

# Unsupervised Classification of Sea Surface Temperature (SST) in the Tropical Atlantic Using Spatial and Functional Data Analysis

Ogoudjobi François Adjibode

International Chair in Mathematical Physics and Applications (ICMPA-UNESCO Chair), University of Abomey-Calavi, Cotonou, Republic of Benin

Email: francois.ajibode@gmail.com

**How to cite this paper:** Adjibode, O.F. (2025) Unsupervised Classification of Sea Surface Temperature (SST) in the Tropical Atlantic Using Spatial and Functional Data Analysis. *Applied Mathematics*, **16**, 482-502. <https://doi.org/10.4236/am.2025.166027>

**Received:** May 15, 2025

**Accepted:** June 23, 2025

**Published:** June 26, 2025

Copyright © 2025 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

---

## Abstract

In this study, we employ a spatial unsupervised classification technique to analyze the spatio-temporal variability of Sea Surface Temperature (SST) in the tropical African zone. The methodology we propose considers both the spatial dimensions of the data and their functional characteristics, distinguishing it from conventional approaches. The results demonstrate noteworthy fluctuations in SST across spatial and temporal scales. This variability signifies a detected anomaly in SST within the study area, which can be attributed to the impacts of climate change.

## Keywords

Clustering, Unsupervised Classification, Functional Data, Spatio-Temporal Data, Climate Change, SST

---

## 1. Introduction

Nowadays, various extreme environmental events such as droughts, floods, and fires are being observed in different parts of the Earth, significantly impacting millions of people worldwide [1]-[3]. These events lead to the destruction of fauna and pose a threat to marine life by causing a decrease in oxygen concentration [4]-[11]. The decline in oxygen levels in seawater is primarily attributed to climate change, which is a critical concern. This phenomenon of deoxygenation poses a severe threat to marine life and undermines the benefits that humans derive from marine ecosystems [7] [12]-[14]. Although the prevention of such events is not currently feasible, their prediction across various time and spatial scales can help mitigate potential damages stemming from their occurrence [15].

Sea surface temperature (SST), in conjunction with pollution and climate change, serves as a robust indicator of marine resource productivity [16]-[18]. SST refers to the temperature of a significant layer near the sea surface, playing a crucial role in the development of meteorological systems as well as the biomass of diverse marine organisms at different depths. This includes vital organisms like phytoplankton and pelagic fish. Additionally, SST facilitates energy exchanges between the sea and the atmosphere, making it an essential parameter to monitor and understand. Hence, understanding SST holds significance for weather prediction, offering insights into the prospective development of systems and aquatic organisms [18]-[21]. In the specific study area under consideration (West African region, with a particular focus on Benin), there exists a notable dearth of research on environmental challenges, despite the abundant potential issues concerning the enhancement of the quality of aquatic and agricultural resources, which hold immense importance in the lives of the population. Due to its influence on the growth and spatial distribution of species, SST anomalies have the potential to impose stress on fish populations [19] [22]-[24].

To enhance the monitoring of fisheries resources, Sea Surface Temperature (SST) is modeled in relation to other climatic variables and fish abundance, within the context of climate change. Oceanographers have dedicated significant attention to modeling SST as a climatic variable using an ecosystemic approach. Linear inverse models have been employed to predict SST, as demonstrated by [25] in the Niño 3 regions and [26] off tropical Atlantic. In their work, [27] has presented various modeling methods (interpolation, spectral analysis, filtering estimation, gradient, regression model, etc.) to analyze, among other aspects, how SST responds to the damages caused by the effects of climate change. Additionally, [28]-[30] have utilized supervised machine learning tools to predict microbial diversity and composition in response to SST.

However, the methods mentioned above do not fully consider the spatial and temporal information inherent in SST data. Analyzing interactions within oceanological systems in marine ecosystems also necessitates the consideration of air-ocean interactions. Extensive and complex data with dynamic spatial and/or temporal components have been generated to study interactions within oceanological systems in marine ecosystems (refer to [31] [32]). Such data are abundant across various fields, particularly in the description of oceanological systems. Understanding the relationships between variables represented as high-dimensional vectors and/or functional components is crucial for comprehending the functioning of natural systems.

Therefore, robust methods capable of harnessing the wealth of information contained within such big data are of paramount importance in enhancing the monitoring of SST's response to the effects of climate change. Functional Data Analysis (FDA) presents a suitable methodology for studying such SST data.

FDA pertains to the analysis and theory of data represented as functions, curves, images, shapes, or even more intricate mathematical objects, conceived as smooth

realizations of stochastic processes. Functional data possess an intrinsic, infinite dimensionality. The notable high dimensionality of these data presents challenges in both theoretical understanding and computational handling, with the nature of these challenges varying based on how the functional data were sampled. Functional data can be observed within temporal as well as spatial/spatio-temporal contexts.

FDA utilizes statistical tools to tackle various inquiries, including prediction tasks [33]-[35], estimation of relationships between a primary variable and other variables, and the classification of diverse sets of curves through unsupervised methods or discrimination rules [36]-[39].

Over the past decade, Functional Data Analysis (FDA) has experienced substantial growth across a diverse spectrum of scientific domains. Notably, fields such as medicine [40] [41], ecology and marine biology [33] [35] [42]-[45], as well as environmental sciences and oceanography [46]-[52], have witnessed its profound development. FDA techniques have proven valuable in monitoring networks concerning weather and pollutants (e.g. [53]-[56]), as well as in gas, oil, and petroleum sciences [57] [58], among others.

As alluded to earlier, the application of FDA tools has extended to spatial settings, where data exhibit spatial dependence. Recent research works in this area are exemplified by studies in [59]-[61]. Recognizing the need for advancements in spatially correlated functional data, [62] has extended the spatial autoregressive model and the spatial moving average model to stochastic processes taking values in Hilbert spaces. The utilization of the eigenfunctions basis of the autocovariance operator for projection purposes has been demonstrated in works such as [63] and [64]. In a different vein, [50] expanded hierarchical classification approaches to account for spatial functional correlation, while others have measured similarity between curves using variograms, incorporating spatial correlation through mode and density, as exemplified in [65]. Various methodologies for spatial functional data clustering are presented as well [66], as highlighted in the recent monograph by [66].

The objective of this study is to analyze Sea Surface Temperature (SST) through unsupervised classification using an FDA methodology grounded in Functional Principal Component and clustering analyses. This approach aims to reveal potential heterogeneity in SST across the tropical Atlantic Ocean. The structure of this work is outlined as follows: the spatial functional data analysis and clustering methodology employed are detailed in 2. 3 encompasses the presentation of the SST data from the tropical Atlantic Ocean and the subsequent application of the methodology to this dataset. Finally, 4 is dedicated to the conclusion and discussion of the findings.

## 2. Methodology

We are addressing a measurable spatial process  $X = \{X_s \in \mathbb{R}^N, N \geq 1\}$  defined over some probability space  $(\Omega, \mathcal{A}, \mathbb{P})$  and observed on some spatial region

$\mathcal{I} \subseteq \mathbb{R}^N$  of cardinal  $n$ ,  $\mathcal{I} = \{s_1, \dots, s_n\}$ ,  $s_i \in \mathbb{R}^N, i = 1, \dots, n$ . We assume that for each location  $s$ , the random variables  $X_s$  takes its values in a semi-metric space  $(\mathcal{E}, d)$ . The space  $(\mathcal{E}, d)$  is an infinite dimension space and the random variables,  $X_s, s \in \mathcal{I}$ , are locally identically distributed. This means that when a spatial location  $u$  is sufficiently close to another one  $v$ , the variables  $X_u$  and  $X_v$  have identical or similar distributions. This hypothesis is less restrictive than strict stationarity. It is motivated by the fact that it is possible to imagine that variables located on neighbouring sites may be similar and have the same local distribution that may be different to the local distribution of another set of variables at other locations. In the classical framework of FDA, the space  $\mathcal{E}$  is a space of functions, typically the space of squared integrable functions defined on some finite interval  $\mathcal{T} = [0, T], T > 0$ . Let denote with  $S$  the set of the  $n$  curves,  $S = \{X_s, s \in \mathcal{I}\}$  (renamed in an arbitrary way as  $S = \{X_1, \dots, X_n\}$  in the following).

### 2.1. Model-Based Clustering for Spatial Functional Data

In this section, we apply a model-based clustering developed by [66] to the SST data described in the upcoming section. Clustering is an unsupervised learning technique that aims to identify clusters with homogeneous characteristics. Within the clustering framework, the model-based techniques assume the existence of a latent categorical random variable  $Z$  defining  $G$  clusters of data. This variable  $Z$  leads to a probability distribution of data as a mixture of cluster distributions. Let  $f$  denote the probability distribution of  $X$  and by  $f_g$  represent the probability distribution of  $X$  given  $Z = g$ . Consequently, the mixture model is expressed as

$$f(x) = \sum_{g=1}^G \pi_g f_g(x); \quad (1)$$

where  $\pi_g = P(Z = g)$  represents the prior probability of cluster  $g$ .

In the context of spatial dependency, the model given in Equation (1) has been extended to incorporate the location  $s$  into the prior probabilities of clusters. This modification transforms the mixture model into:

$$f(x/s) = \sum_{g=1}^G \pi_g(s; \beta) f_g(x); \quad (2)$$

where  $\beta$  represents a parametrization of the spatial prior. Consequently, given the cluster  $Z = g$ , the distribution of observations within the cluster becomes independent of location. All spatial dependencies are accounted for the priors  $\pi_g(s; \beta)$ . This concept is utilized in [67] for clustering spatio-temporal data. This paper introduces multinomial logistic regression as a model for the  $\pi_g(s; \beta)$ :

$$\ln \left( \frac{\pi_g(s; \beta)}{\pi_C(s; \beta)} \right) = \beta_{0g} + \langle \beta_g; s \rangle_{\mathbb{R}^N}. \quad (3)$$

Within a parametric framework, the conditional distribution  $f_g$  depends on parameters  $\theta_g$ . For instance, in the Gaussian model,  $\theta_g$  represents the mean

and the covariance matrix of cluster  $g$ . Let  $\theta$  denote the set of all parameters, which also encompasses those defining the  $\pi_g(s; \beta)$ . As a result, the model is transformed into:

$$f(x/s; \theta) = \sum_{g=1}^G \pi_g(s; \beta) f_g(x; \theta_g). \tag{4}$$

In a finite-dimensional context, the multivariate probability density function serves as the primary tool for estimating such a model using the EM algorithm. However, for functional random variables, the concept of a probability density isn't well-defined due to the infinite dimension of the data. To address this challenge, [66] employs the expansion coefficients of  $X$  with respect to a finite basis of functions. This approach enables the derivation of a well-defined probability density function based on these coefficients. The use of functional principal component analysis helps define an approximation of the probability density for functional data.

Assuming a spatial autoregressive dynamic for the random effect, [66] introduces a functional classification criterion to identify local spatially homogeneous regions. In the subsequent section, we assume that given  $Z = g$ ,  $X$  follows a Gaussian process. Within cluster  $g$ , a pseudo-density is employed [68]:

$$f_g^{(q_g)}(x; \theta_g) = \prod_{j=1}^{q_g} f_{g_j}(C_{g_j}(x), \lambda_{g_j}) \prod_{j'=q_g+1}^d f_{g_{j'}}(C_{g_{j'}}(x); \bar{\lambda}_g), \tag{5}$$

where  $f_{g_j}$  represents the probability density of the  $j$ -th major component  $C_{g_j}$  of  $X$  within cluster  $g$ . The random variables  $C_{g_j} (j=1, \dots, q_g)$  are independent Gaussian variables with zero mean and variances equal to the eigenvalues  $\lambda_{g_j}$  of the covariance operator of  $X$ . Similarly, the random variables  $C_{g_{j'}} (j' = q_g + 1, \dots, d)$  are independent Gaussian variables with zero mean and variances equal to the mean eigenvalues  $\bar{\lambda}_g$  of eigenvalues  $\lambda_{g_{j'}} (j' = q_g + 1, \dots, d)$  of the  $X$  covariance operator. Consequently, the parameters  $\theta_g = (\lambda_{g_1}, \dots, \lambda_{g_{q_g}}, \bar{\lambda}_g), q_g$  and  $d$  must be appropriately chosen.

Indeed, the surrogate density proposed can be regarded as an actual density when the functional data belong to a finite-dimensional space of functions spanned by a basis  $\{\phi_1, \dots, \phi_d\}, d \geq 1, i.e.$

$$X(t) = \sum_{j=1}^d \alpha_j \phi_j(t), \quad t \in [0, T], \quad T > 0.$$

Hence, we will choose  $d$  as the dimension of the basis used for data smoothing. In this scenario, the principal components  $C_{k_j}$  of the functional PCA can be derived by conducting PCA on the expansion coefficients of  $X$  in the metric  $M$  defined by the inner product of the basis functions.

### 2.2. The Expectation-Maximization (EM) Algorithm

Let us now outline the EM algorithm for estimating  $\theta$  and, consequently, similar to the finite setting, and based on Equation (5), the likelihood of the sample of curves  $S = \{X_s, s \in \mathcal{I}\}$  is:

$$l(\theta, S) = \prod_{s \in \mathcal{I}} \left( \sum_{g=1}^G \pi_g(s, \beta) f_g^{(q_g)}(x_s, \theta_g) \right). \tag{6}$$

A common approach for maximizing the likelihood when data are missing (such as the variable  $Z$ ) is to employ the iterative EM algorithm to maximize the likelihood (6), and modify it for update the principal components scores of each group as well as the parameters  $\beta$  define  $\pi_g(s)$  in (3).

The algorithm involves maximizing the approximate complete log-likelihood. Let  $Z_g(s)$  denote the indicator random variable for the cluster  $g$  at location  $s$ . Thus, the completed log-likelihood is as follows:

$$L_c(\theta; S, Z) = \sum_{s \in \mathcal{I}} \sum_{g=1}^G Z_g(s) \left( \log \pi_g(s, \beta) + \log f_g^{(q_g)}(x_s; \theta_g) \right); \tag{7}$$

This version is known to be easier to maximize than its incomplete counterpart. Let  $\theta^{(h)}$  represent the estimated value of the parameter at iteration  $h \geq 0$  of the algorithm [66].

**E Step:**

Since the groups to which  $Z_g(s)$  belong unknown, the E step involves calculating the conditional expectation of the approximated completed log-likelihood:

$$\begin{aligned} Q(\theta, \theta^{(h)}) &= \mathbb{E}_{\theta^{(h)}} [L_c(\theta, S, Z) / S] \\ &= \sum_{s \in \mathcal{I}} \sum_{g=1}^G t_g^{(h+1)}(s) \left( \log \pi_g(s; \beta) + \log f_g^{(q_g)}(x_s; \theta_g) \right), \end{aligned} \tag{8}$$

where  $t_g^{(h+1)}(s)$  represents the probability that the curve  $X_s$  belongs to the cluster  $g$  given  $C_{gj} = c_{gj}(x_s); j = 1, \dots, q_g$  at iteration  $h + 1$ :

$$\begin{aligned} t_g^{(h+1)}(s) &= \mathbb{E}_{\theta^{(h)}} [L(Z_g(s)) / s] \\ &= \frac{\pi_g(s; \beta^{(h)}) f_g^{(q_g)}(x_s; \theta_g^{(h)})}{\sum_{l=1}^G \pi_l(s; \beta^{(h)}) f_l^{(q_l)}(x_s; \theta_l^{(h)})}. \end{aligned} \tag{9}$$

**M step:**

The M step involves maximizing the conditional expectation of the completed log-likelihood with respect to  $\theta$

$$\begin{aligned} \theta_g^{(h+1)} &= \arg \max_{\theta_g} \sum_{s \in \mathcal{I}} t_g^{(h+1)}(s) \log f_g^{(q_g)}(x_s; \theta_g); \\ \beta_g^{(h+1)} &= \arg \max_{\beta} \sum_{s \in \mathcal{I}} t_g^{(h+1)}(s) \log \pi_g(s; \beta). \end{aligned} \tag{10}$$

Observe that  $\beta_g^{(h+1)}$  is obtained as a solution of a weighted logistic regression. The EM algorithm commences with an initial random partition of the data  $S$  into  $G$  clusters. It's important to note that in homoscedastic models, there's a modification of the update  $\theta_g^{h+1}$ . Further details can be found in [66].

**2.3. Selection Method**

To determine the number of clusters  $G$  when  $q_g (g = 1, \dots, G)$  are known, we

suggest maximizing the Bayesian Information Criterion (BIC) criterion defined as:

$$\text{BIC}(G) = \log l(G) - \frac{v_G}{2} \log(n); \tag{11}$$

where

$$v_G = (N + 1)(G - 1) + Gd + \sum_{g=1}^G (q_g [d - (q_g - 1)/2] + 1)$$

here  $v_G$  is the number of parameters in the model (including spatial mixing proportions, center means, principal scores and variances) and  $n = |\mathcal{I}|$  represents the number of points involved. When the values  $q_g (g = 1, \dots, G)$  are unknown, they can be determined by maximizing the BIC criterion. This can be achieved through the following modified M step, which aims to maximize the conditional expectation of the BIC criterion:

$$(q_g, \theta_g^{(h+1)}) = \arg \max_{(q_g, \theta_g)} \sum_{s \in \mathcal{I}} t_g^{(h+1)}(s) \log f_g^{(q_g)}(x_s, \theta_g) - \frac{v_{q_g}}{2} \log(n), \tag{12}$$

where  $v_{q_g} = q_g (d - (q_g - 1)/2)$  represents the additional number of parameters needed for the model with  $q_g$  main components, as discussed in [66].

### 2.4. Determination of the Number of Clusters

In functional data analysis, directly applying cluster analysis to observations is not often recommended. This caution arises for valid reasons; the discrete measurements intervals of the observations might be irregular, and the measurement intervals could differ among different functional observations. Consequently, conducting cluster analysis directly on such data can present challenges. To address this, a practical approach is to carry out cluster analysis based on the primary functional principal component (FPC) scores [69].

K-centers functional clustering (KCFC) is a method grounded in the computation of principal components. In this approach, the elements of each cluster are drawn with consideration to better approximation by the first principal components. The method can be outlined in the following steps:

- The process  $(X_{(s)}(\cdot))$  constitutes a mixed sub-process in  $L^2(T)$ , characterized by:

$$X_{(s)}^{(c)}(t) = \mu(s) + \sum_{k=1}^{+\infty} \xi_k(s) \rho_k(t). \tag{13}$$

- Each sub-process is associated with a cluster.
- The random variable  $C$  denotes cluster membership.

The covariance function is defined as follows:

$$\Gamma^{(c)}(s, s') = \text{cov}(X_{(s)}(t), X_{(s)}(t') / C = c). \tag{14}$$

From (13) and (14), we obtain

$$\langle \Gamma^{(c)}(s, s'), \rho_k^{(c)} \rangle = \lambda_k^{(c)} \rho_k^{(c)},$$

and

$$\xi_k^{(c)}(s) = \langle X(t, s) - \mu(s), \rho_k^{(c)} \rangle.$$

Given that the data of interest are functional in nature, dimension reduction is necessary for efficient computation. A common approach involves expanding each curve using a finite number of principal components [66] [70].

## 2.5. Clustering Using the Principal Component Scores

Recalling that clustering and supervised classification are valuable tools in traditional multivariate data analysis, they present challenges in the context of functional data analysis. Clustering involves grouping a dataset into configurations where data within clusters are more similar to each other than across clusters, based on a defined metric. In contrast, supervised classification assigns an individual to a predefined group or class using labeled observations.

In machine learning terms, functional data clustering is an unsupervised learning process, while supervised classification employs a discriminant function or classifier to assign new data to predetermined groups. Functional classification typically uses training data with functional predictors and associated multi-class labels for each data point.

In the application, it's essential to determine the percentage of variance to be explained and subsequently establish the number  $K_c$  of principal components required. Equation (13) is then modified as follows:

$$\hat{X}_{(s)}^{(c)}(t) = \mu(s) + \sum_{k=1}^{K_c} \xi_k(s) \rho_k(t). \quad (15)$$

To avoid making additional distribution assumptions, the cluster membership for an observation  $X_{(t,s)}$  is determined by:

$$\begin{aligned} C^*(X) &= \arg \min_{c \in \{1, \dots, C\}} \left\| X_{(s)}^{(c)}(\cdot) - \hat{X}_{(s)}^{(c)}(\cdot) \right\| \\ &= \arg \min_{c \in \{1, \dots, C\}} \left\| \sum_{k=K_c+1}^{+\infty} (\xi_k(s) \rho_k(\cdot))^{1/2} \right\|, \end{aligned} \quad (16)$$

which determines the cluster that can represent the observation with the smallest error. For the purpose of grouping, it's essential to initially estimate the moments, eigenfunctions, eigenvalues, and functional principal component (FPC) scores. The KCFC algorithm builds upon an initial cluster assignment based on the FPC scores  $(\hat{\xi}_{i,1}, \dots, \hat{\xi}_{i,K_c})$ , where  $i = 1, \dots, n$ , one common approach is to use a standard classification procedure such as K-means clustering with  $K_c$  representing the number of main components considered. Once the initial clustering is established, the algorithm operates as follows:

Suppose  $g_i^{(l)} \in \{1, \dots, C\}$  is the cluster membership of the  $i$ -th observation in the  $l$ -th iteration  $G^{(l)} = \{g_i^{(l)}, i = 1, \dots, n\}$  all the clusters, we have:

1) Choose  $i \in \{1, \dots, n\}$  and we calculate  $\hat{\mu}_{(-i)}^{(c)}$  and  $\hat{g}_{(-i)}^{(c)}$  based on observations with  $g_j^{(l)}$  with  $j \neq i$ .

2) Calculate the  $i$ -th predicted observation for cluster  $c$ .

$$\hat{X}_i^{(c)}(t) = \hat{\mu}_{(-i)}^{(c)}(s) + \sum_{k=1}^{K_c} \hat{\xi}_k(s) \hat{\rho}_k(t). \quad (17)$$

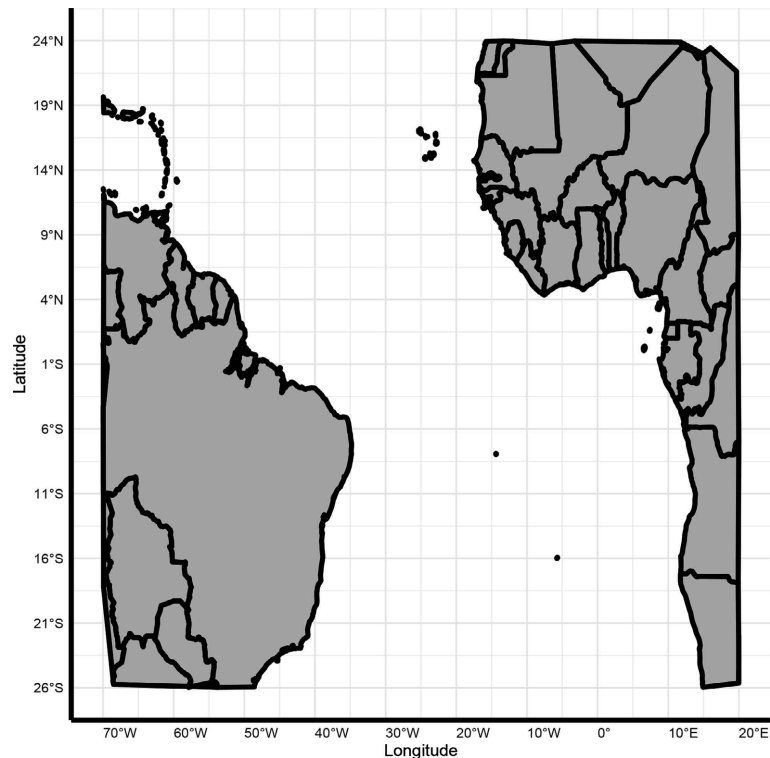
3) Observation number  $i$  is assigned to the closest cluster.

4) Steps 1 to 3 are repeated until there is no further reclassification.

### 3. Unsupervised Classification of SST in the Tropical African Zone

#### 3.1. Data Description

The data come from NCDC/NOAA (National Climatic Data Center/National Climatic Data Center) <https://psl.noaa.gov/data/gridded/data.noaa.ersst.v4.html>. They are monthly measurements of sea surface temperature (SST) off tropical African zone from January 1, 1854, to February 29, 2020. This area ( $\mathcal{I} \subset \mathbb{R}^2$ ) of interest, see **Figure 1**, is covered by longitude  $-70^\circ$  to  $20^\circ$  and latitude  $-26^\circ$  to  $24^\circ$ .



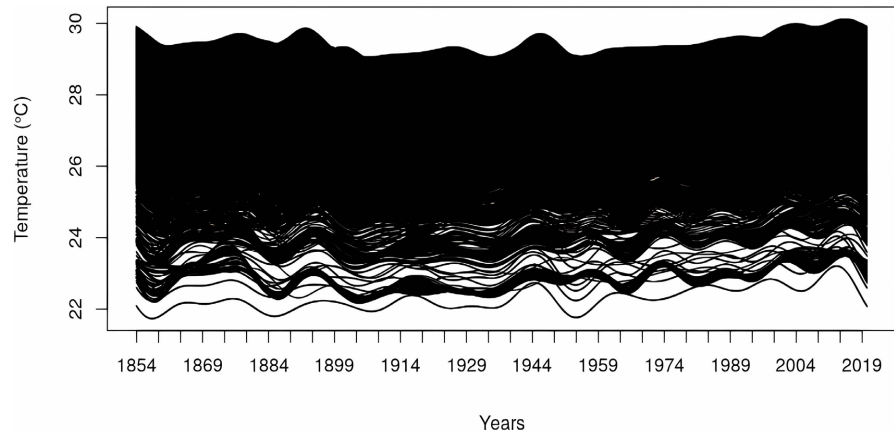
**Figure 1.** Study area.

This study area includes most of the countries of West Africa, Central Africa and especially the coastal countries.

This area is divided into 4309 geographical points. At each of these points, monthly sea surface temperatures are recorded from January 1, 1854, to February 29, 2020.

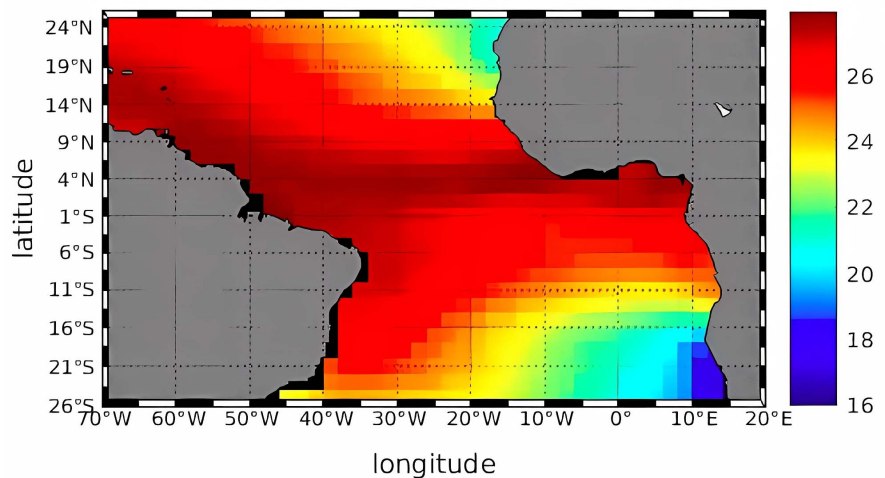
Let  $s_i \in \mathcal{I}$ ,  $i = 1, \dots, 4309$  be the locations. From January 1, 1854, to February

29, 2020, we consider monthly SST at given locations  $s_i$ . Then the temporal index is  $t \in [0, 1994]$ ;  $t = \text{month}$ . The 4309 observations recorded in these measurement sites, are transformed into a functional object using B-splines (**Figure 2**), see [71]-[73] for more details.



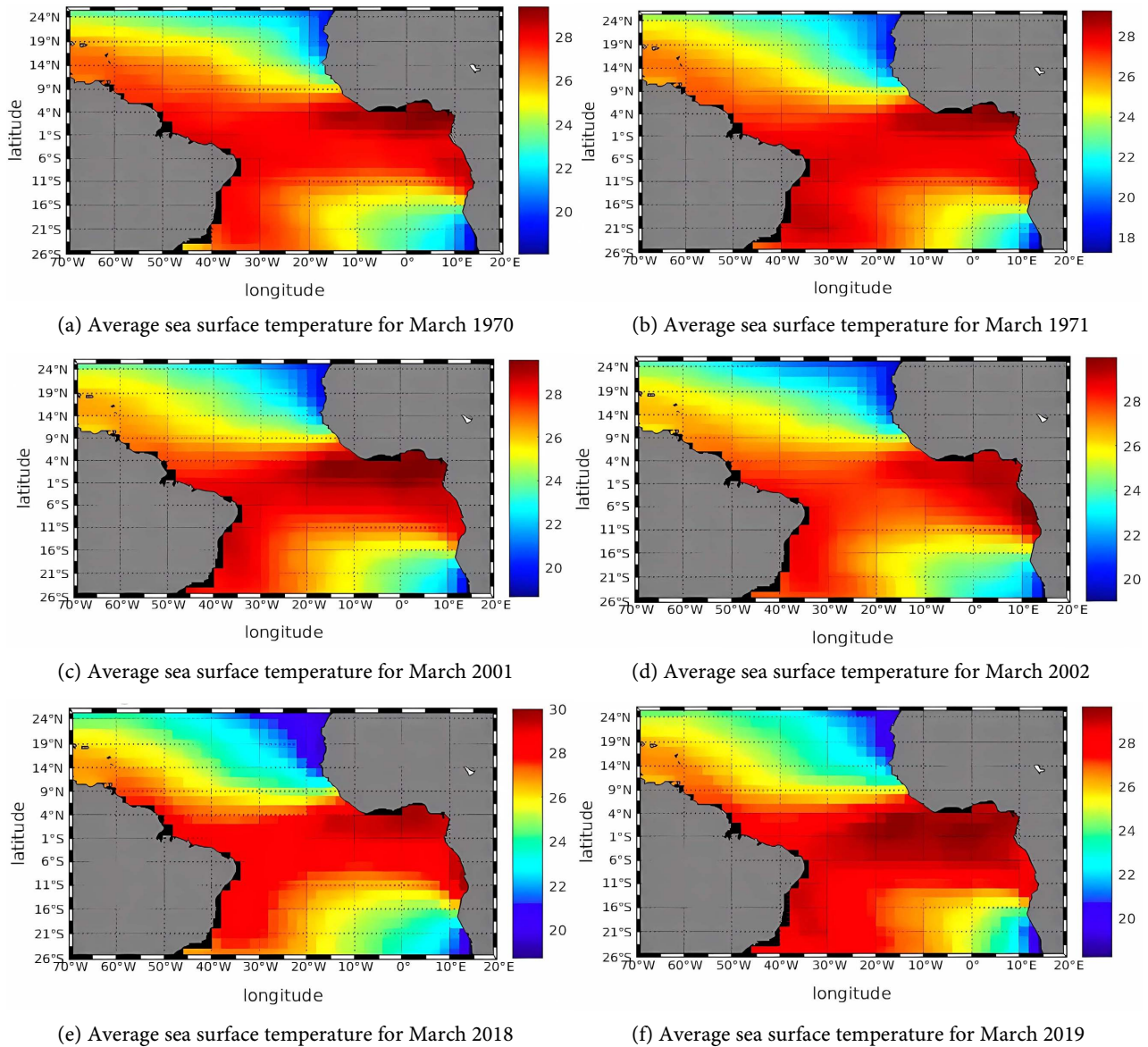
**Figure 2.** Smoothing of SST observations for all curves use B-splines.

We have 4309 sites where SST measurements were taken. **Figure 2** illustrates that not all SST curves are overlaid. The temporal temperature variation differs across the various sites, indicating spatial temperature heterogeneity. On average, the curves share a similar. When examining the curves shapes, there appears to be a suggestive periodicity.



**Figure 3.** Average sea surface temperature (SST) of the tropical African zone.

This **Figure 3** displays a heterogeneous spatial distribution of SST in the tropical zone, with higher temperatures observed in the central area and lower temperatures in the eastern extremes. **Figure 3** corroborates the spatial heterogeneity observed in 2.



**Figure 4.** Average sea surface temperature off tropical African zone for March corresponding: 1970 (a), 1971 (b), 2001 (c), 2002 (d), 2018 (e) and 2019 (f).

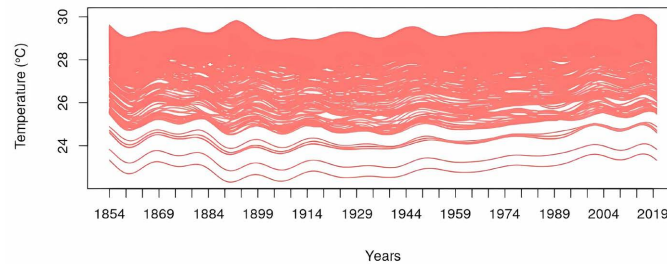
Panels (a), (b), (c) (d), (e), and (f) of **Figure 4** depict distinct SST trends for the respective years 1970, 1971, 2001, 2002, 2018 and 2019. By focusing on the month of March across these six years, it becomes apparent that the SST distribution across the off-tropical African sub-zones varies in terms of spatial scale. Notably, the spatial configuration of the off-tropical African zone in March 2018 differs from that in March 2019. The clustering method outlined in Section 2 is subsequently applied to the SST functional spatial data (as shown in **Figure 2**) to discern the heterogeneity of SST.

### 3.2. Results

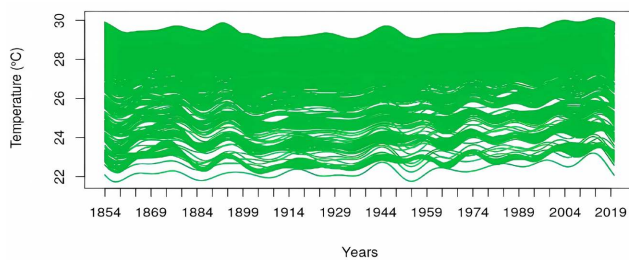
In each step of the EM algorithm, and for each value of  $q$ , BIC is computed using

Equation (12).

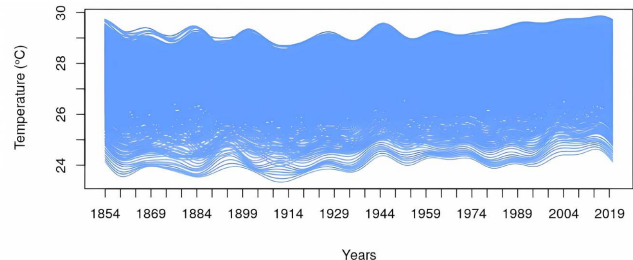
While the curves appear to share the same shape, **Figure 5** depicts three distinct classes of curves. An analysis of this figure reveals that the clustering of the sea surface temperatures (SST) off tropical Atlantic consists of three groups: one distinct cluster and a combination of two clusters. To gain a clearer view of these classes, we aim to extract and represent them separately.



(a) First class temperature curves



(b) Second class temperature curves



(c) Third class temperature curves

**Figure 5.** Clustering with three clusters.

**Figure 5** displays the outcomes of the unsupervised classification involved in three groups portraying the spatial and temporal structure of SST off tropical Atlantic.

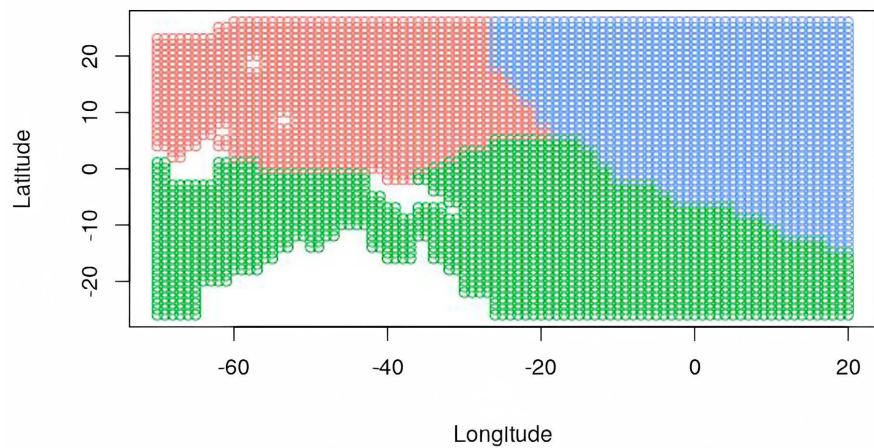
In panel (c) of **Figure 5** it is observed most curves exhibit temperature variations between the ranges:  $24^{\circ}$  and  $30^{\circ}$ . Similarly, in panel (d) of **Figure 5** an analysis reveals that most curves undergo temperature changes within the interval of:  $22^{\circ}$  and  $30^{\circ}$ . Similarly an examination of panel (e) in **Figure 5** demonstrates that most curves experience temperature fluctuations within the range of  $24^{\circ}$  to  $30^{\circ}$ .

An analysis of the panel **Figure 6** illustrates the spatial distribution of the measurement sites for the three temperature classes.

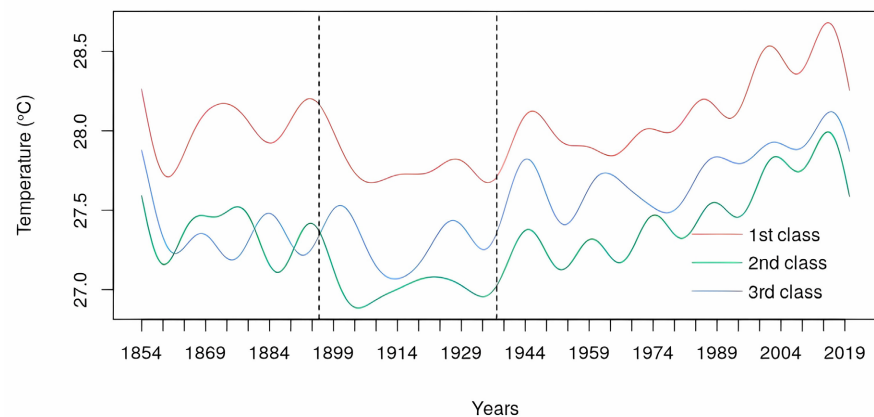
The average curves of the three classes (**Figure 7**) demonstrate distinct three phases in the SST. Each phase is characterized by abrupt changes in SST. Notably, during the initial phase, the red and blue classes are intermingled, whereas in the subsequent phases, they are clearly separated. Furthermore, the red curve class dominates as the primary class, followed by the blue curve class as the intermediate class, and the green curve class as the least prominent class.

The first phase of the red curve spans from 1854 to August 1897 (at  $t = 500$ ). The second phase, marked by a sharp SST decline, extends from September 1897

to April 1939. The final phase, characterized by an SST increase, covers the period from May 1939 to February 2020.



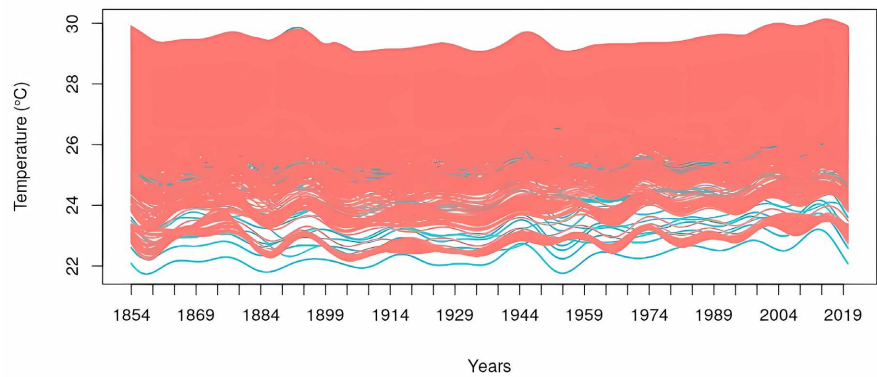
**Figure 6.** Scatter plot of locations by three clusters.



**Figure 7.** Cluster mean curves for the 3 groups clustering.

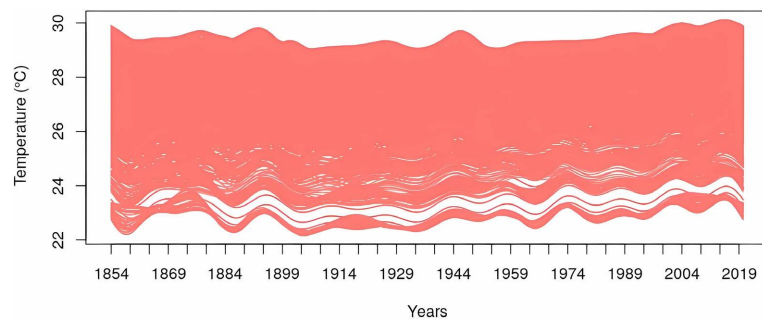
The three phases of SST variation in the green curve align with those of the red curve. A slight distinction is observed in the phases of variation of the blue curve. Notably, its first phase is longer than the first two phases of the other classes (red and green curves), extending until the year 1900. This suggests that global warming might have commenced around 1939. In summary, the descriptive analysis of **Figure 6** and **Figure 7** reveals the spatial distribution of measurement sites across the three distinct classes: a very hot zone (red), a moderately hot zone (blue), and a relatively less hot zone (green).

A more detailed analysis of the differences in SST curves could be beneficial through a grouping of SSTs that enables the clear differentiation of two classes (**Figure 8** and **Figure 9**). In each class, sites with similar SST curves are grouped together. Furthermore, by considering the average curves within the classes, these can be divided into two categories: the hot class and the non-hot class (**Figure 10** and **Figure 11**).

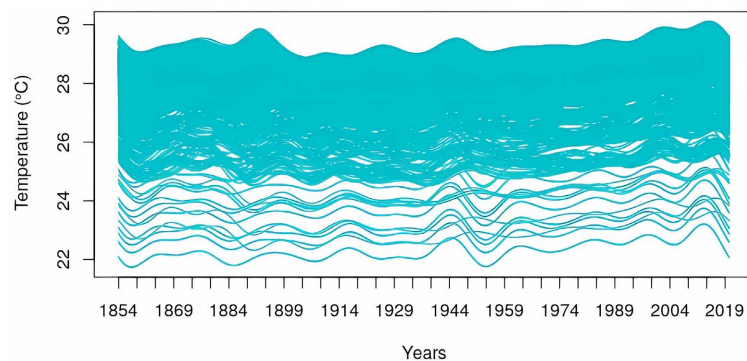


**Figure 8.** Clustering with two clusters.

An analysis of the graph in **Figure 8** reveals that the SST range fluctuates between  $16^{\circ}$  and  $28^{\circ}$ . Regarding the classification into three classes, **Figure 9** distinctly illustrates the heterogeneous nature of SST. To enhance visibility of the two classes, they will be presented separately in two panels (**Figure 9**).



(a) First class temperature curves



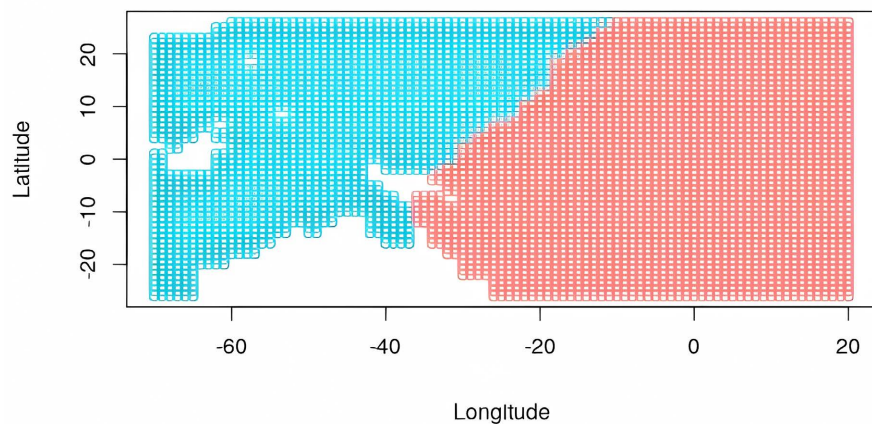
(b) Second class temperature curves

**Figure 9.** Clustering with two clusters.

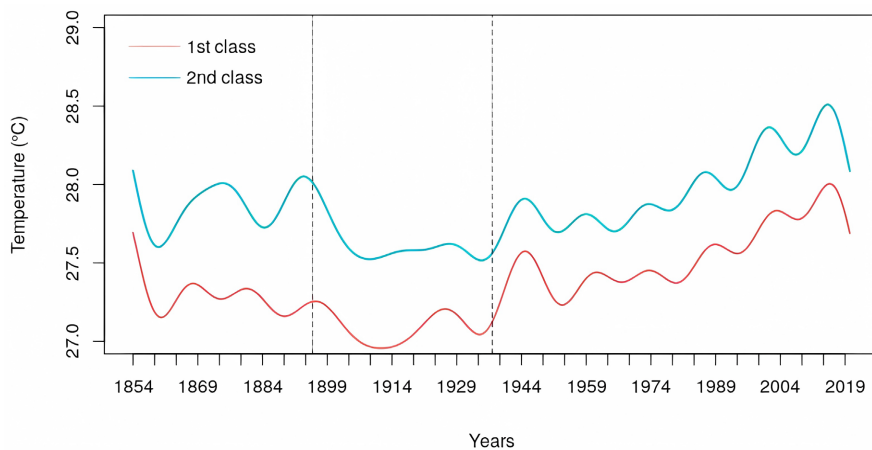
**Figure 9** illustrates the outcomes of unsupervised classification using 2 groups to represent the spatial and temporal structure of SST of the tropical Atlantic.

**Figure 10** and **Figure 11** present two distinctly discernible clusters, demonstrating the heterogeneity of SST across both spatial and temporal scales. It's note-

worthy that these figures highlight the evident spatial and temporal heterogeneity of SST within the tropical zone.



**Figure 10.** Scatter plot of locations by two clusters.



**Figure 11.** Mean cluster curves for the 2 groups clustering.

The comprehensive analysis of the two curves in **Figure 11** reveals three distinct phases of sea surface temperature (SST) change. The first phase spans from 1854 to August 1897 ( $t = 500$ ). The second phase exhibits a sudden SST drop and covers the period from September 1897 to April 1939. The final phase extends from May 1939 to the end of February 2020. Throughout these phases, the two SST classes (represented by the red and blue curves) exhibit clear separation. The warmer class corresponds to the blue curve, while the cooler class corresponds to the red curve.

#### 4. Conclusion and Discussion

This contribution introduces a novel technique, unsupervised classification, to analyze spatial functional data and delve into the spatial and temporal dynamics of Sea Surface Temperature (SST) off tropical Africa. Considering the range of ap-

plications involving multivariate methods and machine learning in oceanic data analysis, it is evident that unsupervised classification has transformed the traditional manual approach to SST data analysis. It has not only enhanced the efficiency of spatial functional data analysis but also provided tailored solutions for specific scientific research questions within this field.

This new method is particularly significant in identifying some possible anomalies in the ocean, using SST as an indicative factor of such physical or environmental parameter irregularities. It comprehensively encompasses temporal dynamics and spatial of the variation of SST off the tropical Atlantic, setting. The proposed approach apart from conventional multivariate space-time series analyses. The outcomes presented in **Figures 4-11** depict distinct SST anomalies, highlighting by the temporal and spatial variations of SST spanning from 1854 to February 2020. These anomalies might be attributed to the influence of climate change. However, it is crucial to characterize the different phases noted in the temporal evolution of SST.

This study has revealed that the sea surface temperature from January 1854 to February 2020 can be delineated into three distinct phases. The first phase spans from 1854 to August 1897, followed by a decline in temperature observed from September 1897 to April 1939. The third phase, extending from May 1939 to February 2020, represents the most significant upward trend, signifying the contemporary climate warming. This result suggests that global warming commenced following the Second World War.

Given the significance and complexity of the results we have attained, alongside the ongoing advancements in machine learning and ocean observation technology, it would be prudent in the very near future to expand this study to encompass whole off African coast. This expansion could involve employing supervised classification methods while considering the local specifics of each country.

## Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

## References

- [1] Tabari, H. (2021) Extreme Value Analysis Dilemma for Climate Change Impact Assessment on Global Flood and Extreme Precipitation. *Journal of Hydrology*, **593**, Article ID: 125932. <https://doi.org/10.1016/j.jhydrol.2020.125932>
- [2] Perkins-Kirkpatrick, S.E., Stone, D.A., Mitchell, D.M., Rosier, S., King, A.D., Lo, Y.T.E., *et al.* (2022) On the Attribution of the Impacts of Extreme Weather Events to Anthropogenic Climate Change. *Environmental Research Letters*, **17**, Article ID: 024009. <https://doi.org/10.1088/1748-9326/ac44c8>
- [3] Ebi, K.L., Vanos, J., Baldwin, J.W., Bell, J.E., Hondula, D.M., Errett, N.A., *et al.* (2021) Extreme Weather and Climate Change: Population Health and Health System Implications. *Annual Review of Public Health*, **42**, 293-315. <https://doi.org/10.1146/annurev-publhealth-012420-105026>
- [4] Ury, E.A., Yang, X., Wright, J.P. and Bernhardt, E.S. (2021) Rapid Deforestation of a

- Coastal Landscape Driven by Sea-Level Rise and Extreme Events. *Ecological Applications*, **31**, e02339. <https://doi.org/10.1002/eap.2339>
- [5] Mann, J., Foroughirad, V., McEntee, M.H.F., Miketa, M.L., Evans, T.C., Karniski, C., *et al.* (2021) Elevated Calf Mortality and Long-Term Responses of Wild Bottlenose Dolphins to Extreme Climate Events: Impacts of Foraging Specialization and Provisioning. *Frontiers in Marine Science*, **8**, Article No. 219. <https://doi.org/10.3389/fmars.2021.617550>
- [6] Johnson, A.J., Shields, E.C., Kendrick, G.A. and Orth, R.J. (2020) Recovery Dynamics of the Seagrass *Zostera Marina* Following Mass Mortalities from Two Extreme Climatic Events. *Estuaries and Coasts*, **44**, 535-544. <https://doi.org/10.1007/s12237-020-00816-y>
- [7] Sampaio, E., Santos, C., Rosa, I.C., Ferreira, V., Pörtner, H., Duarte, C.M., *et al.* (2021) Impacts of Hypoxic Events Surpass Those of Future Ocean Warming and Acidification. *Nature Ecology & Evolution*, **5**, 311-321. <https://doi.org/10.1038/s41559-020-01370-3>
- [8] Xu, G., Kong, H., Chang, X., Dupont, S., Chen, H., Deng, Y., Hu, M. and Wang, Y. (2021) Gonadal Antioxidant Responses to Seawater Acidification and Hypoxia in the Marine Mussel *Mytilus coruscus*. *Environmental Science and Pollution Research International*, **28**, 53847-53856.
- [9] Ettinger, N.P., Larson, T.E., Kerans, C., Thibodeau, A.M., Hattori, K.E., Kacur, S.M., *et al.* (2020) Ocean Acidification and Photic-Zone Anoxia at the Toarcian Oceanic Anoxic Event: Insights from the Adriatic Carbonate Platform. *Sedimentology*, **68**, 63-107. <https://doi.org/10.1111/sed.12786>
- [10] Pitcher, G.C., Aguirre-Velarde, A., Breitburg, D., Cardich, J., Carstensen, J., Conley, D.J., *et al.* (2021) System Controls of Coastal and Open Ocean Oxygen Depletion. *Progress in Oceanography*, **197**, Article ID: 102613. <https://doi.org/10.1016/j.pocean.2021.102613>
- [11] Kroeker, K.J. and Sanford, E. (2022) Ecological Leverage Points: Species Interactions Amplify the Physiological Effects of Global Environmental Change in the Ocean. *Annual Review of Marine Science*, **14**, 75-103. <https://doi.org/10.1146/annurev-marine-042021-051211>
- [12] Graves, C.A., Powell, A., Stone, M., Redfern, F., Biko, T. and Devlin, M. (2021) Marine Water Quality of a Densely Populated Pacific Atoll (Tarawa, Kiribati): Cumulative Pressures and Resulting Impacts on Ecosystem and Human Health. *Marine Pollution Bulletin*, **163**, Article ID: 111951. <https://doi.org/10.1016/j.marpolbul.2020.111951>
- [13] Laffoley, D., Baxter, J.M., Amon, D.J., Claudet, J., Hall-Spencer, J.M., Grorud-Colvert, K., *et al.* (2020) Evolving the Narrative for Protecting a Rapidly Changing Ocean, Post-Covid-19. *Aquatic Conservation: Marine and Freshwater Ecosystems*, **31**, 1512-1534. <https://doi.org/10.1002/aqc.3512>
- [14] Chauhan, A., Singh, R.P., Dash, P. and Kumar, R. (2021) Impact of Tropical Cyclone “Fani” on Land, Ocean, Atmospheric and Meteorological Parameters. *Marine Pollution Bulletin*, **162**, Article ID: 111844. <https://doi.org/10.1016/j.marpolbul.2020.111844>
- [15] Salles, R., Mattos, P., Iorgulescu, A.D., Bezerra, E., Lima, L. and Ogasawara, E. (2016) Evaluating Temporal Aggregation for Predicting the Sea Surface Temperature of the Atlantic Ocean. *Ecological Informatics*, **36**, 94-105. <https://doi.org/10.1016/j.ecoinf.2016.10.004>
- [16] Hernández-Padilla, J.C., Zetina-Rejón, M.J., Arreguín-Sánchez, F., del Monte-Luna,

- P., Nieto-Navarro, J.T. and Salcido-Guevara, L.A. (2021) Structure and Function of the Southeastern Gulf of California Ecosystem during Low and High Sea Surface Temperature Variability. *Regional Studies in Marine Science*, **43**, Article ID: 101686. <https://doi.org/10.1016/j.rsma.2021.101686>
- [17] Feng, J., Stige, L.C., Hessen, D.O., Zuo, Z., Zhu, L. and Stenseth, N.C. (2021) A Threshold Sea-Surface Temperature at 14°C for Phytoplankton Nonlinear Responses to Ocean Warming. *Global Biogeochemical Cycles*, **35**, e2020GB006808. <https://doi.org/10.1029/2020gb006808>
- [18] Sambah, A.B., Muamanah, A., Harlyan, L.I., Lelono, T.D., Iranawati, F. and Sartimbul, A. (2021) Sea Surface Temperature and Chlorophyll-a Distribution from Himawari Satellite and Its Relation to Yellowfin Tuna in the Indian Ocean. *Aquaculture, Aquarium, Conservation & Legislation*, **14**, 897-909.
- [19] Perry, R.I., Young, K., Galbraith, M., Chandler, P., Velez-Espino, A. and Baillie, S. (2021) Zooplankton Variability in the Strait of Georgia, Canada, and Relationships with the Marine Survivals of Chinook and Coho Salmon. *PLOS ONE*, **16**, e0245941. <https://doi.org/10.1371/journal.pone.0245941>
- [20] Chen, S., Wu, R. and Chen, W. (2021) Influence of North Atlantic Sea Surface Temperature Anomalies on Springtime Surface Air Temperature Variation over Eurasia in CMIP5 Models. *Climate Dynamics*, **57**, 2669-2686. <https://doi.org/10.1007/s00382-021-05826-5>
- [21] Frémont, P., Gehlen, M., Vrac, M., Leconte, J., Delmont, T.O., Wincker, P., *et al.* (2022) Restructuring of Plankton Genomic Biogeography in the Surface Ocean under Climate Change. *Nature Climate Change*, **12**, 393-401. <https://doi.org/10.1038/s41558-022-01314-8>
- [22] Sheppard, C. (2018) World Seas: An Environmental Evaluation: Volume III: Ecological Issues and Environmental Impacts. Academic Press.
- [23] Kalyan, D., Mandar, N., Mohit, A., Manickam, N., Sambhaji, M. and Baban, I. (2021) Application of Remotely Sensed Sea Surface Temperature for Assessment of Recurrent Coral Bleaching (2014-2019) Impact on a Marginal Coral Ecosystem. *Geocarto International*, **37**, 4483-4508.
- [24] Kaur, S., Kumar, P., Weller, E. and Young, I.R. (2021) Positive Relationship between Seasonal Indo-Pacific Ocean Wave Power and SST. *Scientific Reports*, **11**, Article No. 17419. <https://doi.org/10.1038/s41598-021-97047-3>
- [25] Penland, C. and Magorian, T. (1993) Prediction of Niño 3 Sea Surface Temperatures Using Linear Inverse Modeling. *Journal of Climate*, **6**, 1067-1076. [https://doi.org/10.1175/1520-0442\(1993\)006<1067:ponsst>2.0.co;2](https://doi.org/10.1175/1520-0442(1993)006<1067:ponsst>2.0.co;2)
- [26] Penland, C. and Matrosova, L. (1998) Prediction of Tropical Atlantic Sea Surface Temperatures Using Linear Inverse Modeling. *Journal of Climate*, **11**, 483-496. [https://doi.org/10.1175/1520-0442\(1998\)011<0483:potass>2.0.co;2](https://doi.org/10.1175/1520-0442(1998)011<0483:potass>2.0.co;2)
- [27] Autret, E. (2014) Analyse de champs de température de surface de la mer à partir d'observations satellite multi-sources. Theses, Télécom Bretagne; Université de Rennes 1.
- [28] Abdul Azeez, P., Raman, M., Rohit, P., Shenoy, L., Jaiswar, A.K., Mohammed Koya, K., *et al.* (2020) Predicting Potential Fishing Grounds of Ribbonfish (*Trichiurus lepturus*) in the North-Eastern Arabian Sea, Using Remote Sensing Data. *International Journal of Remote Sensing*, **42**, 322-342. <https://doi.org/10.1080/01431161.2020.1809025>
- [29] Tonelli, M., Signori, C.N., Bendia, A., Neiva, J., Ferrero, B., Pellizari, V., *et al.* (2021) Climate Projections for the Southern Ocean Reveal Impacts in the Marine Microbial

- Communities Following Increases in Sea Surface Temperature. *Frontiers in Marine Science*, **8**, Article ID: 636226. <https://doi.org/10.3389/fmars.2021.636226>
- [30] Lou, R.R., *et al.* (2021) Application of Machine Learning in Ocean Data. In: Xu, C.S., Ed., *Multimedia Systems*, Springer, 1815-1824.
- [31] Zanna, L. (2012) Forecast Skill and Predictability of Observed Atlantic Sea Surface Temperatures. *Journal of Climate*, **25**, 5047-5056. <https://doi.org/10.1175/jcli-d-11-00539.1>
- [32] Hou, M. (2022) Mori-Zwanzig Formalism Based Reduced-Order Modeling for Decision-Making in Marine Autonomy. PhD Thesis, Georgia Institute of Technology.
- [33] Yang, W., Wikle, C.K., Holan, S.H. and Wildhaber, M.L. (2013) Ecological Prediction with Nonlinear Multivariate Time-Frequency Functional Data Models. *Journal of Agricultural, Biological, and Environmental Statistics*, **18**, 450-474. <https://doi.org/10.1007/s13253-013-0142-1>
- [34] Aguilera-Morillo, M.C., Durbán, M. and Aguilera, A.M. (2016) Prediction of Functional Data with Spatial Dependence: A Penalized Approach. *Stochastic Environmental Research and Risk Assessment*, **31**, 7-22. <https://doi.org/10.1007/s00477-016-1216-8>
- [35] Ndiaye, M., Dabo-Niang, S., Ngom, P., Thiam, N., Fall, M. and Brehmer, P. (2020) Nonparametric Prediction for Spatial Dependent Functional Data: Application to Demersal Coastal Fish off Senegal. In: Manou-Abi, S.M., Dabo-Niang, S. and Salone, J.-J., Eds., *Mathematical Modeling of Random and Deterministic Phenomena*, ISTE Ltd., 31-51.
- [36] Jiménez-Cordero, A. and Maldonado, S. (2020) Automatic Feature Scaling and Selection for Support Vector Machine Classification with Functional Data. *Applied Intelligence*, **51**, 161-184. <https://doi.org/10.1007/s10489-020-01765-6>
- [37] Carrizosa, E., Molero-Río, C. and Romero Morales, D. (2021) Mathematical Optimization in Classification and Regression Trees. *TOP*, **29**, 5-33. <https://doi.org/10.1007/s11750-021-00594-1>
- [38] Richards, J.A. (2022) Clustering and Unsupervised Classification. In: Richards, J.A., Ed., *Remote Sensing Digital Image Analysis*, Springer International Publishing, 369-401. [https://doi.org/10.1007/978-3-030-82327-6\\_9](https://doi.org/10.1007/978-3-030-82327-6_9)
- [39] Levantesi, S., Nigri, A. and Piscopo, G. (2022) Clustering-Based Simultaneous Forecasting of Life Expectancy Time Series through Long-Short Term Memory Neural Networks. *International Journal of Approximate Reasoning*, **140**, 282-297. <https://doi.org/10.1016/j.ijar.2021.10.008>
- [40] Sørensen, H., Goldsmith, J. and Sangalli, L.M. (2013) An Introduction with Medical Applications to Functional Data Analysis. *Statistics in Medicine*, **32**, 5222-5240. <https://doi.org/10.1002/sim.5989>
- [41] Ieva, F. and Paganoni, A.M. (2016) Risk Prediction for Myocardial Infarction via Generalized Functional Regression Models. *Statistical Methods in Medical Research*, **25**, 1648-1660. <https://doi.org/10.1177/0962280213495988>
- [42] Elliott, J.M. (1993) The Self-Thinning Rule Applied to Juvenile Sea-Trout, *Salmo Trutta*. *The Journal of Animal Ecology*, **62**, 371-379. <https://doi.org/10.2307/5368>
- [43] Yen, J.D.L., Thomson, J.R., Paganin, D.M., Keith, J.M. and Mac Nally, R. (2014) Function Regression in Ecology and Evolution: Free. *Methods in Ecology and Evolution*, **6**, 17-26. <https://doi.org/10.1111/2041-210x.12290>
- [44] Di Battista, T., Fortuna, F. and Maturo, F. (2016) Parametric Functional Analysis of Variance for Fish Biodiversity Assessment. *Journal of Environmental Informatics*, **28**,

- 101-109. <https://doi.org/10.3808/jei.201600348>
- [45] Caruso, G. and Fortuna, F. (2021) Mediterranean Diet Patterns in the Italian Population: A Functional Data Analysis of Google Trends. In: Soitu, D., *et al.*, Eds., *Decisions and Trends in Social Systems*, Springer International Publishing, 63-72. [https://doi.org/10.1007/978-3-030-69094-6\\_6](https://doi.org/10.1007/978-3-030-69094-6_6)
- [46] Bolger, T. and Connolly, P.L. (1989) The Selection of Suitable Indices for the Measurement and Analysis of Fish Condition. *Journal of Fish Biology*, **34**, 171-182. <https://doi.org/10.1111/j.1095-8649.1989.tb03300.x>
- [47] Lorenzen, K. (1996) The Relationship between Body Weight and Natural Mortality in Juvenile and Adult Fish: A Comparison of Natural Ecosystems and Aquaculture. *Journal of Fish Biology*, **49**, 627-642. <https://doi.org/10.1111/j.1095-8649.1996.tb00060.x>
- [48] Giraldo, R., Delicado, P. and Mateu, J. (2010) Ordinary Kriging for Function-Valued Spatial Data. *Environmental and Ecological Statistics*, **18**, 411-426. <https://doi.org/10.1007/s10651-010-0143-y>
- [49] Torres, J.M., Nieto, P.J.G., Alejano, L. and Reyes, A.N. (2011) Detection of Outliers in Gas Emissions from Urban Areas Using Functional Data Analysis. *Journal of Hazardous Materials*, **186**, 144-149. <https://doi.org/10.1016/j.jhazmat.2010.10.091>
- [50] Dabo-Niang, S., Yao, A., Pishedda, L., Cuny, P. and Gilbert, F. (2009) Spatial Mode Estimation for Functional Random Fields with Application to Bioturbation Problem. *Stochastic Environmental Research and Risk Assessment*, **24**, 487-497. <https://doi.org/10.1007/s00477-009-0339-6>
- [51] Curceac, S., Ternynck, C., Ouarda, T.B.M.J., Chebana, F. and Niang, S.D. (2019) Short-Term Air Temperature Forecasting Using Nonparametric Functional Data Analysis and SARMA Models. *Environmental Modelling & Software*, **111**, 394-408. <https://doi.org/10.1016/j.envsoft.2018.09.017>
- [52] Boudreault, J., St-Hilaire, A., Chebana, F. and Bergeron, N.E. (2021) Modelling Fish Physico-Thermal Habitat Selection Using Functional Regression. *Journal of Ecohydrology*, **6**, 105-120. <https://doi.org/10.1080/24705357.2020.1840313>
- [53] Escabias, M., Aguilera, A.M. and Valderrama, M.J. (2004) Modeling Environmental Data by Functional Principal Component Logistic Regression. *Environmetrics*, **16**, 95-107. <https://doi.org/10.1002/env.696>
- [54] Ignaccolo, R., Ghigo, S. and Bande, S. (2012) Functional Zoning for Air Quality. *Environmental and Ecological Statistics*, **20**, 109-127. <https://doi.org/10.1007/s10651-012-0210-7>
- [55] Xu, B., Luo, L. and Lin, B. (2016) A Dynamic Analysis of Air Pollution Emissions in China: Evidence from Nonparametric Additive Regression Models. *Ecological Indicators*, **63**, 346-358. <https://doi.org/10.1016/j.ecolind.2015.11.012>
- [56] Xiao, W. and Hu, Y. (2018) Functional Data Analysis of Air Pollution in Six Major Cities. *Journal of Physics: Conference Series*, **1053**, Article ID: 012131. <https://doi.org/10.1088/1742-6596/1053/1/012131>
- [57] Zhou, P., Sang, H., Jin, L. and Lee, W.J. (2017) Application of Statistical Methods to Predict Production from Liquid-Rich Shale Reservoirs. *Proceedings of the 5th Unconventional Resources Technology Conference*, Austin, 24-26 July 2017, 2999-3017. <https://doi.org/10.15530/urtec-2017-2694668>
- [58] Anifowose, F., Adeniyi, S., Abdulraheem, A. and Al-Shuhail, A. (2016) Integrating Seismic and Log Data for Improved Petroleum Reservoir Properties Estimation Using Non-Linear Feature-Selection Based Hybrid Computational Intelligence Models.

- Journal of Petroleum Science and Engineering*, **145**, 230-237.  
<https://doi.org/10.1016/j.petrol.2016.05.019>
- [59] Jorge, M. and Romano, E. (2016) Advances in Spatial Functional Statistics. Stochastic Environmental Research and Risk Assessment.
- [60] Mateu, J. and Romano, E. (2017) Advances in Spatial Functional Statistics.
- [61] Ndiaye, M., Dabo-Niang, S. and Ngom, P. (2022) Nonparametric Prediction for Spatial Dependent Functional Data under Fixed Sampling Design. *Revista Colombiana de Estadística*, **45**, 391-428. <https://doi.org/10.15446/rce.v45n2.98957>
- [62] Ruiz-Medina, M.D. (2011) Spatial Autoregressive and Moving Average Hilbertian Processes. *Journal of Multivariate Analysis*, **102**, 292-305.  
<https://doi.org/10.1016/j.jmva.2010.09.005>
- [63] Ruiz-Medina, M.D. and Espejo, R.M. (2012) Spatial Autoregressive Functional Plug-In Prediction of Ocean Surface Temperature. *Stochastic Environmental Research and Risk Assessment*, **26**, 335-344. <https://doi.org/10.1007/s00477-012-0559-z>
- [64] Ruiz-Medina, M.D., Anh, V.V., Espejo, R.M., Angulo, J.M. and Frías, M.P. (2013) Least-Squares Estimation of Multifractional Random Fields in a Hilbert-Valued Context. *Journal of Optimization Theory and Applications*, **167**, 888-911.  
<https://doi.org/10.1007/s10957-013-0423-4>
- [65] Hörmann, S., Kidziński, Ł. and Hallin, M. (2014) Dynamic Functional Principal Components. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, **77**, 319-348. <https://doi.org/10.1111/rssb.12076>
- [66] Mateu, J. and Giraldo, R. (2021) Geostatistical Functional Data Analysis: Theory and Methods. John Wiley and Sons.
- [67] Cheam, A.S.M., Marbac, M. and McNicholas, P.D. (2017) Model-Based Clustering for Spatiotemporal Data on Air Quality Monitoring. *Environmetrics*, **28**, e2437.  
<https://doi.org/10.1002/env.2437>
- [68] Delaigle, A. and Hall, P. (2010) Defining Probability Density for a Distribution of Random Functions. *The Annals of Statistics*, **38**, 1171-1193.  
<https://doi.org/10.1214/09-aos741>
- [69] Hörmann, S. and Kokoszka, P. (2012) Supplement to “Consistency of the Mean and the Principal Components of Spatially Distributed Functional Data”.
- [70] Antoniadis, A. and Oppenheim, G. (2012) Wavelets and Statistics, Volume 103. Springer Science & Business Media.
- [71] Ramsay, J.O. (2004) Functional Data Analysis. Encyclopedia of Statistical Sciences, 4.
- [72] Ramsay, J.O. and Silverman, B.W. (2005) Principal Components Analysis for Functional Data. In: Ramsay, J.O. and Silverman, B.W., Eds., *Functional Data Analysis*, Springer, 147-172.
- [73] Ramsay, J.O. and Silverman, B.W. (2007) Applied Functional Data Analysis: Methods and Case Studies. Springer, 191 p.