

동북아시아 근해역 영양염의 시공간 모델링

Spatio–Temporal modeling of Nutrient in
Northeast Asia

2021. 11. 30.

과학기술연합대학원대학교

[보고서식 제2호] 최종보고서

UST Young Scientist 양성 사업 최종보고서

UST Young Scientist Research Program Final Report

과제번호 Grant Number		2020YS27					
과제명 Project Title	국문 Korean	동북아시아 근해역 영양염의 시공간 모델링					
	영문 English	Spatio-Temporal modeling of Nutrient in Northeast Asia					
스쿨/캠퍼스 School · Campus		한국해양과학기술원		전공 Major	응용해양과학		
UST 학생 UST Student	성명 Name	이기섭	학위과정 Course	박사과정	연락처 Contact	(E-mail) freelgs7@kiost.ac.kr (Mobile) 010 9378 6757	
지도교수 Advisor	성명 Name	조홍연	직급/직위 Position	책임연구원	연락처 Contact	(E-mail) hycho@kiost.ac.kr (Mobile) 010 2417 5711	
연구기간 Period	2020.12.01.~2021.11.30.			총연구비 Grants	10,000,000	원 KRW	

2020년도 UST Young Scientist 양성 사업 최종보고서를 붙임과 같이 제출합니다. 보고서에는 사실과 다른 내용이 포함되지 아니하였으며, 만약 허위사실이나 중대한 오류가 발견될 경우에는 그에 상응하는 불이익을 감수할 것임을 서약합니다.

I hereby submit the Final Report of UST Young Scientist Research Program 2020 as attached. To the best of my knowledge, all of the information included in my report is true and I will take any penalty if false information or serious error is found in my report.

첨부 : 최종보고서 1부.

Attachment: Final Report

2021. 11. 30.

UST 학생 : 이기섭 (서명)
UST Student : 이기섭 Signature

지도교수 : 조홍연 (서명)
Advisor : 조홍연 Signature

과학기술연합대학원대학교 총장 귀하

To the President of University of Science and Technology

요 약 문

동북아시아 인근 해역은 생산성이 높은 해역으로, 1차 생산을 주도하는 영양염 농도에 대한 관심이 높은 지역이다. 그러나 영양염 항목은 수온, 염분 등과 같이 수치 모델을 이용한 재분석 자료의 활용에 한계가 있는 실정이며, 현상의 분석을 주로 관측 자료에 의존한다. 따라서 관측 자료 기반의 시공간 통계 모델링을 이용한 신뢰할 만한 격자형 자료 생산에 대한 수요가 높으나 고급 시공간 통계 기법의 적용은 비전문가 영역에서 수행이 어려울 수 있다. 본 과제에서는 이러한 요구에 부응하고자 Spatio-Temporal Kriging 기법을 이용하여 동북아시아 인근 해역의 영양염 자료를 보간하는 연구를 수행했다. 보간은 10분 간격 평면 격자에 1개월 간격의 시간 해상도로 수행했다. 수온, 질산염, 인산염 관측 자료를 이용하여 1980년부터 2019년까지 40년 동안의 변화 양상을 모델링 했으며, 10 Fold Cross Validation을 통해 검증한 오차는 평균절대오차를 기준으로 수온 1.42, 질산염 1.87, 인산염 0.15로 나타났으며 수정된 결정계수는 0.93에서 0.97 수준으로 높게 나타났다. 본 연구를 통해 생산된 격자형 보간 자료는 검증 결과를 참고하여 동아시아 인근 해역에서 주요 영양염인 질산염과 인산염 항목의 장기 변화 양상 연구에 활용 가능할 것으로 판단된다. 또한 본 연구 과정에서 작성된 Spatio-Temporal Kriging 코드는 S/W 등록 및 연구자 Github에 등록하여 조건 없이 사용 가능하도록 했다. 이를 통해 적절한 시공간 모델링 기법의 활용 가능성을 높이며, 지속적인 수정을 통해 해양 관측 자료를 이용한 연구 분야의 발전할 수 있을 것으로 생각한다.

1. 연구개요

사회 전반에 걸쳐 공개된 자료의 활용 가치가 높다는 인식이 확산된 이후, 국가기관에서 생산되는 많은 자료들을 ‘공공데이터개방포털’과 같은 자료공개 플랫폼에서 쉽게 이용할 수 있는 환경이 되었지만 자연과학 분야의 경우 연구자 개인이 직접 관측한 자료가 대부분이고 연구 활동을 통해 가공된 개방 자료들은 부족함. 그림 1과 같이 JAMSTEC에서 공개하고 있는 Yasunaka et al.(2016)의 연구결과나 여러 위성 프로젝트에서 Level 2 이상으로 가공된 위성자료를 연구자가 이용하는 것처럼 ‘신뢰할 만한 방법으로’, ‘사용하기 쉽게’ 처리된 자료의 가치는 높음.

최근 국내외에서 자료 공유 체계를 통해 다양한 해양 자료들이 활발하게 이용되고 있으나 관측 자료가 부족한 영양염 항목은 그 수가 상대적으로 적음. 해류, 수온 등과 같이 역학 모델을 이용한 자료 생산과 비교하여 상대적으로 관측이 적고 재분석 자료로 예측한 값을 검증하는 과정이 쉽지 않기 때문에 소수의 관측 자료에 의존하여 간단한 보간 기법들을 사용하고 있음.

그럼에도 불구하고, 다양한 공간 보간 모델링 연구들이 수행되고 있으며, 대표적으로 Yasunaka et al.(2016)와 Zhou et al.(2014)의 사례가 있음. Yasunaka et al.(2016)에서는 북태평양 선박 관측 자료를 이용한 표층 영양염을 보간했으며, Zhou et al.(2014)에서는 Chesapeake Bay의 빈산소수괴 부피 변화를 연구함. 이 중 Yasunaka et al.(2016)는 연구에서 그치지 않고 JAMSTEC 홈페이지 등을 통해 연구 결과를 공유함.

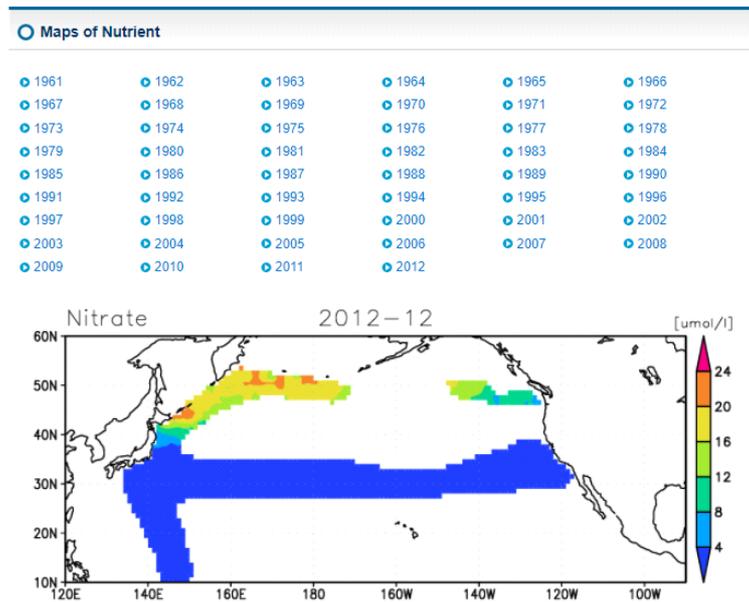


그림 1 JAMSTEC에서 Yasunaka et al.(2016)의 연구결과로 공유된 보간자료.

Zhou et al.(2014)에서는 3차원 Kriging 기법을 이용하여 용존산소 농도를 공간 모델링 했고, 이를 통해 빈산소수괴의 부피 변화를 정량적으로 분석하는 연구를 수행함. 그러나 관측 간격이 비교적 촘촘하여(약 15일) 시간에 대한 모델링은 포함하지 않음.

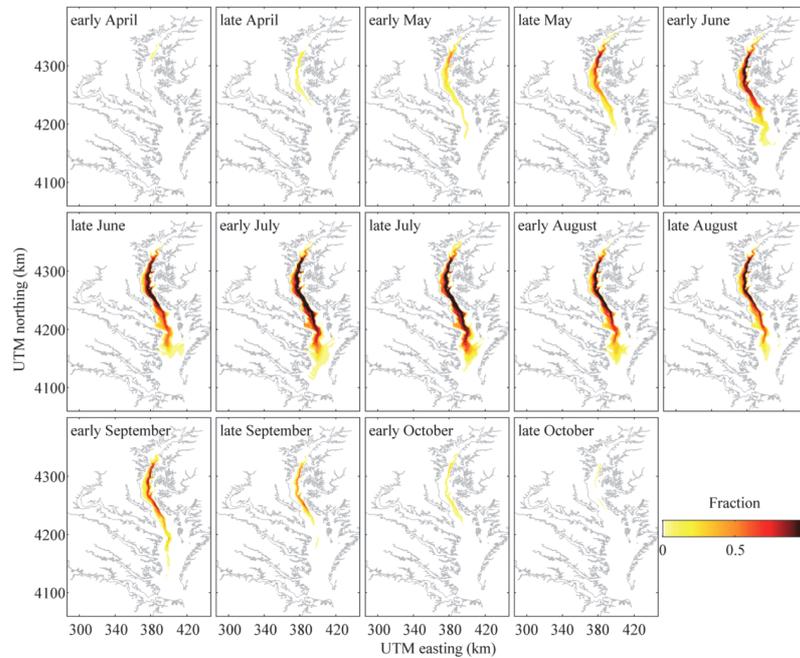


Fig. 2. Estimated frequency of hypoxia for early April to late October. Frequency is expressed as the fraction of years from 1985 to 2010 when the dissolved oxygen (DO) concentration is estimated using UK to have been below 2 mg L^{-1} at a given location for at least one point in the water column.

그림 2 Zhou et al.(2014)의 빈산소 수괴 변동 양상 연구(평면)

상기 사례들을 참고하여, 동아시아 연근해를 대상으로 해양 시공간 관측자료에 대하여 Spatio-Temporal Modeling을 통해 시공간적으로 불균질한 자료들을 격자형태로 재생산하고, 분석 코드와 결과 자료를 공유하고자 함.

II. 연구 수행내용 및 과정

자료수집 및 기본 정보

연구에 사용한 자료 중 모델링 대상 관측 자료는 국립수산과학원(National Institute of Fisheries Science, NIFS) 정선관측 자료(이하 NIFS 자료)와 일본 기상청(Japan Meteorological Agency, JMA) 연구선 관측 자료(이하 JMA 자료)이다. NIFS 자료와 JMA 자료는 다음 링크를 통해 다운로드 받을 수 있다(2021-11-17 기준).

NIFS 자료

https://www.nifs.go.kr/kodc/soo_list.kodc

JMA 자료

https://www.data.jma.go.jp/gmd/kaiyou/db/vessel_obs/data-report/html/ship/ship_e.php

NIFS 자료는 엑셀 파일(.xls)로 다운로드 가능하며 Table 형태로 가공되어 있다. JMA 자료는 126 bytes의 ASCII 자료로 .E 확장자 파일이며 관측 정보와 관측 항목을 일정 간격으로 배치하는 Fixed Width Format이다. 자료 구조에 대한 설명은 JMA에서 제공하는 메타정보를 참조한다(JMA, 2015).

JMA 자료의 경우 선박별, 항차별 자료가 나누어져 있기 때문에 해당 자료를 수동으로 다운로드 받기 위해서는 상당한 시간이 소요될 것으로 판단했다. 본 연구에서 1965년부터 2020년 일부 자료를 취합한

결과 총 2063개의 Hydrographic data가 존재했으며 해당 자료들을 효율적으로 수집하기 위해 Crawling 코드를 작성해서 활용했다. 연구 수행에 많은 부분을 차지하는 자료 수집 과정을 자동화를 통해 획기적으로 줄였다.

각 자료는 각각 우리나라와 일본 인근에서 관측한 수심 방향 Profile이며, 관측 좌표, 시간, 수심이 함께 기록된 4차원 자료이다. 위경도로 기록된 좌표 체계는 거리 계산의 편의를 위해 TM좌표계로 변환했으며, 투영 방법은 우리나라 주변 지역에 대해 왜곡이 적은 것으로 보고된 Lambert 정적방위도법을 사용했다(이와 조, 2012). 또한 여러 관측 항목들 중 수온, 염분, ph, 용존산소, 인산염, 아질산염, 질산염, 규산염 자료를 선택하여 취합했다.

연직 방향 좌표를 설정하기 위한 모델링 대상 지역의 수심 자료는 미해양대기청(National Oceanic and Atmospheric Administration, 이하 NOAA)의 ETOPO1 자료를 사용했다. 해당 자료는 R 패키지 'marmap'을 통해 원하는 해상도로 조정하여 다운로드 받을 수 있으며 모델링 해상도에 맞추어 10" 격자로 조정했다.

수평, 수직 자료 분포는 다음 그림과 같으며 시공간적으로 불균질한 분포 특성을 보임. 또한 수심 방향으로는 약 1,000m 이내에 대부분의 자료가 분포한다.

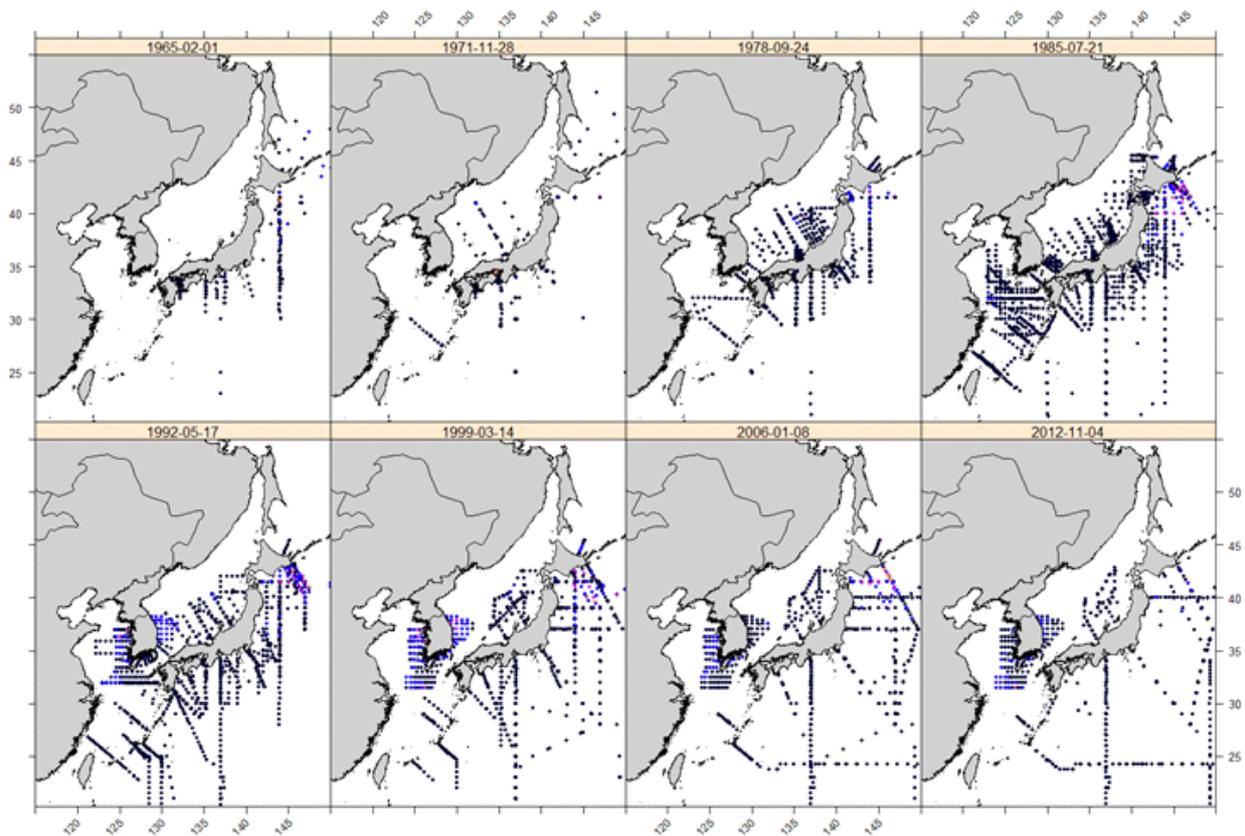


그림 3 관측 자료의 시간에 따른 공간 분포 변화

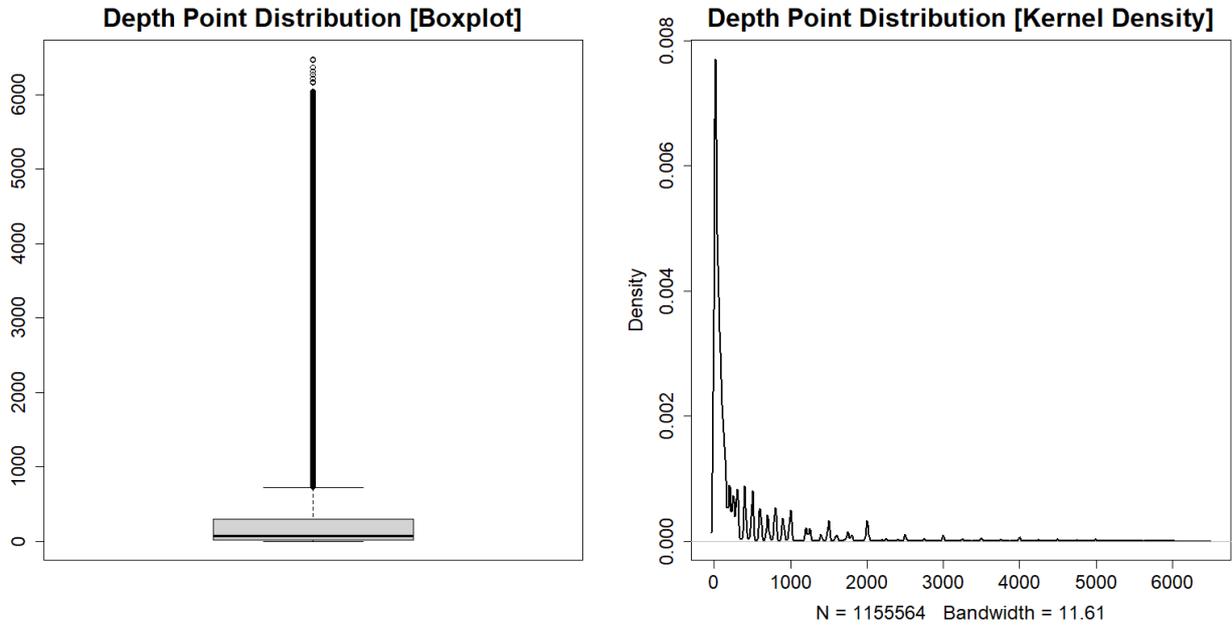


그림 4 관측 자료의 수심에 따른 분포. 1,000m 이하 수심에 대부분의 자료가 분포함.

Spatio-Temporal Kriging

관측된 시공간 자료를 이용하여 미관측 지점을 예측하기 위해서, 다음과 같이 분산을 이용한 계수 추정 과정을 거침. 이를 Variogram 모델링이라 하며 거리에 따른 자료의 공분산 변화를 이용하여 미관측 지점을 계산할 때 관측 지점들의 가중치를 계산하는 방법임.

$$\sum_{i=1}^n w_i \sigma_{li}^2 = \sigma_{0l}^2; l = 1, 2, \dots, n$$

이를 행렬식으로 간략하게 표현하면 다음과 같음.

$$C \begin{bmatrix} W \\ \Lambda \end{bmatrix} = C_0$$

$$\begin{bmatrix} \sigma_{11}^2 & \sigma_{12}^2 & \dots & \sigma_{1n}^2 & -1 \\ \sigma_{21}^2 & \sigma_{22}^2 & \dots & \sigma_{2n}^2 & -1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \sigma_{n1}^2 & \sigma_{n2}^2 & \dots & \sigma_{nn}^2 & -1 \\ 1 & 1 & \dots & 1 & 0 \end{bmatrix} = C, \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix} = W, [\lambda] = \Lambda, \begin{bmatrix} \sigma_{01}^2 \\ \sigma_{02}^2 \\ \vdots \\ \sigma_{0n}^2 \\ 1 \end{bmatrix} = C_0$$

위 식은 관측 자료만을 사용하여 가중합을 계산하는 Simple Kriging 방정식이며, 여기에 가중치의 합이 1이 되도록 하는 제약조건과 공간의 평균적인 변화를 반영하는 External Regressor를 추가한 형태가 아래와 같은 Universal Kriging임.

$$\sum_{i=1}^n w_i \sigma_{ii}^2 + \sum_{j=1}^m \lambda_j f_j(x_0) = \sigma_{0i}^2; i = 1, 2, \dots, n$$

$$\begin{bmatrix} C & X \\ X^T & 0 \end{bmatrix} \begin{bmatrix} W \\ \Lambda \end{bmatrix} = \begin{bmatrix} C_0 \\ X_0 \end{bmatrix}$$

위 식에서 관측 자료의 공분산 행렬 C 는 특정 형태를 가정한 모델로 최적 계수를 찾아 적합시키며 본 연구에서는 아래와 같은 구형모델을 사용함. 여기서 a 는 상관거리를 나타내는 Range, b 는 공분산 하한선을 나타내는 Nugget, $C_0 + b$ 는 공분산 상한선을 나타내는 Sill 매개변수임.

$$C_0 \left(1.5 \left(\frac{h}{a} \right) - 0.5 \left(\frac{h}{a} \right)^3 \right) + b$$

위 모델을 각각 시간, 공간, Joint Variogram의 합으로 표현하면 아래 식과 같으며 Joint Variogram 항은 κ 매개변수를 이용하여 공간(h)에 대한 시간(u) 차원의 결합 가중치 조정으로 모델의 형태를 조정함. 이러한 형태를 Sum-Metric 모델이라고 하며 Product-Sum 모델과 함께 Spatio-Temporal Kriging 분야에서 많이 사용되는 결합 방식임(Pebesma and Heuvelink, 2016).

$$C_{sm}(h, u) = C_s(h) + C_t(u) + C_{jt} \left(\sqrt{h^2 + (\kappa u)^2} \right)$$

시공간 Variogram인 $C_{sm}(h, u)$ 의 매개변수는 L-BFGS-B 최적화 기법을 이용해 관측 자료를 이용하여 계산한 Experimental Variogram과 Sum-Metric 함수 $C_{sm}(h, u)$ 와의 오차 제곱합을 최소로 하는 매개변수를 결정함.

자료의 보간 간격은 시간적으로, 1980-2019년 자료에 대하여 모의 대상 기간 전후로 30일씩 Buffer를 부여하여 격월 또는 그 이상 간격으로 관측된 영양염 자료를 1개월 간격으로 보간했으며, 공간적으로는 경도 약 122-155.8, 위도 약 20.5-45.6 영역의 자료를 사용함. 이 과정에서 m 단위의 거리 계산을 위하여 TM 좌표계로 변환하였으며 투영법은 위도 34.53333, 경도 137.0698를 원점으로 Lambert 정적방 위도법을 사용하여 투영함.

III. 연구 수행결과 및 목표달성 수준

연구 수행 결과

앞서 제시한 방법을 사용하여 계산된 Spatio-Temporal Variogram은 그림 5와 같음.

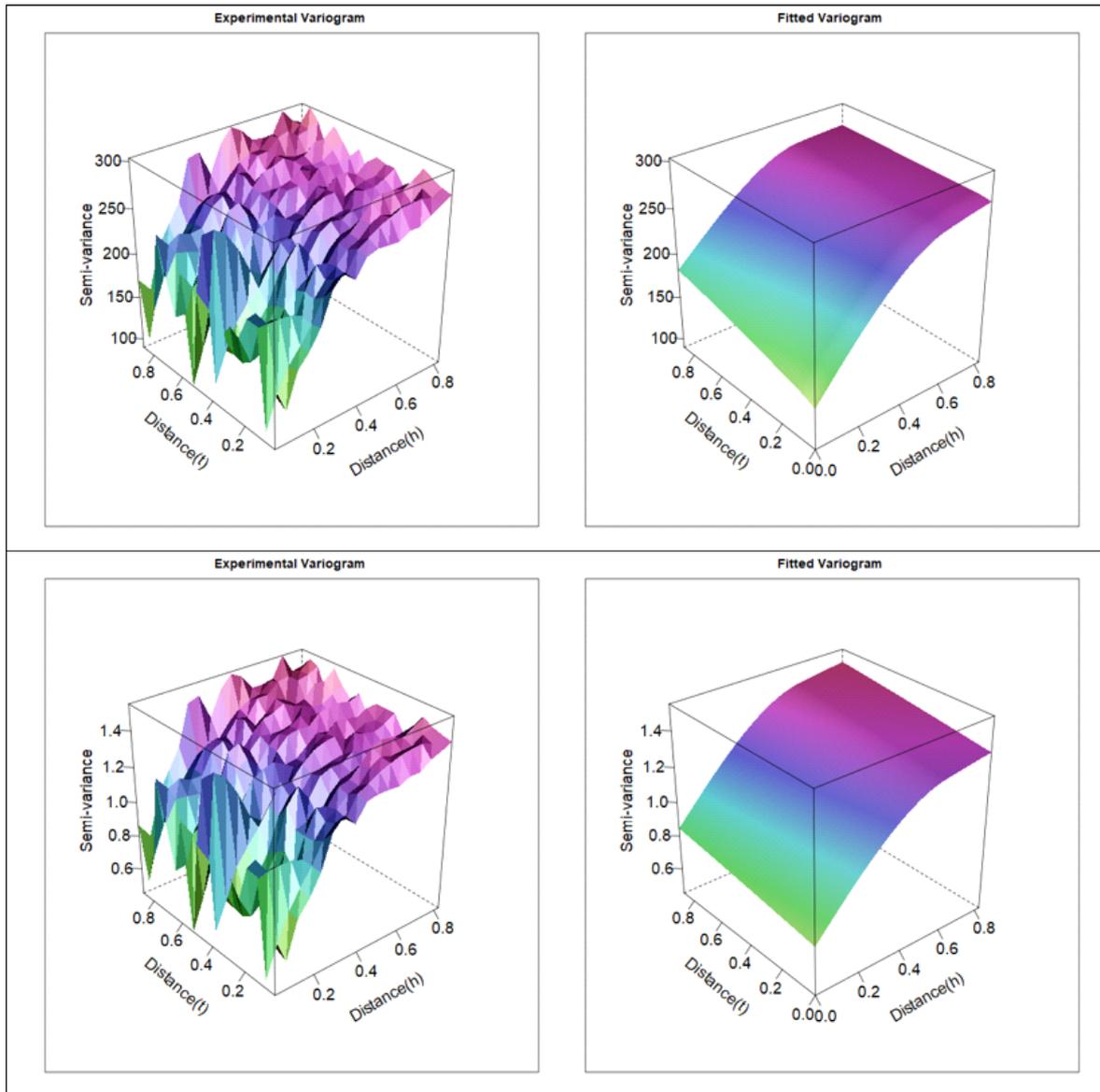


그림 5 2012년도 시공간 Variogram 모델링 결과. 상단의 그림은 질산염, 하단은 인산염의 Variogram이며 각 그림의 좌측은 Experimental Variogram, 우측은 Modeled Variogram.

상기 Variogram 모델을 이용하여 보간한 결과는 그림 6과 같음. Kriging의 특성인 추정 오차의 정도를 함께 계산할 수 있으며, 추정오차는 NIFS의 정점과 JMA의 연구선 운항 경로에 가까울수록 적은 것을 알 수 있음. Kriging은 원하는 위치의 값을 계산 가능하므로, 보간된 질산염 자료는 ETOPO 자료의 수심값에 따라 각 격자에서 최대 수심까지만 계산하였고, 이에 따라 깊은 수심 영역을 mapping 할 때 blank로 나타남.

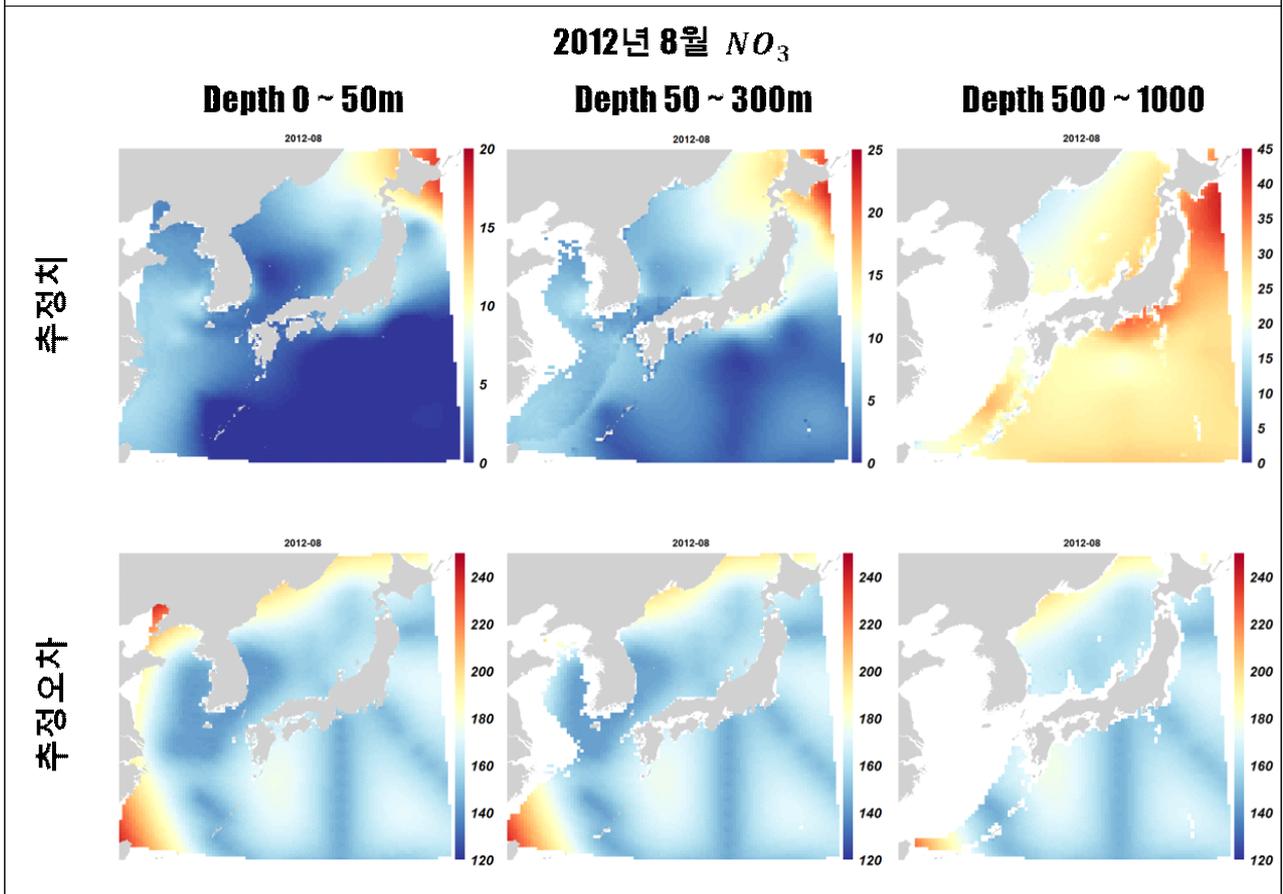
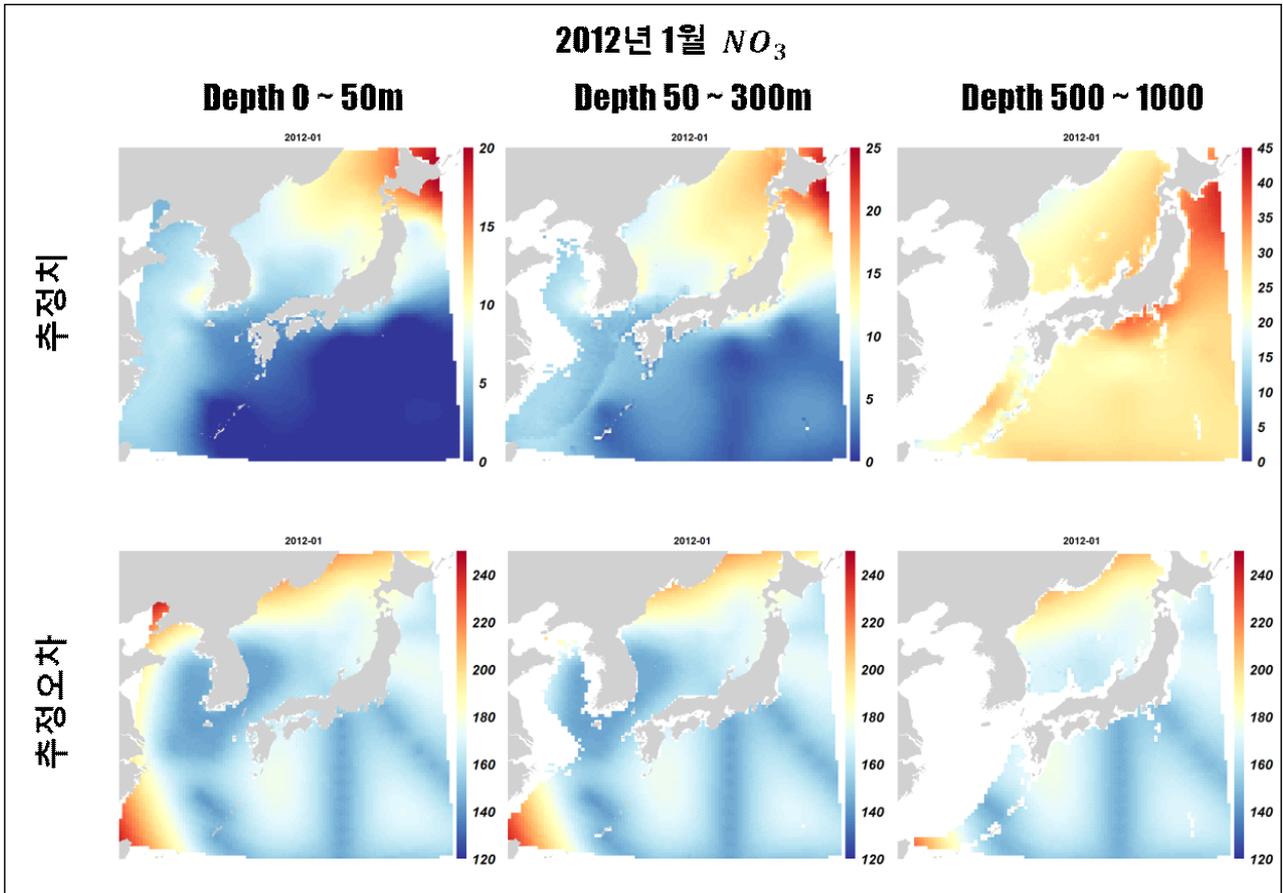


그림 6 2012년 1월(위)과 8월(아래)의 질산염 보간 결과(각 월별 상단 그림)와 추정 오차 지도(각 월별 하단 그림)

계산 결과에 대한 오차 평가는 10-Fold Cross Validation을 사용함. 1년 단위 자료를 10등분하여 1개 그룹을 제외하고 나머지 자료를 이용하여 보간한 뒤, 제외했던 그룹의 자료들의 위치의 추정값과 관측값을 비교하는 방식으로 성능평가를 수행함. 수행 결과는 그림 7 및 표 1과 같음. 수온, 질산염, 인산염에 대한 오차 평가 결과 평균제곱근편차(Root Mean Squared Error, RMSE)는 각각 2.05, 2.79, 0.22로 나타났으며, 일부 자료에서 분산이 크게 나타나는 현상으로 인해 평균절대오차(Mean Absolute Error, MAE)를 검토한 결과 RMSE에 비해 감소하는 것으로 나타남. 이는 1:1 대응선에 가까운 영역에 대부분의 자료가 모여 있기 때문으로 판단됨. 항목별 관측 범위를 벗어나는 값은 일반크리깅의 특성으로 인해 외삽 영역에서 발생하는 것으로 판단되므로, 적절한 경험적 상하한 경계값을 부여하는 방법을 고려할 수 있음. 보정된 결정계수(Adjusted R Squared)는 0.93-0.97로 높은 값을 나타냄.

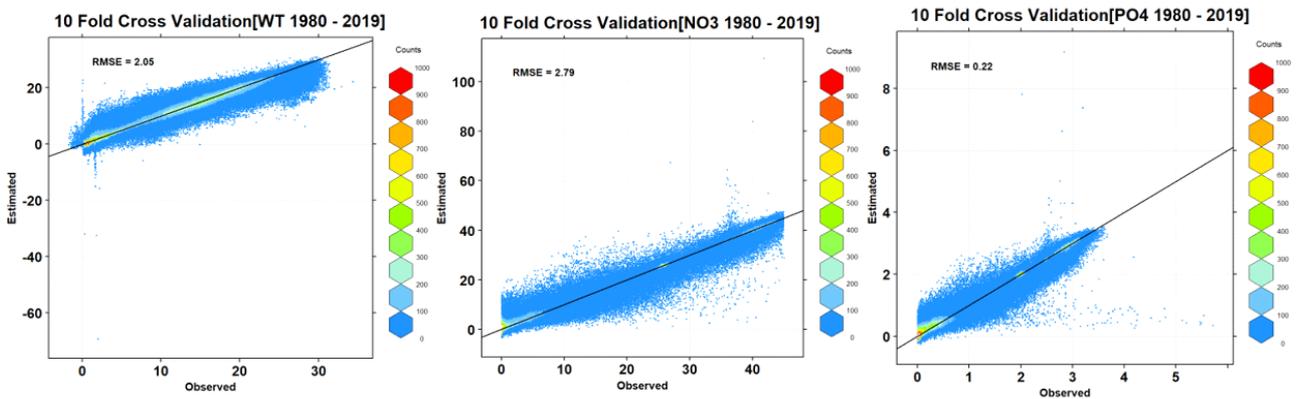


그림 7 10 Fold-Cross Validation으로 관측값과 예측값을 비교한 결과. 일부 자료들은 1:1 대응선(검은 선)을 벗어나지만 자료의 밀도는 대응선 근처에서 높게 나타남.

표 2 10-Fold Cross Validation을 이용한 Spatio-Temporal Kriging 오차 평가 결과

평가 지표	WT(°C)	NO ₃ (μmol/L)	PO ₄ (μmol/L)
RMSE	2.05	2.79	0.22
MAE	1.42	1.87	0.15
R_{adj}^2	0.93	0.97	0.96

목표 달성 수준

Spatio-Temporal Kriging 기법을 이용한 4차원 관측 자료를 이용한 격자형 자료 재생산 연구를 수행했고, 자료 수집, 모델 선정 및 성능 평가를 거쳐 재현 가능한 계산 코드를 github에 업로드 및 S/W로 등록하여, 당초 제안서에 제시한 연구 목표를 대부분 달성함.

다만, Yasunaka et al.(2016)와 같이 생산된 자료의 배포 단계까지 구현은 미비함. 향후 관련 분야 전문가와 상의를 통해 해당 보간 결과에 대한 타당성 검토 후 공유할 수 있는 적절한 플랫폼을 찾아 공개할 계획.

IV. 연구 성과 및 관련 분야 기여도

본 연구는 주 연구 주제인 동아시아 인근 해역의 영양염 자료 뿐만 아니라 다양한 해양 관측 자료에 대하여 활용할 수 있을 것으로 판단됨. 평면 공간 좌표, 수심, 시간이 기록된 형태의 관측 자료라면 모두 적용 가능하며, 보간 기법 또한 Kriging 기초 이론을 숙지한 사람이 사용이 용이한 수준으로, 최소한의 입력자료를 받아 실행되도록 디자인 했으며 오픈소스 프로그램인 R을 이용하여 작성된 코드를 공개하여 무료 사용 및 수정이 가능함. 향후 보간된 자료의 업로드 및 모델 개선에 따라 자료를 갱신하면서 보다 정확한 재생산 자료를 제공할 예정이며, 동 자료를 이용한 장기 경향 분석에 활용될 수 있을 것으로 판단됨.

V. 연구결과의 활용방안

본 연구를 통해 중권역 지역해의 관측 자료를 격자형의 빈틈없는 자료로 재생산하고 해당 자료의 해양학적 활용성을 평가하기 위해 해양학회의 화학세션에서 소개함. 또한 개인 분석 코드 수준에서 함수화를 거친 S/W개발을 통해 당초 제시한 공간분석 분야 비전공자들의 활용성 증대에 기여함.

본 연구 성과를 이용하여 학위논문과 국외 학술지 투고를 위한 원고를 작성 중이며, 향후 해양 분야의 다양한 자료에 관련 기법을 적용할 수 있을 것으로 판단됨.

후속 과제로 Zhou et al.(2014)의 연구를 참고하여 국내 빈산소수괴 부피 변화를 추정하고 원인을 분석하는 연구와 단순 보간 시 문제가 발생하는 태풍 발생 시 기압의 시공간 분포 예측 연구를 기획 중임.

VI. 참고문헌

Pebesma, E., & Heuvelink, G. (2016). Spatio-temporal interpolation using gstat. *RFID Journal*, 8(1), 204-218.

Yasunaka, S., Ono, T., Nojiri, Y., Whitney, F. A., Wada, C., Murata, A., ... & Hosoda, S. (2016). Long-term variability of surface nutrient concentrations in the North Pacific. *Geophysical Research Letters*, 43(7), 3389-3397.

Zhou, Y., Scavia, D., & Michalak, A. M. (2014). Nutrient loading and meteorological conditions explain interannual variability of hypoxia in Chesapeake Bay. *Limnology and Oceanography*, 59(2), 373-384.

이상일, & 조대현. (2012). 대한민국 주변도 제작을 위한 최적의 지도 투영법 선정: GIS-기반 투영 왜곡 분석. *한국지도학회지*, 12(3), 1-16.

VII. 주요 연구실적

1. 연구실적 요약

학술지 논문	학회논문	특허 출원	특허 등록	대내외 수상
건	1 건	건	1 건	건

2. 학술지 논문

No.	논문제목	학술지명	저자순	게재연월	국내/국외	SCI 여부

3. 학회 논문

No.	논문제목	학회명	저자순	게재연월	국내/국외	발표 유형
1	동아시아 인근 해역 영양염의 시공간 모델링	한국해양학회	1	2021.11.4	국내	Oral

4. 특허출원 · 등록

No.	특허명	출원일	등록일	국가명	등록번호	비고
1	4차원 자료 보간을 위한 Spatio-Temporal Kriging 코드	2021.11.30		대한민국		

5. 대내 · 외 수상

No.	수상명	수상등급	수여기관	수여일자

VIII. 국외출장 현황

No.	출장기간	출장자	국가	지역	출장목적	학회/ 행사명	발표제목	연구비 사용여부	
								참가비	여비
1	코로나19로 국외출장 없음								

IX. 전문가 교류내역

No.	전문가 정보			교류내용	교류일자	비고
	성명	소속	직위			
1	박노욱	인하대학교	교수	자문	2021.4.15	
2	이정호	SK텔레콤	매니저	자문	2021.9.1	
3						
4						
5						

X. 연구비 집행 현황

세 목	책정금액(A)	집행금액(B)	집행률(B/A)*100
학생인건비	2,000,000	2,000,000	100
연구시설·장비비	0	0	100
연구재료비	0	0	100
연구활동비	8,000,000	4,270,940	57.29
합 계	10,000,000	5,805,051	61.17

XI. 지도교수 의견

주어진 과제명에 적합한 시공간 모델링을 수행한 것으로 판단됨. 수행결과에 대하여 전문가의 자문과 수정보완 과정을 거쳤으며, 학술발표 등을 통하여 검증을 받았음. UST 연구비 지원에 감사드립니다.