



(19) 대한민국특허청(KR)

(12) 등록특허공보(B1)

(45) 공고일자 2023년12월01일

(11) 등록번호 10-2608357

(24) 등록일자 2023년11월27일

(51) 국제특허분류(Int. Cl.)

G01W 1/10 (2006.01) G01W 1/02 (2022.01)

G01W 1/08 (2006.01) G01W 1/18 (2006.01)

G06N 20/00 (2019.01) G06Q 50/26 (2012.01)

(52) CPC특허분류

G01W 1/10 (2013.01)

G01W 1/02 (2022.01)

(21) 출원번호 10-2022-0127092

(22) 출원일자 2022년10월05일

심사청구일자 2022년10월05일

(56) 선행기술조사문헌

KR101219508 B1

(뒷면에 계속)

(73) 특허권자

울산과학기술원

울산광역시 울주군 언양읍 유니스트길 50

(72) 발명자

임정호

울산광역시 울주군 언양읍 유니스트길 50 울산과학기술원

장은나

울산광역시 울주군 언양읍 유니스트길 50 울산과학기술원

성대준

울산광역시 울주군 언양읍 유니스트길 50 울산과학기술원

(74) 대리인

특허법인지원

전체 청구항 수 : 총 4 항

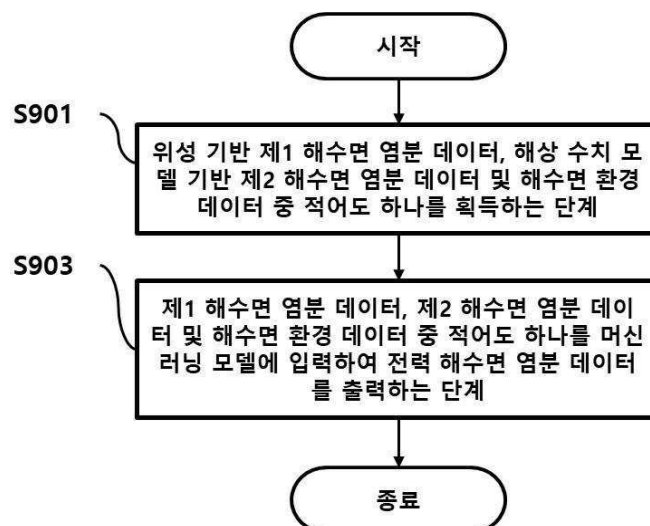
심사관 : 이민우

(54) 발명의 명칭 위성 기반 해수면 염분 데이터와 해양 수치 모델 기반 해수면 염분 데이터를 머신러닝 모델에 적용하여 해수면 염분을 예측하기 위한 방법 및 장치

(57) 요약

본 발명은 위성 기반 해수면 염분 데이터와 해양 수치 모델 기반 해수면 염분 데이터를 머신러닝 모델에 적용하여 해수면 염분을 예측하기 위한 방법 및 장치에 관한 것이다. 본 발명의 일 실시예에 따른 해수면 염분을 예측하기 위한 방법은, (a) 위성 기반 제1 해수면 염분 데이터, 해양 수치 모델 기반 제2 해수면 염분 데이터 및 해수면 환경 데이터 중 적어도 하나를 획득하는 단계; 및 (b) 상기 획득된 제1 해수면 염분 데이터, 제2 해수면 염분 데이터 및 해수면 환경 데이터 중 적어도 하나를 머신러닝 모델에 입력하여 전역(global) 해수면 염분 데이터를 출력하는 단계;를 포함할 수 있다.

대표도 - 도9



(52) CPC특허분류

G01W 1/08 (2013.01)
G01W 1/18 (2013.01)
G06N 20/00 (2021.08)
G06Q 50/26 (2013.01)
G01W 2201/00 (2013.01)

(56) 선행기술조사문헌

KR102437344 B1
KR102492075 B1
Dae-won Kim 외 3명, Estimation of Hourly Sea Surface Salinity in the East China Sea Using Geostationary Ocean Color Imager Measurements, Remote Sensing, 2020.02.25., pp1-19. 1부.*
N. Reul 외 29명, Sea surface salinity estimates from spaceborne L-band radiometers: An overview of the first decade of observation (2010-2019), Remote Sensing of Environment, 2020.06.01., pp1-37. 1부.*
Nicolas Guillou 외 2명, Predicting sea surface salinity in a tidal estuary with machine learning, Oceanologia, 2022.08.10, ISSN 0078-3234, pp1-15. 1부.*
Niels Skou, Faraday rotation and L band oceanographic measurements, RADIO SCIENCE, 2003.5.21., VOL.38, NO.4, pp1-8. 1부.*

*는 심사관에 의하여 인용된 문헌

이 발명을 지원한 국가연구개발사업

과제고유번호	1711160096
과제번호	2018-0-01424-005
부처명	과학기술정보통신부
과제관리(전문)기관명	정보통신기획평가원
연구사업명	정보통신방송혁신인재양성(R&D)
연구과제명	위성정보 처리 및 융합 서비스 기술 개발
기 여 율	1/1
과제수행기관명	아주대학교 산학협력단
연구기간	2018.06.01 ~ 2023.12.31
공지예외적용 : 있음	

명세서

청구범위

청구항 1

(a) 위성 기반 제1 해수면 염분 데이터, 해양 수치 모델 기반 제2 해수면 염분 데이터, 및 해수면 환경 데이터를 수신하는 단계; 및

(b) 상기 획득된 제1 해수면 염분 데이터, 제2 해수면 염분 데이터, 및 해수면 환경 데이터를 머신러닝 모델에 입력하여 전역(global) 해수면 염분 데이터를 출력하는 단계;

를 포함하고,

상기 제1 해수면 염분 데이터는, 상기 위성으로부터 획득된 해수면에 대한 마이크로파의 밝기 온도, 상기 위성 에 의해 직접 관찰된 마이크로파의 수평 편광 및 수직 편광에 대한 밝기 온도, 상기 수평 편광 및 수직 편광 각각에 대한 밝기 온도의 비율(ratio) 및 토지 부분(land fraction)에 기반하여 산출되고,

상기 토지 부분은, 해상에서의 해수면 염분 데이터에서 육상 지역으로 인한 노이즈의 포함 비율을 포함하고,

상기 제2 해수면 염분 데이터는, 해수면에 대한 센서로부터 측정된 현장 염분 데이터를 상기 해양 수치 모델에 적용하여 산출된 해수면 염분 데이터를 포함하고,

상기 해수면 환경 데이터는, 육지로부터의 거리, 해수면 온도 데이터, 풍속 데이터, 강수 데이터 및 파고 데이터를 포함하는,

해수면 염분을 예측하기 위한 방법.

청구항 2

삭제

청구항 3

삭제

청구항 4

삭제

청구항 5

제1항에 있어서,

상기 (b) 단계 이전에,

제1 해수면 염분 데이터, 제2 해수면 염분 데이터, 해수면 환경 데이터 및 해수면에 대한 센서로부터 측정된 현장 염분 데이터를 이용하여 상기 머신러닝 모델을 학습시키는 단계;

를 더 포함하는,

해수면 염분을 예측하기 위한 방법.

청구항 6

위성 기반 제1 해수면 염분 데이터, 해양 수치 모델 기반 제2 해수면 염분 데이터 및 해수면 환경 데이터를 수신하는 통신부; 및

상기 획득된 제1 해수면 염분 데이터, 제2 해수면 염분 데이터 및 해수면 환경 데이터를 머신러닝 모델에 입력하여 전역(global) 해수면 염분 데이터를 출력하는 제어부;

를 포함하고,

상기 제1 해수면 염분 데이터는, 상기 위성으로부터 획득된 해수면에 대한 마이크로파의 밝기 온도, 상기 위성 에 의해 직접 관찰된 마이크로파의 수평 편광 및 수직 편광에 대한 밝기 온도, 상기 수평 편광 및 수직 편광 각각에 대한 밝기 온도의 비율(ratio) 및 토지 부분(land fraction)에 기반하여 산출되고,

상기 토지 부분은, 해상에서의 해수면 염분 데이터에서 육상 지역으로 인한 노이즈의 포함 비율을 포함하고,

상기 제2 해수면 염분 데이터는, 해수면에 대한 센서로부터 측정된 현장 염분 데이터를 상기 해양 수치 모델에 적용하여 산출된 해수면 염분 데이터를 포함하고,

상기 해수면 환경 데이터는, 육지로부터의 거리, 해수면 온도 데이터, 풍속 데이터, 강수 데이터 및 파고 데이터를 포함하는,

해수면 염분을 예측하기 위한 장치.

청구항 7

삭제

청구항 8

삭제

청구항 9

삭제

청구항 10

제6항에 있어서,

상기 제어부는,

제1 해수면 염분 데이터, 제2 해수면 염분 데이터, 해수면 환경 데이터 및 해수면에 대한 센서로부터 측정된 현장 염분 데이터를 이용하여 상기 머신러닝 모델을 학습시키는,

해수면 염분을 예측하기 위한 장치.

발명의 설명

기술 분야

[0001] 본 발명은 해수면 염분을 예측하기 위한 방법 및 장치에 관한 것으로, 더욱 상세하게는 위성 기반 해수면 염분 데이터와 해양 수치 모델 기반 해수면 염분 데이터를 머신러닝 모델에 적용하여 해수면 염분을 예측하기 위한 방법 및 장치에 관한 것이다.

배경 기술

[0003] 해수면 염분(Sea surface salinity, SSS)은 해양 역학(지구 물 순환 및 해양 순환) 및 대기-해양 상호 작용의 변동성에 대한 정보를 제공하여 지구 기후의 중요한 변화를 식별하고 예측한다.

[0004] 위성 관측을 통한 전역(global) SSS 모니터링은 L-대역 마이크로파 복사계를 사용하여 가능했으나, 무선 주파수 간섭, 낮은 해수면 온도 및 강풍과 같은 조건에서 산출 알고리즘(retrieval algorithm)에 의해 성능이 제한된다.

[0005] 그러나 L-밴드 마이크로파 센서 산출 SSS 프로덕트에는 몇 가지 제한 사항이 있다.

- [0006] 인공위성 산출 SSS의 불확실성은 인위적인 무선 주파수 간섭(radio frequency interference, RFI) 및 육지-바다 오염(land-sea contamination, LSC)으로 인해 오류가 증가하기 때문에 해안 지역에서 상대적으로 높다. SSS에 대한 Tb의 민감도 감소는 더 낮은 SST에서 오류를 증가시킨다.
- [0007] 또한, 강한 풍속과 집중호우가 해수면 거칠기에 영향을 미쳐 산출 오류를 더욱 증가시킨다.
- [0008] 이에, 전역 해수면 염분을 산출하기 위한 많은 연구가 진행되고 있으나 미흡한 실정이다.

선행기술문헌

특허문헌

- [0010] (특허문헌 0001) [특허문헌 1] 한국공개특허 제10-2021-0085030호

발명의 내용

해결하려는 과제

- [0011] 본 발명은 전술한 문제점을 해결하기 위하여 창출된 것으로, 위성 기반 해수면 염분 데이터와 해양 수치 모델 기반 해수면 염분 데이터를 머신러닝 모델에 적용하여 전역(global) 해수면 염분을 예측하기 위한 방법 및 장치를 제공하는 것을 그 목적으로 한다.
- [0012] 본 발명의 목적들은 이상에서 언급한 목적들로 제한되지 않으며, 언급되지 않은 또 다른 목적들은 아래의 기재로부터 명확하게 이해될 수 있을 것이다.

과제의 해결 수단

- [0014] 상기한 목적들을 달성하기 위하여, 본 발명의 일 실시예에 따른 위성 기반 해수면 염분 데이터와 해양 수치 모델 기반 해수면 염분 데이터를 머신러닝 모델에 적용하여 해수면 염분을 예측하기 위한 방법은, (a) 위성 기반 제1 해수면 염분 데이터, 해양 수치 모델 기반 제2 해수면 염분 데이터 및 해수면 환경 데이터 중 적어도 하나를 수신하는 단계; 및 (b) 상기 획득된 제1 해수면 염분 데이터, 제2 해수면 염분 데이터 및 해수면 환경 데이터 중 적어도 하나를 머신러닝 모델에 입력하여 전역(global) 해수면 염분 데이터를 출력하는 단계;를 포함할 수 있다.
- [0015] 실시예에서, 상기 제1 해수면 염분 데이터는, 상기 위성으로부터 획득된 해수면에 대한 마이크로파의 밝기 온도(brightness temperature) 데이터에 기반하여 산출된 해수면 염분 데이터를 포함할 수 있다.
- [0016] 실시예에서, 상기 제2 해수면 염분 데이터는, 해수면에 대한 센서로부터 측정된 현장 염분 데이터를 상기 해양 수치 모델에 적용하여 산출된 해수면 염분 데이터를 포함할 수 있다.
- [0017] 실시예에서, 상기 해수면 환경 데이터는, 해수면 온도 데이터, 풍속 데이터, 강수 데이터 및 파고 데이터 중 적어도 하나를 포함할 수 있다.
- [0018] 실시예에서, 상기 해수면 염분을 예측하기 위한 방법은, 상기 (b) 단계 이전에, 상기 제1 해수면 염분 데이터, 제2 해수면 염분 데이터, 해수면 환경 데이터 및 해수면에 대한 센서로부터 측정된 현장 염분 데이터를 이용하여 상기 머신러닝 모델을 학습시키는 단계;를 더 포함할 수 있다.
- [0019] 실시예에서, 해수면 염분을 예측하기 위한 장치는, 위성 기반 제1 해수면 염분 데이터, 해양 수치 모델 기반 제2 해수면 염분 데이터 및 해수면 환경 데이터 중 적어도 하나를 수신하는 통신부; 및 상기 획득된 제1 해수면 염분 데이터, 제2 해수면 염분 데이터 및 해수면 환경 데이터 중 적어도 하나를 머신러닝 모델에 입력하여 전역(global) 해수면 염분 데이터를 출력하는 제어부;를 포함할 수 있다.
- [0020] 실시예에서, 상기 제1 해수면 염분 데이터는, 상기 위성으로부터 획득된 해수면에 대한 마이크로파의 밝기 온도(brightness temperature) 데이터에 기반하여 산출된 해수면 염분 데이터를 포함할 수 있다.
- [0021] 실시예에서, 상기 제2 해수면 염분 데이터는, 해수면에 대한 센서로부터 측정된 현장 염분 데이터를 상기 해양 수치 모델에 적용하여 산출된 해수면 염분 데이터를 포함할 수 있다.

- [0022] 실시예에서, 상기 해수면 환경 데이터는, 해수면 온도 데이터, 풍속 데이터, 강수 데이터 및 파고 데이터 중 적어도 하나를 포함할 수 있다.
- [0023] 실시예에서, 상기 제어부는, 상기 제1 해수면 염분 데이터, 제2 해수면 염분 데이터, 해수면 환경 데이터 및 해수면에 대한 센서로부터 측정된 현장 염분 데이터를 이용하여 상기 머신러닝 모델을 학습시킬 수 있다.
- [0024] 상기한 목적들을 달성하기 위한 구체적인 사항들은 첨부된 도면과 함께 상세하게 후술될 실시예들을 참조하면 명확해질 것이다.
- [0025] 그러나, 본 발명은 이하에서 개시되는 실시예들에 한정되는 것이 아니라, 서로 다른 다양한 형태로 구성될 수 있으며, 본 발명의 개시가 완전하도록 하고 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자(이하, "통상의 기술자")에게 발명의 범주를 완전하게 알려주기 위해서 제공되는 것이다.

발명의 효과

- [0027] 본 발명의 일 실시예에 의하면, 모든 지역에서 균일하게 개선된 글로벌 SSS를 산출할 수 있다.
- [0028] 본 발명의 효과들은 상술된 효과들로 제한되지 않으며, 본 발명의 기술적 특징들에 의하여 기대되는 잠정적인 효과들은 아래의 기재로부터 명확하게 이해될 수 있을 것이다.

도면의 간단한 설명

- [0030] 도 1a는 본 발명의 일 실시예에 따른 아르고 기반 현장 SSS 관찰의 평균 개수를 도시한 도면이다.
- 도 1b는 본 발명의 일 실시예에 따른 TSG 기반 현장 SSS 관찰의 평균 개수를 도시한 도면이다.
- 도 2a는 본 발명의 일 실시예에 따른 위성 기반 해수면 염분 데이터와 해양 수치 모델 기반 해수면 염분 데이터를 머신러닝 모델에 적용하여 해수면 염분을 예측하기 위한 과정을 도시한 도면이다.
- 도 2b는 본 발명의 일 실시예에 따른 입력 변수의 예를 도시한 도면이다.
- 도 3a는 본 발명의 일 실시예에 따른 위성 기반 제1 해수면 염분 데이터의 분포 그래프를 도시한 도면이다.
- 도 3b는 본 발명의 일 실시예에 따른 해양 수치 모델 기반 제2 해수면 염분 데이터의 분포 그래프를 도시한 도면이다.
- 도 3c는 본 발명의 일 실시예에 따른 위성 기반 제1 해수면 염분 데이터의 전체 데이터 분포를 도시한 도면이다.
- 도 3d는 본 발명의 일 실시예에 따른 해양 수치 모델 기반 제2 해수면 염분 데이터의 전체 데이터 분포를 도시한 도면이다.
- 도 3e는 본 발명의 일 실시예에 따른 위성 기반 제1 해수면 염분 데이터와 현장 염분 데이터의 차이에 대한 공간 분포를 도시한 도면이다.
- 도 3f는 본 발명의 일 실시예에 따른 해양 수치 모델 기반 제2 해수면 염분 데이터와 현장 염분 데이터의 차이에 대한 공간 분포를 도시한 도면이다.
- 도 3g는 본 발명의 일 실시예에 따른 각 해양 지역 별 SMAP SSS와 HYCOM SSS의 RMSD 및 MB를 도시한 도면이다.
- 도 4a는 본 발명의 일 실시예에 따른 GBRT 기반 해수면 염분 데이터의 공간 분포를 도시한 도면이다.
- 도 4b는 본 발명의 일 실시예에 따른 위성 기반 제1 해수면 염분 데이터의 분포 그래프를 도시한 도면이다.
- 도 4c는 본 발명의 일 실시예에 따른 위성 기반 제1 해수면 염분 데이터의 공간 분포를 도시한 도면이다.
- 도 4d는 본 발명의 일 실시예에 따른 해양 수치 모델 기반 제2 해수면 염분 데이터의 분포 그래프를 도시한 도면이다.
- 도 4e는 본 발명의 일 실시예에 따른 해양 수치 모델 기반 제2 해수면 염분 데이터의 공간 분포를 도시한 도면이다.
- 도 4f는 본 발명의 일 실시예에 따른 GBRT 기반 해수면 염분 데이터의 분포 그래프를 도시한 도면이다.

도 4g는 본 발명의 일 실시예에 따른 GBRT 기반 해수면 염분 데이터의 공간 분포를 도시한 도면이다.

도 4h는 본 발명의 일 실시예에 따른 머신러닝 모델의 훈련 및 검증 결과를 도시한 도면이다.

도 5a는 본 발명의 일 실시예에 따른 평균 GBRT 기반 해수면 염분 데이터의 공간 분포를 도시한 도면이다.

도 5b는 본 발명의 일 실시예에 따른 평균 위성 기반 제1 해수면 염분 데이터의 공간 분포를 도시한 도면이다.

도 5c는 본 발명의 일 실시예에 따른 GBRT 기반 해수면 염분 데이터와 위성 기반 제1 해수면 염분 데이터의 차이를 도시한 도면이다.

도 5d는 본 발명의 일 실시예에 따른 평균 해양 수치 모델 기반 제2 해수면 염분 데이터의 공간 분포를 도시한 도면이다.

도 5e는 본 발명의 일 실시예에 따른 GBRT 기반 해수면 염분 데이터와 해양 수치 모델 기반 제2 해수면 염분 데이터의 차이를 도시한 도면이다.

도 5f는 본 발명의 일 실시예에 따른 위성 기반 제1 해수면 염분 데이터와 해양 수치 모델 기반 제2 해수면 염분 데이터의 차이를 도시한 도면이다.

도 5g는 본 발명의 일 실시예에 따른 각 해양 지역에 대한 현장 염분 데이터의 RMSD 및 MB를 도시한 도면이다.

도 6a 내지 6i는 본 발명의 일 실시예에 따른 현장 TSG 데이터, 위성 기반 제1 해수면 염분 데이터, 해양 수치 모델 기반 제2 해수면 염분 데이터 및 GBRT 기반 해수면 염분 데이터의 시간 플롯 및 공간 분포를 도시한 도면이다.

도 7은 본 발명의 일 실시예에 따른 GBRT 모델을 사용하여 식별된 특징 중요도를 도시한 도면이다.

도 8a는 본 발명의 일 실시예에 따른 GBRT 모델에 대한 SHAP 값의 플롯을 도시한 도면이다.

도 8b는 본 발명의 일 실시예에 따른 ARP 지역의 플롯을 도시한 도면이다.

도 8c는 본 발명의 일 실시예에 따른 위성 기반 제1 해수면 염분 데이터의 SHAP 종속 플롯을 도시한 도면이다.

도 8d는 본 발명의 일 실시예에 따른 해양 수치 모델 기반 제2 해수면 염분 데이터의 SHAP 종속 플롯을 도시한 도면이다.

도 9는 본 발명의 일 실시예에 따른 위성 기반 해수면 염분 데이터와 해양 수치 모델 기반 해수면 염분 데이터를 머신러닝 모델에 적용하여 해수면 염분을 예측하기 위한 방법을 도시한 도면이다.

도 10은 본 발명의 일 실시예에 따른 위성 기반 해수면 염분 데이터와 해양 수치 모델 기반 해수면 염분 데이터를 머신러닝 모델에 적용하여 해수면 염분을 예측하기 위한 장치를 도시한 도면이다.

발명을 실시하기 위한 구체적인 내용

- [0031] 본 발명은 다양한 변경을 가할 수 있고, 여러 가지 실시예들을 가질 수 있는 바, 특정 실시예들을 도면에 예시하고 이를 상세히 설명하고자 한다.
- [0032] 청구범위에 개시된 발명의 다양한 특징들은 도면 및 상세한 설명을 고려하여 더 잘 이해될 수 있을 것이다. 명세서에 개시된 장치, 방법, 제법 및 다양한 실시예들은 예시를 위해서 제공되는 것이다. 개시된 구조 및 기능상의 특징들은 통상의 기술자로 하여금 다양한 실시예들을 구체적으로 실시할 수 있도록 하기 위한 것이고, 발명의 범위를 제한하기 위한 것이 아니다. 개시된 용어 및 문장들은 개시된 발명의 다양한 특징들을 이해하기 쉽게 설명하기 위한 것이고, 발명의 범위를 제한하기 위한 것이 아니다.
- [0033] 본 발명을 설명함에 있어서, 관련된 공지기술에 대한 구체적인 설명이 본 발명의 요지를 불필요하게 흐릴 수 있다고 판단되는 경우, 그 상세한 설명을 생략한다.
- [0034] 이하, 본 발명의 일 실시예에 따른 위성 기반 해수면 염분 데이터와 해양 수치 모델 기반 해수면 염분 데이터를 머신러닝 모델에 적용하여 해수면 염분을 예측하기 위한 방법 및 장치를 설명한다.
- [0036] 도 1a는 본 발명의 일 실시예에 따른 아르고 기반 현장 SSS 관찰의 평균 개수를 도시한 도면이다. 도 1b는 본 발명의 일 실시예에 따른 TSG 기반 현장 SSS 관찰의 평균 개수를 도시한 도면이다.

- [0037] 도 1a 및 1b를 참고하면, 본 발명의 경우, K-최근접 이웃(K-nearest neighbor), 지원 벡터 회귀(support vector regression), 인공 신경망, 랜덤 포레스트(random forest), 익스트림 그래디언트 부스팅, 라이트 그래디언트 부스팅 모델 및 그래디언트 부스트 회귀 트리(gradient boosted regression tree, GBRT)의 7가지 머신러닝 접근 방식을 기반으로 다중 소스 데이터를 사용하여 글로벌(global) 해수면 염분(Sea surface salinity, SSS) 데이터를 산출할 수 있다.
- [0038] 여기서, 글로벌 해수면 염분 데이터는 전역 해수면 염분 데이터, 글로벌 SSS 데이터 또는 이와 동등한 기술적 의미를 갖는 용어로 지칭될 수 있다.
- [0039] 일 실시예에서, 위성 기반 제1 해수면 염분 데이터, 해양 수치 모델 기반 제2 해수면 염분 데이터 및 해수면 환경 데이터 중 적어도 하나를 머신러닝 모델에 입력하여 전역(global) 해수면 염분 데이터를 출력할 수 있다.
- [0040] 예를 들어, 위성 기반 제1 해수면 염분 데이터는 SMAP(Soil Moisture Active Passive) 프로덕트(product), SMAP SSS, SMAP 위성 데이터 또는 이와 동등한 기술적 의미를 갖는 용어로 지칭될 수 있다. 예를 들어, 위성 기반 제1 해수면 염분 데이터는 위성으로부터 획득된 해수면에 대한 마이크로파의 밝기 온도(brightness temperature) 데이터에 기반하여 산출된 연속적인 해수면 염분 데이터를 2차원 이미지로 표현한 형태로 구현될 수 있다.
- [0041] 또한, 해양 수치 모델 기반 제2 해수면 염분 데이터는 HYCOM(Hybrid Coordinate Ocean Model) 프로덕트, HYCOM SSS 또는 이와 동등한 기술적 의미를 갖는 용어로 지칭될 수 있다. 예를 들어, 해양 수치 모델 기반 제2 해수면 염분 데이터는 전역 공간에 대한 연속적인 해수면 염분 데이터를 2차원 이미지로 표현한 형태로 구현될 수 있다.
- [0042] 또한, 해수면 환경 데이터는 보조 데이터 또는 이와 동등한 기술적 의미를 갖는 용어로 지칭될 수 있다.
- [0043] 머신러닝 모델 중 최고 성능의 GBRT 모델은 SMAP SSS 프로덕트와 비교하여 검증 데이터 세트에 대한 제공 평균 제공근 차이를 1.062에서 0.259 실제 염분 단위로 줄일 수 있다.
- [0044] 본 발명에 따라 산출된 전역 해수면 염분 데이터는 전세계 해양의 저염분 및 고염분 해역 모두에 대한 현장 염분 데이터와의 상관관계를 증가시켰고, 따라서 전지구 SSS에 대한 모니터링을 효과적으로 수행하기 위해 사용될 수 있다.
- [0045] SSS에 대한 종래의 방식은 로컬 또는 지역 영역에만 집중하여 일반화를 제한했지만, 현장 염분 데이터를 사용하여 SMAP SSS 프로덕트 및 HYCOM SSS 프로덕트를 검증 및 분석하고, 머신러닝 모델을 기반으로 SMAP 및 HYCOM SSS 프로덕트를 사용하여 글로벌 SSS 데이터를 생성하고, 입력 변수의 중요성과 상호 작용을 분석할 수 있다.
- [0046] 일 실시예에서, 현장(In situ) 염분 데이터가 사용될 수 있으며, 현장 염분 데이터는 해수면에 위치한 센서(예: 부이(buoy))로부터 측정된 센서가 위치한 현장(예: 지점(point))의 해수면 염분 데이터를 의미할 수 있다. 여기서, 현장 염분 데이터의 해수면 염분 데이터는 해당 지점에서의 해수면 염분을 나타내는 불연속적인 점의 집합으로 구현될 수 있다.
- [0047] 예를 들어, 현장 염분 데이터는 Argo 데이터 또는 TSG(thermosalinograph) 데이터를 포함할 수 있다.
- [0048] Argo 데이터는 표류 프로파일링 플로트를 사용하여 2000m 이하의 해양 환경을 모니터링하는 글로벌 어레이를 의미할 수 있다.
- [0049] TSG 데이터는 Argo 데이터로 잘 대표되지 않는 해안 지역 근처에서 측정되었기 때문에 글로벌 모델 적용 가능성을 보장하기 위해 사용될 수 있다.
- [0050] 모든 현장 염분 데이터는 품질 관리 플래그(양호 또는 아마도 양호에 해당하는 1 또는 2)를 사용하여 스크리닝될 수 있다.
- [0051] 도 1a 및 1b를 참고하면, $\leq 5m$ 깊이의 현장 염분 데이터가 사용될 수 있다.
- [0052] 현장 염분 데이터는 SMAP 프로덕트와의 코로케이션을 기반으로 선택될 수 있다.
- [0053] 위성 기반 제1 해수면 염분 데이터와 현장 염분 데이터 간의 거리는 $< \pm 12.5km$ (25km의 SMAP 픽셀 크기 기준), $< \pm 6h$ 의 시간 차이로 유지될 수 있다.
- [0054] 도 1a 및 1b의 각 이미지의 동일한 픽셀에서 여러 번 읽지 않도록 주어진 그리드 픽셀 내에 분산된 여러 현장

염분 데이터의 평균을 산출할 수 있다.

- [0055] 일정 기간 동안 총 207,853개의 글로벌 현장 SSS 데이터 포인트(Argo floats, 171,363개; TSG, 36,490개)가 사용될 수 있다.
- [0056] 일 실시예에서, SMAP 위성 데이터가 사용될 수 있다. 일 실시예에서, 위성에 의해 직접 관찰된 마이크로파의 각 복사계 편파(H-편광 및 V-편광)에 대한 밝기 온도(Tb), 토지 부분(land fraction) 및 유도된 SSS 추정치 및 보조 데이터는 위성 기반 제1 해수면 염분 데이터를 생성하기 위해 사용될 수 있다.
- [0057] 일 실시예에서, 토지 부분은 해상에서의 해수면 염분 데이터에서 육상 지역으로 인한 노이즈의 포함 비율을 의미할 수 있다.
- [0058] 일 실시예에서, SMAP SSS에 대해 품질 플래그가 0('SSS 사용 가능')인 픽셀만 사용될 수 있다.
- [0059] 일 실시예에서, HYCOM 데이터가 사용될 수 있다. 해양 수치 모델 기반 제2 해수면 염분 데이터는 일련의 원시 방정식(primitive equation)을 기반으로 하는 해양 일반 순환 모델에 기반하여 생성될 수 있다.
- [0060] 일 실시예에서, 해수면에 대한 여러 현장 염분 데이터(즉, 현장 온도 및 염분 프로파일, 원격으로 감지된 SST 및 고도계 해수면 높이 이상)를 동화(assimilate)하여 해양 수치 모델 기반 제2 해수면 염분 데이터를 생성할 수 있다.
- [0061] 예를 들어, Argo 데이터(현장 염분)은 HYCOM SSS의 동화 과정에서 사용되므로, TSG 데이터는 본 발명에 따른 방식을 평가하기 위한 독립적인 검증 데이터 소스로 사용될 수 있다.
- [0062] 예를 들어, 해양 수치 모델 기반 제2 해수면 염분 데이터는 1/12° 해상도의 3h SSS 프로덕트를 포함할 수 있다.
- [0063] 일 실시예에서, SMAP 프로덕트 분해능(예: 25km)에 따라 리샘플링된 HYCOM SSS 추정치가 사용될 수 있다.
- [0064] 일 실시예에서, 보조 데이터가 사용될 수 있다. 예를 들어, 보조 데이터는 NOAA(National Oceanic and Atmospheric Administration) OISST(Optimal Interpolation Sea Surface Temperatures), NOAA NCEP(National Centers for Environment Prediction) GFS(Global Forecast System) 대기 풍속 모델 및 NOAA WaveWatch III SWH(significant wave height) 중 적어도 하나를 포함할 수 있다.
- [0065] 일 실시예에서, 1일 이하(6시간마다) NOAA 풍속 및 SWH 데이터는 시간에 따라 선형으로 보간되었으며, 일일 NOAA OISST 데이터는 추가 시간 보간 없이 사용될 수 있다.
- [0066] 강우 중에는 해수면의 염분이 감소하고 해수면 거칠기에도 영향을 미치며, 이 두 가지 모두 위성에서 산출한 SSS 프로덕트의 성능에 영향을 미칠 수 있다. 따라서 강수 데이터는 SSS 보정 모델에 입력될 수 있다.
- [0067] GPM 강수 데이터는 SMAP 데이터와 일치시키기 위해 쌍선형 보간법을 통해 25km로 다시 샘플링될 수 있다.
- [0068] 열대 강우 측정 미션(Tropical Rainfall Measuring Mission, TRMM) 다중 위성 강수 분석에서 25km x 25km 공간 분해능으로 제공한 육지-바다 마스크도 사용될 수 있다. 육지로부터의 거리는 TRMM 육지-바다 마스크를 사용하여 계산될 수 있다.
- [0070] 도 2a는 본 발명의 일 실시예에 따른 위성 기반 해수면 염분 데이터와 해양 수치 모델 기반 해수면 염분 데이터를 머신러닝 모델에 적용하여 해수면 염분을 예측하기 위한 과정을 도시한 도면이다. 도 2b는 본 발명의 일 실시예에 따른 입력 변수의 예를 도시한 도면이다.
- [0071] 도 2a를 참고하면, SMAP 프로덕트, HYCOM SSS 추정치 및 보조 데이터는 머신러닝 모델을 기반으로 글로벌 SSS 데이터를 생성하기 위해 사용될 수 있다.
- [0072] 머신러닝 모델에 입력되는 입력 변수는 SMAP SSS, 마이크로파의 H-편광(H-pol)의 Tb, 마이크로파의 V-편광(V-pol)의 Tb, H-pol:V-pol의 Tb 비율, 토지 비율, HYCOM SSS, OISST, NCEP 풍속, 육지로부터의 거리, GPM 강수량 및 WaveWatch III SWH(도 2a 및 2b)를 포함할 수 있다.
- [0073] 일 실시예에서, 본 발명에 따른 머신러닝 모델은 K-NN(K-Nearest Neighbor), SVR(Support Vector Regression), ANN(Artificial Neural Network), RF(random forest), XGBoost(Extreme Gradient Boosting), LGBM(Light Gradient Boosting Model) 및 GBRT(Gradient Boosted Regression Tree)를 포함할 수 있으나, 이에 제한되지 않고 다양한 모델을 포함할 수 있다.

- [0074] 일 실시예에서, 변수 및 데이터 전처리가 수행될 수 있다.
- [0075] H-pol의 Tbs, V-pol의 Tbs 및 OISST는 SSS 산출 알고리즘에서 직접 사용될 수 있다. H-pol의 Tb와 V-pol의 Tb의 민감도는 환경 정보(즉, SST, SSS, 풍속)에 따라 달라지므로, H-pol과 V-pol 사이의 Tb 비율이 사용될 수 있다.
- [0076] 실시간으로 산출되는 SMAP SSS와 달리, HYCOM SSS는 해양 수치 모델을 사용하여 될 수 있다.
- [0077] SMAP 및 HYCOM SSS 프로덕트는 시너지 효과를 낼 수 있다. Tbs는 해수면의 거칠기(roughness)에 민감하기 때문에, 풍속과 SWH 데이터가 입력 변수로 될 수 있다.
- [0078] L-band 마이크로파 센서 산출 SSS 프로덕트의 불확실성은 RFI 및 LSC로 인해 해안 지역에서 증가하기 때문에, 토지로부터의 거리 및 토지 부분도 사용될 수 있다.
- [0079] 일 실시예에서, 트리(tree) 기반 머신러닝 모델은 각 특징(feature)이 무작위로 교란될 때 모델 성능이 감소하는 정도를 계산하여 식별된 상대적 특징 중요도(feature importance)를 산출할 수 있다. 모델 성능의 현저한 감소는 큰 특징 중요도 값에 해당할 수 있다.
- [0080] 트리 기반 머신러닝 모델의 모든 하이퍼 파라미터는 그리드 산출을 사용하여 최적화될 수 있다.
- [0082] 성능 평가 및 특징 분석을 위하여, 현장 측정과 위성 데이터가 일치하는 총 207,853개의 샘플을 두 그룹으로 나누어 각 모델을 평가할 수 있다.
- [0083] 샘플의 80%는 교차 검증을 통해 하이퍼 파라미터를 최적화하기 위한 모델의 훈련 데이터로 사용되었고, 나머지 20%는 모델 성능을 독립적으로 검증하는데 사용될 수 있다.
- [0084] 훈련 및 검증 데이터 세트는 모델 시간 이동성을 평가하기 위해 날짜(DOY)(훈련의 경우 167,052개, 검증의 경우 40,801개)를 기준으로 분할될 수 있다.
- [0085] 10일마다 마지막 2일(9일과 10일)의 데이터는 검증 데이터로 사용될 수 있다(1일부터 8일까지는 보정 데이터로 사용).
- [0086] 다양한 데이터 소스가 모델 성능에 미치는 영향을 조사하기 위해 현장 염분 데이터 소스(즉, Argo 및 TSG)를 사용한 검증도 수행될 수 있다.
- [0087] 현장 측정, SMAP SSS, HYCOM SSS 및 제안된 모델의 결과 간의 차이를 평가하기 위해 세 가지 메트릭이 사용될 수 있다. 세 가지 메트릭은 <수학식 1>과 같은 RMSD(root mean square deviation), <수학식 2>와 같은 결정 계수(coefficient of determination) 및 <수학식 3>과 같은 평균 바이어스(MB, 평균 차이)를 포함할 수 있다.

수학식 1

$$\text{RMSD} = \sqrt{\frac{\sum (M - O)^2}{n}}$$

[0088]

수학식 2

$$R^2 = \left(\frac{\sum (M - \bar{M})(O - \bar{O})}{\sqrt{\sum (M - \bar{M})^2 \sum (O - \bar{O})^2}} \right)^2$$

[0089]

수학식 3

$$MB = \frac{1}{n} \sum (M - O)$$

[0090]

[0091]

여기서, O는 현장 데이터(in situ observation data), M은 SSS 프로덕트, O와 M은 각각 시간 및 공간 도메인에
서 함께 수집된 모든 관측 및 추정 SSS의 평균값을 의미할 수 있다.

[0092]

SHAP(Shapley additive explanation) 값은 모델 성능에 대한 입력 특징의 기여도와 상호 작용을 분석하는데 사
용될 수 있다.

[0093]

SHAP 값은 게임 이론에 기반한 부분 데이터 세트와 전체 데이터 세트를 사용하여 입력 특징 간의 관계를 탐색하
여 모델 복잡성에 대한 이해를 향상시킬 수 있다.

[0094]

그들은 크기와 방향(양수 또는 음수) 측면에서 모든 데이터 샘플에 대한 각 특징의 기여도 외에도 교육 데이터
(예상 값)에 대한 평균 예측을 산출할 수 있다.

[0095]

이러한 방식으로 SHAP 값은 모델에 내재된 특징 중요도 정보의 한계를 보완할 수 있다.

[0097]

도 3a는 본 발명의 일 실시예에 따른 위성 기반 제1 해수면 염분 데이터의 분포 그래프를 도시한 도면이다. 도
3b는 본 발명의 일 실시예에 따른 해양 수치 모델 기반 제2 해수면 염분 데이터의 분포 그래프를 도시한 도면이
다. 도 3c는 본 발명의 일 실시예에 따른 위성 기반 제1 해수면 염분 데이터의 전체 데이터 분포를 도시한 도면
이다. 도 3d는 본 발명의 일 실시예에 따른 해양 수치 모델 기반 제2 해수면 염분 데이터의 전체 데이터 분포를
도시한 도면이다. 도 3e는 본 발명의 일 실시예에 따른 위성 기반 제1 해수면 염분 데이터와 현장 염분 데이터
의 차이에 대한 공간 분포를 도시한 도면이다. 도 3f는 본 발명의 일 실시예에 따른 해양 수치 모델 기반 제2
해수면 염분 데이터와 현장 염분 데이터의 차이에 대한 공간 분포를 도시한 도면이다. 도 3g는 본 발명의 일 실
시예에 따른 각 해양 지역 별 SMAP SSS와 HYCOM SSS의 RMSD 및 MB를 도시한 도면이다.

[0098]

SMAP 및 HYCOM SSS 프로덕트와 현장 염분 데이터를 비교할 수 있다.

[0099]

SMAP 및 HYCOM SSS 프로덕트를 일정 기간 동안의 현장 염분 데이터 앙상블과 비교할 수 있다.

[0100]

SMAP 및 HYCOM 일일 SSS를 현장 염분 데이터의 비교(도 3a 및 3b), SMAP 및 HYCOM 일일 SSS 프로덕트의 평균
분포(도 3c 및 d) 및 SSS 프로덕트와 현장 SSS 데이터 사이의 평균 차이(도 3e 및 f)를 확인할 수 있다.

[0101]

SMAP 및 HYCOM 일간 SSS 오차 분포를 구체적으로 평가하기 위해, 도 3g와 같이 전지구 해양을 8개 영역으로 구
분할 수 있다.

[0102]

전체 해양에 대한 R2 및 RMSD 값은 현장 측정과 비교할 때 SMAP SSS의 경우 각각 0.90 및 1.060 psu, HYCOM
SSS의 경우 0.98 및 0.396 psu일 수 있다.

[0103]

SMAP SSS의 R2, RMSD 및 MB는 각각 0.50, 0.970 psu 및 0.08 psu인 반면, HYCOM SSS의 R2, RMSD 및 MB는 32
psu 이상의 고염분 지역에 대해 0.93, 0.280 psu 및 0.02 psu일 수 있다.

[0104]

그러나 10psu와 32psu 사이의 저염분 영역에서 SMAP SSS의 R2, RMSD 및 MB 값은 각각 0.52, 1.798 psu, 0.55
psu이고, HYCOM SSS에 대해 0.01, 3.050 psu, 1.85 psu일 수 있다.

[0105]

HYCOM SSS는 염분이 높은 지역에서 SMAP SSS보다 현장 염분 데이터와 더 많은 상관관계가 있는 반면, 염분이 낮
은 지역에서는 SMAP SSS에 해당할 수 있다.

[0106]

현장 염분 데이터는 HYCOM SSS에 통합되어 있기 때문에, HYCOM SSS와 현장 염분 데이터 사이에 높은 상관관계가
존재할 수 있다.

[0107]

예를 들어, HYCOM과 Argo SSS 데이터 사이의 RMSD는 0.275psu인 반면, 전자(former)와 독립 TSG 사이의 RMSD는
~0.735psu로 상대적으로 높을 수 있다.

[0108]

대조적으로, SMAP SSS 데이터는 Argo 및 TSG SSS 데이터와 유사한 상관관계를 보여 각각 0.995 psu 및 1.325
psu의 RMSD를 나타낼 수 있다.

- [0109] SMAP SSS는 극지방에서 상당한 불확실성을 초래한 반면, HYCOM SSS는 하천 지배 해안 지역(river-dominated coastal region)에서 상당한 불확실성을 나타낼 수 있다(도 3e 및 3f).
- [0110] 10 psu(1970 데이터 포인트) 미만의 현장 SSS 데이터는 스웨덴, 핀란드, 폴란드를 포함한 국가로 둘러싸인 반폐쇄 바다인 보스니아 만과 발트해에서 측정되어 RFI 및 LSC로 인한 SMAP SSS 데이터의 성능이 좋지 않을 수 있다.
- [0111] SMAP SSS와 현장 염분 데이터의 상관관계는 북인도양(NIO) 및 ARP(Amazon River Plume)과 같이 상대적으로 낮은 SSS 값을 갖는 하천 지배 지역에서 높은 반면, HYCOM SSS는 동일한 지역에서 과대평가될 수 있다(도 3f의 주황색 원 및 도 3g).
- [0112] ARP 영역의 경우 SMAP SSS의 RMSD는 0.647이고 HYCOM SSS의 RMSD는 1.258 psu일 수 있다.
- [0113] HYCOM 모델은 현장 염분 데이터와 SSS 기후학을 기반으로 SSS를 추정하고, 실시간 관측이 가능한 위성에서 파생된 SSS 프로덕트를 사용하지 않을 수 있다.
- [0114] 하천 유출수(River runoff)는 해양 수치 모델에서 각 강에 대한 월별 평균 하천 배출량 값으로 사용될 수 있다.
- [0115] 해양 수치 모델에 사용된 참조 SSS 기후학 및 하천 유입은 너무 거칠어서 좁은 지역에서 저염분 물의 급격한 변화를 포착하지 못할 수 있다.
- [0116] 도 3e 및 3g를 참고하면, SMAP SSS 데이터는 상대적으로 낮은 SST(전 지구 평균 SST가 19.69 °C인 반면, 극지방의 SST는 ARO에서 7.99 °C, SOO에서 4.20 °C 임)로 인해 극지방(북극해(Arctic Ocean, ARO) 및 남극해(Southern Ocean, SOO))에서 상당한 불확실성을 가질 수 있다.
- [0117] SST가 감소함에 따라 염분에 대한 방사율의 민감도가 감소하여 SMAP SSS의 오류가 증가할 수 있다.
- [0118] 북태평양(NOP) 지역도 낮은 SST와 RFI로 인해 SMAP SSS 데이터에 상당한 불확실성이 있다.
- [0119] 그러나 SMAP SSS와 in situ 데이터의 상관관계는 SST가 상대적으로 높은 적도 지역(열대해, TRO, 적도 동부, EEP, 태평양, PAO, NIO, ARP)에서 증가할 수 있다.
- [0120] SMAP SSS, HYCOM SSS, in situ 데이터의 차이는 위치와 환경에 따라 다르므로 실증적 머신 러닝 접근 방식을 기반으로 하는 3가지 프로덕트와 보조 데이터의 시너지 활용을 통해 고품질의 글로벌 SSS를 생산할 수 있다.
- [0122] 도 4a는 본 발명의 일 실시예에 따른 GBRT 기반 해수면 염분 데이터의 공간 분포를 도시한 도면이다. 도 4b는 본 발명의 일 실시예에 따른 위성 기반 제1 해수면 염분 데이터의 분포 그래프를 도시한 도면이다. 도 4c는 본 발명의 일 실시예에 따른 위성 기반 제1 해수면 염분 데이터의 공간 분포를 도시한 도면이다. 도 4d는 본 발명의 일 실시예에 따른 해양 수치 모델 기반 제2 해수면 염분 데이터의 분포 그래프를 도시한 도면이다. 도 4e는 본 발명의 일 실시예에 따른 해양 수치 모델 기반 제2 해수면 염분 데이터의 공간 분포를 도시한 도면이다. 도 4f는 본 발명의 일 실시예에 따른 GBRT 기반 해수면 염분 데이터의 분포 그래프를 도시한 도면이다. 도 4g는 본 발명의 일 실시예에 따른 GBRT 기반 해수면 염분 데이터의 공간 분포를 도시한 도면이다. 도 4h는 본 발명의 일 실시예에 따른 머신러닝 모델의 훈련 및 검증 결과를 도시한 도면이다.
- [0123] 도 4h를 참고하면, 시간적 홀드아웃 검증이 적용된 7개의 머신러닝 모델을 사용하여 SMAP과 HYCOM SSS 프로덕트를 시너지 효과로 생성한 글로벌 SSS를 비교한 결과를 확인할 수 있다.
- [0124] HYCOM SSS 데이터는 Argo 데이터와 독립적이지 않을 수 있다. 따라서 본 발명에 따른 모델은 HYCOM SSS와 독립적인 다른 현장 염분 데이터 소스인 TSG 데이터를 사용하여 평가될 수 있다.
- [0125] 도 4h를 참고하면, 7가지 기계 학습 모델은 SMAP SSS 프로덕트에 비해 RMSD 및 MB가 감소하고 R2 값이 증가함을 확인할 수 있다.
- [0126] 5개 모델(ANN, RF, XGBoost, LGBM 및 GBRT)이 HYCOM SSS를 능가했지만, K-NN 및 SVR 모델만 그렇지 않음을 확인할 수 있다.
- [0127] 7개의 보정 모델 중 모델 구조가 유사한 4개의 앙상블 트리 기반 기계 학습 방법(RF, XGBoost, LGBM, GBRT)은 유사한 평가 메트릭으로 양적으로 좋은 성능을 보임을 확인할 수 있다.
- [0128] 검증 데이터가 훈련 데이터로부터 일시적으로 분리되는 동안 머신러닝 모델은 설득력 있는 검증 결과를 생성할 수 있다.

- [0129] 전체 데이터 세트에서 GBRT 모델은 시간 홀드아웃 검증 결과로 R2 0.99, RMSD 0.230 psu, MB 0.00 psu를 가졌으며, SMAP SSS 데이터는 R2 0.90, RMSD 1.060 psu, 0.07 psu의 MB를 가질 수 있다.
- [0130] 저염분 샘플(< 32 psu)의 경우, GBRT가 0.833 psu의 가장 낮은 RMSD로 최고의 보정 성능을 제공하고 XGBoost 모델이 0.895 psu로, RF가 0.919 psu로 뒤따릅니다. 이것은 3.776 psu의 SMAP SSS 및 1.643 psu의 HYCOM SSS와 비교될 수 있다.
- [0131] HYCOM SSS의 저염분 데이터에 대한 과대 평가가 감소하고(SMAP SSS로 인해), 고염분 데이터의 성능이 향상됨에 따라(HYCOM SSS의 결과), 머신러닝 모델의 전체 오류는 1.062 psu에서 SMAP 및 HYCOM SSS 데이터와 각각 비교할 때 0.334 psu ~ 0.259 psu로 감소함을 확인할 수 있다(도 4b, 4d 및 4f).
- [0132] 도 4a 내지 4f 및 아래 도 5를 참고하면, SMAP 불확실성이 높은 고위도 지역과 HYCOM SSS 데이터가 현장 염분 데이터와 낮은 상관관계를 갖는 하천 지배 지역 모두에 대한 GBRT 모델로 잘 수정된 결과를 확인할 수 있다.
- [0133] 해안 지역에 상대적으로 많은 데이터 포인트가 있고(도 1a 및 1b), 저염분 샘플(< 32 psu)이 많은 TSG는 Argo 데이터보다 SMAP 및 HYCOM SSS와 덜 상관관계가 있음을 확인할 수 있다(도 4h).
- [0135] 도 5a는 본 발명의 일 실시예에 따른 평균 GBRT 기반 해수면 염분 데이터의 공간 분포를 도시한 도면이다. 도 5b는 본 발명의 일 실시예에 따른 평균 위성 기반 제1 해수면 염분 데이터의 공간 분포를 도시한 도면이다. 도 5c는 본 발명의 일 실시예에 따른 GBRT 기반 해수면 염분 데이터와 위성 기반 제1 해수면 염분 데이터의 차이를 도시한 도면이다. 도 5d는 본 발명의 일 실시예에 따른 평균 해양 수치 모델 기반 제2 해수면 염분 데이터의 공간 분포를 도시한 도면이다. 도 5e는 본 발명의 일 실시예에 따른 GBRT 기반 해수면 염분 데이터와 해양 수치 모델 기반 제2 해수면 염분 데이터의 차이를 도시한 도면이다. 도 5f는 본 발명의 일 실시예에 따른 위성 기반 제1 해수면 염분 데이터와 해양 수치 모델 기반 제2 해수면 염분 데이터의 차이를 도시한 도면이다. 도 5g는 본 발명의 일 실시예에 따른 각 해양 지역에 대한 현장 염분 데이터의 RMSD 및 MB를 도시한 도면이다.
- [0136] GBRT에서 파생된 SSS는 Argo 및 TSG 데이터와 높은 상관관계를 나타낼 수 있다.
- [0137] GBRT에서 파생된 SSS는 북위 60° 이상의 극지방, 큰 호수 및 육지로 둘러싸인 바다(예: 카스피해, 흑해, 지중해 및 허드슨만)에서의 분포를 제공하지 않을 수 있다(도 5a).
- [0138] 이는 입력 변수 중 하나인 SWH가 이러한 영역에 데이터를 제공하지 않기 때문일 수 있다. SWH가 있거나 없는 결과에는 큰 차이가 없기 때문에(즉, SWH가 있는 GBRT 모델은 시간 홀드아웃 검증 결과로 R2가 0.99, RMSD가 0.259psu, MB가 0.00psu인 반면 SWH가 없는 GBRT 모델은 R2가 0.99, RMSD가 0.266psu, MB가 0.00psu임) SWH가 없는 모델의 결과로 빈 영역을 보완할 수 있다.
- [0139] GBRT 기반 SSS는 SMAP SSS에 비해 극지방(60° N 이상 및 60° S 미만)에서 과대 평가될 수 있다(도 5a 내지 5c).
- [0140] SMAP L2B 데이터에서 제공하는 HYCOM SSS는 북위 80° 이상의 북극 지역에서 현저히 낮아(도 5d), 이는 추가적인 개선이 필요할 수 있다.
- [0141] 4가지 트리 기반 앙상블 기계 학습 방법(RF, XGBoost, LGBM 및 GBRT)은 다른 3가지 모델(k-NN, SVR 및 ANN)보다 더 나은 성능을 보임을 확인할 수 있다.
- [0142] 이러한 앙상블 방법은 수많은 의사 결정 트리의 결과를 통합할 수 있다. 서로 다른 특징을 가진 여러 트리를 결합(즉, 앙상블)하여 최종 예측 값을 생성하여 모델의 신뢰도를 높일 수 있다.
- [0143] 이러한 트리 기반 앙상블 모델은 종종 다른 유형의 기계 학습 방법보다 더 나은 성능을 생성할 수 있다.
- [0144] 트리 기반 앙상블 모델의 구조는 유사하지만 데이터 특징 및 파라미터 튜닝에 따라 성능이 다를 수 있다.
- [0145] 4가지 앙상블 모델 중 GBRT가 가장 좋은 성능을 보여 RMSD가 가장 낮았으며 고위도 지역과 강이 많은 지역 모두에서 SSS를 잘 개선할 수 있다.
- [0146] GBRT 기반 SSS 오류 분포를 추가로 평가하기 위해 앞서 설명된 8개 영역에 대한 GBRT 기반 SSS와 현장 염분 데이터 간의 차이를 확인할 수 있다(도 5g).
- [0147] GBRT 기반의 SSS 데이터는 SMAP과 HYCOM SSS 데이터의 차이에 따라 지역별로 개선되는 결과를 확인할 수 있다.
- [0148] GBRT 기반 SSS는 모든 지역에서 SMAP SSS보다 현장 염분 데이터와의 차이가 낮지만 전자는 NOP 및 SOO에서

HYCOM SSS보다 약간 높은 차이를 확인할 수 있다.

- [0149] SOO 및 NOP에서 GBRT 기반 SSS 데이터(각각 0.111 및 0.224 psu의 RMSD 및 0.00 psu의 MB)는 HYCOM SSS(RMSD 0.103 및 0.215 psu 및 MB 0.02 및 0.05 psu)와 비교하여 현장 염분 데이터와 유사한 차이를 나타낼 수 있다.
- [0150] GBRT 기반 SSS와 현장 염분 데이터(RMSD 0.111~0.309 psu 및 MB 0.00 psu) 사이의 상관 관계는 SMAP SSS가 낮은 SST와 높은 RF로 인해 현장 염분 데이터와 덜 상관관계가 있는 ARO, SOO 및 NOP 영역(RMSD 1.052~1.922 psu 및 MB 0.08~0.14 psu)에서 크게 개선됨을 확인할 수 있다(도 4c, g, 도 5c).
- [0151] HYOM SSS와 현장 염분 데이터(RMSD 1.021 psu 및 MB 0.21 psu) 사이의 상관관계가 낮은 하천 지배 지역인 ARP 지역은 GBRT 모델(RMSD 0.591 psu 및 MB 0.05 psu)을 통해 잘 수정될 수 있다.
- [0152] GBRT 기반 SSS는 SMAP SSS(RMSD 0.649 psu 및 MB 0.07 psu) 및 HYCOM SSS보다 우수함을 확인할 수 있다(도 4e 및 도 5e의 주황색 원).
- [0153] SMAP 및 HYCOM SSS 데이터 모두 저위도(PAO 지역)의 외양에서 현장 측정과 높은 상관관계를 갖고 있음에도 불구하고 GBRT 모델은 두 프로덕트를 능가할 수 있다.
- [0155] 도 6a 내지 6i는 본 발명의 일 실시예에 따른 현장 TSG 데이터, 위성 기반 제1 해수면 염분 데이터, 해양 수치 모델 기반 제2 해수면 염분 데이터 및 GBRT 기반 해수면 염분 데이터의 시간 플롯 및 공간 분포를 도시한 도면이다.
- [0156] 도 6a 내지 6i를 참고하면, 시간 연속 TSG 시계열 데이터는 추가 검증을 위해 사용될 수 있다.
- [0157] SMAP SSS 불확실성이 높은 지역에서 GBRT 기반 SSS는 HYCOM SSS 그래디언트를 따라 현장 TSG 데이터와 높은 상관관계를 나타낼 수 있다(도 6b, c, d, e).
- [0158] 반면, HYCOM SSS 불확실성이 높은 해안 지역에서 GBRT 기반 모델은 SMAP SSS 데이터를 사용한 현장 측정과 유사한 SSS를 생성할 수 있다(도 6h 및 i).
- [0159] GBRT 기반의 SSS 데이터는 적도 지역에서도 TSG 데이터와 높은 상관관계를 보였으며, HYCOM과 SMAP SSS와 현장 염분 데이터의 상관도가 높음을 확인할 수 있다(도 6f 및 g).
- [0160] 다양한 경우에 개선된 GBRT 기반 모델은 시연속 TSG 시계열 데이터와 높은 상관관계를 보여 더 짧은 시간적, 더 미세한 공간적 스케일에 적합함을 입증할 수 있다.
- [0161] 글로벌 SSS 모델의 높은 성능은 다중 입력 변수의 시너지, 지역에 따라 다양한 물리적 및 환경적 특징을 가진 데이터의 통합, 기계 학습 모델의 최적화 때문일 수 있습니다.
- [0162] 이러한 결과는 본 발명에 따른 GBRT 기반 보정 모델이 SMAP과 HYCOM SSS를 모두 활용하여 보다 정확한 전역 SSS를 제공할 수 있음을 나타낼 수 있다.
- [0163] 본 발명에 따른 모델은 특히 극지방에서 추가 현장 염분 데이터를 활용하여 더욱 향상될 수 있다.
- [0164] 여러 머신러닝 모델을 사용하는 경우 해당 앙상블은 개별 모델을 사용하는 것보다 일반적으로 더 나은 성능을 나타낼 수 있다.
- [0165] 그러나 7가지 기계 학습 모델의 단순 평균 앙상블은 단일 머신러닝 모델을 사용하는 것보다 더 나은 성능을 나타내지 않을 수 있다.
- [0167] 도 7은 본 발명의 일 실시예에 따른 GBRT 모델을 사용하여 식별된 특징 중요도를 도시한 도면이다.
- [0168] 도 7을 참고하면, GBRT 기반 글로벌 SSS 모델에서 사용된 입력 특징의 특징 중요도를 보여주며 결과적으로 최고의 성능을 보임을 확인할 수 있다.
- [0169] 4개의 앙상블 트리 기반 모델(RF, XGBoost, LGBM 및 GBRT)은 모델에 유사한 입력 특징 기여도를 가질 수 있다.
- [0170] HYCOM SSS와 SMAP SSS는 GBRT 모델에 가장 많은 기여함을 확인할 수 있다. V-편광에 대한 SMAP Tb가 세 번째 기여 변수(contributing variable)로 확인되었으며, SMAP 토지 부분과 OISST가 그 뒤를 이을 수 있다.
- [0171] 고염분 샘플(전체 데이터 세트의 98%)에서 높은 성능을 보인 HYCOM SSS는 SMAP SSS보다 모델에 더 많이 기여함을 확인할 수 있다.
- [0172] 32 psu 이상의 고염분 샘플의 경우 HYCOM SSS의 RMSD는 0.280 psu, R²는 0.93, SMAP SSS의 RMSD는 각각 0.970

psu와 0.50일 수 있다.

- [0173] SMAP Tb 중 V-편광에 대한 Tb가 H-편광에 대한 Tb보다 모델에 더 많이 기여함을 확인할 수 있다.
- [0174] 일 실시예에서, SSS에 대한 Tb의 반응은 H-편광보다 V-편광에 약간 더 민감할 수 있다.
- [0175] 토지분율 값(land fraction value)은 SMAP Tb의 토지 보정을 위해 시야(field of view)에서 토지의 비율로 사용될 수 있다.
- [0176] SMAP SSS 프로덕트의 불확실성은 LSC로 인해 해안 지역에서 증가하기 때문에, 토지 비율을 모델에 대한 기여 특징(contributing feature)으로 가질 수 있으며, 토지 부분과 유사한 특징을 갖는 토지로부터의 거리가 뒤따를 수 있다.
- [0177] 또한, SSS는 SSS, SST 및 Tb 간의 관계를 사용하여 위성 데이터에서 산출되기 때문에, SST는 SSS 및 Tb에 이어 기여 특징일 수 있다.
- [0178] L-밴드 마이크로파 센서의 표면 민감도는 SST의 영향을 크게 받으며, SST가 높을수록 증가하고 차가운 바다보다 따뜻한 바다에서 SSS 산출에 더 나은 성능을 제공할 수 있다.
- [0179] SMAP SSS는 SST가 낮은 극지방에서 상당한 불확실성을 가질 수 있다(도 3g).
- [0180] SSS와 SST 사이의 상관 관계는 주로 서로 다른 수괴(water masses)의 존재, 이류 및 혼합으로 인해 변하기 때문에, 둘 사이의 상관 관계는 종종 지역과 계절에 따라 다를 수 있으며, SSS 정량화에 대한 SST의 다른 기여도를 도출할 수 있다.
- [0181] SMAP 및 HYCOM SSS(특징 중요도의 대부분을 차지하는)만을 입력 변수로 사용하여 훈련된 GBRT 모델의 시간적 홀드아웃 검증 결과는 R2가 0.99, RMSD가 0.319 psu, MB가 0.00 psu일 수 있다.
- [0182] 저염분 샘플(< 32 psu)에 대한 R2, RMSE 및 MB는 0.99, 1.341 psu 및 0.47 psu였으며, 이는 본 발명(전체 검증 데이터에 대한 0.230 psu의 RMSD 및 저염분 샘플에 대한 0.833 psu의 RMSD, 도 4h)에서 11개의 입력 변수로 훈련된 GBRT 모델보다 나쁜 결과일 수 있다.
- [0183] 이는 SMAP 및 HYCOM SSS가 매우 높은 특징 중요도를 가졌지만, 나머지 변수도 어느 정도 모델에 기여했음을 확인할 수 있다.
- [0185] 도 8a는 본 발명의 일 실시예에 따른 GBRT 모델에 대한 SHAP 값의 플롯을 도시한 도면이다. 도 8b는 본 발명의 일 실시예에 따른 ARP 지역의 플롯을 도시한 도면이다. 도 8c는 본 발명의 일 실시예에 따른 위성 기반 제1 해수면 염분 데이터의 SHAP 종속 플롯을 도시한 도면이다. 도 8d는 본 발명의 일 실시예에 따른 해양 수치 모델 기반 제2 해수면 염분 데이터의 SHAP 종속 플롯을 도시한 도면이다.
- [0186] 도 8a 내지 8c를 참고하면, GBRT 기계 학습 모델에 대한 SHAP 값의 요약 플롯(summary plot)과 종속 플롯(dependency plot)을 보여줍니다.
- [0187] GBRT 모델(도 7)에 의해 확인된 매우 중요한 V-편광에 대한 SMAP Tb, HYCOM SSS, SMAP SSS, OISST 및 SMAP 토지 부분(land fraction)의 SHAP 값은 높았으며, 이러한 특징이 크게 기여했음을 확인할 수 있다.
- [0188] 각 특징이 기여하는 양수 또는 음수 방향은 SHAP 값을 사용하여 해석할 수 있다.
- [0189] 각 변수에는 샘플이 변경됨에 따라 다른 SHAP 값과 기여 방향이 있을 수 있다. 전체 데이터 세트에 대한 SHAP 값의 요약 플롯(도 8a)에서 높은 HYCOM SSS는 GBRT 기반 모델에 긍정적으로 기여함을 확인할 수 있다. 즉, 평균값에서 SSS를 증가시키는 방향으로, HYCOM SSS의 높은 성능 염분이 높은 지역을 나타낼 수 있다.
- [0190] 이러한 결과는 도 3g(도 8b)의 8개 영역 중 HYCOM SSS의 성능이 가장 낮은 ARP 영역에 대한 요약 플롯에서 더욱 명확하게 나타날 수 있다.
- [0191] SMAP과 관련된 특징(즉, V 및 H 편광에 대한 SMAP SSS 및 Tbs)은 다른 특징보다 음수 값이 더 높을 수 있다.
- [0192] 낮은 SMAP SSS는 HYCOM SSS 데이터와 비교하여 음으로(즉, 평균값에서 SSS 감소, 저염분 영역에서 SMAP SSS의 고성능) 기여할 수 있다.
- [0193] 도 8c와 8d는 SMAP SSS와 HYCOM SSS가 토지 부분 값에 따라 전체 데이터 세트(즉, 글로벌)에 대한 GBRT 모델에서 상호 작용하는 방법을 나타낼 수 있다.

- [0194] SMAP SSS의 SHAP 값은 HYCOM SSS의 SHAP 값에 비해 일반적으로 (즉, 대양에 위치한) 낮은 토지 부분 값(land fraction value)을 갖는 고염분 샘플(> 32 psu, 그림 8(c))에 대해 0에 가까울 수 있다(도 8d).
- [0195] 이것은 GBRT 모델이 낮은 토지 부분 값을 가진 고염분 샘플에 대해 SMAP SSS 대신 HYCOM SSS 및 OISST와 같은 다른 입력 특징을 사용하여 설정할 수 있음을 의미할 수 있다.
- [0196] 대부분의 고염분 샘플은 HYCOM SSS 데이터(도 3g)와 높은 상관관계가 있었기 때문에 SMAP SSS의 기여도는 상대적으로 낮을 수 있다.
- [0198] 본 발명에 따르면, SMAP과 HYCOM 데이터의 시너지 활용을 통해 글로벌 SSS를 생성하기 위한 머신러닝 모델이 사용될 수 있다.
- [0199] 7개의 머신러닝 알고리즘을 평가하여 고품질의 글로벌 SSS를 생성했으며, GBRT가 가장 좋은 머신러닝 모델임을 확인할 수 있다.
- [0200] 지역에 따라 변하는 입력 특징 간의 관계를 일반화하는 것은 어렵지만, 모든 지역에서 GBRT 기반 SSS의 성능이 HYCOM SSS 및 SMAP SSS보다 높음을 확인할 수 있다.
- [0201] 다양한 기후 및 해양 모델에서 본 발명에 따른 글로벌 해수면 염분 데이터를 사용하면 머신러닝 모델에 기반하여 더 정확한 계절 예측 및 기후 예측이 가능할 수 있다.
- [0203] 도 9는 본 발명의 일 실시예에 따른 위성 기반 해수면 염분 데이터와 해양 수치 모델 기반 해수면 염분 데이터를 머신러닝 모델에 적용하여 해수면 염분을 예측하기 위한 방법을 도시한 도면이다. 일 실시예에서, 도 9의 각 단계는 도 10의 해수면 염분을 예측하기 위한 장치(1000)에 의해 수행될 수 있다.
- [0204] 도 9를 참고하면, S901 단계는, 위성 기반 제1 해수면 염분 데이터, 해양 수치 모델 기반 제2 해수면 염분 데이터 및 해수면 환경 데이터 중 적어도 하나를 수신하는 단계이다.
- [0205] 일 실시예에서, 위성 기반 제1 해수면 염분 데이터, 해양 수치 모델 기반 제2 해수면 염분 데이터 및 해수면 환경 데이터 중 적어도 하나는 위성, 외부 서버 및 센서 장치 중 적어도 하나로부터 수신될 수 있다.
- [0206] 일 실시예에서, 제1 해수면 염분 데이터는, 위성으로부터 획득된 해수면에 대한 마이크로파의 밝기 온도(brightness temperature) 데이터에 기반하여 산출된 해수면 염분 데이터를 포함할 수 있다.
- [0207] 일 실시예에서, 제2 해수면 염분 데이터는, 해수면에 대한 센서로부터 측정된 현장 염분 데이터를 해양 수치 모델에 적용하여 산출된 해수면 염분 데이터를 포함할 수 있다.
- [0208] 일 실시예에서, 해수면 환경 데이터는, 해수면 온도 데이터, 풍속 데이터, 강수 데이터 및 파고 데이터 중 적어도 하나를 포함할 수 있다.
- [0209] S903 단계는, 제1 해수면 염분 데이터, 제2 해수면 염분 데이터 및 해수면 환경 데이터 중 적어도 하나를 머신러닝 모델에 입력하여 전역(global) 해수면 염분 데이터를 출력하는 단계이다.
- [0210] 일 실시예에서, S903 단계 이전에, 제1 해수면 염분 데이터, 제2 해수면 염분 데이터, 해수면 환경 데이터 및 해수면에 대한 센서로부터 측정된 현장 염분 데이터를 이용하여 머신러닝 모델을 학습시킬 수 있다.
- [0211] 일 실시예에서, 전역 해수면 염분 데이터는 글로벌 지역에 대한 연속적인 해수면 염분 데이터를 2차원 이미지로 표현한 형태로 구현될 수 있다.
- [0213] 도 10은 본 발명의 일 실시예에 따른 위성 기반 해수면 염분 데이터와 해양 수치 모델 기반 해수면 염분 데이터를 머신러닝 모델에 적용하여 해수면 염분을 예측하기 위한 장치(1000)를 도시한 도면이다. 일 실시예에서, 장치(1000)는 사용자 단말(예: 데스크탑, 태블릿 PC, 노트북, 스마트폰) 또는 서버로 구현될 수 있다.
- [0214] 도 10을 참고하면, 장치(1000)는 통신부(1010), 제어부(1020), 저장부(1030) 및 표시부(1040)를 포함할 수 있다.
- [0215] 일 실시예에서, 통신부(1010)는 위성 기반 제1 해수면 염분 데이터, 해양 수치 모델 기반 제2 해수면 염분 데이터 및 해수면 환경 데이터 중 적어도 하나를 수신할 수 있다.
- [0216] 일 실시예에서, 통신부(1010)는 유선 통신 모듈 및 무선 통신 모듈 중 적어도 하나를 포함할 수 있다. 통신부(1010)의 전부 또는 일부는 '송신부', '수신부' 또는 '송수신부(transceiver)'로 지칭될 수 있다.

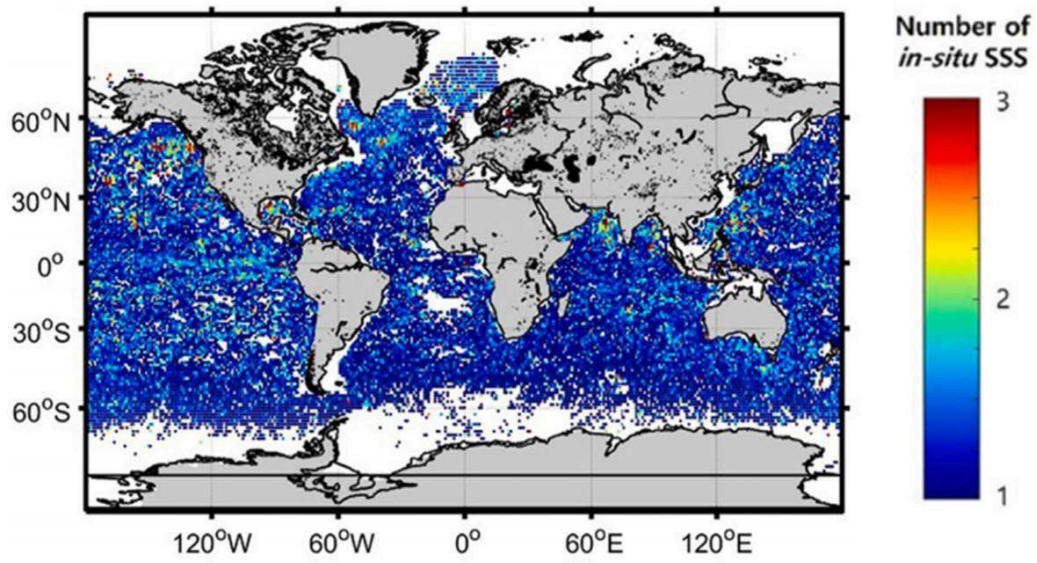
- [0217] 제어부(1020)는 제1 해수면 염분 데이터, 제2 해수면 염분 데이터 및 해수면 환경 데이터 중 적어도 하나를 머신러닝 모델에 입력하여 전역 해수면 염분 데이터를 출력할 수 있다.
- [0218] 일 실시예에서, 제어부(1020)는 적어도 하나의 프로세서 또는 마이크로(micro) 프로세서를 포함하거나, 또는, 프로세서의 일부일 수 있다. 또한, 제어부(1020)는 CP(communication processor)라 지칭될 수 있다. 제어부(1020)는 본 발명의 다양한 실시예에 따른 장치(300)의 동작을 제어할 수 있다.
- [0219] 저장부(1030)는 위성 기반 제1 해수면 염분 데이터, 해양 수치 모델 기반 제2 해수면 염분 데이터 및 해수면 환경 데이터 중 적어도 하나를 저장할 수 있다.
- [0220] 저장부(1030)는 머신러닝 모델 및 전역 해수면 염분 데이터를 저장할 수 있다.
- [0221] 일 실시예에서, 저장부(1030)는 휘발성 메모리, 비휘발성 메모리 또는 휘발성 메모리와 비휘발성 메모리의 조합으로 구성될 수 있다. 그리고, 저장부(1030)는 제어부(1020)의 요청에 따라 저장된 데이터를 제공할 수 있다.
- [0222] 표시부(1040)는 전역(global) 해수면 염분 데이터를 디스플레이할 수 있다.
- [0223] 일 실시예에서, 표시부(1040)는 장치(1000)에서 처리되는 정보를 나타낼 수 있다. 예를 들면, 표시부(1040)는 액정 디스플레이(LCD; Liquid Crystal Display), 발광 다이오드(LED; Light Emitting Diode) 디스플레이, 유기 발광 다이오드(OLED; Organic LED) 디스플레이, 마이크로 전자기계 시스템(MEMS; Micro Electro Mechanical Systems) 디스플레이 및 전자 종이(electronic paper) 디스플레이 중 적어도 어느 하나를 포함할 수 있다.
- [0224] 도 10을 참고하면, 장치(1000)는 통신부(1010), 제어부(1020) 및 저장부(1030) 및 표시부(1040)를 포함할 수 있다. 본 발명의 다양한 실시 예들에서 장치(1000)는 도 10에 설명된 구성들이 필수적인 것은 아니어서, 도 10에 설명된 구성들보다 많은 구성들을 가지거나, 또는 그보다 적은 구성들을 가지는 것으로 구현될 수 있다.
- [0226] 이상의 설명은 본 발명의 기술적 사상을 예시적으로 설명한 것에 불과한 것으로, 통상의 기술자라면 본 발명의 본질적인 특성이 벗어나지 않는 범위에서 다양한 변경 및 수정이 가능할 것이다.
- [0227] 본 명세서에 개시된 다양한 실시예들은 순서에 관계없이 수행될 수 있으며, 동시에 또는 별도로 수행될 수 있다.
- [0228] 일 실시예에서, 본 명세서에서 설명되는 각 도면에서 적어도 하나의 단계가 생략되거나 추가될 수 있고, 역순으로 수행될 수도 있으며, 동시에 수행될 수도 있다.
- [0229] 본 명세서에 개시된 실시예들은 본 발명의 기술적 사상을 한정하기 위한 것이 아니라, 설명하기 위한 것이고, 이러한 실시예들에 의하여 본 발명의 범위가 한정되는 것은 아니다.
- [0230] 본 발명의 보호범위는 청구범위에 의하여 해석되어야 하며, 그와 동등한 범위 내에 있는 모든 기술 사상은 본 발명의 권리범위에 포함되는 것으로 이해되어야 한다.

부호의 설명

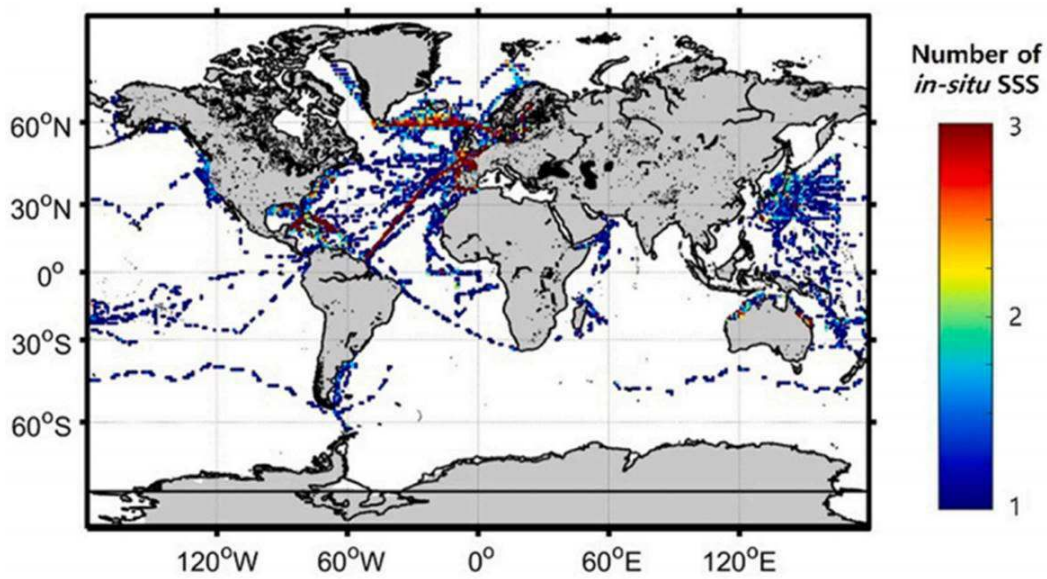
- [0232] 1000: 장치
- 1010: 통신부
- 1020: 제어부
- 1030: 저장부
- 1040: 표시부

도면

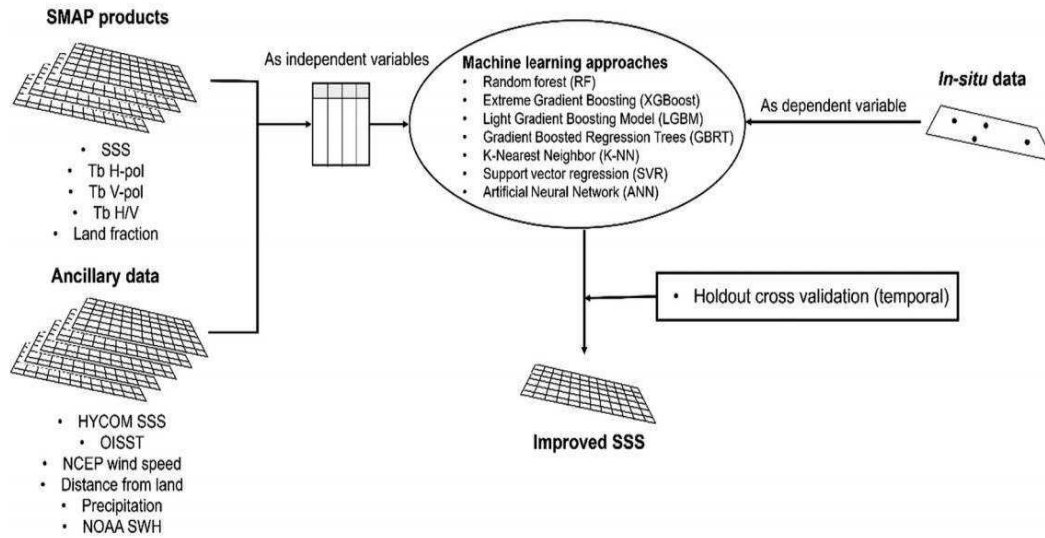
도면1a



도면1b



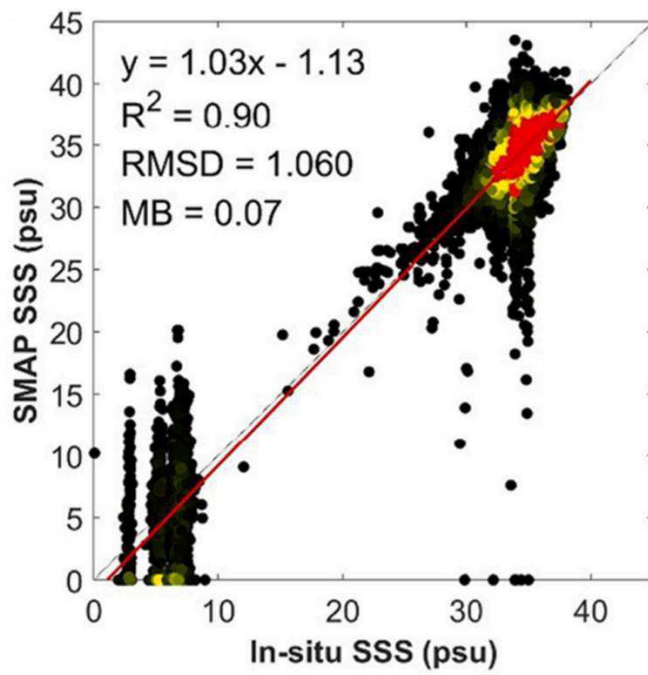
도면2a



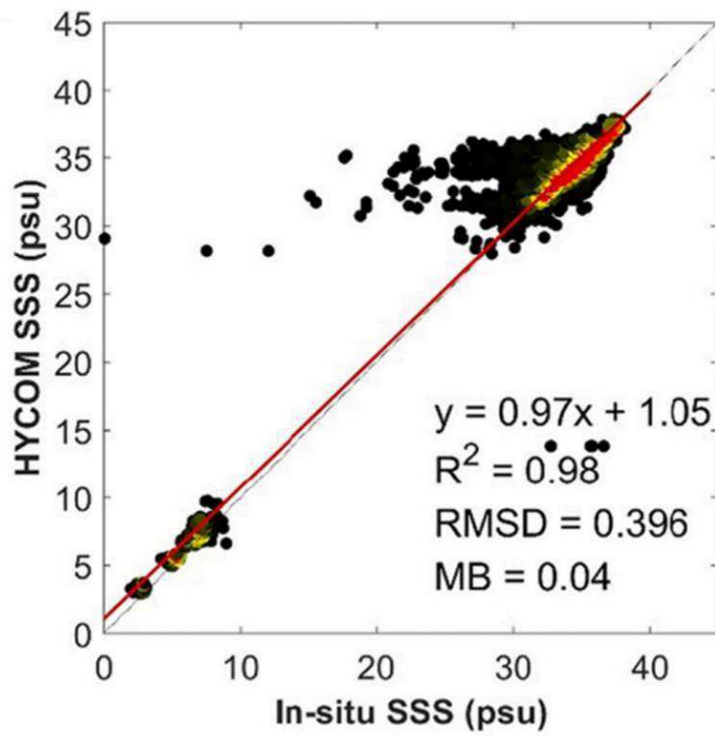
도면2b

Input Data	Data Source	Variable (unit)
Distance from land	TRMM TMPA	Distance from land (km)
SMAP SSS	SMAP	Sea surface salinity (psu)
Tb H-pol		Brightness temperature of H-polarization (K)
Tb V-pol		Brightness temperature of V-polarization (K)
Tb H/V		The Tb ratio between H- and V-pol
Land fraction		Land fraction
HYCOM SSS	HYCOM model	Sea surface salinity (psu)
OISST	NOAA OISST	Sea surface temperature (K)
NCEP wind speed	NOAA NCEP GFS model	Wind speed (m/s)
GPM precipitation	GPM 3IMERGDL	Precipitation (mm)
NOAA SWH	NOAA WaveWatch III	Sea wave height (m)

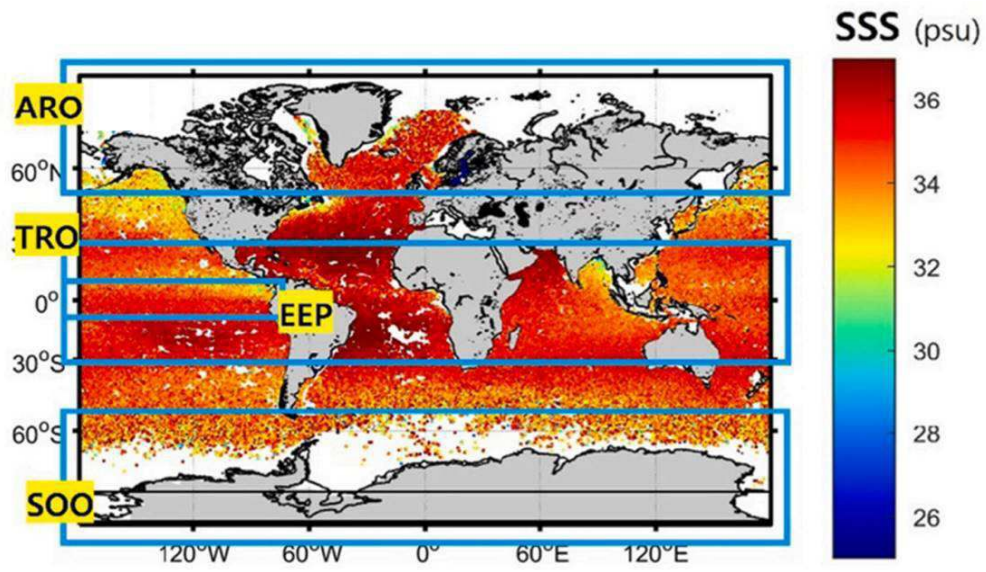
도면3a



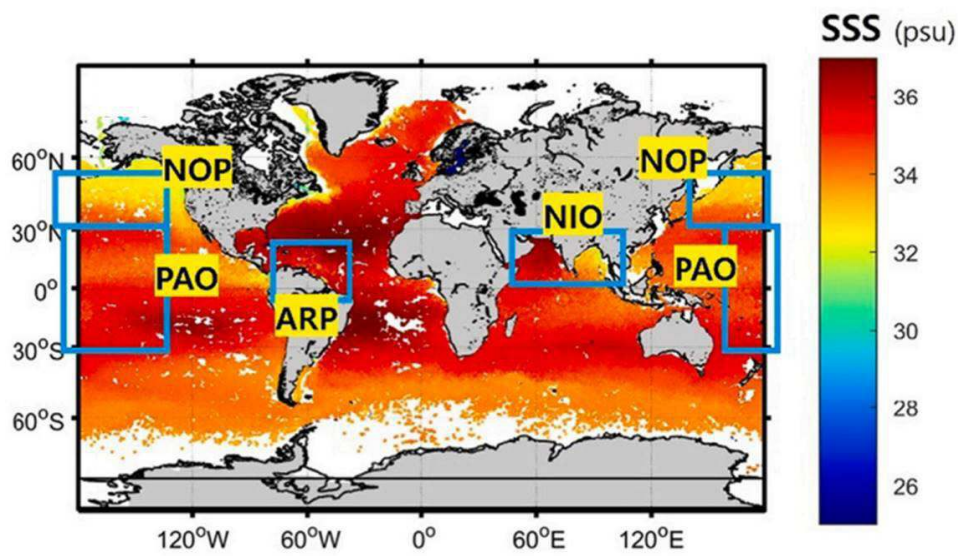
도면3b



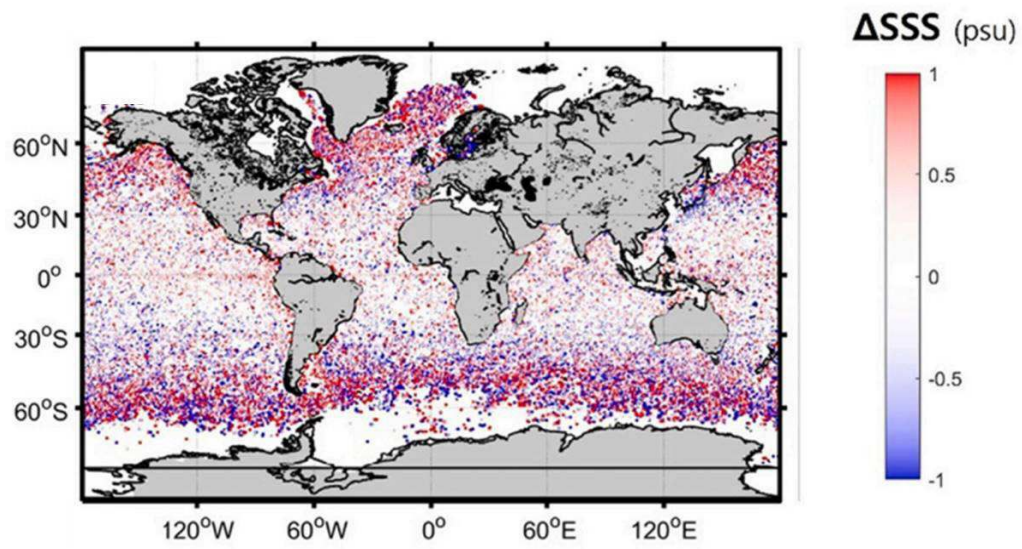
도면3c



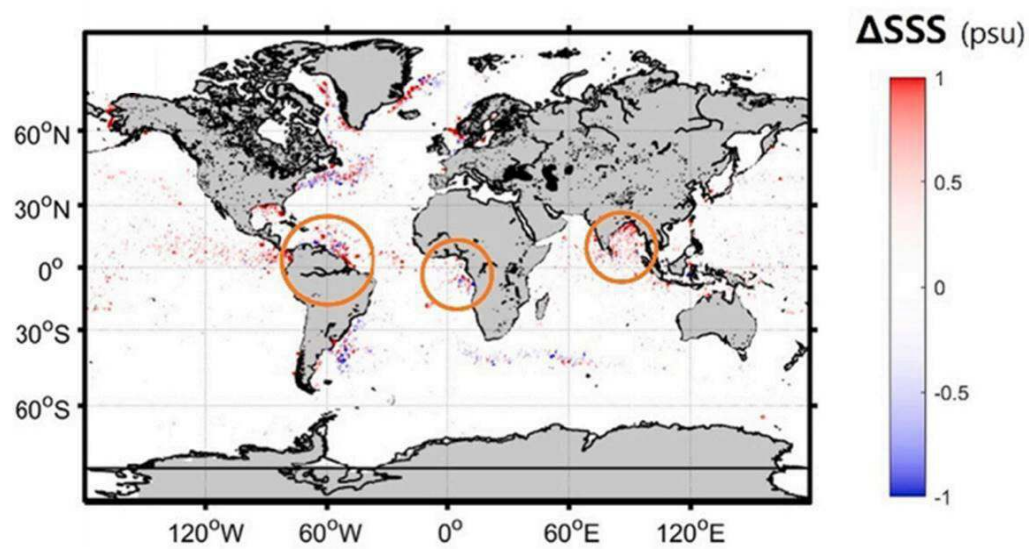
도면3d



도면3e



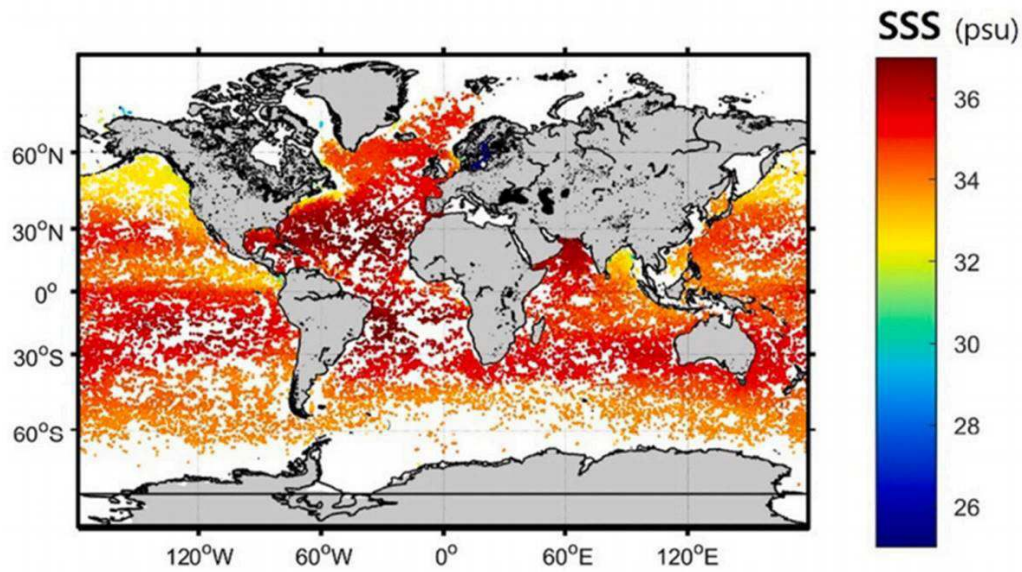
도면3f



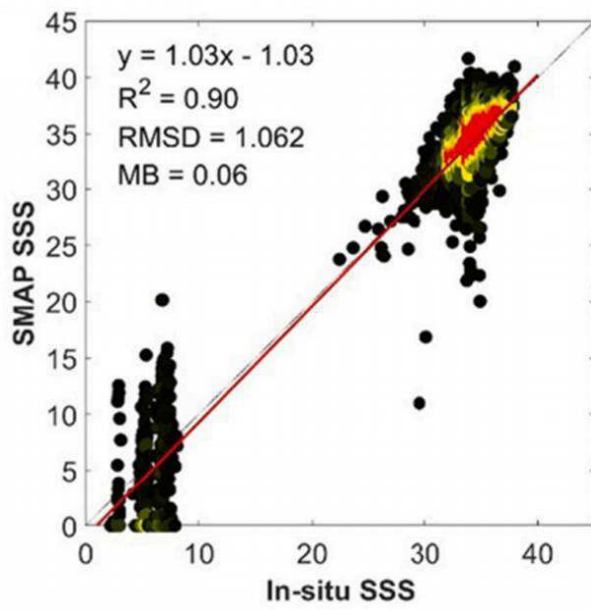
도면3g

Region	Latitude	Longitude	N ^a	SST (°C)	SMAP		HYCOM	
					RMSD (psu)	MB (psu)	RMSD (psu)	MB (psu)
Global	whole data		207,853	19.69	1.060	0.07	0.396	0.04
	SSS ≥ 32		204,361	19.79	0.970	0.08	(0.440)	(0.04)
	10 < SSS < 32		1524	18.54	1.798	0.55	(0.347)	(0.02)
Pacific Ocean (PAO)	30° S – 30° N	180° W – 140° W, 160° E – 180° E	23,853	27.15	0.563	0.04	3.050	1.85
Tropical Ocean (TRO)	30° S – 30° N	180° W – 180° E	102,911	26.59	0.591	0.07	(3.032)	(1.78)
Arctic Ocean (ARO)	50° N – 90° N	180° W – 180° E	22,721	7.99	1.963	0.14	0.218	0.07
Southern Ocean (SOO)	90° S – 50° S	180° W – 180° E	15,803	4.20	1.725	0.08	(0.228)	(0.08)
North Pacific (NOP)	30° N – 50° N	180° W – 120° W, 140° E – 180° E	13,160	15.15	1.052	0.13	0.439	0.07
Eastern Equatorial Pacific (EEP)	10° S – 10° N	180° W – 80° W	16,611	26.92	0.565	0.15	(0.452)	(0.06)
Northern Indian Ocean (NIO)	0° N – 25° N	50° E – 110° E	8376	28.55	0.634	0.11	0.449	0.08
Amazon River Plume (ARP)	5° S – 20° N	75° W – 30° W	4046	27.47	0.647	0.07	(0.728)	(0.10)
							0.098	0.02
							(0.105)	(0.02)
							0.217	0.05
							(0.209)	(0.05)
							0.306	0.09
							(0.312)	(0.10)
							0.464	0.13
							(0.732)	(0.04)
							1.258	0.25
							(1.320)	(0.27)

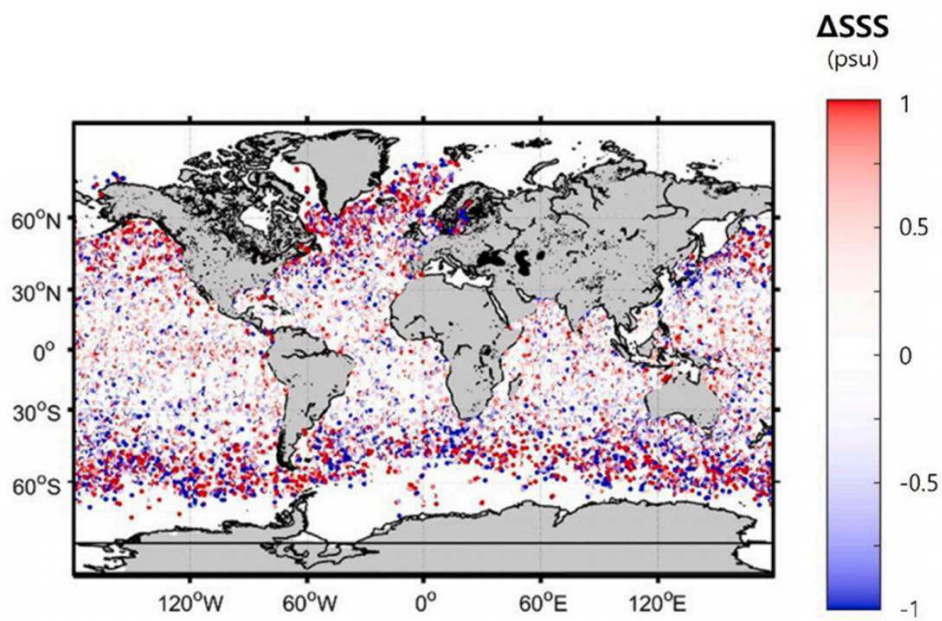
도면4a



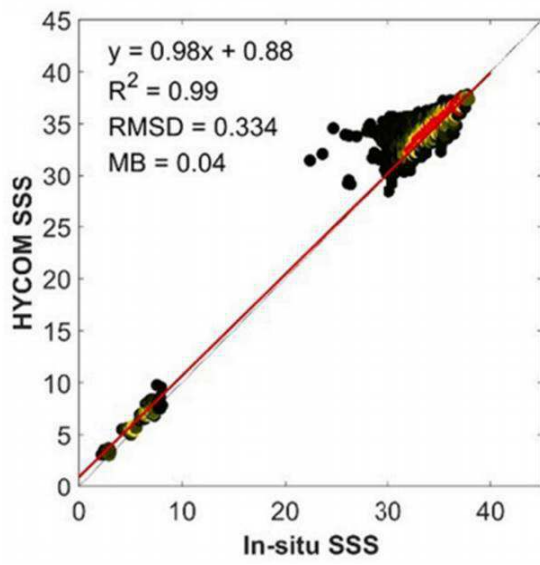
도면4b



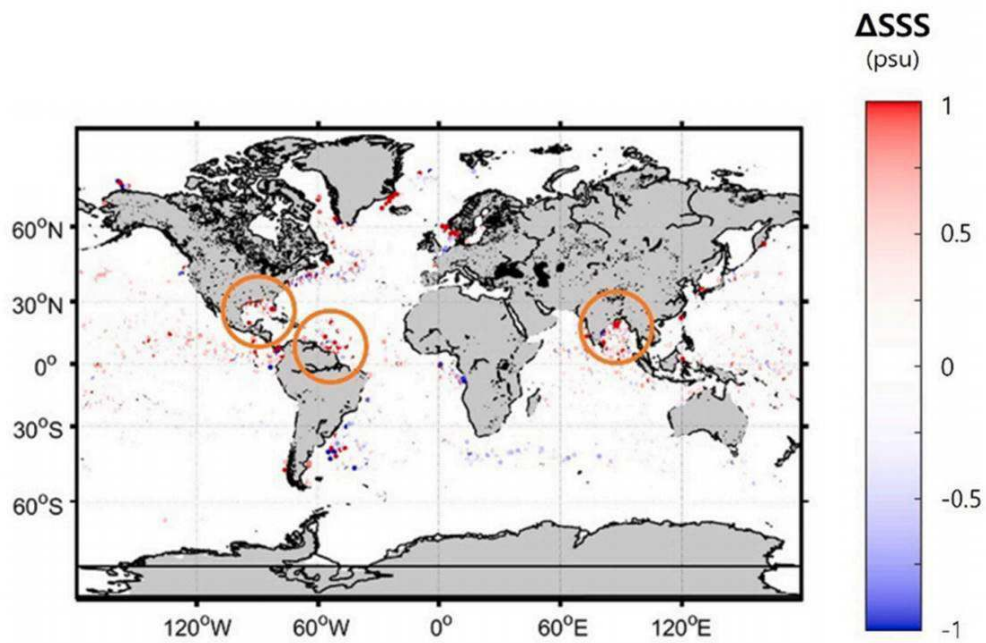
도면4c



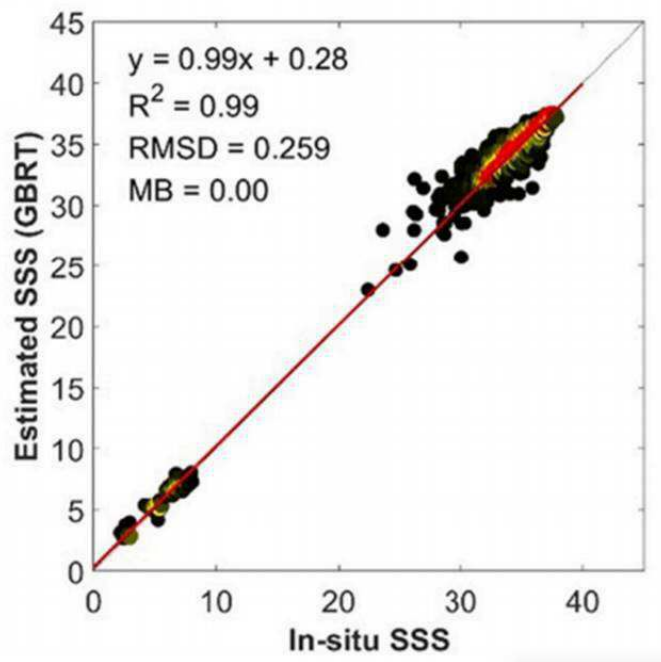
도면4d



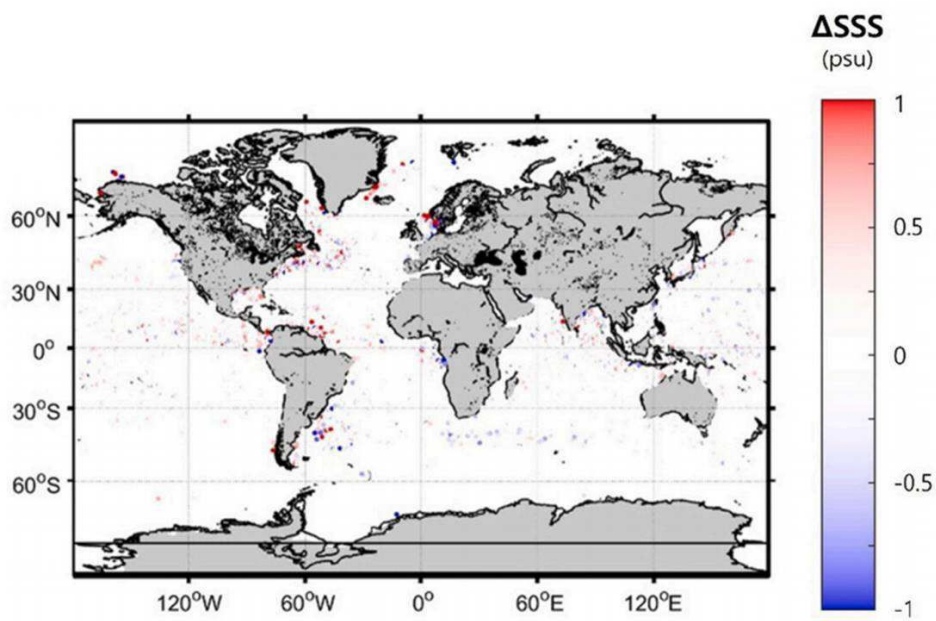
도면4e



도면4f



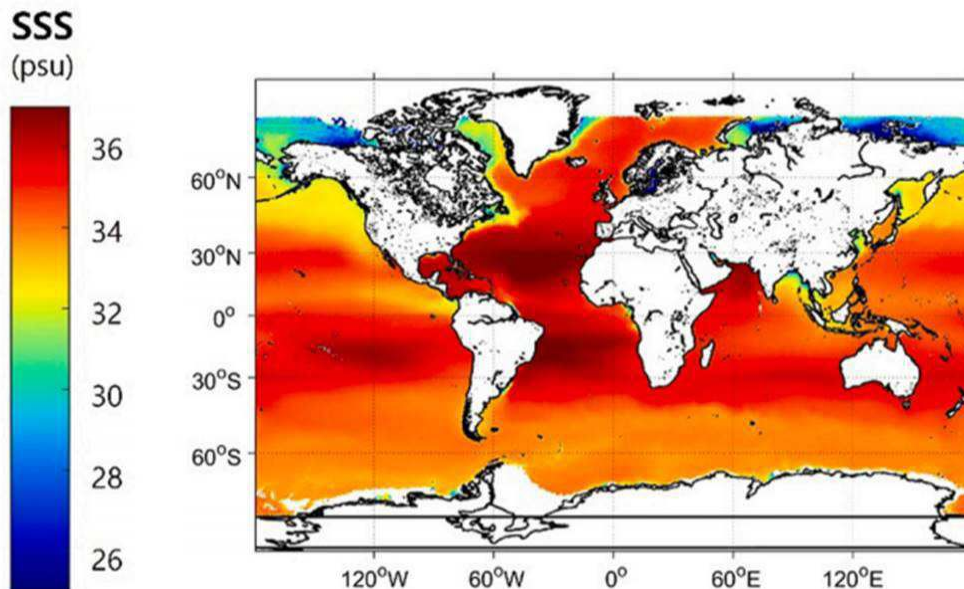
도면4g



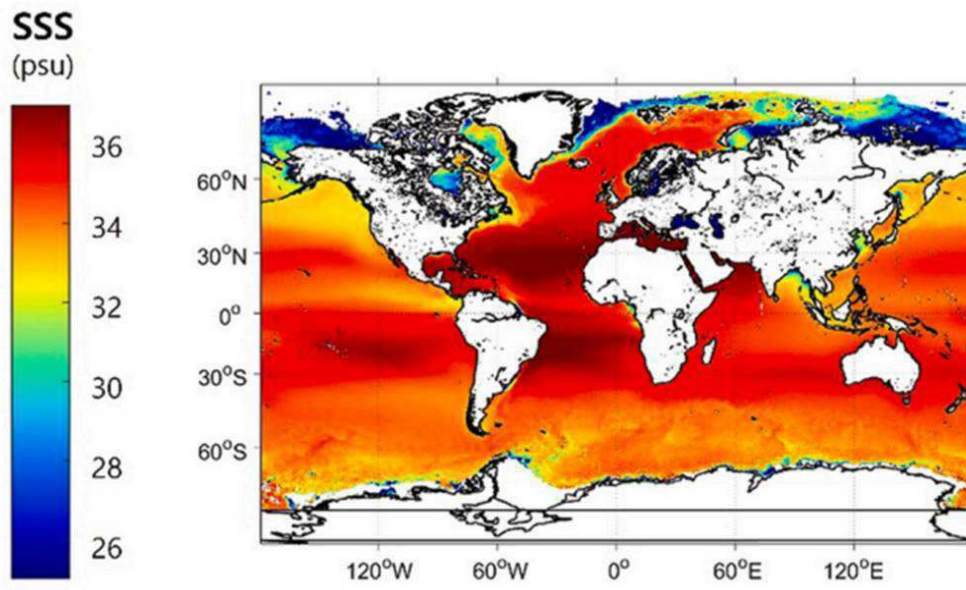
도면4h

Approach			RMSD (psu)		R ²		MB (psu)	
			SSS < 32	whole	SSS < 32	whole	SSS < 32	whole
SMAP SSS	Training		3.363	1.060	0.95	0.90	-0.57	0.07
	Validation	Whole	3.776	1.062	0.93	0.90	-0.60	0.06
		Argo	3.086	0.982	0.95	0.77	-0.34	0.06
		TSG	4.120	1.387	0.89	0.96	-0.75	0.08
HYCOM SSS	Training		2.158	0.398	0.98	0.98	1.04	0.04
	Validation	Whole	1.643	0.334	0.99	0.99	0.87	0.04
		Argo	1.354	0.264	0.99	0.98	0.70	0.03
		TSG	1.788	0.563	0.99	0.99	0.96	0.09
K-NN	Training		0.966	0.377	1.00	0.98	0.34	0.02
	Validation	Whole	1.153	0.452	0.99	0.98	0.48	0.01
		Argo	1.293	0.436	0.99	0.94	0.67	0.02
		TSG	1.064	0.525	0.99	0.99	0.36	-0.04
SVR	Training		1.898	0.524	0.98	0.97	-0.09	-0.17
	Validation	Whole	1.506	0.486	0.99	0.98	-0.26	-0.17
		Argo	1.261	0.455	0.99	0.94	-0.26	-0.14
		TSG	1.631	0.768	0.99	0.99	-0.26	-0.30
ANN	Training		1.487	0.340	0.99	0.99	0.52	0.07
	Validation	Whole	1.181	0.297	0.99	0.99	0.37	0.07
		Argo	0.975	0.254	1.00	0.98	0.30	0.07
		TSG	1.285	0.451	0.99	1.00	0.41	0.08
RF	Training		0.358	0.098	1.00	1.00	0.11	0.00
	Validation	Whole	0.919	0.265	1.00	0.99	0.31	0.00
		Argo	0.831	0.232	1.00	0.98	0.26	0.00
		TSG	0.966	0.388	1.00	1.00	0.34	0.02
XGBoost	Training		0.885	0.264	1.00	0.99	0.31	0.00
	Validation	Whole	0.895	0.264	1.00	0.99	0.31	0.00
		Argo	0.841	0.234	1.00	0.98	0.25	-0.01
		TSG	0.925	0.379	1.00	1.00	0.35	0.03
LGBM	Training		1.035	0.275	0.99	0.99	0.31	0.00
	Validation	Whole	1.338	0.295	0.99	0.99	0.32	0.00
		Argo	0.959	0.238	0.99	0.98	0.24	0.00
		TSG	1.513	0.486	0.98	0.99	0.36	0.03
GBRT	Training		0.584	0.230	1.00	0.99	0.20	0.00
	Validation	Whole	0.833	0.259	1.00	0.99	0.29	0.00
		Argo	0.861	0.234	1.00	0.98	0.27	0.00
		TSG	0.816	0.358	1.00	1.00	0.29	0.02

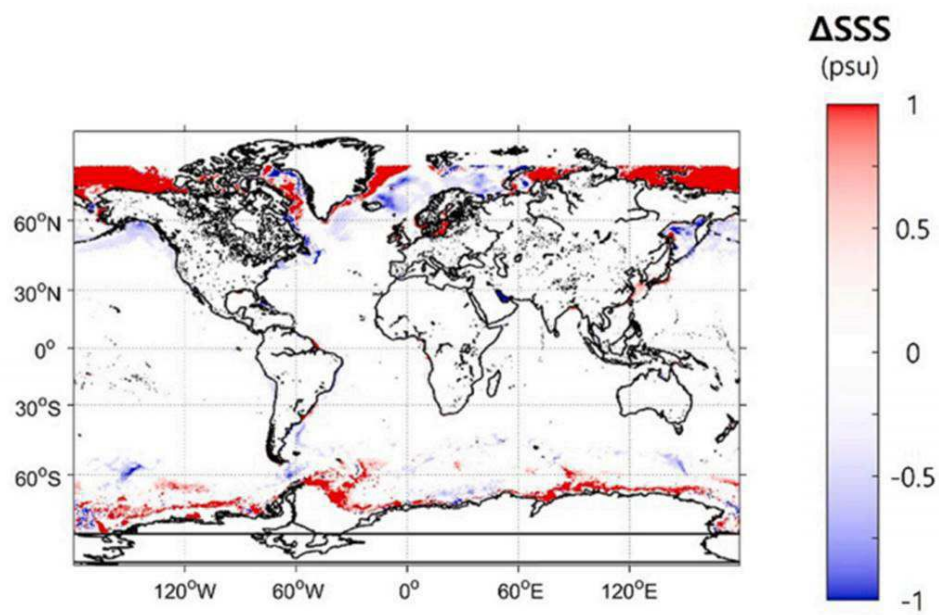
도면5a



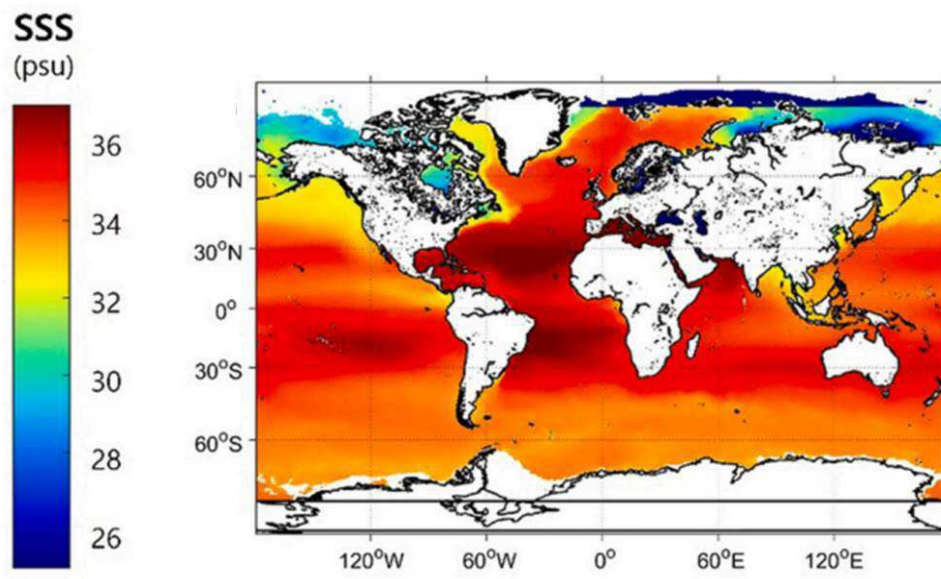
도면5b



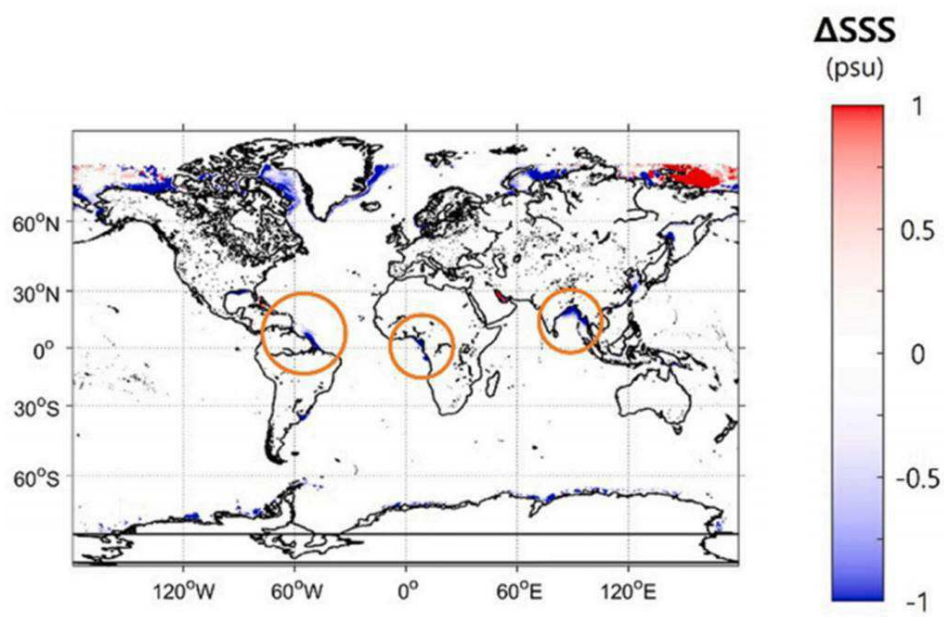
도면5c



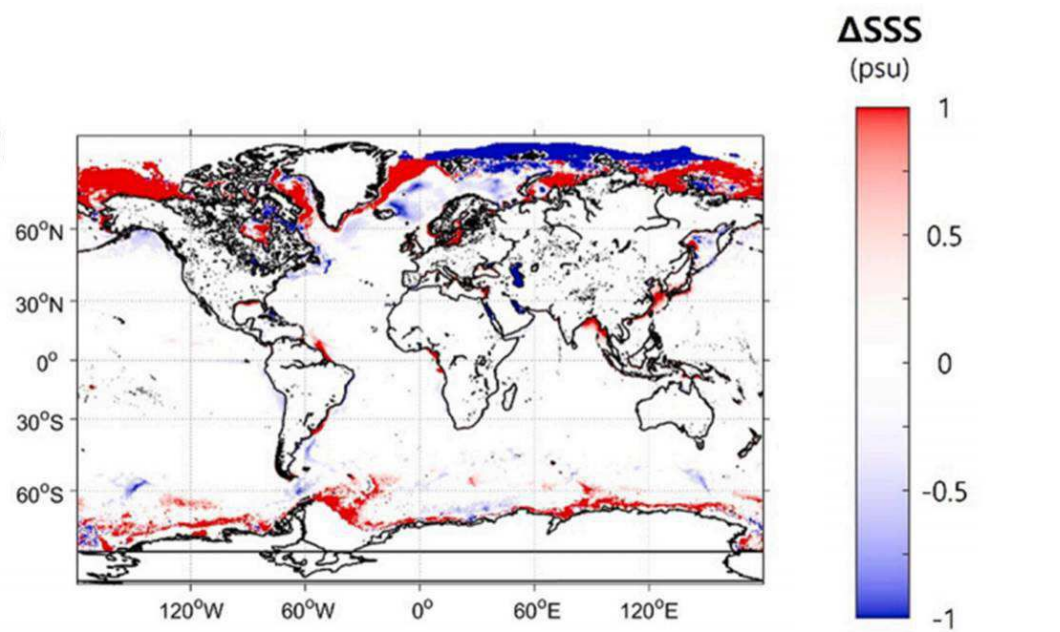
도면5d



도면5e



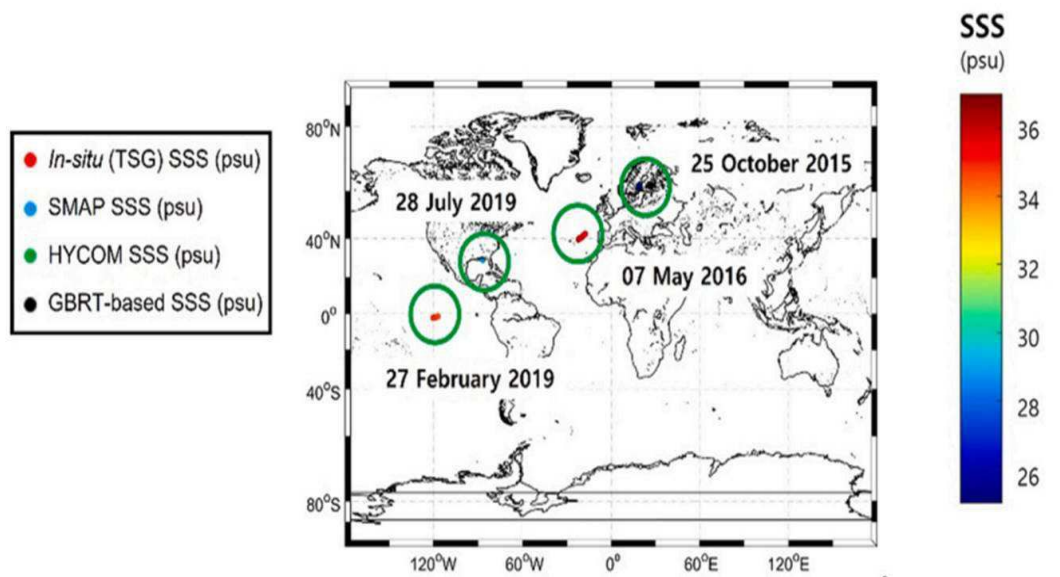
도면5f



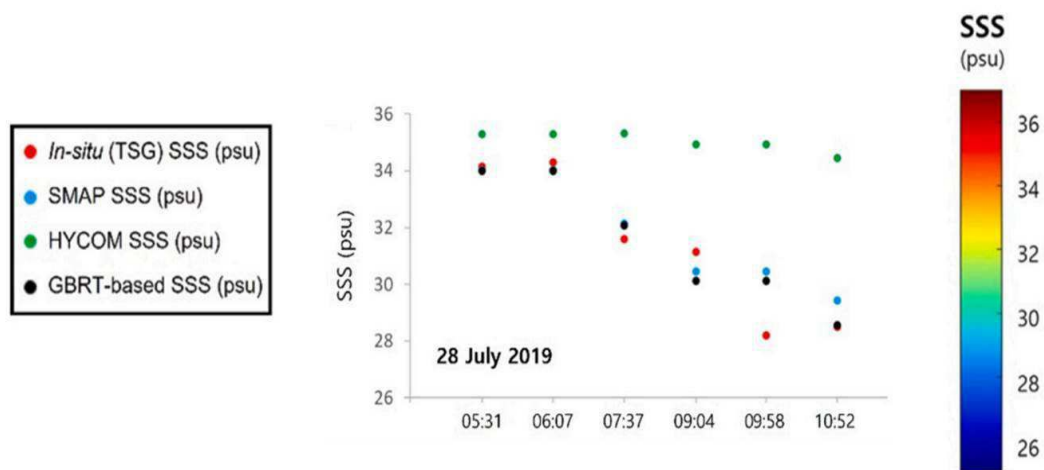
도면5g

Region	N ^a	SMAP SSS		HYCOM SSS		GBRT-based SSS	
		RMSD (psu)	MB (psu)	RMSD (psu)	MB (psu)	RMSD (psu)	MB (psu)
Global	40,801	1.062	0.06	0.334	0.04	0.259	0.00
Pacific Ocean (PAO)	4644	0.559	0.04	0.218	0.07	0.199	0.00
Tropical Ocean (TRO)	20,121	0.585	0.08	0.381	0.07	0.270	0.01
Arctic Ocean (ARO)	4815	1.922	0.14	0.392	0.06	0.309	0.00
Southern Ocean (SOO)	3103	1.713	0.08	0.103	0.02	0.111	0.00
North Pacific (NOP)	2708	1.052	0.10	0.215	0.05	0.224	0.01
Eastern Equatorial Pacific (EEP)	3178	0.575	0.14	0.298	0.09	0.247	0.02
Northern Indian Ocean (NIO)	1803	0.632	0.11	0.507	0.13	0.352	0.01
Amazon River Plume (ARP)	829	0.649	0.07	1.021	0.21	0.591	0.05

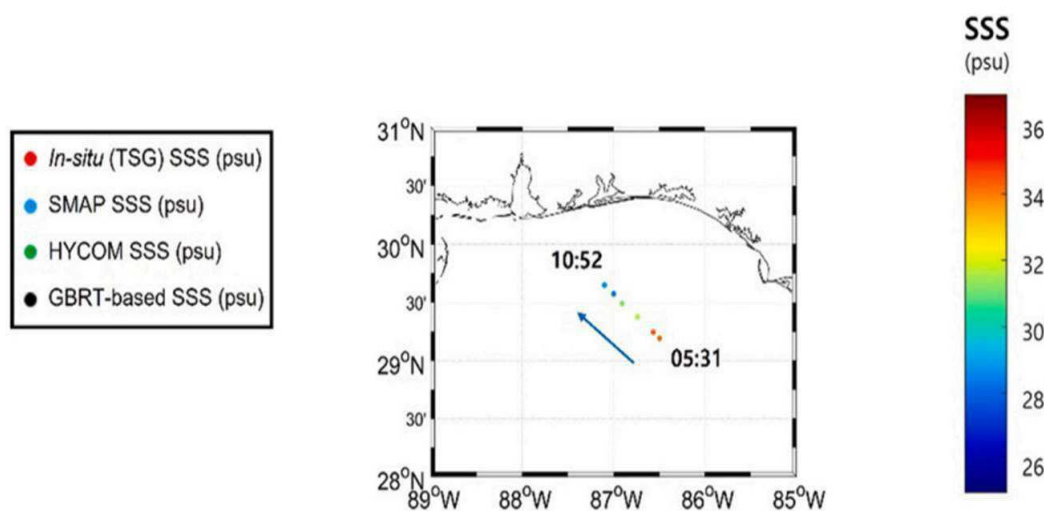
도면6a



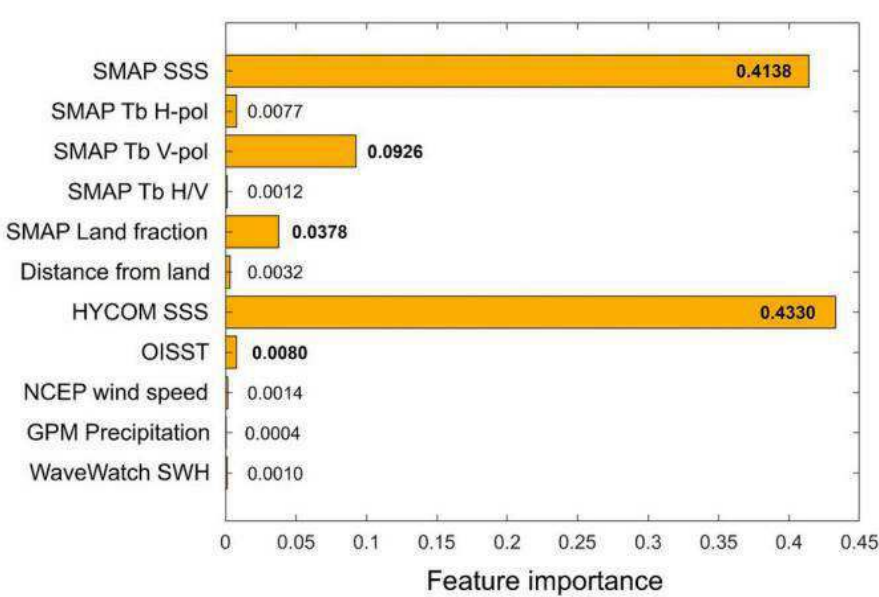
도면6h



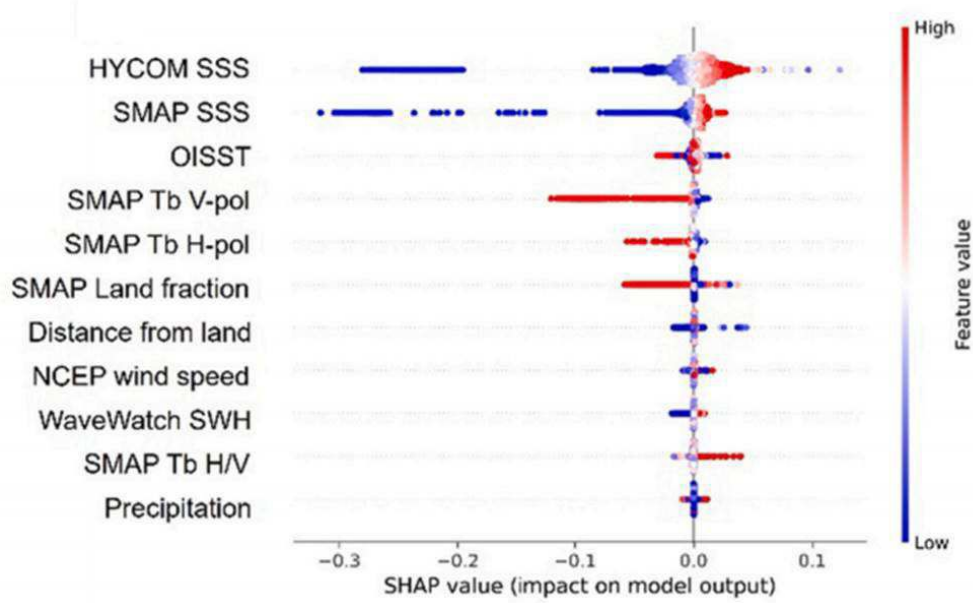
도면6i



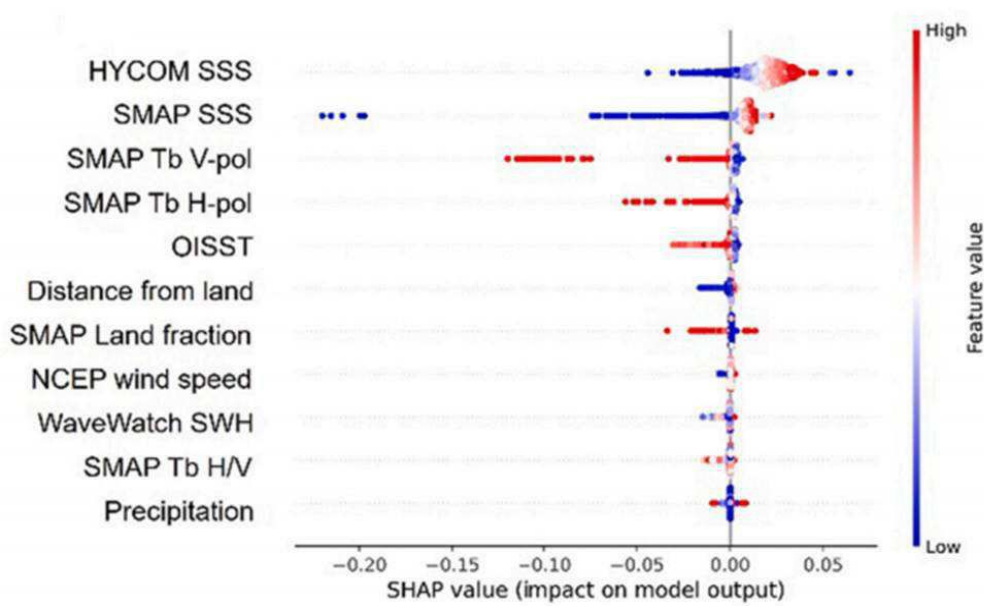
도면7



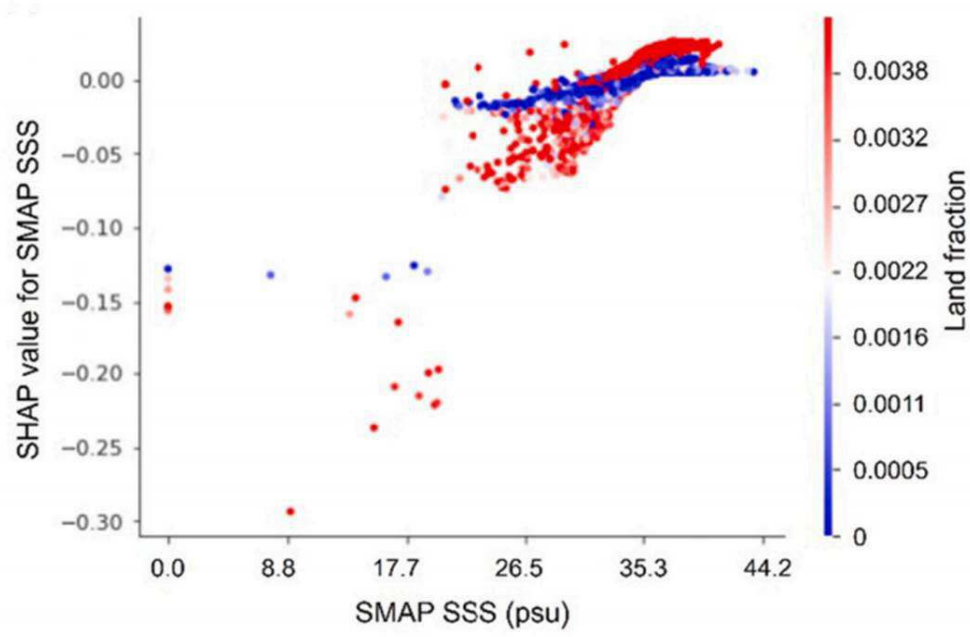
도면8a



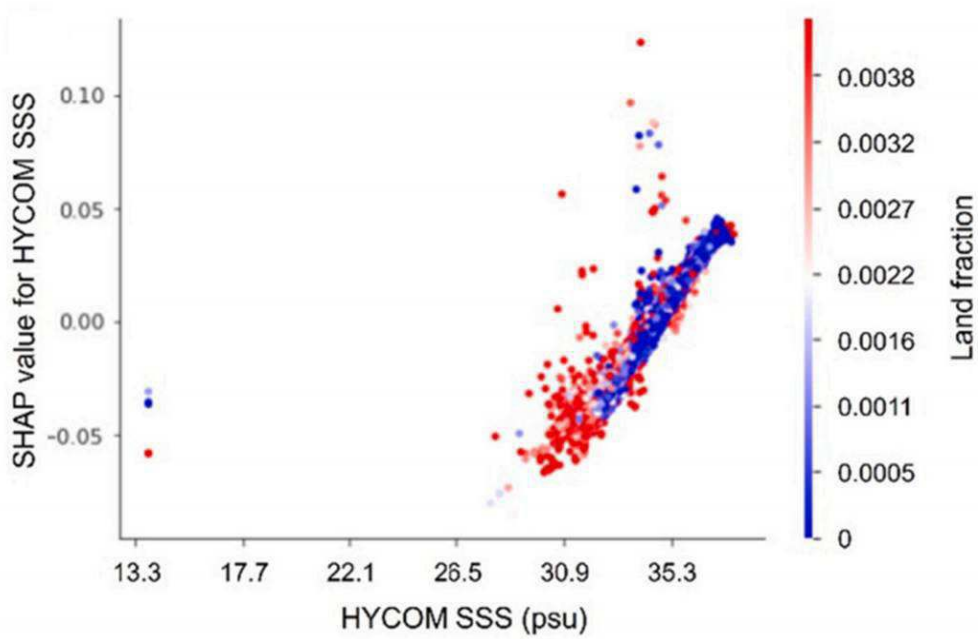
도면8b



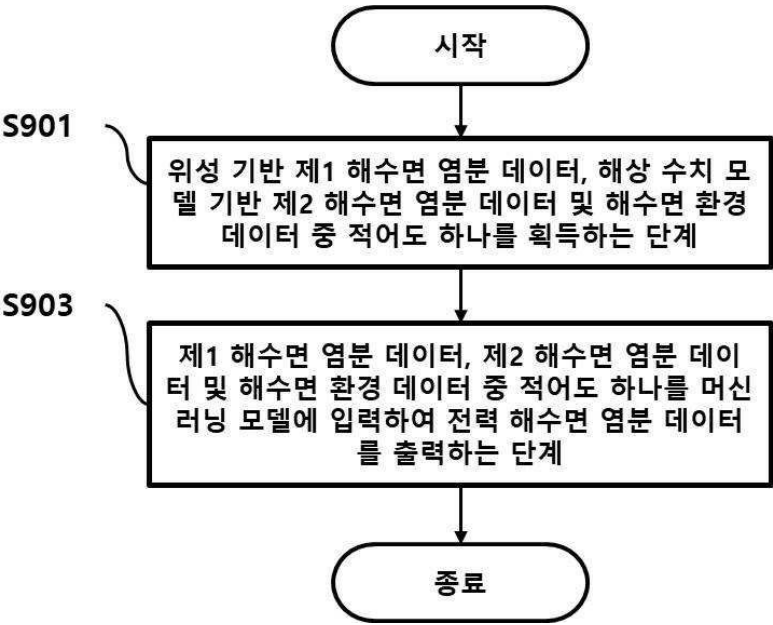
도면8c



도면8d



도면9



도면10

