Applied Ocean Research 170 (2026) 105038

Contents lists available at ScienceDirect

Applied Ocean Research

journal homepage: www.elsevier.com/locate/apor

Deep learning for sea surface temperature reconstruction under cloud occlusion

Andrea Asperti [a], Ali Aydogdu [b,*], Angelo Greco [a], Fabio Merizzi [a], Pietro Miraglio [b], Beniamino Tartufoli [b], Alessandro Testa [a], Nadia Pinardi [b,c], Paolo Oddo [b,c]

[a] University of Bologna, Department of Informatics - Science and Engineering (DISI), Via Mura Anteo Zamboni 7, Bologna, 40126, Italy
[b] CMCC Foundation - Euro-Mediterranean Center on Climate Change, Italy
[c] University of Bologna, Department of Physics and Astronomy (DIFA), Viale Carlo Berti Pichat 6/2, Bologna, 40127, Italy

ARTICLE INFO

Keywords:
Sea Surface Temperature reconstruction
Cloud filling techniques
Neural networks
Deep learning

ABSTRACT

Sea Surface Temperature (SST) reconstructions from satellite images affected by cloud gaps have been extensively documented in the past three decades. Here we describe several deep learning models to fill the cloud-occluded areas starting from MODIS Aqua nighttime L3 images in the Italian Seas. To tackle this challenge, after testing different models and methodologies, we employed a type of Convolutional Neural Network model (U-Net) to reconstruct cloud-covered portions of satellite imagery while preserving the integrity of observed values in cloud-free areas. We demonstrate the high precision of U-Net with respect to available products done using OI interpolation algorithms. Our results are promising with respect to some earlier studies while suggesting further investigation for more robust intercomparison.

1. Introduction

Sea Surface Temperature (SST) is one of the essential climate variables for understanding and modeling the Earth's climate system. Air-sea heat fluxes, depending on the sea surface and air temperature difference, are a major driver of the atmosphere-ocean coupled dynamics. Satellite SST products, interpolated on regular grids and cloud gaps filled (Reynolds et al., 2002; Konik et al., 2019), continue to be used today to force atmospheric model analyses and reanalyses (Donlon et al., 2012). The influence of high-resolution sea surface temperatures (SST) on the accuracy of atmospheric reanalysis has been demonstrated (Parfitt et al., 2017), along with the role of SST fronts in driving climate variability (Larson et al., 2024). Therefore, we aim to investigate various methods to fill cloud-occluded areas without introducing smoothing effects on the cloud free pixels.

The SST satellite measurement is typically made by sensing the ocean radiation in many wavelengths within the near infrared and microwave parts of the electromagnetic spectrum. One difficulty in getting a global product using only infrared imagery is that there are large regions with clouds (Wylie et al., 2005), and reconstructing SST under cloud occlusion conditions is a challenging and lively research topic. Traditional statistical techniques, such as Optimal Interpolation (OI; Bretherton et al., 1976), Empirical Orthogonal Function (EOF; Alvera-Azcárate et al., 2011), and other similar techniques (Jung et al., 2022), fill missing values based on spatial/temporal correlations between observed SST points. The present-day SST Level 4 (L4) product from Copernicus Marine Service is produced by a spacetime OI scheme (Nardelli et al., 2013).

A problem of statistical objective mapping techniques is that they often struggle to resolve fine-scale features, resulting in smooth reconstructions (Chin et al., 2017; Fablet et al., 2017; Barth et al., 2020). These techniques typically assume linearity, which limits their ability to capture the complex dynamics of SST fields. Particularly in the EOF-based methods (Alvera-Azcárate et al., 2005), truncation of smaller modes results in the over-weighting of large scale signals which also leads to smoothing of the reconstructed field. One of the most advanced SST multi-sensor reconstructions Chin et al. (2017) currently employs adaptable time windows to fill image gaps caused by cloud cover. However, this approach has several issues, including the potential for mesoscale signals beneath the clouds to be biased by older data from several days earlier, leading to contamination of the reconstructed area.

Consequently, there has been growing interest in recent years in applying deep learning techniques to address this challenge relying on the rapidly increasing amount of data and computational resources. A recent survey of both statistical and AI-based reconstruction methods for oceanographic data is available in Ćatipović et al. (2023). Specifically, convolutional autoencoders, such as Data INterpolating Convolutional Auto-Encoder (DINCAE), have been successfully used

* Corresponding author.
E-mail address: ali.aydogdu@cmcc.it (A. Aydogdu).

https://doi.org/10.1016/j.apor.2026.105038
Received 30 June 2025; Received in revised form 15 March 2026; Accepted 19 March 2026

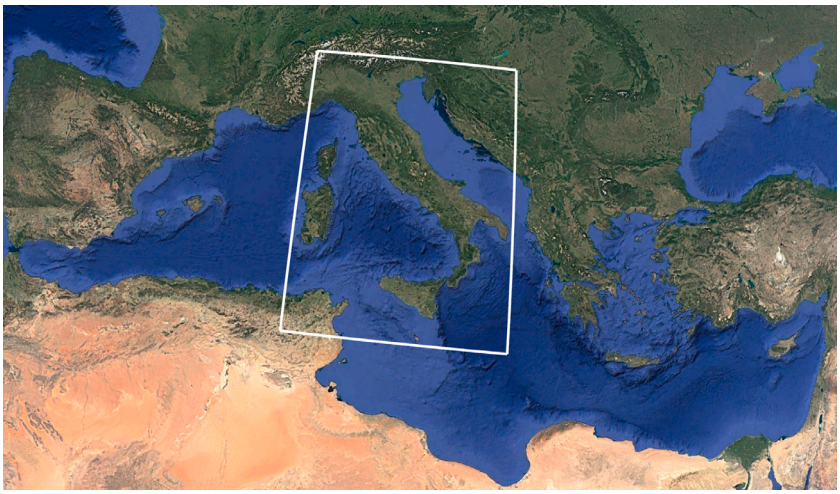

Fig. 1. The white polygon describes the region of our investigation, with latitude between 35.33° and 46.0°N and longitude between 7.92° and 18.58°E.

to reconstruct SST and chlorophyll concentrations (Barth et al., 2020, 2022; Han et al., 2020), and vision transformers trained with a masked autoencoder approach have also been explored (Masked Autoencoders for Sea Surface Temperature Reconstruction under Occlusion; Goh et al., 2024, MAESSTRO). Muc et al. (2025) blends both approaches in CRITER 1.0 and reconstructs low-frequency SST components utilizing a Vision Transformer-based model, then to recovering high-frequency details, a U-Net type of network iteratively refines the estimate.

Wang et al. (2025a) proposed GDCM, Generalized data completion model for satellite observations, in which they utilized CNNs and attention mechanism, to leverage spatio-temporal information within remote sensing data to fill missing regions. They use AMSR2/SSMI for global coverage and MODIS dataset for the local one. Wang et al. (2025b) instead developed seamless global daily SST (SDG-SST) in which they used Visible Infrared Imaging Radiometer Suite (VIIRS) in Deep spatio-temporal fusion model based on CNNs in U-Net architecture. DINCoDE (Yang et al., 2025) adopts a hybrid dual-encoder architecture in the marginal seas of the Northwest Pacific Ocean that integrates a convolutional neural network (CNN) and a Transformer encoder, enabling the joint extraction of local and global spatiotemporal features. A composite loss function is employed to ensure both pixel-level accuracy and structural fidelity in the reconstructed data. Recently, Fourier Neural Operators have also been used to reconstruct sea surface temperature. MSF_FNO (Chen et al., 2025) integrates multi-scale feature fusion and frequency-domain neural operator technology to effectively overcome the limitations of single-scale feature processing and image-domain reconstruction in existing methods. Their approach extracts critical features across multiple scales, ensuring global consistency and detailed features in reconstruction results besides capturing SST frequency-domain information and extracting structured features of SST images. FDNIN, Fourier Diffusion Neural Inpainting Network, integrates Fourier neural operators into a generative diffusion model to leverage the advantages of the frequency domain (Zuo et al., 2025). FDNIN employs a strip-stream structure to utilize global consistency and local details from historical data, improving SST reconstruction. CARE-SST: context-aware reconstruction diffusion model for sea surface temperature using visible infrared imaging radiometer suite (VIIRS) provides a superior feature resolution compared to the other SST products (Choo et al., 2025).

More recently, new architectures such as Swin-Transformers and denoising diffusion models are also used efficiently to reconstruct ocean fields. Wang et al. (2026) employed a dual-residual SwinTransformer (SwinOcean3D) to reconstruct the 3D fields of the ocean in the upper 1000 m. Ye et al. (2024) combined optimal interpolation and SwinU-Net (OI-SwinUnet) architectures to reconstruct daily chlorophyll-a data. OI-SwinUnet method was used to reconstruct the MODIS chlorophyll-a concentration products of the South China Sea. Barth et al. (2024) employed a denoising diffusion model to reconstruct chlorophyll a concentration from the Ocean and Land Color Instrument (OLCI) sensor (aboard the satellites Sentinel-3 A and Sentinel-3B) on a small area of the Black Sea and compared with DINCAE.

In a more general context, Pauthenet et al. (2022) introduced OSnet (Ocean Stratification network), is a bootstrapped multilayer perceptron trained to predict simultaneously temperature and salinity (T-S) profiles down to 1000 m and the mixed-layer depth (MLD) from surface data covering 1993 to 2019. Sammartino et al. (2025) developed 4DMED-bionet to reconstruct the 3D fields of the ocean using satellite and in-situ observations in a model built on CNN and LSTM layers. We note that in this work, SST L4 products are used therefore, cloud occultation is not addressed. ReconMOST (Song et al., 2020) proposed a diffusion model trained on CMIP6 and EN4 datasets for multi-layer sea temperature reconstruction extending SST reconstruction to a global, multi-layer setting, handling over 92.5% missing data while maintaining reconstruction accuracy, spatial resolution, and superior generalization capability. CLuster Optimal Interpolation Neural Network (CLOINet; Cutolo et al., 2024) combines sparse in-situ observations into a full 3D field leveraging shape information from ocean remote sensing images. Beyond these applications using earth observations, image completion -or image inpainting- is a well-researched area in image processing (Iizuka et al., 2017; Liu et al., 2018; Peng et al., 2021; Wan et al., 2021; Zheng et al., 2022; Jain et al., 2023), with various techniques successfully applied across different domains.

In this study, we evaluate some of these image completion techniques for reconstructing SST in areas occluded by clouds. The method is applied to SST Level 3 (L3) images at the resolution of 4 km from the MODIS Aqua satellite and infrared sensor; particular attention is given to the model configuration and the calibration of the parameters. After this, we apply the best algorithm to another L3 product from Copernicus Marine Service.

The analysis is done on a region comprising the Italian Seas, described in the white box of Fig. 1. The area covers 256 × 256 observation points with a latitude between 35.33° and 46.0°E and a longitude between 7.92° and 18.58°N, for a total domain extension of 1020 x 1020 square kilometers. The reconstruction obtained by our model was validated against the current L4 reconstructions from Copernicus Marine Service (Nardelli et al., 2013; Pisano et al., 2022) and the recent DINCAE model (Barth et al., 2020, 2022), testifying in both cases an improvement.

The article is structured in the following way. In Section 2, we describe the satellite dataset used for the initial training of the neural

network. This section also contains an investigation of data, comprising gradients (Section 2.2), persistence (Section 2.3), and climatology (Section 2.4). The methodology is explained in Section 3, where we introduce the main classes of models investigated in this article, namely U-Net models (Section 3.1) and Visual Transformers (Section 3.2). In this section we also discuss the artificial cloud generator (Section 3.3), and training (Section 3.4). The model intercomparison and selection is illustrated in Section 4. Section 5 applies the trained network to an operational L3S dataset for comparison with established products. We discuss the results and draw the conclusions in Section 6.

2. Satellite data set for algorithm development

The SST training dataset is taken from the products derived from Moderate-resolution Imaging Spectroradiometer (MODIS; Werdell et al., 2013) (available at https://podaac.jpl.nasa.gov/dataset/MODIS_AQUA_L3_SST_THERMAL_DAILY_4K_NIGHTTIME_V2014.0, last accessed Sept 2024). Data are acquired by the NASA TERRA and AQUA satellite platforms, launched in 1999 and 2002 respectively. For our investigation, we used the daily products at 4 km spatial resolution relative to nighttime passes. We reconstructed the MODIS-AQUA data, using MODIS-TERRA for validation purposes. The MODIS-AQUA data set contains SST values with missing data due to cloud occlusions. All nighttime measurements from 7/4/2002 to 12/31/2023 were used. Part of the data, from 7/4/2021 to 12/31/2023, was used for testing purposes. The data set also contains quality flags for each grid point. The flags go from 0 (best) to 5 (worst). We used only values of quality 0, 1 and 2, amounting respectively to 79%, 21% and 0.2% of grid points for our region of interest. Our study relies on MODIS Acqua nighttime L3 data since we want to learn foundation temperature which is closer to night time SST. This choice is done to be able to test our model with daily products in Copernicus Services which usually release foundation temperature (see Yang et al., 2021). This choice follows also the upgrade in DINCAE2.0 (Barth et al., 2022) which uses MODIS based high-resolution products after the experience using lower resolution AVHRR products in DINCAE1.0 (Barth et al., 2020). Moreover, the product that we target on Copernicus Marine Service provides daily (night time), merged multi-sensor (L3S), and optimally interpolated (L4) satellite-based estimates of the foundation SST (Nardelli et al., 2013). Therefore, our investigation is built on night time products available. We note that these products completely miss the diurnal cycle of the SST therefore cannot be used to study the variation of the SST during the day.

2.1. Data set analysis

In this section, we investigate the datasets, pointing out a few critical aspects of collected data, typically due to problems of the signal in proximity of the coast, or at the border of clouds. The minimum, maximum and average values for all nighttime SST grid values from MODIS-AQUA are 0.1, 31.1 and 20.5 °C respectively. The SST of the Italian Seas never falls below 5 °C, thus the minimum temperature is likely due to the cloud borders incorrectly associated with seawater by the cloud detection algorithm. These outliers are very few in number (0.0002% of the total data) and have negligible impact on training or evaluation. As expected for the Italian Seas, there is a large amplitude seasonal cycle (Fig. 2) and a suitable seasonal climatology should be calculated, as described in Section 2.4.

2.2. SST gradients

Here, we define gradients as differences in the SST grid points, both in time and space. Gradients are defined as the difference between two consecutive nighttime values in the same grid point and the difference between two consecutive grid points every night (4 km grid spacing). Table 1 presents statistics related to these gradients.

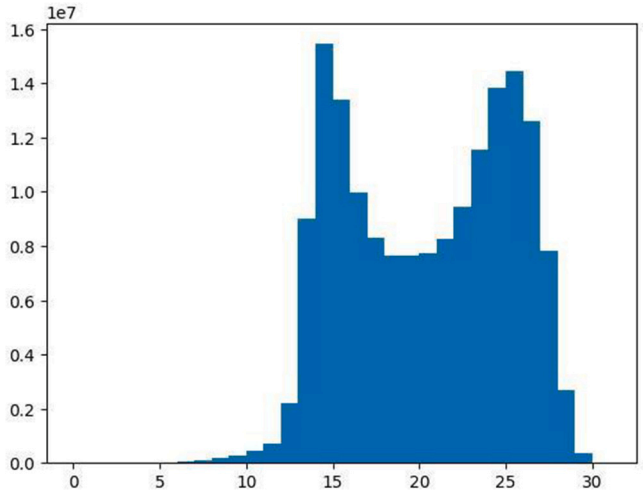

Fig. 2. Histogram relative to the distribution of MODIS-AQUA nightly temperatures, cumulative overall spatial positions and all years.

Table 1
Statistics on the spatial and temporal sea surface temperature gradients (in °C/4km and °C/day respectively), measured as differences between consecutive nights and neighboring points in space within the same night calculated on all data from 2002 to 2023.

	temporal axis				spatial axis			
	max	avg max	mean	std	max	avg max	mean	std
Night	8.2	4.	0.4	0.4	8.1	3.5	0.2	0.2

Table 2
Frequency of high spatial and temporal gradients in MODIS-AQUA data in the region of interest and for the time period 2002–2023. In the spatial case, the "gradient" refers to the difference in SST between two neighboring grid points, 4 km apart; similarly, for the temporal case, it is the difference in SST between two consecutive days at a given spatial position.

	> 1 °C	> 1.5 °C	> 2 °C	> 2.5 °C	> 3 °C
night:time	1.1%	1%	0.9%	0.8%	0.7%
night:spatial	0.3%	0.1%	–	–	–

Spatial gradients are generally low, with an average fluctuation around 0.1/0.2 °C. However, these fluctuations are not uniformly distributed: larger gradients are typically observed near coastal areas (see Fig. 3), particularly along the western coast of the Adriatic Sea and the southern Sicilian coast.

In some days, large gradients are found in the open ocean around large scale oceanic features. An example is given in Fig. 3, where the maximum gradients are around the southern border of the Northern Tyrrhenian cyclonic gyre (Pinardi et al., 2015), east of the Strait of Bonifacio. In other cases, the extreme gradient values are around the cloud borders, due to the accuracy of satellite cloud removal. The mean temporal variation from day to day is approximately 0.4 °C (Table 1).

Table 2 shows the percentage of points with extreme gradients, considering thresholds between 1 and 3 °C. There are almost no nighttime gradients greater than 1 °C.

2.3. Filling cloudy pixels with temporal interpolation

As in many classical reconstructions of SST below the clouds (Chin et al., 2017), it is common to use for each target night a temporal sequence of a few consecutive days to temporally extrapolate/interpolate the SST values in the cloudy pixel. This approach can be seen as a form of persistence filling algorithm, where we use the closest available data to fill the gaps. Unfortunately, the approximation provided by the data

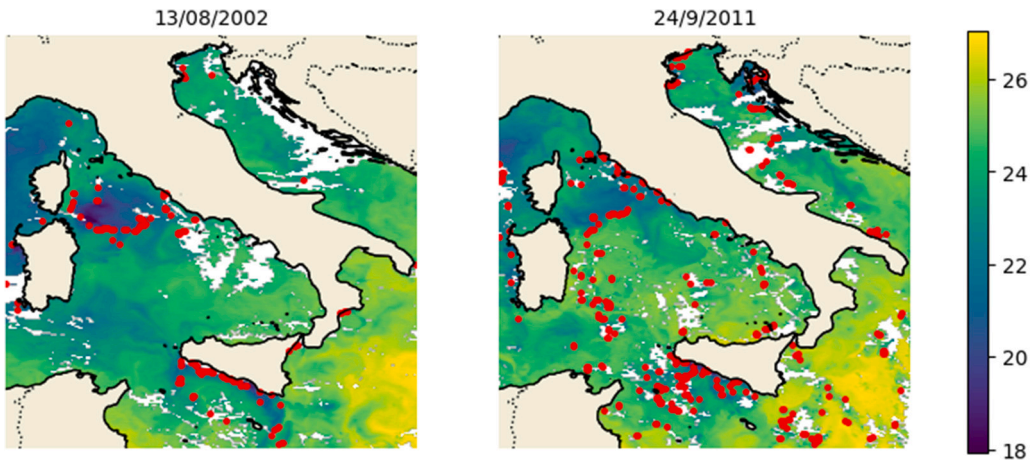

Fig. 3. Large Spatial Gradients (> 1.5 °C locations shown in red dots) relative to different days of the year. Units are °C. Maps cover the area between 35.33° - 46.0°N and 7.92° and 18.58°E.

Table 3
Average distance, in terms of Mean Absolute Difference (MAD) and Rooted Mean Squared Difference (RMSD).

	1d	2d	3d	4d	5d
MAD	0.4°	0.5°	0.6°	0.7°	0.8°
RMSD	0.5°	0.7°	0.8°	1°	1.1°

from the previous days is very inaccurate. Table 3 shows the difference for nighttime data, expressed in terms of both Mean Absolute Difference (MAD) and Root Mean Square Difference (RMSD), for increasing temporal gaps, ranging from 1 to 5 days. The latter is the temporal window used with the current SST reconstruction methods (Ćatipović et al., 2023; Chin et al., 2017). The difference is measured as an average over sea locations that are uncontaminated by clouds on both days. The RMSD and MAD increase very rapidly, and the quality of reconstruction using data from previous days is doubtful.

According to our experiments, reported in Section 4, considering temporal sequences longer than 4 consecutive days does not improve the reconstruction.

2.4. Seasonal climatology

As is clear from Fig. 2, climatology has a large seasonal cycle. In statistical analysis, it is important to subtract the quasi-periodic signals in the time series, such as the seasonal cycle. Thus, we compute the seasonal climatology as the time mean SST across all the 21 years dataset. The anomalies are then calculated by subtracting from each day the seasonal climatology.

The computed daily climatology still suffers from small time gaps where SST values are absent for a particular day of the year, due to the presence of clouds. To fill the climatology gaps, we use interpolation, specifically a Gaussian blur, which uses nearby spatial values to estimate missing ones. The Gaussian function gives more weight to closer values, while gradually reducing the influence of distant points. The algorithm is briefly described in Appendix. We also considered an "unbiased" version of the climatology, where we adjust the mean SST of the baseline towards the observed daily temperature from non-cloudy pixels.

Fig. 4 shows an example of daily climatology and its unbiased version. As is evident from the figure, both the daily climatology and its unbiased version have limitations because the climatology can differ significantly from the specific day under consideration. The unbiased version is usually closer to reality but still not capable of capturing the single day values. For our experiments, we trained the model to learn the residual information with respect to climatology.

3. Models

We tested several different neural network architectures, including variants of U-Net (Ronneberger et al., 2015), Visual Transformers (ViT; Dosovitskiy et al., 2020), and Diffusion Models (Song et al., 2020; Ho et al., 2020). Each model was evaluated across a wide range of configurations, varying input dimensions, the number of channels, network depth, and incorporating specific modules such as attention layers or inception modules. We tested numerous additional variants of U-Net and ViT, tuning their architectures and incorporating more sophisticated layers. Specifically, we experimented with inception modules (Szegedy et al., 2015), Bottleneck Attention Modules (BAM; Park et al., 2018), Convolutional Block Attention Modules (CBAM; Woo et al., 2018), and AttentionAware layers (Zheng et al., 2022). However, none of these mechanisms led to notable performance improvements. So far, we have not achieved satisfactory results with diffusion models, so we will not report on those results.

Special attention was paid to determining the optimal size of the geographical area under investigation. Experimentally, we found that splitting the original 256×256 region into four smaller areas, each 128×128 in size, and training four separate models resulted in better performance. Another key focus of the experimentation was determining the appropriate length of the temporal sequence of consecutive days to be used as input to the model.

In this section, we briefly introduce the two main classes of models: U-Net and ViT.

3.1. U-Net

The U-Net is a type of convolutional neural network (CNN) originally designed for biomedical image segmentation. Its architecture is structured as a U-shaped network, consisting of two main parts: the contracting path (encoder) and the expansive path (decoder). The encoder progressively reduces the spatial dimensions of the input image through convolutional and pooling layers, capturing increasingly abstract and high-level features. The decoder, in contrast, upsamples the feature maps to the original input size, allowing for precise localization in the reconstructed output. All convolutional layers use kernels of dimension 3×3. We tested several different activation functions, without noticing notable differences; the current version relies on the swish activation function (Ramachandran et al., 2018).

The number of downsampling layers and their respective number of channels are key hyperparameters of the network. For example, a U-Net with the structure [64, 128, 256, 512] refers to a model with three downsampling layers that progressively halve the spatial dimensions

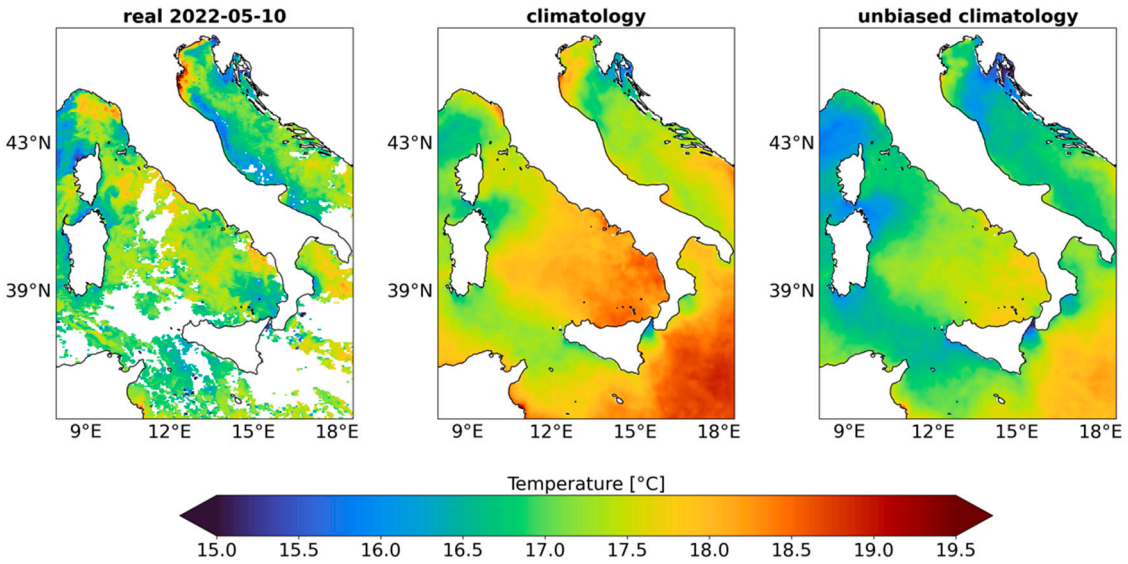

Fig. 4. On the left, the nighttime SST data from a sample day, in this case 10/05/2022; in the middle, the climatology for May 10; on the right, the climatology adjusted (shifted) to the mean of the specific day.

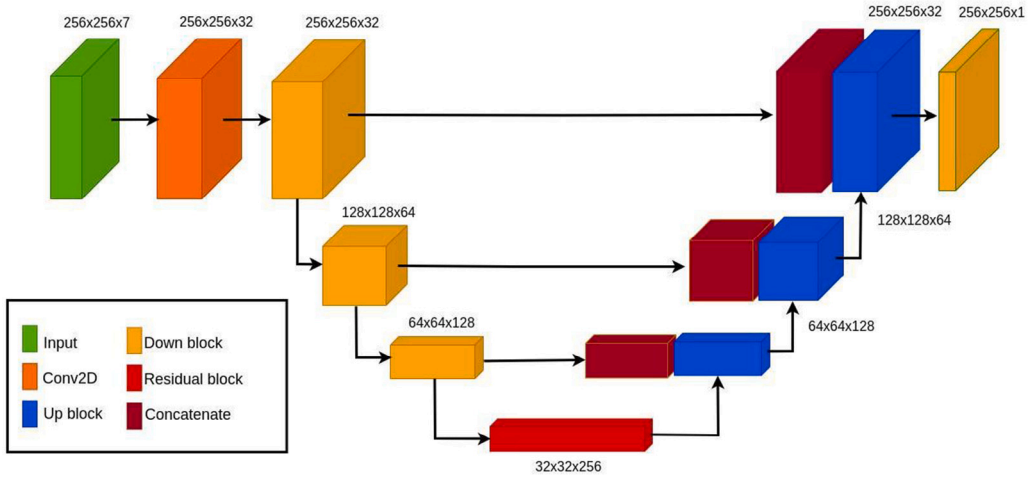

Fig. 5. Basic U-Net. In our terminology, this U-Net has a [32,64,128,256] structure, meaning that it is composed of three downsampling blocks progressively halving the spatial dimension, and increasing the channel dimension to 64, 128 and 256. The initial spatial dimension is 256 × 256. The initial number of channels is 7, corresponding to three input days with the associated masks and the land-sea mask.

while increasing the depth from the initial 64 channels to 128, 256, and 512 channels, respectively.

A key innovation of U-Net is the use of skip connections between corresponding layers in the encoder and decoder. These connections transfer high-resolution feature maps from the encoder to the decoder, allowing the model to combine both coarse and fine-grained information during reconstruction. This design makes U-Net highly effective for tasks where detailed output is essential. The U-Net was also already used by Barth et al. (2020, 2022) in its cloud filling algorithm using two U-Nets in sequence.

The detailed architecture of our models is described in Fig. 5. Downsampling and upsampling blocks are composed of a short, configurable sequence of Residual Blocks, as described in Fig. 6.

3.2. Visual transformer

The Visual Transformer (ViT; Dosovitskiy et al., 2020) is a deep learning architecture designed for image recognition tasks, leveraging the transformer model (Vaswani et al., 2017), which was originally developed for natural language processing. Unlike traditional convolutional neural networks (CNNs) that rely on convolutions to capture spatial information, ViTs use self-attention mechanisms (Bahdanau et al., 2015) to model the relationships between different parts of an image. Our ViT architecture is described in Fig. 7.

The input image is divided into fixed-size patches, which are then flattened and projected into embeddings, similar to how words are handled in transformers for language tasks. In our case, we use patches of dimension 8×8, so an input of dimension $256 \times 256 \times 11$ (relative to 5 consecutive days) is spatially divided into 1024 patches, and reshaped into a vector of dimension (1024×704). The 704 features are projected into an embedding space with dimension 128. These patch embeddings are then processed by the transformer layers, which apply multi-head self-attention to capture global dependencies across the entire image (we use 8-heads). Transformer layers typically do not alter the input–output dimension, allowing to be used in cascade. Specifically, we use a stack of 6 transformer layers. As it is customary for transformer layers, we used the gelu-activation function (Hendrycks and Gimpel, 2016).

At the end of the transformation layers, the sequence of patches is restructured into their original bidimensional spatial structure, and

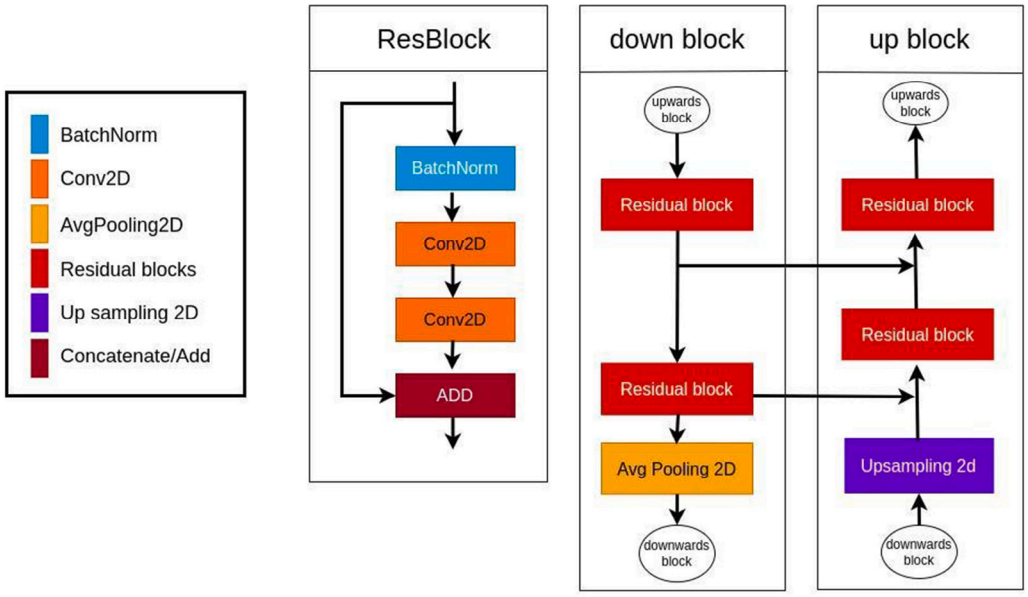

Fig. 6. Upsampling and Downsampling blocks consist of Residual Blocks, exploiting residual connections. All convolutional layers use kernels of dimension 3×3.

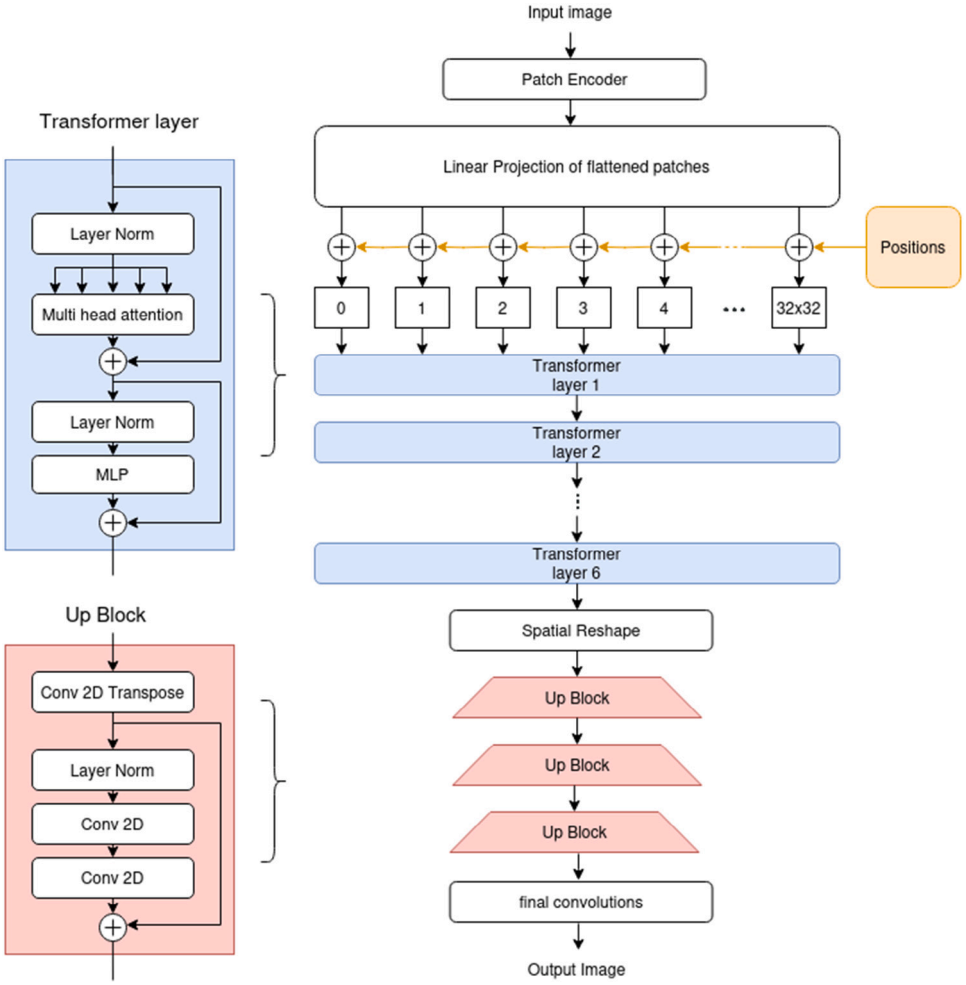

Fig. 7. ViT Model. The source image is divided into fixed-size patches that, after embedding, are then processed by the transformer layers. Transformer layers use multi-head self-attention to capture global dependencies across the entire image. Observe the final upsampling blocks, peculiar to our implementation.

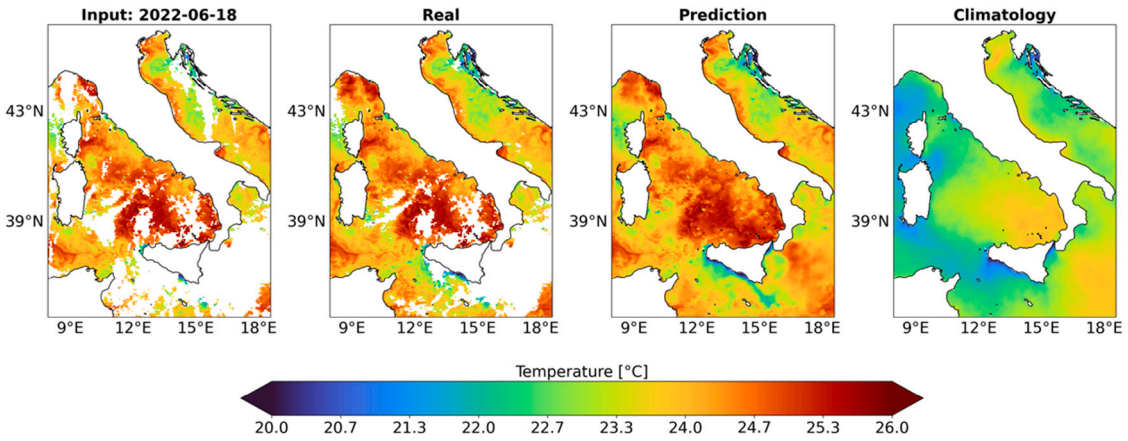

Fig. 8. The image on the left is the SST model input created by the generator adding artificial occlusions to the second image, which is the real or observed L3 image; the third image is the model prediction; the last on the right is the climatology.

then transformed into the expected output dimension. To this aim, we used a stack of three convolutional layers with kernel size 3×3.

Relative to the topic of cloud occlusion, a ViT model was used in Goh et al. (2024), where a random subset of SST patches is masked or removed at training time, mimicking a larger cloud occlusion. Afterward, a set of learnable mask tokens is added to the encoded patches before they are passed to the decoder, that reconstructs the original SST tile in pixel format.

In our case, the additional masking at training time did not result in additional advantages, probably due to our alternative introduction of artificial occluded areas described in the next section.

3.3. Generator for training and evaluation

The training/evaluation of the reconstruction model is not straightforward since we do not have a ground truth for comparison, i.e. we do not know the SST under the cloud on each specific night. Here we use the approach used also in Barth et al. (2020, 2022) and Goh et al. (2024), the so-called generator, which involves creating an artificial occluded area in the source image and restricting the evaluation to the region of the artificial clouds. The generator begins by selecting a random day from the nighttime dataset, ensuring that the chosen image has at least 40% of the visible sea, to avoid working with insufficiently informative images. For each selected day, SST measurements relative to a given, configurable number of previous days are also retrieved. This approach enables the network to capture both spatial and temporal information, allowing the model to use historical data to reconstruct missing areas in the current image. The clouds are selected in such a way as to guarantee a minimum percentage of visible sea (typically 5%), while ensuring that the artificially occluded area covers at least 10% of the sea area. These two ranges are easily configurable. The overall procedure is meant to ensure that the image has enough occluded areas to support meaningful training and validation. Artificial masks are also applied to the previous days, maintaining temporal correlation. The average percentage of visible sea in nightly MODIS data relative to the Italian seas region is around 46%. The generator produces an average visibility of around 25%, with an average artificial cloud occlusion above 40%. This is good for training since we expose the model to relatively challenging situations, but it is a bit unrealistic during testing. This is a delicate point since, not surprisingly, the performance of the model depends on the degree of occlusion of the input image, which becomes a crucial parameter of the evaluation. Starting with the real image (Fig. 8) the generator computes the real occlusion mask. The real occluded area is changed by superimposing an artificial mask for clouds of a different day (Fig. 8). The difference between the artificial mask and the real mask will define the region of the input where the reconstruction will be assessed. This approach risks introducing biases since the sea temperature under clouds is usually different from the temperature under a clear sky, but this bias is lower during the night and it is small relative to RMSE, as we will show in Section 5. The reconstruction is given in Fig. 8. It is interesting to observe that, despite the heavy occlusion of the input, the model can correctly reconstruct many details of the real image. Qualitative and quantitative evaluations will be given in Section 4.

3.4. Training

The models have been developed in the Tensorflow/Keras framework and trained using the recent AdamW optimizer, which adapts the learning rate during training, combining it with weight decay. The starting learning rate was 1e-4.

Exploratory experiments indicated that training was robust to moderate variations of learning rate and batch size, with no appreciable impact on convergence stability or final reconstruction accuracy. Training typically converged within fewer than 100 epochs.

A typical training history is reported in Fig. 9, showing stable convergence of both loss and RMSE. Using an NVIDIA A100-SXM-64GB GPU, each training epoch requires approximately 100 s, and full training is completed in less than 4 h. At inference time, reconstruction of a single SST image requires approximately 2 s on a single CPU core (Intel Xeon Platinum 8360Y), with a peak RAM usage of about 711 MB, including model loading and data processing.

4. Model comparison and selection

This section describes the numerical experiments performed with the different models and configurations to choose the optimal configuration. We primarily compare three models: two U-Nets and one ViT. The two U-Nets, called U-Net32 and U-Net64, differ in the number of channels, with the latter having double the number of channels. Both models have three downsampling layers, with a topology of [32,64,128,256] for U-Net32 and [64,128,256,512] for U-Net64. The numbers 32 and 64 in the network names correspond to the initial number of channels before downsampling. The number of parameters for the three models is provided in Table 4.

For each model, we consider two variants with different spatial dimensions: 128×128 and 256×256. The 128×128 model is a combination of four models, each trained on a different region of the input image. In this case, the evaluation metric is the average performance across the four models. Additionally, we vary the number of "s" consecutive input days used for reconstructing the SST, from the day "t" (current day) to day "t-s". We do not consider future days in the

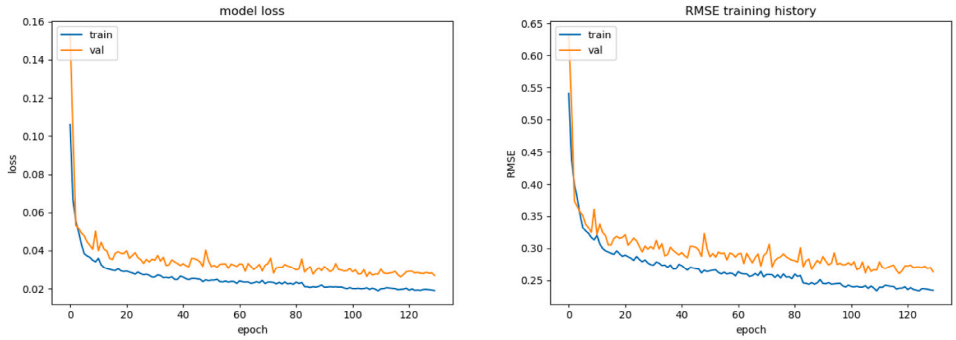

Fig. 9. Training history for the loss function (mse) and the RMSE metric. The former is relative to normalized data.

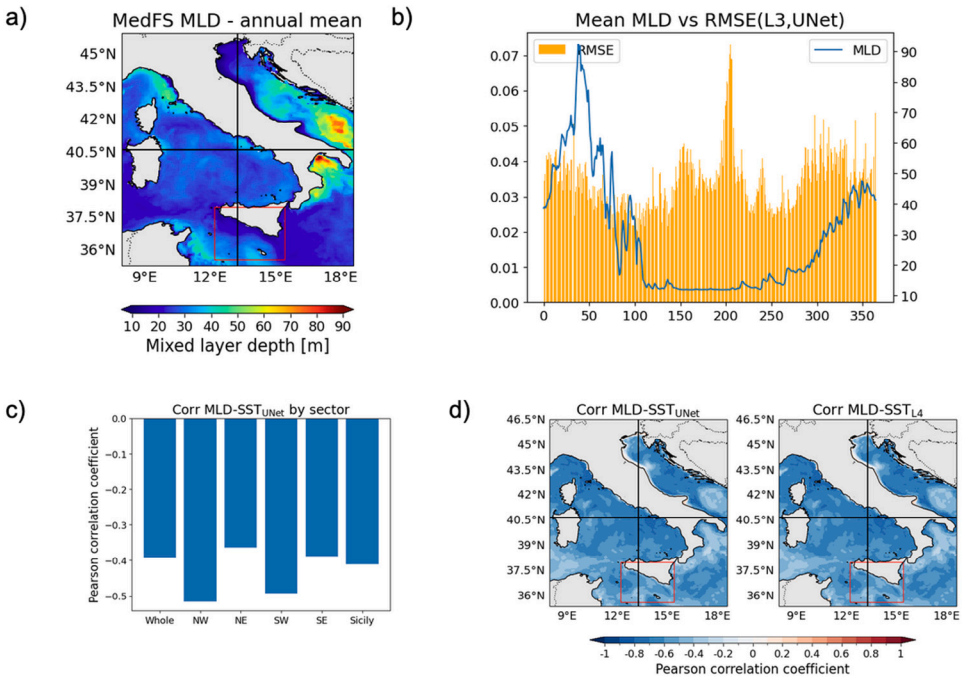

Fig. 10. (a) 2023 Mean mixed layer depth (MLD) from MedFS on the study domain. (b) The time evolution of MLD and RMSE of U-Net against L3 in 2023. (c) Correlation coefficient between MLD and U-Net SST for the four quadrants of the study domain and upwelling region near Sicily. (d) time correlation in 2023 MLD and SST from UNET (left) and L4 (right).

reconstruction. The performance score used is the Root Mean Square Error (RMSE) calculated as the difference between the reconstructed SST under an artificial cloud occlusion and the real SST at that point. The average is measured using the test generator across a total of 50 batches, each consisting of 32 samples (for a total of 1600 days). As explained in Section 3.2, the cloud occlusion generator was set to provide an average percentage of the visible sea of around 46%, which is similar to real data. The results reported in Table 5 show that the quality improvement given by increasing the number of previous days saturates after 4 days.

This is consistent with our investigation of persistence (Section 2.3). On the other hand, splitting the model into smaller geographical regions of dimension 128 × 128 each results also increases the performance. In Table 6 we report the details of the RMSE for the four quadrants of Italy; the values refer to our best model, namely U-Net64 with 4 days in input.

Reconstruction accuracy was highest in the Northern and Central Adriatic (NE) region, encompassing most of the Adriatic Sea, and lower in the Southern Adriatic (SE), which includes the Ionian Sea. To understand possible causes of these subregional differences, the 2023 MLD structure is shown in Fig. 10(a), estimated from the Copernicus Marine Service Mediterranean Forecasting System (Coppini et al., 2023, MedFS). MLD significantly impacts SST retrieval capacity by determining the effective heat capacity of the upper ocean, which dictates how sensitive the surface temperature is to atmospheric forcing. MLD acts as a moderator: a shallow mixed layer warms or cools rapidly, resulting in high-amplitude SST changes, while a deep mixed layer acts as a heat sink, keeping SST stable. The SE quadrant is dominated by a shallow MLD, except in the northern part of the Gulf of Taranto and it has the largest errors. Thus, our algorithm is not capable of capturing correctly the high-frequency time variability of SST. We argue that shallow MLD areas have the largest errors due to high-frequency processes not resolved by the input data sets as well as our algorithm as can be seen in Fig. 10(b). This is further shown by Fig. 10(c) and 10(d) where

Table 4
Number of parameters for the models. The numbers refer to the versions with 11 input channels (5 days). Shorter sequences do not notably change the total number of parameters.

	U-Net32	U-Net64	ViT
Parameters	4,259,489	17,022,273	5,918,081

Table 5
Performance of the different models, measured in RMSE (°C). The different rows refer to the number of consecutive days passed as input to the model. Values refer to data with an average percentage of visible sea around 46%. Splitting the model in 4 models with lower spatial dimension consistently gives better results.

Days	U-Net32		U-Net64		ViT	
	256 × 256	128 × 128 (×4)	256 × 256	128 × 128 (×4)	256 × 256	128 × 128 (×4)
1	0.36	0.33	0.35	0.32	0.38	0.36
2	0.35	0.32	0.34	0.31	0.36	0.34
3	0.35	0.31	0.34	0.30	0.35	0.33
4	0.34	0.31	0.33	0.30	0.35	0.33
5	0.34	0.31	0.33	0.30	0.35	0.33
6	0.34	0.31	0.33	0.30	0.35	0.33

Table 6
RMSE subdivided by quadrants. NW: North Tyrrhenian Sea, NE: Adriatic Sea, SW: South Tyrrhenian Sea, SE: Ionian Sea. The values refer to U-Net64, with 4 days in the past of input.

Quadrant	NW	NE	SW	SE	Mean
RMSE	0.305	0.283	0.304	0.314	0.302

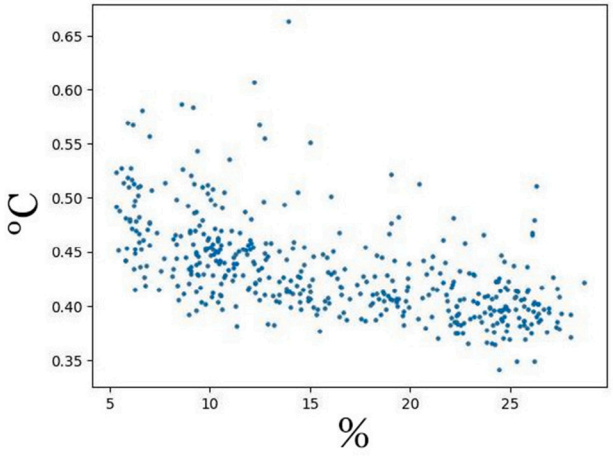

Fig. 11. Degradation of the reconstruction error (in °C) with the percentage of visible sea. Each point is a batch of 32 days.

correlation between MLD and SST is large in shallow MLD areas, while it is low in deep MLD areas where the heat dumping effects are effective both for U-Net and L4 products.

A positive feature of our configurable cloud occlusion generator is its ability to easily adjust the maximum percentage of visible sea in the model's input. Fig. 11 illustrates the relationship between visible sea percentage and RMSE, with each point representing a batch of 32 days. The plot corresponds to a U-Net32 model with 4 input days. As expected, performance degrades with higher levels of occlusion, but this degradation is nearly linear and not particularly severe. According to our investigations, there is an incremental error of approximately 0.005 °C for each additional percentage point of sea occlusion.

As mentioned in Section 3, we tested also diffusion models (Ho et al., 2020; Song et al., 2020), which are recently successfully applied to downscaling (Merizzi et al., 2024) and precipitation nowcasting (Asperti et al., 2025). Despite our efforts, we were unable to achieve competitive performance in this case.

From a methodological perspective, the subtraction of the unbiased climatology to compute anomalies gave us the best performance for the reconstruction of SST with artificial cloud occlusion. While the results without subtracting the unbiased seasonal cycle were acceptable, they consistently showed a performance decrease of around 10%. In this scenario, subtracting seasonality generally led to faster and more stable training as well as improved final performance.

Table 7
Comparison of RMSE reconstruction errors (units °C) between our 4-day input days models and DINCAE.

DINCAE	U-Net32		U-Net64	
	256 × 256	128 × 128	256 × 256	128 × 128
0.54	0.45	0.43	0.44	**0.42**

4.1. Verification with DINCAE

In this section, we compare the performance of our model with that of another state-of-the-art data-driven model: the Data INterpolating Convolutional Auto-Encoder (DINCAE2.0; Barth et al., 2022). The comparison data are daily and at the same resolution as the data on which our model was trained, originating from nighttime SST measurements by the MODIS-TERRA satellite from 1/1/2003 to 12/31/2016. These data are divided into two datasets: one with artificially added coverage and the other with original data. Unlike our approach, the additional coverage is fixed and is not configurable. The DINCAE2.0 test focuses on the Northern Adriatic Sea, the analysis area differs from the one used by our model but largely overlaps. This area extends in latitude from 40° to 46° and in longitude from 12° to 19°. From the perspective of the model, DINCAE (Barth et al., 2020) is essentially a double U-Net, where two networks are composed in sequence. Unlike our model, which only uses the available SST information, DINCAE incorporates additional indicators, including the date and wind speed, which are further extended in DINCAE2.0 to account for satellite chlorophyll.

In Table 7, we compare the RMSE of DINCAE (best model, with chlorophyll) and our models with 4 days in the past as input.

This rough comparison proves promising skill of our model while further investigation is needed for more robust intercomparison.

5. Application to operational input data sets

In this section, we explore the applicability of our U-Net64 model with 4 days input data to the real time Copernicus Marine Core Service L3 product. The latter is a calibrated cloud occluded image at 1/16° resolution, merging of several available thermal imaging sensors observations. We retrained the model between 7/4/2002 and 12/31/2023 on 1/16° resolution (see Section 2). The reconstructions are performed using L3S NRT product as input and then we compare our reconstruction with the L4 NRT product. Our purpose is to test the best configuration against the NRT products since those are used in Copernicus Marine operational analysis and forecasting systems. We refer to Pisano et al. (2022) for the detailed description of these datasets.

In Fig. 12, we compare the error by the two different reconstruction methods and the L3 input over visible regions of the sea. The error is relative to the year 2022, and is shown as a function of month. Specifically, the error of L4 is in blue, with an average RMSE of 0.14 °C, while the average error of our reconstruction (in orange) is around 0.04 °C. The error of the U-Net64 is more uniform, and particularly low in the period from January to March. Moreover, we anticipate the

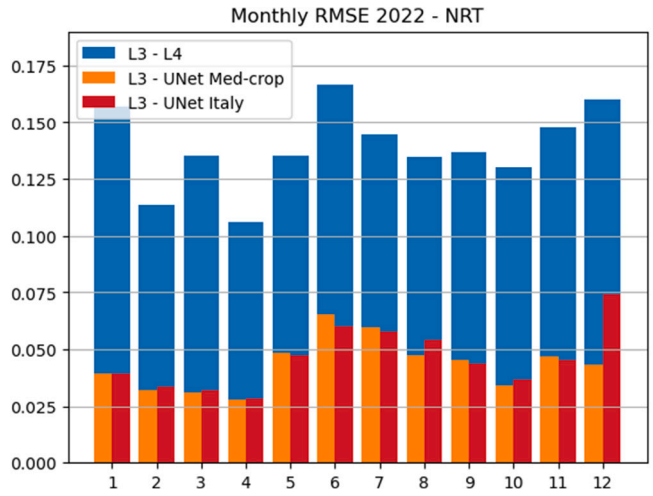

Fig. 12. (Blue) RMSE in (°C) for all months of 2022 computed over the regions not cloud occluded on the Italian Seas using L3 as reference data set. (Orange) RMSE relative to the Italian Seas only reconstruction. (Red) RMSE relative to the entire Mediterranean Sea reconstruction on the overlapping area of the Italian Seas.

Table 8
Summary of model skill statistics: Mean Bias (MB in °C), Mean Absolute Error (MAE in °C) and spatial correlation coefficient (CC).

MB	MAE	CC
0.006	0.18	0.87

preliminary comparison of the extended version of U-Net into the entire Mediterranean Sea as shown later in Fig. 14 and discussed as future work. The extended Mediterranean Sea model seems comparible to U-Net64 over the Italian Seas even though a degradation is observed in December.

Further mean statistics of our model is summarized in Table 8. Mean bias is much smaller than mean absolute error as anticipated in Section 3.3.

In Fig. 13 we qualitatively compare our reconstructions with those offered by the L4 product for two days in July 2021 and January 2022.

Our reconstruction looks more faithful in the cloud free areas, maintaining frontal regions in a manner very similar to the original data. We conclude that the trained network on the MODIS-AQUA data set also performs very well with different input data sets probably because the cloud occlusion geometry is similar in the two data sets.

6. Discussion and conclusions

This study investigated deep neural networks to reconstruct SST data gaps caused by cloud occlusion, focusing on improving data completeness and reliability. Our findings indicate that deep learning approaches, if properly tuned on spatial and temporal dimensions, improve SST reconstruction accuracy. Comparisons with existing methods, including L4 statistical interpolation reconstructions highlight the effectiveness of deep learning in achieving reliable SST reconstructions. Our selected model architectures is made of a U-Net64 algorithm with 3 or 4 days in the past input data and the subtraction of a long term unbiased seasonal cycle. Training in the subregions using smaller domains helps to improve the skill possibly because it avoids spurious remote correlations between different basins. A rough comparison with multivariate DINCAE model shows slight improvements possibly because not involving spurious correlations from other variables.

To assess the generalizability of the proposed method, we are testing its application to a larger domain, comprising the Mediterranean Sea and a portion of the adjacent Atlantic sector. This area corresponds to the Mediterranean region, as provided by CMEMS L3s datasets which were used for both training and inference. Fig. 14 presents an example of the reconstruction for 1 January 2022, corresponding to the case shown in the bottom panel of Fig. 13. (To ensure a consistent comparison, the same U-Net64 architecture and identical temporal periods were adopted for training, validation, and testing.) The results indicate that the quality of the reconstruction remains visually comparable to that achieved in the smaller domain. (Moreover, when computing the mean skill statistics of the model trained on the larger domain and restricting the evaluation to the Italian region, the performance is comparable to that of the model trained directly on the smaller domain.) Preliminary tests of the methodology have also been started for other regions such as Black Sea and South China Sea. A detailed description and validation of these additional case studies will be presented in future work.

In the future, both for training and validation purposes, we would also like to incorporate data from microwave sensors, which provide lower resolution products compared to the ones from infrared sensors, with the advantage of not being occluded by clouds. This could further enhance reconstruction accuracy, allowing models to resolve the daily cycle SST dynamics. In summary, our results underscore the potential of deep learning to enhance SST data completeness and accuracy, offering promising applications in the marginal seas.

CRediT authorship contribution statement

Andrea Asperti: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Ali Aydogdu:** Writing – review & editing, Visualization, Validation, Methodology, Investigation, Formal analysis, Data curation. **Angelo Greco:** Visualization, Validation, Investigation, Formal analysis, Data curation. **Fabio Merizzi:** Investigation, Formal analysis. **Pietro Miraglio:** Visualization, Formal analysis, Data curation. **Beniamino Tartufoli:** Visualization, Validation, Data curation. **Alessandro Testa:** Visualization, Validation, Formal analysis, Data curation. **Nadia Pinardi:** Writing – review & editing, Supervision, Methodology, Conceptualization. **Paolo Oddo:** Writing – review & editing, Supervision, Investigation, Conceptualization.

Code and data

The code developed in this work is available in the GitHub repository at the following url:https://github.com/asperti/SST_reconstruction. Main datasets may be accessed trough the notebooks in the repository or downloaded from the sites specified in the work.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This research was partially funded and supported by the following Projects:

- Future AI Research (FAIR) project of the National Recovery and Resilience Plan (NRRP), Mission 4 Component 2 Investment 1.3 funded from the European Union - NextGenerationEU.
- ISCRA Project "AI for weather analysis and forecast" (AIWAF2)
- Mediterranean Sea Monitoring and Forecasting Centre (Contract No. 24252L05-COP-MFC MED-5200) within the Copernicus Marine Service, funded by the European Union, Implemented by Mercator Ocean International.

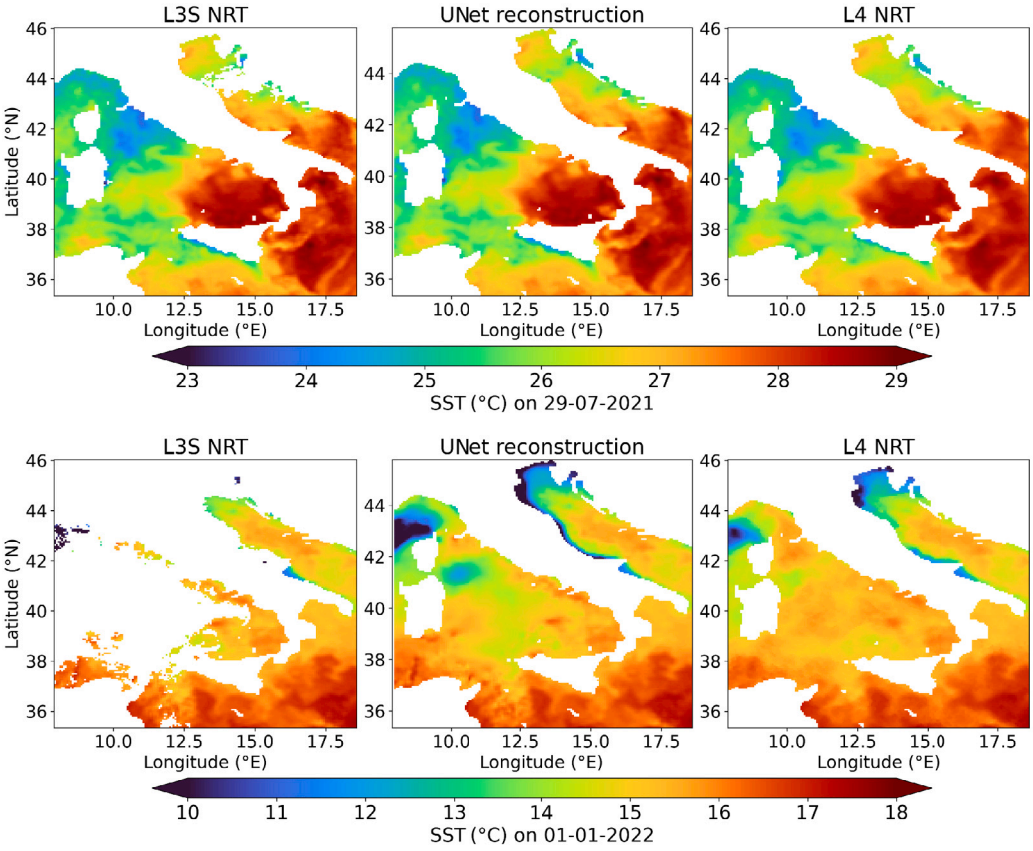

Fig. 13. Visual comparison between our reconstruction (U-Net32 with 4 input days) and L4 product from Copernicus Marine Service (marine.copernicus.eu) for 29 July 2021 (top) selected days and 1 January 2022 (bottom).

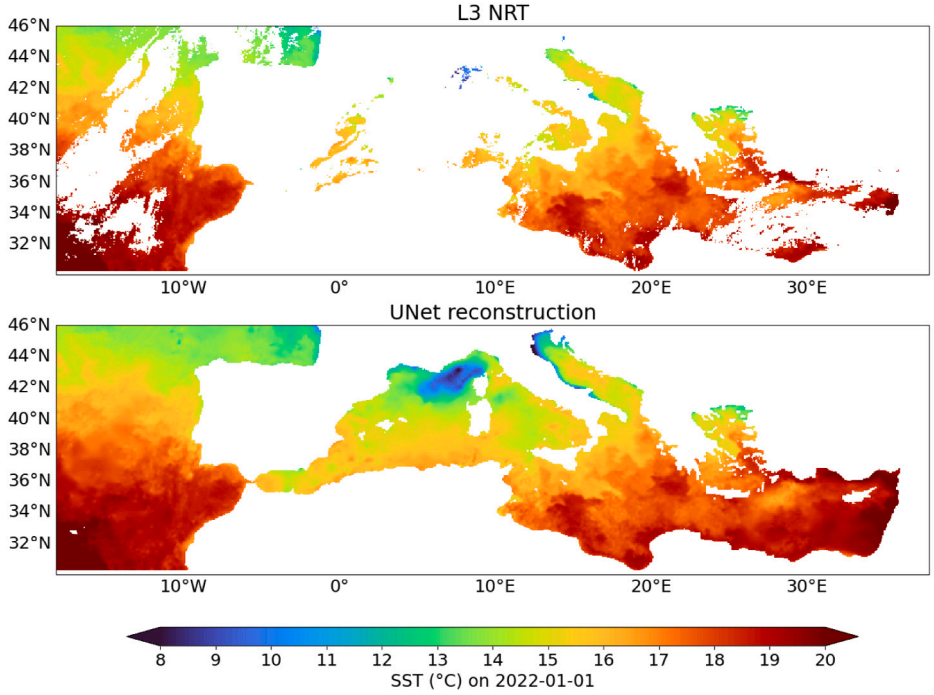

Fig. 14. Example of reconstruction with U-Net64 on the Copernicus Marine Service Mediterranean Forecasting System domain, for 1 January 2022.

• Black Sea Monitoring and Forecasting Centre (Contract No. 24252L04-COP-MFC BLK-5400) within the Copernicus Marine Service, funded by the European Union, Implemented by Mercator Ocean International.

Appendix

The application of a Gaussian blur to a matrix with no values (NaN) is not entirely straightforward. Simply replacing the no values with zeros would treat them as valid values, flattening the result. Therefore, we need to re-weight the result according to the average of the filter weights corresponding to meaningful (non-zero values) locations.

The interpolation process is divided into the following stages:

1. Replacement of missing values. The missing values (NaN) are replaced with zeros in the data matrix.
2. Application of the Gaussian filter. A Gaussian filter is applied to the resulting matrix, creating a "blurred" (biased) version of the data.
3. Generation of a binary mask. A second matrix is created to track the originally known points, assigning a weight of 0 to points with missing values and 1 to known points.
4. Application of the Gaussian filter to the mask. The resulting weights matrix describes the actual contribution of the region under the filter.
5. Division of matrices. The blurred data matrix is divided by the weights matrix to correct the original bias.

In the final division, the denominator may be 0, corresponding to a region in the input entirely composed of NaNs. In this case, mathematical libraries usually produce a NaN. After applying the filter, a few NaNs may still remain, which can be filled by iterating the technique or using other interpolation methods. It is worth mentioning that we interpolate both spatially and temporally to take advantage of persistence and produce a smoother climatology along the temporal axis. The interpolation algorithm used to create a climatology from satellite SST nighttime values should account for the fact that certain grid points consistently lack values due to cloud coverage and coastal, shallow water low temperatures. Therefore, an extrapolation procedure needs to be developed. In this study, the following code, written in NumPy style, was employed.

Algorithm 1 Gaussian Blur Interpolation with NaNs

Input: Data matrix D with NaNs, Gaussian filter G
Output: Interpolated matrix $D_{interpolated}$
$D_{copy} = D.copy()$ ▷ make a copy of D
$D_{zeroed} = D_{copy}[np.isnan(D)] = 0$ ▷ replace NaN with zero
$D_{blurred} = gaussian\_filter(D_{zeroed}, G)$ ▷ apply G to D_{zeroed}
$M = np.where(np.isnan(D), 0, 1)$ ▷ create a bynary mask
$W = gaussian\_filter(M, G)$ ▷ apply G to the mask M
$D_{interpolated} = D_{blurred}/W$ ▷ correct the bias

Data availability

All data and code used in this study can be accessed openly.

References

Alvera-Azcárate, A., Barth, A., Rixen, M., Beckers, J., 2005. Reconstruction of incomplete oceanographic data sets using empirical orthogonal functions: application to the adriatic sea surface temperature. Ocean. Model. (ISSN: 1463-5003) 9 (4), 325–346. http://dx.doi.org/10.1016/j.ocemod.2004.08.001, https://www.sciencedirect.com/science/article/pii/S1463500304000472.

Alvera-Azcárate, A., Barth, A., Sirjacobs, D., Lenartz, F., Beckers, J.-M., 2011. Data interpolating empirical orthogonal functions (dineof): a tool for geophysical data analyses. Mediterranean Marine Sci. 5–11.

Asperti, A., Merizzi, F., Paparella, A., Pedrazzi, G., Angelinelli, M., Colamonaco, S., 2025. Precipitation nowcasting with generative diffusion models. Appl. Intell. 55 (2), 1–21.

Bahdanau, D., Cho, K., Bengio, Y., 2015. Neural machine translation by jointly learning to align and translate. In: 3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings. http://arxiv.org/abs/1409.0473.

Barth, A., A. Alvera-Azcárate, M. Licer, Beckers, J.-M., 2020. Dincae 1.0: a convolutional neural network with error estimates to reconstruct sea surface temperature satellite observations. Geosci. Model. Dev. 13 (3), 1609–1622.

Barth, A., A. Alvera-Azcárate, C. Troupin, Beckers, J.-M., 2022. Dincae 2.0: multivariate convolutional neural network with error estimates to reconstruct sea surface temperature satellite and altimetry observations. Geosci. Model. Dev. 15 (5), 2183–2196.

Barth, A., Brajard, J., A. Alvera-Azcárate, B. Mohamed, Troupin, C., Beckers, J.-M., 2024. Ensemble reconstruction of missing satellite data using a denoising diffusion model: application to chlorophyll *a* concentration in the black sea. Ocean. Sci. 20 (6), 1567–1584. http://dx.doi.org/10.5194/os-20-1567-2024, https://os.copernicus.org/articles/20/1567/2024/.

Bretherton, F.P., Davis, R.E., Fandry, C., 1976. A technique for objective analysis and design of oceanographic experiments applied to mode-73. In: Deep Sea Research and Oceanographic Abstracts, vol. 23, Elsevier, pp. 559–582.

Ćatipović, L., Matić, F., Kalinić, H., 2023. Reconstruction methods in oceanographic satellite data observation—a survey. J. Mar. Sci. Eng. 11 (2), 340.

Chen, X., Zuo, Z., Nie, J., Li, X., Diao, Y., Liang, X., 2025. Sea surface temperature image completion method based on multiscale fourier fusion neural operator. IEEE Geosci. Remote. Sens. Lett. 22, 1–5. http://dx.doi.org/10.1109/LGRS.2025.3576674.

Chin, T.M., Vazquez-Cuervo, J., Armstrong, E.M., 2017. A multi-scale high-resolution analysis of global sea surface temperature. Remote Sens. Environ. 200, 154–169.

Choo, M., Jung, S., Im, J., Han, D., 2025. Care-sst: context-aware reconstruction diffusion model for sea surface temperature. ISPRS J. Photogramm. Remote Sens. (ISSN: 0924-2716) 220, 454–472. http://dx.doi.org/10.1016/j.isprsjprs.2025.01.001, https://www.sciencedirect.com/science/article/pii/S0924271625000012.

Coppini, G., Clementi, E., Cossarini, G., Salon, S., Korres, G., Ravdas, M., Lecci, R., Pistoia, J., Goglio, A.C., Drudi, M., Grandi, A., Aydogdu, A., Escudier, R., Cipollone, A., Lyubartsev, V., Mariani, A., S. Cretì, F. Palermo, Scuro, M., Masina, S., Pinardi, N., Navarra, A., Delrosso, D., Teruzzi, A., Biagio, V. Di, Bolzon, G., Feudale, L., Coidessa, G., Amadio, C., Brosich, A., A. Miró, E. Alvarez, Lazzari, P., Solidoro, C., Oikonomou, C., Zacharioudaki, A., 2023. The mediterranean forecasting system – part 1: Evolution and performance. Ocean. Sci. 19 (5), 1483–1516. http://dx.doi.org/10.5194/os-19-1483-2023, https://os.copernicus.org/articles/19/1483/2023/.

Cutolo, E., Pascual, A., Ruiz, S., Zarokanellos, N.D., Fablet, R., 2024. Cloinet: ocean state reconstructions through remote-sensing, in-situ sparse observations and deep learning. Frontiers in Marine Science (ISSN: 2296-7745) 11–2024, http://dx.doi.org/10.3389/fmars.2024.1151868, https://www.frontiersin.org/journals/marine-science/articles/10.3389/fmars.2024.1151868.

Donlon, C.J., Martin, M., Stark, J., Roberts-Jones, J., Fiedler, E., Wimmer, W., 2012. The operational sea surface temperature and sea ice analysis (ostia) system. Remote Sens. Environ. 116, 140–158.

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., et al., 2020. An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929.

Fablet, R., Viet, P.H., Lguensat, R., 2017. Data-driven models for the spatio-temporal interpolation of satellite-derived sst fields. IEEE Trans. Comput. Imaging 3 (4), 647–657.

Goh, E., Yepremyan, A., Wang, J., Wilson, B., 2024. Maesstro: Masked autoencoders for sea surface temperature reconstruction under occlusion. Ocean. Sci. 20 (5), 1309–1323.

Han, Z., He, Y., Liu, G., Perrie, W., 2020. Application of dincae to reconstruct the gaps in chlorophyll-a satellite observations in the south china sea and west philippine sea. Remote. Sens. 12 (3), 480.

Hendrycks, D., Gimpel, K., 2016. Gaussian error linear units (gelus). CoRR. abs/1606.08415. http://arxiv.org/abs/1606.08415.

Ho, J., Jain, A., Abbeel, P., 2020. Denoising diffusion probabilistic models. Adv. Neural Inf. Process. Syst. 33, 6840–6851.

Iizuka, S., Simo-Serra, E., Ishikawa, H., 2017. Globally and locally consistent image completion. ACM Trans. Graph. (ToG) 36 (4), 1–14.

Jain, J., Zhou, Y., Yu, N., Shi, H., 2023. Keys to better image inpainting: Structure and texture go hand in hand. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. pp. 208–217.

Jung, S., Yoo, C., Im, J., 2022. High-resolution seamless daily sea surface temperature based on satellite data fusion and machine learning over kuroshio extension. Remote. Sens. 14 (3), 575.

Konik, M., Kowalewski, M., Bradtke, K., Darecki, M., 2019. The operational method of filling information gaps in satellite imagery using numerical models. Int. J. Appl. Earth Obs. Geoinf. 75, 68–82.

Larson, J.G., Thompson, D.W., Hurrell, J.W., 2024. Signature of the western boundary currents in local climate variability. Nature 634 (8035), 862–867.

Liu, G., Reda, F.A., Shih, K.J., Wang, T.-C., Tao, A., Catanzaro, B., 2018. Image inpainting for irregular holes using partial convolutions. In: Proceedings of the European Conference on Computer Vision. ECCV, pp. 85–100.

Merizzi, F., Asperti, A., Colamonaco, S., 2024. Wind speed super-resolution and validation: from era5 to cerra via diffusion models. Neural Comput. Appl. 36 (34), 21899–21921.

Muc, M. Zupančič, Zavrtanik, V., Barth, A., Alvera-Azcarate, A., Ličer, M., Kristan, M., 2025. Criter 1.0: a coarse reconstruction with iterative refinement network for sparse spatio-temporal satellite data. Geosci. Model. Dev. 18 (17), 5549–5573. http://dx.doi.org/10.5194/gmd-18-5549-2025, https://gmd.copernicus.org/articles/18/5549/2025/.

Nardelli, B.B., Tronconi, C., Pisano, A., Santoleri, R., 2013. High and ultra-high resolution processing of satellite sea surface temperature data over southern european seas in the framework of myocean project. Remote Sens. Environ. 129, 1–16.

Parfitt, R., Czaja, A., Kwon, Y.-O., 2017. The impact of sst resolution change in the era-interim reanalysis on wintertime gulf stream frontal air-sea interaction. Geophys. Res. Lett. 44 (7), 3246–3254.

Park, J., Woo, S., Lee, J.-Y., Kweon, I.S., 2018. Bam: bottleneck attention module. arXiv preprint arXiv:1807.06514.

Pauthenet, E., Bachelot, L., Balem, K., Maze, G., A.-M. Tréguier, F. Roquet, Fablet, R., Tandeo, P., 2022. Four-dimensional temperature, salinity and mixed-layer depth in the gulf stream, reconstructed from remote-sensing and in situ observations with neural networks. Ocean. Sci. 18 (4), 1221–1244. http://dx.doi.org/10.5194/os-18-1221-2022, https://os.copernicus.org/articles/18/1221/2022/.

Peng, J., Liu, D., Xu, S., Li, H., 2021. Generating diverse structure for image inpainting with hierarchical vq-vae. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 10775–10784.

Pinardi, N., Zavatarelli, M., Adani, M., Coppini, G., Fratianni, C., Oddo, P., Simoncelli, S., Tonani, M., Lyubartsev, V., Dobricic, S., et al., 2015. Mediterranean sea large-scale low-frequency ocean variability and water mass formation rates from 1987 to 2007: A retrospective analysis. Prog. Oceanogr. 132, 318–332.

Pisano, A., Ciani, D., Marullo, S., Santoleri, R., new operational mediterranean diurnal optimally interpolated sea surface temperature product within the copernicus marine service, B. Buongiorno Nardelli. A, 2022. Earth Syst. Sci. Data 14 (9), 4111–4128.

Ramachandran, P., Zoph, B., Le, Q.V., 2018. Searching for activation functions. In: 6th International Conference on Learning Representations, ICLR 2018, Vancouver, BC, Canada, April 30 - May 3, 2018, Workshop Track Proceedings. OpenReview.net, https://openreview.net/forum?id={Hkuq2EkPf}.

Reynolds, R.W., Rayner, N.A., Smith, T.M., Stokes, D.C., Wang, W., 2002. An improved in situ and satellite sst analysis for climate. J. Clim. 15 (13), 1609–1625.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18. Springer international publishing, pp. 234–241.

Sammartino, M., Cioppa, L. Della, Colella, S., Nardelli, B. Buongiorno, 2025. A physics-informed deep neural network for the joint prediction of 3d chlorophyll-a and hydrographic fields in the mediterranean sea. Environ. Model. Softw. (ISSN: 1364-8152) 194, 106660. http://dx.doi.org/10.1016/j.envsoft.2025.106660, https://www.sciencedirect.com/science/article/pii/S1364815225003445.

Song, J., Meng, C., Ermon, S., 2020. Denoising diffusion implicit models. arXiv preprint arXiv:2010.02502.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A., 2015. Going deeper with convolutions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 1–9.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I., 2017. Attention is all you need. In: Advances in Neural Information Processing Systems, vol. 30.

Wan, Z., Zhang, J., Chen, D., Liao, J., 2021. High-fidelity pluralistic image completion with transformers. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 4692–4701.

Wang, A., Tang, Z., Huang, Z., Xia, X.-G., Su, H., 2026. Cross-scale 3-d thermohaline modeling via dual-residual swin transformer with multisource ocean observations. Int. J. Digit. Earth 19 (1), 2607902. http://dx.doi.org/10.1080/17538947.2025.2607902.

Wang, Q., Zhang, Q., Xie, H., Liu, Z., Dong, Y., 2025b. Sgd-sst: Seamless global daily sea surface temperature products reconstruction and validation via deep spatio-temporal fusion model. Expert Syst. Appl. (ISSN: 0957-4174) 293 (128703), http://dx.doi.org/10.1016/j.eswa.2025.128703, https://www.sciencedirect.com/science/article/pii/S0957417425023218.

Wang, H., Zhou, Y., Li, X., 2025a. Gdcm: Generalized data completion model for satellite observations. Remote Sensing Environ (ISSN: 0034-4257) 324 (114760), http://dx.doi.org/10.1016/j.rse.2025.114760, https://www.sciencedirect.com/science/article/pii/S0034425725001646.

Werdell, P.J., Franz, B.A., Bailey, S.W., Feldman, G.C., Boss, E., Brando, V.E., Dowell, M., Hirata, T., Lavender, S.J., Lee, Z., et al., 2013. Generalized ocean color inversion model for retrieving marine inherent optical properties. Appl. Opt. 52 (10), 2019–2037.

Woo, S., Park, J., Lee, J.-Y., Kweon, I.S., 2018. Cbam: Convolutional block attention module.. In: Proceedings of the European Conference on Computer Vision. ECCV, pp. 3–19.

Wylie, D., Jackson, D.L., Menzel, W.P., Bates, J.J., 2005. Trends in global cloud cover in two decades of hirs observations. J. Clim. 18 (15), 3021–3031.

Yang, X., Fan, D., Hu, W., He, H., Fu, B., 2025. Dincode: A data interpolation network with a collaborative dual encoder for reconstructing missing sea surface temperature data. IEEE Trans. Geosci. Remote Sens. 63, 1–12. http://dx.doi.org/10.1109/TGRS.2025.3638341.

Yang, C., Leonelli, F.E., Marullo, S., Artale, V., Beggs, H., Nardelli, B.B., Chin, T.M., Toma, V.D., Good, S., Huang, B., Merchant, C.J., Sakurai, T., Santoleri, R., Vazquez-Cuervo, J., Zhang, H.-M., Pisano, A., 2021. Sea surface temperature intercomparison in the framework of the copernicus climate change service (c3s). J. Clim. 34 (13), 5257–5283. http://dx.doi.org/10.1175/JCLI-D-20-0793.1, https://journals.ametsoc.org/view/journals/clim/34/13/JCLI-D-20-0793.1.xml.

Ye, H., Yang, C., Dong, Y., Tang, S., Chen, C., 2024. A daily reconstructed chlorophyll-*a* dataset in the south china sea from modis using oi-swinunet. Earth Syst. Sci. Data 16 (7), 3125–3147. http://dx.doi.org/10.5194/essd-16-3125-2024, https://essd.copernicus.org/articles/16/3125/2024/.

Zheng, C., Cham, T.-J., Cai, J., Phung, D., 2022. Bridging global context interactions for high-fidelity image completion. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11512–11522.

Zuo, Z., Nie, J., Liang, X., 2025. Fdnin: Fourier diffusion neural inpainting network for cloud-occluded missing area in sea surface temperature imagery. IGARSS 2025-2025 IEEE International Geoscience and Remote Sensing Symposium. IEEE, pp. 6186–6190.