Modeling Earth Systems and Environment
https://doi.org/10.1007/s40808-024-02117-w

ORIGINAL ARTICLE

Unraveling climate trends in the mediterranean: a hybrid machine learning and statistical approach

Mutaz AlShafeey[1]

Received: 30 June 2024 / Accepted: 25 July 2024

Abstract

This study presents a comprehensive spatiotemporal analysis of sea surface temperatures (SST) and surface air temperatures (TAS) across 15 Mediterranean coastal stations, leveraging centennial-scale data to analyze regional climate dynamics. The modeling framework integrates three sequential phases: data preprocessing, statistical analysis, and advanced machine learning techniques, creating a robust analytical pipeline. The data preprocessing phase harmonizes diverse datasets, addresses missing values, and applies transformations to ensure analytical consistency. The statistical modeling employs the Pettitt test for change point detection and linear trend analysis to unveil underlying patterns. The machine learning phase utilizes K-means clustering for climate regime classification and implements tailored Convolutional Neural Networks (CNNs) for cluster-specific future climate anomaly projections. Results unveil a marked anthropogenic climate signal, with contemporary observations consistently surpassing historical baselines. Breakpoint analyses and linear trend assessments reveal heterogeneous climatic shifts, with pronounced warming in the northern Mediterranean. Notably, Nice and Ajaccio exhibit the highest SST increases (0.0119 and 0.0113 °C/decade, respectively), contrasting with more modest trends in Alexandria (0.0052 °C/decade) and Antalya (0.0047 °C/decade) in the eastern Mediterranean. The application of clustering and CNN projections provides granular insights into differential warming trajectories. By 2050, cooler northwestern Mediterranean zones are projected to experience dramatic SST anomalies of approximately 3 °C above the average, with corresponding TAS increases of 2.5 °C. In contrast, warmer eastern and southern regions display more subdued warming patterns, with projected SST and TAS increases of 1.5–2.5 °C by mid-century. This research's importance is highlighted by its potential to inform tailored adaptation strategies and contribute to the theoretical understanding of climate dynamics, advancing climate modeling and analysis efforts.

Keywords Mediterranean climate change · Sea surface temperature · Surface air temperatures · Machine learning in climate analysis · Climate modeling

Introduction

In the realm of climate science, the Mediterranean region is a very well-known area due to its combination of subtropical arid and temperate environments, stemming from its unique location. However, this positioning makes it especially vulnerable to the effects of climate change, emphasizing its importance as a key climate change hotspot (Cramer et al. 2018; Lionello and Scarascia 2018). This region, which includes a semi-enclosed sea and its surrounding lands, has experienced significant changes in climate patterns in the last few decades. These changes are supported by evidence in the Mediterranean region, manifested in altered temperature trends, shifting precipitation regimes, and an uptick in the frequency and severity of extreme weather events (Insua-Costa et al. 2022; Androulidakis et al. 2023). The consequences of these climate changes go far beyond meteorological issues, severely affecting crucial socioeconomic sectors. Agriculture, water resource management, coastal zone planning, and the protection of fragile ecosystems all face considerable issues due to these changes (Raihan 2023; Noto et al. 2023). As a result, there is an urgent need to understand the intricate patterns of climatic

✉ Mutaz AlShafeey
mutaz.alshafeey@uni-corvinus.hu

[1] Institute of Data Analytics and Information Systems, Corvinus University of Budapest, Fővám tér 13-15, Budapest H-1093, Hungary

Published online: 06 August 2024

variability and predict their future trajectories. This understanding is critical for creating strong adaptation strategies and making informed policy decisions to limit the effects of climate change in this highly reactive region (Mastrorillo et al. 2024).

Recent data and investigations add a new engine to the urgency of addressing climate change in this region, revealing an alarming trend of accelerated warming. As a shred of evidence, many studies indicate that the temperature rise in this area surpasses the global average by approximately 20% (Lionello and Scarascia 2018), highlighting a critical issue demanding immediate attention. This accelerated warming stems from a complex interplay of factors, including anthropogenic activities such as deforestation and industrial emissions, as well as natural fluctuations in atmospheric and oceanic circulation patterns (Urdiales-Flores et al. 2023; Longobardi et al. 2016). The consequences of this rapid warming are particularly significant for two crucial climatic indices: Sea Surface Temperature (SST) and Temperature at Surface Air (TAS). These metrics serve as vital indicators of terrestrial and marine ecosystem health and play key roles in climate modeling and adaptation strategies (Karmalkar and Bradley 2017; Tanaka and Van Houtan 2022). Changes in SST can have a cascade effect on marine biodiversity, including coral bleaching and shifts in fish populations (Duvat et al. 2021). Concurrently, changes in TAS might cause more frequent and severe heat waves, affecting human health, agriculture, and natural ecosystems (Celik 2020). Furthermore, SST and TAS are fundamentally connected to the hydrological cycle, impacting precipitation patterns and hence freshwater availability (Benestad et al. 2022). This interrelationship highlights the far-reaching consequences of temperature fluctuations for numerous areas of the Mediterranean ecosystem and its inhabitants.

The precise monitoring and analysis of temperature patterns is the foundation of climate science and policy development. Quantifying changes in SST and TAS is critical for understanding and mitigating the effects of climate change, both globally and in specific regions such as the Mediterranean (Guinaldo et al. 2023; Pisano et al. 2020). Precise monitoring of SST changes is critical for forecasting shifts in marine ecosystems, sea level rise, and extreme weather events (O'carroll et al. 2019). Correspondingly, TAS measurements offer critical insights into terrestrial climate dynamics, including the frequency and intensity of heat waves, drought patterns, and biodiversity changes (Alomar et al. 2022). The integration of both SST and TAS data enables researchers to develop a comprehensive understanding of the complex feedback mechanisms between oceanic and atmospheric systems. This holistic approach facilitates more accurate climate modeling and supports informed decision-making processes. By analyzing these interconnected temperature trends, scientists and policymakers can better assess the full scope of climate change impacts and devise more effective mitigation and adaptation strategies for vulnerable regions like the Mediterranean.

In the Mediterranean context, quantifying SST and TAS changes is particularly critical due to the region's high responsiveness to climate change consequences (Patti et al. 2022). The Mediterranean Sea's semi-enclosed nature, along with its complex bathymetry and intricate coastline, make it particularly vulnerable to localized temperature changes and marine heatwaves. These phenomena profoundly impact marine ecosystems, tourism industries, and coastal communities (Martínez et al. 2023). Furthermore, the Mediterranean region's diverse topography, encompassing coastal zones, inland areas, and mountainous terrain, necessitates comprehensive monitoring of TAS changes. This wide-ranging analysis is crucial for assessing potential risks to key sectors such as agriculture and water resources, as well as evaluating threats to public health. By accurately quantifying and analyzing SST and TAS trends, policymakers and stakeholders can develop tailored adaptation strategies and targeted mitigation measures. These evidence-based approaches are essential for protecting the Mediterranean's unique ecosystems and safeguarding its diverse economic sectors (Bonaldo et al. 2023). The precise measurement of these temperature indicators thus serves as a fundamental tool in crafting resilient and sustainable policies for this climatically sensitive region.

Climate change analysis requires sophisticated methodologies to accurately detect and interpret long-term environmental shifts. The quantification of climate variable shifts typically involves a comprehensive analysis of long-term datasets (Duan et al. 2022). Historically, researchers utilized a variety of statistical methods to examine climate patterns (Chandler and Scott 2011). However, many studies have failed to identify breakpoints at which large changes in climatic patterns become apparent. While global temperature trends have often been the focus of climate research, these analyses frequently report only on general increases in mean temperatures over defined historical periods. This approach neglects the nuanced examination of underlying trends and overlooks critical aspects such as the quasi-periodic heterogeneity in variance observed in temperature data (Duan et al. 2022). Nonetheless, employing breakpoint analysis is crucial as it identifies specific years or periods marking significant changes in climatic trends. This methodology not only enhances our understanding of the timing and dynamics of climate variability but also provides essential insights for predicting future climate conditions in the Mediterranean region and beyond. This modified focus on temporal shifts is critical for various reasons. It promotes a better understanding of previous interactions within the climate

system, improves the accuracy of future climate forecasts, and gives a more comprehensive picture of climate development, allowing for more focused mitigation and adaptation methods. This technique allows researchers to move beyond broad generalizations and establish a more detailed knowledge of climate change patterns, boosting the effectiveness of climate models and policy recommendations.

The field of climate modeling is undergoing a transformative shift, driven by the integration of cutting-edge computational methods. Advanced computational approaches, particularly machine learning, are becoming increasingly important in climate research (Huntingford et al. 2019; Beucler et al. 2024). Sophisticated algorithms like K-means clustering and Convolutional Neural Networks (CNNs) excel at modeling complex, non-linear patterns that traditional statistical methods fail to fully capture (Han et al. 2024). These powerful tools can process vast datasets to forecast future changes with high accuracy, providing researchers with invaluable means to simulate and comprehend potential climate scenarios under various models. This evolution in methodology represents a pivotal shift towards a more dynamic and predictive approach to climate science. By leveraging machine learning techniques, researchers can uncover subtle patterns in climate data, generate more accurate long-term predictions, and develop comprehensive simulations of future climate states. The integration of these advanced computational methods enables scientists to formulate more precise adaptation and mitigation strategies, allowing policymakers and stakeholders to make better-informed decisions to address the challenges posed by climate change.

The Mediterranean basin has been extensively studied for temperature trends and climate change impacts, with research employing diverse methodological approaches and data sources. Studies have varied in scope, focusing on specific sub-regions like the western Mediterranean (Schroeder et al. 2016; Méndez-Cea et al. 2023; Insua-Costa et al. 2022), the eastern Mediterranean (Zittis et al. 2022; Hochman et al. 2022), or the Adriatic Sea (Scarponi et al. 2022; Bonacci et al. 2021). Others have examined the entire basin but with a particular emphasis on SST trends (Pisano et al. 2020; Pastor et al. 2020) or TAS patterns (García-Monteiro et al. 2022; Seker and Gumus 2022). Despite providing valuable insights, many of these studies rely on limited methodological approaches such as linear trend analyses or basic statistical techniques, which often fail to capture the full complexity of climatic processes and inherent non-linearities (Corduas 1994). To address these limitations, a growing number of researchers have advocated for the use of machine learning algorithms in climate trend analysis (Fahad et al. 2023; Zhu et al. 2023). These approaches have the potential to capture complex non-linear patterns, handle high-dimensional data, and provide more nuanced projections of future climate trajectories Climate-invariant machine learning (Beucler et al. 2024).

Addressing gaps in the existing literature, this study aims to provide a comprehensive modeling of SST and TAS trends and climate variability across the Mediterranean region. By harmonizing observational data from the HadISST and CRUTEM5 datasets, spanning over a century of observations from 15 coastal stations, holistic modeling of the complex regional climate dynamics is offered. The comprehensive modeling and analysis accounts for spatial heterogeneity through clustering, addressing the limitations of previous studies. To achieve this, a three-phase modeling methodology is employed. The first phase involves data preprocessing to establish a clean, structured dataset suitable for in-depth analysis. Statistical techniques in the second phase, including the Pettitt test and linear trend analysis, enable the detection of abrupt changes and the deciphering of underlying patterns in the data. The final phase leverages advanced machine learning methods, with K-Means clustering accounting for spatial heterogeneity and CCNs tailored for cluster-specific projections. This integrative approach combines machine learning capabilities with spatial considerations, facilitating robust climate modeling within the Mediterranean context.

The novelty of this study is derived from three distinct aspects that set it apart from previous research efforts. Firstly, the study's scope encompasses a network of 15 coastal stations distributed throughout the Mediterranean basin, providing a holistic perspective on the spatial variability of climate dynamics within this region. Additionally, the integration of both sea SST and TAS datasets in the models offers an opportunity to explore the intricate relationships between marine and terrestrial climate systems, a facet often overlooked in existing studies. Secondly, the modeling framework employed in this study is innovative, as it seamlessly integrates statistical techniques with advanced machine learning algorithms. While most studies typically employ either statistical or machine learning methods in isolation, the utilized approach represents an extremely rare hybrid, synergistically combining these two methodologies. The incorporation of CNNs alongside statistical modeling pushes the boundaries of climate trend analysis, enabling the capture of complex non-linear patterns and projections that traditional methods often fail to detect. Thirdly, the implementation of the Pettitt test addresses a critical gap in climate research. Many studies frequently report only on general increases in mean temperatures over defined historical periods, neglecting the nuanced examination of underlying trends and the quasi-periodic heterogeneity in variance observed in temperature data. By explicitly employing the Pettitt test, this study aims to detect abrupt changes and

shifts in the SST and TAS data, providing a comprehensive understanding of climate dynamics that transcends the limitations of conventional approaches.

The importance of this study is multifaceted, as it addresses a critical knowledge gap in understanding the heterogeneous nature of climate change impacts within the Mediterranean region. By elucidating the spatial and temporal variability of temperature trends, the findings provide a solid foundation for developing tailored adaptation strategies and informing decision-makers on region-specific climate challenges. Furthermore, the insights gained from this research contribute to advancing the theoretical understanding of the complex interplay between local environmental factors and large-scale climatic processes, fostering further scientific inquiry and refinement of climate models.

Methods

The methodology utilized in this study, as depicted in Fig. 1, is organized into three interconnected phases to comprehensively model climate trends. The first phase involves data collection and preprocessing, where raw data from various sources is harmonized, missing values are addressed, and data is transformed to create a clean, structured dataset. In the second phase, this processed data undergoes statistical analysis using the Pettitt test to identify abrupt change points and linear trend analysis to uncover underlying patterns. The final phase employs machine learning techniques, specifically K-means clustering to categorize the data into distinct climatic groups, and Convolutional Neural Networks (CNNs) to predict future climate anomalies for each cluster. This phased approach ensures a robust modeling process, with the CNNs providing precise, cluster-specific

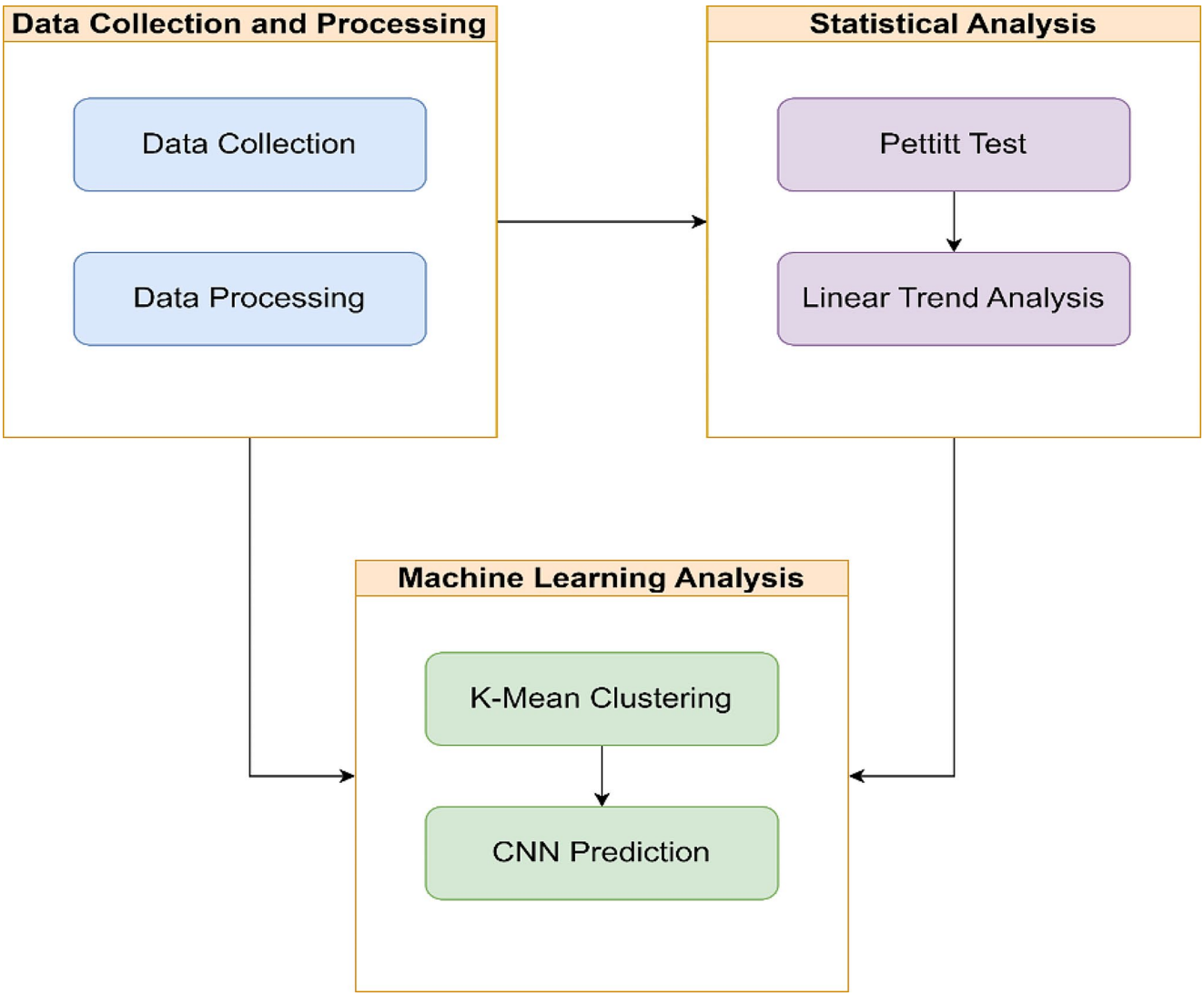


Fig. 1 Schematic overview of the hybrid statistical-machine learning analysis methodology

predictions. All methods, graphs, maps, and equations were implemented using the Python Jupyter environment, ensuring clarity and reproducibility.

Data collection and preprocessing

The initial phase of the study involves systematic data collection, where datasets are sourced and compiled based on predefined criteria relevant to the research objectives. Upon acquisition, the data undergoes rigorous preprocessing, which includes cleaning (removal of missing values), normalization, and transformation operations to ensure consistency and suitability for subsequent analysis.

In this paper, the utility and reliability of two datasets in the domain of climate research are explored: The Hadley Centre Sea Ice and Sea Surface Temperature (SST) dataset (HadISST) and the Climatic Research Unit Temperature dataset (CRUTEM5) which provides surface air temperature (TAS). The HadISST dataset, compiled and maintained by the UK Met Office, provides an exhaustive monthly record of SST and sea ice concentrations dating back to the year 1870 (Rayner et al. 2003). It amalgamates in situ measurements from the Met Office Marine Data Bank (MDB) with satellite observations, ensuring a detailed depiction of global oceanic conditions across a 1-degree latitude-longitude grid.

Simultaneously, the CRUTEM5 dataset is a critical resource for analyzing near-surface air temperatures over land, with records spanning from 1850 to the present (Osborn et al. 2021). This dataset, a collaborative effort between the Climatic Research Unit at the University of East Anglia and the Met Office Hadley Centre, is distinguished by its robust spatial resolution and global coverage. It aggregates temperature records from a multitude of meteorological stations, providing monthly mean temperatures with a spatial resolution refined to 5° by 5° latitude and longitude. The dataset facilitates a granular investigation into regional temperature fluctuations.

In this research, SST data for the entire Mediterranean Sea were initially collected using a 1°x1° latitude-longitude grid covering its surface area. Subsequently, TAS data were gathered for all available stations within the Mediterranean region to compile a comprehensive dataset from various coastal stations. The extraction process began with the identification of all relevant Mediterranean coastal stations from the CRUTEM5 dataset. However, the selection of stations was constrained by data availability. Many stations, particularly in the southern and eastern Mediterranean, had significant gaps or outdated records with more than 10 consecutive missing or outdated data points. After thoroughly evaluating data availability and the presence of missing values, only 15 stations were deemed suitable for inclusion due to the lack of reliable and continuous data from other stations. For these 15 stations, the nearest available SST data from the Mediterranean Sea grid were subsequently extracted from the HadISST dataset, ensuring consistency between the two types of climate observations. This approach ensured that the selected stations had the most reliable and continuous data.

Figure 2 shows the final fifteen coastal Mediterranean stations selected after the thorough evaluations of data availability. Although SST data is available from 1870, TAS data for most of these stations is available only from January 16, 1961, to April 16, 2023. Consequently, the dataset used for the statistical and machine learning phases aligns with the temporal range of the TAS data, starting from January 16, 1961.

Statistical analysis

After preprocessing, the data underwent a thorough exploratory statistical analysis. Initially, the Pettitt test, known for its robustness as a non-parametric method, was applied to identify change points within the time series data of TAS and SST. This test pinpointed the break years, marking significant shifts in TAS and SST values. Subsequently, linear trend analysis was conducted to delineate potential linear relationships in the data. This step was crucial to discern the trends preceding and succeeding the detected breakpoints, thereby providing insights into long-term trends. Such a statistical approach is instrumental in revealing the data's structural intricacies and in detecting shifts in both mean and variance over the timeline.

Pettitt test

The Pettitt test is a robust, non-parametric method designed to detect a significant shift in the mean of a time series when the precise timing of the change is unknown. For a series of observed data points $x_1, x_2, x_3, \ldots, x_n$, the test hypothesizes a change point at time t, such that the portion of $x_1, x_2, x_3, \ldots, x_t$ is characterized by a cumulative distribution function $F_1(x)$, distinct from $F_2(x)$, which describes the distribution of the subsequent segment $x_{t+1}, x_{t+2}, x_{t+3}, \ldots, x_n$ (de Assis PAIVA and SÁFADI 2021).

The core of Pettitt's test lies in the computation of the non-parametric test statistic U_t (critical value used to assess the presence and significance of the change point), which involves assessing the differences between all pairs of observations in the series. Specifically, for each position k in the series, the test calculates $U(k)$, the sum of signs of differences between the $k-th$ observation and every other observation, as defined by the sgn function $sgn(x_i - x_j)$

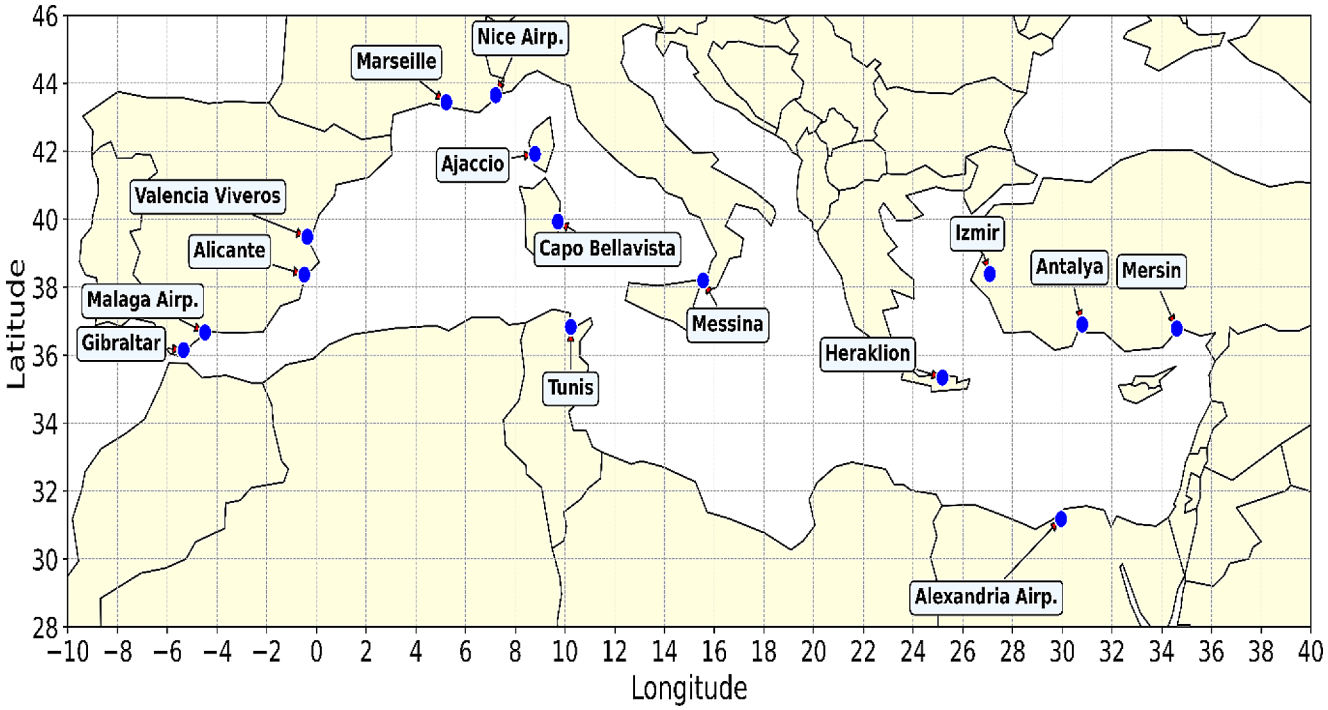


Fig. 2 Selected coastal stations in the Mediterranean

(Pettitt 1979). The sgn function is critical for this computation, returning:

$$sgn\,(x) = \begin{cases} -1 & if\ x < 0, \\ 0 & if\ x = 0, \\ 1 & if\ x > 0, \end{cases} \tag{1}$$

Which discerns the directionality of the difference between any two data points. The test statistic $U\,(k)$ is formulated as in 2 (Pettitt 1979):

$$U\,(k) = \sum\nolimits_{i=1}^{k} \sum\nolimits_{j=k+1}^{n} sgn(x_i - x_j) \tag{2}$$

The cumulative sums $U\,(k)$ are evaluated (as in 3) to identify the maximum $U\,(t)$, the value of which suggests a change point's presence.

$$U\,(t) = \max |U\,(k)| \ \ for\ \ k = 1,\ 2,\ 3, \ldots,\ n \tag{3}$$

The location K of the potential change point is pinpointed by the index k at which this maximum $U\,(t)$ is observed.

$$K\ =\ arg\,max_k\,|\,U\,(k)| \tag{4}$$

Here, $argmax$ represents the 'argument of the maximum', a mathematical operation used to identify the index k at which the function $|\ U\,(k)\ |$ reaches its maximum value.

To ascertain the statistical significance of the change detected at K, a p-value is computed based on the distribution of $U\,(t)$ under the null hypothesis (no change) as in 5 (Zhang and Wu 2023):

$$p - value = 2 * \left(1 - \varphi\ \left(\frac{U_t}{\sqrt{6n}}\right)\right) \tag{5}$$

Where φ signifies the standard normal cumulative distribution function (CDF). This $p - value$ quantifies the likelihood that the observed change at K is not due to random variation, with lower values, indicating a higher statistical significance of the detected change point (Rybski and Neumann 2011).

Linear trend analysis

Following the identification of significant change points within the time series data for SST and TAS via the Pettitt test, a detailed linear trend analysis was conducted to quantitatively assess the shifts in trends across these breakpoints, i.e. evaluate the changes in SST and TAS trends before and after the detected breakpoints.

For each interval and variable, a linear regression model was fitted to delineate the trend as in 6.

$$y = mx + b \tag{6}$$

Where y represents the variable under consideration (in this study, SST or TAS), m signifies the trend's slope, and

b denotes the intercept. For observations preceding the breakpoint, a linear regression was applied to calculate the slope (m_{before}) and intercept (b_{before}), thereby elucidating the trend preceding the change. Similarly, for observations after the breakpoint, another linear regression provided the slope (m_{after}) and intercept (b_{after}), highlighting the trend following the change.

Machine learning analysis

Upon establishing the foundational statistics of the datasets, the machine learning phase took place. The processed data is partitioned into distinct clusters using the K-Mean Clustering algorithm. This unsupervised learning method groups the data into k clusters, each represented by the centroid of the points belonging to the cluster. The optimal number of clusters is determined based on the evaluation of cluster validity indices such as the elbow method. Each cluster identified by the K-Mean algorithm is then independently analyzed using a CNN. The CNN architecture is designed to capture the complex patterns and features within each cluster. It comprises multiple layers, including convolutional layers for feature extraction and fully connected layers for prediction. The CNN is trained, validated, and tested using cluster-specific datasets to ensure that the predictive model is attuned to the nuances of each cluster's characteristics. This approach allows for customized predictions that account for the unique properties identified within each cluster, thereby enhancing the predictive performance of the model.

K-Mean algorithm

In the preliminary phase of the clustering analysis, a critical step involved determining the optimal number of clusters that accurately reflects the underlying structure of our dataset while maintaining a balance between model simplicity and clustering effectiveness.

The investigation commenced with the establishment of a candidate range for the number of clusters (2–7). This range was chosen to encompass a broad spectrum of potential cluster structures, from oversimplified to excessively granular groupings. The core of this method relies on the concept of inertia, a metric that quantifies the compactness of clusters formed by the K-Means algorithm. Inertia is defined as the total sum of squared distances between each point in a cluster and the cluster's centroid (Nan et al. 2022). Mathematically, for each potential cluster number f within the predefined range, inertia (u) is calculated as in 7:

$$u = \sum {}_{i=1}^{n} min_{\mu_j} \in C \left(\|x_i - \mu_j\|^2 \right) \quad (7)$$

Where $_n$ represents the number of observations in the dataset, $_{x_i}$ is an individual observation, μ_j denotes the centroid of the cluster C_j, and $?x_i - \mu_j?^2$ is the squared Euclidean distance between $_{x_i}$ and μ_j. For each candidate cluster number (f), the K-Means algorithm partitions the dataset into f clusters. Following the computation of inertia for each clustering configuration, these inertia values are analyzed. The optimal number of clusters is inferred from the plot of inertia values against the number of clusters, applying the "elbow method". The "elbow" point represents a threshold beyond which increases in the number of clusters result in diminishing reductions in inertia. This point signifies an optimal balance, indicating that additional clusters do not substantially enhance the model's ability to capture more detailed cluster structures. In this study, Fig. 3 presents the inertia graph from which the optimal number of clusters was identified. The analysis revealed an elbow at 4 clusters, which led to determining that setting the number of clusters to 4 was most appropriate for the analysis.

In conducting the K-Means clustering analysis, variables obtained (for each location) from Phase 1: Data Collection and Processing, and Phase 2: Statistical Analysis, were integrated. Variables from Phase 1 encompassed geographical coordinates, specifically Latitude and Longitude, providing the spatial context essential for the analysis. Phase 2 contributed variables that encapsulate significant statistical insights into SST and TAS, including breakpoints and trends pre- and post-breakpoint, offering a temporal analysis lens. The precise variables incorporated, along with their phase of derivation, are cataloged in Table 1.

Upon defining the variables, the K-Means algorithm was tasked with partitioning the dataset into four clusters (see Fig. 3). The algorithm's objective, minimizing the within-cluster sum of squares (WCSS), is mathematically encapsulated as in Eq. 8, which emphasizes the algorithm's iterative refinement of cluster centroids (μ_j) to achieve a minimized WCSS, effectively grouping data points ($_x$) into cohesive clusters based on the specified variables (Ahmed et al. 2020).

$$WCSS = \sum {}_{i=1}^{k} \sum {}_{x \in C_i} \left(\|x - \mu_j\|^2 \right) \quad (8)$$

In the analysis, for each identified cluster, the approach involved aggregating data across all locations within the cluster to calculate a collective average, effectively representing the cluster's overall characteristics.

1-Dimensional convolutional neural networks (1D CNNs)

Initially, the series data (SST and TAS) was aggregated from a monthly to an annual scale by calculating the mean for

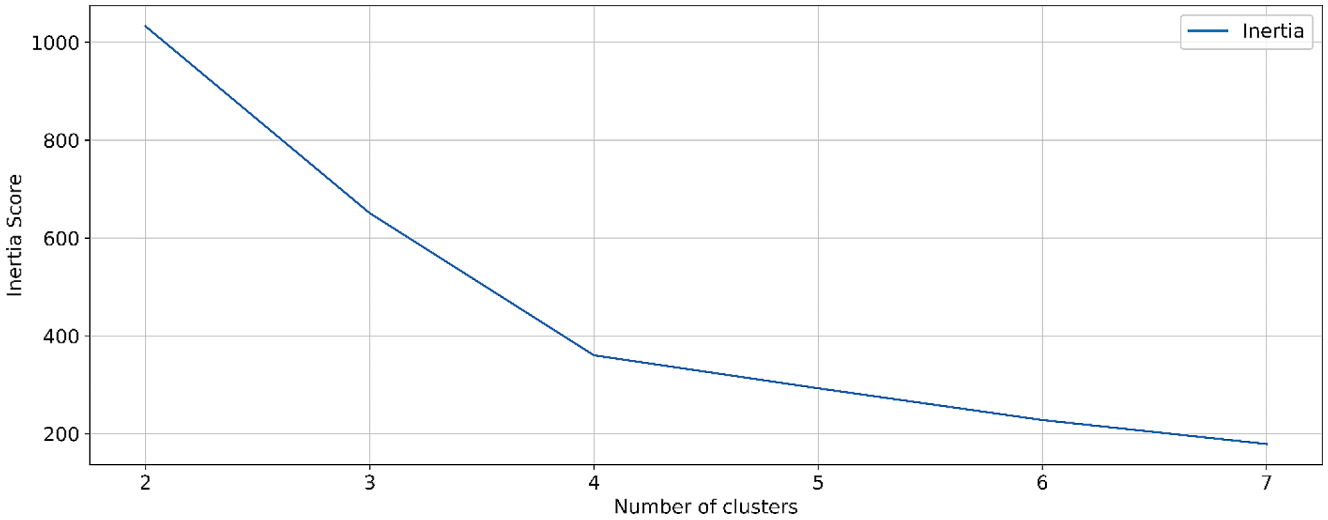


Fig. 3 Inertia score

Table 1 Variables incorporated into K-means clustering analysis and their phase of derivation

Variable	Derived From Phase
Latitude	Phase 1. Data Collection and Processing
Longitude	Phase 1. Data Collection and Processing
Breakpoint for SST	Phase 2. Statistical Analysis
Breakpoint for TAS	Phase 2. Statistical Analysis
SST trend before Breakpoint	Phase 2. Statistical Analysis
SST trend after Breakpoint	Phase 2. Statistical Analysis
TAS trend before Breakpoint	Phase 2. Statistical Analysis
TAS trend after Breakpoint	Phase 2. Statistical Analysis

each year where $_{x\,avr}$ is the average target variable (SST then TAS) for each year as in 9:

$$Annual\ data_{x\,avr\ } = \frac{1}{12}\sum {}_{i=1}^{12} x_{month\,i} \tag{9}$$

Next, deviations from the average SST and TAS up to the year 2000 (as this study focuses on predicting the anomaly of these variables compared to 2000) were calculated. This baseline average is given by *10*. Note that $_{n}$ is the number of years up to 2000:

$$Average\ till\ 2000\ = \frac{1}{n}\sum x\ avr\ {}_{year \leq\ 2000} \tag{10}$$

The objective of this study is to analyze the anomalies in SST and TAS, hence it is essential to calculate deviations from a historical baseline (2000). The deviations for each year were then calculated as:

$$x\ avr\ dev = \ x\ avr - \ \ Average\ till\ 2000 \tag{11}$$

These deviations were normalized using MinMaxScaler to ensure that the input values fit well into the neural network model:

$$Scaled\ x\ avr\ dev = \frac{x\ avr\ dev - \min\left(x\ avr\ dev\right)}{\max\left(x\ avr\ dev\right) - \min\left(x\ avr\ dev\right)} \tag{12}$$

The scaled TAS and SST deviations were organized into sequences of 10 years to facilitate the modeling of dependencies over decades. The CNN model architecture was defined with the flexibility to test various configurations through hyperparameter tuning. The convolutional layer of the model is expressed as:

$$C_l\left(x\right) = ReLU\left(Conv1D\left(x;\ filters_l,\ kernel\ size_l\right)\right) \tag{13}$$

Where $C_l\left(x\right)$ is the convolution operation on layer l, $filters_l$ are the number of filters in the convolution layer l, and $kernel\ size_l$ is the size of the kernel in the convolution layer l. ReLU (Rectified Linear Unit) serves as the activation function, providing non-linear capabilities to the model.

Following each convolution operation, a MaxPooling layer reduces the dimensionality of the data, summarizing the most significant features:

$$P_l\left(x\right) = MaxPooling1D\left(x\right) \tag{14}$$

Where P_l represents the max pooling operation applied in the $l-$ th layer.

Models were compiled with the Adam optimizer, which is known for its efficiency in large datasets and sparse

gradients. The learning rate ($_{\eta}$) was tuned to optimize the convergence. Equation 15 was utilized aiming to minimize the mean squared error (MSE). where N is the total number of samples, $\widehat{y}_i$ are the predicted values, $_{y_i}$ and are the actual values.

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - \widehat{y}_i)^2 \tag{15}$$

Hyperparameter optimization was executed using the Hyperband algorithm, a method based on the multi-armed bandit problem that dynamically allocates resources to configurations that show promise. This efficient strategy minimizes the validation MSE, optimizing model parameters for best performance. Table 2 summarizes the key parameters of the CNN used in the study, providing explanations for each parameter and the range of values tested during hyperparameter tuning. Note that for the CNN models, the data was preprocessed by splitting it into training and test sets following an $80-20$ scheme. This approach ensures that the models are trained on a substantial portion of the available data (80%). The remaining (20%) of the data is set aside as an independent test set, providing an unbiased evaluation of the model's generalization capabilities on unseen data.

Results

The results of this article systematically reflect the structured methodology delineated in Fig. 1, which is organized into three sequential phases. The initial findings begin with the "Dataset Overview and Temporal Trends" where the focus lies on examining the broad patterns and temporal variations observed in the SST and TAS data, spotting how these variables change over time. The analysis then progresses to "Statistical analyses findings". In this phase, the Pettitt test identifies significant change points, while linear trend analysis elucidates the underlying patterns and trends across the data. Finally, the "Machine learning findings" section presents the outcomes of employing advanced machine learning strategies. K-Mean clustering is used to segment the data into clusters, while CNN is deployed to make predictions specific to each cluster.

Dataset overview and temporal trends

The Mediterranean Sea has undergone significant changes in SST over the past century and a half. The analysis of SST data spanning from January 16, 1870, to April 16, 2023, reveals distinct spatial and temporal patterns across the basin. Figure 4 illustrates the spatial distribution of average SST across the Mediterranean region for this extended period. A clear temperature gradient is evident, with the coastal waters along southern France, the Alboran Sea, and

Table 2 Hyperparameters of the 1D Convolutional Neural Network and their ranges for optimization

Parameter	Explanation	Tested Parameters
Number of layers	The number of 1D layers in the CNN. This affects the depth of the network, influencing the ability to learn and represent complex patterns in the data.	1 to 5 layers
Filters	The number of filters in each convolutional layer. Filters capture different features of the input data, with more filters allowing the network to capture a broader range of features.	32 to 264
Kernel size	The size of the kernel in each convolutional layer. The kernel size determines the extent of the window over which the network aggregates information, affecting the granularity of the features extracted.	2 to 10, incremented by 2
Learning rate	The step size at each iteration while moving toward a minimum of a loss function. A suitable learning rate ensures convergence to a minimum during training.	0.0001 to 0.01, logarithmic sampling
Activation	The activation function used in the convolutional layers. ReLU is commonly used for its efficiency in non-linear transformations and avoiding the vanishing gradient problem.	ReLU
Pooling type	The type of pooling layer utilized after each convolutional layer, typically used to reduce the spatial dimensions of the input volume for the next layer.	MaxPooling1D, pool size of 2
Optimizer	The method used to update weights in the network. Adam optimizer is utilized for its adaptive learning rate capabilities, which helps in converging faster.	Adam
Loss function	The function used to measure the model's prediction error. MSE (Mean Squared Error) quantifies the difference between the predicted and actual values.	MSE

parts of the Balearic Sea exhibiting cooler SSTs (16–18 °C). In contrast, the eastern Mediterranean, including the Levantine Basin and coasts of Egypt and southern Turkey, displays warmer SSTs (21–22 °C). This pattern aligns with known oceanographic characteristics: the eastern basin retains heat due to its enclosed nature and warm water inflows from the southeast, while the western basin is moderated by cooler Atlantic inflow through the Strait of Gibraltar.

The temporal evolution of SST is depicted in Fig. 5, which presents the historical progression of monthly SST from 1870 to 2023. A distinct long-term warming trend is evident, with recent observations frequently exceeding 21 °C, a threshold rarely crossed in earlier records. Prior to the 1980s, monthly SSTs often fell below 12 °C, but such low temperatures have become increasingly rare in recent decades. This shift indicates a fundamental change in the Mediterranean Sea's thermal regime, with warming accelerating notably since the late 20th century.

Table 3 provides a detailed overview of SST data for specific coastal locations across the Mediterranean. Mersin, Turkey, exhibits the highest mean SST (22.4 °C) and maximum recorded SST (30.4 °C), indicative of the warmer conditions in the eastern basin. Conversely, Marseille, France, in the western Mediterranean, shows one of the lower mean SSTs (17.3 °C). The range of SST variability differs markedly between locations, from a minimum of 8.9 °C in Gibraltar to a maximum of 15.5 °C in Messina, Italy. Skewness values are predominantly positive, suggesting a slight bias towards warmer SSTs in most locations. Negative kurtosis values across all sites indicate flatter distributions compared to normal, implying fewer extreme SST events.

Surface air temperature (TAS) data for coastal Mediterranean stations, presented in Table 4, further illustrates the region's climatic diversity. Alexandria, Egypt, records the highest mean (20.5 °C), while Marseille, France, shows the lowest (15.0 °C). Extreme temperatures range from a maximum of 31.0 °C in Mersin, Turkey, to a minimum of 2.0 °C in Marseille. The greatest TAS range is observed in Marseille (26.2 °C), indicating substantial seasonal variability, while Gibraltar exhibits the smallest range (14.2 °C), suggesting a more moderate climate.

Statistical analyses findings

Statistical modeling was employed to discern and elucidate the Mediterranean's climate trends, revealing specific temporal and spatial variations in both SST and TAS. Table 5 presents the key outputs of this modeling stage, providing an in-depth comparative analysis of the temporal shifts in SST and TAS across the studied stations.

The model identifies breakpoint years, where a notable change in the temperature trend is observed, and quantifies

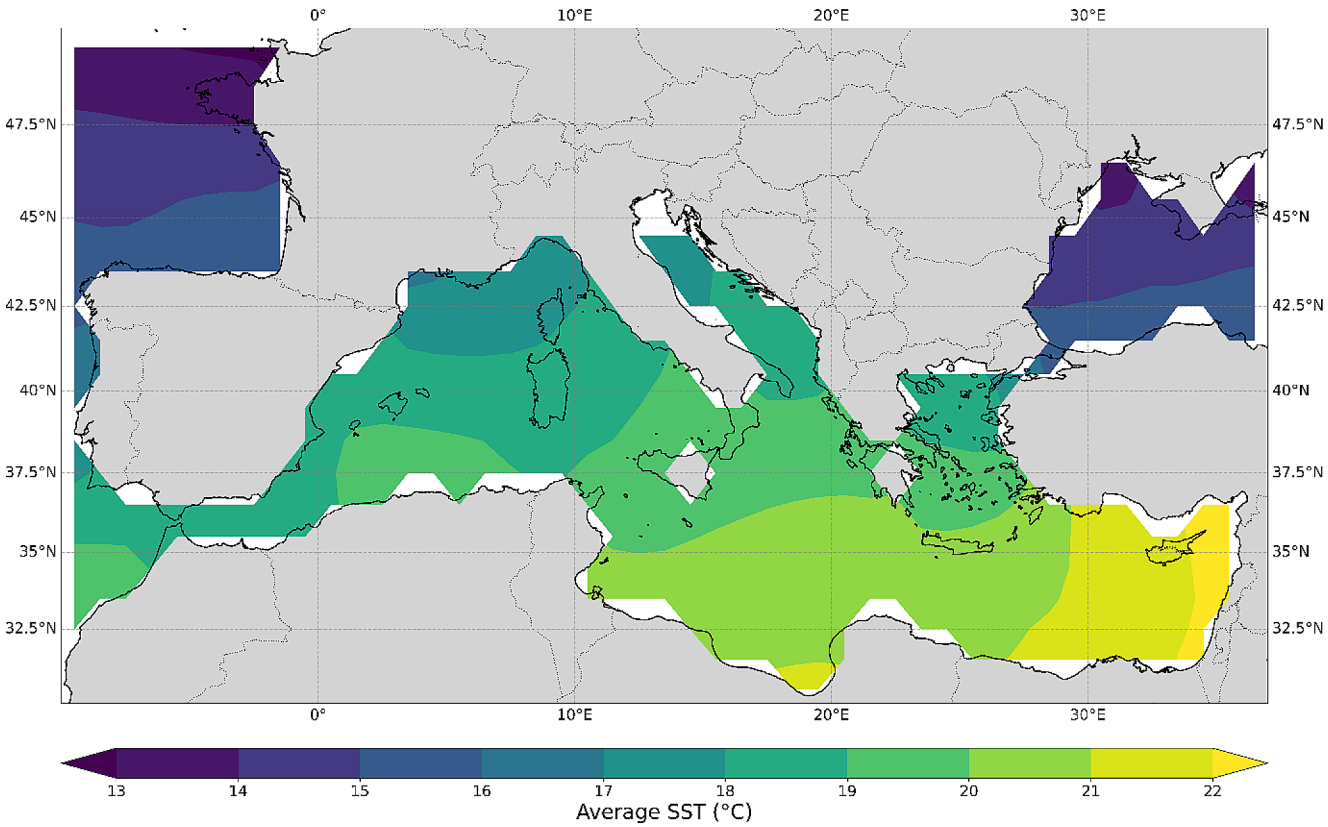


Fig. 4 Spatial distribution of average SST in the Mediterranean Sea and surrounding waters (1870–2023)

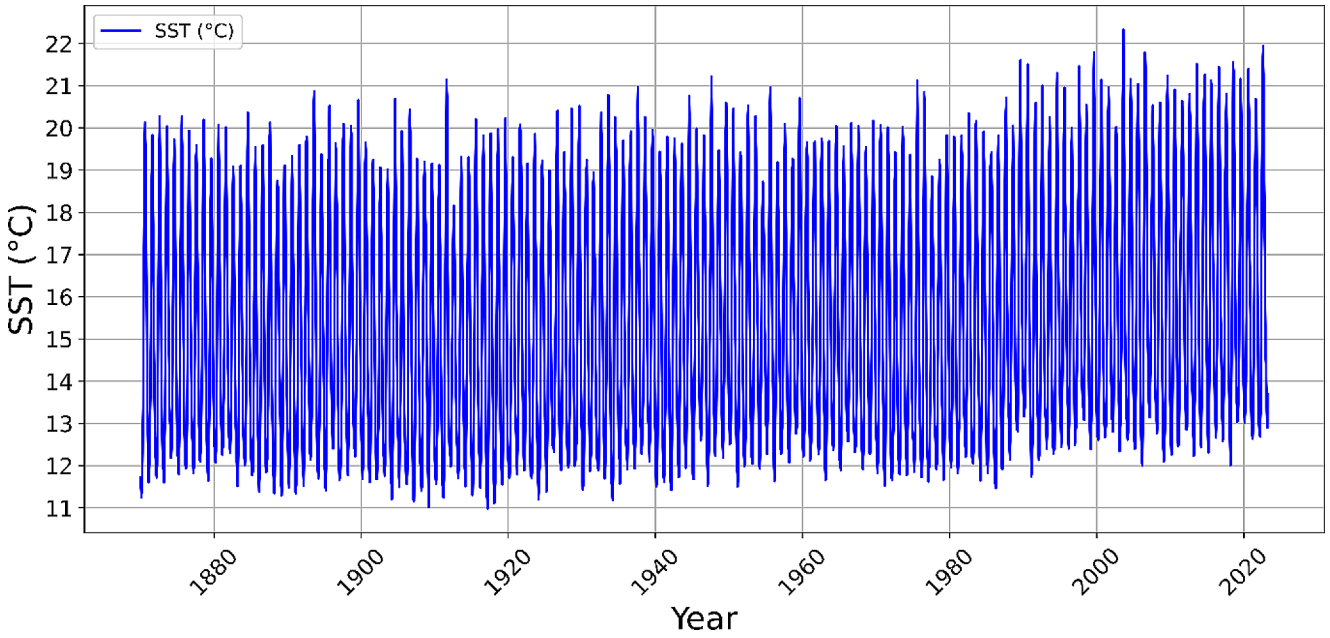


Fig. 5 Historical trends and recent elevations in monthly SST of the Mediterranean Sea (1870–2023)

Table 3 Summary of sea surface temperature (SST) data for Coastal Mediterranean cities (1961–2023)

City	Latitude Longitude	Country	Mean	Min	Max	Range	Skewness	Kurtosis	Missing Values
Ajaccio	41.92 8.79	France	18.2	12.7	27.3	14.6	0.3	-1.3	0
Alexandria	31.18 29.95	Egypt	21.4	15.7	28.1	12.4	0.1	-1.5	0
Alicante	38.37 -0.49	Spain	19.2	13.3	27.7	14.4	0.3	-1.3	0
Antalya	36.9 30.8	Turkey	21.4	15.2	29.0	13.8	0.2	-1.4	0
Bellavista	39.93 9.71	Italy	18.7	12.6	27.8	15.2	0.3	-1.3	0
Gibraltar	36.15 -5.35	Gibraltar	18.7	14.9	23.8	8.9	0.3	-1.3	0
Heraklion	35.34 25.18	Greece	20.2	14.7	27.0	12.3	0.2	-1.5	0
Izmir	38.39 27.08	Turkey	19.0	13.2	26.0	12.8	0.2	-1.4	0
Malaga	36.67 -4.48	Spain	18.6	14.7	24.2	9.5	0.4	-1.2	0
Marseille	43.44 5.22	France	17.3	12.2	26.4	14.2	0.4	-1.2	0
Mersin	36.78 34.6	Turkey	22.4	15.1	30.4	15.3	0.2	-1.5	0
Messina	38.2 15.55	Italy	19.7	13.1	28.6	15.5	0.3	-1.4	0
Nice	43.65 7.21	France	17.7	12.7	26.9	14.2	0.4	-1.2	0
Tunis	36.83 10.23	Tunisia	19.5	13.2	28.1	14.9	0.3	-1.4	0
Valencia	39.48 -0.37	Spain	18.8	12.9	27.3	14.4	0.3	-1.4	0

Table 4 Summary of surface air temperature (TAS) data for Coastal Mediterranean cities

City	Latitude Longitude	Country	Mean	Min	Max	Range	Skewness	Kurtosis	Missing Values
Ajaccio	41.92 8.79	France	15.2	5.9	26.5	20.6	0.2	-1.3	0
Alexandria	31.18 29.95	Egypt	20.5	12.1	29.3	17.2	0.0	-1.4	31
Alicante	38.37 -0.49	Spain	18.1	9.0	28.2	19.2	0.2	-1.3	0
Antalya	36.9 30.8	Turkey	18.8	6.7	30.8	24.1	0.1	-1.3	4
Bellavista	39.93 9.71	Italy	17.5	8.1	29.8	21.7	0.3	-1.3	14
Gibraltar	36.15 -5.35	Gibraltar	18.4	11.9	26.1	14.2	0.2	-1.3	10
Heraklion	35.34 25.18	Greece	18.9	9.6	28.6	19.0	0.1	-1.4	0
Izmir	38.39 27.08	Turkey	18.0	5.4	30.7	25.3	0.1	-1.4	2
Malaga	36.67 -4.48	Spain	18.6	10.2	28.3	18.1	0.3	-1.3	0
Marseille	43.44 5.22	France	15.0	2.0	28.2	26.2	0.1	-1.2	0
Mersin	36.78 34.6	Turkey	19.2	6.1	31.0	24.9	0.0	-1.3	26
Messina	38.2 15.55	Italy	18.6	8.4	29.5	21.1	0.2	-1.3	15
Nice	43.65 7.21	France	15.6	4.7	27.4	22.7	0.2	-1.2	0
Tunis	36.83 10.23	Tunisia	19.1	9.1	30.5	21.4	0.2	-1.3	17
Valencia	39.48 -0.37	Spain	17.0	4.7	27.7	23.0	0.1	-1.2	2

the rate of change both before and after these breakpoints (in °C/decade). For SST, the most pronounced post-breakpoint increase is recorded in Nice, with a trend shift of 0.0119 °C/decade, signaling a significant uptick in temperature rise. In contrast, the lowest post-breakpoint SST trend is seen in Mersin, where the trend actually decreases by -0.0032 °C/decade. As for TAS, the sharpest post-breakpoint rise occurs in Antalya, accelerating to an annual trend of 0.0126 °C/decade, while Messina exhibits a post-breakpoint deceleration in TAS trend, dropping − 0.0002 °C/decade.

These modeling results reveal that breakpoint years vary across the dataset, with the earliest SST breakpoint observed in Messina (1987) and the latest in Bellavista (2003). A similar temporal range is seen for TAS, with Messina again marking the earliest breakpoint in 1985. Notably, 1998 stands out as the year when several stations concurrently observed changes in TAS, marking it as a notable period of climatic shift for the region.

Machine learning findings

To further elucidate the spatial patterns emerging from the initial statistical modeling, an unsupervised machine learning approach was employed. Specifically, K-means clustering was applied to the derived SST and TAS trends (both pre-and post-breakpoint), in conjunction with geographical coordinates. This multivariate clustering technique aimed to identify coherent thermal regimes across the Mediterranean basin, effectively synthesizing the complex spatio-temporal patterns revealed by the breakpoint analysis. Figure 6 presents the results of this advanced modeling, revealing four distinct clusters that capture the nuanced spatial variability of temperature patterns.

Cluster 0 is primarily identified along the coasts of France and Spain in the northwestern Mediterranean. In this cluster, stations exhibit cooler SSTs, ranging from 17 to 19 °C (see Fig. 4). Cluster 1 is concentrated in the eastern Mediterranean, particularly in the Levantine Basin, and along the coasts of Turkey and Egypt. Corresponding to the warmest SSTs on the map, cities in this cluster often experience SSTs exceeding 20 °C, with some areas reaching up to around

Table 5 Comparative analysis of SST and TAS breakpoints and trends across Mediterranean cities

City	Breakpoint for SST	Breakpoint for TAS	SST trend before Breakpoint	SST trend after Breakpoint	TAS trend before Breakpoint	TAS trend after Breakpoint
Valencia	1994	1994	0.001	0.0026	0.0096	0.0114
Alicante	1994	1986	0.0013	0.0047	0	0.0057
Malaga	1994	1994	0.0002	0.0037	0.005	0.0063
Gibraltar	1994	1986	0.0013	0.0042	0.0036	0.0027
Marseille	1997	1989	-0.0001	0.0094	0.0042	0.0075
Nice	2000	1987	-0.0007	0.0119	0.0057	0.0073
Messina	1987	1985	-0.0054	0.0049	0.0043	-0.0002
Ajaccio	2000	1987	-0.0043	0.0113	-0.0024	0.0097
Bellavista	2003	1994	-0.0039	0.0063	-0.0046	0.0063
Heraklion	1998	1998	-0.0034	0.0024	-0.0071	0
Izmir	1998	1998	-0.0026	0.0002	-0.0003	0.0023
Antalya	1993	1998	-0.0085	0.0047	-0.0148	0.0126
Mersin	1998	1992	-0.0044	-0.0032	0.0098	0.0052
Alexandria	1993	1998	-0.0035	0.0052	-0.0017	0.0072
Tunis	1987	1994	-0.006	0.0051	0.003	0.0007

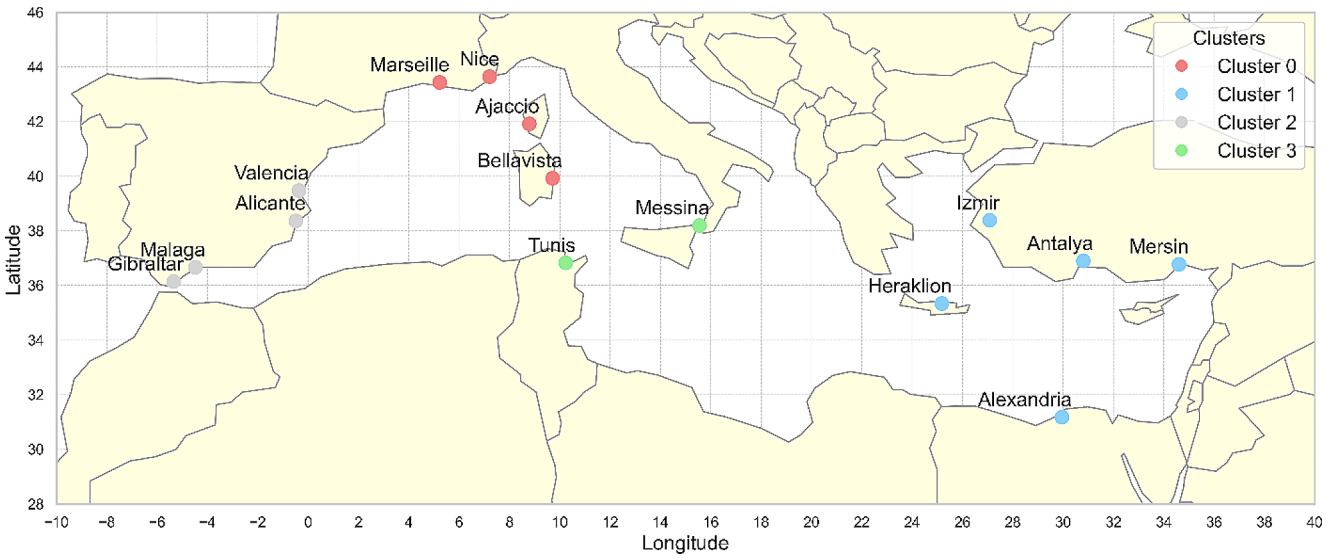


Fig. 6 K-Means climatic clustering of the studied locations across the Mediterranean

22 °C. Cluster 2 encompasses regions within the western Mediterranean, such as the Alboran Sea. This cluster's SSTs have cooler temperatures than those seen in Cluster 1 yet slightly warmer than Cluster 0. Cluster 3 spans the central Mediterranean, including parts of the Tyrrhenian Sea, Ionian Sea, and Strait of Sicily. This cluster's SSTs, while slightly warmer than Cluster 0 and Cluster 2, are cooler than Cluster 1 SSTs. The map in Fig. 6 highlights the effective differentiation of temperature regimes across the Mediterranean. It delineates a clear gradient from the cooler western Mediterranean (Clusters 0 and 2) to the warmer eastern regions (Cluster 1), punctuated by the transitionally moderate temperatures of the central Mediterranean (Cluster 3).

Building upon the clustering analysis presented in Fig. 6, a more detailed examination of SST and TAS trends within each identified cluster was conducted. This analysis aimed to clarify the temporal dynamics of temperature changes before and after the climatic breakpoints identified in the initial statistical modeling. Figure 7 presents this comprehensive analysis through a series of four box plots (a-d), each corresponding to a distinct cluster and temporal phase.

Subplot (a) of Fig. 7 reveals the heterogeneity of SST trends before the breakpoint across clusters. Notably, Cluster 0, primarily associated with the northwestern Mediterranean (as shown in Fig. 6), exhibits the highest variability, including some negative trends. In contrast, Clusters 1 and 3, representing the eastern and central Mediterranean respectively, show more constrained interquartile ranges, suggesting greater consistency in pre-breakpoint SST trends within these regions.

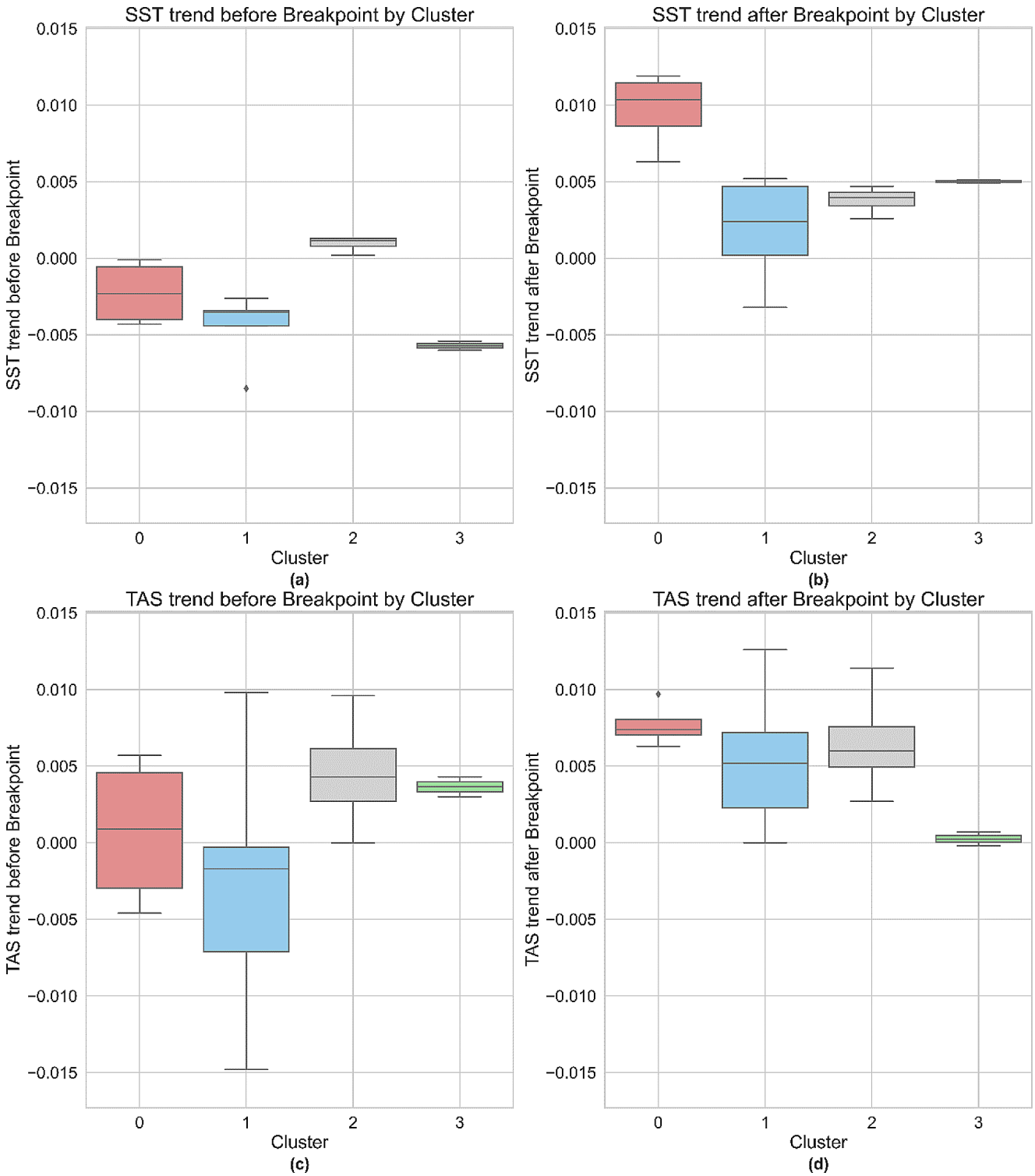


Fig. 7 Decadal variations in SST and TAS across Mediterranean clusters (°C/decade) (a) SST trend before breakpoint. (b) SST trend after breakpoint. (c) TAS trend before breakpoint. (d) TAS trend after breakpoint

Post-breakpoint SST trends, illustrated in subplot (b), demonstrate a convergence across clusters, indicative of a more uniform warming pattern throughout the Mediterranean basin after the identified breakpoints. This convergence aligns with the accelerated warming trend observed in the long-term SST data. Interestingly, Cluster 0 exhibits the most pronounced post-breakpoint increase, despite its pre-breakpoint variability, suggesting a significant shift in the thermal regime of the northwestern Mediterranean.

TAS trends, depicted in subplots (c) and (d), reveal distinct patterns before and after the breakpoint. Pre-breakpoint TAS trends (subplot c) show considerable variability in Clusters 0 and 1, while Clusters 2 and 3 exhibit more uniform, slightly positive trends. Post-breakpoint (subplot d), all clusters demonstrate positive TAS trends, with Cluster 0 again showing the most substantial increase. This pattern corroborates the overall warming trend observed in the TAS data presented in Table 4.

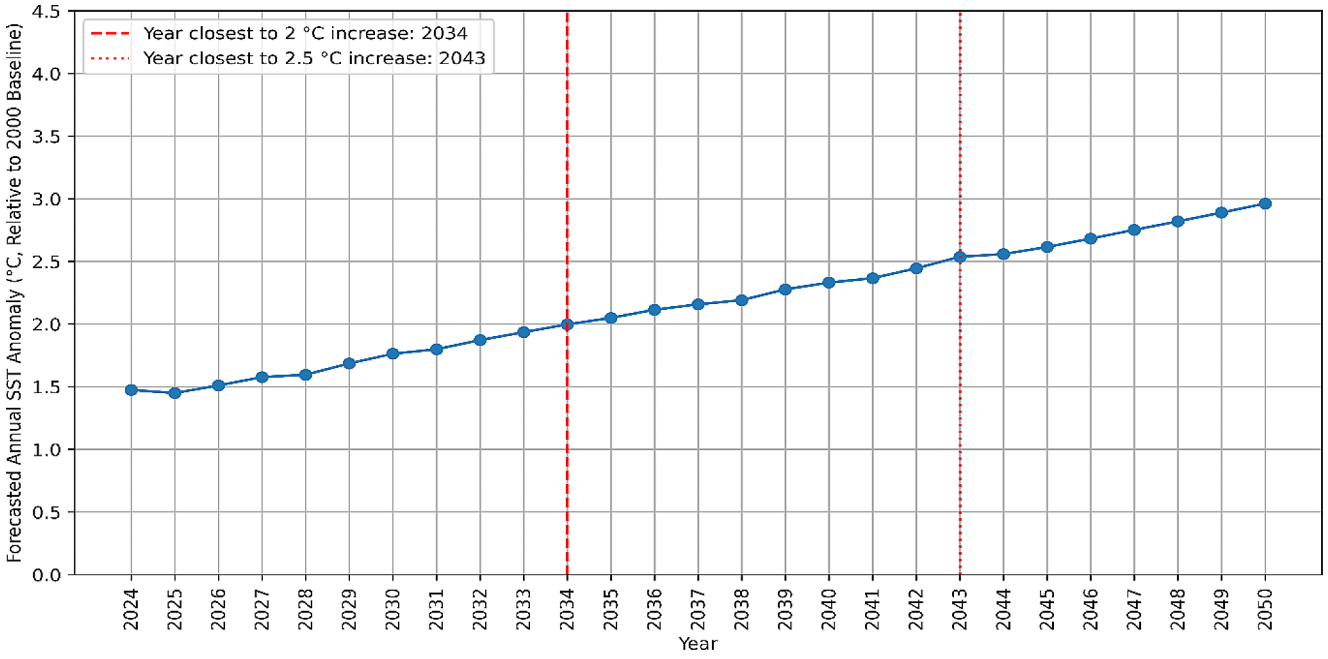
The consistent upward trend in both SST and TAS across all clusters post-breakpoint, as evident in Fig. 7, provides strong evidence for a basin-wide warming trend in the Mediterranean. The analysis suggests that while pre-breakpoint trends were spatially heterogeneous, post-breakpoint warming has been more uniform across the basin, albeit with varying intensities among the identified clusters.

Building upon the insights gained from the linear trend analysis and clustering, a more sophisticated modeling approach was implemented to capture the complex, nonlinear patterns in SST and TAS anomalies. One-dimensional Convolutional Neural Networks (1D CNNs) were employed to complement the linear analysis, offering a nuanced exploration of potential accelerations or deviations from the established linear trends identified earlier. For each cluster delineated in Fig. 6, two distinct 1D CNN models were developed and optimized: one for SST and another for TAS anomaly predictions. These models were trained on historical data and subsequently used to forecast annual anomalies from 2024 to 2050. This approach allows for a more granular understanding of the temporal evolution within each spatial cluster identified in our earlier modeling.

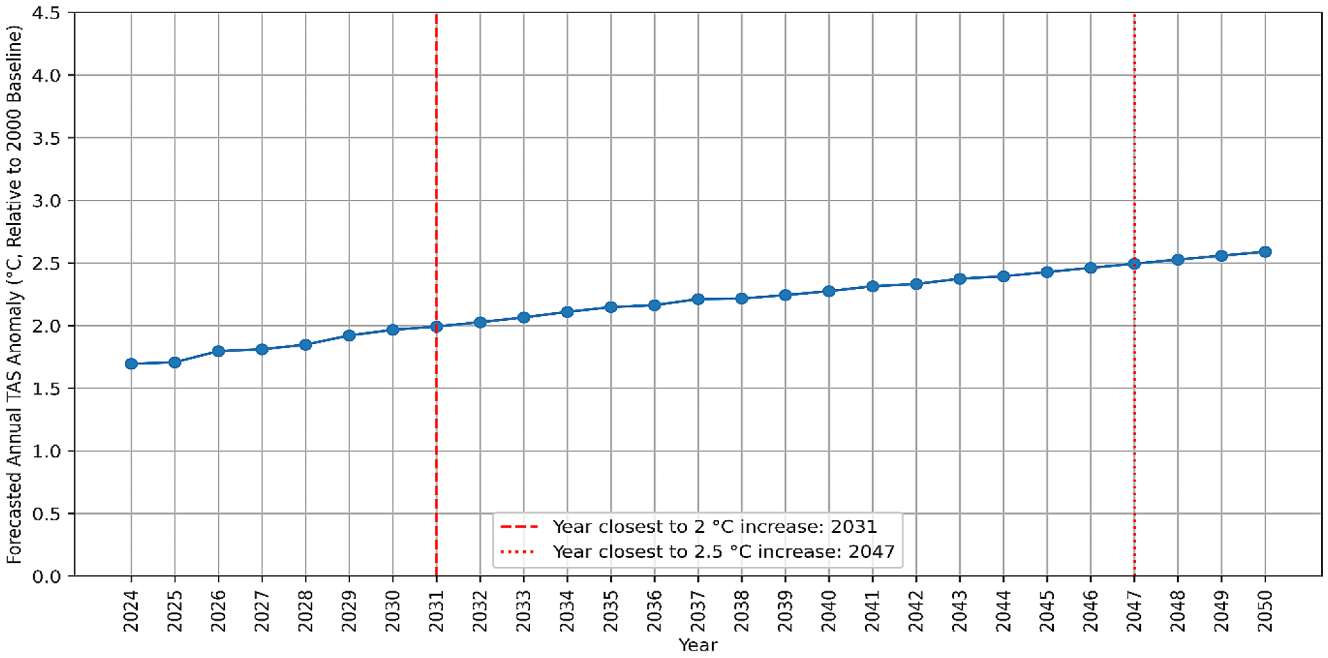
For Cluster 0, the SST-optimized model utilized a single-layer setup with 128 filters and a kernel size of 10, with a learning rate of 0.000109. This model achieved an MSE of 0.0185. The TAS prediction model was fine-tuned through two convolutional layers with optimized parameters: the first layer with 224 filters and a kernel size of 8, and the second layer with 64 filters and a kernel size of 4, supported by an optimal learning rate of 0.000841. This configuration yielded an MSE of 0.0191 on the test dataset. The two CNN models were employed to forecast annual SST and TAS anomalies for the period spanning from 2024 to 2050, benchmarked against the baseline average till the year 2000. Figure 8 (a) illustrates the anticipated annual SST anomaly, displaying a gradual upward trend with the CNN model predicting the year 2034 as the point where the SST anomaly is closest to a 2 °C increase, signifying a significant warming milestone. A further escalation is projected, with the SST anomaly nearing a 2.5 °C increase by the year 2043, indicating an accelerated rate of change. Figure 8 (b) delineates the predicted annual TAS anomaly, showing a similar progressive increase. The model forecasts the year 2031 as the juncture at which the TAS anomaly approaches a 2 °C increase. The TAS anomaly is expected to continue rising, approximating a 2.5 °C increase around 2047, mirroring the escalating trend observed in SST predictions.

For Cluster 1, the SST prediction model was streamlined with a single convolutional layer equipped with 64 filters and a kernel size of 4, fine-tuned with a learning rate of 0.005167. This model achieved an MSE of 0.005808 on the test data. The TAS-optimized model incorporated two convolutional layers, the first with 96 filters and a kernel size of 8, followed by a second layer with 32 filters and a kernel size of 6. The learning rate for this model was optimized at 0.000723. On the test data, the TAS model reported an MSE of 0.005594. The two CNN models were employed to forecast annual SST and TAS anomalies for the period spanning from 2024 to 2050, benchmarked against the baseline year 2000. Figure 9 (a) presents the forecasted annual SST anomaly, indicating a steady upward trend. The CNN model pinpoints the year 2047 as the moment when the SST anomaly will approximate a 1.5 °C increase over the baseline, marking a critical threshold in warming trends. Figure 9 (b) showcases the projected annual TAS anomaly, revealing a higher pattern of escalation. The year 2029 is identified as the threshold year when the TAS anomaly is expected to be closest to a 2 °C increase, highlighting the rapid progression of warming. This upward trend in TAS is anticipated to continue, with a 2.5 °C increase likely by 2045, consistent with the increasing trajectory noted in SST projections.

For Cluster 2, the SST prediction model was refined with a two-layer setup. This model's first and second convolutional layers both feature 256 filters with kernel sizes of 8 and 10, respectively. The model's learning rate was finely tuned to 0.000846. This SST model reported an MSE of 0.0081 on the test data. The TAS-optimized model implemented a three-layer convolutional architecture. The first layer consists of 160 filters with a kernel size of 2, the second layer has 96 filters with a kernel size of 10, and the third layer comprises 224 filters with a kernel size of 4. An optimal learning rate of 0.004071 was determined for the model. On the test dataset, the TAS model achieved an MSE of 0.008386. The two CNN models were employed to forecast annual SST and TAS anomalies for the period spanning from 2024 to 2050, benchmarked against the baseline



(a)



(b)

Fig. 8 CNN projected annual SST and TAS anomalies for cluster 0 (2024–2050). (a) Forecasted annual SST anomaly. (b) Forecasted annual TAS anomaly

year 2000. Figure 10 (a) displays the predicted annual SST anomaly trends for Cluster 2, revealing a consistent increase over the forecasted period. The 1D CNN model estimates that by the year 2043, the SST anomaly is likely to reach a 1.5 °C increase relative to the 2000 baseline, signifying a critical juncture in the region's warming pattern. The projections suggest a continued rise in SST. Figure 10 (b) depicts the anticipated annual TAS anomaly trends, which exhibit a more pronounced upward trajectory. The TAS anomaly is forecasted to approach a 2 °C increase as early as 2031,

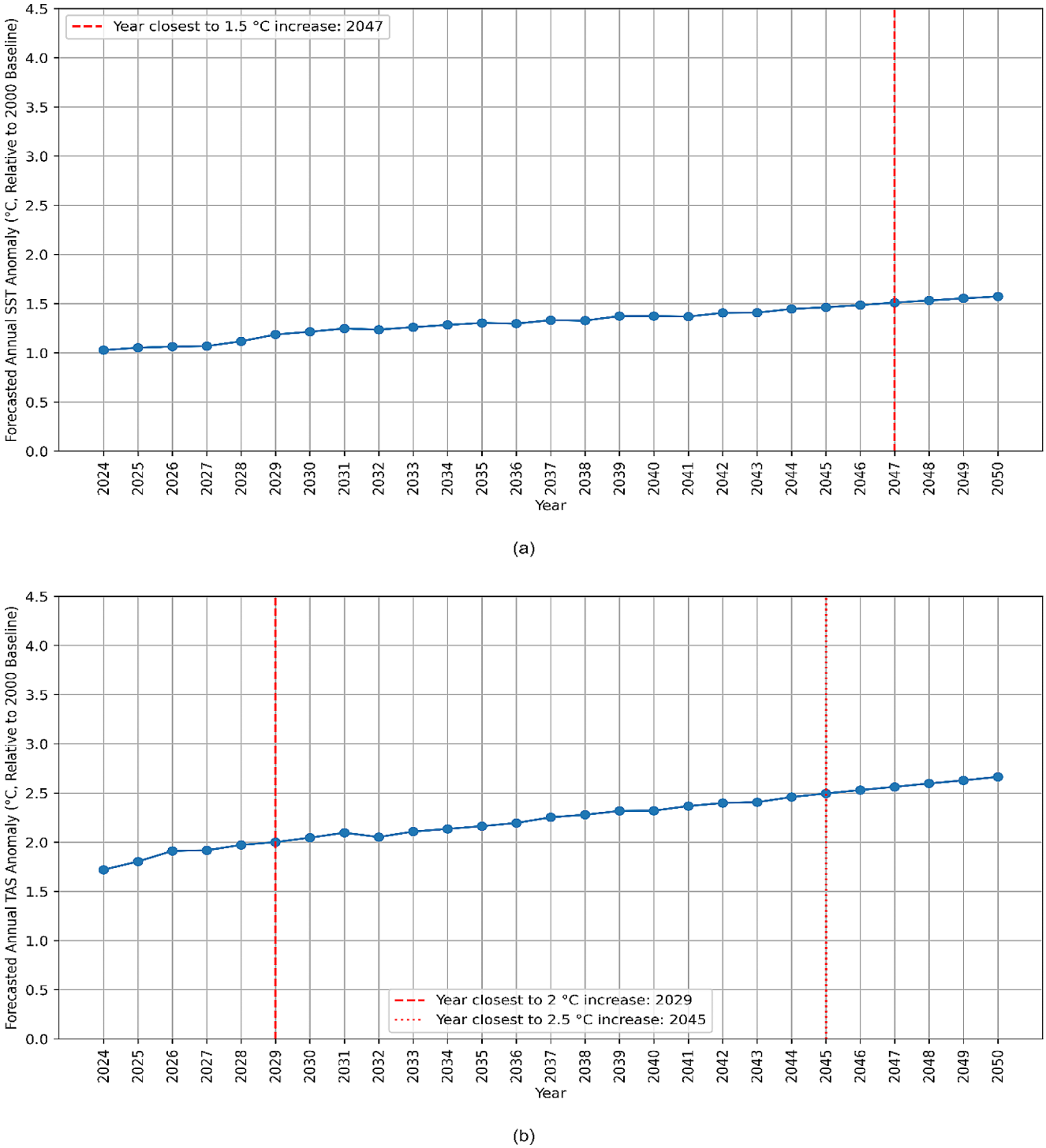


Fig. 9 CNN projected annual SST and TAS anomalies for cluster 1 (2024–2050). (a) Forecasted annual SST anomaly. (b) Forecasted annual TAS anomaly

indicating an expedited warming trend. The increase is projected to persist, with the TAS anomaly anticipated to approximate a 2.5 °C increase by 2039. Moreover, the projections suggest a continued rise in TAS with the anomaly expected to exceed a 3 °C increase by 2050.

For Cluster 3, the SST-optimized model employed a complex four-layer convolutional structure, with the initial layer comprising 256 filters with a kernel size of 4, followed by subsequent layers featuring 64, 224, and again 64 filters, with kernel sizes of 6, 4, and 2, respectively. The learning

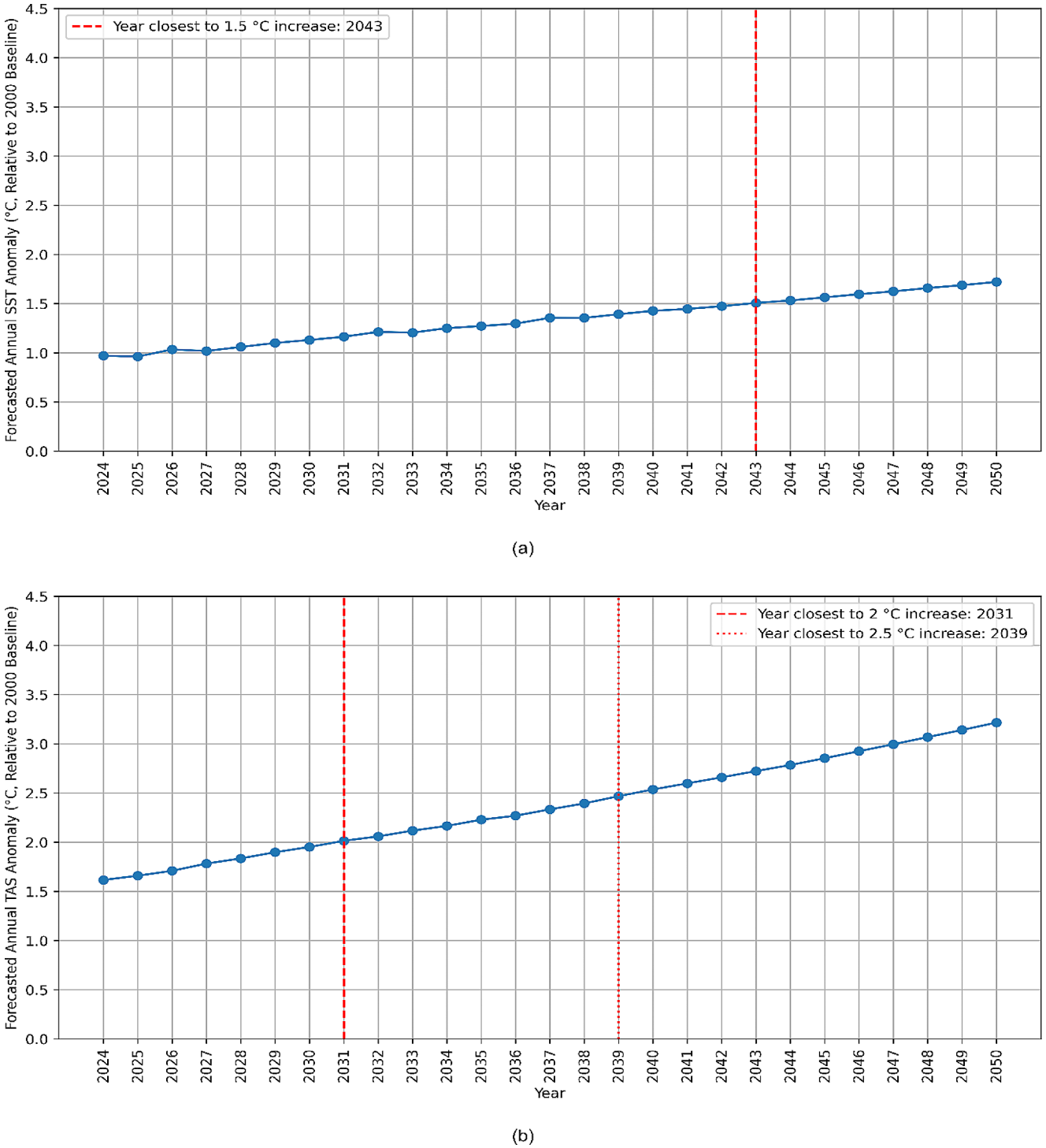


Fig. 10 CNN projected annual SST and TAS anomalies for cluster 2 (2024–2050). (a) Forecasted annual SST anomaly. (b) Forecasted annual TAS anomaly

rate was fine-tuned to an optimal value of 0.000499, with the model yielding an MSE of 0.00927 on the test data. Concurrently, the TAS prediction model was configured with a single-layer architecture equipped with 128 filters of a kernel size of 10, optimized for a learning rate of 0.00238. This TAS model demonstrated its effectiveness with an MSE of 0.00507 on the test data. Figure 11 (a) displays the predicted yearly anomalies in SST, ascertained by a CNN model for Cluster 3. The projections show a horizontal trend, with the model forecasting the year 2050 as the turning point when

the SST anomaly is expected to reach a 1.5 °C increase compared to the baseline. On the contrary, Fig. 11 (b) illustrates the anticipated yearly anomalies in TAS, following a more stable pattern. In contrast to the SST forecast, no significant increase is detected within the timespan considered, suggesting a possible stabilization in the TAS anomaly trend for this cluster.

These CNN-based projections provide a nuanced view of potential future climate trajectories in the Mediterranean, building upon the spatial patterns identified in the clustering analysis and the temporal trends observed in the initial

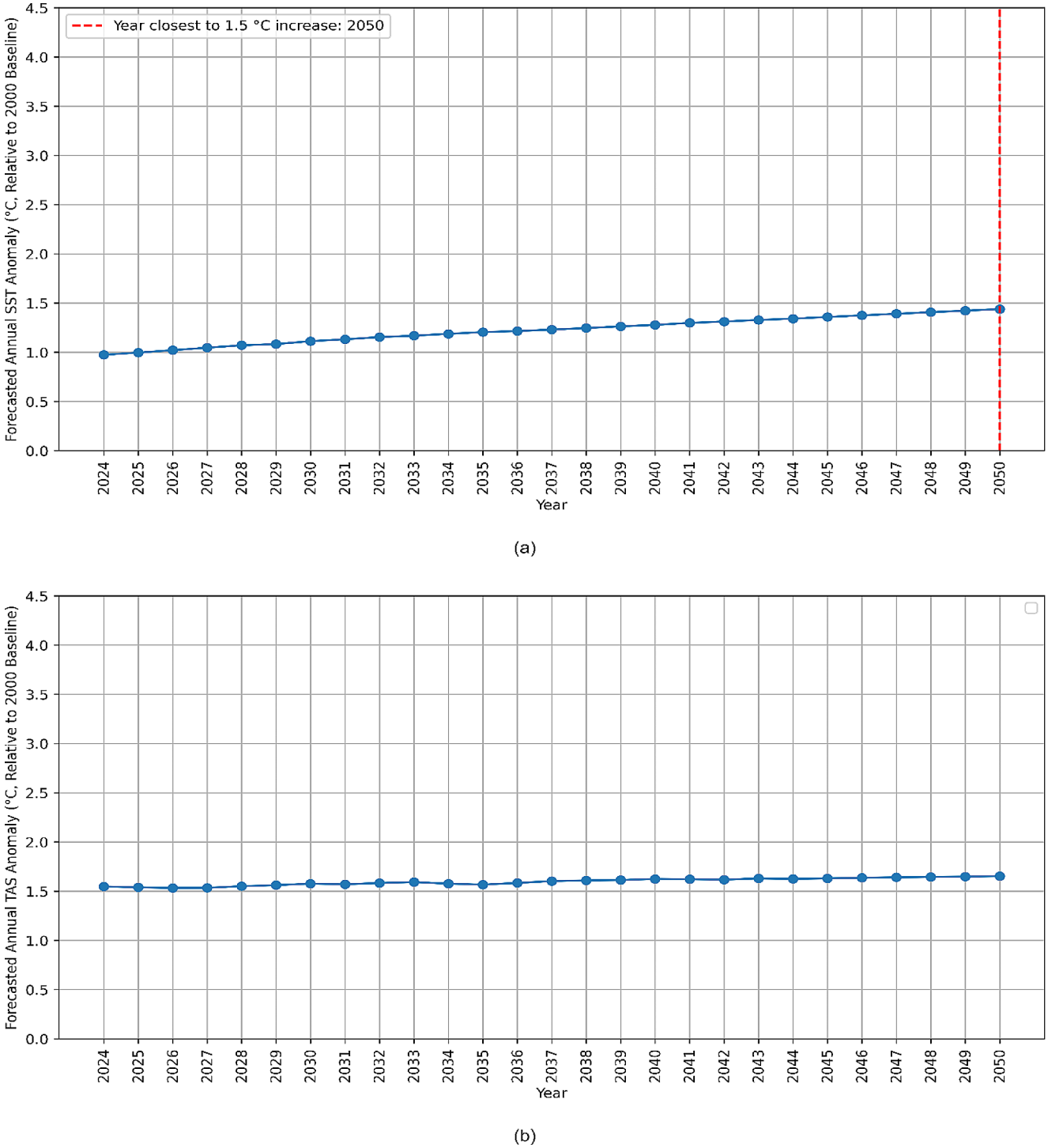


Fig. 11 CNN projected annual SST and TAS anomalies for cluster 3 (2024–2050). (a) Forecasted annual SST anomaly. (b) Forecasted annual TAS anomaly

statistical modeling. The varied projections across clusters underscore the spatial heterogeneity of climate change impacts in the region. While all clusters indicate an overall warming trend, the differences in the magnitude and timing of SST and TAS anomalies highlight the complexity of regional climate dynamics. This detailed understanding is crucial for developing tailored adaptation strategies to mitigate the diverse impacts of climate change across the Mediterranean.

Discussion

The overarching goal of this study is to elucidate the intricate patterns and dynamics directing climate trends within the Mediterranean region, a domain characterized by significant spatial and temporal heterogeneity. By employing a synergistic approach that harmonizes observational data analysis, statistical modeling techniques, clustering algorithms, and advanced machine learning models, this study aims to unravel the complexities inherent in this climatic system, offering a holistic perspective on the region's climatic evolution.

The spatial distribution of SSTs across the Mediterranean basin serves as a foundational observation, revealing distinct thermal regimes that align with well-established oceanographic patterns. The eastern regions, encompassing the Levantine Basin and adjacent coastal areas, exhibit consistently warmer temperatures compared to the cooler western and central Mediterranean zones, which are consistent with other works such as (Pastor et al. 2020) and (Aboelkhair et al. 2023). This spatial variability is rooted in the interplay of various factors, including water inflow characteristics, basin morphology, and regional atmospheric dynamics, collectively shaping the unique thermal landscapes within the Mediterranean (Estournel et al. 2021). However, the historical progression of monthly SSTs unveils a striking upward shift, with recent observations frequently exceeding 21 °C, a range that was rarely recorded prior to the 1980s. This pronounced warming trend, particularly evident since the late 20th century, underscores the profound impact of global climate change on the Mediterranean's thermal regime.

To delve into the temporal dynamics underpinning these observed changes, statistical modeling techniques were employed to identify breakpoint years when notable shifts in temperature trends occurred. The varying breakpoint years across monitoring stations highlight the spatial heterogeneity in climatic shifts within the region, reflecting the intricate interplay between local environmental factors and large-scale climatic processes. Stations in the northern Mediterranean, such as Nice, Ajaccio, and Marseille, experienced the most pronounced post-breakpoint increasing trends in SST, while eastern stations like Mersin, Izmir, and Heraklion exhibited more even trends. This underscores the localized nature of these changes and the need for region-specific assessments. Regarding TAS, stations in the northern Mediterranean, such as Nice, Ajaccio, and Marseille, also experienced the most pronounced post-breakpoint increasing trends. Only one station (Antalya) shared a similar high-increasing pattern with these northern stations. On the other hand, stations in the central Mediterranean, such as Messina and Tunis, exhibited stable temperature trends. These findings suggest that the colder water areas in the northern Mediterranean experienced more severe warming in both SST and TAS, contrasting with the warmer eastern Mediterranean region. While the eastern Mediterranean still experienced some warming, it was milder and more subdued compared to the northern part, underscoring significant spatial variability in the magnitude and rate of climatic changes across the Mediterranean basin. These results are corroborated by studies such as (Reynolds et al. 2007) and (Pisano et al. 2020), which confirm an overarching warming trend across the basin, documenting an annual temperature increase ranging from 0.007 to 0.006 °C/year, substantiating the pervasive nature of warming, albeit at a rate that is notably more accelerated than what has been observed here. Conversely, a contrasting spatial distribution of temperature increases is revealed when comparing with the findings of (Von Schuckmann et al. 2019), who identified the Eastern Mediterranean, particularly near Rhodes, Crete, and the northern Aegean Sea, as regions experiencing the highest warming rates of up to 0.055–0.06 °C/year, while the analysis in this work indicates that the most pronounced warming is occurring in the colder northwestern waters near France, with lesser warming observed in the Eastern Mediterranean.

The integration of machine learning modeling, particularly the use of 1D CNNs, significantly enhances the understanding of climatic patterns by capturing non-linearities and complex interactions that traditional statistical methods might overlook. The clustering algorithm initially segmented the Mediterranean into distinct climatic zones, allowing for tailored modeling approaches within each cluster. This segmentation was crucial as it accounted for the spatial heterogeneity that characterizes the Mediterranean climate. The CNN models were then specifically trained for each cluster to predict SST and TAS anomalies. This cluster-specific modeling is a notable advancement as it acknowledges and incorporates the inherent variability within the Mediterranean basin, providing more accurate and localized modeling. This modeling approach revealed four distinct climatic clusters within the Mediterranean region, each characterized by unique SST and TAS patterns and projections. These projections suggest a steadier, yet significant warming trend that aligns with regional atmospheric and

oceanic dynamics. For instance, in the western Mediterranean (Cluster 2), the SST model projects anomalies reaching 1.5 °C by 2043, while the TAS model predicts a more rapid increase, with anomalies reaching 3 °C by 2050. This differential warming rate between SST and TAS highlights the complex interaction between sea and air temperatures, indicating that land-based temperatures may be more sensitive to climatic changes. This emphasizes the need for targeted climate adaptation strategies that address both marine and terrestrial ecosystems to mitigate the impacts of such differential warming.

The central Mediterranean (Cluster 3) demonstrates the most stable projections. The SST model suggests a gradual increase, reaching 1.5 °C by 2050, while the TAS model indicates minimal change over the forecast period. This relative stability could be due to regional factors such as specific oceanic currents, local atmospheric conditions, or topographical features that buffer these areas from more extreme temperature changes. However, this stability underscores the importance of continuous monitoring to detect any emerging trends, as even regions with currently stable conditions could experience sudden changes due to shifting climatic dynamics.

Comparatively, the northwestern Mediterranean (Cluster 0) is projected to experience significant warming trends. The SST-optimized CNN model forecasts anomalies approaching 3 °C above the 2000 average by 2050, and the TAS model predicts a similar trajectory, indicating a critical warming threshold around 2034 for SST and 2031 for TAS. This rapid increase in temperature anomalies highlights the severe impact of climate change on this region, necessitating urgent adaptation measures. The significant warming in this cluster contrasts sharply with the more moderate increases observed in the eastern Mediterranean (Cluster 1), where the SST model projects a 1.5 °C increase by 2047 and the TAS model indicates a 2 °C rise by 2029 and 2.5 °C by 2045. The differences between these clusters underscore the spatial heterogeneity of climate impacts within the Mediterranean basin.

The CNN-based projections provide a nuanced view of future climate trajectories, enhancing the spatial and temporal resolution of our climate models. By predicting the timing and magnitude of critical warming thresholds, these models are particularly valuable for developing proactive and region-specific climate adaptation strategies. For example, the identification of a 2 °C SST anomaly around 2034 in Cluster 0 and a similar TAS anomaly by 2031 underscores the urgent need for early intervention in these regions. Similarly, the moderate but steady warming trend in Cluster 1 suggests a need for ongoing adaptation measures that can accommodate gradual changes over time.

Furthermore, the findings of this study are consistent with broader literature, reinforcing the validity of the modeling approach. The projections for the eastern Mediterranean align with those reported by (Zittis et al. 2022), who estimated anomalies of 2.5–3 °C above the average temperatures recorded during 1986–2005. Similarly, the study's results correspond with (Cos et al. 2022), noting higher temperature anomalies in the northwestern Mediterranean compared to the Levant. However, the projections from this analysis suggest even higher temperature increases, potentially reaching up to 3.5 °C in the northwestern Mediterranean and 1.5 °C in the eastern Mediterranean. This discrepancy underscores the importance of continuous model refinement and regional assessments to accurately capture the nuances of climate change impacts, ensuring that adaptation strategies are based on the most precise and up-to-date information.

Conclusions

This study provides a comprehensive analysis of historical trends in SST and TAS across 15 Mediterranean coastal stations, utilizing data spanning over a century. By integrating data preprocessing, statistical modeling, and advanced machine learning techniques, the research offers a robust modeling framework for understanding and projecting climate dynamics within the Mediterranean region. The findings underscore the profound impact of global climate change on the Mediterranean's thermal regime, with recent observations frequently exceeding historical norms. The statistical modeling phase, employing the Pettitt test and linear trend analysis, revealed significant temporal shifts in SST and TAS across various stations. Stations in the northern Mediterranean, such as Nice, Ajaccio, and Marseille, demonstrated the highest post-breakpoint increases in both SST and TAS, indicating pronounced localized warming. In contrast, central Mediterranean stations like Messina and Tunis exhibited more stable temperature trends, highlighting the region's spatial heterogeneity. Machine learning insights further elucidated these spatial patterns. The K-means clustering algorithm effectively segmented the Mediterranean into four distinct climatic clusters, each characterized by unique SST and TAS trends. This segmentation allowed for the tailored application of 1D CNNs to provide detailed projections of SST and TAS anomalies for each cluster. The CNN models offered nuanced insights, revealing significant regional differences in the rate and magnitude of warming. For instance, clusters encompassing the cooler northwestern Mediterranean (Cluster 0) and transitional temperature zones (Cluster 2) are projected to experience rapid increases in SST and TAS anomalies, surpassing critical warming

thresholds (2 °C and 2.5 °C) within the next few decades. In contrast, Cluster 3, representing relatively stable TAS trends, and Cluster 1, encompassing the warmer eastern regions, exhibit less extreme warming patterns. Notably, the northern and western Mediterranean regions are projected to experience the most severe warming trends, both in terms of SSTs and TAS, in stark contrast to the relatively milder warming observed in the warmer eastern and mid-southern parts. These insights are vital for developing localized climate adaptation strategies and enhancing the theoretical frameworks that underpin climate modeling and analysis. While this study offers significant contributions, certain limitations should be acknowledged. The network of stations analyzed, although extensive, could be further expanded to encompass a more comprehensive representation of the entire Mediterranean basin, particularly in the eastern and southern regions. Moreover, the availability of longer-term observational TAS data would enable more robust analyses and projections, accounting for potential cyclical patterns or long-term oscillations.

Author contributions This research was solely conducted by Mutaz AlShafeey. Mutaz AlShafeey was responsible for the conception and design of the study, data collection and analysis, development of the machine learning models, interpretation of the results, and the writing and revision of the manuscript. All aspects of this work were carried out independently by the author.

Funding Open Access funding provided by Hungarian Electronic Information Services National Programme (EISZ) - Corvinus University of Budapest.
Open access funding provided by Corvinus University of Budapest.

Data availability The datasets generated and analyzed during the current study are available in the Hadley Centre Sea Ice and Sea Surface Temperature data set (HadISST) repository, https://www.metoffice.gov.uk/hadobs/hadisst/.
https://www.metoffice.gov.uk/hadobs/crutem5/.
Disclaimer All maps and borders presented in this paper are generated using the Python packages utilized in the analysis. These representations are solely based on the default settings of the software employed and do not reflect the views or positions of the author or the journal.

Declarations

Ethical approval The author agrees to ethical approval and understands its related rules and content.

Consent to participate Not applicable.

Consent for publish This manuscript has not been submitted or published in other journals, and the author agrees to consent to publish.

Competing interests The author has no relevant financial or non-financial interests to disclose.

References

Aboelkhair H, Mohamed B, Morsy M, Nagy H (2023) Co-occurrence of atmospheric and oceanic heatwaves in the eastern mediterranean over the last four decades. Remote Sens 15:1841

Ahmed M, Seraj R, Islam SMS (2020) The k-means algorithm: a comprehensive survey and performance evaluation. Electronics 9:1295

Alomar MK, Khaleel F, Aljumaily MM, Masood A, Razali SFM, AlSaadi MA, Al-Ansari N, Hameed MM (2022) Data-driven models for atmospheric air temperature forecasting at a continental climate region. PLoS ONE 17:e0277079

Androulidakis Y, Makris C, Mallios Z, Pytharoulis I, Baltikas V, Krestenitis Y (2023) Storm surges and coastal inundation during extreme events in the Mediterranean Sea: the IANOS Medicane. Nat Hazards 117:939–978

Benestad RE, Lussana C, Lutz J, Dobler A, Landgren O, Haugen JE, Mezghani A, Casati B, Parding KM (2022) Global hydro-climatological indicators and changes in the global hydrological cycle and rainfall patterns. PLoS Clim 1:e0000029

Beucler T, Gentine P, Yuval J, Gupta A, Peng L, Lin J, Yu S, Rasp S, Ahmed F, O'Gorman PA (2024) Climate-invariant machine learning. Sci Adv 10:eadj7250

Bonacci O, Bonacci D, Patekar M, Pola M (2021) Increasing trends in air and sea surface temperature in the Central Adriatic Sea (Croatia). J Mar Sci Eng 9:358

Bonaldo D, Sanchez-Arcilla A, Samaras AG, Snoussi M (2023) Climate change impacts on Mediterranean coastal and transitional areas: assessment, projection, and adaptation. Frontiers Media SA

Celik S (2020) The effects of climate change on human behaviors. Environment, climate, plant and vegetation growth:577–589

Chandler R, Scott M (2011) Statistical methods for trend detection and analysis in the environmental sciences. Wiley

Corduas M (1994) Nonlinearity tests in time series analysis. J Italian Stat Soc 3:291–313

Cos J, Doblas-Reyes F, Jury M, Marcos R, Bretonnière P-A, Samsó M (2022) The Mediterranean climate change hotspot in the CMIP5 and CMIP6 projections. Earth Sys Dyn 13:321–340

Cramer W, Guiot J, Fader M, Garrabou J, Gattuso J-P, Iglesias A, Lange MA, Lionello P, Llasat MC, Paz S (2018) Climate change and interconnected risks to sustainable development in the Mediterranean. Nat Clim Change 8:972–980

de Assis PAIVAD, SÁFADI T (2021) Study of tests for trend in time series. Brazilian J Biometrics 39:311–333

Duan Q, McGrory CA, Brown G, Mengersen K, Wang Y-G (2022) Spatio-temporal quantile regression analysis revealing more nuanced patterns of climate change: a study of long-term daily temperature in Australia. PLoS ONE 17:e0271457

Duvat VK, Magnan AK, Perry CT, Spencer T, Bell JD, Wabnitz CC, Webb AP, White I, McInnes KL, Gattuso JP (2021) Risks to future atoll habitability from climate-driven environmental changes. Wiley Interdisciplinary Reviews: Clim Change 12:e700

Estournel C, Marsaleix P, Ulses C (2021) A new assessment of the circulation of Atlantic and Intermediate Waters in the Eastern Mediterranean. Prog Oceanogr 198:102673

Fahad S, Su F, Khan SU, Naeem MR, Wei K (2023) Implementing a novel deep learning technique for rainfall forecasting via climatic variables: an approach via hierarchical clustering analysis. Sci Total Environ 854:158760

García-Monteiro S, Sobrino J, Julien Y, Sòria G, Skokovic D (2022) Surface temperature trends in the Mediterranean Sea from MODIS data during years 2003–2019. Reg Stud Mar Sci 49:102086

Guinaldo T, Voldoire A, Waldman R, Saux Picart S, Roquet H (2023) Response of the sea surface temperature to heatwaves during the France 2022 meteorological summer. Ocean Sci 19:629–647

Han H, Neira-Molina H, Khan A, Fang M, Mahmoud HA, Awwad EM, Ahmed B, Ghadi YY (2024) Advanced series decomposition with a gated recurrent unit and graph convolutional neural network for non-stationary data patterns. J Cloud Comput 13:20

Hochman A, Marra F, Messori G, Pinto JG, Raveh-Rubin S, Yosef Y, Zittis G (2022) Extreme weather and societal impacts in the eastern Mediterranean. Earth Sys Dyn 13:749–777

Huntingford C, Jeffers ES, Bonsall MB, Christensen HM, Lees T, Yang H (2019) Machine learning and artificial intelligence to aid climate change research and preparedness. Environ Res Lett 14:124007

Insua-Costa D, Senande-Rivera M, Llasat MC, Miguez-Macho G (2022) A global perspective on western Mediterranean precipitation extremes. Npj Clim Atmospheric Sci 5:9

Karmalkar AV, Bradley RS (2017) Consequences of global warming of 1.5 C and 2 C for regional temperature and precipitation changes in the contiguous United States. PLoS ONE 12:e0168697

Lionello P, Scarascia L (2018) The relation between climate change in the Mediterranean region and global warming. Reg Envriron Chang 18:1481–1493

Longobardi P, Montenegro A, Beltrami H, Eby M (2016) Deforestation induced climate change: effects of spatial scale. PLoS ONE 11:e0153357

Martínez J, Leonelli FE, García-Ladona E, Garrabou J, Kersting DK, Bensoussan N, Pisano A (2023) Evolution of marine heatwaves in warming seas: the Mediterranean Sea case study. Front Mar Sci 10:1193164

Mastrorillo M, Scartozzi CM, Pacillo G, Menza G, Desai B, Maviza G, Jaskolski M, Schapendonk F, Meddings G, Carneiro B (2024) Towards a common vision for climate change, security and migration in the Mediterranean

Méndez-Cea B, García-García I, Linares JC, Gallego FJ (2023) Warming appears as the main risk of non-adaptedness for western Mediterranean relict fir forests under expected climate change scenarios. Front Plant Sci 14:1155441

Nan W, Dawei W, Lixia W, Qiang G, Hao C Design of improved K-medoids algorithm for adaptive clustering number selection. In: International Conference on Advanced Algorithms and Neural Networks (AANN 2022), 2022. SPIE, pp 96–102

Noto L, Cipolla G, Pumo D, Francipane A (2023) Climate change in the Mediterranean Basin (Part II): a review of challenges and uncertainties in climate change modeling and impact analyses. Water Resour Manage 37:2307–2323

O'carroll AG, Armstrong EM, Beggs HM, Bouali M, Casey KS, Corlett GK, Dash P, Donlon CJ, Gentemann CL, Høyer JL (2019) Observational needs of sea surface temperature. Front Mar Sci 6:420

Osborn TJ, Jones PD, Lister DH, Morice CP, Simpson IR, Winn J, Hogan E, Harris IC (2021) Land surface air temperature variations across the globe updated to 2019: the CRUTEM5 data set. J Geophys Research: Atmos 126:e2019JD032352

Pastor F, Valiente JA, Khodayar S (2020) A warming Mediterranean: 38 years of increasing sea surface temperature. Remote Sens 12:2687

Patti B, Fiorentino F, Fortibuoni T, Somarakis S, García-Lafuente J (2022) Impacts of environmental variability related to climate change on biological resources in the Mediterranean. Front Mar Sci 9:1059424

Pettitt AN (1979) A non-parametric approach to the change-point problem. J Roy Stat Soc: Ser C (Appl Stat) 28:126–135

Pisano A, Marullo S, Artale V, Falcini F, Yang C, Leonelli FE, Santoleri R, Buongiorno Nardelli B (2020) New evidence of Mediterranean climate change and variability from sea surface temperature observations. Remote Sens 12:132

Raihan A (2023) A review of the global climate change impacts, adaptation strategies, and mitigation options in the socio-economic and environmental sectors. J Environ Sci Econ 2:36–58

Rayner N, Parker DE, Horton E, Folland CK, Alexander LV, Rowell D, Kent EC, Kaplan A (2003) Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. J Geophys Research: Atmos 108

Reynolds RW, Smith TM, Liu C, Chelton DB, Casey KS, Schlax MG (2007) Daily high-resolution-blended analyses for sea surface temperature. J Clim 20:5473–5496

Rybski D, Neumann J (2011) A review on the Pettitt test. In Extremis: Disruptive events and trends in climate and hydrology:202–213

Scarponi D, Nawrot R, Azzarone M, Pellegrini C, Gamberi F, Trincardi F, Kowalewski M (2022) Resilient biotic response to long-term climate change in the Adriatic Sea. Glob Change Biol 28:4041–4053

Schroeder K, Chiggiato J, Bryden H, Borghini M, Ben Ismail S (2016) Abrupt climate shift in the Western Mediterranean Sea. Sci Rep 6:23009

Seker M, Gumus V (2022) Projection of temperature and precipitation in the Mediterranean region through multi-model ensemble from CMIP6. Atmos Res 280:106440

Tanaka KR, Van Houtan KS (2022) The recent normalization of historical marine heat extremes. PloS Clim 1:e0000007

Urdiales-Flores D, Zittis G, Hadjinicolaou P, Osipov S, Klingmüller K, Mihalopoulos N, Kanakidou M, Economou T, Lelieveld J (2023) Drivers of accelerated warming in Mediterranean climate-type regions. Npj Clim Atmospheric Sci 6:97

Von Schuckmann K, Le Traon P-Y, Smith N, Pascual A, Djavidnia S, Gattuso J-P, Grégoire M, Nolan G, Aaboe S, Aguiar E (2019) Copernicus Marine service ocean state report, issue 3. J Oper Oceanogr 12:S1–S123

Zhang H, Wu Z (2023) The generalized Fisher's combination and accurate p-value calculation under dependence. Biometrics 79:1159–1172

Zhu J-J, Yang M, Ren ZJ (2023) Machine learning in environmental research: common pitfalls and best practices. Environ Sci Technol 57:17671–17689

Zittis G, Almazroui M, Alpert P, Ciais P, Cramer W, Dahdal Y, Fnais M, Francis D, Hadjinicolaou P, Howari F (2022) Climate change and weather extremes in the Eastern Mediterranean and Middle East. Reviews of geophysics 60:e2021RG000762