

표지

(뒷면)	(측면)	(앞면)
↑ 7cm ↓ 주의 (편집 순서8) (16 포인트 고딕체) ↑ 7cm ↓	딥러닝 기법에 기반한 지구시스템 모델 해양 자료동화 기법 개발 서울대 학교 산학협 력단 ↑ 5cm ↓	↑ 7cm ↓ 딥러닝 기법에 기반한 지구시스템 모델 해양 자료 동화 기법 개발 Development of Oceanic Data Assimilation System for Earth System Model Using Deep Learning 극지연구소 서울대학교 산학협력단 ↓ 7cm ↑

제 출 문

극지연구소장 귀하

본 보고서를 “지구시스템모델 기반 북극-한반도 통합 재해기상 예측 시스템 (KP0PS-Earth)의 개발 및 활용(본과제명)” 과제의 위탁연구 “딥러닝 기법에 기반한 지구시스템 모델 해양 자료 동화 기법 개발(위탁과제명)” 과제의 최종보고서로 제출합니다.



(본과제) 총괄연구책임자 : 김주홍

위탁연구기관명 : 서울대학교 산학협력단

위탁연구책임자 : 함유근

위탁참여연구원 : 이수빈

“ : 김지광

“ : 이정길

“ : 성민규

“ : 이현정

“ : 강근형

보고서 초록

위탁연구과제명	딥러닝 기법에 기반한 지구시스템 모델 해양 자료 동화 기법 개발				
위탁연구책임자	함유근	해당단계 참여연구원수	6	해당단계 연구비	147,000,000
연구기관명 및 소속부서명	서울대학교 산학협력단		참여기업명		
국제공동연구	상대국명 : 상대국연구기관명 :				
요약(연구결과를 중심으로 개조식 500자이내)				보고서 면수	62
본 연구는 딥러닝 기반 자료동화 기법을 활용하여 전 지구 3차원 해양 온도 및 염분 재분석장과 초기 조건을 생성할 수 있는 해양 초기화 시스템을 개발하는 것을 목표로 수행되었다. 해양 관측 자료는 공간적·시간적으로 불연속적으로 분포하는 특성이 있어, 이를 효과적으로 반영하기 위한 자료동화 기법의 고도화가 요구된다. 본 연구에서는 이러한 한계를 극복하기 위해 부분 합성곱 신경망과 적대적 생성 신경망을 결합한 딥러닝 기반 자료동화 구조를 설계하고, 전 지구 규모의 3차원 해양 온도 및 염분 자료동화 시스템을 구축하였다. 1차년도에는 관측 결측을 고려할 수 있는 딥러닝 기반 자료동화 기법을 정립하고, 전 지구 3차원 해양 온도 자료동화 시스템을 구축하였다. 2차년도에는 이를 확장하여 해양 염분 자료동화 및 온도-염분 다변량 자료동화 기법을 개발하고, 전 지구 해양 재분석장을 시범적으로 생산하여 품질을 검증하였다. 3차년도에는 지구시스템 모델 CESM2 기반 해양 성분 모형에 적용 가능한 3차원 너징 시스템을 구축하고, 딥러닝 기반 재분석장을 활용한 장기간 해양 초기장을 생산 및 검증하였다. 그 결과, 본 연구에서는 전 지구 규모에서 해양 온도 및 염분을 동시에 고려한 초기화 기술을 확보하였으며, CESM2 기반 예측 시스템으로의 연계 가능성을 확인하였다. 본 연구 성과는 향후 계절 예측 및 기후 예측 정확도 향상에 기여할 수 있을 것으로 기대된다.					
색 인 어 (각 5개 이상)	한 글	딥러닝, 자료동화, 해양온도, 해양염분, 다변량 자료동화, 해양 초기화			
	영 어	Deep learning, Data assimilation, Ocean temperature, Ocean salinity, Multivariate data assimilation, Ocean initialization			

요 약 문

I. 딥러닝 기법에 기반한 지구시스템 모델 해양 자료 동화 기법 개발

II. 연구개발의 목적 및 필요성

극 지역 해양의 물리적 변동성은 해빙 변화와 대기-해양 상호작용을 통해 중위도 기후 변동성과 계절 예측 성능에 직접적인 영향을 미친다. 따라서 극 지역 해양 상태의 정확한 초기화는 극 지역뿐만 아니라 중위도 기후 예측 정확도 향상을 위해서도 필수적이다. 그러나 극 지역은 관측 인프라가 제한적이어서 장기적·정량적 해양 변동성 분석에 큰 불확실성이 존재한다.

이러한 한계를 극복하기 위한 방안으로 관측 자료와 수치모형을 결합한 자료동화 기반 재분석장 생산의 필요성이 지속적으로 제기되고 있다. 기존 앙상블 칼만 필터와 같은 동적 오차 추정 기법은 높은 정확도를 제공할 수 있으나, 막대한 계산 비용으로 인해 고해상도 해양 모형에의 적용에는 현실적인 제약이 따른다. 이에 따라 계산 효율성과 비선형 학습에 강점을 지닌 딥러닝을 활용한 3차원 해양 자료동화 프레임워크 개발이 요구되며, 이는 극 지역 해양 및 해빙 예측과 중위도 계절 예측 성능 향상에 기여할 것으로 기대된다.

III. 연구개발의 내용 및 범위

본 연구는 딥러닝 기반 자료동화 기법을 활용하여 전 지구 3차원 해양 온도 및 염분 초기화 기술을 개발하고, 이를 지구시스템 모델(CESM2)에 적용 가능한 형태로 확장하는 것을 목표로 수행되었다. 연구는 단계적으로 진행되었으며, 1차년도에는 관측 결측을 고려할 수 있는 딥러닝 기반 해양 온도 자료동화 기법을 구축하고, 전 지구 해양 관측·재분석 자료를 수집·데이터베이스화하였다. 2차년도에는 해양 염분 자료동화 및 온도-염분 다변량 자료동화 기법을 개발하고, 이를 적용한 전 지구 해양 재분석장을 시범적으로 생산하였다. 3차년도에는 딥러닝 기반 재분석장을 CESM2 해양 모형에 연계하기 위한 3차원 너징 시스템을 구축하고, 장기간 전 지구 해양 초기장을 생산 및 검증하였다. 이를 통해 본 연구는 전 지구 규모의 해양 초기화 기술 개발부터 지구시스템 모델 적용까지를 포괄하는 연구 범위를 설정하고, 예측 시스템으로의 확장 가능성을 확보하였다.

IV. 연구개발결과

본 연구에서는 딥러닝 기반 자료동화 기법을 활용하여 전 지구 3차원 해양 온도 및 염분 재분석장과 초기 조건을 생성할 수 있는 해양 초기화 시스템을 구축하였다. 관측 결측을 고려한 딥러닝 자료동화 구조를 통해 계산 효율성과 재분석 품질을 동시에 확보하였으며, 다변량 자료동화 기법을 적용하여 해양 내부 구조의 물리적 일관성을 향상시켰다. 또한, 개발된 재분석장을 지구시스템 모델(CESM2)에 연계 가능한 3차원 너징 시스템으로 확장하고, 장기간 전 지구 해양 초기장을 안정적으로 생산함으로써 예측 시스템 적용 가능성을 검증하였다.

V. 연구개발결과의 활용계획

본 연구는 극 지역 해양 및 해빙 변동성 분석과 중위도 계절 예측 성능 향상을 위한 기초 자료로 활용될 수 있다. 특히 딥러닝 기반 해양 초기화 기술은 계절·중장기 기후 예측 모델의 초기장 생성 과정에 적용되어 예측 정확도 개선에 기여할 수 있다. 또한, 본 연구에서 생산된 전 지구 해양 재분석장은 해양·기후 연구, 정책 수립, 예측 정보 서비스 등 다양한 분야로의 확장이 가능하며, 향후 지구시스템 모델 기반 예측 시스템과의 연계를 통해 실질적인 활용 가치가 높을 것으로 기대된다.



S U M M A R Y

(영 문 요 약 문)

I. Development of Oceanic Data Assimilation System for Earth System Model Using Deep Learning

II. Purpose and Necessity of R&D

Physical variability in polar oceans directly affects mid-latitude climate variability and seasonal prediction skill through changes in sea ice and atmosphere-ocean interactions. Therefore, accurate initialization of polar ocean states is essential not only for polar regions but also for improving seasonal climate prediction over mid-latitudes. However, due to limited observational infrastructure in polar regions, substantial uncertainties remain in long-term and quantitative analyses of ocean variability. To overcome these limitations, the production of reanalysis fields based on data assimilation that integrates observational data with numerical models has been increasingly emphasized. While dynamic error estimation methods such as the Ensemble Kalman Filter can provide high accuracy, their substantial computational cost imposes practical limitations on applications to high-resolution ocean models. Consequently, the development of a three-dimensional ocean data assimilation framework based on deep learning, which offers high computational efficiency and strong capability for nonlinear learning, is required. Such a framework is expected to contribute to improved prediction of polar ocean and sea ice conditions as well as enhanced seasonal prediction skill in mid-latitude regions..

III. Contents and Extent of R&D

This study aimed to develop a deep learning-based data assimilation technique for global three-dimensional ocean temperature and salinity initialization and to extend it into a form applicable to an Earth system model (CESM2). The research was conducted in a phased manner. In the first year, a deep learning-based ocean temperature data

assimilation method capable of accounting for observational gaps was developed, and global ocean observational and reanalysis datasets were collected and databaseized. In the second year, ocean salinity data assimilation and temperature-salinity multivariate data assimilation techniques were developed, and global ocean reanalysis fields were produced on a pilot basis. In the third year, a three-dimensional nudging system was constructed to link the deep learning-based reanalysis fields with the CESM2 ocean model, and long-term global ocean initial conditions were generated and validated. Through these efforts, this study established a research scope encompassing global ocean initialization technology development through application to an Earth system model, thereby securing the potential for expansion into a prediction system.

IV. R&D Results

This study successfully constructed an ocean initialization system capable of generating global three-dimensional ocean temperature and salinity reanalysis fields and initial conditions using deep learning-based data assimilation techniques. By employing a data assimilation structure that accounts for observational gaps, both computational efficiency and reanalysis quality were achieved. Furthermore, the application of multivariate data assimilation enhanced the physical consistency of internal ocean structures. The developed reanalysis fields were extended into a three-dimensional nudging system applicable to an Earth system model (CESM2), and long-term global ocean initial conditions were stably produced, thereby demonstrating the feasibility of application to a prediction system.

V. Application Plans of R&D Results

The outcomes of this study can be utilized as fundamental data for analyzing polar ocean and sea ice variability and for improving seasonal prediction performance in mid-latitude regions. In particular, the deep learning-based ocean initialization technology can be applied to the generation of initial conditions for seasonal and medium- to long-term climate prediction models, contributing to enhanced prediction accuracy. In addition, the global ocean reanalysis fields produced in this study can be extended to various fields, including ocean and climate research, policy development, and prediction information services. Through integration with Earth system model-based prediction systems, these results are expected to offer substantial practical value.

목 차

제 1 장 서론

제 2 장 국내외 기술개발 현황

제 3 장 연구개발수행 내용 및 결과

제 4장 연구개발목표 달성도 및 대외기여도

제 5 장 연구개발결과의 활용계획

제 6 장 연구개발과정에서 수집한 해외과학기술정보

제 7 장 참고문헌



제 1 장 서론

1절 연구개발의 목적 및 필요성

해양의 물리적 변동성은 극 지역 해양 및 해빙 변수의 계절 예측 성능에 직접적인 영향을 미칠 뿐만 아니라, 대기의 저주기(low-frequency) 변동성을 조절하는 핵심 요인으로 작용. 특히 극 지역에서 발생하는 해양 순환 변화, 수온 및 염분 구조의 변동, 해빙의 생성과 소멸 과정은 대기-해양 상호작용을 통해 중위도 기후 시스템에까지 영향을 미치는 것으로 알려져 있음. 이러한 극-중위도 원격 연계(teleconnection) 메커니즘은 계절 규모의 기후 예측 정확도를 좌우하는 중요한 요소로 작용하며, 이에 따라 극 지역 해양 상태를 정확하게 초기화하는 기술은 극 지역뿐만 아니라 중위도 지역의 계절 예측 성능 향상을 위해서도 필수적인 요소로 인식되고 있음.

극 지역 해양의 변동성과 이로 인해 유발되는 해빙 변수들의 경년 변동성(interannual variability)은 중위도 대기 순환 패턴 및 기후 변동성을 결정짓는 주요 요인 중 하나로 보고되고 있음. 예를 들어, 해빙 면적과 두께의 변화는 알베도(albedo) 및 해양-대기 열 교환 과정을 변화시켜 중위도 지역의 기온, 강수 패턴, 극 소용돌이(polar vortex) 강도 등에 영향을 미침. 이러한 연구 결과들은 극 지역 해양 변동의 정확한 표현이 중위도 기후 예측 성능 개선에 직결됨을 시사하며, 이를 위해서는 현실적인 극 지역 해양 상태의 초기화가 필수적임을 의미.

그러나 극 지역은 관측 인프라가 매우 제한적이며, 장기간 연속 관측 자료 확보가 어려움. 위성 관측의 경우에도 구름 등의 영향으로 자료 품질과 신뢰도에 한계가 존재. 이로 인해 극지 해양 환경 변화에 대한 정량적 분석과 장기적 변동성 평가에는 상당한 불확실성이 수반됨. 이러한 관측 한계를 극복하기 위한 대안으로, 관측 자료와 수치모형을 결합하는 극지 해양 자료동화 기법을 활용한 장기간 재분석장(reanalysis) 생산의 필요성이 지속적으로 제기. 자료동화 기반 재분석장은 시공간적으로 일관된 극지 해양 및 해빙 변수를 제공함으로써 기후 변동성 분석 및 예측 연구의 신뢰도를 향상시킬 수 있음.

시공간적으로 발달하는 모형 오차를 실시간으로 추정할 수 있는 앙상블 칼만 필터(Ensemble Kalman Filter, EnKF)와 같은 방법은 이론적으로 매우 높은 정확도의 초기장을 생성할 수 있는 장점을 가짐. 그러나 다수의 앙상블 멤버를 동시에 적분해야 하는 계산 구조로 인해 막대한 계산 비용이 요구되며, 이는 고해상도 해양 순환 모형에 적용하는 데 현실적인 제약으로 작용. 이러한 계산 자원 문제로 인해 실제 운영 또는 연구용 해양 자료동화 시스템에서는 EnKF와 같은 동적 오차 추정 기법의 적용이 제한적이며, 대부분의 시스템에서 시간에 따라 고정된 배경장 오차 공분산을 사용하고 있는 실정임. 하지만 이러한 고정 오차 가정은 시공간적으로 변화하는 해양 상태에 따른 오차 성장 및 구조 변화를 충분히 반영하지 못하는 한계를 가짐. 이는 연산 효율성을 유지하면서도 시공간적으로 변화하는 해양 오차 특성을 자료동화에 효과적으로 반영할 수 있는 새로운 기술 개발의 필요성을 시사함.

한편, 최근 기후 예측 및 기후 모형 개발 과정에서 딥러닝 기법이 활발히 활용되고 있으며, 복잡한 비선형 관계 학습을 통해 기존 수치모형의 한계를 보완하는 성과들이 보고되고 있음. 딥러닝 기반 접근법은 대규모 자료 처리와 패턴 인식에 강점을 가지며, 예측 정확도 향상과 계산 효율성 측면에서 큰 잠재력을 보이고 있음. 그러나 이러한 기술 발전에도 불구하고, 딥러닝을 활용한 3차원 해양 자료동화 시스템의 개발 및 적용 사례는 아직 매우 제한적인 상황임. 대부분의 연구는 단순한 2차원 표층 변수 예측이나 통계적 보정에 국한되어 있으며, 해양 내부 구조를 포함하는 3차원 동화 시스템으로의 확장은 아직 초기 단계에 머물러 있음. 따라서 딥러닝의 장점을 활용하여 해양 내부 구조까지 반영할 수 있는 새로운 형태의 3차원 해양 자료동화 프레임워크 개발이 요구되며, 이는 극 지역 해양 및 해빙 예측 성능 향상에 중요한 기여를 할 것으로 기대됨.



제 2 장 국내외 기술개발 현황

1절 국내 기술개발 현황

전 지구 결합 모형 기반의 기후 예측 시스템은 국내 주요 연구 기관을 중심으로 활발히 개발·운영되고 있으며, 이와 더불어 해양 자료동화 기법 또한 지속적으로 발전하고 있음. 한국해양과학기술원(KIOST)은 자체적으로 구축한 해양-대기 결합 모형을 활용하여 해양 초기화 기법을 개발하였으며, 이를 기반으로 엘니뇨(El Niño) 예측 실험을 수행하고 있음. 해당 연구는 열대 태평양 해양 상태의 정확한 초기화가 ENSO 예측 성능 향상에 핵심적인 역할을 함을 입증하며, 국내 기후 예측 연구의 중요한 성과로 평가됨.

APEC 기후센터(APEC Climate Center; APCC)는 자체 기후 예측 시스템(SCoPS, Seasonal Climate Prediction System) 구축을 목표로 넛징(Nudging) 기반의 해양 자료동화 시스템을 개발하여 운영하고 있음. 넛징 기법은 모형 상태를 관측값에 점진적으로 수렴시키는 방식으로, 비교적 계산 비용이 적으면서도 안정적인 초기장 생성을 가능하게 한다는 장점이 있음. 이를 통해 APCC는 다중 모형 기반 계절 예측 시스템 운영을 위한 해양 초기화 기술을 지속적으로 개선하고 있음.

한국형수치예보모델링사업단(KIAPS)은 계절 예측을 위한 전 지구 결합 모형 개발과 함께 해양 자료동화 시스템 구축을 추진하고 있음. 특히 고해상도 해양 모형과의 결합을 목표로 하여, 다양한 관측 자료(위성, 부이, Argo 등)를 효과적으로 활용할 수 있는 동화 기법 개발에 집중하고 있으며, 이를 통해 국내 독자적인 기후 예측 역량 확보를 추진하고 있음.

기상청은 현업 기후 예측 시스템(GloSea6) 운영을 위해 영국 기상청(Met Office)에서 개발된 3차원 변분(3DVAR) 기반 해양 자료동화 시스템을 도입하였음. 해당 시스템은 고해상도 NEMO 해양 모형의 초기화를 위해 활용되고 있으며, 이를 통해 해양 상태의 현실적인 초기화를 수행하고 있음. 이는 국내 기후 예측 현업 시스템에 해양 자료동화를 본격적으로 적용한 대표적인 사례로 평가됨.

부경대학교 김영호 교수 연구팀은 앙상블 최적 내삽법(Ensemble Optimal Interpolation, EnOI)에 기반한 해양 자료동화 기법을 개발하고, 이를 활용하여 장기간 해양 재분석 자료를 생산하였음(Kim et al., 2015). 해당 연구는 비교적 적은 계산 비용으로도 안정적인 동화 성능을 확보할 수 있음을 보이며, 국내 해양 자료동화 연구의 기반을 마련.

전남대학교 함유근 교수 연구팀 역시 EnOI 기반의 해양 자료동화 시스템을 개발하여 고해상도 NEMO 해양 모형에 적용하였음. 이를 통해 연안 및 대양 규모의 해양 변동성을 보다 정밀하게 재현할 수 있는 가능성을 제시하였으며, 국내 고해상도 해양 자료동화 연구의 확장에 기여하고 있음.

최근에는 관측 자료의 공간적 불균일성과 빈번한 결측 문제를 효과적으로 처리하기 위한 딥러닝 기반 기법들이 활발히 개발되고 있음. 특히 partial convolution 기법은

관측이 존재하는 격자점의 정보를 선택적으로 강조하여 반영하고, 결측 영역에서는 주변 격자의 물리적 특성을 활용해 합리적으로 값을 보완하는 구조를 가지는 것이 특징임. 실제로 Ham et al. (2024)에서는 partial convolution 기반 기법을 해양 자료동화 프레임워크에 성공적으로 적용하여, 관측 결측이 심한 조건에서도 신뢰도 높은 초기장 생성이 가능함을 입증하였음. 이는 딥러닝 기반 자료동화 기법이 기존 전통적 방법의 한계를 보완할 수 있는 실질적인 대안임을 보여주며, 향후 해양 초기화 기술로의 활용 가능성을 제시함.

2절 국외 기술개발 현황

국외에서는 자체 해양 자료동화 시스템을 구축·운용하는 기관들이 다수 존재하며, 일본의 JAMSTEC, 미국 NASA/GSFC의 GMAO(Global Modeling and Assimilation Office), 호주 기상청(BoM), 유럽중기예보센터(ECMWF), 미국 국립환경예측센터(NCEP) 등 세계 주요 기관들이 해양 자료동화를 현업 및 연구 목적으로 활용하고 있음. 이들 기관은 해양 재분석 자료 생산, 엘니뇨 실태 분석, 계절 및 중장기 기후 예측 시스템 초기화 등에 해양 자료동화를 적극적으로 적용하고 있음.

NASA/GSFC는 앙상블 최적 내삽법(EnOI) 및 앙상블 칼만 필터(EnKF) 기반의 해양 자료동화 시스템을 구축하여 병행 운용하고 있음. EnOI 기법은 사전 계산된 앙상블 통계를 활용하여 계산 효율성을 확보하는 반면, EnKF는 시간에 따른 오차 발달 과정을 동적으로 추정할 수 있다는 장점을 가짐. 특히 EnKF 시스템은 모형 상태의 시공간적 변화에 따라 오차 공분산을 실시간으로 갱신할 수 있어, 보다 현실적인 초기장 생성이 가능하다는 강점을 가짐. 그러나 높은 계산 비용으로 인해 전 지구 고해상도 모형 적용에는 여전히 제약이 존재하며, 이에 따라 EnOI와 EnKF를 병행 운용하는 전략이 활용되고 있음.

ECMWF와 NCEP 역시 4DVAR 및 하이브리드 변분-앙상블 자료동화 기법을 해양 시스템에 적용하여, 관측 정보를 보다 정교하게 반영하는 방향으로 기술을 고도화하고 있음. 이러한 하이브리드 기법은 변분법의 안정성과 앙상블 기법의 유연성을 결합함으로써, 계산 효율성과 동화 성능을 동시에 확보하고자 하는 시도로 평가됨.

이와 같이 국외 주요 기관들은 고성능 컴퓨팅 자원을 기반으로 다양한 해양 자료동화 기법을 실험·운용하고 있으며, 해양-대기 결합 예측 시스템의 정확도 향상을 위한 핵심 기술로 해양 자료동화를 지속적으로 발전시키고 있음.

제 3 장 연구개발수행 내용 및 결과

1절 자료동화에 활용 가능한 딥러닝 기반 연구

Kadow et al. (2020)에서는 부분 합성곱 신경망(Partial Convolutional Neural Network)을 활용하여 지역적으로만 존재하는 관측 온도 자료로부터 전지구 격자점 온도 복원을 성공적으로 수행. 일반적인 합성곱 과정에서는 결측값이 계산에 포함되는 것과는 다르게, 부분 합성곱 과정은 결측값을 제외하고 계산되는 특징이 있음. 이러한 특징에 기반하여 부분 합성곱 신경망을 활용하여 이미지를 생성할 경우 값이 스무딩 되는 것을 최소화하고 원본에 가까운 이미지를 생산 가능. 해양 자료동화 시 관측은 전지구 그리드 포인트 모두에 값이 존재하는 것이 아니라 지역적으로 존재. 따라서, 간헐적으로 존재하는 관측을 모형의 적분 결과에 적절히 반영시키기 위해서는 부분 합성곱 기법을 활용하는 것이 적합하다는 것을 의미.

적대적 생성 신경망(Generative Adversarial Network)은 이미지 생성 분야에서 활발히 활용·개발되고 있는 기법으로, 생성자와 판별자 간의 경쟁적인 학습을 통해서 이미지 생성 능력을 향상시켜 나가는 방법임. 이러한 특징에 의해 적대적 생성 신경망은 생성된 이미지가 실제 이미지와 통계적으로 거의 구분이 되지 않도록 강제하게 되고, 높은 품질의 이미지를 생성 가능.

pytorch를 기반으로 Chang et al(2019) 에서 Convolution Neural Network를 비디오 인페인팅에 활용(그림 1). 위 논문에 따르면 Gated Convolution은 마스킹 정보가 많은 비디오 인페인팅 작업에서 특히 효과적이며, Partial Convolution보다 우수한 성능을 보임. Gated Convolution 기법은 유효한 픽셀에 주의(attention)를 기울여 연산을 수행하는 방법임. 이 기법은 입력 이미지의 각 픽셀이 얼마나 중요한지를 판단하여, 중요한 픽셀에 더 많은 가중치를 부여한다. 이를 통해 모델이 불필요한 정보에 영향을 덜 받고, 중요한 정보에 더 집중할 수 있음.

이러한 특성을 고려할 때, 관측 밀도가 낮아 공간적 불균형이 있는 염분의 3D 복원 작업에도 gated convolution을 적용하는 것이 효과적일 것으로 예상됨. Gated Convolution은 입력된 데이터에서 0과 1 사이의 값으로 관측의 신뢰도(confidence)를 반영. 관측 데이터의 신뢰도가 높은 지역은 1에 가까운 값을, 신뢰도가 낮은 지역은 0에 가까운 값을 가지며, 이러한 정보를 바탕으로 모델이 학습. 이는 관측이 많고 신뢰도가 높은 데이터를 우선적으로 반영하되, 관측이 부족하거나 불확실성이 높은 지역에서도 주변 정보를 적절히 활용.

또한 해양 염분 데이터는 해양 온도 데이터에 비해 관측 개수가 적고, 관측 밀도가 낮아 공간적 불균형 문제가 발생할 수 있음. Dilation convolution은 커널 내 픽셀 간 간격을 확장하여 더 넓은 수용 영역을 확보할 수 있는 기법으로, 관측 밀도가 낮은 해양 염분 데이터에서도 주변 정보를 효과적으로 반영할 수 있음. 이를 통해 해상도를 유지하면서 넓은 범위의 정보를 처리하여 데이터 손실을 줄이고 계산 효율성을 높일 수 있음.

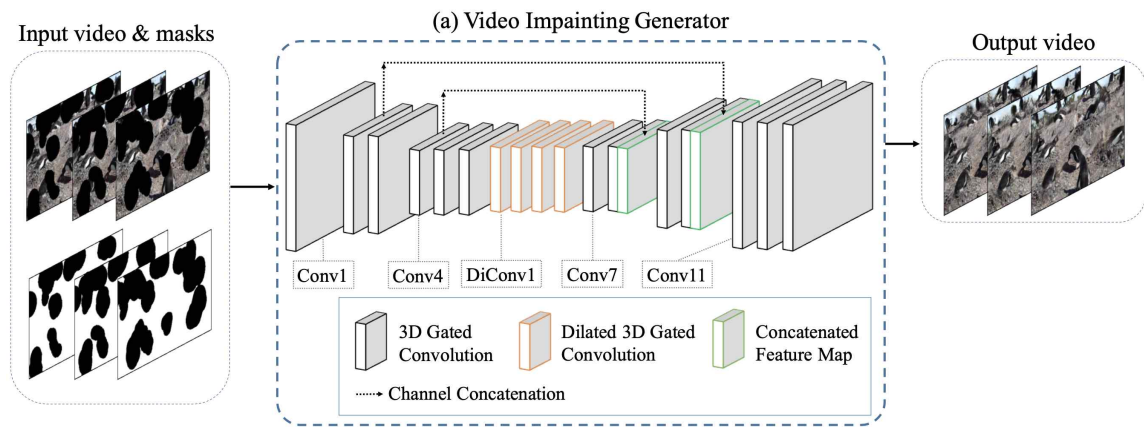


그림 1. Example of video restoration using the gated convolution approach (Chang et al., 2019). The input video and corresponding missing-information masks are provided as inputs, and the restored video is produced through a U-Net - based architecture.



2절 해양 관측 자료 수집 및 데이터베이스화

딥러닝 모형의 학습을 위해서는 충분한 수의 학습 샘플 수를 확보하는 것이 중요. Ham et al. (2019, Nature)에서는 재분석 자료에 더해 기후 모형 장기 적분 데이터를 딥러닝 모형의 학습에 활용. 이를 차용하여, 다양한 자료 동화를 위한 딥러닝 모형의 학습 샘플로 활용하기 위해 Community Earth System Model 2 (CESM2) large ensemble project에서 제공하는 다중 앙상블 자료를 확보..

딥러닝 모형의 학습 과정에서, 실제 관측 네트워크들의 분포를 학습시키고, 관측 네트워크를 기반으로 재분석장을 생산하기 위해 Met Office의 Hadley Centre Integrated Ocean Database (HadIOD)를 확보.

딥러닝 모형의 자료동화 품질을 평가하기 위해 NOAA의 Optimum Interpolation SST (OISST)와 Met Office의 EN4 등 가용한 모든 재분석 자료들을 확보.

3절 전지구 3차원 해양 온도 자료동화 기법 개발

구글에서 개발한 TensorFlow를 기반으로 하여 부분 합성곱 신경망과 적대적 생성 신경망 기법을 결합한 형태의 전지구 3차원 해양 온도 자료동화 시스템을 구축하였다.(DeepDA). DeepDA 모형은 실제 자료동화에 사용될 수 있도록, 재귀적인 과정을 갖도록 구성하였다(그림 2).

현재 시점 t 에서의 3차원 해양 자료동화는, 이전 시점 $t-1$ 에서 생성자 G 에 의해 출력된 분석장(그림 2.b)과 현재 시점 t 의 관측장(그림 2.c)을 생성자 G 의 입력으로 사용하여 현재 시점 t 의 재분석장(그림 2d)을 얻음으로써 수행되었다.

학습은 현재 시점 t 에 대한 라벨(참값)과의 차이로부터 얻을 수 있는 손실함수와 판별자 D 로부터 얻을 수 있는 손실함수를 모두 사용.

생성자 G 의 구조는 image-to-image 생성에 널리 사용되는 U-net을 기반으로 함. 이 구조는 다운 샘플링과 업 샘플링을 통해 입력과 똑같은 크기를 갖는 출력을 생성. 다운 샘플링은 세 번의 합성곱(혹은 부분 합성곱)과 두 번의 최대값 풀링 과정으로 구성했으며, 업 샘플링은 네 번의 전치 합성곱(transposed convolution)으로 구성.

판별자 D 는 PatchGAN을 차용, 다섯 번의 합성곱으로 된 완전 합성곱 구조로 구성. 여기서, 마지막 합성곱 계층은 각 격자점에 대한 확률(참 혹은 거짓)을 출력하는데, 이는 큰 이미지에 대해 효과적인 방법임. 모형의 학습을 위해, 실제 관측 네트워크(HadIOD) 자료에서 관측이 있는 격자점은 1, 없는 격자점은 0으로 마스킹하고 이를 CESM2 LE 자료에 곱하여 실제 관측 네트워크 크가 주어진 상황을 모사함.

여기서, 실제 관측에는 오차가 있고 자료동화 시스템은 이를 고려해야 하므로, 평균이 0이고 표준편차가 0.2인 표준화된 난수값을 더하여 학습을 위한 가상의 관측장으로 사용.

DeepDA 시스템의 성능을 테스트하기 위해, 1974년부터 2014년까지의 CESM2 LE 데이터셋을 활용하여 관측 시스템 시뮬레이션 실험(observing system simulation

experiments, OSSE)을 수행하였음.

이때, 생성자의 초기 background state는 학습 데이터에서 무작위로 선택된 state가 사용됨. 따라서, 처음에는 true state 정보만 제공되는 것임. 재분석장은 1° 의 수평해상도의 3차원 전지구로 표현되며, 해수면으로부터 700m깊이까지 37개의 수직 레벨을 포함함. 자료동화는 5일의 시간 간격으로 수행됨.

그림 2.에 나타낸 전지구 지도들은 OSSE의 첫 번째 주기에서 DeepDA 시스템에 의해 도출된 스냅샷을 의미.

현재 시점 t 에서, 실제 상태(ground truth, 그림 2e)는 태평양에서 엘니뇨 상태, 양의 인도양 쌍극자 패턴, 따뜻한 열대 대서양, 차가운 북태평양을 보임. 관측값(그림 2c)은 설계한 대로, 실제 상태와 유사하고 희박한 공간 분포를 보임. 반면, 생성자 G의 이 보정 효과는 관측 정보가 상대적으로 적어 보정 정도가 표층에 비해 적지만 관측 표층 하부에서도 잘 나타남(그림 3).



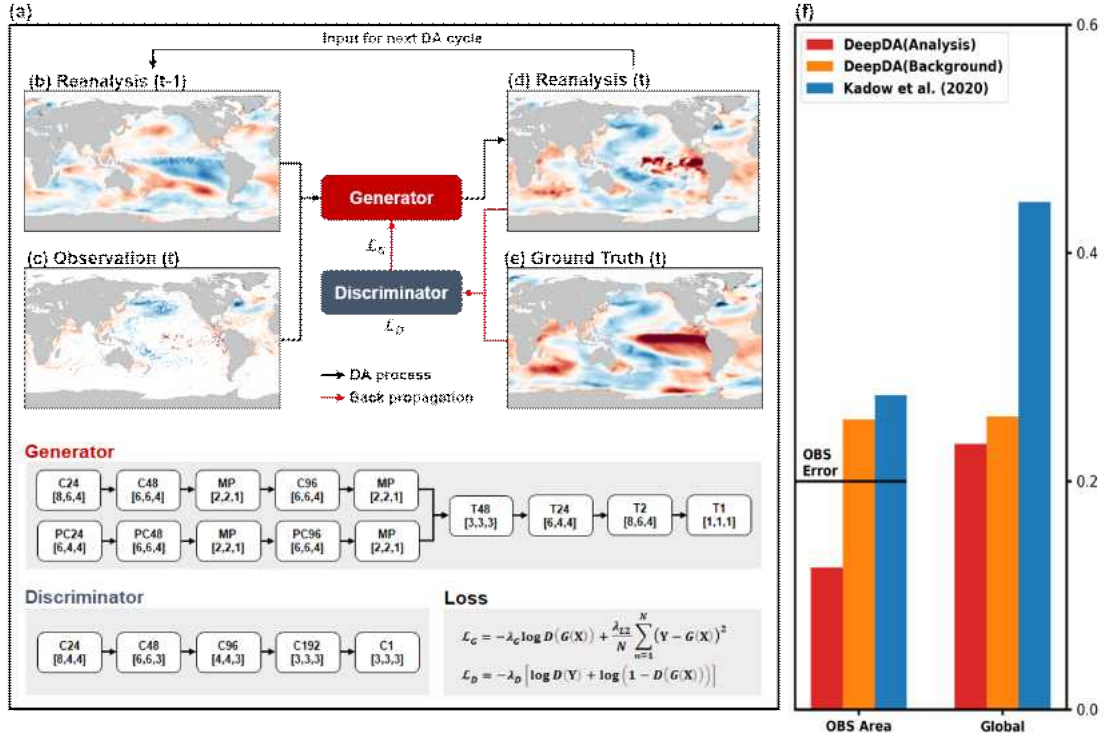


그림 2. Description of the DeepDA system and the observing system simulation experiments(OSSEs). (a) Description of the deep-learning-based DA system. Sea surface temperature anomalies of the (b) background states, (c) observations, (d) analysis states, and (e) ground truth at the first DA cycle in the OSSEs. (f) Observed-locations-averaged (left), and global-averaged (right) RMSEs of the 5-day-averaged 0-700m temperature analysis in the DeepDA (red); the results obtained using background values in the DeepDA (orange) and the partial convolutional system reported in K2022 (blue) are also shown. The prescribed RMSE of the random observational errors (i.e., 0.2) is denoted as a black line in panel (f).

분석장의 해수면 온도는 관측이 존재하는 격자점뿐만 아니라, 인접한 곳에서도 관측과 유사함. 예를 들어, 북대서양 북쪽의 따뜻한 해수면 온도는 관측이 전체를 커버하지

못하고 있지만, 분석장은 실제 상태와 유사하게 나타남. 또한, 많은 관측이 선박 기반 관측에 의해 특정한 직선을 따라 나타나지만, 분석장은 그러한 특징을 보이지 않음.

이는 관측 정보가 인접한 격자점으로 잘 확산되고 있으며, 각 대양에서 단일 지점 관측을 이용한 민감도 실험을 통해서 이러한 특징을 확인했음 (그림 4). 더 중요한 것은 분석 증분의 공간적 규모와 세부 패턴이 시간과 지역에 따라 다르다는 것인데, 이는 DeepDA의 상태 의존적 특징을 잘 보여줌.

시간에 따라 자료 동화 사이클을 반복함으로써, 분석장은 축적된 관측 정보에 의해 점차 실제 상태에 맞게 조정됨. 분석장과 테스트 기간(즉, 3개월의 스핀업 기간을 제외한 1974년 4월부터 2014년 12월까지의 기간) 동안 얻은 DeepDA 기간 3차원 온도 분석장의 RMSE는 관측값이나 배경 상태보 다 작음(그림 2f).

관측이 존재하는 지역에서 분석 값의 RMSE는 0.12로 감소하여 배경(0.25) 또한 관측치(0.2)보다 작음(그림 2f 왼쪽 막대들). 이는 DeepDA가 두가지 가능한 선택적 확률(즉, 배경 및 관측 분포)을 결합하여 실제 상태에 대한 최적의 추정치를 적절히 찾아내고 있음을 의미함.

분석장의 전지구 평균 RMSE도 배경장의 평균으로 체계적으로 감소하여 자료동화 과정을 통한 체계적인 오차 감소를 보여줌(그림 2f 오른쪽 막대들). Kadow et al. (2020) 방법을 사용할 경우 DeepDA의 약 두 배에 달하는 RMSE를 보임.

이는 기존 딥러닝 기반 자료동화 시스템과 비교했을 때, DeepDA가 월등한 성능을 보임을 명확히 보여줌.

1980부터 2020년까지 41년에 걸친 실제 현장 해양 관측자료를 이용, 실제 현장 관측을 활용하여 해양 재분석 자료를 생성하였음. 첫해인 1980년의 결과는 스핀업 기간으로 간주하여 분석에서 제외하였음. 이 경우, 현장 프로파일 관측과 함께, 위성 기반 격자형 SST 자료(즉, OISST V2)도 함께 동화됨.

재분석장을 섭동시키고, 드롭아웃 기법을 활용하여 초기조건과 모형 산출의 불확실성을 각각 추정하여 100개의 앙상블 멤버를 생성하였음.

1981-2020년 동안 현장 HadIOD 관측과 DeepDA 배경 상태 간 RMSE는 해수면과 수심 100m 부근에서 최대 1°C 정도이며(그림 5a), 이는 다른 해양 재분석 결과와 비슷한 정도임.

동화 후의 RMSE는 명확히 감소하며, 이는 DeepDA 시스템이 주어진 현장 관측을 사용하여 성공적으로 동화한다는 것을 보여줌.

HadIOD 관측 위치에 대한 앙상블 확산은 0.8°C로 수심 100m에서 최대가 되며, RMSE 정도로 명확하게 감소하지는 않지만, 깊이가 깊어질수록 감소함.

Jan 01 - Jan 05, 1974 Temp anomaly

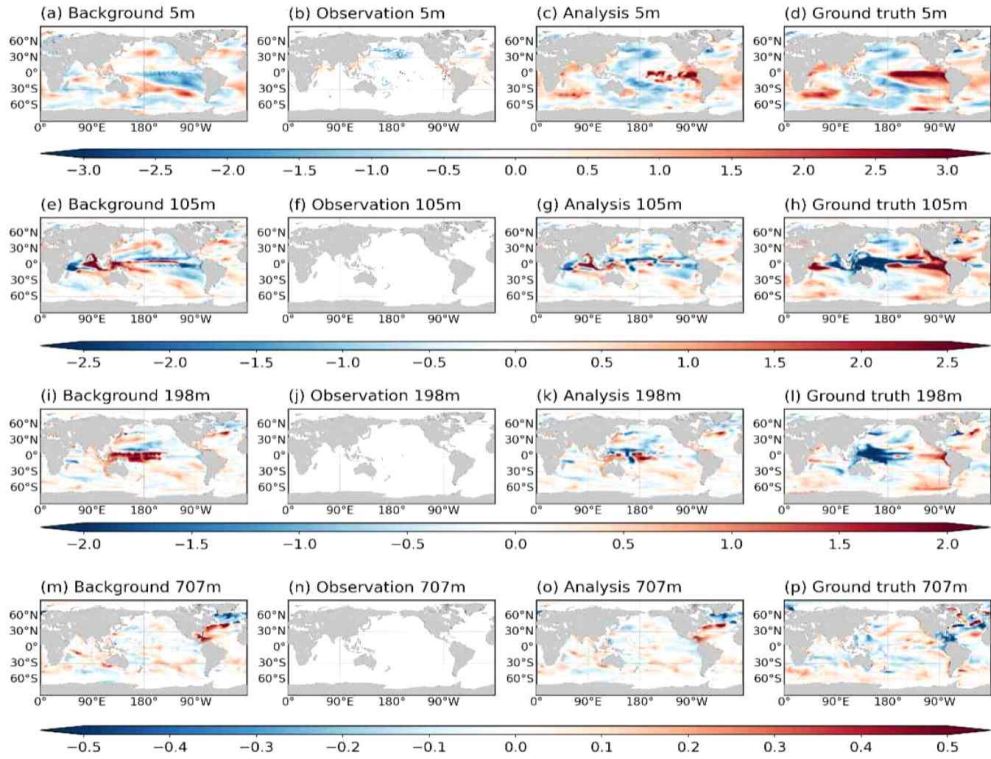


그림 3. Initial adjustment of DeepDA-produced subsurface temperature anomalies in the OSSEs. The temperature anomalies of the (a) ground truth, (b) observations, (c) analysis, and (d) background states at 5 m at first data assimilation cycle in the Observing System Simulation Experiment (OSSEs). (e)-(h) same as (a)-(d), but for 105 m. (i)-(l) same as (a)-(d), but for 198 m. (m)-(p) same as (a)-(d), but for 707m.

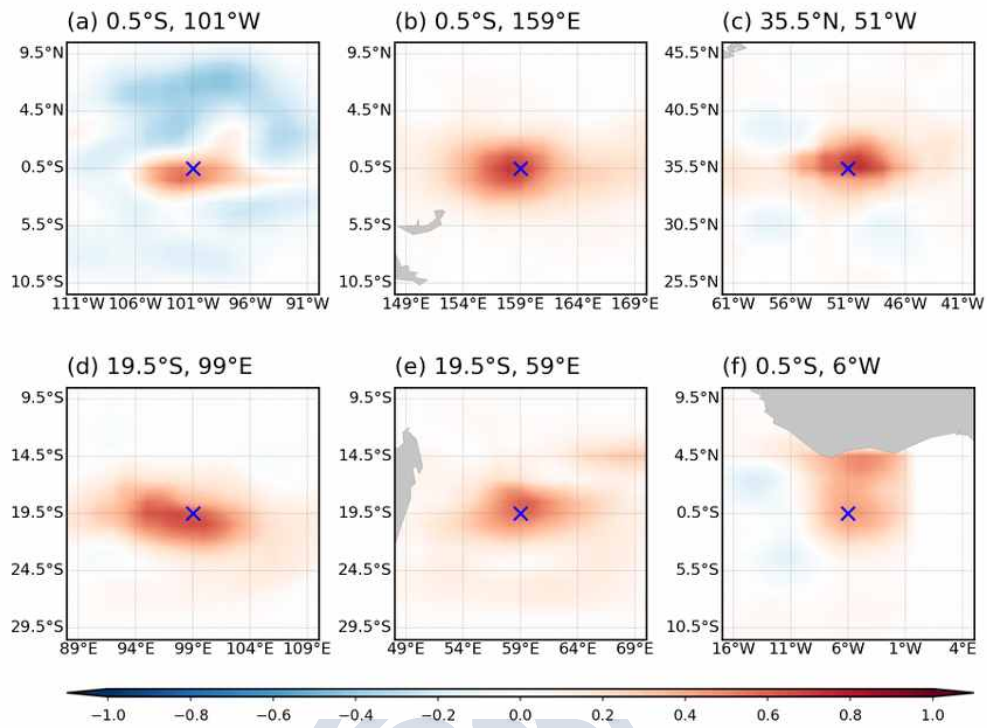


그림 4. Horizontal distribution of the analysis increment in the OSSEs using the DeepDA.

현장 ARGO/TAO 온도 프로파일에서 월평균 DeepDA 재분석의 RMSE를 비교하였음(그림 5b). DeepDA의 전지구 평균 온도 RMSE는 모든 수심에서 1°C 미만이며, 가장 작은 수준의 RMSE를 보이는 편이고, 해수면에서 수심 500m 깊이까지 다른 재분석

결과들의 범주 내에 있음.

래퍼런스 위성 관측(즉, ERSST V5)의 전지구 평균 해수면 온도는 온난화 추세에 따라 19°C에서 20°C사이를 보임(그림 5c). DeepDA에서도 이러한 전지구 평균 해수면 온도의 진폭과 변동이 전반적으로 잘 복원되어, 관측이 재분석장에 잘 반영되고 있음을 보임.

따라서, DeepDA의 전지구 평균 해수면 온도 시계열은 상관관계와 변동성 측면에서 가장 높은 일관성을 보이는 그룹에 속함(그림 5d). 여기서, 모든 격자점의 해수면 온도는 위성 기반 관측이 모두 커버하므로 DeepDA의 전지구 평균 해수면 온도의 양상을 확산은 무시될 수 있음.

관측의 지구온난화 추세는 학습 데이터셋과 다른 진폭을 보임에도 불구하고, DeepDA 시스템이 관측 지구온난화 추세를 성공적으로 재현한 것은 주목할만함. 이는 불규칙한 위치의 값을 데이터에 투영하는 방법을 학습하는 DeepDA의 능력을 보여주는 것으로, 적절한 수의 실제 관측이 제공되는 한, DeepDA는 구조적 오차의 영향을 덜 받음을 의미.

DeepDA에서 전지구 평균 해양 열용량은 수심 500m까지의 밀도 가중 수직 적분된 온도(OHC500으로 표현)으로 정의되며, 1990년 이후 뚜렷한 증가를 보이는 반면, 1980년대에는 온난화 추세가 명확하지 않음(그림 5e). DeepDA에 의해 도출된 OHC500은 EN4.2.2에 비해 1990년대에는 과대평가, 2010년 대에는 과소평가하였음에도 불구하고, 래퍼런스 데이터(즉 EN4.2.2)와 어느 정도 일치함.

재분석장의 공간 분포를 조사하기 위해, DeepDA기반 연평균 해수면 온도와 T300(즉, 수심 300m까지의 수직 평균 온도)을 각각 ERSST V5 및 ARGO 프로파일 기반

관 분석장과 비교하였음(그림 6).

연평균 해수면 온도는 열대 서태평양, 인도양, 북대서양에서 약한 음의 편향을 보임(그림 6a). 이는 학습 데이터셋에서 일관되게 나타나며 그 크기는 상당히 감소했음.

DeepDA와 ERSST V5의 연평균 해수면 온도 패턴 상관관계는 최대 0.99에 달하며, 이는 분석에 사용된 모든 재분석 산출물들 중 가장 큰 값 중 하나임(그림 6b).

연평균 T300은 북태평양과 북대서양에서 음의 편향을 보였으며, 최대 1°C의 크기까지 나타났음(그림 6c). 적도 반류 지역, 일본 동쪽 연안, 남빙양에서는 양의 편향을 보였음.

DeepDA 재분석에서 편향의 공간적 분포는 CESM을 사용한 기존 자료동화 시스템에서와 매우 유사함. DeepDA과 EN4.2.2의 T300 간의 패턴 상관관계는 0.9로, 다른 재분석 결과와 유사함(그림 6). T300의 전지구 평균된 표준편차 또한 잘 모의되었음.

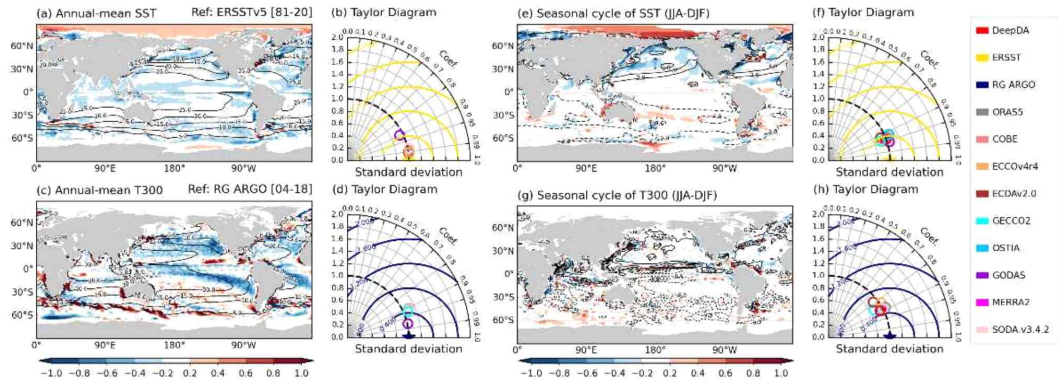


그림 6. Climatological fields of the DeepDA reanalysis. Annual-mean (a) SST, and (c) vertically-averaged temperature from 0 to 300 m (T300) of the DeepDA (contour) and its bias (shading) from ERSST V5, and Roemmich-Gilson Argo climatology during 1994–2018, respectively. Taylor diagram of the annual-mean (b) SST, and (d) T300 climatology during 1994–2018 for DeepDA and various oceanic reanalysis products. Difference in the climatological (e) SST, and (g) T300 between JJA and DJF seasons during 1994–2018, and the corresponding Taylor diagram for the seasonal climatology difference in (f) SST, and (h) T300 for the DeepDA and various other oceanic reanalysis products. Note that, due to the relative short integration period, a period for calculating the climatology in GECCO2, ECCO v4r4, and SODA v3.4.2 is 1994–2016, and 1994–2017, and 1994–2017, respectively.

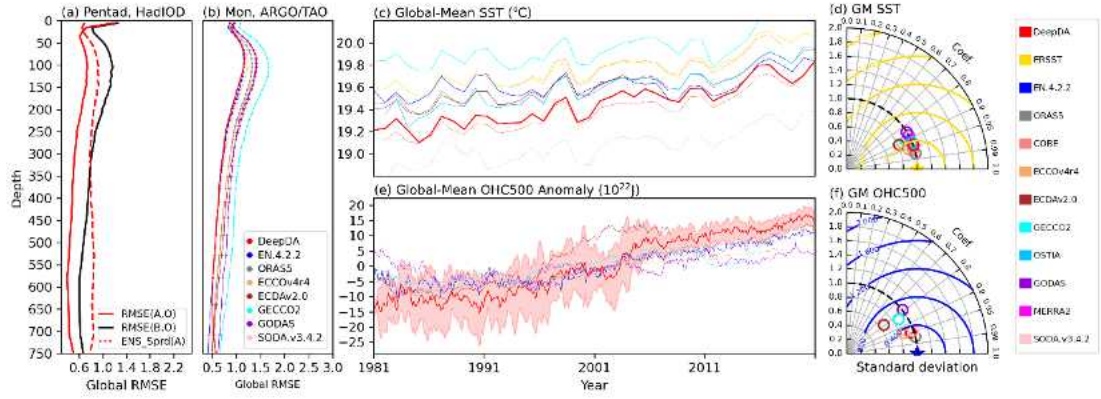


그림 5. Comparison of the DeepDA reanalysis to other oceanic reanalysis datasets. (a) Vertical profiles of 5-day-averaged temperature RMSE (in unit of $^{\circ}\text{C}$) between in situ HadIOD observation and DeepDA background state (black), or DeepDA reanalysis state (red) during 1981–2020 over the global domain (89.5°S – 89.5°N , 0° – 360°E). Dashed red line denotes the ensemble spread in DeepDA reanalysis using 100 ensemble members. (b) Vertical profiles of the monthly-averaged temperature RMSE (in unit of $^{\circ}\text{C}$) between ARGO/TAO observations and the DeepDA or other oceanic reanalysis products over the global domain. The RMSE in panels (a) and (b) are calculated over the in situ observed locations. (c) Monthly time-series of the global-mean SST from 1981 to 2020 and (d) Taylor diagram of the global-mean monthly SST time-series in the DeepDA and various oceanic reanalysis products. (e) Monthly time-series of global-mean oceanic heat content which is defined as density-weighted vertical averages of the temperature from surface to 500 m (OHC500) and (f) Taylor diagram of the global-mean monthly OHC500 time-series in the DeepDA and various oceanic reanalysis products. The light red area in panel (e) denotes a two-tailed 95 % confidence interval of the OHC500 defined as the interval between 2nd largest and smallest values among 100 ensemble members in DeepDA reanalysis.

연평균장 외에도, DeepDA는 SST와 T300의 계절 주기에 대해서도 잘 모의하였음(그림 5e-h). 연평균장과 마찬가지로, 여름철(6-8월 평균)과 겨울철(12-2월 평균)의 차이로 정의되는 해수면 온도의 계절 주기는 래퍼런스 데이터와 거의 동일함(그림 6e, f).

대부분의 지역에서, 편향의 정도가 0.2°C 미만이므로, T300의 계절 주기 또한 DeepDA 시스템에 의해 성공적으로 복원되었다고 볼 수 있음(그림 6g). 이에 대한 패턴 상관관계는 다른 재분석 산출물과 비슷한 수준인 0.9에 달함(그림 6).

각 대양에 대한 첫 번째 경험적 직교 함수 EOF의 고유 벡터(그림 7b, c) 또한 래퍼런스 데이터와 유사한 공간 분포를 보이는 것으로 보아, DeepDA의 해수면 온도 변동성은 현실적으로 모의되고 있음을 확인할 수 있음.

DeepDA와 EN4.2.2의 T300 아노말리 간 상관관계 또한 계산하였음(그림 7d).

DeepDA의 T300 아노말리의 변동은 일반적으로 열대 및 북태평양에서 상대적으로 높은 상관계수를 보이는 반면, 열대 대서양과 남빙양에서는 상관계수가 감소.

남태평양과 대서양에서 구조적 오차가 나타나지만, DeepDA에서 T300의 첫 번째 EOF 고유 벡터는 전반적으로 래퍼런스 데이터와 유사함 (그림 7e, f).

관측 경년 변동성에 대한 모의 품질을 검증하기 위해, DeepDA 재분석과 ERSST V5 간 월별 해수면 온도 아노말리의 시간적 상관관계를 계산하였음(그림 7a).

시간적 상관계수는 현장 관측의 횟수가 상대적으로 적은 남빙양을 제외한 대부분의 대양에서 0.9 이상으로 나타났으며, 이 특징은 다른 재분석 산출물에서도 유사하게 나타남(그림 8).

각 대양에 대한 첫 번째 경험적 직교 함수 EOF의 고유 벡터(그림 7b, c) 또한 래퍼런스 데이터와 유사한 공간 분포를 보이는 것으로 보아, DeepDA의 해수면 온도 변동성은 현실적으로 모의되고 있음을 확인할 수 있음.

DeepDA와 EN4.2.2의 T300 아노말리 간 상관관계 또한 계산하였음(그림 8d).

DeepDA의 T300 아노말리의 변동은 일반적으로 열대 및 북태평양에서 상대적으로 높은 상관계수를 보이는 반면, 열대 대서양과 남빙양에서는 상관계수가 감소(그림 9).

남태평양과 대서양에서 구조적 오차가 나타나지만, DeepDA에서 T300의 첫 번째 EOF 고유 벡터는 전반적으로 래퍼런스 데이터와 유사함 (그림 7e, f).

정리하면, DeepDA 기반 3차원 해양 온도 재분석의 품질은 전지구 해양에 대한 기후값과 계절 주기, 경년 변동성, 지구온난화 추세를 사실적으로 모의함으로써, 다른 최첨단 재분석 산출물과 비교해도 손색이 없음을 확인하였음.

부분 합성곱 기법을 활용함으로써, 불균질하고 산발적인 관측 정보를 격자점에 성공적으로 반영시킬 수 있었음. 이전 시점에서의 관측 정보를 축적하는 DeepDA 시스템의 재귀적인 접근 방식은 기존 선행 연구(Kadow et al., 2020)와 비교하여 더욱 향상된 성능을 제공함.

DeepDA는 최첨단에 이른 전지구 재분석장을 생산할 뿐만 아니라, 기존 역학 예측의 초기 조건을 생성할 때 발생하는 기술적 단점을 완화하여 역학 예측의 새로운 시대를 열 수 있음.

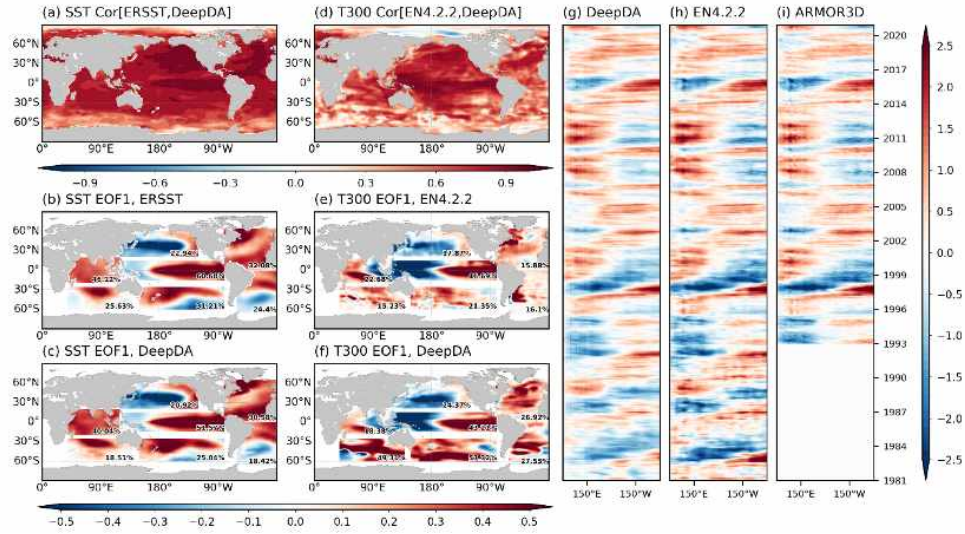


그림 7. Interannual variations simulated in the DeepDA. (a) Temporal correlation coefficients of the monthly SST anomalies between the DeepDA and the ERSST V5 during 1981–2020. The first Empirical Orthogonal Function (EOF) eigen-vector of the monthly SST anomalies in (b) ERSST V5, and (c) DeepDA. Note that the EOF is performed for each ocean basins, then merged into a global map, and the percentage denoted in each ocean basin indicates an explained variance of the first EOF. (d)–(f) Same as (a)–(c), but for T300 anomalies with EN4.2.2 as a reference dataset. Hovmuller diagram of the 5-day-averaged T300 anomalies at the equator in (g) DeepDA, (h) EN4.2.2, and (i) ARMOR3D from 1981 to 2020.

Correlation with ERSSTv5

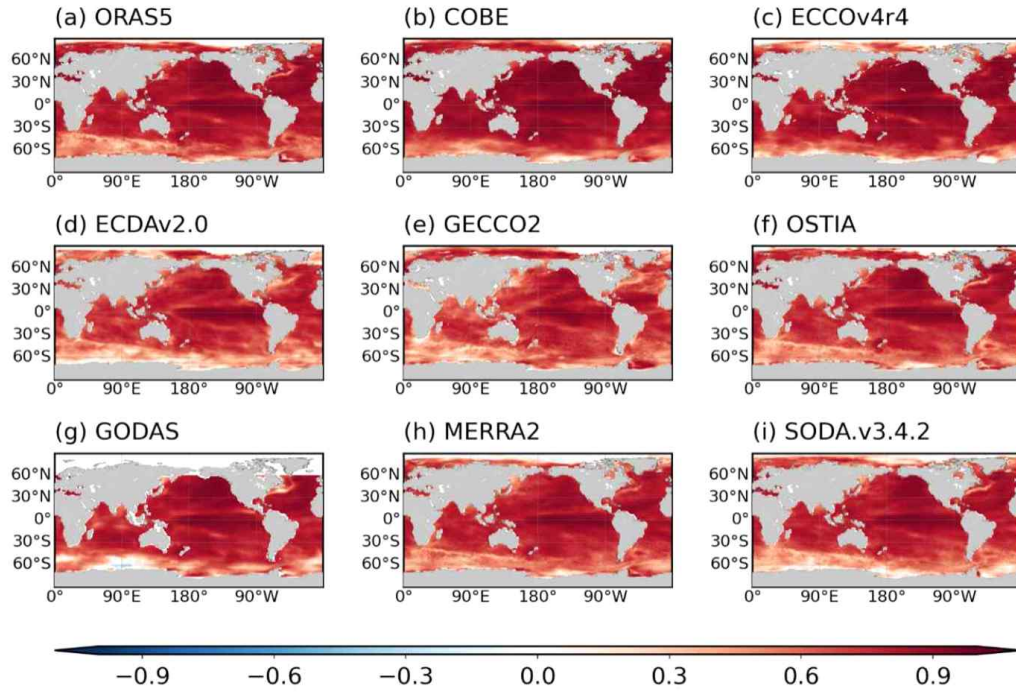


그림 8. Correlation skill of SST anomalies in various oceanic reanalysis products. Temporal anomaly correlation coefficients of the monthly SST anomalies between the ERSST V5 and (a) ORAS5, (b) COBE, (c) ECCO v4r4, (d) ECDA v2, (e) GECCO2, (f) OSTIA, (g) GODAS, (h) MERRA2, or (i) SODA v3.4.2 during 1981-2020.

Correlation with EN4.2.2

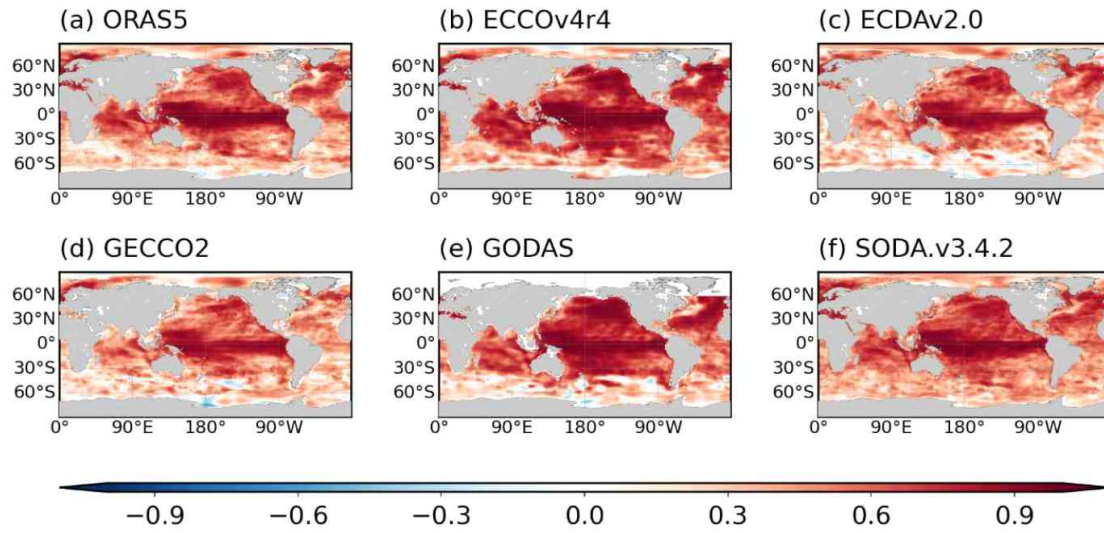


그림 9. Correlation skill of T300 anomalies in various oceanic reanalysis products. Temporal anomaly correlation coefficients of the monthly T300 anomalies between the EN4.2.2 and (a) ORAS5, (b) ECCO v4r4, (c) ECDA v2, (d) GECCO2, (e) GODAS, or (f) SODA v3.4.2 during 1981–2020.



4절 전지구 3차원 해양 염분 자료동화 기법 개발

Gated Convolution 기법은 유효한 픽셀에 주의(attention)를 기울여 연산을 수행하는 방법임. 이 기법은 입력 이미지의 각 픽셀이 얼마나 중요한지를 판단하여, 중요한 픽셀에 더 많은 가중치를 부여함. 이를 통해 모델이 불필요한 정보에 영향을 덜 받고, 중요한 정보에 더 집중할 수 있음. 이러한 특성을 고려할 때, 관측 밀도가 낮아 공간적 불균형이 있는 염분의 3D 복원 작업에도 gated convolution을 적용하는 것이 효과적일 것으로 예상됨.

Gated Convolution은 입력된 데이터에서 0과 1 사이의 값으로 관측의 신뢰도(confidence)를 반영함. 관측 데이터의 신뢰도가 높은 지역은 1에 가까운 값을, 신뢰도가 낮은 지역은 0에 가까운 값을 가지며, 이러한 정보를 바탕으로 모델이 학습됨. 이는 관측이 많고 신뢰도가 높은 데이터를 우선적으로 반영하되, 관측이 부족하거나 불확실성이 높은 지역에서도 주변 정보를 적절히 활용할 수 있도록 함.

또한 해양 염분 데이터는 해양 온도 데이터에 비해 관측 개수가 적고, 관측 밀도가 낮아 공간적 불균형 문제가 발생하므로, 이를 보완하기 위해 Dilation convolution을 적용.

Dilation convolution은 커널 내 픽셀 간 간격을 확장하여 더 넓은 수용 영역을 확보할 수 있는 기법으로, 관측 밀도가 낮은 해양 염분 데이터에서도 주변 정보를 효과적으로 반영할 수 있음. 이를 통해 해상도를 유지하면서 넓은 범위의 정보를 처리하여 데이터 손실을 줄이고 계산 효율성을 높임

Gated Convolution과 Dilation Convolution을 기반으로 형태의 전지구 3차원 해양 염분 자료동화 시스템을 구축하였으며, 이를 GatedDA_sal라고 칭함.

재귀적인 구조를 갖는 GatedDA_Sal 시스템의 개발을 위해, GATED Convolution과 PARTIAL Convolution 기법을 비교하여 해양 염분의 재구성 성능을 평가하였음. 실험 결과(그림 11), GATED 기법이 PARTIAL 기법에 비해 RMSE가 더 낮고 복원 정확도가 높아, GATED 기법을 기반으로 한 GatedDA_Sal 시스템을 구축하게 되었음.(그림10)

Community Earth System Model 2 (CESM2) large ensemble 프로젝트에서 제공한 전지구 5일 평균 해양 염분 데이터를 기반으로 딥러닝 모델을 훈련. 1892년부터 1932년까지의 자료를 모델 학습에 활용하였고, 1933년부터 1973년까지의 기간을 검증에 사용함. 이 자료를 통해 재분석장을 생성하였으며, 공간 해상도는 1도 간격, 시간 해상도는 5일 간격으로 설정.

모델에는 3차원 합성곱 신경망을 도입하여, 수평과 수직 방향의 정보를 동시에 학습할 수 있는 필터를 적용.

실제 관측 데이터에는 불확실성이 존재하므로, 자료동화 시스템에서는 이를 고려해야 함. 이를 반영하기 위해, 평균이 0이고 표준편차가 0.005인 정규 분포에서 생성된 랜덤 에러 값을 더해, 학습용 가상 관측장을 구성함.

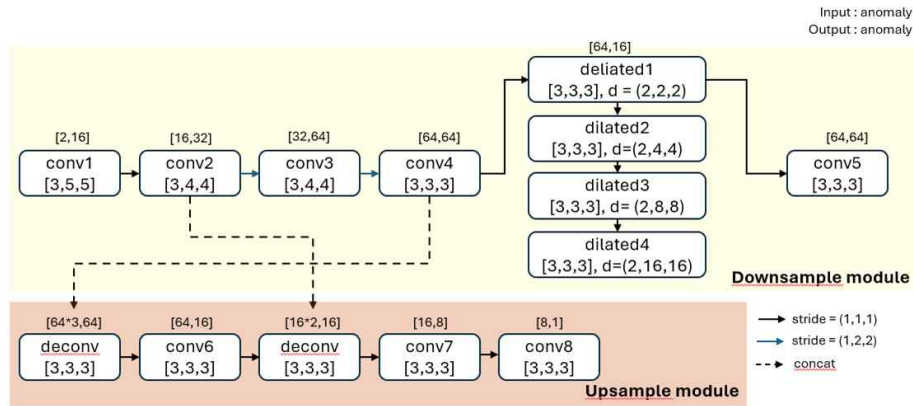


그림 10. Architecture of the downsample and upsample modules used for processing masked salinity data. The downsample module reduces the dimensionality through a series of convolutional and dilated convolution layers. The upsample module then restores the data back to its original resolution using deconvolution and convolution layers. Solid arrows represent different strides used in the convolution layers, while dashed arrows indicate concatenation operations. The input and output of the entire process are salinity anomaly fields.



손실 함수로는 평균 제곱 오차(MSE, Mean Squared Error) loss를 사용하여 모델의 예측값과 실제값 사이의 오차를 최소화하도록 학습함.

깊이가 105m와 305m로 갈수록 관측 데이터의 부족으로 인해 Partial 모델의 성능이 저하됨을 확인할 수 있음(그림 11). 그림 11 우측 열의 상관계수 차이(Gated - Partial)를 살펴보면, 깊이가 깊어질수록 GATED 모델이 Partial 모델보다 훨씬 더 좋은 성능을 보임. 특히 305m 깊이에서는 GATED 모델이 전반적으로 더 나은 상관계수를 기록했으며, 이는 관측 데이터가 부족한 환경에서 GATED 모델의 데이터 재구성 능력이 더욱 탁월함을 시사함. 관측 데이터가 상대적으로 풍부한 얕은 층 뿐만 아니라, 관측 데이터가 적은 깊은 층에서도 GATED 모델이 더 안정적이고 일관된 성능을 보임.

이는 GATED 모델이 적은 관측 데이터 환경에서도 효과적으로 데이터를 재구성할 수 있음을 의미함.

그림12 은 해양 염분 자료 동화 모델에서 마스크 패턴이 GATED와 PARTIAL 모델의 RMSE 분포에 미치는 영향을 비교한 것임. 그림 12a는 GATED 모델과 PARTIAL 모델 간의 RMSE 차이를 나타냄. 파란색 영역은 GATED 모델이 PARTIAL 모델에 비해 RMSE가 더 낮은 지역을, 빨간색 영역은 PARTIAL 모델이 더 나은 성능을 보이는 지역을 나타냄. 특히, 관측 지점이 없는 지역(즉, 관측 밀도가 낮거나 전혀 없는 지점)에서 GATED 모델이 PARTIAL 모델보다 더 낮은 RMSE를 보여 더 나은 성능을 보이고 있음을 알 수 있음. 이 결과는 GATED 모델이 관측이 없을 때 더 효과적으로 정보 격차를 채워넣고 있음을 의미함. 그림 12b 는 그림12a의 관측위치를 반전한 결과임.

그림 12 그래프는 GATED와 PARTIAL 모델의 RMSE 값을 경도(120°E에서 240°E) 별로 평균하여 비교한 것임. 관측 밀도가 낮은 지역에서 GATED 모델의 RMSE가 PARTIAL 모델보다 더 낮게 나타나는 구간이 관찰됨. 특히 경도 180°E 근처의 구간에서 두 모델 간 RMSE 차이가 두드러짐. 관측 위치가 반전된 후에도(그림 12d), 관측 밀도가 매우 낮은 지역에서 GATED 모델이 PARTIAL 모델에 비해 더 나은 성능을 보임을 확인할 수 있음.

Mask가 반전된 상태에서도 GATED 모델이 관측이 적거나 없는 지역에서 더 낮은 RMSE를 유지하는 이유는, Gated Convolution이 단순히 지형적 위치에 의존하지 않고, 관측 신뢰도를 기반으로 정보를 처리하기 때문임.

이 사실은, GATED 모델의 성능 차이가 특정 지역의 지형적 요인이 아닌, 모델의 내부 메커니즘에 의해 주도된다는 것을 의미함. 즉, gating value는 관측 신뢰도에 따라 각 픽셀의 중요도를 조절하며, 이로 인해 관측 밀도가 낮은 지역에서도 일관된 성능 향상이 가능하다는 점을 보여줌.

결과적으로, GATED 모델이 관측 신뢰도와 정보 처리 방식에 의해 성능이 결정된다는 것을 이 실험은 뒷받침함.

OSSE Correlation: GATED Model vs CESM2 Test Set (1974-2014) by Depth

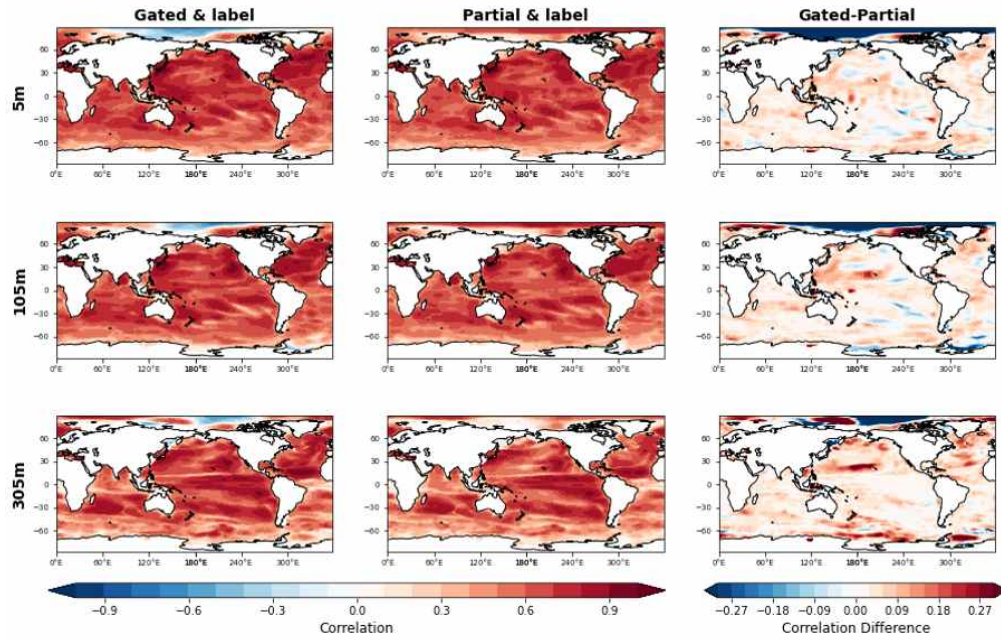


그림 11. OSSE(Observing System Simulation Experiment) correlation. Left (Gated Correlation): The left column shows the correlation values between the Gated model and the reference data at three different depths (5m, 105m, 305m). The color scale represents the correlation values. Middle (Partial Correlation): The middle column displays the correlation values for the Partial model at the same three depths (5m, 105m, 305m). Right (Difference: Gated - Partial Correlation): The right column illustrates the difference in correlation between the Gated and Partial models. Red areas indicate that the Gated model has a higher correlation (better performance), while blue areas show that the Partial model performs better (higher correlation).

Impact of Mask Pattern on RMSE: GATED vs. PARTIAL Models

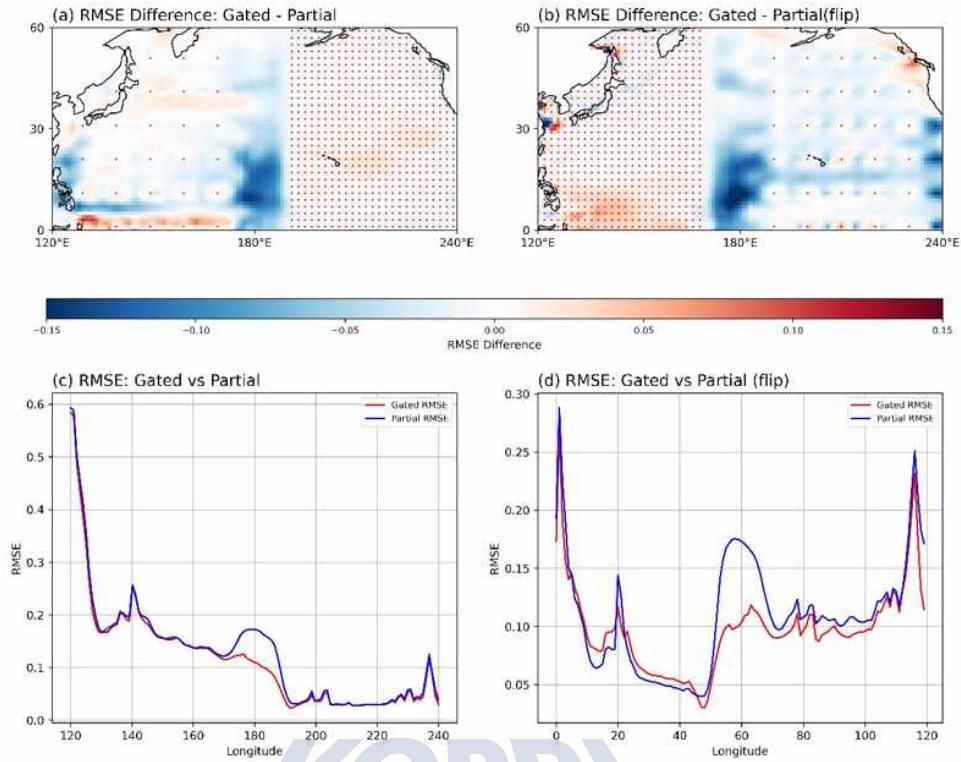


그림 12. (a) RMSE Difference: Gated - Partial This subplot illustrates the spatial distribution of RMSE differences between the GATED and PARTIAL models. The color shading represents the difference in RMSE value. The red dots show the locations where observations are available. (b) RMSE Difference: Gated - Partial(flip) This subplot shows the RMSE differences between the GATED and PARTIAL(flip) models. In this case, the PARTIAL(flip) model uses the inverse of the observation mask from (a), meaning the locations with observational data are flipped. (c) Longitude-Averaged RMSE: Gated vs Partial This plot shows the longitude-averaged RMSE values for the GATED model (red line) and the PARTIAL model (blue line). The x-axis represents the range of longitudes, and the y-axis shows the RMSE values averaged across longitudes. These values correspond to the regions shown in subplot (a). (d) Longitude-Averaged RMSE: Gated vs Partial(flip) This plot is similar to (c) but corresponds to the flipped observation regions shown in (b).

DeepDA vs EN4.2.2: NINO3.4 and North Atlantic Salinity Comparison

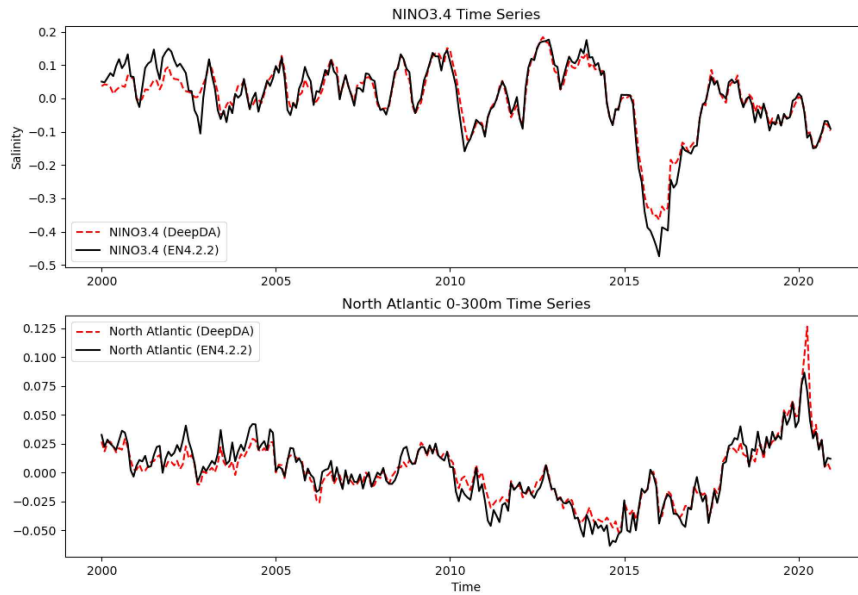


그림 13. Time series comparison of salinity between GatedDA_sal and EN4.2.2 datasets. The top panel shows the NINO3.4 region's(5m) salinity, and the bottom panel represents the North Atlantic (0-300m mean) salinity. The dashed red line corresponds to the GatedDA_sal dataset, and the solid black line corresponds to the EN4.2.2 dataset.



그림13은 NINO3.4 지역(그림 13 상단 패널)과 북대서양 0-300m 지역(그림 23 하단 패널)에서 GatedDA_Sal와 EN4.2.2 데이터셋의 염분 시계열을 비교한 것임. GatedDA_Sal 시스템은 실제 관측 자료인 HADIOD 데이터를 입력으로 사용하여 재분석장을 생성하였으며, 이를 EN4.2.2와 비교하였음. 빨간색 점선은 GatedDA_Sal(Data Assimilation, 자료 동화) 결과를, 검은색 실선은 EN4.2.2 데이터셋(관측 재분석 자료)을 나타냄.

NINO3.4 지역은 엘니뇨(El Nino)를 모니터링하는데 중요한 역할을 함. 엘니뇨는 이 지역의 해수면 온도가 상승하는 현상으로, 전 세계적인 기후 변동에 큰 영향을 미침.

반면, 북대서양은 AMOC(대서양 자오선 반전 순환, Atlantic Meridional Overturning Circulation)과 밀접하게 관련된 지역으로, 북대서양의 염분 및 온도 변화는 AMOC의 강도와 변동성을 나타내는 중요한 지표임.

AMOC는 전 지구적 해양 순환 시스템에서 핵심적인 역할을 하며, 기후 변화와 연계되어 있음. 따라서, NINO3.4와 북대서양의 염분 시계열을 분석하는 것은 엘니뇨와 AMOC 변동성에 대한 이해를 높이는 중요한 과정임.

두 데이터셋을 비교해 보면, NINO3.4 지역에서 나타난 큰 염분 감소나 북대서양 지역에서의 변동과 같은 주요 이벤트에서 GatedDA_Sal가 EN4.2.2의 염분 변화 패턴과 크기를 잘 재현하고 있음을 확인할 수 있음

전반적으로 GatedDA_Sal는 EN4.2.2 데이터셋과 매우 유사한 경향을 보이며, 이는 GatedDA_Sal의 자료 동화 과정이 성공적이었음을 의미함. 특히 두 지역에서의 시계열이 잘 일치한다는 것은 GatedDA_Sal 시스템이 관측 데이터를 효과적으로 반영하고, 해양 상태를 잘 재구성했음을 보여줌.

이러한 결과는 GatedDA_Sal 재분석장이 향후 과학적 연구 및 다양한 응용에 있어 신뢰할 수 있는 자료임을 뒷받침해줌.

따라서, GatedDA_Sal는 EN4.2.2와 비교했을 때 성공적으로 염분장을 재현한 자료 동화 시스템으로 평가될 수 있음.

동화 후의 RMSE(O-A)가 배경 상태의 RMSE(O-B)보다 감소한 것은 본 연구에서 개발한 딥러닝 기반 3차원 해양 염분 자료 동화 시스템이 주어진 현장 관측 데이터를 효과적으로 동화했음을 나타냄.(그림 14)

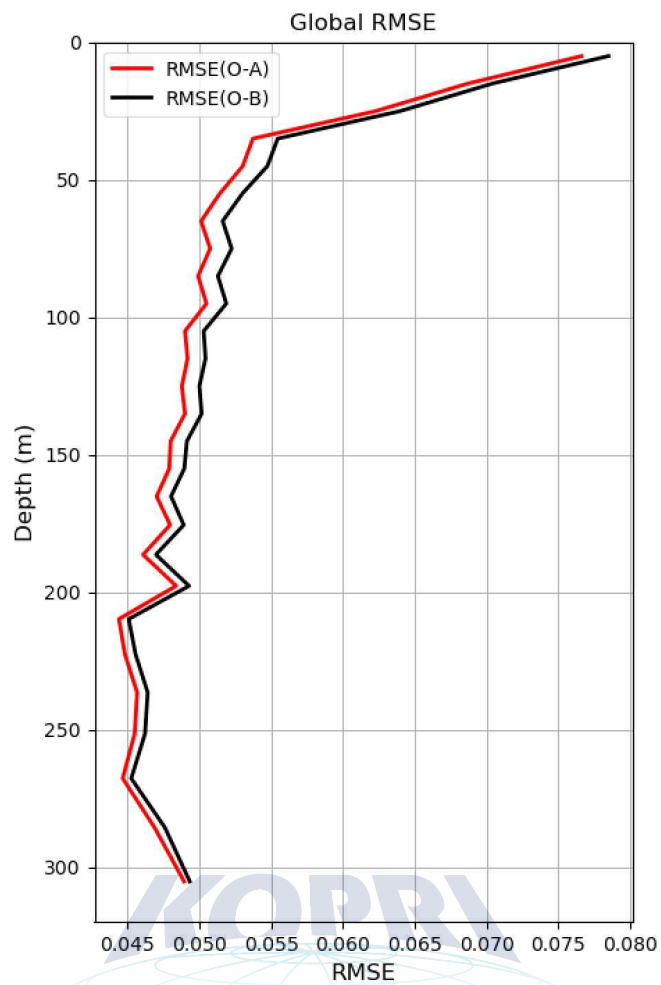


그림 14. Global, depth - dependent O - B and O - A statistics of observations (HadIOD) used in deep-learning-based three-dimensional ocean salinity data assimilation (1981 - 2020).

5절 전지구 3차원 다변량 자료동화 기법 개발

연구에서 개발한 자료동화 모델은 해양 온도(Temperature)와 염분(Salinity)을 동시에 고려하는 다변량 구조로 설계됨(그림 15).

모델 아키텍처는 Gated Convolution과 Dilation Convolution을 결합하여 온도와 염분 간의 상호작용과 공간적 연속성을 효율적으로 반영할 수 있도록 설계되었음.(그림 15.)

이를 기반으로 전지구 3차원 해양 온도 및 염분 자료동화 시스템(GatedDA_TS)를 개발하였음. GatedDA_TS는 입력된 관측 정보와 배경 상태를 바탕으로 재분석장을 생성하며, 실제 해양자료동화 과정에 활용할 수 있도록 순환 구조를 갖춘 형태로 설계되었음. 이 모델은 해양 온도 및 염분을 동시에 동화할 수 있도록 설계되었음.

Dilated Convolution은 특정 맵의 해상도를 충분히 축소한 이후, 넓은 공간 범위의 정보를 효과적으로 추출하기 위해 도입된 구조임. 이러한 dilated convolution은 수용 범위를 확장함으로써 지역적인 정보를 넘어서 전 지구적인 해양 구조를 학습할 수 있게 하며, 단순한 깊은 계층 구성보다 효율적으로 넓은 맥락을 포착할 수 있다는 장점을 갖음.skip-connection 형태를 통해, 고해상도 정보의 손실을 최소화하고 복원 성능을 향상시킴.

모델의 입력에는 배경장(background field)과 관측장(observation field)이 모두 포함되며, 각 입력 변수에는 해당 시점의 게이팅 값(gating value), 즉 자료의 신뢰도 정보(confidence)가 함께 제공됨(표1).

이를 통해 모델은 온도와 염분 각각에 대해 관측 및 배경 정보의 상대적 신뢰도를 반영하여 분석장을 산출할 수 있도록 구성됨.

배경장은 초기 단계에서는 랜덤 초기값(randomly initialized background)을 사용하여 설정하였으며, 이후 자료동화 사이클부터는 이전 시점에서 산출된 분석 결과를 다음 시점의 배경장으로 순차적으로 전달하는 반복 구조를 적용함. 이러한 구조를 통해 시간에 따른 분석장의 연속성과 안정성을 확보하였음.

관측장에 부여되는 게이팅 값은 앞서 정의한 방법에 따라 관측 오차를 기준으로 산출된 값을 사용함.구체적으로는 관측 오차를 1일 예측 오차(one-day forecast error)의 크기를 기준으로 정규화하여 계산하였으며, 이를 통해 관측 자료의 신뢰도가 자료동화 과정에 정량적으로 반영되도록 하였음.

Community Earth System Model 2 (CESM2) large ensemble 프로젝트에서 제공한 전지구 5일 평균 해양 온도와 염분 데이터를 기반으로 딥러닝 모델을 훈련. 훈련 기간은 1892년부터 1932년까지, 검증 기간은 1933년부터 1973년까지로 설정하였다. 자료는 1도 간격의 공간 해상도와 5일 간격의 시간 해상도를 가지며, 이 기간 동안 모델이 다양한 시공간 스케일의 해양 구조를 학습할 수 있도록 구성하였음.

관측값에는 오차와 불확실성이 존재하기 때문에, 자료동화 과정에서 이를 반영하기 위해 본 연구에서는 평균이 0이고 표준편차가 각각 온도 0.002℃, 염분 0.005psu의 정규분포에서 생성된 랜덤 오차를 실제 관측값에 더하여 관측오차를 반영하였음.

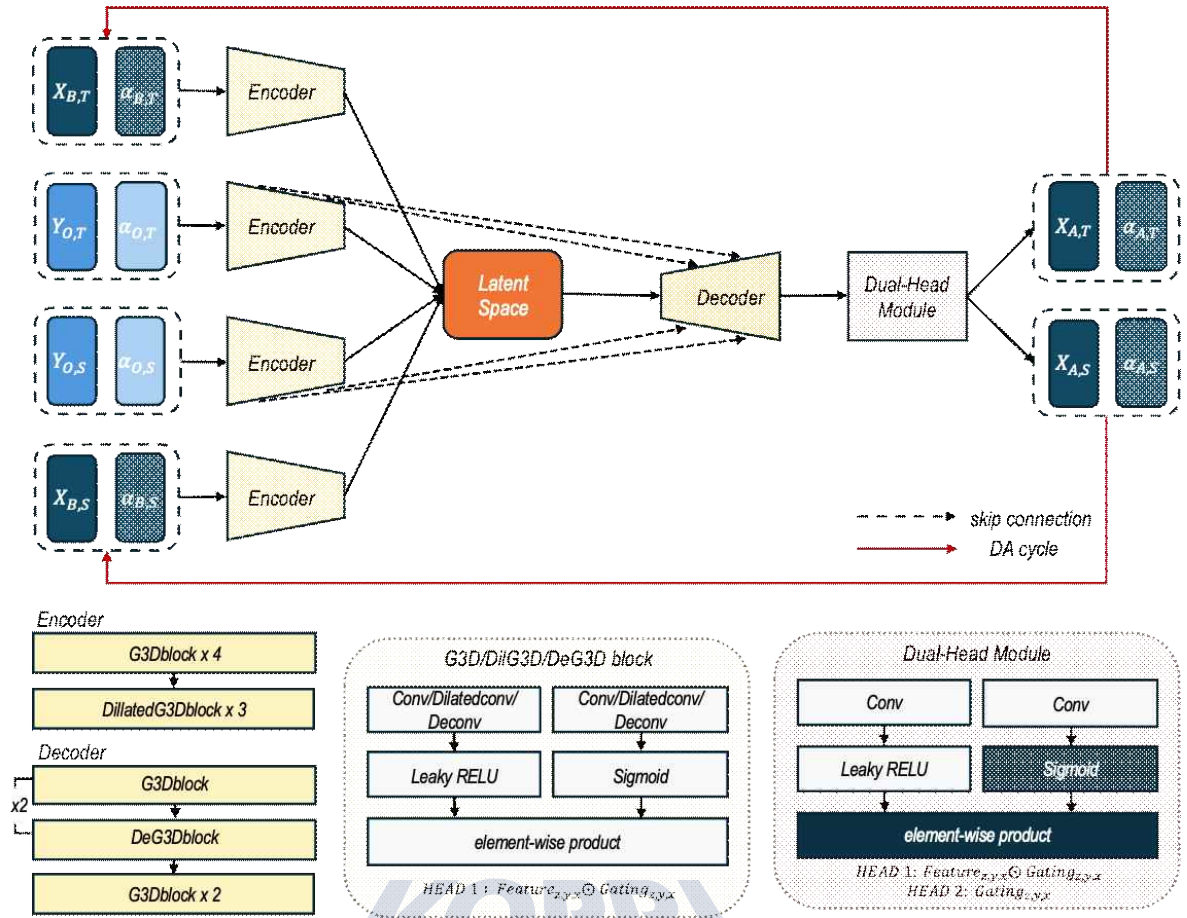


그림 15. Schematic diagram of the proposed deep learning - based multivariate data assimilation framework. The model simultaneously assimilates ocean temperature and salinity by encoding background and observation fields together with their corresponding gating values (confidence information). Multiple encoders project the inputs into a shared latent space, followed by a decoder and a dual-head module that separately estimates the analysis increments and gating-related outputs. Skip connections are used to preserve multiscale features, and the data assimilation (DA) cycle is implemented by feeding the analysis results back as background fields for the next assimilation step.

Experimental Design	
Data Domain	CESM2 -LE Horizontal : Global (360x180) Vertical : (37 lev, 0-700m) Period : 1985.01-2014.12
Variable	5-day mean Temperature, 5-day mean Salinity
Assimilation Frequency	Pentad (5-day)

표 1. Summary of the experimental design and CESM2-LE data configuration used for the multivariate data assimilation experiments



온도 관측이 염분 분석에 미치는 영향을 평가하기 위해 수행한 관측 시스템 모의 실험(OSSE, Observing System Simulation Experiment)을 함.

실제 해양 관측 환경에서는 염분 관측의 수가 온도 관측에 비해 현저히 적으며, 이로 인해 염분 분석에는 상대적으로 큰 불확실성이 존재함.

그러나 해양 수괴는 온도와 염분의 물리적 관계를 통해 정의되므로, 온도 - 염분 구조를 기반으로 할 경우 온도 정보는 염분 분석을 보완할 수 있는 중요한 역할을 수행할 수 있음.

이에 따라 본 연구에서는 온도 관측 정보가 염분 분석 정확도 향상에 기여하는 정도를 정량적으로 평가하기 위한 OSSE를 수행함.

실험은 두 가지 경우를 비교하는 방식으로 설계됨. 첫 번째는 CTL 실험으로, 온도와 염분 관측을 모두 동화한 다변량 자료동화 실험임. 두 번째는 noT 실험으로, 염분 관측만을 동화하고 온도 관측 정보는 0으로 설정하여 온도 정보의 영향을 배제한 실험임.

두 실험의 비교를 통해 온도 관측 정보가 염분 분석에 미치는 영향을 분리하여 평가할 수 있으며, 이를 통해 다변량 자료동화 구조에서 변수 간 상호작용의 중요성을 확인함.

그림 17은 게이팅 값(gating value)이 실제 분석 불확실성을 적절히 반영하는지를 검증하기 위해 수행한 OSSE(Observing System Simulation Experiment) 결과를 나타냄.

본 실험에서는 학습 단계에서 사용하지 않은 독립적인 앙상블 멤버를 활용하여 자료동화를 수행함으로써, 게이팅 값의 일반화 성능을 평가하였음.

그림 17의 상단 패널은 온도(좌)와 염분(우)에 대한 RMSE의 공간 분포를 나타내며, 하단 패널은 동일 시점에 대해 게이팅 값으로부터 계산된 불확실성(Gating Uncertainty)을 시간 평균한 결과를 제시함.

여기서 게이팅 불확실성은 학습된 게이팅 값을 기준으로 변환된 상대적 불확실성 지표로, 자료동화 과정에서 각 격자점의 신뢰도를 정량적으로 표현함.

두 변수 모두에서 RMSE가 크게 나타나는 해역일수록 게이팅 불확실성 역시 높게 분포하는 경향이 확인되며, 특히 서안경계류 및 변동성이 큰 해역에서 이러한 특성이 두드러짐.

RMSE와 게이팅 불확실성 간의 공간 패턴 상관계수는 온도에서 0.71, 염분에서 0.65로 나타나, 두 지표 간에 높은 공간적 일관성이 존재함을 보여줌.

이러한 결과는 본 연구에서 학습된 게이팅 값이 단순한 경험적 가중치가 아니라, 실제 분석 오차의 공간적 분포를 효과적으로 반영하는 불확실성 지표로 활용 가능함을 정량적으로 입증함.

즉, 게이팅 기반 자료동화 구조는 해양 상태의 지역별 예측 신뢰도를 표현하고, 향후 가변 오차 공분산 또는 불확실성 기반 자료동화 기법으로 확장될 수 있는 잠재력을 가짐.

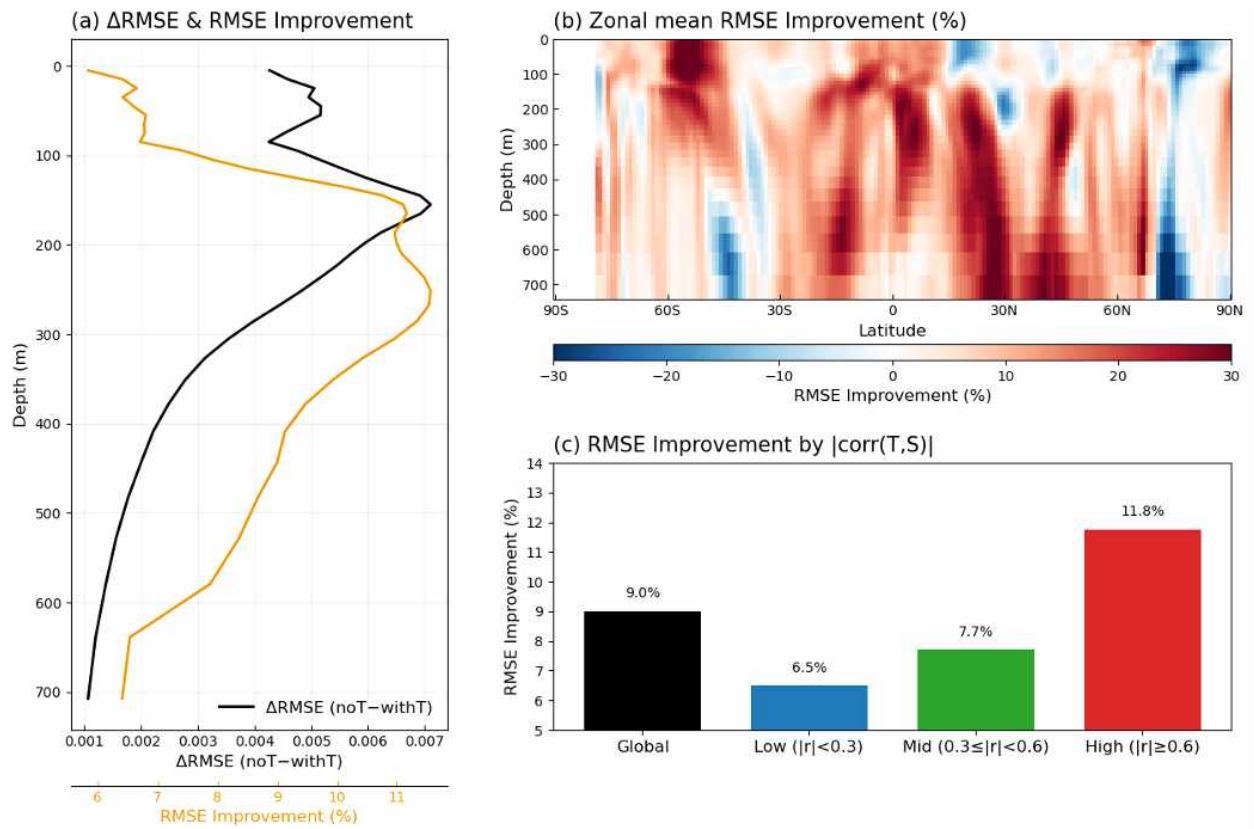


그림 16. Salinity RMSE improvement due to assimilating temperature observations in a multivariate data assimilation framework, shown as (a) vertical profiles, (b) zonal-mean latitude - depth structure, and (c) dependence on temperature - salinity correlation strength.

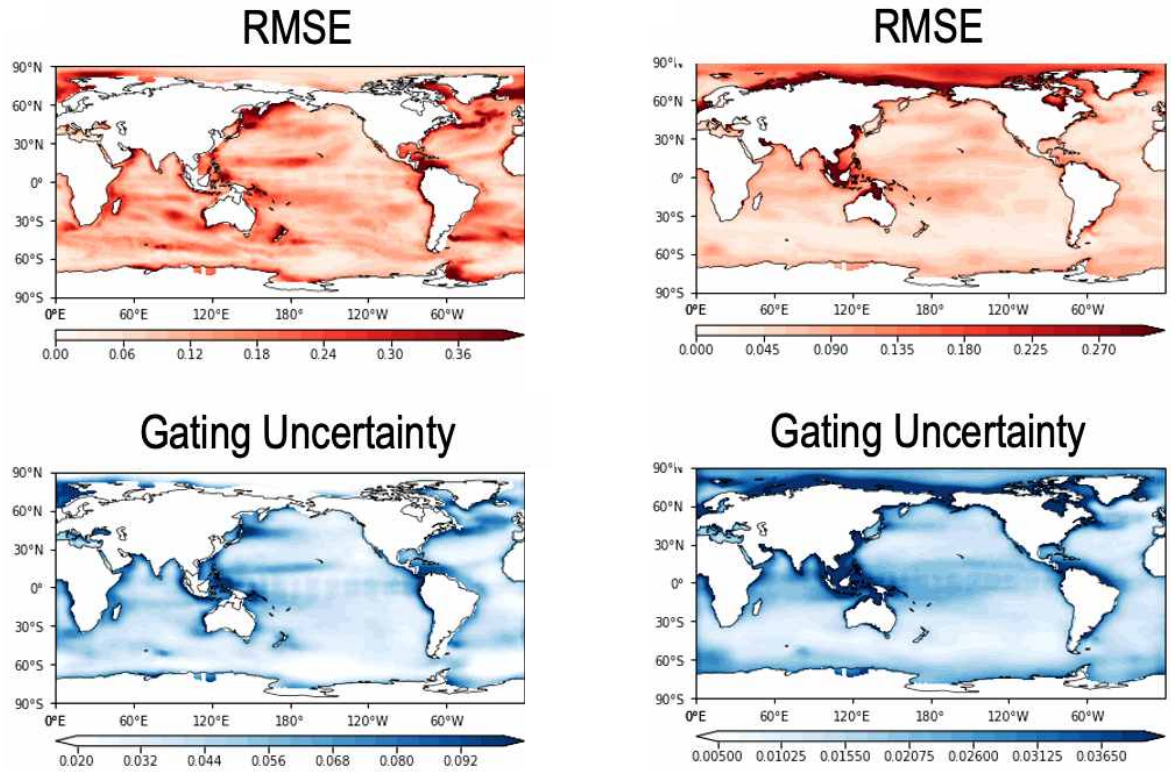


그림 17. Validation of gating values as an uncertainty metric using OSSE experiments. Spatial distributions of RMSE (top) and time-mean gating-based uncertainty (bottom) are shown for temperature (left) and salinity (right). Pattern correlation coefficients between RMSE and gating uncertainty are 0.71 for temperature and 0.65 for salinity, indicating strong spatial consistency. Regions with larger RMSE tend to exhibit higher gating uncertainty, demonstrating that the learned gating values effectively represent analysis uncertainty.

Experimental Design	
Data Domain	Horizontal : Global (360x180,Global) Vertical : (37 lev, 0-700m) 1995.01-2024.12 (1-yr spin-up time)
Variable	5-day mean Oceanic Temperature, 5-day mean Oceanic Salinity
Assimilation Frequency	Pentad (5-day)
Oceanic In-situ observation	TOGA/TAO moored buoy ARGO float CMEMS drifting buoy Bathythermograph, Expendable bathythermograph Digital bathythermograph CTD, expendable CTD Merchant ship Ocean station vessel Coastal/island station
SST observation	OISST V2
Initial Condition	Arbitrary selected from CESM2 LE

Figure 2. Observational datasets and data configuration for real-observation-based multivariate ocean data assimilation. The assimilation was conducted over the global ocean domain with a horizontal resolution of 360×180 and 37 vertical levels from the surface to 700 m, covering the period January 1995 to December 2024 with a one-year spin-up. Five-day mean oceanic temperature and salinity were assimilated at a pentad (5-day) frequency. In-situ observations include TOGA/TAO moored buoys, ARGO floats, CMEMS drifting buoys, bathythermographs (BT, XBT, DBT), CTD and expendable CTD measurements, merchant ships, ocean station vessels, and coastal/island stations. Sea surface temperature observations were obtained from OISST v2. Initial conditions were selected from the CESM2 Large Ensemble.

표 2는 본 연구에서 실제 해양 관측(real-ocean observations)을 활용한 자료동화에 사용된 입력 자료 구성과 적용 범위를 요약한 것임.

자료동화는 전 지구 해양 영역을 대상으로 수행되었으며, 수평 해상도는 360×180 , 연직 방향으로 표층부터 수심 700 m까지 총 37개 층으로 구성하였음.

분석 기간은 1995년 1월부터 2024년 12월까지이며, 초기 스핀업 효과를 최소화하기 위해 1년의 spin-up 기간을 포함하였음. 자료동화에 사용된 변수는 5일 평균 해양 온도와 염분으로, 자료동화 주기는 pentad(5일) 단위로 설정하였음.

자료동화 초기 조건은 CESM2 Large Ensemble에서 임의로 설정하였음.

이러한 관측 자료 구성은 전 지구 해양의 시공간적 특성을 포괄적으로 반영할 수 있도록 설계되었으며, 본 연구에서 제안한 다변량 자료동화 기법의 실질적인 적용 가능성을 검증하는 기반을 제공함.

그림 18은 실제 HadIOD 관측 자료의 개수와 게이팅 값으로부터 추정된 불확실성 간의 관계를 나타냄. 그림 18의 왼쪽 패널은 해양 온도, 오른쪽 패널은 해양 염분에 대한 결과를 제시한다. 검은색 선은 해당 격자 및 시점에서의 관측 개수(observation count)를 나타내며, 파란색과 빨간색 선은 게이팅 값으로부터 계산된 불확실성 지표의 의미함.

두 변수 모두에서 관측 개수가 증가함에 따라 불확실성이 감소하는 뚜렷한 경향이 확인됨.

이는 관측 정보가 풍부한 구간에서 게이팅 값이 높은 신뢰도를 부여하고, 관측이 부족한 구간에서는 상대적으로 높은 불확실성을 반영하고 있음을 의미함.

이러한 결과는 게이팅 값이 실제 관측 자료의 양에 따라 자료 신뢰도 및 불확실성을 합리적으로 표현하고 있음을 검증하며, 본 연구에서 제안한 게이팅 기반 다변량 자료동화 구조의 타당성을 뒷받침함.

그림 19는 실제 해양 관측 자료를 적용한 자료동화 결과를 바탕으로 해양 온도와 염분 분석 성능을 평가한 결과를 제시함.

그림 19의 왼쪽 패널은 상관계수 기반 공간 분포를, 오른쪽 패널은 전 지구 평균 시계열 비교 결과를 나타냄.

성능 비교에 사용된 자료는 다음과 같음. GatedDA는 본 연구에서 개발한 딥러닝 기반 자료동화 결과이며, EN4.2.2는 관측에 기반한 해양 분석 자료임. 또한 MME_Others는 ORAS5를 포함한 여러 해양 재분석 자료를 결합한 다중모델 앙상블 결과를 의미함.

그림 19의 왼쪽 패널에서 등치선(contour)은 GatedDA와 EN4.2.2 간의 상관계수를 나타내며, 음영(shading)은 GatedDA - EN4.2.2 상관계수와 MME_Others - EN4.2.2 상관계수 간의 차이를 의미함.

전반적으로 GatedDA는 다중모델 앙상블에 비해 높은 상관계수를 보이며, 특히 남극 주변 해역, 서안경계류 지역, 그리고 북대서양에서 상관성이 뚜렷하게 개선된 것으로 나타남.

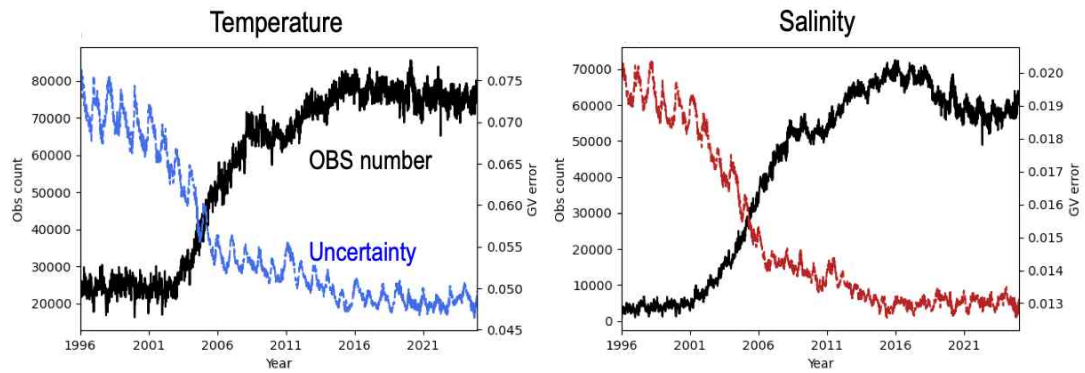


그림 18. Relationship between the number of HadIOD observations and uncertainty estimated from gating values. Results are shown for (left) temperature and (right) salinity. The black line represents the observation count, while the blue and red lines denote uncertainty metrics derived from the gating values. A clear inverse relationship is observed, with uncertainty decreasing as the number of observations increases, indicating that the gating values appropriately capture observation-dependent uncertainty



그림 19의 오른쪽 패널은 수심 5 - 300 m 및 5 - 700 m 구간에 대해 평균한 전 지구 평균 시계열을 비교한 결과이다. GatedDA(적색)는 관측 기반 EN4.2.2(흑색)의 장기 변동 추세를 가장 잘 재현하며, MME_Others(청색)에 비해 보다 안정적이고 일관된 시계열 변화를 나타냄.

이러한 경향은 특히 염분(Salinity)에서 더욱 뚜렷하게 나타나며, 이는 다변량 자료동화 구조와 게이팅 기반 신뢰도 반영 기법이 염분 분석 성능 향상에 효과적으로 기여하고 있음을 시사함.

그림 20의 왼쪽 패널은 실제 관측을 적용한 자료동화 결과에서 해양 온도의 주요 공간 변동성을 EOF 분석을 통해 비교한 결과를 나타냄. 표층(5 m) 및 5 - 300 m 평균에 대해 EN4.2.2, MME_Others, 그리고 GatedDA의 EOF 제1모드 공간 분포를 제시하였음.

표층(5 m)과 5 - 300 m 평균 모두에서 GatedDA는 EN4.2.2에서 나타나는 주요 대규모 변동 패턴을 전반적으로 잘 재현하는 것으로 나타남.

특히 열대 태평양과 인도양, 그리고 중위도 해역에서의 공간 구조가 다중모델 앙상블(MME_Others)에 비해 EN4.2.2와 보다 유사한 형태를 보임.

이는 딥러닝 기반 자료동화가 온도 관측 정보를 효과적으로 반영하여 전 지구 규모의 온도 변동성을 안정적으로 분석하고 있음을 시사함.

지역별 주성분(PC) 상관계수 분석 결과에서도 GatedDA는 대부분의 해역에서 MME_Others와 유사하거나 다소 높은 상관을 보이며, 특히 열대 태평양 및 인도양에서 관측 기반 EN4.2.2와의 시계열 일관성이 유지되는 것으로 나타남.

이러한 결과는 본 연구에서 제안한 자료동화 기법이 온도 변동성의 주요 시공간 특성을 안정적으로 재현하고 있음을 보여줌.

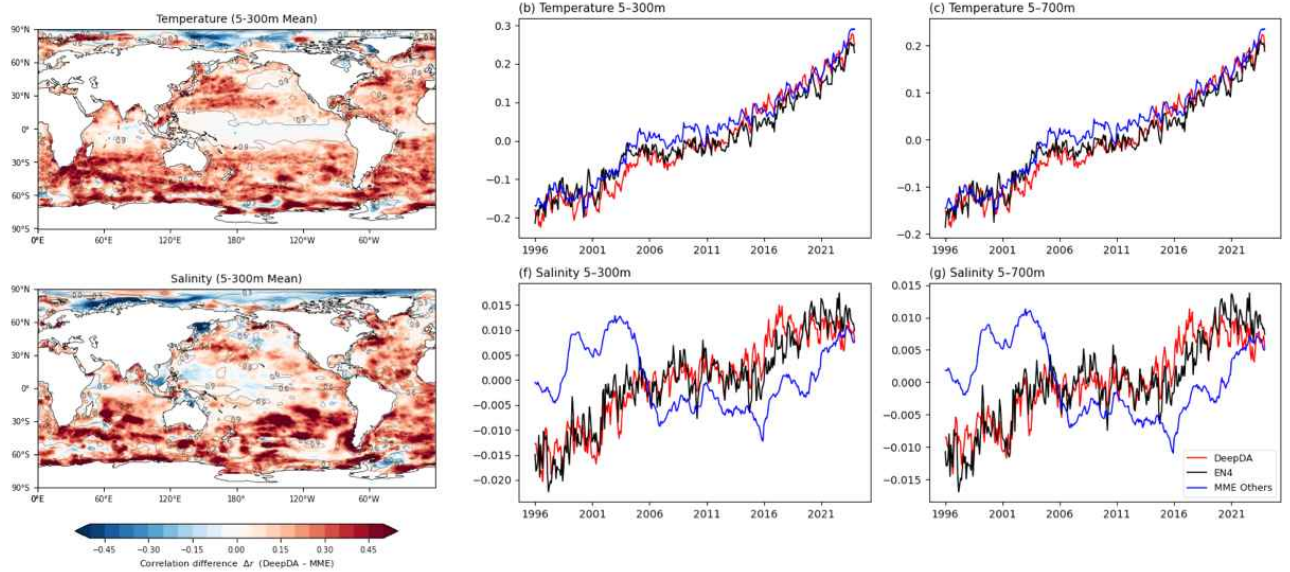


그림 19. Performance evaluation of temperature and salinity analyses based on real-observation data assimilation. (a, d) Spatial distribution of correlation differences (Δr) between GatedDA and the multi-model ensemble (MME) with respect to EN4.2.2 for (a) temperature and (d) salinity, averaged over the 5 - 300 m depth range. Positive values indicate higher correlation of GatedDA with EN4.2.2 than MME. (b, c) Global mean time series of temperature anomalies averaged over the (b) 5 - 300 m and (c) 5 - 700 m depth ranges. (f, g) Same as (b, c) but for salinity. Red, black, and blue lines denote GatedDA, EN4.2.2, and MME_Others, respectively. GatedDA shows closer agreement with EN4.2.2 and more stable long-term variability, with particularly pronounced improvements for salinity.

극지연구소

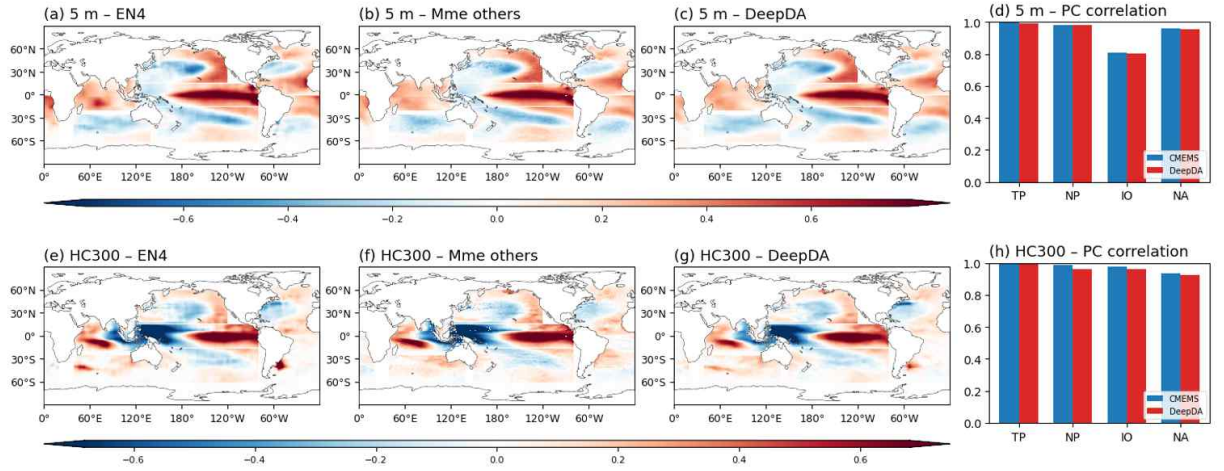


그림 20. EOF analysis of ocean temperature based on real-observation data assimilation. (a - c) Spatial patterns of the first EOF mode of temperature at 5 m depth derived from (a) EN4.2.2, (b) MME_Others, and (c) GatedDA. (d) Regional principal component (PC) correlation coefficients between each dataset and EN4.2.2 for the 5 m temperature field. (e - g) Same as (a - c) but for vertically averaged temperature over the upper 300 m (HC300). (h) Same as (d) but for HC300. GatedDA exhibits spatial patterns more consistent with EN4.2.2 than the multi-model ensemble, along with higher PC correlations across most ocean basins, indicating improved representation of dominant temperature variability.



그림 21의 왼쪽 패널은 해양 염분에 대한 EOF 제1모드 공간 분포 및 주성분 상관계수 분석 결과를 나타냄.

염분의 경우 온도에 비해 관측 자료가 상대적으로 제한적이며, 이에 따라 재분석 간 불확실성이 큰 변수로 알려져 있음.

EOF 공간 패턴을 살펴보면, GatedDA는 EN4.2.2의 주요 염분 변동 구조를 MME_Others에 비해 더욱 명확하게 재현하는 것으로 나타남.

특히 북대서양, 서안정계류 인근, 그리고 고위도 해역에서 GatedDA의 공간 패턴은 EN4.2.2와 높은 유사성을 보이며, 다중모델 앙상블에서는 상대적으로 약화되거나 분산된 구조가 나타남.

주성분(PC) 상관계수 분석에서도 이러한 경향이 뚜렷하게 확인됨. GatedDA는 북대서양을 포함한 여러 해역에서 EN4.2.2와의 PC 상관계수가 MME_Others보다 현저히 높게 나타나며, 이는 염분 변동성의 시간적 진화를 보다 일관되게 재현하고 있음을 의미함.

이러한 개선은 게이팅 값을 통해 관측 불확실성을 조절하고, 온도 - 염분 간의 다변량 상관관계를 자료동화 과정에 반영한 결과로 해석될 수 있음.

종합적으로, EOF 분석 결과는 GatedDA가 온도뿐만 아니라 염분에서도 주요 공간·시간 변동성을 효과적으로 재현하며, 특히 변동성이 크고 관측이 제한적인 해역에서 분석 안정성을 향상시키는 데 기여함을 시사함.

그림 22은 HadIOD 실측 관측 자료를 기준으로 DeepDA 배경장과 분석장의 RMSE를 비교한 결과를 나타냄. 분석 기간은 1년의 스핀업 기간을 제외한 1996년부터 2023년까지이며, 모든 RMSE는 관측이 존재하는 지점에서만 계산하였음.

그림 22왼쪽 패널은 해양 온도, 오른쪽 패널은 해양 염분에 대한 연직 RMSE 분포를 보여줌. 검은색 선은 배경장(background)의 RMSE를, 빨간색 선은 자료동화 이후 산출된 분석장(analysis)의 RMSE를 의미함.

두 변수 모두에서 전 수심에 걸쳐 분석장의 RMSE가 배경장보다 일관되게 감소한 것으로 나타남. 온도의 경우, 표층부터 중층까지 분석장의 RMSE 감소가 뚜렷하게 나타나며, 이는 관측 정보가 배경장 대비 효과적으로 반영되었음을 의미함. 염분에서도 유사한 경향이 확인되며, 특히 관측이 상대적으로 제한적인 염분 변수임에도 불구하고 분석장에서 RMSE가 전반적으로 감소한 점은 다변량 자료동화 구조와 게이팅 기반 불확실성 조절 기법의 효과를 보여줌.

이러한 결과는 본 연구에서 개발한 DeepDA 기반 자료동화 시스템이 실제 해양 관측 정보를 안정적으로 흡수하여 배경장의 오차를 감소시키고, 보다 정확한 해양 상태 분석장을 생성하고 있음을 정량적으로 입증함.

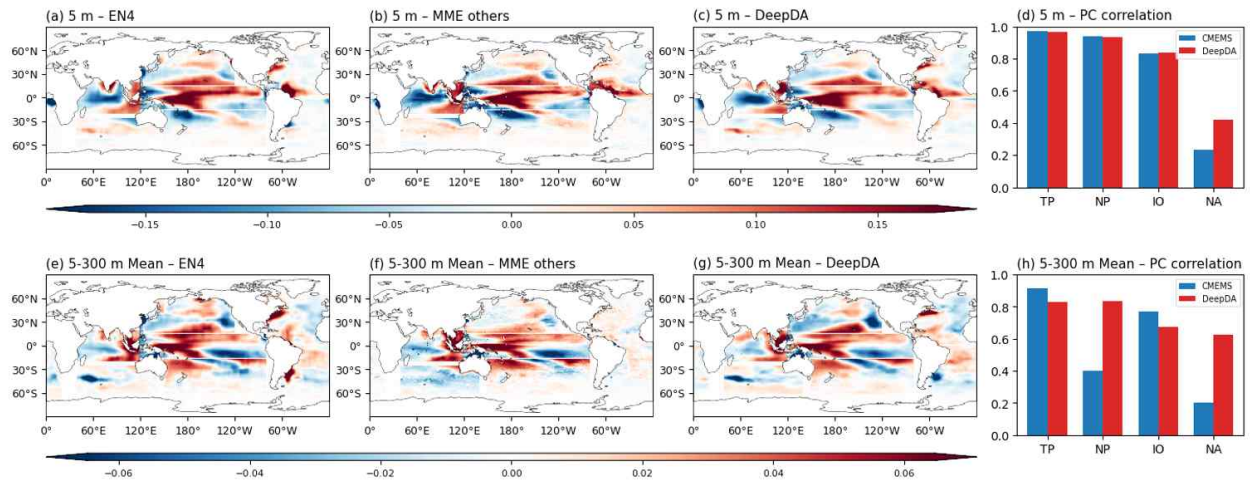


그림 21. EOF analysis of ocean salinity based on real-observation data assimilation.

(a - c) Spatial patterns of the first EOF mode of salinity at 5 m depth derived from (a) EN4.2.2, (b) MME_Others, and (c) GatedDA. (d) Regional principal component (PC) correlation coefficients between each dataset and EN4.2.2 for the 5 m salinity field.

(e - g) Same as (a - c) but for vertically averaged salinity over the upper 300 m (HC300). (h) Same as (d) but for HC300. GatedDA shows substantially improved agreement with EN4.2.2 compared to the multi-model ensemble, particularly in regions with strong salinity variability, as evidenced by more coherent EOF spatial patterns and higher PC correlations across most ocean basins.



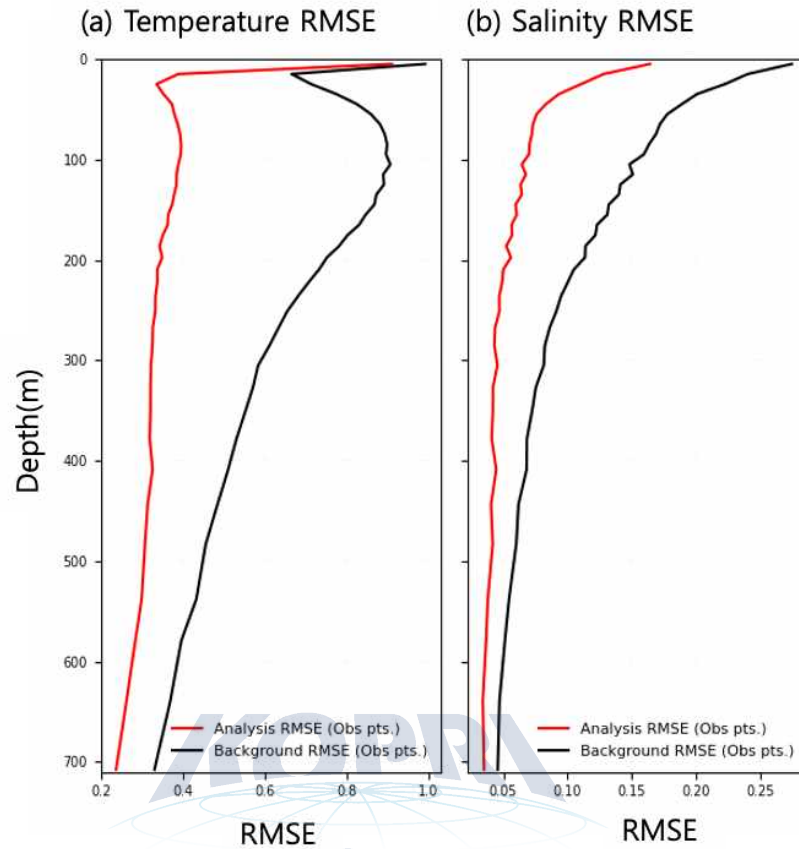


그림 22. vertical profiles of RMSE for ocean temperature and salinity with respect to HadIOD in-situ observations. RMSEs are computed over the period 1996 - 2023 after a one-year spin-up. Red lines denote analysis RMSE, while black lines indicate background RMSE, both evaluated at observation locations. Results are shown for (a) temperature and (b) salinity. Reduced RMSE in the analysis compared to the background demonstrates the effective assimilation of observational information.

6절 지구시스템 모델(CESM2) 기반 3차원 해양 온도 및 염도 너징 시스템 개발

1. 3차원 너징 시스템 개발 및 안정도 실험 수행

전 지구 결합 모형 중 하나인 국립대기연구센터 (NCAR; National Center for Atmospheric Research)에서 개발 및 제공하는 지구시스템 모델 (CESM2; Community Earth System Model) 활용.

모형의 정상적인 구동 여부 확인 및 규준 실험 확보를 위한 적분 수행.

3차원 해양 온도 및 염분 너징 시스템을 개발하기 위해, 해양모델인 POPv2 (Parallel Ocean Program version 2)의 예단 변수 구조와 소스 코드 분석.

특히, 딥러닝 기반 자료동화 모델에서 산출된 온도 및 염분 재분석장이 POPv2의 입력 격자(pop grid)와 일치하도록 변환하는 작업을 수행.

해양 자료동화를 통해 생산된 재분석장의 너징 시스템 내 입력 체계 구축.

너징 시스템의 성공적인 구축 여부를 위한 모형 예단 변수를 입력 재분석 자료로 대체 (replace) 하는 이상 실험 (idealized experiment) 수행.

그림 13은 딥러닝 기반으로 생성된 nudging input이 기존 CESM2 모형 출력과 얼마나 잘 일치하는지를 시각적으로 비교한 결과임. (a)는 CESM2의 해수면 온도(SST) 출력 결과를 나타내며, (b)는 딥러닝 모형의 output을 POP 격자에 맞게 변환한 nudging input. (c)에서는 (a)와 (b)의 차이를 보여주며, 두 필드 간의 공간적 패턴이 정합적으로 일치함을 확인. (d)는 최종 nudging input으로, 실제 POPv2 모형 적분 시 사용됨.

비교적 단순한 구조로 연산 효율적이며 재분석장의 안정적인 입력이 가능한 너징 (Nudging or Newtonian relaxation) 기법 기반의 자료동화 시스템 구축.

너징 기법은 모형 내 입력된 관측치에 모형 예단변수가 비슷해지도록 하나의 항을 진단 방정식에 삽입하는 과정으로 아래와 같은 모형 내 진단 방정식 활용.

$$\frac{\partial T, S_n}{\partial t} = \frac{\partial T, S}{\partial t} + \frac{T, S_O - T, S_M}{\tau} a$$

좌변은 nudging을 통해 보정된 해양 온도 및 염분의 시간 변화율이며, 우변의 첫 번째 항은 물리 모형에서 예단되는 시간 변화율, 두 번째 항은 2차년도에 개발한 딥러닝 기반 해양 자료동화 시스템을 통해 생산된 해양 온도 및 염분 재분석장을 너징으로 처방해주는 보정 항임.

너징 시스템의 성공적인 구축 여부를 위한 모형 예단 변수를 입력 재분석 자료로 대체 (replace) 하는 이상 실험 (idealized experiment) 수행.

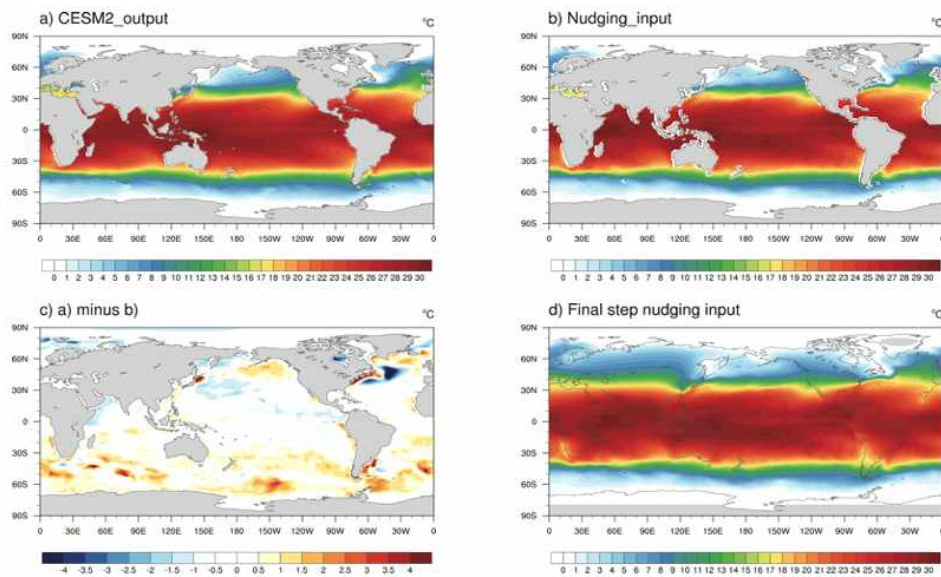


그림 23. Comparison of sea surface temperature (SST) fields between CESM2 output and the nudging input generated from the deep learning-based data assimilation system. (a) CESM2 model SST output (b) nudging input interpolated to the POP model grid using deep learning analysis output (c) difference field between (a) and (b), confirming the consistency of spatial patterns (d) final step nudging input.



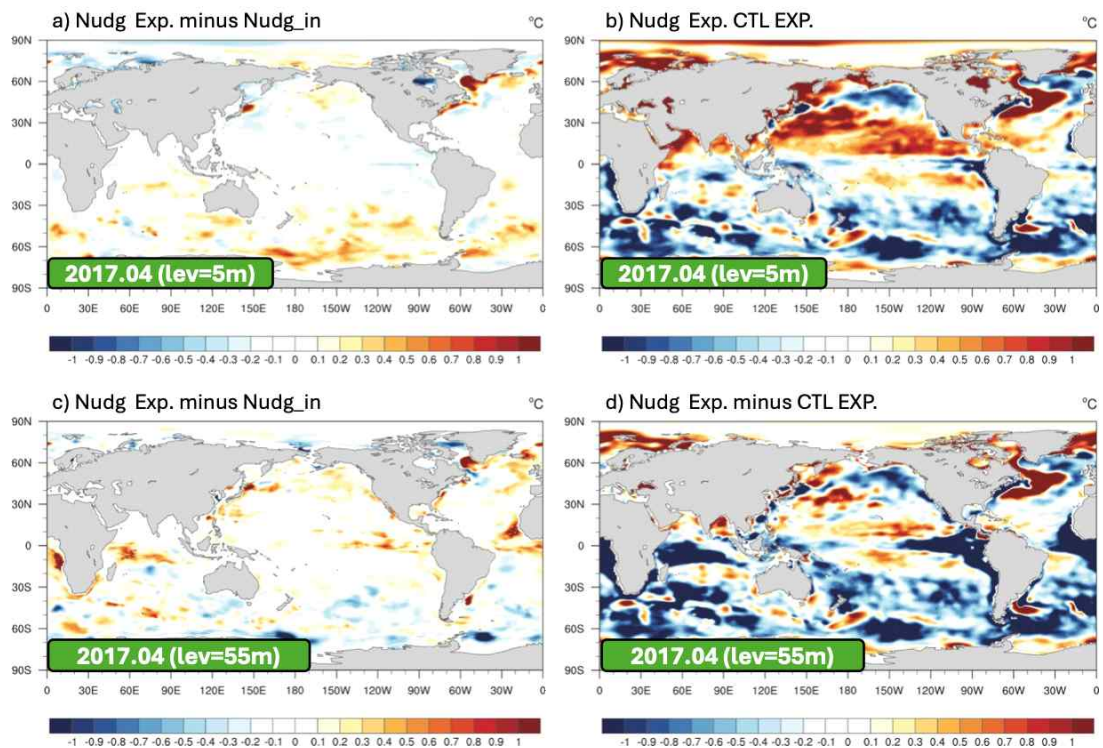


그림 24. Monthly mean temperature differences at 5m depth (lev=5m) for April 2017. (a) Difference between the nudging experiment and the GatedDA-based nudging input (b) nudging experiment minus control experiment. Monthly mean temperature differences at approximately 55 meters depth (lev=55m) for April 2017. (c) Nudging experiment minus GatedDA-based nudging input. (d) Nudging experiment minus control experiment. system at deeper levels.

본 연구에서는 2차년도에 개발된 딥러닝 기반 자료동화 시스템(GatedDA)을 통해 생성된 3차원 해양 온도 재분석장을 활용하여, CESM2 해양모델에서 nudging 기반의 온도 자료동화를 수행(그림 24).

너징 실험은 2017년 4월 1일부터 30일까지 총 1개월간 수행되었으며, nudging 타임 스케일은 하루(1 day)로 설정. 또한, 단순히 표층 온도만을 보정하는 데 그치지 않고, 수직 방향으로 nudging 효과가 반영되도록 설정하기 위해 약 55m 수심에 해당하는 레벨 5(lev=5)까지 너징을 적용

이를 통해 재분석장의 정보가 수직적으로도 모형에 충분히 전달되도록 하였으며, 삼차원 구조를 갖는 해양 온도 nudging 시스템의 적용 가능성을 실험적으로 평가하고자 함.

그림 24는 이러한 nudging 실험의 결과를 다양한 비교를 통해 시각화한 것으로, 각 그림은 실험 간의 해양 온도 차이를 월평균 기준으로 나타냄. 그림 24.(a)는 nudging 실험 결과와 GatedDA 기반 nudging 입력장 간의 월평균 해양 온도 차이를 시각화한 것임.

해당 그림에서는 일부 지역을 제외하면 대부분의 해역에서 차이가 매우 작게 나타나며, 이는 nudging을 통해 모형의 예단장이 재분석 입력장에 성공적으로 수렴하였음을 의미함.

이는 CESM2 해양모델이 nudging을 통해 입력 재분석장으로의 수렴을 잘 수행했음을 보여주는 정량적 결과이며, 딥러닝 기반 재분석장이 실제 모델 내에서 안정적인 nudging 입력장으로 활용될 수 있음을 나타냄. 그림 (b)는 nudging 실험 결과와 control 실험 결과 간의 차이를 나타냄.

control 실험은 동일한 CESM2 해양모델을 사용하되, nudging 항이 포함되지 않은 상태로 적분한 실험임. 즉, 동일한 초기조건과 외부 강제력 하에서 nudging만을 제외한 실험 결과로, nudging 기법의 효과를 직접적으로 평가할 수 있음. 그림 (b)에 나타난 바와 같이, 너징이 적용된 실험에서는 전 해역에 걸쳐 control 실험과 유의한 차이가 나타남

이러한 차이는 수심 5m뿐만 아니라, 수심 55m에 해당하는 레벨에서도 일관되게 관측되었으며(그림 24), 이는 nudging 효과가 수직적으로도 잘 전달되었음을 의미한다. 그림 24(d)는 수심 55m에서의 nudging 실험 결과와 control 실험 결과 간의 월평균 해양 온도 차이를 나타낸 것으로, nudging 입력장이 해양 내부까지 안정적으로 확산되었음을 보여준다. 이는 nudging이 단지 표층에만 작용한 것이 아니라, 수심 방향으로도 입력 정보가 효과적으로 반영되었음을 의미함.

이는 딥러닝 기반 재분석장이 모델 내 물리장에 효과적으로 영향을 주었음을 의미하며, nudging 시스템이 온도 예단 변수에 대해 성공적으로 작동했음을 시사함.

nudging input은 딥러닝 모델의 출력을 해양모형의 격자계에 맞게 변환하여 생성된 것이며, 물리적 연속성이 고려되지 않으면 모델 적분에 불안정성을 초래할 수 있음. 본 실험 결과는 딥러닝 기반 자료동화 모델이 생산한 재분석장이 해양모형에 안정적

으로 연동될 수 있다는 것을 나타내며 그 결과가 기존 모형 출력(control)과 비교하여 실질적인 차이를 만들어낼 수 있다는 점을 동시에 보여줌.

2. 너징 민감도 실험

앞선 실험을 통해 딥러닝 기반 재분석장을 활용한 해양 온도 nudging 시스템이 CESM2 해양모형 내에서 성공적으로 구현되었음을 확인하였음.

이에 추가적으로, nudging 반영 강도에 따른 시스템의 민감도를 평가하기 위해 완화 시간규모(nudging time scale)에 대한 민감도 실험을 추가로 수행.

이전 실험에서는 time scale을 1일(24시간)로 설정하였으나, 본 실험에서는 이를 12시간으로 줄여 더 강한 nudging 효과가 모형 내에 작용하도록 구성하였음. 시간 규모를 줄이는 것은 모형이 재분석장에 더 빠르게 수렴하도록 만드는 방식임.

그림 25은 nudging time scale을 1일에서 12시간으로 줄여 더 강하게 너징을 수행한 실험의 결과를 수심 5m 기준으로 나타낸 것임. 이 실험은 딥러닝 기반 재분석장이 해양모형 예단 변수에 얼마나 효과적으로 반영되는지를 평가하고, 너징 강도 변화에 따른 모델 반응을 확인하기 위해 수행되었음.

그림 25의 좌측 패널(a)은 너징 실험 결과와 딥러닝 기반 입력 재분석장 간의 해양 온도 차이를, 그림 25의 우측 패널(b)은 너징 실험 결과와 너징이 적용되지 않은 control 실험 간의 차이를 각각 월평균 기준으로 시각화한 것임.

실험 결과, 열대 및 아열대 해역을 중심으로 두 자료 간의 차이가 현저히 줄어들었음. 이는 time scale을 짧게 설정할수록 입력장에 대한 수렴이 더욱 빠르고 강하게 이루어졌음을 보여줌. 또한, 일본 동쪽 해역과 북대서양 연안 등 주요 해류 경계 부근에서도 차이 감소가 두드러지게 나타났음.

그림 25의 하단패널은 nudging time scale을 12시간으로 설정한 민감도 실험 결과를 수심 약 100m를 시각화한 것이다. 그림을 통해 딥러닝 기반 재분석장이 수심 방향으로도 정보를 포함하고 있을 때, 해당 입력이 실제 해양모형의 내부 수온 구조까지 효과적으로 반영되는지를 평가할 수 있음.

그림 25(c)는 너징 실험 결과와 입력 재분석장 간의 월평균 온도 차이를 나타내며, 그림 25(d)는 너징 실험 결과와 control 실험 간의 차이를 보여줌.

그림 25(c)를 통해 확인할 수 있듯이 딥러닝 기반으로 생성된 3차원 재분석장이 단지 표층에만 국한되지 않고, 수심 약 100m 수준까지 포함되어 모델에 효과적으로 전달되었음을 의미함. 또한 그림 25(d)에서는 control 실험과의 온도 차이가 전 해역에 걸쳐 강하게 나타나며, nudging 효과가 모델 수온 구조 전반에 반영되었음을 확인할 수 있음.

이는 강한 너징 조건 하에서 입력된 3차원 재분석장이 해양모형 내부 구조에까지 안정적으로 주입되었음을 보여주는 결과로, 본 연구에서 구현한 3차원 너징 시스템이 표층을 넘어 해양 내부까지 작동했음을 입증함.

이러한 결과는 완화시간규모를 줄임으로써 nudging 강도를 증가시켰을 때, 딥러닝

기반 입력장이 해양모델에 더 빠르고 강하게 전달됨을 보여줌.

더불어, 입력장이 3차원적으로 구성되어 있었기 때문에, 수심 100m 수준에서도 모델이 재분석 정보를 효과적으로 반영할 수 있었으며, 이는 해양 전 층에 걸친 3차원적 nudging 시스템의 구현 가능성과 안정성을 입증하는 중요한 결과로 해석됨.

3. 지구시스템 모델(CESM2) 모형에 기반한 장기간 (20년 이상) 전지구 해양 초기장 생산

딥러닝 기반 해양 온도 염분 다변량 자료동화 모델의 재분석장을 활용하여, 해양모델에 3차원적으로 온도와 염분 정보를 동화시키는 nudging 시스템을 완성하였으며, 이를 기반으로 장기간(20년 이상) 해양모델 적분이 가능한 전 지구 해양 초기장 생산 체계를 확보하였음.

특히, 재분석장의 품질 검증과 너징 민감도 실험을 통해 입력장으로서의 안정성과 물리적 신뢰도를 확보하였으며, 이는 향후 다양한 기간에 대한 모형 초기장 생성에 직접 활용될 수 있음.

딥러닝 기반으로 생성된 3차원 해양 온도·염분 재분석장을 해양모델(CESM2 POPv2)에 적용한 nudging 시스템이 성공적으로 구현되었으며, 입력 자료가 모델 구조에 효과적으로 반영됨을 다양한 수심에서 확인하였음.



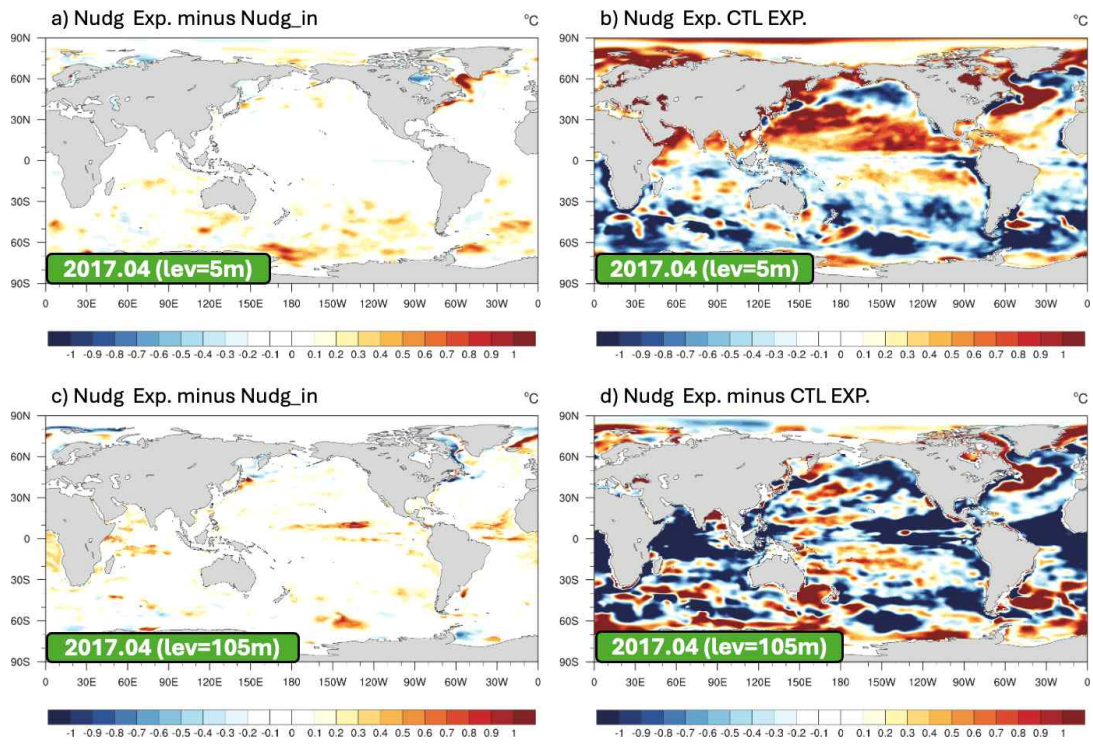


그림 25. Monthly mean temperature difference between nudging experiment and nudging input in April 2017 at ocean with a reduced nudging time scale (12 hours). (a)Nudging experiment minus GatedDA-based nudging input. (b)Nudging experiment minus control experiment. (c)Nudging experiment minus GatedDA-based nudging input at 105m depth. (d)Nudging experiment minus control experiment at 105m depth.

제 4장 연구개발목표 달성도 및 대외기여도

본 연구는 연도별 연구목표 및 평가착안점에 따라 단계적으로 수행되었으며, 각 연차별 연구개발목표를 계획대로 달성함으로써 딥러닝 기반 전지구 3차원 해양 자료동화 및 초기화 기술의 기술적 완성도를 확보함과 동시에 관련 분야의 기술 발전에 기여함.

1절 연도별 연구개발목표 달성도

1. 1차년도

1차년도에는 관측 자료의 공간적 결손 문제를 효과적으로 처리할 수 있는 딥러닝 기반 자료동화 기법에 대한 기초 연구를 수행함.

부분 합성곱 신경망(Partial Convolutional Neural Network)의 결측값 처리 특성을 분석하고, 간헐적으로 존재하는 해양 관측 자료를 전지구 격자에 안정적으로 반영할 수 있는 자료동화 기법으로서의 적용 가능성을 검토함.

또한, 적대적 생성 신경망(Generative Adversarial Network)의 이미지 생성 특성을 자료동화 문제에 적용함으로써, 생성된 재분석장이 통계적으로 관측 자료와 유사한 분포를 갖도록 유도할 수 있는 기반을 마련함. 이를 통해 고품질의 재분석장 생산을 위한 딥러닝 기반 자료동화 구조를 정립함.

딥러닝 모형 학습을 위한 충분한 학습 샘플 확보를 위해 CESM2 large ensemble 자료를 수집하고, 장기 적분 자료를 데이터베이스화함.

또한, 실제 관측 네트워크의 공간적 분포를 반영하기 위해 HadIOD 관측 자료를 확보하였으며, 자료동화 결과 검증에 위해 NOAA OISST 및 EN4 등 주요 재분석 자료를 수집함.

장기 적분 자료, 관측 자료, 재분석 자료를 통합적으로 데이터베이스화함으로써, 딥러닝 기반 자료동화 시스템 개발 및 검증을 위한 기초 자료 구축 목표를 달성함.

TensorFlow 기반으로 부분 합성곱 신경망과 적대적 생성 신경망을 결합한 전지구 3차원 해양 온도 자료동화 시스템을 구축함.

3차원 합성곱 구조를 적용하여 수평 및 연직 방향의 물리적 일관성을 고려하였으며, 이전 사이클의 재분석장과 현재 관측을 입력으로 하는 반복 구조의 자료동화 시스템을 구현함.

이를 통해 공간 해상도 약 $1^{\circ} \times 1^{\circ}$, 시간 해상도 5일 간격의 전지구 3차원 해양 온도 재분석장을 신속하게 생산할 수 있는 기술적 기반을 확보함.

또한, OSSE 실험을 통해 관측 자료가 제한적인 극지 해역에서의 재분석 성능을 정량적으로 평가함으로써, 개발된 자료동화 기법의 신뢰성을 검증함.

2. 2차년도

1차년도에 구축한 전지구 3차원 해양 온도 자료동화 기법을 확장하여, 전지구 3차원 해양 염분 자료동화 기법을 개발함.

염분 관측의 동화가 해양 온도 구조 및 대기 - 해양 상호작용에 미치는 영향을 고려하여, 온도 - 염분 연계 자료동화 구조를 구현함으로써 해양 초기장의 물리적 일관성을 향상시킴.

온도와 염분을 개별적으로 동화하는 방식의 한계를 극복하기 위해, 딥러닝 기반 전지구 3차원 해양 다변량 자료동화 기법을 개발함.

Receptive Field Block(RFB)을 적용하여 다양한 공간 규모의 변수 간 상호 관련성을 효과적으로 학습할 수 있도록 네트워크 구조를 설계함.

이를 통해 연산량 증가 없이 변수 간 상호작용을 고려한 다변량 자료동화가 가능함을 확인하였으며, 기존 자료동화 기법 대비 계산 효율성과 안정성을 동시에 확보함.

개발된 다변량 자료동화 시스템을 적용하여 전지구 해양 온도 및 염분 재분석장을 시범적으로 생산함.

재분석 자료와의 비교·분석을 통해 생산된 재분석장의 품질을 검증하였으며, 전지구 규모 적용 가능성을 확인함.

3. 3차년도

NCAR에서 개발한 지구시스템 모델 CESM2의 원형 소스 코드를 수집하고 극지연구소 자체 서버에 이식하여 정상 구동을 확인함.

해양 성분 모형 POPv2의 구조를 분석하고, 3차원 해양 온도 및 염분 너징 시스템을 구축함.

딥러닝 기반 자료동화를 통해 생산된 재분석장을 CESM2 모형 내 입력 자료로 연계하였으며, 이상 실험을 통해 너징 시스템의 안정성을 검증함.

완화시간규모에 대한 민감도 실험을 수행하여 최적 너징 강도를 도출함. 이를 바탕으로 20년 이상 장기간 전지구 3차원 해양 온도 및 염분 초기장을 생산하고, CESM2 모형 내 초기 입력 안정성을 확인함.

2절 최종 연구목표 달성도

본 연구에서는 딥러닝 기반 전 지구 3차원 해양 온도·염분 자료동화 시스템을 단계적으로 개발하고, 이를 CESM2 해양 성분 모형에 적용 가능한 3차원 너징 시스템과 연계함으로써 최종 연구목표를 달성함.

특히, 장기간(20년 이상) 전 지구 해양 초기장을 안정적으로 생산하고, CESM2 모형 내 초기 입력 안정성을 검증함으로써, 실제 예측 시스템으로의 적용 가능성을 확인함.

이를 통해 본 연구에서 개발한 해양 초기화 기술은 총괄 과제의 지구시스템 모델 기반 예측 시스템 구축을 직접적으로 지원할 수 있는 수준의 기술적 완성도를 확보함.

제 5 장 연구개발결과의 활용계획

1절 학술적 활용계획

본 연구에서 개발된 딥러닝 기반 전 지구 3차원 해양 초기화 시스템은 해양 온도 및 염분의 정밀한 재분석장을 생성할 수 있는 기반 기술로, 향후 극지 및 중위도 해역의 기후 변화 메커니즘 분석에 활용할 수 있음.

생산된 재분석장은 계절 및 연간 규모의 해양 변동성을 정량화하는 데 활용되며, 극지 해양 환경 변화의 원인을 규명하고, 중위도와의 상호작용 분석 연구에 기반 자료로 제공될 수 있음.

본 시스템을 계절 예측 모델의 초기장 생성 과정에 적용함으로써, 예측 정확도 향상에 기여할 수 있음. 해수면 온도, 혼합층 깊이, 해빙 면적 등 주요 해양 예측 인자의 동화 정확도 향상에 따른 응용 연구로 확장될 수 있음

2절 정책적·산업적 활용계획

본 연구의 해양 초기화 기술은 극지 해역의 해양 및 해빙 상태를 정밀하게 예측할 수 있는 기반으로 활용되며, 특히 북극 항로 예측 모델에 적용함으로써 항로 운영의 안정성과 경제성 확보에 기여할 수 있음.

생산된 재분석장은 해양 기후 관련 정책 수립에 필요한 기초 자료로 제공될 수 있으며, 장기 해양 환경 변화 분석을 통해 국제 협력 및 기후 대응 전략 수립에 활용될 수 있음.

또한, 본 기술은 해양 실측 자료의 활용 범위를 확대하여 관측망의 운영 효율성을 높이는 데 기여할 수 있으며, 실시간 예측 시스템과의 연계를 통해 산업 현장 및 국가 기상재해 대응 체계에도 직접 활용이 가능함.

북극-중위도 간 기후 연결성과 관련된 연구 확장을 통해 대국민 기후 인식 제고 및 교육 자료로도 활용할 수 있는 기반을 제공할 수 있음

3절 추가연구 필요성

본 연구에서 구축한 딥러닝 기반 전 지구 3차원 해양 초기화 시스템은 해양 온도와 염분을 중심으로 설계되었으나, 해양 순환 및 기후 변동성을 보다 정밀하게 재현하기 위해서는 해류, 해수면 높이 등 추가 물리 변수를 포함한 다변량 확장이 필요함.

특히 극지 해역에서는 관측 자료의 시·공간적 결손이 크기 때문에, 관측 오차 특성에 대한 체계적인 분석과 이를 반영한 자료동화 기법의 고도화가 요구됨.

또한, 본 연구에서 적용한 딥러닝 기반 초기화 기법의 장기 안정성과 계절·연간 예측 성능에 대한 검증을 위해, 다양한 해양·기후 시나리오를 대상으로 한 장기 재분석 및 예측 실험이 추가적으로 수행될 필요가 있음.

4절 응용가능성 및 추진방안

본 연구에서 개발된 딥러닝 기반 전 지구 3차원 해양 초기화 시스템은 전 지구 해양 온도 및 염분 재분석장을 안정적으로 생산할 수 있는 기반 기술로서, 다양한 해양·기후 연구 및 예측 분야로의 응용이 가능함.

특히 극지 및 중위도 해역의 해양 변동성 분석, 북극-중위도 간 기후 연결성 연구, 해빙-해양 상호작용 규명 등 기초·응용 연구 전반에 활용될 수 있음.

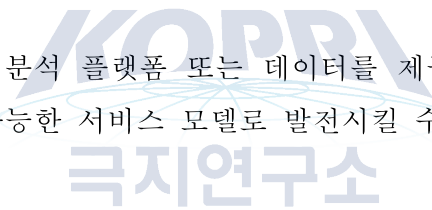
또한 본 시스템은 계절 예측 및 중장기 기후 예측 모델의 초기장 생성 과정에 적용 가능하며, 해수면 온도, 혼합층 깊이, 해빙 면적 등 주요 예측 인자의 동화 정확도 향상을 통해 예측 성능 개선에 기여할 수 있음.

이를 통해 기존 수치모델 기반 예측 연구와의 결합을 통한 융합 연구로 확장될 수 있음.

북극 항로 예측, 해빙 위험도 평가, 해양 환경 변화 모니터링 등 정책·산업 수요가 높은 분야를 중심으로 공공 및 민간 부문에서의 활용이 기대됨.

향후에는 공공기관 및 연구기관과의 협력을 통해 시범 적용을 수행하고, 단계적으로 민간 해양·물류·에너지 분야를 대상으로 한 맞춤형 해양 정보 제공 서비스로 확장할 수 있음.

아울러, 클라우드 기반 분석 플랫폼 또는 데이터를 제공함으로써, 산업 현장에서 실시간 의사결정 지원이 가능한 서비스 모델로 발전시킬 수 있음.



제 6 장 연구개발과정에서 수집한 해외과학기술정보

본 연구를 수행하는 과정에서 전 지구 해양 자료동화, 해양 재분석, 딥러닝 기반 예측 및 초기화 기술과 관련된 해외 과학기술 정보를 조사·분석하였음.

특히 유럽 및 미국을 중심으로 수행되고 있는 해양 재분석 시스템과 기후 예측 모델 개발 동향을 파악하고, 이를 본 연구의 모델 설계 및 실험 구성에 참고함.

구체적으로는 ECMWF, NOAA, NASA 등 해외 주요 기상·해양 연구기관에서 공개한 해양 재분석 자료 및 기술 문서를 검토하였으며, 딥러닝 기반 자료동화 및 초기화 기법과 관련된 최신 해외 학술 논문을 분석하여 연구 방법론 수립에 반영함.

이를 통해 해양 온도 및 염분을 중심으로 한 초기화 전략과, 전 지구 규모 적용을 위한 모델 구조 설계 방향을 설정함.

또한, 계절 예측 및 중장기 기후 예측 분야에서 활용되고 있는 해외 해양 - 대기 결합 모델의 적용 사례를 조사하여, 본 연구 결과의 향후 확장 가능성과 응용 방향을 검토함.

이러한 해외 과학기술정보 분석을 통해 본 연구의 기술적 타당성을 확보하고, 국제 연구 동향과의 정합성을 유지함.



제 7 장 참고문헌

Kim, Y. H., Hwang, C., & Choi, B. J. (2015). An assessment of ocean climate reanalysis by the data assimilation system of KIOST from 1947 to 2012. *Ocean Modelling*, 91, 1-22.

Ham, Y. G., Joo, Y. S., Kim, J. H., & Lee, J. G. (2024). Partial-convolution-implemented generative adversarial network for global oceanic data assimilation. *Nature Machine Intelligence*, 6(7), 834-843.

Kadow, C., Hall, D. M., & Ulbrich, U. (2020). Artificial intelligence reconstructs missing climate information. *Nature Geoscience*, 13(6), 408-413.

Chang, Y. L., Liu, Z. Y., Lee, K. Y., & Hsu, W. (2019). Free-form video inpainting with 3d gated convolution and temporal patchgan. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 9066-9075).



뒷 면

주 의

1. 이 보고서는 극지연구소 위탁과제 연구결과보고서입니다.
2. 이 보고서 내용을 발표할 때에는 반드시 극지연구소에서 위탁연구과제로 수행한 연구결과임을 밝혀야 합니다.
3. 국가과학기술 기밀유지에 필요한 내용은 대외적으로 발표 또는 공개하여서는 안됩니다.