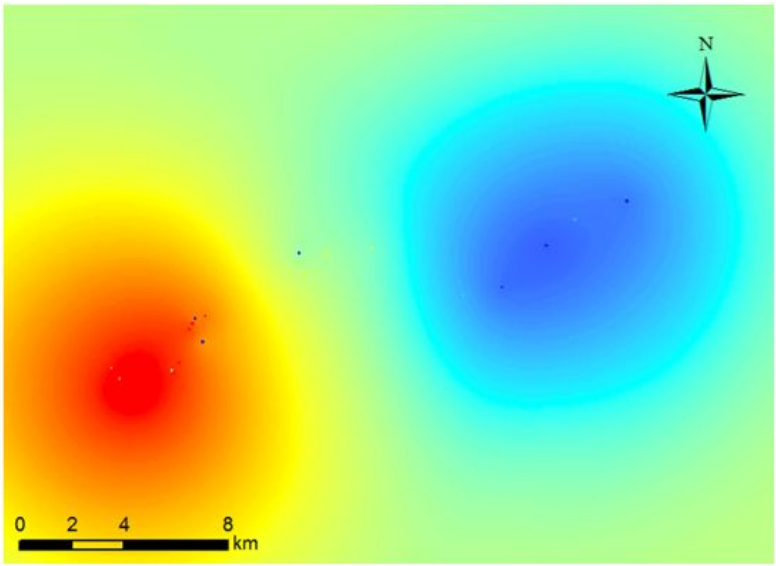
5.4. 오염농도 예측 및 검증

이론적 베리오그램을 이용하여 대상지역에서의 Zn 오염농도 값을 예측하였다. 먼저 일반적인 크리깅 기법을 적용하여 연구지역의 모든 지점에 대해서 값을 예측하였고, 하천의 경로를 고려한 예측기법을 적용한 경우에는 전체영역이 아닌 하천의 경로에 대한 예측결과를 얻는다. Fig. 5.6은 20개의 모든 샘플을 이용하여 오염농도를 예측한 결과로서 베리오그램 모델링 과정(Fig. 5.4)에서 구형모델에 너깃모델을 결합하였기 때문에 하나의 샘플에 큰 영향을 받지 않고 대체로 완만하게 값이 변화하는 것을 확인할 수 있다. 일반적인 크리깅 방법(Fig. 5.6(a))과 본 연구에서 제안한 방법(Fig. 5.6(b))의 결과를 비교하면 두 경우 모두 광미장이 위치한 서쪽의 오염농도 값이 크고 동쪽으로 갈수록 값이 작아지는 경향을 나타낸다. 그러나 샘플이 없는 지점에서 그 차이가 두드러진다. 일반적인 크리깅 방법의 경우 모든 지점의 예측 과정에서 20개 샘플이 모두 영향을 주기 때문에 하천의 네트워크를 전혀 고려할 수 없으며 특히 오염되지 않은 하천을 구별할 수 없다는 단점이 있다. 본 연구에서 제안한 기법은 하천의 네트워크를 고려하여 예측 지점에 영향을 주는 샘플을 선택하기 때문에 오염되지 않은 하천 줄기를 예측할 수 있다는 장점을 가진다. 대상지역 서쪽에서 남동쪽으로 뻗어있는 하천 줄기는 상대적으로 높은 오염농도를 가지는 것으로 예측되었는데 대조군으로서 취득된 샘플은 1개인데 비하여 그 하류에서 취득된 샘플들이 높은 오염농도를 가지기 때문이다. 따라서 이러한 하천 줄기는 예측값이 과장될 수 있기

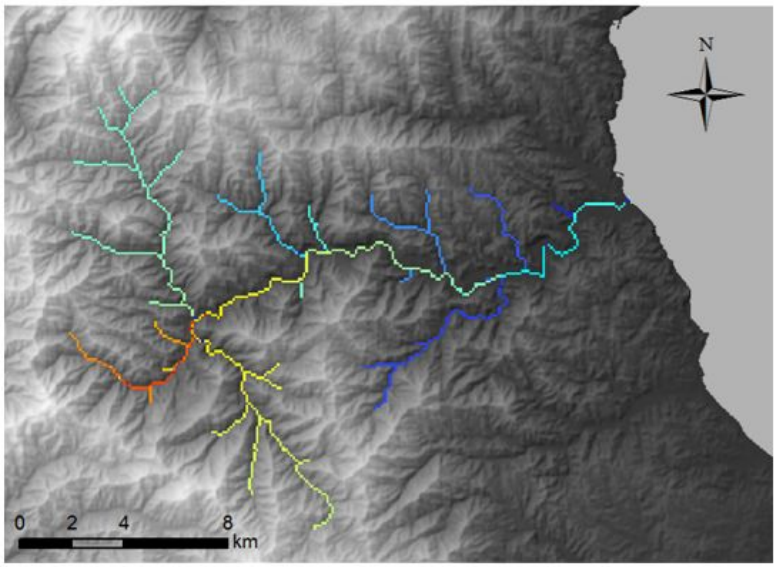
때문에 추가적인 샘플의 취득이 필요한 구간이라고 판단된다.

대조군 샘플과 오염 샘플은 가까운 거리에서도 값의 차이가 크기 때문에 베리오그램 모델링 과정에서 방해요인이 될 수 있다. 따라서 대조군을 제외한 베리오그램(Fig. 5.5)에 대해서도 오염농도 예측을 수행하였다(Fig. 5.7). 일반적인 크리깅을 적용한 경우에 더 짧은 상관거리를 적용한 차이 외에는 시각적으로 큰 차이가 없는 것으로 나타났으며, 하천을 고려한 경우에도 오염 구간이 짧아진 것 외에는 시각적으로 큰 차이는 없는 것으로 나타났다. 이 때 대조군을 제외하고 베리오그램 모델링을 수행했기 때문에 주 오염경로 외의 하천 줄기 예측은 하지 않았다.

예측능력 측면에서는 어떠한 차이가 있는지 확인하기 위하여 Table 5.5와 같이 교차검증을 수행하였다. 모든 경우에서 일반적인 크리깅을 적용한 경우보다 하천을 고려한 경우가 MAE와 RMSE 값이 더 작은 값을 보이므로 본 연구에서 제안한 기법을 통해 예측능력이 향상된 것을 확인할 수 있다. 20개의 모든 샘플을 분석한 경우 오차가 큰 대조군으로 인해 전체적인 오차가 크게 나타날 수 있으므로 대조군을 제외한 예측 결과도 비교하였다. 대조군으로서 취득된 샘플은 아니지만 1번 샘플도 댐골 광미장의 오염경로에서 벗어났기 때문에 비교대상에서 제외하고 14개의 샘플에 대한 분석결과를 비교하였다. 처음부터 대조군을 제외하고 베리오그램을 모델링한 결과가 더 우수한 예측 능력을 보이는 것을 확인할 수 있다. 따라서 오염경로에 대해서 예측의 정확도를 높이고 싶은 경우에는 대조군을 제외하고 베리오그램을 모델링하는 것이 효율적임을 알 수 있다. 그러나 대조군을 제외하였기 때문에 주변의 하천 줄기에 대한 합리적인 예측을 제공하지 못한다는 단점이 있다.

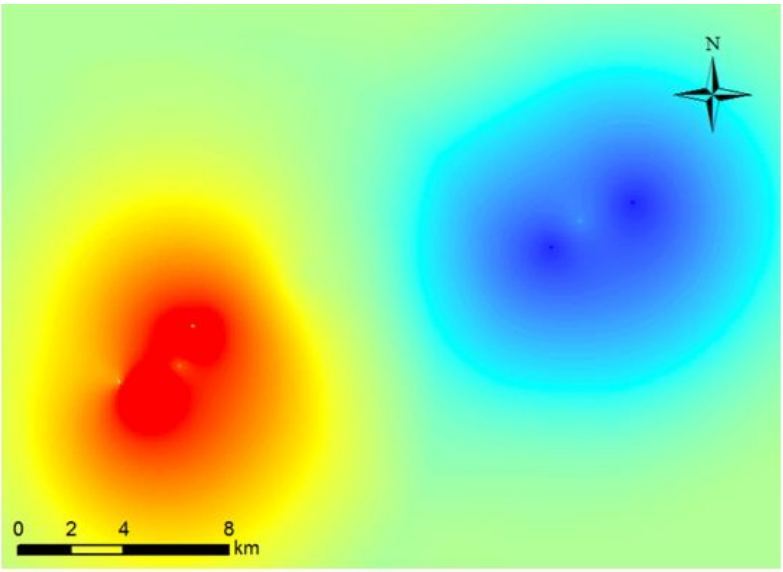


(a)

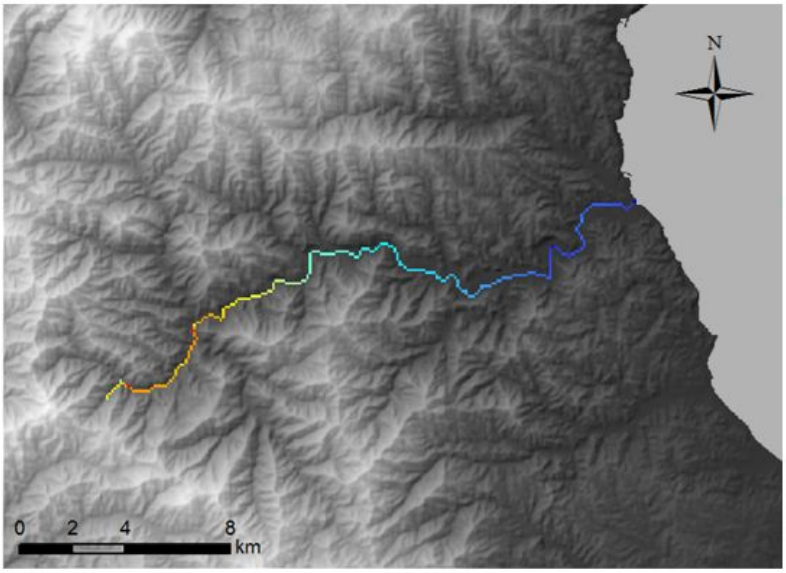


(b)

Fig. 5.6. Prediction of Zn concentrations for the study area using (a) Euclidean distance and (b) flow kriging with variogram model for 20 samples.



(a)



(b)

Fig. 5.7. Prediction of Zn concentrations for the study area using (a) Euclidean distance and (b) flow kriging with variogram model for 14 samples.

Table. 5.5. Cross validation results of general kriging and flow kriging for study area (shaded number is the most similar value to original value).

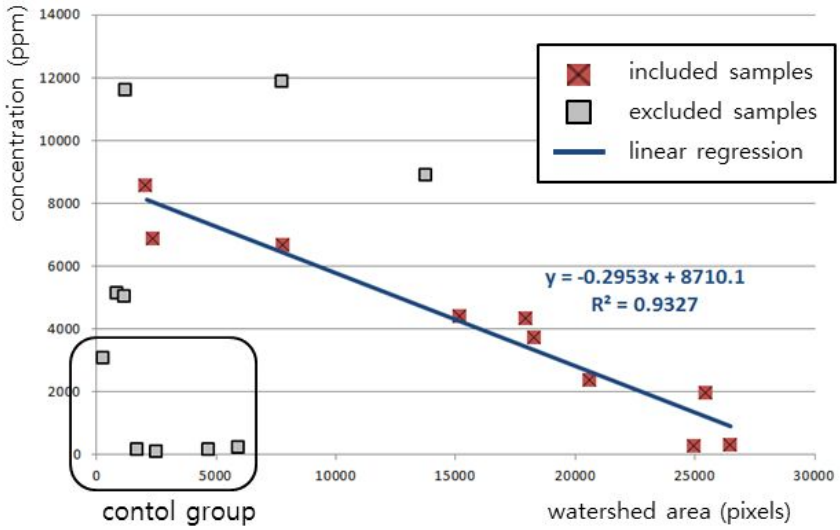
(unit: mg/kg)

sample id	original value	general kriging		flow kriging	
		for 20 samples	for 14 samples	for 20 samples	for 14 samples
1	5158	6367.3	-	6402.2	-
2	5046	**6614.5**	10886.5	**6584.6**	10057.9
3	11607	5101.8	**5630.2**	5678.5	**5721.4**
4	8558	5454.2	**7782.3**	6461.4	**7905.4**
5*	3090	6533.0	-	5852.8	-
6	6859	**5703.3**	9078.4	**6274.3**	8916.4
7	11882	4725.1	**7149.4**	5082.6	**7341.6**
8*	151	6280.1	-	5905.0	-
9	6684	**5438.3**	10379.9	**5588.5**	9826.4
10*	229	6409.1	-	4258.6	-
11	8911	4644.9	**7748.6**	4626.4	**7912.5**
12	4385	3351.3	**5305.0**	**4298.6**	5235.1
13*	161	4359.3	-	2620.5	-
14	4347	3017.4	**4189.5**	3454.4	**4082.7**
15	3726	2723.0	**3530.4**	3397.9	**3720.3**
16	2380	1919.8	**2224.6**	3268.2	**2920.3**
17*	109	2225.5	-	822.0	-
18	280	**1918.8**	2125.0	2783.1	**2494.9**
19	1956	**1466.0**	518.0	**2585.3**	835.0
20	307	**2554.5**	2713.2	3146.4	**3112.3**
ME (for 20 samples)		48.9	-	163.3	-
ME (for 14 samples)		-1592.5	166.6	-978.4	225.2
MAE (for 20 samples)		2824.0	-	2373.0	-
MAE (for 14 samples		2371.8	2251.5	2178.2	2149.3
RMSE (for 20 samples)		3535.0	-	3078.6	-
RMSE (for 14 samples		3148.1	2987.7	2981.7	2816.0

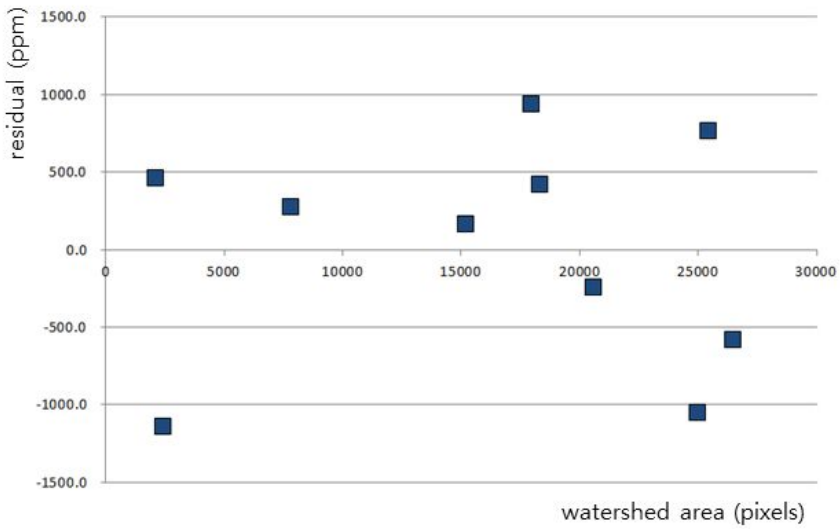
* control group

집수구역을 고려한 SKLM의 현장 적용 가능성을 확인하기 위해 Fig. 5.8(a)와 같이 샘플들의 집수구역 면적과 Zn 함량을 축으로 하여 그래프로 나타냈다. 대조군 샘플은 분석대상에서 제외하였으며 전체적인 경향에서 값이 크게 벗어나는 샘플은 회귀식에 영향을 주어 전체적인 경향을 반영하기 어렵게 하므로 회귀식 계산 과정에서 제외하였다. 집수구역이 작은 구간에서는 값의 진동이 상대적으로 크지만 전체적으로 집수구역의 면적과 오염농도가 음의 상관관계를 가지는 것을 확인할 수 있다. 이러한 경향을 반영하기 위해 선형 회귀식을 적용하였으며 회귀식에 대한 샘플들의 잔차 분포는 Fig. 5.8(b)와 같다.

SKLM을 적용하기 위해서는 회귀식을 이용하여 지역평균 값을 구하고 잔차를 크리깅한 결과를 더해야 하기 때문에 먼저 잔차를 이용하여 베리오그램을 모델링하였다(Fig. 5.9). 원래의 Zn 오염농도 분포에 비해서 잔차는 지역적 상관성이 약하기 때문에 실험적 베리오그램이 전형적인 형태를 보이지는 않는다. 본 연구에서는 구형모델을 적용한 결과와 구형모델과 너깃모델을 결합하여 모델링한 결과를 비교하였다. 가중최소자승법을 이용하여 이론적 베리오그램을 모델링 하였고 상관거리는 12,000 미터로 설정하였다.



(a)



(b)

Fig. 5.8. (a) Correlation between watershed areas and heavy metal concentrations of study area and (b) the residual distribution according to watershed area.

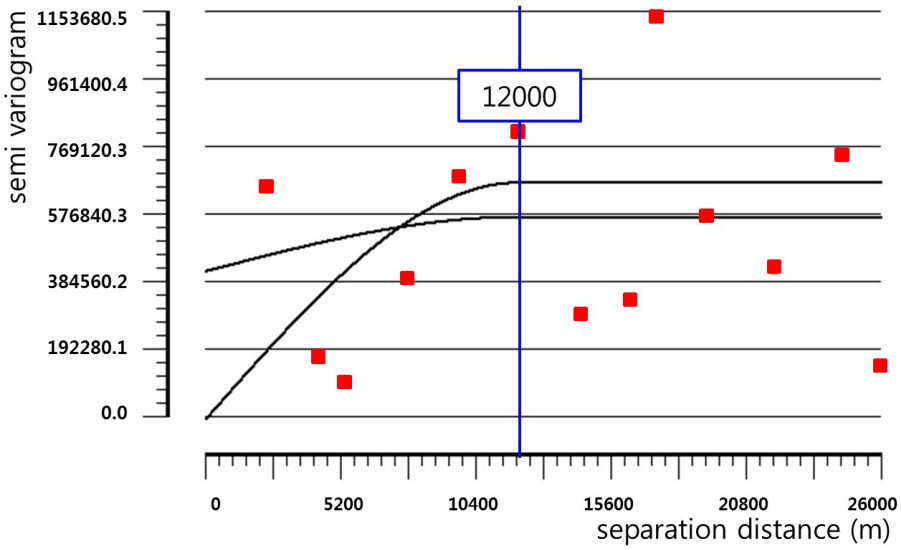
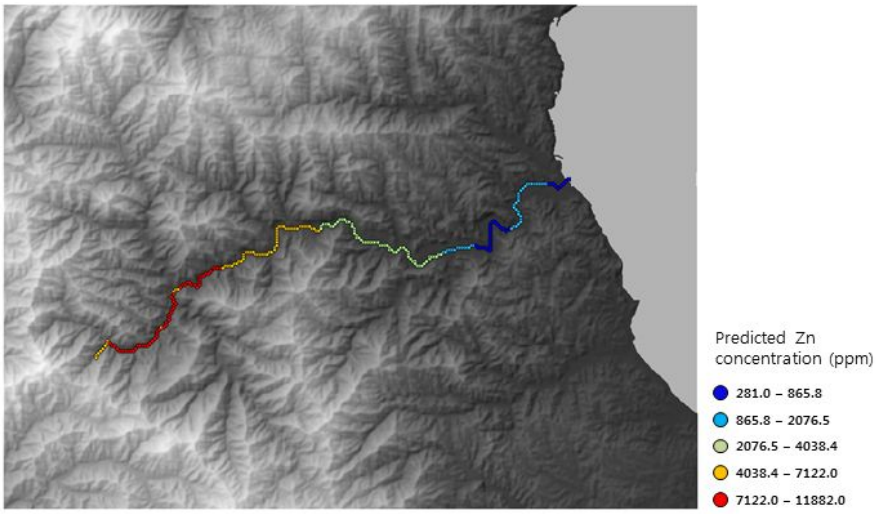


Fig. 5.9. Empirical variogram and theoretical variogram of residuals.

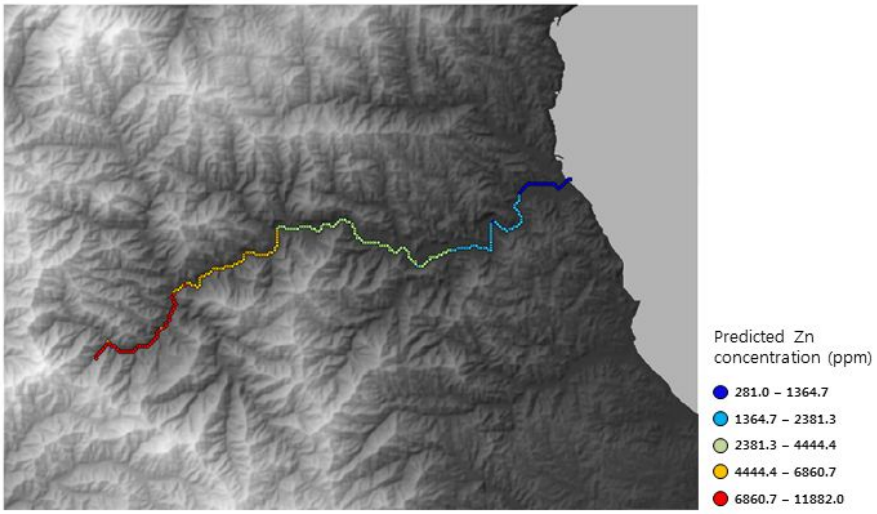
다음 과정으로 회귀 분석을 통해 얻은 관계식을 이용하여 연구 지역 전체에 걸쳐 집수구역의 면적 값을 오염농도로 변환하였다. 이 값은 각 지점의 지역평균으로서 이용된다. 여기에 잔차의 크리깅 결과를 합산하면 Fig. 5.10과 같은 오염농도 예측값을 구할 수 있다. Fig. 5.10(a)는 구형모델을 적용한 결과이고, Fig. 5.10(b)는 구형모델과 선형모델을 결합한 결과이다. 오염농도 예측은 주 오염경로에 대해서 수행하였고 래스터 자료를 포인트 자료로 변환하여 가시화하였다. 전체적인 경향은 유사하나 너깃모델을 결합한 결과에 비해서 구형모델만 적용한 결과는 가까운 자료의 영향을 많이 받는 것을 확인할 수 있다.

적용한 모델에 따른 차이를 더 자세히 확인하기 위해서 Fig. 5.11과 같이 거리에 따른 오염농도 값의 변화를 그래프로 나타냈다. 가로축은 댐골 광미장으로부터의 거리를 격자 개수로 나타낸 것이고 세로축은 해당 격자의 오염농도를 의미한다. 사각형 모양의 포인트는 샘플이 있는 지점으로서 크리깅의 특성상 모든 분석 결과가 정확히 원래의 값을 예측하는 것을 확인할 수 있다. 두 개의 SKLM 적용 결과와 함께 Zn 오염농도를 바로 예측한 결과(Fig. 5.7(b))도 함께 그래프로 나타냈다. 구형모델만 적용한 SKLM 기법이 샘플 주변의 값에 영향을 많이 받는 반면에 너깃모델을 함께 적용한 SKLM 기법의 결과는 가까운 샘플의 값보다 전체적인 경향에 영향을 많이 받는 것을 알 수 있다. 구형모델과 너깃모델을 결합하여 Zn 오염농도를 바로 예측한 결과는 대체로 중간적인 결과를 보인다. 교차검증을 수행한 결과 MAE는 SKLM 기법(구형모델)이 2,677, SKLM 기법(구형모델과 너깃모델 결합)이 1,843, 오염농도를 바로 예측한 방법(구형모델과 너깃모델 결합)이

2,149로 나타났다. 따라서 교차검증을 통한 예측능력은 구형모델과 너깃모델을 결합한 SKLM 방법이 가장 우수하게 나타났다. 그러나 너깃모델을 적용할 경우 샘플 주변에서 예측값과 샘플값의 차이가 커질 수 있다는 점에 주의하여야 한다.



(a)



(b)

Fig. 5.10. Prediction of Zn concentrations for the study area using flow kriging with (a) spherical variogram model and (b) spherical variogram model combined with nugget model.

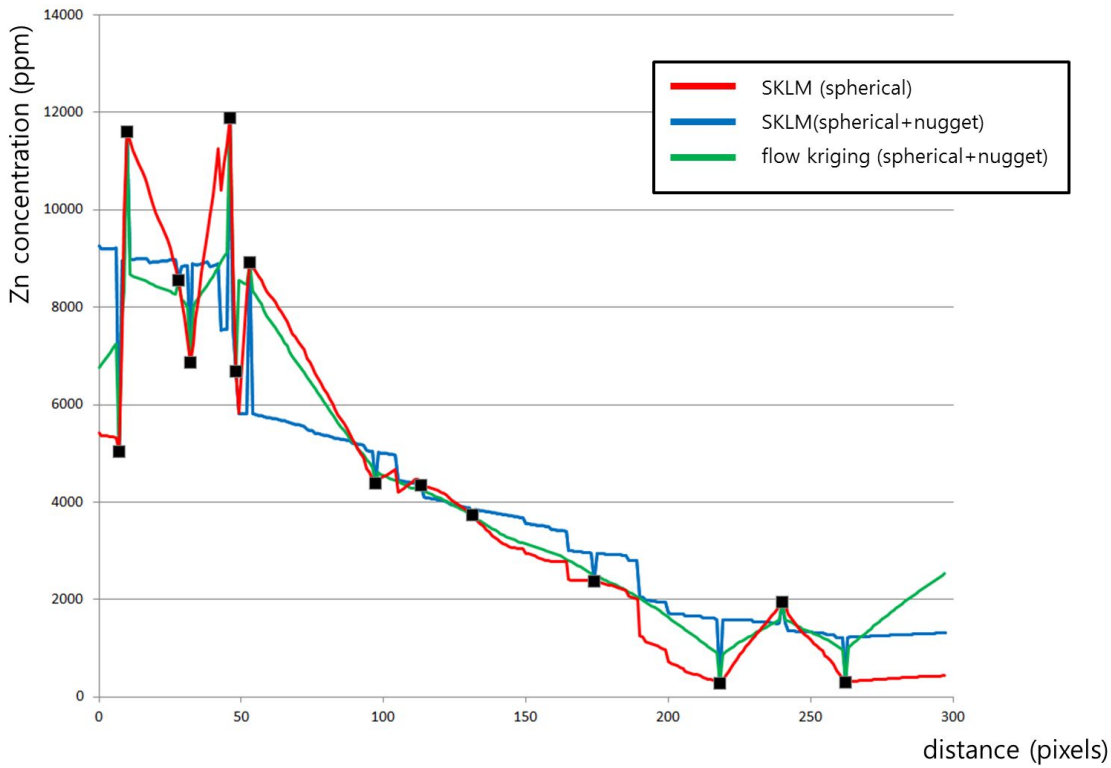


Fig. 5.11. Variation of predicted Zn concentrations for different methods according to distance from tailings dam.

6. 결론

본 연구에서 제안한 하천을 고려한 크리깅 기법을 견본자료와 현장에 적용한 결과 일반적인 크리깅에 비해서 하상퇴적물에 대한 예측능력이 향상되는 것을 확인할 수 있었다. 실제 하상퇴적물의 이동을 반영하여 하천에서의 거리를 사용하기 때문에 교차검증 수행시 예측의 정확도가 향상되었고, 특히 하천의 네트워크를 고려하여 하천 줄기마다 영향을 주는 샘플을 구별하여 분석이 가능하다는 장점을 가진다.

예측능력 검증은 샘플을 하나씩 제거하고 제거한 샘플의 값을 예측하는 교차검증을 통해서 수행하였다. 베리오그램 모델에 따라서 교차검증 결과의 차이가 큰 것을 확인할 수 있으며 너깃모델을 결합한 결과가 대체로 전체적인 오차가 작은 것을 알 수 있다. 너깃모델을 적용할 경우 가까운 샘플의 영향은 작아지고 전체적인 평균값의 영향이 커지므로 극단적인 오차가 적기 때문으로 판단된다. 마찬가지 이유로 구형모델을 적용한 경우에 선형모델이나 가우스모델을 적용한 결과보다 전체적인 오차가 작은 경향을 보였다. 그러나 교차검증 결과 오차가 작다고 하여 반드시 우수한 예측을 의미하는 것은 아니므로 주의할 필요가 있다. 교차검증을 위해 제거하는 샘플의 영향을 무시할 수 없으며 특히 전체 샘플의 수가 적을수록 제거한 샘플의 영향이 크게 작용할 수 있다. 따라서 예측의 신뢰도를 향상시키기 위해서는 궁극적으로는 다량의 자료수 확보가 최선의 방법이다.

그러나 일반적으로 양질의 자료는 부족할 수 밖에 없고 따라서

제한된 자료로부터 최적의 예측을 하기 위해서는 예측 기법의 우수성 못지 않게 분석자의 판단이 중요하다. 크리깅의 경우 어떤 이론적 베리오그램을 이용하여 예측을 수행하는지에 따라 최종적인 예측값이 달라지기 때문에 분석자의 판단 하에 주변 샘플의 값을 중요시 할지 전체적인 샘플의 경향을 중요시 할지 등을 고려해야 한다.

본 연구에서는 제안된 알고리즘에 SKLM 기법을 결합하여 집수구역이 커짐에 따라 오염농도가 작아지는 경향을 고려하였다. 하천 네트워크 전체에 걸쳐서 분석을 수행할 경우 오염되지 않은 하천이 과장되게 예측되는 경향이 있으나 오염된 하천 줄기만을 대상으로 적용할 경우에는 오염농도 값만 고려한 경우보다 예측능력이 향상되었다. 샘플의 수가 충분히 확보될 경우에는 각각의 하천 줄기마다 다른 회귀식을 적용하여 분석한다면 유용하게 활용될 수 있을 것이다. SKLM을 합리적으로 적용하기 위해서는 회귀식 예측 과정에서 상관관계가 분명히 존재해야 하며 잔차의 지역적 상관성이 떨어질 경우에는 오차가 커지는 원인이 될 수 있으므로 주의하여야 한다.

제안된 알고리즘은 다음과 같은 한계점을 지니고 있다. 본 연구에서는 미지의 값을 예측할 때 상류 지역의 샘플과 하류 지역의 샘플을 모두 고려하여 가중치를 부여한다. 상류의 오염은 하류에 영향을 주는 것이 분명하기 때문에 고려하는 것이 당연하다. 그러나 하류가 오염되었을 경우에는 상류가 오염되었을지 오염되지 않았을지 알 수가 없다. 예를 들어 오염된 샘플의 상류에 여러 개의 하천 줄기가 연결되어 있다면 어떠한 하천 줄기가 오염되었는지는 알 수가 없다. 각각의 하천 줄기에서 충분한 샘플이 취득되었다면

그 샘플들의 영향으로 합리적인 분석이 가능하지만 그렇지 않을 경우에는 하류의 오염된 샘플로 인해 오염되지 않은 상류 지점의 예측값이 과장될 수 있다. 서로 다른 하천 줄기에 대해서는 하류의 샘플을 고려하지 않는 후속연구를 통해서 이러한 문제점을 개선할 수 있을 것으로 보인다.

제안된 예측기법은 하천 네트워크에 적용하여 하천 줄기별 오염 복구우선순위를 산정할 때 유용하게 활용될 수 있을 것으로 기대된다. 폐광산 지역 뿐만 아니라 목축지나 도심지, 공장지대로 인한 오염 등 다양한 상황에서 범용적으로 활용될 수 있으며 하천의 네트워크가 복잡하고 굴곡이 심할수록 상대적 활용도가 높을 것이다. 개발로 인한 오염이 상대적으로 적은 저개발 국가에 적용할 경우에는 광화지역을 찾을 때에도 유용하게 활용하여 자원개발에 기여할 수 있을 것으로 기대한다.

참고문헌

김경웅, 이현구, 1996, 대전지역 하천수 및 하상퇴적물에서의 중금속 오염, 한국자원공학회지, Vol. 33, pp. 266-273.

김상현, 전효택, 1993, 삼보 연-아연-중정석 광산 주변 토양에서의 중금속 오염연구, 한국자원공학회지, Vol. 30, pp. 228-237.

김성민, 최요순, 박형동, 김태혁, 2012, 지표수의 확산 흐름을 고려한 광산 침출수 유출경로 예측, 한국자원공학회지, Vol. 49, pp. 736-745.

박노욱, 장동호, 2008, 수치표고모델과 다변량 크리깅을 이용한 기온 및 강수 분포도 작성, 대한지리학회지, Vol. 43, pp. 1002-1015.

박천영, 박영석, 정연중, 1995, 광양광산 주변 토양의 중금속 오염에 관한 연구, 한국자원공학회지, Vol. 32, pp. 163-174.

박천영, 정연중, 김성구, 2002, 함백폐탄광 지역의 광산배수와 하상퇴적물에 대한 지구화학적 특성, 한국자원공학회지, Vol. 39, pp. 35-50.

안주성, 전효택, 손아정, 김경웅, 1999, 구봉 금은광산 주변지역의 비소 및 중금속에 의한 환경오염과 벼작물의 흡수특성, 한국자원공학회지, Vol. 36, pp. 159-169.

이평구, 2004, 체계적인 폐광산지역 중금속 오염 평가와 경제적인 복원을 위한 문제제기, 제130회 정책&지식 포럼, http://www.know.or.kr.

정상용, 이강근, 1995, 난지도 매립지 일대의 지하수위 분포 추정을 위한 복합 크리깅의 응용, 지하수환경, Vol. 2, pp. 58-63.

최종근, 2013, 지구통계학, 시그마프레스, 419p.

한국광해관리공단, 2007, 삼척 제2연화광산 광물찌꺼기 유실방지사업 기본 및 실시설계 보고서, 508p.

한국광해관리공단, 2009, 제2연화광산 광미유실 실태파악 합동조사 결과 보고서

Ayad, Y., 2006, Prioritizing acid mine drainage stream remediation, GIS best practices, ESRI, USA, http://www.esri.com/library/bestpractices/mining.pdf.

Bijanzadeh, E., Mokarram, M. and Naderi, R., 2014, Applying Spatial Geostatistical Analysis Models for Evaluating Variability of Soil Properties in Eastern Shiraz, Iran, Iran Agricultural Research, Vol. 33, pp. 35-46.

Bonham-Carter, G.F. and Goodfellow, W.D., 1984, Autocorrelation structure of stream sediment geochemical data: Interpretation of Zn and Pb anomalies, Nahanni river area, Yukon - Northwest Territories, Canada, Geostatistics for Natural Resources Characterization, Reidel, Dordrecht, pp. 817-823.

Carranza, E.J.M and Hale, M., 1997, A Catchment Basin Approach to the Analysis of Reconnaissance Geochemical-geological Data from Albay Province, Philippines, Journal of Geochemical Exploration, Vol. 60, pp. 157-171.

Cressie, N., 1985, Fitting variogram models by weighted least squarees, Mathematical Geology, Vol. 17, pp. 563-586.

Curriero, F., 1996, The use of non-euclidean distance in geostatistics, Ph. D. Thesis, Kansas State University.

Dent, C.L. and Grimm, N.B., 1999, Spatial heterogeneity of stream water nutrient concentrations over successional time, Ecology, Vol. 80, pp. 2283-2298.

Dumanski, J., Pettapiece, W.W. and McGregor, R.J., 1998, Relevance of scale dependent approaches for integrating biophysical and socio-economic information and development of agroecological indicators, Nutrient Cycling in Agroecosystems, Vol. 50, pp. 13-22.

Goovaerts, P., 2000, Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall, Journal of Hydrology, Vol. 228, pp. 113-129.

Heathwaite, A.L., Quinn, P.F. and Hewett, C.J.M., 2005, Modelling and managing critical source areas of diffuse pollution from agricultural land using flow connectivity simulation, Journal of Hydrology, Vol. 304, pp. 446-461.

Hengl, T., Heuvelink, G.B.M. and Stein, A., 2004, A generic framework for spatial prediction of soil variables based on regression-kriging, Geoderma, Vol. 120, pp. 75-93.

Jenson, S.K. and Domingue, J.O., 1988, Extracting topographic structure from digital elevation data for geographic information system analysis, Photogrammetric Engineering & Remote Sensing, Vol. 54, pp. 1593-1600.

Kalivas, D.P., Triantakonstantis, D.P. and Kollias, V.J., 2002, Spatial Prediction Of Two Soil Properties Using Topographic Information, Global Nest Journal, Vol. 4, pp. 41-49.

Kim, S.M., Choi, Y., Suh, J., Oh, S., Park, H.D., Yoon, S.H. and Go, W.R., 2012, ArcMine: A GIS extension to support mine reclamation planning, Computers & Geosciences, Vol. 46, pp. 84-95.

Korea Meteorological Administration, Annual Climatological Report (2005 - 2014), http://www.kma.go.kr.

Lin, Y.P., chang, T.K. and Teng, T.P., 2001, Characterization of soil lead by comparing sequential Gaussian simulation, simulated annealing simulation and kriging methods, Environmental Geology, Vol. 41, pp. 189-199.

Little, L.S., Edwards, D. and Porter, D.E., 1997, Kriging in estuaries: as the crow flies, or as the fish swims?, Journal of Experimental Marine Biology and Ecology, Vol. 213, pp. 1-11.

Liu, X.M., Xu, J.M., Zhang, M.K., Huang, J.H., Shi, J.C. and Yu, X.F., 2004, Application of geostatistics and GIS technique to characterize spatial variabilities of bioavailable micronutrients in paddy soils, Environmental Geology, Vol. 46, pp. 189-194.

McCool, D.K., Brown, L.C. and Foster, G.R., 1987, Revised slope steepness factor for the Universal Soil Loss Equation, Transactions of the ASAE, vol. 30, pp. 1387-1396.

O'Callaghan, J.F. and Mark, D.M., 1984, The extraction of drainage networks from digital elevation data, Computer Vision. Graphics and Image Processing, Vol. 28, pp. 323-344.

Raaflaub, L.D., 2002, The Effect of Error in Gridded Digital Elevation Models on Topographic Analysis and on the

Distributed Hydrological Model TOPMODEL, UCGE Report No. 20163, Department of Geomatics Engineering, University of Calgary, Calgary, Alberta, 113p..

Rathbun, S.L., 1998, Spatial modeling in irregularly shaped regions: kriging estuaries, Environmetrics, Vol. 9, pp. 109-129.

Salgueiro, A.R., Avila, P.F., Pereira, H.G. and Santos Oliveira J.M., 2008, Geostatistical estimation of chemical contamination in stream sediments: The case study of Vale das Gatas mine (northern Protugal), Journal of Geochemical Exploration, Vol. 98, pp. 15-21.

Skøien, J.O., Merz, R. and Bl¨oschl, G., 2006, Top-kriging-geostatistics on stream networks, Hydrology and Earth System Sciences, Vol. 10, pp. 277-287.

Smith, R.A., Schwarz, G.E. and Alexander, R.B., 1997, Regional interpretation of water-quality monitoring data, Water Resources Research, Vol. 33, pp. 2781-2798.

Steiger, B., Webster, R., Schulin, R. and Lehmann, R., 1996, Mapping heavy metals in polluted soil by disjunctive kriging, Environmental Pollution, Vol. 94, pp. 205-215.

Tamas, J. and Kovacs, E., 2003, Evaluation of Migration of Heavy Metal Containing Sediment Resulting from Water Erosion using a Geo Information Model, EFITA 2003 Conference, Debrecen, Hungary, July 5-9, pp. 796-800.

Thornton, I., 1983, Applied Environmental Geochemistry, Academic Press, San Diego, USA, 501p.

Tobler, W., 1970, A computer movie simulating urban growth in the Detroit region, Economic Geography, Vol. 46, pp. 234-240.

Torgersen, C.E., Gresswell, R.E., and Bateman, D.S., 2004, Pattern detection in stream networks: quantifying spatial variability in fish distribution, Proceedings of the Second Annual International Symposium on GIS/Spatial Analyses in Fishery and Aquatic Sciences, Fishery GIS Research Group, Saitama, Japan.

Ver Hoef, J.M., Peterson, E. and Theobald, D., 2006, Spatial statistical models that use flow and stream distance, Environmental and Ecological Statistics, Vol. 13, pp. 449-464.

White, J.G., Welch, R.M. and Norvell, W.A., 1997, Soil zinc map of USA using geostatistics and geographic information systems, Soil Science Society of America Journal, Vol. 61, pp. 185-194.

Yenilmez, F., Kyter, N., Emil, M.K. and Aksoy, A., 2011, Evaluation of pollution levels at an abandoned coal mine site in Turkey with the aid of GIS, International Journal of Coal Geology, Vol. 86, pp. 12-19.

Yuan, L.L., 2004, Using spatial interpolation to estimate stressor levels in unsampled streams, environmental monitoring and assessment, Vol. 94, pp. 23-38.

Abstract

Geostatistical prediction of heavy metal concentration considering the path of stream

Sung-Min Kim
Dept. of Energy Systems Engineering
The Graduate School
Seoul National University

Erosion of heavy metals in mine waste rocks and tailings can contaminate the surrounding environment including soil and water. To prevent and reduce the damage, estimating the degree of pollution of stream and predicting the transport pathway are important. However, it is difficult to assess the contamination at the site because of topographic complexity and limited data. By using Geographic Information Systems (GIS), faster, less subjective and more reproducible prediction for extensive area can be obtained from a Digital Elevation Model (DEM). This study presents GIS modeling techniques to predict the heavy metal concentrations of the stream sediments for assessing the contamination and planning the reclamation. To estimate heavy

metal concentrations over the stream, a new kriging method which calculates the stream distances between samples was suggested, taking into account the topography and networks of streams. Combined with GIS-based flow analysis, the result presents more accurately predicted concentrations than other general kriging does. This paper describes the concept and details of the new technique, and outlines applications for this technique using synthetic and real-world datasets. The validation data indicate that the new technique is an effective approach for the prediction of heavy metal concentrations of stream sediments. It is also verified that catchment areas of samples derived from DEM can be used as an additional data for prediction if there is some correlation between the catchment area and the concentration of heavy metal. The proposed techniques can prioritize environmental hazards and give useful information for reclamation of stream networks. It is applicable to abandoned mines, industrial areas, and other various areas and countries.

Keywords : GIS, geostatistics, heavy metal, kriging, stream

Student Number : 2009-23182