Alma Mater Studiorum — Università di Bologna

Dipartimento di Fisica e Astronomia "Augusto Righi"
Laurea Magistrale in Fisica del Sistema Terra

Sea Temperature Monitoring From Air Temperature Data, a Neural Networks Approach

Presentata da:
Beniamino Tartufoli

Relatore:
Prof. Paolo Oddo

Correlatori:

Prof. Andrea Asperti
Dr. Jacopo Chiggiato

Anno Accademico 2023 - 2024
Appello IV

Sommario

Questa tesi analizza la possibilità di ricostruire la temperatura del mare utilizzando dati di temperatura dell'aria provenienti da stazioni meteorologiche costiere. Modelli neural networks sono stati utilizzati per la loro capacità di modellare relazioni non lineari. Lo studio si concentra sul Mar Adriatico settentrionale, utilizzando i dati delle stazioni meteorologiche di Ancona, Venezia e Trieste per ricostruire la temperatura del mare a diverse scale temporali e spaziali. L'indagine comprende tre esperimenti principali: (1) ricostruzione di misure puntuali dalla piattaforma oceanografica Acqua Alta utilizzando diversi approcci di elaborazione temporale (dati orari, orari filtrati e giornalieri), (2) ricostruzione di temperature mediate spazialmente dai dati di rianalisi CMEMS su diversi livelli verticali e domini spaziali, e (3) ricostruzione puntuale dell'intero campo di temperatura. Mentre l'analisi ha rivelato limitazioni nella ricostruzione della temperatura misurata ad Acqua Alta, ulteriormente approfondita attraverso l'analisi in frequenza, risultati più promettenti sono stati ottenuti nella ricostruzione sia delle temperature mediate spazialmente che di quelle puntuali dai dati di rianalisi CMEMS. I risultati sono stati analizzati in termini di distribuzione spaziale e temporale delle metriche di performance, evidenziando pattern sistematici nell'accuratezza della ricostruzione. Un'analisi comparativa tra temperatura osservata da satellite, rianalisi e ricostruzione tramite neural networks fornisce indicazioni sui punti di forza e le differenze relativi di questi tre approcci alla stima della temperatura.

Abstract

This thesis investigates the feasibility of reconstructing sea temperature using air temperature data from coastal weather stations. Neural networks have been used for their capability of modelling non linear relationship. The study focuses on the Northern Adriatic Sea, utilizing data from weather stations in Ancona, Venezia, and Trieste to reconstruct SST at different temporal and spatial scales. The investigation comprises three main experiments: (1) reconstruction of point measurements from the Acqua Alta oceanographic tower using various temporal processing approaches (hourly, filtered hourly, and daily data), (2) reconstruction of spatially averaged temperatures from CMEMS reanalysis data with different vertical levels and horizontal domains, and (3) pointwise reconstruction of the full temperature field. While the analysis revealed limitations in reconstructing Acqua Alta measured temperature, further investigated through coherence analysis, more promising results were achieved in reconstructing both spatially averaged and pointwise temperature fields from the CMEMS reanalysis. The results were analyzed through spatial and temporal distributions of performance metrics, revealing systematic patterns in reconstruction accuracy. A comparative analysis between satellite observed temperature, reanalysis and neural networks reconstruction provides insights into the relative strengths and differences of these three approaches to temperature estimation.

Contents

1 Introduction **3**
- 1.1 Adriatic Sea . . . 9
- 1.2 Sea temperature monitoring . . . 13
- 1.3 Deep learning . . . 16
- 1.4 Motivation and objective of the study . . . 18

2 Literature review **19**
- 2.1 Neural networks for filling gaps in satellite observations . . . 20
- 2.2 Neural networks and data assimilation . . . 21

3 Data **23**
- 3.1 Weather stations data . . . 23
- 3.2 Acqua Alta oceanographic tower data . . . 25
- 3.3 CMEMS data . . . 27
  - 3.3.1 Reanalysis data . . . 27
  - 3.3.2 Satellite data . . . 29

4 Methods **31**
- 4.1 Deep learning framework . . . 31
  - 4.1.1 Neural Network Layers . . . 31
  - 4.1.2 Optimization . . . 33
- 4.2 Preprocessing . . . 34
  - 4.2.1 Climatologies and Anomalies . . . 34
  - 4.2.2 Acqua Alta data cleaning . . . 35
- 4.3 Acqua Alta Temperature reconstruction . . . 37
- 4.4 CMEMS averaged temperature reconstruction . . . 42
- 4.5 CMEMS pointwise temperature reconstruction . . . 45

5 Results **49**
- 5.1 Acqua Alta temperature reconstruction . . . 49
- 5.2 CMEMS averaged temperature recontruction . . . 59

5.3 CMEMS pointwise temperature reconstruction 65
5.3.1 Model Performances 65
5.3.2 Comparison with spatial averaged temperature 75
5.3.3 Comparison with satellite 76
5.4 Comparison with Acqua Alta statistics 83

6 Conclusions **86**

Bibliography **89**

Chapter 1

Introduction

The ocean plays a leading role shaping Earth's weather and climate patterns. It has strong influence on the distribution of rainfall, droughts, floods, regional climate, and the development of storms, hurricanes, and typhoons [1]. The sea surface temperature (SST) is a crucial parameter in physical oceanography, for understanding, monitoring and predicting processes that happen at a variety of temporal and spatial scales. SST governs complex interactions between the atmosphere and oceans, such as fluxes of heat, momentum and gases, fundamentals for the understanding of atmosphere and ocean phenomena. It has been defined by World Meteorological Organization (WMO) as one of the essential climate variables (ECVs) contributing to the characterization of Earth's climate. SST serves as an indicator of the ocean's capacity to store thermal energy. This property is crucial in understanding climate dynamics, as a mere 3.5 meter layer of ocean water holds an equivalent amount of heat energy to the entire atmospheric column above it [2]. SST plays a vital role in climate science, offering insights into climate change and variability while also impacting weather systems, ocean currents, and the distribution of marine life. Monitoring this parameter is therefore essential.

Starting in 1981, satellite observations of SST became available, and along with *in-situ* measurements, formed the modern-era ocean observing system. Over the years more and more applications and studies having needs of high quality observations have been developed, while the existing data couldn't satisfy those needs. So minimum data specification for the use in ocean models has been set by the Global Data Assimilation Experiment (GODAE): global coverage, spatial resolution of 10 km or higher, accuracy of 0.2 °C or better. The task of achieving these requirements was taken over by the Group for High Resolution Sea Surface Temperature (GHRSST), that over the past years has reached the previously imposed specifications, with the exception of accuracy.

Despite all the improvements achieved in data quality, the continuous tech-

nological advances that allow increasingly precise and reliable measurements, and the development of new techniques to study and evaluate those observations, the monitoring of SST is still facing the problems of lack of data, that make difficult to study the ocean in some regions. Figure 1.1 shows the spatial coverage of the major shipping lanes in the 1955-2012 period, as reported by the World Ocean Atlas in 2013, along with the temporal distribution of SST. As can be seen from the figure, in the high latitudes regions there is clear scarcity of observations. This is true still today and caused by several issues that make, both *in-situ* and satellite, measurements in those regions difficult, due to the unique environmental conditions. In ice-free areas, *in-situ* measurements are made using various methods as ships, Argo floats, surface drifters and gliders (see Section 1.2) used also in other regions of the ocean. On the other hand in ice-covered areas only instruments developed specifically for work in those conditions can operate and transmit data in near real-time. These instruments have high installation and maintenance costs, so only a limited number of such measurement systems are available, causing wide gaps in the coverage of Arctic ocean ice measurements. The scarcity of *in-situ* measurements and the often lack of coincident atmospheric conditions observations affects also the satellite SST retrieval, that use *in-situ* measurements for algorithm development. Relying on too sparse data, the algorithms introduce unknown errors in conditions that differs from the relatively few observed ones. In addition there are the inherent limitations of remote sensing observations. Instruments operating in the infrared (IR) are sensible to the presence of clouds, that invalidate the measure of temperature and so they should masked. In the high latitudes this issues is particularly evident, due to the persistent cloudiness, resulting in extended gaps in the satellite products [4]. A complication, derives from the difficulty in distinguishing clouds from snow and ice. This latter can form quickly over large areas and may not be accurately mapped by daily sea ice maps. The instruments operating in the microwaves (PMW) are not affected by the presence of clouds, but most PWM missions lack a channel at 6.9 GHz, required in polar regions, due to its sensitivity below 13 °C.

Other areas where measuring SST is challenging are the inland seas, lakes and the coastal zones. The main reasons that makes difficult to estimate temperature from satellites are the greater variability in atmospheric conditions, like water vapor concentration, temperature and aerosol, than over the most of the ocean. In addition, observations in those areas should account for land contamination, due to the unresolved scales of SST sensors, water surface contamination and turbidity, that could interact with cloud detection. This can lead to observation bias, specially in the warming spring-summer period, when products face the problem of over masking, due to the increased temperature difference between

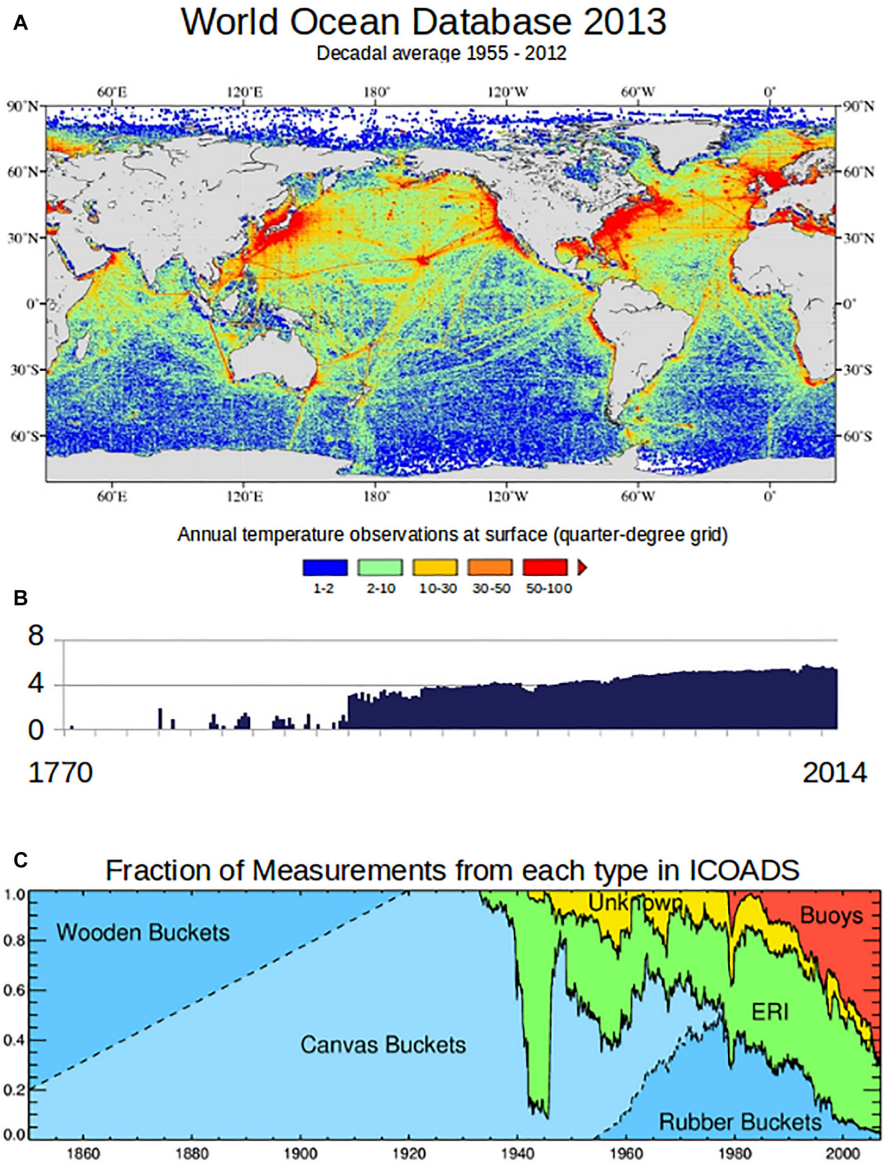


Figure 1.1: Distribution of SST measurements and by method. (A) Shows the spatial distribution of SST measurements as included in the World Ocean Atlas in 2013. (B) Shows the temporal distribution of SST measurements in the World Ocean Atlas in 2013 (from Freeman et al., 2017). (C) How the measurement methods changed over time. ERI are Engine Room Intake measurements, Buoy measurements include drifting and moored buoys (from Rayner et al., 2018). Please note that the spatial and temporal time lines are different, 1955–2012 for the spatial distribution shown in plate (A) and 1770–2014 for the temporal distribution shown in plate (B). After [3].

turbid warm water and cold clear one, possibly identifying the latter as clouds [3].

In addition to area-specific issues, there is a global growing need of high resolution SST measures due to the increasing effort of the physical oceanography community in the study of sub-mesoscale dynamic, relevant for the large scale dynamic and thus for weather and ocean variability. Those high resolution fields can be provided by IR instruments in clear sky conditions, while the PMW instruments, having lower spatial resolution, can be used to partially overcome the cloud cover problem, along with multi-scale techniques or the use of high resolution data within a larger time window. This techniques has still some limitations as: over smoothed SST fields in both time and space, retrievals errors, or errors in masking erroneous retrievals, as in the case of the presence of clouds, rain, sea ice or radio frequency-interference [3].

The need of higher spatial resolution products is particularly important in the Mediterranean Sea, where the high temporal and spatial variability of the basin is characterized by fine-scale processes, that cannot be captured otherwise. In addition the *in-situ* measurement network has significant gaps and asymmetries throughout the basin as can be seen by looking the catalog of observing platforms offered by the Mediterranean Operational Network for the Global Ocean Observing System (MONGOOS), which spatial distribution is shown in Figure 1.2 and Figure 1.3. For example, on the western European Mediterranean coast the number of tide gauges and wave buoys is sufficient for most of the applications, while on the Eastern side sensors are scarce and along the African coast measuring platform are almost completely absent, making not possible estimating the mean sea level simply from tide gauges.

On the other hand there are some variables that are under sampled throughout all the basin, like currents, temperature and salinity. For the these two latter variables Argo and gliders provide additional data that mitigate the situation, even though the number of platform are not able to provide a sufficient amount of data to properly contribute to data assimilation in numerical models [5].

All these limitations and the increasing need for monitoring solutions motivate the exploration of novel approaches.

The following sections will provide an overview of the Adriatic Sea's characteristics, discuss SST monitoring techniques, introduce the deep learning main features, and a description of the motivation and structure of this study.

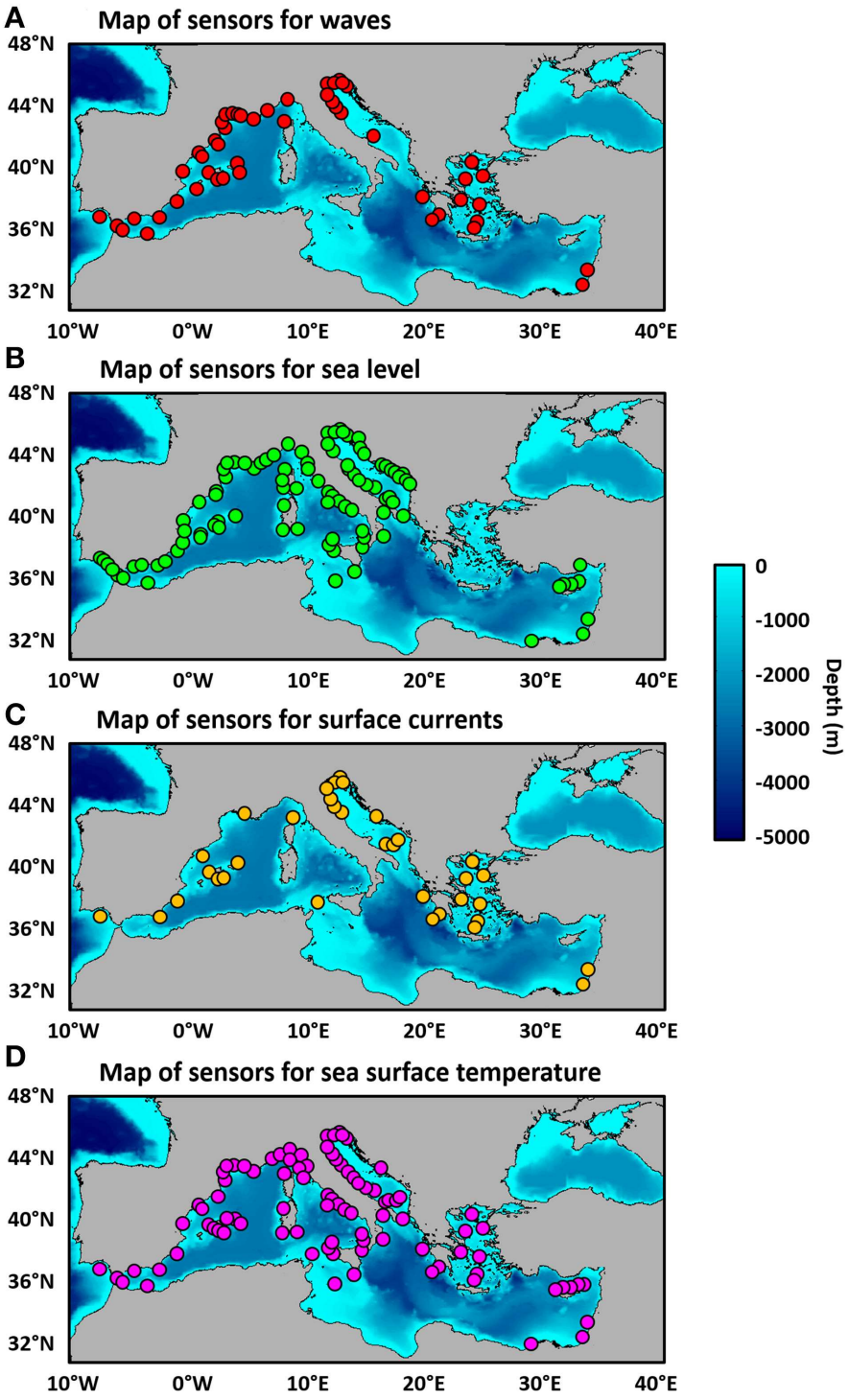


Figure 1.2: Map of sensors deployed in the Mediterranean Sea for monitoring purposes (as in 2017): (A) waves, (B) sea level, (C) surface currents, and (D) sea surface temperature. After [5].

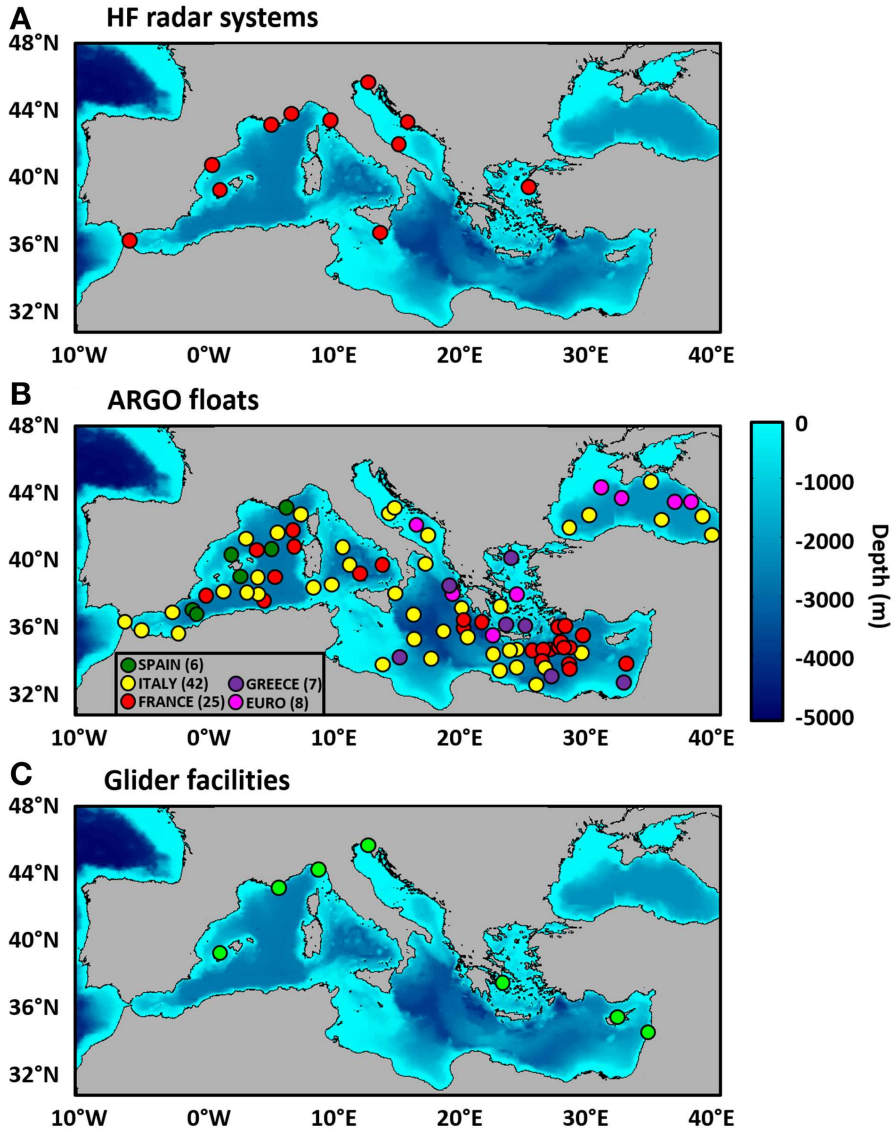


Figure 1.3: Map of sensors deployed in the Mediterranean Sea for monitoring purposes: (A) High Frequency radars, (B) Argo Floats as in 2018, and (C) Gliders. After [5].

1.1 Adriatic Sea

In a context where there is always a lack of observations, adding other source of data is surely valuable. For this reason, in this study, a new way of monitoring SST is investigated. This approach consists in the use of proxy air temperature data from coastal weather stations to reconstruct sea temperature, through the use of artificial neural networks.

The method is applied in the North Adriatic Sea region, due to the availability of data and to the crucial importance of SST monitoring in this area, due to its sensibility to climate change and its influence on weather patterns. In [6] SST variations are studied over a period of 37 years, from 1982 to 2018, finding that the Adriatic Sea is experiencing a more steep warming trend than the rest of the Mediterranean. The study computed a trend of 0.041(6) °C yr^{-1} in the Mediterranean, sensibly higher compared to the trend found in the Northeastern Atlantic, adiacent to the Mediterranean, of 0.027(80) °C yr^{-1}. The study analyzes also the different trends across different regions of the basin, finding the higher values in the Levantine Sea and in the Adriatic Sea, with respective values of 0.048(60) °C yr^{-1} and 0.045(70) °C yr^{-1}. In [7] the study of a severe event that hit Italy on 10 July 2019, causing heavy damage because of giant hailstones reaching the ground, is conducted using numerical simulations. This study evidenced the key role of the SST anomaly over the Adriatic, along with the topography, in the development of the event, causing the instability of the environment, and thus highlighting the importance of the monitoring of the sea temperaure, also for its impact on regional weather.

The Adriatic Sea is a part of the eastern Mediterranean Sea located between the Italian peninsula and the Balkan peninsula. It borders six countries: Italy, Slovenia, Croatia, Bosnia and Herzegovina, Montenegro and Albania, and it's connected the Ionian Sea to the southeast, trough the Strait of Otranto. It has a surface area of 138 600 km^2, a volume of 35 000 km^3 [8], with an extreme length of about 800 km and mean width of about 160 km [9], with its major axis in the northwest–southeast direction. It constitutes the northernmost part of the Mediterranean, extending as far North as 45°47'N.

The Adriatic longitudinal variations in the bathymetry define three zones: northern, middle and southern Adriatic. The northern part of the basin is very shallow with an average bottom depth of about 30 m. Starting with a depth of about 15 m along the Venice-Trieste coastline, increasing southward, it extends till the 100 m isobath. The depth reach about 270 m in the middle Adriatic, in the two depressions called the Pomo Pits [8]. To the South of this pit the bathymetry rises in the Palagruža sill, separating the middle Adriatic from the southern Adriatic. In the southern Adriatic there is the deepest part of the basin, with

a maximum depth of 1270 m. This depression is often referred as the South Adriatic Pit. Further South, the bottom rises again, forming the Otranto Sill (780 m), which separates the Adriatic from the Ionian Sea [9]. There are also transversal variations in the bathymetry: western coastline is mostly regular and smooth, with no islands and a gentle shelf, while the eastern side is characterized by the presence of many islands and irregular bottom, with a steep shelf [8]. The bathymetry is shown in Figure 1.4, along with the main currents.

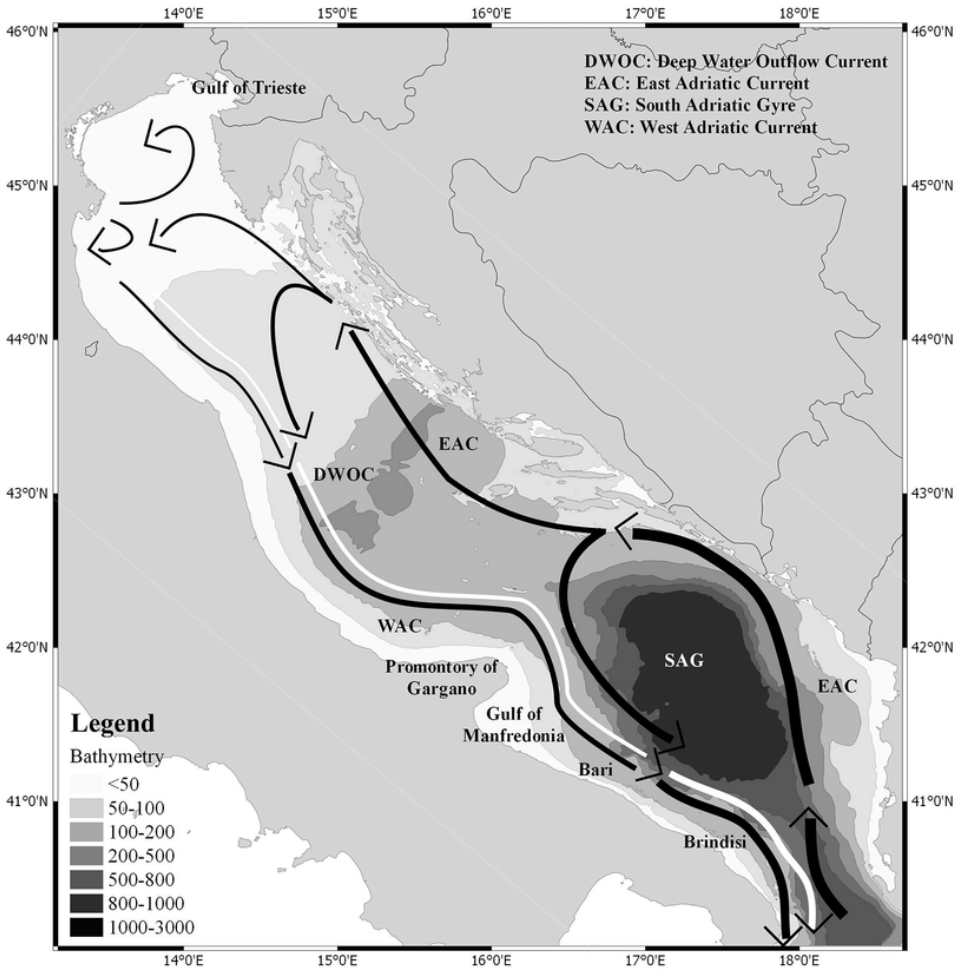


Figure 1.4: Adriatic Sea surface circulation. After [10].

The dominant winds on the Adriatic Sea are the bora and sirocco, that characterize winter wind regime. The name bora is generally used for the strong northeasterly cold winds that enters the Adriatic Sea through the Trieste gap and other gaps of the Dinaric Alps on the eastern shore of the Adriatic in the pres-

ence of strong pressure gradients between the two side of the mountain range. This condition can happen in the presence of two primary weather patterns: the anticyclonic pattern and the cyclonic pattern. In the anticyclonic pattern a large high pressure center is present over central Europe without a well defined low to the South. In the cyclonic pattern a low pressure center is present over the Tyrrhenian Sea, the southern Adriatic Sea or the Ionian Sea. Bora winds are most common during the cool season (November through March). On the other hand, sirocco is the name commonly used for winds of continental tropical origin when blowing over the Mediterranean Sea. Sirocco blows over the Adriatic Sea, after originating from the deserts of North Africa and Arabia and passing over the Ionian Sea. Therefore sirocco wind brings humid and warm air, since the it has picked up moisture during its passage over the Ionian Sea, often accompanied by Saharan dust. It tends to occur more frequently in the southern regions of the Adriatic with a decrease in frequency northward.

During summer the prevalent winds are northwesterly winds, the etesian (or meltemi) winds and the mistral. The former blow mostly over the Aegean Sea, when there is the presence of a continental depression centered over southwester Asia, but can extend to the southern Adriatic basin. The mistral is a strong wind that originates in the rear of cold fronts. Along the coasts daily sea-breeze variations prevails over those winds [8].

Being subjected to interaction with air masses of different origins, the Adriatic Sea is characterized by substantial heat exchanges with the atmosphere. The total heat flux Q_T (positive when it's directed from the air to the sea) is the sum of four components, namely

$$Q_T = Q_S + Q_B + Q_H + Q_E \tag{1.1}$$

where Q_S is the incident solar radiation flux, Q_B the backward long-wave radiation flux, Q_H the sensible heat flux, and Q_E the latent heat flux. The annual climatological Q_T over the Adriatic Sea is negative, so, on average, the basin loses heat to the atmosphere, implying that, under steady-state conditions, it imports heat from the Mediterranean Sea through the Strait of Otranto. Several studies have been made to quantify the average annual heat loss, over different periods, with values of $Q_T = -22\,\mathrm{W\,m^{-2}}$ for the period 1954-1988, $Q_T = -19(10)\,\mathrm{W\,m^{-2}}$ for the period 1980-1988 [11], $Q_T = -17\,\mathrm{W\,m^{-2}}$ for the period 1991-1994 [12] and $Q_T = -26\,\mathrm{W\,m^{-2}}$ in the 1998-2001 period [13]. In the latter two studies also interannual variability has been investigated, and found a big variation in the 1991-1994 period, with an average value of $-24\,\mathrm{W\,m^{-2}}$ for 1991-1993 and a value of $4\,\mathrm{W\,m^{-2}}$ in the 1994.

The direct water exchange between sea and atmosphere consists of evaporation and precipitation. Evaporation is related to the latent heat flux, which is

maximum in autumn and winter and minimum in summer. The evaporation rate, E, is usually estimated from the expression $E = Q_E/L_v$, where Q_E is the latent heat flux and L_v the latent heat of vaporization. Precipitation regimes differs over the region: in the northern region, winter is the driest season and maximum precipitation is observed in autumn, whereas in the remaining part of the basin the precipitation minimum occurs in summer and maximum in winter. Considering the basin average, precipitation maximum occur in late autumn, while minimum is in summer.

The Adriatic Sea receives a significant portion of the Mediterranean's freshwater inflow, accounting for up to one-third of the total. The major contributor to this influx is the Po River that account for the 28% of the total freshwater input. Flow rates are modulated by the precipitation and snow melting regimes characterizing the different sections of the drainage basin. The primary flow rate maximum is related to the precipitation maximum in autumn, while snow melting contributes to the spring peak. The minimum flow rate is in summer, except for the northern coast rivers, which exhibit a winter minimum.

The water budget is the net flux of water entering or leaving the basin. By taking into account only evaporation (E) and precipitations (P) we can define the surface water budget $W' = E - P$ (positive for a water flux from the sea to the atmosphere). The long-term annual surface water budget W' is positive both in the Adriatic and in the whole Mediterranean Sea. By taking into account also the river runoff (R) we define the total water budget

$$W = W' - R = E - P - R. \tag{1.2}$$

The annual estimates of W are negatives for the Adriatic Sea, denoting it as a dilution basin. In [11] the water budget is estimated to be $W = -1.14(20)\,\mathrm{m\,yr^{-1}}$. On the other hand annual values of W for the whole Mediterranean Sea are positive, making it a concentration basin [8].

The primary circulation pattern of the Adriatic Sea is characterized by thermohaline dynamics, driven by buoyancy gain, due to river runoff, heating and precipitation, and loss, caused by the evaporation and cooling. There are two kind of thermohaline circulation in basins, depending on the density of water in the basin with respect to the external water. When the density in the basin is less that the outside, this is called positive thermohaline circulation, in the opposite case, negative thermohaline circulation. The Adriatic exhibits both types of circulation [14]. The Norhern Adriatic receives substantial river water inflow which promotes positive circulation. Simultaneously, strong winter evaporation leads to the formation of denser water, responsible of the negative circulation. The coexistence of the two phenomena is possible because riverine waters flow mainly along the coast and only a small part mix with the interior waters, without inhibiting the dense water production. The positive circulation occurs when

light water accumulates in the basin, leading to a rising of the sea level and a consequent pressure gradient force directed outside the basin. As result of this force, the geostrophic response leads to a convergence toward the boundary on the right side (Italian coast), creating consequently a rise in the sea level along the coast. The resulting current is called the Western Adriatic Current (WAC) and is mainly confined on the shallow italian shelf. The negative circulation is driven by the formation of dense water that sinks. Dense water induces a baroclinic pressure gradient force that combined with the geostrophic response creates a bottom current that flows outside the basin following the western boundary. This current is called the Deep Water Outflow Current (DOWC). The outflowing of the WAC and the DOWC causes a lowering of the sea level inside the basin that in turn causes the formation of an inflowing current from the Ionian Sea through the Strait of Otranto. Since both outflowing currents export water on the western side, the inflowing current take place on the eastern side, and it's called the Eastern Adriatic Current (EAC). Differently from WAC, EAC is not confined nearshore but it's wider and it spread in the basin accordingly to the sea level distribution. In addiction to these main currents, to this main currents, observations have revailed the presence of gyres. A cyclonic gyre is present in the northern Adriatic, with prevalence in the autumn. More extended gyres are present also in the middle Adriatic and southern Adriatic.

A schematic representation of the circulation is shown in Figure 1.4

1.2 Sea temperature monitoring

Observations of the ocean can be divided in *in-situ* observations, that are direct measurements of sea water properties made using a combination of sensors on a wide range of platforms, and remote observations, made using remote sensing techniques from satellites.

***in-situ* observations** Measurements of sea surface temperature have been made for over 200 years but they were very sparse until 1850's when measurements became more systematic [15]. First observations were made from ships, by collecting sample of water in a bucket and measuring its temperature, leading to various inaccuracies and inconsistencies, due to many factors as the material and shape of the bucket and the atmospheric conditions. Figure 1.1B shows the evolution of the measuring instruments over time as recorded by the International Comprehensive OceanAtmosphere Data Set (ICOADS). Nowadays ship-based measurements relies on measuring instruments that automatically collect data, like thermosalinograph (TSG) and expendable bathythermograph

(XBT). TSG measures sea surface temperature and conductivity, used for computing salinity, from a water intake, while also keeping track of GPS data. XBT is a probe that dropped from a ship measures the temperature as it falls through the water. The probe is designed to fall at a known rate, so that the depth of the temperature profile can be inferred from the time since it enters the water.

The prevalent source of *in-situ* measurements comes from drifting buoys that collect data following the ocean currents providing a dynamic and wide-ranging view of surface conditions. Maintaining an evenly distribution is challenging, due to the difficulty of deployment at high latitudes and the tendency to cluster in the zones of ocean currents convergence.

On the other hand, moored buoys provide continuous, high-frequency measurements at fixed locations, complementing the spatial coverage of drifting buoys with long-term time series data at specific points of interest.

The Argo program further expanded oceanic monitoring capabilities with an array of approximately 4,000 floats. These autonomous devices operate on a 10-day cycle: drifting at 1,000 m depth for 9 days, then descending to 2,000 m before ascending to the surface while measuring temperature and salinity throughout the water column. In the context of the Mediterranean Sea the Argo program has been developed to adapt the characteristics of these devices to the features of this basin, in the program called MedArgo. The profilers are thus programmed to execute 5-day cycles, drifting at a modified parking depth of 350 m before to execute the vertical profiling, from either 700m or 2000m up to the surface [16].

Complementing the Argo program are ocean gliders, which are autonomous underwater vehicles, that can be equipped with various sensors to measure temperature, salinity and ocean currents. They move through the water column, ascending and descending with changes in buoyancy while they can also move horizontally and be programmed to follow specific routes.

Since 2004 several hundred thousand of temperature and salinity profiles have been collected by CTD (conductivity, temperature and depth) instruments mounted on marine mammals. This data are very useful for ocean modeling and sea ice verification in high latitudes [17].

In the Adriatic sea an important source of *in-situ* observations is the oceanographic tower called *Acqua Alta*, located 8 nautical miles off the coast opposite Venice. It has been operative since 1970 and it's the world's only offshore operating platform that allows researchers and technicians to remain on board for extended periods during measurement campaigns.

Satellite observations Since the beginning of operational satellite SST in 1981 [17] satellite observations have provided data SST data on global scale, constitut-

ing a crucial part of the ocean monitoring. The retrieval of SST from satellite is made based on the Plank equation, that links the thermodynamic temperature to the radiation emitted, for an ideal blackbody:

$$B_\lambda(T) = 2hc^2\lambda^{-5}\left(e^{\frac{hc}{\lambda kT}} - 1\right)^{-1} \tag{1.3}$$

where h is the Planck constant, c is the speed of light in vacuum, k is the Boltzmann's constant, λ is the wavelength, and T is the temperature in Kelvin. $B_\lambda(T)$ is the radiant energy flux for unit wavelength. The temperatures derived from the radiance measurements of satellite radiometers, using Planck's Function are called brightness temperatures.

Even though ocean surface radiant energy flux is close to the blackbody one, in particular at certain wavelength [18], its emissions are a fraction of $B_\lambda(T)$. This fraction is called emissivity ϵ_λ and is wavelength dependent. The emitted radiation interacts with the atmosphere, where it's scattered or absorbed, and only a fraction f_λ reach the top of the atmosphere. Also the atmosphere emissions L_λ should be taken into account, so the total radiation measured by the satellite sensor is $L_\lambda + f_\lambda\epsilon_\lambda B_\lambda(T)$.

There are two kind of satellite sensor for SST retrieval: infrared (IR) and passive microwaves (PMW) radiometers. Infrared radiation is emitted from a skin ocean layer with thickness less than 0.1 mm. This layer is cooler than the water beneath, so IR radiometers derive temperature lower of 0.17 K on average, with respect to the subskin layer [19]. SST retrieved from IR sensor has a typical spatial resolution of from 0.75 to 4 km at nadir, higher than PMW based observations.

The main issue of IR observations is that infrared radiation emitted from the sea surface highly interact with atmosphere and this lead to several problems. First of all the brightness temperature measured by a satellite is not the same as would be measured by a radiometer of comparable accuracy just above the sea surface. This is due to the high variability of the trasmissivity of clear-sky atmosphere in the IR, both with wavelength and with concentration of atmospheric gases. Thus SST retrieval in the IR band is it possible only in the so called atmospheric windows, where the atmosphere is relatively transparent.

Another issue of the satellite measurements is that in the presence of clouds the SST retrieval process can produce significant errors, so effective cloud screening algorithms are used to exclude cloudy pixels from the SST field [20] [21]. Also in clear sky conditions, atmosphere molecules absorb sea surface emission and also emit radiation. Being at lower temperature than SST they have a cooling effect on the brightness temperature at satellite height. A variety of algorithm have been developed for retrieving SST, taking into account this effect [19].

On the other hand PMW radiometer measures the temperature at the subskin layer, at approximately 1 mm depth. As previous mentioned, they have coarser spatial resolution of about 25 km [17]. A great advantage of PWM derived SST measures is that they are largely insensitive to the presence of clouds, except in the presence of heavy rainfall [3]. Microwave radiation from sea surface can be contaminated by radio frequency interference from ground based sources, geostationary satellites and communication satellites.

1.3 Deep learning

Deep learning is a subset of machine learning that employs artificial neural networks models to find pattern in the data. One approach of machine learning is the supervised learning, in which the aim is to infer the relationship between inputs and targets in a set of labeled data. Let $\mathbf{x} \in \mathbb{R}^M$ be the vector of input variables, called *features* and $\mathbf{y} \in \mathbb{R}^N$ the vector of the targets, the task of a machine learning model is to find the mapping from features to targets, given by the function $f(\cdot)$ such as:

$$\mathbf{y} = f(\mathbf{x}). \tag{1.4}$$

Machine learning models doesn't solve the problem analytically, instead they find the best representation of $f(\cdot)$, using a parameterized function f_θ where θ are the model's parameters, such that the output of the model is:

$$\hat{\mathbf{y}} = f_\theta(\mathbf{x}). \tag{1.5}$$

The model's parameters θ are optimized by minimizing a *loss function* $\mathcal{L}(\mathbf{y}, \hat{\mathbf{y}})$, that quantify the misfit between the true target $\mathbf{y}$ and the output of the model $\hat{\mathbf{y}}$. The parameters are not optimized in a single step, but instead the optimization take place in a iterative way in which the parameters are updated at each step. This process is the learning phase of the model, called *training*.

Artificial neural networks are mathematical models that mimics the information processing in brain, where each of the billions of neurons is connected to other neurons by thousand of synapses, forming a complex network with around 10^{14} connections [22]. The ancestor of the modern neural networks is the *perceptron*, firstly introduced by the scientist Frank Rosemblatt in the 1950s and 1960s [23]. The perceptron can be described as a model that takes in input a set of input variables $\{x_1, \ldots, x_M\}$ and performs a linear combination, assigning a weight w_i to each variable, and adding also a constant value b called bias. The output of the perceptron is binary, and is given by the application of a step

function to the sum of the linear combination and the bias:

$$\text{output} = \begin{cases} 0, & \text{if} \quad \sum_{i=1}^{M} w_i x_i + b \leq 0 \\ 1, & \text{if} \quad \sum_{i=1}^{M} w_i x_i + b > 0. \end{cases} \tag{1.6}$$

The perceptron is the basic unit of neural networks but its capabilities are limited. It's possible to overcome this limitations by stacking together various perceptron units, by forming the so called multi-layer perceptron (MLP). The basic perceptron use the step function to mimic the biological neuron activation, so if the neuron sends a signal or not. This function is then called *activation function* and in artificial neural networks serves to introduce non linearity into the network, which is crucial for learning complex patterns. Without activation functions, neural networks would essentially be just a series of linear transformations, limiting their ability to learn and represent complex relationships in data. In MLP activation functions other than the step function must be used. This is primarily due to the use of gradient-based optimization techniques such as gradient descent and backpropagation, which require differentiable activation functions. The step function, being non-differentiable, is unsuitable for these methods. Common activation function are the sigmoid σ, the rectified linear unit (ReLU), the hyperbolic tangent, defined as:

$$\text{ReLU}(x) = \max(0, x) \tag{1.7}$$

$$\sigma(x) = \frac{1}{1 + e^{-x}} \tag{1.8}$$

$$\tanh(x) = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}} \tag{1.9}$$

A layer is formed by stacking in parallel N units, each with their own weights and biases. So a layer of this kind has M inputs and N outputs, with each of the outputs being connected to all inputs, forming the so called fully connected layer or *dense* layer. Layers are then concatenated, with the outputs of one layer forming the input of the next one. Layers between the input and the output are called hidden layers and a network with more than one hidden layer is called deep network.

Modern deep learning has evolved developing several architectures like Convolutional Neural Networks (CNNs) for image processing and gridded data, Recurrent Neural Networks (RNNs) for sequential data, Transformers for natural language processing, and Graph Neural Networks (GNNs) for graph structured data.

In oceanography and environmental sciences, deep learning has shown his results in several studies. Neural networks have found application in prediction,

optimization of the remotely sensed data and improve spatial resolution of sea surface variables, like temperature, salinity, currents, height. They have also been used in the identification in the identification of sea ice, and track detection and prediction of mesoscale eddies, fronts and internal waves [24].

1.4 Motivation and objective of the study

The aforementioned *in-situ* and satellite based methods for measuring sea temperature give a very good description of this quantity and have significantly advanced our understanding. However they still face some limitations as the gaps in the data retrieved by satellite measures in the presence of clouds, and the sparsity of *in-situ* measurements.

This study proposes an approach to sea temperature reconstruction using air temperature data, from coastal weather stations. Neural networks are used to achieve this, because of their ability to capture complex, nonlinear relationships between inputs and targets. This approach has several advantages and can potentially give some interesting insights.

- Neural networks can potentially fill gaps in satellite SST data where direct measurements are unavailable or infrequent, particularly valuable for regions frequently obscured by cloud cover.
- By leveraging the existing coastal weather stations network, this method potentially allow for more comprehensive sea temperature monitoring without significant increases in operational expenses.
- There may be longer historical records of air temperature from weather stations compared to sea temperature measurements, allowing for potential reconstruction of historical data.
- Weather station data is often available in real-time, whereas satellite and *in-situ* data may have longer processing and distribution delays.

The primary objective of this study is to develop and validate a neural network model for reconstructing SST in the North Adriatic Sea using air temperature data from three weather stations located at Ancona, Venezia and Trieste. These stations were selected because of their strategic coastal locations and the potential influence of local weather conditions on sea temperature.

Three experiments were conducted, varying the target spatial extension.

Chapter 2

Literature review

Deep learning and neural networks have seen an increasing number of interest in recent years and found applications in a variety of different fields. One of these is oceanography, where the number of published paper that involve deep learning in the oceanographic science is greatly increased in the last years [24], as shown in Figure 2.1.

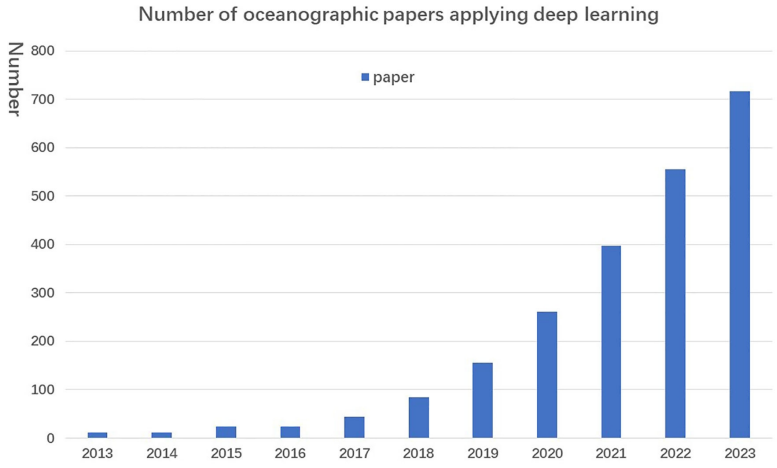


Figure 2.1: Trends in the number of papers published on the application of deep learning in oceanography retrieved from the Web of Science each year since 2012. After [24].

Among hundred of publications in this section two of them are presented that involve the use of neural networks in a data assimilation framework and the implementation of a convolutional neural network for filling gaps in SST satellite observations.

2.1 Neural networks for filling gaps in satellite observations

One of the popular application of neural networks in oceanography is the reconstruction of gaps in the satellite observations. In [25] and [26] a convolutional neural network is used to fill missing data due to the presence of clouds or gaps between tracks in satellite data. The architecture they used, called DINCAE (Data INterpolating Convolutional Auto-Encoder), can handle non gridded data in input, and provide gridded fields as output with also variances of the reconstructed pixels. In order to do this, instead of using a standard L2 or L1 loss function they used the negative log-likelihood of a Gaussian with mean $\hat{y}$ and variance $\hat{\sigma}^2$:

$$J(\hat{y}_{ij}, \hat{\sigma}^2_{ij}) = \frac{1}{2N} \sum_{ij} \left[\left(\frac{y_{ij} - \hat{y}_{ij}}{\hat{\sigma}_{ij}} \right)^2 + \log(\hat{\sigma}^2_{ij}) + 2\log(\sqrt{2\pi}) \right] \tag{2.1}$$

where the sum runs over all grid points with valid values and N is the number of these values. The first term of the right-hand side of the equation is the MSE, but scaled by the estimated error standard deviation, the second term penalizes any overestimation of the error standard deviation.

The structure of the network is constituted of two sequential Unet auto-encoders, where the output of the first is concatenated to the inputs and passed to the second auto-encoder. This approach is called refinement. An intermediate loss funtion is computed with the output of the first auto-encoder and combined with the loss function at the final auto-encoder. The final loss is therefore:

$$J_r = \alpha J(\hat{y}_{ij}, \hat{\sigma}^2_{ij}) + \alpha' J(\hat{y}'_{ij}, \hat{\sigma}'^2_{ij}) \tag{2.2}$$

Where $\hat{y}'$ and $\hat{\sigma}'$ are the outputs of the second auto-encoder and the weights α and α' controls the importance given to the intermediate output.

Input data of the network are: the values of the variables divided by the corresponding error variance and the inverse of the error variance. In this way missing data are treated as data with infinitely large errors and therefore the corresponding input is zero.

The final layer of the network produces a 3D array $T_{ijk}$where the first two dimensions represents the size of the grid and the third dimension contains the output from which the values of the variable and their error variances are derived as:

$$\hat{\sigma}^2_{ij} = \frac{1}{\max(\exp(\min T_{ij1,\gamma})\mu)} \tag{2.3}$$

$$\hat{y}_{ij} = T_{ij2}\hat{\sigma}^2_{ij} \tag{2.4}$$

where the parameters γ and μ and the max and min function are introduced to prevent the division for a value close to zero or the exponentiation for a too large number.

The problem of non gridded data arise from the convolutional layer of the neural network, that requires data discretized on a rectangular grid. The approach to address this issue has been to compute the interpolation of the weights of the first convolutional layer, discretized on a regular grid, to the data location of the non gridded data. Instead of interpolating the weights directly, the adjoint of the linear interpolation is applied to the non-gridded values. In this way the convolution can be implemented using the optimized functions of the network library.

This approach has been compared with a more classical method based on empirical orthogonal function, proving an effective improvement of the score in the reconstruction.

2.2 Neural networks and data assimilation

In [27] neural networks are used in a data assimilation experiment, that consist in the assimilation of transmission loss (TL) data into an oceanic model. The study compare the capability of both a neural network (NN) and canonical correlation analysis (CCA) in the reconstruction of the TL profile using temperature data, and use the relative models as observational operator in the cost function of the data assimilation scheme. Namely, the two statistical methods aim to build the operator such as:

$$\mathbf{y}^{TL} = H^{AC}(\mathbf{x}) + \epsilon^{TL} \tag{2.5}$$

where $\mathbf{y}^{TL}$ is the vector of the acoustic observations, $\mathbf{x}$ is the temperature field and ϵ^{TL} is the associated error. And

$$H^{AC}(\mathbf{x}) - H^{AC}(\mathbf{x^b}) \simeq \mathbf{H}^{AC}(\mathbf{x} - \mathbf{x^b}) = \mathbf{H}^{AC}\delta\mathbf{x} \tag{2.6}$$

where $\mathbf{H}^{AC}$ is the tangent-linear observational operator, linearized around the background state $\mathbf{x^b}$. The models are trained using an ensemble of NEMO model simulations that provide the temperature fields for the input of the NN, while sound speed fields have been used to run the acoustic model RAM and obtain the transmission losses, used as output. The NN model implemented in the study is composed of two dense layers, with ReLU activation function, and an output layer without activation, and trained using root mean error loss function. Since the assimilation scheme needs a linearized operator, the NN model had to be linearized around the background fields $\mathbf{x^b}$. The results of the reconstruction of TL data from temperature data shown that NN outperform CCA in this task,

also in the linearized version. In the assimilation experiment both methods shown improvements compared to the control run, with better results for the NN implementation in reducing the RMSE of temperature and mixed layer depth. Another advantage of the NN approach is that it's able to capture the temporal changes of the covariations between input and output, while adapting CCA to be flow dependent is unfeasible for real world applications.

Chapter 3

Data

In order to obtain input features and target to train and evaluate the neural network models we need historical atmospheric data and sea temperature data. Coastal weather stations have been the source of air temperature, wind and precipitation data, while sea temperatures have been obtained from two source: Acqua Alta oceanographic tower for puntual observations and CMEMS reanalysis for gridded data. In addition, satellite data are used to perform a cross comparison between the results obtained from the model reconstruction, the reanalysis data and satellite observations

3.1 Weather stations data

Atmospheric data have been obtained from weather stations located at Ancona, Venezia and Trieste. Since those stations are located in three different regions, data are stored in databases of the respective Protezione Civile and Arpa websites, namely:

- Protezione Civile Marche, Sistema Informativo Regionale Meteo-Idro-Pluviometrico (SIRMIP): http://app.protezionecivile.marche.it/sol/indexjs.sol?lang=it.
- Arpa Veneto (ARPAV): https://www.ambienteveneto.it/datiorari/.
- Arpa Friuli Venezia Giulia, Osservatorio meteorologico regionale (OSMER): https://www.meteo.fvg.it/archivio.php?ln=&p=dati.

Ancona Regarding temperature measurements at Ancona, the measurement station selected is "Ancona Regione", located at 43.610°N, 13.508°E, at an elevation of 91 m a.s.l and with sensor placed at 2 m from the ground [28]. Data

from this weather station start from September 2003 and the frequency of measurements depends on the variable selected: half hourly for temperature (T) and wind (w), and every 15 minutes for precipitation (p). After selecting the sensor and the kind of data, data can be downloaded by selecting the start and end dates, in a maximum range of six months. To avoid manually selecting all dates ranges, an automated script is adopted, using Selenium package for Python.

Venezia Regarding Venezia data, the station name is "Venezia - Istituto Cavanis". The station is located at 45.420°N, 12.328°E, at 18 m a.s.l., and the temperature sensor is at 2 m from the ground. Accessing to the web portal, you can select the stations and have access to all data measured by that station, at hourly time steps. A maximum range of 365 days can be selected and data are displayed in a table, and they can be visualized also in csv or xml format, in the latter are showed also the station's coordinates. Also in this case the process of selecting dates and saving data to a single file has been automatized.

Trieste Data of Trieste has been collected by the OSMER portal, using data of "Trieste molo F.lli Bandiera" station, located at 45.650°N, 13.752°E, at 1 m a.s.l., with temperature sensor at 10 m from ground [29]. Measurements are available for download since 1993, but because other data's start date is more recent, only data since 2003 have been downloaded. OSMER provides hourly data from weather station. Data can be visualized and downloaded only one day at a time, so the process of download several years of data can be long and also in this case has been automatized.

A summary of the weather stations' features is presented in Table 3.1

Station name	Coordinates	Altitude a.s.l.	Variable	Sensor height	Data frequency	Start time	End time
Ancona Regione	(43.610°N, 13.508°E)	91 m	T	2 m	1/30 min	2003-09-17	2024-02-10
			w		1/30 min		
			p		1/15 min		
Venezia - Istituto Cavanis	(45.420°N, 12.328°E)	18 m	T	2 m	1/60 min	2010-01-01	2024-02-10
			w		1/60 min		
			p		1/60 min		
Trieste molo F.lli Bandiera	(45.650°N, 13.752°E)	1 m	T	10 m	1/60 min	2003-01-01	2024-02-10
			w		1/60 min		
			p		1/60 min		

Table 3.1: Summary of weather stations data.

Locations of the three weather stations is shown in Figure 3.1, along with position of Acqua Alta tower.

Figure 3.1: Locations of the weather stations and of the Acqua Alta oceanographic platform.

3.2 Acqua Alta oceanographic tower data

The Acqua Alta oceanographic tower (AAot), operated by Consiglio Nazionale delle Ricerche (CNR), has been in operation in the northern Adriatic Sea since March 1970 and it is the only open-sea platform in the world that allows researchers and technicians to stay on board for extended periods, conducting measurements even in adverse weather conditions. It is located about 8 nautical miles off the coast of Venice, at 45.314°N, 12.508°E, in 16 meters of water. Location of the tower is shown in Figure 3.1, along with locations of the weather stations. The platform consists of a laboratory module and living quarters, supported by a steel structure embedded 22 meters into the seabed. It is equipped with advanced meteorological and oceanographic instruments, with data being both recorded onboard and transmitted in real time to shore. These instruments continuously monitor essential variables such as sea level, sea temperature, wind speed, wave direction, and other oceanographic and atmospheric parameters. AAot platform plays a critical role in operational forecasting for high water (acqua alta) in Venice. Its long, uninterrupted time series of wave measurements, dating back to 1978, is among the longest in the world, providing extremely valuable data for climate change studies. The platform has also been

used to calibrate satellite instruments, such as the ERS-1 altimeter and the SeaWiFS and OLTS optical sensors. Furthermore, it contributes to international research networks like DANUBIUS-ESFRI and JERICO-RI and collaborates with the European Commission's Joint Research Centre. The platform is equipped with sophisticated technologies, including a broadband wireless communication system that allows real-time data transfer and internet access for onboard scientists. High-resolution webcams monitor sea conditions, and underwater cameras provide continuous surveillance of marine life and ocean phenomena. It also supports complex *in-situ* experiments, reducing the risk of instrument loss due to the secure nature of the platform. Wave measurement campaigns with on-board personnel and highly sophisticated instrumentation have clarified unknown aspects of the dynamics of wave generation by the wind. This has led to a marked improvement in operational sea forecasting worldwide. The platform has been also essential in analyzing extreme events such as the Vaia storm of 2018 and the exceptional high water event of November 12, 2019 [30] [31].

Provided dataset include different variables like sea temperature, density, salinity, at 3 m, 6 m, 14 m, 17 m depths and with various sampling frequencies and time ranges:

- a long series of sea temperature (T) measurements at 3 m depth, ranging from the end of 2002 to beginning of 2024, with sampling frequency of 5 minutes;
- records of temperature (T), conductivity (C), salinity (S), density anomaly (σ), optical dissolved oxygen (DO), percentage of oxygen saturation (% DO), chlorofylle (Chl) and turbidity (NTU) at 3 m, 6 m, 14 m depths, with sampling frequency of 10 minutes, from 2020 to 2024.
- long series if records of temperature at sea bottom (17 m) from 2008 to 2019, recorded with various frequency
- measurements from 2019 to 2022 of temperature (T) at sea bottom, pressure (P), sound speed (c), heading (H) and pitch (θ), from 2019 to 2022, recorded every 30 minutes

A summary of various measurement available from AAot is reported in Table 3.2.

Variables	Sensor depth	Data frequency	Start time	End time
T	3 m	$1/5$ min	2002-12-04	2024-01-25
T, C, S, DO, % DO, Chl, NTU	3, 6, 14 m	$1/10$ min	2020-08-07	2024-02-09
T	17 m	$1/30$ min	2019-05-07	2022-06-29
T, c, H, θ	17 m	various	2008-06-30	2019-04-11

Table 3.2: Summary of Acqua Alta platform data.

3.3 CMEMS data

The Copernicus Marine Environment Monitoring Service (CMEMS) is a key component of the European Union's Earth Observation Program, Copernicus, aimed at providing regular and systematic information about the state of the world's oceans and seas. Its primary goal is to provide comprehensive, real-time, and historical ocean data, including parameters like sea temperature, salinity, currents, and chlorophyll concentration, at both global and European regional levels. CMEMS integrates data from satellite observations, *in-situ* measurements, and numerical models, offering a wide range of products and forecasts that are indispensable for understanding ocean dynamics, marine resources, and environmental changes. Its data products are freely available at CMEMS website and they can be downloaded after registration.

3.3.1 Reanalysis data

The Mediterranean Sea Physical Reanalysis Product (MEDSEA_REANALYSIS_PHYS_006_004) [32], part of the CMEMS catalog, offers a comprehensive reanalysis dataset for the Mediterranean Sea, spanning the period from 1987 to July 2022. Even though data of reanalysis goes back to 1987, only temperature data since 2010 have been downloaded, beacause of the availability of Venezia weather station that start from 2010. A summary of temporal coverage of the various temperature datasets is reported in Figure 3.2.

A detailed description of the product is given in [33]. The reanalysis system is composed of three main components: an ocean model, that solves the hydrodynamic equations, a data assimilation scheme, to incorporate observations into the model, and the assimilated data.

Model The ocean model is the Nucleus for European Modelling of the Ocean (NEMO) v3.6 model, which has been implemented in the Mediterranean at a spatial resolution of $1/24°$ (approximately 4.5 km) with 141 unevenly spaced vertical levels, allowing for detailed simulations of the physical dynamics of the

Mediterranean basin. The model also extends into the Atlantic in order to better resolve the exchanges with the Atlantic Ocean at the Strait of Gibraltar.

Data Assimilation The data assimilation scheme is the OceanVar scheme [34]. This scheme uses the variational formulation of data assimilation where a cost function has to be minimized in order to find the best estimation (analysis) of the state variable $\mathbf{x}$. The cost function quantify the distance between the analysis and both the physical model output (background) and the observations:

$$J(\mathbf{x}) = \frac{1}{2}(\mathbf{x} - \mathbf{x}_b)^T \mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}_b) + \frac{1}{2}(\mathcal{H}(\mathbf{x}) - \mathbf{y})^T \mathbf{R}^{-1}(\mathcal{H}(\mathbf{x}) - \mathbf{y}) \tag{3.1}$$

where $\mathbf{x}$ is the analysis state vector, $\mathbf{x}_b$ is the background state vector, $\mathbf{y}$ is the vector of the observations, $\mathbf{B}$ is the background error covariance matrix, $\mathbf{R}$ is the observational error covariance matrix and H is the non-linear observational operator that maps the analysis state into the observational space. The first part represents the distance between analysis and background, while the second part measures the difference between analysis and observations. Eq. 3.1 is linearized around the background state, obtaining:

$$J(\delta\mathbf{x}) = \delta\mathbf{x}^T \mathbf{B}^{-1}\delta\mathbf{x} + \frac{1}{2}(\mathbf{H}\delta\mathbf{x} - \mathbf{d})^T \mathbf{R}^{-1}(\mathbf{H}\delta\mathbf{x} - \mathbf{d}) \tag{3.2}$$

Where $\mathbf{d} = \mathbf{y} - \mathcal{H}(\mathbf{x}_b)$ is the innovation vector, $\mathbf{H}$ is the linearized observational operator, evaluated at $\mathbf{x}_b$ and $\delta\mathbf{x} = \mathbf{x} - \mathbf{x}_b$ are the increments. In this formulation the variables to be corrected are temperature (T), salinity (S) and sea surface height (η), so the state vector is $\mathbf{x} = (T, S, \eta)$. In practice the fully non-linear operator $\mathcal{H}$ is used only once for computing the innovation vectors employing the background fields closer to the observation time, using the so called First Guess at Appropriate Time variation of 3DVar (FGAT-3DVar).

Assimilated data The system assimilates along track sea level anomalies (SLA) from satellite altimetry (SEALEVEL_EUR_PHY_L3_REP_OBSERVATIONS_008_-061 from CMEMS), *in-situ* temperature and salinity profiles, given by a combination of INSITU_GLO_NRT_OBSERVATIONS_013_030 from CMEMS and *in-situ* dataset from SeaDataNet.

Domain of interest The domain of interest is the Northern Adriatic, where the three selected coastal weather stations are located. So the domain of the reanalysis data downloaded extends from (12E-17E, 42.52N-45.98N) for a total of 84×121 grid points. The vertical levels considered are 5, from 1 m to 10.5 m.

3.3.2 Satellite data

For the comparison with satellite temperatures the dataset used is the Reprocessed Mediterranean L4 dataset, SST_MED_SST_L4_REP_OBSERVATIONS_010_021 [35]. It provides a stable and consistent long-term SST time series over the Mediterranean Sea and the adjacent North Atlantic box. This product consists of daily (nighttime), optimally interpolated (L4), satellite-based estimates of the foundation SST (namely, the temperature free, or nearly-free, of any diurnal cycle) at 0.05° resolution grid covering the period from January 1st 1982 to present. In order to make the comparison with reconstructed temperature by neural network models, in the test sets, only data in the period from 2020 to July 2022 have been downloaded.

Data processing The L4 SST product is obtained by processessing multiple satellite data sources through various quality control and merging steps before producing the final interpolated product. Input data are Level-3 Collated (L3C) nighttime SST data from multiple sources:

- ESA CCI SST data (mid-1981 to 2016)
- C3S data (2017 to present)
- PFV53 data (late 1981 to 2014)

L3C data are selected based on the quality flags and confidence values associated with each pixel. Additional cloud contamination masking is made by flagging the pixels that result to be colder (by a fixed threshold) than the previous day value. The L3C data are used to build a single multi-sensor merged (L3S) image per day by selecting only the highest quality measure available for each grid point. The quality of the measurements is established by sensor validation statistics that identify a hierarchy of sensors, performed by Met Office. Large-scale biases are estimated and removed through comparison with already merged data as reference. The final processing stage employs a space time optimal interpolation (OI) scheme using daily decadal climatology as first guess. Input data are selected only within a limited sub-domain: temporal window of ten days and a spatial influential radius ranging between 300 and 900 km. The validation is based on a co-location procedure that operates on a daily basis, identifying the nearest neighbor grid points in both L3S and L4 products to each *in-situ* measurement. The temporal matching window spans from 21:00 local time of the previous day to 06:00 local time of the validation day, ensuring appropriate temporal correspondence. This comprehensive validation covers a substantial period from January 2005 to December 2021, providing robust long-term performance assessment.

Domain of interest Since reanalysis dataset and satellite dataset are not defined on the same grid points, the domain considered for this latter is from (11.96E-17.06E, 42.41N-46.03N)

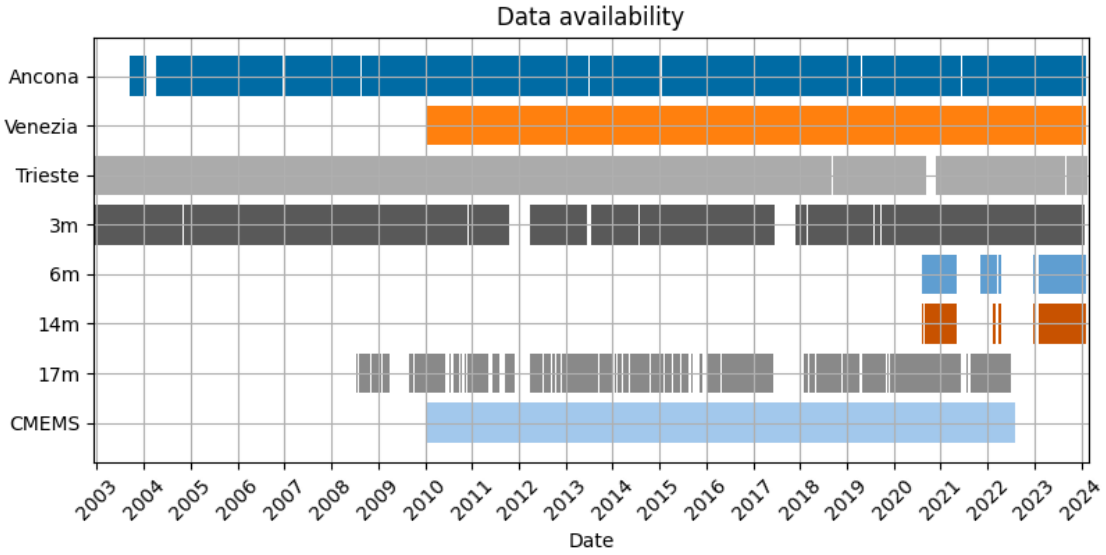


Figure 3.2: Temporal coverage of temperature datasets from different sources: weather stations (first three rows), Acqua Alta platform measurements at various depths (middle rows), and CMEMS reanalysis data (bottom row).

Chapter 4

Methods

This study is divided in three parts, that differ in data source, spatial extension and data frequency:

- **Single point**: reconstruction of hourly sea temperature measured by Acqua Alta oceanographic tower.
- **Spatial average**: reconstruction of daily spatially averaged sea temperature obtained from reanalysis of Copernicus Marine Service (CMEMS).
- **Gridded temperature**: Reconstruction of monthly gridded temperature obtained from CMEMS reanalysis.

For each of these parts, neural network models have been implemented and trained.

4.1 Deep learning framework

The neural network models were implemented using TensorFlow 2.17.1 and Keras 3.5.0 on Google Colab's cloud computing platform. TensorFlow provides high-level APIs for building and training neural networks, while Keras serves as its primary interface for deep learning implementations.

4.1.1 Neural Network Layers

The model architectures incorporate several types of neural network layers:

- Dense layers: Fully connected layers implementing the operation:

$$\mathbf{y} = \phi(\mathbf{W}\mathbf{x} + \mathbf{b}) \tag{4.1}$$

where **x** is the input vector, **y** is the output vector, **W** is the weights matrix, **b** is the bias vector, and ϕ is the activation function. In the Keras implementation the main parameters are:

  - number of units: represents the dimension of the output vector y
  - activation function

- Dropout layers: Regularization layers that randomly deactivate a fraction of neurons during training according to:

$$\mathbf{y} = \mathbf{m} \odot \mathbf{x} \tag{4.2}$$

  where **m** is a binary mask and $\odot$ represents element-wise multiplication. This layer serves to discourage overfitting during training. This operation happen at train time, while at test time, no units are dropped out; instead, the layer's output values are scaled down by a factor equal to the dropout rate, to balance for the fact that more units are active than at training time. In the Keras implementation the main parameter is the rate, that indicates the fraction of units to deactivate randomly.

- Conv2D layers: Two-dimensional convolutional layers performing:

$$y_{ij} = \sum_{m=1}^{M} \sum_{n=1}^{N} w_{mn} x_{i+m,j+n} + b \tag{4.3}$$

  where w_{mn} are the learnable filter weights and b is a bias term. An activation function is then applied to the result. The main parameters of the Keras implementation are:

  - filters: the dimension of the output space, i.e. the number of the filters applied;
  - kernel_size: the size of the filters
  - stride: the step of the application of the filter
  - padding: whether or not pad the input shape with zeros

- MaxPooling2D layers: Downsampling operations that reduce spatial dimensions:

$$y_{ij} = \max_{(m,n) \in R_{ij}} x_{mn} \tag{4.4}$$

  where R_{ij} is a rectangular region centered at position (i, j). The main parameter is the pool_size, determining the size of R_{ij}.

- Conv2DTranspose layers: Transpose convolution layers for upsampling, effectively performing the reverse operation of Conv2D layers.

4.1.2 Optimization

To train artificial neural networks, i.e., to find an optimal set of parameters θ that minimize the loss function $\mathcal{L}$, a fundamental technique called gradient descent is employed. The basic idea of gradient descent is to iteratively update the parameters in the direction of steepest descent of the loss function. Mathematically, this can be expressed as:

$$\theta^{(\tau+1)} = \theta^{(\tau)} - \eta \nabla_\theta \mathcal{L}(\theta^\tau) \tag{4.5}$$

Where $\theta^{(\tau+1)}$ is the updated parameter vector, $\theta^{(\tau)}$ is the current parameter vector, η is the learning rate, a hyperparameter that controls the step size, $\nabla_\theta \mathcal{L}$ is the gradient of the loss function with respect to the parameters. The gradient $\nabla_\theta \mathcal{L}$ provides the direction of steepest ascent, so we move in the opposite direction to minimize the loss.

Evaluate the gradient can become extremely inefficient if there are many data points in the training set because each error function or gradient evaluation requires the entire data set to be processed. So the improve efficiency, the loss function and the gradient are evaluated in small subset of data, called mini-batches. A complete pass through the whole training set is known as a training epoch.

The models were trained using the Adam (Adaptive Moment Estimation) optimizer, which combines two approaches for gradient descent optimization: momentum and RMSprop (Root Mean Square propagation). Momentum approach consists in adding a term to 4.5, made by the parameter update at the previous step, $-\eta \nabla_\theta \mathcal{L}(\theta^{(\tau-1)})$, multiplied by a scaling factor μ, $0 \leq \mu \leq 1$. This term tend to accumulate, resulting in an effective increase of the learning rate from η to $\frac{\eta}{1-\mu}$, that help to speed up the convergence toward the minimum of the loss function. In zones of high curvature of the loss function, as near a minimum, the gradient sign tends to oscillate, so successive iterations lead the momentum terms to cancel out and tending to the original learning rate. RMSprop adapts the learning rate for each parameter by dividing the current gradient by the square root of an exponentially decaying average of squared gradients. This scaling helps handle parameters that have different scales of gradients, preventing too large updates for frequently occurring features and too small updates for infrequent ones.

The loss function employed is the Mean Squared Error (MSE):

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^{N} (\hat{y}_i - y_i)^2 \tag{4.6}$$

where $\hat{y}$ and y are the output of the network and the true value, respectively.

To prevent overfitting, an Early Stopping callback monitored the validation loss and halted training when no improvement was observed for a specified number of epochs, called patience.

4.2 Preprocessing

4.2.1 Climatologies and Anomalies

In order to perform the analysis and the computations described in the next sections, the periodic part of the signals has been taken out. This has the aim of focusing on the underlying patterns by removing predictable seasonal variations. Let's define the key terminology used throughout the text:

- Climatology represents the expected or typical state of a variable for a given time of year, calculated from historical data.
- Anomaly is the deviation from this expected state (actual value minus climatology).

Two kind of climatologies have been used to perform this analysis: hourly and daily climatology. Hourly climatology is is the average temperature for each hour of the year, averaged over all available years. Let's consider hourly data and call $T(y, m, d, h)$ the original time series of temperature and $\overline{T}(m, d, h)$ the hourly climatology, where y, m, d, h indicates respectively year, month, day and hour. The hourly climatology doesn't depend anymore from the year. In formulae:

$$\overline{T}(m, d, h) = \frac{1}{N_y} \sum_{y=1}^{N_y} T(y, m, d, h) \tag{4.7}$$

where N_y is the number of years considered. Daily climatology is the average temperature for each calendar day of the year, averaged over all available years, represented by $\overline{T}(m, d)$ and depending only on the month and day. In formula:

$$\overline{T}(m, d) = \frac{1}{N_y \cdot 24} \sum_{y=1}^{N_y} \sum_{i=0}^{23} T(y, m, d, h) \tag{4.8}$$

where 24 is the number of hours in a day. For daily data it is possible to define only the daily climatology as:

$$\overline{T}(m, d) = \frac{1}{N_y} \sum_{y=1}^{N_y} T(y, m, d). \tag{4.9}$$

The anomaly is computed by removing the climatology from the original value:

$$T' = T - \overline{T} \tag{4.10}$$

4.2.2 Acqua Alta data cleaning

By examining the plot of data collected from AAot (first panel of Figure 4.1), several inconsistent values are evident. A systematic data cleaning procedure was implemented through the following steps:

- Physical thresholding: data points were constrained to the physically reasonable temperature range of 0 °C to 35 °C. The results of this initial thresholding are displayed in the second panel of Figure 4.1.
- Statistical outlier removal: an iterative process was applied until convergence (no further outliers detected):
  - Construction of hourly climatology
  - Calculation of hourly anomalies
  - Computation of standardized anomalies (z-scores) using:

    $$z = \frac{x - \mu}{\sigma}$$

    where μ represents the mean and σ the standard deviation
  - Removal of extreme values exceeding three standard deviations:

    $$|z| \leq 3$$

The results after the complete procedure are shown in Figure 4.1.

Data aligning

The temperature datasets of AAot, Ancona, Venezia, and Trieste were collected at varying time intervals: 5 minutes, 30 minutes, 1 hour and 1 hour, respectively. To ensure consistency in the analysis, it was necessary to align these datasets onto a common temporal grid. Given that the lowest frequency data was collected hourly, an hourly time step has been used for all datasets. This means that data collected at higher frequencies (5-minute and 30-minute) were regridded to hourly values. This it's been made by using an hourly time grid and select the nearest observation for each time step, in a range of 1 hour. The same procedure is applied to Ancona wind data, originally sampled at 30 minutes. For the Ancona precipitation data, sampled at 15 minutes intervals, aggregation is made by summing the precipitation measured in the previous hour. Missing data are replaced with nans.

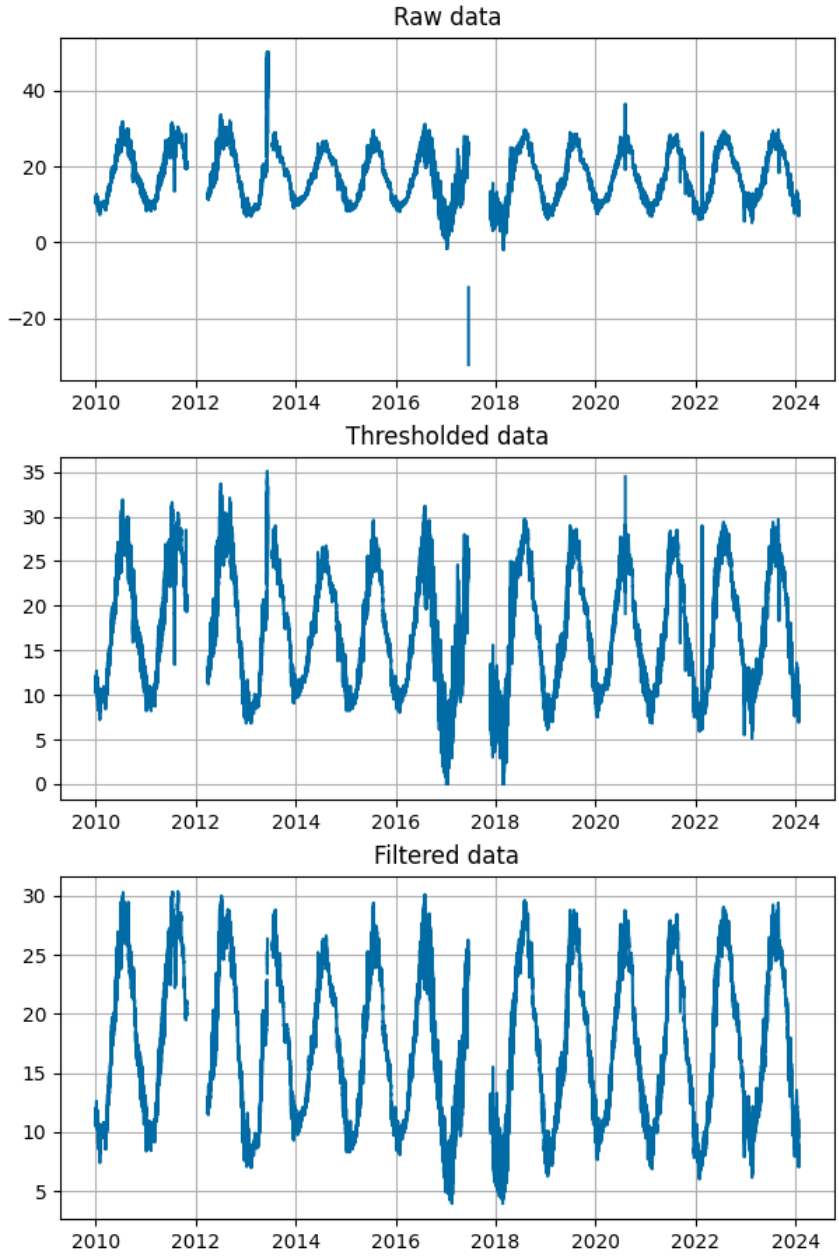


Figure 4.1: Acqua Alta temperature time series at different cleaning level. Top panel: raw data. Middle panel: data thresholded in the the physical reasonable range. Bottom panel: data after the cleaning procedure.

4.3 Acqua Alta Temperature reconstruction

In this section a detailed description of the methods adopted to reconstruct Acqua Alta temperature is provided.

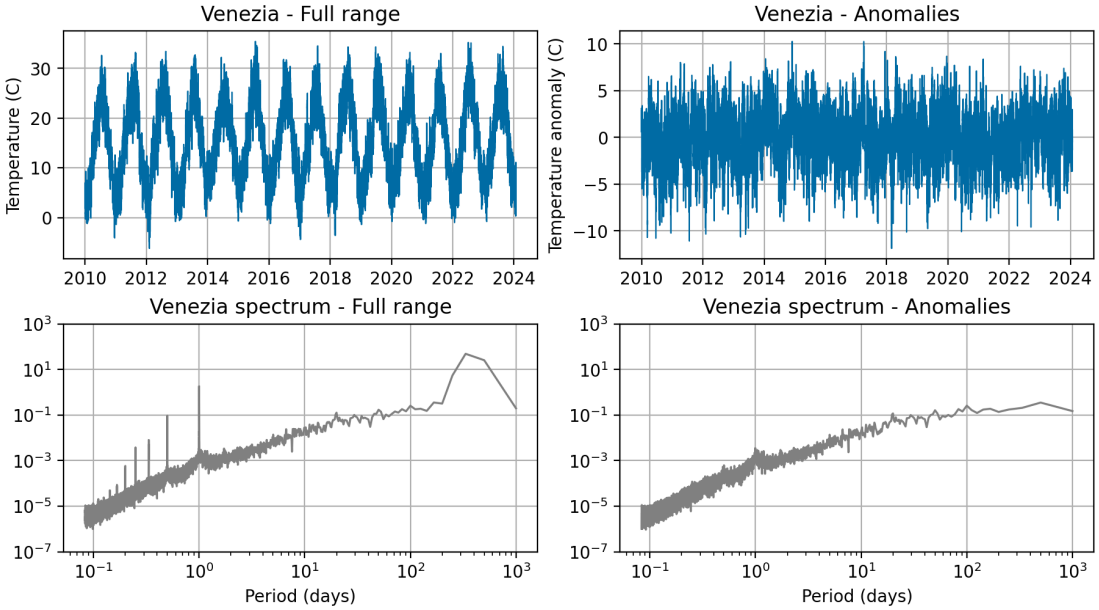
Anomalies As mentioned before, the climatological part of the signal it has been taken out, to focus on the anomalies. The results of the removal of the hourly climatology is shown in Figure 4.2, for the air temperatures measured at Venezia weather station and for the sea temperature measured by AAot. The figure shows the original temperature observations, the power spectrum of the original signal, the temperature anomalies (deviations from hourly climatology) and the power spectrum of the anomalies. The power spectrum of the original signal shows peaks at seasonal frequency (1 year period), at 1 day period and at higher frequency, in both the air and sea temperatures. By computing the anomalies great part of this signals should be taken out. It can be seen that in the power spectrum of Venezia temperature anomalies, the main peaks are removed, even though still remain some periodical signal, mainly at daily frequency. In the sea temperature anomaly power spectrum, instead the peaks are attenuated but are still well visible. The presence of these higher frequency signals led to try using also filtered data and daily mean data.

Signal Processing Approaches Given the persistent high-frequency components in the anomaly signals, three different temporal processing approaches were investigated:

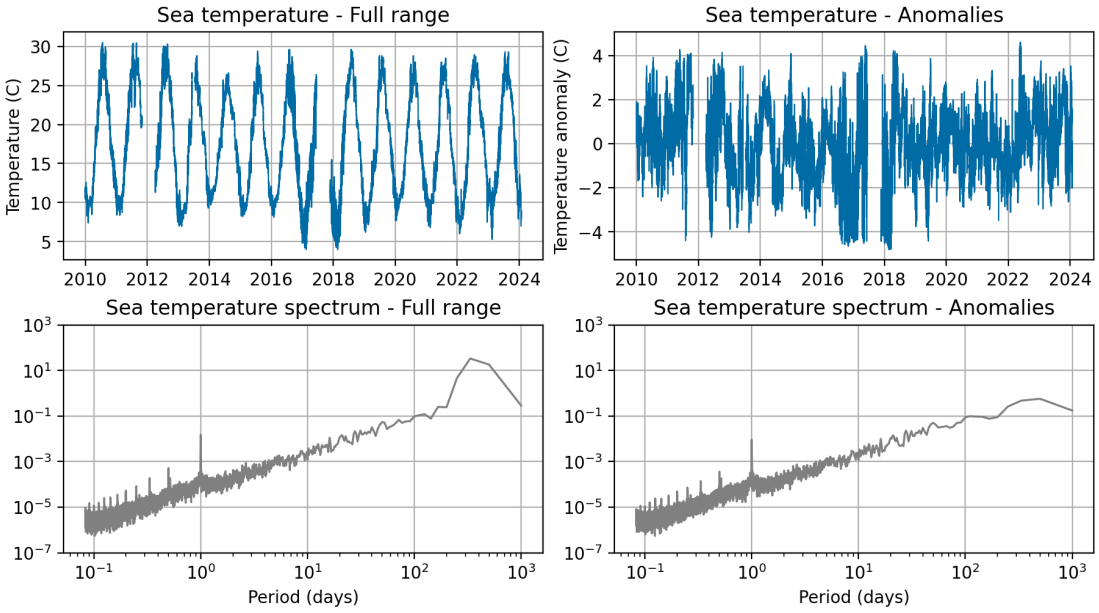
- Raw hourly data
- Filtered hourly data using a fourth-order Butterworth low-pass filter (implemented via SciPy's signal.filtfilt method that apply the filter by processing the input data in both the forward and reverse directions to avoid phase distortion)
- Daily averaged data. Daily mean air temperatures has been computed excluding days with more than 4 hours of missing data, to reduce the probability of having temperatures that are not representative of the entire day.

The Butterworth filter, given its maximally flat frequency response in the passband, it's suitable for preserving the lower-frequency temperature variations of interest while removing higher-frequency fluctuations. The cutoff frequency used is $f_c = \frac{1}{26h}$, to reduce all subdaily frequencies.

For each of the different temporal processing four different experiments have been made, varying the model and the input features:



(a) Venezia time series and spectrum.



(b) Acqua Alta time series and spectrum.

Figure 4.2: Observed temperature and anomaly time series of (a) Venezia and (b) Acqua alta with their respective power spectra. Each subfigure is divided in four panels. Top left: Original temperature observations. Bottom left: Power spectrum of the original signal. Top right: Temperature anomalies (deviations from hourly climatology). Bottom right: Power spectrum of the anomalies.

Models The architectures employed are three:

- Linear model: A single Dense layer with one unit, and no activation function, thus performing a linear regression.
- Dense model: a succession of Dense and Dropout layers, with the following parameters:
  - Dense, 64 units, ReLU activation
  - Dropout, rate=0.2
  - Dense, 256 units, ReLU activation
  - Dropout, rate=0.2
  - Dense, 64 units, ReLU activation
  - Dropout, rate=0.2
  - Dense, 1 unit, no activation
- Linear+Dense, architecture combining linear and Dense pathways. The final output is obtained by summing the outputs of the two paths.

These models are trained using Adam optimizer with learning rate of 10^{-4} and MSE loss. Early stopping callback has been implemented with a patience of 10 epochs. The batch size used is 32.

Input features The main input features utilized are:

- air temperature anomaly for Ancona, Venezia, Trieste weather stations;
- discounted sum of the anomalies, defined as

$$T'_n + \sum_{i=n-1}^{0} \alpha^{n-i} T'_i \tag{4.11}$$

  where T_n is the temperature at time t_n and the sum is performed over all past temperature from $i = n - 1$ to $i = 0$, scaled by the exponentially decaying, with past observation, factor α^{n-i}. Missing values in the time series are replaced with 0 for the calculation. This features serves in order to take into account also past observations of air temperature that can have influence to the sea temperature at the time step of reconstruction. Indeed this value accumulate the anomalies in the past observations and grow if they are of the same sign or tends to cancel out if the sign is varying. Two different values are used for hourly and daily data:

  - $\alpha = 0.99$ for hourly data
  - $\alpha = 0.9$ for daily data.

  The smaller value of α for daily values is chosen to account less for observations of previous days, compared to the previous hours.

- Time features encoded in sine and cosine. For hourly data it has been used the hour of the year for representing time:

$$h_{\sin} = \sin\left(\frac{2\pi \cdot h_y}{24 \cdot 365.25}\right) \tag{4.12}$$

$$h_{\cos} = \cos\left(\frac{2\pi \cdot h_y}{24 \cdot 365.25}\right). \tag{4.13}$$

  with h_y representing hour of the year and 365.25 is the average number of days in a year. For daily data, it has been used the calendar day of the year, d_y:

$$h_{\sin} = \sin\left(\frac{2\pi \cdot d_y}{365.25}\right) \tag{4.14}$$

$$h_{\cos} = \cos\left(\frac{2\pi \cdot d_y}{365.25}\right). \tag{4.15}$$

To explore the potential influence of additional weather variables, an experiment has been conducted adding also wind and precipitation data, using the Dense model. In particular the additional features are:

- wind speed at the three weather stations, measured in $\mathrm{m\,s^{-1}}$;
- wind direction at the three weather stations, give by the two components of the versor, V_x and V_y;
- cumulated precipitation, measured in mm;

Each feature has been normalized by removing its mean and dividing by the standard deviation.

The number of experiments for each of temporal processing is four:

- Linear model - temperature data
- Dense model - temperature data
- Linear + Dense model - temperature data
- Dense model - temperature, wind and precipitation data

Target The target for all experiments is the sea temperature anomaly measured by AAot, with temporal processing according to the input features.

Evaluation The model evaluation strategy consisted of two phases: a cross-validation procedure to assess model performance across various years and a final evaluation on an independent test set.

The consistent period where all input variables and target data are available spans from January 2010 to January 2024.

The cross-validation was performed using a leave-one-year-out approach on the 2010-2021 period. For each iteration:

- One year was held out as the validation set
- The remaining 11 years were used as the training set
- The model was trained on the 11-year training period
- Performance was evaluated on the held-out validation year using RMSE

This process was repeated 12 times, with each year from 2010 to 2021 serving once as the validation set, ensuring independence between training and validation data and assessment of the model's ability to generalize across different annual conditions.

For the final model evaluation, the dataset was partitioned as follows:

- Training period: 2010-2019 (10 years)
- Validation period: 2020-2021 (2 years)
- Test period: 2022-January 2024 (approximately 2 years)

The final model was trained on the 2010-2019 period, with 2020-2021 used for validation and early stopping to prevent overfitting. Early stopping monitored the model's MSE on the validation set during training and stopped when no improvement was observed for 10 epochs. The model weights that achieved the best performance on the validation set were retained.

The test set (2022-January 2024) remained completely independent from both the cross-validation procedure and the final model training, serving as an unbiased evaluation of model performance on unseen data.

Coherence analysis To investigate the reasons behind the model performance limitations, coherence analysis has been performed between the input and target time series. This analysis helps understand the frequency-dependent relationships. The coherence analysis consists of three main components: coherence, phase, and gain, derived from the cross-spectral analysis of two time series. Given two time series $x(t)$ and $y(t)$, their power spectra $P_{xx}(f)$ and $P_{yy}(f)$, and their cross-spectrum $P_{xy}(f)$, the coherence $C_{xy}(f)$ is defined as:

$$C_{xy}(f) = \frac{|P_{xy}(f)|^2}{P_{xx}(f)P_{yy}(f)} \tag{4.16}$$

where f is the frequency. The coherence ranges from 0 to 1, with 1 indicating perfect linear correlation at that frequency and 0 indicating no correlation.

The phase spectrum $\phi(f)$ represents the phase difference between the two signals at each frequency:

$$\phi(f) = \arg(P_{xy}(f)) \tag{4.17}$$

where arg denotes the principal argument of the complex number. The phase can be expressed in time, by dividing the phase in radians by $2\pi f$, $\phi_t = \frac{\phi_{\text{rad}}}{2\pi f}$.

The gain spectrum $G(f)$ represents the amplitude ratio between the output and input signals:

$$G(f) = \frac{|P_{xy}(f)|}{P_{xx}(f)} \tag{4.18}$$

The gain provides information about how the amplitude of oscillations at different frequencies in one time series relates to those in the other time series.

These spectral estimates were computed using SciPy's signal.csd method for computing the cross spectral density P_{xy} and signal.welch for computing the power spectral densities. Both method use the Welch's method and the Hann window.

4.4 CMEMS averaged temperature reconstruction

This section presents the methodology adopted for the reconstruction of temperature derived from CMEMS reanalysis data over the domain (12E-17E, 42.52N-45.98N) and spatially averaged.

The target for the reconstruction are computed considering various vertical levels, horizontal domains, and time lags. Since the CMEMS reanalysis dataset provide daily data, as input for the models daily averaged air temperature data have been used. Daily climatologies and anomalies has been calculated for both air temperature and sea temperature at each grid point.

The targets consist on spatial averages on the vertical levels and on the horizontal domain; the dimensions considered are:

- Vertical levels
  - Surface level (v0): at 1 m depth
  - First three levels average (v02): from 1 m to 5.5 m
  - First five levels average (v04): from 1 m to 10.5 m
- Horizontal domains
  - Complete domain: entire study area
  - Threshold-based domain: regions where mean correlation coefficient exceeds 0.5

Also the following time lags between air and sea temperature are considered

- No lag (l0): same-day for air and sea temperatures
- One-day lag (l1): sea temperature response after 1 day
- Two-day lag (l2): sea temperature response after 2 days

Vertical averaging For the vertical levels average (v02 and v04), vertical means were computed using the trapezoidal integration method. Let $(h_0, h_1, \ldots, h_N)$ be the unevenly spaced vertical levels where sea temperature is defined, and $(T'_0, T'_1, \ldots, T'_N)$ the corresponding temperature anomalies. The vertical integral is computed as:

$$\int_{h_0}^{h_N} T'(z)dz \approx \sum_{i=0}^{N-1} \frac{1}{2}(h_{i+1} - h_i)(T'_i + T'_{i+1}) \tag{4.19}$$

And the vertical average as:

$$\langle T \rangle_0^N = \frac{1}{h_N - h_0} \sum_{i=0}^{N-1} \frac{1}{2}(h_{i+1} - h_i)(T'_i + T'_{i+1}) \tag{4.20}$$

Horizontal domain selection Two approaches were used for spatial averaging:

- Complete domain averaging: including all grid points in the study area
- Threshold-based averaging: including only grid points where the mean correlation coefficient (r) exceeded 0.5, indicating strong air-sea temperature coupling. Those areas are shown in Figure 5.7.

For each combination of vertical level (v0, v02, v04) and time lag (l0, l1, l2), correlation coefficients have been computed as follows:

- For each grid point (x, y) in the CMEMS domain, Pearson correlation coefficient has been computed between:
  - The air temperature anomaly time series from each weather station: $T'_{air,s}(t)$, where s indicates the station (Ancona, Venezia, Trieste)
  - The sea temperature anomaly time series at that grid point: $T'_{sea}(x, y, t)$

  This yields three correlation coefficients for each grid point: $r_s(x, y)$ where s represents each station.
- The final correlation coefficient for each grid point is obtained by averaging the correlation coefficients from the three stations:

$$r(x, y) = \frac{1}{3} \sum_{s=1}^{3} r_s(x, y) \tag{4.21}$$

Model The model used is the same Dense model described in the previous section, constituted of three Dense layers with 64, 256, and 64 units respectively, each followed by a Dropout layer with a rate of 0.2, and the final Linear layer with a single unit.

The model is trained using Adam with the same configuration: learning rate of 10^{-4}, MSE loss and batch size of 32. Early stopping is also implemented with a patience of 10.

Input features Input features are:

- daily air temperature anomalies from the three weather stations
- discounted sum of the anomalies as defined before, with $\alpha = 0.9$
- sine and cosine of the day of year

Each feature has been normalized by removing its mean and dividing by the standard deviation.

Target One experiment has been run for each of the combination of vertical domains, time lags and horizontal domain. The total number of targets is 18: 3 vertical domain, 3 time lags and 2 horizontal domains

Evaluation The evaluation procedure followed the same approach, based on cross-validation, previously described, adapted to accommodate the shorter data availability period (2010 to July 2022).

The leave-one-year-out cross-validation is done using data from 2010 to 2019. For the final model assessment, the dataset has been split in:

- Training period (2010-2017): Eight years of data for model parameter optimization
- Validation period (2018-2019): Two years used for early stopping during model training
- Test period (2020-2022): Approximately 2.5 years reserved for final model evaluation

4.5 CMEMS pointwise temperature reconstruction

This section describes the methodology adopted to reconstruct sea temperature at each individual grid point in the domain. This approach aims to capture the full spatial variability of temperature patterns.

The input features remain consistent with our previous analysis, while the output dimension expands to match the spatial resolution of the domain.

Models

Two neural network architectures are tested for this reconstruction task:

- Dense model, with the same structure as before, but with the output shape adjusted: three Dense layers of 64-256-64 units, each followed by a Dropout layer with rate of 0.2. The final layer is a Dense layer with no activation function and $84 \cdot 121$ units, corresponding to the number of grid poibts in the domain.
- Dense-Unet model. This second architecture combines Dense layers with a Unet structure to potentially capture spatial relationships and temperature patterns. The Unet architecture derives its name from its characteristic U-shaped structure, consisting of a down-sampling path (encoder) and an up-sampling path (decoder) connected by skip connections. This

design make it particularly effective for tasks requiring both global context and local detail preservation. The down-sampling path progressively reduces spatial dimensions while increasing the number of feature channels, allowing the model to capture increasingly abstract representations. The up-sampling path essentially mirrors the downsampling path path, and the output from each down-sampling layer is concatenated with the corresponding up-sampling layer, giving those layers access to higher-resolution spatial information. The down-sampling path is composed of down-sampling layers made as follow:

- Conv2D, 3×3, ReLU
- Conv2D, 3×3, ReLU
- MaxPooling2D, 2×2

The upsampling layers are made as

- Conv2DTranspose, 3×3
- Conv2D, 3×3, ReLU
- Conv2D, 3×3, ReLU

The number of filters in the convolutional layers is (64-128-256) for the down-sampling path and the reverse for the upsampling path (256-128-64). The two path are connected by a bottleneck composed of two Conv2D(3×3)layers with 512 filters. The whole Unet architecture can be referred synthetically using the number of his filters as (64-128-256-512). The Unet takes in input the output of the Dense model described before.

Both models have been trained using Adam optimizer with learning rate di 10^{-4} and early stopping callback has been implemented with a patience of 20 epochs. The batch size used is 32. Grid points with nan values in the target, representing land points, have been handled by implementing a custom MSE loss that mask those points, setting to zero the corresponding values.

Input Features

The input feature set remains consistent with the previous spatial averaging approach:

- Daily air temperature anomalies from Ancona, Venezia, and Trieste weather stations
- Discounted historical temperature sums with $\alpha = 0.9$

- Annual cycle encoded through sine and cosine of the day of year

Each feature has been normalized by removing its mean and dividing by the standard deviation.

Training and Evaluation

Given the increased complexity of the grid-point reconstruction task, a chronological split of the dataset is used:

- Training period: 2010-2017
- Validation period: 2018-2019
- Test period: 2020-2022

The performance of both models is evaluated not only through RMSE, but also using derived metrics. Following [36], MSE can be decomposed as:

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(m_i - o_i)^2 = MB^2 + SDE^2 + 2\sigma_m\sigma_o(1 - CC) \tag{4.22}$$

where m_i and o_i are respectively the i^{th} modeled and observed variables, with corresponding standard deviations σ_m and σ_o; MB is the mean bias, SDE is the standard deviation error and CC is the correlation coefficient, defined as:

$$MB = \overline{m} - \overline{o} \tag{4.23}$$

$$SDE = \sigma_m - \sigma_o \tag{4.24}$$

$$CC = \frac{1}{\sigma_o}\frac{1}{\sigma_m}\sum_{i=1}^{N}(m_i - \overline{m})(o_i - \overline{o}) \tag{4.25}$$

where $\overline{m}$ and $\overline{o}$ are the mean of the modeled and observed variables respectively. In addiction also the unbiased RMSE (uRMSE) is evaluated:

$$uRMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left((m_i - \overline{m})(o_i - \overline{o})\right)^2} \tag{4.26}$$

These metrics are used to evaluate the performance of the reconstruction both spatially and temporally. Indeed each sum can be computed on the spatial

dimension or in the time dimension. For spatial analysis, the statistics are computed across all timestamps T for each spatial point (x,y), e.g.:

$$\overline{m}(x,y) = \frac{1}{T}\sum_{t=1}^{T} m(x,y,t) \tag{4.27}$$

Conversely, for temporal analysis, the statistics are computed across all spatial points $N = X \cdot Y$ for each timestamp t, e.g.:

$$\overline{m}(t) = \frac{1}{N}\sum_{x=1}^{X}\sum_{y=1}^{Y} m(x,y,t) \tag{4.28}$$

Comparison with satellite observations

In order to quantify the differences present between reanalysis and satellite measurement and between satellite and neural networks reconstrucion, spatial and temporal analyses were performed using MB, SDE, CC, and uRMSE metrics to evaluate the agreement between the datasets Since satellite data are defined on a different spatial grid than reanalysis and model data, satellite observations were linearly interpolated onto the reanalysis grid to enable comparison between the three temperature fields. To perform a consistent comparison, the anomalies of satellite dataset are computed by removing to the observed values its climatology.

Chapter 5

Results

This chapter presents the findings of the experiments previously described. Each experiment is designed to test the possibility of reconstructing sea temperature from weather stations data, across different spatial and temporal resolution, using neural networks models. For each experiment the performance of the models is evaluated, along with an analysis of the strength and limitations of the approach.

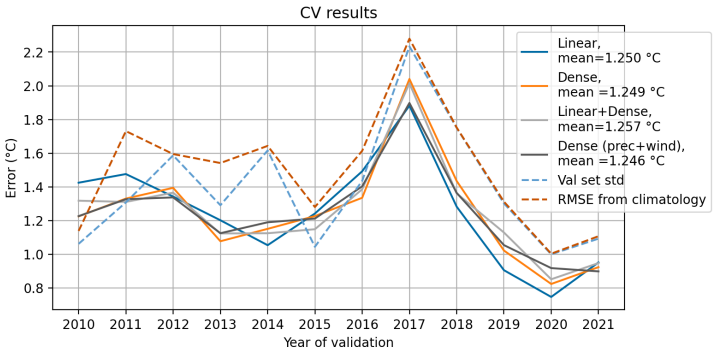
5.1 Acqua Alta temperature reconstruction

This section analyzes the reconstruction of temperature measurements from AAot using three temporal processing approaches: hourly data, filtered hourly data, and daily averages. The analysis aims to identify optimal methods for sea surface temperature reconstruction and to study the sensitivity of the proposed methodology to high frequency energy content in the input signals.

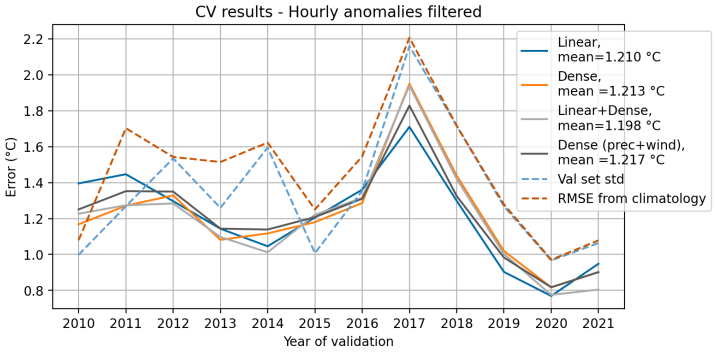
Hourly data

The first approach has been to use hourly data. The performance of the various models has been assessed trough cross validation as described in Section 4.3.

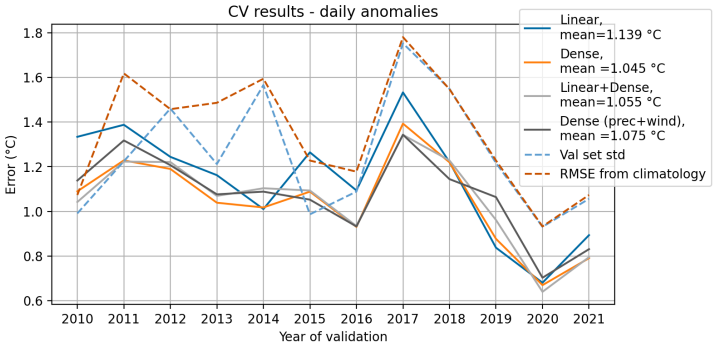
Cross validation scores Figure 5.1a shows the results of this procedure, in the years ranging from 2010 to 2021. On the x-axis there is the year of the validation set, on which the performance of the model have been evaluated, and on the y-axis the score, measured by the RMSE between the model's prediction and the actual values. The solid lines represents the performances of the different models throughout the validation years, while the dashed light blue line represent the standard deviation of the validation set on the individual years and the dashed



(a) Hourly anomalies.



(b) Hourly anomalies filtered.



(c) Daily anomalies.

Figure 5.1: Results of the cross validation in the reconstruction of AAot temperature, for the three different temporal processing approaches. Solid lines represent the RMSE of the models on the validation year. Light blue dashed line is the standard deviation of the anomalies. Orange dashed line is the RMSE of the anomaly from the climatology.

orange the RMSE when using the climatology as prediction. The plot shows that the performances fluctuates from year to year, roughly following the standard deviation of the validation set. There isn't a model that performs consistently better than the others throughout the years, and also the mean RMSE are similar. It's interesting to note that even though the performances are not optimal all the models reconstructions constitutes still an improvement from the climatology, with exception for the year 2010.

Training history Figure 5.2 shows the RMSE history during the training phase. Solid lines represents the RMSE computed on the training set while dashed lines are the RMSE on the validation set; different colors represents the year of the validation set. We can see that the Linear model RMSE on the training set stabilizes quickly, reaching its stable value just at second epoch. The RMSE on the validation set also stabilizes, but with small oscillations, that reset the patience and make the training continue for some more epochs. For the other models instead the RMSE on the training set keep decreasing over the epochs, reaching lower values than the Linear model, while the validation RMSE oscillate and in some cases increase. This indicate that the models learn quickly the relevant patterns and then keep overfitting on the training data. It is possible to note by looking at epoch 13, for example, that this behavior is even more accentuated in the model using precipitation and wind data, where the RMSE during training, decrease more rapidly. Thus the added features make the model capable of capturing more non-linearity in the training data, but that it fails to generalize to unseen data. The use of the early stopping, serves to prevent the model to continue the training while no more improvements are made to the validation set and and avoiding overfitting, saving the weights of the model's best performance. As additional confirmation is also possible to see that, for each model, while training loss converge roughly to the same value for all train-validation split, the validation RMSE vary a lot, as seen also in Figure 5.1a.

Results on test set In Figure 5.3 the output of the models (in orange) are plotted against the true values (in blue), in the test period, from 2022 to 2024, after being trained on the 2010-2019 period and being evaluated on the 2020-2021 years, in order to met the conditions for early stopping. It can be noted that the Linear model is the one that achieve better score in terms of RMSE. By looking at the reconstructed temperatures (orange line) we can explain this better performance, noting that the Linear model try to capture the general trend of the data, without reproducing its high frequency variability, while the others methods capture more variability of the data, but it's not well reproduced. This suggest that the target data may contain some patterns that are not explained

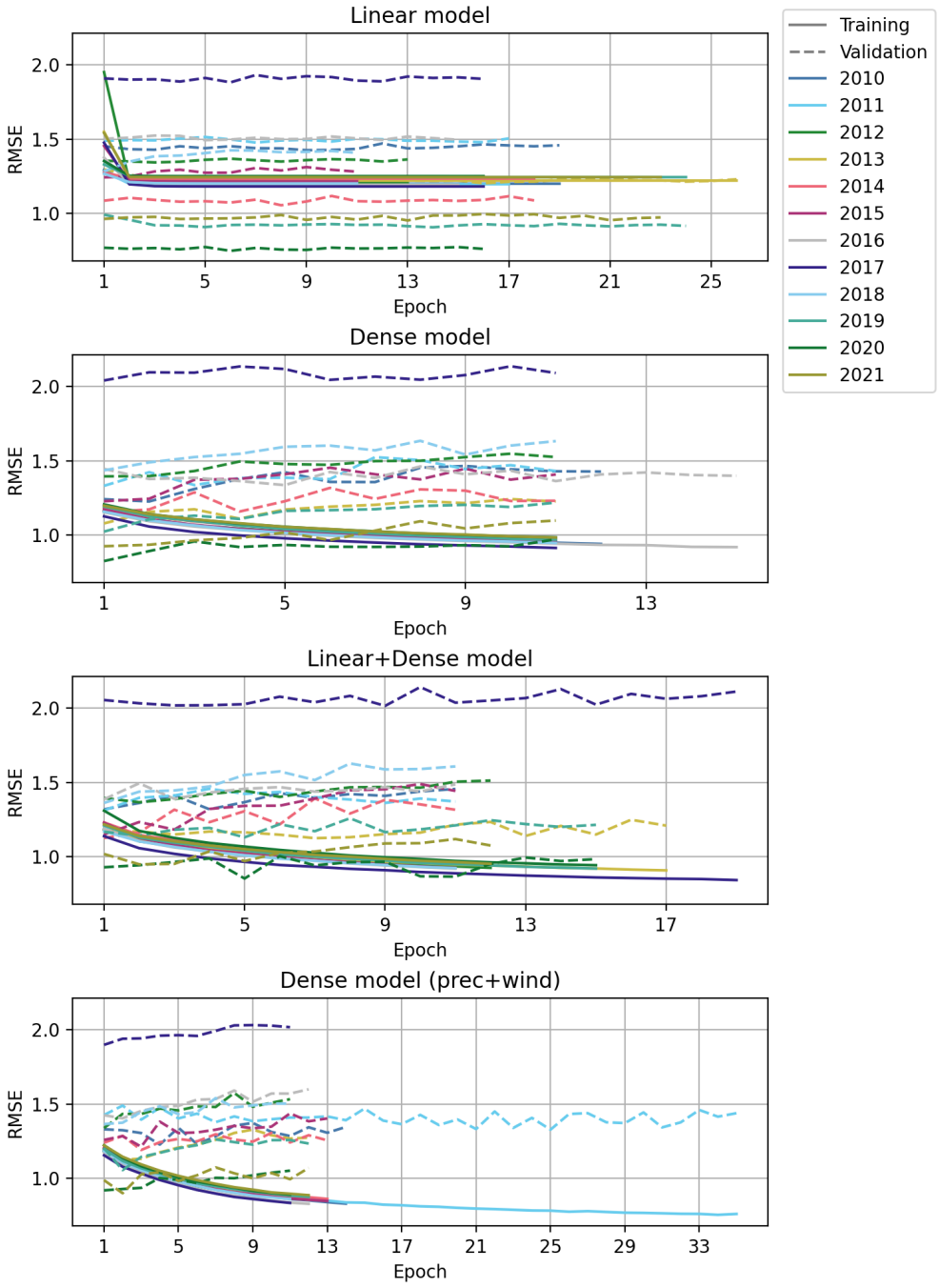


Figure 5.2: RMSE history during epochs. Solid line is the RMSE on the training set. Dashed lines is the RMSE on the validation set. Different colors represents the validation year.

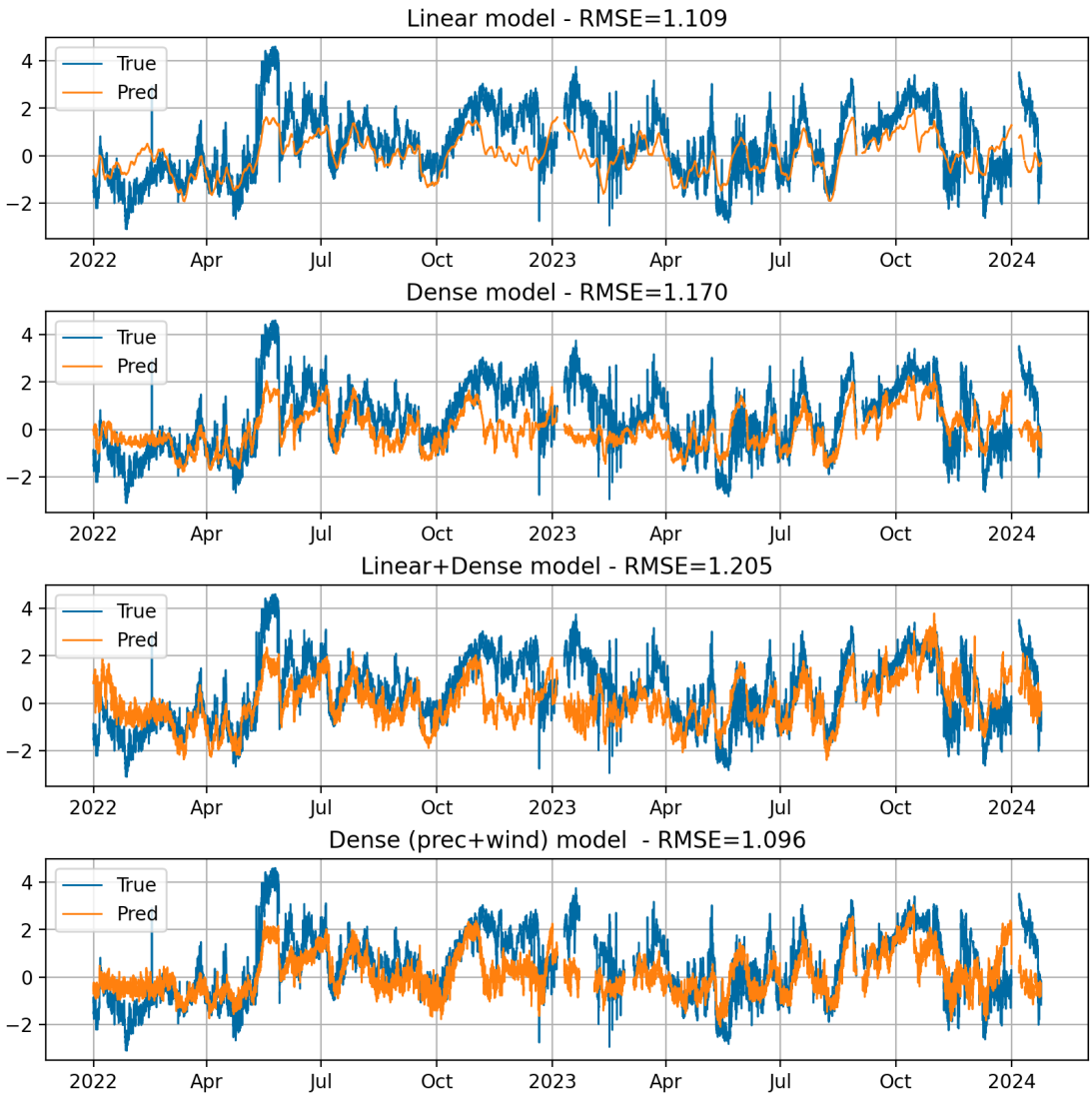


Figure 5.3: Results of the reconstruction on the test set. Top panel: Linear model. Second panel: Dense model. Third panel: Linear+Dense model. Fourth panel: Dense model with precipitation and wind features.

by the input features. The aim of using a combination of the Linear and Dense models is to let the linear path learn quickly the general trend while the dense path learn the residual components; however Figure 5.2 suggest that the model is capable to capture the non linear relationships in the training set, that are not reproduced in the validation set. Figure 5.4 shows a zoom of Figure 5.3 on 13 days-window on the period from 15^{th} to 27^{th} of June 2022, where the true and recostructed temperatures differs significantly, along with a plot of the air temperatures in the same period. It can be observed that in both the sea and air temperatures there is a periodicity of approximately one day. This signals are also evident in Figure 4.2, by looking at the spectral components of the temperature anomaly. The spectrum of the sea temperature anomaly shows also higher-frequency components that persist even after subtracting the climatology, indicating that some signals are not fully removed.

Filtered hourly data

These high-frequency signals could interfere with the models' ability to reproduce the desired output accurately. To adress this a low-pass Butterworth filter with a cutoff frequency $f_c = \frac{1}{26\,\text{h}}$ is applied to both the inputs and outputs. Figure 5.5 shows the results of the filtering in terms of the spectrum of the sea temperature anomaly. The cross validation procedure has been repeated with the filtered signals and the results are shown in Figure 5.1b. The RMSE trend across cross-validation iterations remains similar to that observed with the non-filtered data (Figure 5.1a), but there is a slight improvement in the models' mean performance. The average RMSE decreases from 1.25 °C with non-filtered data to 1.21 °C with filtered data. However, despite this marginal improvement, the results remain unsatisfactory, leading to the decision to try to use daily averages instead.

Daily data

By using daily averages, we effectively ignore all intradaily variability, shifting the focus entirely to longer-term trends and smoother patterns. This approach aims to reduce the influence of short-term fluctuations, which the models may struggle to capture accurately, and instead prioritize capturing the underlying daily patterns that may be more stable and predictable.

As shown in Figure 5.1c, using daily averages yields a slight further improvement in performance, with the average RMSE decreasing to 1.08 °C. This improvement suggests that the model is better able to generalize and capture the broader patterns in sea temperature over daily timescales, rather than attempting to follow rapid, intra-daily changes that may be influenced by factors

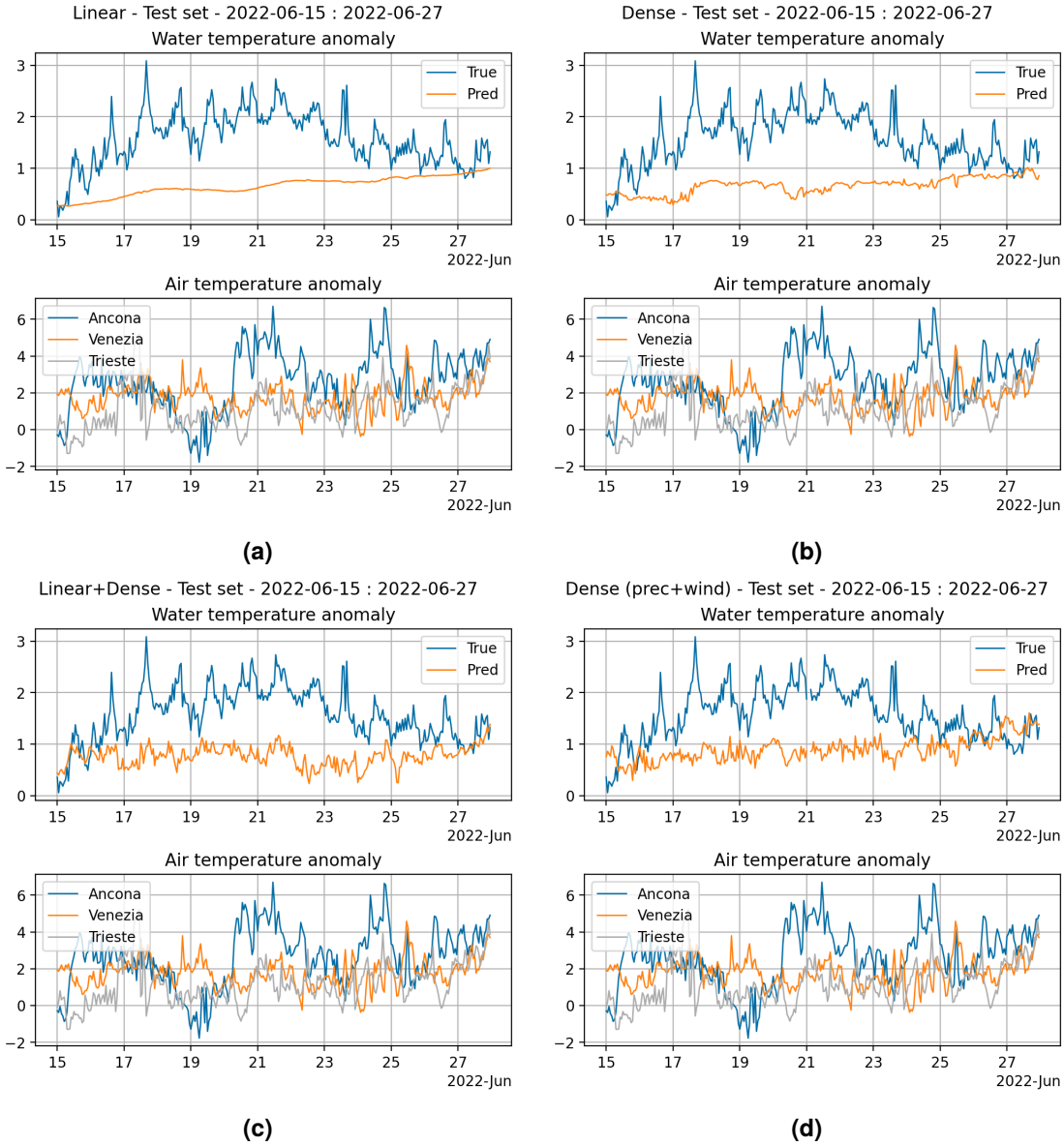


Figure 5.4: Zoom on a 13-days window of the results on the test set.

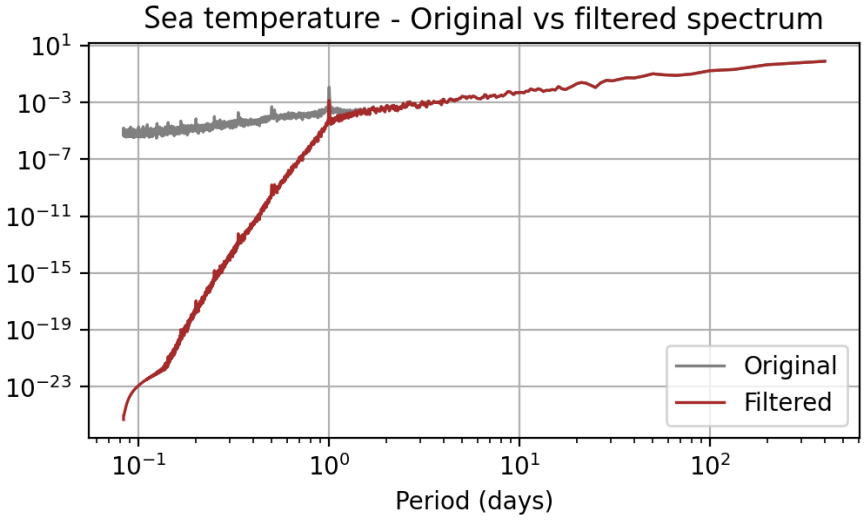


Figure 5.5: Power spectrum of hourly anomaly time series before and after the application of the low-pass filter.

not accounted for in the input features.

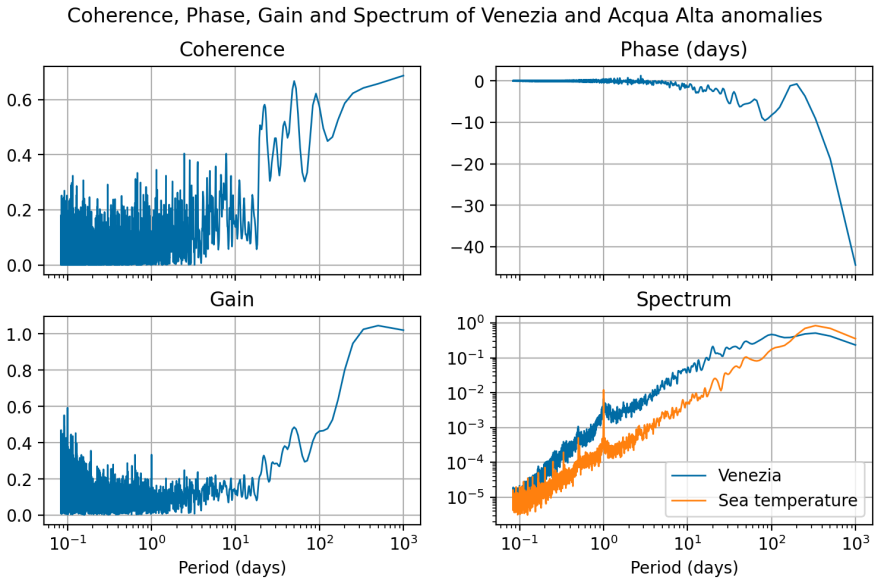
Although there is an improvement, an average RMSE of 1.08 °C indicates that the model is still not accurately capturing all the important dynamics of sea temperature variability. This suggests that there are critical dynamics,particularly at higher frequencies, that the model struggles to account for.

Coherence analysis

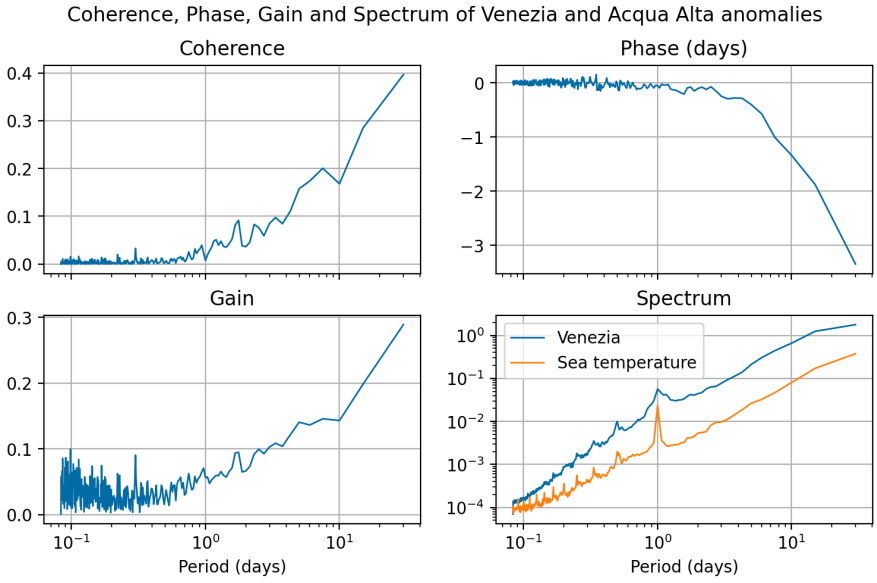
The coherence, phase, gain, and spectrum plots in Figure 5.6 between Venezia air temperature and temperature of the sea at 3m depth measured by AAot, offer additional insights into this performance limitation due the relationship between air and sea temperature variability at different temporal scales. In Figure 5.6a, which spans from high to seasonal frequencies, we observe a gradual increase in coherence as the period lengthens. This trend indicates that the air and sea temperatures are more strongly correlated at lower frequencies, particularly on multi-day or seasonal scales. The gain plot in both figures further supports this observation, showing that the influence of air temperature on sea temperature becomes more pronounced over longer timescales. The phase plot shows that the sea temperature signal is always delayed with respect to the air temperature, at long periods. To further understand the coupling between the two temperatures in Figure 5.6b the parameters of the coherence analysis have been optimized to analyze higher frequency window component. The figure shows clearly that the coherence decrease steeply as the signals period goes toward small

periods indicating a weaker relationship between the two variables at these scales, meaning the sea temperature fluctuations may be driven by other factors, such as water advection processes, rather than by air temperature alone. This lack of coherence implies that the model, which relies heavily on air temperature as a predictor, may struggle to capture these sea temperature variations, resulting in residual error even with daily averaging. A confirmation of this can be found also by comparing air and sea temperature anomaly in Figure 5.4 and noting that the two seem to be decoupled on a period of about two weeks.

Indeed the Northern Adriatic is an hot-spot for the dynamic of the Adiatic and temperature measured by AAot can be influenced by the circulation patterns. In fact, it is in this area that gives rise to two of the main currents of the Adriatic Sea: the WAC and DWOC. The first is due to the rather strong positive water balance due to the concentration of riverine discharge; the second is linked to the formation of the Northern Adriatic dense water (NAdw) in winter, during intense Bora winds. This cold, dry northeasterly winds are particularly effective at dense water formation because they blow over the broad, shallow northern shelf, where waters are isolated from the stored heat of deeper layers and can thus cool more efficiently. The cooled waters become denser than surrounding waters and sink, creating a pool. The accumulation of these dense waters sustains a baroclinic pressure force outward of the basin. The geostrophic response to this force creates a convergence against the western boundary that develops into a vertically sheared, bottom boundary current flowing southward. Additionally, a secondary cyclonic circulation frequently develops in the Gulf of Venice. This secondary circulation forms due to the presence of remnant winter dense water at the bottom that generates a local low in the sea surface height distribution, creating a cyclonic circulation around it. The southern branch of this cyclone transports freshened surface waters towards the Istrian coast [14].



(a)



(b)

Figure 5.6: Coherence, phase, gain and power spectra of Venezia air temperature and AAot sea temperature.

5.2 CMEMS averaged temperature recontruction

As discussed in the previous section the issues arisen in the reconstruction of temperature measured by AAot from air temperature data may reside in the predominantly dependence of temperature in that point from advection. To validate this hypothesis spatially averaged temperatures have been used as target to train the Dense model. In this section the results of the reconstruction of temperature from CMEMS reanalysis on the domain (12E-20E, 36.02N-45.98N), averaged spatially on different combination of vertical levels and horizontal extension of the domain are reported. Also a combination of time lags is experimented. Three vertical levels have been considered: 1m (the first model level), the average from 1 m to 5.5 m and the average from 1 m to 10.5 m. To indicate these level it will be used the notation v0, v02, v04 respectively. Time lag experimented are from 0 to 2 days, indicated by the abbreviation l0, l1, and l2 respectively. The domains considered for the horizontal averages are of two kind: the entire domain and geographical subset based on a threshold conditon. The threshold is applied to the mean correlation coefficient between the temperature measured from the weather stations and the sea temperature, $r > 0.5$. Since the latter is a threshold criteria, the extension and shape of the horizontal area considered change with the variation of the others parameter.

Horizontal domain These areas are represented in Figure 5.7, in red: from left to right there are the three vertical domains considered and from top to bottom there are the three time lags.

The area where $r > 0.5$, varies with both the vertical domains and time lags, and its extension decrease at the increase of vertical levels considered, and in the top right panel it's reduced to only few points. This is quite expected because the correlation between air and sea temperatures reduces with the increasing depth. On the other hand, the red area extension increase at the increase of the time lag between air and sea temperatures. This can be due to the lag in the water response to the changes in the air temperature, due to its greater thermal capacity.

Cross validation scores The results of the cross validation for all the targets are reported in Figure 5.8, in terms of distribution of the absolute errors and RMSE. On the x-axis there are the various combination of vertical domains and time lags, denoted with the previous mentioned abbreviations. The distinction between the different horizontal areas considered is represented by the colors blue and orange on the plots, blue for the average on the entire horizontal area, and orange for the average on the area selected by the threshold $r > 0.5$. The

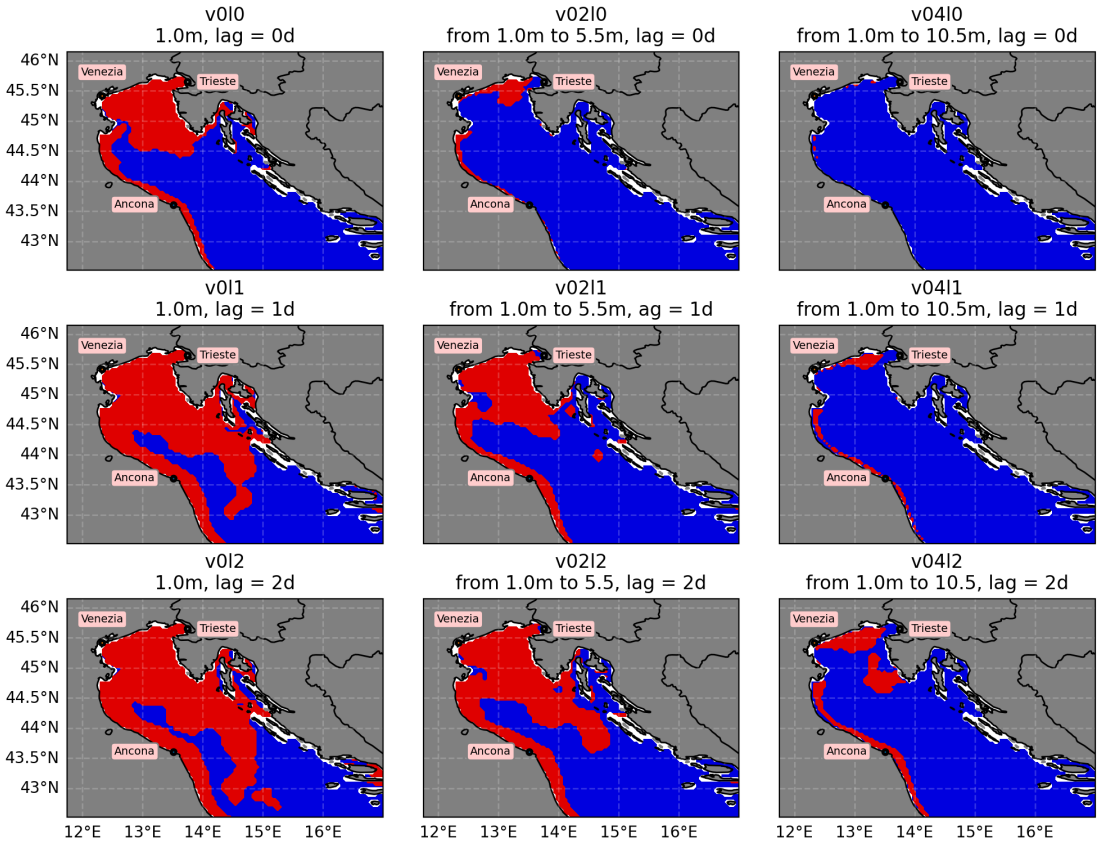


Figure 5.7: Domain considered for the spatial averaging. In red the areas where $r > 0.5$.

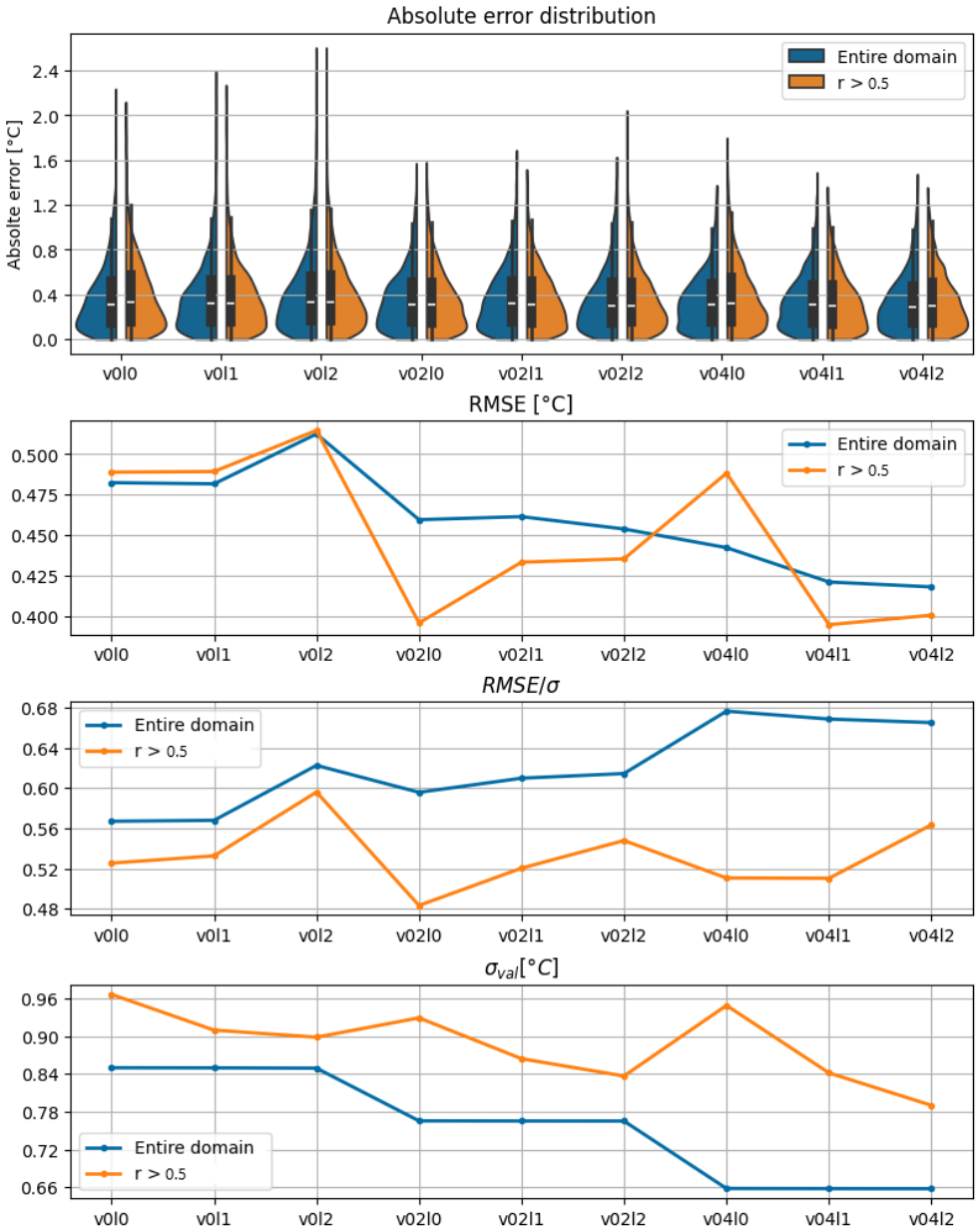
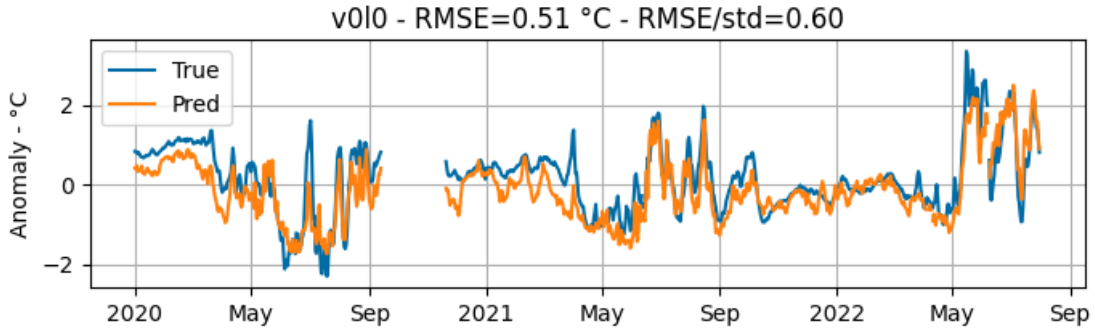


Figure 5.8: Results of the cross validation for the spatial averaged temperatures. First panel: absolute error distribution over all cross validation splits. Second panel: mean RMSE. Third panel: normalized RMSE. Fourth panel: standard deviation of the target.

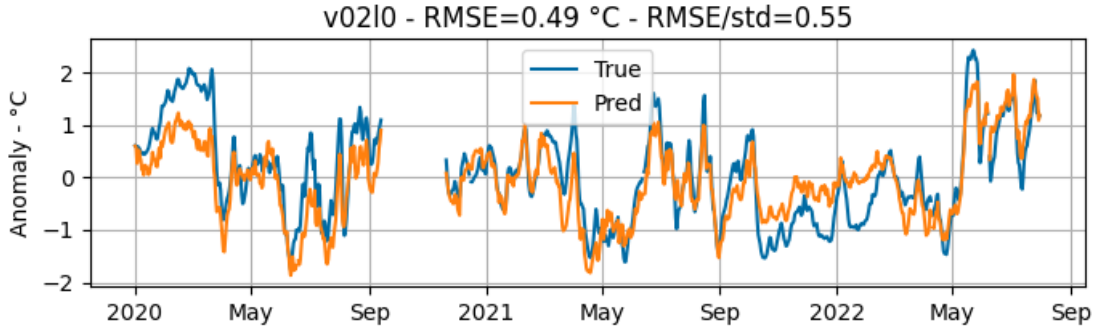
first panel of Figure 5.8 represents the distribution of the absolute errors across all validation years, obtained during the cross validation procedure. For each combination of vertical domain and time lag there is a violin plot with each side for one of the horizontal domain considered. The inner bar of the distribution shows a box-and-whiskers plot. The thicker bar represents the 1^{th} and 3^{th} quartiles, Q_1 and Q_3, and the white tick is the median. The thinner black line represents the whiskers and extends from Q_1 to $Q_1 - 1.5$IQR and from Q_3 to $Q_3 + 1.5$IQR, where IQR is the interquartile range $Q_3 - Q_1$. Because the thin black line should end at observed values, the extension at the two sides are different and the lower boundary is zero. The median for all different targets is always below 0.4 °C and Q_3 is around 0.6 °C. The interquartile range is pretty similar between the various combinations, with some differences in the v0l0 and v04l0 between the two horizontal domains, with the domain with $r > 0.5$ having a bit larger Q_3, meaning that the absolute errors are a bit more spread toward bigger values. In the v0 vertical levels there is an increasing trend in the maximum absolute error and in the upper whiskers, with increasing time lag. In v02 there isn't a clear trend but there is a peak in the in the maximum error at 2 days time lag in the $r > 0.5$ area average, while in v04 the peak is at 0 days time lag.

The second panel in Figure 5.8 shows the mean RMSE across all validation sets in cross validation. The RMSE on v0 has the same increasing trend, as seen before, with the time lag. Same trend also in v02 in $r > 0.5$ horizontal domain, while the opposite is found in v04 entire horizontal domain. The minimum values of RMSE are found for v02l0 and v04l1, with the values of around 0.39 °C. Since the average temperature is computed on different spatial domains, the variability of the data is different between the various domains. So in order to compare score obtained in the reconstruction, the RMSE is normalized by dividing it by the standard deviation of the specific target data. The third and fourth panel in Figure 5.8 shows respectively the normalized RMSE, RMSE/σ, and the standard deviation of the targets σ. As expected the average temperature on the entire horizontal domain average has a lower standard deviation that the temperature averaged on the $r > 0.5$ areas. Observing the normalized RMSE, it's possible to note a general increasing trend with the time lag, with exception of v04 on the entire domain. The areas selected by the correlation threshold ($r > 0.5$) consistently show lower normalized RMSE compared to the full domain averages, indicating that these regions are better represented by the air temperatures, without that spatial averaging compromise excessively the variability .

The best score is achieved with v02l0 configuration in the $r > 0.5$ domain, yielding an RMSE of 0.48 °C, while for the entire domain, the best performance is obtained with the surface-only configuration (v0l0) with an RMSE of 0.57 °C.



(a) v0l0, entire domain horizontal average.



(b) v02l0, $r > 0.5$ horizontal average.

Figure 5.9: Results of the reconstruction on the test set for the two best performing target. v0l0 on the entire domain horizontal average and v02l0 on the $r > 0.5$ area average.

Results on test set Figure 5.9 shows the result of the reconstruction in the test period, for the best performing spatial domain average. Both reconstructions show varying performance across season. There seems to be a tendency on underestimate the temperature anomaly in v0l0 particularly evident in early 2020 and and early 2021. In the same period of early 2020 also v02l0 temperature anomaly is underestimated, but there is also a long period in winter 2021-2022 in which there is overestimation of anomaly.

Coeherence Comparison To further investigate the significant difference in performance when reconstructing the Acqua Alta (AAot) temperature compared to the spatially averaged temperature, a coherence analysis was performed between the Venezia air temperature anomaly and both the AAot temperature and the spatially averaged temperature. The best performance obtained in cross-validation for the reconstruction of the AAot temperature was achieved using daily data, resulting in an RMSE value of 1.04°C (see Figure 5.1). This RMSE is significantly higher than the best RMSE score of 0.48°C obtained for the spatially averaged temperature using the v02l0 dataset for the area with r>0.5 (Figure 5.8). Figure 5.10 shows the coherence plot on period shorter than 30 days, comparing the Venezia air temperature anomaly with the AAot temperature (left panel) and the spatially averaged temperature v02l0 with r>0.5 (right panel). The coherence plot reveals that the spatially averaged temperature (v02l0 r>0.5) exhibits a higher coherence with the Venezia air temperature anomaly compared to the AAot temperature. While the AAot temperature shows near zero coherence for periodicity shorter than 3 days, the spatially averaged temperature demonstrates stronger coherence (0.1) even at these short timescales. The difference becomes more pronounced at increasing periods, where the v02l0 coherence raches values over 0.6, while the AAot coherence remains confined under 0.3 This support the hypothesis of advective dominated processes in the AAot temperature, that are less present in the v02l0 temperature, given the choice of the domain, based on correlation coefficient with air temperature, and the effect of the spatial averaging.

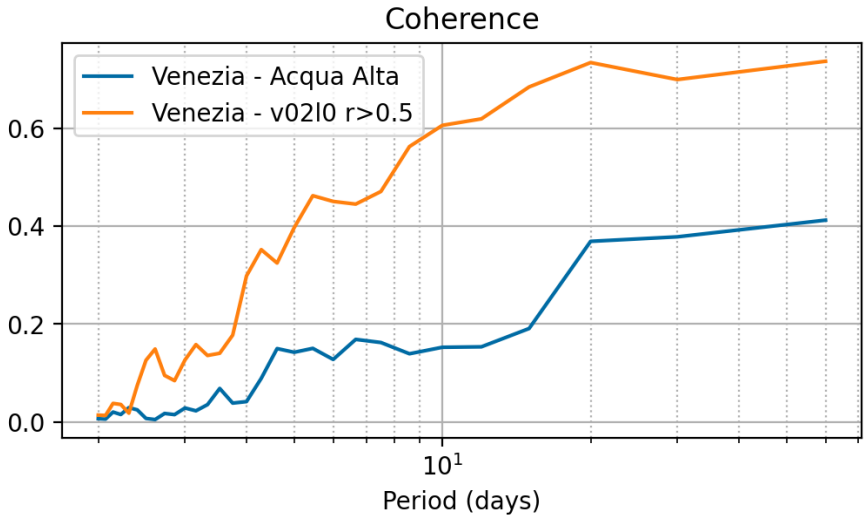


Figure 5.10: Comparison of the coherence between Venezia-Acqua Alta temperature and Venezia-v02l0 $r > 0.5$ spatially averaged temperature.

5.3 CMEMS pointwise temperature reconstruction

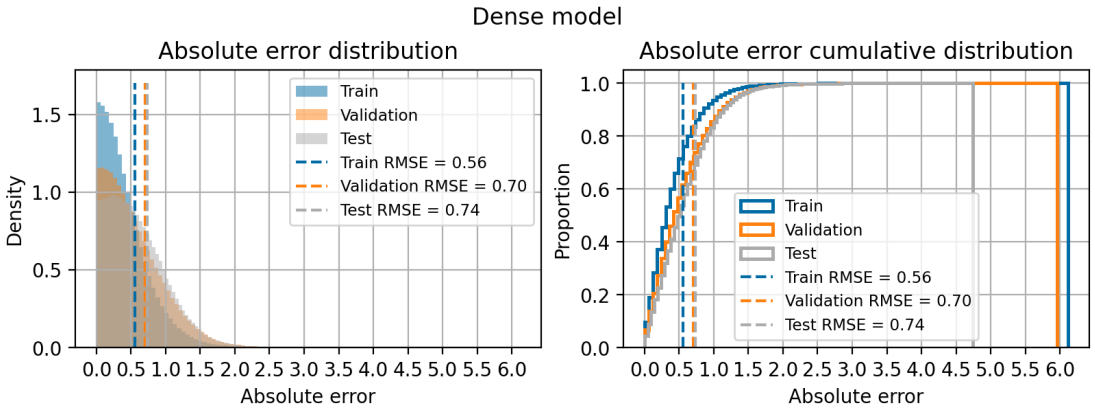
In this section are reported the results of the reconstruction of the reanalysis temperature from CMEMS, pointwise in the domain (12E-20E, 36.02N-45.98N). The performance of the reconstructions of the two models tested are compared.

5.3.1 Model Performances

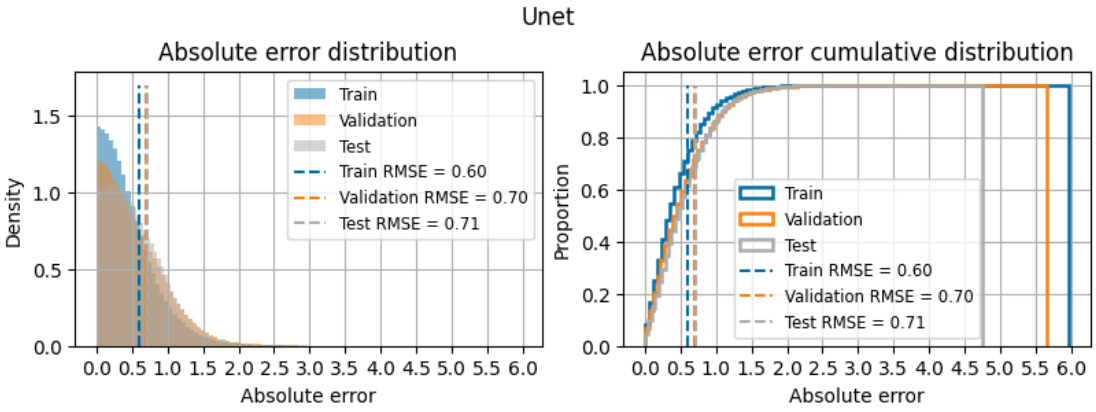
Distribution of errors

Figure 5.11 shows the distribution of absolute errors and their corresponding cumulative distributions across training, validation, and test sets for each model architecture. Both architectures demonstrated comparable performance characteristics, with the Dense model achieving training, validation, and test RMSE values of of 0.56, 0.70, and 0.74 °C respectively, while the Unet model exhibited RMSE values of 0.60, 0.70, and 0.71 °C. These metrics suggest similar generalization capabilities between the two approaches, with only marginal differences in their learning patterns. The absolute error distributions for both models exhibit right-skewed patterns, with the highest density of errors concentrated in the 0.0-1.0 range. This asymmetric distribution indicates that while most predictions maintain relatively low error, there exists a tail of bigger errors extending beyond 2.0, as we can see in the cumulative plot. The maximum error in the training and validation sets is slightly higher in the Dense model, while in the

test is very similar. It can be also noted that around the 70% of all predictions have a lower absolute error than the RMSE.



(a)



(b)

Figure 5.11: Absolute error distribution in the training, validation and test sets for the Dense model and Unet model.

In both models the training set has the lowest RMSE and error rates, as evidenced by the leftmost peaks in the distribution plots and the steeper rise in cumulative distributions. The validation and test sets show slightly higher RMSE, with highly similar error patterns. This suggest a slightly level of overfitting, but the similar performances on the validation and test sets, suggest a stable capacity of generalization.

Spatial distributions of model performances

The performances of both models (Dense and Unet) have evaluated on the test set using the four statistical metrics: Mean Bias (MB), Standard Deviation Error (SDE), Correlation Coefficient (CC), and unbiased Root Mean Square Error (uRMSE). Spatial patterns of the metrics, are shown in Figure 5.12.

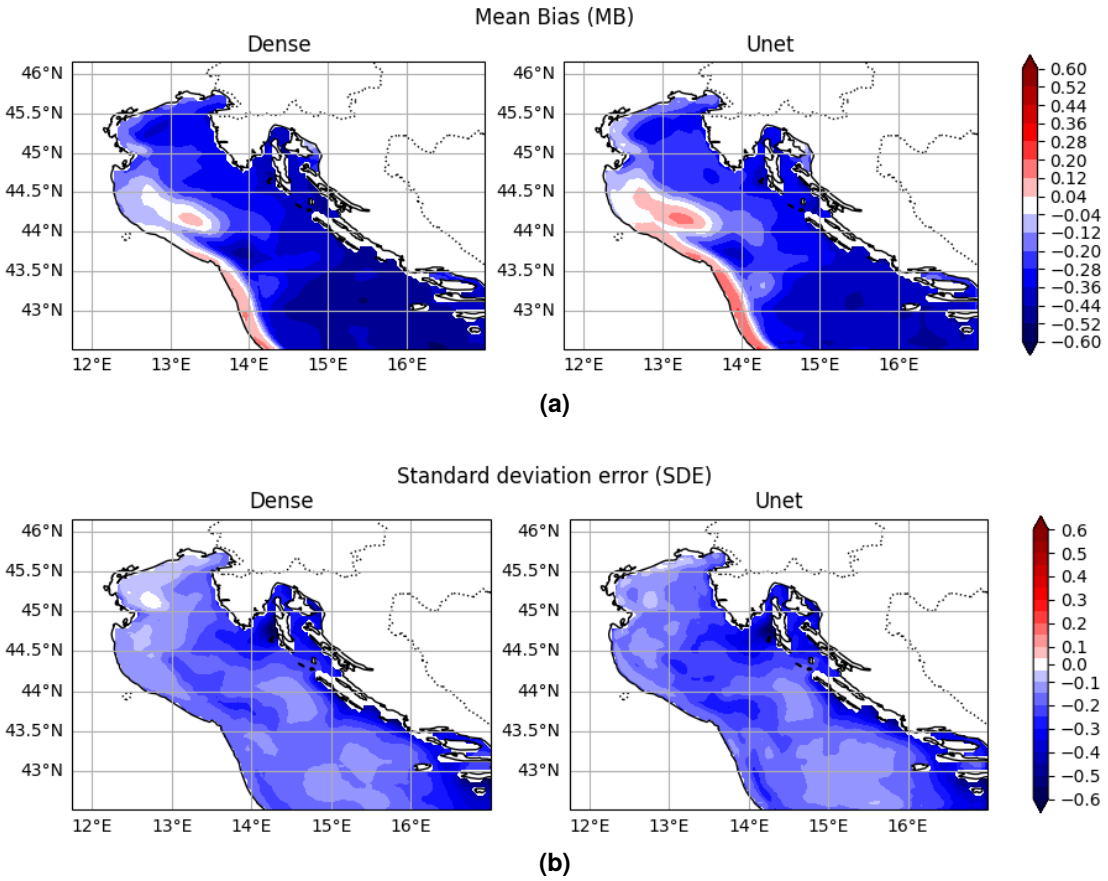


Figure 5.12: Comparison of the spatial distribution of the evaluation metrics (part 1).

Mean Bias (MB) Both models exhibit predominantly negative bias, with more negative values southward. We can argue that this feature can be due to the increasing distance from the meteorological stations and to the advective component arriving from the southern part of the basin. The Unet model shows generally reduced negative bias compared to the Dense model, as evidenced by the lighter blue coloring across the domain. A distinct region of positive bias appears at around 44°-44.5°N, 13°-13.5°E in both models and along the western coast, though the Unet model shows a more concentrated pattern of this

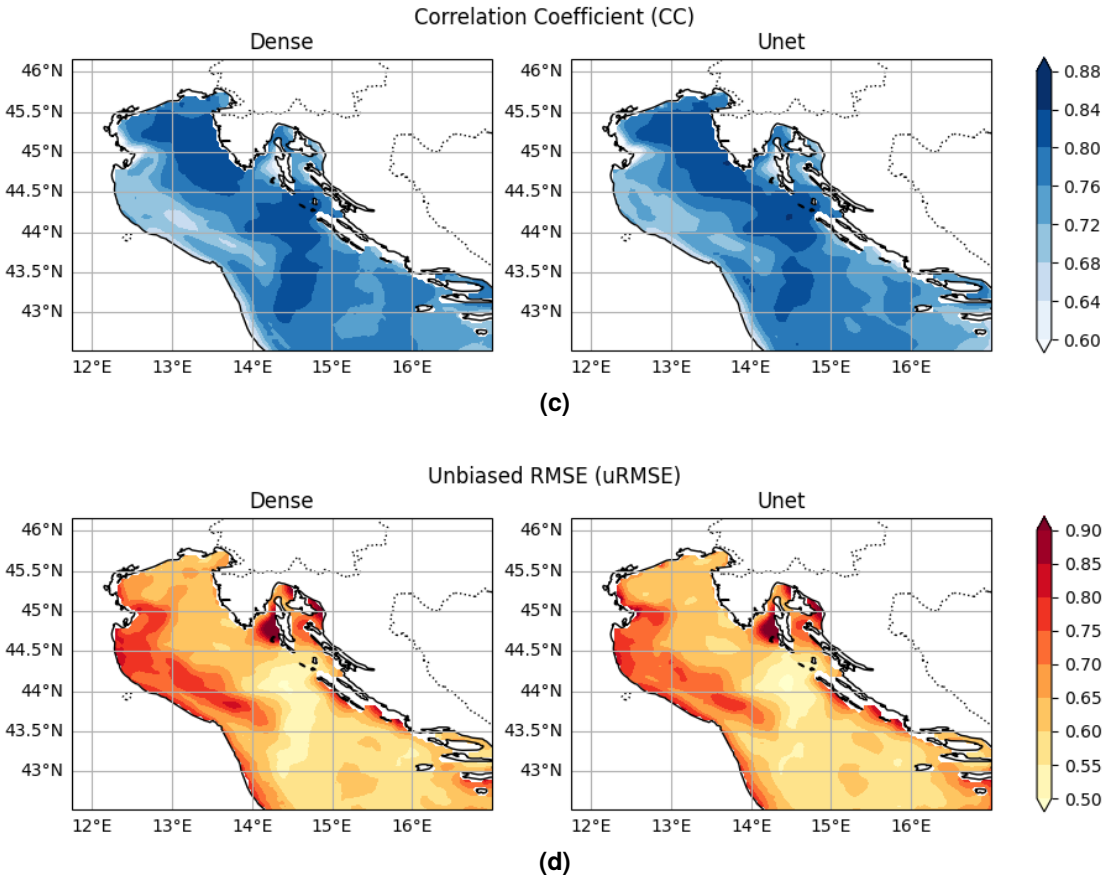


Figure 5.12: Comparison of the spatial distribution of the evaluation metrics (part 2).

	MB	SDE	CC	uRMSE
Dense	−0.325	−0.300	0.104	0.468
Unet	−0.264	−0.325	0.052	0.474

Table 5.1: Mean of temporal metrics. Comparison between Dense and Unet.

positive anomaly. This suggests that while both models tend to underestimate sea temperatures, the Unet architecture partially mitigates this systematic error, at the cost of slightly higher positive bias in these regions. Here the error is probably related to the ocean dynamics and presence of the WAC. We know that the intensity and extension of this current is poorly related to the atmospheric temperature.

Standard Deviation Error (SDE) The SDE patterns are similar between both models, with general underestimation of standard deviation. Smaller errors are found in the North while larger negative SDE values are observed on the eastern coast, under Istria peninsula. We can assume that this is due to the presence of complex coastal geometry that can lead to sub-mesoscale processes and thus to an enhanced variability, not directly driven by atmospheric temperature.

Correlation Coefficient (CC) Both models demonstrate generally good correlation coefficients throughout the domain, with most of the values in the range 0.6-0.85. Regions with lower values are found in the region south of Istria and around the Po mouth. In this latter the significant freshwater input can lead to decorrelation with air temperatures. In the western coast is recognizable a pattern with lower CC than the surrounding, that can be associated to the presence of the WAC.

Unbiased RMSE (uRMSE) The uRMSE values range from approximately 0.5 to 1.1 for both models, with generalized slightly lower error for the Unet model. Highest errors (in red) are observed in the eastern coast under Istria peninsula where also the SDE is bigger, while lower errors (lighter orange) are present in the central and southern regions. In the western coast there is a pattern of higher uRMSE than the surrounding, similar to the the CC distribution. This indicates that the model lack in the ability of capture both the sign of the temporal fluctuations (CC) and the intensity (uRMSE), in these regions.

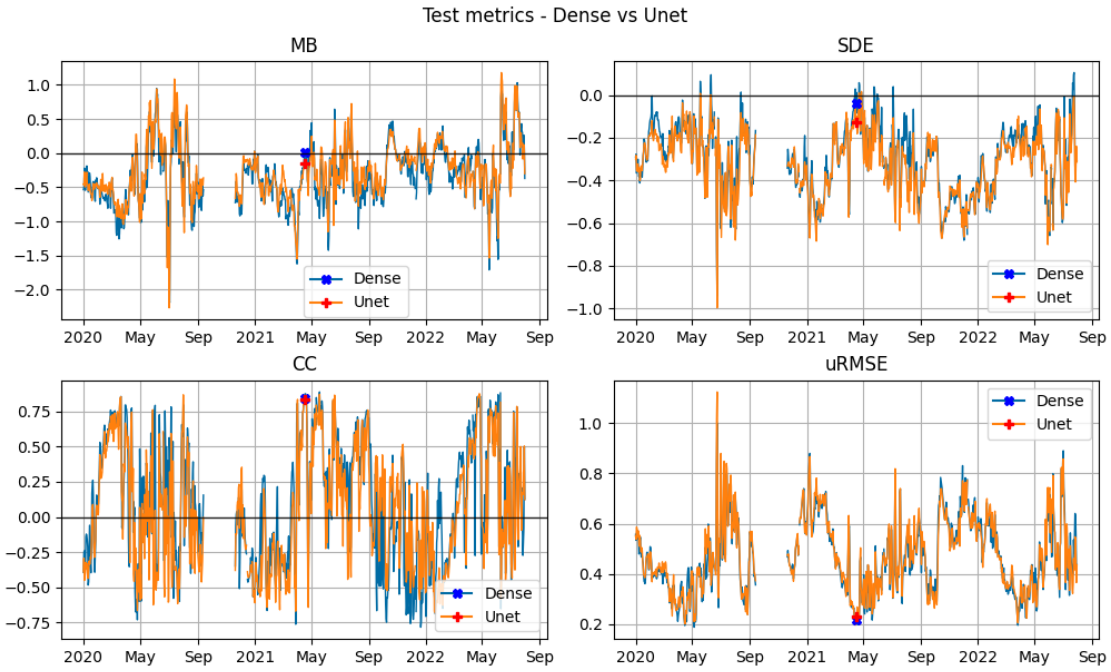


Figure 5.13: Temporal distribution of evaluation metrics. Markers indicate the 17th April, day with optimal reconstruction scores.

Temporal Evolution of Models Performance

The temporal evolution of performance metrics for both Dense and Unet models over the test period, is shown in Figure 5.13, providing insights into the models' temporal behavior and consistency. Mean values of the metrics are shown in Table 5.1.

Mean Bias (MB) Both models show temporal variability in bias, generally fluctuating between -1°C and +0.5°C, with a tendency of the Dense model toward lower values. The most extreme temperature biases occur during summer months. During this period, positive biases can exceed +1°C, while negative biases can drop below -1.5°C. In summer 2020 there is the worst performance in terms of mean bias, with values exceeding -2°C.

Standard Deviation Error (SDE) The SDE values predominantly oscillate between −0.5°C and +0°C, denoting a constant underestimation of the standard deviation, confirming a the tendency of the models to fail to capture well the variability of the sea temperature, both spatially and temporally. Both models show similar temporal variability, with periodic fluctuations that might be

related to seasonal patterns.

Correlation Coefficient (CC) The CC values for both models fluctuate considerably over time, ranging from approximately -0.75 to +0.8, indicating varying levels of spatial coherency between the true and reconstructed temperature. Both models show similar temporal patterns in their correlation structure, with periods of high (>0.5) low (<0.5) and negative correlation. This suggests that the reconstruction skill varies significantly with time. In particular periods of negative CC indicate that the models predict a positive anomaly where there is a negative anomaly and vice versa, on most of the spatial domain. In some periods there are slightly better performances of one model over the other, as in May-September 2020, with Dense model, having higher CC, or in September-December 2021, with Unet having better score.

Unbiased RMSE (uRMSE) The uRMSE values are significantly fluctuating, with minimum values around 0.2°C and usually over 0.4°C. Values often exceed 0.8°C with a peak at 1.2°C between May and September 2020. A stagionality seems to be present, with minimum values around May, and peaks in summer and winter

Both models show similar temporal variability patterns, suggesting that their performance is influenced by common underlying factors, possibly related to specific ocean-atmosphere coupling. In particular the seasonal cycle, more evident in the SDE and uRMSE plots, can be due to the seasonality of the intensity of sub-mesoscale processes.

Temperature fields visualization

Anomalies In Figure 5.14 the reconstruction results of the models are plotted and compared to the true anomaly field. 17^{th} of April 2021 has been selected as it represents the day with the best overall model performance. This date was identified by finding the minimum of a composite score that combines the considered statistical metrics, for both models:

$$\text{score} = |MB| + |SDE| - CC + uRMSE \tag{5.1}$$

The performance metrics can be visualized in Figure 5.13, marked with a blue cross for the Dense model and with a red plus for the Unet. The good performances of both models are confirmed by looking at Figure 5.14, where we can see that the north-south gradient is well reproduced, with stronger negative anomalies around -2°C in the north, progressively weakening to values around

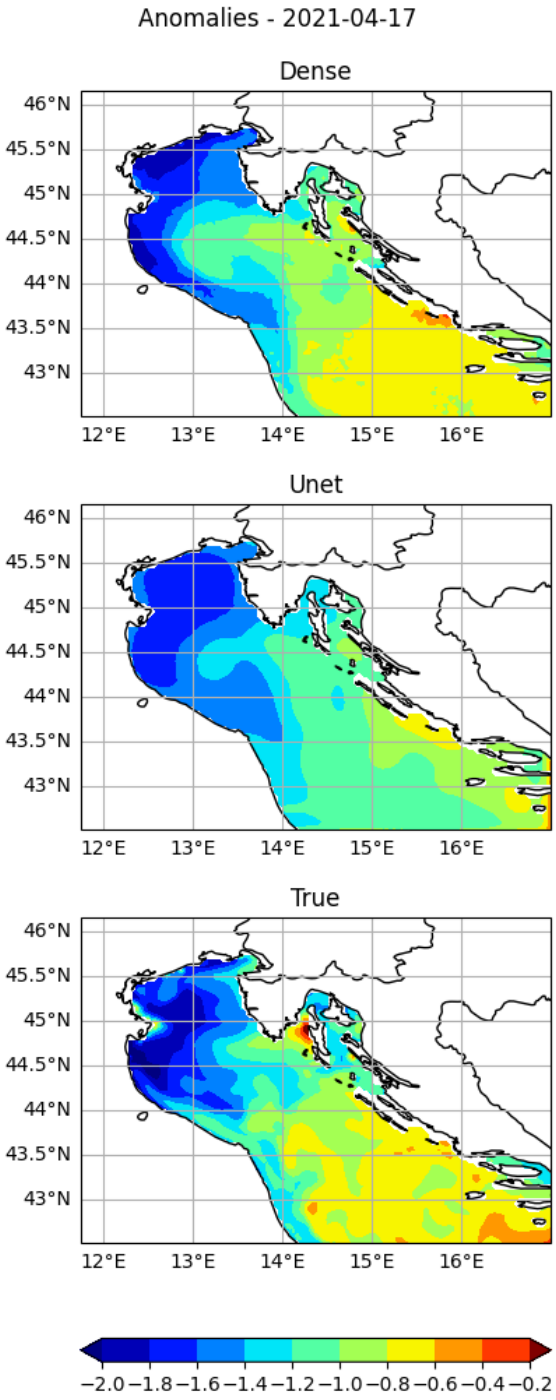


Figure 5.14: Temperature anomalies fields. Top: Dense model reconstruction. Center: Unet reconstruction. Bottom: True anomalies field.

-0.6°C towards the southern region. Of particular note is the approximate reconstruction of key features such as the cold pool (anomalies between -1.4°C and -2°C) confined between the Po River delta and the Istrian peninsula, as well as the meandering pattern that extends along the western coast down to Ancona, even though the magnitude is overestimated. The Dense model shows an overal better performance, characterized by the better representation of the temperature ranges in both the northern and southern areas.

However, both models exhibits some limitations in reproducing finer-scale structures present in the true field, this is because small structures are decoupled from the atmospheric conditions. While the general magnitude of the anomalies is well captured, the reconstruction tends to produce slightly more negative values (by about 0.2-0.4°C) in the southern region. The most significant discrepancies are observed in the archipelago area south of the Istrian peninsula. This region consistently shows poor reconstruction quality, as confirmed by the spatial distribution of evaluation metrics in Figure 5.12, where it is characterized by elevated SDE and uRMSE values, along with reduced CC.

Full range Figure 5.15 presents the full-range temperature fields obtained by adding back the climatology to both the reconstructed and true anomaly fields. A comparison with the climatological field for that day is included to evaluate the model's improvement over this baseline reference. Both patterns look very similar, with colder waters (11-12°C) in the northern Adriatic basin gradually transitioning to warmer waters (14-15°C) in the central and southern regions. In contrast, the climatological field shows a markedly different pattern with substantially warmer temperatures overall, particularly in the southern portion of the domain where temperatures exceed 15°C. Of particular note is the consistency between the observed and modeled fields in representing the cold pool in the northern basin and the temperature front in the central Adriatic, features that are absent in the climatological field. Indeed the climatology in this particular day is quite inaccurate. Figure 5.16 shows the standard deviation with respect to the climatology for the 17^{th} of April. The standard deviation have noticeable variability with the southeastern part of the region having lower values, under 0.43 °C, and increasing westward and northward, reaching a maximum over 1.39 °C in the northern area, indicating a large inter-annual variability in this region.

The analysis of the spatial and temporal distribution of the metrics, along with the visualization of the reconstructed temperature field highlighted the similar behavior of the two models, with only slightly differences. So the increased complexity of the Unet doesn't contribute to a significant improvement and we can hypothesize that this is due to fact that the information processed

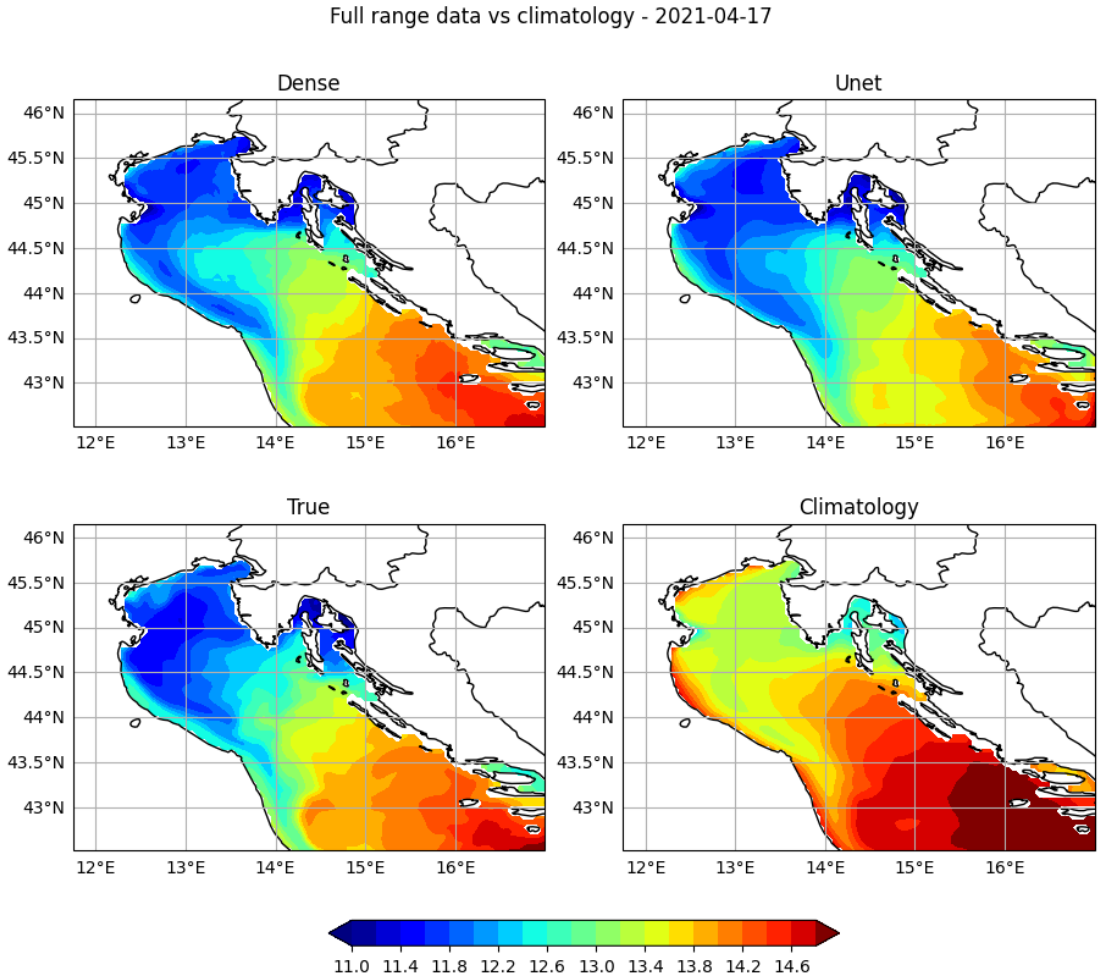


Figure 5.15: Full range temperature fields. Top left: Dense model reconstruction. Top left: Unet model reconstruction. Bottom left: true temperature field. Bottom right: climatology.

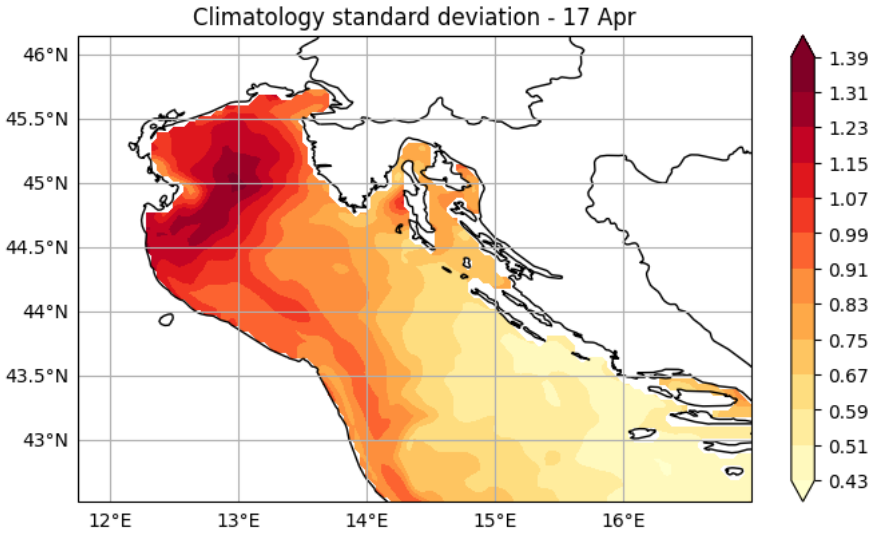


Figure 5.16: Standard deviation with respect to the daily climatology for the 17^{th} of April.

by the Unet passes trought the bottleneck of the Dense model, establishing the capability of reconstruction.

5.3.2 Comparison with spatial averaged temperature

Here a comparison between the reconstruction of the area averaged temperature and the recontruction made pointwise and then spatially averaged over the test period is presented. In order to use consistent spatial domains, v0l0, i.e. the surface level with zero time lag, averaged horizontally on the entire domain is used in this comparison. Figure 5.17 shows the time evolution of the mean bias over time. For v0l0 the the mean temperature has been computed before the reconstruction, so the mean bias is only the difference between the output of the model and the target. For the pointwise reconstruction the horizontal averages are computed after the reconstruction. The temporal behavior of the MB is similar among the three methods. The oscillations of v0l0 tend to be a slightly less accentuated; the Unet MB tend tends to follow the v0l0 trend, with slightly more fluctuations and the Dense model MB tends more to negative values. The mean values confirm this ovservation.

However the consistent trend among the three methods indicates a good ability of reconstructing the averaged temperatures, even from a pointwise reconstruction, for both models.

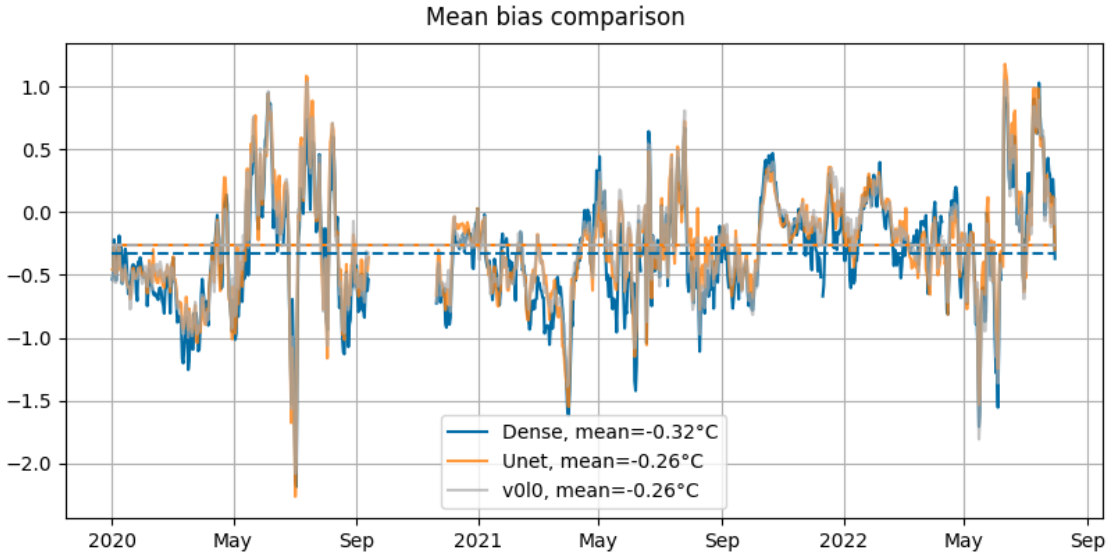


Figure 5.17: Comparison of time evolution of mean bias, between v0l0 and pointwise reconstruction of the Dense model and Unet model.

5.3.3 Comparison with satellite

In this section a comparison between the satellite, neural network models' output and the reanalysis is performed. Spatial and temporal characteristics of the previously adopted metrics is analyzed. The obtain a consistent analysis with previous ones, the metrics are computed on the anomalies, in the test period. Anomalies of the satellite temperatures have been computed by removing its own climatology.

Spatial analysis

Mean Bias MB spatial distribution is shown in Figure 5.18a. First panel shows the comparison between reanalysis and satellite, that evidence a pretty marked general positive MB, with the exception of little spot where MB is slightly negative. All the western coast shows positive bias more pronounced than other regions. In the bottom panels the comparison between the neural network models and the satellite shows instead a diffuse negative bias with a marked positive bias in the western coast. This is due to the general underestimation of the anomalies of the neural networks reconstruction of the reanalysis, except on the western coast. Between the two models the Unet has slightly better performance due to the less severe underestimation.

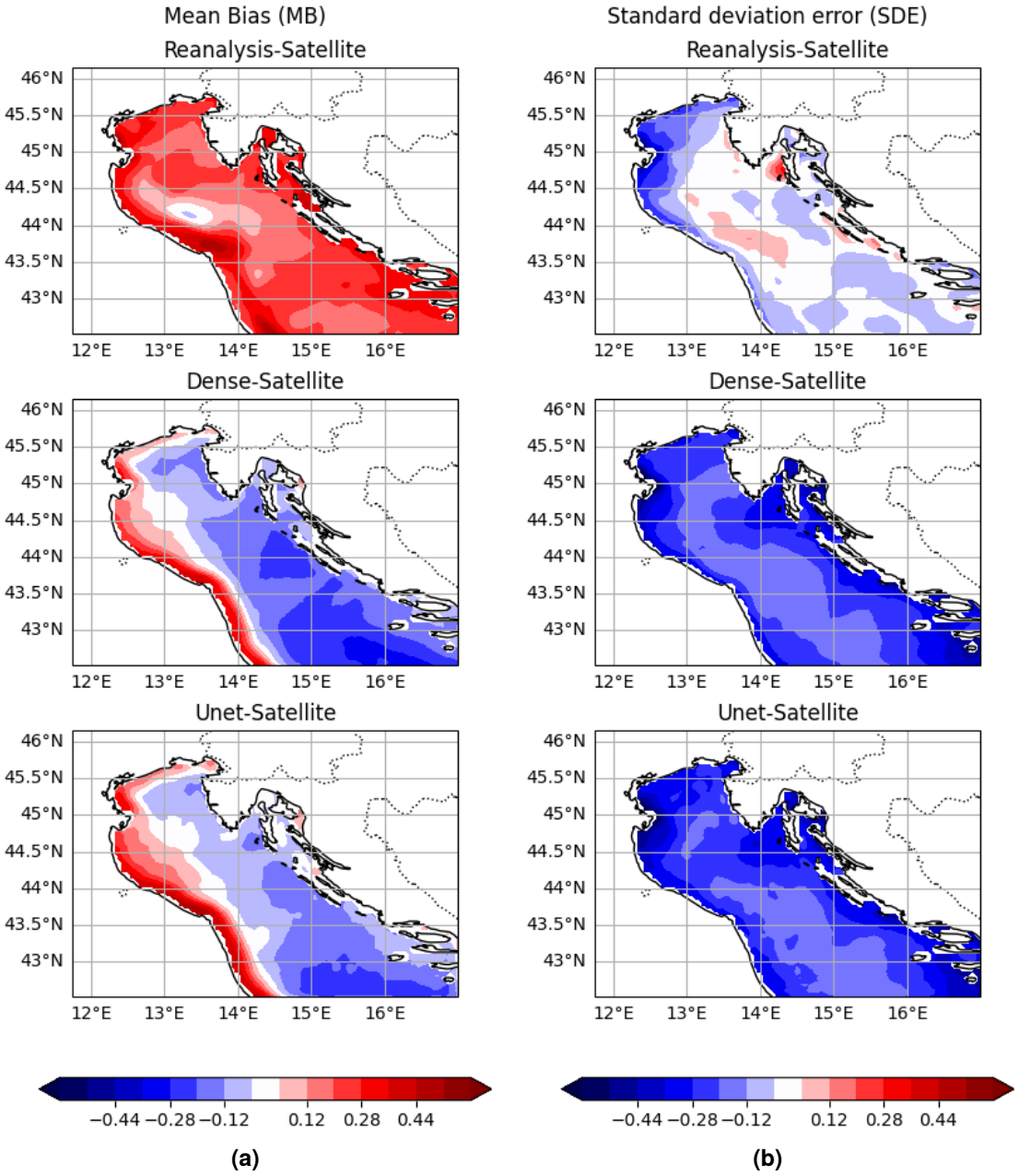


Figure 5.18: Spatial distribution of the evaluation metrics in the comparison between reanalysis and satellite, Dense model reconstruction and satellite, Unet reconstruction and satellite (part 1).

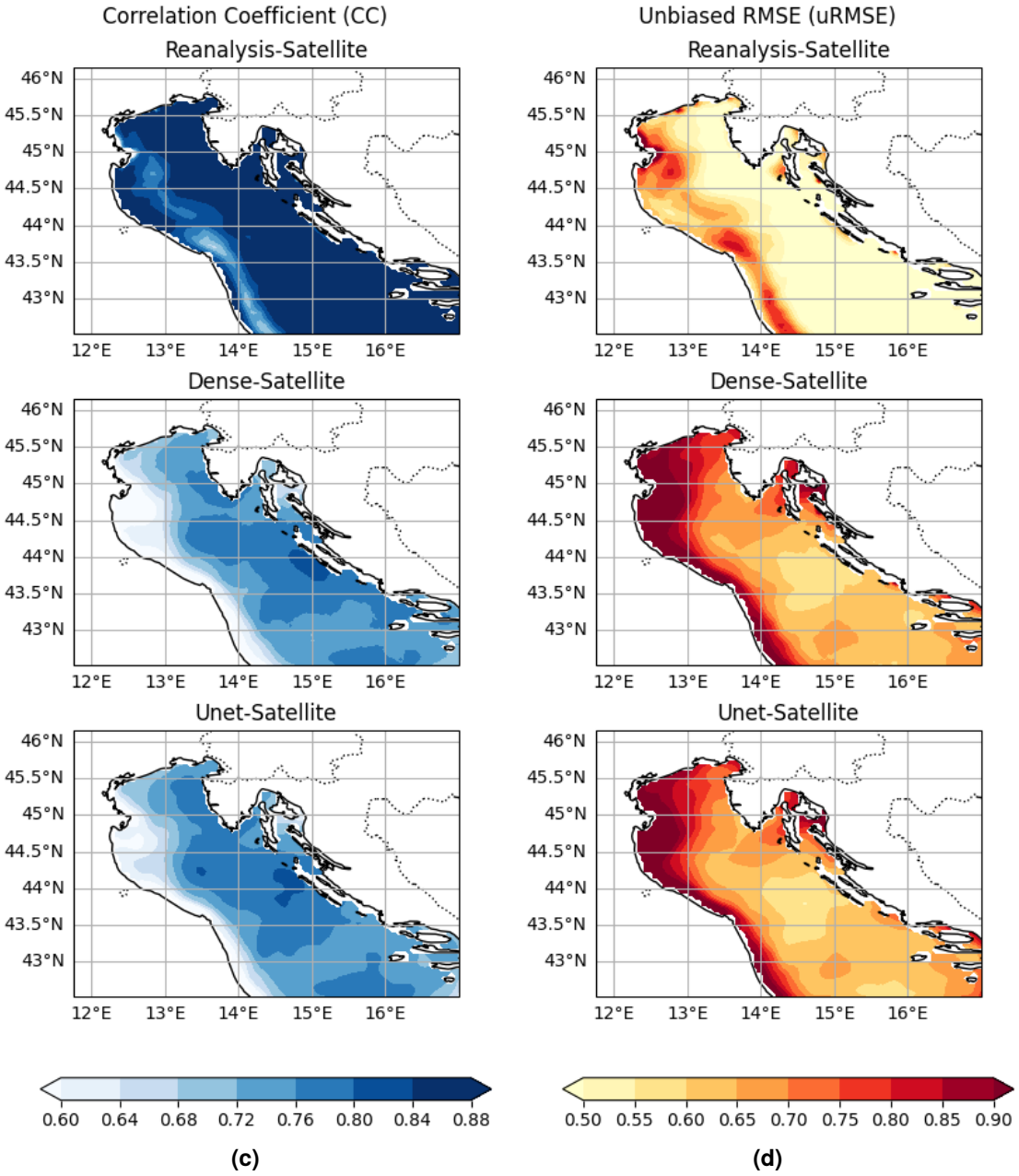


Figure 5.18: Spatial distribution of the evaluation metrics in the comparison between reanalysis and satellite, Dense model reconstruction and satellite, Unet reconstruction and satellite (part 2).

	MB	SDE	CC	uRMSE
Dense-Sat	−0.103	−0.349	0.091	0.553
Unet-Sat	−0.042	−0.373	0.074	0.515
Rean-Sat	0.227	−0.051	0.517	0.453

Table 5.2: Mean values of the temporal metrics.

Standard deviation error Figure 5.18b shows the SDE spatial distribution. The general low values of SDE between reanalysis and satellite indicate a consistent ability to capture variability of the sea temperature. However there are some areas with significant SDE values, as in the northern region where there is an underestimation of the variability observed from the satellite, in the reanalysis, and in the island under the Istria peninsula, where the reanalysis overestimate the variability. This can be due to the complex dinamic in this region, that the satellite struggle to capture due to the proximity with the indented coastline and the issues of land contamination effects. On the other hand, both neural network models strongly underestimate the variability, with larger effect in the northern areas.

Correlation coefficient Correlation coefficient is shown in Figure 5.18c. Reanalysis and satellite exhibit a strong CC throughout all the region, with values higher than 0.88 and lower values only in a band near the western coast, that can be associated with the dynamic of the WAC. In this region CC of the neural network models with the satellite present its lower values, with better performances eastward.

Unbiased RMSE Figure 5.18d shows the uRMSE spatial distribution. Patterns in the spatial distribution resemble the pattern of the CC with generalized low error, under 0.5°C for the reanalysis-satellite comparison and higher values the near the western coast, with values reaching around 0.8°C. The error in the neural network models is model considerably higher, especially in the northwest region, with values over 0.9°C.

These patterns, highlights the presence of bias and inconsistency in some areas, among all methods. The reanalysis and satellite have a good degree of agreement, but the MB is quite marked almost all domain. MB between neural network reconstruction and satellite is instead mitigated by the underestimation of reanalysis temperatures.

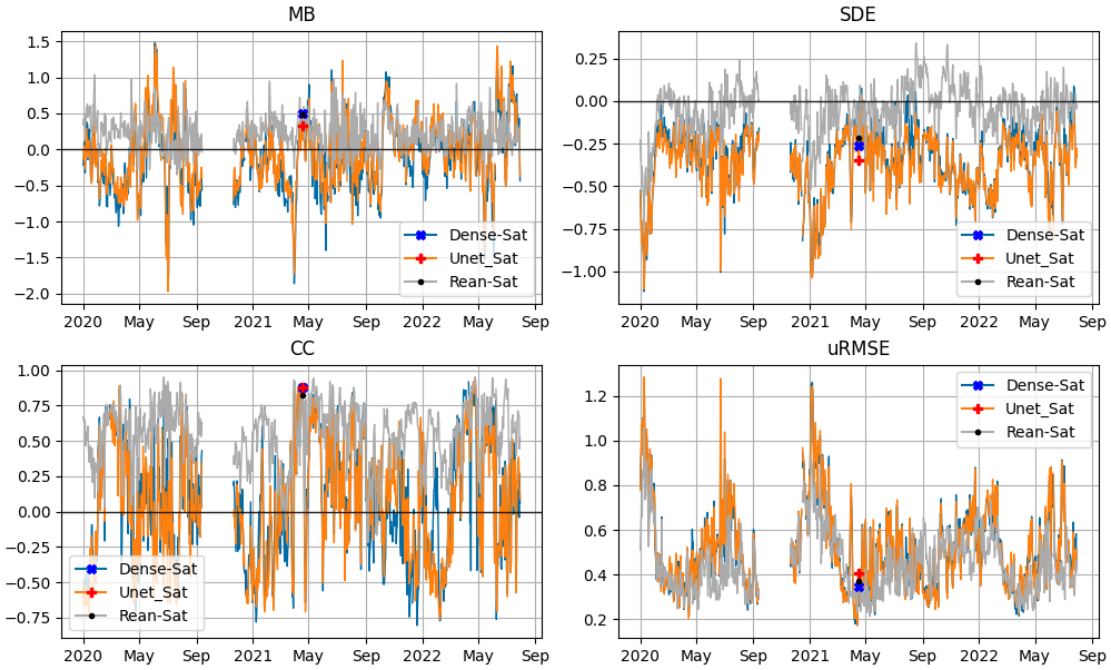


Figure 5.19: Temporal distribution of the evaluation metrics in the comparison between reanalysis and satellite, Dense model reconstruction and satellite, Unet reconstruction and satellite.

Temporal analysis

Figure 5.19 show the temperal behavior of the metrics, with mean values shown in Table 5.2.

Mean Bias In satellite-reanalysis comparison, MB is confined between around -1°C and +0.5°C with relatively small variations. MB for the neural networks-satellite comparison generally fluctuates between -1°C and +1°C but often shows larger negative fluctuations exceeding -2°C with several pronounced spikes, particularly in mid-2020 and mid-2022. Temporal mean of MB in Table 5.2 shows a slight tendency of the reanalysis to overestimate MB. The Satellite-Unet comparison exhibits the largest variations, with MB values ranging from -1°C to often over +1°C and frequent positive spikes, reaching and exceeding +2°C throughout the entire period. These figures evidence that the satellite is significantly more in agreement with the reanalysis regarding the mean temperature over the basin, with a confined behavior of the mean bias throughout time. The neural network models sometimes struggle to reconstruct the right mean temperature, as already evidenced. These errors are often of the same sign of the one between satellite and reanalysis, and thus are amplified in the mean bias between satellite and Unet. This is the case of the March-April 2021 period in which errors between neural network models and Unet exceed 1°C in magnitude while in the neural networks-satellite comparison they are bigger than 2°C in magnitude. The same happen in October-Novemeber 2022.

Standard Deviation Error Satellite-reanalysis DE range between -0.5 and 0.25°C with some spikes exceeding these values, in particular in early 2020 when SDE reaches around -0.75°C. Neural networks-satellite SDE is generally shifted towards negative values, with values oftwn exceeding -0.5 °C and a peak under -1°C in early 2020. As for the MB some underestimates and overestimates of the SDE happen in the same period between satellite and neural networks, with respect to the reanalysis, leading to an increased SDE between the two.

Correlation Coefficient All three comparisons shows a similar temporal pattern of CC, with higher values in winter and lower values in summer. CC for the reanalysis-satellite comparison is usually over 0, except for some values as in spring and autumn 2021. On the other hand values for neural networks are largely fluctuating, ranging between -0.75 to over +0.75, indicating varying spatial consistency in the temperature distribution.

Unbiased RMSE All uRMSE plots shows the same temporal behavior, similar to the one already seen berween the neural network models and the reanalysis, in Figure 5.13, but with more pronounced peaks in winter. Reanalysis-Satellite uRMSE values are manily in the range 0.2°C-0.8°C, with peaks in early 2020 and early 2021. Neural network uRMSE pattern is similar but with larger amplitude, with values occasionally exceeding 1.2°C.

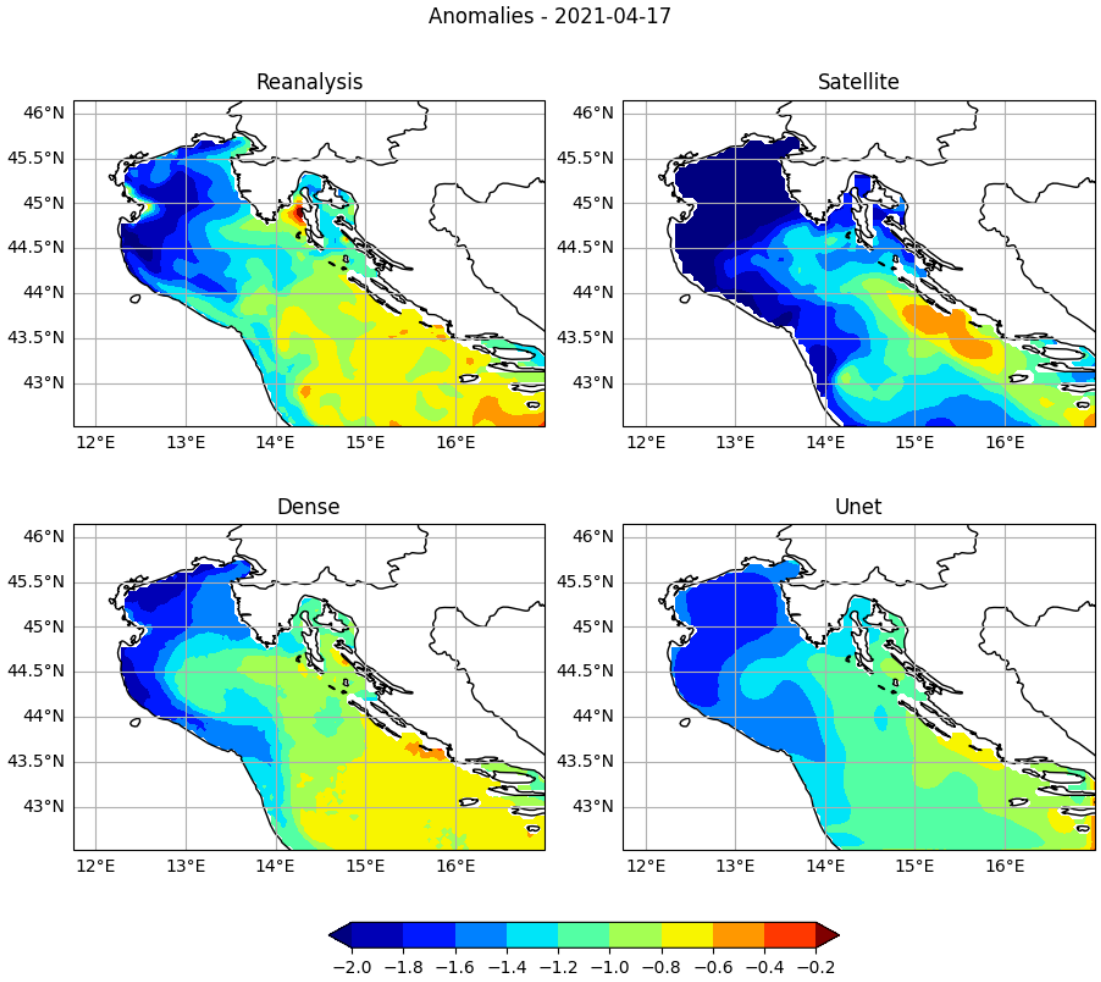


Figure 5.20: Temperature anomalies fields of the reanalysis and satellite, compared with the reconstruction of the neural network models.

Anomalies field

Here a comparison between reanalysis and satellite is reported to highlight the possible inconsistencies between the two and potential goodness of the neural

networks' reconstruction. Figure 5.20 shows the temperature anomalies fields for the reanalysis and satellite, on the 17th April 2021. This day has been chosen as example of a good performance in the neural networks reconstruction, as described in Section 5.3.1. In the figure is possible to see that in this day reanalysis and satellite fields have significant differences. In the northern region the temperature observed by the satellite is consistently lower than the reanalysis, with values ranging between -1.2°C and -2°C for the former and values smaller than -2°C for the latter. The temperature is generally overestimated by the reanalysis, with the exception of a spot in the western coast where the satellite measure a warmer anomaly, so we expect a positive MB, as confirmed by Figure 5.19, where the value is highlighted by a black dot. The variability, and the spatial distribution of the features, are quite consistent so we can see that SDE, CC, and uRMSE have good values as expected. It is worth noting that for this specific day the neural network models exhibit a greater agreement with the reanalysis than the reanalysis with the satellite.

5.4 Comparison with Acqua Alta statistics

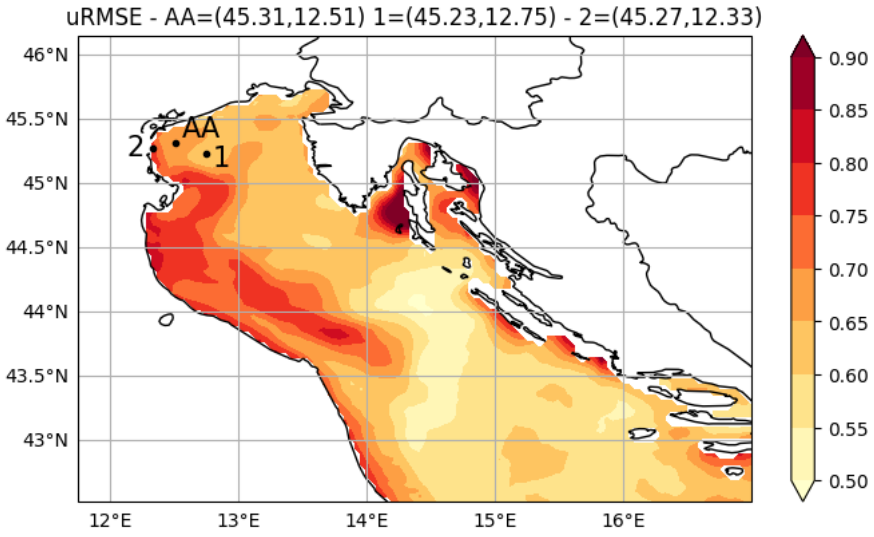


Figure 5.21: Positions of AAot and of the two points used to compare the statistics with, superimposed to the uRMSE spatial distribution of the Dense model reconstruction.

As evidenced by the previous analysis, the performance of the pointwise reconstruction has substantial spatial variability. In this section the statistics of the reconstruction of the AAot temperature are compared with the statistics of

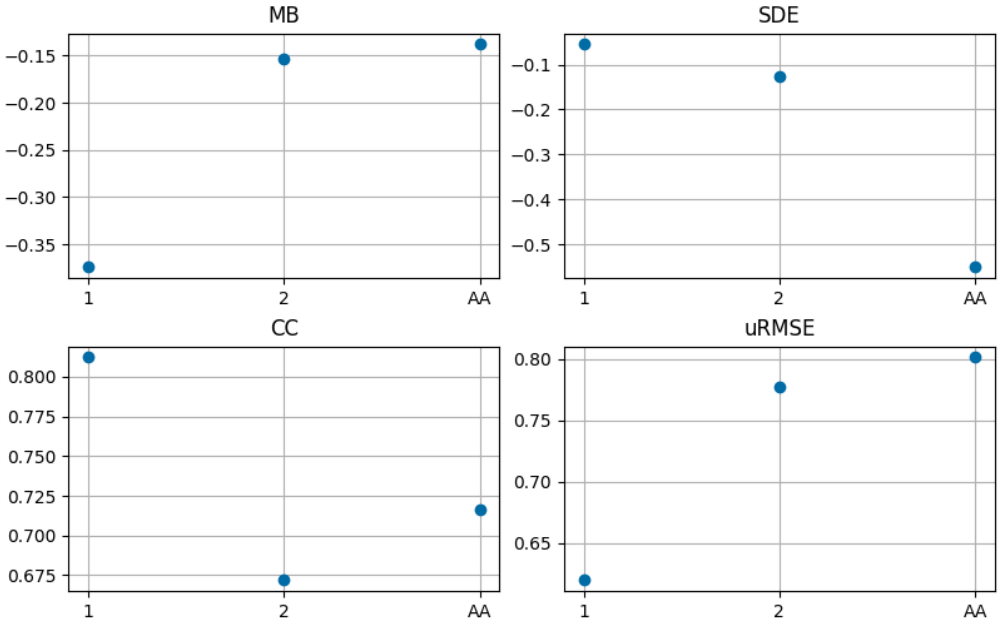


Figure 5.22: Statistics of the temperature reconstruction in points 1 and, compared to the statistics of the reconstruction of AAot temperature.

the temperature reconstruction in two points in the proximity of the platform. The data used to compute the statistics for AAot are obtained by using the Dense model and the daily mean data of the first experiment, with training, validation and test splits consistent with the pointwise experiment. The position of the two points and of AAot are shown in Figure 5.21. These points have been chosen for their relatively high and low values of uRMSE despite being geographically close, to highlight that the non optimal performance in the first experiment can be due to the characteristics of that area. Indeed, even though the uRMSE in the pointwise reconstruction has not a maximum in the AAot point, the neighbor point 2 shows an higher uRMSE than the surrounding. Point 1 serve as term of comparison for statistics in a region with different characteristics. The statistics of the reconstruction of the two points and the AAot temperatures are shown in Figure 5.22. MB for point 1 is larger in magnitude than MB in both point 2 and AAot, that have similar values. On the other hand the variability of point 1 is better captured while in point 2 and especially for AAot is underestimated, as shown by the SDE values. CC is higher in point 1, with lower values for the AAot reconstruction ad in point 2. The uRMSE shows similar values for both point 2 and AAot.

The statistics values show that in the reanalysis there are points with similar characteristics to AAot. The lack of exact spatial coincidence may be due to

a representation error in the reanalysis. However, the spatial distribution of performance metrics confirms the complexity of deriving sea temperature from air temperature in this region, explaining the sub-optimal results in the first experiment.

Chapter 6

Conclusions

This thesis presents an approach to reconstructing SST using atmospheric data from coastal weather stations. This approach has several advantages, that make it interesting for further development. In fact, once tuned, this method is independent from the existing monitoring methods of sea temperature, and thus presents a complementary alternative, that doesn't suffer from the same limitations. *In-situ* measurements are sources of reliable data but these are often sparse and they ere not able to give a simultaneous coverage of a wide area. On the other hand satellite observations, that have the advantage of having a wider spatial coverage, suffer from some limitations, derived by the interaction of the radiation emitted by the sea with the atmosphere. They measure a different SST from the *in-situ* measured one and have a major issue give by the gaps in the observed fields, given by the presence of clouds for IR measurements or by contamination of other radio frequency sources for PMW observations.

An air temperature-derived sea temperature has also time-related characteristics: it can potentially be used to reconstruct historical data, where direct measurements were unavailable and has the advantage of exploiting weather data, that are often available in real time.

In addition, by leveraging coastal weather stations the neural network approach doesn't require significant additional operational costs, emulating the spirit of the ship-of-opportunity program exploiting already available resources, in a method that can be called "data-of-opportunity".

This study aim to apply the approach to the Northern Adriatic sea, using atmospheric data from Ancona, Venezia and Trieste weather stations. The study encompassed a range of experiments, each designed to test the feasibility of SST reconstruction at different spatial and temporal scales.

The capability of reconstructing sea temperature measured at AAot has been investigated, using different temporally processed data, hourly, low-pass filtered hourly and daily data, different architectures, Linear, Dense and a combination

of the two, and different input features, air temperatures only or additional wind and precipitation data. The different models' architecture and additional features didn't obtain significantly different results, while the effect of the temporal processing had the effect of reducing the mean RMSE on obtained in the cross validation from 1.25°C for hourly data, to 1.21°C for filtered data and 1.08°C for daily data, with an improvement of around the 13% from hourly to daily data, leading to the hypothesis that the large errors obtained in the reconstruction are partially due to the higher frequency fluctuations, that are not captured by the air temperatures. The quite large RMSE obtained also with daily data, led the the hypothesis that the temperature measured by AAot was driven mainly by advection process, instead of heat exchanges with the atmosphere.

To further investigate the hypothesis that advective processes influence the temperature measured by AAot, we employed spatially averaged CMEMS reanalysis data. A Dense model was used to reconstruct spatially averaged temperature targets, obtained by considering various combinations of vertical levels, time lags, and horizontal areas. The optimal reconstruction performance was achieved using a target temperature averaged over the vertical levels from 1m to 5.5m, with no time lag (0 days), and considering a horizontal area where the correlation coefficient between air and sea temperature exceeded 0.5 ($r > 0.5$). The RMSE for this specific target reconstruction was 0.48°C, which is 55% lower than the RMSE obtained when reconstructing the daily temperature data from AAot. These findings support the hypothesis that advective processes play a significant role in shaping the temperature measured by AAot. The spatial averaging of temperature data likely smooths out the local advective effects, making it easier for the model to capture the relationship between atmospheric variables and sea temperature.

As last experiment, the CMEMS reanalysis temperature at 1.0m has been reconstructed pointwise, using two different architectures: Dense model, and a Unet model. The performances of the two models have been analyzed through the distribution of absolute errors and the analysis of spatial and temporal distribution of MB, SDE, CC, uRMSE. The models obtained similar statistics, with a lower RMSE in the training and larger in the validation and test set. In this latter the Dense and Unet models obtained an RMSE of 0.74°C and of 0.71°C respectively. The spatial and temporal distribution of MB, SDE, CC and uRMSE presents the same characteristics between the two models, as also highlighted by the visualization of the reconstructed temperature fields. This evidenced that the added complexity of the Unet doesn't bring additional information. The time pattern of MB for both model is compared with the MB of the averaged spatial temperature reconstruction, evidencing the good consistence in the estimation of basin averaged temperature even by computing it from a pointwise

reconstruction.

As final evaluation, a comparison, on spatial and temporal performance of MB, SDE, CC and uRMSE, is made between satellite measured temperature, reanalysis and modeled temperature. This spatial analysis evidenced systematic differences also between the satellite and reanalysis temperature, although they are generally in agreement. The temporal analysis highlighted the presence of variable consistency through time, also between the reanalysis and satellite, with similar patterns of the neural networks and reanalysis differences.

In conclusion this study has highlighted some of the issues and strengths of the neural networks approach of reconstructing sea temperature from atmospheric data, proving the potential values of this approach in future applications. Indeed even if this method has several limitations we have seen that also already navigated methods of observing sea temperature have biases and differences between them. In addition this approach has room for several future investigations and improvements, one of which may be the inclusion of prior knowledge of the sea state at a certain time, like a cloud free satellite observation, from which the neural network starts to evolve the temperature field, by using only atmospheric data at present time.

In summary, despite its current limitations, the neural networks approach can add value to the ocean observing framework by providing an additional source of sea temperature estimates in a context where data scarcity remains a persistent challenge.

Bibliography

[1] R. H. Stewart, "Introduction to physical oceanography", 2008. [Online]. Available: https://hdl.handle.net/1969.1/160216.

[2] C. Deser, M. A. Alexander, S.-P. Xie, and A. S. Phillips, "Sea surface temperature variability: Patterns and mechanisms", *Annual Review of Marine Science*, vol. 2, pp. 115–143, 2010, ISSN: 1941-1405. DOI: 10.1146/annurev-marine-120408-151453. pmid: 21141660.

[3] A. G. O'Carroll *et al.*, "Observational Needs of Sea Surface Temperature", *Frontiers in Marine Science*, vol. 6, Aug. 20, 2019, ISSN: 2296-7745. DOI: 10.3389/fmars.2019.00420.

[4] S. L. Castro, G. A. Wick, and M. Steele, "Validation of satellite sea surface temperature analyses in the Beaufort Sea using UpTempO buoys", *Remote Sensing of Environment*, vol. 187, pp. 458–475, Dec. 15, 2016, ISSN: 0034-4257. DOI: 10.1016/j.rse.2016.10.035.

[5] J. Tintoré *et al.*, "Challenges for Sustained Observing and Forecasting Systems in the Mediterranean Sea", *Frontiers in Marine Science*, vol. 6, Sep. 13, 2019, ISSN: 2296-7745. DOI: 10.3389/fmars.2019.00568.

[6] A. Pisano *et al.*, "New Evidence of Mediterranean Climate Change and Variability from Sea Surface Temperature Observations", *Remote Sensing*, vol. 12, no. 1, p. 132, 1 Jan. 2020, ISSN: 2072-4292. DOI: 10.3390/rs12010132.

[7] A. Ricchi *et al.*, "Impact of the SST and topography on the development of a large-hail storm event, on the Adriatic Sea", *Atmospheric Research*, vol. 296, p. 107 078, Dec. 1, 2023, ISSN: 0169-8095. DOI: 10.1016/j.atmosres.2023.107078.

[8] B. Cushman-Roisin, M. Gačić, P.-M. Poulain, and A. Artegiani, Eds., *Physical Oceanography of the Adriatic Sea*. Dordrecht: Springer Netherlands, 2001, ISBN: 978-90-481-5921-5 978-94-015-9819-4. DOI: 10.1007/978-94-015-9819-4.

[9] P. Oddo and A. Guarnieri, "A study of the hydrographic conditions in the Adriatic Sea from numerical modelling and direct observations (2000–2008)", *Ocean Science*, vol. 7, no. 5, pp. 549–567, Sep. 6, 2011, ISSN: 1812-0784. DOI: 10.5194/os-7-549-2011.

[10] S. Malavolti *et al.*, "Distribution of Engraulis encrasicolus eggs and larvae in relation to coastal oceanographic conditions (the South-western Adriatic Sea case study)", *Mediterranean Marine Science*, vol. 19, no. 1, pp. 180–192, 1 May 23, 2018, ISSN: 1791-6763. DOI: 10.12681/mms.14402.

[11] A. Artegiani, E. Paschini, A. Russo, D. Bregant, F. Raicich, and N. Pinardi, "The Adriatic Sea General Circulation. Part I: Air–Sea Interactions and Water Mass Structure", Aug. 1, 1997, ISSN: 1520-0485. [Online]. Available: https://journals.ametsoc.org/view/journals/phoc/27/8/1520-0485_1997_027_1492_tasgcp_2.0.co_2.xml.

[12] A. Maggiore, M. Zavatarelli, M. G. Angelucci, and N. Pinardi, "Surface heat and water fluxes in the Adriatic Sea: Seasonal and internanual variability", *Physics and Chemistry of the Earth*, vol. 23, no. 5, pp. 561–567, Jan. 1, 1998, ISSN: 0079-1946. DOI: 10.1016/S0079-1946(98)00070-6.

[13] J. Chiggiato, M. Zavatarelli, S. Castellari, and M. Deserti, "Interannual variability of surface heat fluxes in the Adriatic Sea in the period 1998–2001 and comparison with observations", *Science of The Total Environment*, vol. 353, no. 1-3, pp. 89–102, Dec. 2005, ISSN: 00489697. DOI: 10.1016/j.scitotenv.2005.09.031.

[14] T. S. Hopkins, A. Artegiani, C. Kinder, and R. Pariante, "A discussion of the northern Adriatic circulation and flushing as determined from the ELNA hydrography", *The Adriatic Sea, Ecosystem Research Report No. 32, 1999, EUR 18834, European Commission, Brussels*, vol. 32, Jan. 1, 1999.

[15] E. C. Kent, J. J. Kennedy, D. I. Berry, and R. O. Smith, "Effects of instrumentation changes on sea surface temperature measured in situ", *WIREs Climate Change*, vol. 1, no. 5, pp. 718–728, 2010, ISSN: 1757-7799. DOI: 10.1002/wcc.55.

[16] P.-M. Poulain *et al.*, "MedArgo: A drifting profiler program in the Mediterranean Sea", *Ocean Science*, vol. 3, no. 3, pp. 379–395, Aug. 8, 2007, ISSN: 1812-0784. DOI: 10.5194/os-3-379-2007.

[17] E. Alvarez Fanjul *et al.*, *Implementing Operational Ocean Monitoring and Forecasting Systems*. IOC-UNESCO, GOOS-275, 2022, 392 pages. DOI: 10.48670/ETOOFS.

[18] C. J. Merchant *et al.*, "Global Sea Surface Temperature", in *Taking the Temperature of the Earth*, Elsevier, 2019, pp. 5–55, ISBN: 978-0-12-814458-9. DOI: 10.1016/B978-0-12-814458-9.00002-2.

[19] P. Minnett *et al.*, "Half a century of satellite remote sensing of sea-surface temperature", *Remote Sensing of Environment*, vol. 233, p. 111 366, Nov. 2019, ISSN: 00344257. DOI: 10.1016/j.rse.2019.111366.

[20] C. J. Merchant, A. R. Harris, E. Maturi, and S. Maccallum, "Probabilistic physically based cloud screening of satellite infrared imagery for operational sea surface temperature retrieval", *Quarterly Journal of the Royal Meteorological Society*, vol. 131, no. 611, pp. 2735–2755, 2005, ISSN: 1477-870X. DOI: 10.1256/qj.05.15.

[21] K. A. Kilpatrick, G. Podestá, E. Williams, S. Walsh, and P. J. Minnett, "Alternating Decision Trees for Cloud Masking in MODIS and VIIRS NASA Sea Surface Temperature Products", Mar. 1, 2019. DOI: 10.1175/JTECH-D-18-0103.1.

[22] C. M. Bishop and H. Bishop, *Deep Learning: Foundations and Concepts*. Cham: Springer International Publishing, 2024, ISBN: 978-3-031-45468-4. DOI: 10.1007/978-3-031-45468-4.

[23] M. A. Nielsen, *Neural Networks and Deep Learning*. Determination Press, 2015. [Online]. Available: http://neuralnetworksanddeeplearning.com.

[24] Q. Zhao, S. Peng, J. Wang, S. Li, Z. Hou, and G. Zhong, "Applications of deep learning in physical oceanography: A comprehensive review", *Frontiers in Marine Science*, vol. 11, Jul. 15, 2024, ISSN: 2296-7745. DOI: 10.3389/fmars.2024.1396322.

[25] A. Barth, A. Alvera-Azcárate, M. Licer, and J.-M. Beckers, "DINCAE 1.0: A convolutional neural network with error estimates to reconstruct sea surface temperature satellite observations", *Geoscientific Model Development*, vol. 13, no. 3, pp. 1609–1622, Mar. 27, 2020, ISSN: 1991-959X. DOI: 10.5194/gmd-13-1609-2020.

[26] A. Barth, A. Alvera-Azcárate, C. Troupin, and J.-M. Beckers, "DINCAE 2.0: Multivariate convolutional neural network with error estimates to reconstruct sea surface temperature satellite and altimetry observations", *Geoscientific Model Development*, vol. 15, no. 5, pp. 2183–2196, Mar. 15, 2022, ISSN: 1991-959X. DOI: 10.5194/gmd-15-2183-2022.

[27] A. Storto, G. De Magistris, S. Falchetti, and P. Oddo, "A Neural Network-Based Observation Operator for Coupled Ocean-Acoustic Variational Data Assimilation", *Monthly Weather Review*, Apr. 13, 2021, ISSN: 0027-0644, 1520-0493. DOI: 10.1175/MWR-D-20-0320.1.

[28] P. Sandroni, "Annali Idrologici - Parte prima", 2022.

[29] "Golfo di Trieste: Molo Fratelli Bandiera", www.ismar.cnr.it. [Online]. Available: https://www.ismar.cnr.it/web-content/cosa-facciamo/infrastrutture-oceanografiche/golfo-di-trieste-molo-fratelli-bandiera/.

[30] "Piattaforma Oceanografica Acqua Alta", www.ismar.cnr.it. [Online]. Available: https://www.ismar.cnr.it/web-content/cosa-facciamo/infrastrutture-oceanografiche/piattaforma-oceanografica-acqua-alta/.

[31] "Piattaforma oceanografica Acqua Alta | Consiglio Nazionale delle Ricerche". [Online]. Available: https://www.cnr.it/it/piattaforma-oceanografica-acqua-alta.

[32] R. Escudier *et al.*, *Mediterranean Sea Physical Reanalysis (CMEMS MED-Currents, E3R1 system): MEDSEA_MULTIYEAR_PHY_006_004*, version 1, Copernicus Monitoring Environment Marine Service (CMEMS), 2020. DOI: 10.25423/CMCC/MEDSEA_MULTIYEAR_PHY_006_004_E3R1.

[33] R. Escudier *et al.*, "A High Resolution Reanalysis for the Mediterranean Sea", *Frontiers in Earth Science*, vol. 9, Nov. 24, 2021, ISSN: 2296-6463. DOI: 10.3389/feart.2021.702285.

[34] S. Dobricic and N. Pinardi, "An oceanographic three-dimensional variational data assimilation scheme", *Ocean Modelling*, vol. 22, no. 3-4, pp. 89–105, Jan. 2008, ISSN: 14635003. DOI: 10.1016/j.ocemod.2008.01.004.

[35] European Union-Copernicus Marine Service, *Mediterranean Sea - High Resolution L4 Sea Surface Temperature Reprocessed*, Mercator Ocean International, 2015. DOI: 10.48670/MOI-00173.

[36] P. Oddo *et al.*, "Evaluation of different Maritime rapid environmental assessment procedures with a focus on acoustic performance", *The Journal of the Acoustical Society of America*, vol. 152, no. 5, pp. 2962–2981, Nov. 1, 2022, ISSN: 0001-4966, 1520-8524. DOI: 10.1121/10.0014805.