





Article

다중 위성자료와 인공지능 기법을 이용한 한반도 주변 해역의 고해상도 해수면온도 자료 생산

정시훈 ¹⁾ · 추민기 ¹⁾ · 임정호 ^{2)†} · 조동진 ¹⁾

Generation of Daily High-resolution Sea Surface Temperature for the Seas around the Korean Peninsula Using Multi-satellite Data and Artificial Intelligence

Sihun Jung ¹⁾ · Minki Choo ¹⁾ · Jungho Im ^{2)†} · Dongjin Cho ¹⁾

Abstract: Although satellite-based sea surface temperature (SST) is advantageous for monitoring large areas, spatiotemporal data gaps frequently occur due to various environmental or mechanical causes. Thus, it is crucial to fill in the gaps to maximize its usability. In this study, daily SST composite fields with a resolution of 4 km were produced through a two-step machine learning approach using polar-orbiting and geostationary satellite SST data. The first step was SST reconstruction based on Data Interpolate Convolutional AutoEncoder (DINCAE) using multi-satellite-derived SST data. The second step improved the reconstructed SST targeting in situ measurements based on light gradient boosting machine (LGBM) to finally produce daily SST composite fields. The DINCAE model was validated using random masks for 50 days, whereas the LGBM model was evaluated using leave-one-year-out cross-validation (LOYOCV). The SST reconstruction accuracy was high, resulting in R^2 of 0.98, and a root-mean-square-error (RMSE) of 0.97°C. The accuracy increase by the second step was also high when compared to in situ measurements, resulting in an RMSE decrease of 0.21–0.29°C and an MAE decrease of 0.17–0.24°C. The SST composite fields generated using all in situ data in this study were comparable with the existing data assimilated SST composite fields. In addition, the LGBM model in the second step greatly reduced the overfitting, which was reported as a limitation in the previous study that used random forest. The spatial distribution of the corrected SST was similar to those of existing high resolution SST composite fields, revealing that spatial details of oceanic phenomena such as fronts, eddies and SST gradients were well simulated. This research demonstrated the potential to produce high resolution seamless SST composite fields using multi-satellite data and artificial intelligence.

Key Words: Sea surface temperature, Reconstruction, Artificial intelligence, DINCAE, LGBM, Korea peninsula

Received October 5, 2022; Revised October 17, 2022; Accepted October 21, 2022; Published online October 31, 2022

¹⁾ 울산과학기술원 도시환경공학과 석박사통합과정생(Combined MS/PhD Student, School of Urban and Environmental Engineering, Ulsan National Institute of Science and Technology, Ulsan, Republic of Korea)

²⁾ 울산과학기술원 도시환경공학부 교수(Professor, Department of Urban and Environmental Engineering, Ulsan National Institute of Science and Technology, Ulsan, Republic of Korea)

† Corresponding Author: Jungho Im (ersgis@unist.ac.kr)

This is an Open-Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<https://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

요약: 위성기반 해수면온도는 광역 모니터링이 가능한 장점이 있지만, 다양한 환경적 그리고 기계적 이유로 인한 시공간적 자료공백이 발생한다. 자료공백으로 인한 활용성의 한계가 있으므로, 공백이 없는 자료 생산이 필수적이다. 따라서 본 연구에서는 한반도 주변 해역에 대해 극궤도와 정지궤도 위성에서 생산되는 해수면온도 자료를 두 단계의 기계학습을 통해 융합하여 4 km의 공간해상도를 가지는 일별 해수면온도 합성장을 만들었다. 첫번째 복원 단계에서는 Data INterpolate Convolutional AutoEncoder (DINCAE) 모델을 이용하여 다중 위성기반 해수면온도 자료를 합성하여 복원하였고, 두번째 보정 단계에서는 복원된 해수면온도 자료를 현장관측자료에 맞춰 Light Gradient Boosting Machine (LGBM) 모델로 학습시켜 최종적인 일별 해수면온도 합성장을 만들었다. 개발된 모델의 검증을 위해 복원 단계에서 무작위 50일의 자료 중 일부분을 제거하여 복원된 뒤 제거된 영역에 대해 검증하였으며, 보정 단계에서는 Leave One Year Out Cross Validation (LOYOCV) 기법을 이용하여 현장자료와의 정확도를 검증하였다. DINCAE 모델의 해수면온도 복원 결과는 상당히 높은 정확도 ($R^2=0.98$, $bias=0.27^\circ\text{C}$, $RMSE=0.97^\circ\text{C}$, $MAE=0.73^\circ\text{C}$)를 보였다. 두번째 단계의 LGBM 보정 모델의 정확도 개선은 표층 뜰개 부이와 계류형 부이 현장자료와의 비교에서 모두 상당한 향상($RMSE=\Delta 0.21-0.29^\circ\text{C}$, $rRMSE=\Delta 0.91-1.65\%$, $MAE=\Delta 0.17-0.24^\circ\text{C}$)을 보여주었다. 특히, 모든 현장 자료를 이용한 보정 모델의 표층 뜰개 부이와의 정확도는 동일한 현장 자료가 동화된 기존 해수면온도 합성장보다 나은 정확도를 보였다. 또한 LGBM 보정 모델은 랜덤포레스트(random forest)를 사용한 선행연구에서 보고된 과적합의 문제를 상당부분 해결하였다. 보정된 해수면온도는 기존의 초고해상도 해수면온도 합성장들과 유사한 수준으로 수온 전선과 와동 등의 중규모 해양현상을 뚜렷하게 모의하였다. 본 연구는 다중위성 자료와 기계학습 기법을 사용해 시공간적 공백 없는 고해상도 해수면온도 합성장 제작 방법을 제시하였다는 점에서 가치가 있다.

1. 서론

주기적으로 생산되는 위성기반 해수면온도 자료는 넓은 영역에 대한 모니터링이 가능하며, 기계학습과 역학 기반의 해양 예측 모델의 주요한 입력자료로 사용된다(Good *et al.*, 2020; Jung *et al.*, 2020; Xiao *et al.*, 2019). 위성기반 해수면온도는 관측 센서의 파장에 따른 특성으로 인해 유효한 자료 양에서 차이가 발생한다. 열적외선 센서는 상세한 공간해상도의 해수면온도 관측이 가능하지만, 구름아래에서의 온도를 관측할 수 없다(Donlon *et al.*, 2012). 마이크로파 센서는 구름으로 가려진 영역의 해수면온도를 산출할 수 있지만, 공간해상도가 낮고, 강한 바람과 강수가 있는 경우 관측이 어렵다는 한계가 있다(Wentz *et al.*, 2000). 위성을 통한 해수면온도 산출은 위와 같은 관측의 한계로 인해 일부 지역에 대해 시공간적 자료공백이 발생해 그 활용성이 감소한다. 따라서, 해수면온도 자료의 활용성을 증가시키기 위해 공백이 없는 자료를 생산할 필요가 있다.

공백이 없는 자료를 생산하기 위해 자료동화, 신호처리 그리고 통계적인 방법 기반의 연구들이 수행되어 왔다(Beckers and Rixen, 2003; Chin *et al.*, 2017; Donlon *et al.*, 2012; Kim *et al.*, 2017; Park and Kim, 2009). 한국 기상청과

유럽중기예보센터(European Centre for Medium-Range Weather Forecasts) 수치예보모델의 입력자료로 사용되는 영국 기상청(Met Office)의 Operational Sea Surface Temperature and Sea Ice Analysis (OSTIA) 합성장 자료는 다중위성과 현장 관측 자료들을 사용해서 자료동화 기법 중 하나인 최적보간을 통해 0.05° 공간해상도로 생산되며 현장 자료와의 비교에서 높은 정확도를 보여준다(Donlon *et al.*, 2012). 미국 해양대기청(National Oceanic and Atmospheric Administration, NOAA)의 Multi-scale Ultra-high Resolution SST (MUR SST) 자료는 신호처리 방법인 웨이블릿 분해기반의 다중해상도 변동분석(multi-resolution variational analysis) 기법을 사용하여 공간해상도 0.01° 의 초고해상도로 생산된다(Chin *et al.*, 2017). 대표적인 통계적인 기법으로는 해수면온도, 클로로필 농도, 그리고 부유물질 농도 등 다양한 해양 파라미터 자료 복원 연구에서 우수한 성능을 보인 경험적 교함수 기반의 Data INterpolate Empirical Orthogonal Function (DINEOF)가 있다(Alvera-Azcárate *et al.*, 2016; Alvera-Azcárate *et al.*, 2007; Beckers and Rixen, 2003; Liu and Wang, 2018; Nechad *et al.*, 2011).

기존 방법들은 선형적인 연산, 노이즈 제거, 그리고 장기간 누락된 자료의 기후장 대체로 인해 복원된 결과

에서 평활화(smoothing)가 발생하여 메소스케일(meso-scale)의 해양현상을 모의함에 한계가 있다(Barth *et al.*, 2020; Donlon *et al.*, 2012). 최근에는 기계학습 기법이 발전함에 따라 기계학습 기반의 누락된 해양 자료 복원 연구가 활발히 진행되고 있다(Barth *et al.*, 2020; Ouala *et al.*, 2018; Sunder *et al.*, 2020). Barth *et al.* (2021)는 아노말리 기반의 최적보간(optimal interpolation) 개념을 활용한 합성곱 인공 신경망인 Data INterpolate Convolutional AutoEncoder (DINCAE)를 고안하였다(Barth *et al.*, 2020). DINCAE는 해양 분야에서 우수한 복원 성능을 보인 DINEOF 기법보다 높은 복원 정확도와 해양현상의 시공간 변동성을 더 잘 모의할 수 있음을 제시하였다(Barth *et al.*, 2020). DINCAE를 이용해 북서태평양 해수면온도를 복원한 연구에서 4 km 해상도의 복원된 해수면온도장은 수평해상도 50 km 수준의 메소스케일 해양현상을 복원할 수 있음을 보였다(Jung *et al.*, 2022). DINCAE는 외해뿐만 아니라 남중국해, 서해 그리고 지중해 등의 복잡한 연안선을 가진 연구지역에서도 해수면온도, 클로로필, 그리고 해수면높이 자료 복원에 사용되어 높은 정확도와 메소스케일의 해양현상을 복원하였다(Barth *et al.*, 2020; Barth *et al.*, 2022; Han *et al.*, 2020; Ji *et al.*, 2021; Jung *et al.*, 2022; Luo *et al.*, 2022).

해수면온도 합성장에 사용되는 위성기반 해수면온도는 낮 시간 관측의 일주기 가열(diurnal warming) 현상, 물리기반 산출식과 현장자료와의 측정 수심의 차이로 인한 편차(bias)가 발생한다(Merchant *et al.*, 1999; O'Carroll *et al.*, 2008). 그러므로, 해수면온도 합성장의 입력자료인 위성기반 해수면온도자료는 현장자료를 참조자료로 활용하여 편차를 보정한다(Donlon *et al.*, 2012; Stark *et al.*, 2007). Jung *et al.* (2022)는 DINCAE를 사용하여 다중위성 해수면온도를 복원하였으며, 앙상블 기계학습을 통해 복원된 해수면온도 자료를 융합하여 현장 자료와의 일치도를 높였다. 하지만, 해당 연구에서 사용된 랜덤포레스트 기계학습은 현장자료와의 검증 정확도를 향상시켰지만, 학습단계에서 다소 과적합(overfitting)을 야기하였다.

본 연구에서는 극궤도와 정지궤도를 포함한 다중 위성자료와 DINCAE를 사용해 다양한 해양환경 특성을 가진 한반도 주변 해역의 해수면온도를 일별 복원하였다. 이어 최근 많이 활용되고 있는 부스팅(boosting) 기반 기계

학습 방법 중 높은 학습 효율성과 효과적인 과적합 방지를 보이는 Light Gradient Boosting Machine (LGBM)을 이용하여 현장 관측자료를 타겟으로 하여 복원된 해수면온도를 학습시킴으로써 현장 관측자료에 대한 일치성을 향상시켰다(Choi *et al.*, 2021; Jang *et al.*, 2022; Kang *et al.*, 2022; Park *et al.*, 2021).

2. 연구 자료

1) 위성기반 해수면온도 자료

밤시간 해수면온도는 낮 시간에서와 같은 태양으로 인한 해수면 가열 현상이 없으므로, 낮보다 수직적 구조 안정성이 높다(Donlon *et al.*, 2002). 현존하는 해수면온도 합성장은 해수면온도의 일주기 가열과 냉각 특성을 보정한 자료를 사용하여 일 해수면온도 평균장(foundation SST)의 형태로 생산되기 때문에 본 연구에서도 동일하게 일 평균장을 생산하고자 한다(Chin *et al.*, 2017; Donlon *et al.*, 2012). 한반도 해역의 해수면온도 평균 합성장 생산을 위한 해수면온도 자료로 극궤도위성기반 L3S_LEO_PM-STAR-v2.80 (이후에는 L3S로 지칭), 정지궤도 위성 기반 Himawari-8, 그리고 GK-2A의 밤 시간 해수면온도 자료를 활용하였다. 미국 NOAA에서 제공하는 극궤도 다중 열적외 위성 기반의 L3S는 Moderate Resolution Imaging Spectroradiometer (MODIS) 와 Visible Infrared Imaging Radiometer Suite (VIIRS) 센서를 탑재한 위성 자료를 사용하였다. VIIRS 센서를 탑재한 Joint Polar Satellite System (JPSS) 와 NOAA-20위성과 MODIS 센서를 탑재한 Aqua위성의 Level 3 밤 시간(Local Solar Time [LST] 01:30 AM) 해수면온도 자료들을 합성(collate)하여 산출된다(Jonasson *et al.*, 2021). L3S 해수면온도 자료는 다중위성 swath 자료들을 정규격자에 합성할 때 발생하는 공간적 불연속성(spatial discontinuity)과 구름으로 인한 노이즈 제거를 위해 공간 평균을 활용해 편차(bias) 보정하였다(Jonasson *et al.*, 2021). L3S 아피층(Sub-skin) 해수면온도 자료는 2 km의 공간해상도와 하루의 시간해상도를 가진다(NOAA/STAR, 2021). 2012년 2월 1일부터 2021년 12월 31일까지의 L3S 해수면온도 자료를 수집하였으며 해당 자료는 NASA Physical Oceanography Distributed Active Archive Center (PODAAC)에서 받을

수 있다(https://podaac.jpl.nasa.gov/dataset/L3S_LEO_PM-STAR-v2.80).

일본 기상청은 정지궤도 기상위성인 Himawari-8 위성의 Advanced Himawari Imager (AHI) 센서로부터 수집된 자료를 이용하여 준물리기반 알고리즘을 통해 아피층 해수면온도를 산출한다(Kurihara *et al.*, 2016). Himawari-8 해수면온도 자료는 2 km의 공간해상도와 1시간의 시간해상도를 가진다. L3S 해수면온도 자료의 생산에 사용되는 극궤도 위성들이 한반도 지역을 관측하는 시간은 지방시로 평균 02:00 AM이다(Levy *et al.*, 2018). 그러므로, L3S 자료와 동일한 관측 시간대인 UTC 17시(KST 02:00 AM)의 Level 3 밤 시간 SST 자료를 2015년 7월 7일부터 2021년 12월 31일까지 수집하였으며, 해당 자료는 일본 우주항공연구개발기구(Japan Aerospace Exploration Agency)에서 받을 수 있다(<https://earth.jaxa.jp/en/data/index.html>).

한국 기상청(Korea Meteorological Administration, KMA)에서 제공하는 해수면온도 자료는 정지궤도 기상위성인 천리안 2A호(GeoKompsat-2A, GK-2A)의 Advanced Meteorological Imager (AMI) 센서로부터 관측된 자료를 이용하여 생산된다(Park, 2019). GK-2A 해수면온도 자료

는 2 km의 공간해상도와 10분의 시간해상도를 가진다. 연구기간 중 해당 위성이 가용한 2019년 7월 19일부터 2021년 12월 31일까지 UTC 17시(KST 02:00 AM)의 해수면온도 자료를 수집하였다(<https://nmsc.kma.go.kr/>).

2) 현장관측 자료

본 연구에서 생산되는 해수면온도 합성장 자료를 검보정하기 위해 새벽 2시(UTC 17시)에 관측된 기상청 해양 기상 부이(mooring buoy), in situ QUALity Monitor (iQUAM)의 표층 뜰개 부이(drifting buoy)에서 수집한 해수면온도 자료를 사용하였다. 본 연구의 보정에는 2012년부터 2021년까지의 iQUAM과 기상청 부이자료를 사용하였으며, 검증에는 2012년부터 2021년까지 iQUAM, 기상청 그리고 국립해양조사원의 현장관측 자료 모두를 사용하였다. iQUAM과 기상청 현장관측 자료는 기존 해수면온도 합성장자료 생산에 활용된다. 반면, 국립해양조사원의 계류형 부이자료는 기존 합성장자료 생산에 사용되지 않으므로 기존 합성장자료와 동일한 비교를 위한 검증용으로만 활용하였다. 기상청 해양 기상 부이는 한반도 주변 해역의 20개 관측 지점에서 해수면온도를 수집하고 있으며, 관측 수심은 0.2-

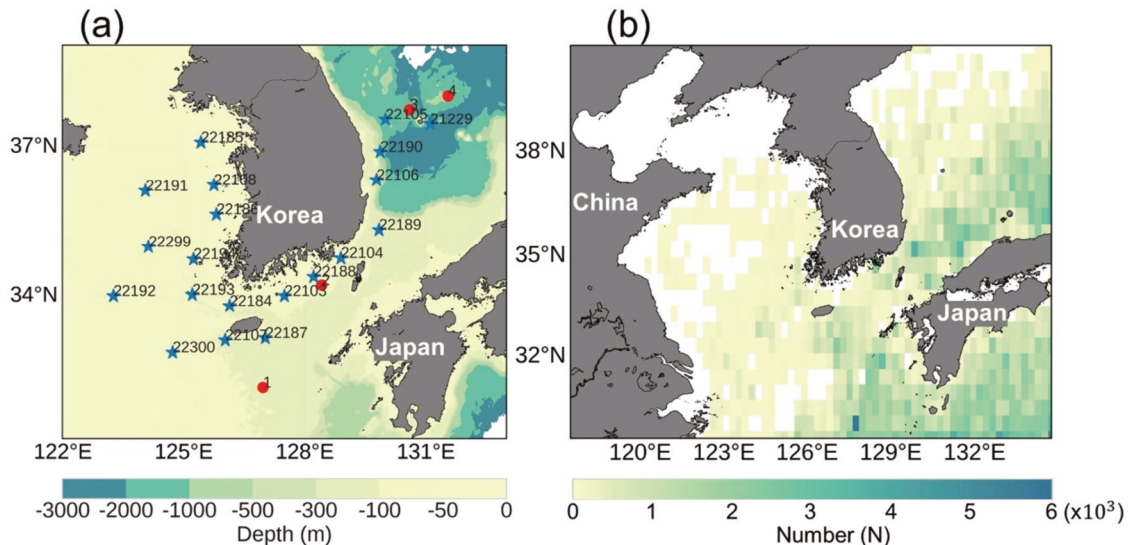


Fig. 1. Map of the in situ measurements for the validation and correction. (a) location of two types of fixed (or two types of mooring buoys) in situ monitoring sites. The Korea Meteorological Administration (KMA) buoys are shown in blue stars, while the Korea Hydrographic and Oceanographic Agency (KHOA) buoys are shown in red circles with their IDs as the labels. The bathymetry downloaded from General Bathymetry Chart of the Oceans (GEBCO) (www.gebco.net) was used as the background image. (b) accumulated acquisition of iQUAM drifting buoys. The density is aggregated into 0.25°.

Supplementary Table 1. The location of Korea Meteorological Administration (KMA) and Korea Hydrographic and Oceanographic Agency (KMOA) mooring buoys

Service administration	Buoy name	ID	Location		Validation time period
			Latitude (°N)	Longitude (°E)	
KMA	Ulleungdo	21229	37.4554	131.1144	01 Jan 2012 to 31 Dec 2021
	Donghae	22105	37.5442	130.0	06 Feb 2012 to 31 Dec 2021
	Pohang	22106	36.35	129.7833	01 Jan 2012 to 31 Dec 2021
	Ulsan	22189	35.3453	129.8414	22 Dec 2015 to 31 Dec 2021
	Uljjin	22190	36.9069	129.8744	09 Dec 2015 to 31 Dec 2021
	Geomundo	22103	34.0014	127.5014	01 Jan 2012 to 31 Dec 2021
	Geojeodo	22104	34.7667	128.9	01 Jan 2012 to 31 Dec 2021
	Marado	22107	33.0833	126.0333	01 Jan 2012 to 31 Dec 2021
	Chujado	22184	33.7936	126.1411	15 Jan 2014 to 31 Dec 2021
	Seogwipo	22187	33.1281	127.0228	22 Dec 2015 to 31 Dec 2021
	Tongyeong	22188	34.3917	128.225	22 Dec 2015 to 31 Dec 2021
	Namhae 239	22300	32.83	124.73	01 Dec 2015 to 31 Dec 2021
	Oeyeondo	22108	36.25	125.75	01 Jan 2012 to 31 Dec 2021
	Incheon	22185	37.0917	125.4289	22 Dec 2015 to 31 Dec 2021
	Buan	22186	35.6586	125.8139	22 Dec 2015 to 31 Dec 2021
	Seohae 170	22191	36.1333	124.0569	01 Jan 2020 to 31 Dec 2021
	Seohae 206	22192	34.0	123.2625	01 Jan 2020 to 31 Dec 2021
	KHOA	Gageodo	22193	34.0275	125.2147
Hongdo		22194	34.7467	125.2456	01 Jan 2020 to 24 Feb 2020
Seohae 190		22299	35.0	124.13	01 Jan 2021 to 31 Dec 2021
Jeju nambu		1	32.0902	126.9658	01 Jan 2012 to 31 Dec 2021
	Namhae dongbu	2	34.2222	128.4188	01 Jan 2012 to 31 Dec 2021
	Ulleungdo bukseo	3	37.7425	130.6011	01 Jan 2012 to 31 Dec 2021
	Ulleungdo bukdong	4	38.0072	131.5525	23 Apr 2012 to 28 Dec 2021

0.4 m이다. 국립해양조사원의 해양 관측 부이는 기상청 부이와 표층 뜰개 부이보다 상대적으로 깊은 수심(~3 m)에서 온도를 관측하며, 본 연구에서는 동해, 남해의 4 개 지점에서 관측된 자료를 보정을 제외한 검증에만 사용하였다. 표층 뜰개 부이는 수심 0.2 m에서 해수면온도를 측정하기 때문에 위성기반 해수면온도 자료들의 검증 자료로 주로 활용되어 왔다(Donlon *et al.*, 2002; Xu and Ignatov, 2014; 2016). 그러므로, 본 연구에서는 위성 해수면온도 자료의 품질 관리 데이터 베이스인 iQUAM의 표층 뜰개 부이 자료를 사용하였다(Xu and Ignatov, 2014; 2016). 수집된 기상청과 국립해양조사원의 부이 자료의 위치정보와 사용기간은 Supplementary Table 1에 자세히 나타내었고, 부이의 위치와 표층 뜰개 부이 자료의 밀도는 Fig. 1과 같다.

3. 방법

본 연구에서는 두 단계로 나누어 다중위성자료와 기계학습 기법을 이용하여 일별 해수면온도 합성장을 만들었다(Fig. 2). 첫 단계는 DINCAE를 사용하여 다중위성자료로부터 자료공백이 있는 해수면온도를 시공간적으로 복원하는 것이고, 이어 두 번째 단계는 LGBM을 이용하여 복원된 해수면온도를 현장관측 자료에 맞춰 보정하는 것이다. 위성기반 아피층 해수면온도를 현장 관측 부이 자료에 맞춘 보정이므로, 보정된 해수면온도는 현장 자료 관측 수심에 맞춰진 표층 해수면온도 합성장이다. 개발된 해수면온도 합성장 자료를 국립해양조사원(KHOA) 현장관측 자료로 정확도를 평가하였고, 현존하는 고해상도 해수면온도 합성장인 MUR SST, OSTIA 자료와 공간분포를 비교하였다.

1) 해수면온도 복원 및 보정 모델 구축

첫 번째 단계(Step 1)에서 우선 DINCAE 모델의 입력 자료인 극궤도와 정지궤도 위성기반 해수면온도 자료를 동일하게 L3S 2 km 격자에 맞춘 뒤 합성하였다. 다중 위성 자료를 합성하는 경우 각 위성자료의 특성과 처리 방법이 달라서 해수면온도의 차이가 발생한다(Jonasson *et al.*, 2021). 그러므로, 해수면온도 합성장을 만들 때에는 각 위성 해수면온도와 현장관측자료 또는 배경장과의 차이를 입력 위성자료의 가중치로 사용하여 합성하

였다(Donlon *et al.*, 2012; Kim *et al.*, 2017). 본 연구에서는 다중위성 해수면온도를 합성할 때, 위성자료간 차이를 줄이는 전처리(편차보정 및 정규화)가 완료된 L3S를 우선적으로 사용하였으며, 극궤도 위성이 관측하지 못한 영역에 대해 유효한 정지궤도 위성 기반 해수면온도 자료들을 평균하여 사용하였다. 이어 연산자원의 한계로 인해 2 km 해상도로 생산된 입력자료를 선형보간 기법을 이용하여 4 km 해상도로 재격자화 하였다. 선행연구에서 4 km 해상도의 복원된 해수면온도 합성장은 50 km 수준의 메소스케일 해양현상을 모의할 수 있음을

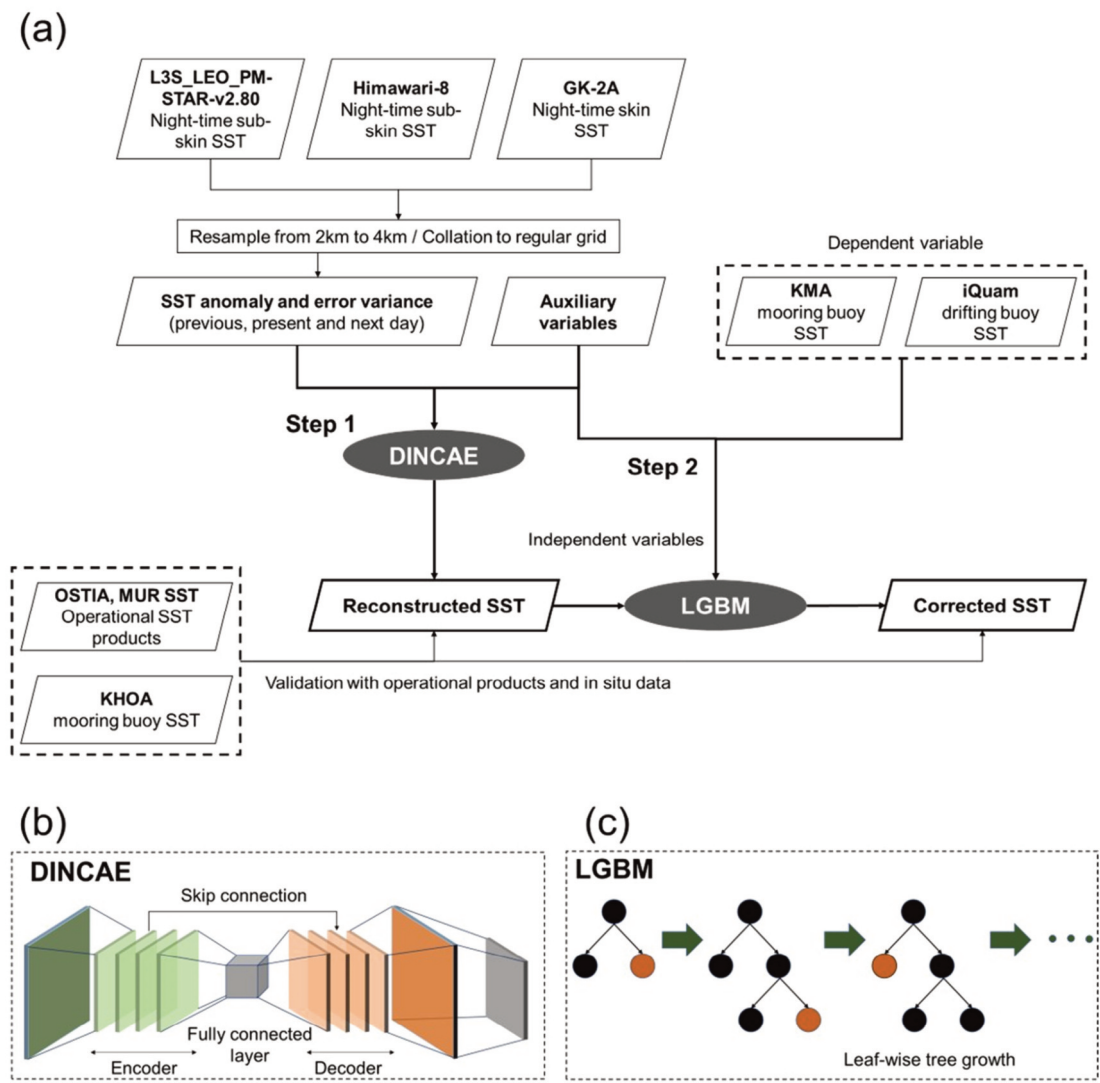


Fig. 2. Overall process and structure of machine learning in this study. (a) flow diagram of the proposed reconstruction and correction methods, (b) structure of the Data INterpolate Convolutional AutoEncoder (DINCAE) model, and (c) structure of the Light Gradient Boosting Machine (LGBM) model.

Table 1. Input variables used in the proposed two models: Data INterpolate Convolutional AutoEncoder (DINCAE) and Light Gradient Boosting Machine (LGBM)

Model	Variable
DINCAE	Anomalies of SST scaled by the inverse of the error variance of the present day
	Inverse of the error variance of the present day
	Anomalies of SST scaled by the inverse of the error variance of the previous day
	Inverse of error variance of the previous day
	Anomalies of SST scaled by the inverse of the error variance of the next day
	Inverse of error variance of the next day
	Normalized longitude (from -1 to 1)
	Normalized latitude (from -1 to 1)
	Cosine of the day of the year divided by 365.25
	Sine of the day of the year divided by 365.25
LGBM	Reconstructed SST from DINCAE
	Normalized latitude (from -1 to 1)
	Sine of the day of the year divided by 365.25

보였으므로, 본 연구에서도 입력자료와 합성장의 공간 해상도를 4 km에 맞추었다(Jung *et al.*, 2022). 기존 합성장들은 구름아래 관측이 가능한 마이크로파센서 기반 해수면온도 자료를 사용하여 합성되는 자료의 양을 증가시켰다(Donlon *et al.*, 2012). 하지만, 마이크로파 기반 해수면온도 자료들은 육지에서 100 km 이내를 관측할 수 없으며, 낮은 해상도(~25 km)로 인해 본 연구지역에는 적합하지 않으므로 사용하지 않았다(Wentz *et al.*, 2000). 합성한 해수면온도 자료를 사용해 DINCAE의 입력자료인 아노말리와 오차분산(error variance)을 계산하였다(Barth *et al.*, 2020). DINCAE 모델의 입력자료는 Table 1에 정리하였다.

두 번째 단계(Step 2)에서는 복원된 해수면온도 자료와 현장관측 자료 사이의 일치도를 높이기 위해 현장관측 자료(iQUAM의 표층 뜰개 부이와 기상청의 계류형 부이자료)를 타겟으로 한 학습을 진행하였다. 현장자료의 관측지점에 해당되는 위성격자 해수면온도 자료를 추출하여 입력자료로 사용하였다. 입력자료로 활용한 위성기반 해수면온도는 아피층(sub-skin) 자료이므로 부이가 측정하는 0.2-0.4 m 깊이의 표층 해수면온도와 차이가 있다. 이러한 차이는 해상풍(sea surface wind)의 영향을 받게 되므로 보정할 필요가 있다(Donlon *et al.*, 2002). 하지만, 위성기반 해상풍 자료는 좁은 관측 범위(swath), 그리고 긴 관측 주기(revisit time)로 인해 준 실시간으로 제공되지 않는다. 또한, 낮은 공간해상도의 해상

풍 자료를 사용하여 연구지역에 대한 피층-표층의 온도차이를 보정하기에는 그 오차가 크기 때문에 본 연구에서는 사용하지 않았다(Park *et al.*, 2018). 최종적으로 LGBM 모델에서 생산되는 보정된 표층 해수면온도(corrected SST), 즉 일별 해수면온도 합성장은 첫 번째 단계에서 복원된 해수면온도와 동일한 시공간 해상도를 가진다. 보정단계에서 사용된 LGBM 모델의 입력자료는 Table 1에 정리하였다.

2) 기계학습 모델

(1) Data INterpolate Convolutional AutoEncoder (DINCAE)

Barth *et al.* (2020)가 제안한 DINCAE 모델은 일부분의 유효한 자료를 통해 누락된 지역을 복원한다(Barth *et al.*, 2020). DINCAE 모델은 시계열 자료의 아노말리를 합성곱 인공 신경망의 입력자료로 활용하였다. DINCAE 모델은 합성곱 인공 신경망 구조 중 차원 감소(dimension reduction)를 통해 주요한 특징(feature)을 추출하는데 우수한 성능을 보이는 오토인코더(autoencoder) 구조를 사용한다(Barth *et al.*, 2020). 본 연구에서는 각각 4개의 합성곱 인공 신경망 층으로 이루어진 인코더(encoder)와 디코더(decoder), 그리고 완전 연결층(fully connected)으로 DINCAE 모델을 구성하였다. DINCAE 모델은 각 배치(mini-batch)마다 아노말리 입력자료에서 일부분을 제거하여 정답(reference)으로 활용하여, 그 차이를 줄이

는 방향으로 학습한다. 더 나아가 현재시점의 복원 정확도를 향상시키기 위해 전날과 다음날의 시계열 정보를 추가적인 입력자료로 활용하였다(Barth *et al.*, 2020; Jung *et al.*, 2022). 입력자료(Table 1)들을 통해 DINCAE 모델에서 산출되는 결과물은 두가지로, 첫 번째는 예측된 오차 분산의 역수(T_{ij}) 그리고 두 번째는 아노말리를 오차분산으로 나눈 값(T_{ij2})이다(식(1), (2)). 예측된 오차 분산($\hat{\sigma}_{ij}^2$)은 식(1)의 연산을 통해 산출되며, 복원된 해수면온도의 정확도를 나타낸다. 예측된 아노말리(\hat{y}_{ij})는 식(2)에 따라 계산되며, 복원된 해수면온도는 예측된 아노말리와 $mean(SSTs)$ 은 전체 연구기간의 해수면온도 자료의 평균을 더하여 산출한다.

$$\hat{\sigma}_{ij}^2 = \frac{1}{\text{maximum}(\text{exponential}(\text{minimum}(T_{ij}, \gamma)), \delta)} \quad (1)$$

$$\hat{y}_{ij} = T_{ij2} * \hat{\sigma}_{ij}^2 \quad (2)$$

$$R_{ij} = \hat{y}_{ij} + \text{mean}(SSTs) \quad (3)$$

위 식에서 $\gamma=10$, 그리고 $\delta=10^{-3}C^{-2}$ 이다. i, j 첨자는 각 픽셀의 위치정보이다.

DINCAE 모델은 Geohydrodynamics and environment research Github (<https://github.com/gher-ulg/DINCAE>)에서 제공되며, Python3언어와 tensorflow 패키지로 구성되어 있다(Barth *et al.*, 2020). DINCAE 모델에는 다양한 하이퍼파라미터가 존재하는데 학습의 최적화까지 소요되는 연산시간과 연산량에 영향을 주는 Number of convolution layers of the encoder and decoder 파라미터를 다양한 값들로 테스트한 뒤 가장 적은 연산시간을 보인 값을 우선적으로 선택하였다. 이후, 나머지 하이퍼파라미터에 대해 다양한 조합을 시험하여 가장 우수한 복원

Table 2. Hyperparameters determined in the DINCAE model

Hyper-parameters	Selection
Window of timeseries SST	3 days
Epoch	600
Image resize method in the decoder layers	Nearest neighbor
Pooling layer in the encoder layers	Average
Number of convolution layers of the encoder and decoder	4 / 4
Number of filters for the encoder (same in decoder layers)	16 / 24 / 36 / 54
Activation function of convolutional layers	Leaky ReLu
Optimizer	Adam

정확도와 해양현상 모의수준을 보인 하이퍼파라미터 조합을 선정하였다(Table 2).

(2) Light Gradient Boosting Machine (LGBM)

기존 부스팅 방법인 gradient boosting decision tree (GBDT) 알고리즘은 입력 자료의 종류와 수가 증가하면 순차적인(sequential) 연산으로 인해 소요시간이 길어지며, 효율성과 확장성이 감소한다. 이에 비해 LGBM 모델은 Gradient-based One Side Sampling (GOSS) 부스팅 방법과 Exclusive Feature Bundling (EFB)를 적용하여, 기존 부스팅(i.e., GBDT) 방법보다 효율성과 확장성이 우수하다(Ke *et al.*, 2017). LGBM 모델 구현을 위해 Python3 환경에서 구동가능한 LGBMRegressor을 사용하였다. LGBM 모델도 다양한 하이퍼파라미터를 사용하는데, 대표적인 것으로 maximum depth of tree, minimum data in leaf, number of leaves, 그리고 number of trees가 있다. Maximum depth of tree와 minimum data in leaf는 각각 학습하는 트리들의 깊이와 한 개의 leaf에 할당되는 샘플의 수를 조절하는 파라미터이며, 과적합을 조절할 수 있는 중요한 파라미터이다. Number of leaves는 leaf-wise 모델인 LGBM에서 적합한 학습을 위한 가장 중요한 변수이다. 종종 과적합을 야기하는 깊은 트리로부터 발전하는 것을 막기 위해 $2^{\text{Maximum depth of tree}}$ 보다 작게 설정할 필요가 있다(Park *et al.*, 2021). 모든 하이퍼파라미터는 grid search 방법을 사용해 결정하였다(Table 3). Grid search는 다수의 하이퍼파라미터 후보값 조합에 대한 교차 검증을 통해 최적의 하이퍼파라미터를 도출한다. 본 연구에서는 DINCAE와 LGBM 모델 구동을 위해 Intel (R) Xeon (R) Silver 4215R CPU @ 3.20GHz와 NVidia Quadro RTX 8000 GPU (48 GB of memory)를 사용하였다.

Table 3. Hyperparameters determined in the LGBM model

Hyper-parameters	Selection
Boosting type	GOSS
Maximum depth of tree	12
Minimum data in leaf	100
Number of trees	100
Number of leaves	28

3) 검증 방법

첫번째 단계의 DINCAE 모델의 복원 성능을 평가하

기 위해 입력자료의 일정 부분을 임의로 제거한 뒤 이를 검증자료로 사용하였다. 우선 2012년 1월 1일부터 50 일간의 자료에서 구름 마스크 부분을 추출하였다. 추출한 50개의 구름 마스크를 연구기간 중 랜덤하게 50일의 자료에 적용하여 유효한 해수면온도 값이 있는 부분을 제거한 뒤, 이를 DINCAE 모델의 복원 정확도 평가에 사용하였다. 두번째 단계에서 보정 모델인 LGBM의 성능을 검증하기 위해 시간적 교차검증 방법인 Leave-One-Year-Out-Cross-Validation (LOYOCV)을 사용하였다. LOYOCV는 전체 연구기간 중 특정 해의 자료를 제외하고 나머지 해의 자료로 학습하여 제외된 일년 자료를 검증하는 교차 방법이다. 연구 기간이 10년이므로 LOYOCV를 10회 반복해서 수행한다. 개발된 해수면온도 자료를 평가하기 위해 아래의 다섯가지 정확도 지수를 통해 검증 결과를 정량적으로 도출하였다(식(4)-(8)).

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}, \bar{y} = \frac{\sum_{i=1}^n y_i}{n} \quad (4)$$

$$Bias(^{\circ}C) = \frac{\sum_{i=1}^n (\hat{y}_i - y_i)}{n} \quad (5)$$

$$RMSE(^{\circ}C) = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{n}} \quad (6)$$

$$rRMSE(\%) = \frac{100 * RMSE}{\bar{y}_i} \quad (7)$$

$$MAE(^{\circ}C) = \sum_{i=1}^n |\hat{y}_i - y_i| \quad (8)$$

수식의 n 은 정확도 계산에 사용되는 샘플 수, y_i 는 참값

(보간: 검증을 위해 제거된 부분; 보정: 현장 관측 자료), \hat{y}_i 는 모델로부터 산출된 예측값, 그리고 \bar{y}_i 는 참값의 평균이다.

4. 결과 및 토의

1) 해수면온도 복원 검증

Fig. 3은 임의로 제거된 50일의 검증 자료에 대한 해수면온도 복원 정확도를 산점도와 오차 공간분포로 표현한 그림이다. Fig. 3(a)에 따르면 복원된 결과와 검증 자료는 높은 상관성($R^2=0.98$, $bias=0.27^{\circ}C$, $RMSE=0.97^{\circ}C$, $MAE=0.73^{\circ}C$)을 보인다. 위 결과는 서해가 연구지역이며 동일한 극궤도 해수면온도 자료를 사용한 선행 연구($R^2=0.94$, $RMSE=1.09^{\circ}C$) 보다 상대적으로 높은 복원 정확도를 보여주었다(Ji *et al.*, 2021). 복원 결과(Fig. 3(a))에서 해수면온도가 $5^{\circ}C$ 이하인 경우 과대추정 하였으나, $5^{\circ}C$ 이상인 경우에는 과대 또는 과소추정 경향이 보이지 않았다. 복원 결과에서 전반적으로 과소추정하는 경향을 보이는 선행연구보다 안정된 결과를 보였다(Ji *et al.*, 2021; Jung *et al.*, 2022). 이러한 차이는 입력자료인 아노말리를 생산하기 위한 해수면온도 평균장이 연구기간(본 연구: 10년, 선행연구: 10년 이하)에 따라 모델의 학습도에 영향을 준 것으로 판단된다. 열적외 센서 기반 자료를 사용할 경우 높은 자료 소실률로 인해 평균장 생산에 영향을 주기 때문에 연구기간이 늘어남에 따

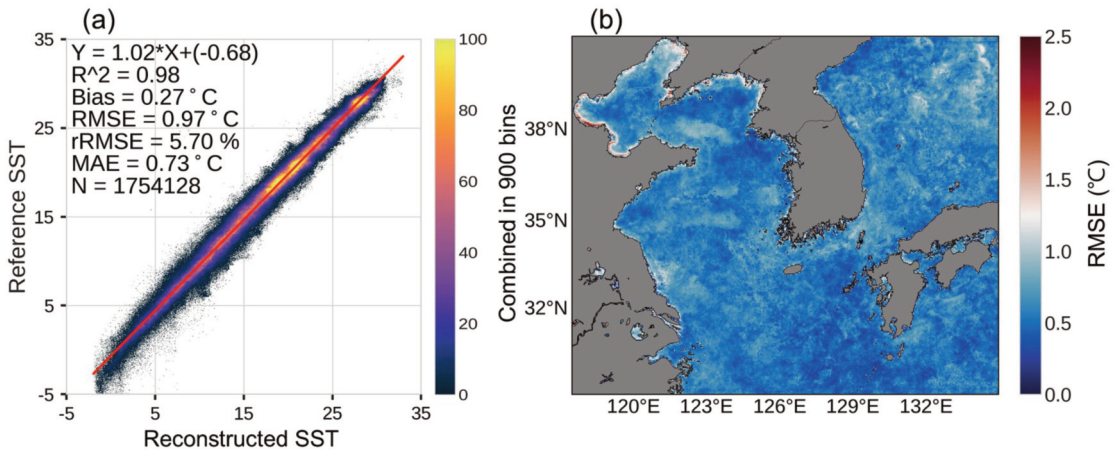


Fig. 3. Validation accuracy of the SST reconstructed through DINCAE. (a) scatterplot between the reconstructed and reference SST data and (b) spatial distribution of validation RMSE.

라 모델의 학습도가 상승될 수 있다(Ji *et al.*, 2021; Jung *et al.*, 2022; O'Carroll *et al.*, 2019). Fig. 3(b)는 복원 정확도 중 RMSE의 공간분포이며, 연안지역이 외해지역보다 상대적으로 높은 오차를 보인다. 특히, 보하이해(bohai sea) 연안의 복원 정확도는 RMSE가 1.2°C 이상으로 다른 지역과 비교해서 높은 오차를 보인다. 시계열 아노말리를 통해 복원하는 DINCAE 모델은 장기간 일정한 변동성을 가진 지역에 대한 모의능력은 우수하지만, 시공간 변동성이 큰 연안에서는 복원 정확도가 감소하는 경향을 보여주었다(Jung *et al.*, 2022). 반면, 위도가 감소함에 따라 복원 정확도는 증가하는데, 이는 선행연구와 유사한 경향을 보인다(Jung *et al.*, 2022). 저위도로 갈수록 해수면온도의 시공간적 변동성이 감소되므로 모델이 모의해야 하는 아노말리의 범위가 감소하게 되며, 이는 다시 정확도 향상에 영향을 준 것으로 판단된다(Bulgin *et al.*, 2020; Jung *et al.*, 2022).

Fig. 4는 원본, 검증자료, 복원된 해수면온도의 예시이다. 복원된 해수면온도(Fig. 4(c), (f))를 보면, 울릉도 부근과 일본 남부의 쿠루시오 해류의 전선, 와동, 그리고 온도 구배가 원본(Fig. 4(a), (d))과 유사하게 복원되었음을 알 수 있다. 원본(Fig. 4(d))의 보하이만 연안은 외해보

다 상대적으로 낮은 해수면온도를 보이는데, 복원 결과(Fig. 4(f))에서도 비슷한 공간 분포를 보인다.

2) 현장자료와의 비교

Table 4는 기존 해수면온도 합성장과 본 연구에서 개발된 해수면온도 합성장을 표층 뜰개 부이(iQUAM drifting buoy) 현장관측 자료와 비교한 정확도를 보여주고 있다. 복원된 해수면온도(Reconstructed SST: $R^2=0.92$, $bias=-0.53^\circ\text{C}$, $RMSE=1.45^\circ\text{C}$, $rRMSE=6.31\%$, $MAE=1.14^\circ\text{C}$)는 OSTIA, MUR SST보다 상대적으로 다소 낮은 정확도를 보인다. 이는 기존 합성장 자료들을 생산할 때 검증으로 사용된 표층 뜰개 부이 자료들을 활용하기 때문이다(Chin *et al.*, 2017; Donlon *et al.*, 2012; Stark *et al.*, 2007). LGBM을 통해 보정된 해수면온도의 정확도(Corrected SST: $R^2=0.92$, $bias=0.10^\circ\text{C}$, $RMSE=1.23^\circ\text{C}$, $rRMSE=5.35\%$, $MAE=0.97^\circ\text{C}$)는 복원된 해수면온도보다 향상($RMSE=\Delta 0.22^\circ\text{C}$, $rRMSE=\Delta 0.93\%$, $MAE=\Delta 0.17^\circ\text{C}$)되었다. 기존 합성장 자료와 동일한 조건으로 모든 현장 관측 자료를 활용한 모델은 OSTIA와 MUR SST보다 상대적으로 높은 정확도($R^2=0.97$, $bias=0.00^\circ\text{C}$, $RMSE=0.81^\circ\text{C}$, $rRMSE=3.52\%$, $MAE=0.61^\circ\text{C}$)를 보인다. 또한, 본 연구에서 사용

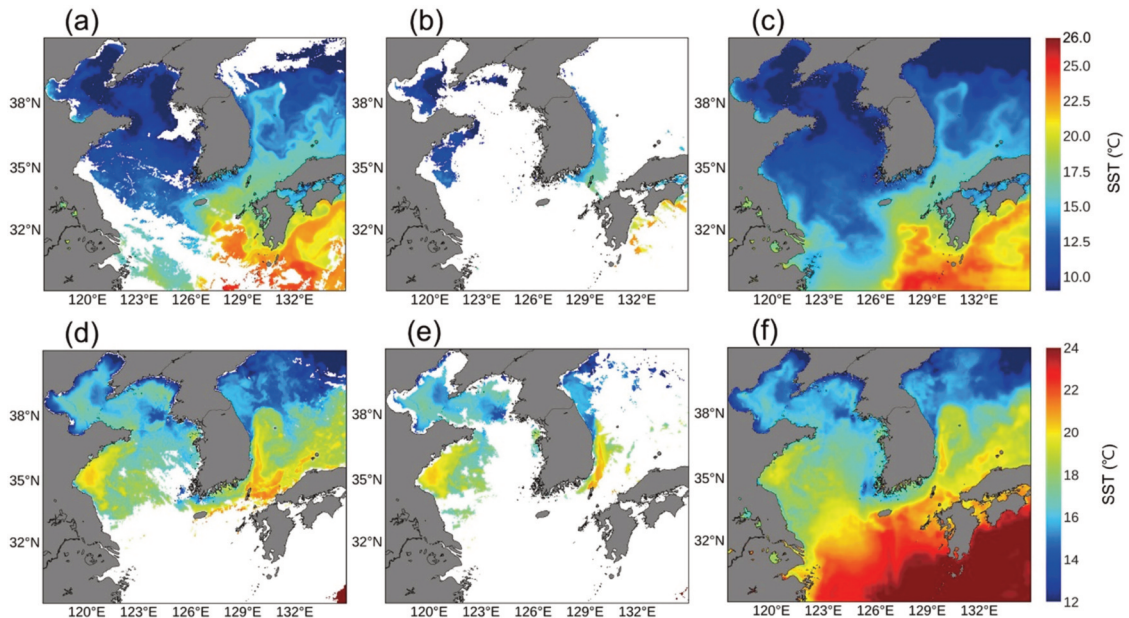


Fig. 4. Spatial distribution of SSTs. (a) original SST, (b) original SST with the occlusion mask for validation, and (c) the reconstructed SST on 5 May 2013. (d) original SST, (e) original SST with the occlusion mask for validation, and (f) the reconstructed SST on 30 Oct. 2015.

Table 4. Accuracy metrics of the operational SST products and SSTs produced in this study when compared to iQUAM drifting buoy data

SSTs	Coefficient of determination (R ²)	Bias (°C)	RMSE (°C)	rRMSE (%)	MAE (°C)	The number of data
OSTIA	0.94	0.00	1.07	4.69	0.81	627,981
MUR SST	0.94	0.08	1.13	4.93	0.86	
Reconstructed SST (step1)	0.92	-0.53	1.45	6.31	1.14	
Corrected SST (step2)	0.92	0.10	1.23	5.35	0.97	
Corrected SST using entire in situ data (step2)	0.97	0.00	0.81	3.52	0.61	

Table 5. Accuracy metrics of the operational SST products and SSTs produced in this study when compared to KMA mooring buoy data

SSTs	Coefficient of determination (R ²)	Bias (°C)	RMSE (°C)	rRMSE (%)	MAE (°C)	The number of data
OSTIA	0.98	-0.01	0.69	3.88	0.49	40,765
MUR SST	0.98	0.15	0.85	4.78	0.59	
Reconstructed SST (step1)	0.97	-0.64	1.30	7.30	1.01	
Corrected SST (step2)	0.96	0.05	1.09	6.11	0.84	
Corrected SST using entire in situ data (step2)	0.97	0.06	0.99	5.56	0.76	

Table 6. Accuracy metrics of the operational products and SSTs produced in this study when compared to KHOA buoy data

SSTs	Coefficient of determination (R ²)	Bias (°C)	RMSE (°C)	rRMSE (%)	MAE (°C)	The number of data
OSTIA	0.97	-0.01	0.74	4.14	0.54	11,493
MUR SST	0.97	0.11	0.80	4.46	0.57	
Reconstructed SST (step1)	0.94	-0.79	1.49	8.33	1.16	
Corrected SST of training entire data (step2)	0.94	0.38	1.20	6.68	0.92	

한 LGBM 모델은 선행연구(Jung *et al.*, 2022)에서 사용한 랜덤포레스트의 문제점인 과적합의 정도(RMSE=Δ0.71°C, rRMSE=Δ3.89%, MAE=Δ0.53°C; 본 연구: RMSE=Δ0.42°C, rRMSE=Δ1.83%, MAE=Δ0.36°C)를 상당히 감소시켰다.

Table 5는 해수면온도 자료들을 기상청의 계류형 부이 현장관측 자료와 비교한 정확도를 보여주고 있다. 기상청의 계류형 부이 자료와의 정확도는 표층 뜰개 부이 자료와의 검증 결과와 동일하게 LGBM 모델을 통해 뚜렷하게 향상됨을 보여주었다. 보정된 해수면온도(단계 2)는 복원된 해수면온도(단계 1)보다 향상된 정확도(RMSE=Δ0.21°C, rRMSE=Δ1.19%, MAE=Δ0.17°C)를 나타내었다. 보정된 해수면온도(단계 2)의 정확도는

기상청 부이 자료(RMSE=Δ0.21°C, rRMSE=Δ0.91%, MAE=Δ0.17°C)와 iQUAM 표층 뜰개 부이 자료(RMSE=Δ0.22°C, rRMSE=Δ0.96%, MAE=Δ0.17°C)를 검증자료로 사용했을 때 유사하게 상승하였다. 모든 현장 관측 자료를 학습하여 보정된 해수면온도는 정확도가 더욱 상승(RMSE=Δ0.31°C, rRMSE=Δ1.74%, MAE=Δ0.25°C) 하였지만, 기존의 해수면온도 합성장 자료보다는 약간 낮은 정확도를 보였다. 이러한 결과의 주요한 이유는 앞선 뜰개 부이 자료를 사용하였을 때와 동일하게 기상청 부이 자료가 전세계 해양관측 데이터베이스(global telecommunications system)에 포함되어 기존 해수면온도 합성장을 생산할 때 사용되었기 때문일 것으로 판단된다.

앞서 검증한 iQUAM과 기상청의 현장 관측 자료들은 기존 해수면온도 합성장 자료 생산에 활용되었지만, 국립해양조사원의 부이 자료는 사용되지 않았다. 그러므로, 본 연구지역에 대해 개발된 모델을 국립해양조사원의 현장 관측 자료에 대해 검증을 수행하였다(Table 6). 복원된 해수면온도자료의 정확도(RMSE=1.49°C, rRMSE=8.33%, MAE=1.16°C)는 앞서 검증한 iQUAM과 기상청 자료와의 정확도보다 낮다. 기상청(관측 수심: 0.2-0.4 m)과 iQUAM(관측 수심: 0.2 m) 부이에 비해 국립해양조사원(관측 수심: 3 m) 부이는 상대적으로 깊은 수심에서 온도를 측정한다. 이는, 아피층의 해수면온도를 복원한 결과의 정확도에 영향을 주었을 것으로 보인다. 또한, 보정모델의 정답으로 기상청과 iQUAM 부이 자료만을 사용했기 때문에 국립해양조사원 부이 자료의 특성이 학습되지 않은 점이 정확도가 낮게 나온 요인으로 사료된다. 기상청과 iQUAM의 모든 자료로 학습

한 보정된 해수면온도의 현장 자료와의 정확도는 국립해양조사원의 경우(Table 6)가 앞서 비교한 iQUAM과 기상청과의 결과(Table 4, 5)보다 향상(RMSE=Δ0.29°C, rRMSE=-1.65%, MAE=Δ0.24°C)하였다.

3) 공간분포 비교

Fig. 5와 6은 기존 해수면온도 합성장과 본 연구에서 개발된 해수면온도 자료들의 공간 분포 예시를 보여주고 있다. 복원된 해수면온도(Fig. 5(c))는 초고해상도 합성장인 MUR SST (Fig. 5(b)), 원본 해수면온도(Fig. 4(a))와 비슷한 공간분포를 보이지만, MUR SST와 원본에 비해 전반적으로 해수면온도를 과소추정하였다. 반면, 보정된 해수면온도(Fig. 5(d))는 동해의 와동(eddy)과 전선 등의 해양현상을 복원된 해수면온도보다 뚜렷하고 상세하게 모의했으며, MUR SST와 유사한 온도분포를 보여주었다. 반면, 최적보간 기반의 OSTIA (Fig. 5(a)) 자료

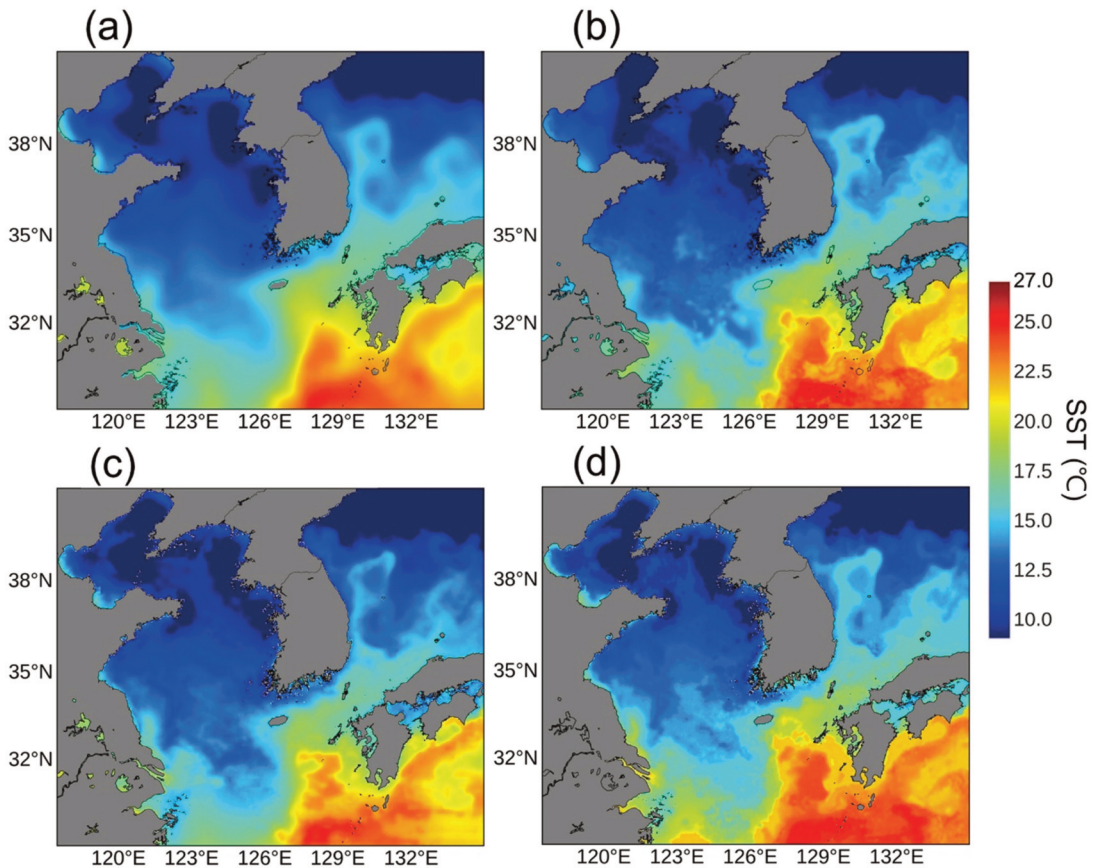


Fig. 5. Spatial distribution of SSTs on 5 May 2013. (a) OSTIA, (b) MUR SST, (c) reconstructed SST, and (d) corrected SST. The spatial resolutions of the SSTs are 0.05°, 1, 4, and 4 km, respectively.

에는 메소스케일의 해양현상들이 평활화(smoothing)로 인해 뚜렷하게 모의되지 않았다(Donlon *et al.*, 2012; Woo and Park, 2020). 본 연구의 보정과정은 전반적으로 복원된 해수면온도를 상승시키는 쪽으로 모의하였다. 이는 야간 해수면 냉각으로 인해 위성의 열적외 센서가 측정하는 해수면온도가 현장 관측센서로 측정하는 해수면온도보다 낮기 때문인 것으로 유추할 수 있다(Donlon *et al.*, 2002; Park *et al.*, 2008). 본 연구에서 생산된 최종 해수면온도 합성장(Fig. 5(d))은 복원된 해수면온도(Fig. 5(c))에서 기존 합성장 산출물에 비해 저추정된 동해 와동과 일본 남쪽의 쿠루시오 해류 부분을 더욱 명확하게 모의하였다. 하지만, Fig. 6에서 두 단계를 거쳐 보정된 해수면온도는 복원된 해수면온도를 과대추정하여 전반적으로 높은 온도 분포를 보이며, 특히 제주도 남쪽해역에 대해 복원된 해수면온도 보다 1°C 이상 상승하는 경향을 보인다. 이는, 학습에 사용된 현장 자료들이 38°N

이상의 연구지역에 부재하기 때문에 보정 결과에 편차(bias)를 야기했을 것으로 판단된다(Fig. 1). 또한, 보정에 사용된 입력변수가 적어 보정에 사용된 LGBM 모델의 성능을 높이는 데 한계가 있었을 것으로 보인다(Ke *et al.*, 2017). 선행연구에서는 위성 마이크로파 센서기반 해수면온도 산출물을 입력 변수로 추가하여 보정 모델의 성능과 안정성을 향상시켰다(Jung *et al.*, 2022). 하지만, 마이크로파 센서기반 해수면온도는 육지로 인한 노이즈를 제거하기 위해 연안으로부터 100 km까지 자료가 산출되지 않아 본 연구에서는 사용할 수 없다. 추후, 위성 해수면온도 산출(retrieval)과 합성장 생산에서 오차 보정과 품질 개선을 위해 사용되는 Optimum Interpolated Sea Surface Temperature (OISST)를 입력자료로 활용한다면 모델의 성능을 향상시킬 수 있을 것으로 보인다.

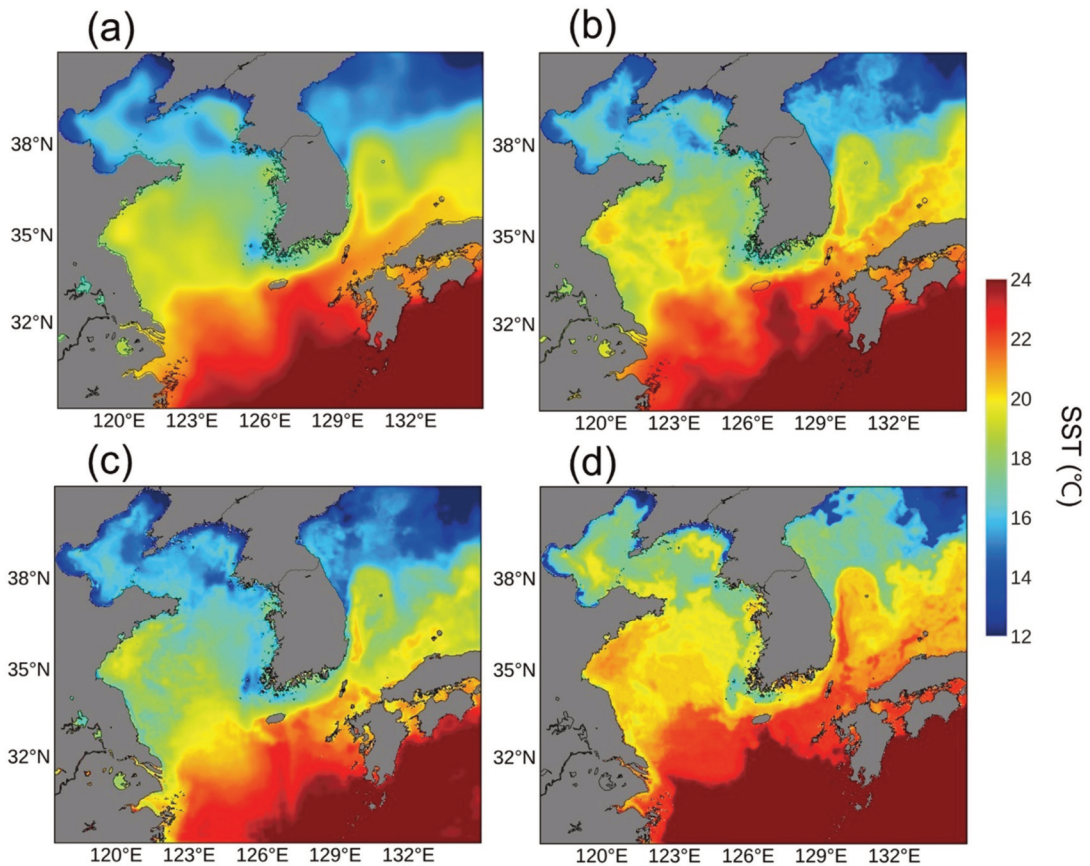


Fig. 6. Spatial distribution of SSTs on 30 Oct. 2015. (a) OSTIA, (b) MUR SST, (c) reconstructed SST, and (d) corrected SST. The spatial resolutions of the SSTs are 0.05°, 1, 4, and 4 km, respectively.

4. 결론

본 연구에서는 DINCAE 모델을 사용하여 다중위성 기반 한반도 연안 해수면온도를 복원하고 이어 LGBM 기계학습을 사용하여 현장 자료와의 일치도를 높이는 보정을 수행하였다. 첫 번째 단계에서 복원된 해수면온도 자료의 정확도는($R^2=0.98$, $RMSE=0.97^\circ C$)였고, 동해의 와동, 전선, 그리고 쿠루시오해류의 사행과 같은 메소스케일의 해양현상을 초고해상도 해수면온도 합성장인 MUR SST와 유사하게 모의하였다. 보정된 해수면온도(Step 2)의 현장자료와의 정확도는 기상청 계류형 부이($RMSE=\Delta 0.21^\circ C$, $rRMSE=\Delta 0.91\%$, $MAE=\Delta 0.17^\circ C$), iQUAM 뜰개 부이($RMSE=RMSE=\Delta 0.22^\circ C$, $rRMSE=\Delta 0.96\%$, $MAE=\Delta 0.17^\circ C$), 그리고 국립해양조사원 계류형 부이($RMSE=\Delta 0.29^\circ C$, $rRMSE=\Delta 1.65\%$, $MAE=\Delta 0.24^\circ C$) 모두 상승하였다. 또한, LGBM을 통한 보정은 과적합을 보인 선행연구의 문제점을 유의미하게 해결했다. 모든 현장 자료들을 이용해 보정된 해수면온도의 표층 뜰개 부이와의 정확도는 다수의 다중위성, 현장 자료 활용, 그리고 오차를 줄이는 다수의 작업을 통해 생산되는 기존의 해수면온도 합성장 자료보다 나은 정확도를 보였다. 이는, 기계학습을 통해 피층을 관측하는 위성과 표층을 관측하는 현장 자료 간의 차이를 줄이는 기여를 할 수 있을 것으로 기대한다. 본 연구의 연구기간(2012-2021년) 동안 기존 합성장 자료 생산에 활용된 해수면온도 자료의 갯수는 OSTIA가 6개, 그리고 MUR SST의 경우 8개이다(Chin *et al.*, 2017; Good *et al.*, 2020). 기존 합성장 자료들은 각 위성 별 고유의 오차와 다중 자료를 합성할 때 발생하는 오차를 보정하는 다수의 단계를 통해 생성된다(Chin *et al.*, 2017; Good *et al.*, 2020). 본 연구의 기계학습을 통해 생산된 해수면온도 합성장은 기존 자료동화 기반의 합성장보다 적은 자료(3개)를 활용하고 복원과 보정 두 단계로 이루어진 비교적 단순한 과정을 통해 생산되므로, 향후 해수면온도 합성장 생산 시 효율성 증대에 기여를 할 것으로 보인다.

사사

본 연구는 2022년도 정부(해양수산부)의 재원으로

해양수산과학기술진흥원의 ‘해양위성영상 분석 활용 기술 개발(KIMST-20210046)’, ‘과학기술기반 해양환경 영향평가 기술개발(KIMST-20210427)’ 사업, 그리고 행정안전부 재난안전 부처협력 기술개발사업(20009742) 지원연구사업의 지원을 받아 수행된 연구임.

References

- Alvera-Azcárate, A., A. Barth, G. Parard, and J.-M. Beckers, 2016. Analysis of SMOS sea surface salinity data using DINEOF, *Remote Sensing of Environment*, 180: 137-145. <https://doi.org/10.1016/j.rse.2016.02.044>
- Alvera-Azcárate, A., A. Barth, J. M. Beckers, and R. H. Weisberg, 2007. Multivariate reconstruction of missing data in sea surface temperature, chlorophyll, and wind satellite fields, *Journal of Geophysical Research: Oceans*, 112(C3). <https://doi.org/10.1029/2006JC003660>
- Barth, A., A. Alvera-Azcárate, M. Licer, and J.-M. Beckers, 2020. DINCAE 1.0: a convolutional neural network with error estimates to reconstruct sea surface temperature satellite observations, *Geoscientific Model Development*, 13(3): 1609-1622. <https://doi.org/10.5194/gmd-13-1609-2020>
- Barth, A., A. Alvera-Azcárate, C. Troupin, and J.-M. Beckers, 2022. DINCAE 2.0: multivariate convolutional neural network with error estimates to reconstruct sea surface temperature satellite and altimetry observations, *Geoscientific Model Development*, 15(5): 2183-2196. <https://doi.org/10.5194/gmd-15-2183-2022>
- Beckers, J.-M. and M. Rixen, 2003. EOF calculations and data filling from incomplete oceanographic datasets, *Journal of Atmospheric and Oceanic Technology*, 20(12): 1839-1856. [https://doi.org/10.1175/1520-0426\(2003\)020<1839:ECADFF>2.0.CO;2](https://doi.org/10.1175/1520-0426(2003)020<1839:ECADFF>2.0.CO;2)
- Bulgin, C.E., C.J. Merchant, and D. Ferreira, 2020.

- Tendencies, variability and persistence of sea surface temperature anomalies, *Scientific Reports*, 10(1): 1-13. <https://doi.org/10.1038/s41598-020-64785-9>
- Chin, T.M., J. Vazquez-Cuervo, and E.M. Armstrong, 2017. A multi-scale high-resolution analysis of global sea surface temperature, *Remote Sensing of Environment*, 200: 154-169. <https://doi.org/10.1016/j.rse.2017.07.029>
- Choi, H., Y. Kang, and J. Im, 2021. Estimation of TRO POMI-derived Ground-level SO₂ Concentrations Using Machine Learning Over East Asia, *Korean Journal of Remote Sensing*, 37(2): 275-290 (in Korean with English abstract). <https://doi.org/10.7780/kjrs.2021.37.2.8>
- Donlon, C., P. Minnett, C. Gentemann, T. Nightingale, I. Barton, B. Ward, and M. Murray, 2002. Toward improved validation of satellite sea surface skin temperature measurements for climate research, *Journal of Climate*, 15(4): 353-369. [https://doi.org/10.1175/1520-0442\(2002\)015<0353:TIVOSS>2.0.CO;2](https://doi.org/10.1175/1520-0442(2002)015<0353:TIVOSS>2.0.CO;2)
- Donlon, C.J., M. Martin, J. Stark, J. Roberts-Jones, E. Fiedler, and W. Wimmer, 2012. The operational sea surface temperature and sea ice analysis (OSTIA) system, *Remote Sensing of Environment*, 116: 140-158. <https://doi.org/10.1016/j.rse.2010.10.017>
- Good, S., E. Fiedler, C. Mao, M. J. Martin, A. Maycock, R. Reid, J. Roberts-Jones, T. Searle, J. Waters, and J. While, 2020. The current configuration of the OSTIA system for operational production of foundation sea surface temperature and ice concentration analyses, *Remote Sensing*, 12(4): 720. <https://doi.org/10.3390/rs12040720>
- Han, Z., Y. He, G. Liu, and W. Perrie, 2020. Application of DINCAE to Reconstruct the Gaps in Chlorophyll-a Satellite Observations in the South China Sea and West Philippine Sea, *Remote Sensing*, 12(3): 480. <https://doi.org/10.3390/rs12030480>
- Jang, E., Y.J. Kim, J. Im, Y.-G. Park, and T. Sung, 2022. Global sea surface salinity via the synergistic use of SMAP satellite and HYCOM data based on machine learning, *Remote Sensing of Environment*, 273: 112980. <https://doi.org/10.1016/j.rse.2022.112980>
- Ji, C., Y. Zhang, Q. Cheng, and J. Y. Tsou, 2021. Investigating ocean surface responses to typhoons using reconstructed satellite data, *International Journal of Applied Earth Observation and Geoinformation*, 103: 102474. <https://doi.org/10.1016/j.jag.2021.102474>
- Jonasson, O., I. Gladkova, A. Ignatov, and Y. Kihai, 2021. Algorithmic improvements and consistency checks of the NOAA global gridded supercollated SSTs from low Earth orbiting satellites (L3S-LEO), *Proc. of SPIE 11752 Ocean Sensing and Monitoring XIII*, May 25, vol. 11752, pp. 5-18. <https://doi.org/10.1117/12.2585819>
- Jung, S., Y.J. Kim, S. Park, and J. Im, 2020. Prediction of sea surface temperature and detection of ocean heat wave in the South Sea of Korea using time-series deep-learning approaches, *Korean Journal of Remote Sensing*, 36(5-3): 1077-1093 (in Korean with English abstract). <https://doi.org/10.7780/kjrs.2020.36.5.3.7>
- Jung, S., C. Yoo, and J. Im, 2022. High-Resolution Seamless Daily Sea Surface Temperature Based on Satellite Data Fusion and Machine Learning over Kuroshio Extension, *Remote Sensing*, 14(3): 575. <https://doi.org/10.3390/rs14030575>
- Kang, Y., M. Kim, E. Kang, D. Cho, and J. Im, 2022. Improved retrievals of aerosol optical depth and fine mode fraction from GOCI geostationary satellite data using machine learning over East Asia, *ISPRS Journal of Photogrammetry and Remote Sensing*, 183: 253-268. <https://doi.org/10.1016/j.isprsjprs.2021.11.016>
- Ke, G., Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T.-Y. Liu, 2017. Lightgbm: A highly efficient gradient boosting decision tree, *Advances*

- in neural information processing systems, *Proc. of 31st Conference on Neural Information Processing Systems*, Long Beach, CA, vol. 30, pp. 3146-3154.
- Kim, T., S. Chung, C.Y. Chung, and S. Baek, 2017. An estimation of the composite sea surface temperature using COMS and polar orbit Satellites data in Northwest Pacific Ocean, *Korean Journal of Remote Sensing*, 33(3): 275-285 (in Korean with English abstract). <https://doi.org/10.7780/kjrs.2017.33.3.3>
- Kurihara, Y., H. Murakami, and M. Kachi, 2016. Sea surface temperature from the new Japanese geostationary meteorological Himawari-8 satellite, *Geophysical Research Letters*, 43(3): 1234-1240. <https://doi.org/10.1002/2015GL067159>
- Levy, R.C., S. Mattoo, V. Sawyer, Y. Shi, P.R. Colarco, A.I. Lyapustin, Y. Wang, and L.A. Remer, 2018. Exploring systematic offsets between aerosol products from the two MODIS sensors, *Atmospheric Measurement Techniques*, 11(7): 4073-4092. <https://doi.org/10.5194/amt-11-4073-2018>
- Liu, X. and M. Wang, 2018. Gap filling of missing data for VIIRS global ocean color products using the DINEOF method, *IEEE Transactions on Geoscience and Remote Sensing*, 56(8): 4464-4476. <https://doi.org/10.1109/TGRS.2018.2820423>
- Luo, X., J. Song, J. Guo, Y. Fu, L. Wang, and Y. Cai, 2022. Reconstruction of chlorophyll-a satellite data in Bohai and Yellow sea based on DINCAE method, *International Journal of Remote Sensing*, 43(9): 3336-3358. <https://doi.org/10.1080/01431161.2022.2090872>
- Merchant, C., A. Harris, M. Murray, and A. Zavody, 1999. Toward the elimination of bias in satellite retrievals of sea surface temperature: 1. Theory, modeling and interalgorithm comparison, *Journal of Geophysical Research: Oceans*, 104(C10): 23565-23578. <https://doi.org/10.1029/1999JC900105>
- Nechad, B., A. Alvera-Azcarate, K. Ruddick, and N. Greenwood, 2011. Reconstruction of MODIS total suspended matter time series maps by DINEOF and validation with autonomous platform data, *Ocean Dynamics*, 61(8): 1205-1214. <https://doi.org/10.1007/s10236-011-0425-4>
- NOAA/STAR, 2021. *GHRSSST NOAA/STAR ACSPO v2.80 0.02 degree L3S Dataset from Afternoon LEO Satellites (GDS v2)*, Physical Oceanography Distributed Active Archive Center (PODAAC), CA, USA. <https://doi.org/10.5067/GHLPM-3SS28>
- O'Carroll, A.G., E.M. Armstrong, H.M. Beggs, M. Bouali, K.S. Casey, G.K. Corlett, P. Dash, C.J. Donlon, C.L. Gentemann, and J.L. Hoyer, 2019. Observational needs of sea surface temperature, *Frontiers in Marine Science*, 6: 420. <https://doi.org/10.3389/fmars.2019.00420>
- O'Carroll, A.G., J.R. Eyre, and R.W. Saunders, 2008. Three-way error analysis between AATSR, AMSR-E, and in situ sea surface temperature observations, *Journal of Atmospheric and Oceanic Technology*, 25(7): 1197-1207. <https://doi.org/10.1175/2007JTECHO542.1>
- Ouala, S., R. Fablet, C. Herzet, B. Chapron, A. Pascual, F. Collard, and L. Gaultier, 2018. Neural network based Kalman filters for the Spatio-temporal interpolation of satellite-derived sea surface temperature, *Remote Sensing*, 10(12): 1864. <https://doi.org/10.3390/rs10121864>
- Park, J., D.-W. Kim, Y.-H. Jo, and D. Kim, 2018. Accuracy evaluation of daily-gridded ASCAT satellite data around the Korean Peninsula, *Korean Journal of Remote Sensing*, 34(2-1): 213-225 (in Korean with English abstract). <https://doi.org/10.7780/kjrs.2018.34.2.1.5>
- Park, K. and Y.-H. Kim, 2009. A Methodology for 3-D Optimally-Interpolated Satellite Sea Surface Temperature Field and Limitation, *Journal of the Korean Earth Science Society*, 30(2): 223-233 (in Korean with English abstract). <https://doi.org/10.7780/kjrs.2009.30.2.223>

- org/10.5467/JKES.2009.30.2.223
- Park, K., 2019. *GK-2A AMI Algorithm Theoretical Basis Document (ATBD) Sea Surface Temperature*, National Meteorological Satellite Center (NMSC), Jincheon, Korea.
- Park, K., F. Sakaida, and H. Kawamura, 2008. Oceanic skin-bulk temperature difference through the comparison of satellite-observed sea surface temperature and in-situ measurements, *Korean Journal of Remote Sensing*, 24(4): 273-287 (in Korean with English abstract). <https://doi.org/10.7780/kjrs.2008.24.4.273>
- Park, S., M. Kim, and J. Im, 2021. Estimation of Ground-level PM 10 and PM 2.5 Concentrations Using Boosting-based Machine Learning from Satellite and Numerical Weather Prediction Data, *Korean Journal of Remote Sensing*, 37(2): 321-335 (in Korean with English abstract). <https://doi.org/10.7780/kjrs.2021.37.2.11>
- Stark, J.D., C.J. Donlon, M.J. Martin, and M.E. McCulloch, 2007. OSTIA: An operational, high resolution, real time, global sea surface temperature analysis system, *Proc. of Oceans 2007-Europe*, Aberdeen, UK, Jun. 18-21, pp. 1-4. <https://doi.org/10.1109/OCEANSE.2007.4302251>
- Sunder, S., R. Ramsankaran, and B. Ramakrishnan, 2020. Machine learning techniques for regional scale estimation of high-resolution cloud-free daily sea surface temperatures from MODIS data, *ISPRS Journal of Photogrammetry and Remote Sensing*, 166: 228-240. <https://doi.org/10.1016/j.isprsjprs.2020.06.008>
- Wentz, F.J., C. Gentemann, D. Smith, and D. Chelton, 2000. Satellite measurements of sea surface temperature through clouds, *Science*, 288(5467): 847-850. <https://doi.org/10.1126/science.288.5467.84>
- Woo, H.J. and K. Park, 2020. Inter-comparisons of daily sea surface temperatures and in-situ temperatures in the coastal regions, *Remote Sensing*, 12(10): 1592. <https://doi.org/10.3390/rs12101592>
- Xiao, C., N. Chen, C. Hu, K. Wang, Z. Xu, Y. Cai, L. Xu, Z. Chen, and J. Gong, 2019. A spatiotemporal deep learning model for sea surface temperature field prediction using time-series satellite data, *Environmental Modelling & Software*, 120: 104502. <https://doi.org/10.1016/j.envsoft.2019.104502>
- Xu, F. and A. Ignatov, 2014. In situ SST quality monitor (iQuam), *Journal of Atmospheric and Oceanic Technology*, 31(1): 164-180. <https://doi.org/10.1175/JTECH-D-13-00121.1>
- Xu, F. and A. Ignatov, 2016. Error characterization in iQuam SSTs using triple collocations with satellite measurements, *Geophysical Research Letters*, 43(20): 10826-10834. <https://doi.org/10.1002/2016GL070287>