NOVA SCHOOL OF
SCIENCE & TECHNOLOGY

DEPARTMENT OF
COMPUTER SCIENCE

ALEXANDRE GUERREIRO MARTINS

BSc in Computer Science

UNSUPERVISED SPATIO-TEMPORAL ANALYSIS OF COASTAL UPWELLING FROM SEA SURFACE TEMPERATURE IMAGES

MASTER IN COMPUTER SCIENCE

NOVA University Lisbon
March, 2022

DEPARTMENT OF
COMPUTER SCIENCE

UNSUPERVISED SPATIO-TEMPORAL ANALYSIS OF COASTAL UPWELLING FROM SEA SURFACE TEMPERATURE IMAGES

ALEXANDRE GUERREIRO MARTINS
BSc in Computer Science

Adviser: Susana Maria Nascimento
Assistant Professor, NOVA University of Lisbon

Co-adviser: Paulo Relvas
Assistant Professor, Campus de Gambelas, Centro de Ciências do Mar (CCMAR), Universidade do Algarve

Examination Committee

Chair: Carla Maria Gonçalves Ferreira
Associate Professor, FCT-NOVA

Rapporteur: Cláudia Antunes
Associate Professor, IST

MASTER IN COMPUTER SCIENCE

NOVA University Lisbon
March, 2022

Unsupervised Spatio-Temporal Analysis of Coastal Upwelling from Sea Surface Temperature Images

---

This document was created with the (pdf/Xe/Lua)LaTeX processor and the NOVAthesis template (v6.9.4) [49].

Acknowledgements

Gostaria de começar por agradecer à professora Susana Nascimento pela ajuda e orientação prestadas no decorrer do desenvolvimento desta dissertação.

Não poderia também deixar de agradecer à ajuda providenciada pelo professor Boris Mirkin no que toca à fundamentação do algoritmo de Core-Shell clustering, e finalmente ao meu co-orientador professor Paulo Relvas e ao professor Joaquim Luís pela ajuda relativamente a assuntos teóricos referentes ao tema em estudo e às avaliações prestadas aos resultados obtidos.

É fundamental agradecer, também, a todos os professores do departamento de Informática da FCT pela ajuda prestada e conhecimentos transmitidos ao longos dos últimos 5 anos.

Passando para agradecimentos pessoais, gostaria de agradecer à minha namorada por toda a ajuda e carinho dados, nos melhores e piores momentos. Gostaria também de agradecer à minha família mais próxima pela preocupação que tiveram ao longo deste meu percurso académico.

Por último e com a mesma importância que os agradecimentos anteriores, gostaria de agradecer a todos os colegas de curso e faculdade, e amigos que conheci ao longo do meu percurso académico, que me ajudaram a alcançar o sucesso académico obtido.

Abstract

Oceanic dynamics are a hot topic in the scientific literature. When equatorward winds blow parallel to the coastline, these can form oceanic currents that can consequently create many phenomena, one of these being coastal upwelling. The automatic analysis, segmentation, study and understanding of such structures and phenomena is of major importance in various application areas mainly due to the time-consuming work needed to manually analyze such data. The segmentation and tracking of the dynamics of such structures can be performed by the use of a novelty clustering concept introduced in [20] where dynamic Spatio-Temporal (ST) clusters are studied and identified using consecutive Sea Surface Temperature (SST) images over a period of time.

The aim of this dissertation was to develop a new algorithm called Core-Shell clustering algorithm, which is an extension of the previously developed Sequential Self Tuning Seeded Expanding Cluster (S-STSEC) algorithm, proposed in [56, 59]. This new algorithm aims at the automatic recognition, definition and ST characterization of coastal upwelling from SST images. A suitable experimental protocol for SST image preprocessing, tests and validation of the Core-Shell clustering algorithm through two unsupervised measures was developed. Collections of images from the years 2007, 2015 and 2019 were chosen to total 69 SST images.

The segmentations' results obtained by the S-STSEC algorithm of the various SST instants were manually evaluated as accurate and of high quality. The two unsupervised evaluation measures, when applied, were used to evaluate the quality of the Core-Shell clusters created by the Core-Shell clustering algorithm when compared to the correspondent SST instants' segmentations, with mean values higher than 85%. Time series were extracted and segmented in an unsupervised manner for the computation of upwelling spans. It was concluded that the results obtained correctly represent the behavior of coastal upwelling regions and their dynamics.

Keywords: coastal upwelling, spatio-temporal clustering, image segmentation validation, sea surface temperature images

Resumo

A dinâmica oceânica é uma temática bastante debatida na literatura científica. Quando surgem ventos que sopram em direção ao equador, paralelos à costa, é possível ocorrer a formação de correntes oceânicas que, por sua vez, podem levar à criação de variados fenómenos, um dos quais o afloramento costeiro. A análise automática, segmentação, estudo e compreensão de tais fenómenos são de grande importância em diversas áreas de aplicação, devido principalmente ao trabalho demorado que é necessário para analisar manualmente esses mesmos dados. A segmentação e rastreio das dinâmicas destas estruturas podem ser realizados através da utilização de um conceito de agrupamento inovador introduzido em [20] onde grupos dinâmicos espaço-temporais são estudados e identificados a partir de imagens de satélite de temperatura da superfície do oceano (imagens SST) consecutivas, ao longo de um determinado período de tempo.

O objetivo desta dissertação foi desenvolver um novo algoritmo designado de *Core-Shell clustering*, sendo este uma extensão de um algoritmo anteriormente desenvolvido, o algoritmo S-STSEC, introduzido em [56, 59]. Este novo algoritmo tem como objetivo o reconhecimento automático, definição e caracterização espaço-temporal do afloramento costeiro a partir de imagens SST. Foi desenvolvido um protocolo experimental adequado, de pré-processamento de imagens SST, de testes e de validação do algoritmo *Core-Shell clustering* através de duas medidas não supervisionadas. Foram escolhidas coleções de imagens dos anos 2007, 2015 e 2019, perfazendo 69 imagens SST.

As segmentações obtidas pelo algoritmo S-STSEC dos variados instantes SST foram manualmente avaliadas como precisas e de grande qualidade. As duas medidas de avaliação não supervisionadas, quando aplicadas, foram usadas para avaliar a qualidade dos *Core-Shell clusters* obtidos pelo algoritmo *Core-Shell clustering* em relação às segmentações dos instantes SST obtidas, tendo sido obtidos valores médios superiores a 85%. Foram extraídas séries temporais e identificados de forma não supervisionada períodos de afloramento costeiro. Concluiu-se que os resultados obtidos representam de maneira correta o comportamento de regiões de afloramento costeiro e suas dinâmicas.

Palavras-chave: afloramento costeiro, agrupamento espaço-temporal, validação de segmentação de imagens, imagens de temperatura de superfície oceânica

Contents

Appendices

Annexes

List of Figures

List of Tables

List of Algorithms

Acronyms

ARI	Adjusted Rand Index xvii, 61–64, 71
AVHRR	Advanced Very High Resolution Radiometer 1, 12
CAM	Classification Activation Map 10
CPD	Change Point Detection 26, 27
CTA	Computed Tomography Angiography 10
CUI	Coastal Upwelling Index 18
DB-SCAN	Density-Based Spatial Clustering of Applications with Noise 3, 22–24
DEM	Digital Elevation Model 10
EKM	Enhanced K-Means 8
EM	Expectation Maximization 7, 8
GK	Gustafson-Kessel 13
GMT	Generic Mapping Tools 30, 47
IAP	Iterated Anomalous Patterns 37, 38, 43, 47, 56, 62, 71–73
KS	Kulczynski Similarity 63
MODIS	Moderate-Resolution Imaging Spectroradiometer 1
MRI	Magnetic Resonance Imaging 8, 10
PSO	Particle Swarm Optimization 13
S-STSEC	Sequential Self Tuning Seeded Expanding Cluster iv, v, x, xvii, 2–4, 16, 29, 31–36, 40, 42–47, 49–51, 56, 60–62, 65, 66, 71–73, 75
SAR	Synthetic Aperture Radar 8

SEC	Seeded Expanding Cluster 2, 4, 15, 16
SRG	Seeded Region Growing 2, 4, 6, 9–11, 14, 72
SSC	Sea Surface Chlorophyll 2, 13
SST	Sea Surface Temperature iv, v, ix–xi, xvii, 1–4, 10, 12–14, 17, 18, 29–35, 37–39, 43–57, 59–64, 66–73, 75, 76
ST	Spatio-Temporal iv, v, ix, 2–4, 18–24, 35
ST-SEC	Self Tuning Seeded Expanding Cluster 2, 15, 16
TPI	Topographic Position Index 18
WKS	Weighted Kulczynski Similarity 63, 64, 71

1

Introduction

1.1 The Problem and its Importance

Oceanic dynamics are a hot topic in the scientific literature. When equatorward winds blow parallel to the coastline, these can form oceanic currents that can consequently create many phenomena, one of these being coastal upwelling events.

These winds may cause warmer surface waters to move offshore, and consequently provoke colder waters to rise to the ocean surface. Coastal upwelling events are characterized by the presence of the aforementioned cold waters in the coastal region and warmer offshore waters, usually with a drastic contrast separating these. The colder waters have a lower salt content and a richer nutrient composition when compared with the offshore waters, which increases massively the biomass in coastal waters, attracting large animal populations and thus making its detection crucial to the fishery commercial sector.

The analysis, study and understanding of such structures and phenomena is of major importance in many application areas, from fishery and coastal monitoring, to climatic change studies, pollutant detection and oceanic dynamics. This is possible by the ever-growing amount of available remote sensed data, usually collected in the form of SST images by the Moderate-Resolution Imaging Spectroradiometer (MODIS) or the Advanced Very High Resolution Radiometer (AVHRR) sensors, for example. An example of an SST image taken of the Portuguese coast can be seen in Figure 1.1.

The increase in data quantity and the uncertainty implicit in the subjective recognition of the upwelling patterns by human experts requires researchers and developers to create and produce automatic methods for the analysis and posterior segmentation of the upwelling regions and its fronts, since manual analysis of such amounts of data is not efficient neither cost-effective. For the aforementioned reasons, remote sensing is a common method used to overcome these problems.

Several works and approaches have been presented to deal with the automatic segmentation problem of coastal upwelling from remote sensing imagery. In [18] by Chaudhari, neural networks were trained for the detection and segmentation of upwelling regions from AVHRR data. In [68] by Fernández et al., neural networks were used alongside

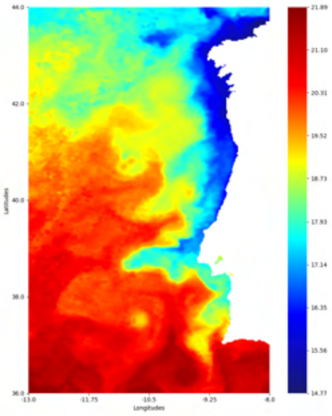


Figure 1.1: SST image of the Portuguese coast

Bayesian Nets for automatic ocean structure recognition. Supervised approaches like these involve major manual work from the user where each data sample needs to be labelled and identified for training purposes, therefore requiring a lot of time and money when compared to unsupervised approaches. Regarding the latter, Abidi et al. in [1] used a combination of SST images with Sea Surface Chlorophyll (SSC) images to overcome usual cloud noise present when segmenting upwelling regions. In [12], Aouni et al. proposed a preprocessing method to deal with the over segmentation problems present when using Seeded Region Growing (SRG) approaches for image segmentation. Such problems when using approaches are commonly referred as leaking or explosion problems.

Nascimento et al. proposed in [56] the Seeded Expanding Cluster (SEC) algorithm and a self tuning version of it, the Self Tuning Seeded Expanding Cluster (ST-SEC) algorithm, for deriving upwelling areas from SST images as an extension of the SRG algorithm presented in [3]. This approach based its segmentation not only in the pixels' temperature but also in their spatial context, achieving great and promising results when tested in SST images of the Portuguese coast. The ST-SEC version also uses an automatic computation of the homogeneity's threshold, being this an unique feature when compared to other adaptive threshold approaches that are used in the same application area. In [59], Nascimento et al. introduced the S-STSEC algorithm, where the previous work was improved and extended in order to iteratively extract clusters without providing the number of final clusters to extract, thus allowing to segment noncontinuous upwelling regions.

Moreover, along with the detection and segmentation process, most upwelling systems demonstrate seasonal, monthly and yearly spatial and temporal variability. ST data has the potential to revolutionize diverse fields where there is a need to extract and understand complex ST phenomena and their dynamics. Additionally, the data in many scientific domains like the one being studied here tend to be large and unlabeled. This highlights the importance of unsupervised methods in monitoring ST dynamics with little or no human supervision [20].

In [41] by Kalnis et al., this concept of dynamic clustering was studied when looking at moving groups of animals, where the Density-Based Spatial Clustering of Applications with Noise (DB-SCAN) algorithm [26] and the Jaccard index were used to evaluate similarities between clusters of consecutive time steps and find relationships between these. In [20], a novel approach to identify dynamic ST clusters for detecting surface water dynamics was proposed, where the core and the uncertain points concepts were introduced. The core points compose steady clusters, which are identified using an extension of the traditional DB-SCAN algorithm, and the uncertain points are points located most likely at the clusters' boundaries where dynamics occur. The concept of dynamic clusters was also introduced, these being clusters that appear in consecutive snapshots such that two consecutive spatial clusters share many common objects.

Thus, an extension to the algorithm developed in [59] by Nascimento et al. will be developed, where time series of SST images will be used in order to approach the clustering mechanism differently. Based on the concepts introduced and explored in [20] by Chen et al., dynamic clusters will be identified. It will not only take into account spatial proximity between pixels when forming clusters, like the S-STSEC already does, but also temporal proximity when building these, comparing cluster being built with complete clusters of previous time steps.

Compared to spatial data analysis, ST data analysis is tougher and more laborious, since the latter involves analysis of much larger volumes of data and study of dependency patterns in space and in time that are far more complex. A better understanding of these spatial and temporal variations is essential to adequately manage and preserve the ecology and economy of areas near coastal regions. Finding and understanding periodic spatial and temporal behaviors of upwelling regions is also of crucial importance to derive climatological indices and reason about climatic changes.

Consequently, a technique to allow the tracking of upwelling regions from SST remote sensing data will also be developed and tested using images from the Portuguese Upwelling System. Weekly images, which will be an arrangement of 8 daily images, will be used in order to study and track first the seasonal upwelling variations, and then the work will be focused on the tracking of several years of SST images taken of that coast.

1.2 Objectives of the Dissertation

Since there is almost non-existing work in the ST analysis setting where data is in the form of ST fields and the clusters are dynamic, this motivated us in proposing a method to automatically detect and segment dynamic coastal upwelling regions from ST data. Moreover, along with the detection and segmentation process, a mechanism of tracking using this ST information will be proposed, where these regions will be monitored over time, in a seasonal, monthly and yearly time windows.

Following and completing the aforementioned information, the objectives for this dissertation are the following:

1. To implement and test a robust preprocessing pipeline to prevent over segmentation

problems when applying the S-STSEC algorithm and to improve the quality of a set of SST images;

2. To test and assess the quality of the S-STSEC algorithm's segmentations when applied to yearly sets of SST preprocessed images;

3. To use an unsupervised algorithm as a time series segmentation method to be applied to several sets of S-STSEC segmentation results, and test such approach with several validity measures for the evaluation and validation of the obtained time series' segmentation results;

4. To define, propose and implement a new algorithm called Core-Shell clustering algorithm for a new proposed model with the intent of automated detection and posterior characterization of coastal upwelling events;

5. To fine tune hyperparameters of the algorithm to the problem of upwelling tracking, mainly on defining a proper time window to track the upwelling dynamics;

6. To design and develop a proper experimental protocol, including validation, to apply the algorithm of Core-Shell clustering to a representative collection of SST images of the Portuguese upwelling system.

1.3 Organization of the Document

The rest of the document is organized as follows. Chapter 2 gives an introduction to some better known image segmentation techniques, such as clustering, where three main approaches are overviewed, followed by SRG techniques and finally automatic threshold approaches. For each category, several works and studies are presented for various application areas. Still in this chapter, a few supervised and unsupervised approaches for automatic recognition of coastal upwelling are presented. Finally, two approaches for the automatic unsupervised upwelling segmentation are explored, these being the SEC and S-STSEC algorithms [56, 59] by Nascimento et al.. In Chapter 3 an introduction to ST clustering is made, which is a major topic in the area of ST data mining. An overview on data types and clustering methods for each data type is presented, followed by a survey of some applications present in the current literature and finally an introduction to time series segmentation. Chapter 4 introduces the proposed approach for this dissertation, where firstly the proposed preprocessing pipeline is presented, followed by the introduction of a new technique for time series segmentation for the obtention of upwelling spans and the introduction of a new clustering algorithm called Core-Shell clustering algorithm. In Chapter 5, the experimental study is explored, where every important and relevant analysis are made and discussed, regarding the produced segmentations, the Core-Shell clusters derived from such segmentations and time series created with features extracted from both the segmentations and the Core-Shell clusters. The conclusions and work to be continued in the future are available in Chapter 6.

2

Related Work

2.1 Introduction to Image Segmentation

Image segmentation can be defined as the use of techniques or algorithms to divide an image into smaller and meaningful segments, which ideally represent different objects or regions of interest. Image processing strategies and segmentation techniques have become a part of many applications nowadays.

Many fields can take advantage of these kinds of techniques: medical imaging, content based image retrieval, fishing, landscape monitoring, climatic changes, natural disasters, traffic control systems, object detection, machine vision, recognition tasks, video surveillance, and several others. Surveys on various image segmentation techniques have been carried out in several research articles, such as [9] by Anjna and Kaur, [42] by Kaur and [101] by Zaitoun and Aqel.

Supervised approaches can be more accurate than unsupervised approaches, but the former depend heavily on the training phase. Creating training datasets for supervised segmentation requires manual delineation of labels which is labor-intensive and costly for a dataset that is sufficiently large for training purposes. Furthermore, new image types are constantly being developed and discovered, making the creation of the manual training sets needed for supervised segmentation an expensive and ongoing task. Unsupervised segmentation, however, has the advantage of using unlabeled data only to segment images and therefore is helpful in the absence of manually-labeled datasets. Unsupervised segmentation methods are more generally applicable and more robust to atypical or unseen situations [5].

In this Section we focus on three popular unsupervised approaches to image segmentation: Clustering, SRG, and Automatic thresholding. Proposed and published works using one or more of these approaches will be analyzed and summarized, interesting results obtained will be mentioned and problems found of the developed approaches will be discussed.

2.1.1 Clustering

One unsupervised approach that is commonly used in image segmentation is Clustering, which is used to divide pixels into a number of clusters, a number that can be received as an input or can be computed by an algorithm in order to automate the process. Clustering can also be considered as the partition of data sets into subsets, in such a way that the data in each subset should exhibit some common properties [23].

Finding the optimal clusters' number to be used is one of the major challenges in clustering methods. This parameter is one of many used in algorithms of this kind, so a proper initialization needs to be studied and implemented in order to obtain the most optimal results possible.

Of all the types of clustering present in the literature, the most popular partitional clustering methods are generally divided into two main categories: hard clustering and fuzzy clustering.

- **Hard Clustering:** Hard clustering techniques group data in an exclusive way, so that if a certain data point belongs to a definite cluster then it could not be included in any other cluster [15]. In other words, hard clustering is about grouping the data items such that each item is only assigned to one cluster.

- **Fuzzy Clustering:** In fuzzy clustering, fuzzy sets are used to cluster data, so that each point may belong to two or more clusters with different degrees of membership. In this case, data will be associated to an appropriate membership value. In many situations, fuzzy clustering is more natural than hard clustering. Objects on the boundaries between several classes are not forced to fully belong to one of the classes, but rather are assigned membership degrees between 0 and 1 indicating their partial membership [15].

2.1.1.1 K-Means

One of the most known and used hard clustering algorithms is k-means by MacQueen et al. [52], a classic algorithm to solve the clustering analysis problems and widely used in image segmentation. Given a fixed number of K clusters, observations are assigned to clusters so that the means across clusters are as different from each other as possible and the means inter cluster are as similar as possible. This method finds non overlapping clusters, where each cluster is described by its members and by its cluster centroids.

2.1.1.2 Expectation Maximization

The Expectation Maximization (EM) algorithm [22] is an unsupervised method where the maximum likelihood estimate of the parameters is computed from a given data set. The EM algorithm alternates between two steps. The first step is called the expectation process, where missing variables in the data set are estimated, or in other words, an estimation of the probability distribution over completions of missing data given the current model is made. The second step is called maximization process, where the maximum likelihood estimates of the parameters are computed. This last step can also be explained as the re-estimation of the model parameters using the completions of the first step. The most common termination rule used in this method is when the values computed in both steps converge.

2.1.1.3 Fuzzy C-Means

Fuzzy C-Means, presented by Dunn in [25], is an unsupervised technique based on the fuzzy logic clustering that is frequently implemented in the image segmentation field, very similar to the k-means method aforementioned. This method allows each data point to be in multiple clusters with different probabilities assigned to each one. These probabilities are called memberships and are values in a range between 0 and 1. Therefore, each pixel in an image has a membership corresponding to each existing cluster.

2.1.1.4 Clustering applications to Image Segmentation

Clustering is a very useful technique to accomplish several image segmentation tasks. As image clustering is an unsupervised learning based approach, it is capable to handle unlabeled image data [96].

Looking at medical imagery clustering, Ghane et al. [31] proposed an approach that combined Otsu and k-means to extract white blood cells and its nucleus in order to help in the diagnostic process of blood-related diseases. Nasir et al. [61] introduced the concept of variance into k-means, creating an Enhanced K-Means (EKM) clustering algorithm to acquire good segmentation of malaria parasite species from malaria images. Qaddoura et al. [69] presented a method for dental radiography segmentation where the EM algorithm was combined with the grasshopper optimization algorithm [76]. Kaur and Sharma [43] introduced an automatic detection method of tumor regions in brain Magnetic Resonance Imaging (MRI) images, where properties of the segmented regions obtained from fuzzy c-means were used in order to keep or discard segmentations of an image.

Moving onto remote sensing, Niharika et al. [62] presented an algorithm where Otsu was applied on the k-means' output with the goal to solve the speckle noise problem when segmenting Synthetic Aperture Radar (SAR) imagery. Zhang et al. [103] introduced an extension to the fuzzy c-means algorithm for extraction of land cover information from remotely sensed imagery, where a novel fuzzy local similarity measure is defined to replace a fixed parameter in an enhanced version of the fuzzy c-means algorithm. Suji et al. [84] proposed a novel method for polar ice sheet segmentation from radar imagery using a fuzzy c-means approach combined with Gaussian mixture models. Reza et al. [73] introduced a combination of k-means clustering with kernel conjugate gradient algorithm to separate rice grains from background objects using unmanned aerial vehicle images and therefore improve the accuracy of rice yield estimation.

Regarding the industrial sector, Riaz et al. [74] proposed a method where Gaussian filters and k-means were used to detect casting surface defects. Soligon et al. [82] introduced a k-means based approach for corrosion detection in steel power transmission lattice towers. Noh et al. [63] introduced a work for automatic detection of cracks in concrete surfaces, where morphological operations were used to make the cracks more visible and then fuzzy c-means was applied in order to segment these. Zhang et al. [104] presented a robust framework for automatic fabric defect detection based on L0 gradient minimization and fuzzy c-means.

2.1.2 Seeded Region Growing

Seeded Region Growing is a region based image segmentation method, also identified as a pixel based segmentation method, introduced by Adams and Bischof in [3].

In this algorithm, an image is segmented with respect to a set of initial selected points called seeds, from which regions will be grown. Each step of the SRG algorithm consists in adding one pixel to one of the aforementioned regions. In the most simple version of this algorithm, each pixel is assigned, one by one, to the region that minimizes a given similarity condition. When a pixel is assigned to a region, its neighbors that are not inserted into one of these regions will be added to a set in order to be analyzed later. The algorithm finishes whenever every pixel is assigned to a region or whenever a different termination condition is met.

Seed Selection: In order to run a SRG algorithm, there needs to be an initial set of points from which the growth of the regions starts. This initial seed can be selected in two ways:

- **Manual seed selection:** The first way and the simplest one is manual seed selection. This process involves human interaction and requires prior information about the image being segmented. Park in [66] presented a graph method to segment water present in river and reservoir images where prior knowledge was used to manually place the initial seeds. In [100] by Yasar et al., a method to detect cancerous tissue in stomach gastroscopy images was proposed based on SRG, where a seed was selected manually whenever a region stopped growing. In [39], Ijiri et al. proposed a method for computed tomography images segmentation in order to extract 3D heart models. Three types of manually placed seeds were introduced, these being spheres, cylinders and negative seeds.

- **Automatic seed selection:** The second method of seed selection is known as automatic seed selection, where algorithms, some more complex than others, are used to find optimal seeds. An automatic seed selection approach is used in [38] by Huang et al., where semantic segmentation networks were trained to find seed cues, then a method called Classification Activation Map (CAM) was used to select from these the initial seeds for the background and foreground. An example where the authors take advantage of the domain's knowledge is in [56] by Nascimento et al., where the seeds are automatically selected as the coldest temperature pixel in the SST map.

Similarity Condition: Whenever growing a region starting from a seed, a pixel will be added to the region based on a similarity criterion. In [3] by Adams and Bischof, the similarity condition presented was very simple, this being the difference between the gray level of the pixel being analyzed and the mean of gray levels of an adjacent region. In [4] by Afifi et al., the method used was changed in order to avoid the explosion problem when segmenting brain MRI images. Instead of using the mean intensity value of an adjacent region, the intensity value of the initial seed of an adjacent region is used.

2.1.2.1 Applications of SRG

The SRG algorithm has been studied and extended in order to be used in many real world applications.

In remote sensing segmentation, Xianchen et al. [98] used SRG with the goal to segment Digital Elevation Model (DEM) using remote sensing images. Brilliant et al. [17] developed a method to detect land cover changes using land satellite images, where the initial seeds were placed in the maximum gray value on the grayscale converted photos. Zhang et al. [105] had the goal to provide an inexpensive method of fast coastline mapping from remotely sensed imagery in coastal sectors with intense human interference. Instead of using individual pixels to start the region growing step, whole regions automatically computed were used as the initial seeds. Chen et al. [19] presented a method for automatic road extraction from remote sensing imagery using Gabor wavelets to extract Gabor features and given the possibility of objects being present in the roads, connection analysis was performed to the edges in order to fully connect these features.

Regarding medical image segmentation, Fan et al. [27] published a breast cancer computer aided diagnosis where iterative quadtree decomposition was used for the seed selection phase. Taori et al. [90] proposed a method where retina images were used in order to segment and extract the macular region with the help of black and white top-hat functions. The seeds were found automatically with a priory knowledge of the structure of the eye. Seada et al. [77] presented a modified SRG approach for the automatic segmentation of the ascending aorta from Computed Tomography Angiography (CTA) images, where a ray based region growing was used and all the pixels from the seed pixel to the edge of the aorta were analyzed.

Finally looking at color image segmentation, Shimodaira [80] presented a color image segmentation method that combines a SRG method and a merging method based on square elemental regions. Wang et al. [94] introduced the concept of initial regions as initial seeds obtained using a watershed segmentation algorithm proposed in [93].

2.1.3 Automatic Thresholding

Threshold based approaches are among the most popular and commonly used methods due to their simplicity and efficiency, consisting in image pixels being divided using intensity levels of an image. This family of methods is mainly used to distinguish the foreground objects from background ones. These approaches can be divided into 3 main categories:

- **Global thresholding:** A threshold value is computed based on the whole image, and this is the value used to segment the entire image.

- **Adaptive thresholding:** In adaptive thresholding, different threshold values for different local areas are used.

- **Multiple thresholding:** More than one threshold is used in order to segment an image.

One of the most known thresholding algorithms is called Otsu presented by Otsu [65]. In its most basic form, this global thresholding method is used to automatically find an intensity value that best divides an image into two classes. This is determined by minimizing intra-class intensity variance, or in other words, by maximizing inter-class variance.

Other automatic thresholding methods that are used and mentioned in image segmentation are Kittler and Illingworth's method [46], a multiple thresholding approach which is an efficient solution to the minimum error thresholding problem, Ridler and Calvard's method [75], where an optimum threshold is chosen automatically through an iterative process, and Chow and Kaneko's method [21], an adaptive threshold method used to separate desirable foreground image objects from the background based on the difference in pixel intensities of each region.

Automatic thresholding is a very important step in the segmentation of image features. Singla and Patra [81] presented a threshold selection technique for any kind of image where energy curves of the images were used to incorporate spatial contextual information in the threshold selection process along with a genetic algorithm to detect the optimum values for thresholds. Feng et al. [28] introduced a novel thresholding algorithm based on 3D Otsu and multiscale image representation for medical image segmentation. Septiarini et al. [78] proposed a method for optic disc and cup segmentation for glaucoma evaluation based on border mask, Otsu's thresholding and Sobel methods. Lu and Lu [51] introduced a combination of three methods based on bimodal thresholding techniques, Intermode, Ridler and Otsu, for the automatic segmentation of bruises in apples.

2.2 Automatic Recognition of Coastal Upwelling

2.2.1 Supervised Approaches

Supervised approaches are based on the idea that a user can select or target regions of interest in an image and direct the image processing software to use these as training for the classification of other regions in other images. Some of these approaches, such as neural nets, have a good abstraction about the temperature ranges over a time period, yielding consistent classification of temperature bands when compared to individually applied clustering methods [18].

Chaudhari in [18] proposed a novel method for the detection and segmentation of upwelling regions in AVHRR SST data using a feed-forward back propagation neural network. Images with and without upwelling regions were segmented using k-means and the outputs were used to train a neural network. In [67], Piedra et al. implemented and compared a set of supervised methods, including a neuro-fuzzy system and a hybrid system for the automatic recognition of oceanic structures including coastal upwelling. Rojas et al. in [33] introduced a method for automatic recognition of oceanic structures, including coastal upwelling regions. It includes an automatic cloud mask system developed using a neural approach, an iterative segmentation region builder based in intelligent threshold detection and a high-level information classifier which uses functional and operational principles derived from artificial neural networks and rule-based knowledge systems. In [68] by Fernández et al., a set of methods to be used for automatic ocean structure recognition was proposed, such as a neuro-fuzzy system called Nefclass, the fuzzy lattice neural map (σ-FLNMAP) and a proposed hybrid hierarchical model composed by three classifiers, these being two Bayesian Nets and one σ-FLNMAP.

Supervised approaches can be more accurate than unsupervised approaches, but they depend heavily on the training phase. Massive amounts of labelled information are needed to achieve good results when doing classification. This involves major manual work from the user where each data sample needs to be labelled and identified for training purposes, therefore requiring a lot of time and money when compared to unsupervised approaches. Taking as an example [33], it took one full hour to process a single image.

2.2.2 Unsupervised Approaches

When looking at unsupervised approaches, there are no training stages, and therefore it is possible to handle unlabeled image data. This helps to save valuable time when running experiments or doing any kind of research and therefore reduce the amount of data needed to acquire, store and process.

In [57] by Nascimento and Franco and in [60] by Nascimento et al., a novel approach named FuzzyUPWELL for the automatic segmentation and recognition of coastal upwelling was presented using a fuzzy clustering method in conjunction with the anomalous clustering initialization method.

In [85], Tamim et al. had the goal to segment the main upwelling front in the Moroccan Atlantic Coast. Fuzzy c-means was used and was then compared to the more traditional Otsu. In [89] by Tamim et al., fuzzy c-means was used doing purposefully an overestimation. After the segmentation, adaptive cluster merging was applied in order to get the final segmentation. In that same year, [88] by Tamim et al., a two phase approach to segment upwelling regions in the Southern Moroccan Atlantic Coast was presented. In the first phase, an extension of the fuzzy c-means algorithm was used called Gustafson-Kessel (GK) [34], where the Mahalanobis distance was used instead of the Euclidean distance and therefore allowing the detection of clusters with different shapes. In [87], Tamim et al. worked on a segmentation algorithm for the upwelling segmentation using SST images in the Southern Moroccan Atlantic Coast that made use of the expectation maximization algorithm. Later in [86] by Tamim et al., k-means and fuzzy c-means were tested against each other. The number of clusters used was derived from the analysis of 19 validity indices and a merge procedure was introduced using the mean temperatures of the obtained clusters.

In [1], Abidi et al. used both biological and physical satellite images for the segmentation of upwelling regions. Both types of images were used in order to overcome the cloud noise and dust that might cause discontinuity in the images and the temperatures of the water. Fuzzy c-means was applied to both images and these were merged using an AND operation.

In [11] by Aouni et al., the goal was to develop an upwelling segmentation method from SST and SSC images while trying to overcome the problem of over segmentation that usually occurs in the northern Moroccan coast. To solve this, each image was divided into subregions. The Particle Swarm Optimization (PSO) algorithm was then used at each subregion. These subregions were then grouped into a bigger subregion, creating the upwelling region. Aouni et al. [12] tried to improve the aforementioned work. Each image was partitioned into a number of r lines which were perpendicular to the coast, and each line was then normalized with respect to its maximum temperature found offshore. After the normalization, fuzzy c-means was applied to the normalized images. Very accurate and robust results were obtained throughout the entire coast of Morocco, surpassing successfully the explosion problem present in that region and possibly solving this problem in other parts of the globe where the upwelling events are studied.

2.3 A Spatial Clustering Algorithm and its Extension

2.3.1 Seeded Expanding Cluster Algorithm

In 2015, Nascimento et al. [56] proposed a novel seed expanding cluster algorithm for deriving upwelling areas from sea surface temperature images as an extension of the SRG algorithm presented in [3]. Here, the goal was to overcome several issues that came along with using a SRG approach: (i) manual seed selection when initializing all the parameters needed to run the algorithm; (ii) definition of the non-homogeneity threshold used to select which pixels would pass or fail the homogeneity test; (iii) segmentation results being dependent on the order at which the pixels were tested.

Pre Processing: Before starting the algorithm itself, a numerical preprocessing phase is used in order to center the data values. Let $A(L,C)$ be a SST map, with L rows/lines and C columns. The average temperature of the whole image $t^* = mean(A(L,C))$ was subtracted to each pixel (i,j), meaning that the new average temperature for each image will be zero. The new centered pixel values were denoted as $t(i,j)$.

Cluster Initialization: This algorithm will find a cluster Z in the format of a binary map $Z(L,C)$, where each element $z_{i,j}$ takes the value of 1 if $(i,j) \in Z$ or 0 if $(i,j) \notin Z$. The algorithm also uses a window W where $W(i,j)$ is a squared window of the size 7 x 7 centered at the pixel (i,j).

After the numerical preprocessing phase, the algorithm starts by selecting a single seed pixel $o = (i_o, j_o)$. This single seed is the pixel with the lowest temperature value of the entire image. If there are several pixels with the same temperature value, the seed pixel o is selected randomly. The cluster Z is therefore initialized as a set containing the initial seed o and any other pixel within the window $W(i_o, j_o)$ that satisfies the similarity condition:

$$c \times t(i,j) \geq \pi, \tag{2.1}$$

where c is the temperature of the initial seed pixel o and π is the similarity threshold, which will be later discussed.

Boundary Cluster Initialization: Alongside the initialization of the cluster Z, a set F referred as the cluster boundary of Z is also initialized, and it will contain pixels that have their 8 neighborhood intersecting at least one pixel that is in Z. This boundary set F is defined as:

$$F = \{(i', j') \notin Z \mid N(i', j') \cap Z \neq \emptyset\}, \tag{2.2}$$

with $N(i', j')$ being the 8-neighborhood of the pixel (i,j).

Region Growing Process: The algorithm then starts growing the cluster Z iteratively until the boundary set F becomes empty or this doesn't change between two consecutive iterations. For each boundary pixel $(i', j') \in F$, a boundary expand region is defined as the set of pixels $(i,j) \in Z$ that intersect the exploring window $W(i', j')$, or in other words, the boundary expand region is defined as $(i,j) \in W(i', j') \cap Z$, and the average temperature of this region is defined as c^*.

For a boundary pixel $(i', j') \in F$ to be added to the cluster Z, it needs to satisfy two homogeneity conditions. The first condition, called the temperature similarity condition, is as follows:

$$c^* \times t(i', j') \geq \pi \tag{2.3}$$

The second homogeneity condition is a density condition defined as:

$$\frac{|W(i',j') \cap Z|}{|W(i',j')|} \geq \alpha, \tag{2.4}$$

where $|W(i',j') \cap Z|$ is the density of points at a boundary pixel (i',j'), defined as the number of pixels inside the window $W(i',j')$, centered at the pixel (i',j'), that intersect the cluster Z. This condition is used to ensure the cluster Z being built has a smooth shape without irregular fragments.

If both conditions are satisfied, the point (i',j') is added to a secondary set Z' and its 8-neighborhood $N_8(i',j')$ that does not already belong to the cluster Z is added to a secondary set F'. Whenever F becomes empty, meaning that an iteration ended, all the pixels in Z' are moved into Z, all the pixels in F' are moved into F and a new iteration begins. At the end of an iteration, if no pixel satisfied both homogeneity conditions and Z' is empty, the algorithm stops.

2.3.1.1 Self Tuning SEC

In the base version of the SEC algorithm, the threshold π used in the eqs. (2.1) and (2.3) is given to the algorithm as an input and is used as a constant throughout the whole run.

To solve that and in order to automatize the process, a self tuning version was implemented, ST-SEC, where the threshold is derived from the cluster approximation criterion. Here, the threshold π for the similarity condition is derived as being the half squared average temperature over the cluster Z intersecting with the window W centered at a boundary point (i',j'):

$$\pi = \frac{mean(T(W(i',j') \cap Z))^2}{2} \tag{2.5}$$

A more or less smooth shape of the growing region is warranted by the averaging nature of the similarity criterion and by involving windows around all pixels under consideration. This new structure of the clustering process allows to abandon the density condition (2.4).

2.3.1.2 Experimental Results

When experimenting, the results obtained from the base SEC and the ST-SEC versions were compared with manual segmented reference images, since there is no generally accepted methodology for evaluating segmentation algorithms.

The base SEC algorithm was tested using two methods to define the threshold π in both eqs. (2.1) and (2.3). In the first method called the supervised approach, a range of values between 0.01 and 0.15 was tested by running the algorithm several times with an incremental step of 0.01, and the value that maximized the F-score was used as the π threshold. For the unsupervised approach, Otsu's method was used. In this case, the optimum threshold value was fixed as the one presenting the minimum intra-class variance when dichotomizing the pixels in a bimodal histogram.

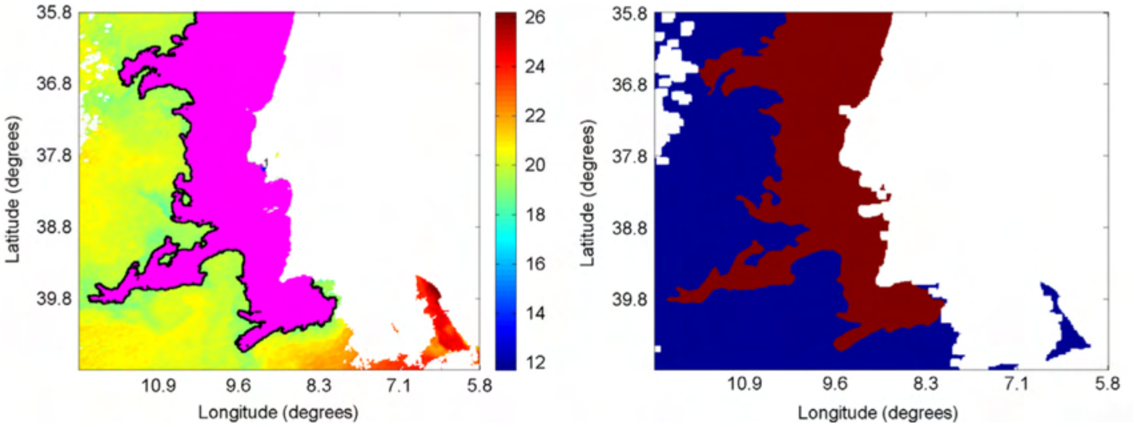


Figure 2.1: Upwelling area retrieved by the SEC algorithm with supervised thresholding (left column) and binary ground-truth map of that image (right column) [56].

The base SEC algorithm with supervised tuning of the parameters got 93% of the F-scores between 0.768 and 0.985, having only 2 images with F-scores lower than 0.7. The base SEC algorithm with Otsu's method got 82% with F-scores between 0.7 and 0.977, having 5 images with F-scores lower than 0.6. The final version of the algorithm, with the self tuning of the parameters, got 75% of the tests with a F-score higher than 0.7.

The images that got lower F-scores were images that either had discontinuity along the upwelling region or images that obtained over segmented results.

2.3.2 S-STSEC Algorithm

In 2020, Nascimento et al. [59] presented a new approach that extended the previous algorithm ST-SEC, explained above. The method discussed here, named S-STSEC, is a sequential version of the ST-SEC that extracts clusters one by one until a stop condition is reached, without providing the number of final clusters.

2.3.2.1 Improvements to the ST-SEC version

This extension was proposed in order to overcome the possible discontinuity along the upwelling region since coastal upwelling can cover two or more contiguous regions.

Instead of looking at the 8-neighborhood when selecting the pixels to go into the set F, the 4-neighborhood was instead used since it showed to produce better segmentation results. Therefore, when defining F, $N(i', j')$ will correspond now to the 4-neighborhood.

Unlike the prior version, the algorithm does not stop when no pixel satisfies both homogeneity conditions and C' is empty. This version will keep building clusters one by one until a different termination condition is reached. This condition is defined based on a priori knowledge of the studied domain. Another difference is that candidate seed pixels are the coldest pixels that do not exceed a predefined distance δ to the coastline.

Other than that, it was experimentally verified that the difference between the average temperature of the first extracted cluster $mean(C^1)$ and the lowest temperature value of the current cluster $t(i_o, j_o)$ decreases with the increase of the number of clusters. A second sub-condition was then established which guarantees that the aforementioned difference can not be less than a threshold τ, automatically calculated by the Otsu's method.

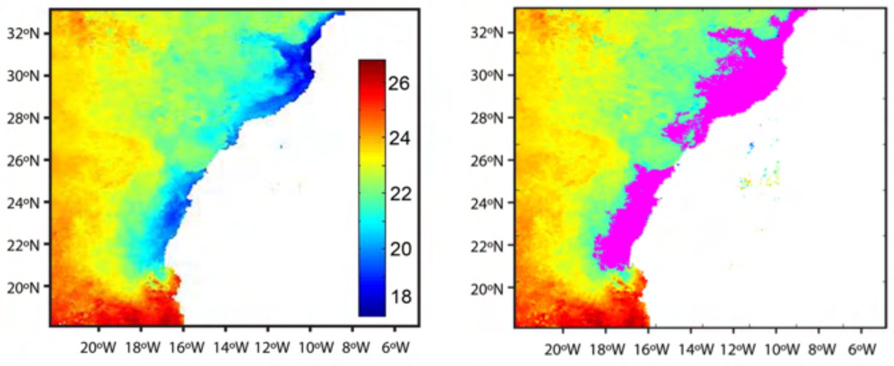


Figure 2.2: SST image of Canary System (left image); Upwelling area retrieved by the S-STSEC when dealing with discontinuities (right image) [56].

The algorithm has shown to be effective and efficient in automatically delineating continuous and non-continuous coastal upwelling regions. This version showed good and precise results in more than one region, these being in the coastal upwelling of the Portuguese coast and in the upwelling system of the Canary system, as it can be seen in Figure 2.2. Therefore, this method tackles one major drawback of the majority of the approaches presented until the present, which is approaches being specific to only one region and not working in other regions than the one where the study focused on. The leaking or explosion problem is still up to further investigation.

2.4 Coastal Upwelling Tracking

A better understanding of spatial and temporal variations is essential to adequately manage and preserve the ecology and economy of areas near coastal regions. Finding and understanding periodic ST behaviors of upwelling regions is also of crucial importance to derive climatological indices and reason about climatic changes.

In [71] by Ramanantsoa et al. and [53] by Marcello et al., two approaches were presented for the understanding of the seasonal and interannual variability of coastal upwelling in Madagascar and West African coast, respectively. Both methods used a Coastal Upwelling Index (CUI) to determine the intensity of the upwelling regions studied, being such index defined as the difference between coastal water and offshore water at the same latitude.

In [64], Oerder et al. proposed a method to automatically identify the front and define characteristics of coastal upwelling regions from SST images in Central Chile. The upwelling detection method is rather simple and region specific, where the main front is detected if the difference between the coastal (at 1km of the coast) and offshore (at 300km of the coast) mean latitude's temperatures is greater than 1 degree. Broad upwelling regions can't be detected with this approach since the aforementioned condition can not be satisfied. The intensity analysis is handled in the same way as the 2 aforementioned methods. It can only handle an analysis in a 0.5º area. The upwelling regions are not segmented in detail here, only the front is detected as being the isotherm corresponding to the mean SST over the frontal zone detected. Doesn't work when upwelling regions are broad neither when noticeable cloud noise is present.

In [37] by Huang and Wang, a semi-automatic technique to map upwelling areas was presented and further analysis of its behavior was studied. A combination of the Topographic Position Index (TPI) and the multi-resolution algorithm was used in order to segment the upwelling regions, where good and precise segmentation were obtained. Area of influence, SST anomaly and upwelling speed were the upwelling characteristics studied in this work. The upwelling index was also analyzed in order to study mechanism behind the phenomena and wind data was taken into account to compute this value. Month to month, seasonal and yearly variability was studied here. Some level of human intervention is needed in the selections of TPI window size, segmentation parameters and SST threshold, and the removal of a few upwelling candidates.

All the aforementioned techniques presented are simple in terms of implementation and concepts, causing them to be too region specific and therefore needing to set too many parameters manually as well as possibly giving imprecise results due to the way fronts are detected in the analysis. Another weak point of these approaches is the lack of relationships between the upwelling regions detected over time, not taking into account how several upwelling regions evolve. This is really important given the dynamic characteristics of the event being studied.

3

Spatio-Temporal Clustering

Spatio-temporal clustering is a major topic in the area of spatio-temporal data mining. Spatio-temporal data mining is a powerful approach with the potential to extract useful and unknown information from immense amounts of data. Spatio-temporal data mining is becoming more important every day in the big data era that we live in. Due to the low cost of geographic positioning sensors, more and more applications are being produced and engineered. This allows to collect the usual data along with spatial and temporal information, which is called ST data. This data is becoming increasingly available in many forms like maps, virtual globes, remote-sensing images and GPS trajectories [95].

Spatio-temporal data is data that is collected along with spatial and temporal information, which is information that is always present in any kind of measurements made [13]. This extra spatial and temporal information introduces dependencies among measurements, allowing a new kind of analysis to be made, and new results can be perceived and discovered. When working with this kind of data, it is really important to not forget about these new properties or else it can lead to misinterpretation of results when used with traditional approaches.

Common methods of interpreting this kind of data are clustering, visualization of data, classification, forecasting and anomaly detection, which can be applied in several application domains namely environment and climate vigilance, public safety, human mobility and many more.

3.1 Introduction

The presence of space and time introduces a rich variety of kinds of ST data types and representations, varying depending on the application domain. There are two main characteristics of ST data.

The first main characteristic of this kind of data is auto-correlation. Measures made in a spatial and / or temporal proximity are correlated with each other. This can cause traditional data mining approaches to malfunction, because many of these have the assumption of uncorrelated or independent data.

The second main characteristic is related with heterogeneity. Another assumption made by traditional data mining approaches is that every instance belongs to the same population and hence is identically distributed. ST datasets can show heterogeneity both in space and in time.

3.1.1 Data Types

The classification of data types is based on two dimensions. The first is the spatial dimension, and this describes if a given data object has a fixed location or if its location is dynamic. The second one is the temporal dimension, and it describes if an evolution of the data is captured. ST data can be classified into five different types: event data, trajectory data, geo-referenced data items, geo-referenced time series and moving objects [10].

- **Event data:** Event data can be characterized by where and when an event occurred. Both spatial and temporal information is static since no evolution nor movement is present in these measures. Apart from these two characteristics, event data can have other variables called marked variables, which provide additional information about the given event. A collection of ST events is called a spatial point pattern.

- **Trajectory data:** Trajectory data can be described as the path that a given body takes in space over a period of time. This data may not only have the path information throughout time, but also associated marked variables to further characterize the trajectory.

- **Geo-referenced data item:** A geo-referenced data item is usually the measure, at a fixed location, of a given characteristic over a period of time. In this kind of data, usually only the latest measure is kept.

- **Geo-referenced time series:** A geo-referenced time series is the measure of a continuous ST field over a set of fixed reference points over a period of time. The fixed locations can be organized like a grid, e.g. pixels in an image, or can be placed in a random spatial pattern. Observations are recorded at a set of time stamps, which can be regular delays between each recording or can be irregularly spaced

- **Moving objects:** A moving object changes its spatial location over a period of time. In this kind of data, the most recent data observations are maintained while the past observations may or not be forgotten. Usually a moving object has an id in order to trace the movement of a particular individual.

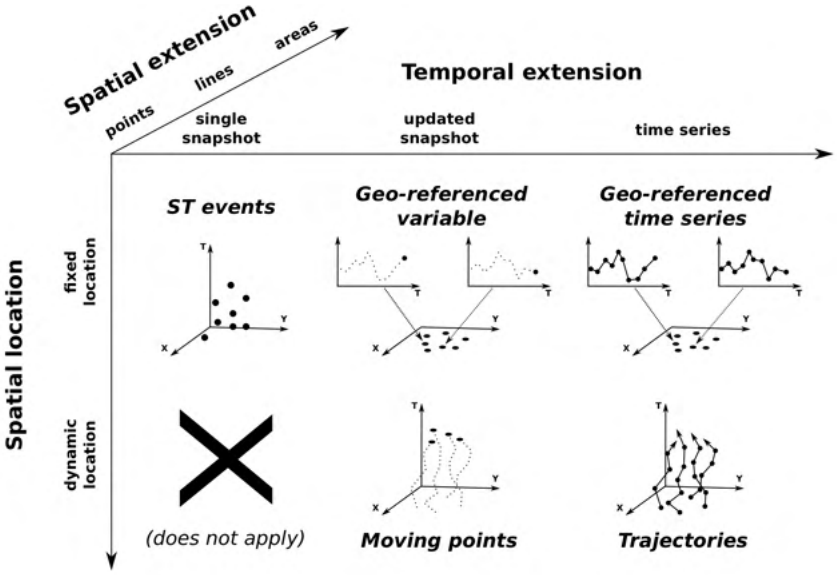


Figure 3.1: Overview of ST data types. Image taken from [45]

3.1.2 Clustering Methods

Spatial and temporal aspects introduce a novelty when clustering ST data instances. When doing so, pre-processing of data needs to receive a lot of attention. The possibility of existing irrelevant attributes might change the results when looking at spatial or temporal distances. When looking at the different kinds of data instances, one may suggest that there are different goals and approaches to cluster each kind of data.

- **Event Clustering:** One goal of event clustering is to find hot-spots: clusters with a high density of ST event points. This problem is known as event detection, and it consists in finding data points that are close to each other in terms of spatial position and time. Another goal of this kind of clustering is to find clusters of points that have similar non ST attributes.

- **Trajectory Clustering:** When clustering trajectories, the goal can be to group trajectories with similar paths or even group trajectories that have similar sub-paths between them. Two important points of this kind of clustering are the choice of the distance method used and the choice of the clustering technique.

- **Geo-referenced data items clustering:** Clustering geo-referenced data items consists in finding groups of objects with similar spatial attributes at any given time. It can also be used to find groups of objects similar among them regarding non-spatial attributes at any timestamp.

- **Geo-referenced time series clustering:** One of the main goals of clustering time series is to find spatial locations with similar temporal activity. Another goal of this kind of clustering is to group objects based on spatial closeness and associated time series similarity among objects.

- **Moving clusters:** In moving clusters, the identity of a moving cluster does not change although its location and content may change over time [10]. One main idea of this kind of clustering is to find a majority of points that are similar to each other and therefore compose a cluster, and then at each timestamp, evaluate how this cluster's content and/or location changes without having to recompute the whole clustering process again, by using information from previous time steps.

3.1.3 Domains of Application

Surveys on this relatively young field of research have been made in [13] by Atluri et al., [10] by Ansari et al., [79] by Shi and Cheng, where a state of the art is presented, and many application areas are discussed. Some main sectors where ST data mining and clustering are employed are climate science, neuroscience, epidemiology, traffic dynamics and crime data.

Climate Science: The understanding and monitoring of earth science sensor datasets offer unique data-centric and algorithmic challenges imposed by the volume, variety and richness. In [14] by Birant and Kut, ST-DBSCAN was introduced, a density-based clustering algorithm based on DB-SCAN that creates clusters with arbitrary shapes based on non-spatial, spatial and temporal attributes. In [6] by Agrawal et al., ST-OPTICS was proposed where micro clusters are created in a clustering phase and are then merged in an agglomerating phase. In [48], Liu et al. proposed the DcSTCA, an approach to detect and cluster marine anomaly variations. Here, a ST grid cube formulation is used, which takes into account spatial adjacency, attribute homogeneity, and time continuity of marine anomaly variations.

Neuroscience: Continuous neural activity captured using a variety of technologies is commonly studied in neuroscience [13]. In [24] by Doborjeh and Kasabov, a method for clustering of ST brain data based on NeuCube spiking neural network architecture was presented. Clustering is performed using spiking neural networks and a functional analysis between the clusters obtained is performed in order to understand brain functional dynamics.

Epidemiology: Traditional epidemiology surveillance comprises a set of epidemiological procedures that monitor the spread of a disease and determine how it is spreading [32]. In [32] by Gomide et al., a method that accomplished active surveillance of Dengue epidemic looking at Twitter information was proposed using regression models and ST-DBSCAN. In [7], Alkhamis et al. explored the temporal and spatio-temporal dynamics in the beginning of the COVID-19 pandemic in Kuwait using the effective time-dependent reproductive numbers model, a likelihood method used in the SARS pandemic in 2002, and multi variable permutation scan statistic implemented in SatScan.

Traffic Dynamics: Looking at traffic dynamics, in [29] by Fiori et al., a novel approach called DeCoClu was developed to overcome the lack of network topology knowledge regarding position signals of buses and bus stops. In [35] by Higgs and Abbas, research was presented to investigate characteristics of different driving behaviors linking drivers' states to actions performed by them using segment lengths and segment centroids for the segmentation phase and k-means for the clustering one.

Crime Data: Finding crime hotspots and related crime risks and interactions between crimes are some results obtained when applying ST data mining in this field. In [55] by Nakaya and Yano, a novel approach for interpretation of ST patterns of crime hotspots in Kyoto was proposed, by creating a three-dimensional mapping of crime events in a space-time cube using ST data. In [36], Hu et al. proposed a ST framework for predictive hotspot mapping and evaluation of crimes in Louisiana, where ST kernel density estimation was used to estimate density of an event and predictive accuracy index was used to evaluate predictive accuracy of the identified crime hotspots.

3.2 Dynamic Clustering

A common clustering problem that is of interest when dealing with ST data is to identify subregions that show coherent measurements in both spatial and temporal dimensions, termed dynamic ST clusters[20]. In other words, dynamic ST clusters can be defined as a sequence of spatial clusters that appear in consecutive snapshots such that two consecutive spatial clusters share many common objects [41].

The discovery of dynamic ST clusters can help in detecting phenomena that only influence a subset of locations during a subset of time points. For example, water bodies that grow and shrink in space over time, e.g., lakes and reservoirs, can be identified as dynamic ST clusters in remote sensing data, as they appear as coherent observations in subsets of space and time. Note that a dynamic ST cluster may evolve over time and thus change its shape, size, and appearance as we progress in time. Hence, while some locations are retained across consecutive time stamps, the cluster assignments of locations are dynamic and the clusters can grow and shrink over time [13].

It is known that in several dynamic settings, there's an evolution occurring in the clusters' compositions, but there is usually a set of points that does not change its membership for a long period of time.

In the work by Kalnis et al. [41], the concept of dynamic clustering was applied to moving groups of animals. In this context, during the movement of a group of animals, some animals might leave the group and others might join the group, but the majority of its composition will not change for a long period of time. Clusters were found at each time step using DB-SCAN and Jaccard index was used to evaluate similarities between clusters of consecutive time steps, and therefore verifying if a cluster of a time step s is an extension of a cluster detected at the time step $s-1$.

3.2.1 Dynamic Density-Based Clustering

In [20] by Chen et al., a novel approach to identify dynamic ST clusters for detecting surface water dynamics was proposed, while helping to deal with missing data and noise, which are some of the biggest challenges when dealing with ST clustering in real world applications.

The first step of this method consists in finding points that for a certain time don't have their membership changed and therefore have the most certain cluster memberships, called core points. These core points compose steady clusters, which are identified using an extension of the traditional DB-SCAN algorithm. The goal is to associate pairs of points that have similar feature values over a non-trivial time window. These are denoted as ϵ-neighbors, points with spatial adjacency and with a temporal distance between them smaller than ϵ. If a point has more than m ϵ-neighbors, then it is considered a core point. Then, a second step is used where the uncertain points are labelled, these being most likely at the clusters' boundaries where dynamics occur.

This approach considers two main aspects these being the definition of temporal similarity and a method for growing and refining clusters, which will be described next.

3.2.1.1 Temporal Similarity

Two time series are considered similar if they have the same expected value for a given time interval. There is the assumption that real data observations $x(t)$ are the result of a sum between the true value of the data $\hat{x}(t)$ and a white noise $n(x,t)$. This sum is called the additive white noise model:

$$x(t) = \hat{x}(t) + n(x,t)$$

When two objects, x and y, are temporally similar, their true value is identical and therefore:

$$x(t) - y(t) = n(x,t) - n(y,t)$$

Assuming the white noise model, the expected noise value is zero and for this reason the estimation of temporal similarity between two time series x and y can be given by the p-value of the test:

$$H0 : E(x - y) = 0$$

$$Ha : E(x - y) \neq 0$$

When data is susceptible to outliers, instead of the expected value, an alternate test using the median can be used. Under the white noise assumption, two time series are similar if the difference between them given a time interval is centered around zero. Then the p-value of this hypothesis test can be used as the similarity measure.

Even if the target clusters' and background's temporal profile differ from each other only during some small period of time, the temporal similarity over a long period of time introduced in this method will help to deal with ST heterogeneity.

The temporal similarity measure is very sensitive to the time window w used. If w is equal to 1, this approach would be similar to traditional approaches that do not take into account time. If the w value is too high, the uncertainty introduced would be higher from the highly variable data signal or the changing properties of the dynamic cluster, resulting in a higher quantity of uncertain points.

3.2.1.2 Growing and refining clusters

After applying the two steps mentioned in Section 3.2.1, these being the labelling process of core and uncertain points, there might be points that still have no label assigned, thus being called uncertain points. These points need to be assigned to clusters and errors that might have happened in the previous steps need to be corrected.

At a certain time step t, the cluster assignments from the previous time step $t-1$ are used to build a spatial predictive model. This model is used to predict cluster memberships of uncertain points based on their spatial proximity and feature values at the current time step t. Assuming that k stable clusters were found in the time step $t-1$, a k class classifier is built to estimate the probability of an uncertain point being a part of one of the k stable clusters at the current time step t. When classifying an uncertain point, if it has the same label as its neighbor from a stable cluster, the label is accepted and the next uncertain point is analyzed. If the uncertain point has a different label than the neighbor from a stable cluster, this point is then classified as unknown, delaying the decision until there is more information about other neighboring points.

For this approach, a local Bayesian classifier was used and for each neighborhood, a Bayesian classifier was trained. The data in the same cluster is considered to follow a normal distribution and its conditional probability for any cluster C is:

$$p(x \mid x \in C, t) \sim N(\ \mu(C_R, t)\ ,\ \sigma(C_R, t)\ )\ ,$$

where $\mu(C_R, t)$ is the mean of all points belonging to the cluster C in a region R, centered in the point x at time t, and identically $\sigma(C_R, t)$ is the standard deviation of all the points belonging to the cluster C in a region R. The prior probability of $p(x \in C \mid t)$ is given as:

$$p(x \in \mathrm{C} \mid t) = e^{-\frac{\|x, C\|_{space}}{\delta_d^2}}\ ,$$

where $\|x, C\|_{space}$ is the spatial distance between the point x and the cluster C. The parameter δ_d^2 controls the weight of the spatial constraint. As this value increases, the smaller impact the spatial distance has on the prior probability.

Concluding this step, each uncertain point is finally assigned to the cluster with the highest probability.

When analyzing the results of this approach and comparing them with two other approaches, Normal-cut which is an image segmentation technique and Gao et al.'s method, a k-means based approach, all three methods performed similarly well when data had small amounts of missing data. As the amount of missing data increased, the method here presented outperformed the other two.

3.3 Overview of time series segmentation

One of the most frequent forms of presenting temporal data is through time series [50]. A time series is an ordered sequence of data points measured over time, where each data point can be of univariate or multivariate types.

When analyzing time series with the goal of finding patterns or changes in the data's behavior through time, one can approach such a challenge by finding a certain group of points called change points. A change point is a point located where the behavior of a given time series changes. In other words, change points can be seen as abrupt variations in a time series where transitions between states occur.

Change Point Detection (CPD) is the process of finding abrupt changes in the data being studied where some property of the time series' data changes. Such problem can also be denominated as edge detection or event detection [8].

Looking at CPD approaches, the number of change points to be detected can be known a priori or the assumption of an unknown number of change points in the data can be made.

Another way to achieve the same goal as the one aforementioned is by doing time series segmentation, where a time series is segmented into different segments without purposely finding the change points' locations. An example of change points and obtained segments can be seen in Figure 3.2.

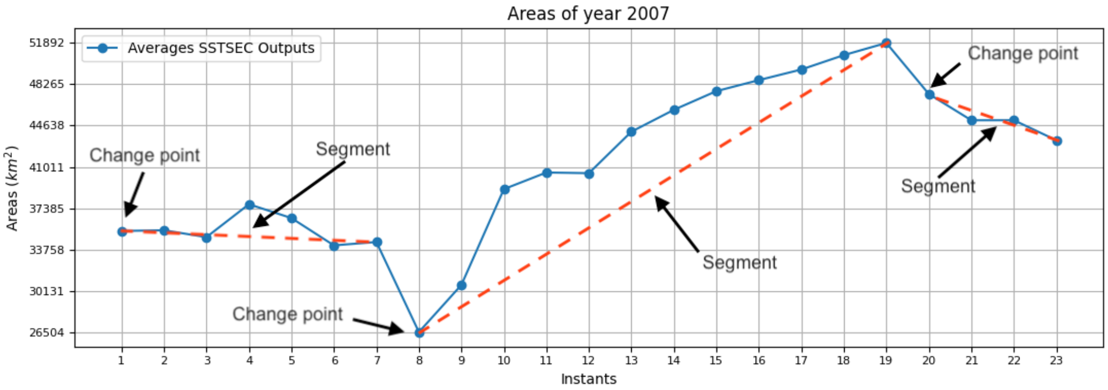


Figure 3.2: Example of change points and time series segmentation results

Such segments or change points' locations can be detected using online approaches, where a real time stream of data is being analyzed as it is collected, or using an offline approach, where the scanning and division of the entirety of a time series into a number of segments or the detection of change points' locations is performed while the available number of data points remains unchanged during the execution time of the algorithm [50].

Taking such characteristics and descriptions into account, time series segmentation and CPD methods can be divided into two major categories, these being supervised and unsupervised approaches.

Supervised Approaches: Supervised time series segmentation methods can be defined as machine learning algorithms trained for binary or multi-class classifiers for the detection of single or multiple change points.

Such methods can be multi class classifiers, to which decision trees, nearest neighbor, support vector machine, bayesian nets, hidden markov models and gaussian mixture models approaches belong to; binary class classifiers, which can be divided into support vector machine, naive bayes and logistic regression approaches; and finally virtual classifier methods.

Unsupervised Approaches: Unsupervised methods, however, have the advantage of not requiring training data, thus being more generally applicable and more robust to atypical or unseen situations. Given the study being conducted in this dissertation, this kind of approaches is of greater importance and relevance.

A great amount of approaches uses statistical models to predict and estimate change points positions, where probability functions and distributions are analyzed regarding candidate change points. Such methods can be divided into several approach types like cumulative sum, change finder and semi-parametric log-likelihood change detector methods. Other approaches in the same realm of application where there is the estimation of some sort of parameter are bayesian methods. A final probabilistic method usually utilized in this kind of problems is the gaussian process, where normal distribution predictions are made for CPD [8].

Kernel based methods where the homogeneity of data is tested using sliding windows and graph based methods where a generalized dissimilarity measure is used to compute and create a graph structure are two other techniques explored for CPD and time series segmentation.

One kind of approach that is of great interest to be studied and discussed is the use of clustering techniques for time series segmentation. Rakthanmanon et al. in [70] introduces a subsequence grouping based on minimum description length for data extraction and posterior bit stream bottom-up clustering of the extracted data, where the clustering algorithm is based on three operators: create, add and merge. In [92] by Tran, an online change detection algorithm using reactive clustering with the use of two moving windows is introduced, where a change point is detected whenever an instance or portion of data does not belong to the current cluster being built and where K-means is used along with the euclidean distance for the clustering procedure. Bose and Mali [16] and Song et al. [83] introduced two time series segmentation approaches based on rough Fuzzy C-means, noting that the latter approach used L2-norm to describe the temporal relationship between mining machinery data and a Lagrangian multiplier method to minimize the segmentation objective function and thus obtain an optimal segmentation result.

Based on and adding to the concepts aforementioned, there’s other three types of unsupervised segmentation methods of great importance, these being the top-down, bottom-up and sliding window approaches [50].

Top-Down Approach: A top-down approach, also referred to as divide and conquer method, can be described as a set of methods where the starting point is a complete time series. The first step is defined as dividing a time series into two segments where these segments are as different as possible form each other considering a certain metric. After this, both of the sub-segments are then tested from the aspect of the level of the approximation error. If the approximation error of the observed segments is above a user-defined threshold, the segmentation procedure stops, and the tested segments are accepted. If not, each sub-segment is divided and processed like the original time series was. The division stops until a requested number of segments is found or the approximation error of the observed segment is over the user-defined threshold.

Bottom Up Approach: Moving to the bottom up approach, which is also referred to as the iterative merge, is an algorithm where firstly the original time series is divided into very small segments with equal lengths. After this initial step, and taking into account an error measure, the two consecutive segments that cause the minimal increase in the error are merged into a bigger segment. The merging process stops until a requested number of segments is found or the approximation error of the observed segment is over the user-defined threshold.

Sliding Window Approach: Ending with the sliding window algorithm or one pass algorithm, this online method iterates the entirety of the time series from start to finish. When a specified error measure for the current segment being built is greater than a user defined threshold, that segment is defined as being a final segment, and a new segment starts to be built. These steps are repeated until there are no more data points available.

Combining the bottom up and the sliding window algorithms, Keogh et al. in [44] introduced the SWAB algorithm, which combines the online aspect of the sliding window algorithm with the accuracy of the bottom up approach to perform piece wise linear approximation. Such a method only requires constant space and produces high quality linear approximations of the data.

4

Proposed Approach

This chapter is focused on presenting the proposed approach for this dissertation. In section 4.2, a new preprocessing pipeline will be presented, which has the goals of overcoming the over segmentation problem as mentioned in previous Sections and preparing the data set for the dynamic clustering procedure. Following this, in Section 4.3 an overview of extracted features from the obtained S-STSEC segmentations of the fully preprocessed SST images is made, for the unsupervised definition of several upwelling time periods where the upwelling activity follows alike patterns which is explained in Section 4.4. After this, in Section 4.5, the developed algorithm for the construction of Core-Shell clusters is discussed. In Section 4.6, an explanation of features extracted from the obtained Core-Shell clusters is made, in order to finalize the overview of the entirety of the developed workflow.

4.1 Proposed Workflow

The entire developed pipeline can be analyzed in Figure 4.1. In summary, a set of N SST images referent to an upwelling season are preprocessed using the preprocessing pipeline described in Section 4.2. The SST maps resulting from this are called SST instants, and these are given as input to the S-STSEC algorithm in order to extract the correspondent coastal upwelling regions.

With the obtained coastal upwelling regions, several features are extracted and time series of these are created, in order to perform an analysis of the upwelling regions through time and also to perform time series segmentation. This segmentation allows grouping consecutive SST instants with similar coastal upwelling behavior, groups that are called upwelling spans.

Then, T SST instants referent to each upwelling span are used as input to the Core-Shell clustering algorithm, which produces a new segmentation called Core-Shell cluster. These structure allow to analyze a more constant region of a coastal upwelling event called a Core and to study the dynamics of the coastal upwelling present in that upwelling span called Shells. To do so, several features are extracted from these new structures in order to analyze such and study how the constant and dynamic portions of a coastal upwelling event evolve over time.

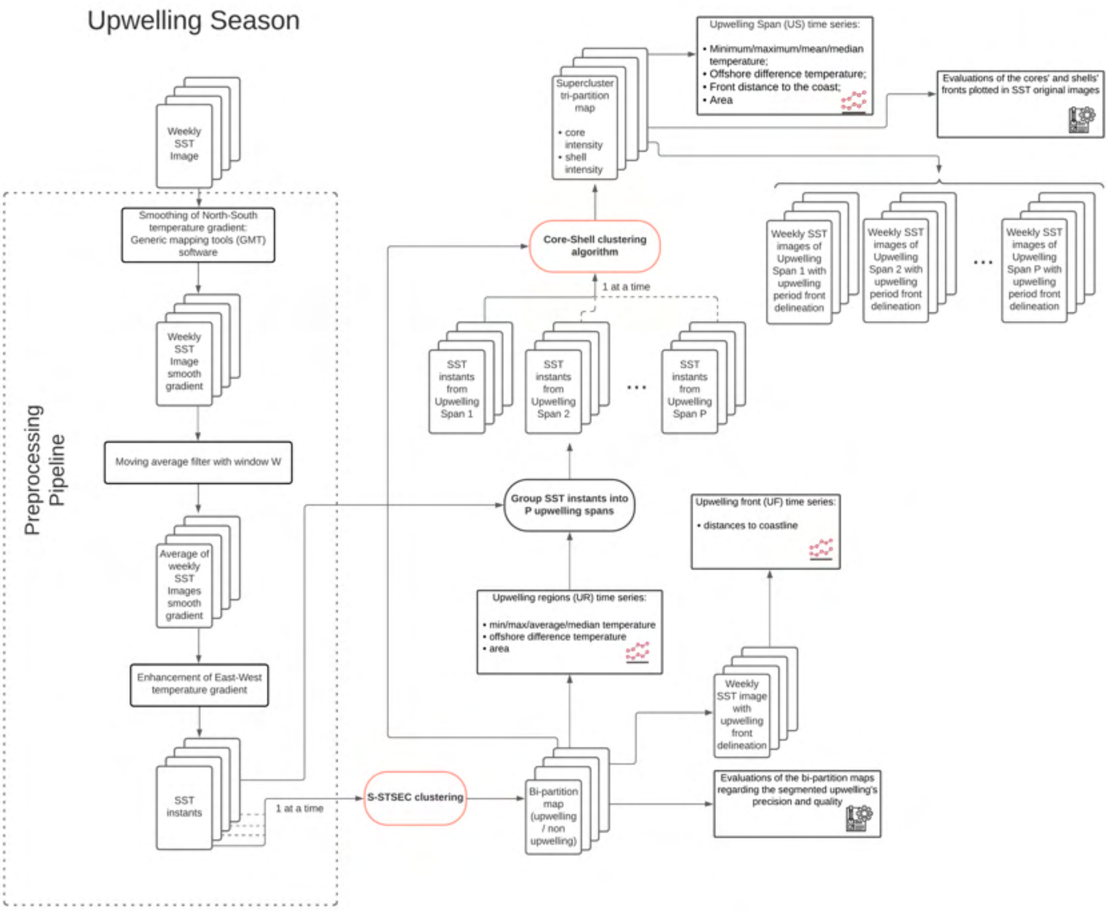


Figure 4.1: Full experiment pipeline

4.2 Preprocessing Methodology

Image preprocessing is the processing of images using one or several algorithms in order to prepare a set of images to be used in a certain context. Being more specific, in the present study, image preprocessing was used to overcome a very common problem which is known as the explosion problem observed along several studies [12, 56, 2], where over segmented results are obtained by a given segmentation algorithm. Such processing was also used to manipulate a set of SST imagery data with the goal of improving the images' quality while concealing any unwanted deformation of the data itself.

In this Section, a full preprocessing pipeline was developed with the intent of preparing a set of SST images to be used by the newly developed Core-Shell clustering algorithm. This pipeline can be seen in Figure 4.2. With the set of weekly SST images received as input, the first phase is the smoothing of the north-south temperature gradient, explained in Section 4.2.1, then a moving average filter is used in order to produce a set of SST averaged images, explained in Section 4.2.2, and finally the images go through a temperature preprocessing method which enhances the east-west temperature gradient, which is described in detail in Section 4.2.3. This temperature preprocessing method was developed to overcome the existing over segmentation problems of the S-STSEC algorithm whenever the front of upwelled waters was not well-defined.

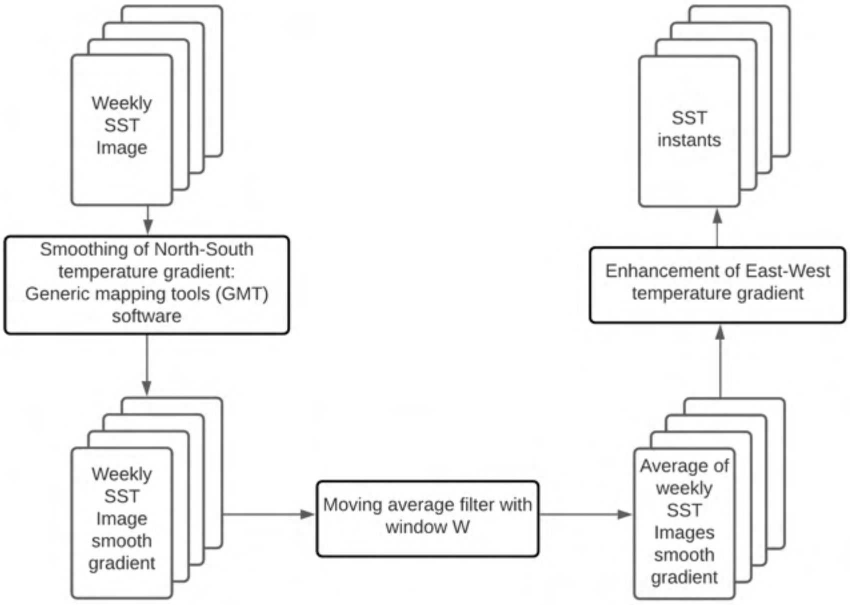


Figure 4.2: Entire preprocessing pipeline

4.2.1 Smoothing of North-South temperature gradient

In order to remove the north-south gradient present in the Portuguese SST images and thus help to improve the segmentation results, Generic Mapping Tools (GMT) [91] was used with the Julia programming language. GMT is a set of modules widely used to process geophysical data and project maps mainly present in Earth, Ocean and Planetary sciences.

To do so, the grdtrend() function from GMT was used, which fits a low order polynomial to the data provided. The number of model parameters N to be used was set to 3 in order to only use linear parameters. When applying such transformation to the original Portuguese SST images, the north-south gradient is removed as it can be seen in Figure 4.3.

In order to evaluate if this transformation helped in the segmentation of an upwelling region using the S-STSEC algorithm, this algorithm was used to segment one image twice. The first segmentation was made using the original SST image without any transformation. The second segmentation was done on the same image with the north-south gradient

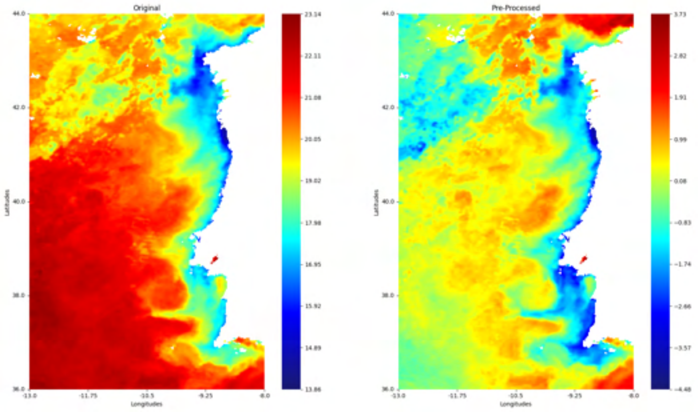


Figure 4.3: Original image on the left and the same image after smoothing the north-south temperature gradient on the right

removed. As one can observe in Figure 4.4, the segmentation using the image without the north-south gradient, this being the Figure on the right, presents a more precise result when compared to the middle image, avoiding some over segmented artifacts present in the segmentation with the original image.

These results were observed in the majority of the segmentations, which allows to conclude that the north-south gradient smoothing helps with the over segmentation problem aforementioned, therefore improving the segmentation results.

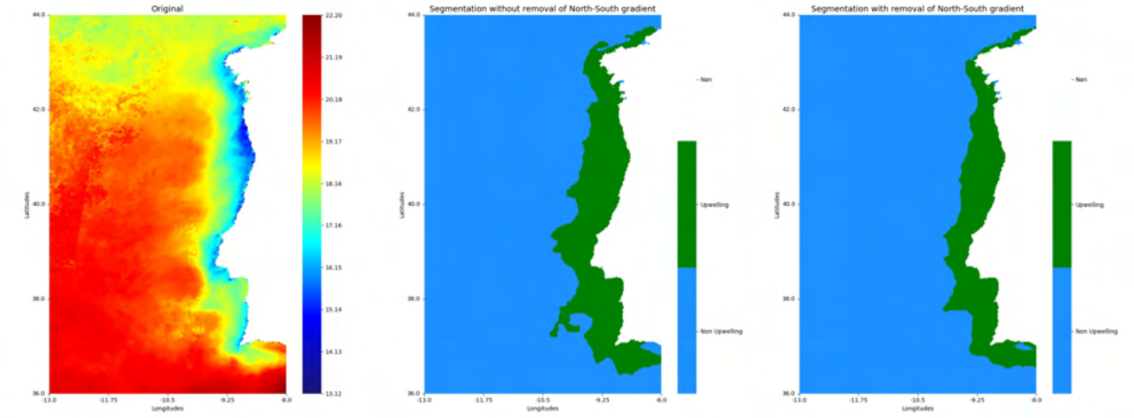


Figure 4.4: Comparison of segmentation results using a non preprocessed SST image in the middle and a SST image which suffered the transformation mentioned in Section 4.2.1 on the right

4.2.2 Moving Average Filter

The previously described preprocessing method, which helps to get more precise segmentation results obtained by the S-STSEC algorithm, was included into a preprocessing pipeline in order to prepare a data set of weekly SST images to be used by the Core-Shell clustering algorithm.

For this, a moving average filter was applied to a set of weekly SST images in order to create a new time series of SST images. This new time series is a smoother set of data which is comprised on the original SST images and which better exposes the target behavior to be studied.

This procedure also helps to remove white noise from the images, result of measure errors and atmospheric events. A complete study was made, where the temporal variation of the mean temperatures regarding averaged SST images was observed. Knowing that one year of images is composed by 46 8 days weekly SST images, a moving average of window size from 2 to 25 was studied in order to analyze which window size kept the maximal temporal resolution while smoothing out the curve of the correspondent time series. Note that for every time window studied, a step of 1 was used in the moving average procedure.

As expected, with greater window sizes the resulting averaged images had to low of a variability between them, implying that the temporal resolution was lost due to using too much information to compose a single image. With a smaller window size, the temporal resolution was kept, but the smoothing of data was not enough and therefore a lot of variability was still observed in the results as well as fragments of white noise.

As a result of these observations, a window size of 5 was set to be used in this study. After applying such moving average filter, the mean temperatures, the correspondent standard deviations and the median temperatures of such averaged images were extracted and can be seen in Figure 4.5, which present a smooth curve of temperature values throughout the year while maintaining enough resolution.

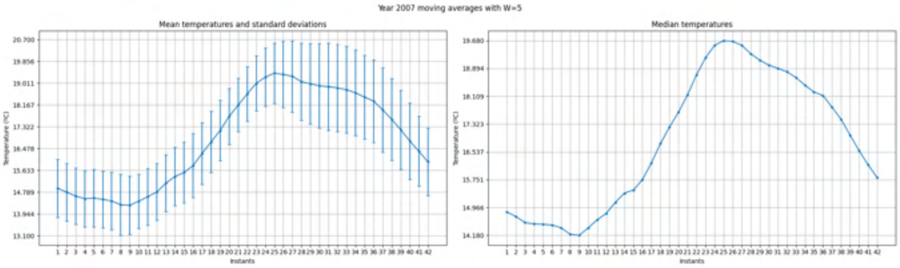


Figure 4.5: Mean, standard deviations and median upwelling temperatures' series for the year of 2017

4.2.3 Enhancement of East-West temperature gradient

To deal with the still occurring over segmentation problem like the one present in Figure 4.6, where the region growing algorithm S-STSEC over segmented the northern region of the image, the preprocessing pipeline described until here was extended adding a third phase of preprocessing.

In the work proposed by Aouni et al. [12], a SST image processing algorithm was developed to tackle the over segmentation problem present in region growing algorithms when applied to upwelling detection. The main goal of this approach was to enhance the

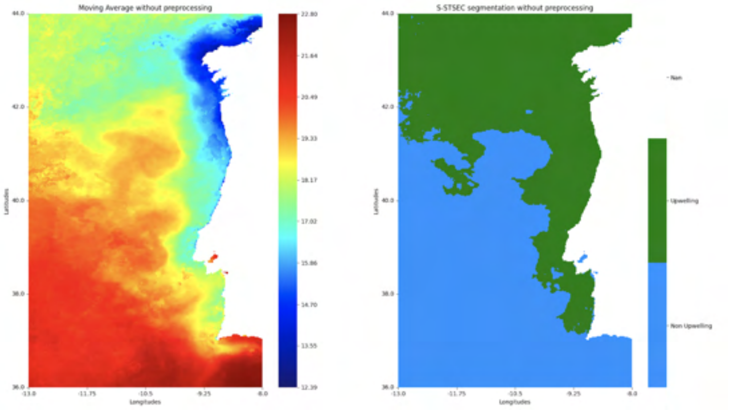


Figure 4.6: Example of an over segmentation situation when using the S-STSEC algorithm in a Portuguese coastal SST image

gradient between the upwelling regions to be extracted and the offshore waters. Thus, whenever applying the similarity condition in the border region of the target area, a greater distinction between upwelling pixels and non-upwelling ones is present which helps to ignore the latter and therefore preventing over segmented results.

Therefore, a preprocessing algorithm was developed based on the work presented by Aouni et al. with the goal of focusing the study in the Portuguese coast.

Method Summary: This method boils down to processing a given SST map using perpendicular lines to a given coastline.

The method starts by finding the coastline orientation for the image to be segmented. Since the western coastal region of Portugal has a somewhat constant longitude, the coastline was defined as being a column of the image matrix.

Having the coastline orientation, the next step is to compute a set of perpendicular equations to the coastline. In order to do so, each row of an SST image $A(L,C) = \alpha_{i,j}$ was used as a perpendicular line to the coast, with $\alpha_{i,j}$ being the temperature of the pixel (i,j). Therefore, for a SST image with L lines and C columns, L perpendicular lines to the coast are computed and L correspondent sets s are initialized.

Moving to the transformation phase, for each set of points s, the max temperature t^s_{max} is computed and, for each pixel (i,j) belonging to s, its temperature value $\alpha_{i,j}$ is updated:

$$\alpha_{i,j} = \alpha_{i,j} - t^s_{max} \tag{4.1}$$

After this update, one last step is performed where a mean filter is used in order to smooth the result from the transformation phase, which can leave some unnatural lines in the final result. The result of such preprocessing can be seen in Figure 4.7

Improving temperature gradient: Since the gradient obtained in the previous method is not strong enough and therefore did not avoid the S-STSEC algorithm to over segment the upwelled waters, the eq. (4.1) was modified. Instead of using the max value of a set

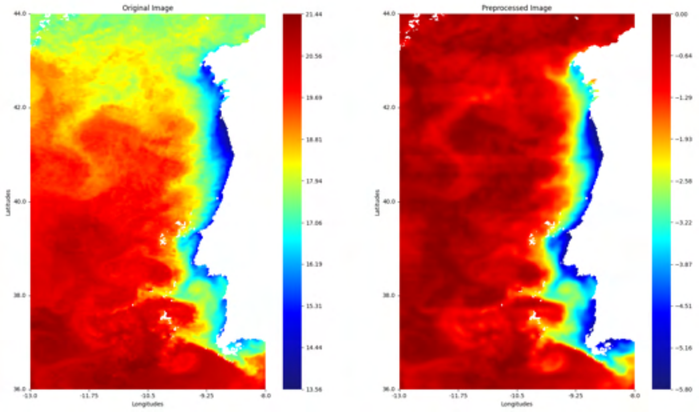


Figure 4.7: Original image on the left and corresponding preprocessed image on the right

s to update the value of a pixel $\alpha_{i,j}$, the mean value of a set s is used instead, and the update is only applied in certain situations.

All the steps until the assignment of each pixel to the correspondent set s were performed as explained before. After these, for each set of points s, the mean temperature t^s_{mean} is computed. For each pixel (i,j) belonging to a set s, if its temperature is higher than the computed mean t^s_{mean}, the value assigned to that pixel is 0. If not, its temperature value $\alpha_{i,j}$ is updated:

$$\alpha_{i,j} = \alpha_{i,j} - t^s_{mean} \tag{4.2}$$

The improved method produces a stronger gradient between the upwelling and non-upwelling regions, allowing the S-STSEC algorithm to segment a very accurate result and avoiding over segmented results like before. An example of the improved preprocessing method of a Portuguese SST image can be observed in Figure 4.8, which result can be compared with the previously obtained preprocessed image in Figure 4.7.

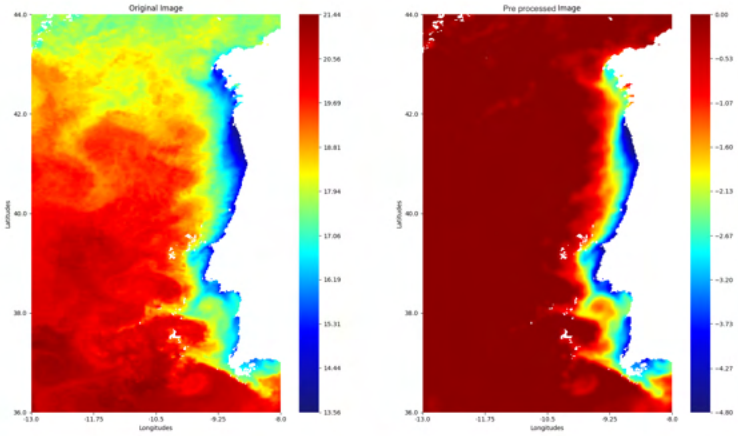


Figure 4.8: Original image on the left and corresponding preprocessed image on the right

4.2.4 Summary: Preprocessing workflow

All the aforementioned stages of the preprocessing pipeline can be seen in Figure 4.2. Given a set of N weekly SST images, the north-south gradient of each image is removed as described in Section 4.2.1. Then, every W images are averaged into a single W weeks averaged SST image like explained in Section 4.2.2. Finally, each averaged SST image is preprocessed with the temperature preprocessing method explained in Section 4.2.3.

This pipeline creates fully preprocessed SST averaged images, which are called SST instants. Such instants are then used as inputs for the S-STSEC algorithm to produce binary segmentation maps.

4.3 Feature Extraction from S-STSEC instants' segmentations

With the SST instants of a given upwelling season segmented by the S-STSEC algorithm, several features were empirically chosen and extracted, with the intent of studying the behavior of the obtained upwelling regions.

To do so, upwelling regions (UR) time series were built, these being composed by the areas and average temperatures time series of the obtained upwelling regions. In addition to these, the latitudes of the northernmost and southernmost areas of the upwelling regions were obtained, in order to study and analyze such characteristics of the upwelling regions through time.

When studying such features, it was noticed that the segmented upwelling regions did not gradually evolve through time. Instead, it was found the presence of several time periods where in each one the upwelling behavior was similar and followed an alike pattern. It was also noted that the upwelling events were of different intensity and had a different morphology when compared to other time periods. Such differences can be seen in Figure 4.9, where a clear distinction of upwelling area morphology can be seen between SST instants 1 and 2 and SST instants 13 and 14 for the year 2007.

Therefore, it was concluded that several groups of consecutive SST instants needed to be obtained. Such groups of SST instants are called upwelling spans.

The bi-partition maps obtained by the S-STSEC algorithm are also used to automatically delineate the upwelling fronts and thus to create the upwelling front (UF) time series. From this time series, the pixels that belong to the frontier of the upwelling region are obtained and their distance to the coastline is computed.

4.4 Iterative Anomalous Pattern for unsupervised Time Series segmentation

In order to create such upwelling spans in an unsupervised manner, the features aforementioned are used with an unsupervised method introduced in [54] by Mirkin called Iterated Anomalous Patterns (IAP).

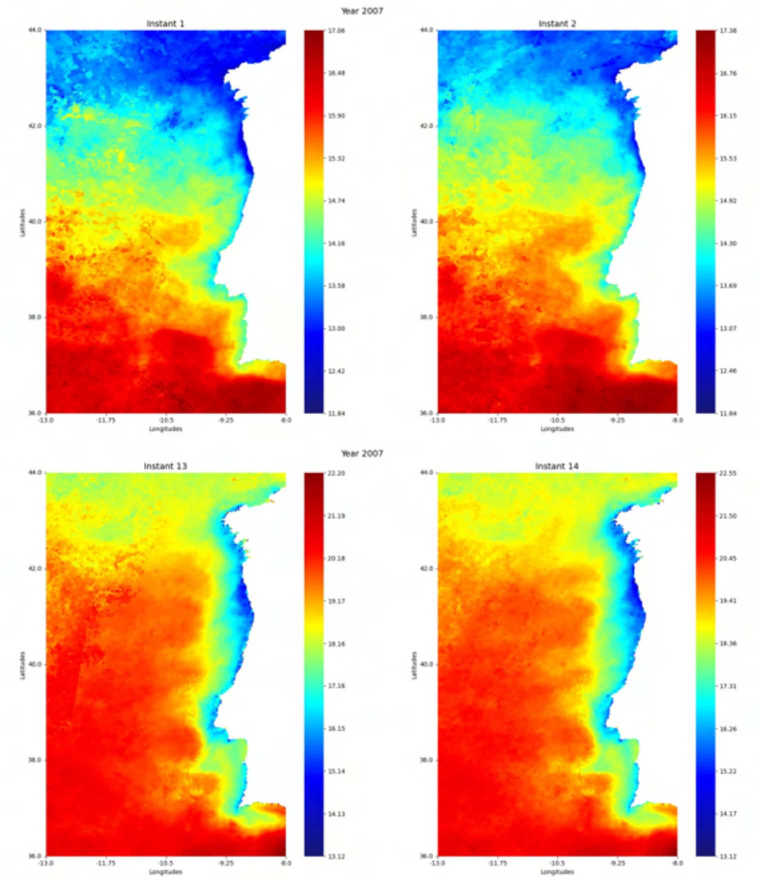


Figure 4.9: SST instants 1 and 2 on the top row and SST instants 13 and 14 on the bottom row for the year 2007, where a clear distinction in upwelling region characteristics can be noticed

Iterated Anomalous Patterns: This clustering algorithm is categorized as a divide and conquer method and can be included in the top-down approaches for the segmentation of time series. It is also classified as an unsupervised approach in this context since it does not receive as an input the number of spans to be extracted.

This algorithm uses 2 seeds, a constant reference point and a cluster seed which at first is the furthest point to the reference point. Then, a cluster is built as being the set of points that are closer to the cluster seed than to the reference point. The cluster seed is updated to the new cluster's gravity and the previous two steps are repeated until one of four stop conditions is reached:

1. All entities are clustered: There are no unclustered points left;
2. Cumulative contribution: The clusters' cumulative contribution is greater than a pre-defined threshold;
3. Relative contribution: The cluster's contribution is too small;
4. Cardinality of clusters: The number of extracted clusters reached a given threshold.

The full pseudocode for such algorithm can be seen below:

Algorithm 1: Iterated Anomalous Patterns

```
Preprocessing: Specify a reference point $\alpha$, which can be the grand mean, and standardize the original data in order to shift the space origin to the reference point $\alpha$;
Initial setting: Initialize a centroid $c$ as being the most distant unclustered point to $\alpha$;
Cluster update: Determine the cluster set $S$ as being the set of points that are closer to the centroid $c$ than to the reference point $\alpha$;
Centroid update: Calculate the within S mean $c'$ and check if it differs from the previous centroid $c$. If it does, the centroid $c'$ is set as the new centroid $c$, and return to step 3. Else, go to step 5;
Stop condition check: The stop conditions are tested. If any stop condition is verified, go to step 6. Else, save the set S and go to step 2;
Output: Outputs the saved sets S's, the centroids c's and further aiding information.
```

For the present study, the IAP algorithm will be used as a top-down approach for the task of time series segmentation. The goal of this algorithm is the creation of several upwelling spans using an unsupervised clustering algorithm, where each upwelling span is composed by several consecutive SST instants.

IAP algorithm inputs: To produce such upwelling spans, like aforementioned, the IAP algorithm will be used with the following features as input: the areas' and average temperatures' time series of the N upwelling regions obtained by segmenting N SST instants in their chronological order, and the latitudes of the northernmost and southernmost region of the detected upwelling.

IAP algorithm stopping condition: The IAP algorithm was used to segment a given upwelling season time series until no more data is left to be analyzed, this being the condition number 1 aforementioned. The use of such condition can be justified by the fact that the goal here is to include every instant of a time series into an upwelling span.

After running the IAP algorithm, several clusters with different sizes are generated, some of them with sizes smaller than a pre-defined threshold of 3 maps. This threshold was defined with the thought that an upwelling span can not be defined by less than 3 SST averaged maps.

Therefore, for each cluster obtained which size was smaller than the defined threshold, an automatic correction is made. For each image instant that belongs to a smaller cluster, it is added to the closest consecutive cluster with the lowest contribution value. After this step, each cluster is composed by three or more consecutive instants.

Finally, each one of these clusters is considered an upwelling span, which is composed by several consecutive SST instants. Such SST instants that belong to a certain upwelling span are then used by the Core-Shell clustering algorithm as inputs for the creation of a single Core-Shell cluster.

4.5 Core-Shell Clustering

In this Section a novel dynamic clustering model for the automated detection and tracking of coastal upwelling events from SST remote sensing data is introduced.

To this end, the segmentations resulting from the S-STSEC algorithm were included into a dynamic Core-Shell clustering algorithm based on the method presented in Section 3.2.1. This proposal is integrated into a broader field of research called dynamic ST clustering, which embodies the identification of subregions in space and in time that show coherent measurements, denominated dynamic ST clusters.

The main goal of this dissertation was to conceive an adaptive dynamic ST clustering algorithm based upon SST remote sensed imagery to segment pretended coastal upwelling regions. With this purpose, a new algorithm called Core-Shell clustering algorithm was developed which goal is to iteratively maximize a clustering criterion G until all the SST instants are processed and a local maximum value for the clustering criterion is found, creating a new segmentation result called a Core-Shell cluster.

As discussed in Section 3.2, dynamic ST clustering involves real challenges on data representation and data analysis methodologies. One of these is the spatial and the temporal resolutions, where the integration of domain knowledge when selecting such is recommended since there are no objective guidelines provided by the current literature [13].

Adding to this, another common challenge present is the need of ad-hoc parameter specification. The requirement of such parameters can be explained by the existence of features that are not informative enough, caused by common absence of data and/or presence of noise in it.

4.5.1 The Model

Given T SST instants $A(L,C) = \alpha_{i,j}^{t}$, with $i \in L$, $j \in C$, $t = 1,2,3,...,T$, a Core-Shell cluster is represented by the union of $T+1$ binary maps. Of these binary maps, one is the upwelling Core $R(L,C) = r_{i,j}$ and the rest are upwelling Shells $S(L,C) = s_{i,j}^{t}$.

A Core is a steady cluster composed by a set of points that have the most constant membership, with $r_{i,j} = 1$ if a pixel (i,j) belongs to the Core and $r_{i,j} = 0$ otherwise.

The union of the remaining binary maps is called an upwelling Shell, characterized by a binary matrix $S(L,C) = \cup_{t=1}^{T} s_{i,j}^{t}$. A Shell is a cluster comprised of points with uncertain memberships, which represent the dynamics of an upwelling event through time. In this binary cluster, $s_{i,j}^{t} = 1$ if a pixel (i,j) belongs to the Shell at time step t and $s_{i,j}^{t} = 0$ otherwise.

Note that a Core pixel can not belong to any upwelling Shell at any instant $t = 1, 2, 3, ..., T$. The same happens when a pixel belongs to any or several upwelling Shells at any moment t, where this can not belong to the upwelling Core:

$$r_{i,j} * s^t_{i,j} = 0, \quad t = 1, 2, ..., T \tag{4.3}$$

Given a SST instant $A(L, C) = \alpha^t_{i,j}$, the proposed model is defined by:

$$\alpha^t_{i,j} = \lambda^t s^t_{i,j} + (\lambda^t + \mu^t)\, r_{i,j} + e^t_{i,j}, \tag{4.4}$$

where λ^t is the Shell's intensity value at time step t, $\lambda^t + \mu^t$ is the Core's intensity value at time step t and $e^t_{i,j}$ are the residual values. The Core's intensity should be always smaller than the Shell's intensity for any $t \in T$ since the Core's temperatures are lower than the Shell's temperatures. The values $e^t_{i,j}$ are to be minimized according to the least squares criterion:

$$\Delta = \sum_{t=1}^{T} \sum_{i=1}^{L} \sum_{j=1}^{C} (\alpha^t_{i,j} - (\lambda^t + \mu^t)\, r_{i,j} - \lambda^t s^t_{i,j})^2 \rightarrow min \tag{4.5}$$

Using the first order optimality conditions, which states that for a smooth function to achieve a minimum its gradient must be zero:

$$\frac{\partial \Delta}{\partial \lambda_t} = 0 \ ; \ \frac{\partial \Delta}{\partial \mu_t} = 0 \tag{4.6}$$

These conditions lead to the optimal intensities of both Core and Shell to be equal to the mean upwelling temperatures, retrieved from the SST instants according to the corresponding binary matrices:

$$\lambda_t = \frac{\sum_{i=1}^{L} \sum_{j=1}^{C} \alpha^t_{i,j} s^t_{i,j}}{\sum_{i=1}^{L} \sum_{j=1}^{C} s^t_{i,j}} \ ; \ \lambda_t + \mu_t = \frac{\sum_{i=1}^{L} \sum_{j=1}^{C} \alpha^t_{i,j} r_{i,j}}{\sum_{i=1}^{L} \sum_{j=1}^{C} r_{i,j}} \tag{4.7}$$

When replacing the derived expressions 4.7 in eq. (4.5) and algebraically simplifying the expression, the Δ criterion can be written as:

$$\Delta = D - G, \tag{4.8}$$

with D and G being, respectively:

$$D = \sum_{t=1}^{T} \sum_{i=1}^{L} \sum_{j=1}^{C} (\alpha^t_{i,j})^2 \ ; \ G = \sum_{t=1}^{T} ((\lambda_t + \mu_t)^2 * |R| + (\lambda_t)^2 * |S^t|), \tag{4.9}$$

where $|R|$ is the number of pixels in the Core of interest and $|S^t|$ is the number of pixels in the Shell at time step t being analyzed.

Since the term D is constant, one can conclude that minimizing the criterion (4.8), which is the goal of this algorithm, is equivalent to maximizing the term G.

4.5.2 The Algorithm

Given that a Core is composed by the pixel points with higher and most constant memberships, the initial Core set R is instantiated as being the intersection of every S-STSEC segmentation result received as input:

$$R = \cap_{t=1}^{T} C^t, \tag{4.10}$$

with T being the number of the of SST instants under study and C^t being the t'th cluster obtained by the S-STSEC algorithm. Afterwards, the shells S^t are defined as being the set differences between each S-STSEC segmentation result C^t and the computed core set R:

$$S^t = C^t - R, \tag{4.11}$$

With R and S^t computed, for $t = 1, 2, ..., T$, a set B is computed as being the union between the core R, all the shells S^t computed, and the set of pixels forming a 4 neighborhood F_4 with the region defined by the union of the core R and the shells S^t:

$$B = R \cup_{t=1}^{T} S^t \cup F_4, \tag{4.12}$$

A set B contains all the pixels of interest to be visited and analyzed by the Core-Shell clustering algorithm, whose goal is to find a suboptimal Core-Shell cluster $R \cup_{t=1}^{T} S^t$ and the Core's and Shells' intensities. With all the initializations finished, the Core's and Shells' intensities are computed from the sets obtained in eq. (4.10) and eq. (4.11) and an initial G value is calculated.

All of these steps can be seen in Algorithm 2 (lines:1-3)

In order to maximize the increase of the criterion G, specified in eq. (4.9), $1 + 2^T$ scenarios need to be analyzed in order to find a suboptimal one. These scenarios are the following:

1. $r_{i,j} = 1$: For this single scenario, a pixel (i, j) is seen as belonging to the Core. Using the criterion specified in eq. (4.3), the pixel can not belong to any shell for $t = 1, 2, ..., T$. Therefore, $s_{i,j}^t = 0$ for $t = 1, 2, ..., T$ (lines:5-9);

2. $r_{i,j} = 0$: If a pixel (i, j) does not belong to the Core, it can belong to up to T different Shells for $t = 1, 2, ..., T$. Therefore, there are 2^T different combinations to be tested, where each states to which Shell's partitions a pixel belongs to (lines:10-17).

In summary, in order to find a suboptimal solution to the criterion in eq. (4.8), all the pixels (i, j) from set B need to be iterated and for each pixel, the Core values $r_{i,j}$ and Shells' possible combinations $s_{i,j}^t$ aforementioned need to be tested and analyzed.

The combinations that cause a bigger increment in the clustering criterion G are saved (lines:18-19), and when all the pixels (i, j) that belong to the set B are visited, for each decision saved, the assignments to the Core's and Shells' values are made (lines:20-21).

With the completion of this last step, the final Core's and Shells' intensities are computed, and the final Core-Shell cluster is computed (lines:23-24) as being:

$$SC = R' \cup_{t=1}^{T} S'^{\,t}, \tag{4.13}$$

where R' is the Core after all the assignments made and $S'^{\,t}$ are the Shells with all the assignments made.

In conclusion, for each pixel (i, j) that belongs to the set B, $2^T + 1$ different scenarios are analyzed where the pixel is added and removed from the Core and different T Shells, in order to find the combination of $r_{i,j}$ and $s_{i,j}^t$ values that maximize the clustering criterion G formulated in eq. (4.9).

The outputs of such algorithm are the intensity values of the obtained Core and Shells, as well as the computed Core-Shell cluster, in a tri-partition map form. The pseudocode for such algorithm can be seen and inspected in Algorithm 2.

4.6 Feature Extraction from Core-Shell clusters

A time series based on the tri-partition maps obtained for a whole upwelling season called Upwelling Span (US) time series is built, with the intent of evaluating the evolution of the minimum, maximum and average temperatures of the obtained Core-Shell clusters, as well as offshore difference temperatures and areas of such computed structures.

Such time series are of great importance in order to comprehend the upwelling behaviors and dynamics through time of the Core-Shell clusters obtained from the Core-Shell clustering algorithm discussed until this point for a given upwelling season. It is also important to study the dynamics captured by the obtained Shells for the different upwelling spans.

Algorithm 2: Core-Shell Clustering Algorithm

```
Inputs:
  - T SST instants $A^t = (a^t_{i,j}), i = 1, 2, ..., L; j = 1, 2, ..., C; t = 1, 2, ..., T$.
  - T bi-partition maps obtained by S-STSEC algorithm over the T consecutive $A^t$ maps.

get $R$, $S^t$ for t=1,2,...,T, and $B$;
get starting intensities $\lambda_t$ and $\lambda_t + \mu_t$;
get correspondent $G$ value;

for i, j in $B$:
  # Case where pixel belongs to the Core and not to any Shell
  $r_{i,j} \leftarrow 1$;
  $s^t_{i,j} \leftarrow 0$, for t=1,2,...,T;
  get new intensities $\lambda_t$ and $\lambda_t + \mu_t$;
  get new $G$ value;

  # Cases where pixel belongs to any number of Shells and not to the Core
  $r_{i,j} \leftarrow 0$;
  # Gets all possible combinations of 0's and 1's of T digits.
  $combinations = get\_combinations(2^T)$;
  for $combination$ in $combinations$:
    $s_{i,j} \leftarrow combination$;
    get new intensities $\lambda_t$ and $\lambda_t + \mu_t$;
    get new $G$ value;

    if new $G$ value greater than any other $G$ value:
      save correspondent decisions for $r_{i,j}$ and $s_{i,j}$;

for $decision$ in saved decisions:
  assign decisions to $R$ and $S^t$;

get intensities $\lambda_t$ and $\lambda_t + \mu_t$;
return $R \cup_{t=1}^{T} S^t$, $\lambda_t$, $\lambda_t + \mu_t$, with t = 1, 2,..., T;
```

5

Experimental Study

In this chapter we present and discuss the experimental results. The chapter is organized as follows: in Section 5.1, the goals of the current study will be presented. The imagery data is explained in Section 5.2 and the settings for the conducted study can be consulted in Section 5.3. The experimental study will be discussed starting in Section 5.4, where the segmentation results of the algorithm S-STSEC are analyzed when applied to preprocessed and non preprocessed SST images. Following this, experts' reviews of such segmentations can be seen in Section 5.5 and the upwelling regions time series obtained from such segmentations can be examined in Section 5.6. The obtained upwelling spans, posterior analysis of the obtained Core-Shell clusters and analysis of Cores' and Shells' fronts can be explored in Section 5.7. In Section 5.8, time series created using the computed Core-Shell clusters are explored. A final summary of the results obtained for such experimental study can be found in Section 5.9.

5.1 Goals of the Study

The goals of the experimental study are:

1. Development of a robust preprocessing stage able to overcome the problems of North-South and East-West temperature gradients in order to overcome the problems of explosion present in S-STSEC algorithm's segmentations, to smooth a set of time series SST images and to improve SST data quality, explored in Section 4.2 and referred as the preprocessing pipeline in Figure 4.1;

2. Assessment and analysis of not only the quality of S-STSEC segmentations applied to complete upwelling seasons of the Portuguese coast, but as well as the upwelling features extracted from those segmentations, referred as bi-partition maps and upwelling regions time series in Figure 4.1 and mentioned at the start of Section 4.5. Posterior reflection and opinion of oceanography experts about obtained results;

3. Application of the IAP algorithm discussed in Section 4.4 to several groups of SST images coupled with S-STSEC segmentations to retrieve unsupervised upwelling

spans through time series segmentation, method presented in the full experimental pipeline 4.1 as "Group SST instants into P spans";

4. Exploration of validity measures to assess the quality of the unsupervised retrieved upwelling spans;

5. Application of the Core-Shell clustering algorithm to the retrieved upwelling spans using the SST instants and their S-STSEC segmentations to extract one Core-Shell cluster per upwelling span, alongside with Core's and Shells' information of each Core-Shell cluster as explained in Section 4.5.1 and shown in Figure 4.1 as "Core-Shell Clustering algorithm";

6. Extraction of several upwelling features from the upwelling spans and creation of features' time series called Upwelling Span time series in the full experimental pipeline 4.1, features such as areas of the Cores, Shells and the whole Core-Shell clusters, mean, minimum and maximum temperatures of the obtained regions, and posterior analysis of the upwelling trends along different upwelling seasons.

5.2 Imagery Data

Each image used in this study is a SST map with a resolution of 401 pixels by 251 pixels. Each pixel of such matrix represents a float value of temperature in degrees Celsius and has a spatial resolution of 2km by 2km. These images are from the Portuguese coast, ranging from a latitude of 36ºN to a latitude of 44ºN, and a longitude from 8ºW to 13ºW. Each one of the SST images is a composition of 8 daily SST images, which were processed and combined to form a single weekly SST image.

For the whole collection of images, whenever a measure for a pixel (i, j) is collected with some error associated, such as meteorological events related errors or collected data from a non-sea surface, a value of NaN is instead saved for that pixel (i, j).

The entire available data set is comprised by 16 different SST image collections from 16 different years, since 2004 until 2019, the most recent one. For each year, 46 8 days weekly SST images are available.

It is known that off western Iberia an upwelling season develops between March and October [97, 30]. Taking into account the temporal resolution of each image and the aforementioned knowledge, only the images from approximately the 30^{th} of March until the 30^{th} of October were taken into consideration for the study here conducted.

One important detail regarding the experimental study is how SST instants relate to the dates of the SST images. Given that an instant is a SST averaged image composed by 5 SST weekly images, and the upwelling seasons under study begin at the 30^{th} of March, the first instant corresponds to the SST original images from the 30^{th} of March until the 8^{th} of May, the last day. After the first instant, each instant starts 8 days after the previous instant's start, and ends 8 days after the end of the previous instant. Formulating this, each instant i starts x days after the 30^{th} of March, with x being:

$$x = (i - 1) * 8 \tag{5.1}$$

Figure 5.1 shows two SST original images as an example within the time window determined in the paragraph above. The image on the left is referent to an 8 days week starting on the 30^{th} of March 2009, where one can not see a well-defined upwelling system. This is very common since the upwelling regions tend to increase in strength starting at around this time. The image on the right is referent to an 8 days week staring on the 24^{th} of October 2017, where one can see a drastic difference in the upwelling strength when compared to the left image.

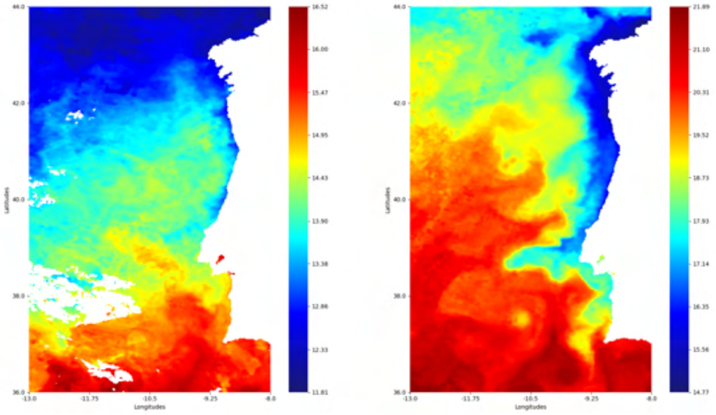


Figure 5.1: Original SST weekly image of the Portuguese upwelling system from the 30^{th} of March 2009 on the left image and an original SST weekly image of the Portuguese upwelling system from the 24^{th} of October 2017 on the right image

Other kinds of SST images and bi-partition maps used and analyzed in this experimental study are the SST instants, which are created from the procedure described in Section 4.2, and the binary maps segmentations' results when applying the S-STSEC algorithm to the fully preprocessed data. An example of such images can be found in Figure 5.2.

In order to analyze uncorrelated years of SST imagery regarding upwelling regions, there was the caution to select non-consecutive years in order to work and study different kinds of data and coastal upwelling trends. Therefore, a very important aspect of the study conducted was related to the preliminary analysis of the entire collection of 736 SST images, divided between 16 collections of SST images.

5.3 Setting of the Experiments

Another variable that needed to be studied was the quality of available data. Due to cloud cover and other meteorological events, a lot of weekly SST images present more than 50% of NaN values, thus having very low quantity of information available.

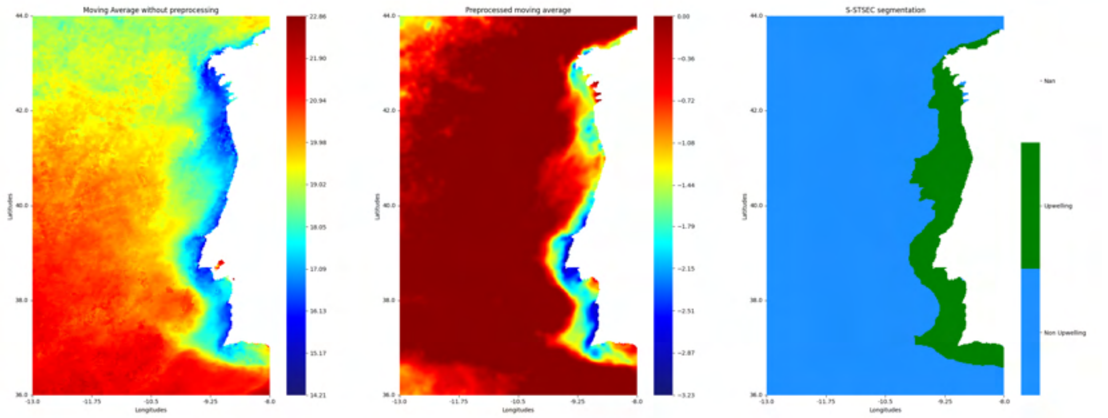


Figure 5.2: Instant 15 of the year 2019, with a SST averaged image of the Portuguese upwelling system on the left, its preprocessing result on the middle image and its S-STSEC segmentation bi-partition map on the right image

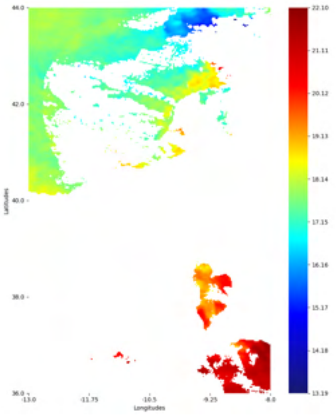


Figure 5.3: Weekly SST image from 2013, where almost no information regarding temperature values near the coastline is available.

As an example, in Figure 5.3 there is almost no information regarding temperature values near the coastline, the target area to be studied. In order to avoid that problem, the collections of SST images correspondent to the years 2007, 2015 and 2019 were selected as being the years with the highest quality of data. These sets of images were also selected due to none of these being consecutive years, which introduces the uncorrelated upwelling episodes between such sets of data.

Like mentioned before, only the images from the 30th of March until the 30th of October were taken into consideration for the study here conducted, totaling 27 weekly SST images. After applying the full preprocessing pipeline, 23 SST instants per year were obtained, giving a total of 69 SST instants for the exploratory study conducted on 3 upwelling seasons for the years 2007, 2015 and 2019.

The following experimental setting will follow the order of the workflow described in Section 4.

Smoothing of North-South temperature gradient: In order to smooth the north-south gradient from the images, the grdtrend() function from the GMT library [91] was used with a parameter $N = 3$ for the use of only linear parameters for the polynomial fit.

Moving average filter: For the moving average filter included into the preprocessing pipeline, a window size of 5 was used alongside with a moving step of 1. Therefore, for every 5 consecutive SST images, 1 averaged SST image was produced.

Enhancement of East-West temperature gradient: Regarding the enhancement of East-West temperature gradient method, all the images were preprocessed using the improved method explained in Section 4.2.3. For the smoothing of the final results, a mean filter with a window size of 5 and a step of 1 both vertically and horizontally were used for the smoothing of the preprocessing result.

Upwelling Spans definition: The upwelling spans were studied and discovered using the unsupervised method IAP, and a threshold value of 3 was set to the minimum size of an upwelling span. Mean upwelling temperatures, areas of the obtained upwelled waters and latitudes of the northernmost and southernmost pixels of the S-STSEC segmentation results were the features used and explored for the computation of such upwelling spans.

Time Series analysis: When analyzing the results of the S-STSEC segmentations for all the SST instants for a give upwelling season and the results of the Core-Shell clustering algorithm, which both are binary matrices, the average temperatures and areas of such structures are extracted, and time series are created with such features in order to study the obtained behavior for the upwelling season.

5.4 SST Image Preprocessing

One of the first steps of the work here produced was to find a proper preprocessing pipeline for the SST images, where the goal was to overcome the over segmentation problem and prepare the imagery data for the Core-Shell clustering algorithm.

In order to evaluate the impact of such preprocessing method on the S-STSEC algorithm results, this algorithm was applied to the SST averaged images with the original scale of values and to the SST instants, and the results of both sets of image's segmentations were compared.

After a thorough examination, 69 out of 69 segmentations obtained with the S-STSEC algorithm using SST non preprocessed images were over segmented on the northern region, while at the same time showed impressive quality and precision on the segmentations on the southern upwelling region, which can be seen on the middle column images of Figure 5.4. Such SST images were selected due to the representation of upwelling scenarios with different intensities.

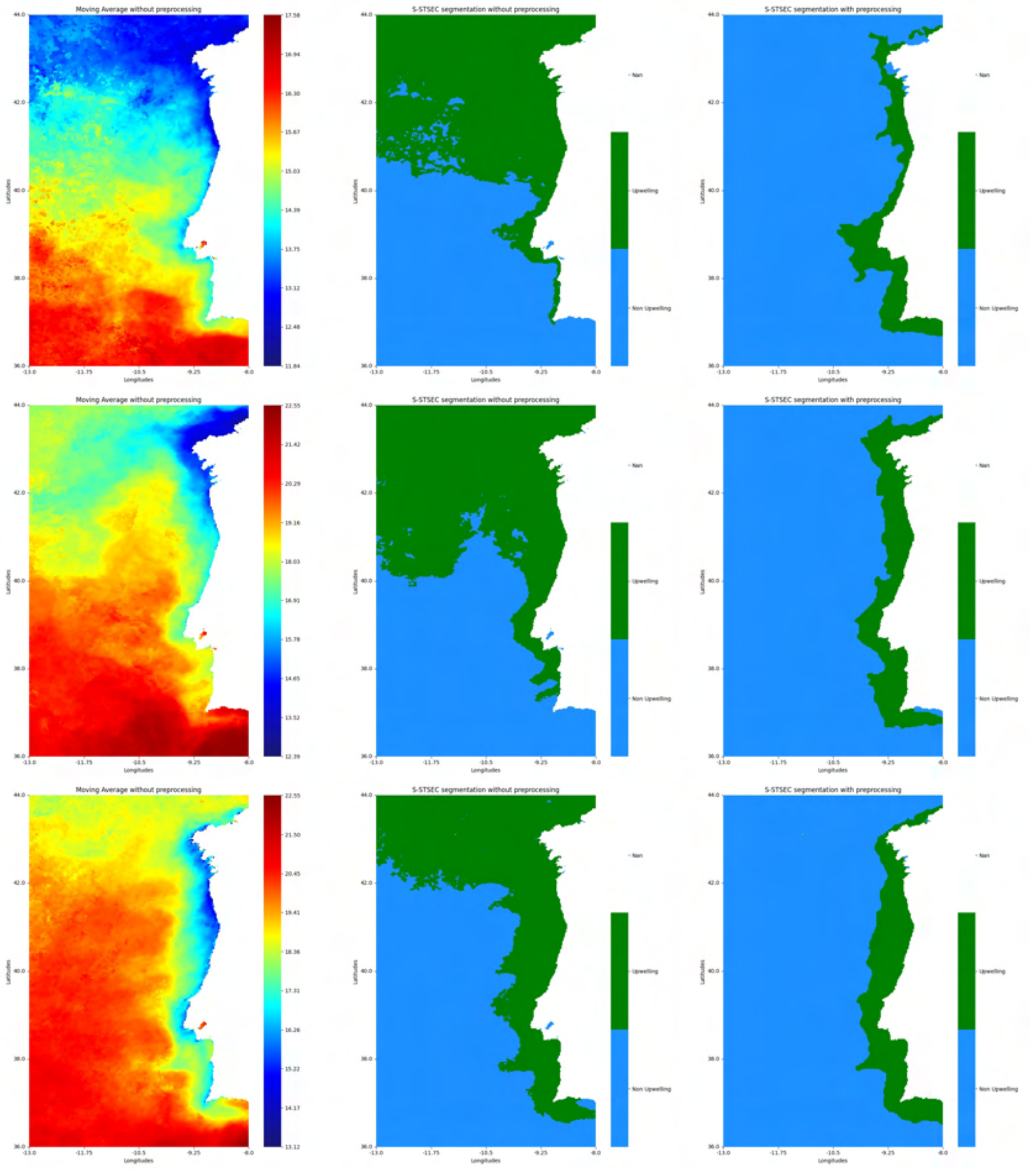


Figure 5.4: Comparisons between segmentations of non preprocessed and preprocessed SST images

After applying the preprocessing method and segmenting the new SST instants, 69 out of 69 images did not present over segmented regions like the ones obtained before, results that can be perceived on the right column of Figure 5.4.

Despite that, it was noticed that in some particular cases a loss of quality in the upwelling segmentation results occurred. This can be seen when comparing the last row's images, where the segmentation obtained for the southern region of the Portuguese coast has more detail on the frontier of the non preprocessed image when compared with the segmentation of the SST instant.

It was also noted that some over segmented results occurred, while less frequent and of lower dimensions when compared with the prior segmentation results. Such behavior can be observed on the northern region of the first row's image of Figure 5.4, where there is no upwelling presence but still was recognized a small area of upwelled waters.

With that being said, it is safe to conclude that the preprocessing method discussed is of great quality and helps to overcome the major problem of over segmentation present in methods like the S-STSEC algorithm. It is of great interest to improve this preprocessing technique, in order to improve the quality of the obtained segmentations and to avoid the low amount of over segmentations obtained.

5.5 Automatic Recognition of Upwelling regions

The 69 SST segmentation results obtained by the S-STSEC algorithm were provided to and evaluated by two oceanographers with the goal of assessing the quality of the obtained upwelling regions and how well the segmented regions represented the original real upwelling regions.

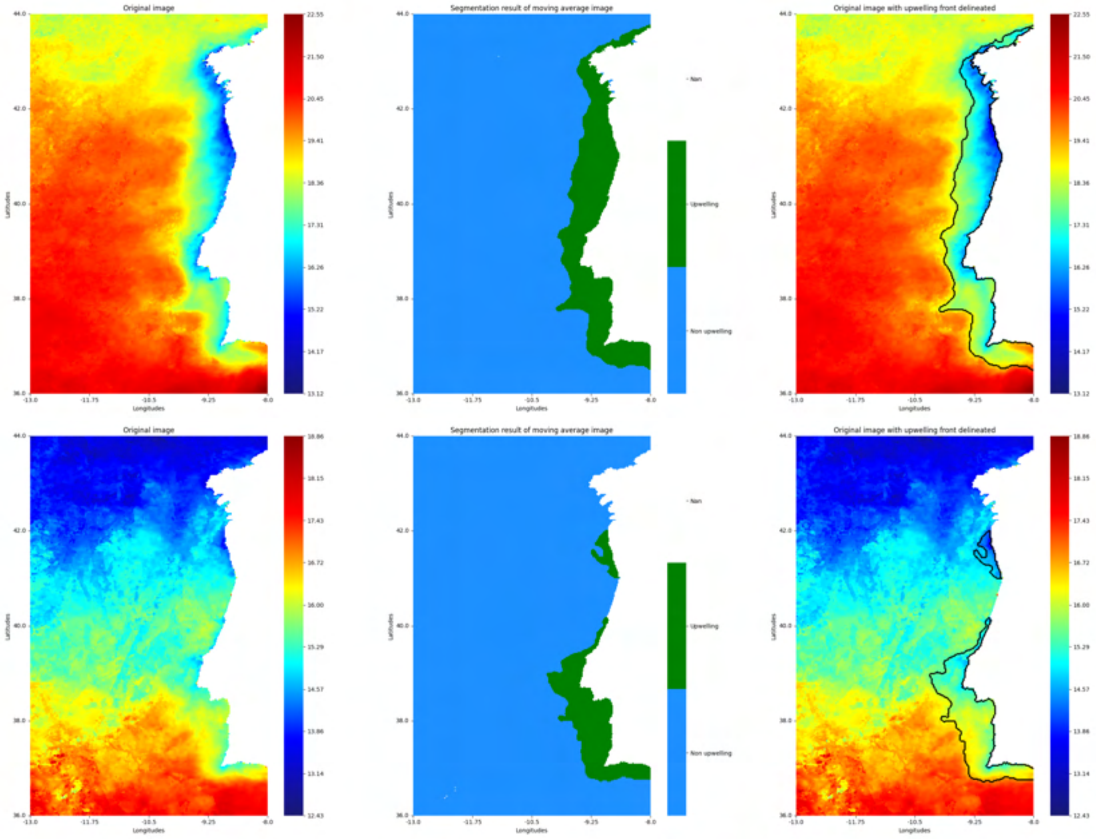


Figure 5.5: Segmented results of two SST instants for the years of 2007 and 2019, respectively

Two examples of segmentation results can be seen in Figure 5.5, where the high quality of such can be perceived after applying the preprocessing pipeline and different numbers of upwelling regions segmented can be counted.

In order to evaluate the segmented upwelling regions, a scale of five grades was used composed by: 1 (bad), 2 (poor), 3 (fair), 4 (good) and 5 (excellent). The grades 1 (bad) and 2 (poor) were provided to images where the upwelling regions obtained did not correspond or did not correspond well to the real upwelling regions. The grade 3 (fair) was provided to results where the upwelling regions obtained had some over or under segmentation problems or small regions wrongly recognized as upwelling. The grade 4 (good) was given to images where a good representation of the real upwelling region was obtained by the S-STSEC algorithm and a grade of 5 (excellent) was provided to a segmentation result which represented perfectly the existing upwelling region.

Thus, in order to evaluate the quality of the segmentations' results created in this study of the SST instants for the years 2007, 2015 and 2019, available in annex I.1, these were provided to the team of oceanographers and the obtained evaluations can be analyzed in Table A.1.

Before analyzing the grades provided by the specialists' team, note that the study being made here is an objective approach for a subjective problem, thus the grades sometimes differed from different experts for a given binary segmentation result.

Looking at the results for the year of 2007, a mean classification of 3.93 was obtained, with the lowest value of 2.5 being given to the instants 8 and 10. Such results can be explained by the over segmented area present in the southern region of the Portuguese coast and in the instant 8, an upwelling region was detected north of Lisbon where in fact there is no existing upwelling, which can be seen in Figure I.8.

The year 2015 was the year when the worst results were obtained, but still with a mean grade of 3.37, where the lowest ratings were given to the first 6 instants. This is evident when looking to the segmentations' results, like the one in Figures I.24 and I.27, where in such instants the upwelling behavior was of low intensity, with low gradients throughout the whole map and therefore causing the S-STSEC algorithm to detect big upwelling regions where in fact there are none or of small intensity and size.

Concluding this analysis with the best results, the year of 2019 obtained a mean grading of 4.13. The worst result for this year was given to the instant 5 which obtained a grade of 2.0. Looking at Figure I.51, while the obtained southern segmentation region for such instant was of great precision, other two upwelling regions were obtained north of the first one which were not present in the original image. Other than that, the majority of the results for this year were of high quality, with the majority of grades higher than 4.0.

In summary, the majority of the results were graded as precise and were considered very compelling to the experts. Whenever the upwelling event was of moderate to high intensity, the upwelling segmentation results were overall classified as excellent, with grades above or equal to 4.0. The noticed downsides of the segmentation algorithm S-STSEC together with the full preprocessing pipeline was the slight over segmentation of some existing upwelling regions and the recognition of non-existing small upwelling

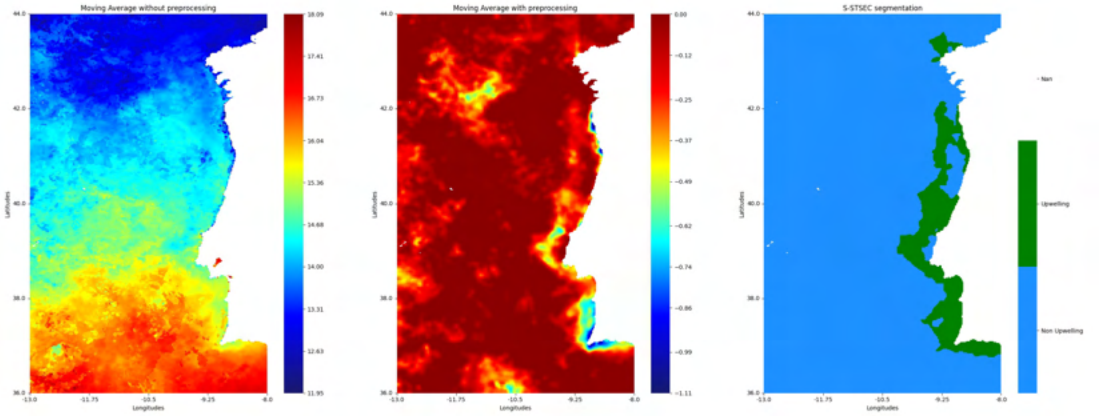


Figure 5.6: First instant of year 2015, where an over segmented upwelling area was obtained in a SST image with a small upwelling presence throughout the entire coastal region

areas whenever the upwelling presence was weak throughout an entire SST map, as seen in Figure 5.6. Another downside of the temperature preprocessing method introduced in Section 4.2.3 is the handling of the northern and southern coast of Portugal, where the average orientation of the coastline is no longer a vertical line. This issue can compromise the results in such areas and will be handled in future works.

5.6 Analysis of Upwelling Regions Time Series

After the segmentation of the SST instants using the S-STSEC algorithm, a study was made regarding the areas and the temperatures of the segmented upwelling regions through time. This analysis is fundamental and of high importance since there is the interest of extracting features to characterize the upwelling trends.

5.6.1 Upwelling vs offshore temperatures

The first study to be conducted was the comparison of the temperatures of the segmented upwelling regions with the temperatures of the non-upwelling regions. Attending to the knowledge of coastal upwelling regions, these are characterized as being regions with generally colder waters when compared to non-upwelling or offshore waters.

When looking at the year 2007 on the top image of Figure 5.7, the mean temperatures of the upwelling regions obtained by the S-STSEC algorithm are always colder when compared to the mean temperature of non-upwelling waters. The upwelled waters' temperatures also tend to be constant from the instant 13, which represents the moving average of the weekly SST images starting in August.

When analyzing the mean upwelled waters temperatures' time series of the years 2015 and 2019 in the middle and bottom graphs respectively of Figure 5.7, these present a similar pattern to the one in the year of 2007. However, both of these time series exhibit an unusual behavior in the first 3 instants, where the mean segmented upwelled waters' temperatures are higher than the mean non upwelled waters' temperatures (mean temperature of non upwelled waters).

When analyzing the SST averaged images without any preprocessing, these being the averaged SST images with the original range of temperature values, one can conclude that the northern non-upwelling regions have a high amount of low temperature pixels, which is a result of the moving average filter used. Thus, when performing the mean temperature of the non-upwelling regions, these tend to have a lower mean temperature than the upwelling regions. It is important to note that such differences are rather small when compared to how warmer the non-upwelling regions are when compared to the segmented upwelling regions from the instant 4 onward.

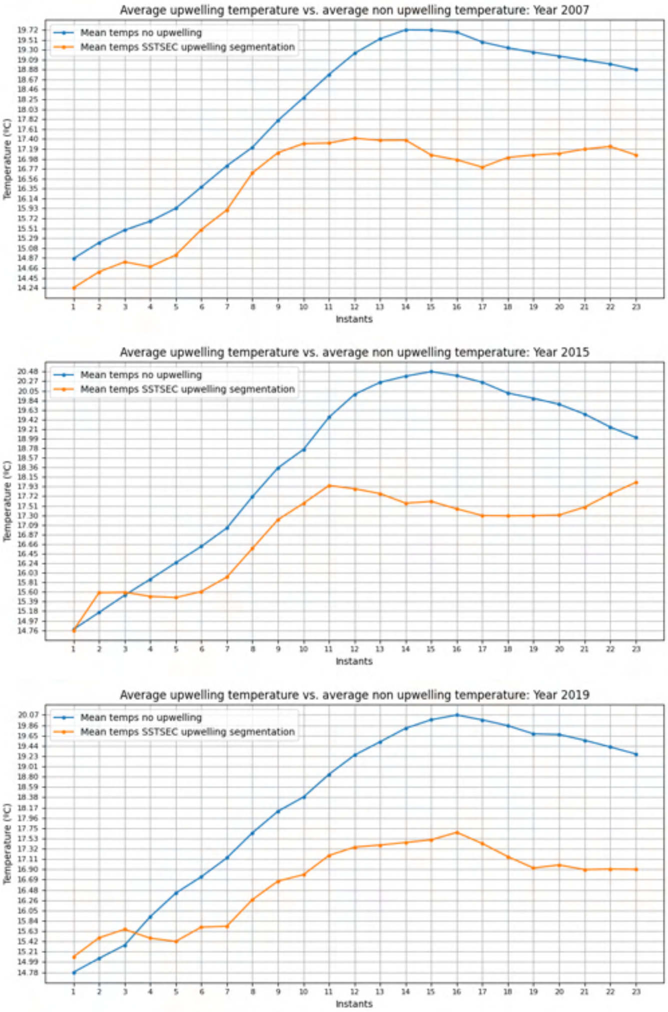


Figure 5.7: Upwelling mean temperatures vs. non-upwelling mean temperatures from 2007, 2015 and 2019 respectively

5.6.2 Upwelling area analysis

After the temperatures' time series analysis, a study was made regarding the areas' time series. This study was made while looking at the entirety of the SST maps of Portugal, but also to the regions north of and south of Cabo da Roca, since this is geographic region of major importance due to upwelling having different behavior north of and south of such area.

When looking at the areas' time series for the year 2007 in Figure 5.8, a detail of major importance was detected. When comparing the areas' information throughout the series between the whole SST map and the region north of Cabo da Roca, it was concluded that these follow a similar pattern. While evaluating the areas of the full SST images, Figure on the top, one can conclude that from instants 1 to 7, the upwelling areas follow a rather constant pattern. After this, starting on instant 8 where a minimum of upwelling area is reached, the upwelling area almost doubles its value until the instant 19, where it reaches to the maximum upwelling area value for this year. The instant 19 refers to the time window from approximately the 21^{st} of August 2007 until the 29^{th} of September of the same year. Then, the upwelling area tends to decrease until the end of the upwelling season being studied.

Figure 5.8: Areas for the segmented upwelling regions from the SST instants of the year 2007

It is also important to note that while the upwelling on the region south of Cabo da Roca tend to increase 38.7% from its minimum value to its maximum value, the upwelling area north of Cabo da Roca increases 242.3% from its minimum value to its maximum value, this meaning that a lot of dynamics happened on the region north of Cabo da Roca in 2007 when compared to the region south of Cabo da Roca.

Moving to the year of 2015, in Figure 5.9, a similar pattern is present between the whole SST map's upwelling areas and the upwelling areas on the region north of Cabo da Roca, like what was stated in the year of 2007. The upwelling behavior throughout the upwelling season studied is very different when compared with the year of 2007.

Figure 5.9: Areas for the segmented upwelling regions from the SST instants of the year 2015

When looking at the whole SST images' areas, the upwelling area tends do increase right from the beginning, ignoring the drop in area from the first instance to the second, hitting its peak at instant 10, a time window starting approximately on the 10^{th} of June 2015 and ending on the 19^{th} of July 2015, which is around 2 months earlier than the maximum area value from 2007. After this maximum, the upwelling areas decrease in small steps throughout the rest of the upwelling season.

When analyzing the whole SST map, the upwelling area grew 95.8% from its lowest to the highest area values, whether in the regions north of and south of Cabo da Roca, the increase from the minimum area value to the maximum area value was of 236.3% and 77.4% respectively. Once again, the dynamics on the north of Cabo da Roca occurred with more intensity than on the region south of Cabo da Roca, meeting the 2007 results discussed previously.

Finally, when analyzing the upwelling areas' time series of the year 2019, in Figure 5.10, some patterns observed in the previous years persist. The variation in areas of the full Portuguese upwelling area follows a similar behavior when compared to the upwelling areas north of Cabo da Roca. This year presents a behavior closer to the one studied in the year of 2007, where there is an increase of the upwelling areas over time and where the greater upwelling region is detected at instant 20, correspondent to the time frame from the 29^{th} of August 2019 until the 7^{th} of October 2019.

For this year, the maximum variation detected for the full upwelling regions throughout the entirety of the time window being studied was of 104.7%, the maximum variation throughout the year for the region north of Cabo da Roca was of 411.2% and the same metric for the region south of Cabo da Roca was of 30.9%, noticing again the most variability of upwelling area and thus greater dynamics detected on the region north of Cabo da Roca.

Figure 5.10: Areas for the segmented upwelling regions from the SST instants of the year 2019

In summary, it was noticed that in the 3 years here studied and analyzed, the northern region of Cabo da Roca always presented a more dynamic upwelling activity, where it grew more from its lowest value when compared to the southern region of Cabo da Roca.

The year 2015 had the most abrupt growth in upwelling area, taking only 3 instants to go from its lowest area value to the highest area value, whereas the years of 2007 and 2019 had a more conservative growth of segmented upwelling areas throughout the time. It was also noted that the behavior on the region south of Cabo da Roca never followed a recurrent pattern during the 3 years studied neither had similarities when compared with the region north of Cabo da Roca.

5.7 Modelling Upwelling Spans

5.7.1 Unsupervised Time Series Segmentation

In order to evaluate the obtained Core-Shell clusters in an unsupervised manner by the IAP algorithm and correspondent time series regarding mean temperatures and areas of such clusters, a small analysis to the computed upwelling spans, which can be observed in Table 5.1, will be provided.

Upwelling Spans	Year 2007	Year 2015	Year 2019
1	1 to 7	1 to 3	1 to 3
2	8 to 12	4 to 8	4 to 7
3	13 to 20	9 to 14	8 to 13
4	21 to 23	15 to 23	14 to 23

Table 5.1: Upwelling spans' composition for each year

While the upwelling seasons for the 3 years were separated into four different upwelling spans, the years of 2015 and 2019 had very similar upwelling spans. It is important to note that the smallest obtained upwelling spans are of length 3, correspondent to the last span of the year 2007 and to the first span of the years 2015 and 2019, and the largest upwelling span obtained was the last one of the year 2019 with a length of 10 instants.

For the entirety of the study here produced, the upwelling spans presented in Table 5.1 are the upwelling spans that are to be analyzed and referred to.

5.7.2 Core-Shell Clusters

In order to demonstrate how the construction of such Core-Shell clusters works, the second upwelling span of the year 2007 will be used as an example. The SST instants of such upwelling span and correspondent S-STSEC segmentations can be seen in Figure 5.11 and 5.12, respectively.

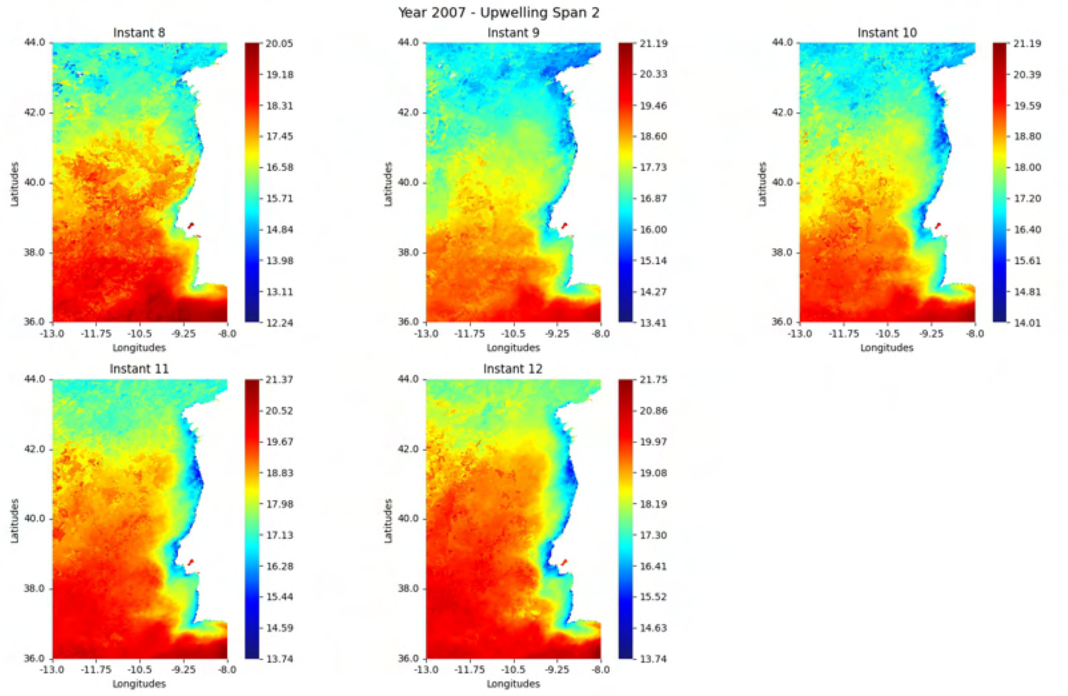


Figure 5.11: SST instants that belong to the second upwelling span for the year 2007

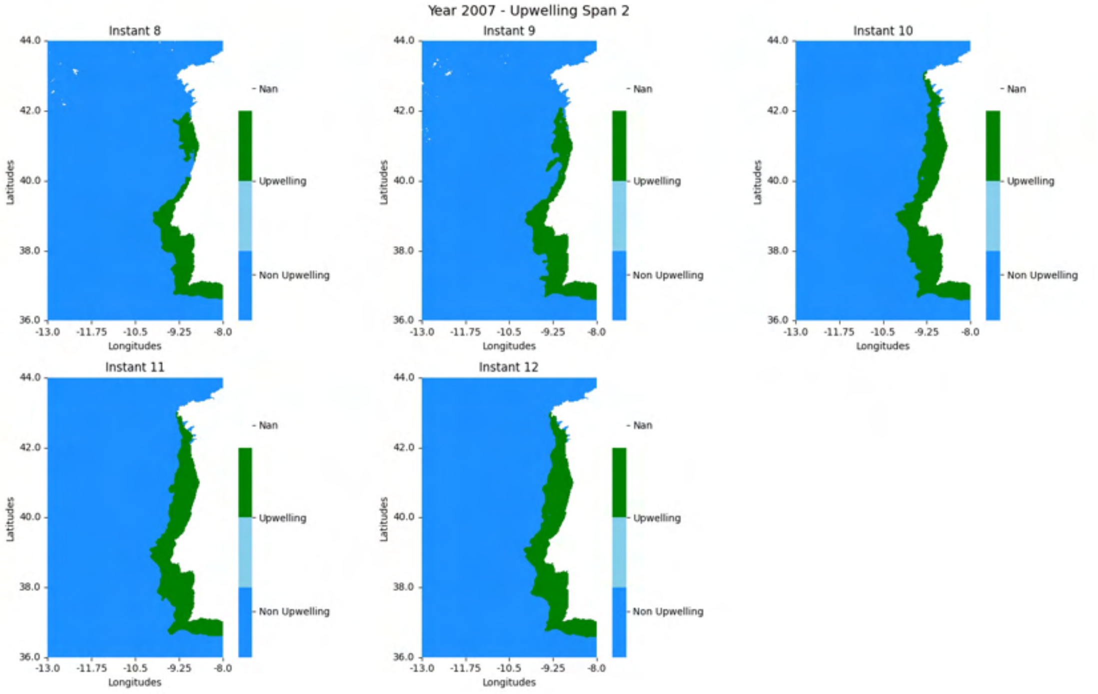


Figure 5.12: Bi-partition maps resulted from the segmentation of the second upwelling span' SST instants for the year 2007

After segmenting such SST images, the obtained segmentation maps are used as the input of the Core-Shell clustering algorithm to build the Core-Shell cluster correspondent to such upwelling span.

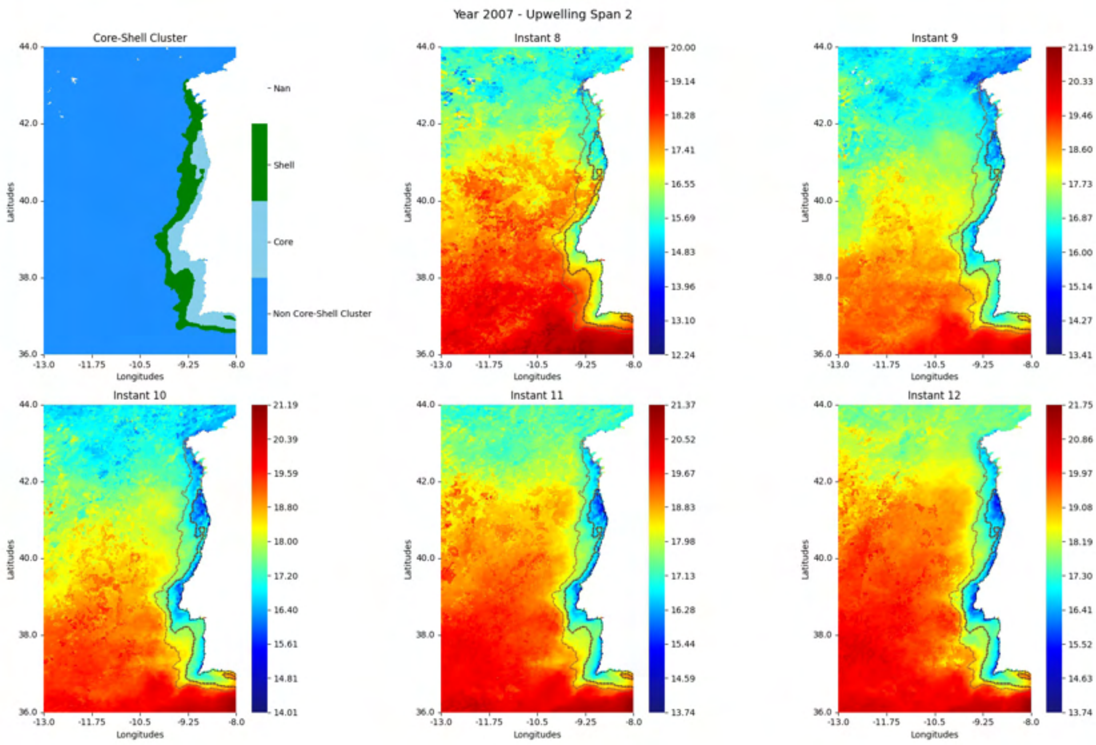


Figure 5.13: Core-Shell cluster for the second upwelling span of the year 2007, with its fronts plotted on top of the instants that belong to that upwelling span

Moving onto the analysis of the obtained Core-Shell clusters, the output of the Core-Shell clustering algorithm for the second upwelling span of the year 2007 is shown in Figure 5.13 on the top left image, as well as its frontiers on top of the instants that are a part of that same upwelling span. Note that the Core and Shell are represented in two different colors in the Figure, but the final Core-Shell cluster is defined as the union of these two structures. The plots of the Core-Shell clusters' fronts for the other three upwelling spans of the year 2007 and for the years 2015 and 2017 can be inspected in Annex I.2.

As pretended, the Core represents the pixels with the most constant membership, these being regions closer to the coastline and represented in light blue, and the obtained Shell represents the dynamics of every SST instant for the correspondent upwelling span and is represented in green. The same results were obtained for the other three upwelling spans of the year 2007, which can be seen in Figure A.9, as well as for the year 2015 and 2019, which can be seen in Figures A.10 and A.11, respectively.

According to the Section 4.5.1, the Core-Shell clustering algorithm has the goal of maximizing the clustering criterion G presented in eq. (4.9) when building a Core-Shell cluster. Therefore, looking at Figure 5.14, one can observe the evolution of the G criterion when building the second Core-Shell cluster for the year of 2007.

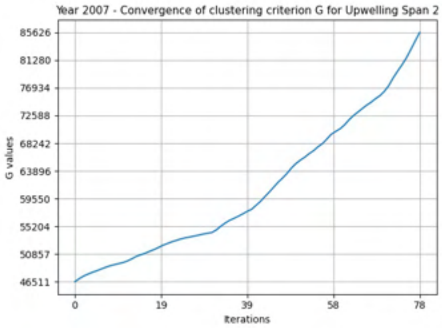


Figure 5.14: Evolution of the clustering criterion G when building the second Core-Shell cluster for the year 2007

As expected, the clustering criterion G converges to a local maximum when building the Core-Shell cluster for the second upwelling span of the year 2007, as well as when building the remaining Core-Shell clusters referent to the first, third and fourth upwelling spans of that same year, which plots can be observed in Figure A.12.

The evolution of the clustering criterion G for the years of 2015 and 2019 can be seen in Appendix, in Figures A.13 and A.14, respectively, where the same behavior as the one mentioned for the year 2007 occurred. All the clustering criterion G here analyzed converged to a local maximum and followed an alike pattern, therefore concluding that the goal of minimizing the residual values of the proposed model was reached and suboptimal Core-Shell clusters were obtained.

5.7.3 Analysis of weekly SST images with upwelling span front delineation

In order to further inspect the results of the Core-Shell clustering algorithm, a fine analysis was made regarding the Cores and Shells obtained for every upwelling season studied. To do so, for each instant i, with $i = 1, 2, \ldots, 23$, the correspondent Core and Shell were obtained automatically and plotted with the intention of helping the specialists' analysis of such results.

Since each instant i is a preprocessed SST averaged image, and each one of these instants is composed by 5 original weekly SST images, the Core's and Shell's frontiers for the instant i were plotted in each of these 5 images. An example of such composition of images can be seen in Figure 5.15. In I.3, the same compositions of images can be seen for the rest of the year 2007 and for the years of 2015 and 2019.

All the evaluations made by the experts to all the Cores' and Shell's frontiers plotted on top of the images that compose the correspondent SST instant can be observed and analyzed in Table A.2. Note that the Core plotted for a given instant is the Core for the upwelling span that the instant belongs to, and the Shell plotted is the Shell S^t, where t is the t^th instant of the correspondent upwelling span.

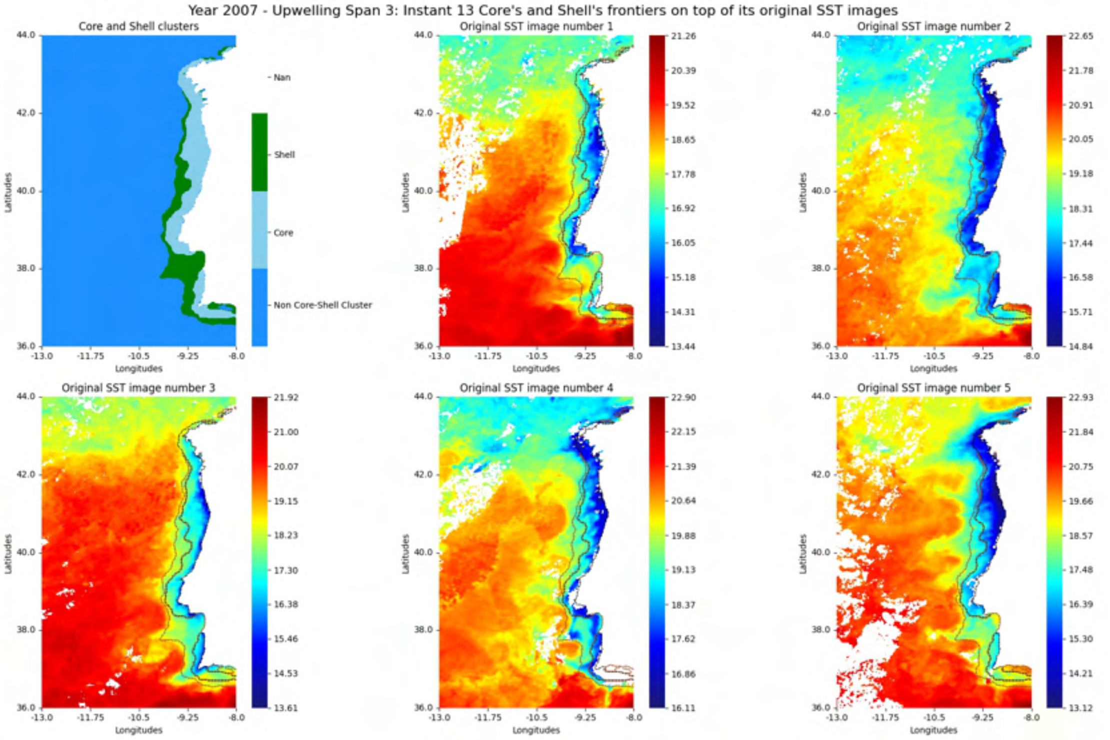


Figure 5.15: Computed Core and Shell of instant 13 of the year 2007, with fronts plotted on top of the original weekly images that compose the SST instant 13

Like the results analyzed in section 5.5, the year when the grades were the lowest was the year 2015. Given that the Core-Shell clusters are built using the segmentation results of the correspondent SST instants and since these segmentations for the year 2015 were the ones with the lowest grades and therefore the least accurate, such inaccuracy was propagated to the final results of the Core-Shell clusters. Nevertheless, this year received an average result of 4.26, meaning that even with some inaccurate binary segmentation results the obtained Cores and Shells represented the upwelling region really well.

The other two years, these being the years 2007 and 2019, received both an average score of 4.39 and 4.52 respectively. It was noted by the specialists that the Cores' and Shells' fronts did usually fit better the second, third and fourth SST original images of a given SST instant, while not fitting so well the first and fifth SST original image. It was also noted that the fronts usually fitted the best the center and southern region of the Portuguese coastline, usually failing on the northern region. This can be explained, again, by the over segmented results obtained by the S-STSEC algorithm in the beginning of the upwelling season on the northern region of the Portuguese coastline when segmenting the SST instants that then are input to the Core-Shell clustering algorithm.

It is important to note that these comparisons made between a front of the Core and the Shell and 5 SST original images is a tough comparison to do, due to the fact that some fronts fit well in the northern region of some images, while in other images the fronts fit better the central or southern region of the SST images.

Assessment of Core-Shell segmentation results

When analyzing the quality and precision of the computed Core-Shell clusters by comparing these structures to the SST instants that belong to their correspondent spans, two similarity measures were used. Note that the inputs for both similarity measures are S-STSEC segmentations, i.e. binary maps, which are used as ground truths.

Adjusted Rand Index: The first similarity measure to be studied was the ARI [72], which is a very important and common used index to compare clustering results against some external criteria. It can also be defined as a measure of agreement between a given partition and the target data.

In general, this index has values between 0 and 1, a value of 0 whenever two sets being compared have no agreement between any pair of points, and a value of 1 when two data clusters are equal to each other.

Given a ground truth $P^* = \{C_1^*, ..., C_K^*\}$ and clustering results $P = \{C_1, ..., C_K\}$ generated by a clustering algorithm, the ARI between these is defined, using the Permutation model:

$$ARI(P^*, P) = \frac{\sum_{i,j}\binom{N_{ij}}{2} - \left[\sum_i \binom{N_i}{2} \sum_j \binom{N_j}{2}\right] / \binom{N}{2}}{\frac{1}{2}\left[\sum_i \binom{N_i}{2} + \sum_j \binom{N_j}{2}\right] - \left[\sum_i \binom{N_i}{2} \sum_j \binom{N_j}{2}\right] / \binom{N}{2}}, \tag{5.2}$$

where N is the number of data points in a given data set, N_{ij} is the number of data points of the class label $C_j^* \in P^*$ assigned to cluster C_i in partition P, N_i is the number of data points in cluster C_i of partition P and N_j is the number of data points in class C_j^* [99].

In this case of study, it was used to measure the similarity between each Core-Shell cluster and each one of the SST instants belonging to the correspondent upwelling span.

When analyzing the ARI scores for the three years under analysis, which can be observed in Table 5.2, one can conclude that the majority of the ARI scores obtained are higher than 0.7 and the mean values of such scores are close to 0.8, which indicates a high similarity between each Core-Shell cluster and the SST instants' segmentations that composed it. The highest ARI scores obtained for each year are highlighted in a bold font.

For the year of 2007, there is not a single instant where the ARI index was lower than 0.7. Adding to that, it is possible to note that the SST instants correspondent to the first upwelling span and thus, to the first Core-Shell cluster, are the ones that are generally less similar when compared to the correspondent Core-Shell cluster, but still with a similarity index greater than the threshold value of 0.7.

Moving on to the year 2015's results, only the ARI indexes regarding the instants 2 (Figure I.25), correspondent to the first upwelling span, and 23 (Figure I.46), which corresponds to the fourth upwelling span, are lower than the 0.7 threshold value. When manually inspecting the reason for these lower ARI values, one can conclude that both instants' segmentations do not have a great upwelling presence in the region north of Cabo

Upwelling Instants	Year 2007	Year 2015	Year 2019
1	0.726	0.766	**0.923**
2	0.735	0.652	0.799
3	0.725	**0.862**	0.802
4	0.776	0.703	0.663
5	0.752	0.833	0.787
6	0.724	0.829	0.714
7	0.731	0.822	0.748
8	0.707	0.831	0.775
9	0.789	0.819	0.811
10	0.926	0.820	0.781
11	**0.939**	0.775	0.813
12	0.928	0.775	0.830
13	0.766	0.772	0.843
14	0.789	0.753	0.790
15	0.812	0.764	0.791
16	0.825	0.776	0.808
17	0.835	0.773	0.805
18	0.841	0.747	0.811
19	0.851	0.721	0.798
20	0.791	0.722	0.819
21	0.911	0.709	0.821
22	0.913	0.725	0.769
23	0.888	0.676	0.763
Mean	0,812 ± 0,076	0,766 ± 0,055	0,794 ± 0,048

Table 5.2: ARI scores when Core-Shell clusters were compared to the correspondent upwelling instants

da Roca while the correspondent Core-Shell clusters present higher upwelling intensity in such region.

Finally, regarding the year 2019, only the ARI score for the instant 4 (Figure I.50), correspondent to the second upwelling span, is under the 0.7 threshold. The northern region of the Core-Shell cluster for the second instant has high presence on the northern region which extends to the offshore waters. This behavior occurred due to the obtained S-STSEC segmentation of instant 5 (Figure I.51), which got an over segmented result on the northern region of Cabo da Roca and thus contributed to the presence of such a big upwelling region on the northern region of the obtained Core-Shell cluster. For the rest of the year, the ARI index scores show a very good similarity between the Core-Shell clusters obtained and the SST instants' segmentations that belong to each Core-Shell cluster's upwelling span.

When looking at the overall results regarding the ARI similarities, all the years got an average for these scores above 0.76 and standard deviations lower than 0.08, which shows that overall the partitions obtained by the IAP method are of high quality without significant deviations from the average results obtained, which allows to obtain Core-Shell clusters for each upwelling span that represent very well the SST segmentations of the instants that belong to it.

Kulczynski Similarity Index: After inspecting the ARI index scores, a more precise evaluation analysis was conducted, where a similarity measure was used to validate the obtained Core-Shell clusters based on the Kulczynski Similarity (KS) index [47].

This new measure was introduced in [102] by Zakani et al. called the Weighted Kulczynski Similarity (WKS) index. This index has an advantage when compared to the original measure, which is allowing the comparison of one single cluster, for example a segmentation result, with a set of associated ground-truth clusters.

Let $\{S_1,...,S_n\}$ be the set of ground truths where each $S_i = \{R^1_{S_i},...,R^k_{S_i}\}$ is a ground truth segmentation and $A = \{R^1_A,...,R^m_A\}$ the result partition from a given segmentation algorithm. First, for each R^i_A, its optimal correspondent segments $R^{i_t}_{S_j}$ is searched for every available ground truth as follows:

$$i_t = max_j(max_k \|R^i_A \cap R^k_{S_j}\|) \tag{5.3}$$

Then, the Kul Similarity index between each pair R^i_A and $R^{i_t}_{S_j}$ is computed using the formula:

$$KS(R^i_A, R^{i_t}_{S_j}) = \frac{1}{2}\left(\frac{\|R^i_A \cap R^{i_t}_{S_j}\|}{\|R^i_A \cap R^{i_t}_{S_j}\| + \|R^i_A \setminus R^{i_t}_{S_j}\|} + \frac{\|R^i_A \cap R^{i_t}_{S_j}\|}{\|R^i_A \cap R^{i_t}_{S_j}\| + \|R^{i_t}_{S_j} \setminus R^i_A\|}\right), \tag{5.4}$$

with $\|x\|$ being the cardinality of the set x and $\setminus$ the set difference operation. The WKS is then computed as the average of all the Kul Similarity indexes computed:

$$WKS(A, \{S_n\}) = \frac{\sum_{i=1}^{m} KS(R^i_A, R^{i_t}_{S_j})}{m} \tag{5.5}$$

For the present study, this is an interesting measure because it allows to compare each obtained Core-Shell cluster with all the SST instants' segmentations from which it was built.

Moving on to the analysis of the obtained WKS index, in the study made from Zakani et al., the results were classified as very good segmentations when these had values above 0.8 for the WKS index, like the scale used in the Jaccard similarity index [40], and were classified as perfect segmentations when these values were closer to a value of 1.

It is yet important to note that the segmentations being analyzed in the present study are of greater difficulty and thus more challenging due to the fact that the upwelling patterns are blurred and with a not well-defined front separating these regions from a non-upwelling region.

Core-Shell Clusters	Year 2007	Year 2015	Year 2019
1	0.857	0.866	0.895
2	0.902	0.882	0.855
3	0.885	0.875	0.882
4	0.919	0.857	0.878
Mean	0.891 ± 0.026	0.870 ± 0.011	0.878 ± 0.017

Table 5.3: Weighted Kulczynski similarity indexes when comparing each Core-Shell cluster with the SST instants' segmentations that composed it for each year

When looking at the obtained WKS scores, one can conclude that the results for all the similarities between each Core-Shell cluster and its correspondent SST instants' segmentations are greater than 0.8. Note that the second upwelling span of the year 2019 was the one with the lowest WKS index value despite still being well above the 0.8 threshold.

Looking at the average WKS scores for each year, one can conclude that these are very close to 0.9, with a low standard deviation that is never above 0.026, confirming the quality of the obtained results.

In summary the WKS similarities between the Core-Shell clusters and the correspondent SST instants' segmentations received a classification between relatively good results and perfect results, which is concordant with the ARI similarities' analysis.

5.8 Analysis of Upwelling Spans Time Series

Having all the Core-Shell clusters correspondent to all the upwelling spans for a given year, several features were extracted and used to create several new time series, each one of these being called an upwelling span time series.

Like the time series analyzed in Section 5.6, the mean, minimum, maximum and median temperatures of the Cores, Shells and complete Core-Shell clusters were obtained, along with the temperature differences between the Core's region and the non-upwelling regions, and the areas of such obtained structures.

All the collected information about each feature for each year was divided into 3 time series: 1 for the full map of the Portuguese coast, 1 for the region north of Cabo da Roca, and 1 for the region south of Cabo da Roca.

5.8.1 Analysis of Core upwelling vs offshore temperatures over upwelling spans

The first analysis to be made was the comparison of the mean Cores' temperatures with the mean non upwelled waters' temperatures, which plots can be observed in Figure 5.16 for the years 2007, 2015, 2019.

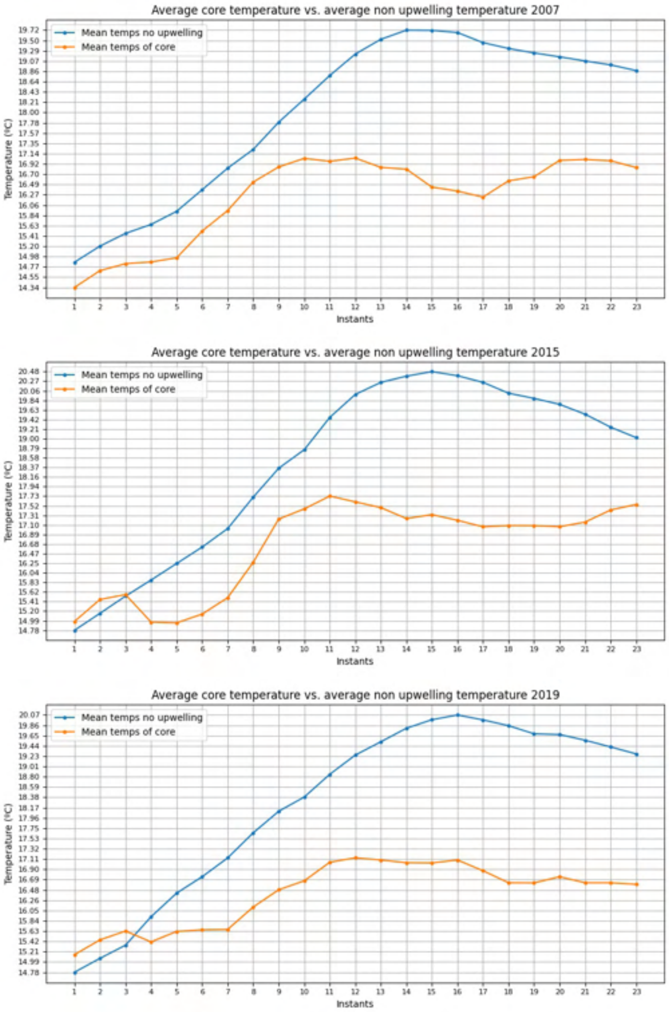


Figure 5.16: Comparison of the mean core's temperatures with the mean non-upwelling waters' temperatures for the years 2007, 2015 and 2019

The results are as expected regarding the year 2007, where the average Core's temperatures are very distinct and of lower values than the mean non upwelled waters' temperatures. This difference is smaller when the upwelling presence is of lower intensity, but as the upwelling activity tends to increase throughout the year, the differences between the obtained average Cores' and non upwelled waters' temperatures increase as well.

The same pattern can be observed for the years of 2015 and 2019 in the middle and bottom Figures, where the mean Cores' temperatures are generally lower when compared to the mean non upwelled waters' temperatures. Such does not happen in the first three instants for both years due to small over segmented results regarding the S-STSEC segmentations of such instants, like it was noted in Section 5.5.

Regarding all the three years under study, it was also noted that a strong increase in the gradient between the Core's and non upwelled waters' temperatures between the instants 9 and 11 occurs, which coincides with summer where the upwelling events tend to get stronger.

Overall, the results presented for the three years being studied were as expected when looking at the comparison between the Core's temperatures and the mean non upwelled waters' temperatures, thus validating the method for obtention of the Cores of the Core-Shell clusters.

5.8.2 Analysis of Core-Shell temperatures over upwelling spans

Moving on to the analysis regarding the comparison and analysis between the S-STSEC instants segmentations', whole Core-Shell clusters', Cores' and Shells' mean temperatures throughout a given year, the temperatures are analyzed looking at the entire Portuguese coastline as whole, and also to the regions north of and south of Cabo da Roca.

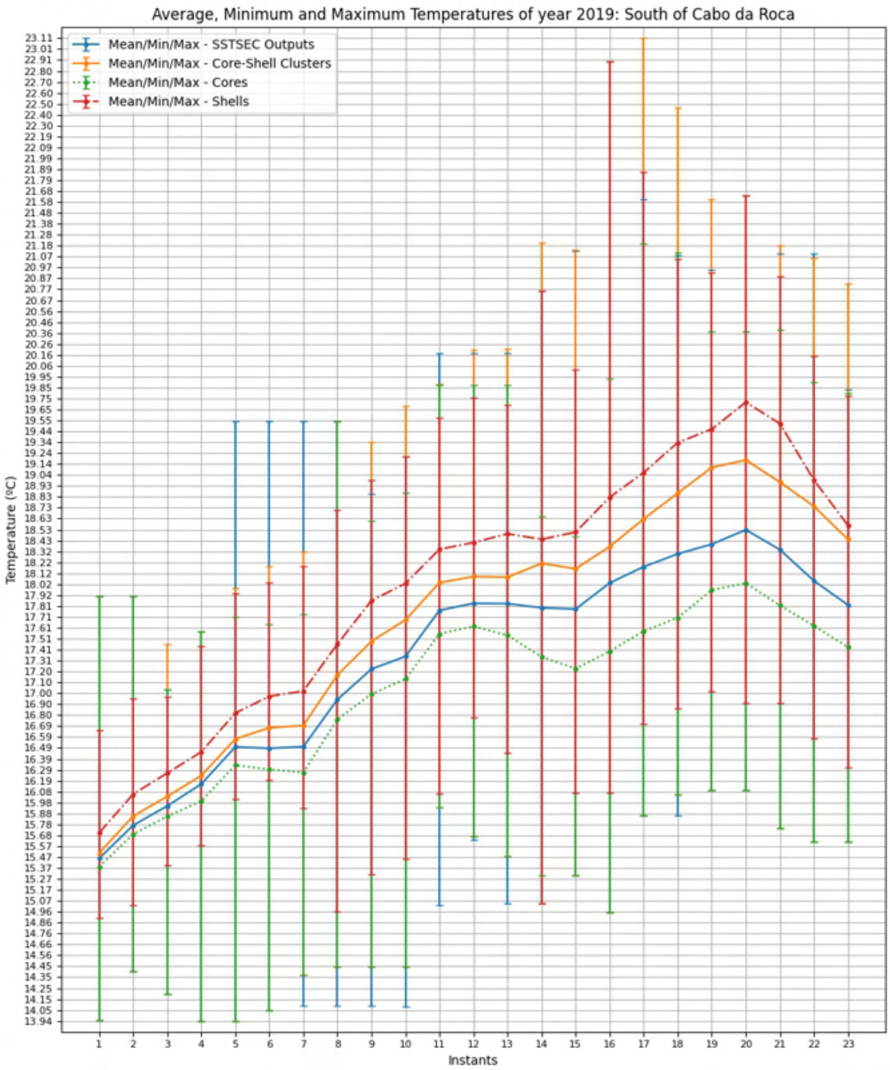


Figure 5.17: Mean temperatures for the segmented upwelling regions, Cores, Shells and Core-Shell clusters of the year 2019 on the region south of Cabo da Roca. Upwelling span 1 contains instants 1 to 3, upwelling span 2 contains instants 4 to 7, upwelling span 3 contains instants 8 to 13 and upwelling span 4 contains instants 14 to 23

When looking at the average temperatures of the structures aforementioned for the years of 2007, in Figures A.1, A.2 and A.3, 2015 in Figures A.4, A.5 and A.6 and 2019 in Figures A.7, A.8 and 5.17, all of these show a steady increase of mean water temperatures throughout the defined upwelling season, reaching a maximum average temperature value in the second half of each season.

For the years of 2007 and 2019, the average temperatures of the Cores' and Shells' regions are very similar in the first half of the upwelling season. From then on, the differences between these tend to increase and maintain a greater difference for the second half of the upwelling season. Regarding the year of 2015, the average temperatures of the Cores and Shells keep a constant difference throughout the entire upwelling season.

Regarding the Cores' and Shells' average temperatures, it is important to remember that the Cores are supposed to have a lower average temperature when compared to the Shells' when looking at the same instant. When looking at the average temperature on the regions north of and south of Cabo da Roca, it is important to note that this behavior happens much more clearly on the regions south of Cabo da Roca, while on the regions North of Cabo da Roca, the sea surface temperatures of the Shells regions are sometimes lower than the Cores' temperatures. This can be explained by the obtained S-STSEC segmentations of the first few instants of the upwelling seasons studied on the northern region of the Portuguese coastal waters, where like aforementioned, over segmented results were obtained.

The desired behavior regarding the temperatures of the obtained Cores' and Shells' regions can be well observed in Figure 5.17, which refers to the region south of Cabo da Roca for the year of 2019 and where the temperatures of the Cores are always considerably lower than the temperatures of the Shells.

5.8.3 Analysis of Core-Shell areas over upwelling spans

Finalizing the analysis with the area's time series, a comparison was made between the areas of 4 different derived structures: the SST instants segmentations' results, the obtained Core-Shell clusters, and the Cores and Shells obtained for the previously mentioned Core-Shell clusters.

When looking at all the information regarding the computed areas for the years 2007, 2015 and 2019 correspondent to the Figures 5.18, 5.19 and 5.20 respectively, there are some existing details in common between the 3 years. Note that since a single upwelling span only has a single Core-Shell cluster, composed by 1 core and N shells, for each instant $n \in N$, the area of the Core-Shell cluster is the result of the sum between the Core's area and the Shell's area of such instant n.

One detail that was perceived in this analysis was that the Shells have a varying area through time with a pattern very similar to the Core-Shell clusters' and SST instants segmentations' areas. This happens because the Shells represent the dynamics of a given

Areas of year 2007

Areas of year 2007: North of Cabo da Roca

Areas of year 2007: South of Cabo da Roca

Figure 5.18: Computed areas for the whole Portuguese coast and regions north of and south of Cabo da Roca for the year 2007 regarding the SST instants segmentation results, the obtained Core-Shell clusters, and the Cores and Shells of the respective Core-Shell clusters

upwelling region during an upwelling season. With this statement, it is possible to conclude that, for each year here studied, the Shells are correctly representing the dynamics of the entire upwelling season.

When comparing the areas of the full Portuguese coast with the regions south of and north of Cabo da Roca, it is possible to observe that the Cores' and Shells' areas of the region north of Cabo da Roca, for the 3 years, follow a similar pattern to the Cores' and Shells' areas of the full Portuguese coast.

Unlike the region north of Cabo da Roca, the region south of Cabo da Roca presents a very distinct behavior when compared to the region north of Cabo da Roca and the full Portuguese coast, thus not presenting any common behavior when comparing to these two time series.

Concluding such analysis of the upwelling span time series for the 3 years being studied, it is important to note that the Shells capture correctly the pretended dynamic behavior of the upwelling events under research, and the areas north of Cabo da Roca of

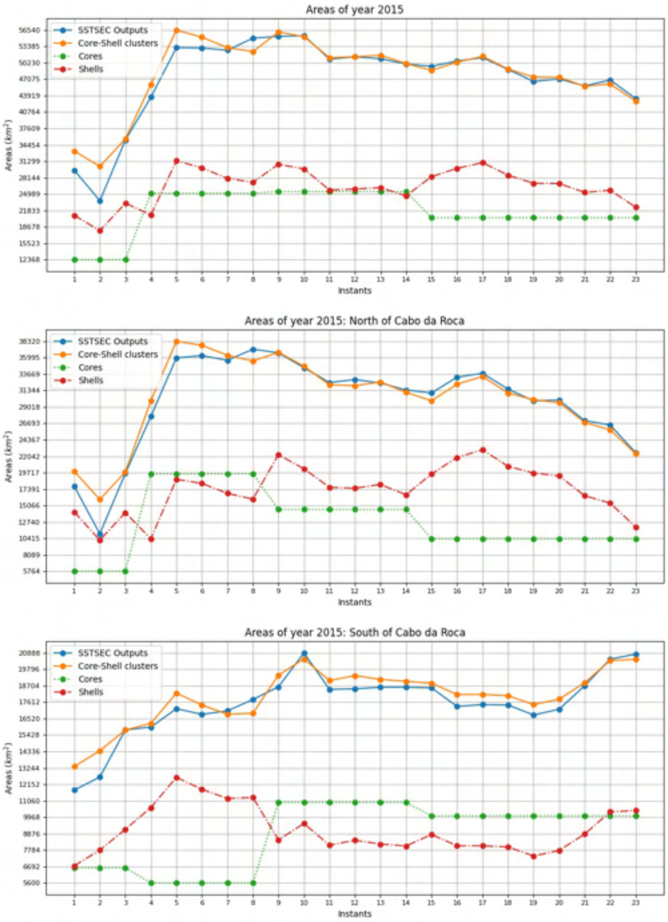


Figure 5.19: Computed areas for the whole Portuguese coast and regions north of and south of Cabo da Roca for the year 2015 regarding the SST instants segmentation results, the obtained Core-Shell clusters, and the Cores and Shells of the respective Core-Shell clusters

all the structures under analysis in this Section follow an alike behavior when compared to the entirety of the Portuguese coast region.

5.9 Summary

All the experiments belonging to the experimental study were done using a 15-inch MacBook Pro 2018, with a 2.2 GHz 6-Core Intel Core i7 processor, 16 GB 2400 MHz DDR4 of memory and a Radeon Pro 555X 4 GB graphics card, on the macOS Big Sur version 11.6 operating system.

5.9.1 Running time of experiments

The average time of running each step of the experimental study for one SST image, excluding the Core-Shell clustering algorithm step where the times are regarding the whole experiment, are as follows:

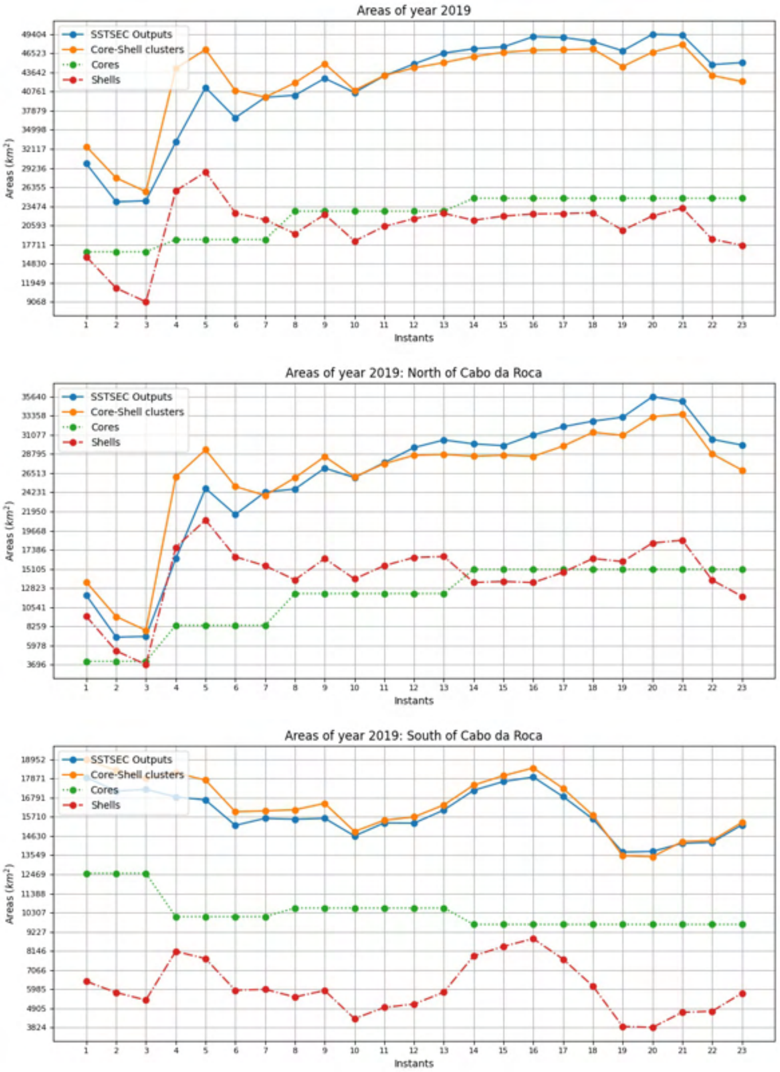


Figure 5.20: Computed areas for the whole Portuguese coast and regions north of and south of Cabo da Roca for the year 2019 regarding the SST instants segmentation results, the obtained Core-Shell clusters, and the Cores and Shells of the respective Core-Shell clusters

- Preprocessing pipeline: 12.6 seconds
- S-STSEC segmentations: 15.6 seconds
- Definition of upwelling spans by IAP: 0.47 seconds
- Core-Shell clustering algorithm:
  - 2007: 5:57 mins
  - 2015: 10:50 mins
  - 2019: 18:37 mins

It is important to note that the variable which impacted the most the running time of the experiments for each year was the length T of the defined upwelling spans. Since in this step there is an exponential number of combinations to be tested 2^T, the average running time increases exponentially whenever the length T of a defined upwelling span increases by 1. Therefore, looking at the running time of the Core-Shell clustering algorithm for different T values:

- T=3: 00:08 mins;
- T=4: 00:14 mins;
- T=5: 00:26 mins;
- T=6: 00:54 mins;
- T=7: 01:42 mins;
- T=8: 03:43 mins;
- T=9: 09:13 mins;
- T=10: 17:27 mins.

5.9.2 Outlook of the obtained results

Looking at the results obtained, the preprocessing pipeline developed demonstrated to be very successful, not only eliminating and overcoming almost completely the issue of over segmentation but also removing any unwanted noise from a set of SST images. Such method shown to not work as well on the regions where the coast had a non-vertical orientation, these being the northern and southern coasts of Portugal. Despite these small setbacks, the results obtained were already very promising and of high quality.

With a good preprocessing method implemented, the S-STSEC algorithm shown once again its precision and efficiency when segmenting upwelling regions from SST imagery data, obtaining very good grades by the team of experts who evaluated the obtained segmentations. This algorithm shown to be more effective whenever the upwelling event was of greater intensity, causing small over segmented results in very few occurrences of low intensity upwelling events.

For the unsupervised definition of the upwelling spans, the IAP algorithm shown to be appropriate for such task. Having the upwelling spans, the Core-Shell clustering algorithm obtained Cores and respective Shells that captured very well the constant and dynamic regions of the upwelling events studied. When the results were analyzed using the ARI and WKS measures, these shown that the use of the IAP algorithm for the unsupervised upwelling span discovery and the Core-Shell clustering algorithm is of high proficiency and quality to model upwelling spans.

6

Conclusion and Future Work

6.1 Main Contributions

The work here developed started with the creation of a preprocessing pipeline, with the goals of overcoming the explosion problem present in segmentations provided by SRG based algorithms, like the S-STSEC algorithm, and also with the goal of preparing a set of SST images for the Core-Shell clustering algorithm also developed.

A moving average filter was implemented to prepare a set of weekly SST images to be used by the Core-Shell clustering algorithm. Such procedure also helped to remove white noise from the images. This method produced a smoother time series of averaged SST images, ready to be preprocessed. The averaged SST images looked natural and with good quality, allowing for an easier perception of the evolution of the upwelling regions through time.

The smoothing of the North-South temperature gradient and the Enhancement of the East-West temperature gradient were the two methods developed to tackle the explosion problem. These methods, when used together, allowed to get much improved segmentation results when using the S-STSEC algorithm, avoiding the over segmented areas on the northern region of the Portuguese coast. Such segmentations were reviewed by two experts, who stated that the results were very promising and of good precision. When using the S-STSEC algorithm alongside with the two preprocessing methods aforementioned, the segmentation results tend to be better when the upwelling presents a higher intensity. Whenever an upwelling episode was of lower intensity, the segmentation results were of lower but still of satisfactory quality.

Several features were obtained from the S-STSEC segmentations' time series, like the upwelled waters' area, mean temperatures of the upwelled waters and mean temperatures of the offshore waters, and such were analyzed and compared with the real world behavior, coming to the conclusion that the obtained upwelling segmentations represented very well the real world behavior of upwelling regions, regarding not only the areas but also the upwelling temperatures for the region of Portugal through time.

Such features were also used to perform time series segmentation using the IAP algorithm, with the intent of obtaining several upwelling spans for each upwelling season where in each one the upwelling behavior was similar and followed a constant pattern.

When looking at the Core-Shell clusters created by the Core-Shell clustering algorithm from the upwelling spans created by the IAP algorithm, these represented very well the correspondent upwelling spans. Such statement is backed by the experts' evaluations of such representations.

The obtained Shells for the Core-Shell clusters managed to represent properly the dynamics of the upwelling events when looking at the time series of the Shells' and Cores' areas and temperatures through time.

It is also possible to conclude the effectiveness of the unsupervised upwelling span discovery method, due to the fact that when comparing the obtained Core-Shell clusters with the SST images from the correspondent spans, the average similarity measures obtained were higher than 85%.

6.2 Future Work

The future work following the progress made in the current dissertation is as follows:

1. Improve the method for the East-West temperature gradient enhancement, to overcome the slight over segmented results that rarely occur when segmenting SST images using the S-STSEC algorithm;

2. Improve the approach when applied to massive data of a given geographical location;

3. Extend the approach for other geographical locations of the globe. A good adaptability is expected due to previous results obtained by the S-STSEC algorithm [58];

4. Develop a more robust unsupervised approach for the time series segmentation procedure, with the goal of describing coastal upwelling profiles;

5. Improve the Core-Shell clustering method and validation protocol;

6. Develop a comparison study between the IAP algorithm and the SWAB algorithm for the task of time series segmentation.

Bibliography

[1] Z. E. Abidi et al. “An Efficient Detection of Moroccan Coastal Upwelling Based on Fusion of Chlorophyll-a and Sea Surface Temperature Images With a New Validation Index”. In: *IEEE Geoscience and Remote Sensing Letters* (2020), pp. 1–5. DOI: `10.1109/LGRS.2020.3002473` (cit. on pp. 2, 13).

[2] Z. E. Abidi et al. “Delineation of Moroccan Coastal Upwelling Using The Principal Component Analysis Fusion Algorithm on SSC and SST Images”. In: *2018 9th International Symposium on Signal, Image, Video and Communications - ISIVC 2018*. 2018, pp. 174–178. ISBN: 978-1-5386-8173-2. DOI: `10.1109/ISIVC.2018.8709227` (cit. on p. 30).

[3] R. Adams and L. Bischof. “Seeded region growing”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 16.6 (1994), pp. 641–647. DOI: `10.1109/34.295913` (cit. on pp. 2, 9, 10, 13).

[4] A. Afifi et al. “New Region Growing based on Thresholding Technique Applied to MRI Data”. In: *IJCNIS* 7.7 (June 2015), pp. 61–67. DOI: `10.5815/ijcnis.2015.07.08` (cit. on p. 10).

[5] I. Aganj et al. “Unsupervised Medical Image Segmentation Based on the Local Center of Mass”. In: *Scientific Reports* 8.1 (2018), pp. 1–8. ISSN: 20452322. DOI: `10.1038/s41598-018-31333-5` (cit. on p. 6).

[6] K. P. Agrawal et al. “Development and validation of OPTICS based spatio-temporal clustering technique”. In: *Information Sciences* 369 (2016), pp. 388–401. ISSN: 0020-0255. DOI: `https://doi.org/10.1016/j.ins.2016.06.048`. URL: `http://www.sciencedirect.com/science/article/pii/S0020025516304765` (cit. on p. 22).

[7] M. A. Alkhamis et al. “Spatiotemporal dynamics of the COVID-19 pandemic in the State of Kuwait”. In: *International Journal of Infectious Diseases* 98 (2020), pp. 153–160. ISSN: 1201-9712. DOI: `https://doi.org/10.1016/j.ijid.2020.06.078`. URL: `http://www.sciencedirect.com/science/article/pii/S1201971220305129` (cit. on p. 22).

[8] S. Aminikhanghahi and D. J. Cook. “A survey of methods for time series change point detection”. In: *Knowledge and Information Systems* 51 (2 May 2017), pp. 339–367. ISSN: 02193116. DOI: 10.1007/s10115-016-0987-z (cit. on pp. 26, 27).

[9] E. Anjna and E. Kaur. “Review of Image Segmentation Technique”. In: *International Journal of Advanced Research in Computer Science* 8.4 (May 2017). ISSN: 0976-5697 (cit. on p. 6).

[10] M. Y. Ansari et al. “Spatiotemporal clustering: a review”. In: *Artificial Intelligence Review* 53.4 (2020), pp. 2381–2423. ISSN: 1573-7462. DOI: 10.1007/s10462-019-09736-1. URL: https://doi.org/10.1007/s10462-019-09736-1 (cit. on pp. 20, 22).

[11] A. Aouni et al. “Physical and Biological Satellite Observations of the Northwest African Upwelling: Spatial Extent and Dynamics”. In: *IEEE Transactions on Geoscience and Remote Sensing* 99 (Oct. 2019), pp. 1–13. DOI: 10.1109/TGRS.2019.2946300 (cit. on p. 13).

[12] A. E. Aouni et al. “Robust Detection of the North-West African Upwelling From SST Images”. In: *IEEE Geoscience and Remote Sensing Letters* 99 (Apr. 2020), pp. 1–4. DOI: 10.1109/LGRS.2020.2983826 (cit. on pp. 2, 13, 30, 33).

[13] G. Atluri, A. Karpatne, and V. Kumar. “Spatio-temporal data mining: A survey of problems and methods”. In: *arXiv* 51.4 (2017). ISSN: 23318422. arXiv: 1711.04710 (cit. on pp. 19, 22, 23, 39).

[14] D. Birant and A. Kut. “ST-DBSCAN: An algorithm for clustering spatial–temporal data”. In: *Data Knowledge Engineering* 60.1 (2007), pp. 208–221. ISSN: 0169-023X. DOI: https://doi.org/10.1016/j.datak.2006.01.013. URL: http://www.sciencedirect.com/science/article/pii/S0169023X06000218 (cit. on p. 22).

[15] D. J. Bora and D. A. K. Gupta. “A Comparative study Between Fuzzy Clustering Algorithm and Hard Clustering Algorithm”. In: *International Journal of Computer Trends and Technology* 10.2 (2014), pp. 108–113. ISSN: 23490829. DOI: 10.14445/22312803/ijctt-v10p119 (cit. on p. 7).

[16] M. Bose and K. Mali. “A novel data partitioning and rule selection technique for modeling high-order fuzzy time series”. In: *Applied Soft Computing Journal* 63 (Feb. 2018), pp. 87–96. ISSN: 15684946. DOI: 10.1016/j.asoc.2017.11.011 (cit. on p. 27).

[17] M. Brilliant et al. “Land Cover Changes Detection Using Region Growing Segmentation”. In: *Journal of Physics: Conference Series* 1529.2 (2020). ISSN: 17426596. DOI: 10.1088/1742-6596/1529/2/022066 (cit. on p. 10).

[18] S. Chaudhari, R. Balasubramanian, and A. Gangopadhyay. “Upwelling detection in avhrr sea surface temperature (sst) images using neural-network framework”. In: *International Geoscience and Remote Sensing Symposium (IGARSS)* 4.1 (2008), pp. 926–929. DOI: `10.1109/IGARSS.2008.4779875` (cit. on pp. 1, 12).

[19] L. Chen et al. “Road extraction from VHR remote-sensing imagery via object segmentation constrained by Gabor features”. In: *ISPRS International Journal of Geo-Information* 7.9 (Sept. 2018). ISSN: 22209964. DOI: `10.3390/ijgi7090362` (cit. on p. 10).

[20] X. C. Chen et al. “Clustering Dynamic Spatio-Temporal Patterns in the Presence of Noise and Missing Data”. In: *Proceedings of the 24th International Conference on Artificial Intelligence*. IJCAI’15. Buenos Aires, Argentina: AAAI Press, 2015, pp. 2575–2581. ISBN: 9781577357384 (cit. on pp. iv, v, 2, 3, 23, 24).

[21] C. K. Chow and T. Kaneko. “Automatic boundary detection of the left ventricle from cineangiograms”. In: *Computers and Biomedical Research* 5.4 (1972), pp. 388–410. ISSN: 0010-4809. DOI: `https://doi.org/10.1016/0010-4809(72)90070-5`. URL: `https://www.sciencedirect.com/science/article/pii/0010480972900705` (cit. on p. 11).

[22] A. P. Dempster, N. M. Laird, and D. B. Rubin. “Maximum Likelihood from Incomplete Data Via the EM Algorithm”. In: *Journal of the Royal Statistical Society: Series B (Methodological)* 39.1 (1977), pp. 1–22. DOI: `10.1111/j.2517-6161.1977.tb01600.x` (cit. on p. 7).

[23] N. Dhanachandra and Y. J. Chanu. “A Survey on Image Segmentation Methods using Clustering Techniques”. In: *European Journal of Engineering Research and Science* 2.1 (2017), p. 15. DOI: `10.24018/ejers.2017.2.1.237` (cit. on p. 7).

[24] M. G. Doborjeh and N. Kasabov. “Dynamic 3D Clustering of Spatio-Temporal Brain Data in the NeuCube Spiking Neural Network Architecture on a Case Study of fMRI Data”. In: *Neural Information Processing*. Ed. by S. Arik et al. Springer International Publishing, 2015, pp. 191–198. ISBN: 978-3-319-26561-2 (cit. on p. 22).

[25] J. C. Dunn. “A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters”. In: *Journal of Cybernetics* 3.3 (1973), pp. 32–57. ISSN: 00220280. DOI: `10.1080/01969727308546046` (cit. on p. 8).

[26] M. Ester et al. “A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise”. In: *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*. AAAI Press, 1996, pp. 226–231. ISBN: 1-57735-004-9 (cit. on p. 3).

[27] H. Fan et al. “A novel breast ultrasound image automated segmentation algorithm based on seeded region growing integrating gradual equipartition threshold”. In: *Multimedia Tools and Applications* 78 (June 2019), pp. 27915–27932. DOI: `10.1007/s11042-019-07884-8` (cit. on p. 10).

[28] Y. Feng et al. “A multi-scale 3D Otsu thresholding algorithm for medical image segmentation”. In: *Digital Signal Processing: A Review Journal* 60 (2017), pp. 186–199. ISSN: 10512004. DOI: `10.1016/j.dsp.2016.08.003`. URL: `http://dx.doi.org/10.1016/j.dsp.2016.08.003` (cit. on p. 11).

[29] A. Fiori, A. Mignone, and G. Rospo. “DeCoClu: Density consensus clustering approach for public transport data”. In: *Information Sciences* 328 (2016), pp. 378–388. ISSN: 00200255. DOI: `10.1016/j.ins.2015.08.054`. URL: `http://dx.doi.org/10.1016/j.ins.2015.08.054` (cit. on p. 23).

[30] A. F. G. Fiúza. “Upwelling Patterns off Portugal”. In: *Coastal Upwelling Its Sediment Record: Part A: Responses of the Sedimentary Regime to Present Coastal Upwelling*. Ed. by E. Suess and J. Thiede. Boston, MA: Springer US, 1983, pp. 85–98. ISBN: 978-1-4615-6651-9. DOI: `10.1007/978-1-4615-6651-9_5`. URL: `https://doi.org/10.1007/978-1-4615-6651-9_5` (cit. on p. 45).

[31] N. Ghane et al. “Segmentation of White Blood Cells from Microscopic Images Using a Novel Combination of K-Means Clustering and Modified Watershed Algorithm”. In: *Journal of Medical Signals and Sensors* 7.2 (2017), pp. 92–101. ISSN: 22287477. DOI: `10.4103/2228-7477.205503` (cit. on p. 8).

[32] J. Gomide et al. “Dengue Surveillance Based on a Computational Model of Spatio-Temporal Locality of Twitter”. In: *Proceedings of the 3rd International Web Science Conference*. WebSci ’11. Koblenz, Germany: Association for Computing Machinery, 2011. ISBN: 9781450308557. DOI: `10.1145/2527031.2527049`. URL: `https://doi.org/10.1145/2527031.2527049` (cit. on p. 22).

[33] F. Guindos-Rojas et al. “Automatic Recognition of Ocean Structures From Satellite Images By Means of”. In: *Proceedings of ESA-EUSC* 1 (2004), pp. 2–6 (cit. on p. 12).

[34] D. Gustafson and W. Kessel. “Fuzzy Clustering with a Fuzzy Covariance Matrix”. In: *Proceedings of the IEEE-CDC*. Vol. 2. 1978, pp. 761–766. DOI: `10.1109/CDC.1978.268028` (cit. on p. 13).

[35] B. Higgs and M. Abbas. “Segmentation and Clustering of Car-Following Behavior: Recognition of Driving Patterns”. In: *IEEE Transactions on Intelligent Transportation Systems* 16.1 (2015), pp. 81–90. DOI: `10.1109/TITS.2014.2326082` (cit. on p. 23).

[36] Y. Hu et al. "A spatio-temporal kernel density estimation framework for predictive crime hotspot mapping and evaluation". In: *Applied Geography* 99 (2018), pp. 89–97. ISSN: 0143-6228. DOI: `https://doi.org/10.1016/j.apgeog.2018.08.001`. URL: `http://www.sciencedirect.com/science/article/pii/S0143622818300560` (cit. on p. 23).

[37] Z. Huang and X. H. Wang. "Mapping the spatial and temporal variability of the upwelling systems of the Australian south-eastern coast using 14-year of MODIS data". In: *Remote Sensing of Environment* 227.September 2018 (2019), pp. 90–109. ISSN: 00344257. DOI: `10.1016/j.rse.2019.04.002`. URL: `https://doi.org/10.1016/j.rse.2019.04.002` (cit. on p. 18).

[38] Z. Huang et al. "Weakly-Supervised Semantic Segmentation Network with Deep Seeded Region Growing". In: *Conference on Computer Vision and Pattern Recognition - CVPR 2018*. 2018, pp. 7014–7023. ISBN: 978-1-5386-6421-6. DOI: `10.1109/CVPR.2018.00733` (cit. on p. 9).

[39] T. Ijiri et al. "Spatially controllable region growing for segmenting heart chambers". In: *Advanced Biomedical Engineering* 9 (2020), pp. 172–180. ISSN: 21875219. DOI: `10.14326/ABE.9.172` (cit. on p. 9).

[40] P. Jaccard. *HEW PHYTOIiOGIST. THE DISTRIBUTION OF THE FLORA IN THE ALPINE ZONE.' Professor at the Federal Polyteehnic, Ziirich* (cit. on p. 64).

[41] P. Kalnis, N. Mamoulis, and S. Bakiras. "On Discovering Moving Clusters in Spatio-temporal Data". In: *Advances in Spatial and Temporal Databases*. Ed. by C. Bauzer Medeiros, M. J. Egenhofer, and E. Bertino. Berlin, Heidelberg: Springer Berlin Heidelberg, 2005, pp. 364–381. ISBN: 978-3-540-31904-7 (cit. on pp. 2, 23).

[42] H. Kaur. "Review of Remote Sensing Image Segmentation Techniques". In: *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)* 4.4 (4 2015), pp. 1667–1674. ISSN: 2278-1323. URL: `http://ijarcet.org/wp-content/uploads/IJARCET-VOL-4-ISSUE-4-1667-1674.pdf` (cit. on p. 6).

[43] H. Kaur and R. Sharma. "Segmentation of Tumor Region from Brain Mri Images Using Fuzzy C-Means Clustering And Seeded Region Growing". In: *IOSR Journal of Computer Engineering* 18 (May 2016), pp. 20–24. DOI: `10.9790/0661-1805012024` (cit. on p. 8).

[44] E. Keogh et al. "An online algorithm for segmenting time series". In: *Proceedings 2001 IEEE International Conference on Data Mining*. 2001, pp. 289–296. DOI: `10.1109/ICDM.2001.989531` (cit. on p. 28).

[45] S. Kisilevich et al. "Spatio-Temporal Clustering". In: *Data Mining and Knowledge Discovery Handbook* (2010). DOI: `10.1007/978-0-387-09823-4` (cit. on p. 21).

[46] J. Kittler and J. Illingworth. "Minimum error thresholding". In: *Pattern Recognition* 19.1 (1986), pp. 41–47. ISSN: 0031-3203. DOI: `https://doi.org/10.1016/0031-3203(86)90030-0`. URL: `https://www.sciencedirect.com/science/article/pii/0031320386900300` (cit. on p. 11).

[47] S. Kulczynski. "Die Pflanzenassoziationen der Pieninen". In: *Bulletin International de l'Academie Polonaise des Sciences et des Lettres, Classe des Sciences Mathematiques et Naturelles, B (Sciences Naturelles)* II (1927), pp. 57–203 (cit. on p. 64).

[48] J. Liu et al. "Dual-Constraint Spatiotemporal Clustering Approach for Exploring Marine Anomaly Patterns Using Remote Sensing Products". In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 11.11 (2018), pp. 3963–3976. DOI: `10.1109/JSTARS.2018.2873216` (cit. on p. 22).

[49] J. M. Lourenço. *The NOVAthesis LaTeX Template User's Manual*. NOVA University Lisbon. 2021. URL: `https://github.com/joaomlourenco/novathesis/raw/master/template.pdf` (cit. on p. ii).

[50] M. Lovrić, M. Milanović, and M. Stamenković. *ALGORITHMIC METHODS FOR SEGMENTATION OF TIME SERIES: AN OVERVIEW*. 2014, pp. 31–53 (cit. on pp. 26, 27).

[51] Y. Lu and R. Lu. "Histogram-based automatic thresholding for bruise detection of apples by structured-illumination reflectance imaging". In: *Biosystems Engineering* 160 (2017), pp. 30–41. ISSN: 15375110. DOI: `10.1016/j.biosystemseng.2017.05.005`. URL: `http://dx.doi.org/10.1016/j.biosystemseng.2017.05.005` (cit. on p. 11).

[52] J. MacQueen et al. "Some methods for classification and analysis of multivariate observations". In: *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability* 1.14 (1967), pp. 281–297 (cit. on p. 7).

[53] J. Marcello et al. "Seasonal and temporal study of the northwest African upwelling system". In: *International Journal of Remote Sensing* 32.7 (2011), pp. 1843–1859. ISSN: 13665901. DOI: `10.1080/01431161003631576` (cit. on p. 17).

[54] B. G. (G. Mirkin. *Clustering for data mining : a data recovery approach*. Chapman Hall/CRC, 2005, p. 266. ISBN: 1584885343 (cit. on p. 36).

[55] T. Nakaya and K. Yano. "Visualising Crime Clusters in a Space-time Cube: An Exploratory Data-analysis Approach Using Space-time Kernel Density Estimation and Scan Statistics". In: *Transactions in GIS* 14.3 (2010), pp. 223–239. DOI: `https://doi.org/10.1111/j.1467-9671.2010.01194.x`. URL: `https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9671.2010.01194.x` (cit. on p. 23).

[56] S. Nascimento, S. Casca, and B. Mirkin. "A seed expanding cluster algorithm for deriving upwelling areas on sea surface temperature images". In: *Computers Geosciences* 85 (2015), pp. 74–85. ISSN: 0098-3004. DOI: `https://doi.org/10.1016/j.cageo.2015.06.002` (cit. on pp. iv, v, 2, 4, 10, 13, 16, 17, 30).

[57] S. Nascimento and P. Franco. "Segmentation of Upwelling Regions in Sea Surface Temperature Images via Unsupervised Fuzzy Clustering". In: *Intelligent Data Engineering and Automated Learning - IDEAL 2009*. Vol. 5788. Sept. 2009, pp. 543–553. ISBN: 978-3-642-04394-9. DOI: `https://doi.org/10.1007/978-3-642-04394-9_66` (cit. on p. 12).

[58] S. Nascimento and N. Madaleno. "Unsupervised Initialization of Archetypal Analysis and Proportional Membership Fuzzy Clustering". In: *Intelligent Data Engineering and Automated Learning – IDEAL 2019*. Ed. by H. Yin et al. Cham: Springer International Publishing, 2019, pp. 12–20. ISBN: 978-3-030-33617-2 (cit. on p. 74).

[59] S. Nascimento, S. Mateen, and P. Relvas. "Sequential Self-tuning Clustering for Automatic Delimitation of Coastal Upwelling on SST Images". In: *Intelligent Data Engineering and Automated Learning – IDEAL 2020*. Switzerland: Springer, Cham, 2020, pp. 434–443. ISBN: 978-3-030-62365-4. DOI: `https://doi.org/10.1007/978-3-030-62365-4_41` (cit. on pp. iv, v, 2–4, 16).

[60] S. Nascimento et al. "Automated computational delimitation of SST upwelling areas using fuzzy clustering". In: *Computers & Geosciences - COMPUT GEOSCI* 43 (2012), pp. 207–216. DOI: `10.1016/j.cageo.2011.10.025` (cit. on p. 12).

[61] A. S. A. Nasir, M. Y. Mashor, and Z. Mohamed. "Enhanced k-means clustering algorithm for malaria image segmentation". In: *Journal of Advanced Research in Fluid Mechanics and Thermal Sciences* 42.1 (2018), pp. 1–15. ISSN: 22897879 (cit. on p. 8).

[62] E. Niharika et al. "K-means based noisy SAR image segmentation using median filtering and Otsu method". In: *IEEE International Conference on IoT and its Applications, ICIOT 2017* (2017), pp. 3–6. DOI: `10.1109/ICIOTA.2017.8073630` (cit. on p. 8).

[63] Y. Noh et al. "Automatic crack detection on concrete images using segmentation via fuzzy C-means clustering". In: *Proceedings of the 2017 IEEE International Conference on Applied System Innovation: Applied System Innovation for Modern Technology, ICASI 2017* (2017), pp. 877–880. DOI: `10.1109/ICASI.2017.7988574` (cit. on p. 9).

[64] V. Oerder et al. "Coastal upwelling front detection offCentral Chile (36.5-37°S) and spatio-temporal variability of Frontal characteristics". In: *Remote Sensing* 10.5 (2018). ISSN: 20724292. DOI: `10.3390/rs10050690` (cit. on p. 17).

[65] N. Otsu. “A Threshold Selection Method from Gray-Level Histograms”. In: *IEEE Transactions on Systems, Man, and Cybernetics* 9.1 (1979), pp. 62–66. DOI: 10.1109/TSMC.1979.4310076 (cit. on p. 11).

[66] S. Park. “Water Region Segmentation Scheme using Seeded Region Growing”. In: *Journal of JAITC* 8 (July 2018), pp. 53–62. DOI: 10.14801/jaitc.2018.8.1.53 (cit. on p. 9).

[67] J. A. Piedra et al. “Pattern recognition in AVHRR images by means of Hibryd and Neuro-fuzzy Systems”. In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 3643 LNCS (2005), pp. 373–378. ISSN: 03029743. DOI: 10.1007/11556985_48 (cit. on p. 12).

[68] J. A. Piedra-Fernández, M. Cantón-Garbín, and F. Guindos-Rojas. “Application of fuzzy lattice neurocomputing (FLN) in ocean satellite images for pattern recognition”. In: *Studies in Computational Intelligence* 67 (2007), pp. 215–232. ISSN: 1860949X. DOI: 10.1007/978-3-540-72687-6_11 (cit. on pp. 1, 12).

[69] R. Qaddoura et al. “Dental radiography segmentation using expectation-maximization clustering and grasshopper optimizer”. In: *Multimedia Tools and Applications* 79.29-30 (2020), pp. 22027–22045. ISSN: 15737721. DOI: 10.1007/s11042-020-09014-1 (cit. on p. 8).

[70] T. Rakthanmanon et al. “Time series epenthesis: Clustering time series streams requires ignoring some data”. In: 2011, pp. 547–556. ISBN: 9780769544083. DOI: 10.1109/ICDM.2011.146 (cit. on p. 27).

[71] J. D. Ramanantsoa et al. “Coastal upwelling south of Madagascar: Temporal and spatial variability”. In: *Journal of Marine Systems* 178 (2018), pp. 29–37. ISSN: 09247963. DOI: 10.1016/j.jmarsys.2017.10.005. URL: http://dx.doi.org/10.1016/j.jmarsys.2017.10.005 (cit. on p. 17).

[72] W. M. Rand. “Objective criteria for the evaluation of clustering methods”. In: *Journal of the American Statistical Association* 66 (336 1971), pp. 846–850. ISSN: 1537274X. DOI: 10.1080/01621459.1971.10482356 (cit. on p. 62).

[73] M. N. Reza et al. “Rice yield estimation based on K-means clustering with graph-cut segmentation using low-altitude UAV images”. In: *Biosystems Engineering* 177 (2019), pp. 109–121. ISSN: 15375110. DOI: 10.1016/j.biosystemseng.2018.09.014. URL: https://doi.org/10.1016/j.biosystemseng.2018.09.014 (cit. on p. 8).

[74] F. Riaz et al. “An inspection approach for casting defects detection using image segmentation”. In: *2017 International Conference on Mechanical, System and Control Engineering, ICMSC 2017* (2017), pp. 101–105. DOI: 10.1109/ICMSC.2017.7959451 (cit. on p. 9).

[75] T. W. Ridler and S. Calvard. “Picture Thresholding Using an Iterative Selection Method”. In: *IEEE Transactions on Systems, Man, and Cybernetics* 8.8 (1978), pp. 630–632. DOI: 10.1109/TSMC.1978.4310039 (cit. on p. 11).

[76] S. Saremi, S. Mirjalili, and A. Lewis. “Grasshopper Optimisation Algorithm: Theory and application”. In: *Advances in Engineering Software* 105 (2017), pp. 30–47. ISSN: 0965-9978. DOI: https://doi.org/10.1016/j.advengsoft.2017.01.004. URL: http://www.sciencedirect.com/science/article/pii/S0965997816305646 (cit. on p. 8).

[77] N. A. Seada, S. Hamad, and M. G. Mostafa. “Automatically seeded region growing approach for automatic segmentation of ascending aorta”. In: *ACM International Conference Proceeding Series* (2016), pp. 127–132. DOI: 10.1145/2908446.2908479 (cit. on p. 10).

[78] A. Septiarini et al. “Optic disc and cup segmentation by automatic thresholding with morphological operation for glaucoma evaluation”. In: *Signal, Image and Video Processing* 11.5 (2017), pp. 945–952. ISSN: 18631711. DOI: 10.1007/s11760-016-1043-x (cit. on p. 11).

[79] Z. Shi and L. Pun-Cheng. “Spatiotemporal Data Clustering: A Survey of Methods”. In: *ISPRS International Journal of Geo-Information* 8 (2019). DOI: 10.3390/ijgi8030112 (cit. on p. 22).

[80] H. Shimodaira. “Automatic Color Image Segmentation Using a Square Elemental Region-Based Seeded Region Growing and Merging Method”. In: *arXiv* (2017). ISSN: 23318422 (cit. on p. 10).

[81] A. Singla and S. Patra. “A fast automatic optimal threshold selection technique for image segmentation”. In: *Signal, Image and Video Processing* 11.2 (2017), pp. 243–250. ISSN: 18631711. DOI: 10.1007/s11760-016-0927-0 (cit. on p. 11).

[82] A. Soligon, J. Neo, and X. Duan. “Automated Detection of Corrosion Damage in Power Transmission Lattice Towers Using Image Processing”. In: *Structures Congress 2017* 2013 (2017), pp. 474–482 (cit. on p. 9).

[83] X. Song et al. “A new fuzzy c-means clustering-based time series segmentation approach and its application on tunnel boring machine analysis”. In: *Mechanical Systems and Signal Processing* 133 (Nov. 2019). ISSN: 10961216. DOI: 10.1016/j.ymssp.2019.106279 (cit. on p. 27).

[84] R. A. Suji, D. B. Anand, and R. L. Babu. “Polar ice image segmentation using improved estimation and normalization of illumination”. In: *Cluster Computing* 22 (2019), pp. 3811–3819. ISSN: 15737543. DOI: 10.1007/s10586-018-2351-4. URL: https://doi.org/10.1007/s10586-018-2351-4 (cit. on p. 8).

[85] A. Tamim et al. “A simple and efficient approach for coarse segmentation of Moroccan coastal upwelling”. In: *Proceedings of the 21st European Signal Processing Conference - EUSIPCO 2013* (Sept. 2013) (cit. on p. 13).

[86] A. Tamim et al. “Automatic Detection of Moroccan Coastal Upwelling Zones using Sea Surface Temperature Images”. In: *International Journal of Remote Sensing* 40 (Oct. 2018). DOI: `10.1080/01431161.2018.1528513` (cit. on p. 13).

[87] A. Tamim et al. “Detection of Moroccan coastal upwelling in SST images using the Expectation-Maximization”. In: *Intelligent Systems and Computer Vision - ISCV 2015*. Dec. 2015. ISBN: 978-1-4799-7511-2. DOI: `10.1109/ISACV.2015.7106195` (cit. on p. 13).

[88] A. Tamim et al. “On Detectability of Moroccan Coastal Upwelling in Sea Surface Temperature Satellite Images”. In: *International Symposium on Visual Computing - ISVC 2014*. Dec. 2014, pp. 386–395. ISBN: 978-3-319-14363-7. DOI: `10.1007/978-3-319-14364-4_37` (cit. on p. 13).

[89] A. Tamim et al. “Upwelling Detection in SST Images Using Fuzzy Clustering with Adaptive Cluster Merging”. In: (Nov. 2014) (cit. on p. 13).

[90] A. M. Taori et al. “Segmentation of macula in retinal images using automated seeding region growing technique”. In: *2016 International Conference on Inventive Computation Technologies (ICICT)*. Vol. 2. 2016, pp. 1–5. DOI: `10.1109/INVENTIVE.2016.7824792` (cit. on p. 10).

[91] *The Generic Mapping Tools*. 2019. URL: `https://www.generic-mapping-tools.org` (visited on 11/11/2021) (cit. on pp. 31, 48).

[92] D.-H. Tran. “Automated Change Detection and Reactive Clustering in Multivariate Streaming Data”. In: (Nov. 2013). URL: `http://arxiv.org/abs/1311.0505` (cit. on p. 27).

[93] L. Vincent and P. Soille. “Watersheds in digital spaces: an efficient algorithm based on immersion simulations”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 13.6 (1991), pp. 583–598. DOI: `10.1109/34.87344` (cit. on p. 10).

[94] J. Wang et al. “A Region-Based SRG Algorithm for Color Image Segmentation”. In: *2007 International Conference on Machine Learning and Cybernetics*. Vol. 3. 2007, pp. 1542–1547. DOI: `10.1109/ICMLC.2007.4370390` (cit. on p. 10).

[95] S. Wang, J. Cao, and P. S. Yu. “Deep learning for spatio-temporal data mining: a survey”. In: *IEEE Transactions on Knowledge and Data Engineering* 14.8 (2020). ISSN: 23318422. DOI: `10.1109/tkde.2020.3025580`. eprint: `1906.04928` (cit. on p. 19).

[96] S. Wazarkar and B. N. Keshavamurthy. "A survey on image data analysis through clustering techniques for real world applications". In: *Journal of Visual Communication and Image Representation* 55.October 2017 (2018), pp. 596–626. ISSN: 1047-3203. URL: `https://doi.org/10.1016/j.jvcir.2018.07.009` (cit. on p. 8).

[97] W. Wooster, A. Bakun, and D. McLain. "SEASONAL UPWELLING CYCLE ALONG THE EASTERN BOUNDARY OF THE NORTH ATLANTIC." English (US). In: *Journal of Marine Research* 34.2 (Jan. 1976), pp. 131–141. ISSN: 0022-2402 (cit. on p. 45).

[98] S. Xianchen, W. Baoyun, and L. Kunxiang. "Dem and valley segmentation in remote sensing image based on region growing algorithm". In: *Journal of Physics: Conference Series* 1693 (Dec. 2020). DOI: `10.1088/1742-6596/1693/1/012174`. URL: `https://doi.org/10.1088/1742-6596/1693/1/012174` (cit. on p. 10).

[99] Y. Yang. *Temporal Data Clustering*. 2017. DOI: `10.1016/b978-0-12-811654-8.00003-8` (cit. on p. 62).

[100] A. Yasar, I. Saritas, and H. Korkmaz. "Determination of Intestinal Mass By Region Growing Method". In: (May 2018). ISSN: 2018050449. DOI: `10.20944/preprints201805.0449.v1` (cit. on p. 9).

[101] N. M. Zaitoun and M. J. Aqel. "Survey on Image Segmentation Techniques". In: *Procedia Computer Science* 65 (2015), pp. 797–806. ISSN: 18770509. DOI: `10.1016/j.procs.2015.09.027`. URL: `http://dx.doi.org/10.1016/j.procs.2015.09.027` (cit. on p. 6).

[102] F. R. Zakani et al. "Kulczynski similarity index for objective evaluation of mesh segmentation algorithms". In: vol. 0. IEEE Computer Society, Apr. 2017, pp. 12–17. ISBN: 9781509051465. DOI: `10.1109/ICMCS.2016.7905611` (cit. on p. 64).

[103] H. Zhang et al. "A novel adaptive fuzzy local information C-means clustering algorithm for remotely sensed imagery classification". In: *IEEE Transactions on Geoscience and Remote Sensing* 55.9 (2017), pp. 5057–5068. ISSN: 01962892. DOI: `10.1109/TGRS.2017.2702061` (cit. on p. 8).

[104] H. Zhang et al. "Fabric Defect Detection Using L0 Gradient Minimization and Fuzzy C-Means". In: *Applied Sciences* 9.17 (2019). ISSN: 2076-3417. DOI: `10.3390/app9173506`. URL: `https://www.mdpi.com/2076-3417/9/17/3506` (cit. on p. 9).

[105] T. Zhang et al. "Extraction of coastline in aquaculture coast from multispectral remote sensing images: Object-based region growing integrating edge detection". In: *Remote Sensing* 5.9 (2013), pp. 4470–4487. ISSN: 20724292. DOI: `10.3390/rs5094470` (cit. on p. 10).

A

Appendix 1

A.1 Instants S-STSEC segmentations' evaluations

Upwelling Instants	Evaluation 2007	Evaluation 2015	Evaluation 2019
1	4.5	2.0	3.5
2	4.5	2.0	4.0
3	4.5	2.0	4.0
4	4.0	2.5	3.5
5	4.0	3.0	2.0
6	3.5	2.5	3.5
7	4.0	3.0	4.0
8	2.5	3.5	4.0
9	3.5	3.5	4.5
10	2.5	3.5	4.5
11	3.5	4.0	4.5
12	4.0	4.0	4.5
13	4.0	3.5	5.0
14	5.0	3.5	5.0
15	4.5	3.5	4.0
16	4.5	4.5	4.5
17	4.0	4.0	4.5
18	4.0	4.5	4.0
19	3.5	3.5	4.0
20	4.0	4.0	4.5
21	4.5	3.5	4.0
22	3.5	4.0	4.5
23	4.0	3.5	4.5
Mean	3.93 ± 0.80	3.37 ± 0.97	4.13 ± 0.78

Table A.1: Evaluations of the S-STSEC segmentations regarding the fully preprocessed SST moving averages

Upwelling Instants	Evaluation 2007	Evaluation 2015	Evaluation 2019
1	4.0	2.0	3.0
2	4.0	3.0	4.0
3	5.0	3.0	4.0
4	5.0	3.0	4.0
5	5.0	4.0	4.0
6	5.0	4.0	5.0
7	5.0	4.0	5.0
8	5.0	5.0	5.0
9	5.0	5.0	5.0
10	5.0	5.0	5.0
11	4.0	5.0	5.0
12	5.0	5.0	5.0
13	5.0	4.0	5.0
14	5.0	5.0	4.0
15	5.0	4.0	5.0
16	4.0	5.0	5.0
17	4.0	5.0	4.0
18	3.0	5.0	4.0
19	3.0	5.0	5.0
20	3.0	5.0	5.0
21	4.0	4.0	5.0
22	4.0	4.0	4.0
23	4.0	4.0	4.0
Mean	4.39 ± 0.72	4.26 ± 0.86	4.52 ± 0.59

Table A.2: Evaluations of the Core's and Shell's delineation on the 5 original SST images that compose each SST average map

A.2 Time Series

Figure A.1: Mean temperatures for the segmented upwelling regions, Cores, Shells and Core-Shell clusters of the year 2007. Upwelling span 1 contains instants 1 to 7, upwelling span 2 contains instants 8 to 12, upwelling span 3 contains instants 13 to 20 and upwelling span 4 contains instants 21 to 23

Figure A.2: Mean temperatures for the segmented upwelling regions, Cores, Shells and Core-Shell clusters of the year 2007 on the region north of Cabo da Roca. Upwelling span 1 contains instants 1 to 7, upwelling span 2 contains instants 8 to 12, upwelling span 3 contains instants 13 to 20 and upwelling span 4 contains instants 21 to 23

Figure A.3: Mean temperatures for the segmented upwelling regions, Cores, Shells and Core-Shell clusters of the year 2007 on the region south of Cabo da Roca. Upwelling span 1 contains instants 1 to 7, upwelling span 2 contains instants 8 to 12, upwelling span 3 contains instants 13 to 20 and upwelling span 4 contains instants 21 to 23

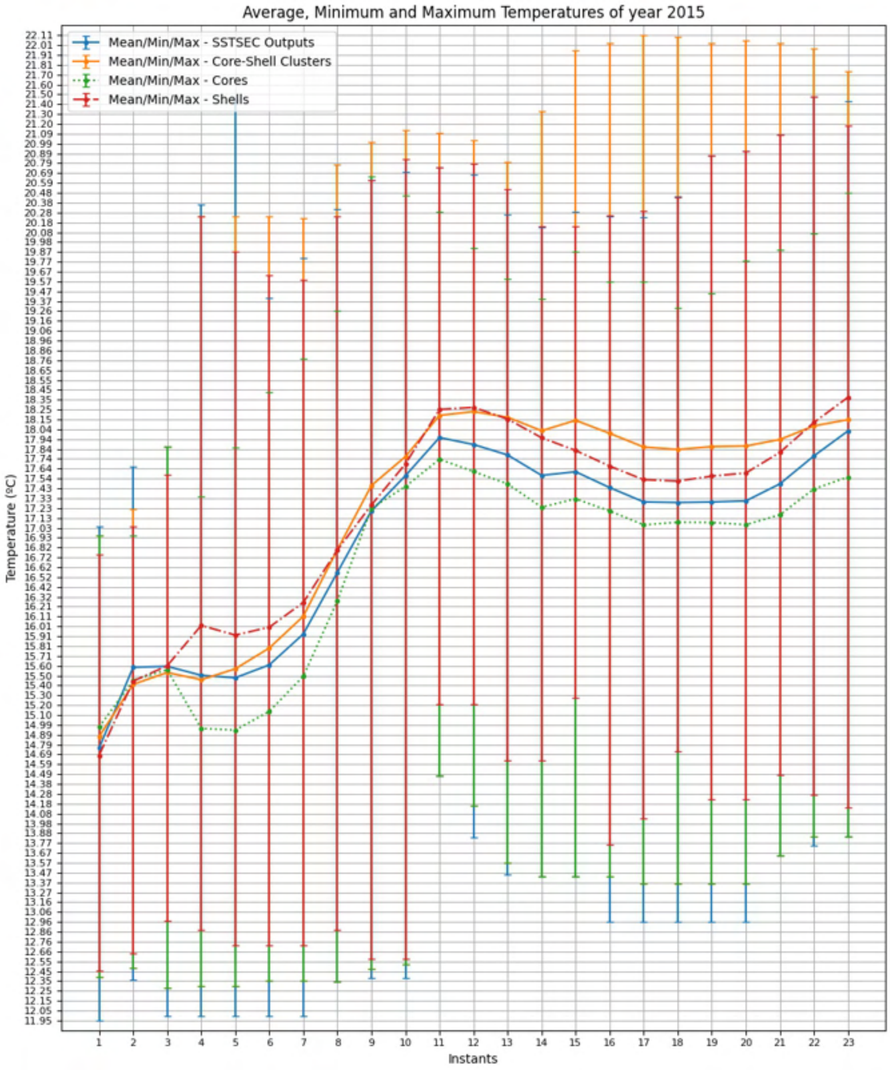


Figure A.4: Mean temperatures for the segmented upwelling regions, Cores, Shells and Core-Shell clusters of the year 2015. Upwelling span 1 contains instants 1 to 3, upwelling span 2 contains instants 4 to 8, upwelling span 3 contains instants 9 to 14 and upwelling span 4 contains instants 15 to 23

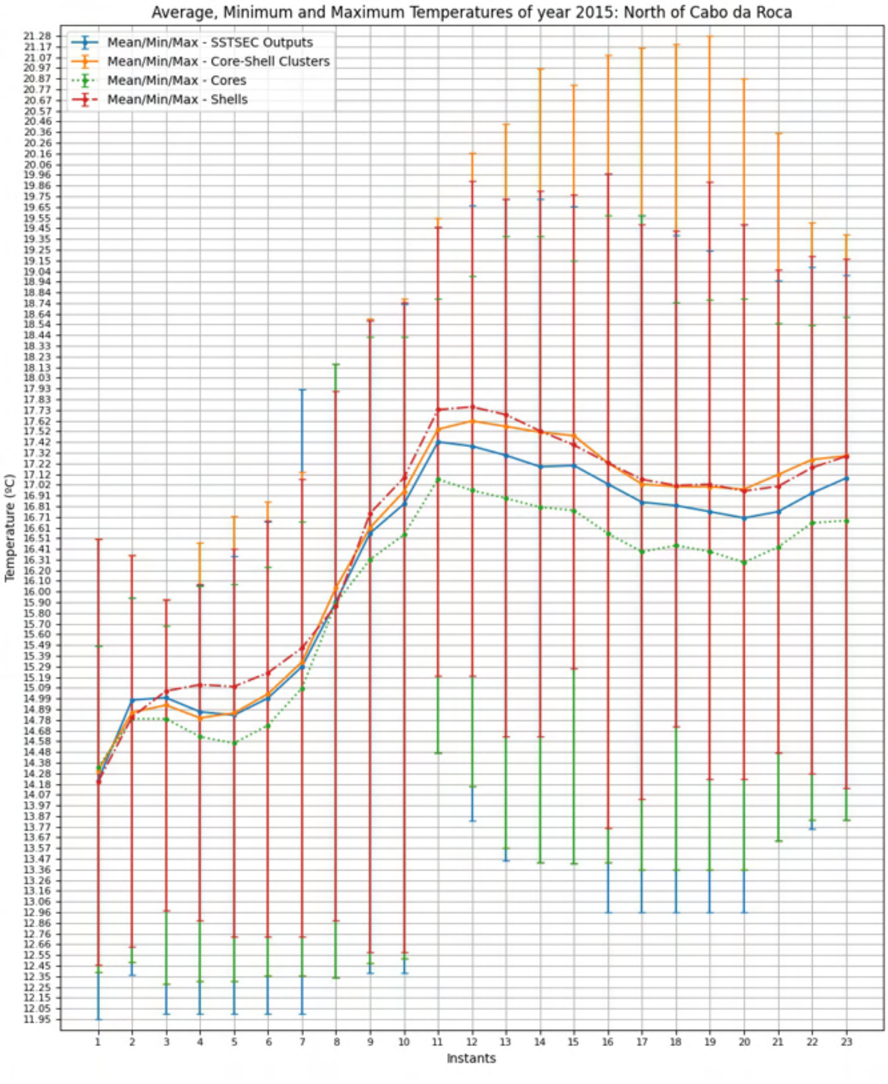


Figure A.5: Mean temperatures for the segmented upwelling regions, Cores, Shells and Core-Shell clusters of the year 2015 on the region north of Cabo da Roca. Upwelling span 1 contains instants 1 to 3, upwelling span 2 contains instants 4 to 8, upwelling span 3 contains instants 9 to 14 and upwelling span 4 contains instants 15 to 23

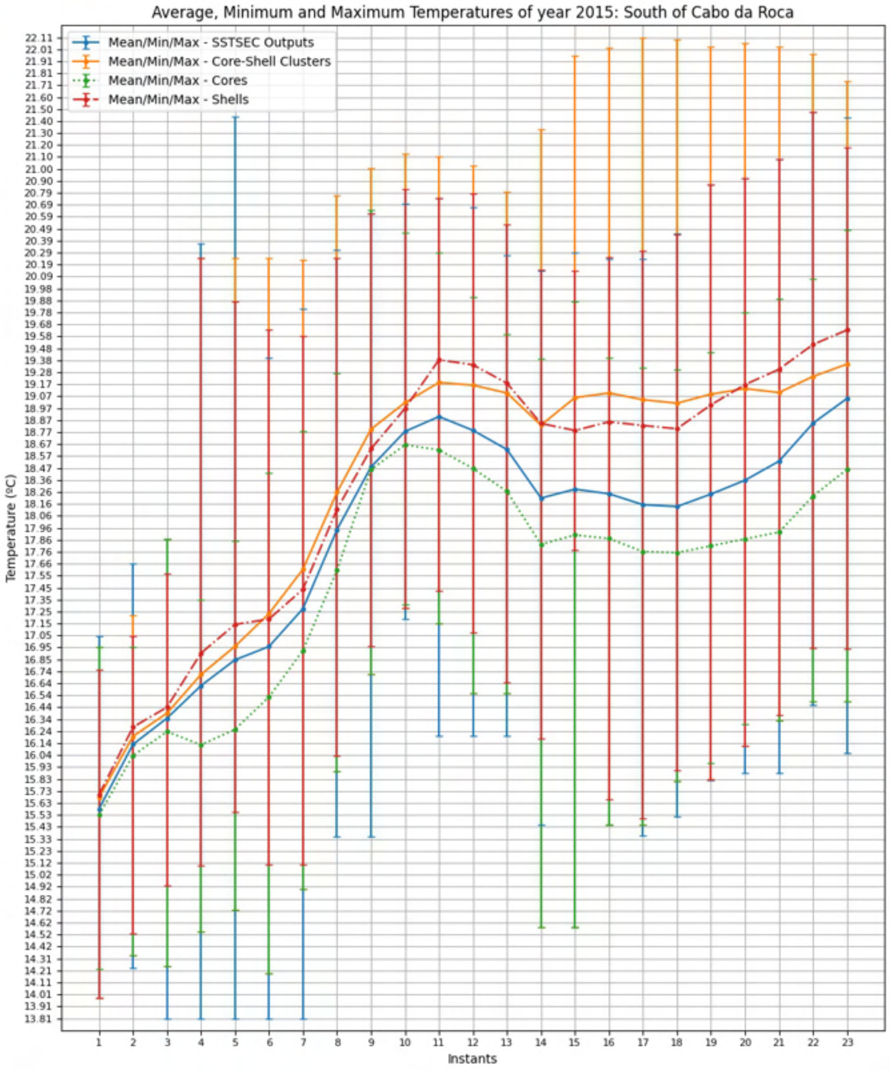


Figure A.6: Mean temperatures for the segmented upwelling regions, Cores, Shells and Core-Shell clusters of the year 2015 on the region south of Cabo da Roca. Upwelling span 1 contains instants 1 to 3, upwelling span 2 contains instants 4 to 8, upwelling span 3 contains instants 9 to 14 and upwelling span 4 contains instants 15 to 23

Figure A.7: Mean temperatures for the segmented upwelling regions, Cores, Shells and Core-Shell clusters of the year 2019. Upwelling span 1 contains instants 1 to 3, upwelling span 2 contains instants 4 to 7, upwelling span 3 contains instants 8 to 13 and upwelling span 4 contains instants 14 to 23

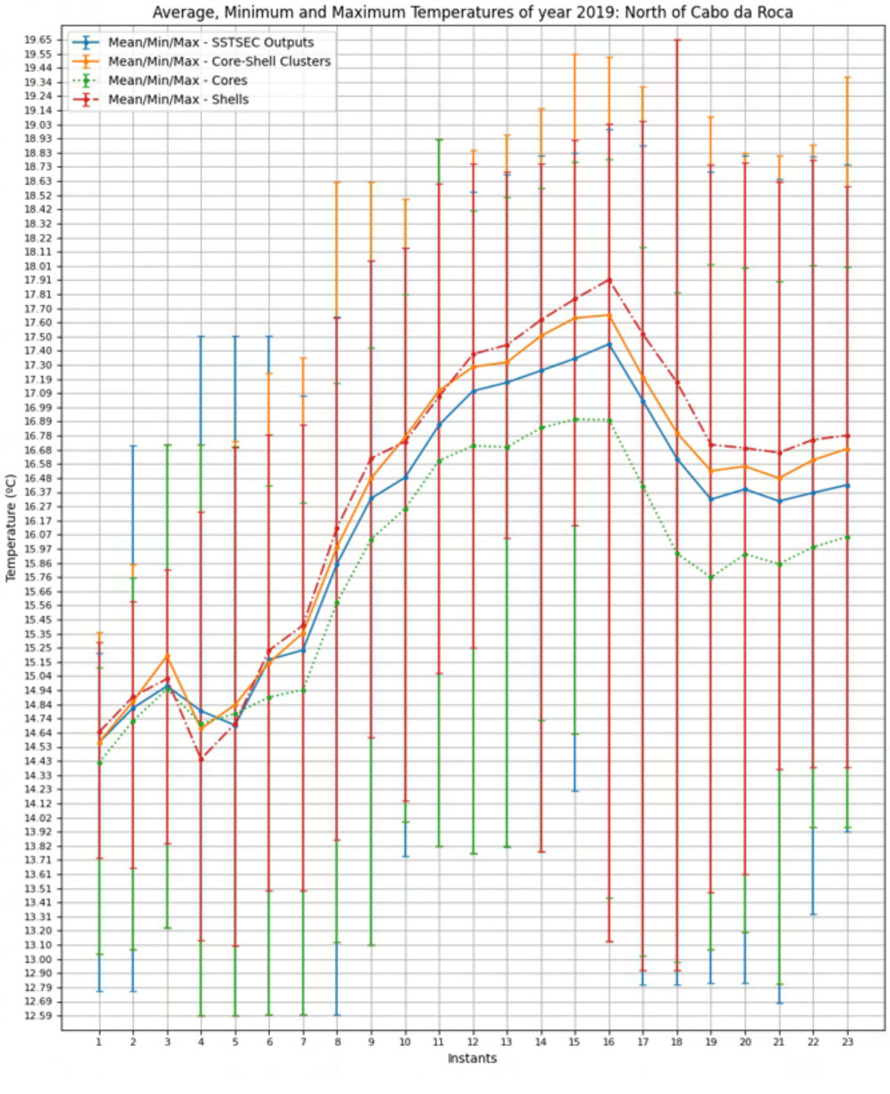


Figure A.8: Mean temperatures for the segmented upwelling regions, Cores, Shells and Core-Shell clusters of the year 2019 on the region north of Cabo da Roca. Upwelling span 1 contains instants 1 to 3, upwelling span 2 contains instants 4 to 7, upwelling span 3 contains instants 8 to 13 and upwelling span 4 contains instants 14 to 23

A.3 Core-Shell Clustering Algorithm Results

A.3.1 Core-Shell clusters results

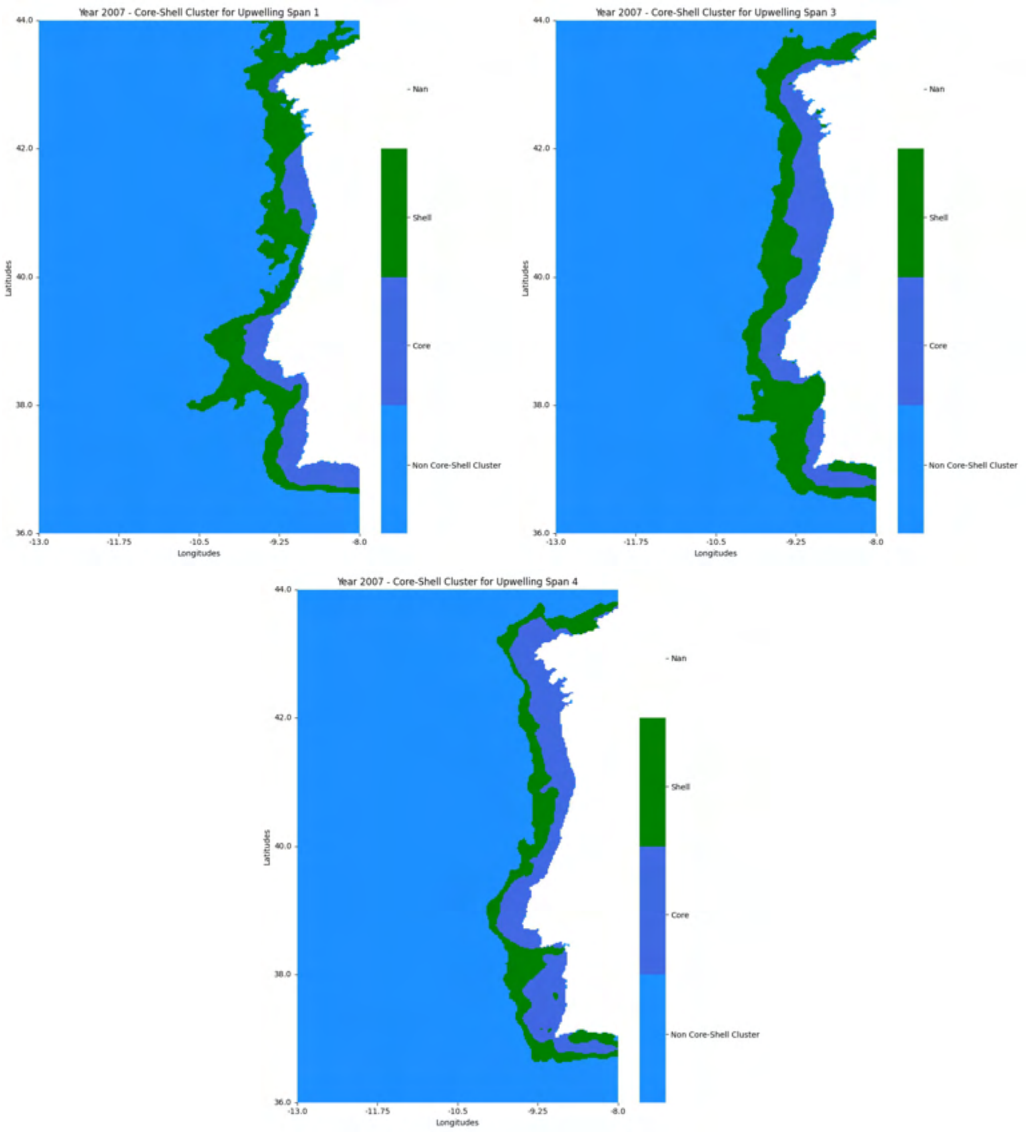


Figure A.9: First, third and forth Core-Shell clusters for the year 2007

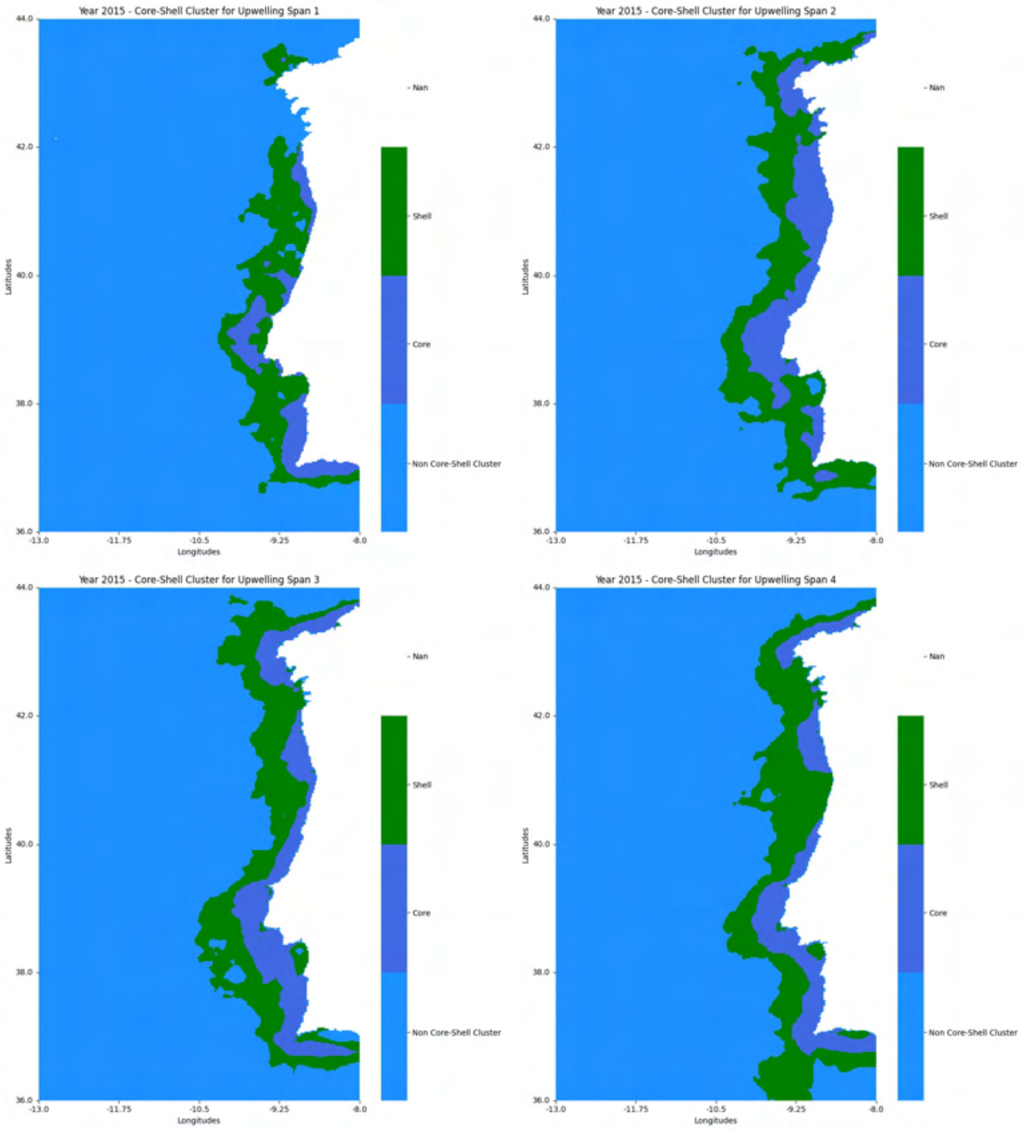


Figure A.10: Core-Shell clusters for the year 2015

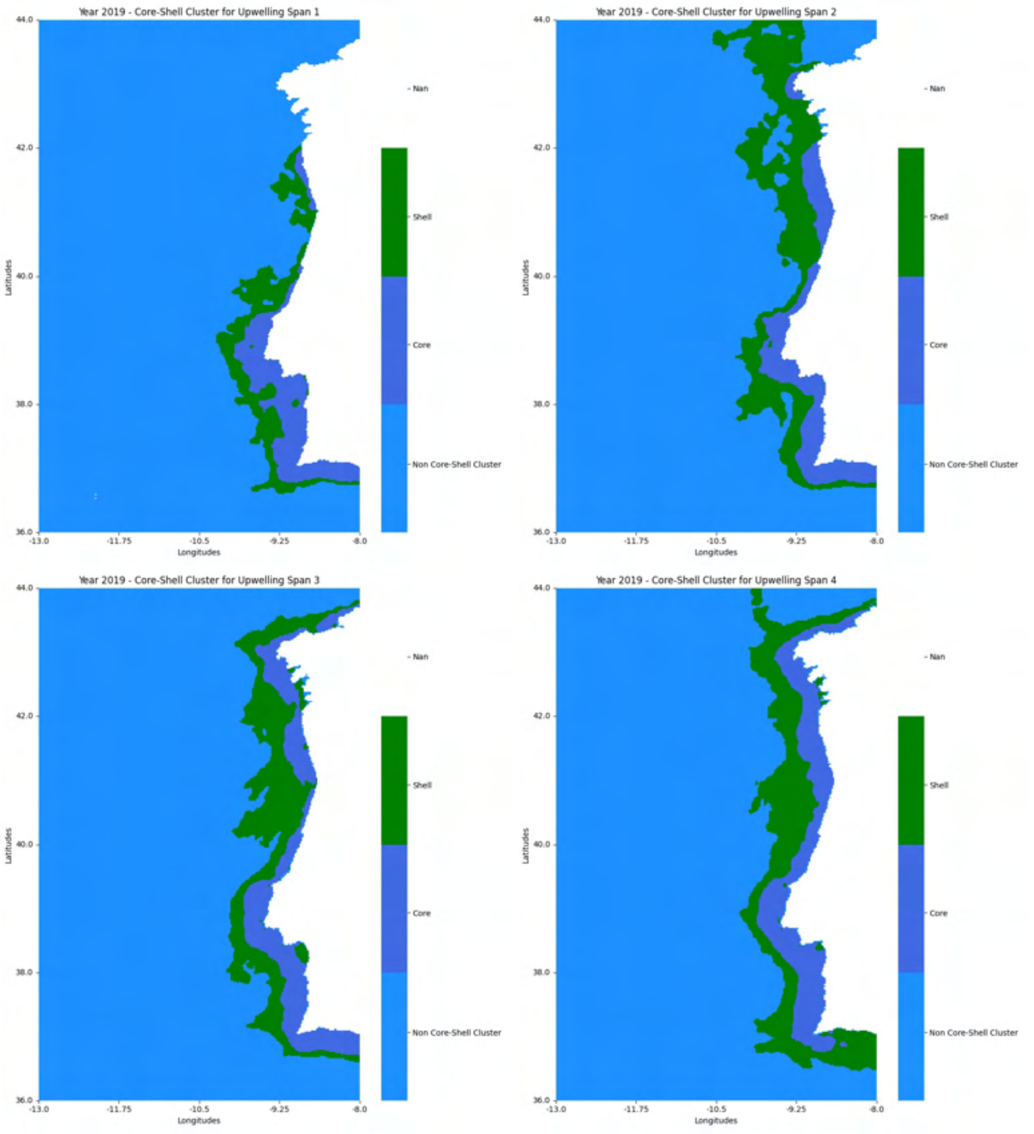


Figure A.11: Core-Shell clusters for the year 2019

A.3.2 Convergence of the clustering criterion G

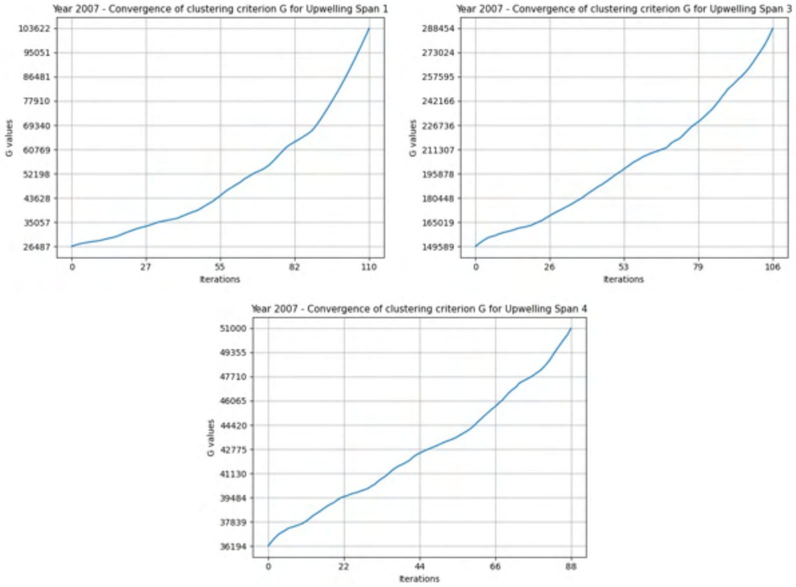


Figure A.12: Convergence of the clustering criterion G when building the first, third and fourth Core-Shell clusters for the year 2007

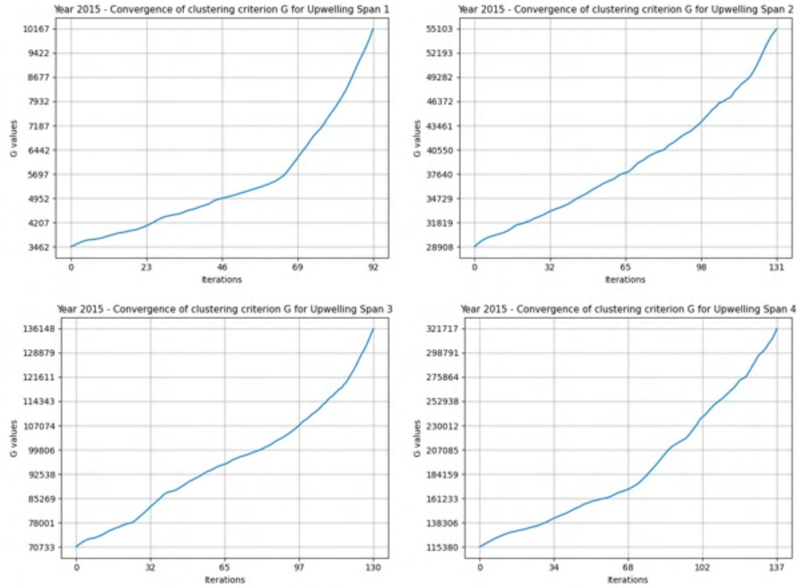


Figure A.13: Convergence of the clustering criterion G when building the four Core-Shell clusters for the year 2015

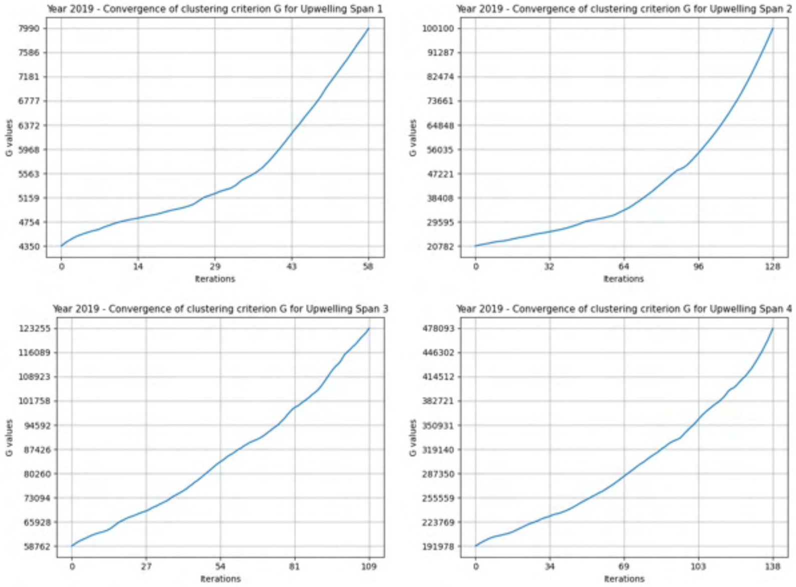


Figure A.14: Convergence of the clustering criterion G when building the four Core-Shell clusters for the year 2019

I

Annex 1

I.1 Instants S-STSEC segmentations' results

I.1.1 Year 2007

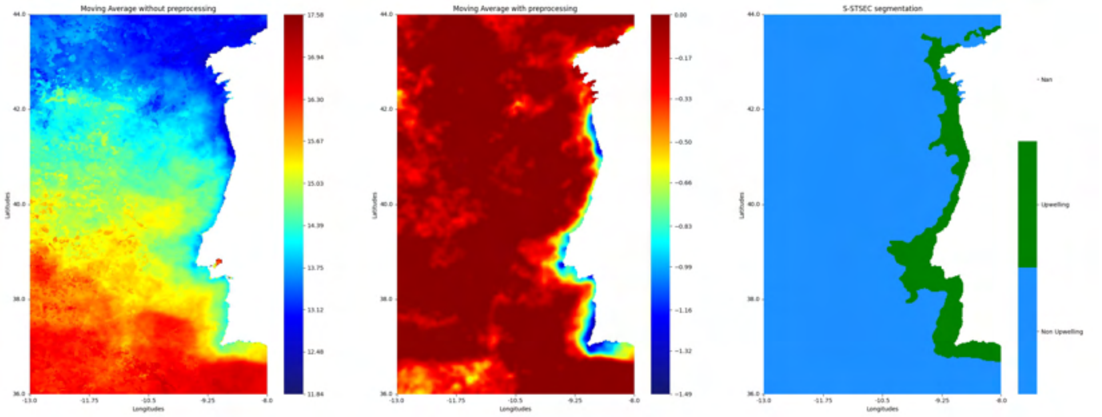


Figure I.1: Segmentation of instant 1 from 2007

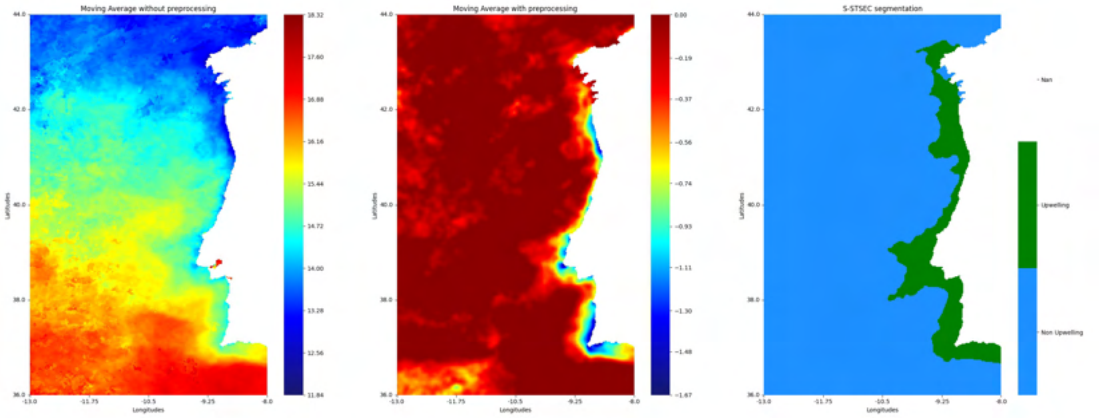


Figure I.2: Segmentation of instant 2 from 2007

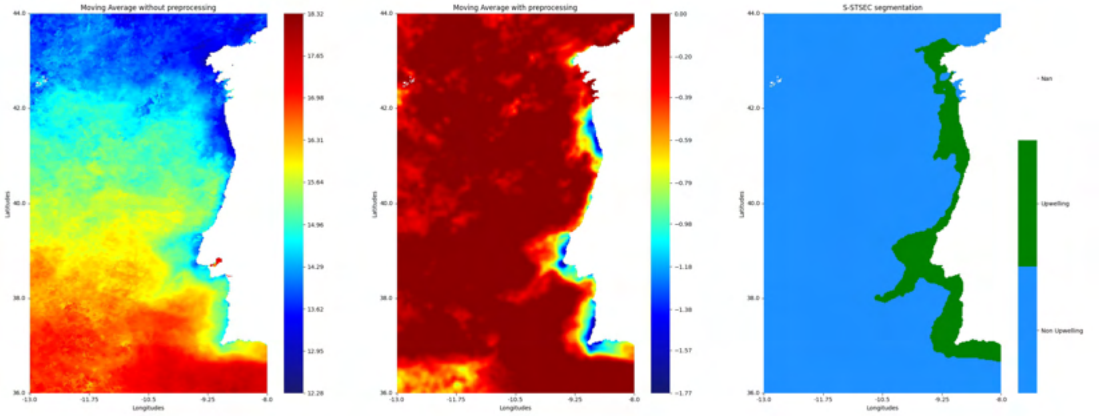


Figure I.3: Segmentation of instant 3 from 2007

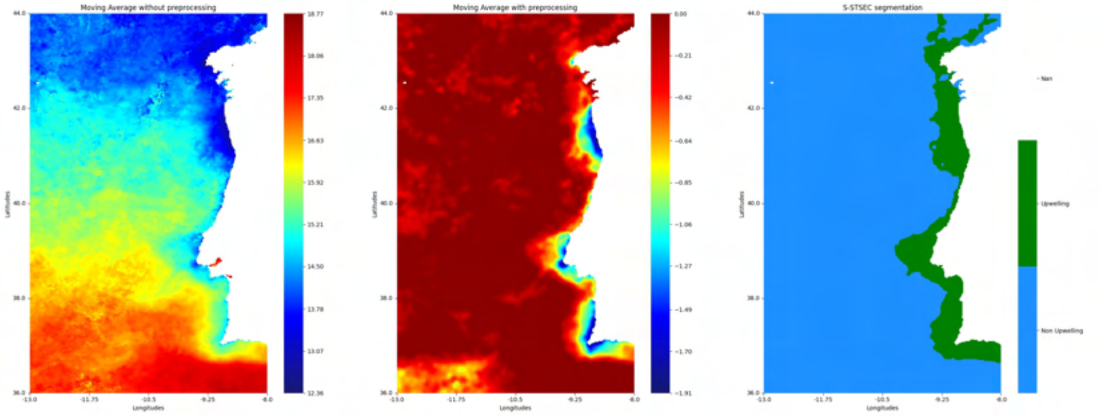


Figure I.4: Segmentation of instant 4 from 2007

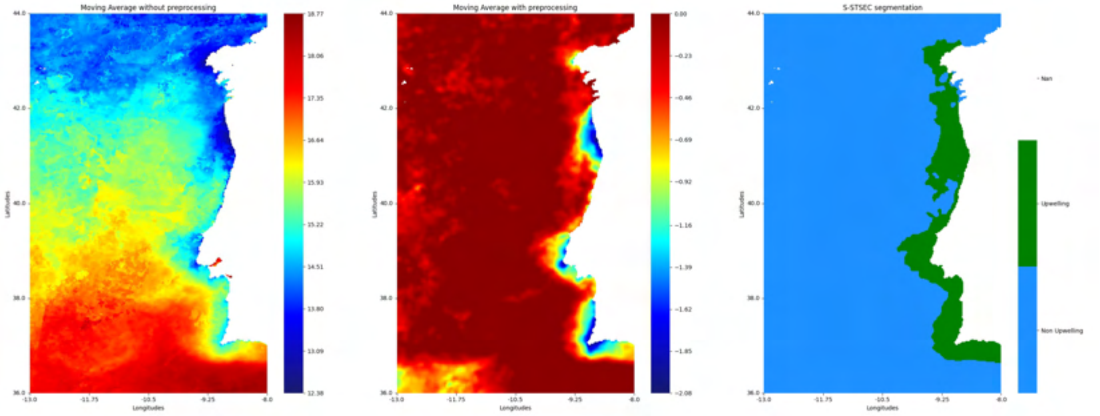


Figure I.5: Segmentation of instant 5 from 2007

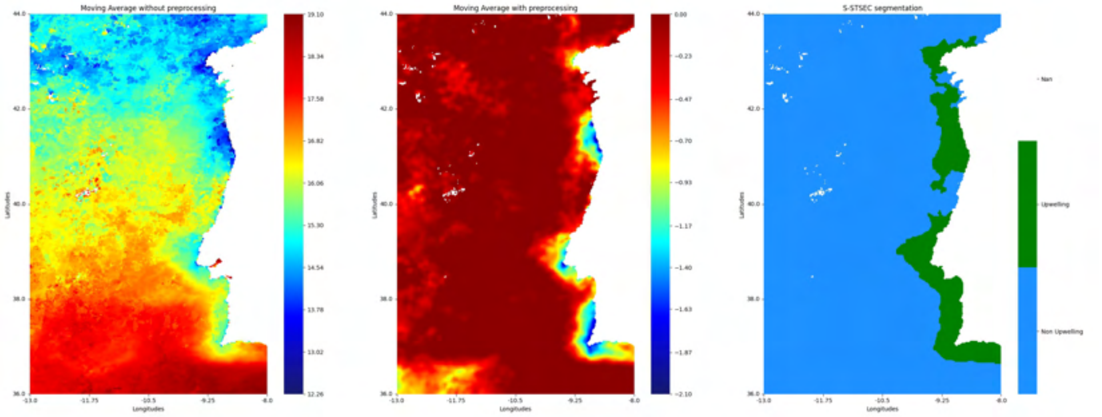


Figure I.6: Segmentation of instant 6 from 2007

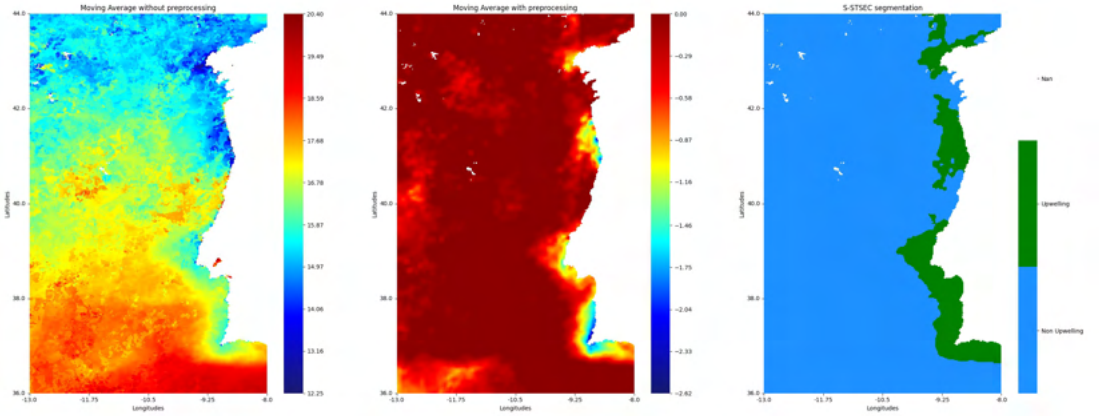


Figure I.7: Segmentation of instant 7 from 2007

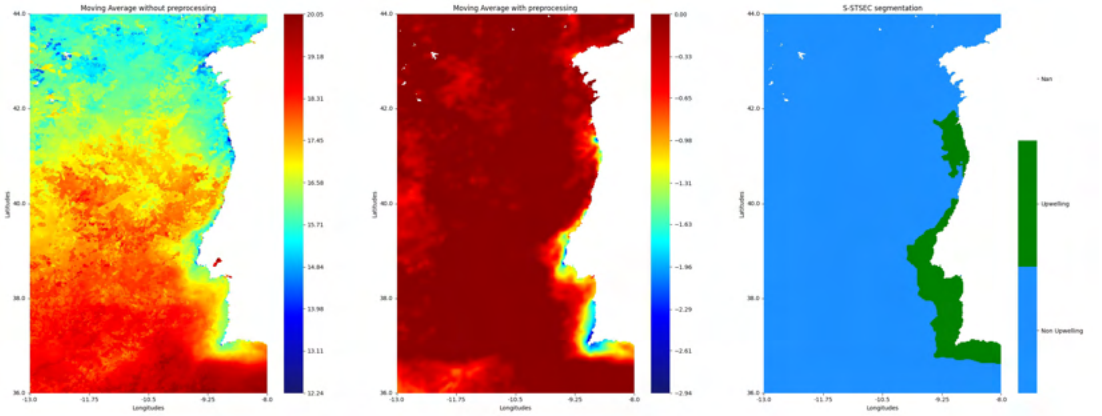


Figure I.8: Segmentation of instant 8 from 2007

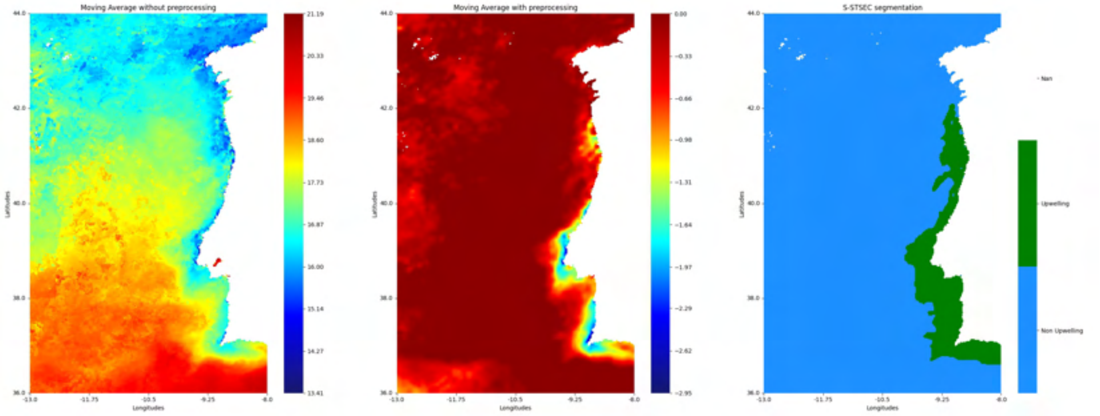


Figure I.9: Segmentation of instant 9 from 2007

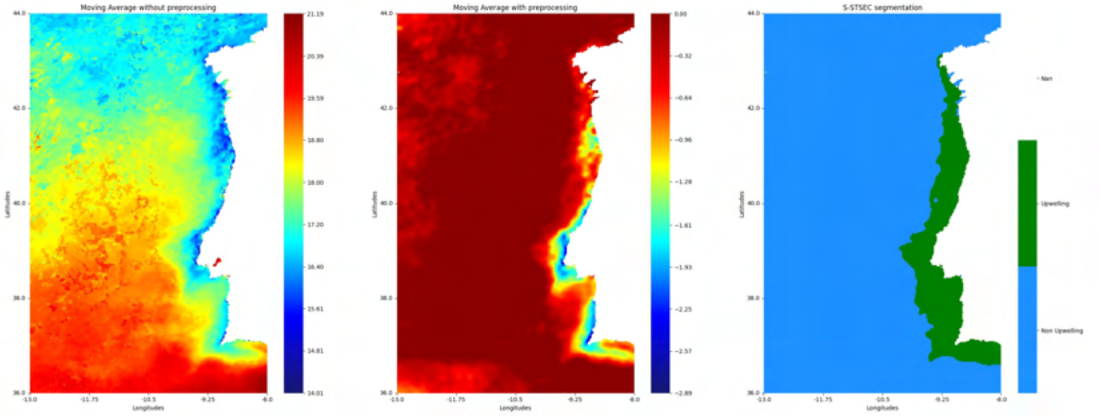


Figure I.10: Segmentation of instant 10 from 2007

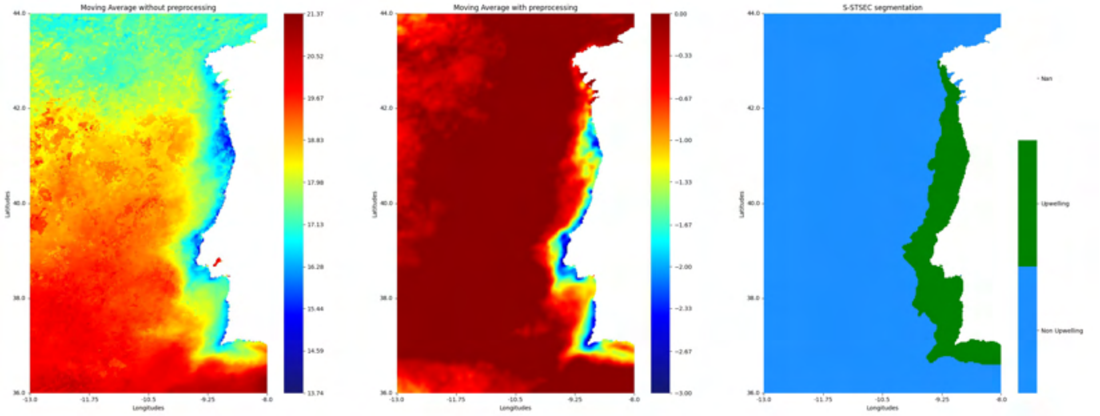


Figure I.11: Segmentation of instant 11 from 2007

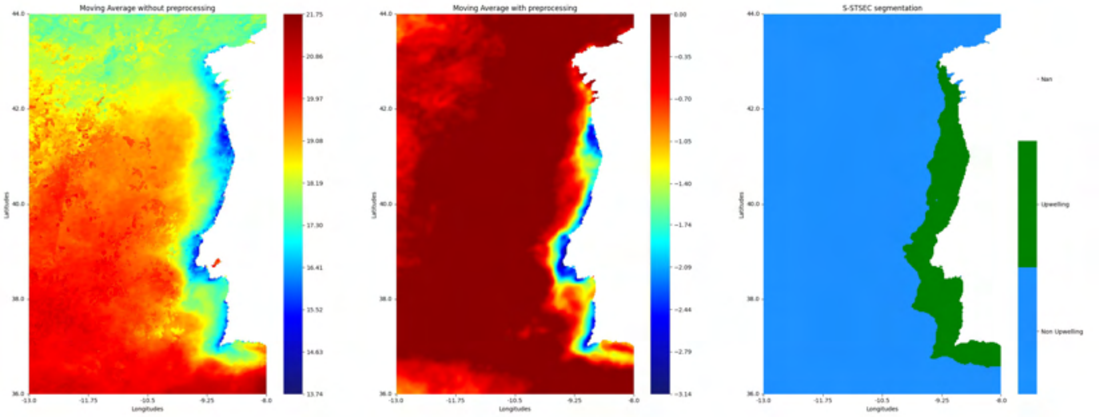


Figure I.12: Segmentation of instant 12 from 2007

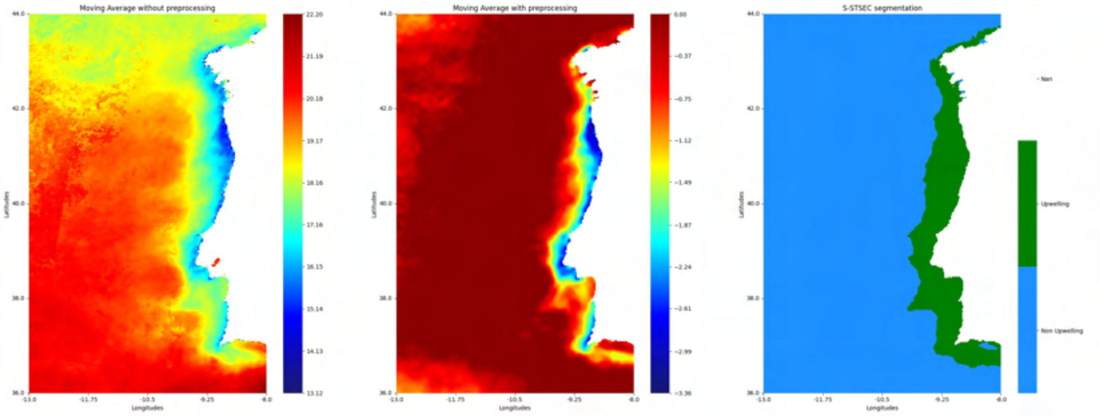


Figure I.13: Segmentation of instant 13 from 2007

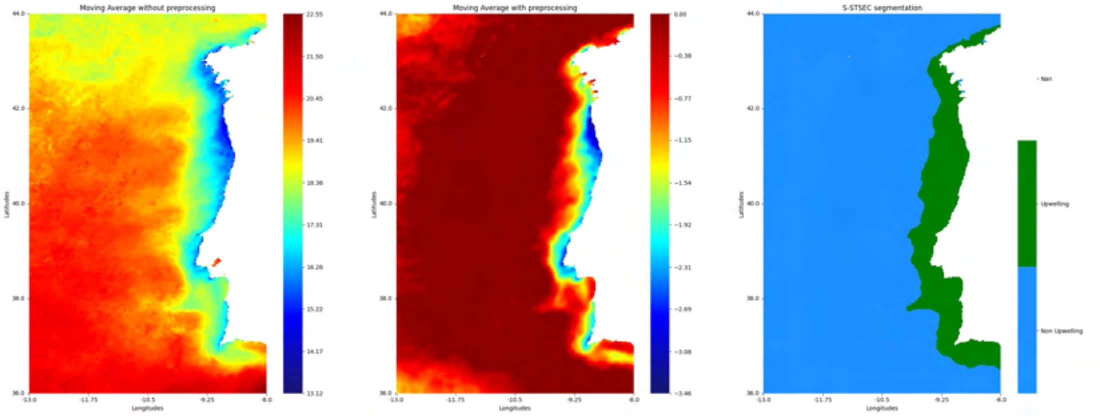


Figure I.14: Segmentation of instant 14 from 2007

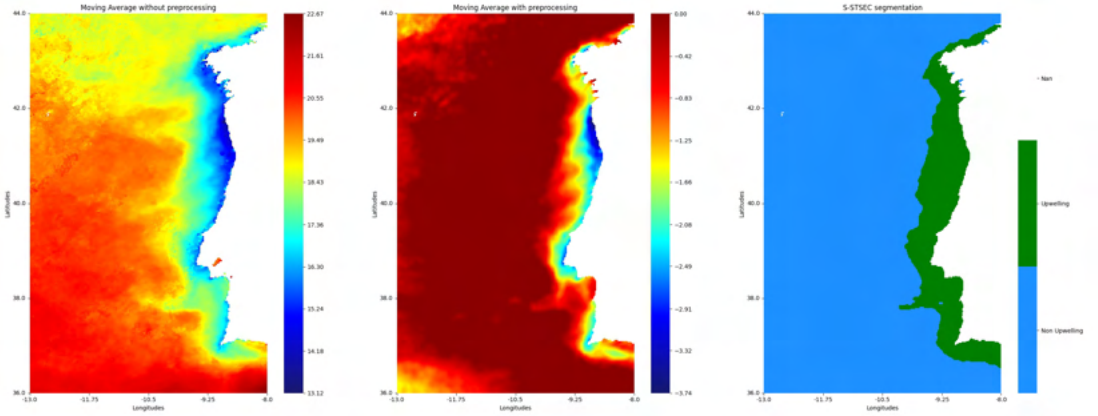


Figure I.15: Segmentation of instant 15 from 2007

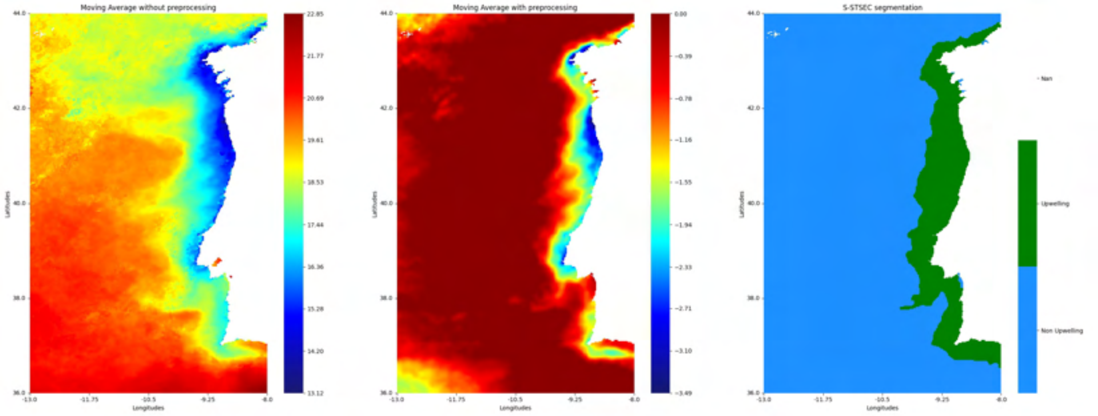


Figure I.16: Segmentation of instant 16 from 2007

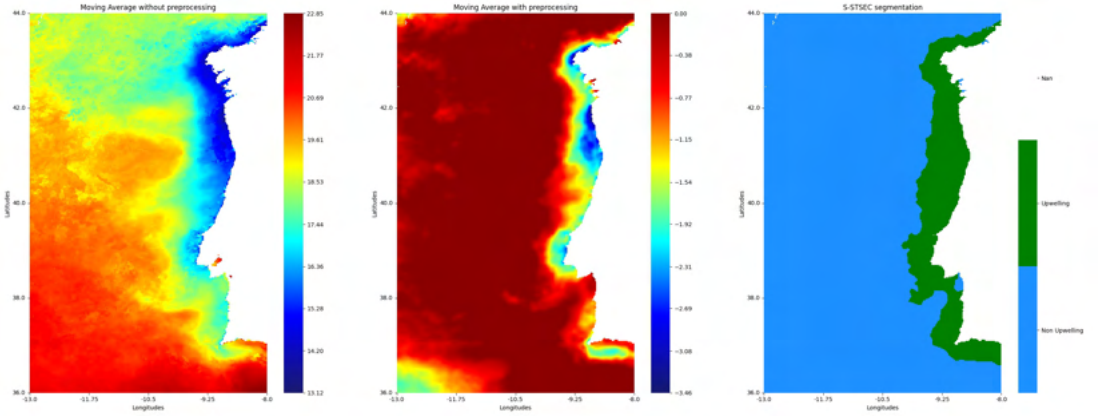


Figure I.17: Segmentation of instant 17 from 2007

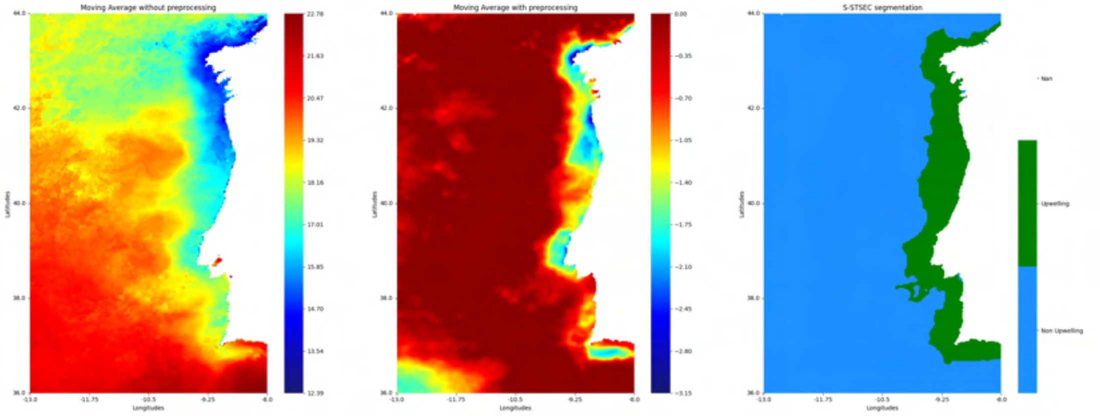


Figure I.18: Segmentation of instant 18 from 2007

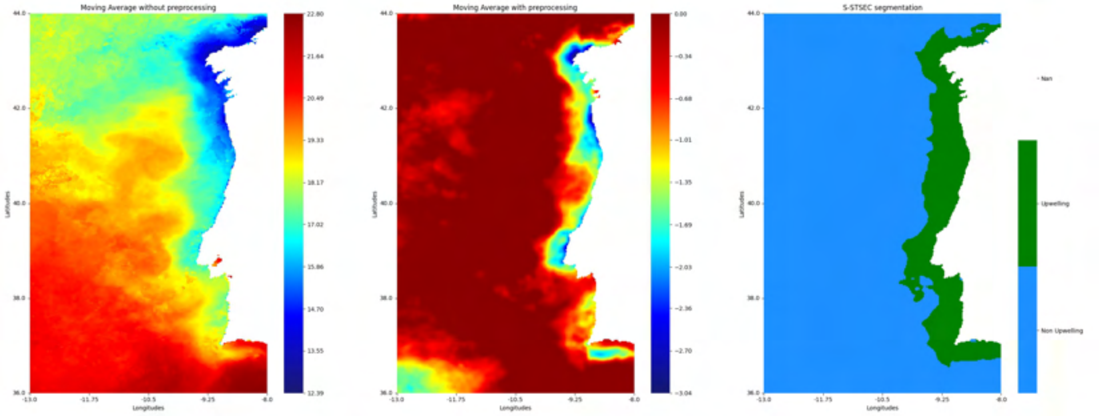


Figure I.19: Segmentation of instant 19 from 2007

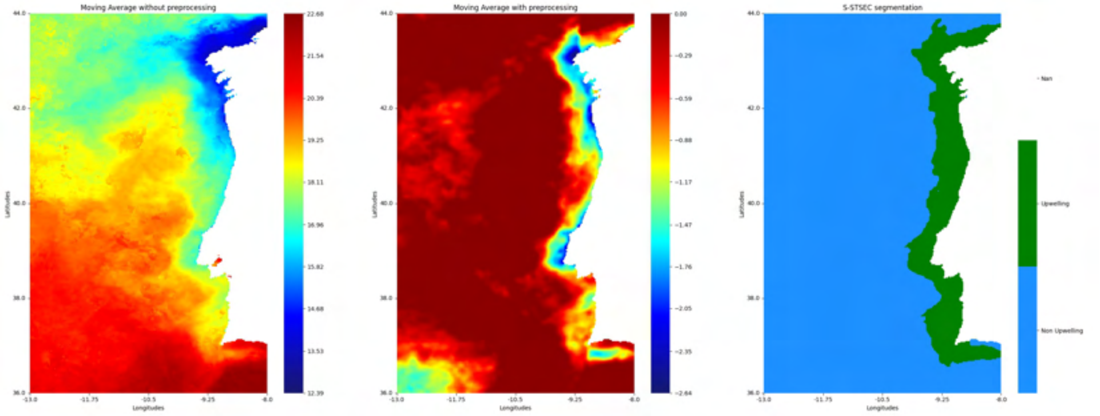


Figure I.20: Segmentation of instant 20 from 2007

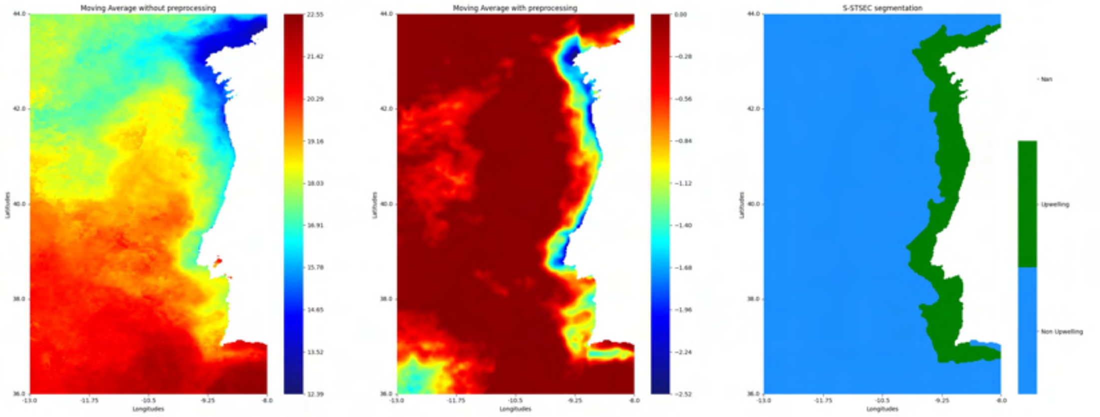


Figure I.21: Segmentation of instant 21 from 2007

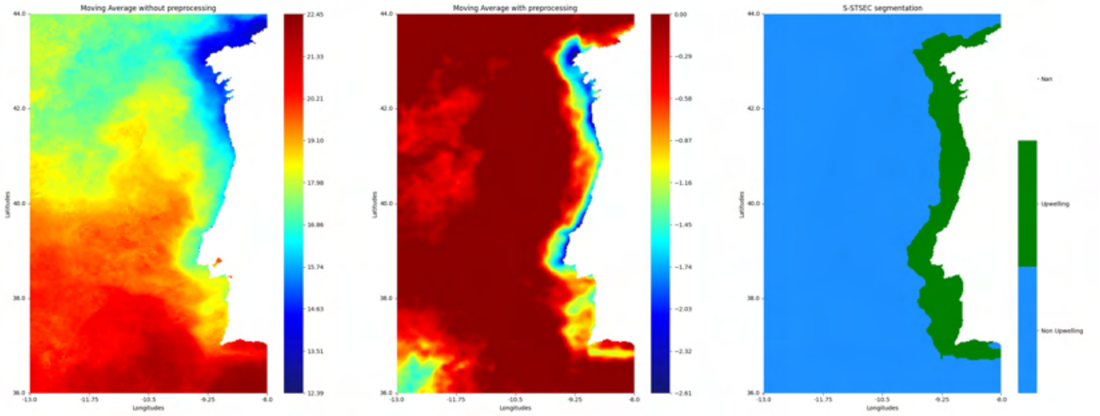


Figure I.22: Segmentation of instant 22 from 2007

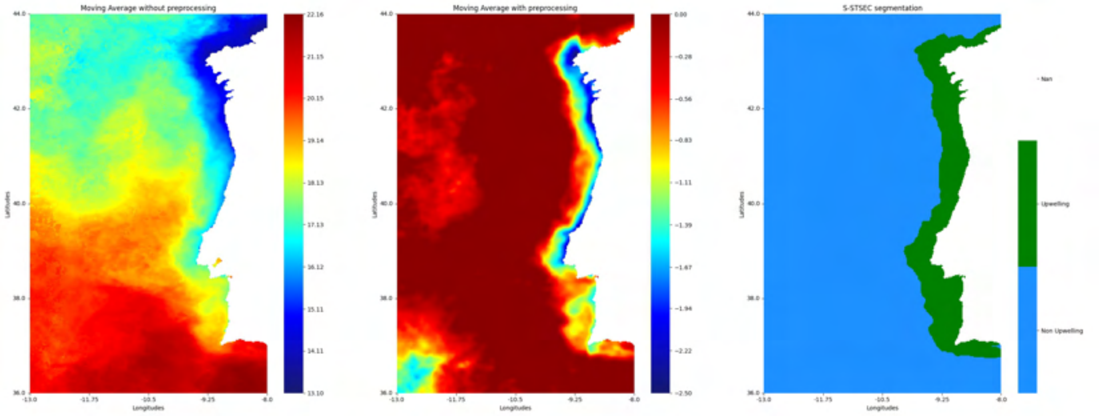


Figure I.23: Segmentation of instant 23 from 2007

I.1.2 Year 2015

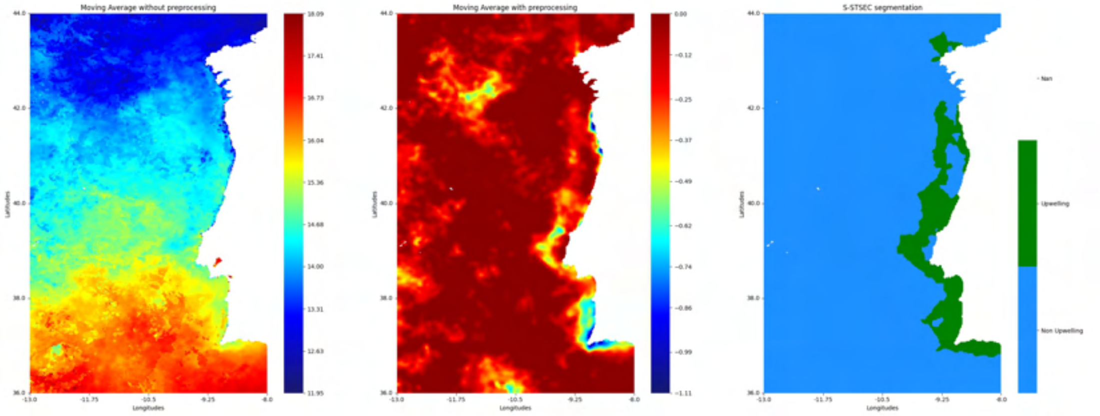


Figure I.24: Segmentation of instant 1 from 2015

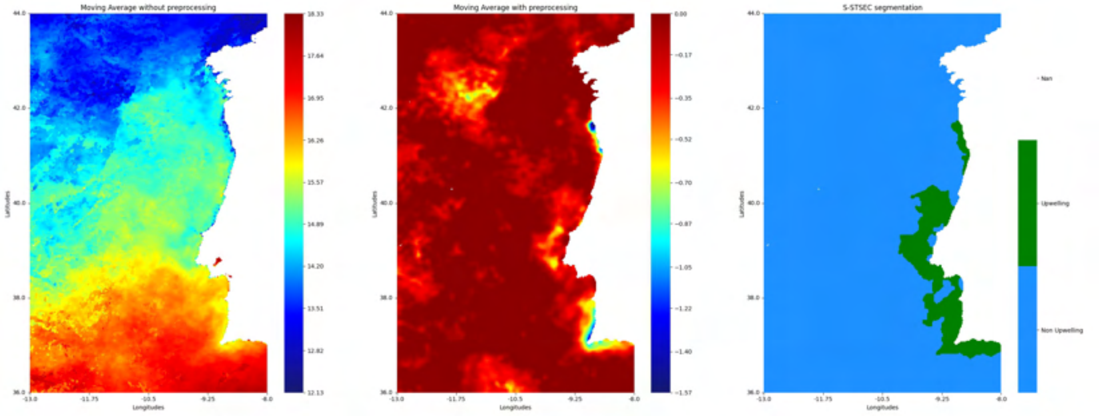


Figure I.25: Segmentation of instant 2 from 2015

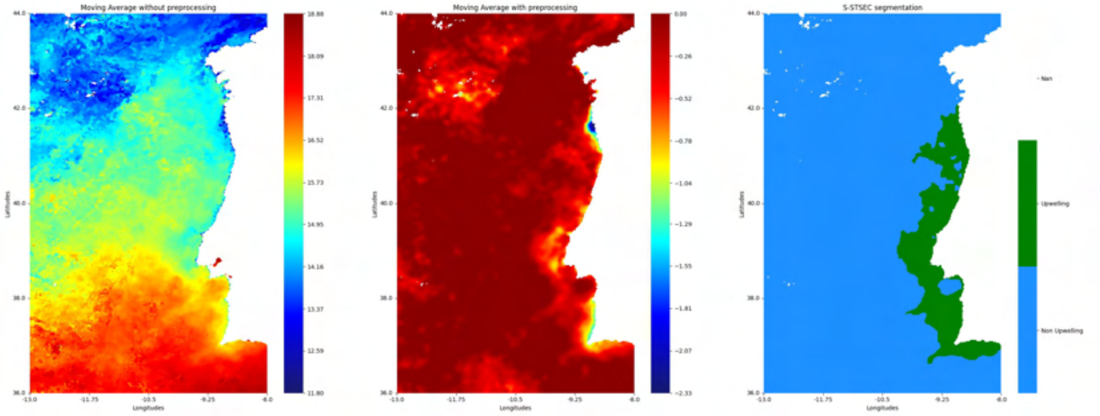


Figure I.26: Segmentation of instant 3 from 2015

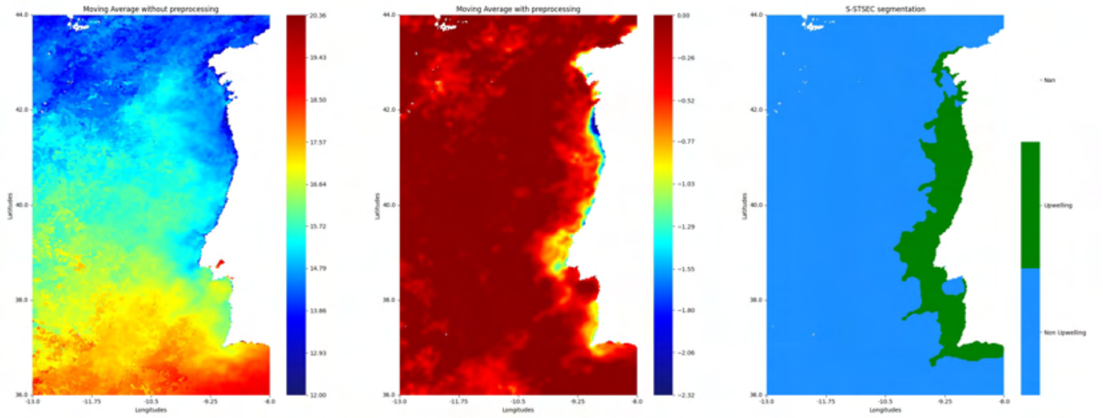


Figure I.27: Segmentation of instant 4 from 2015

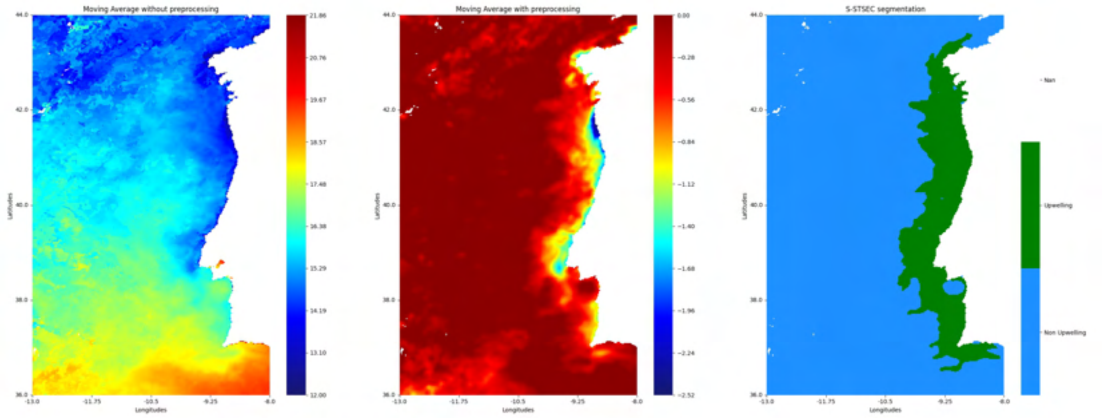


Figure I.28: Segmentation of instant 5 from 2015

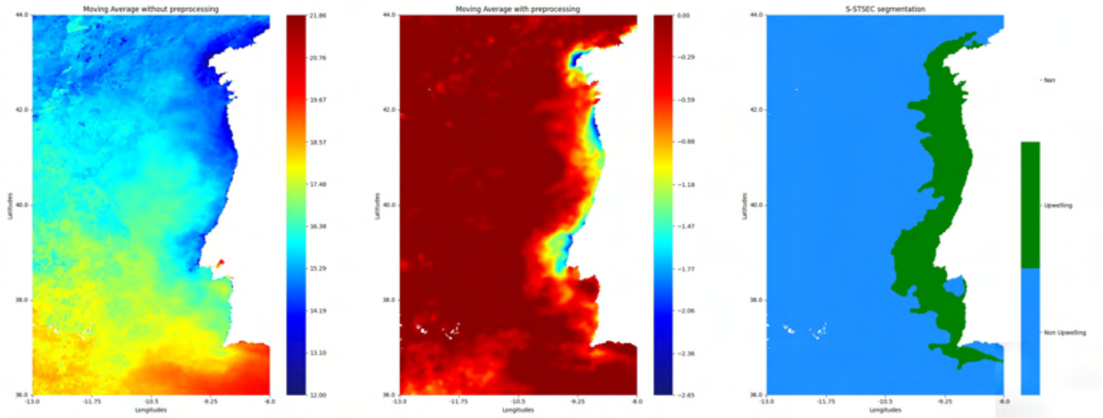


Figure I.29: Segmentation of instant 6 from 2015

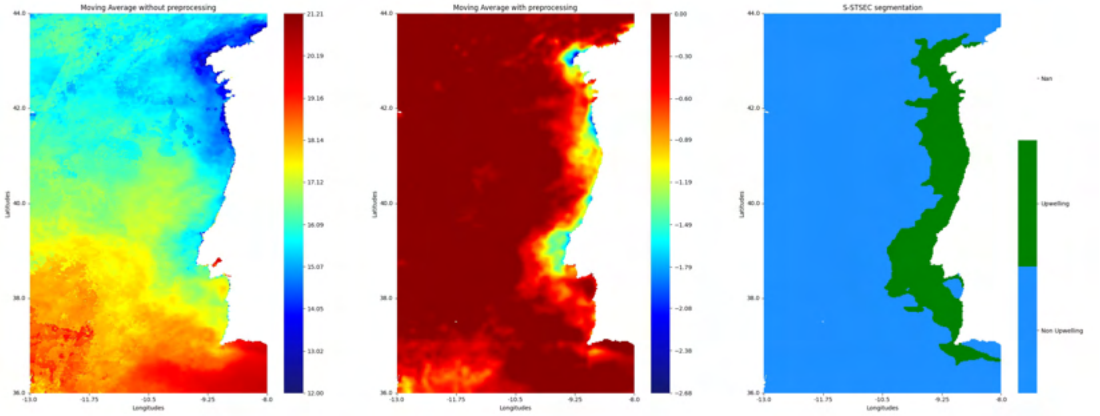


Figure I.30: Segmentation of instant 7 from 2015

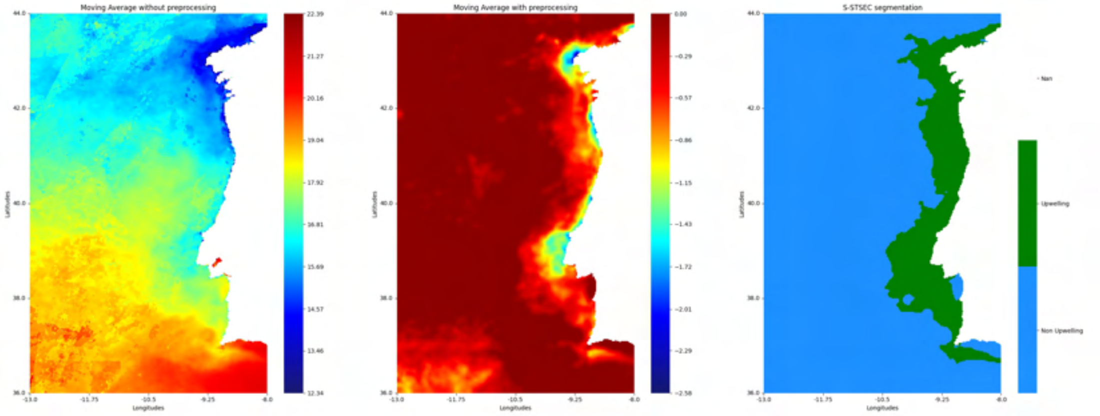


Figure I.31: Segmentation of instant 8 from 2015

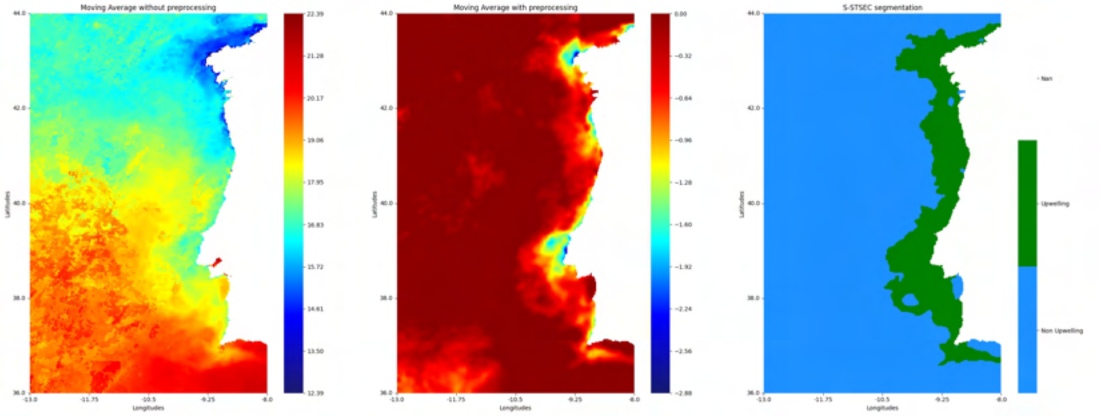


Figure I.32: Segmentation of instant 9 from 2015

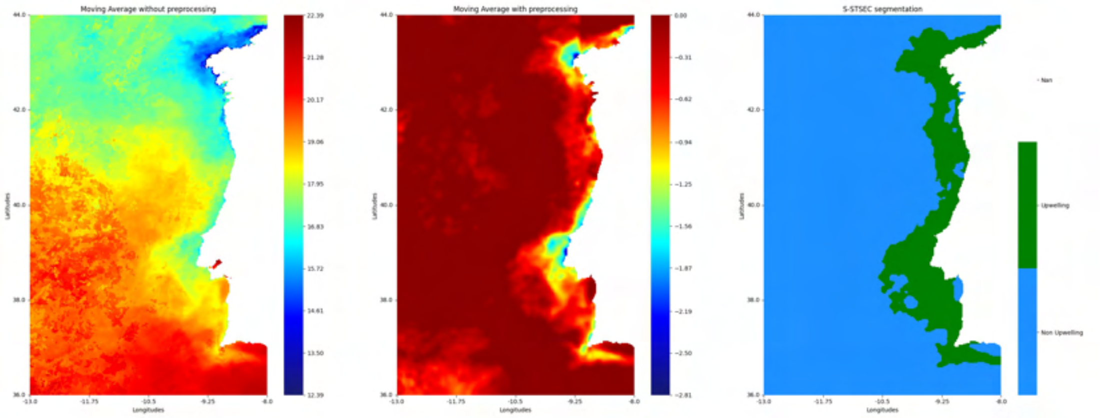


Figure I.33: Segmentation of instant 10 from 2015

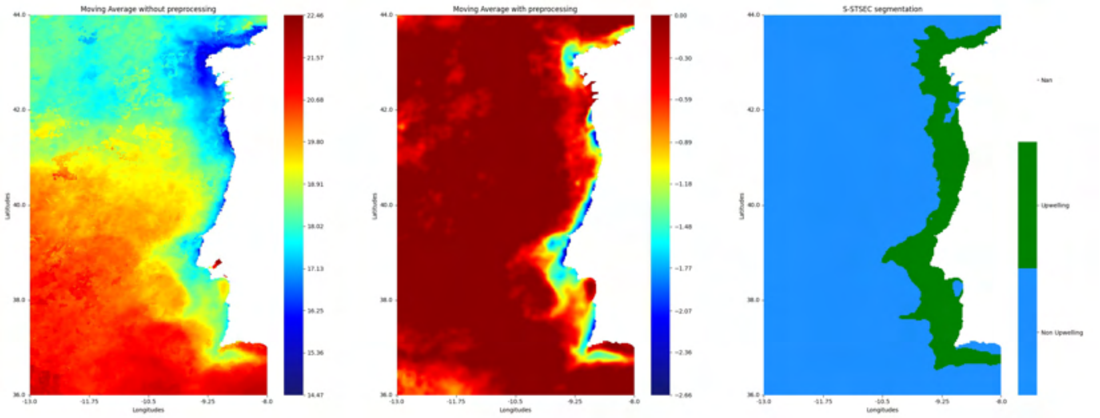


Figure I.34: Segmentation of instant 11 from 2015

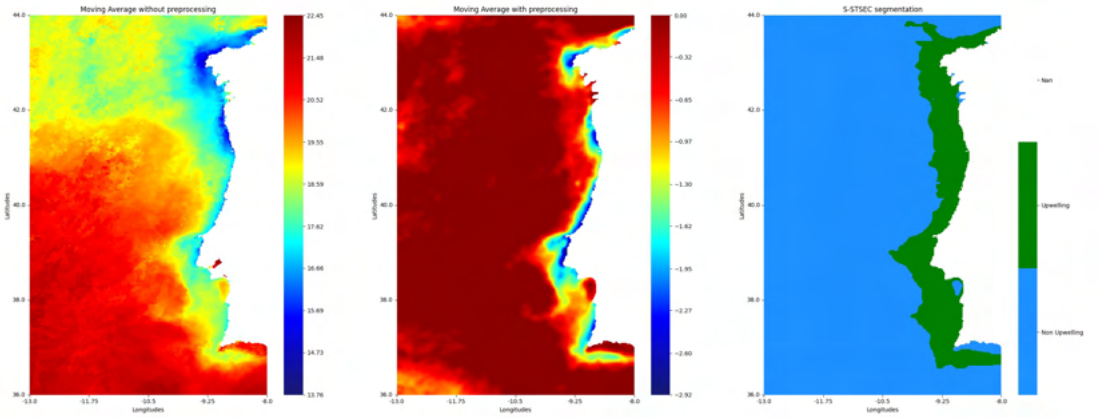


Figure I.35: Segmentation of instant 12 from 2015

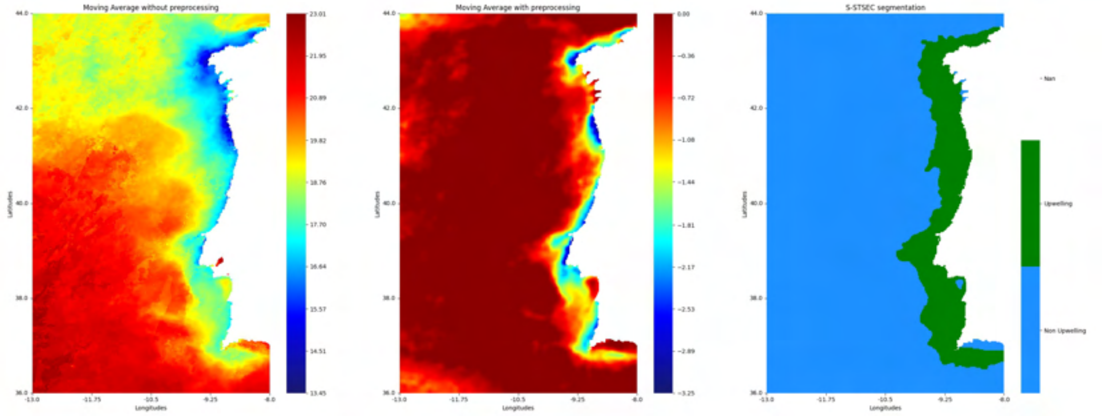


Figure I.36: Segmentation of instant 13 from 2015

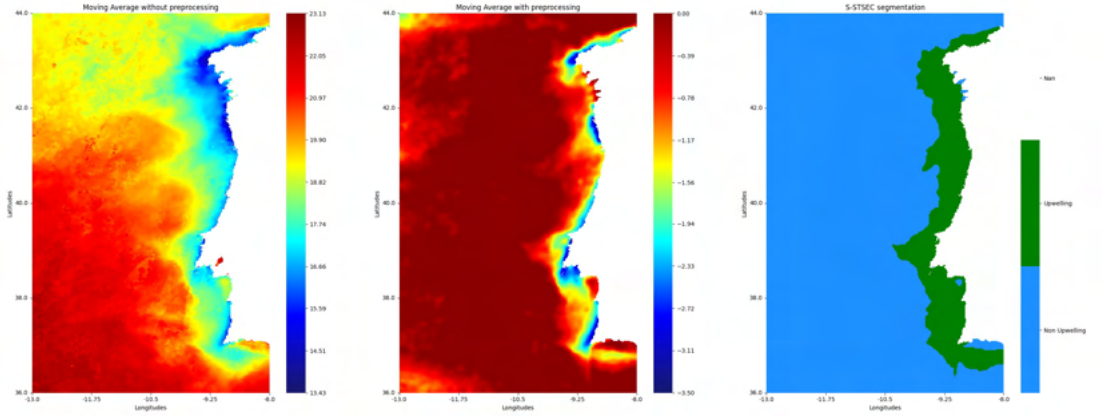


Figure I.37: Segmentation of instant 14 from 2015

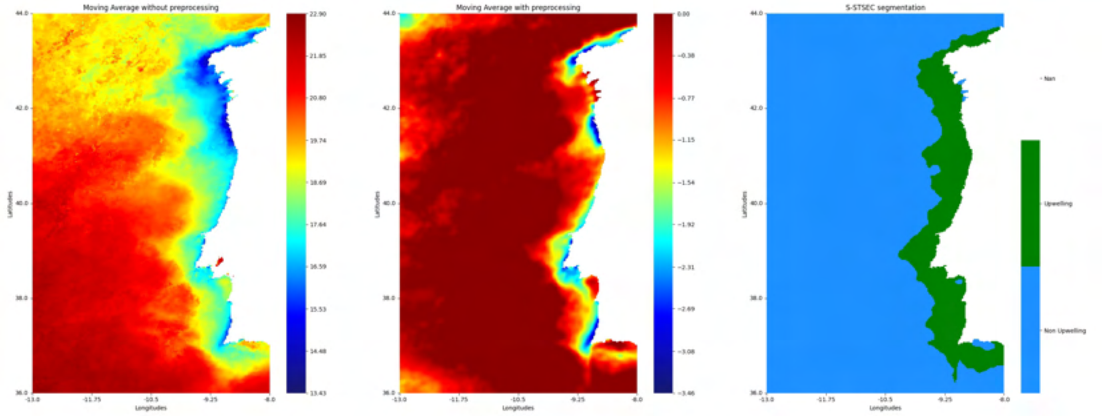


Figure I.38: Segmentation of instant 15 from 2015

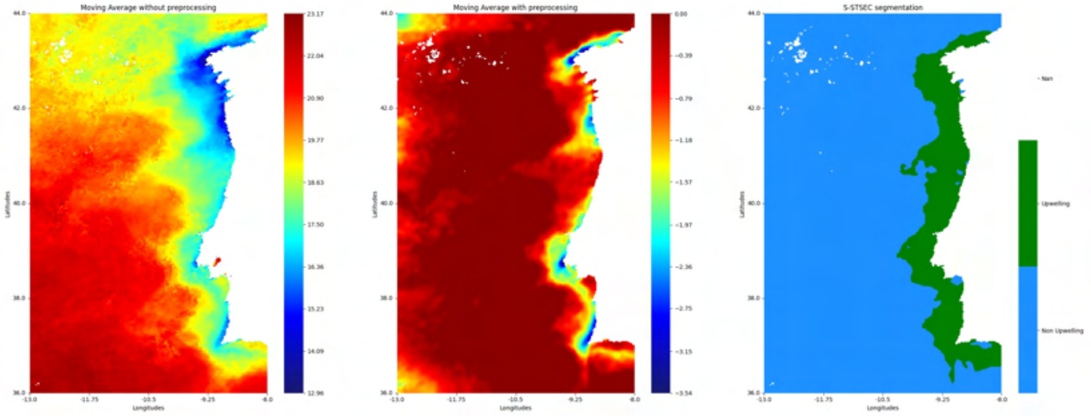


Figure I.39: Segmentation of instant 16 from 2015

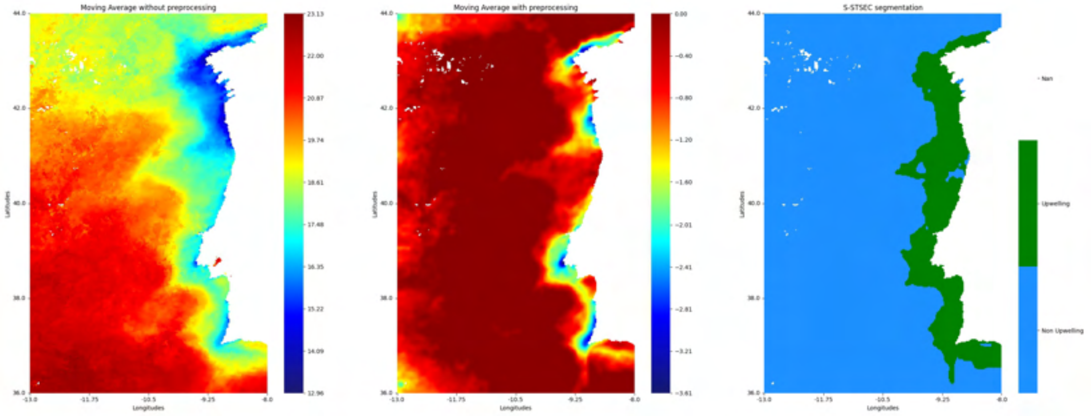


Figure I.40: Segmentation of instant 17 from 2015

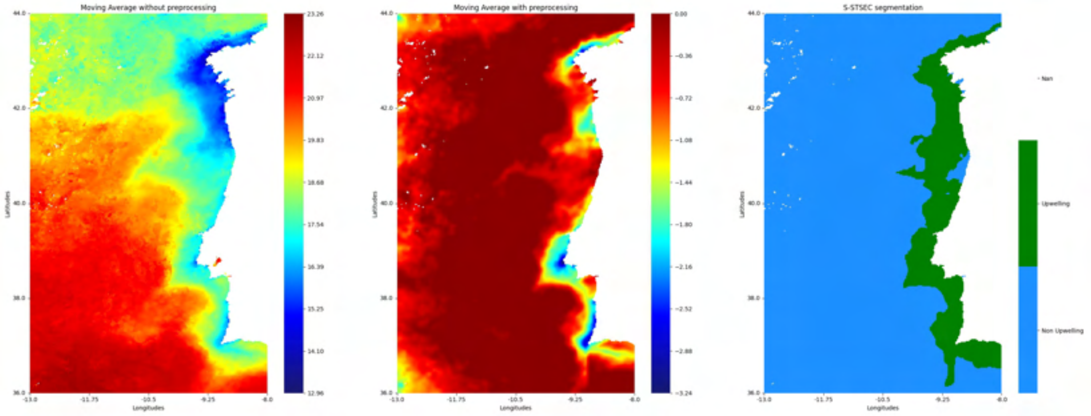


Figure I.41: Segmentation of instant 18 from 2015

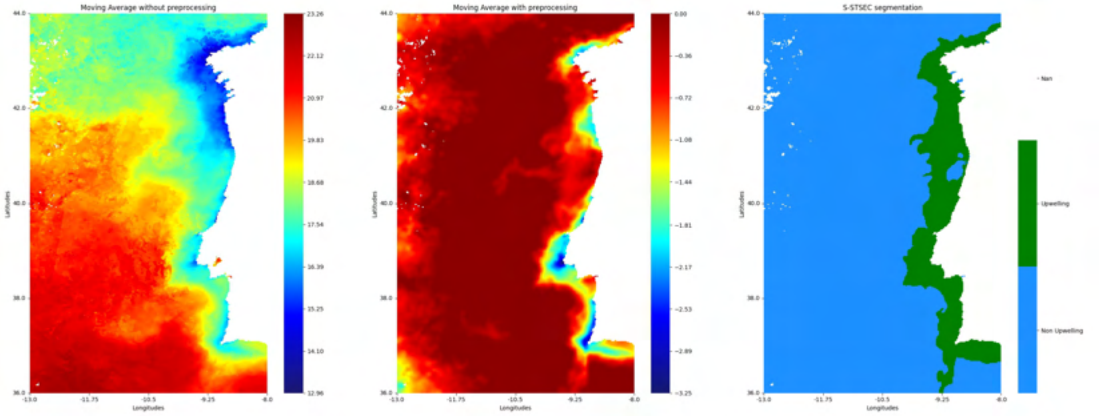


Figure I.42: Segmentation of instant 19 from 2015

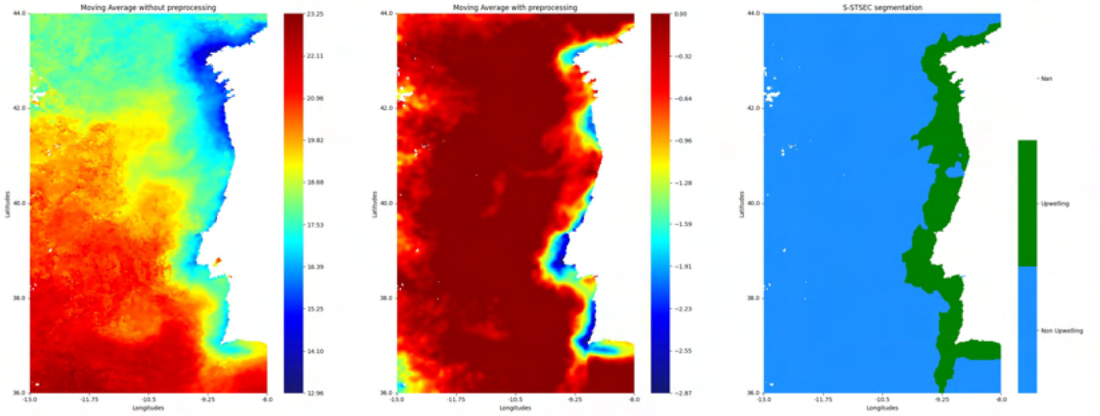


Figure I.43: Segmentation of instant 20 from 2015

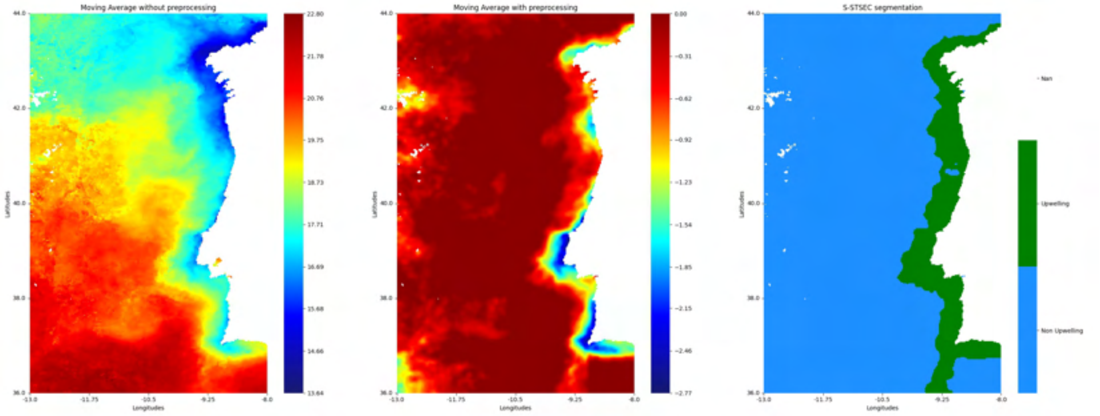


Figure I.44: Segmentation of instant 21 from 2015

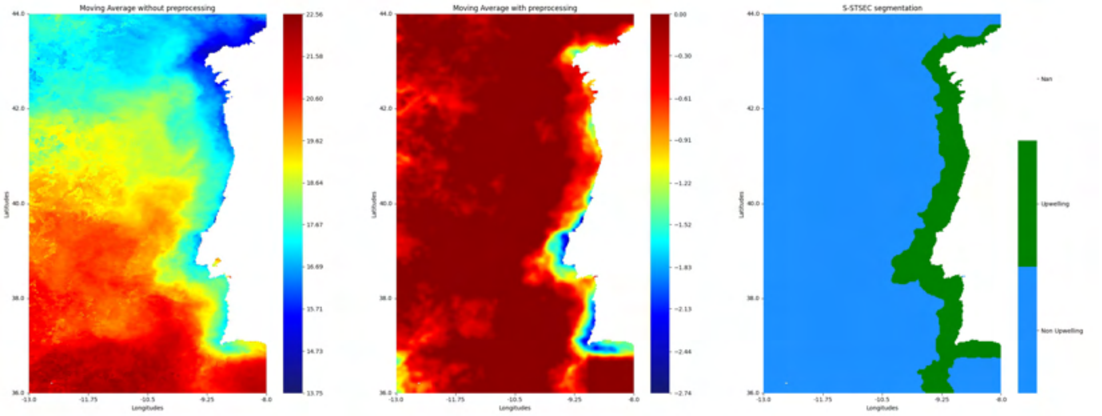


Figure I.45: Segmentation of instant 22 from 2015

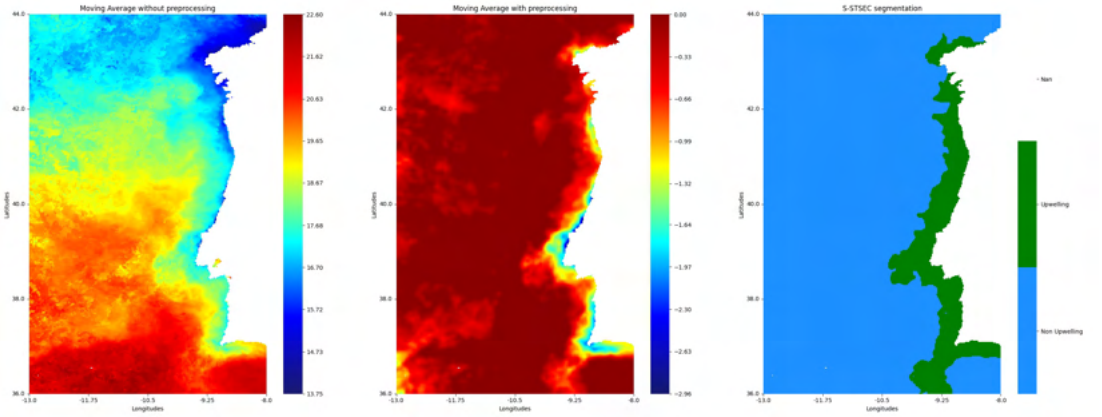


Figure I.46: Segmentation of instant 23 from 2015

I.1.3 Year 2019

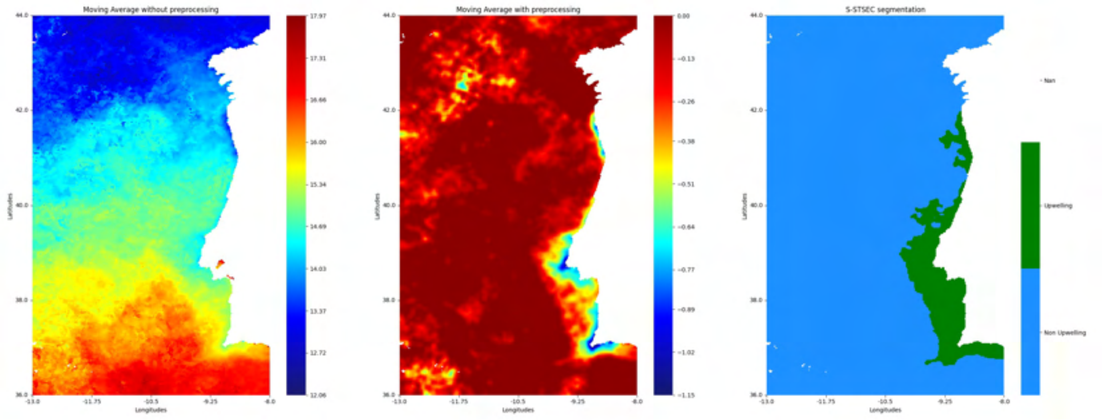


Figure I.47: Segmentation of instant 1 from 2019

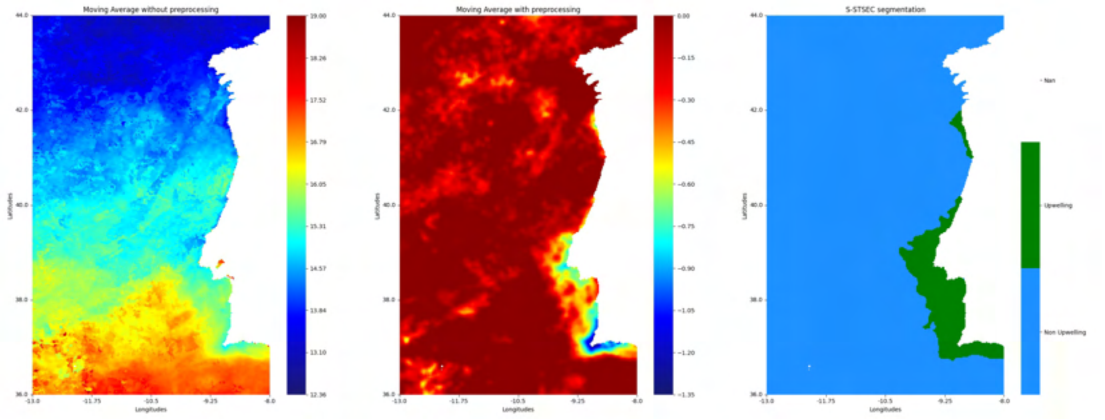


Figure I.48: Segmentation of instant 2 from 2019

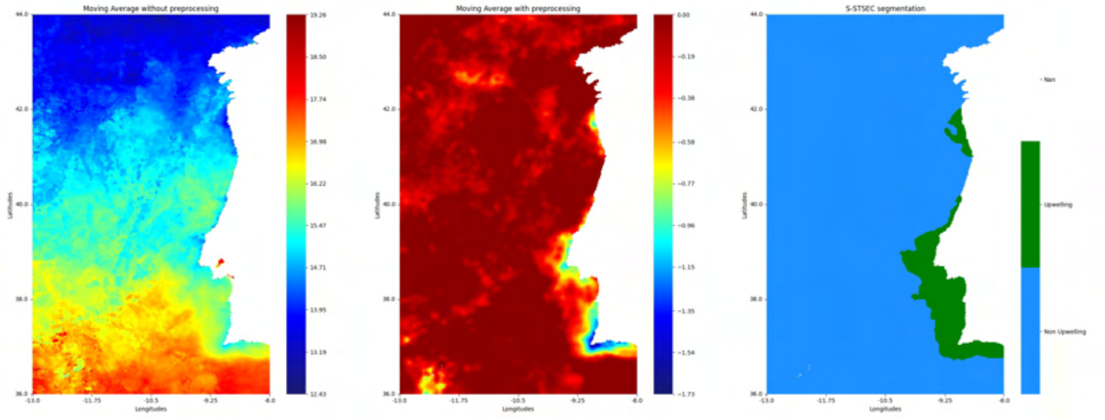


Figure I.49: Segmentation of instant 3 from 2019

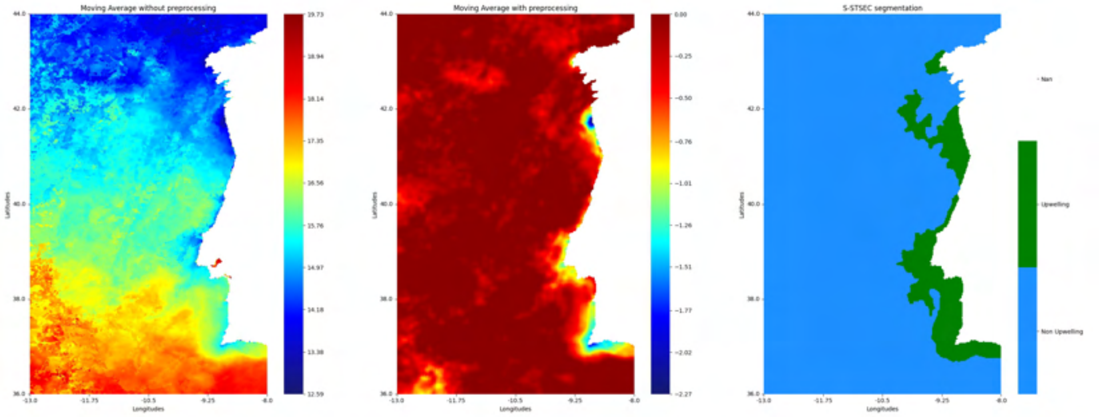


Figure I.50: Segmentation of instant 4 from 2019

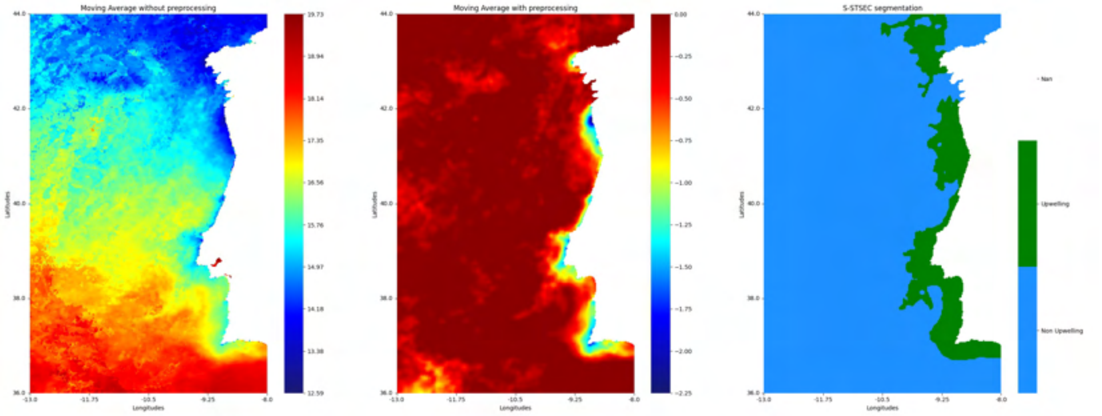


Figure I.51: Segmentation of instant 5 from 2019

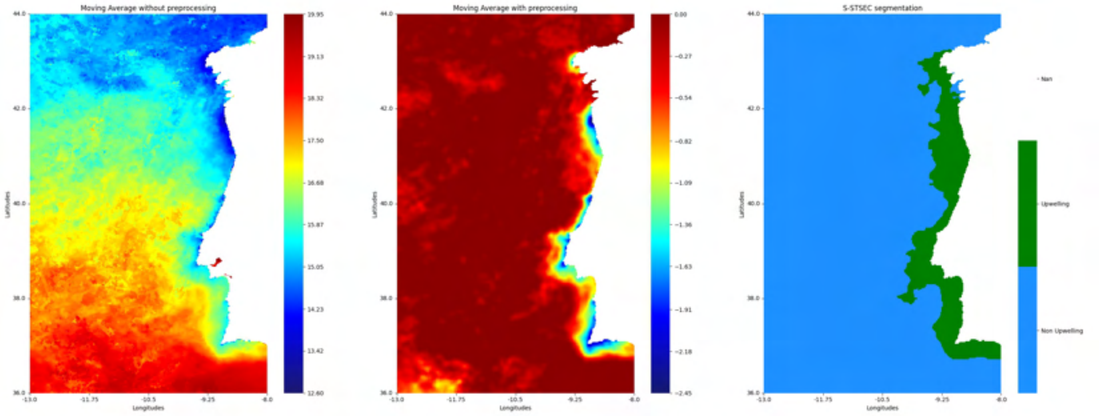


Figure I.52: Segmentation of instant 6 from 2019

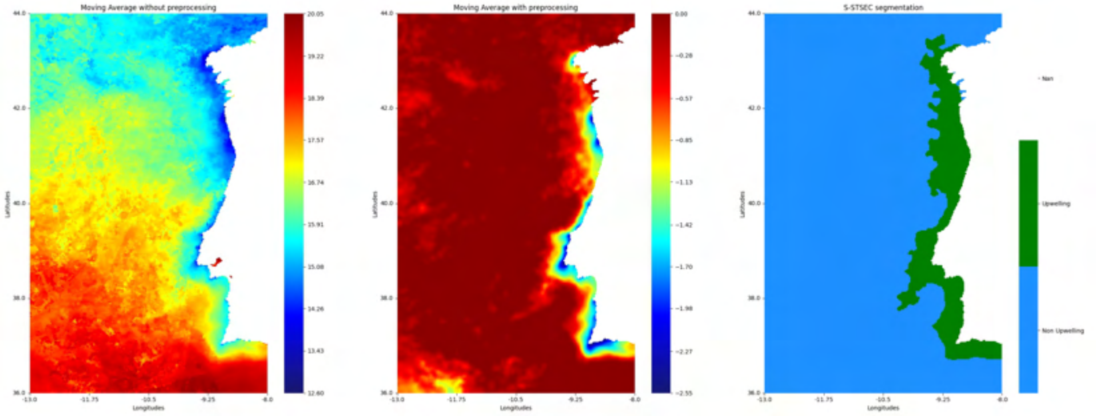


Figure I.53: Segmentation of instant 7 from 2019

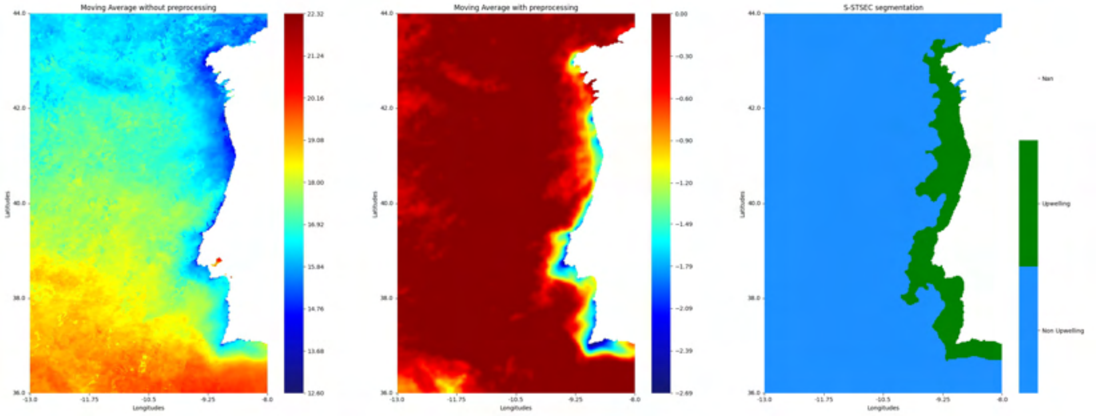


Figure I.54: Segmentation of instant 8 from 2019

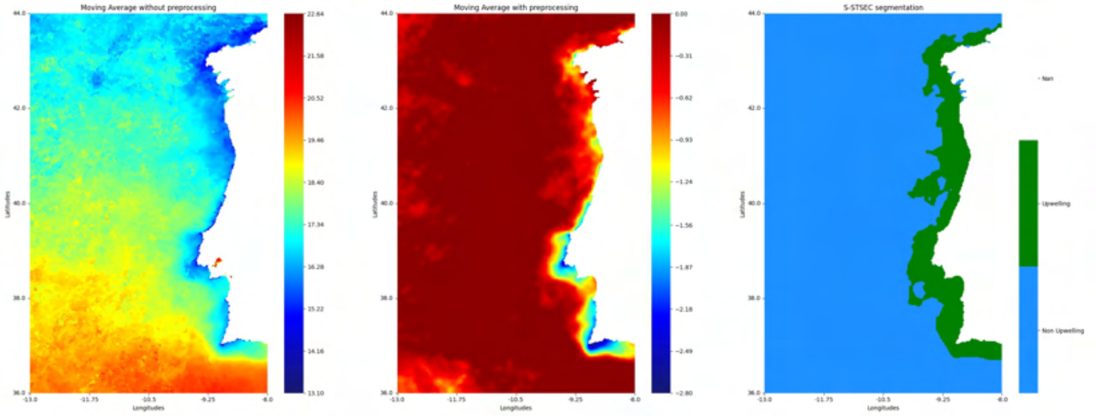


Figure I.55: Segmentation of instant 9 from 2019

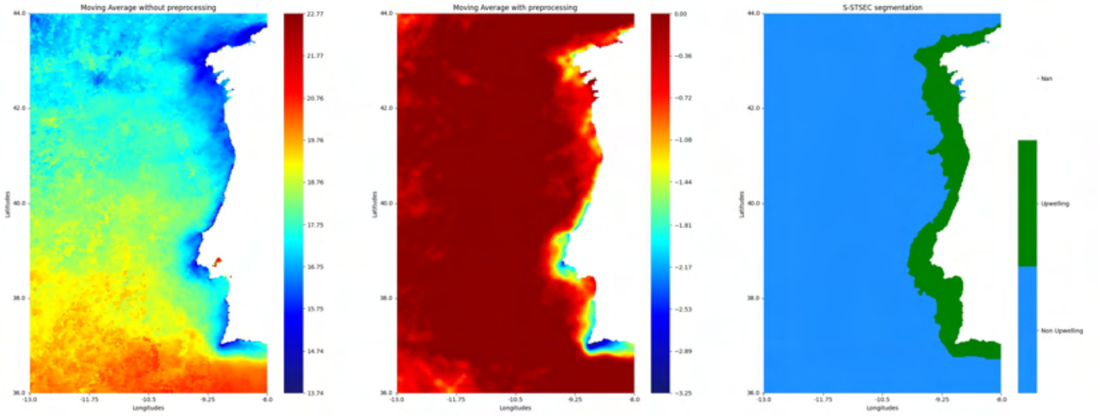


Figure I.56: Segmentation of instant 10 from 2019

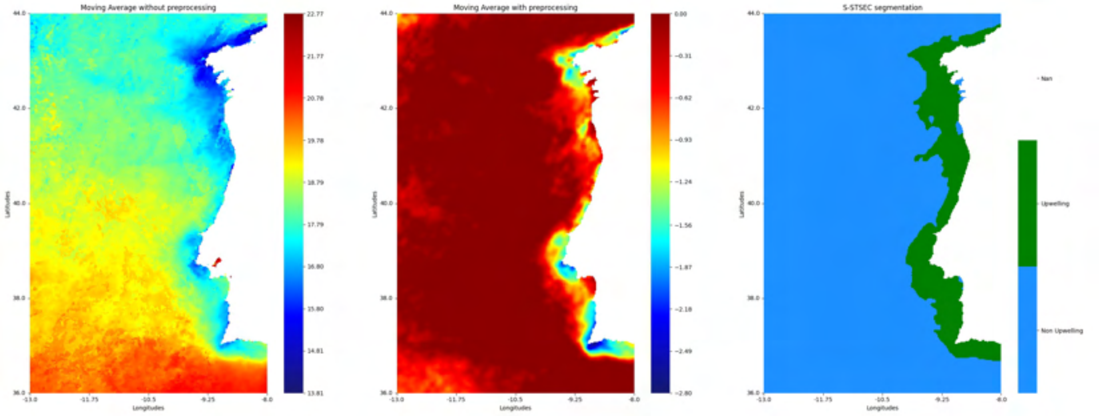


Figure I.57: Segmentation of instant 11 from 2019

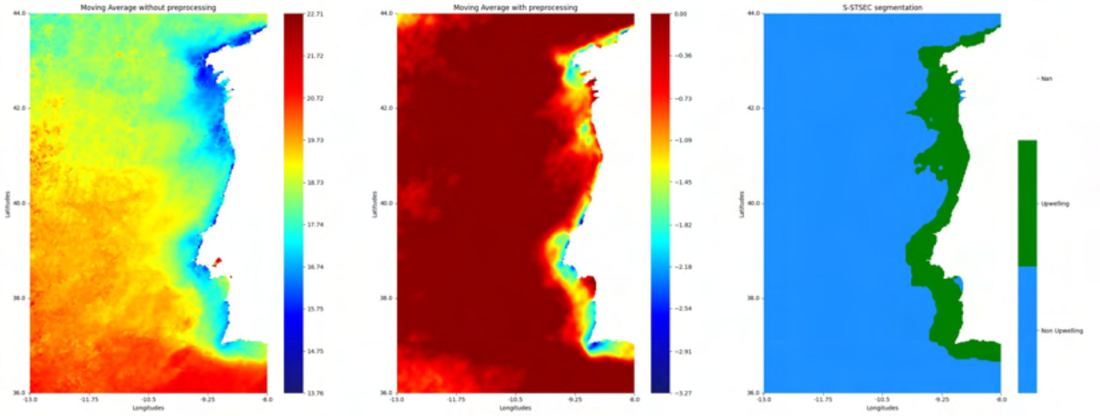


Figure I.58: Segmentation of instant 12 from 2019

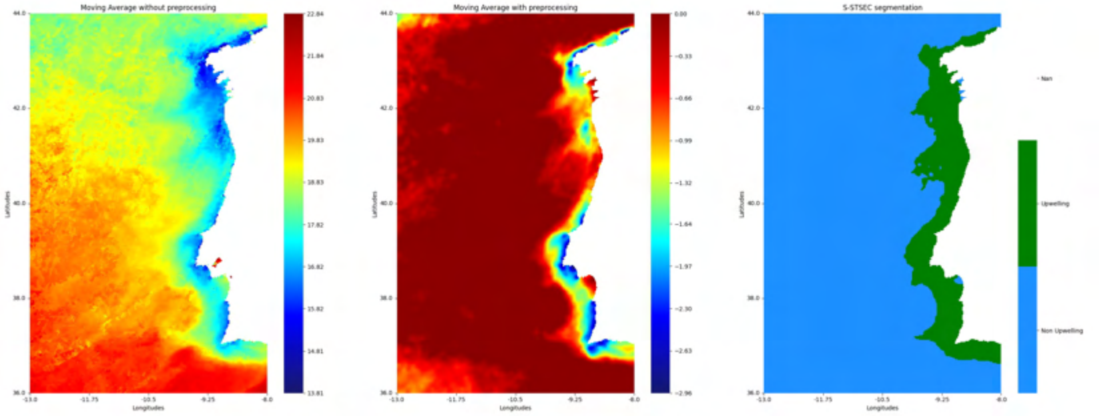


Figure I.59: Segmentation of instant 13 from 2019

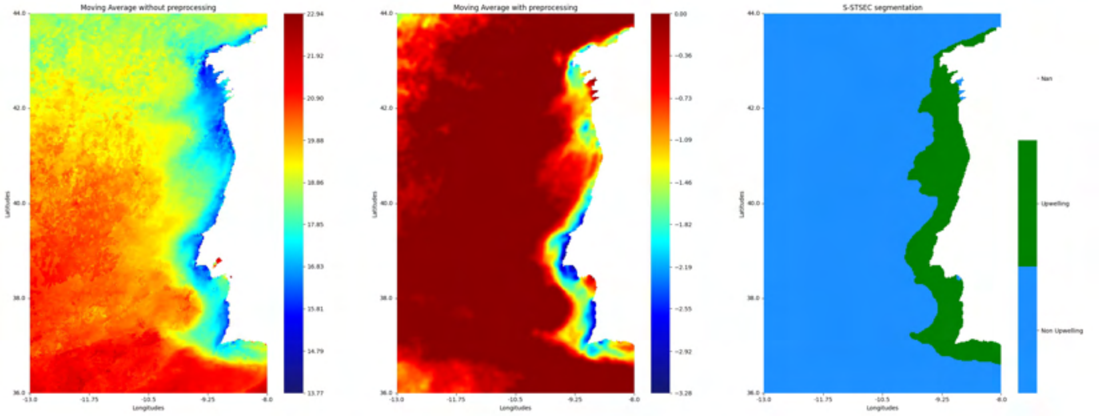


Figure I.60: Segmentation of instant 14 from 2019

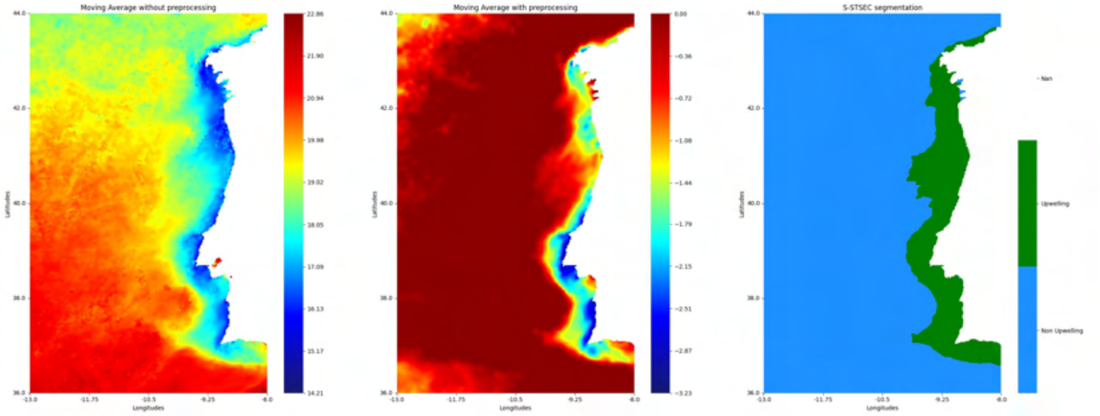


Figure I.61: Segmentation of instant 15 from 2019

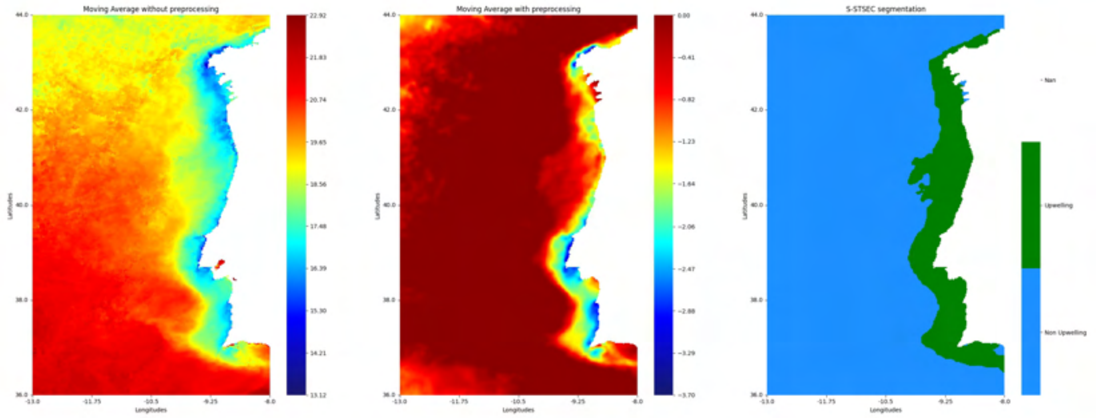


Figure I.62: Segmentation of instant 16 from 2019

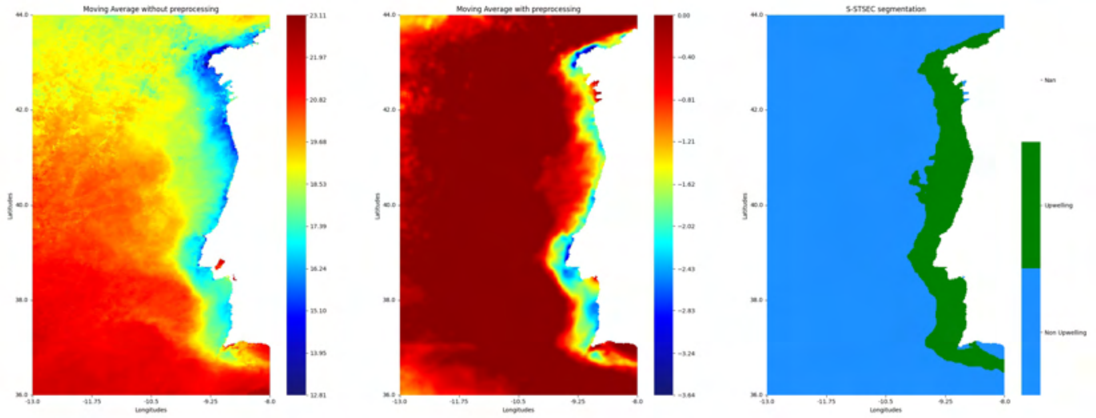


Figure I.63: Segmentation of instant 17 from 2019

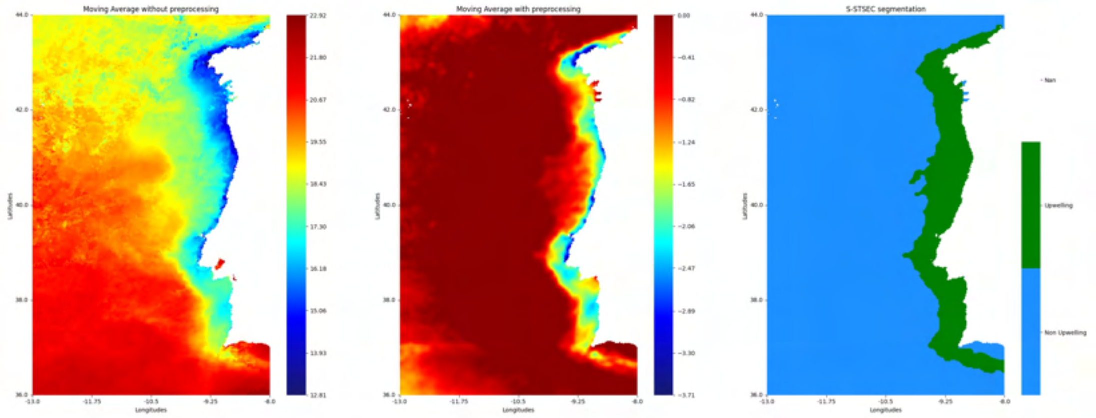


Figure I.64: Segmentation of instant 18 from 2019

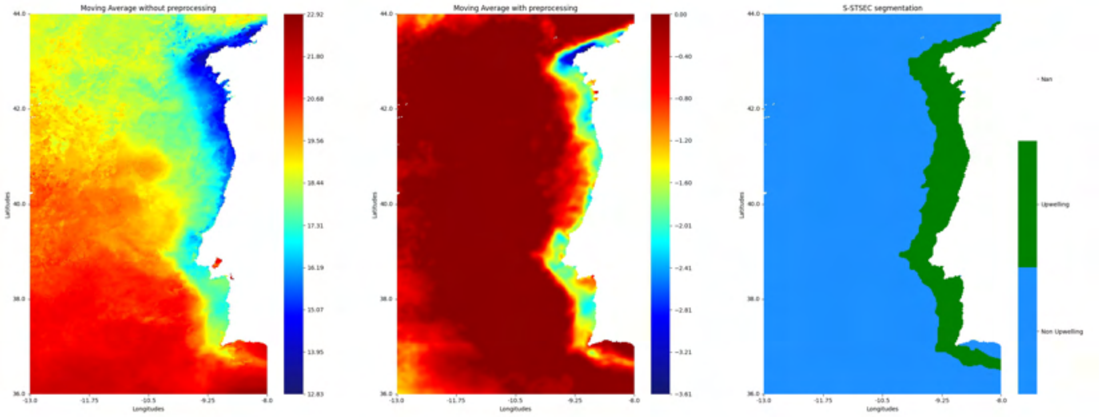


Figure I.65: Segmentation of instant 19 from 2019

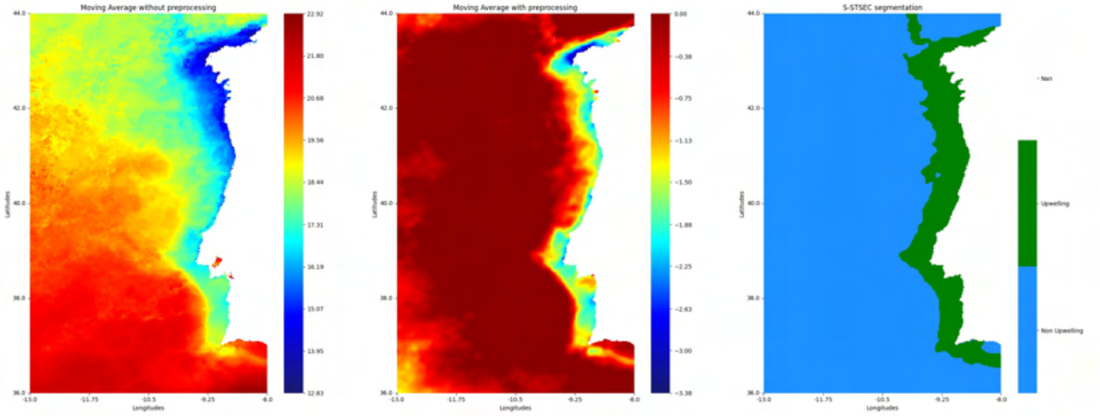


Figure I.66: Segmentation of instant 20 from 2019

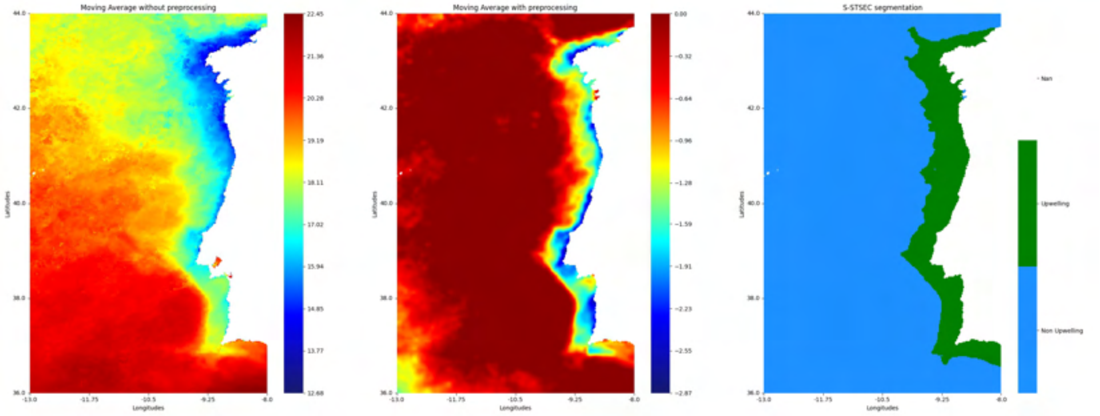


Figure I.67: Segmentation of instant 21 from 2019

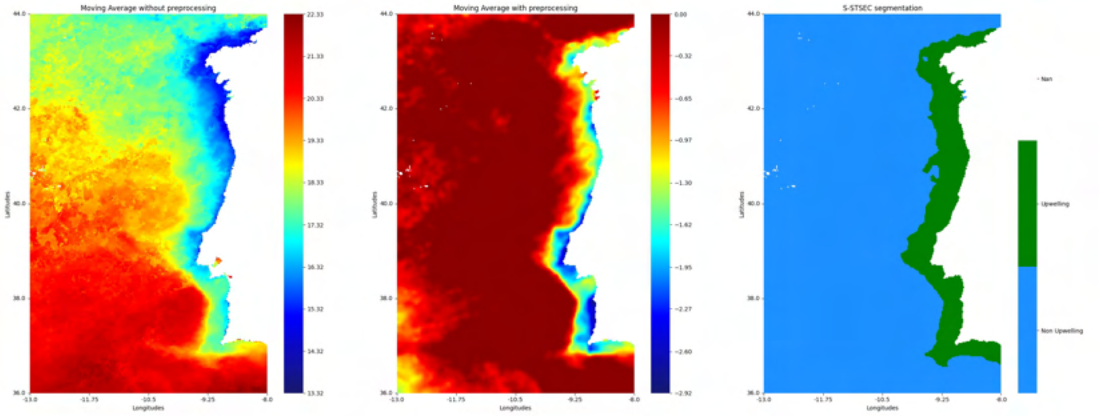


Figure I.68: Segmentation of instant 22 from 2019

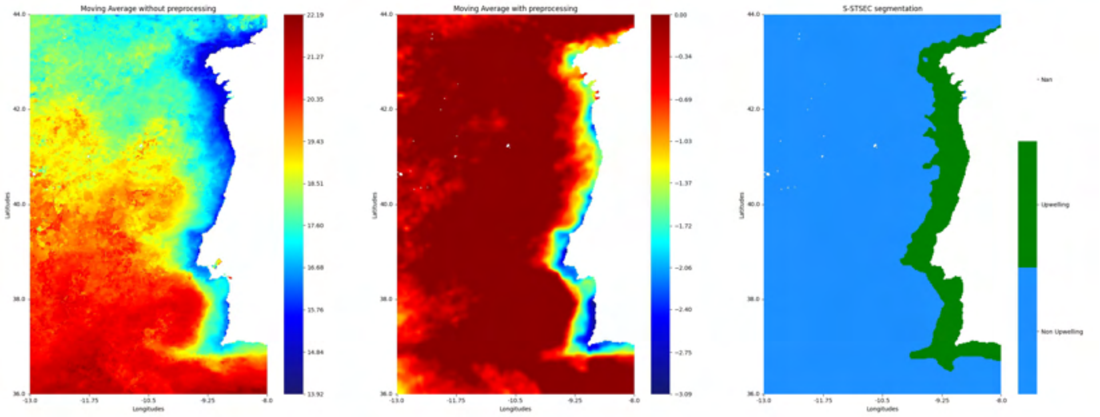


Figure I.69: Segmentation of instant 23 from 2019

I.2 Core-Shell Clusters Results over Upwelling Spans

I.2.1 Year 2007

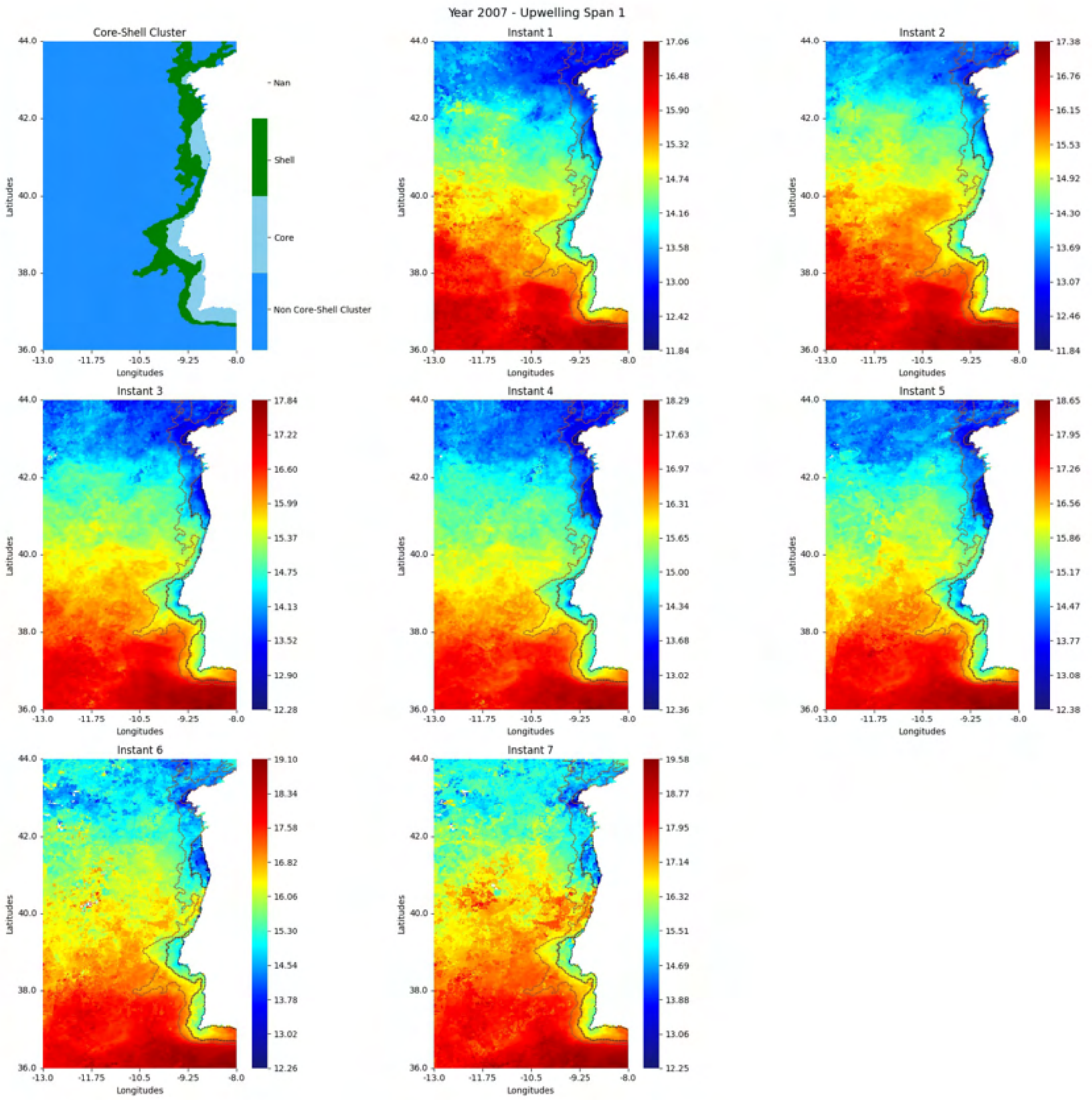


Figure I.70: Core-Shell cluster for the first upwelling span of the year 2007

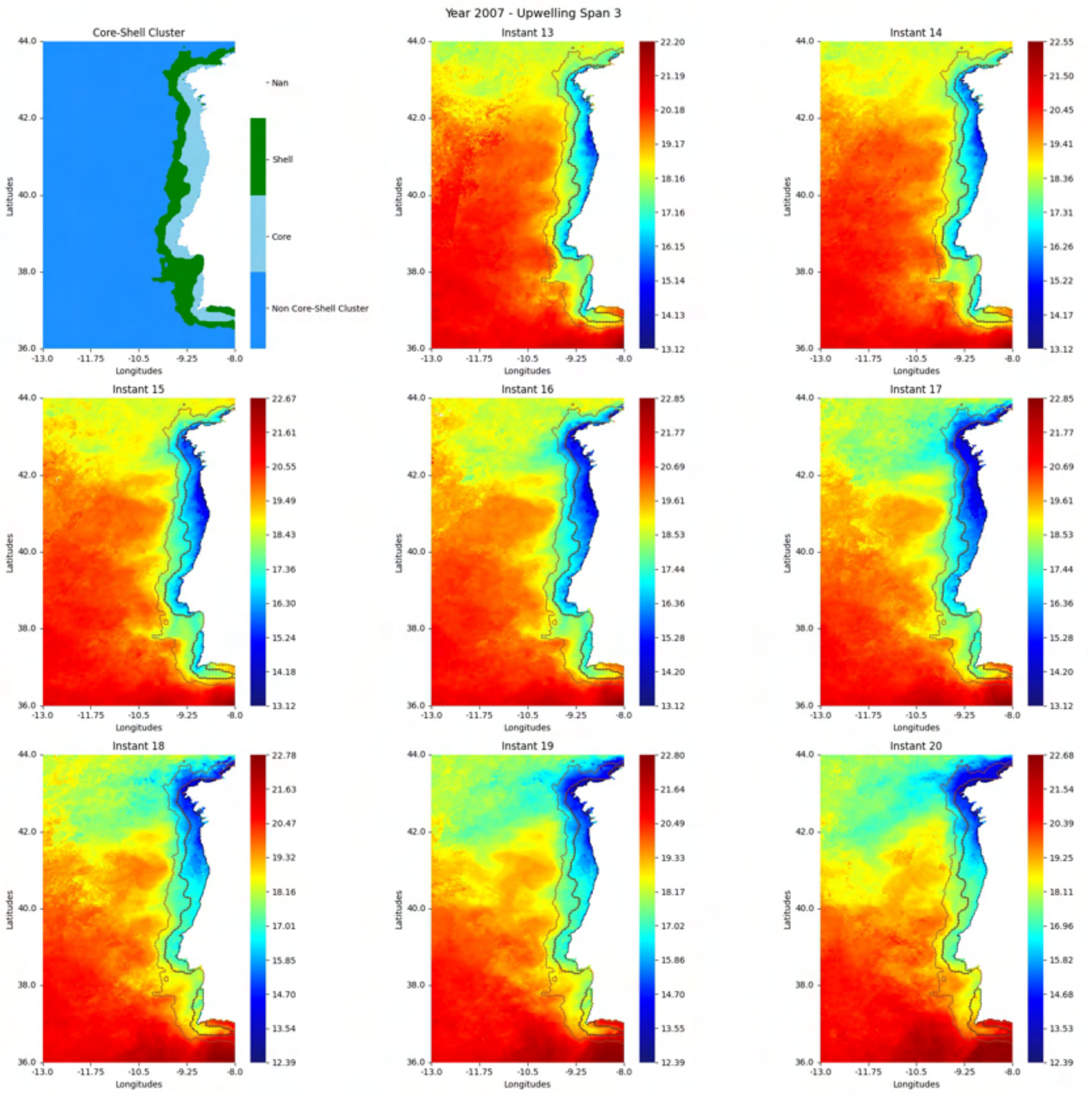


Figure I.71: Core-Shell cluster for the third upwelling span of the year 2007

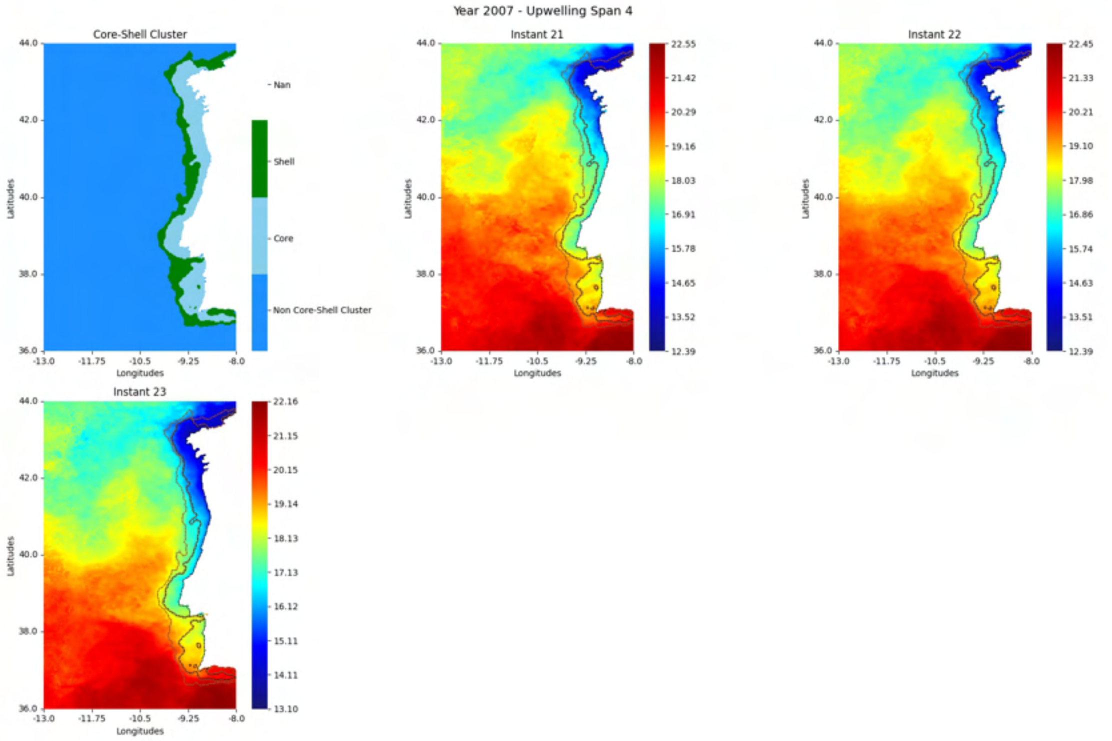


Figure I.72: Core-Shell cluster for the fourth upwelling span of the year 2007

I.2.2 Year 2015

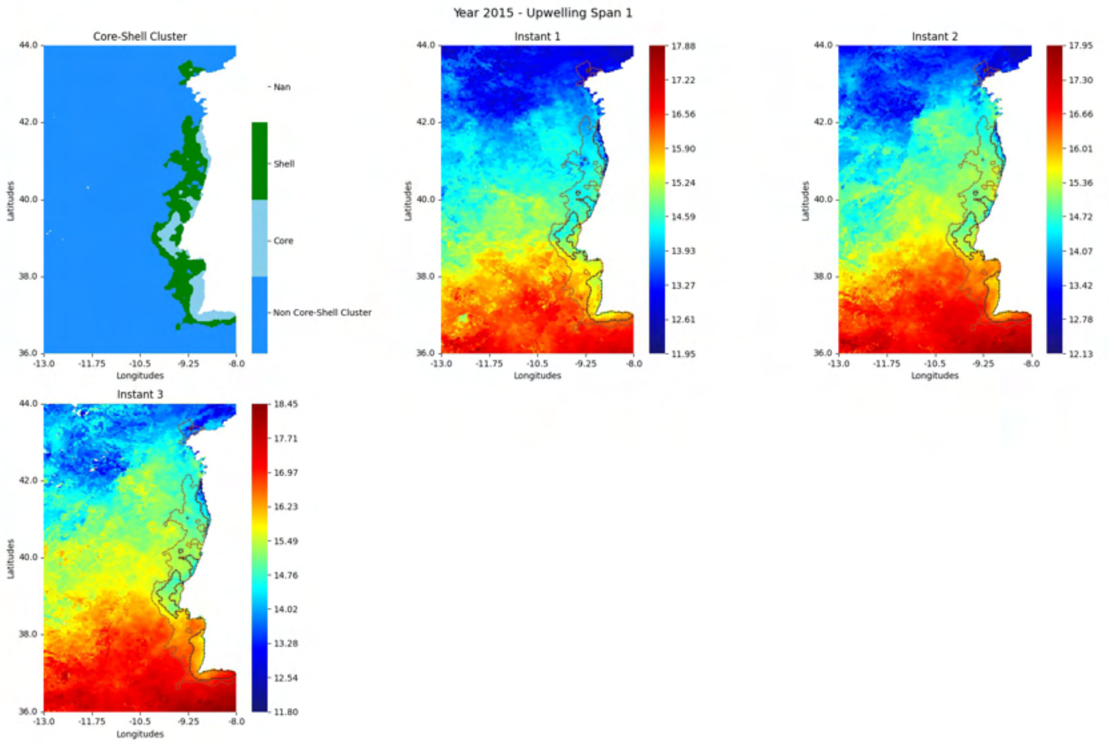


Figure I.73: Core-Shell cluster for the first upwelling span of the year 2015

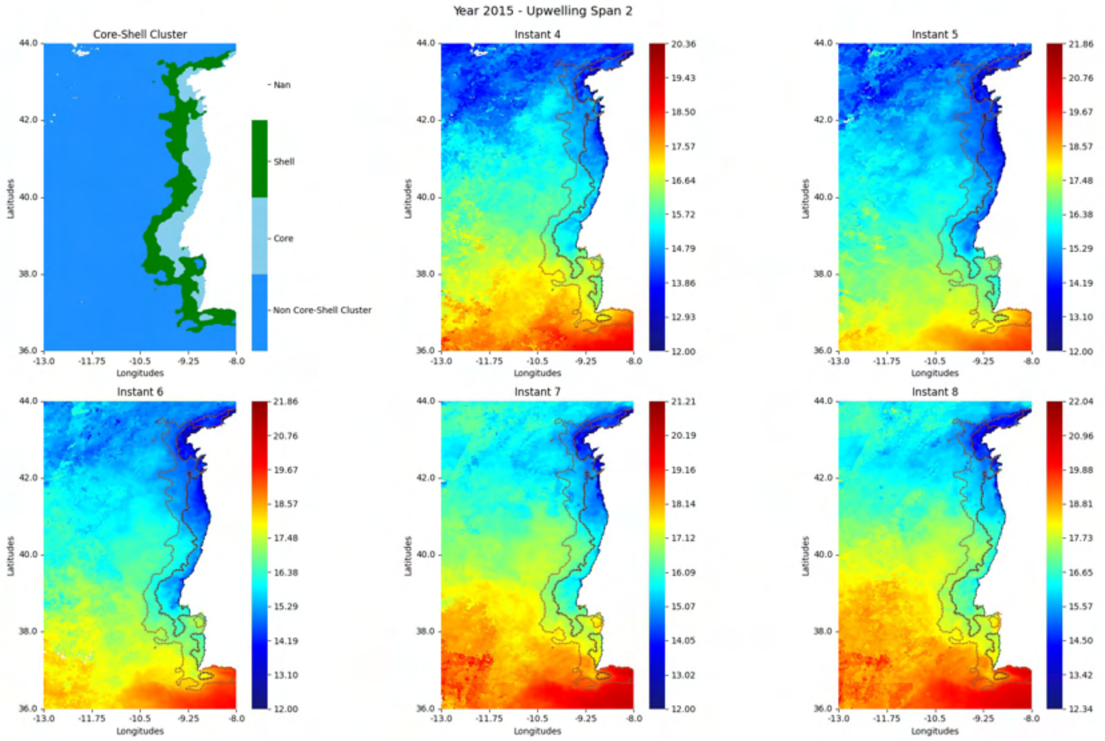


Figure I.74: Core-Shell cluster for the second upwelling span of the year 2015

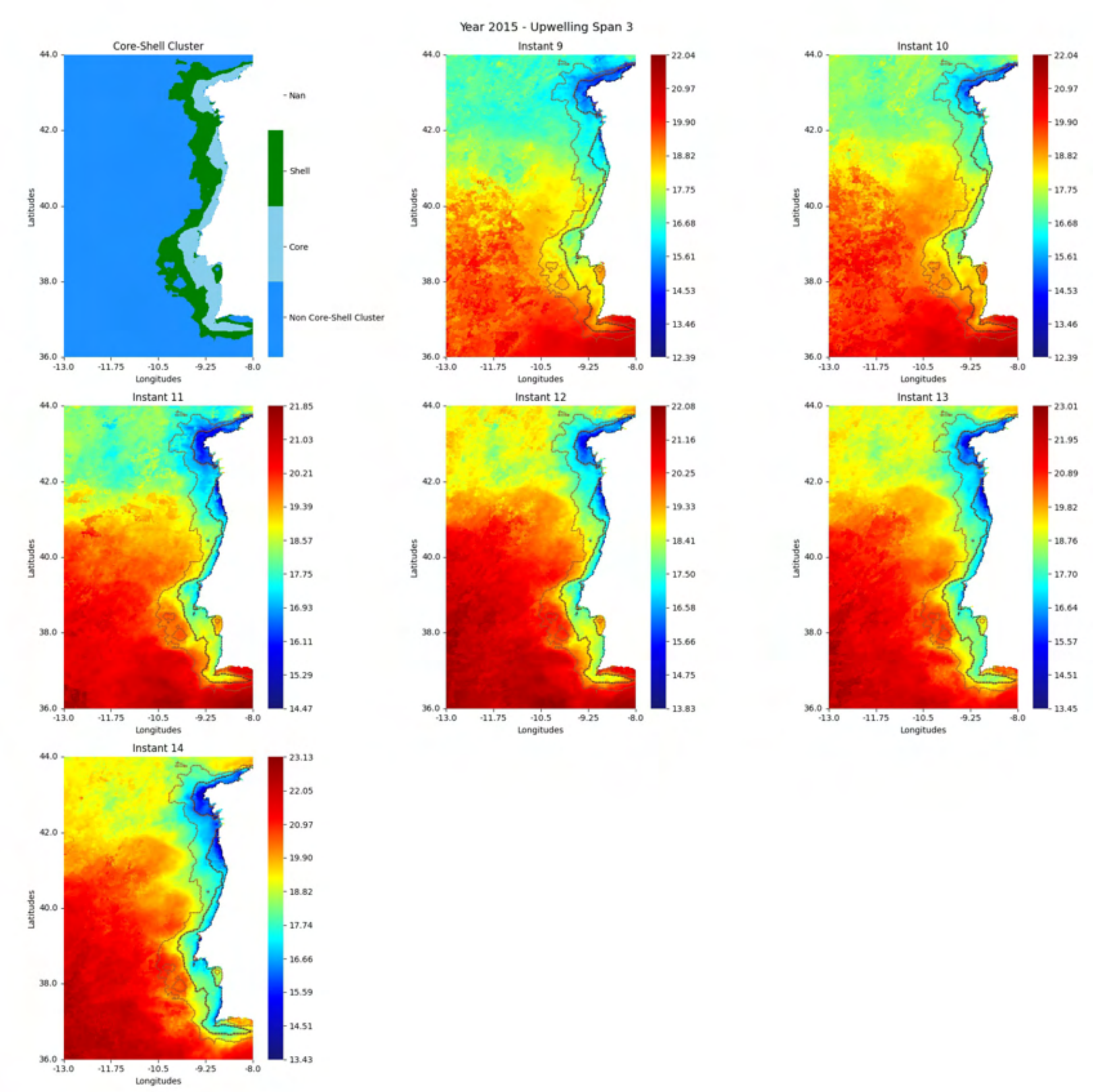


Figure I.75: Core-Shell cluster for the third upwelling span of the year 2015

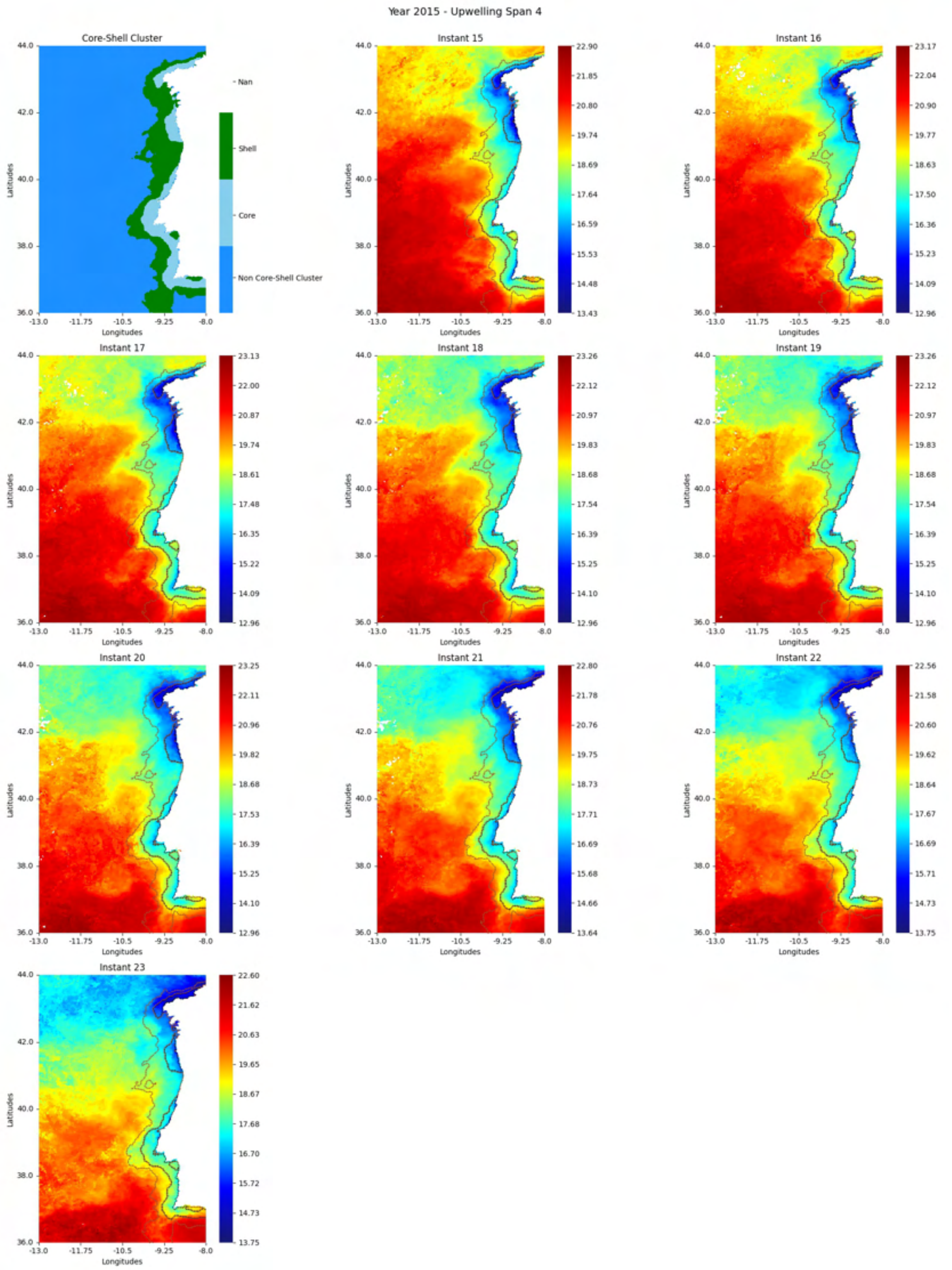


Figure I.76: Core-Shell cluster for the fourth upwelling span of the year 2015

I.2.3 Year 2019

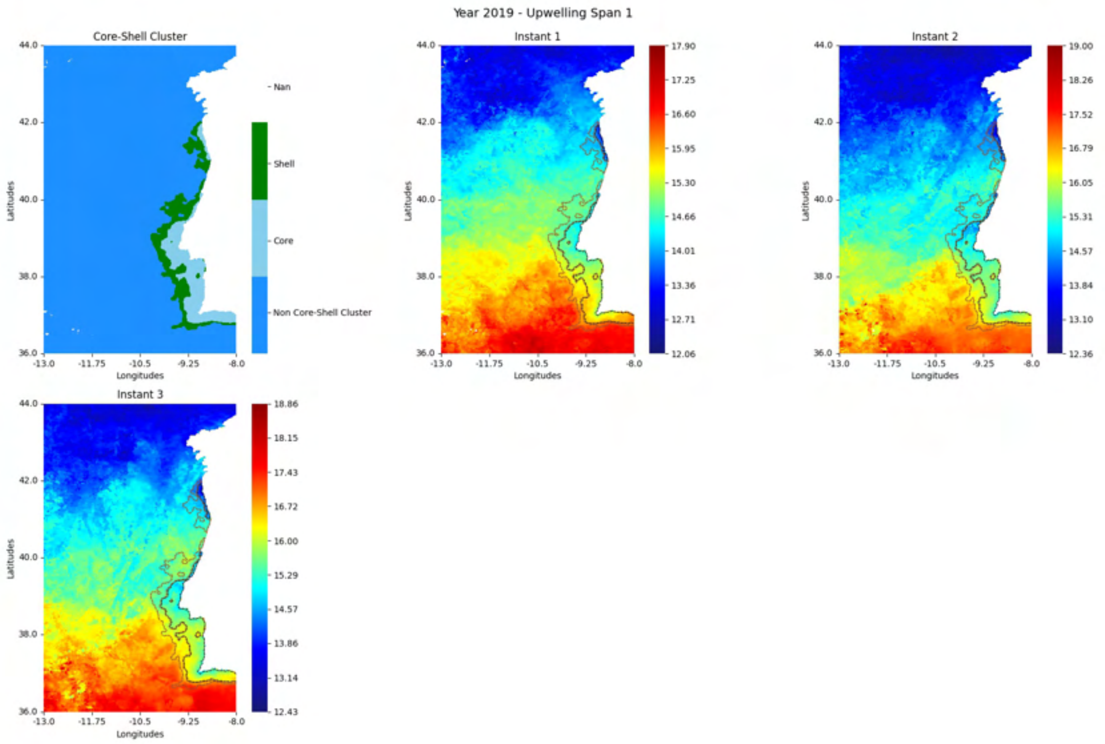


Figure I.77: Core-Shell cluster for the first upwelling span of the year 2019

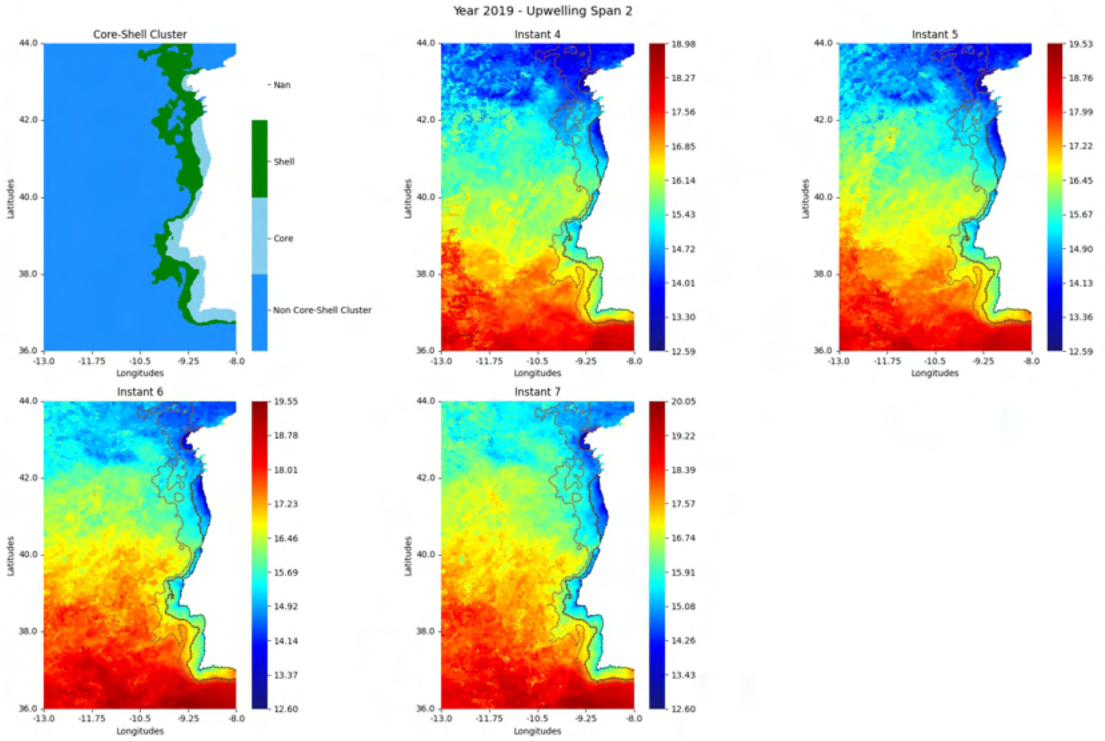
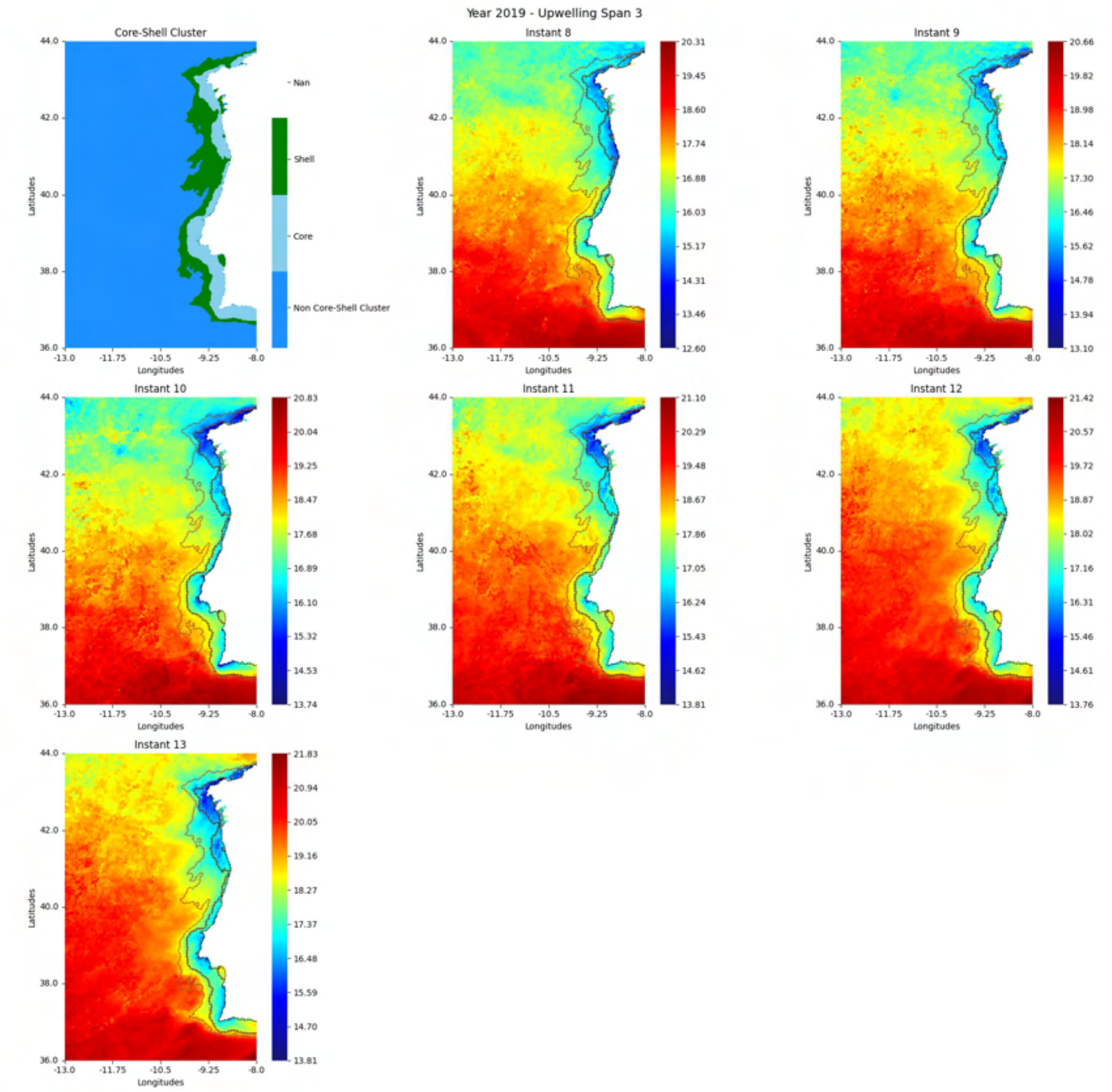


Figure I.78: Core-Shell cluster for the second upwelling span of the year 2019



Year 2019 - Upwelling Span 3

Core-Shell Cluster

Instant 8

Instant 9

Instant 10

Instant 11

Instant 12

Instant 13

Figure I.79: Core-Shell cluster for the third upwelling span of the year 2019

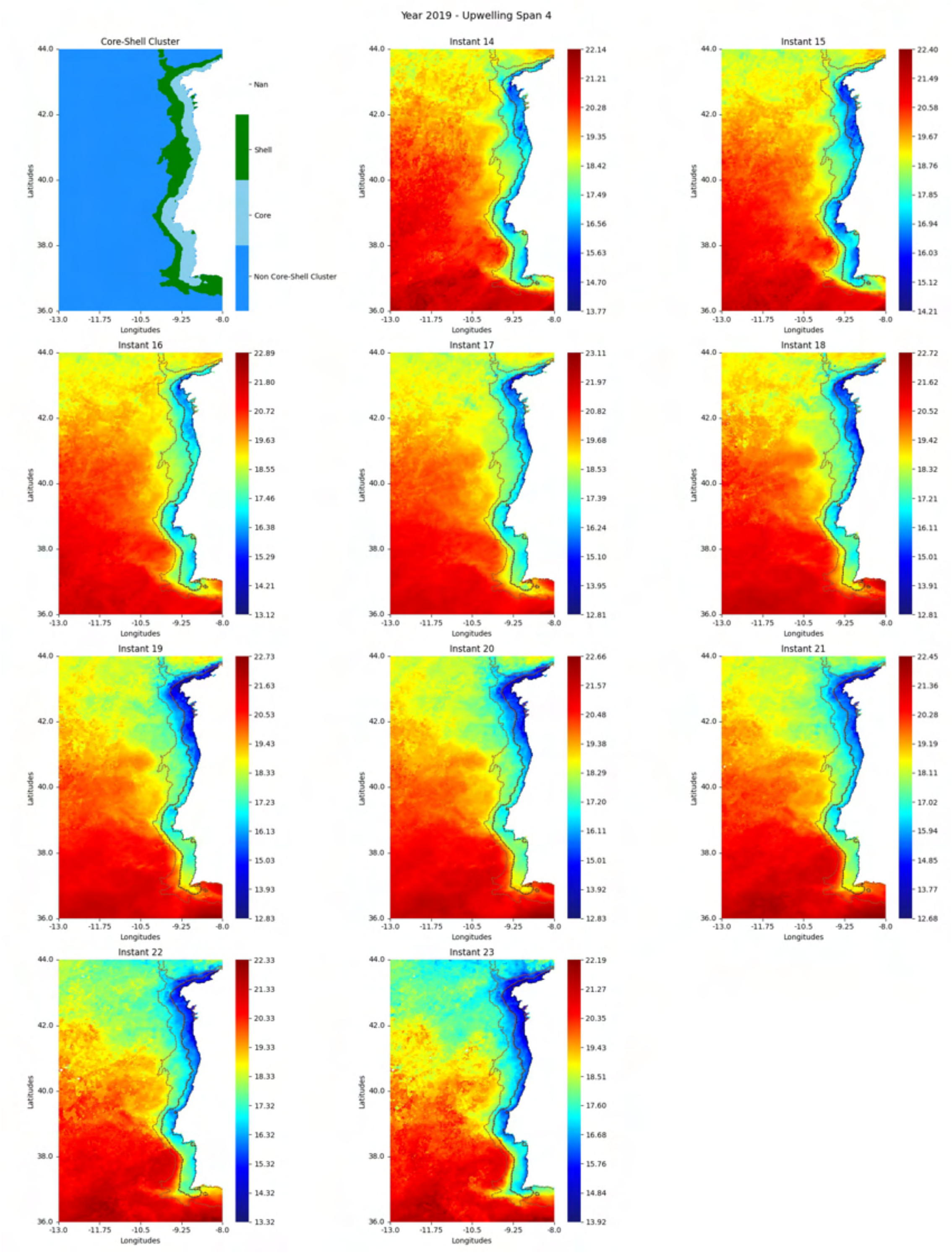


Figure I.80: Core-Shell cluster for the fourth upwelling span of the year 2019

I.3 Original SST images with Core and Shell front delineation

I.3.1 Year 2007

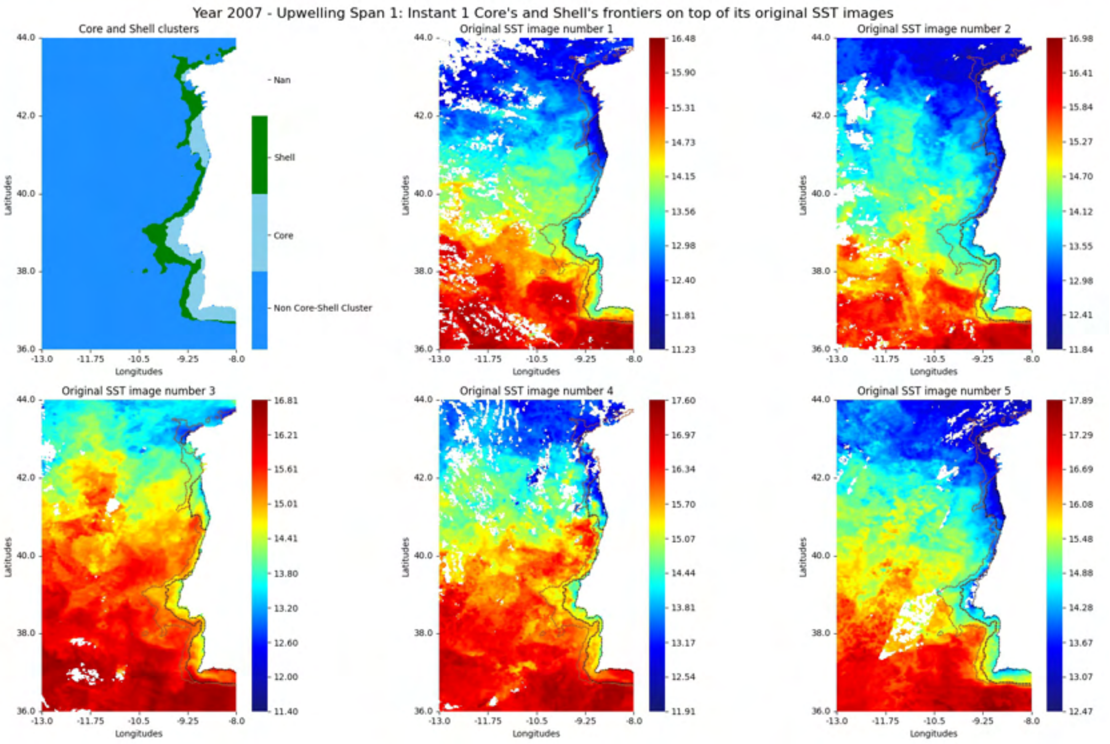


Figure I.81: Fronts regarding instant 1 of year 2007

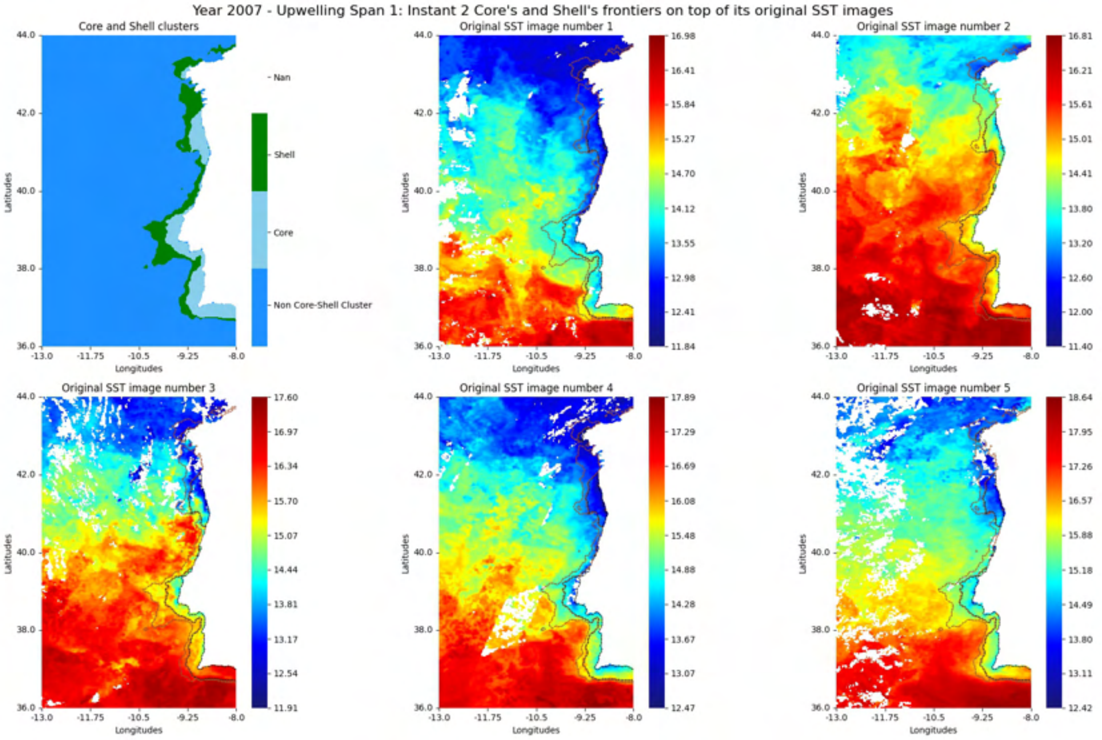


Figure I.82: Fronts regarding instant 2 of year 2007

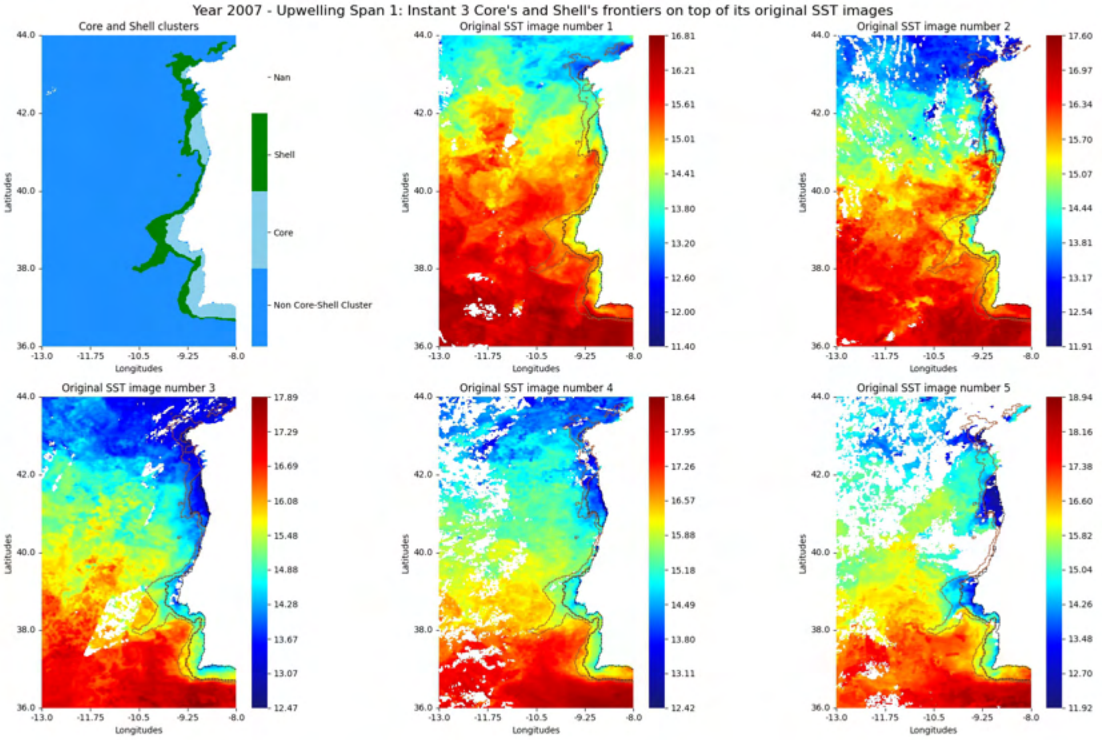


Figure I.83: Fronts regarding instant 3 of year 2007

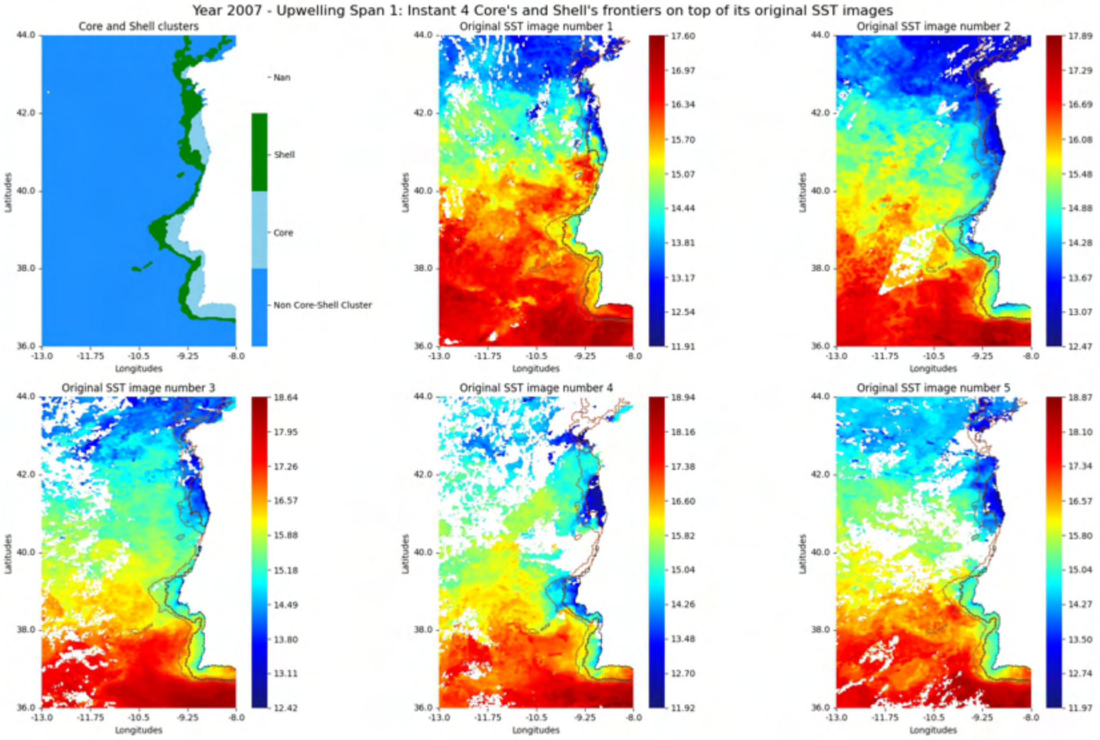


Figure I.84: Fronts regarding instant 4 of year 2007

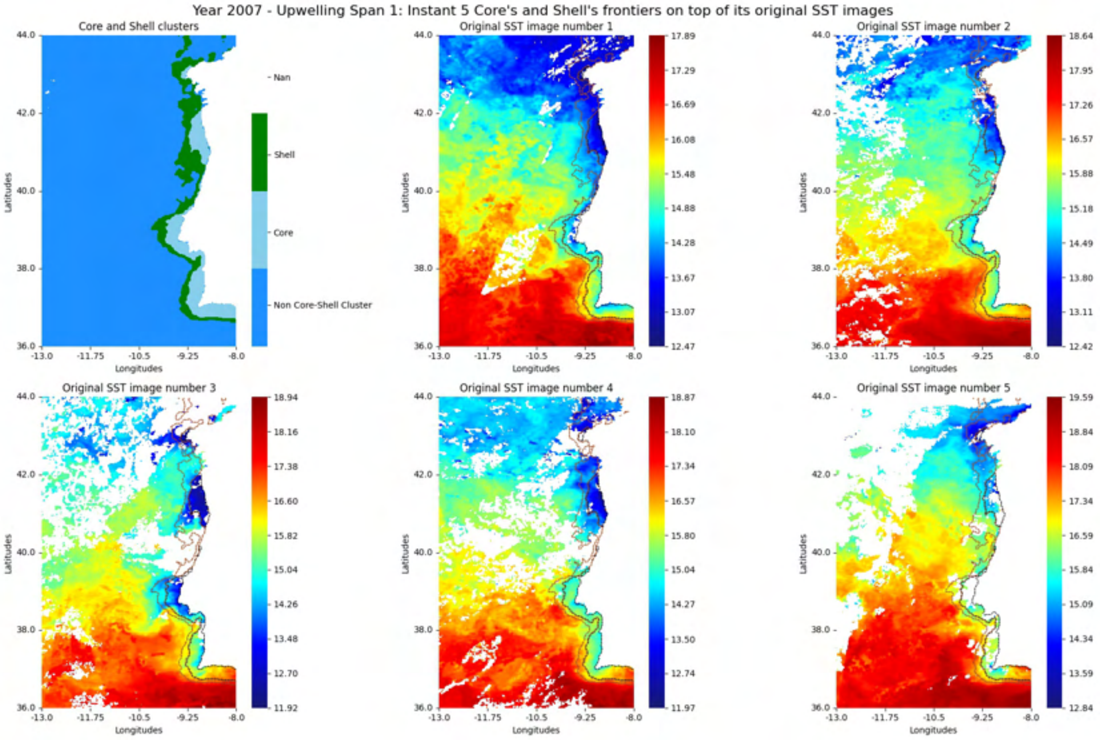


Figure I.85: Fronts regarding instant 5 of year 2007

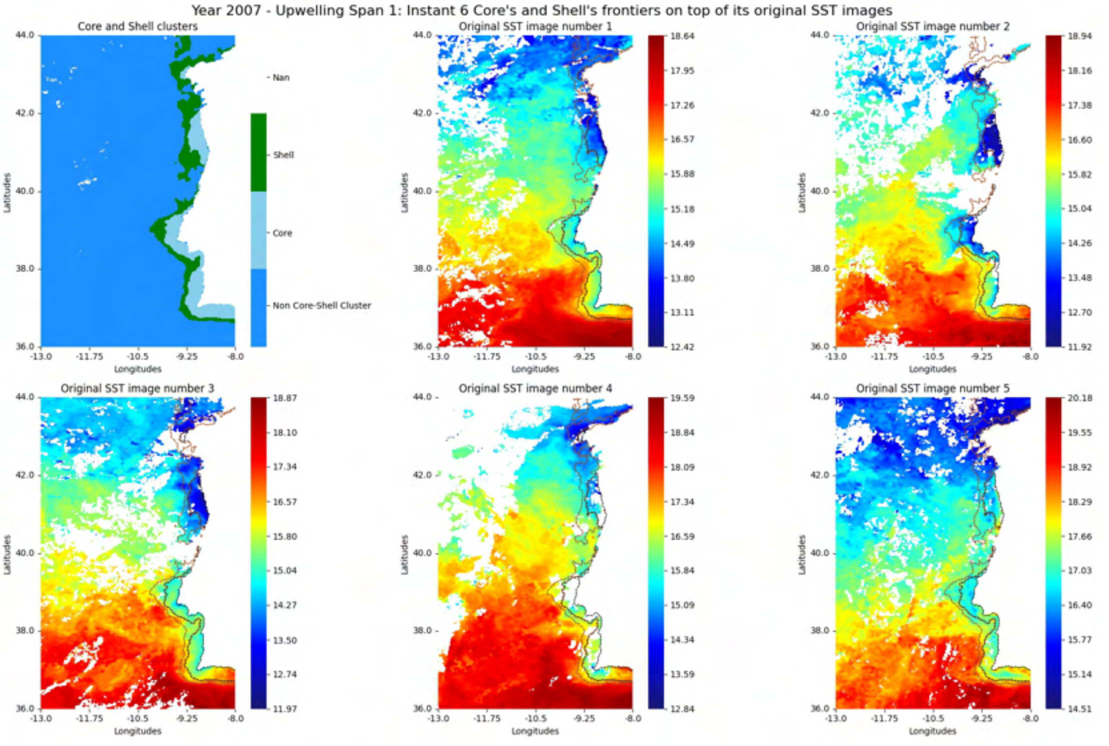


Figure I.86: Fronts regarding instant 6 of year 2007

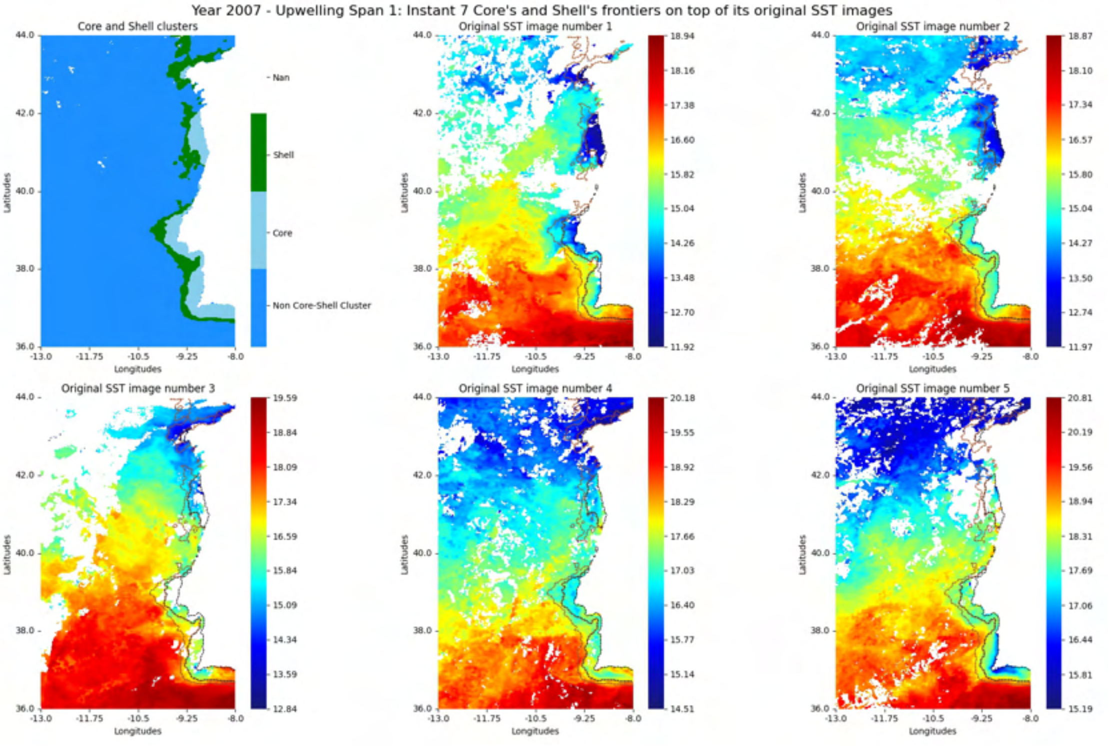


Figure I.87: Fronts regarding instant 7 of year 2007

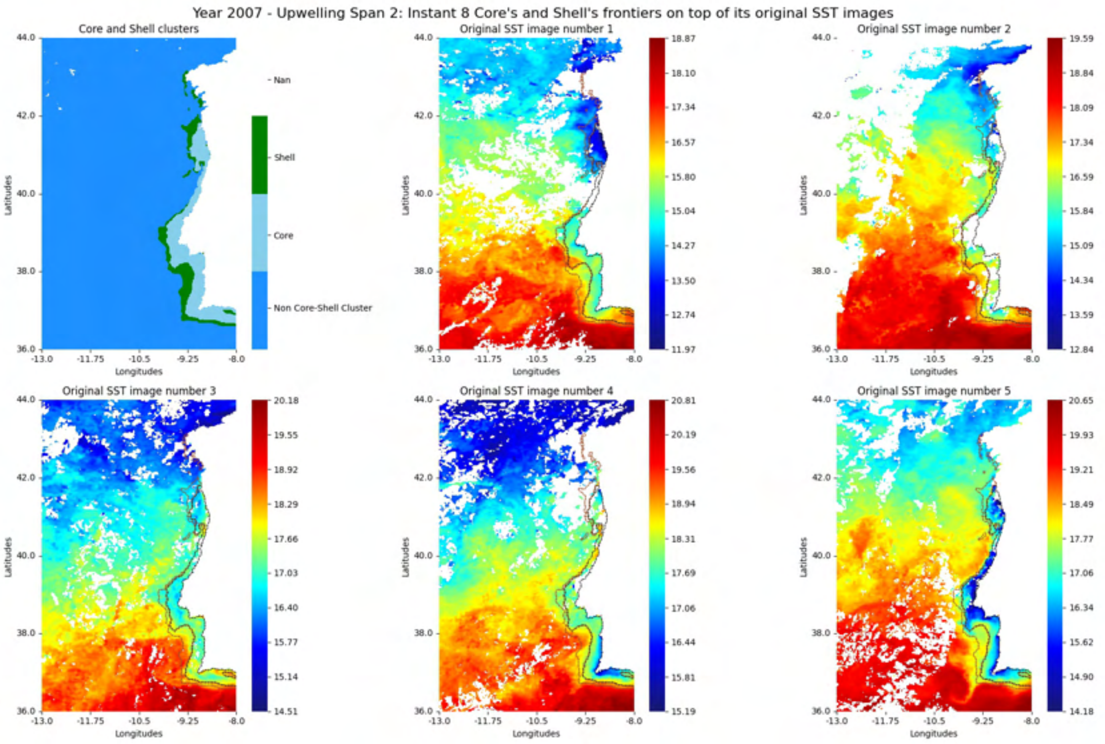


Figure I.88: Fronts regarding instant 8 of year 2007

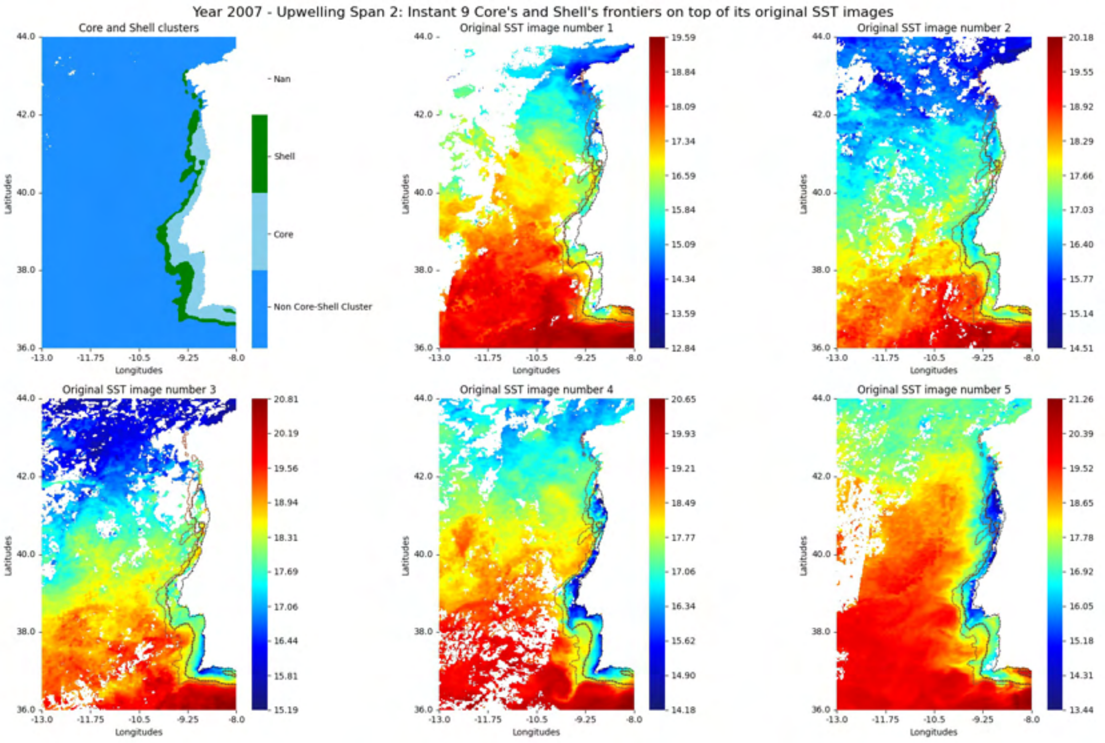


Figure I.89: Fronts regarding instant 9 of year 2007

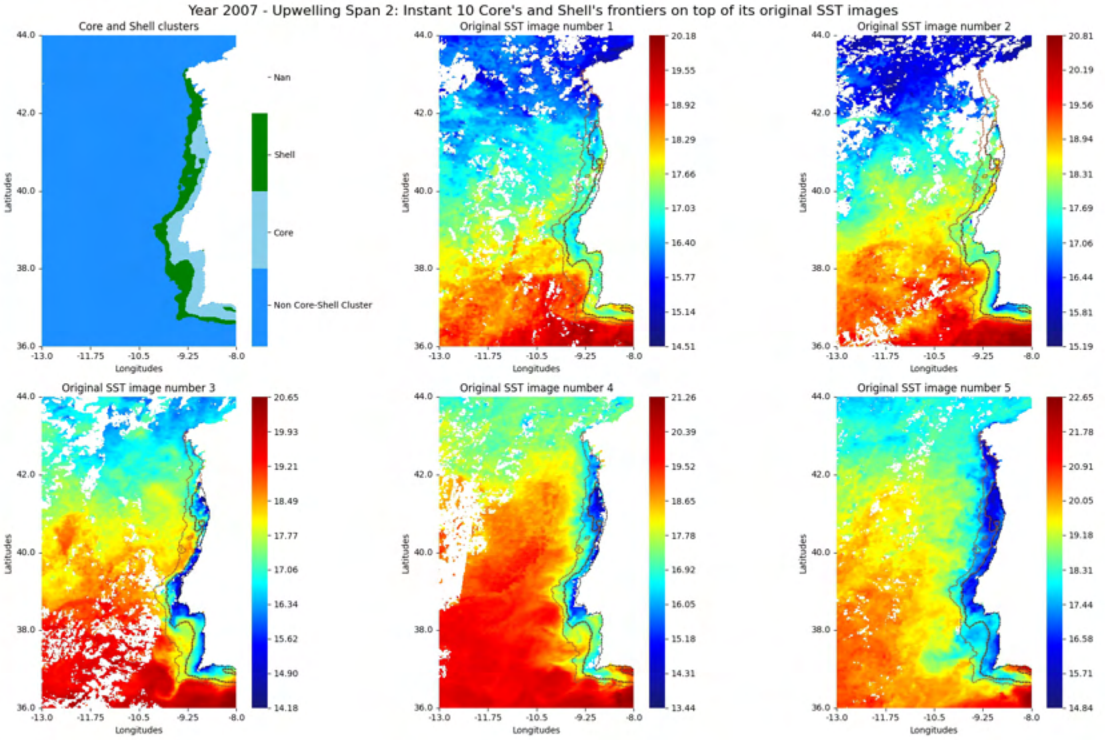


Figure I.90: Fronts regarding instant 10 of year 2007

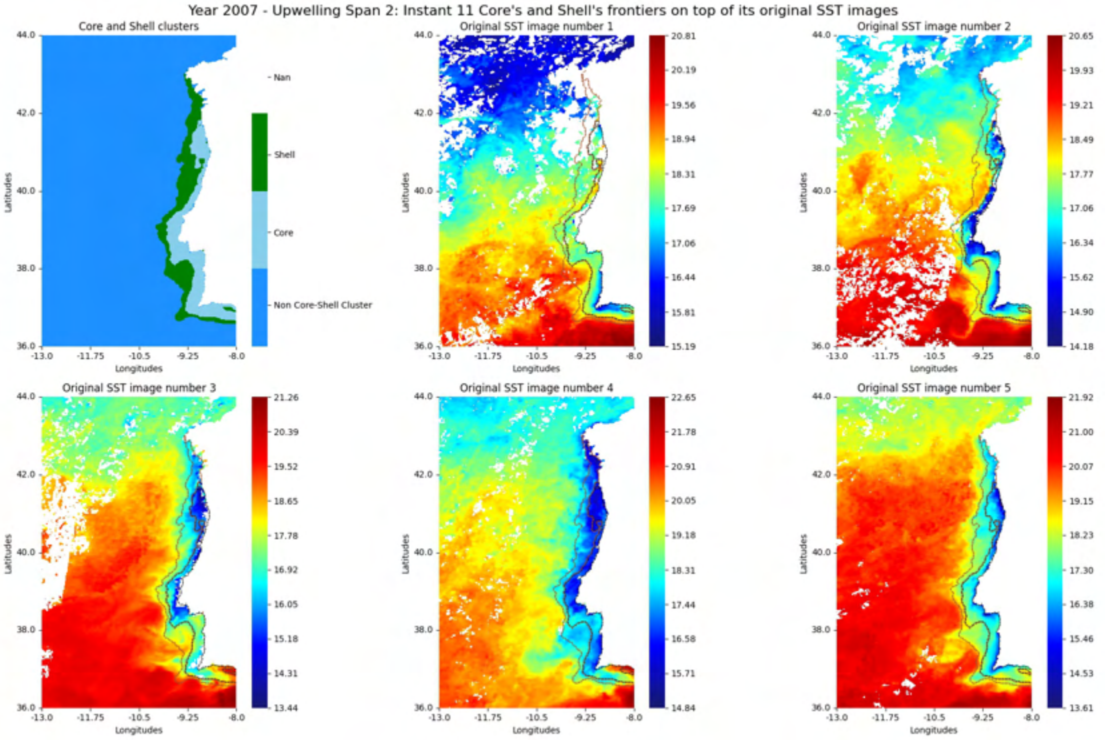


Figure I.91: Fronts regarding instant 11 of year 2007

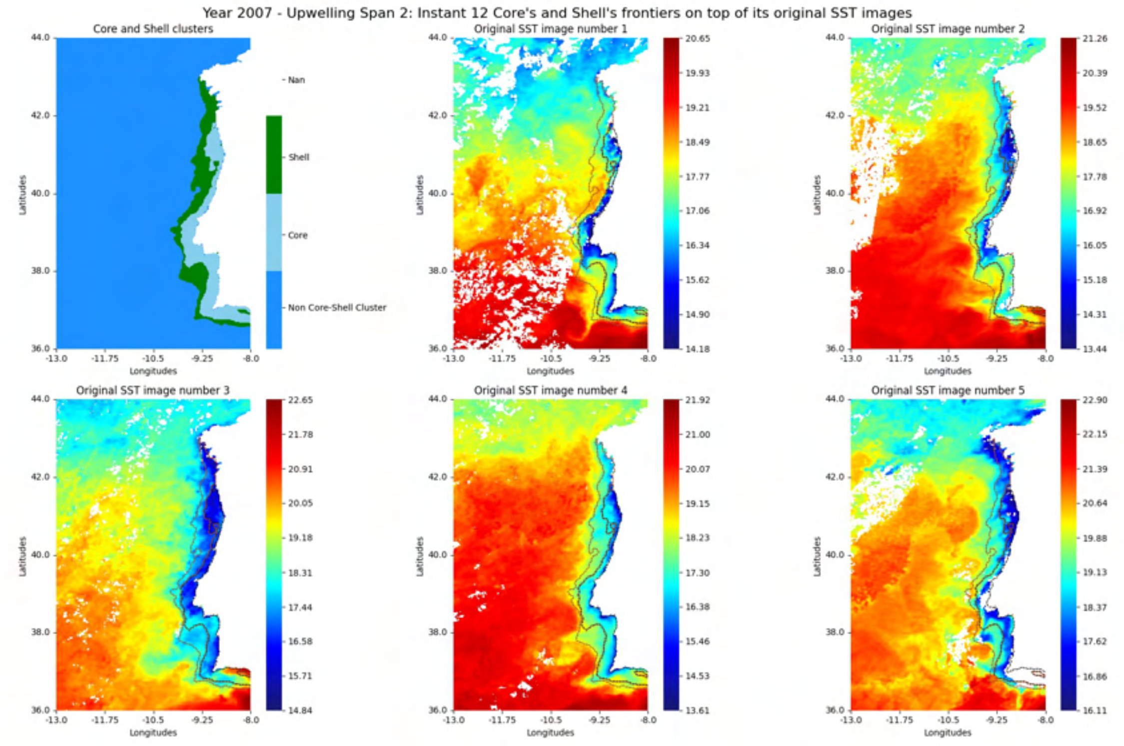


Figure I.92: Fronts regarding instant 12 of year 2007

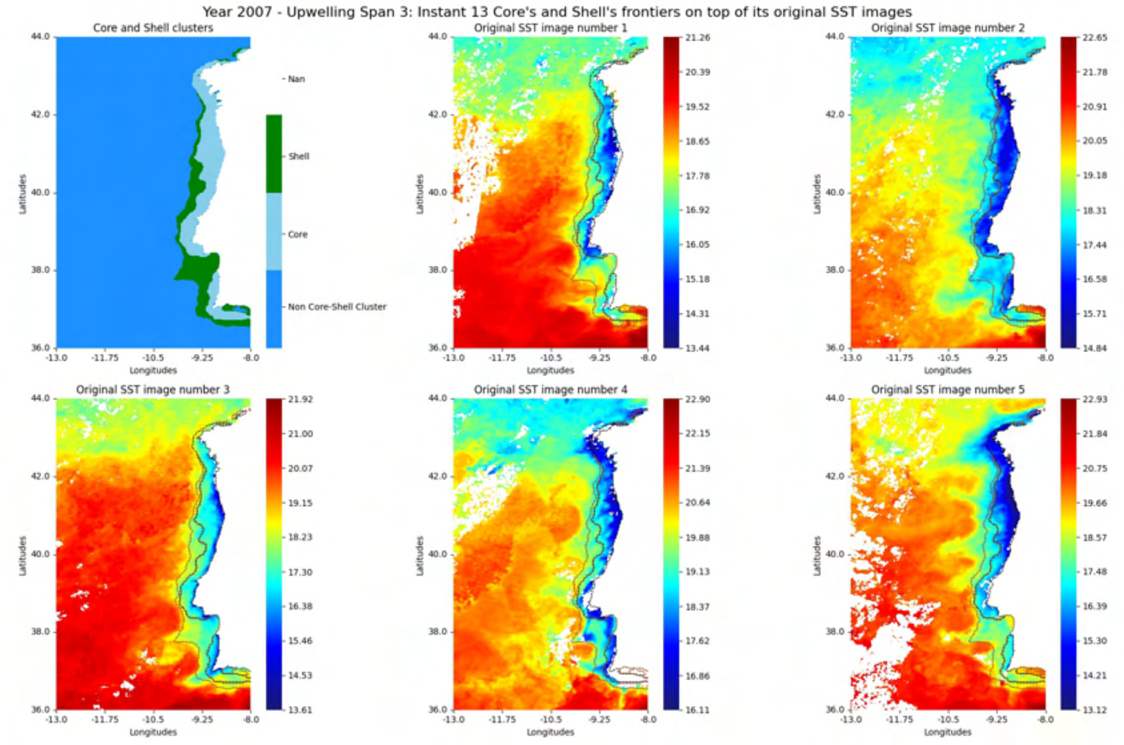


Figure I.93: Fronts regarding instant 13 of year 2007

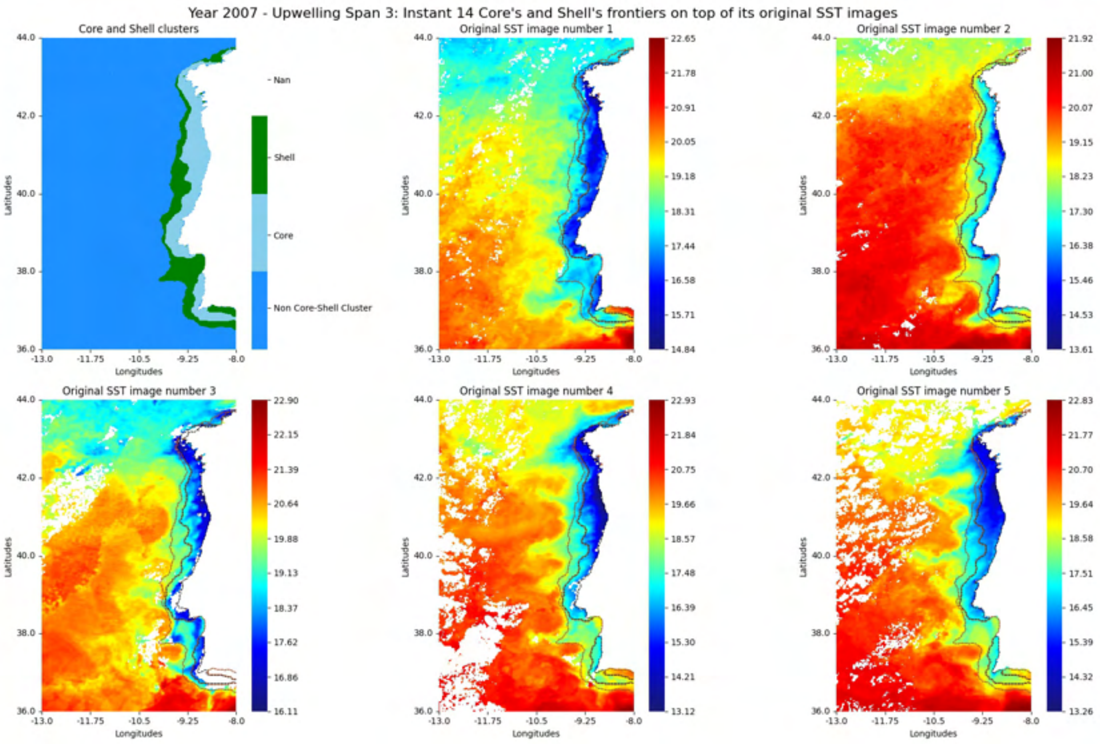


Figure I.94: Fronts regarding instant 14 of year 2007

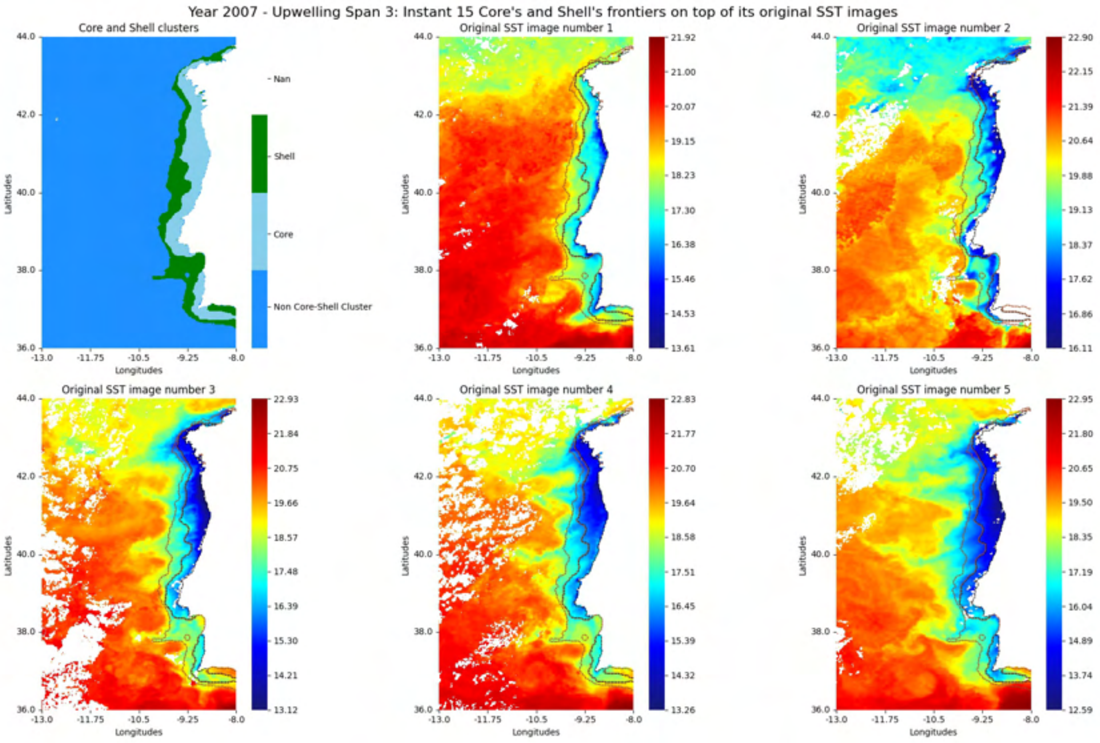


Figure I.95: Fronts regarding instant 15 of year 2007

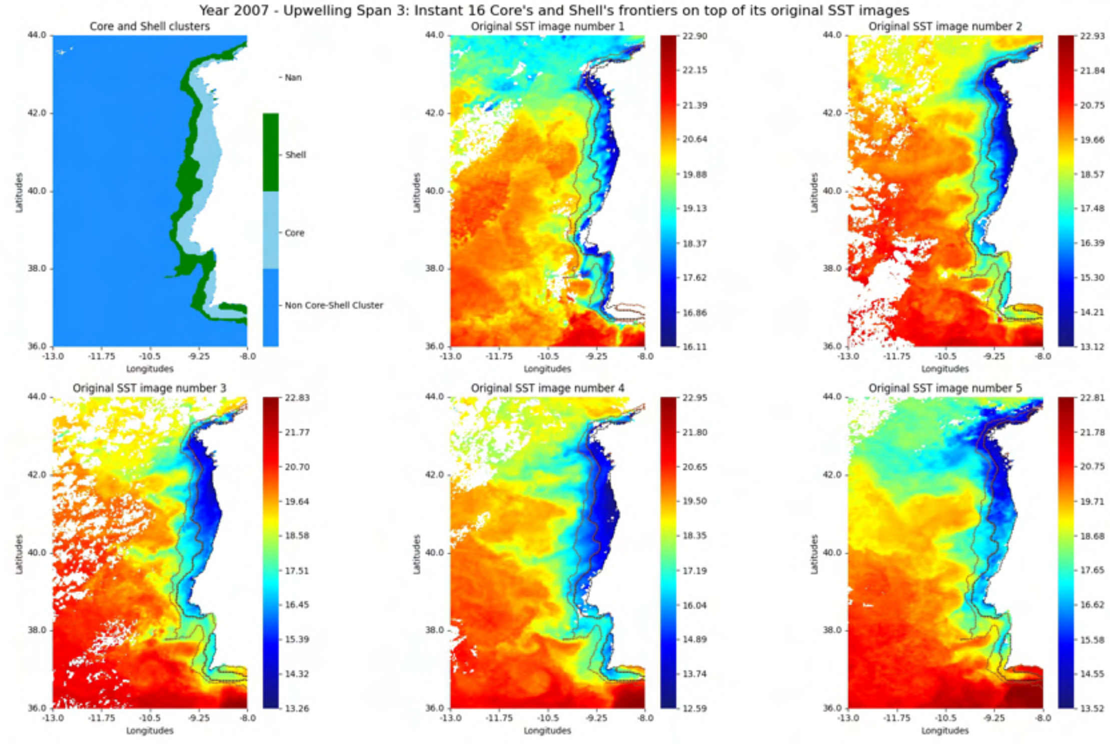


Figure I.96: Fronts regarding instant 16 of year 2007

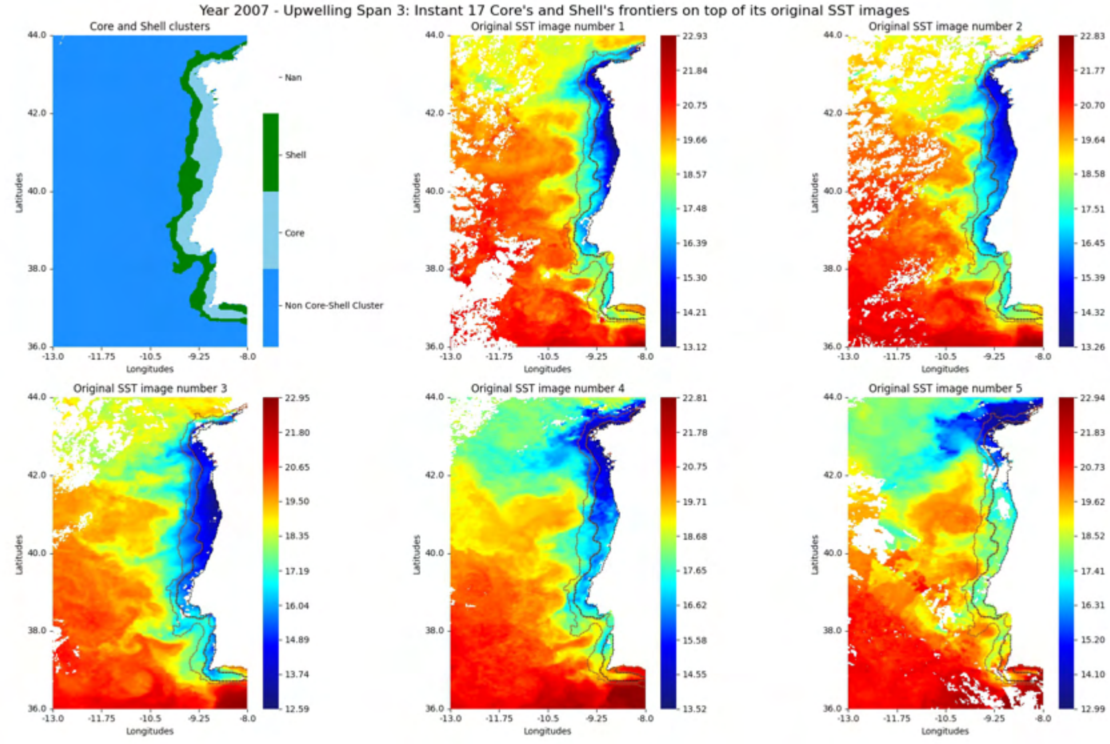


Figure I.97: Fronts regarding instant 17 of year 2007

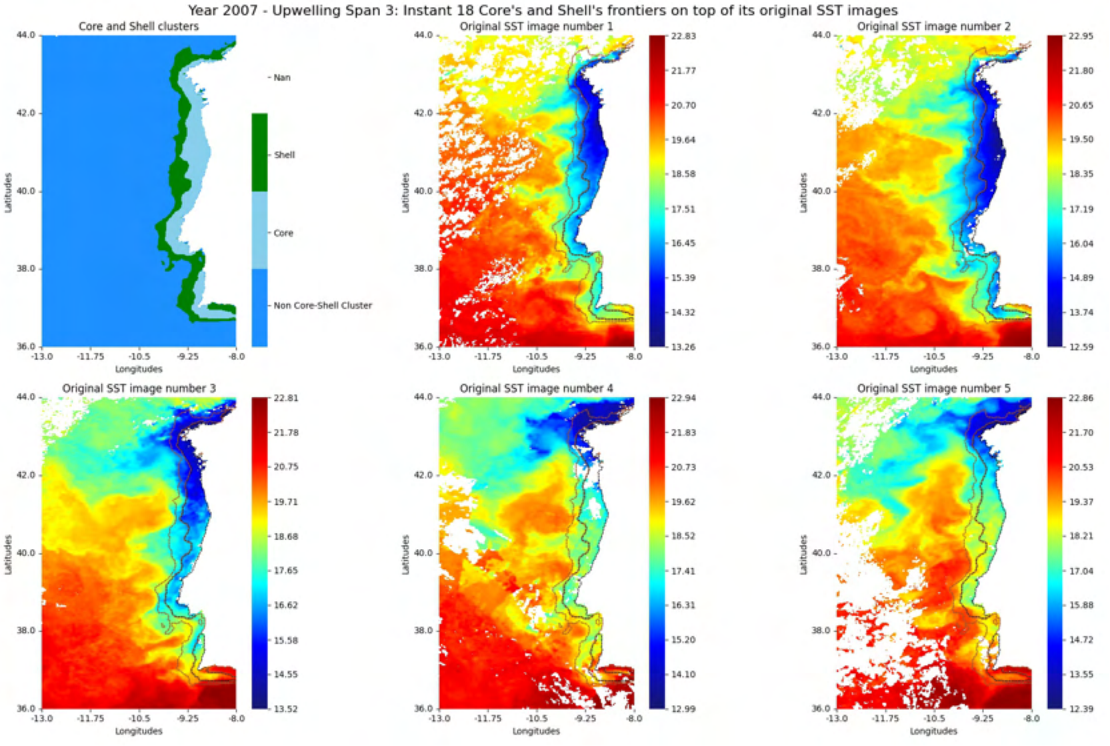


Figure I.98: Fronts regarding instant 18 of year 2007

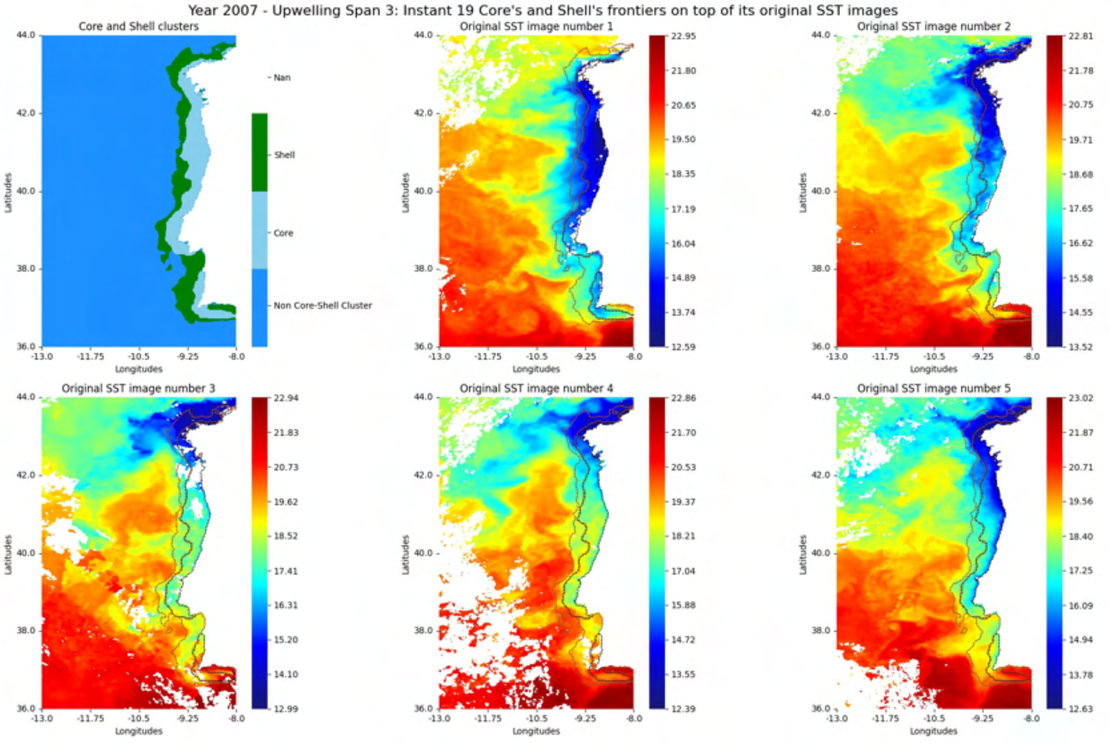


Figure I.99: Fronts regarding instant 19 of year 2007

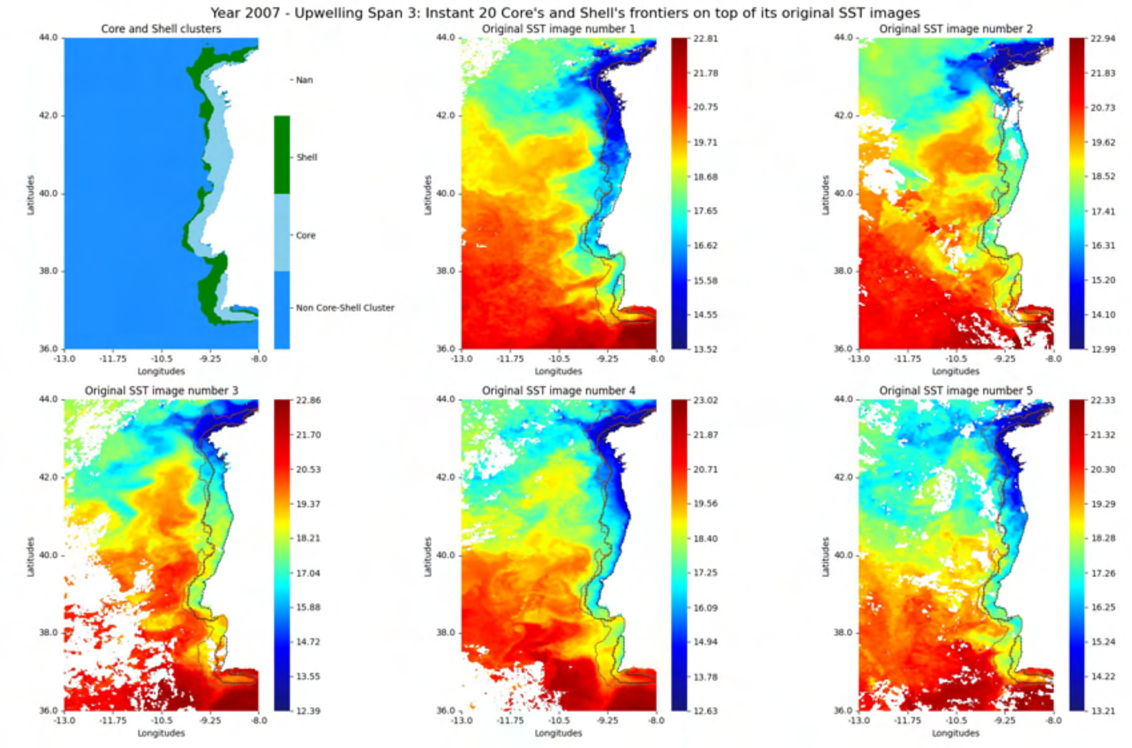


Figure I.100: Fronts regarding instant 20 of year 2007

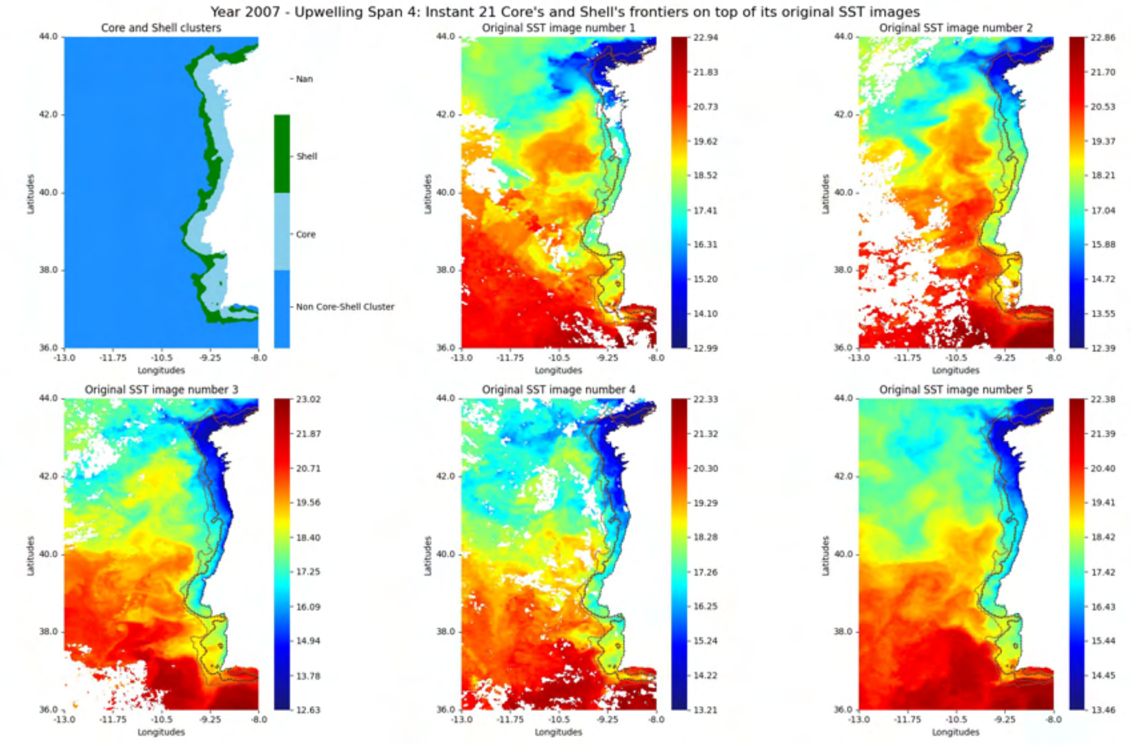


Figure I.101: Fronts regarding instant 21 of year 2007

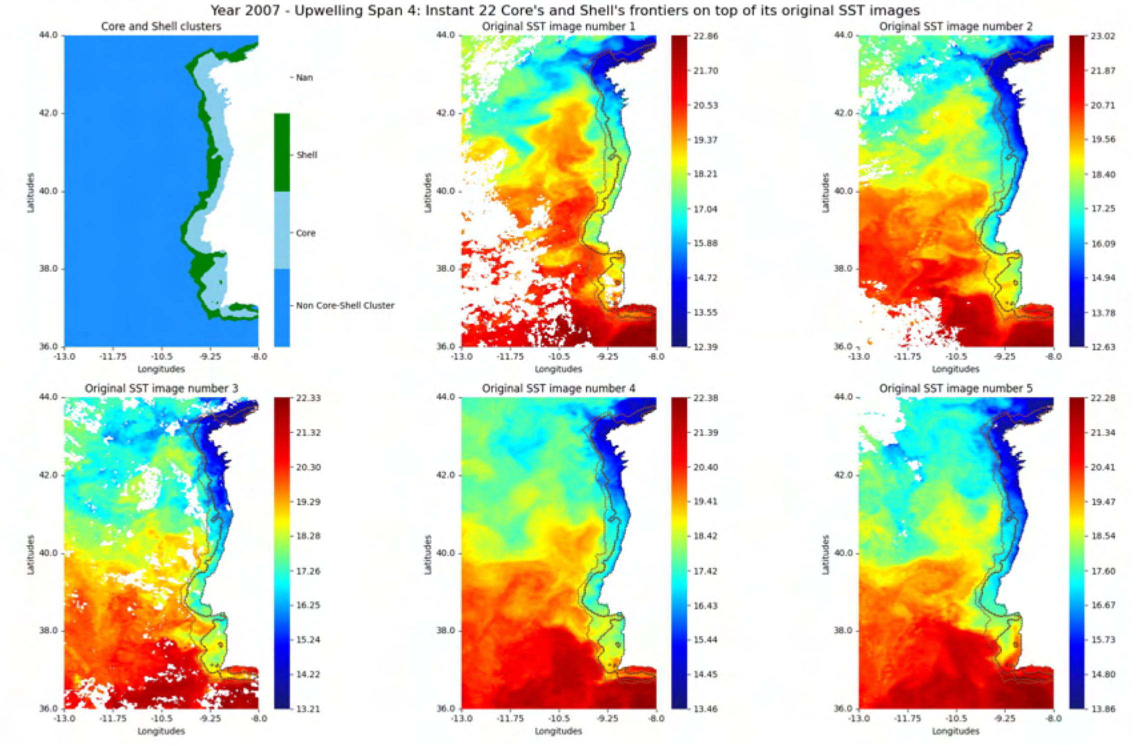


Figure I.102: Fronts regarding instant 22 of year 2007

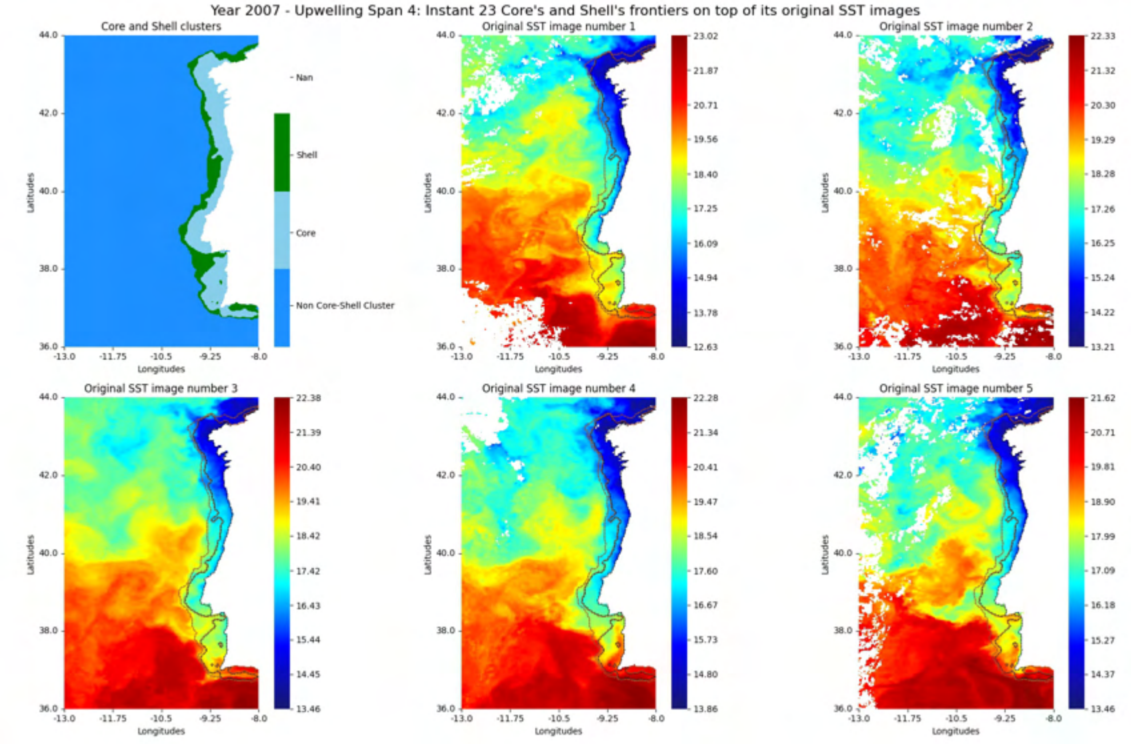


Figure I.103: Fronts regarding instant 23 of year 2007

I.3.2 Year 2015

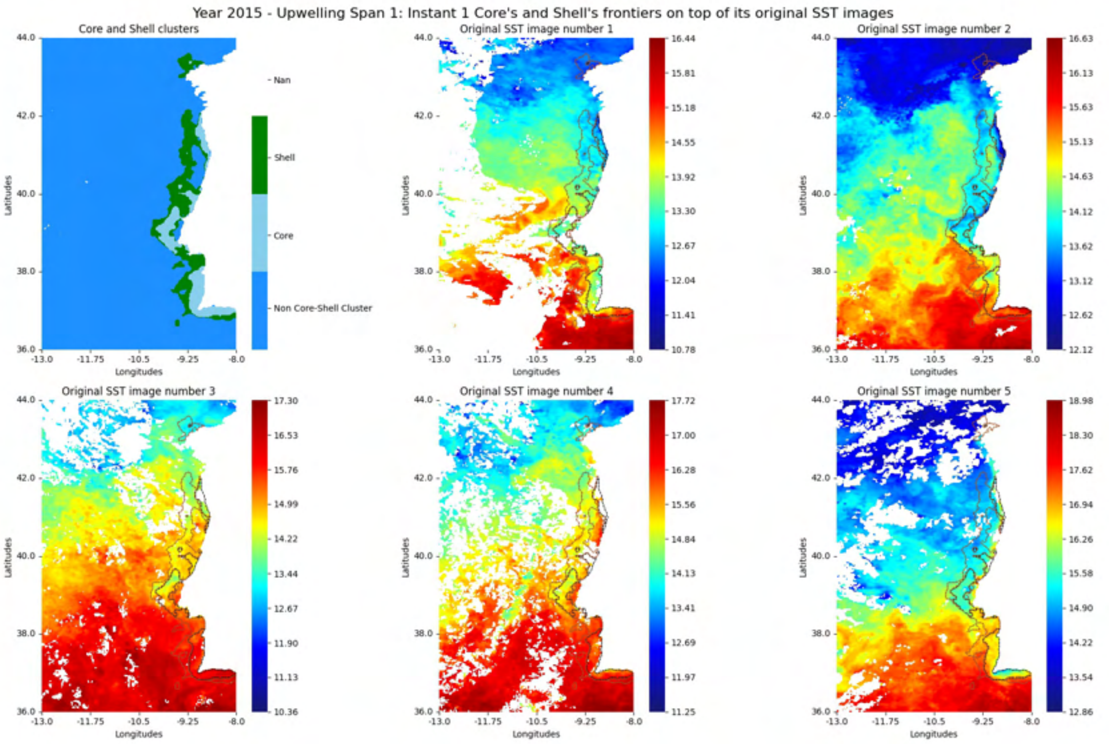


Figure I.104: Fronts regarding instant 1 of year 2015

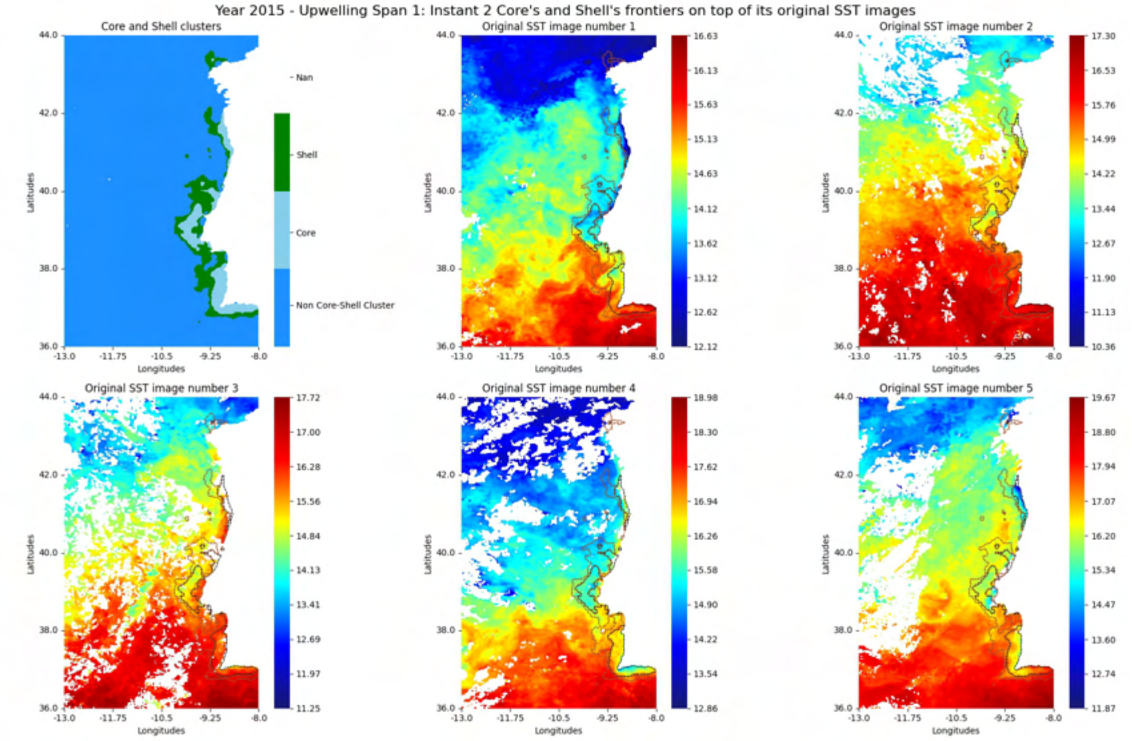


Figure I.105: Fronts regarding instant 2 of year 2015

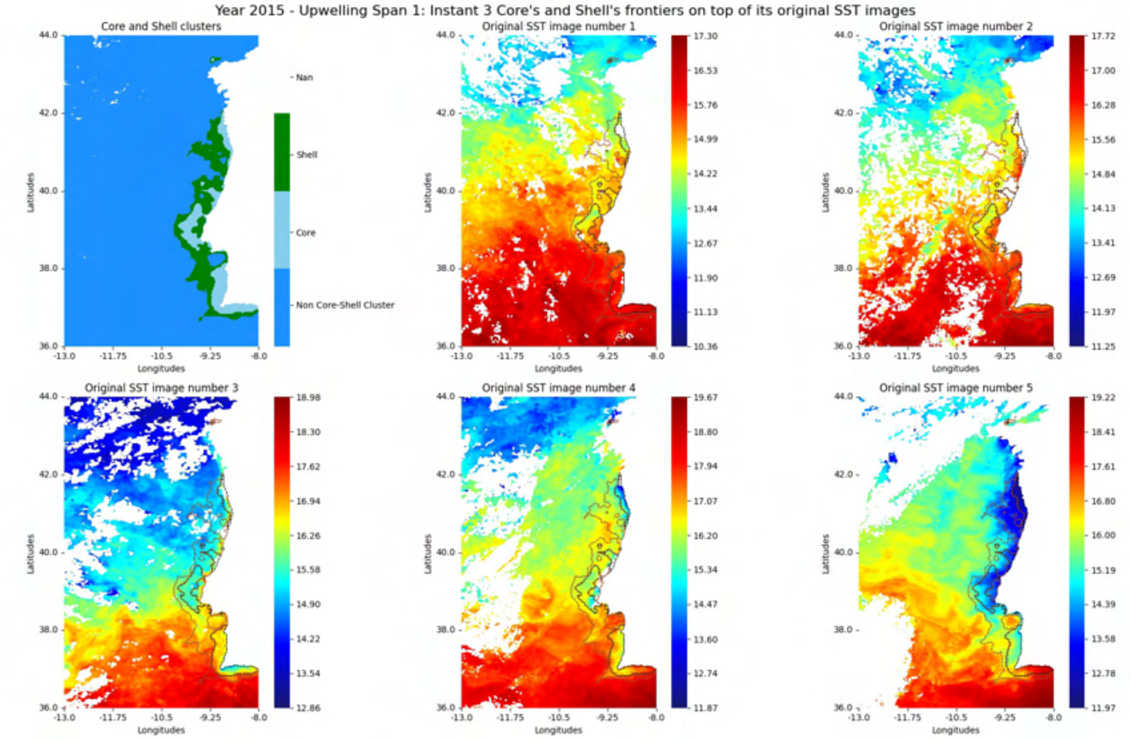


Figure I.106: Fronts regarding instant 3 of year 2015

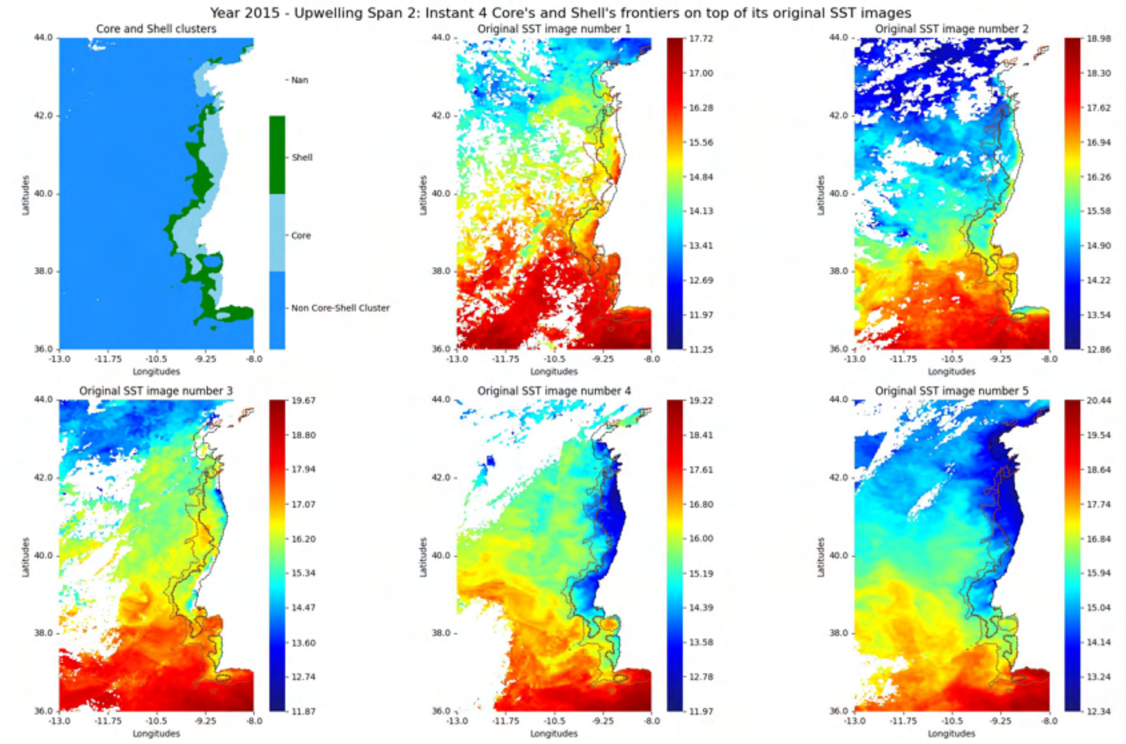


Figure I.107: Fronts regarding instant 4 of year 2015

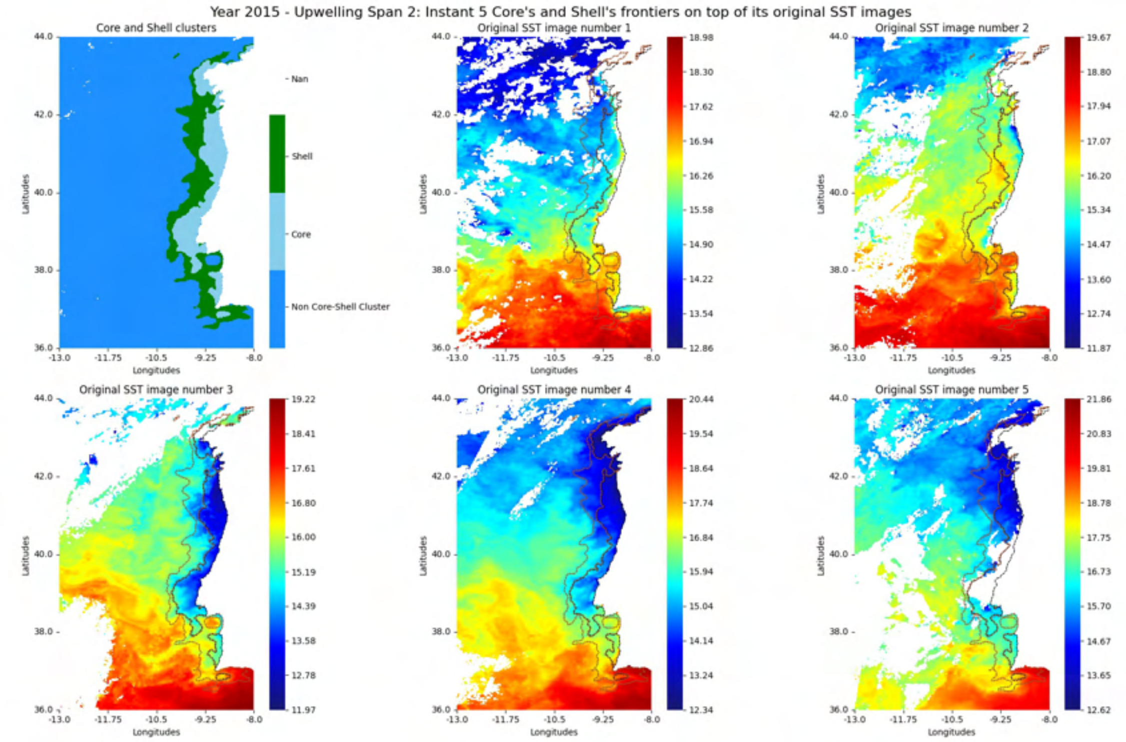


Figure I.108: Fronts regarding instant 5 of year 2015

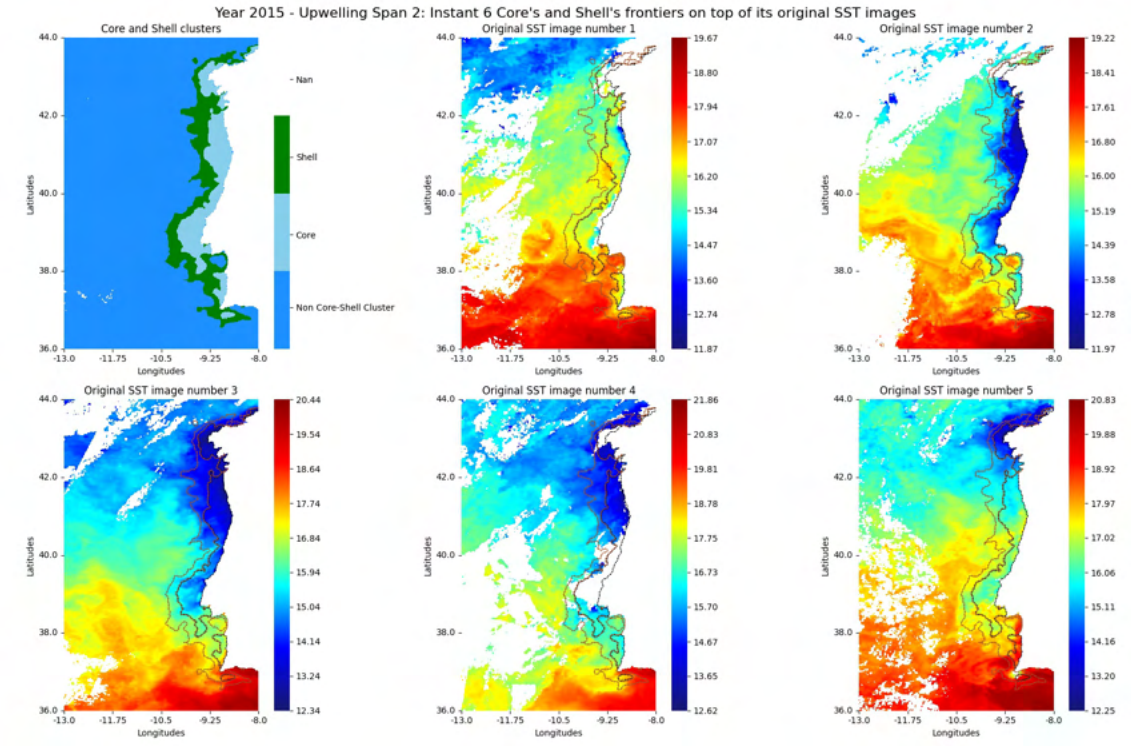


Figure I.109: Fronts regarding instant 6 of year 2015

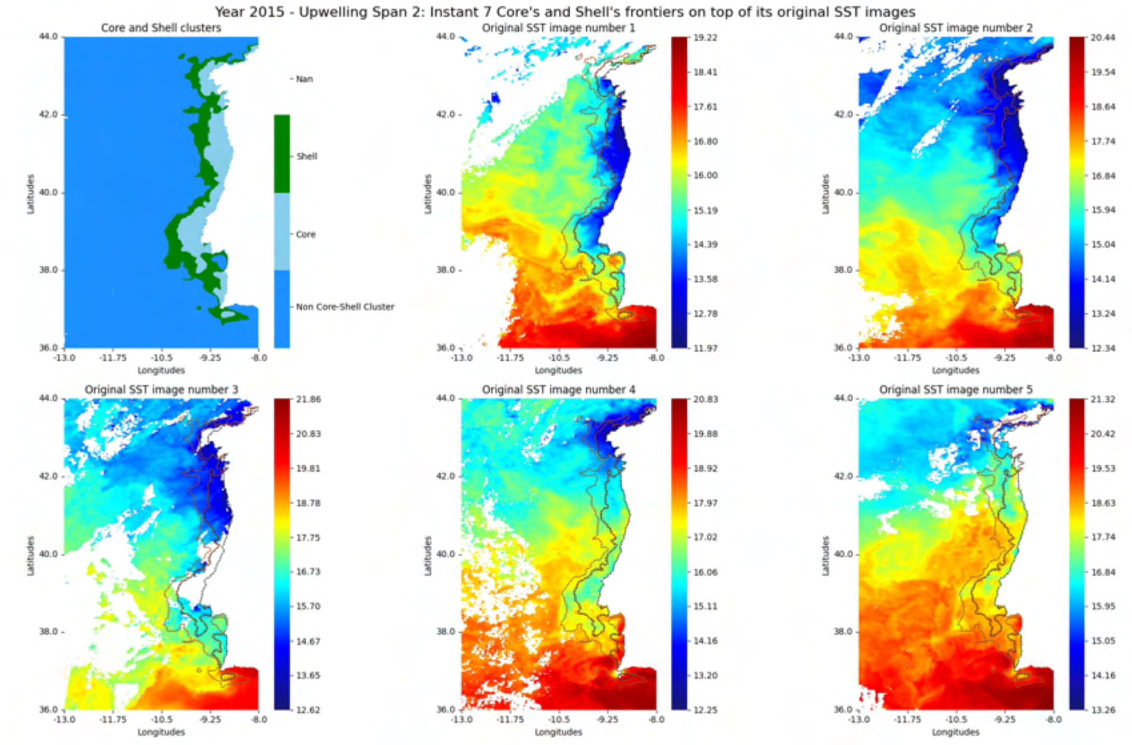


Figure I.110: Fronts regarding instant 7 of year 2015

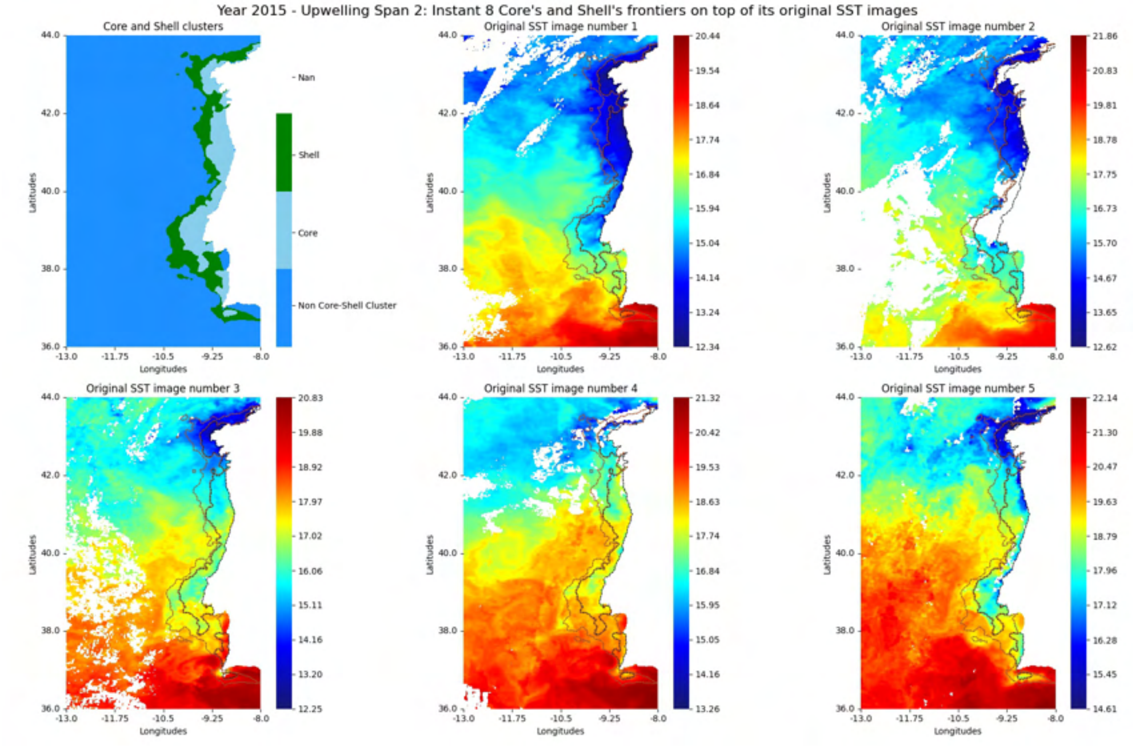


Figure I.111: Fronts regarding instant 8 of year 2015

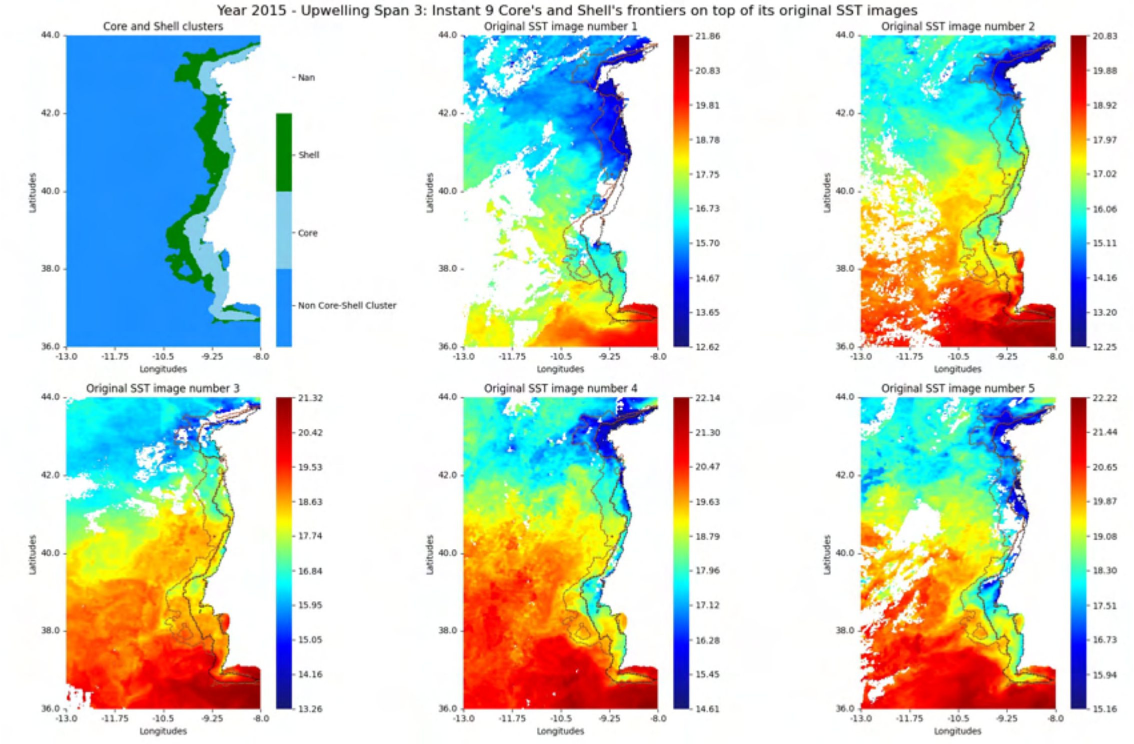


Figure I.112: Fronts regarding instant 9 of year 2015

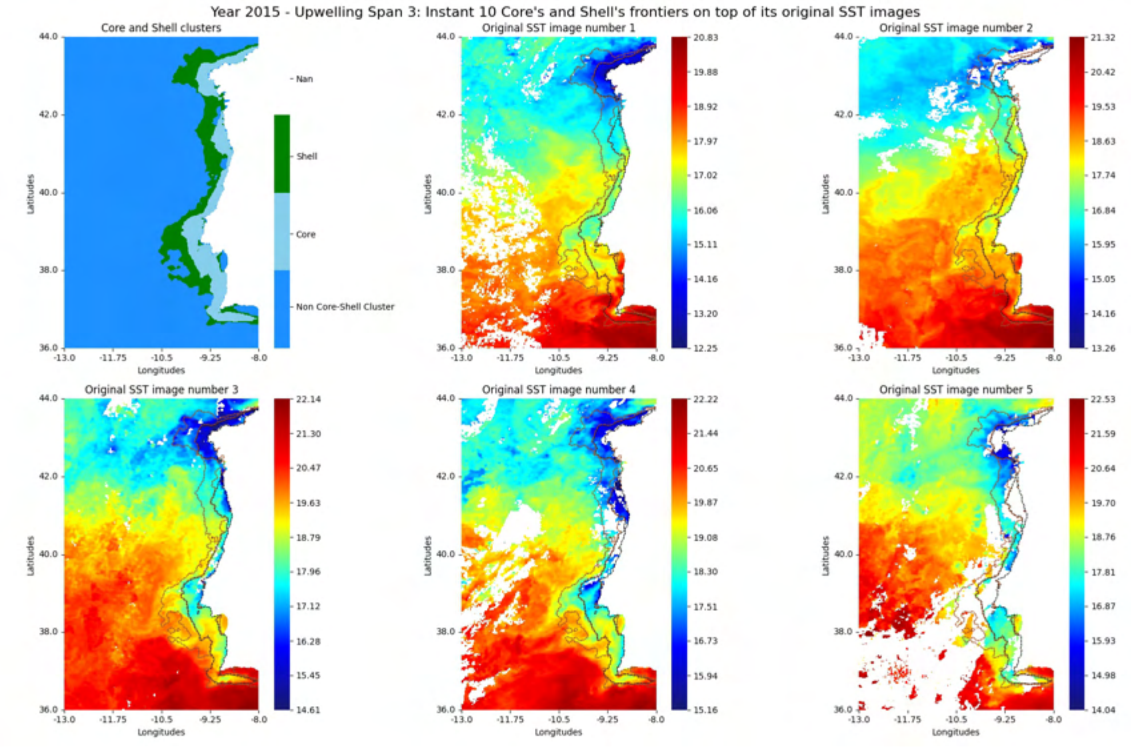


Figure I.113: Fronts regarding instant 10 of year 2015

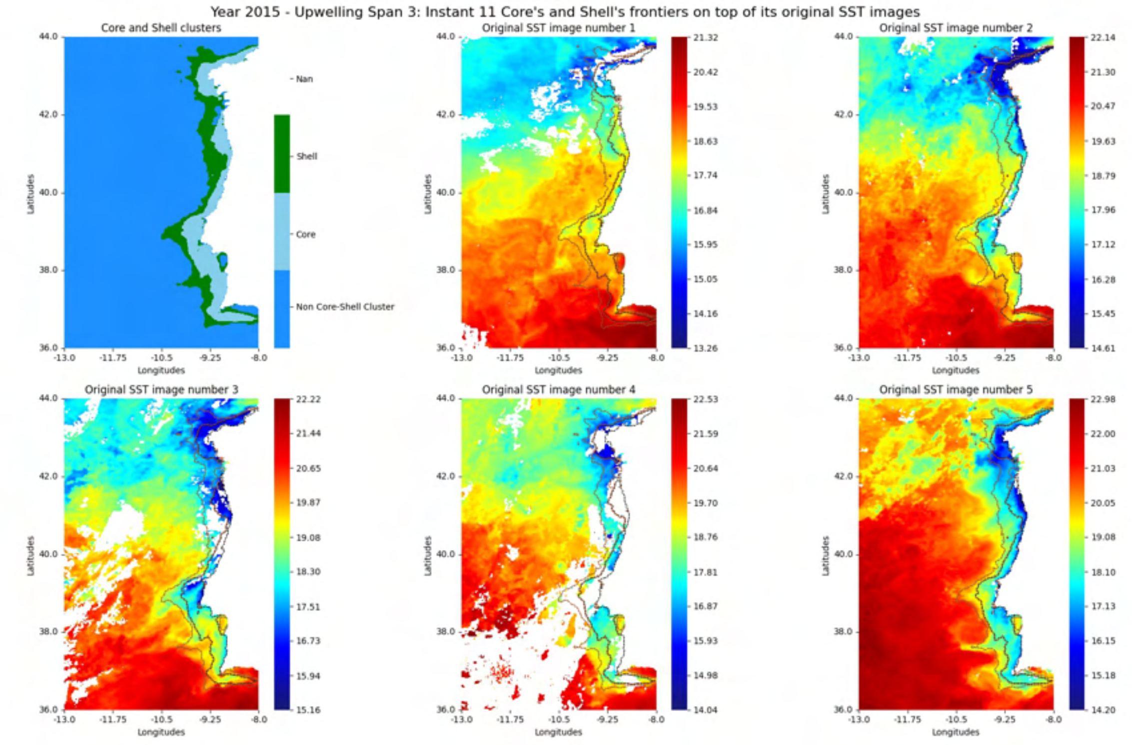


Figure I.114: Fronts regarding instant 11 of year 2015

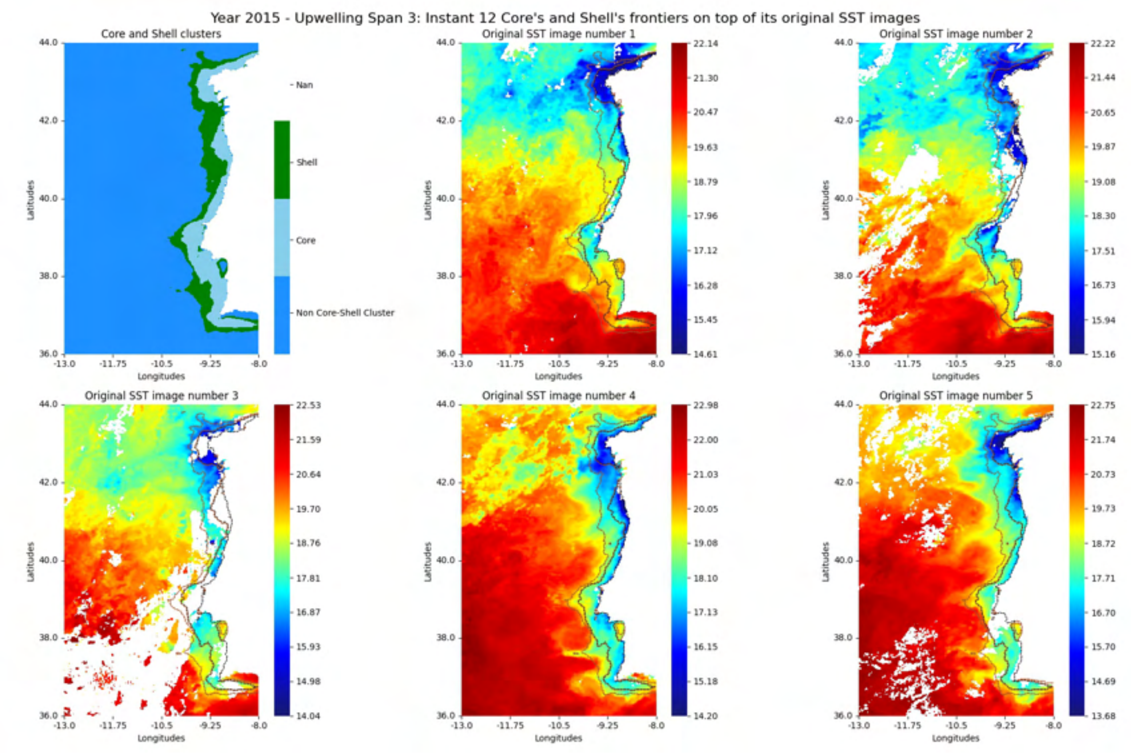


Figure I.115: Fronts regarding instant 12 of year 2015

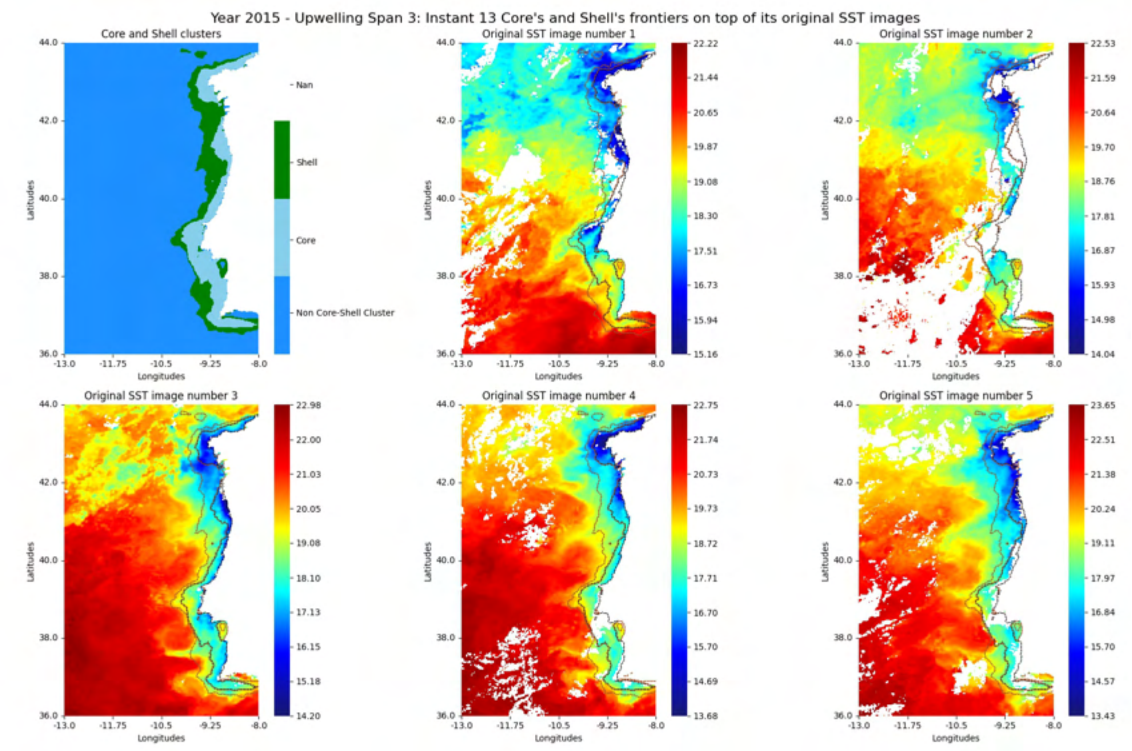


Figure I.116: Fronts regarding instant 13 of year 2015

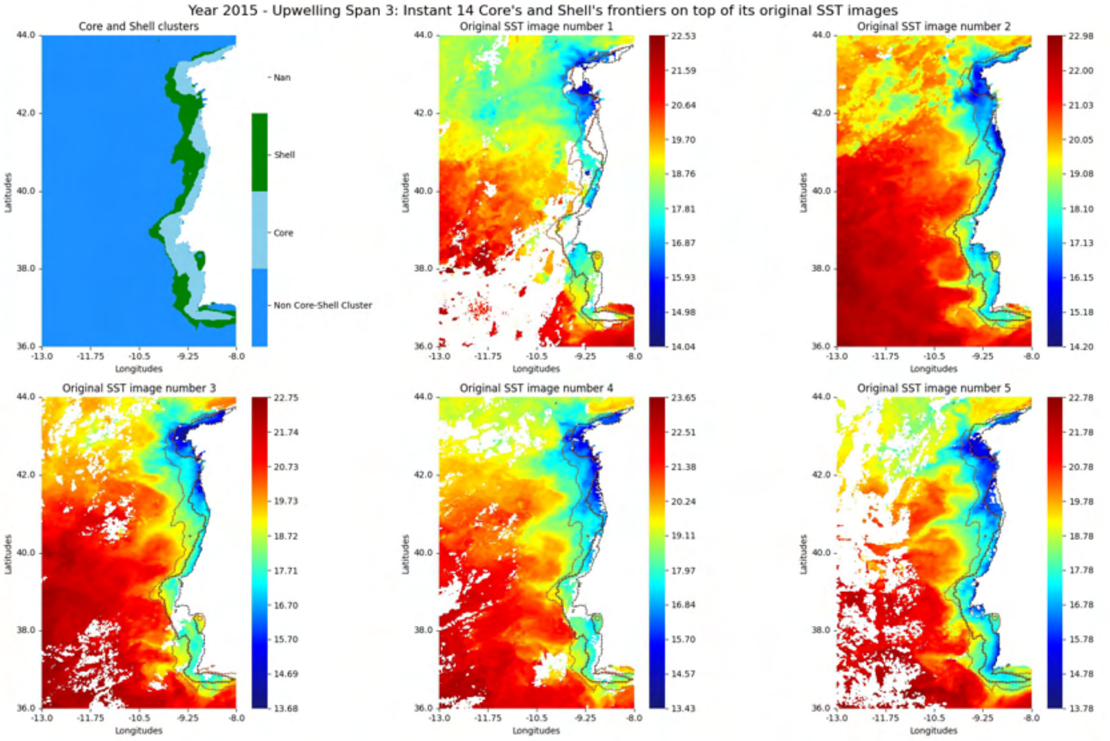


Figure I.117: Fronts regarding instant 14 of year 2015

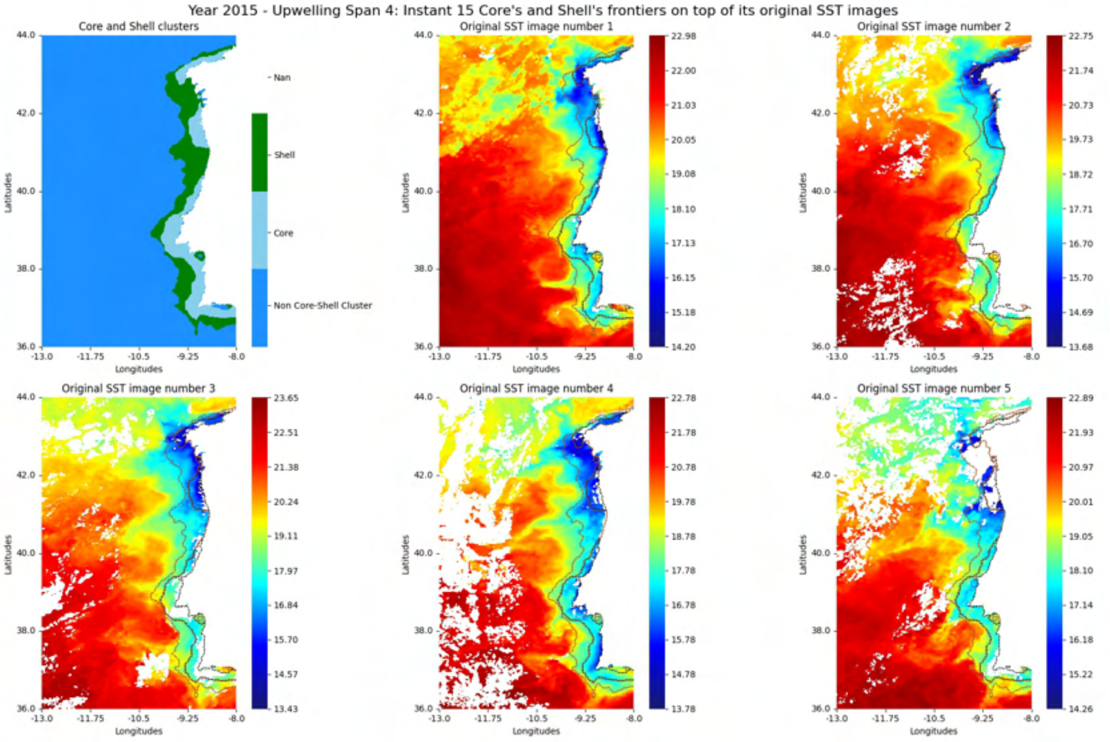


Figure I.118: Fronts regarding instant 15 of year 2015

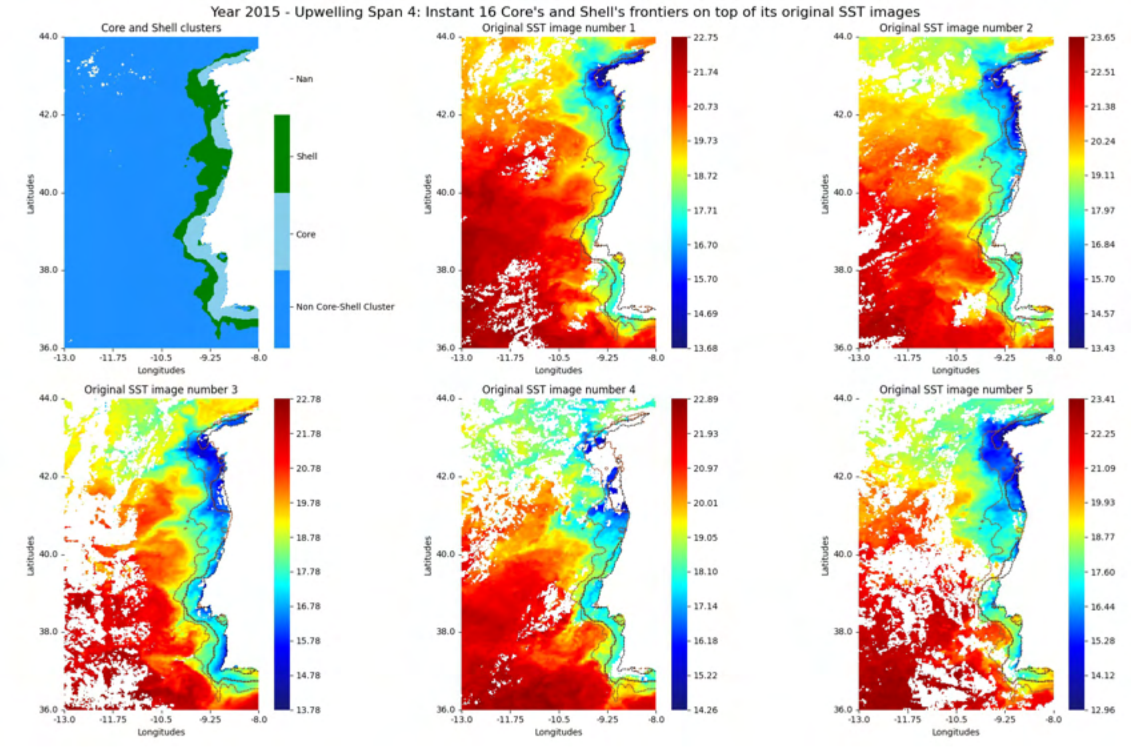


Figure I.119: Fronts regarding instant 16 of year 2015

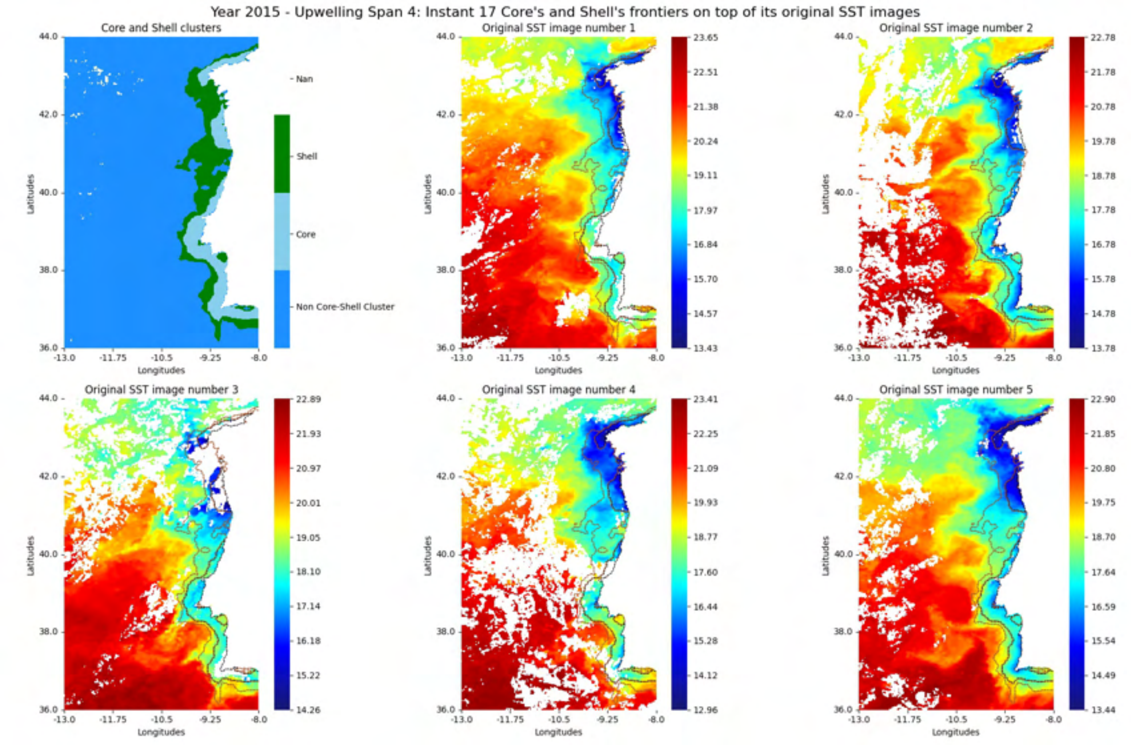


Figure I.120: Fronts regarding instant 17 of year 2015

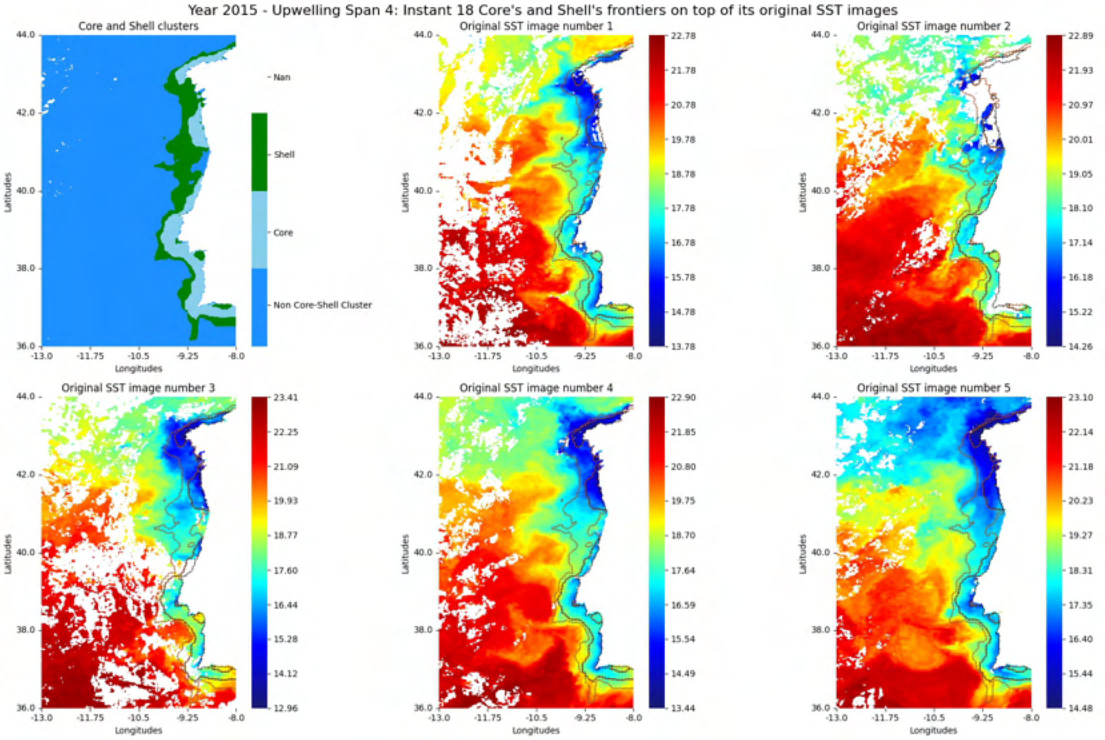


Figure I.121: Fronts regarding instant 18 of year 2015

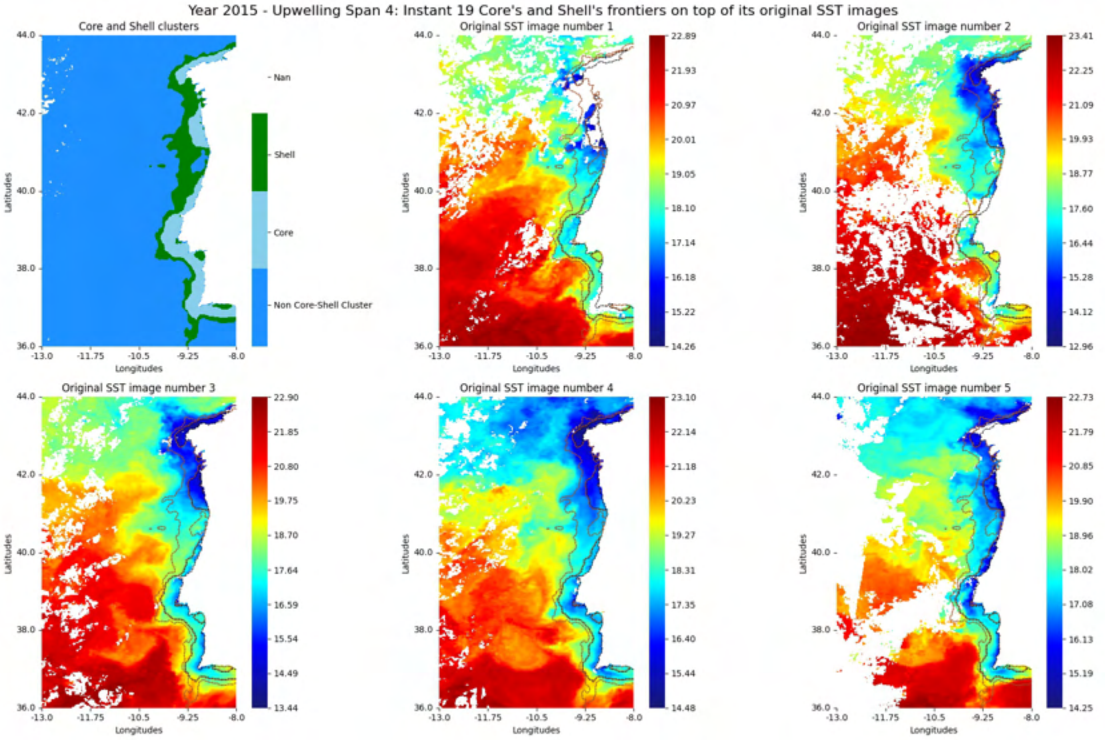
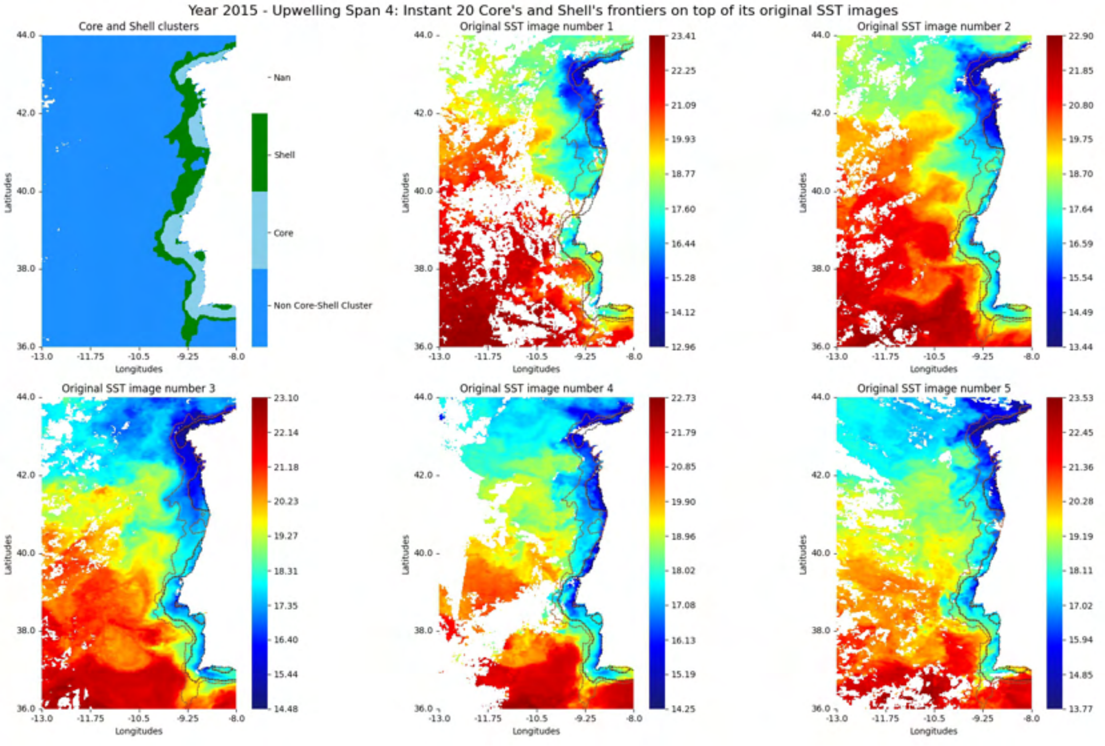


Figure I.122: Fronts regarding instant 19 of year 2015



Year 2015 - Upwelling Span 4: Instant 20 Core's and Shell's frontiers on top of its original SST images

Core and Shell clusters

Original SST image number 1

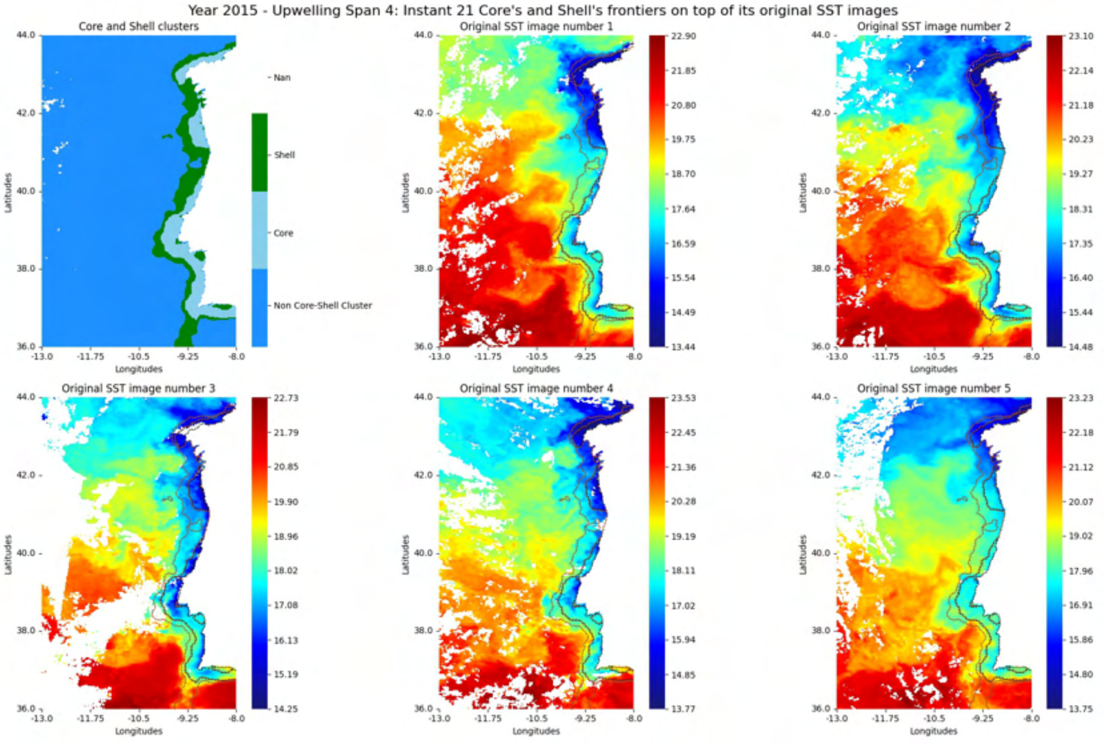
Original SST image number 2

Original SST image number 3

Original SST image number 4

Original SST image number 5

Figure I.123: Fronts regarding instant 20 of year 2015



Year 2015 - Upwelling Span 4: Instant 21 Core's and Shell's frontiers on top of its original SST images

Core and Shell clusters

Original SST image number 1

Original SST image number 2

Original SST image number 3

Original SST image number 4

Original SST image number 5

Figure I.124: Fronts regarding instant 21 of year 2015

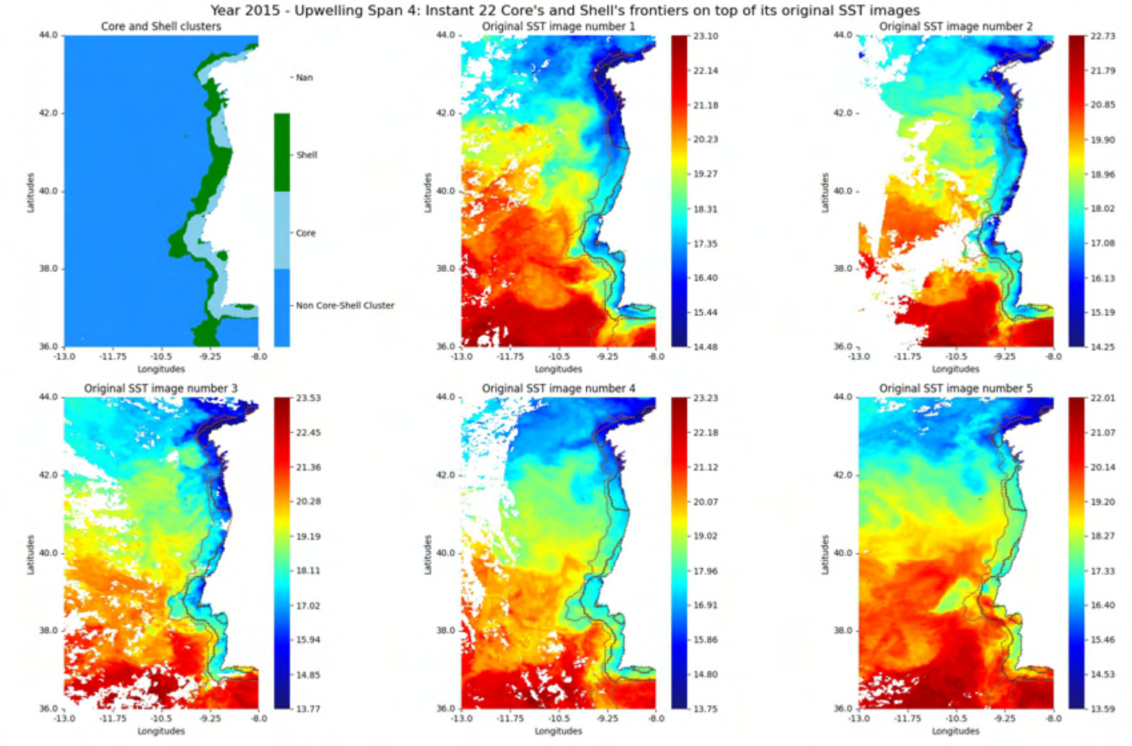


Figure I.125: Fronts regarding instant 22 of year 2015

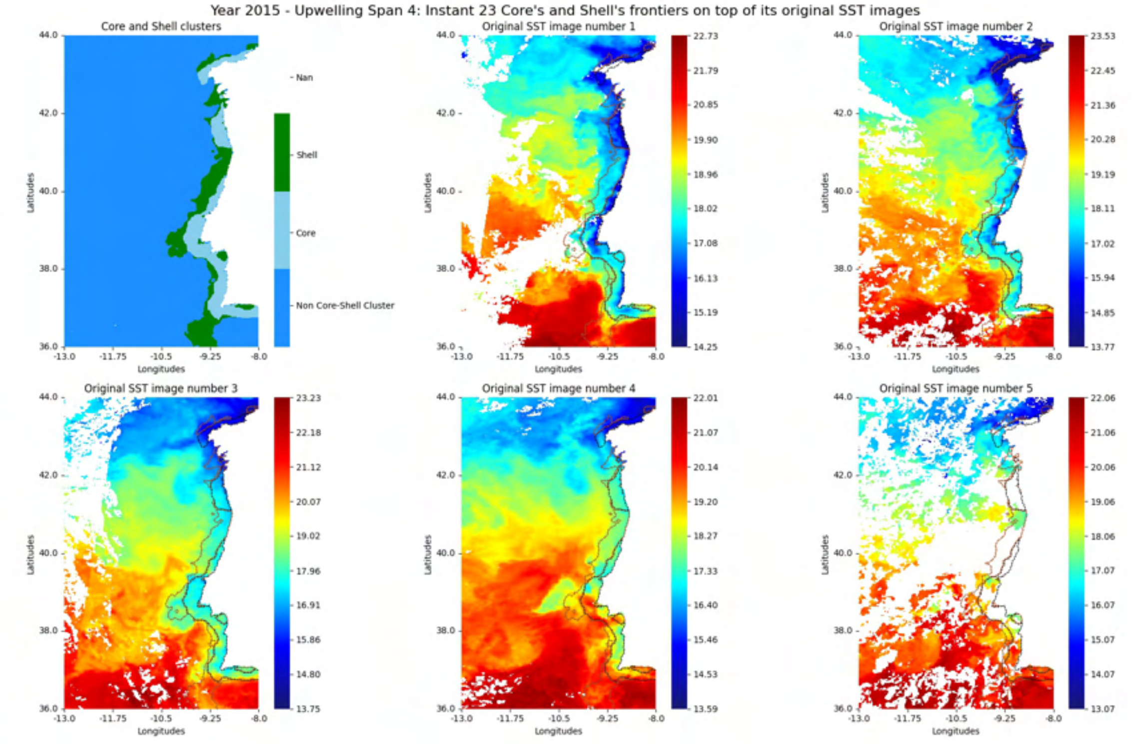


Figure I.126: Fronts regarding instant 23 of year 2015

I.3.3 Year 2019

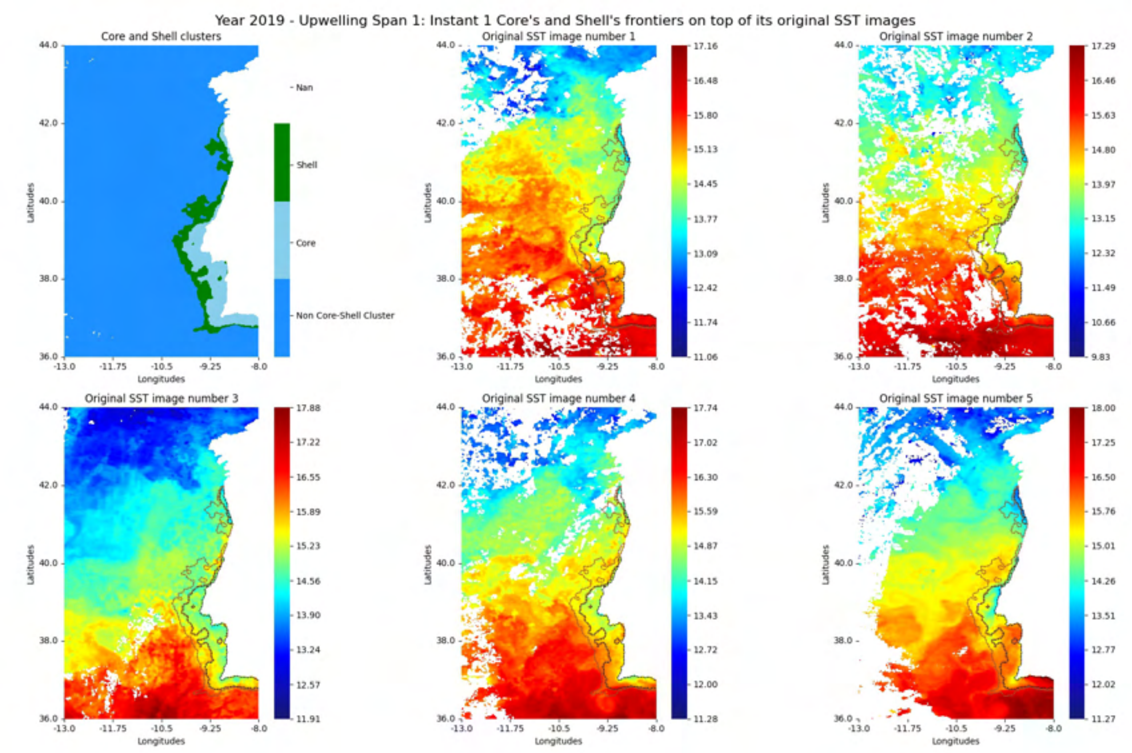


Figure I.127: Fronts regarding instant 1 of year 2019

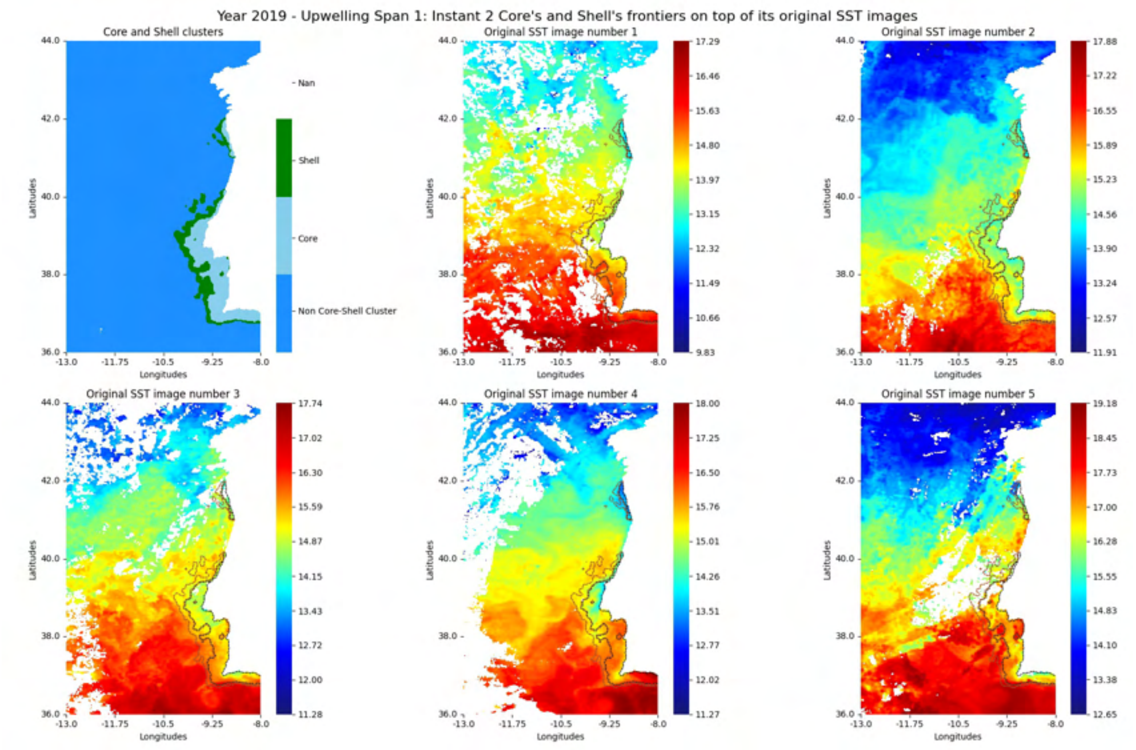


Figure I.128: Fronts regarding instant 2 of year 2019

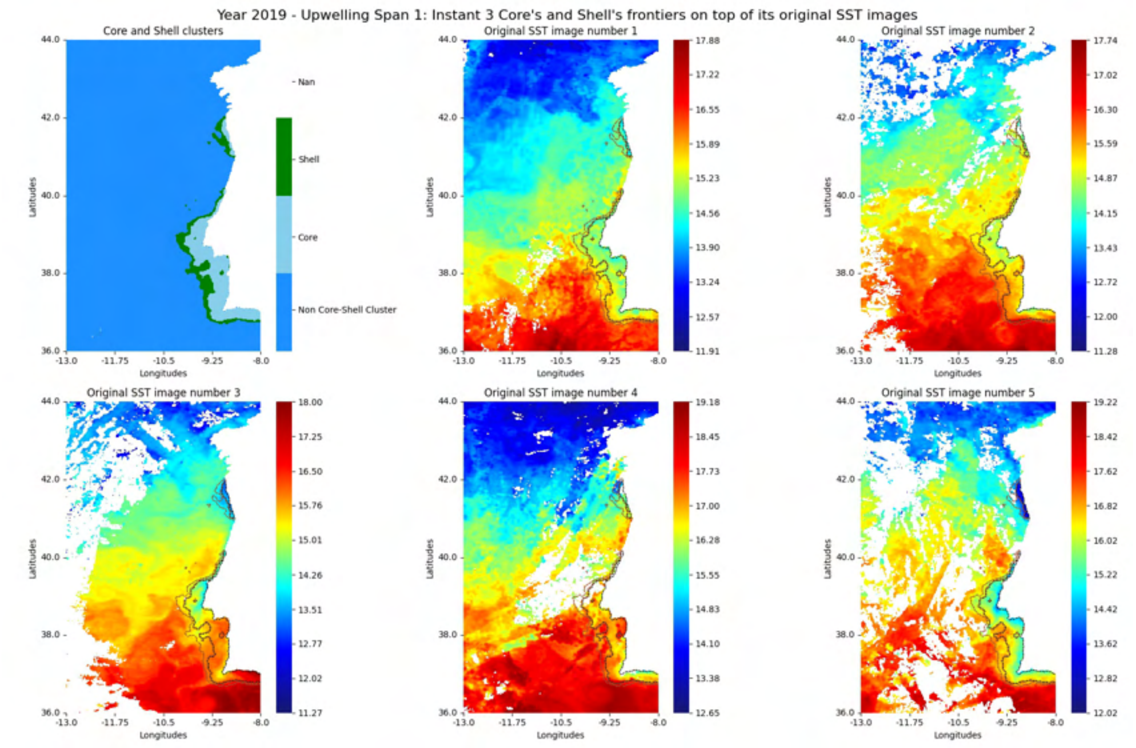
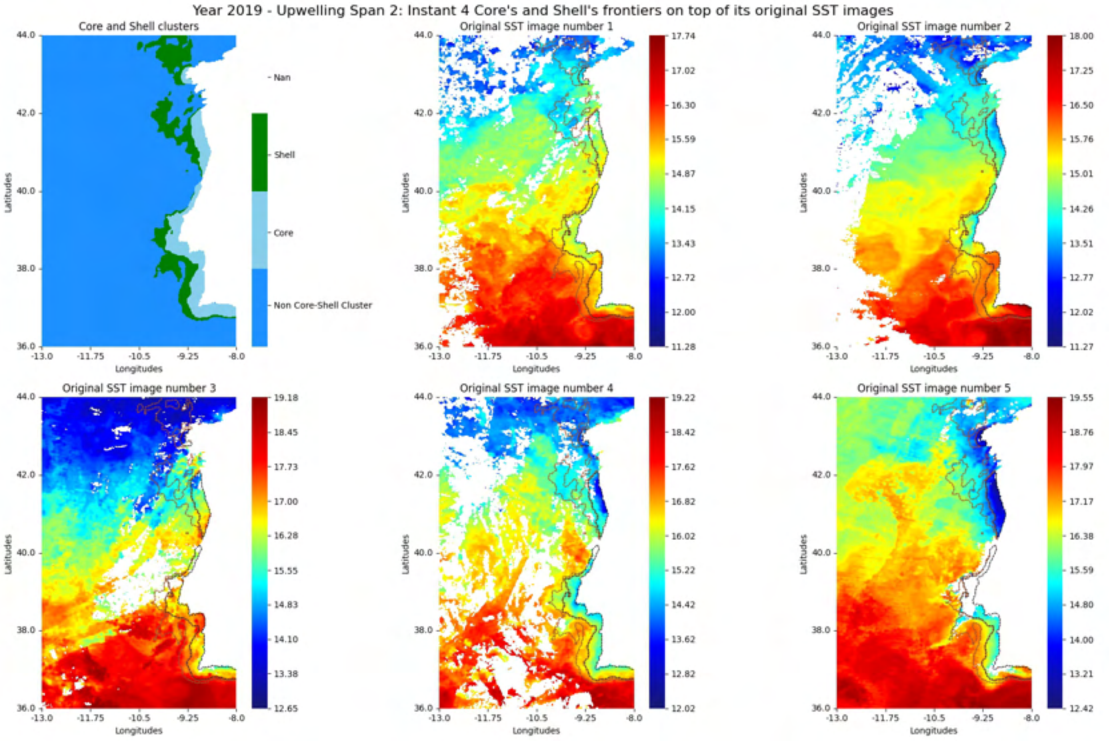


Figure I.129: Fronts regarding instant 3 of year 2019


Year 2019 - Upwelling Span 2: Instant 4 Core's and Shell's frontiers on top of its original SST images

Core and Shell clusters

Original SST image number 1

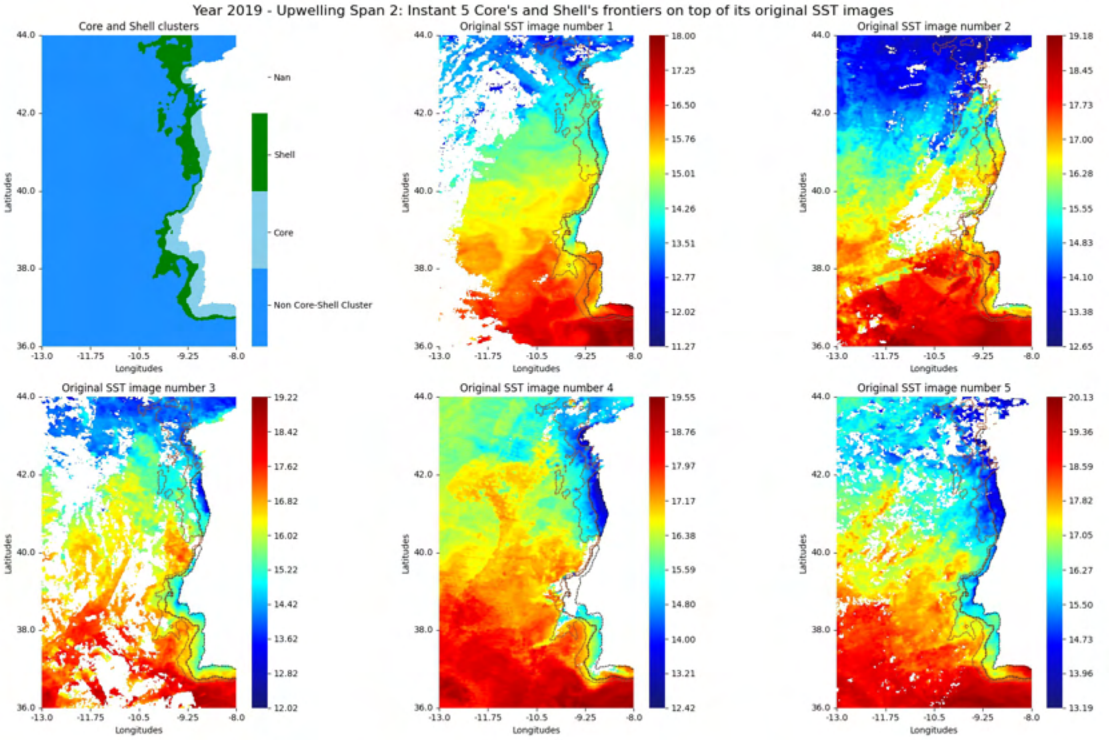
Original SST image number 2

Original SST image number 3

Original SST image number 4

Original SST image number 5

Figure I.130: Fronts regarding instant 4 of year 2019


Year 2019 - Upwelling Span 2: Instant 5 Core's and Shell's frontiers on top of its original SST images

Core and Shell clusters

Original SST image number 1

Original SST image number 2

Original SST image number 3

Original SST image number 4

Original SST image number 5

Figure I.131: Fronts regarding instant 5 of year 2019

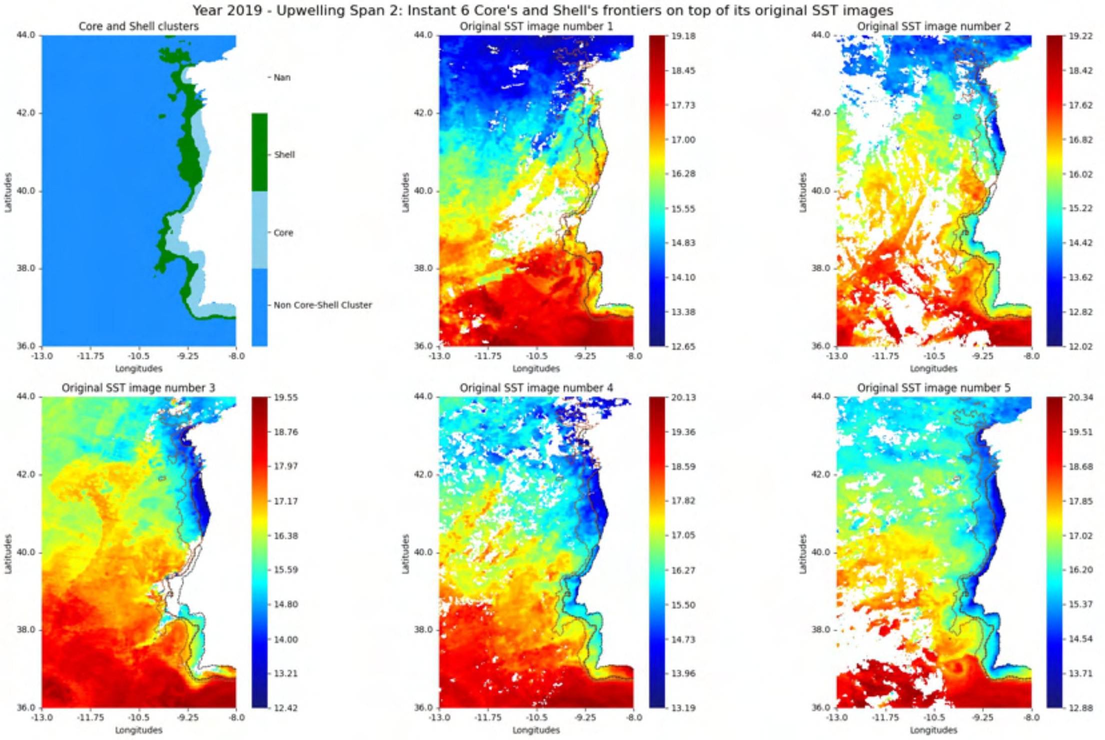


Figure I.132: Fronts regarding instant 6 of year 2019

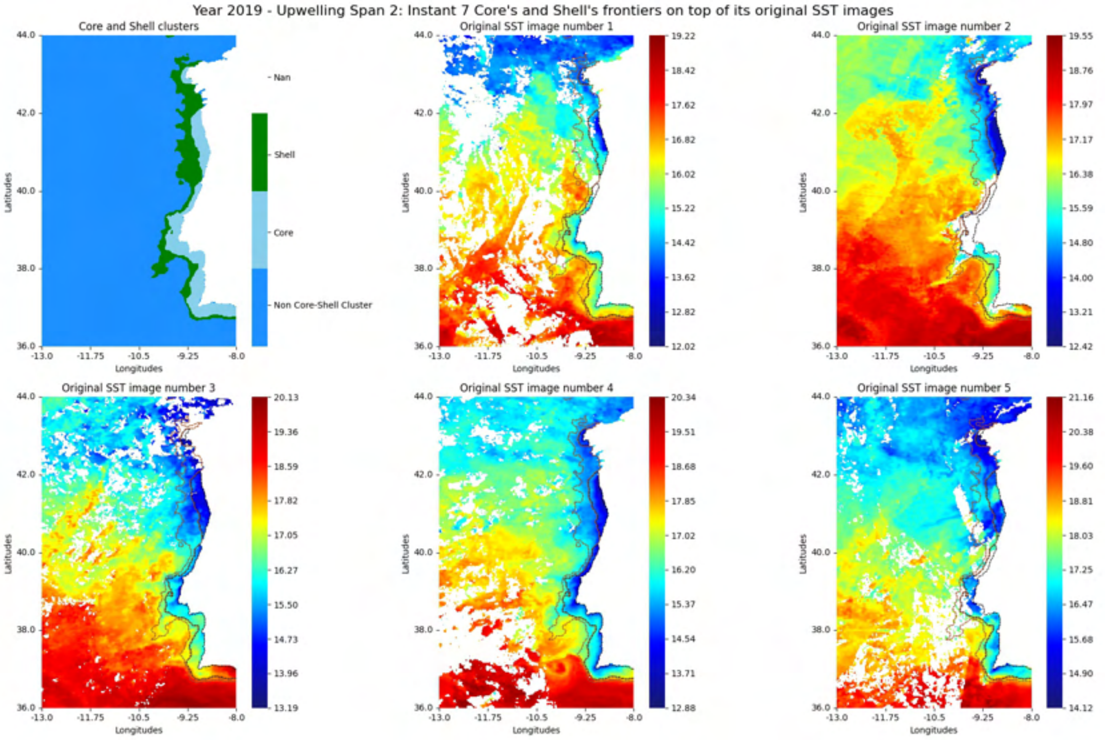


Figure I.133: Fronts regarding instant 7 of year 2019

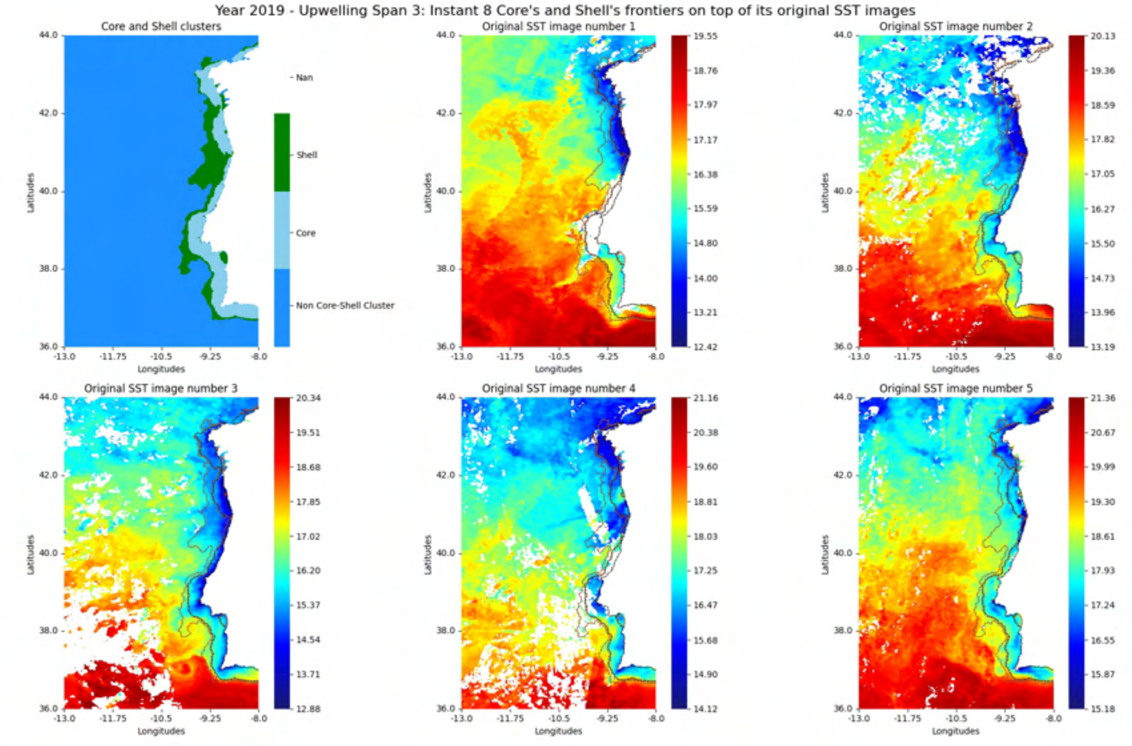


Figure I.134: Fronts regarding instant 8 of year 2019

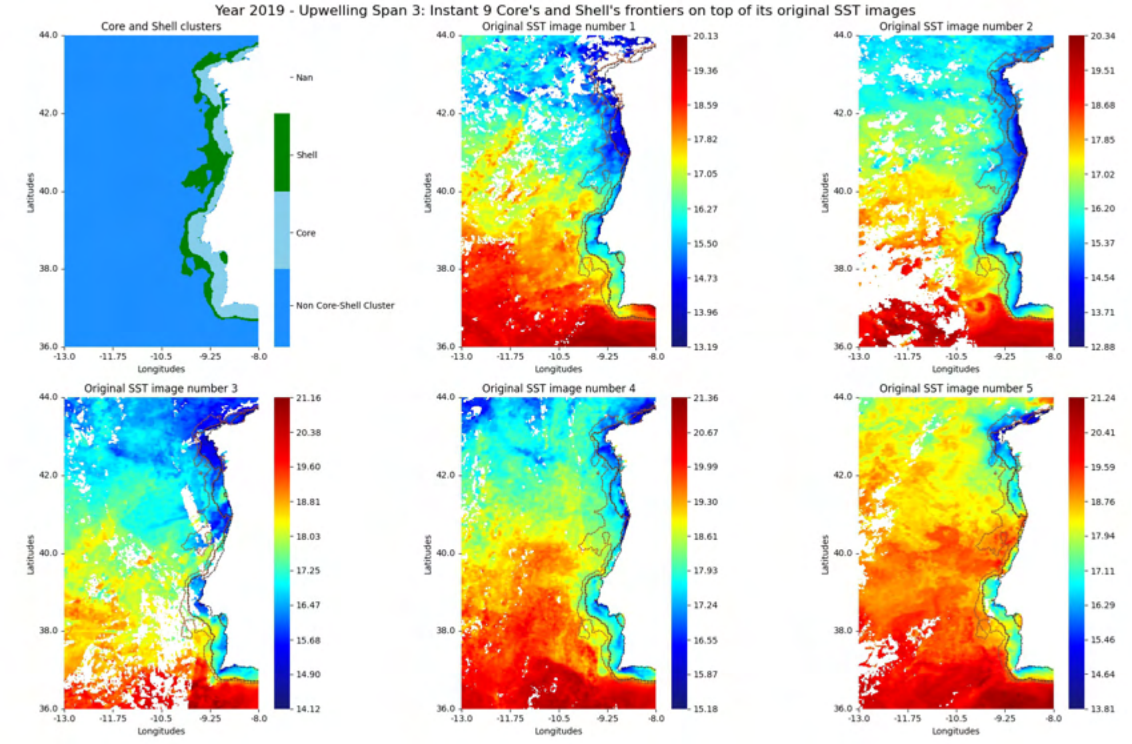
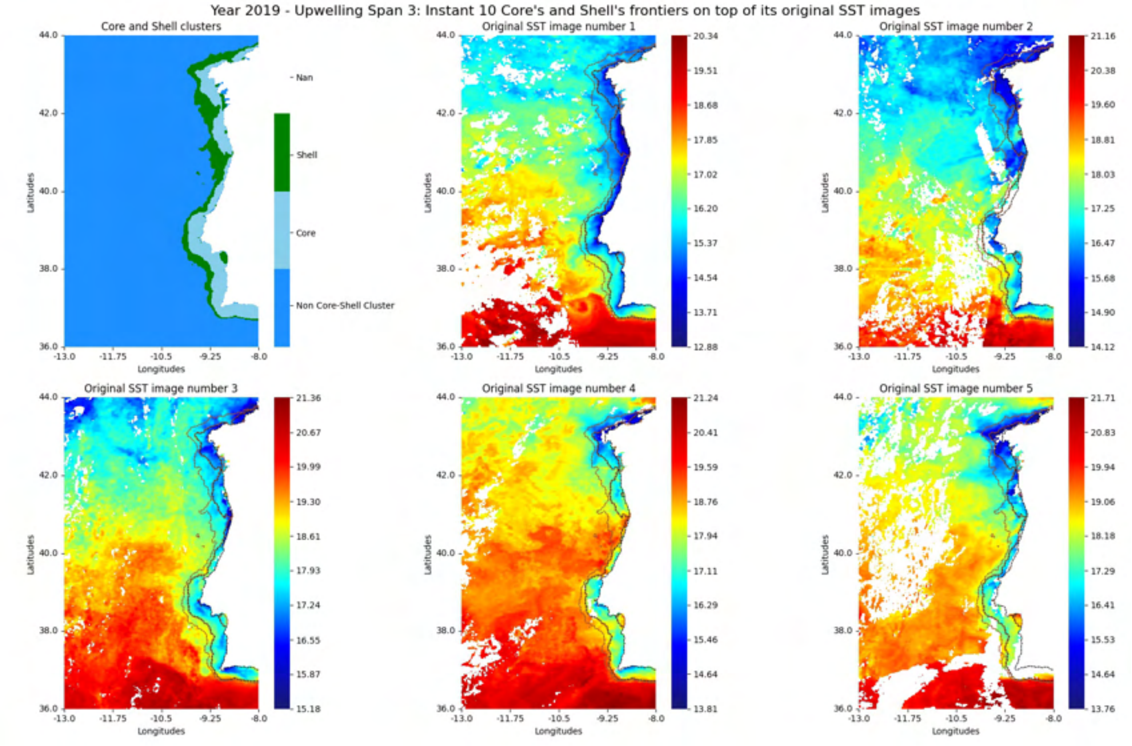


Figure I.135: Fronts regarding instant 9 of year 2019



Year 2019 - Upwelling Span 3: Instant 10 Core's and Shell's frontiers on top of its original SST images

Core and Shell clusters

Original SST image number 1

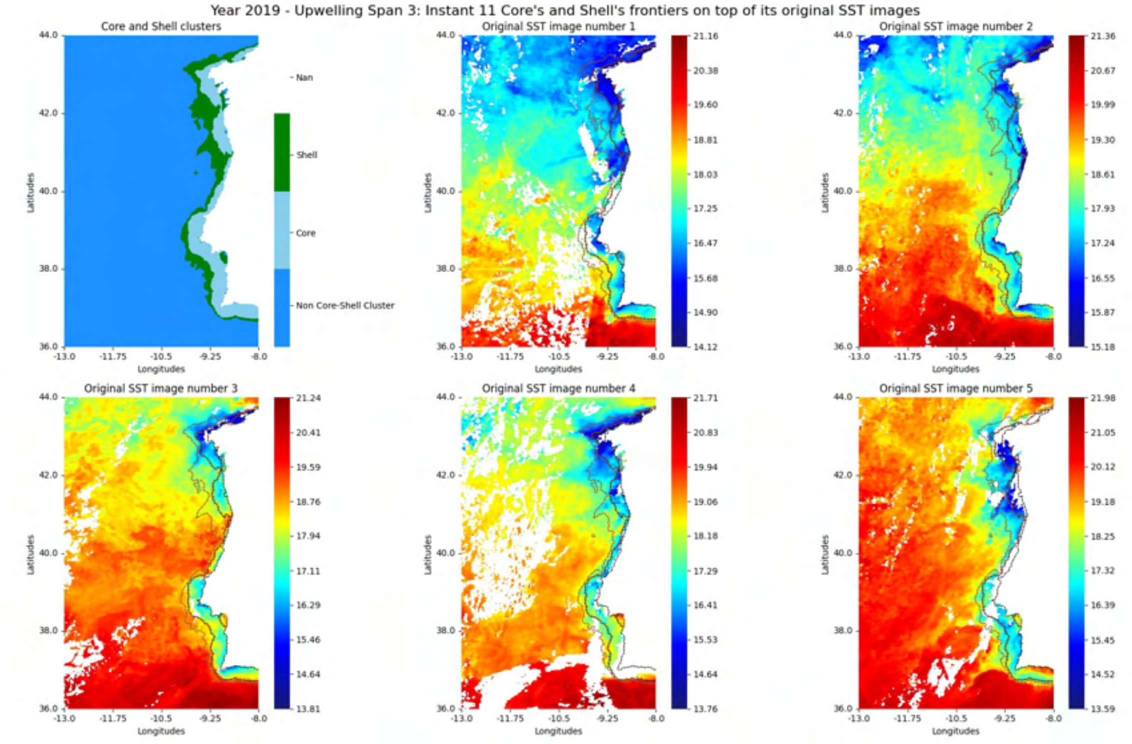
Original SST image number 2

Original SST image number 3

Original SST image number 4

Original SST image number 5

Figure I.136: Fronts regarding instant 10 of year 2019



Year 2019 - Upwelling Span 3: Instant 11 Core's and Shell's frontiers on top of its original SST images

Core and Shell clusters

Original SST image number 1

Original SST image number 2

Original SST image number 3

Original SST image number 4

Original SST image number 5

Figure I.137: Fronts regarding instant 11 of year 2019

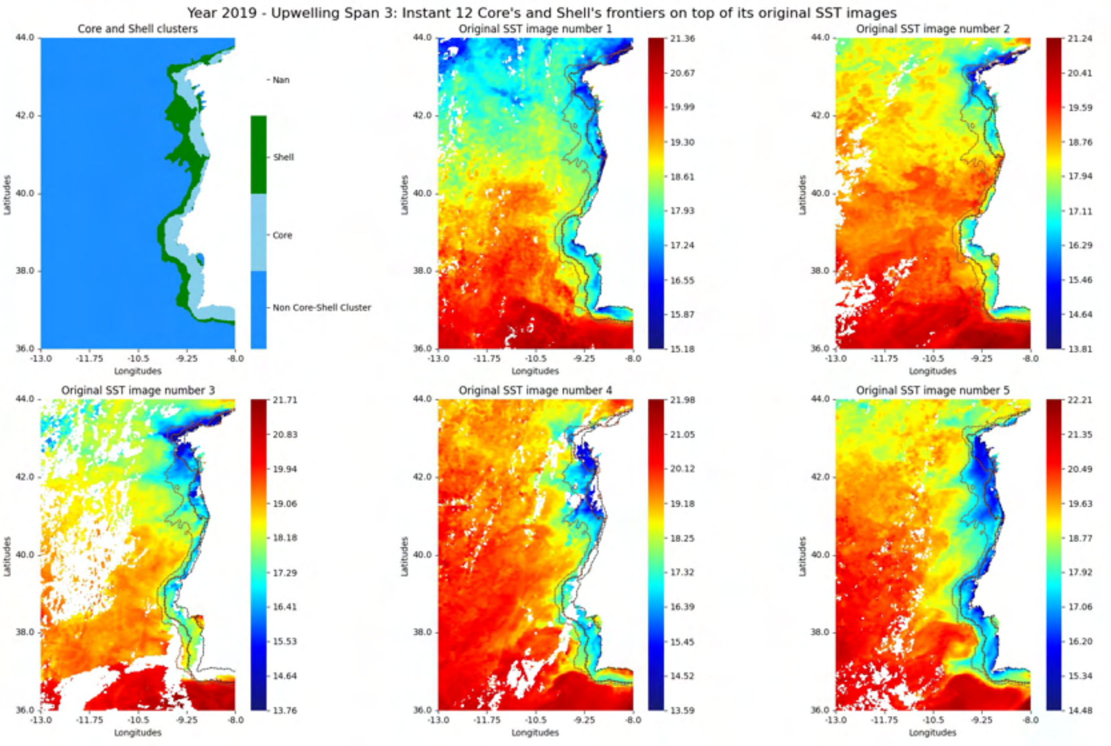


Figure I.138: Fronts regarding instant 12 of year 2019

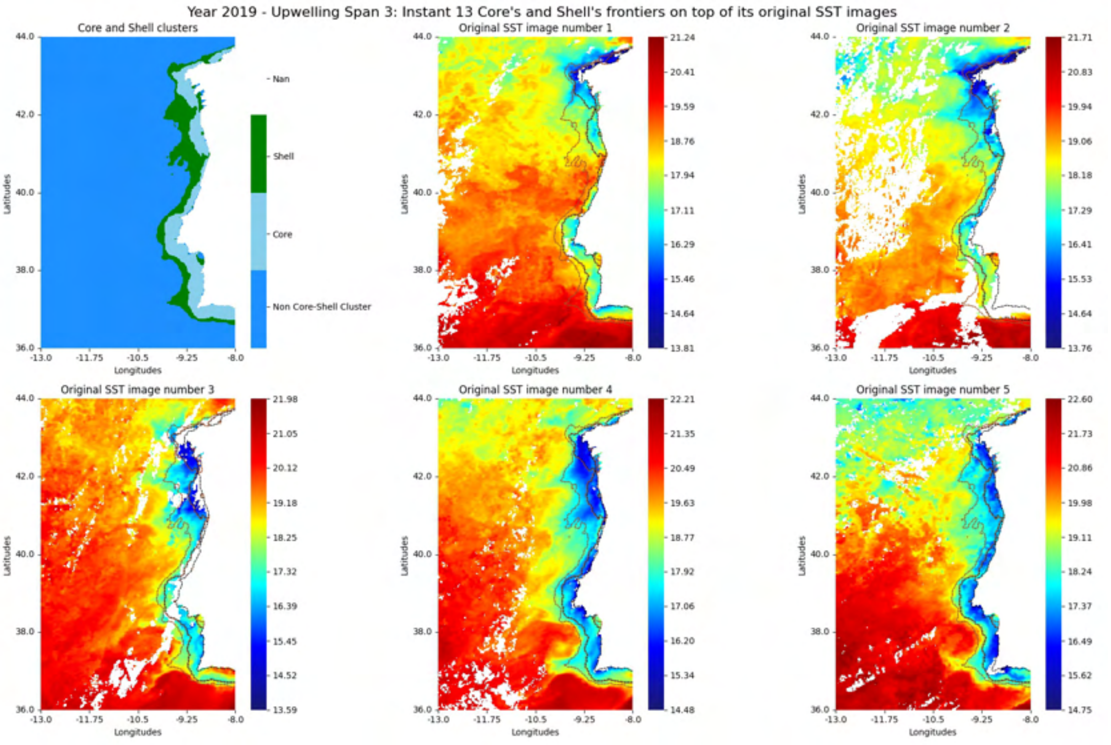
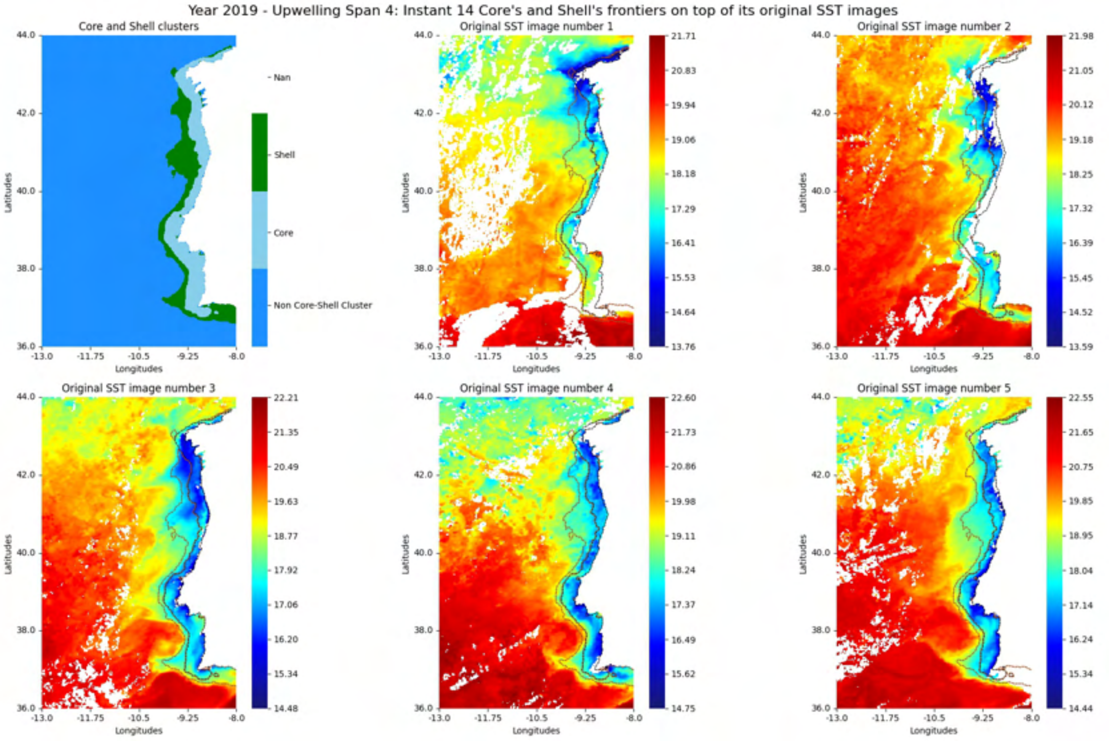


Figure I.139: Fronts regarding instant 13 of year 2019



Year 2019 - Upwelling Span 4: Instant 14 Core's and Shell's frontiers on top of its original SST images

Core and Shell clusters

Original SST image number 1

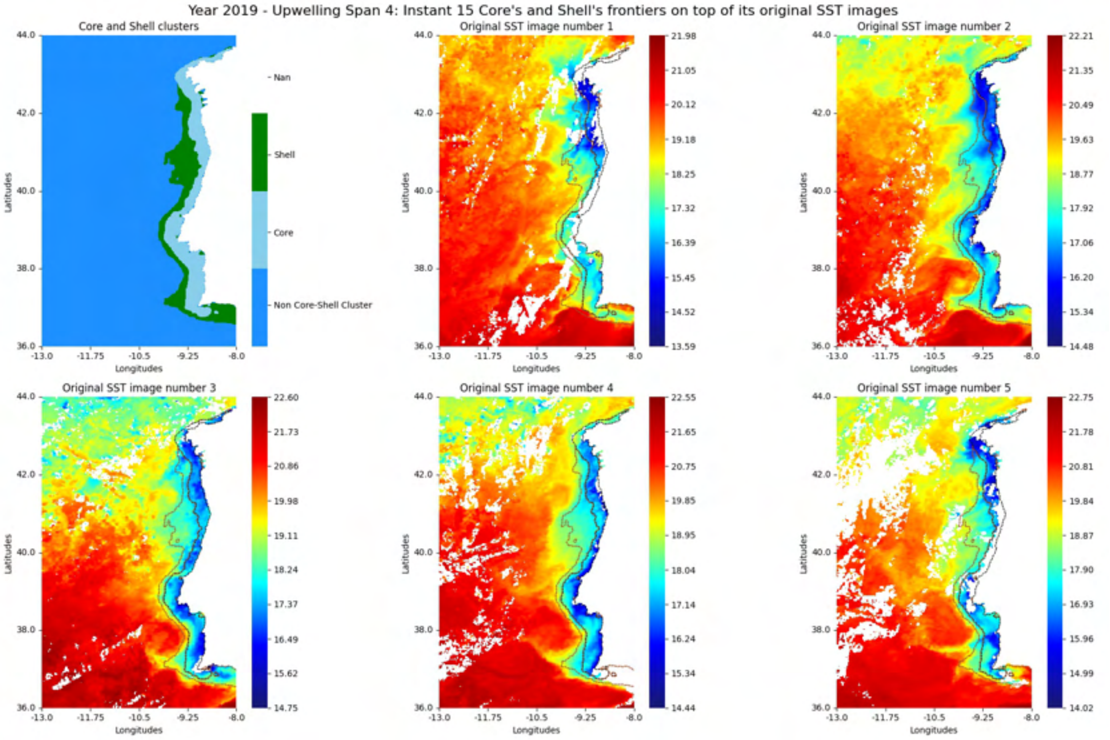
Original SST image number 2

Original SST image number 3

Original SST image number 4

Original SST image number 5

Figure I.140: Fronts regarding instant 14 of year 2019



Year 2019 - Upwelling Span 4: Instant 15 Core's and Shell's frontiers on top of its original SST images

Core and Shell clusters

Original SST image number 1

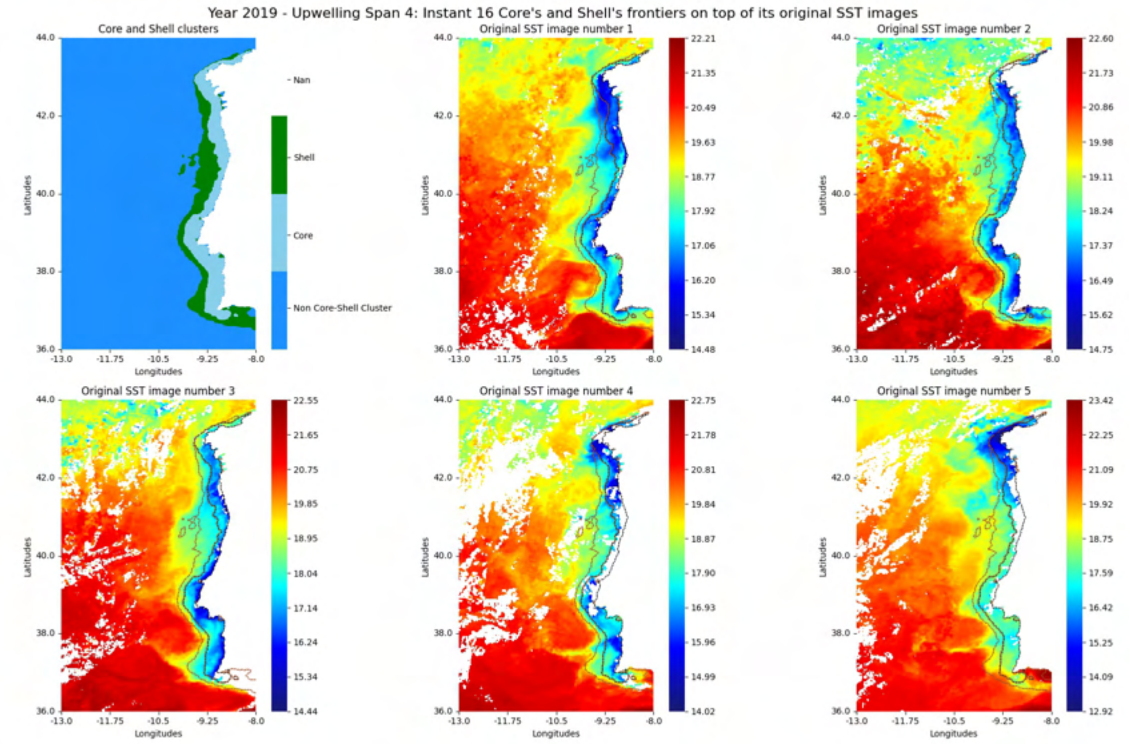
Original SST image number 2

Original SST image number 3

Original SST image number 4

Original SST image number 5

Figure I.141: Fronts regarding instant 15 of year 2019


Year 2019 - Upwelling Span 4: Instant 16 Core's and Shell's frontiers on top of its original SST images

Core and Shell clusters

Original SST image number 1

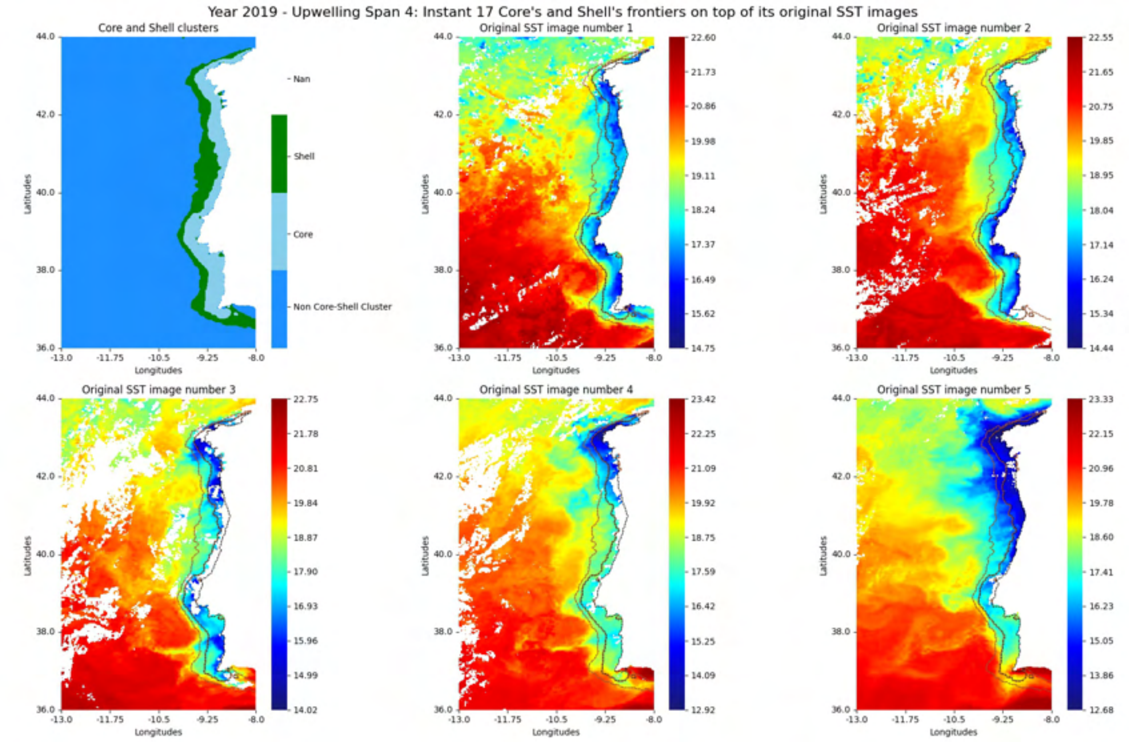
Original SST image number 2

Original SST image number 3

Original SST image number 4

Original SST image number 5

Figure I.142: Fronts regarding instant 16 of year 2019


Year 2019 - Upwelling Span 4: Instant 17 Core's and Shell's frontiers on top of its original SST images

Core and Shell clusters

Original SST image number 1

Original SST image number 2

Original SST image number 3

Original SST image number 4

Original SST image number 5

Figure I.143: Fronts regarding instant 17 of year 2019

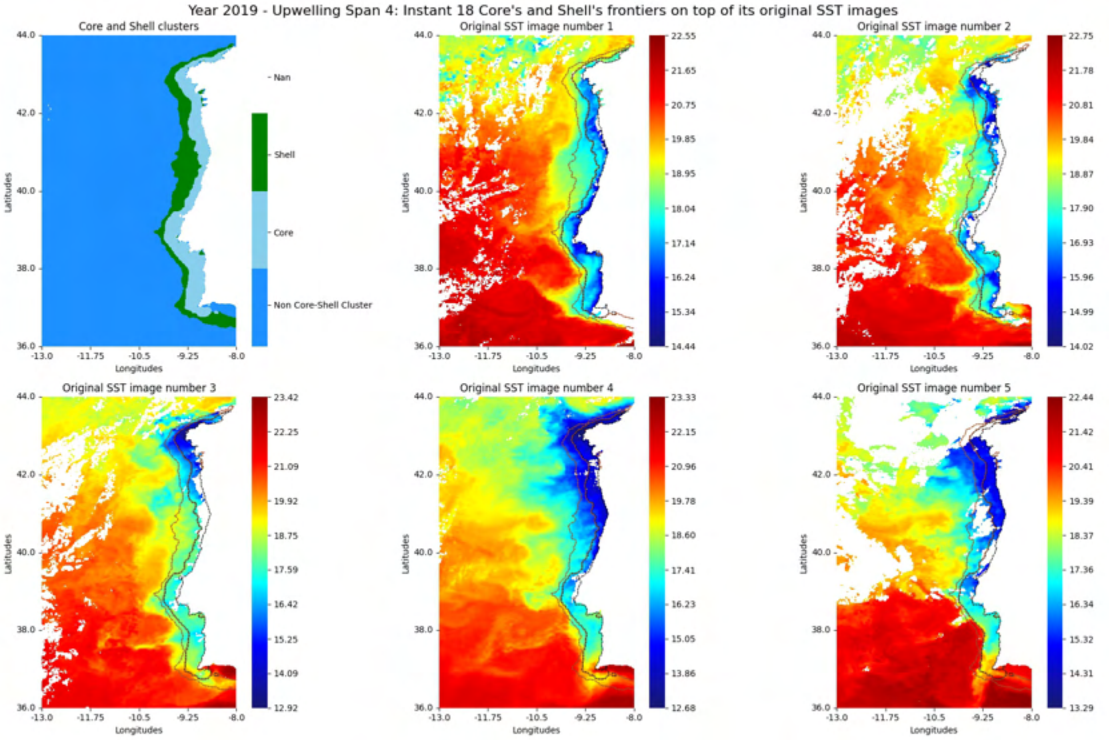


Figure I.144: Fronts regarding instant 18 of year 2019

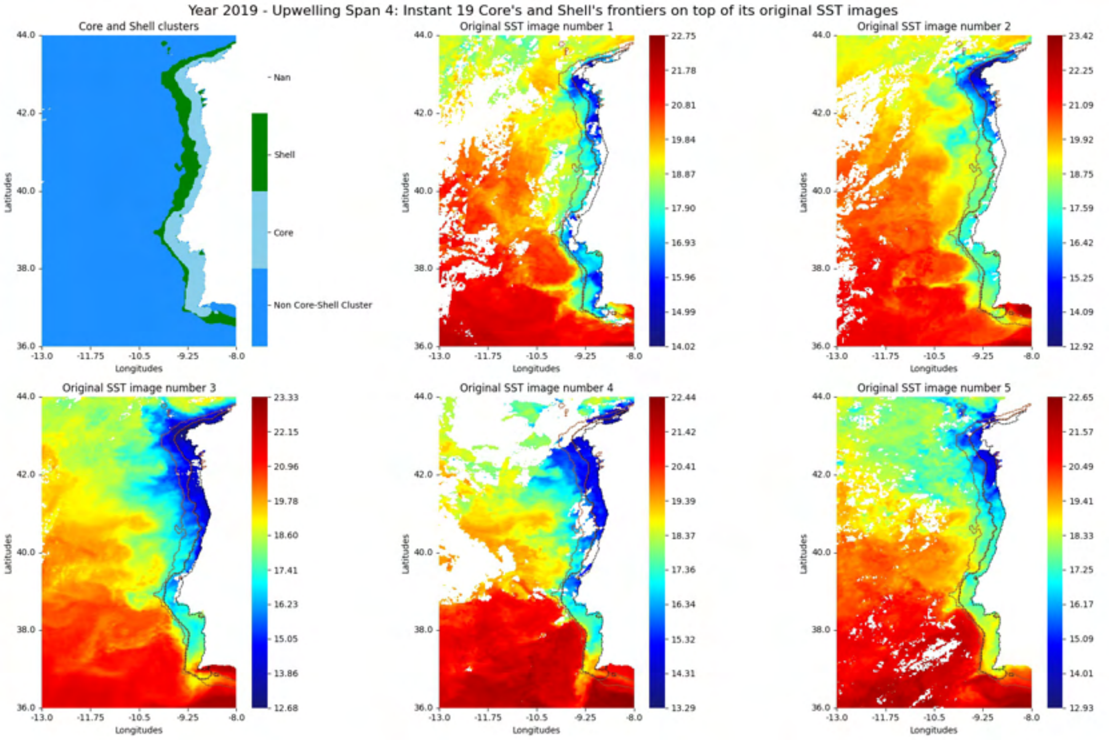
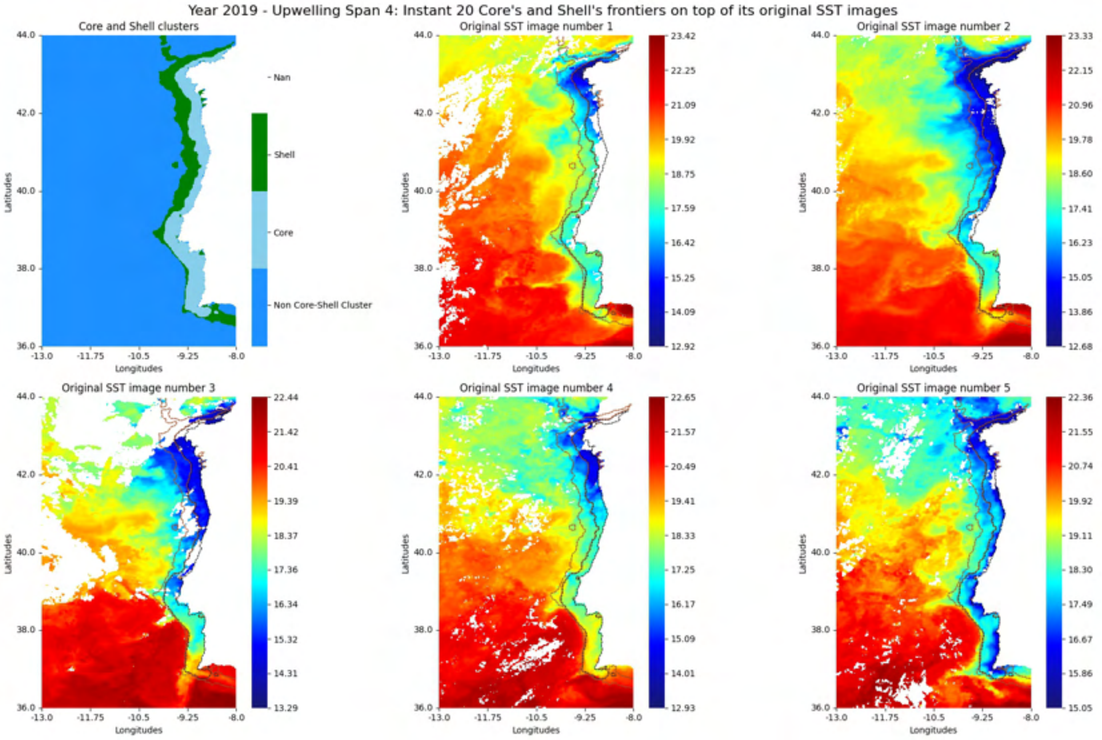


Figure I.145: Fronts regarding instant 19 of year 2019


Year 2019 - Upwelling Span 4: Instant 20 Core's and Shell's frontiers on top of its original SST images

Core and Shell clusters

Original SST image number 1

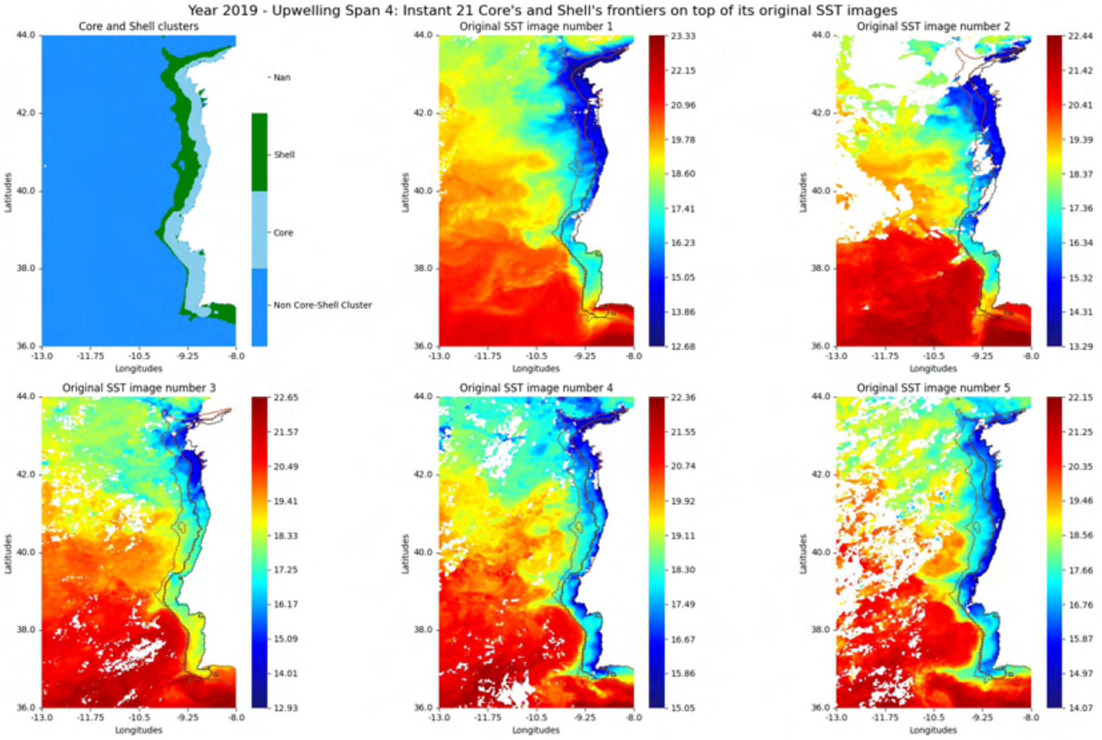
Original SST image number 2

Original SST image number 3

Original SST image number 4

Original SST image number 5

Figure I.146: Fronts regarding instant 20 of year 2019


Year 2019 - Upwelling Span 4: Instant 21 Core's and Shell's frontiers on top of its original SST images

Core and Shell clusters

Original SST image number 1

Original SST image number 2

Original SST image number 3

Original SST image number 4

Original SST image number 5

Figure I.147: Fronts regarding instant 21 of year 2019

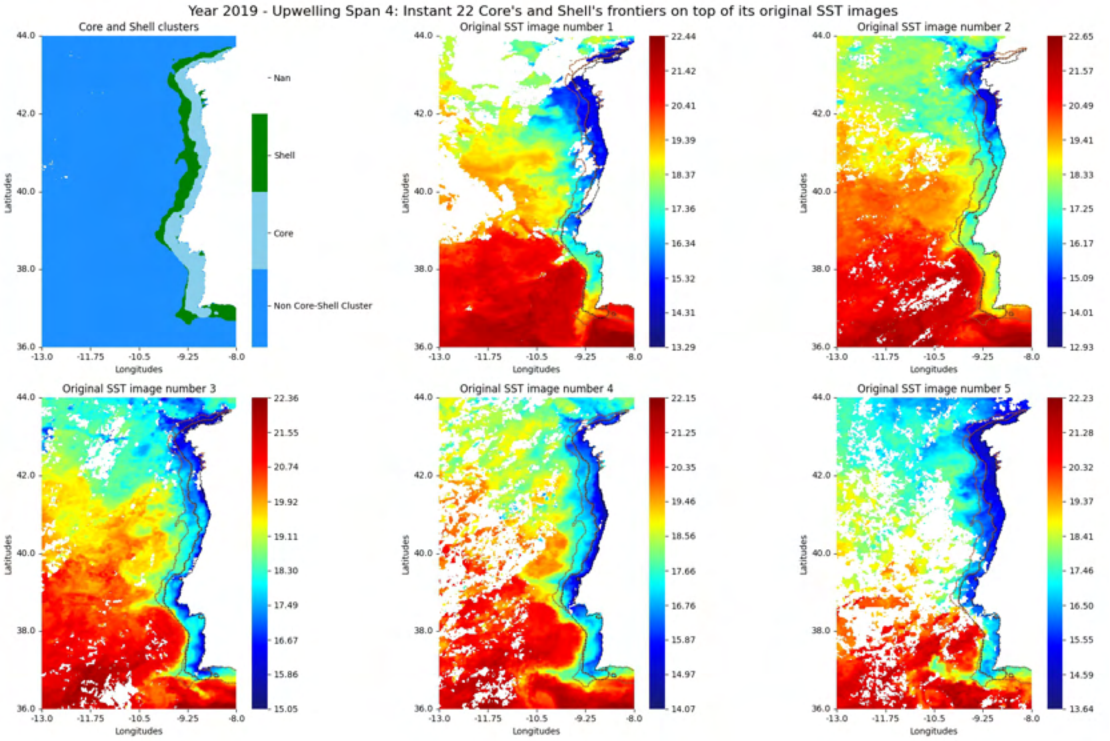


Figure I.148: Fronts regarding instant 22 of year 2019

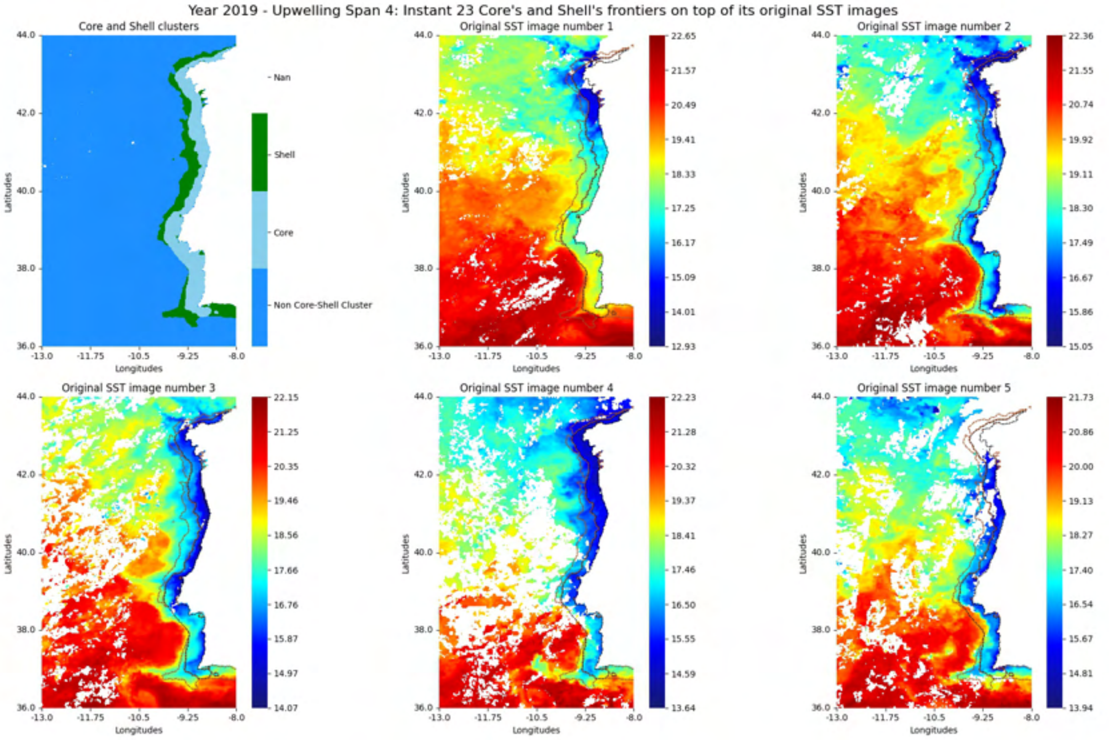


Figure I.149: Fronts regarding instant 23 of year 2019

II

Subsidiary Results

II.1 Distances of upwelling fronts to the coastline for the year 2007

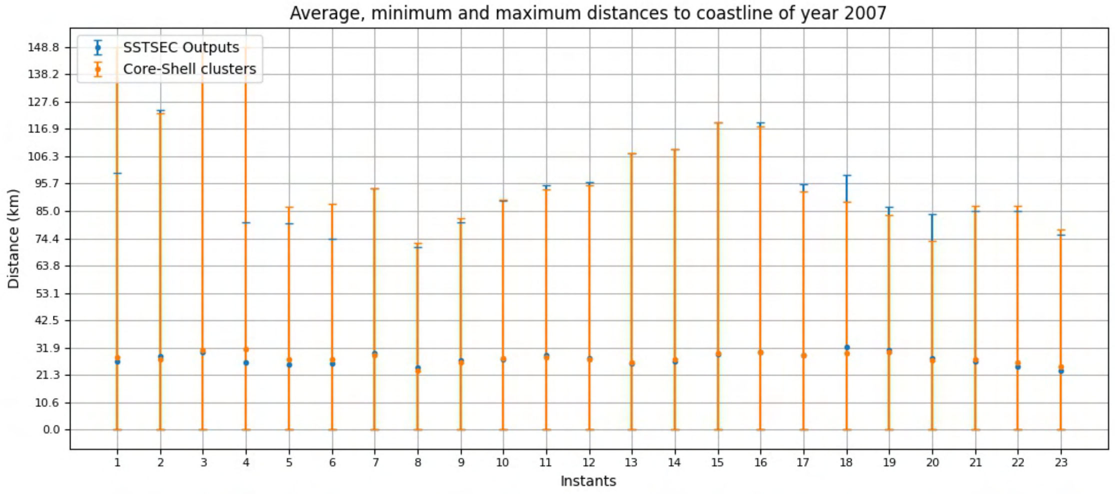


Figure II.1: Mean, minimum and maximum distances of upwelling fronts to the coastline for the entirety of the Portuguese coast in 2007

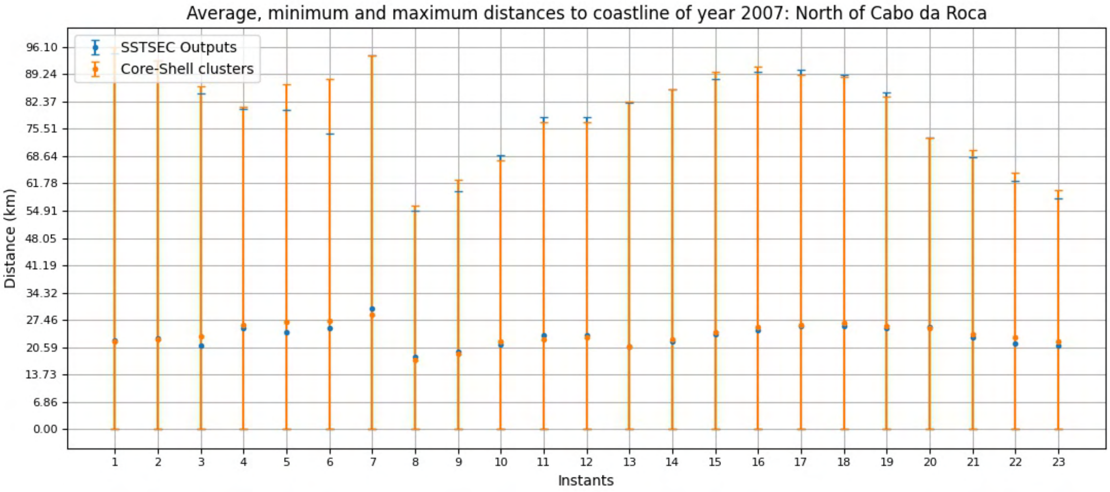


Figure II.2: Mean, minimum and maximum distances of upwelling fronts to the coastline for the region north of Cabo da Roca in 2007

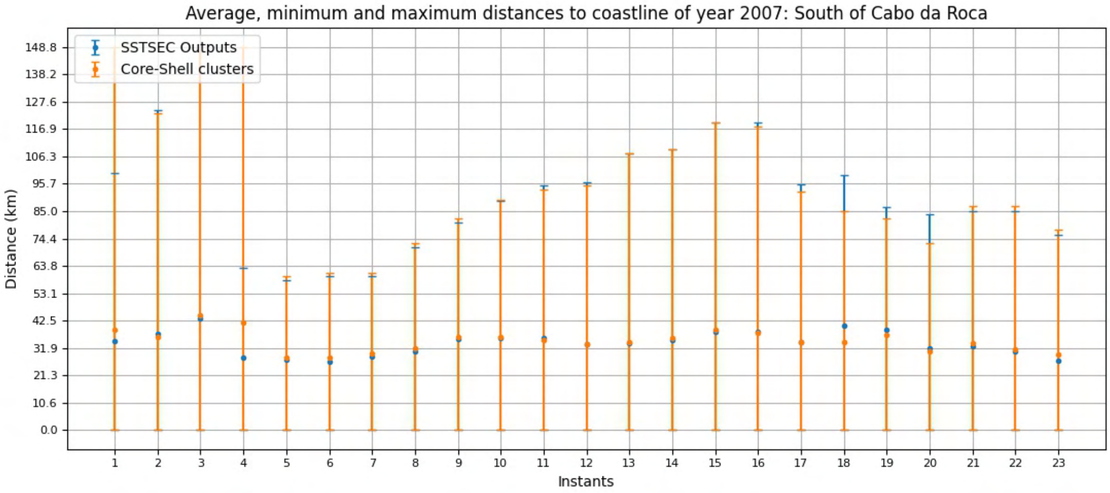


Figure II.3: Mean, minimum and maximum distances of upwelling fronts to the coastline for the region south of Cabo da Roca in 2007

II.2 Distances of upwelling fronts to the coastline for the year 2015

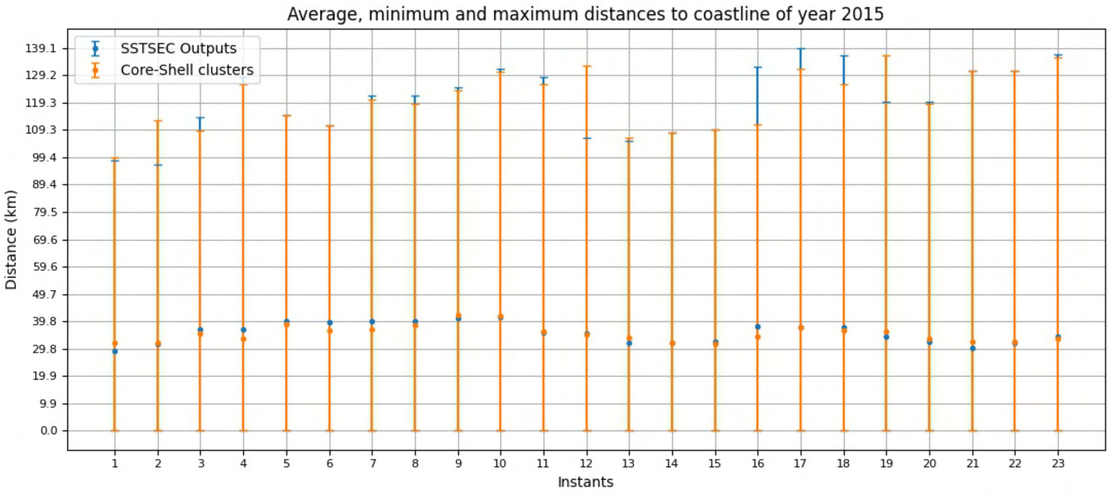


Figure II.4: Mean, minimum and maximum distances of upwelling fronts to the coastline for the entirety of the Portuguese coast in 2015

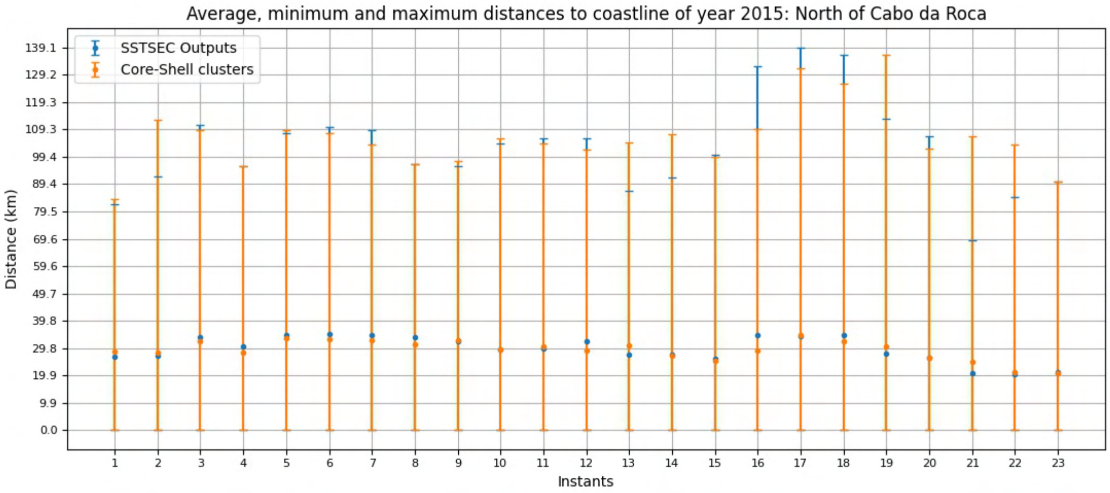


Figure II.5: Mean, minimum and maximum distances of upwelling fronts to the coastline for the region north of Cabo da Roca in 2015

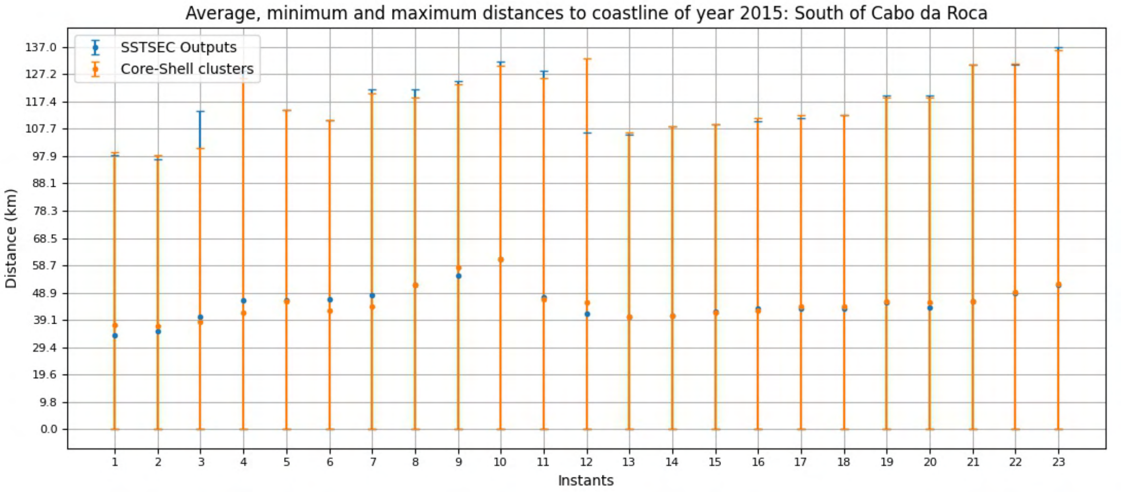


Figure II.6: Mean, minimum and maximum distances of upwelling fronts to the coastline for the region south of Cabo da Roca in 2015

II.3 Distances of upwelling fronts to the coastline for the year 2019

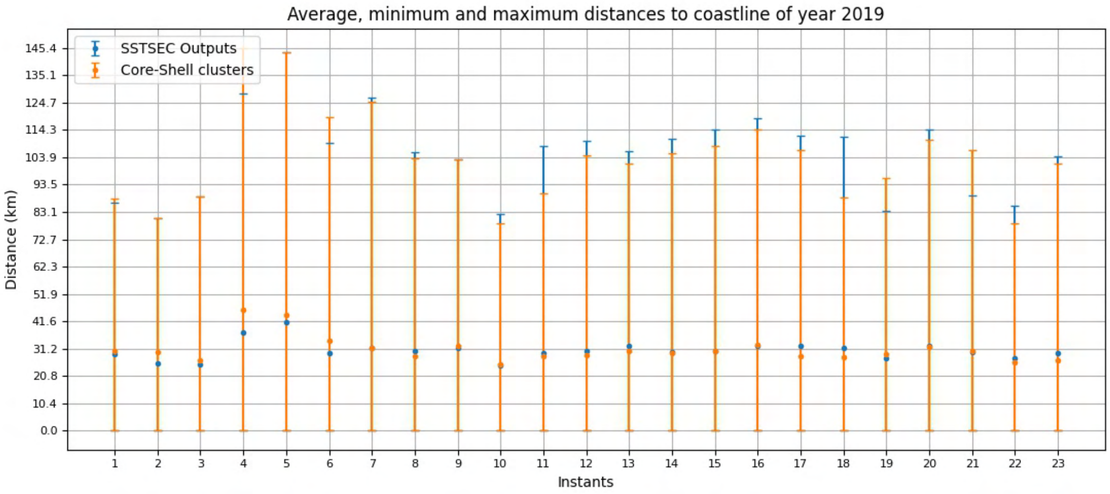


Figure II.7: Mean, minimum and maximum distances of upwelling fronts to the coastline for the entirety of the Portuguese coast in 2019

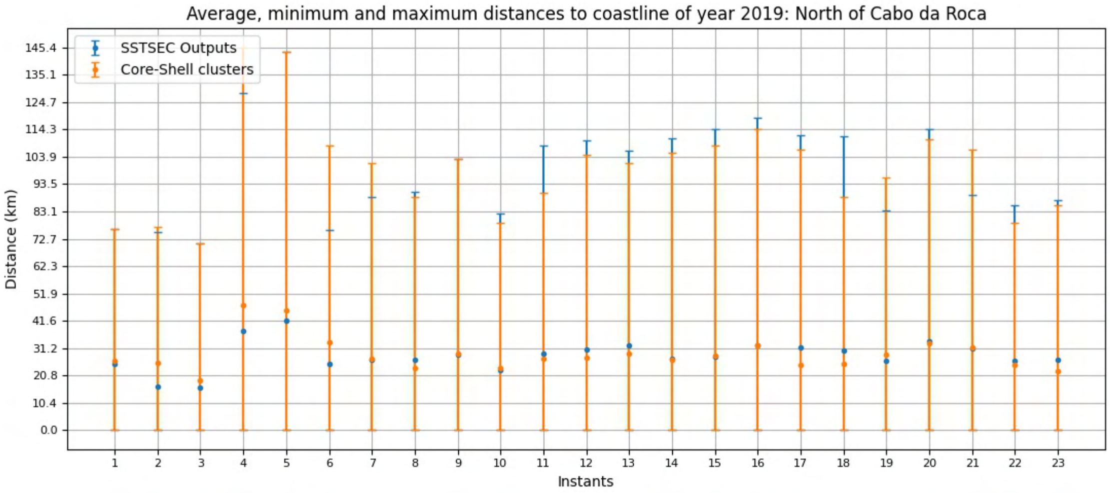


Figure II.8: Mean, minimum and maximum distances of upwelling fronts to the coastline for the region north of Cabo da Roca in 2019

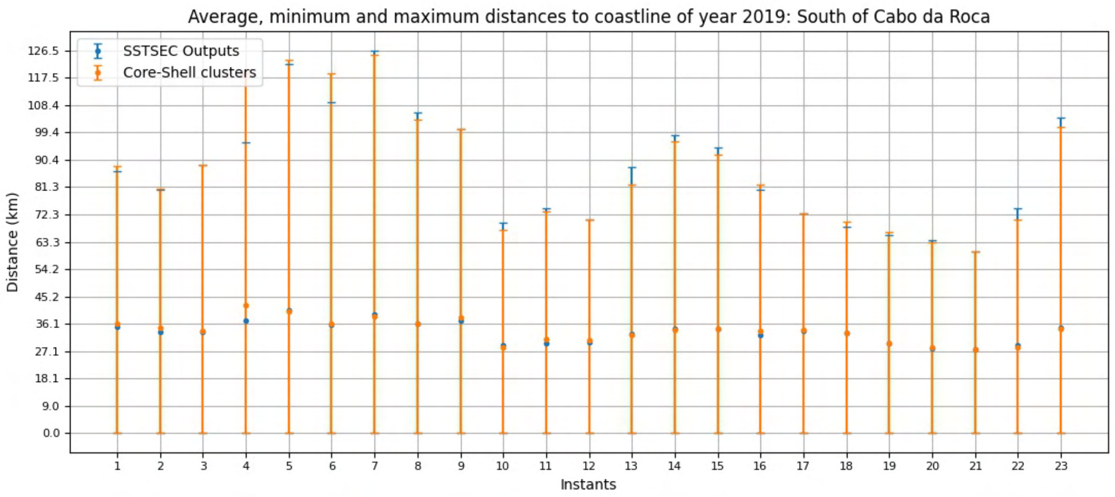
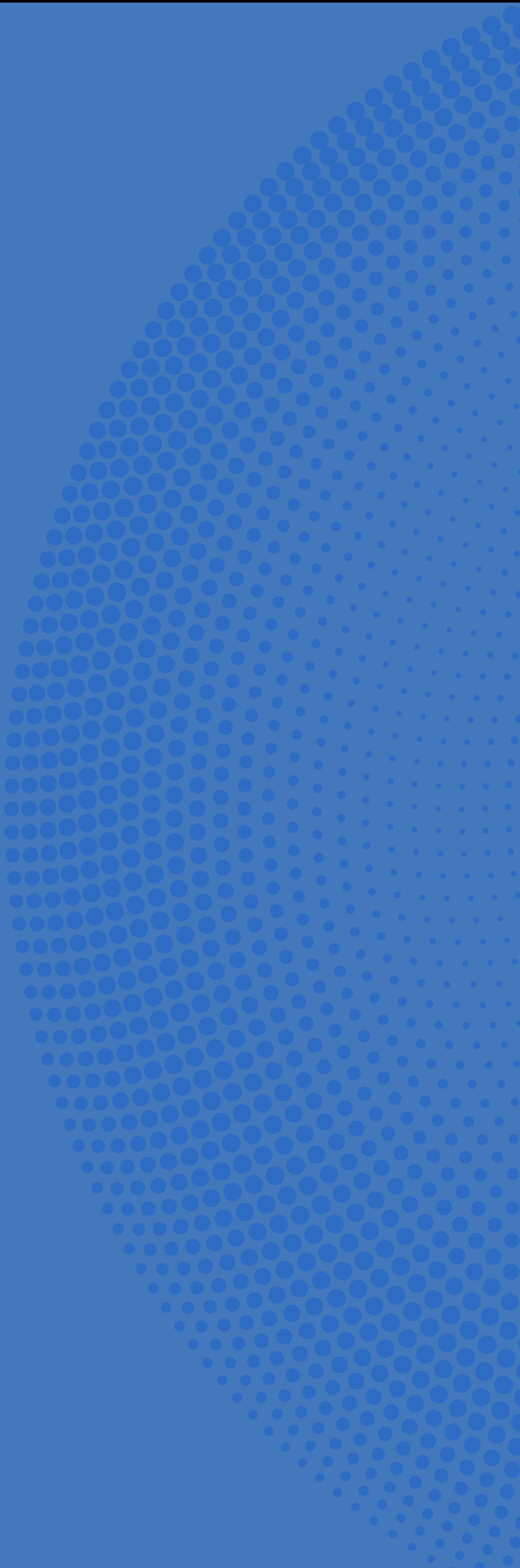


Figure II.9: Mean, minimum and maximum distances of upwelling fronts to the coastline for the region south of Cabo da Roca in 2019



2022

UNSUPERVISED SPATIO-TEMPORAL ANALYSIS OF COASTAL UPWELLING FROM SEA SURFACE TEMPERATURE IMAGES

Alexandre Martins

NOVA

NOVA SCHOOL OF